Bibliometrics, Change Detection and Multimodal Learning on Earth Observation Data : The Case of Deforestation

Nathalie Neptune

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Nathalie NEPTUNE

Le 15 juillet 2022

Bibliométrie, détection de changement et apprentissage multimodal sur des données d'observation de la Terre - Le cas de la déforestation

---

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et Télécommunications de Toulouse**

Spécialité : **Informatique et Télécommunications**

Unité de recherche :
IRIT : Institut de Recherche en Informatique de Toulouse

Thèse dirigée par
Josiane MOTHE

Jury

M. Mathieu ROCHE, Rapporteur
Mme Muriel VISANI, Rapporteure
Mme Josiane MOTHE, Directrice de thèse
Mme Florence SèDES, Présidente

Contents

Acknowledgements

Having reached the end of my PhD journey, looking back at this enriching experience, I would like to thank God for His blessings during this time.

To my director, Professor Josiane Mothe, I would like to express my deepest gratitude and highest appreciation, for her supervision, guidance and counsel, first during my masters and then again during my PhD. My deepest gratitude and appreciation to her for the support that I needed to successfully complete my PhD. She demonstrated extreme patience towards me as she taught me about all aspects of research.

I would also like to thank Mr. Julius Akinyemi, my mentor, his advice and counsel helped guide my work and his strong support made this thesis possible.

Thank you to the members of the jury, for the time devoted to reviewing this thesis, for their precious remarks, their interest and their involvement.

I thank Professors Edgard Etienne, Janain Jadotte and Justin Casimir of the board of directors of the Faculté des Sciences of the State University of Haiti for supporting me throughout this thesis.

I would also like to thank Professor Bernard Dousset for our many exchanges and his insightful suggestions which have helped me a lot with my research.

I am grateful for the professors, Olivier Teste, André Péninou, Jean Michel Bruel, Françoise Adreit, Nathalie Hernandez, and Stephane Isnard who allowed me to share their experience as teachers and trusted me with their courses.

My gratitude also goes to all the members of the SIG team, professors, post-docs, PhD students and interns. Thank you for having made my time at IRIT an enriching and enjoyable experience.

Thank you to Victor and Bissmella and all the interns I had the pleasure to work with for the enthusiasm they brought to our collaborative work.

I want to thank my colleagues, PhD students Oumaima, Léa, Tianyi, Michele, Olivier, Chaima, Elvis, Abder, and doctors Clément, Nabil, Faneva, Reshma, Daria, Elliot, Frank, Mahmoud, Mahdi, Ngoc, Thiziri, Luis, Wafa, Oihana, Yacine, Philippe, for the many discussions, lunches, coffee breaks and all the good times shared.

Thank you to my friends outside of IRIT, in particular, Hanna, Eric, Francisca, Behi, Carole, Webens, Malia, Citra, Josué, Yohan, Aina, Merveille, Prince, Fanja, Toky, Cecilia, Okhino, Sarah, Daniel, Loïs, Michel, and all my EPET family too numerous to name, for their support especially as I went through those really

rough patches, for being there for me, for their encouragements and understanding as they saw less of me when I was taking extra time to focus on the thesis. Thank you to Victoria for being my biggest cheerleader, always finding time to be present and always having encouraging words.

I am grateful for my long time office mate Zia, for being always willing and enthusiastic to share with me his knowledge on all matters IR and beyond.

A particular thanks to Jacques Thomazeau, for generously taking time to reply to my numerous requests regarding the OSIRIM platform, his support has been instrumental to my work.

I also want to thank the members of the administration at IRIT, Chloé Bourbon, Catherine Blanc, Clémentine Roger, Jean-Christophe Barbelet, Julie Mballa, for helping me with all administrative matters.

Thank you to Agnès Requis at the doctoral school for her help with all the administrative process throughout the past few years.

A special thanks to my family, my parents, my siblings, my nieces and nephews, for their unconditional, unwavering love and support.

Résumé

Dans cette thèse, nous présentons des méthodes automatiques pour analyser des images satellites et des documents scientifiques, séparément et conjointement. Nous utilisons le cadre de la déforestation pour valider notre approche. En effet, la déforestation et de la dégradation affectant de nombreuses zones forestières rend nécessaire l'utilisation de méthodes permettant de détecter et de surveiller l'état des forêts automatiquement.

Ce travail, appliqué à la déforestation, reste générique et peut s'appliquer à d'autres domaines dans lesquels des images d'observation peuvent être annotées à partir de publications au sujet du phénomène à annoter.

Nous présentons notre travail d'analyse des publications liées à la déforestation couvrant plusieurs décennies. Nos méthodes offrent une nouvelle approche prometteuse pour l'analyse des données environnementales afin d'étudier à grande échelle des informations sur la déforestation ou d'autres sujets.

Nous proposons d'extraire les meilleurs mots-clés d'un corpus de publications scientifiques sur un même sujet, après avoir retiré les documents les moins pertinents de ce corpus en fonction de leur titre. Pour ce faire, nous utilisons des plongements de phrases et des mesures de similarité sémantique. En utilisant notre approche sur les corpus liés à la déforestation, nous obtenons les meilleurs mots-clés principalement liés à ce sujet.

Nous proposons d'annoter des paires d'images satellites de forêts ayant subi des changements, avec des mots-clés de publications scientifiques. Les paires d'images sont annotées avec les mots qui leur ressemblent le plus. Nous utilisons une représentation commune des images et des mots-clés extraits des publications, à l'aide de réseaux de neurones. Nous trouvons les mots-clés les plus similaires à la paire d'images dans cet espace commun avec la mesure de similarité du cosinus. Avec notre approche, nous constatons que les corpus qui sont liés aux forêts en général, et à la région d'intérêt plus spécifiquement, peuvent être utilisés pour annoter automatiquement et de façon pertinente les images. Nous montrons que ces corpus donnent de meilleurs résultats que l'utilisation d'un corpus plus générique comme Wikipedia.

Abstract

In this thesis, we present automatic methods to analyze satellite images and scientific documents, separately and jointly. We use the deforestation framework to validate our approach. Indeed, deforestation and degradation affecting many forest areas make it necessary to use automatic methods to detect and monitor the state of forests.

This work, applied to deforestation, remains generic and can be applied to other fields in which observation images can be annotated from publications about the phenomenon of interest.

We present our analysis work on publications related to deforestation covering several decades. Our methods offer a promising new approach for the analysis of environmental data to study large-scale information on deforestation or other topics.

We propose to extract the best keywords from a corpus of scientific publications on the same subject, after having removed the least relevant documents from this corpus according to their title. To do this, we use sentence embeddings and semantic similarity measures. By using our approach on corpora related to deforestation, we obtain the best keywords mainly related to this subject.

We propose to annotate pairs of satellite images of forests having undergone changes, with keywords of scientific publications. Pairs of images are annotated with the words that most closely resemble them. We use a common representation of images and keywords extracted from scientific publications, using neural networks. We find the most similar keywords to the image pair in this common space, with the cosine similarity measure. With our approach, we find that corpora that are related to forests in general, and to the region of interest more specifically, can be used to automatically and meaningfully annotate images. We show that these corpora give better results than using a more generic corpus like Wikipedia.

Publications

[Akinyemi 2018] Julius Akinyemi, Josiane Mothe et Nathalie Neptune. *Fouille de publications scientifiques pour une analyse bibliométrique de l'activité de recherche sur la déforestation.* In EGC-Atelier Fouille du Web, pages 11–23, 2018.

[Huynh 2018] Duy Huynh et Nathalie Neptune. *Automatic image annotation: the case of deforestation.* In Actes de la Conférence TALN. Volume 2-Démonstrations, articles des Rencontres Jeunes Chercheurs, ateliers DeFT, pages 101–116, 2018.

[Akinyemi 2019] Julius Akinyemi, Josiane Mothe et Nathalie Neptune. *Jeux de Données d'Observation de la Terre pour la Détection des Changements dans les Forêts.* Information Retrieval, Document and Semantic Web, vol. 19, no. 1, 2019.

[Neptune 2020] Nathalie Neptune. *Automatic Annotation of Change in Earth Observation Imagery.* In Proceedings of the Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020), 2020. [Online; accessed 9-March-2020].

[Neptune 2021] Nathalie Neptune et Josiane Mothe. *Automatic Annotation of Change Detection Images.* Sensors, vol. 21, no. 4, page 1110, 2021.

Chapter 1

Introduction

Summary

Abstract.

Ending deforestation is one of the targets of the Sustainable Development Goals laid out by the United Nations. Monitoring this requires the collection and analysis of a variety of data on the state of the Earth's forests. There are presently many Earth observation satellites that provide images of the planet including of forested areas. There are also many scientists and researchers who have been investigating questions related to deforestation for a long time, and have produced a wealth of publications on the subject. However, there still remains a need to combine these sources of information together for a better view of the phenomenon of deforestation. Our work proposes to address this need. In this introduction, we present the motivation for our work, the research questions we try to answer, and our contributions.

1.1 Motivation

In the context of the Sustainable Development Goals, set out by the United Nations in 2015 [Assembly 2015] to guide communities globally towards a more sustainable future, environmental monitoring and preservation are essential. One of the key Sustainable Development Goals' targets, in relation to forests, is to end

deforestation by halting forest loss and increasing forest cover through restoring lost forest areas (reforestation) and establishing new forest areas (afforestation) [Assembly 2015]. In fact, many developing countries face numerous environmental challenges among which the loss of forest for other types of land use, which is also referred to as deforestation (according to the Food and Agriculture Organization of the United Nations 2001 definition[1]). There have been multiple initiatives taken in order to address the issues of deforestation and degradation, in particular in tropical forests, including the Reducing Emissions from Deforestation and forest Degradation (REDD+) international framework [Mora 2013].

In this context, gathering and analyzing information about deforestation, and the state of forests in general, can prove important. The information about the changes occurring in forests can be gathered from a variety of sources. Earth observation satellites capture images of the surface of the Earth including forest areas. A lot of documents on the state and evolution of forests are produced in different forms such as scientific papers, news articles, conservation agency reports, government official reports and many more. All this information provides an abundant source of heterogeneous data. Stakeholders such as conservationists, researchers, and decision makers can use this information to make timely and relevant decisions about forest conservation.

To perform an integrated analysis of this environmental data, special methods are needed in particular to combine the information from satellite images and text documents. While both sources of information are valuable on their own, combining them could provide additional insight that might not come out when they are considered separately. This is especially true for instance when considering that a phenomenon happening within a forest may not be readily visible from images captured by a particular satellite sensor. The information on this "non-visible from space" phenomenon could however be widely mentioned in documents related to this area. Even for visible phenomena, using text data as an additional information source could provide supplementary information not available from the satellite images but mentioned in the texts. The combination of the image and text information can therefore be valuable for stakeholders by potentially giving them a more complete picture of the area of interest.

The amount of data from satellite images and text documents alike may pose a challenge for stakeholders who wish to analyze them. Satellites have been capturing images of the Earth for several decades, and their number keeps increasing, this creates a need for automatic methods to perform some analyses

[1] https://www.fao.org/3/j9345e/j9345e07.htm - Definitions of deforestation.

that can guide stakeholders in their tasks.

The goal of this thesis is to provide methods for analyzing both satellites images and text from scientific documents for monitoring changes in forest as a starting point to building solutions for environmental stakeholders and decision makers. In this work we propose generic approaches that can be applied to other contexts, and other subjects of interest, where there is a need to combine data from a set of images and a collection of related documents.

1.2 Research Questions

Our primary research question is whether we can learn more by combining images and text. We essentially are looking for a way to better combine the two modalities (image and text) to extract insights. Knowing that we are working with satellite imagery in the case of deforestation, we need a related text corpus to combine with our images. This brings up a related question about the corpus. Could a focused sub-corpus be more relevant and better suited for our goal of extracting insights than the original full corpus? If we look at keywords as a first level of insight that we can extract from our corpus, we want to find out if we can possibly improve keyword extraction if we reduce the size of the corpus in a targeted way. This leads to another related question about what can be extracted from such a corpus. What other insights can we extract from the corpus in addition to keywords?

1.3 Research Focus

Sensors that are on-board Earth Observation (EO) satellites capture images of the Earth at specific times. An increasing number of operating EO satellites are currently orbiting our planet. In the context of environmental conservation, it is useful to have automated methods to analyze the images and extract the relevant forest-related information. In fact, deforestation can be measured, reported and verified by using available satellite and forest inventory data [Mitchell 2017]. Likewise, scientists are publishing an increasingly large number of research papers on changes occurring in forests [Akinyemi 2018, Aleixandre-Benavent 2018]. Automated methods for analyzing the information contained within those documents can be useful to environmental stakeholders.

This thesis focuses primarily on scientific publications related to deforestation

and satellite images of forested areas that are undergoing change. Scientific publications have particular characteristics, they are structured (divided into sections), factual (evidence-based), and come with standard metadata. Satellite images have spectral characteristics that are sensor-specific, they have specific spatial and temporal resolutions, are georeferenced and have standardized metadata. The particular characteristics of these data make them appropriate for a variety of automatic analysis tasks using their metadata, their features and their content.

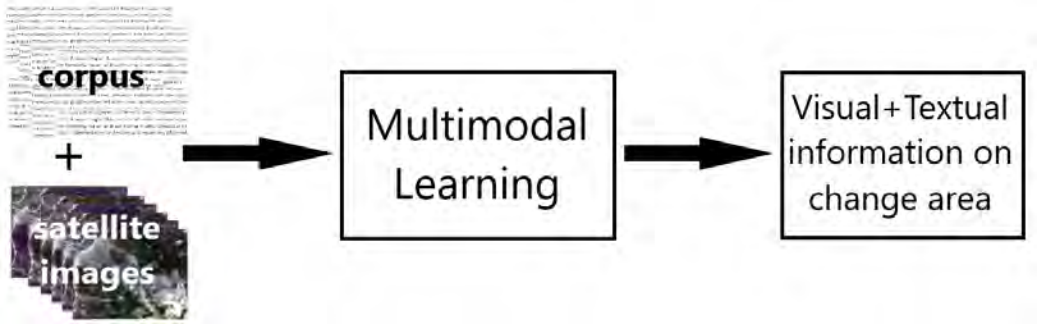

Figure 1.1: Overview of the proposed approach to learn from text corpora and satellite images for extracting information on forest change. Inputs are a text corpus and satellite images, outputs are change maps and annotations from the corpus.

Existing text analysis methods such as topic modelling [Deerwester 1990, Blei 2003, Grootendorst 2020a] and keyword extraction [Jones 1972, Mothe 2018, Campos 2020] allow to find if a document is related to deforestation by looking at the most important terms from the document. Combining current pixel-classification based change detection methods [El Amin 2016, Peng 2019] with text analysis methods could provided a different type of insight into geo-phenomena like deforestation. With this combination, we can find terms that the scientists have used in their publications in relation to deforestation in addition to the fact that there is deforestation in a given area. Achieving this combination, requires not only using the existing methods designed for joint text-image analysis, but also taking into consideration our particular interest in changes happening within forests requiring us to compare images taken at different times. Our approach aims to combine the modalities, image and text, to present a more complete picture of deforestation in a given area for a given time period, which cannot be provided by each modality taken individually. Figure 1.1 provides an overview of our approach showing satellite images and text corpus that serve as inputs for multimodal learning. The outputs that are produced are change maps learned from the images and annotations taken from the text. When combined, the text

and image elements give complementary information on the area and the changes that are detected in that area.

We hypothesize that combining the text from the publications with the satellite images, by using keywords extracted from the publications as annotations for the images, will help us learn more about the phenomenon of deforestation than if we do not combine them. The images are provided with a limited set of annotations or labels that can be used for training learning models. Previous work [Frome 2013, Uzkent 2019] have relied on large text sources like Wikipedia to be able to predict image labels. We aim to show that scientific publications, in the case of changes in forests, can provide a better source for image annotations than Wikipedia.

Deforestation is an important research topic in environmental sciences, therefore, many scientists have published on the topic. We perform bibliometric analyses of scientific publications, on the topic of deforestation, as an example of using bibliometrics to uncover the geographic structure and the coverage of research activity on deforestation. We want to get insight into what scientists are saying about the problem and how they collaborate with each other on the topic. We also want to get a view on countries involved in the research and the ones under study. We center our analysis around topics such as: (1) the evolution of the trend of scientific production on deforestation with time, (2) the countries where authors are based in versus the locations they are studying, (3) the countries and regions under study with their associated keywords, (4) the collaborative structure of the work on the topic of deforestation, (5) the keyword extraction methods that work best at a corpus level and can help evaluate the suitability of a corpus for subsequent learning tasks, and (6) the impact of a chosen corpus on the alignment of visual and textual features for change image annotation.

1.4 Contribution

To combine the detection of change in satellite images of forests with relevant words extracted from related scientific publications, we propose a series of tasks: bibliometric analyses (Chapter 3), corpus keyword extraction (Chapter 4), and change detection with image annotation (Chapter 5).

For performing bibliometric analyses on scientific literature on the topic of deforestation, our work uses text mining, statistics and network analyses. This helps us answer our secondary research question about the types of insight we can extract from a corpus. We use an approach similar to previous studies in

bibliometrics of a research topic [Mothe 2006, Pautasso 2015, Juárez-Orozco 2017] but focus on an understudied topic namely deforestation. We also analyze more data compared to other work on the same topic [Aleixandre-Benavent 2018]. Our main data source is the Web of Science Core Collection[2], which contains publications on deforestation from 1975. Our analysis focuses on the years from 1975 to 2016 and the publications were collected in October 2017. We also use data from the Food and Agriculture Organization (FAO)[3] and the Worldbank[4], for 2016, on forest area and gross domestic product by country respectively. We perform most of our bibliometric analyses using Tetralogie [Dousset 2009], a platform made for scientific and technological monitoring, developed at the Institut de Recherche en Informatique de Toulouse. For the first analysis topic about the evolution of production in research on deforestation with time, we count the number of publications and chart the results per year. We also look at the proportion of change from the earlier years to the latest decade. For the topic on the countries and authors performing research and the locations under study, we count the number of publications per author and country to find the top ones and we extract locations names from the publications to find the most frequently mentioned ones. We also look at how the publication count of top countries compare to their Gross Domestic Product (GDP) and how often they are mentioned. We address the third topic related to the countries and regions under study with their associated keywords, by extracting the countries and regions that are most often mentioned in the corpus. For each country or region, we find the top keywords with which it co-occurs. For the fourth topic related to the collaborative structure of the work on the topic of deforestation, we perform network analyses to look at collaborations among countries and others. Our work on the bibliometric analyses of scientific literature of the topic of deforestation is presented in Chapter 3.

Given a set of satellite images of a forest area that has undergone changes, we need to find a corpus of scientific publications that we can use with these images to extract relevant keywords. This requires that we extract keywords representing the whole corpus as opposed to keywords for individual publications separately. Before using the corpus for future tasks, we can pre-process it and examine its suitability by keeping only the most relevant publications and looking

[2] `https://www.webofscience.com/wos/woscc/basic-search` - Web of Science - A global citation database.

[3] `http://www.fao.org/home/en/` - Food and Agriculture Organization of the United Nations

[4] `https://data.worldbank.org/` - World Bank Open Data

at the top keywords. To answer our secondary research question about whether a more focused sub-corpus would be better suited for extracting keywords than the full corpus, and address our fifth analysis topic, we propose a novel title-topic similarity selection approach to select publications from a corpus for keyword extraction. This approach is based on using similarity measures between word and sentence embeddings [Bojanowski 2017, Reimers 2019], to find the publications that are most similar to the topic of the corpus. We find keywords representing the whole corpus by aggregating document-level keywords, which are extracted using current methods [Jones 1972, Mothe 2018, Campos 2020, Grootendorst 2020b]. We also propose to combine keywords extracted with existing methods [Jones 1972, Grootendorst 2020b] to obtain an improved keyword list. We use data from the Web of Science[5], Elsevier [Kershaw 2020], and Pubmed [Aronson 2000] in our experiments. Our work on corpus keyword extraction is presented in Chapter 4.

Given a satellite image set and a set of corpora, we want to detect the changes by comparing images from before and after the changes occurred, and we want to find the words from our corpora that are most relevant to the pre and post-change images taken together. We want to annotate the change images with annotations from our corpora. To answer our primary research question about learning more insights from the combination of text and images, and address our sixth analysis topic, we propose to use visual semantic embeddings [Frome 2013, Faghri 2017], generally used for image captioning, they allow us to learn the representations of the images and the words from the corpora in the same embedding space. Having the images and the words in the same space allows us to compare them using similarity measures like cosine similarity to find, for example, for a given image pair the words most similar to it, and vice versa. Unlike the previous work [Frome 2013, Faghri 2017], we are not working with single images from cameras but with pairs of satellite images. We use word and sentence embeddings from state of the art models [Bojanowski 2017, Reimers 2019] and learn to represent image pairs in the word/sentence embedding space. To the best of our knowledge this is the first time this approach has been used for change annotation in satellite images. We use an early fusion deep neural network change detection model, based on an early fusion model [Daudt 2018a] previously used for urban change detection, performing pixel-level classification. We use the encoding part of the early fusion deep neural network change detection model as the visual semantic model, based on the deep visual semantic embedding approach [Frome 2013] to

[5]https://www.webofscience.com/wos/woscc/basic-search - Web of Science - A global citation database.

predict the vector representation of the images in the embedding space. From the resulting image vectors we find, from within the corpus, additional relevant annotations through image-text retrieval. With image-text retrieval, for a given image pair we find the top words that are most similar to it from the words present in our chosen corpus. We effectively pose the annotation problem as a text retrieval problem from an image query. With this approach we are both detecting changes in forests and finding related topics through the keywords extracted from the corpus. We also investigate the impact of the chosen corpus on the annotation by experimenting with a variety of corpora and by aligning the corpora with each other and examining the resulting annotations. We use data from the European Space Agency[6], the Landsat Program[7], the Web of Science[8] and Wikipedia[9]. Our work on change detection and visual semantic embeddings for image annotation is presented in Chapter 5.

The remainder of this thesis is organized as follows: we present, in Chapter 2, a review of the literature on the research streams related to our work, namely bibliometric analyses of scientific documents, corpus keyword extraction, change detection in satellite images, and visual semantic learning models. We end the chapter with a review of relevant datasets from the literature. Chapter 3 contains our work on the bibliometric analyses conducted on a large deforestation corpus. In Chapter 4 we present our work on corpus keyword extraction, describing our proposed title-topic similarity selection method along with the related experiments. Our method for combining change detection and image annotation with text from an unpaired corpus is described in Chapter 5. Our conclusion is presented in the Chapter 6.

[6]https://www.esa.int/ - The European Space Agency

[7]https://landsat.gsfc.nasa.gov/ - Landsat Program

[8]https://www.webofscience.com/wos/woscc/basic-search - Web of Science - A global citation database.

[9]https://www.wikipedia.org/ - Wikipedia - The Free Encyclopedia

Chapter 2

Litterature Review

Summary

Abstract.

Combining scientific text and satellite images to extract information on geo-phenomena requires connecting computer vision with natural language processing. Attaching annotations to changes that are detected in satellite images is one of the applications of such approaches. It is the main goal of our work. Many machine learning approaches addressing this task rest on two common steps of encoding the image and using a language model for the text. Those two key elements have seen many variations but the core approach rests on finding a way to match the representations of text and image together. Furthermore, there has been a number of datasets that have been produced to test the proposed methods. In this chapter we provide an overview of approaches and datasets for bibliometric analyses of scientific documents, change detection on satellite imagery and visual semantic learning. Our goal is to highlight the most relevant current approaches for our proposed tasks and the remaining gaps that we will address in our work.

2.1 Introduction

As we aim to extract information from scientific text and Earth observation satellites, we conduct a review of the literature on information extraction approaches for both text and images. We review text mining in the context of bibliometric analysis of a scientific domain. We focus on the sub-task of keyword extraction from a scientific corpus as a way of finding candidate annotations for our images. We focus on change detection methods for satellite imagery as well as land cover mapping and more general image segmentation methods from computer vision used in change detection. We go over both traditional approaches and those based on deep neural networks. We propose to combine the two modalities, text and images with a visual semantic learning model. We present general purpose models models and their applications to satellite imagery. We end this chapter with a presentation of datasets available for the different tasks involved, bibliometric analysis, change detection in satellite imagery and visual semantic representation learning.

In the following sections, we present the three main research areas related to our work: bibliometric analyses of scientific documents (Section 2.2), change detection in satellite imagery (Section 2.3) and visual semantic learning models (Section 2.4). We also review datasets for each task (Section 2.5).

2.2 Bibliometric Analyses of Scientific Documents

Our work on bibliometric analysis is mainly related to two research axes: text mining for bibliometric analyses and the bibliometric analysis of a research domain. When bibliometrics is performed for a specific scientific discipline or domain it is concerned with scientific information. [Glänzel 2003] calls it "an extension of science information using metrics". This is where it touches on quantitative research in information retrieval.

2.2.1 Text Mining for Bibliometric Analyses

Text mining has been defined as performing exploratory data analyses over text; its main purpose is the discovery of new information from text collections [Hearst 1999]. In order to perform text mining we need to extract numerical representations for our text documents. There exist different numerical representations of text that have been proposed over time to facilitate automatic analyses

of textual data by computer programs.

To perform such analyses with automatic methods several word and document representation models have been proposed over time, starting from the vector space model [Salton 1965], originally proposed for information retrieval systems, in which vectors of attribute weights represent documents and queries. The attributes also called terms, are commonly keywords or phrases and the weights of each attribute are computed based on the presence/absence of the attribute. The goal of an information retrieval system is to match a query with relevant documents from a document collection, retrieved documents are returned by the system from most relevant to least relevant. In the vector space model, documents and queries vectors are compared with a similarity measure, the cosine similarity is commonly used.

In a boolean model [Lancaster 1973], the attribute weights account only for the presence of absence of an attribute. Frequency-based weighting account for how frequently an attribute is present, allowing to capture the relative importance of an attribute within a document collection. The term frequency inverse document frequency weighting (TF-IDF) [Jones 1972] is a commonly used frequency-based weighting statistic. It is calculated as the inverse proportion of the frequency of a term in a document to the percentage of documents the term appears in.

As the dimension of the document-attribute matrix increases significantly when the number of documents and features becomes really high, in large datasets, dimension reduction techniques were applied to the original vector space model, such as latent semantic indexing (LSI) [Deerwester 1990] and Principal Component Analysis (PCA). In LSI [Deerwester 1990], instead of using the large albeit sparse attribute-document matrix, resulting from using all the terms for each document, a low-rank approximation based on the singular value decomposition of the matrix is used.

In more recent years, with big datasets available for training, deep learning based methods have been proposed for representing words and documents. Vector representations based on neural language models called word embeddings have been proposed such as [Mikolov 2013, Bojanowski 2017] in which vocabulary terms are represented by vectors of real numbers embedded from a high dimension space to a lower dimension space. Skip-gram is an unsupervised learning model that given a word, learns to predicts its context which is its surrounding words [Mikolov 2013]. Words with similar meaning will have a similar vector representation when learned with the skip-gram model. In [Bojanowski 2017], which is an extension of [Mikolov 2013], instead of using whole words, n-grams are used

and each n-gram has its own vector, resulting in each word being represented by the sum of its n-grams.

One limitation of these word embeddings is that they only provide a single representation for a word and therefore do not account for polysemy. If the dataset on which the algorithm is trained is big and diverse enough there might be at least a word with more than one meaning but this word will only have a single vector representation. Contextualized word vector representations are provided by the most recent deep neural network language models [Peters 2018, Devlin 2018, Radford 2019]. These models can be used to learn contextual word embeddings where a single word can have different representations based on the contexts in which it appears.

Bibliometrics was defined by Alan Pritchard as quantifying written communication using mathematical and statistical methods [Pritchard 1969]. Bibliometric analyses based on network analysis can be performed based on different statistics methods and data mining techniques considering different types of networks. The basic bibliometric methods are frequency-based. The number of publications for example is a simple measure that can be used to analyse a data set of scientific publications [Broadus 1987]. Frequency counts can be done on other data and metadata from the dataset such as authors, keywords, topics and countries [Broadus 1987]. Other more elaborated measures such as the average number of publications, and the journal impact factor [Garfield 2006], which is an index reflecting the number of citations the papers published in a journal receive, have also been used. These measures inform on how active researchers are in terms of written scientific production and on their impact within the community [Glänzel 2003].

Co-occurrence analyses of authors (co-authors) are done to analyze collaboration among researchers [Glänzel 2003]. This practice is based on the principle that formal collaborations between researchers are well documented and result in co-authorship [Glänzel 2003]. Aggregations can be done at different levels (country, research unit, research domain, authors etc..) to analyse collaboration between individuals, researcher units, or countries [Borgman 1989]. Co-occurrence analyses also apply to terms present in titles, abstracts, keywords and full text of publications. Furthermore, multidimensional approaches have been proposed using co-occurrence matrices of terms or other variables [Glänzel 2003].

Several tools have been proposed to perform these bibliometric analyses in an integrated manner such as [Dousset 2003, Dousset 2009] which implements text mining algorithms for bibliometric analyses.

In our work we focus on the bibliometric analysis of a specific domain of study, which we discuss in the following section.

2.2.2 Bibliometric Analysis of a Research Domain

Bibliometric analysis techniques have been used to measure the importance of collaborative activities between co-authors forming networks in a given research field [Logan 1991]. Bibliometric tools have also been used to analyse research topics and streams in a specific domain [Milojević 2011]. In the field of information retrieval, for example, [Smeaton 2003] analyzed the co-authors of the Association for Computing Machinery (ACM) Special Interest Group on Information Retrieval (SIGIR) conference publications over a period of 25 years in order to reveal the evolution of the themes/subjects of this conference publications and to find the most central authors by considering the graph of co-authors. [Mothe 2006] showed how Tétralogie [Dousset 2003] can be used to combine text mining and geographic information system features to uncover the geographic structure of a domain. The authors presented a use case on the proceeding of the ACM SIGIR conference. Geographic maps were used to visually represent the geographic dimension revealed through text mining.

Domain specific analyses can focus on a single researcher in particular within a field. For example, [Skupin 2014] used bibliometrics and network visualization jointly to highlight domain structure as well as communities based on co-citations with the publications of the author David Mark. This approach made it possible to carry out a visual analysis of the dimension of David Mark's influence and its persistence over time in the field of geographic information systems. The focus of a domain specific analysis might also be a research unit. [Neptune 2014] used bibliometric analyses of scientific publications to analyse the research activity of a specific research unit. That work focused on all the publications from the research unit's database. Using data on the organisation and the personnel of the laboratory, the analyses on the production by team as well as the collaborations among teams and with authors outside of the researcher unit were performed.

[Kergosien 2018] preformed a cased study on an heterogeneous scientific corpus made of documents of different types (theses, papers, reports and others) and from different sources (ISTEX [Cuxac 2017], Agritop[1] and the ANRT[2]), on

[1]https://agritrop.cirad.fr/ - Open repository of publications of the CIRAD research unit.

[2]http://www.diffusiontheses.fr/content/4-anrt-lille-reproduction-theses - Atelier National de Reproduction des Thèses.

the topic of climate change in Senegal and Madagascar. The corpus contained publications in both English and French. Metadata and abstracts were used to standardize the documents using the Metadata Object Description Schema[3] (MODS) format. Spatial, temporal and thematic entities were annotated in the abstract. Spatial entities are annotated based on linguistic patterns. Thematic entities were annotated by extracting domain vocabulary using semantic resources for lexical annotation. Temporal annotations were created using a rule-based time-sensitive labeling system for temporal expressions. The data were then indexed from subsequent processing, analysis and for information retrieval. This work was part of a larger project to extract territories and areas under study from the publications, to find the academic disciplines, and to analyze the evolution of research paradigms.

In our work we are interested in analysing scientific publications related to the topic of deforestation. Bibliometric analyses have been performed on topics related to forests such as forest health [Pautasso 2015], forest fires [Juárez-Orozco 2017] and more broadly on forestry journals [Dobbertin 2010, Bojović 2014]. Limited work has been done on the specific topic of deforestation such as [Aleixandre-Benavent 2018] which is concurrent with our initial work on the subject.

2.2.3 Keyword Extraction from a Scientific Corpus

In our research we perform text mining on text data and meta-data of scientific publications related to deforestation. We do this by performing term frequency analyses and collocation analyses. We believe this approach to be suitable to evaluate whether a chosen corpus can be used for subsequent tasks using statistical or machine learning models such as keyword extraction. We further use word embeddings of words extracted from the scientific publications as input to a visual semantic model for learning annotations for satellite images of areas undergoing change.

A number of statistical methods have been proposed to extract keywords from documents one of the first such methods is TF-IDF [Jones 1972]. For each term in the corpus, the value of TF-IDF is calculated for each document. TF-IDF estimates the importance of a term in a document relative to the corpus. A more recent statistical method YAKE [Campos 2020] extracts and ranks keywords based on a number of pre-defined features. As word embeddings have become more popular

[3] `https://www.loc.gov/standards/mods/` - Metadata Object Description Schema

and successful for a variety of natural language processing tasks, they have also been included in keyword extraction methods. [Mothe 2018] performs keyword and keyphrase extraction on documents using a graph method in which the words were replaced by their embeddings.

When the goal is to extract keywords representing a corpus, the keywords extracted from each document are aggregated. In our work we mostly use embedding-based keyword extraction [Bojanowski 2017, Reimers 2019] on a corpus of scientific publications, and we aggregate the document-level keywords at the corpus level to have a unique set of keywords representing the whole corpus. The top corpus-level keywords allow to evaluate how relevant a corpus is to our topic of interest, namely deforestation. Additionally, the extracted keywords serve as candidate annotations for satellite images of forests where deforestation has occurred.

2.3 Change Detection in Satellite Images

Our work on change detection in satellite images is related to three main research areas, namely image segmentation techniques typically used for land cover mapping, change detection methods and neural networks for change detection.

2.3.1 Land Cover Mapping and Image Segmentation

Classification methods are commonly used for land cover mapping and satellite image segmentation. [Lagrange 2015] showed that, in the image domain, for pixel classification, the best performance is obtained when using deep convolutional neural networks as opposed to expert classifiers. [Kussul 2017] proposed a method combining supervised and unsupervised neural networks for the segmentation and classification of multi-source satellite images. [El Amin 2016] proposed a change detection method for satellite images based on the difference of image features extracted using a deep CNN model. In the Planet[4] competition "Planet: Understanding the Amazon from Space" [Kaggle 2017] where the goal was to classify images of the Amazon forest, the winning submission was an ensemble of state of the art CNN image classification architectures.

[4] `https://www.planet.com/` - Planet Labs is an earth observation satellite company based in San Francisco.

2.3.2 Traditional Change Detection Methods

Many techniques have been proposed over the years to detect changes in satellite imagery. These techniques can be categorized into different approaches such as algebra, transformation, classification, advanced models, geographical information system (GIS) approaches, or visual analysis [Lu 2004].

The very first algebra technique was univariate image differencing. This straightforward technique detects the change by applying a threshold to the difference in pixel value between first-date image and the second-date image [Lu 2004]. This technique was widely used in change detection problems, particularly for detecting forest changes [Miller 1978, Singh 1989]. Another well-known algebra technique was image regression. First, the method assumed that pixels in the same location are related by a linear function in time. Thus, the pixels values in the second-date image can be predicted according to the regression function. Finally, a threshold was applied to the difference between the true second-date value and the predicted second-date value. This technique showed better performance than the image differencing technique [Singh 1989].

The second group of techniques uses transformations such as Principal Component Analysis (PCA), Multitemporal Kauth-Thomas (MKT), Gramm–Schmidt (GS), and Chi-square transformations [Lu 2004]. [Collins 1996] examined PCA, MKT, GS methods to the problem of forest change due to conifer mortality and concluded that PCA and MKT give better results than GS.

The third group of change detection techniques is made of classification approaches which have been used for both image-pair change detection and for time series change detection with maximum-likelihood-based estimation [Mertens 1997] or a random forest classifier [Olofsson 2016].

2.3.3 Neural Network Change Detection Models

Binary change detection methods allow systems to detect if and where a change occurred, whereas semantic change detection methods also specify semantic information about the change that is detected [Lu 2004]. Both types of methods have been applied in the literature to satellite images to detect land cover changes with the most recent ones using deep learning models [Daudt 2018a, Peng 2019]. These latest models are all based on the U-Net architecture [Ronneberger 2015], which is an encoder–decoder architecture with skip connections between the encoding and decoding streams. U-Net was initially proposed for the segmentation of biomedical images.

Three variations of U-Net were proposed by [Daudt 2018a]. The first variation performs early fusion by taking the concatenation of the images as input, effectively treating them as different color channels. In this case, the change detection problem is posed as an image segmentation task with two classes: change and no-change. The second and third model variations are siamese versions of the U-Net architecture where the encoder part is duplicated to encode each image separately, and skip connections are used in two ways, by concatenating either the skip connections from both encoding streams or the absolute value of their difference. The three models were tested on four datasets. Based on the F1 measure, the first fully convolutional early fusion model outperformed the siamese models on two out of four datasets used for evaluation. The second siamese model with the concatenation of the difference values outperformed the other models on the remaining two datasets. Another variation of U-Net with early fusion, for change detection, which uses dense skip connections was proposed by [Peng 2019], it was shown to outperform [Daudt 2018a] on only one dataset [Lebedev 2018].

In our work, we use a model similar to the early fusion model proposed by [Daudt 2018a] to perform the binary change detection task. It is a model that is particularly well suited for learning global image features from image pairs at once making it appropriate for learning visual semantic embeddings.

2.4 Visual Semantic Learning Models

Global land cover maps are a key resource for scientists and a variety of users looking for data in many sectors from disaster relief to ecosystem and biodiversity conservation, to name a few [Mora 2014]. However, when subtle changes are occurring at fine scale, they may not appear in global maps, and might in fact be hard to detect by remote sensing [Houet 2010]. The detection and analysis of certain types changes might therefore require additional data sources. We propose to add information from a scientific corpus to the information from satellite images. Our visual semantic learning model work is related to three research streams, traditional visual semantic methods, visual semantic embeddings and in particular visual semantic embeddings for satellite images.

2.4.1 Traditional Models

Several methods have been proposed to represent images and text together in order for both modalities to have similar representations that can be compared,

typically to perform subsequent tasks.

Methods based on canonical correlation analysis (CCA) [Hotelling 1936] have been proposed to align elements from text and images such as [Hardoon 2004, Blaschko 2008, Socher 2010]. All three methods use kernelized CCA (kCCA) to match images and text, for retrieving images with a text query [Hardoon 2004], for clustering images and text in a latent space, and for annotating image segments [Socher 2010]. Except for [Socher 2010] all the other methods use a parallel corpus with the images. [Socher 2010] also addresses the problem of insufficient labeled data for annotation and semantic segmentation. While methods like [Socher 2010] use handcrafted features, more recent multi-model approaches have leveraged convolution networks trained on a classification task to extract image features and address the problem of not having sufficient labeled data. In these visual semantic embedding methods, embeddings for both image and text are generated in a joint embedding space.

2.4.2 Visual Semantic Embeddings

With visual semantic embeddings, we learn to represent textual and visual data in the same vector space. The closer those data points are semantically, the smaller their distance is in this joint space. Several approaches have been proposed including the joint embedding of images and words into a common low-dimension space [Frome 2013] for image classification, and the embedding of images and sentences into a common space for image description [Socher 2014]. More recently, a neural network named CLIP [Radford 2021] has been proposed for learning visual representations from natural language supervision. CLIP [Radford 2021] uses a transformer model to learn visual features and a causal language model for the text features. It is pre-trained on a large dataset of text-image pairs taken from the internet. The main purpose of CLIP [Radford 2021] is to be used without retraining (in a zero-shot manner) for tasks such as image classification and description.

By using joint image and text embeddings, we propose to automatically assign relevant annotations to image pairs where changes are detected. Therefore, we perform two core tasks: change detection and the annotation of the image pairs. We propose to use scientific publications as a source of annotations, for which we learn the vector representations using a neural language model. These annotations can be used in subsequent tasks such as image indexing and retrieval. While change detection can be applied to any type of image pairs of the same scene or

location, in our work we focus on the case of changes occurring in forest areas to test our approach.

2.4.3 Visual Semantic Change Embeddings for Satellite Images

When the semantic information about the classes present in EO images is inconsistent or lacking, different solutions have been proposed to extract that information from other sources such as ontologies [Bouyerbou 2014] or geo-referenced Wikipedia articles [Uzkent 2019]. Another solution is to use visual semantic embeddings by representing the images and text in the same vector space and learning the classes of the unlabeled images based on the similarity between the vector representations across the image and text modalities [Frome 2013]. Our proposed approach is in this line of work.

By integrating an ontology to the segmentation process of pre- and post-disaster images, authors in [Bouyerbou 2014] showed that overall accuracy went from 67.9% to 89.4% for images of their test area. With a reduced number of samples (200), authors in [Uzkent 2019] demonstrated that using Wikipedia annotations for the task of semantic segmentation, the Intersection-over-Union score (a measure of the similarity and diversity of sample sets that essentially takes the ratio of their intersection over their union) was 51.70% compared to 50.75% when pre-training on ImageNet. In both cases, the methods were tested on images of urban areas. While the use of the ontology created by experts in [Bouyerbou 2014] improved greatly the accuracy of the classification algorithm, it came at the high cost of expert hours. The crowdsourcing approach using Wikipedia data in [Uzkent 2019] while promising, resulted only in modest improvement for the semantic segmentation task. A version of CLIP [Radford 2021] fine-tuned on satellite imagery [Lu 2017] has also been made available [5]. It was evaluated on image-text retrieval tasks with satellite images, and shown to outperform the base CLIP [Radford 2021] model. CLIP-like models require that candidate labels be provided for the images when performing the image annotation task.

In our case, the scientific publications written by researchers can be seen as both a source of expert knowledge and a crowdsourced resource as they are coming from a large number of scientists. We propose to use expert knowledge through relevant scientific papers from which annotations are extracted. Our

[5] `https://github.com/arampacha/CLIP-rsicd` - Repository for CLIP model fine-tuned on RSICD data.

method therefore performs change detection, in satellite image pairs by predicting change pixels, and it also performs semantic annotation of the image pair by predicting its labels. We use a deep learning network architecture based on U-Net [Ronneberger 2015]. U-Net is an encoder-decoder deep neural network architecture. Unlike the original U-Net we are using an encoder built with residual blocks. Our network architecture is detailed in Chapter 5. We therefore propose a way of finding relevant candidate labels for change images as well as ways to match the best candidate labels with each image pair.

The specificity of our work is in that we perform annotation learning by using a corpus that is not aligned to our images to learn additional annotations beyond the image labels. We specifically apply our work to the case of change detection in satellite imagery which has not been done previously. We consider the impact of learning representations for pre and post change images jointly instead of learning single image representations individually as in the common visual semantic embedding pipeline. The intuition is that the joint representation will enhance the difference between change portions of the images and unchanged portions, making the change image pairs easier to differentiate from the non-change image pairs, and facilitating the annotation task. Using before and after images also allows us to know in which time period the change occurred.

2.5 Text and Image Data Sources

There are datasets previously made available in the form of corpora and satellite images that can be used for bibliometric analyses of scientific documents [Juárez-Orozco 2017, Kershaw 2020], change detection in satellite images [Bourdis 2011, Lebedev 2018, Daudt 2018a, Ji 2018] and visual semantic embeddings respectively [Lin 2014, Young 2014, Russakovsky 2015]. We present in this section three types of data sets. Table 2.1 shows datatasets from the literature that have been used (or could be used) for bibliometric analyses, change detection or for learning visual semantic embeddings.

For the full range of our experiments, we created new datasets for bibliometric analyses, and for combining change detection and learning visual semantic embeddings which are later described in 3.3 and 5.3.1. In our work, we combine change detection with learning visual semantic embeddings to add semantic annotation to satellite images of forest areas that have undergone changes. We propose to extract candidate annotation from scientific literature. To perform the experiments required to validate our proposed approaches we need to use data

Dataset	Domain	Public	Suitable for
[Dobbertin 2010]	Forest	No	Bibliometrics
[Bojović 2014]	Forest	No	Bibliometrics
[Juárez-Orozco 2017]	Forest	No	Bibliometrics
[Uribe-Toril 2019]	Forest	No	Bibliometrics
[Kershaw 2020]	Multi-domain	Yes	Bibliometrics
[Bourdis 2011]	Multi-domain	Yes	Binary Change Detection
[Lebedev 2018]	Multi-domain	Yes	Binary Change Detection
[Daudt 2018a]	Urban	Yes	Binary Change Detection
[Ji 2018]	Urban	Yes	Segmentation, Change Detection
[Shimabukuro 2000]	Forest	Yes	Change Maps
[Hansen 2013]	Forest	Yes	Change Maps
[Wheeler 2014]	Forest	Yes	Change Maps
[Shimada 2014]	Multi-domain	Yes	Land Cover Map
[ESA 2017a]	Multi-domain	Yes	Land Cover Map
[ESA 2017b]	Multi-domain	Yes	Land Cover Map
[Lin 2014]	Multi-domain	Yes	Object Detection, Segmentation, Key-point Detection, Image Captioning
[Young 2014]	Multi-domain	Yes	Image Captioning
[Russakovsky 2015]	Multi-domain	Yes	Image Classification, Object Detection
[Lu 2017]	Multi-domain	Yes	Image Captioning
[Helber 2019]	Multi-domain	Yes	Image Classification

Table 2.1: Datasets from the literature that have been used for bibliometric analyses, change detection, visual semantic embeddings learning, along with land cover and change maps are listed.

from scientific publications and satellite images. We need these data to be at least related to each other thematically. Ideally, we would also like to find publications and images that are linked geographically and temporally as well. We also need the images to be at least partially labeled. Because we want to perform change detection we need to have either image pairs or time series covering the periods before and after the changes. An existing dataset meeting all those requirements did not exist.

For change detection on satellite imagery, the datasets used in the literature are mostly urban change detection or other non-forested areas. While several change maps are publicly available, they do not typically come with pre and post-change image pairs. Our criteria for selecting the satellite data were that they are real images (as opposed to synthetic images), contain before and after image pairs (non-composite), and that they are publicly available and can thus be freely reused to reproduce our work. For the scientific publications our criteria was that they would be real publications (not computer generated) and be as related as possible to the images selected. While our criteria might appear strict our priority was to show that our approach can be successfully applied to a real-world setting. We created new datasets to perform our experiments, reusing, repurposing, and combining the data that were already available.

We also investigate the impact of the corpus on the visual semantic learning task and propose ways to build a domain-specific corpora to be used instead of (or jointly with) commonly used large corpora such as Wikipedia or Common Crawl[6].

2.5.1 Datasets for Bibliometric Analyses

For bibliometric analyses of forest-related topics, [Dobbertin 2010, Bojović 2014, Juárez-Orozco 2017, Uribe-Toril 2019] have each created a corpus with relevant scientific publication data. Except for [Juárez-Orozco 2017], the list of publications used is not given. In all cases the dataset is not readily available but needs to be recreated following the procedures described by the authors. [Dobbertin 2010, Bojović 2014] focus on forestry journals over 29 and 4 years respectively, while [Uribe-Toril 2019] focuses on a single journal over a period of 8 years. [Juárez-Orozco 2017] is more topic specific and focuses on fires in tropical rainforests over a period of 36 years.

Bibliometric datasets created with records provided by an indexing platform

[6]https://commoncrawl.org/ - Common Crawl.

such as in [Dobbertin 2010, Bojović 2014, Juárez-Orozco 2017], are typically not made publicly available due to copyright restrictions. However, the datasets underlying bibliometric studies can, most of the time, be recreated by following the collection and processing procedures of the dataset creator. A more recent initiative by an academic publisher [Kershaw 2020] is an open access dataset with full papers from a variety of journals from the publisher, over a 6 year period. This dataset is not specifically related to forests but contains publications in Earth and environmental sciences.

2.5.2 Datasets for Change Detection

Several change detection datasets have been used in previous work such as [Bourdis 2011, Lebedev 2018]. These two aforementioned datasets are made of synthetic images either entirely for [Bourdis 2011] or in part for [Lebedev 2018]. Datasets with only real satellite images have also been used [Daudt 2018a, Ji 2018], for urban change detection. All the previously cited datasets are provided with a binary change map but without image or pixel-level semantic labels for the changes.

[Shimabukuro 2000, Hansen 2013, Wheeler 2014] are change datasets that provide data on changes occurring in forests. These datasets can be used as change maps for change detection in forests but they are not provided with the satellite images used to create the change maps. Land cover maps have been made available covering large portions of the globe at different times [Shimada 2014, ESA 2017b, ESA 2017a]. While these maps provide annotations for satellite images they only provide a snapshot of the areas at one point in time. Such large scale land cover maps can be useful in assessing global and regional environmental change, some of them suffer from relatively low accuracy for some classes [Mora 2014].

2.5.3 Datasets for Visual Semantic Embeddings

Some well-known datasets have been used for learning visual semantic embeddings such was the case in [Lin 2014, Young 2014, Russakovsky 2015]. MS-COCO [Lin 2014] is a collection of images with rich annotations that can be used for various tasks such as image classification, semantic segmentation, and image captioning. Flickr-30K [Young 2014] is a dataset of annotated images, similar in terms of content but smaller then [Lin 2014], that can be used for image captioning tasks. ImageNet [Russakovsky 2015] is a very large (14.2 million images and

20000 classes) dataset of annotated images that can be used for image classification and object detection. All three datasets can be used to learn visual semantic embeddings at the image or object level. They all contain images from the web featuring scenes from everyday life captured on camera.

RSICD [Lu 2017] is a dataset proposed specifically for remote sensing image captioning made of aerial images, each with five captions created by human volunteers with knowledge of remote sensing and annotation experience. More recently, EuroSAT [Helber 2019], a satellite image dataset with single-label image patches has been released. Each image therein is labelled with its respective land cover class.

All these datasets can be used to learn visual semantic embeddings. None of these datasets can be used for learning changes in satellite images directly, however, they can be used for transfer learning in a change detection context if training data is lacking.

Chapter 3

Bibliometric Analysis of Research on Deforestation

Summary

Abstract.

Deforestation is a very widespread phenomenon affecting sizeable portions of territories, especially in the tropical regions. Remote sensing allows the monitoring and analysis of the spatial-temporal evolution of this phenomenon. Using text and metadata mining on scientific publications on the subject of deforestation, we identify the locations of scientific production on deforestation. We find out how researchers are connected to each other with network analysis. With keyword analysis, we identify the sites affected by deforestation in which researchers are interested, namely tropical forests and the Amazon, for the most part, as well as related subjects linked to the environment and health. We conclude that a corpus made of scientific publications, on the topic of deforestation, is appropriate for finding annotations for satellite images on which deforestation can be observed visually. In this case, the text might not be a direct description of the images but it is related to them based on the topic, location, and time of interest. In this chapter we present the results of our work on the bibliometric analyses of research activity on deforestation.

3.1 Introduction

Deforestation is an environmental phenomenon that can have a negative impact on the ecosystem of the earth [Foley 2005]. It is defined by the FAO as "the change of intended land use from forest to non-forest[1]." As early as 1992, reviewing the links between the processes that lead to deforestation and its consequences in the Amazon basin of Brazil, [Diegues 1992] estimated that the rate of deforestation was high and increasing rapidly and dangerously.

Performing text mining on scientific publications related to deforestation makes it possible to quantify the information about these publications as well as their content [Pritchard 1969]. We carry out an analysis on these texts to see the geographical and temporal evolution of scientific research on deforestation. This study makes it possible to identify the main players and their location. In addition, the subjects on which they focus their work on are also highlighted.

Collaborations between researchers is an important aspect of research. Due to the nature of the phenomenon, research on deforestation often calls for expertise in various disciplines. It follows that an analysis of collaborations in publications on deforestation can highlight the multidisciplinary nature of this research.

The results of our bibliometric analyses guide our work on finding relevant publications for specific geo-phenomena to be used in experiments for image-text learning.

3.1.1 Analysis Objectives

A phenomenon that has been studied by scientists will typically be published in a formal way. Finding out if there is, over time, increasing interest on a specific topic, such as deforestation, is a way to evaluate whether that topic is still relevant and if enough information can potentially be gathered through publications on that topic. We therefore, begin by looking at the evolution of the volume of scientific output on deforestation with time and ask the following question:

I. How has the number of publications evolved over time?

We then have a closer look at the scientific output by country, and at how the countries where researchers are based compare to countries under study, and ask the following questions:

II. Which countries or locations are the focus of research on deforestation? And which countries or locations are most frequently mentioned?

[1] http://www.fao.org/3/i0440e/i0440e02.htm - FAO | Deforestation, land-use change and REDD

We have a closer look at the topics associated to those countries or locations by asking the question:

III. What keywords are associated to a given frequently mentioned country/location over time?

Then we examine the publications' authorship to find out how it relates to the country and the disciplines involved:

IV. Which authors publish the most on deforestation? What is the importance and nature of collaborations between researchers from different countries? And to which scientific disciplines do the authors belong?

3.1.2 Summary of Contributions

The following contributions are included in this chapter:

- We conduct an extensive bibliometric analysis of the topic of deforestation in scientific literature covering more than two decades of publications.
- We propose an approach to extract insights by crossing the data from a corpus on deforestation with socio-economical and environmental data from public data sources.
- We propose to combine bibliometric analyses with text mining to compare and contrast countries publishing on deforestation and countries related to deforestation based on the keywords extracted from the publications.

3.2 Related Work

We conduct analyses of a corpus in which publications from various universities, laboratories, research units and other institutions can be found. The publications are all related to the same topic of deforestation.

[Mothe 2006] illustrated how the Tetralogie [Dousset 2003] platform can combine data mining with the functionalities of geographic information systems to discover the geographic structure of a domain. The authors presented a case study using the proceedings of the Association for Computing Machinery (ACM) Special Interest Group on Information Retrieval (SIGIR) conference. They used geographic maps to visually represent the geographic dimension revealed by the data mining task. In [Neptune 2014] we have already used bibliometric analyses of scientific publications to analyze research activities within a specific scientific

unit, the Toulouse Computer Science Research Institute. This work covered all the publications present in the database of the research unit, all themes combined. Using data on laboratory organization and staff, analyses of team production as well as inter-team collaborations and collaboration with authors outside the unit were carried out. We have shown how bibliometric analysis can be used for certain aspects of the evaluation of a scientific unit such as scientific production, outreach, involvement in training through research and scientific perspectives.

The analyses presented in this chapter follow on from this work. We use scientific publication data and metadata mining to analyze research activities on the subject of deforestation. We are interested in the geographic dimension present in the data not only in relation to the location of the authors but also in relation to the areas and regions on which they focus their research.

[Kang 1990] proposed a feasibility study on the Feature-Oriented Domain Analysis (FODA) method for domain analysis. This method creates a model of the domain, including performing an analysis of the extent of the domain. [Skupin 2014] jointly used bibliometrics and network visualization to bring out the structure of a domain as well as communities based on co-citations with the publications of author David Mark. This approach made it possible to visually analyze David Mark's persistent influence in the field of geographic information systems.

The analyses presented here, are based on bibliometics, and they shed light on the field of deforestation with the perspective of guiding future work on data related to this theme.

3.3 Data and Methods

We take the data from the Web of Science Core Collection [2]. We carried out a topic search using the term "deforest*". All publications dated from 1975 to 2016 were collected on October 23, 2017. This approach makes it possible to have an overview of the field. More sophisticated queries could have been used in particular to find publications which do not explicitly mention deforestation while being linked to it, however such approach would be more error-prone. The analyses are performed with Tetralogie [3], a platform for scientific and technological monitoring which was developed at the Toulouse Computer Science Research Institute [Dousset 2009].

[2] https://webofknowledge.com - Clarivate | Web of Science

[3] http://atlas.irit.fr/ - Tetralogie, a technology watch software used in research and teaching.

A total of $16,136$ publications were collected with $31,772$ authors in 149 countries and territories.

We carried out univariate and multivariate analyses including publication counts (per year, per country per year, per author), keyword counts, author co-occurrence counts, country co-occurrence counts. The results obtained were then used to produce scatter diagrams, networks of co-authors and research categories. Finally, we crossed data from the Food and Agriculture Organization of the United Nations and the World Bank to see the productivity of the countries that publish the most in relation to the number of inhabitants and to the gross domestic product, for the year 2016. This crossover puts into perspective the scientific production on the theme of deforestation, for a country, in relation to the size of its population and in relation to the size of its economy, for the year 2016. This crossover gives an idea of the human and financial effort represented, for each country, by their contribution to scientific production on deforestation. The authors' countries are extracted from the address in the data provided by the Web of Science. The different spellings of the names of the countries are taken into account. For some publications, the author's address may be missing.

3.4 Results

3.4.1 Production Study

3.4.1.1 Overall Production on the Topic of Deforestation

Figure 3.1 shows the number of publications found for each year. For the first years with less than 10 publications per year until 1982 which has 12 publications, the number of publications present in the collection is very low. From the beginning of the 1990s a remarkable increase is noted in the annual production.

This upward trend in production continues until 2016. This trend is similar to the evolution observed in other fields, in particular the natural sciences and the health sciences, [Bornmann 2015] reported an exponential growth in publication output for the years 1980 to 2012. Over the last 10 years covered by our collection, from 2006 to 2016, the number of publications on the topic of deforestation increased by 220%. Another topic-specific example is from [Sampaio 2013] who analyzed publications on the tropical disease Leishmaniasis from two publication databases PubMed and PASCAL. For PubMed, they found that the number of publications increased steadily from the years 2000 to 2012, from 474 to 938.

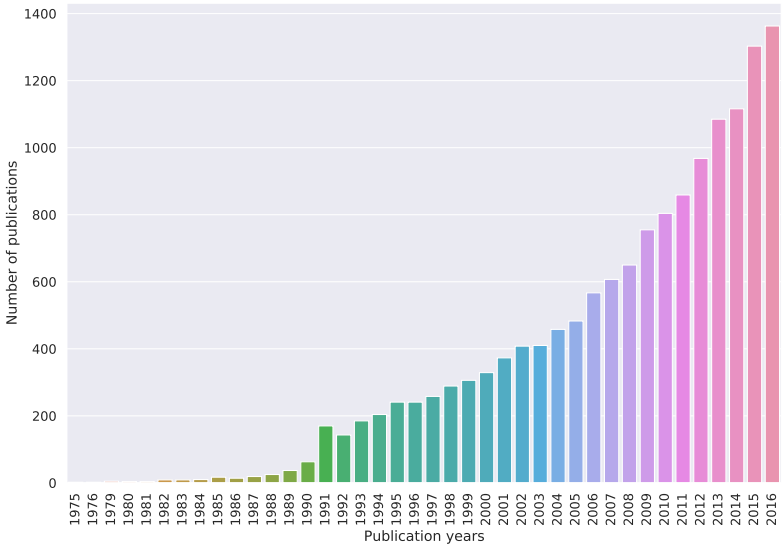

Figure 3.1: The number of publications on deforestation has grown exponentially in the last decade, in our corpus extracted from the Web of Science. We can see the increasing numbers over time from 1978 to 2016.

3.4.1.2 Production by author and network of co-authors.

The authors who published the most appear in the Table 3.1, with a Brazilian author in the lead (Fearnside), followed by an author from Australia (Laurance) and an American author (Houghton) in second and third positions respectively. Two authors are therefore from the two countries with the highest overall production while the third comes from one of the countries with the highest production per person for 2016. Some authors were affiliated with institutions in several countries during the period covered by our collection, only the top countries are reported.

The graph of co-authors in Figure 3.2 provides insight into collaboration trends. It shows the many groups formed by the authors who collaborate on the subject. The largest collaboration networks are formed around the most prolific authors.

author	publications	countries
Fearnside, PM	97	Brazil
Laurance, WF	73	Australia, Brazil and Panama
Houghton, RA	63	United States
Lambin, EF	58	United States and Belgium
Shimabukuro, YE	57	Brazil
Koh, LP	56	Australia, Switzerland and United States
Herold, M	49	Netherlands
Asner, GP	49	United States
Achard, F	48	Italy
Peres, CA	44	United Kingdom

Table 3.1: Publications of the 10 authors who published the most and their country of affiliation, in descending order of publication count. Fearnside from Brazil is the most prolific author.

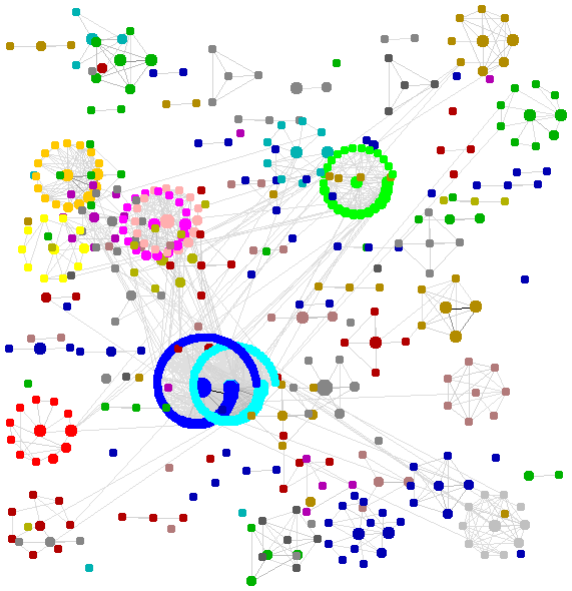

Figure 3.2: Research on deforestation is highly collaborative as shown by the network of co-authors with many collaboration groups of various sizes. Large collaboration groups can be seen, formed around the most prolific authors.

3.4.2 Country Study

In this section, we carry out an analysis of contributions by country, and then present the collaborative network formed by the countries. An author's country (or countries) is the country of affiliation as stated in the author metadata of the publication. Finally, we show the number of publications for each country in relation to their gross domestic product and their population for the year 2016.

3.4.2.1 Contributions and Collaborations

Looking at the data by country, in Figure 3.3, it emerges that among the ten countries with the most publications on deforestation, over the period 1996-2016, three are emerging countries, and two of them are from Latin America: Brazil, China and Mexico. This is an atypical trend which is not observed in other areas. For example, for publications on the broader topic of geosciences, only two emerging countries (China and India), none of which from Latin America, are in the top ten in number of publications (see Figure 3.4). South America has the largest share of forest in protected areas in the world [UNEP 2020]. Of the top ten countries with the most tree species reported by [Beech 2017], 6 are in Latin America (5 in South America plus Mexico). The presence of Brazil and Mexico among the top countries publishing on the topic of deforestation is therefore not surprising in this context. In comparison, [Sampaio 2013] found on the PubMed database, for the topic of Leishmaniasis, among the top 10 publishing countries, there was only one country from Latin America, Brazil. China, Australia, Mexico and Japan were absent from their list, replaced by India, Spain, Iran, and Italy.

The evolution of the production of each country can also be evaluated by year in relation to the total production of the country. This is shown in Figure 3.5, for the last 30 years, from 1996 to 2016. Each point represents a percentage that is calculated by dividing the number of publications in the country for the year by the sum of publications, for the country, for all years, from 1975 to 2016. From 1998, the United States and Canada lead the way and maintain an increase in production over each previous year. However, this trend changed starting from 2008. From 2012 to 2016, the countries that increased their production the most were Germany, Australia, China and Brazil.

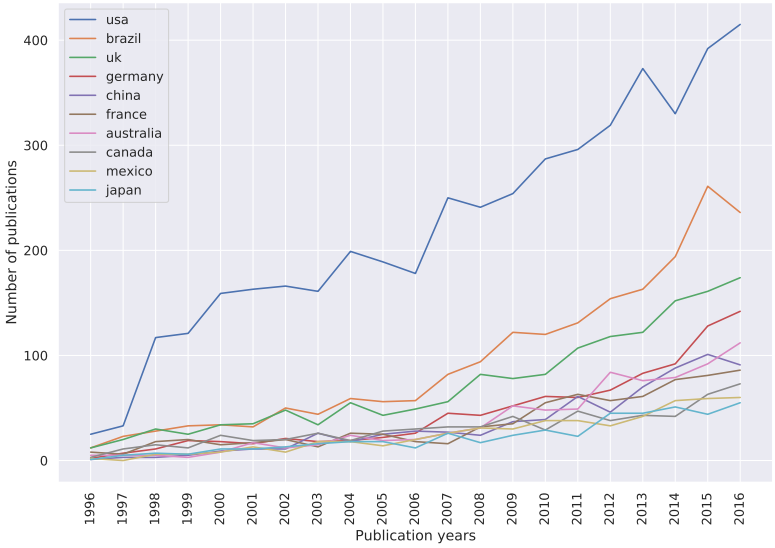

Figure 3.3: For the period from 1996 to 2016, the United States of America have the highest number of publications followed by Brazil. The ten countries with the most publications are shown. The years 1975-1995 are not shown due to the low number of publications for these years.

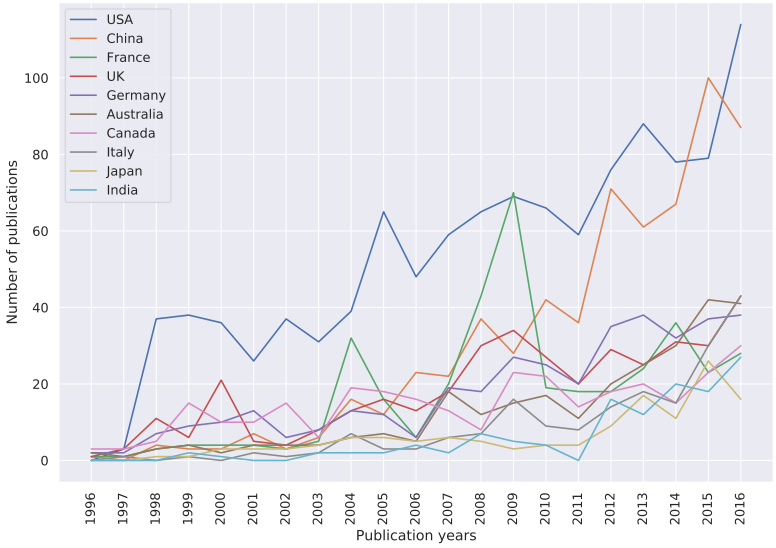

Figure 3.4: From 1996 to 2016 the the United States of America had the highest number of publications on the topic of geosciences in most years except 2009 and 2015 when it was surpassed by France and China respectively. A total of 4641 publications were collected with 13699 authors in 87 countries and territories. In 2016, the United States of America had the highest number of publications followed by China.

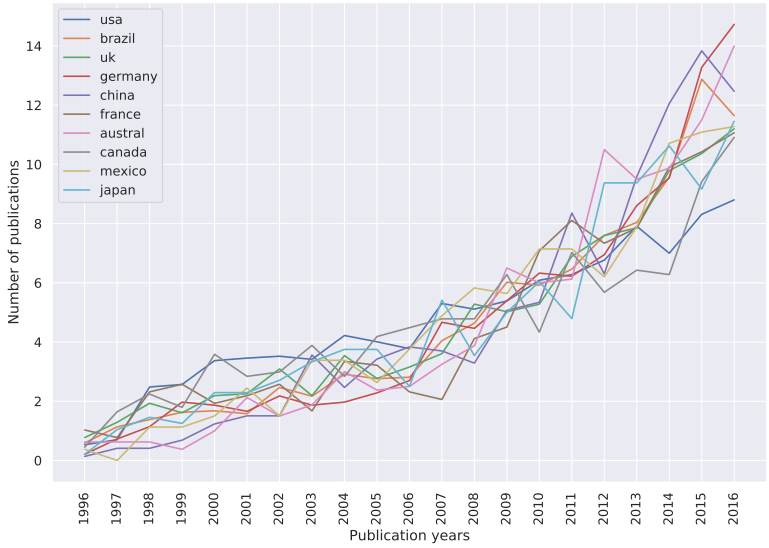

Figure 3.5: Ratio of the number of publications for each country to the total number of publications for the country, by year, from 1996 to 2016 as a percentage. Until 2007, the United States and Canada had the highest increase in production from one year to another. Then from 2008, Germany and Australia became the two countries with the highest increase over the years.

3.4.2.2 Collaboration Network Between Countries

The author collaboration network shows a high level of collaboration between authors from different countries (Figure 3.6). Each node represents a country and the strength of the links between them represents the number of publications co-authored by authors from different countries. The size of each node represents the number of publications for the country. We can observe that all the countries are found in a large group centered mainly around the United States of America.

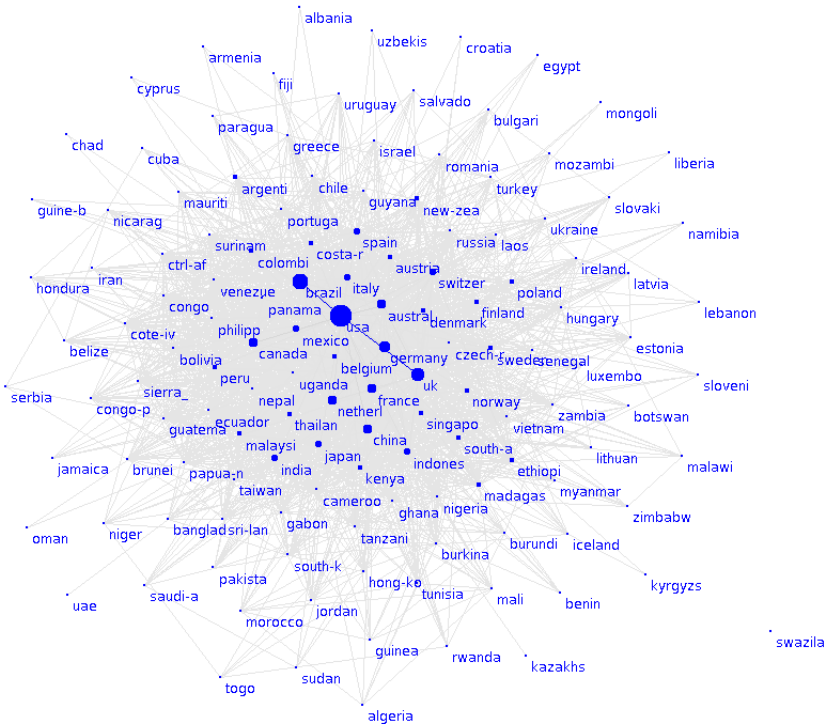

Figure 3.6: The countries publishing on deforestation form a large group mainly centered around the United States for the years 1975 to 2016.

Figure 3.7 shows a closer view of the countries with the most collaborations with the United States. Brazil turns out to be the country that has collaborated most often with the United States, followed by the United Kingdom and Germany. It is possible that the very high position of the United Kingdom and Germany in the ranking of the countries publishing the most on deforestation is in part due to their very strong collaboration with researchers from the United States. [Malhado 2014] has shown that the proportion of publications about the Amazon by authors from the Amazon region, particularly Brazilians, has increased over time but also, the proportion of articles on the Amazon not involving any author

in the region also increased. It may be that even by considerably increasing their participation in scientific production on deforestation, Brazilians are not necessarily able to take the leadership roles.

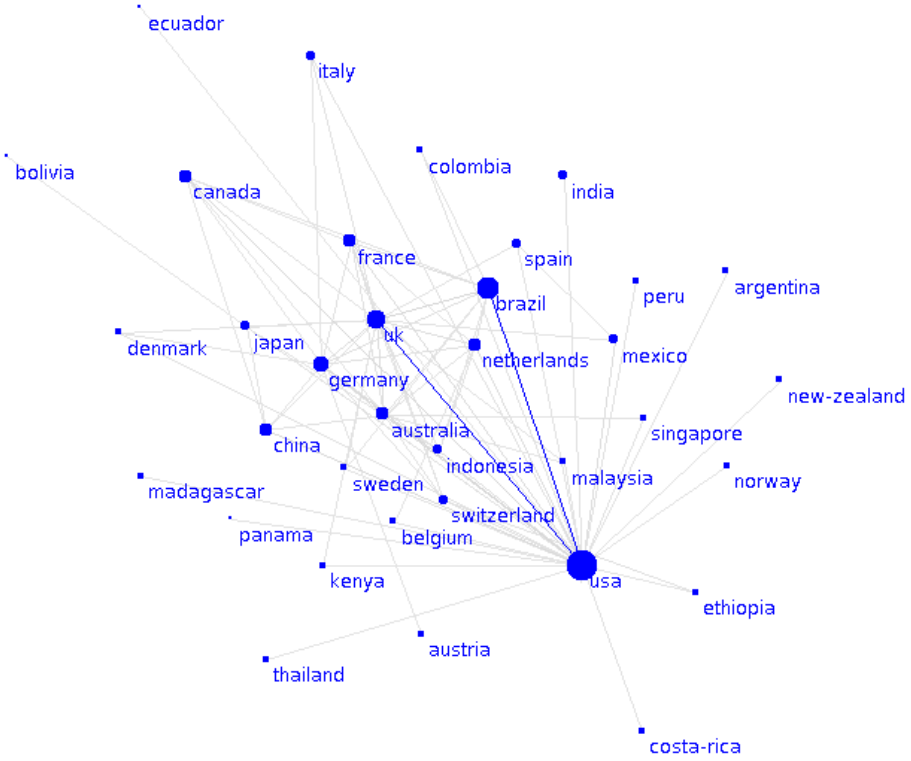

Figure 3.7: For the period from 1975 to 2016, the countries that collaborated the most with the United States were Brazil, the United Kingdom and Germany, as shown by the collaborative network around the United States.

3.4.2.3 Number of publications compared to gross domestic product and population in 2016

We calculate the number of publications per capita and per gross domestic product expressed in billions of US dollars, using data provided by the World Bank [4], for year the 2016.

In Table 3.2, we see that Australia and the Netherlands stand out as being the countries producing the most publications per capita. In the ranking of the number of publications relative to their Gross Domestic Product (GDP), Brazil

[4]https://data.worldbank.org/ - World Bank Open Data.

country or region	count	pop.	count/pop.	GDP'	count/GDP'
Brazil	236	207	1.14	1800	0.13
Australia	112	24	4.64	1200	0.09
Netherlands	72	17	4.23	771	0.09
UK	174	65	2.65	2620	0.07
Canada	73	36	2.01	1530	0.05
Germany	142	82	1.72	3470	0.04
France	86	66	1.29	2470	0.03
India	65	1324	0.05	2260	0.03
USA	415	323	1.28	18600	0.02
China	91	1378	0.07	11200	0.01

Table 3.2: Number of publications for each of the 10 countries with the most publications relative to population and GDP (in billions of US dollars) for 2016. The first column "count" represents the number of publications for the year 2016. The second column "pop." represents the population in millions of inhabitants. The third column "count/pop." represents the ratio between the number of publications and the population expressed in millions. The fourth column "GDP'" represents the gross domestic product in billions of US dollars. The fifth column "count/GDP'" represents the number of publications divided by the gross domestic product in billions. Australia and the Netherlands have the highest publication count per capita. Brazil has the highest publication count relative to GDP.

comes first followed by Australia and the Netherlands. The latter perhaps sees their leadership in deforestation research limited by the relative small size of their economy compared to the United States.

3.4.3 Objects of studies of publications

3.4.3.1 Countries and regions under study

For analyzing the objects of studies we consider only the publications in the English language which represent a total of $13,819$. We have considered the metadata of the publications, we will now look at the data, in other words, the content, which in our case are the titles and abstracts. The content of the publications also provide information on the countries, regions or territories in which the authors were most interested. We are looking for specific insight into the particularities of each country/region beyond the more general trends.

We extract the countries and regions that are mentioned within the text, we do this using Tetralogie [Dousset 2009]. They are highlighted in table 3.3. The Amazon and Brazil take the lead, which is an expected result given the large number of Brazilian contributions and the fact that the Amazon rainforest is the most important in Brazil. Although other countries and regions are mentioned less often, it is interesting to note that all the continents appear in this list. Other countries and regions have under 300 count. The total forest area in 1000 ha as well as the forest conversion from the FAO database[5] are reported for the year 2016. Forest conversion is defined as the portion of forest area converted to another type of land cover in the year 2016. We can see that for 2016, the largest forest areas are reported in the Americas which includes Brazil. The second largest area is found in Europe which includes Russia. The highest number of ha of converted forest was found in Africa, slightly higher than the Americas. When we look at the proportion of conversion compared to the forest area we see that Africa and Indonesia have the two highest conversion to area ratio 0.63 and 0.61 respectively. They are top 4 and top 7 of the most mentioned areas respectively, this ratio might explain, at least in part, their high position in the list.

Another way to look at these countries and regions is through their correlation with the word "deforestation" within the corpus. To do so we take the count of each word for each year and compute the pairwise Pearson correlation of the counts of two given words. For two words with their counts over n years represented

[5]FAO. Emission database. License: CC BY-NC-SA 3.0 IGO. Extracted from: http://www.fao.org/faostat/en/#data/GF/. Date of Access: 22-06-2021.

	country or region	number of mentions	forest area 2016	forest conversion 2016	conversion ratio in %
1	Amazon	3403	n/a	n/a	n/a
2	Brazil	2008	502431	1710	0.34
3	China	1417	212231	0	0.00
4	Africa	1256	652513	4108	0.63
5	America	950	1608212	3996	0.25
6	Mexico	911	66203	131	0.20
7	Indonesia	899	94448	578	0.61
8	India	867	71094	0	0.00
9	Europe	416	1016136	405	0.04
10	Costa Rica	398	2969	0	0.00
11	Malaysia	338	19394	70	0.36
12	Australia	305	133276	5	0.00

Table 3.3: The top 12 countries and regions most often mentioned in publication summaries for the period 1975 to 2016. Forest area and forest conversion (from the FAO database[6]) are reported in 1000 ha, for the year 2016. The most often mentioned region is the Amazon. The most mentioned country is Brazil. The largest forest area is in the Americas. The largest converted forest area as well as the highest ratio of forest conversion are in Africa.

by two variables $X = x_i, ..., x_n$ and $Y = y_i, ..., y_n$, their correlation noted r_{XY} is given by $r_{XY} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$, where $\bar{x}$ and $\bar{y}$ are the means of X and Y respectively.

Indeed deforestation is the conversion of forest to another type of land cover, in other words, it is the change in land cover from "forest" to another land cover or land use. In our corpus the words "deforestation" and "change" are highly correlated at 0.99. Table 3.4 shows the most mentioned countries and regions with their correlation with the word "deforestation" for the years 1996 to 2016. From this table we can see that in some cases there is agreement with the trend highlighted in table 3.3. Indonesia and Africa have similar correlations, they also have high conversion ratios that are very close for the year 2016. In some other cases there seems to be disagreement such as with Europe where the correlation is high at 0.91 while the conversion ratio was only 0.04. It is possible that in cases of continents and regions such as Europe, America and Amazon the data from the FAO and the words from the publication may not be referring to the same entities. In the FAO data Europe includes Russia this may or may not be the case in publications. Likewise, in the FAO data America (Americas originally in

	name of country or region	number of mentions	correlation with "deforestation" 1996-2016
1	Amazon	3403	0.96
2	Brazil	2008	0.98
3	China	1417	0.90
4	Africa	1256	0.94
5	America	950	0.95
6	Mexico	911	0.82
7	Indonesia	899	0.94
8	India	867	0.90
9	Europe	416	0.91
10	Costa Rica	398	0.50
11	Malaysia	338	0.87
12	Australia	305	0.86

Table 3.4: The countries and regions most often mentioned in publication summaries and their correlation with the term "deforestation" for the period from 1996 to 2016. Brazil is the most correlated with deforestation at 0.98 followed by the Amazon at 0.96.

the data) refers to North, Central and South America, while in the publications the word America might be used more often when referring to South or Latin America and less often to North or Central America. We also see that Costa Rica has a the lowest correlation with deforestation even though it is mentioned more often than Malaysia and Australia for instance which have higher correlations. The country mentions with the lowest correlation with deforestation are Costa Rica, Mexico, Australia and Malaysia with 0.50, 0.82, 0.86 and 0.87 respectively. These countries also had low conversion ratios as seen in table 3.3.

While the correlation ratio gives us an idea of how the different countries/regions mentioned were correlated with the term "deforestation", we can view yearly trends by looking at mentions over time for the top 12 countries/regions. Figure 3.8 shows the relative frequencies of the term "deforestation" for each country/region, calculated as the number of times the term is present for the year divided by the total number of terms for that year. The values shown are multiplied by 10^4. We can see that the countries follow different trends, upwards or downwards depending on the period. A higher number of mentions of the country's name would suggest that it is more affected by deforestation and a lower number of mentions would suggest that it is less affected by the phenomenon. For example, we see that from 1996 to 2005 the mentions of "Costa Rica" were high except for a dip in the year 2000 similar to its initial level in 1996. Costa Rica is the only country with numbers consistently low from 2008 onward. Inversely, "Malaysia" had fewer mentions from 1996 to 2004 except for a peak in 2001 similar to its initial level in 1996. However, from 2003 to 2005 "Malaysia" started trending upward. A similar trend can be observed for Indonesia, from 2007 onward, its numbers have followed an upward trend while they were low for earlier years. This might explain why the correlation between deforestation and "Costa Rica" is so low while it is much higher for "Malaysia". This seems to indicate that deforestation in Costa Rica, or the documentation of it, has been decreasing in the last decade covered by our dataset. While around the same period, deforestation has been increasing in Malaysia and Indonesia.

3.4.3.2 Top Keywords per Country/Region over the Years

One way to look at the topics related to each country or region of interest, is through the keywords that are assigned by the authors. We take the same list of countries/regions listed in tables 3.4 and 3.3. For each country/region we take the most frequent keywords. We remove common keywords such as "deforestation",

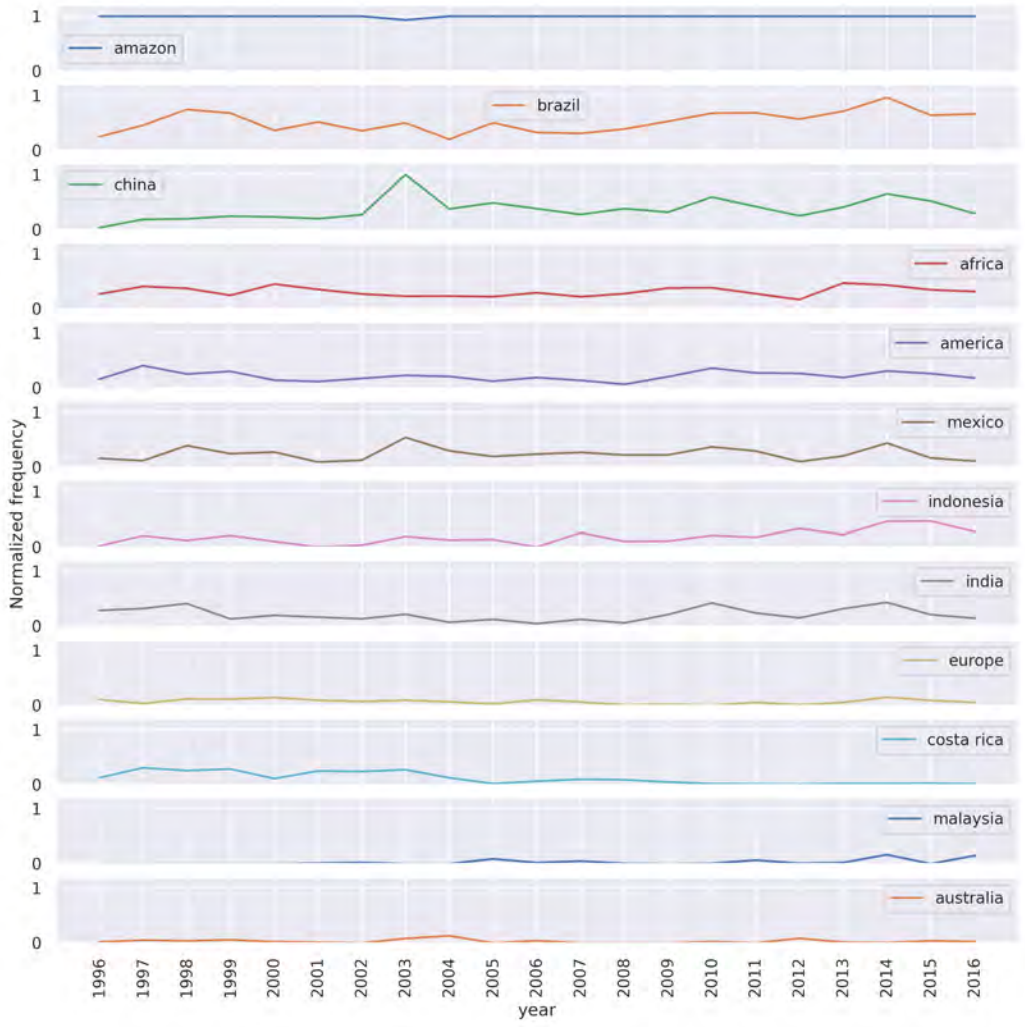

Figure 3.8: Normalized relative frequency of most mentioned areas and regions. The normalize frequency is the number of times the country/region is mentioned in the corpus for the year divided by the total number of terms for that year. The values shown are multiplied by 10^4. Amazon is the most mentioned region every year except 2003 where it was surpassed by China. Brazil and China are the following two most mentioned countries. Mentions of Costa Rica have fallen to very low levels from 2009, while Malaysia is seeing and uptick in numbers of mentions, in the later years. Other countries seeing increased mentions in later years compared to earlier years are India and Indonesia.

"forest", "land cover", "change", which are present in practically all the documents. For each year, we look at the top five keywords and keep five in total, picking roughly one word for each time period. While this does not give us all the important keywords, it still allows us to look into some of the important topics and their evolution overtime. The list of the keywords selected for each country is shown in Table 3.5. We tracked certain words such as "climate", "carbon" and "conservation" for most countries/regions because they were among the most frequent words year after year. Some countries/regions had top words that were particular and rarely if ever found in other countries such as "mangrove" for Australia and "fuelwood" for Africa and "pollen" for Europe.

Figure 3.9, shows the evolution of the five keywords for Indonesia by year. Indonesia is the country on our list with the highest forest conversion ratio for 2016 as reported by the FAO[7].

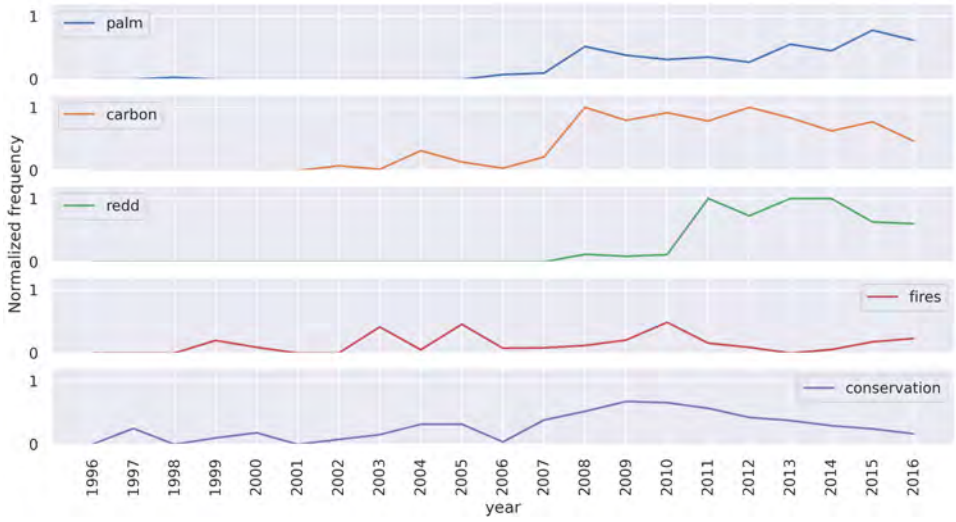

Figure 3.9: A sample of five of the top yearly keywords for the Indonesia, shown with their normalized frequency over time. The normalize frequency is the number of times the keyword is counted for the year divided by the total number of terms for that year in all the publications where "Indonesia" is present. The values shown are multiplied by 10^4. "Palm" and "carbon" are showing high counts mostly in the second half of the timeline. "Redd" also reaches high counts from 2011 onward. "Fires" is present throughout the timeline at moderately high levels with a few peaks in 2003, 2005 and 2010, and its trending upward again into 2015 and 2016. "Conservation" had its highest counts from 2008 to 2011 and then it went on a downtrend.

[7]FAO. Emission database. License: CC BY-NC-SA 3.0 IGO. Extracted from: http://www.fao.org/faostat/en/#data/GF/. Date of Access: 22-06-2021.

amazon	brazil	china	africa
1. pasture	1. pasture	1. conservation	1. climate
2. agriculture	2. carbon	2. climate	2. carbon
3. redd	3. conservation	3. carbon	3. redd
4. carbon	4. redd	4. fragmentation	4. fuelwood
5. biomass	5. amazon	5. erosion	5. nitrogen

america	mexico	india	europe
1. climate	1. habitat	1. conservation	1. climate
2. conservation	2. conservation	2. climate	2. governance
3. biodiversity	3. carbon	3. carbon	3. carbon
4. carbon	4. biodiversity	4. biodiversity	4. pollen
5. logging	5. community	5. fuel	5. conservation

indonesia	malaysia	australia	costa rica
1. palm	1. palm	1. carbon	1. conservation
2. carbon	2. redd	2. habitat	2. environmental
3. redd	3. carbon	3. climate	3. biodiversity
4. fires	4. conservation	4. mangrove	4. carbon
5. conservation	5. management	5. environment	5. forestry

Table 3.5: The five keywords tracked for each country/region. Those keywords are taken from frequent keywords in the early, mid and late years, for each country. Certain words like "conservation" and "carbon" are common to most countries/regions. Certain keywords stand out as being more specific to a country/region such as "fuelwood" for Africa, "fragmentation" for China, "habitat" for Mexico and Australia, "fuel" for India, "pollen" for Europe, "palm" for Indonesia and Malaysia, "fires" for Indonesia, and "mangrove" for Australia.

3.4.4 The disciplines represented

The network of the most represented scientific disciplines is built with the categories defined by the Web of Science (WC field). Since a publication can be found in several categories, it is possible to calculate the co-occurrences of the categories and to use the result to build a network. Thus, it emerges that most of the disciplines are grouped around the environment and ecology. A second group is also formed which relates to medicine, comprising in particular tropical medicine and parasitology.

Table 3.6 shows the disciplines in which the most publications have been classified (WC field of WoS). Thus, the majority of publications are classified under the category of environmental sciences. By looking at the other disciplines it is possible to see which sub-disciplines of environmental sciences are the most represented. Ecology and geosciences have the most publications. We can conclude that publications on deforestation tend to be interdisciplinary, with environmental sciences themselves being interdisciplinary and bringing together, among other disciplines, ecology and geosciences.

	publications
Environmental Sciences	3890
Ecology	2812
Geosciences, Multidisciplinary	1571
Environmental Studies	1491
Biodiversity Conservation	1298
Forestry	1262
Meteorology & Atmospheric Sciences	1104
Geography, Physical	907
Remote Sensing	901
Multidisciplinary Sciences	807

Table 3.6: Most of the publications in our corpus on deforestation are in the Environmental Sciences category. The 10 categories (WC field of Web of Science) with the most publications are shown in the table.

3.5 Conclusion

We performed text mining of data and metadata on scientific publications related to deforestation, providing an overview of the evolution of research activity on this subject over the years, as well as the almost generalized collaboration between countries, and the many collaborative networks between authors. An almost regular increase in the number of publications is observed from one year to the next. Brazil stands out as an important player both through the contribution of its authors and through the references made to the country in the publications analyzed.

By performing these analyses, we found that we can automatically find names of locations along with other words describing some particular topic related to deforestation in that area. Therefore, such corpus can be used to find annotations for satellite images on which deforestation can be observed visually. This can be done even if the text is not a direct description of the images but only related to them based on the location, the time and the event of interest.

The data collected augmented with external data from the Organization of the United Nations for Food and Agriculture and the World Bank help to answer questions related to the link between the research theme being investigated, namely deforestation, and the way it is experienced in a country or region based on this external data. The extraction of named entities such as countries and other locations facilitates the identification of the most concerned areas, and it also allows to find the other topics related to deforestation in those areas as well as how the mention of those topics evolved over time. We found that some of the top topics were common for most countries and most time periods. However some special topics emerged for certain countries and periods. Overall we have found that in our corpus, which we built with publications related to deforestation, the terms "deforestation" and "change" have a very high correlation close to one. This tells us that we can indeed, from the text, confirm that deforestation is seen as a "change" happening. We know, based on the definition from the FAO, that deforestation is "a change in the land use from forest to non-forest." Therefore, we can anticipate that we can link the topics associated with deforestation in the text to changes happening in forests areas from satellite images. We can attempt to establish this link by using images and text that are related to the same area and to the same time period. Which is why we present here this analysis of areas, and their related topics as they are mentioned with time.

In the following chapters, we will compare our corpus with another defor-

estation related corpus covering more recent years and we will explore various methods for extracting the top keywords from them (in Chapter 4). Then, we will make the link between keywords extracted from the corpus introduced in this chapter with change detection on satellite images through image annotation (in chapter 5).

Chapter 4

Corpus Keyword Extraction

Summary

Abstract.

When we are trying to find a set of keywords to represent a whole corpus of scientific publications, we call it corpus-level keyword extraction, as opposed to publication-level keyword extraction where a set of keywords is extracted for a single publication. Extracted keywords can provide an overview of the content of a corpus without having to manually inspect each publication separately. If the corpus is very large, and the keyword extraction method of choice requires that each publication be processed separately, there may be a need for reducing the size of the corpus, to decrease processing time. This corpus size reduction should be done without losing the most representative keywords, thus the most relevant publications, in the process. We show that in the case of a topic-specific corpus, like a corpus on deforestation, by keeping only the publications with titles that are most similar to the topic in terms of their vector representations, we can successfully extract top corpus-level keywords, and even improve on the precision at 25 scores of several keyword extraction methods compared to using all the publications in the corpus. We obtain this result with a variety of keyword extraction methods from word frequency counts to sentence-embedding-based methods, while reducing the corpus size to 76%-36% of its full size. We conclude that we could keep the reduced set of publications for combining it with images.

4.1 Introduction

How do two scientific corpora on the same topic compare? Does it make a difference whether one corpus or the other is used for extracting words relevant to the topic, or will both corpora give similar results? Can they be used interchangeably in downstream tasks such as information retrieval? In this chapter we present a method for comparing corpora on the same topic using keyword extraction methods along with bibliometric statistics. By automatically extracting the top keywords for a corpus, we are performing a task similar to topic modeling [Blei 2003] or keyword-based document cluster labelling [Grootendorst 2020a].

Our approach uses word and sentence similarities as a pre-processing step to filter scientific documents before performing keyword extraction. While the whole corpus is relevant for the comparison, when dealing with large corpora it might be useful to reduce the size before applying keyword extraction. A smaller corpus can reduce the time/resources needed to extract keywords. Reducing the corpus size can therefore be particularly useful when dealing with very large corpora. By using only the title of the documents and computing their similarity with the topic of the corpus, we are able to effectively select a sub-corpus within the main corpus on which keywords can be extracted. This process can be done as a preliminary step before or instead of more extensive or computationally expensive analyses. By using only the title of the publications in this preliminary step, we are processing much less data than if we were using the abstract and/or text body.

The documents with titles highly similar with the topic are assumed to be the ones most relevant to the corpus topic and therefore most likely to contain the main keywords that would best represent the corpus. The documents with titles that have a low similarity with the topic are assumed to be those that are the least representative of the topic of the corpus. In brief, the method is as follows: first, compute the representation of the title and the topic word with a word embedding model (if using word embedding representations instead of character strings); then compute the similarity measure of each title representation with the topic representation; finally, take the top titles by similarity value. We test two vector representations, namely fastText [Bojanowski 2017] and BERT [Reimers 2019], along with two distance measures, cosine, and Wasserstein. We also test with Levenshtein distance of the words as character strings directly. We evaluate quantitatively using the precision at 25 and at 50, qualitatively we visually compare the top keywords extracted by each method. Our results show

that we can reduce the corpus size to about 76%-36% of its original size and obtain the same of higher precision at 25 or 50 depending on the extraction method used.

4.1.1 Motivation and Objective

Keywords extracted from a multi-document set or a corpus can provide a compact summary of the content of that corpus.

Extracting information from a large corpus of scientific documents is a useful but tedious task for humans to do manually. It is therefore important to have high performing automatic methods to extract relevant keywords for a large collection of documents. The task of extracting key terms for a corpus is done for various purposes such as topic modeling [Blei 2003] or keyword-based document cluster labelling [Grootendorst 2020a]. Reducing the number of publications in the corpus can be beneficial as it reduces processing time if enough publications are removed. However, it is important in removing publications from the corpus that the quality of the keywords subsequently extracted is not negatively affected.

Our goal is to later use a corpus in visual semantic learning, where we will be matching words extracted from the corpus with a set of images of forested areas undergoing change. Therefore, we want to create a topic-specific word embedding space matching the topic related to the images we will be using in subsequent tasks. The corpora used to train the word embeddings can impact results obtained when using those embeddings in a downstream natural language understanding task. Better results were obtained with embeddings trained on a specialized corpus as shown by [Hadifar 2018, Dal Pont 2020, Neuraz 2020]. It is therefore important to use a corpus that is well aligned with our topic. We can assess the alignment with our topic of interest by examining top keywords extracted from the corpus.

Furthermore, we seek to limit the impact of publications that might not be as related to the topic as the rest of the corpus. Those publications might yield keywords that are not the most relevant to the topic of the corpus. We therefore want to detect them and remove them from the corpus before performing keyword extraction. In the era of big data, we want to do this in an efficient manner using the least data possible on each publication. To address this issue, we propose a title-topic selection approach (TT-SS) to select only the publications with titles that are close to the topic semantically as measured by the distance between their embeddings. The core idea is to use the title as a summary of the content

of the publication and select the publication based on its title alone. While the abstract of a publication provides a more complete summary than its title, using the abstract in the selection process would require processing much more data per publication which quickly adds up when using a very large corpus.

We consider the two related tasks: publication selection and keyword extraction. First, the selection of publications (to include in the subsequent keyword extraction task) which is done by measuring the similarity between each publication title and the keyword representing the topic of the corpus. Then, from this sub-corpus we extract the top keywords. Our objective is to reduce the size of the original corpus in a significant way while maintaining or improving on the performance of the keyword extraction task.

Our proposed method is specifically targeted at extracting keywords to represent a corpus as opposed to extracting keywords for a single document. We consider corpus-level keyword extraction to be an aggregation of document-level keyword extraction. A corpus being a collection of documents, corpus-level keywords can be viewed as a collection of document-level keywords. In the same way that a set of keywords extracted from a single document can represent that document, a set of keywords extracted from a corpus can represent that corpus. Corpus-level keyword extraction might be of interest to those working with corpora and wanting to extract information from these corpora without having to look at individual documents. Corpus-level keywords can provide a quick way to evaluate the content of a corpus as a preliminary step before deciding to look into the corpus further.

The top keywords extracted from a corpus related to our main topic of interest, namely deforestation, can also provide information on locations affected by this phenomenon and other related sub-phenomena. In this work we are mainly interested in corpora on deforestation and on how to extract the most relevant keywords from them. However, we propose a method that can be applied to any other topic.

For evaluation, we use data from the Web of Science[1], Elsevier Open Access Journals [2] and Pubmed[3]. We use an aggregation of the author keywords as the ground truth for the keyword extraction task. We measure the performance of our proposed selection method by the performance of keyword extractions methods on the original corpus and the sub-corpus. We keep the author keywords from

[1]https://www.webofscience.com/
[2]https://www.elsevier.com/open-access/open-access-journals
[3]https://pubmed.ncbi.nlm.nih.gov/

the original corpus as the reference to which we compare the extracted keywords from the sub-corpus created with the title-topic selection approach.

4.1.2 Overview of Our Method

Fig. 4.1 shows an overview of our publication selection and keyword extraction method. A corpus of publications on a specific topic is the input of our method. We use a similarity measure to find the titles that are most similar to the defined topic, represented by a key term. If the similarity is over a certain threshold we select the publication for our sub-corpus. The later is then used to extract keywords using state of the art methods.

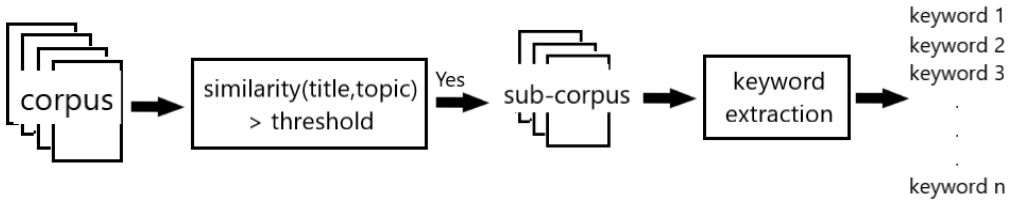

Figure 4.1: Overview of the proposed title topic selection and keyword extraction method. The inputs are the publications from the topic-specific corpus. We select the publications with titles that have a similarity with the topic above a certain threshold. We then extract the keywords from the resulting sub-corpus.

Given a corpus of scientific publications on a specific topic, we use the title of the publications and compare them to the topic itself. We define a threshold for the similarity between the topic and the title of the publications. If the similarity for a title is beyond this threshold, the publication is selected for the sub-corpus that will be used to extract keywords. The keyword extraction is performed only on this sub-corpus using state of the art keyword extraction methods. We describe the core elements of our approach as well as the main task of keyword extraction in the following subsections.

4.1.2.1 Title-Topic Similarity Selection (TT-SS)

We provide a brief introduction to the proposed TT-SS approach here while in Section 4.3 we report on the experiments in which we apply TT-SS to the evaluation datasets (Section 4.2 presents the publication selection process).

In summary, TT-SS reduces the size of a topic-specific scientific corpus by selecting only the publications most similar to the topic of the corpus. It allows

for more efficient corpus-level keyword extraction, as opposed to using the whole corpus, by reducing the amount of data used in the extraction process. With TT-SS as a preprocessing step, the number of correct keywords extracted using unsupervised keyword extraction methods is typically either maintained or increased depending on the method. TT-SS can be use on corpora of different sizes as demonstrated in our experiments in section 4.3. While we apply TT-SS on specific topics in our experiments, this approach can be used with different topics.

The input to TT-SS is a corpus of scientific publications made of documents on the same topic. Once we have the corpus, we get the vector representation of the topic and the titles of the publications through word and sentence embeddings [Bojanowski 2017, Reimers 2019]. We then use distance measures between the topic embedding and the publication title embeddings to find the publications most semantically similar to the topic, using the title as a proxy for the whole publication. The most similar publications found are selected as part of a sub-corpus. We then proceed to use that sub-corpus for our keyword extraction task, having used TT-SS as a preprocessing step.

Sentence embeddings have been used for keyphrase extraction on single documents, using embeddings pre-trained on a large corpora for example by [Bennani-Smires 2018, Grootendorst 2020b]. The distance between the embeddings in the vector space is used to measure their semantic relatedness [Bennani-Smires 2018, Grootendorst 2020b]. We use word and sentence embeddings to represent document titles and corpus topics. We then compute the distance measure to compare publication titles and corpus topics embeddings. We define a threshold for the distance, over which publications are selected as part of the resulting sub-corpus that will be used in the downstream task of keyword extraction.

The word embeddings that we use result from models pre-trained on large corpora from [Bojanowski 2017, Reimers 2019] allowing for domain independent and general purpose embeddings. These embeddings also provide stable results when repeated experiments are performed, due to the fact that the models do not need to be retrained on each new corpus or for each new experiment.

TT-SS works with minimal pre-processing and no training. It is meant to be used for a collection of documents on a specific topic.

4.1.2.2 Extracting corpus keywords

For extracting the keywords we use state of the art keyword extraction methods that perform well with a collection of documents. We propose to reduce the size of the corpus before extracting keywords with TT-SS, in order to reduce the amount of data that needs to be processed and also to increase the number of correct keywords retrieved. To further improve the precision of extracted keywords we propose to combine the keywords extracted with TF-IDF [Jones 1972] with those extracted with BERT embeddings [Reimers 2019, Grootendorst 2020b]. While both methods share most of their top keywords there are a few keywords that the BERT-based method [Reimers 2019] is more likely to find, based on our observations on the corpora we tested. The goal is to find a way to include the best keywords of both methods, while avoiding to add the keywords that were not correctly placed in the top by each method individually. On average, the two methods showed a higher precision at 25 than the other methods that we have tested, for all the corpora used. We propose a way to intersperse the two lists to get the best precision on the final combined list. We chose this approach because it is likely to include, in the resulting list, some words that only BERT had placed in a top position while keeping the highest positioned words from TF-IDF. In many cases this would improve the precision at 25 or at least keep it unchanged. The interspersing also allows us to combine the lists without knowing their composition. The experiments in Section 4.3 demonstrate that our proposed combination method can achieve a good performance in terms of precision at 25 and outperform other state of the art methods.

4.1.3 Summary of Contributions

The following contributions are included in this chapter:

- We conduct a comparison of keyword extraction methods on a set of topic-specific corpora and examine how the choice to build a corpus based on title, abstract, body or keywords impacts the keywords that are extracted.

- We propose a Title-Topic Similarity Selection approach to select publications to use for keyword extraction, which allows to reduce the size of large corpora while keeping or improving the performance of keyword extraction methods [Jones 1972, Bojanowski 2017, Reimers 2019, Campos 2020].

- We examine the influence of selecting publications with different similarity measures on the performance of the title-topic similarity selection.
- We propose novel topic-specific datasets for evaluating keyword extraction methods along with the baseline performance of applied keyword extraction methods [Jones 1972, Bojanowski 2017, Reimers 2019, Campos 2020].
- We propose a combination of the keywords from TF-IDF [Jones 1972] and BERT [Reimers 2019] as an improved keyword list with higher precision, outperforming other methods tested [Jones 1972, Bojanowski 2017, Reimers 2019, Campos 2020].

4.2 Publication Selection

In this section, we present our formulation of the problem of selection publications for creating a sub-corpus, and our proposed solution using word and sentence embeddings with similarity measures.

4.2.1 Problem Formulation

With the TT-SS, our goal is to select from a single-topic corpus a subset that will best describe the corpus based on extracted corpus keywords. We do this by removing publications that have low semantic similarity of their title with the corpus topic. We show examples of titles of publications removed in Table 4.1, titles are shown in decreasing order of their distance (dissimilarity) with the topic.

In most cases, titles encapsulate the essential description of the publication. Semantic similarity allows us to find words and text that are semantically similar. We make the hypothesis that a publication that is most related to the topic of the corpus will have a title that is semantically close to that topic, we therefore use the topic as a proxy for the abstract and the whole body of the publication.

Formally, we are given a corpus of N publications, $C = \{p_i; i = 1, ..., N\}$ and a similarity function F such as $F(p_i) = Similarity(t_i, \tau)$, where t_i is the title of publication i, and τ is the topic of the corpus or its vector representation. We are looking for a subset X of C, by defining a threshold θ such that: $X \subset C$ with $X = \{p_i \mid F(p_i) > \theta \text{ and } P_X >= P_C\}$, where P_X and P_C are the scores of the extracted keywords from the subset X and the corpus C respectively. We define θ based on the histogram of the similarities of the titles with the topic.

Sample titles with low similarity to the topic of "deforestation"	Distance
Macroecological patterns of American Cutaneous Leishmaniasis transmission across the health areas of Panamá (1980–2012)	0.0685
Mapping major land cover types and retrieving the age of secondary forests in the Brazilian Amazon by combining single-date optical and radar remote sensing data	0.0646
Rapid integrated clinical survey to determine prevalence and co-distribution patterns of lymphatic filariasis and onchocerciasis in a Loa loa co-endemic area: The Angolan experience	0.0628
Renting legality: How FLEGT is reinforcing power relations in Indonesian furniture production networks	0.0596
Mapping tropical disturbed forests using multi-decadal 30m optical satellite imagery	0.0506

Table 4.1: A sample of 5 titles from the corpus with high distance to the topic of deforestation show low similarity to the topic, in the Elsevier corpus. The titles are shown ordered by the Wasserstein distance of their BERT embeddings to the embedding of the topic, from most distant to least distant. The higher the distance the lower the similarity with the topic. The titles appear to be somewhat related to forests in general but having a different main focus. Some titles focus on disease, others on mapping land cover and emphasize the technology used.

From this histogram we find a value that separates the majority of the publications with a minority that has lower than average similarity with the topic. We define P_X and P_C as the precision at k for the subset X and the corpus C, which we use to evaluate the keyword extraction.

4.2.2 Text Representation

In order to computer the similarities between the titles and the topic of the corpus, we use different text representations.

When the original character string representation is used, the similarity can be calculated with a string distance function. We might also use word and sentence embeddings to represent the text. Word embeddings, which are vectors of real

numbers representing words, rely on neural network architectures to train vector space models [Mikolov 2013, Bojanowski 2017, Peters 2018, Devlin 2018, Radford 2019]. Words are mapped into a lower dimension space from the higher dimension vector space.

FastText embeddings [Bojanowski 2017], which are character-based, are derived from word2vec [Mikolov 2013] word embeddings. With word2vec, the word representations in vector space are obtained with a two-layer neural network. This network (language model) is trained on a corpus to capture the linguistic context of words. With fastText [Bojanowski 2017], the word2vec model is updated to add subword information by splitting words into character n-grams. This allows the model to handle words that it was not trained on (out of vocabulary) provided they are made of n-grams from known words. One limitation of a word2vec-type models is that for a given word, it will generate a single vector representation. However, within a same corpus, a word might exhibit polysemy, meaning it might have several meanings depending on its context. This is not handled by word2vec.

BERT [Devlin 2018] embeddings are obtained from a pre-trained transformer [Vaswani 2017] model, which is a deep learning model that handles sequential data using a self-attention mechanism. These embeddings allow for a given word to have multiple representation based on its context, if it has several meanings in the corpus on which the model is trained. Sentence-BERT (sBERT) [Reimers 2019] is a modified version of BERT specially designed for sentence-pair regression tasks. It presents a siamese architecture where two BERT networks have shared weights. The sentence embeddings obtained can be compared using a similarity measure between two vectors (like cosine similarity).

We use all three text representations, character strings (which are just the words in their original form), fastText [Bojanowski 2017] embeddings, and BERT [Devlin 2018, Reimers 2019] embeddings, in the different variations of our proposed TT-SS approach.

4.2.3 Similarity Measures

Levenshtein distance. The Levenshtein distance finds the distance between two strings by comparing their characters, computing the number of character edits between the two strings. An edit is a substitution, an insertion or a deletion performed to transform one string into the other. For example to transform the string "car" to "tar", you need to make one substitution and replace "c" with "t", therefore the edit distance is 1. To transform "cover" to "powers" you need to make

two substations and one insertion: "cover" -> "pover" -> "power" -> "powers", a total of three operations for an edit distance of 3. The lower the Levenshtein distance between two strings, the higher their similarity.

Cosine similarity. The cosine similarity is the cosine of the angle between two vectors in an inner product space. Between two vectors X and Y, the cosine similarity $cossim(X, Y)$ is given by:

$$cossim(X, Y) = \frac{X \cdot Y}{||X|| \times ||Y||} \tag{4.1}$$

Wasserstein distance. The first Wasserstein (W_1) distance measures similarity between distributions. Also known as the Earth mover's distance it has been also used for natural language processing tasks such as document classification [Kusner 2015]. In one dimension, with two distribution of the same size n, the distance is given by:

$$W_1(X, Y) = \sum_{i=1}^{n} ||X_{(i)} - Y_{(i)}|| \tag{4.2}$$

A lower Wasserstein distance between two embeddings indicates a higher similarity.

We use the Levenshtein distance with the titles as character strings in one variation of TT-SS. In the other variations we use Wasserstein distance and cosine distance (1- cosine similarity) with both fastText and BERT embeddings.

4.2.4 Empirical Evaluation of Assumptions in TT-SS

In TT-SS we assume that the title of a publication is as a compact summary of the whole document. Compared to the abstract which is a the longer, multi-sentence summary, the title could be viewed as a summarized abstract. We verify this assumption by looking at the histogram of the similarities of the titles with the topic, for three Web of Science corpora (Wos-TC, WoS-KW, WoS-TA) and for one Elsevier corpus (ELS-BIG-TA) in Figure 4.2. For the WoS corpora, we can see that the fastText embeddings result in bimodal distributions with both Wasserstein and cosine distance. In both cases, there is a major and a minor mode with high peaks. The values around the minor mode are the ones with the highest distance thus the lowest similarity to the topic. The BERT embeddings with Wasserstein distance results in a right skewed distribution. A large part of the titles are highly similar to the topic but there is a number of values in the tail corresponding to the titles

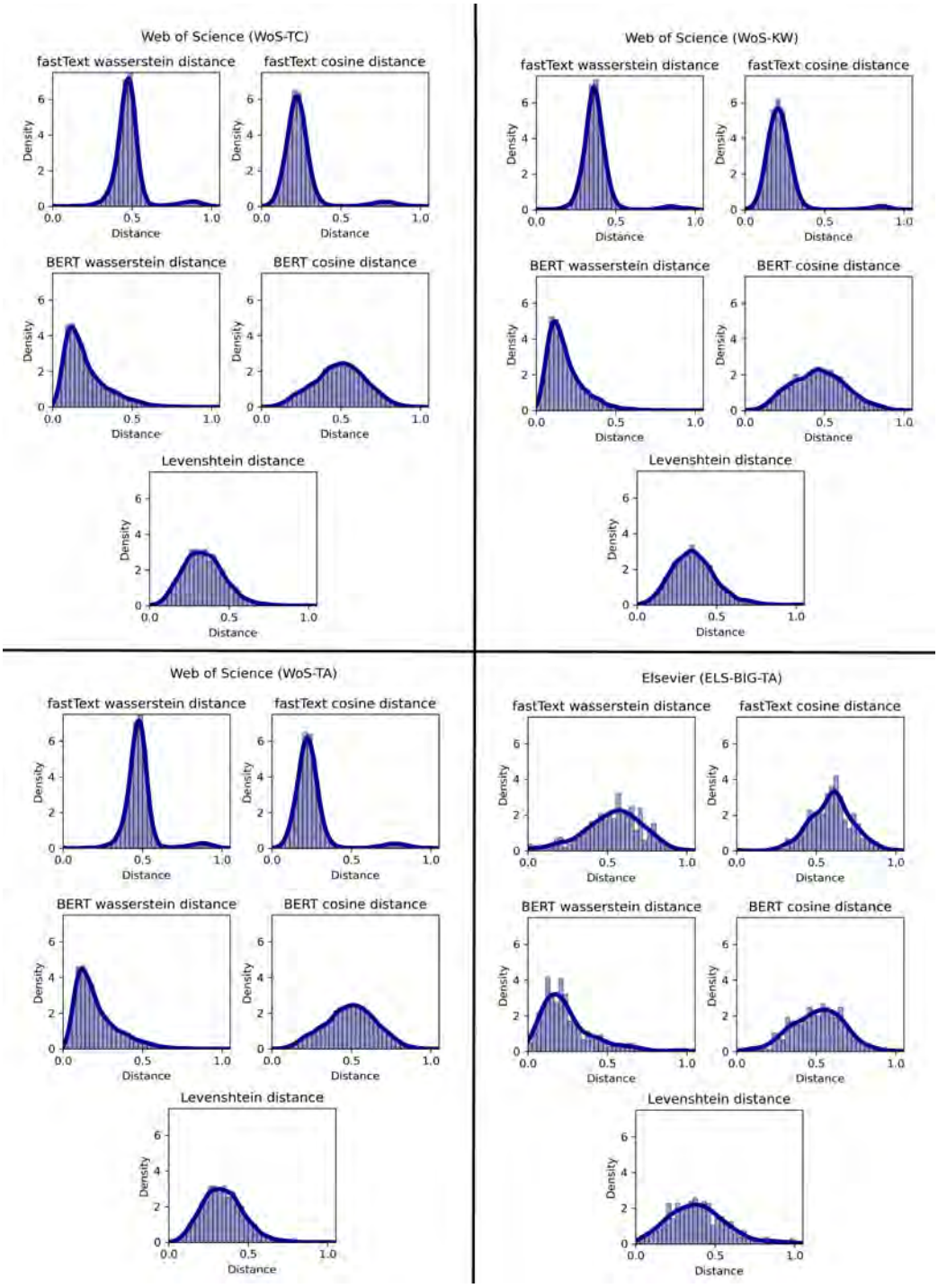

Figure 4.2: The histograms of the distance of publication titles to the topic for four corpora show similar trends when the same embeddings and same distance measures are used.

with the lowest similarities. The BERT embedding with cosine distance results in an almost normal distribution with a low peak, most values are around the mean. The character strings with Levenshtein result in a right skewed distribution. A large part of values are lower than the mean indicating a majority of titles highly similar to the topic and a minority of less similar ones in the right tail. With the Elsevier corpus we do not observe the bimodal distribution with the fastText embedding. All histograms tend to resemble normal distributions, with the BERT embeddings with Wasserstein distance histogram being right skewed as we have seen in the WoS corpora. Similarly, the histogram of the character strings with Levenshtein distance is also right skewed. For all the text representations and distance measures featured in Figure 4.2, there seems to be an agreement on the overall trend of the distributions of the similarities of the titles.

Titles with high similarity with the topic (lowest distances)	Wasserstein distance
Livestock and the Environment: What Have We Learned in the Past Decade?	0.0126
Socio-economic importance, domestication trends and in situ conservation of wild Citrus species of Northeast India	0.0127
Implementation of Forest Canopy Density Model to Monitor Tropical Deforestation	0.0130
Fuel switching from wood to LPG can benefit the environment	0.0134
Tree biomass equations for tropical peat swamp forest ecosystems in Indonesia	0.0140
Titles with low similarity with the topic (highest distances)	**Wasserstein Distance**
REDD Mitigation	0.1318
Avifauna of Hingol National Park, Balochistan	0.1159
Taxonomic observations regarding four genera of Afrotropical robber flies, Choerades Walker, 1851, Laphria Meigen, 1803, Nannolaphria Londt, 1977 and Notiolaphria Londt, 1977, and the description of Ericomyia gen. n. (Diptera, Asilidae, Laphriinae)	0.1151
Reappearance of Anopheles minimus in Singhbum hills of East-Central India	0.1142
New Density Estimates of a Threatened Sifaka Species (Propithecus coquereli) in Ankarafantsika National Park	0.1112

Table 4.2: The titles most similar to "deforestation" are relation to the environment, forests and conservation in general while the least similar titles are, for the most part related to insects. Titles shown are from the Web of Science corpus. The Wasserstein distance of their BERT embeddings to the BERT embedding of the word "deforestation" is also shown.

Table 4.2 shows the titles most similar and least similar to the topic of de-

forestation in the Web of Science (WoS-TC) corpus, based on the Wasserstein distance of their BERT embeddings. We find at the top of the least similar titles, one with only two words and no explicit mention of deforestation. The other least similar titles are focused on animals. The most similar titles are related to forests, the environment and conservation. This seems to validate our assumption that we are able to find the publications that are most closely related to our topic by looking at the titles.

4.3 Experiments

4.3.1 Datasets

Table 4.3 lists the three datasets that we use in our experiments:

(1) **Web of Science Deforestation Dataset (WoS)** [Akinyemi 2018]. We collected publications from the Web of Science on the topic of deforestation. These publication contain the term "deforest*" either in their title, abstract, author keywords or keywords plus®, the later are frequent terms that appear in the titles of a publication's references but not in the publication's own title. We use several subsets of the WoS corpus. The full corpus with the documents matching the topic is referred to as WoS-TC. We extract a sub-corpus where the topic term is included either in the title or abstract and we refer to it as WoS-TA. Another sub-corpus where the topic term is included in the keyword is referred to as WoS-KW. The WoS-TC datasets contains 9722 publications, the WoS-TA 6897, and WoS-KW 2772.

(2) **Elsevier**. The Elsevier OA CC-BY Corpus [Kershaw 2020] is a publicly available corpus of open access scientific research publications in a variety of Elsevier's journals. This dataset is under a creative commons license[4], it contains 40001 articles. We created two subcorpora on the topic of deforestation with this corpus. The first sub-corpus contains publications containing the key term "deforest*" in the title or abstract referred to as ELS-MINI. The second sub-corpus is made of publications with the key term either in the title, abstract or article body, later referred to as ELS-BIG. The total number of publications is 54 for ELS-MINI and 379 for ELS-BIG. In our experiment we use both datasets with and without the full text body

[4]https://data.mendeley.com/public-files/datasets/zm33cdndxs/files/3a6bd579-aed4-48c2-8d86-da7b07b10ca3/file_downloaded/ - Elsevier OA CC-BY Corpus licence

for the keyword extraction task. When only the title and abstract are used, the corpora are referred to as ELS-MINI-TA and ELS-BIG-TA. When the body is used they are referred to as ELS-MINI-TA-B and ELS-BIG-TA-B respectively.

(3) **PubMed**. The PubMed dataset [Aronson 2000]. The PubMed dataset is a collection of publications from PubMed Central. It contains 500 documents from which we select 161 containing our target keyword "protein" among the author keywords. We use the title, abstract and body of the publications, in the keyword-extracting experiments, for this dataset. We use this corpus to show the potential of our proposed approach on other topic-specific corpora non-related to our initial topic of interest namely deforestation.

	Number of Publications	Document Sections	Topic Keyword Found in
Web Of Science			
WoS-TC	9722	title, abstract	WoS topic
WoS-TA	6897	title, abstract	title or abstract
WoS-KW	2772	title, abstract	author keyword
Elsevier			
ELS-MINI-TA	54	title, abstract	title or abstract
ELS-MINI-TA-B	54	title, abstract, body	title or abstract
ELS-BIG-TA	379	title, abstract	title or abstract or body
ELS-BIG-TA-B	379	title, abstract, body	title or abstract or body
PubMed			
PubMed	161	title, abstract, body	author keywords

Table 4.3: List of datasets used for corpus keyword extraction experiments.

Annotation: We use the author keywords as reference keywords for the Elsevier corpora and Web of Science corpora, similarly to [Campos 2020]. For PubMed, the keywords given are the Medical Subject Headings (MeSH), which is a controlled vocabulary thesaurus that is used to index the PubMed publications.

4.3.2 Keyword Extraction Methods

We perform our experiments using a variety of keyword extraction methods. We use statistical methods including word frequency, TF-IDF [Jones 1972], and YAKE

[Campos 2020], embedding-based methods with fastText [Bojanowski 2017] and BERT embeddings [Reimers 2019], and a mixed graph-based with embeddings [Mothe 2018].

Frequency Method. The most straightforward method to extract keywords from a document is to count the frequency of the words present within the document. It is one of the simplest statistical methods for extracting keywords. To get the frequency of the words for the whole corpus, we combine all the publications into a single document. We then count each word and list them from most frequent to least frequent.

TF-IDF. The term frequency inverse document frequency or TF-IDF [Jones 1972] is a statistic that estimates how important a term is in a document while taking into account the whole corpus. We use Scikit-learn [Pedregosa 2011] to compute the TF-IDF statistic, for each word and each publication in the corpus, resulting in a matrix of TF-IDF features. We then sum the TF-IDF over all the documents and finally obtain the list of keywords from highest to lowest summed value. Let t be a term, d a document, n the total number of documents. The frequency of a term in a document is noted $tf(t, d)$. The inverse document frequency of a term $idf(t)$ is given by: $idf(t) = \log \frac{1+n}{1+df(t)} + 1$. Where $df(t)$ is the number of documents in the corpus containing t. The TF-IDF of a term for a given document is thus $tf - idf(t, d) = tf(t, d) * idf(t)$.

YAKE. YAKE! [Campos 2020] is a keyword extraction method based on statistics and heuristics. For a given document YAKE extracts the relevant keywords following a number of predefined steps such as preprocessing, feature extraction and term ranking. Among the features used by YAKE are the term casing, the position and a normalized frequency. Other features such as the number of times the term appears in different sentences and the number of different terms it appears with, are also used. YAKE outputs the keywords from most important to least important. We use the YAKE python package[5] in our experiments. We extract the YAKE keywords for each document and then aggregate the results for the whole corpus by taking the most frequent keywords.

fastText embeddings. We use fastText [Bojanowski 2017] word and sentence embeddings in a manner similar to [Bennani-Smires 2018] who used Word2vec-based [Mikolov 2013] embeddings to embed documents and candidate keywords. We use a publicly available pre-trained fastText model[6] which was trained on

[5]YAKE! online repository - https://github.com/LIAAD/yake

[6]Pre-trained fastText model - https://fasttext.cc/docs/en/crawl-vectors.html

Wikipedia[7] and Common Crawl[8]. We calculate the mean vector of document sentences to get the document embedding. We calculate the cosine similarity of each candidate keyword with each document and the most similar keywords are returned as the top keywords for the document. We then aggregate the result for the whole corpus, taking the most frequent top keywords.

BERT embeddings. We also extract keywords with the more recent transformer-based embeddings [Devlin 2018] similarly to [Grootendorst 2020b]. We use a BERT-based pre-trained sentence transformer model [Reimers 2019] to embed the words and the documents that has been shown to perform well on semantic similarity tasks[9]. As we do with fastText vectors, we find the candidate keyword that are most similar with the documents based on cosine similarity. We get the overall result for the whole corpus by taking the keywords that most frequently appear among the top for all publications.

Graph-based plus word embeddings. In a graph-based method, a non-directed graph is built with the adjectives and nouns obtained from part-of-speech tagging. The nodes are connected based on cooccurrence in a window of words within a document. Node (word) ranking is performed based on graph-based ranking algorithm with adjacent nodes as candidate keyterms. When multiterms (keyphrases) are present, their ranking is equal to the sum of their single terms' ranking. The keywords and keyphrases are extracted based on their ranking from highest to lowest. When word embeddings are added to the graph [Mothe 2018], instead of character strings, the cosine similarity is used along with cooccurence as the edge weights, to connect the nodes, and also for the ranking algorithm.

Our combined TF-IDF and BERT. We propose two methods to combine the outputs of the keyword extraction methods based on TF-IDF and BERT embeddings into a single list of keywords. First, we use the two methods as described previously then we combine the resulting keyword lists by interspersing one list into the other. The motivation for this combination is the fact that both methods often perform better than the others, in our experiments, however they sometimes differ in the words they correctly find in their top 25. We want to find a way to keep as many correct top 25 words as possible by looking into both lists. The challenge is that we do not know, in advance, for any position in the list whether the word is correct or not. Our goal is therefore to combine the list by increasing

[7]Wikipedia - https://www.wikipedia.org/

[8]Common Crawl - https://commoncrawl.org/the-data/get-started/

[9]Sentence transformer model - https://huggingface.co/sentence-transformers/xlm-r-bert-base-nli-stsb-mean-tokens/tree/main

the number of words correctly put in top positions. We propose to do this by interspersing every other element of the BERT list into the TF-IDF list, starting the new list with the first element of the TF-IDF list. After this combined list is created we keep only the unique values. The second way in which we combine the two methods is by defining a window of three elements from the BERT list and randomly picking one element to intersperse into the TF-IDF list, sliding over one element at a time and not picking the same word twice. Both combinations work well, and perform comparably, we prefer the non-random list as its outcome is more predictable. Table 4.4 contains the list of the keyword extraction methods that we use.

Keyword extraction method
Frequency
TF-IDF [Jones 1972]
YAKE! [Campos 2020]
fastText embeddings [Bojanowski 2017, Bennani-Smires 2018]
BERT embeddings [Devlin 2018, Grootendorst 2020b]
Graph plus word embeddings [Mothe 2018]
Combined TF-IDF and BERT embeddings *
Combined TF-IDF and BERT embeddings with random pick *

Table 4.4: Keyword extracted methods used on the evaluation corpora. (*) Denotes our proposed methods.

4.3.3 Experiment I: Corpus Keyword Extraction

In this experiment we conduct keyword extraction on the Elsevier, Web of Science and PubMed corpora, previously described in Section 4.3.1. We take all the documents in each corpus and extract keywords from them, using the methods described in Section 4.3.2. For each method we get the list of keywords in order of importance.

Evaluation Protocol. When evaluating the performance of the keyword extraction methods, we use single terms and therefore split keywords when they contain multiple words. While we could have matched using multiterm keywords, we perform this evaluation with the more straightforward single terms allowing us to have keyword lists made of unique words. We compute the relative

frequency of each keyword. We take the frequency of the keyword and divide it by the number of publications with that keyword. We use the precision at k to evaluate the performance of each keyword extraction method. We use 25 and 50 as the values of k, which are the number of keywords to consider in the evaluation. The average number of keywords per publications varies from 5.7 to 16.6 depending on the corpus, including both single-term and multiterms, and between 10.4 and 34.6 when considering single terms only.

[Dieng 2020] defined the diversity of topics generated by a topic model as the percentage of unique words found in the top 25 words of all the topics. We make a similar assumption that the top 25 (and 50) words produced by our keyword extraction methods inform us on the quality of the whole list. As in topic models, our keyword lists, for the most part, could be as long as the number of unique words in a given corpus. Therefore we need to set a cutoff number k at which we can evaluate them and obtain results that can also be qualitatively evaluated by visually examining the keyword lists. The top 25 keywords provide a good summary of the dominant themes of the corpus. The top 50 words show how well the methods work as the expected number of keywords gets longer.

With the precision at k for each method and each corpus, we establish a baseline, to which we will compare the values after using TT-SS (in subsequent experiments, in Sections 4.3.4 and 4.3.5).

The author keywords serve as ground truth, they are therefore the reference keywords to which we compare the keywords returned by the keyword extraction methods that we tested. Because we want to evaluate the ability of the methods to extract keywords representing the whole corpus, we aggregate the publication author keywords. This aggregation is done, for each corpus, by computing the TF-IDF of the author keywords for all the publications in the corpus. The reference keywords are listed in order of the sum of their TF-IDF over all the publications, from highest to lowest. This provides us with an aggregated author keyword list that is slightly different from a pure frequency-based list. It will lower the rank of potential corpus stop words that might be included in the author keywords, which might happen if the list was built based on the frequency.

To compute the precision at k we take the first k words returned by each keyword extraction method and we compare them to the first k ground truth words. The precision at k is given by: $precision\ at\ k = \frac{true\ positives\ at\ k}{true\ positives\ at\ k + false\ positives\ at\ k}$, where the true positives at k are the keywords, among the top k, that match the ground truth, the false positives at k are the non-matching keywords, among the top k.

We find the precision at k appropriate as it rewards words correctly placed in the top k positions without penalizing based on order. Consider the following ground truth keywords in the top 3: "forest", "conservation", "land". Two methods returning "land", "forest", "conservation" and "conservation", "land", "forest" would both have a score of 1 (3/3) for precision at 3. Taking the order into account would have given both lists a score of 0 (0/3). Since we are only evaluating on a limited number (k) of returned keyword and not on the complete list of extracted keywords we do not run the risk of having most methods returning all the ground truth words, which most do eventually, albeit out of order.

Tested Methods. All methods previously described in 4.3.2 are compared on their performance on each corpus, namely:

(1) Total frequency computed over the whole corpus as a single text;

(2) TF-IDF [Jones 1972] sum where the TF-IDF value for each word is calculated for each document and then summed across all documents;

(3) YAKE [Campos 2020] which uses statistical features to extract keywords from each document;

(4) Embedding based keyword extraction with BERT [Reimers 2019, Grootendorst 2020b] finding within each document the keywords that are semantically most similar to it, using BERT from a pre-trained model;

(5) fastText [Bojanowski 2017, Bennani-Smires 2018], similarly to (4), finding within each document the keywords that are semantically most similar to it, using pre-trained fastText embeddings, and cosine similarity between the embeddings;

(6) Graph-based key phrase extraction with word embeddings [Mothe 2018];

(7) Combined TF-IDF [Jones 1972] and BERT [Grootendorst 2020b], where the resulting keyword lists from (2) and (4) are combined into one list.

Parameter Settings. For all methods, except (1), the keywords are extracted per publication. All the keywords are combined into a common list and the number of occurrences of each keyword is counted. The keywords are then placed in order of their count, from highest to lowest, in the final list for a given keyword extraction method. For YAKE we use default parameter values except for the number of keywords that we set to 50. Max ngram range is 3, deduplication threshold is 0.9, deduplication function is seqm and window size is 1.

For all methods, we remove stop words from NLTK [Bird 2009] and the scikit-learn's [Pedregosa 2011] English stop words lists. All the text is converted to lower case. When keyphrases with more than one word are returned we break them up into individual keywords.

Precision at 25								
Corpus	Freq.	TF-IDF	YAKE	BERT	fastText	Garph + Emb.	TF-IDF + BERT	Random TF-IDF + BERT
WoS-TC	0.64	**0.68**	0.60	0.64	0.56	0.64	**0.68**	**0.68**
WoS-TA	0.60	**0.68**	0.60	0.64	0.56	N/A	0.64	0.64
WoS-KW	0.60	**0.68**	0.56	0.64	0.48	N/A	**0.68**	**0.68**
ELS-MINI-TA	0.52	0.56	0.52	0.56	0.52	0.48	**0.60**	**0.60**
ELS-MINI-TA-B	0.36	0.40	0.44	0.44	0.36	N/A	**0.48**	0.44
ELS-BIG-TA	0.48	0.60	0.60	**0.72**	0.44	0.48	0.68	0.64
ELS-BIG-TA-B	0.36	0.52	0.52	**0.72**	0.28	N/A	0.60	0.56
PubMed	0.40	0.40	0.44	0.40	0.24	N/A	**0.48**	**0.48**
Precision at 50								
WoS-TC	0.58	0.62	**0.64**	**0.64**	0.54	0.58	**0.64**	**0.64**
WoS-TA	0.56	0.60	0.60	0.62	0.56	N/A	**0.64**	**0.64**
WoS-KW	0.52	0.58	0.56	0.58	0.52	N/A	**0.60**	**0.60**
ELS-MINI-TA	0.40	0.46	0.48	**0.56**	0.42	0.46	0.52	0.52
ELS-MINI-TA-B	0.38	0.38	0.40	**0.42**	0.30	N/A	0.40	**0.42**
ELS-BIG-TA	0.62	0.64	0.64	**0.70**	0.62	0.62	0.68	0.66
ELS-BIG-TA-B	0.48	0.54	0.54	0.60	0.38	N/A	**0.62**	0.58
PubMed	0.32	0.34	**0.42**	0.38	0.24	N/A	0.34	0.36

Table 4.5: TF-IDF outperforms other methods in precision at 25 for the largest corpus from the Web of Science (WoS-TC) and its derived corpora (WoS-TA and WoS-KW), without TT-SS. For the other corpora, the BERT-embedding-based method outperforms or performs equally as TF-IDF. The combined TF-IDF and BERT methods have comparable performances. They outperform other methods on the smaller Elsevier corpora (ELS-MINI) and on PubMed. They are outperformed by BERT on the larger Elsevier corpora (ELS-BIG-TA and ELS-BIG-TA-B). At 50, YAKE outperforms all the other methods on PubMed. BERT outperforms the other methods on two Elsevier corpora and the combined TF-IDF and BERT methods outperform the other methods on two Web of Science corpora and one Elsevier corpus. The highest precision for each corpus is in bold. The methods shown in the columns, in order, are: Frequency, TF-IDF, YAKE, BERT, fastText, Graph method with word embeddings, TF-IDF combined with BERT method, and TF-IDF combined with BERT random pick.

Comparison Results. Table 4.5 shows the precision at 25 and 50 for each corpus and each keyword extraction method. We see that adding the body reduces the performance of all methods. There is an overall decline in performance

between the precision at 25 and at 50 when the body of publications is included in the corpus. The performance on the PubMed corpus is relatively low for all methods, with the max precision at 25 of 0.44 for YAKE. This might be due to the fact that the PubMed corpus contains the body of the publications unlike the WoS corpora for example. The combined TF-IDF and BERT methods have the highest precision at 25 on the PubMed corpus. YAKE has the highest precision at 50 on the PubMed corpus. No single method outperforms all the others for all the corpora. This shows that the keyword extraction method is dependent on the characteristics of the corpus it is being used on, including the number of publications in the corpus and whether the body of the publications are included or not. The larger corpora tend to have higher precision values than smaller ones and the corpora without the publication body tend to have higher precision than corpora with the publication body. Our two proposed methods which combine the results of TF-IDF and BERT methods have overall higher precision at 25 than the other method. Our first proposed method reached the top precision at 25 five out of eight times and our second method four out of eight times. Comparatively, TF-IDF has the highest recall at 25 three times, and BERT only twice. At 50 our proposed methods reach the top precision four times, once on an Elsevier corpus and on all three WoS corpora. BERT also has the top precision at 50 four out of eight times, but on one WoS corpus and three Elsevier corpora.

4.3.4 Experiment II: Title-Topic Similarity Selection

In this experiment, we conduct publication selection using TT-SS to create a sub-corpus prior to performing keyword extraction on this sub-corpus. We include both quantitative evaluation based on precision at 25 and qualitative evaluation based on the top 25 keywords from each method. We perform the experiments on the same datasets, as in experiment I (in section 4.3.3).

1) ***Quantitative Evaluation.*** We show that our proposed sub-corpus creation approach with TT-SS can improve the precision at 25 for the corpus keyword extraction task when used as a preprocessing step. After performing TT-SS, we apply five of the keyword extraction methods described in section 4.3.2: the frequency method, TF-IDF [Jones 1972], YAKE [Campos 2020], BERT [Grootendorst 2020b] and fastText [Bojanowski 2017, Bennani-Smires 2018]. We measure the precision at 25 on the outputs of each method.

Tested Methods. We test several variations of TT-SS with different text

representations and distance measures as previously described (in section 4.2). In this section, we report the results for the following versions of TT-SS:

(1) TT-SS with BERT embeddings (TT-SS BERT) and Wasserstein distance, which selects publications based on the Wasserstein distance between the BERT embedding of their title and the BERT embedding of the topic.

(2) TT-SS BERT with cosine distance, which is similar to (1) but uses the cosine distance (1- cosine similarity).

(3) TT-SS with fastText embeddings (TT-SS fastText) and Wasserstein distance, which selects publications based on the Wasserstein distance between the fastText embedding of their title and the fastText embedding of the topic.

(4) TT-SS fastText with cosine distance, which is similar to (3) but uses cosine distance (1- cosine similarity).

(5) TT-SS fastText and BERT with Wasserstein distance, which keeps the publications that were selected by both TT-SS BERT and TT-SS fastText with Wasserstein distance in (1) and (3).

(6) TT-SS fastText and BERT with cosine distance, which is similar to (5) but keeps the common selections of TT-SS BERT and TT-SS fastText with cosine distance in (2) and (4).

(7) TT-SS with Levenshtein distance, which selects publications based on the Levenshtein distance of their title with the topic.

Parameter Settings. We set the threshold at the mean value of the distribution of the similarities in all versions of TT-SS except for TT-SS fastText and BERT with Wasserstein distance selection (5). We therefore select the publications with a title that is a distance greater than the mean of the distances of all titles to the topic. This threshold allows us to keep only the publications with more than average similarity to the topic based on the distance measure. These publications are more likely than the non-selected ones to contribute topic-relevant keywords, when we apply keyword extraction. For the TT-SS fastText and BERT with Wasserstein distance selection (5), we set the similarity threshold θ at the mean value plus one standard deviation of the similarities between the titles and the topic, for a given corpus. When the threshold is set at the mean for (5), the number of selected publications is very low. Adding one standard deviation to the mean of the distance allows for a higher number of publications to be selected.

Results. Table 4.6 shows the results of two versions TT-SS versions with Wasserstein distance (TT-SS BERT and TT-SS fastText+BERT), on all the corpora we tested, for five keyword extraction methods (Frequency, TF-IDF, YAKE, BERT and fastText). We see that most methods reach a higher precision at 25, on average, either with TT-SS BERT or with TT-SS fastText+BERT. The only exception being ELS-MINI-TA-B and ELS-BIG-TA-B which do not improve, on average, because the best performing methods for these two corpora, namely YAKE and BERT, under-perform with TT-SS. Methods that start with a relatively low value for precision at 25 more often show increased values of precision after applying TT-SS, such is the case with Frequency and fastText methods. 10 out 16 cases of using TT-SS result in an increase in precision at 25 with the Frequency. 15 out of 16 cases result in an increase in precision at 25 with the fastText method with TT-SS. For methods that have highest precision at 25 without TT-SS, such as TF-IDF and BERT, there are more instances where the value of the precision at 25 with TT-SS remains the same. For these methods this results in fewer increases in precision at 25 with TT-SS. For TF-IDF there are 2 out of 16 increases and 4 out of 16 for BERT. Compared to 10 out of 16 instances of similar precision with and without TT-SS for TF-IDF and 7 out of 16 for BERT. The TT-SS fastText+BERT version with Wasserstein is made in such a way that it results in a higher number of publications being selected than TT-SS BERT with Wasserstein. In some cases this results in higher precision at 25. It is particularly effective on the PUBMED corpus where all keyword extraction methods increased their precision at 25 with TT-SS fastText+BERT compared to not using TT-SS.

To further compare the effect of using TT-SS, we also show the precision at 25, for all the versions of TT-SS on one Web of Science corpus (WoS-TA) corpus in Table 4.7. On average, compared to not using TT-SS, the precision at 25 increases with TT-SS whichever variation of TT-SS is used, with the exception of fastText embeddings with Wasserstein distance. This result is even more impressive considering in some cases less than half the corpus remains after the selection. The combiend fastText and BERT embeddings with cosine similarity configuration results in only 36% of the original corpus being used to extract the keywords.

Corpus	Number of Publications	Freq.	TF-IDF	YAKE	BERT	fastText
WoS-TC	9722					
Without TT-SS	9722	0.64	0.68	0.60	0.64	0.56
After TT-SS BERT	6065	0.60	0.68	0.60	0.68	0.60
After TT-SS fastText+BERT	7185	0.60	0.68	0.60	0.64	0.60
WoS-TA	6897					
Without TT-SS	6897	0.60	0.68	0.60	0.64	0.56
After TT-SS BERT	4298	0.64	0.68	0.64	0.72	0.60
After TT-SS fastText+BERT	5064	0.60	0.68	0.60	0.64	0.60
WoS-KW	2772					
Without TT-SS	2772	0.60	0.68	0.56	0.64	0.48
After TT-SS BERT	1698	0.64	0.68	0.60	0.64	0.52
After TT-SS fastText+BERT	2084	0.60	0.68	0.60	0.64	0.50
ELS-MINI-TA	54					
Without TT-SS	54	0.52	0.56	0.52	0.56	0.52
After TT-SS BERT	35	0.52	0.56	0.52	0.60	0.56
After TT-SS fastText+BERT	41	0.56	0.52	0.52	0.52	0.56
ELS-MINI-TA-B	54					
Without TT-SS	54	0.36	0.40	0.44	0.44	0.36
After TT-SS BERT	35	0.40	0.36	0.36	0.36	0.40
After TT-SS fastText+BERT	41	0.40	0.40	0.40	0.36	0.36
ELS-BIG-TA	379					
Without TT-SS	379	0.48	0.60	0.60	0.72	0.44
After TT-SS BERT	242	0.56	0.64	0.60	0.68	0.56
After TT-SS fastText+BERT	281	0.52	0.60	0.64	0.72	0.48
ELS-BIG-TA-B	379					
Without TT-SS	379	0.36	0.52	0.52	0.72	0.28
After TT-SS BERT	242	0.44	0.48	0.56	0.68	0.32
After TT-SS fastText+BERT	281	0.44	0.48	0.56	0.72	0.32
PUBMED	161					
Without TT-SS	161	0.40	0.40	0.44	0.40	0.24
After TT-SS BERT	85	0.40	0.40	0.40	0.40	0.32
After TT-SS fastText+BERT	122	0.44	0.44	0.48	0.48	0.28

Table 4.6: Precision at 25 after TT-SS BERT and TT-SS fastText+BERT with Wasserstein distance on each corpus for each keyword extraction method. Most methods reach a higher precision at 25, on average, either with TT-SS BERT or with TT-SS fastText+BERT. The two methods that benefit the most from TT-SS are Frequency and fastText. TT-SS BERT results in a higher precision at 25 with the Frequency method in 5 out of 8 corpora, the same is observed with TT-SS fastText BERT. For the fastText method, the precision at 25 increases in 15 out of all the 16 cases of using TT-SS. When the precision at 25 was already relatively high, as with TF-IDF and BERT methods, we see fewer instances of increase with TT-SS. We see 2 out of 16 increases for TF-IDF and 4 for BERT.

Precision at 25 for the corpus WoS-TA									
	Number of Publications	%	Freq.	TF-IDF	YAKE	BERT	fastText	TF-IDF +BERT	Random TF-IDF +BERT
Without TT-SS	6897	100%	0.60	0.68	0.60	0.64	0.56	0.64	0.64
After TT-SS BERT Wasserstein	4298	62%	**0.64**	0.68	**0.64**	**0.72**	0.60	0.64	0.68
After TT-SS fastText +BERT Wasserstein	5064	73%	0.60	0.68	0.60	0.64	0.60	0.68	0.68
After TT-SS fastText Wasserstein	3075	45%	0.56	0.64	0.60	0.64	0.56	0.68	0.68
After TT-SS BERT cosine	3363	49%	**0.64**	**0.72**	**0.64**	0.68	0.60	0.68	0.68
After TT-SS fastText +BERT cosine	2497	36%	0.60	0.68	**0.64**	0.68	**0.64**	**0.72**	**0.72**
After TT-SS fastText cosine	3554	52%	0.60	**0.72**	0.60	0.68	0.60	0.68	**0.72**
After TT-SS Levenshtein	3605	52%	**0.64**	0.68	**0.64**	0.68	0.60	0.64	0.64
Precision at 50 for the corpus WoS-TA									
	Number of Publications	%	Freq.	TF-IDF	YAKE	BERT	fastText	TF-IDF +BERT	Random TF-IDF +BERT
Without TT-SS	6897	100%	**0.56**	**0.60**	0.60	**0.62**	0.56	**0.64**	**0.64**
After TT-SS BERT Wasserstein	4298	62%	0.54	0.58	**0.62**	**0.62**	**0.58**	0.60	0.62
After TT-SS fastText +BERT Wasserstein	5064	73%	0.54	**0.60**	**0.62**	**0.62**	**0.58**	0.62	0.60
After TT-SS fastText Wasserstein	3075	45%	0.52	**0.60**	**0.62**	**0.62**	0.56	0.62	0.60
After TT-SS BERT cosine	3363	49%	0.52	0.56	0.58	0.60	0.54	0.60	0.60
After TT-SS fastText +BERT cosine	2497	36%	0.50	0.56	0.56	0.58	0.54	0.58	0.58
After TT-SS fastText cosine	3554	52%	0.50	0.58	0.58	0.60	0.52	0.62	0.62
After TT-SS Levenshtein	3605	52%	0.52	**0.60**	0.58	0.60	0.56	0.62	0.60

Table 4.7: Precision at 25 on the WoS-TA corpus before and after TT-SS. The highest values of precision at 25 and 50 are highlighted, for each keyword method. Applying the proposed combinations of TT-SS are applied to WoS-TA shows that on average the precision at 25 increases compared to not using TT-SS at all. At 25, each keyword extraction method has at least one version of TT-SS that resulted in a higher precision. At 50, 4 out of 7 keyword extraction methods saw equal or higher precision. For this corpus, TT-SS has a consistent positive effect on the precision at 25, however the results on the precision at 50 are higher for YAKE and fastText methods.

2) ***Qualitative Evaluation.*** We perform the qualitative evaluation of TT-SS by looking at a sample of the top k words extracted.

Results. Table 4.8 shows the top 25 keywords extracted using the BERT embeddings method [Grootendorst 2020b] before and after using TT-SS with BERT embeddings and Wasserstein distance on the WoS-TA corpus. We also include the reference keywords for the corpus, which are the author keywords. Words correctly found with and without TT-SS, matching the reference keywords in the top 25, are highlighted in green. Words found only with TT-SS are highlighted in yellow. We can see that the two lists differ by two words while they share the remaining 23 correctly found words. With TT-SS two additional words were found in the top 25, "amazon" and "biomass". Without TT-SS, 16 out of 25 words are correctly found while with TT-SS 18 out of 25 correct words are found. With TT-SS more correct keywords were placed at higher positions in the keyword list.

Reference Keywords				
1. forest	6. climate	11. soil	16. management	21. biomass
2. deforestation	7. tropical	12. amazon	17. biodiversity	22. species
3. land	8. conservation	13. environmental	18. forests	23. policy
4. change	9. redd	14. remote	19. degradation	24. model
5. carbon	10. cover	15. sensing	20. analysis	25. fragmentation
BERT Keywords without TT-SS				
1. forest	6. carbon	11. cover	16. study	21. vegetation
2. deforestation	7. change	12. climate	17. degradation	22. biodiversity
3. land	8. soil	13. changes	18. area	23. management
4. forests	9. conservation	14. tropical	19. emissions	24. water
5. species	10. environmental	15. areas	20. agricultural	25. global
BERT Keywords with TT-SS BERT with Wasserstein distance				
1. forest	6. soil	11. species	16. emissions	21. study
2. deforestation	7. change	12. tropical	17. degradation	22. management
3. land	8. conservation	13. cover	18. agricultural	23. biomass
4. carbon	9. environmental	14. changes	19. biodiversity	24. water
5. forests	10. climate	15. areas	20. global	25. amazon

Table 4.8: Top 25 keywords extracted using BERT with and without TT-SS on the WoS-TA corpus. The reference keywords are the top 25 keywords from the author keywords. The top 25 keywords extracted are shown, first without TT-SS then with TT-SS with BERT and Wasserstein distance. An extracted keyword is considered correct for the precision at 25 if it is found in the first 25 reference keywords. With TT-SS the BERT keyword extraction method was able to find two additional correct keywords in the top 25, "biomass" and "amazon", compared to not using TT-SS. All the words correctly found without TT-SS were also found when TT-SS was used.

WoS-TA		ELS-MINI-TA		ELS-BIG-TA	
Reference keywords	BERT Was. keywords	Reference keywords	BERT Was. keywords	Reference keywords	BERT Was. keywords
1. forest	1. forest	1. land	1. deforestation	1. land	1. forest
2. deforestation	2. deforestation	2. forest	2. forest	2. forest	2. land
3. land	3. land	3. deforestation	3. cooking	3. change	3. environmental
4. change	4. carbon	4. change	4. forests	4. climate	4. carbon
5. carbon	5. forests	5. carbon	5. biomass	5. conservation	5. climate
6. climate	6. soil	6. climate	6. carbon	6. carbon	6. global
7. tropical	7. change	7. amazon	7. environmental	7. policy	7. change
8. conservation	8. conservation	8. energy	8 brazil	8. ecosystem	8. conservation
9. redd	9. environmental	9. ecosystem	9. vegetation	9. environmental	9. agricultural
10. cover	10. climate	10. services	10. conservation	10. water	10. biodiversity
11. soil	11. species	11. solar	11. stove	11. energy	11. energy
12. amazon	12. tropical	12. madagascar	12. soil	12. assessment	12. water
13. environmental	13. cover	13. environmental	13. ecological	13. services	13. development
14. remote	14. changes	14. cover	14. solar	14. management	14. management
15. sensing	15. areas	15. cooking	15. mangrove	15. biodiversity	15. ecosystem
16. management	16. emissions	16. forests	16. agricultural	16. biomass	16. policy
17. biodiversity	17. degradation	17. data	17. sustainability	17. agriculture	17. soil
18. forests	18. agricultural	18. africa	18. biogas	18. food	18. forests
19. degradation	19. biodiversity	19. policy	19. cattle	19. sustainability	19. food
20. analysis	20. global	20. development	20. beef	20. cover	20. sustainable
21. biomass	21. study	21. conservation	21. biodiversity	21. africa	21. study
22. species	22. management	22. intensification	22. production	22. sustainable	22. production
23. policy	23. biomass	23. soil	23. management	23. soil	23. cover
24. model	24. water	24. sustainable	24. agriculture	24. deforestation	24. species
25. fragmentation	25. amazon	25. biodiversity	25. global	25. governance	25. data

Table 4.9: Comparison between top 25 reference keywords from three corpora after TT-SS. Comparison between top 25 reference keywords from the three corpora, WoS-TA, ELS-MINI-TA and ELS-BIG-TA, and the top 25 keywords extracted after using TT-SS with BERT embedings and Wassertein distance. The reference keywords are the author keywords. The keywords extracted by the BERT [Grootendorst 2020b] keyword extraction method are shown. An extracted keyword is considered correct if it is found in the top 25 reference keywords.

Corpora comparison based on their keywords. We show in table 4.9 the top 25 keywords for the following three corpora, WoS-TA, ELS-MINI-TA and ELS-BIG-TA. The number of publications originally in each corpus is 6897 for WoS-TA, 54 for ELS-MINI-TA and 379 for ELS-BIG-TA. Results are shown after applying TT-SS with BERT embeddings and Wasserstein distance, for the BERT keyword extraction method [Grootendorst 2020b]. The precision at 25 for each corpus is 0.72 for WoS-TA, 0.60 for ELS-MINI-TA and 0.68 for ELS-BIG-TA, for the BERT-base keyword extraction method shown. The refernce keywords are the author keywords. An extracted keyword is considered correct with respect to the precision at 25 if it is found in the top 25 reference keywords. In all three cases the majority of the top 25 keywords were correctly found by the BERT extraction method. We can see from the keyword lists that all three corpora are related to forests as the word "forest" is either the top 1 or top 2 word in both the reference keywords and the extracted keywords. We also see that the word "deforestation" is in the top 3 keywords for WoS-TA and ELS-MINI-TA but was not found as one of the top 25 keywords by the keyword extractor for ELS-BIG-TA. We can see that while all three corpora are about forests, climate and conservation, ELS-BIG-TA is not as specifically about deforestation, as the other two corpora. Some geographic regions appear in the top 25 keywords. In the case of WoS-TA it is the Amazon, both in the reference keyword list and in the extracted keyword list. For ELS-MINI-TA, Brazil appears in the extracted keyword list while the Amazon, Madagascar and Africa appear in the reference keyword list. Africa also appears in the reference keyword list for ELS-BIG-TA but not in the extracted keyword list. For someone interested in information about deforestation in the Amazon, WoS-TA and ELS-MINI-TA are two good candidate corpora based on the top 25 keywords. For someone mostly interested in deforestation in Africa, ELS-MINI-TA or ELS-BIG-TA might be more interesting (ELS-MINI-TA is a subset of ELS-BIG-TA).

4.3.5 Experiment III: Random Selection

To support the fact that our method works better than random chance we conduct random selection experiments where publications are selected randomly to form the sub-corpus (later used for keyword extraction). The first drawback of this random selection is knowing the number of publications to select. A second drawback of the random selection is the fact that results are likely to change each time even if the same number of publication is selected. We perform the random

selection experiments on the WoS-TA corpus. We perform each random selection experiment 10 times and report the average precision at 25 of the 10 experiments for each keyword extraction method.

We perform two variations of this experiment, one with the same number of publications as selected by TT-SS fastText + BERT embeddings and Wasserstein distance, and the second with the number of publications selected by TT-SS fastText + BERT embeddings and cosine distance. These two variations were selected because they resulted in the highest and lowest number of selected publications, 5064 and 2497 respectively out of a total of 6897. A higher number of selected publications, closer to the total number of publications in the corpus may not make a big difference in the precision at 25 for a corpus of several hundred publications. However, a low number of selected publications (under 50%) is more likely to impact the precision at 25.

The obtained results are shown in Table 4.10. We report both TT-SS fastText + BERT embeddings and Wasserstein distance and TT-SS fastText + BERT embeddings and cosine results along with the results of random selections made with the same number of publications as these two versions of TT-SS.

With the two TT-SS versions reported, we see that the precision at 25 either increased or stay the same, it never decreased, for all keyword extraction methods. With the random selection the first experiment, which selects 5064 publications (73% of the corpus) the results are somewhat similar to TT-SS. This should be expected, because the higher the percentage of publication selected the more likely the results are to resemble what is obtained on the original corpus, this seems to hold true with or without TT-SS. There is one case where TT-SS did better (fastText) and one case where the random selection did better (YAKE). For the other methods, the results remain the same.

We can see that, for the same number of publications, representing about 62% of the corpus, we get a higher precision at 25 with TT-SS and BERT with Wasserstein than with random selection. With TT-SS and the combined fastText and BERT embeddings with Wasserstein, which selects around 73% of the corpus the results are similar to random selection on average. This suggests that TT-SS might work better with a lower number of publications. If we only remove a very small number of publications from a large corpus, we are not likely to make a big different on the aggregated keywords extracted from that corpus as we are working with a sub-corpus that is not so different from the original full corpus. We expect to see more of a difference when comparing a random selection to TT-SS on a smaller corpus.

	Nb. Pubs.	%	Freq.	TF-IDF	YAKE	BERT	fastText	TF-IDF +BERT	Random TF-IDF +BERT
WoS-TA	6897	100%	0.60	0.68	0.60	0.64	0.56	0.64	0.64
After random selection 1 (average of 10 results)	5064	73%	0.60	0.68	0.61	0.64	0.55	0.64	0.64
After TT-SS fastText +BERT Wasserstein	5064	73%	0.60	0.68	0.60	0.64	0.60	0.68	0.68
After random selection 2 (average of 10 results)	2497	36%	0.60	0.66	0.62	0.65	0.56	0.65	0.65
After TT-SS fastText +BERT cosine	2497	36%	0.60	0.68	0.64	0.68	0.64	0.72	0.72

Table 4.10: Random selection of publication compared to TT-SS on WoS-TA corpus for precision at 25. Selecting publications with TT-SS using BERT embeddings and Wasserstein distance results in higher precision at 25 than performing a random selection on the same number of publications. When 73% of the publications are selected by TT-SS with the combined fastText and BERT embeddings with Wasserstein distance, the precision at 25 is the same on average. For the random selections 1 and 2, the number reported is the average of 10 experiments for each.

4.4 Discussion and Conclusion

We presented our approach for comparing single-term-topic corpora of scientific publications and evaluating them in the prospect of using them in keyword extraction tasks. Our title-topic similarity selection (TT-SS) approach uses the similarity measure between the embedding of the title of the publication and the embedding of the target topic term as a criteria to select the publications that will be used for keyword extraction. We show that this clustering technique can improve the precision at 25 and 50 score of state of the art keyword extraction methods at the corpus level. This approach is potentially of value for tasks where there is a need to reduce the size of the corpus before performing keyword extraction without losing the top keywords. We also examined the effect of similarity measures and text representations, corpus composition, and adding the body of the documents to the extraction process. We found that BERT embeddings generally perform better than fastText embeddings (as seen in the results on the WoS-TA corpus in Table 4.7). One reason for this result might be that the BERT model used is pre-trained on a larger corpus than the fastText model.

We also found that when the corpus is made of publications having the topic keyword in title or abstract, our proposed title topic selection method results in

greater improvements in the precision at 25 for keyword extraction (as see in Table 4.6 when comparing the Web of science corpora WoS-TC versus Wos-TA versus WoS-KW). This is probably due to the fact that TT-SS is very dependent on how relevant the title of the publications are to the topic. By design a corpus that is made by matching titles and/or abstracts to a topic is a good candidate to use with TT-SS. The more relevant a publication is to the topic the more likely it will contain the topic in its title and the more likely it is to be selected by TT-SS.

On the two corpora we tested, we found that adding the body results in lower precision at 25 (as shown for the Elsevier corpora with and without the body added, in Table 4.6, comparing the results for ELS-MINI-TA versus ELS-MINI-TA-B and ELS-BIG-TA versus ELS-BIG-TA-B). The body of the publications seem to add noise to the keyword selection process. This results in lower precision at 25 and 50 without TT-SS compared to not using the body in the keyword extraction process. Applying TT-SS before extracting the keywords is not enough to narrow the gap.

Among the keyword extraction methods we tested, the ones that started out with the lowest precision at 25 were more likely to improve after using TT-SS (as seen in Table 4.6 for the Frequency and fastText keyword extraction methods). This is likely due to the fact that with TT-SS fewer relevant publications are removed reducing the likelihood that non-relevant words would be extracted by any given method, most notably the ones that have low precision before TT-SS. The methods with the highest precision at 25 would either improve or remain at the same level with TT-SS, they would rarely underperform with TT-SS (as shown in Tables 4.6 and 4.7).

We also presented a way to combine TF-IDF and BERT extracted keywords in order to more consistently reach top precision at 25 across corpora, without TT-SS (as shown in Table 4.5). One way in which we could improve on TT-SS is by fine-tuning the selection of the similarity threshold. In our experiments, we mostly use the mean value (and in one case we use the mean value plus the standard deviation) as a threshold for the distance under which we select publications to be included in the sub-corpus used in the keyword extraction task. While this works reasonably well, we could try to optimise the value of this threshold for each corpus, each text representation and each distance measure. We only use a topic represented by a single keyword in our work, however, TT-SS can be applied in the same way with a multi-word topic and evaluated on multi-term keyword extraction.

One of the limits of TT-SS is dealing with non-explicit titles. For instance the

BERT embedding of the title "REDD Mitigation", in the WoS-TC corpus, has a Wasserstein distance of 0.13 to the topic "deforestation", compared to an average distance of 0.04 for all titles. This means that it is deemed not very similar to the word "deforestation". However, REDD is an acronym that stands for "**R**educing **E**missions from **D**eforestation and forest **D**egradation". Therefore, the title "REDD Mitigation" is actually very similar to our topic and even contains the word "deforestation" implicitly. This type of problem could be addressed by pre-training the word embeddings on more specialized corpora. In our approach we only use embeddings pre-trained on general purpose corpora. Another type of publications that might not work well with our approach are those with catchy titles, as their catchiness might be competing with their informativeness [Lopez 2014].

We have found that different combinations of text representations and distance measures used in the title-topic selection process have different levels of performance, in terms of precision at 25 and 50, for different corpora. This means that a given combination of representation and distance may not always be the best for a given corpus. However, we found certain combinations to perform well on average, with the right similarity threshold. We demonstrated the effectiveness of our approach using minimal preprocessing and no new training on the data. One of our goals is indeed to reduce the volume of data that will be used for extracting keywords, without sacrificing precision, by using TT-SS as a light-weight pre-processing step that itself requires little data relative to the volume of the dataset.

In the next chapter we will show how we combine satellite images and text documents in a multi-modal learning task. This task involves matching keywords from a corpus to images, requiring that we use a corpus that contains keywords relevant to the images. Extracting top keywords from our corpus (as a preliminary step) allows us to compare them to the class labels of our images as a way to assess how relevant a corpus is to our images.

Chapter 5

Annotation of Satellite Images in the Context of Change Detection

Summary

Abstract.

Earth observation satellites have been capturing a variety of data about our planet for several decades, making many environmental applications possible such as change detection. Recently, deep learning methods have been proposed for urban change detection. However, there has been limited work done on the application of such methods to the annotation of unlabeled images in the case of change detection in forests. This annotation task consists of predicting semantic labels for a given image of a forested area where change has been detected. Currently proposed methods typically do not provide other semantic information beyond the change that is detected. To address these limitations we first show that deep learning methods can be effectively used to detect changes in a forested area with a pair of pre- and post-change satellite images. We show that by using visual semantic embeddings we can automatically annotate the change images with labels extracted from scientific documents related to the study area.

5.1 Introduction

An increasing number and variety of Earth observation (EO) satellites are orbiting our planet providing a wealth of data for those who need to perform environmental monitoring at various scales [Turner 2015]. This wide coverage is particularly useful for the monitoring of large or very remote areas where on-site data acquisition is impractical. Indeed, the impact of environmental events such as deforestation [Shimabukuro 2000, Vargas 2019], wildfires [Van Leeuwen 2010], and other natural disasters [Bouyerbou 2014, Du 2013] can be assessed with data from EO satellites. With change detection techniques, the various changes that are happening on the Earth's surface can be automatically detected by analyzing images of a given area taken at different times [Singh 1989]. Such techniques have been used to monitor loss and disturbances in forests [Hansen 2016, Vargas 2019], to track change in urban areas [Daudt 2018a], and also to map out areas affected by natural disasters [Bouyerbou 2014, Du 2013]. Figure 5.1 shows an area in the Amazon forest with visible change from 2017 to 2018.

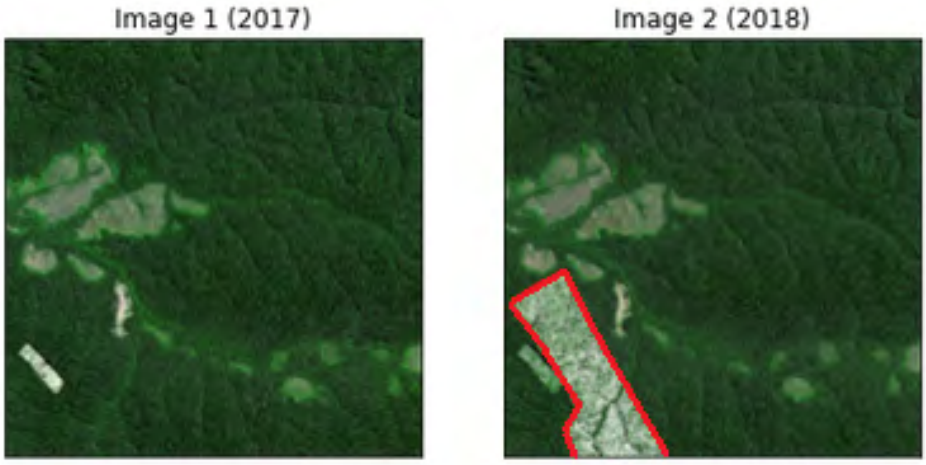

Figure 5.1: An example of forest change in images from the Brazilian Amazon from 2017 to 2018. The change regions are highlighted in the 2018 images. The deforestation that can be seen in 2017 does not count as change, only the new deforestation that appears in the 2018 is considered change for the time period between the two images.

When a change detection method neither provides nor needs additional semantic information beyond the change/no-change pixels, it is called a binary change detection method. Deep learning has been used for binary change detection in urban areas and forests [Daudt 2018a, Ortega Adarme 2020, de Bem 2020], and it has been shown to provide improved results compared to traditional methods

(such as image differencing [Miller 1978] or random forest [Müller 2016]), while requiring less post-processing [Ortega Adarme 2020, de Bem 2020].

Annotation (or semantic labeling), is needed to add semantic information to EO data. In fact, EO images without ground truth labels (or annotations) are plentiful. The American Landsat [1] and European Copernicus [2] programs, for instance, provide free access to the images produced by their respective satellite missions with new images made available every day. These open data policies enable an increasing number of applications to be developed, with those data, in particular for ecology and biodiversity conversation [Turner 2015]. Hence, detecting changes that have occurred in an area of interest at a specific time might require the use of satellite images that have not yet been annotated. In the absence of semantic labels, changes can still be detected by comparing the images; however, semantic information about those changes will be missing. Without this information it is not possible to tell anything more about the area of interest beyond the fact that some change was detected at specific locations. Having labels for each image, acquired either through human (expert) annotators or automatic methods, solves this problem. In a context where expert annotators are not available, and few or no annotated images exist, for an area of interest, automatic annotations can fill the gap.

In this chapter we present our approach to learning annotations for satellite image pairs in the context of change detection. We use pairs of images of a forested area that has undergone change, with one image captured before the change took place and one image captured after. By performing change detection on the image pair we are able to detect the changes that occurred in the time period between the dates the images were taken. This process can be done with one or several pairs of image.

In our approach, we first perform image segmentation on the images where we classify each pixel as change or no-change pixels. The output of this task is a change map showing the pixels where change has occurred on the image. We perform the second task of learning annotations for the image pair by using a visual semantic embedding network. The change map with the annotation shows us where the change occurred and the semantic label(s) of the change.

We evaluate our proposed approach method quantitatively by using the recall at 1, 5 and 10 for image to text retrieval (the annotation task) and text to image retrieval. Qualitatively, we visually examine the change maps produced by the

[1] https://www.usgs.gov/land-resources/nli/landsat - Landsat Missions|U.S. Geological Survey
[2] https://www.copernicus.eu/en - Copernicus | European Union's Earth Observation Program

change detection model and the keywords returned by the image to text retrieval task as annotations.

We show that by using an image pair for the annotation task we improve the recall at 1,5, and 10 compared to using only a single image (post-change). We also show qualitatively that the corpus can provide additional annotations not learned during the training process of the visual semantic model. We also show that using a model trained with fastText embeddings, which was trained on our corpus [Akinyemi 2018], reaches higher recall at one for annotation retrieval than a model trained with BERT embeddings from a pre-trained BERT model.

Finally, we show that given the same set of candidate keywords from a corpus, our models outperform the state of the art CLIP [Radford 2021] model in the annotation retrieval task in recall at 1.

5.1.1 Motivation and Objective

Supervised machine learning has been successfully used for semantic change detection [Daudt 2018a]. Such models are trained on images along with their semantic change masks. In the case of deep learning models in particular, the scale of the data needed for training makes it impractical to have experts manually annotate all the images. Crowdsourcing has been used to provide image annotations at very large scale [Russakovsky 2015]. A similar approach is not well suited for EO images because some expertise is required to properly identify and differentiate among classes. As a result, automatic and semi-automatic approaches are commonly used to build large labeled EO data sets [Shimabukuro 2000, Hansen 2016].

Scientific literature published by researchers who work with EO images is undoubtedly a source of expert knowledge in the field. Publications in Earth sciences therefore can be seen as a very large source of expertise that could be leveraged for adding semantic information to EO images. Furthermore, it is available at a large scale. In fact, across all scientific disciplines, the number of publications has grown exponentially in the past decades [Bornmann 2015]. Using the text from those publications, we can train a neural network to learn word vector representations or embeddings. Word embeddings [Mikolov 2013, Bojanowski 2017, Devlin 2018] are vectors of real numbers, which can be learned by a neural network in an unsupervised way, from a text corpus, without any annotation. For example, Word2vec [Mikolov 2013] learns word vectors with the skip-gram model. With Word2vec, the words with similar meaning will have a similar vector representation.

Visual semantic embeddings allow us to learn to represent textual and visual data in the same vector space. Data points that are semantically closer have a smaller distance between them in this joint space. Different approaches have been proposed to learn visual semantic embeddings for tasks such as image classification [Frome 2013, Radford 2021] and image description [Socher 2014], using the joint embedding of images and words [Frome 2013, Radford 2021] or the joint embedding of images and sentences [Socher 2014] into a common space.

Our goal is to provide a method to learn annotations for satellite image pairs that can be used for change detection. We want to do so by using keywords from a related corpus as candidate annotations, by learning joint image and text embeddings. Therefore, we perform the two core tasks of change detection and annotation of the image pairs. We propose using scientific publications as a source for the annotations, and learning the vector representations of these annotations with a neural language model. Such annotations can be used later in tasks like image indexing and retrieval.

We test our method on images from Sentinel and Landsat missions, and text from the Web of Science, including our largest deforestation corpus presented in Chapter 3. The data used are described in detail in section 5.3.1.

5.1.2 Overview of our Method

Figure 5.2 shows an overview of our approach. While change detection can be applied to any type of image pairs of the same scene or location, we are focusing on the case of changes occurring in forest areas to test our approach. Given a pair of images, we use a change detection model to detect pixels that have changed from one image to the other. The change detection model outputs a change map. With the same image pair we use a visual semantic embedding model to learn the representation of the image pair in the same space as the embbedings of its label. We then use this representation to retrieve the corresponding annotation for the image pair from a corpus, by finding the words embeddings most similar to the image pair representation.

5.1.2.1 Change Detection in Image Pairs

Two satellite images of the same area taken at different times can be compared to find if (and where) changes happened during the time period between the first and the second image. We aim to detect changes in pairs of satellite images of forests, for specific time periods. By doing so we can not only confirm that

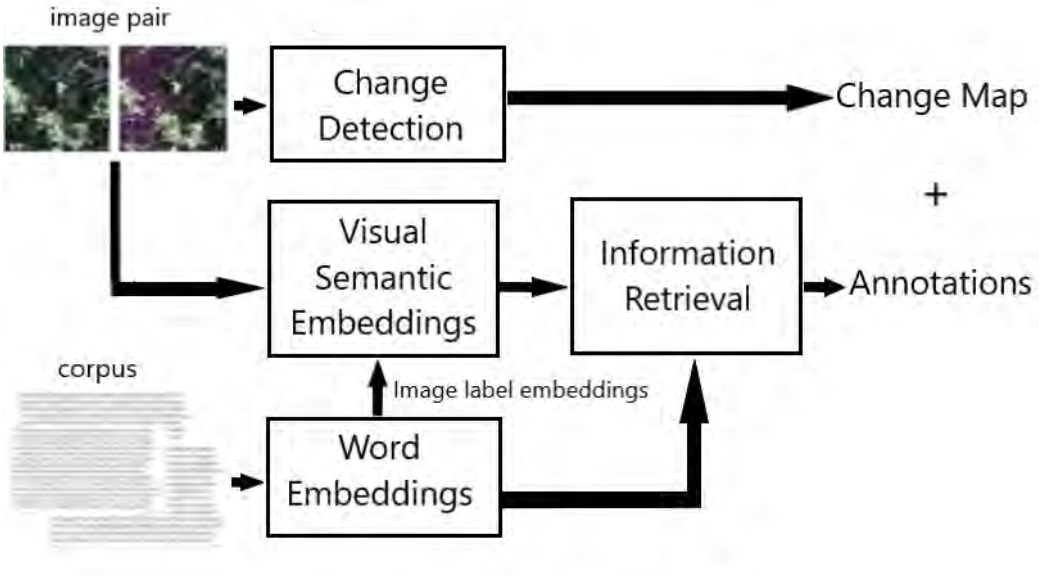

Figure 5.2: Overview of our method for annotating an image pair with words extracted from a corpus of scientific publications. Change detection is performed, with an encoder–decoder model, to predict a change map for an image pair. Word Embeddings are used, to learn the vector representations of all the words in a corpus. Visual Semantic Embeddings are used to learn the feature vectors of the images in the same vector space as the word embeddings. We obtain the image annotations by performing Information Retrieval; given the vector representation of the image pair as the query, we retrieve its annotations from the corpus word vectors.

change happened in those areas, for these periods, but also visibly show where the changes happened. For change detection we use state of the art deep neural network models, which have been shown to perform well on change detection for satellite images. We perform change detection first to find the model that is best at predicting change. We propose to use part of this same model for learning and predicting annotations.

5.1.2.2 Annotation of Satellite Image Pairs

Image annotations provide information about what is present in the image. The same principle applies to an image pair where we assign an annotation jointly to the two images in the pair. We aim to perform the task of automatically annotating image pairs by training a regression model to correctly learn representations for the image pairs that are similar to the representations of their respective annotations. We propose to use the encoder of the same network used in the change detection task (introduced previously) as the model that learns representations of our image pairs. In doing so, we hope to learn image representations that emphasize the change between the two images, because we want the annotations to be related to the changes when they are present. We propose to use those representations in a text retrieval task (where the query is an image pair and the result a word), to find the annotations by searching through candidate words from a given corpus. This allows the possibility of annotations to be any relevant word from the corpus whether the regression model had seen it during training or not.

5.1.3 Summary of Contributions

This chapter includes the following contributions:

(1) We propose a text retrieval approach to annotate image pairs, which allows the matching of words with a pair of images, resulting in improved performance compared to annotating single images.

(2) We examine the influence of using different corpora as the source of candidate annotations.

(3) We propose novel multimodal datasets with satellite image pairs and scientific text to evaluate visual semantic embedding models, along with baseline performance on our models.

5.2 Annotating Satellite Images in a Change Detection Context

In this section we present the problem of change detection and image pair annotation and our proposed solution using common text and image representations also called visual semantic embeddings.

5.2.1 Problem Statement

Our goal is to be able to detect changes that can be seen on satellite image pairs and also assign annotations to the image pair that can correctly label the images but also potentially give us additional relevant information. We achieve this by performing several tasks, namely change detection, visual semantic embedding learning and information retrieval.

Given a pair of satellite images of the same area captured at different times, we want to be able to find if and where change has occurred. We do this by using a supervised change detection learning model. Given the same image pair we want to represent them in a way that makes them similar to their semantic label. We use a supervised visual semantic embedding learning model to learn the representation of the images in the same vector space as the embedding of their label(s). Finally, given the image pair we want to find annotations that are most relevant to it. We do this by performing information retrieval with the representation of the image pair as the query we search among the representations of the words from a relevant corpus for the ones that are most similar to our images. The text used for training the word embedding model and for the retrieval task play an important role in our proposed approach. We need abundant text for the word embedding model and very relevant text for information retrieval. We use scientific publications related of our change type and area of interest. If our chosen corpus is insufficient to satisfactorily train a word embedding model, we take a larger corpus and align it to our smaller corpus in order to improve results. Our proposed approach produces a change map and its related annotations given an image pair as its input.

5.2.2 Change Detection Method for Image Pairs

Our approach for annotating the changes in satellite images uses a change detection method for image pairs. We are therefore using a bi-temporal approach

to change detection where we only consider two images of the same area taken at two different times, once before the change event occurred and once after the change event occurred. We are only using images with a single broadly defined type of change such as wildfire or deforestation. Therefore we are performing binary change detection in which the change map indicates only whether a pixel has changed or not.

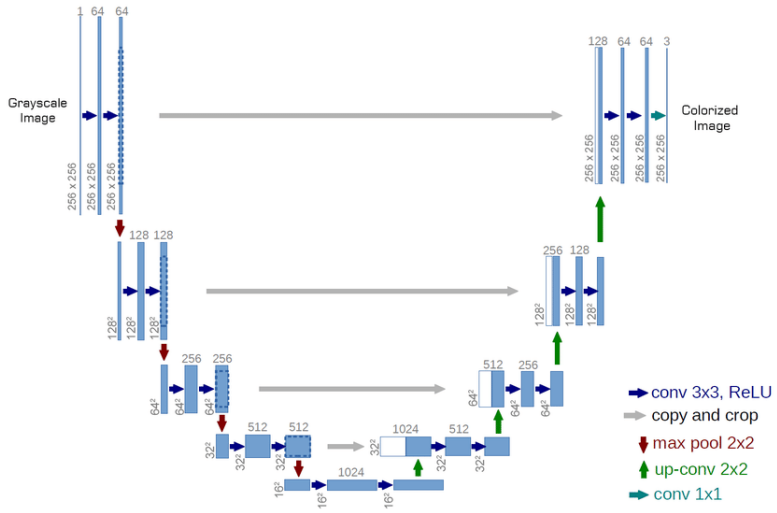

Figure 5.3: The U-Net architecture (from [Ronneberger 2015]). The blue boxes are multi-channel feature maps. The white boxes are copied feature maps. The gray arrows are the skip connections. On top of each box is its number of channels. The vertical numbers are the input size. The left side of the network is the encoder and the right side is the decoder.

For change detection, several deep learning models based on U-Net [Ronneberger 2015] have been proposed [Daudt 2018b, Daudt 2018a, Peng 2019]. The U-Net architecture is a fully convolutional neural network combining an encoder and a decoder connected with skip connections. A skip connection (or shortcut connection) connects two layers of the networks while skipping layers in between. The role of the encoder is to extract features at different spatial resolutions using convolutional filters, generating a downsampled feature map of the original input. The skip connections propagate information from the encoder to the decoder to define the output. Figure 5.3 shows an illustration of the U-Net architecture.

Different encoders can be used with the U-Net architecture. Two such encoders are the Very Deep Convolutional network from the Oxford Visual Geometry Group or VGG [Simonyan 2014], and the Residual Network or ResNet [He 2016]. VGG [Simonyan 2014] is a deep convolutional network that supports up to 19

layers. In fact, the deeper the convolutional network, the more difficult it becomes to train. ResNet [He 2016] was proposed as a way to increase the depth of deep networks while improving accuracy and performance. The first proposed version of ResNet contained 34 layers. ResNets are made of Residual Blocks in which skip connections are used, typically skipping two or three layers at a time. These residual blocks are stacked on top of each other to form the residual network. Adding the skip connections helps avoid the degradation of performance as the network gets deeper. Figure 5.4 illustrates the architectures of VGG19 and ResNet34 along with a plain 34 layer network (without skip connections).

In this work, we use a model similar to the U-Net-based early fusion model that was initially proposed by [Daudt 2018a] to perform the binary change detection task. This network uses a ResNet encoder and takes a concatenation of two images as its input as in [Daudt 2019]. Adding residual blocks to the network has been shown to improve its performance on the image segmentation task for change detection [Daudt 2019]. This model the most simple and generic architecture that outperformed other tested models such as siamese networks, on the binary change detection task [Daudt 2018a, Daudt 2019]. An overview of our change detection approach is shown in Figure 5.5. In addition, we use attention blocks in the decoder as proposed by [Roy 2018]. The attention mechanism was introduced by [Bahdanau 2014], it provides a connection between the encoder and decoder to share information from every encoder hidden state. These attention blocks have been shown to improve the performance of fully convolutional network architectures such as U-Net for image segmentation [Roy 2018]. Figure 5.6 shows an illustration of our encoder decoder architecture with residual and attention blocks.

5.2.3 Visual Semantic Embeddings

To learn annotations for our images we want to use the representation of the images to be able to compare them to the representations of annotations. Visual semantic embeddings create joint representations for image and text in a shared embedding space.

Our approach is built on visual semantic embeddings for annotating changes to add semantic labels to binary change detection. We use deep learning models to predict the binary change map and the vector representation of the images in the word vector space. The encoder of the U-Net [Ronneberger 2015] architecture is then used, with an approach similar to [Frome 2013]. A regression head is

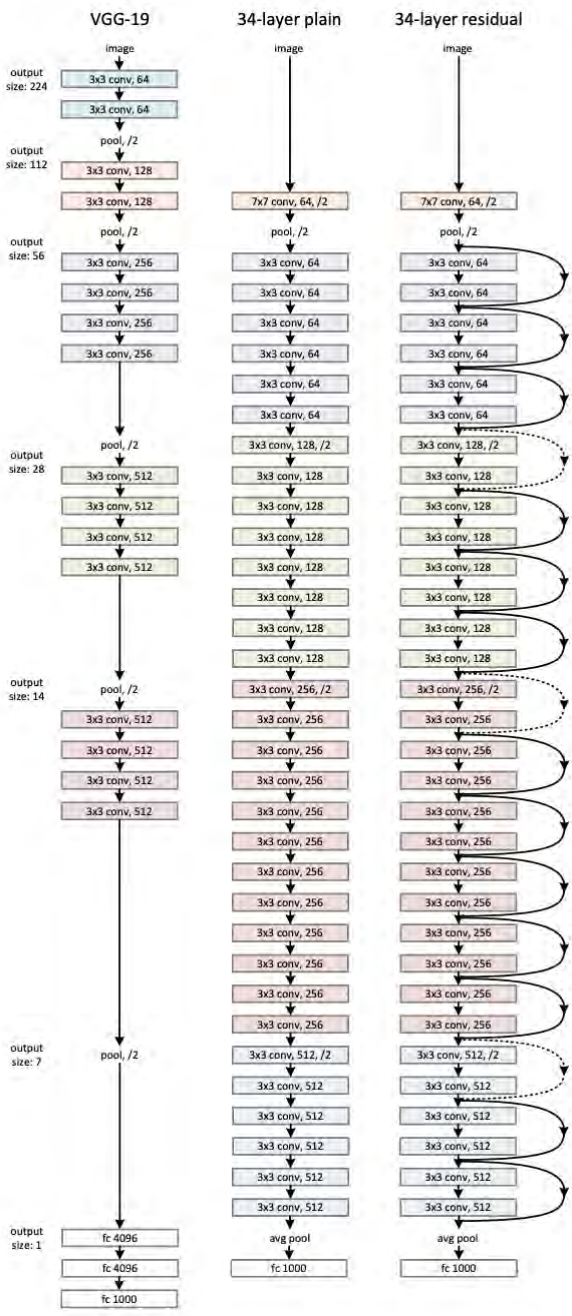

Figure 5.4: By adding skip connections to a plain 34 layer network, ResNet34 avoids training issues that occur in very deep networks. The architectures of VGG19 and ResNet34 are illustrated along with a plain 34 layer network for comparison. Skip connections are shown in ResNet34, the dotted connection increase dimensions. (Illustration from [He 2016]).

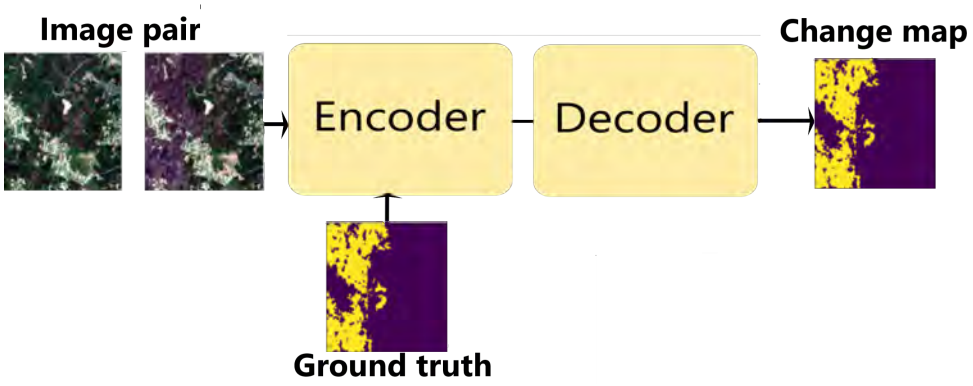

Figure 5.5: Change detection on a satellite image pair. An encoder-decoder model is used to predict a change map for an image pair.

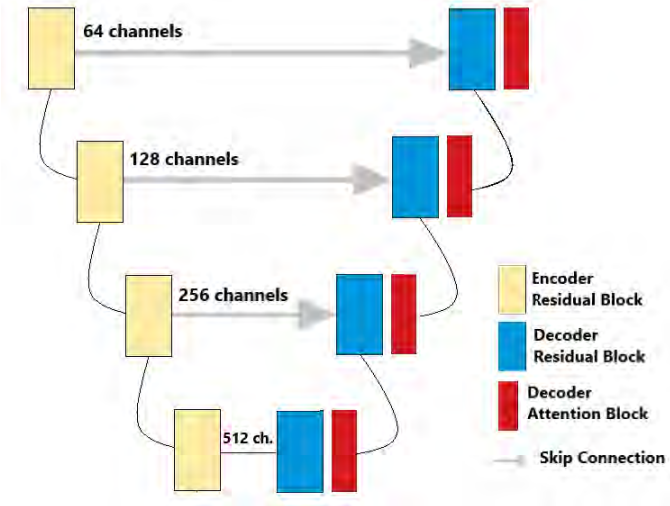

Figure 5.6: An illustration of our encoder decoder network with residual blocks, and attention.

added on top of the encoder to learn the feature vector of the image pair in the same dimension as the vector of its label (Figure 5.7). In [Frome 2013], the vector representations of images are projected into vectors of the same dimensions as the word vectors, and the model predicts the label vector using a similarity metric. In our case, we use a text corpus made of publications related to the area and the type of change of interest to train a word embedding model. We use the word embeddings of the image labels to train the regression head.

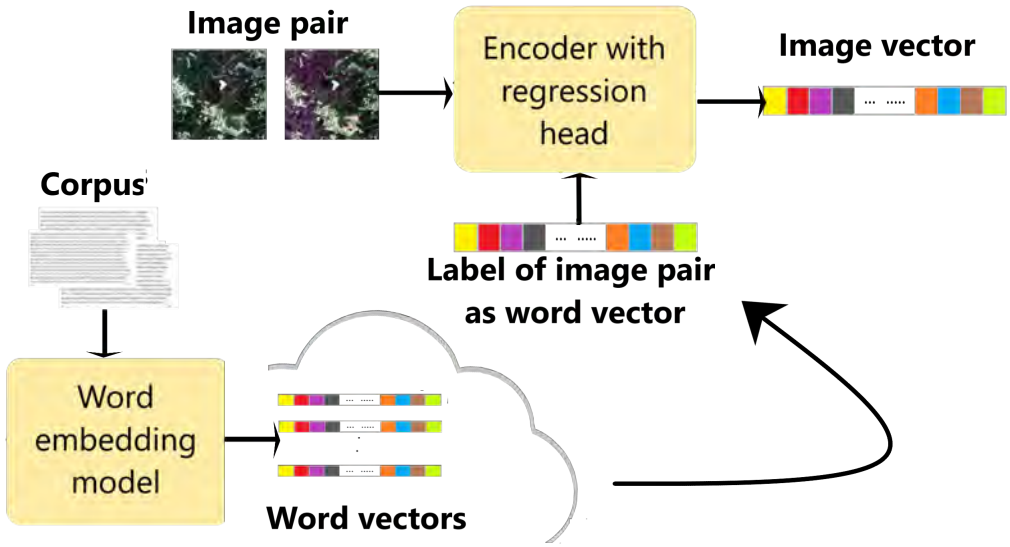

Figure 5.7: Visual semantic embedding. A word embedding model is used to learn the vector representations of all the words in a corpus; a regression head is added to the encoder used in the change detection task (Figure: 5.5), to predict a vector for the image pair, based on the word vector corresponding to its label, which is used as the ground truth.

When making predictions, we search for an annotation among all the word embeddings learned by the word embedding model (Figure 5.8). Predicted annotations can, therefore, be among labels present during training, but they might also be among words that have not been seen during training but are nearest neighbors of the image label in the word vector space.

We use two types of word embeddings, fastText [Bojanowski 2017] and BERT [Reimers 2019]. FastText [Bojanowski 2017] is an extension of the Word2vec model [Mikolov 2013]. With Fasttext (unlike Word2vec) words are broken into n-grams, which are portions of words. For example, the word "forest" will have 5-grams such as "fores" and "orest". Each n-gram will have its own vector, and

the full word will have a vector that is the sum of all its n-gram vectors. BERT [Devlin 2018] is a transformer model that uses a self-attention mechanism. Transformers accept a sequence as input to produce an output. In this case the input is a sequence of words or a sentence. Transformers process all the elements in the sequence together using self-attention. The self-attention mechanism is used to associate each word in a sentence with every other word. The self-attention of a word in a sentence is a function of every word in the sentence (similar to a weighted average). With BERT (unlike fastText), a word can have multiple embedding representations based on its context, this is useful for words that may carry multiple meanings within a same corpus.

We use different corpora to investigate how results might differ with a larger, more general text corpus as opposed to a smaller, more relevant corpus.

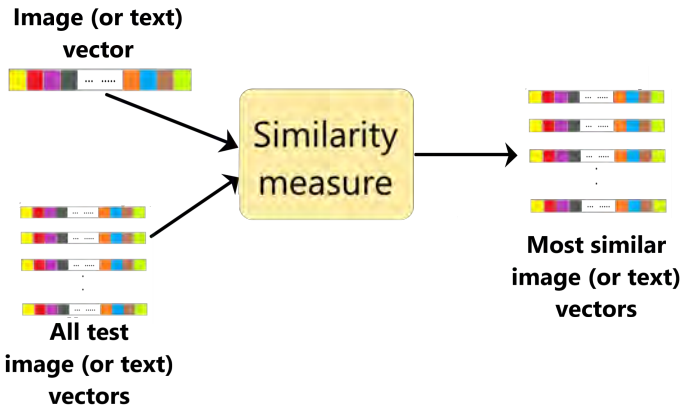

Figure 5.8: Information Retrieval. The proposed method is tested by performing text and image retrieval tasks with the predicted vectors. The vector of an image pair, referred to as an image vector, is compared to all the word vectors learned from the corpus, using a similarity metric; the vector of the label of an image pair, referred to as a text vector, is compared to all the predicted image vectors using the similarity metric. When the input is an image vector, the most similar text vectors are found. When the input is a text vector, the most similar image vectors are found.

Unlike [Daudt 2018a, Peng 2019], our approach is suitable for a change detection dataset that is not fully annotated, meaning that some annotations may be missing or incorrect. We are also not manually building an ontology like [Bouyerbou 2014, Bouyerbou 2019]. Our approach is closer to [Uzkent 2019]; however, they do not perform the change detection task but classify and segment individual EO images.

We want to have a model that is suitable for environmental applications;

therefore, we test on these types of data first. We test on optical satellite imagery, but our proposed approach can be adapted to other types of remotely sensed data, with an adapted network architecture, as needed. Additionally, applications in domains other than environmental sciences are possible, as long as we can find a dataset of image pairs with a corpus of relevant documents.

Evaluation. We assess the task of automatically annotating a pair of EO images used in change detection. Given a pair of images, we want to automatically predict the correct labels for it. Labels are deemed correct if they match the ones assigned by the human annotator. We train the visual semantic model with a single label per image pair and treat this as a single-label multiclass classification problem. We use the cosine similarity to measure the similarity between the predicted annotations and target annotations. We reported the recall at k, which is commonly used for text-image/image-text retrieval tasks [Faghri 2017, Wang 2018], to evaluate the visual semantic embeddings. For the purpose of adding semantic information to the change detected in the images, we also look at predicted annotations that were not an exact match with the target label but were among its closest word vectors.

5.3 Experiments

5.3.1 Evaluation Datasets

5.3.1.1 Portugal Forest Fire Datasets

We use two satellite image datasets and several publication collections. The first image dataset that we use is of the area of Pedrógão Grande in Portugal. The images are from June and July 2017 of the area of Pedrógão Grande in Portugal, which was affected by wildfires in June 2017. The images were captured by the MultiSpectral Instrument of the Sentinel-2[3] satellite. The first image is from 14 June 2017, and the second image is from 4 July 2017. Both images are from the Sentinel-2 tile T29TNE. In this work, we use the three red, green, and blue (RGB) spectral channels, which are the B4, B3, and B2 bands, respectively, for Sentinel-2. The data, provided by the European Space Agency [4] (ESA), were preprocessed and therefore atmospherically corrected and resampled at 10 m. These images are openly available for download from the Copernicus Open Access Hub [5].

[3]https://sentinel.esa.int/web/sentinel/missions/sentinel-2 - Sentinel-2 Mission | Sentinel Online
[4]https://www.esa.int/ - The European Space Agency
[5]https://scihub.copernicus.eu/ - Copernicus Open Access Hub

For the ground truth, the Normalized Burn Ratio (NBR) [García 1991] was used to identify the burnt areas and label the pixels accordingly to obtain a binary segmentation mask. The Normalized Burn Ratio is presented as a reliable means to detect burnt areas by comparing a first pre-fire image to a second post-fire image. The NBR is defined as the ratio between the Near Infrared (*NIR*) and the Shortwave Infrared ($SWIR$) spectral bands, $(NIR - SWIR)/(NIR + SWIR)$. The difference between the two images in terms of NBR decreases with time because of vegetation regrowth. In our case, the image of the burnt areas was taken only about two weeks after the fires, which limits regrowth. Furthermore, the burnt areas are mostly large continuous areas, and a visual evaluation of the performance of the NBR can also be performed.

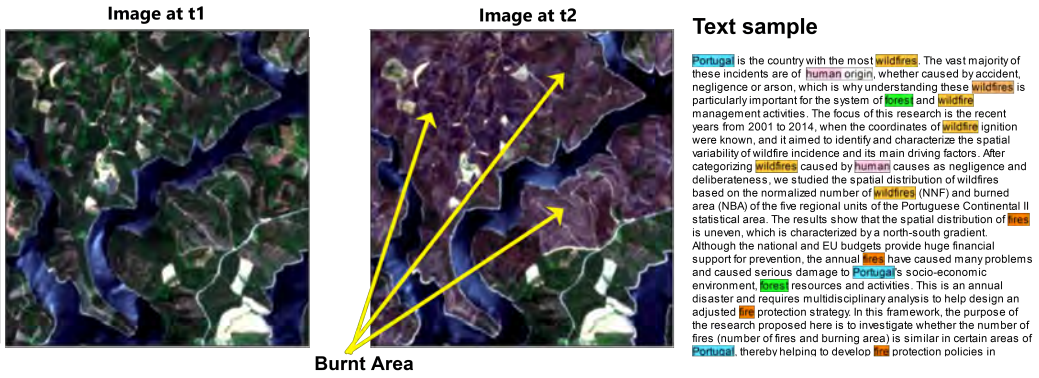

Figure 5.9: Sample images from the dataset with an example text containing relevant keywords for annotating the images. A pair of images is shown with the first image taken at t1 (14 June 2017), and the second one taken at t2 (4 July 2017) showing apparent burnt areas. A portion of text is shown to illustrate the content of the scientific publications being used, and relevant keywords within the text are highlighted.

We collected 16 publications with their title and abstract from the Web of Science [6], which were published from 2017 to 2019 using the topic "Portugal forest fire June 2017". We refer to this corpus as the Portugal Fire (PF) corpus. As there have been other instances of wildfires in the country, we restricted our corpus to the documents more closely related to the June 2017 events. Figure 5.9 shows sample image patches from the dataset from before and after the wildfires along side a text sample with the most relevant keywords highlighted as candidate keywords for the annotations. Because the number of publications that match exactly our dataset is small, whereas word embedding models are better trained on

[6] https://www.webofknowledge.com - Web of Science

large corpora, we also used our deforestation corpus [Akinyemi 2018] introduced in Chapter 3, initially created for investigating deforestation in scientific literature, containing 16136 publications from the years 1975 to 2016. We refer to this corpus as the "Forest" corpus. While this larger corpus does not exactly match our event of interest, it is nevertheless thematically related to it and is appropriate to train a word embedding model.

5.3.1.2 Amazon Deforestation Datasets

The second dataset that we use contains images of a site in the Brazilian Amazon. The images are from the Landsat 8 satellite mission and were captured by its Operational Land Imager sensor. We use the scene 230_65 with images captured on June 21 2017, June 24 2018 and July 13 2019. The images were downloaded from the United States Geological Survey's EarthExplorer [7]. The ground truth masks were created by [de Bem 2020] using data from the Brazilian Institute of Space Research's Project for Deforestation Mapping [Shimabukuro 2000]. In our experiment we only use the Red, Green and Blue bands (bands 4, 3 and 2). In the ground truth, all changes from forest to another land cover type are marked as positive for change (deforestation). Figure 5.10 shows a patch from our second image dataset with the change mask. To create a corpus more related to our second image dataset than the corpora we already have, we collected an additional 446 publications from the Web of Science using the topic keywords "Amazon Brazil deforestation" and restricting our search to the years 2017 to 2020. We refer to this corpus as the "Amazon" corpus. This allows us to include publications about deforestation in Brazil that are contemporary to the images in the dataset and that do not overlap with our largest deforestation corpus [Akinyemi 2018] (the "Forest" corpus), which does not contain publications beyond 2016.

We perform three sets of experiments with the data. The first set on change detection where we are learning to detect change on image pairs and generating the change map. The goal is to find the network that provides the best results and use its encoder in the following experiments. The second set of experiments are on learning the visual semantic embeddings and finding the annotations for the image pair using the encoder from the previous experiment as a visual feature extractor. The third and final set of experiments is on learning annotations using only the post-change image to compare with the results of using image pairs and show the benefit of the later approach.

[7] https://earthexplorer.usgs.gov/ - EarthExplorer

Figure 5.10: Images from the Amazon forest showing changes from 2017 to 2018. The change map shows the areas that have undergone change between the two years. The 2017 images already shows deforested regions. Additional deforestation can be seen in 2018 accounting for the change that can be seen in the change map.

5.3.2 Experiment I : Change Detection for Image Pairs

The change detection task was treated as a binary image segmentation task where two images are segmented as a pair. We trained a U-net [Ronneberger 2015] with pairs of images of the same area taken at different times, as input, and the segmentation map (positive or negative label) of the pixels as ground truth. The two RGB images were concatenated and passed to the network as a single six-channel input. The output was the predicted segmentation map. Pixels that were positive in the segmentation map are the pixels where change occurred. The model was thus trained to learn to differentiate between positive (change) and negative (no-change) pixels in the image pair.

The change detection model used is a fully convolutional neural network (U-Net) [Ronneberger 2015] with a residual network (ResNet34) [He 2016] encoder and decoder attention [Roy 2018].

We used the segmentation models implemented by [Yakubovskiy 2020] with Pytorch [Yakubovskiy 2020] version 1.6.0 and Python version 3 to train the model with the following hyperparmeters: dice loss, Adam optimizer with default parameters, 200 epochs, and a 0.001 learning rate.

We trained the network from scratch without any pretraining. We report the precision, recall, F1 score and mean Intersection over Union (mIoU) obtained from training the U-net model with residual network encoders (ResNet) [He 2016] and very deep convolutional networks originally from the Oxford Visual Geometry Group (VGG) [Simonyan 2014]. Table 5.1 shows the results for binary change detection on the images. The values of the F1 scores varied from 0.71 to 0.85. For

Encoder	Precision	Recall	F1	mIoU
Portugal Forest Fire Images				
ResNet18	0.78	0.89	0.83	0.70
ResNet34	0.78	0.90	0.83	0.72
ResNet50	0.77	0.89	0.83	0.71
VGG11	**0.79**	0.91	**0.85**	**0.73**
VGG16	0.75	0.90	0.82	0.70
VGG19	0.58	**0.93**	0.71	0.55
Amazon Deforestation Images				
ResNet18	**0.85**	0.77	0.81	0.68
ResNet34	0.83	0.82	**0.83**	**0.70**
ResNet50	0.73	0.76	0.74	0.59
VGG11	0.77	**0.86**	0.81	0.68
VGG16	0.70	0.84	0.76	0.62
VGG19	0.69	0.83	0.75	0.60

Table 5.1: ResNet and VGG encoders yield comparable performance measures for the binary change detection task. The values of F1 and mIoU differ only by a few points for the Portugal Forest Fire images for all the encoders. The only exception is VGG19, which had poorer performance than the other networks especially in terms of recall. VGG11 has the highest precision on the Portugal Forest Fire images while Resnet18 has the highest precision on the Amazon Deforestation images. Overall the smaller networks have higher precision than the larger networks. VGG19 has the highest recall on the Portugal Forest Fire images. VGG11 has the highest recall on the Amazon Deforestation images. In terms of F1 and mIoU, the best encoder for the Forest Fire image is VGG11, in terms of F1 and mIoU the best encoder for the Amazon Forest Fire is ResNet34. For the Amazon Deforestation images, ResNet18, ResNet34 and VGG11 have comparable results in terms of F1 and mIoU. For these images, ResNet50, VGG16 and VGG19 underperform.

the mIoU, the values were between 0.55 and 0.73. While the overall performance varied with each network, many of them had comparable performance on the Portugal Forest Fire images except for VGG19 which had much lower precision thatn all the other networks. The results are a bit more varied on the Amazon Deforestation images where larger networks, VGG19, VGG16 and ResNet50 had F1 between 0.74 and 0.76, and mIoU between 0.59 and 0.62. While smaller networks, VGG11, ResNet18 and ResNet34 had higher F1 values between 0.81 and 0.83, and mIoU values between 0.68 and 0.70.

On the basis of the results from the binary change detection task, we chose a ResNet34 encoder for the visual semantic embedding learning task. While it did yield slightly lower F1 and mIoU scores than the VGG11 on the Portugal Forest Fire images it had the best performance on the Amazon Deforestation images.

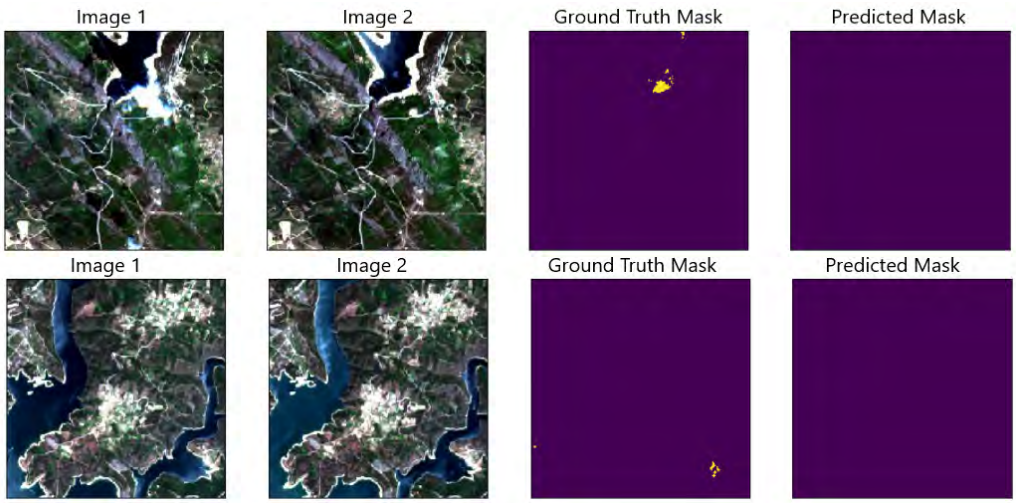

Figure 5.11: The model correctly predicts negative pixels inside water bodies when the ground truth mask has them marked as positive. The positive pixels are the yellow spots. For these image pairs, the model does not make the same mistake as the ground truth mask, and has not marked any pixel as positive as this image does not show burned areas.

In terms of qualitative results, we found that in some cases the model did better than the ground truth mask at predicting negative pixels for the Portugal Forest Fire images. In fact, in some cases, the Normalized Burn Ratio used to create the ground truth mask erroneously marked pixels inside water bodies as burnt vegetation. The deep learning model seemed to be less prone to make the same mistake, as illustrated in Figure 5.11.

5.3.3 Experiment II : Visual Semantic Embedding for Image Pair Annotation

The goal of the visual semantic task is to find a common representation (essentially a vector of real numbers) for images and text in which the images and texts that are related are similar. An image and a text are deemed similar if they relate to the same concept. For example, an image of an ocean would have a similar representation of the word "ocean". With this common representation, we can then find words similar to the images that we wish to annotate and choose the needed annotations from those words.

Given an image pair, we learn the vector representation of that pair in the word vector space. We do this by performing regression with the encoder used for the binary change detection task. We add additional layers to the encoder to predict a single vector for the image features. This image feature vector is of the same dimension as the word vector for the label of the image pair, which we obtain from a word embedding model.

For learning the visual semantic embeddings needed for our annotation task we use a convolutional neural network encoder with a regression head. The regression head is a small neural network added on top of the encoder. This network is made of two fully connected layers with batch normalization, dropout regularization at 25% then 50%, and rectified linear activation function (ReLU). It takes the output of the encoder as its input, then applies adaptive max pooling to reduce the number of dimensions, then flattens the resulting tensor, then passes it through the linear layers, and outputs a vector the same size as the word vectors.

We calculate annotation retrieval metrics to evaluate the performance of our visual semantic model as is common for visual semantic learning models [Frome 2013, Wang 2018]. We therefore report both annotation retrieval (also referred to as text retrieval) and image retrieval metrics.

We report the average recall at k ($R@k$) with k taking values 1, 5, and 10. The recall at k is calculated for text-image retrieval where the query is an annotation, and the result is the corresponding image pairs; for image-text retrieval where the image pair is the query, the result is the corresponding text. For a given text/image (in our case annotation/image pair), the recall at k was set to 1 if the target text/image was present in the top k-nearest neighbors and 0 if not. For the image-text retrieval task, the recall at 1 was equal to the $R - precision$ in our case, because for each image pair we only had a single annotation as its ground truth.

For the text retrieval only, we report the average R-precision (R-Prec), where R represents the number of image pairs with a given annotation, and A_c is the number of correctly predicted annotations; the R-precision is given by A_c/R. For each annotation, the R-precision is therefore the proportion of top R image pairs that were correctly found to match this label, based on the similarity between the predicted vector of the image pair and the vector of the label. R is the total number of image pairs with that label in the dataset.

We perform several experiments on the Portugal Forest Fire data and on the Amazon Deforestation data, which will be described in the following sections.

5.3.3.1 Learning Annotations for the Portugal Forest Fire Images

The images in the Portugal Forest Fire dataset are labeled on the basis of their land cover and land use classes. We define a total of six unique label values: 'agriculture', 'city', 'forest', 'ground', 'wildfire', and 'water'. Each image pair has one or several labels, based on its content. We use one label per image pair to test our method. The label was selected as follows: if wildfire was detected, the image pair was labeled with 'wildfire', if not it was labeled with one of its other labels. The input of the model is the image pair similarly to the change detection task, the difference is that the ground truth is now a word vector, and the training objective is to maximize the similarity between the word vector and the image vector. We use cosine similarity as the vector similarity metric that we are trying to maximize. The word vectors are obtained using fastText [Bojanowski 2017].

We report the evaluation of the change detection task (Table 5.1), the image retrieval and text retrieval (annotation retrieval) tasks, of the image pairs with visual semantic embeddings (Tables 5.2 and 5.3).

We report the values of R-precision (in Tables 5.2 and 5.6) as the averages over all tested keywords for each model. To perform image retrieval we proceed as follows, for each representation (embedding) of a ground truth annotation, we found the representations of the image pair from the test dataset that were most similar to it using the k-nearest neighbors algorithm.

We use two training strategies for learning visual semantic embeddings. The first strategy is to use the encoder to perform each task independently, once for the change detection and once for the visual semantic embedding learning. The second strategy was first to train the model for the binary change detection task, and then use that trained model to train the encoder on the visual semantic embedding learning. In both strategies we performed the change detection task

first. We apply our two training strategies with the different combinations of word embeddings. In addition to word embeddings from our Portugal Fire (PF) corpus (Section 5.3.1), we also test our method with pretrained Wikipedia (Wiki) word embeddings (from FastText [Bojanowski 2017]). Additionally, we test our method with our deforestation corpus (Forest) from [Akinyemi 2018] to find out if having a relatively big corpus, which is more thematically close to our images than Wikipedia, will lead to better predictions. The results can be found in Tables 5.2 and 5.3.

In tables 5.2, 5.3, 5.6, and 5.7, the training strategy indicates whether the model was first pre-trained on the change detection task or not. Aligned word vectors are noted with the '&' symbol, meaning the vectors on the left were aligned to the vectors on the right using [Smith 2017].

For image retrieval evaluation (see Table 5.2), when the query is a word and the result is an image pair, the highest recall at 1, on average, is obtained when the network is not pre-trained. The network trained on the Forest corpus, without pre-training, has the highest recall at 1 for image retrieval. Pre-training the network on the change detection task slightly improves R-precision and recall at 10 for image retrieval, on average. Recall at 1 decreases, on average, when the network is pre-trained on the change detection task. This might be due to the fact that the ground truth annotations do not always capture the differences between the two images, which is essentially what the change detection task does. In fact, for most image pairs, there is no difference to be found. It is likely that by emphasizing the features related to change in the image pair, the pre-training resulted in lower performance for the retrieval task when there are no changes in the image pair.

Pre-training on the change detection task can be beneficial when using aligned corpora for training the visual semantic embeddings. It increases the recall at 1, from 0.25 to 0.50 for Wiki & PF, and recall at 5, from 0.50 to 0.75 for Forest & PF, for image retrieval.

For text retrieval evaluation, when the query is an image pair and the result is a word, considering only recall at 1, the models trained with the corpora that were aligned with the PF corpus obtained the best results in the no-pretraining strategy, the difference for Wiki & PF is the highest of the two, at 0.13. For the pre-training strategy, the model trained on the Forest corpus reached the best recall at 1 of 0.55; however, it performed less well than the top performers under the strategy, without pre-training, for this same task, where the highest recall at 1 was 0.57. The text retrieval results show one limitation of our approach, as we do

		Image Retrieval			
Training Strategy	**Word Vectors Trained on**	**R-Prec**	**R@1**	**R@5**	**R@10**
No Pre-training on Change Detection	PF	0.17	0.25	0.25	0.25
	Forest	0.36	0.75	0.75	0.75
	Wiki	0.39	0.50	0.75	0.75
	Forest & PF	0.40	0.25	0.50	0.75
	Wiki & PF	0.30	0.25	0.75	0.75
	Wiki & Forest	0.34	0.25	0.75	0.75
Pre-training on Change Detection	PF	0.18	0.00	0.50	0.50
	Forest	0.27	0.25	0.50	0.75
	Wiki	0.40	0.50	0.75	0.75
	Forest & PF	0.36	0.25	0.75	0.75
	Wiki & PF	0.37	0.50	0.75	0.75
	Wiki & Forest	0.40	0.50	0.75	0.75

Table 5.2: Choice of corpus and training strategy both influence the performance of the visual semantic model. For image retrieval, when the query is a word and the result is an image pair, recall at 1 is higher, on average, when the network is not pre-trained. Pre-training the network on the change detection task slightly improves R-precision and recall at 10. The models trained with the corpora that are aligned with the PF (Portugal Fire) corpus obtain their best results under the pre-training strategy. The model trained on the Forest corpus reached the highest recall at 1, 5 and 10 under the no-pre-train strategy. Except for the models trained on the PF corpus, all models performed equally well in recall at 10, getting the same score of 0.75 under both the pre-training and no-pre-training strategies.

Training Strategy	Word Vectors Trained on	Text Retrieval R@1	R@5	R@10
No Pre-training on Change Detection	PF	0.00	0.10	0.29
	Forest	0.47	0.82	0.90
	Wiki	0.50	0.68	0.72
	Forest & PF	0.57	0.61	0.61
	Wiki & PF	0.57	0.69	0.69
	Wiki & Forest	0.51	0.56	0.56
Pre-training on Change Detection	PF	0.00	0.04	0.18
	Forest	0.55	0.65	0.68
	Wiki	0.50	0.68	0.72
	Forest & PF	0.52	0.60	0.63
	Wiki & PF	0.44	0.45	0.45
	Wiki & Forest	0.51	0.51	0.51

Table 5.3: Text retrieval results for the visual semantic model trained on image pairs. For text retrieval, when the query is an image pair and the result is a word, recall is higher, on average, when the network is not pre-trained. The highest values for recall at 1 are found with the two corpora aligned with PF, in the no-pre-training strategy. At 5 and 10 the highest recall are found with the Forest corpus under the no-pre-training strategy.

not perform any re-ranking or merging of the predicted annotations, we simply use the nearest neighbors; this results in many variations of the same word in our predicted annotations in many cases (Figure 5.12). One area of improvement would be to post-process our results and filter out words that are only variations of the same word or synonyms.

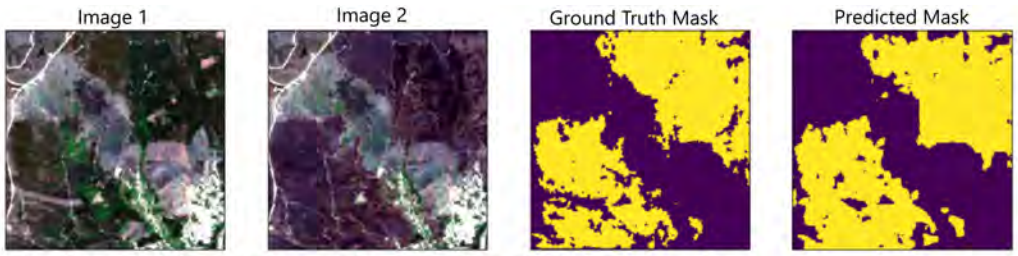

Figure 5.12: An image pair with its ground truth mask and the mask predicted by the change detection model (Section 5.3.2) along with the top five annotations predicted by three models. The results from the models trained on PF without pre-training, PF under the change detection pre-training strategy and Forest aligned with PF without pre-training are shown. The model trained on PF after pre-training on the change detection task is predicting "burn" related words while the non-pre-train model is not. The model trained on the Forest corpus aligned with PF predicts the true annotation as the top 1 annotation.

Models trained with words from the PF corpus, which is the most related to our images, perform less well than the other models in almost all metrics. We can try to qualitatively evaluate samples of the predictions made by these models to see whether they could still be used to add semantic information to the changes detected in the images. As shown in Figure 5.12, while models trained on the PF corpus failed to predict the correct image-pair annotation, they had related words in their top five predictions that add semantic information to the images. Additionally, the effect of the pre-training can be seen in the differences between the top words predicted by each model. The model with change detection pre-training is predicting "burn"-related words, whereas the model that was not pre-trained is not. The third model which was trained on the Forest corpus aligned with the PF corpus (without pre-training) predicts the image annotation correctly. The next four words predicted by this third model are the plural form of the word "wildfire" and words containing either its first part "wild" or its last part "fire".

No much additional information can be learned from those five words that is not already known. These observations lead us to presume that a corpus that is very thematically related to the images is likely to predict annotations that add semantic information, but in order to predict the correct top annotation this corpus should not be too small.

5.3.3.2 Learning Annotations for the Amazon Deforestation Images

We use two labels for the images in the Amazon Deforestation dataset. When an image pair is positive for deforestation it is labeled "deforestation" if not, it is labeled "forest". We are therefore not taking other classes that might be present into account as we do not have any other reference labels for this dataset. However, we find that this allows us to represent the majority classes of our dataset. Four variations of the visual semantic model were used in our experiments, a model trained with fastText embeddings, which were trained on a the Forest corpus (fastText-Forest), a model trained with fastText embeddings, which were trained on the Amazon corpus (fastText-Amazon), a model trained with fastText embeddings pre-trained on Common Crawl and Wikipedia (fastText-CCWiki), and a model with BERT embeddings pre-trained on Web data (BERT-Web). Pre-trained word embedding models were not retrained on our corpora but used as is. Candidate annotations, for all models except fastText-Forest, were taken from the top 25 words for the Amazon corpus extracted using fastText (when fastText embeddings are use) and BERT (when BERT embeddings are used) using methods described in Chapter 3. For fastText-Forest, the candidate annotations were taken from the Forest corpus in the same manner. We used top keywords as candidate annotations to limit the possible annotations (and the possibility of erroneous predictions) while still keeping the most relevant candidates. We report image to text retrieval (annotation retrieval) metrics only for the Amazon Deforestation Images. Table 5.4 shows the results obtained with fastText and BERT embeddings, with visual semantic models trained from scratch. For recall at 1, the highest value of 0.70 is obtained with the fastText-Forest and fastText-CCWiki models. For recall at 5, the BERT-Web model reaches the highest value of 0.99. For recall at 10 the fastText-Amazon model reaches the highest value of 1.

We can evaluate our models qualitatively by looking at a sample of the obtained results. In Figure 5.13 we see the change detection map obtained by the U-Net-ResNet34 change detection model along with the annotations predicted by the variations of the visual semantic model (ResNet34 with regression head). Only

Model Name	Word Embeding Model	Word Embeddings Trained on	R@1	R@5	R@10
fastText-Forest	fastText	Forest	0.70	0.95	0.96
fastText-Amazon	fastText	Amazon	0.68	0.68	1.00
fastText-CCWiki	fastText	CC. + Wiki	0.70	0.94	0.94
BERT-Web	BERT	Web Data	0.65	0.99	0.99

Table 5.4: Recall at 1, 5 and 10 for the visual semantic models predicting annotations for the Amazon Deforestation images. The word embedding models are the models used for the word embeddings. The corpora are listed on the third column showing what data the word embeddings were trained on. The visual semantic model uses the embeddings of the labels of the images as the target in a regression task with the image pair as input. The models were trained from scratch without pre-training on the change detection task. For recall at 1 the models using fastText embeddings outperform the one using BERT. For recall at 5, the BERT-based model outperforms all the others. The fastText model trained on the smallest corpus (fastText-Amazon) performs less well than the ones trained on larger corpora for recall at 1 and at 5, but it outperforms them at 10.

the fastText-Forest and BERT-Web models predicted the true annotation correctly as the first annotation.

Figure 5.13: An image pair with its ground truth mask and the mask predicted by the change detection model (Section 5.3.2) along with the top five annotations predicted by three visual semantic models. fastText-Forest and BERT-Web correctly predicted the true annotation "deforestation" as the top 1 annotation. fastText-CCWiki produced the correct annotation in second position. These three models predicted forest related words. The fastText-Amazon model does not predict the correct annotation in the top 5 but it predicts the relevant annotation "amazon" in fourth position.

We compare our models to CLIP-RSICD [Radford 2021, Lu 2017] by testing on the case where the candidate annotations are only the labels present in the dataset that were are using for the tests. CLIP [Radford 2021] is a visual semantic model that uses a visual transformer to extract image features, and a causal language model to extract text features. These features are then projected, with the same dimensions, into a latent space. In that new space the similarity between the visual and text features is calculated. CLIP-RSCID is a version of CLIP [Radford 2021] trained on the RSICD dataset [Lu 2017] and other high resolution satellite imagery [Yang 2010]. We report the recall at 1, for this case we do not report recall at 5 or 10 because in the case of the Amazon data set there are only two possible annotations. Table 5.5 shows the recall at 1 for all the models tested. CLIP-RSICD reaches a recall at 1 value of 0.49 compared to our models that reach values from 0.65 to 0.75.

Model Name	R@1
fastText-Forest	0.71
fastText-Amazon	0.68
fastText-CCWiki	0.75
BERT-Web	0.65
CLIP-RSICD*	0.49

Table 5.5: When only using the two labels as candidate annotations, the fastText model trained on Common Crawl and Wikipedia improved its performance from 0.70 to 0.75 in recall at 1 compared to when it used the top 25 words of the Amazon corpus as the candidate annotations. The BERT model and the fastText models trained on the Amazon corpus show the same performance as when they were using the top 25 words as candidate annotations. The fastText model trained on the Forest corpus improve its performance slightly from 0.70 to 0.71. The CLIP-RSICD model underperforms our models. (*) The CLIP-RSICD model was trained on single images and tested on the post-change image, contrary to our models, which were trained and tested on image pairs.

5.3.4 Experiment III : Visual Semantic Embedding Learning for Post-Change Image Annotation

We further test our approach using single images instead of image pairs on the Portugal Fire Dataset. The goal is to show the benefit of using pairs for the

annotation task. We use the post-change image in the following experiments and report the image retrieval results with the no-pre-training strategy.

Results of this set of experiments are presented in tables 5.6, for image retrieval. Recall at 1 for Wiki & Forest went from 0.25 to 0.50. Recall at 10 for Wiki & Forest went from 0.75 to 1.00. Recall at 1 for PF went from 0.25 to 0.50. Recall at 5 for Forest & PF went from 0.50 to 0.75. R-precision values have also all increased.

		Image Retrieval			
Training Strategy	**Word Vectors**	**R-Prec**	**R@1**	**R@5**	**R@10**
No Pre-training on Change Detection	PF	0.19	0.25	0.50	0.50
	Forest	0.42	0.75	0.75	0.75
	Wiki	0.41	0.50	0.75	0.75
	Forest & PF	0.42	0.25	0.75	0.75
	Wiki & PF	0.37	0.25	0.75	0.75
	Wiki & Forest	0.40	0.50	0.75	1.00

Table 5.6: Image retrieval results when the visual semantic model is trained on a a single post-change image. Training the visual semantic model only on the second (post-change) image improves image retrieval scores. The model performs better in image retrieval than when trained on the image pair.

For text retrieval, results are shown in Table 5.7. Overall results are lower for all models. Most notably, for PF, recall is 0.00 at 1, 5 and 10 while it was 0.10 and 0.29 at 5 and 10 when trained on the image pair. For Forest recall at 1, 5 and 10 are 0.42, 0.54, 0.54 compared to 0.47, 0.82, 0.90 (when trained on the image pair).

We can make a more direct comparison with CLIP because only the post-change image is used in our experiments. When all the words from the corpus are provided as candidate labels, as is the case with our models, CLIP scores 0 for recall at 1. We tested by providing only the four labels found in the test set as candidate words and CLIP scored 0.41. This is still lower than three out of six models as shown in Table 5.7.

We tested to see if the model can learn the annotations when it is trained only on a single (post-change) image. We found, in that case, the model performs slightly better in the image retrieval tasks, where the goal is to predict an image for a given annotation. However, it underperforms in text retrieval tasks, where the goal is to predict an annotation for an image (Table 5.7).

		Text Retrieval		
Training Strategy	**Word Vectors**	**R@1**	**R@5**	**R@10**
No Pre-training on Change Detection	PF	0.00	0.00	0.00
	Forest	0.42	0.54	0.54
	Wiki	0.47	0.59	0.61
	Forest & PF	0.40	0.41	0.41
	Wiki & PF	0.39	0.49	0.49
	Wiki & Forest	0.49	0.53	0.54

Table 5.7: Text retrieval results when the visual semantic model is trained on a a single post-change image. Training the visual semantic model only on the second (post-change) image yields lower values for recall at k for the text retrieval (annotation prediction task). When the same visual semantic embedding model was trained on a single image taken after the change event, it reached lower scores than when trained on the image pair, for all the corpora, in text retrieval.

5.4 Conclusions

We propose a new method to predict relevant annotations for pairs of satellite images in areas undergoing visible change that can be detected by comparing two images. Our goal is to have a model that can annotate unlabeled image pairs, and then be able to provide additional semantic information on the area of interest as extra annotations. We showed that this can be done using state-of-the-art deep neural networks. Since both the change detection and annotation tasks rely on feature extraction from the images, they can share the same feature extractor, i.e., the same convolutional neural network encoder. We attempted to use word vectors learned from a small corpus relevant to the area and the changes of interest to predict the most relevant keywords. We showed the limitation of this basic approach with image-text retrieval metrics and demonstrated how, using a larger, less relevant corpus that is aligned to the initial small corpus, we can achieve better performances on the retrieval tasks, in particular in the image to text retrieval task, which is in essence what annotation prediction is doing (i.e., given an image predict the top words associated with it).

Overall, we emphasized the role of the corpora on which the word embedding model is trained in the performance of the visual semantic model. To the best of our knowledge, this is the first work to propose a method for learning to predict annotations of change detection image pairs, with word vectors learned from a

scientific publication corpus.

Our proposed method can be applied to a dataset of images presenting several types of changes. The only requirement, for this dataset is that the image pairs used for training are labeled with their respective change types. For the corpus, any collection of documents with text related to the types of changes of interest can be used.

Our approach can be used as is, or easily adapted to other types of input data including from other types of sensors such as radar. Future studies could continue to explore broader applications of our method, for example, to larger and more diverse satellite image datasets, with a variety of changes. Along with these new image datasets, the use of new types of corpora could be explored such as news articles.

Our method is generic although we only used it on deforestation data in our work. It can be used with data from other domains where there are available images and related text.

We wanted to show how our method works on real data; therefore, we used satellite images from real change events and corpora of scientific publications related to both the areas and the types of changes that occurred there. Consequently, the size of our image datasets is small. This might limit the ability for our models to generalize on images of different forests, with different types of changes or from different sensors. One possible solution could be to use models trained of larger and more varied data such as CLIP [Radford 2021]. However we showed that our models perform better than the version of CLIP specifically pre-trained on satellite images on the annotation prediction task.

Chapter 6

Conclusion

In this thesis we present methods for performing change detection on satellite images, and adding annotations to the images, by extracting keywords from an unpaired corpus. We show (in Chapter 5) the impact of the corpus choice and how to improve the performance on the annotation task by aligning relevant corpora.

We take deforestation as an example of change that may occur on the land cover and that can be detected by using satellite images. First, we present an examination of deforestation in scientific literature with detailed bibliometric analyses of a corpus on the topic. We conduct several quantitative analyses of scientific publications on the topic of deforestation to visualize scientific production as well as collaboration networks. We perform keyword and network analyses as well as metadata analyses to find locations of interests and author and country collaboration. For the top countries and regions of interest, we also analyse their top keywords over time to identify the ones more specifically relevant to each country/region.

Then, we present a method to select publications from a corpus prior to extracting keywords. We propose a title-topic similarity selection (TT-SS) approach for creating a sub-corpus that is used to extract the keywords. TT-SS uses the similarity between the title of the publication and a word representing the topic of the corpus to find and thus select the publication with high enough similarity. We show that using BERT embeddings to represent the topic word, the titles and the abstracts, we can extract keywords for a single-topic corpus with high precision by finding the words most similar to the title and abstract. We further improve the precision by combining BERT-based keywords with TF-IDF keywords.

We investigate the close relation between images and the semantic labels of those images in a change detection context. By focusing specifically on deforestation we were able to adopt a single-topic approach and use a topic-specific corpus to extract candidate annotations for our images. In that settings the corpus keywords proved not only informative but also relevant to the images. We conclude that a topic-specific corpus of scientific publications is appropriate and

even beneficial as a source of candidate keywords to annotate images related to the same topic.

Finally, we present a method for detecting change and providing annotations for change images combining image segmentation, word embeddings and visual semantic embeddings. The later learns the similarity between an image pair and the vector representation of its label by projecting the representation of the image pair into the same vector space as its label vector. By doing so, we are able to find additional words (annotations) that relate to the image pair using the cosine similarity between the vector of the images and all the word vectors from a related corpus. We evaluate the performance of our proposed method using image-text retrieval. We show how the quality of these annotations varies with the choice of corpus by comparing results from Wikipedia, and other corpora.

One area of complementary research to the work presented in this thesis would be perform semantic change detection on time series of satellite images. This would require images with semantic change labels for training. Having the time series would allow to follow the evolution of the changes over time. The same could be done with a related corpus covering the same area for the same time periods by tracking the semantic labels (keywords). While finding real data for these tasks might be challenging it could result in finding a broader variety of annotations than the ones we could find when limited to a single type of change, and give insight into the dynamics between the different types of changes observed.

We have attempted to better understand the phenomenon of deforestation with innovative approaches based on deep learning and statistical analyses applied to a mix of text corpora and satellite imagery datasets. We focused our work on an environmental domain but there are other domains also interested in matching text with images. For example, in the medical field, medical imagery produced for patients are accompanied with annotations and reports produced by medical doctors. There is also an extensive scientific medical literature that is available. The approaches presented in our work could be applied in such context especially in cases involving the production of repeated follow up images being produced for the same patient.

List of Figures

List of Tables

Bibliography

[Aleixandre-Benavent 2018] Rafael Aleixandre-Benavent, José Luis Aleixandre-Tudó, Lourdes Castelló-Cogollos et José Luis Aleixandre. *Trends in global research in deforestation. A bibliometric analysis.* Land Use Policy, vol. 72, pages 293–302, 2018.

[Aronson 2000] Alan R Aronson, Olivier Bodenreider, H Florence Chang, Susanne M Humphrey, James G Mork, Stuart J Nelson, Thomas C Rindflesch et W John Wilbur. *The NLM Indexing Initiative.* In Proceedings of the AMIA Symposium, page 17. American Medical Informatics Association, 2000.

[Assembly 2015] UN General Assembly. *Transforming our world : the 2030 Agenda for Sustainable Development.* A/RES/70/1, 21 October, 2015. Available at: https://www.refworld.org/docid/57b6e3e44.html [accessed: 17 March 2022].

[Bahdanau 2014] Dzmitry Bahdanau, Kyunghyun Cho et Yoshua Bengio. *Neural machine translation by jointly learning to align and translate.* arXiv preprint arXiv:1409.0473, 2014.

[Beech 2017] E Beech, M Rivers, S Oldfield et PP Smith. *GlobalTreeSearch: The first complete global database of tree species and country distributions.* Journal of Sustainable Forestry, vol. 36, no. 5, pages 454–489, 2017.

[Bennani-Smires 2018] Kamil Bennani-Smires, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl et Martin Jaggi. *Simple Unsupervised Keyphrase Extraction using Sentence Embeddings.* In Proceedings of the 22nd Conference on Computational Natural Language Learning, pages 221–229, 2018.

[Bird 2009] Steven Bird, Ewan Klein et Edward Loper. Natural language processing with python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.", 2009.

[Blaschko 2008] Matthew B Blaschko et Christoph H Lampert. *Correlational spectral clustering.* In 2008 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, 2008.

[Blei 2003] David M Blei, Andrew Y Ng et Michael I Jordan. *Latent dirichlet allocation.* the Journal of machine Learning research, vol. 3, pages 993–1022, 2003.

[Bojanowski 2017] Piotr Bojanowski, Edouard Grave, Armand Joulin et Tomas Mikolov. *Enriching word vectors with subword information.* Transactions of the Association for Computational Linguistics, vol. 5, pages 135–146, 2017.

[Bojović 2014] Srđan Bojović, Rada Matić, Zorica Popović, Miroslava Smiljanić, Milena Stefanović et Vera Vidaković. *An overview of forestry journals in the period 2006–2010 as basis for ascertaining research trends.* Scientometrics, vol. 98, no. 2, pages 1331–1346, 2014.

[Borgman 1989] C. L. Borgman. *Bibliometrics and scholarly communication.* Communication Research, vol. 16, page 583, 1989.

[Bornmann 2015] Lutz Bornmann et Rüdiger Mutz. *Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references.* Journal of the Association for Information Science and Technology, vol. 66, no. 11, pages 2215–2222, 2015.

[Bourdis 2011] Nicolas Bourdis, Denis Marraud et Hichem Sahbi. *Constrained optical flow for aerial image change detection.* In 2011 IEEE International Geoscience and Remote Sensing Symposium, pages 4176–4179. IEEE, 2011.

[Bouyerbou 2014] Hafidha Bouyerbou, Kamal Bechkoum, Nadjia Benblidia et Richard Lepage. *Ontology-based semantic classification of satellite images: Case of major disasters.* In 2014 IEEE Geoscience and Remote Sensing Symposium, pages 2347–2350. IEEE, 2014.

[Bouyerbou 2019] Hafidha Bouyerbou, Kamal Bechkoum et Richard Lepage. *Geographic ontology for major disasters: methodology and implementation.* International Journal of Disaster Risk Reduction, vol. 34, pages 232–242, 2019.

[Broadus 1987] Robert N Broadus. *Toward a definition of "bibliometrics".* Scientometrics, vol. 12, no. 5, pages 373–379, 1987.

[Campos 2020] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Jorge, Célia Nunes et Adam Jatowt. *YAKE! Keyword extraction from single documents using multiple local features.* Information Sciences, vol. 509, pages 257–289, 2020.

[Collins 1996] John B Collins et Curtis E Woodcock. *An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data.* Remote sensing of environment, vol. 56, no. 1, pages 66–77, 1996.

[Cuxac 2017] Pascal Cuxac et Nicolas Thouvenin. *Archives numériques et fouille de textes: le projet ISTEX.* Atelier TextMine, EGC 2017 (Extraction et Gestion des Connaissances), Grenoble, France, January 24, vol. 27, page 2017, 2017.

[Dal Pont 2020] Thiago Raulino Dal Pont, Isabela Cristina Sabo, Jomi Fred Hübner et Aires José Rover. *Impact of text specificity and size on word embeddings performance: An empirical evaluation in brazilian legal domain.* In Brazilian Conference on Intelligent Systems, pages 521–535. Springer, 2020.

[Daudt 2018a] Rodrigo Caye Daudt, Bertr Le Saux et Alexandre Boulch. *Fully convolutional siamese networks for change detection.* In 2018 25th IEEE International Conference on Image Processing (ICIP), pages 4063–4067. IEEE, 2018.

[Daudt 2018b] Rodrigo Caye Daudt, Bertr Le Saux, Alexandre Boulch et Yann Gousseau. *Urban change detection for multispectral earth observation using convolutional neural networks.* In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pages 2115–2118. IEEE, 2018.

[Daudt 2019] Rodrigo Caye Daudt, Bertrand Le Saux, Alexandre Boulch et Yann Gousseau. *Multitask learning for large-scale semantic change detection.* Computer Vision and Image Understanding, vol. 187, page 102783, 2019.

[de Bem 2020] Pablo Pozzobon de Bem, Osmar Abílio de Carvalho Junior, Renato Fontes Guimarães et Roberto Arnaldo Trancoso Gomes. *Change Detection of Deforestation in the Brazilian Amazon Using Landsat Data and*

Convolutional Neural Networks. Remote Sensing, vol. 12, no. 6, page 901, 2020.

[Deerwester 1990] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer et Richard Harshman. *Indexing by latent semantic analysis.* Journal of the American society for information science, vol. 41, no. 6, pages 391–407, 1990.

[Devlin 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee et Kristina Toutanova. *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018.

[Diegues 1992] Antônio Carlos Sant'Ana Diegues, Paulo Kageyama et Virgilio Viana. The social dynamics of deforestation in the brazilian amazon: an overview, volume 36. United Nations Research Institute for Social Development, 1992.

[Dieng 2020] Adji B Dieng, Francisco JR Ruiz et David M Blei. *Topic modeling in embedding spaces.* Transactions of the Association for Computational Linguistics, vol. 8, pages 439–453, 2020.

[Dobbertin 2010] Michèle Kaennel Dobbertin et Michael Peter Nobis. *Exploring research issues in selected forest journals 1979–2008.* Annals of Forest Science, vol. 67, no. 8, page 800, 2010.

[Dousset 2003] B Dousset. *Intégration de méthodes interactives de découverte de connaissances pour la veille stratégique.* Habilitation à diriger des recherches, Université Toulouse, vol. 3, page 20, 2003.

[Dousset 2009] Bernard Dousset. *TETRALOGIE: Software for monitoring Science and Technology.* International Journal of Competitive Intelligence, Strategic, Scientific and Technology Watch (SCI&WATCH), pages 13–21, 01 2009.

[Du 2013] Lingtong Du, Qingjiu Tian, Tao Yu, Qingyan Meng, Tamas Jancso, Peter Udvardy et Yan Huang. *A comprehensive drought monitoring method integrating MODIS and TRMM data.* International Journal of Applied Earth Observation and Geoinformation, vol. 23, pages 245–253, 2013.

[El Amin 2016] Arabi Mohammed El Amin, Qingjie Liu et Yunhong Wang. *Convolutional neural network features based change detection in satellite images.*

In First International Workshop on Pattern Recognition, volume 10011, page 100110W. International Society for Optics and Photonics, 2016.

[ESA 2017a] ESA. *Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017).*, 2017.

[ESA 2017b] ESA. *S2 prototype Land Cover 20m map of Africa 2016.*, 2017.

[Faghri 2017] Fartash Faghri, David J Fleet, Jamie Ryan Kiros et Sanja Fidler. *Vse++: Improving visual-semantic embeddings with hard negatives.* arXiv preprint arXiv:1707.05612, 2017.

[Foley 2005] Jonathan A. Foley, Ruth DeFries, Gregory P. Asner, Carol Barford, Gordon Bonan, Stephen R. Carpenter, F. Stuart Chapin, Michael T. Coe, Gretchen C. Daily, Holly K. Gibbs, Joseph H. Helkowski, Tracey Holloway, Erica A. Howard, Christopher J. Kucharik, Chad Monfreda, Jonathan A. Patz, I. Colin Prentice, Navin Ramankutty et Peter K Snyder. *Global Consequences of Land Us.* Science, vol. 309, pages 570–574, 2005.

[Frome 2013] Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Marc'Aurelio Ranzato et Tomas Mikolov. *Devise: A deep visual-semantic embedding model.* In Advances in neural information processing systems, pages 2121–2129, 2013.

[García 1991] MJ López García et V Caselles. *Mapping burns and natural reforestation using Thematic Mapper data.* Geocarto International, vol. 6, no. 1, pages 31–37, 1991.

[Garfield 2006] Eugene Garfield. *Citation indexes for science. A new dimension in documentation through association of ideas.* International journal of epidemiology, vol. 35, no. 5, pages 1123–1127, 2006.

[Glänzel 2003] W. Glänzel et Magyar Tudományos Akadémia. Kutatásszervezési Intézet. Bibliometrics as a research field: A course on theory and application of bibliometric indicators. Magyar Tudományos Akadémia, Kutatásszervezési Intézet, 2003.

[Grootendorst 2020a] Maarten Grootendorst. *BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics.*, 2020.

[Grootendorst 2020b] Maarten Grootendorst. *KeyBERT: Minimal keyword extraction with BERT.*, 2020.

[Hadifar 2018] Amir Hadifar et Saeedeh Momtazi. *The impact of corpus domain on word representation: a study on Persian word embeddings.* Language Resources and Evaluation, vol. 52, no. 4, pages 997–1019, 2018.

[Hansen 2013] Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, SV Stehman, Scott J Goetz, Thomas R Loveland*et al. High-resolution global maps of 21st-century forest cover change.* science, vol. 342, no. 6160, pages 850–853, 2013.

[Hansen 2016] Matthew C Hansen, Alexander Krylov, Alexandra Tyukavina, Peter V Potapov, Svetlana Turubanova, Bryan Zutta, Suspense Ifo, Belinda Margono, Fred Stolle et Rebecca Moore. *Humid tropical forest disturbance alerts using Landsat data.* Environmental Research Letters, vol. 11, no. 3, page 034008, 2016.

[Hardoon 2004] David R Hardoon, Sandor Szedmak et John Shawe-Taylor. *Canonical correlation analysis: An overview with application to learning methods.* Neural computation, vol. 16, no. 12, pages 2639–2664, 2004.

[He 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren et Jian Sun. *Deep residual learning for image recognition.* In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[Hearst 1999] Marti A Hearst. *Untangling text data mining.* In Proceedings of the 37th Annual meeting of the Association for Computational Linguistics, pages 3–10, 1999.

[Helber 2019] Patrick Helber, Benjamin Bischke, Andreas Dengel et Damian Borth. *Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 7, pages 2217–2226, 2019.

[Hotelling 1936] Harold Hotelling. *Relations between two sets of variates.* Biometrika, vol. 28, no. 3/4, pages 321–377, 1936.

[Houet 2010] Thomas Houet, Thomas R Loveland, Laurence Hubert-Moy, Cédric Gaucherel, Darrell Napton, Christopher A Barnes et Kristi Sayler. *Exploring subtle land use and land cover changes: a framework for future landscape studies.* Landscape Ecology, vol. 25, no. 2, pages 249–266, 2010.

[Ji 2018] Shunping Ji, Shiqing Wei et Meng Lu. *Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set.* IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 1, pages 574–586, 2018.

[Jones 1972] Karen Sparck Jones. *A statistical interpretation of term specificity and its application in retrieval.* Journal of documentation, 1972.

[Juárez-Orozco 2017] SM Juárez-Orozco, C Siebe et D Fernández y Fernández. *Causes and effects of forest fires in tropical rainforests: a bibliometric approach.* Tropical Conservation Science, vol. 10, page 1940082917737207, 2017.

[Kaggle 2017] Kaggle. *Planet: Understanding the Amazon from Space | Kaggle.* https://www.kaggle.com/c/planet-understanding-the-amazon-from-space, 2017. Accessed: 2018-01-25.

[Kang 1990] Kyo C Kang, Sholom G Cohen, James A Hess, William E Novak et A Spencer Peterson. *Feature-oriented domain analysis (FODA) feasibility study.* Rapport technique, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst, 1990.

[Kergosien 2018] Eric Kergosien, Amin Farvardin, Maguelonne Teisseire, Marie-Noëlle Bessagnet, Joachim Schöpfel, Stéphane Chaudiron, Bernard Jacquemin, Annig Le Parc-Lacayrelle, Mathieu Roche, Christian Sallaberry*et al. Automatic identification of research fields in scientific papers.* arXiv preprint arXiv:1806.03144, 2018.

[Kershaw 2020] Daniel Kershaw et Rob Koeling. *Elsevier OA CC-BY Corpus.* https://data.mendeley.com/datasets/zm33cdndxs/1, August 2020.

[Kusner 2015] Matt Kusner, Yu Sun, Nicholas Kolkin et Kilian Weinberger. *From word embeddings to document distances.* In International conference on machine learning, pages 957–966. PMLR, 2015.

[Kussul 2017] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun et Andrii Shelestov. *Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data.* IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pages 778–782, 2017.

[Lagrange 2015] A. Lagrange, B. Le Saux, A. Beaupère, A. Boulch, A. Chan-Hon-Tong, S. Herbin, H. Randrianarivo et M. Ferecatu. *Benchmarking classification of earth-observation data: From learning explicit features to convolutional networks.* In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pages 4173–4176, July 2015.

[Lancaster 1973] Frederick Wilfrid Lancaster et Emily Gallup. *Information retrieval on-line.* Rapport technique, Melville Publishing Company, 1973.

[Lebedev 2018] MA Lebedev, Yu V Vizilter, OV Vygolov, VA Knyaz et A Yu Rubis. *CHANGE DETECTION IN REMOTE SENSING IMAGES USING CONDITIONAL ADVERSARIAL NETWORKS.* International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, vol. 42, no. 2, 2018.

[Lin 2014] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár et C Lawrence Zitnick. *Microsoft coco: Common objects in context.* In European conference on computer vision, pages 740–755. Springer, 2014.

[Logan 1991] Elisabeth L Logan et WM Shaw. *A bibliometric analysis of collaboration in a medical specialty.* Scientometrics, vol. 20, no. 3, pages 417–426, 1991.

[Lopez 2014] Cédric Lopez, Violaine Prince et Mathieu Roche. *How can catchy titles be generated without loss of informativeness?* Expert systems with applications, vol. 41, no. 4, pages 1051–1062, 2014.

[Lu 2004] Dengsheng Lu, Paul Mausel, Eduardo Brondizio et Emilio Moran. *Change detection techniques.* International journal of remote sensing, vol. 25, no. 12, pages 2365–2401, 2004.

[Lu 2017] Xiaoqiang Lu, Binqiang Wang, Xiangtao Zheng et Xuelong Li. *Exploring models and data for remote sensing image caption generation.* IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 4, pages 2183–2195, 2017.

[Malhado 2014] A. C. M. Malhado, R. S. D. de Azevedo, P. A. Todd, A. M. C. Santos, N. N. Fabré, V. S. Batista, L. J. G. Aguiar et R. J Ladle. *Geographic and Temporal Trends in Amazonian Knowledge Production.* Computers, Environment and Urban Systems, vol. 46, pages 6–13, 2014.

[Mertens 1997] Benoît Mertens et Eric F Lambin. *Spatial modelling of deforestation in southern Cameroon: spatial disaggregation of diverse deforestation processes.* Applied Geography, vol. 17, no. 2, pages 143–162, 1997.

[Mikolov 2013] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado et Jeff Dean. *Distributed representations of words and phrases and their compositionality.* In Advances in neural information processing systems, pages 3111–3119, 2013.

[Miller 1978] L. D. (Lee Durward) Miller, Kaew Nualchawee, C. (Craig) Tom et Goddard Space Flight Center. Analysis of the dynamics of shifting cultivation in the tropical forests of northern thailand using landscape modeling and classification of landsat imagery. Greenbelt, Md. : National Aeronautics and Space Administration, Goddard Space Flight Center, 1978. "May 1978.".

[Milojević 2011] Staša Milojević, Cassidy R Sugimoto, Erjia Yan et Ying Ding. *The cognitive structure of library and information science: Analysis of article title words.* Journal of the American Society for Information Science and Technology, vol. 62, no. 10, pages 1933–1953, 2011.

[Mitchell 2017] Anthea L Mitchell, Ake Rosenqvist et Brice Mora. *Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+.* Carbon balance and management, vol. 12, no. 1, pages 1–22, 2017.

[Mora 2013] Brice Mora et Martin Herold. *REDD+ Measuring, Reporting and Verification–Science solutions to policy challenges*, 2013.

[Mora 2014] Brice Mora, Nandin-Erdene Tsendbazar, Martin Herold et Olivier Arino. *Global land cover mapping: Current status and future trends.* In Land Use and Land Cover Mapping in Europe, pages 11–30. Springer, 2014.

[Mothe 2006] J. Mothe, C. Chrisment, T. Dkaki, B. Dousset et S. Karouach. *Combining mining and visualization tools to discover the geographic structure of a domain.* Computers, Environment and Urban Systems, vol. 30, pages 460–484, 2006.

[Mothe 2018] Josiane Mothe, Faneva Ramiandrisoa et Michael Rasolomanana. *Automatic Keyphrase Extraction Using Graph-Based Methods.* In Proceedings of the 33rd Annual ACM Symposium on Applied Computing, SAC '18, page 728–730, New York, NY, USA, 2018. Association for Computing Machinery.

[Müller 2016] Hannes Müller, Patrick Griffiths et Patrick Hostert. *Long-term deforestation dynamics in the Brazilian Amazon—Uncovering historic frontier development along the Cuiabá–Santarém highway.* International Journal of Applied Earth Observation and Geoinformation, vol. 44, pages 61–69, 2016.

[Neptune 2014] N. Neptune. ánalyses bibliométriques des publications de l'irit. Mémoire de master, Université Paul Sabatier, 2014.

[Neuraz 2020] Antoine Neuraz, Bastien Rance, Nicolas Garcelon, Leonardo Campillos Llanos, Anita Burgun et Sophie Rosset. *The Impact of Specialized Corpora for Word Embeddings in Natural Langage Understanding.* In MIE, pages 432–436, 2020.

[Olofsson 2016] Pontus Olofsson, Christopher E Holden, Eric L Bullock et Curtis E Woodcock. *Time series analysis of satellite data reveals continuous deforestation of New England since the 1980s.* Environmental Research Letters, vol. 11, no. 6, page 064002, 2016.

[Ortega Adarme 2020] Mabel Ortega Adarme, Raul Queiroz Feitosa, Patrick Nigri Happ, Claudio Aparecido De Almeida et Alessandra Rodrigues Gomes. *Evaluation of Deep Learning Techniques for Deforestation Detection in the Brazilian Amazon and Cerrado Biomes From Remote Sensing Imagery.* Remote Sensing, vol. 12, no. 6, page 910, 2020.

[Pautasso 2015] Marco Pautasso, Markus Schlegel et Ottmar Holdenrieder. *Forest health in a changing world.* Microbial Ecology, vol. 69, no. 4, pages 826–842, 2015.

[Pedregosa 2011] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg*et al.* *Scikit-learn: Machine learning in Python.* the Journal of machine Learning research, vol. 12, pages 2825–2830, 2011.

[Peng 2019] Daifeng Peng, Yongjun Zhang et Haiyan Guan. *End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++.* Remote Sensing, vol. 11, no. 11, page 1382, 2019.

[Peters 2018] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee et Luke Zettlemoyer. *Deep contextualized word representations.* arXiv preprint arXiv:1802.05365, 2018.

[Pritchard 1969] A. Pritchard. *Statistical Bibliography or Bibliometrics.* Journal of Documentation, vol. 25, pages 348–349, 1969.

[Radford 2019] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei et Ilya Sutskever. *Language models are unsupervised multitask learners.* OpenAI blog, vol. 1, no. 8, page 9, 2019.

[Radford 2021] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark*et al.* *Learning transferable visual models from natural language supervision.* In International Conference on Machine Learning, pages 8748–8763. PMLR, 2021.

[Reimers 2019] Nils Reimers et Iryna Gurevych. *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.* In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3982–3992, 2019.

[Ronneberger 2015] Olaf Ronneberger, Philipp Fischer et Thomas Brox. *U-net: Convolutional networks for biomedical image segmentation.* In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.

[Roy 2018] Abhijit Guha Roy, Nassir Navab et Christian Wachinger. *Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks.*

In International conference on medical image computing and computer-assisted intervention, pages 421–429. Springer, 2018.

[Russakovsky 2015] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein*et al. Imagenet large scale visual recognition challenge.* International journal of computer vision, vol. 115, no. 3, pages 211–252, 2015.

[Salton 1965] Gerard Salton et Michael E. Lesk. *The SMART automatic document retrieval systems - an illustration.* Commun. ACM, vol. 8, no. 6, pages 391–398, 1965.

[Sampaio 2013] Ricardo Barros Sampaio, Bernard Dousset et Brigitte Gay. *A comparative study between scientific publications and patents: a case of a neglected disease, Leishmaniasis.* In Colloque International VSST'2013: Veille stratégique scientifique & technologique, page 0, 2013.

[Shimabukuro 2000] Yosiu Edemir Shimabukuro, Valdete Duarte, Eliana Maria Kalil Mello et José Carlos Moreira. *Presentation of the Methodology for Creating the Digital PRODES.* Rapport technique, INPE, São José dos Campos, 2000.

[Shimada 2014] Masanobu Shimada, Takuya Itoh, Takeshi Motooka, Manabu Watanabe, Tomohiro Shiraishi, Rajesh Thapa et Richard Lucas. *New global forest/non-forest maps from ALOS PALSAR data (2007–2010).* Remote Sensing of Environment, vol. 155, pages 13–31, 2014.

[Simonyan 2014] Karen Simonyan et Andrew Zisserman. *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.

[Singh 1989] Ashbindu Singh. *Review article digital change detection techniques using remotely-sensed data.* International journal of remote sensing, vol. 10, no. 6, pages 989–1003, 1989.

[Skupin 2014] André Skupin. *Making a Mark: a computational and visual analysis of one researcher's intellectual domain.* International Journal of Geographical Information Science, vol. 28, no. 6, pages 1209–1232, 2014.

[Smeaton 2003] Alan F Smeaton, Gary Keogh, Cathal Gurrin, Kieran McDonald et Tom Sødring. *Analysis of papers from twenty-five years of SIGIR conferences: what have we been doing for the last quarter of a century?* In ACM SIGIR Forum, volume 37, pages 49–53. ACM New York, NY, USA, 2003.

[Smith 2017] Samuel L Smith, David HP Turban, Steven Hamblin et Nils Y Hammerla. *Offline bilingual word vectors, orthogonal transformations and the inverted softmax.* arXiv preprint arXiv:1702.03859, 2017.

[Socher 2010] Richard Socher et Li Fei-Fei. *Connecting modalities: Semi-supervised segmentation and annotation of images using unaligned text corpora.* In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 966–973. IEEE, 2010.

[Socher 2014] Richard Socher, Andrej Karpathy, Quoc V Le, Christopher D Manning et Andrew Y Ng. *Grounded compositional semantics for finding and describing images with sentences.* Transactions of the Association for Computational Linguistics, vol. 2, pages 207–218, 2014.

[Turner 2015] Woody Turner, Carlo Rondinini, Nathalie Pettorelli, Brice Mora, Allison K Leidner, Zoltan Szantoi, Graeme Buchanan, Stefan Dech, John Dwyer, Martin Herold*et al. Free and open-access satellite data are key to biodiversity conservation.* Biological Conservation, vol. 182, pages 173–176, 2015.

[UNEP 2020] FAO UNEP. *The State of the World's Forests 2020. Forests, biodiversity and people.* Forests, Biodiversity and People. FAO, 2020.

[Uribe-Toril 2019] Juan Uribe-Toril, José Luis Ruiz-Real, Julia Haba-Osca et Jaime de Pablo Valenciano. *Forests' first decade: a bibliometric analysis overview.* Forests, vol. 10, no. 1, page 72, 2019.

[Uzkent 2019] Burak Uzkent, Evan Sheehan, Chenlin Meng, Zhongyi Tang, Marshall Burke, David Lobell et Stefano Ermon. *Learning to interpret satellite images in global scale using wikipedia.* arXiv preprint arXiv:1905.02506, 2019.

[Van Leeuwen 2010] Willem JD Van Leeuwen, Grant M Casady, Daniel G Neary, Susana Bautista, José Antonio Alloza, Yohay Carmel, Lea Wittenberg, Dan Malkinson et Barron J Orr. *Monitoring post-wildfire vegetation response*

with remotely sensed time-series data in Spain, USA and Israel. International Journal of Wildland Fire, vol. 19, no. 1, pages 75–93, 2010.

[Vargas 2019] Christian Vargas, Joselyn Montalban et Andrés Alejandro Leon. *Early warning tropical forest loss alerts in Peru using Landsat.* Environmental Research Communications, vol. 1, no. 12, page 121002, 2019.

[Vaswani 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser et Illia Polosukhin. *Attention is all you need.* Advances in neural information processing systems, vol. 30, 2017.

[Wang 2018] Liwei Wang, Yin Li, Jing Huang et Svetlana Lazebnik. *Learning two-branch neural networks for image-text matching tasks.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pages 394–407, 2018.

[Wheeler 2014] David Wheeler, Dan Hammer, Robin Kraft et Aaron Steele. *Satellite-Based Forest Clearing Detection in the Brazilian Amazon: FORMA, DETER, and PRODES In.* World Resources Institute Issue Brief, pages 1–24, 2014.

[Yakubovskiy 2020] Pavel Yakubovskiy. *Segmentation Models Pytorch.* `https://github.com/qubvel/segmentation_models.pytorch`, 2020.

[Yang 2010] Yi Yang et Shawn Newsam. *Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification.* GIS '10, page 270–279, New York, NY, USA, 2010. Association for Computing Machinery.

[Young 2014] Peter Young, Alice Lai, Micah Hodosh et Julia Hockenmaier. *From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions.* Transactions of the Association for Computational Linguistics, vol. 2, pages 67–78, 2014.