Modeling, Learning, and Transferring Anatomical Representations in Medical Imaging using AI

Pietro Gori

INSTITUT POLYTECHNIQUE DE PARIS

NNT : XXXXXXXX

Habilitation à diriger les recherches

Institut Mines-Télécom

TELECOM Paris

IP PARIS

Modeling, Learning, and Transferring Anatomical Representations in Medical Imaging using AI

Thèse d'habilitation à diriger les recherches de l'Institut Polytechnique de Paris
préparée à Télécom Paris

École doctorale n°626 de l'Institut Polytechnique de Paris (EDIPP)
Spécialité : Informatique, Données et IA

Thèse présentée et soutenue à Palaiseau, le 05 Septembre 2024, par

PIETRO GORI

Composition du Jury :

Olivier Bernard Full Professor, INSA, Lyon	Rapporteur
Francois Rousseau Full Professor, IMT Atlantique	Rapporteur
Ender Konukoglu Full Professor, ETH Zürich	Rapporteur
Diana Mateus Full Professor, Ecole Centrale Nantes	Examinatrice
Caroline Petitjean Full Professor, University of Rouen Normandy	Examinatrice
Bertrand Thirion Research Director, Inria, Paris-Saclay	Examinateur

Summary

Introduction

Recent advances in computer vision and statistical learning, particularly in deep learning, have fostered research in (anatomical) medical imaging in recent years.[1]. However, it has been reported that State-Of-The-Art (SOTA) algorithms from computer vision do not necessarily show the same performance when employed on natural or medical imaging applications [Dufumier et al., 2021a], and that a simple supervised pre-training from ImageNet, as largely employed for natural images, does not necessarily work well on medical tasks [Mustafa et al., 2021, Matsoukas et al., 2022, Raghu et al., 2019, Konz et al., 2022, Konz and Mazurowski, 2024]. The main reasons that have prevented a simple and naive transfer between natural and medical imaging applications are:

- The important "visual" domain gap between natural and medical images (i.e., a 2D image of a cat is quite different from a 3D volume of a brain).
- The different relevant and irrelevant sources of variation (i.e., geometric factors, like size, scale or location are usually irrelevant discriminative factors for natural images whereas they can be very important for medical problems).
- The link between pixel intensity and physics (e.g., differently from natural images, intensity values in X-ray or CT scans have a precise anatomical meaning).
- The difference in size between the datasets (e.g., ImageNet [Deng et al., 2009] has more than 14 million images whereas medical imaging datasets have usually around 1-2 thousands images).
- The more subtle differences distinguishing clinical groups (i.e., a cat is easy to distinguish from a car for a human eye, whereas identifying a schizophrenic subject from a brain MRI scan is difficult even for a radiologist).
- The prior knowledge about the information content (i.e., prior medical knowledge can enrich and/or accurately describe the information content of a medical image and improve the downstream task performance. Prior knowledge about natural objects is not usually known or important for the final downstream task.)

Based on that, I have focused my research efforts on developing AI methods that answer specific needs and constraints of the medical imaging data (e.g., low-data regime, data biases, physical specificity, lack of labeling data) leveraging clinical knowledge and (healthy) unlabeled data. From a methodological perspective, my research has followed three main axes (highlighted in Fig.1):

1. *Modeling* medical knowledge and anatomy and integrating it into machine learning models.
2. *Learning* compact, relevant and explanatory representations of anatomical imaging data.
3. *Transferring* anatomical representations between domains (*i.e.,* different modalities, data-sets, populations) to increase downstream performance (*e.g.,* segmentation, classification, regression) or to discover new pathological biomarkers.

[1] In this manuscript, we will always refer to anatomical imaging modalities, in particular CT and MR scans, if otherwise stated.

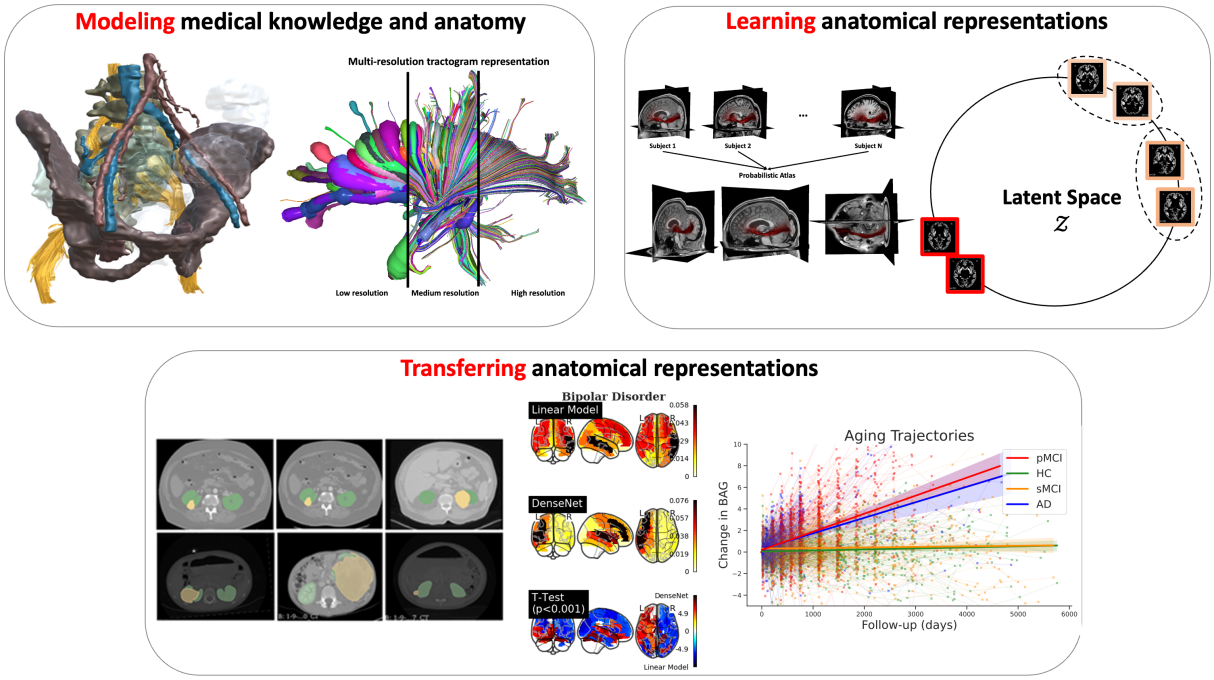


Figure 1: Schematic view of my main three research axes.

In terms of clinical applications, I have only worked with anatomical data, MRI and CT scans, from the brain, chest, abdomen and pelvic area. Please note that I also had a collaboration with l'Oreal, where we worked on makeup synthesis and transfer, and thus with natural images of human faces [Kips et al., 2020, Kips et al., 2021, Kips et al., 2022a, Kips et al., 2022b]. In the following chapters, I will present some of my work using brain MRI data based on the three clinical applications I have worked on the most. In the last part of this manuscript, I will describe perspectives related to new research directions, novel clinical applications and research valorization of previous works.
The first three chapters will be about:

1. *Detection of anatomical signatures predictive of brain disorders* in collaboration with Neurospin (CEA).
   In this Chapter, I will present new representation learning methods for the identification of new anatomical biomarkers prognostic of clinical course of psychiatric disorders, such as schizophrenia, autism and bipolar disorder, and for parsing their heterogeneity.

2. *Glioblastoma atlas estimation* in collaboration with MAP5 and St. Anne hospital.
   In this Chapter, a new framework for estimating a 3D atlas of glioblastoma using MR brain images will be presented. To reach such a challenging goal, new methods for segmenting and registering brain MR images with tumors will be presented.

3. *Brain white matter tractogram analysis* in collaboration with LIX, ENS Paris-Saclay and St. Anne hospital.
   In this Chapter, I will describe a new multi-scale, geometric representation for fast, robust and reliable processing, comparison and visualization of white matter tractograms of the brain. Two new segmentation methods, based on symbolic AI and optimal transport, will also be presented.

Chapter 1

Detection of anatomical signatures predictive of brain disorders

This chapter has been published in [Dufumier et al., 2021b, Dufumier et al., 2021c, Dufumier et al., 2023, Dufumier et al., 2024, Barbano et al., 2023a, Barbano et al., 2023b, Louiset et al., 2021, Louiset et al., 2024b, Louiset et al., 2024a, Carton et al., 2024] and is based on the PhD theses of B. Dufumier and R. Louiset, co-directed with E. Duchesnay (NeuroSpin, CEA), and C. Barbano, co-directed with M. Grangetto (University of Turin) and I. Bloch (Télécom Paris). Part of the work presented here has also been produced during the Post-Doc of F. Carton.

Contents

1.1 Context

The physio-pathology of mental and neurodevelopmental disorders, like schizophrenia and autism spectrum disorders, as well as neurodegenerative diseases, like Alzheimer's disease, is still poorly understood. Brain disorders can be complex and highly heterogeneous, showing clinical, biological, and environmental inter-subjects variations [Wolfers et al., 2018], that make their neurobiological characterization even more challenging. Furthermore, in most cases, there is currently a lack of objective quantitative measures to guide the clinician in choosing the right therapeutic treatment.

Finding anatomical patterns characterizing a disease could increase our understanding of its pathological mechanisms and pave the way towards a personalized medicine for brain disorders. Machine learning, and in particular deep learning (DL), have the potential to automatically learn such patterns. Indeed, there is now a consensus on the benefit of DL in addressing many medical

imaging tasks, such as object detection and image segmentation. However, its performance in single-subject predictions, based on neuroanatomical data, has not yet achieved the expected results (*i.e.*, AUC $\geq$ 90) [Dufumier et al., 2021a]. Furthermore, recent studies [Schulz et al., 2020, Peng et al., 2021, Abrol et al., 2021] yielded contradictory results when comparing DL with Standard Machine Learning (SML) on top of classical feature extraction [Dufumier et al., 2024].

The emergence of large-scale neuroimaging datasets, like UKBioBank [Bycroft et al., 2018], HCP [Van Essen et al., 2013], ABIDE [Di Martino et al., 2014] and OpenBHB [Dufumier et al., 2022], gives a unique opportunity for studying the neuroanatomical signatures of such disorders. However, these datasets contain mostly healthy subjects and there is still a lack of pathological imaging data. Indeed, most of the current (and past) neuro-anatomical research works presented results based on a training set composed of less than 2k pathological imaging data. Furthermore, the study of neurodegenerative and psychiatric disorders involves the use of various data modalities to better understand the underlying pathological mechanisms, identify biomarkers, and develop effective treatments. Besides the "prior" clinical and medical knowledge of the disease, as described in medical and anatomical books, the most used data modalities include: a) several non-invasive imaging modality, such as Magnetic Resonance Imaging (MRI), Compued Tomography (CT) scans or electroencephalography (EEG); b) genetic and omics data (*e.g.,* genomics, transcriptomics, proteomics, and metabolomics); c) cognitive and behavioral assessments and d) digital health recordings (*e.g.,* smartphone apps, wearable devices, and remote monitoring systems). Only by integrating data from all these modalities, it would be possible to have an holistic view of the disease and gain a comprehensive understanding of its underlying mechanisms, which could ultimately lead to improved diagnosis, treatment, and management strategies [Schulz et al., 2024].

1.2 Challenges

When working with medical data in a "clinical" context, the first challenges comprise: the choice of the acquisition protocol, data quality assessment, data cleaning, data anonymization and harmonization. All these tasks are usually time-consuming and require large and experienced manpower.

To reach a larger scientific community and foster methodological development, well adapted to the medical imaging problems, large "research" datasets were made available. Differently from "clinical" datasets, "research" datasets are usually: anonymized, quality checked, accessible and quite homogeneous. In the following, we will focus *solely* on "research" datasets, discussing some of the most important challenges and research questions that we worked on.

Small pathological datasets

The first challenge concerns the small number of pathological samples. In supervised learning, when dealing with a small labeled dataset, the most used and well-known solution is supervised Transfer Learning from ImageNet (or other large vision datasets). However, it has been recently shown that this strategy is useful, namely features are re-used, only when there is a high visual similarity between the pre-train and target domain (*e.g.,* low Fréchet inception distance (FID)) [Mustafa et al., 2021, Neyshabur et al., 2020, Raghu et al., 2019, Matsoukas et al., 2022, Li et al., 2024]. This is not the case when comparing natural and medical images. Furthermore, many medical images, and in particular brain MRI scans, are 3D volumes, differently from the 2D images of ImageNet. This entails a great domain gap between the large labeled datasets used in computer vision and medical images.

Another approach, usually employed when the labeling procedure is complex, time-consuming

and/or costly, comprises self-supervised learning (SSL) methods. This class of methods leverages an annotation-free pretext task to provide a surrogate supervision signal for feature learning. Pretext tasks should only use the visual information and context of the images and recent examples comprise: context prediction [Doersch et al., 2015], generative models [He et al., 2022, Donahue and Simonyan, 2019], contrastive learning [Chen et al., 2020], teacher/student methods [Grill et al., 2020], information maximization [Zbontar et al., 2021, Bardes et al., 2022]. Nonetheless, these methods still need large (unannotated) datasets, which should comprise, to reduce the domain gap, data similar to the ones in the (labeled) target dataset, namely pathological patients. However, the large majority of images currently stored in hospitals and clinical laboratories belong to healthy subjects. Indeed, the largest datasets currently available (*e.g.,* UKBioBank [Bycroft et al., 2018] and OpenBHB [Dufumier et al., 2022]) mostly contain data of healthy subjects. Furthermore, these datasets usually comprise one or multiple imaging modalities, as well as clinical data, such as age, gender and weight. The research challenge thus becomes how to leverage large datasets of healthy subjects and combine the heterogeneous sources of information (i.e., clinical and imaging data) to improve the diagnostic and understanding of brain disorders.

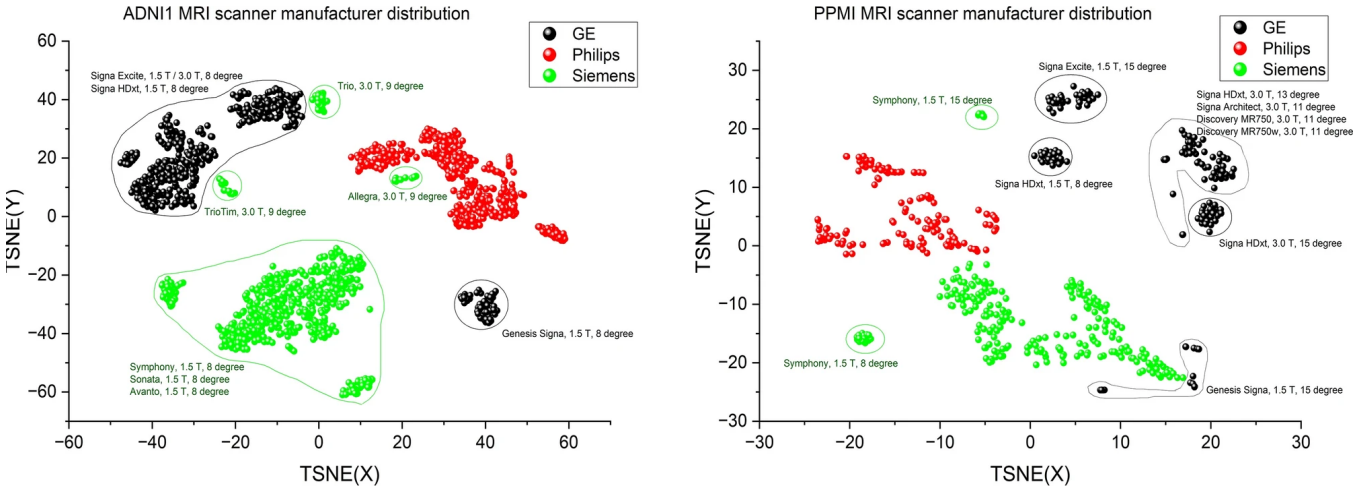


Figure 1.1: Example of site effect: a simple clustering on imaging features primarily finds the scanner manufacturers. MRI volumes from ADNI1[a] (left) and PPMI[b] (right) datasets are first pre-processed using MRQy [Sadri et al., 2020] and then mapped to a 2D space using the t-SNE method [Van der Maaten and Hinton, 2008]. Image taken from [Kushol et al., 2023].

a. https://adni.loni.usc.edu/ b. https://www.ppmi-info.org/

Data biases - site effect

A second challenge concerns the data biases. In our work, we define data biases as the visual patterns that correlate with the target task and/or are easy to learn, but are not relevant for the target task. For instance, the *site effect* in MRI images refers to systematic variations or discrepancies in feature distributions across different imaging sites, that arise from differences in equipment, protocols, or settings , and are not related to a disease (i.e., target task) [Bayer et al., 2022]. When working with MRI samples in a binary classification problem (healthy Vs patients), these spurious differences can be visually more accentuated, and thus easy to learn, than the relevant differences between the two classes (see Fig. 1.1). This can result in a biased model, whose predictions majorly rely on the bias attributes and not on the true, generalizable, and discriminative features. In addition, it has been shown that neural networks tend to rely on simple and easy-to-learn patterns to make their

decisions [Geirhos et al., 2019, Li et al., 2021c]. This means that a network trained on a dataset comprising images from a single acquisition site might have a drop in performance when being tested on images from another acquisition sites, as shown in [Wachinger et al., 2021, Glocker et al., 2019]. Furthermore, most classification networks are based on the minimization of the cross-entropy loss, which can be affected by biases in the data, as shown in [Alvi et al., 2018, Kim et al., 2019, Sagawa et al., 2019, Tartaglione et al., 2021, Torralba and Efros, 2011], or suffer from noise and corruption in the labels [Elsayed et al., 2018, Graf et al., 2021].

Several harmonization methods were proposed for multi-site MRI samples. Two of the most used techniques are two linear mixed models: Linear Adjusted Regression [Wachinger et al., 2021] and ComBat [Johnson et al., 2007, Fortin et al., 2018, Bayer et al., 2022, Marzi et al., 2024],which adds a multiplicative non-linear effect on the residual noise. These models generally require to have access to all imaging sites during training, which might not always been the case and it prevents a correct analysis of the generalization error on external sites (*i.e.,* not used during training). Recently, more advanced learning-based methods have started to emerge, which are usually based on image-to-image translation (IIT) and/or (unsupervised) domain adaptation (DA) techniques. First methods proposed to translate images across sites using generative deep learning methods. These methods are mainly based on paired/"travelling heads" datasets [Zhao et al., 2019, Dewey et al., 2019]. To avoid learning pairwise mappings between sites, which would require learning $N(N-1)$ mappings for N sites, a unified but disentangled representation can be learned across sites [Zuo et al., 2021, Liu and Yap, 2024]. The representation of each image is decomposed into anatomical content, invariant across sites, and appearance style (e.g., intensity and contrast), which depends on the site. Generative methods need large training datasets and are difficult to train and validate since there is still a lack of metrics that can accurately evaluate the quality of generated MR images and that are broadly accepted by the medical scientific community.

Other authors have proposed to avoid generating images and directly learn a scanner-invariant representation, thus having a single network for all sites. Different strategies have been proposed based, for instance, on variational autoencoders (VAE) [Moyer et al., 2020] and domain adaptation techniques [Dinsdale et al., 2021]. These strategies are always based on a (hidden) implicit hypothesis to drive the disentanglement between site-invariant anatomical content, which should be relevant to the downstream task, and the spurious information specific to each site. For instance, the disentanglement can be driven by the reconstruction error [Moyer et al., 2020] or by a domain classifier [Dinsdale et al., 2021]. In both cases, the optimization loss, network architecture and training procedure can be difficult to choose and might not be adapted to the subtle imaging patterns characterizing brain disorders. Furthermore, differently from the previous IIT methods, these strategies are trained for a specific downstream task and can not be used in a generic context. Indeed, IIT methods can be employed as pre-processing to transform all images to a specific site and then run any algorithm (e.g., segmentation, regression, classification) trained on images from that specific site.

Mental disorders are heterogeneous

In psychiatry, mental disorder diagnoses are typically established through a combination of interviews, questionnaires, and observations. These evaluations are designed to ascertain the presence, severity, frequency, and duration of psychiatric symptoms, which are subsequently linked to specific mental disorders according to standardized classification systems, such as the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [American Psychiatric Association and American Psychiatric Association, 2013] or the International Classification of Diseases (ICD-11) [Harrison et al., 2021]. Even if these systems are constantly modified to refine the etiology of mental disor-

ders, they still exhibit significant heterogeneity in their symptoms presentations. This has raised concerns about the reliability of a nosology only based on the assessment of exterior cognitive, behavioral, emotional, and physical symptoms [Insel and Quirion, 2005]. There is a need for quantitative, consistent and reliable biomarkers based on the anatomy and functioning of the brain that, together with current behavioral and cognitive assessments, could improve the definition of mental disorders. In particular, we could better parse their heterogeneity and thus pave the way towards a more accurate therapeutic strategy.

Existing methods can be divided into two groups, based on their underlying hypothesis about the pathological heterogeneity: categorical (i.e., subtypes) or continuous (i.e., dimensional). The former assumes that pathological subtypes form distinct clusters in the feature space, well separated among them and from the healthy population. The second group of methods assume instead that there is a continuum between the healthy population and the different subtypes, and it focuses on estimating the (latent) feature dimensions that better describe the pathological heterogeneity.

Finding consistent and reproducible subtypes within a mental disorder is complex since the most important anatomical variations within a pathology are shared with the normal, healthy population. This means that using a standard clustering algorithm, such as K-means, within pathological patients might result in clusters that actually reflect the healthy inter-individual heterogeneity, possibly due to confounds variables, such as age or sex.

When the effect of confounds variables is known, residualization methods may be used [Wachinger et al., 2021, Fortin et al., 2018, Glocker et al., 2019]. These approaches aim at producing neuroanatomical features that are not driven by aging or gender, for example. However, they are usually based on simplistic assumptions, such as a linear relationship between confounding factors and input features, and it is usually hard to know all confounding variables a priori. This can thus produce pathological subtypes that, even if they are based on covariate-adjusted neuroanatomical features, still exhibit variations shared with healthy subjects, as shown in [Iftimovici, 2021].

Another technique that can be used to parse the anatomical heterogeneity is normative modeling [Marquand et al., 2016, Marquand et al., 2019]. It first estimates a normative (i.e., reference) model of the healthy population with respect to pre-selected covariates, and then infer the deviation indices of the patients for each covariate. These deviation indices (or z-scores) can be further used as features to analyze the inter-individual heterogeneity not driven by the chosen covariates and thus identify pathological subtypes. However, as for residualisation methods, normative models are based on known covariate variables and they can not automatically reduce the dimensionality of the input data.

Instead than disregarding confounding covariates, that are not necessarily known or available, pathological heterogeneity can be parsed by *contrasting* it with the general, healthy population variability. In the following, we will discuss how this new perspective, combined with the learning capacity of deep learning, can be adapted for both hypotheses: categorical and continuum.

1.3 Contributions

In this Chapter, our main contributions are:

1. A geometric approach for contrastive learning that can be used in all settings: unsupervised (i.e., no labels), supervised (i.e., class labels) and weakly-supervised (i.e., weak attributes or regression). It is well adapted to integrate prior information, such as weak attributes or representations learned from generative models, and can thus be used to learn a representation of the healthy population by leveraging both clinical and imaging data.

2. Based on the proposed geometric approach, we show why recent contrastive losses (InfoNCE, SupCon, etc.) can fail when dealing with biased data and derive a new supervised contrastive loss and debiasing regularization loss, that work well even with extremely biased data.

3. Two Contrastive Subgroup Discovery methods, entitled UCSL and Deep UCSL. By contrasting controls with patients, we identify subgroups that stem only from the pathological variability specific to the disease, while disregarding the common variability shared with the controls.

4. Three new Contrastive Analysis methods based on: Variational AutoEncoders (VAE), Generative Adversarial Network (GAN) and Contrastive Learning.

1.3.1 Contrastive Learning - a geometric approach

Let $x \in \mathcal{X}$ be an original sample (i.e., anchor), x_i^+ a similar (positive) sample, x_j^- a dissimilar (negative) sample and P and N the number of positive and negative samples respectively. Contrastive learning methods look for a parametric mapping function $f : \mathcal{X} \to \mathbb{S}^{d-1}$ that maps "semantically" similar samples close together in the representation space (a (d-1)-sphere) and dissimilar samples far away from each other. Once pre-trained, f is fixed and its representation is evaluated on a downstream task, such as classification, through linear evaluation on a test set. In general, positive samples x_i^+ can be defined in different ways depending on the problem: using transformations of x (unsupervised setting) [Chen et al., 2020], samples belonging to the same class as x (supervised) [Khosla et al., 2020] or with similar image attributes of x (weakly-supervised) [Dufumier et al., 2021c, Dufumier et al., 2023, Barbano et al., 2023a]. The definition of negative samples x_j^- varies accordingly.

We define $s(f(a), f(b))$ as a similarity measure (e.g., cosine similarity) between the representation of two samples a and b. Please note that since $||f(a)||_2 = ||f(b)||_2 = 1$, using a cosine similarity is equivalent to using a L2-distance ($d(f(a), f(b)) = ||f(a) - f(b)||_2^2$).
Similarly to [Chopra et al., 2005, Hadsell et al., 2006, Schroff et al., 2015, Sohn, 2016, Wang et al., 2014, Wang et al., 2019, Weinberger and Saul, 2009, Yu and Tao, 2019], we propose to use an ϵ-margin metric learning approach, which allows us to better formalize recent contrastive losses, such as InfoNCE [Chen et al., 2020, Oord et al., 2018], InfoL1O [Poole et al., 2019] and SupCon [Khosla et al., 2020], and derive new losses that better approximate the mutual information. Probably the simplest contrastive learning formulation is looking for a mapping function f such that the following ϵ-condition is always satisfied:

$$\underbrace{d(f(x), f(x^+))}_{d^+} - \underbrace{d(f(x), f(x_j^-))}_{d_j^-} < -\epsilon \iff \underbrace{s(f(x), f(x_j^-))}_{s_j^-} - \underbrace{s(f(x), f(x^+)}_{s^+} \leq -\epsilon \quad \forall j \tag{1.1}$$

where $\epsilon \geq 0$ is a margin between positive and negative samples and we consider, for now, a single positive sample x^+. If we also considered a single negative sample x^-, we would obtain the *triplet contrastive loss*, that was initially used in [Weinberger and Saul, 2009, Wang et al., 2014, Schroff et al., 2015] to extend the *pairwise contrastive loss* [Chopra et al., 2005, Hadsell et al., 2006].

Derivation of InfoNCE

The constraint of Eq. 1.1 can be transformed in an optimization problem using, as it is common in contrastive learning, the max operator and its smooth approximation *LogSumExp*:

$$s_j^- - s^+ \leq -\epsilon \quad \forall j$$
$$\arg\min_f \max(0, \{s_j^- - s^+ + \epsilon\}_{j=1,\dots,N}) \approx \arg\min_f \underbrace{-\log\left(\frac{\exp(s^+)}{\exp(s^+ - \epsilon) + \sum_j \exp(s_j^-)}\right)}_{\epsilon - InfoNCE} \tag{1.2}$$

Please note that another loss could be $\arg\min_f \sum_{j=1}^N \max(0, s_j^- - s^+ + \epsilon)$, which is a lower-bound of Eq.1.2. When these losses are equal to 0 (i.e., minimized), the conditions $s_j^- - s^+ \leq -\epsilon$ are fulfilled $\forall j$. Furthermore, we can notice that when $\epsilon = 0$, we retrieve the InfoNCE loss, also known as N-Pair loss [Sohn, 2016], whereas when $\epsilon \to \infty$ we obtain the InfoL1O loss. It has been shown in [Poole et al., 2019] that these two losses are lower and upper bound of the Mutual Information $I(X^+, X)$ respectively:

$$\mathop{\mathbb{E}}_{\substack{(x,x^+)\sim p(x,x^+) \\ x_j^- \sim p(x^-)}} \left[\underbrace{\log \frac{\exp s^+}{\exp s^+ + \sum_j \exp s_j^-}}_{InfoNCE}\right] \leq I(X^+, X) \leq \mathop{\mathbb{E}}_{\substack{(x,x^+)\sim p(x,x^+) \\ x_j^- \sim p(x^-)}} \left[\underbrace{\log \frac{\exp s^+}{\sum_j \exp s_j^-}}_{InfoL1O}\right] \tag{1.3}$$

By changing the value of $\epsilon \in [0, \infty)$, one might find a tighter approximation of $I(X^+, X)$ since the exponential function at the denominator $\exp(-\epsilon)$ monotonically decreases as ϵ increases.

Inclusion of multiple positives

The inclusion of multiple positive samples (s_i^+) can lead to different formulations. Some of them can be found in [Barbano et al., 2023b]. One of the simplest is:

$$s_j^- - s_i^+ \leq -\epsilon \quad \forall i, j$$
$$\sum_i \max(-\epsilon, \{s_j^- - s_i^+\}_{j=1,\dots,N}) \approx -\sum_i \log\left(\frac{\exp(s_i^+)}{\exp(s_i^+ - \epsilon) + \sum_j \exp(s_j^-)}\right) \tag{1.4}$$

It's interesting to notice that Eq. 1.4 is similar to $\mathcal{L}_{out}^{sup}$, which is one of the two supervised contrastive losses (SupCon) proposed in [Khosla et al., 2020], but they differ for a sum over the positive samples at the denominator. The $\mathcal{L}_{out}^{sup}$ loss, presented as the "most straightforward way to generalize" the InfoNCE loss, actually contains another *non-contrastive constraint* on the positive samples: $s_t^+ - s_i^+ \leq 0 \quad \forall i, t$[1]. Fulfilling this condition alone would force all positive samples to collapse to a single point in the representation space, thus loosing the intra-class variability that could be important for the downstream task. As shown in Table 1.1, the proposed loss, called ϵ-SupInfoNCE and presented in Eq. 1.4, outperforms all other losses in a supervised setting. In particular, it performs better than SupCon (i.e., $\mathcal{L}_{out}^{sup}$). We conjecture that the lack of the non-contrastive term leads to an increased robustness. Further considerations and results can be found in [Barbano et al., 2023b].

Debiasing with FairKL

Satisfying the ϵ-condition can generally guarantee good downstream performance. However, it does not take into account the presence of biases (*e.g.,* data or selection biases). A model could therefore take its decision based on certain visual features that are correlated with the target downstream

[1] we call it non-contrastive since it does not take into account negative samples but only positive ones

Dataset	Network	SimCLR	Max-Margin	SimCLR*	CE*	SupCon*	ϵ-SupInfoNCE*
CIFAR-10	ResNet-50	93.6	92.4	91.74±0.05	94.73±0.18	95.64±0.02	**96.14**±0.01
CIFAR-100	ResNet-50	70.7	70.5	68.94±0.12	73.43±0.08	75.41±0.19	**76.04**±0.01
ImageNet-100	ResNet-50	-	-	66.14±0.08	82.1±0.59	81.99±0.08	**83.3**±0.06

Table 1.1: Accuracy on vision datasets. SimCLR and Max-Margin results are taken from [Khosla et al., 2020]. Results denoted with * are (re)implemented with mixed precision due to memory constraints.

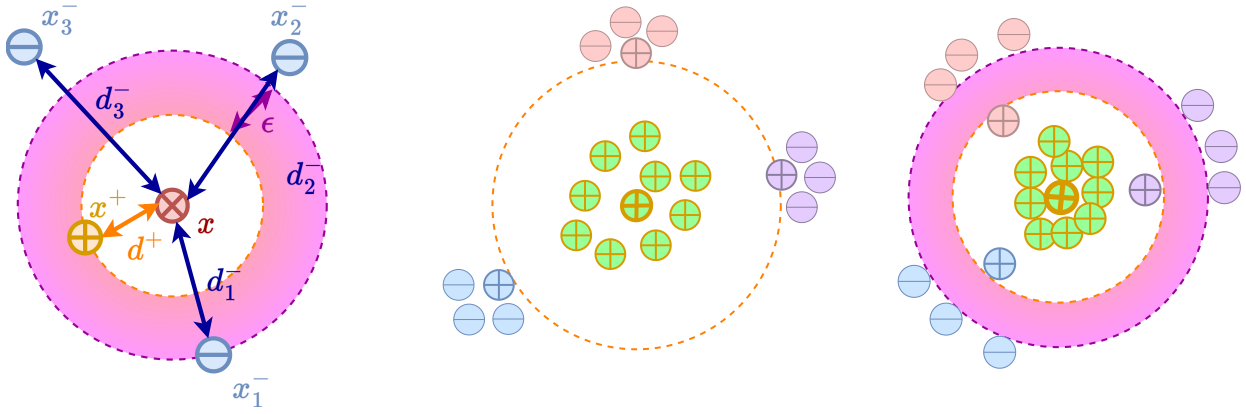


Figure 1.2: In a) we show a visual explanation of ϵ-SupInfoNCE. We aim at increasing the minimal margin ϵ, between the distance d^+ of a positive sample x^+ (+ symbol inside and yellow color) from an anchor x and the distance d^- of the closest negative sample x^- (− symbol inside and blue color). By increasing the margin, we can achieve a better separation between positive and negative samples. In b) and c), we show two different scenarios without margin (b) and with margin (c). Filling colors of datapoints represent different biases. In both b) and c) the contrastive conditions are fulfilled and thus the loss is minimized (i.e., positives are closer to the anchor than the negatives). However, we observe that, without imposing a margin, biased clusters might appear containing both positive and negative samples (b). This issue can be mitigated by increasing the ϵ margin (c).

task but don't actually characterize it. This means that the same bias features would probably have a worse performance if transferred to a different dataset (e.g. different acquisition sites or protocol or image quality). Specifically, in contrastive learning, this can lead to settings where we are still able to minimize a contrastive loss, but with a biased representation and thus (probably) degraded classification performance (see explanation in Fig. 1.2).

(a) Anchor x **(b)** Bias-aligned $x^{+,b}$ **(c)** Bias-conflicting $x^{+,b'}$

Figure 1.3: Biased MNIST [Bahng et al., 2020]. A positive bias-aligned sample $x^{+,b}$ is semantically similar (positive) to the anchor (same digit) but it has also the same bias b (yellow color). A positive bias-conflicting sample shares the same digit but it has a different bias b'(different color). Here, the color is defined as a data bias since it's a visual feature that is correlated with the semantic content related to the target task (digit recognition), but it doesn't characterize it.

To formally explain and tackle the bias problem, we employ the notion of *bias-aligned* sample $x^{\cdot,b}$ and *bias-conflicting* sample $x^{\cdot,b'}$, as in [Nam et al., 2020]. Using the biased MNIST example [Bahng

et al., 2020] in Fig.1.3, a bias-aligned sample shares the same bias attribute b (here color) of the anchor, while a bias-conflicting sample has a different bias attribute b'. Here, we assume that the bias attributes are either known *a priori* or that they can be estimated using a bias-capturing model, such as in [Hong and Yang, 2021].

Given an anchor x, if the bias is "strong" and easy-to-learn, a *positive bias-aligned* sample $x^{+,b}$ will probably be closer to the anchor x in the representation space than a *positive bias-conflicting* sample (of course, the same reasoning can be applied for the negative samples). This is why, even in the case in which the ϵ-condition is satisfied and the ϵ-SupInfoNCE is minimized, as in Fig. 1.2c, we could still be able to distinguish between bias-aligned and bias-conflicting samples. Hence, we say that there is a bias if we can identify an ordering on the learned representations, such as:

$$\underbrace{d(f(x), f(x_i^{+,b}))}_{d_i^{+,b}} < \underbrace{d(f(x), f(x_k^{+,b'})}_{d_k^{+,b'}} \leq \underbrace{d(f(x), f(x_t^{-,b}))}_{d_t^{-,b}} - \epsilon < \underbrace{d(f(x), f(x_j^{-,b'}))}_{d_j^{-,b'}} - \epsilon \quad \forall i, k, t, j \tag{1.5}$$

This represents the worst-case scenario, where the ordering is total (i.e., $\forall i, k, t, j$). Of course, there can also be cases in which the bias is not as strong, and the ordering may be partial.
To tackle this issue,we proposed in [Barbano et al., 2023b] the FairKL regularization technique: a set of debiasing constraints that prevent the use of the bias features within the proposed metric learning approach.

Ideally, we would enforce the conditions $d_k^{+,b'} - d_i^{+,b} = 0 \quad \forall i, k$ and $d_t^{-,b'} - d_j^{-,b} = 0 \quad \forall t, j$, meaning that every positive (resp. negative) bias-conflicting sample should have the same distance from the anchor as any other positive (resp. negative) bias-aligned sample. However, in practice, this condition is very strict, as it would enforce uniform distance among all positive (resp. negative) samples. A more relaxed condition would instead force the distributions of distances, $\{d_k^{;b'}\}$ and $\{d_i^{;b}\}$, to be similar. Here, we propose two new debiasing constraints for both positive and negative samples using either the first moment (mean) of the distributions or the first two moments (mean and variance). Using only the average of the distributions, we obtain:

$$\frac{1}{P_a}\sum_i d_i^{+,b} - \frac{1}{P_c}\sum_k d_k^{+,b'} = 0 \iff \frac{1}{P_c}\sum_k |s_k^{+,b'}| - \frac{1}{P_a}\sum_i |s_i^{+,b}| = 0 \tag{1.6}$$

where P_a and P_c are the number of positive bias-aligned and bias-conflicting samples, respectively[2]. Denoting the first moments with $\mu_{+,b} = \frac{1}{P_a}\sum_i d_i^{+,b}$, $\mu_{+,b'} = \frac{1}{P_c}\sum_k d_k^{+,b'}$, and the second moments of the distance distributions with $\sigma^2_{+,b} = \frac{1}{P_a}\sum_i (d_i^{+,b} - \mu_{+,b})^2$, $\sigma^2_{+,b'} = \frac{1}{P_c}\sum_k (d_k^{+,b'} - \mu_{+,b-})^2$, and making the hypothesis that the distance distributions follow a normal distribution, we can define a new set of debiasing constraints using, for example, the Kullback–Leibler divergence:

$$D_{KL}(\{d_i^{+,b}\}||\{d_k^{+,b'}\}) = \frac{1}{2}\left[\frac{\sigma^2_{+,b} + (\mu_{+,b} - \mu_{+,b'})^2}{\sigma^2_{+,b'}} - \log\frac{\sigma^2_{+,b}}{\sigma^2_{+,b'}} - 1\right] = 0 \tag{1.7}$$

In practice, one could also use another distribution such as the log-normal, the Jeffreys divergence $(D_{KL}(p||q) + D_{KL}(q||p))$, or a simplified version, such as the difference of the two statistics (e.g., $(\mu_{+,b} - \mu_{+,b'})^2 + (\sigma_{+,b} - \sigma_{+,b'})^2$). The proposed debiasing constrains can be easily added to any contrastive loss, using the method of the Lagrange multipliers, as a regularization term $\mathcal{R}^{FairKL} = D_{KL}(\{d_i^{+,b}\}||\{d_k^{+,b'}\})$. In our experiments, we used: $\mathcal{L} = \mathcal{L}_{\epsilon-SupInfoNCE} + \lambda\mathcal{R}^{FairKL}$ where λ is a positive trade-off hyperparameter. Results on the Biased MNIST dataset are shown in Table 1.2, where the proposed strategy outperforms all other state-of-the-art (SOTA) methods. More results and discussion can be found in [Barbano et al., 2023b].

[2]The same reasoning can be applied to negative samples (omitted for brevity.)

Method	0.999	0.997	0.995	0.99
CE [Hong and Yang, 2021]	11.8±0.7	62.5±2.9	79.5±0.1	90.8±0.3
LNL [Kim et al., 2019]	18.2±1.2	57.2±2.2	72.5±0.9	86.0±0.2
ϵ-SupCon	24.36±3.23	74.35±0.09	84.13±1.31	91.12±0.35
ϵ-SupInfoNCE	33.16±3.57	73.86±0.81	83.65±0.36	91.18±0.49
EnD [Tartaglione et al., 2021]	59.5±2.3	82.70±0.3	94.0±0.6	94.8±0.3
BiasCon+BiasBal* [Hong and Yang, 2021]	30.26±11.08	82.83±4.17	88.20±2.27	95.04±0.86
BiasBal [Hong and Yang, 2021]	76.8±1.6	91.2±0.2	93.9±0.1	96.3±0.2
BiasCon+CE* [Hong and Yang, 2021]	15.06±2.22	90.48±5.26	95.95±0.11	97.67±0.09
CE + FairKL	79.9±4.29	93.86±1.13	94.85±0.55	95.92±0.17
ϵ-SupCon + FairKL	89.45±1.82	95.75±0.16	96.31±0.81	96.72±0.2
ϵ-SupInfoNCE + FairKL	**90.51**±1.55	**96.19**±0.23	**97.00**±0.06	**97.86**±0.02

Table 1.2: Top-1 accuracy (%) on Biased-MNIST. Reference results from [Hong and Yang, 2021]. Results denoted with * are re-implemented without color-jittering and bias-conflicting oversampling.

Weakly-supervised setting

When samples have weak attributes, namely supplementary information about the data not defining a proper class, how can we include them in the previous contrastive framework ?

We first need to distinguish two cases: discrete/categorical weak attributes and continuous ones. The former can be easily included by modifying the positive and negative sampling procedure. Using as example the gender weak attribute (Male/Female), one could consider as (possibly) positive only the samples that have the same gender as the anchor and negative otherwise. However, when considering continuous weak attributes, the previous losses are not well adapted, as it is not possible to determine a hard boundary between positive and negative samples. All samples are somehow positive and negative at the same time.

Given the continuous weak attribute y for the anchor x and y_k for a sample x_k, one could threshold the distance d between y and y_k at a certain value τ in order to create positive and negative samples (*i.e.,* k is positive if $d(y, y_k) < \tau$). The problem would then be how to choose τ. Differently, we propose to define a degree of "positiveness" between samples using a kernel function $w_k = K(y - y_k)$, where $0 \leq w_k \leq 1$. Our goal is thus to learn a parametric function $f : \mathcal{X} \to \mathbb{S}^d$ that maps samples with a high degree of positiveness ($w_k \sim 1$) close in the latent space and samples with a low degree ($w_k \sim 0$) far away from each other.
As first approach [Dufumier et al., 2021b], we proposed to consider as "positive" only the samples that have a degree of positiveness greater than 0, and align them with a strength proportional to the degree, namely:

$$\begin{gathered} \frac{w_k}{\sum_j w_j}(s_t - s_k) \leq 0 \quad \forall j, k, t \neq k \in A \\ \arg\min_f \sum_k \max(0, \frac{w_k}{\sum_j w_j}\{s_t - s_k\}_{\substack{t=1,\dots,N \\ t\neq k}}) \approx \mathcal{L}^{y-aware} = -\sum_k \frac{w_k}{\sum_j w_j} \log\left(\frac{\exp(s_k)}{\sum_{t=1}^N \exp(s_t)}\right) \end{gathered} \tag{1.8}$$

where we have normalized the kernel so that the sum over all samples is equal to 1 and we denote with A the indices of samples in the minibatch distinct from the anchor x. Due to the non-hard boundary between positive and negative samples, both s_t and s_k are defined over the entire minibatch. This loss has been used in [Dufumier et al., 2021c, Dufumier et al., 2024] to estimate the

general variability of healthy subjects, merging the information of both imaging and (weak) clinical attributes, such as age and sex. This new pre-training step should learn an accurate representation of the biological and environmental variability of the healthy brain, that can be then used in a second fine-tuning phase to discover pathological patterns of a brain disease. A visual explanation of this new paradigm is shown in Fig.1.4.

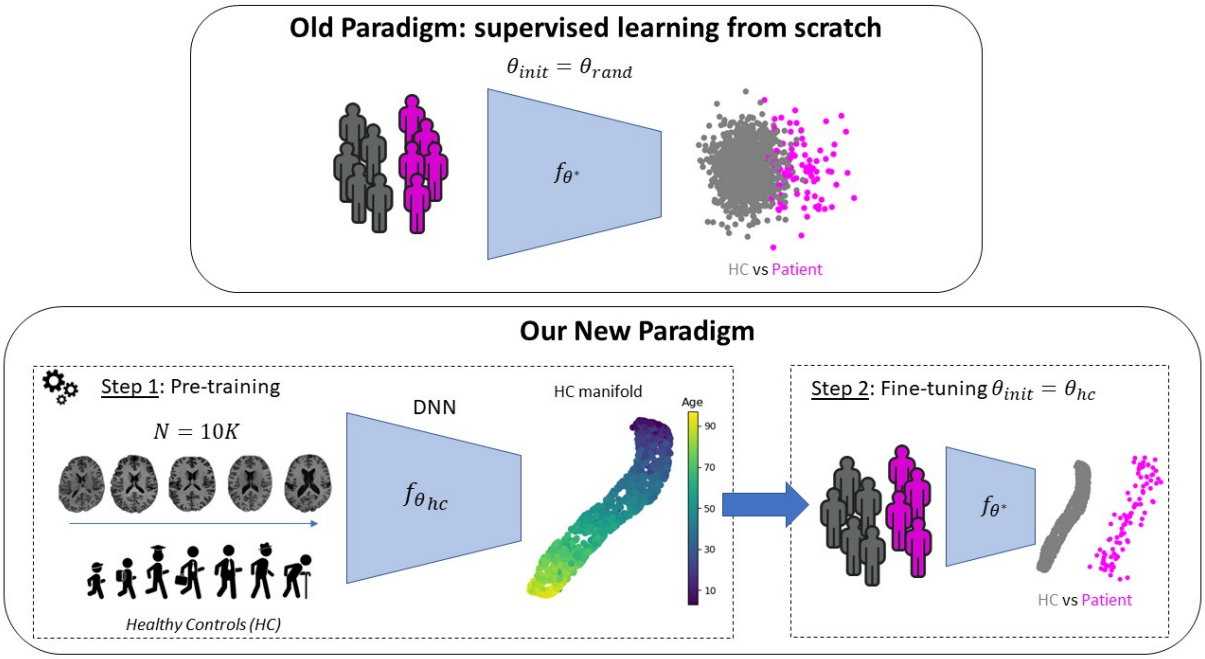


Figure 1.4: New paradigm for discriminating psychiatric disorders at the subject-level combining imaging and clinical data. In a pre-training phase, a Deep learning network f_θ is trained to learn a low-dimensional embedding from a large brain imaging dataset of healthy controls, discovering the general variability associated with non-specific weak attributes, such as age and sex. This pre-training can be performed either with i) self-supervised tasks (e.g., contrastive learning [Dufumier et al., 2021c, Chen et al., 2020]) ii) generative modeling (e.g., VAE [Kingma and Welling, 2014]) or iii) discriminative tasks (e.g., age prediction [Bashyam et al., 2020]). In the second step, the model is initialized with pre-trained weights $\theta_{init} = \theta_{hc}$ and fine-tuned to discriminate between patients and controls. Our main hypothesis is that the representation learned during pre-training will allow easier discovery of the specific variability associated with the pathology of interest (e.g., abnormal cortical atrophy in temporal and pre-fontal regions for schizophrenia or ASD).

We tested our new loss presented in Eq. 1.8 by only using age as auxiliary information, thus calling it Age-Aware, and using a Radial Basis Function as kernel K. We also compared it with: 1) a supervised model without pre-training (Baseline); 2) unsupervised contrastive learning (SimCLR [Chen et al., 2020]); 3) another self-supervised model for medical imaging based on context-based restoration (Model Genesis [Zhou et al., 2021b]); 4) Variational AutoEncoder (VAE [Kingma and Welling, 2014]) considered as SOTA generative model (easier to train than GAN [Goodfellow et al., 2014] or diffusion models [Nichol and Dhariwal, 2021, Dhariwal and Nichol, 2021, Ho et al., 2020, Rombach et al., 2022] and obtain an encoder that can be fine-tuned); 5) a discriminative supervised model trained on age prediction. Results on three mental disorders, schizophrenia (SCZ), Bipolar Disorder (BD), and Autism Spectrum Disorders (ASD), are shown in Table 1.3. We use both an internal (same sites as training data) and external (different sites as training data) test sets to evaluate the performance. The proposed pre-training strategy outperforms all other methods on the three clinical datasets.

Importantly, age information is only used during pre-training and it is *never* used during fine-tuning. All these models are pre-trained using the healthy subjects of the datasets: OpenBHB [Dufumier et al., 2022], HCP[3], ICBM[4] and OASIS3[5]. The final pre-training dataset comprises 9116 3D MRI scans of healthy brains coming from 42 sites and covering the entire life-spectrum in terms of age (i.e., both young and old subjects), thus promoting heterogeneity in the healthy population under study. Hyper-parameter values, such as the kernel σ^2 and the learning rate, are selected using a Monte-Carlo Cross-Validatation strategy, where both training and validation sets are stratified on age, sex and site. More details about the validation strategy, datasets, augmentations and robustness to hyper-parameters can be found in [Dufumier et al., 2021c, Dufumier et al., 2024].

Task	**Test Set**		**Pre-training Strategies**				
			Weakly Self-Supervised	Self-Supervised		Generative	Discriminative
		Baseline	Age-Aware	Model Genesis	SimCLR	VAE	Age Sup.
SCZ vs. HC ↑	Internal ($N = 118$)	$85.27_{\pm 1.60}$	$\mathbf{85.17_{\pm 0.37}}$	$76.31_{\pm 1.77}$	$82.31_{\pm 2.03}$	$82.56_{\pm 0.68}$	$83.05_{\pm 1.36}$
$N_{train} = 933$	External ($N = 133$)	$75.52_{\pm 0.12}$	$\mathbf{77.00_{\pm 0.55}}$	$67.40_{\pm 1.59}$	$75.48_{\pm 2.54}$	$75.11_{\pm 1.65}$	$74.36_{\pm 2.28}$
BD vs. HC ↑	Internal ($N = 107$)	$76.49_{\pm 2.16}$	$\mathbf{78.81_{\pm 2.48}}$	$76.25_{\pm 1.48}$	$72.71_{\pm 2.06}$	$71.61_{\pm 0.81}$	$77.21_{\pm 1.00}$
$N_{train} = 832$	External ($N = 131$)	$68.57_{\pm 4.72}$	$\mathbf{77.06_{\pm 1.90}}$	$65.66_{\pm 0.90}$	$71.23_{\pm 3.05}$	$71.70_{\pm 0.23}$	$73.02_{\pm 2.66}$
ASD vs. HC ↑	Internal ($N = 186$)	$65.74_{\pm 1.47}$	$66.36_{\pm 1.14}$	$63.58_{\pm 4.35}$	$61.92_{\pm 1.67}$	$59.67_{\pm 2.04}$	$\mathbf{67.11_{\pm 1.76}}$
$N_{train} = 1526$	External ($N = 207$)	$62.93_{\pm 2.40}$	$\mathbf{68.76_{\pm 1.70}}$	$54.95_{\pm 3.58}$	$61.93_{\pm 1.93}$	$57.45_{\pm 0.81}$	$62.07_{\pm 2.98}$

Table 1.3: Fine-tuning results. All pre-trained models use a data-set of 9116 3D MRI scans of healthy brains. We report average AUC(%) for all models and the standard deviation by repeating each experiment three times. Baseline uses a DenseNet121 backbone optimized in a supervised way. The same backbone is also used for the Contrstive self-supervised methods.

Decoupled uniformity

The InfoNCE loss, and similarly all previously presented losses, can be divided into two terms: $\underbrace{-s^+}_{\text{aligment}} + \underbrace{\log\left(\exp(s^+) + \sum_j \exp(s_j^-)\right)}_{\text{uniformity}}$, as proposed in [Wang and Isola, 2020, Dufumier et al., 2021b], where we have used a single positive and batch for simplicity. Using this formulation we can easily see that positive samples (s^+) are both attracted (alignment) and repelled (uniformity) at the same time. This is called the *negative-positive coupling* problem, which can be solved by removing the positive sample at the denominator, as proposed in [Yeh et al., 2022] (equivalent to using an $\epsilon \to \infty$ in Eq. 1.2). Another characteristic of the InfoNCE loss is the need for large batch sizes and thus many negative examples. Indeed, it has been shown that this loss works well only when the batch size is very large (e.g., $>$ 2048 for ImageNet [Chen et al., 2020]). In an unsupervised setting (i.e., no labels), this brings to another issue related to the number of *False Negative samples*. If the number of latent classes in the training set is low (for instance 2) and they are balanced (similar number of samples per class), then the number of false negative will be probably high ($\frac{2N-2}{2}$ if we augment once all samples N). Here, a false negative is a sample that is considered as a negative, but that it actually belongs to same (latent) class as the anchor. To solve both problems, we have proposed in [Dufumier et al., 2023] a new solution by imposing uniformity only between centroids

[3] https://www.humanconnectome.org/study/hcp-young-adult
[4] https://ida.loni.usc.edu/login.jsp
[5] https://sites.wustl.edu/oasisbrains/

$\mu_i = \frac{1}{V}\sum_{v=1}^{V} f(x_i^v)$, defined as the average representation between several views x_i^v of the same image x_i. The new loss, that we called Decoupled Uniformity, is defined as:

$$\mathcal{L}_{unif}^{de}(f) = \log \frac{1}{N(N-1)} \sum_{i \neq j} \exp\left(-||\mu_i - \mu_j||^2\right) \tag{1.9}$$

where N is the size of the batch, i and j are two different samples of the batch and V the number of views (same for all samples). From a metric learning point-of-view, minimizing Eq. 1.9 is equivalent to looking for an encoder f such that the sum of similarities of all views from the same image (s_i^+ and s_j^+) are higher than the sum of similarities between views from different samples (s_{ij}^-):

$$\underbrace{s_i^+ = ||\mu_i||^2 = \frac{1}{V^2}\sum_{v,v'} s(f(x_i^v), f(x_i^{v'}))}_{\text{similarities between views of } x_i} \quad \underbrace{s_j^+ = ||\mu_j||^2 = \frac{1}{V^2}\sum_{v,v'} s(f(x_j^v), f(x_j^{v'}))}_{\text{similarities between views of } x_j} \quad \underbrace{s_{ij}^- = \langle \mu_i, \mu_j \rangle = \frac{1}{V^2}\sum_{v,v'} s(f(x_i^v), f(x_j^{v'}))}_{\text{similarities between views of } x_i \text{ and } x_j}$$

$$s_i^+ + s_j^+ > 2s_{ij}^- + \epsilon \quad \forall i \neq j$$

$$\arg\min_f \log\left(\exp(-\epsilon) + \sum_{i \neq j} \exp\left(-s_i^+ - s_j^+ + 2s_{ij}^-\right)\right) \tag{1.10}$$

At $\lim_{\epsilon \to \infty}$, it results $\log\left(\sum_{i \neq j} \exp\left(-s_i^+ - s_j^+ + 2s_{ij}^-\right)\right)$ and by adding the constant term $\log\left(\frac{1}{N(N-1)}\right)$, we obtain exactly $\mathcal{L}_{unif}^{de}(f)$. This loss implicitly optimizes alignment between positives through the maximization of $||\mu||^2$ and thus we do not need to explicitly add an alignment term. Since we apply it to couple of samples, we solve the False Negative problem. Furthermore, we show in [Dufumier et al., 2023] that this loss does not suffer from the positive-negative coupling problem. We empirically demonstrate the benefits of our loss in Table 1.4 comparing it with both InfoNCE [Chen et al., 2020] and DC [Yeh et al., 2022] losses. Furthermore, we show in Table 1.5 that our loss is less sensitive to batch size than InfoNCE, thanks to its decoupling between positives and negatives. More results can be found in [Dufumier et al., 2023].

Dataset	Network	$\mathcal{L}_{InfoNCE}$	$\mathcal{L}_{DC}$	$\mathcal{L}_{unif}^{de}$
CIFAR-10	ResNet18	$82.18_{\pm 0.30}$	$84.87_{\pm 0.27}$	$\mathbf{85.05}_{\pm 0.37}$
CIFAR-100	ResNet18	$55.11_{\pm 0.20}$	$58.27_{\pm 0.34}$	$\mathbf{58.41}_{\pm 0.05}$
ImageNet100	ResNet50	68.76	73.98	**77.18**

Table 1.4: Comparison of Decoupled Uniformity without prior with InfoNCE[Chen et al., 2020] and DC[Yeh et al., 2022] loss. Batch size $n = 256$. All models are trained for 400 epochs.

Datasets	Loss	$n = 128$	$n = 512$	$n = 1024$	$n = 2048$
CIFAR10	InfoNCE	78.89	79.40	80.02	80.06
	Decoupled Unif	82.67	82.12	82.74	82.33
CIFAR100	InfoNCE	49.53	53.46	54.45	55.32
	Decoupled Unif	54.61	54.12	55.56	55.20

Table 1.5: Linear evaluation accuracy (%) after training for 200 epochs with a batch size n, ResNet18 backbone and latent dimension $d = 128$.

Similarly to the previously presented *y-aware loss* (Eq. 1.8), $\mathcal{L}_{unif}^{de}$ can also integrate prior knowledge, like weak attributes, modeled as scalar or vector values $z(x)$ for each sample x, using a kernel K_σ. Based on the conditional mean embedding theory [Song et al., 2013], we define:

Definition 1.3.1. (Empirical Kernel Decoupled Uniformity Loss) Let $(x_i)_{i \in [1..N]}$, the N samples of a batch with their V views x_i^v and $K_N = [K_\sigma(z(x_i), z(x_j))]_{i,j \in [1..N]}$, the Kernel prior matrix, where K_σ is a standard kernel (*e.g.*, Gaussian or Cosine). We define the new centroid estimator as $\hat{\mu}_j = \frac{1}{V}\sum_{v=1}^{V}\sum_{i=1}^{N} \alpha_{i,j} f(x_i^v)$ with $\alpha_{i,j} = ((K_n + \lambda N \mathbf{I}_N)^{-1} K_N)_{ij}$, $\lambda = O(n^{-1/2})$ a regularization constant. The **empirical Kernel Decoupled Uniformity loss** is then: $\hat{\mathcal{L}}_{unif}^{de}(f) \overset{\text{def}}{=} \log \frac{1}{n(n-1)} \sum_{i \neq j} \exp(-||\hat{\mu}_i - \hat{\mu}_j||^2)$

The computational cost added is roughly $O(n^3)$ (to compute the inverse matrix of size $n \times n$) but it remains negligible compared to the back-propagation time using classical stochastic gradient descent. Importantly, the gradients associated to $\alpha_{i,j}$ are not computed.

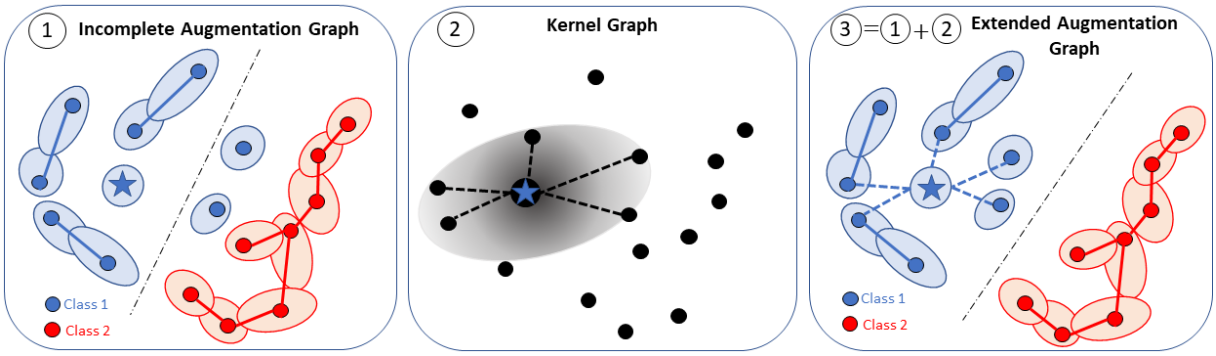


Figure 1.5: Visual explanation of the augmentation graph and illustration of the proposed method. Each point is an original image x. Two points are connected if they can be transformed into the same augmented image, via the chosen augmentations/transformations. Colors represent semantic (latent, unknown) classes and light disks represent the support of augmentations for each sample x. From an incomplete augmentation graph (1) where intra-class samples are not connected, we reconnect them using a kernel defined on prior information. The extended augmentation graph (3) is the union between the (incomplete) augmentation graph (1) and the kernel graph (2). In (2), the gray disks indicate the set of points x' that are close to the anchor (blue star) in the kernel space.

Most recent theories about CL [Wang et al., 2022a, HaoChen et al., 2021] make the hypothesis that samples from the same semantic class have overlapping augmented views, to provide guarantees on the downstream task when optimizing InfoNCE [Chen et al., 2020] or Spectral Contrastive loss [HaoChen et al., 2021]. This assumption, known as *intra-class connectivity hypothesis*, is very strong and only relies on the distributions of the augmentations (to create the different views). In particular, augmentations should not be "too weak", so that all intra-class samples are connected among them, and at the same time not "too strong", to prevent connections between inter-class samples and thus preserve the semantic information. In [Dufumier et al., 2023], we proved that we can relax this hypothesis if we can provide a kernel that is "good enough" to relate intra-class samples not connected by the augmentations (see Fig. 1.5). We show that $\hat{\mathcal{L}}^{de}_{unif}$ can tightly bound the supervised classification risk by assuming that the *extended* augmentation graph is class-connected and not the augmentation graph, as in previous works [Wang et al., 2022a, HaoChen et al., 2021]. This implies that we do not need *optimal augmentations* to have tight bounds, as in previous works, but we just need a "good enough" kernel to reconnect the disconnected intra-class samples. Furthermore, we don't require perfect alignment for f nor L-smoothness, as in [Saunshi et al., 2019]. Please note that the previous *y-aware loss*, being based on the InfoNCE loss, need stronger assumptions to have tight bounds on the classification loss, as shown in [Wang et al., 2022a, HaoChen et al., 2021].

Besides clinical attributes, as for the *y-aware loss*, we also tested representations of generative models as prior information. We first show results on natural images from ImageNet100 in Table 1.6, where we leverage the prior representation of the BigBiGAN network [Donahue and Simonyan, 2019] pre-trained on ImageNet. We use it to define a kernel $K_{GAN}(x_i, x_j) = K(z(x_i), z(x_j))$ (with K a RBF kernel and $z(\cdot)$ the BigBiGAN's encoder). Results clearly indicate that our method outperforms all other SOTA methods. Then, we also show results on the bipolar disorder detection (BD) using the BIOBD dataset [Hozer et al., 2020]. It contains 356 healthy controls (HC) and 306 patients with BD. As before, we use the Open BHB dataset [Dufumier et al., 2021c] for pre-training, which contains ∼9k 3D images of healthy subjects. As prior representation, we use a pre-trained

VAE to define $K_{VAE}(x_i, x_j) = K(\mu(x_i), \mu(x_j))$ where $\mu(\cdot)$ is the mean Gaussian distribution of x in the VAE latent space and K is a standard RBF kernel. In Table 1.7, we show that the proposed method outperforms previous methods. These results show that generative models can provide good prior when augmentations are too weak or insufficient to remove easy-to-learn noisy features. More details and results can be found in [Dufumier et al., 2023]

Model	ImageNet100
SimCLR [Chen et al., 2020]	68.76
BYOL [Grill et al., 2020]	72.26
CMC* [Tian et al., 2020b]	73.58
DCL* [Chuang et al., 2020]	74.6
AlignUnif [Wang and Isola, 2020]	76.3
DC [Yeh et al., 2022]	73.98
SwAV (w/o multi-crop) [Caron et al., 2020]	73.5
BigBiGAN [Donahue and Simonyan, 2019]	72.0
Decoupled Unif	<u>77.18</u>
K_{GAN} Decoupled Unif	**78.02**
Supervised	$82.1_{\pm 0.59}$

Table 1.6: Linear evaluation accuracy (%) on ImageNet100 using ResNet50 trained for 400 epochs with batch size $n = 256$ for all methods. *Results from paper.

Model	BD vs HC
SimCLR [Chen et al., 2020]	$60.46_{\pm 1.23}$
BYOL [Grill et al., 2020]	$58.81_{\pm 0.91}$
MoCo v2 [He et al., 2020]	$59.27_{\pm 1.50}$
Model Genesis [Zhou et al., 2021b]	$59.94_{\pm 0.81}$
VAE [Kingma and Welling, 2014]	$52.86_{\pm 1.24}$
K_{VAE} Decoupled Unif (ours)	$\mathbf{62.19}_{\pm 1.58}$
Supervised	$67.42_{\pm 0.31}$

Table 1.7: Linear evaluation AUC scores(%) on BD detection using a 5-fold leave-site-out CV with DenseNet121 as backbone.

Contrastive learning for regression

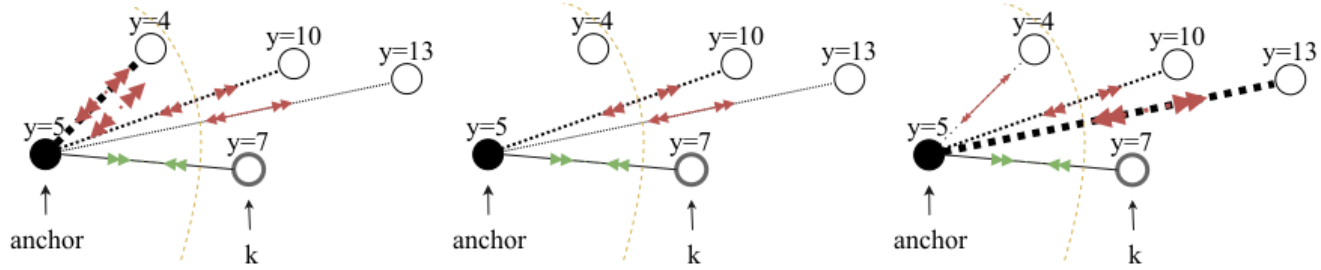
The previously presented y-aware loss (Eq. 1.8) can also be used for regression tasks. However, one of its limitation is that, while the numerator aligns x_k, in the denominator, the uniformity term (as defined in [Wang and Isola, 2020]) focuses more on the closest samples in the representation space. This could be undesirable, as these samples might have a greater degree of positiveness than the considered x_k. To avoid that, we formulate a second loss, called $\mathcal{L}^{thr}$, which limits the uniformity term (i.e., denominator) to the samples that are more distant from the anchor than the considered x_k in the kernel space. Omitting the kernel normalization, we obtain:

$$\begin{aligned} &w_k(s_t - s_k) \leq 0 \quad \text{if } w_t - w_k \leq 0 \quad \forall k, t \neq k \in A(i) \\ &\mathcal{L}^{thr} = -\sum_k \frac{w_k}{\sum_t \delta_{w_t < w_k} w_t} \log \left(\frac{\exp(s_k)}{\sum_{t \neq k} \delta_{w_t < w_k} \exp(s_t)} \right) \end{aligned} \tag{1.11}$$

Ideally, $\mathcal{L}^{thr}$ avoids repelling samples more similar than x_k. However, it still focuses more on the closest sample "less positive" than x_k, i.e. x_t s.t $w_t > w_x$ and $w_t \leq w_j \; \forall j \neq k$. In a classification task, increasing the margin with respect to the closest "negative" sample sounds correct from a theoretical point of view and we have also shown that this is the case in our experiments (see previous Table 1.1 and experiments in [Barbano et al., 2023b]). However, it might not be best suited for (continuous) weak-supervision and regression. For this reason, we propose a third formulation ($\mathcal{L}^{exp}$) that takes an opposite approach. Instead of focusing on repelling the closest "less positive" sample, we increase the repulsion strength of each sample proportionally to its distance from the anchor in the kernel space, obtaining:

$$\begin{aligned} &w_k[s_t(1 - w_t) - s_k] \leq 0 \quad \forall k, t \neq k \in A(i) \\ &\mathcal{L}^{exp} = -\frac{1}{\sum_t w_t} \sum_k w_k \log \frac{\exp(s_k)}{\sum_{t \neq k} \exp(s_t(1 - w_t))} \end{aligned} \tag{1.12}$$

In the resulting $\mathcal{L}^{exp}$ formulation, the weighting factor $1 - w_t$ acts like a temperature value, by giving more weight to the samples that are further away from the anchor in the kernel space. Also, for a proper kernel choice, samples closer than x_k will be repelled with very low strength ($\sim$0). We argue that this approach is more suited for continuous attributes (i.e., regression task), as it enforces the fact that samples close in the kernel space will be close in the representation space. A visual explanation of the three losses can be found in Fig.1.6. It's also interesting to notice that these losses, differently from the InfoNCE loss, do not asymptotically optimize (global) Alignment and (global) Uniformity [Wang and Isola, 2020], but a conditional version of them. Indeed, as shown in [Dufumier et al., 2021b], these losses do not align and repel all samples in the same way, but proportionally to the kernel value.



(a) y-aware **(b)** threshold **(c)** exp

Figure 1.6: Comparison between the proposed contrastive learning losses for continuous weak attributes (weak-supervision and regression) and their effect on the representations. Samples are aligned ($\gg \ll$) and repelled ($\ll\gg$) with varying strength (line thickness) based on the continuous label y.

The proposed losses are tested on the OpenBHB Challenge [Dufumier et al., 2022] whose goal is age prediction (without being biased by the site-effect). The dataset contains 5330 3D brain T1-w MRI scans of different subjects, from 71 different acquisition sites and pre-processed using Voxel-Based Morphometry (VBM) [Ashburner and Friston, 2000]. The evaluation is performed using two private test sets (internal and external). The internal test set contains the same sites as training, the external contains unseen ones. For every model, we evaluate the mean absolute error (MAE) and the balanced accuracy (BAcc) for site prediction, training a logistic regression on the model representations. The final challenge score is computed as $\mathcal{L}_c = \text{BAcc}^{0.3} \cdot \text{MAE}_{\text{ext}}$. Results are shown in Table 1.8 where the proposed losses, and in particular $\mathcal{L}^{exp}$, outperform a baseline model [Dufumier et al., 2022] trained with the L1 loss, and ComBat [Fortin et al., 2018], a site harmonization algorithm developed for MRIs. More details can be found in [Barbano et al., 2023a].

Method	Model	Int. MAE ↓	BAcc ↓	Ext. MAE ↓	$\mathcal{L}_c$ ↓
Baseline (ℓ_1)	DenseNet	2.55±0.01	8.0±0.9	7.13±0.05	3.34
	ResNet-18	2.67±0.05	6.7±0.1	4.18±0.01	1.86
	AlexNet	2.72±0.01	8.3±0.2	4.66±0.05	2.21
ComBat	DenseNet	5.92±0.01	2.23±0.06	10.48±0.17	3.38
	ResNet-18	4.15±0.01	**4.5**±0.0	4.76±0.03	1.88
	AlexNet	3.37±0.01	6.8±0.3	5.23±0.12	2.33
$\mathcal{L}^{exp}$	DenseNet	2.85±0.00	5.34±0.06	4.43±0.00	1.84
	ResNet-18	**2.55**±0.00	5.1±0.1	**3.76**±0.01	**1.54**
	AlexNet	2.77±0.01	5.8±0.1	4.01±0.01	1.71
$\mathcal{L}^{y-aware}$	ResNet-18	2.66±0.00	6.60±0.17	4.10±0.01	1.82
$\mathcal{L}^{thr}$	ResNet-18	2.95±0.01	5.73±0.15	4.10±0.01	1.74

Table 1.8: Final scores on the OpenBHB Challenge leaderboard using a 3D ResNet-18. **MAE**: Mean Absolute Error. **BAcc**: Balanced Accuracy for site prediction. **Challenge score**: $\mathcal{L}_c = \text{BAcc}^{0.3} \cdot \text{MAE}_{\text{ext}}$.

1.3.2 Contrastive Subgroup Discovery

In the past decades, unsupervised and self-supervised learning techniques have proven to be particularly effective at identifying relevant patterns and factors of variation within a dataset. Combined with powerful Neural Networks (NNs), these methods can produce semantically rich representations [Chen et al., 2020, He et al., 2020, Zheng et al., 2020]. Notably, unsupervised Deep Clustering (DC) methods [Caron et al., 2018, Asano et al., 2020, Caron et al., 2020, Li et al., 2021b, Van Gansbeke et al., 2020] seek to produce a suitable representation space for identifying homogeneous latent clusters based on the *general variability* of the entire dataset (i.e., imaging patterns common to all samples).
With a different perspective, Subgroup Discovery (SD) in medical applications [Atzmueller, 2015, Klösgen, 1996, Yang et al., 2021] aims at identifying relevant latent subtypes/subgroups that arise from the *pathological variability* of the diseased population and not from the irrelevant common variability that may exist in both healthy subjects (i.e., controls) and diseased patients.

For instance, in clinical research, it is essential to identify subtypes of patients with a given disorder (red dots in Fig. 1.7). However, the general variability that stems from age or sex, and which is observed in both healthy controls (grey dots in Fig. 1.7) and patients, will probably drive the clustering of patients to a non-specific solution (2^{nd} plot in Fig. 1.7). Instead, subtypes should be defined only by the modes of variation (horizontal arrow Fig. 1.7) specific to the pathology, thus discarding non-specific variability and emphasizing more disease-related differences.

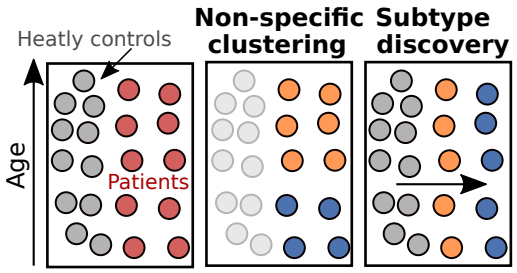


Figure 1.7: Subtype discovery in clinical research.

In Fig. 1.8, we use an intuitive toy example based on the MNIST dataset to better clarify the differences between Deep Clustering and Subtype Discovery. We consider the digit "7" as the pathological group and all the other digits as the healthy group. Results show how Deep Cluster's subgroups [Caron et al., 2018] of the digit "7" are only defined by the most predominant characteristics (i.e.: boldness of the digit) common to all digits. Instead, a Subtype Discovery method, such as the proposed Deep UCSL, disregards these common characteristics and uses only the specific patterns of the digit "7" (i.e., the presence of the crossing middle bar) to define the subgroups.

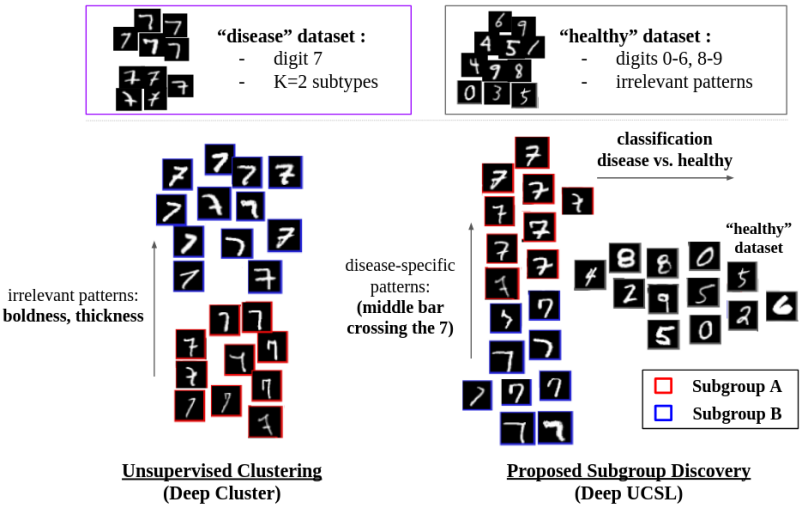


Figure 1.8: Comparison between a Deep Clustering method (Deep Cluster [Caron et al., 2018]) and one of the proposed Subgroup Discovery method (Deep UCSL) on a subtype discovery task within the digit 7. We show 2D PCA plots of the representation spaces learnt by the two methods.

As already explained in the Introduction, mental disorders are heterogeneous. This motivates the need for new machine-learning methods to help human experts discover or validate subgroups, while relying on reproducible, data-

driven, and objective imaging patterns. By contrasting controls with patients, we proposed two new methods to identify subgroups that stem only from the pathological variability specific to the disease, while disregarding the common variability shared with the controls. The objectives of both methods are: a) correctly discover the pathological subgroups, b) encourage healthy samples not to belong to a pathological subgroup, and c) accurately discriminate each subgroup from the healthy class. Here, we present Deep Unsupervised Clustering driven by Supervised Learning (Deep UCSL) [Louiset, 2024], that is an extension of our previous method UCSL, presented in [Louiset et al., 2021].

Deep Unsupervised Clustering driven by Supervised Learning (Deep UCSL)

Let $(X, Y) = \{(x_i, y_i)\}_{i=1}^{N}$ be a labeled dataset composed of N samples. We will restrict to the binary (e.g., patient/control) classification paradigm, $y_i \in \{-1, +1\}$. We will denote with N^+ and N^- ($N = N^+ + N^-$) the number of positive and negative samples, respectively. Our objective is to estimate the latent pseudo-labels[6] of subgroups within disease samples ($y_i = +1$). The membership of each sample i to latent subgroups is modeled via a latent categorical variable $c_i \in C = \{1, ..., K\}$, where K is the number of subgroups. Here, K is assumed to be known. We look for a discriminative model that maximizes the joint conditional likelihood:

$$\sum_{i=1}^{n} \log p(y_i|x_i) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} p(y_i, c_i = k|x_i) \tag{1.13}$$

To attain the previously described objectives, we need to optimize Eq. 1.13 with respect to both $p(c_i|x_i, y_i)$ and $p(y_i|x_i, c_i)$. Indeed, we need to identify the subgroups only within the diseased samples (thus knowing y) and to accurately discriminate the healthy class from each subgroups (thus knowing c). However, developing the joint conditional likelihood in Eq. 1.13 would result in either $p(c_i|x_i, y_i)$ or $p(y_i|x_i, c_i)$, but not in both. To solve that, as in UCSL [Louiset et al., 2021], we introduce a probability distribution Q over the subgroups C, so that $\sum_{k=1}^{K} Q(c_i = k) = 1 \quad \forall i$, and use the Jensen inequality to obtain a tractable, lower bound of Eq. 1.13:

$$\sum_{i=1}^{n} \log \sum_{k=1}^{K} Q(c_i = k) \frac{p(y_i, c_i = k|x_i)}{Q(c_i = k)} \geq \sum_{i=1}^{n} \sum_{k=1}^{K} Q(c_i = k) \log \left(\frac{p(y_i, c_i = k|x_i)}{Q(c_i = k)} \right) \tag{1.14}$$

where equality holds when: $Q(c_i = k) = \frac{p(y_i, c_i = k|x_i)}{\sum_{k=1}^{K} p(y_i, c_i = k|x_i)} = \mathbf{p(c_i = k|x_i, y_i)}$. Then, Eq. 1.13 can be rewritten with respect to both $p(y_i|x_i, c_i)$ and $Q(c_i)$ (estimated to approximate $p(c_i|x_i, y_i)$):

$$\underbrace{\sum_{i=1}^{n} \sum_{k=1}^{K} Q(c_i = k) \log \mathbf{p(y_i|x_i, c_i = k)}}_{\text{Mixture-of-Classifying Experts term}} \underbrace{- D_{KL}(Q(c)||p(c|x))}_{\substack{\text{Clustering} \\ \text{Regularization term}}} \tag{1.15}$$

Our goal is to learn a single representation space where both the classifying experts $p(y_i|x_i, c_i = k)$ and the disease subgroup $p(c_i = k|x_i, y_i = +1)$ can be accurately estimated. To this end, we propose using a deep encoder f_θ with parameters θ for feature extraction and two neural networks with parameters ϕ and ψ for the classifying experts $p_{\theta,\phi}(y_i|c_i = k, x_i)$ and the unsupervised clustering head $p_{\theta,\psi}(c_i = k|x_i)$, respectively. An overview of the proposed method can be seen in Fig.1.9. To optimize the proposed cost function (Eq. 1.15), we use an EM algorithm that alternatively:

[6] we call them latent pseudo-labels, since we assume that subgroups labels are not known at training

1. estimates Q as $p(c_i = k|x_i, y_i)$ (E-step) at the end of each epoch, freezing the encoder f_θ. Since we assume that only the positive class ($y_i = +1$) contains subgroups, we compute $p_\theta(c_i = k|x_i, y_i = +1)$ using a regularized K-means algorithm[7], fixing a uniform clustering probability distribution (*i.e.*, $\frac{1}{K}$) for the healthy samples ($y_i = -1$).

2. estimates $p(y_i|x_i, c_i = k)$ and $p(c_i = k|x_i)$ batch-wise by maximizing Eq. 1.15 (M-step) at the beginning of each epoch, freezing Q.

The minimization of the clustering regularization term in the M-step brings to a representation space more suited for subgroup discovery. Indeed, healthy samples should be encoded in the representation space as points equidistant from the subgroup centroids, since their membership probability should be the same for all subgroups (i.e., $1/K$). Furthermore, positive samples should be clustered as in $Q(c_i)$, namely as if the "unsupervised" clustering algorithm was only considering the pathological/positive variations. This regularization thus promotes a representation space where the general variability (common to both negative and positive classes) is discarded for the identification of subgroups.

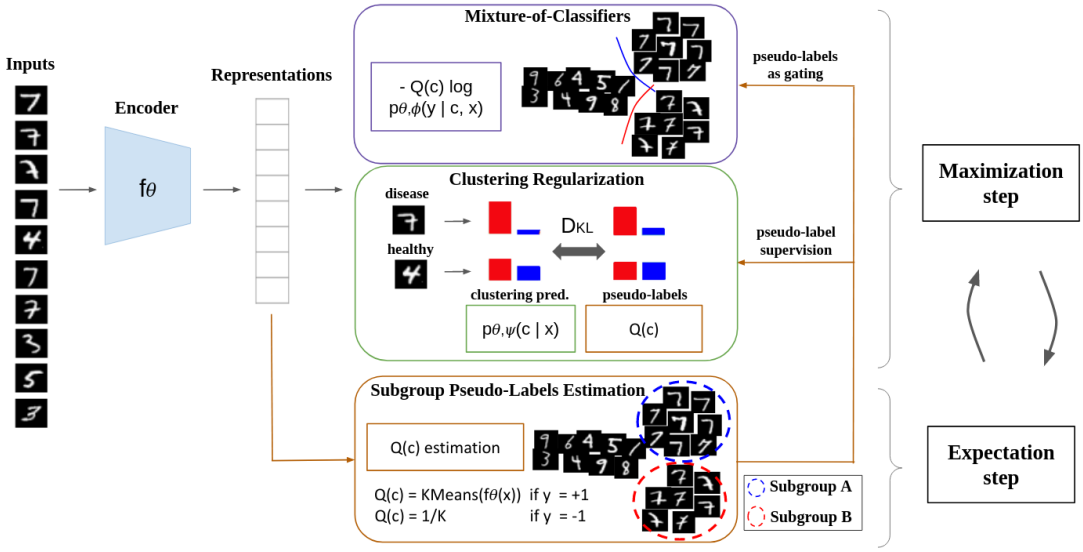


Figure 1.9: A schematic diagram of Deep UCSL with $K = 2$ subgroups (red and blue). At each epoch, K-Means produces subgroup pseudo-labels during the Expectation step (in brown). These pseudo-labels are then used to weight a classification Mixture-of-Experts (in purple) between the "healthy" class (digits 0-6, 8-9) and the "disease" class (digit 7). Additionally, the pseudo-labels are also used for the clustering regularization (in green), where uniform pseudo-labels (i.e.: $\frac{1}{K}$) are used to regularize the healthy class distribution, so that healthy samples are equidistant from all the diseased subgroups. This forces the learnt representation to disregard the general variability, common to both healthy and diseased samples.

Comparison with UCSL

This mathematical framework is similar to the one of our preliminary work UCSL [Louiset et al., 2021], but with significant differences.
First, Deep UCSL uses a deep feature encoder, instead of user-defined features, and two neural-networks for classification and subgroup estimate, instead of linear models.

[7] any clustering algorithm could be used here. Since the number of subgroups K is assumed to be known, K-means is a reasonable and simple choice.

Second, we do not assume that $p_\theta(c|x) = Q(c)$, as in UCSL, but we force it by explicitly introducing and minimizing the clustering regularization term $KL(Q(c)||p_{\theta,\psi}(c|x_i))$. This guarantees the monotonic convergence of the optimization procedure, which was not the case in UCSL.
Third, since we want to estimate subgroups only within positive/diseased samples, all negative/healthy samples are assigned a uniform probability for all subgroups. This strategy, also not proposed in UCSL, encourages the features encoder f_θ to produce a representation space where negative samples do not belong to (positive) subgroups. This new contribution implies that $p_\theta(c_i|x_i)$, the estimated clustering distribution, is not simply extended to all samples regardless their label y, as in UCSL, but the representation space is estimated so that the unsupervised clustering $p_\theta(c_i|x_i)$ gives the same result as the "supervised" subgroups estimation $p(c_i = k|y_i, x_i)$, namely knowing the label y. This entails an encoder f and a representation space where the general variability, common to both positive and negative samples, is discarded and the subtype estimation only depends on the specific variability of the positive class.

Evaluation

To test the usefulness of the proposed methods, we create a dataset for subgroup identification comprising 3D brain MRI T1-w images. There are two classes: one of Healthy Controls (HC=686) and one of patients comprising two Mental Disorders (MD) (i.e., subgroups): 1) patients with Schizophrenia (SZ=275), from SCHIZCONNECT [Wang et al., 2016a], and 2) patients with Bipolar Disorder (BD=307), from BIOBD dataset [Sarrazin et al., 2018]. Images are pre-processed with Voxel-based morphometry (VBM) using CAT12 [Gaser and Dahnke, 2016]. Furthermore, we also compute 142 features by averaging Gray Matter (GM) values over the Regions-of-Interest (ROIs) of the Neuromorphometrics atlas. For Deep Learning methods, we use the pre-processed GM-only images as inputs of a 3D-DenseNet deep encoder, as in [Dufumier et al., 2021c]. For UCSL, we consider the GM ROIs features. In Table 1.9, we show the subgroup identification capability of Deep UCSL compared with related works. All evaluation criteria are computed on an independent TEST set (199 HC, 190 SZ, 116 BP), coming from the BSNIP cohort [Tamminga et al., 2014], with different acquisition sites. Controls and patients share common (thus irrelevant) sources of variations (e.g.: age, sex, acquisition site).

Algorithm	Subgroup B-ACC	Class B-ACC	Overall B-ACC
Deep Cluster - v2 [Caron et al., 2018]	0.517±0.010	×	×
PCL [Li et al., 2021b]	0.542±0.030	×	×
SwAV [Caron et al., 2020]	0.522±0.008	×	×
SCAN [Van Gansbeke et al., 2020]	0.509±0.008	×	×
SimCLR [Chen et al., 2020]	0.571±0.017	×	×
BYOL [Grill et al., 2020]	0.508±0.006	×	×
VAE [Kingma and Welling, 2014] + UCSL [Louiset et al., 2021]	0.5348±0.016	0.588±0.013	0.459±0.018
BCE + K-Means	0.507±0.005	0.653±0.025	0.428±0.038
SupCon [Khosla et al., 2020]	0.550±0.014	0.656±0.017	0.458±0.017
GM ROI features [Gaser and Dahnke, 2016] + UCSL	**0.590±0.016**	0.653±0.012	0.525±0.011
Deep UCSL	**0.589±0.011**	**0.671±0.018**	**0.543±0.014**
CE (upper bound)	0.615±0.007	×	×

Table 1.9: Results on Neuro-psychiatry task (BP/SZ) on an independent TEST set. Top methods are trained on [SZ+BP] only.

We train all methods using only the class label y (healthy vs disease), but **not** the subgroup labels c. Then, to quantitatively evaluate performance, we use a test set where we know both the class label y and the subgroup label c. About the representation/contrastive learning methods that do *not* have a classification head (e.g., Deep Cluster, SimCLR), we test their performance only in subgroups identification with a K-means algorithm fitted only on target samples (as if they had a perfect classification head). We use three different metrics:

1) *Class Balanced Accuracy (Class B-ACC):* which is the binary Balanced Accuracy between true labels y_j and class predictions $p(y_j|x_j)$.
2) *Subgroup Balanced Accuracy (Subgroup B-ACC):* Balanced Accuracy between true subgroups c_j and inferred ones $p(c_j|x_j)$.
3) *Overall B-ACC*: takes into account both class and subgroup prediction errors: $\frac{1}{2}\frac{TP}{TP+FN}+\frac{1}{2}\frac{TN}{TN+FP}$, where TN and FN are the class true and false negatives, namely the number of healthy and disease samples classified as healthy, respectively. TP is the number of disease samples correctly classified *AND* assigned to the right subgroup. FP is the number of healthy samples classified as disease *OR* disease samples correctly classified but assigned to the wrong subgroup.

To compare with an upper bound, we train a Deep Neural Network to classify between SZ and BD in a fully supervised manner with a Binary Cross-Entropy (BCE). Interestingly, it seems that UCSL's performance highly depends on the feature extraction step. In particular, when using as features the latent vectors of a Variational AutoEncoder (VAE) [Kingma and Welling, 2014], the performance decreases. On the other hand, when using highly specific features obtained from more than 20 years of research (GM ROI features with age confound effect correction), the performance is among the best. We argue that Deep UCSL provides an end-to-end subgroup discovery method that needs no prior knowledge about the feature extraction step and leads to better or similar performances. More results and details can be found in [Louiset et al., 2021, Louiset, 2024].

1.3.3 Contrastive Analysis

Differently from Subgroup discovery methods, in this subsection we will not assume the existence of *distinct* clusters of patients, but we will focus on estimating the (latent) feature dimensions that better describe the pathological heterogeneity, implicitly assuming that there might be a continuum between the healthy population and the different subtypes.

Contrastive Analysis (CA) deals with the discovery of what is common and what is distinctive (i.e., added or modified) of a target domain, here patients, compared to a background one, healthy controls. Both the target (patients) and the background (healthy) datasets are supposed to share uninteresting (healthy) variations. The goal is thus to *identify* and *separate*, in an unsupervised way, the generative factors **common** to both populations from the ones distinctive (i.e., **salient**) only of the target dataset. Here, we present three methods based on 1) Variational AutoEncoders (VAE), 2) Generative Adversarial Networks (GANs) and 3) Contrastive Learning that have been proposed in [Louiset et al., 2024a], [Carton et al., 2024], [Louiset et al., 2024b], respectively.

The most recent CA methods are based on the VAE model and they are called Contrastive VAE (CA-VAE). All these methods share the same general mathematical formulation, which derives from the standard VAE. However, they all either ignore a term of the proposed loss (e.g., KL loss in [Abid and Zou, 2019, Ruiz et al., 2019]) or they don't enforce important assumptions (e.g., independence between common and salient factors in [Weinberger et al., 2022]), which may lead to sub-optimal solutions where salient factors are mistaken for common ones (or viceversa). Chronologically, we have thus first worked on a new CA-VAE method [Louiset et al., 2024a] to overcome such shortcomings. However, all VAE methods, ours included, share a typical downside: a blurry and poor quality image generation. That is why we have then moved towards a generative model with better image quality generation: GAN models. We have thus proposed a novel Contrastive method [Carton et al., 2024] which leverages the high-quality synthesis of GANs and the separation power of InfoGAN [Chen et al., 2016]. To the best of our knowledge, this was the first GAN based method proposed in the context of Contrastive Analysis. Working with generative models, such as GAN and VAE, is particularly suitable for generation and image-level manipulations. However, as shown in [Phuong et al., 2018], VAE and GAN can fail to learn meaningful latent representations, or even learn trivial representations when the decoder is too powerful [Chen et al., 2017]. Conversely, Contrastive Learning (CL) methods have demonstrated outstanding results in many domains producing representations more robust and expressive than VAEs or GANs. This performance gap might be explained by the fact that: 1) CL representations are invariant to user-defined transformations, to which generative models, as VAE, might be highly sensitive, and 2) CL methods implicitly maximize the Mutual Information (MI) between input data and latent features, whereas VAE maximize the log-likelihood, which is only a function of the marginal distribution of the input data and *not* of the latent representations. This motivated our last contribution [Louiset et al., 2024b], where we "harmonized" Contrastive Learning and Contrastive Analysis to better separate common from salient representations.

Contrastive Analysis - Mathematical Framework

Let $(X, Y) = \{(x_i, y_i)\}_{i=1}^N$ be a data-set of images x_i associated with labels $y_i \in \{0, 1\}$, 0 for background and 1 for target. Both background and target samples are assumed to be i.i.d. from the *same* conditional distribution $x_i \sim p_\theta(x|y_i, c_i, s_i)$, that is parameterized by unknown parameters θ and depends on two latent variables: $c_i \in \mathbf{R}^L$ and $s_i \in \mathbf{R}^M$. Our objective is to have a generative model so that: 1- the **common** latent vectors $C = \{c_i\}_{i=1}^N$ capture the common generative factors of variation between the background and target distributions and fully encode the background samples and 2- the **salient** latent vectors $S = \{s_i\}_{i=1}^N$ capture the distinct generative factors of

variation of the target set (*i.e.,* patterns that are only present in the target dataset and not in the background dataset). The separation between c and s can be considered as a weakly supervised learning problem, since the only level of supervision is the population-based label $y_i \in \{0, 1\}$. The user has no knowledge about the common and salient generative factors at training (or test) time.

SepVAE - Contrastive VAE (CA-VAE)

Similarly to previous CA-VAE works [Abid and Zou, 2019, Weinberger et al., 2022, Zou et al., 2022], we assume the generative process: $p_\theta(x, y, c, s) = p_\theta(x|c, s, y)\ p_\theta(c)p_\theta(s|y)p(y)$. Since $p_\theta(c, s|x, y)$ is hard to compute in practice, we approximate it using an auxiliary parametric distribution $q_\phi(c, s|x, y)$ and derive the Evidence Lower Bound (ELBO) of the marginal log-likelihood $\log p(x, y)$: $-\log p_\theta(x, y) \leq \mathbf{E}_{c,s\sim q_{\phi_c,\phi_s}(c,s|x,y)} \log \frac{q_{\phi_c,\phi_s}(c,s|x,y)}{p_\theta(x,y,c,s)}$.

Then, we can develop the lower bound into three terms, a conditional reconstruction term, a common space prior regularization, and a salient space prior regularization (see Eq.1.16). Here, as usually done in previous CA-VAE methods, we assume the independence of the auxiliary distributions (*i.e.:* $q_{\phi_c,\phi_s}(c, s|x, y) = q_{\phi_c}(c|x)q_{\phi_s}(s|x, y)$) and prior distributions (*i.e.:* $p_\theta(c, s) = p_\theta(c)p_\theta(s)$). Both $p_\theta(x|y_i, c_i, s_i)$ (i.e., single decoder) and $q_{\phi_c}(c|x)q_{\phi_s}(s|x, y)$ (i.e., two encoders) are assumed to follow a Gaussian distribution parametrized by a neural network.

To reinforce the independence assumption between c and s, we introduce a Mutual Information regularization term $KL(q(c, s)||q(c)q(s))$. This property is desirable in order to ensure that the information is well separated between the latent spaces. Theoretically, this term is similar to the one in [Abid and Zou, 2019]. However, in [Abid and Zou, 2019], the Mutual Information estimation and minimization are done simultaneously [8], which is theoretically wrong since one needs an independent optimizer, as correctly proposed in [Louiset et al., 2024a].

Differently from previous works, to further reduce the overlap of target and common distributions on the salient space, we also introduce a salient classification loss defined as $\mathbf{E}_{s\sim q_{\phi_s}(s|x,y)} \log p(y|s)$. An illustration of the proposed method, called *SepVAE*, is shown in Fig.1.10. By combining all these losses together, we obtain the final loss $\mathcal{L}$:

$$\begin{aligned}\mathcal{L} = &\underbrace{-\mathbf{E}_{c,s\sim q_{\phi_c,\phi_s}(c,s|x,y)} \log p_\theta(x|c, s, y)}_{\textbf{a) Conditional Reconstruction}} + \underbrace{KL(q(c, s)||q(c)q(s))}_{\textbf{e) Mutual Information}} - \underbrace{\mathbf{E}_{s\sim q_{\phi_s}(s|x,y)} \log p_\theta(y|s)}_{\textbf{d) Salient Classification}} \\ &+ \underbrace{KL(q_{\phi_c}(c|x)||p_\theta(c))}_{\textbf{b) Common Prior}} + \underbrace{KL(q_{\phi_s}(s|x, y)||p_\theta(s|y))}_{\textbf{c) Salient Prior}}\end{aligned} \tag{1.16}$$

Conditional reconstruction The reconstruction term is $-\mathbf{E}_{c,s\sim q_{\phi_c,\phi_s}(c,s|x,y)} \log p_\theta(x|c, s, y)$. Given an image x (and a label y), a common and a salient latent vector can be drawn from q_{ϕ_c,ϕ_s} with the help of the reparameterization trick. We assume that $p(x|c, s, y) \sim \mathcal{N}(d_\theta([c, ys + (1 - y)s'], I)$, *i.e:* $p_\theta(x|c, s, y)$ follows a Gaussian distribution parameterized by θ, centered on $\mu_{\hat{x}} = d_\theta([c, ys+(1-y)s'])$ with identity covariance matrix, and d_θ is the decoder and $[., .]$ denotes a concatenation. Therefore, by developing the reconstruction loss term, we obtain the mean squared error between the input and the reconstruction: $\mathcal{L}_{\text{rec}} = \sum_{i=1}^{N} ||x - d_\theta([c, ys + (1 - y)s'])||_2^2$. Importantly, as in [Abid and Zou, 2019], we set the salient latent vectors of background samples to **s'** = 0. This choice enables isolating the background factors of variability in the common space only.

[8] In [Abid and Zou, 2019], Alg. 1 suggests that the MI estimation and minimization depend on two distinct parameter updates. However, in their code, a single optimizer is used. Moreover, in Sec. 3, authors write: "discriminator is trained simultaneously with the encoder and decoder".

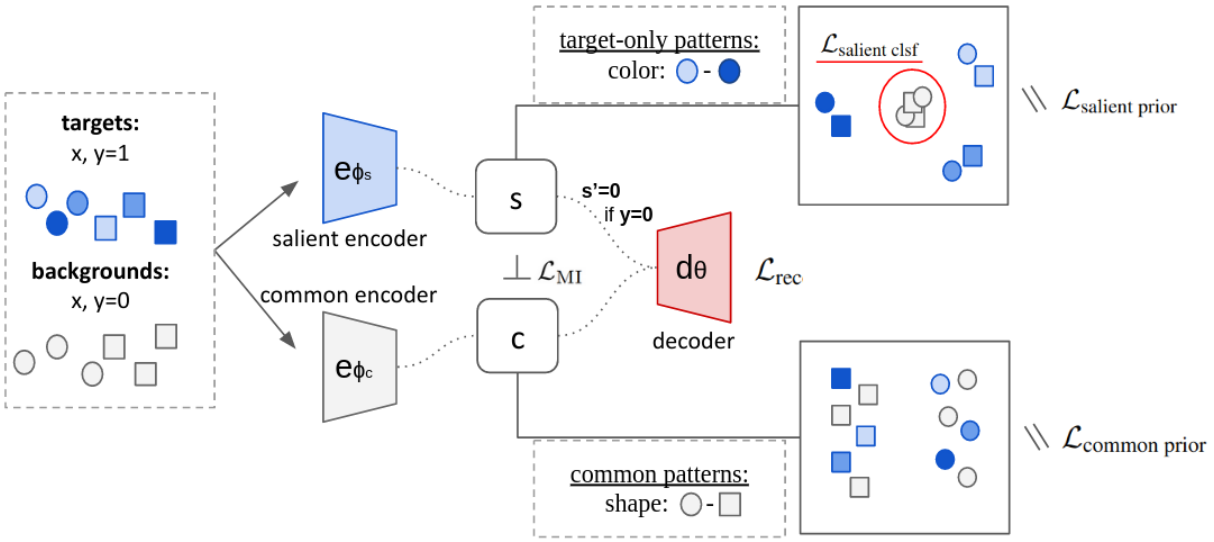


Figure 1.10: Illustration of SepVAE training. Target ($y = 1$) and background ($y = 0$) images are encoded with the same encoders e_{ϕ_s} and e_{ϕ_c}. The first encoder e_{ϕ_s} estimates the salient factors of variation s of the target samples. Background samples' salient space is set to an informationless value $s' = 0$. The second encoder e_{ϕ_c} estimates the common factors c. Images are reconstructed using a single decoder d_θ fed with the concatenation of $\mathbf{c}$ and $\mathbf{s}$. The common space $\mathbf{c}$ should only capture common factors of variability (shape), while the salient space $\mathbf{s}$ should model target-only factors of variability (color).

Common prior Assuming $p(c) \sim \mathcal{N}(0, I)$ and $q_{\phi_c}(c|x) \sim \mathcal{N}(\mu_\phi(x), \sigma_\phi(x, y))$, the KL loss has a closed form solution, as in usual VAE. Here, both $\mu_\phi(x)$ and $\sigma_\phi(x, y)$ are the outputs of the encoder e_{ϕ_c}. This loss is also used in [Abid and Zou, 2019, Weinberger et al., 2022].

Salient prior First, we develop $p_\theta(s) = \sum_y p(y)p_\theta(s|y)$, where $p(y)$ follows a Bernoulli distribution with probability equal to 0.5. This allows us to distinguish the salient priors of background samples ($p(s|y = 0)$) and target samples ($p(s|y = 1)$). Similar to other CA-VAE methods, we assume that $p(s|y = 1) \sim \mathcal{N}(0, I)$ and , as in [Zou et al., 2022], that $p(s|x, y = 0) \sim \mathcal{N}(s', \sqrt{\sigma_p}I)$, with $s' = 0$ and $\sqrt{\sigma_p} < 1$, namely a Gaussian distribution centered on an informationless reference s' with a small constant variance σ_p. We preferred it to a Delta function $\delta(s = s')$ (as in [Weinberger et al., 2022]) because it eases the computation of the KL divergence (i.e., closed form) and it also means that we tolerate a small salient variation (e.g., noisy/erroneous diagnosis labels) in the background samples.

Salient classification The salient prior regularization encourages BG and TG salient factors to match two different Gaussian distributions centered in $s' = 0$, but with different covariance. To further reduce the overlap of target and common distributions on the salient space, we propose to minimize a Binary Cross Entropy (BCE) loss to distinguish the target from background samples in the salient space. Assuming that $p(y|s)$ follows a Bernoulli distribution parameterized by $f_\xi(s)$, a 2-layers classification Neural Network, we obtain a BCE loss between true labels y and predicted labels $\hat{y} = f_\xi(s)$. This loss is *not* used in previous works.

Mutual Information To promote independence between c and s, we minimize their mutual information, defined as the KL divergence between the joint distribution $q(c, s)$ and the product of their marginals $q(c)q(s)$. As in [Abid and Zou, 2019], we use the density-ratio trick [Kim and Mnih, 2018] but, differently from [Abid and Zou, 2019], we correctly implement it.
More information in [Louiset et al., 2024a].

Double InfoGAN - Contrastive GAN

In Double InfoGAN, we use a generative model similar to the one proposed for SepVAE but, to simplify the presentation, we employ a slightly different nomenclature: we call $X = \{\mathbf{x}_i\}$ and $Y = \{\mathbf{y}_j\}$ the healthy and patient data-sets of images, respectively. Thus, differently from before, y_j now refers to a target image and not to a binary label.

In [Chen et al., 2016], differently from standard GAN [Goodfellow et al., 2014], authors propose a new method, called InfoGAN, where they decompose the input noise vector of GANs into two parts: 1) $\mathbf{z}$, which is considered as a nuisance and incompressible noise and 2) $\mathbf{c}$, which should model the salient semantic features of the data distribution. The generator of this new model, $G(\mathbf{z}, \mathbf{c})$, takes as input both $\mathbf{z}$ and $\mathbf{c}$ to generate samples $\mathbf{x}$.

Here, differently from InfoGAN, and similarly to CA-VAE, we change the generative model and decompose the input of G into: 1) $\mathbf{c}$, which should capture the generative factors common to both X and Y, and 2) $\mathbf{s}$, that should model the salient factors proper only to Y. As in GAN [Goodfellow et al., 2014], we introduce a generator $G(\mathbf{c}, \mathbf{s})$ to generate both x and y and a discriminator. The generator G should generate samples that are indistinguishable from the true ones, whereas the discriminator is divided into two modules. The first (and standard) one D is trained to discriminate between fake and real samples. The second module C is trained to correctly classify real samples (i.e., X or Y). As in InfoGAN, we also use one encoder, divided into two modules, Q_c and Q_s, to reconstruct the latent factors $\mathbf{c}$ and $\mathbf{s}$. The discriminator, D and C, and the encoder, Q_c and Q_s, are parametrized as neural networks, that share all layers but the output one. As for SepVAE, we set the salient latent vector of healthy samples to a constant value s' (e.g., $s' = 0$), thus enforcing $\mathbf{c}$ to fully encode alone X. Let $\mathbf{x} = G(\mathbf{c}, \mathbf{s} = s')$ and $\mathbf{y} = G(\mathbf{c}, \mathbf{s})$ be the generated samples. We suppose, and force it in practice, that the latent variables $\mathbf{c} = \{z_1, ..., z_L\}$ and $\mathbf{s} = \{s_1, ..., s_M\}$ are independent and follow a factorized distribution: $P(\mathbf{c}) = \prod_{i=1}^{L} P(c_z)$ and $P(\mathbf{s}) = \prod_{j=1}^{M} P(s_j)$, for X and Y. To correctly estimate both $\mathbf{c}$ and $\mathbf{s}$, we minimize:

$$\min_{G,Q_c,Q_s C} \max_{D} \quad w_{Adv}\mathcal{L}_{Adv}(G, D) + w_{Class}\mathcal{L}_{cl}(G, C) - w_{Info}\mathcal{L}_{Info}(G, Q_c, Q_s) + w_{Im}\mathcal{L}_{Im}(G, Q_c, Q_s) \tag{1.17}$$

In the following, we will describe each term.

Adversarial GAN Loss As in [Goodfellow et al., 2014], G and D are trained together in a *min-max* game using the original nonsaturating GAN (NSGAN) formulation:

$$\mathcal{L}_{Adv}(D, G) = w_{bg}\Big(-\mathbb{E}_{\mathbf{x}_R \sim P(\mathbf{x}_R)}\Big[\log(D(\mathbf{x}_R)\Big] - \mathbb{E}_{z \sim P_x(\mathbf{c})}\Big[\log(1 - (D(G(\mathbf{c}, 0))))\Big]\Big) + w_t\Big(-\mathbb{E}_{\mathbf{y}_R \sim P(\mathbf{y}_R)}$$
$$\Big[\log(D(\mathbf{y}_R)\Big] - \mathbb{E}_{\mathbf{c},\mathbf{s} \sim P_y(\mathbf{c},\mathbf{s})}\Big[\log(1 - (D(G(\mathbf{c}, \mathbf{s}))))\Big]\Big) \tag{1.18}$$

where $D(I)$ indicates the probability that I is real or fake and $\mathbf{x}_R \sim P(\mathbf{x}_R)$ and $\mathbf{y}_R \sim P(\mathbf{y}_R)$ are real images. Furthermore, we choose the same factorized prior distribution $P(\mathbf{c})$ for both X and Y (i.e., $P_x(\mathbf{c}) = P_y(\mathbf{c}) = P(\mathbf{c})$), namely a Gaussian $\mathcal{N}(0, 1)$. We also tested a uniform distribution $\mathcal{U}_{[-1,1]}$ but the results were slightly worse. Instead, about $P(\mathbf{s})$, it should be different between X and Y. We use a Dirac delta distribution centered at 0 for X (i.e., $P_x(\mathbf{s}) = \delta(\mathbf{s} = 0)$) and we have tested several distributions for $P_y(\mathbf{s})$. Depending on the data and related assumptions, one could use, for instance, a factorized uniform distribution, $\mathcal{U}_{(0,1]}$, or a factorized Gaussian $\mathcal{N}(0, 1)$ (ignoring the samples equal to 0). In our experiments, results were slightly better when using $\mathcal{N}(0, 1)$.

Class Loss To make sure that generated images belong to the correct class, we propose to add a second discriminator module C. It is trained on real images to predict the correct class: X or Y.

At the same time, G is trained to produce images correctly classified by C. We (arbitrarily) assign 0 (resp. 1) for class X (resp. Y) and use the binary cross entropy ($\mathcal{B}$). The loss is:

$$\begin{aligned}\mathcal{L}_{cl}(C) =& \mathbb{E}_{\mathbf{x}_R\sim P(\mathbf{x}_R)}\Big[\mathcal{B}(C(\mathbf{x}_R),0)\Big] + \mathbb{E}_{\mathbf{y}_R\sim P(\mathbf{y}_R)}\Big[\mathcal{B}(C(\mathbf{y}_R),1)\Big]\\ \mathcal{L}_{cl}(G) =& \mathbb{E}_{\mathbf{c}\sim P_x(\mathbf{c})}[\mathcal{B}(C(G(\mathbf{c},0)),0)] + \mathbb{E}_{\mathbf{c},\mathbf{s}\sim P_y(\mathbf{c},\mathbf{s})}[\mathcal{B}(C(G(\mathbf{c},\mathbf{s})),1)]\end{aligned} \tag{1.19}$$

Info Loss Similarly to InfoGAN [Chen et al., 2016], we propose two regularization terms based on mutual information, $I((\mathbf{c},\mathbf{s});\mathbf{y})$ and $I((\mathbf{c},\mathbf{s}=s');\mathbf{x})$, to encourage informative latent codes. However, in our case, these two terms are *not* added to disentangle between informative and nuisance generative factors, as in InfoGAN [Chen et al., 2016], but to enforce *the separation* between common and salient factors. Indeed, the maximization of these two regularity terms should enforce $\mathbf{c}$ to fully encode X and at the same time to be informative for the generation of Y. In parallel, $\mathbf{s}$ should only encode distinctive semantic information of Y. Since $\mathbf{c}$ and $\mathbf{s}$ are independent *by construction*, the mutual information $I((\mathbf{c},\mathbf{s});\cdot)$ can be decomposed into the sum of the two mutual information $I(\mathbf{c};\cdot) + I(\mathbf{s};\cdot)$. Thus, similarly to InfoGAN, we can retrieve four lower bounds.

$$\begin{aligned}I(\mathbf{c};\mathbf{y}) &\geq \mathbb{E}_{\mathbf{c}\sim P_y(\mathbf{c}),\mathbf{s}\sim P_y(\mathbf{s}),\mathbf{y}\sim G(\mathbf{c},\mathbf{s})}\log(Q_c(\mathbf{c}|\mathbf{y})) + H(\mathbf{c})\\ I(\mathbf{s};\mathbf{y}) &\geq \mathbb{E}_{\mathbf{c}\sim P_y(\mathbf{c}),\mathbf{s}\sim P_y(\mathbf{s}),\mathbf{y}\sim G(\mathbf{c},\mathbf{s})}\log(Q_s(\mathbf{s}|\mathbf{y})) + H(\mathbf{s})\\ I(\mathbf{c};\mathbf{x}) &\geq \mathbb{E}_{\mathbf{c}\sim P_x(\mathbf{c}),\mathbf{s}\sim P_x(\mathbf{s}),\mathbf{x}\sim G(\mathbf{c},\mathbf{s})}\log(Q_c(\mathbf{c}|\mathbf{x})) + H(\mathbf{c})\\ I(\mathbf{s};\mathbf{x}) &\geq \mathbb{E}_{\mathbf{c}\sim P_x(\mathbf{c}),\mathbf{s}\sim P_x(\mathbf{s}),\mathbf{x}\sim G(\mathbf{c},\mathbf{s})}\log(Q_s(\mathbf{s}|\mathbf{x})) + H(\mathbf{s})\end{aligned} \tag{1.20}$$

As in [Chen et al., 2016, Lin et al., 2020], to promote stability and efficiency, we model the two auxiliary distributions, Q_c and Q_s, as factorized distributions. Beside a factorized Gaussian distribution with identity covariance, we have also tested a factorized Laplace distribution $\mathbf{L}(\mu, b)$ with $b = 1$. This brings to a $l1$ reconstruction loss instead of a standard $l2$, and showed better performance in practice. To better train Q_s, and since we know that $\mathbf{s}$ should be equal to 0 for real images of domain X (i.e., $\mathbf{x}_R \sim P(\mathbf{x}_R)$), we also add as regularization the lower bound of the mutual information $I(\mathbf{s};\mathbf{x}_R)$. As before, we fix $P_x(\mathbf{s}) = \delta(\mathbf{s} = 0)$. The sum of these five lower bounds defines the $\mathcal{L}_{Info}$ loss:

$$\begin{aligned}&\mathcal{L}_{Info}(G,Q_c,Q_s) = w_{bg}\mathbb{E}_{\mathbf{c}\sim P_y(\mathbf{c})}\Big[w^c_{Info}|(Q_c(G(\mathbf{c},0)) - \mathbf{c}| + w^s_{Info}|Q_s(G(\mathbf{c},0)) - 0|\Big] +\\ &w_t\mathbb{E}_{\mathbf{c},\mathbf{s}\sim P_y(\mathbf{c},\mathbf{s})}\Big[w^c_{Info}|(Q_c(G(\mathbf{c},\mathbf{s})) - \mathbf{c}| + w^s_{Info}|Q_s(G(\mathbf{c},\mathbf{s})) - \mathbf{s}|\Big] + w^{real}_{Info}\mathbb{E}_{\mathbf{x}_R\sim P(\mathbf{x}_R)}\Big[|(Q_s(\mathbf{x}_R)) - 0|\Big]\end{aligned} \tag{1.21}$$

Image reconstruction loss Differently from usual GAN models, we also propose to maximize the log-likelihood $\log(P(\mathbf{y}))$ (and $\log(P(\mathbf{x}))$) of the generated images based on the proposed model. Indeed, no likelihood is generally available for optimizing the generator G in a GAN model [Goodfellow et al., 2014]. However, here, given a real image $\mathbf{y}_R$ (or $\mathbf{x}_R$), we can use the auxiliary encoder $Q = (Q_s, Q_c)$ to estimate the latent factors $\hat{c}$ and $\hat{s}$ that should generate $\mathbf{y}_R$ (or $\mathbf{x}_R$) and then maximize (an approximation) of the log-likelihood of the generated images $\mathbf{y} = G(\hat{c},\hat{s})$ (or $\mathbf{x} = G(\hat{c},0)$):

$$\log P(\mathbf{y}) \geq \mathbb{E}_{\mathbf{y}_R\sim P(\mathbf{y}_R),(\mathbf{c},\mathbf{s})\sim Q(\mathbf{c},\mathbf{s}|\mathbf{y}_R)}\log P(\mathbf{y}|\mathbf{c},\mathbf{s},\mathbf{y}_R) - \mathbb{E}_{\mathbf{y}_R\sim P(\mathbf{y}_R)}KL(Q(\mathbf{c},\mathbf{s}|\mathbf{y}_R)||P(\mathbf{c},\mathbf{s}|\mathbf{y}_R)) \tag{1.22}$$

We notice that the second term should tend towards 0 during training thanks to the previous Info Loss.[9] We can thus approximate $\log P(\mathbf{y})$ by computing only the left term and modeling $P(\mathbf{y}|\mathbf{c},\mathbf{s},\mathbf{y}_R)$ as a Laplace distribution $\mathbf{L}(\mu, b)$ with $b = 1$. We use a Laplace distribution, instead of

[9] Lower bounds become tight as Q resembles the true P.

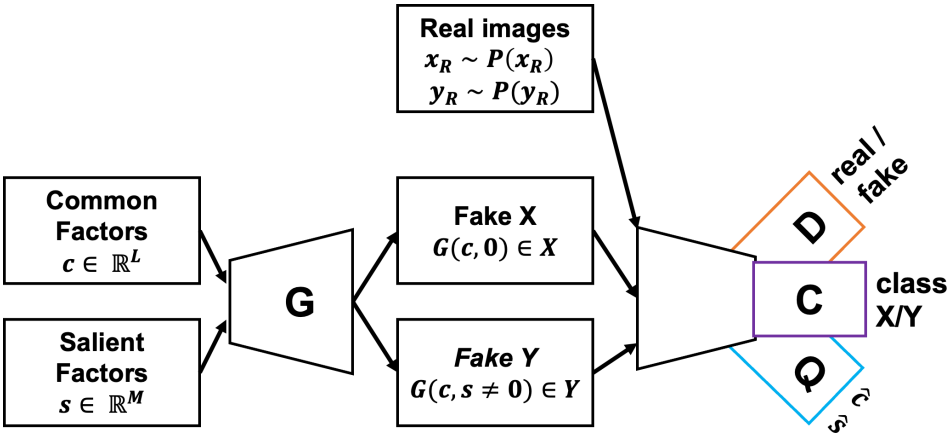


Figure 1.11: Double InfoGAN. Our model takes two inputs: **c** (common factors) and **s** (salient factors). The generator G produces fake images that, together with the real images, are passed to a discriminator and encoder. The discriminator has two modules: D for detecting real from fake images, and C for classyfing images in the correct domain (i.e., X or Y). The encoder Q has two modules, Q_c and Q_s, to reconstruct the latent factors $(\hat{\mathbf{c}}, \hat{\mathbf{s}})$. D, C and Q share all layers but the last one.

a Gaussian one, since it has been shown, for instance in [Isola et al., 2017], that a $l1$-loss encourages sharper and better image reconstructions than a $l2$-loss. Similar computations can be done for $\log P(\mathbf{x})$. We define $\mathcal{L}_{Im}(G, Q_c, Q_s) = \log P(\mathbf{x}) + \log P(\mathbf{y})$:

$$\mathcal{L}_{Im}(G, Q_c, Q_s) = w_{bg} \mathop{\mathbb{E}}_{\substack{\mathbf{x}_R \sim P(\mathbf{x}_R) \\ \hat{c}=Q_c(\mathbf{x}_R)}} \Big[|G(\hat{c}, 0) - \mathbf{x}_R| \Big] + w_t \mathop{\mathbb{E}}_{\substack{\mathbf{y}_R \sim P(\mathbf{y}_R) \\ \hat{c},\hat{s}=Q(\mathbf{y}_R)}} \Big[|G(\hat{c}, \hat{s}) - \mathbf{y}_R| \Big] \tag{1.23}$$

A visual overview of the method is shown in Fig.1.11

SepCLR - Contrastive Representation Learning

As last method, we present SepCLR, where we use the same generative model as before and the nomenclature of Double InfoGAN (i.e., y_j is a patient/target sample).

Differently from the two previous generative models, we propose to reformulate the Contrastive Analysis problem under the lens of the well-known InfoMax principle [Bell and Sejnowski, 1995, Hjelm et al., 2019] and to leverage the representation power of Contrastive Learning (CL) to estimate the MI terms of our newly proposed Contrastive Analysis setting. Since we want the common factors c to be representative of both datasets, we propose to maximize the mutual information I between **c** and both datasets X and Y. Similarly, we propose maximizing the mutual information between the salient factors **s** and **only** the target samples Y. Furthermore, since we want the background samples x to be fully encoded by **c**, we enforce the salient factors **s** of x to be always equal to a constant value s' (*i.e.,* no information): $x_i \sim p_\theta(x|c_i, s_i = s')$. Mathematically, we do that by minimizing the Kullback–Leibler divergence D_{KL} between $p(\mathbf{s}|x)$ and $\delta(s')$, a Dirac Delta distribution centered at s'. Eventually, to enforce the separation (*i.e.,* independence) between **c** and **s**, we also propose to use $I(\mathbf{c}, \mathbf{s}) = 0$ as a regularization constraint. This choice, differently from simply minimizing the MI, avoids undesirable results which could bring to a trivial solution, as shown in [Louiset et al., 2024b], where **c** and/or **s** would contain no information. Instead, we propose a new method, called *kernel-based Joint Entropy Maximization (k-JEM)*, to estimate and maximize their joint entropy $H(\mathbf{c}, \mathbf{s})$, without requiring any assumptions about the form of its pdf nor a neural network-based approximation. More information can be found in [Louiset et al., 2024b].

Our objective is to *separate* and *infer* the common $\mathbf{c}$ and salient $\mathbf{s}$ factors given the input data X and Y. We use two probabilistic encoders, f_{θ_c} and f_{θ_s}, parameterised by θ_c and θ_s, to approximate the conditional distributions $p(c|\cdot)$ and $p(s|\cdot)$ respectively. The two encoders are shared between X and Y. Furthermore, as commonly done in recent representation learning papers, we assume to have multiple views v of each image x (or y) generated via a stochastic augmentation function t: $v = t(\cdot)$. By denoting $\mathbf{c} = f_{\theta_c}(v)$, $\mathbf{s} = f_{\theta_s}(v)$, $\mathbf{s}_x = f_{\theta_s}(t(x))$, our goal becomes finding the optimal parameters $\theta^* = \{\theta_c^*, \theta_s^*\}$ that maximize the following cost function:

$$\arg\max_{\theta} \underbrace{\lambda_C(I(\mathbf{x};\mathbf{c}) + I(\mathbf{y};\mathbf{c}))}_{\text{Common InfoMax}} + \underbrace{\lambda_S I(\mathbf{y};\mathbf{s})}_{\text{Salient InfoMax}} \quad \text{s.t.} \quad \underbrace{D_{KL}(\mathbf{s}_x||\delta(s')) = 0}_{\text{Information-less hyp.}} \text{ and } \underbrace{I(\mathbf{c},\mathbf{s}) = 0}_{\text{Independence hyp.}} \tag{1.24}$$

We propose to estimate the MI terms, $I(\mathbf{x};\mathbf{c})$, $I(\mathbf{y};\mathbf{c})$ and $I(\mathbf{y};\mathbf{s})$, via a formulation similar to the alignment and uniformity terms introduced in [Wang and Isola, 2020]. Let f_{θ_C} be the common encoder and $\mathbf{c} \sim f_{\theta_C}(t(.))$ be the common representations. The MI $I(\mathbf{x};\mathbf{c})$ (same reasoning is also valid for the other MI terms) can be decomposed into:

$$I(\mathbf{x};\mathbf{c}) = \underbrace{-\mathbf{E}_{\mathbf{x}\sim p_x} H(\mathbf{c}|\mathbf{x})}_{\text{Alignment}} + \underbrace{H(\mathbf{c})}_{\text{Entropy}} \tag{1.25}$$

Entropy (Uniformity). As in [Wang and Isola, 2020], the entropy can be computed with a non-parametric estimator described in [Ahmad and Lin, 1976]. To do so, we compute the approximate density function $\hat{p}(c_i)$ with a Kernel Density Estimator as in [Parzen, 1962, Rosenblatt, 1956], based on views v_j (random augmentation of an image with index j) uniformly sampled from both the target dataset $f_{\theta_c}(t(\mathbf{y})) \sim p(\mathbf{c}|\mathbf{y})$ and the background dataset $f_{\theta_c}(t(\mathbf{x})) \sim p(\mathbf{c}|\mathbf{x})$. We choose a von Mises-Fischer kernel with concentration parameter $\frac{1}{\tau}$. As in [Wang and Isola, 2020], we optimize a lower bound of this estimator called $-\mathcal{L}_{\text{unif}}$:

$$\mathcal{L}_{\text{unif}} = \log \frac{1}{N_X+N_Y} \sum_{i=1}^{N_X+N_Y} \frac{1}{N_X+N_Y} \sum_{j=1}^{N_X+N_Y} \exp \frac{-||f_{\theta_C}(v_i) - f_{\theta_C}(v_j)||_2^2}{2\tau} + \underbrace{\log\sqrt{2\pi\tau}}_{\text{Constant term}} \tag{1.26}$$

Alignment: Differently from [Wang and Isola, 2020], we propose to estimate the conditional entropy $-H(c|x)$ with a re-substitution entropy estimator. We compute the approximate density function $\hat{p}(c_i|x_i)$ with a Kernel Density Estimator based on samples uniformly drawn from the conditional distribution $c_i^k \sim p(\mathbf{c}|x_i)$, where $c_i^k = f_\theta(v_i^k)$ and v_i^k are K views obtained via the stochastic process $t(.)$. As for the entropy term, we choose a von Mises-Fischer kernel with concentration parameter $\frac{1}{\tau}$ to derive an L2 distance between the views. Our formulation generalizes [Wang and Isola, 2020], as we directly retrieve a multi-view alignment term between K positive views of the same image and not a single-view alignment as in [Wang and Isola, 2020]. Combining the background alignment $-H(c|x)$ and the target alignment $-H(c|y)$, we obtain:

$$\mathcal{L}_{\text{align}} = -\frac{1}{N_X+N_Y} \sum_{i=1}^{N_X+N_Y} \log \frac{1}{K} \sum_{k=1}^{K} \exp \frac{-||f_{\theta_C}(v_i) - f_{\theta_C}(v_i^k)||_2^2}{2\tau} + \underbrace{\log(\sqrt{2\pi\tau})}_{\text{Constant term}} \tag{1.27}$$

On the relation with $\mathbf{I}(\mathbf{f}_\theta(\mathbf{v}), \mathbf{f}_\theta(\mathbf{v}'))$. Many recent representation learning works ([Chen and He, 2021, Wang and Isola, 2020]) maximize the MI between two views v and v' of $\mathbf{x}$: $I(f_\theta(v), f_\theta(v')$. Inspired by the InfoMax principle, we propose instead maximizing $I(f_\theta(v), \mathbf{x})$. As shown in [Tschannen et al., 2020], by directly applying the *data processing inequality*, one can demonstrate that $I(f_\theta(v), f_\theta(v')$ is a lower bound of $I(f_\theta(v), \mathbf{x})$.

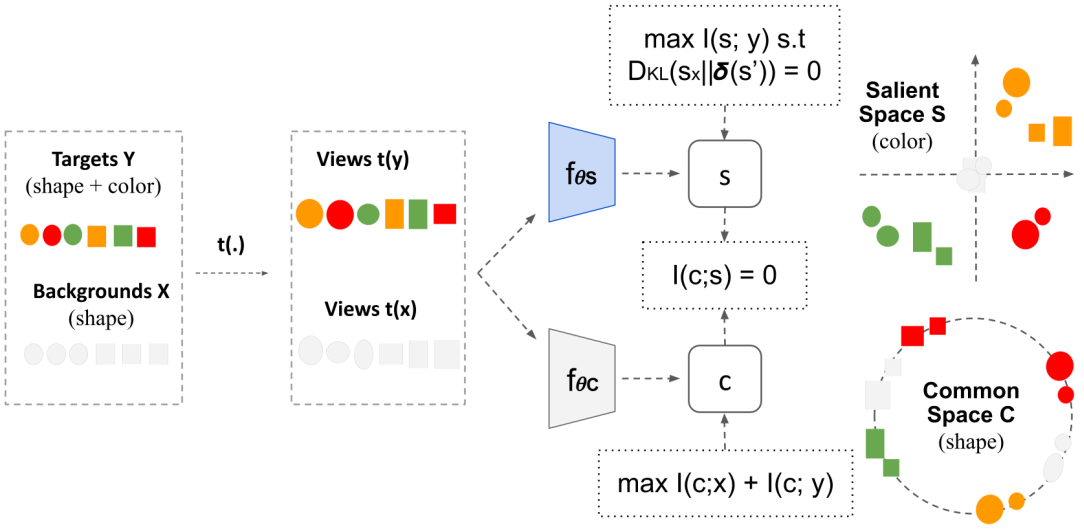


Figure 1.12: SepCLR is trained to identify and separate the salient patterns (color variations) of the target dataset Y from the common patterns (shape) shared between background X and target dataset Y. Views (transformations $t(\cdot)$) of both datasets are fed to two different encoders, one for the salient space (f_{θ_s}) and one for the common space (f_{θ_c}). In the hyperspherical common space, C, embeddings of views of the same image (from both X and Y) are aligned, while embeddings from different images are repelled (max $I(c;x) + I(c;y)$). This enforces C to represent the shared patterns (shape). In the salient space S, which is a Euclidean space, in order not to capture background variability (*i.e:* shape), background embeddings are aligned onto an information-less null vector **s'** ($D_{KL}(s_x||\delta(s')) = 0$). Furthermore, embeddings of views of the same image (only from Y) are aligned while embeddings from different images are pushed away from each other, and they are all repelled from **s'** (max $I(s;y)$). This enforces S to capture only the salient patterns of Y (color). To limit the information leakage between C and S, their MI is constrained to be null, *i.e*: $I(c;s=0)$.

Evaluation

Schizophrenia Here, we evaluate the performance of our methods in separating healthy from pathological latent mechanisms that drive neuro-anatomical variability in schizophrenia. The goal is to capture the pathological factors of variability in the salient space, that should correlate with clinical scales, such as positive symptoms (SAPS), and negative symptoms (SANS), while isolating in the common space the patterns related to demographic variables, such as age and sex, or acquisition sites. For each experiment, we gather T1w anatomical VBM [Ashburner and Friston, 2000] pre-processed images of both schizophrenic patients and healthy controls from the datasets SCHIZCONNECT-VIP [Wang et al., 2016b] and BSNIP [Tamminga et al., 2014]. We divide them into 5 TRAIN, VAL splits (0.75, 0.25) and evaluate in a cross-validation scheme the average performance of SOTA CA-VAEs as well as the proposed methods. Results in Tab. 1.10 show that the salient factors estimated using our methods better predict schizophrenia-specific variables of interest: SAPS (Scale of Positive Symptoms), SANS (Scale of Negative Symptoms), and diagnosis. On the other hand, salient features are shown to be poorly predictive of demographic variables: age, sex, and acquisition site. More details can be found in [Louiset et al., 2024a, Louiset et al., 2024b].

CelebA with accessories To evaluate the generative performance of Double InfoGAN and SepVAE[10], we used the the CelebA with attributes dataset [Liu et al., 2015b], where the target set (Y) contains images of celebrities wearing glasses or hats while background images X show no accessories. In Fig. 1.13 and 1.13, we can see that both methods correctly retrieve the salient factors (glasses and hats), since these are kept when swapping the salient features (i.e., use $\hat{s}_y$ instead than 0 for X and 0 instead than $\hat{s}_y$ for Y). This is also confirmed by quantitative results presented

[10] we did not use a decoder in SepCLR since it was not the goal of the paper and it decreased the performance

	Age MAE		Sex B-ACC		Site B-ACC	
	C ↓	S ↑	C ↑	S ↓	C ↑	S ↓
cVAE	6.43±0.18	7.27±0.25	75.06±3.48	74.99±2.15	65.12±4.06	59.62±5.42
ConVAE	6.40±0.26	7.46±0.18	74.45±1.80	72.72±1.32	60.42±3.67	54.46±2.46
MM-cVAE	6.55±0.18	7.10±0.34	72.80±3.95	72.15±2.47	63.24±1.41	56.69±9.84
SepVAE	**6.40±0.13**	**7.98±0.25**	74.19±1.81	72.61±2.19	63.89±2.16	44.10±5.78
SepCLR-k-JEM	6.64±0.21	7.72±0.45	**76.5±1.98**	**70.85±1.89**	**66.94±5.06**	**42.40±4.91**
	SANS MAE		SAPS MAE		Diagnosis	
	C ↑	S ↓	C ↑	S ↓	C ↓	S ↑
cVAE	5.89±0.67	4.35±0.26	4.65±0.34	2.98±0.18	60.66±2.63	68.24±5.42
ConVAE	6.17±0.45	3.95±0.28	4.50±0.37	2.76±0.18	61.85±2.60	58.53±4.87
MM-cVAE	6.78±0.54	4.92±0.58	4.52±0.33	3.16±0.05	64.25±2.98	70.94±4.08
SepVAE	7.05±0.67	4.14±0.39	4.79±0.67	2.60±0.27	60.90±1.75	79.15±3.39
SepCLR-k-JEM	**9.17±2.49**	**3.74±0.12**	**5.54±0.70**	**2.52±0.16**	**60.16±1.19**	**79.90±1.57**

Table 1.10: Average performance on the prediction of disorder-specific variables (*i.e.*, SANS, SAPS, and diagnosis) and common variables (Age, Sex, Site) using the estimated salient S and common C factors of validation images. MAE=Mean Absolute Error. B-ACC=Balanced Accuracy. Best in **bold**.

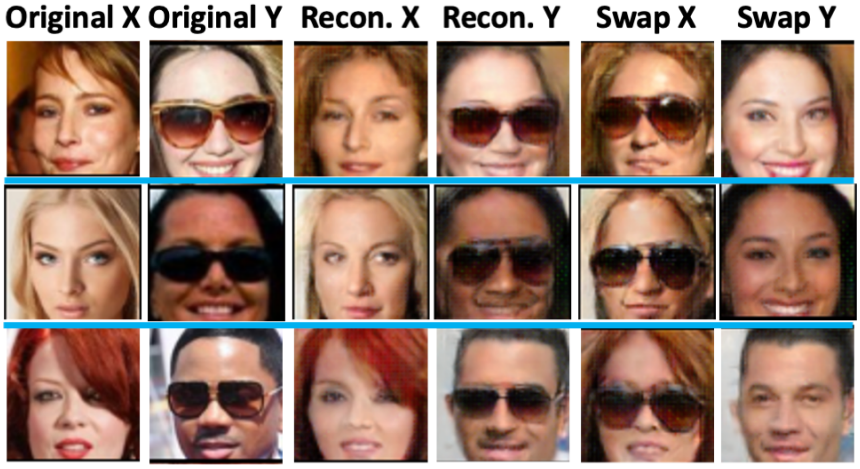


Figure 1.13: Double InfoGAN: image reconstruction and salient feature swap with the CelebA with accessories dataset, where the target set (Y) contains images of celebrities wearing glasses or hats while background images X show no accessories.

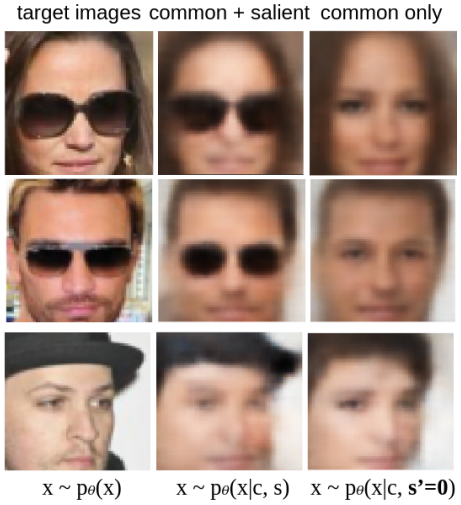


Figure 1.14: SepVAE: reconstructions with CelebA accessories dataset.

in [Carton et al., 2024, Louiset et al., 2024a]. However, the generative quality of Double InfoGAN is definitely better than the one of SepVAE, as expected. This was also confirmed by quantitative measures, such as the FID score, presented in [Carton et al., 2024].

1.4 Conclusions and Perspectives

In this Chapter, we presented several methodological contributions whose final goals were 1) identifying anatomical patterns predictive of brain disorders and 2) parsing their heterogeneity to estimate either distinct subtypes or relevant generative factors. Thanks to a recently proposed geometric approach for Contrastive Learning, we proposed new losses well adapted to integrate prior clinical information and learn a complete and robust representation of the healthy population. We showed that, by transferring it to smaller-scale clinical datasets, we can improve the discriminative performance, and established the new state-of-the-art prediction performance on bipolar disorder

detection from brain anatomical images ($> 78\%$ AUC on both internal and external tests, with 1173 healthy controls and 471 patients). Furthermore, we also improved current state-of-the-art results in Subtype Detection and Contrastive Analysis, which are two promising frameworks to analyze brain disorder heterogeneity. Aggregating other modalities (e.g., functional or diffusion MRI, genetics) to perform representation learning remains an exciting challenge that might be solved with (geometric) contrastive learning, or other self-supervised methods. It would improve our understanding of brain disorders and possibly pave the way towards personalized medicine in psychiatry through predictive models of clinical outcome.

Longitudinal data

The methodological contributions presented in the previous sections were developed for cross-sectional studies. However, the natural aging of the brain and the neurodevelopmental characteristic of some brain disorders would favor longitudinal analyses. The main reason why most of the studies in the literature (ours included) focus on a single time point is mainly due to data availability. Indeed, large datasets with multiple time points are much more costly and difficult to obtain. Recently, few research longitudinal datasets have emerged, such as UKBiobank [Littlejohns et al., 2020] and Alzheimer's Disease Neuroimaging Initiative (ADNI) [Petersen et al., 2010], where at least two time points are usually available. Adapting the previously presented geometric framework to longitudinal data would be a promising research direction to fully leverage the anatomical similarity of the same subject through successive time points, similarly to [Zhao et al., 2021, Ouyang et al., 2022b, Ouyang et al., 2021a, Zeghlache et al., 2023, Sun et al., 2023, Ren et al., 2022]. Furthermore, we could also adapt the presented Contrastive methods, such as SepCLR, to disentangle the natural anatomical variations due to aging from the pathological ones, as proposed in [Ouyang et al., 2022a, Zhao et al., 2021, Ouyang et al., 2021b, Ouyang et al., 2023, Zeghlache et al., 2024]. Average anatomical brain representations conditioned on the chronological age, as estimated in this Chapter, could be used as reference point for the longitudinal analysis.

Debiasing losses for regression

In Table 1.8, we can notice that $\mathcal{L}^{exp}$ shows a better external MAE than the Baseline and ComBat but it has also a low debiasing capability since the Balanced Accuracy should be equal to random chance, namely $1/n_{\text{sites}} = 1/64 \approx 1.56$. Unfortunately, simply adding the previously presented FairKL regularization loss does not improve the debiasing capability (BAcc=5.2 with a ResNet-18) of the $\mathcal{L}^{exp}$ loss, even if it improves the external MAE (3.56) and the challenge score (1.47). Finding an appropriate debiasing loss for regression is an interesting and important future research direction.

Disentanglement in Contrastive Analysis

Learning disentangled (or factorized) representations in Contrastive Analysis would increase interpretability and thus clinical utility. Most of the recent disentanglement methods [Bengio et al., 2013, Higgins et al., 2017, Burgess et al., 2017, Kim and Mnih, 2018, Chen et al., 2016, Chen et al., 2018, Locatello et al., 2020b] assume the existence of independent generative factors that capture distinct, noticeable and semantically meaningful variations in the datasets. Since unsupervised disentanglement has been shown to be impossible [Locatello et al., 2019], inductive biases, class labels or weak information is required [Locatello et al., 2020a, Shu et al., 2020]. These methods have shown good results in toy datasets that have been built with independent factors, such as dSprites [Matthey et al., 2017]. However, we have shown in [Carton et al., 2024] that disentangling salient (or common) factors in toy datasets for Contrastive Analysis is much more difficult than in a single

data-set. Indeed, we proposed a new toy dataset where background X images consisted of 4 MNIST digits regularly placed in a square, while target Y images had dSprites element added on top of the same 4 MNIST digits. We adapted the Contrastive Regularizer (CR) module of InfoGAN-CR [Lin et al., 2020] for our model, obtaining a maximum fvae score of 0.47. For comparison, InfoGAN-CR achieves a fvae score of 0.88 on the dsprite dataset alone. Furthermore, there is an important difference between factorized toy datasets, such as dSprites, and real datasets of medical images: generative factors might be correlated. For instance, it is known that age or sex, thus common factors, might be correlated with salient factors in certain psychiatric disorders, such as schizophrenia. Exploring disentanglement regularizations more suited for CA, and in particular when dealing with unknown correlated factors, like in [Träuble et al., 2021], is left as future work.

Identifiability in Contrasive Analysis

An important question in Contrastive Analysis (CA), is the identifiability of the models. Namely, under which conditions can the models recover the true latent factors of the underlying data-generating process. Recent works have shown that non-linear models, VAEs included, are generally not identifiable. To obtain identifiability, two different solutions have been proposed: 1) either regularizing [Kivva et al., 2022] the encoder or 2) introducing an auxiliary variable so that the latent factors are conditionally independent given the auxiliary variable [Hyvarinen et al., 2019, Khemakhem et al., 2020]. In CA, neither of these solutions may be used [11]. Even though all the proposed methods effectively *separate* common from salient factors, they do not assure that *all* true generative factors have been identified (like all other existing CA methods). This is a serious limitation of all CA methods that we leave as future work. Inspired by [Wyner, 1975, Huang and Gamal, 2024], a possible research direction would be adding an information-theoretic loss that quantifies the common and salient information content so that, under realistic assumptions, the model could be identifiable. Another interesting direction is given by multi-modal (called also multi-view) causal representation learning [Yao et al., 2024]. In such works, authors assume that we have multiple views (modalities) for each sample and different views are functions of only some of the generative factors. We could leverage these theoretical results by seeing X and Y as different "views" generated by different factors.

[11] The dataset label could be considered as an auxiliary variable, but it does not make c and s independent

Chapter 2

Glioblastoma atlas estimation

This chapter has been published in [François et al., 2021, François et al., 2022, Maillard et al., 2022, Hu et al., 2020] and is based on the PhD theses of M. Maillard and A. Francois, co-directed with I. Bloch (Télécom Paris) and J. Glaunès (MAP5), respectively.

Contents

2.1 Clinical context

Glioblastoma (GBM) is a type of aggressive brain cancer that is still considered incurable [Tykocki and Eltayeb, 2018] and it accounts for more than 60% of all brain tumours in adults [Rock et al., 2012]. The symptoms include headaches, focal neurologic deficits, confusion, memory loss, personality changes, or seizures [Alifieris and Trafalis, 2015]. Furthermore, the median overall survival ranges from 12 to 20 months depending on the study [Lacroix et al., 2001, Stummer et al., 2006, Pallud et al., 2015]. In the United States, the 5 and 10-year survival rate is estimated to be respectively 5% and 2.6% [Ostrom et al., 2014]. A common treatment against GBM is the resection of the tumor followed by radiotherapy and adjuvant therapy [Alifieris and Trafalis, 2015]. Despite this aggressive treatment, recurrence almost always occurs in proximity to the original lesion (75 to 90 percent of patients according to [Tykocki and Eltayeb, 2018]). The low survival rate and negative prognosis have fostered a lot of research for a better understanding of the behavior of this kind of tumor. Clinical evidence suggests that tumor *size*, *location*, and *shape* could be important factors related to recurrence and seizures.

Indeed, tumour location is a key parameter in the care of patients with glioblastoma because it correlates with demographic characteristics, symptoms, surgical management, delivery of subsequent oncologic treatments, and, ultimately affects the patient's prognosis. [Roux et al., 2019]. Previous pathogenesis research has shown that the most frequent location is the cerebral hemispheres. 95%

of glioblastoma arise in the supratentorial region (upper part), while only a few in the cerebellum, brainstem and spinal cord [Nakada et al., 2011]. At a macroscopic scale, glioblastomas are quite heterogeneous in form and irregularly shaped but usually arise in white matter [Nakada et al., 2011]. It has been shown that depending on the lobe where the tumour arises, symptoms vary. For example, patients with a glioblastoma located in the temporal lobe often show hearing and visual problems, while those who have one in the frontal lobe might demonstrate personality change [Hanif et al., 2017]. Furthermore, the distribution of edema/necrosis leads to different secondary effects in the patient. For instance, a gradual increase in tumor size and surrounding edema might lead to a shift in intracranial contents, resulting in headaches. A tumor that resides in the eloquent cortex brain stem or basal ganglia cannot go through surgery and these patients usually have worse prognoses [Mrugala, 2013]. Several studies [Bilello et al., 2016, Parisot et al., 2016, Simpson et al., 1993] saw no difference in survival for different tumour sizes and asses that patients with frontal lobe tumours survived longer than those with temporal or parietal lobe lesions, concluding that *localisation is a crucial prognosis indicator*.

The standard research protocol to detect brain tumors is Magnetic Resonance Imaging (MRI) [Thust et al., 2018] as it constitutes a non-ionizing and non-invasive method to produce detailed images of the brain internal structures. Different MRI modalities are usually used as they provide different contrast between tissues, highlighting specific tumor parts. The commonly acquired modalities are T1, T1 contrast-enhanced (T1ce), T2, and Flair [Menze and others, 2015]. As seen in Figure 2.1, the contrast in each modality is different and each one highlights different tumor regions. For instance, the T1ce image shows the necrotic region and the enhancing tissue, while the Flair and T2 better reveal the edema. However, acquiring multiple modalities is usually not possible in a clinical setting due to a limited number of physicians and scanners, and to limit costs and scan time. Most of the time, only one modality is acquired.

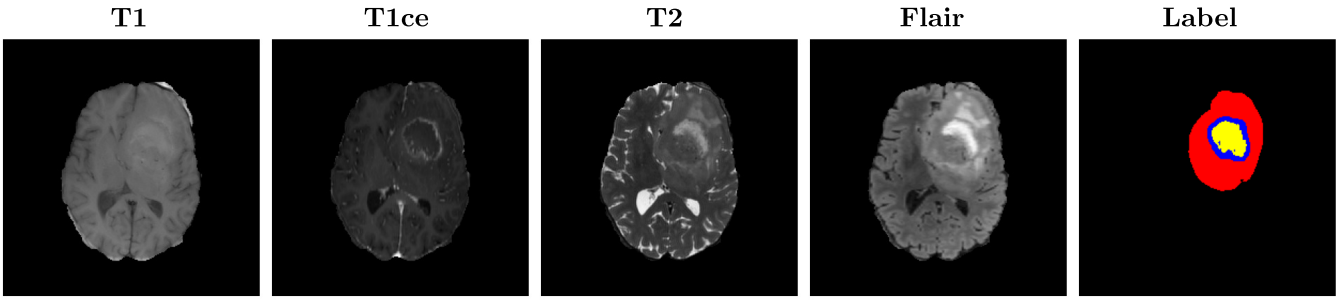


Figure 2.1: Example of the four MRI modalities commonly used to study brain tumors. Last figure presents the corresponding manual segmentation of the brain glioblastoma. The yellow part is the necrotic tumor, blue is the tumor core, and red is the edema.

In the following, we will use the notation of the challenge Brats[1], where authors have subdivided the tumor into four types of intra-tumoral structures: 1) Necrosis, 2) Edema, 3) Non-Enhancing tumor, and 4) Enhancing Tumor (see Fig.2.2 for an exemple and please refer to [Menze and others, 2015] for more information).

2.2 Clinical Goal and Challenges

The goal of this chapter is to propose a method to estimate a *3D atlas* of glioblastoma using a population of MR brain images. In medical imaging, a statistical atlas is usually defined as an aver-

[1]https://www.med.upenn.edu/cbica/brats/

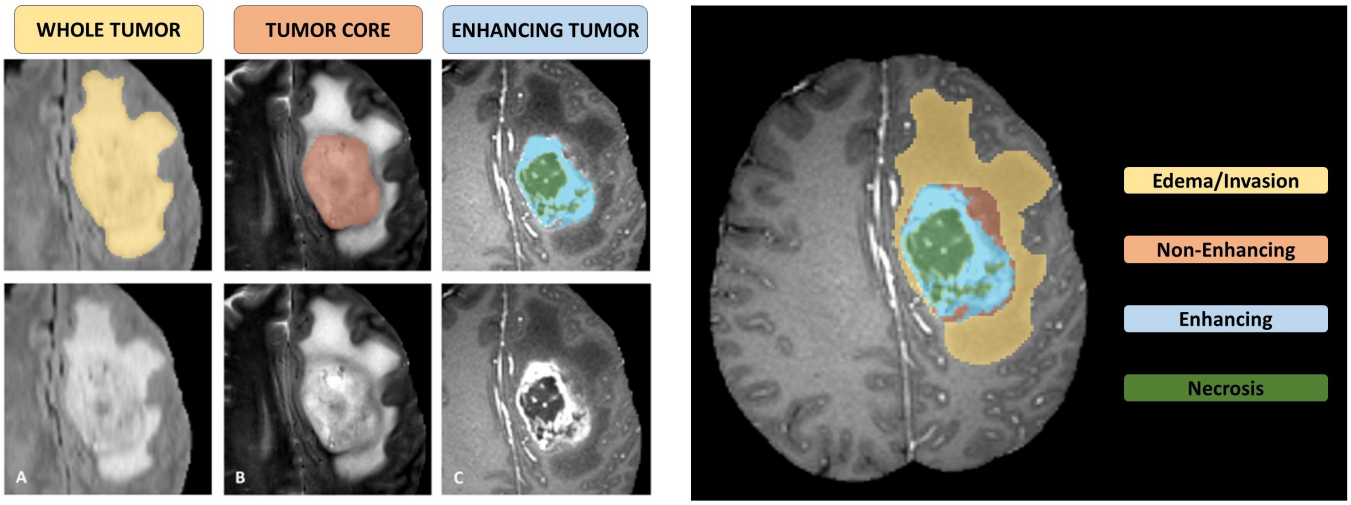


Figure 2.2: BraTS annotations. On the left: the whole tumor visible in FLAIR (A), the tumor core visible in T2 (B), the enhancing tumor structures visible in T1c (blue), surrounding the necrotic components of the core (green) (C). On the right: labels of the tumor structures: edema (yellow), non-enhancing tumor core (red), enhancing tumor core (blue), necrotic/cystic core (green). Figure taken from [Menze and others, 2015].

age image and a set of deformations of the average. The deformations should model the variability within the population. Most of the works in the literature focus on the *morphological* variability, namely the variations in shape of the anatomical structures. This analysis is relevant for modeling the healthy anatomical variability, as well as pathological variations that only concern the anatomy (*e.g.,* atrophy in Alzheimer's disease) [Gori et al., 2013, Gori et al., 2017, Gori et al., 2015, Ashburner and Friston, 2011, Joshi et al., 2004]. Furthermore, most of the works proposed in the literature define the deformations as diffeomorphisms, which are differentiable (smooth and continuous) bijective transformation (one-to-one) with differentiable inverse. The main reason is the anatomical plausibility of the produced deformations, since they preserve the topology and spatial organization, namely no intersection, folding or shearing may occur. Such methods include Large Deformation Diffeomorphic Metric Mapping (LDDMM) [Dupuis et al., 1998, Beg et al., 2005, Vialard et al., 2011], diffeomorphic B-splines [Rueckert et al., 2006], diffeomorphic Demons [Vercauteren et al., 2009, Lorenzi et al., 2013], Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL) [Ashburner, 2007] and symmetric image normalization (SyN) [Avants et al., 2008]. Furthermore, diffeomorphic deep learning-based methods have also recently emerged [Detlefsen et al., 2018, Dalca et al., 2019, Krebs et al., 2019, Mok and Chung, 2020]. The network outputs a vector field and the scaling-and-squaring algorithm is used [Arsigny et al., 2006] to generate a diffeomorphic deformation. Diffeomorphic deep learning methods offer faster registration at inference than classical methods at the cost of a relatively large training set.

Diffeomorphisms are not enough

However, the presence of tumors induce two sources of variation that can not be taken into account by diffeomorphisms: topological and appearance changes. The first is due to the presence of tumors, since two subjects may have a different number of tumors at different locations. Appearance differences are instead due to the infiltration of the tumors causing the edema (see Fig.2.2). This means that previous methods, mainly based on diffeomorphisms or splines deformations, can not be used to estimate a 3D atlas of glioblastoma. The deformation models and the definition itself of statistical atlas need to be revised. Indeed, since the tumor location can vary among subjects,

standard definitions of average (e.g., geometrical average, Karcher average, etc.) can not be used.

An early approach for the registration of images with a different topology is the cost function masking (CFM) [Brett et al., 2001, Stefanescu et al., 2004], where the tumor/lesion region is ignored when evaluating the cost function. This strategy has also been combined with the creation of an intermediate, cohort-specific template in [Pappas et al., 2021]. However, the CFM method falls short with large tumors/lesions [Kim et al., 2007]. To cope with that, geometric metamorphosis [Niethammer et al., 2011] adds a specific deformation to the masked area, but it works only when the lesion/tumor is present in both source and target images. Additionally, segmentation masks are required for both images.

In the context of aligning a healthy image with one showing a tumor or a lesion, it has been proposed to first make both images topologically identical and then perform the registration. A first approach has been to simulate the growth of the tumor in the healthy image with a biophysical model [Zacharaki et al., 2009, Gooya et al., 2012, Scheufele et al., 2019], and then register it onto the pathological scan. This strategy requires user initialization, and extensive computations to estimate the model parameters, which are specific to a particular kind of tumor. Although a recent fully-automatic method was introduced in [Scheufele et al., 2021], it is based on a rather simplistic biophysical growth model. In [Shen et al., 2019], authors use a similar perspective with a non-biophysical growth model computed simultaneously with the diffeomorphic warping. Despite being more generic than the previous methods, it still requires user initialization and extensive computations. An opposite strategy consists in removing the tumor to generate a healthy image. In [Liu et al., 2015a, Yang et al., 2016, Han et al., 2017, Tang et al., 2019], the pathological region is removed by synthesizing a quasi-normal image via low-rank approaches. This approach can effectively recover tumor regions, but at the same time distort or blur the healthy regions. Furthermore, it is a statistical technique that needs lesions to be homogeneously (and randomly) distributed across the population [Liu et al., 2015a], which is not the case for all kinds of lesions or tumors (*e.g.,* brain glioblastoma). With a similar perspective, inpainting techniques on brain MRI have also been proposed [Sdika and Pelletier, 2009, Almansour et al., 2021]. However, with a strong mass effect (deformation of healthy tissues surrounding the tumor), the inpainting of a tumor might not produce realistic results [Almansour et al., 2021].

Concurrently, a mathematical elegant method, called Metamorphosis [Trouvé and Younès, 2005, Holm et al., 2009, Younes, 2010], has been developed to align images with different shapes and appearances. It does not assume a one-to-one correspondence between source and target images. Metamorphosis can be seen as a relaxed version of diffeomorphisms, where small intensity variations are added to the diffeomorphic flow, therefore allowing for appearance and topological changes. This model can theoretically align any couple of images since it allows for both iconographic and morphological changes. However, in particular in medical imaging, interpretability and explicability are two important properties for a model to be trustworthy and accepted by clinicians. In our case, to fulfill these properties, we need to guarantee a proper disentanglement between shape (*i.e.* geometric) and appearance (*i.e.* intensity) transformations. This means that the appearance transformation should only account for the tumor core (topological difference) and infiltration (edema) but it should not deal with shape changes due to anatomical differences or tumor mass effect. Morphological variations should be taken into account only by the diffeomorphic deformations. This is critical for correctly interpreting the estimated alignment and for using the computed transformations in further statistical analysis, such as the atlas construction [Gori et al., 2017]. The main drawbacks of Metamorphosis are that: 1) it's computationally cumbersome [François et al., 2021] and 2) finding the parameters that perfectly disentangle shape and appearance is rather difficult.

Multi-modal data are not always available

Segmenting the tumor in the MR image can be very important to disentangle shape and appearance variations and thus build a clinically relevant and accurate 3D atlas of glioblastoma. Multi-modal segmentation models represent the state-of-the-art technique to detect brain tumors. However, it is often difficult to obtain multiple modalities in a clinical setting due to a limited number of physicians and scanners, and to limit costs and scan time. In many cases, especially for patients with pathologies or in case of emergency, only one modality is acquired. This means that there is large gap between multi-modal, high-quality *research* datasets and *uni-modal*, low-quality clinical datasets. Since our final goal is to build an atlas using clinical datasets, we have asked ourselves the following question "Can we leverage multi-modal, high-quality *research* datasets to improve the tumor segmentation in *uni-modal*, low-quality clinical datasets ?". That is to say, segmenting brain tumors using only one modality at test time, while multi-modal data are available during training.

Two main strategies have been proposed in the literature to deal with such a problem. The first one is to train a generative model to synthesize the missing modalities and then perform multi-modal segmentation. In [van Tulder and de Bruijne, 2015], the authors have shown that using a synthesized modality helps improving the accuracy of classification of brain tumors. Ben Cohen et al. [Ben-Cohen et al., 2018] generated PET images from CT scans to reduce the number of false positives in the detection of malignant lesions in livers. Generating a synthesized modality has also been shown to improve the quality of the segmentation of white matter hypointensities [Orbes-Arteaga et al., 2018]. The main drawback of this strategy is that it is computationally cumbersome, especially when many modalities are missing, since one needs to train one generative network per missing modality, in addition to a multi-modal segmentation network. Furthermore, the optimization can be difficult and the choice of the most adapted architecture quite tedious.

The second strategy consists in learning a modality-invariant feature space that encodes the multi-modal information during training, and that allows for all possible combinations of modalities during inference. Within this second strategy, Havaei et al. proposed HeMIS [Havaei et al., 2016], a model that, for each modality, trains a different feature extractor. The first two moments of the feature maps are then computed and concatenated in the latent space from which a decoder is trained to predict the segmentation map. Dorent et al. [Dorent et al., 2019], inspired by HeMIS, adapted the Multi-modal Variational Auto-Encoders (MVAE [Wu and Goodman, 2018]) architecture to the missing modality case. They proposed U-HVED where they introduced skip-connections by considering intermediate layers, before each down-sampling step, as a feature map and the modality-specific latent spaces were assumed to follow a Gaussian distribution. This network outperformed HeMIS on BraTS 2018 dataset. In [Chen et al., 2019], instead of fusing the layers by computing mean and variance, the authors learned a mapping function from the multiple feature maps to the latent space. They claimed that computing the moments to fuse the maps is not satisfactory since it makes each modality contribute equally to the final result, which is inconsistent with the fact that each modality highlights different zones. They obtained better results than HeMIS on BraTS 2015 dataset. Similarly, [Zhou et al., 2021a] designed a vector fusion procedure by extracting spatial and channel attention. This second strategy has good results only when one or two modalities are missing. However, when only one modality is available, it has worse results than a model trained on this specific modality. This kind of methods is therefore not suitable for a clinical setting.

2.3 Contributions

Previous works have focused only on the spatial distribution of the tumors by estimating a frequency distribution map onto a healthy template (e.g. MNI) [Bilello et al., 2016, Parisot et al., 2016, Roux

et al., 2019]. Here, we make a further step by proposing a new theoretical paradigm to estimate a statistical atlas of 3D glioblastoma. We plan to first divide the brain into relevant anatomical regions (e.g. lobes or more precise brain parcellation), so that the location variability of the tumors is highly reduced in each region and can be considered as "constant". Then, we propose to estimate one atlas per region using, for each atlas, only the images with tumors present in the respective region. As average image, we can use a pre-estimated healthy template, such as the MNI one [Fonov et al., 2009, Fonov et al., 2011]. In this way, for each region, we can estimate the average shape and variability of the tumor (i.e. geometric model), and the morphological variations of all surrounding anatomical structures, due to the mass-effect. Furthermore, by using a well adapted deformation model (metamorphosis, explained in the next paragraph), we can also estimate how a tumor usually infiltrates, given a certain anatomical location. This creates an ensemble of atlases, where each atlas corresponds to an individual anatomical region. From a topological point of view, it's like modeling the brain as a manifold divided into individual charts (anatomical regions/parcellations). To increase reliability and reduce the dependence on the chosen parcellation, one should also force the smoothness of the transition functions between charts. This means that the transformations estimated in one region should be related to the transformations estimated in the adjacent regions, so that there would be a smooth transition of the variations due to the mass-effect and tumor infiltration between regions.

Such an atlas could improve our understanding of the pathophysiology of glioblastoma and thus be used for surgical and chemotherapeutic planning or to better understand the association between glioblastoma and refractory epileptic seizures [Pallud et al., 2013]. Furthermore, we could also include into our analysis white matter fiber bundles obtained from tractography and segmented with the algorithm presented in Chapter 3. This could be used to test the hypothesis that tumors tend to predominantly grow along the white matter [Esmaeili et al., 2018].
This new theoretical definition of atlas requires a deformation model that can take into account not only the morphological changes but also the appearance and topological variations. As previously explained, metamorphosis is a good candidate. However, its original implementation is rather complicated and computationally cumbersome and the shape/appearance disentanglement can be tricky. To this end, we have proposed two new implementations and regularization strategies, published in [François et al., 2021, François et al., 2022, Maillard et al., 2022].

As last point, we would like our method to be generic and compatible with a clinical context, where a single MRI modality (T1w or T1ce) is usually acquired.

In the following Sections, we will describe our methodological contributions to estimate an atlas of glioblastoma using clinical 3D MR images. In particular, we will present:

1. A new framework, called **KD-Net**, to transfer knowledge from a multi-modal segmentation network (Teacher) to a mono-modal one (Student) [Hu et al., 2020]. The student network produces a precise segmentation of all tumor areas taking as input only images from a single modality. This method can thus be used in a clinical setting, leveraging the rich datasets and computational resources available in research laboratories.

2. Two implementations of the Metamorphosis image registration method based on a new semi-Lagrangian scheme [François et al., 2021]. The first uses classical numerical integration schemes [François et al., 2022] while the second employs a deep learning architecture (i.e., ResNet) [Maillard et al., 2022]. Both methods leverage the KD-Net segmentation method to correctly disentangling appearance and morphological variations.

2.3.1 KD-Net

In [Hu et al., 2020], differently from previous methods, we proposed a new strategy that consists in *distilling the knowledge* of a trained multi-modal teacher network into a uni-modal student model. It is based on the concept of *generalized knowledge distillation* [Lopez-Paz et al., 2016], which is a combination of distillation [Hinton et al., 2015] and privileged information [Vapnik and Izmailov, 2015], where one uses distillation to extract useful knowledge from the privileged information of the Teacher [Lopez-Paz et al., 2016]. This method has originally been designed for classification problems to make a small network (Student) learn from an ensemble of networks or from a large network (Teacher). It has also been applied to image segmentation in [Liu et al., 2019, Xie et al., 2018], but always to "compress" the Teacher model information, since the *same input modalities* have been used for the Teacher network and the Student network.

To the best of our knowledge, we proposed in [Hu et al., 2020] the first method that adapted the concept of generalized knowledge distillation to guide the learning of a mono-modal segmentation network using a multi-modal teacher network. It was the first work based on generalized knowledge distillation where Student and Teacher networks learned from *different* input modalities. Teacher and Student have the same architecture (i.e. same number of parameters) but the Teacher can learn from multiple input modalities (additional information), whereas the Student from only one. The proposed framework is based on two encoder-decoder networks, which have demonstrated to work well in image segmentation [Isensee et al., 2021], one for the Student and one for the Teacher. Importantly, the proposed framework is generic since it can work for any architecture of the encoders and decoders. Each encoder summarizes its input space to a latent representation that captures important information for the segmentation. Since the Teacher and the Student process different inputs but aim at extracting the same information, we make the assumption that their first layers should be different, whereas the last layers and especially the latent representations (i.e. bottleneck) should be similar. By forcing the latent space of the Student to resemble the one of the Teacher, we make the hypothesis that the Student should learn from the additional information of the Teacher. The proposed method, called KD-Net, is illustrated in Figure 2.3.

We first train the Teacher, using only the reference segmentation as target. Then, we train the Student using three different losses: the knowledge distillation term (KD loss), the dissimilarity between the latent spaces (KL loss), and the reference segmentation loss (GT loss, a combination of cross-entropy and Dice loss). Note that the weights of the Teacher are frozen during the training of the Student and the error of the Student is not back-propagated to the Teacher. The first two terms allow the Student to learn from the Teacher by using the soft prediction of the latter as target (KD loss) and by forcing the encoded information (i.e. bottleneck) of the Student to be similar to the one of the Teacher (i.e., we minimize the Kullback-Leibler (KL) divergence between the teacher and student's bottlenecks). The last GT loss makes the predicted segmentation of the Student similar to the reference segmentation.

At first, in [Hu et al., 2020], we evaluated the performance of the proposed framework on the publicly available dataset from the BraTS 2018 Challenge [Menze and others, 2015], which contains 285 patients. We compared it to the baseline nnU-Net and to two other models, U-HVED [Dorent et al., 2019] and HeMIS [Havaei et al., 2016], using only T1ce as input. Our method outperformed all other methods in the segmentation of all three tumor components. We also provided an ablation study showing that both the KL and KD loss functions improved the results with respect to the baseline model, especially for the enhanced tumor and tumor core.

However, when applying our method on a larger dataset (BraTS 2021, $N = 1251$), we found that KD-Net did not significantly improve the results compared to the baseline. We also tried two other knowledge-transfer strategies that have been proposed in the literature: Attention Transfer

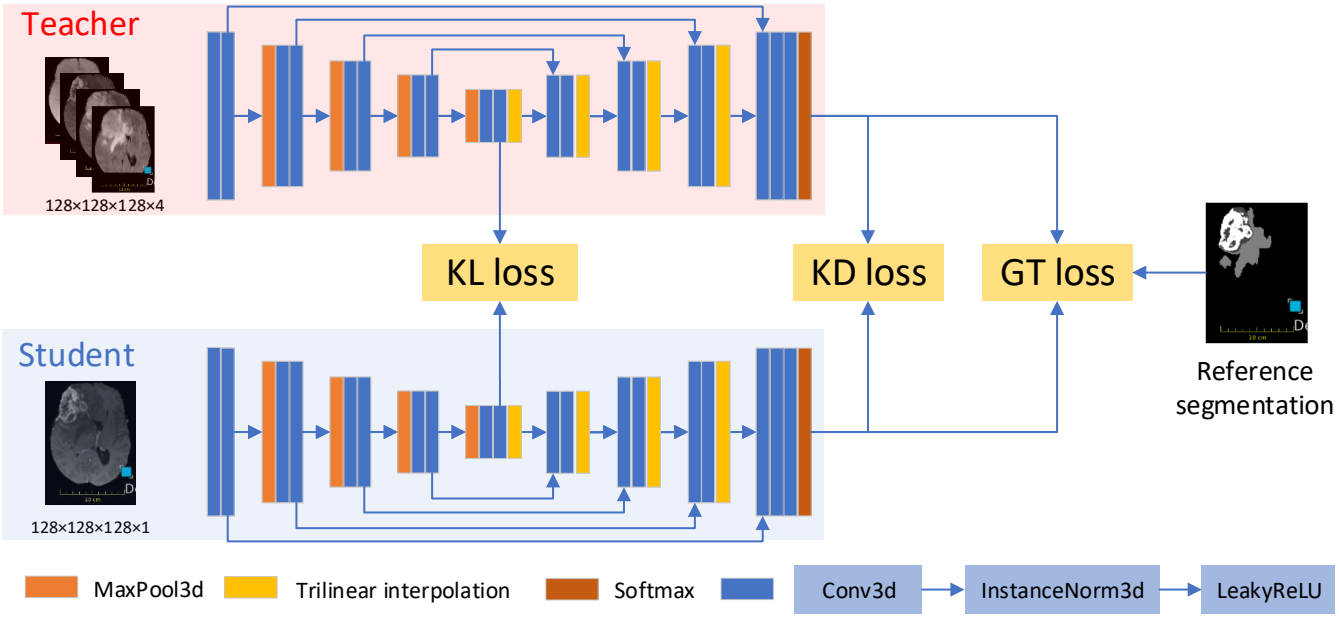


Figure 2.3: Illustration of the proposed framework. Both Teacher and Student have the same architecture adapted from nnUNet [Isensee et al., 2021]. First, the Teacher is trained using only the reference segmentation (GT loss). Then, the student network is trained using all proposed losses: KL loss, KD loss and GT loss.

(Att) [Zagoruyko and Komodakis, 2017, Qin et al., 2021, Cho and Kang, 2022] and Contrastive Distillation (CT) [Chen et al., 2022]. Results are shown in Table 2.1, where we vary the training set size while keeping fixed the test set across all experiments (178 patients). It is interesting to notice that the T1ce modality highlights the enhancing tumor and the necrotic tumor core, therefore the baseline reaches comparable results with the teacher for ET and TC but has a significantly lower WT score. Thus, we are primarily interested in improving the results for the WT label. Here, our framework KD-Net corresponds to the model trained with the KD+KL loss.

Our results indicate that the teacher-student framework is beneficial only when little data is available, see Table 2.1, and it mainly helps students to segment small structures (whole tumor or tumor parts), see Fig.2.4. In the following section, to compute the segmentation masks employed as regularization for the metamorphic image registration model, we will employ a mono-modal U-Net architecture trained on BraTS 2021 with no teacher supervision.

It's interesting to notice that BraTS is an exceptional dataset that required the collaboration of multiple international institutions and the manual annotation of more than 50 experts. Not all anatomical regions and pathologies benefited from such attention. For instance, in myocardial pathology segmentation, the only publicly available multi-modal dataset contains only 45 annotated subjects [Li et al., 2023]. For this type of applications, where only a small database of annotated images is available, the current teacher-student knowledge distillation approach should be beneficial.

More details can be found in [Maillard, 2023].

2.3. Contributions

Table 2.1: Dice score and Hausdorff distance for the models trained on six training sets, $N = 834, 417, 208$ from BraTS2021 and $N = 190, 95, 47$ from BraTS 2018. Bold indicates the best score. The symbol $*$ (respectively †) indicates that the improvement (respectively deterioration) with respect to the baseline is statistically significant ($p < 0.05$). The statistical significance of the differences with the baseline are evaluated with a paired t-test.

Training set	Model	Dice			Hausdorff		
		ET	TC	WT	ET	TC	WT
N=834	Teacher	87.61	89.95	91.44	6.05	5.63	9.27
	Baseline	86.96	89.68	77.86	6.9	6.1	13.51
	Att	**87.39**	**90.19**	77.77	**6.18**	5.49	**11.48***
	KL	86.75	89.84	77.43	7.33	5.75	11.51*
	KL + KD	86.81	90.06	77.58	6.48	**5.37***	12.89
	KD	87.33	90.1	**77.93**	6.26	5.76	12.28*
	CT	87.08	89.92	76.3†	6.63	5.57	12.59*
	CT + KD	86.63	90.11	76.37†	6.86	5.89	13.42
N=417	Teacher	86.9	89.31	90.5	7.87	7.16	11.54
	Baseline	86.02	88.8	76.77	7.13	6.42	12.79
	Att	86.47	89.42	**77.68***	7.11	6.39	**12.66**
	KL	**86.5**	**89.5**	77.29	7.03	**6.31**	13.9†
	KL + KD	86.44	89.48	75.85†	**6.72**	6.74	12.75
	KD	85.02†	87.84†	75.74†	8.21†	8.31†	13.87†
	CT	85.83	88.59	75.44†	6.8	6.37	12.98
	CT + KD	85.9	89.06	75.23†	7.6	7.3†	13.47†
N=208	Teacher	84.85	86.88	89.23	9.06	7.9	11.1
	Baseline	84.67	87.45	74.1	8.25	7.64	13.3
	Att	85.12	87.8	**77.09***	8.14	6.6*	12.76
	KL	83.6†	86.48†	74.77	8.8	7.17	**12.57***
	KL + KD	84.05	87.51	76.41*	8.78	6.74*	13.6
	KD	85.1	87.96	75.88*	7.67	7.1	13.25
	CT	84.64	87.54	76.17*	8.15	6.94*	12.96
	CT + KD	**85.25**	**88.3**	76.78*	**7.59**	**6.29***	12.78
N=190	Teacher	82.88	85.19	83.69	8.73	7.58	15.4
	Baseline	82.77	85.68	71.05	8.25	7.4	14.08
	Att	82.59	85.43	73.68*	7.84	7.31	13.48
	KL	**83.32**	86.38	**74.86***	7.52	6.69*	**12.28***
	KL + KD	82.48	86.42	74.19*	7.76	6.69*	12.95*
	KD	83.15	**87.2***	73.14*	**7.29***	**6.34***	13.19*
	CT	83.2	85.47	68.51†	9.2†	8.33†	13.52
	CT + KD	82.66	85.27	73.39*	9.05	8.15†	13.87
N=95	Teacher	79.92	83.83	82.14	12.46	9.69	14.06
	Baseline	78.8	82.6	70.0	10.89	9.27	13.8
	Att	79.53	**84.35***	72.15*	9.2*	**7.74***	**12.26***
	KL	79.37	83.75*	70.42	**8.41***	7.94*	12.46*
	KL + KD	**79.75***	82.2	**73.24***	9.57*	9.55	13.28
	KD	79.0	82.76	72.21*	9.89*	9.22	13.73
	CT	**79.75***	82.21	70.09	10.45	9.6	12.96*
	CT + KD	79.19	82.59	70.32	9.52*	9.25	13.66
N=47	Teacher	76.08	77.66	77.25	13.9	14.43	13.88
	Baseline	75.36	75.98	64.35	15.96	15.56	16.73
	Att	74.66	78.6*	66.49*	12.62*	**11.3***	**13.81***
	KL	74.22	76.72	66.88*	12.88*	11.38*	15.51*
	KL + KD	75.69	78.25*	68.09*	12.89*	11.49*	14.16*
	KD	**76.44**	**78.76***	**70.04***	13.06*	11.65*	14.65*
	CT	71.74†	74.83	64.15	15.36	14.52	15.95
	CT + KD	73.1†	76.35	67.52*	**12.24***	11.68*	14.24*

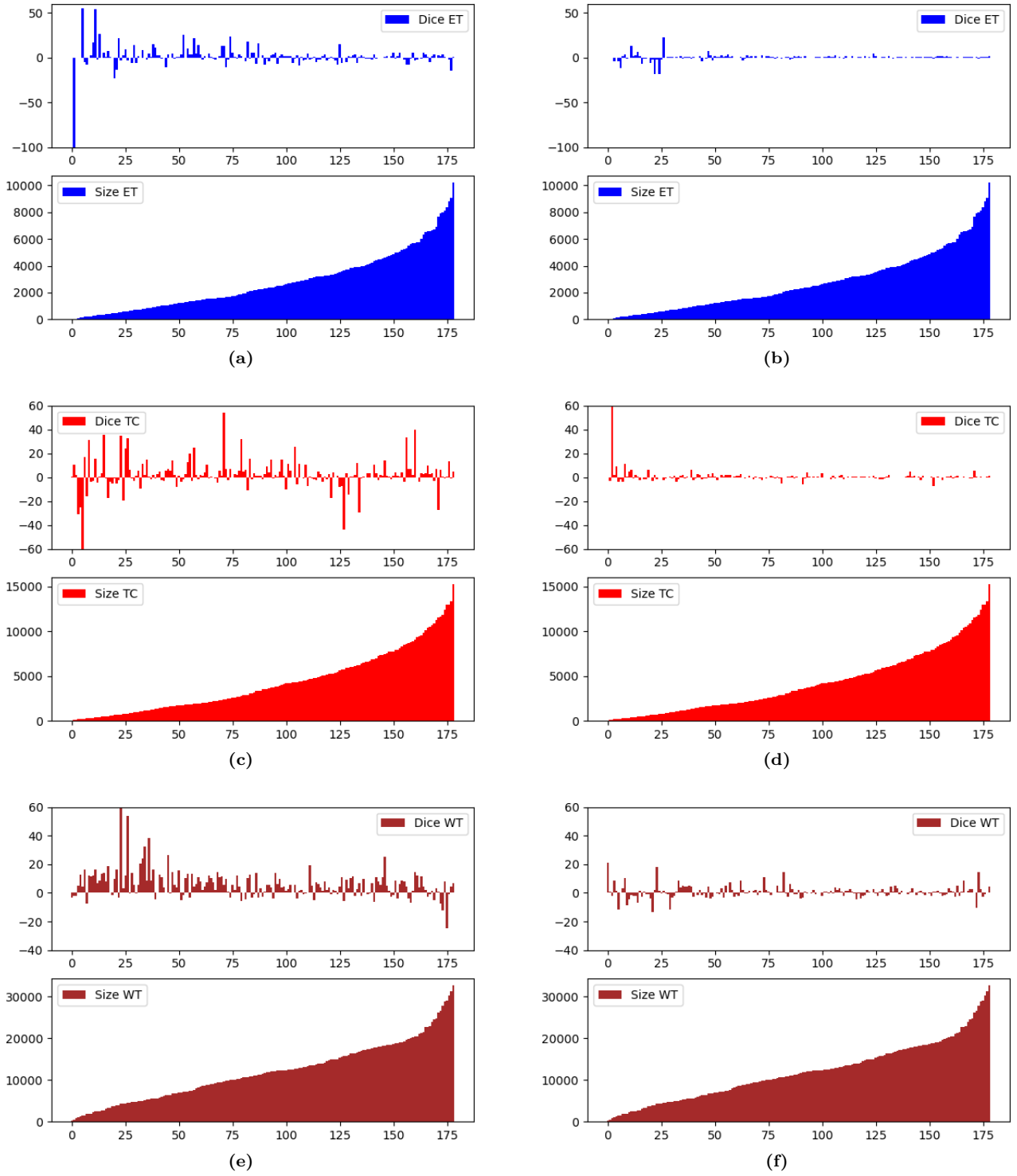


Figure 2.4: Improvement (or deterioration) of the Dice score of the three tumors labels (*ET* first row, *TC* second row and *WT* third row) with respect to the baseline when using KDNet. The results are sorted by the size of the label (i.e., size of the tumor part). For each tumor label, the improvement/deterioration is in the top row while the tumor size is in the bottom row. The x-axis represents the index of the test subject. Figures (a), (c) and (e) show the model trained with 47 subjects. Figures (b), (d) and (f) the one trained with 834 subjects. Please note that when the size is very small (few voxels), even one mislabeled voxel can drastically change the Dice score.

2.3.2 Metamorphic Image registration

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, with $d \in \{2, 3\}$. Let V be a Reproducible Kernel Hilbert space (RKHS) with kernel K of vector fields with support Ω and T times continuously differentiable, where $T \in \mathbb{N}^*$. Let I and J be the source (moving) and target (fixed) gray-scale images, both defined on Ω. We suppose that both are square-integrable and differentiable. As discussed before, when using images with varying topology, diffeomorphic deformations, such as LDDMM [Dupuis et al., 1998], may not correctly align them. Indeed, since there are different components in the images, one cannot compute a one-to-one correspondence between the images. For such cases, Metamorphosis [Trouvé and Younès, 2005] has been developed. The model joins diffeomorphic deformations with additive and infinitesimal intensity changes so that it can perfectly align images with morphological, topological, and appearance differences. Similarly to [Trouvé and Younès, 2005], the evolution of the image I at time $t \in [0, 1]$ is defined:

$$\frac{\partial I_t}{\partial t} = v_t \cdot I_t + \mu^2 z_t = -\langle \nabla I_t, v_t \rangle + \mu^2 z_t \quad \text{s.t. } I_0 = I \text{ and } I_1 = J \text{ and } \mu \in \mathbb{R} \tag{2.1}$$

where $v_t \cdot I_t$ implies that I_t is deformed by an infinitesimal, smooth vector field $v_t \in V$, producing a flow of diffeomorphisms ϕ_t in $t \in [0, 1]$ as in LDDMM, and $z_t : \Omega \to \mathbb{R}$ is the additive part corresponding to the infinitesimal intensity variation (called the residual image or momentum). The hyperparameter $\mu^2 \in \mathbb{R}^+$ balances the intensity and geometric changes.

The goal of Metamorphosis is to compute the minimal geodesic path by minimizing the energy of the transformation, $\int_0^1 ||v_t||_V^2 + \rho^2 ||z_t||_2^2 dt$, under the condition in Eq. 2.1. As shown in [Trouvé and Younès, 2005, Holm et al., 2009, Younes, 2010], by computing the Euler-Lagrange equations, one obtains the following geodesic equations for Metamorphosis:

$$\begin{cases} v_t = -\dfrac{\rho^2}{\mu^2} K_\sigma * (z_t \nabla I_t) & \text{(2.2a)} \\ \partial_t z_t = -\nabla \cdot (z_t v_t) & \text{(2.2b)} \\ \partial_t I_t = -\langle \nabla I_t, v_t \rangle + \mu^2 z_t & \text{(2.2c)} \end{cases}$$

where $\nabla \cdot (zv) = \text{div}(zv)$ is the divergence of the field v times z at each pixel and $||v_t||_V^2 = \langle K_\sigma * (z_t \nabla I_t), z_t \nabla I_t \rangle = \langle v_t, L v_t \rangle$, where, as for LDDMM, K_σ is a (usually Gaussian) kernel. By setting $\rho = \mu$ and letting $\mu \to 0$, one recovers the geodesic equations for LDDMM [Dupuis et al., 1998].

From this system of equations, we can notice that v_t is completely defined by z_t and I_t. Since I_0 is given, the only unknowns are the z_t. The momentum z_t has therefore a double role. It represents the additive intensity variation *and* it is also the parameter of the deformation. This eases the computation but at the same time it makes the disentanglement between geometry and intensity variations more difficult.

Finding a perfect alignment that verifies $I_1 = J$ is not always desirable, due to potential noise or artifacts. Thus, as it's commonly done in the literature, we cast the metamorphic registration as an inexact matching problem, minimizing the cost function:

$$E = \frac{1}{2} ||I_1 - J||_2^2 + \lambda \left[\int_0^1 ||v_t||_V^2 + \rho^2 ||z_t||_2^2 dt \right] \tag{2.3}$$

where the first term is the classical L_2 data term (please note that other data terms could be used as well) and the second term, weighted by λ, is the total energy of the transformation, which can be seen as a regularization.

Geodesic Shooting

We can first notice that, by following the geodesic paths, the energies (i.e., squared norms over time) are conserved and thus one can actually optimize only the initial norms in Eq.2.3. Furthermore, Eq.2.2b makes z_0 the only parameter of the entire system. This means that the boundary value problem can be reduced to an initial value problem, and thus one can use the shooting method to solve it [Stoer and Bulirsch, 2002]. Shooting methods were proposed and used for LDDMM [Beg et al., 2005, Vialard et al., 2011, Miller et al., 2006, Ashburner and Friston, 2011]. However, to the best of our knowledge, the only shooting method proposed in the literature for image Metamorphosis is the one in [Richardson and Younes, 2016]. It is based on a Lagrangian frame of reference and therefore it is not well suited for large images showing complicated deformations, as it could be the case when registering healthy templates to patients with large tumors. Eulerian schemes are the most natural candidate for flow integration over an image. However, they are notoriously numerically unstable and very slow since they need a large number of time steps (Courant–Friedrichs–Lewy(CFL) condition) or interpolations between grid points (e.g., Runge-Kutta methods).

As a solution, in [François et al., 2021], we proposed the first semi-Lagrangian scheme for Metamorphosis. Practically, we compute the deformation of a grid corresponding to a small displacement $\mathrm{Id} - \delta t\, v_t$, and then interpolate the values of the image I_t on the grid. This can be summarized by $I_{t+\delta t} \approx I_t \circ (\mathrm{Id} - \delta t\, v_t)$. Semi-Lagrangian schemes are stable and don't need many iterations.

Using the Pytorch framework, we proposed the first two Python-based implementations of Metamorphic image registration: an optimization-based one[2] and a learning-based one[3]. In the former, thanks to the automatic differentiation of Pytorch, we bypassed the extensive and delicate work of deriving the backward adjoint equations and proposed different optimization methods. We call this implementation Meta.

In the latter, similarly to [Krebs et al., 2019, Dalca et al., 2019], we use a UNet [Ronneberger et al., 2015], taking I and J as input, to estimate z_0. We also add an inverse-consistency term to further reinforce the diffeomorphic property of the deformation. Interestingly, although the model is primarily meant to be used in a learning context, it is possible to use it on a single pair of images and optimize its parameters by repetitively minimizing the cost. We call this method MetaMorph-G.

More information can be found in [François et al., 2021, Maillard, 2023].

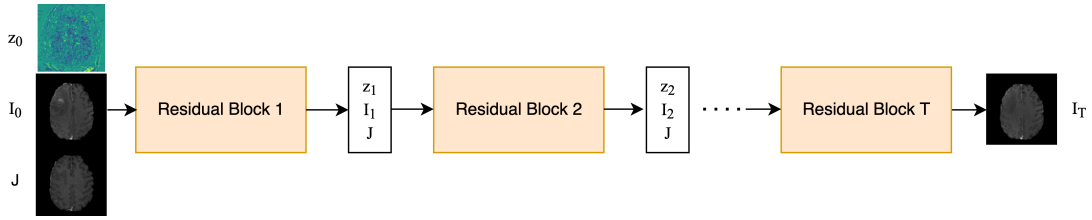
MetaMorph-R - Resnet integration

With a different perspective, we also propose to directly estimate all z_t. Inspired by [Amor et al., 2023, Rousseau et al., 2020], we propose to use a residual neural network (ResNet) to find the solution of the system of differential Equations 2.2. We take advantage of the similarity between ResNets and the numerical solutions of ODEs using Euler's method (given an initial value). Indeed, the numerical integration of Equation 2.2b, using discrete time steps t, is: $z_{t+1} = z_t - \delta\nabla \cdot (z_t v_t)$ for $t \in 0, .., T-1$, where T is the number of steps and δ is the integration step equal to $\frac{1}{T}$. By replacing the divergence operator with a neural network, we obtain a ResNet: $z_{t+1} = z_t - \delta f_{\theta_t}(z_t, I_t, J)$, where f_{θ_t} is a convolutional neural network with parameters θ_t.

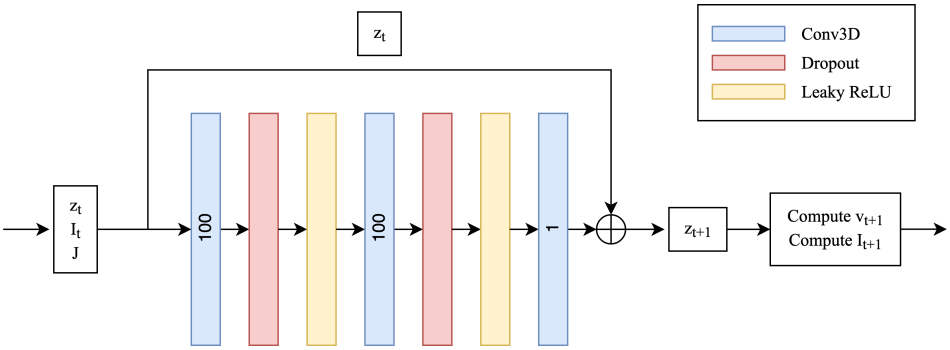
The benefit of using a neural network is that Metamorphosis can be applied in a learning context rather than just in an optimization scheme. For that reason, the source and target images are also given as input to f_{θ_t}. The network is built as a sequence of T convolutional blocks f_{θ_t}. At each time step t, z_{t+1} is computed using the previous Eq.. Subsequently, v_{t+1} is calculated directly with Eq. 2.2a and one determines I_{t+1} by applying the geometric transformation induced by v_t and

[2] https://github.com/PietroGori/Demeter_metamorphosis
[3] https://github.com/PietroGori/MetaMorph

adding the residuals z_t as in Eq. 2.2c. The architecture of the model is detailed in Fig. 2.5. The parameters of this model are optimized by minimizing the same data and regularization terms as before. Furthermore, we can also add an inverse-consistency term. We call this implementation MetaMorph-R.



(a) Overall design of the neural network.



(b) Composition of a residual block. z_t, I_t and J are concatenated before the convolution. The output of the block is added to z_t to form z_{t+1}. The numbers on the blue layers are the number of channels of the output tensor.

Figure 2.5: The residual network is composed of T residual blocks. All residual blocks have the same architecture.

Local regularization

The main problem with Metamorphosis is that it is hard to control the disentanglement between shape and appearance changes. For instance, a trivial solution would be to set the overall geometrical deformation function to the identity (no geometrical change) and the overall appearance deformation map to $J - I_0$. In that case, the L_2 distance between the deformed image and J would be 0 but it would not be a satisfactory result since homologous structures should be matched using only geometric deformations whereas appearance and topological changes (*i.e.*, new components) should be taken into account by the intensity modifications. The disentanglement can be controlled by tuning the hyper-parameters μ and λ. However, finding the right ones is a difficult task and they are different for each setting. If they are not correctly chosen, the appearance map could, for instance, modify the shape of the image, thus distorting the results and their interpretations.

To this end, we propose to restrict the intensity modifications (i.e. z) only to the regions showing a topological or appearance difference between the source and target images. Here, we do that by multiplying z by a (pre-computed) mask m of the region where the topological/appearance changes occur (a tumor for instance). Equation 2.2c then becomes: $\partial_t I_t = v_t \cdot I_t + \mu^2 m_t z_t$, with $m_0(x) = 1$ if x is a voxel in the selected region and 0 otherwise. Since the region varies along t with the source image, the mask must follow the deformation generated by the velocity fields. Consequently, the mask is not fixed but it follows the equation: $\partial_t m_t = v_t \cdot m_t$. The transformation of the image is then: $\partial_t I_t = v_t \cdot I_t + \mu^2 m_t z_t$. Using this equation and $||\sqrt{m_t} z_t||_2^2$ as regularization term for z_t, we obtain the same geodesic Eq. 2.2a and 2.2b, as shown in [François et al., 2022].

Evaluation on BraTS 2021

For evaluation, we use as before the BraTS 2021 dataset comprising MR brain images with tumors of 1251 subjects. The experiments are only conducted using the T1-w modality. We register the scans on the healthy sri24 template [Rohlfing et al., 2010]. As preprocessing, we perform histogram matching on every scan, with the template as target image, and crop the volumes to the size of $192 \times 192 \times 144$ voxels. We randomly pick 34 MR T1-w from the dataset to form a test set. We use the segmentation of the tumor as mask for local regularization. To evaluate the registration, we use the $\mathcal{L}_2$ distance between the entire source and target images and the $\mathcal{L}_2$ distance only outside the mask (i.e., w/o tumor). Additionally, we manually segment the ventricles of all 34 test images and warp them with the computed deformation to measure the overlap with the (already segmented) ventricles of the target image (i.e., sri24 template) using the Dice score.

We compare our implementations with rigid registration (Rigid), symmetric normalization (SyN) [Avants et al., 2008], and voxelmorph (VM) [Balakrishnan et al., 2019] to show that one-to-one methods are not adapted in this context. Additionally, we compare them with their cost masking versions (Rigid-CFM, SyN-CFM, VM-CFM). For VM-CFM, since it is a learning-based method, source images are masked with the tumor segmentation during both training and test.

From Table 2.2, we can notice that our implementations of Metamorphosis, Meta and MetaMorph-R, outperform all other methods both in terms of Dice and $\mathcal{L}_2$ distances.

Fig. 2.6 shows the visual comparison for three different subjects (rows) of MetaMorph, Voxelmorph, and SyN-CFM. On all three patients, our model better aligns the ventricles and it generates rather realistic healthy images, although some edges of the tumor mask can be spotted due to a sudden intensity change between healthy and masked regions.

Furthermore, as demonstrated in Fig. 2.7, the introduction of the local regularization (i.e., segmentation mask m) makes the disentanglement between shape and appearance transformations easier. Indeed, for different values of the hyper-parameters λ and μ, the results without masking are more variable. For a low value of μ, the intensity transformation is non-existent whereas the shape deformation is too small to properly align the images. For higher values, the appearance transformation modifies the shape of the images, namely it changes the topology of healthy tissues. Similar behavior occurs when changing λ. On the other hand, the masked method obtains better and less varying results for the various hyper-parameter combinations.

More results can be found in [Maillard et al., 2022, Maillard, 2023].

Limitations

The main limitations of our work are the computational time and memory. Indeed, on the one hand the optimization based method needs around 15-20 minutes per registration (depending on the size of the images) while the deep learning based method needs several hours for training but less than a second at inference time. On the other hand, the optimization based method needs less than 8GB of memory (thus fitting most of the GPU cards) while the deep learning based method requires 30 GB of VRAM for training, which only a few GPUs verify. Thus, the choice between the two methods mainly depends on the computational resources (and time) at disposal.

Another limitation is the fact that the momentum z_t has a double role: it models the additive intensity variation and it parameterizes the entire transformation. This simplifies modeling and computation but it makes the analysis of the results more complex. Furthermore, using a mask m to restrict the intensity variation makes sense but why should we restrict the spatial area of the deformation parameters? Even if we did not observe a degradation in performance of the diffeomorphic matching when using a mask, we believe that a model that separates the intensity

Method	Dice	$\mathcal{L}_2$	$\mathcal{L}_2$ w/o tumor	$\lvert J_\phi\rvert < 0$
Rigid	43.9(12.3)	8.1(1.0)	7.4(0.9)	**0(0)**
SyN	55.7(12.9)	5.5(1.0)	4.9(0.8)	**0(0)**
VM	65.5(7.2)	4.4(0.8)	*3.3(0.6)*	4427(1821)
Rigid-CFM	43.9(12.3)	8.1(1.1)	7.5(0.9)	**0(0)**
SyN-CFM	56.4(12.8)	5.5(1.0)	4.9(0.9)	**0(0)**
VM-CFM	61.5(8.1)	4.9(1.0)	3.4(0.6)	3423 1733)
Meta [François et al., 2022]	**70.3(5.3)**	*4.2(0.6)*	3.9(0.6)	**0 (0)**
MetaMorph-G [Maillard, 2023]	65.32(6.25)	5.1(0.7)	4.9(0.7)	283(666)
MetaMorph-R [Maillard et al., 2022]	*69.2(6.7)*	**3.4(0.57)**	**3.1(0.55)**	366(594)

Table 2.2: Results on BraTS 2021. Classical registration methods are compared with their cost function masking version and our methods. The "Dice" column provides the Dice scores between the segmentation of the ventricles in both images after registration. $\mathcal{L}_2$ is the $\mathcal{L}_2$ distance between the deformed and target image. $\mathcal{L}_2$ w/o tumor excludes the tumor region when computing the $\mathcal{L}_2$ distance. $|J_\phi| < 0$ measures the number of folds in the image. Bold and italic numbers indicate the first and second best scores. The $\mathcal{L}_2$ scores are divided by 10^4 for better readability.

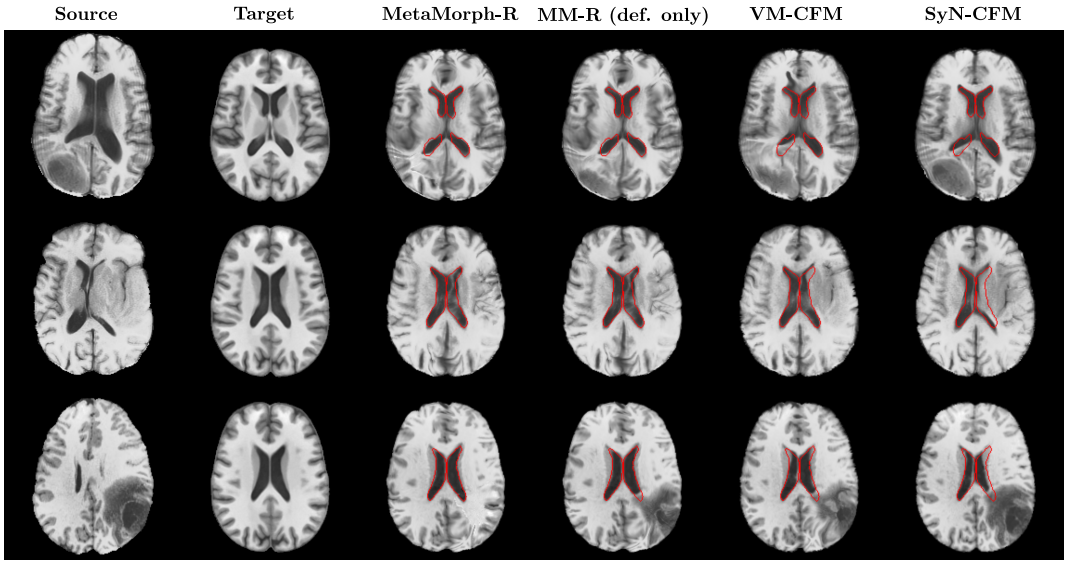


Figure 2.6: Results on three test subjects (each row is a subject) from the BraTS 2021 dataset using: MetaMorph-R (total transformation and deformation only), Voxelmorph (VM), and symmetric normalization with cost function masking (SyN-CFM). Red lines delineate the ventricles of the target image superposed on the deformed source image.

variation from the deformation parameters would be more interpretable. Indeed, the values of z are not easy to interpret due to its dual effect.

2.4 Conclusions and Perspectives

In this Chapter, we have presented new methods for 1- transferring knowledge from a multi-modal segmentation network to a uni-modal one and 2- registering images with morphological, appearance and topological differences. These algorithms represent two important steps for our final goal, which is the estimate of a 3D atlas of glioblastoma using uni-modal clinical images.

Our main conclusions are that knowledge distillation, and more in general existing techniques for transferring knowledge, are useful only in a low data regime. However, as soon as the dataset

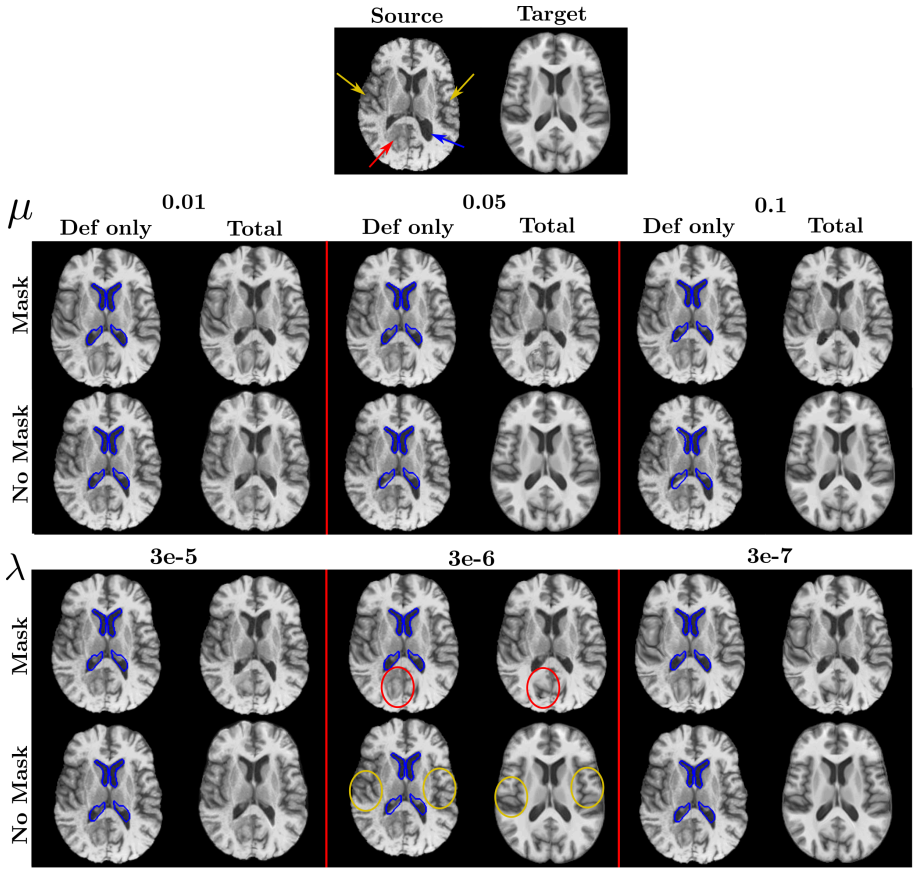


Figure 2.7: Deformation and total transformation for the masked and non-masked versions of MetaMorph-R with 3 values of μ (first two rows) and λ (last two rows). When μ is varying, λ = 3e-6 and when λ is varying $\mu = 0.04$. Red arrow indicates the tumor, blue arrow shows the ventricle that is incorrectly aligned without masking (manual segmentation in blue), yellow arrows show healthy tissues that are incorrectly modified by the appearance transformation without masking.

is big enough, directly learning a uni-modal segmentation network is as accurate as leveraging the multi-modal Teacher knowledge, but more efficient.

About the metamorphic image registration methods, we showed that our new implementations, based on a semi-Lagrangian scheme, are more stable and accurate than current image registration methods, when aligning images with different topology and appearance. Furthermore, the use of a regularization mask eases the disentanglement between morphological and appearance changes, thus increasing the clinical usefulness of our method and making it well adapted for atlas construction. Finally, the method is not specific to a certain imaging modality or anatomical location. We used it on T1 and T1ce MRI scans of the brain, but it could also be used with other modalities, such as CT or PET, and with pathological images of other anatomical areas, such as the abdomen [La Barbera et al., 2022, La Barbera et al., 2021].

Knowledge Distillation in the large data regime

Our results show that current methods for distilling and transferring knowledge between student and teacher networks are not useful in the large data regime. We believe that this is probably due to the fact that current Teacher-Student methods do not actually transfer or distill knowledge, but they "simply" help the student to focus on the information present in its input modality. If the information is not present in the imaging modality, the student can not learn it. This kind of

methods thus helps the student only when the number of training data is too low for the student to correctly learns alone.

To overcome that, we believe that the student should also be helped by generating the missing knowledge. However, this should not happen in the pixel space, which would require large data-sets and powerful generative models, but probably more in the feature space. Indeed, we would not need to generate the entire image of the missing modality, but just the missing information useful to the student. This would avoid learning how to generate redundant or irrelevant information in the pixel space. Recent self-supervised methods, like [Assran et al., 2023] , have employed a similar strategy, showing impressive results.

Metamorphic image registration with multiple modalities

One limitation of our metamorphic image registration method is that it does not apply to multi-modal data. Indeed, it has been conceived for a clinical context, where a single modality is usually available. However, the framework could be extended to multimodal data by considering the image I_t as a function from $\Omega \subset \mathbb{R}^d$ to $\mathbb{R}^C$, with C being the number of modalities. Furthermore, appearance changes should be specific to each modality (*i.e.* different z for each modality) but the shape deformation should be the same across modalities (*i.e.* same v for every modality). Thus, one could set $z_t : \Omega \to \mathbb{R}^C$, and a single velocity-fields v_t, as before. With this formulation, we would have that, for each modality c, the deformation of image I_t^c is induced by the common velocity field v_t and the intensity transformation is specific to that modality with z_t^c. A first (simple) model has been proposed in the PhD thesis of M. Maillard [Maillard, 2023].

Leveraging bio-physical tumor growth models

The metamorphic image registration methods proposed in this Chapter are not based on a bio-physical tumor growth model, to mimic the growth of a tumor into a healthy image, but on a rather simplistic model. Indeed, the evolution of the tumor mask, used as regularization term to disentangle morphological from appearance variations, is not estimated via a bio-physical model but by simply registering the segmentation mask towards a (infinitesimal) small ball positioned in the center of the tumor (which should represent the starting point of the tumor). This is a simple and effective solution which produces a good (final) alignment but its evolution (flow between $t = 0$ to $t = 1$) is not biologically relevant and thus clinically interpretable. Theoretically, one should consider how the tumor evolves in the brain, considering the different tissues (e.g., white and gray matter), brain location and tumor-specific parameters. This could be done using bio-physical models of tumor growth, that usually comprise a system of complex Partial Differential Equations (PDE) [Gooya et al., 2012, Scheufele et al., 2019, Mang et al., 2020]. However, their solution is slow, computationally heavy, and needs several imaging modalities to estimate all the tumor-specific parameters. Furthermore, some methods, like GLISTR [Gooya et al., 2012], also need manual inputs from the user (e.g., starting point of the tumor). Adding a bio-physical growth model into our framework, to drive the evolution of the mask, for instance, could thus improve the clinical relevance and interpretability. However, it would 1) require more computational resources, 2) to have access to the tissue segmentation of the whole brain and, possibly, 3) to several imaging modalities.

An interesting research direction could be combining ensemble inversion schemes, as in [Subramanian et al., 2023], with deep learning frameworks, as in [Pati et al., 2021], to learn how to (quickly) infer the most important tumor parameters from a single imaging modality. These parameters could then be used to estimate a biologically-relevant evolution of the regularization mask.

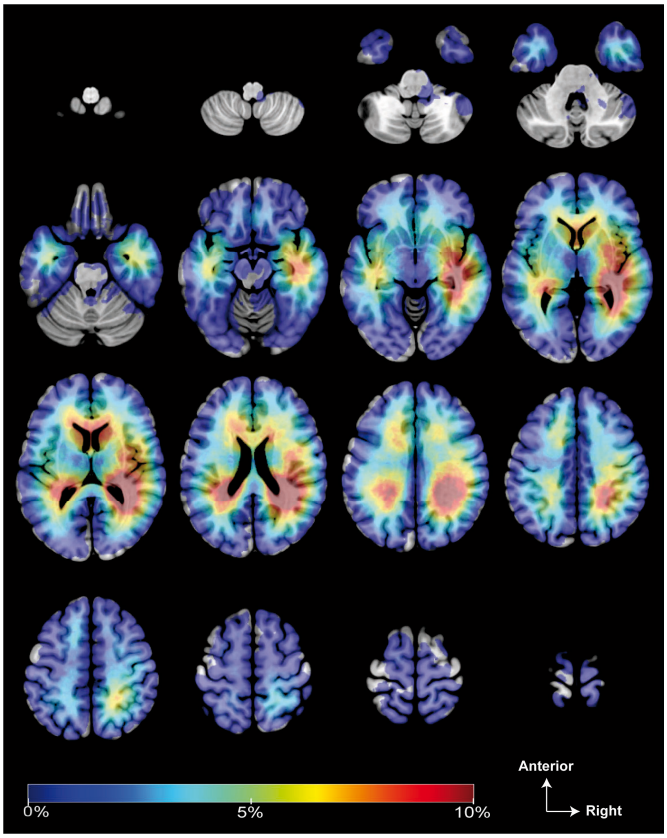


Figure 2.8: Frequentist atlas of glioblastoma showing the tumor frequency at each brain location using 392 training subjects. Figure from [Roux et al., 2019].

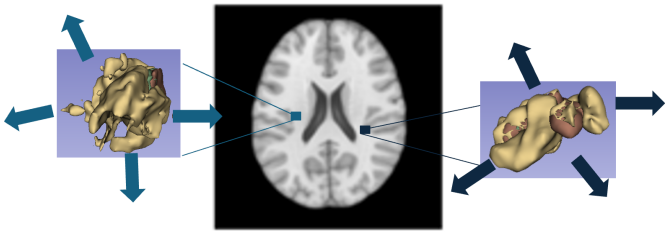


Figure 2.9: Example of a 3D atlas of glioblastoma. At each voxel or anatomical area, one could analyze the average and main variations of the tumor parts (core, oedema, etc.)

Towards a probabilistic 3D atlas

In [Roux et al., 2019], we presented a frequentist atlas (see Fig.2.8) showing the tumor frequency on the (healthy) MNI template. This atlas was obtained by first aligning 392 T1-MR volumes of subjects with glioblastoma onto the MNI template and then computing, at each voxel, the probability of tumor presence. We used the cost-function masking method of [Andersen et al., 2010] and it would be interesting to re-estimate it using our metamorphic image registration methods. This should provide more accurate and anatomically relevant alignments avoiding, for instance, an estimated presence of tumor in the ventricles, which is clinically impossible.
Furthermore, as previously explained, our final goal would be to estimate an actual 3D statistical atlas, modeled as a combination of local atlases. Preliminary results, estimating one atlas per lobe, can be seen in Fig.2.9, where we computed the average and main variations of the different tumor parts (core and edema) at each voxel.

Chapter 3

Brain white matter tractogram analysis

This chapter has been published in [Mercier et al., 2018, Delmonte et al., 2019, Feydy et al., 2019a] and is based on: 1) the PhD thesis of C. Mercier, co-directed with with I. Bloch, J.M. Thiery and D. Rohmer, 2) the Master theses of A. Del Monte and A. Di Girolamo, and 3) the post-doc of P. Roussillon.

Contents

3.1 Context

Tractography from diffusion MRI is currently the only technique able to non-invasively explore the white matter architecture of the brain. It results in a tractogram, which is a set of 3D polylines (*i.e.,* lists of ordered 3D points), usually called streamlines, which are estimates of the trajectories of large groups of nerves (axons). Indeed, current diffusion MR machines have a spatial resolution in the millimeter (mm) scale, and therefore it's not possible to perfectly reconstruct each axon, whose diameter is typically in the micrometer (μm) scale. This means that reconstructed streamlines might approximate more than 10^5 axons in each voxel [Saliani et al., 2017]. Nevertheless, even with such low reconstruction resolution, tractography has proven to be an invaluable tool for clinicians and researchers. It is nowadays used on a daily basis by neurosurgeons for pre-operative planning and during surgical operations [Jeurissen et al., 2019]. It also offers important information for studying pathological processes in neurological and psychiatric diseases [Ciccarelli et al., 2008] and aging [Davis et al., 2009].

When visualized, a whole-brain tractogram, namely the estimate of all the nerve streamlines in the white matter of the brain, might seem a highly intricate and complex wiring system, where it's difficult to recognize specific bundles and a well-structured organization (see Fig.3.1 for an exemple).

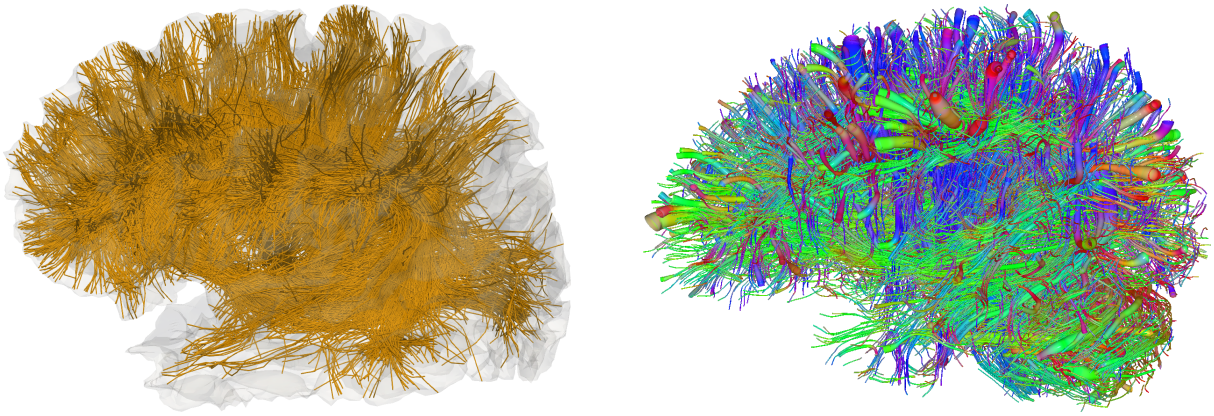


Figure 3.1: Example of whole-brain tractogram visualization using streamlines (left) or tubes (right).

However, thanks to numerous postmortem studies [Dejerine, 1895, Klingler and Gloor, 1960, Glickstein, 2006, Nieuwenhuys et al., 2015, Yendiki et al., 2022], the white matter architecture of the brain has been deciphered into a set of biologically plausible pathways (also called fasciculi or tracts) [Schmahmann and Pandya, 2006, Schmahmann and Pandya, 2007, Catani et al., 2002, Wakana et al., 2004], that have helped understanding the functional expression of the cerebral activity. Tractogram segmentation can thus be defined as the subdivision of whole-brain tractograms into anatomically relevant and reproducible tracts with well-known structural and diffusion properties.

3.2 Challenges

Recent tractography methods may produce whole-brain tractograms composed of millions of streamlines [Tournier et al., 2011]. This can complicate their visualization and interpretation, thus limiting the aforementioned clinical applications. Furthermore, the considerable number of streamlines can make computationally intractable processes such as segmentation [Delmonte et al., 2019], nonlinear registration [Gori et al., 2018] or atlas construction [Gori et al., 2016], which are important for research purposes. The presence of spurious streamlines (outliers) and the high inter-subject variability are two other challenges that can make the analysis of tractograms even more complicated. Lastly, the anatomical definition of a tract, as documented in the postmortem studies [Ebeling and von Cramon, 1992, Sarubbo et al., 2013], is usually qualitative, since it is based on spatial relationships, such as "anterior to" or "close to", that are difficult to model in a quantitative and accurate way.

Tractograms are redundant

A whole-brain tractogram is usually redundant since many streamlines might have a similar trajectory and connectivity. For this reason, several authors have proposed new geometric representations and visualization techniques to simplify tractograms. One of the most popular approaches consists in grouping similar streamlines into clusters [Gori et al., 2016, Garyfallidis et al., 2012, Guevara et al., 2011, Maddah et al., 2007, Zhang et al., 2008], which are then approximated with one representative streamline, usually called *prototype* [Garyfallidis et al., 2012, Guevara et al., 2011], that represents the average trajectory of the streamlines of the clusters. Prototypes can be computed as the mode [Zvitia et al., 2010] or mean[Garyfallidis et al., 2012] of the streamlines, if there is a point-correspondence, or according to a streamline dissimilarity measure [O'Donnell et al., 2009, Guevara

et al., 2011]. Representative streamlines are mainly used to ease the interpretation and visualisation of a bundle and to reduce the memory footprint and computational time for segmentation, shape analysis and registration.

Other authors have also proposed to represent the spatial extent of the clusters using an encompassing geometry (i.e., isosurfaces) [Maddah et al., 2007]. These methods are usually controlled by one parameter, e.g. a threshold [Zhang et al., 2003], thus presenting only one level of resolution at a time and some important spatial information, such as the number of streamlines or the spatial extent (i.e. the volume) of the cluster, might be lost in the process. Furthermore, isosurface representations can only be used for tubular-shaped bundles that can be modeled as convex envelopes. Other bundles, such as the corpus callosum and the rostral part of the corticospinal tract, have a different topology and they are defined as sheet-like bundles. In [Maddah et al., 2011], authors proposed to represent those bundles as 3D surface meshes whereas in [Yushkevich et al., 2008] it was suggested to use deformable medial models (cm-reps). In both cases, the medial surface representations are employed only for visualisation and clustering and to provide statistics about diffusion coefficients. A different representation, which can be employed for any kind of bundle, is the tract probability map [Hua et al., 2008, Bürgel et al., 2006, Wassermann et al., 2010]. It indicates the probability of a voxel to belong to a given bundle. This method is very concise but it is not based on a geometrical primitive and it has been used for visualisation, interpretation and clustering. A last example is the sparse representation based on the matching pursuit algorithm for currents presented in [Durrleman et al., 2011]. In the framework of currents [Vaillant and Glaunès, 2005], a bundle is considered as a single mathematical object composed of disconnected oriented points which model the local orientation of the streamlines. The approximation presented in [Durrleman et al., 2011] represents a bundle with a sparse set of oriented points. This representation is very concise but it has the drawback to accurately approximate only the areas of the bundle characterized by a high density of streamlines, like the central mass of the bundle. Thus, the small fascicles may not be well approximated.

Tractograms are difficult to segment

The anatomical definition of a white matter tract is vague and not always consistent across studies and clinicians [Bullock et al., 2022]. This is why the segmentation of whole brain tractograms is a difficult and not well posed problem.

The most common technique for identifying a tract is the virtual dissection technique, where an expert manually delineates Regions of Interest (ROIs) and select (or exclude) the streamlines that pass through them [Wakana et al., 2007]. This method is tedious, time-consuming and not easily reproducible, especially for tracts with convoluted trajectories [Zhang et al., 2010].

A second class of methods consists of machine or deep learning strategies that use either the geometry of the streamlines [Zhang et al., 2019, Dumais et al., 2023] or their voxel-wise principal directions [Wasserthal et al., 2018]. These methods require large, annotated data-sets, are hardly explainable, and are prone to data biases, such as the site-effect [Bayer et al., 2022], due to specific protocols or scanners.

A third technique is based on the transfer of manually segmented ROIs from one (or multiple) training images or atlases to test subjects via image-based non-linear deformations [Zhang et al., 2010]. The resulting segmentation highly depends on the quality of the registration, which might not be accurate when training and test images do not share the same anatomical topology (e.g. due to a tumor or illness). To this end, streamline-based registration methods have been shown to be more robust to topological differences (e.g., presence of tumors) [Guevara et al., 2012, Garyfallidis et al., 2018, Sharmin et al., 2018, O'Donnell et al., 2016]. However, these methods usually need a pre-

processing steps where streamlines are clustered to simplify the tractogram so that the registration becomes computationally feasible [Gori et al., 2016, Garyfallidis et al., 2012, Guevara et al., 2011, Maddah et al., 2007, Siless et al., 2018, O'Donnell and Westin, 2007, Chen et al., 2023]. Furthermore, the resulting segmentation highly depends on several user-tuned hyperparameters (e.g. size or number of clusters, kernel size).

All previous methods do not take into account the intrinsic vagueness of the definitions of the tracts. Indeed, differently from other anatomical structures, such as bones or organs, it is almost impossible to clearly delineate the boundaries and contours of white matter tracts.

That is why, with a different perspective, few recent methods have tried to directly model the "qualitative" anatomical spatial relations defining the white matter tracts. For instance, in [Wassermann et al., 2016], authors proposed a query language (WMQL) to mathematically model simple spatial relationships and logical operations used to define the white matter tracts of the brain. This method is fast and easy to use but it is based on simple binary relations and bounding boxes that can not correctly segment small and convoluted tracts and can produce tracts of different size and shapes across subjects.

3.3 Contributions

Here, we present three original contributions to tackle the aforementioned challenges:

1. A new parsimonious and multi-resolution geometric representation for tractograms, called Neural Meta Tracts [Mercier et al., 2018]
2. Two fast and scalable methods to automatically segment tractograms, one based on symbolic AI [Delmonte et al., 2019] and one on optimal transport [Feydy et al., 2019a].

3.3.1 Neural Meta Tracts

In [Mercier et al., 2018], taking inspiration from error-driven surface mesh simplification [Garland and Heckbert, 1997], we propose a progressive merging strategy for grouping streamlines into generalized cylinders (see Fig.3.2 for a visual explanation of the entire pipeline). The proposed method reduces the redundancy of the tractogram, producing a multi-resolution structure, which is organized into a nested hierarchy of levels of detail. Every fusion of streamlines (or cylinders) represents a new level of resolution. Once the entire multi-resolution representation is computed, it is possible for the user to navigate through different levels of detail in a continuous fashion and in real-time, while maintaining the overall structure of the original tractogram. Furthermore, we also propose an efficient implementation based on a Delaunay tetrahedralization which makes it possible to use our method on large tractograms containing millions of streamlines. In this way, we can determine adequate candidates for merging in a very efficient way and using any distance/similarity measure between streamlines. Differently from previous methods, we do not focus on single resolution approximations (i.e. clustering and prototypes), but propose a multi-resolution representation based on progressive merging, that preserves the overall structure of the tractogram and whose continuous levels of resolution can be traversed in real-time (see a comparison with QuickBundles [Garyfallidis et al., 2012], a well-known approximation algorithm for brain white matter tractograms based on prototypes, in Fig.3.3). From a technical point of view, our two main contributions are: 1) a multi-resolution representation for tractograms based on a progressive decimation algorithm; and 2) a combinatorial strategy based on a Delaunay tetrahedralization to make it computationally tractable.

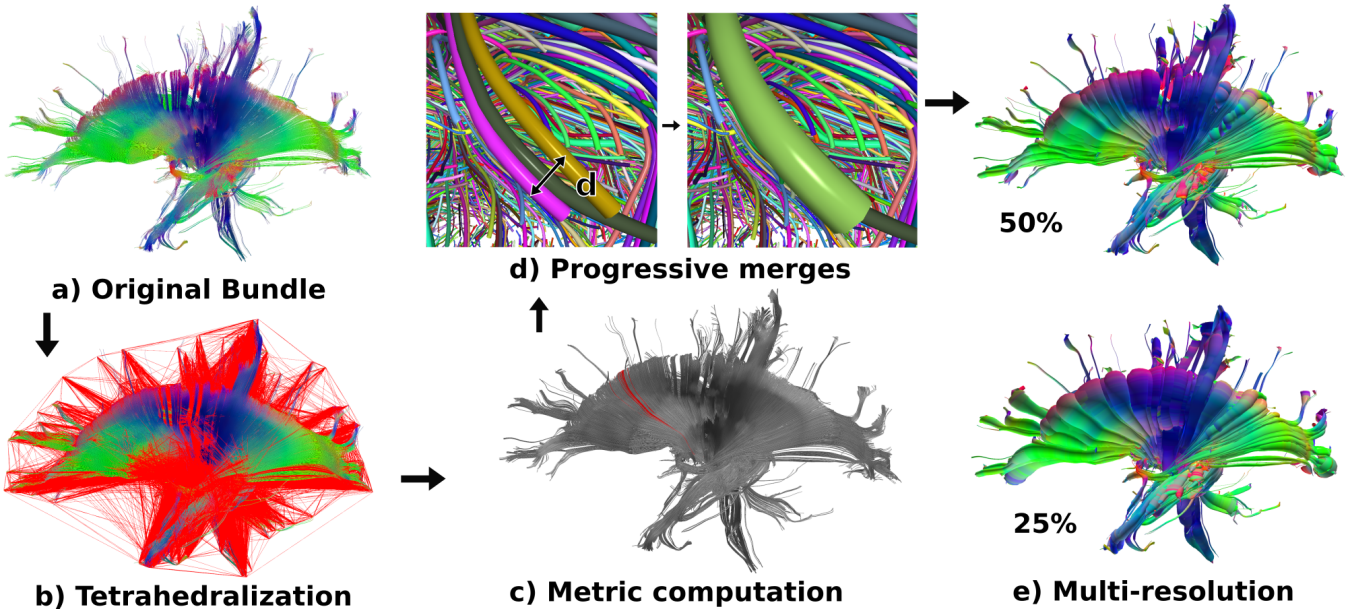


Figure 3.2: Multi-resolution pipeline: a) original bundle tractogram, here the thalamocortical one, b) connections created by the Delaunay tetrahedralization based on the extremities of the streamlines, c) the similarity is computed for each streamline, taking into account only the neighbors found in step b), d) couples of closest streamlines are progressively merged into generalized cylinders, e) the final multi-resolution representation makes it possible to navigate through the different levels of detail in real-time. The percentage refers to the fraction of employed generalized cylinders compared to the original number of streamlines. Color code depends on the orientation of the streamline: red for left-right, blue for inferior-superior and green for anteroposterior.

Visualizing groups of similar streamlines as single generalized cylinders and being able to easily change the level of resolution may be very useful for clinicians. For instance, it can help neurosurgeons better understand the organization of the white matter and identify relevant anatomical tracts (i.e., manual segmentation) which should not be severed during the operation, thus reducing post-operative complications and improving the clinical outcome.

In the next section, we will make a step forward proposing an interactive segmentation method that combines the proposed multi-resolution geometric representation with a symbolic AI method, that models the qualitative and vague anatomical definitions of the tracts. This will result in an explainable and trustworthy segmentation method where all streamlines will have a (normalized) membership score for each anatomical tract summarizing all qualitative descriptions.

3.3.2 White Matter Segmentation

Here, we present two methods to segment whole-brain tractograms, one based on symbolic AI [Delmonte et al., 2019] and one based on optimal transport [Feydy et al., 2019a]. Both methods take advantage of the anatomical knowledge about white matter tracts.

The first symbolic AI method leverages logic and fuzzy sets to directly model the anatomical relations defining a tract. This is probably the "rawest" form of clinical knowledge since it is what medical doctors learn and it should not be biased by inter-subject variability. However, anatomical definitions are usually vague and qualitative and not always complete, namely it might be hard in some anatomical regions to precisely describe the tract with spatial relations.

Another way of modeling clinical knowledge is the use of manual segmentation. Manual masks do not only reflect the clinical knowledge from the books (as modeled by the previous method)

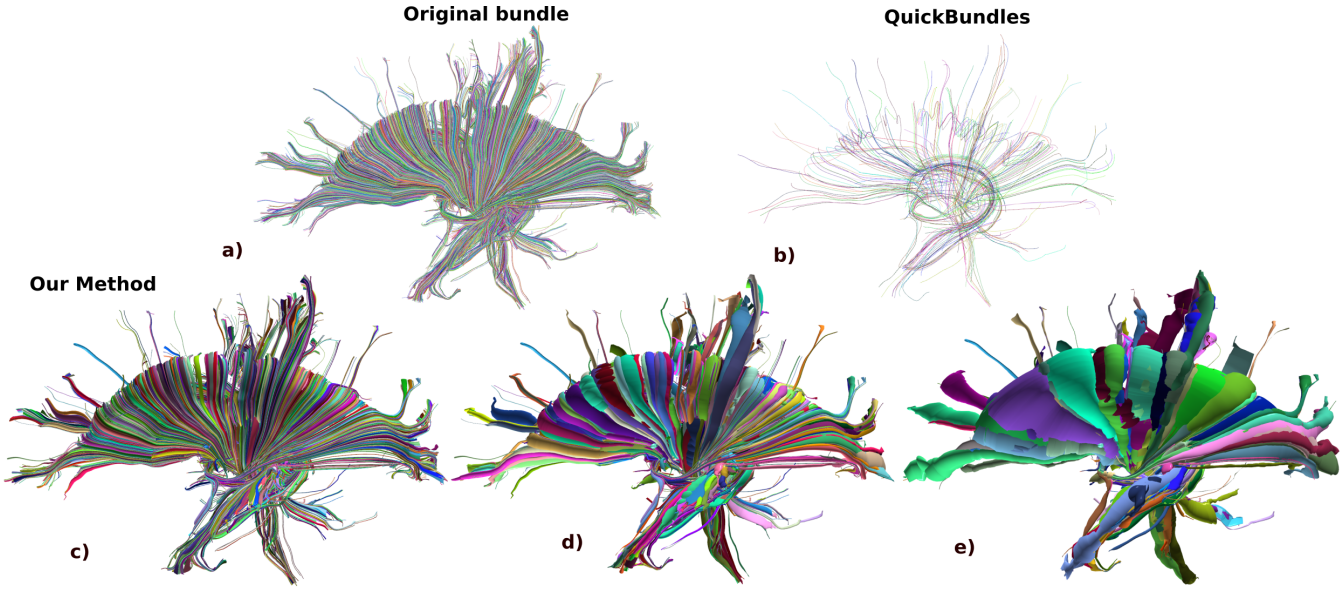


Figure 3.3: a) *Thalamocortical* bundle with 19,782 streamlines, b) reduced to 275 prototypes with QuickBundles [Garyfallidis et al., 2012] (Threshold=7mm), c), d) and e) reduced to 6925, 1422 and 275 cylinders with our method.

but they also contain the rater's knowledge and experience, which might "fill the gap" when the definitions are not clear or complete. This might thus be beneficial, when the rater is experienced, since it might add relevant information not clearly expressed in the books, but at the same time it might also be detrimental if the rater's experience is low or if two raters disagree. The inter-rater variability , as the intra-rater variability, can thus introduce a bias (and errors) in the manual segmentation.

Nevertheless, manual segmentation is probably the most used way of modeling clinical knowledge (and experience). The second method, based on optimal transport, leverages expert knowledge, in the form of a labeled atlas, that is then mapped to a subject tractogram. In this way, one obtains the probability of belonging to a certain tract, segmented in the atlas, for each streamline of the subject tractogram.

An important similarity between the two methods is the fact that they both produce a soft segmentation and not an hard one, namely they produce for each streamline a membership score or probability to belonging to a certain tract. This is quite important since it means that streamlines that are in between two tracts or at the border will have a high probability score for both tracts and they will not be assigned to just one of them.

Furthermore, both methods are fast, can be used with full whole-brain tractograms and are based on few and easy-to-tune hyper-parameters, making them reliable and trustworthy to clinicians.

Symbolic AI - Fuzzy set

As first segmentation method, we propose to directly model qualitative anatomical definitions, as in WMQL [Wassermann et al., 2016], but within the richer framework of first order modal logic. We also propose to associate it with fuzzy semantics [Bloch, 2005], in order to cope with the intrinsic imprecision of anatomical descriptions and of spatial relations. Furthermore, fuzzy representations [Bloch and Ralescu, 2023] inherently solve the semantic gap, establishing links between abstract clinical concepts/definitions and spatial image information. The general idea is to define for each point in the space the degree to which it satisfies a given relation with respect to a reference object

(i.e. an anatomical structure).

Leveraging the efficient approach based on fuzzy dilation proposed in [Bloch, 1999], we can model the directions anterior, posterior, superior, inferior, right and left as well as the relations "lateral" and "medial", which are commonly used in the neuro-anatomical literature. For these two last relations, we use as reference the mid-sagittal plane which is automatically detected using the method described in [Tuzikov et al., 2003]. Furthermore, a white matter tract is usually described as a logic combination of several relations, using operators such as *AND* and *OR*. The proposed fuzzy models of spatial relations are combined using fuzzy *AND* (using t-norms) and fuzzy *OR* (using t-conorms). Here, we use the minimum for *AND* and the maximum for *OR*, computed voxel-wise. A membership value μ^* describing the degree of satisfaction of the combined relations is computed for every point P in the space (i.e. every voxel). Then, a fuzzy score *FS* is assigned to each streamline of the tractogram by computing the weighted average of the membership values μ^* of the voxels the streamline passes through. Weights are computed as the proportion of the length of the streamline within each voxel.

In addition to the relative directions, we also model another common anatomical definition about the location of the tract terminations (e.g. "streamlines terminate in temporal lobe"). Let f be one of the endpoints of a streamline and M the region of the ending area, we define the degree of rightness as: $EP = \min_{m \in M} \exp -\frac{||f-m||_2^2}{\lambda^2}$, where λ is a fixed parameter. When the definition involves only one region, f is the endpoint closer to M. Otherwise, when using two ending areas, the streamline orientation is the one minimizing the sum of the distances between the ending points and the regions (each extremity being linked to a different region).

Eventually, all qualitative relations describing a tract are combined together in a conjunctive way, $ACS = FS \times EP$, resulting into a single, quantitative membership score called "Anatomical Coherence Score" (*ACS*), which is assigned to every streamline of the tractogram.

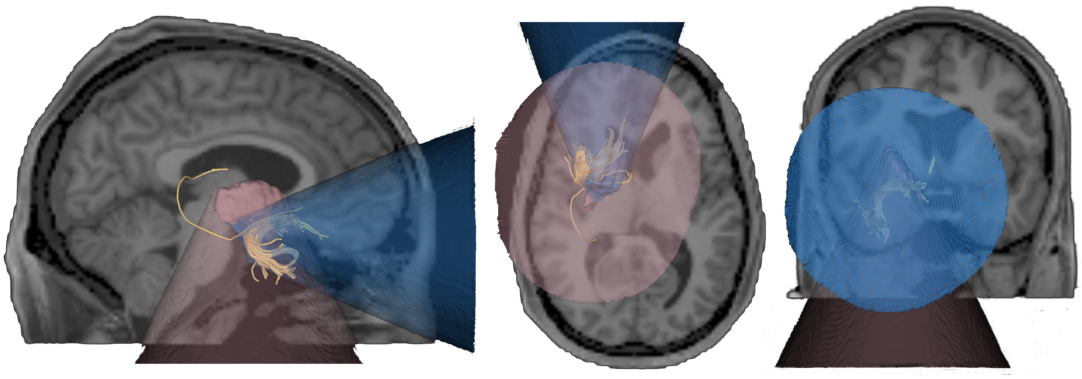


Figure 3.4: Fuzzy spaces in 3D modeling the relation $inferior(Putamen)$ and $anterior(Amygdala)$. Both spaces have been thresholded and visualized as cones.

Interpreting or choosing a threshold value for ACS might be quite hard when working with tractograms composed of millions of streamlines. This is particularly the case for applications demanding a high accuracy like surgical planning. To this end, exploiting the previously presented simplification method [Mercier et al., 2018], we progressively group the most similar streamlines into generalized cylinders with elliptical basis, producing a nested hierarchy of resolution levels, where every level of detail corresponds to the fusion of two streamlines (or cylinders).

Here, we propose to compare streamlines using an extension of the computational model of Weighted Currents (WC_{ext}) [Gori et al., 2016]. Two streamlines are considered similar, and thus merged together, only if their trajectories are alike, their endpoints are close to each other and their ACS values are similar. We also propose an automatic stopping criterion for the multi-resolution

to prevent oversimplification (e.g. a single cylinder). We use the inner product in the Weighted Currents space to compute angles between streamlines and cylinder center-lines. Two cylinders/streamlines are not merged together if they are almost orthogonal (angle $> 89°$). Streamlines that were never merged in the whole process are then considered outliers and discarded. The proposed technique simplifies the geometric representation, preserving at the same time the overall structure of the original tractogram (i.e. shape, connectivity and ACS).

We also provide a GUI where the user can navigate in real-time through different levels of detail and at the same time select only the streamlines/cylinders with an ACS value above a user-defined threshold. This can help clinicians to better understand the structure of the tracts and for surgery preparation. Visual examples at different resolutions and ACS thresholds for the Uncinate Fasciculus (UF) and Inferior Fronto-Occipital Fasciculus (IFOF) can be found in Fig. 3.5 and Fig. 3.6.

Fuzzy definitions are implemented in Python and the computational time is about 10-15s per definition. The source code is publicly available at https://github.com/PietroGori/FuzzyTracts. Based on both imaging and dissection studies [Wakana et al., 2007, Sarubbo et al., 2013, Ebeling and von Cramon, 1992, Catani and Thiebautdeschotten, 2008] and with the help of an experienced neurosurgeon, we defined and modeled 12 white matter tracts. The definitions and modeling of the UF and IFOF can be found in [Delmonte et al., 2019], while for the 10 other tracts, they can be found in the Master's thesis of A. Di Girolamo [Di Girolamo, 2019].

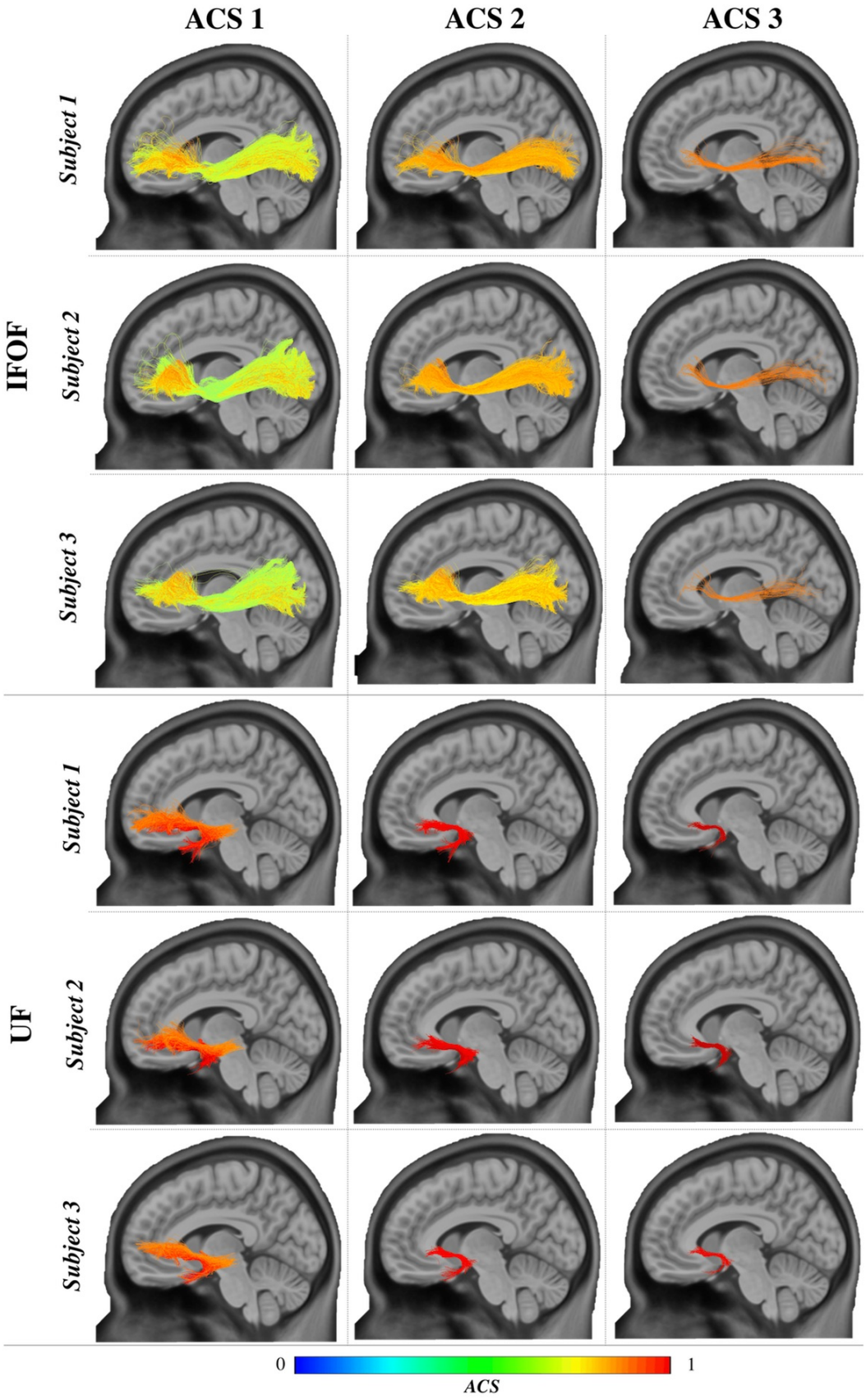


Figure 3.5: Segmentation of IFOF and UF bundles of three subjects using three different thresholds for the ACS (0.5, 0.65, 0.7 for the IFOF and 0.7, 0.85, 0.9 for the UF respectively). Results are shown on the MNI152 T1w image.

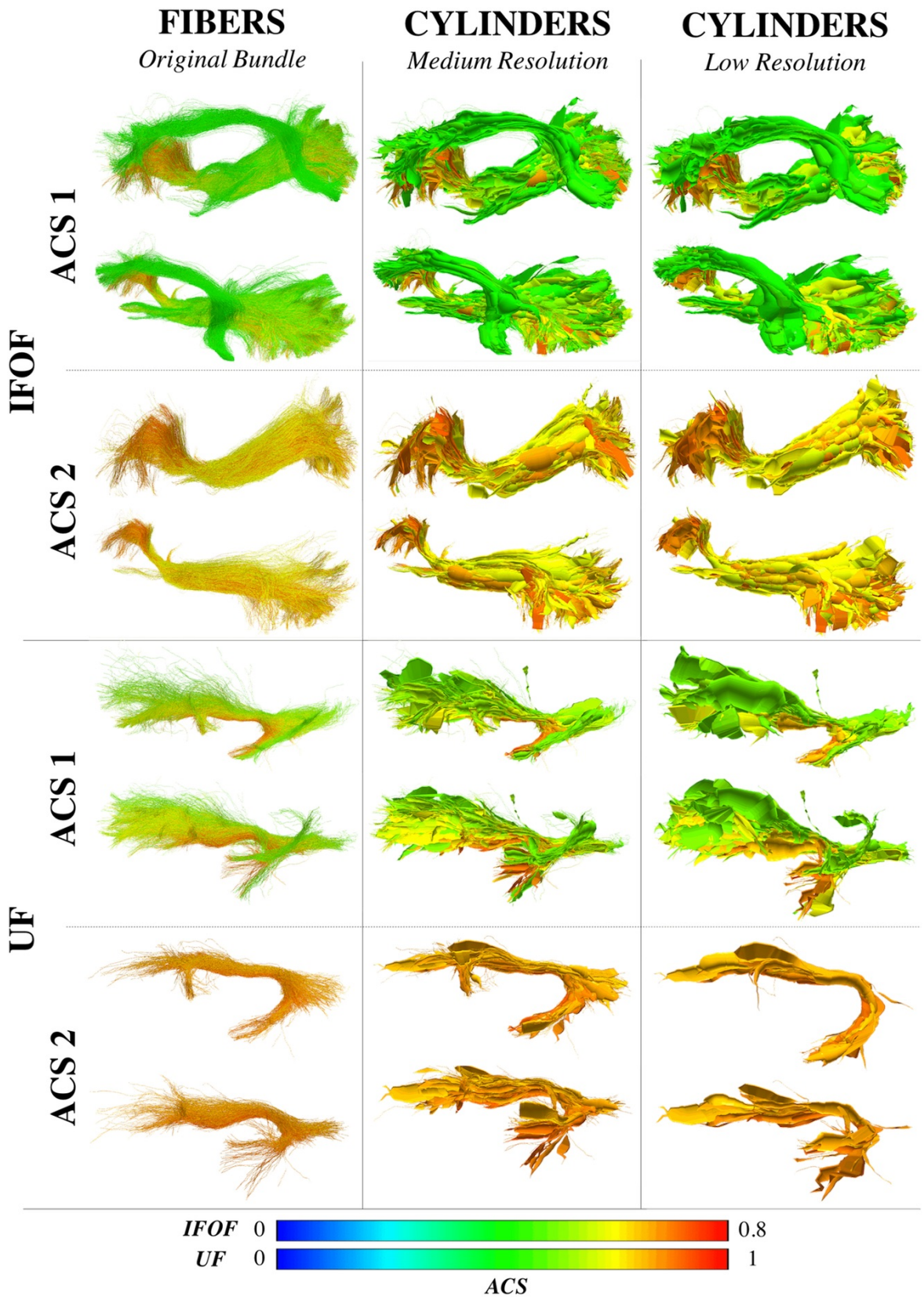


Figure 3.6: IFOF and UF bundles visualized at three different resolution levels and segmented with two different ACS thresholds (0.4, 0.57 for the IFOF and 0.6, 0.81 for the UF respectively).

Optimal Transport

As second method, we have presented in [Feydy et al., 2019a] an efficient, fast and scalable algorithm that solves the regularised (entropic) Optimal Transport problem for transferring the labels of a labeled atlas to a subject tractogram. By leveraging an efficient, GPU-based and multi-resolution implementation, we are able to directly segment a whole brain tractogram without any pre-processing or clustering. The entire algorithm only depends on two meaningful hyperparameters, the *blur* and *reach* scales, that define the minimum and maximum distances at which two streamlines are compared. Intuitively, *blur* is the resolution of the finest details that we try to capture, while *reach* acts as an upper bound on the distance that points may travel to meet their targets – instead of seeing them as *outliers*. On top of label transfer, we also propose to estimate a probabilistic atlas, as a Wasserstein barycenter, of a population of tracts modeled as track density maps [Wassermann et al., 2010], where each map contains, for every voxel in the space, the probability that a specific track (e.g. IFOF) passes through.

Optimal Transport (OT) looks for a transportation plan between two probability distributions, α and β, that minimizes a cost metric under marginal constraints, ensuring a full covering of the input data. When working with discrete data, as streamlines or track density maps, the distributions are encoded as weighted sums of Dirac: $\alpha = \sum_{i=1}^{\mathrm{N}} \alpha_i \, \delta_{x_i}$ and $\beta = \sum_{j=1}^{\mathrm{M}} \beta_j \, \delta_{y_j}$ with weights $\alpha_i, \beta_j \geqslant 0$ and samples' locations $x_i, y_j \in \mathcal{X} = \mathbb{R}^{\mathcal{D}}$. In most applications, the feature space $\mathcal{X}$ is the ambient space $\mathbb{R}^3$ endowed with the standard Euclidean metric. This is the case, for instance, when using *track density maps* where α_i and β_j are the probabilities associated to the voxel locations x_i and y_j, respectively. Meanwhile, when using streamline tractograms, a usual strategy is to resample each streamline to the same number of points P. In this case, the feature space $\mathcal{X}$ becomes $\mathbb{R}^{\mathrm{P}\times 3}$ and x_i, y_j are the N and M streamlines that constitute the source and target tractograms with uniform weights $\alpha_i = \frac{1}{\mathrm{N}}$ and $\beta_j = \frac{1}{\mathrm{M}}$, respectively. Each streamline, modeled as a polyline, is thus embedded into a feature vector by simply concatenating its points. Distances can be computed using the standard Euclidean L^2 norm – normalized by $1/\sqrt{\mathrm{P}}$ – and we alleviate the problem of *streamline orientation* by augmenting our tractograms with the mirror flips of all streamlines. This corresponds to the simplest of all encodings for unoriented curves.

Based on previous works on Optimal Transport[Chizat et al., 2018, Peyré et al., 2019, Cuturi, 2013, Feydy and Trouvé, 2018, Feydy et al., 2019b], we consider a generalization of the original Kantorovitch formulation where α and β don't have the same total mass or may contain *outliers*, which is typically the case when working with streamline bundles, and propose to minimize the *unbiased* Sinkhorn divergence, which defines a *positive*, *definite* and *convex* loss function, as shown in [Feydy et al., 2019b]. Furthermore, to tend towards the $O(n\log(n))$ complexity of multiscale methods, we use a coarse subsampling of the x_i's and y_j's in the first few iterations. Here, we use a simple K-means algorithm to group together similar locations and use only their average during the first iterations (instead than all locations), but other strategies could be employed. In this way, when moving from this coarse subsampling to the full resolution, we can *prune out* useless computations and thus speed-up the entire algorithm. By heavily relying on the `KeOps` library [Feydy et al., 2020], we proposed the *first GPU implementation of a multiscale OT solver*. It provides a `x1,000` speed-up when compared to simple `PyTorch` GPU implementations of the Sinkhorn loop [Cuturi, 2013], while keeping a linear (instead of quadratic) memory footprint. It is freely available on the `PyPi` repository (`pip install geomloss`) and at `www.kernel-operations.io/geomloss`.

In Fig.3.7, we show how OT plans can be used to transport labels from a streamline atlas to a subject tractogram. This method takes into account the whole organization of the bundles – unlike standard nearest neighbours or clustering algorithms –, detects outliers and is not hampered by streamline crossings, differently from standard registration algorithms (e.g., LDDMM). In Fig.3.8,

we show the second application of the proposed method, where we use Sinkhorn divergences to estimate a *geometric* average of track density maps, which can be seen as a soft atlas. More details can be found in [Feydy et al., 2019a].

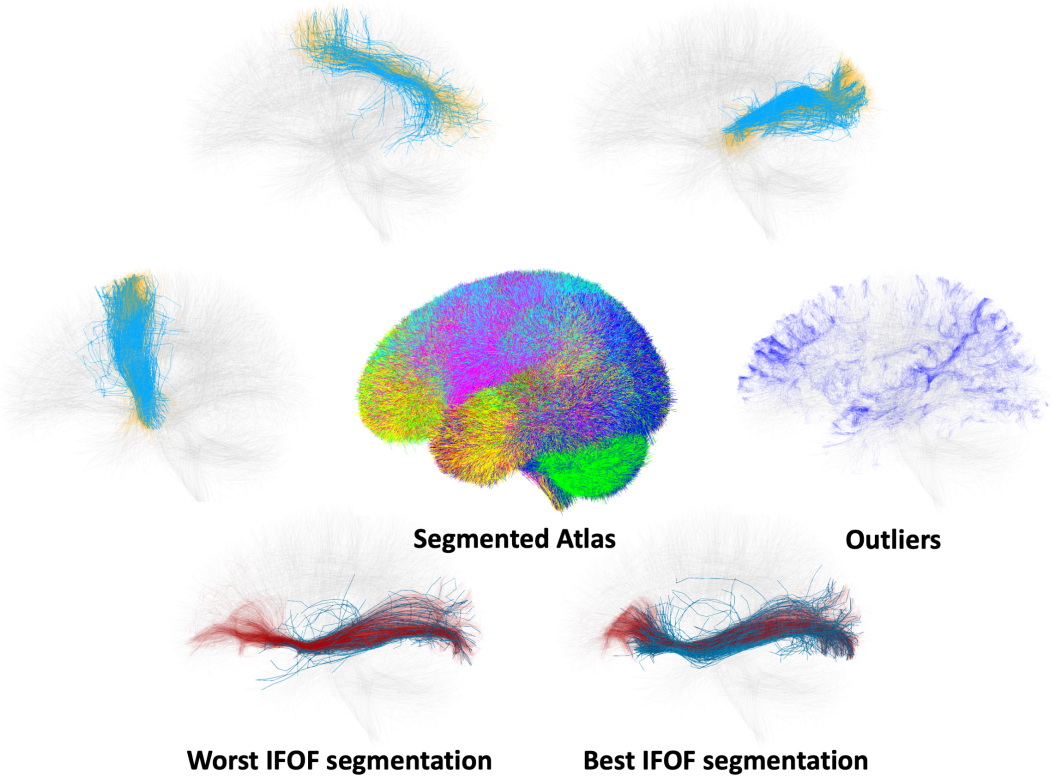


Figure 3.7: Label transfer between the segmented atlas (in the middle) and the subject tractograms. Top: some clusters of the atlas (in orange) with their respective segmentations (in light blue) of one random subject. Detected outliers are on the right (dark blue). Bottom: worst and best segmentation of the left IFOF, among the five tested subjects, compared to a manual segmentation.

3.4 Conclusions, Limitations and Perspectives

In this Chapter, we have presented two multi-scale methods to model, simplify and segment white matter tractograms. They are fast, interpretable and explainable. Furthermore, they depend on few hyper-parameters that can be easily tuned by the user based on the resolution of the image, and the anatomy of the tracts under analysis. Another important characteristic of the proposed methods is the fact that they are "soft" and not "hard", namely they provide a membership score or probability for each streamline and not a one-hot encoding. We believe that this is quite important since it implicitly encodes the uncertainty of the algorithm. If the algorithm gives to a streamline a 50% score between two tracts, it probably means that the streamline is at the border between them or that it's in an area not well defined, and thus uncertain, by the definitions of the tracts. Furthermore, knowing that there is still a poor inter- and intra-user reproducibility for the manual segmentation of white matter tractograms [Zhang et al., 2010] makes us believe that looking for a clear, one-hot encoding segmentation does not make much sense.

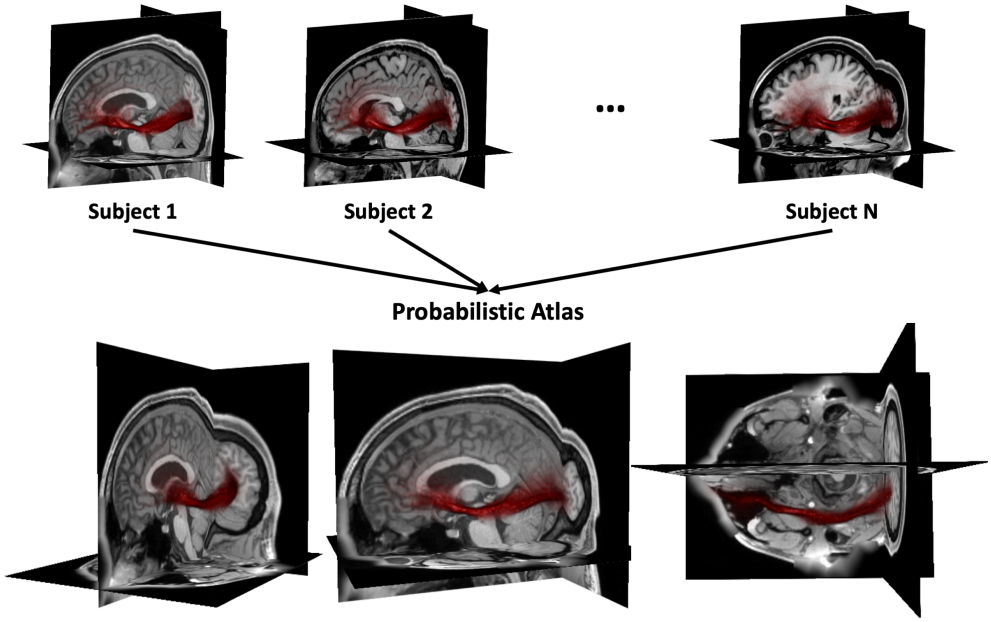


Figure 3.8: Probabilistic atlas of the left IFOF from 5 track density maps (in red). Top row: the densities of the five subjects, shown with their T1-wMRI. Bottom row: three views of the obtained atlas, alongside the T1-wMRI of one subject.

Limitation

The preliminary results shown in this Chapter are promising but we still lack a quantitative analysis and comparison with other state-of-the-art methods on a large clinical data-set. Unfortunately, these works were developed during the Covid-19 pandemic, which limited the time and availability of our clinical partners from St. Anne hospital.

Do we need Deep Learning?

Interestingly, the proposed methods do *not* use deep learning architectures. This makes them more interpretable, explainable and they do not need a large, labeled data-set for training. Furthermore, while deep learning methods provide fast segmentations, their results can still be unsatisfactory, as shown in [Bertò et al., 2021], and are not robust to variations in bundle sizes, tracking methods and data quality. This is mainly because there is still a lack of large, open, annotated, and multi-site datasets with several tractograms for each subject obtained with different tractography algorithms. This prevents a rigorous analysis of the generalizability and reliability of the existing methods on external (i.e., out-of-distribution) test sets. Furthermore, many works use different techniques to generate the reference segmentations and there is still no consensus on the best metric to validate the results [Joshi et al., 2024].

To fill that gap, some recent works have (finally) released their annotations and datasets, such as TractSeg [Wasserthal et al., 2018] and TractoInferno [Poulin et al., 2022]. This has set a very good example for the community and gave the opportunity to more researchers to came up with interesting solutions by using, for instance, self-supervised learning on un-annotated datasets [Chen et al., 2023, Xue et al., 2023, Joshi et al., 2024, Ghazi et al., 2023]. However, the number of available data is still not comparable with computer vision datasets (e.g., ImageNet), where the tasks and validation metrics are usually well defined and accepted by the community. We believe that an interesting research direction could be using an hybrid AI algorithm by merging deep learning

segmentation methods with symbolic AI (fuzzy sets and logic). The latter would give a coarse (fuzzy) prior segmentation, based on prior anatomical definitions, which would be refined by the learning capacity of deep neural networks.
Instead than directly modeling the prior knowledge using symbolic AI, another interesting research direction would be to leverage pre-trained language models, like BERT [Devlin et al., 2019] or GPT4 [OpenAI, 2023], to directly encode the qualitative anatomical definitions. Vision-language pre-training has been recently used in the medical domain [Tiu et al., 2022, Lu et al., 2024, Wu et al., 2023, You et al., 2023, Boecking et al., 2022, Huang et al., 2021], in particular to leverage clinical (e.g., radiology or pathology) reports as weak, but free and available, annotations. However, in our case, we would not have paired vision-language data, namely reports describing the semantic contents of images, and therefore this strategy could not be used. Differently, we could, for instance, compare the representations of the language encoder, modeling the anatomical definitions, with the ones given by a chatbot (using a similar language model as before) describing the segmentation results. The divergence between the prior anatomical definitions and the semantic description of the chatbot could be used as (weak) supervision signal to correct the results of the segmentation network.

Extending to pelvic nerves

Nerve imaging from diffusion MRI and reconstruction using tractography are mainly applied to the brain and the central nervous system. However, there is a need to plan complex surgical interventions in the pelvic region, where nerve damage can cause significant complications, like genito-urinary dysfunctions. Currently, no licensed softwares nor research works propose an automatic solution for modeling and segmenting the pelvic nerves. Furthermore, to the best of our knowledge, there is no labeled dataset or atlas freely available, which hampers the use of learning strategies (e.g., deep learning) or transferring algorithms (e.g., optimal transport). That is why we believe that an interesting perspective would be to use the proposed fuzzy-based segmentation method for pelvic nerves segmentation. A preliminary work (based on manual segmentation) has shown great promise [Muller et al., 2019].

New questions could be addressed to better adapt our model to the definitions of the pelvic nerves. For instance: 1) How to adapt and extend existing definitions of spatial relations to the case of very elongated structures (like some organs and bones in the pelvic region)? 2) How to account for portions of streamlines satisfying a relation in a gradual manner? 3) How to account for the shape of organs in the definition of spatial relations? 4) How to combine the degrees of satisfaction of relations along fibers into a value representing to which degree a fiber is part of a given nerve tract, and derive a decision on nerve recognition from it?

Statistical Shape Analysis

In the previous section, we have proposed a method, based on optimal transport, to compute a probabilistic atlas (estimated as a Wasserstein barycenter, see Fig. 3.8) of white matter tracts represented as track density maps. Another interesting perspective would be to leverage the proposed multi-resolution geometric representation to statistically analyze the extracted tracts. Tracts could be compared using transformations applied first at a coarse level, and then refined using higher levels. This methodology has proven to enforce robustness in similar applications from Computer Graphics [Hoppe, 1996, Lee et al., 1999, Manson and Schaefer, 2011]. In medical imaging, deformations are usually defined as diffeomorphisms since they preserve the anatomical organization of the brain. However, they only consider the morphology of the structures, completely ignoring possible

functional signals mapped onto them. Indeed, each streamline "carries" an important quantity of information that goes beyond its trajectory through the voxels of the MR image. It connects two different areas of the brain (*i.e.,* anatomical connectivty) and one can map voxel-wise functional signals onto it, such as: brain activity time series (e.g., MEG, fMRI), metabolic imaging data (e.g., PET, MR spectroscopy) or quantities describing the microstructure of the brain (e.g., Fractional Anisotropy). Geometry, connectivity and functional signals have been shown to be crucial in the characterization of the pathophysiological processes underlying a condition (i.e., tumor) or a disease [Horsfield and Jones, 2002, Smith, 2016].

To this end, following recent works about cascades of diffeomorphisms [Gori et al., 2015], metamorphoses [Charlier et al., 2017, François et al., 2021, François et al., 2022, Maillard et al., 2022] and functional maps [Ovsjanikov et al., 2012, Li et al., 2022], we could investigate a fast and reliable deformation scheme for the proposed multi-resolution representation in order to combine the statistical analysis of shape and functional signals into a single and unifying framework.

As data-term losses, we could also leverage our previous work on robust losses [Roussillon et al., 2019] for white matter tracts, where we used a robust L^p-RKHS norm:

$$A(Q, Q') = \sum_{i=1}^{n} \min_{j=1,\dots,m} \left(||q_i - q'_j||_V^2 \right)^{p/2} \tag{3.1}$$

where the norm $||\cdot||_V$ could be the one of varifold [Charon et al., 2020], $Q = \{q_i, i = 1, \dots, n\}$ and $Q' = \{q'_j, : j = 1, \dots, m\}$ are two tracts and q_i and q'_j are two streamlines of Q and Q' respectively. In [Roussillon et al., 2019], we have shown that usual metrics for streamlines, such as the L^2 or Varifold/Currents [Charon et al., 2020] metrics, suffer from the so-called "shrinking phenomenon". This happens when one wants to compute the distance between two streamlines that are far away from each other, with respect to the used norm. Indeed, the L2 distance or varifold distance is defined as: $||q - q'||_V^2 = ||q||_V^2 + ||q'||_V^2 - 2\langle q, q' \rangle_V$. If the inner product $\langle q, q' \rangle_V$ is very small, then the optimization process will concentrate on minimizing the norm of the source streamline $||q||_V^2$, since the norm of the target streamline $||q'||_V^2$ is constant. This makes the streamline curl up/shrink as it can be seen in Fig. 3.9. By using the proposed robust $L^p - RKHS$ metric, we can avoid this phenomenon.

Furthermore, it would also be interesting to extend this robust metric to functional and geometric data, similarly to weighted/functional currents [Charon and Trouvé, 2014, Gori et al., 2016].

Eventually, once defined the transformations, we could also automatically estimate the average (and its associated variability) of the tracts using the atlas construction strategy [Gori et al., 2017].

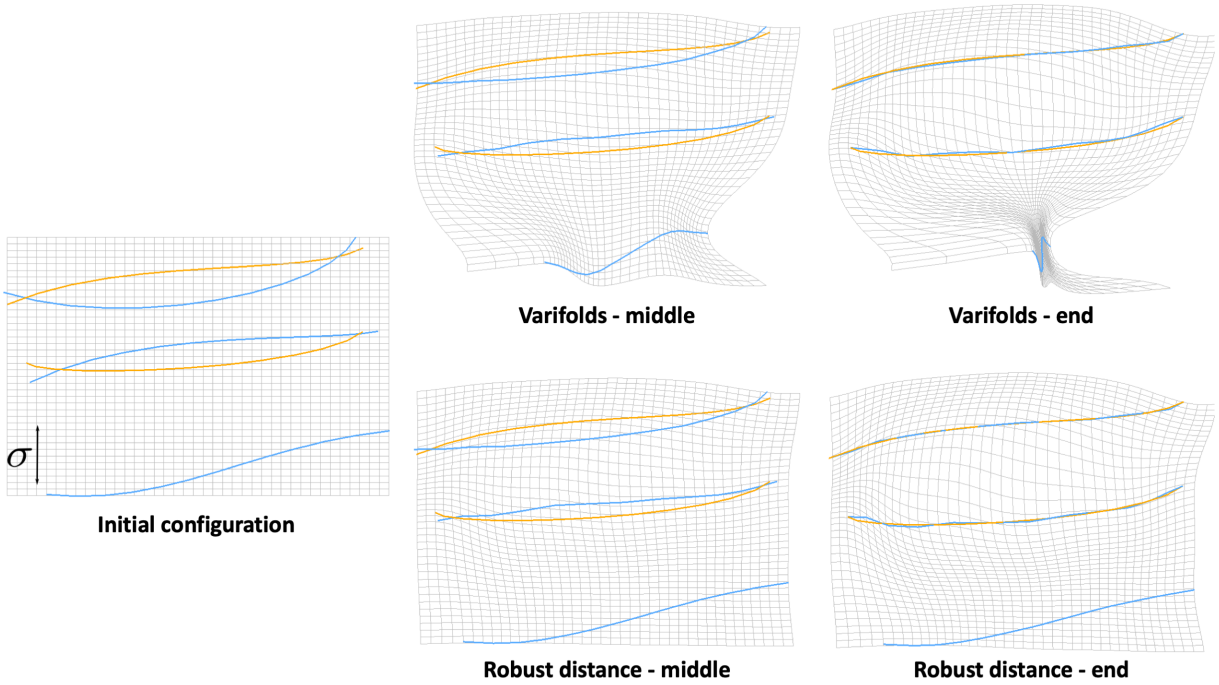


Figure 3.9: Diffeomorphic matching from three curves (blue) to two curves (orange) with a data attachment term using varifolds (first row) and with a robust varifolds distance (second row, Eq. 3.1 with $p = 0.1$). In the first row, the grid shows that the deformation is strongly distorted by the outlier curve (bottom blue curve). In the second row, the outlier is not taken into account during registration, without any pre-processing or new hyper-parameter.

Conclusions and perspectives

In this manuscript, I have summarized the majority of the research I have conducted since arriving at Télécom Paris in 2017. I could accomplish all these works thanks to collaborations with colleagues from IPParis (LTCI and LIX), hospitals (St Anne and Necker), other laboratories (NeuroSpin, MAP5, Uni. Trento, Uni. Torino) and companies (Guerbet, Incepto, Philips). In particular, I had the pleasure to work with 10 Master students, 13 PhD students and 2 post-docs.

During the last 7,5 years, I have worked on new symbolic and learning AI methods, related to fuzzy logic, knowledge distillation, contrastive learning, transfer learning, image registration, and image-to-image translation, for 1- modeling prior medical knowledge and anatomy, 2- learning relevant representations of anatomical imaging data and 3- transferring representations between domains. All methods were developed taking into account specific constraints of the medical imaging data (e.g., dataset size, biases, physics, etc.) and the needs of a precise clinical application.

The presented manuscript was structured into three chapters. Each chapter discussed one of the three main applications I have worked on, along with their respective perspectives. In the following, I will introduce other projects I have started, focusing on their clinical significance, challenges, and research prospects.

Towards multi-modal foundation models for medicine

Recent medical datasets contain data from multiple and heterogeneous sources of information (images, genetic, multi-omic, clinical, etc.) to obtain a complete and holistic view about the health of patients with the hope to identify more discriminative, and probably multi-modal, pathological biomarkers. This calls for new self-supervised multi-modal methods, since most of these datasets contain healthy or unlabeled data. Eventually, these methods could be used to create general medical foundation models for carrying out diverse downstream tasks using little or no task-specific labeled data [Moor et al., 2023]. In computer vision, many methods have already been proposed, for instance well adapted to videos, text or 3D rendering (e.g., CLIP [Radford et al., 2021] or [Alayrac et al., 2020, Yuan et al., 2021]). In medical imaging, most methods focused on datasets with paired images and reports, such as GLoRIA [Huang et al., 2021], Con VIRT [Zhang et al., 2022b], CheXzero [Tiu et al., 2022], MedClip [Wang et al., 2022b] or combine several imaging modalities for a single downstrem task, such as segmentation [Ma et al., 2024] or registration [Pielawski et al., 2020]. Other methods have also been proposed for combining images and genetics, such as ContIG [Taleb et al., 2022]. Even if many methods have been recently published, see [Zong et al., 2023, Qiu et al., 2023] for an exhaustive list, all these multi-modal/foundation methods either use (or slightly adapt) existing self-supervised methods (e.g., SimCLR, DINO), that have not been conceived for multi-modal data, or are developed for a specific task, such as segmentation [Ma et al., 2024], report creation [Yang et al., 2024], clustering [Lin et al., 2021, Trosten et al., 2021, Xu et al., 2022]. Only few, recent works, such as [Tian et al., 2020a, Federici et al., 2020, Liu et al., 2021, Ke et al., 2023, Liang et al., 2023], propose new methods or approaches for merging and organizing

multi-modal information into common and task-relevant factors.
I believe that many future biomarkers will be multi-modal and that, even if medical images will play an important role, they will need to be combined with other data. An interesting research direction will thus be studying how to leverage and adapt the proposed geometric framework for a truly self-supervised multi-modal method, that could be employed in several clinical applications, like brain disorder detection. This avenue will open up several interesting questions that we plan to work on in the coming years, such as: how to merge heterogeneous sources of information (e.g., early-, middle-, late- fusion, architecture) ? How to use unpaired data or tackle missing modality ? How common information between two modalities should be defined ? Should we create task-specific models or more generic ones ? and many others.

A new clinical application: Histopathology

In collaboration with the St. Joseph hospital, I have recently started working on a new clinical application: Whole Slide histopathology Image (WSI) analysis [Gurcan et al., 2009]. In particular, we have started a project whose goal is the development of an innovative multimodal deep learning method that, by combining histological images with clinical/biological data, should improve the diagnostic accuracy of the Sjögren's syndrome (SjS) [Liao et al., 2022], a rare disorder of the immune system, by identifying new interpretable biomarkers.
The gigapixel size of WSIs, which can easily reach billions of pixels, makes their manual analysis very time-consuming and presents significant challenges for conventional Deep Learning (DL) methods, as they are not designed to support such large images [Cheplygina et al., 2019, Srinidhi et al., 2021]. To address that, a simple approach involves dividing the WSI into smaller patches that DL methods can easily handle [Hou et al., 2016, Wei et al., 2019]. However, patch-level annotations are rarely available and, as in our case, one usually uses (weak) slide-level labels. Multiple-Instance Learning (MIL) methods, coupled with Deep Learning Feature Extraction (FE), have thus emerged as the most prominent solution in the field of WSI classification, where FE avoids the costly and experience-based feature engineering part, while MIL eliminates the need for patch-level (or pixel-level) annotations [Tschuchnig et al., 2022, Qu et al., 2022]. Under MIL formulation, each WSI is treated as a "bag" containing multiple instances in the form of patches, which are embedded with Convolutional Neural Network (CNN) or Vision Transformer (ViT) backbones. The bag is labeled positive (i.e., diseased) if at least one of its patches is positive, or negative if all patches are negatives [Campanella et al., 2019, Ilse et al., 2018]. In general, the existing methodologies follow a two-step pipeline: 1) feature extraction (FE) from individual patches, and 2) MIL aggregation through a pooling operation to predict the slide label [Li et al., 2018, Lu et al., 2021, Li et al., 2021a, Shao et al., 2021, Zhang et al., 2022a, Guan et al., 2022]. The first step is usually performed by leveraging existing self-supervised learning methods, such as DINO [Caron et al., 2021], MOCO-v3 [He et al., 2020] or Barlow Twins [Zbontar et al., 2021]. In the second step, MIL aggregation methods can be categorized into two groups: *instance-based* and *embedding-based* methods. Instance-based methods use an instance-level classifier, which predicts a score for each patch, and then a simple pooling operator (usually average or max) to make the final prediction for the entire slide. These methods are highly interpretable, easily explainable and with very few parameters. However, they highly depend on the quality of the features extracted during the first step. To increase reliability, researchers proposed to aggregate features instead than scores, moving the classification head after the pooling. These are called embedding-based methods whose pooling mechanisms, usually based on attention or self-attention, are more complex (more parameters) than instance-based ones [Ilse et al., 2018, Lu et al., 2021, Shao et al., 2021, Li et al., 2021a]. On the one hand, this means

that the model has a greater capability of learning how to correctly aggregate the features and thus might have a greater prediction power. On the other hand, interpretability and explainability can decrease[1] and at the same time computational complexity and overfitting might increase (i.e., number of needed training samples increases).
Many interesting research questions, in particular related to the proposed SjS project, will be addressed. Concerning the first FE step of the pipeline, almost no self-supervised methods has been specifically conceived for histological images. Most of the proposed methods, such as the latest Foundation model UNI [Chen et al., 2024], use existing methods (e.g., Dino-v2) that have been proposed for natural images. Very few works explored specific solutions for histological images and, most of them, as in [Kang et al., 2023], only explore new data augmentation tailored for histopathology images (e.g., stain changes). An interesting research direction will thus be proposing new self-supervised methods specifically tailored for histopathology images. These methods should mimic the decision mechanism of pathologists by leveraging the pyramidal nature (multiple magnifications) of the images, accounting for their variations (and biases) in stain, anatomical location and cell size, and using complementary information from clinical and biological data. Related to that, we will also investigate whether the existing available histopathological imaging datasets, which mostly come from The Cancer Genome Atlas (TCGA) initiative and thus describe cancerous tissue, are also useful for non-cancerous, rare diseases, such as SjS.
About the second step, our preliminary results [Mammadov et al., 2024] show that simple instance-based MIL methods, when combined with robust self-supervised feature extractors, are on par or even outperform complex embedding-based methods in WSI classification. This indicates that our future efforts should be focused on developing simple, yet relevant and discriminative, MIL methods that will leverage the previously described (new) self-supervised methods. This will produce robust, highly interpretable and explainable algorithms, whose results will thus be more trustworthy and accepted by physicians (and regulators).

Going from a clinical research project to a Start-Up

Personalized medicine in surgery is based on the development of 3D, precise and patient-specific models of the anatomical structures and pathologies, like tumors or malformations. These individual models are of uttermost importance for pre-operative planning and per-operative guidance. They are usually built from segmentations obtained via deep learning methods trained on structural medical images, such as MRI or CT scans. These methods are highly developed and perform well when the training datasets are large and the anatomical structures present a low heterogeneity in size and pose with high-contrast and clear boundaries. This is usually the case for dense structures, like bones, in adults. However, there are still challenging segmentation tasks, like in pediatrics or in the abdomen and pelvis, where deep learning algorithms may fail.
Since 2017, when I joined Télécom Paris, we have proposed, through collaborations with the Necker hospital and companies (Philips, Incepto and Guerbet), several solutions for improving the segmentation of abdominal and pelvic structures. In particular, we have worked on pediatric segmentation (kidney, renal tumor [La Barbera et al., 2021], renal tubular structures [La Barbera et al., 2022, La Barbera et al., 2023] and pelvic vessels [Virzì et al., 2018, Virzì et al., 2020]), pancreas [Vétil et al., 2022] and prostate cancer segmentation [Ruppli et al., 2023]. Part of these works have produced two patents [Delmonte et al., 2023, Vétil et al., 2024] and some of them have been gathered into a software that can, almost automatically, segment all pelvic anatomical structures (e.g. bones, organs, vessels), tumors and malformations and even some pelvic nerves. These segmenta-

[1] even though attention and self-attention mechanisms can be exploited in that regard.

tions are then used to produce a 3D, anatomical digital twin of the patient's pelvis which is currently used at the Necker hospital for pre-operative planning and per-operative guidance in daily clinical practice (Prototype TRL4). Indeed, there was an urgent need for a solution helping the surgeons, since there is currently no licensed software nor research work proposing an automatic solution for modeling and segmenting all pelvic structure, including the nerves. The initial user feedback was very positive, indicating an easy-to-use solution, highly useful in practice. Furthermore, a market research concluded that many other surgeons, from various domains, expressed a high interest in our software. Indeed, we received many requests of collaborations (e.g., urethral, maxillofacial, gynecological surgery) with their own specificity in terms of data, protocol and difficulties. This motivated us to step out of the researcher's comfort zone and move towards the market. We have thus assembled a great team of 8 people and we are currently in the process of creating a Start-Up. This opens up several challenges and opportunities, from both a research and business point of view, with, as final goal, the creation of a 3D, automatic, accurate and patient-specific digital twin of the entire human body, including nerves and vessels [2].

[2]this is probably more of a dream than a goal, but apparently you need to make investors dream in order to secure funding...

Bibliography

[Abid and Zou, 2019] Abid, A. and Zou, J. (2019). Contrastive Variational Autoencoder Enhances Salient Features. 29, 30, 31

[Abrol et al., 2021] Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., and Calhoun, V. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature communications*, 12(1):1–17. Publisher: Nature Publishing Group. 8

[Ahmad and Lin, 1976] Ahmad, I. and Lin, P.-E. (1976). A nonparametric estimation of the entropy for absolutely continuous distributions (Corresp.). *IEEE Transactions on Information Theory*, 22(3):372–375. 35

[Alayrac et al., 2020] Alayrac, J.-B., Recasens, A., Schneider, R., Arandjelović, R., Ramapuram, J., De Fauw, J., Smaira, L., Dieleman, S., and Zisserman, A. (2020). Self-Supervised MultiModal Versatile Networks. In *Advances in Neural Information Processing Systems*, volume 33, pages 25–37. Curran Associates, Inc. 75

[Alifieris and Trafalis, 2015] Alifieris, C. and Trafalis, D. T. (2015). Glioblastoma multiforme: Pathogenesis and treatment. *Pharmacology & Therapeutics*, 152:63–82. 41

[Almansour et al., 2021] Almansour, M., Ghanem, N. M., and Bassiouny, S. (2021). High-Resolution MRI Brain Inpainting. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–6. 44

[Alvi et al., 2018] Alvi, M., Zisserman, A., and Nellåker, C. (2018). Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0. 10

[American Psychiatric Association and American Psychiatric Association, 2013] American Psychiatric Association and American Psychiatric Association, editors (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association, Washington, D.C, 5th ed edition. 10

[Amor et al., 2023] Amor, B. B., Arguillère, S., and Shao, L. (2023). ResNet-LDDMM: Advancing the LDDMM Framework Using Deep Residual Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(03):3707–3720. Publisher: IEEE Computer Society. 52

[Andersen et al., 2010] Andersen, S. M., Rapcsak, S. Z., and Beeson, P. M. (2010). Cost Function Masking during Normalization of Brains with Focal Lesions: Still a Necessity? *Neuroimage*, 53(1):78–84. 58

[Arsigny et al., 2006] Arsigny, V., Commowick, O., Pennec, X., and Ayache, N. (2006). A log-euclidean framework for statistics on diffeomorphisms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 924–931. Springer. 43

[Asano et al., 2020] Asano, Y. M., Rupprecht, C., and Vedaldi, A. (2020). Self-labelling via simultaneous clustering and representation learning. In *ICLR*. 24

[Ashburner, 2007] Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113. 43

[Ashburner and Friston, 2000] Ashburner, J. and Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage*, 11(6):805–821. 22, 36

[Ashburner and Friston, 2011] Ashburner, J. and Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *NeuroImage*, 55(3):954–967. 43, 52

[Assran et al., 2023] Assran, M., Duval, Q., Misra, I., Bojanowski, P., Vincent, P., Rabbat, M., LeCun, Y., and Ballas, N. (2023). Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15619–15629, Vancouver, BC, Canada. IEEE. 57

[Atzmueller, 2015] Atzmueller, M. (2015). Subgroup discovery. *WIREs Data Mining and Knowledge Discovery*, 5(1):35–49. 24

[Avants et al., 2008] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41. Publisher: Elsevier. 43, 54

[Bahng et al., 2020] Bahng, H., Chun, S., Yun, S., Choo, J., and Oh, S. J. (2020). Learning debiased representations with biased representations. In *International Conference on Machine Learning*, pages 528–539. PMLR. 14, 15

[Balakrishnan et al., 2019] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, a. A. V. (2019). VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE TMI*, 38(8):1788–1800. 54

[Barbano et al., 2023a] Barbano, C. A., Dufumier, B., Duchesnay, E., Grangetto, M., and Gori, P. (2023a). Contrastive learning for regression in multi-site brain age prediction. In *IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. 7, 12, 22

[Barbano et al., 2023b] Barbano, C. A., Dufumier, B., Tartaglione, E., Grangetto, M., and Gori, P. (2023b). Unbiased Supervised Contrastive Learning. In *International Conference on Learning Representations (ICLR)*. 7, 13, 15, 21

[Bardes et al., 2022] Bardes, A., Ponce, J., and LeCun, Y. (2022). VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. In *ICLR*. 9

[Bashyam et al., 2020] Bashyam, V. M., Erus, G., Doshi, J., Habes, M., Nasralah, I., Truelove-Hill, M., Srinivasan, D., Mamourian, L., Pomponio, R., Fan, Y., and others (2020). MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain*, 143(7):2312–2324. Publisher: Oxford University Press. 17

[Bayer et al., 2022] Bayer, J. M. M., Thompson, P. M., Ching, C. R. K., Liu, M., Chen, A., Panzenhagen, A. C., Jahanshad, N., Marquand, A., Schmaal, L., and Sämann, P. G. (2022). Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multisite neuroimaging analyses. *Frontiers in Neurology*, 13. 9, 10, 61

[Beg et al., 2005] Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157. 43, 52

[Bell and Sejnowski, 1995] Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159. 34

[Ben-Cohen et al., 2018] Ben-Cohen, A., Klang, E., Raskin, S., Soffer, S., Ben-Haim, S., Konen, E., Amitai, M., and Greenspan, H. (2018). Cross-Modality Synthesis from CT to PET using FCN and GAN Networks for Improved Automated Lesion Detection. *Engineering Applications of Artificial Intelligence*, 78:186–194. 45

[Bengio et al., 2013] Bengio, Y., Courville, A., and Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE TPAMI*. 38

[Bertò et al., 2021] Bertò, G., Bullock, D., Astolfi, P., Hayashi, S., Zigiotto, L., Annicchiarico, L., Corsini, F., De Benedictis, A., Sarubbo, S., Pestilli, F., Avesani, P., and Olivetti, E. (2021). Classifyber, a robust streamline-based linear classifier for white matter bundle segmentation. *NeuroImage*, 224:117402. 71

[Bilello et al., 2016] Bilello, M., Akbari, H., Da, X., Pisapia, J. M., Mohan, S., Wolf, R. L., O'Rourke, D. M., Martinez-Lage, M., and Davatzikos, C. (2016). Population-based MRI atlases of spatial distribution are specific to patient and tumor characteristics in glioblastoma. *NeuroImage: Clinical*, 12:34–40. 42, 46

[Bloch, 1999] Bloch, I. (1999). Fuzzy relative position between objects in image processing: a morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):657–664. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. 65

[Bloch, 2005] Bloch, I. (2005). Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, 23(2):89–110. 64

[Bloch and Ralescu, 2023] Bloch, I. and Ralescu, A. (2023). *Fuzzy Sets Methods in Image Processing and Understanding: Medical Imaging Applications*. Springer International Publishing, Cham. 64

[Boecking et al., 2022] Boecking, B., Usuyama, N., Bannur, S., Castro, D. C., Schwaighofer, A., Hyland, S., Wetscherek, M., Naumann, T., Nori, A., Alvarez-Valle, J., Poon, H., and Oktay, O. (2022). Making the Most of Text Semantics to Improve Biomedical Vision–Language Processing. In Avidan, S., Brostow, G., Cissé, M., Farinella, G. M., and Hassner, T., editors, *Computer Vision – ECCV 2022*, pages 1–21, Cham. Springer Nature Switzerland. 72

[Brett et al., 2001] Brett, M., Leff, A. P., Rorden, C., and Ashburner, J. (2001). Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking. *NeuroImage*, 14(2):486–500. 44

[Bullock et al., 2022] Bullock, D. N., Hayday, E. A., Grier, M. D., Tang, W., Pestilli, F., and Heilbronner, S. R. (2022). A taxonomy of the brain's white matter: twenty-one major tracts for the 21st century. *Cerebral Cortex*, 32(20):4524–4548. 61

[Burgess et al., 2017] Burgess, C. P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., and Lerchner, A. (2017). Understanding disentangling in beta-VAE. In *NIPS Workshop on Learning Disentangled Representations*. 38

[Bycroft et al., 2018] Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., and Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209. Publisher: Nature Publishing Group. 8, 9

[Bürgel et al., 2006] Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M., and Zilles, K. (2006). White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. *NeuroImage*, 29(4):1092–1105. 61

[Campanella et al., 2019] Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., Brogi, E., Reuter, V. E., Klimstra, D. S., and Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, 25(8):1301–1309. 76

[Caron et al., 2018] Caron, M., Bojanowski, P., Joulin, A., and Douze, M. (2018). Deep Clustering for Unsupervised Learning of Visual Features. In *ECCV*. 24, 27

[Caron et al., 2020] Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., and Joulin, A. (2020). Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. In *Advances in Neural Information Processing Systems*, volume 33. 21, 24, 27

[Caron et al., 2021] Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A. (2021). Emerging Properties in Self-Supervised Vision Transformers. In *ICCV*. 76

[Carton et al., 2024] Carton, F., Louiset, R., and Gori, P. (2024). Double InfoGAN for Contrastive Analysis. In *International Conference on Artificial Intelligence and Statistics (AISTATS),*. 7, 29, 37, 38

[Catani et al., 2002] Catani, M., Howard, R., Pajevic, S., and Jones, D. K. (2002). Virtual in Vivo Interactive Dissection of White Matter Fasciculi in the Human Brain. *NeuroImage*, 17(1):77–94. 60

[Catani and Thiebautdeschotten, 2008] Catani, M. and Thiebautdeschotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, 44(8):1105–1132. 66

[Charlier et al., 2017] Charlier, B., Charon, N., and Trouvé, A. (2017). The Fshape Framework for the Variability Analysis of Functional Shapes. *Foundations of Computational Mathematics*, 17(2):287–357. 73

[Charon et al., 2020] Charon, N., Charlier, B., Glaunès, J., Gori, P., and Roussillon, P. (2020). Fidelity metrics between curves and surfaces: currents, varifolds, and normal cycles. In Pennec, X., Sommer, S., and Fletcher, T., editors, *Riemannian Geometric Statistics in Medical Image Analysis*, pages 441–477. Academic Press. 73

[Charon and Trouvé, 2014] Charon, N. and Trouvé, A. (2014). Functional Currents: A New Mathematical Tool to Model and Analyse Functional Shapes. *Journal of Mathematical Imaging and Vision*, 48(3):413–431. 73

[Chen et al., 2019] Chen, C., Dou, Q., Jin, Y., Chen, H., Qin, J., and Heng, P.-A. (2019). Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement and Gated Fusion. In *MICCAI*, volume LNCS 11766, pages 447–456, Cham. Springer. 45

[Chen et al., 2022] Chen, C., Dou, Q., Jin, Y., Liu, Q., and Heng, P. A. (2022). Learning With Privileged Multimodal Knowledge for Unimodal Segmentation. *IEEE Transactions on Medical Imaging*, 41(3):621–632. 48

[Chen et al., 2024] Chen, R. J., Ding, T., Lu, M. Y., Williamson, D. F. K., Jaume, G., Song, A. H., Chen, B., Zhang, A., Shao, D., Shaban, M., Williams, M., Oldenburg, L., Weishaupt, L. L., Wang, J. J., Vaidya, A., Le, L. P., Gerber, G., Sahai, S., Williams, W., and Mahmood, F. (2024). Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30(3):850–862. Publisher: Nature Publishing Group. 77

[Chen et al., 2018] Chen, R. T. Q., Li, X., Grosse, R. B., and Duvenaud, D. K. (2018). Isolating Sources of Disentanglement in Variational Autoencoders. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc. 38

[Chen et al., 2020] Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR. 9, 12, 17, 18, 19, 20, 21, 24, 27

[Chen et al., 2016] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., and Abbeel, P. (2016). InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc. 29, 32, 33, 38

[Chen and He, 2021] Chen, X. and He, K. (2021). Exploring Simple Siamese Representation Learning. In *CVPR*. 35

[Chen et al., 2017] Chen, X., Kingma, D. P., Salimans, T., Duan, Y., Dhariwal, P., Schulman, J., Sutskever, I., and Abbeel, P. (2017). Variational Lossy Autoencoder. In *ICLR*. 29

[Chen et al., 2023] Chen, Y., Zhang, C., Xue, T., Song, Y., Makris, N., Rathi, Y., Cai, W., Zhang, F., and O'Donnell, L. J. (2023). Deep fiber clustering: Anatomically informed fiber clustering with self-supervised deep learning for fast and effective tractography parcellation. *NeuroImage*, 273:120086. 62, 71

[Cheplygina et al., 2019] Cheplygina, V., de Bruijne, M., and Pluim, J. P. W. (2019). Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis*, 54:280–296. 76

[Chizat et al., 2018] Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2018). An interpolating distance between optimal transport and Fisher–Rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44. Publisher: Springer. 69

[Cho and Kang, 2022] Cho, Y. and Kang, S. (2022). Class Attention Transfer for Semantic Segmentation. In *2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pages 41–45. 48

[Chopra et al., 2005] Chopra, S., Hadsell, R., and LeCun, Y. (2005). Learning a Similarity Metric Discriminatively, with Application to Face Verification. In *CVPR*. 12

[Chuang et al., 2020] Chuang, C.-Y., Robinson, J., Lin, Y.-C., Torralba, A., and Jegelka, S. (2020). Debiased Contrastive Learning. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 8765–8775. Curran Associates, Inc. 21

[Ciccarelli et al., 2008] Ciccarelli, O., Catani, M., Johansen-Berg, H., Clark, C., and Thompson, A. (2008). Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. *The Lancet Neurology*, 7(8):715–727. 59

[Cuturi, 2013] Cuturi, M. (2013). Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc. 69

[Dalca et al., 2019] Dalca, A. V., Balakrishnan, G., Guttag, J., and Sabuncu, M. R. (2019). Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis*, 57:226–236. Publisher: Elsevier. 43, 52

[Davis et al., 2009] Davis, S. W., Dennis, N. A., Buchler, N. G., White, L. E., Madden, D. J., and Cabeza, R. (2009). Assessing the effects of age on long white matter tracts using diffusion tensor tractography. *NeuroImage*, 46(2):530–541. 59

[Dejerine, 1895] Dejerine, J. (1895). *Anatomie des Centres Nerveux*. Rueff et Cie, Paris. 60

[Delmonte et al., 2019] Delmonte, A., Mercier, C., Pallud, J., Bloch, I., and Gori, P. (2019). White Matter Multi-Resolution Segmentation Using Fuzzy Set Theory. In *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, pages 459–462. ISSN: 1945-8452. 59, 60, 62, 63, 66

[Delmonte et al., 2023] Delmonte, A., Sarnacki, S., Bloch, I., Gori, P., Muller, C., Virzi, A., and Barbera, G. L. (2023). Automatic generation of 3D anatomical models. 77

[Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee. 5

[Detlefsen et al., 2018] Detlefsen, N. S., Freifeld, O., and Hauberg, S. (2018). Deep Diffeomorphic Transformer Networks. In *CVPR*, pages 4403–4412. IEEE. 43

[Devlin et al., 2019] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs]. 72

[Dewey et al., 2019] Dewey, B. E., Zhao, C., Reinhold, J. C., Carass, A., Fitzgerald, K. C., Sotirchos, E. S., Saidha, S., Oh, J., Pham, D. L., Calabresi, P. A., van Zijl, P. C. M., and Prince, J. L. (2019). DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. *Magnetic Resonance Imaging*, 64:160–170. 10

[Dhariwal and Nichol, 2021] Dhariwal, P. and Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794. Curran Associates, Inc. 17

[Di Girolamo, 2019] Di Girolamo, A. (2019). Reproducible segmentation of white matter tractograms using artificial intelligence and spatial fuzzy sets. Technical report, Politecnico di Torino. 66

[Di Martino et al., 2014] Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D. A., Gallagher, L., Kennedy, D. P., Keown, C. L., Keysers, C., Lainhart, J. E., Lord, C., Luna, B., Menon, V., Minshew, N. J., Monk, C. S., Mueller, S., Müller, R.-A., Nebel, M. B., Nigg, J. T., O'Hearn, K., Pelphrey, K. A., Peltier, S. J., Rudie, J. D., Sunaert, S., Thioux, M., Tyszka, J. M., Uddin, L. Q., Verhoeven, J. S., Wenderoth, N., Wiggins, J. L., Mostofsky, S. H., and Milham, M. P. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6):659–667. Publisher: Nature Publishing Group. 8

[Dinsdale et al., 2021] Dinsdale, N. K., Jenkinson, M., and Namburete, A. I. (2021). Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *NeuroImage*, 228:117689. Publisher: Elsevier. 10

[Doersch et al., 2015] Doersch, C., Gupta, A., and Efros, A. A. (2015). Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE international conference on computer vision*, pages 1422–1430. 9

[Donahue and Simonyan, 2019] Donahue, J. and Simonyan, K. (2019). Large Scale Adversarial Representation Learning. In *Advances in Neural Information Processing Systems*. 9, 20, 21

[Dorent et al., 2019] Dorent, R., Joutard, S., Modat, M., Ourselin, S., and Vercauteren, T. (2019). Hetero-Modal Variational Encoder-Decoder for Joint Modality Completion and Segmentation. In *MICCAI*, volume LNCS 11765, pages 74–82. Springer. 45, 47

[Dufumier et al., 2023] Dufumier, B., Barbano, C. A., Louiset, R., Duchesnay, E., and Gori, P. (2023). Integrating Prior Knowledge in Contrastive Learning with Kernel. In *International Conference on Machine Learning (ICML)*. 7, 12, 18, 19, 20, 21

[Dufumier et al., 2021a] Dufumier, B., Gori, P., Battaglia, I., Victor, J., Grigis, A., and Duchesnay, E. (2021a). Benchmarking CNN on 3D Anatomical Brain MRI: Architectures, Data Augmentation and Deep Ensemble Learning. 5, 8

[Dufumier et al., 2024] Dufumier, B., Gori, P., Petiton, S., Louiset, R., Mangin, J.-F., Grigis, A., and Duchesnay, E. (2024). Exploring the potential of representation and transfer learning for anatomical neuroimaging: Application to psychiatry. *NeuroImage*, page 120665. 7, 8, 16, 18

[Dufumier et al., 2021b] Dufumier, B., Gori, P., Victor, J., Grigis, A., and Duchesnay, E. (2021b). Conditional Alignment and Uniformity for Contrastive Learning with Continuous Proxy Labels. In *MedNeurIPS, Workshop NeurIPS*. 7, 16, 18, 22

[Dufumier et al., 2021c] Dufumier, B., Gori, P., Victor, J., Grigis, A., Wessa, M., Brambilla, P., Favre, P., Polosan, M., McDonald, C., Piguet, C. M., Phillips, M., Eyler, L., and Duchesnay,

E. (2021c). Contrastive Learning with Continuous Proxy Meta-data for 3D MRI Classification. In de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., and Essert, C., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI*, Lecture Notes in Computer Science, pages 58–68, Cham. Springer International Publishing. 7, 12, 16, 17, 18, 20, 27

[Dufumier et al., 2022] Dufumier, B., Grigis, A., Victor, J., Ambroise, C., Frouin, V., and Duchesnay, E. (2022). OpenBHB: a Large-Scale Multi-Site Brain MRI Data-set for Age Prediction and Debiasing. *NeuroImage*, 263:119637. 8, 9, 18, 22

[Dumais et al., 2023] Dumais, F., Legarreta, J. H., Lemaire, C., Poulin, P., Rheault, F., Petit, L., Barakovic, M., Magon, S., Descoteaux, M., and Jodoin, P.-M. (2023). FIESTA: Autoencoders for accurate fiber segmentation in tractography. *NeuroImage*, 279:120288. 61

[Dupuis et al., 1998] Dupuis, P., Grenander, U., and Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, pages 587–600. Publisher: JSTOR. 43, 51

[Durrleman et al., 2011] Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., and Ayache, N. (2011). Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NeuroImage*, 55(3):1073–1090. 61

[Ebeling and von Cramon, 1992] Ebeling, U. and von Cramon, D. (1992). Topography of the uncinate fascicle and adjacent temporal fiber tracts. *Acta Neurochirurgica*, 115(3-4):143–148. 60, 66

[Elsayed et al., 2018] Elsayed, G. F., Krishnan, D., Mobahi, H., Regan, K., and Bengio, S. (2018). Large Margin Deep Networks for Classification. *Advances in Neural Information Processing Systems*, 2018-December:842–852. Publisher: Neural information processing systems foundation. 10

[Esmaeili et al., 2018] Esmaeili, M., Stensjazen, A. L., Berntsen, E. M., Solheim, O., and Reinertsen, I. (2018). The Direction of Tumour Growth in Glioblastoma Patients. *Scientific Reports*, 8(1):1199. 46

[Federici et al., 2020] Federici, M., Dutta, A., Forré, P., Kushman, N., and Akata, Z. (2020). Learning Robust Representations via Multi-View Information Bottleneck. In *International Conference on Learning Representations (ICLR)*. arXiv. 75

[Feydy et al., 2020] Feydy, J., Glaunès, A., Charlier, B., and Bronstein, M. (2020). Fast geometric learning with symbolic matrices. In *Advances in Neural Information Processing Systems*, volume 33, pages 14448–14462. Curran Associates, Inc. 69

[Feydy et al., 2019a] Feydy, J., Roussillon, P., Trouvé, A., and Gori, P. (2019a). Fast and Scalable Optimal Transport for Brain Tractograms. In Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., Zhou, S., Yap, P.-T., and Khan, A., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI*, Lecture Notes in Computer Science, pages 636–644. Springer International Publishing. 59, 62, 63, 69, 70

[Feydy et al., 2019b] Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-I., Trouvé, A., and Peyré, G. (2019b). Interpolating between Optimal Transport and MMD using Sinkhorn divergences. In *AiStats*. 69

[Feydy and Trouvé, 2018] Feydy, J. and Trouvé, A. (2018). Global divergences between measures: from Hausdorff distance to Optimal Transport. In *ShapeMI, MICCAI workshop*, pages 102–115. 69

[Fonov et al., 2011] Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., Group, B. D. C., and others (2011). Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 54(1):313–327. 46

[Fonov et al., 2009] Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., and Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (47). 46

[Fortin et al., 2018] Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., and others (2018). Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*, 167:104–120. Publisher: Elsevier. 10, 11, 22

[François et al., 2021] François, A., Gori, P., and Glaunès, J. (2021). Metamorphic Image Registration Using a Semi-lagrangian Scheme. In Nielsen, F. and Barbaresco, F., editors, *Geometric Science of Information*, Lecture Notes in Computer Science, pages 781–788, Cham. Springer International Publishing. 41, 44, 46, 52, 73

[François et al., 2022] François, A., Maillard, M., Oppenheim, C., Pallud, J., Bloch, I., Gori, P., and Glaunès, J. (2022). Weighted Metamorphosis for Registration of Images with Different Topologies. In Hering, A., Schnabel, J., Zhang, M., Ferrante, E., Heinrich, M., and Rueckert, D., editors, *Biomedical Image Registration (WBIR)*, Lecture Notes in Computer Science, pages 8–17, Cham. Springer International Publishing. 41, 46, 53, 55, 73

[Garland and Heckbert, 1997] Garland, M. and Heckbert, P. S. (1997). Surface Simplification Using Quadric Error Metrics. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 209–216, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co. 62

[Garyfallidis et al., 2012] Garyfallidis, E., Brett, M., Correia, M. M., Williams, G. B., and Nimmo-Smith, I. (2012). QuickBundles, a Method for Tractography Simplification. *Frontiers in Neuroscience*, 6. 60, 62, 64

[Garyfallidis et al., 2018] Garyfallidis, E., Côté, M.-A., Rheault, F., Sidhu, J., Hau, J., Petit, L., Fortin, D., Cunanne, S., and Descoteaux, M. (2018). Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage*, 170:283–295. 61

[Gaser and Dahnke, 2016] Gaser, C. and Dahnke, R. (2016). CAT-a computational anatomy toolbox for the analysis of structural MRI data. *HBM*, 2016:336–348. 27

[Geirhos et al., 2019] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., and Brendel, W. (2019). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In *International Conference on Learning Representations*. 10

[Ghazi et al., 2023] Ghazi, N., Aarabi, M. H., and Soltanian-Zadeh, H. (2023). Deep Learning Methods for Identification of White Matter Fiber Tracts: Review of State-of-the-Art and Future Prospective. *Neuroinformatics*, 21(3):517–548. 71

[Glickstein, 2006] Glickstein, M. (2006). Golgi and Cajal: The neuron doctrine and the 100th anniversary of the 1906 Nobel Prize. *Current biology*, 16(5):R147–151. 60

[Glocker et al., 2019] Glocker, B., Robinson, R., Castro, D. C., Dou, Q., and Konukoglu, E. (2019). Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects. *Medical Imaging meets NeurIPS Workshop*. 10, 11

[Goodfellow et al., 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*. 17, 32, 33

[Gooya et al., 2012] Gooya, A., Pohl, K. M., Bilello, M., Cirillo, L., Biros, G., Melhem, E. R., and Davatzikos, C. (2012). GLISTR: Glioma Image Segmentation and Registration. *IEEE TMI*, 31(10):1941–1954. 44, 57

[Gori et al., 2018] Gori, P., Colliot, O., Kacem, L. M., Worbe, Y., Routier, A., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2018). Double Diffeomorphism: Combining Morphometry and Structural Connectivity Analysis. *IEEE Transactions on Medical Imaging*, 37(9):2033–2043. Conference Name: IEEE Transactions on Medical Imaging. 60

[Gori et al., 2016] Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., De Vico Fallani, F., Chavez, M., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2016). Parsimonious Approximation of Streamline Trajectories in White Matter Fiber Bundles. *IEEE Transactions on Medical Imaging*, 35(12):2609–2619. Conference Name: IEEE Transactions on Medical Imaging. 60, 62, 65, 73

[Gori et al., 2017] Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2017). A Bayesian framework for joint morphometry of surface and curve meshes in multi-object complexes. *Medical Image Analysis*, 35:458–474. 43, 44, 73

[Gori et al., 2015] Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., Routier, A., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2015). Joint Morphometry of Fiber Tracts and Gray Matter Structures Using Double Diffeomorphisms. In Ourselin, S., Alexander, D. C., Westin, C.-F., and Cardoso, M. J., editors, *Information Processing in Medical Imaging (IPMI)*, Lecture Notes in Computer Science, pages 275–287. Springer International Publishing. 43, 73

[Gori et al., 2013] Gori, P., Colliot, O., Worbe, Y., Marrakchi-Kacem, L., Lecomte, S., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2013). Bayesian Atlas Estimation for the Variability Analysis of Shape Complexes. In Mori, K., Sakuma, I., Sato, Y., Barillot, C., and Navab, N., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pages 267–274. Springer Berlin Heidelberg. 43

[Graf et al., 2021] Graf, F., Hofer, C., Niethammer, M., and Kwitt, R. (2021). Dissecting supervised contrastive learning. In *International Conference on Machine Learning*, pages 3821–3830. PMLR. 10

[Grill et al., 2020] Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., Piot, B., kavukcuoglu, k., Munos, R., and Valko, M. (2020). Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 21271–21284. Curran Associates, Inc. 9, 21, 27

[Guan et al., 2022] Guan, Y., Zhang, J., Tian, K., Yang, S., Dong, P., Xiang, J., Yang, W., Huang, J., Zhang, Y., and Han, X. (2022). Node-aligned graph convolutional network for whole-slide image representation and classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18813–18823. 76

[Guevara et al., 2012] Guevara, P., Duclap, D., Poupon, C., Marrakchi-Kacem, L., Fillard, P., Le Bihan, D., Leboyer, M., Houenou, J., and Mangin, J. F. (2012). Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *NeuroImage*, 61(4):1083–1099. 61

[Guevara et al., 2011] Guevara, P., Poupon, C., Rivière, D., Cointepas, Y., Descoteaux, M., Thirion, B., and Mangin, J. F. (2011). Robust clustering of massive tractography datasets. *NeuroImage*, 54(3):1975–1993. 60, 61, 62

[Gurcan et al., 2009] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., and Yener, B. (2009). Histopathological image analysis: a review. *IEEE reviews in biomedical engineering*, 2:147–171. 76

[Hadsell et al., 2006] Hadsell, R., Chopra, S., and LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE. 12

[Han et al., 2017] Han, X., Yang, X., Aylward, S., Kwitt, R., and Niethammer, M. (2017). Efficient registration of pathological images: A joint PCA/image-reconstruction approach. In *IEEE 14th International Symposium onBiomedical Imaging (ISBI)*, pages 10–14. 44

[Hanif et al., 2017] Hanif, F., Muzaffar, K., Perveen, K., Malhi, S. M., and Simjee, S. U. (2017). Glioblastoma multiforme: A review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pac. J. Cancer Prev.*, 18(1):3–9. 42

[HaoChen et al., 2021] HaoChen, J. Z., Gaidon, A., Wei, C., and Ma, T. (2021). Provable Guarantees for Self-Supervised Deep Learning with Spectral Contrastive Loss. In *Advances in Neural Information Processing Systems*. 20

[Harrison et al., 2021] Harrison, J. E., Weber, S., Jakob, R., and Chute, C. G. (2021). ICD-11: an international classification of diseases for the twenty-first century. *BMC Medical Informatics and Decision Making*, 21(6):206. 10

[Havaei et al., 2016] Havaei, M., Guizard, N., Chapados, N., and Bengio, Y. (2016). HeMIS: Hetero-Modal Image Segmentation. In *MICCAI*, volume LNCS 9901, pages 469–477. Springer. 45, 47

[He et al., 2022] He, K., Chen, X., Xie, S., Li, Y., Dollar, P., and Girshick, R. (2022). Masked Autoencoders Are Scalable Vision Learners. In *CVPR*. 9

[He et al., 2020] He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738. 21, 24, 76

[Higgins et al., 2017] Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. (2017). beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations (ICLR)*. 38

[Hinton et al., 2015] Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the Knowledge in a Neural Network. *Deep Learning and Representation Learning Workshop: NIPS 2015*. 47

[Hjelm et al., 2019] Hjelm, R. D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., and Bengio, Y. (2019). Learning deep representations by mutual information estimation ... In *ICLR*. 34

[Ho et al., 2020] Ho, J., Jain, A., and Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc. 17

[Holm et al., 2009] Holm, D. D., Trouvé, A., and Younes, L. (2009). The Euler-Poincaré Theory of Metamorphosis. *Quarterly of Applied Mathematics*, 67(4):661–685. Publisher: Brown University. 44, 51

[Hong and Yang, 2021] Hong, Y. and Yang, E. (2021). Unbiased Classification through Bias-Contrastive and Bias-Balanced Learning. In *Advances in Neural Information Processing Systems*. 15, 16

[Hoppe, 1996] Hoppe, H. (1996). Progressive Meshes. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 99–108, New York, NY, USA. ACM. 72

[Horsfield and Jones, 2002] Horsfield, M. A. and Jones, D. K. (2002). Applications of diffusion-weighted and diffusion tensor MRI to white matter diseases – a review. *NMR in Biomedicine*, 15(7-8):570–577. 73

[Hou et al., 2016] Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E., and Saltz, J. H. (2016). Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2424–2433, Las Vegas, NV, USA. IEEE. 76

[Hozer et al., 2020] Hozer, F., Sarrazin, S., Laidi, C., Favre, P., Pauling, M., Cannon, D., McDonald, C., Emsell, L., Mangin, J.-F., Duchesnay, E., and others (2020). Lithium prevents grey matter atrophy in patients with bipolar disorder: an international multicenter study. *Psychological medicine*, pages 1–10. Publisher: Cambridge University Press. 20

[Hu et al., 2020] Hu, M., Maillard, M., Zhang, Y., Ciceri, T., La Barbera, G., Bloch, I., and Gori, P. (2020). Knowledge Distillation from Multi-modal to Mono-modal Segmentation Networks. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., and Joskowicz, L., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI*, Lecture Notes in Computer Science, pages 772–781, Cham. Springer International Publishing. 41, 46, 47

[Hua et al., 2008] Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C. M., and Mori, S. (2008). Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *NeuroImage*, 39(1):336–347. 61

[Huang et al., 2021] Huang, S.-C., Shen, L., Lungren, M. P., and Yeung, S. (2021). GLoRIA: A Multimodal Global-Local Representation Learning Framework for Label-efficient Medical Image

Recognition. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3922–3931, Montreal, QC, Canada. IEEE. 72, 75

[Huang and Gamal, 2024] Huang, T.-H. and Gamal, H. E. (2024). Efficient Solvers for Wyner Common Information with Application to Multi-Modal Clustering. arXiv:2402.14266 [cs, math]. 39

[Hyvarinen et al., 2019] Hyvarinen, A., Sasaki, H., and Turner, R. E. (2019). Nonlinear ICA Using Auxiliary Variables and Generalized Contrastive Learning. In *AISTATS*. 39

[Iftimovici, 2021] Iftimovici, A. (2021). *Analyses d'apprentissage supervisé appliquées en neuro-imagerie et épigénétique pour prédire la transition psychotique : vers un modèle neurodéveloppemental*. PhD thesis, Université Paris Cité. 11

[Ilse et al., 2018] Ilse, M., Tomczak, J., and Welling, M. (2018). Attention-based Deep Multiple Instance Learning. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2127–2136. PMLR. ISSN: 2640-3498. 76

[Insel and Quirion, 2005] Insel, T. R. and Quirion, R. (2005). Psychiatry as a Clinical Neuroscience Discipline. *JAMA : the journal of the American Medical Association*, 294(17):2221–2224. 11

[Isensee et al., 2021] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211. Publisher: Nature Publishing Group. 47, 48

[Isola et al., 2017] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, Honolulu, HI. IEEE. 34

[Jeurissen et al., 2019] Jeurissen, B., Descoteaux, M., Mori, S., and Leemans, A. (2019). Diffusion MRI fiber tractography of the brain. *NMR in Biomedicine*, 32(4):e3785. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/nbm.3785. 59

[Johnson et al., 2007] Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1):118–127. Publisher: Oxford University Press. 10

[Joshi et al., 2024] Joshi, A., Li, H., Parikh, N. A., and He, L. (2024). A systematic review of automated methods to perform white matter tract segmentation. *Frontiers in Neuroscience*, 18:1376570. 71

[Joshi et al., 2004] Joshi, S., Davis, B., Jomier, M., and Gerig, G. (2004). Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160. 43

[Kang et al., 2023] Kang, M., Song, H., Park, S., Yoo, D., and Pereira, S. (2023). Benchmarking Self-Supervised Learning on Diverse Pathology Datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3344–3354. 77

[Ke et al., 2023] Ke, G., Yu, Y., Chao, G., Wang, X., Xu, C., and He, S. (2023). Disentangling Multi-view Representations Beyond Inductive Bias. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, pages 2582–2590, New York, NY, USA. Association for Computing Machinery. 75

[Khemakhem et al., 2020] Khemakhem, I., Kingma, D., Monti, R., and Hyvarinen, A. (2020). Variational Autoencoders and Nonlinear ICA: A Unifying Framework. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 2207–2217. PMLR. 39

[Khosla et al., 2020] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., and Krishnan, D. (2020). Supervised Contrastive Learning. In *Advances in Neural Information Processing Systems*. 12, 13, 14, 27

[Kim et al., 2019] Kim, B., Kim, H., Kim, K., Kim, S., and Kim, J. (2019). Learning not to learn: Training deep neural networks with biased data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9012–9020. 10, 16

[Kim and Mnih, 2018] Kim, H. and Mnih, A. (2018). Disentangling by Factorising. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2649–2658. PMLR. 31, 38

[Kim et al., 2007] Kim, J., Avants, B., Patel, S., and Whyte, J. (2007). Spatial normalization of injured brains for neuroimaging research: An illustrative introduction of available options. *NCRRN Methodology Papers*. 44

[Kingma and Welling, 2014] Kingma, D. P. and Welling, M. (2014). Auto-encoding variational bayes. In *International Conference on Learning Representations*, pages 1–14. _eprint: 1312.6114. 17, 21, 27, 28

[Kips et al., 2022a] Kips, R., Bokaris, P.-A., Perrot, M., Gori, P., and Bloch, I. (2022a). Hair color digitization through imaging and deep inverse graphics. In *Electronic Imaging*, volume 34, pages 1–5. Society for Imaging Science and Technology. 6

[Kips et al., 2020] Kips, R., Gori, P., Perrot, M., and Bloch, I. (2020). CA-GAN: Weakly Supervised Color Aware GAN for Controllable Makeup Transfer. In Bartoli, A. and Fusiello, A., editors, *AIM20 (ECCV20 Workshop)*, Lecture Notes in Computer Science, pages 280–296, Cham. Springer International Publishing. 6

[Kips et al., 2022b] Kips, R., Jiang, R., Ba, S., Duke, B., Perrot, M., Gori, P., and Bloch, I. (2022b). Real-time Virtual-Try-On from a Single Example Image through Deep Inverse Graphics and Learned Differentiable Renderers. *Computer Graphics Forum*, 41(2):29–40. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14456. 6

[Kips et al., 2021] Kips, R., Jiang, R., Ba, S., Phung, E., Aarabi, P., Gori, P., Perrot, M., and Bloch, I. (2021). Deep Graphics Encoder for Real-Time Video Makeup Synthesis from Example. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3884–3888. ISSN: 2160-7516. 6

[Kivva et al., 2022] Kivva, B., Rajendran, G., Ravikumar, P., and Aragam, B. (2022). Identifiability of deep generative models without auxiliary information. In *NeurIPS*. 39

[Klingler and Gloor, 1960] Klingler, J. and Gloor, P. (1960). The connections of the amygdala and of the anterior temporal cortex in the human brain. *Journal of Comparative Neurology*, 115(3):333–369. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.901150305. 60

[Klösgen, 1996] Klösgen, W. (1996). Explora: a multipattern and multistrategy discovery assistant. In *Advances in knowledge discovery and data mining*, pages 249–271. American Association for Artificial Intelligence, USA. 24

[Konz et al., 2022] Konz, N., Gu, H., Dong, H., and Mazurowski, M. A. (2022). The Intrinsic Manifolds of Radiological Images and Their Role in Deep Learning. In Wang, L., Dou, Q., Fletcher, P. T., Speidel, S., and Li, S., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 684–694, Cham. Springer Nature Switzerland. 5

[Konz and Mazurowski, 2024] Konz, N. and Mazurowski, M. A. (2024). The Effect of Intrinsic Dataset Properties on Generalization: Unraveling Learning Differences Between Natural and Medical Images. In *International Conference on Learning Representations (ICLR)*. 5

[Krebs et al., 2019] Krebs, J., Delingette, H., Mailhé, B., Ayache, N., and Mansi, T. (2019). Learning a probabilistic model for diffeomorphic registration. *IEEE transactions on medical imaging*, 38(9):2165–2176. Publisher: IEEE. 43, 52

[Kushol et al., 2023] Kushol, R., Parnianpour, P., Wilman, A. H., Kalra, S., and Yang, Y.-H. (2023). Effects of MRI scanner manufacturers in classification tasks with deep learning models. *Scientific Reports*, 13(1):16791. Publisher: Nature Publishing Group. 9

[La Barbera et al., 2022] La Barbera, G., Boussaid, H., Maso, F., Sarnacki, S., Rouet, L., Gori, P., and Bloch, I. (2022). Anatomically constrained CT image translation for heterogeneous blood vessel segmentation. In *British Machine Vision Conference (BMVC)*. 56, 77

[La Barbera et al., 2023] La Barbera, G., Rouet, L., Boussaid, H., Lubet, A., Kassir, R., Sarnacki, S., Gori, P., and Bloch, I. (2023). Tubular structures segmentation of pediatric abdominal-visceral ceCT images with renal tumors: Assessment, comparison and improvement. *Medical Image Analysis*, 90:102986. 77

[La Barbera et al., 2021] La Barbera, G. L., Gori, P., Boussaid, H., Belucci, B., Delmonte, A., Goulin, J., Sarnacki, S., Rouet, L., and Bloch, I. (2021). Automatic Size And Pose Homogenization With Spatial Transformer Network To Improve And Accelerate Pediatric Segmentation. In *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1773–1776. ISSN: 1945-8452. 56, 77

[Lacroix et al., 2001] Lacroix, M., Abi-Said, D., Fourney, D. R., Gokaslan, Z. L., Shi, W., DeMonte, F., Lang, F. F., McCutcheon, I. E., Hassenbusch, S. J., Holland, E., and others (2001). A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *Journal of neurosurgery*, 95(2):190–198. Publisher: Journal of Neurosurgery Publishing Group. 41

[Lee et al., 1999] Lee, A. W. F., Dobkin, D., Sweldens, W., and Schröder, P. (1999). Multiresolution Mesh Morphing. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, pages 343–350, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co. 72

[Li et al., 2021a] Li, B., Li, Y., and Eliceiri, K. W. (2021a). Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328. 76

[Li et al., 2021b] Li, J., Zhou, P., Xiong, C., and Hoi, S. C. H. (2021b). Prototypical Contrastive Learning of Unsupervised Representations. In *ICLR*. 24, 27

[Li et al., 2022] Li, L., Donati, N., and Ovsjanikov, M. (2022). Learning Multi-resolution Functional Maps with Spectral Attention for Robust Shape Matching. In *Advances in Neural Information Processing Systems*. 73

[Li et al., 2023] Li, L., Wu, F., Wang, S., Luo, X., Martín-Isla, C., Zhai, S., Zhang, J., Liu, Y., Zhang, Z., Ankenbrand, M. J., Jiang, H., Zhang, X., Wang, L., Arega, T. W., Altunok, E., Zhao, Z., Li, F., Ma, J., Yang, X., Puybareau, E., Oksuz, I., Bricq, S., Li, W., Punithakumar, K., Tsaftaris, S. A., Schreiber, L. M., Yang, M., Liu, G., Xia, Y., Wang, G., Escalera, S., and Zhuang, X. (2023). MyoPS: A benchmark of myocardial pathology segmentation combining three-sequence cardiac magnetic resonance images. *Medical Image Analysis*, 87:102808. 48

[Li et al., 2018] Li, R., Yao, J., Zhu, X., Li, Y., and Huang, J. (2018). Graph CNN for Survival Analysis on Whole Slide Pathological Images. In Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C., and Fichtinger, G., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Lecture Notes in Computer Science, pages 174–182, Cham. Springer International Publishing. 76

[Li et al., 2024] Li, W., Yuille, A., and Zhou, Z. (2024). How well do supervised 3D models transfer to medical imaging tasks? In *International Conference on Learning Representations (ICLR)*. 8

[Li et al., 2021c] Li, Y., Yu, Q., Tan, M., Mei, J., Tang, P., Shen, W., Yuille, A., and xie, c. (2021c). Shape-Texture Debiased Neural Network Training. In *International Conference on Learning Representations*. 10

[Liang et al., 2023] Liang, P. P., Deng, Z., Ma, M. Q., Zou, J. Y., Morency, L.-P., and Salakhutdinov, R. (2023). Factorized Contrastive Learning: Going Beyond Multi-view Redundancy. In *Advances in Neural Information Processing Systems*, volume 36, pages 32971–32998. 75

[Liao et al., 2022] Liao, R., Yang, H.-T., Li, H., Liu, L.-X., Li, K., Li, J.-J., Liang, J., Hong, X.-P., Chen, Y.-L., and Liu, D.-Z. (2022). Recent Advances of Salivary Gland Biopsy in Sjögren's Syndrome. *Frontiers in Medicine*, 8. 76

[Lin et al., 2021] Lin, Y., Gou, Y., Liu, Z., Li, B., Lv, J., and Peng, X. (2021). COMPLETER: Incomplete Multi-view Clustering via Contrastive Prediction. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11169–11178, Nashville, TN, USA. IEEE. 75

[Lin et al., 2020] Lin, Z., Thekumparampil, K., Fanti, G., and Oh, S. (2020). InfoGAN-CR and ModelCentrality: Self-supervised Model Training and Selection for Disentangling GANs. In *Proceedings of the 37th International Conference on Machine Learning*. 33, 39

[Littlejohns et al., 2020] Littlejohns, T. J., Holliday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J. D., Boultwood, C., Collins, R., Conroy, M. C., Crabtree, N., Doherty, N., Frangi, A. F., Harvey, N. C., Leeson, P., Miller, K. L., Neubauer, S., Petersen, S. E., Sellors, J., Sheard, S., Smith, S. M., Sudlow, C. L. M., Matthews, P. M., and Allen, N. E. (2020). The UK Biobank imaging enhancement of 100,000 participants:. *Nature Communications*. 38

[Liu and Yap, 2024] Liu, S. and Yap, P.-T. (2024). Learning multi-site harmonization of magnetic resonance images without traveling human phantoms. *Communications Engineering*, 3(1):1–10. Publisher: Nature Publishing Group. 10

[Liu et al., 2015a] Liu, X., Niethammer, M., Kwitt, R., Singh, N., McCormick, M., and Aylward, S. (2015a). Low-Rank Atlas Image Analyses in the Presence of Pathologies. *IEEE TMI*, 34:2583–2591. 44

[Liu et al., 2019] Liu, Y., Chen, K., Liu, C., Qin, Z., Luo, Z., and Wang, J. (2019). Structured knowledge distillation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2604–2613. 47

[Liu et al., 2021] Liu, Y., Fan, Q., Zhang, S., Dong, H., Funkhouser, T., and Yi, L. (2021). Contrastive Multimodal Fusion with TupleInfoNCE. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 734–743, Montreal, QC, Canada. IEEE. 75

[Liu et al., 2015b] Liu, Z., Luo, P., Wang, X., and Tang, X. (2015b). Deep Learning Face Attributes in the Wild. In *Proceedings of International Conference on Computer Vision (ICCV)*. 36

[Locatello et al., 2019] Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., and Bachem, O. (2019). Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations. In *Proceedings of the 36th International Conference on Machine Learning*, pages 4114–4124. PMLR. 38

[Locatello et al., 2020a] Locatello, F., Poole, B., Rätsch, G., Schölkopf, B., Bachem, O., and Tschannen, M. (2020a). Weakly-Supervised Disentanglement Without Compromises. In *ICML*. 38

[Locatello et al., 2020b] Locatello, F., Tschannen, M., Bauer, S., Rätsch, G., Schölkopf, B., and Bachem, O. (2020b). Disentangling Factors of Variation Using Few Labels. In *ICLR*. 38

[Lopez-Paz et al., 2016] Lopez-Paz, D., Bottou, L., Schölkopf, B., and Vapnik, V. (2016). Unifying distillation and privileged information. In *ICLR*. 47

[Lorenzi et al., 2013] Lorenzi, M., Ayache, N., Frisoni, G., and Pennec, X. (2013). LCC-Demons: A robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage*, 81:470–483. 43

[Louiset, 2024] Louiset, R. (2024). *Learning pathological representations in neuroimaging:predicting psychiatric diagnosis by integrating heterogeneity constraints.* PhD thesis, Université Paris-Saclay. 25, 28

[Louiset et al., 2024a] Louiset, R., Duchesnay, E., Grigis, A., Dufumier, B., and Gori, P. (2024a). SepVAE: a contrastive VAE to separate pathological patterns from healthy ones. In *Medical Imaging with Deep Learning (MIDL)*. 7, 29, 30, 31, 36, 37

[Louiset et al., 2024b] Louiset, R., Duchesnay, E., Grigis, A., and Gori, P. (2024b). Separating common from salient patterns with Contrastive Representation Learning. In *International Conference on Learning Representations (ICLR)*. 7, 29, 34, 36

[Louiset et al., 2021] Louiset, R., Gori, P., Dufumier, B., Houenou, J., Grigis, A., and Duchesnay, E. (2021). UCSL : A Machine Learning Expectation-Maximization Framework for Unsupervised Clustering Driven by Supervised Learning. In Oliver, N., Pérez-Cruz, F., Kramer, S., Read, J., and Lozano, J. A., editors, *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*, Lecture Notes in Computer Science, pages 755–771. Springer International Publishing. 7, 25, 26, 27, 28

[Lu et al., 2024] Lu, M. Y., Chen, B., Williamson, D. F. K., Chen, R. J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L. P., Gerber, G., Parwani, A. V., Zhang, A., and Mahmood, F. (2024). A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874. Publisher: Nature Publishing Group. 72

[Lu et al., 2021] Lu, M. Y., Williamson, D. F., Chen, T. Y., Chen, R. J., Barbieri, M., and Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570. Publisher: Nature Publishing Group UK London. 76

[Ma et al., 2024] Ma, J., He, Y., Li, F., Han, L., You, C., and Wang, B. (2024). Segment anything in medical images. *Nature Communications*, 15(1):654. Publisher: Nature Publishing Group. 75

[Maddah et al., 2011] Maddah, M., Miller, J. V., Sullivan, E. V., Pfefferbaum, A., and Rohlfing, T. (2011). Sheet-Like White Matter Fiber Tracts: Representation, Clustering, and Quantitative Analysis. In Fichtinger, G., Martel, A., and Peters, T., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, number 6892 in Lecture Notes in Computer Science, pages 191–199. Springer Berlin Heidelberg. 61

[Maddah et al., 2007] Maddah, M., Wells, W. M., Warfield, S. K., Westin, C.-F., and Grimson, W. E. L. (2007). Probabilistic Clustering and Quantitative Analysis of White Matter Fiber Tracts. *Information Processing in Medical Imaging*, 20:372–383. 60, 61, 62

[Maillard, 2023] Maillard, M. (2023). *Towards the generation of glioblastoma atlases with deep learning methods : Tumor segmentation and metamorphic image registration.* These de doctorat, Institut polytechnique de Paris. 48, 52, 54, 55, 57

[Maillard et al., 2022] Maillard, M., François, A., Glaunès, J., Bloch, I., and Gori, P. (2022). A Deep Residual Learning Implementation of Metamorphosis. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. ISSN: 1945-8452. 41, 46, 54, 55, 73

[Mammadov et al., 2024] Mammadov, A., Folgoc, L. L., Adam, J., Buronfosse, A., Hayem, G., Hocquet, G., and Gori, P. (2024). Self-Supervision Revives Simple Multiple Instance Classification Methods in Pathology. In *Medical Imaging with Deep Learning - Short Paper*. 77

[Mang et al., 2020] Mang, A., Bakas, S., Subramanian, S., Davatzikos, C., and Biros, G. (2020). Integrated Biophysical Modeling and Image Analysis: Application to Neuro-Oncology. *Annual Review of Biomedical Engineering*, 22(Volume 22, 2020):309–341. Publisher: Annual Reviews. 57

[Manson and Schaefer, 2011] Manson, J. and Schaefer, S. (2011). Hierarchical Deformation of Locally Rigid Meshes. *Computer Graphics Forum*, 30(8):2387–2396. 72

[Marquand et al., 2019] Marquand, A. F., Kia, S. M., Zabihi, M., Wolfers, T., Buitelaar, J. K., and Beckmann, C. F. (2019). Conceptualizing mental disorders as deviations from normative functioning. *Molecular psychiatry*, 24(10):1415–1424. Publisher: Nature Publishing Group. 11

[Marquand et al., 2016] Marquand, A. F., Rezek, I., Buitelaar, J., and Beckmann, C. F. (2016). Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biological Psychiatry*, 80(7):552–561. 11

[Marzi et al., 2024] Marzi, C., Giannelli, M., Barucci, A., Tessa, C., Mascalchi, M., and Diciotti, S. (2024). Efficacy of MRI data harmonization in the age of machine learning: a multicenter study across 36 datasets. *Scientific Data*, 11(1):115. Publisher: Nature Publishing Group. 10

[Matsoukas et al., 2022] Matsoukas, C., Haslum, J. F., Sorkhei, M., Söderberg, M., and Smith, K. (2022). What Makes Transfer Learning Work for Medical Images. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9215–9224. 5, 8

[Matthey et al., 2017] Matthey, L., Higgins, I., Hassabis, D., and Lerchner, A. (2017). dSprites: Disentanglement testing Sprites dataset. 38

[Menze and others, 2015] Menze, B. H. and others (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024. 42, 43, 47

[Mercier et al., 2018] Mercier, C., Gori, P., Rohmer, D., Cani, M.-P., Boubekeur, T., Thiery, J.-M., and Bloch, I. (2018). Progressive and Efficient Multi-Resolution Representations for Brain Tractograms. In *Eurographics Workshop VCBM*. Accepted: 2018-09-19T15:19:28Z ISSN: 2070-5786. 59, 62, 65

[Miller et al., 2006] Miller, M. I., Trouvé, A., and Younes, L. (2006). Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24(2):209–228. Publisher: Springer. 52

[Mok and Chung, 2020] Mok, T. C. and Chung, A. (2020). Fast symmetric diffeomorphic image registration with convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4644–4653. 43

[Moor et al., 2023] Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., and Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956):259–265. Publisher: Nature Publishing Group. 75

[Moyer et al., 2020] Moyer, D., Ver Steeg, G., Tax, C. M. W., and Thompson, P. M. (2020). Scanner invariant representations for diffusion MRI harmonization. *Magnetic Resonance in Medicine*, 84(4):2174–2189. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.28243. 10

[Mrugala, 2013] Mrugala, M. M. (2013). Advances and challenges in the treatment of glioblastoma: a clinician's perspective. *Discovery medicine*, 15(83):221–230. 42

[Muller et al., 2019] Muller, C. O., Mille, E., Virzi, A., Marret, J.-B., Peyrot, Q., Delmonte, A., Berteloot, L., Gori, P., Blanc, T., Grevent, D., Boddaert, N., Bloch, I., and Sarnacki, S. (2019). Integrating tractography in pelvic surgery: a proof of concept. *Journal of Pediatric Surgery Case Reports*, 48:101268. 72

[Mustafa et al., 2021] Mustafa, B., Loh, A., Freyberg, J., MacWilliams, P., Wilson, M., McKinney, S. M., Sieniek, M., Winkens, J., Liu, Y., Bui, P., Prabhakara, S., Telang, U., Karthikesalingam, A., Houlsby, N., and Natarajan, V. (2021). Supervised Transfer Learning at Scale for Medical Imaging. 5, 8

[Nakada et al., 2011] Nakada, M., Kita, D., Watanabe, T., Hayashi, Y., Teng, L., Pyko, I. V., and Hamada, J.-I. (2011). Aberrant signaling pathways in glioma. *Cancers*, 3(3):3242–3278. Publisher: Molecular Diversity Preservation International (MDPI). 42

[Nam et al., 2020] Nam, J., Cha, H., Ahn, S., Lee, J., and Shin, J. (2020). Learning from Failure: Training Debiased Classifier from Biased Classifier. In *Advances in Neural Information Processing Systems*. 14

[Neyshabur et al., 2020] Neyshabur, B., Sedghi, H., and Zhang, C. (2020). What is being transferred in transfer learning? In *Advances in Neural Information Processing Systems*. 8

[Nichol and Dhariwal, 2021] Nichol, A. Q. and Dhariwal, P. (2021). Improved Denoising Diffusion Probabilistic Models. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8162–8171. PMLR. ISSN: 2640-3498. 17

[Niethammer et al., 2011] Niethammer, M., Hart, G. L., Pace, D. F., Vespa, P. M., Irimia, A., Horn, J. D. V., and Aylward, S. R. (2011). Geometric Metamorphosis. In *MICCAI*, volume LNCS 6892, pages 639–646. 44

[Nieuwenhuys et al., 2015] Nieuwenhuys, R., Broere, C. A. J., and Cerliani, L. (2015). A new myeloarchitectonic map of the human neocortex based on data from the Vogt–Vogt school. *Brain Structure and Function*, 220(5):2551–2573. 60

[O'Donnell and Westin, 2007] O'Donnell, L. and Westin, C.-F. (2007). Automatic Tractography Segmentation Using a High-Dimensional White Matter Atlas. *IEEE Transactions on Medical Imaging*, 26(11):1562–1575. 62

[O'Donnell et al., 2009] O'Donnell, L. J., Westin, C.-F., and Golby, A. J. (2009). Tract-based morphometry for white matter group analysis. *NeuroImage*, 45(3):832–844. 61

[Oord et al., 2018] Oord, A. v. d., Li, Y., and Vinyals, O. (2018). Representation Learning with Contrastive Predictive Coding. 12

[OpenAI, 2023] OpenAI (2023). GPT-4 Technical Report. Technical report, OpenAI. 72

[Orbes-Arteaga et al., 2018] Orbes-Arteaga, M., Cardoso, M. J., Sørensen, L., Modat, M., Ourselin, S., Nielsen, M., and Pai, A. (2018). Simultaneous synthesis of FLAIR and segmentation of white matter hypointensities from T1 MRIs. In *MIDL*. 45

[Ostrom et al., 2014] Ostrom, Q. T., Gittleman, H., Liao, P., Rouse, C., Chen, Y., Dowling, J., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J. (2014). CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. *Neuro-oncology*, 16(suppl_4):iv1–iv63. Publisher: Oxford University Press. 41

[Ouyang et al., 2023] Ouyang, J., Zhao, Q., Adeli, E., Peng, W., Zaharchuk, G., and Pohl, K. M. (2023). LSOR: Longitudinally-Consistent Self-Organized Representation Learning. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI*, pages 279–289, Cham. Springer Nature Switzerland. 38

[Ouyang et al., 2021a] Ouyang, J., Zhao, Q., Adeli, E., Sullivan, E. V., Pfefferbaum, A., Zaharchuk, G., and Pohl, K. M. (2021a). Self-supervised Longitudinal Neighbourhood Embedding. In de Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., and Essert, C., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 80–89, Cham. Springer International Publishing. 38

[Ouyang et al., 2022a] Ouyang, J., Zhao, Q., Adeli, E., Zaharchuk, G., and Pohl, K. M. (2022a). Disentangling Normal Aging From Severity of Disease via Weak Supervision on Longitudinal MRI. *IEEE Transactions on Medical Imaging*, 41(10):2558–2569. Conference Name: IEEE Transactions on Medical Imaging. 38

[Ouyang et al., 2022b] Ouyang, J., Zhao, Q., Adeli, E., Zaharchuk, G., and Pohl, K. M. (2022b). Self-supervised learning of neighborhood embedding for longitudinal MRI. *Medical Image Analysis*, 82:102571. 38

[Ouyang et al., 2021b] Ouyang, J., Zhao, Q., Sullivan, E. V., Pfefferbaum, A., Tapert, S. F., Adeli, E., and Pohl, K. M. (2021b). Longitudinal Pooling & Consistency Regularization to Model Disease Progression From MRIs. *IEEE Journal of Biomedical and Health Informatics*, 25(6):2082–2092. Conference Name: IEEE Journal of Biomedical and Health Informatics. 38

[Ovsjanikov et al., 2012] Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., and Guibas, L. (2012). Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics*, 31(4):30:1–30:11. 73

[O'Donnell et al., 2016] O'Donnell, L. J., Suter, Y., Rigolo, L., Kahali, P., Zhang, F., Norton, I., Albi, A., Olubiyi, O., Meola, A., Essayed, W. I., Unadkat, P., Ciris, P. A., Wells, W. M., Rathi, Y., Westin, C.-F., and Golby, A. J. (2016). Automated white matter fiber tract identification in patients with brain tumors. *NeuroImage : Clinical*, 13:138–153. 61

[Pallud et al., 2015] Pallud, J., Audureau, E., Noel, G., Corns, R., Lechapt-Zalcman, E., Duntze, J., Pavlov, V., Guyotat, J., Hieu, P. D., Le Reste, P.-J., Faillot, T., Litre, C.-F., Desse, N., Petit, A., Emery, E., Voirin, J., Peltier, J., Caire, F., Vignes, J.-R., Barat, J.-L., Langlois, O., Dezamis, E., Parraga, E., Zanello, M., Nader, E., Lefranc, M., Bauchet, L., Devaux, B., Menei, P., Metellus, P., Neurochirurgie, C. d. N.-O. o. t. S. F. d., Lahoud, G. A., Andreiuolo, F., Borha, A., Busson, A., Capelle, L., Chapon, F., Chassoux, F., Catry-Thomas, I., Chrétien, F., Colin, P., Czorny, A., Derlon, J.-M., Diebold, M.-D., Duffau, H., Edjlali-Goujon, M., Eskandari, J., Fustier, A., Gantois, C., Gadan, R., Geffrelot, J., Gimbert, E., Godard, J., Godon-Hardy, S., Gueye, M., Guillamo, J.-S., Heil, N., Hoffmann, D., Jovenin, N., Kalamarides, M., Katranji, H., Khouri, S., Koziak, M., Landré, E., Leon, V., Liguoro, D., Mandonnet, E., Mann, M., Méary, E., Meder, J.-F., Mellerio, C., Michalak, S., Miquel, C., Mokhtari, K., Monteil, P., Naggara, O., Nataf, F., Oppenheim, C., Quintin-Roue, I., Page, P., Paquis, P., Pedenon, D., Peruzzi, P., Riem, T., Rigau, V., Rigaux-Viodé, O., Rougier, A., Roux, F.-X., Salon, C., Théret, E., Turak, B., Trystram, D., Vandenbos, F., Varlet, P., Viennet, G., and Vital, Anne, C. d. N.-O. o. t. S. F. d. N. (2015). Long-term results of carmustine wafer implantation for newly diagnosed glioblastomas: a controlled propensity-matched analysis of a French multicenter cohort. *Neuro-Oncology*, 17(12):1609–1619. _eprint: https://academic.oup.com/neuro-oncology/article-pdf/17/12/1609/7639841/nov126.pdf. 41

[Pallud et al., 2013] Pallud, J., Capelle, L., and Huberfeld, G. (2013). Tumoral epileptogenicity: How does it happen? *Epilepsia*, 54(s9):30–34. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.12440. 46

[Pappas et al., 2021] Pappas, I., Hector, H., Haws, K., Curran, B., Kayser, A. S., and D'Esposito, M. (2021). Improved normalization of lesioned brains via cohort-specific templates. *Hum Brain Mapp*, 42(13):4187–4204. 44

[Parisot et al., 2016] Parisot, S., Darlix, A., Baumann, C., Zouaoui, S., Yordanova, Y., Blonski, M., Rigau, V., Chemouny, S., Taillandier, L., Bauchet, L., and others (2016). A probabilistic atlas of diffuse WHO grade II glioma locations in the brain. *PloS one*, 11(1):e0144200. Publisher: Public Library of Science San Francisco, CA USA. 42, 46

[Parzen, 1962] Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076. 35

[Pati et al., 2021] Pati, S., Sharma, V., Aslam, H., Thakur, S. P., Akbari, H., Mang, A., Subramanian, S., Biros, G., Davatzikos, C., and Bakas, S. (2021). Estimating Glioblastoma Biophysical Growth Parameters Using Deep Learning Regression. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes (Workshop)*, volume 12658, pages 157–167. 57

[Peng et al., 2021] Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A., and Smith, S. M. (2021). Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, 68:101871. Publisher: Elsevier. 8

[Petersen et al., 2010] Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., and Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neurology*, 74(3):201–209. Publisher: Wolters Kluwer. 38

[Peyré et al., 2019] Peyré, G., Cuturi, M., and others (2019). Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607. Publisher: Now Publishers, Inc. 69

[Phuong et al., 2018] Phuong, M., Welling, M., Kushman, N., Tomioka, R., and Nowozin, S. (2018). The Mutual Autoencoder: Controlling Information in Latent Code Representations. In *Arxiv*. 29

[Pielawski et al., 2020] Pielawski, N., Wetzer, E., Öfverstedt, J., Lu, J., Wählby, C., Lindblad, J., and Sladoje, N. (2020). CoMIR: Contrastive Multimodal Image Representation for Registration. In *Advances in Neural Information Processing Systems*, volume 33, pages 18433–18444. Curran Associates, Inc. 75

[Poole et al., 2019] Poole, B., Ozair, S., Oord, A. V. D., Alemi, A., and Tucker, G. (2019). On Variational Bounds of Mutual Information. In *ICML*. 12, 13

[Poulin et al., 2022] Poulin, P., Theaud, G., Rheault, F., St-Onge, E., Bore, A., Renauld, E., de Beaumont, L., Guay, S., Jodoin, P.-M., and Descoteaux, M. (2022). TractoInferno - A large-scale, open-source, multi-site database for machine learning dMRI tractography. *Scientific Data*, 9(1):725. Publisher: Nature Publishing Group. 71

[Qin et al., 2021] Qin, D., Bu, J.-J., Liu, Z., Shen, X., Zhou, S., Gu, J.-J., Wang, Z.-H., Wu, L., and Dai, H.-F. (2021). Efficient Medical Image Segmentation Based on Knowledge Distillation. *IEEE Transactions on Medical Imaging*, 40(12):3820–3831. 48

[Qiu et al., 2023] Qiu, J., Li, L., Sun, J., Peng, J., Shi, P., Zhang, R., Dong, Y., Lam, K., Lo, F. P.-W., Xiao, B., Yuan, W., Wang, N., Xu, D., and Lo, B. (2023). Large AI Models in Health Informatics: Applications, Challenges, and the Future. *IEEE Journal of Biomedical and Health Informatics*, 27(12):6074–6087. 75

[Qu et al., 2022] Qu, L., Liu, S., Liu, X., Wang, M., and Song, Z. (2022). Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis. *Physics in Medicine & Biology*. Publisher: IOP Publishing. 76

[Radford et al., 2021] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8748–8763. PMLR. ISSN: 2640-3498. 75

[Raghu et al., 2019] Raghu, M., Zhang, C., Kleinberg, J., and Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. In *Advances in neural information processing systems*, pages 3347–3357. 5, 8

[Ren et al., 2022] Ren, M., Dey, N., Styner, M. A., Botteron, K. N., and Gerig, G. (2022). Local Spatiotemporal Representation Learning for Longitudinally-consistent Neuroimage Analysis. In *Neural Information Processing Systems*. 38

[Richardson and Younes, 2016] Richardson, C. L. and Younes, L. (2016). Metamorphosis of images in reproducing kernel Hilbert spaces. *Adv Comput Math*, 42(3):573–603. 52

[Rock et al., 2012] Rock, K., McArdle, O., Forde, P., Dunne, M., Fitzpatrick, D., O'Neill, B., and Faul, C. (2012). A clinical review of treatment outcomes in glioblastoma multiforme—the validation in a non-trial population of the results of a randomised Phase III clinical trial: has a more radical approach improved survival? *The British journal of radiology*, 85(1017):e729–e733. Publisher: The British Institute of Radiology. 131–151 Great Titchfield Street, London 41

[Rohlfing et al., 2010] Rohlfing, T., Zahr, N. M., Sullivan, E. V., and Pfefferbaum, A. (2010). The SRI24 multichannel atlas of normal adult human brain structure. *Human brain mapping*, 31(5):798–819. Publisher: Wiley Online Library. 54

[Rombach et al., 2022] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685, New Orleans, LA, USA. IEEE. 17

[Ronneberger et al., 2015] Ronneberger, O., P.Fischer, and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer. 52

[Rosenblatt, 1956] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27(3):832–837. 35

[Rousseau et al., 2020] Rousseau, F., Drumetz, L., and Fablet, R. (2020). Residual Networks as Flows of Diffeomorphisms. *JMIV*, 62(3):365–375. 52

[Roussillon et al., 2019] Roussillon, P., Thiery, J.-M., Bloch, I., and Gori, P. (2019). Appariement difféomorphique robuste de faisceaux neuronaux. In *GRETSI 2019*, Lille, France. 73

[Roux et al., 2019] Roux, A., Roca, P., Edjlali, M., Sato, K., Zanello, M., Dezamis, E., Gori, P., Lion, S., Fleury, A., Dhermain, F., Meder, J.-F., Chrétien, F., Lechapt, E., Varlet, P., Oppenheim, C., and Pallud, J. (2019). MRI Atlas of IDH Wild-Type Supratentorial Glioblastoma: Probabilistic Maps of Phenotype, Management, and Outcomes. *Radiology*, 293(3):633–643. Publisher: Radiological Society of North America. 41, 46, 58

[Rueckert et al., 2006] Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V., and Hammers, A. (2006). Diffeomorphic Registration Using B-Splines. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, pages 702–709, Berlin, Heidelberg. Springer Berlin Heidelberg. 43

[Ruiz et al., 2019] Ruiz, A., Martinez, O., Binefa, X., and Verbeek, J. (2019). Learning Disentangled Representations with Reference-Based Variational Autoencoders. In *ICLR workshop on Learning from Limited Labeled Data*, pages 1–17, New Orleans, United States. 29

[Ruppli et al., 2023] Ruppli, C., Gori, P., Ardon, R., and Bloch, I. (2023). Decoupled conditional contrastive learning with variable metadata for prostate lesion detection. In *MILLanD workshop (MICCAI)*. 77

[Sadri et al., 2020] Sadri, A. R., Janowczyk, A., Zhou, R., Verma, R., Beig, N., Antunes, J., Madabhushi, A., Tiwari, P., and Viswanath, S. E. (2020). MRQy — An open-source tool for quality control of MR imaging data. *Medical Physics*, 47(12):6029–6038. 9

[Sagawa et al., 2019] Sagawa, S., Koh, P. W., Hashimoto, T. B., and Liang, P. (2019). Distributionally Robust Neural Networks. In *International Conference on Learning Representations*. 10

[Saliani et al., 2017] Saliani, A., Perraud, B., Duval, T., Stikov, N., Rossignol, S., and Cohen-Adad, J. (2017). Axon and Myelin Morphology in Animal and Human Spinal Cord. *Frontiers in Neuroanatomy*, 11. 59

[Sarrazin et al., 2018] Sarrazin, S., Cachia, A., Hozer, F., McDonald, C., Emsell, L., Cannon, D. M., Wessa, M., Linke, J., Versace, A., Hamdani, N., and others (2018). Neurodevelopmental subtypes of bipolar disorder are related to cortical folding patterns: An international multicenter study. *Bipolar disorders*, 20(8):721–732. Publisher: Wiley Online Library. 27

[Sarubbo et al., 2013] Sarubbo, S., De Benedictis, A., Maldonado, I. L., Basso, G., and Duffau, H. (2013). Frontal terminations for the inferior fronto-occipital fascicle: anatomical dissection, DTI study and functional considerations on a multi-component bundle. *Brain Structure and Function*, 218(1):21–37. 60, 66

[Saunshi et al., 2019] Saunshi, N., Plevrakis, O., Arora, S., Khodak, M., and Khandeparkar, H. (2019). A theoretical analysis of contrastive unsupervised representation learning. In *International Conference on Machine Learning*, pages 5628–5637. PMLR. 20

[Scheufele et al., 2019] Scheufele, K., Mang, A., Gholami, A., Davatzikos, C., Biros, G., and Mehl, M. (2019). Coupling brain-tumor biophysical models and diffeomorphic image registration. *Comput Methods Appl Mech Eng*, 347:533–567. 44, 57

[Scheufele et al., 2021] Scheufele, K., Subramanian, S., and Biros, G. (2021). Fully Automatic Calibration of Tumor-Growth Models Using a Single mpMRI Scan. *IEEE Transactions on Medical Imaging*, 40(1):193–204. 44

[Schmahmann and Pandya, 2006] Schmahmann, J. D. and Pandya, D. N. (2006). *Fiber Pathways of the Brain*. Oxford University Press. 60

[Schmahmann and Pandya, 2007] Schmahmann, J. D. and Pandya, D. N. (2007). Cerebral White Matter — Historical Evolution of Facts and Notions Concerning the Organization of the Fiber Pathways of the Brain. *Journal of the History of the Neurosciences*, 16(3):237–267. Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1080/09647040500495896. 60

[Schroff et al., 2015] Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. In *CVPR*. 12

[Schulz et al., 2024] Schulz, M.-A., Bzdok, D., Haufe, S., Haynes, J.-D., and Ritter, K. (2024). Performance reserves in brain-imaging-based phenotype prediction. *Cell Reports*, 43(1):113597. 8

[Schulz et al., 2020] Schulz, M.-A., Yeo, B. T., Vogelstein, J. T., Mourao-Miranada, J., Kather, J. N., Kording, K., Richards, B., and Bzdok, D. (2020). Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nature communications*, 11(1):1–15. Publisher: Nature Publishing Group. 8

[Sdika and Pelletier, 2009] Sdika, M. and Pelletier, D. (2009). Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Hum. Brain Mapp.*, 30(4). 44

[Shao et al., 2021] Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., and others (2021). Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147. 76

[Sharmin et al., 2018] Sharmin, N., Olivetti, E., and Avesani, P. (2018). White Matter Tract Segmentation as Multiple Linear Assignment Problems. *Frontiers in Neuroscience*, 11. 61

[Shen et al., 2019] Shen, Z., Vialard, F.-X., and Niethammer, M. (2019). Region-specific Diffeomorphic Metric Mapping. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc. 44

[Shu et al., 2020] Shu, R., Chen, Y., Kumar, A., Ermon, S., and Poole, B. (2020). Weakly Supervised Disentanglement with Guarantees. In *International Conference on Learning Representations (ICLR)*. 38

[Siless et al., 2018] Siless, V., Chang, K., Fischl, B., and Yendiki, A. (2018). AnatomiCuts: Hierarchical clustering of tractography streamlines based on anatomical similarity. *NeuroImage*, 166:32–45. 62

[Simpson et al., 1993] Simpson, J., Horton, J., Scott, C., Curran, W., Rubin, P., Fischbach, J., Isaacson, S., Rotman, M., Asbell, S., Nelson, J., and others (1993). Influence of location and extent of surgical resection on survival of patients with glioblastoma multiforme: results of three consecutive Radiation Therapy Oncology Group (RTOG) clinical trials. *International Journal of Radiation Oncology* Biology* Physics*, 26(2):239–244. Publisher: Elsevier. 42

[Smith, 2016] Smith, S. (2016). Linking cognition to brain connectivity. *Nature Neuroscience*, 19(1). 73

[Sohn, 2016] Sohn, K. (2016). Improved Deep Metric Learning with Multi-class N-pair Loss Objective. In *Advances in Neural Information Processing Systems*. 12, 13

[Song et al., 2013] Song, L., Fukumizu, K., and Gretton, A. (2013). Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models. *IEEE Signal Processing Magazine*, 30(4):98–111. Publisher: IEEE. 19

[Srinidhi et al., 2021] Srinidhi, C. L., Ciga, O., and Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67:101813. 76

[Stefanescu et al., 2004] Stefanescu, R., Commowick, O., Malandain, G., Bondiau, P.-Y., Ayache, N., and Pennec, X. (2004). Non-rigid Atlas to Subject Registration with Pathologies for Conformal Brain Radiotherapy. In *MICCAI*, volume LNCS 3216, pages 704–711. 44

[Stoer and Bulirsch, 2002] Stoer, J. and Bulirsch, R. (2002). *Introduction to Numerical Analysis*, volume 12 of *Texts in Applied Mathematics*. Springer. 52

[Stummer et al., 2006] Stummer, W., Pichlmeier, U., Meinel, T., Wiestler, O. D., Zanella, F., Reulen, H.-J., Group, A.-G. S., and others (2006). Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *The lancet oncology*, 7(5):392–401. Publisher: Elsevier. 41

[Subramanian et al., 2023] Subramanian, S., Ghafouri, A., Scheufele, K. M., Himthani, N., Davatzikos, C., and Biros, G. (2023). Ensemble Inversion for Brain Tumor Growth Models With Mass Effect. *IEEE Transactions on Medical Imaging*, 42(4):982–995. 57

[Sun et al., 2023] Sun, Y., Wang, F., Shu, J., Wang, H., Wang, L., Meng, D., and Lian, C. (2023). Dual Meta-Learning with Longitudinally Generalized Regularization for One-Shot Brain Tissue Segmentation Across the Human Lifespan. In *International Conference on Computer Vision (ICCV)*, pages 21061–21071. IEEE. 38

[Taleb et al., 2022] Taleb, A., Kirchler, M., Monti, R., and Lippert, C. (2022). ContIG: Self-supervised Multimodal Contrastive Learning for Medical Imaging with Genetics. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20876–20889, New Orleans, LA, USA. IEEE. 75

[Tamminga et al., 2014] Tamminga, C. A., Pearlson, G., Keshavan, M., Sweeney, J., Clementz, B., and Thaker, G. (2014). Bipolar and schizophrenia network for intermediate phenotypes: outcomes across the psychosis continuum. *Schizophrenia bulletin*, 40:S131–S137. 27, 36

[Tang et al., 2019] Tang, Z., Yap, P., and Shen, D. (2019). A New Multi-Atlas Registration Framework for Multimodal Pathological Images Using Conventional Monomodal Normal Atlases. *IEEE Transactions on Image Processing*, 28(5):2293–2304. 44

[Tartaglione et al., 2021] Tartaglione, E., Barbano, C. A., and Grangetto, M. (2021). EnD: Entangling and Disentangling deep representations for bias correction. In *CVPR*. 10, 16

[Thust et al., 2018] Thust, S., Heiland, S., Falini, A., Jäger, H. R., Waldman, A., Sundgren, P., Godi, C., Katsaros, V., Ramos, A., Bargallo, N., and others (2018). Glioma imaging in Europe: a survey of 220 centres and recommendations for best clinical practice. *European radiology*, 28(8):3306–3317. Publisher: Springer. 42

[Tian et al., 2020a] Tian, Y., Krishnan, D., and Isola, P. (2020a). Contrastive multiview coding. In *European Conference on Computer Vision*, pages 776–794. Springer. 75

[Tian et al., 2020b] Tian, Y., Krishnan, D., and Isola, P. (2020b). Contrastive Representation Distillation. In *International Conference on Learning Representations*. 21

[Tiu et al., 2022] Tiu, E., Talius, E., Patel, P., Langlotz, C. P., Ng, A. Y., and Rajpurkar, P. (2022). Expert-level detection of pathologies from unannotated chest X-ray images via self-supervised learning. *Nature Biomedical Engineering*, 6(12):1399–1406. Publisher: Nature Publishing Group. 72, 75

[Torralba and Efros, 2011] Torralba, A. and Efros, A. A. (2011). Unbiased look at dataset bias. In *CVPR*, pages 1521–1528. 10

[Tournier et al., 2011] Tournier, J.-D., Mori, S., and Leemans, A. (2011). Diffusion tensor imaging and beyond. *Magnetic Resonance in Medicine*, 65(6):1532–1556. 60

[Trosten et al., 2021] Trosten, D. J., Lokse, S., Jenssen, R., and Kampffmeyer, M. (2021). Reconsidering Representation Alignment for Multi-view Clustering. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1255–1265, Nashville, TN, USA. IEEE. 75

[Trouvé and Younès, 2005] Trouvé, A. and Younès, L. (2005). Metamorphoses Through Lie Group Action. *Foundations of Computational Mathematics*, 5(2):173–198. 44, 51

[Träuble et al., 2021] Träuble, F., Creager, E., Kilbertus, N., Locatello, F., Dittadi, A., Goyal, A., Schölkopf, B., and Bauer, S. (2021). On Disentangled Representations Learned from Correlated Data. In *Proceedings of the 38th International Conference on Machine Learning*, pages 10401–10412. PMLR. ISSN: 2640-3498. 39

[Tschannen et al., 2020] Tschannen, M., Djolonga, J., Rubenstein, P. K., Gelly, S., and Lucic, M. (2020). On Mutual Information Maximization for Representation Learning. In *International Conference on Learning Representations (ICLR)*. 35

[Tschuchnig et al., 2022] Tschuchnig, M. E., Grubmüller, P., Stangassinger, L. M., Kreutzer, C., Couillard-Després, S., Oostingh, G. J., Hittmair, A., and Gadermayr, M. (2022). Evaluation of Multi-Scale Multiple Instance Learning to Improve Thyroid Cancer Classification. In *Eleventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. 76

[Tuzikov et al., 2003] Tuzikov, A. V., Colliot, O., and Bloch, I. (2003). Evaluation of the symmetry plane in 3D MR brain images. *Pattern Recognition Letters*, 24(14):2219–2233. 65

[Tykocki and Eltayeb, 2018] Tykocki, T. and Eltayeb, M. (2018). Ten-year survival in glioblastoma. A systematic review. *Journal of Clinical Neuroscience*, 54:7–13. Publisher: Elsevier. 41

[Vaillant and Glaunès, 2005] Vaillant, M. and Glaunès, J. (2005). Surface Matching via Currents. In Christensen, G. E. and Sonka, M., editors, *Information Processing in Medical Imaging*, number 3565 in Lecture Notes in Computer Science, pages 381–392. Springer Berlin Heidelberg. 61

[Van der Maaten and Hinton, 2008] Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(11). 9

[Van Essen et al., 2013] Van Essen, D., Smith, S., Barch, D., Behrens, T., Yacoub, E., and Ugurbil, K. (2013). The WU-Minn Human Connectome Project: an overview. *NeuroImage*, 80. 8

[Van Gansbeke et al., 2020] Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., and Van Gool, L. (2020). Scan: Learning to classify images without labels. In *European Conference on Computer Vision*, pages 268–285. Springer. 24, 27

[van Tulder and de Bruijne, 2015] van Tulder, G. and de Bruijne, M. (2015). Why Does Synthesized Data Improve Multi-sequence Classification? In *MICCAI*, volume LNCS 9349, pages 531–538, Cham. Springer. 45

[Vapnik and Izmailov, 2015] Vapnik, V. and Izmailov, R. (2015). Learning Using Privileged Information: Similarity Control and Knowledge Transfer. *Journal of Machine Learning Research*, 16(61):2023–2049. 47

[Vercauteren et al., 2009] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1, Supplement 1):S61–S72. 43

[Vialard et al., 2011] Vialard, F.-X., Risser, L., Rueckert, D., and Cotter, C. J. (2011). Diffeomorphic 3D Image Registration via Geodesic Shooting Using an Efficient Adjoint Calculation. *International Journal of Computer Vision*, 97:229–241. 43, 52

[Virzì et al., 2018] Virzì, A., Gori, P., Muller, C. O., Mille, E., Peyrot, Q., Berteloot, L., Boddaert, N., Sarnacki, S., and Bloch, I. (2018). Segmentation of Pelvic Vessels in Pediatric MRI Using a Patch-Based Deep Learning Approach. In Melbourne, A., Licandro, R., DiFranco, M., Rota, P., Gau, M., Kampel, M., Aughwane, R., Moeskops, P., Schwartz, E., Robinson, E., and Makropoulos, A., editors, *PIPPI MICCAI Workshop*, Lecture Notes in Computer Science, pages 97–106. Springer International Publishing. 77

[Virzì et al., 2020] Virzì, A., Muller, C. O., Marret, J.-B., Mille, E., Berteloot, L., Grévent, D., Boddaert, N., Gori, P., Sarnacki, S., and Bloch, I. (2020). Comprehensive Review of 3D Segmentation Software Tools for MRI Usable for Pelvic Surgery Planning. *Journal of Digital Imaging*, 33(1):99–110. 77

[Vétil et al., 2024] Vétil, R., Abi-Nader, C., Bône, A., Rohé, M.-M., Gori, P., and Bloch, I. (2024). Method for characterizing an organ of a patient in a medical image. 77

[Vétil et al., 2022] Vétil, R., Bône, A., Vullierme, M.-P., Rohé, M.-M., Gori, P., and Bloch, I. (2022). Improving the Automatic Segmentation of Elongated Organs Using Geometrical Priors. In *IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. ISSN: 1945-8452. 77

[Wachinger et al., 2021] Wachinger, C., Rieckmann, A., Pölsterl, S., Initiative, A. D. N., and others (2021). Detect and correct bias in multi-site neuroimaging datasets. *Medical Image Analysis*, 67:101879. Publisher: Elsevier. 10, 11

[Wakana et al., 2007] Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., Hua, K., Zhang, J., Jiang, H., Dubey, P., Blitz, A., van Zijl, P., and Mori, S. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *NeuroImage*, 36(3):630–644. 61, 66

[Wakana et al., 2004] Wakana, S., Jiang, H., Nagae-Poetscher, L. M., van Zijl, P. C. M., and Mori, S. (2004). Fiber Tract–based Atlas of Human White Matter Anatomy. *Radiology*, 230(1):77–87. 60

[Wang et al., 2016a] Wang, B., Prastawa, M., Irimia, A., Saha, A., Liu, W., Goh, S. Y. M., Vespa, P. M., Horn, J. D. V., and Gerig, G. (2016a). Modeling 4D pathological changes by leveraging normative models. *Computer Vision and Image Understanding*, 151:3–13. 27

[Wang et al., 2014] Wang, J., Song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J., Chen, B., and Wu, Y. (2014). Learning Fine-Grained Image Similarity with Deep Ranking. In *CVPR*. 12

[Wang et al., 2016b] Wang, L., Alpert, K. I., Calhoun, V. D., Cobia, D. J., Keator, D. B., King, M. D., Kogan, A., Landis, D., Tallis, M., Turner, M. D., Potkin, S. G., Turner, J. A., and Ambite, J. L. (2016b). SchizConnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration. *NeuroImage*, 124:1155–1167. 36

[Wang and Isola, 2020] Wang, T. and Isola, P. (2020). Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*, pages 9929–9939. PMLR. 18, 21, 22, 35

[Wang et al., 2019] Wang, X., Hua, Y., Kodirov, E., and Robertson, N. M. (2019). Ranked List Loss for Deep Metric Learning. In *CVPR*. 12

[Wang et al., 2022a] Wang, Y., Zhang, Q., Wang, Y., Yang, J., and Lin, Z. (2022a). Chaos is a Ladder: A New Theoretical Understanding of Contrastive Learning via Augmentation Overlap. In *ICLR*. 20

[Wang et al., 2022b] Wang, Z., Wu, Z., Agarwal, D., and Sun, J. (2022b). MedCLIP: Contrastive Learning from Unpaired Medical Images and Text. In *EMNLP*. 75

[Wassermann et al., 2010] Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., and Deriche, R. (2010). Unsupervised white matter fiber clustering and tract probability map generation. *NeuroImage*, 51(1):228–241. 61, 69

[Wassermann et al., 2016] Wassermann, D., Makris, N., Rathi, Y., Shenton, M., Kikinis, R., Kubicki, M., and Westin, C.-F. (2016). The white matter query language: a novel approach for describing human white matter anatomy. *Brain Structure and Function*, 221(9):4705–4721. 62, 64

[Wasserthal et al., 2018] Wasserthal, J., Neher, P., and Maier-Hein, K. H. (2018). TractSeg - Fast and accurate white matter tract segmentation. *NeuroImage*, 183:239–253. 61, 71

[Wei et al., 2019] Wei, J. W., Tafe, L. J., Linnik, Y. A., Vaickus, L. J., Tomita, N., and Hassanpour, S. (2019). Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Scientific Reports*, 9(1):3358. Number: 1 Publisher: Nature Publishing Group. 76

[Weinberger et al., 2022] Weinberger, E., Beebe-Wang, N., and Lee, S.-I. (2022). Moment Matching Deep Contrastive Latent Variable Models. In *AISTATS*. arXiv. 29, 30, 31

[Weinberger and Saul, 2009] Weinberger, K. Q. and Saul, L. K. (2009). Distance Metric Learning for Large Margin Nearest Neighbor Classification. *JMLR*. 12

[Wolfers et al., 2018] Wolfers, T., Doan, N. T., Kaufmann, T., Alnæs, D., Moberget, T., Agartz, I., Buitelaar, J. K., Ueland, T., Melle, I., Franke, B., and others (2018). Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA psychiatry*, 75(11):1146–1155. Publisher: American Medical Association. 7

[Wu et al., 2023] Wu, C., Zhang, X., Zhang, Y., Wang, Y., and Xie, W. (2023). MedKLIP: Medical Knowledge Enhanced Language-Image Pre-Training for X-ray Diagnosis. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 21315–21326, Paris, France. IEEE. 72

[Wu and Goodman, 2018] Wu, M. and Goodman, N. (2018). Multimodal Generative Models for Scalable Weakly-Supervised Learning. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 5575–5585. Curran Associates, Inc. 45

[Wyner, 1975] Wyner, A. (1975). The common information of two dependent random variables. *IEEE Trans. Inform. Theory*, 21(2):163–179. 39

[Xie et al., 2018] Xie, J., Shuai, B., Hu, J.-F., Lin, J., and Zheng, W.-S. (2018). Improving Fast Segmentation With Teacher-student Learning. In *British Machine Vision Conference (BMVC)*. 47

[Xu et al., 2022] Xu, J., Tang, H., Ren, Y., Peng, L., Zhu, X., and He, L. (2022). Multi-level Feature Learning for Contrastive Multi-view Clustering. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16030–16039. IEEE. 75

[Xue et al., 2023] Xue, T., Zhang, F., Zhang, C., Chen, Y., Song, Y., Golby, A. J., Makris, N., Rathi, Y., Cai, W., and O'Donnell, L. J. (2023). Superficial white matter analysis: An efficient point-cloud-based deep learning framework with supervised contrastive learning for consistent tractography parcellation across populations and dMRI acquisitions. *Medical Image Analysis*, 85:102759. 71

[Yang et al., 2021] Yang, J., Angelini, E. D., Balte, P. P., Hoffman, E. A., Austin, J. H. M., Smith, B. M., Barr, R. G., and Laine, A. F. (2021). Novel Subtypes of Pulmonary Emphysema Based on Spatially-Informed Lung Texture Learning: The Multi-Ethnic Study of Atherosclerosis (MESA) COPD Study. *IEEE Transactions on Medical Imaging*, 40(12):3652–3662. Conference Name: IEEE Transactions on Medical Imaging. 24

[Yang et al., 2024] Yang, L., Xu, S., Sellergren, A., Kohlberger, T., Zhou, Y., Ktena, I., Kiraly, A., Ahmed, F., Hormozdiari, F., Jaroensri, T., Wang, E., Wulczyn, E., Jamil, F., Guidroz, T., Lau, C., Qiao, S., Liu, Y., Goel, A., Park, K., Agharwal, A., George, N., Wang, Y., Tanno, R., Barrett, D. G. T., Weng, W.-H., Mahdavi, S. S., Saab, K., Tu, T., Kalidindi, S. R., Etemadi, M., Cuadros, J., Sorensen, G., Matias, Y., Chou, K., Corrado, G., Barral, J., Shetty, S., Fleet, D., Eslami, S. M. A., Tse, D., Prabhakara, S., McLean, C., Steiner, D., Pilgrim, R., Kelly, C., Azizi, S., and Golden, D. (2024). Advancing Multimodal Medical Capabilities of Gemini. arXiv:2405.03162 [cs]. 75

[Yang et al., 2016] Yang, X., Han, X., Park, E., Aylward, S., Kwitt, R., and Niethammer, M. (2016). Registration of pathological images. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 97–107. Springer. 44

[Yao et al., 2024] Yao, D., Xu, D., Lachapelle, S., Magliacane, S., Taslakian, P., Martius, G., von Kügelgen, J., and Locatello, F. (2024). Multi-View Causal Representation Learning with Partial Observability. In *International Conference on Learning Representations (ICLR)*. 39

[Yeh et al., 2022] Yeh, C.-H., Hong, C.-Y., Hsu, Y.-C., Liu, T.-L., Chen, Y., and LeCun, Y. (2022). Decoupled Contrastive Learning. In *European Conference on Computer Vision (ECCV)*. 18, 19, 21

[Yendiki et al., 2022] Yendiki, A., Aggarwal, M., Axer, M., Howard, A. F. D., van Walsum, A.-M. v. C., and Haber, S. N. (2022). Post mortem mapping of connectional anatomy for the validation of diffusion MRI. *NeuroImage*, 256:119146. 60

[You et al., 2023] You, K., Gu, J., Ham, J., Park, B., Kim, J., Hong, E. K., Baek, W., and Roh, B. (2023). CXR-CLIP: Toward Large Scale Chest X-ray Language-Image Pre-training. In Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., and Taylor, R., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 101–111, Cham. Springer Nature Switzerland. 72

[Younes, 2010] Younes, L. (2010). *Shapes and Diffeomorphisms*. Springer Berlin Heidelberg. 44, 51

[Yu and Tao, 2019] Yu, B. and Tao, D. (2019). Deep Metric Learning With Tuplet Margin Loss. In *International Conference on Computer Vision (ICCV)*. 12

[Yuan et al., 2021] Yuan, X., Lin, Z., Kuen, J., Zhang, J., Wang, Y., Maire, M., Kale, A., and Faieta, B. (2021). Multimodal Contrastive Training for Visual Representation Learning. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6991–7000. IEEE. 75

[Yushkevich et al., 2008] Yushkevich, P. A., Zhang, H., Simon, T. J., and Gee, J. C. (2008). Structure-specific statistical mapping of white matter tracts. *NeuroImage*, 41(2):448–461. 61

[Zacharaki et al., 2009] Zacharaki, E. I., Hogea, C. S., Shen, D., Biros, G., and Davatzikos, C. (2009). Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. *NeuroImage*, 46(3):762–774. 44

[Zagoruyko and Komodakis, 2017] Zagoruyko, S. and Komodakis, N. (2017). Paying More Attention to Attention: Improving the Performance of Convolutional Neural Networks via Attention Transfer. In *International Conference on Learning Representations (ICLR)*. 48

[Zbontar et al., 2021] Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. (2021). Barlow Twins: Self-Supervised Learning via Redundancy Reduction. In *Proceedings of the 38th International Conference on Machine Learning*, pages 12310–12320. PMLR. ISSN: 2640-3498. 9, 76

[Zeghlache et al., 2023] Zeghlache, R., Conze, P.-H., Daho, M. E. H., Li, Y., Boité, H. L., Tadayoni, R., Massin, P., Cochener, B., Brahim, I., Quellec, G., and Lamard, M. (2023). Longitudinal Self-supervised Learning Using Neural Ordinary Differential Equation. In Rekik, I., Adeli, E., Park, S. H., Cintas, C., and Zamzmi, G., editors, *Predictive Intelligence in Medicine*, pages 1–13, Cham. Springer Nature Switzerland. 38

[Zeghlache et al., 2024] Zeghlache, R., Conze, P.-H., Daho, M. E. H., Li, Y., Boité, H. L., Tadayoni, R., Massin, P., Cochener, B., Rezaei, A., Brahim, I., Quellec, G., and Lamard, M. (2024). LaTiM: Longitudinal representation learning in continuous-time models to predict disease progression. arXiv:2404.07091 [cs]. 38

[Zhang et al., 2019] Zhang, F., Hoffmann, N., Karayumak, S. C., Rathi, Y., Golby, A. J., and O'Donnell, L. J. (2019). Deep White Matter Analysis: Fast, Consistent Tractography Segmentation Across Populations and dMRI Acquisitions. In Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., Zhou, S., Yap, P.-T., and Khan, A., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Lecture Notes in Computer Science, pages 599–608, Cham. Springer International Publishing. 61

[Zhang et al., 2022a] Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S. E., and Zheng, Y. (2022a). DTFD-MIL: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18802–18812. 76

[Zhang et al., 2008] Zhang, S., Correia, S., and Laidlaw, D. H. (2008). Identifying White-Matter Fiber Bundles in DTI Data Using an Automated Proximity-Based Fiber Clustering Method. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):1044–1053. 60

[Zhang et al., 2003] Zhang, S., Demiralp, C., and Laidlaw, D. (2003). Visualizing diffusion tensor MR images using streamtubes and streamsurfaces. *IEEE Transactions on Visualization and Computer Graphics*, 9(4):454–462. Conference Name: IEEE Transactions on Visualization and Computer Graphics. 61

[Zhang et al., 2022b] Zhang, Y., Jiang, H., Miura, Y., Manning, C. D., and Langlotz, C. P. (2022b). Contrastive Learning of Medical Visual Representations from Paired Images and Text. In *Proceedings of the 7th Machine Learning for Healthcare Conference*, pages 2–25. PMLR. ISSN: 2640-3498. 75

[Zhang et al., 2010] Zhang, Y., Zhang, J., Oishi, K., Faria, A. V., Jiang, H., Li, X., Akhter, K., Rosa-Neto, P., Pike, G. B., Evans, A., Toga, A. W., Woods, R., Mazziotta, J. C., Miller, M. I., van Zijl, P. C. M., and Mori, S. (2010). Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *NeuroImage*, 52(4):1289–1301. 61, 70

[Zhao et al., 2019] Zhao, F., Wu, Z., Wang, L., Lin, W., Xia, S., Shen, D., and Li, G. (2019). Harmonization of Infant Cortical Thickness Using Surface-to-Surface Cycle-Consistent Adversarial Networks. In Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., Zhou, S., Yap, P.-T., and Khan, A., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 475–483, Cham. Springer International Publishing. 10

[Zhao et al., 2021] Zhao, Q., Liu, Z., Adeli, E., and Pohl, K. M. (2021). Longitudinal self-supervised learning. *Medical Image Analysis*, 71:102051. 38

[Zheng et al., 2020] Zheng, Y., Fan, J., Zhang, J., and Gao, X. (2020). Exploiting Related and Unrelated Tasks for Hierarchical Metric Learning and Image Classification. *IEEE Transactions on Image Processing*. 24

[Zhou et al., 2021a] Zhou, T., Canu, S., Vera, P., and Ruan, S. (2021a). Latent correlation representation learning for brain tumor segmentation with missing MRI modalities. *IEEE Transactions on Image Processing*, 30:4263–4274. Publisher: IEEE. 45

[Zhou et al., 2021b] Zhou, Z., Sodha, V., Pang, J., Gotway, M. B., and Liang, J. (2021b). Models Genesis. *Medical Image Analysis*, 67:101840. 17, 21

[Zong et al., 2023] Zong, Y., Mac Aodha, O., and Hospedales, T. (2023). Self-Supervised Multimodal Learning: A Survey. 75

[Zou et al., 2022] Zou, K., Faisan, S., Heitz, F., and Valette, S. (2022). Joint Disentanglement of Labels and Their Features with VAE. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1341–1345, Bordeaux, France. IEEE. 30, 31

[Zuo et al., 2021] Zuo, L., Dewey, B. E., Liu, Y., He, Y., Newsome, S. D., Mowry, E. M., Resnick, S. M., Prince, J. L., and Carass, A. (2021). Unsupervised MR harmonization by learning disentangled representations using information bottleneck theory. *NeuroImage*, 243:118569. 10

[Zvitia et al., 2010] Zvitia, O., Mayer, A., Shadmi, R., Miron, S., and Greenspan, H. K. (2010). Co-registration of White Matter Tractographies by Adaptive-Mean-Shift and Gaussian Mixture Modeling. *IEEE Transactions on Medical Imaging*, 29(1):132–145. Conference Name: IEEE Transactions on Medical Imaging. 60