# Automatic Annotation of Change in Earth Observation Imagery

Nathalie Neptune
nathalie.neptune@irit.fr
Under the supervision of
Josiane Mothe
IRIT, UMR 5505, CNRS
Toulouse, France

## ABSTRACT

Earth observation images are plentiful and have been used for several decades for many tasks such as classifying land cover, monitoring agriculture and more generally for detecting changes happening on the surface of the Earth. There exist however several challenges in using these images in information systems. One of them is the annotation of the images with semantic descriptors. To address this, as our PhD main objective, we propose a new approach for the semantic annotation of changes detected in satellite images using annotations from an unaligned text corpus. Using scientific publications on the areas of interest and related to the changes observed in the images, we will extract candidate keywords for the annotations. State of the art image change detection techniques will be used to find the regions within the images where the changes have occurred. The keywords and change pixels will be matched by jointly embedding them into a low dimension space.

## CCS CONCEPTS

**• Computing methodologies → Matching**; **• Applied computing → Annotation**; **Environmental sciences**.

## KEYWORDS

Image annotation, Multimodal learning, Text-image matching

## 1 INTRODUCTION

The comparison of two earth observation (EO) images of an area, taken at different times, allows to detect whether changes have occurred in that area during that time frame. This process is called change detection [21]. Various applications call for the use of change detection techniques such as the monitoring of forest disturbance and loss [11, 25], the tracking of urban change [23], and the mapping of changes caused by natural disasters [3, 7].

Semantic annotation of images involves associating them with semantic keywords [5]. Likewise, adding semantic information to changes detected in EO images results in semantic change detection. There exist several collections of annotated EO data that can be used for semantic change detection such as [11], [8], [17] and [9]. However, many are either limited by their geographical coverage [17] [9], or by their temporal coverage [9] or by their topical coverage [11]. Moreover, all the previously cited image collections are provided on a yearly basis only. Therefore, using these datasets for semantic change detection may not be suitable in cases where the goal is to detect changes in a timely manner.

Therefore, to detect changes that are happening in an area of interest at a specific time, one might need to use images that have not yet been annotated. In fact, unannotated EO images are plentiful. The Landsat and Copernicus programs provide free access to the images produced by their satellite missions with new images made available every day. However, using these images for semantic change detection poses the challenge of correctly identifying the classes for each pixel of each image. Without labels, changes can still be detected by comparing the images, however the semantic information about those changes will be missing. Without this information it is not possible to tell for example that a piece of land previously covered by trees has now changed into a road. Which is why class labels or annotations for each pixel of each image need to be acquired either through human (expert) annotators or automatic methods. The latter can fill the gap in a context where expert annotators are not available and few or no annotated images exist for a given area. In fact, machine learning algorithms can learn to automatically annotate images from previously annotated examples. A sufficient number of examples must still be provided to train these algorithms.

Supervised machine learning has been successfully used for semantic change detection [6]. Such models are trained on images along with their semantically segmented 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 [19]. A similar approach is not well suited for EO images because some expertise or training might be required to properly identify and differentiate among classes. As a result, automatic and semi-automatic approaches for the annotation task are commonly used to build large EO data sets such as [20] and [11].

Scientific literature published by researchers who work with earth observation 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 interpreting 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 [2].

With visual semantic embeddings we learn to represent textual and visual data in the same latent space. The closer those data points are semantically, the smaller their distance is in the joint space. Several approaches have been proposed including the joint embedding of images and words into a common low dimension space [10] for image classification, and the embedding of images and sentences into a common space for image description [22].

https://www.usgs.gov/land-resources/nli/landsat
https://www.copernicus.eu/en

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 tree loss cover in tree covered areas to test our approach. For these cases, the texts may not be available at testing time. The model needs to learn from the labels and make predictions without extracting new annotations when images are provided without accompanying text at testing time. The text data in this case is considered privileged information as defined by [24] and is used only when training the model.

By using joint image and text embeddings we will automatically assign relevant annotations to image pairs and segments where changes are detected. Therefore we will perform two core tasks, change detection and the annotation of the changes. We propose to use scientific publications as a source of annotations which will be extracted using a neural language model. These annotations can be used in subsequent tasks such as image indexing and retrieval.

## 2 BACKGROUND

The first step in the semantic annotation of changes in EO images is to locate the changes within the images. Binary change detection methods only detect if and where a change occurred, while semantic change detection methods also specify semantic information about the change that is detected [14]. Both types of methods have been applied to satellite images to detect land cover changes with the most recent ones using deep learning models [6, 16].

Dault *et al.* [6] proposed a change detection deep learning model for satellite images based on U-Net [18] which is an encoder-decoder architecture with skip connections between the encoding and decoding streams. The U-Net architecture was initially proposed for the segmentation of biomedical images. Three variations of the model are proposed. The first model 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 models are siamese variations 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. Another variation of U-Net with early fusion which uses dense skip connections was proposed by [16].

When the semantic information about the classes present in the image pair is inconsistent or lacking, one solution is to include other sources of information such as ontologies [3] or geo-referenced Wikipedia articles[23].

Our proposed approach combines visual semantic embeddings for annotating changes, and adds the semantic label to binary change detection. The semantic labels will be extracted from a text corpus made of publications related to the area and the type of change of interest in time and space, using word embeddings. The changes will be detected using change detection methods from the state of the art [6, 16] with the extracted labels added as privileged information in the learning process using heteroscedastic dropout which was proposed by [12] as a way to learn using privileged information in deep neural networks. The variance of the heteroscedastic dropout is a function of the privileged information.

When a machine learning algorithm can, at testing time, identify classes that were not previously seen during training, it is performing "zero-shot" classification. Few-shot learning happens when the number of examples of a class is very low, for training. Both zero-shot and few-shot learning methods have been proposed for learning classes with missing labels or with insufficient examples in the training data for supervised land cover and land use classification models [13].

Our proposed approach differs from [13] in that we will perform change detection with a deep learning model and our word vectors will also be trained on a specialized corpus of scientific documents. This should make our model more scalable while providing more relevant annotations specifically for change detection as opposed to classification alone. Unlike [6, 16] our model 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 [4]. Our approach is closer to [23], however they only take into account the location of the text that is used to learn the annotations, they do not consider the temporal relation between the text and the images which is useful for change detection.

## 3 RESEARCH QUESTIONS

Our primary goal is to perform semantic change detection on satellite images by combining image change detection with text information extraction, adding semantic annotations to the detected changes. To the best of our knowledge this is the first time this type of approach has be proposed for the semantic change detection problem. In addition, we seek to improve the performance of the image pair change detection task by using these extracted annotations when training the model.

### 3.1 Hypotheses

When related scientific documents are used, relevant annotations for change detection in image pairs can be extracted and used for semantic change detection in the absence of labels provided by experts. This implies that the most important changes have been studied and documented by scientists and can be found in their publications. With a visual-semantic model, we can make the link between the changed segments and the labels in zero and few-shot learning cases. By using these annotations as privileged information when training the model, the change detection algorithm should be able to detect changes with higher accuracy compared to when images alone are used.

### 3.2 Experimental setup

To test our model, we will use images from an area of interest in Madre de Dios, in Peru, corresponding to the Tile 19LCF from the Sentinel satellite missions. We will use all available Sentinel-1 and Sentinel-2 images of the area from 2015 to 2017. This region was chosen because it is an area where there are a lot of reported deforestation incidents. Madre de Dios was the department in Peru with the second largest area affected by deforestation in 2017, according to a report by the Ministry of Environment of Peru [15].

https://sentinel.esa.int/web/sentinel/missions

We collected abstracts of publications from the Web of Science to create our corpus. We retrieved a total of 298 publications using the keywords "Madre de Dios" for the topic and taking only articles and proceeding papers for the years 2000 to 2019. Using the name of the area allowed us to exclude articles that were not relevant to the location. By only taking papers published in this 19 year period we avoid those that were written too early before the Sentinel missions started in 2014. However, limiting the collection only to articles published after 2014 yields too few results. Articles published after 2017 might also include changes observed after our most recent image, but including them allows for better coverage of changes present in our image dataset because they are more likely to appear in a publication months or years later.

For the image (change/no-change) labels we use data from two sources, the GLAD alert system [11] and the Geobosques system [25]. The alerts will be used as labels for a supervised change detection learning algorithm applied to images of tree covered areas. For each pixel these alerts indicate whether there was tree cover loss or not. We will use the definition of tree cover as vegetation that is at least 5 meters tall with a minimum of 60% canopy cover as it is defined in the GLAD alert system.

Our deep learning model will be based on binary change detection methods [6, 16] for image pairs. The model will be adapted to use the semantic information extracted from our corpus to learn the annotations for the images and annotate images with classes that may not have been seen during training.

Let us consider an example where there are confirmed tree cover loss alerts for January 2, 2017, in our area of interest. The alerts are images of the area with all pixels at 0 except for those for which an alert was generated. To learn to detect those change pixels, one image for each satellite is taken before and after the date of the alert. The images from different satellites are stacked together to create a single image. In this case, one image from Sentinel-1 is taken on December 27, 2016 and on from Sentinel-2 on December 28, 2016. Then two other images are similarly taken from January 2 and January 3, 2017, to see the area after the alerts. The model learns to identify the change pixels from the alerts by comparing both images. In parallel, we take the corpus made of the publications about the same area and extract the word vectors. We then find the word vectors from the corpus which are most similar to our change class word vectors. This will give us a first set of additional labels for our change map. Then we cluster the word vectors from the corpus to find the ones that more often occur with our change word vectors.

## 4 METHOD

By integrating an ontology to the segmentation process of pre and post-disaster images, [3] showed that overall accuracy went from 67.9% to 89.4% for images of their test area. With a reduced number of samples (200), [23] demonstrated that using Wikipedia annotations for the task of semantic segmentation, the Intersection-over-Union (IoU) score 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 [3] improved greatly the accuracy of the classification algorithm, it came at the high cost of expert hours. The crowdsourcing approach using Wikipedia data in [23] while promising, resulted only in modest improvements for the semantic segmentation task.

Ideally, we would like to combine the benefits of knowledge from experts and big data from crowdsourcing. We propose to use the expert knowledge through relevant scientific articles from which annotations will be extracted. Adapted versions of state of the art change detection models from [6] and [16] will be used, to test on images from our area of interest. Both Sentinel-2 optical imagery and Sentinel-1 radar imagery will be included. Using images from a tropical rainforest area means that there will be significant cloud cover making a large number of pixels, in the optical images, non usable for change detection. The radar imagery will help mitigate this problem.

Our proposed method will therefore perform change detection, in satellite image pairs by predicting change pixels. The method will also perform semantic annotation of the detected changes by predicting their labels.

For the annotations, we will use word embeddings trained on Wikipedia and align them with our text dataset using the RCSLS criterion [1]. These embeddings cover a large vocabulary on many topics, most of which are not relevant to our case. This is why we will realign them with embeddings from our specific corpus to ensure that the distribution of words matches our intended topic.

The word embeddings are used to look up our change labels. Using an approach similar to [10], for each pixel, the model will predict its label as its vector representation. In [10], the vector representation 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.

We will also add the text as privileged information during training using heteroscedastic dropout [12]. This heteroscedastic droptout will be used in order to improve the performance of the model on our relatively small dataset.

While the model does not explicitly take into account the temporal relation between the text and the image we will use documents from a limited time period that covers the time before, during and after the changes occur. In doing so, we expect to limit the number of extracted words that are temporally out of scope.

We want to have a model that is suitable for environmental applications and therefore we will test on this type of data first. However, applications in other domains should be possible, provided there exist a relevant corpus that can be used with the images. We will evaluate the performance of our proposed method using precision and recall statistics and the Intersection over Union score.

We will compare our method to state of the art methods for change detection [6, 16] on the Madre de Dios data set described in section 3.2.

## 5 CONCLUSION AND DISCUSSION

The automatic annotation of change in satellite imagery using annotations extracted from relevant scientific literature will demonstrate how expert knowledge can be gathered from text, in an unsupervised way to add semantic information to changes detected in images. Furthermore, for the change detection task, the use of

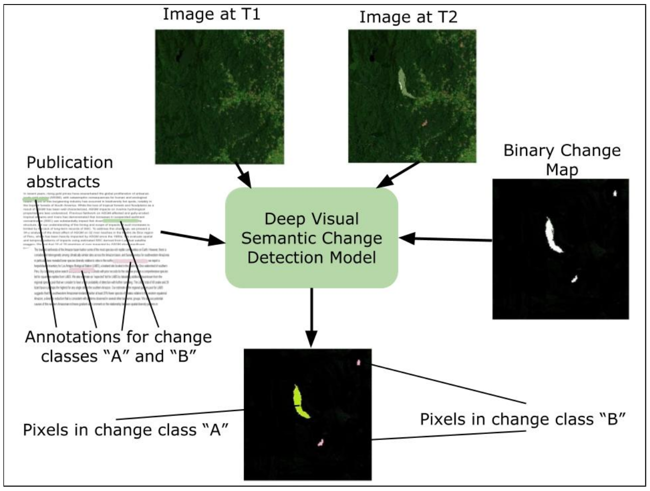
Image at T1
Image at T2
Publication abstracts
Binary Change Map
Deep Visual Semantic Change Detection Model
Annotations for change classes "A" and "B"
Pixels in change class "A"
Pixels in change class "B"

**Figure 1: Overview of our proposed approach for automatic annotation of changes in EO images.**

these annotations as privileged information when training the algorithm should result in higher overall accuracy. This may provide a method that could be used in tools for non-experts to find and identify changes in satellite images of their areas of interest, using open data even when no expert annotations are available. An extension of this work could be to extract triples of events with location and time from the full text of the publications, and using them to construct a timeline of change events with the corresponding time series of satellite images. Unlike the loose temporal relation of the currently proposed method, the temporally annotated triples should provide more exact matches to the changes observed on the images as they would represent a phenomenon along with its location and its temporal information. Our hypothesis is that this would also make the candidate labels less noisy overall.

## ACKNOWLEDGMENTS

The author would like to thank Dr. Brice Mora and Julius Akinyemi for their feedback and advice on the research project. We are grateful for the suggestions and comments of the anonymous reviewers. This work is supported by the Schlumberger Foundation hrough the Faculty for the Future program.

## REFERENCES

[1] Piotr Bojanowski, Onur Celebi, Tomas Mikolov, Edouard Grave, and Armand Joulin. 2019. Updating Pre-trained Word Vectors and Text Classifiers using Monolingual Alignment. *arXiv preprint arXiv:1910.06241* (2019).

[2] Lutz Bornmann and Rüdiger Mutz. 2015. 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* 66, 11 (2015), 2215–2222.

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

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

[5] Gustavo Carneiro and Nuno Vasconcelos. 2005. Formulating semantic image annotation as a supervised learning problem. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2. IEEE, 163–168.

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

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

[8] ESA. 2017. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017). maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

[9] ESA. 2017. S2 prototype Land Cover 20m map of Africa 2016. http://2016africalandcover20m.esrin.esa.int/

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

[11] Matthew C Hansen, Alexander Krylov, Alexandra Tyukavina, Peter V Potapov, Svetlana Turubanova, Bryan Zutta, Suspense Ifo, Belinda Margono, Fred Stolle, and Rebecca Moore. 2016. Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters* 11, 3 (2016), 034008.

[12] John Lambert, Ozan Sener, and Silvio Savarese. 2018. Deep learning under privileged information using heteroscedastic dropout. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8886–8895.

[13] Aoxue Li, Zhiwu Lu, Liwei Wang, Tao Xiang, and Ji-Rong Wen. 2017. Zero-shot scene classification for high spatial resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 55, 7 (2017), 4157–4167.

[14] Dengsheng Lu, Paul Mausel, Eduardo Brondizio, and Emilio Moran. 2004. Change detection techniques. *International journal of remote sensing* 25, 12 (2004), 2365–2401.

[15] MINAM. 2018. *Cobertura y deforestación en los bosques húmedos amazónicos 2017.* Technical Report. Programa Nacional de Conservación de Bosques para la Mitigación del Cambio Climático del Ministerio del Ambiente.

[16] Daifeng Peng, Yongjun Zhang, and Haiyan Guan. 2019. End-to-End Change Detection for High Resolution Satellite Images Using Improved UNet++. *Remote Sensing* 11, 11 (2019), 1382.

[17] DA Roberts, M Toomey, I Numata, TW Biggs, J Caviglia-Harris, MA Cochrane, C Dewes, KW Holmes, RL Powell, CM Souza, et al. 2013. LBA-ECO ND-01 Landsat 28.5-m Land Cover Time Series, Rondonia, Brazil: 1984-2010. *ORNL DAAC* (2013).

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

[19] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115, 3 (2015), 211–252.

[20] Yosiu Edemir Shimabukuro, Valdete Duarte, Eliana Maria Kalil Mello, and José Carlos Moreira. 2000. *Presentation of the Methodology for Creating the Digital PRODES.* Technical Report. São José dos Campos.

[21] Ashbindu Singh. 1989. Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing* 10, 6 (1989), 989–1003.

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

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

[24] Vladimir Vapnik. 2006. *Estimation of dependences based on empirical data.* Springer Science & Business Media.

[25] Christian Vargas, Joselyn Montalban, and Andrés Alejandro Leon. 2019. Early warning tropical forest loss alerts in Peru using Landsat. *Environmental Research Communications* 1, 12 (2019), 121002.