



Mabel Ximena Ortega Adarme

**A comparison of Deep Learning techniques for
deforestation detection in the Brazilian
Amazon and Cerrado biomes from remote
sensing imagery**

Dissertação de Mestrado

Thesis presented to the Programa de Pós-graduação em of
PUC-Rio in partial fulfillment of the requirements for the degree
of Mestre em .

Advisor : Prof. Raul Queiroz Feitosa
Co-advisor: Dra. Alessandra Rodrigues Gomes

Rio de Janeiro
July 2019



Mabel Ximena Ortega Adarme

**A comparison of Deep Learning techniques for
deforestation detection in the Brazilian
Amazon and Cerrado biomes from remote
sensing imagery**

Thesis presented to the Programa de Pós-graduação em of
PUC-Rio in partial fulfillment of the requirements for the degree
of Mestre em . Approved by the Examination Committee.

Prof. Raul Queiroz Feitosa

Advisor

Departamento de Engenharia Elétrica – PUC-Rio

Dra. Alessandra Rodrigues Gomes

Co-advisor

Instituto Nacional de Pesquisas Espaciais – INPE

Dr. Claudio Aparecido De Almeida

Instituto Nacional de Pesquisas Espaciais – INPE

Prof. Marcelo Gattass

Tecgraf – PUC-Rio

Rio de Janeiro, July the 31st, 2019

All rights reserved.

Mabel Ximena Ortega Adarme

The author received her bachelor's degree in Electronic Engineering at the Universidad de Nariño, Colombia in 2017. Since then, she has worked in the field of Digital Image Processing, Remote Sensing and Machine Learning.

Bibliographic data

Ortega Adarme, Mabel Ximena

A comparison of Deep Learning techniques for deforestation detection in the Brazilian Amazon and Cerrado biomes from remote sensing imagery / Mabel Ximena Ortega Adarme; advisor: Raul Queiroz Feitosa; co-advisor: Alessandra Rodrigues Gomes. – Rio de Janeiro: PUC-Rio, Departamento de Engenharia Elétrica, 2019.

v., 70 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Detecção de Desmatamento;. 3. Biomas Brasileiros;. 4. Sensoriamento Remoto;. 5. Aprendizado Profundo;. 6. Redes Neurais Convolucionais..

I. Queiroz Feitosa, Raul.. II. Rodrigues Gomes, Alessandra. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 66898

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Prof. Raul Queiroz Feitosa, for the encouragement, his advice, stimulating talks and generous support throughout the development of this thesis.

Besides my advisor, I am truly grateful to Dr. Patrick Happ, for his advice, continuous support, motivation, and his valuable guidance throughout the development of this thesis.

In addition, I would like to express my gratitude to my parents, Melba and Ernesto, my sister, Marcela and, my brother Diego, for their support and unconditional love.

I want to thank all my colleagues from the Computer Vision Lab in Pontifical Catholic University of Rio de Janeiro - PUC-Rio for their companionship and valuable scientific discussions.

Finally, I also gratefully acknowledge the financial support of CNPq.

Abstract

Ortega Adarme, Mabel Ximena; Queiroz Feitosa, Raul. (Advisor); Rodrigues Gomes, Alessandra (Co-Advisor). **A comparison of Deep Learning techniques for deforestation detection in the Brazilian Amazon and Cerrado biomes from remote sensing imagery.** Rio de Janeiro, 2019. 70p. Dissertação de mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Deforestation is one of the main causes of biodiversity reduction, climate change, among other destructive phenomena. Thus, early detection of deforestation processes is of paramount importance. Techniques based on satellite images are one of the most attractive options for this application. However, many works developed include some manual operations or dependency on a threshold to identify regions that suffer deforestation or not. Motivated by this scenario, the present dissertation presents an evaluation of methods for automatic deforestation detection, specifically Early Fusion (EF) Convolutional Network, Siamese Convolutional Network (SN), Convolutional Support Vector Machine (CSVM) and Support Vector Machine (SVM), taken as the baseline. These methods were evaluated in regions of Brazilian Amazon and Cerrado Biomes. Two Landsat 8 images acquired at different dates were used in the experiments, and the impact of training set size was also analyzed. The results demonstrated that Deep Learning-based approaches clearly outperformed the SVM baseline in our approaches, both in terms of F1-score and Overall Accuracy, with the superiority of SN and EF over CSVM and SVM. In the same way, a reduction of the *salt-and-pepper* effect in the generated probabilistic change maps was noticed due, mainly, to the increase of samples in the training sets. Finally, an analysis was carried out to assess how the methods can reduce the time invested in the visual inspection of deforested areas.

Keywords

Deforestation Detection; Brazilian Biomes; Remote Sensing; Deep learning; Convolutional Neural Network.

Resumo

Ortega Adarme, Mabel Ximena; Queiroz Feitosa, Raul.; Rodrigues Gomes, Alessandra. **Comparação de técnicas de Deep Learning para detecção de desmatamento em biomas da Amazônia e Cerrado brasileiros a partir de imagens de sensoriamento remoto.** Rio de Janeiro, 2019. 70p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O desmatamento é uma das principais causas de redução da biodiversidade, mudança climática e outros fenômenos destrutivos. Assim, a detecção antecipada de desmatamento é de suma importância. Técnicas baseadas em imagens de satélite são uma das opções mais interessantes para esta aplicação. No entanto, muitos trabalhos desenvolvidos incluem algumas operações manuais ou dependência de um limiar para identificar regiões que sofrem desmatamento ou não. Motivado por este cenário, a presente dissertação apresenta uma avaliação de métodos para detecção automática de desmatamento, especificamente de Early Fusion (EF) Convolutional Network, Siamese Convolutional Network (SN), Convolutional Support Vector Machine (CSVM) e Support Vector Machine (SVM), o último tomado como baseline. Todos os métodos foram avaliados em regiões dos biomas brasileiros Amazônia e Cerrado. Duas imagens Landsat 8 adquiridas em diferentes datas foram utilizadas nos experimentos, e também o impacto do tamanho do conjunto de treinamento foi analisado. Os resultados demonstraram que as abordagens baseadas no Deep Learning superaram claramente o baseline SVM em termos de pontuação F1-score e Overall Accuracy, com uma superioridade de SN e EF sobre CSVM e SVM. Da mesma forma, uma redução do efeito *sal e pimenta* nos mapas de mudança gerados foi notada devido, principalmente ao aumento de amostras nos conjuntos de treinamento. Finalmente, realizou-se uma análise visando avaliar como os métodos podem reduzir o esforço humano na inspeção visual das áreas desmatadas.

Palavras-chave

Detecção de Desmatamento; Biomas Brasileiros; Sensoriamento Remoto; Aprendizado Profundo; Redes Neurais Convolucionais.

Table of contents

1	INTRODUCTION	16
1.1	Objectives	18
1.1.1	General Objective	18
1.1.2	Specific Objectives	18
1.2	Contributions	18
1.3	Organization of the remaining parts of this thesis	19
2	RELATED WORKS	20
3	FUNDAMENTALS	24
3.1	Remote Sensing (RS)	24
3.2	Convolutional Neural Networks (CNNs)	25
3.3	Siamese Networks (SNs)	28
3.4	Convolutional SVM (CSVM)	29
4	METHODS	33
4.1	Early Fusion (EF)	33
4.2	Siamese Network (SN)	33
4.3	Convolutional SVM (CSVM)	34
5	EXPERIMENTAL ANALYSIS	37
5.1	Datasets	37
5.1.1	Amazon Biome	37
5.1.2	Cerrado Biome	39
5.2	Experimental Setup	39
5.3	Accuracy Assessment	45
5.4	Results	46
5.4.0.1	Amazon Biome	47

5.4.0.2 Cerrado Biome	54
6 CONCLUSIONS	62
References	65

List of figures

Figure 1 - Forest formation in Brazil from 1985 to 2017 . Mapbiomas (Available at: http://www.mapbiomas.org/stats).	16
Figure 2 - Passive and Active sensors. Passive sensors does not require any external power source to produce output signal while active sensor requires an external power source to operate.	25
Figure 3 - A <i>CNN</i> basic architecture with two convolutional layers. Adapted from [45].	26
Figure 4 - Illustration of the most common non-linear activation functions.	27
Figure 5 - Example of <i>max-pooling</i> applying a 2×2 filter with stride equal to 2. Each element of the output matrix is the max of the region in the original input.	27
Figure 6 - Dropout in a Neural Network. (a) A standard neural net with two hidden layers. (b) An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped [47].	28
Figure 7 - Architecture of the Siamese Network. The two CNNs share the same structure and weights.	29
Figure 8 - SVM convolutional layer. The procedure starts with the construction of a training set, followed by the training of SVM filter banks and ends with the generation of the convolutional feature maps.	30
Figure 9 - Generation of training set for the first convolutional layer of CSVM.	30
Figure 10 - Feature maps obtained from the first convolutional layer of CSVM.	31
Figure 11 - Creation of the training set for the final classifier.	31

- Figure 12 - EF method. Images at different dates (T1 and T2) are concatenated to produce an image pair; then, patches are extracted and fed to the CNN model. 34
- Figure 13 - Siamese network. Patches of each image (T1 and T2) are extracted and fed to the CNN model independently. The two branches in the network share exactly the same architecture and parameters values. 34
- Figure 14 - Procedure to train the SVMs filter bank. Mini-patches are extracted from input patches and they are vectorized to compose the training set of SVMs. 35
- Figure 15 - Generation of feature maps. 36
- Figure 16 - Region of the Brazilian Amazon, located in the Pará State, Brazil. *NIR – G – B* composition of the study area at dates T1 (a) and T2 (b). c) Deforestation reference from 2016 to 2017. 38
- Figure 17 - Region of the Brazilian Cerrado biome, located in the Maranhão State, Brazil. *NIRGB* composition of the study area at dates T1 (a) and T1 (b). c) Deforestation reference from 2017 to 2018. 40
- Figure 18 - Distribution of the Amazon database. The region was divided into fifteen tiles. Four tiles was used for training (1, 7, 9, 13). Two for validation (5, 12). The rest of the them were used for testing (2, 3, 4, 6, 8, 10, 11, 14, 15). 41
- Figure 19 - Distribution of the Cerrado database. The region was divided into fifteen tiles. Four tiles was used for training (1, 5, 12, 13). Two for validation (6, 10). The rest of the them were used for testing (2, 3, 4, 7, 8, 9, 11, 14, 15). 41
- Figure 20 - Architecture of EF and SN. 43
- Figure 21 - Architecture of CSVM. 44
- Figure 22 - *F1-score* of Amazon Biome obtained from SVM, EF, SN, and CSVM using one, two, three, and four tiles for training. 47

Figure 23 - Overall Accuracy of Amazon Biome obtained from SVM, EF, SN, and CSVM using one, two, three, and four tiles for training.	48
Figure 24 - Predicted maps of the second tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.	49
Figure 25 - Predicted maps of the sixth tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.	50
Figure 26 - Predicted maps of fourteenth tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.	51
Figure 27 - Recall vs Area classified as deforested for Amazon biome using one tile for training.	52
Figure 28 - Recall vs Area classified as deforested for Amazon biome using two tiles for training.	53
Figure 29 - Recall vs Area classified as deforested for Amazon biome using three tiles for training.	53
Figure 30 - Recall vs Area classified as deforested for Amazon biome using four tiles for training.	54
Figure 31 - <i>F1-score</i> of Cerrado Biome obtained from SVM, EF, SN, and CSVM using one, two, three, and four tiles for training.	55
Figure 32 - Overall Accuracy of Cerrado Biome obtained from SVM, EF, SN, and CSVM using one, two, three, and four tiles for training.	55
Figure 33 - Predicted maps of the second tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.	57
Figure 34 - Predicted maps of the eighth tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.	58
Figure 35 - Predicted maps of the eleventh tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.	59
Figure 36 - Recall vs Area classified as deforested for Cerrado biome using one tile for training.	60

Figure 37 - Recall vs Area classified as deforested for Cerrado biome using two tiles for training.	60
Figure 38 - Recall vs Area classified as deforested for Cerrado biome using three tiles for training.	61
Figure 39 - Recall vs Area classified as deforested for Cerrado biome using four tiles for training.	61

List of tables

Table 1 - Number of samples in the training, validation and test sets for Amazon database.	42
Table 2 - Number of samples in the training, validation and test sets for Cerrado database.	42
Table 3 - Training tiles used for the Amazon database.	44
Table 4 - Training tiles used for the Cerrado database.	45
Table 5 - Confusion matrix for a binary classification problem.	45

List of Abbreviations

ANN	Artificial Neural Networks
ASGD	Average Stochastic Gradient Descen
BLA	Brazilian Legal Amazon
BN	Batch Normalization
CM	Confusion Matrix
CNN	Convolutional Neural Networks
CSVM	Convolutional Support Vector Machine
CVA	Change Vector Analysis
DETER	Near Real-Time Deforestation Detection Project
DL	Deep Learning
DNN	Deep Neural Networks
EF	Early Fusion
EM	Electromagnetic
EO	Earth Observation
FC	Fully Connected
FCN	Fully Convolutional Networks
INPE	National Institute for Space Research
k -NN	K-Nearest Neighbor
KT	Tasselled Cap
LSMM	Linear Spectral Mixture Model
LSTM	Long Short-Term Memory
MODIS	Moderate-Resolution Imaging Spectroradiometer
MRF	Markov Random Fields
MSI	MultiSpectral Instrument
NDVI	Normalized Difference Vegetation Index
NIR	Near InfraRed
NN	Neural Network
OA	Overall Accuracy
OBIA	Object-Based Image Analysis
PRODES	Amazon Deforestation Monitoring Project
RBF	Radial Basis Function
RF	Random Forest

RNN	Recurrent Neural Networks
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SN	Siamese Network
SVM	Support Vector Machine
SWIR	Short-Wave InfraRed
T1	Time 1
T2	Time 2
UAV	Unmanned Aerial Vehicle
USGS	United States Geological Survey

1

INTRODUCTION

Deforestation is one of the largest sources of anthropogenic CO₂ emissions. It is a wide-reaching problem including reduction of carbon storage, greenhouse gas emissions, and other environmental issues such as biodiversity losses and climate change [1]. Currently, one of the worldwide highest deforestation rates occurs in South America [2], where the major statistics of tree losses is concentrated in Brazil. This country comprises the biggest portion of tropical rainforest with 60% of its total territory [3]. Indeed, the last three decades it has lost about 660'000 square kilometers (see Figure 1).

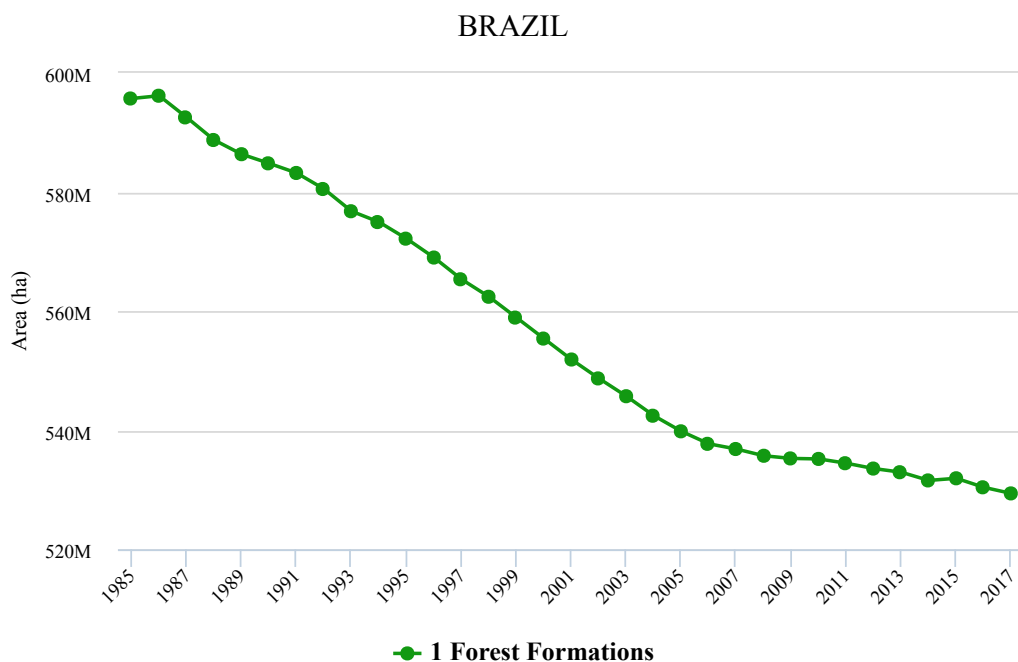


Figure 1: Forest formation in Brazil from 1985 to 2017 . Mapbiomas (Available at: <http://www.mapbiomas.org/stats>).

Brazil comprises six terrestrial biomes: Amazon, Cerrado, Atlantic Forest, Caatinga, Pampa, and Pantanal with an area of 49.3%, 23.9%, 13%, 9.9%, 2%, and 1.7%, respectively [3]. In this sense, Amazon and Cerrado Brazilian biomes cover an area of about 73%. Both biomes together comprise an area around 6.2 million square kilometers of the Brazilian territory. These

biomes accommodate large biodiversity and they are home to a large number of species, including endemic and endangered flora and fauna [4]. Therefore, they provide essential resources for the maintenance of our planet and its preservation is of paramount importance.

For many years the Amazon biome has faced several threats as a result of unsustainable economic development. These threats are primarily caused by the extension of agricultural activities at industrial scale (e.g., soybeans, cattle), slash-and-burn land grabbing by underprivileged rural communities, forest fires, illegal gold mining and logging, expansion of informal settlements, and infrastructure construction such as roads and train tracks [5–7]. According to the National Institute for Space Research (INPE) [8], deforestation accelerated significantly during the 1990s and early 2000s in the Amazon biome. Likewise, the World Wildlife Fund (WWF) [9] estimates that more than a quarter of the biome will be without trees by 2030 if the current rate of deforestation continues.

Following the Amazon biome, Cerrado is the Brazilian biome that suffered the most changes with human occupation [10]. Despite its biological importance, it faces intensive land use pressure, losing over 40% of natural vegetation due to agricultural expansion [11, 12]. Moreover, the Cerrado has been one of the least studied biomes in Brazil. It has not received great attention compared to other Brazilian biomes, such as the Amazon and the Atlantic Forest [4, 11]. In fact, information about land use and land cover change of this biome is still limited. In this sense, the monitoring of this biome is necessary to track natural disasters and human activities, as well as to enforce public policies to avoid illegal activities in the region, achieving an ecological balance and contributing to the mitigation of climate change [13]. In this scenario, Remote Sensing (RS) has proven to be a cost-effective solution to monitor these regions.

Traditionally, change detection techniques based on image algebra, such as image differencing, image rationing and Change Vector Analysis (CVA) [14], have been widely used to detect changes in multi-temporal images. However, they strongly depend on manually selected thresholds to define what is considered a change [15]. More recently, Deep Learning (DL) techniques have become state-of-the-art in many application fields, including RS. Through the usage of Deep Neural Networks (DNNs), it is possible to learn multiple levels of data representation, which usually correspond to more informative features and often allow for better results than what can be achieved by using domain-specific handcrafted features.

In this context, this work evaluates and compares Deep Learning based

techniques for deforestation detection in the Brazilian Amazon and Cerrado biomes. The aim is to reduce the human effort involved in monitoring programs such as the Amazon Deforestation Monitoring Project (Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite-PRODES) and the Cerrado Monitoring Project. In addition, this work aims at contributing to improve accuracy and reduce the subjectivity inherent to human photointerpretation, as well as reduce the time needed to generate results.

1.1 Objectives

1.1.1 General Objective

The general objective of this work is to test state-of-the-art Deep Learning techniques for deforestation detection in Brazil from optical images acquired at two different dates.

1.1.2 Specific Objectives

The specific objectives of this work are the following:

1. Compare Siamese Networks, Early Fusion and Convolutional SVM for automatic deforestation detection;
2. Assess the aforementioned methods in datasets of two regions with different deforestation patterns: the Amazon and Cerrado biomes;
3. Evaluate how the amount of training samples impacts the performance of each method;
4. Analyze how semi-automatic approaches could be designed based on these methods to reduce human intervention with minimal accuracy loss.

1.2 Contributions

The main contributions of this work are:

1. An evaluation of state-of the art Deep Learning techniques for automatic deforestation detection in Brazilian Amazon and Cerrado biomes;
2. An assessment of these methods' accuracy under scarce training samples;

3. An estimation for each method of the relation: area assigned as deforestation vs. area of true deforestation.

1.3

Organization of the remaining parts of this thesis

Chapter 2 describes the related work available in the literature for change detection and deforestation detection using traditional approaches as well as models based on Deep Learning.

Chapter 3 presents the fundamental concepts and theory underlying the methods used in this work.

Chapter 4 introduces and explains the deep learning methods used for deforestation detection.

Chapter 5 presents the study areas used in this work, the experimental setup followed in the experiments and the results obtained from each method.

Chapter 6 summarizes the conclusions derived from the performed experiments and provides directions for further development of the proposed method.

This chapter presents an overview of different works related to change detection techniques for remote sensing applications. Most of them focused on deforestation detection in the Brazilian biomes as well as other areas commonly studied in the literature.

With the development of RS, the dynamic earth observation (EO) has led to a great deal of available, detailed, accurate and up-to-date change information for use in learning about and monitoring our planet [16, 17]. Numerous change detection techniques have been proposed thus far. Some of the traditional unsupervised methods are based on image algebra such as Image Differencing [18], Image Ratioing [19], Regression Analysis [20] and Change Vector Analysis (CVA) [21]. In addition, techniques based on transformations such as Principal Component Analysis (PCA) [22] and Tasselled cap (KT) [23] have been also used for this purpose. However, these methods require the selection of a proper threshold to identify the changed regions. Besides, the features adopted by these conventional algorithms are hand-crafted, which may lead to poor image representations [24].

For the supervised methods, Support Vector Machine (SVM) is one of the most popular algorithms used in satellite image classification [25, 26] due to its good performance and robustness in scenarios where there are few available labeled samples. Nevertheless, a proper setup of its hyper parameters, including the appropriate kernel function, is essential to achieve good generalization. In general, this process involves to consider several kernel types and compare them via cross-validation or other methods [27]. Similarly, tree based [28] classifiers and methods based on Artificial Neural Networks (ANN) are widely used for image classification [29]. However, one of the drawbacks of decision trees is its high computational cost during the tree building because the training samples have to traverse repeatedly each node of the tree [30]. For ANN, is necessary to make modeling assumptions to properly select the number of hidden layers and number of their units. Indeed, there are not a general rule to select the optimal setup of these parameters.

On the other hand, the Brazilian National Institute for Space Research (Instituto Nacional de Pesquisas Espaciais-INPE) has developed and maintains

a number of projects to provide surveillance reports about deforestation over the Brazilian Legal Amazon (BLA). The best known-action is the Amazon Deforestation Monitoring Project (PRODES) [31], which supervises the deforestation in areas with native vegetation of BLA since 1988. They estimate deforested areas using the Linear Spectral Mixture Model (LSMM), which uses the original Landsat TM bands 3, 4 and 5 to generate synthetic images of vegetation, soil, and shade. Next, an image segmentation based on the region growing algorithm is performed followed by a classification stage using the clustering algorithm ISOSEG, but it is only applied to the segmented shade image. After these procedures, it is necessary to identify errors in the classified image manually. Thus, this task is highly dependent on experienced interpreters.

The near real-time deforestation detection project (DETER) [32], was developed to support land use policies in BLA and controls the illegal deforestation, forest degradation and wood exploration. This process is done by LSMM and visual interpretation based on five main elements (color, tonality, texture, shape and context). Finally, the Land Use and Land Cover Mapping of Amazon Deforested Areas (TerraClass) project is responsible for qualifying deforestation in BLA and investigating the possible causes of logging. De Almeida C and et al. [33], showed that in the Amazonian post-deforestation landscape pasture is dominant, followed by clear cut, used for silviculture, agriculture or pasture. Eventually, part of them regenerates and turns into forest again.

Similar to BLA, the Cerrado biome has also been monitored by INPE. Brito et al. [34] present a methodology used for mapping deforestation in the Cerrado biome since 2000. The process is entirely manual, and is carried out by visual interpretation by taking into account five main elements: color, tonality, texture, shape and context. TerraClass Cerrado is another project coordinated by the Brazil's environment minister in cooperation with the Brazilian Agricultural Research Corporation (Empresa Brasileira de Pesquisa Agropecuária - Embrapa) and INPE. With this project, they map the use and cover of deforested areas of the Brazilian Amazon and they enable the characterization of the areas mapped by PRODES, using satellite images from 2013 ¹.

Similarly, MapBiomass ² is an initiative that analyzes the Brazilian territory by mapping the land use and land cover since 1985. MapBiomass has the potential to monitor primary forest changes. The methodology adopted

¹<http://www.dpi.inpe.br/tccerrado/>

²<http://mapbiomas.org/>

by MapBiomass is presented in [35]. Its dataset comprises images of all Brazilian biomes and they have been collected by Landsat 5-6-7 sensors. The methodology also involves extracting statistical features to train a Random Forest (RF) classifier. In a post-classification stage, spatial and temporal filters remove classification noise and fill information gaps due to cloud. In this methodology, all procedures are performed using the Google Earth Engine. However, the final validation stage is based on visual interpretation.

Further works have been developed by the RS community. Bueno et al. [4] present a study for detecting deforestation in highly seasonal areas of the Cerrado biome. They adopted the object-based image analysis (OBIA) methodology, applied to Landsat OLI (The Operational Land Imager) time series with the aim to find out the best spectral bands and/or vegetation indices for discrimination of true deforestation from seasonal changes using a RF classifier. Likewise, Machado et al. [36] present a study of mapping the deforested areas using images of MODIS sensor (or Moderate Resolution Imaging Spectroradiometer). A maximum likelihood classifier implemented in ERDAS software is used for the task.

Recently, Deep Learning (DL) techniques have been successfully applied to RS image analysis. Using Deep Neural Networks (DNNs), it is possible to learn multiple levels of data representation and to extract more robust and abstract features [24], which usually provide more meaningful information than hand-crafted ones. They are also able to capture spatial context information using different kernels in convolution operations.

Zagoruyko and Komodakis [37], proposed and explored different CNN architectures to learn similarity functions between image pairs that implicitly suffered some transformations and other kinds of effects (due to e.g., rotation, translation, illumination). These algorithms presented good performances in comparison to methods based on hand-crafted feature descriptors. Examples of such algorithms are the Early Fusion, Siamese CNN and Pseudo-Siamese, which were also used in [38] to detect changes in urban areas. In this work the authors compared the Siamese CNN and Early Fusion techniques, and evaluated the impact of using different spectral channels as inputs.

In a similar work [39] a Siamese CNN was successfully applied to identify building and tree changes, and also to distinguish between real changes from false ones caused by misregistration errors or false matches. However, the patches assigned as "change" go through a post-processing stage, then they are grouped and analyzed as individual objects to perform a filtering and segmentation stage, and finally detect individual object-level changes.

Zhan et al. [24] proposed a supervised change detection method based

on a deep Siamese CNN for optical aerial images. As in previous approaches, the authors of this work improve the preliminary classification result in post-processing stage. Essentially, the score map produced by the SN is segmented using a thresholding segmentation. The generated segments are then classified using a k-nearest neighbor (k -NN) approach. In a similar approach, the authors of [40] integrate the advantages of CNN and RNN to learn joint spectral-spatial-temporal features and solve a multispectral image change detection problem, achieving encouraging results. A drawback of LSTM is the number of parameters to be learned, which increases significantly compared to CNN.

3 FUNDAMENTALS

This chapter provides the basics for a proper understanding of the methods selected for deforestation detection. First, a brief introduction about Remote Sensing is given. Then, some concepts related to Deep Learning (DL) are described, centering on Convolutional Neural Networks (*CNN*), Siamese Networks (*SN*) and Convolutional SVM (*CSVM*).

3.1 Remote Sensing (RS)

Remote sensing is the process of acquiring information about an object, area, or phenomenon without being in physical contact with it [41]. This process is carried out by measuring emitted radiation by the sensor and reflection radiation by the object of interest. One of the most important advantages of RS is the possibility of collecting data of dangerous or inaccessible areas. This property makes RS a very useful tool to monitor deforestation, glaciers, and depth sounding of coastal and ocean depths.

In RS there are two basic types of sensors, passive and active. Passive sensors measure the proportion of the electromagnetic (EM) emitted by the sun, which is reflected by the imaged objects. Hence, orbital optical sensors can only obtain measurements in the day time. They are sensitive to weather conditions and also very vulnerable to clouds and shadows. On the other hand, active systems are equipped with their own source of energy to illuminate objects of interest. Active sensors emit their own EM energy towards the earth and measures the energy reflected from it. In contrast to the passive counterparts, active sensors can obtain measurements anytime (Day and Night) and operate in different environmental conditions (see Figure 2).

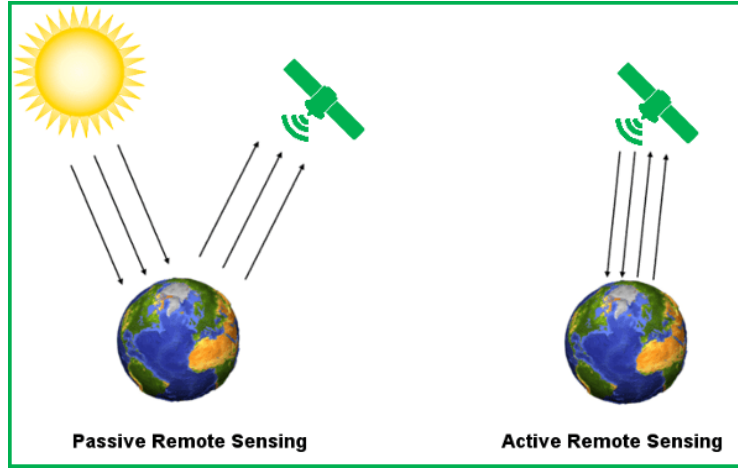


Figure 2: Passive and Active sensors. Passive sensors does not require any external power source to produce output signal while active sensor requires an external power source to operate.

3.2

Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (*CNNs*) are a kind of Neural Networks (*NN*) specialized in processing data with an inherent grid-like topology in which nearby entries are correlated, such as time series or a two-dimensional (2D) image [42]. They are comprised of units, called neurons, which have learnable weights and biases, and they are focused in processing information associated to a restricted location of the *visual field*, known as the receptive field. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity.

CNNs were introduced by LeCun et al. [43], outperforming other approaches and becoming the state-of-the art in areas such as image recognition, natural language processing and classification.

The basic *CNN* comprises one or more convolutional layers, each followed by a pooling layer which reduces the data resolution. Finally, there are one or more fully-connected layers, where the last one —the output layer—delivers class scores. For *CNNs* the layers are organized in three dimensions: width (w), height (h) and depth ($N_{features}$). Additionally, the neurons in one layer do not connect to all the neurons in the next layer but only to a small region of it [44]. The final output will be reduced to a single vector of probability scores.

Figure 3 illustrates a basic *CNN* architecture with two convolutional layers (with kernels sizes k_1 and k_2) + 2×2 pooling, a fully-connected layer, and finally, the output layer with m neurons, where m is the number of classes.

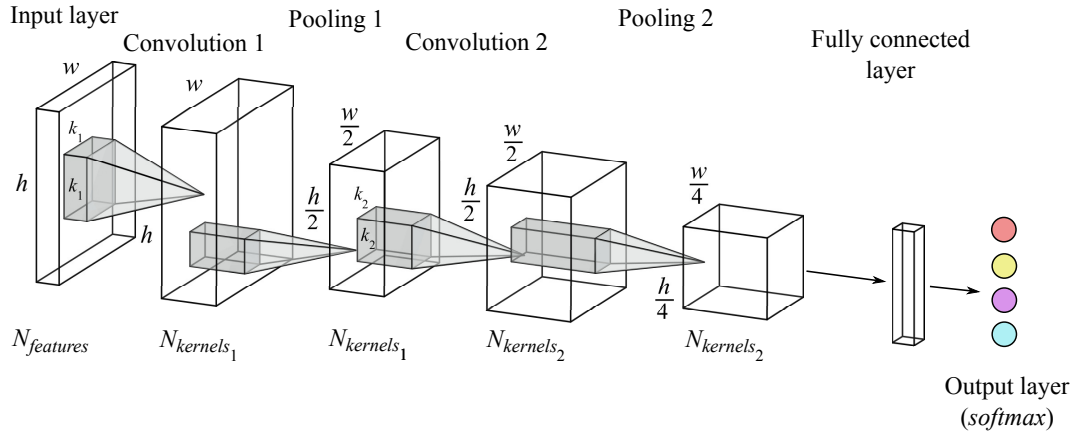


Figure 3: A *CNN* basic architecture with two convolutional layers. Adapted from [45].

In the following, each block is detailed, as well as other layers commonly used in *CNN* architectures.

- **Convolutional layer:** the term convolution refers to the mathematical operation on two functions to produce a third one. The arguments involved in this operation are the input data and the kernels, followed by a non-linear activation function. Each kernel is a $k \times k \times N_{features}$ multidimensional array of parameters (weights) that slides over the input data to produce a feature map, performing an element-wise multiplication and then summing up the results into a single output pixel. The output is a 3D tensor which dimension depends on the spatial dimension of the input data (w and h), the stride size, the padding size, and the number of selected kernels.
- **Batch Normalization (BN):** It is a trainable normalization layer that increases the stability of a neural network and reduces the dependency on the initialization [46]. *BN* normalizes the output of a previous activation layer by subtracting the mean and dividing by the standard deviation over each training mini-batch. Consequently, *BN* adds two trainable parameters to each layer, so the normalized output is multiplied by a standard deviation parameter (*gamma*) and add a mean parameter (*beta*). However, this normalization might not be beneficial for all cases. It depends on the data, the network architecture and the batch size.
- **Activation Function:** the output of a convolutional layer, called activation map, goes through a nonlinear activation function. Figure 4 illustrates the most common activation functions, *softmax*, *sigmoid*, *tanh*, *ReLU*, *leaky ReLU*, and *exponential LU*.

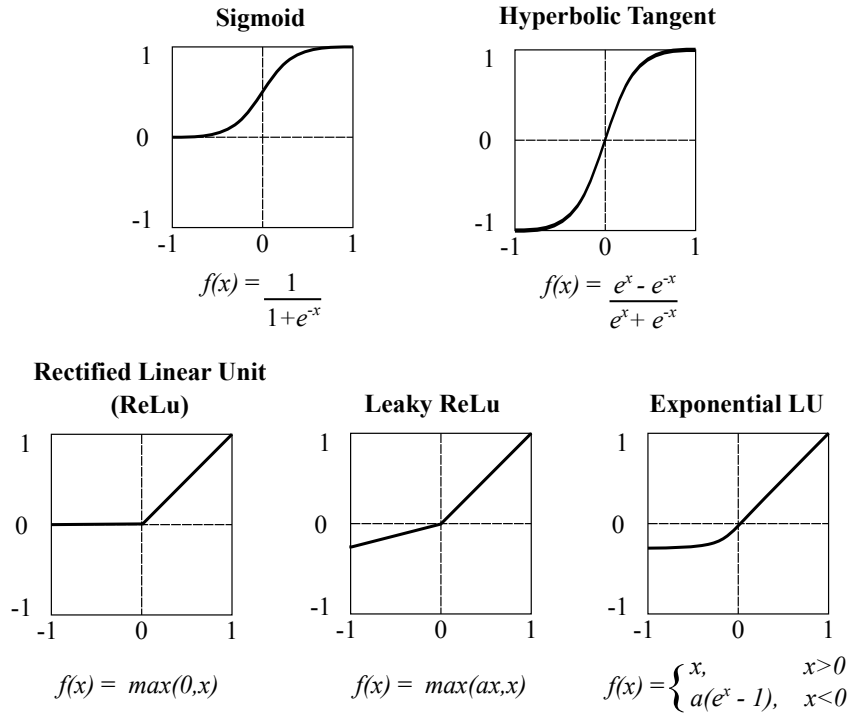
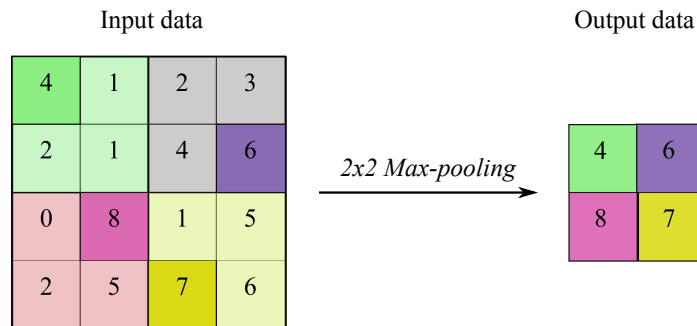


Figure 4: Illustration of the most common non-linear activation functions.

- **Pooling layer:** this layer summarizes the feature map over a neighborhood around each pixel. This operation usually reduces the spatial resolution of the feature maps, to reduce the amount of parameters, and improve the computational efficiency of the network, as well as controls overfitting [44]. The most common variant is the 2×2 *max-pooling*, which applies a 2×2 filter with a stride of 2 to every depth slice of the feature map, replacing all values covered by the filter in each location by the maximum value among them (see Figure 5).

Figure 5: Example of *max-pooling* applying a 2×2 filter with stride equal to 2. Each element of the output matrix is the max of the region in the original input.

- **Fully-connected layer (FC):** in this layer, each neuron has connections to all activations of the preceding layer, similar to a regular

NeuralNetwork. Usually, the last *FC* comprises a *softmax* activation function, which performs the classification task based on the features extracted by the previous layers, and assigning a posterior probability of belonging to each class.

- **Dropout:** is a regularization technique that reduces over-fitting [47] during the feature-learning procedure. Specifically, dropout can be considered analogous of training a large ensemble of models consisting of all sub-networks that share their parameters [42]. In this process, some randomly neurons are turned off in the forward pass, as well as their connections, during the training phase. Figure 6 shows an example of the application of dropout to a neural network with two hidden layers.

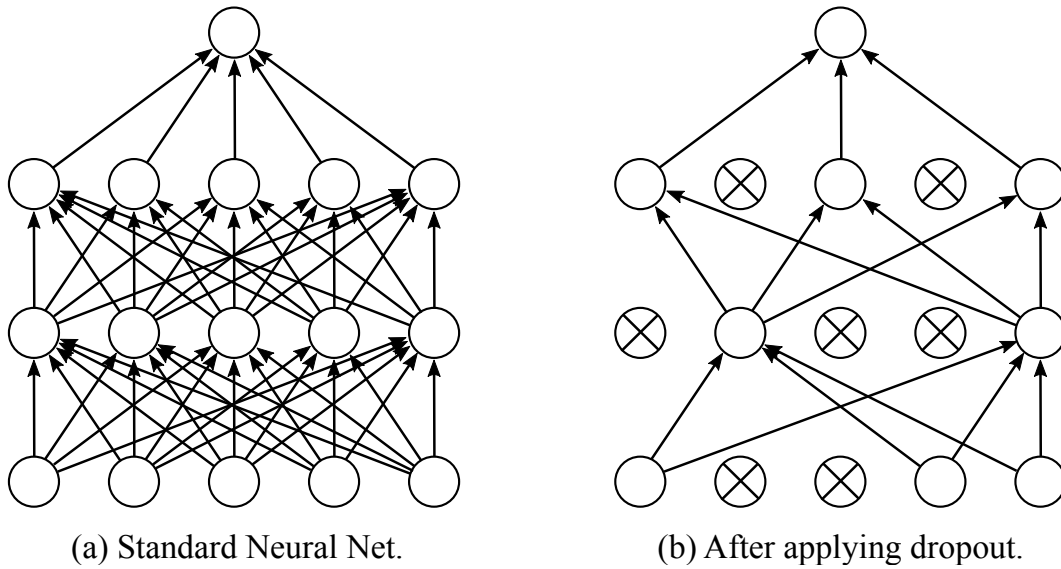


Figure 6: Dropout in a Neural Network. (a) A standard neural net with two hidden layers. (b) An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped [47].

3.3

Siamese Networks (SNs)

The Siamese Network is an adaptation of a traditional CNN. This network was first introduced in the early 1990s by Bromley et al. [48] for signature verification as an image matching problem. SN comprises two or more identical subnetwork components to estimate the similarity between two inputs and has the ability to learn their discriminative features. They also share the same hyperparameters and weights values [39], meaning that the descriptor vector of the two input data are extracted using the same approach [24]. Figure 7 shows the basic architecture of a SN.

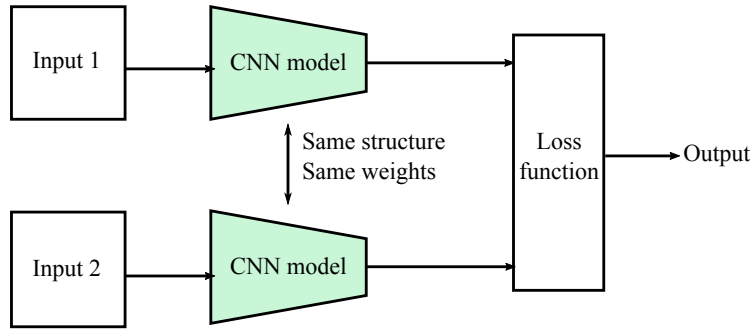


Figure 7: Architecture of the Siamese Network. The two CNNs share the same structure and weights.

Usually, SNs perform a binary classification at the output, classifying if the inputs correspond to the same class or not. Different loss functions may be used during training. One of the most popular loss functions is the *binary cross-entropy loss*, which is defined by:

$$L = -y \log p + (1 - y) \log(1 - p) \quad (3-1)$$

where L is the loss function, y the class label (0 or 1) and p is the prediction of the network.

3.4

Convolutional SVM (CSVM)

The CSVM, proposed by [49], is an alternative DL approach based on SVMs. This method was tested for object detection from Unmanned Aerial Vehicles (UAVs) imagery and performed well in the task of discerning between instances of object of interest and the background. In the reported experiments [49], CSVMs outperformed other deep learning based approaches for small training sample sizes. This characteristic motivated the inclusion of CSVM in this study.

Analogous to a traditional CNN, a CSVM architecture is composed of a set of convolutional and pooling layers followed by a classification layer at the end [49]. But in contrast to CNN, CSVM does not rely on the backpropagation algorithm for training, it trains the set of linear SVMs in a layerwise fashion. The description of each layer of CSVM for the task of image classification is presented in the following.

- **SVM Convolutional Layer:** this layer uses a set of linear SVMs as convolutional filters for generating the feature maps. Likekise, the weights of these filters are learned in a forward supervised way [49]. The training process of the CSVM model is summarized in Figure 8.

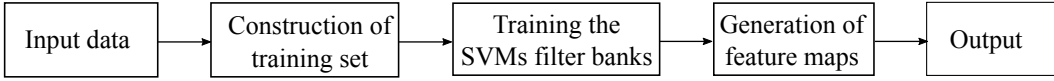


Figure 8: SVM convolutional layer. The procedure starts with the construction of a training set, followed by the training of SVM filter banks and ends with the generation of the convolutional feature maps.

Construction of training set

Let's assume henceforth that positive images contain the object of interest, whereas the negative ones correspond to the complementary class in a binary classification problem. For each training image, patches of size (h_1, h_1, c) are randomly selected, where h_1 corresponds to the spatial dimensions, and c is the number of bands of the input data. Each selected patch is flattened into a vector of size $d = h_1 \times h_1 \times c$. The vectors obtained this way form the set of samples available for training a bunch of SVMs filters in the next step (see Figure 9).

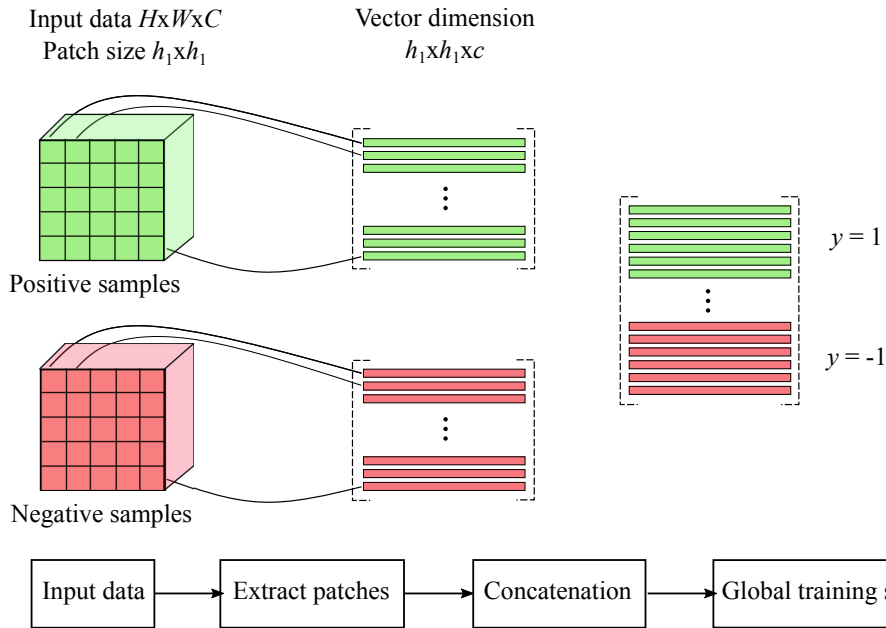


Figure 9: Generation of training set for the first convolutional layer of CSVM.

Training the SVMs filter bank

In this stage, an ensemble of SVMs is trained. The diversity among the SVMs in the ensemble is obtained by training each each one of them upon

a different random subset of the samples collected as explained before. The subsets are composed of n samples per class, which are randomly selected from the global training set. The weights of the m SVMs filters obtained after training are grouped into a tensor of dimensions $(h_1 \times h_1 \times c \times m)$.

Generation of feature maps

In this step, each training sample is convolved with the SVMs filters to generate features maps (see Figure 10). Then, they are fed to a pooling layer followed by a non linear activation function and by a batch normalization layer. The output is the input for the next convolution layer. The same procedure is adopted to build each layer.

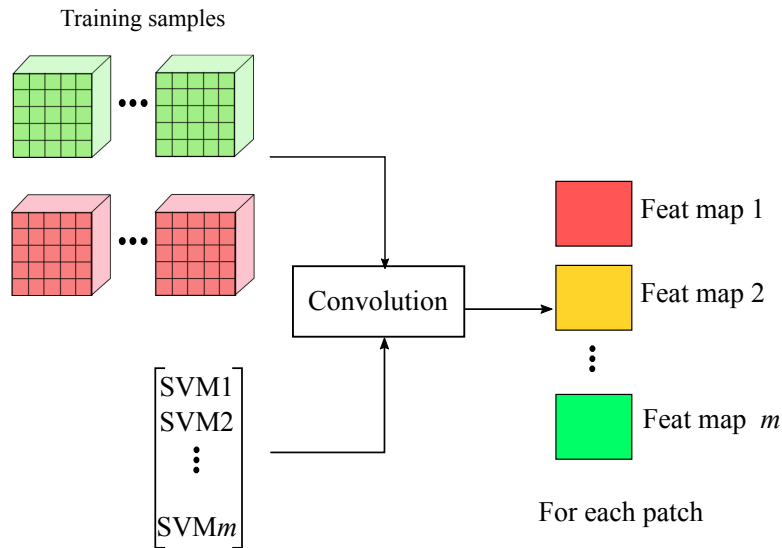


Figure 10: Feature maps obtained from the first convolutional layer of CSVM.

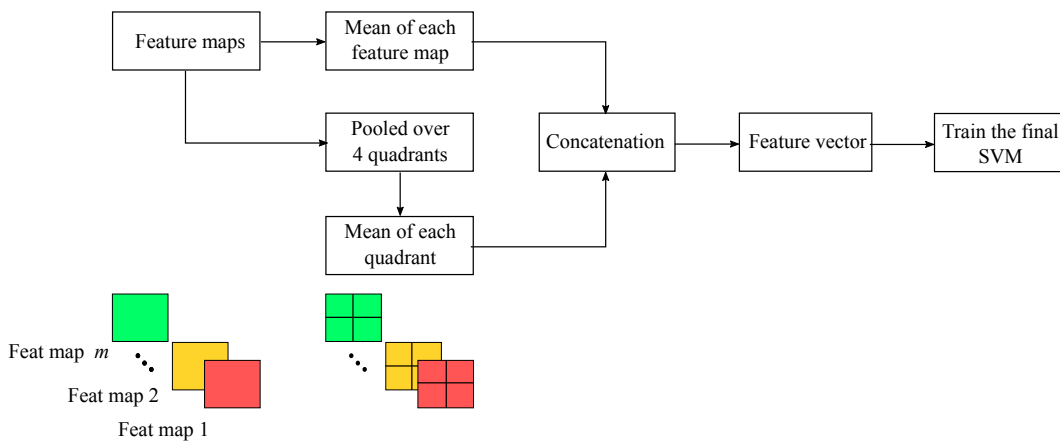


Figure 11: Creation of the training set for the final classifier.

- **Reduction layer:** this layer performs the same function as the spatial pooling in CNNs. It is applied after the convolutional layer and reduces the spatial size of the feature maps by selecting the most representative features for the next layers.
- **Feature generation and classification:** at this step features are extracted from each patch, which are then fed to a binary SVM classifier. A five dimensional feature vector is computed from each feature map. The first element is given by the mean of the entire feature map. The next four elements are obtained by first splitting each feature into four quadrants and then averaging each one, as illustrated in Figure 11. Thus, the dimension of the complete feature vector computed for each patch is $(4 \times m) + m$, where m is the number of SVMs in each layer. This process is applied to each training and test sample. Finally, the SVM assigns the final class label (1, -1).

4 METHODS

This chapter complements the description presented in chapter 3 about the methods evaluated in this work for deforestation detection: Early Fusion (EF), Siamese Network (SN), and Convolutional SVM (CSVM). Implementation details, as well as the particular setups, are clarified in the following sections.

For all three methods, the inputs are pairs of co-registered patches of two optical images acquired at different dates, denoted henceforth T1 and T2. The classification result is assigned to the patch central pixel. A sliding window approach is adopted to classify all pixels of the target site.

4.1 Early Fusion (EF)

EF is a CNN based model composed of a series of convolutions and pooling layers, followed by fully connected (FC) layers, whereby the last one is a Softmax layer having as many outputs as the number of classes. Softmax assigns posterior probability values to each class in a classification problem, which add up to 1. For deforestation detection, this layer has two outputs, related with the "deforestation" and "no-deforestation" classes and the final label is assigned based on the class corresponding to the maximum probability.

The EF architecture used in this thesis was inspired on the CNN model proposed in [38] used for detecting changes in urban areas with good reported performance. The architecture involves the concatenation of both images, before applying the CNN model. Hence, the two images (T1 and T2) are stacked along their spectral dimension to generate a unique input image for patch extraction. Then, each patch is a tensor of size h -by- h -by- $2c$, denoting the patch height, width and depth respectively. The procedure is illustrated in Figure 12.

4.2 Siamese Network (SN)

As explained in chapter 3, a SN architecture contains two identical subnetworks that estimate the similarity between two inputs. The architecture

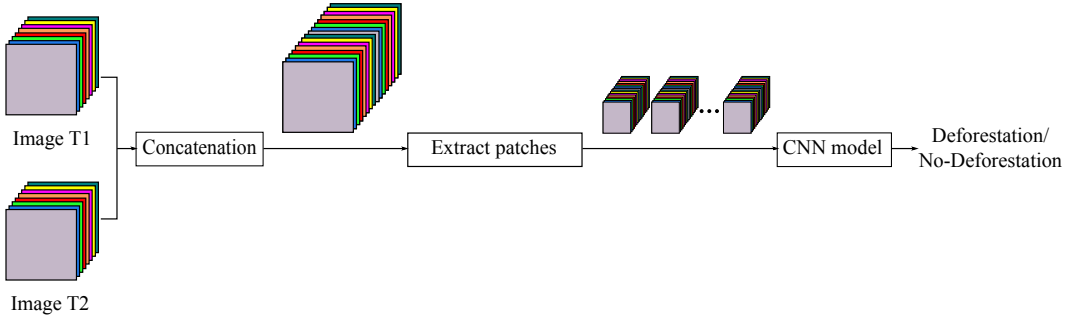


Figure 12: EF method. Images at different dates (T1 and T2) are concatenated to produce an image pair; then, patches are extracted and fed to the CNN model.

used in this work was brought from [38], also applied for urban changing detection. In this method, both input images are treated independently and each SN subnetwork receives as input one patch cropped from corregistered image pair (T1 and T2). The features obtained from each SN subnetwork are concatenated to produce the final feature vector. This vector is also followed by two (FC) layers, where the last one is a Softmax layer with two outputs that assigns it to a class: "deforestation" and "no-deforestation". This process is summarized in Figure 13.

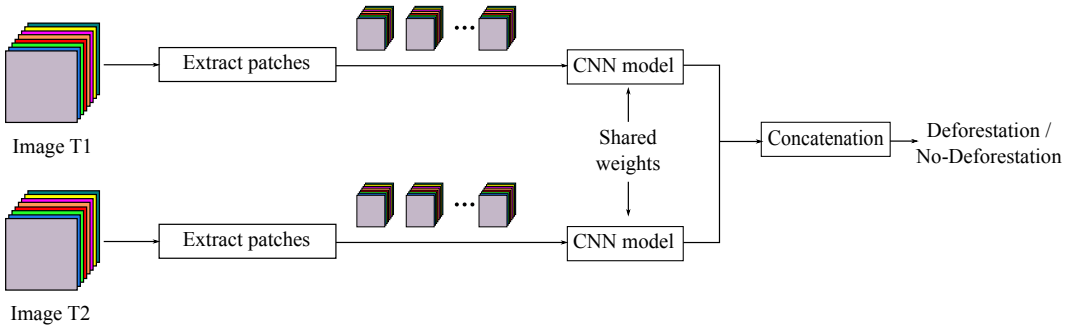


Figure 13: Siamese network. Patches of each image (T1 and T2) are extracted and fed to the CNN model independently. The two branches in the network share exactly the same architecture and parameters values.

4.3 Convolutional SVM (CSVM)

Similar to the EF method, the two input images (T1 and T2) are concatenated along their spectral dimension. Again, as in EF, in the CSVM approach we classify patches of size h -by- h -by- $2c$ whose classification output is assigned to the patch central pixel. Next, we describe how the method proposed

by Bazi and Melgani [49] for image classification was adapted in this work for pixel-wise deforestation detection.

Construction of training set

Following patches extraction, the training set is created for learning the SVMs filters. Then, the input patches are divided into nonoverlapping rectangular sections, which are mini-patches of size h_1 -by- h_1 -by- $2c$. Next, they are vectorized to form the global training set. This procedure is showed in Figure 14 -a.

Training the SVMs filter bank

After the creation of the global training set, subsets of m random selected samples are created to train m SVMs filter. These m subsets are composed of n samples per class, which are randomly selected from the global training set. In our study, the value of n corresponds to the ratio of number of training samples and the amount of SVMs filters (see Figure 14-b).

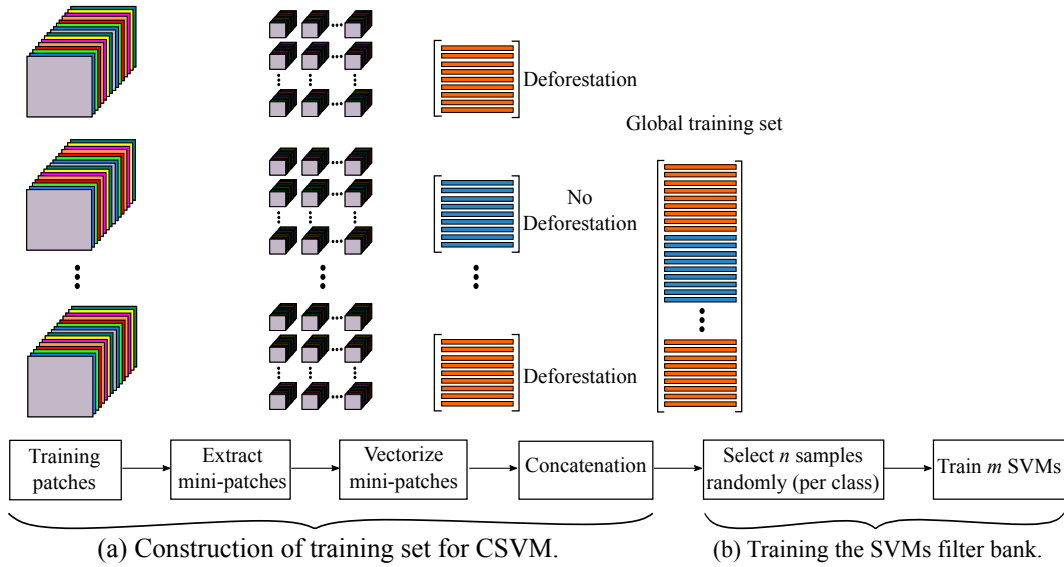


Figure 14: Procedure to train the SVMs filter bank. Mini-patches are extracted from input patches and they are vectorized to compose the training set of SVMs.

Generation of feature maps

In this step, the input patch pairs are convolved with the learned SVM filters to generate the feature maps, which are fed to a pooling layer followed by

a non linear activation function. The output is the input to the next convolution layer. The procedure is repeated until the desired number of layer is reached (see Figure 15).

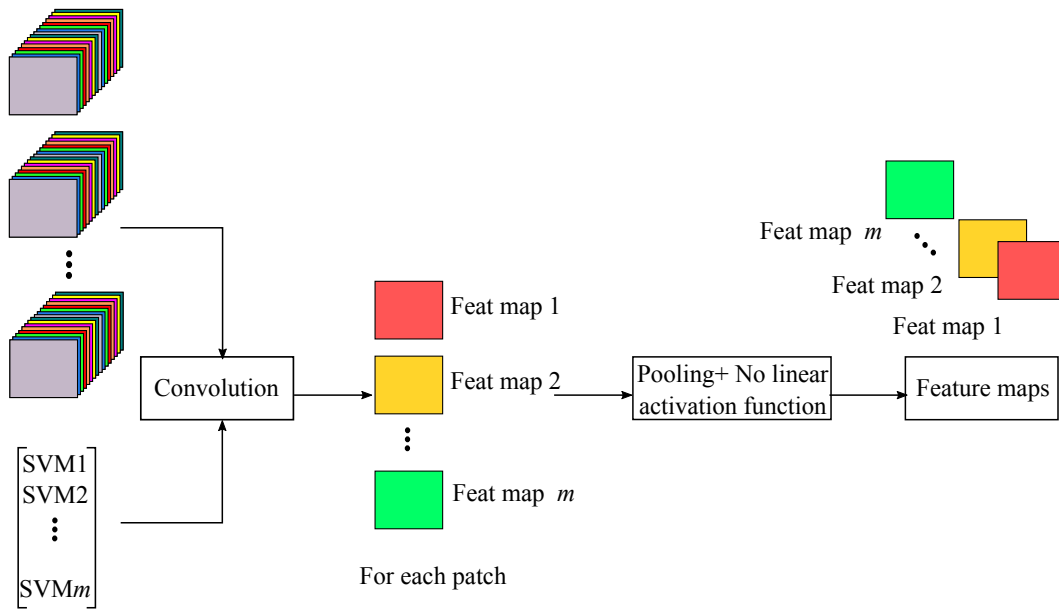


Figure 15: Generation of feature maps.

Classification

As mentioned before, the feature maps obtained in the last layer are fed again to a final binary SVM classifier that identifies the class label to each patch central pixel, either as "deforestation" or "no-deforestation". In contrast to the original CSVM, where a vector of means was obtained, the feature descriptor is a vector obtained after flattening the output of each convolutional layer. This procedure was carried out after experimental analysis where the results presented a better performance as well as a reduction in the inference time.

5 EXPERIMENTAL ANALYSIS

This chapter reports experiments carried out to evaluate the methods introduced in the previous chapter. First, the datasets used in the experiments are presented. Then, the experimental protocol and the parameter setup are detailed. Finally, the results obtained in all experiments are reported and discussed.

5.1 Datasets

Two study areas from the Brazilian Biomes were selected to evaluate the methods. The first one is an Amazon region, and the second one is in Cerrado. The detailed description of each one is presented in the following.

5.1.1 Amazon Biome

The first study area corresponds to a region of the Amazon Biome, more specifically localized in the Pará State, Brazil, centered on coordinates of 3° 17' 23" South and 50° 55' 08" West, Figure 16. Pará state comprises 26% of Brazilian Amazon [50] and most of it is covered with dense tropical rainforest. This territory has faced a significant deforestation process, and its progress has been tracked and monitored by PRODES [31].

The reference change map used in our experiments refers to the deforestation occurred between August 2016 and August 2017. This information was downloaded from the INPE site, which is freely available at the PRODES database ¹. For this reference the following considerations were taken into account:

- Some polygons were not considered because they refer to areas deforested in previous years.
- A buffer of two pixels around the polygons of class "deforestation" was not considered for the training, validation and test. The reason was to avoid the impact of inaccuracies that may have occurred as photointerpreters delineated these polygons.

¹Available at: <http://terrabrasilis.dpi.inpe.br/map/deforestation>

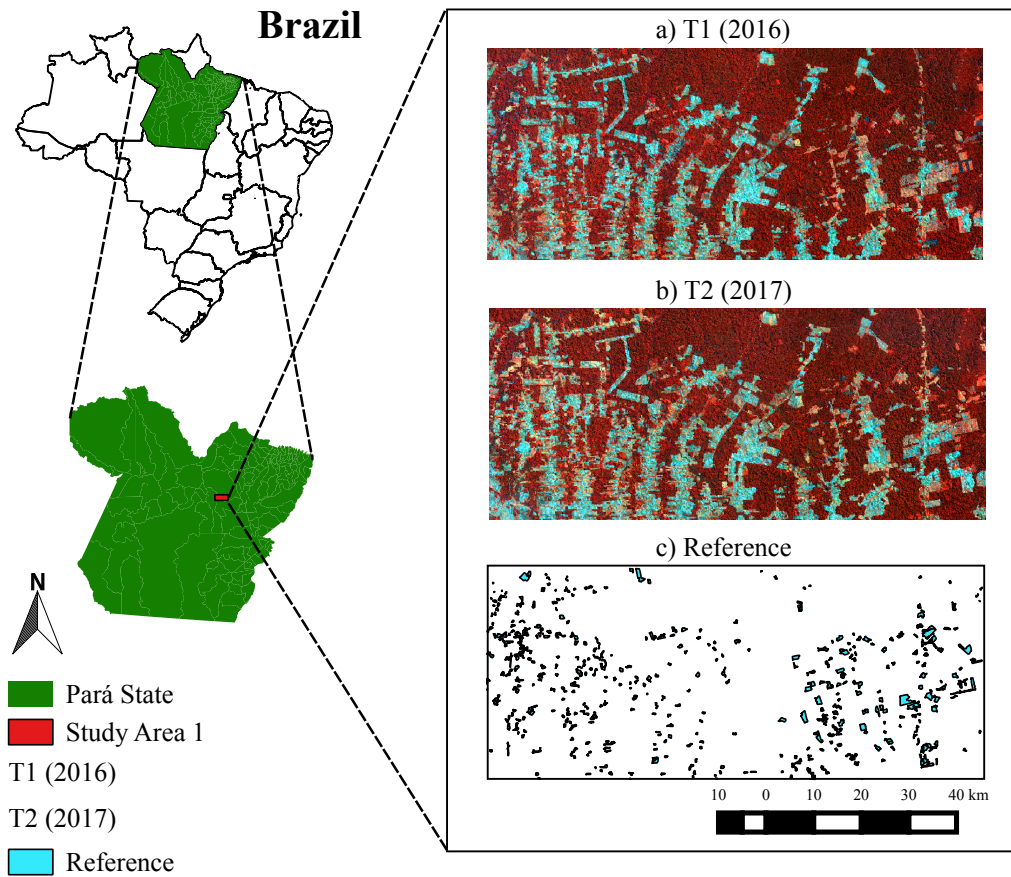


Figure 16: Region of the Brazilian Amazon, located in the Pará State, Brazil. *NIR – G – B* composition of the study area at dates T1 (a) and T2 (b). c) Deforestation reference from 2016 to 2017.

- Areas lower than 6.25 ha (69 pixels) were also not considered in our evaluation because PRODES data does not record deforestation areas smaller than that.

The dataset comprises a pair of Landsat 8-OLI, with 30m spatial resolution. The images were acquired from the United States Geological Survey (USGS) Earth Resources Observation and Science Center. Additionally, each scene was atmospherically corrected to each scene, and then, we clipped them to the target area. The final images are 1100×2600 pixels large with seven spectral bands (Coastal/Aerosol, Blue, Green, Red, NIR, SWIR-1, and SWIR-2). The first image is from August 2nd, 2016 and the second one from July 20th, 2017 (Figure 16). These dates were selected based on PRODES reference date, which computes the annual deforestation rate from August 1st of each year, during the dry season (June to September), when the cloud

cover, a major problem over all BLA region, is minimum.

5.1.2

Cerrado Biome

The second study area belongs to the Brazilian Cerrado biome, localized in the Maranhão State, Brazil, centered on coordinates of 4° 58' 53" S and 43° 49' 41" W. Figure 17 illustrates this study area. Maranhão State contains 64% of Cerrado vegetation and has suffered a significant agricultural expansion, most of it over native vegetation [51]. The deforestation in this biome has been also monitored by PRODES². The dataset also comprises a pair of Landsat 8-OLI images with seven spectral bands, pre-processed in the same way as in the Amazon dataset. The size of the images is 1717 × 1442 pixels. For this database, the first image is from September 3rd, 2017 and the second one is from September 22th, 2018. Due to the reference provided by PRODES was also obtained during the dry season, the reference used in this case does not contain all the deforested areas. Then, the reference had the following adaptations:

- Some areas that suffered a deforestation after the PRODES report were included in the reference. The added polygons were reviewed and approved by an expert photointerpreter. The final reference change map of the Cerrado is presented in Figure 17.
- A buffer of two pixels around the samples of class "deforestation" was not considered in our evaluation to avoid the aforesaid inaccuracy problem along the borders.
- Areas lower than 1 ha (11 pixels) were not considered in the computation of the accuracy metrics because PRODES data does not consider deforested areas smaller than this value.

5.2

Experimental Setup

For all the methods, two optical images acquired at different dates were used. In addition, the *Normalized Difference Vegetation Index (NDVI)* was calculated, as given in Equation 5-1. This index quantifies the presence and quality of vegetation and it is calculated using bands 5 and 4 for Landsat 8, corresponding to the spectral reflectance measurements acquired in the near-infrared and red regions.

²Available at: <http://www.obt.inpe.br/cerrado/>

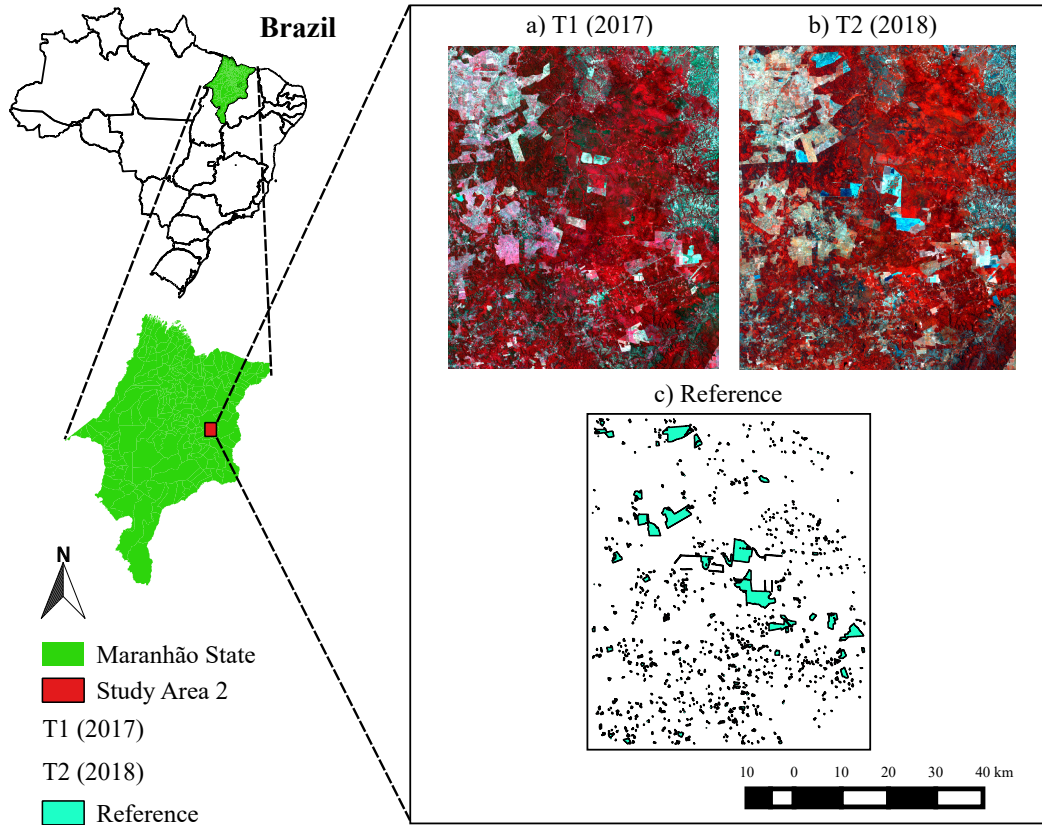


Figure 17: Region of the Brazilian Cerrado biome, located in the Maranhão State, Brazil. *NIRGB* composition of the study area at dates T1 (a) and T1 (b). c) Deforestation reference from 2017 to 2018.

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (5-1)$$

The NDVI was stacked along the spectral dimension of the corresponding images, resulting in images with eight bands. Then, the spectral bands of each image were normalized to zero mean and unit variance.

The input to EF and CSVM was a tensor of size of 15-by-15-by-16, for SN a tensor of size of 15-by-15-by-8 in each branch and for SVM a vector of size of $15 \times 15 \times 16$. The patches were extracted following the overlapping sliding windows procedure with stride of three. The patch size and the stride size were selected empirically. In all methods, the input is an image patch and the classification outcome is assigned to the patch central pixel.

Similar to [39], the input images were divided into tiles. Then, from the Amazon and Cerrado databases, 15 tiles were obtained as shown in Figure 18 and Figure 19 respectively. From each image, four tiles were used for training, two tiles for validation, and nine tiles for testing.

As both datasets are highly unbalanced (see Figures 16-c and 17-c), a data augmentation on deforestation samples was adopted. Each training pair

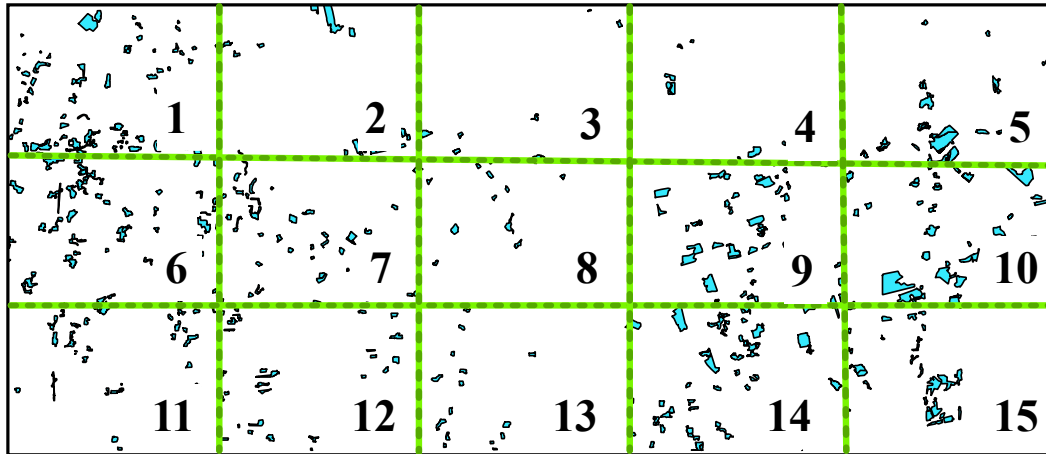


Figure 18: Distribution of the Amazon database. The region was divided into fifteen tiles. Four tiles was used for training (1, 7, 9, 13). Two for validation (5, 12). The rest of the them were used for testing (2, 3, 4, 6, 8, 10, 11, 14, 15).

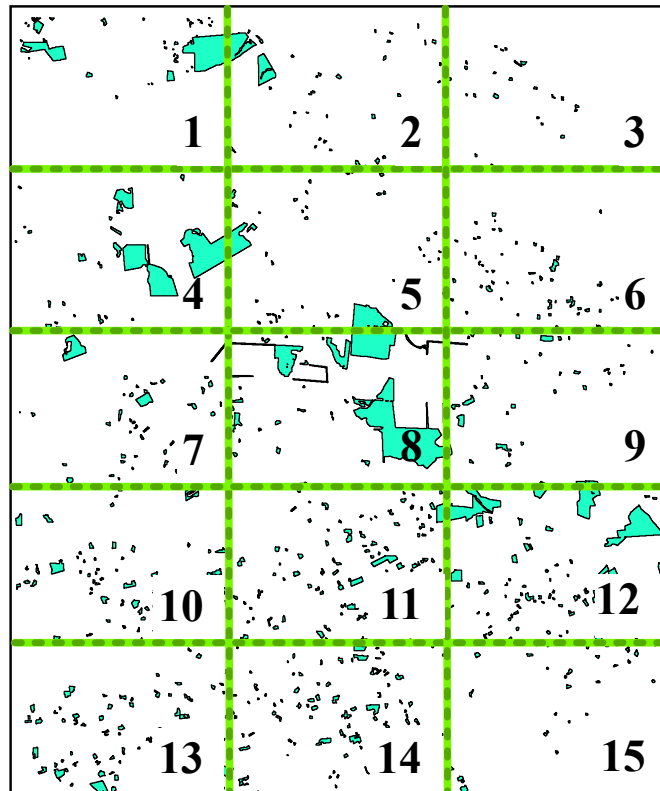


Figure 19: Distribution of the Cerrado database. The region was divided into fifteen tiles. Four tiles was used for training (1, 5, 12, 13). Two for validation (6, 10). The rest of the them were used for testing (2, 3, 4, 7, 8, 9, 11, 14, 15).

was rotated 90° and flipped over the horizontal and vertical axis. Likewise, an under-sampling technique on the majority class was applied in order to balance the number of training pairs for both classes. Just part of all "no-deforestation"

samples available in the training tiles was randomly selected for training and validation, to balance both classes.

Table 1 and Table 2 present the number of available patches pairs of training, validation and test for the Amazon and Cerrado databases. Tables also present the number of patches obtained after apply the balancing procedure for both classes, "deforestation" and "no-deforestation".

Table 1: Number of samples in the training, validation and test sets for Amazon database.

Set	Tiles	Available Def. Samples	Available No-def. Samples	Balanced Samples (per class)	Total Samples
Training	1, 7, 9, 13	2706	78431	8118	16236
Validation	5, 12	963	39697	2889	5778
Test	2, 3, 4, 6, 8, 10, 11, 14, 15	40392	1675608	-	1716000

Table 2: Number of samples in the training, validation and test sets for Cerrado database.

Set	Tiles	Available Def. Samples	Available No-def. Samples	Balanced Samples (per class)	Total Samples
Training	1, 5, 12, 13	4182	65717	12546	25092
Validation	6, 10	663	34658	1989	3978
Test	2, 3, 4, 7, 8, 9, 11, 14, 15	68983	1416278	-	1485261

The network architectures used to implement the EF and SN approaches was adapted from [37] and [38]. The architecture is composed of three convolutional layers (Conv) with ReLU as activation function, two Max-pooling (MaxPool) layers and two Fully Connected layers (FC), being the last one a softmax with two outputs, associated to "deforestation" and "no-deforestation" classes. The kernels and output size of each layer are illustrated in Figure 20. For training we used the following setup: batch size equals to 32 with 100 number of epochs, early stopping to break after 10 epochs without improvement (over the validation set) and a dropout with rate set to 0.2 in the last fully connected layer. In contrast to [38], where

Average Stochastic Gradient Descent (ASGD) was used, in this work the Adam optimizer was employed, which presented a better performance in the preliminary experiments with learning rate of 10^{-3} and weight decay of 0.9. As loss function we used the binary cross entropy.

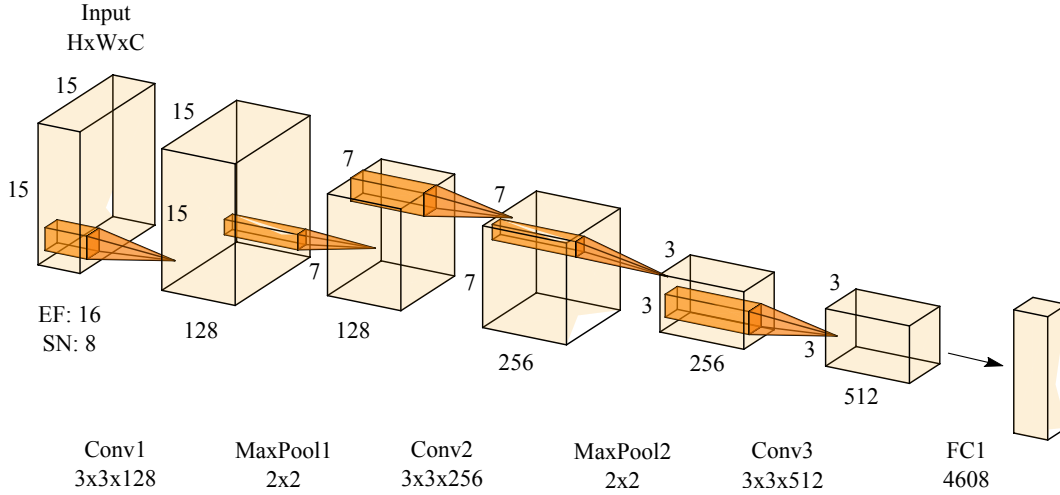


Figure 20: Architecture of EF and SN.

For the CSVM approach, the architecture comprised three convolutional layers (Conv) with ReLU activation function and three Max-pooling (MaxPool) layers. The output size from each layer is illustrated in Figure 21. In this method and for the baseline, the validation samples were added to the training set. For the computation of the weights of the SVM filters, the multicore Liblinear software package [52] was used. The parameter setup of the CSVM was: stride equal to one for the Conv and MaxPool layers, 12 SVMs used in each Conv layer. The global training was splitted between the number of SVM, in such a way that each SVM has the same number samples for both classes. The size of the mini-patches used for learning the SVMs was equal to $3 \times 3 \times 16$ for the first convolutional layer and $3 \times 3 \times 12$ for the second and third layers. These parameters were selected after experimental analysis. Finally, the estimation of the penalty parameter C for each SVM was performed via a threefold cross-validation procedure in the range $[10^{-1}, 10^3]$.

The buffer of both references was obtained applying the morphological dilation, using a disk as structuring elements with radius equal to two. This buffer creates an outer edge, and the patches with the central pixel in this region were not considered for training and validation samples, as well as to obtain the metrics from the test set.

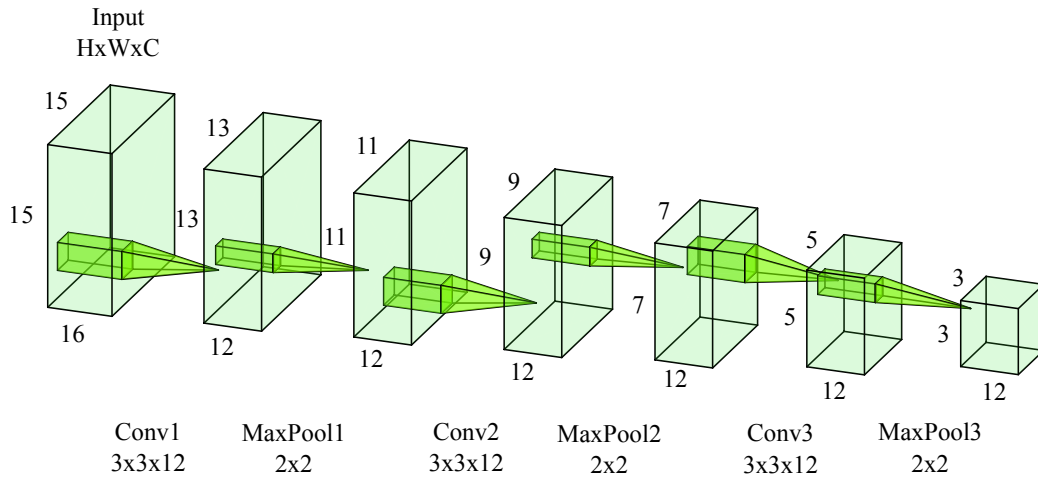


Figure 21: Architecture of CSVM.

Influence of the number of training samples

In order to evaluate the influence of the number of training samples, three additional scenarios were considered, using training samples from a single tile, from two tiles, from three tiles and the initial training set from four tiles, denoted as $N1$, $N2$, $N3$, and $N4$ respectively. The tiles used in each scenario were selected based on the condition $N1 < N2 < N3 < N4$. For EF and SN methods, validation set (val) was used to stop the training once the loss starts to increase (early stopping), but for CSVM and SVM the samples in this set were added to training set (tr). The number of training samples for the Amazon and Cerrado databases are presented in Table 3 and Table 4, respectively.

Table 3: Training tiles used for the Amazon database.

Training Set	Tiles	Available Def. Samples	Available No-def. Samples	Balanced Samples (per class)	Total Samples (tr+val)
1 Tile	13	239	20306	717	1434+5778
2 Tiles	1, 13	709	40515	2127	4254+5778
3 Tiles	1, 7, 13	1807	59102	5421	10842+5778
4 Tiles	1, 7, 9, 13	2706	78431	8118	16236+5778

Table 4: Training tiles used for the Cerrado database.

Training Set	Tiles	Available Def. Samples	Available No-def. Samples	Balanced Samples (per class)	Total Samples (tr+val)
1 Tile	5	671	17370	2013	4026+3978
2 Tiles	5, 13	1240	33760	3720	7440+3978
3 Tiles	1, 5, 13	2287	50273	6861	13722+3978
4 Tiles	1, 5, 12, 13	4182	65717	12546	25092+3978

5.3

Accuracy Assessment

The metrics selected to evaluate the results were *F1-Score*, *Overall Accuracy* (*OA*) and *Alert Area* (*AA*). These metrics are obtained from the *Confusion matrix* (*CM*), which is a common scheme to measure the performance of a classification algorithm. Each row of the matrix represents the predicted class sample,s while the columns represent the actual samples of each class. Table 5 shows a typical *CM* for a binary classification problem.

Table 5: Confusion matrix for a binary classification problem.

	Actual positive (<i>P</i>)	Actual negative (<i>N</i>)
Predicted positive	<i>tp</i>	<i>fp</i>
Predicted negative	<i>fn</i>	<i>tn</i>

True positives (*tp*) are defined as the correct classifications of the interest class, *False positives* (*fp*) refer to the samples erroneously assigned to the interest class. Analogously, *true negatives* (*tn*) and *false negatives* (*fn*) correspond to samples correctly and incorrectly assigned to the complementary class, respectively. *P* and *N* correspond to the total samples of the positive and negative classes to predict, which in this work represent the "deforestation" and "no-deforestation" classes respectively.

- *Overall Accuracy* (*OA*): is a global metric that indicates the percent of samples correctly classified with respect to the total samples. This

metrics is expressed as a percent, and it varies in a range of 0 to 1, being 0% the worst classification and 100% a perfect classification. It is defined by:

$$OA = \frac{tp + tn}{P + N} \times 100 \quad (5-2)$$

- *Precision*, also known as *Correctness*, represents the proportion of samples assigned by the classifier to the interest class, which truly belongs to the that class.

$$Precision = \frac{tp}{tp + fp} \quad (5-3)$$

- *Recall*, also known as *Completeness*, gives the proportion of all samples of the interest class recognized by the classifier as such:

$$Recall = \frac{tp}{tp + fn} \quad (5-4)$$

- *F1-score*: is given by the harmonic mean of *Precision* and *Recall* and it also varies in a range of 0 to 1. This metric is defined by:

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100 \quad (5-5)$$

- *Alert Area*: this metric is computed to measure portion of the area being monitored that comprises the samples classified as the interest class. The above with the aim to quantify how much the methods reduce the human effort during the photointerpreter tasks. We defined this metric by the rate of *tp* and *fp* between the total *P* and *N* samples in the test set.

$$AA = \frac{tp + fp}{P + N} \times 100 \quad (5-6)$$

5.4 Results

The experimental results are presented in the following subsections, first for Amazon, then for Cerrado. For both databases, we ran five times each experiment and the results presented a variance about ± 1 .

5.4.0.1

Amazon Biome

Figure 22 summarizes the results of the experiments on the Amazon biome in terms of *F1-score* for the deforestation class. The Figure shows that SN and EF achieved the best performance in terms of *F1-score* in all experiments. As expected, the methods improved their performance as the number of training samples increased. However, CSVM was only able to reach the baseline performance in the last experiment, when four tiles were used for training. It presented low scores in comparison with the other methods, as well as different behavior in each scenario. Using a single tile for training, the performance was similar after the second and the third layer; using two and four tiles for training, the best performance was obtained after the second layer and using three tiles for training, the performance decreased as more layers were added.

The *F1-score* recorded in the first scenario, using a single tile for training was equal to 46%, 51% and 53%, 40%, 43%, 43.5% for SVM, EF, SN and CSVM (after the first layer (L1), second layer (L2) and third layer (L3)), respectively. EF and SN outperformed SVM in 5% and 7%, respectively, but not CSVM. When two, three and four tiles were taken for training, the SN and EF presented better performance than SVM and CSVM, with a gain margin of about 10%. With four training tiles, EF and SN outperformed SVM in 13% and 11%, respectively.

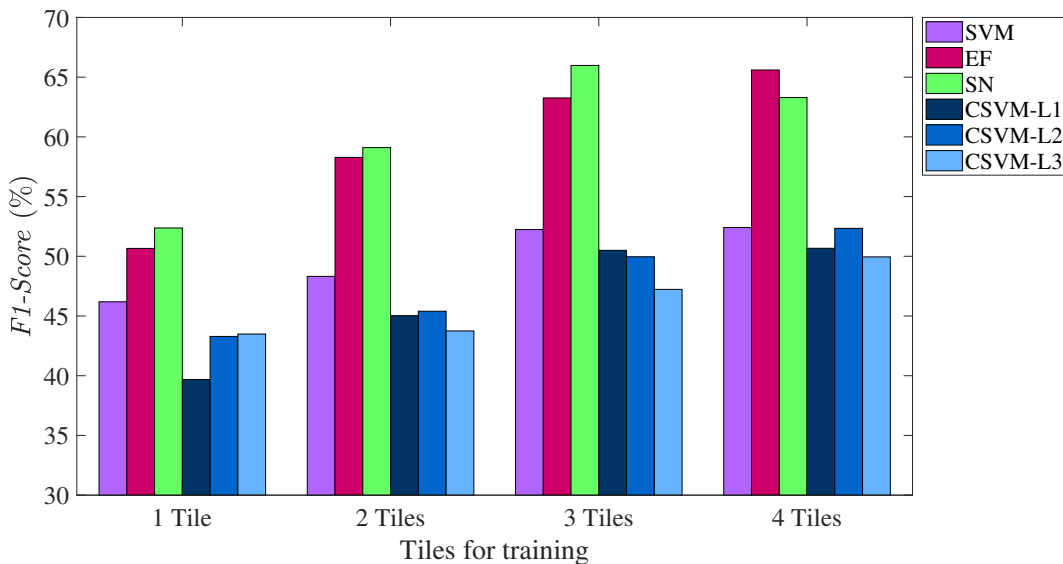


Figure 22: *F1-score* of Amazon Biome obtained from SVM, EF, SN, and CSVM using one, two, three, and four tiles for training.

The results in terms of *Overall Accuracy* (OA) are presented in Figure 23. Similar to the *F1-score*, the results improved when the number of training tiles

increased. In all scenarios, scores above 90% were achieved. The scores of EF and SN went from about 95%, for one tile, to 98% when four tiles were used, while SVM and CSVM remained in the range of 94% and 96%. Again, CSVM presented lower scores in comparison to other methods and its performance was similar to *F1-score* in each scenario. However, the higher values for *OA* were related to the higher number of no-deforestation samples that were correctly classified. Indeed, about 97% of the test area corresponds to "no-deforestation".

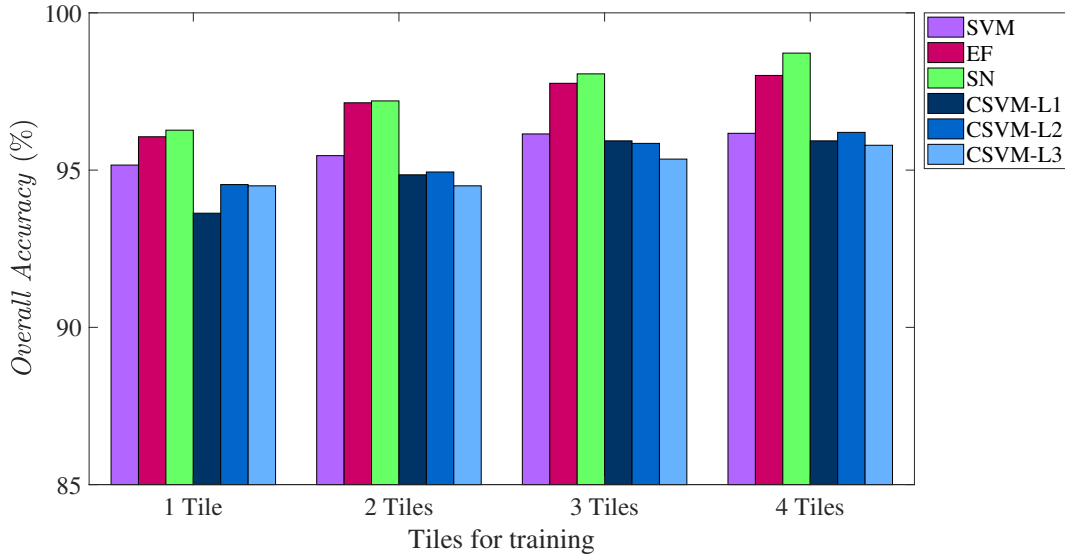


Figure 23: Overall Accuracy of Amazon Biome obtained from SVM, EF, SN, and CSVM using one, two, three, and four tiles for training.

Figures 24, 25, and 26 show the *NIR-G-B* composition (Near Infrared, Green and Blue bands) at both dates (T1 and T2), the reference and the probability maps of the second, sixth and, fourteenth tiles, respectively. These tiles are part of the test set. Columns correspond to methods SVM, EF, SN, and CSVM (after layer 2) and rows correspond to the output of each one using one, two, three and four tiles for training. Blue color represents the lowest probability of deforestation, while Red color represents the highest probability.

As in the *F1-score* and *OA* plots, the probability maps improved and the *salt-and-pepper* effect reduces as the number of training tiles increased. In the first scenario, when a single tile for training was used, SVM presented more false deforestation detection, followed by CSVM, causing a more noticeable *salt-and-pepper* effect. However, the CSVM was more confident in comparison to SVM, which delivered more probability values in an intermediate range. Actually, among all tested methods, SVM was the most uncertain in its results. Similar to CSVM, the probability output of EF and SN were close to one (deforestation) and zero (no-deforestation). Hence, these methods were more confident about their predictions. Besides, all methods presented probabilistic

values in the intermediate range mainly close to the polygon borders, which could result from inaccuracies in the delimitation of deforested areas in the reference.

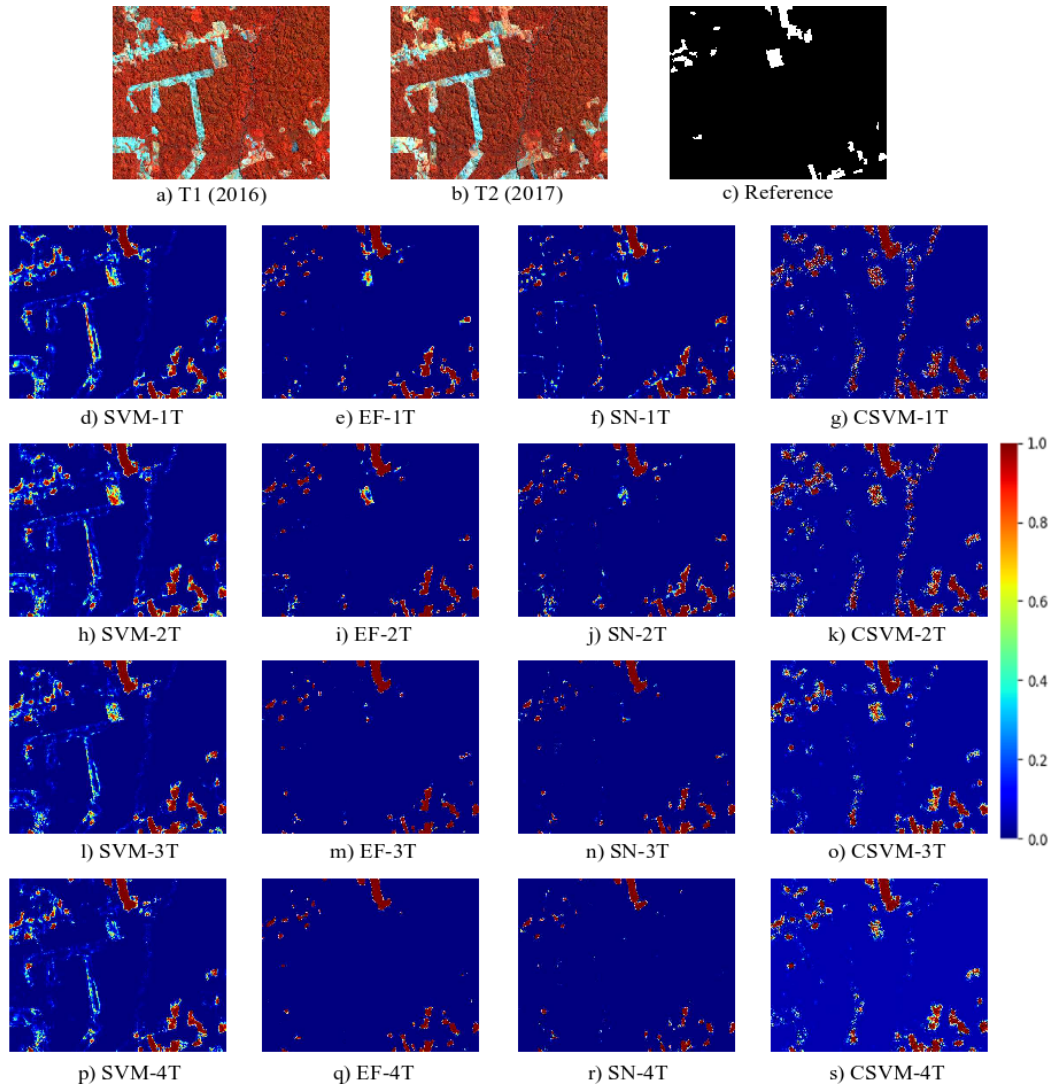


Figure 24: Predicted maps of the second tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.

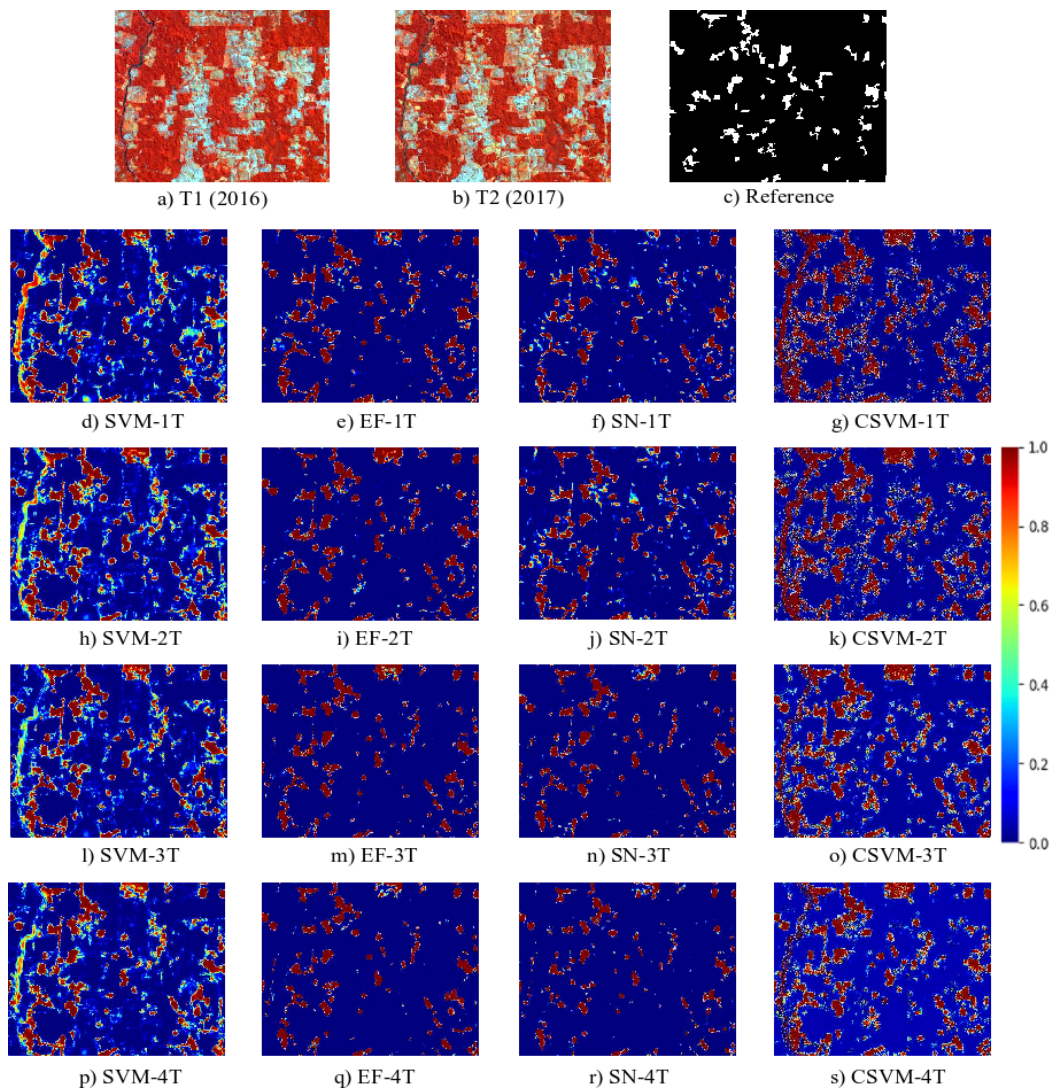


Figure 25: Predicted maps of the sixth tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.

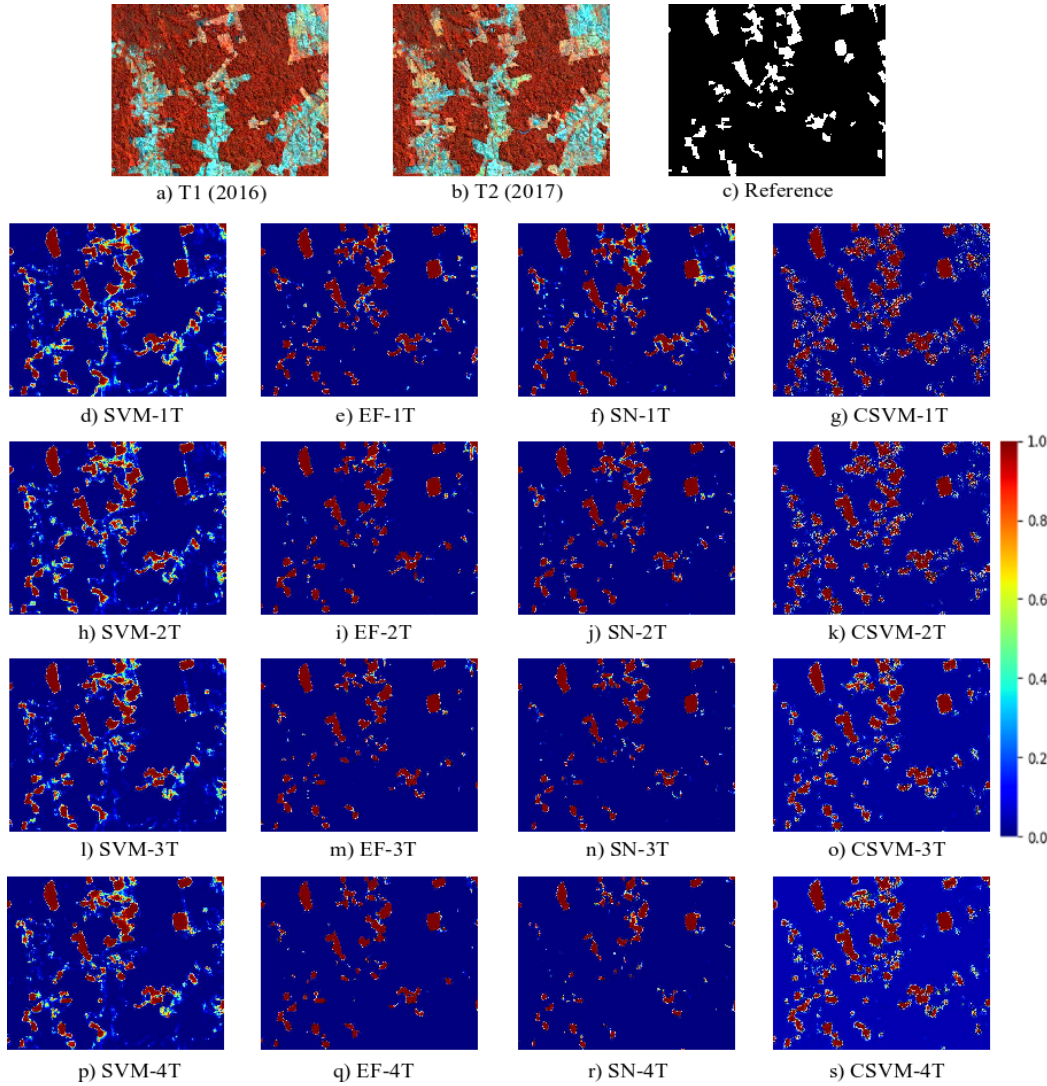


Figure 26: Predicted maps of fourteenth tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.

Next, we evaluate the use of the methods as an alarm scheme. In this scheme, the underlying classifier would indicate areas where deforestation is likely to have occurred. A photointerpreter would then visually analyze on the image, or an inspector could be sent to the indicated areas to check what was real deforestation and what was just false alarm. The main benefit of this procedure would be to restrict the human effort to just a portion of the area being monitored. On the other hand, in this scheme part of the deforested areas could not be detected by the classifier and would go unnoticed. Thus, two metrics are critical in this analysis: first, the proportion of monitored area flagged as potentially deforested and, second, the proportion of total deforestation concentrated in the areas indicated by the classifier. The first metric is defined in Equation 5-6, whereas the second metric is the *Recall* defined in Equation 5-4.

Both metrics will depend on the deforestation probability value assigned by the classifier above which a site should deserve attention. The higher this threshold, the smaller the area under alarm and the smaller the *Recall*. The appropriate threshold value will be determined by the operational demands at each time and each region.

Therefore, the following analysis focuses on the behavior of these two metrics as the deforestation probability threshold varies.

Specifically, we present the curves *Recall* versus *Alert Area* for each method. Each point in the curve corresponds to a threshold imposed on the deforestation probability produced by each tested method. Small area to be observed at a high *Recall*, is the desired profile.

In the first case, using one tile for training, all methods achieved *Recall* values of about 90% when looking only at 10% of the whole imaged area. It means that the 90% of the true deforestation correctly identified are contained in the 10% of the image. Hence, instead of looking at the entire image, the analyst would only have to focus on 10% of it, reducing the human work in 90% .

As expected, as *Recall* increased the area to be observed also increased. But, in this particular case, CSVM (after the third layer) presented the best results, when the *Recall* was between 88% and 96% (see Figure 27-a). However, we didn't manage to record values for CSVM when *Recall* exceeded 96% because the corresponding threshold values in such case were very close to zero, all samples were classified as deforestation and the area to be observed went up to 100% as can be observed in Figure 27-b.

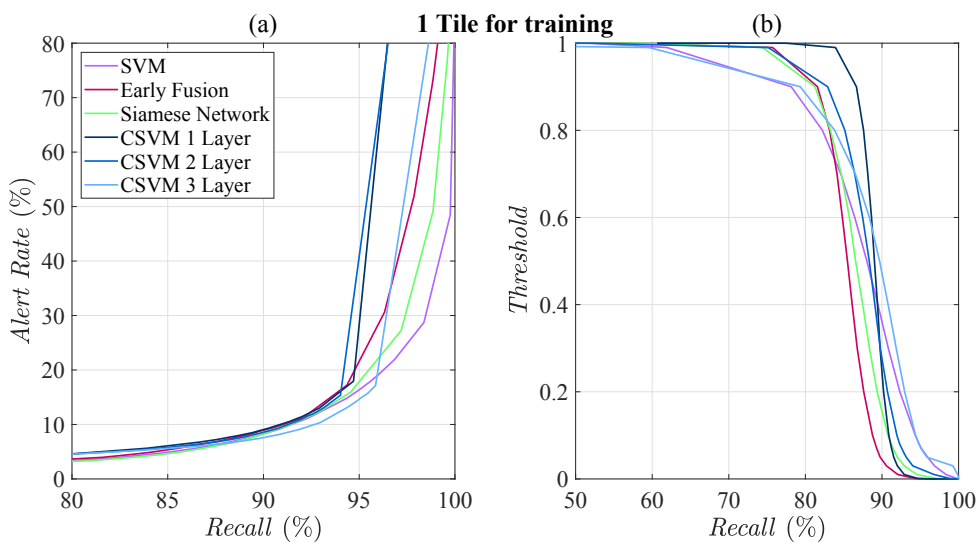


Figure 27: Recall vs Area classified as deforested for Amazon biome using one tile for training.

Using two, three, and four tiles for training, all methods presented a similar profile until 96% *Recall* but, after this value, SVM presented the best performance. It managed to classify more deforested samples correctly, with minimum increase in the area to be observed. Analogous to results for one training tile, when thresholds values are very close to zero, all samples were classified as deforestation and the *Recall* became 100%, as well as the area to be observed.

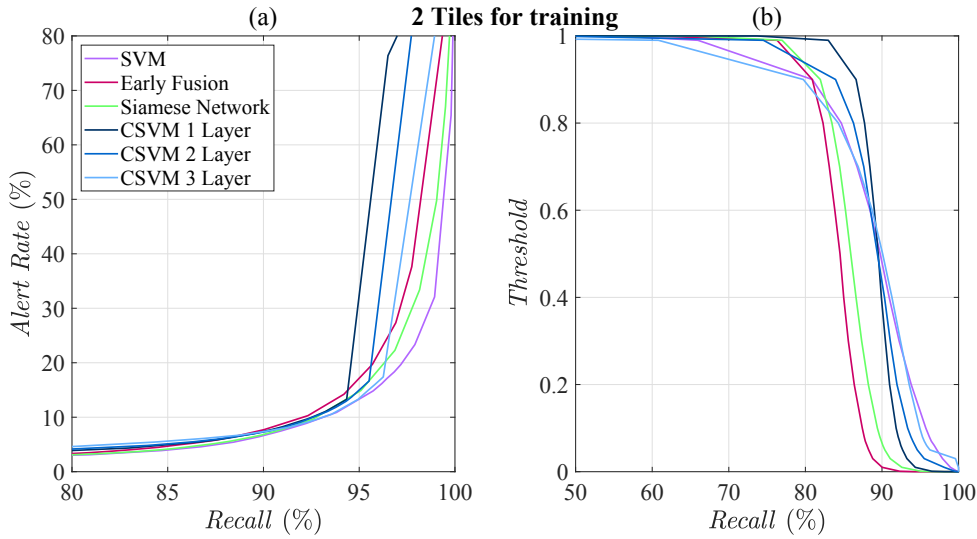


Figure 28: Recall vs Area classified as deforested for Amazon biome using two tiles for training.

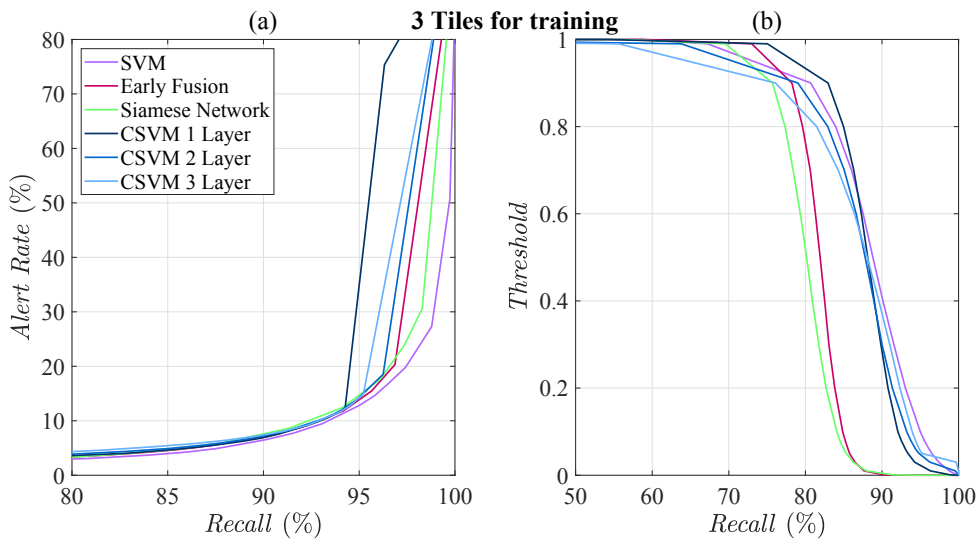


Figure 29: Recall vs Area classified as deforested for Amazon biome using three tiles for training.

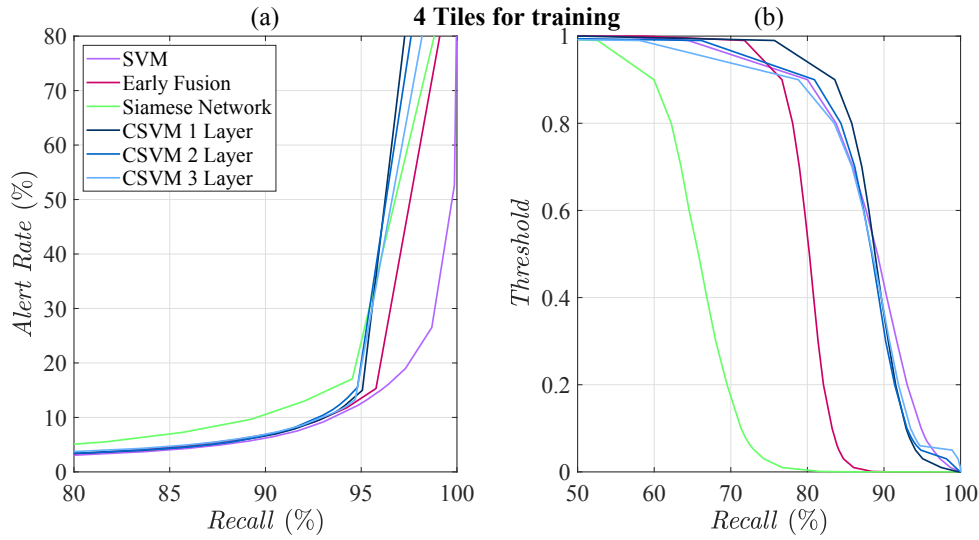


Figure 30: Recall vs Area classified as deforested for Amazon biome using four tiles for training.

5.4.0.2 Cerrado Biome

The results for Cerrado in terms of $F1$ -score and OA is summarized in Figures 31 and 32, respectively. Similar to the Amazon database, EF and SN presented the best performance in all experiments. Using a single tile for training, EF and SN outperformed SVM in 3% and 2% respectively. The best performance achieved by CSVM was 51%, which was obtained after the first layer. However, it did not reach the baseline. Using two tiles for training, EF outperformed SVM with a larger difference, about 7%, and also SN, but with a slight difference of 1%. In this case, SVM outperformed CSVM in 9%. Using three tiles for training, EF and SN outperformed SVM in 5% and 3%, respectively, and CSVM in the second layer came very close to SVM. In the last scenario, using four training tiles, the DL based methods were better than SVM. EF and SN and CSVM (first layer) achieved $F1$ -score of 77% 78% and 76%, overcoming SVM in 2%, 3%, and 1% respectively.

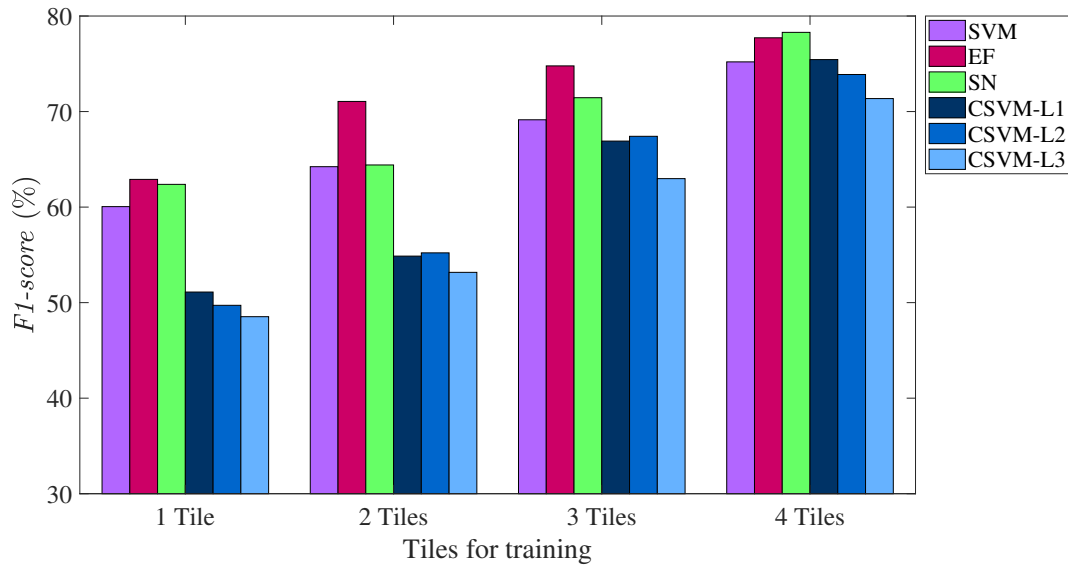


Figure 31: $F1$ -score of Cerrado Biome obtained from SVM, EF, SN, and CSVM using one, two, three, and four tiles for training.

In terms of OA, the results presented a similar trend observed on $F1$ -score. Scores above 90% were achieved in all scenarios. However, EF and SN obtained the best performance in all the experiments. Analogous to the experiments on the Amazon database, CSVM presented lower scores in comparison with other methods. Only in the last case, using four training tiles, CSVM matched SVM at 97%. As in the Amazon experiments, the higher values achieved in term of OA were related to the higher number of correctly classified no-deforestation samples.

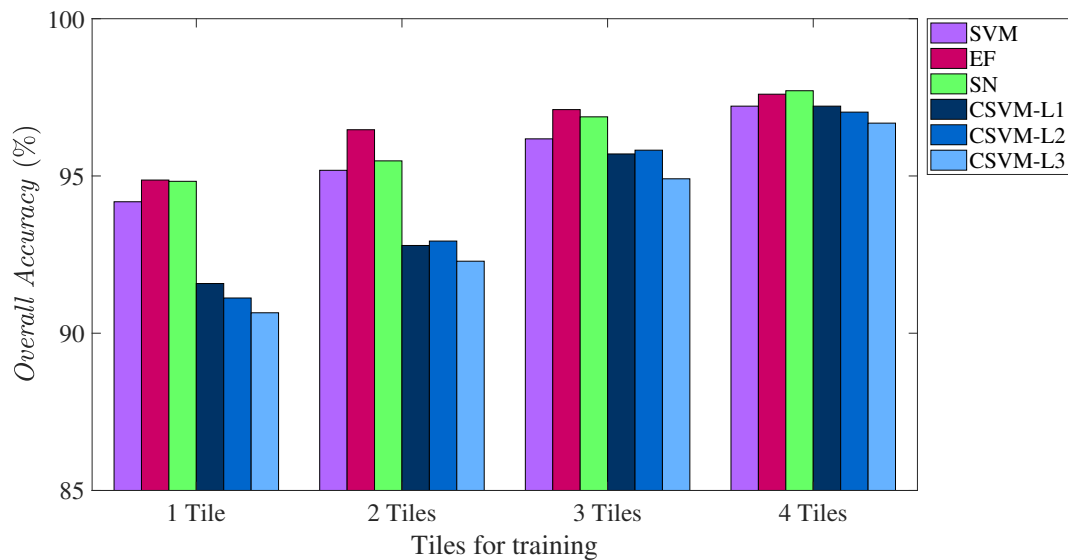


Figure 32: Overall Accuracy of Cerrado Biome obtained from SVM, EF, SN, and CSVM using one, two, three, and four tiles for training.

Figures 33, 34, and 35 show the NIR - G - B composition (Near Infrared,

Green and Blue bands) at both dates (T1 and T2), the reference and the probability maps of the second, sixth and, fourteenth tiles, respectively. Again, columns correspond to methods SVM, EF, SN, and CSVM (after layer 1) and rows correspond to the results one, two, three and four training tiles. Blue represents the lowest probability of deforestation, while Red represents the highest probability.

Like the results recorded on Amazon database, the probability maps improve and the *salt-and-pepper* effect reduces as the number of tiles for training increase. If we observe the first scenario, where a single tile is used for training, all maps present a large number of false positives and a notable *salt-and-pepper* effect. Likewise, EF, SN, and CSVM are more confident, assigning values close to one for pixels of class "deforestation", and values close to zero to pixels of "no-deforestation" class. Contrarily, the probability maps delivered by SVM contain pixels with probability values in the intermediate range.

As observed in the previous experiment series, the probability maps show that all methods are less confident, i.e., present probability values around 50%, close to the borders of the reference polygons. As mentioned before, this is possibly related to inaccuracies in the delimitation of deforestation polygons in the reference.

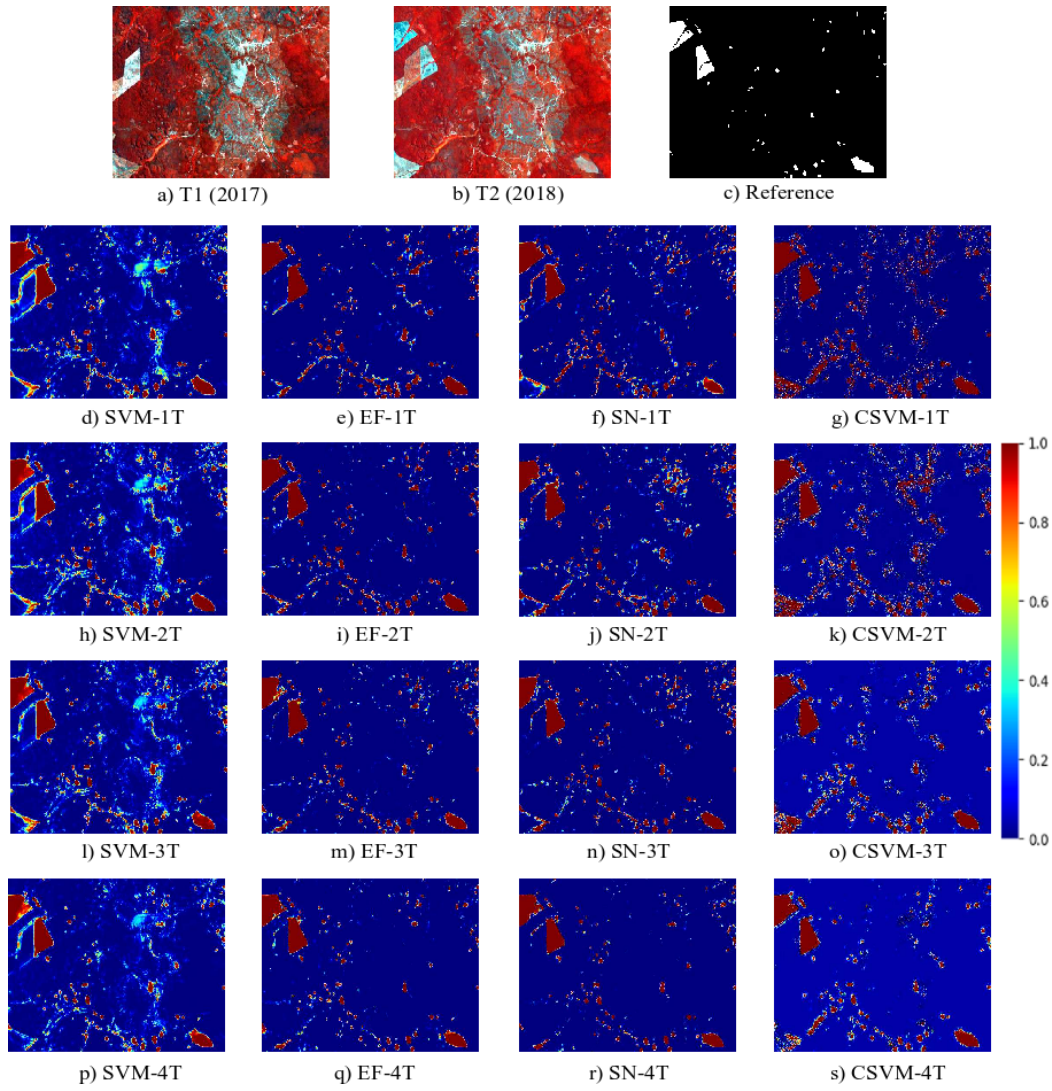


Figure 33: Predicted maps of the second tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.

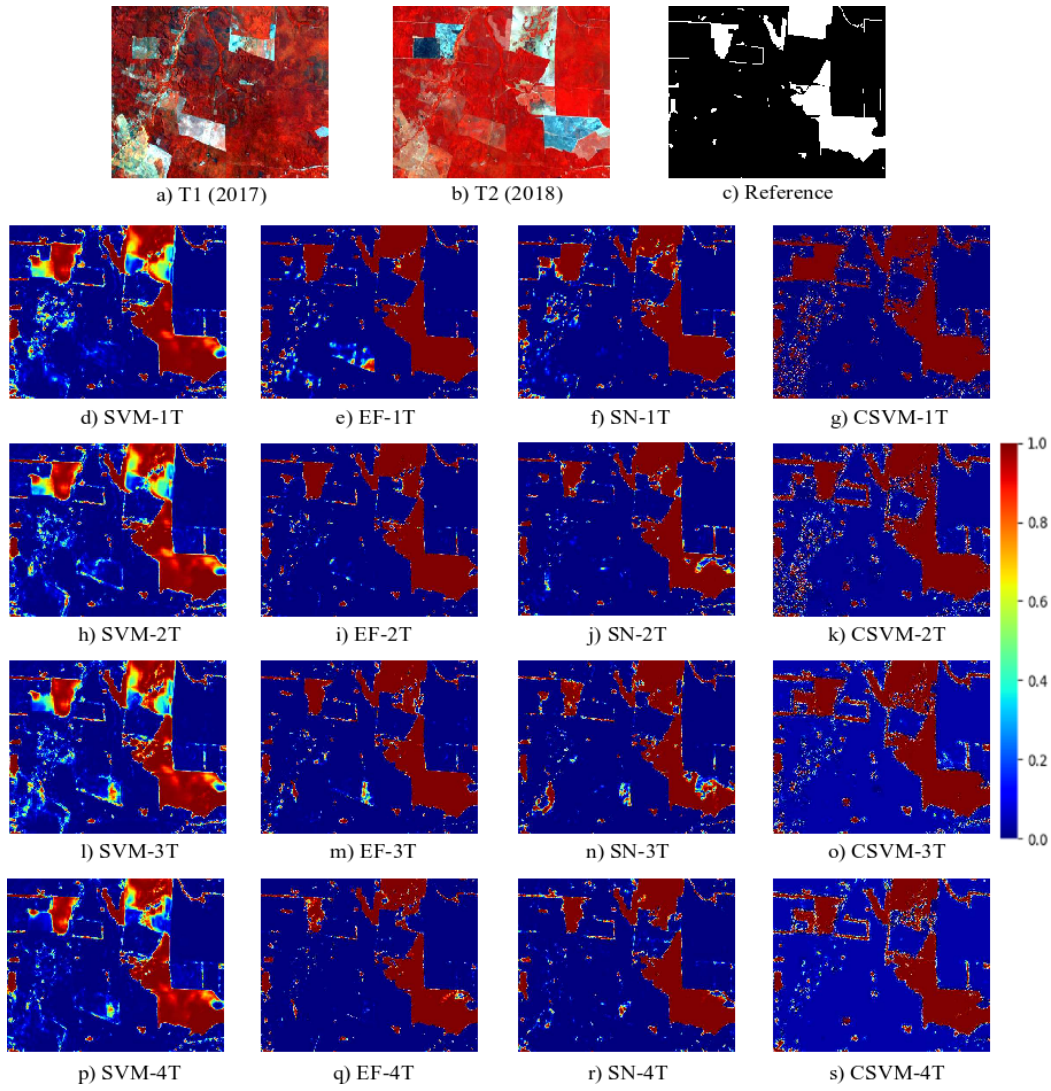


Figure 34: Predicted maps of the eighth tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.

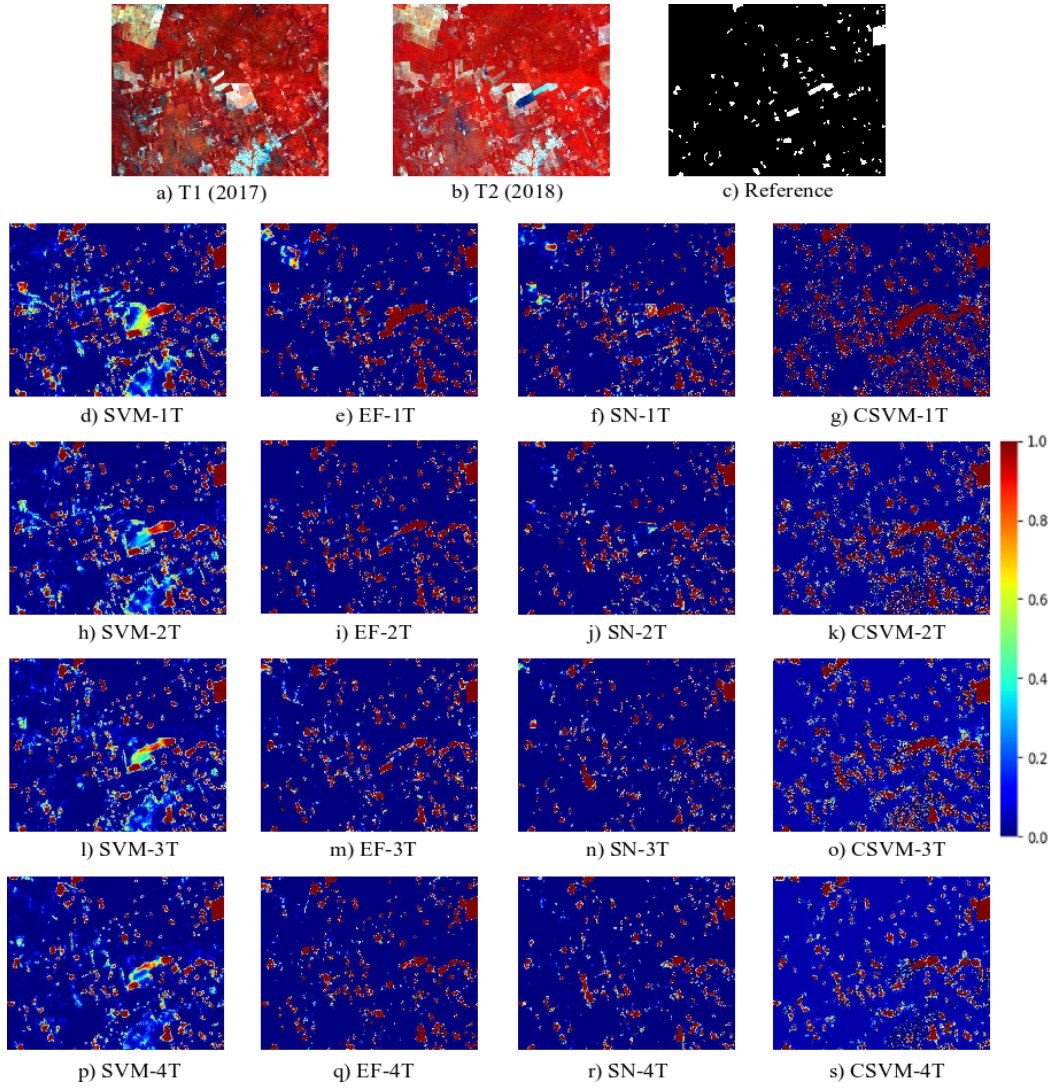


Figure 35: Predicted maps of the eleventh tile obtained from SVM, EF, SN and CSVM using one, two, three, and four tiles for training.

The analysis under the perspective of an alarm system are presented in Figures 36, 37, 38, and 39 for one, two, three and four training tiles, respectively. For this database, in the four scenarios, the best performance was obtained by EF. Although with a single training tile the performance was similar for all methods, EF was slightly superior. Using two tiles for training, EF, also was also the best performing method, followed by CSVM and SVM, which presented very similar results. Finally, using three and four tiles for training, EF and CSVM achieved better results: they correctly classified more samples of deforested class and the area to be observed is lower. According to the graphs at 95% of *Recall*, the area to be observed is reduced to 10% of the entire image (see Figures 36-a, 37-a, 38-a, and 39-a). In the same case of the Amazon database, for threshold values close to zero, all samples are classified as deforestation class, the value of *Recall* is about 100% and the area

to be observed is the entire image, as can be seen in Figures 36-b, 37-b, 38-b, and 39-b.

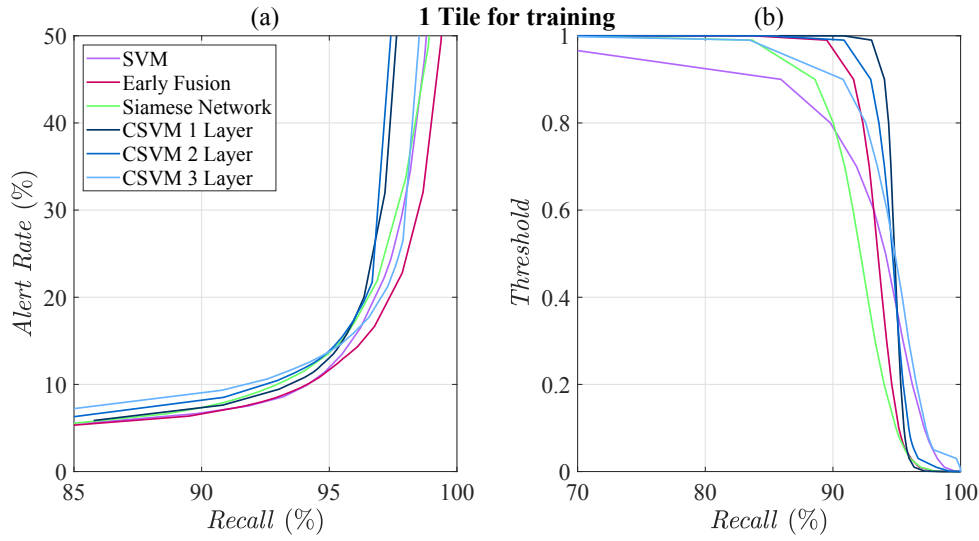


Figure 36: Recall vs Area classified as deforested for Cerrado biome using one tile for training.

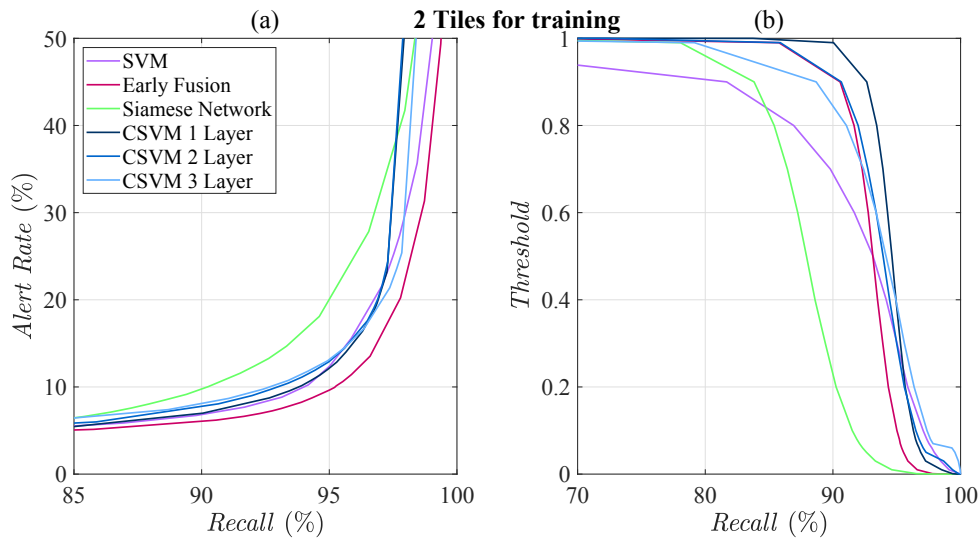


Figure 37: Recall vs Area classified as deforested for Cerrado biome using two tiles for training.

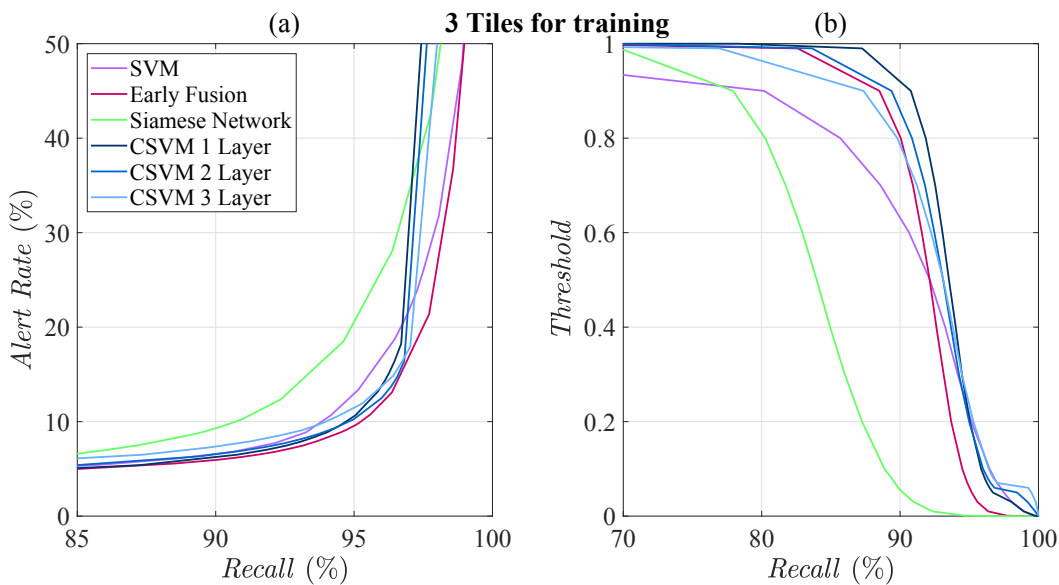


Figure 38: Recall vs Area classified as deforested for Cerrado biome using three tiles for training.

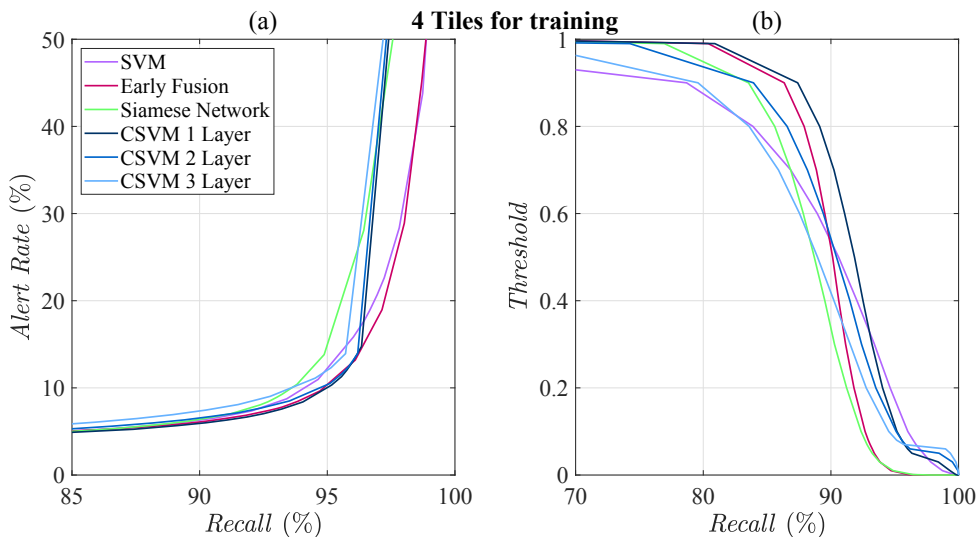


Figure 39: Recall vs Area classified as deforested for Cerrado biome using four tiles for training.

This work reported an evaluation of three state-of-the-art deep learning techniques for deforestation detection: Early Fusion (EF), Siamese Network (SN) and Convolutional SVM (CSVM). Additionally, the performance of these methods was compared against a baseline based on probabilistic Support Vector Machine (SVM), which is one of most popular machine learning techniques for change detection.

Experiments were carried out using two areas of the Brazilian biomes. The first one corresponds to a region of the Amazon biome, and the second one corresponds to the Cerrado biome. The references used in this work were taken from the Amazon Deforestation Monitoring Project (PRODES) developed by the National Institute for Space Research (INPE). The methodology employed to accomplish this task involves significant human intervention. This dissertation aimed at reducing human intervention, assessing state-of-the-art method towards a more automatic deforestation detection.

The experimental analysis relied on two LANDSAT 8/OLI optical images acquired at dates about one year apart from each other. Four different scenarios were considered, using one, two, three, and four tiles training.

EF and SN presented the best performance in most experiments. In few cases, CSVM outperformed SVM. The accuracy obtained by EF and SN in experiments were up to 95% in terms of *Overall Accuracy* (OA) and up to 63% in terms of *F1-score* for Amazon, and up to 97% in terms of OA and 78% in terms of *F1-score* for Cerrado, demonstrating the capacity of the deep learning methods to detect deforestation. As expected, the performance of all methods increased with the number of training samples. This trend was more remarkable for EF and SN .

Besides, the probability maps indicated that EF, SN and CSVM were more confident in their outcomes. Most posterior probabilities delivered by these methods were concentrated close to one and zero, for deforestation and non-deforestation, respectively, whereas the posteriors computed by SVM took intermediate values over comparatively many areas.

The main motivation for including CSVM in this study was the good performance reported in a recent paper under small training sample size.

The experimental analysis did not confirm this expectation. Indeed, in our experiments CSVM was consistently outperformed by EF and SN, and in few cases also by SVM.

Regarding CSVM, some additional experiments for the final classification layer were performed. The first one was the usage of the flattening feature maps obtained after each convolutional layer to train the binary SVM, instead of pooling them over four quadrants and calculate the means. The second experiment involved the selection of the final classifier. We tested a Softmax layer classifier and SVM with an RBF (Radial Basis Function) and a linear kernel. However, the best results were obtained using a linear kernel.

It is worth to mention that despite CSVM did not overcome the baseline in most cases, it presents an advantage concerning EF and SN, it is a CPU-based method, then it does not require GPU to carry out the experiments. GPU is much more costly, and it relies on powerful supplementary equipment to support it.

Although the evaluated methods were tested deforestation detection, they can be easily adapted to other change detection applications. These methods are effective research directions to monitor and control environmental issues that nowadays is of paramount importance.

Future directions

Future works are intended to fine-tune the hyperparameters of the tested methods in order to reduce the false deforestation rate, as well as to evaluate others deep architectures, such as Fully Convolutional Networks (FCN) and Recurrent Neural Networks (RNN).

An additional consideration is to explore other loss functions, for instance, with focus on optimizing a function in terms of *Recall* and area with deforested regions to be observed. This, with the goal of reduce the visual inspection for the analyst to identify the deforested areas with a better accuracy.

Another investigation direction relates to the usage of freely available data of other sensors. An attractive alternative is the Sentinel-2 data that provides better temporal and spatial accuracy than LANDSAT-8. Also, the usage of Synthetic Aperture Radar (SAR) would allow monitoring deforestation independent on weather conditions. Indeed, the Brazilian biomes are covered by clouds most of the year, which prevents the use of optical data. Under these conditions, SAR data becomes an attractive alternative.

A critical issue is still the amount of training samples required by deep

learning based methods to achieve their full potential. Techniques based on domain adaptation seem another promising research direction to mitigate this hindrance.

References

- 1 DE SY, V.; HEROLD, M.; ACHARD, F.; BEUCHLE, R.; CLEVERS, J.; LINDQUIST, E. ; VERCHOT, L.. **Land use patterns and related carbon losses following deforestation in south america.** Environmental Research Letters, 10(12):124004, 2015.
- 2 MACDICKEN, K.; JONSSON, Ö.; PIÑA, L.; MAULO, S.; CONTESSA, V.; ADIKARI, Y.; GARZUGLIA, M.; LINDQUIST, E.; REAMS, G. ; D'ANNUNZIO, R.. **Global forest resources assessment 2015: how are the world's forests changing?** Food and Agricultural Organization of the United Nations (FAO), 2016.
- 3 OF GEOGRAPHY, B. I.; (IBGE), S.. **Mapa de Biomas e de Vegetação.** Available in: <https://ww2.ibge.gov.br/home/presidencia/noticias/21052004biomashtml.shtm>, 2004.
- 4 BUENO, I. T.; ACERBI JÚNIOR, F. W.; SILVEIRA, E. M.; MELLO, J. M.; CARVALHO, L. M.; GOMIDE, L. R.; WITHEY, K. ; SCOLFORO, J. R. S.. **Object-based change detection in the cerrado biome using landsat time series.** Remote Sensing, 11(5):570, 2019.
- 5 GOODMAN, R.; ARAMBURU, M.; GOPALAKRISHNA, T.; PUTZ, F.; GUTIÉRREZ, N.; ALVAREZ, J.; AGUILAR-AMUCHASTEGUI, N. ; ELLIS, P.. **Carbon emissions and potential emissions reductions from low-intensity selective logging in southwestern amazonia.** Forest Ecology and Management, 439:18–27, 2019.
- 6 MALINGREAU, J.; EVA, H. ; DE MIRANDA, E.. **Brazilian amazon: a significant five year drop in deforestation rates but figures are on the rise again.** Ambio, 41(3):309–314, 2012.
- 7 BARRETO, P.; SOUZA JR, C.; NOGUERON, R.; ANDERSON, A.; SALOMÃO, R. ; OTHERS. **Human pressure on the brazilian amazon forests.** World Resources Institute, Washington, DC, 2006.
- 8 FOR SPACE RESEARCH (INPE), N. I.. **Monitoring of the Brazilian Amazonian Forest by Satellite.** Available in: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>, 1988.

- 9 (WWF), W. W. F.. **Amazon deforestation.** Available in: http://wwf.panda.org/our_work/forests/deforestation_fronts/deforestation_in_the_amazon/, 1988.
- 10 OF THE ENVIRONMENT (MMA), M.; OF ENVIRONMENT, B. I. ; (IBAMA), R. N. R.. **Monitoramento do desmatamento nos biomas brasileiros por satélite.** Available in: http://www.mma.gov.br/estruturas/sbf_chm_rbbio/_arquivos/relatoriofinal_cerrado_2010_final_72_1.pdf, 2011.
- 11 BEUCHLE, R.; GRECCHI, R. C.; SHIMABUKURO, Y. E.; SELIGER, R.; EVA, H. D.; SANO, E. ; ACHARD, F.. **Land cover changes in the brazilian cerrado and caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach.** *Applied Geography*, 58:116 – 127, 2015.
- 12 BONANOMI, J.; TORTATO, F. R.; RAPHAEL DE SOUZA, R. G.; PENHA, J. M.; BUENO, A. S. ; PERES, C. A.. **Protecting forests at the expense of native grasslands: Land-use policy encourages open-habitat loss in the brazilian cerrado biome.** *Perspectives in Ecology and Conservation*, 17(1):26–31, 2019.
- 13 SATHLER, D.; ADAMO, S. ; LIMA, E.. **Deforestation and local sustainable development in brazilian legal amazonia: an exploratory analysis.** *Ecology and Society*, 23(2):1–36, 2018.
- 14 LYU, H.; LU, H. ; MOU, L.. **Learning a transferable change rule from a recurrent neural network for land cover change detection.** *Remote Sensing*, 8(6):506, 2016.
- 15 XIAOLU, S.; BO, C.. **Change detection using change vector analysis from landsat tm images in wuhan.** *Procedia Environmental Sciences*, 11:238–244, 2011.
- 16 SINGH, A.. **Review article digital change detection techniques using remotely-sensed data.** *International journal of remote sensing*, 10(6):989–1003, 1989.
- 17 ASOKAN, A.; ANITHA, J.. **Change detection techniques for remote sensing applications: a survey.** *Earth Science Informatics*, 12(2):143–160, 2019.

- 18 JENSEN, J. R.; TOLL, D. L.. **Detecting residential land-use development at the urban fringe.** *Earth Resources and Remote Sensing*, 48:629–643, 1982.
- 19 HOWARTH, P. J.; WICKWARE, G. M.. **Procedures for change detection using landsat digital data.** *International Journal of Remote Sensing*, 2(3):277–291, 1981.
- 20 LUDEKE, A.; MAGGIO, R. C. ; REID, L.. **An analysis of anthropogenic deforestation using logistic regression and gis.** *Journal of Environmental Management*, 31(3):247–259, 1990.
- 21 NACKAERTS, K.; VAESEN, K.; MUYS, B. ; COPPIN, P.. **Comparative performance of a modified change vector analysis in forest change detection.** *International Journal of Remote Sensing*, 26(5):839–852, 2005.
- 22 DENG, J. S.; WANG, K.; DENG, Y. H. ; QI, G. J.. **Pca-based land-use change detection and analysis using multitemporal and multisensor satellite data.** *International Journal of Remote Sensing*, 29(16):4823–4838, 2008.
- 23 HAN, T.; WULDER, M. A.; WHITE, J. C.; COOPS, N. C.; ALVAREZ, M. F. ; BUTSON, C.. **An efficient protocol to process landsat images for change detection with tasselled cap transformation.** *IEEE Geoscience and Remote Sensing Letters*, 4(1):147–151, 2007.
- 24 ZHAN, Y.; FU, K.; YAN, M.; SUN, X.; WANG, H. ; QIU, X.. **Change detection based on deep siamese convolutional network for optical aerial images.** *IEEE Geoscience and Remote Sensing Letters*, 14(10):1845–1849, 2017.
- 25 DHINGRA, S.; KUMAR, D.. **A review of remotely sensed satellite image classification.** *International Journal of Electrical & Computer Engineering* (2088-8708), 9(3), 2019.
- 26 KRANJČIĆ, N.; MEDAK, D.; ŽUPAN, R. ; REZO, M.. **Support vector machine accuracy assessment for extracting green urban areas in towns.** *Remote Sensing*, 11(6):655, 2019.
- 27 GUNN, S. R.; OTHERS. **Support vector machines for classification and regression.** *ISIS technical report*, 14(1):5–16, 1998.
- 28 PAL, M.. **Random forest classifier for remote sensing classification.** *International Journal of Remote Sensing*, 26(1):217–222, 2005.

- 29 MAXWELL, A. E.; WARNER, T. A. ; FANG, F.. **Implementation of machine-learning classification in remote sensing: An applied review.** *International Journal of Remote Sensing*, 39(9):2784–2817, 2018.
- 30 ANYANWU, M. N.; SHIVA, S. G.. **Comparative analysis of serial decision tree classification algorithms.** *International Journal of Computer Science and Security*, 3(3):230–240, 2009.
- 31 VALERIANO, D. M.; MELLO, E. M. K.; MOREIRA, J. C.; SHIMABUKURO, Y. E.; DUARTE, V.; SOUZA, I. M.; DOS SANTOS, J. R.; BARBOSA, C. C. F. ; DE SOUZA, R. C. M.. **Monitoring tropical forest from space: the prodes digital project.** *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 35:272–274, 2004.
- 32 SHIMABUKURO, Y.; DUARTE, V.; ANDERSON, L.; VALERIANO, D.; ARAI, E.; FREITAS, R.; RUDORFF, B. F. ; MOREIRA, M.. **Near real time detection of deforestation in the brazilian amazon using modis imagery.** *Ambiente e Agua-An Interdisciplinary Journal of Applied Science*, 1(1):37–47, 2007.
- 33 DE ALMEIDA, C.; COUTINHO, A. AND ESQUERDO, J.; ADAMI, M.; VENTURIERI, A.; DINIZ, C.; DESSAY, N.; DURIEUX, L. ; GOMES, A.. **High spatial resolution land use and land cover mapping of the brazilian legal amazon in 2008 using landsat-5/tm and modis data.** *Acta Amazonica*, 46(3):291–302, 2016.
- 34 BRITO, A.; OTHERS. **Metodologia da detecção do desmatamento no bioma cerrado. mapeamento de áreas antropizadas com imagens de média resolução espacial.** São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2018.
- 35 SOUZA, C.; AZEVEDO, T.. **Mapbiomas general handbook.** MapBiomas: São Paulo, Brazil, p. 1–23, 2017.
- 36 MACHADO, R. B.; RAMOS NETO, M. B.; PEREIRA, P. G. P.; CALDAS, E. F.; GONÇALVES, D. A.; SANTOS, N. S.; TABOR, K. ; STEININGER, M.. **Estimativa de perda da área do cerrado brasileiro.** Relatório técnico não publicado. Conservação Internacional, Brasília, DF., p. 1–25, 2004.
- 37 ZAGORUYKO, S.; KOMODAKIS, N.. **Learning to compare image patches via convolutional neural networks.** In: *PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION*, p. 4353–4361, 2015.

- 38 DAUDT, R. C.; LE SAUX, B.; BOULCH, A. ; GOUSSEAU, Y.. Urban change detection for multispectral earth observation using convolutional neural networks. In: IGARSS 2018-2018 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, p. 2115–2118. IEEE, 2018.
- 39 ZHANG, Z.; VOSSelman, G.; GERKE, M.; TUIA, D. ; YANG, M. Y.. Change detection between multimodal remote sensing data using siamese cnn. arXiv preprint arXiv:1807.09562, 2018.
- 40 MOU, L.; BRUZZONE, L. ; ZHU, X.. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 57(2):924–935, 2019.
- 41 LILLESAND, T.; KIEFER, R. W. ; CHIPMAN, J.. Remote sensing and image interpretation. John Wiley & Sons, 2015.
- 42 GOODFELLOW, I.; BENGIO, Y. ; COURVILLE, A.. Deep Learning. MIT Press, 2016. <http://www.deeplearningbook.org>.
- 43 LECUN, Y.; BOTTOU, L.; BENGIO, Y. ; HAFFNER, P.. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, Nov 1998.
- 44 LI, F.-F.; KARPATY, A. ; JOHNSON, J.. CS231n: Convolutional Neural Networks for Visual Recognition 2016. Available in: <http://cs231n.stanford.edu/>, 2016.
- 45 ACHANCCARAY, P. M.. Crop Recognition in Tropical Regions based on spatio-temporal Conditional Random Fields from multi-temporal and multi-resolution sequences of remote sensing images (Doctoral thesis). Pontifical Catholic University of Rio de Janeiro Rio de Janeiro, Brazil, 2019.
- 46 IOFFE, S.; SZEGEDY, C.. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167, 2015.
- 47 SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I. ; SALAKHUTDINOV, R.. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15:1929–1958, 2014.

- 48 BROMLEY, J.; GUYON, I.; LECUN, Y.; SÄCKINGER, E. ; SHAH, R..
Signature verification using a" siamese" time delay neural network. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, p. 737–744, 1994.
- 49 BAZI, Y.; MELGANI, F..
Convolutional svm networks for object detection in uav imagery. IEEE transactions on geoscience and remote sensing, 56(6):3107–3118, 2018.
- 50 CASSEB, A. D. R.; CHIANG, J. O.; MARTINS, L. C.; SILVA, S. P. D.; HENRIQUES, D. F.; CASSEB, L. M. N. ; VASCONCELOS, P. F. D. C..
Alphavirus serosurvey in domestic herbivores in pará state, brazilian amazon. Revista Pan-Amazônica de Saúde, 3(4):43–48, 2012.
- 51 STEINWEG, T.; GERARD, R. ; THOUMI, G..
Cargill: Zero-deforestation approach leaves room for land clearing inbrazil's maranhão. Chain Reaction Research. Available in: <https://chainreactionresearch.com/wp-content/uploads/2018/04/Cargill-report-April-2018.pdf>, p. 1–18, 2018.
- 52 FAN, R.-E.; CHANG, K.-W.; HSIEH, C.-J.; WANG, X.-R. ; LIN, C.-J..
Liblinear: A library for large linear classification. Journal of machine learning research, 9(Aug):1871–1874, 2008.