Annotation automatique d'images: le cas de la déforestation

Duy Huynh[1] Nathalie Neptune[1]
(1) Institut de Recherche en Informatique de Toulouse, UMR5505 CNRS
Université de Toulouse
118 Route de Narbonne, F-31062 Toulouse Cedex 9, France
duy.huynh@irit.fr, nathalie.neptune@irit.fr
Under the supervision of
Josiane MOTHE (1)

RÉSUMÉ

Cet article correspond à un état de l'art sur le thème de l'annotation automatique d'images d'observation de la terre pour la détection de la déforestation. Nous nous intéressons aux différents challenges que recouvre le domaine et nous présentons les méthodes de l'état de l'art puis les pistes de recherche que nous envisageons.

ABSTRACT

Automatic image annotation : the case of deforestation.

This paper aims to present the state of the art of the methods that are used for automatic annotation of earth observation image for deforestation detection. We are interested in the various challenges that the field covers and we present the state of the art methods and the future research that we are considering.

MOTS-CLÉS : Recherche d'information, annotation d'images, réseaux de neurones convolutionnels, détection de la déforestation.

KEYWORDS: Information retrieval, image annotation, CNN, convolutional neural networks, deforestation detection.

1 Introduction

According to the National Geographic, forest covers about 30% of the planet. Forest ecosystems play an essential role in supporting life on the earth such as supplying wood, water regulation, preventing storms and soil erosion and forests store rare genetic resources for our planet. Deforestation affects the environment in a multitude of ways. The most obvious effects are global warming and loss of biodiversity. From the photosynthetic function of trees, forests release oxygen and absorb carbon dioxide. The fewer forests, the more carbon dioxide entering the atmosphere, increasing the speed of global warming. In addition, earth's forests are home to over 80% of plants and animals but deforestation destroys these habitats, diminishing biodiversity and causing the extinction of four to six thousand rainforest species every year (Geographic, 2017). Direct causes of deforestation are agricultural expansion, logging and wood extraction, bio fuels from palm oil, infrastructural expansion such as road and urbanization, and mining (Tariq & Aziz, 2015) (see also `http://www.`

theworldcounts.com/stories/deforestation-facts-and-statistics).

Satellite remote sensing makes it possible to observe the earth and help detect deforestation in a faster way than ever. For example, the Copernicus program provides images from any region every 5 days [1]. Moreover, with the advances of computing power and machine learning techniques, it becomes possible to detect deforestation automatically from earth observation (EO) images (Achard *et al.*, 2002; O'Connor *et al.*, 2015).

This paper reviews the state of the art of the methods that are used in the domain of automatic deforestation detection. When using remote sensing, deforestation detection is an application of change detection. It generally consists in two steps : the annotation of satellite images in order to identify the land cover and the change detection on zones that are identified as forest. In these two steps, convolution neural networks (CNN) were proven successful. This is because CNN allow for a large learning capacity(Krizhevsky *et al.*, 2012) ; moreover their automatic feature extraction capability is very useful in image analysis (Witten *et al.*, 2016). Figure 1 shows the basic workflow of automatic deforestation detection using two images of the same area captured at different times. The first step is linked to information retrieval and can be paralleled with the tasks developed in ImageCLEF for automatic association of concepts to images in medicine (Ionescu *et al.*, 2017), although it goes a step further since image segmentation is usually considered to annotate the images in the case of deforestation detection while it is usually an index only in information retrieval.

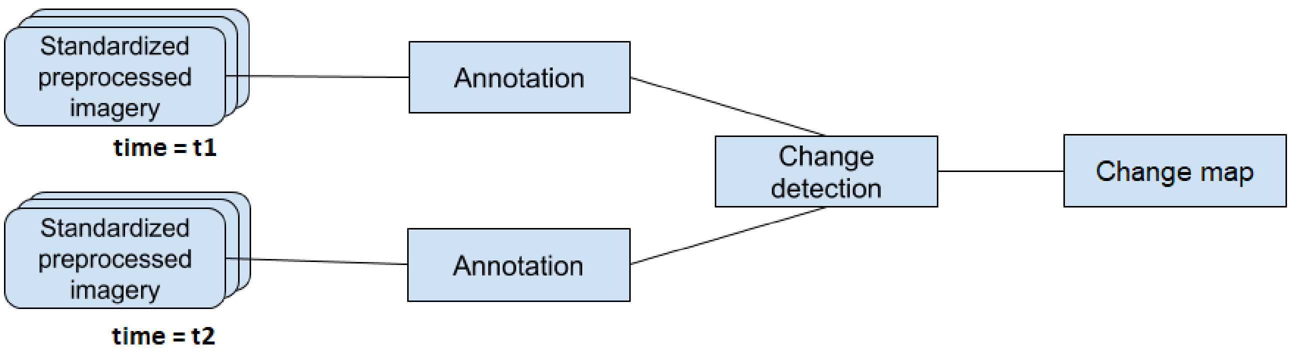

FIGURE 1: Basic workflow for annotation and change detection in satellite imagery of the same area taken at two different times.

The remainder of this paper is as follows : sections 2 and 3 present the state of the art of the two steps of the automatic deforestation detection problem. In section 4, we present the main current challenges which also correspond to the ones that we would like to tackle in our future work. Section 5 concludes this paper. Additionally, a brief explanation of how convolutional neural networks work is presented in the appendix.

1. "Copernicus, previously known as GMES (Global Monitoring for Environment and Security), is the European Programme for the establishment of a European capacity for earth Observation." www.copernicus.eu

2 Annotation of land cover

Image annotation is a general problem which is not specific to forest detection but rather that covers a large variety of domains such as information extraction from medical images for disease detection (Ionescu *et al.*, 2017),(Mothe *et al.*, 2017), and image retrieval (Babenko *et al.*, 2014). In the case of deforestation detection, what we need is land cover information. The automatic annotation of earth observation images to associate land cover types with areas is typically done with classification methods. Our intention is not to provide a state of the art of image classification methods but rather to detail how these methods have been used in the case of land cover annotation with an emphasis on deforestation detection.

As in many fields, deep learning has revolutionised the domain of image annotation for deforestation detection in remote sensing. Without aiming to be exhaustive, in this section, we first present some of the approaches that have been developed for land cover annotation of satellite imagery before the deep neural networks arose, then we present how deep learning is used in this domain.

2.1 Methods used before deep learning

Classification-based methods are very commonly used when having to annotate satellite images for deforestation detection. (Shimabukuro *et al.*, 1998) and (Müller *et al.*, 2016) are two examples of such methods, the first one opting for an unsupervised classification approach by segment while the second one is using a supervised pixel-by-pixel approach.

To generate deforestation maps and provide related information on areas experiencing deforestation, (Shimabukuro *et al.*, 1998) proposed an approach based on shade fraction image generated by a spectral mixture model, then "region growing" segmentation and unsupervised classification (clustering) of fields were applied. At the time, the state of the art was visual interpretation of satellite imagery or classification based on pixel-by-pixel analysis without contextual information. However, visual interpretation is laborious and thus costly especially when dealing with a large number of small surfaces to label in the same area. Using shade fraction images instead of working with the Landsat images directly allowed for reduced processing time for classification and post-processing time for manual removal of undetermined classes. Their results were validated against results from conventional techniques in use at the time on images from Rondônia, in Brazil.

Yearly deforestation patterns for an area of the Amazon forest in Mato Grosso and Pará were derived from Landsat TM [2] and ETM+ [3] images captured between 1985 and 2012, by (Müller *et al.*, 2016). The proposed approach uses a random forest classifier on the training data made of labelled samples of the values of reflectance images and minimum tasseled cap wetness (TCW) observations, for each pixel. An overall accuracy of 85% was reached with 95% confidence interval margin of $\pm 2\%$. Classification error tended slightly towards late detections. Higher deforestation rates were found compared to the state of the art deforestation datasets for the same region and over the same time

2. Landsat Thematic Mapper (TM) is a an advanced multi-spectral scanning sensor carried by Landsat 4 and 5 and featuring seven spectral bands. With band 6 being a thermal infrared radiation sensor. The TM sensor has a spatial resolution of 30 m X 30 m (120 m for band 6) and a temporal resolution of 16 days. https://landsat.gsfc.nasa.gov/the-thematic-mapper/

3. Landsat Enhanced Thematic Mapper Plus (ETM+) is an enhanced version of Landsat TM carried by Landsat 7. ETM+ is a multi-spectral scanning radiometer with eight bands. It has a spatial resolution of 30 m X 30 m (60 m for the thermal band, 15 m for the panchromatic band). https://landsat.gsfc.nasa.gov/the-enhanced-thematic-mapper-plus/

period.

The next subsections review the approaches in the literature that use convolutional neural networks (CNN) to annotate remote sensor data, in particular for the detection and analysis of deforestation.

2.2 Methods using convolutional neural networks

Several approaches using neural networks, including convolutional neural networks have been proposed for annotating satellite images, such as (Kussul *et al.*, 2017) and (Zhang *et al.*, 2017). In the first approach the annotation is done by segment while it is done by pixel in the second approach.

(Kussul *et al.*, 2017) proposed a deep learning approach for the classification of multi-source satellite images. They used a combination of supervised and unsupervised neural networks to segment and classify satellite imagery from Landsat 8 and Sentinel 1A. Testing was done with data from the Joint Experiment of Crop Assessment and Monitoring (JECAM) test site in Ukraine. While this study focused mostly on crop identification, the area was labelled with eleven land cover types among which the "forest" type. This approach was tested for overall accuracy against the following approaches : random forest (RF), and an ensemble of neural networks (ENN) made of multilayer perceptrons. The proposed ensemble of 2D CNN outperformed these two other methods in overall accuracy. However, for the forest class, the difference is only a few decimal points because the random forest and ENN had already reached over 99% accuracy.

The approach from (Zhang *et al.*, 2017) classifies fine resolution images from remote sensors with a model integrating a CNN and a multilayer perceptron (MLP). This model was compared to standalone standard pixel-based CNN and MLP classifiers and a pixel-based texture MLP based on the standard Grey Level Co-occurrence Matrix (GLCM-MLP). The authors tested the method on data from Southampton and its surroundings, in the UK. A total of eight land cover classes were detected and large patches of forest in rural areas. While they did not have a forest class, forest patches were put in the "Trees" class which is described as "large patches of deciduous trees" and "patches of tree species". The overall accuracy of the MLP-CNN approach, including for the Trees class, on the two test sites was found to be higher than that of the other three methods.

In the next subsection we discuss a Kaggle [4] competition aimed at automatically classifying forest images for deforestation detection and the prevalence of CNN models.

2.3 The Kaggle competition for deforestation detection

Deforestation can also be detected in a single image by detecting known deforestation patterns in a forested area. The "Planet Deforestation Detection" Kaggle competition launched by Planet Labs [5] in April 2017, proposed such task.

The competition aimed to find effective methods to track forest changes by using high resolution satellite imagery. The images were provided as image chips and the task was to automatically annotate each image chip with its corresponding labels for atmospheric condition and land cover. A total of 17 possible labels were defined. The competition started on April 20, 2017 and ended on July 20, 2017 (Kaggle, 2017).

4. Kaggle is a data science and machine learning competition platform. `https://www.kaggle.com/`
5. Planet Labs is an earth observation satellite company based in San Francisco `https://www.planet.com/`

The chips are extracted from satellite images and represent approximately one square kilometre (Scott, 2017). These chips are from Planet's full-frame analytic scene products and were provided in GeoTiff formats with four bands of data and in JPG format. The scenes included were exclusively from the Amazon basin (?). The training set contained over 40000 images while the test set contained over 20000 images.

The F2-score, which is a weighted average of precision and recall, was used to rank the submissions.

The formula for calculating the F2-score was given as follows (Kaggle, 2017) :

$(1+\beta^2)\frac{P.R}{\beta^2 P+R}$ where $P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}, \beta = 2.$

P is the precision and R is the recall, TP stands for true positives, FP for false positives and FN for false negatives, β is a float which in this case is equal to 2 to indicate that a higher weight is given to the recall than to the precision.

There were some data quality issues with TIFF images not matching their corresponding JPG images in the test set and there were labels incorrectly assigned as well. Ambiguous labelling was in several cases due to the fact that from the image alone certain classes could not be told apart by visual interpretation. The winning model used exclusively the JPG images with several data augmentation techniques to pre-process the images such as image rotation and haze removal (He *et al.*, 2011). A model made of 11 CNN was used with existing CNN architectures such as Inception, Resnet and Densenet. Then ridge regression was used on the probabilities obtained for each label. The CNN were then combined with a ridge regression model. Finally, the loss function was designed to take into account the F2-score used for evaluating the submissions [6] (Kaggle, 2017).

A total of 938 submissions were made for this competition (Kaggle, 2017). Of the top 16 teams from the private leader-board, 7 reported using CNN models [7]. (Lagrange *et al.*, 2015) had found that for semantic labelling of earth observation images, the best performance is obtained when using deep convolutional neural networks compared to expert classifiers and spectral support-vector classification.

Image annotation is the first step in change detection, and deep learning can be used for automatically learning features and annotating the images. The result of the image annotation task is images that are semantically labelled and can be compared to detect changes (Hirschmugl *et al.*, 2017). In the next section we will go over the methods that are commonly used for change detection in remote sensing imagery.

3 Change detection in remote sensing images

In remote sensing, the use of satellite images to assess and then map deforestation is one of the applications of change detection. Two methods are applied for change detection with optical images : image to image change detection and time series analysis. Image-to-image change detection requires a minimum of two images captured at different times while time series analysis requires a series

6. `http://blog.kaggle.com/2017/10/17/planet-understanding-the-amazon-from-space-1st-`
7. `https://www.kaggle.com/c/planet-understanding-the-amazon-from-space/discussion/36732`

of images captured over a period of time. Image-to-Image change detection is more commonly used. Performing change detection over long time periods may require the use of images captured by different remote sensing devices with different characteristics which then requires data fusion techniques to combine these data together.

3.1 Methods prior to deep learning

Deforestation detection is one of the major applications of satellite image change detection. The problem has been studied for decades and many techniques have been proposed. These techniques can be categorized into different approaches such as algebra, transformation, classification, advanced models, geographical information system (GIS) approaches, or visual analysis (Lu *et al.*, 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 *et al.*, 2004). This technique was widely used in change detection problems, particularly for detecting forest changes as in (Miller *et al.*, 1978) and others (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 *et al.*, 2004).

(Collins & Woodcock, 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 to image change detection and for time series change detection like in (Mertens & Lambin, 1997) and (Olofsson *et al.*, 2016) respectively.

(Mertens & Lambin, 1997) proposed a model to detect deforestation in southern Cameroon based on remote sensing data from Landsat MSS sensor [8] for the years 1973 and 1986. Photographs, digitized topographic maps, aerial photos and other remote sensing data and population data were used as well. In addition, ground observation data were collected and used to validate the land cover and land cover change maps. A maximum likelihood classifier was used to generate land cover maps for 1973 and 1986, independently for the areas common to the two Landsat images. The classifier was trained on field observations and interpretations of aerial photos of unchanged areas for 1979. The accuracy was evaluated with field observation data and low-altitude aerial photos. For the classification task, 90% accuracy was reached for 1973 but only 85% for 1986 due to haze. Five classes were derived among which the "dense forest" which the authors define as "evergreen or moist deciduous forest zones, dominated by trees at least 5 m high and with a forest-cover proportion of 30 per cent or more", the four other classes were for different non-forested areas.

8. The Landsat MSS sensor is a multi spectral scanner carried by Landsat satellites one through five, with a spatial resolution of 68 m X 83 m and a temporal resolution of 18 days. `https://landsat.gsfc.nasa.gov/the-multispectral-scanner-system/`

A time series analysis was performed by (Olofsson *et al.*, 2016) to reveal deforestations trends in the New England area in the North East of the United States with the goal of modelling the impact that forest changes have on the carbon balance of the planet. To map the land cover change, the Continuous Change Detection and Classification (CCDC) algorithm was used on pixel-level time series of Landsat data available for the area from 1985 to 2011. Areas of forest harvest were not added to the deforestation estimate. A random forest classifier was used on the training data with the attributes from the time series prediction model for each time series. To account for bias caused by classification errors, the mapped area is estimated from a random sample of reference observations then visual examination is performed to confirm the labels. This study detected the land cover changes but also capture the evolution of forest loss over time. Accelerated deforestation was found to have taken place in the 1990 then around 2007 the deforestation rate stabilized.

While change detection with optical imagery is fairly common, less work has been done with radar imagery especially with time series. However there have been some contributions on the subject such as in (Barreto *et al.*, 2016).

A change detection method based on object detection for high resolution Synthetic Aperture Radar (SAR) data was proposed by (Barreto *et al.*, 2016). This method involves three steps. First, multi-temporal Xband high resolution SAR image segmentation, followed by feature extraction and finally, area detection and classification. The image is segmented into superpixels with a simple linear iterative clustering algorithm (SLIC). The features are extracted with the object correlation images (OCI) framework and with gray-level cooccurrence matrix (GLCM). Areas are detected and classified into unchanged, deforestation and other changes classes with a multilayer perceptron. Experts manually annotated a set of multi-temporal Xband SAR images captured from August to November 2015 with 2,263 regions of interest. Results showed an improvement of 10% in accuracy, compared to the state of the art approaches, in change detection and classification for the deforested areas. The proposed approach was compared to the following state of the art object-based approaches : Optimum Path Forest (OPF) clustering for image segmentation, OCI alone for feature extraction, SVM and OPF-classifier for classification.

3.2 Method using deep learning

(Khan *et al.*, 2017) proposed a new deep CNN model for object-based change detection in incomplete satellite imagery. Their approach performs two tasks, first data recovery to fill data that is missing due to limited camera aperture, cloud cover, and sensor artefacts, then change detection which is treated as a regions classification problem. For this second task, object-based change detection was used without domain knowledge to extract features, instead, all features were learned with a deep neural network.

The model was used to analyse satellite data on the north-east region of Melbourne, in Victoria, Australia. The authors were able to detect forest cover change on images taken from 1987 to 2015. This approach outperforms baseline techniques in temporal change detection and patch-wise classification tasks.

For start-time and end-time predictions for detected change events, this approach outperforms the state-of-the-art approaches, one based on hand-crafted features for classification and the second one based on bag of visual words for classification. To establish the baseline with these two state-of-the-art approaches, dense scale invariant feature transform (SIFT) descriptors were used as a baseline for change detection and linear SVM, kernel SVM and random forest (RF) were used for prediction.

3.3 Summary of state of the art approaches and cited papers

Table 1 shows the papers that were cited for annotation and change detection with earth observation images. The papers are categorized by the type of approach used. Among the papers cited, most have used a classification-based approach, including the ones using CNN.

TABLE 1: Publications cited and the methods that they propose for each annotation and change detection tasks.

METHOD	TASK	
	Annotation	**Change detection**
Algebra based approach		(Miller *et al.*, 1978)
Classification based	*Prior to CNN*	
	(Miller *et al.*, 1978) (Mertens & Lambin, 1997) (Shimabukuro *et al.*, 1998) (Müller *et al.*, 2016) (Olofsson *et al.*, 2016) (Barreto *et al.*, 2016)	(Müller *et al.*, 2016) (Olofsson *et al.*, 2016) (Barreto *et al.*, 2016)
	CNN	
	(Khan *et al.*, 2017) (Kussul *et al.*, 2017) (Zhang *et al.*, 2017) (Kaggle, 2017)	(Khan *et al.*, 2017)
Transformation		(Collins & Woodcock, 1996) (Müller *et al.*, 2016)
Advanced model (spectral mixture)	(Shimabukuro *et al.*, 1998)	
Other methods	(Achard *et al.*, 2002)	(Achard *et al.*, 2002)

Deforestation detection using earth observation data implies mainly two complementary tasks : (a) detecting in a given image what is forest and what is not (i.e. image annotation) (b) detecting changes between images. While deep learning model have shown promising results in these two tasks, remote sensing suffers from a limited availability of annotated data for training such models. Moreover, it is likely that using different sources of evidence can help in the accuracy of the detection. The next section is related to these issues.

4 Research questions

From our readings and from related work, we elaborate what we think are the main research questions at this stage in the domain of automatic detection of deforestation using earth observation imagery.

4.1 Data fusion : how to combine various sources of evidence ?

A current trend in remote sensing is data fusion (Zhang, 2010). The main idea is to combine data from multiple sources in order to produce better quality results taking advantage of the information each source carries. While data fusion is not new in remote sensing (Pohl & Van Genderen, 1998), the most recent research in the field focuses on high-level fusion in place of pixel-level fusion. (Joshi *et al.*, 2016) reviewed 112 studies for various types of applications on fusing optical and radar data and concluded that the main methods used are pre-classification fusion followed by pixel-level inputs in traditional classification algorithms.

Various types of data can be fused. For example, (Reiche *et al.*, 2015) presented a fusion approach (MulTiFuse) that exploits the full observation density of optical and Synthetic Aperture Radar (SAR) time series to detect deforestation in the tropics. (Schmidt *et al.*, 2015) used coarse spatial resolution MODIS data combined with finer spatial resolution Landsat data to map forest and agricultural elements of an area in central southeast Queensland, Australia. In (Reiche *et al.*, 2018), the authors show that spatial and temporal accuracies of the multi-sensor approach were higher than the single-sensor results for near real-time deforestation detection in tropical dry forests (the authors combined Sentinel-1 C-band SAR time series with ALSO-2 PALSAR-2 L-band SAR, and Landsat-7/ETM+ and 8/OLI).

Current data fusion techniques focus on fusing data at various levels of resolution, various number of layers for multi-spectral imagery but a few consider other sources of evidence such as statistics, scientific publications, etc. While satellite data is one source of evidence for deforestation detection, it may not be enough to explain the cause of deforestation and monitor its evolution as well as other related geo-phenomena. To get a step further we would like to combine earth observation imagery with other sources of evidence such as statistics related to wood exportation, type of manufacturing goods implying wood or derivatives, level of urbanisation and population growth. Another source of evidence we would like to use in cross analyses are reports and scientific papers. We believe that scientific papers on sustainable development contain relevant information about deforestation and the related geo-phenomena and that this information can be extracted and combined with data from earth observation imagery for analysis purposes. The main issue to fuse these sources is their high level of heterogeneity which is an open question.

4.2 Transfer learning with CNN : how effective is it ?

Transfer learning is the process of learning features by training a network on a large data set and then transferring these features to a different dataset. This approach has also been used in remote sensing and showed promising results by (Hu *et al.*, 2015) and (Salberg, 2015). The ability for learned features to be transferred across domains accounts for the success of transfer learning in deep learning and makes it a promising approach to cope with the problem of limited training data. Transfer learning is also used to overcome the problem of the lack of annotated data for training.

The annotation of remote sensing images is done either by human annotators or automatically by a computer program. In both cases, ground truths are the reference against which these annotations can be validated. However, going on the ground to collected theses ground truths may be costly and impractical. Consequently, the amount of labelled data with validated labels remains limited compared to the large amount of earth observation data that is available.

To overcome the problem of limited labelled data available for training, several approaches using transfer learning have been proposed in the remote sensing field. The effectiveness of deep learning approaches in remote sensing considering the insufficient training data is still an open question. So far, limited work has been done on change detection in forests with deep learning and CNN in particular. We aim to propose a new approach to detect and map deforestation in tropical forests using transfer learning.

Our goal is to propose a model to track the evolution of deforestation over time with time series analysis. By using transfer learning, we aim to find the best suited image dataset for this task by testing with various types of images.

We aim to provide a general model that can be reused in different countries and areas. This will require training our model on a large and diverse dataset. There is a massive amount of spatial data available though mostly unlabelled. We therefore aim to propose an unsupervised approach to annotate these images and validate the results with very high resolution images and ground truth data when available.

4.3 How to adapt CNN architectures to multi-spectral images ?

Satellite images are multi-spectral and pose the challenge of high dimensionality for machine learning. We will propose a CNN model that is adapted to this particular type of data. We aim to propose a dimension reduction approach that is optimal for our problem of detecting and tracking deforestation in tropical forests.

4.4 How to get annotated satellite images for our areas of interest ?

Due to the fact that annotated satellite data for most tropical forests is scarce, another contribution we aim to make is provide a new data set of labelled earth observation data for one of our areas of interest. To achieve this, we will use very high resolution images and ground truth data from local organizations if available or from field missions.

5 Conclusion

This paper presented the state of the art related to automatic detection of deforestation using earth observation images from remote sensors. We present the methods that have been developed in the two main steps of automatic detection of deforestation : image annotation which includes image segmentation and classification, and change detection. For both steps, we chose to present first the methods that were developed before deep neural networks then the methods which use deep learning. As in many other fields, neural networks are now very commonly used in remote sensing. Finally, we mentioned the research questions we would like to tackle in our future work and for two of them, namely data fusion and transfer learning, we detailed related work.

Acknowledgements

The work of Nathalie Neptune is supported by the Schlumberger Foundation through the Faculty for the Future fellowship. We would also like to thank the reviewers for their valuable comments. This work also benefits support from FabSpace 2.0 project ; FabSpace 2.0 received funding from the European Union's Horizon 2020 Research and Innovation programme under the Grant Agreement number 693210.

6 Appendix - Convolutional neural networks

Convolutional Neural Networks are one of the deep neural network classes that give the best current results in most computer vision problems such as classification and object recognition (Krizhevsky *et al.*, 2012), (Witten *et al.*, 2016). Basically, CNN is a straightforward Artificial Neural Network (ANN) in which the architecture consists of several layers connected together in a multi-tiered structure (LeCun *et al.*, 1998). There are three main types of layers : convolution layer,activation layer, pooling layer and fully connected layer (Krizhevsky *et al.*, 2012).

Before going into the basic components of CNN, we would like to look at an example describing how a feedforward ANN processes an input information. From that, we can draw on the effects of CNN components.

Suppose that we have a 200 x 200 image processed by a fully connected ANN. Each neuron needs 40,000 parameters to be trained which is costly. To reduce the number of parameters it is necessary to reduce the number of connections between layers ; this is the objective of the CNN convolution component. The idea is that each neural only needs to be connected to a local area of the image instead of the entire image. This feature enlarges CNN learning capacity (Krizhevsky *et al.*, 2012) which is the basic need to deal with large dataset problem such as earth observation images.

According to the literature of the domain (see for example (Krizhevsky *et al.*, 2012) there are 4 types of layers in a CNN which are defined as follows :

Convolution layer
The convolution layer plays the main role in the architecture of a CNN. Each neuron of the layer is formed by doing a convolution between a kernel and an image, applying convolution to the images aims at extracting important features such as edges, direction, color (Witten *et al.*, 2016).

Activation layer
This layer is usually placed right after the convolutional layer. This layer uses an activation function such as sigmoid, tanh, softplus, rectified linear unit (ReLU). However, ReLU ($f(x) = \max(0, x)$) is used most recently. The function converts all negative values in the result obtained from the convolutional layer to the value 0. The meaning of this setting is to make the model non-linear (Witten *et al.*, 2016). Using an activation layer also increases the learning rate (Krizhevsky *et al.*, 2012).

Pooling Layer
The goal of this layer is to reduce the matrix size but still highlight the features that are present in the input matrix. Max pooling is often used (Guo *et al.*, 2016). In terms of meaning, Max-Pooling determines where the strongest signal is when applying a filter. It is done by taking the maximum value of the neurons within the pooling region (Witten *et al.*, 2016).

Fully Connected Layer
This layer is similar to the feedforward ANN : all the nodes of the current layer are fully linked to the nodes from the next layer. After the image is processed by the previous layers, the image data will no longer be too large compared to the ANN model. This layer is placed at the end part of the CNN (Witten *et al.*, 2016).

A convolution neural network is formed by putting the above layers together. The model always starts with the convolutional layer. The activation layer usually follows right after the convolutional layer or even merges both layers into a layer. The next layer can be convolutional or pooling. This pattern can be repeated depending on the architecture. The output of these layers then may feeded to fully connected layers. The final layer of CNN usually uses the softmax function ($\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}$) which forces the output of the network represent a probability distribution across discrete alternatives (Krizhevsky *et al.*, 2012). Figure 2 represents a general CNN which uses the ReLU function as an activation function.

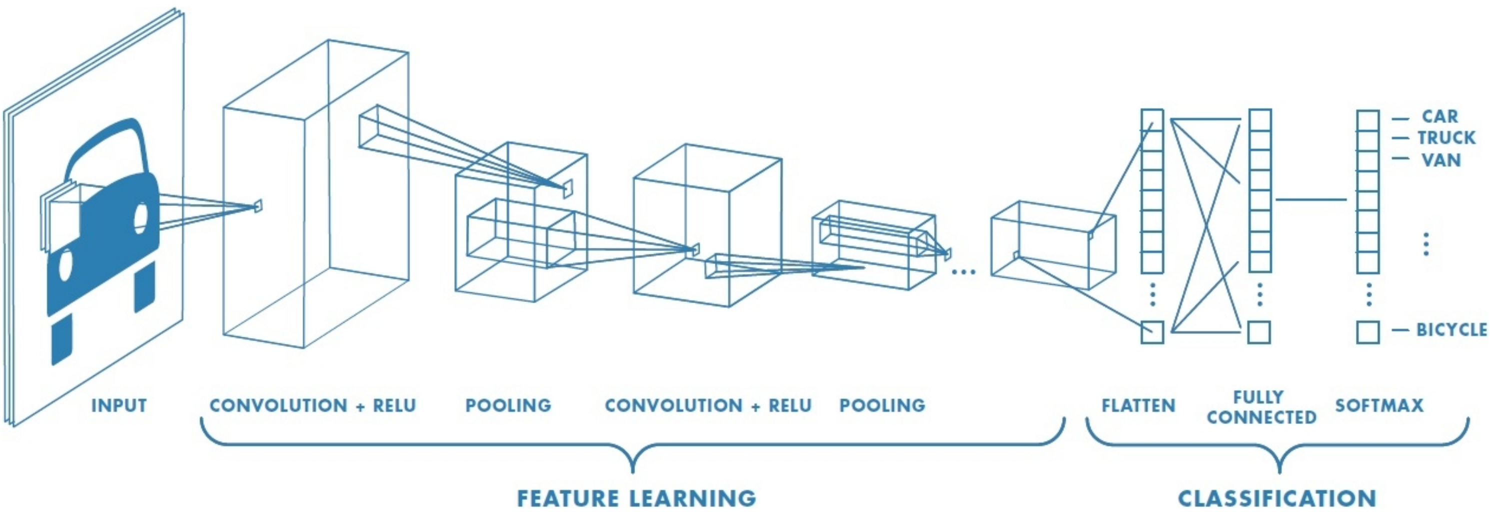

FIGURE 2: A general CNN architecture which is merges the convolution layer and activation layer into one (Mathworks, 2016)

Références

ACHARD F., EVA H. D., STIBIG H.-J., MAYAUX P., GALLEGO J., RICHARDS T. & MALINGREAU J.-P. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science*, **297**(5583), 999–1002.

BABENKO A., SLESAREV A., CHIGORIN A. & LEMPITSKY V. (2014). Neural codes for image retrieval. In *European conference on computer vision*, p. 584–599 : Springer.

BARRETO T. L., ROSA R. A., WIMMER C., MOREIRA J. R., BINS L. S., CAPPABIANCO F. A. M. & ALMEIDA J. (2016). Classification of Detected Changes From Multitemporal High-Res Xband SAR Images : Intensity and Texture Descriptors From SuperPixels. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **9**(12), 5436–5448.

COLLINS J. B. & WOODCOCK C. E. (1996). An assessment of several linear change detection techniques for mapping forest mortality using multitemporal landsat tm data. *Remote sensing of environment*, **56**(1), 66–77.

GEOGRAPHIC N. (2017). Climate 101 : Deforestation. https://www.nationalgeographic.com/environment/global-warming/deforestation/.

GUO Y., LIU Y., OERLEMANS A., LAO S., WU S. & LEW M. S. (2016). Deep learning for visual understanding : A review. *Neurocomputing*, **187**, 27–48.

HE K., SUN J. & TANG X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**(12), 2341–2353.

HIRSCHMUGL M., GALLAUN H., DEES M., DATTA P., DEUTSCHER J., KOUTSIAS N. & SCHARDT M. (2017). Methods for mapping forest disturbance and degradation from optical earth observation data : a review. *Current Forestry Reports*, **3**(1), 32–45.

HU F., XIA G.-S., HU J. & ZHANG L. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, **7**(11), 14680–14707.

IONESCU B., MÜLLER H., VILLEGAS M., ARENAS H., BOATO G., DANG-NGUYEN D.-T., CID Y. D., EICKHOFF C., DE HERRERA A. G. S., GURRIN C., ISLAM B., KOVALEV V., LIAUCHUK V., MOTHE J., PIRAS L., RIEGLER M. & SCHWALL I. (2017). Overview of imageclef 2017 : Information extraction from images. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, p. 315–337 : Springer.

JOSHI N., BAUMANN M., EHAMMER A., FENSHOLT R., GROGAN K., HOSTERT P., JEPSEN M. R., KUEMMERLE T., MEYFROIDT P., MITCHARD E. T. A., REICHE J., RYAN C. M. & WASKE B. (2016). A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, **8**(1).

KAGGLE I. (2017). Planet : Understanding the amazon from space | kaggle. https://www.kaggle.com/c/planet-understanding-the-amazon-from-space. Accessed : 2018-01-25.

KHAN S. H., HE X., PORIKLI F. & BENNAMOUN M. (2017). Forest change detection in incomplete satellite images with deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **55**(9), 5407–5423.

KRIZHEVSKY A., SUTSKEVER I. & HINTON G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, p. 1097–1105.

KUSSUL N., LAVRENIUK M., SKAKUN S. & SHELESTOV A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, **14**(5), 778–782.

LAGRANGE A., SAUX B. L., BEAUPÈRE A., BOULCH A., CHAN-HON-TONG A., HERBIN S., RANDRIANARIVO H. & FERECATU M. (2015). Benchmarking classification of earth-observation data : From learning explicit features to convolutional networks. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 4173–4176.

LECUN Y., BOTTOU L., BENGIO Y. & HAFFNER P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**(11), 2278–2324.

LU D., MAUSEL P., BRONDIZIO E. & MORAN E. (2004). Change detection techniques. *International journal of remote sensing*, **25**(12), 2365–2401.

MATHWORKS (2016). Convolutional neural network. https://fr.mathworks.com/discovery/convolutional-neural-network.html.

MERTENS B. & LAMBIN E. F. (1997). Spatial modelling of deforestation in southern cameroon : spatial disaggregation of diverse deforestation processes. *Applied Geography*, **17**(2), 143–162.

MILLER L. D. L. D., NUALCHAWEE K., TOM C. C. & CENTER G. S. F. (1978). *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. "May 1978.".

MOTHE J., NY HOAVY N. & RANDRIANARIVONY M. I. (2017). IRIT & MISA at Image CLEF 2017 - Multi label classification. In *International Conference of the CLEF Association, CLEF 2017 Labs Working Notes*, volume 1866 of *ISSN 1613-0073* : CEUR Workshop Proceedings.

MÜLLER H., GRIFFITHS P. & HOSTERT P. (2016). 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*, **44**, 61–69.

O'CONNOR B., SECADES C., PENNER J., SONNENSCHEIN R., SKIDMORE A., BURGESS N. D. & HUTTON J. M. (2015). Earth observation as a tool for tracking progress towards the aichi biodiversity targets. *Remote sensing in ecology and conservation*, **1**(1), 19–28.

OLOFSSON P., HOLDEN C. E., BULLOCK E. L. & WOODCOCK C. E. (2016). Time series analysis of satellite data reveals continuous deforestation of new england since the 1980s. *Environmental Research Letters*, **11**(6), 064002.

POHL C. & VAN GENDEREN J. L. (1998). Review article multisensor image fusion in remote sensing : concepts, methods and applications. *International journal of remote sensing*, **19**(5), 823–854.

REICHE J., HAMUNYELA E., VERBESSELT J., HOEKMAN D. & HEROLD M. (2018). Improving near-real time deforestation monitoring in tropical dry forests by combining dense sentinel-1 time series with landsat and alos-2 palsar-2. *Remote Sensing of Environment*, **204**, 147 – 161.

REICHE J., VERBESSELT J., HOEKMAN D. & HEROLD M. (2015). Fusing landsat and sar time series to detect deforestation in the tropics. *Remote Sensing of Environment*, **156**, 276–293.

SALBERG A. B. (2015). Detection of seals in remote sensing images using features extracted from deep convolutional neural networks. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 1893–1896.

SCHMIDT M., LUCAS R., BUNTING P., VERBESSELT J. & ARMSTON J. (2015). Multi-resolution time series imagery for forest disturbance and regrowth monitoring in queensland, australia. *Remote Sensing of Environment*, **158**, 156–168.

SCOTT K. (2017). Forest recognition : Planet launches kaggle competition. `https://www.planet.com/pulse/forest-recognition-planet-launches-kaggle-competition/`. Accessed : 2018-01-25.

SHIMABUKURO Y. E., BATISTA G., MELLO E., MOREIRA J. & DUARTE V. (1998). Using shade fraction image segmentation to evaluate deforestation in landsat thematic mapper images of the amazon region. *International Journal of Remote Sensing*, **19**(3), 535–541.

SINGH A. (1989). Review article digital change detection techniques using remotely-sensed data. *International journal of remote sensing*, **10**(6), 989–1003.

TARIQ M. & AZIZ R. (2015). An overview of deforestation causes and its environmental hazards in khyber pukhtunkhwa. *Journal of Natural Sciences Research*, **5**(1).

WITTEN I. H., FRANK E., HALL M. A. & PAL C. J. (2016). *Data Mining : Practical machine learning tools and techniques*. Morgan Kaufmann.

ZHANG C., PAN X., LI H., GARDINER A., SARGENT I., HARE J. & ATKINSON P. M. (2017). A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*.

ZHANG J. (2010). Multi-source remote sensing data fusion : status and trends. *International Journal of Image and Data Fusion*, **1**(1), 5–24.