



HAL
open science

Exploring the potential of landmarks detection to reduce disorientation in pan-scalar maps

Quentin Potié

► **To cite this version:**

Quentin Potié. Exploring the potential of landmarks detection to reduce disorientation in pan-scalar maps. Geography. Université Gustave Eiffel, 2025. English. ⟨NNT : 2025UEFL2072⟩. ⟨tel-05479166⟩

HAL Id: tel-05479166

<https://theses.hal.science/tel-05479166v1>

Submitted on 27 Jan 2026

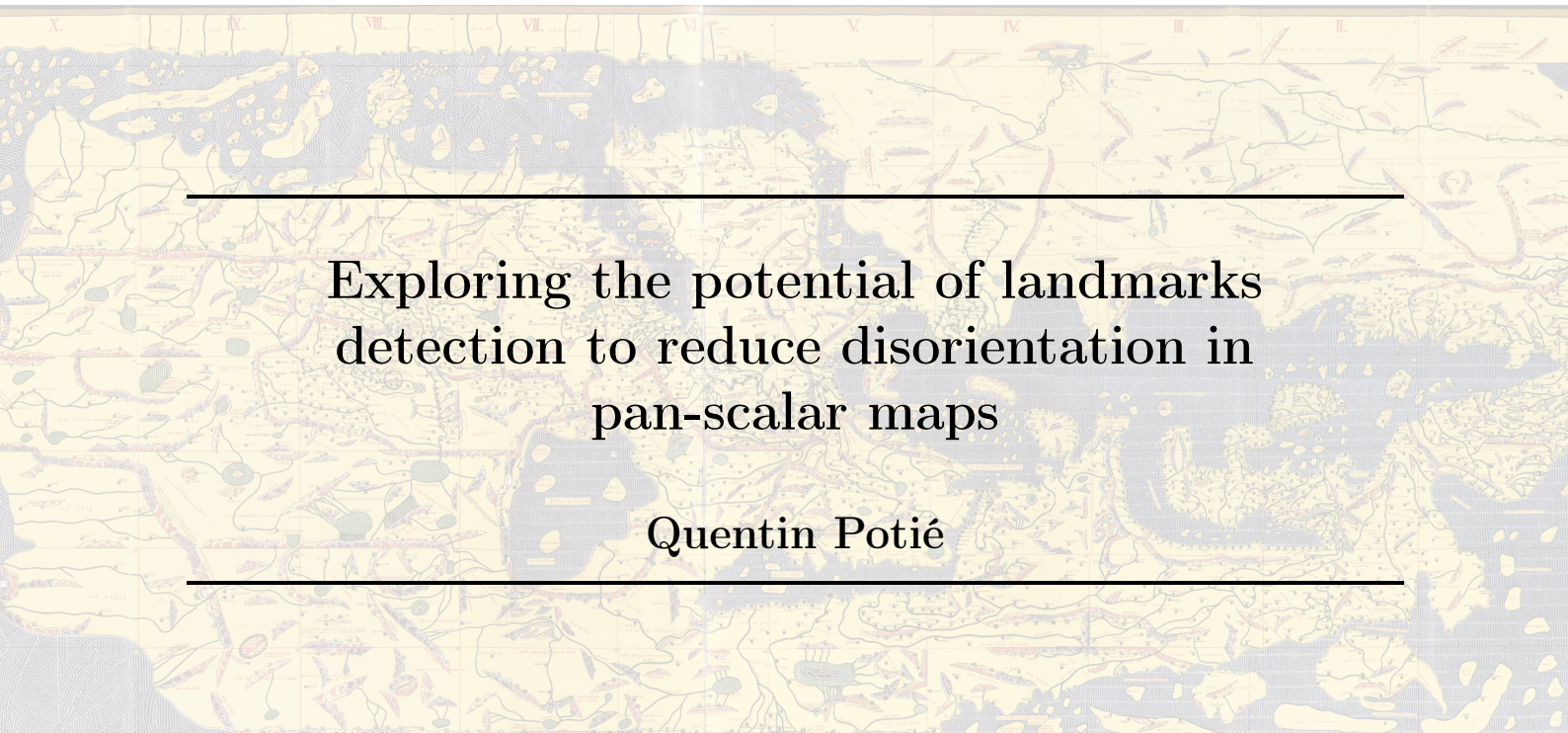
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Doctoral thesis from Université Gustave Eiffel
Doctoral school MSTIC
Specialty: Geographic information sciences



**Exploring the potential of landmarks
detection to reduce disorientation in
pan-scalar maps**

Quentin Potié

Publicly defended on June 26th, 2025

Jury

Pr Christophe Claramunt,	Arts et Métiers Institute of Technology	Chair of the jury, Reviewer
Pr Arzu Çöltekin,	University of Applied Sciences & Arts, Northwestern Switzerland	Reviewer
Pr Anne Ruaz,	Université Gustave Eiffel	Examiner
Mr Cyril De Runz,	Université de Tours	Examiner
Mr Guillaume Touya,	Université Gustave Eiffel	Supervisor
Pr William Mackaness	University of Edinburgh	Supervisor

« The unreal is more powerful than the real. Because nothing is as perfect as you can imagine it. Because it's only intangibles, ideas, concepts, beliefs, fantasies that last. Stone crumbles. Wood rots. People, well, they die. But things as fragile as a thought, a dream, a legend, they can go on and on. If you can change the way people think. The way they see themselves. The way they see the world. You can change the way people live their lives. »

— Chuck Palahniuk, *Choke*

Abstract

Pan-scalar maps are interactive digital maps that allow users to navigate through space at varying levels of detail and scales. While these maps are now commonplace, there is still room for design improvements. The LostInZoom (LiZ) project aims to enhance navigation in such maps by improving their content and display, and by exploring new interaction methods.

As part of the LiZ project, this thesis investigates the concept of pan-scalar landmarks—specific features or groups of features that remain salient or recognizable across multiple zoom levels, thereby supporting user navigation. Through a series of user experiments and AI training, this work explores methods for detecting such landmarks in order to provide feedback to cartographers and map designers on how and where their maps can be improved. In doing so, it also offers insights into the cognitive processes that govern how users distinguish landmarks from other map elements.

The thesis is structured around two main research axes. In the first axis, we classify landmarks in different categories, list them, and focus on specific categories. In particular, we conduct a perceptual study on ring roads and discuss how such experiments can help build better models. Additionally, we present an experiment aimed at detecting remarkable buildings in a database using a multimodal approach with a deep learning model.

In the second axis, we focus on a more global approach by training segmentation models (U-Net and Segformer) to detect areas of high anchoring potential (AHAPs) in map images. This involves creating a dedicated training dataset based on existing pan-scalar maps and developing a methodology for evaluating the performance of these models.

This work contributes to the growing body of research at the intersection of cartographic design, human-computer interaction, and cognitive science, and is novel in the use of neural networks to identify cognitively salient features in map images. More widely, it reaffirms the need for additional map design considerations when it comes to using maps in highly interactive environments.

❖ Keywords :

landmarks, visual cognition, map generalisation, pattern recognition, machine learning.

Résumé

Les cartes pan-scalaires sont des cartes numériques interactives qui permettent aux utilisateurs de naviguer à travers l'espace et les échelles. Bien que ces cartes soient désormais courantes, il reste encore des possibilités d'améliorations en matière de conception. Le projet LostInZoom (LiZ) vise à améliorer la navigation dans ces cartes en optimisant leur contenu et leur affichage, et en explorant de nouvelles méthodes d'interaction.

Dans le cadre du projet LiZ, cette thèse s'intéresse au concept de "landmarks pan-scalaires" — des éléments ou groupes d'éléments dans une carte qui sont saillants ou reconnaissables à plusieurs niveaux de zoom, et qui aident ainsi les utilisateurs à naviguer dans la carte. À travers une série d'expériences utilisateurs et d'entraînements de modèles d'intelligence artificielle, ce travail explore des méthodes pour détecter ces landmarks afin de fournir un retour aux cartographes et concepteurs de cartes sur les moyens d'améliorer leurs cartes. Ce faisant, ce travail offre également certains éclairages sur les processus cognitifs par lesquels les utilisateurs distinguent les landmarks des autres éléments cartographiques.

La thèse s'articule autour de deux axes de recherche. Dans le premier axe, nous classons les ancres en différentes catégories et nous nous concentrons sur des catégories spécifiques. En particulier, nous menons une étude perceptuelle sur les périphériques urbains (ring roads) et discutons de la manière dont ce genre d'expériences peuvent contribuer à la construction de meilleurs modèles. Nous présentons ensuite une expérience visant à détecter des bâtiments remarquables dans une base de données en utilisant une approche multimodale combinant des données tabulaires et des images via un modèle d'apprentissage profond.

Dans le second axe, nous adoptons une approche plus globale en entraînant des modèles de segmentation (U-Net et Segformer) pour détecter les ancres dans les images de cartes. Cela implique la création d'un jeu de données d'entraînement basé sur des cartes pan-scalaires existantes et la création d'une méthodologie pour évaluer la performance de ces modèles.

Ce travail contribue au corpus croissant de recherches à l'intersection du design cartographique, de l'interaction humain-machine et des sciences cognitives, et se distingue notamment par l'utilisation de réseaux de neurones pour identifier des éléments cognitivement saillants dans des images de cartes. Plus largement, il réaffirme la nécessité d'intégrer davantage de considérations de design cognitif dans la conception des cartes utilisées dans des environnements numériques hautement interactifs.

❖ Mots clefs:

points de repère, cognition visuelle, généralisation cartographique, reconnaissance de formes, apprentissage machine.

❖ Un résumé long, en français, est disponible à la fin du manuscrit.

Contents

Phd Context	15
Acknowledgements — Remerciements	17
Disclaimer on the use of AI in the research process	25
Introduction	27
I The Role of Landmarks in Pan-Scalar Maps	33
Chapter A. Literature review	35
A.1 Landmarks and disorientation	35
A.1.1 What landmarks are	35
A.1.2 Landmarks and spatial cognition	37
A.1.3 Landmarks in our thesis	39
A.2 Pan-scalar maps	42
A.2.1 What pan-scalar maps are	42
A.2.2 Disorientation in pan-scalar maps	45
A.2.3 Designing and building pan-scalar maps	50
A.2.4 Focus on map generalisation	56
A.3 Landmark detection in maps	60
A.3.1 Human perception of the map: from perception to meaning	60
A.3.2 Pattern detection	64
A.3.3 Conclusion on landmark detection	68
A.4 Conclusion of the chapter	68
A.5 Thesis orientation	69
A.5.1 Problems	69
A.5.2 Overview of the experiments	69
Chapter B. Preliminary studies	71
B.1 What are the pan-scalar landmarks in contemporary pan-scalar maps?	71
B.1.1 Methodology	71
B.1.2 Results	72
B.1.3 Conclusion	74
B.2 What are the optimal densities for pan-scalar landmarks and visual cues in maps?	74
B.2.1 Literature review	75

B.2.2	Minimum visual cues for good orientation	76
B.3	Pan-scalar map interfaces evolution: a two-decade perspective	81
B.3.1	Context	81
B.3.2	Map design evolution	82
B.3.3	Zooming and panning methods	83
B.3.4	Conclusion	88
B.3.5	Bonus: Modelling navigation tasks in pan-scalar maps	89
B.4	Conclusion of the chapter	91
 II Focusing on the Detection of Specific Landmarks		93
 Chapter C. When is a ring road a ring road? A perceptual study		95
C.1	Introduction	95
C.1.1	Ring roads potential	95
C.1.2	Ring road preliminary definition	97
C.1.3	Literature on ring road detection	98
C.2	Perceptual study	100
C.2.1	Objectives of the study and hypothesis	100
C.2.2	Methodology of the experiment	100
C.2.3	Analysis of the results	105
C.2.4	Limits of the survey	108
C.3	Measuring ringroadness	108
C.3.1	Characterization properties and congruence	108
C.3.2	Ring road ontology	111
C.4	Conclusion of the chapter	114
 Chapter D. Multimodal learning for remarkable buildings detection		117
D.1	Motivations	117
D.2	Project history	118
D.2.1	Fieldwork student project	118
D.2.2	Research initiation project	118
D.2.3	Master thesis internship	121
D.3	About remarkable building detection	121
D.4	About multimodal learning	123
D.5	Experiment	123
D.5.1	Main idea and objectives	123
D.5.2	Building an experimental dataset	124
D.5.3	Model implementations and results	130
D.5.4	Discussion	143
D.6	Conclusion of the chapter	145
 III Towards the Detection of Areas of High Anchoring Potential in Map Images		149
 Chapter E. AHAPs – Concept and preliminary experiments		151
E.1	Introduction to AHAPs	151
E.1.1	Definition	151

E.1.2 Hypothesis	152
E.1.3 Motivations	153
E.1.4 Related concepts	154
E.1.5 Drawing as data annotation	155
E.1.6 Preliminary experiment with Gribouillon	156
E.2 Preparing an evaluation dataset	158
E.2.1 The challenge of evaluation	158
E.2.2 Selecting geographic areas for the dataset	158
E.2.3 Annotation protocol	159
E.2.4 Output and first analysis	160
E.2.5 Establishing a common truth with STAPLE	162
E.3 Measuring inter-rater reliability	165
E.3.1 Indicators	165
E.3.2 Results and discussions	166
E.4 Conclusion of the chapter	168
Chapter F. Detecting areas of high anchoring potential with deep learning	171
F.1 About image segmentation	171
F.1.1 Definition	171
F.1.2 CNNs and U-Net	171
F.1.3 Transformers and Segformer	173
F.2 Training dataset	174
F.3 Experiment 1: U-Net	176
F.3.1 Dataset pre-treatment	176
F.3.2 Implementation and Training	177
F.4 Experiment 2: Segformer	179
F.4.1 Dataset pre-treatment	179
F.4.2 Implementation and training	179
F.4.3 Challenges and workarounds	180
F.5 Predictions and metrics	181
F.6 Discussion on the results	182
F.7 Ideas for further experiments with AHAPs	186
F.8 Limitations	187
F.9 Conclusion of the chapter	188
Thesis Conclusion	191
Glossary	199
Bibliography	201
Publications	213
Appendices	215
A Ring road survey part 1	215
B Ring road survey part 2	221
C Ring road survey part 3	227
D Metrics monitored during MLP model training	230

E	Metrics monitored during CNN model training	236
F	Multimodal model summary and monitored metrics	241
G	AHAP segmentation evaluation	247
	Résumé substantiel (French comprehensive summary)	259

Phd Context

This thesis project could not have come to fruition without a stimulating work environment. It is the result of three years and a half of work at LASTIG, one of the research laboratories on geographic information within the French National Institute of Geographic and Forest Information (IGN). As the national mapping agency of France, IGN has been producing both paper and digital topographic maps since 1940, when the Army Geographic Service (SGA) became a civilian organization. At the time of writing, approximately 70 researchers work at LASTIG on a variety of topics ranging from satellite image analysis to 3D modeling of urban spaces, with artificial intelligence playing an increasingly significant role.

Specifically, this thesis is part of the Lost In Zoom (LIZ) project, funded by an ERC grant (2021-2026) and led by Guillaume Touya, one of the thesis supervisors. The LIZ project focuses on the design of pan-scalar maps. The team studies these maps and seeks to establish innovative new theories to improve them, with particular attention to human perception, map construction, and user interfaces. Notably, the project delves into the cognitive aspects of users: how they perceive pan-scalar maps and how they construct mental maps. Additional experiments examine user attention and pre-attention mechanisms, exploring which elements of the map draw focus and which movements users perform when completing specific tasks.

While the work presented in this thesis is primarily my own, many of the ideas emerged through discussions with other members of the LIZ team, whether during our weekly meetings or informal conversations over coffee. Furthermore, the visibility and impact of the LIZ project facilitated numerous connections with other researchers, enriching my work through conferences and collaborations.

Acknowledgements — Remerciements

Some parts are in English, et certains passages sont en français, selon.

For readers who do not know me and who may come across this manuscript, whether through research or by chance. To you, welcome! Thank you for having the curiosity and courage to explore and learn. I hope you will find here a spark of inspiration, a few fresh ideas, and, ideally, even a touch of entertainment. But before that, here are a few words about the people who have shaped my work.

Au même titre que la majorité des travaux scientifiques, cette thèse a nécessité la participation directe ou indirecte d'un grand nombre de personnes dont l'ensemble des noms ne tiendrait pas sur la première page. C'est leur implication et leur soutien qui m'ont permis d'aller au bout du travail le plus ambitieux de ma vie jusqu'à présent. À ce titre, j'ai à la fois l'envie et le devoir de les remercier. Je m'excuse d'avance pour tous ceux que je vais oublier de mentionner ici — il y en aura forcément, tant j'ai eu la chance d'être bien entouré.

First of all, my heartfelt thanks go to my two supervisors, Guillaume Touya and William Mackaness, for guiding me throughout this journey and inspiring me through their approach to science, their cheerful personalities, their constant sense of wonder, and, of course, their sense of humor. These two don't just do good science — they make you want to do science with them. They elevate the projects and ideas of those around them, and their boundless motivation and infectious enthusiasm make every step of the journey more exciting. I couldn't have imagined better supervision.

Merci Guillaume de m'avoir fait confiance et de m'avoir donné l'opportunité de m'intéresser à l'intelligence artificielle malgré mes compétences encore balbutiantes sur le sujet, ce qui m'a permis d'apprendre tant de choses. Merci également de m'avoir donné la chance, à travers cette thèse, de rencontrer tant de personnes plus intéressantes les unes que les autres. Merci d'avoir trouvé les mots pour m'encourager et me rassurer quand c'était nécessaire, et d'avoir su trouver les mots, difficiles, pour me guider lorsque je m'aventurais dans de mauvaises directions. Merci enfin pour l'énergie que tu consacres aux nouvelles générations de scientifiques, et bravo pour la juste posture que tu sais adopter, tant dans l'encadrement du quotidien que lors des conférences.

Thank you, William, for your patience, your courage, your humility, and your wise advice. Although our interactions were less frequent, they were no less meaningful and encouraging to me. I am glad and proud to be one of your last students, and I will do my best to be worthy of it in the years to come. Be assured that your influence will continue to resonate long after your retirement, through the many people you have undoubtedly inspired throughout your career. Take a well-deserved rest — But not too much, just enough!

Then I would like to thank the wider GIS research community, which definitely seems to be one of the more welcoming ones. I am grateful to all the researchers, of all ages, who made the conferences I had the chance to attend so lively, enjoyable, and instructive: Monika Sester, Robert Weibel, Yu Feng, Anthony C. Robinson, and Arzu Çöltekin, just to name a few pillars of the field.

Talking about conferences, I want to give special thanks to Merve Keskin for

the friendship we built during these little interludes of life, for her attentiveness, originality, and dedication to the community — and also to Lars De Sloover for our brief but hilarious encounter; Italian-style flirting clearly holds no secrets for us anymore.

I would also like to acknowledge the anonymous reviewers of my articles, who took the time to read my work and provide detailed, insightful feedback, and the examiners of my thesis who, not content with merely reading and commenting on my scribbles, also stayed awake during my defence and even asked some very pertinent questions.

Then I want to warmly thank the science communicators and teachers who dedicate themselves to sharing knowledge freely online, and who helped me and others to reach our goals. In particular, the following YouTube channels have inspired and supported me throughout this thesis: *MachineLearnia* for introducing me to machine learning (in French), *ZEISS arivis* for its teachings on deep learning and computer vision (in English), and *Monsieur Phi* (in French) for showing me that scientific rigor can (and, in my opinion, should) be applied to philosophy, even though this is too rarely the case in French education.

Merci à celles et ceux qui ont partagé mon quotidien à l'ENSG ou à l'IGN et qui ont contribué à la bonne ambiance générale.

Merci aux collègues doctorants et professeurs qui n'ont cessé de m'étonner chaque jour. Sébastien pour ton accueil chaleureux et pour tes blagues, et Cécile pour ta sagacité et les conseils que tu m'as fournis. Merci à vous deux de brandir vos valeurs comme vous le faites et de motiver, directement ou indirectement, votre entourage à passer à l'action. Merci Emmanuel F. pour ton accueil également, pour tous les services, pour ton insatiable curiosité, la finesse de ton humour, ton analyse historique et géopolitique, et bien sûr pour Ralph, la contre-enquête, et toutes ces autres références dont tu fais l'éloge avec brio. Merci Nicolas d'avoir partagé nos repas, de m'avoir emmené aux archives trouver de belles cartes anciennes tout en me racontant leurs origines. Merci Victor et Vincent de m'avoir fait signe de vous rejoindre au Descartes lors de ma première semaine à l'école, et ensuite pour votre bonne humeur dans l'organisation des cours de Git et de Python qui m'ont rendu le travail très agréable. Merci Samuel pour le Batz et toutes les aventures qui en ont découlé, et Bruno, pour nous y avoir amenés en 2024, et pour les échanges musicaux et philosophiques que nous avons eus à cette occasion. Merci, d'ailleurs, Lâman pour ces discussions sur le fromage, la photographie et les bandes dessinées, pour ton écoute, et pour ce si beau déguisement que tu nous as concocté. Merci Florent et Melvin pour l'ambiance que vous avez mise au labo et pour votre motivation et votre ouverture d'esprit lors de l'organisation du LYRD. Merci Paul B. et Paul C. pour les parties de ping-pong endiablées. Merci Raphaël d'être venu jusqu'à Paris pour ma soutenance, et merci Chahine, Solenn, Helen, Damien, Laurent, Teng, Alexane, Emmanuel C. et tous les membres de l'IGN avec qui j'ai eu des conversations enrichissantes pendant ces trois ans et dont je ne saurais citer tous les noms.

Merci aussi à celles et ceux qui travaillent indirectement à la réussite de la recherche et de l'enseignement. En particulier Nathalie et Christine pour le support administratif que vous m'avez fourni, Léon pour les multiples services, et Steve, Safia et Louisa pour avoir pris le relais de sainte Jeanine avec brio à la cafétéria (un

sacré défi).

J'en profite également pour adresser mes remerciements au personnel du parc régional naturel des Alpilles, qui m'a accueilli pendant plus d'un an, avant mon doctorat, alors que je faisais mes premiers pas avec les SIG. Je tiens à les saluer pour leur courage dans un monde où lutter pour le respect de l'environnement peut parfois sembler sisyphéen.

J'en viens maintenant à mes partenaires de travail, qui ont connu mes états d'âme et m'ont accompagné dans les hauts comme dans les bas.

Merci Bérénice pour la passion avec laquelle tu communique, pour l'idée géniale de comparer les cartes à la BD et au cinéma, pour les sessions roller-code (même quand on code pas), et pour tes post-it qui m'ont donné le courage d'aller au bout. Merci Laura pour la raclette, ta bonne humeur, ton engagement dans le labo et pour les élèves, et pour ta passion contagieuse pour les Pokémon. Merci Justin pour ton flegme apaisant, pour tes recommandations de films, pour ton sourire en coin quand on fait des blagues le midi, et évidemment pour Mapdraw et pour toute l'aide que tu nous as apportée sur nos projets. Merci Jérémie d'être toujours si enjoué, et pour ton engagement envers la communauté open source. Merci Paul et Vanessa pour la fraîcheur que vous avez apportée à notre équipe de recherche, par vos esprits autant enjoués qu'affûtés, et pour nos sorties estivales de fin de journée si importantes pour décompresser. Et puis merci Éric L. de m'avoir supporté comme directeur de stage pendant six mois, d'avoir apporté ton enthousiasme au projet et à l'équipe, de nous avoir partagé tes connaissances en plantes carnivores, et d'avoir oublié ton téléphone chez les militaires — je me souviendrai longtemps de cette excursion !

Enfin, un grand merci à Azelle et Maïeul pour toutes ces discussions que nous avons eues pendant les trois années où nous avons partagé le même bureau. Grâce à vous, je m'y suis senti chez moi, et j'étais content de me lever pour vous y retrouver, même le matin.

Merci Azelle, pour ton écoute, ton humour et ta bienveillance quotidienne, pour ta rigueur et ta précision dans les mots et les concepts, et pour toutes tes aides précieuses, qu'elles soient personnelles, scientifiques ou administratives. Merci de nous avoir fait confiance, à Maïeul et à moi, et merci pour ton rire si communicatif. Merci également pour ton template LaTeX qui m'a enlevé une sacrée épine du pied. Notre séjour à Vilnius est gravé dans ma mémoire, et je suis content d'avoir pu partager ce moment avec toi. Je te souhaite toute une quantité de belles et surprenantes choses pour les années qui t'attendent !

Merci Maïeul, pour ces années de soutien mutuel, pour ces rigolades, ces moments de partage honnêtes et forts, pour les bières et frites du Descartes, pour tes lasagnes maison, et pour les parties d'échecs impossibles à gagner. Notre connivence m'a fortement aidé à tenir jusqu'au bout, et je n'imagine pas ce qu'aurait été ma thèse sans toi. J'espère réussir un jour à te suivre en randonnée, et revoir ce bout de Bretagne que tu m'as présenté. D'ici là, je te souhaite également tout le bonheur du monde pour la suite !

En dehors de la sphère professionnelle, je veux également remercier toutes les personnes qui m'ont soutenu, qu'il s'agisse de rencontres fugaces ou de soutiens de longue date. Dans le plus grand désordre :

Merci Morgan pour ta sensibilité, ton entrain débordant, nos fous rires à l'entraînement,

la connivence que nous avons construite en quelques années, et pour m’avoir initié aux subtilités du poker. Merci Julie pour ta franchise et ton énergie, pour les bons repas que tu nous as cuisinés, et pour les cours de renforcement musculaire. Et merci à tous les deux pour votre ouverture d’esprit et pour ces soirées “jeux. . . mais finalement on discute”, qui comptent parmi mes préférées. Merci Apoline pour ton aide sur les choix de vêtements, tes prises de position fortes, tes éclats de rire, les tests de personnalité, et pour m’avoir considéré à la fois comme une amie avec qui on peut partager et comme un mec qu’il faut critiquer. Merci Léo pour les cours de rattrapage de maths en prépa, pour m’avoir fait découvrir la musique classique et m’avoir invité à tes concerts, pour ton rire éclatant, et pour héberger sans relâche les Grandes Rencontres Cinématographiques Internationales d’Orsay, de la Vallée de Chevreuse, et de tout l’Essonne. Merci Lisa pour tes sourires entraînants, tes questions sincères qui subliment les gens, et pour tous ces beaux moments que tu as organisés à la colloKB. Merci Bastien de m’avoir accueilli et épaulé durant ma première vraie expérience professionnelle, d’avoir d’abord pris cette place de mentor, puis de collègue, et enfin d’ami. Merci Adrien pour notre trajet Toulouse-Paris, notre aventure au parc des Beaumonts, et pour la confiance que tu m’as accordée dès notre rencontre. Merci Maro pour ton soutien dans le sprint final et pour ce fameux : « Tu as testé de séquencer ? » qui m’a permis de finaliser ma dernière expérience. Merci Jean-Mi pour les conduites aux matchs et pour être venu à ma soutenance. Merci Swan de m’avoir ouvert les portes de la programmation, et pour tes streams divertissants. Et puis merci Mathilde de m’avoir parlé pour la première fois de géomatique dans ce restaurant en 2019.

On respire.

Merci Marc, pour ta curiosité, pour les cours de musique, pour nos discussions sincères, et pour l’énergie que tu investis à rendre notre groupe de copains et copines heureux. Merci Jess d’avoir excusé mes dérapages, privilégié notre amitié, et de m’avoir offert une tasse Disney. Merci Juliette pour la confiance que tu sembles m’accorder, pour les draps propres, et pour m’avoir dit : « C’est simple, il suffit sûrement de concaténer des couches épaisses issues d’un CNN d’un côté et d’un MLP de l’autre, et de coller un perceptron en sortie pour déterminer la classe », un vocabulaire qui me paraissait très cryptique à l’époque. Merci Sandy pour ces discussions à des moments charnières de nos vies, pour l’amitié que nous avons construite, et pour tes encouragements. Merci Mike pour ton calme, ta vision rafraîchissante des festivals, et pour nos discussions sur le game design. Merci Jojo d’avoir conservé cette personnalité et cette innocence unique après toutes ces années. Merci Aymeric d’avoir conservé cette personnalité et ce flegme unique après toutes ces années. Merci Cindy d’avoir conservé cette personnalité et ce côté apaisant unique après toutes ces années. Merci Tereza pour tes questions sincères et l’intérêt que tu portes encore à notre amitié après plus de douze ans. Merci Janou-du-bout-du-monde pour la bièraubeurre réussie, la soirée *Detroit: Become Human*, et les nouvelles venues d’au-delà des océans. Merci Elisa de t’occuper si bien de ma mamie, pour ton humour, ton courage, et de m’avoir partagé les difficultés que rencontre ton pays. Merci Villy de m’avoir rassuré quand je me suis fait mordre par une chauve-souris, pour ton accueil à Sofia, et pour nos conversations qui me prouvent la confiance que tu m’accordes. Merci Sophie d’être si rayonnante, pour ton rire communicatif, et de dormir si bien la nuit. J’ai hâte de rencontrer l’enfant, puis l’adulte, que tu deviendras. Et merci

Clément d’avoir été mon seul ami en primaire après le départ de Yoma, et pour tous ces moments réconfortants passés devant les jeux, dans la rue, ou dans un restaurant en haut d’une tour; à faire tourner des stylos, recycler des vaisseaux, faire combattre des poireaux, écouter de la musique, ou tout simplement à discuter. Merci d’avoir gardé ce quelque chose qui fait que tu es toi après tout ce temps.

Merci ensuite aux deux zigotos avec qui je partage une conversation depuis des années, et avec qui la relation dure, contre vents et marées. Merci pour tous vos efforts, et merci pour votre amour.

Merci à Pierre, dont je reprends éhontément la structure de ses remerciements. Merci d’avoir été là au lycée pour me faire me rendre compte de ce qui était vraiment important pour moi. Merci pour toute l’inspiration que tu m’as fournie, de m’avoir soutenu et poussé à être la meilleure version de moi-même sans me brusquer. Merci pour ton intarissable passion pour l’enseignement, les sujets scientifiques, et l’Histoire, pour tes encouragements en prépa, de t’être tordu le cou pour moi, et pour toutes ces nuits où nous avons trop peu dormi à relater les événements de notre monde pour essayer d’y mettre du sens. Merci d’avoir pardonné ma suffisance et ma mémoire défaillante toutes ces années, et de répondre aux messages, toujours, et avec toujours plus d’entrain. Merci de m’avoir fait rencontrer tant de gens si différents, et merci pour cette page 14. Soyons honnête, je n’aurais jamais fait cette thèse si ce n’était en partie pour te suivre dans cette expérience du monde de la recherche, et partager une expérience de plus avec toi.

Merci Eric pour ton accueil au collège, pour ces LANs, ces bodegas pendant lesquelles tu étais ma lanterne, ces clés USB remplies de musique, et pour toutes ces nuits où nous avons trop peu dormi à relater les événements de notre monde pour essayer d’y mettre du sens. Merci d’être resté dans le bus pour m’attendre le 1er avril, merci d’avoir été là avant, pendant et après ma soutenance, et de manière générale merci d’être là dans les moments importants. Merci pour cette amitié solide et pour les efforts que tu investis dedans. Les cicatrices font partie de notre histoire, et rendent d’autant plus forts les moments de connivence que nous partageons aujourd’hui.

Enfin, merci Laurie. Pour toutes ces épreuves qu’on a traversées ensemble, et pour toutes les aventures aussi. Des rues parisiennes noires de monde au calme humide de la Lozère, en passant par les temples malais, la canicule croate, ou même la porte de Baldur, j’ai fait le tour des mondes avec toi. Je pourrais dire que ces voyages m’ont construit en tant qu’adulte, mais ce serait faux. Ce qui m’a véritablement fait grandir, c’est toi. Alors merci d’avoir tenu bon face à ma fierté, à mes retards, à mes pires humeurs, et face à mon rapport de stage il y a dix ans, alors que je partais en vacances de mon côté. Merci de m’avoir introduit au féminisme, aux sciences sociales, à l’art de la communication, et à tant d’autres sujets qui m’ont rendu un peu moins innocent, mais surtout un peu moins con. Merci de montrer l’exemple en suivant tes convictions, et pour ta considération envers les générations actuelles et futures. J’ai découvert ta force avec les années, mais je ne cesse d’être étonné. Merci donc pour l’inspiration et le courage que tu transmets, sans lesquels je ne me serais certainement pas lancé dans ce doctorat. Et, plus personnellement, merci pour la confiance que tu m’accordes, pour tes encouragements, ta franchise, tous ces bons petits plats, et pour ton amour sincère.

Je ne peux finir mes élucubrations sans dire un mot à ma famille qui, malgré sa taille menue et les distances que j'ai prises si jeune, m'a tant apporté et n'a jamais cessé de me soutenir.

Merci à mon arrière-grand-père René, avec qui j'ai eu la chance de partager mes premières années et qui m'a appris à rire, à peindre de la pâte à sel, à faire du vélo, des origamis, du ping-pong assis dans la cuisine, à aimer les caramels, les jeux de Super Nintendo trop difficiles (mais qui a fini Tintin au Tibet ?!), et toutes les choses dont pourrait rêver un enfant.

Merci, marraine Françoise, pour les soirées passées dans tes maisons quand j'étais enfant, pour la glace au chocolat devant les films, et pour les masques effrayants accrochés aux murs qui m'ont valu tant de cauchemars.

Merci, Édouard, de m'avoir ouvert ton antre d'adolescent et fait découvrir tant de films, de jeux et de mangas qui n'étaient clairement pas de mon âge ; merci pour les dessins de Batman à colorier, les batailles de GI Joe, les nuits de Mario Party et, plus tard, de m'avoir fait découvrir Palahniuk qui a tant influencé ma pensée.

Merci, mamie Jo, d'avoir pris soin de moi, de m'avoir conduit à tant d'endroits et autorisé l'achat de tant de jeux et jouets dans les braderies (mais uniquement les bonnes affaires !). Merci pour les sandwiches à la sortie de l'école, la recette du rata, des escargots et de la langue de bœuf. Merci pour tout le soutien que tu as apporté à notre famille durant tes belles années, et pour l'amour que tu as toujours donné à tes petits-enfants.

Merci à mes sœurs, bien sûr, que j'ai la chance d'avoir, qui m'inspirent malgré la distance et dont je suis si fier d'être le frère.

Merci, Annelise, d'avoir égayé notre enfance dans le Nord avec ton énergie débordante, même si je ne savais pas toujours comment l'accueillir en tant que pré-adolescent (et que je n'avais pas envie que tu entendes mes copains dire des gros mots). Préoccupé par ma propre émancipation, je n'ai — je le regrette — pas pris le temps de t'observer grandir, et qu'est-ce que tu as grandi ! Merci pour la solidité que tu apportes à notre famille, pour tes rires et ton humour subtil, ta fraîcheur, ton intelligence et ton engagement à rendre le monde meilleur. Merci d'être devenue cette personne si forte et si agréable et pour tous les efforts que tu n'as jamais cessé de fournir pour ton entourage.

Au passage, merci, Guillaume, pour ta sympathie, ton humour également, tes talents d'intégration déconcertants, tes phrases rassurantes, et pour l'amour sincère que tu sembles porter à ma sœur. Et merci à vous deux d'avoir assemblé mon meuble d'entrée.

L'autre, maintenant. Un sacré bout. Encore une qui m'impressionne. Merci, Cyrielle, pour tes défilés en déguisement, tes grimaces sur les photos, ton flegme au quotidien, et pour les tartines et les desserts que tu daignes partager. Merci pour nos discussions et nos promenades avec les chiens pendant lesquelles j'ai pu découvrir ta clairvoyance sur les rapports sociaux, ton esprit d'analyse, et ta sensibilité. Merci pour cette franchise qui te définit, sur les sujets légers comme sur les plus complexes, et surtout merci de me faire sentir accepté pour qui je suis.

Enfin, merci à mes parents qui m'ont toujours soutenu malgré nos différences, et dont la force de volonté est inspirante. Ils m'ont offert le cadre et le soutien nécessaires pour que je me sente libre de faire ce qu'il me plaît. Sans cela, je n'aurais

certainement pas osé reprendre les études, ni exploré autant de métiers différents, et je n'aurais probablement pas eu l'opportunité de me lancer dans la recherche scientifique.

Merci, papa, pour toutes ces initiations, de m'avoir lu des histoires quand j'étais petit, de m'avoir appris les échecs, de m'avoir installé Adibou puis Starcraft sur ton ordinateur, de m'avoir fait découvrir les grands classiques de la musique de ta génération et laissé à disposition ta collection de CD, de m'avoir dit « ceci est une clé plate » ou « viens, je te montre comment on passe une gaine », et, de manière générale, de m'avoir transmis le goût des sciences et des sujets techniques. Bon, le voilier, ça n'a pas pris, mais j'imagine qu'on ne peut pas grand-chose face à nos phobies. Merci aussi pour toutes ces fois où tu nous as sauvé les miches à l'étranger, qu'on soit perdus au milieu de la mer des Caraïbes ou coincés dans un taxi fou en Thaïlande. Si la communication n'a pas toujours été facile, j'ai ressenti, à travers toutes ces choses, l'amour caché derrière la pudeur caractéristique des pères.

Et puis merci, maman. Pour les milliers de repas, de conduites, et pour toutes les petites et grandes attentions que j'ai trop longtemps considérées comme acquises. Merci pour toute cette énergie (débordante) que tu investis dans la famille, mais aussi pour tes sourires, pour tes câlins que j'ai trop refusés, pour les photos, l'organisation des voyages, pour tes appels qui font vivre notre relation, pour tes explications sur le fonctionnement des banques et de la fiscalité, et pour les chiens que tu as déjà sauvés¹. Tu m'as appris à foncer sans avoir peur, à être fier de mes valeurs, et à vivre la vie que j'ai choisie. À mon échelle, c'est cette trace que j'essaie de suivre aujourd'hui.

¹Merci Pepette, Pipeau, Argos et Djack pour toute l'affection et le réconfort que vous avez apportés à notre famille. Soyez assurés, en lisant ces lignes, que notre amour pour vos poils est — ou fut — bien réel.

Disclaimer on the use of AI in the research process

This thesis occurred at a critical time in the evolution of artificial intelligence. Large language models (LLMs) became famous during the thesis, with the advent of ChatGPT (OpenAI) in late 2022 and Gemini (Google) one year later. Since, it has been demonstrated that these tools were increasingly used in scientific writings (Liang et al. 2024).

Because of their tendency to claim false information with confidence, these LLMs should be used carefully. Moreover the energy consumption of generative models is far from being negligible and already has an environmental impact.

Despite those facts, I believe, through my experience with this thesis, that AI might bring a lot to the research community by increasing research speed (especially to code experiments) and redaction quality. I think it is important, for transparency reason, to claim each time how AI have been used. So here we are.

During the second half of this thesis, I used LLMs as assistants for multiple purpose:

- it helped me to learn new things, especially about coding and AI architectures and implementation,
- it allowed me to code faster, in a more standardised and clean way by explaining hard-to-understand errors or even suggesting entire chunk of code,
- it corrected the content I wrote in English to a suitable scientific form (for example using passive voice instead of active voice, or proposing words that I wasn't aware of),
- it translated in English some content I wrote in French,
- it helped me organise my ideas by suggesting better order and transitions between paragraphs and sections of my manuscript.

Needless to say, all textual content retrieved from the LLMs was inspected in detail and adapted before being used in the manuscript. I apologise to the reader if the sentences are sometimes wordy, I have made every effort to ensure that the end result is as clear as possible.

Overall I think that AI helped me to produce better science, and that this manuscript carries my work in a much clearer way than the version I would have produced without this technology.

Have an enjoyable reading.

❖ **This text was not generated nor corrected by ChatGPT.**

Introduction

This thesis focuses on the design of pan-scalar maps, i.e. contemporary interactive and zoomable maps such as Google Maps. Pan-scalar maps are now commonplace and billions of people use them frequently. Still, they are relatively new in the history of maps, and their design is a recent domain of research. While map designers can rely on some best practices from paper maps, designing a pan-scalar map is not straightforward. It builds on the foundations of classical cartography, human-computer interface design, and human perception and cognition. Integrating all these disciplines to create a high-quality map is a challenging task and we can observe design faults in need of rectification. Before diving into those problems, let us quickly recap the major advancements in the history of cartography, highlighting their influence in the design of our modern maps.

Preamble

The earliest maps, which appeared several millennia ago, were rudimentary representations of the known world at the time, engraved on clay tablets or drawn on animal skins. The Babylonians created some of the first known maps around 2300 BCE, often centered around their cities and depicting surrounding rivers and mountains. The ancient Greeks made significant advancements in cartography between 600 and 100 BCE, developing more detailed maps and geographical theories. Muslim scholars then developed that knowledge between 800 and 1200, and mixed it with what merchants learned in their travels to make the best world maps known at the time (Figure 1).

In the 15th century, the invention of the mechanical printing press by Johannes Gutenberg revolutionized cartography, making the production and dissemination of maps on an unprecedented scale possible. This era also saw the emergence of the first planispheres. Since then, printing techniques have continued to improve, particularly with the industrial and energy revolutions of the 19th and 20th centuries. But it was not until the 1960s that cartography underwent a new revolution—the digital revolution.

In the early 1960s, the digitization of maps primarily involved scanning existing paper maps to make them available on digital platforms. This initial step, while revolutionary for its time, was quite limited in terms of functionality and interactivity. However, in the 1980s, the production of digital maps began to rely on geographic databases, transitioning maps from static images to dynamic compositions built from information extracted from these databases. This new approach greatly facilitated the production of both topographic and thematic maps, and enabled the creation of



✿ Figure 1 : The Tabula Rogeriana, an ancient world map drawn by Muhammad al-Idrisi for Roger II of Sicily in 1154. Feeling lost? Try turning it upside down. Source: Bibliothèque Nationale de France (BNF)

maps of virtually infinite size, like those we use daily today.

From this point on, the development of digital maps accelerated rapidly. Initially restricted to professional use, digital maps became accessible to the general public in the early 2000s with the advent of web maps. In less than twenty years, the transition from physical map formats to digital platforms (on computer and then on smartphone) allowed world-covering maps to be accessible anywhere, on-demand, by anyone. This transition also enabled new interactions - such as zooming to obtain additional information or typing an address in the search bar to get a proposed route. So arose the pan-scalar maps mentioned at the beginning of this introduction.

Research focus

Despite their widespread use, pan-scalar maps are relatively new, with only about 30 years of development compared to the millennia-long history of paper maps. This brief period of evolution means that there are still numerous areas for refinement in their design, such as user-friendliness, aesthetics or resource consumption. Specifically, new problems emerge from the new ability of map users to navigate through scales and map interaction is helping to identify areas for improvement. One of them that particularly interests us is the issue of user disorientation when navigating through different scales, particularly due to abrupt changes in data representation (Figure 2).

The Lost In Zoom project, to which this thesis contributes, explores ideas to mitigate this phenomenon of disorientation (Touya 2020b). Drawing from the anchor-point theory of cognition, we believe that certain specific elements on the map that are salient at multiple scales can help the users to orient themselves during their exploration of the map. We call these salient features that act as reference points "pan-scalar landmarks", and we think they can provide a consistent sense of spatial orientation across different zoom levels.

In this thesis, we focus on finding elements within the map that can act as pan-scalar landmarks. We propose exploratory experiments that combine user perception of the map with automatic detection of cartographic elements, sometimes using artificial intelligence. Our objectives are twofold: we aim to study the data from our experiment participants to better understand how they perceive the map, and also to propose methodologies for enriching our databases with information that could subsequently be used directly by cartographers or integrated into automatic generalization processes that give visual emphasis to certain features in order to reduce the disorientation associated with large changes in zoom level.



☞ Figure 2 : Southeast region of France as displayed on the Plan IGN at zoom levels 9 (top left), 10 (top right), and 11 (bottom). A drastic change in representation is noticeable at zoom level 9, as well as a significant gap in detail at zoom level 10. These changes may induce disorientation.

Structure of the manuscript

Part 1 introduces the basic concepts and provides a general state-of-the-art overview of the main themes of the thesis: digital cartography and cognition (the technical state-of-the-art related to pattern detection and artificial intelligence are presented in Part 3, when relevant).

- Chapter A presents the concept of pan-scalar maps, introduces related terminology, and explore the problem of disorientation that such maps can create. We then discuss cognitive aspects of map use, introducing the notion of cognitive anchors. This is followed by a discussion on the design of pan-scalar maps, including the map generalization process. The chapter concludes with an outline of the thesis objectives and an overview of the conducted experiments.
- Chapter B presents three preliminary studies conducted early in the thesis. The first aims to identify what qualifies as a landmark in contemporary digital maps. The second explores the concept of landmark density and proposes a thought experiment to estimate its lower bound. The third examines the evolution of digital map interfaces and discusses how these interfaces may impact user navigation and contribute to disorientation.

As a way of focusing the problem, Part 2 examines specific anchors and proposes new approaches to improve their detection.

- Chapter C presents a study on the perception of urban ring roads. We hypothesize that while perceptions of ring roads vary across individuals, certain attributes are universally recognized by map users. Based on these findings, we construct an ontology of ring roads, which we argue can enhance detection algorithms.
- Chapter D focuses on remarkable buildings and explores how existing machine learning models based on tabular data could be improved by incorporating map imagery to provide additional contextual information.

Part 3 proposes a comprehensive approach for detecting a variety of potential cartographic anchors in map images using deep learning models.

- Chapter E introduces the concept of Areas of High Anchoring Potential (AHAPs) and explains the rationale for detecting them in map images. It also presents our evaluation methodology for AHAPs detection in map images, including the creation of a dedicated dataset built from expert annotations.
- Chapter F introduces the concept of image segmentation and describes the two models we used to detect AHAPs (U-Net and Segformer). It then presents the training datasets and the models' results, before concluding with a discussion of the limitations and potential improvements of our approach.

Part I

The Role of Landmarks in Pan-Scalar Maps

Chapter A

Literature review

This chapter, parts of which are inspired by our co-written article (Touya, Potié, and Mackaness 2023), presents the essential definitions necessary to understand our work. The objective is to establish the technical foundations that will be referenced throughout the entire manuscript and to provide an overview of the conceptual challenges encountered in designing pan-scalar maps.

A.1 Landmarks and disorientation

Who has never felt a momentary sense of disorientation while exploring a digital map on their phone or computer? Even if only for a few seconds, that moment of uncertainty before zooming out or panning to familiar territory can be unsettling. But how does one regain their bearings? Perhaps it is the recognition of a broad avenue ending in a roundabout, the familiar sight of a restaurant by a river, or the name of a neighboring village. More likely, it is the combination of multiple elements that helps us reestablish our sense of location. These points of reference, known as "landmarks", guide us in all various navigational contexts: navigating at sea, on land whether on foot or by vehicle, navigating through a book with chapters and page numbers, on a webpage, or within maps.

This section provides a review of the existing literature on the concept of landmarks, examining their role and significance in both cartography and spatial cognition. We will explore how landmarks function as anchors in the user's mental map, and their importance in mitigating disorientation during map navigation. The concept of landmark is hard to pin down, yet Richter and Winter 2014 offers a comprehensive review that we draw upon in this section. For readers interested in the topic, we highly recommend consulting their book.

A.1.1 What landmarks are

The term "landmark" carries a multitude of definitions, which vary slightly depending on the author and context. It is therefore important that we clarify our stance. When we look at dictionary definitions of the word "landmark", two primary meanings often emerge. The first (1) defines landmarks as prominent, distinguishable

features in the real world, such as a particularly high building, which help people recognize a place (e.g., Figure A.1). These types of landmarks are often used when giving directions or describing locations (e.g., "the park next to the Hôtel des Invalides"). The second common definition (2) refers to landmarks as dates or events that mark a significant moment in time (e.g., "the day we met, on 14 July").



⌘ Figure A.1 : The distinctive design of the famous sculpture "L'axe de la Terre" makes it a notable landmark that is only seeable on the Champs-sur-Marne campus.

Although these two definitions encompass the most common uses of the term "landmark", they miss some important nuances. For example, a place can become a personal landmark for someone if it is tied to a meaningful event, even if it lacks distinctive physical characteristics (e.g., "the bar where we watched the World Cup final"). Indeed, studies on the impact of emotions on spatial memory are emerging and show promising potential (Rasse, Gyselinck, and Guegan 2023). The previous definitions also tend to focus on the nature of landmarks (buildings, areas, events, etc.), while it might be more fitting to consider their function -helping individuals orient themselves in space or time - and how the human mind perceives them (Richter and Winter 2014, Section 1.1). By shifting the focus from the physical properties of landmarks to their cognitive and functional aspects, we can better understand their role in navigation and memory.

In the context of the map, we are interested in spatial landmarks (those that help us to find our bearings in space) rather than temporal landmarks. Lynch 1960, in his seminal work *The Image of the City*, was the first to define the concept of "landmarks" in the context of spatial cognition and urban navigation. During an experiment in which he asked participants to sketch maps of their hometowns, he observed that these drawings often contained similar structures, suggesting shared spatial perceptions among individuals. This experiment led him to define landmarks as "prominent objects that serve as points of reference for individuals, either in real or virtual environments". This definition includes the function of landmarks and thus

seems much closer to reality than the definition usually proposed in dictionaries. Still, this definition is not fully satisfactory for us as it is not really linked with what occurs in the mind of individuals.

Fortunately, in the same work, Lynch expanded on the theory of mental maps (which he called “mental picture” or “mental image” back then). Let us take a brief detour to explain mental maps, a key concept in spatial cognition which is crucial for formulating our own definition of landmarks and explaining the phenomenon of disorientation.

A.1.2 Landmarks and spatial cognition

A.1.2.1 Mental maps and the phenomenon of disorientation

Understanding the notion of mental map is important in the context of our thesis because the phenomenon of disorientation is linked to it. From a cognitive perspective, **disorientation** can be seen as the difficulty in reconciling sensory inputs with the mental map, leading to a sense of confusion or a loss of direction (Touya, Gruget, and Muehlenhaus 2023).

A **mental map**, also known as a **cognitive map**, is an internal representation of the spatial and topological relationships and features in an environment, formed within the human brain. It serves as a mental model that helps individuals navigate through and understand their surroundings. Mental maps are not literal “maps” in the sense of being visual or metrically accurate representations. Instead, they are cognitive frameworks, composed of places and regions that are linked in our minds by connections representing spatial relationships (Figure A.2). They do not store all details—such as the exact location of a specific road or building—but are instead composed of key concepts that emerge from the combination of these geographic elements (Montello 2002).



✿ Figure A.2 : A drawn representation of a mental map, by Kevin (17). Source : Trell and Hoven 2010

The concept of mental maps was first introduced by Tolman 1948. In his paper,

Tolman argued that both animals and humans develop mental representations of their environment, which they use to navigate and understand spatial relationships. This concept has since been extensively explored and refined in fields such as psychology, cognitive science, and cartography. By the late 1970s, certain properties of mental maps were already well established: regionalization (the division of space into meaningful regions), route segmentation (the linking of regions through spatial or functional relationships), hierarchical organization of information, and what interests us the most: the importance of salient cues in constructing mental maps (Lynch 1960; Stea 1973; Gould and White 1974; Siegel and White 1975). Around the same time, a theory emerged, which attempts to connect the properties of landmarks with the structure of mental maps, suggesting that these anchor points play a crucial role in spatial cognition and navigation.

A.1.2.2 The Anchor Point Theory

The Anchor Point Theory was introduced by Couclelis, Golledge, Gale, and Tobler 1987 and has gained considerable recognition in the field of spatial cognition research since then. This theory suggests that cognitive maps are organized around "anchor points" that are defined as follow:

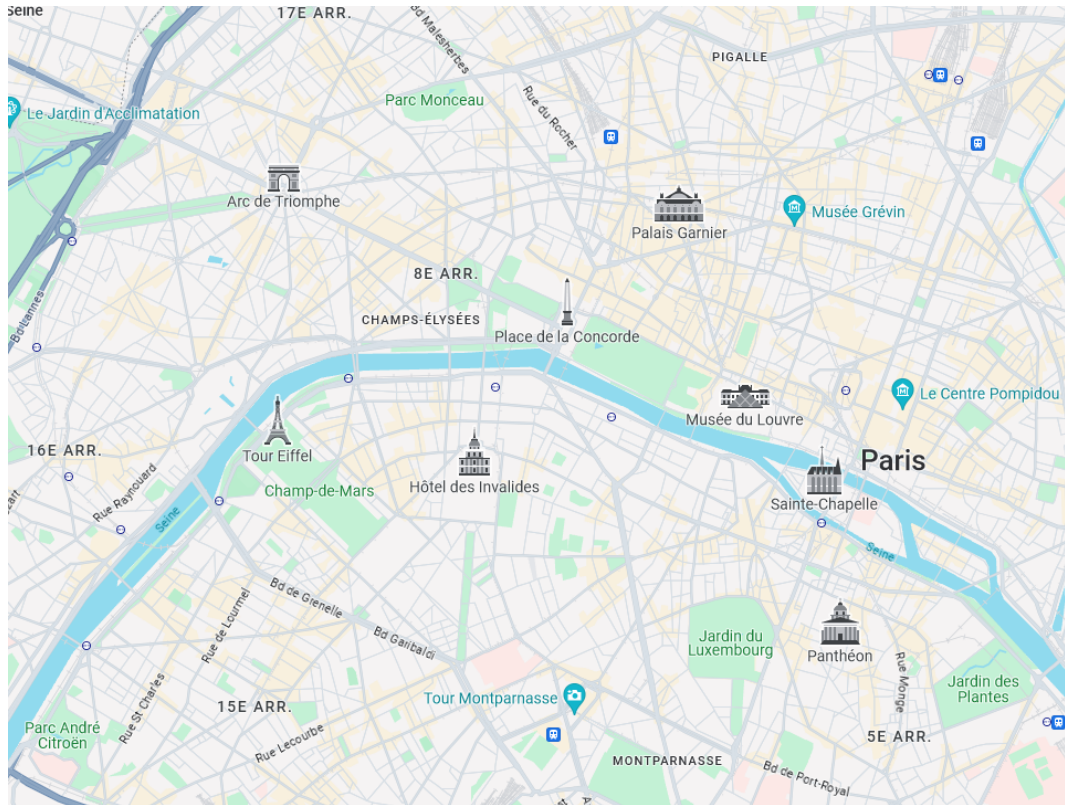
Anchor points (or “anchors”) are reference points in the cognitive space of individuals that help structure spatial information in the mind, facilitating navigation and distance estimation. These anchor points can be mental representations of landmarks, or more abstract and personal concepts.

According to Couclelis, the anchors are not all equal in terms of hierarchy, with the most prominent ones being referred to as "primary anchors". The concept of prominence is close to the concept of “**saliency**” that is central in all definitions of landmarks and that has been developed by Sorrows and Hirtle 1999. In their paper, Sorrows and Hirtle explain that a landmark can be salient because it visually stands out (a pink house in the city center), because it is structurally important in its environment (the menu of a webpage), or because it is related to someone’s experience (one’s house or workplace); the most salient of them often possessing two or three of these characteristics.

To sum up, landmarks in the environment (real or virtual) can create anchor points in the mind of individuals, and the more salient the landmarks are, the stronger the anchors are in the users’ mind. Golledge, Jacobson, Kitchin, and Blades 2000 even claimed that landmarks not only support the construction of the mental map by creating reference points that individuals may recognise later, but also assist individuals in locating less important features near them (for example remembering the location of a specific store because it is near the Eiffel tower). If this statement holds true, it becomes crucial to consider it in cartographic design in order to enhance the user experiment.

One could imagine that an indirect objective of cartographers (or application designers in general) would be to facilitate the emergence of anchor points in the

mental maps of their users by designing specific landmarks in the map¹. This could be achieved, for example, by modifying the symbology or the generalization of certain elements on the map, as Google did with some historical buildings in the European capitals (Figure A.3). In this thesis, our experiments are designed and our results analyzed with the anchor point theory in mind.



❁ Figure A.3 : View of Paris on Google Maps. Some important buildings, which serve as landmarks in the real world, have been represented with symbols to become landmarks on the map as well.

A.1.3 Landmarks in our thesis

A.1.3.1 Landmark definition

In the context of this thesis, we set aside personal landmarks—those tied to an individual user’s personal knowledge (Nuhn and Timpf 2017)—and focus instead on studying landmarks that have the potential to be recognized by most map users.

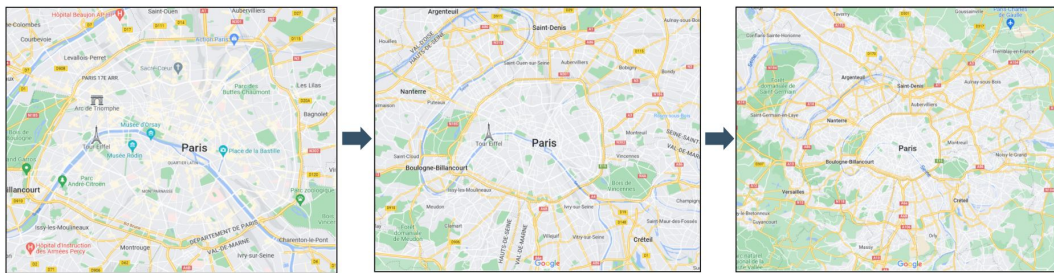
We propose the following definition of landmarks for this thesis:

¹For example, Schwering 2019 suggests adding descriptions of landmarks in wayfindings instructions to the existing turn-by-turn instructions in order for the user to build an incidental knowledge of the environment and feed its mental map.

Landmarks are prominent features on the map that have the potential to create anchors in most users' mental maps, providing them with the necessary reference points for navigation.

A.1.3.2 Landmarks in pan-scalar maps

As we study the phenomenon of disorientation in pan-scalar maps, we are particularly interested in landmarks that are visible at multiple scales and we call them **pan-scalar landmarks**². Just as landmarks in traditional maps, pan-scalar landmarks are of multiple forms (buildings, transportation networks, natural features...). Each pan-scalar landmark has a **scale-range**, ie a range of scales on which the landmark is visible on the pan-scalar map. Depending on how they are represented on the pan-scalar map, we believe they can help navigation and help users build vertical spatial relations between map features in their mental maps (Figure A.4).



☞ Figure A.4 : Paris at multiple scales when zooming out. The ring road and the toponym 'Paris' remain visible and act as pan-scalar landmarks, while the Seine River and the monument symbols disappear. Would displaying the Seine more prominently at smaller scales help users navigate between scales more effectively?

Pan-scalar landmarks can emerge naturally from the environment, or be intentionally designed to facilitate navigation and understanding of the map. Actually, specific elements of the map can effectively be turned into landmarks by enhancing their symbology, for example by representing them with pictograms (Figure A.3).

A.1.3.3 Note on salient visual cues

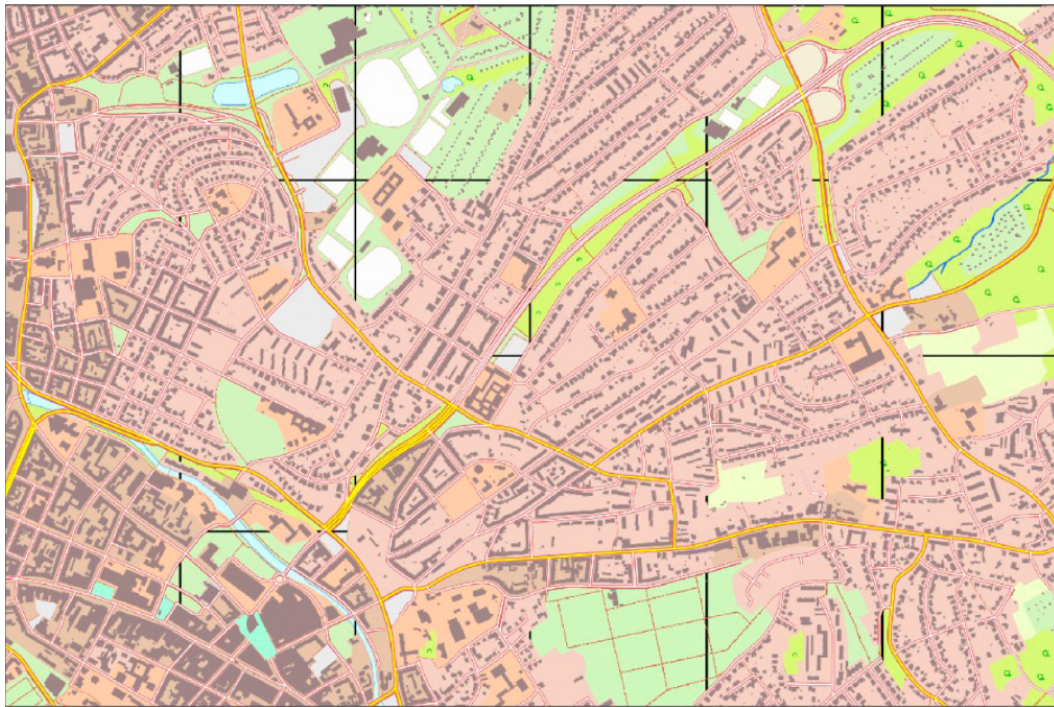
We define salient cues as visual elements on the map that provide users with immediate visual feedback on their movements or scale changes. While non-personal landmarks in the map are often salient visual cues, not all salient visual cues are landmarks. They may not be specific or memorable enough to influence users' mental maps significantly and may lack the characteristics necessary to form an anchor in their mind. For instance, a cue might be too abstract, e.g., color shading on a topographic map, or not specific to a particular location, e.g., a grid overlay (Figure

²In other publications within the Lost In Zoom project, the term "pan-scalar anchor" is more frequently used to refer to the pan-scalar landmarks. However, in this thesis, we prefer to differentiate between landmarks and cognitive anchors.

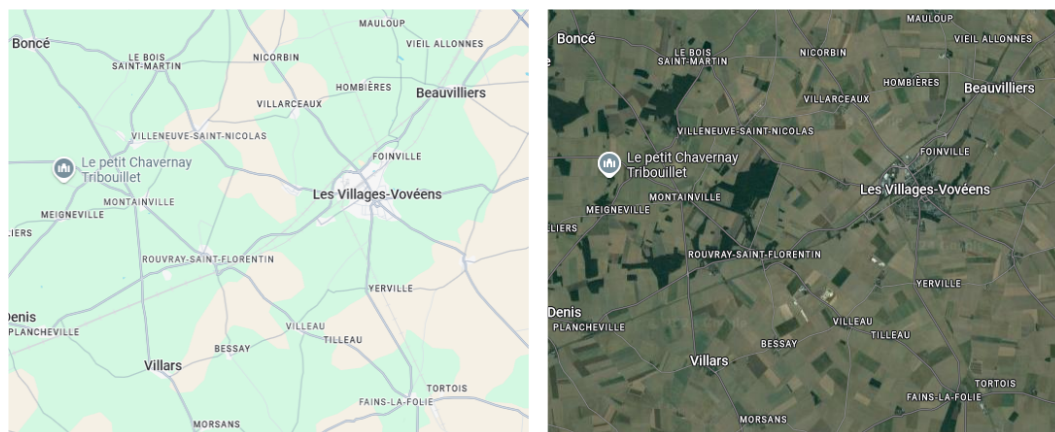
A.5).

For illustration, we see in the vector view of Figure A.6 a village on the right, surrounded by color patches corresponding to land use, although they are difficult to interpret. The beige patch serves as a visual cue that helps someone familiar with the map to locate the village quickly. However, it is unlikely that this colored patch will become a strong mental anchor for the user. Despite its visual prominence, this feature lacks the structural qualities typically necessary to form a lasting mental landmark (Sorrows and Hirtle 1999). This is even more apparent when looking at the satellite view on the right, where these color patches are hard to relate to real-world terrain features.

It would be a mistake to only see those cues as lesser landmarks however. They can still help users quickly locate and orient themselves within the map or find specific objects (Dickmann, Edler, Bestgen, and Kuchinke 2017). As we know reducing cognitive load associated with map use is key for better navigation (Bunch and Lloyd 2006), leveraging visual cues that assist in navigation without disrupting the user’s mental map may open up opportunities for map design. This thesis mainly focuses on landmarks, but some of our experiments deal with visual cues as well.



✿ Figure A.5 : Example of a partial grid used as a visual cue in the map, from (Dickmann, Edler, Bestgen, and Kuchinke 2017). Although this type of cue can be helpful for performing specific tasks, it is unlikely that the grid (or parts of it) will become a mental anchor for the user.



☞ Figure A.6 : Two map views of the same area in Google Maps, the first using the vector map and the second using the satellite imagery.

A.2 Pan-scalar maps

Now that we have introduced the concepts of landmark and disorientation, we must discuss pan-scalar maps, which provide the context and framework for our study.

A.2.1 What pan-scalar maps are

A.2.1.1 Pan-scalar map definition

Historically, each paper map had only one scale, displaying features relevant to that scale. However, the advent of digital maps, and later web maps, introduced the possibility of integrating all scales into a single interactive map, allowing for a seamless user experience (Haklay, Singleton, and Parker 2008). Roth 2013 uses the term multi-scale interactive maps for those, but we prefer to call them pan-scalar maps to highlight the fact that those multiple scales need to be used together in order to work well. We use the definition proposed by Gruget, Touya, and Muehlenhaus 2023 for pan-scalar³ maps:

A **pan-scalar map** is an interactive, multi-scale, zoomable map designed as a single cartographic product – i.e. an application, a map website, or a cartographic interface on a navigation device.

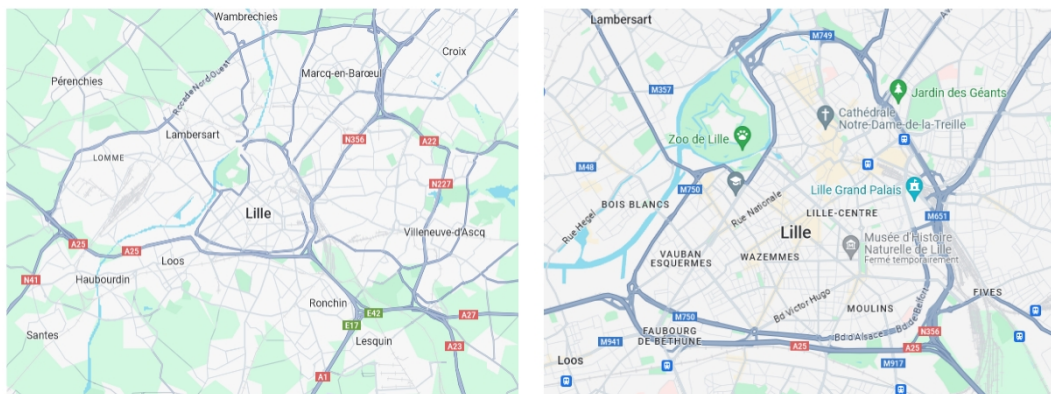
Pan-scalar maps differ from traditional one-scale maps as they encourage the users to interact with the map and explore it by **panning** (moving around) and **zooming** (changing scale). The continuous three-dimensional spaces (coordinates (x, y) and scale) of pan-scalar maps allows map designers to spread the information

³The prefix "pan" originate from the Greek word "παν" (pan), meaning "all" or "every". The term "scalar" is derived from the Latin word "scala", meaning "ladder", which is relevant to the concept of scale in maps.

across the different scales, thus reducing the amount of information visible at one scale for making map reading a more manageable task. It also allows map designers to use the scale of appearance to convey hierarchy (Gruget, Touya, and Muehlenhaus 2023) by making some map elements appear before others instead of just using the size or color of the symbols (Figure A.7). Because of this design, the cartographic representations at different scales are never meant to be consumed individually; each scale is designed to ‘work’ with maps at other scales.

A.2.1.2 Pan-scalar map uses

Pan-scalar maps are used for a wide variety of tasks. Boér, Coltekin, and Clarke 2013, based on the earlier research by Carter 2005, derived eight primary uses of maps: self-location, locating objects, route planning, navigation, identifying points of interest (POI), communication, information storage, and exploration. Ideally, the design of pan-scalar map design would adapt dynamically and propose a specific experience depending of what task the user is aiming to accomplish. In modern maps, this adaptation is mostly done by the interaction tools pan-scalar maps typically support: route planning menus for route navigation, search bars for locating objects, tooltips to identify POIs etc. Adapting the content of the map depending on the task is also a possibility, for example by enhancing some landmarks near the planned routes (Lokka and Çöltekin 2019) or by adapting the level of abstraction of the map (Boér, Coltekin, and Clarke 2013). However, this approach implies knowing what tasks the user wants to do and is not implemented yet in current maps.



☞ Figure A.7 : The city of Lille on Google Maps at zoom level 11 (left) and zoom level 13 (right). The most important information (from Google’s point of view) appears sooner and less important information appears later when we continue to zoom in.

A.2.1.3 The concept of scale and scale perception

The concept of scale is key in pan-scalar maps and in cartography in general, but it has multiple meanings that should not be confused. Montello 2001 distinguishes three: Cartographic scale refers to the ratio between the size of a feature represented on a map and its corresponding size in the real world. Analysis scale is the size of the unit at which some problem is analyzed, such as the country or state level.

Phenomenon scale (also called geographic scale) refers to the size or extent at which a specific phenomenon appears. In this thesis, we use the term “scale” meaning cartographic scale.

Cartographic scale has historically been indicated on maps numerically by a fraction (for example, 1:10,000) or by a verbal statement such as "1 cm equals 1 km". However, when maps are displayed on screens, the scale must be expressed in "meters per pixel" to remain valid because the size of the screen may vary depending on the user's device. A scale bar can also indicate the scale as it is not affected by screen size (Figure A.8). Map users can theoretically rely on these scale indicators during navigation to gain a more objective understanding of the scale at which they are currently viewing the map, but the absence of scale indicators in Apple's "Maps" application in 2024 suggests that such indicators may not be as essential for navigation as traditionally thought.



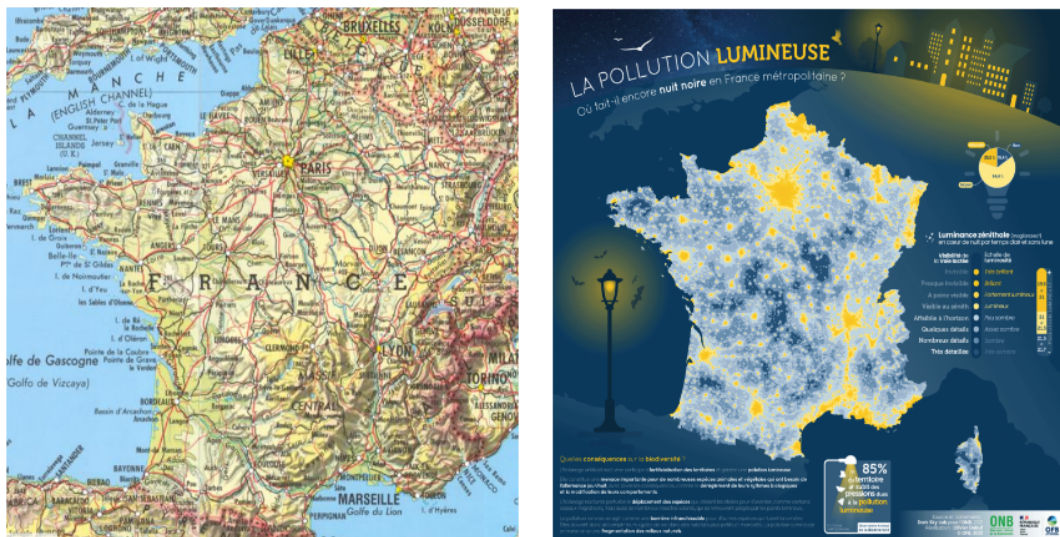
✿ Figure A.8 : Scale indicators on IGN map (top left), OSM (top right), Google Maps (bottom). We notice that only IGN displays the verbal statement, and that Google displays a scroll bar related to scales (on the right of the picture).

In fact, the study of scale perception in maps has a long history and started with paper maps, and it is established that paper map users typically infer map scale from the features and symbology present on the map rather than from artificial indicators such as scale bars (Eastman 2013). This observation likely holds true for pan-scalar maps, where users frequently rely on the relative size and representation of features across different zoom levels. Therefore, it is essential to design map content in a

way that enhances scale perception for users, allowing them to intuitively grasp their position within the map’s scale without needing explicit indicators. More information about how scales are implemented in pan-scalar maps is presented in Section A.2.3.3.

A.2.1.4 Thematic maps and topographic maps

A distinction is often made between two main types of maps: thematic maps and topographic maps (also called reference maps). Thematic maps focus on specific data and often convey a particular message, highlighting certain features or trends. In contrast, topographic maps aim to describe the physical reality of the terrain as accurately as possible, providing detailed information that can be used in various ways by different users (Figure A.9). The majority of the pan-scalar maps we use today are topographic maps, and our work focuses on these. However, it is also possible to create thematic pan-scalar maps. If the study of a particular theme requires the visualization of data at different and continuous scales, constructing a thematic pan-scalar map can be beneficial. For instance, Le Mao B., a LostInZoom project member, worked on a pan-scalar thematic map to help firefighters⁴.



✿ Figure A.9 : On the left, a topographic map of France (source: IGN). On the right, a thematic map of France showing light pollution (source: Office National de la Biodiversité)

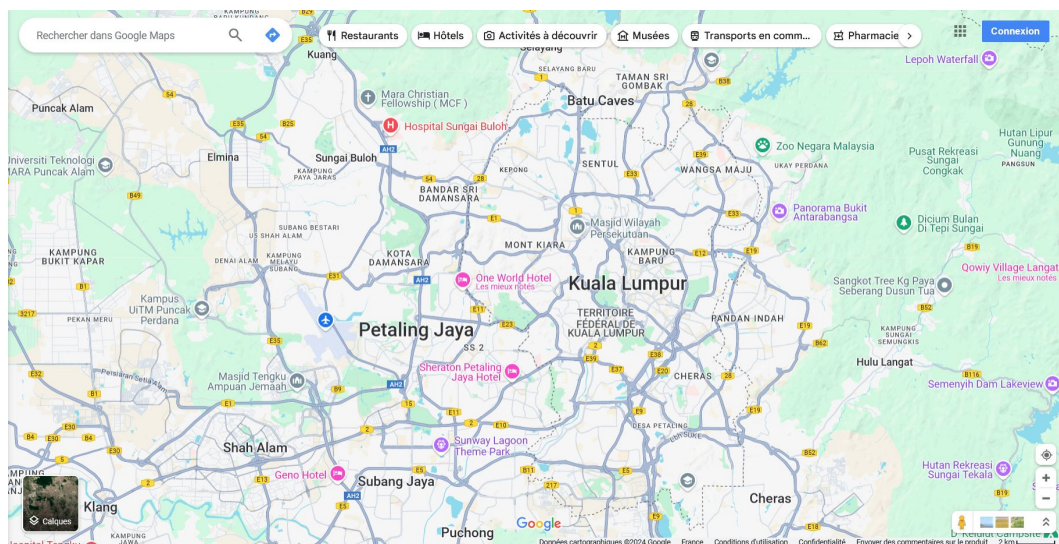
A.2.2 Disorientation in pan-scalar maps

Pan-scalar maps have now been around for about twenty years, but they continue to present challenges for users, particularly when it comes to maintaining orientation during navigation. One of them is the disorientation phenomenon occur-

⁴The associated article is not out yet, but the map is available here: <https://lostinzoom.huma-num.fr/seism/>

ring when zooming through scales or moving across unfamiliar areas on the map (Touya 2020b). Disorientation in pan-scalar maps can be explained by a disconnection between the user’s mental map—an internal representation of spatial relationships—and the changing visual cues provided by the map interface (Touya, Gruget, and Muehlenhaus 2023). As users explore new areas, they must continually reconcile these transitions with their existing mental map, which may not always align smoothly with the new visual information.

Disorientation typically manifests when users struggle to recognize familiar features or landmarks, or when they find it difficult to link the different views they encounter during navigation, making it harder for users to maintain their sense of position or scale. Interestingly, when users first encounter an unfamiliar area on the map, they do not experience disorientation in the same way. Instead, they begin constructing a new mental map based on the visible landmarks, which then serve as their anchors for navigation (Figure A.10). However, disorientation tends to arise as users move beyond these initial anchor points, struggling to integrate new views into their evolving mental map.



☞ Figure A.10 : The initial map view presented when searching for Kuala Lumpur in Google Maps search bar. The city is easily identifiable by its toponym and the dense road network. Petaling Jaya appears as another significant city, distinguished by the size of its label. To the east, we observe what appears to be a natural area. For someone unfamiliar with this part of the world, these elements might form their first cognitive anchors related to this region.

While some disorientation may be inevitable due to individual defective mental maps⁵, well-designed maps can mitigate this issue by offering consistent visual cues across scales and facilitating the creation of mental anchors. Before we explore specific design solutions, let us discuss three of the main causes of disorientation in

⁵It seems difficult to imagine that even the best-designed map could counteract the disorientation felt by someone who has always believed the Earth to be flat when confronted with a globe.

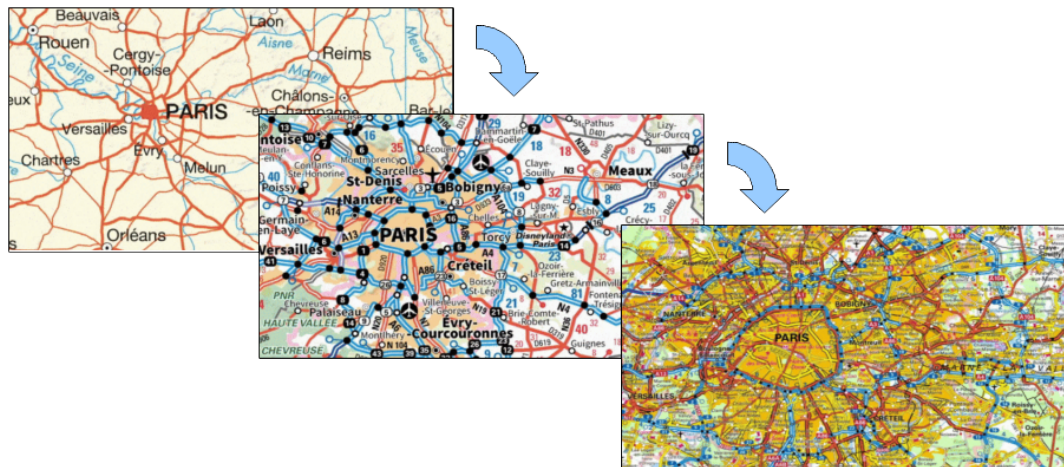
pan-scalar maps.

A.2.2.1 Cause 1: Changes in data representation

Disorientation can be experienced when a map user faces a change in the representation of the features in the map, for example when a paper map user switches from a map that represents the country to a more specific map of a region. The importance of this problem increases with pan-scalar maps, as users are encouraged to navigate within the scales. In pan-scalar maps, the features in the database are adapted to the different scales through a process called “map generalization” (see Section A.2.4). This process is mandatory to keep the map readable while displaying the most important information, but implies that what is displayed on the map changes when a user is zooming.

The first attempt by IGN to construct a pan-scalar map clearly illustrates this phenomenon of disorientation following a change in data representation. It involved overlaying a series of scanned paper maps in a web application, displaying them sequentially based on the viewing scale. The main drawback of this implementation was that when users reached a transition point and moved from one data representation to another, they could experience a loss of orientation (Figure A.11).

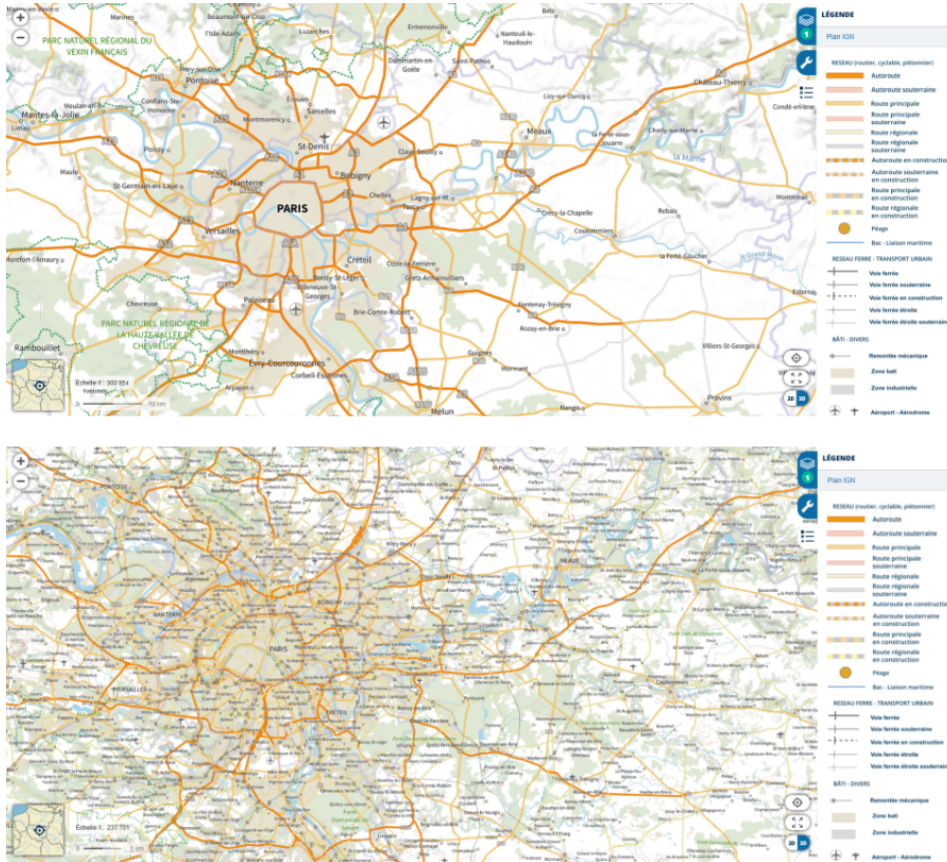
Fortunately, pan-scalar maps now maintain a much more consistent style across all scales. However, this does not entirely eliminate the potential for significant gaps in representation between certain scales (Figure A.12), particularly when zooming through multiple zoom levels at once.



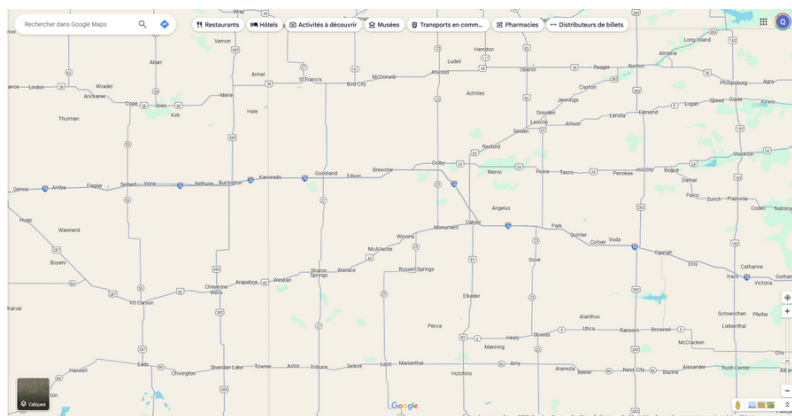
☞ Figure A.11 : Zooming on Paris in the “IGN classic” pan-scalar map. This web application was based on paper maps.

A.2.2.2 Cause 2: Lack of reference points

Disorientation is also enhanced by a lack of reference points (landmarks) in certain areas. This can occur in areas with very little information, such as deserts or rural regions, or in highly dense areas where nothing stands out, such as in certain large-scale urban representations (Figure A.13). During navigation, the lack of reference



☼ Figure A.12 : Two cartographic views of Paris and its surroundings from the IGN pan-scalar map, with less than one zoom level of difference between them. Yet the data representation changes significantly: town names appear (even small town's ones), new road and rail networks appear, and the generalization of roads and rivers changes. Source: <https://www.geoportail.gouv.fr/donnees/plan-ign-v2>

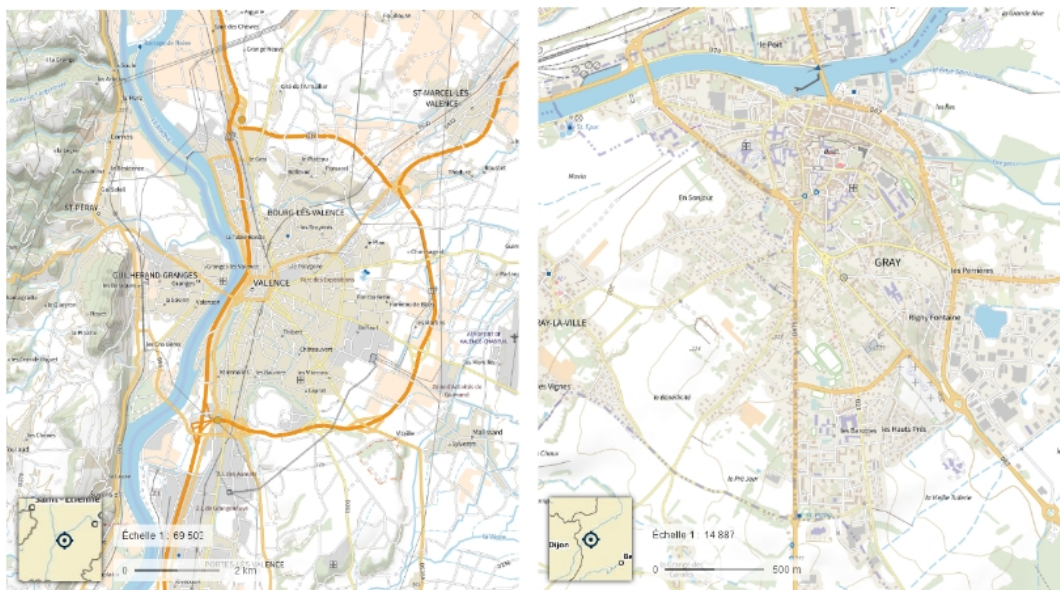


☼ Figure A.13 : A map view from the middle of the US, on Google Maps. In this area no element stands out and the same pattern repeats in all directions, making it difficult to know exactly where we were previously in the navigation and where we are exactly now.

points makes it hard to infer how much we zoomed or panned and to know where the previous view was, and thus induce disorientation.

A.2.2.3 Cause 3: Confusion between two locations or scales

Finally, disorientation can occur when users confuse two cartographic areas that look alike. More specifically, it arises when the user starts to sense that something is wrong but has not yet identified the nature of the error. It is interesting to note that these similar areas do not necessarily belong to the same scale, meaning that a user may mistakenly believe they are navigating at a particular scale when, in fact, they are at another. For example, it is hard to infer the zoom level of the two views presented in Figure A.14 just from the features they display as they share the same ones: a city with the same screen size that is next to a river, with a circular road network that goes north, south and east. This phenomenon can be particularly disruptive when zooming in or out across various scales, further complicating the user's orientation.



☼ Figure A.14 : Two map views that look alike, although one is taken from zoom level 13 and one is taken from zoom level 15.

A.2.2.4 Desert fog

Causes 1 and 2 can lead users to experience a "desert fog" effect. The theoretical problem of desert fog was introduced by Jul and Furnas 1998 in the context of human-computer interactions, even before pan-scalar maps emerged. The authors define **desert fog** as "a condition wherein a view of an information world contains no information on which to base navigational decisions". In digital cartography, a **view** refers to the specific portion of the map visible on the user's screen at any given moment.

Whether users lack information due to change in data representation or simply

because no meaningful information is present, the result is the same: a sense of disorientation. But the effect is more insidious when there is a confusion between two locations or scales, as users may not immediately realize they are lost.

The LostInZoom project (Touya 2020b) tries to address the phenomenon of disorientation by proposing an "anchor-based zooming". This approach seeks to highlight landmarks during zooming interactions, for example by prolonging their scale range or adjusting their colour or transparency, to better link different scales together and provide consistent reference points for users.

The project is structured around four main challenges:

1. Better understand the desert fog effect and measure it
2. Identify which landmarks serve as the best anchors and find a way to encode them
3. Build a pan-scalar map that may host these landmarks
4. Design a zooming interaction that makes use of the landmarks to minimize user disorientation

This thesis primarily focuses on addressing the second challenge of the project. But if we believe anchor-based zooming can help mitigating disorientation, we are also convinced that studying landmarks to improve map generalisation in order to build more continuous pan-scalar maps is key in the process (more details about map generalisation in Section A.2.4).

A.2.3 Designing and building pan-scalar maps

A.2.3.1 Map design: scale oriented versus data oriented

A map scale only exists in the context of a map displayed as an image. The data by itself does not have a particular scale, and can be displayed at several different scales (Carral, Scheider, Janowicz, Vardeman, Krisnadhi, and Hitzler 2013). For traditional single-scale, non-interactive topographic maps, the creation process typically begins with selecting the area to be represented and determining the appropriate scale. The next step involves deciding which information needs to be conveyed at that scale, choosing the data to align with the common uses of the map at that particular scale. For example, a 1:25,000 scale map is typically suited for hiking and urban planning, while a 1:100,000 scale map (regional scale) is more focused on transportation networks and key points within regions of France. Once the data relevant to that scale's intended uses has been identified, the design process focuses on integrating this information as clearly and effectively as possible in the map.

The process of creating pan-scalar maps, however, works in the opposite direction to this traditional approach (Courtial and Touya 2023). Instead of starting with a fixed scale and then determining the relevant information to display, the creation of pan-scalar maps is more similar to that of thematic maps. Here, the process begins by identifying the key information that needs to be shown, followed by the selection of an appropriate scale (or range of scales) to represent that information

effectively. This reverse approach allows for a dynamic adjustment of scales to suit the content, which is crucial for maps designed for continuous zooming and varying levels of detail.

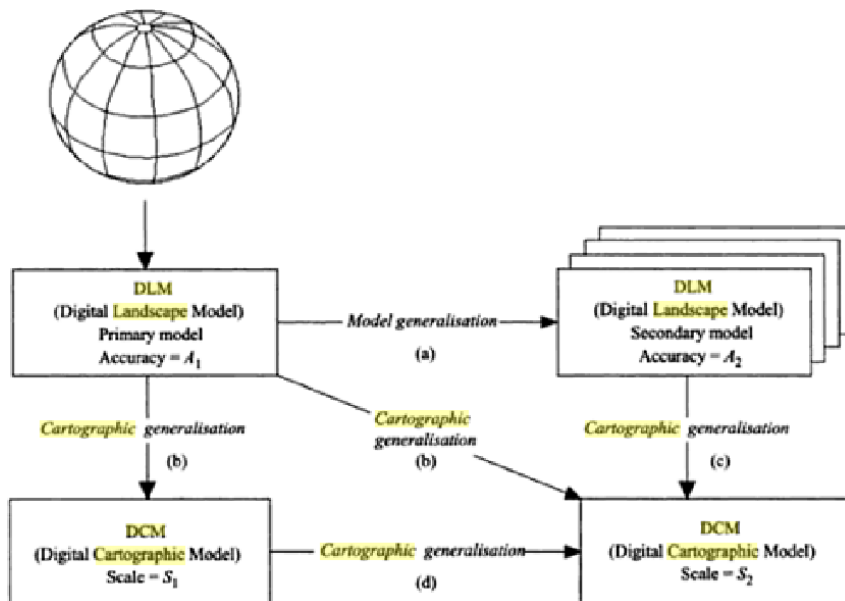
Although cartographic scale is continuous, allowing for any value to be selected, certain scale thresholds are more significant to users because they align with familiar concepts such as city, regional, or national scales. These scales can be more effective for conveying information because users are more likely to have seen other maps at these same scales, making it easier for them to recognize locations and draw connections with other maps. NMA maps are traditionally offered at scales of 1:25,000 (neighbourhoods), 1:50,000 (urban areas), 1:100,000 (regions), 1:250,000 (states), and 1:1,000,000 (continents) (Burghardt, Duchêne, and Mackaness 2014). We postulate that particular attention should be given to these key scales when designing a pan-scalar map, as well as ensuring smooth transitions between these scales to maintain coherence and usability across different levels of detail.

A.2.3.2 Pan-scalar map databases

Pan-scalar maps are dynamic compositions built from very large cartographic databases. These databases usually contain billions of georeferenced elements, each with precise coordinates, geometric shapes, and specific attributes, such as the number of lanes on a road or the height of a building. To efficiently manage all this information, National Mapping Agencies (NMAs) use Geographic Information Systems (GIS) software. They allow automation of many of the cartographic processes, enabling the integration and manipulation of large datasets necessary for building detailed and scalable maps. NMAs have historically utilised multiple database architectures for pan-scalar maps databases, and there is no consensus yet, but some practices have become more widespread.

The geographic databases used to produce pan-scalar maps are usually divided into two categories: Digital Landscape Models (DLM) and Digital Cartographic Models (DCM) (Stoter, Meijers, Oosterom, Grunreich, and Kraak 2010). A **Digital Landscape Model** contains information directly captured from the real world by diverse processes, such as deriving geometries from orthophotos. A **Digital Cartographic Model** is derived from a DLM, adapted to produce an optimal visual end-product. This transformation, called cartographic generalization, implies processes such as modifying the size and positioning of symbols in order to improve the clarity of the map (more about the generalization process on Section A.2.4).

High-accuracy DLMs can include fine details, such as individual houses, rural roads, or even traffic lights, which are not necessarily useful at smaller scales in pan-scalar maps. Lower-accuracy DLMs can be created by generalizing a high-accuracy DLM, a process known as model generalization. From each DLM, a corresponding DCM can be derived to be displayed at specific scales, although a lower-accuracy DCM may also be directly derived from a high-accuracy DLM. A schematic illustration of these relationships is shown in Figure A.15. Different structures are possible and studied to derive DCMs from one or more DLMs. There is no consensus yet but some good principles are described in Courtial and Touya 2024.



✿ Figure A.15 : Schematization of DLM and DCM concept. Source: João 1999.

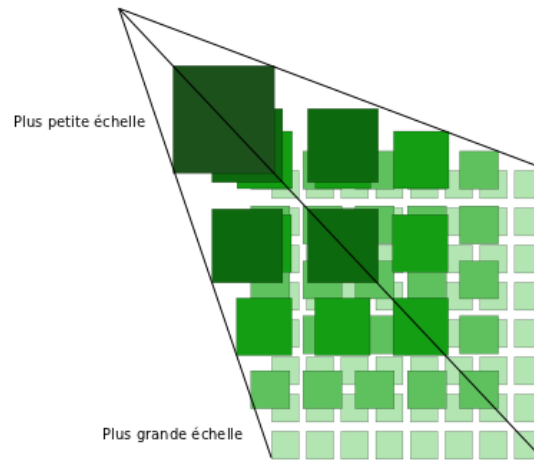
A.2.3.3 WMTS structure and zoom levels

WMTS (Web Map Tile Service) is a standard protocol developed by the Open Geospatial Consortium (OGC) to efficiently deliver map data over the internet in small, manageable "tiles." Rather than downloading an entire map as a single image, WMTS breaks the map into a grid of smaller tiles. These tiles are stored on servers and can be requested with a HTTP query and loaded individually as needed. This protocol allows for quicker loading times and make the data interoperable across different applications and platforms.

Map tiles are stored in a quadtree (pyramid) structure (Figure A.16). The **zoom levels** correspond to the different levels of this quadtree. Each zoom level corresponds to a different level in the quadtree. At the top of this pyramid is zoom level 0, which represents the broadest view (often encompassing entire continents). As we move down the pyramid, each subsequent zoom level doubles the number of tiles, increasing detail and narrowing the map's focus (Figure A.17). At first, pan-scalar maps only displayed maps at integer zoom levels (i.e., 10, 11, 12) but recent maps allow the users for "smooth zooming" (Van Wijk and Nuij 2003) to stop at decimal zoom level (i.e., 10.3 or 11.7), which gives a smoother user experience. This provides a smoother user experience, even though these decimal zoom levels often share the same data representation as their corresponding integer zoom levels in most pan-scalar maps.

A.2.3.4 Where in the Map Production process does anchor point modelling take place?

Given the complexity of pan-scalar maps, which integrate multiple levels of detail and aim to cover extensive areas—potentially even the entire world—the map creation and update processes have been partially automated. Today, specialized software assists cartographers, allowing them to focus on key areas that require human ex-



✿ Figure A.16 : Illustration of the zoom-level pyramid in pan-scalar maps, with the largest tile at the top representing zoom level 0, followed by four tiles below corresponding to zoom level 1, and so on. Source: IGN

Level	# Tiles	Tile width (° of longitudes)	m / pixel (on Equator)	~ Scale (on screen)	Examples of areas to represent
0	1	360	156 543	1:500 million	whole world
1	4	180	78 272	1:250 million	
2	16	90	39 136	1:150 million	subcontinental area
3	64	45	19 568	1:70 million	largest country
4	256	22.5	9 784	1:35 million	
5	1 024	11.25	4 892	1:15 million	large African country
6	4 096	5.625	2 446	1:10 million	large European country
7	16 384	2.813	1 223	1:4 million	small country, US state
8	65 536	1.406	611.496	1:2 million	
9	262 144	0.703	305.748	1:1 million	wide area, large metropolitan area
10	1 048 576	0.352	152.874	1:500 thousand	metropolitan area
11	4 194 304	0.176	76.437	1:250 thousand	city
12	16 777 216	0.088	38.219	1:150 thousand	town, or city district
13	67 108 864	0.044	19.109	1:70 thousand	village, or suburb
14	268 435 456	0.022	9.555	1:35 thousand	
15	1 073 741 824	0.011	4.777	1:15 thousand	small road
16	4 294 967 296	0.005	2.389	1:8 thousand	street
17	17 179 869 184	0.003	1.194	1:4 thousand	block, park, addresses
18	68 719 476 736	0.001	0.597	1:2 thousand	some buildings, trees
19	274 877 906 944	0.0005	0.299	1:1 thousand	local highway and crossing details
20	1 099 511 627 776	0.00025	0.149	1:5 hundred	A mid-sized building

✿ Figure A.17 : Zoom levels and their associated scales in OSM. Source: https://wiki.openstreetmap.org/wiki/Zoom_levels

pertise. The goal is to reserve manual intervention for critical elements that cannot be automated, while the bulk of the map creation process is handled by advanced computational tools. In this subsection, we contribute to the existing state of the art by sharing additional insights gathered during the course of this thesis.

❖ **Interview with Plan IGN technical team**

To better understand the current process of building a pan-scalar map, we met Hervé Quinquenel, who is part of the IGN technical team composed of geographical information scientists. Hervé's team is a group of 10 people that conceive cartographic processes. They manage the data with postgres/postgis, make the queries, and furnish the results to the cartographers. If they need to create a new application or to manage servers, they ask the developer team. We conducted two free-form interviews of two hours each which provided us with further insights into the IGN Plan and pan-scalar map design in general, and highlighted the key issues they consider most important regarding their design. The key points from those discussions are summarized below.

❖ **About Plan IGN**

The "Plan IGN" is an exclusively online product, launched in 2019, and it has not undergone significant updates since. This product is often referred to as "pyramid cartography" because it utilizes multiple data layers, each corresponding to a specific zoom level. Depending on the scale, the map draws from various thematic layers of the BD Topo database, with some large-scale information coming from cadastral data as well. To produce the Plan IGN, an intermediate database—a cartographic database—is created from selected BD Topo information. While automated generalization algorithms (especially elastic beams) were implemented and used by IGN in the early 2000s to address issues such as feature superposition, highway interchange generalization, and toponym placement, the current process involves only minimal generalization (mainly for toponym placement) and instead relies heavily on data selection.

❖ **The four major challenges for plan IGN**

1. Placement of toponyms: While some algorithms exist and are implemented in ArcGIS, it needs human verification and usually intervention as well in cluttered areas to maintain readability.
2. Road overlapping with other networks: After generalization and stylization, roads often overlap with other networks such as railways, rivers, and embankments, which complicates visual clarity.
3. Mountain Roads Generalization: Generalizing mountain roads presents unique challenges due to terrain complexity and the need to preserve road connectivity and visual coherence.
4. Locating problems in the map: Last but not least, the IGN team told us that correcting map issues manually was usually not difficult, but that the primary challenge lies in detecting where these issues occur in the first place.

❖ Pan-scalar map building steps

We also discussed with the production team the key steps for building a pan-scalar map such as plan IGN. The technical architecture of a pan-scalar map can vary from one implementation to another, but the fundamental steps remain largely the same:

1. Define the purposes of the map, the possible user interactions, and the geographical scope (is the map limited to a specific location or aim to display the entire world?)
2. Gather the geographical data that is the most suited to the project, and structure them within a database
3. Decide which features appear at which scale (including toponyms), how they will be symbolized, and what degree of generalization will be involved. This can be done through a Scalemaster diagram⁶
4. Create the interface through which the users will interact with the map
5. Render the map on an engine, usually using a tiling system to break the map into smaller, manageable tiles
6. Update the map and its interface iteratively with new geographical data and user feedback

Compared to other GIS projects, steps 3 and 4 are specific to pan-scalar maps and are particularly influential in addressing the issue of disorientation. Our work focuses on the feature representation aspect (step 3), but we discuss the evolution of map interfaces in Section B.3. Feature representation involves several key sub-steps (not ordered):

- Determine the number of zoom levels and what data will be represented at each
- Develop a consistent visual language for the map, including the choice of symbols used to represent various geographic elements. This includes:
 - Point Symbols: For points of interest such as cities or monuments
 - Line Symbols: For features such as roads and rivers
 - Area Symbols: For regions such as parks and forests
 - Text and Labels: To provide additional information
- Carefully select colors and contrasts to enhance readability and distinguish between different data types
- Establish rules and processes for cartographic generalization

Additionally, this step may involve data enrichment tailored to the specific needs of feature representation and cartographic generalization (see Section A.3.2.1).

⁶Brewer and Buttenfield 2007 proposed a tool to assist in the design of multi-scale maps called the Scalemaster. It consists of a diagram that describes the behavior of map feature symbols across the different scales of the map being constructed. This tool helps cartographers quickly understand what elements should appear at each scale, facilitating informed design decisions.

A.2.4 Focus on map generalisation

Map generalization, that we already mentioned a few times, is (or should be?) so important in pan-scalar map design that it deserves a focus. Formal research on map generalization has been reviewed extensively from the 80s (Brassel and Weibel 1988; Mackaness, Ruas, and Sarjakoski 2007; Burghardt, Duchêne, and Mackaness 2014), and a recent review of automated map generalization techniques has been made in Zhang, Touya, and Meijers 2024. This section is not about detailing the literature as generalization is not core in our thesis but about introducing the concept and how it may impact user disorientation.

A.2.4.1 Introduction

❖ Origin

When representing geographic information at a small scale (whether on paper or on screen), cartographers quickly encounter challenges such as clutter, where an overwhelming amount of information makes the map difficult to read (see Section B.2.1.2), and overlapping, where the proximity of elements leads to partial or complete superposition of map elements, obscuring critical details. To maintain the readability of a map while conveying the maximum amount of relevant information, cartographers engage in a process of representing information more effectively through abstraction, known as cartographic generalization.

By nature, cartographic generalization has existed for as long as cartography itself. The discipline began to be formalized during the 20th century by National Mapping Agencies (NMAs), which required more rigorous methodologies to mass-produce maps and update them efficiently. With the advent of digital cartography, the field expanded significantly as the volume of collected data and the frequency of its updates increased dramatically. This development led NMAs to explore the possibility of automating the generalization process.

❖ Generalisation is not just simplification

Cartographic generalisation involves the careful selection and representation of map elements, sometimes requiring deviations from strict geographical accuracy to enhance clarity and usability. A well-generalized map not only helps users navigate more easily but also aids in forming a good mental representation of the space. This process is especially critical for topographic maps, which are inherently dense as they aim to comprehensively represent the terrain. While cartographic generalization is sometimes viewed as a mere simplification of data, we believe it is more accurate to see it as a method of delivering the right information, tailored to each scale (Touya, Potié, and Mackaness 2023). Map generalization should also seek to preserve or even enhance the visualization of relationships between map elements⁷, as these relationships are crucial for conveying meaning (see Section A.3.1.3).

Depending on the scale, this may involve omitting less relevant details or em-

⁷Mackaness et al., in "Abstracting Geographic Information in a Data-Rich World," page 6, stated: "If we consider a map to be a system of relationships, then generalization is fundamentally about modeling, filtering, and visualizing those relationships."

phasizing certain features that become significant only at broader perspectives. For instance, while airports may not be particularly important on a city-scale map, they become interesting landmarks on a national map as map users may wish to know if a specific city is accessible by flight (Figure A.18). If map generalization changes the representation of map features, it is crucial however that users can easily connect a generalized feature they observe at one scale with the same feature, which may be generalized differently or not at all, at another scale.



❁ Figure A.18 : Lyon-Bron airport at multiple generalization levels on Plan IGN. While zooming out, the representation of the airport takes up more and more space relative to the size of the city and thus becomes more visible.

❖ Definition

Having those ideas in mind, we propose a more comprehensive definition of cartographic generalization:

Cartographic generalization is the process of adapting the map's content according to the scale of representation to effectively convey the geographical information appropriate for that scale.

A.2.4.2 Operators

Cartographic generalization encompasses a variety of techniques referred to as **operators**. It is important to distinguish between generalization operations, which define the conceptual processes, and generalization algorithms, which are their specific implementations. Several taxonomies for generalization operations have been proposed (Mustière 2001; Foerster, Stoter, and Köbben 2007; Roth, Brewer, and Stryker 2011). Below is a description of some of the most common operations:

- **Simplification:** Reducing the complexity of lines and shapes to enhance readability at smaller scales.

- **Selection:** Choosing which elements to display based on their significance or relevance to the map’s purpose.
- **Combination (Aggregation):** Grouping similar elements to minimize clutter and simplify the map’s visual presentation.
- **Exaggeration:** Emphasizing important features by making them more prominent on the map.
- **Displacement:** Adjusting the positions of elements to prevent overlap and maintain clarity.

A generalization process applies these operators in sequence or sometimes all at once (Harrie and Weibel 2013; Regnauld, Touya, Gould, and Foerster 2014). The key challenges lie in determining the appropriate intensity of each operator, the optimal order in which they should be applied, and resolving the various conflicts or problems that these operators can create. Significant efforts have been made to automate these processes, particularly through the development of ontologies for generalization algorithms, such as the Scalemaster 2.0 framework proposed in Touya and Girres 2013, and Touya 2020a.

A.2.4.3 Problems

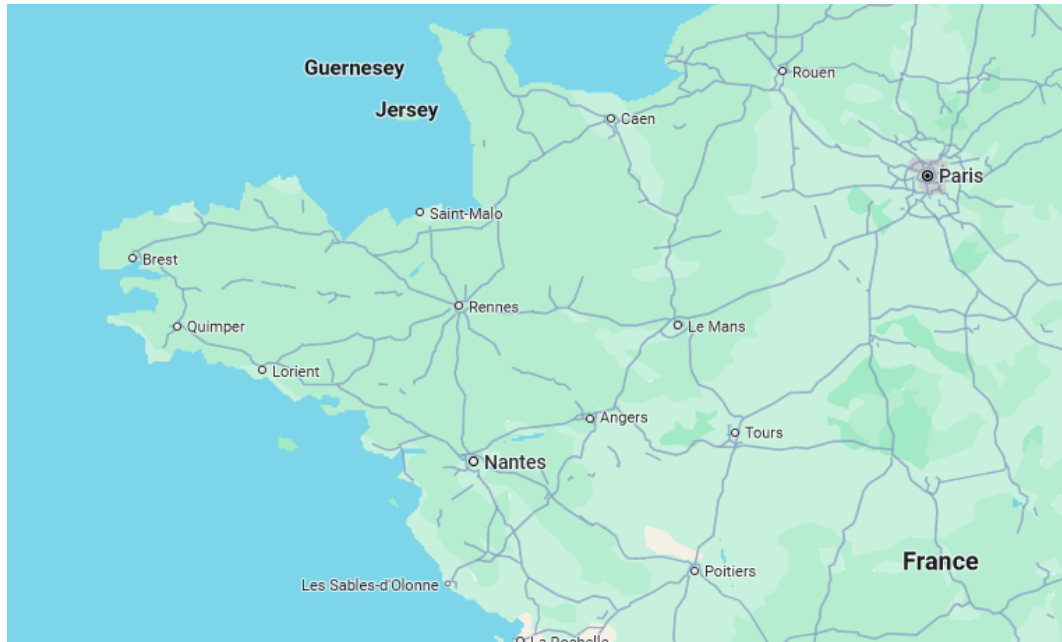
In addition to the two map generalization issues we have already discussed—the change in data representation and the possible absence of landmarks in certain areas—pan-scalar maps introduce a range of smaller problems that, when combined, can significantly hinder users’ ability to navigate. Although some generalization challenges, such as symbol overlapping, are less problematic in pan-scalar maps due to the ability to zoom in for more detail, these minor issues accumulate and impact overall usability.

❖ Example 1: Incomplete road network

Let’s take the example of generalizing a road network. When applying a progressive selection process based on criteria such as road classification, there is a risk of creating a disconnected network—where segments of roads become isolated, with no connection to other parts of the network (Figure A.19). This scenario can undermine the map’s utility, as it presents an incomplete and fragmented representation of the infrastructure. First, users may struggle to determine whether one city is connected to another by a road. Moreover, the absence of visible connections in some areas may fail to invite users to zoom in, as these empty regions provide no indication that additional details are available at larger scales. Secondly, such network disconnections result in major inconsistencies between road representations across different scales, potentially inducing a "desert fog" effect for users. Finally, such anomalies may capture users’ attention and distract them from their initial task⁸.

⁸Attentional capture is a concept in the scientific literature that refers to situations where observers are unable to disregard a stimulus that is irrelevant to their task. However, recent research has proposed a more nuanced definition to refine the scope and direction of investigations in this area (Anderson 2021).

Proper attention must thus be given to maintaining connectivity to ensure that the map remains functional and accurately represents the real-world network.



❁ Figure A.19 : Western France on Google Maps at zoom level 6. Disconnected road segments can be seen, isolated from the main network. Unfortunately, it is not possible to determine the specific road selection process used by Google.

❖ Example 2: River meanders

When trying to preserve certain patterns a dilemma may arise, as it may sometimes require sacrificing the exact geographical accuracy. For instance, in the generalization of river meanders for smaller cartographic scales⁹, it's important to thicken the river's representation to ensure it remains distinguishable at smaller scales, but it is also important to maintain the river's sinuous nature so that users can easily recognize it. However, retaining every twist and turn of the river can lead to overlaps due to the increased line thickness, which can clutter the map. In this case, the best solution for representing this river at a smaller scale may be to reduce the number of its meanders, even if it means deviating from the exact geographical reality (Figure A.20). To determine the extent to which a river's meanders should be simplified and how to approach the process, it is first necessary to characterize the unique features of the river. Stanislawski, Kronenfeld, Buttenfield, and Shavers 2023 provides a highly detailed methodology for this purpose.

⁹We can only reduce the size of a feature that much for it to stay understandable. Moreover the eyes do not perceive details on screens as well as on printed paper (Ledermann 2023).

**Original
scale**

**Reduced
scale**

**Generalised
version**



⌘ Figure A.20 : Example of a river representation and its generalised version at smaller scale. While the main pattern is conserved, some meanders are removed in order to increase the readability. Source: Ordnance Survey (Great Britain’s NMA)

A.3 Landmark detection in maps

Now that we have reviewed the concepts of pan-scalar maps and cartographic generalization, along with their associated literature, it is time to focus on the scientific domain at the core of our research: the detection of patterns in geolocated data, with a particular emphasis on landmarks.

A.3.1 Human perception of the map: from perception to meaning

Landmarks, as we have defined them, are intrinsically linked to human cognition. Therefore, to better understand the various existing approaches to detecting landmarks, we must first explore how humans perceive landmarks and maps.

A.3.1.1 Map abstraction

The power of maps¹⁰ lies in their abstraction, allowing cartographers to represent the seemingly infinite details of the real world using a finite amount of ink or pixels (Touya, Potié, and Mackaness 2023). It is the capacity to abstract the world that allows cartographers to make maps more comprehensible than orthophotos (Figure A.21). However, this abstraction only becomes meaningful if users can successfully interpret it. This raises an important question: how do humans perceive maps and derive meaning from them?

¹⁰..or “super-power of maps”, as Georg Gartner suggested during the Eurocarto 2024 conference



✿ Figure A.21 : Aerial photograph (left) of the Rhône River and the Alpilles mountain range, compared with a map (right) of the same area generated from vector data. While modern technology allows us to photograph the entire Earth, a photograph alone does not convey the same level of detailed information as a carefully designed map, which can effectively highlight important geographical features and relationships. Source: Plan IGN, géoportail (<https://www.geoportail.gouv.fr/carte>)

A.3.1.2 Cartographic communication and cartographic knowledge

The link between map perception and interpretation and human cognition has been established and have been studied for decades (Board and Taylor 1977). When users look¹¹, their brains process the visual information, interpreting it as geographic data based on their understanding of cartographic language and their prior knowledge of geography.

The field of cartography has been greatly influenced through the second half of the 20th century by Robinson 1952 and the concept of “cartographic communication”. **Cartographic communication** involves interactions between cartographers and users and the effective transmission of spatial information, ideas, and messages through visual elements in maps to ensure that the intended meaning is clearly understood. Theoretical cartographic communication models have been proposed to describe the interactions between the cartographers, the maps, the users, and a so-called “cartographic language” (Koláčný 1969). The **cartographic language** refers to the set of visual and symbolic conventions used in maps to communicate spatial information effectively. Just as spoken or written language uses words and grammar to convey meaning, cartographic language employs symbols, colors, lines, and shapes to represent geographical features and spatial relationships. While those models remain mostly theoretical, Beitlova, Popelka, Konopka, and Macku 2023 proposes an eye tracking experiment to demonstrate the accuracy of cartographic communication models.

¹¹While there are some projects aiming to incorporate senses like touch and sound (Brock and Jouffrais 2015; Wabiński, Mościcka, and Touya 2022) at a map, most maps rely solely on visual elements and the vast majority of map perception occurs through visual processing.

A.3.1.3 Meaning emerges from structures and patterns

It is the spatial and topological arrangement of the map symbols that conveys meaning to the user (MacEachren and Garter 1990; Mackaness and Edwards 2002). For instance, when we examine the map on the right of figure (Figure A.21), we don't merely see isolated orange lines; instead, our minds connect these lines to reveal the road network as a coherent structure. We also understand that the main road passes to the north of the Alpilles massif, skirting the towns of Beaucaire and Tarascon, which are themselves separated by the Rhône River. These elements can act as landmarks and are fundamentally important for the entire process of map understanding and building our geographic understanding.

In fact, many notable elements of the map are compositions of seemingly minor objects that together form a larger element on the map, like train stations, airports, hospitals, etc (Figure A.22). These macro objects are referred to as **structures** and have different relations with their environment depending on the scale we are looking at them (Touya, Potié, and Mackaness 2023). Additionally, some elements of the map stand out from others of the same type because of their pattern. We use the definition of Mackaness and Edwards 2002 for “**pattern**”: “a pattern is a property within an object, or between objects that is repeated with sufficient regularity”. Such repeated properties may be shape, orientation, connectedness, density or distribution. For example, common patterns in maps are buildings alignments or river meanders.



❁ Figure A.22 : The Montpellier train station is a map landmark composed of a large blue building intersected by railways. Source: Plan IGN

So geographic patterns and structures, such as road networks, river systems, or building layouts, are fundamental to how users understand maps and may be themselves pan-scalar anchors. Map generalization should therefore aim to preserve these patterns but also model the spatial relationships between them. This includes preserving topological, geometric, and semantic relationships to ensure that the overall

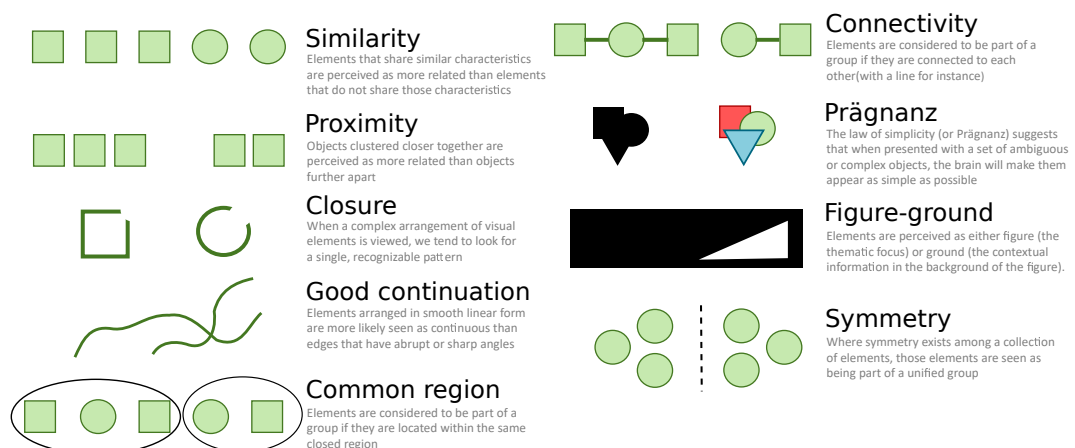
map still makes sense to the user at different scales (Mackaness and Edwards 2002). The human brain processes structures and patterns pre-attentively, making the task feel effortless to us. However, in the context of automated map design, understanding precisely how the brain identifies and interprets these patterns becomes crucial in order to develop maps that are more intuitive and efficient for users to navigate. Let us therefore discuss the Gestalt theory.

A.3.1.4 The Gestalt theory

Gestalt theory explores how the brain groups symbols together based on their visual properties, such as proximity, colour, and shape. Gestalt psychology began in 1912 with the work of Max Wertheimer. The concept gained significant prominence in the second half of the 20th century, particularly influencing many cartographers. However, as modern cognitive sciences developed, Gestalt's approach gradually became overshadowed, though its impact on the field remains notable (Wagemans, Feldman, Gepshtein, Kimchi, Pomerantz, Helm, and Leeuwen 2012).

Several phenomena have been recognized and presented as "Gestalt laws," some of which are illustrated in Figure A.23. These laws are central to the perception and interpretation of maps, and some have already been integrated by researchers and cartographers to improve the generalization process, ensuring that maps remain coherent and informative. For example, these laws are applied to preserve specific arrangements in urban areas (Li, W, Ai, and Chen 2004) or to maintain the continuity of road networks.

From a designer perspective, these laws suggest that added landmarks do not need to be entirely visible to be effective in aiding orientation and navigation. On the flip side, if we aim to detect landmarks algorithmically in maps, it is essential to consider these Gestalt principles, which can introduce a layer of complexity in their implementation.



✿ Figure A.23 : Illustration of Gestalt so-called laws. Source: Touya, Potié, and Mackaness 2023 (made by Guillaume Touya)

A.3.2 Pattern detection

Our thesis focuses on identifying landmarks as potential anchors to enhance pan-scalar maps. However, landmarks often emerge from patterns and structures rather than standalone elements. Detecting landmarks or map features that could serve as landmarks thus involves identifying these underlying patterns and structures. This section introduces methodologies and techniques for pattern detection and is inspired from the literature review we did in the third part of Touya, Potié, and Mackaness 2023. We invite interested readers to read the paper for more details and references.

A.3.2.1 The importance of data enrichment

Richter and Winter (2014) highlight a key limitation of conventional databases: while people naturally perceive and recall environments as interconnected patterns, geographic databases often store individual objects without explicitly representing these structures. Data enrichment is a key step that seeks to bridge this gap, as Mackaness, Burghardt, and Duchêne 2014 describe at page 6:

“**Data enrichment** is the process by which we make explicit those relationships implicit in the patterns and associations among features represented on the map [...]”

Data enrichment may operate at two levels:

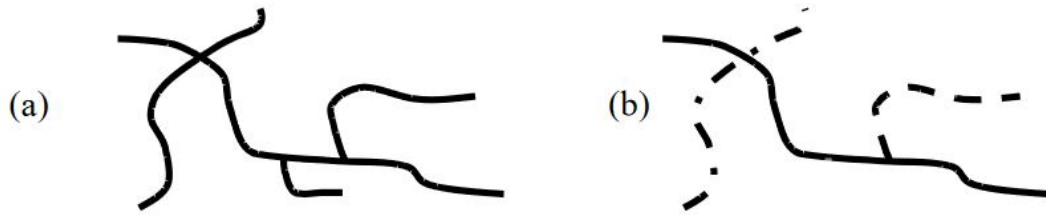
- **Enhancing Individual Objects:** Adding attributes to individual database elements, such as computing building orientation (Yang 2008) or many other building properties (Elias 2003).
- **Creating New Composite Objects:** Identifying patterns or structures formed by multiple objects, such as road strokes (Thomson 2006) (Figure A.24), road intersections (Savino, Rumor, Zanon, and Lissandron 2010), highway interchanges (Touya and Lokhat 2020), or building alignments (Christophe and Ruas 2002; Li, W, Ai, and Chen 2004).

Data enrichment is core in cartographic generalization and in pattern detection. Depending on the goal of the enrichment, the enriched information may be directly used by another algorithm or stored in a database for later use.

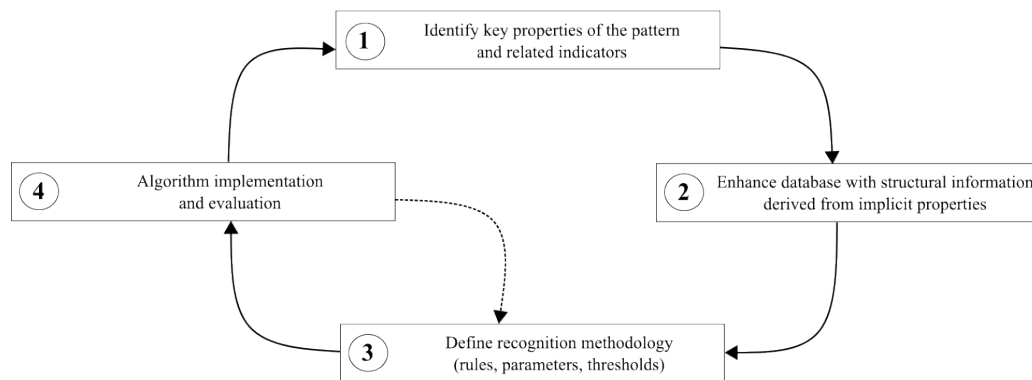
A.3.2.2 Pattern detection algorithms draw inspiration from one another

While the task of detecting patterns may seem complex given the variety of different patterns in maps, it is encouraging that various pattern detection algorithms share common processes, drawing from spatial analysis, graph theory (when studying networks or specific agencements of objects) and sometimes machine learning techniques. By studying the literature, we conclude that a standard process for building a new algorithm that allows us to detect structures and patterns is usually done in four steps (Figure A.25), which we detail in Touya, Potié, and Mackaness 2023.

As another encouraging point, we should notice that many patterns exhibit fractal-like properties, recurring at various geographic scales (Lam and Quattrochi



✿ Figure A.24 : Strokes consist of linear elements that seem continuous through junctions. The road network example (a) consists of 8 arcs and 9 nodes, but 3 strokes may be discerned as we can see in (b). Source: Thomson 2006



✿ Figure A.25 : A four-step approach to implementing a pattern recognition algorithm. Source: Touya, Potié, and Mackaness 2023

1992). For instance, the star-shaped pattern formed by roads around a village can also be found around a capital city, albeit at a much larger scale. Similarly, small streams flow into rivers, which in turn flow into larger rivers, creating a branching network that resembles a tree, regardless of the scale at which the phenomenon is observed. This suggests that findings from studies on patterns at one scale may apply to others, as with ring roads discussed in Chapter 3.

A.3.2.3 Pattern detection with machine learning

Machine learning has seen rapid advancements in cartography during the first quarter of the 21st century, particularly in the field of pattern recognition (Harrie, Touya, Oucheikh, Ai, Courtial, and Richter 2024). Traditional pattern detection techniques based on spatial analysis require researchers to manually define the properties that characterize the structures to be detected, as well as any thresholds that distinguish between categories (e.g., buildings larger than a certain size or within a specific distance from a road). This approach works well when the characteristics of the target objects are clearly understood.

Machine learning, however, offers an alternative that reduces this manual effort by creating algorithms that rely on example data rather than predefined parameters. In supervised learning, the algorithm learns to identify patterns based on labeled training data, while unsupervised learning enables algorithms to independently distinguish different categories within a dataset without prior labeling. This flexibility

allows for the detection of objects even when their defining properties are unclear to cartographers or researchers. For instance, as we will see in Chapter C, a user may intuitively recognize a circular road as a ring road, but it remains challenging to define precise spatial properties for such identification (Potié, Mackaness, and Touya 2022). This is where supervised learning becomes particularly valuable. In this thesis, we will explore supervised learning approaches for detecting patterns in geographic data.

A.3.2.4 Deep learning for pattern detection and map generalization

Deep learning, a subset of machine learning, employs large artificial neural networks to perform complex tasks such as identifying structures within vast datasets (LeCun, Bengio, and Hinton 2015). This approach has revolutionized computer vision, significantly improving image classification and interpretation compared to traditional methods. For example, segmentation algorithms like U-Net (Ronneberger, Fischer, and Brox 2015) are capable of recognizing pixels belonging to specific objects in images, from either aerial photographs (Khan and Jung 2024), historical maps (Weinman, Chen, Gafford, Gifford, Lamsal, and Niehus-Staab 2019), or digital map images (Touya and Lokhat 2020). More recently, a novel deep learning model architecture using transformers—previously popularized in natural language processing—has also shown promise for image processing tasks (Dosovitskiy et al. 2021).

However, while images offer rich visual information, they lack the tabular details embedded in geographic databases. But deep learning has also been applied directly to geographic data, with notable successes in simplifying geometries, such as building footprints (Sester, Feng, and Thiemann 2018). Yet, these methods often struggle to account for spatial relationships through their relative positioning on the map. One of the main challenges lies in the vast number of examples required to train deep learning models to achieve accurate results. Producing these datasets is particularly complex for tasks such as cartographic generalization, where evaluation metrics remain subjective and task-specific (Touya and Lokhat 2020).

The encoding of geographic data into tensors—the input format for deep learning models—is crucial in the process of implementing a deep learning model, as it directly affects how well the model can incorporate and link information. When training models based on map images, (Courtial, Touya, and Zhang 2022) proposed separating geographic information into distinct layers that are then fused during encoding. This technique ensures the model retains access to all information, even in overlapping areas, such as when a bridge passes over a river. Without this layered approach, a simple vector-based conversion might lead the model to perceive a river as discontinuous. We think that this principle could be adapted for landmark detection, as landmarks in images are often complex structures that may be crossed by others.

Despite its potential, one major drawback of deep neural network (DNN) architectures is their "black-box" nature—it is often difficult to understand why a model produces a specific result, even if the output appears accurate. Recent advancements in explainable AI (XAI) seek to address this issue by analyzing how DNNs trained on image datasets recognize shapes (Biscione and Bowers 2023) and what geographic knowledge they acquire (Fu, Zhou, Xin, and Weibel 2024). Preliminary findings

suggest that current models may not yet achieve the high-level abstraction seen in human cognition. They are, however, capable of recognizing basic concepts such as proximity between elements. This progress hints at a future where AI models may develop more advanced spatial reasoning capabilities and pushed us to experiment with deep learning model to detect landmarks in Chapter 5 and 6.

A.3.2.5 Landmark detection

In the literature on using machine learning to detect landmarks, Elias 2003 proposed a pioneering methodology for identifying remarkable buildings near road intersections, aiming to utilize them as landmarks in navigation software (Figure A.26). This approach involved enriching building data with numerous attributes, followed by two types of machine learning algorithms. First, a classification algorithm (supervised learning) was employed to distinguish buildings into two categories—landmarks and non-landmarks. Second, a clustering algorithm (unsupervised learning) was used to identify distinctive buildings at intersections without relying on labeled examples. This work laid the foundation for further research on enriching navigation instructions with landmarks (Duckham, Winter, and Robinson 2010; Rousell and Zipf 2017). A comprehensive review of this topic is available in (Yesiltepe, Conroy Dalton, and Ozbil Torun 2021).



✿ Figure A.26 : Selection of building candidates for landmarks near a road intersection. Source: (Elias 2003)

In Particular, (Lin, Cai, Gong, Kang, and Li 2019) and (Kang, Liu, Wang, Li, and

Weng 2022) demonstrated the effectiveness of random forest classifiers for extracting urban landmarks, a method we also explore in Chapter 4. While our objective differs—it focuses on leveraging landmarks to improve navigation within pan-scalar maps rather than providing instructions during GPS-guided navigation—we draw inspiration from the enrichment and distinction methodologies developed in these studies.

Some researchers have also explored detecting landmarks based on how people refer to them in social media posts and online forums (Quesnot and Roche 2015). This semantic approach shows promise, as it leverages public discourse to identify culturally or contextually significant landmarks. However, it diverges from the primary focus of this thesis, which centres on detecting landmarks directly within maps and their associated databases. While the semantic approach may complement map-based methodologies, our work emphasizes spatial and structural properties over user-generated textual content.

A.3.3 Conclusion on landmark detection

Encoding the concept of a landmark in computational terms is a significant challenge, as it is intrinsically tied to human perception. Machine learning algorithms appear to be well-suited for this task, given their ability to work with implicit and hard-to-define concepts. However, the creation of training datasets for supervised algorithms remains a critical challenge in the field.

Landmark detection has been particularly studied in the context of navigation applications, focusing on identifying distinctive elements at intersections. This approach aims to enrich wayfinding instructions with meaningful references. Additionally, landmark detection is also a topic of interest in cartographic generalization research. However, it focuses more on preserving shapes and arrangements to minimize visual and structural disruptions between different levels of generalization rather than on identifying potential anchors.

A.4 Conclusion of the chapter

In this chapter, we introduced the concept of (pan-scalar) map landmarks, which assist users in navigation, and the concept of (pan-scalar) anchors, their equivalent within the user’s mental map. We introduced the concept of pan-scalar maps and discussed how disorientation occurs when there is a mismatch between the user expectations and the actual content displayed on the map, especially during scale transitions in the exploration of pan-scalar maps. Lastly, we reviewed pattern detection techniques employed to enrich databases with landmark information, which offers new avenues for improving the generalization and design of pan-scalar maps in a more systematic and scalable way.

A.5 Thesis orientation

This thesis is guided by results from the conclusion that disorientation issues in pan-scalar maps are a genuine concern, but that they can be mitigated through improved cartographic generalization and zooming interaction design. Although there has been substantial work on the role of patterns in map perception and on preserving these patterns during the generalization process, few studies attempt to bridge these two domains and our work tries to fill this gap. In particular, the anchor point theory from Couclelis, Golledge, Gale, and Tobler 1987 makes us feel there is something to be achieved by exploring the connection between pan-scalar landmarks in maps, which can be enhanced by cartographers, and the cognitive anchors formed in users' mental maps.

A.5.1 Problems

Our central problem arises from this parallel between landmarks in maps and cognitive anchors in mental maps:

Can we design pan-scalar map landmarks in a way that influences the formation of cognitive anchors in users' mental maps?

From this problem, we derive three core research questions:

- Can user perception studies help determine which map features should be emphasized, and under what contexts?
- How can we translate findings from perception studies into material adapted to be used in generalization and detection algorithms?
- To what extent can supervised machine learning models be utilized for detecting landmarks in maps?

To address these questions, we will explore a combination of user perception studies, cartographic data enrichment processes, and machine learning-based pattern recognition techniques. Through the experiments conducted in this thesis, we aim not only to explore new techniques to enrich geographic databases to improve map generalization, but also to delve into the realm of spatial cognition. By questioning participants about why certain map features stand out more than others, we seek to understand which elements are most likely to crystallize into anchors within their mental maps.

A.5.2 Overview of the experiments

In addition to the preliminary studies proposed in Chapter B, this thesis will be driven by the following experiments:

❖ **When is a ring road a ring road? (Chapter C)**

Urban ring roads, or "city ring roads", appear to be high-potential pan-scalar landmarks. However, not everyone has the same definition of an urban ring road. We conducted a study on the perception of these ring roads based on a questionnaire with 28 participants to better characterize this potential anchor and thus help in their detection.

❖ **Multimodal learning for remarkable building detection (Chapter D)**

Salient buildings are real life landmarks and have the potential to be good map landmarks if their representation is enhanced. In this machine learning experiment, we associate each building with a topographic map image where it is centered and use a multimodal deep learning model to classify remarkable buildings from unremarkable ones.

❖ **Potential anchor detection in a map image with deep learning (Chapter E and F)**

The human eye quickly differentiates important elements from less significant ones in a map image. In this experiment, we train deep learning models like CNNs and Segformers to evaluate map images and test our models on a custom evaluation set created from expert annotations.

Chapter B

Preliminary studies

In this chapter, we present three preliminary studies we conducted. The first study, conducted in collaboration with Maïeul Gruget, analyzes the existing pan-scalar landmarks and attempts to create an inventory of those already present in pan-scalar maps. The second study investigates the question of the optimal density of landmarks in pan-scalar maps. Finally, the third study provides a review of the evolution of pan-scalar map interfaces and possible user interactions since 2005. While this last study is not directly focused on landmarks, we believe that understanding how people interact with pan-scalar maps is crucial for improving their design.

B.1 What are the pan-scalar landmarks in contemporary pan-scalar maps?

Before discussing how we could add or enhance landmarks to improve the maps, we must first examine what already exists. In this preliminary study, we analyzed the widely used pan-scalar maps available to us¹ to identify the elements that we believe serve as the most frequent pan-scalar landmarks. The idea behind this study is to attempt to list all types of pan-scalar landmarks comprehensively so that each type can later be examined individually in further detail. By cataloging these different types, we hope to better understand how they contribute to the overall map experience and how they may become cognitive anchors in the mental maps of users.

B.1.1 Methodology

My colleague Maïeul Gruget² and I systematically explored Google Maps, OSM (OpenStreetMap), and the IGN map, zooming in and out across various locations worldwide (Except for the IGN map which is limited to France), between zoom levels 5 and 20. For each map and at each zoom level, we intentionally navigated through

¹Unfortunately, we didn't have an Apple machine available to carry out this study and the other experiments on the Apple map.

²Also a PhD student in the LostInZoom project.

urban, semi-urban, and rural areas. During these explorations, we focused on identifying the major types of pan-scalar landmarks that could help users orient themselves across different scales. We focused on recurring geographic features, avoiding landmarks that are unique to specific areas, and we tested them in various locations to confirm their consistency. We also deliberately avoided areas we were personally familiar with to reduce bias. Our findings were recorded in a Scalemaster-like diagram (Brewer and Buttenfield 2007), where the rows represent the types of features on the map and the columns indicate the zoom levels. The coloured bands do not simply show when the features are visible, but rather the range of zoom levels at which the features are salient enough to serve as pan-scalar landmarks in contemporary maps (Figure B.1).

B.1.2 Results

We observed that the pan-scalar landmarks across the three studied maps—Google Maps, OSM, and the IGN map—are largely similar, with only a few specific landmarks unique to Google Maps (country names, High summits and points of interest). Unsurprisingly, the main features of the pan-scalar maps are road networks of various sizes, rivers, lakes, and other water bodies, urban areas, and specific natural spaces such as national parks and forests.

Interestingly, we note that toponyms (place names) also serve as important landmarks. According to IGN cartographers, the placement of toponyms is one of the challenges of automated cartographic generalization, as they need to be positioned in a way that clearly indicates the locations they represent without overlap on other critical map information. Adding to this complexity, we found that toponyms act as pan-scalar landmarks, meaning their position needs to be also preserved across different zoom levels to avoid disorienting users. Work has already been conducted in this domain. For example, Been, Nöllenburg, Poon, and Wolff 2010 proposed modelling labels as solids in a three-dimensional space, using scale as the third dimension, and Wu, Takahashi, Poon, and Arikawa 2017 introduced a method for automatically positioning toponyms across scales using a genetic algorithm.

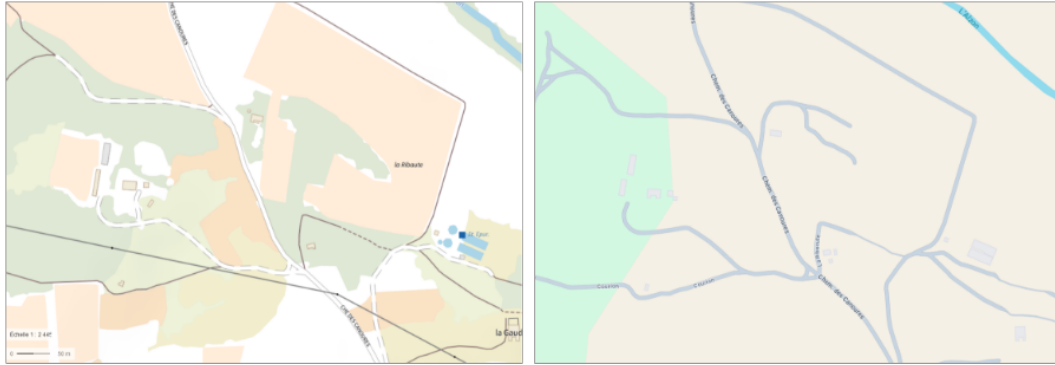
Another observation worthy of mention is that land cover patches (areas with distinct color-coding representing terrain or vegetation types) serve as visual cues across scales. However, as mentioned in Section A.1.3.3, the ambiguity in interpreting their meaning and the subjective nature of their boundaries make them unlikely to form cognitive anchors in users’ mental maps (Figure B.2).

Based on our observations, all scales include a certain number of landmarks, with issues arising primarily in areas that are either particularly empty or overly dense. In the context of our thesis, where we aim to leverage some of these elements to improve zooming interactions, the question of detection becomes critical. Some landmarks we identified are more easily detectable than others from an algorithmic perspective.

Among the potential landmarks we identified, mountain ranges appear to be less straightforward to highlight. However, a study of Digital Elevation Models (DEMs) could be considered to emphasize certain terrain features in map design. On the other hand, elements based on graph networks are generally more readily analyzable, such as ring roads (which will be the focus of Chapter C), highway interchanges, or railway lines.

	Landmark			Potential landmark if representation is improved			Landmark specific to Google Maps				
Zoom level	16	15	14	13	12	11	10	9	8	7	6
Country boundaries											
Country shapes											
Capitals											
Country names											
Mountain ranges											
Large land cover patches											
Oceans and seas											
Large islands (e.g. Corsica, or Jamaica)											
Coastline											
Large cities >300k											
Small seas, large lakes											
Secondary islands (e.g. Cypruss)											
Highways											
Large natural parks/regions/spaces											
Main rivers											
main train lines											
Secondary cities (>100k)											
Large conurbations											
ring roads											
Small islands											
Secondary rivers											
large forests											
Medium lakes											
High summits											
Small towns (>20k)											
Summits, passes											
Secondary roads											
Sinuuous rivers											
Isolated roads											
points of interest											
Road interchange											

☞ Figure B.1 : Summary of the pan-scalar landmarks and their range of apparition in Google Maps, OSM, and Plan IGN.



⌘ Figure B.2 : Rural area north of the town of Colias as seen on the IGN map (left) and Google Maps (right). Land cover patches can serve as visual cues during navigation. However, the colors are often difficult to interpret and vary between different cartographic representations, making them less likely to become anchors in users' mental maps than other map features.

B.1.3 Conclusion

One might naively conclude that the features identified in our table are the map elements with the most potential and should therefore be the primary focus of further analysis. However, it is crucial to remember that the symbology used to represent these elements significantly influences our perception of them. The map designers have deliberately chosen to emphasize these elements in the studied maps. Therefore, we cannot directly conclude that these specific types of elements inherently possess the greatest potential. Nonetheless, we can observe that these features are among the most prominent in modern digital maps, which means that users are accustomed to perceiving them across various scales. This habitual exposure to these features must be considered when analyzing their role in map navigation and cognition and when designing future maps.

B.2 What are the optimal densities for pan-scalar landmarks and visual cues in maps?

We have seen in chapter A that, for good navigation, map users need 1) to see landmarks to recognize the area by connecting them to their mental map and 2) to get instant feedback about how much they pan or zoom by seeing visual cues.

With automated map generalization, it is technically possible to modify the representation of map features based on their surrounding context, allowing, in theory, the automatic creation of landmarks by emphasizing certain features. This capability poses an important question for map designers: under what circumstances should

this enhancement be applied?

Determining the ideal number of landmarks and visual cues in maps is complex and involves multiple factors such as navigation task, geographic location, zoom level, screen size, user experience and fatigue, or the symbology used for the landmarks (Coltekin and Clarke 2011). To simplify the question, we propose to explore the lower and upper limits of a reasonable density of features in pan-scalar maps by reviewing the relevant literature and conducting a small thought experiment.

B.2.1 Literature review

B.2.1.1 Landmark density and map complexity

While we did not find a study that directly proposes a specific optimal density for landmarks in maps, several works explore the impact of landmarks on navigation tasks, such as street view itinerary following and route planning (Steck and Malot 2000; Kapaj, Lin, and Lanini-Maggi 2022; Cheng, Wunderlich, Gramann, Lin, and Fabrikant 2022). In particular, Bruns and Chamberlain 2019 and Keil, Edler, Kuchinke, and Dickmann 2020 demonstrated that, during route planning, landmarks closest to the planned route are more frequently used and remembered by users, as opposed to those further away but still visible on the screen. Furthermore, Keil et al. found that the lower the visual complexity of a map, the more users tend to search for landmarks located farther from the route. This suggests that the optimal density of landmarks should correlate with the map's complexity. For instance, fewer landmarks may be needed in rural areas, while more may be necessary in urban environments due to their higher complexity.

B.2.1.2 Clutter and visual overload

Though there is almost no technical limit to the number of elements that can be placed on a single cartographic view, practical and cognitive limitations must be considered when designing maps. An overabundance of visual elements can cause confusion and hinder the ability to locate specific reference points (consider, for example, the overwhelming effect in "Where's Waldo?" images). In maps, an excess of visual information can result in clutter, making the map harder to interpret. Rosenholtz, Li, and Nakano 2007 define **clutter** as "a state in which excess items, or their representation or organization, lead to a degradation of performance at some task." They also provide a review of methods for measuring clutter in images, while Touya, Hoarau, and Christophe 2016 discusses specific approaches for quantifying clutter in maps. Clutter is an important factor to consider in the design of landmarks, as higher levels of clutter make it increasingly challenging to introduce a visually distinct element. When the visual space is crowded, it becomes harder for users to identify and focus on a new standout feature. This is particularly relevant for small screens, where space is limited, and users need clear, concise information without overlapping elements or excessive detail.

B.2.1.3 Competition between landmarks

Chamizo, Manteiga, Rodrigo, and Mackintosh 2006's experiments with rats showed that there is not only a hierarchy of landmarks based on their proximity to a target, but also competition between landmarks during the learning process. This implies that the memorability of a landmark can be reduced when another landmark is introduced, which has important implications for cartographic design. Adding too many landmarks may interfere with the user's ability to retain critical ones. Chamizo's findings align with Miller 1956's famous article on cognition, "The Magical Number Seven, Plus or Minus Two," which demonstrates the limitations of the human mind in processing multiple pieces of information, typically around five to seven items depending on the task. Similar numbers were observed in experiments on itinerary following in 3D environments, where users remembered only a limited number of landmarks. Hence, cognitive competition must be considered when deciding how many and which landmarks to emphasize on a map.

B.2.2 Minimum visual cues for good orientation

B.2.2.1 Problem

While map users are typically responsible for initiating zooms and pans within pan-scalar maps, they may not always be fully aware of how far they have moved, either spatially or across different zoom levels. This issue can be observed in GIS or drawing software. Imagine a blank workspace with only a horizontal line representing a river, devoid of any other reference points. If you pan horizontally across the river, especially if the river is straight, it is almost impossible to know how far you have moved without any additional landmarks. Similarly, if you zoom in or out without other reference points, estimating the current scale becomes difficult. However, if a house was added next to the river, you could more easily judge your horizontal pan distance based on the movement of the house and gauge the scale based on the change in house size.

This scenario is more relevant to real maps than one might assume. When users quickly scroll through maps, they only process a limited subset of the available information. The brain can only handle and record a certain number of visual inputs at a time, and when the user has to rely on less prominent features for orientation, it results in a higher cognitive load. Therefore if, as map designers, we aim to emphasize specific landmarks, it would be ideal if they were sufficient to help users orient themselves both in terms of space and scale.

B.2.2.2 Thought experiment

In the second chapter of Richter and Winter 2014, the authors demonstrate the importance of landmarks by proposing a thought experiment in which a community is placed in a completely empty environment referred to as "the void." The navigational capabilities of this community are then explored as landmarks are progressively introduced. Inspired by this concept, we sought to determine the minimum number of visual cues required to maintain orientation while zooming or panning within a pan-scalar map. For each of the three main types of cartographic objects — point

symbols, lines, and polygons — we envisioned them in a blank pan-scalar map, through which a user would attempt to navigate (Figure B.3). Next, we tested various combinations of visual cues to observe how they interact with each other (Figure B.4).

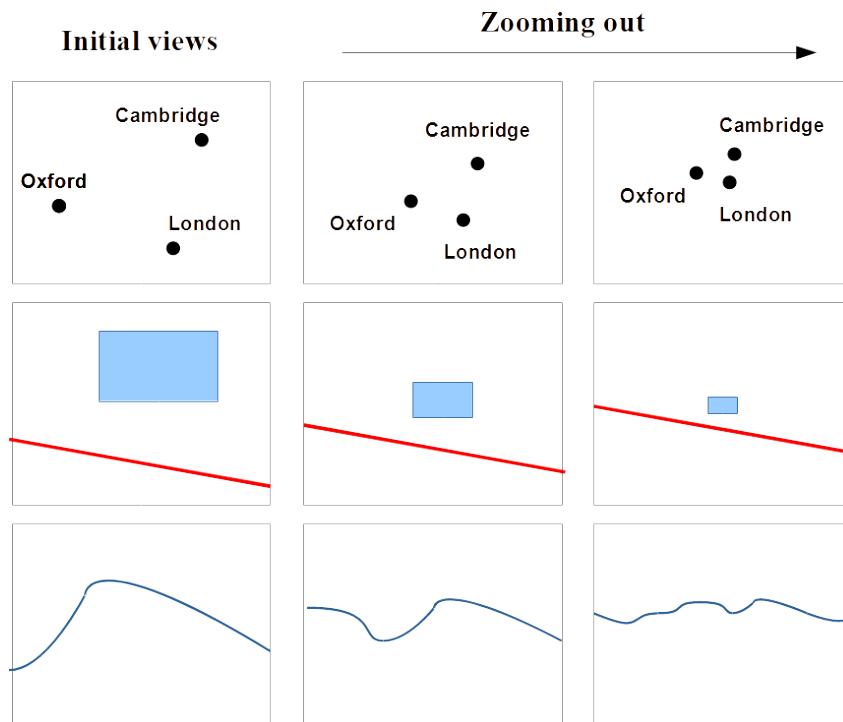


✿ Figure B.3 : Impact of zoom on three simple shapes commonly found in maps. Point and linear vector elements remain unchanged during zooming.

Our first observation is that the three primary types of cartographic objects do not provide the same level of orientation across scales and space³. Points and straight lines, when they are alone, do not allow users to comprehend the extent of a zoom. However, the change in size of a polygon can help users estimate the change in scale, even if the shape of the polygon affects the degree to which users can successfully complete this task.

Our second observation is that visual cues combine to provide more information. For example, two points placed side by side give insight into the change in scale through the variation in the visual distance between them during zooming. As zoom levels are usually built on a quad tree structure, the distance between two points double for every zoom level we zoom in. We also observe that in non-linear elements, such as rivers with bends, the bends can serve as landmarks to indicate where you are along the river as well as the scale at which you are viewing it. However, there is a risk that the user may confuse different bends in the river, leading to disorientation due to this ambiguity (see Section A.2.2.3).

³In fact, it would certainly be possible to theoretically study the geometries and examine their invariances through homothety, rotation, or translation along certain axes. However, this does not seem necessary to draw conclusions here.



✿ Figure B.4 : Impact of zooming on a combination of elements (lines 1 and 2) and on a non-linear element (line 3). The change in scale is discernible in all three cases.

We observed several properties specific to lines. As illustrated in the top-left image of Figure B.5, simply adding lines is not sufficient to convey information about the zoom level. If the lines intersect at the same point, the resulting star-shaped figure remains invariant during zooming. Instead, it is the number of intersections at different locations that provides meaningful information. Notably, a single intersection offers the same amount of information as an isolated point, meaning that two intersections are theoretically sufficient to indicate changes in scale. Additionally, bifurcations and inflection points along a single line share these properties, as demonstrated in the third and fourth lines of the figure.

Interestingly, we noticed that cartographic elements can also combine with interface elements. The most common zoom interfaces center the zoom action around the current position of the mouse cursor (see Section B.3). In this case, The mouse⁴ cursor can also serve as a temporary point of reference, acting as an additional landmark during zooming.

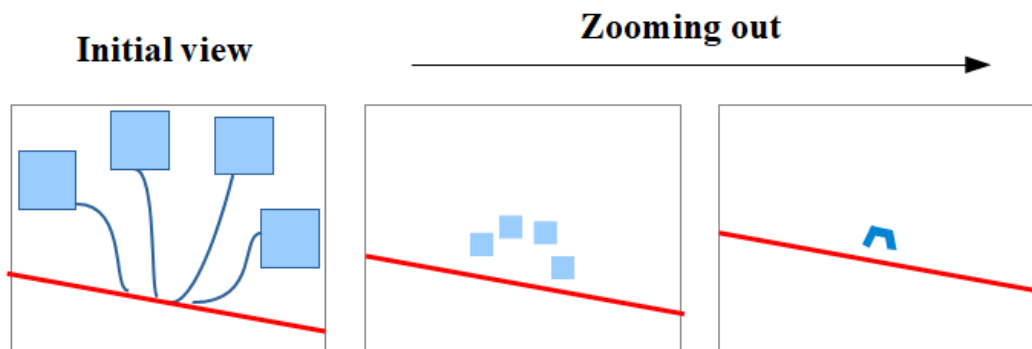
Beyond geometric shapes and their combinations, the change in representation of elements across scales, resulting from their generalization, can itself provide users with cues about the scale they are viewing. Figure B.6 illustrates what might represent buildings connected to a main road via secondary paths. In a pan-scalar map, if one were to zoom out, the disappearance of these secondary paths could likely be observed at a certain zoom level, followed by the aggregation of the buildings into a single polygon. These generalization steps provide experienced users with clues about the scale. This phenomenon challenges the previously emphasized importance

⁴Or at the center of the pinch when using a smartphone, see Section B.3.

of smooth transitions in the generalization process.



⌘ Figure B.5 : Intersections and bifurcations are what matters in order to indicate changes in scales, not the number of lines.



⌘ Figure B.6 : Generalization steps may change the representation of visual cues, but change in representation may be used by map users to infer the current scale.

B.2.2.3 Conclusion

To conclude this brief experiment, we can assert that a single point or rectilinear object does not provide enough information to ensure good orientation. Other shapes, such as polygons and non-linear lines, theoretically suffice, but in practice, it is challenging for the map user to fully grasp changes in spacing or the scaling of certain shapes⁵. However, with two visible cues, regardless of their form, users obtain sufficient information to theoretically gauge the extent of pans and zooms. Thus, Google’s decision to highlight multiple prominent buildings in capital cities seems to be a good choice (Figure A.4). Not only do these buildings have the potential to create strong anchors individually, but the combination of them also helps users better estimate distances and scales.

B.2.2.4 Limitations and further considerations

The main limitation of this thought experiment lies in its static nature, as it considers cartographic elements within a specific view, while navigation in pan-scalar maps involves constant movement. Users are not confined to a single static view—they zoom in and out, pan across the map, and explore different areas dynamically. As a result, it would be inadequate to focus solely on a fixed view to precisely determine the optimal density of visual cues in a pan-scalar map. While the view can offer measurable insights, it should be treated more as an intuition than as a strict metric. Moreover, we do not address the question about the maximal density of cues here.

An additional consideration is the distribution of visual cues across both time and space during navigation. As users explore a map dynamically, their perception of visual cues changes not only with the map view but also with the sequence of movements they make. For example, a cue that might be insignificant in a static view could gain importance as it transitions into or out of focus during a pan or zoom. Similarly, cues that appear in successive views could form a mental trajectory, reinforcing the user’s orientation through continuity. Understanding how these transitional dynamics influence users’ spatial cognition could refine the design of landmarks and other visual aids in pan-scalar maps. Regarding the management of visual cues in maps, one approach to controlling their density across zoom levels could be to distribute them spatially and across scales using a quad-tree structure, similar to how WMTS tiles are managed.

Another avenue for extending this experiment is to explore how the semantics of visual cues interact with their geometry. In a real-world map, the elements are not just abstract shapes but also represent objects with specific meanings, such as buildings, roads, or rivers. Users’ familiarity with these objects and their semantic significance likely play a critical role in how effectively they are used for orientation. For instance, a hospital symbol may serve as a stronger cue than a generic building symbol, even if their geometric properties are identical. Future studies could investigate how semantic and geometric properties combine to optimize orientation in pan-scalar maps.

⁵For simplicity, we focused here on abstract shapes, but in a real map, the elements represent real-world features such as buildings or transport infrastructure. The user’s knowledge of the actual size of these features also provides valuable cues about the scale they are currently viewing.

B.3 Pan-scalar map interfaces evolution: a two-decade perspective

It is important to acknowledge that this research lies within the realm of human computer interaction and interface design (Zeng 2009). Therefore in this section, we digress from landmarks to delve into the interactions between users and map browsers by revisiting the systematic approach proposed by Harrower and Sheesley 2005. Additionally, we examine how the evaluated browsing methods have fared in 2024, given technological advancements and changing user expectations. Furthermore, we propose a novel model for user interactions with maps during zooming or panning, which offers a refined approach for evaluating efficiency criteria. While our work does not aim to comprehensively review the literature on Human-Computer Interaction (HCI) or multi-scale maps, it serves to highlight key insights, providing a solid foundation to conduct experiments on interactive map design.

B.3.1 Context

New devices, particularly smartphones, are akin to artificial organs that enhance the human body. These devices enable 'superpowers' such as instantaneous communication with individuals worldwide or access to any shared piece of information at any time. However, these new devices do not benefit from the millions of years of evolutionary refinement characteristic of natural organs. The designers' work is to create and refine those devices until they become indistinguishable from any of our natural organs.

Efficient and user-friendly map browsers are essential for facilitating effective navigation and spatial exploration, but contrary to the common assumption, the most commonly used solutions are not inherently the most effective. Effectiveness often hinges on how well the interface aligns with users' habits, preferences, and cultural expectations; therefore assessing the efficacy of such browsers remains challenging.

(Harrower and Sheesley 2005) proposed a systematic approach to compare various zooming and panning methods in map browsers. This approach, outlined in their paper "Designing Better Map Interfaces: A Framework for Panning and Zooming", introduced four functionality criteria (sequentiality, precision, presence of orientation cues, feedback) and two efficiency criteria (interface workload and information to interface screen estate ratio). By applying these criteria, Harrower and Sheesley evaluated nine map browsing methods to determine their effectiveness. Nearly two decades later, we propose to revisit their paper by delving into the evolution of map design, assessing the remaining relevance of the criteria proposed in 2005, and assessing performance of contemporary browsing methods compared to their predecessors.

B.3.2 Map design evolution

B.3.2.1 From paper maps to pan-scalar maps

Map design has undergone significant evolution from the pre-digital era to the present day. In the past, cartographers meticulously crafted paper maps with a focus on printing quality and user legibility, prioritising certain details while generalising others. Map design was restricted to the creation of a restricted set of topographic and thematic paper maps. The cost of production required them to have the widest utility.

In contrast, contemporary map design encompasses just-in-time mapping and intuitive interfaces, often on small devices. As a result, issues of multimodal interaction modelling (M2M, H2M interfaces) have become integral to the design of the map. Nowadays, most maps are pan-scalar maps, i.e. interactive, multi-scale, zoomable maps designed with the assumption that the user will navigate seamlessly and intuitively across many levels of detail in order to generate contextual understanding with the least cognitive effort (Gruget, Touya, and Muehlenhaus 2023). They are tailored to accommodate a wide range of tasks and user environments, but this paradigm shift in interaction and map design presents new challenges. First, maps now integrate significantly more data than before. To prevent overwhelming users, efforts on map generalisation aim to suppress and selectively enhance content in order to minimise navigational trauma. Secondly, additional tasks (itinerary creation, landmark selection..) are now mediated through the map. The challenge is in keeping interfaces simple while expanding interaction possibilities (Gedicke and Haunert 2023).

B.3.2.2 Task-based approach for evaluation

The evaluation of a map design and its map browser is a multi-variate challenge. The importance of task and context of use is reflected in the continued use of task-based frameworks, offering a structured approach to categorising tasks and defining functionality and modes of interactions (Norman 1991; Roth 2013).

Harrower and Sheesley, in particular, proposed a distinction between **precise** map navigation tasks and **fuzzy** map navigation tasks. In precise navigation tasks, users possess sufficient information about the desired location to utilise the search bar to input an address or coordinates. Conversely, in fuzzy navigation tasks, users lack locational knowledge and must directly interact with the map. In the context of this thesis, our research is focused on effective and efficient interface and map design in support of fuzzy navigation tasks. Nonetheless, it is noteworthy that this task classification may evolve in the future, as search bars make it possible to carry out increasingly complex tasks (Khellaf, Schlicht, Bayer, Bouwmeester, Miral, and Wagner 2023).

B.3.3 Zooming and panning methods

B.3.3.1 Evaluation criteria

In order to evaluate the different zooming and panning interactions, Harrower and Sheesley proposed four criteria based on functionality and two criteria concerning efficiency.

❖ **Functionality Criteria 1: Sequential versus non-sequential map browsing**

Does the method allow “jumping” to a new location or scale with a single input, irrespective of how far the destination is from the starting point? Today’s browsers only support this task by facilitating zooming out, panning at the small scale and zooming in again.

❖ **Functionality Criteria 2: Precision**

Does the method allow panning or zooming with precision? Precision navigation need minimising the latency and requires algorithms capable of high levels of abstraction/data handling.

❖ **Functionality Criteria 3: Local-global orientation cues**

Does the method give local-global orientation cues to users? A local-global orientation cue refers to a displayed element or interaction mechanism that aids users in understanding the spatial context of their current view within the overall map. This criterion is more readily applicable to the entire map browser rather than specific zooming or panning methods. As interactive maps became virtually limitless, orientation cues can no longer relate to the border of the map, but users can still orient themselves relative to map features they recall from current or previous explorations. This problem of maintaining the spatial awareness of users to prevent “navigational trauma” has been considered an important challenge for a long time (Rodgers 2001).

❖ **Functionality Criteria 4: Live-linked manipulation**

Does the method provide instant feedback? Today’s users are intolerant of any latency - which can add significantly to the risk of navigational trauma. Such a criteria places huge demands on rendering algorithms.

❖ **Efficiency criteria 1: Interface Workload**

How exhausting is it to use this mode of interaction? The measurement and modelling of cognitive workload is complex (Kosch, Karolus, Zagermann, Reiterer, Schmidt, and Woźniak 2023). A count of the ‘interactions’ in the execution of a task is a simple measure of workload. It might be the number of mouse clicks, gaze trajectory or object fixation time, but accounting for user strategies, prior experience, and task definition all compound the complexity of measuring the cognitive effort of interaction. This remains a critical area of research.⁶

⁶The website <https://userinyerface.com/> offers a unique demonstration of interface workload.

❖ **Efficiency criteria 2: Information-to-interface ratio**

Do we minimise screen area devoted to interface functionality, whilst maximising space devoted to geographic content? Such a criteria remains ever more important given the screen size of handheld devices. Multimodal/dialogue/gesture-based interfaces are significantly improving this ratio, whilst offering the chance for more intuitive and immersive environments (Hurst, Degbelo, and Kray 2024).

B.3.3.2 Zooming and panning methods review

We have supplemented the nine zooming and panning methods outlined by Harrower and Sheesley (2005) with four new methods and features. In the following discussion, we will examine these additions and assess their implementation in six prominent map browsers: Google Maps, OpenStreetMap, Plan IGN v2 (from the french mapping agency), Bing Maps, Mapbox and QGIS. The results of our analysis are presented in Table B.1.

Functionalities	Map browsers					
	GMAP	OSM	PIGNv2	BING	MAPBOX	QGIS
1. Directly Re-position the Map (Grab and Drag)	X	X	X	X	X	X
2. Smart Scroll Bars						
3. Rate-Based Scrolling						
4. Keyboard Controls	X	X	X	X	X	X
5. Zoom and Re-centre Under Mouse Click	X	X	X	X	X	X
6. Navigator Tabs / Interactive Compass				X ^a		
7. Navigator Window Interaction (Detail + context)			X ^b			
8. Specify Explicit Coordinates or Scale	X	X	X	X	X	X
9. Zoom Box (simultaneous pan and zoom)						X
10. Scroll wheel to zoom, centred on the mouse	X	X	X	X	X	X
11. Pinch to zoom, centred on the fingers (touchscreen)	X	X	X	X	X	NR
12. Continuous transition when zooming or panning	X	X	X	X	X	
13. Inertial drift after releasing the grab	X	X	X	X	X	

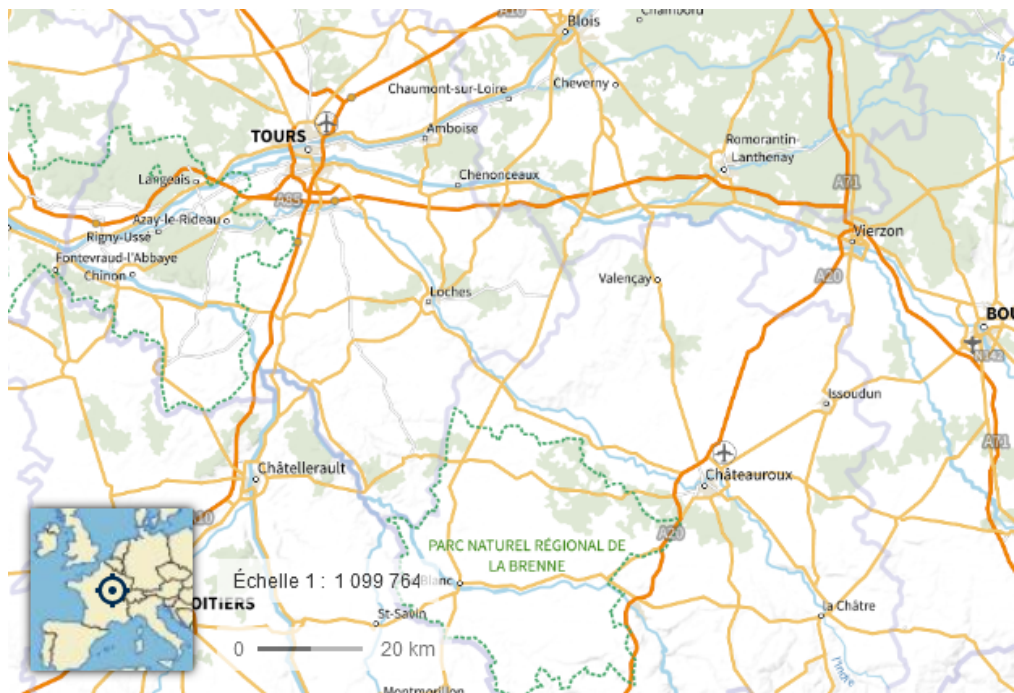
^a only for rotation and inclination, ^b allows to click but not to draw a

✱ Table B.1 : Panning and zooming methods and functionalities currently implemented in map browsers.

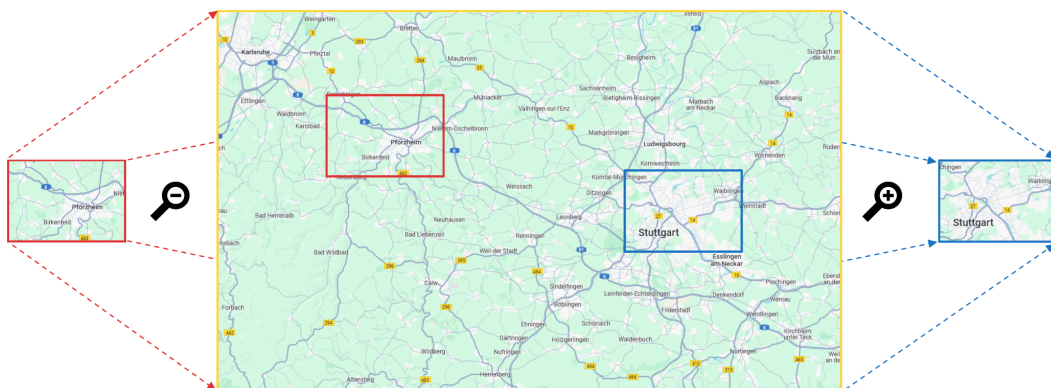
❖ **Navigation methods from 2005**

1. *Directly Re-position the Map (Grab and Drag)*: Grab and drag is now standard in all map browsers.
2. *Smart Scroll Bars*: Scroll bars that dynamically adjust in size based on the length of the document have become a standard feature in web browsers and document viewers. However, they do not readily translate to pan-scalar maps which are conceptually infinite in extent and detail.
3. *Rate-Based Scrolling*: This method has been abandoned for maps but remains classical on web browsers and on document browsers. Nowadays, we need to activate it with the middle mouse button, which was not standardised in 2005.
4. *Keyboard Controls*: All web map browsers facilitate keyboard arrows to pan and the “+” and “-” keys to zoom. Harrower and Sheesley observe that this method doesn’t allow the user to precisely choose the map extent or the level of zoom, but keystrokes remain a very popular mode of interaction for gamers. One approach to enhancing precision with keyboard controls could involve adjusting the velocity of the screen movement rather than moving the screen by a predetermined amount of space.
5. *Zoom and Re-center Under Mouse Click*: A double-click (or double-tap on a touchscreen) zooms the map within the map browser. Considered less precise compared to using the mouse wheel, it has been retained due to its convenience for mobile users, allowing them to zoom with just one thumb while holding the phone.
6. *Navigator Tabs / Interactive Compass*: This method involves using on-screen buttons to zoom or pan by a predetermined amount. Harrower and Sheesley evaluated it as being inefficient. In today’s browsers zoom in and out buttons remain, though rarely utilised (a task increasingly performed by the mouse wheel).
7. *Navigator Window Interaction*: Although categorised as a navigation method by Harrower and Sheesley, it corresponds to a specific map browser interface commonly known as “overview+detail” (Cockburn, Karlson, and Bederson 2009). Here the interface allows users to click on the inset minimap to recenter. The “overview+detail” window has been a standard feature in video games since the 1990s and recent studies analyse them to deduce good practices for cartography (Zagata and Medyńska-Gulij 2024). Harrower and Sheesley argued that enabling users to draw a zoom box over the minimap would be the most efficient method for map navigation. IGN France’s interface is one example of a browser that offers an “overview+detail” window, though it does not allow a zoom box to be drawn (Figure B.7).
8. *Specify Explicit Coordinates or Scale*: All map browsers have a search bar that enables users to enter specific addresses or coordinates. The search bar has since evolved into a powerful tool, offering an additional range of versatile features and functionalities.

9. *Zoom Box (simultaneous pan and zoom)*: This method has not been retained as a primary feature. It is only found in specific 'zoom' options within GIS software and by pressing 'ctrl' in maps built with OpenLayers. Its only advantage over using the scroll wheel arise when users aim to reach a specific extent, a functionality generally unnecessary for non-professional use. Moreover, un-zooming with this method can feel disorienting.



☼ Figure B.7 : The "overview+detail" interface of the web map browser Plan IGN v2 is situated in the bottom left corner of the screen (<https://www.geoportail.gouv.fr/donnees/plan-ign-v2>).



☼ Figure B.8 : By pointing the cursor to the top left of the screen and scrolling to zoom out, followed by pointing the cursor to the right part of the screen and scrolling to zoom in, the user transitions from Pforzheim to Stuttgart.

❖ **New navigation methods**

10. *Scroll wheel to zoom in or out*: Though not explored by Harower and Sheesley, this method has emerged as the primary means by which we interact with maps. In current implementations, zooming is centred on the mouse rather than on the centre of the screen, despite potentially being less efficient (You, Chen, Liu, and Lin 2007). Nevertheless, this approach enables users to zoom and pan with a single interaction, requiring minimal input once mastered (Figure B.8). This method offers symmetric zooming in or out, providing an advantage over the zoom box method.
11. *Pinch to zoom in or out*: This method is now a standard feature in all touch screen interfaces. Pinching feels very similar to using the scroll wheel, which allows the user to quickly adapt from one method to the other.
12. *Rotation*: Although this interaction is often criticized by cartographers because it breaks one of the most classic conventions of orientation (north at the top of the map), it is offered by all touch-based interfaces, especially on smartphones. Typically, this action is performed with two fingers, one moving up and the other moving down. This functionality invariably adds to the risk of disorientation.

❖ **New map browser functionalities**

12. *Continuous transition when zooming or panning*: Smooth transitions are a significant improvement to the zooming and panning methods, as they allow the users to comprehend more local-global clues (criteria 3).
13. *Inertial drift after releasing the grab*: The drift allows users to go in a given direction with less dragging at the cost of precision. Professional tasks often need more precise control which is perhaps why it has not been implemented in GIS software.

Table B.1 suggests that various map browsers have converged towards a shared set of interaction modes, leveraging users' prior experiences—a form of lateral application of users' prior knowledge—that renders new experiences intuitive. When new interaction modalities are introduced, such as in the navigation of pan-scalar maps, there arises a collective need to augment users' computer literacy to include these novel interaction modes.

B.3.4 Conclusion

Over the past two decades, map browsing has seen a paradigm shift, marked by an expansion in the scope and purpose of map usage, with smartphones now dominating usage. Despite these changes, the six criteria (Section B.3.3.1) established by Harower and Sheesley in 2005 continue to hold relevance in contemporary map design practices, with one notable exception being the criterion for instant feedback, which has evolved from a desirable feature to an essential requirement. Concerning zooming and panning interactions, a standardisation has been observed across map browsers, with the 'Grab and drag' interaction emerging as the prevailing method

for panning, both on desktop and mobile (tactile) devices, while mouse scrolling and pinch gestures have become standard for zooming. This standardisation is an opportunity for the Lost In Zoom project as we ultimately seek to improve the zoom interaction. It means we can use the centred zoom as a baseline for our experiments. Additionally, we noticed that AI advancements have bolstered the functionality of search bars in map browsers, enabling certain tasks that previously required manual map exploration before to be delegated to the computer.

B.3.5 Bonus: Modelling navigation tasks in pan-scalar maps

The majority of navigation tasks performed with a map involve multiple zoom and pan interactions. To delve deeper into the analysis of user behaviours in real conditions, we propose a new model that describes the steps of a user who interacts with a map browser.

B.3.5.1 Static phases and transition phases

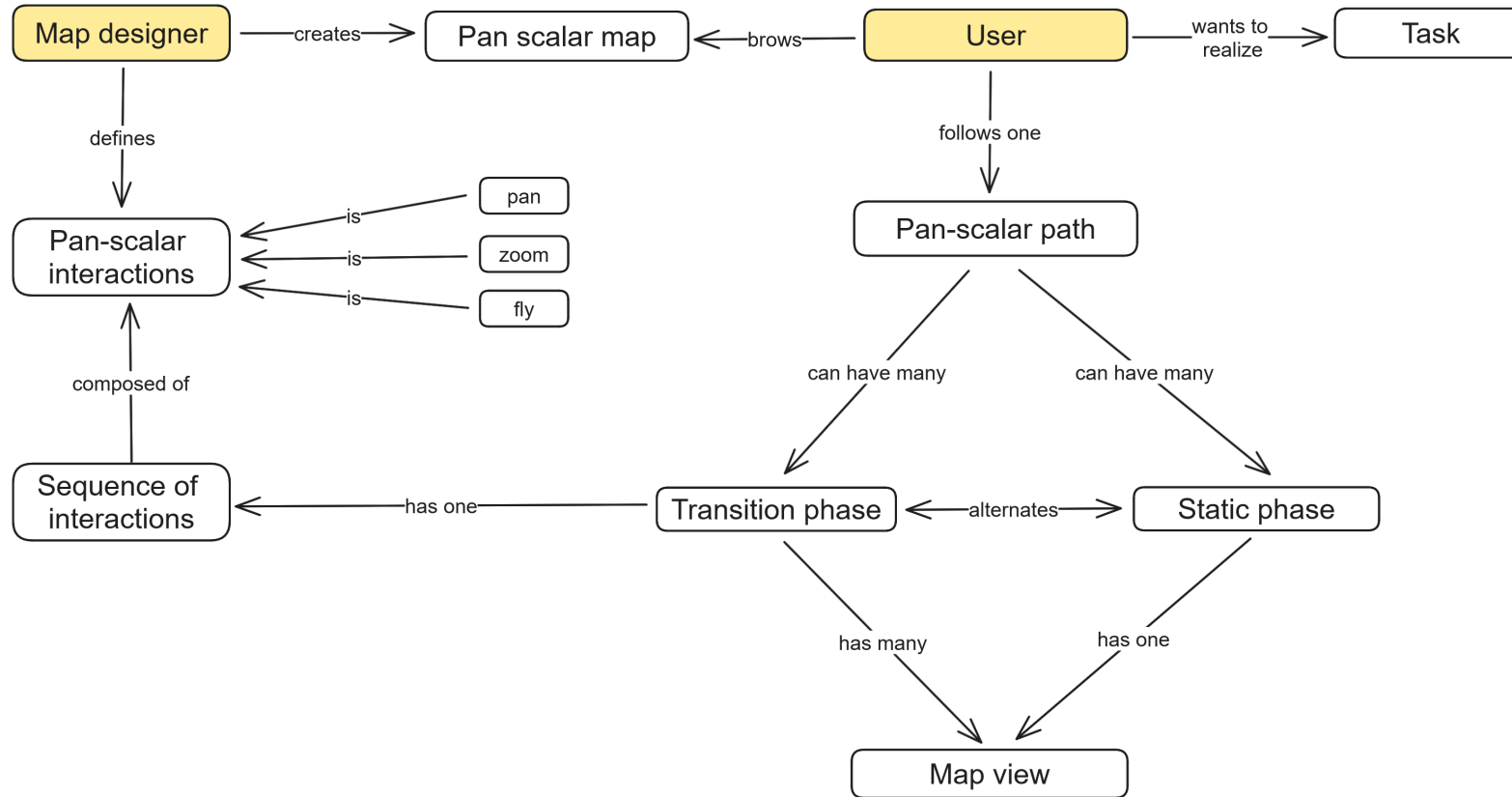
By observing user experiments in map interaction, we have identified two distinct alternating phases that occur during a map exploration: **static phases**, where the user does not interact with the map browser for a significant period, usually because he or she is observing the map, and **transition phases**, where the user engages in multiple interactions with the map browser. While the map view (i.e., the displayed map extent) remains unchanged during static phases, it undergoes rapid changes during transition phases. We think that modelling map exploration as a set of multiple phases could help identify patterns more closely linked to the process of comprehension and task execution.

B.3.5.2 The model

Inspired by the definition of a pan-scalar exploration proposed in Touya, Gruget, and Muehlenhaus 2023, our model (Figure B.9) centres around the concept of the "Pan-scalar path", which represents the trajectory a user follows while completing a task with a map browser. It is composed of successive static phases and transition phases. Each transition phase includes a sequence of pan-scalar interactions.

B.3.5.3 Building experiments

Pan-scalar interactions, transition phases, and static phases can be identified solely by analysing the evolution of the map extents displayed in the map browser. This can be accomplished by tracking map events during user experiments with a custom OpenLayer script. To differentiate between static and dynamic phases, we propose a minimum duration of one second for static phases. Further analysis includes counting of additional mouse inputs and positions, as well as tracking participants' gaze using eye-tracking devices (Wenclik and Touya 2023).



⊛ Figure B.9 : Model of users interactions browsing a pan-scalar map. Source: Potié, Touya, Wenlik, and Mackaness 2024.

B.3.5.4 Why it matters

We think that this model could offer valuable insights into mapping cognition and user behaviours in the consumption of map content. By categorising and describing user behaviours, researchers can better understand the cognitive processes underlying different types of map interactions and tasks. By leveraging unsupervised machine learning, the model enables the classification of navigation archetypes, enhancing experimental design and optimising map browsing interfaces by tailoring interactions for each phase. If the model proves robust, it could be taken into account by designers to develop interactions specific to each phase. For example, we could imagine displaying specific information during transition phases, thus improving users' understanding of map features and overall browsing efficacy.

B.4 Conclusion of the chapter

In this chapter, we conducted an inventory of the types of landmarks present in current pan-scalar maps and observed that they are relatively consistent across different maps. While this study does not identify which landmarks have the greatest potential to form cognitive anchors, it provides a solid foundation for further exploration. We then examined the ideal number of landmarks and visual cues to include in pan-scalar maps. The literature on landmarks highlights that landmarks compete for cognitive attention during navigation and shows that the number of landmarks remembered in a route-following scenario is limited. However, additional research is needed to transition from determining the ideal number of landmarks in route-following scenarios to identifying the optimal number of landmarks in pan-scalar maps. Still, we proposed a thought experiment to explore the minimum number of visual cues required in a map view and concluded that a single element is not always sufficient for users to gauge the extent of their zoom or pan movements.

Finally, we reviewed the range of interactions available in current pan-scalar maps and traced their evolution since 2005. We observed a convergence toward interactions that can be performed similarly across devices, whether from a computer or a smartphone, such as "grab and drag" and zooming centred on the mouse or via pinching gestures. Moreover, we noticed a pattern related to user interactions with the map which alternated between static phases and transition phases, and we proposed a model to describe it, hoping it would help designing future map interactions and interfaces. These findings provide insights into the potential use of carefully designed landmarks to reduce user disorientation during map navigation. In the next chapter, we will focus on a specific type of landmark observed in our first study: ring roads.

Part II

Focusing on the Detection of Specific Landmarks

Chapter C

When is a ring road a ring road? A perceptual study

In this chapter, we present a user study from (Potié, Mackaness, and Touya 2022) aimed at gaining a deeper understanding of users' perceptions of ring roads, with the objective of identifying potential approaches for their detection. We take advantage of the extended format of this thesis to elaborate on certain ideas and integrate the findings into the broader narrative of our research.

C.1 Introduction

C.1.1 Ring roads potential

C.1.1.1 Ring roads may become cognitive anchors

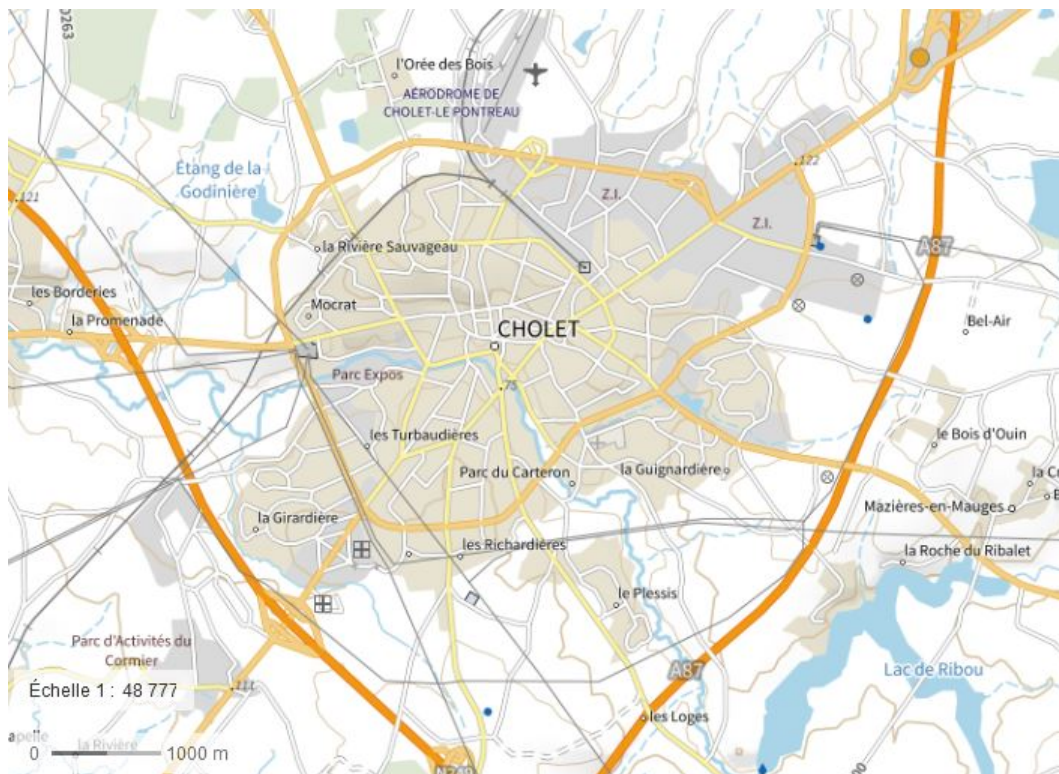
Among the types of pan-scalar landmarks identified in our study in the previous chapter, urban ring roads stand out as some of the most promising candidates for creating cognitive anchors in users' mental maps. Ring roads have a strong visual identity due to their circular shape, play a structural role within cities by facilitating accessibility across neighbourhoods (Nes 2021), and are present at several key scales frequently used by map users. Notably, the largest ring roads become visible as early as zoom level 7 (a scale at which one can see the entirety of France) and can serve as reference points at increasingly detailed scales, even when they are no longer fully visible at the highest zoom levels.

Moreover, ring roads are already included in most maps, suggesting that further highlighting them would not necessarily increase visual clutter—and might even clarify the layout of the map. Finally, their distinctive circular structure allows them to combine effectively with other landmarks to help map users gauge scale as they zoom (see Chapter 2, Section B.2.2 about minimum visual cues).

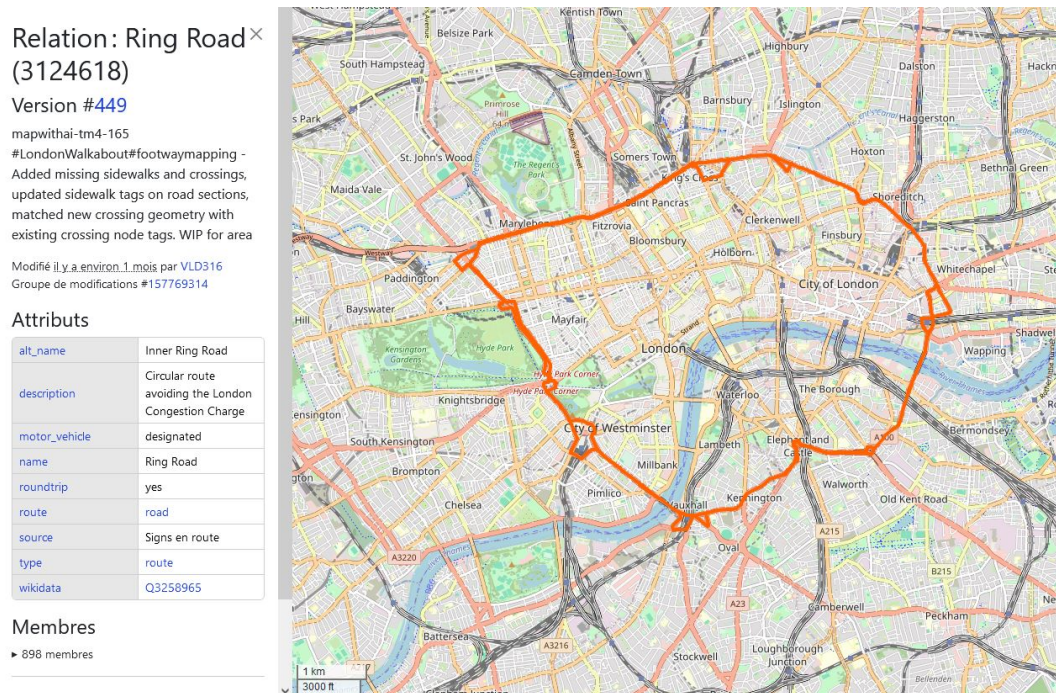
C.1.1.2 The Symbolisation of Ring Roads – in Need of Special Treatment?

Currently in the IGN map and OSM, urban ring roads follow the same representation rules and classification processes as other roads. However, we may observe that roads encircling cities are often naturally “up-ranked” by classification algorithms, which results in a more prominent representation (Figure C.1). This prioritization is not based on their circular shape or classification as “ring roads” in the database but rather on other intrinsic characteristics used to determine road importance, such as the number of lanes, speed limits, or road width. Google Maps also appears to use a similar approach, as there is no noticeable difference in the depiction of ring roads between it and the IGN map.

The fact that most ring roads already stand out naturally on the map on a city-wide scale, due to their typically higher road class compared to the rest of the network and their circular shape, may explain the lack of interest from cartographers in treating them as a distinct category. However, the absence of a "ring road" classification in the database (Figure C.2) means that if we aim to highlight ring roads by adjusting their symbology during the zooming process or to use them during generalization processes, it would not be possible without an initial detection step.



☼ Figure C.1 : The city of Cholet as it appears on the IGN map at zoom level 13. A ring road is visible in light orange, the same color used for major departmental roads extending outward from the city.



✿ Figure C.2 : London ring road in OSM. OSM does not have a specific "ring road" attribute for roads. The only (few) ring roads available in the database are manually created "relations" that include "ring road" (or "circular road") in their name and description.

C.1.2 Ring road preliminary definition

It may also be the challenge of providing a clear definition of what constitutes a ring road that has prevented their inclusion in databases thus far. If there is no universal definition, how can we develop algorithms that could reliably enrich the database with ring road information?

Ring roads are not just circular roads. Some circular roads are perceived as ring roads and some others are not, and people do not always agree on what to consider as a ring road or not. They are complex objects because of their many characteristics: their shape, their size, their type of roads, their location etc. We started this study with this definition of ring roads¹:

A “ring road” is a road or a collection of roads that has a circular aspect and that is known for its property to go around a city or a village.

This definition is of course fuzzy, but one goal of this study is to improve this definition, either through better contextualisation or by making it more precise.

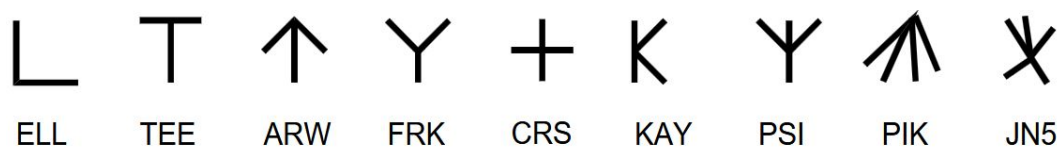
¹While the term "ring road" is commonly used in the UK, the term "beltway" is often preferred in the US. In French, the equivalents for "ring road" are "route périphérique urbaine", "périph", or "tour de ville".

C.1.3 Literature on ring road detection

C.1.3.1 Pattern detection in road networks

The detection of road structures and importance through automated methods has been a well-explored field, with significant applications in urban planning, architecture, database enrichment, and map generalization. In map generalization, such methods assist in detecting and preserving key characteristics of a place across varying scales by extracting implicit information from road networks.

While most pattern detection techniques focus on images, detecting patterns in road networks is particularly well-suited to the use of graph structures, where roads are represented as edges and intersections as nodes. A typology of graph nodes was proposed by Sester 1995 and may serve as a foundational building block for identifying more complex patterns in graphs (Figure C.3). Identifying nodes with high degrees (hubs) and enriching edges with weights derived from their connections to these hubs allow us to detect key structures like major road axes. This is essential to ensure their preservation through map generalisation processes.



☛ Figure C.3 : Node types proposed in Sester 1995 (in German), and discussed in Heinzle and Anders 2007.

An insightful state-of-the-art review by Heinzle and Anders 2007 discusses various pattern detection in road network "spaghetti" dataset (collections of geometric elements like points, lines, and polygons). Their article describe strokes, stars, grids and circular roads as basic patterns and explain methodologies to identify them in a graph. Touya 2010 proposes a methodology that make use of such pattern detection to select rivers and roads during the map generalisation process.

As explained in Heinzle's article, Thomson and Richardson 1999 introduced a method to group road segments with strong continuity into "strokes," providing a foundation for further spatial analysis (see Section A.3.2.1). This approach addresses the issue of road segmentation when converting a road network into a graph, where each intersection splits a road into multiple edges, even if it remains a single road in reality.

Grids are also common patterns in road network that can be described by two sets of parallel lines intersecting with an angle that is more or less 90 (Figure C.4). He, Ai, Yu, and Zhang 2017 demonstrate that grid detection is feasible through a combination of spatial analysis and machine learning technics.

Zhang 2004 further advanced road enrichment by modelling patterns as objects and proposing an indicator for the network density in a road graph, while Savino, Rumor, Zanon, and Lissandron 2010 and Scheider and Possin 2012 focuses on detecting specific road junctions and highway carriageway. Concerning road importance, Morisset and Ruas 1997 proposed a method based on agents to simulate the road



✿ Figure C.4 : Grid patterns detected in the city of Zurich. Source: Heinzle and Anders 2007.

use in a road network to determine the roads of high frequency usage.

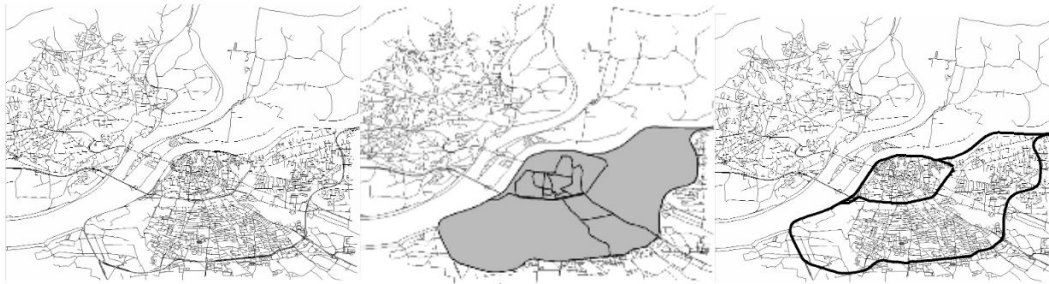
Other advancements in road enrichment include modeling patterns as objects and introducing network density indicators (Zhang 2004), as well as detecting specific road junctions and highway interchanges, either from graphs (Savino, Rumor, Zanon, and Lissandron 2010; Scheider and Possin 2012) or from map images (Touya and Lokhat 2020). Regarding road importance, Morisset and Ruas 1997 proposed a method using agent-based simulations to identify roads with high usage frequency, and spatial analysis using space syntax (Hillier and Hanson 1984) has proved to be a very useful tool for predicting road use (Czerkauer-Yamu and Voigt 2011).

C.1.3.2 Circular road detection

The detection of ring roads is part of the pattern detection in road networks field. A notable study by Heinzle, Anders, and Sester 2006 demonstrated an effective computational approach for detecting circular roads, focusing on the geometry of roads (Figure C.5). Their model identifies strokes in the road network and aggregates the resulting road polygons (les polygones formes a partir du reseau de routes) while maintaining the integrity of major road strokes, thereby simplifying the network sufficiently to analyse and select cycles exhibiting ring road properties. However, this model was based on specific assumptions about what constitutes a ring road, with subjective criteria for aggregating road polygons and selecting ring roads. Unlike the detection of simple geometric shapes, identifying ring roads is more nuanced, akin to the challenge of detecting urban centers, which lack clear boundaries and depend

on external factors such as the proximity of commercial areas or public perception.

To refine the approach of Heinzle, Anders, and Sester 2006, we aim to identify additional characteristics that more accurately define ring roads from a human perspective. To this end, we conducted a perceptual study to explore the fundamental question: "When does a ring road qualify as a 'ring road'?". Our study seeks to bridge the gap between computational detection and human perception, enhancing our understanding and identification of ring roads.



✿ Figure C.5 : The French city of Avignon and the two circular roads roads detected based on road blocks aggregation. Source: Heinzle, Anders, and Sester 2006.

C.2 Perceptual study

C.2.1 Objectives of the study and hypothesis

Our experiment aims to explore the perceptual characteristics of urban ring roads. By understanding how these roads are perceived, and what attributes contribute to their meaningful characterisation, we aim to improve their detection, with the long term objective to use them to create navigational landmarks in pan-scalar maps.

- Assumption: if people recognise a ring road and name it as such, they may anchor it in their mental map

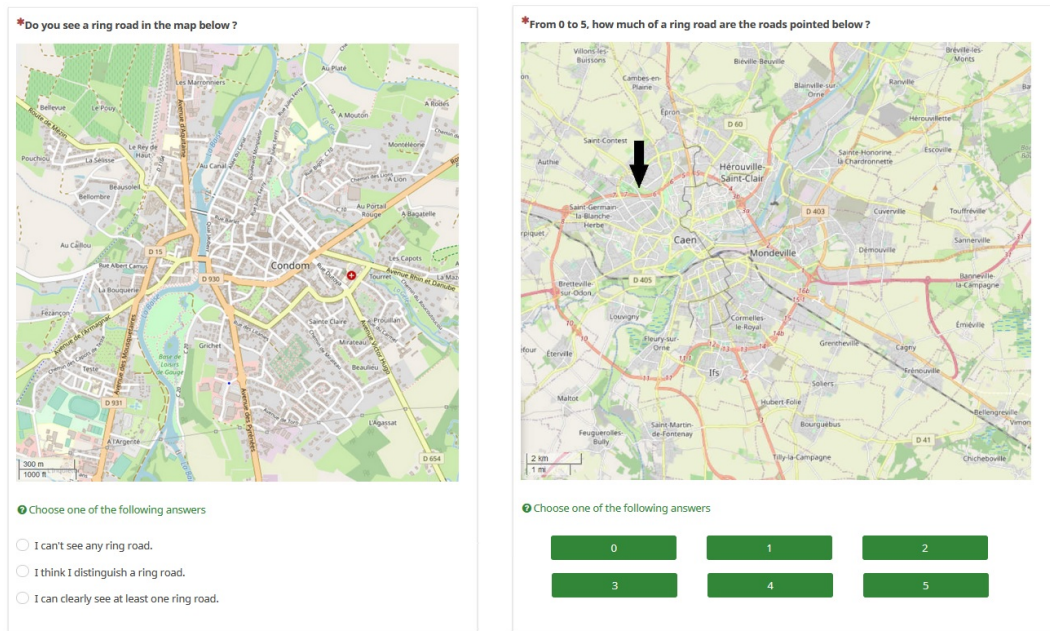
Hypothesis to be verified: while the definition of ring roads is fuzzy, there is some kind of consistency between users concerning what is a ring road and what is not, potentially enough to build a detection algorithm.

C.2.2 Methodology of the experiment

The experiment consists of a user survey on the Limesurvey platform² in which we show (non interactive) map images to users and ask them questions about their perception of ring roads contain within them (Figure C.6). The survey was divided

²<https://anchormap.limesurvey.net/632458?lang=en>

into 3 parts in which user were asked various questions. No definition of a “ring road” was provided. All maps and results are available online³.



✿ Figure C.6 : Two sample views of the survey as displayed to participants responding via a web browser.

C.2.2.1 Participants

28 participants went through the entire survey 14 of them responded online, and the 14 others went through the survey in the office under our supervision. Those ones were asked to explain their reasoning in a « think aloud » way, which is discussed in the « qualitative results » section.

The pool of participants was composed of researchers from our spatial data team and students from the university, as well as online participants, including fellow researchers and volunteers recruited from our networks via social media. One-third worked in the field of GIS, but we did not perceive obvious differences in how they responded to the survey compared to participants who were not familiar with GIS.

Considering genders, the group of 28 participants was composed of, 17 male, 9 female, 1 non-binary, and 1 preferred not to answer. We did not perceive obvious differences between males and females. Considering age, 10 aged 18-25, 14 aged 26-40, and 4 aged 41-60. We noticed a small difference in how older people analyse the map compared to younger people.

Also, two-third of the participants were French, few participants were native English speakers, and the rest were non-English native speakers from other countries. This may have an impact on how they understood the concept of the ring road. We

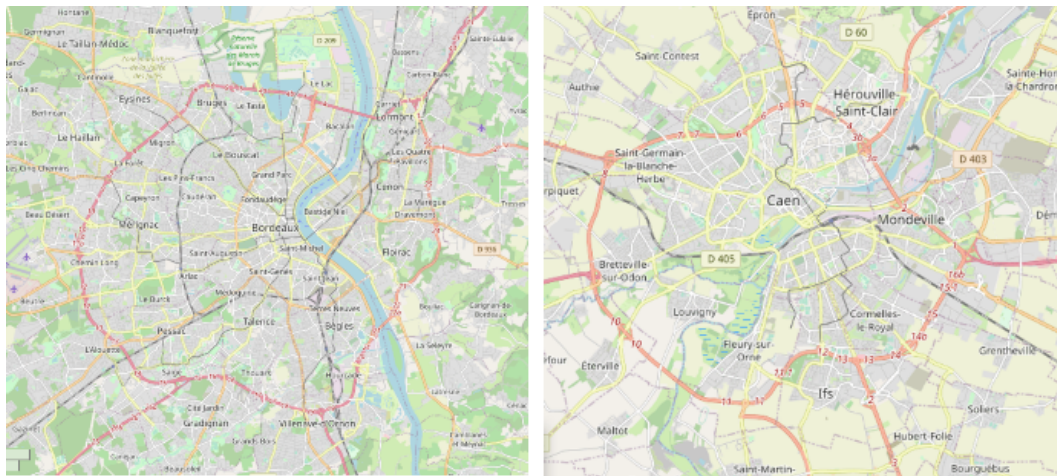
³<https://zenodo.org/records/6538873>

talk about these specificities in Section C.2.4.

Coutrot et al. 2018 highlights differences in individuals’ ability to navigate based on whether they have lived in urban or rural environments. Unfortunately, we did not include this question in our survey, but it could have been insightful to investigate whether such differences influence participants’ perception of what constitutes a ring road. For instance, it would be interesting to see if this background affects their perception of the minimum size required for a road to be considered a ring road. We may also have asked the town where they grew up to verify if it has a well defined ring road or not.

C.2.2.2 Maps

The map images used in the 3 parts of the survey were extracted from OSM and the locations were all in France. We chose to centre each map image on a city, trying to include its whole extent. We selected a pool of cities from zoom level 12 to 15, and tried to diversify the sample by selecting few cities with what we believed were obvious ring roads (Figure C.7), few cities with no ring roads, and many cities where it was hard to tell (Figure C.8). For each part of the survey, we made the maps appear in a random order to minimize bias.

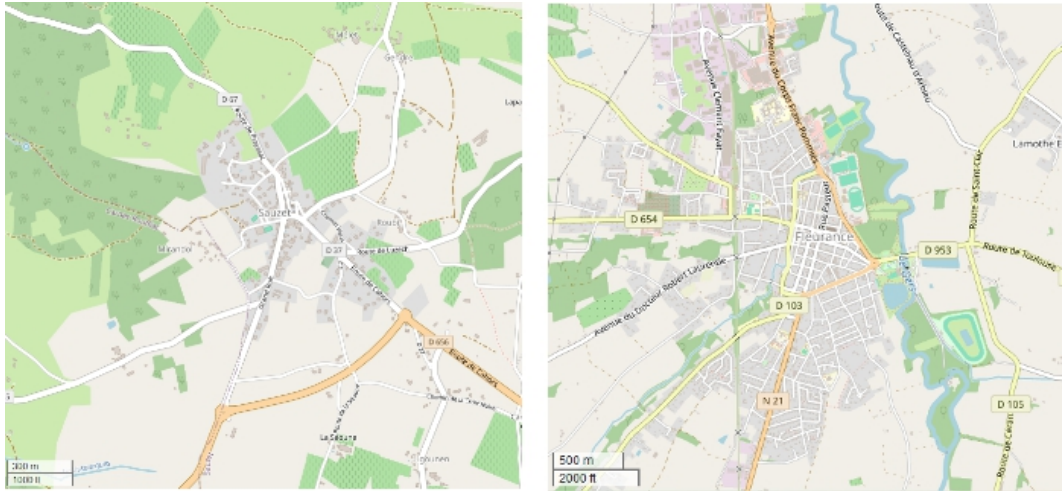


✿ Figure C.7 : Bordeaux and Caen and their distinctive ring roads. Source: OSM

C.2.2.3 Tasks and raw results

In **part 1**, we show 10 map images to the participants, and for each map we ask them: « Do you see a ring road on the map? ». The participants were able to decide between three choices:

- I can clearly see at least one ring road
- I think I distinguish a ring road
- I can’t see any ring road



☼ Figure C.8 : The villages of Sauzet (left) and Fleurance (right). We assumed there was no ring road in the first village and questioned whether the circular road in the second would be perceived as a ring road. Source: OSM

The participants were also allowed not to answer the question, but everyone answered. The map images used for this question and their associated results are available in Appendix A, page 215, and the results are summed up here in Table C.1.

N°	City name	No	Maybe	Yes
1	Bordeaux	0	0	28
2	Brest	9	12	7
3	Caen	0	2	26
4	Carcassonne	1	5	22
5	Chateaubriant	17	6	5
6	Condom	27	0	1
7	Le mans	1	13	14
8	Montbrison	18	4	6
9	St-Rémy-De-Pce	21	4	3
10	Tours	13	10	5

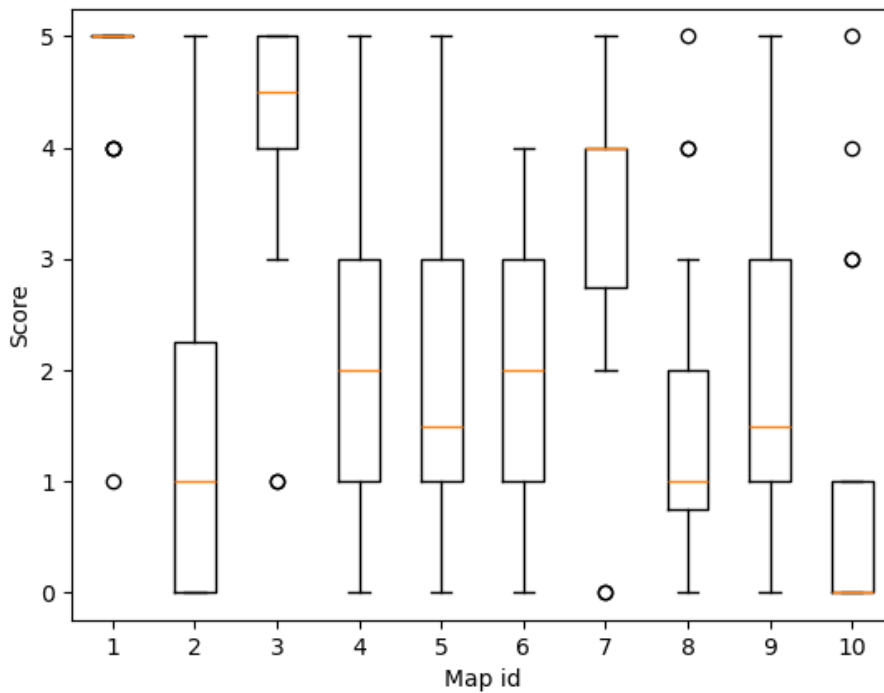
☼ Table C.1 : Results of the first part of the survey : "Do you see a ring road in the map below?"

In **part 2**, we show 10 map images to the participants, but this time in each image there is an arrow pointing to a circular road in the image (see Appendices B). The task was: « Using a score from 0 to 5, rate how much of a ring road the specified roads are. ». The map images used for this question and their associated results are available in Appendix B, page 221, and the results are summed up here in Table C.2, and a boxplot is available in Figure C.9.

In **part 3**, we propose a classification task where the participants were asked to rank 6 map images based on how salient the included ring road was, rank 1 corresponding to the map image with the most salient ring road and rank 6 corresponding

N°	City name	Av. rating /5	Std. deviation
1	Caen	4.75	0.80
2	Carcassonne	1.46	1.64
3	Clermont-Ferrand	4.21	1.10
4	Fleurance	2.29	1.49
5	Issoire (1)	1.71	1.38
6	Issoire (2)	2.04	1.23
7	Lille	3.29	1.27
8	Milau	1.57	1.37
9	Montcuq	1.86	1.60
10	Sauzet	0.68	1.36

⌘ Table C.2 : Results of the second part of the survey on ring roads : "From 0 to 5, how much of a ring road are the indicated roads.



⌘ Figure C.9 : Boxplot of responses from the second part of the survey. There is a strong consensus that Caen has a ring road and that Sauzet does not. Responses for other cities show more variability, indicating room for discussion.

to the map image with the less salient one. The map images used for this question and their associated results are available in Appendix C, page 227, and the results are summed up here in Table C.3. The qualitative analysis of the think-aloud process is particularly interesting in this task, more so than the average ranking of the images.

N°	City name	Average ranking from 1 to 6	R1	R2	R3	R4	R5	R6
1	Clermond-Ferrand	2.57	2	14	6	6	0	0
2	Marne-la-Vallée	2.21	16	3	3	2	1	3
3	Le Mans	2.32	9	9	7	0	1	2
4	Millau	5.00	0	1	3	3	9	12
5	Montbrison	4.71	0	0	3	7	13	5
6	Saint-Etienne	4.18	1	1	6	10	4	6

✿ Table C.3 : Results of the third part of the survey : "Rank the ring roads".

C.2.3 Analysis of the results

C.2.3.1 Quantitative results

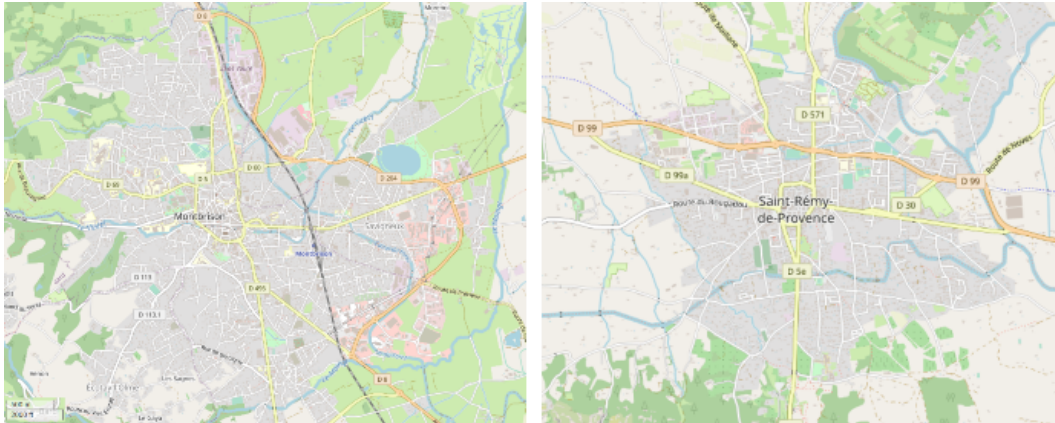
As expected, the ring roads of Bordeaux and Caen were easily recognised by participants in the first part of the survey (Figure C.7). These roads are complete circles, highly circular, and are depicted in a salient colour, making them stand out clearly. On the opposite end of the spectrum, no ring roads were identified in the map image representing the town of Condom, where some salient but non-circular roads can be seen.

For Montbrison (8) and Saint-Rémy-de-Provence (9), these towns lack a road encircling most of their urban area. However, circular roads surrounding their historical town centers are still visible (Figure C.10). Despite this, most participants did not classify these roads as ring roads. More broadly, we noted a trend: the smaller the town, the less likely it was that circular roads were identified as ring roads. Several hypotheses could explain this phenomenon:

- The relatively small size of the circular road compared to the image.
- The actual small size of the town (Do towns need to reach a certain size for surrounding roads to be considered ring roads?).
- The less salient colours used for smaller roads compared to the larger, more prominent ring roads of bigger cities.
- The fact that these roads often encompass only a small portion of the urban area.

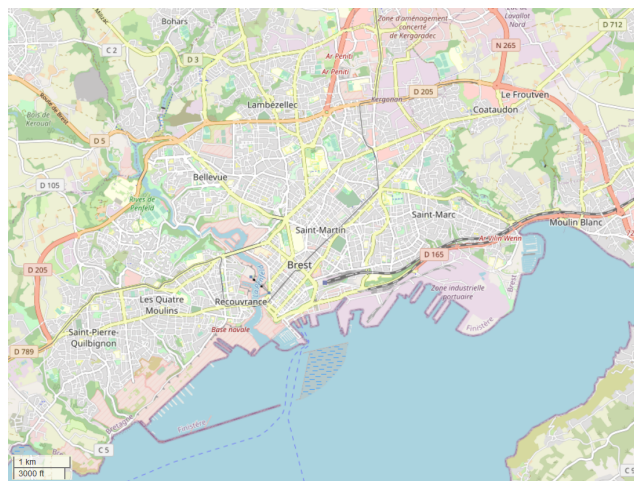
With cases like Châteaubriant (5) and Issoire (2), the incompleteness of ring roads significantly impacted user perception. In these instances, the proposed circular roads did not form a complete circle, leading to mixed participant responses. Issoire was shown twice in the survey, each time with different highlighted roads. The larger, outer, yet incomplete ring road received higher recognition scores than the smaller, more circular inner road. This suggests that circularity alone is not sufficient to identify a ring road, emphasizing the importance of considering partial open rings in our definition.

An intriguing result emerged with the city of Brest (2) (Figure C.11), which features a semi-circular ring road. Multiple participants argued that the coastline



☼ Figure C.10 : The small towns of Montbrison and Saint-Rémy-de-Provence. Both feature a relatively small circular road that goes around their old city center. Source: ©OpenStreetMap contributors

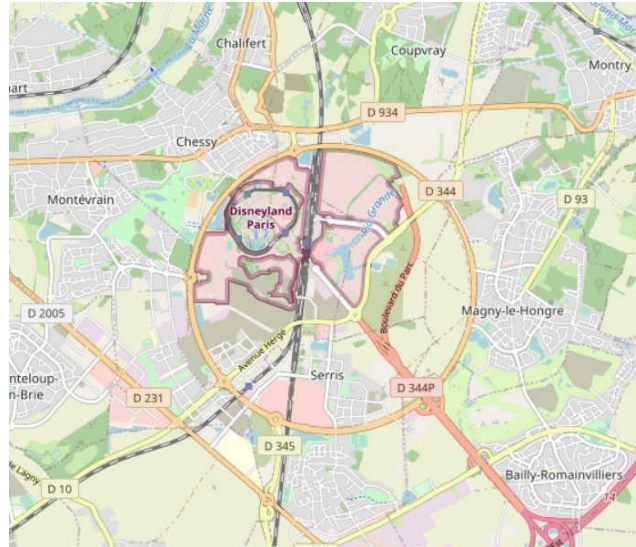
visually "closed" the ring. From their perspective, the coastline completed the loop. This raises questions: Do participants believe a full circle is unnecessary for the ring road to achieve its purpose if a natural boundary like the sea provides closure? Or does the combined visual circularity of the road and coastline influence this perception? While the exact answer is unclear, this example highlights how borders and non-road boundaries can affect ring road perception.



☼ Figure C.11 : The coastal city of Brest. Source: ©OpenStreetMap contributors

Another unique observation came from part 3, focusing on the specific circular road of Marne-La-Vallée (Figure C.12). This road forms a nearly perfect circle, so distinct from typical ring roads that it appeared too different to participants. Additionally, this road does not surround a city but rather the Disneyland resort. Opinions were divided on whether it should be considered a ring road.

Finally, regarding potential differences related to gender, origin, GIS expertise, or profession among participants, no significant influence on responses was detected



✿ Figure C.12 : The perfectly circular road that goes around the Disneyland resort, in Marnes-La-Vallée. Source: ©OpenStreetMap contributors

besides a distinction between younger and older participants that is mentioned in the following section. Moreover, we hypothesize that certain professions, such as truck drivers, might view maps from a unique perspective. Unfortunately, such professions were not represented in our participant pool.

C.2.3.2 Qualitative results

The think aloud method provided valuable insight, in particular:

- Participants tried to interpret the morphology of the city and the function of the roads in each case. A particularly important criteria was how the ring road might help circulation in the city, and how it was connected with major arterial road(s).
- Colour had a big impact on the perception of roads and induced a hierarchy of roads in the mind of the participants. The colour and density of the surrounding entities was also mentioned by candidates particularly in helping them to separate urban from rural since ring roads typically border both.
- Toponyms and annotations also informed the participants and seemed to have an important impact on their decision. They would rate higher a ring road that enclosed a substantial part of the city and some participants used the names to help demarcate the administrative limits of the city.
- Participants were more likely to agree on the presence of a ring road if they knew the city, especially if they had driven the road. This was true even when the roads did not appear to be circular on the map. Perhaps this is a process of reinforcement of the cognitive and structural aspects in the participant's mind despite a possibly poor visual rendering.
- In general, it felt as if younger individuals focused more on the road's visual characteristics, while older participants prioritized its structural role within

the city's road network and development. Given the small sample size, this observation warrants further investigation.

In conclusion, participants had a very similar sense of the defining properties, but they applied them in differing levels depending on their own definition of a ring road and their experience of the city. Participants bounced between the visual and structural properties in reaching their decision.

C.2.4 Limits of the survey

The order in which the images were presented to the participants affected their decisions. We observed that the participants tended to formalize their definition of a ring road early in the study and often kept to this definition until the end. We tried to minimize this bias by making the maps appear in a random order, but we believe it still had a non-negligible impact on the result of the survey.

We acknowledge that the geographical extent of the ring road compared to the total extent of the images impacts on how people perceived the ring roads. Our sample is too small to show a clear correlation between extent and the recognition of ring roads, but choosing a different scale for representation could have changed the results of the survey.

The third part of the survey was about ranking ring roads according to their saliency (distinctiveness). Smaller images were used in this part of the experiment in order to facilitate comparison. This may have encouraged the participants to focus more on the visual aspect of the ring road because they did not have access to as many details of the map compared with the previous parts of the experiment.

Last but not least, linguistic considerations must be taken into account. The survey was written in English, and while all participants were proficient in the language, only a minority were native English speakers. Two-thirds of the participants were native French speakers, which raises the possibility that the assumed definition of "ring roads" among French speakers may differ from that of participants from other linguistic and cultural backgrounds. For the native French speakers who completed the survey under our supervision at the office and asked for clarification on the term "ring roads," we explained that it referred to what they might commonly know as "tour de ville" or "périphérique urbain." This clarification could have influenced their understanding and responses, highlighting the importance of cultural and linguistic context in survey-based research.

C.3 Measuring ringroadness

C.3.1 Characterization properties and congruence

Our perceptual study seems to confirm that there is no singular or universal way of perceiving ring roads on maps. It highlights the difficulty of classifying roads into just two categories: "ring road" or "non-ring road". Instead, we propose defining "**ringroadness**" as a measure of congruence that evaluates how closely a set of roads aligns with the archetype of a ring road. The degree of similarity between a complex

object and its *ideal prototype* is called the "degree of congruence" and can be expressed as a value between 0 and 1. In our case, a very recognizable ring road would have a ringroadness of almost 1 while an object which is barely a ring road would have a ringroadness close to 0. A measure of congruence is especially useful when the studied concept has an unclear definition yet many properties that we can describe individually (Lüscher, Weibel, and Mackaness 2008). This nuanced approach reflects the diversity in user perception and provides a more flexible framework for analysis and classification, keeping the possibility to create a class for ring roads in the database by applying a specific threshold.

The wider ambition is to create an algorithm that automatically detects and quantifies the ringroadness among a collection of roads. From our observations we have identified in Table C.4 a list of ring road properties and alongside, potential measurement techniques and indicators. Some indicators, such as compactness and length, may be computed directly using existing GIS tools, while others, such as the difference in road densities, need to be computed indirectly as explained in Zhang 2004 or by using space syntax techniques⁴ (Czerkauer-Yamu and Voigt 2011). The calculations for some of these indicators are presented in table C.5. We can observe the beginnings of a correlation between ringroadness and certain properties, such as the size of the ring road and its area compared to the urban area's surface. Although Lille's ring road appears quite large, its low compactness (stemming from a lack of roundness) may be the reason for its lower score.

The survey revealed that some properties are mandatory to the definition of ring road, while others less so. These differences need to be further studied and prioritised in any implementation. One approach to compute ringroadness would be to rate all the properties between 0 and 1, sum them, and divide by the number of properties. We could then bring more subtlety to the algorithm by giving each property a weight.

As we see in Table C.4, measuring ringroadness requires us to define (and detect) other complex objects such as "arterial roads". Congruence can again be used to help us classify such objects. We can use congruence recursively until all the complex features are rated (Lüscher, Weibel, and Mackaness 2008).

⁴Space syntax is a method for measuring the relative accessibility of different locations in a spatial system, including street networks (Hillier and Hanson 1984). It combines graph theory, geometry and geography in order to bring out implicit spatial information from a dataset. A nice project is available at <https://spacesyntax-openmapping.netlify.app/> and display a space syntax model of Great Britain's road network.

Property	Landmark type	Category	Measurement techniques / indicators
Is associated with a particular city	Structural	Ontology	Spatial analysis
Is large enough	Visual & Structural	Topology	Length, radius, area
Approximates a circle	Visual	Topology	Compactness, convexity
Connects most parts of the city	Structural	Space syntax	Closeness centrality
Contains X% of the city area	Visual & Structural	Topology	inner city area / city area
Contains the gravitational center of the city	Visual & Structural	Topology	GIS from city polygon
Forms a closed cycle	Visual	Graph theory	Cycles detection
Names of roads include "ring" or equivalent	Cognitive	Dataset	Explicit attribute
Connects with X major arterial roads	Structural	Graph theory	Strokes
Roads internal to the ring road have higher density than external roads	Visual	Graph theory	Graph density
Allows fast flow of traffic	Structural	Space syntax	Betweenness centrality
Minimises the number of changes in direction	Structural	Space syntax	Angular centrality
Has fewer intersections than the average in the road graph of the city	Structural	Graph theory	Graph density

✿ Table C.4 : Properties of a ring road, in order of priority.

	St-Rémy-de-Pce	Issoire (1)	Lille	Caen
Ringroadness from the survey (/5)	not a ring road	1.71	3.29	4.75
Urban area	4.9km ²	8.18km ²	123km ²	65km ²
Ring road area	0.07km ²	0.14km ²	8.5km ²	46km ²
Urban area in ring road area	1%	2%	7%	62%
Ring road length	1km	1.4km	12.7km	26.5km
Compactness	0.89	0.92	0.66	0.83
Closed cycle	Yes	Yes	Yes	Yes
Inner arterial roads connected	0	4	10	12
Outer arterial roads connected	4	3	9	9

⌘ Table C.5 : Some manually calculated indicators of few ring roads from the survey.

C.3.2 Ring road ontology

A conceptual ontology of ring roads was developed to synthesize and organize the insights derived from our study, as well as to provide guidance for potential future implementations of ring road detection algorithms (Figure C.13).

One of the most widely accepted definitions of ontology is provided by Gruber 1995 ⁵:

"An **ontology** is an explicit specification of a conceptualization. [...] A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose."

Participants highlighted the role played by ring roads (easing inner city car congestion). In turn, this functional perspective governed their expectation that the ring road would be circular in form, would largely ‘contain’ the city, be connected to a set of major roads that radiated outwards from the city, have a limited number of connections, and be of a minimum geographical size. Whilst some of these properties are relatively simple to model, other properties are more difficult to model such as the ‘interactions’ between the ring road and 1) other entities (such as rivers, and the surface morphology) and 2) other bounding structures. Participants noted that other bounding structures reinforced the notion of a ring road (for example where the road followed the lake shore). In some cases other network structures ‘complimented’ the path of the road (such as where rail and river networks ran parallel to the road).

The very process of creating an ontology of city ring roads (Figure C.13) helped us to identify other entities and relationships that we need to model. With the ring road concept at its core, our ontology incorporates defining related concepts such as "city"

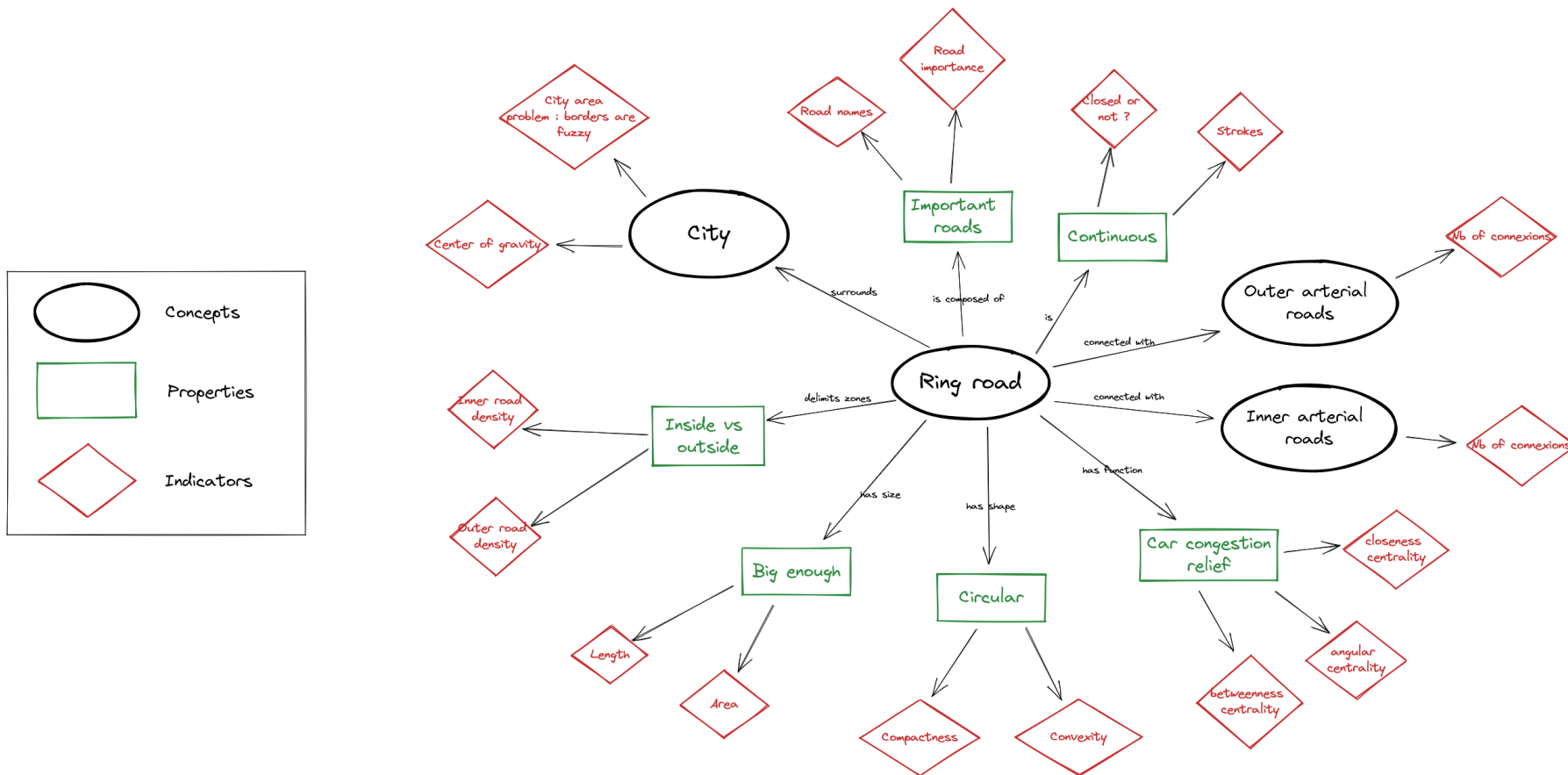
⁵While this definition primarily applies to formal ontologies, which are developed using logic-based languages such as OWL (Web Ontology Language), it may also include conceptual ontologies like the one we propose in this work. For a more in-depth exploration of the use of ontologies in geographic sciences, refer to Couclelis 2010.

and "arterial roads." One challenge in defining what area constitutes a city is the lack of clear boundaries. A practical solution must be selected—for instance, using urban land cover density as an indicator to delineate the city's extent. Once defined, the city's center of gravity can be calculated and compared with the ring road's center of gravity to determine if the ring road is well-centered. Arterial roads can be detected using stroke algorithms and further categorized into two classes: inner arterial roads and outer arterial roads. We hypothesize that the connection between the ring road and both classes of arterial roads is a critical factor in determining its "ringroadness".

For each identified property of a ring road, we propose corresponding indicators:

- **Important Roads:** These can be identified using road classifications and detecting specific road names (e.g., roads with "ring" in their name).
- **Continuity:** This can be measured by analyzing the number of strokes that compose the ring road and checking whether it forms a complete loop.
- **Congestion Relief:** This property can be assessed using space syntax tools for centrality, enriching the road network with metrics that quantify transportation potential and flow.
- **Circularity:** Metrics like convexity and compactness of the polygon formed by the ring road can capture its roundness.
- **Size:** This can be determined by calculating the total length of the road network forming the ring road and the area enclosed by its polygon.
- **City Boundary Delimitation:** This property can be evaluated by analyzing road density within and outside the ring road to measure its effectiveness as a boundary between inner and outer urban areas.

By employing these indicators, we aim to systematically quantify the properties that contribute to a road's "ringroadness," laying a foundation for automated detection and analysis.



☼ Figure C.13 : Ontology of a ring road: Visual and structural characteristics from a human perspective and their linked indicators.

C.4 Conclusion of the chapter

In this chapter, we explored the perception of ring roads in topographic maps at the urban scale. Their naturally distinctive geometry, combined with their functional impact on urban connectivity, positions them as high-potential landmarks in maps and high-potential anchors in mental representations. However, leveraging these ring roads to improve navigation in pan-scalar maps first requires the ability to detect them and store this information in databases. Only then could they be utilized for anchor-based zooming, as proposed in the Lost in Zoom project, incorporated into the generalization process by prioritizing the preservation of their shape over other roads, or assigned with a distinct colour to further enhance their visibility and perception.

While techniques for identifying circular roads have been proposed (Heinzle, Anders, and Sester 2006), the fuzzier concept of a "ring road," which we argue has greater anchoring potential, remains underexplored. To address this gap, we conducted a preliminary perceptual survey involving 28 participants. The results shown that various factors affect the perception of ring roads; they are complex and their perception can be reinforced (and eroded) by the presence of other entities. Despite this complexity, the survey enabled us to establish a list of properties characterizing ring roads. For each property, we identified one or more indicators, drawing from existing literature to facilitate their measurement. Finally, we synthesized these findings into a conceptual ontology, offering a structured framework for further research.

While we did not pursue further development of this topic in the thesis, the next logical step would be to translate these findings into an automated computation of "ringroadness" for a selected group of roads, based on the extracted properties. An important challenge lies in identifying candidate groups of roads for ring road evaluation. A potential filtering method could involve existing algorithms designed to detect circular roads within a network. However, this approach is not without limitations. We observed that not all ring roads form closed circles, meaning such roads could be overlooked. Additionally, in the ideal case where open-circle roads would be included as candidates, calculating properties like "interior" versus "exterior" would become more complex and may require artificially closing the circle for analysis, for example with a convex hull algorithm. Despite these challenges, we believe there are two viable approaches for combining the proposed properties into a "ringroadness" index: (1) using a congruence principle to weight and synthesize the indicators, or (2) employing a machine learning model trained on a dataset enriched with additional inputs from a similar survey. Once a methodology for detecting ring roads within a city's road network is established, an automated detection process could be envisioned. This process would extract a road network subgraph for each city and enrich the database with information in the form of: "ring road for the city of X".

We believe this methodology—using surveys to identify key properties and machine learning to determine the optimal combination of indicators—could be extended to other types of landmarks with potential as cognitive anchors in pan-scalar maps but that are not easily detectable. Examples include urban centers and public squares for city navigation, transportation hubs such as train stations or airports, specific architectural boundaries like city walls and fortifications that mark a city

edges, or salient natural features like mountain ranges and large valleys.

This investigation into ring roads also provides broader insights into the nature of cartographic landmarks. It highlights the fact that landmarks are often fuzzy rather than crisply defined entities: their geometries can vary, and their boundaries are rarely absolute. Moreover, landmarks do not exist in isolation but emerge through interaction with other features—for instance, a ring road only gains meaning in relation to the urban area it encircles. Gestalt theory supports this notion by suggesting that users tend to visually “complete” missing parts of a perceived shape, meaning that even open or incomplete ring roads may be interpreted as closed loops. Additionally, landmarks often involve groupings of features that reinforce one another; in the case of ring roads, parallel infrastructure such as railways or rivers following a similar path can enhance their perceptual salience. Finally, landmarks can be defined through a complex combination of geometric, semantic, and topological properties, underscoring the need for flexible and context-aware detection methods.

Lastly, we cannot conclude this chapter without reflecting on some broader implications of this work. First, the question of whether a simpler, manual approach might be more effective for specific goals such as detecting ring roads. For instance, instead of relying on complex algorithms, one could start by identifying municipalities of a size likely to have a ring road and then directly consulting local authorities for their input on what constitutes the ring road in their area. Given the slow evolution of ring road networks, a manual approach might provide a pragmatic alternative or complement to automated detection. Second, designing maps with a focus on roads reinforces the notion that cars are central to our way of life — a perspective that is increasingly at odds with the principles of sustainability. Do we truly want urban ring roads to become even more prominent landmarks? While platforms like OSM and Google Maps have introduced specialized visualizations for bicycles and public transportation, it remains the case that road maps are still the "default" and most widely used format. This raises important questions about the future direction of cartographic design and its alignment with sustainable development goals.

Chapter D

Multimodal learning for remarkable buildings detection

In this chapter, we address the challenge of detecting remarkable buildings in a cartographic database. We create a training dataset and test various machine learning approaches. This project emerged as an opportunity closely related to the focus of our thesis, as explained in the first section. This work was conducted in collaboration with Eric Lafon as part of his Master’s thesis in geomatics at the University of Toulouse, which I supervised. We worked on this project for six months, with Eric focusing primarily on creating and managing the vectorial dataset and implementing the random forest algorithm, while I concentrated on image extraction, deep learning implementations and result analysis. Both of us contributed to the literature review.

D.1 Motivations

Our motivations for this project were threefold:

As part of its activities, the French army requires functional maps for navigation and object localization, adhering to NATO standards for operational maps (1:50,000 scale). IGN produces and supplies these maps to the Inter-Army Geographic Establishment (EGI) in the form of a multi-scale digital cartographic pyramid (Scan-Défense), ready for printing maps. Few years ago, the EGI highlighted a need to improve IGN maps, particularly in the representation of remarkable buildings. This demand led to three successive student projects from 2022 to 2023, presented in the next section, the latter two of which I supervised.

Second, as detailed in Chapter A, remarkable buildings can serve as landmarks on large-scale maps and as anchors in users’ mental representations. Buildings are crucial features on maps at relatively large scales, and identifying and representing the most significant buildings is a challenge. Some buildings are important due to their function, such as hospitals or town halls, while others gain significance simply because they serve as visual landmarks in the landscape. Discussions with the IGN map production team revealed that these latter buildings are particularly challenging to identify, as they lack specific metadata in the database. Their identification is therefore of interest to us, either to highlight certain buildings on pan-scalar maps or to utilize them in anchored zooming scenarios.

Finally, the detection of remarkable buildings presents a specific challenge that aligns well with the multimodal machine learning techniques we wanted to explore. Remarkable buildings are not only defined by their intrinsic database attributes but also by their geographical context, which plays a significant role in their saliency. However, this contextual information is often missing in the database and difficult to encode (see the literature review). To address this, we saw an opportunity to apply multimodal models, which are capable of integrating multiple types of data to improve prediction accuracy. Our approach involves combining the attribute data of buildings (e.g., size or function) with one or more map images that include the building. These images provide a geographical context, making it easier in theory for the model to identify which buildings are truly remarkable. This technique, therefore, seemed particularly well-suited to the task and allowed us to test the potential of multimodal machine learning in landmark detection.

D.2 Project history

D.2.1 Fieldwork student project

During the summer of 2022, master’s students from ENSG conducted a study on the representation of IGN maps for the EGI, focusing on the French towns of Forcalquier (5,000 inhabitants, see Figure D.1), Manosque (20,000 inhabitants), and Draguignan (60,000 inhabitants). To find remarkable building in databases, the students proposed a simple decision tree based on pre-enriched building properties (primarily criteria related to the function of the building, its height, size, and shape). They also worked on map symbology, offering suggestions to improve the IGN map representation (see Figure D.2).

D.2.2 Research initiation project

This project was subsequently taken up by three other master’s students (Ba R., Rouillard M., Talec C.) as part of a Research Initiation Project (PIR), which I supervised on a schedule of one day per week over three months during the winter of 2022. In this project, the students aimed to extend the previous study further. Building on the methodology of the earlier students, they developed an initial decision tree (T1) to detect remarkable buildings (Figure D.3).

Using this decision tree (T1), they conducted a first selection of buildings, identifying around 3,000 remarkable buildings out of a total of 360,000. They then refined this selection by manually reviewing the buildings to remove those that did not seem remarkable and to add new remarkable buildings that the algorithm had missed in the area. This enriched selection was subsequently used to train a random forest algorithm which in turn gave us a new more complex decision tree (T2). The idea was to use machine learning to integrate human corrections into the decision tree to improve the quality of the selection.

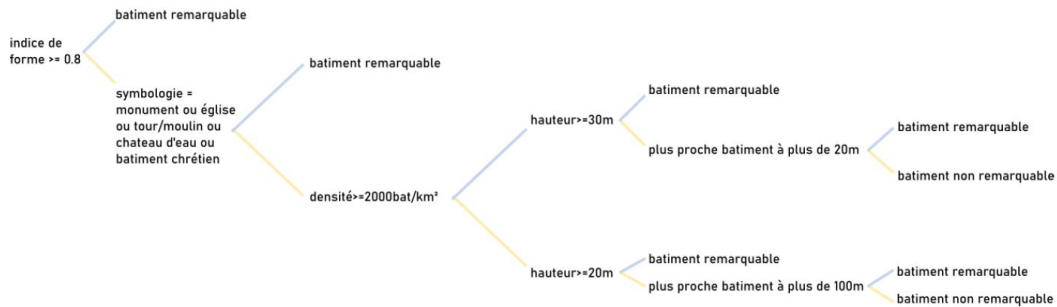
The two decision trees, T1 and T2, were tested on a dataset of 30,000 buildings. Tree T1 categorized 333 buildings as remarkable. Among these, 214 buildings were



✿ Figure D.1 : Photograph of the town of Forcalquier taken from an elevated vantage point. Highlighted in red are the Notre-Dame-du-Bourguet Cathedral (on the left) and the Forcalquier Citadel (on the right), which stand out prominently in the landscape and can serve as landmarks.



✿ Figure D.2 : Map proposal of Draguignan created by the students during the first project, with buildings identified as remarkable (according to their criteria) highlighted in black.



☼ Figure D.3 : Initial decision tree proposed by the students for detecting remarkable buildings. The tree considers the buildings' shape index, function, height, the surrounding building density, and the distance to the nearest building.

also classified as remarkable by T2. However, 119 buildings were downgraded by T2, while T2 added 26 additional buildings as remarkable.

The students then examined the 145 buildings classified differently by T1 and T2 on Google Maps to qualitatively assess the differences between the two algorithms. Their conclusion was that the results were mixed, and the differences in classification did not necessarily favor T2 (Figure D.4).

This outcome did not lead us to conclude that using a random forest algorithm is inherently flawed. Rather, it suggested that the initial training dataset (derived from T1) and the enriched building data might not have been well-suited for this approach. Another possible conclusion is that the features used did not sufficiently capture what makes a building remarkable. Future experiments could explore training datasets created entirely by humans to better evaluate the potential of the algorithm.



☼ Figure D.4 : Two examples of errors made by the random forest decision tree (T2) that were correctly classified by the original decision tree (T1). On the left, a building obscured by trees was classified as remarkable by T2 but not by T1. On the right, a church classified as remarkable by T1 was not recognized as such by T2. Source: Google Street View.

D.2.3 Master thesis internship

Finally, the project presented in this chapter builds upon the work conducted during Eric Lafon’s six-month internship in the summer of 2023. His research aimed to develop a method for detecting remarkable buildings while exploring the potential of multimodal machine learning techniques.

D.3 About remarkable building detection


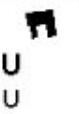
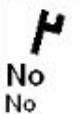



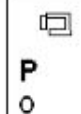
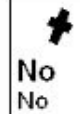
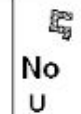
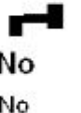

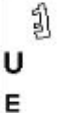





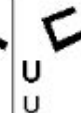


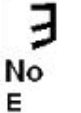

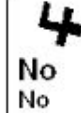




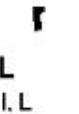
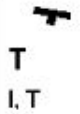


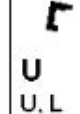


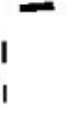
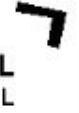
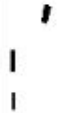


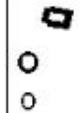
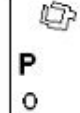

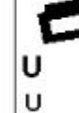
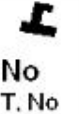

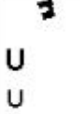

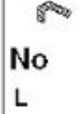
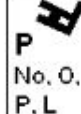

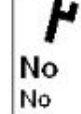
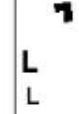

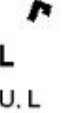
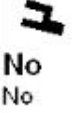

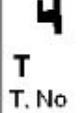

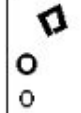

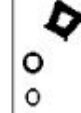

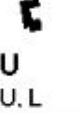
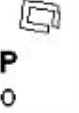
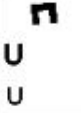

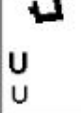
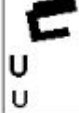
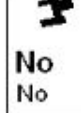
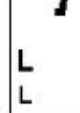
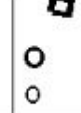

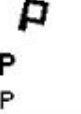
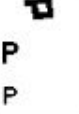

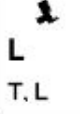

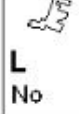
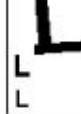
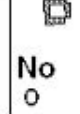
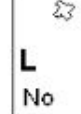
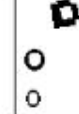
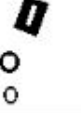

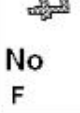


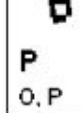
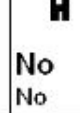

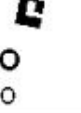
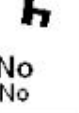

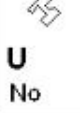
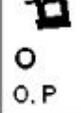

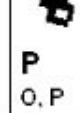

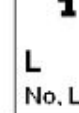
The problem of detecting remarkable buildings falls within the broader challenge of characterizing buildings and identifying patterns they form in maps. These issues have been extensively studied in the field of cartographic generalization (Lichtner 1979; Mackaness, Ruas, and Sarjakoski 2007), where researchers seek to preserve the key characteristics of buildings and building arrangements despite modifications across scales. While remarkable building detection focuses on identifying individual buildings, the problem requires considering its surrounding environment, just as for generalisation.

The first step in detecting remarkable buildings is often enriching the existing database with relevant attributes (see Section A.3.2.1). The most commonly used building characteristics include shape, orientation, and group membership (e.g., alignment with other buildings). Various indicators have been proposed in the literature to capture these features. Biljecki and Chow 2022 conducted a systematic review of building-related metrics, providing a comprehensive synthesis of existing approaches.

Several studies have focused on intrinsic building properties. Duchêne, Bard, Barillot, Ruas, Trévisan, and Holzapfel 2003 introduced a method for calculating building orientation, while Rainsford and Mackaness 2002 developed an automated process to classify buildings by matching them to predefined geometric templates (Figure D.5). Additionally, Wei, Xu, Xiao, Shu, Cheng, Wang, and Liu 2023 explored the use of knowledge graphs to recognize characteristic building patterns.

Concerning group membership, Yang 2008 proposed a method for identifying building groups by combining geometric and spatial analysis, emphasizing the importance of detecting outliers. Similarly, graph-based approaches have been explored for pattern recognition in building arrangements (He, Zhang, and Xin 2018). The principle of data matching has also been applied to detect specific building arrangements. For example, Xing, Wu, Gong, Du, and Liu 2022 proposed a method for identifying collinear building alignments to ensure their preservation during cartographic generalization.

In this chapter’s experiments, we enriched our building dataset by incorporating both intrinsic building attributes and contextual information about the surrounding buildings. However, while we noticed that some buildings gain significance not individually but as part of a larger ensemble, such as hospitals, university campuses or adjoining buildings in the city center, we chose not to delve into building aggregation methods.

 F F	 U U	 No No	 E E	 No F	 I I	 P O	 No No	 No U	 No No
 No No	 T No	 U E	 I I.L	 P P	 U U	 T T	 F F	 U U	 I F
 O O	 L E.F	 No E	 F F	 I I	 No No	 No No	 F F	 L L	 T T
 U No	 L I.L	 No P.L	 T I.T	 F F	 L L	 T I.L.T	 U U.L	 E E	 F F
 I I	 L L	 I I	 P O	 I I	 F F	 O O	 P O	 L L	 U U
 No No	 No T.No	 U U.L	 U U	 F F	 No L	 P No. O. P.L	 O O	 No No	 L L
 T Z	 L U.L	 No No	 P O	 T T.No	 L U	 O O	 L L	 O O	 F T.F
 U U.L	 P O	 U U	 I I.T.F	 U U	 U U	 No No	 L L	 O O	 L No.L
 P P	 P P	 L L	 L T.L	 T I.L.T	 L No	 L L	 No O	 L No	 O O
 O O	 P P	 L U	 No F	 E E	 L I.L	 P O.P	 No No	 P O	 L L
 O O	 No No	 T T	 U No	 O O.P	 O O	 T T	 P O.P	 T T	 L No.L

⌘ Figure D.5 : An inventory of buildings associated with their matching geometric templates proposed in Rainsford and Mackaness 2002, whose elegance and curiosity warrants a full-page display.

D.4 About multimodal learning

Multimodal learning refers to training a machine learning model using data from multiple sources or of different types, such as text, images, spatial information, or audio recordings. The simplest approach to multimodal learning involves leveraging unimodal models—each specialized in a specific data type—and combining their predictions using ensemble learning techniques. **Ensemble learning** is a broad category of machine learning strategies that improve prediction accuracy by combining multiple models, even if each individual model has only moderate predictive power. This can be done by aggregating their predictions in a weighted manner or by fusing different model architectures into a single unified prediction system. A detailed introduction and review of ensemble learning strategies can be found in Sagi and Rokach 2018 and Dong, Yu, Cao, Shi, and Ma 2020.

In cartography, multimodal deep learning is increasingly used in remote sensing for image interpretation, where it combines data from various sources such as multispectral imagery, LiDAR, and radar (Li, Branco, and Zhang 2023). In sociology, Suel, Bhatt, Brauer, Flaxman, and Ezzati 2021 proposed combining satellite imagery with Google Street View photos to assess urban inequalities. Another common application is real estate price prediction, which closely aligns with the approach taken in this chapter. Early models in this field incorporated photographs of buildings alongside traditional tabular data—such as floor area, number of rooms, and construction date (Zhao, Chetty, and Tran 2019; Kucklick and Müller 2021). More recent models integrate multiple data types simultaneously, including structured building attributes, images, textual descriptions, and even geographic coordinates (Li, Branco, and Zhang 2023). Wanga, Chen, Su, Ting-Yun, and Huang 2021 further expanded this approach by incorporating satellite imagery to provide contextual environmental information around buildings.

As introduced in the motivations section, our approach leverages multimodal learning by integrating cartographic images to improve the classification of remarkable buildings. Specifically, we extract a map image for each building to be evaluated, allowing predictive models to leverage additional spatial context.

D.5 Experiment

D.5.1 Main idea and objectives

The goal of this experiment is to assess whether a multimodal model that combines both tabular data and cartographic images can outperform a unimodal model trained solely on tabular data for the detection of remarkable buildings. If successful, this approach could suggest that multimodal learning is a promising method for detecting various types of landmarks more generally.

To perform this comparison, we propose testing the following models:

- A random forest trained on tabular data, a standard baseline in the literature.
- A multilayer perceptron (MLP) trained on tabular data.

- A convolutional neural network (CNN) trained exclusively on extracted map images centered around buildings.
- A multimodal model combining both tabular data and map images for each building.

To achieve this, we will follow these steps:

1. Obtain or construct a vector dataset containing both remarkable and non-remarkable buildings within a given area, annotated by experts. This dataset will serve as training data for our models.
2. Enrich the dataset with additional spatial attributes for the tabular data that could improve detection performance, such as the distance to the nearest building.
3. Extract map images for each building in the training dataset to enable training of the multimodal model.
4. Obtain or construct a test dataset, also containing annotated remarkable and non-remarkable buildings.
5. Implement the random forest, MLP, CNN, and multimodal models and train them.
6. Evaluate and compare model performance to determine whether multimodal learning provides a significant advantage in detecting remarkable buildings.

D.5.2 Building an experimental dataset

The dataset proposed in this section and all the data related to our experiments are available online¹.

D.5.2.1 Definition of a remarkable building

To align our work as closely as possible with the military's needs, we requested an annotated dataset of buildings classified as "remarkable" by their experts. However, they were unable to provide such a dataset. As a result, we had to create our own dataset for our experiments. To better understand what the military considers a "remarkable building", we visited the Inter-Army Geographic Establishment (EGI) for discussions. The definition that emerged from these discussions is as follows:

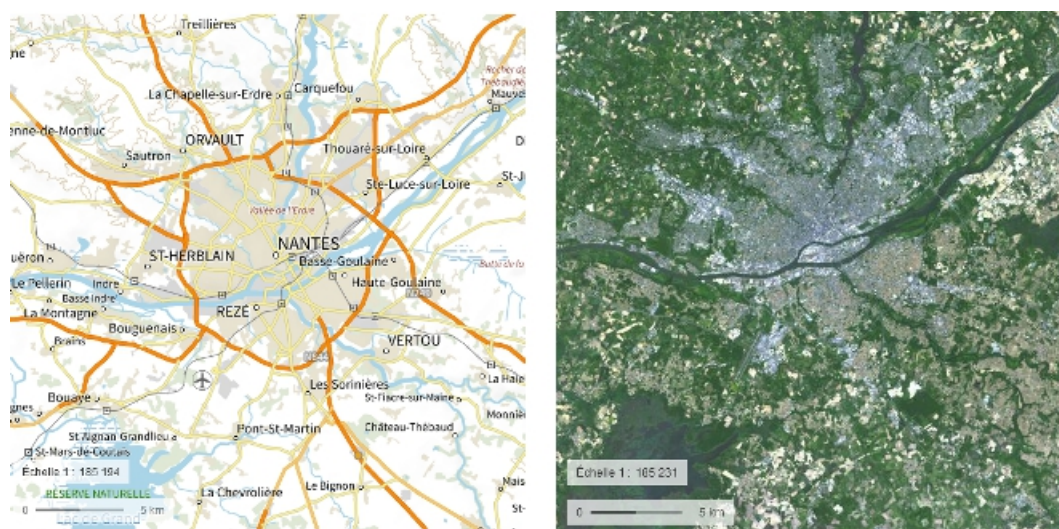
A building is considered **remarkable** if it visually stands out from its surroundings in the real world and can be used as a reference point for orientation or for giving directional instructions.

¹<https://doi.org/10.5281/zenodo.14879919>

It is important to distinguish between a building’s remarkability on a map and its remarkability in the real world. The military’s request specifically focused on improving the representation of buildings in maps based on their real-world prominence.

D.5.2.2 Data and study area

As in the two preceding projects, we chose to use building data from the BD TOPO database, produced by IGN. We aimed to select a study area that included a medium-sized city with both urban and rural zones to ensure a diverse dataset. Nantes was chosen as it met these criteria and because Eric was familiar with the city—a factor we considered an advantage rather than a bias in constructing a high-quality dataset. An overview of the selected study area around Nantes is shown in Figure D.6.



☞ Figure D.6 : The city of Nantes as represented in the IGN map (left) and in an aerial photograph (right). The study area includes a mix of urban, agricultural, industrial, and natural zones, containing approximately 300,000 buildings according to BD TOPO.

D.5.2.3 Data annotation

Given that the semi-automated dataset approach used by students in 2021 did not yield conclusive results, we opted for a fully manual annotation process for this project. The task of extracting and annotating buildings was carried out by Eric. First, he extracted the buildings from BD TOPO within the study area, resulting in approximately 300,000 buildings. He then proceeded with the annotation, dividing the study area into multiple sub-zones to facilitate systematic annotation across different sessions.

The annotation process followed these steps:

1. Selection of a sub-zone.

2. Using Google Earth’s 3D visualization tool, he navigated through the area at an altitude of approximately 10 to 20 meters above the ground.
3. When a remarkable building was identified, he located its corresponding feature in the BD TOPO dataset, visualized in QGIS, and marked it as remarkable in a newly created column (Figure D.7).

We initially annotated the dataset together, discussing our reasoning for considering a building remarkable or not, ensuring alignment with the definition previously established with the military. Once we reached a consensus on the criteria for remarkability, Eric continued the annotation independently. In total, 1,129 buildings were labeled as remarkable over the course of one week, out of the 300,000 buildings in the study area.

Since annotation is inherently subjective, an ideal approach would have involved multiple annotators, allowing for a consensus-based dataset with greater reliability. However, due to time constraints within Eric’s master’s internship, we could not implement this additional step.



⌘ Figure D.7 : Right: A church identified as remarkable in Google Earth’s 3D view. Left: The same church located in QGIS within the BD TOPO dataset.

D.5.2.4 Balancing the dataset

The **1,129 buildings annotated as remarkable** by Eric accounted for only 0.37% of the total number of buildings in our dataset. Training models on such an imbalanced dataset would likely result in them converging toward a trivial solution where almost no buildings are predicted as remarkable. While this would yield high **accuracy** (the percentage of correct predictions out of all predictions), the **recall** (the percentage of actual remarkable buildings correctly identified) would be close to zero, meaning the model would fail to effectively detect remarkable buildings.

To address this issue, various techniques exist to balance datasets and improve classification performance. In this study, we chose downsampling (also called undersampling), which involves reducing the number of instances in the majority class (non-remarkable buildings) to create a more balanced dataset. While a more balanced dataset reduces bias, it also results in fewer training examples, which can impact model performance.

To strike a balance, we arbitrarily decided to retain three non-remarkable buildings for every remarkable one. Specifically, we randomly selected **3,387 non-remarkable buildings**, resulting in a **final downsampled dataset of 4,516 buildings**.

D.5.2.5 Data enrichment - part 1: Spatial analysis

For each building in our downsampled dataset, we computed the following additional attributes. For attributes that consider surrounding buildings, calculations were based on the full dataset of 300,000 buildings.

- Geometric properties:
 - Perimeter
 - Surface area
 - Shape indicator (as proposed by the PIR students, see eq. D.5.2.5)
 - Convexity (building surface area divided by the surface area of its convex hull)
 - Granularity (number of vertices defining the building polygon)
- Contextual properties (calculated within specific buffer zones around the building):
 - Average building height within a 125m, 250m, and 500m radius
 - Average building surface area within a 125m, 250m, and 500m radius

$$\text{shape_indicator} = \frac{\text{building_surface}}{\text{maximum_surface}} \times \frac{\text{building_perimeter}}{\text{maximum_perimeter}} \quad (\text{D.1})$$

In addition to these newly computed attributes, we retained several existing attributes from the BD TOPO database:

- nature, usage1, and usage2 (indicating building usage: industrial, commercial, residential, etc.)
- nb_logts (number of housing units in the building)
- nb_etages (number of floors)
- hauteur (building height)

After enriching our dataset and applying one-hot encoding to categorical labels, we obtained a total of 38 features.

D.5.2.6 Data enrichment - part 2: Extraction of map images

To provide contextual information about each building’s surroundings, we aimed to extract images that could aid the algorithm. After testing different image sizes and resolutions, we settled on square images of **256 × 256 pixels** with a resolution of 2.39 meters per pixel, corresponding to **zoom level 16**.

For each of the 4,516 selected buildings, we extracted a map image from the IGN Plan using these parameters (Figure D.8). To achieve this, we leveraged Python’s Geopandas and Shapely libraries to manage image extents and OWSLib to retrieve map tiles from the IGN Plan service. The related piece of code is available in the Zenodo files.



✿ Figure D.8 : Two images extracted from the IGN Plan, each centered around a building, which is highlighted in black for better visibility. In the actual dataset used for training and evaluation, buildings were kept in their original color.

D.5.2.7 Train-test dataset separation and spatial cross-validation

Ideally, we would have annotated a different geographical area—around another city—to evaluate our detection algorithms with as few biases as possible. However, due to time constraints, we opted to use part of the Nantes dataset we had already built for validation and testing.

Given the small size of our dataset, we employed cross-validation to improve the robustness of our results. **Cross-validation** involves splitting a dataset into multiple complementary training set, validation set pairs and training multiple models on these different pairs before evaluating their performance.

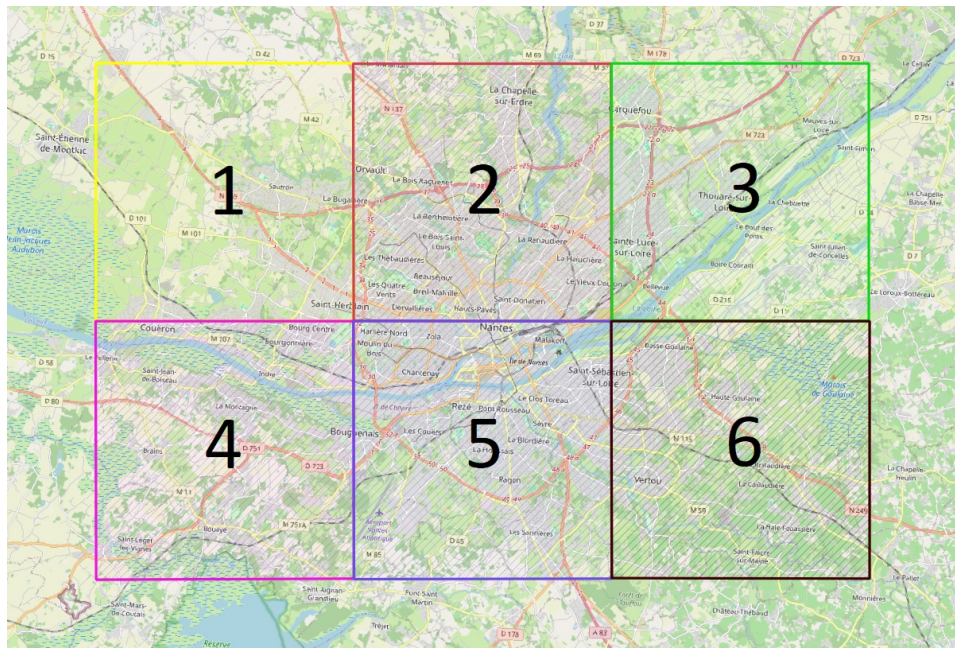
Our initial approach was to perform random cross-validation, where buildings (both remarkable and non-remarkable) were randomly selected for the training and validation sets. However, this method introduced significant spatial biases, particularly when incorporating map images as input. As illustrated in Figure D.9, two nearby buildings often have highly similar associated images. This creates a risk

of data leakage, where models inadvertently gain access to validation data during training, compromising evaluation integrity.



✿ Figure D.9 : Two images associated with buildings located close to each other, demonstrating the risk of data leakage in random cross-validation.

To address this issue, we divided our initial study area into six sub-zones (Figure D.10). We then configured our cross-validation so that each sub-zone was used as a validation dataset once, while the remaining sub-zones were used for training. The number of building and the proportion of remarkable ones for each sub-zone is available in Table D.1.



✿ Figure D.10 : The six sub-zones used for spatial cross-validation.

	Total buildings	Remarkable	% remarkable
Zone 1	385	98	25.5%
Zone 2	1181	282	23.9%
Zone 3	476	75	15.8%
Zone 4	619	189	30.5%
Zone 5	1345	377	28.0%
Zone 6	510	108	21.2%

✿ Table D.1 : List of the experiments compared in this appendix.

Although this spatial cross-validation approach reduces some biases, a few limitations remain:

- Despite spatial separation, all data still comes from Nantes and its surroundings, meaning the geographical characteristics of the validation and training areas remain similar.
- The sub-zones are imbalanced in terms of building counts. Depending on the cross-validation fold, the model may have more or fewer data points available for training and validation, potentially affecting performance consistency.

D.5.3 Model implementations and results

The implementation code for each model is available online² and on Zenodo³. All experiments were conducted on Google Colab, using GeoPandas and NumPy for data processing, and Scikit-Learn and Keras (TensorFlow) for machine learning tasks. To ensure fair comparison of training times across models, we utilized the "GPU L4" option in Google Colab for all model training sessions.

To evaluate our models, we monitored the following performance metrics:

- Accuracy – While commonly used, this metric can be misleading given our dataset’s imbalance (3:1 ratio of non-remarkable to remarkable buildings).
- Precision – Measures how many of the predicted remarkable buildings are actually remarkable.
- Recall – Measures how many of the actual remarkable buildings were correctly identified.
- F1 Score – A balance between precision and recall, making it a particularly relevant indicator for our study, as both metrics are important.
- AUC-ROC (Area Under the Receiver Operating Characteristic Curve) – Evaluates the model’s ability to distinguish between remarkable and non-remarkable buildings across different probability thresholds. This metric is particularly

²<https://github.com/TinQk/LostInZoom>

³<https://doi.org/10.5281/zenodo.14879919>

important for our study since different use cases may require adjusting the sensitivity of remarkable building detection.

- Average Precision (AP) – Assesses the trade-off between precision and recall across varying classification thresholds.

Unless stated otherwise, all reported results in this section correspond to the average performance across the six cross-validation folds for each model. For example, the precision value for the Random Forest model is the mean precision obtained across all six folds of its cross-validation.

In addition to these quantitative metrics, we also:

- Plotted ROC (Receiver Operating Characteristic) curves and PR (Precision-Recall) curves for each model.
- Constructed a combined confusion matrix across all folds. Since each subzone is used once as the validation set, this results in a confusion matrix that accounts for every building exactly once.

D.5.3.1 Random Forest

Random Forest (RF) is a widely used supervised machine learning technique and is often one of the first models tested in machine learning experiments. It has been used in many GIS-related tasks, from building classification to vandalism detection in VGI (Truong, Touya, and Runz 2020). A comprehensive introduction to Random Forests, along with implementations of numerous other machine learning techniques, can be found in [geron_hands-machine_2022](#).

❖ Architecture and hyperparameters

Training time was short enough to allow for extensive hyperparameter testing. Given that variations in hyperparameters did not significantly impact the results, we decided not to automate the tuning process. Instead, we manually adjusted the parameters and settled on the following configuration:

- Number of estimators: 350
- Maximum tree depth: 15
- Minimum samples required to split a node: 4
- Minimum samples per leaf node: 2
- Number of features considered for splitting: Square root of the total number of features

❖ Results

The training results for the Random Forest model are presented in Table D.2. The model achieved an average precision of 65% on the test set with a classification threshold of 0.5, which is a reasonable performance given the inherently subjective nature of identifying remarkable buildings. However, the recall was relatively low at

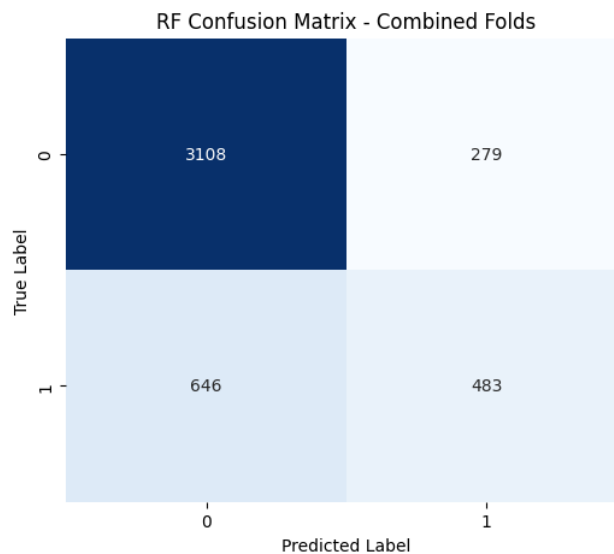
43%, meaning that a significant number of annotated remarkable buildings were not detected.

The F1-score averaged 0.51, and the Average Precision (AP) was 0.42, reflecting the trade-off between precision and recall. The AUC-ROC score of 0.78 indicates that while the model performs well in ranking buildings by remarkability across different decision thresholds, setting a precise boundary between remarkable and non-remarkable buildings remains challenging. The confusion matrix at a threshold of 0.5 is available in Figure D.11.

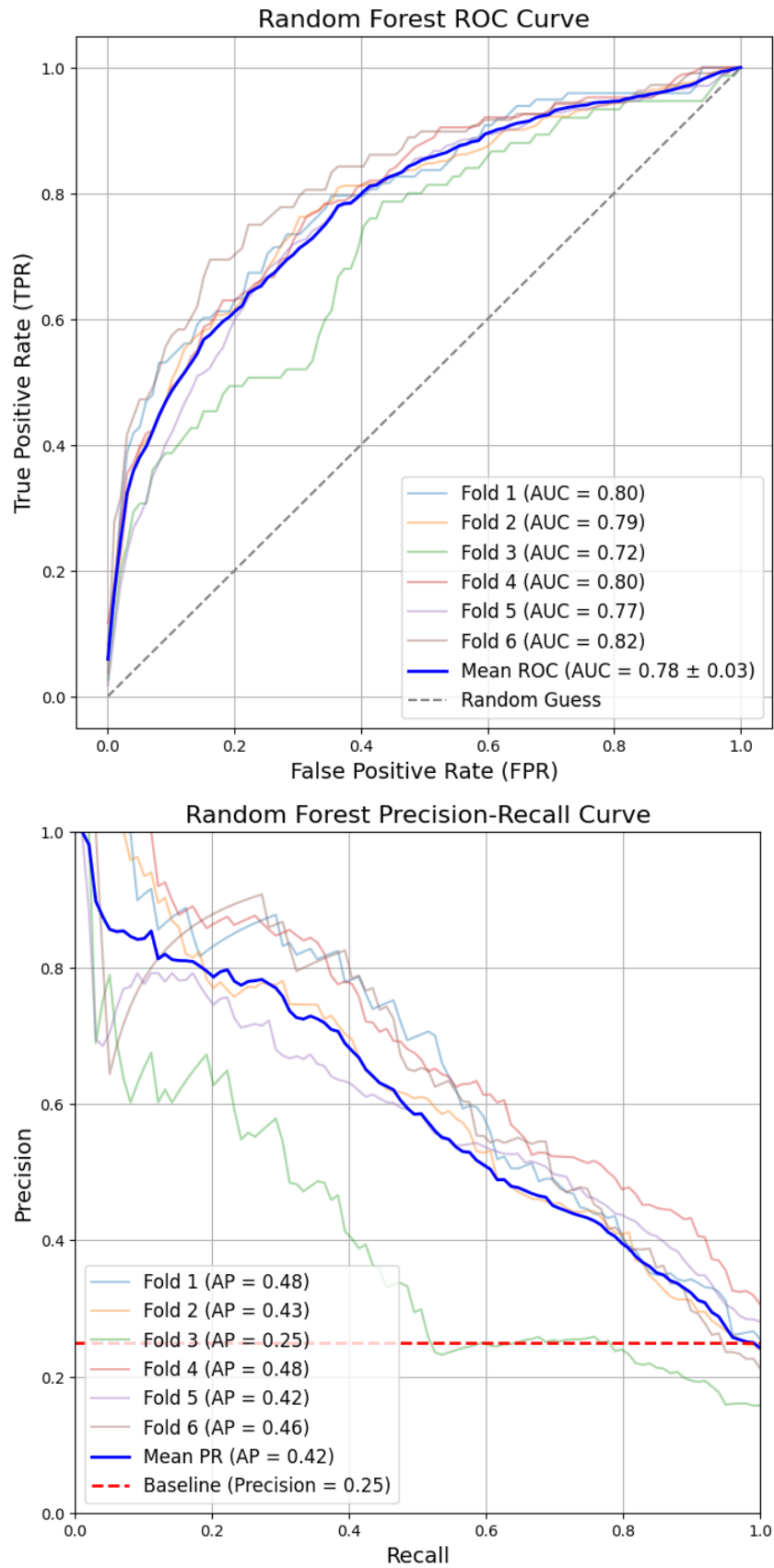
The ROC and Precision-Recall (PR) curves (Figure D.12) reveal that results for fold 3 were noticeably weaker compared to the other folds. To investigate further, we retrained the model multiple times with different random seeds, but the results remained consistent, with fold 3 consistently performing below the others. The exact cause of this discrepancy is difficult to pinpoint. However, we note that fold 3 contains the smallest number of remarkable buildings in its validation set, with only 75 annotated instances.

Metric	Train set	Validation set
Accuracy	0.96	0.80
Precision	0.96	0.65
Recall	0.88	0.44
F1 Score	0.92	0.51
ROC AUC	NC	0.78
Average Precision	NC	0.42

✿ Table D.2 : Evaluation of the random forest model. Each value represents the average across the six folds of the cross-validation. ROC AUC and Average Precision were not computed for the train set.



✿ Figure D.11 : Confusion matrix for the random forest model at a decision threshold of 0.5.



✿ Figure D.12 : ROC curve and PR curve for the random forest model.

D.5.3.2 Multi-Layer Perceptron

The Multi-Layer Perceptron (MLP) is the simplest type of artificial neural network and serves as one of the two branches in our multimodal model. Our implementation is available online⁴.

❖ Architecture and hyperparameters

We designed the model with an initial layer of 32 perceptrons, followed by a second layer of 16 perceptrons. To improve generalization and reduce overfitting, we applied a 30% dropout after each layer. A summary of the model architecture is shown in Figure D.13.

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 37)	0
dense (Dense)	(None, 32)	1,216
dropout (Dropout)	(None, 32)	0
dense_1 (Dense)	(None, 16)	528
dropout_1 (Dropout)	(None, 16)	0
dense_2 (Dense)	(None, 1)	17

Total params: 1,761 (6.88 KB)
 Trainable params: 1,761 (6.88 KB)
 Non-trainable params: 0 (0.00 B)

❖ Figure D.13 : The different layers of our MLP model, implemented with TensorFlow.

❖ Results

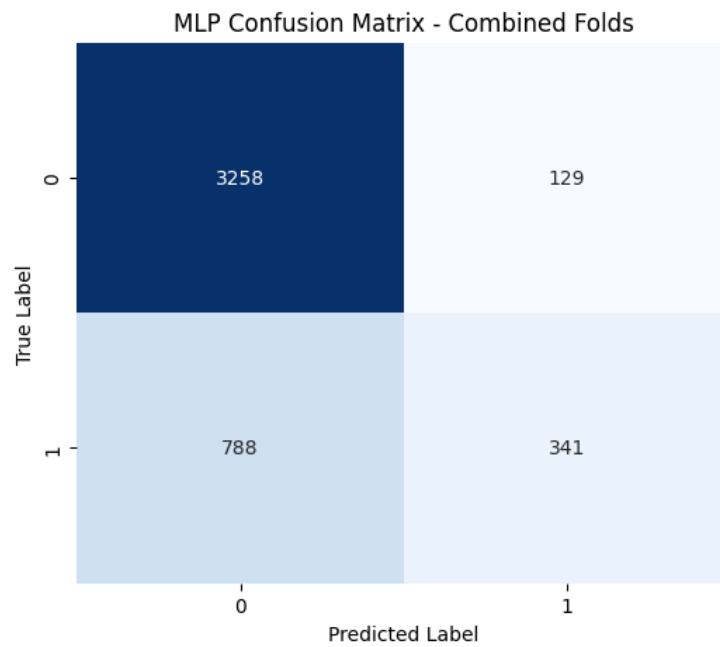
The training results for the MLP model are presented in Table D.3. It is interesting to compare them to the results of the random forest model as both have access to the same data for training (the tabular data). The MLP model performed slightly worse than the RF model, losing a few points in F1 score, ROC AUC and Average Precision. This phenomenon is anticipated by the literature. The ROC and Precision-Recall (PR) curves (Figure D.15) also reveal the same weakness observed in fold 3 for the random forest model. The confusion matrix at a threshold of 0.5 is provided in Figure D.14.

The model was trained for only 30 epochs, as we observed that performance plateaued after just a few epochs. The complete set of metric evolution curves over epochs is available in Appendix D, page 230. We notice on these curves that after a single epoch, the model already reached a reasonable level of performance, but in our plots the initial metric values are not displayed, and the curves start after epoch 1, which is why they appear relatively stable.

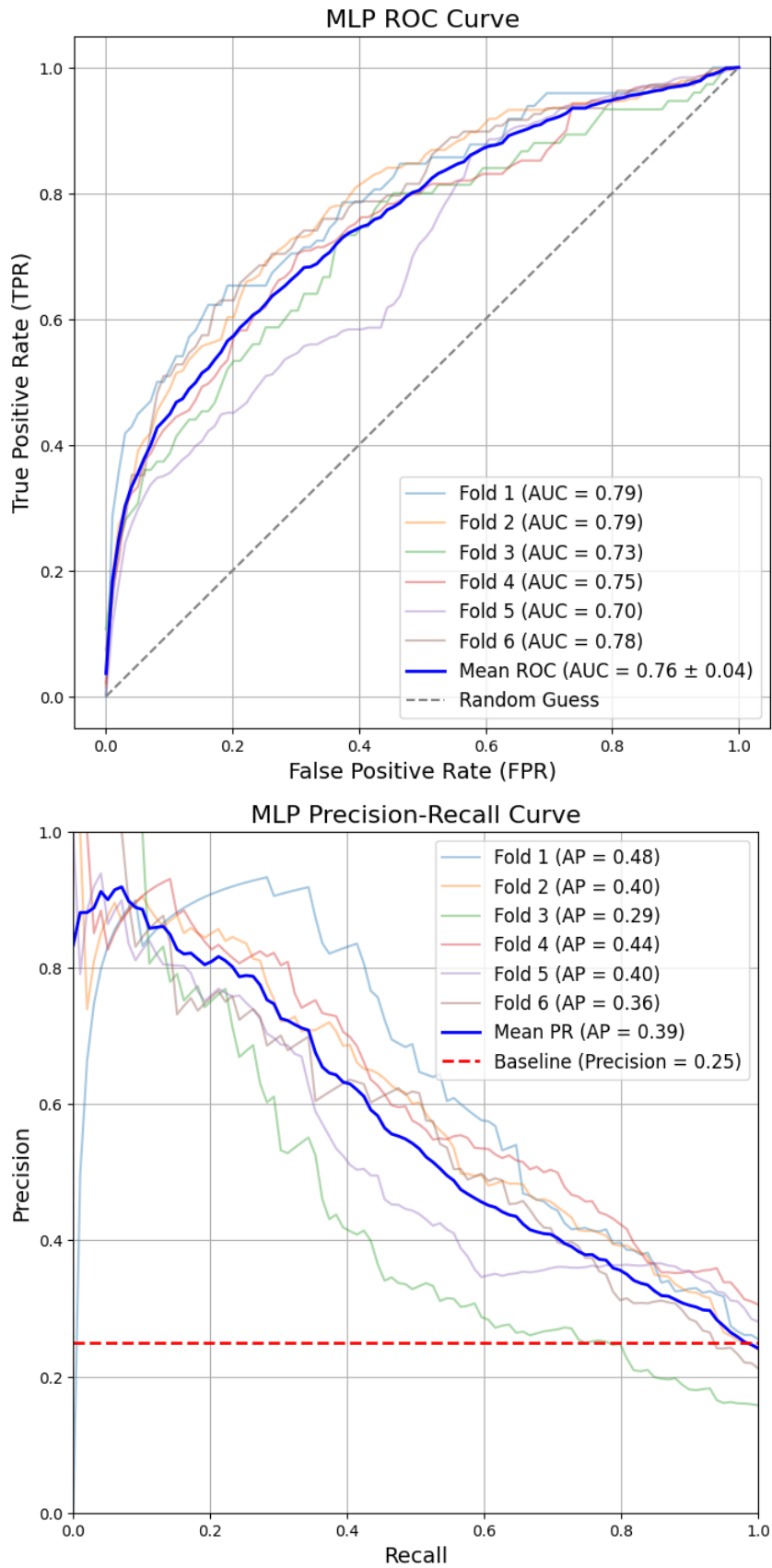
⁴<https://github.com/TinQk/LostInZoom>

Metric	Train set	Validation set
Accuracy	0.81	0.80
Precision	0.75	0.70
Recall	0.38	0.34
F1 Score	0.51	0.45
ROC AUC	NC	0.75
Average Precision	NC	0.40

✿ Table D.3 : Evaluation of the random forest model. Each value represents the average across the six folds of the cross-validation. ROC AUC and Average Precision were not computed for the train set.



✿ Figure D.14 : Confusion matrix for the MLP model at a decision threshold of 0.5.



⊛ Figure D.15 : ROC curve and PR curve for the MLP model.

D.5.3.3 Convolutional Neural Network

The Convolutional Neural Network (CNN) is one of the most widely used deep learning models for image analysis. In our study, it is the second of the two branches of our multimodal model. Our implementation is available on Github⁵.

❖ Architecture and hyperparameters

Our CNN model consists of three sets of convolutional layers, each followed by batch normalization and max pooling layers, and then two fully connected (dense) layers (Figure D.16). No skip connections were implemented between layers.

Initially, we trained the model with early stopping based on validation loss. However, due to instabilities, the validation loss occasionally reached a low point by chance, and the model then struggled to reach this low value again, triggering early stopping too soon. To address this issue, we trained another model, this time for 150 epochs without early stopping. The results discussed here are from the second model.

Layer (type)	Output Shape	Param #
image_input (InputLayer)	(None, 256, 256, 3)	0
conv2d (Conv2D)	(None, 256, 256, 32)	896
batch_normalization (BatchNormalization)	(None, 256, 256, 32)	128
max_pooling2d (MaxPooling2D)	(None, 128, 128, 32)	0
conv2d_1 (Conv2D)	(None, 128, 128, 64)	18,496
batch_normalization_1 (BatchNormalization)	(None, 128, 128, 64)	256
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 64)	0
conv2d_2 (Conv2D)	(None, 64, 64, 128)	73,856
batch_normalization_2 (BatchNormalization)	(None, 64, 64, 128)	512
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 128)	0
flatten (Flatten)	(None, 131072)	0
dense (Dense)	(None, 128)	16,777,344
dropout (Dropout)	(None, 128)	0
cnn_features (Dense)	(None, 64)	8,256
cnn_output (Dense)	(None, 1)	65

Total params: 16,879,899 (64.39 MB)
 Trainable params: 16,879,361 (64.39 MB)
 Non-trainable params: 448 (1.75 KB)

❖ Figure D.16 : The different layers of our CNN model, implemented with TensorFlow.

⁵<https://github.com/TinQk/LostInZoom>

❖ **Results**

The training results for the CNN model are presented in Table D.4, with its confusion matrix shown in Figure D.17. The model performed worse than the two models trained on tabular data. However, as seen in the ROC curve (Figure D.18), the model still manages to distinguish some remarkable buildings from the dataset.

This is an encouraging sign for our multimodal model, as it confirms that implicit information about building remarkability is present in the map images provided to the CNN. Additionally, we observed these same folds variation for our two trained CNN models:

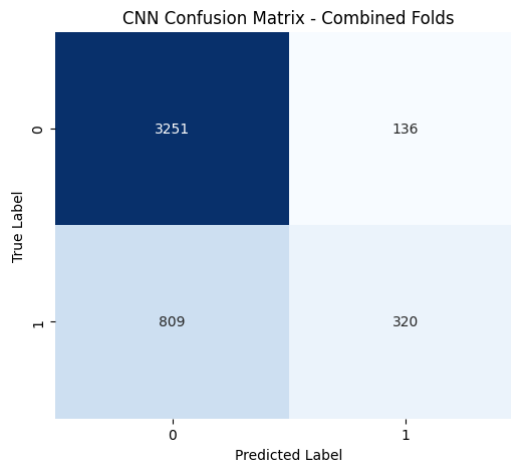
- Fold 5 consistently underperformed across both training attempts,
- Fold 3 achieved above-average results, contrasting with the previous models trained on tabular data.

These observations suggest that the two different data sources (tabular vs. image-based) could complement each other in a multimodal approach.

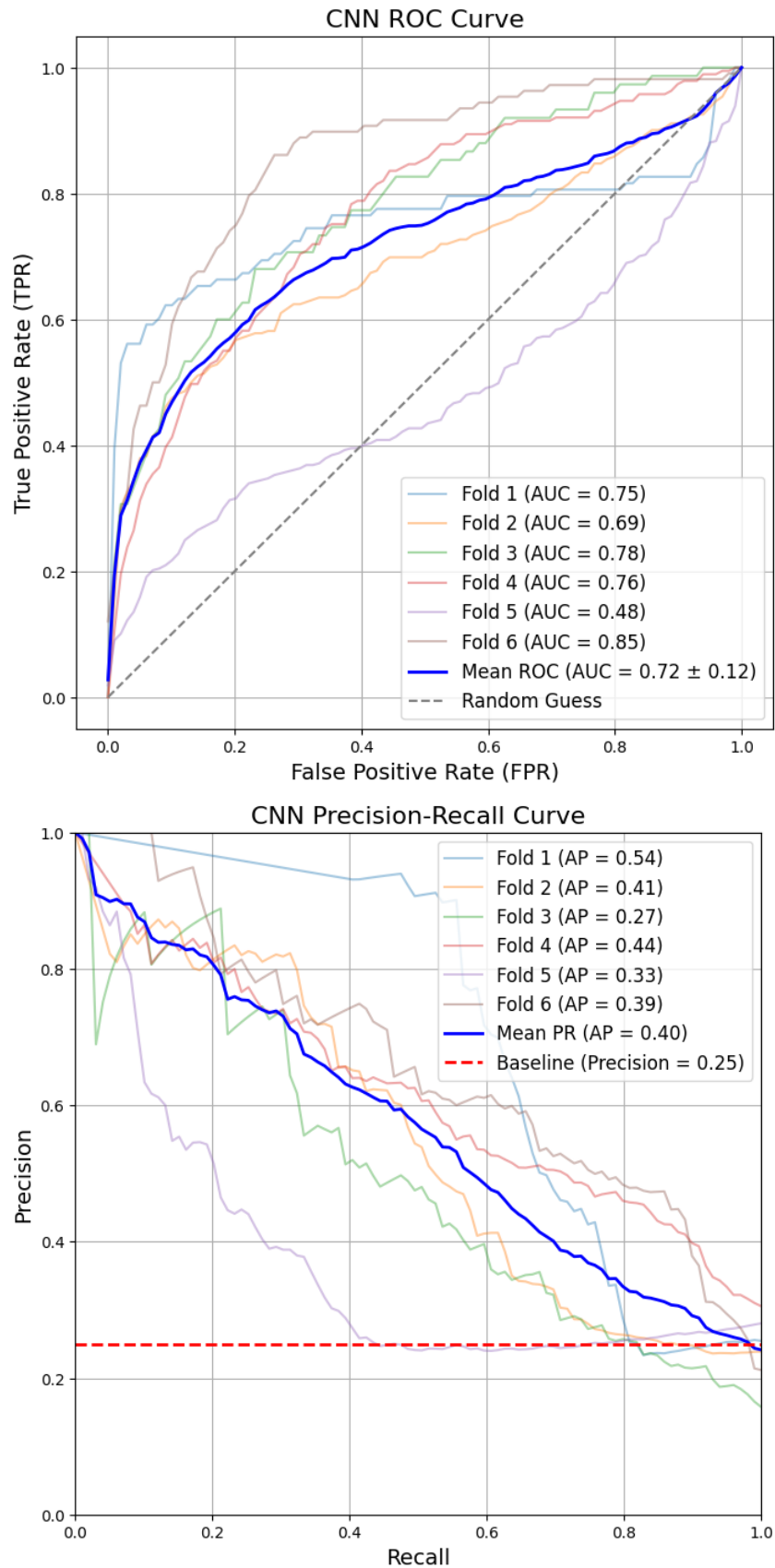
The complete metric evolution curves over epochs, illustrating training instabilities and fold differences, are available in Appendix E, page 236.

Metric	Train set	Validation set
Accuracy	0.89	0.81
Precision	0.96	0.74
Recall	0.58	0.32
F1 Score	0.72	0.42
ROC AUC	NC	0.72
Average Precision	NC	0.40

✂ Table D.4 : Evaluation of the CNN model. Each value represents the average across the six folds of the cross-validation. ROC AUC and Average Precision were not computed for the train set.



✂ Figure D.17 : Confusion matrix for the CNN model at a decision threshold of 0.5.



⌘ Figure D.18 : ROC curve and PR curve for the CNN model.

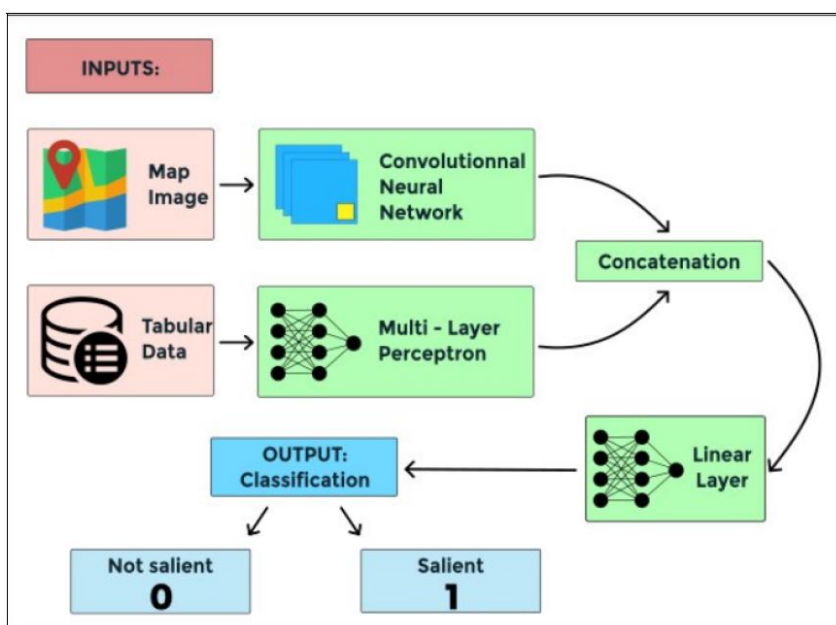
D.5.3.4 Multimodal

❖ Architecture and hyperparameters

A straightforward approach to do multimodal machine learning, and the one we chose, is to combine our two previous architectures by concatenating the final perceptron layer of each model and use a fully connected dense layers to interpret the combination (Figure D.19).

Our implementation is available online⁶. A detailed summary of the model architecture is also provided in Appendix 22, page 241.

Just as with our CNN model, we trained the multimodal model for 150 epochs without early stopping, using a learning rate that progressively decreases over time. To ensure stability, we conducted two separate training runs with different random initializations and obtained consistent results.



⌘ Figure D.19 : Multimodal architecture for remarkable building detection, inspired by the framework proposed in Rosenfelder 2020. Source: Eric Lafon.

❖ Results

The training results for the multimodal model are presented in Table D.5, with its confusion matrix shown in Figure D.20.

When analyzing the F1 score and AUC-ROC, the multimodal model performed worse than the random forest model, but outperformed the CNN model. The ROC and PR curves (Figure D.21) highlight significant variability between folds, mirroring the behavior observed in the CNN model.

Notably, folds 3 and 5 exhibit the same trends as in the CNN model:

- Fold 3 achieved a high AUC-ROC score but had a low average precision.

⁶<https://github.com/TinQk/LostInZoom>

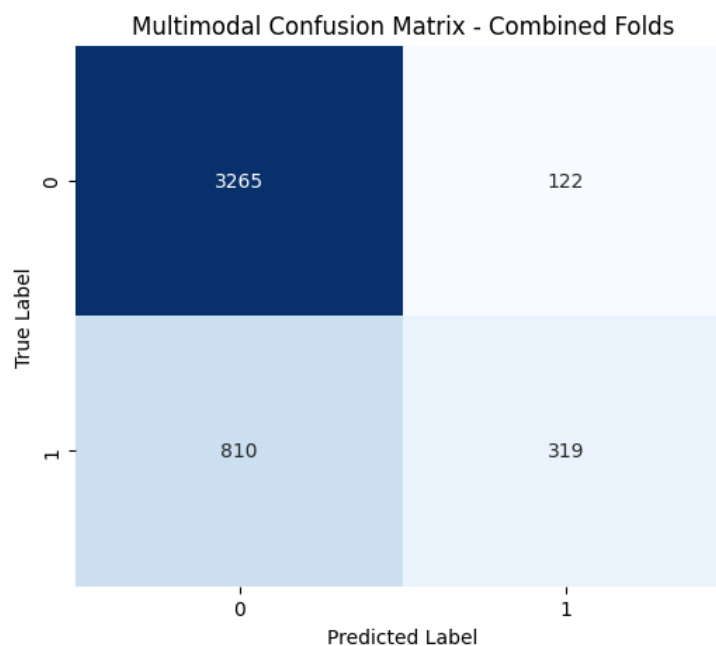
- Fold 5 performed poorly across all metrics.

However, this time, fold 1 yielded strong results, which was not observed in the CNN model.

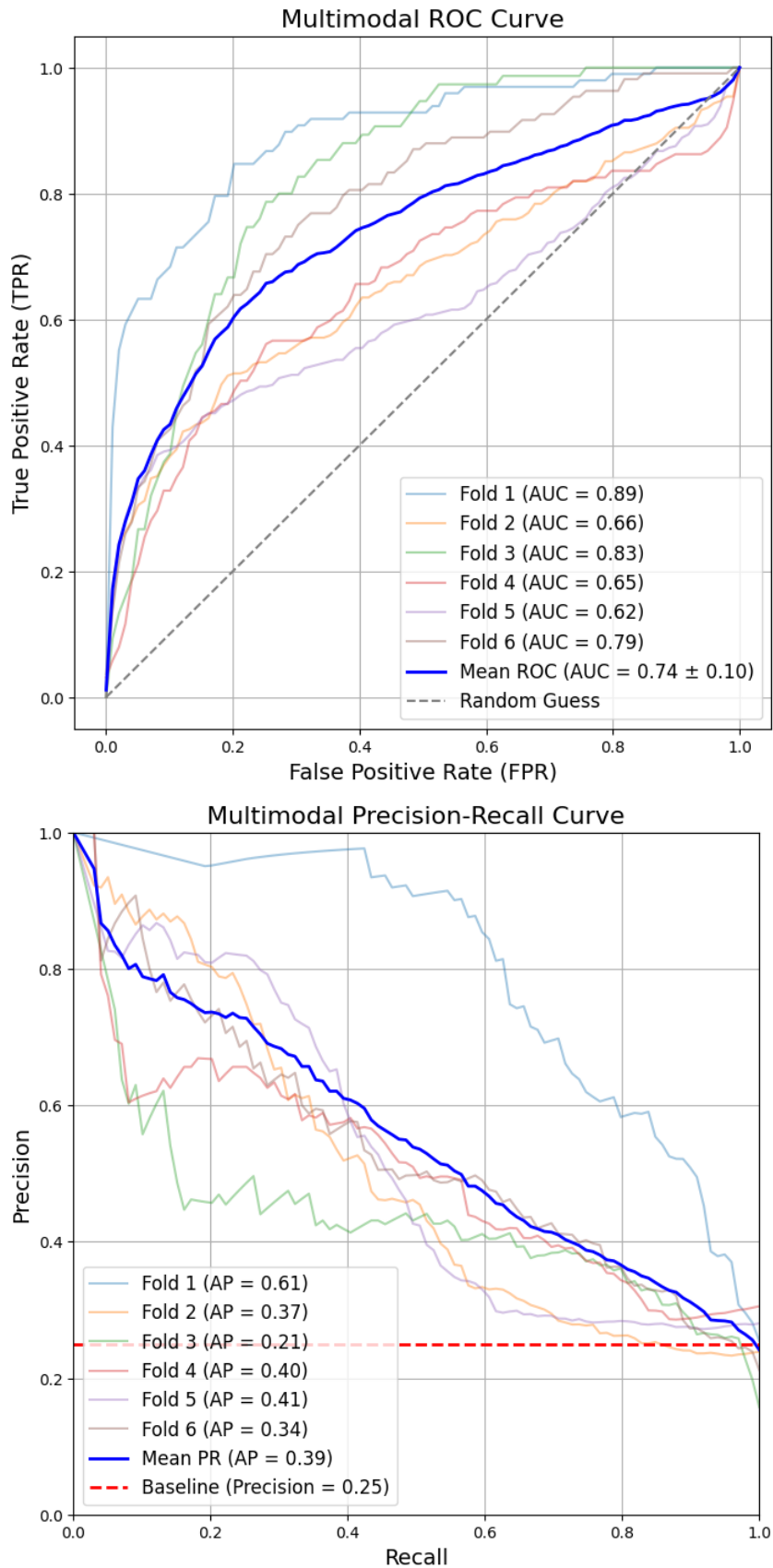
The complete set of metric evolution curves over epochs, illustrating training instabilities and differences between folds, is available in Appendix F, page 241.

Metric	Train set	Validation set
Accuracy	0.91	0.80
Precision	0.95	0.68
Recall	0.69	0.32
F1 Score	0.90	0.42
ROC AUC	NC	0.74
Average Precision	NC	0.39

✿ Table D.5 : Evaluation of the multimodal model. Each value represents the average across the six folds of the cross-validation. ROC AUC and Average Precision were not computed for the train set.



✿ Figure D.20 : Confusion matrix for the multimodal model at a decision threshold of 0.5.



✿ Figure D.21 : ROC curve and PR curve for the multimodal model.

D.5.4 Discussion

D.5.4.1 Model comparison

❖ Performances

As shown in Table D.6, the Random Forest model trained solely on tabular data achieved the highest overall performance across all folds. Our multimodal approach has therefore not succeeded in consistently improving the classification results, even though the CNN model trained exclusively on map images produced non-negligible results.

The strong performance of the CNN and multimodal models on certain folds (see Figure D.18 and Figure D.21) suggests, however, that raster-based map image data has the potential to enhance classification models in specific contexts. Nonetheless, our results do not provide conclusive evidence to confirm this hypothesis.

Model	F1 Score	ROC AUC	Av. Pre.	Training time (minutes)
Random Forest	0.51	0.78	0.42	0.36
MLP	0.45	0.75	0.40	2.35
CNN	0.42	0.72	0.40	109
Multimodal	0.42	0.74	0.39	112

✿ Table D.6 : Summary table of the metrics for each model. The results are based on the average of the results for each fold.

❖ Computational cost considerations

One crucial aspect we have not yet discussed is computational time and its associated energy consumption, which is roughly proportional. While we do not delve into the detailed calculations of energy usage and carbon emissions as explored by Henderson, Hu, Romoff, Brunskill, Jurafsky, and Pineau 2022, Table D.6 shows that training our models using image data requires between 50 and 300 times more computational time than training models based solely on tabular data.

This discrepancy is largely due to the structural complexity of the models and the number of trainable parameters. Our MLP model, for instance, contains approximately 2,000 parameters (which could have been arguably slightly increased), whereas our CNN model has over 16 million parameters. This substantial difference underscores the significant computational cost of incorporating image data, which must be weighed against potential performance gains in future studies.

D.5.4.2 Limitations and potential improvements

Multimodal machine learning is still an emerging field, and it is clear that both our implementation and dataset are far from optimal. While there is always room for improving the model architecture, hyperparameters, or input data to achieve better results, our primary objective was to experiment with a straightforward implementation before considering more complex approaches.

To further refine our approach while keeping the same dataset, a crucial next

step would be to analyze the classification results in detail—examining misclassified buildings and understanding the key differences between model predictions. In particular, it would be helpful to investigate why the CNN and multimodal models performed inconsistently across different folds.

Additionally, a major improvement would be to annotate remarkable buildings in a different city to create a dedicated test set free from any data leakage, rather than relying solely on spatial cross-validation. This would enable us to evaluate model generalizability more effectively. An ensemble learning approach could then be tested, where multiple models trained on different folds are combined to make final predictions on the test set.

Beyond that, several other refinements and directions could be explored:

- **Assessing annotation consistency:** Having another person independently annotate the same Nantes dataset would provide insights into the consistency and generalizability of the remarkable building classification task. Comparing these different annotations could help quantify subjectivity in the labelling process. Annotation subjectivity is discussed in Chapter E for another project.
- **Adding new tabular features:** While we enriched our tabular data with several new features, many others have been explored in the literature for landmark detection, such as the distance to the nearest crossroads or the building’s orientation relative to the closest road (Elias 2003).
- **Merging buildings:** As anticipated, building complexes raised challenges both for annotation and classification. Should all buildings in a complex be labelled as remarkable, or only the most distinctive ones? During annotation, we decided to classify as remarkable all adjacent buildings that contribute to a single large remarkable structure. However, our models struggled with these cases (see Figure D.22). A possible solution could be to merge adjacent buildings into single entities to avoid classifying fragmented building parts separately. However, this process would need to be done carefully to avoid misclassifying large complexes where only a specific section is truly remarkable.
- **Enhancing buildings in image inputs:** One simple modification to try improving CNN performance could be highlighting the target building in the extracted map images. For instance, coloring the building in black while keeping the surrounding environment unchanged could maybe help the CNN distinguish it more clearly.
- **Exploring different scales for extracted images:** A smaller scale (larger geographic area per image) could provide more contextual information around each building, potentially improving classification performance. However, this comes at the cost of lower image resolution, meaning some fine details may no longer be visible. Additionally, larger image sizes increase computational costs during model training, so an optimal balance must be found.
- **Exploring alternative image sources:** Instead of using images extracted from the IGN Plan map, satellite imagery from Google Earth or street-level photos (e.g., Google Street View) might be more informative, especially since Google Earth was the primary tool used during annotation. Previous research,

such as Hou and Biljecki 2022, suggests this could be a promising avenue for further exploration.

- **Using a custom map for image extraction:** Rather than relying on existing maps (such as Google Maps or IGN Plan), a dedicated map could be created specifically for training the models. This would remove biases introduced by the specific design choices of commercial maps. However, any custom map would introduce its own biases, so the trade-offs need to be carefully considered.



✿ Figure D.22 : Large industrial buildings and agricultural greenhouses in the third sub-zone of our dataset. Left: Google Earth image. Right: OpenStreetMap base map with buildings from our vector dataset. Blue buildings were classified as remarkable by the Random Forest model, while yellow buildings were classified as non-remarkable.

D.6 Conclusion of the chapter

In this chapter, we explored the potential of multimodal machine learning using map images for a complex classification task, specifically testing it on remarkable building detection as an opportunity arose. To facilitate this research, we created an annotated dataset of approximately one thousand remarkable buildings, which we have made publicly available online.

We first tested a machine learning approach using two different modalities. The first focused on training models solely with vector-based tabular building data, utilizing both a Random Forest model and an MLP, which are well-established techniques. The second, more experimental approach, involved training a CNN on map images centered on individual buildings. We then combined these two modalities into a simple multimodal architecture to assess whether integrating spatial context from map images could enhance classification performance.

The results of the multimodal model were inconclusive. On average, it underperformed compared to the Random Forest model, despite having access to additional

information through map images during both training and prediction. However, we observed significant variations in performance across the different validation zones. In certain folds, the multimodal model outperformed the Random Forest, suggesting that multimodal machine learning leveraging map images might hold potential in specific, yet-to-be-defined contexts.

Unsurprisingly, training our machine learning models on images required significantly more computational resources than training them on tabular data. Moreover, extracting and storing map images further increases energy consumption. In a context where the ecological impact and energy cost of AI model training have become major global concerns, Schwartz, Dodge, Smith, and Etzioni 2020 advocates for a more responsible and efficient use of machine learning, termed Green AI. We therefore believe that incorporating image data to enhance a model initially designed for tabular data should be reserved for strategic projects where the added value justifies the additional computational cost.

Regarding remarkable building detection specifically, our findings suggest that image-based approaches may not be the most practical path forward. Instead, we propose breaking down the problem into more specific sub-tasks, such as detecting tall buildings separately from isolated buildings. We believe that a well-enriched dataset, combined with simple algorithms like decision trees, could effectively address these sub-tasks, with targeted support from machine learning on tabular data when necessary.

Part III

Towards the Detection of Areas of High Anchoring Potential in Map Images

Chapter E

AHAPs – Concept and preliminary experiments

In this third part, we continue exploring the elements of maps that serve as anchors in users' mental representations. However, we adopt a different approach from the previous part. Instead of focusing on specific landmarks and attempting to detect them within a vector database, we shift our attention to map user perception by analysing map images directly. Our goal is to determine which areas, in map images, are most frequently used for orientation.

In this chapter, we introduce the concept of "Areas of High Anchoring Potential" (AHAP) and discuss the relevance of detecting such areas. We then propose a drawing-based annotation method and present an experiment designed to assess the consistency of these annotations across different individuals.

E.1 Introduction to AHAPs

E.1.1 Definition

We define **Areas of High Anchoring Potential** (AHAPs) as regions within a pan-scalar map that have the potential to serve as reference points when a user navigates through space and across different zoom levels. These areas may contain one or more elements that can themselves be classified as pan-scalar landmarks, as defined in Chapter A.

We believe that using these anchor areas as an intermediary step for detecting pan-scale landmarks is valuable because they allow different types of landmarks, which interact in the user's mental map, to be grouped together through a simple annotation such as highlighting.

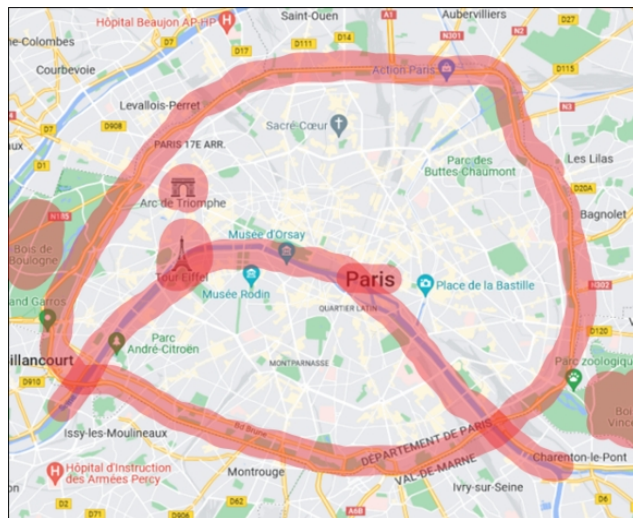
For example, Figure E.1 shows AHAPs highlighted on a map of Paris. These areas encompass the primary landmarks visible at this scale, including the Seine River, the "Paris" toponym, two symbolized monuments, the ring road, and the two large parks on the city's periphery. In this image, the Bois de Boulogne and Bois de Vincennes are considered landmarks not only because their toponyms are

prominently displayed, or because they appear as large green areas, or because they frame the ring road, but because these factors combine and interact with additional elements.

Another advantage of AHAPs is their ability to define a clear beginning and end for linear landmarks such as the Seine, which is not straightforward in a vector-based approach.

❖ **Note:**

We may sometimes use the term "anchors in maps" instead of AHAPs in this manuscript for simplicity. However, this should not be confused with the anchor points discussed in Section A.1.2.2, which exist solely in the human mind.



❖ Figure E.1 : A Google Maps image of Paris with manually highlighted Areas of High Anchoring Potential in red.

E.1.2 Hypothesis

The central hypothesis we aim to investigate in this third part of the manuscript is:

Hypothesis: It is possible to train a machine learning model to predict Areas of High Anchoring Potential in map images.

Our approach relies on leveraging the implicit anchoring properties of landmarks in map images—primarily their salience, spatial structure, and interactions with other landmarks—to train a model capable of recognizing a broad range of landmarks rather than only specific predefined ones.

Ideally, the model should be able to:

- Identify landmark types not explicitly included in the training dataset, learning a more abstract representation of anchoring potential.

- Generalize across different map styles, ensuring robustness across varying cartographic designs.
- Generalize across scales.

E.1.3 Motivations

The study and, more specifically, the automatic detection of Areas of High Anchoring Potential (AHAPs) could serve several purposes:

❖ **Enhancing landmark detection through image segmentation and data Matching**

Analyzing map images offers a more holistic view of spatial relationships and brings us closer to user perception, as it captures the visual choices made during cartographic design. Unlike vector data, images reflect how users actually see and interpret maps. As shown in Section D.4, combining image segmentation with data matching could improve landmark recognition by leveraging both visual and structured geographic information.

❖ **Detecting "Critical Zones" in pan-scalar maps**

Some locations on a pan-scalar map may temporarily lack visible reference points due to the zoom level, leading to potential disorientation for users (see Section E.1.4.2). By identifying critical zones, we could ensure that key reference areas remain visible at appropriate scales by extending their scale range, enhancing user navigation and orientation.

❖ **Identifying problematic areas for correction via change detection techniques**

As discussed in Chapter A, inconsistencies between scales are a key challenge in pan-scalar maps. Yu, Ai, Xu, Yan, and Shen 2023 propose a method to detect inconsistencies in feature representation across scales by working directly on raster data, thus avoiding complex vector data management. Similarly, identifying sudden disappearances or appearances of anchoring zones, for instance, due to inconsistencies in data sources or unintended generalization effects—could assist cartographers in performing targeted manual corrections, ultimately reducing desert fog effects and improving map coherence.

❖ **Scientific Curiosity**

From a broader scientific perspective, recent advances in image segmentation raise an intriguing question: Can modern segmentation models learn to predict Areas of High Anchoring Potential in a way that aligns with human intuition and collective agreement? This research direction not only explores AI's capacity to interpret and replicate human map-reading behavior but also contributes to our understanding of how users rely on visual cues in pan-scalar maps.

E.1.4 Related concepts

E.1.4.1 Visual saliency

The concept we propose—Areas of High Anchoring Potential (AHAPs)—is closely related to **visual saliency**, which refers to key areas in an image that the brain selects in a pre-attentive manner for prioritized analysis before processing the rest of the scene.

Visual saliency has been extensively studied using eye-tracking techniques, as well as drawing-based methods (Borji, Sihite, and Itti 2013). These studies typically rely on photographic datasets, but within the LIZ project, Wenclik and Touya 2024 have tested saliency models from the literature on a dataset composed of map images and eye-tracking data (see Figure E.2).



⌘ Figure E.2 : A map image from OSM representing the city of Châteaubriant and its ring road (left), and the derived saliency map using the Itti-Koch saliency model (right).

Numerous deep learning models have been developed to detect salient regions, currently achieving state-of-the-art results (Kümmerer, Theis, and Bethge 2014; Cornia, Baraldi, Serra, and Cucchiara 2018; Pan, Ferrer, McGuinness, O’Connor, Torres, Sayrol, and Giro-i-Nieto 2018). More recently, Tliba, Kerkouri, Chetouani, Bruno, El Hassouni, and Çöltekin 2024 proposed a self-attention-based saliency model.

These saliency models have already demonstrated their ability to identify unique elements within an image, whether composed of abstract geometric figures, photographs or maps. This provides encouraging evidence for the potential of deep learning models to perform segmentation of AHAPs. However, AHAPs differ from traditional saliency zones in several ways:

- AHAPs are not exclusively pre-attentive; they can be consciously identified after longer deliberation and reflection by the user.
- AHAPs do not necessarily attract direct visual attention. A study within the LIZ project (currently under review) suggests that peripheral vision may be

sufficient for detecting anchors.

- AHAPs are specific to pan-scalar maps, whereas visual saliency applies to all images.

When examining the saliency heatmap in Figure E.2, we can see that all roads are highlighted to a similar degree. However, we believe that a model specialized in detecting AHAPs should emphasize the semicircular ring road surrounding the city center. The city center itself, along with the toponym "Châteaubriant," should also stand out more prominently. The most salient elements in the heatmap are the lakes, but it is unclear whether these would truly be considered important landmarks at this map scale. Additionally, what appears to be a park south of the city center—but within the ring road—seems to have more potential as an anchor than the other green areas outside the ring. These examples highlight the potential differences between a general saliency model and one specifically trained to detect AHAPs. We therefore believe there is a strong case for investigating AHAP detection alongside visual saliency.

E.1.4.2 Critical zones

The AHAP concept also aligns with the notion of critical zones, introduced by Jul and Furnas 1998 in their discussion on desert fog. **Critical zones** refer to areas on a pan-scalar map that contain important or interesting information at large scales, which may disappear when zooming out, leaving no indication that this information exists. To mitigate desert fog, they proposed displaying navigational aids that outline the extent of critical zones in otherwise empty views. This way, when users view a map at a scale where no objects are typically visible, they can still see rectangles marking areas that contain information, ensuring they will find relevant features by zooming in. This approach has directly influenced the LostInZoom (LIZ) project and our broader efforts to detect and highlight landmarks within pan-scalar maps.

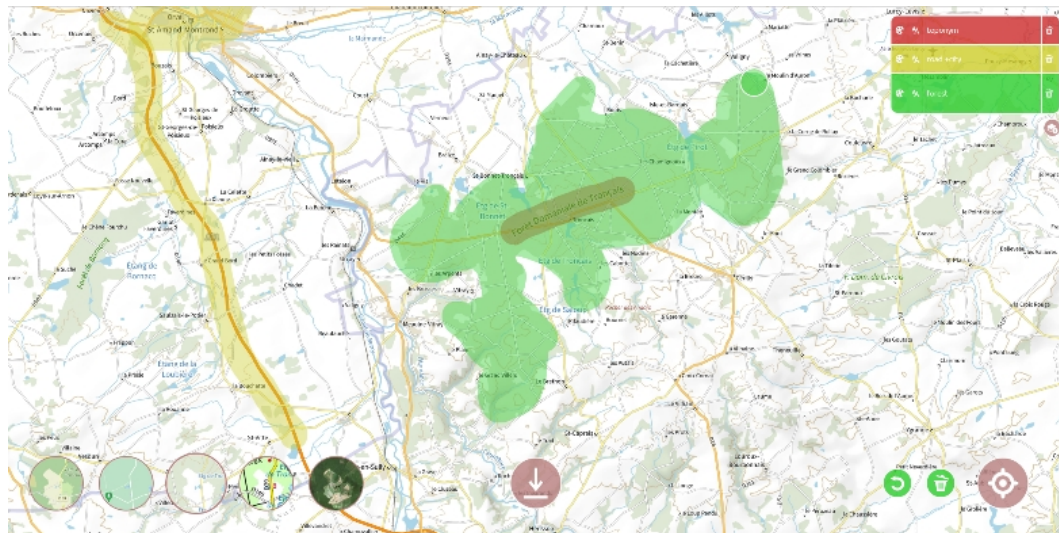
E.1.5 Drawing as data annotation

A dedicated tool for drawing on maps, called MapDraw, was developed early in the LIZ project (Berli, Touya, Gruget, and Potié 2024) (in French) and is available online¹. This web-based tool allows users to navigate within a pan-scalar map, zooming in or out to a desired map view, and then freely highlight specific areas using a drawing tool (Figure E.3). The drawn annotations are then stored as polygons in a PostgreSQL database along with some attributes and metadata, or exported as a GeoJSON file.

We chose to use this map-based drawing tool to collect data on Areas of High Anchoring Potential. Prior research on visual saliency has demonstrated that when participants highlight important map features, their selections strongly correlate with eye-tracking results (Borji, Sihite, and Itti 2013). Since our goal is to detect zones similar to saliency regions, we believe that drawing-based annotation is also a suitable approach.

¹<https://lostinzoom.huma-num.fr/mapdraw/>

Additionally, drawing experiments require less specialized hardware than eye-tracking and can be conducted remotely, making data collection more scalable. Lastly, compared to selecting predefined vector features, freehand drawing (or highlighting) provides greater flexibility, allowing participants to outline not only entire landmarks but also specific parts of them or even groups of features that form meaningful visual anchors.



✿ Figure E.3 : A screenshot from the MapDraw web application. At the bottom left, circular icons allow users to select different map backgrounds, while the menu at the top right enables the creation of layers for drawing, similar to a digital illustration tool.

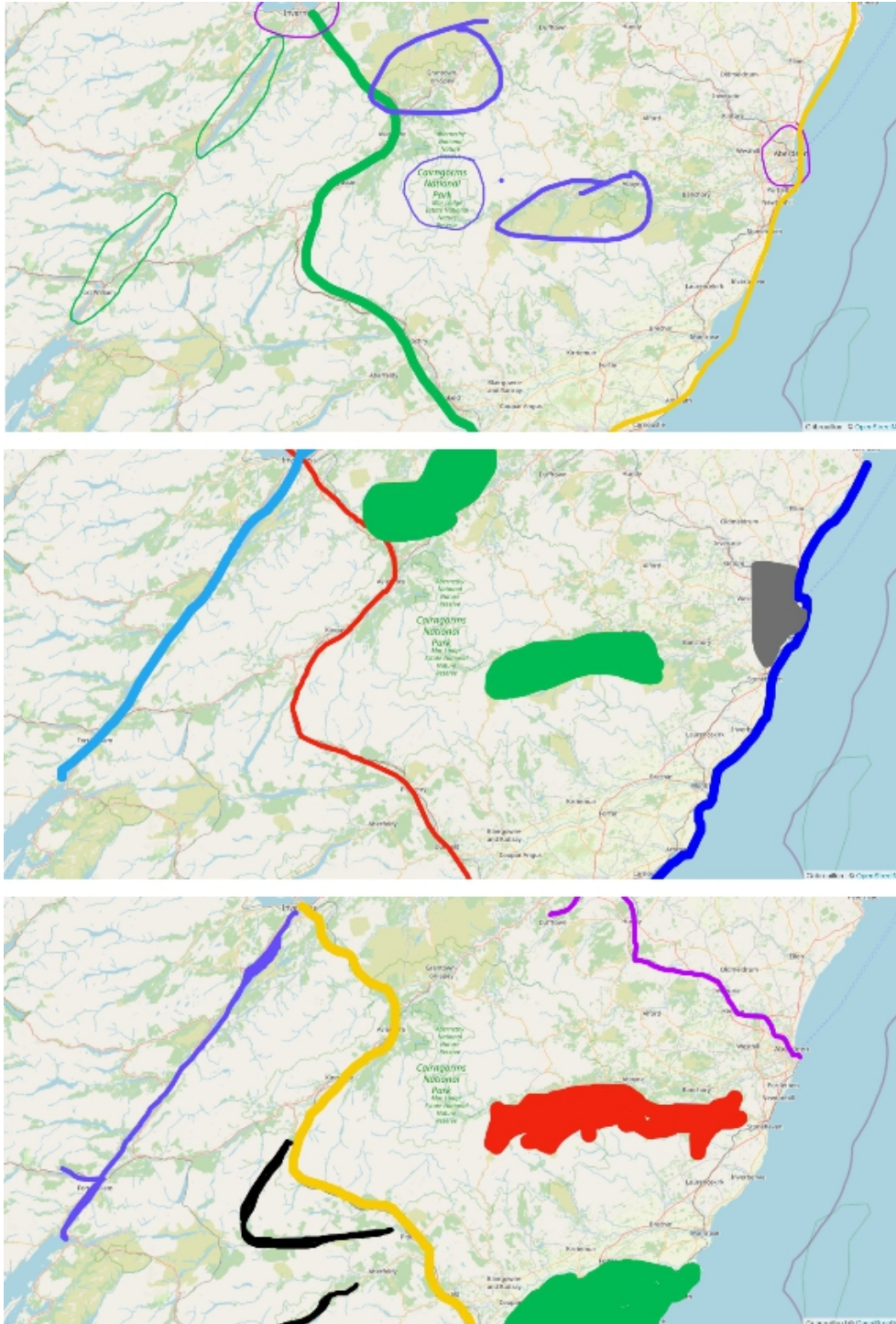
E.1.6 Preliminary experiment with Gribrouillon

Before conducting a larger-scale study, we aimed to verify the consistency of responses among participants given the same instructions and to gather initial feedback on using a coloring tool for annotating maps.

To do this, we carried out a preliminary experiment with members of the LIZ project while MapDraw was still under development. In this experiment, we presented three static OSM maps to eight participants using the online tool Gribrouillon, which allows users to color different areas on a map. Participants were given the following instruction:

- "Highlight in color the elements of the map that you find memorable and that could serve as reference points during a pan-scalar exploration with zooming in and out."

The results show a certain level of agreement in the highlighted areas between participants. As seen in Figure E.4, while each drawing contains some unique features, there are also several common elements across participants. The full set of 24



☼ Figure E.4 : Three annotations of the same area from the Gribrouillon experiment. In the first image, the participant outlined the elements they considered important, while in the other two, participants filled them with color. While this was not an issue for this preliminary experiment, ensuring consistent annotation styles will be crucial for creating a high-quality training dataset.

drawings is available online along with the rest of the project data².

Interestingly, some participants highlighted not individual features but the boundaries between two elements, such as the edge of a forest or the coastline. This raises an important question: Should boundaries between two large elements be considered anchors in their own right when training detection models? Should we ignore such annotations? Or should we associate them with the most relevant feature they separate?

To address this, we chose to make a distinction based on feature size on the screen. In future experiments, participants would be asked to color features directly if they could fit entirely within the screen, while highlighting boundaries was allowed only when marking the separation of very large areas.

E.2 Preparing an evaluation dataset

E.2.1 The challenge of evaluation

To assess the performance of our future models, we aimed to create the most precise evaluation dataset possible, with a well-defined ground truth. However, defining a ground truth is inherently difficult, as anchors—by our definition—are subjective and perception-based. There is no single "correct" answer to what qualifies as an anchor in a map.

To address this issue, we created a set of map images and asked a group of fifteen people to highlight the areas they considered to be Areas of High Anchoring Potential (AHAPs). For this task, we selected individuals with GIS knowledge who were either already familiar with the concept of anchors or were introduced to it beforehand. Our aim doing so was to obtain more reliable and coherent annotations. This approach served two main objectives:

1. Establishing a ground truth: By combining multiple annotations, we aimed to create a reference dataset that reduced individual biases and allowed us to evaluate the performance of our models.
2. Measuring human agreement: By analyzing the similarities between different participants' annotations, we could determine the level of consensus among individuals. This, in turn, would provide a performance benchmark for our models—if a model's predictions matched the ground truth as closely as human annotations did, we could consider the model's performance satisfactory.

E.2.2 Selecting geographic areas for the dataset

Ideally, we wanted to ensure that our models would work across different topographies, various types of anchors (composed of different objects), multiple scales, and different map styles (i.e., varying cartographic symbology).

²<https://zenodo.org/records/15238773>

Concerning the map style, we wanted a unique map style to ensure a more robust validation, and we chose Google Maps as it is the most widely used pan-scalar map interface. Additionally, IGN's data is limited to France, meaning we could not collect more annotations worldwide if ever we wanted.

To introduce geographical diversity, we selected areas in southwestern France, within the Occitanie and Nouvelle-Aquitaine regions. These regions provided a variety of landscapes, including:

- Cities of different sizes
- Plains and mountainous areas
- Forested regions
- Both the Atlantic Ocean and the Mediterranean Sea

In total, we selected 20 areas ranging from zoom level 8 to zoom level 18. We excluded zoom levels below 8, as country-level shapes are already recognizable at that scale, and we did not include levels beyond 18, as all details are already visible at that point.

Images of all 20 areas are available in Appendix G, along with some masks which we will refer to in this chapter and the following.

The geographic coordinates of the selected areas are available in the protocol provided to participants (included in the Zenodo repository). Additionally, blank versions of the selected images (without annotations) are stored in the "base_images" folder.

E.2.3 Annotation protocol

After being introduced to the concept of Areas of High Anchoring Potential and trained on how to use the MapDraw tool, the fifteen participants were given the following instructions:

❖ **Drawing Phase (for each provided map area):**

1. Observe the map area for a few seconds.
2. Highlight elements that could serve as pan-scalar anchors, according to your perception.
3. Multiple elements could be highlighted within the same layer.
4. When highlighting a different type of anchor, create a new layer.
5. Assign each layer to one of five predefined categories.

The reasoning behind proposing to classify different anchor categories was to later refine our model evaluation accordingly. The five AHAP categories participants could assign their annotations to were:

- Artificial landmarks (Symbols, toponyms, administrative borders...)

- Transportation networks
- Cities, buildings, parks...
- Water bodies (Lakes, rivers, coastlines...)
- Forests, mountains, and other natural landmarks

❖ **Additional Guidelines:**

- Participants were advised to spend approximately 3 minutes per area. This time constraint was designed to prevent annotators from examining the area in too much detail in order to simulate a more natural navigation experience, while still giving them enough time to use the coloring tool to highlight relevant zones.
- Colors would not be considered in the analysis. This decision was taken to both simplify the annotation process and avoid biasing participants with the default color order suggested by MapDraw. Annotators were therefore free to use any color for any type of anchor.
- Overlapping drawings from different layers were allowed.
- All drawings within the same layer would be merged. For example, if two roads intersecting at a crossroad were annotated with two separate strokes on the same layer, they would be treated as a single unified region in the collected data.

An annotation example is available in Figure E.5.

E.2.4 Output and first analysis

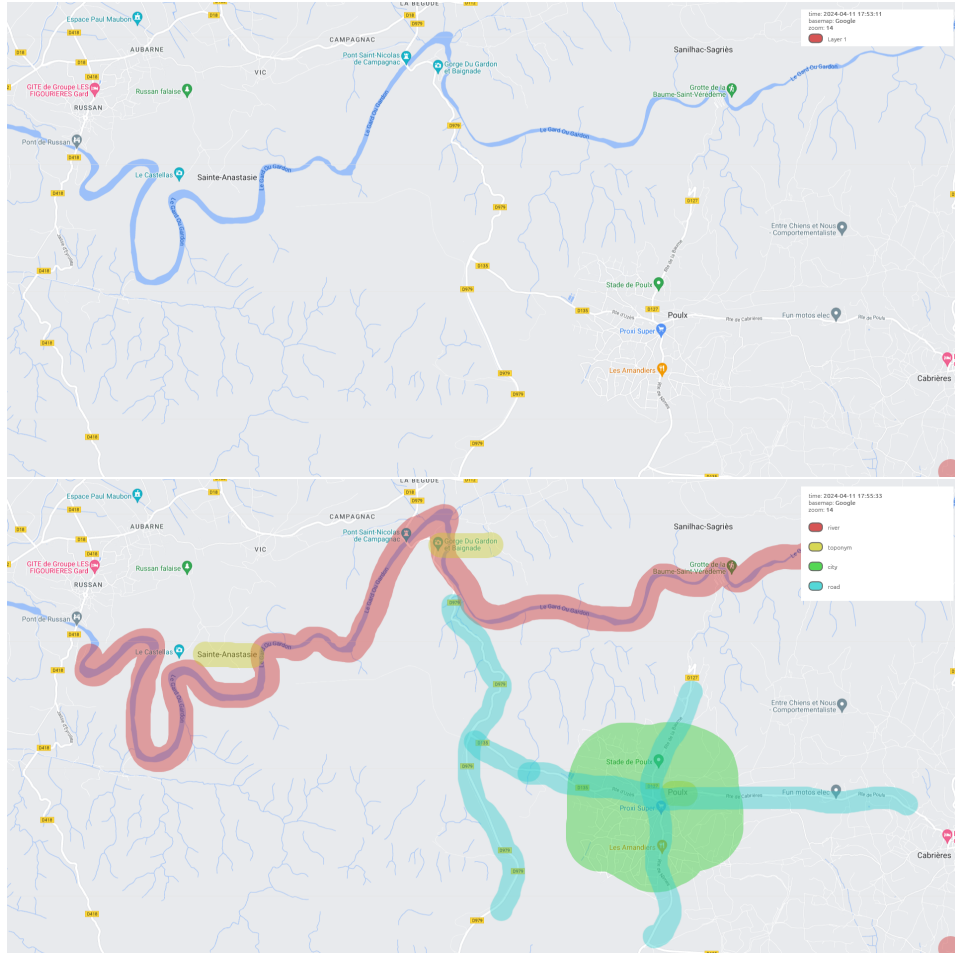
For each zone where a participant drew and for each category of anchor, a black-and-white image representing the highlighted areas was generated, which we refer to as a mask (Figure E.6). All the base map images and the masks created by the participants are available online³.

We counted the number of pixels drawn for each category. As shown in Table E.1, there is a significant difference in the number of pixels drawn across categories, with the surface area annotated for transportation networks being ten times larger than that for artificial landmarks (which mainly consist of toponyms and symbols).

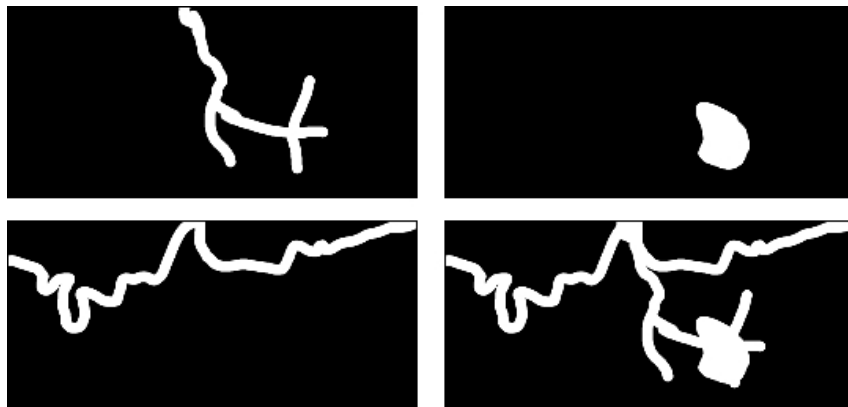
Even though we aimed for a balanced evaluation dataset, this difference is naturally explained by the varying sizes of elements in the maps. For the remainder of our study, we primarily focus on the "complete" masks, obtained by merging all anchor categories, to simplify the problem. However, this discrepancy between categories should be kept in mind for future evaluations, particularly if we wish to assess model performance separately for each category.

Moreover, while toponyms and pictograms are relatively small in size, administrative borders (which we included in our "artificial" category) are linear features

³<https://zenodo.org/records/15238773>



✿ Figure E.5 : Example of annotation using MapDraw on zone 7. Top: Original image. Bottom: The same image with a city, a river, two toponyms, and several roads highlighted.



✿ Figure E.6 : Masks created from a participant's annotation on the same zone as Figure E.5. Top left: Mask for the "road" annotation. Top right: Mask for the "city" annotation. Bottom left: Mask for the "water" annotation. Bottom right: Mask obtained by merging all anchor categories.

Anchor cat.	Pixels drawn	ratio
Artificial	3.1M	4%
Transport	30.8M	39%
Urban	15.8M	20%
Water	20.8M	26%
Nature	8.1M	10%

✿ Table E.1 : Distribution of the anchor categories based on the number of pixels drawn in the annotations.

that occupy a significantly larger portion of the annotations in terms of pixel count. Toponyms—which are known to be crucial landmarks in maps—therefore occupy even less than 4% of the total drawn pixels. This imbalance could impact both our dataset analysis and model training, potentially underrepresenting the importance of toponyms compared to other artificial anchors.

Each participant annotated all 20 map images, but only highlighted a portion of each image. For every participant, we computed the average percentage of the image area they annotated across the 20 maps. The variation in annotation coverage between participants is shown in Figure E.7. A nearly fourfold difference was observed between the participant who highlighted the most pixels and the one who highlighted the least. However, most annotations were concentrated around 15% of the total image area.

A similar analysis was then performed, this time by aggregating the data from all participants to compare the highlighting across different map areas (Figure E.8). The results indicate that for the vast majority of zones, the positively annotated area ranged between 10% and 15% of the image. The only exception was Zone 19, where a significantly larger surface was highlighted. A closer inspection revealed that this was due to multiple participants highlighting the central park in its entirety (Figure E.9).

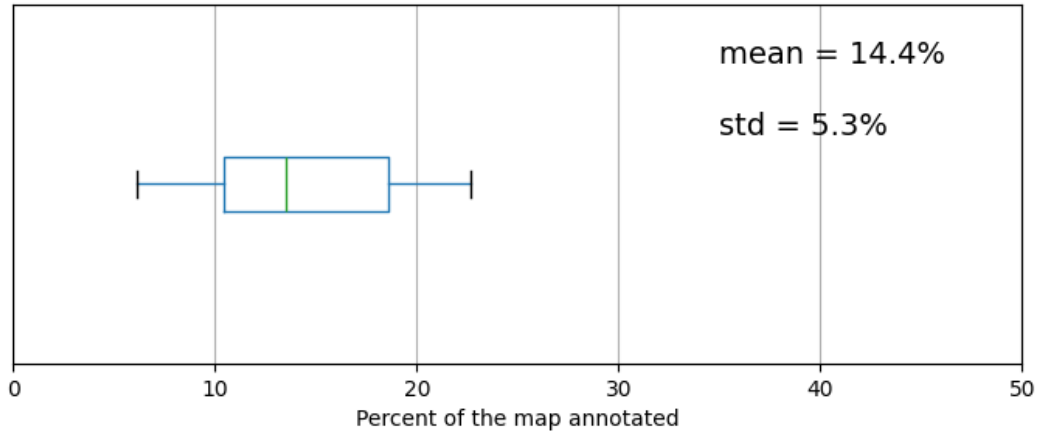
These initial findings suggest that the ratio of annotated surface to image size remains relatively stable for AHAPs. This consistency could enable segmentation models (or post-processing steps) to be calibrated in a way that ideally produces segmented areas of similar proportions, for example by adjusting their classification threshold.

E.2.5 Establishing a common truth with STAPLE

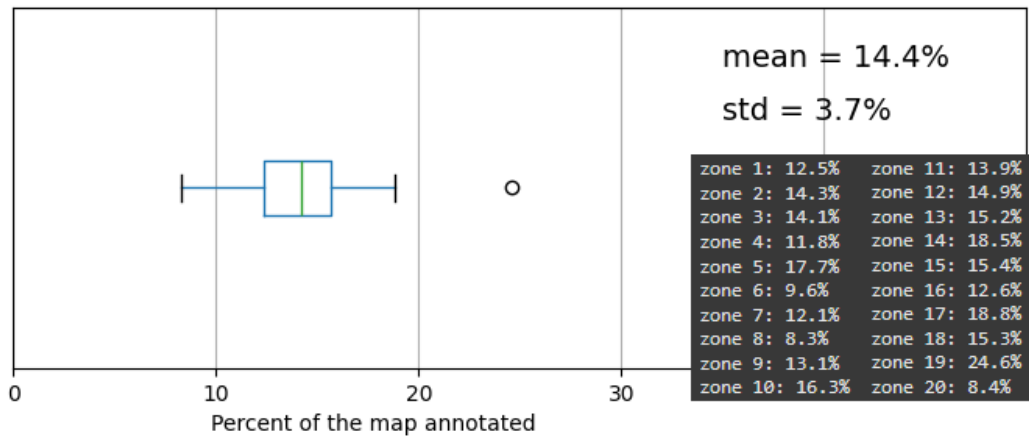
E.2.5.1 Definition

STAPLE (Simultaneous Truth and Performance Level Estimation) is an algorithm proposed by Warfield, Zou, and Wells 2004 that computes a probabilistic estimate of the true segmentation and a performance measure for each segmentation within a set of annotations, whether these annotations come from human experts or automated algorithms.

Originally developed for medical applications, this algorithm is commonly used to merge annotations from doctors on MRI (Magnetic Resonance Imaging) and CT



✿ Figure E.7 : Variation in annotated coverage between participants.



✿ Figure E.8 : Variation in the percentage of the image highlighted across different zones.



✿ Figure E.9 : Zone 19, where a large central green area representing a park was highlighted by many participants.

(Computed Tomography) scans into a single “ground truth” that integrates all provided annotations.

STAPLE is particularly well-suited to our task of detecting anchoring areas, as it is specifically designed to establish a consensus on the presence or absence of structures in an image annotated by multiple experts, which aligns precisely with our objective.

Our models could then be considered to have good performance if their predictions are no further from the "consensus truth" (as determined by STAPLE) than the individual annotations provided by the experts who contributed to establishing that consensus.

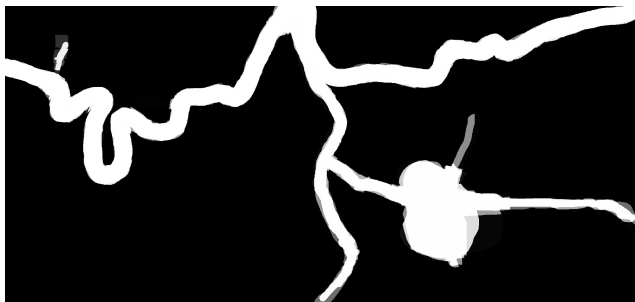
E.2.5.2 Results

To run the STAPLE algorithm, we used the SimpleITK library (Yaniv, Lowekamp, Johnson, and Beare 2018), which provides a built-in implementation of the algorithm. Additionally, the SimpleITK community offers useful notebooks for guidance⁴. Our code is available online⁵.

The STAPLE algorithm was applied to all selected zones, generating results for each type of anchor mask, as well as for the complete masks that merge all anchor categories. The resulting images are all available in Appendix G, top right of each figure, and online⁶.

As illustrated in Figure E.10, the ground truths created by STAPLE encompass most of the elements highlighted by annotators, provided that multiple annotators agree on their presence, but eliminate the least common annotations. This specific behavior must be considered when comparing STAPLE-derived ground truths to predictions made by our deep learning models.

As a last step for producing our common truth, since most segmentation evaluation metrics rely on binary data rather than continuous outputs like those produced by the STAPLE algorithm, we binarized the output by applying a simple 50% probability threshold.



⊛ Figure E.10 : Result of the STAPLE algorithm on zone number 7 (same zone than Figure E.5 and E.6), using annotations of all anchor types to establish a ground truth.

⁴<https://github.com/InsightSoftwareConsortium/SimpleITK-Notebooks?tab=readme-ov-file>

⁵<https://github.com/TinQk/LostInZoom>

⁶<https://zenodo.org/records/15238773>

E.3 Measuring inter-rater reliability

Inter-rater reliability refers to the degree of agreement among independent observers assessing the same phenomenon. In the context of our evaluation dataset, these independent observers are the experts who highlighted anchors in map images. Similarly, the agreement between our model's predictions on the evaluation dataset and the ground truth can be evaluated using the same approach. But how can inter-rater reliability be measured? And how can the similarity between two annotations be quantified?

Yang, Zamzmi, Angara, Rajaraman, Aquilina, Xue, Jaeger, Papagiannakis, and Antani 2023 propose assessing inter-rater reliability in medical image annotation by computing kappa statistics and then comparing individual expert annotations to a reference ground truth generated using the STAPLE algorithm with standard segmentation metrics. Since we already established similarities between anchor annotation in maps and medical image annotation, we adopt and discuss in this section a selection of these metrics, while excluding those that may not be well suited to anchor segmentation.

E.3.1 Indicators

❖ Not Retained: Simple Matching Coefficient (SMC)

The Simple Matching Coefficient (SMC) is calculated as the number of matching attributes divided by the total number of attributes. In the context of binary segmentation, the "matching attributes" correspond to the number of pixels where both sets agree on classification, while the "total attributes" refer to the total number of pixels in the image. However, in our dataset, anchors represent only a small fraction of the total pixels, with most being labeled as "background." This can result in very low SMC scores, making comparisons difficult. Therefore, we prefer using the Jaccard index, which better captures the similarity between segmentations in our experiments.

❖ Retained: Jaccard Index (Intersection Over Union, IoU)

The Jaccard index measures the proportion of overlap between the predicted mask and the ground truth relative to their union, and thus ranges from 0 to 1. Unlike SMC, the Jaccard index does not count "background" pixels as matches.

❖ Retained: F1 Score (Dice Similarity Coefficient, Sørensen Index)

The F1 Score is closely related to the Jaccard index. However, it focuses on the harmonic mean of precision and recall, making it slightly more forgiving than IoU regarding strict overlap requirements.

❖ Not Retained: Indicators Measuring Distance Between Surfaces

A common metric for segmentation evaluation is the Hausdorff distance, which measures the longest distance between any point in one subset and its closest point in another. While effective for comparing segmentations of a single object, it would not

be effective in our case as our dataset images include multiple landmarks. The same limitation applies to other surface distance-based metrics.

❖ **Not Retained: Volume Similarity Indicator**

The volume similarity indicator, which compares the number of pixels annotated as anchors in two images, is also unreliable in our case as the varying geometries of anchor types (symbols, lines, and surfaces) result in significant differences in their "pixel volume." Consequently, this indicator would be heavily biased toward anchors with larger volumes.

❖ **Retained: Cohen’s kappa and Fleiss’ Kappa**

Cohen’s kappa (Cohen 1960) is a statistic that measures inter-rater reliability between two raters while accounting for the possibility of agreement occurring by chance. It is defined as shown in Equation E.1, where p_o is the observed proportion of agreement between the two annotators, and p_e is the expected proportion of agreement by chance, calculated from the marginal probabilities of each annotator.

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{E.1}$$

Fleiss’ kappa extends this measure to account for any fixed number of raters, making it more suitable for our dataset, where multiple experts contribute annotations.

❖ **Future Considerations: Task-Specific Metrics**

Custom metrics could be developed to better evaluate the segmentation of AHAPs. For example, one possible approach is counting the number of detected anchors among the five most important anchors in a given zone. Additionally, a consensus threshold could be established for detecting an anchor, such as considering an anchor detected if at least half of its pixels, as determined by the consensus annotation, are identified. At this stage of the project, we primarily rely on visual inspection to assess such criteria.

E.3.2 Results and discussions

For each zone in the evaluation dataset, we computed Fleiss’ kappa to assess agreement among annotators. Additionally, we calculated the IoU and F1 scores between each individual annotation and the STAPLE-generated ground truth, then averaged these scores across annotators. The results are presented in Table E.2.

Interpreting the value of Fleiss’ kappa in absolute terms is challenging. While Landis and Koch 1977 propose an interpretation scale that would classify our results for the evaluation dataset annotations between "slight agreement" and "fair agreement," this classification is not universally accepted. Actually, an alternative and commonly used interpretation suggests that a value below 0.60 indicates a "significant level of disagreement" (Ranganathan, Pramesh, and Aggarwal 2017).

That said, it is worth noting that the three indicators (Fleiss’ kappa, IoU, and F1) are generally proportional across zones. For example, Zone 5 shows the lowest

score across all three metrics, while Zones 17 and 19 exhibit the highest agreement and performance scores.

When visually inspecting the STAPLE masks shown in Appendix G, we observe that annotations tend to concentrate around four types of features, approximately in the following order of importance:

- Water areas (rivers, lakes, coastlines),
- Major roads (yellow-marked roads), along with some secondary roads and railways when no major roads are present,
- Urban areas (especially at smaller map scales, such as in Zones 9 and 10),
- Natural landmarks that stand out from their surroundings (e.g., city parks, distinct forests, or the meanders seen in Zone 11).

The presence of these elements (particularly water features and salient roads) appears to enhance inter-annotator agreement, as seen in Zones 1, 4, 8, and 19. Conversely, zones with a high density of visual information tend to yield lower agreement. In some cases, annotators likely hesitated between multiple potential landmarks (e.g., Zones 5, 9, 14), while in others, no element clearly stood out (e.g., Zones 15, 18, 20).

Moreover, although the annotation protocol provided guidance on how to handle large landmarks by specifying that annotators should fully color a landmark if it fits within the image, and trace its borders if it extends beyond the image, this rule appears to struggle with edge cases. In modern digital maps, zooming is effectively continuous, which often creates scenarios where a previously useful landmark becomes too large to be identified as a whole, and its boundaries then become the main reference point.

This phenomenon is illustrated in Zone 9 with the city of Toulouse: some annotators colored the entire city center, while others selected only specific parts of it. A similar issue occurs in Zone 14 with the large green area on the left side of the image. These inconsistencies artificially lower inter-annotator consensus and negatively affect the performance scores in Table E.2.

Overall, we believe these variations in focus among users viewing the same map at the same scale arise from multiple factors that are difficult to control. These may include individual perceptual differences, the specific task the user is trying to perform with the map, or even contextual aspects such as screen size (although all annotators in our study used the same screen dimensions) and viewing distance.

Nonetheless, the values obtained for this evaluation dataset show that, for most areas, there is inter-annotator consensus well beyond what would be expected by chance (as we had already observed visually). The inherent fuzziness of AHAPs themselves may at least partially explain the relatively modest scores. This evaluation dataset thus provides a useful initial benchmark for assessing the performance of our segmentation models, which will be presented in the following chapter.

Zone	Fleiss' Kappa	Mean IoU	Std IoU	Mean Dice	Std Dice
01	0.54	0.55	0.13	0.70	0.12
02	0.38	0.46	0.20	0.60	0.21
03	0.42	0.45	0.18	0.59	0.19
04	0.54	0.52	0.14	0.67	0.13
05	0.25	0.37	0.16	0.52	0.16
06	0.50	0.52	0.15	0.67	0.13
07	0.54	0.54	0.13	0.69	0.11
08	0.58	0.59	0.16	0.73	0.13
09	0.30	0.39	0.20	0.52	0.24
10	0.36	0.49	0.20	0.63	0.22
11	0.37	0.46	0.23	0.60	0.25
12	0.45	0.48	0.16	0.64	0.16
13	0.34	0.43	0.16	0.59	0.16
14	0.33	0.42	0.29	0.54	0.27
15	0.32	0.43	0.18	0.57	0.20
16	0.46	0.55	0.14	0.70	0.12
17	0.60	0.62	0.14	0.76	0.11
18	0.32	0.43	0.21	0.57	0.23
19	0.59	0.67	0.21	0.78	0.16
20	0.26	0.42	0.16	0.57	0.18
Mean	0.42	0.49	0.18	0.63	0.17

✱ Table E.2 : Fleiss' Kappa, IoU, and Dice Scores Across Zones.

E.4 Conclusion of the chapter

In this chapter, we introduced the concept of Areas of High Anchoring Potential (AHAPs), which we believe can support landmark detection and the identification of critical zones in maps, as well as offer an additional analytical perspective for understanding how humans perceive and memorize maps.

To assess inter-annotator consensus for AHAPs (1) and provide a benchmark for evaluating future segmentation models (2), we created a reference dataset composed of 20 map images from various locations and scales across France. Each image was annotated by 15 participants with GIS knowledge, all of whom were taught about the concept of anchors. Participants were asked to highlight the potential anchors they identified in the maps, along with their type.

We applied the STAPLE methodology to estimate a consensus ground truth segmentation for each image and anchor type. We then computed performance metrics by comparing each individual annotation to this ground truth, allowing us to quantify the similarity of the annotations across experts.

While the average inter-annotator agreement was moderate according to the evaluation metrics, the masks generated by STAPLE still revealed coherent patterns.

Additionally, we observed significant variation between map areas, which we partially attribute to the nature of the features present in each zone.

Finally, this experiment provided an estimate of the best quantitative results we might expect from models trained to perform this task, and will serve as the foundation for evaluating the AHAP segmentation models discussed in the next chapter.

Chapter F

Detecting areas of high anchoring potential with deep learning

In this chapter, we train two machine learning models, a U-Net and a SegFormer, to detect Areas of High Anchoring Potential (AHAPs) in map images. We begin by introducing the concept of image segmentation and describing the two models used in this study. Next, we detail the process of creating our training dataset. Finally, we outline the implementation details and parameter settings for each model, and present and discuss their respective training results.

F.1 About image segmentation

F.1.1 Definition

Image segmentation refers to the process of labeling each pixel in an image according to a predefined set of classes. It can be binary, such as identifying pixels that correspond to a dog in a photograph, or multi-class, such as classifying pixels in an aerial photograph based on land use.

Segmentation can be performed manually by human annotators, or through algorithmic techniques designed to detect specific features (such as early saliency detection algorithms), or else by using machine learning models trained to identify particular elements in images.

In our study, we focus on binary segmentation, where pixels are classified as either AHAP or non-AHAP. Our objective is to evaluate the potential of supervised learning models for this task. To do so, we selected two existing models from the literature and trained them to assess their performance.

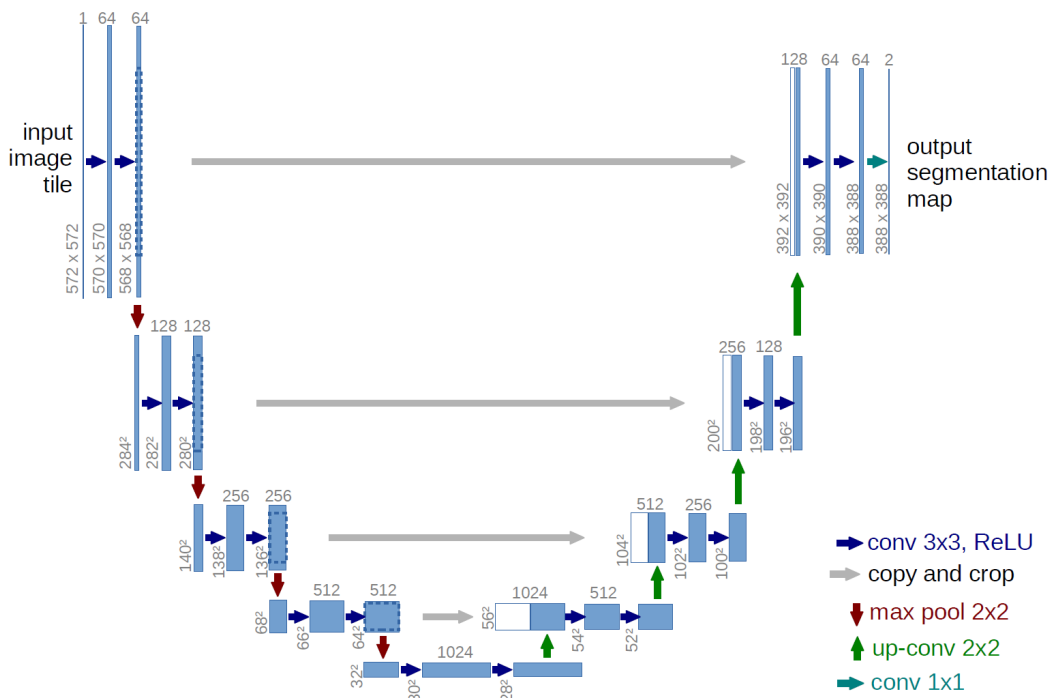
F.1.2 CNNs and U-Net

Historically, Convolutional Neural Networks (CNNs) have been the dominant architecture for image processing tasks, including segmentation. Actually, we ourselves employed a CNN to classify landmark buildings in Chapter D. CNNs are particularly

well-suited for visual tasks due to their ability to extract hierarchical spatial features through convolutional layers.

U-Net (Ronneberger, Fischer, and Brox 2015) is a CNN-based architecture specifically designed for image segmentation, originally developed for biomedical image analysis. It has since been adopted in a wide range of disciplines. Its distinctive U-shaped architecture includes three main components:

1. Encoder: A downsampling path that extracts high-level features through successive convolution and pooling operations.
2. Decoder: An upsampling path that reconstructs the segmentation mask from the encoded features.
3. Skip connections: Direct links between encoder and decoder layers at corresponding resolutions, which help preserve fine spatial information lost during downsampling (Figure F.1).



⊛ Figure F.1 : U-net architecture example from Ronneberger, Fischer, and Brox 2015. We can see the encoding part on the left side and the decoding part on the right side, with the skip connections symbolized by grey arrows.

U-Net has proven highly effective in various segmentation tasks, including in geospatial contexts such as remote sensing (Ren, Yu, and Guan 2020) and the analysis of scanned historical maps (Ekim, Sertel, and Kabadayı 2021). Its ability to learn meaningful segmentation patterns from relatively small datasets makes U-Net a strong candidate for AHAP detection. However, one of our key hypotheses is that anchors influence one another in the mental map of users, meaning that neighboring

AHAPs may interact and reinforce each other. For example, three small lakes arranged in an equilateral triangle may form a more prominent landmark than a single, isolated lake.

This type of spatial interaction presents a limitation for U-Net. The model is constrained to operate on relatively small image tiles, as its computational requirements increase rapidly with image size. While it is technically possible to tile a large map into smaller segments and run predictions independently on each tile, this approach prevents the model from capturing relationships between distant elements. Such limitations could reduce the model’s effectiveness for detecting anchor zones, whose importance may emerge only from their spatial context. While there are some workarounds, such as artificially enlarging U-Net’s receptive field by downsampling the input image (at the cost of resolution), those approaches were not explored in our study.

To address this limitation and complement our analysis, we instead turned to more recent deep learning architectures based on transformers, which can process larger images and are theoretically better equipped to capture global spatial relationships.

F.1.3 Transformers and Segformer

Transformers are a class of deep learning architectures based on the multi-head attention mechanism, which allows the model to weigh and relate all parts of an input sequence to one another. Originally introduced for natural language processing (NLP), transformers process input tokens in parallel and capture contextual relationships over long distances, making them highly effective for tasks such as translation, summarization, and question answering. They power large language models such as GPT and BERT, which learn semantic structure by attending to words in context across an entire text.

Following their success in NLP, transformer architectures were adapted for vision tasks in the early 2020s with the Vision Transformer (ViT) (Dosovitskiy et al. 2021), which treats an image as a sequence of patches instead of a 2D grid. Each patch is embedded as a token and processed through the same attention-based layers used in NLP, enabling the model to capture global spatial dependencies more effectively than traditional convolutional networks.

More recently, transformers have outperformed convolutional networks in a range of medical imaging tasks, largely due to their ability to capture long-range dependencies within visual data (Viriyasaranon, Woo, and Choi 2023). Like U-Net, transformer-based models have quickly been adopted in geospatial applications, including satellite image analysis and historical map interpretation (MacDonald, Jacoby, and Coady 2024; Sertel, Hucko, and Kabadayı 2024).

Concerning our use case, a major advantage is that transformer-based vision models are less constrained by image size than U-Net. This allows us, in theory, to input full map views during training and prediction, enabling the model to learn interactions between distant AHAPs—something U-Net cannot capture when working on small tiles.

Moreover, our specific challenge (detecting AHAPs) is not solely about identifying individual visual elements, but understanding how context contributes to their

salience. Just as large language models (LLMs) can infer the meaning of a sentence and its emotional tone by analyzing how words relate to one another, we hypothesize that transformer-based vision models may be better equipped to detect features that stand out due to their spatial and semantic relationships with their surroundings.

That said, transformer models come with trade-offs. CNNs are generally more sample-efficient and can infer implicit rules with fewer training examples. Transformers, in contrast, often adhere more closely to patterns seen during training, making them more susceptible to overfitting, especially on small or medium-sized datasets (Géron 2022). Given the relatively limited size of our dataset, careful monitoring is hence required to prevent overfitting.

Among the various transformer-based segmentation models such as Segmenter (Strudel, Garcia, Laptev, and Schmid 2021), DETR (Carion, Massa, Synnaeve, Usunier, Kirillov, and Zagoruyko 2020), and MaskFormer (Cheng, Schwing, and Kirillov 2021), we selected SegFormer (Xie, Wang, Yu, Anandkumar, Alvarez, and Luo 2021) for our experiments. This choice was motivated by its ease of use, availability of pre-trained weights, and flexibility in handling input images of varying sizes without the need for model customization.

F.2 Training dataset

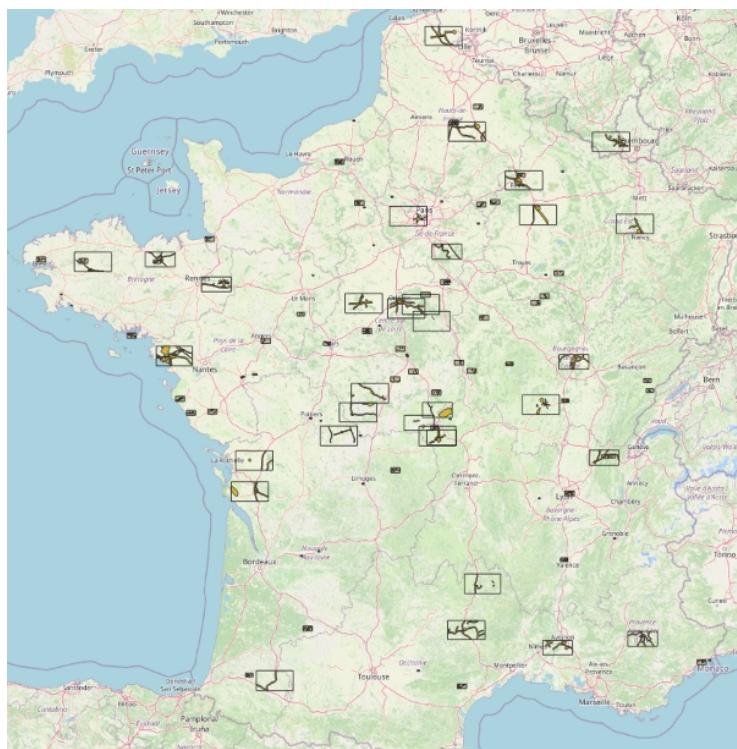
The dataset used to train our models was created using the MapDraw annotation tool, following a protocol very similar to the one used for the evaluation dataset described in Section E.2.3 of the previous chapter. It is composed of three parts, collected at different stages of the PhD, but all designed with segmentation model training in mind. Together, these parts aim to ensure diversity in map styles, geographic coverage, and annotators.

The first part was generated early in the PhD (January 2023) during a collaborative annotation session involving 12 participants from our research team and collaborators, all of whom regularly work with maps. For this initial dataset, we selected the IGN map style due to its compatibility with additional geographic data from the IGN database and its alignment with earlier segmentation experiments conducted on the same map style.

At that time, our anchor taxonomy included 10 categories instead of the current 5. The categorization process was more complex under this earlier taxonomy, which later motivated us to simplify and consolidate the categories for the evaluation dataset. During the session, which lasted approximately 40 minutes, participants were asked to freely annotate any areas they perceived as having high anchoring potential. Each participant annotated as many areas as they wished, and most areas were therefore labeled by a single annotator.

A total of 94 map views from the IGN plan were annotated, evenly distributed across zoom levels 12, 14, and 16, with an average of approximately three anchors per map view (Figure F.2).

The second and third parts of the dataset were created later in the PhD (late 2024 and early 2025), this time by a single annotator (myself), with the aim of increasing the volume of training data and incorporating additional cartographic styles. Although the annotation procedure remained largely similar, these new annotations



⌘ Figure F.2 : An overview of the localisation of the map views annotated during the collaborative session.

did not include anchor categories. This decision was made to save time and because I knew at this point that the models would be trained on a unified set of anchor areas, regardless of category. Specifically, I annotated 106 map views from Google Maps and 36 from OpenStreetMap (OSM). These annotations span zoom levels 12 to 18 evenly, ensuring a representative range of map scales.

To summarize, the final training dataset is composed of:

- 94 map views from the IGN map (12 annotators, old anchor categories, but merged together anyway)
- 106 map views from Google Maps (single annotator, no anchor categories)
- 36 map views from OSM (single annotator, no anchor categories)

For each annotated map view, both the original map image and a corresponding binary mask—representing anchor areas merged across all categories—were exported (see Figures E.5 and E.6). These image-mask pairs were then tiled into smaller patches tailored to the input requirements of each model. The specific tiling strategies used for U-Net and SegFormer are detailed in the following sections. The training datasets are all available online¹.

¹<https://zenodo.org/records/15238773>

F.3 Experiment 1: U-Net

F.3.1 Dataset pre-treatment

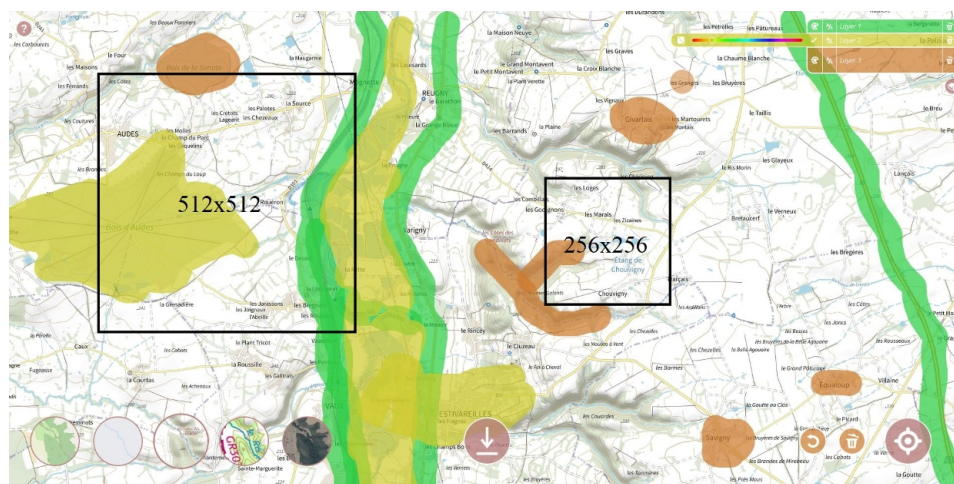
The U-Net implementation we used cannot process images at full screen resolution. Therefore, the training dataset was divided into smaller tiles of size 512×512 pixels, as shown in Figure F.3. This size represents a trade-off between computational efficiency and contextual coverage. Larger tiles significantly increase the computational cost, while smaller tiles capture less spatial context, often resulting in fewer visible anchors per tile, as illustrated in Figure F.4.

To further expand the training dataset and improve model robustness, we applied data augmentation by generating tiles with a 50% overlap. In total, 2,329 tiles were extracted from the annotated maps, each paired with its corresponding segmentation mask.

The evaluation images from the previous chapter were also divided into 512×512 tiles in order to make predictions using our model. We used a sliding window approach with overlap. For each map view, the height was covered by three overlapping tiles and the width by five overlapping tiles, resulting in a total of 15 tiles per evaluation zone. The U-Net predictions obtained from these tiles were then reassembled to reconstruct the full zone by averaging the probabilities in the overlapping regions.



☞ Figure F.3 : Examples of tile/mask pairs extracted from the training dataset in order to train the U-Net model.



❁ Figure F.4 : Comparison between 512px and 256px tiles relative to a full map view. Smaller tiles capture less anchor diversity and interactions.

F.3.2 Implementation and Training

Although several enhanced U-Net variants exist—such as U-Net++ (Zhou, Siddiquee, Tajbakhsh, and Liang 2018), which uses nested skip connections, and Attention U-Net (Oktay et al. 2018), which integrates attention mechanisms—we opted for a simpler architecture closely following the original U-Net design proposed by Ronneberger (Ronneberger, Fischer, and Brox 2015), in order to streamline the implementation. Our code for training both the U-Net model and the Segformer model are available online².

The model was trained from scratch, without pre-trained weights. As for validation, We reserved 10% of the tiles, selected randomly. However, this approach introduces a risk of spatial data leakage between training and validation sets, due to the overlap-based augmentation, meaning that similar spatial patterns could be present in both subsets. This makes the model more prone to overfitting, which is why close monitoring of both training and validation loss was essential to determine when to stop training.

We monitored both the loss function and the mean Intersection over Union (IoU) during training (see Figure F.5). The training loss decreased steadily until around epoch 100, while the validation IoU peaked around epoch 120.

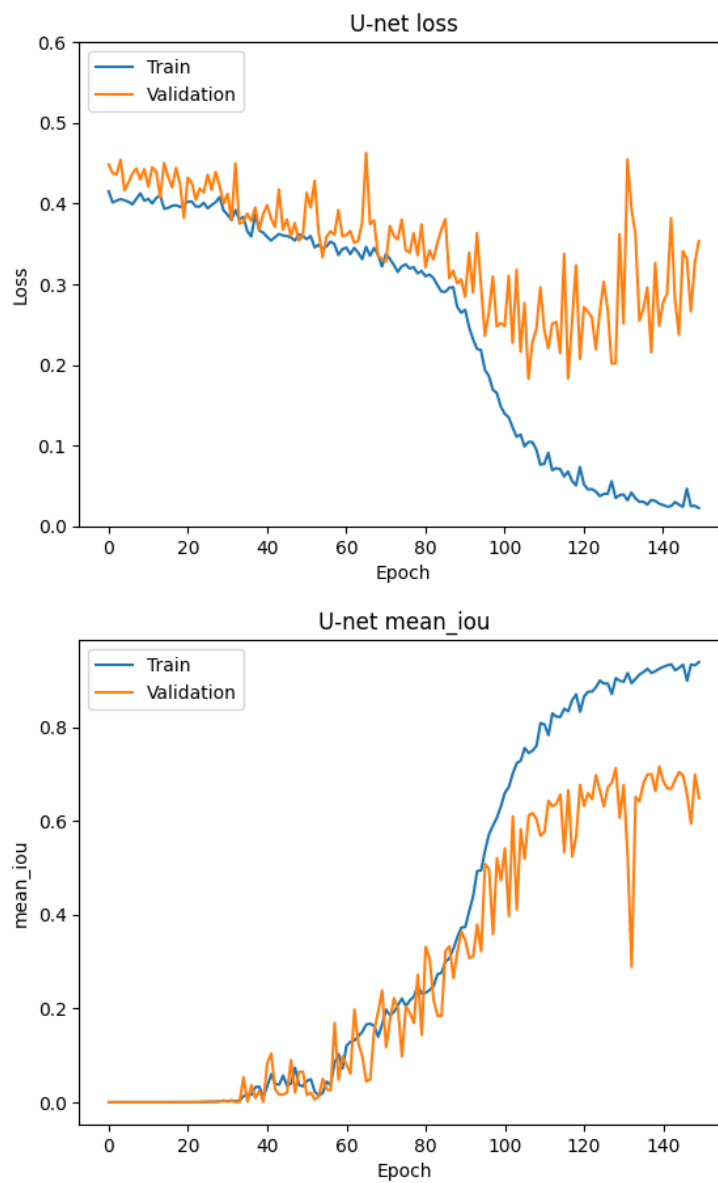
For evaluation, we selected the best-performing model weights, which were saved at epoch 107, when the validation loss reached its minimum (0.23), and the validation mean IoU was approximately 0.62.

❖ Training Configuration Summary:

- Model architecture: U-Net
- Pre-training: No
- Input size: 512×512 pixels

²<https://github.com/TinQk/LostInZoom/>

- Training images: 2,329
- Validation split: 10% (random)
- Batch size: 32
- Hardware: Google Colab (L4 GPU)
- Training time: 135 minutes for 150 epochs
- Evaluation checkpoint: Epoch 107



⌘ Figure F.5 : Evolution of U-Net model loss and IoU during training.

F.4 Experiment 2: Segformer

F.4.1 Dataset pre-treatment

Although the SegFormer architecture is capable of handling full-resolution map views, we chose to artificially expand our dataset by generating smaller cropped images from the original annotated views. For each map view, we extracted five sub-images of size 1200×600 pixels: one centered and four aligned with each corner. This strategy ensured coverage of the central region—where most AHAP annotations are concentrated—while offering sufficient spatial context to learn interactions between anchors. This preprocessing resulted in a total of 1,172 training images.

For the predictions, however, we directly provided the full evaluation images to the model. Since the output predictions were smaller than the original images, we upsampled them using bilinear interpolation on the pixel-wise probability values.

F.4.2 Implementation and training

We used the Hugging Face Transformers library to load and fine-tune a pretrained SegFormer model. Specifically, we selected the Mix-B0 variant recommended in the original article, pretrained on ImageNet-1k, as a compromise between performance and resource usage. Although this pretraining does not include cartographic data, the model can still benefit from learned visual patterns such as contours and geometrical shapes.

As a baseline, we tested the pretrained model without fine-tuning, but it produced only noise on our evaluation set—likely due to the absence of map-related classes in the ImageNet dataset. This confirmed the need for domain-specific fine-tuning.

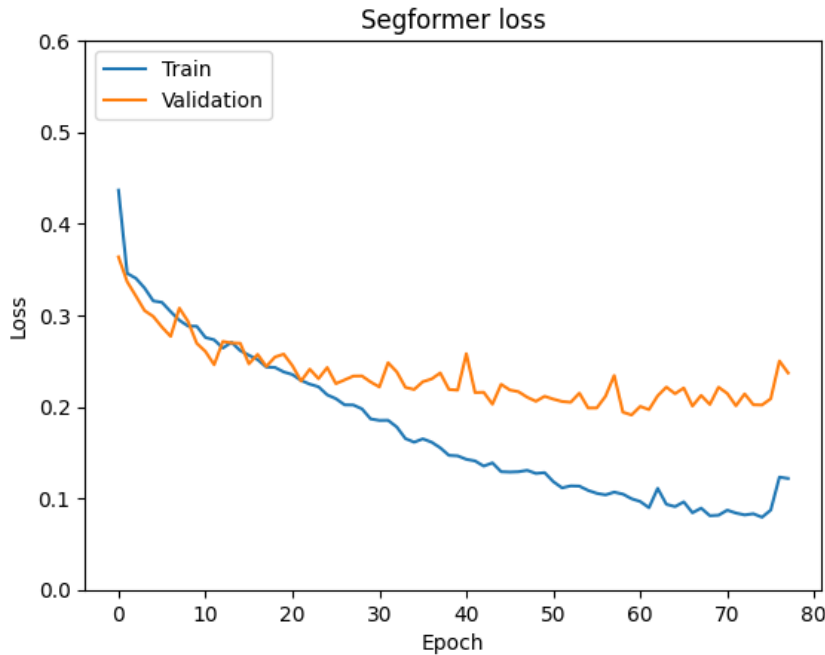
The dataset was randomly split, with 10% (118 images) used for validation. While this approach introduced spatial leakage, since overlapping regions may appear in both training and validation sets, it was a practical compromise given the dataset’s size. To mitigate overfitting, we closely monitored training and validation loss curves.

During training, we observed that the validation loss stabilized around epoch 40–50 while the training loss continued to decline until approximately epoch 70 (Figure F.6). We selected the model weights from epoch 48 for evaluation. At this point, validation loss had plateaued around 0.21, which we considered satisfactory.

❖ Training Configuration Summary:

- Model architecture: SegFormer Mix-B0
- Pre-training: ImageNet-1k
- Input size: 1200×600 pixels
- Training images: 1,172
- Validation split: 10% (random)
- Batch size: 8

- Hardware: Google Colab (L4 GPU)
- Training time: 205 minutes for 78 epochs
- Evaluation checkpoint: Epoch 48



✂ Figure F.6 : Evolution of Segformer model loss during training.

F.4.3 Challenges and workarounds

During training, we encountered a memory leak issue: CPU memory usage increased significantly after each epoch until the runtime crashed. Despite extensive investigation, the root cause remained unidentified. As a workaround, we trained the model in segments of 8 epochs, saving the model and restarting the runtime between sessions to clear memory.

Another unexpected issue was that standard model weight saving and loading did not preserve the model's predictive behavior. Only saving and restoring full TensorFlow checkpoints reliably restored the model's ability to make consistent predictions. This forced us to revise our workflow accordingly for model evaluation.

F.5 Predictions and metrics

Predictions on the evaluation dataset were generated for both models. For each prediction, we produced two outputs:

- A heatmap representing the per-pixel probability of an AHAP classification.
- A binary segmentation map, obtained by applying a 50% threshold to the heatmap (see Figure F.7).



☼ Figure F.7 : Top: Original map image of Zone 2 from the evaluation dataset. Bottom left: Predicted AHAP probability heatmap from Segformer model. Bottom right: Corresponding binary segmentation map.

The complete set of evaluation images, along with the corresponding STAPLE consensus and predictions of both models, is provided in Appendix G, and can be downloaded³. For each zone of the evaluation dataset, the original map image is displayed in the top-left corner, the STAPLE result in the top-right, the U-Net heatmap in the bottom-left, and the SegFormer heatmap in the bottom-right. We chose to display heatmap instead of binary segmentation in this manuscript because we think they bring more information about how well the two models worked.

To quantify model performance, we computed IoU and Dice scores between the predicted binary segmentations and the binary STAPLE-generated masks. These results are summarized in Table F.1, and form the core of our comparative analysis in the following section.

For reference, we also included the mean IoU and Dice scores between the individual annotators and the STAPLE masks. Although these scores are inherently biased upward—since the STAPLE masks were generated from the annotators’ inputs, they still offer useful insight into how closely the models approach human-level consensus.

³<https://zenodo.org/records/15238773>

Zone	Annot IoU	Unet IoU	SegF IoU	Annot F1	Unet F1	SegF F1
01	0.55	0.27	0.19	0.70	0.42	0.31
02	0.46	0.32	0.54	0.60	0.49	0.70
03	0.45	0.18	0.26	0.59	0.31	0.42
04	0.52	0.20	0.27	0.67	0.33	0.42
05	0.37	0.15	0.28	0.52	0.27	0.44
06	0.52	0.03	0.10	0.67	0.05	0.19
07	0.54	0.17	0.34	0.69	0.28	0.50
08	0.59	0.27	0.32	0.73	0.43	0.48
09	0.39	0.11	0.45	0.52	0.20	0.62
10	0.49	0.31	0.58	0.63	0.47	0.73
11	0.46	0.40	0.75	0.60	0.57	0.86
12	0.48	0.32	0.45	0.64	0.48	0.62
13	0.43	0.09	0.33	0.59	0.16	0.49
14	0.42	0.14	0.18	0.54	0.25	0.31
15	0.43	0.41	0.43	0.57	0.58	0.61
16	0.55	0.29	0.45	0.70	0.45	0.62
17	0.62	0.13	0.23	0.76	0.23	0.37
18	0.43	0.01	0.07	0.57	0.02	0.12
19	0.67	0.32	0.50	0.78	0.49	0.67
20	0.42	0.01	0.02	0.57	0.01	0.04
Mean	0.49	0.21	0.34	0.63	0.32	0.48
Std	0.08	0.12	0.18	0.08	0.18	0.21

✱ Table F.1 : Model Performance (IoU and F1 score) for each zone.

F.6 Discussion on the results

It is important to note that the purpose of this section is not to compare the two trained models directly. Their different architectures imply different usage scenarios, and a technical comparison between convolutional models and transformer-based models is best carried out by machine learning specialists. Additionally, we trained the U-Net from scratch, while the Segformer was fine-tuned from a pretrained model. This makes any direct comparison between their performance somewhat imbalanced. Nevertheless, we can highlight the respective strengths and limitations of each model in the context of our specific task: segmenting AHAPs in map images.

As this section contains frequent references to Appendix G, we recommend that readers of the printed version bookmark the relevant appendix section, and that digital readers consider opening the document in two windows or tabs to easily switch between the discussion and the appendix.

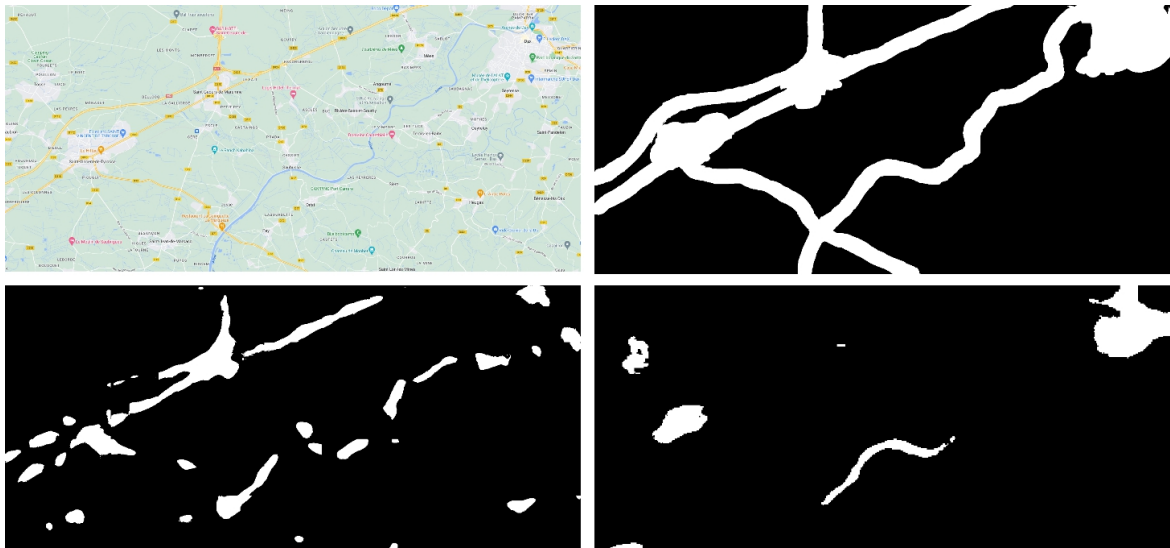
F.6.0.1 General observations

First, we can observe both looking at Table F.1 and to the prediction in the appendix that the two models successfully learned meaningful patterns from the training data. Although the average scores of the U-Net remain relatively low, they are still significantly above random predictions.

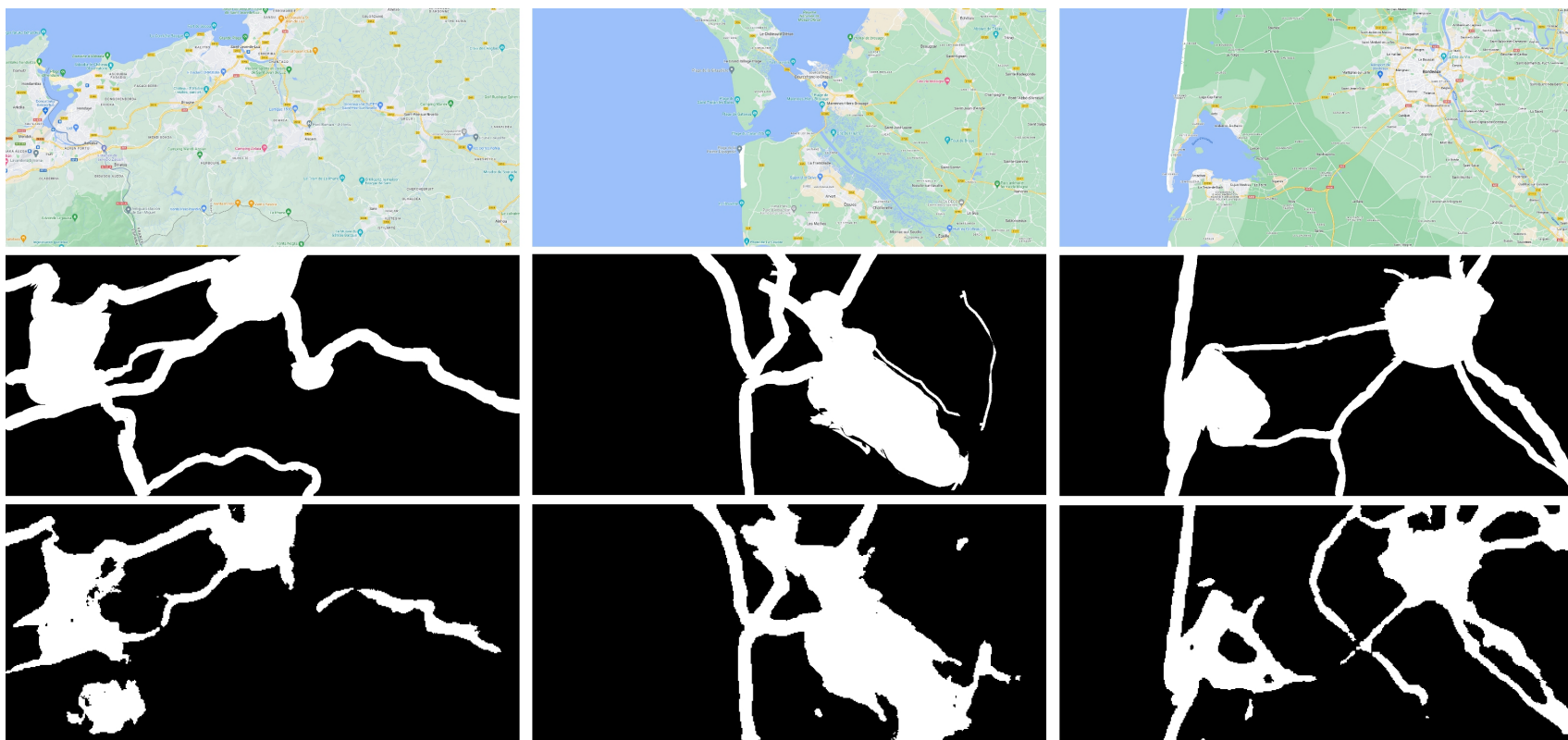
Looking at the heatmaps, Segformer’s predictions tend to closely resemble the STAPLE masks, apart from some missing elements such as (white) roads in Zones 1, 4, 7, 8, 16, and 18, or railways in Zones 17 and 20. In some cases, Segformer predicted elements that did not achieve strong consensus among annotators, such as the green area at the bottom left of Zone 2 or the transition between two green tones in Zone 10. However, inspecting individual annotations reveals that 3 and 2 annotators, respectively, marked these areas similarly, showing that the model aligns with a subset of annotators’ intuitions.

Overall, the Segformer model consistently outperforms the U-Net across all zones, with the exception of Zone 1, where the U-Net achieved better results. In this zone, Segformer particularly underperformed, especially in detecting the river and major roads (see Figure F.8 and Appendix 28). However, it is difficult to identify a clear explanation for this, as the cartographic features in Zone 1 are quite similar to those in other zones where Segformer performed well.

In general, the U-Net’s results were rather mixed, while some of the Segformer predictions were particularly promising (see Figure F.9). The fact that Segformer processes entire map views without the need for tiling seems to have had a significant positive impact on its performance.



✿ Figure F.8 : Base image of Zone 1 (top left), binary STAPLE (top right), U-Net prediction (bottom left), Segformer prediction (bottom right).



* Figure F.9 : From top to bottom and left to right: Base image, binary STAPLE and Segformer predictions on zones 2, 11 and 10, where predictions seemed satisfying.

F.6.0.2 Inter-Annotator Agreement and Model Struggles

Both models show weaker performance in zones that are highly zoomed-in and centered on urban areas (such as Zones 6, 18, and 20), which also correspond to areas of low inter-annotator agreement. Similarly, Zone 5, which is not urban, also led to poor predictions from both models and low annotator agreement, as it is visible in the non-binarized STAPLE mask shown in Figure 32 of the Appendix. At this stage, it remains unclear whether the models struggle because these zones lack clearly defined AHAPs, or because such environments were under-represented in the training set.

F.6.0.3 The problem of highlighting surfaces versus boundaries

As discussed in Section E.3.2, there are nuances in how annotators highlight boundaries versus surfaces. For instance, U-Net introduces a hole in the lake in Zone 14, possibly due to its limited tile view which may have caused it to focus on contours rather than surfaces. Segformer, with access to the entire image, does not make the same mistake.

Moreover, Segformer’s lower performance in Zone 14 may result from the STAPLE mask emphasizing the entire forest on the left side, which may not reflect a consistent annotation practice, given the guidelines provided. Segformer instead highlighted the forest’s edges, suggesting a more boundary-based interpretation. This raises the recurring issue of how users transition their attention from a landmark as a whole to its boundaries, or vice versa. This transition remains critical for the automatic detection of AHAPs.

F.6.0.4 Sensitivity to clusters of nearby anchors

In the Segformer predictions, AHAPs tend to emerge when multiple anchoring elements are located in close proximity. This behavior is visible in several zones. In Zone 5, the combination of a river, roadside, and nearby villages stands out, while more isolated roads are ignored. In Zone 9, dense urban features contribute to strong predictions in the Toulouse city center. Similarly, in Zone 11, urban areas near river meanders are well detected. These examples suggest that spatial interactions between nearby elements may play a role in guiding the model’s attention.

F.6.0.5 Visual analysis by anchor category

- Water-related features (rivers, lakes) are consistently well detected. The models appear to differentiate between features to be fully annotated (e.g., rivers, lakes) and coastlines, where only boundaries are highlighted. This aligns with human annotations.
- Parks and forests (green areas) are well detected by Segformer, which tends to segment them along their boundaries when they are large, rather than in full. In comparison, the U-Net does not appear to have learned to detect them at all.
- Roads are more hit-or-miss. Roads in yellow are frequently detected, while white roads are less consistently segmented or missed altogether (i.e. Zone

7 and zone 8), potentially due to under-representation in training. Interestingly, Segformer sometimes emphasizes certain segments and ignores others. This suggests the model may have learned a form of context-sensitive filtering, raising a promising question for future research: what implicit rules determine whether a road segment is anchor-worthy?

- Urban areas are generally detected when they are prominent and isolated. In Zone 2, two villages (center and south-east) are somewhat ignored by Segformer, while a similarly sized area in Zone 5 stands out. This suggests that the perceived salience of a village as an anchor may depend more on relative isolation or dominance in the surrounding space than on size alone.
- Map toponyms are harder to evaluate. Salient toponyms often appear in zones that also contain other anchoring elements, making it difficult to isolate their effect. More targeted testing will be necessary to better understand how the models treat textual map elements.

F.6.0.6 Note on the nature of segmentation and evaluation metrics

It’s important to note that evaluation metrics such as IoU and F1 score compare predicted and reference areas as binary masks, focusing on surface overlap. In practice, however, it is not always necessary to segment the full surface of a landmark to identify it. For instance, in Zone 1, even though Segformer only segments a part of the diagonal river, this is sufficient for a human observer—or a downstream data matching algorithm—to recognize it as a landmark. This nuance opens the door to future segmentation or matching strategies that tolerate partial coverage or integrating post-processing steps that support more flexible matching.

F.6.0.7 Final thoughts

While results show that the models still need refinement, and the current evaluation setup may not fully reveal their strengths and weaknesses, the outcomes remain encouraging. We observe behaviors that go beyond classic object segmentation and seem to incorporate contextual understanding of how map elements interact to form landmarks.

In particular, the Segformer model—capable of processing larger images and theoretically capturing global relationships—appears to have the potential for more advanced tasks such as data matching and landmark detection.

F.7 Ideas for further experiments with AHAPs

Our experiments stop here for the purpose of this thesis, but several directions could be explored in future work.

One direction would be to test the ability of our models to highlight a salient landmark that they have never encountered during training. This type of situation is critical because many geographic areas are unique, and certain map symbols may

be rare or entirely absent from even a well-populated training set. To simulate such cases, we could manually insert artificial elements (e.g., a road or a symbol) into existing map images and observe whether the model can correctly detect them as potential anchors.

Another valuable test would be to assess whether a model can evaluate the anchoring potential (a notion defined as anchor intensity in Gruget, Touya, and Muehlenhaus 2024) of the same element depending on its context. For example, we would expect an isolated building or a road segment that is part of a ring road to be more likely to serve as a landmark. Ideally, ring roads should be detected as AHAPs more than other roads, although this seems currently unlikely given the results of our models.

These two capabilities — generalization to unseen salient features and sensitivity to contextual salience — are typical of saliency detection models. It would be interesting to evaluate how such models perform on our evaluation dataset. A promising path could be to fine-tune an existing saliency detection model rather than relying on a model pre-trained for generic object segmentation in photographs, as we did in this work.

Another idea would be to incorporate the concept of geometrical regularity into our pipeline in order to obtain smoother segmentation masks. This could be achieved either directly in the model architecture, as suggested in Ravishankar, Venkataramani, Thiruvankadam, Sudhakar, and Vaidya 2017, or through post-processing applied to the predicted masks.

Finally, one of our initial ideas was to distinguish between different types of anchors. We abandoned this path for the scope of this thesis as it would require building a much larger and more precisely annotated training dataset, but doing so could allow models to learn the specific interactions between different categories of anchors and even detect overlapping anchors of different natures. This could be approached using a multi-output regression model, with input images composed of multiple layers, one for each anchor type.

F.8 Limitations

F.8.0.1 Data availability

As a first limitation, training models such as ours requires data that does not currently exist and must be manually produced. This process is time-consuming, expensive, and introduces subjective variability. Even once trained, it remains difficult to predict whether these models will perform well when applied to different territories or map types. The adaptability of models like ours must thus be verified before getting part of a production process.

F.8.0.2 Task-relevance

Another limitation of all our experimental setups in this thesis is that participants were not asked to perform typical map use tasks such as route following, distance estimation, or area comparison. Research shows that the nature of a task influences both attention patterns and brain activity (Qin, Fias, Van de Weghe, and Huang

2024). Although we assume that ring roads and major landmarks would be perceived as such regardless of task, this assumption may not hold true for all anchors. Task-specific perception may affect what is noticed, how it is encoded, and ultimately, how anchors are interpreted.

F.8.0.3 Lack of multi-scale and contextual reasoning

Now if we take a step back, a major limitation of our work is the absence of a multi-scale component in our anchor detection models. AHAPs are inherently multi-scale: a feature may only become a cognitive anchor when viewed at a specific zoom level or in relation to other nearby features. By using single-scale models, we restrict the scope of our conclusions and reduce the reliability of our predictions across varying map scales.

To capture this multi-scale nature, it would be necessary to integrate information from multiple zoom levels, potentially by representing each scale in a separate input channel or designing architectures capable of learning hierarchical representations. This would also help the model learn not just individual anchor features but their spatial and semantic relationships across scales.

Such relational reasoning is already central to other domains like safety and surveillance, where the context of a detected object, rather than the object alone, defines its importance. For instance, identifying a knife in an image does not necessarily signal danger unless it appears in proximity to a person or in a threatening configuration (Jang, Battulga, and Nasridinov 2020). This parallel is highly relevant to cartographic salience: a landmark is rarely defined in isolation but becomes salient within a specific spatial configuration.

Interestingly, some threat detection systems in video analysis leverage temporal context, using sequences of frames to detect emerging risks. A similar paradigm could be applied to cartography by replacing time with zoom level. Comparing map representations of the same region at different levels of generalization could help identify discontinuities or transitions in anchoring potential. This suggests that future models designed to detect AHAPs should integrate both multi-scale data and spatial-semantic context to better reflect the cognitive nature of map reading.

F.8.0.4 Environmental and computational costs

Finally, the environmental impact of training deep learning models cannot be ignored. Although our experiments were relatively lightweight, scaling up to large-scale automatic anchor detection would require significant computational resources and energy consumption. This raises ethical concerns in the context of climate change and biodiversity loss. Any large deployment of such models should first undergo a cost-benefit analysis that includes environmental considerations.

F.9 Conclusion of the chapter

In this chapter, we trained two deep learning models to segment Areas of High Anchoring Potential (AHAPs) in map images and presented their results using the evaluation dataset introduced in Chapter E. Visual examples of the predictions can

be found in Appendix G, while quantitative results are reported in Table F.1.

The U-Net model, trained from scratch and limited to 512×512 image tiles, produced modest results. In contrast, the Segformer-based model delivered more promising outcomes—particularly from a visual analysis perspective—and achieved reasonable quantitative scores, especially when considering the relatively low inter-annotator agreement for this segmentation task.

These results suggest that transformer-based segmentation models, thanks to their ability to capture broader spatial dependencies and contextual information, may have potential for powering downstream workflows such as landmark detection or critical zone identification in cartographic applications.

Building on these insights, several future directions could be explored. These include leveraging existing saliency detection algorithms, designing post-processing strategies for probabilistic heatmaps, and investigating ways to distinguish between different categories of anchors. Such avenues may help refine both the modeling approach and our understanding of what constitutes an anchor in cartographic imagery.

Thesis Conclusion

Reminder of the context and core question

This work is situated within a pragmatic context: the rise of pan-scalar digital maps and the development of automated generalization techniques. These advances enable maps to dynamically adapt to different scales, but they also introduce new cognitive challenges for users, who may become disoriented as visual content changes during zoom operations. Simultaneously, this work belongs to a broader research landscape that brings together cartographic design, human-computer interaction (HCI), and cognitive science—disciplines that all share an interest in how people perceive, interpret, and use spatial representations.

The central question of this thesis was:

Can we design pan-scalar map landmarks in a way that influences the formation of cognitive anchors in users' mental maps?

To approach this question, it is important to clarify what is meant by pan-scalar landmarks and cognitive anchors. Pan-scalar landmarks are features that retain their visual and cognitive salience across multiple levels of zoom, providing continuity in a map that changes dynamically with scale. Cognitive anchors, on the other hand, are the mental reference points users rely on to maintain orientation and construct a coherent spatial understanding of the map. In this thesis, we explore how these concepts intersect: how certain visual or functional elements—such as ring roads or remarkable buildings—can act as pan-scalar landmarks, and how their presence or absence influences users' mental mapping.

Contributions

The first contribution of our thesis is a state-of-the-art review that brings together concepts from cartographic design, map generalization, cognitive science, human-computer interaction, and pattern detection. Rather than offering an exhaustive review (each of these fields being extensive on its own), the thesis provides a focused synthesis aimed to provide a useful entry point for researchers or designers interested in the role of landmarks in pan-scalar map design. It identifies the key ideas and questions that support future research at the intersection of user cognition and map

generalization and highlights the limitations of current approaches, which often favor visual simplification over cognitive coherence.

From this foundation, the thesis proceeds through a series of targeted explorations, each contributing in its own way to a better understanding of how users perceive maps, and encouraging reflection on the pattern detection techniques currently used—or that could be adapted—for this purpose.

❖ **Understanding what constitutes a pan-scalar landmark in contemporary maps**

In the early stages of the thesis, we conducted exploratory studies to better define what kinds of features could be considered as pan-scalar landmarks in current topographic and digital maps. These studies revealed that landmarks are highly contextual and depend not only on feature type (e.g., roads, buildings, parks) but also on their visual contrast, functional importance, and spatial configuration. These findings laid the groundwork for a more flexible and perception-driven definition of landmarks.

❖ **Studying landmark density and cognitive load**

Another study focused on the density of landmarks and the cognitive implications of displaying too many—or too few—landmarks at a given scale. This led to reflections on the optimal balance between maintaining orientation and avoiding visual clutter. We observed that not all types of visual cues provide sufficient information for effective orientation on their own, but that multiple cues, when combined, are often enough to support users in staying oriented.

❖ **Analyzing the evolution of digital map interfaces**

We also examined the evolution of digital map interfaces and how their design impacts user behavior. This investigation showed that while modern interfaces offer flexibility and interactivity, they also introduce challenges by altering the visual consistency of map reading. The ability to zoom and pan freely can result in frequent loss of landmarks, especially when anchor continuity is not explicitly preserved.

❖ **Case study on ring roads as anchors**

A dedicated study explored the case of ring roads, whose distinctive geometry and functional role make them strong candidates for cognitive anchors. Through a perceptual survey and ontology development, we proposed a structured model of ringroadness based on various indicators (closure, contrast, centrality, etc.). This study provided a clear example of how user-based observations can inform formal detection strategies for specific types of anchors.

❖ **Remarkable buildings and multimodal detection**

Another experiment focused on the detection of remarkable buildings, combining attribute-based classification with the use of map images to provide spatial and contextual information. This served as an early foray into multimodal modelling, in which tabular and visual data are combined to improve landmark detection. While our implementation did not successfully improve classification results, the literature from other fields suggests that such an approach remains promising.

❖ Proposal and testing of AHAP segmentation models

In the final part of the thesis, we introduced the concept of Areas of High Anchoring Potential (AHAPs) and proposed a general framework for detecting them using deep learning. We built a dataset annotated by human experts, implemented evaluation metrics, and trained two models (U-Net and Segformer). While performance varied and some limitations emerged, the results demonstrated the feasibility of treating anchor detection as a segmentation problem—and of learning it from image data alone.

Lessons learned and discussions

Clearly, the goal of this thesis was not to achieve state-of-the-art performance in terms of modeling results. Rather, the aim was to open new research avenues and conduct initial, modest experiments to assess whether it is possible to encode geographic information in ways that reflect human perception, and to explore how machine learning might support cognitively-informed cartographic design. Several important insights emerged from our studies and experiments, which are worth highlighting here.

❖ (Pan-scalar) landmarks are rarely isolated features

Their perceived anchoring power often depends on how they relate to their surrounding elements, their contrast with the background, or their role in the larger spatial organization of the map. This was particularly evident in the ring road study, where roads gained or lost salience depending on their relation to a central urban area or to reinforcing features like rivers and railways. This supports a more Gestalt-informed view of landmark perception, where proximity, continuity, and closure shape the user's mental map. Although this characteristic of landmarks has been discussed extensively in the literature, it was particularly striking to observe its persistence across all of our experiments.

❖ (Pan-scalar) landmarks are often fuzzy rather than crisp

They may lack well-defined boundaries (as with city centers or large parks), or be perceived as a whole even when physically disjointed (as with broken ring roads). This challenges traditional approaches in cartographic modeling and calls for more flexible definitions that allow for partial, probabilistic, or context-dependent interpretations. This conceptual ambiguity is actually one of the main reasons we chose to explore the idea of detecting Areas of High Anchoring Potential, which arguably captures the nuanced nature of anchorness in cartographic imagery.

❖ The anchorness of pan-scalar landmarks depends on their behavior across scales

Through our investigations, it became clear that some landmarks were losing their salience when generalized, while others features, not that salient, only became cognitively relevant through their persistence and recognizability across multiple views. This observation reinforces the notion that pan-scalar landmarks must be understood (or detected) not solely based on their intrinsic properties, but also in terms

of how they transition between scales. Unfortunately, this aspect could not be fully explored within the scope of this thesis, but it remains a crucial avenue for future research.

❖ **Machine learning holds potential for modelling cognitive aspects of mapping**

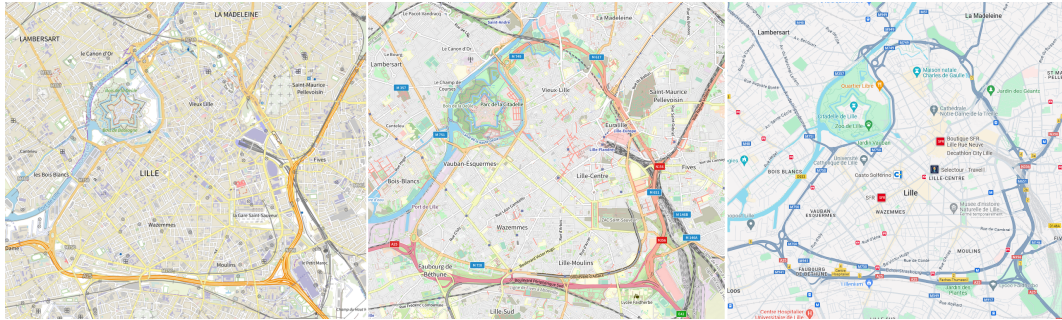
One lesson from this thesis is the difficulty of keeping pace with the rapid evolution of machine learning. The landscape of available models, architectures, and training strategies changes constantly, making it challenging to identify stable or definitive approaches for tasks as complex and nuanced as pan-scalar landmarks detection. Despite this, the progress in artificial intelligence also offers the ability to learn from fuzzy, ambiguous, or context-dependent characteristics that are central to how humans perceive landmarks and construct mental maps. While current models may not yet fully replicate human cognitive processes, their capacity to approximate certain aspects of perception and abstraction suggests that AI could play a key role in cartographic automation.

❖ **Machines do not think the way humans do**

Although explainable AI is still an emerging field, current research suggests that strong performance on tasks such as detecting areas of anchoring potential does not imply that models are learning or reasoning about spatial relations in a cognitively plausible way. In particular, Biscione and Bowers (2023) shows that Gestalt principles tend to emerge only in the later layers of deep learning models, if at all. This contrasts with human cognition, where Gestalt principles like proximity, continuity, and closure are believed to operate from the earliest stages of visual processing. Such findings highlight a fundamental disconnect between model performance and human-like reasoning, and they reinforce the need for caution when interpreting machine-learned patterns as cognitively meaningful.

❖ **(Pan-scalar) landmark design is not entirely neutral**

While this thesis has focused on cognitive aspects of map design, such as the optimization of user orientation and fluidity across scales, it is important to acknowledge that maps are also shaped by political, commercial, and institutional agendas. This is not a novel insight emerging from our work, but rather a well-established understanding in critical cartography that we believe is worth reiterating here. Features that serve navigation and cognitive anchoring may coincide with what is deemed important by the map provider, but this alignment is not guaranteed, as illustrated in Figure F.10 where we can see sponsored content obscuring the train station, which is an important landmark of the eastern part of the city of Lille. As researchers aiming to improve map usability, we must remain aware that landmarks themselves can carry ideological weight depending on how they are selected and represented. For example, privileging certain elements such as road networks risks reinforcing existing transportation norms, instead of supporting alternative or more sustainable design choices.



❁ Figure F.10 : Representation of Lille on different maps. From left to right: IGN Plan, OpenStreetMap, Google Maps.

Future research

Extending the experiments from this thesis

Several threads from this thesis could be pursued further in a direct and incremental manner:

❖ Modeling ring road characteristics

A direction would be to implement and evaluate the indicators proposed in the ontology of ring roads developed in Chapter C. By operationalizing these indicators in computational models, we could assess their capacity to support ring road detection at scale. This would help validate the ontology and provide practical tools for integration into cartographic pipelines.

❖ Expanding multimodal approaches to remarkable building detection

Our initial experiment combining attribute data with map images could be extended using alternative visual inputs such as satellite imagery or street-level photographs. These image sources could offer richer contextual cues to help distinguishing remarkable buildings.

❖ Differentiating types of AHAPs

With a larger and more diverse annotated dataset, it would become possible to train models on different types of AHAPs. It would open the door to investigating whether different AHAP types interact to form composite anchors—configurations in which visual or functional complementarities reinforce the overall anchoring effect. Coupled with annotator interviews, this could lead to a more fine-grained understanding of anchoring potential across map elements.

New perspectives on pan-scalar landmark design

Other research paths we identified throughout our journey include:

❖ **Textual data and knowledge extraction**

While this thesis has focused on vector data and image data, a significant portion of geographic information in today's world exists in textual form. An increasing number of studies explore how geographic entities and spatial relationships can be extracted from unstructured or semi-structured text, sometimes assisted by large language models (LLMs) (Medad, Gaio, Moncla, Mustiere, and Nir 2020; Hu and Adams 2021). These efforts could lead to the identification of strong landmarks that could be enhanced as pan-scalar landmarks.

❖ **Large multimodal models and map image understanding**

Recent breakthroughs in large multimodal models show immense promise for geographic analysis. As of late 2023, LLMs such as GPT-4V were beginning to interpret map images, identifying key elements from legends and recognizing structures directly on the map (Xu and Tao 2024). At the time, their performance relied heavily on carefully crafted prompts and remained limited in tasks involving precise spatial reasoning or quantitative analysis. However, only a year later, these models appear to have made substantial progress. Their ability to describe and interpret even complex pan-scalar map images has improved significantly, raising the question of whether such models could assist cartographers in detecting pan-scalar anchors or identifying breaks in continuity across zoom levels. These capabilities could maybe one day support the semi-automated design of cognitively coherent multi-scale maps, especially in large or poorly documented areas.

❖ **Last but not least: ethical and environmental considerations**

The technological possibilities mentioned above must be weighed against their costs as the environmental impact of training and deploying large-scale models is increasingly a concern. While much of the excitement around AI focuses on performance and capabilities, comparatively little attention has been given (until recently) to the sustainability and ethics of applying such models in various domains, including cartography (Schwartz, Dodge, Smith, and Etzioni 2020). If multimodal models are to be used in cartography, their deployment should be preceded by an assessment of their ethical, social, and ecological implications. How much automation is truly necessary or beneficial? Are large models justifiable for tasks that could be addressed through more lightweight approaches or human–AI collaboration? What values—social, ecological, epistemological—are embedded in the systems we choose to develop and deploy? These questions open an important research direction that bridges technical innovation with responsible design, and calls for interdisciplinary collaboration between AI researchers, cartographers, ethicists, and environmental scientists.

The End

Glossary

- Accuracy** p.126
- Anchor point** p.38
- AHAP** p.151
- Cartographic communication** p.61
- Cartographic generalisation** p.57
- Cartographic language** p.61
- Clutter** p.75
- Cognitive map** p.37
- Critical zones** p.155
- Cross-validation** p.128
- Desert fog** p.49
- Disorientation** p.37
- (Data) enrichment** p.64
- Ensemble learning** p.123
- Fuzzy map navigation task** p.82
- Landmark** p.40
- Mental map** p.37
- Multimodal learning** p.123
- Ontology** p.111
- (Generalisation) operator** p.57
- Pan-scalar map** p.42
- Pattern** p.62
- Precise map navigation task** p.82
- Recall** p.126
- Remarkable building** p.124
- Ringroadness** p.108
- Saliency** p.38
- (Image) segmentation** p.171
- Static phase** p.89
- Structure** p.62
- Transformer** p.173
- Transition phase** p.89
- U-net** p.172
- View** p.49
- Visual saliency** p.154

Bibliography

- Anderson B. A.** (Mar. 2021). “Time to Stop Calling it Attentional “Capture” and Embrace a Mechanistic Understanding of Attentional Priority”. In: *Visual cognition* 29.9, p. 537. DOI: 10.1080/13506285.2021.1892894.
- Been K., Nöllenburg M., Poon S.-H., and Wolff A.** (Apr. 2010). “Optimizing active ranges for consistent dynamic map labeling”. In: *Computational Geometry*. Special Issue on 24th Annual Symposium on Computational Geometry (SoCG’08) 43.3, pp. 312–328. ISSN: 0925-7721. DOI: 10.1016/j.comgeo.2009.03.006.
- Beitlova M., Popelka S., Konopka M., and Macku K.** (Jan. 2023). “Verification of Cartographic Communication Models Using Detection of Map Reading Strategies Based on Eye Movement Recording”. In: *The Cartographic Journal* 60.1. Publisher: Taylor & Francis, pp. 5–24. ISSN: 0008-7041. DOI: 10.1080/00087041.2022.2150362.
- Berli J., Touya G., Gruget M., and Potié Q.** (Oct. 2024). “MapDraw. Un outil libre d’annotation de cartes en ligne”. In: *Mappemonde. Revue trimestrielle sur l’image géographique et les formes du territoire* 138. Number: 138 Publisher: UMR Espace. ISSN: 0764-3470. DOI: 10.4000/12jjo.
- Biljecki F. and Chow Y. S.** (July 2022). “Global Building Morphology Indicators”. In: *Computers, Environment and Urban Systems* 95, p. 101809. ISSN: 0198-9715. DOI: 10.1016/j.compenvurbsys.2022.101809.
- Biscione V. and Bowers J. S.** (Sept. 2023). “Mixed Evidence for Gestalt Grouping in Deep Neural Networks”. In: *Computational Brain & Behavior* 6.3, pp. 438–456. ISSN: 2522-087X. DOI: 10.1007/s42113-023-00169-2.
- Board C. and Taylor R. M.** (1977). “Perception and Maps: Human Factors in Map Design and Interpretation”. In: *Transactions of the Institute of British Geographers* 2.1. Publisher: [Royal Geographical Society (with the Institute of British Geographers), Wiley], pp. 19–36. ISSN: 0020-2754. DOI: 10.2307/622191.
- Boér A., Coltekin A., and Clarke K.** (Aug. 2013). “An evaluation of web-based geovisualizations for different levels of abstraction and realism – what do users predict?” In: DOI: 10.5167/uzh-87810.
- Borji A., Sihite D. N., and Itti L.** (Oct. 2013). “What stands out in a scene? A study of human explicit saliency judgment”. In: *Vision Research* 91, pp. 62–77. ISSN: 0042-6989. DOI: 10.1016/j.visres.2013.07.016.
- Brassel K. E. and Weibel R.** (Jan. 1988). “A review and conceptual framework of automated map generalization”. In: *International journal of geographical information systems* 2.3, pp. 229–244. ISSN: 0269-3798. DOI: 10.1080/02693798808927898.
- Brewer C. and Buttenfield B.** (Jan. 2007). “Framing Guidelines for Multi-Scale Map Design Using Databases at Multiple Resolutions”. In: *Cartography and Geographic Information Science* 34, pp. 3–15. DOI: 10.1559/152304007780279078.
- Brock A. and Jouffrais C.** (Nov. 2015). “Interactive audio-tactile maps for visually impaired people”. In: *SIGACCESS Access. Comput.* 113, pp. 3–12. ISSN: 1558-2337. DOI: 10.1145/2850440.2850441.
- Bruns C. R. and Chamberlain B. C.** (Sept. 2019). “The influence of landmarks and urban form on cognitive maps using virtual reality”. In: *Landscape and Urban Planning* 189, pp. 296–306. ISSN: 0169-2046. DOI: 10.1016/j.landurbplan.2019.05.006.

- Bunch R. L. and Lloyd R. E.** (May 2006). “The Cognitive Load of Geographic Information”. In: *The Professional Geographer* 58.2. Publisher: Routledge, pp. 209–220. ISSN: 0033-0124. DOI: 10.1111/j.1467-9272.2006.00527.x.
- Burghardt D., Duchêne C., and Mackaness W.** (2014). “Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation”. In: *Lecture Notes in Geoinformation and Cartography*. DOI: 10.1007/978-3-319-00203-3.
- Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., and Zagoruyko S.** (May 2020). *End-to-End Object Detection with Transformers*. arXiv:2005.12872 [cs]. DOI: 10.48550/arXiv.2005.12872.
- Carral D., Scheider S., Janowicz K., Vardeman C., Krisnadhi A. A., and Hitzler P.** (2013). “An Ontology Design Pattern for Cartographic Map Scaling”. In: *The Semantic Web: Semantics and Big Data*. Ed. by P. Cimiano, O. Corcho, V. Presutti, L. Hollink, and S. Rudolph. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 76–93. ISBN: 978-3-642-38288-8. DOI: 10.1007/978-3-642-38288-8_6.
- Carter J.** (Jan. 2005). “The many dimensions of map use”. In: *International Cartographic Conference*.
- Chamizo V. D., Manteiga R. D., Rodrigo T., and Mackintosh N. J.** (Jan. 2006). “Competition between landmarks in spatial learning: The role of proximity to the goal”. In: *Behavioural Processes* 71.1, pp. 59–65. ISSN: 0376-6357. DOI: 10.1016/j.beproc.2005.11.003.
- Cheng B., Wunderlich A., Gramann K., Lin E., and Fabrikant S. I.** (Nov. 2022). “The effect of landmark visualization in mobile maps on brain activity during navigation: A virtual reality study”. In: *Frontiers in Virtual Reality* 3. Publisher: Frontiers. ISSN: 2673-4192. DOI: 10.3389/frvir.2022.981625.
- Cheng B., Schwing A. G., and Kirillov A.** (Oct. 2021). *Per-Pixel Classification is Not All You Need for Semantic Segmentation*. arXiv:2107.06278 [cs]. DOI: 10.48550/arXiv.2107.06278.
- Christophe S. and Ruas A.** (2002). “Detecting building alignments for generalisation purposes”. In: *Symposium on Geospatial Theory, Processing and Applications*.
- Cockburn A., Karlson A., and Bederson B. B.** (Jan. 2009). “A review of overview+detail, zooming, and focus+context interfaces”. In: *ACM Computing Surveys* 41.1, 2:1–2:31. ISSN: 0360-0300. DOI: 10.1145/1456650.1456652.
- Cohen J.** (Apr. 1960). “A Coefficient of Agreement for Nominal Scales”. In: *Educational and Psychological Measurement* 20.1. Publisher: SAGE Publications Inc, pp. 37–46. ISSN: 0013-1644. DOI: 10.1177/001316446002000104.
- Coltekin A. and Clarke K.** (Jan. 2011). “Virtual globes or virtual geographical reality”. In: DOI: 10.5167/uzh-59620.
- Cornia M., Baraldi L., Serra G., and Cucchiara R.** (Oct. 2018). “Predicting Human Eye Fixations via an LSTM-Based Saliency Attentive Model”. In: *IEEE Transactions on Image Processing* 27.10. Conference Name: IEEE Transactions on Image Processing, pp. 5142–5154. ISSN: 1941-0042. DOI: 10.1109/TIP.2018.2851672.
- Couclelis H., Golledge R. G., Gale N., and Tobler W.** (June 1987). “Exploring the anchor-point hypothesis of spatial cognition”. In: *Journal of Environmental Psychology* 7.2, pp. 99–122. ISSN: 0272-4944. DOI: 10.1016/S0272-4944(87)80020-8.
- Couclelis H.** (Nov. 2010). “Ontologies of geographic information”. In: *International Journal of Geographical Information Science* 24.12. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13658816.2010.484392>, pp. 1785–1809. ISSN: 1365-8816. DOI: 10.1080/13658816.2010.484392.
- Courtial A. and Touya G.** (2023). “Does Generalisation Matters in Pan-Scalar Maps? (Short Paper)”. In: *LIPICs, Volume 277, GIScience 2023*. Vol. 277. Artwork Size: 6 pages, 1932578 bytes ISSN: 1868-8969 Medium: application/pdf. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 23:1–23:6. ISBN: 978-3-95977-288-4. DOI: 10.4230/LIPICs.GISCIENCE.2023.23.

- Courtial A. and Touya G.** (Sept. 2024). “Strategies for pan-scalar map generalisation”. In: *Abstracts of the ICA* 7, pp. 1–2. DOI: 10.5194/ica-abs-7-25-2024.
- Courtial A., Touya G., and Zhang X.** (June 2022). “Representing Vector Geographic Information As a Tensor for Deep Learning Based Map Generalisation”. In: *AGILE: GIScience Series* 3. Publisher: Copernicus GmbH, pp. 1–8. DOI: 10.5194/agile-giss-3-32-2022.
- Coutrot A. et al.** (Sept. 2018). “Global Determinants of Navigation Ability”. In: *Current Biology* 28.17, 2861–2866.e4. ISSN: 0960-9822. DOI: 10.1016/j.cub.2018.06.009.
- Czerkauer-Yamu C. and Voigt A.** (2011). “Strategic Planning and Design with Space Syntax”. In: *eCAADe*, p. 9.
- Dickmann F., Edler D., Bestgen A.-K., and Kuchinke L.** (July 2017). “Exploiting Illusory Grid Lines for Object-Location Memory Performance in Urban Topographic Maps”. In: *The Cartographic Journal* 54.3. Publisher: Taylor & Francis, pp. 242–253. ISSN: 0008-7041. DOI: 10.1080/00087041.2016.1236509.
- Dong X., Yu Z., Cao W., Shi Y., and Ma Q.** (Apr. 2020). “A survey on ensemble learning”. In: *Frontiers of Computer Science* 14.2, pp. 241–258. ISSN: 2095-2236. DOI: 10.1007/s11704-019-8208-z.
- Dosovitskiy A. et al.** (June 2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. arXiv:2010.11929 [cs]. DOI: 10.48550/arXiv.2010.11929.
- Duchêne C., Bard S., Barillot X., Ruas A., Trévisan J., and Holzapfel F.** (Jan. 2003). “Quantitative and qualitative description of building orientation”. In.
- Duckham M., Winter S., and Robinson M.** (Mar. 2010). “Including landmarks in routing instructions”. In: *Journal of Location Based Services*. Publisher: Taylor & Francis Group. DOI: 10.1080/17489721003785602.
- Eastman R.** (Mar. 2013). “The Perception of Scale Change in Small-Scale Map Series”. In: *The American Cartographer* 8, pp. 5–21.
- Ekim B., Sertel E., and Kabadayı M. E.** (Aug. 2021). “Automatic Road Extraction from Historical Maps Using Deep Learning Techniques: A Regional Case Study of Turkey in a German World War II Map”. In: *ISPRS International Journal of Geo-Information* 10.8. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, p. 492. ISSN: 2220-9964. DOI: 10.3390/ijgi10080492.
- Elias B.** (2003). “Extracting Landmarks with Data Mining Methods”. In: *Spatial Information Theory. Foundations of Geographic Information Science*. Ed. by W. Kuhn, M. F. Worboys, and S. Timpf. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 375–389. ISBN: 978-3-540-39923-0. DOI: 10.1007/978-3-540-39923-0_25.
- Foerster T., Stoter J., and Köbben B.** (Jan. 2007). “Towards a formal classification of Generalization operators”. In.
- Fu C., Zhou Z., Xin Y., and Weibel R.** (June 2024). “Reasoning cartographic knowledge in deep learning-based map generalization with explainable AI”. In: *International Journal of Geographical Information Science* 0.0. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13658816.2024.2369535>, pp. 1–22. ISSN: 1365-8816. DOI: 10.1080/13658816.2024.2369535.
- Gedicke S. and Haunert J.-H.** (2023). “An Empirical Study on Interfaces for Presenting Large Sets of Point Features in Mobile Maps”. In: *The Cartographic Journal* 0.0. Publisher: Taylor & Francis, pp. 1–18. ISSN: 0008-7041. DOI: 10.1080/00087041.2023.2182354.
- Géron A.** (2022). “Chapter 16: Natural Language Processing with Rnns and Attention”. In: *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems*. 3rd ed. O’Reilly.
- Golledge R., Jacobson D., Kitchin R., and Blades M.** (Dec. 2000). “Cognitive Maps, Spatial Abilities, and Human Wayfinding”. In: *GEOGRAPHICAL REVIEW OF JAPAN SERIES B* 73, pp. 93–104. DOI: 10.4157/grj1984b.73.93.
- Gould P. and White R.** (1974). *Mental maps*. Harmondsworth, Penguin. ISBN: 978-0-14-021688-2.

- Gruber T. R.** (Nov. 1995). “Toward principles for the design of ontologies used for knowledge sharing?” In: *International Journal of Human-Computer Studies* 43.5, pp. 907–928. ISSN: 1071-5819. DOI: 10.1006/ijhc.1995.1081.
- Gruget M., Touya G., and Muehlenhaus I.** (Jan. 2023). “Missing the city for buildings? A critical review of pan-scalar map generalization and design in contemporary zoomable maps”. In: *International Journal of Cartography* 0.0. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/23729333.2022.2153467>, pp. 1–31. ISSN: 2372-9333. DOI: 10.1080/23729333.2022.2153467.
- Gruget M., Touya G., and Muehlenhaus I.** (2024). “Exploring the presence and persistence of cartographic anchors in pan-scalar maps”. In: *Spatial Cognition & Computation* 0.0. Publisher: Taylor & Francis, pp. 1–33. ISSN: 1387-5868. DOI: 10.1080/13875868.2024.2383419.
- Gruget M., Touya G., Potié Q., and Muehlenhaus I.** (May 2024). “A methodological inquiry for anchoring pan-scalar maps”. In: *AGILE: GIScience Series* 5. Publisher: Copernicus GmbH, pp. 1–6. DOI: 10.5194/agile-giss-5-27-2024.
- Haklay M., Singleton A., and Parker C.** (2008). “Web Mapping 2.0: The Neogeography of the GeoWeb”. In: *Geography Compass* 2.6, pp. 2011–2039. ISSN: 1749-8198. DOI: 10.1111/j.1749-8198.2008.00167.x.
- Harrie L., Touya G., Oucheikh R., Ai T., Courtial A., and Richter K.-F.** (Feb. 2024). “Machine learning in cartography”. In: *Cartography and Geographic Information Science* 51, pp. 1–19. DOI: 10.1080/15230406.2023.2295948.
- Harrie L. and Weibel R.** (2013). “Modelling the Overall Process of Generalisation”. In: **Harrower M. and Sheesley B.** (2005). “Designing Better Map Interfaces: A Framework for Panning and Zooming”. In: *Transactions in GIS* 9.2, pp. 77–89. ISSN: 1467-9671. DOI: 10.1111/j.1467-9671.2005.00207.x.
- He X., Zhang X., and Xin Q.** (Feb. 2018). “Recognition of building group patterns in topographic maps based on graph partitioning and random forest”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 136, pp. 26–40. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2017.12.001.
- He Y., Ai T., Yu W., and Zhang X.** (Mar. 2017). “A linear tessellation model to identify spatial pattern in urban street networks”. In: *International Journal of Geographical Information Science* 31, pp. 1–21. DOI: 10.1080/13658816.2017.1298768.
- Heinzle F. and Anders K.-H.** (Jan. 2007). “Chapter 12 - Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks”. In: *Generalisation of Geographic Information*. Ed. by W. A. Mackaness, A. Ruas, and L. T. Sarjakoski. International Cartographic Association. Amsterdam: Elsevier Science B.V., pp. 233–253. ISBN: 978-0-08-045374-3. DOI: 10.1016/B978-008045374-3/50014-4.
- Heinzle F., Anders K.-H., and Sester M.** (2006). “Pattern Recognition in Road Networks on the Example of Circular Road Detection”. In: *Geographic Information Science*. Ed. by M. Raubal, H. J. Miller, A. U. Frank, and M. F. Goodchild. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 153–167. ISBN: 978-3-540-44528-9. DOI: 10.1007/11863939_11.
- Henderson P., Hu J., Romoff J., Brunskill E., Jurafsky D., and Pineau J.** (Nov. 2022). *Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning*. arXiv:2002.05651 [cs]. DOI: 10.48550/arXiv.2002.05651.
- Hillier B. and Hanson J.** (1984). *The social logic of space*. Cambridge [Cambridgeshire] ; New York : Cambridge University Press. ISBN: 978-0-521-23365-1 978-0-521-36784-4.
- Hou Y. and Biljecki F.** (Dec. 2022). “A comprehensive framework for evaluating the quality of street view imagery”. In: *International Journal of Applied Earth Observation and Geoinformation* 115, p. 103094. ISSN: 1569-8432. DOI: 10.1016/j.jag.2022.103094.
- Hu Y. and Adams B.** (2021). “Harvesting Big Geospatial Data from Natural Language Texts”. In: *Handbook of Big Geospatial Data*. Ed. by M. Werner and Y.-Y. Chiang.

- Cham: Springer International Publishing, pp. 487–507. ISBN: 978-3-030-55462-0. DOI: 10.1007/978-3-030-55462-0_19.
- Hurst J., Degbelo A., and Kray C.** (May 2024). “FreeMapRetrieve: Freehand Gestures for Retrieve Operations in Large-Screen Map Environments”. In: *AGILE: GIScience Series* 5. Publisher: Copernicus GmbH, pp. 1–13. DOI: 10.5194/agile-giss-5-7-2024.
- Jang S., Battulga L., and Nasridinov A.** (2020). “Detection of Dangerous Situations using Deep Learning Model with Relational Inference”. In: *Journal of Multimedia Information System* 7.3. Publisher: Korea Multimedia Society, pp. 205–214. ISSN: 2383-7632. DOI: 10.33851/JMIS.2020.7.3.205.
- João E. M.** (Nov. 1999). “Causes and Consequences of Map Generalisation”. In: *The Geographical Journal* 165.3, p. 326. ISSN: 00167398. DOI: 10.2307/3060451.
- Jul S. and Furnas G. W.** (Nov. 1998). “Critical zones in desert fog: aids to multiscale navigation”. In: *Proceedings of the 11th annual ACM symposium on User interface software and technology*. UIST '98. New York, NY, USA: Association for Computing Machinery, pp. 97–106. ISBN: 978-1-58113-034-8. DOI: 10.1145/288392.288578.
- Kang M., Liu Y., Wang M., Li L., and Weng M.** (Mar. 2022). “A random forest classifier with cost-sensitive learning to extract urban landmarks from an imbalanced dataset”. In: *International Journal of Geographical Information Science* 36.3. Publisher: Taylor & Francis, pp. 496–513. ISSN: 1365-8816. DOI: 10.1080/13658816.2021.1977814.
- Kapaj A., Lin E., and Lanini-Maggi S.** (2022). “The Effect of Abstract vs. Realistic 3D Visualization on Landmark and Route Knowledge Acquisition”. In: *15th International Conference on Spatial Information Theory (COSIT 2022)*. Ed. by T. Ishikawa, S. I. Fabrikant, and S. Winter. Vol. 240. Leibniz International Proceedings in Informatics (LIPIcs). ISSN: 1868-8969. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 15:1–15:8. ISBN: 978-3-95977-257-0. DOI: 10.4230/LIPIcs.COSIT.2022.15.
- Keil J., Edler D., Kuchinke L., and Dickmann F.** (Mar. 2020). “Effects of visual map complexity on the attentional processing of landmarks”. In: *PLOS ONE* 15.3. Publisher: Public Library of Science, e0229575. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0229575.
- Khan B. A. and Jung J.-W.** (Jan. 2024). “Semantic Segmentation of Aerial Imagery Using U-Net with Self-Attention and Separable Convolutions”. In: *Applied Sciences* 14.9. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 3712. ISSN: 2076-3417. DOI: 10.3390/app14093712.
- Khellaf L., Schlicht I. B., Bayer J., Bouwmeester R., Miraf T., and Wagner T.** (Nov. 2023). *Spot: A Natural Language Interface for Geospatial Searches in OSM*. arXiv:2311.08093 [cs]. DOI: 10.48550/arXiv.2311.08093.
- Koláčny A.** (June 1969). “Cartographic Information—a Fundamental Concept and Term in Modern Cartography”. In: *The Cartographic Journal* 6.1. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1179/caj.1969.6.1.47>, pp. 47–49. ISSN: 0008-7041. DOI: 10.1179/caj.1969.6.1.47.
- Kosch T., Karolus J., Zagermann J., Reiterer H., Schmidt A., and Woźniak P. W.** (July 2023). “A Survey on Measuring Cognitive Workload in Human-Computer Interaction”. In: *ACM Computing Surveys* 55.13s, 283:1–283:39. ISSN: 0360-0300. DOI: 10.1145/3582272.
- Kucklick J.-P. and Müller O.** (May 2021). *A Comparison of Multi-View Learning Strategies for Satellite Image-Based Real Estate Appraisal*. arXiv:2105.04984 [cs]. DOI: 10.48550/arXiv.2105.04984.
- Kümmerer M., Theis L., and Bethge M.** (Nov. 2014). “Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet”. In.
- Lafon E., Potié Q., and Touya G.** (Sept. 2023). “Salient building detection using multimodal deep learning”. In: *CartoAI: AI for cartography*. Leeds, United Kingdom.

- Lam N. and Quattrochi D.** (Feb. 1992). “On the Issues of Scale, Resolution, and Fractal Analysis in the Mapping Sciences*”. In: *The Professional Geographer* 44, pp. 88–98. DOI: 10.1111/j.0033-0124.1992.00088.x.
- Landis J. R. and Koch G. G.** (Mar. 1977). “The measurement of observer agreement for categorical data”. In: *Biometrics* 33.1, pp. 159–174. ISSN: 0006-341X.
- LeCun Y., Bengio Y., and Hinton G.** (May 2015). “Deep learning”. In: *Nature* 521.7553. Publisher: Nature Publishing Group, pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539.
- Ledermann F.** (May 2023). “Minimum dimensions for cartographic symbology – history, rationale and relevance in the digital age”. In: *International Journal of Cartography*. Publisher: Taylor & Francis. ISSN: 2372-9333.
- Li Y., Branco P., and Zhang H.** (Jan. 2023). “Imbalanced Multimodal Attention-Based System for Multiclass House Price Prediction”. In: *Mathematics* 11.1. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 113. ISSN: 2227-7390. DOI: 10.3390/math11010113.
- Li Z., W Y., Ai T., and Chen J.** (July 2004). “Automated building generalization based on urban morphology and Gestalt theory”. In: *International Journal of Geographical Information Science* 18, pp. 513–534. DOI: 10.1080/13658810410001702021.
- Liang W. et al.** (Apr. 2024). *Mapping the Increasing Use of LLMs in Scientific Papers*. arXiv:2404.01268 [cs]. DOI: 10.48550/arXiv.2404.01268.
- Lichtner W.** (Jan. 1979). “Computer-assisted processes of cartographic generalization in topographic maps”. In: *Geo-Processing* 1, pp. 183–199.
- Lin Y., Cai Y., Gong Y., Kang M., and Li L.** (Dec. 2019). “Extracting urban landmarks from geographical datasets using a random forests classifier”. In: *International Journal of Geographical Information Science*. Publisher: Taylor & Francis. ISSN: 1365-8816.
- Lokka I. E. and Çöltekin A.** (Feb. 2019). “Toward optimizing the design of virtual environments for route learning: empirically assessing the effects of changing levels of realism on memory”. In: *International Journal of Digital Earth* 12.2. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/17538947.2017.1349842>, pp. 137–155. ISSN: 1753-8947. DOI: 10.1080/17538947.2017.1349842.
- Lüscher P., Weibel R., and Mackaness W. A.** (2008). “Where is the Terraced House? On the Use of Ontologies for Recognition of Urban Concepts in Cartographic Databases”. In: *Headway in Spatial Data Handling*. Ed. by A. Ruas and C. Gold. Lecture Notes in Geoinformation and Cartography. Berlin, Heidelberg: Springer, pp. 449–466. ISBN: 978-3-540-68566-1. DOI: 10.1007/978-3-540-68566-1_26.
- Lynch K.** (1960). *The Image of the City* | *Semantic Scholar*.
- MacDonald E., Jacoby D., and Coady Y.** (Sept. 2024). *VistaFormer: Scalable Vision Transformers for Satellite Image Time Series Segmentation*. DOI: 10.48550/arXiv.2409.08461.
- MacEachren A. and Ganter J.** (Oct. 1990). “A Pattern Identification Approach to Cartographic Visualization”. In: *Cartographica: The International Journal for Geographic Information and Geovisualization* 27, pp. 64–81. DOI: 10.3138/M226-1337-2387-3007.
- Mackaness W., Burghardt D., and Duchêne C.** (2014). “Map Generalisation: Fundamental to the Modelling and Understanding of Geographic Space”. In: *Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation*. Ed. by D. Burghardt, C. Duchêne, and W. Mackaness. Cham: Springer International Publishing, pp. 1–15. ISBN: 978-3-319-00203-3. DOI: 10.1007/978-3-319-00203-3_1.
- Mackaness W., Ruas A., and Sarjakoski L.** (Jan. 2007). *Generalisation of Geographic Information: Cartographic Modelling and Applications*. ISBN: 978-0-08-045374-3. DOI: 10.1016/B978-0-08-045374-3.X5000-5.
- Mackaness W. A. and Edwards G.** (July 2002). “The Importance of Modelling Pattern and Structure in Automated Map Generalisation”. In: *Proceedings of the Joint ISPRS/ICA Workshop on Multi-Scale Representations of Spatial Data*, pp. 7–8.

- Medad A., Gaio M., Moncla L., Mustiere S., and Nir Y.** (July 2020). “Comparing supervised learning algorithms for Spatial Nominal Entity recognition”. In: *AGILE: GIScience Series* 1, pp. 1–18. DOI: 10.5194/agile-giss-1-15-2020.
- Miller G. A.** (1956). “The magical number seven, plus or minus two: Some limits on our capacity for processing information”. In: *Psychological Review* 63.2. Place: US Publisher: American Psychological Association, pp. 81–97. ISSN: 1939-1471. DOI: 10.1037/h0043158.
- Montello D.** (Dec. 2001). “Scale in Geography”. In: *International Encyclopedia of the Social and Behavioral Sciences*, pp. 13501–13504. DOI: 10.1016/B0-08-043076-7/02473-6.
- Montello D. R.** (Jan. 2002). “Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches”. In: *Cartography and Geographic Information Science* 29.3. Publisher: Taylor & Francis, pp. 283–304. ISSN: 1523-0406. DOI: 10.1559/152304002782008503.
- Morisset B. and Ruas A.** (June 1997). “Simulation and Agent Modelling for Road Selection in Generalisation”. In: *Proceedings of the 18th International Cartographic Conference Stockholm, Sweden, 23–27 June 1997*, pp. 1376–1380.
- Mustière S.** (Jan. 2001). “Apprentissage supervise pour la generalisation cartographique”. These de doctorat. Paris 6.
- Nes A.** (Apr. 2021). “The Impact of the Ring Roads on the Location Pattern of Shops in Town and City Centres. A Space Syntax Approach”. In: *Sustainability* 13, p. 3927. DOI: 10.3390/su13073927.
- Norman D.** (Jan. 1991). “Cognitive Artifacts”. In: *Designing interaction: psychology at the human-computer interface*. Journal Abbreviation: Designing interaction: psychology at the human-computer interface, p. 333. ISBN: 978-0-521-40056-5.
- Nuhn E. and Timpf S.** (2017). “Personal Dimensions of Landmarks”. In: *Societal Geoinnovation*. Ed. by A. Bregt, T. Sarjakoski, R. van Lammeren, and F. Rip. Cham: Springer International Publishing, pp. 129–143. ISBN: 978-3-319-56759-4. DOI: 10.1007/978-3-319-56759-4_8.
- Oktay O. et al.** (May 2018). *Attention U-Net: Learning Where to Look for the Pancreas*. arXiv:1804.03999 [cs]. DOI: 10.48550/arXiv.1804.03999.
- Pan J., Ferrer C. C., McGuinness K., O’Connor N. E., Torres J., Sayrol E., and Giro-i-Nieto X.** (July 2018). *SalGAN: Visual Saliency Prediction with Generative Adversarial Networks*. arXiv:1701.01081 [cs]. DOI: 10.48550/arXiv.1701.01081.
- Potié Q., Mackaness W. A., and Touya G.** (June 2022). “When is a Ring Road a ‘Ring Road’? A Brief Perceptual Study”. In: vol. 3. Copernicus Publications, p. 54. DOI: 10.5194/agile-giss-3-54-2022.
- Potié Q., Touya G., Beladraoui C., El-Moutaouakkil A., and Mackaness W. A.** (Sept. 2022). “Deep learning for anchor detection in multi-scale maps”. In: *EuroCarto 2022*. Ed. by G. Gartner, A. Binn, and O. Ignateva. Vol. 5. Abstracts of the ICA. Vienna, Austria: Copernicus Publications, p. 82. DOI: 10.5194/ica-abs-5-82-2022.
- Potié Q., Touya G., and Mackaness W. A.** (Sept. 2023). “Experiments in the automatic segmentation of anchors using deep learning techniques”. In: *CartoAI: AI for cartography*. Leeds, United Kingdom.
- Potié Q., Touya G., Wenclik L., and Mackaness W.** (Sept. 2024). “Browsing Map Browsers”. In: *Abstracts of the ICA* 7, pp. 1–2. DOI: 10.5194/ica-abs-7-130-2024.
- Qin T., Fias W., Van de Weghe N., and Huang H.** (2024). “Differentiating Everyday Map Tasks: Unique Attention-Related Eye Movements and Electrophysiological Signatures of Map Use”. In: *Annals of the American Association of Geographers* 0.0. Publisher: Taylor & Francis, pp. 1–21. ISSN: 2469-4452. DOI: 10.1080/24694452.2024.2353845.
- Quesnot T. and Roche S.** (Mar. 2015). “Measure of Landmark Semantic Saliency through Geosocial Data Streams”. In: *ISPRS* 4.1. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, pp. 1–31. ISSN: 2220-9964. DOI: 10.3390/ijgi4010001.
- Rainsford D. and Mackaness W.** (2002). “Template Matching in Support of Generalisation of Rural Buildings”. In: *Advances in Spatial Data Handling*. Ed. by D. E. Richardson

- and P. van Oosterom. Berlin, Heidelberg: Springer, pp. 137–151. ISBN: 978-3-642-56094-1. DOI: 10.1007/978-3-642-56094-1_11.
- Ranganathan P., Pramesh C. S., and Aggarwal R.** (2017). “Common pitfalls in statistical analysis: Measures of agreement”. In: *Perspectives in Clinical Research* 8.4, pp. 187–191. ISSN: 2229-3485. DOI: 10.4103/picr.PICR_123_17.
- Rasse T., Gyselinck V., and Guegan J.** (Nov. 2023). “Encountering an emotional landmark on the route for a better spatial memory: What matters, valence or arousal?” In: *Journal of Environmental Psychology* 91, p. 102145. ISSN: 0272-4944. DOI: 10.1016/j.jenvp.2023.102145.
- Ravishankar H., Venkataramani R., Thiruvankadam S., Sudhakar P., and Vaidya V.** (2017). “Learning and Incorporating Shape Models for Semantic Segmentation”. In: *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*. Ed. by M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne. Vol. 10433. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 203–211. ISBN: 978-3-319-66181-0 978-3-319-66182-7. DOI: 10.1007/978-3-319-66182-7_24.
- Regnault N., Touya G., Gould N., and Foerster T.** (Feb. 2014). *Process Modelling, Web Services and Geoprocessing*.
- Ren Y., Yu Y., and Guan H.** (Jan. 2020). “DA-CapsUNet: A Dual-Attention Capsule U-Net for Road Extraction from Remote Sensing Imagery”. In: *Remote Sensing* 12.18. Number: 18 Publisher: Multidisciplinary Digital Publishing Institute, p. 2866. ISSN: 2072-4292. DOI: 10.3390/rs12182866.
- Richter K.-F. and Winter S.** (Jan. 2014). *Landmarks: GIScience for intelligent services*. ISBN: 978-3-319-05731-6. DOI: 10.1007/978-3-319-05732-3.
- Robinson A. H.** (1952). *The look of maps; an examination of cartographic design*. Madison, University of Wisconsin Press.
- Rodgers D.** (Dec. 2001). “A grammar for zooming interfaces: Using interaction design strategies to improve user’s navigation and spatial awareness”. In: *Information Design Journal* 10, pp. 250–257. DOI: 10.1075/idj.10.3.06rog.
- Ronneberger O., Fischer P., and Brox T.** (May 2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. arXiv:1505.04597. DOI: 10.48550/arXiv.1505.04597.
- Rosenfelder M.** (July 2020). *Multi-Input Deep Neural Networks with PyTorch-Lightning - Combine Image and Tabular Data*.
- Rosenholtz R., Li Y., and Nakano L.** (Aug. 2007). “Measuring visual clutter”. In: *Journal of Vision* 7.2, p. 17. ISSN: 1534-7362. DOI: 10.1167/7.2.17.
- Roth R. E.** (Dec. 2013). “An Empirically-Derived Taxonomy of Interaction Primitives for Interactive Cartography and Geovisualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.12. Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 2356–2365. ISSN: 1941-0506. DOI: 10.1109/TVCG.2013.130.
- Roth R. E., Brewer C. A., and Stryker M. S.** (Mar. 2011). “A typology of operators for maintaining legible map designs at multiple scales”. In: *Cartographic Perspectives* 68. Number: 68, pp. 29–64. ISSN: 1048-9053. DOI: 10.14714/CP68.7.
- Rousell A. and Zipf A.** (Mar. 2017). “Towards a Landmark-Based Pedestrian Navigation Service Using OSM Data”. In: *ISPRS International Journal of Geo-Information* 6.3. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 64. ISSN: 2220-9964. DOI: 10.3390/ijgi6030064.
- Sagi O. and Rokach L.** (July 2018). “Ensemble learning: A survey”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.4. ISSN: 1942-4787. DOI: 10.1002/widm.1249.
- Savino S., Rumor M., Zanon M., and Lissandron I.** (2010). “Data enrichment for road generalization through analysis of morphology in the CARGEN project”. In: *Proceedings of 13th ICA Workshop on Generalisation and Multiple Representation*. Zurich, Switzerland.

- Scheider S. and Possin J.** (2012). “Affordance-based individuation of junctions in Open Street Map”. In: *Journal of Spatial Information Science* 2012.4, pp. 31–56. ISSN: 1948-660X (online) DOI:10.5311. DOI: 10.5311/JOSIS.2012.4.54.
- Schwartz R., Dodge J., Smith N. A., and Etzioni O.** (Nov. 2020). “Green AP”. In: *Commun. ACM* 63.12, pp. 54–63. ISSN: 0001-0782. DOI: 10.1145/3381831.
- Schwering A.** (2019). “Wayfinding Through Orientation : Schematizing Landmark , Route and Survey Information in a Single Map”. In.
- Sertel E., Hucko C. M., and Kabadayı M. E.** (Dec. 2024). “Automatic Road Extraction from Historical Maps Using Transformer-Based SegFormers”. In: *ISPRS International Journal of Geo-Information* 13.12. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 464. ISSN: 2220-9964. DOI: 10.3390/ijgi13120464.
- Sester M.** (1995). *Lernen struktureller Modelle für die Bildanalyse*. Beck. ISBN: 978-3-7696-9484-0.
- Sester M., Feng Y., and Thiemann F.** (Sept. 2018). “BUILDING GENERALIZATION USING DEEP LEARNING”. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4*, pp. 565–572. DOI: 10.5194/isprs-archives-XLII-4-565-2018.
- Siegel A. W. and White S. H.** (Jan. 1975). “The Development of Spatial Representations of Large-Scale Environments”. In: *Advances in Child Development and Behavior*. Ed. by H. W. Reese. Vol. 10. JAI, pp. 9–55. DOI: 10.1016/S0065-2407(08)60007-5.
- Sorrows M. E. and Hirtle S. C.** (1999). “The Nature of Landmarks for Real and Electronic Spaces”. In: *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*. Ed. by C. Freksa and D. M. Mark. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 37–50. ISBN: 978-3-540-48384-7. DOI: 10.1007/3-540-48384-5_3.
- Stanislawski L. V., Kronenfeld B. J., Battenfeld B. P., and Shavers E. J.** (Sept. 2023). “At what scales does a river meander? Scale-specific sinuosity (S3) metric for quantifying stream meander size distribution”. In: *Geomorphology* 436, p. 108734. ISSN: 0169-555X. DOI: 10.1016/j.geomorph.2023.108734.
- Stea D., ed.** (1973). *Image and Environment: Cognitive Mapping and Spatial Behavior*. New York: Routledge. ISBN: 978-0-203-78915-5.
- Steck S. D. and Mallot H. A.** (Feb. 2000). “The Role of Global and Local Landmarks in Virtual Environment Navigation”. In: *Presence* 9.1. Conference Name: Presence, pp. 69–83. ISSN: 1054-7460. DOI: 10.1162/105474600566628.
- Stoter J., Meijers B. M., Oosterom P. V., Grunreich D., and Kraak M. J.** (Dec. 2010). “Applying DLM and DCM concepts in a multi-scale data environment”. In.
- Strudel R., Garcia R., Laptev I., and Schmid C.** (Sept. 2021). *Segmenter: Transformer for Semantic Segmentation*. arXiv:2105.05633 [cs]. DOI: 10.48550/arXiv.2105.05633.
- Suel E., Bhatt S., Brauer M., Flaxman S., and Ezzati M.** (May 2021). “Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas”. In: *Remote Sensing of Environment* 257, p. 112339. ISSN: 0034-4257. DOI: 10.1016/j.rse.2021.112339.
- Thomson R.** (Jan. 2006). “The ‘stroke’ Concept in Geographic Network Generalization and Analysis”. In: *Progress in Spatial Data Handling - 12th International Symposium on Spatial Data Handling, SDH 2006*. Journal Abbreviation: Progress in Spatial Data Handling - 12th International Symposium on Spatial Data Handling, SDH 2006, pp. 681–697. ISBN: 978-3-540-35588-5. DOI: 10.1007/3-540-35589-8_43.
- Thomson R. C. and Richardson D. E.** (1999). *The ‘Good Continuation’ Principle of Perceptual Organization applied to the Generalization of Road Networks*.
- Tliba M., Kerkouri M. A., Chetouani A., Bruno A., El Hassouni M., and Çöltekin A.** (June 2024). “Perceptual Evaluation of Masked AutoEncoder Emergent Properties Through Eye-Tracking-Based Policy”. In: *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*. ETRA '24. New York, NY, USA: Association

- for Computing Machinery, pp. 1–3. ISBN: 979-8-4007-0607-3. DOI: 10.1145/3649902.3655638.
- Tolman E. C.** (1948). “Cognitive maps in rats and men.” In: *Psychological Review* 55.4, pp. 189–208. ISSN: 1939-1471, 0033-295X. DOI: 10.1037/h0061626.
- Touya G.** (2010). “Enrichissement automatique de données par analyse spatiale pour la généralisation de réseaux”. In: *Revue Internationale de Géomatique* 20.2, p. 175. DOI: 10.3166/rig.20.175-199.
- Touya G.** (Nov. 2020a). “MapGenOnto: A Shared Ontology for Map Generalisation and Multi-Scale Visualisation”. In: *23rd ICA Workshop on Map Generalisation and Multiple Representation*. Delft, Netherlands: ICA.
- Touya G.** (2020b). “Please, Help Me! I Am Lost in Zoom”. In: p. 4.
- Touya G. and Girres J.-F.** (June 2013). “ScaleMaster 2.0: a ScaleMaster extension to monitor automatic multi-scales generalizations”. In: *Cartography and Geographic Information Science* 40.3. Publisher: Taylor & Francis, pp. 192–200. DOI: 10.1080/15230406.2013.809233.
- Touya G., Gruget M., and Muehlenhaus I.** (Feb. 2023). “Where Am I Now? Modelling Disorientation in Pan-Scalar Maps”. In: *ISPRS International Journal of Geo-Information* 12, p. 62. DOI: 10.3390/ijgi12020062.
- Touya G., Hoarau C., and Christophe S.** (Dec. 2016). “Clutter and Map Legibility in Automated Cartography: A Research Agenda”. In: *Cartographica* 51.4. Publisher: University of Toronto Press, pp. 198–207. ISSN: 0317-7173. DOI: 10.3138/cart.51.4.3132.
- Touya G. and Lokhat I.** (May 2020). “Deep Learning for Enrichment of Vector Spatial Databases: Application to Highway Interchange”. In: *ACM Transactions on Spatial Algorithms and Systems* 6.3, pp. 1–21. ISSN: 2374-0353, 2374-0361. DOI: 10.1145/3382080.
- Touya G., Potié Q., and Mackaness W. A.** (June 2023). “Incorporating ideas of structure and meaning in interactive multi scale mapping environments”. In: *International Journal of Cartography* 9.2. Publisher: Taylor & Francis, pp. 342–372. DOI: 10.1080/23729333.2023.2215960.
- Trell E.-M. and Hoven van B.** (Nov. 2010). “Making sense of place: exploring creative and (inter)active research methods with young people”. In: *Fennia : International Journal of Geography* 188.
- Truong Q. T., Touya G., and Runz C. d.** (Sept. 2020). “OSMWatchman: Learning How to Detect Vandalized Contributions in OSM Using a Random Forest Classifier”. In: *ISPRS International Journal of Geo-Information* 9.9. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 504. ISSN: 2220-9964. DOI: 10.3390/ijgi9090504.
- Van Wijk J. and Nuij W.** (2003). “Smooth and efficient zooming and panning”. In: *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No.03TH8714)*. Conference Name: IEEE Symposium on Information Visualization 2003 ISBN: 9780780381544 Place: Seattle, WA, USA Publisher: IEEE, pp. 15–23. DOI: 10.1109/INFVIS.2003.1249004.
- Viriyasaranon T., Woo S. M., and Choi J.-H.** (Apr. 2023). “Unsupervised Visual Representation Learning Based on Segmentation of Geometric Pseudo-Shapes for Transformer-Based Medical Tasks”. In: *IEEE Journal of Biomedical and Health Informatics* 27.4, pp. 2003–2014. ISSN: 2168-2208. DOI: 10.1109/JBHI.2023.3237596.
- Wabiński J., Mościcka A., and Touya G.** (July 2022). “Guidelines for Standardizing the Design of Tactile Maps: A Review of Research and Best Practice”. In: *The Cartographic Journal* 59.3. Publisher: Taylor & Francis, pp. 239–258. ISSN: 0008-7041. DOI: 10.1080/00087041.2022.2097760.
- Wagemans J., Feldman J., Gepshtein S., Kimchi R., Pomerantz J., Helm van der P., and Leeuwen C.** (July 2012). “A Century of Gestalt Psychology in Visual Perception: II. Conceptual and Theoretical Foundations”. In: *Psychological bulletin* 138, pp. 1218–52. DOI: 10.1037/a0029334.

- Wanga P.-Y., Chen C.-T., Su J.-W., Ting-Yun W., and Huang S.-H.** (Apr. 2021). “Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism”. In: *IEEE Access* PP, pp. 1–1. DOI: 10.1109/ACCESS.2021.3071306.
- Warfield S. K., Zou K. H., and Wells W. M.** (July 2004). “Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation”. In: *Ieee Transactions on Medical Imaging* 23.7, pp. 903–921. ISSN: 0278-0062. DOI: 10.1109/TMI.2004.828354.
- Wei Z., Xu W., Xiao Y., Shu M., Cheng L., Wang Y., and Liu C.** (Oct. 2023). “Enhancing building pattern recognition through multi-scale data and knowledge graph: a case study of C-shaped patterns”. In: *International Journal of Digital Earth* 16.1. Publisher: Taylor & Francis, pp. 3860–3881. ISSN: 1753-8947. DOI: 10.1080/17538947.2023.2259868.
- Weinman J., Chen Z., Gafford B., Gifford N., Lamsal A., and Niehus-Staab L.** (Sept. 2019). *Deep Neural Networks for Text Detection and Recognition in Historical Maps*. Pages: 909. DOI: 10.1109/ICDAR.2019.00149.
- Wenclik L. and Touya G.** (June 2023). “Where do people look at during multi-scale map tasks?” In: *AGILE 2023*. Ed. by P. v. Oosterom, H. Ploeger, A. Mansourian, S. Scheider, R. Lemmens, and B. v. Loenen. Vol. 4. Delft, Netherlands: Copernicus Publications, p. 51. DOI: 10.5194/agile-giss-4-51-2023.
- Wenclik L. and Touya G.** (May 2024). “EyeCatchingMaps, a Dataset to Assess Saliency Models on Maps”. In: *AGILE: GIScience Series* 5. Publisher: Copernicus GmbH, pp. 1–6. DOI: 10.5194/agile-giss-5-51-2024.
- Wu H.-Y., Takahashi S., Poon S.-H., and Arikawa M.** (2017). *Scale-Adaptive Placement of Hierarchical Map Labels*. The Eurographics Association. ISBN: 978-3-03868-043-7.
- Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J. M., and Luo P.** (Oct. 2021). *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*. arXiv:2105.15203 [cs]. DOI: 10.48550/arXiv.2105.15203.
- Xing R., Wu F., Gong X., Du J., and Liu C.** (Aug. 2022). “An axis-matching approach to combined collinear pattern recognition for urban building groups”. In: *Geocarto International* 37.16. Publisher: Taylor & Francis, pp. 4823–4842. ISSN: 1010-6049. DOI: 10.1080/10106049.2021.1899305.
- Xu J. and Tao R.** (Apr. 2024). “Map Reading and Analysis with GPT-4V(ision)”. In: *ISPRS International Journal of Geo-Information* 13.4. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 127. ISSN: 2220-9964. DOI: 10.3390/ijgi13040127.
- Yang F., Zamzmi G., Angara S., Rajaraman S., Aquilina A., Xue Z., Jaeger S., Papagiannakis E., and Antani S. K.** (2023). “Assessing Inter-Annotator Agreement for Medical Image Segmentation”. In: *IEEE access : practical innovations, open solutions* 11, pp. 21300–21312. ISSN: 2169-3536. DOI: 10.1109/access.2023.3249759.
- Yang W.** (2008). “Identify Building Patterns”. In: *ISPRS Commission II, WG II/3*. Beijing, China: ISPRS.
- Yaniv Z., Lowekamp B. C., Johnson H. J., and Beare R.** (June 2018). “SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research”. In: *Journal of Digital Imaging* 31.3, pp. 290–303. ISSN: 1618-727X. DOI: 10.1007/s10278-017-0037-8.
- Yesiltepe D., Conroy Dalton R., and Ozbil Torun A.** (Aug. 2021). “Landmarks in wayfinding: a review of the existing literature”. In: *Cognitive Processing* 22.3, pp. 369–410. ISSN: 1612-4790. DOI: 10.1007/s10339-021-01012-x.
- You M., Chen C.-W., Liu H., and Lin Y.-H.** (Apr. 2007). “A Usability Evaluation of Web Map Zoom and Pan Functions”. In: *International Journal of Design* 1, pp. 15–25.
- Yu J., Ai T., Xu H., Yan L., and Shen Y.** (June 2023). “Inconsistency Detection in Cross-Layer Tile Maps with Super-Pixel Segmentation”. In: *ISPRS International Jour-*

- nal of Geo-Information* 12.6. Number: 6 Publisher: Multidisciplinary Digital Publishing Institute, p. 244. ISSN: 2220-9964. DOI: 10.3390/ijgi12060244.
- Zagata K. and Medyńska-Gulij B.** (Sept. 2024). “Mini-map design as a cartographic element of the user interface - based on video games”. In: *Abstracts of the ICA 7*. Publisher: Copernicus GmbH, pp. 1–2. DOI: 10.5194/ica-abs-7-189-2024.
- Zeng L.** (Sept. 2009). “Designing the User Interface: Strategies for Effective Human-Computer Interaction (5th Edition) by B. Shneiderman and C. Plaisant: Pearson Addison-Wesley, 2009. xviii + 606 pages. ISBN: 978-0-321-53735-5”. In: *International Journal of Human-Computer Interaction* 25.7. Publisher: Taylor & Francis, pp. 707–708. ISSN: 1044-7318. DOI: 10.1080/10447310903187949.
- Zhang Q.** (2004). *Modeling Structure and Patterns in Road Network Generalization*.
- Zhang X., Touya G., and Meijers M.** (Mar. 2024). “Automated Map Generalization: Emerging Techniques and New Trends (Editorial)”. In: *Journal of Geovisualization and Spatial Analysis* 8. DOI: 10.1007/s41651-024-00174-4.
- Zhao Y., Chetty G., and Tran D.** (Dec. 2019). “Deep Learning with XGBoost for Real Estate Appraisal”. In: *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1396–1401. DOI: 10.1109/SSCI44817.2019.9002790.
- Zhou Z., Siddiquee M. M. R., Tajbakhsh N., and Liang J.** (Sept. 2018). “UNet++: A Nested U-Net Architecture for Medical Image Segmentation”. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S...* 11045, pp. 3–11. DOI: 10.1007/978-3-030-00889-5_1.

Publications

Potíé Q., Mackaness W. A., and Touya G. (June 2022). “When is a Ring Road a ‘Ring Road’? A Brief Perceptual Study”. en. In: vol. 3. Copernicus Publications, p. 54. DOI: 10.5194/agile-giss-3-54-2022.

Potíé Q., Touya G., Beladraoui C., El-Moutaouakkil A., and Mackaness W. A. (Sept. 2022). “Deep learning for anchor detection in multi-scale maps”. In: *EuroCarto 2022*. Ed. by G. Gartner, A. Binn, and O. Ignateva. Vol. 5. Abstracts of the ICA. Vienna, Austria: Copernicus Publications, p. 82. DOI: 10.5194/ica-abs-5-82-2022.

Touya G., Potíé Q., and Mackaness W. A. (June 2023). “Incorporating ideas of structure and meaning in interactive multi scale mapping environments”. In: *International Journal of Cartography* 9.2. Publisher: Taylor & Francis, pp. 342–372. DOI: 10.1080/23729333.2023.2215960.

Lafon E., Potíé Q., and Touya G. (Sept. 2023). “Salient building detection using multimodal deep learning”. In: *CartoAI: AI for cartography*. Leeds, United Kingdom.

Potíé Q., Touya G., and Mackaness W. A. (Sept. 2023). “Experiments in the automatic segmentation of anchors using deep learning techniques”. en. In: *CartoAI: AI for cartography*. Leeds, United Kingdom.

Potíé Q., Touya G., Wenclik L., and Mackaness W. (Sept. 2024). “Browsing Map Browsers”. In: *Abstracts of the ICA* 7, pp. 1–2. DOI: 10.5194/ica-abs-7-130-2024.

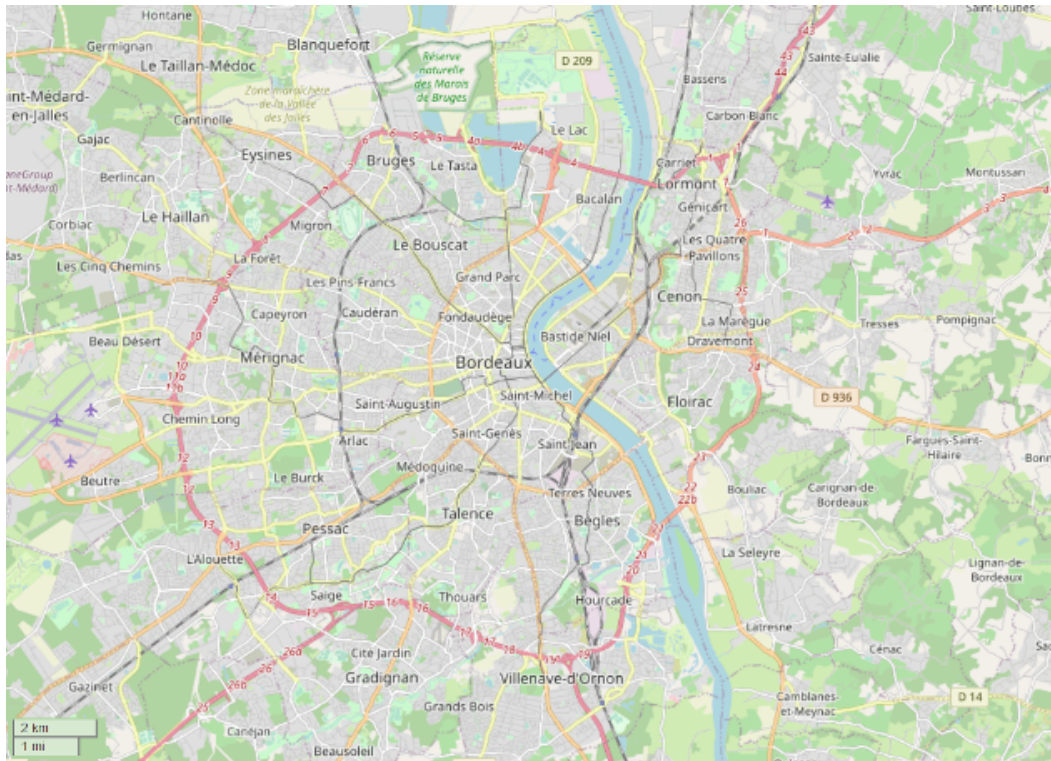
Gruget M., Touya G., Potíé Q., and Muehlenhaus I. (May 2024). “A methodological inquiry for anchoring pan-scalar maps”. English. In: *AGILE: GI-Science Series* 5. Publisher: Copernicus GmbH, pp. 1–6. DOI: 10.5194/agile-giss-5-27-2024.

Appendices

A Ring road survey part 1

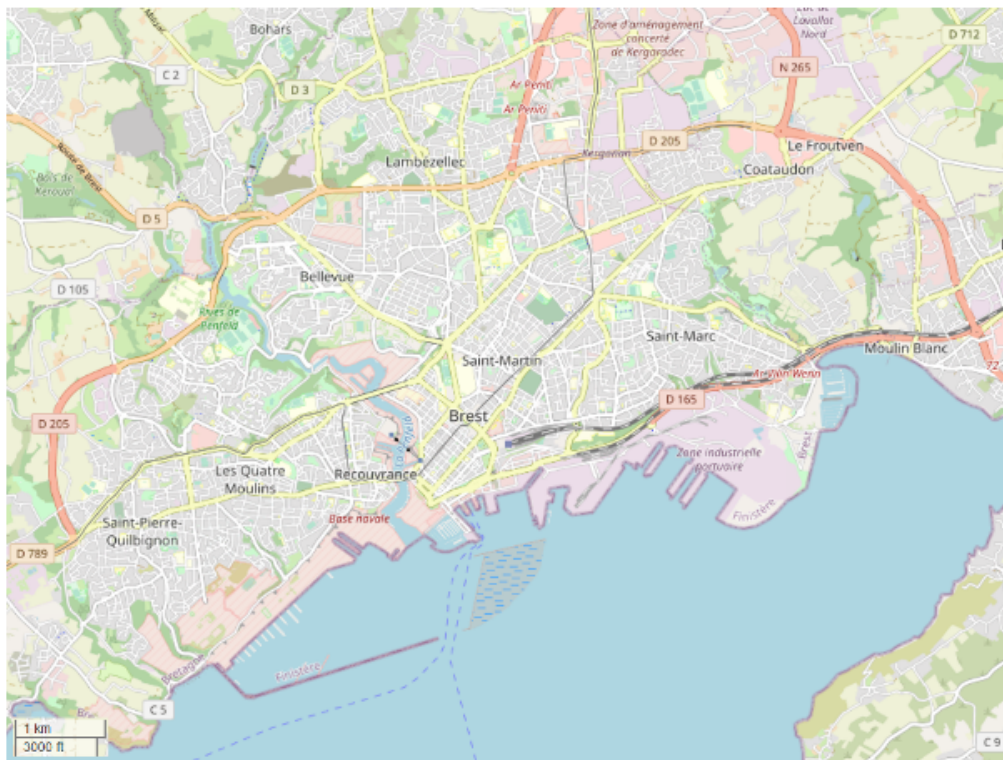
We present here the maps used in the first part of the survey on ring roads described in Section C.2, with their associated results:

- In red: "I can't see any ring road"
- In grey: "I think I distinguish a ring road"
- In blue: "I can clearly see at least one ring road".



28

1 - Bordeaux

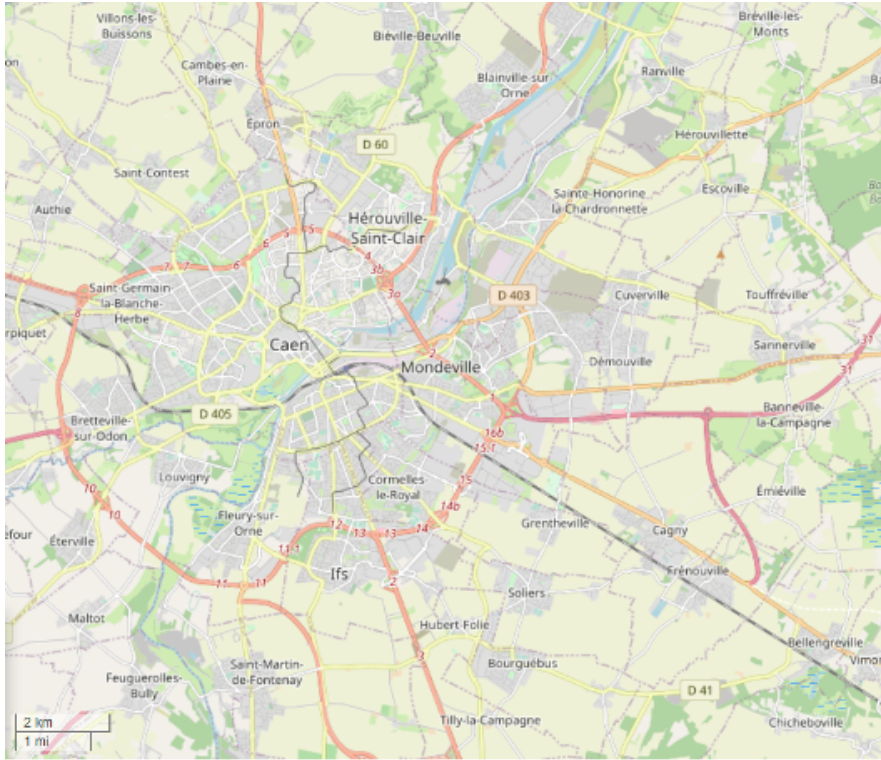


9

12

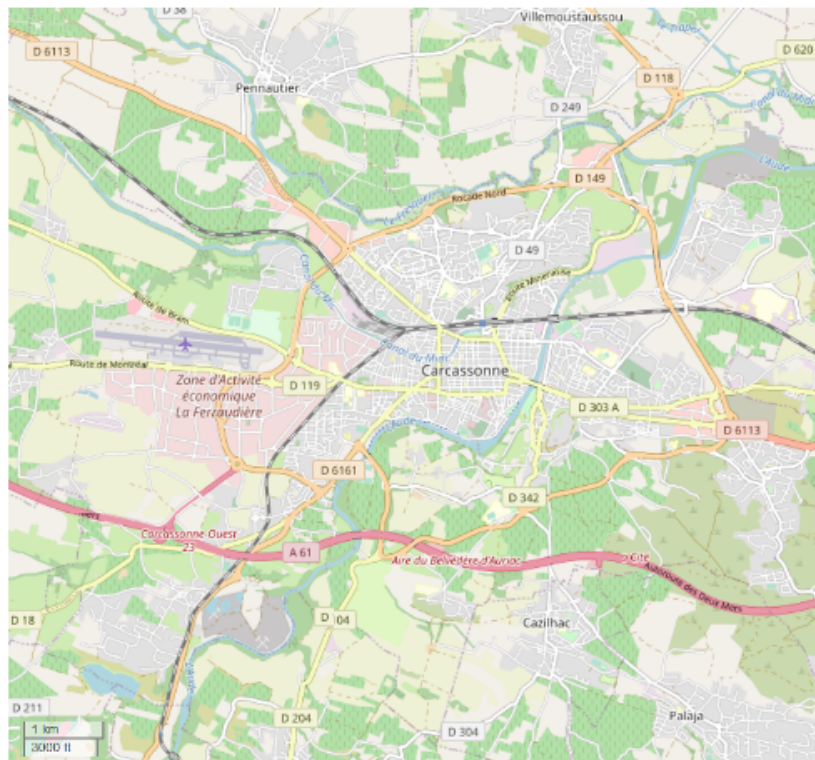
7

2 - Brest



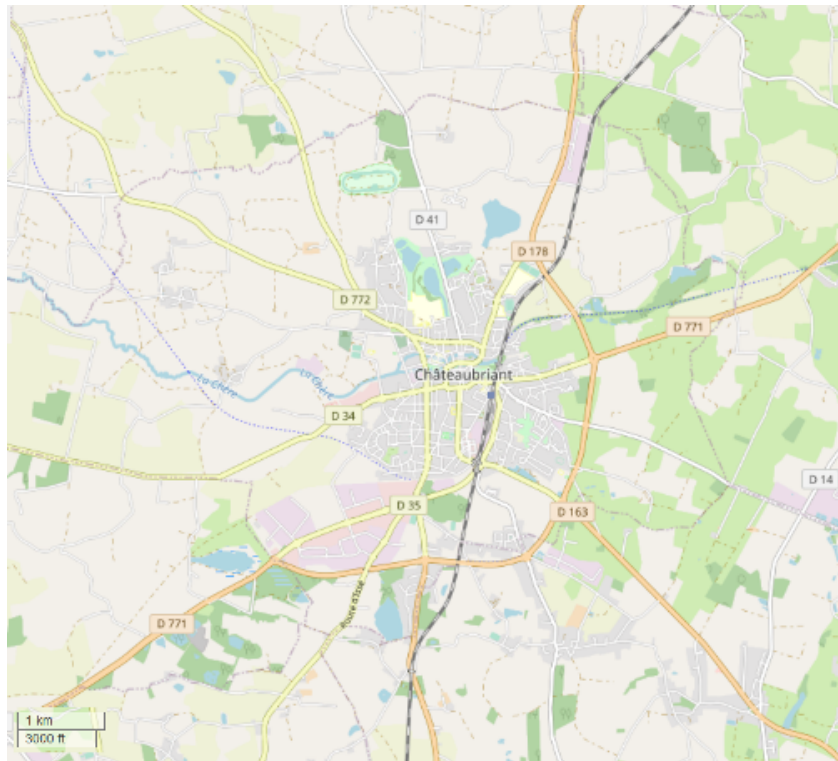
2 **26**

3 - Caen

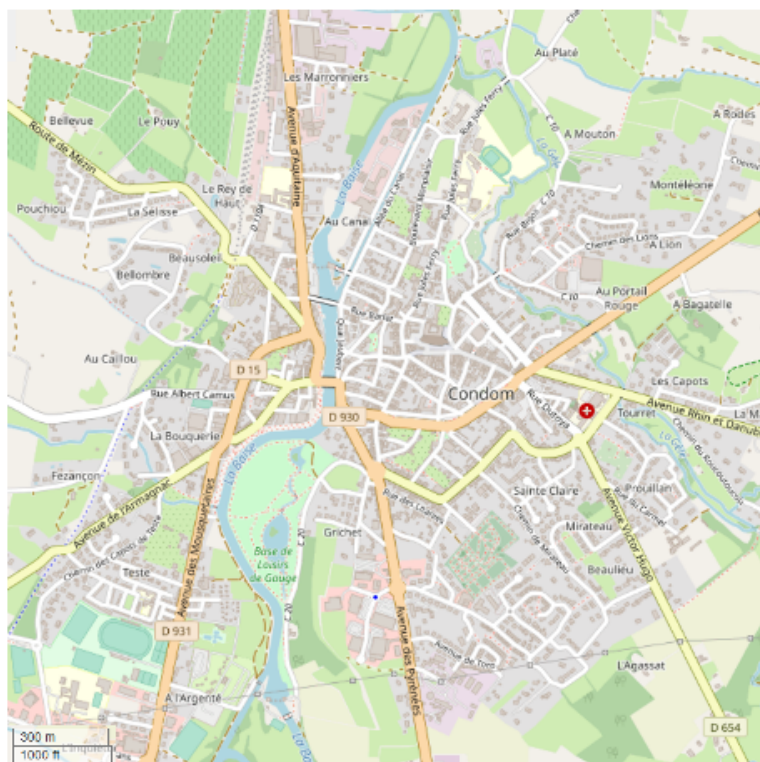


1 **5** **22**

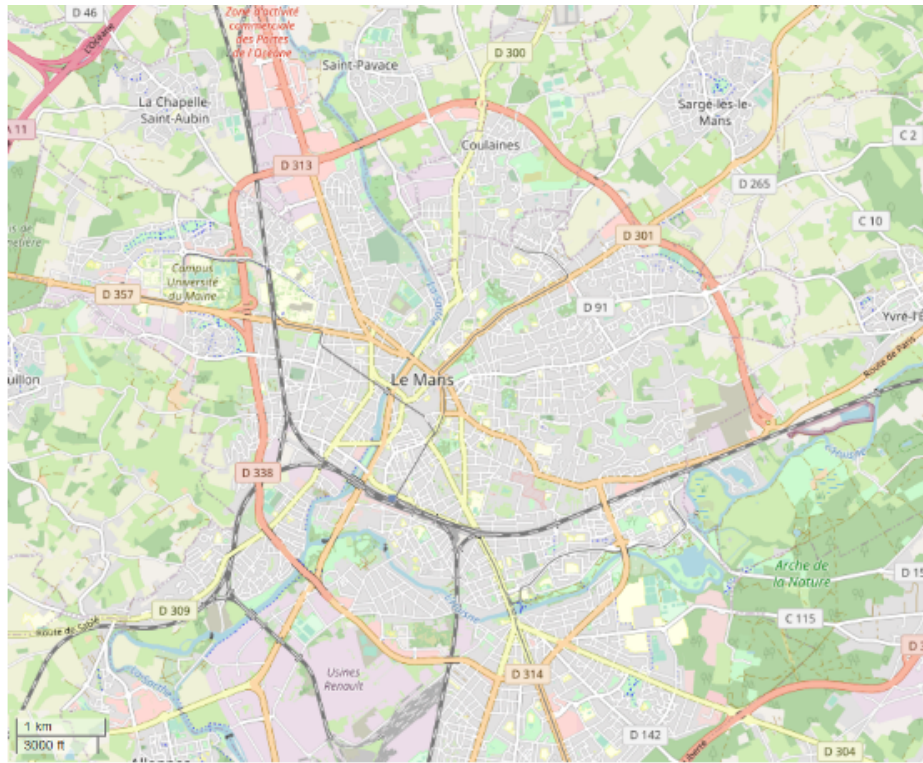
4 - Carcassonne



5 - Chateaubriant

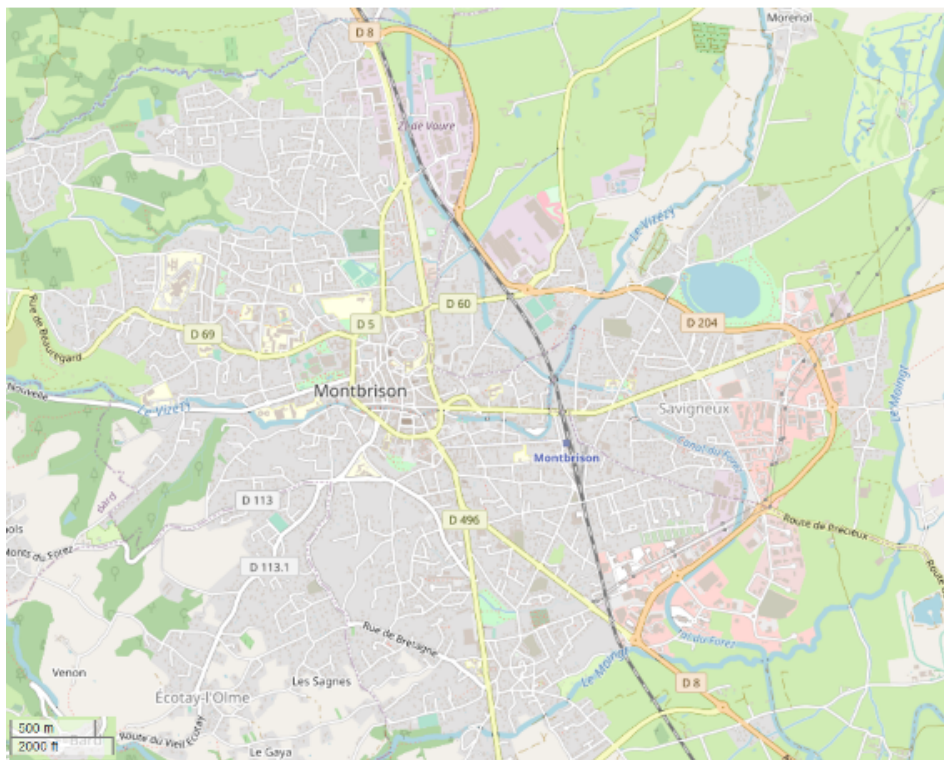


6 - Condom



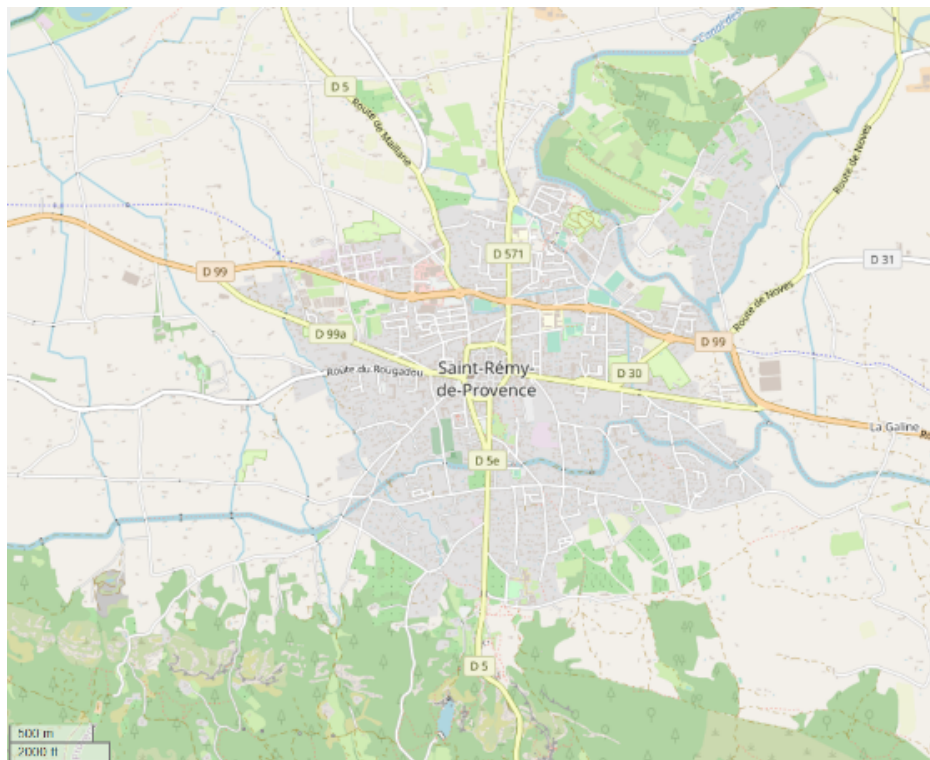
1 13 14

7 - Le Mans



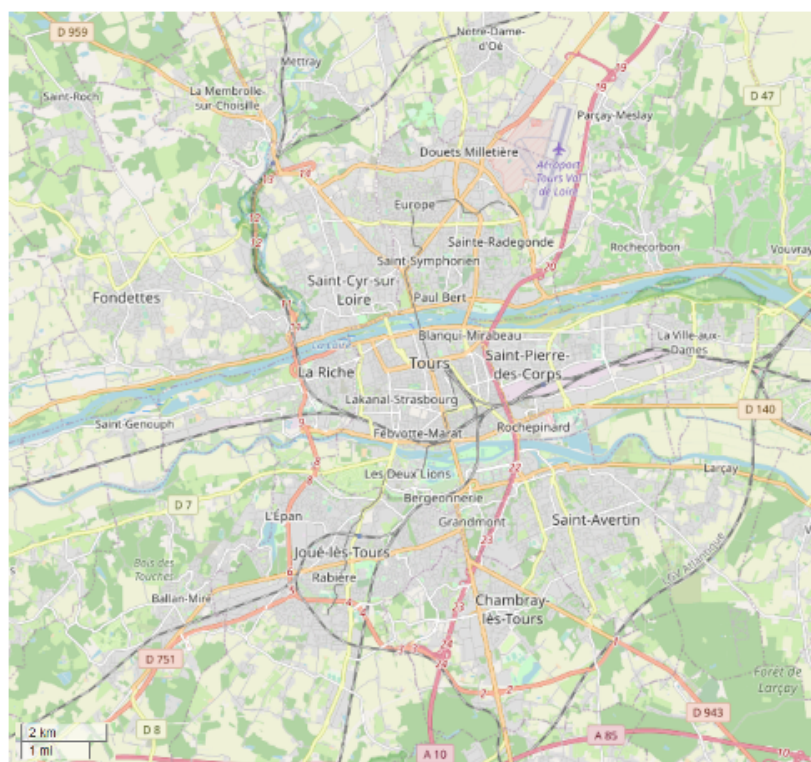
18 4 6

8 - Montbrison



21 4 3

9 - Saint-Rémy-De-Provence

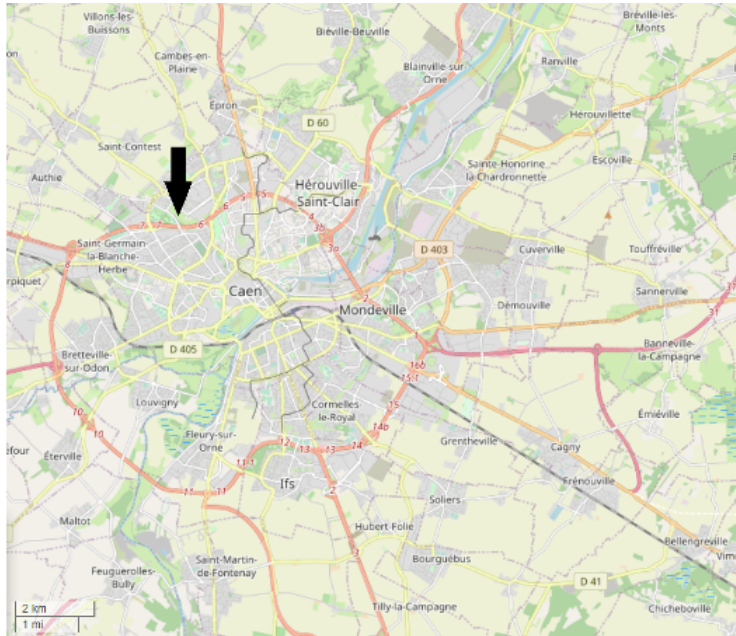


13 10 5

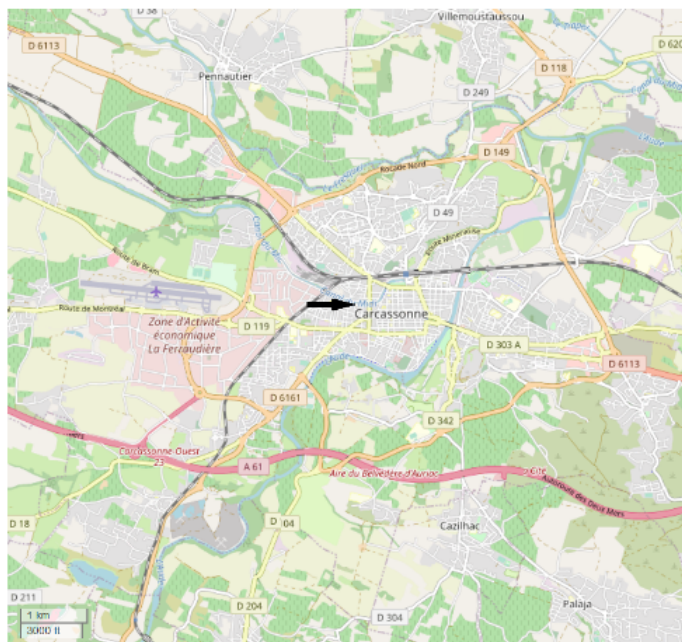
10 - Tours

B Ring road survey part 2

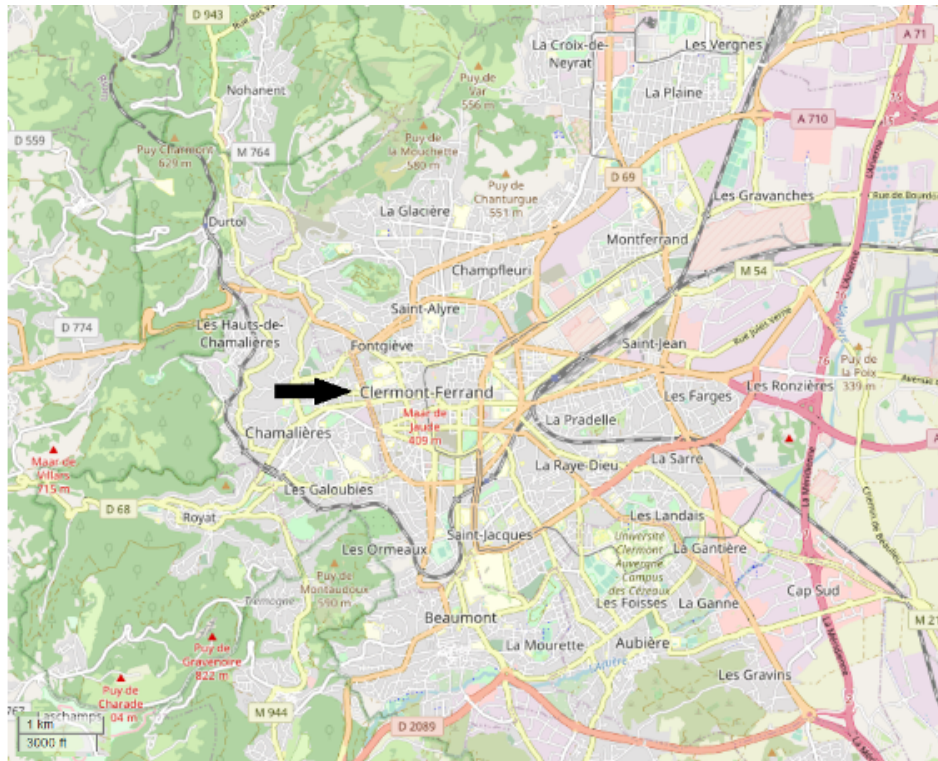
We present here the maps used in the second part of the survey on ring roads described in Section C.2, with their associated results. Every participant rated the pointed road from 0 to 5. We display the average rate and the standard deviation bellow each map.



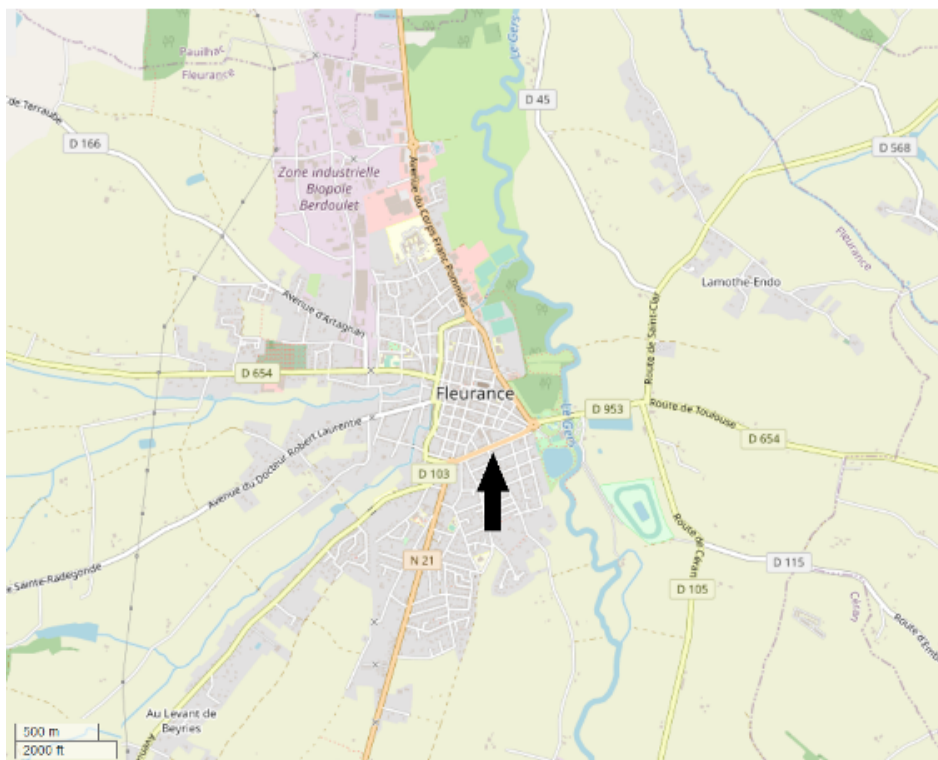
1 – Caen Av 4.75 Std 0.80



2 – Carcassonne Av 1.46 Std 1.64



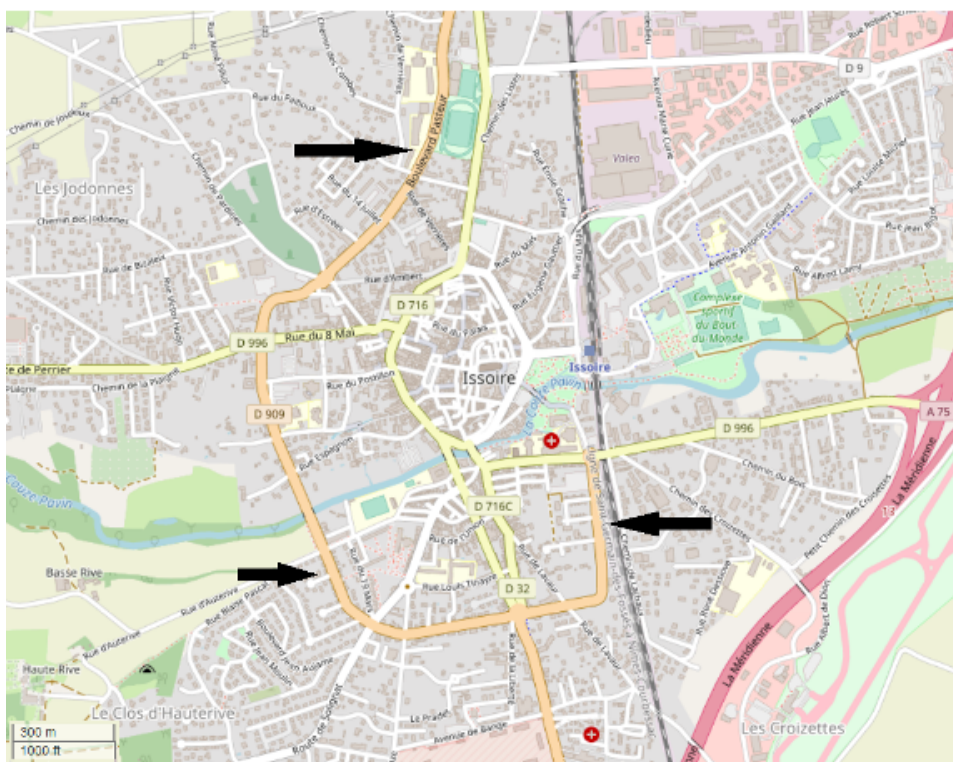
3 – Clermont-Ferrand Av 4.21 Std 1.10



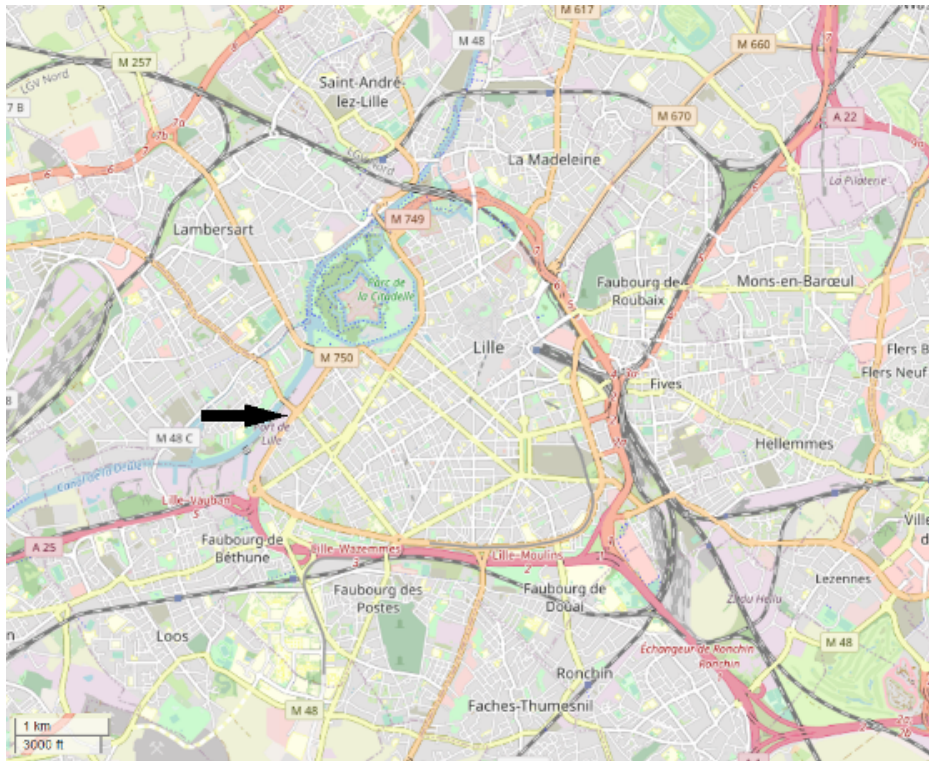
4 – Fleurance Av 2.29 Std 1.49



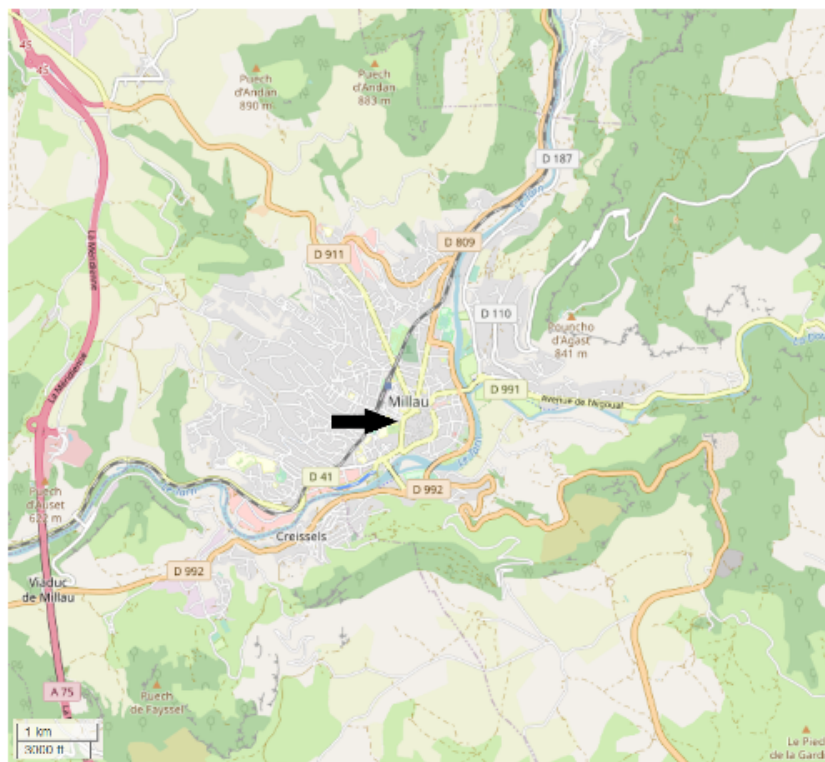
5 – Issoire (1) Av 1.71 Std 1.38



6 – Issoire (2) Av 2.04 Std 1.23



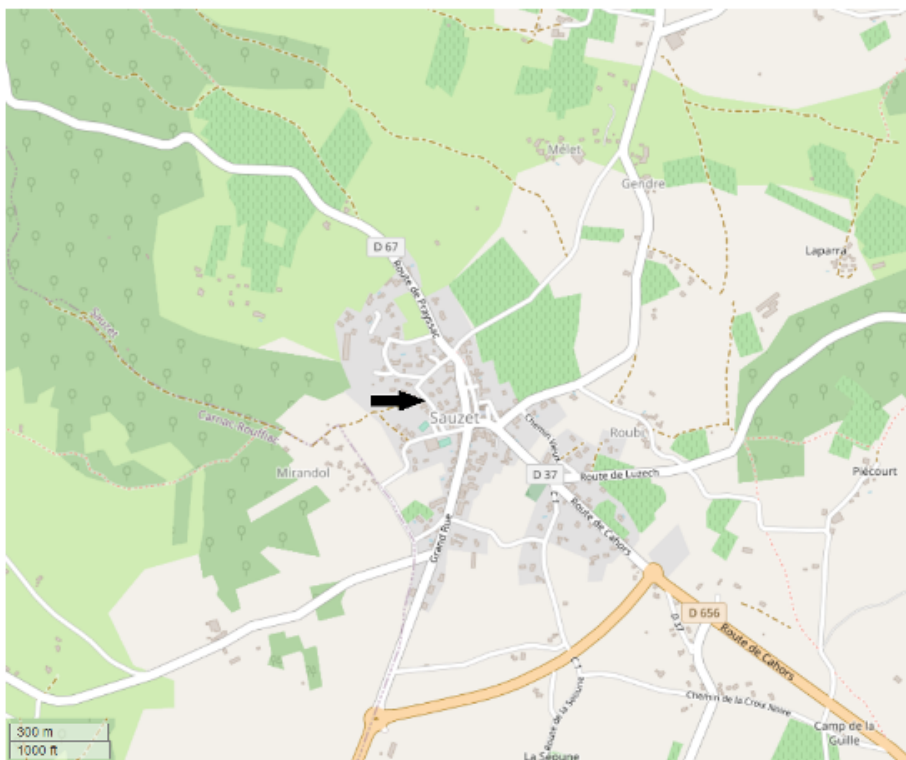
7 – Lille Av 3.29 Std 1.27



8 – Millau Av 1.57 Std 1.37



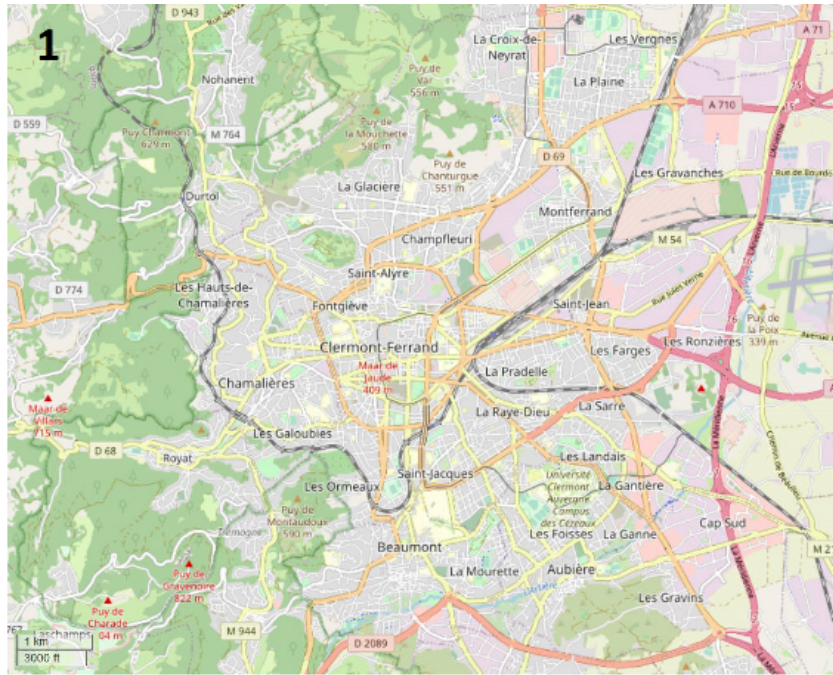
9 – Montcuq Av 1.86 Std 1.60



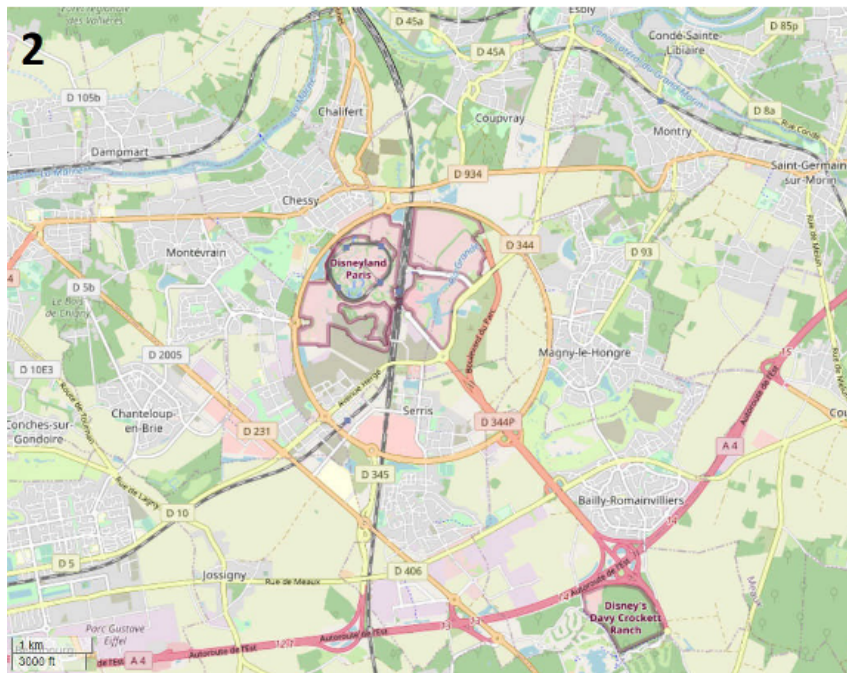
10 – Sauzet Av 0.68 Std 1.36

C Ring road survey part 3

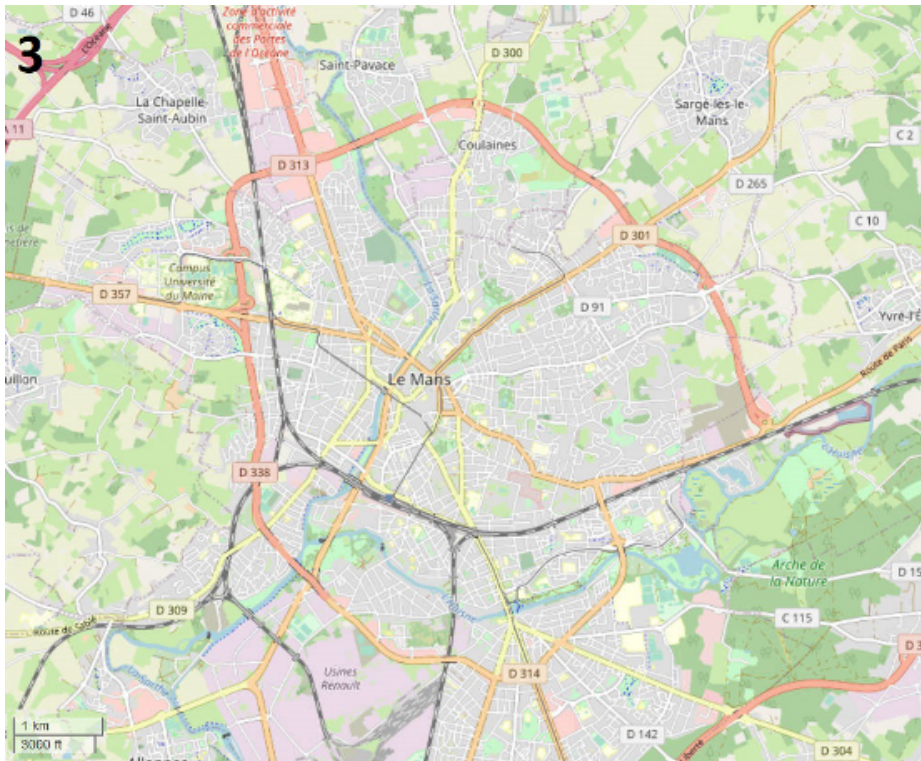
We present here the maps used in the third part of the survey on ring roads



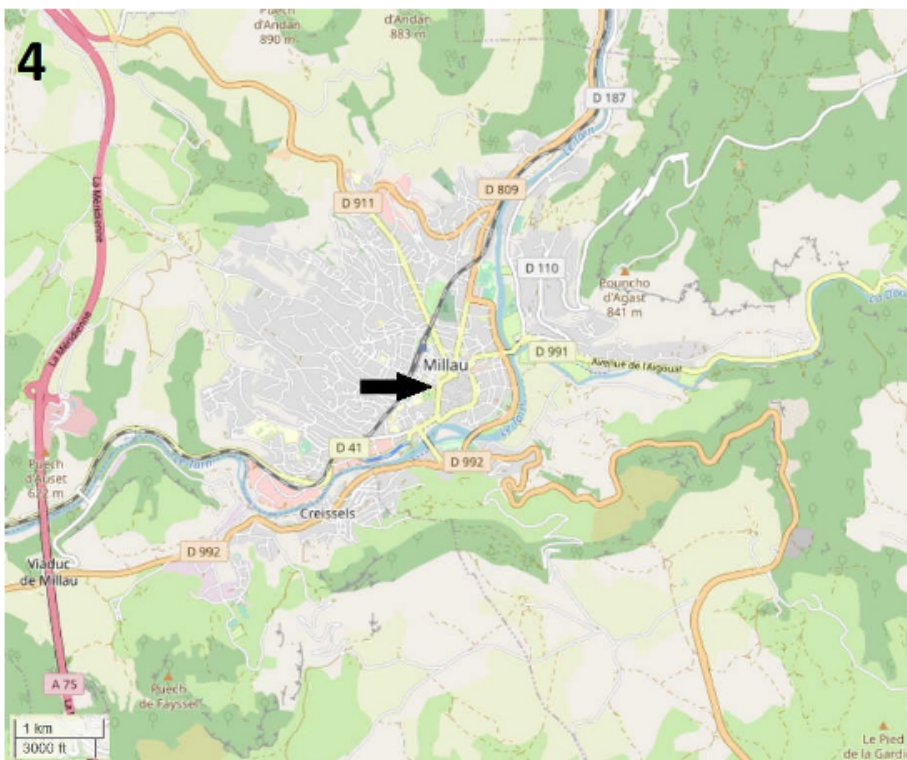
1 - Clermont-Ferrand



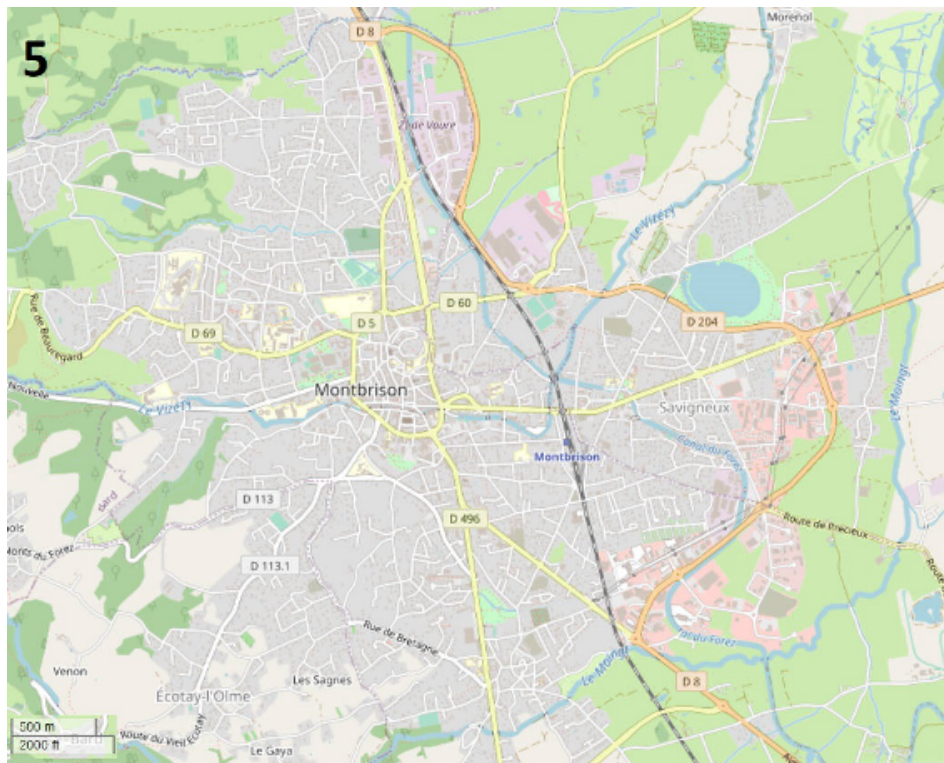
2 - Marne-la-Vallée



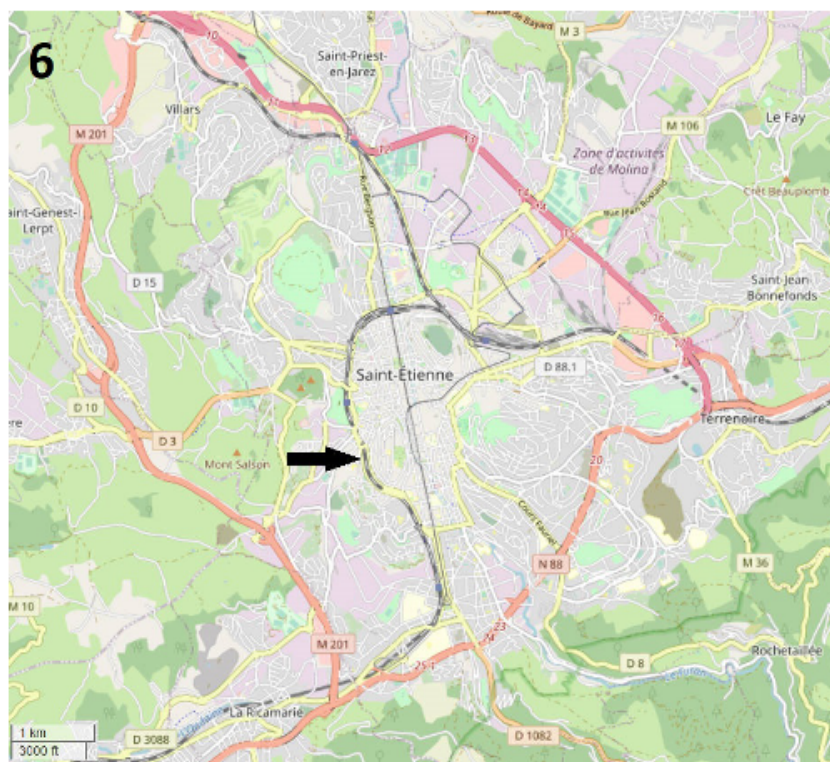
3 – Le Mans



4 – Millau

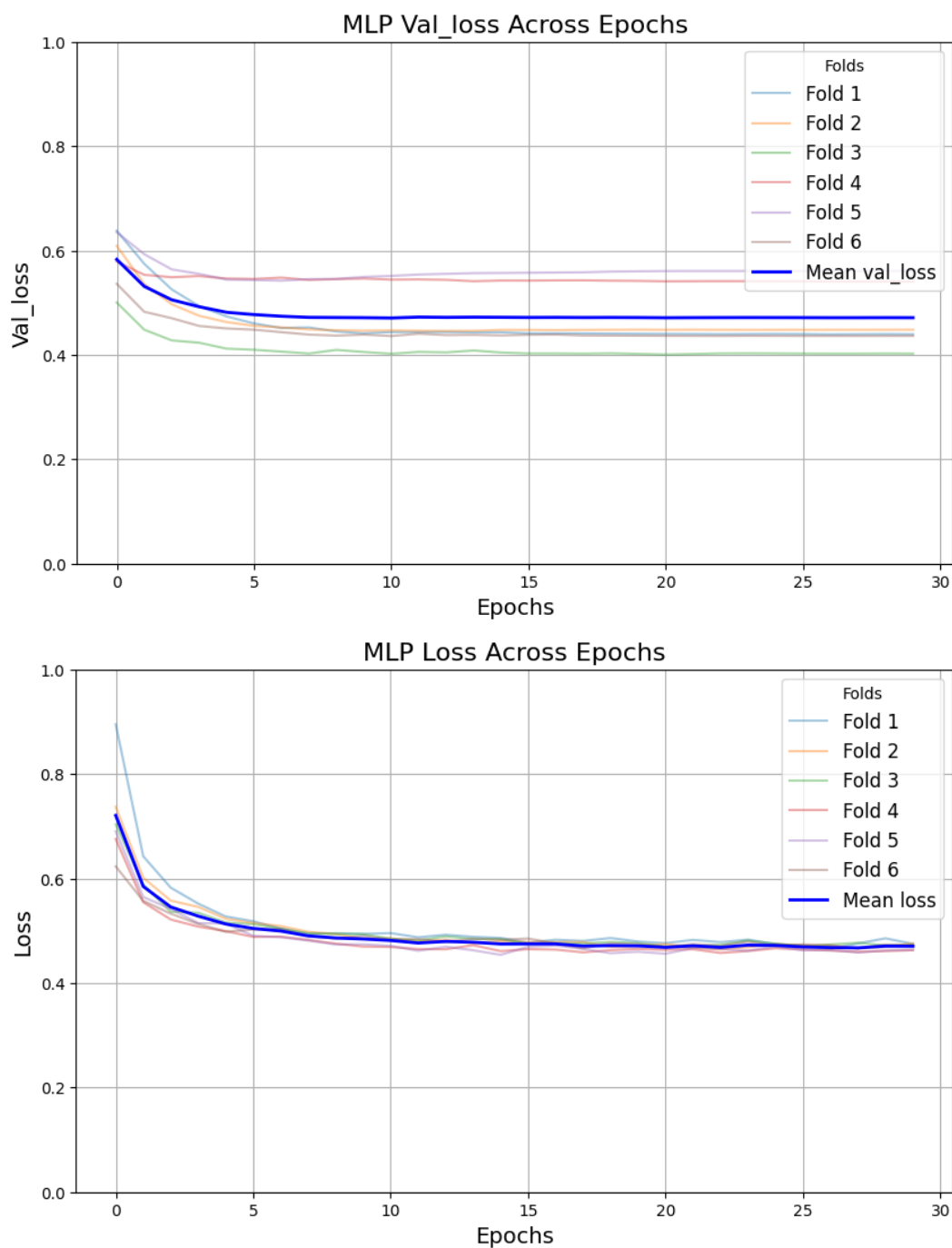


5 – Montbrison

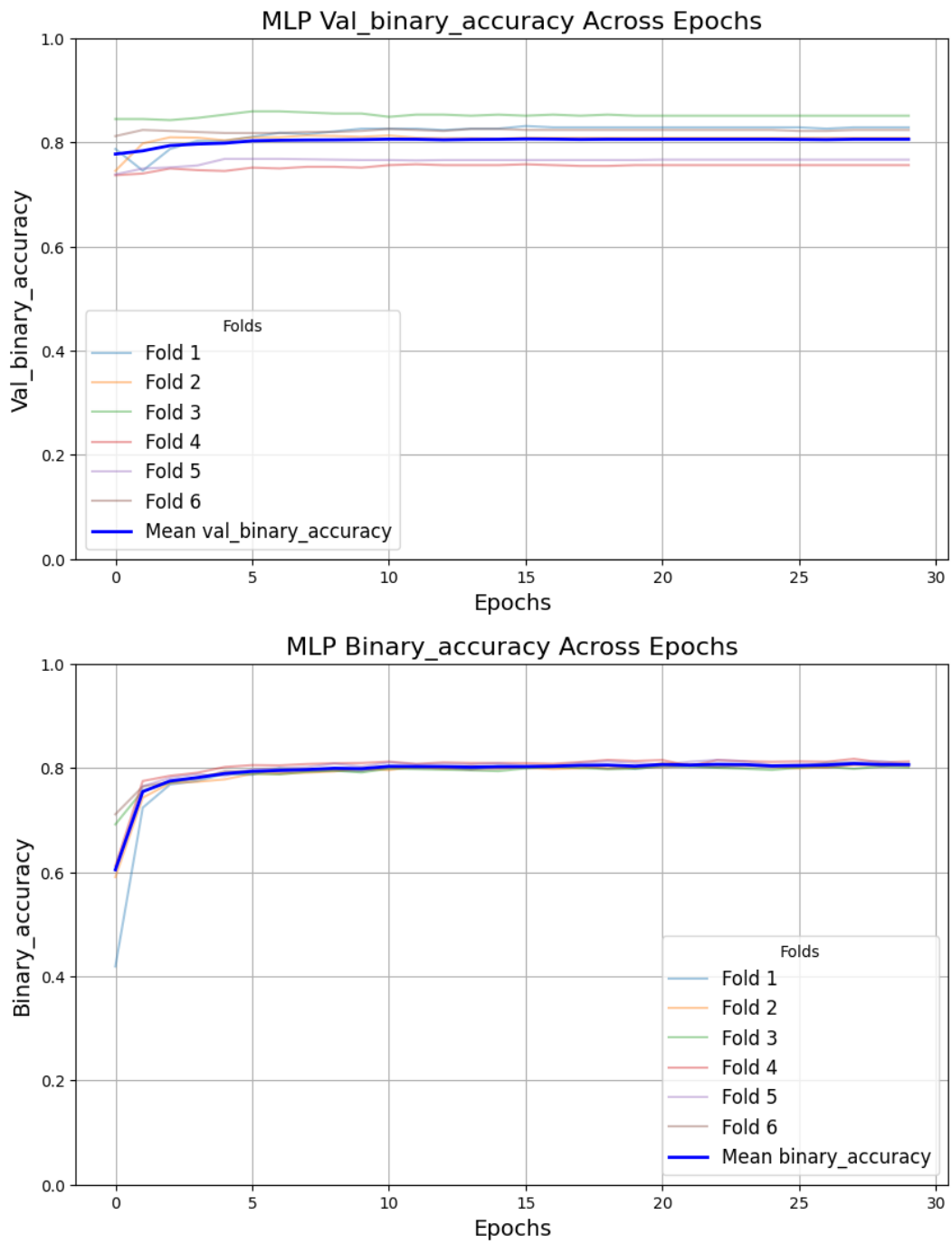


6 – Saint-Etienne

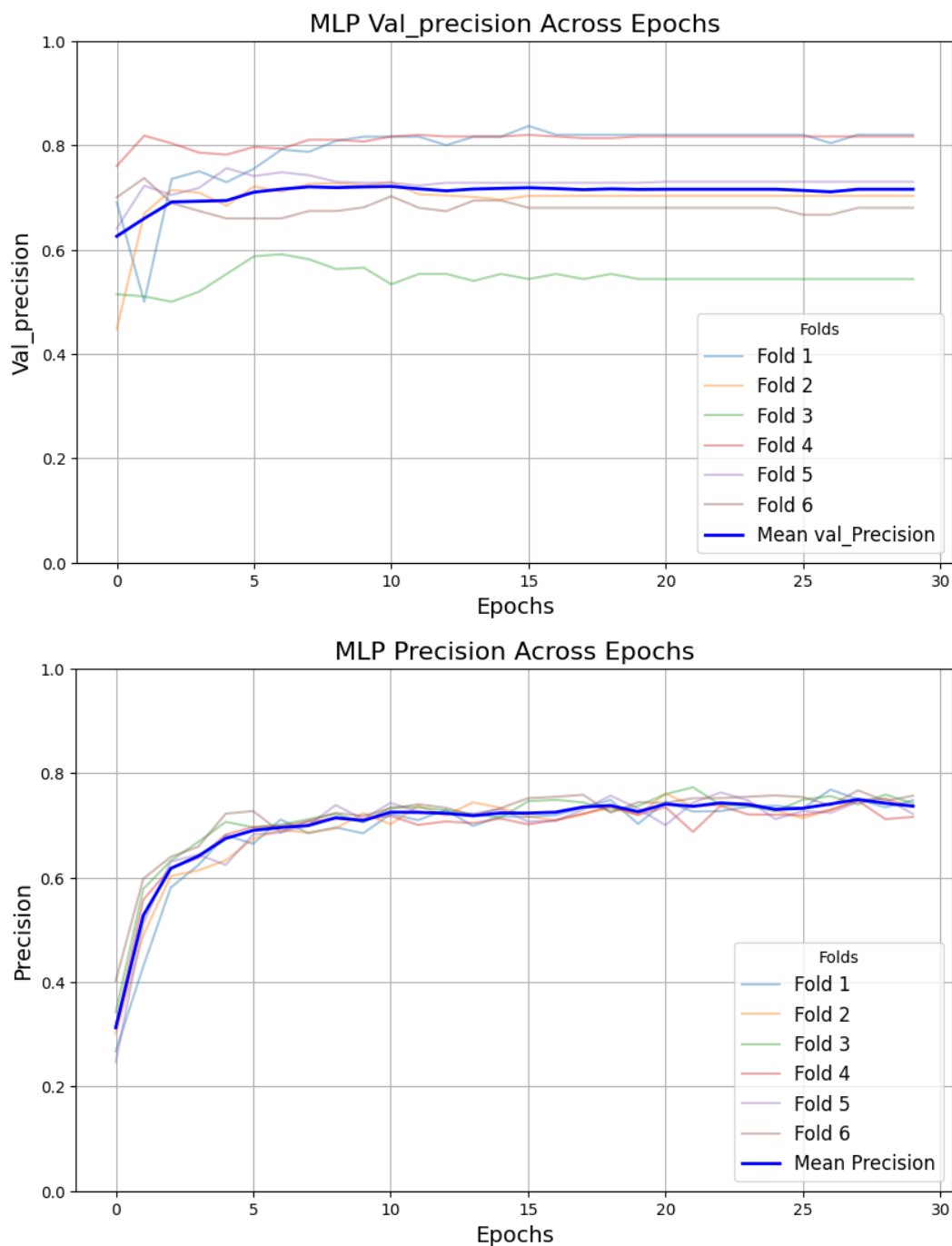
D Metrics monitored during MLP model training



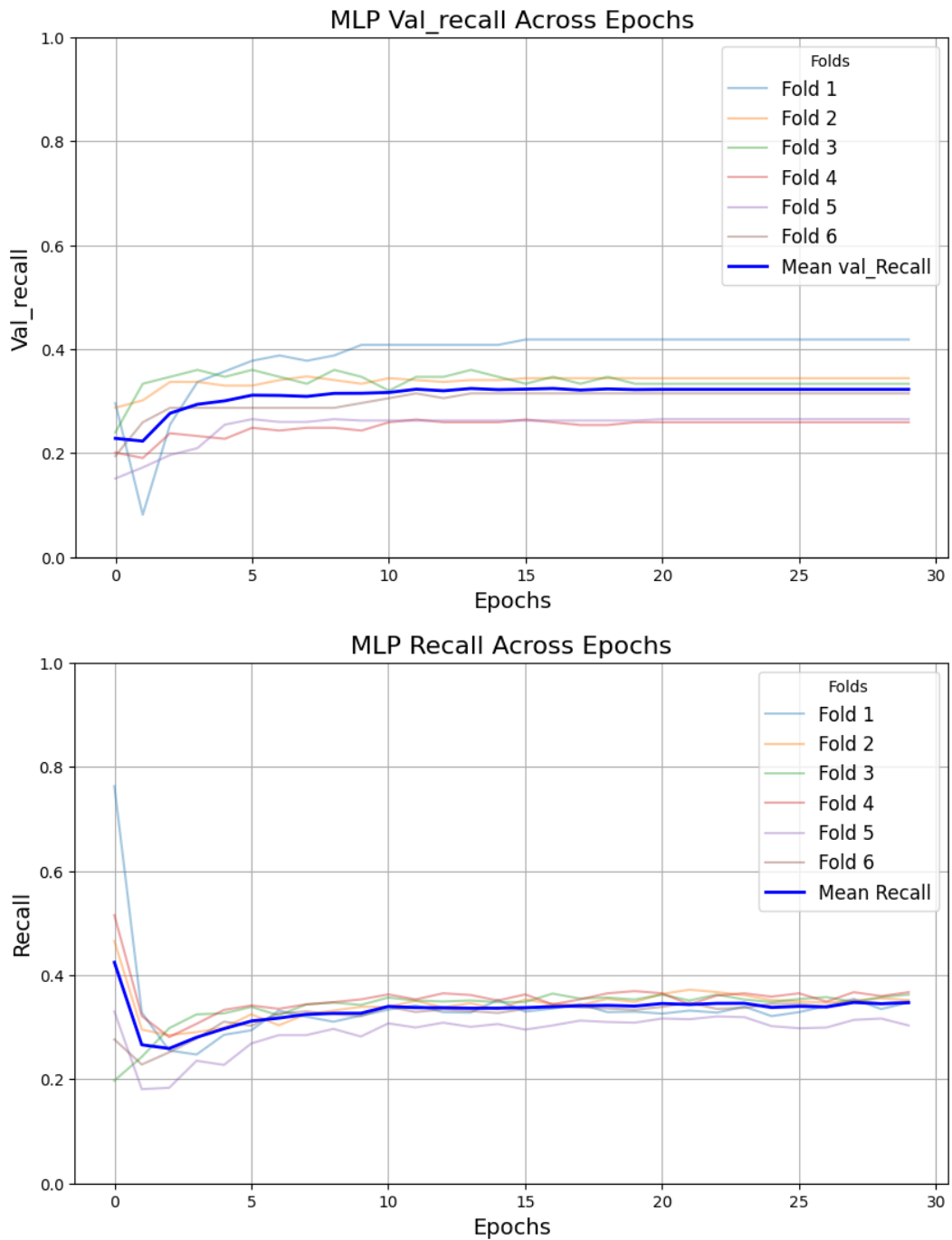
⊗ Figure 11 : Loss evolution over epochs for the MLP model on the training and validation sets.



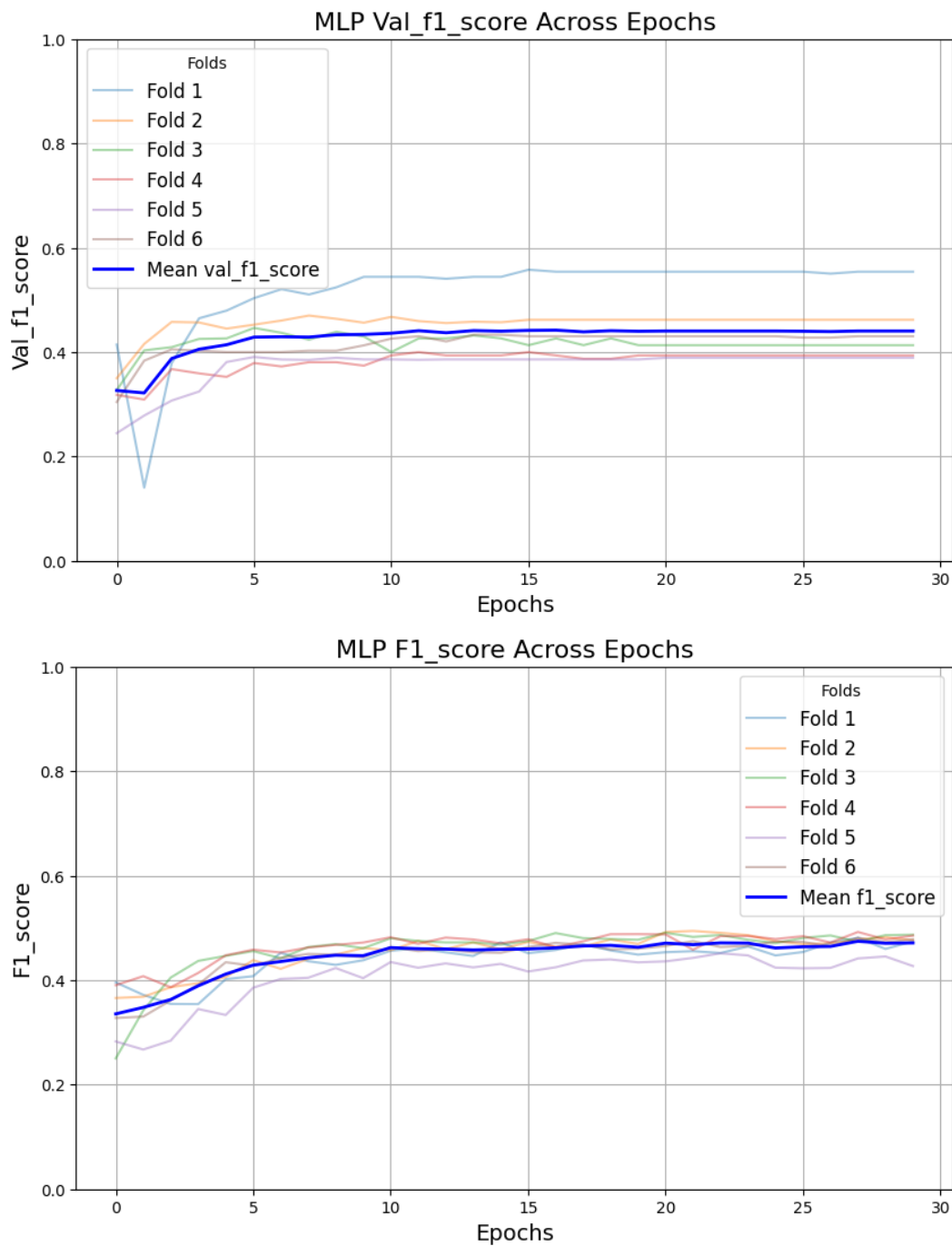
✿ Figure 12 : Accuracy values over epochs for the MLP model on the training and validation sets.



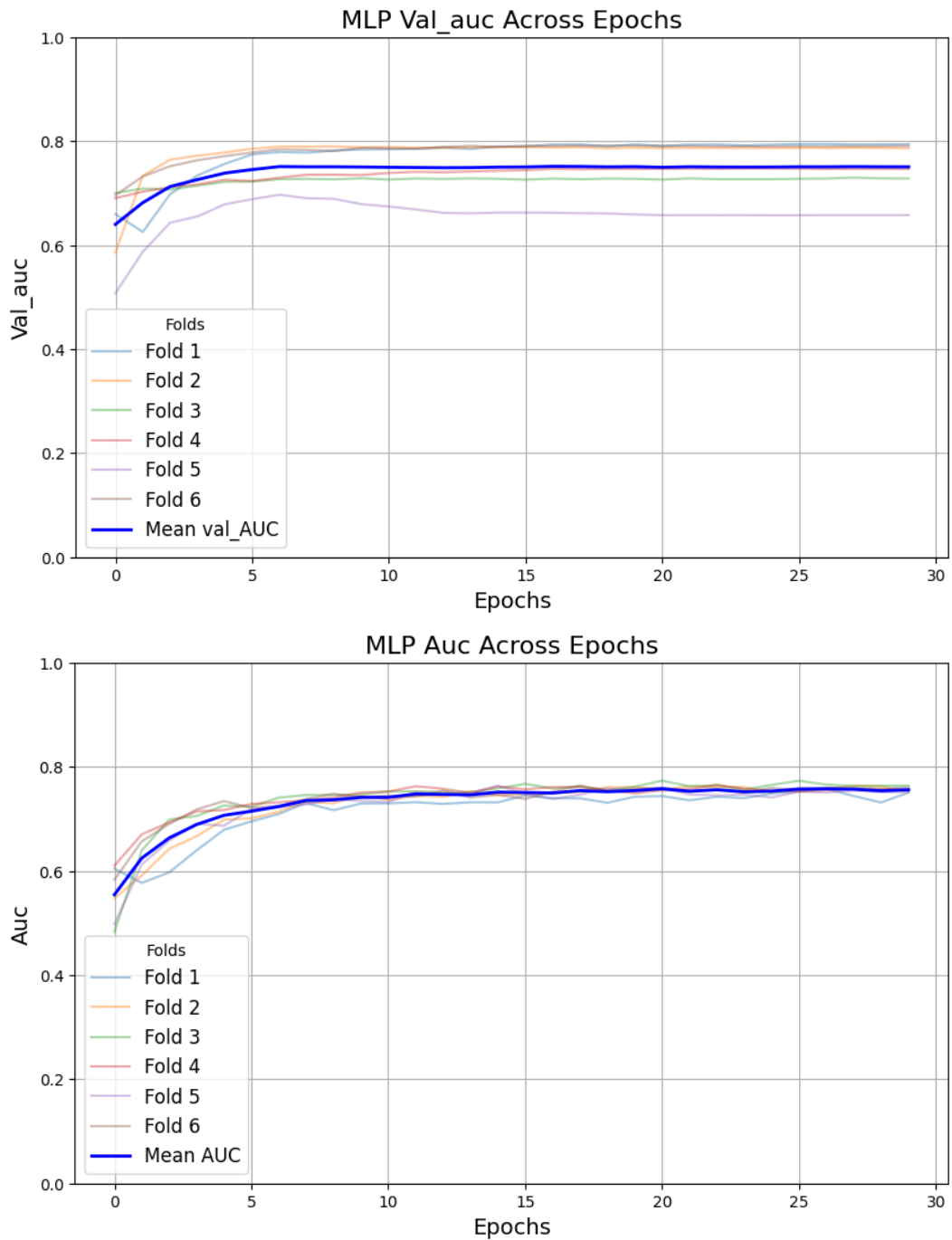
✿ Figure 13 : Precision values over epochs for the MLP model on the training and validation sets.



✿ Figure 14 : Recall values over epochs for the MLP model on the training and validation sets.

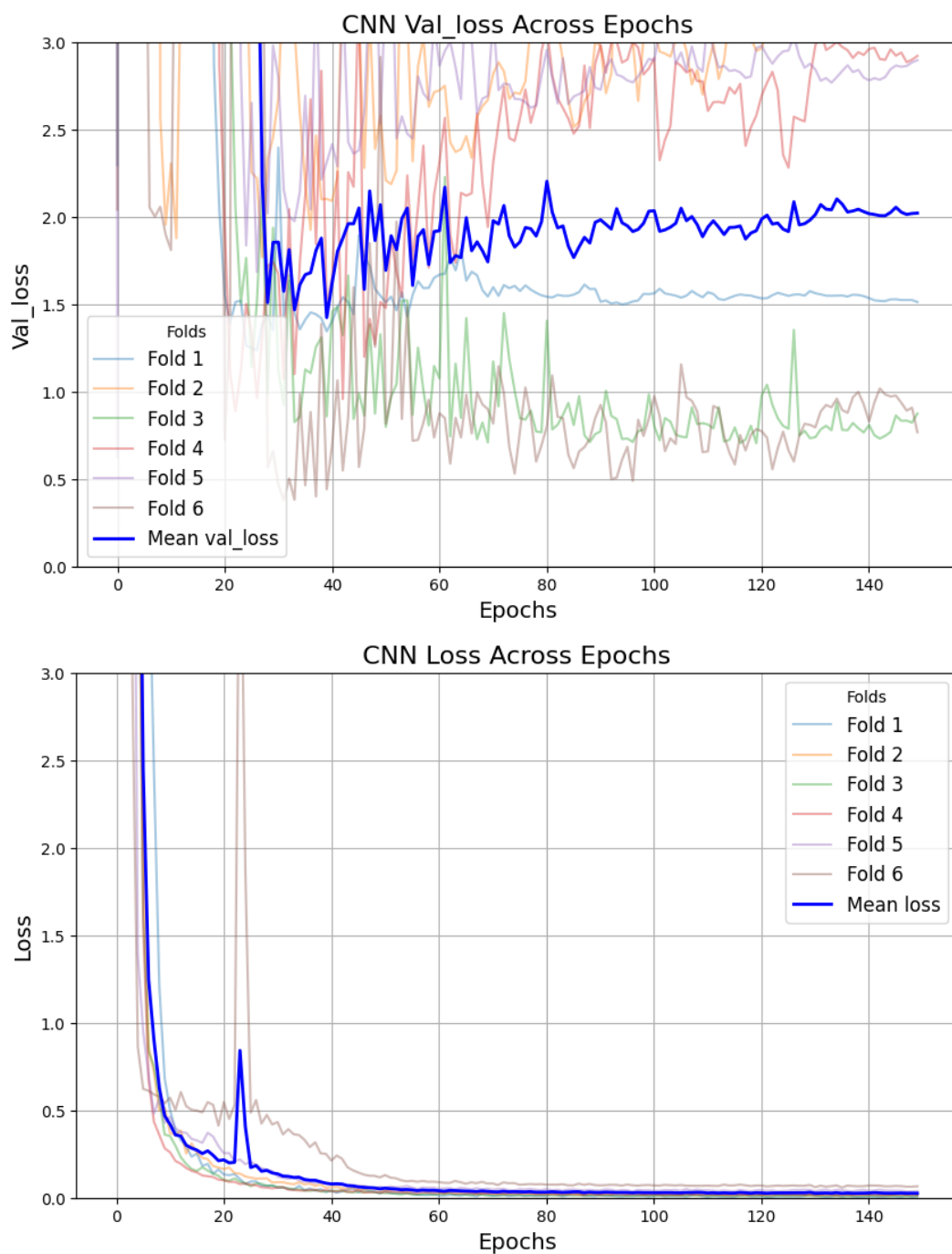


✿ Figure 15 : F1-score values over epochs for the MLP model on the training and validation sets.

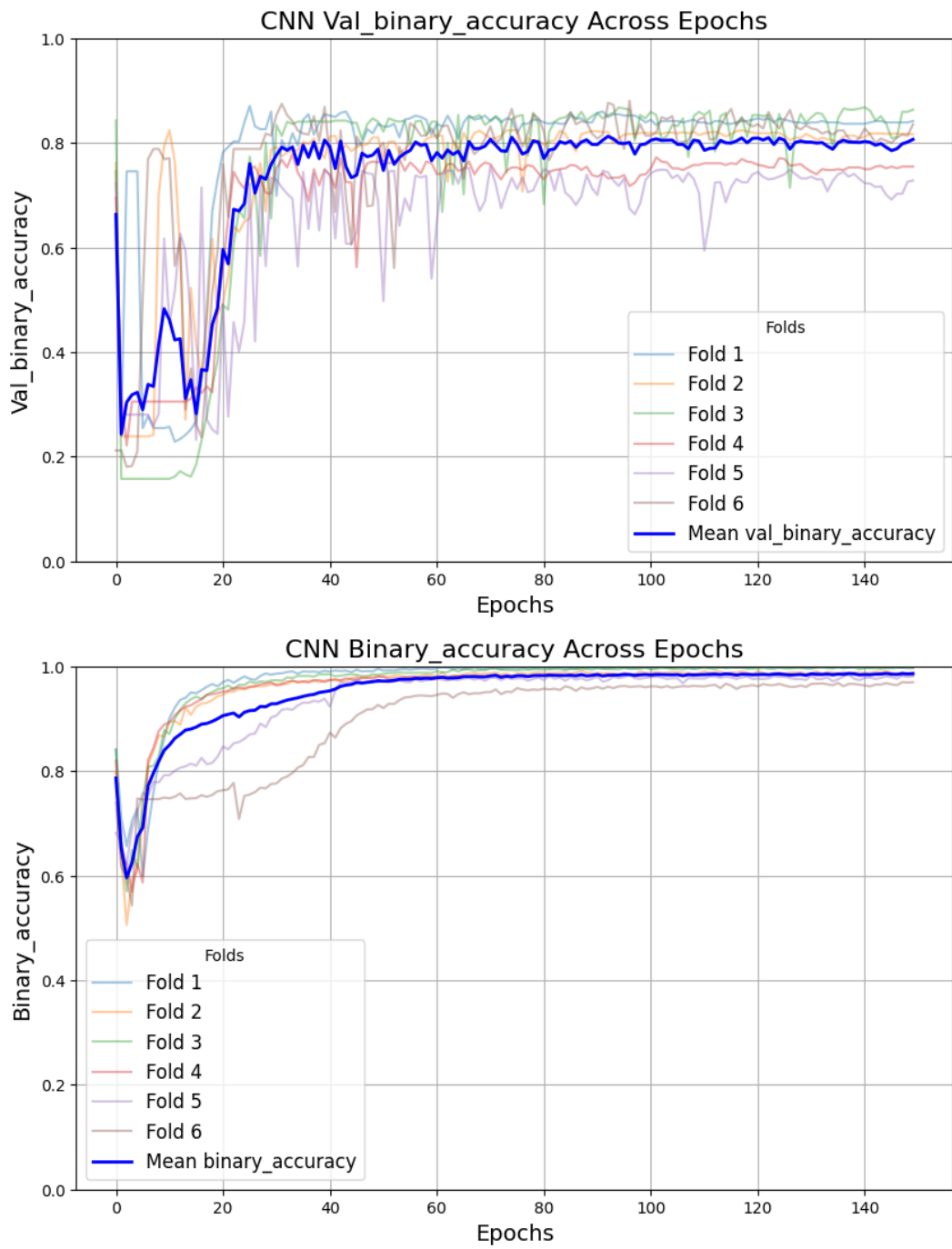


✿ Figure 16 : AUC values over epochs for the MLP model on the training and validation sets.

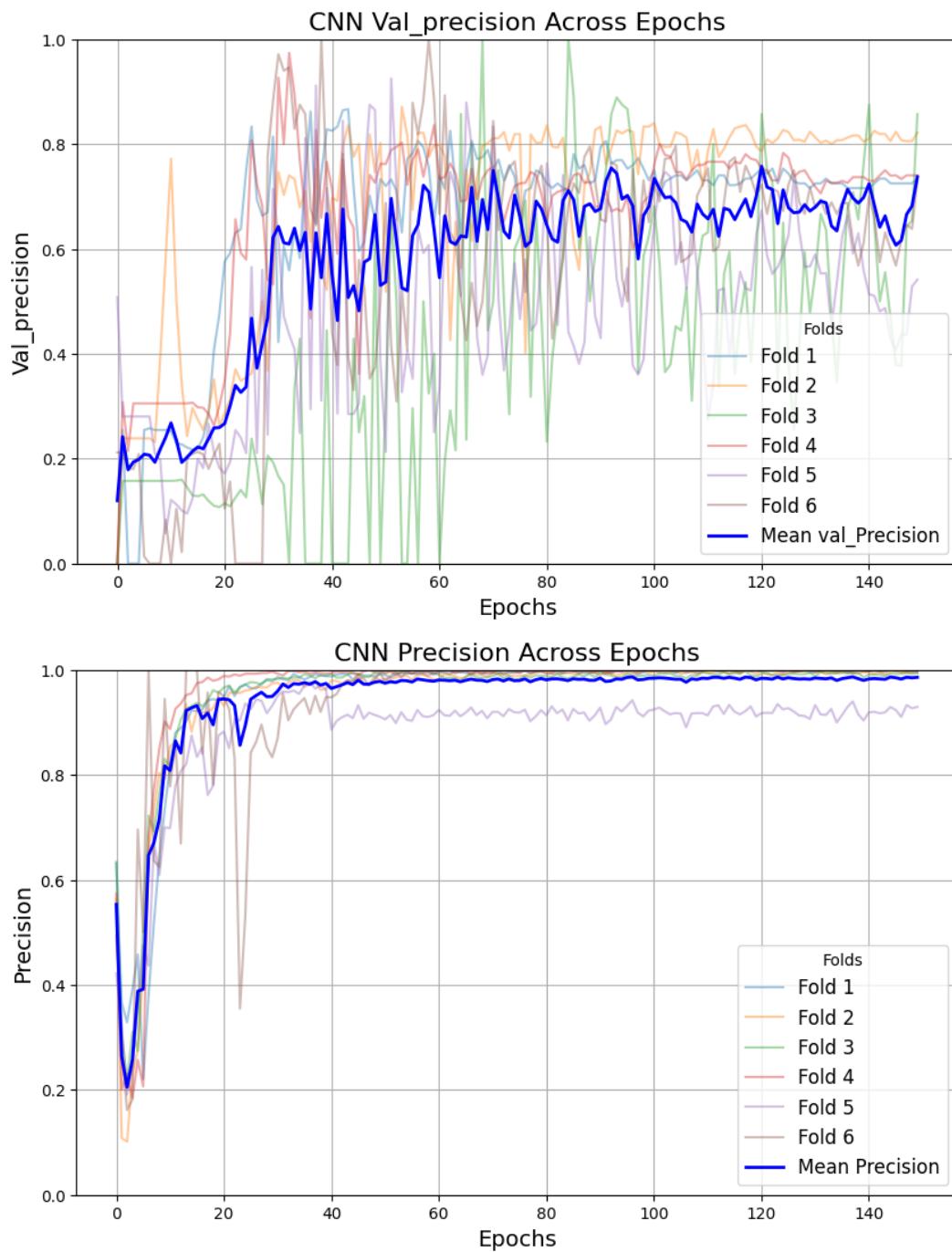
E Metrics monitored during CNN model training



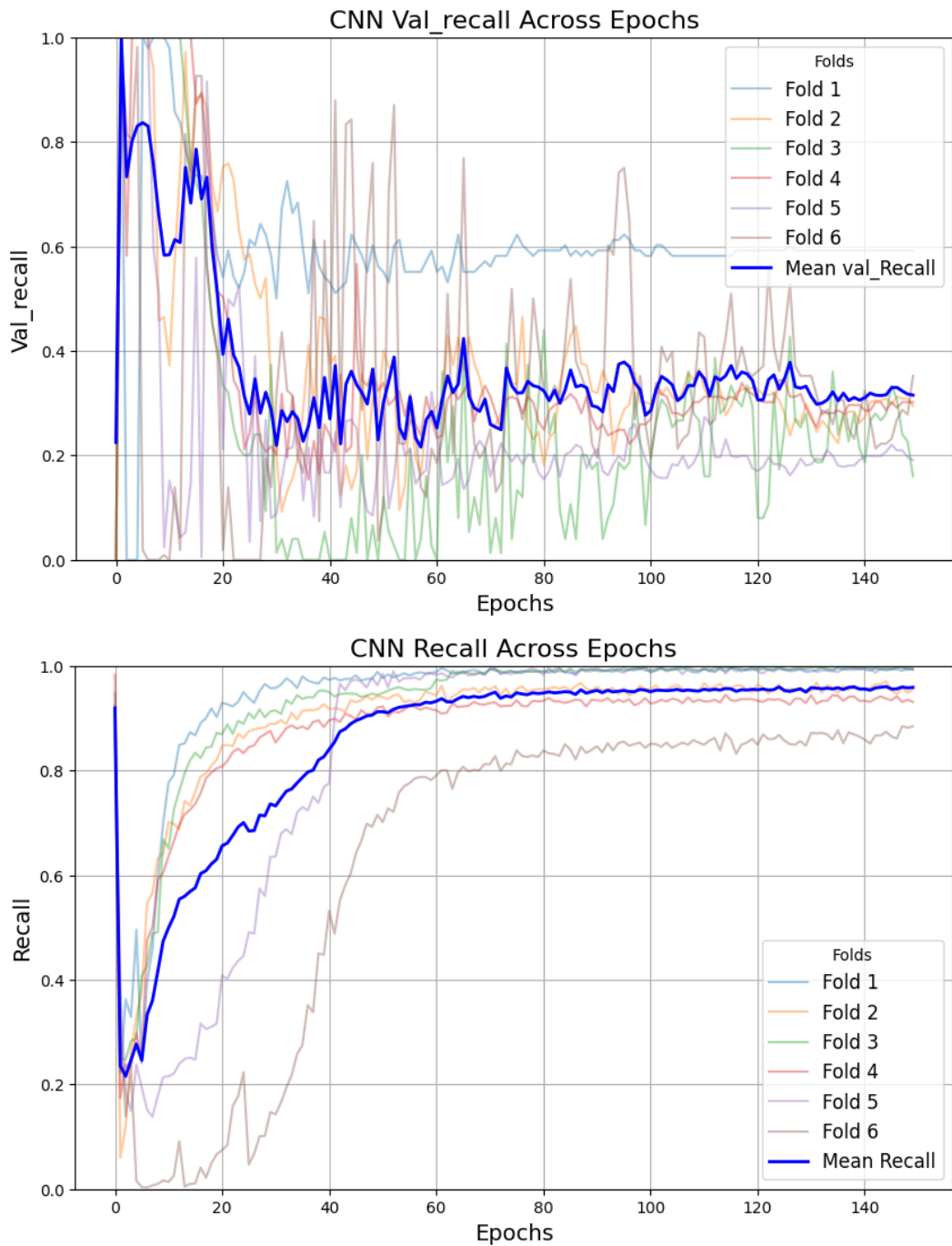
☞ Figure 17 : Loss evolution over epochs for the CNN model on the training and validation sets.



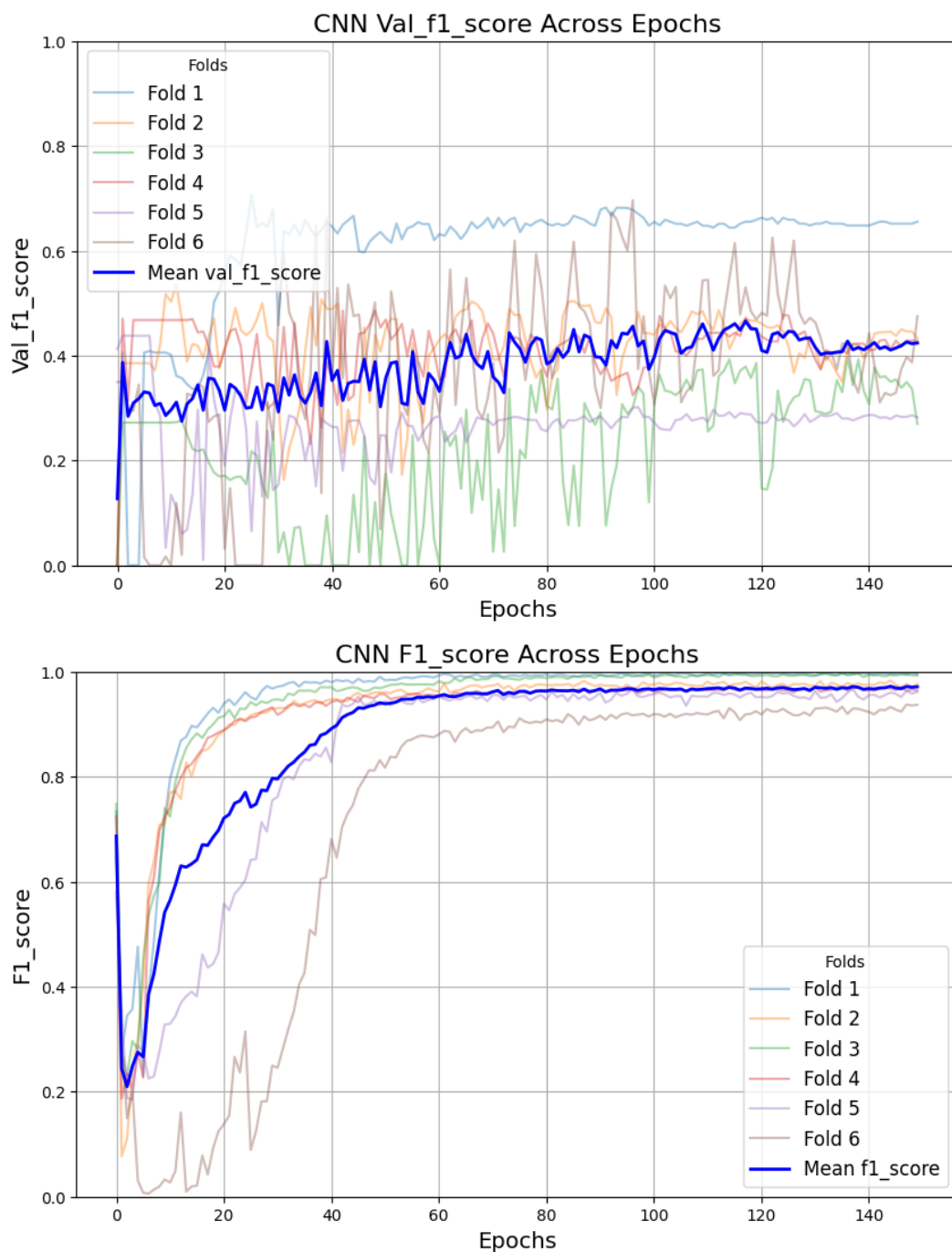
✿ Figure 18 : Accuracy values over epochs for the CNN model on the training and validation sets.



✿ Figure 19 : Precision values over epochs for the CNN model on the training and validation sets.



☼ Figure 20 : Recall values over epochs for the CNN model on the training and validation sets.



✿ Figure 21 : F1-score values over epochs for the CNN model on the training and validation sets.

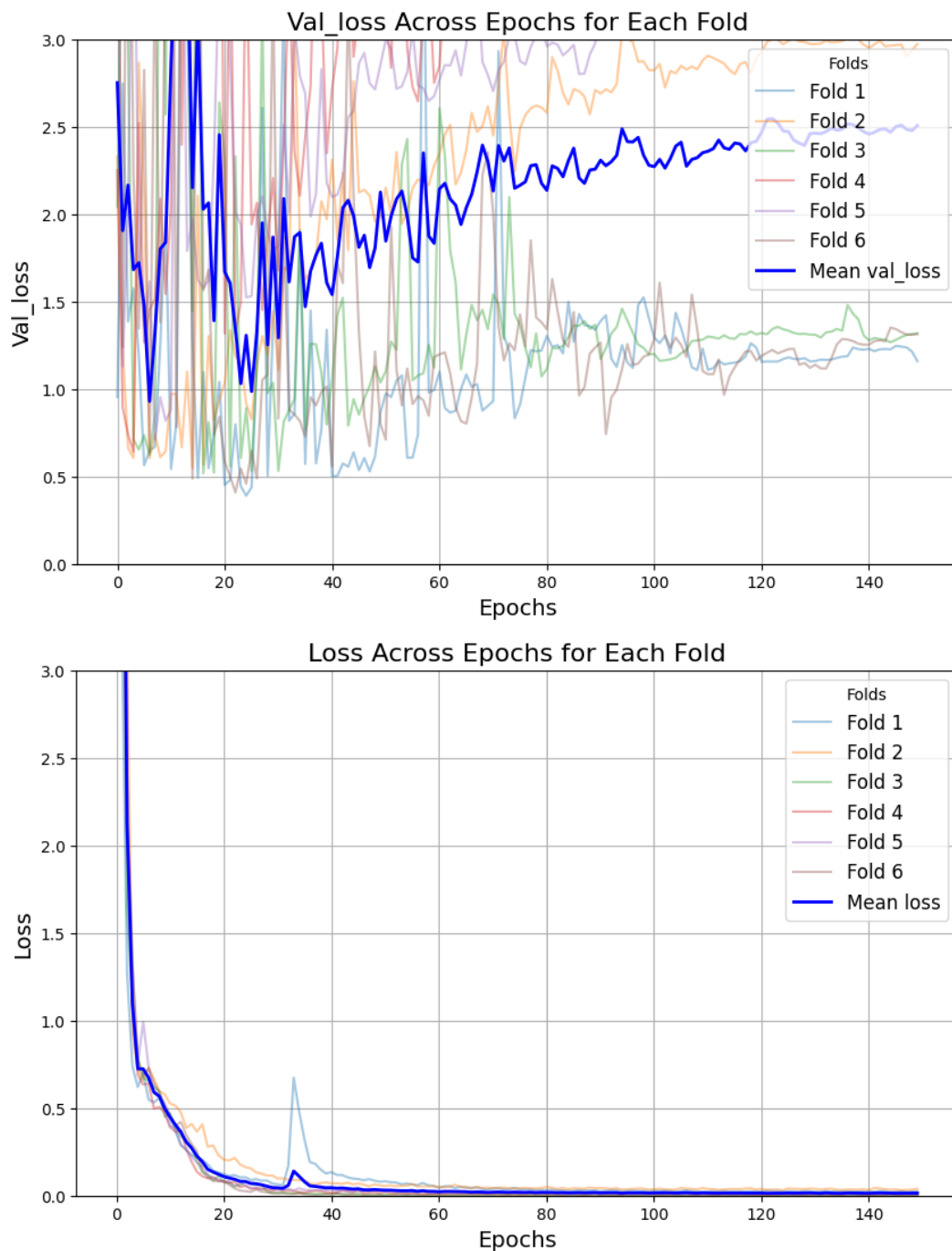
F Multimodal model summary and monitored metrics

Model: "Multimodal_Model"

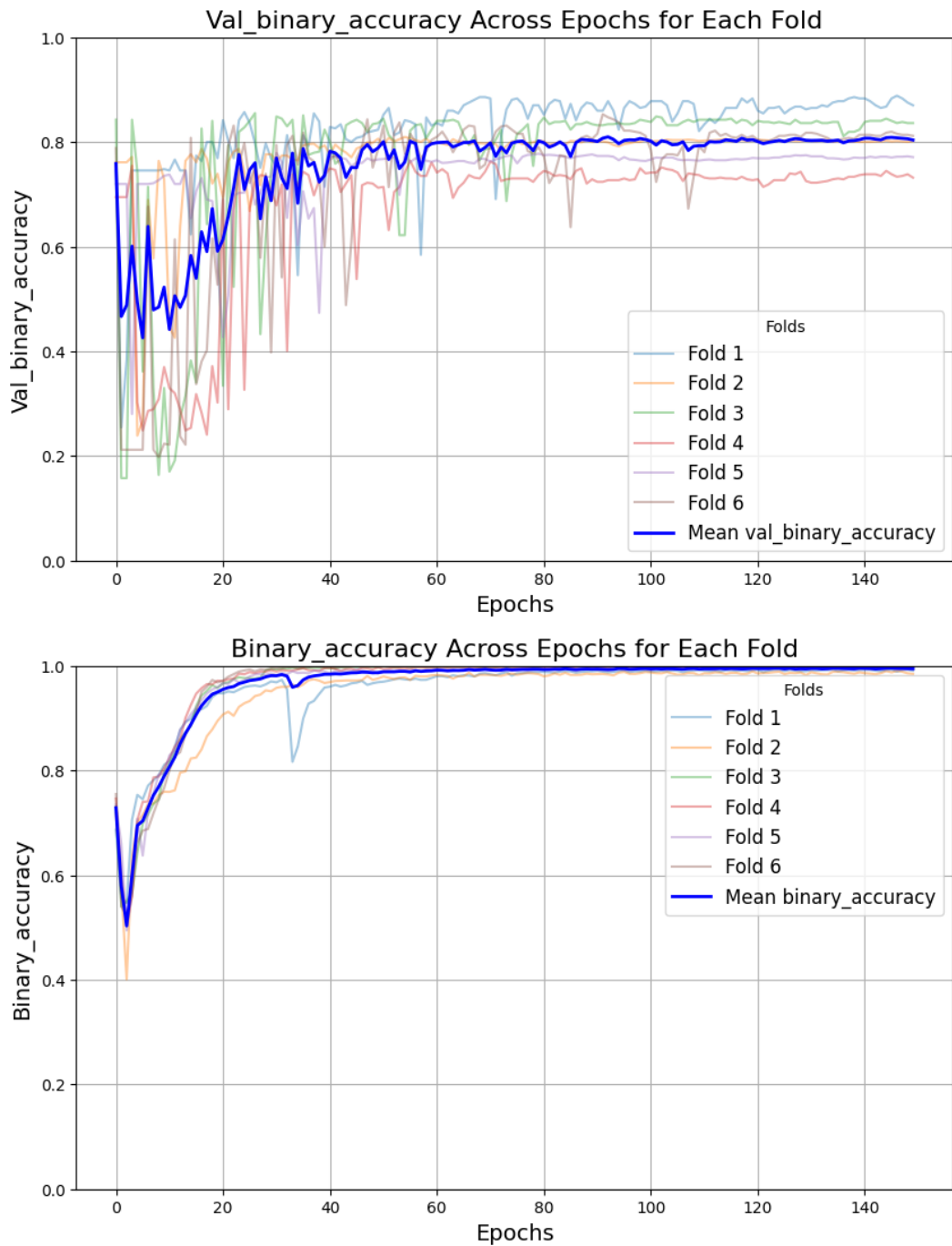
Layer (type)	Output Shape	Param #	Connected to
image_input (InputLayer)	(None, 256, 256, 3)	0	-
conv2d (Conv2D)	(None, 256, 256, 32)	896	image_input[0][0]
batch_normalization (BatchNormalization)	(None, 256, 256, 32)	128	conv2d[0][0]
max_pooling2d (MaxPooling2D)	(None, 128, 128, 32)	0	batch_normalization[0]
conv2d_1 (Conv2D)	(None, 128, 128, 64)	16,496	max_pooling2d[0][0]
batch_normalization_1 (BatchNormalization)	(None, 128, 128, 64)	256	conv2d_1[0][0]
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 64)	0	batch_normalization_1
conv2d_2 (Conv2D)	(None, 64, 64, 128)	73,856	max_pooling2d_1[0][0]
batch_normalization_2 (BatchNormalization)	(None, 64, 64, 128)	512	conv2d_2[0][0]
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 128)	0	batch_normalization_2
flatten (Flatten)	(None, 131072)	0	max_pooling2d_2[0][0]
tabular_input (InputLayer)	(None, 10)	0	-
dense (Dense)	(None, 128)	16,777,344	flatten[0][0]
dense_1 (Dense)	(None, 32)	352	tabular_input[0][0]
dropout (Dropout)	(None, 128)	0	dense[0][0]
dropout_1 (Dropout)	(None, 32)	0	dense_1[0][0]
cnn_features (Dense)	(None, 64)	8,256	dropout[0][0]
mlp_features (Dense)	(None, 16)	528	dropout_1[0][0]
concatenate (Concatenate)	(None, 80)	0	cnn_features[0][0], mlp_features[0][0]
dense_2 (Dense)	(None, 128)	10,368	concatenate[0][0]
dropout_2 (Dropout)	(None, 128)	0	dense_2[0][0]
dense_3 (Dense)	(None, 64)	8,256	dropout_2[0][0]
dropout_3 (Dropout)	(None, 64)	0	dense_3[0][0]
dense_4 (Dense)	(None, 32)	2,080	dropout_3[0][0]
final_output (Dense)	(None, 1)	33	dense_4[0][0]

Total params: 16,901,361 (64.47 MB)
Trainable params: 16,900,913 (64.47 MB)
Non-trainable params: 448 (1.75 KB)

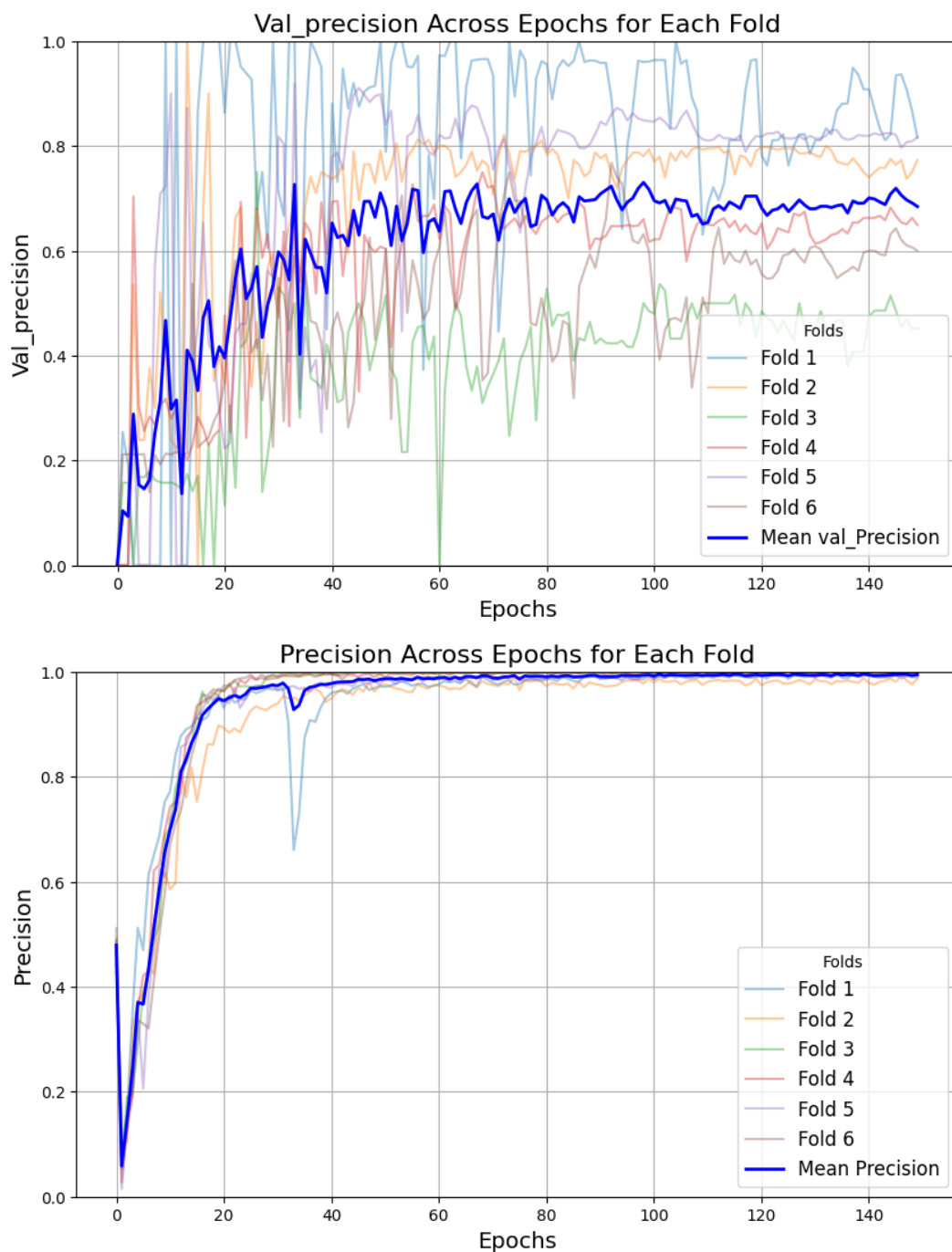
⊗ Figure 22 : The different layers of our multimodal model, implemented with Tensorflow.



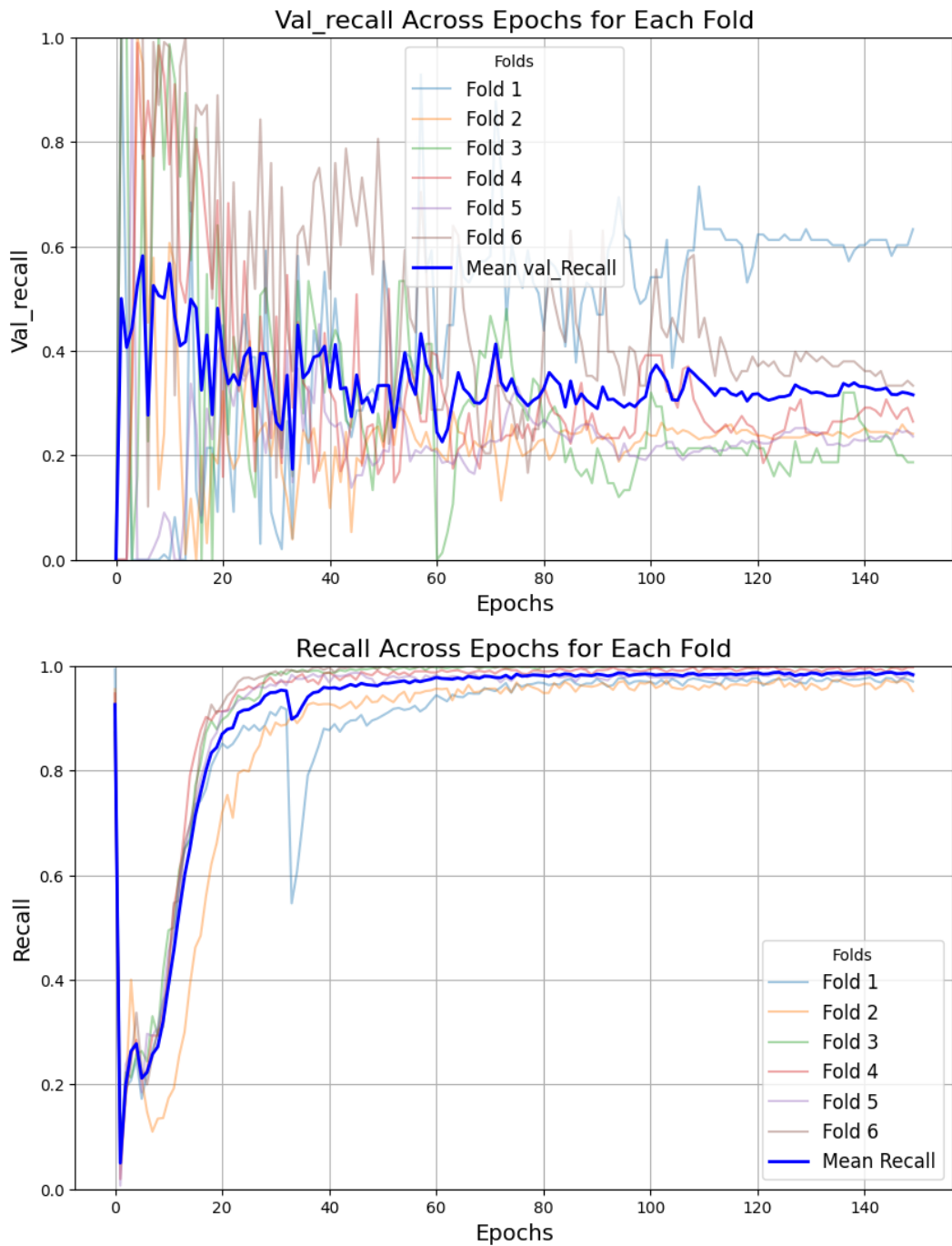
✿ Figure 23 : Loss evolution over epochs for the multimodal model on the training and validation sets.



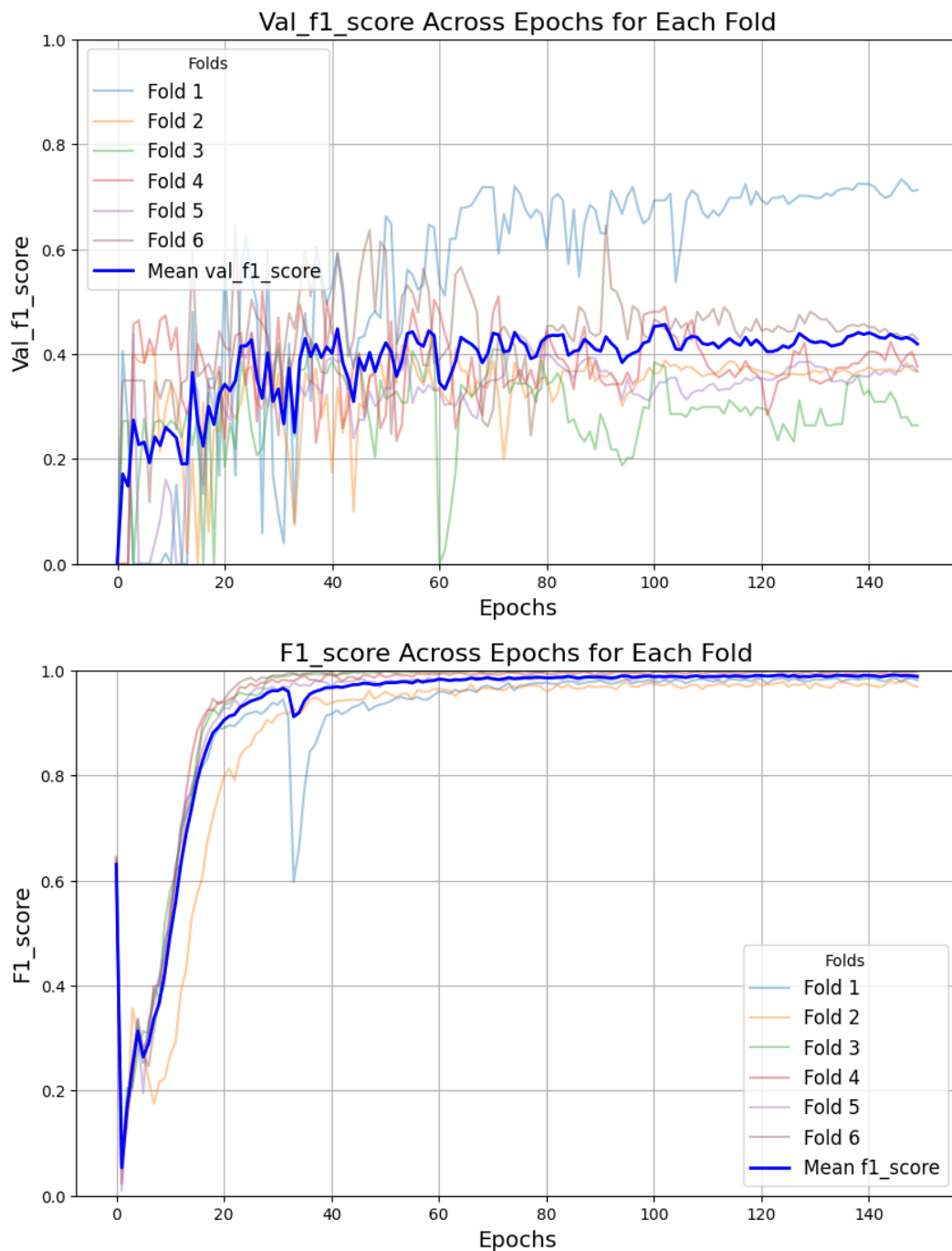
✿ Figure 24 : Accuracy values over epochs for the multimodal model on the training and validation sets.



✿ Figure 25 : Precision values over epochs for the Multimodal model on the training and validation sets.



✿ Figure 26 : Recall values over epochs for the multimodal model on the training and validation sets.

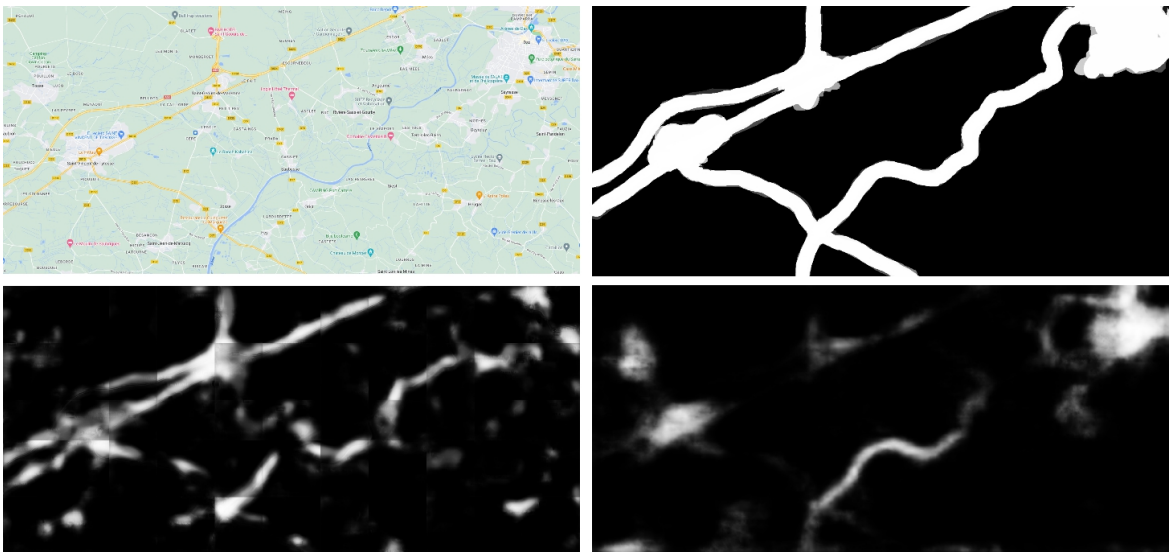


✿ Figure 27 : F1-score values over epochs for the multimodal model on the training and validation sets.

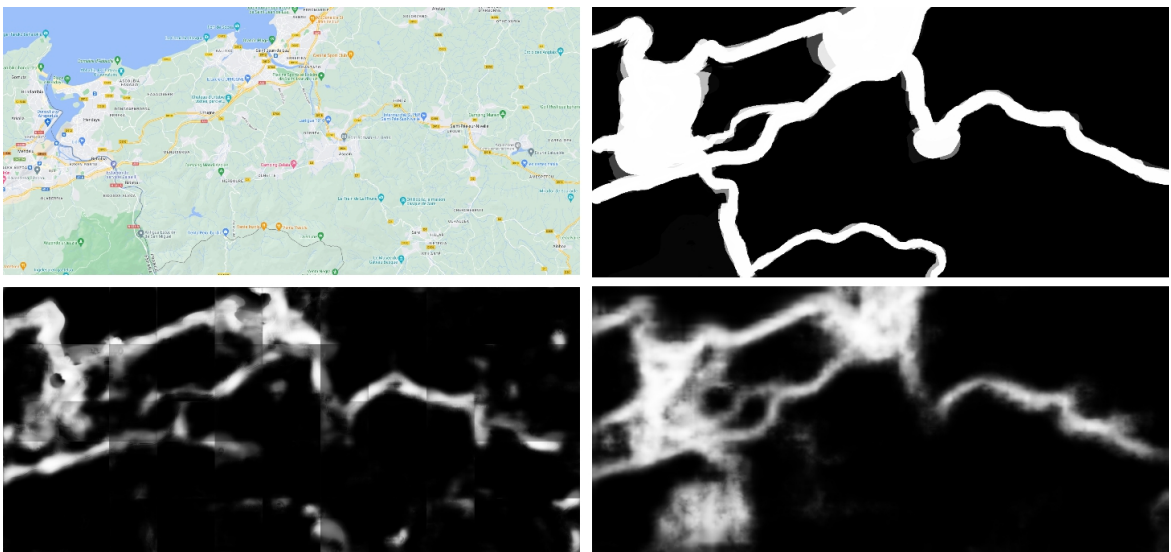
G AHAP segmentation evaluation

In this section of the appendix can be found all 20 zones of the evaluation dataset we propose in Chapter E. Next to each map image, there is:

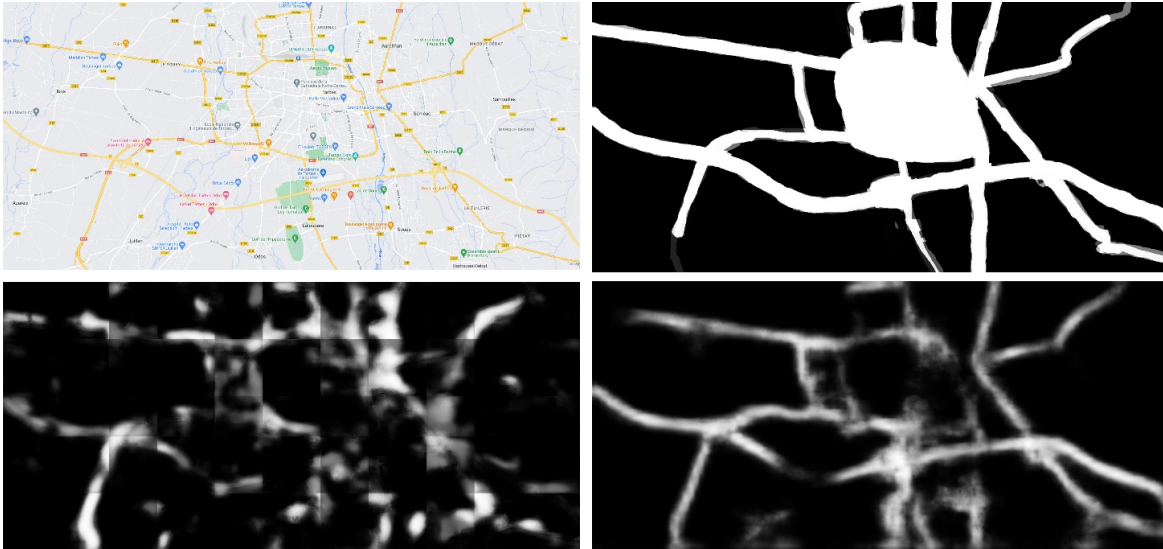
- a STAPLE mask based on participant annotations (top right).
- a prediction from the U-Net model we trained (bottom left).
- a prediction from the Segformer model we finetuned (bottom right).



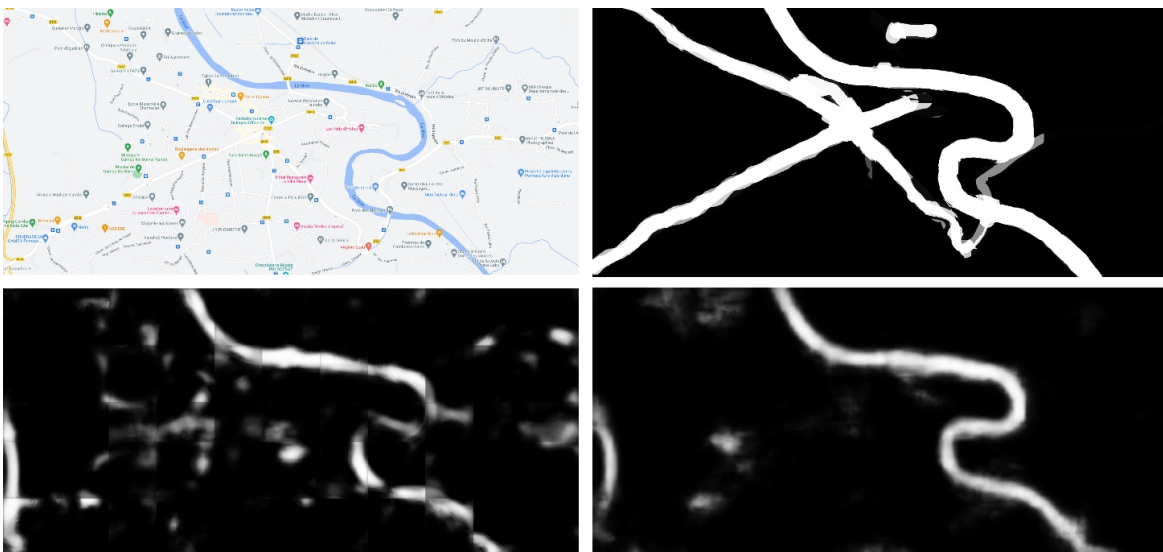
✿ Figure 28 : zone 1



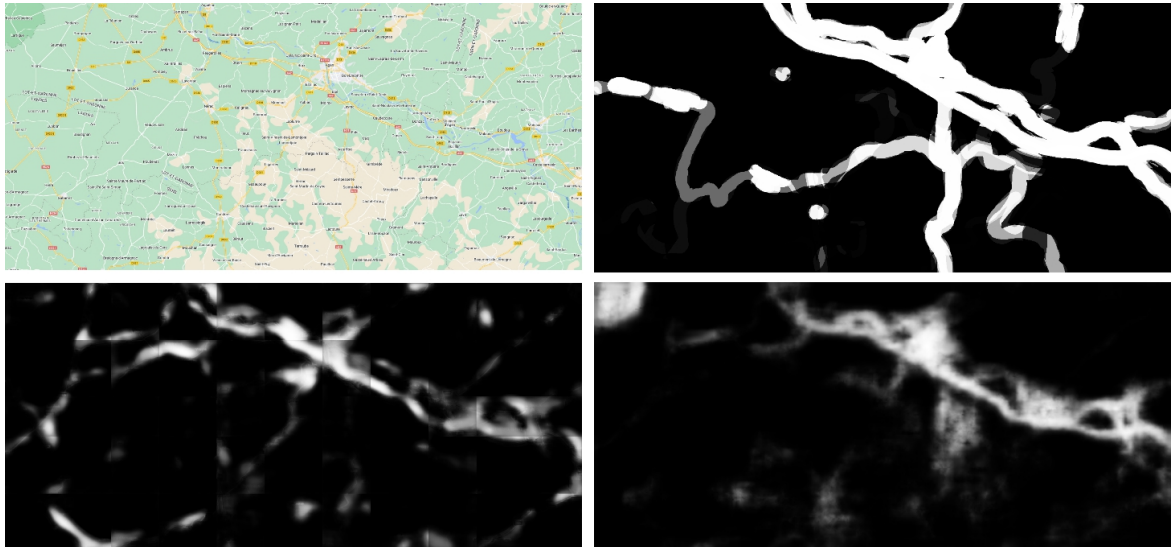
✿ Figure 29 : zone 2



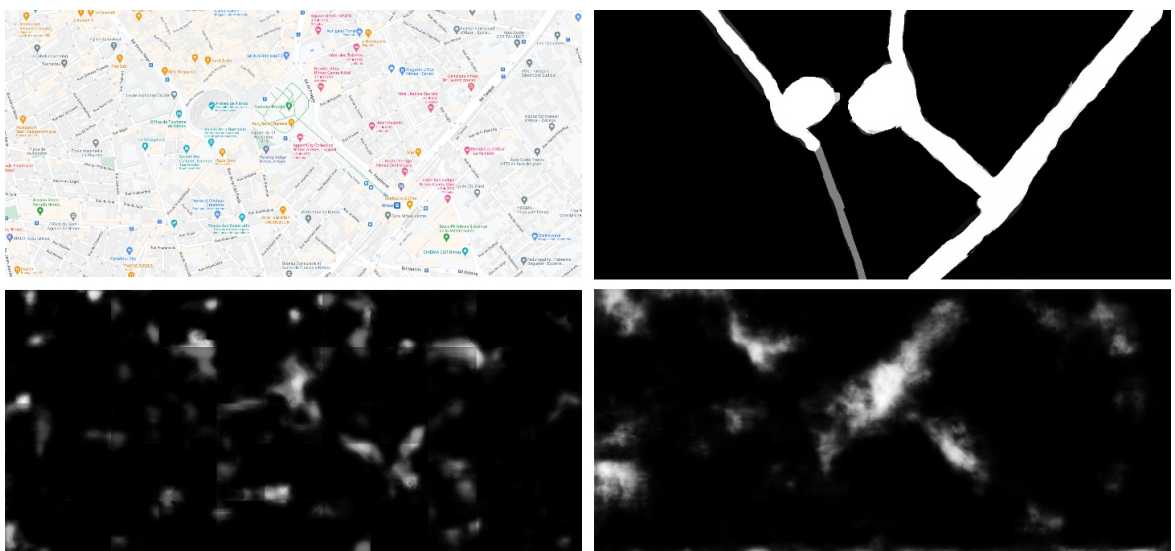
☼ Figure 30 : zone 3



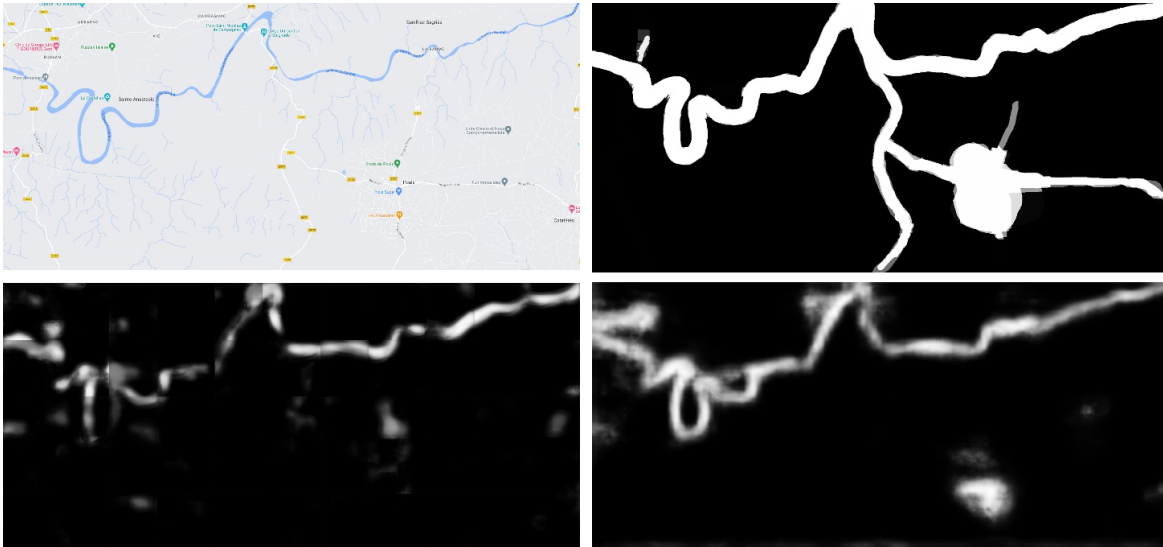
☼ Figure 31 : zone 4



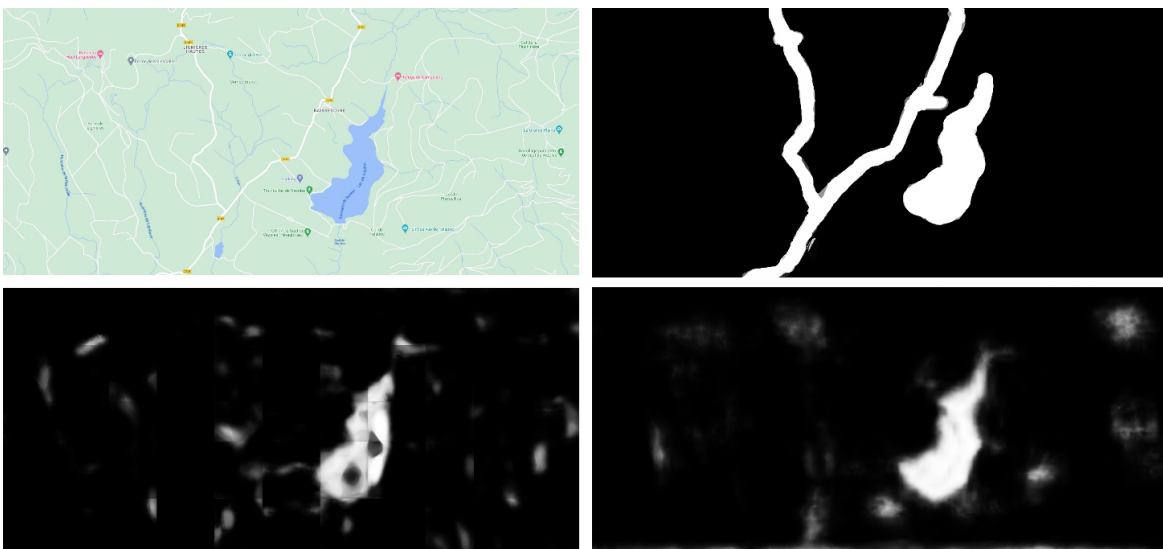
❁ Figure 32 : zone 5



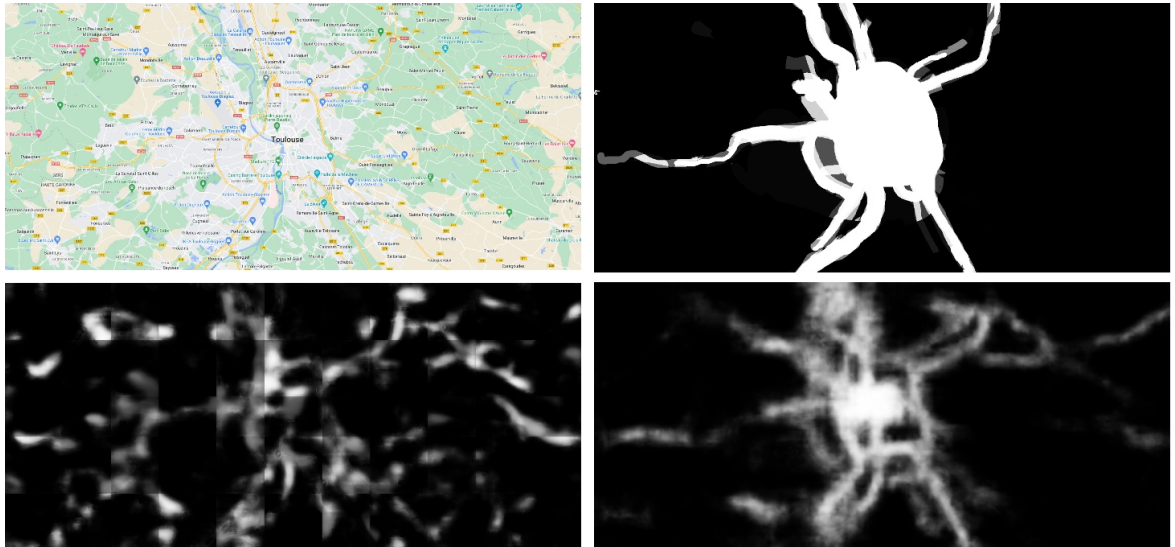
❁ Figure 33 : zone 6



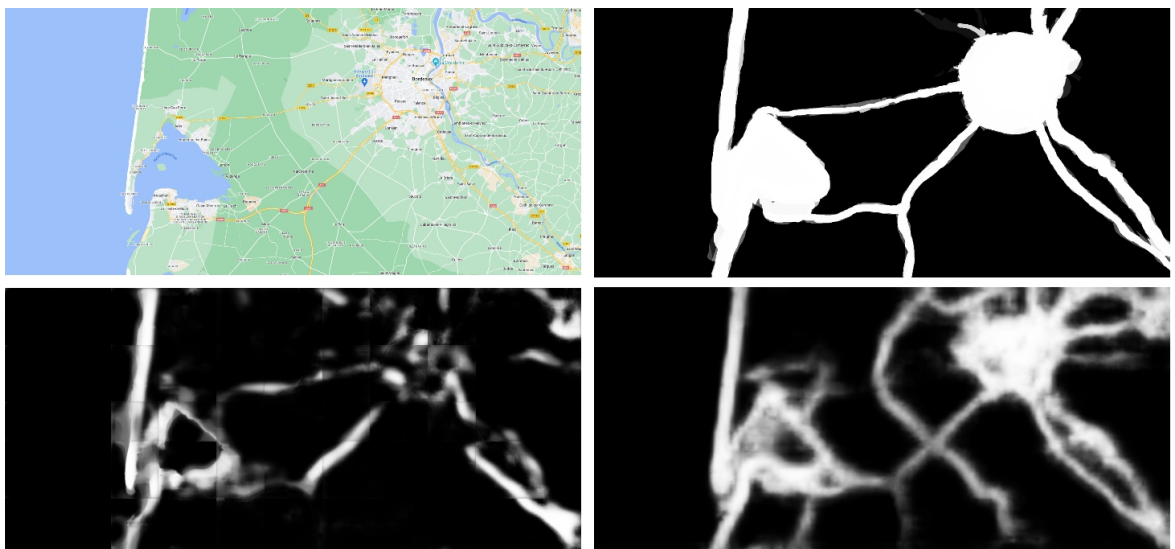
❁ Figure 34 : zone 7



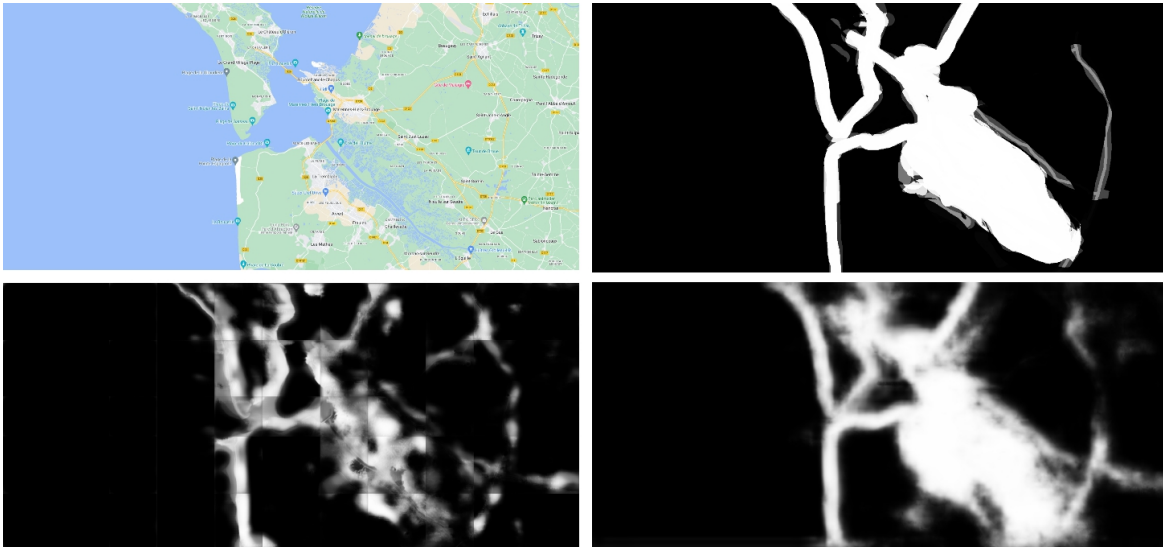
❁ Figure 35 : zone 8



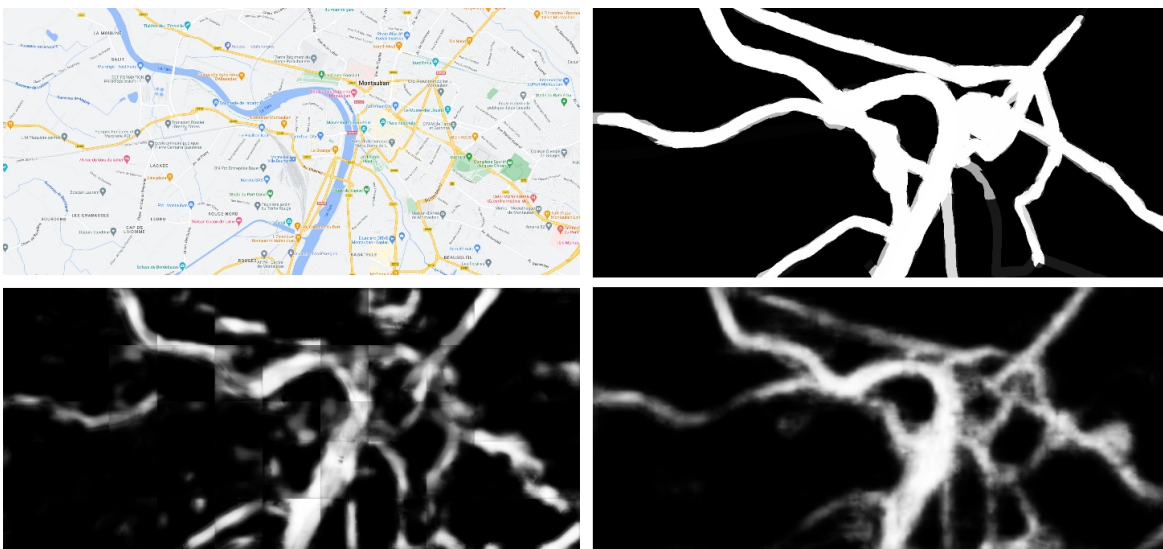
❁ Figure 36 : zone 9



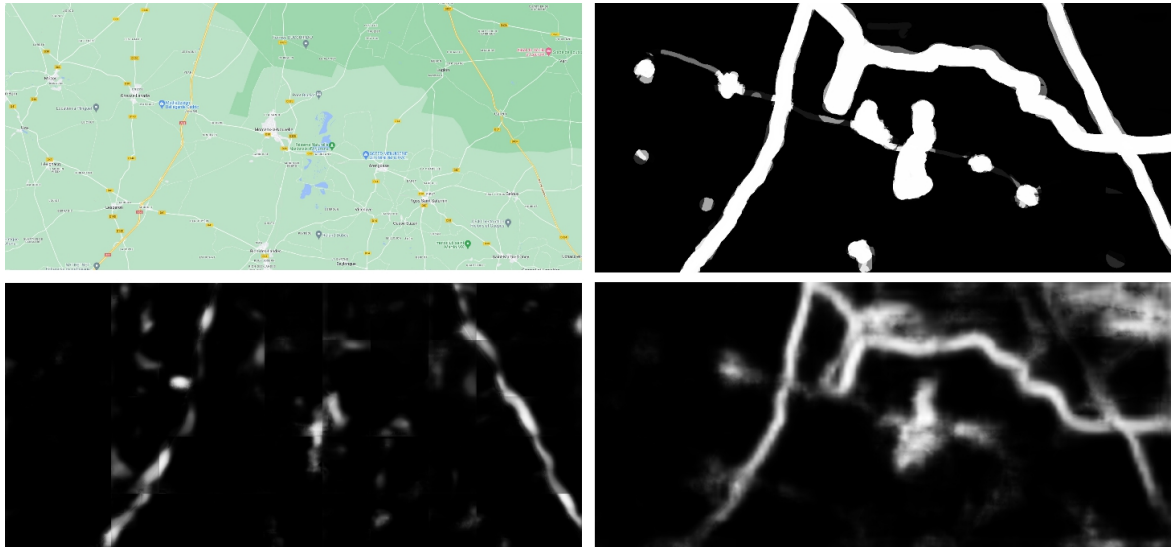
❁ Figure 37 : zone 10



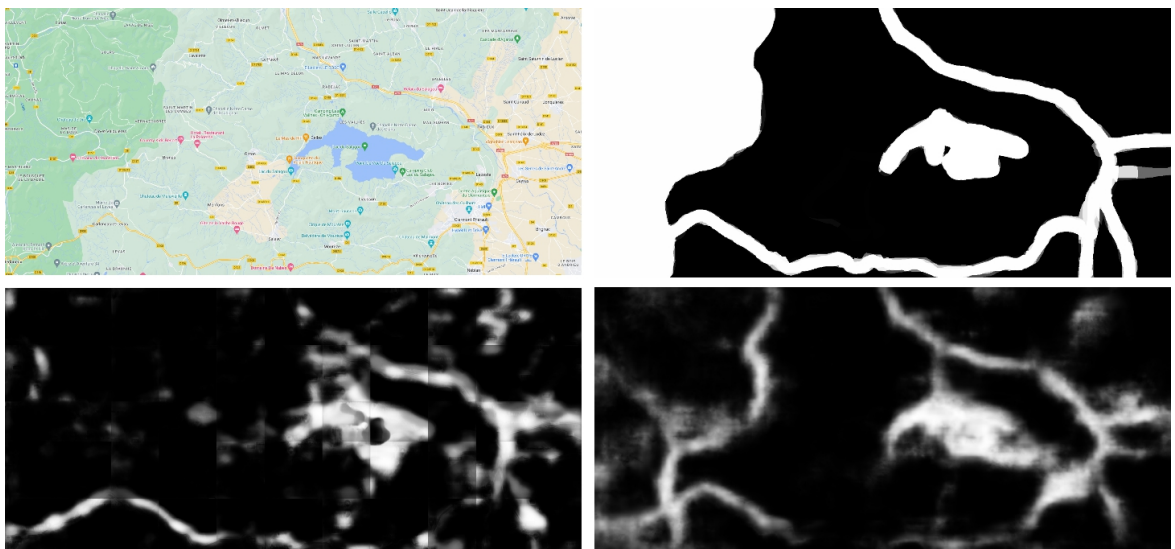
✿ Figure 38 : zone 11



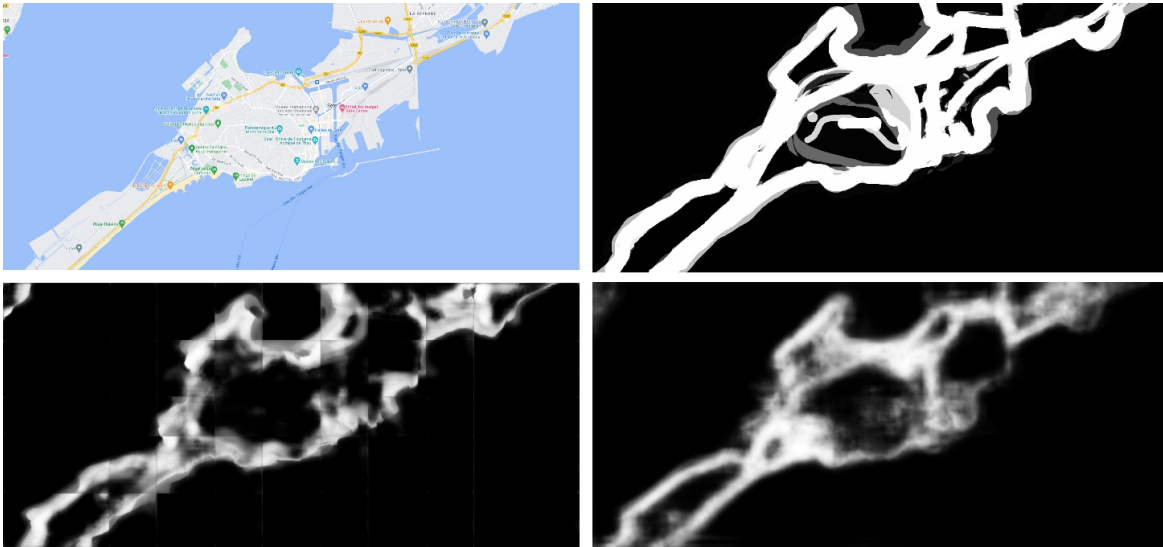
✿ Figure 39 : zone 12



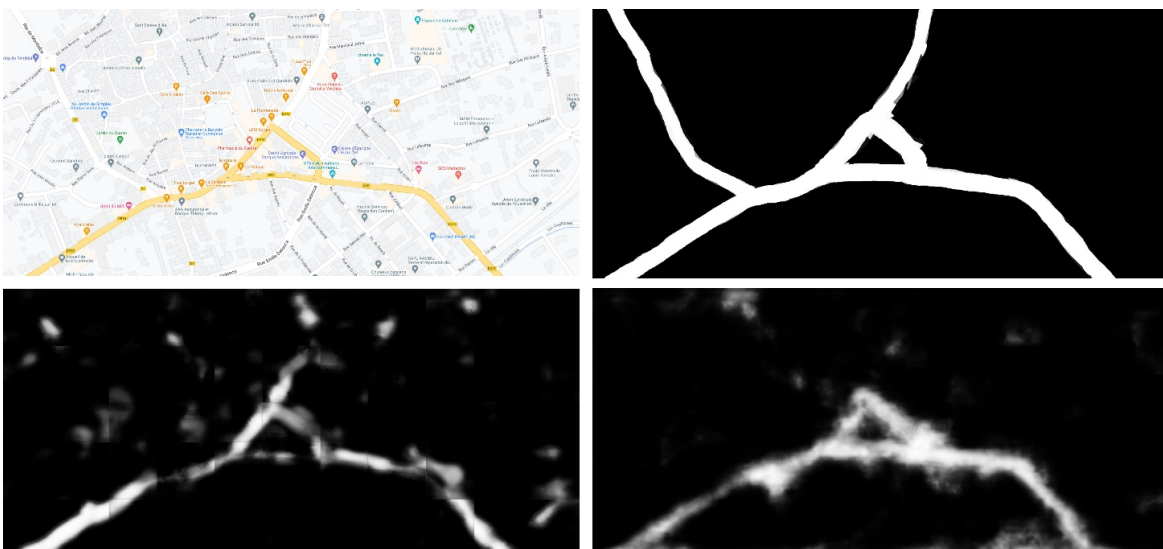
✿ Figure 40 : zone 13



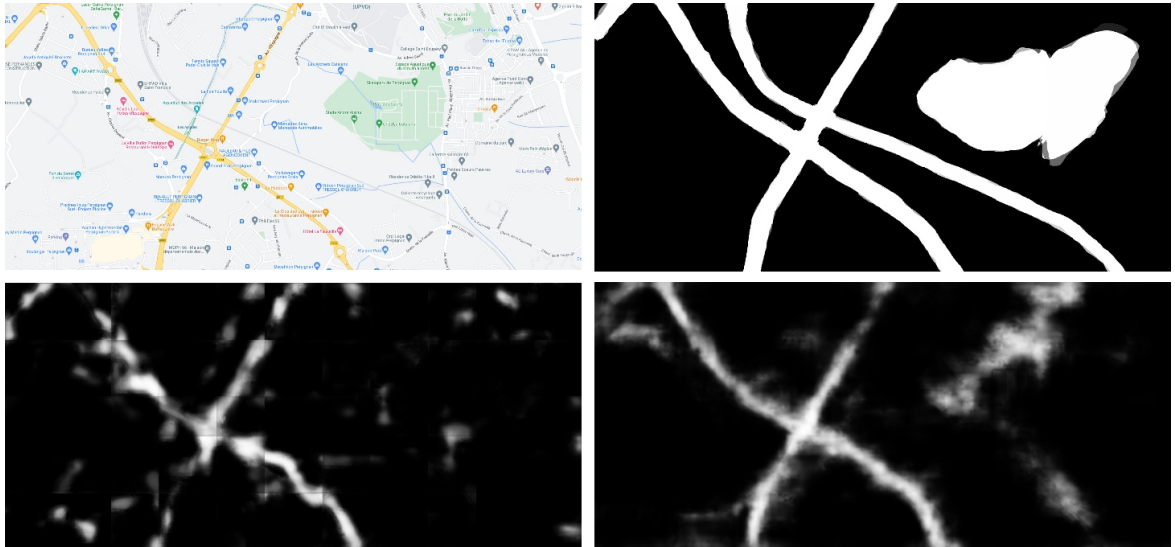
✿ Figure 41 : zone 14



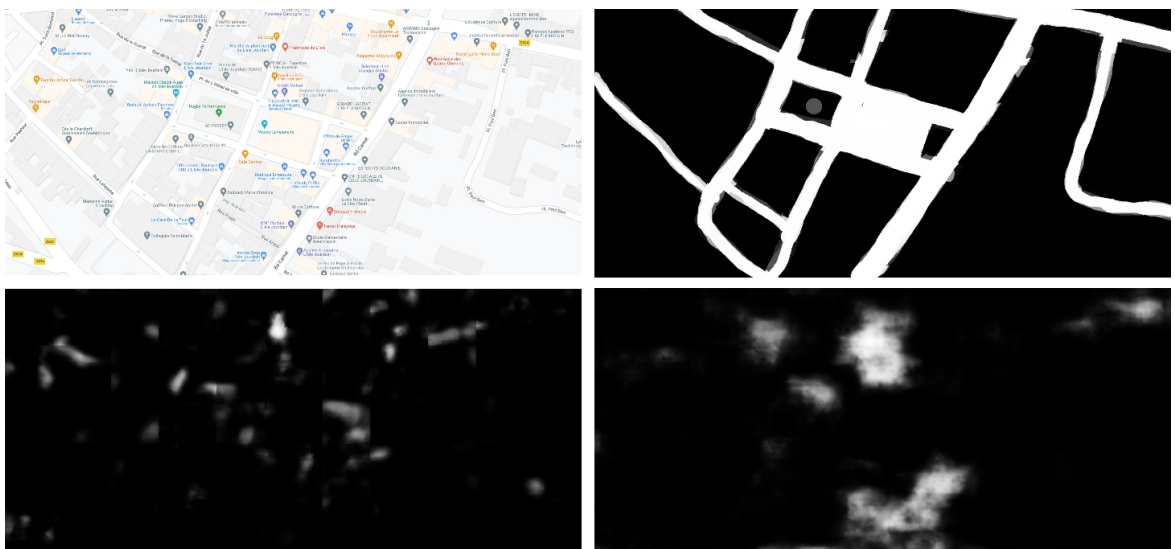
✿ Figure 42 : zone 15



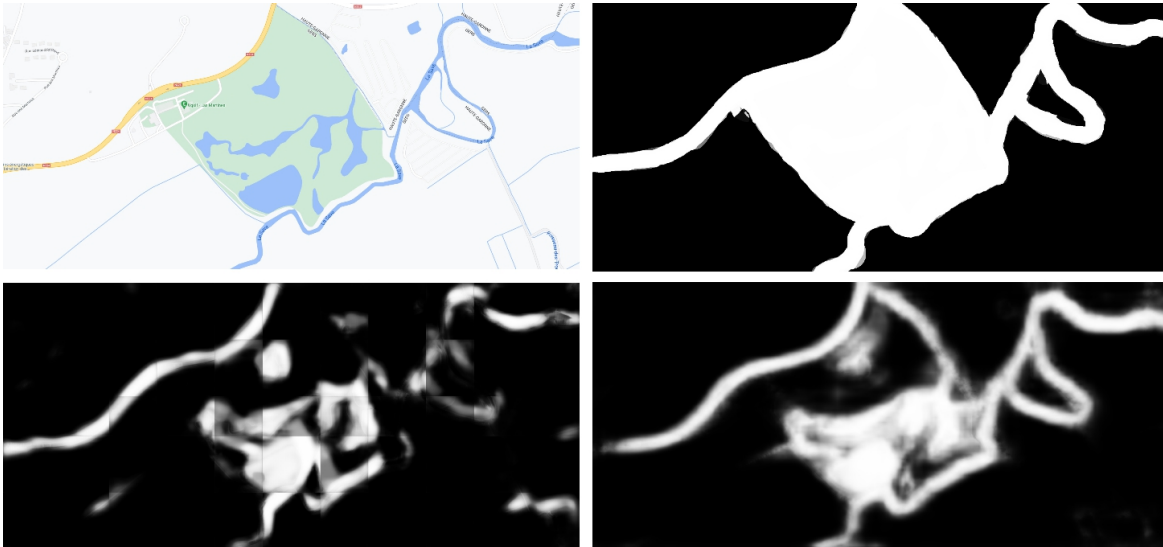
✿ Figure 43 : zone 16



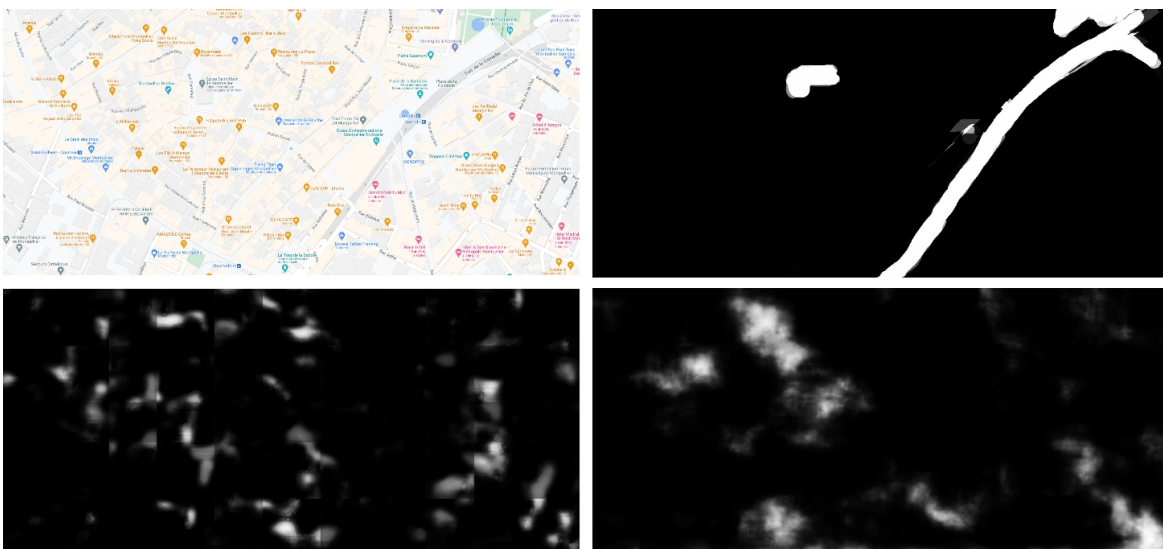
✿ Figure 44 : zone 17



✿ Figure 45 : zone 18



✿ Figure 46 : zone 19



✿ Figure 47 : zone 20

Résumé substantiel : Exploration du potentiel de la détection de points de repère pour limiter la désorientation dans les cartes pan-scalaires

Contexte

Cette thèse s'intéresse à la conception des cartes pan-scalaires, c'est-à-dire des cartes digitales interactives et zoomables, à l'image de Google Maps. Ces cartes sont aujourd'hui omniprésentes et utilisées quotidiennement par des milliards de personnes, pourtant elles demeurent relativement nouvelles dans la grande histoire de la cartographie, ce qui laisse à penser qu'elle peuvent encore faire l'objet d'amélioration. Bien que les concepteurs de cartes puissent s'appuyer sur certaines bonnes pratiques issues de la cartographie papier, la conception d'une carte pan-scalaire (digitale, donc) soulève des défis spécifiques. Elle mobilise des savoirs issus de la cartographie classique, de l'ergonomie des interfaces et des sciences cognitives. Intégrer l'ensemble de ces disciplines pour créer une carte efficace et cohérente est une tâche complexe, et il n'est pas rare d'observer des défauts de conception.

L'un des phénomènes engendré par ces défauts nous intéresse particulièrement : le phénomène de désorientation que les utilisateurs peuvent ressentir lors d'un changement d'échelle, notamment lorsque la représentation des données change brusquement lors de la transition d'une échelle à une autre.

Le projet Lost In Zoom (LiZ), auquel cette thèse contribue, propose d'explorer des pistes pour atténuer ce phénomène de désorientation (TOUYA 2020b). En s'appuyant sur la théorie des points d'ancrage cognitifs, nous faisons l'hypothèse que certains éléments cartographiques, visibles et reconnaissables à plusieurs niveaux de zoom, peuvent aider les utilisateurs à se repérer et s'orienter lorsqu'ils naviguent dans une carte en fournissant une continuité spatiale. Nous appelons ces éléments saillants "points de repère pan-scalaires" (pan-scalar landmarks) et ils sont au centre de ce travail de thèse.

Objectifs et approche

Les objectifs de cette thèse sont d'enquêter sur ce qui fait que certains éléments dans les cartes constituent des points de repère tandis que d'autres non, d'identifier ces éléments dans les cartes actuelles, et de réfléchir à comment nous pourrions transformer certains éléments pour qu'ils deviennent des points de repères. Nous pensons que ces informations et réflexions pourraient aider les cartographes dans leurs choix de généralisation et de hiérarchisation visuelle des éléments de la carte, ou être intégrées dans des algorithmes de généralisation automatique.

Pour répondre à ces objectifs, nous combinons deux approches. D'une part nous proposons des expériences avec des utilisateurs pour mieux comprendre comment ils

perçoivent les cartes, et d'autre part nous entraînons des modèles d'apprentissage supervisé à détecter différents types de points de repères.

Plutôt que de viser des performances de pointe en modélisation, ce travail propose d'ouvrir de nouvelles pistes de recherche et d'évaluer la faisabilité d'une cartographie informée par la perception qui inclurait des informations sur la saillance cognitive des objets cartographiques, tout en explorant le rôle que l'intelligence artificielle pourrait y jouer.

Structure de la thèse et contributions

Revue de la littérature (chapitre A)

La première contribution de cette thèse est un état de l'art qui articule des notions issues du design cartographique, de la généralisation cartographique, des sciences cognitives, de l'interaction homme-machine et de la reconnaissance des formes (au sens large). Cette synthèse vise à offrir un point d'entrée pour les chercheurs et concepteurs intéressés par le rôle des points de repère dans la conception de cartes pan-scalaires en mettant en lumière les concepts clés et les questions centrales, tout en soulignant les limites des approches actuelles en terme de généralisation cartographique, qui privilégient souvent la simplification visuelle au détriment de la cohérence cognitive. Partant de cette base théorique, la thèse se poursuit ensuite par une série d'explorations plus ciblées.

Expérimentations préliminaires (chapitre B)

Nous avons commencé par une exploration systématique de Google Maps, d'OpenStreetMap (OSM) et du plan IGN, en naviguant à différentes échelles à travers des lieux urbains, semi-urbains et ruraux répartis dans le monde entier. À chaque niveau de zoom, nous avons identifié les principaux types de points de repère pan-scalaires pouvant aider à l'orientation et les avons classés dans un tableau. Cette étude nous a permis d'observer une certaine similarité entre les cartes : les réseaux de transport, les plans d'eau, les zones urbaines, les toponymes ainsi que certains espaces naturels incarnent les points de repère importants de chacune d'entre elles.

Nous nous sommes ensuite interrogés sur le nombre optimal de points de repère et d'indices visuels à faire apparaître dans une carte pan-scalaire. Nous avons pour cela proposé une expérience de pensée pour réfléchir au minimum d'éléments visuels nécessaires pour que l'utilisateur perçoive les changements d'échelle. Il en ressort qu'un seul élément est souvent insuffisant, mais qu'un deuxième suffit généralement à en compléter la lecture. Concernant le nombre maximum de points de repère, la littérature montre que ces derniers se font concurrence pour capter l'attention de l'utilisateur, et que le nombre de points de repère mémorisable est limité lors de la navigation. Il reste cependant à adapter ces conclusions au cas des cartes pan-scalaires.

Dans notre dernière étude préliminaire, nous avons retracé l'évolution des interactions disponibles dans les cartes pan-scalaires depuis 2005. Nous avons noté une convergence vers des interactions uniformisées entre ordinateur et téléphone, telles que le glisser-déposer ou le zoom centré via la molette de la souris ou le pincement

sur écran tactile. Nous avons aussi identifié un schéma récurrent d'usage de la carte, alternant entre des phases de transition et des phases statiques, que nous avons formalisé sous forme d'un modèle dans l'espoir qu'il serve à améliorer les interfaces et interactions futures.

Étude sur la perception des tours de villes (chapitre C)

Dans la seconde partie de notre thèse, nous avons commencé par explorer la perception des routes périphériques (ou "rocales" ou "tours de villes", "ring roads" en anglais) dans les cartes topographiques. Leur géométrie distinctive, associée à leur fonction dans la connectivité urbaine, en fait des candidates naturelles pour servir de points de repère. Pour les exploiter dans les cartes pan-scalaires, il est cependant nécessaire de pouvoir les détecter automatiquement et d'enregistrer cette information dans des bases de données.

Bien que des techniques existent pour identifier les routes circulaires, le concept plus flou de tour de ville reste peu étudié, bien qu'il présente selon nous un fort potentiel d'ancrage cognitif. Pour combler cette lacune, nous avons mené une enquête de perception auprès de 28 participants. Celle-ci a montré que la perception des tours de villes dépend de multiples facteurs contextuels et qu'elle peut être renforcée ou affaiblie par la présence d'autres entités. Malgré cette complexité, nous avons pu établir une liste de propriétés caractéristiques des tours de villes, avec des indicateurs mesurables issus de la littérature pour chacune. Ces résultats ont été synthétisés sous forme d'une ontologie conceptuelle, posant les bases d'un cadre structuré pour les recherches futures.

Cette étude offre un exemple de la manière dont les observations issues des utilisateurs peuvent inspirer des stratégies formelles de détection de points de repère spécifiques. La suite logique dans notre cas serait de traduire l'ontologie proposée en un algorithme permettant de calculer automatiquement le "degré de rocadité" (ringroadness) d'un ensemble de routes.

Détection de bâtiments remarquables par apprentissage multimodal (chapitre D)

Dans ce chapitre, nous présentons un projet mené en collaboration avec des étudiants de master portant sur la détection de bâtiments remarquables avec de l'intelligence artificielle. Notre idée était d'utiliser des images de carte centrées sur chaque bâtiment en complément des données tabulaires pour l'entraînement de nos modèles afin d'améliorer leur performance.

Nous avons commencé par produire manuellement un jeu de données d'environ un millier de bâtiments remarquables centrés sur la ville de Nantes, qui a été mis en ligne en libre accès. Ensuite, nous avons entraîné un modèle Random Forest et un MLP sur le jeu de données tabulaire uniquement afin d'avoir une base de comparaison. Ensuite nous avons entraîné un réseau de neurones convolutif (CNN) sur les images de cartes centrées sur chaque bâtiment de notre base de données. Enfin, ces deux modalités ont été combinées dans une architecture multimodale simple, afin d'évaluer si l'ajout d'un contexte spatial issu des cartes permettait d'améliorer les performances de classification des modèles uniquement basés sur les informations tabulaires.

Les résultats se sont révélés mitigés. En moyenne, le modèle multimodal a ob-

tenu de moins bons résultats que le Random Forest, malgré un accès à des informations visuelles supplémentaires. Toutefois, certaines zones de validation ont montré de meilleures performances pour le modèle multimodal. Cette observation, couplée aux travaux issus d'autres domaines, suggère malgré tout que cette approche pourrait quand même être prometteuse dans certains contextes spécifiques qu'il reste à distinguer.

Segmentation des zones à fort potentiel d'ancrage (chapitres E et F)

Dans la dernière partie du manuscrit, nous introduisons le concept de zones à fort potentiel d'ancrage (AHAP - Areas of High Anchoring Potential) : des régions d'une carte pan-scalaire susceptibles de contenir un ou plusieurs points de repère, et qui peuvent servir de références cognitives lors de la navigation à travers l'espace ou les niveaux de zoom. Nous pensons que réussir à détecter ces régions dans les images de cartes pourrait non seulement faciliter la détection des points de repère en base de données, mais également fournir une clé d'analyse supplémentaire pour comprendre la manière dont les utilisateurs perçoivent et mémorisent les cartes.

Pour tenter de détecter les AHAPs, la première étape de ce projet a été de constituer un jeu de données d'entraînement grâce à un outil d'annotation en ligne développé spécifiquement pour le projet LostInZoom. Environ 200 images de cartes ont été annotées puis découpées en tuiles afin d'entraîner nos modèles de segmentation.

Pour évaluer la performance de nos modèle et évaluer le consensus entre différents annotateurs, nous avons constitué un corpus de 20 cartes couvrant divers lieux et échelles en France. Chacune a été annotée par 15 participants invités à surligner les AHAPs qu'ils percevaient. Nous avons ensuite utilisé la méthode STAPLE pour agréger les annotations et produire un consensus pour chaque image, tout en calculant des métriques pour quantifier l'accord inter-participants.

Nous avons ensuite entraîné deux modèles de segmentation. Le premier, basé sur l'architecture U-Net, a obtenu des résultats modestes. Le second, un modèle basé sur l'architecture Segformer combinant transformeurs et mécanismes d'attention, a montré des performances plus prometteuses, à la fois visuellement et quantitativement — en particulier si l'on considère la difficulté de la tâche et l'accord partiel entre annotateurs.

Ces résultats suggèrent que les modèles à base de transformeurs, grâce à leur capacité à capturer des dépendances spatiales étendues, pourraient être pertinents pour des applications comme la détection de points de repères ou l'identification de zones critiques en cartographie.

Conclusions

Certains enseignements ont particulièrement émergé de nos études et expérimentations, et méritent d'être soulignés ici.

Concernant la nature des points de repères pan-scalaires, nous constatons d'abord qu'ils ne sont que rarement perçus comme des entités isolées. Leur potentiel d'ancrage cognitif dépend souvent de leur relation avec les éléments environnant, de leur contraste visuel avec l'arrière-plan ou de leur rôle dans l'organisation spatiale globale de la carte. Cela a été particulièrement visible dans l'étude sur les tours de villes,

où certaines routes gagnaient ou perdaient en saillance selon qu'ils entouraient ou non un centre urbain ou si ils étaient accompagnés d'éléments renforçants comme les rivières ou les voies ferrées. Cette observation renforce une lecture "gestaltiste" de la perception cartographique, où des principes comme la proximité, la continuité ou la clôture structurent les cartes mentales des utilisateurs. Par ailleurs, on constate que certains points de repère pan-scalaires comme les "centres-villes" n'ont pas de délimitations claires. Cette ambiguïté appelle des définitions plus souples et contextuelles pour ce type d'éléments cartographiques. Enfin, nous noterons que la saillance des points de repère évolue selon les échelles : certains éléments perdent en visibilité lorsqu'ils sont généralisés, tandis que d'autres deviennent cognitivement pertinents précisément grâce à leur reconnaissabilité à travers différents niveaux de zoom. La compréhension des points d'ancrage cognitifs nécessite donc une approche multi-échelle.

Concernant l'apprentissage automatique, il semble offrir des perspectives intéressantes pour modéliser les caractéristiques complexes et parfois ambiguës de la perception cartographique. Les modèles récents montrent qu'il est possible d'apprendre à partir de données approximatives ou dépendantes d'un contexte cartographique, ce qui rejoint la manière dont les humains perçoivent les cartes. Cela dit, les machines ne pensent pas comme les humains, et une bonne performance dans la détection de points de repère ne signifie pas nécessairement que les modèles traitent les relations spatiales comme un cerveau humain le ferait. Ce décalage invite à la prudence dans l'interprétation cognitive des résultats des modèles.

Enfin, bien que cette thèse se concentre sur l'optimisation de l'orientation des utilisateurs, nous tenons à rappeler que le design des points de repère n'est pas uniquement une question d'ergonomie et de cognition : il est aussi influencé par des logiques politiques, commerciales ou institutionnelles et les éléments mis en valeur sur une carte reflètent souvent les priorités de ceux qui la produisent, parfois au détriment de la lisibilité.

Perspectives de recherches futures

Cette thèse ouvre plusieurs pistes de recherche dans la continuité directe des expériences menées, et pose les bases pour l'exploration de nouveaux axes prometteurs.

L'ontologie des tours de villes, élaborée dans le chapitre C, constitue une base solide pour une modélisation plus systématique de ces structures. La mise en œuvre informatique des indicateurs proposés permettrait de tester leur capacité à détecter automatiquement les tours de villes à l'échelle nationale ou continentale, validant ainsi leur pertinence tout en fournissant des outils concrets pour l'intégration dans les chaînes de production cartographiques.

Dans un autre registre, les premières expériences en apprentissage multimodal pour la détection de bâtiments remarquables pourraient être enrichies par l'introduction de données visuelles plus variées, telles que les images satellites ou les photographies de rue. Ces sources offriraient un contexte visuel plus riche et pourraient contribuer à une meilleure distinction entre bâtiments ordinaires et remarquables.

De leur côté, les travaux sur les zones à fort potentiel d'ancrage (AHAPs) pourraient évoluer vers une classification plus fine : en élargissant et diversifiant les données d'annotation, il deviendrait envisageable d'entraîner des modèles capables de

reconnaître différentes catégories d'AHAPs. Cela ouvrirait la voie à l'analyse d'ancres composites, dans lesquels plusieurs éléments cartographiques se renforcent mutuellement pour devenir des points de repère.

Au-delà de ces prolongements directs de la thèse, d'autres perspectives existent pour détecter les points de repères potentiels dans les cartes. L'exploitation de données textuelles, par exemple, représente un champ de recherche en pleine expansion. Une part croissante de l'information géographique est aujourd'hui véhiculée par le texte, notamment sur les réseaux sociaux. L'extraction de lieux ou de relations spatiales à partir de ces textes non structurés, parfois avec l'appui des grands modèles de langues (LLMs), pourrait dans le futur être intégré dans les processus de construction des cartes pan-scalaires.

Par ailleurs, les progrès récents des grands modèles multimodaux comme ChatGPT ouvrent des perspectives inédites. Ces modèles commencent à reconnaître des éléments cartographiques complexes directement dans les images de cartes, et leur performance progresse rapidement. À moyen terme, ils pourraient assister les cartographes dans l'identification de points de repère potentiels ou dans la détection de ruptures de continuité visuelle entre les niveaux de zoom, facilitant la conception semi-automatisée de cartes cohérentes à travers les échelles.

Cependant, ces perspectives technologiques ne doivent pas faire oublier les enjeux éthiques et environnementaux associés. Le coût énergétique du développement et du déploiement des modèles d'intelligence artificielle à grande échelle est désormais un problème avéré. Si l'enthousiasme pour ces technologies repose souvent sur leur efficacité, il est essentiel de s'interroger sur leur réelle nécessité pour chaque tâche, ainsi que sur les alternatives moins énergivores envisageables. Une telle réflexion implique de repenser les choix techniques et appelle à une collaboration interdisciplinaire entre chercheurs en IA, philosophes analytiques, scientifiques de l'environnement... et cartographes évidemment.

