Pontifícia Universidade Católica
do Rio de Janeiro

Jose David Bermudez Castro

SYNTHESIS OF MULTISPECTRAL OPTICAL IMAGES FROM SAR/OPTICAL MULTITEMPORAL DATA USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

Tese de Doutorado

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

Advisor : Prof. Raul Queiroz Feitosa
Co-advisor: Dr. Patrick Nigri Happ

Rio de Janeiro
April 2019

Pontifícia Universidade Católica
do Rio de Janeiro

Jose David Bermudez Castro

SYNTHESIS OF MULTISPECTRAL OPTICAL IMAGES FROM SAR/OPTICAL MULTITEMPORAL DATA USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS

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

Prof. Raul Queiroz Feitosa
Advisor
Departamento de Engenharia Elétrica – PUC-Rio

Dr. Patrick Nigri Happ
Co-advisor
Departamento de Engenharia Elétrica – PUC-Rio

Dr. Thales Sehn Körting
Instituto Nacional de Pesquisas Espaciais – INPE

Prof. Hemerson Pistori
Universidade Católica Dom Bosco – UCDB

Dr. Leonardo Alfredo Forero Mendoza
Universidade do Estado do Rio de Janeiro – UERJ

Prof. Wouter Caarls
Departamento de Engenharia Elétrica – PUC-Rio

Rio de Janeiro, April the 25th, 2019

Jose David Bermudez Castro

The author received his engineering degree in Electronic Engineering at the Universidad Del Norte (UniNorte) in 2009. Obtained his master's degree in Electrical Engineering with emphasis on Signal Processing and Control, at the Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) in 2015. Since then, he has worked in the field of Digital Image Processing, Remote Sensing and Machine Learning.

Bibliographic data

Bermúdez Castro, José David

SYNTHESIS OF MULTISPECTRAL OPTICAL IMAGES FROM SAR/OPTICAL MULTITEMPORAL DATA USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS / Jose David Bermudez Castro; advisor: Raul Queiroz Feitosa; co-advisor: Patrick Nigri Happ. – Rio de janeiro: PUC-Rio, Departamento de Engenharia Elétrica, 2019.

v., 76 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Aprendizado Profundo;. 3. Redes Adversárias Generativas;. 4. Reconhecimento de Culturas Agrícolas;. 5. Detecção de Queimadas;. 6. Remoção de Nuvens. I. Feitosa, R. Q.. II. Happ, P. N.. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Acknowledgments

I am truly grateful to my advisor, Prof. Raul Queiroz Feitosa, for the encouragement, his advice, stimulating talks and generous support throughout the development of my thesis.

I am truly grateful to my co-advisor, Dr. Patrick Happ, for the encouragement, his advice, stimulating talks and generous support throughout the development of my thesis.

I thank my parents, Gilberto and Leicy, my Grandparents, Hermenegildo and Carmen, and my sister Maira 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 the companionship and valuable scientific discussion.

I also gratefully acknowledge the financial support of CNPq and NVIDIA's Academic Programs Team for donating a Titan Xp GPU to support my research.

Abstract

Bermúdez Castro, José David; Feitosa, R. Q. (Advisor); Happ, P. N. (Co-Advisor). **SYNTHESIS OF MULTISPECTRAL OPTICAL IMAGES FROM SAR/OPTICAL MULTITEMPORAL DATA USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS**. Rio de Janeiro, 2019. 76p. Tese de doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Optical images from Earth Observation are often affected by the presence of clouds. In order to reduce these effects, different reconstruction techniques have been proposed in recent years. A common alternative is to explore data from active sensors, such as Synthetic Aperture Radar (SAR), as they are nearly independent on atmospheric conditions and solar lighting. On the other hand, SAR images are more difficult to interpret than optical images, requiring specific treatment. Recently, conditional Generative Adversarial Networks (cGANs) have been widely used to learn mapping functions that relate data of different domains. This work proposes a method based on cGANs to synthesize optical data from data of other sources: data of multiple sensors, multitemporal data and data at multiple resolutions. The working hypothesis is that the quality of the generated images benefits from the number of data used as conditioning variables for cGAN. The proposed solution was evaluated in two databases. As conditioning data we used co-registered data from SAR at one or two dates produced by the Sentinel 1 sensor, and optical images produced by the Sentinel 2 and LANDSAT satellite series, respectively. The experimental results demonstrated that the proposed solution is able to synthesize realistic optical data. The quality of the synthesized images was measured in two ways: firstly, based on the classification accuracy of the generated images and, secondly, on the spectral similarity of the synthesized images with reference images. The experiments confirmed the hypothesis that the proposed method tends to produce better results as we explore more conditioning data for the cGANs.

Keywords

Deep Learning; Conditional Generative Adversarial Networks; Remote Sensing; Crop Recognition; Wildfire Detection; Cloud Removal

Resumo

Bermúdez Castro, José David; Feitosa, R. Q.; Happ, P. N.. **SINTETIZAÇÃO DE IMAGENS ÓTICAS MULTIESPECTRAIS A PARTIR DE DADOS SAR/ÓTICOS USANDO REDES GENERATIVAS ADVERSARIAS CONDICIONAIS**. Rio de Janeiro, 2019. 76p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Imagens óticas são frequentemente afetadas pela presença de nuvens. Com o objetivo de reduzir esses efeitos, diferentes técnicas de reconstrução foram propostas nos últimos anos. Uma alternativa comum é explorar dados de sensores ativos, como Radar de Abertura Sintética (SAR), dado que são pouco dependentes das condições atmosféricas e da iluminação solar. Por outro lado, as imagens SAR são mais difíceis de interpretar do que as imagens óticas, exigindo um tratamento específico. Recentemente, as Redes Adversárias Generativas Condicionais (cGANs - Conditional Generative Adversarial Networks) têm sido amplamente utilizadas para aprender funções de mapeamento que relaciona dados de diferentes domínios. Este trabalho, propõe um método baseado em cGANSs para sintetizar dados óticos a partir de dados de outras fontes, incluindo dados de múltiplos sensores, dados multitemporais e dados em múltiplas resoluções. A hipótese desse trabalho é que a qualidade das imagens geradas se beneficia do número de dados utilizados como variáveis condicionantes para a cGAN. A solução proposta foi avaliada em duas bases de dados. Foram utilizadas como variáveis condicionantes dados corregistrados SAR, de uma ou duas datas produzidos pelo sensor Sentinel 1, e dados óticos de sensores da série Sentinel 2 e LANDSAT, respectivamente. Os resultados coletados dos experimentos demonstraram que a solução proposta é capaz de sintetizar dados óticos realistas. A qualidade das imagens sintetizadas foi medida de duas formas: primeiramente, com base na acurácia da classificação das imagens geradas e, em segundo lugar, medindo-se a similaridade espectral das imagens sintetizadas com imagens de referência. Os experimentos confirmaram a hipótese de que o método proposto tende a produzir melhores resultados à medida que se exploram mais variáveis condicionantes para a cGAN.

Palavras-chave

Aprendizado Profundo; Redes Adversárias Generativas; Reconhecimento de Culturas Agrícolas; Detecção de Queimadas; Remoção de Nuvens

Table of contents

List of figures

List of tables

1
INTRODUCTION

Earth Observation using Remote Sensing (RS) technology has become a cost-effective solution for many applications due to the possibility of accessing free satellite imagery with higher spatial resolution and lower revisiting time. It allows gathering information suitable for modeling many environmental processes that exhibit complex spatiotemporal dynamic relationships, such as cropland or deforestation monitoring. However, for fully exploiting the available Earth Observation data, it is required to develop methods capable of dealing with different temporal and spatial resolutions, different sensors and incomplete and noisy data due to adverse atmospheric conditions (e.g., cloud coverage) or sensors malfunctioning.

For instance, malfunctioning of sensors like MODIS or Landsat have been limiting their usage and exploitation for different applications [11]. Aqua MODIS presents on the 1.6-μm channel (band 6) 15 of the 20 detectors either nonfunctional or noisy. This has been a serious problem for Aqua MODIS snow products for instance, which use band 6 primarily for snow detection [12]. The scan line corrector (SLC) of Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor has failed permanently since 2003, inhibiting the retrieval or scanning of 22% of the pixels in each Landsat 7 SLC-off image. This failure has seriously limited the scientific applications and usability of ETM+ data [13].

Cloud coverage represents another critical problem for many remote sensing applications. First, because it cannot be avoided since it results from a natural phenomenon we cannot control. Second, because it may affect optical imagery, which is generally preferred for remote sensing applications, leading to the corruption or missing of the data. But clouds not only represents a problem, the shadow caused by these further stretch the affected area. For example, Figure 1.1 shows an image acquired at the municipality of *Campo Verde*, Mato Grosso state, Brazil affected by clouds. It observed how the affected area is extended because of the shadows caused by clouds. Cloud coverage is especially critical for those applications that demand multitemporal data due to it is required the observation of the environmental processes at regular temporal intervals for modeling the problem precisely. Thus, incomplete data on some dates might dramatically compromise the outcome.

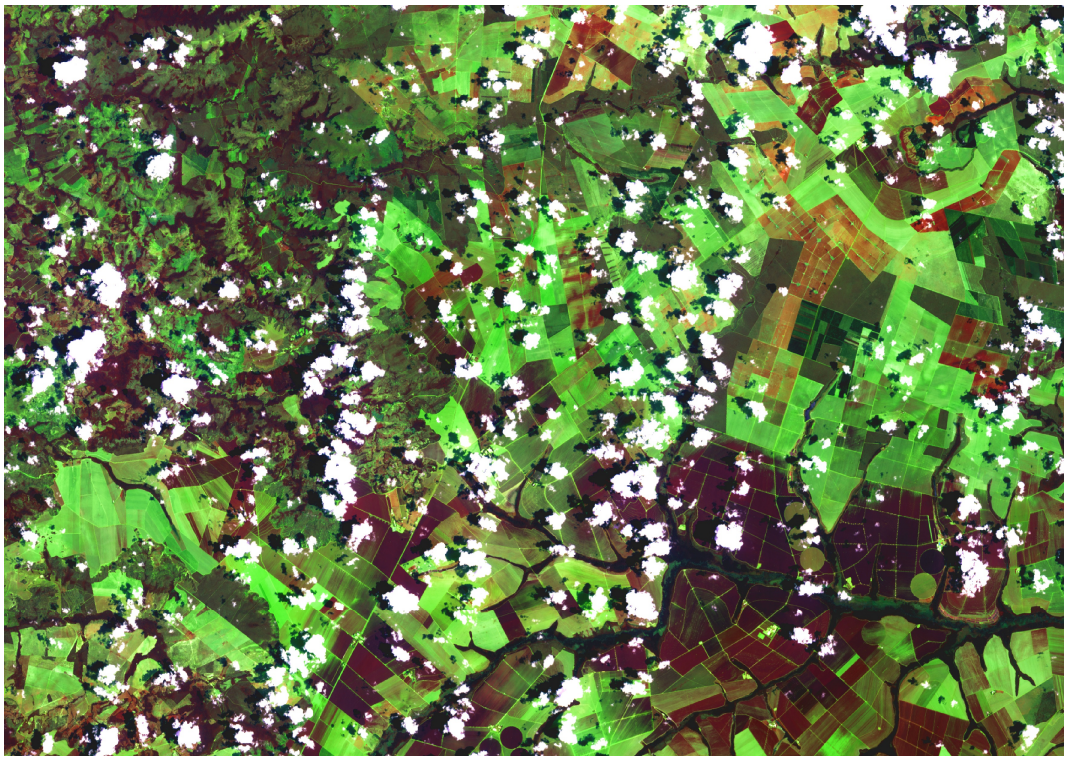

Figure 1.1: Example of an Optical Remote Sensing image corrupted by clouds and shadows. Image acquired at municipality of *Campo Verde*, Mato Grosso, Brazil.

On average, approximately 35% of the global land surface is obscured by clouds, where the major concentration is located in the tropical and sub-tropical regions [14]. For instance, Brazil, which plays an important role in agricultural production worldwide [15], the cloud coverage strongly limits the usefulness of optical imagery for identifying, mapping and monitoring croplands. [16] reports a research to assess cloud cover conditions over four states in the tropical and sub-tropical Center-South region of Brazil. Results showed high seasonality of cloud occurrence within the crop year. In particular, the states close to the Equator line (North) presented the lowest averaged values (15%) of clear sky occurrence during the main crop period (November to February), while in South, the clear sky statistic was around 45%. In these traditional agricultural regions, where approximately 45% of Brazilian agriculture is concentrated, optical satellite data faces severe constraints for mapping summer crops.

Likewise, cloud coverage also represents one of the main difficulties for the design of more accurate methods for monitoring wildfires in the Amazon rainforest [17]. In addition to image corruption, clouds present a similar spectral response to the burned areas, which produces detection/classification

errors. The occurrences of wildfires generate large socioeconomic and ecological impacts [18]. It reduces the forest biomass due to mortality of large trees [19], changes in tree species composition [20], soil impoverishment, loss of biodiversity and economic losses [21], as well as indirect effects such as worsening air quality, which affects human health worldwide [22].

An alternative to optical sensors to circumvent the cloud coverage problem is the usage of active sensors, like Synthetic Aperture Radar (SAR), which is almost independent of atmospheric conditions neither and solar illumination [23]. In spite of these advantages over passive sensors, the usage of SAR data is still challenging because it is less descriptive and more complex to interpret than optical images. Accordingly, a plethora of reconstruction techniques [3, 24–27] has been proposed and used in an attempt to remove the presence of clouds in optical imagery. However, there is still no method able to completely solve this problem.

In recent years, Deep Learning (DL) techniques have become the state-of-art of many machine learning applications [28]. Essentially, DL's paradigm consists of learning meaningful representation from data instead of the traditional manual feature engineering. For RS image analysis, DL techniques have demonstrated better performance than traditional methods based on handcrafted features [29]. The main difficulty for RS adaptation relies on the lack of labeled samples required for optimizing the large set of parameters that defines a DL model.

Currently, Generative Adversarial Networks (GANs) [30] have attracted the attention of the machine community due to their capability for capturing and representing complex probability distributions from data. In some applications, GANs have been used to synthesize missing data or translating data among different domains. For instance, an application based on GANs was developed in [31] for synthesizing photorealistic images of a given input semantic layout. The model can transform simple paint designs of a segmented land scene into highly realistic scenes. Also, in [32] GANs were used for creating art.

For RS, GANs have also been exploited in some applications. In [33], the authors synthesized SAR data using GANs to perform data augmentation of less representative classes of the available training set. Results showed improvements in the classification model after including those synthesized samples for training the classifier. In this particular case, the GANs were trained using the label information as a conditional variable for generating samples of those classes. In the context of recovering missing remote sensed optical data, some works based on GANs have also been exploited. In [34], a

method was proposed that relies on cyclic consistent GANs to clean cloudy images. However, the solution is limited to thin clouds. In [1], the authors proposed a conditional GAN (cGAN)-based algorithm to recover visible RGB image components to exploit the Near Infrared (NIR) data. Again, this method is limited to thin clouds and relies on the NIR, which can only partially penetrate clouds. A later work [35] overcomes part of these shortcomings by exploiting SAR instead of NIR data, but this method is also restricted to thin clouds.

In a recent work [5], we proposed a cGAN-based method that overcomes the aforementioned limitations. Similar approaches were almost simultaneously published in [36], [37] and [38]. Basically, a corresponding multispectral optical image is generated by a cGAN from its SAR counterpart, even in the presence of thick clouds, using cloud-free image patches of the same region for training. However, this method presented difficulties for capturing the high spatial data variability commonly exhibited in some RS applications like the agricultural ones, for instance. These results indicate that using only a SAR image as conditional data could not be enough for synthesizing optical images close to real ones.

Based on that, in this thesis we propose an extension to these methods to synthesize multispectral optical images. Basically, the framework involves taking the SAR counterpart and a SAR-optical pair from the same area in another date as conditioning data for the cGAN. We hypothesize that more realistic outcomes can be produced using a cGAN by exploiting the temporal relations in the optic domain in addition to the SAR-optical correspondence. Our main contribution is a new cGAN-based framework capable to synthesize multispectral optical data in regions where such data is not available, due to thick cloud coverage or any other cause, from coregistered SAR-optical image pairs in different dates.

Besides, the method handles data from satellite sensors of different spatial resolutions using learnable up/downsampling interpolation coefficients. Finally, the quality of the synthesized images is evaluated in terms of similarity metrics, as well as in terms of semantic segmentation performance. Since this is the final goal in many applications, we further propose to use the pixel-wise classification accuracy as an additional quality metric for the synthesized images.

1.1 Objectives

The objectives of this work are the following:

– **General objective:**
Propose a new method for synthesizing remote sensing optical data from multisensor and multitemporal data using conditional generative adversarial networks (cGANs).

– **Specific objectives:**

1. Propose a method to synthesize multispectral optical images from SAR data of the same region acquired in the same date (multisensor synthesis).
2. Propose an extension of the previous method to also take advantage of optical images from the same region acquired at a different date (multitemporal synthesis).
3. Propose an extension of the previous methods to consider data of the same region from SAR images acquired in the same date and optical and SAR images acquired in a different date (multisensor and multitemporal synthesis).
4. Evaluate the proposed methods on critical multitemporal applications.

1.2 Contributions and Novelties

The main contributions of this work are the following:

1. A method able to estimate missing optical data from SAR data.
2. A method able to recover missing optical data from optical and/or SAR data from another date.
3. A cGANs architecture able to work with data of different sensors and spatial resolutions.

1.3 Organization of the remainder parts

Chapter 2 describes the related works available in the literature for cloud removal and synthesis of optical imagery for RS applications.

Chapter 3 provides the theory and fundamental concepts of relevant subjects addressed in this thesis for a better understanding of the proposed method.

Chapter 4 introduces and explains the proposed method for synthesizing remote sensing optical imagery.

Chapter 5 presents the experiments conducted to evaluate the proposed method, including the datasets, the used features, the cGAN architecture, the experimental protocol and the experimental results.

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

2
RELATED WORKS

This chapter summarizes the works related to this research that have been proposed so far. Most of them were developed in the context of cloud removal, since it is one of the main missing data problem in optical imagery.

According to [24], cloud removal techniques can be categorized into spectral-based, spatial-based, temporal-based and hybrid methods.

2.0.1
Spectral-based

These techniques use the multispectral bands' information from the affected image in order to recover the regions covered by clouds. For example, in [39, 40] the authors consider that clouds are mostly composed of spectral low-frequency components and can, in theory, be removed via high-pass filtering. Nevertheless, discovering the optimal cut-off frequency to separate clouds is usually difficult and done empirically. Furthermore, the filtering process also affects the spectral information of cloud-free regions. Because of that, this technique is usually employed to remove thin clouds.

In [27], the authors take into account that the Landsat 8 cirrus band (band 9) provides detection of high-altitude cloud contamination that may not be visible in other spectral bands. Based on that, the authors developed a method for cirrus cloud contamination correction. They used linear regression for estimating a relationship between a visible or infrared band and the cirrus band using data from a homogeneous land cover area, i.e., a region characterized by similar pixel intensity values if it is not contaminated by clouds. Then, the estimated relationship is used to remove the cirrus clouds. The main difficulty of the method is to automatically identify homogeneous regions from cirrus contaminated data. So, it usually requires prior knowledge about the imaged scene for manually selecting these locations.

2.0.2
Spatial-based

The spatial-based methods use the cloud-free local neighboring information for recovering missing data. For instance, [41] proposes the use of inpaint-

ing techniques [42] for filling corrupted regions in RS images. The authors assumed that the affected area shares some statistics with its neighborhood. The basic idea consists of copying the patches from the neighborhood and duplicating them to fill the affected area. The approach first fills the patches positioned on the boundaries. Then, this procedure is repeated iteratively until the whole corrupted area is filled. Nevertheless, the accuracy of the method decreases as the cloudy area increases due to the number of patches that have to be replaced. Additionally, it is necessary to correctly identify the boundaries of the affected regions. In summary, the inpainting image does not retrieve the missing data, but synthesizes plausible data for the affected locations if the corrupted area is small and similar to its neighborhood.

2.0.3 Temporal-based

On the other hand, temporal-based approaches use data from other co-registered cloud-free images acquired at different dates to interpolate the missing information. These approaches have the capacity of dealing with both thin and thick clouds. The simplest approaches are based on image replacement, which consists in replacing the pixels affected by clouds by the pixels located at the same position of another image of the same sensor [25, 43]. Then, a post-processing step reduces the spectral differences among the pixels of the different images. However, depending on the dynamic of the problem, differences in spectral information can be too high to be corrected during this post-processing step. More elaborated approaches use a multitemporal co-registered image sequence in order to build a time series model to infer pixels covered by clouds [44, 45]. The main problem in this approach is to acquire enough cloud-free images in different dates. Additionally, these methods presents problems for dealing with cloud shadows.

2.0.4 Hybrid methods

Lastly, hybrid methods combine different data sources (spectral, spatial and temporal) to design more robust solutions. In [46], a hybrid technique was proposed to model data for recovering data gaps for snow covering estimates. The method uses a Terra/Aqua MODIS time series and a Hidden Markov Random Field framework for integrating MODIS spectral, spatial and temporal contextual information together with an environmental association such as surface topography. The critical issue in this method is the size of the spatiotemporal neighborhood over which spatiotemporal interaction potentials

are defined. [47] employs a time series of Landsat images for filling small and large area gaps in corrupted Landsat data. To do that, the authors proposed an algorithm that uses the spectral-angle-mapper (SAM) [48] similarity metric in both spatial and temporal domain, denoted as SAMSTS. Specifically, for each corrupted pixel, the algorithm searches for an alternative similar pixel located in a non-affected region of the image. The similar pixel locations are determined by the following pipeline. First, they stack the image time series across the spectral band dimension, and the resulting image tensor is then segmented across the time and the spatial domains. After that, the generated segments are clusterized. So, given a specific image of the time series contaminated by clouds, for each cloudy segment, an alternative cloud-free segment is determined based on the cluster information. Finally, the affected pixels of cloudy-segments are replaced by similar ones from cloud-free segments. The method, however, has limitations when rapid changes occur, e.g., due to agricultural harvesting or flooding.

Another strategy is the usage of Synthetic Aperture Radar (SAR) images as auxiliary data source for recovering data of optical remote sensing images. A simple procedure was proposed in [26] which consists of replacing the missing pixels by pixels from the cloud-free regions using an interpolation function. This interpolation function employs the SAR image (from the same acquired date of the corrupted image) as the base for replacing the affected pixels. For each cloudy pixel, the method first looks for a pixel location in the cloud-free image region, where the SAR data are similar. Then, the co-registered optical pixel is selected for replacing the cloudy one. However, differences in the spatial resolutions and speckle noise hardly affect this approach.

Recently, a set of new methods based on the Generative Adversarial Networks (GANs) paradigm has been proposed. GANs were firstly introduced in [30] and have been widely investigated since then by the computer vision community. More recently, conditional Generative Adversarial Networks (cGANs) [49] have been broadly used in different image generation tasks, such as image inpainting [50], image manipulation [51], and image translation [52]. For image translation, for instance, a cGAN learns a nonlinear mapping function capable to transform an image from one domain (say X) to another (say Y). Based on that, [1] employs a cGAN to recover visible light RGB images from multispectral RGB-NIR images. Basically, they assume that the NIR band is almost not affected by thin clouds, and can be used to synthesize cloud-free RGB images. Firstly, they created a set of synthesized cloudy images by using a cloud synthesizing algorithm over a set of multispectral cloud-free images. Each synthesized cloudy RGB-NIR image has its associated ground truth

cloud-free RGB image. Then, a cGAN model is trained to map from images covered by clouds to the corresponding cloud-free counterpart as illustrated in Figure 2.1. However, this approach can not deal with thick clouds and white objects, which are in appearance similar to clouds. This occurs because in practice the NIR band is also sensible to clouds, especially thick clouds. Moreover, the mentioned work does not report any numerical quality assessment and is limited to a mere subjective evaluation.

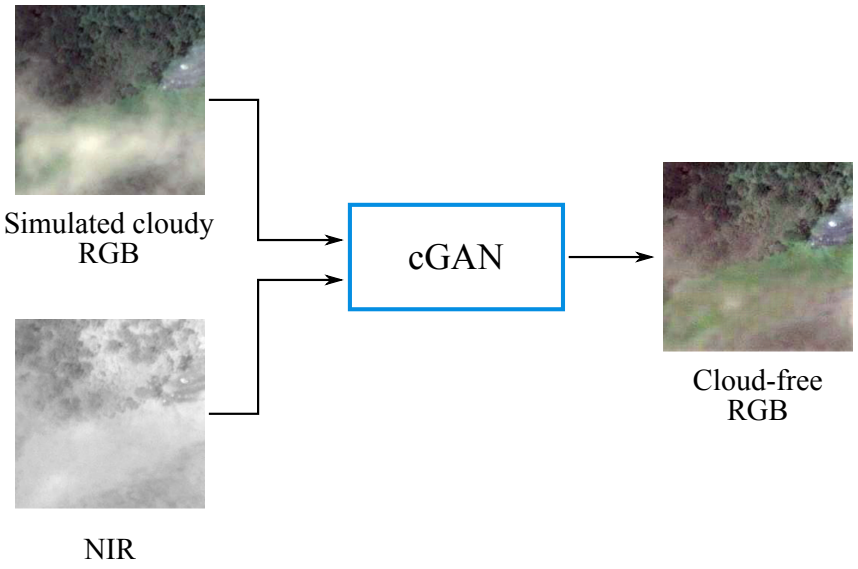

Figure 2.1: Illustration of the approach proposed by [1]. A cGAN model is trained for mapping from simulated RGB-NIR cloudy images to the corresponding cloud-free RGB.

Instead of using cGANs, [2] proposed the use of Cycle-GANs for mapping cloud-free RGB images from cloudy ones. A Cycle-GAN is an extension of the cGANs concept in which, instead of just learning to map from X (cloud-free images) to Y (cloudy images) domain, it also learns to map back from Y to X domain (See Figure 2.2). For some applications, Cycle-GANs perform better than analogous cGANs. The advantage of this approach is that Cycle-GANs allow the use of cloudy - cloud-free pair of images that are not co-registered for training the model. Contrary to [1], this method does not use an algorithm for simulating the cloudy images over the cloud-free. Instead, it builds a dataset by selecting patches from real cloudy and cloud-free optical images. Cycle-GANs are trained for mapping from the cloudy domain to the cloud-free domain and vice-versa. However, the method does not use any cloud-free auxiliary data for conditioning. Thus, the model may synthesize samples that may significantly diverge from a real image. Experiments showed that the model fails in the presence of thick clouds and also for large affected areas. Besides, the fact that this method considers only the RGB spectral bands limits its use.

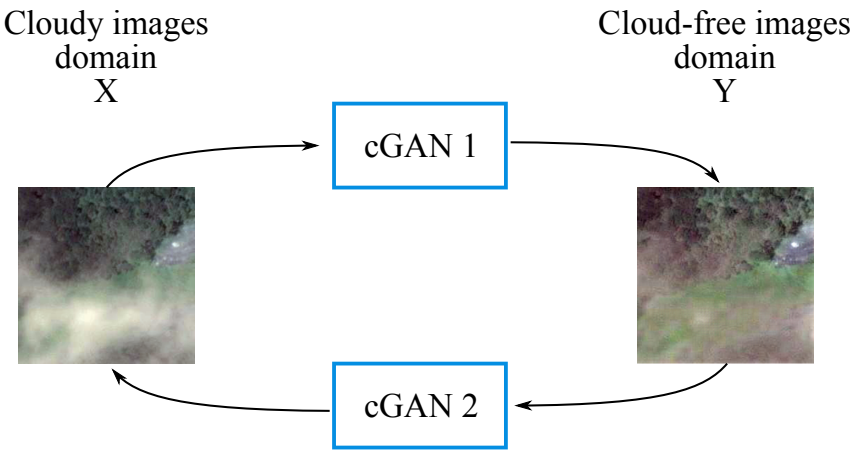

Figure 2.2: Illustration of the approach proposed by [2]. Two cGAN models are trained: one for mapping from RGB cloudy images to the cloud-free RGB images and the other one to map from cloud-free RGB image to RGB cloudy images.

A later work [3] overcomes part of those shortcomings by exploiting SAR as auxiliary data instead of NIR data for recovering missing information in Sentinel-2 optical imagery (See Figure 2.3). The authors also extend the method for multispectral optical images with more than 3 bands. However, the method is also restricted to thin clouds. Similar to [1], they use a cloud synthesizing algorithm over a set of multispectral cloud-free optical images for generating the corresponding set of cloudy images. After that, they stack on each synthesized cloudy image, a co-registered SAR image at approximately the same acquisition date. Finally, the cGANs are trained to map from these cloudy-SAR images to the associated cloud-free version. The method depends on how realistic are the synthesized clouds. In real scenarios the method may not perform well.

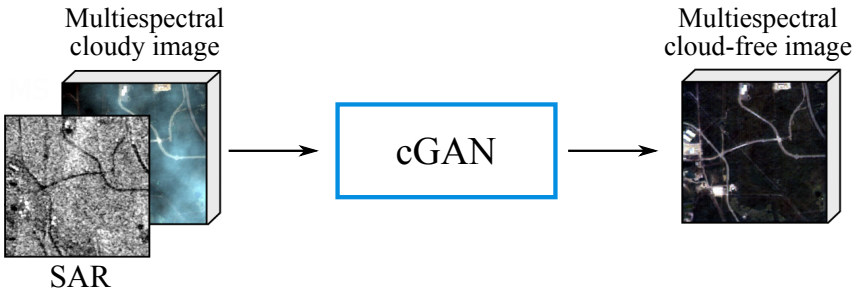

Figure 2.3: Illustration of the approach proposed by [3]. A cGAN model is trained for mapping from simulated Multiespectral-SAR cloudy images to the corresponding cloud-free Multispectral optical.

In [5] and [4] cGANs and SAR data are also used for recovering cloudy

optical images. Instead of using an algorithm for synthesizing clouds, they train a cGAN to learn a direct non-linear mapping function through the generator network, which inputs SAR data and outputs a corresponding plausible optical data (See Figure 2.4). This function is learned by employing co-registered SAR/Optical pairs of patches from cloud-free regions and then, the learned model is used for synthesizing a cloud-free optical image. In [5], the method was evaluated in terms of image classification for agricultural applications, while in [4] it was assessed using similarity performance metrics. Even though the proposed method in [5] outperforms the classification rates of the baseline (the direct classification of the SAR image using GLCM features), the performance of the method is still poor in comparison with the classification of the real optical images (the cloud-free optical images). These results indicate that the cGANs model did not capture the complete spatial data variability of the scene, commonly observed in agricultural applications. In these cases, using only a SAR image as conditional data could not be enough for synthesizing optical images close to real ones.

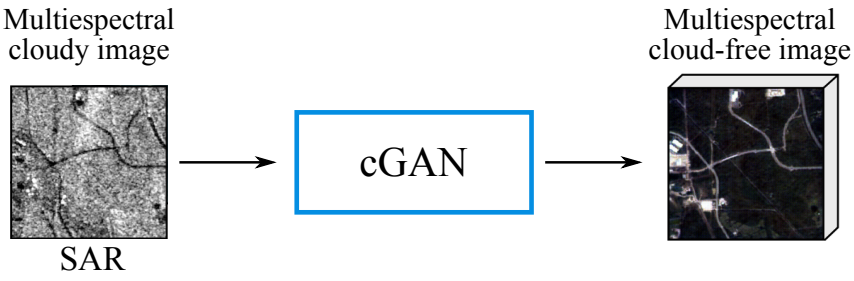

Figure 2.4: Illustration of the approach proposed by [4, 5]. A cGAN model is trained for mapping from SAR images to the corresponding cloud-free Multispectral optical.

To close this review, we quote [37], a recent work developed parallel to ours that contains some of the ideas contained in our proposal. Basically, it uses the adversarial training for pretraining a Fully Convolutional Network (FCN) for semantic segmentation of SAR imagery. First, similar to [5] and [4], the authors propose training a cGAN model to synthesize optical data from SAR. The so trained generator is modified to perform semantic segmentation of SAR imagery. Specifically, they attach a *softmax* classification layer at the end of the generator network. Next, by using a small set of labeled samples, the weights of the last three layers are fine-tuned. Basically, they use the adversarial training as a method for initializing the weights of the FCN in an unsupervised fashion and then the weights of the final layers are improved via fine-tuning. The experimental results showed that the proposed approach performed better than training the classifier from scratch. In addition, they

assessed the stability of the approach in terms of the number of labeled samples used for the supervised part, showing that the performance method is almost not affected in comparison with the models trained from scratch. However, the authors report difficulties to deal with agricultural fields information due to significant differences in appearance between the two evaluated SAR images in those areas. These discrepancies in appearance are related to the differences in acquisition dates of the SAR images (one month of difference). During this period, the vegetation appearance changed affecting the performance of the classifier.

3
FUNDAMENTALS

This chapter aims to present the theoretical foundations for a proper understanding of the proposed method. First, a brief introduction to passive and active sensors for acquiring RS imagery is given. Second, basic concepts associated with Convolutional Neural Networks (CNN) are presented. Finally, the fundamentals of Generative Adversarial Networks are introduced.

3.1
Remote Sensing

Remote Sensing (RS) is "the art, science, and technology of observing an object, scene, or phenomenon by instrument-based techniques without physical contact with the target of interest" [7]. For example, a conventional camera can be considered as a remote sensing instrument because a photo taken with these devices does not involve direct contact with the object. In contrast, an accelerometer, for instance, cannot be regarded as a RS instrument because the sensor must be in contact with the phenomenon to be able to measure it.

Different RS platforms can be used for assisting in inventorying, mapping and monitoring Earth resources. They can be operated from airborne (aircraft, helicopters, and unmanned aerial vehicles (UAVs)) and from space (satellites and space shuttles). Each of them offer particular characteristics that can be advantageous for some applications [53]. Table 3.1 summarizes the main characteristics and advantages of each of these technologies.

RS sensors present a broad variety of spatial, temporal and spectral resolutions. But primarily, they can be categorized into two types: *passive* and *active* sensors. Passive sensors use external energy sources that illuminate the objects, e.g., sunlight, whereas active ones have their own energy source. Figure 3.1 illustrates the physics of how these sensors work.

Table 3.1: Advantages and disadvantages of main platforms for remote sensing data collection. Related costs to each platform are based on [10].

Platform	Characteristics
Satellite	**Advantages:** • Access to some free images. • Clear and stable images. • Large area within each image. • Good historical data. **Disadvantages:** • High cost for high spatial resolution images. • Clouds may hide ground features. • Fixed schedule. • Data may not be collected at critical times.
Aircraft	**Advantages:** • Relative flexible availability. • Relatively high spatial resolution. • Changeable sensors **Disadvantages:** • High cost. • Availability depends on weather condition.
UAV	**Advantages:** • Flexible availability. • Relative low cost. • High and ultra high spatial resolutions. • Changeable sensors. **Disadvantages:** • Relative unstable platform can create blurred images. • Geographic distortion. • May require certification to operate.

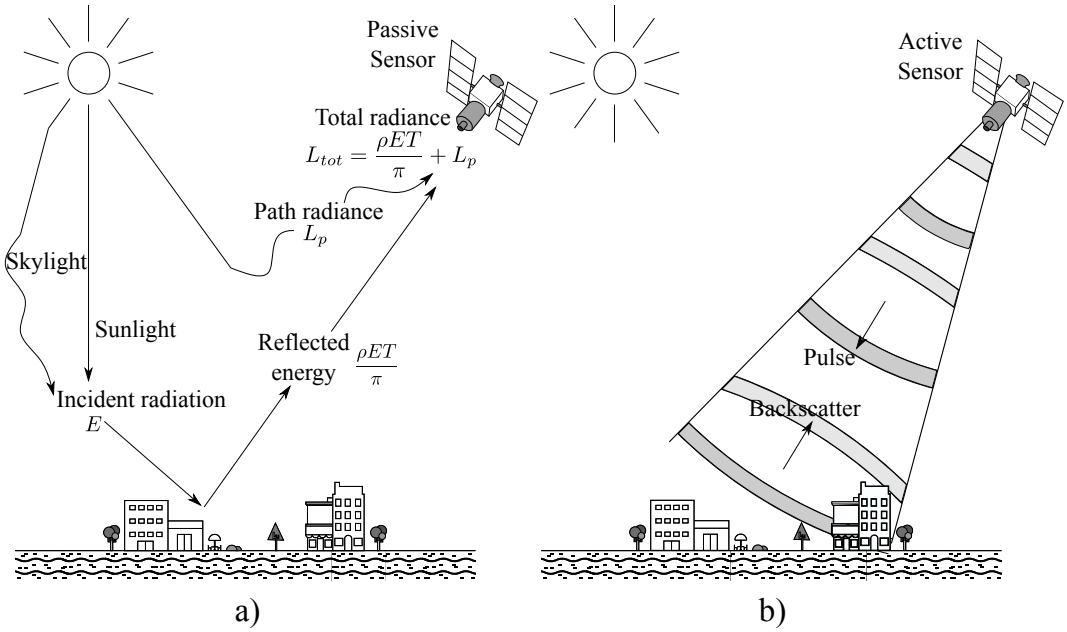

Figure 3.1: Passive vs Active sensors. a) A passive sensor uses external energy sources. b) An active sensor uses its own source of energy. Adapted from [6].

3.1.1 Passive Sensors

Passive sensors measure the energy present emitted or reflected by the scene or the object being monitored. The main source of natural energy measured by passive sensors is the reflected radiation of sunlight. As illustrated in Figure 3.1, the total energy measured by the sensor L_{tot} is composed by the reflected energy (a fraction of the incident radiation E by the sunlight and skylight) and the path radiance L_p (reflected by the atmosphere). The reflected energy is computed as $\rho ET/\pi$, where ρ represents the reflectance of an object and T is the atmospheric transmittance [6]. Other examples are the thermal infrared and passive microwave sensors, which measure natural Earth energy emissions [54].

The radiation is an electromagnetic wave characterized by its wavelength. Not all objects reflect the same wavelength, which depends on the nature of each material [54]. Passive sensors are engineered so that they can measure radiation at different wavelengths. As a matter of fact, they can measure the human visible spectrum light [390-700 nm], infrared [750-1 nm], ultraviolet [100-400 nm] and other other wavelengths [7] . Figure 3.2 illustrates the organization of the electromagnetic spectrum in terms of wavelength ranges.

An example of RS passive sensor based technology is optical imagery, which normally has multiple band sensors specialized in measuring wavelengths within specific ranges. For instance, a multispectral optical image usually

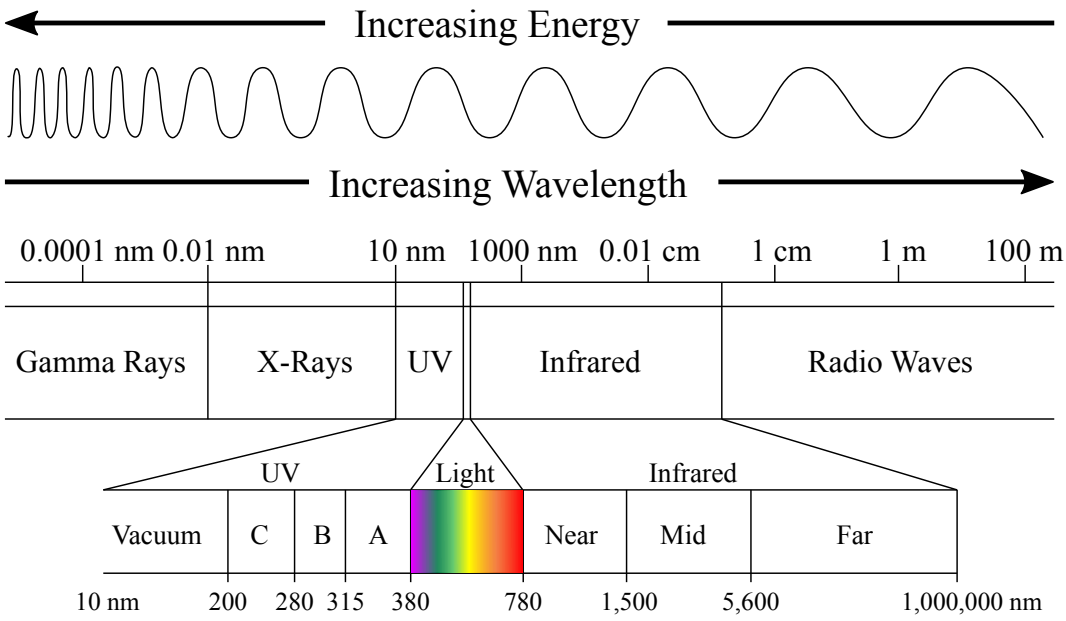

Figure 3.2: Electromagnetic spectrum. Adapted from [7].

has from 3 to 12 bands, while a hyperspectral imaging technology can have hundreds.

The major drawback of optical imagery is that it is affected by the weather conditions. For instance, cloud coverage can disrupt image acquisition. The sunlight is reflected by the clouds and blocked by cloud-shadows. This is a common problem that limits Earth Observation with passive sensors.

3.1.2 Active Sensors

Active Sensors illuminate the object to be observed using their own energy source [7]. Specifically, they emit the radiation toward the target and measure the intensity of waves that are backscattered. Examples of active sensors are the Synthetic Aperture Radar (SAR) and the Light Detection and Ranging (LiDAR). In this work, we focus on SAR due to its capability of producing suitable resolution images as optical sensors, as well as the accessibility to imagery from free satellite platforms. In contrast, LiDAR produces a cloud of points related to the distance between the instrument and the target, which is more suitable for getting information about the elevation of the study area.

This type of sensors has the advantage of being almost independent on the weather and daylight conditions. They operate on the microwave spectrum range, can penetrate the atmosphere and are not backscattered by the clouds, smoke, light rain and snow.

Active Sensors can also offer affordable revisit times in comparison with passive ones.That ability to operate independently on weather conditions is critical in regions like the Amazon Rainforest, where the cloud coverage represents a hindrance to the use of optical imagery. On the other hand, SAR imagery is generally difficult to interpret visually. Thus, for many applications, specifically those focused on vegetation, classification accuracy is usually lower for SAR images than for optical ones.

In addition to the intensity of reflected radiation, the phase information of the electromagnetic waves can also be exploited for performing interferometric analysis. For instance, it can be used for the measurement of small displacements in structures [55], estimates crop height [56], etc.

In short, passive and active sensors present different advantages and disadvantages. Therefore, depending on the application, one of them may be the best source of information. In other cases, the fusion of both technologies may provide the best result.

3.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs or ConvNets) are a family of neural networks specialized in processing data that exhibits a 2D grid-like topology [57]. CNNs are inspired in the animal visual cortex operation where each *neuron/kernel* is focused in processing information associated to a restricted location of the visual field known as the *receptive field*. The entire visual field is covered by the arrange of *receptive fields* of different neurons that partially overlap.

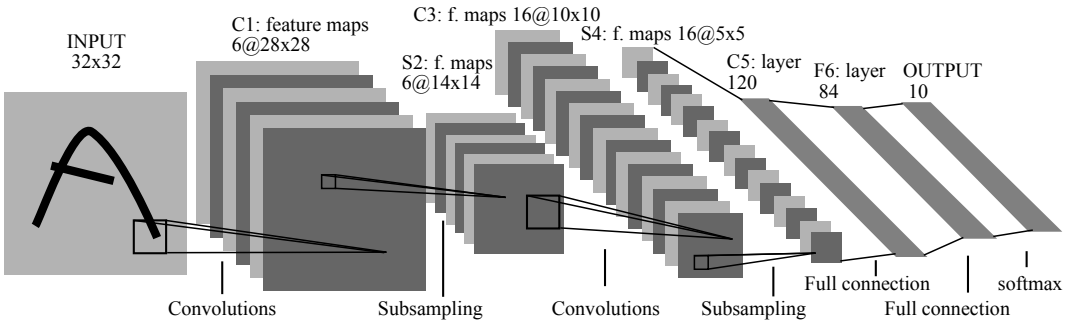

Figure 3.3: LeNet-5 architecture. First proposed CNN's architecture for handwritten and machine-printed character recognition in 1990's. Adapted from [8]

3.2.1 Layers of Convolutional Networks

Figure 3.3 illustrates the first CNN architecture proposed by [8], known as LeNet-5. It consists of seven layers: the *Input* layer, two sets of consecutive *Convolutional* and *Pooling* layers, two successive *Fully-Connected* layers, and an *Output* layer. More complex CNN architectures usually stack many convolutional and pooling layers. The following is a non-exhaustive description of the layers commonly used on the current CNN.

1. **Input**: it is a 3D tensor (multidimensional array of data), arranged as ($width(w)$, $height(h)$, $depth(d)$), which holds the raw pixel values of the evaluated image. The *width* and *height* refer to the spatial dimension of image while the *depth* to the numbers of channels.

2. **Convolutional**: it inputs a $3D$ tensor of data and outputs a $3D$ tensor of neuron's activation, whose channels are known as *feature maps*. It performs the *image convolution* operation between the input data and the *kernels*, followed by a non-linear activation function. Specifically, each kernel is a $k \times k \times d$ multidimensional array of parameters (weights) that slides over the input data, computing the Frobenius inner product at each image location. The values of these parameters are optimized by supervised training. The result is a 3D tensor which dimension depends on the spatial dimension of the input, the stride size, the padding size, and the number of selected kernels.

3. **Batch Normalization**: It is a trainable normalization layer used to improve training convergence [58]. The key idea is to shift and scale the values of a feature map distribution before it is evaluated by the non-linear activation function at each training batch. In simple words, after the convolution, the resulting feature maps are normalized by subtracting the batch mean and dividing it by the batch standard deviation. This process helps to prevent the network from stuck in points of the parameter space where the gradient of activation functions are equal to zero, implying in no parameters' updated.

   However, this normalization might not be beneficial for all cases. It depends on the data, the network architecture and the batch size. So, batch normalization lets the optimization algorithm decides when this process should be applied and the degree of the normalization. Therefore, batch normalization incorporates two trainable parameters to each layer: the gamma parameter, which multiplies the normalized output to control

the scale, and the beta parameter which controls the shift position of the normalized data.

4. **Activation function**: After the convolution or the Batch Normalization, if it was considered, the resultant feature map goes through a non-linear activation function. Figure 3.4 illustrates the most common non-linear activation functions, *sigmoid*, *tanh*, *Rectified Linear Unit (ReLU)*, *Leaky-ReLU* and *exponential LU*.

5. **Pooling**: it produces a summary over a neighborhood around each pixel. Usually, pooling is performed with a stride greater than one, which implies in downsampling the feature maps on the spatial domain (depth remains unchanged) to reduce the number of parameters of subsequent layers, increase the *receptive field*, reduce the computation cost, and also control overfitting. The most common variant is the max-pooling, which consists of filtering the feature maps by applying a 2×2 filter with a stride of 2×2, replacing all values inside the filter by the maximum value among them.

6. **Fully-Connected** (FC): it inputs a $3D$ or a $1D$ tensor and outputs an 1D vector of neuron activations. Likewise regular neural networks, each neuron of this layer is connected to all neurons of the previous layer.

7. **Dropout**: Dropout is a regularization technique for addressing the problem of overfitting in deep neural networks with a large number of parameters [9]. The method consists in randomly dropping neural units from the neural network, as well as their connections, during the training phase. In other words, the set of network parameters to be optimized change randomly at each iteration. This process emulates training a large ensemble of models that share their parameters. Figure 3.5 shows an example of the application of dropout to a neural network with two hidden layers. The resulting network is a "thinned" network in which the number of neural units to be dropped for each layer is specified by the corresponding parameter p_{layer}. Each p_{layer} indicates the probability of a neural unit to be dropped.

   At testing time, the dropout p_{layer} is set to zero and all connections are restored. In order to estimate the contributions of all "thinned" possible trained models, an approximation is made by scaling-down the weights of the network. Specifically, the weights are multiplied for the corresponding parameter p_{layer} defined for each neural unit.

8. **Output**: it is another FC layer, which inputs the feature vector of the previous FC layer and computes the class scores. The result is a vector of dimension equal to the number of classes, where each position corresponds to a class score. Usually, the scores are normalized through a *Softmax* activation function to get the probability distribution for the different classes.

3.2.2 Fully Convolutional Networks

Fully Convolutional Networks (FCN) are a set of CNNs initially designed for the task of pixel-wise dense prediction. As illustrated in Figure 3.6, it inputs an image and outputs another one with the same spatial dimensions. Depending on the application, the output corresponds to a label image (semantic segmentation), a reconstruction of the input image (convolutional autoencoders) or a version from other domain of the input (image translation).

Traditional CNNs involve multiple max-pooling operations which reduce sucessively the spatial resolution of the feature maps. In order to produce a pixel-wise dense prediction, the FCNs introduce the deconvolution layers to recover the original spatial resolution reduced by earlier downsamplings layers. In practice, deconvolution is implemented as the transposed convolution operator. As shown in Figure 3.6, the number of deconvolution layers are usually equal to the number of downsampling operation.

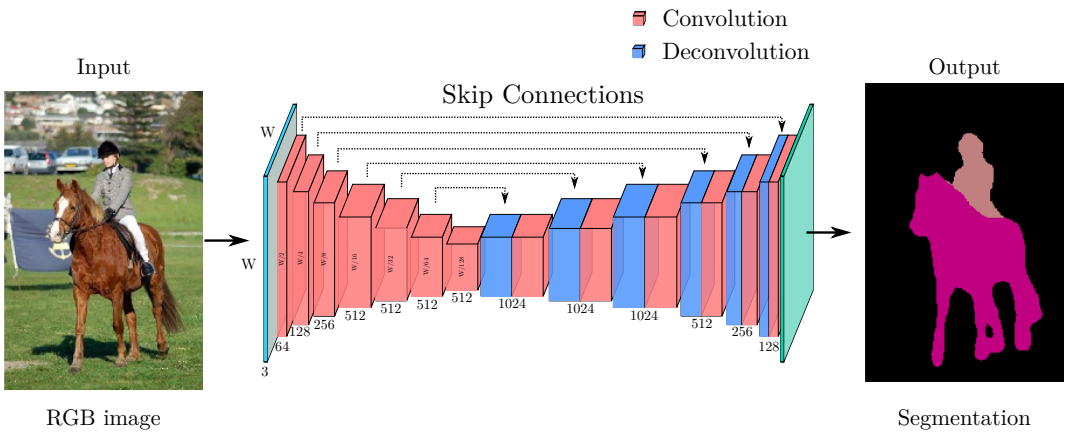

Figure 3.6: Example of a FCN architecture. The network inputs a RGB image and outputs its corresponding segmentation.

Skip Connections: It concatenates the feature map outputs from the deconvolution layers with the corresponding ones from the downsampling stage

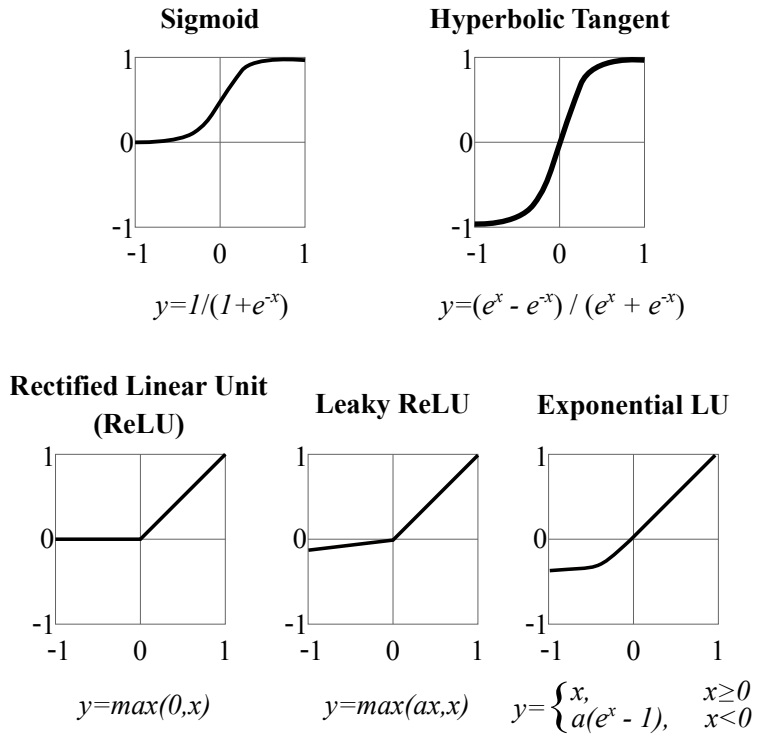

Figure 3.4: Illustration of non-linear activation functions.

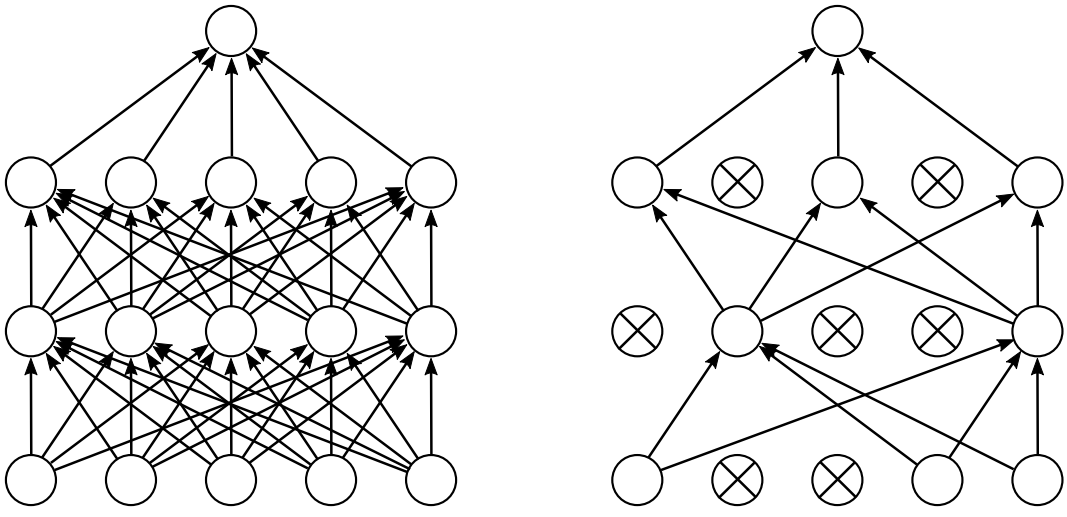

Figure 3.5: Dropout Neural Network Model. (a) A standard neural network with 2 hidden layers. (b): An example of a "thinned" network produced by applying dropout to the network on the left. Crossed units have been dropped. Illustration taken from [9]

.

(see Figure 3.6). It aims to recover fine details that might have vanished due to successive downsamplings.

3.3 Generative Models

Given a set of m training samples $\{x^{(i)}\}$, drawn from an unknown data-generating distribution $p_{data}(x)$, a Generative Model [59] takes the $\{x^{(i)}\}$ samples to learn how to represent an estimate of that distribution by following a particular approach. The resulting model is a probability distribution $p_{model}(x;\theta)$ parameterized by parameters θ. Depending on the method employed, $p_{model}(x;\theta)$ can be used to synthesize samples from $p_{data}(x)$ or/and to estimate $p_{model}(x^{(i)};\theta)$ explicitly.

A plethora of methods for learning generative models from data has been proposed in the recent few years. Classical approaches are based on the principle of **maximum a posteriori** and **maximum likelihood** estimates [60]. A survey about this theme is beyond the scope of this thesis. Instead, we will focus on Generative Adversarial Networks (GANs) [30], which represents the state-of-the-art in the field.

3.3.1 Generative Adversarial Networks (GANs)

GANs [30], are generative models designed initially in the context of modeling image distributions. It is composed of two networks: a Generator (G) that synthesizes images x, and a Discriminator (D) that determines if an image is synthetic or real. Both networks are trained in a two-player adversarial scheme, as can be seen in Figure 3.7: while the Generator tries to learn how to produce realistic images to fool the Discriminator, the Discriminator tries to correctly discriminate between synthesized and real images.

Formally, given any data distribution $p_{data}(x)$, the Generator G learns a distribution $p_{model}(x;\theta)$ such that the Discriminator can hardly distinguish between samples coming from $p_{data}(x)$ and $p_{model}(x;\theta)$.

Generally, $p_{model}(x;\theta)$ is a complex distribution, so that sampling from it is generally not a simple task. GANs circumvent this hindrance by taking a simple distribution $p_z(z)$ easy to sample from (e.g., a normal or an uniform distribution), and learns a function G that maps samples from $p_z(z)$ to samples from $p_{model}(x;\theta)$.

A GAN is trained in a min-max game searching for the optimal mapping function G^*, specifically:

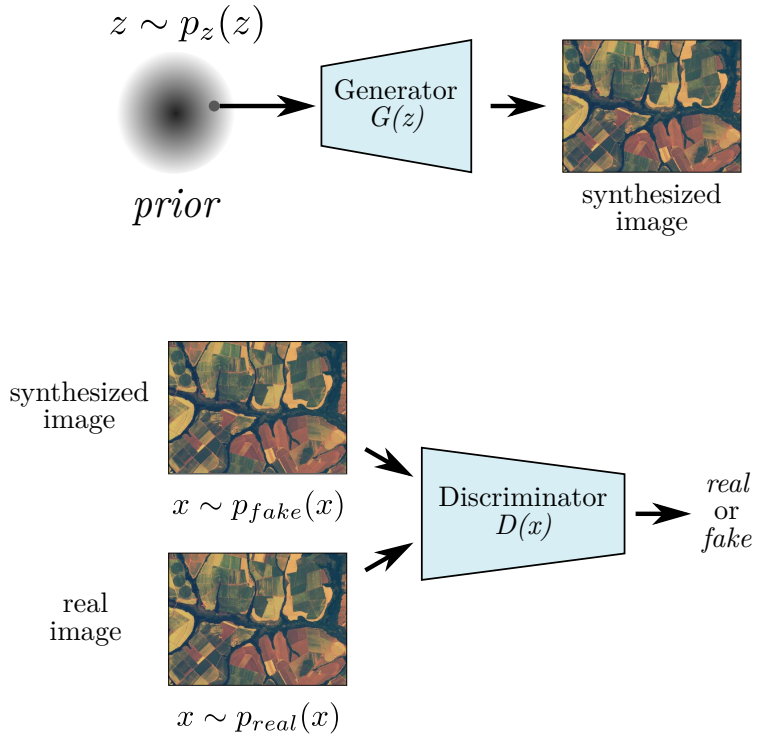

Figure 3.7: GANs training procedure. The Generator learns a function G that maps from a random noise vector z to an output image. The Discriminator learns to distinguish between real and fake (synthetic) images.

$$G^* = \arg\min_G \max_D \mathcal{L}_{GAN}(G, D) \tag{3-1}$$

where $\mathcal{L}_{GAN}(G, D)$ is the GAN objective function defined by,

$$\begin{aligned} \mathcal{L}_{GAN}(G, D) = & E_{x \sim p_{data}(x)}[\log D(x)] \\ & + E_{z \sim p(z)}[\log(1 - D(G(z))] \end{aligned} \tag{3-2}$$

where E and log are the expectation and logarithmic operators, respectively, and z is a random noise vector that follows a prior known noise distribution $p(z)$.

The solution of Equation 3-1 is obtained by training the Generator G and Discriminator D alternately. The Discriminator is trained with real images and with images produced by the last trained Generator. Similarly, the outcome of the last trained Discriminator is used to train the Generator. At the end of several training cycles, the Generator is expected to produce images that the Discriminator is not able to distinguish from real ones.

3.3.2 Conditional Generative Adversarial Networks (cGANs)

Conditional GANs, introduced by Mirza and colleagues [49], are an extension of the GANs concept. cGANs hold many similarities with the original

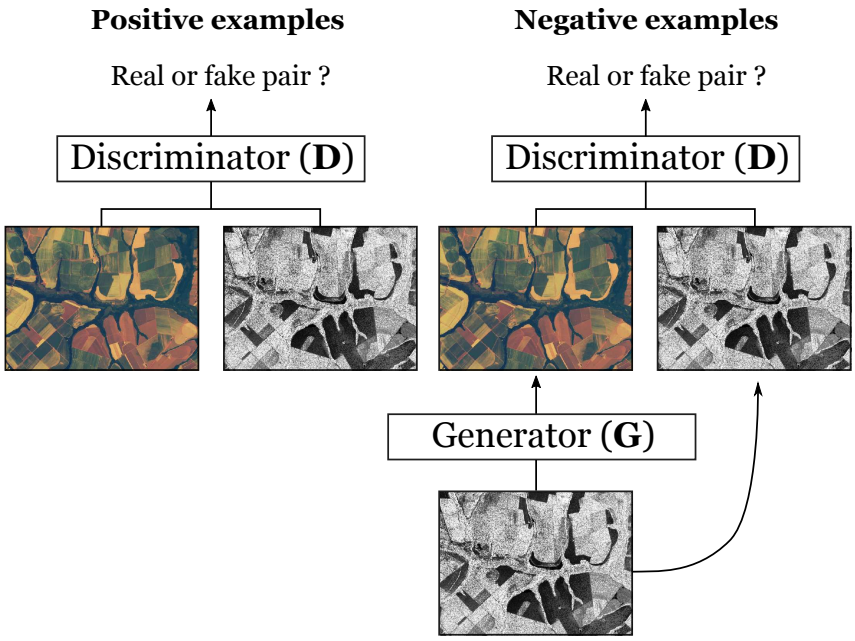

Figure 3.8: cGANs training procedure. The Discriminator learns to classify between real and fake pairs of images. The Generator learns a mapping function that takes as input a real image and outputs a realistic synthetic image from other domain. Illustration inspired in (Isola et al., 2017).

GANs, but instead of dealing with a single image, they handle a pair of co-registered images. The schema is again composed by two networks: the Discriminator, which takes as input a pair of images and learns to correctly identify if they are real-real or real-fake pair, and the Generator that learns how to generate synthetic images capable of fooling the Discriminator. The cGAN model is described in Figure 3.8.

Therefore, the Generator synthesizes images in a very specific condition: it processes a population of real images of a given domain, and learns to generate synthetic images from another domain, that should compose pairs of real-synthetic images realistic enough to fool the Discriminator. Many applications explore this characteristic for image translation, and in this work, we use it in the context of optical from SAR image synthesis.

In a more formal way, the input to the Discriminator consists of samples from two domains (y and x), and the Generator synthesizes samples from one of those domains (say x). Given any conditional probability distribution $p_{data}(y|x)$, the Generator learns a conditional distribution $p_{model}(y|x;\theta)$ given x, such that the Discriminator can hardly distinguish between the associated pair of samples (y and x) coming from $p_{data}(y)$ and $p_{data}(x)$, respectively, and the corresponding pair coming from $p_{model}(y|x;\theta)$ and $p_{data}(x)$. The loss function for conditional GANs is expressed by Equation 3-3.

$$\begin{aligned}\mathcal{L}_{cGAN}(G,D) = E_{x,y\sim p_{data}(x,y)}[logD(x,y)]+ \\ E_{x\sim p_{(x)},z\sim p_{(z)}}[\log(1-D(x,G(x,z))]\end{aligned} \tag{3-3}$$

Usually, a L1 norm distance loss is added to the Generator objective function to drive it to produce less blurred images, as shown in Equation 3-4,

$$G^* = \arg\min_G \max_D \mathcal{L}_{GAN}(G,D) + \lambda\mathcal{L}_{L1}(G) \tag{3-4}$$

where λ is a regularization term, and $\mathcal{L}_{L1}(G)$ is defined as follows,

$$\mathcal{L}_{L1}(G) = E_{x,y\sim p_{data}(x,y),z\sim p_z(z)}[\|y - G(x,z)\|_1] \tag{3-5}$$

4
PROPOSED METHOD

This chapter describes the proposed methods for synthesizing optical satellite images from different domains using cGANs. It is important to emphasize that the method described in the following was initially structured for the context of cloud removal, but it can easily be extended to other problems of missing data.

The proposed method is summarized in Figure 4.1. Let O_a represents an optical satellite image acquired at date t_a with characteristics (e.g., spatial resolution, sensor sensitivity, etc) desirable for a particular application (e.g., image categorization, semantic image segmentation, etc). Let's suppose that data from a collection of M sites $\{o_a^m : 1 \leq m \leq M\} \subset O_a$ is missing, due to sensor malfunctioning or by adverse atmospheric conditions during the image acquisition phase or any other reason.

Let's further assume that there is a set of N variables $\{f_n : 1 \leq n \leq N\}$ for which the mapping function,

$$F : \{f_n^k : 1 \leq n \leq N, \; and \; 1 < k \leq K\} \rightarrow \{o_a^k : 1 \leq k \leq K\} \subset O_a \qquad (4\text{-}1)$$

is unique, where K indicates the set of non-missing sites. Under such assumptions, a cGAN can be regarded as a method to learn the distribution

$$p(o_a|f_1, f_2, ..., f_N) \qquad (4\text{-}2)$$

whereby $\{f_1, f_2, .., f_N\}$ represents the variables that properly define the application scenario. So, equation 4-2 represents the target conditional probability distribution we are actually after. If such distribution is known, realistic data samples for the sites $\{o_a^m\}$ can be drawn, as long as the values of $\{f_1^m, f_2^m, .., f_N^m\}$ are known.

However, in realistic scenarios it is generally impossible to determine all relevant conditioning variables of a given application. Usually, we manage to capture only L, for $L < N$ of such variables. Under these constraints it is possible to estimate $p(o_a|f_1, f_2, ..., f_L)$ instead of $p(o_a|f_1, f_2, ..., f_N)$.

We hypothesize that the more conditioning variables are incorporated in the generative model, the more accurate will be the estimate of the distribution we are interested in. Our work hypothesis can be formally expressed in the

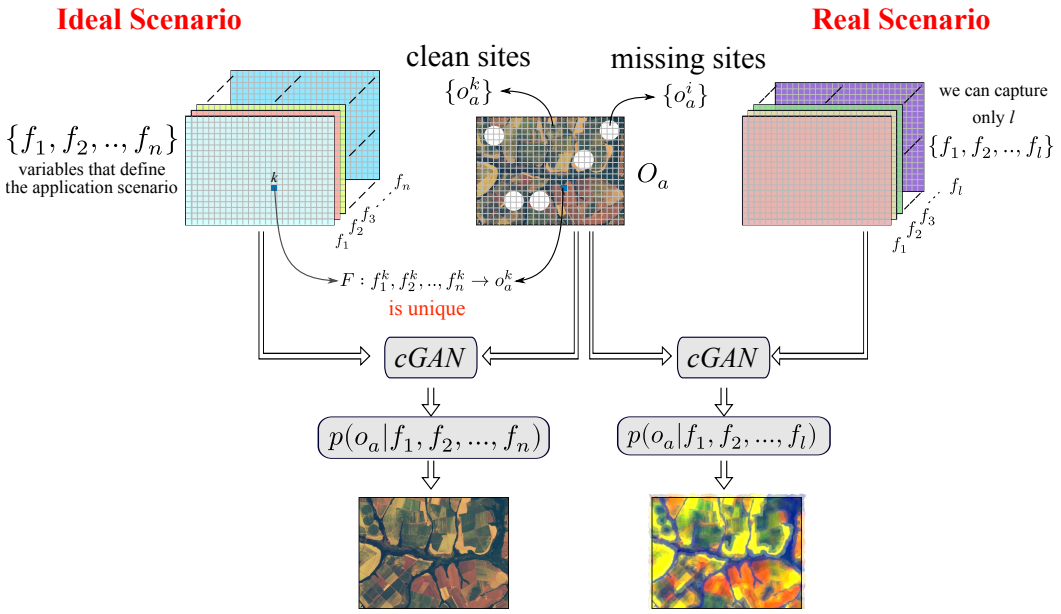

Figure 4.1: Proposed method. A cGANs is used to map from set of $f_1, f_2, ..., f_L)$ observable variables to a non-missing data optical image O_a. White circles represent the sites with missing data.

following way:

$$p(o_a|f_1, f_2, ..., f_L) \rightarrow p(o_a|f_1, f_2, ..., f_N) \;\; as \;\; L \rightarrow N \tag{4-3}$$

To investigate this hypothesis, we propose the use of conditional Generative Adversarial Networks (cGANs) to implicitly estimate $p(o_a|f_1, f_2, ..., f_L)$.

In a typical cGAN, the Generator learns a nonlinear mapping function $G : x \rightarrow y$, which contains an implicit model of the underlying conditional probability distribution $p(y|x)$ learned by training.

Nevertheless, in many reported cGANs applications, the Generator often produces an output incompatible with the underlying application scenario. This is because estimating probability distributions from a limited data set is knowingly an ill-posed problem due to the large number of different distributions that can give rise to the observed samples. To mitigate this problem, constraints have been imposed to the cGAN design. The standard approach consists of adding regularization terms to the loss function, such as the $L1$ norm in Equation 3-4. However, these solutions are often not enough depending on the complexity of the distribution being modeled.

In contrast, we propose to impose more restrictions on the cGANs models by adding conditioning data. Formally, the objective function for training a cGANs model is generalized as follow,

$$G^* = \arg\min_G \max_D E_{f_1,f_2,...,f_L,o_a \sim p_{data}(f_1,f_2,...,f_L,o_a)}[\log D(f_1, f_2, ..., f_L, o_a)] + E_{f_1 \sim p_{(f_1)}, f_2 \sim p_{(f_2)}, ..., f_L \sim p_{(f_L)}, z \sim p_{(z)}}[\log(1 - D(f_1, f_2, ..., f_L, G(f_1, f_2, ..., f_L, z))] + \lambda \mathcal{L}_{L1}(G) \tag{4-4}$$

In this work, we explored both temporal and modal relationships. As for the modal relationship, the goal can be achieved by including co-registered SAR at the same (or approximately the same) date.

As for the temporal relationship, it can be accomplished by exploiting conditioning data from the same domain of the target image, as well as from others domain, but at different acquisition dates.

We examine some variants of the proposed strategy, as illustrated in Figure 4.2. It comprises two pairs co-registered Optical/SAR images acquired at dates t_a and t_b, respectively. The Optical images are represented by O while the corresponding SAR by S. In this scheme, part of the optical image (white color circles) of (O_a) represents the area covered by clouds while the rest of the image is supposed to be cloud-free. (S_a), (S_b) and (O_b) represent the conditioning variables to be used by the cGANs model for synthesizing a plausible O_a sample.

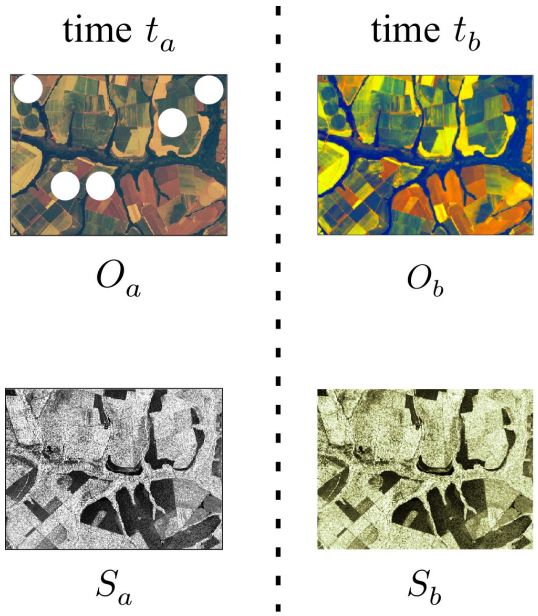

Figure 4.2: Overview of scenario explorer in this thesis. O symbolizes an optical image and S a SAR image. The optical image O_a simulate the image covered by clouds. White circles represent the regions covered by clouds.

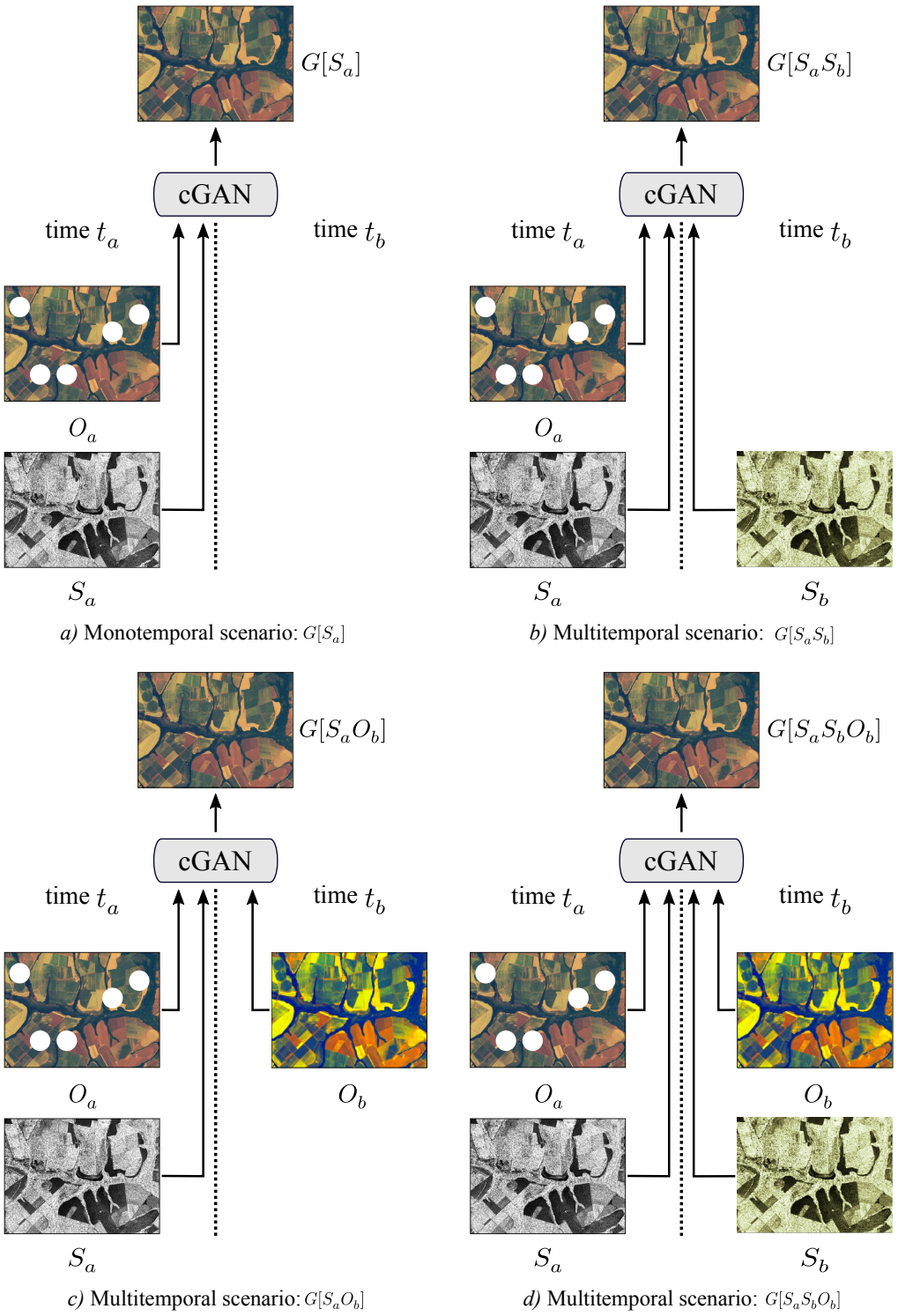

Figure 4.3: Variants exploited of the scenario illustrated in Figure 4.2. White circles represents the regions covered by clouds.

We first investigate the scenario where only the SAR image at date t_a (S_a) is available to condition the cGAN design (Figure 4.3-*a*). In this solution, named hereafter as *monotemporal*, only modal relationships are exploited for

conditioning the Generator. In fact, it represents the particular case of a cGAN model conditioned by just one variable. Then, we examine what we call multitemporal scenarios where data at another date t_b can be used to condition the cGAN. Two variants are explored: either using a SAR (S_b) (Figure 4.3-b) or an optical image (O_b) (Figure 4.3-c) at date t_b. Finally, we consider a variant that exploits all data (Figure 4.3-d). These variants, named hereafter as *multitemporal*, involve modal and temporal relations. At follows we first describe *monotemporal* approach and then the *multitemporal* variants.

4.0.1 Monotemporal approach

As part of this thesis, we published in [5] a research about the *monotemporal* variant. In this work, we proposed a method based on cGANs to synthesize optical images from SAR data for recovering regions covered by clouds. In this context, it was considered the fact that SAR images are nearly independent of atmospheric conditions and solar illumination to learn, via cGANs, a conditional probabilistic model $p(o_a|s_a)$ for mapping SAR data to optical cloud-free images. It is important to emphasize that this method is not restricted to the use of SAR imagery. As we said earlier, data of other sensors could be used or incorporated as conditioning variables as long as it does not present missing data. However, the quality of the synthesized image is expected to be different depending on the capacity of the other sensor technology for capturing relevant information about the environmental process being modeled.

Given a co-registered SAR image (S_a) acquired approximately at the same acquisition date (t_a) of the corrupted image (O_a), in the *monotemporal* approach, the cGAN model is trained using a collection of corresponding SAR/Optical patches $\{s_a, o_a\}$ extracted over the cloud-free region of the optical image. Figure 4.4 illustrates the principal steps of the method processing chain.

For this configuration, the cGAN optimization objective function described in Equation 4-4 takes the following form,

$$\begin{aligned} G^* = \arg\min_G \max_D E_{s_a,o_a\sim p_{data}(s_a,o_a)}[logD(s_a,o_a)] + \\ E_{s_a\sim p_{(s_a)},z\sim p_{(z)}}[log(1-D(s_a,G(s_a,z))] + \lambda\mathcal{L}_{L1}(G) \end{aligned} \tag{4-5}$$

Once this model has been optimized, the conditional probability underlying the Generator network is used for synthesizing plausible optical images $\hat{O}_a$.

4.0.2

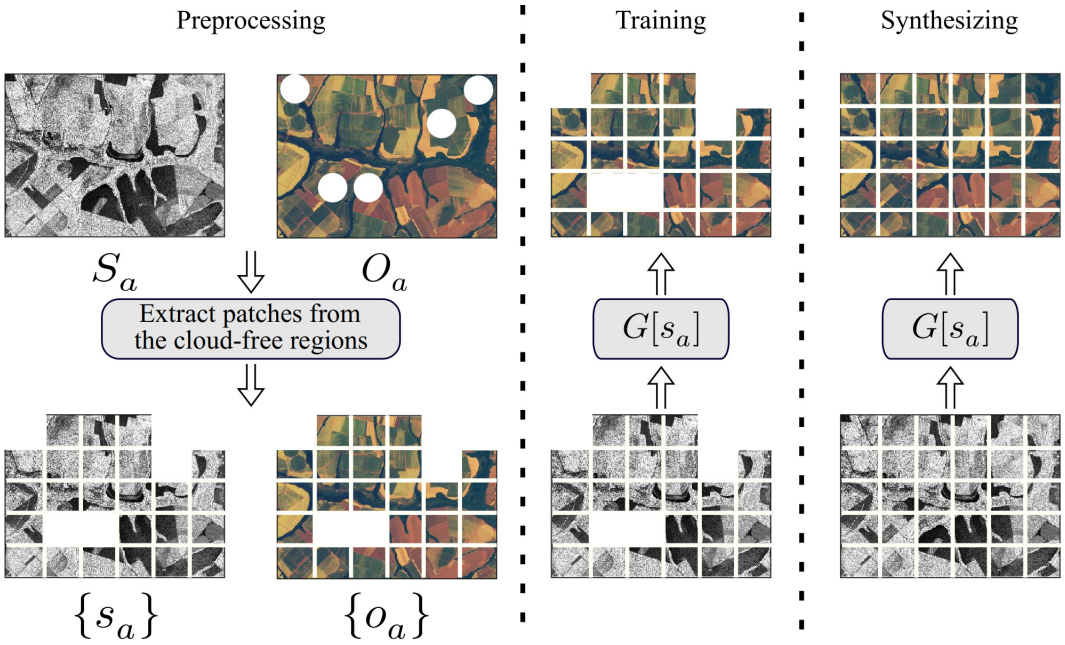

Figure 4.4: *Monotemporal* method for cloud removal of optical satellite images. A cGAN is trained to learn a nonlinear mapping function G that maps from a co-registered SAR image at t_a to a plausible optical image at t_a. White circles represent the regions covered by clouds.

Multitemporal approach

Figure 4.5 illustrates the processing chain of the *multitemporal* approach when the aforementioned images are considered as conditioning variables in the cGAN model.

Figure 4.6 describes the training process of the cGANs model for more than one conditional variable. Essentially, it involves stacking all conditioning images along its channels dimension before being fed to the Generator and Discriminator Networks. This research was published in [61] as part of this thesis.

Given a collection of co-registered $\{o_a^k, s_a^k, o_b^k, s_b^k\}$ training patches, extracted over the cloud-free region of the corrupted optical image O_a, the cGAN optimization function for the *multitemporal* method assumes the following form,

$$G^* = \arg\min_G \max_D E_{s_a,s_b,o_b,o_a \sim p_{data}(s_a,s_b,o_b,o_a)}[\log D(s_a, s_b, o_b, o_a)] + E_{s_a \sim p_{(s_a)}, s_a \sim p_{(s_a)}, s_b \sim p_{(o_b)}, z \sim p_{(z)}}[\log(1 - D(s_a, s_b, o_b, G(s_a, s_b, o_b, z)))] + \lambda \mathcal{L}_{L1}(G) \quad (4\text{-}6)$$

Likewise the *monotemporal* approach, the trained Generator is then used

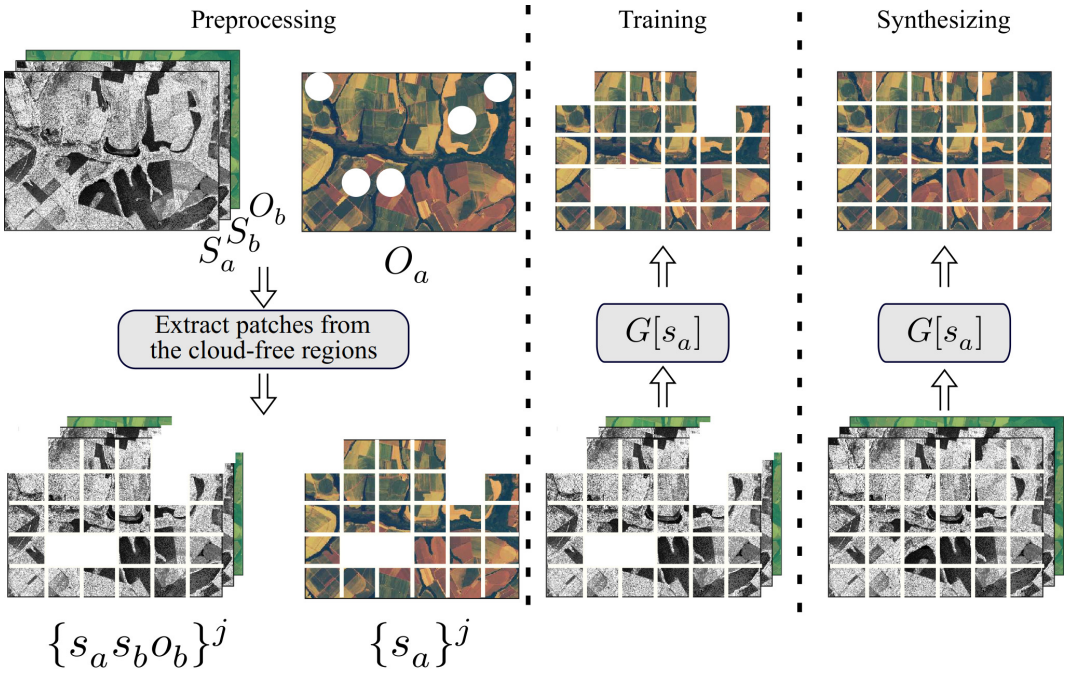

Figure 4.5: Proposed *multitemporal* method for cloud removal in optical satellite images. A cGAN is trained to learn a nonlinear mapping function G that maps a set of three co-registered images (SAR at t_a, and SAR plus optical at t_b) to a plausible optical image at t_a. White circles represent the regions covered by clouds.

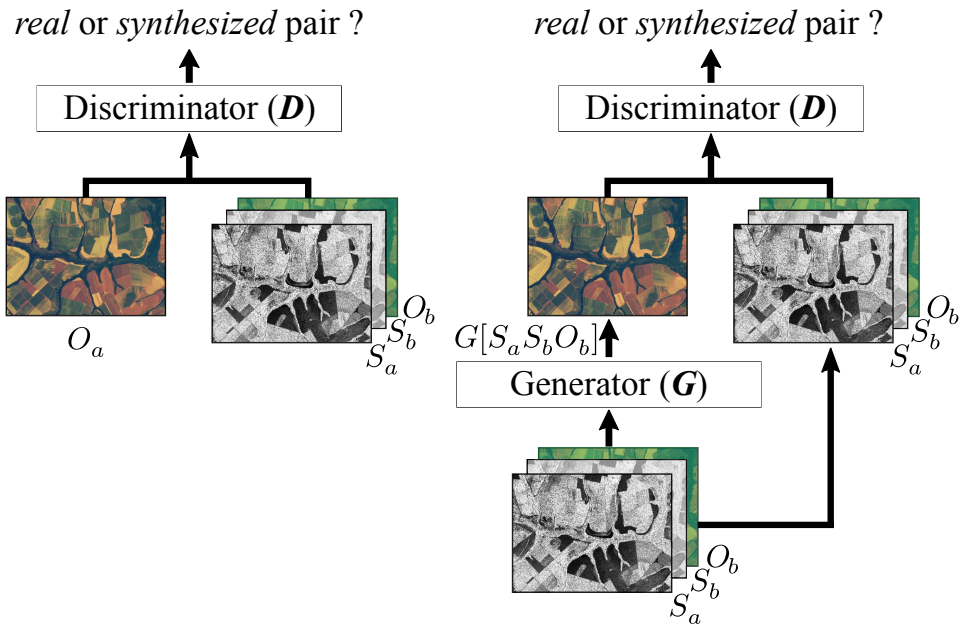

Figure 4.6: The cGAN Generator learns a nonlinear function G that maps a set of three co-registered images (SAR at t_a, and SAR plus optical at t_b) to a plausible optical image at t_a. The cGAN Discriminator learns a function D that separates real from synthetic optical images produced by the Generator.

for synthesizing an estimate of the O_a image.

In summary, both *monotemporal* and *multitemporal* methods involve the following steps.

First, the cloud-free regions are identified via visual observation or by using a cloud detection algorithm like the *Fmask* [62] and *sen2cor* [63] algorithms for Landsat and Sentinel 2 satellite imagery, respectively.

Second, a collection of corresponding patches over the previously identified cloud-free region are extracted through the sliding window procedure with a fixed stride. For the *monotemporal* method we collect pairs of cloud-free co-registered optical/SAR patches from date t_a (O_a, and S_a), whereas for the *multitemporal* method the correspondent co-registered patches at date t_b (O_{t_b} and S_b) are also extracted.

Third, the cGAN model is trained upon the extracted groups of patches, as illustrated in Figure 4.6.

Fourth, once the cGAN has been trained, optical patches over *cloudy* areas at t_a are synthesized by the generator using as input the corresponding SAR patches at t_a, for the *monotemporal* method, as well as correspondent SAR and optical patches at t_b, for the *multitemporal* approach.

Finally, the predicted optical patches are concatenated to build a mosaic over the *cloudy* areas. At patch boundaries, the prediction tends to be less accurate because less spatial context is considered for their generation. This effect can be attenuated by generating overlapping patches and retaining only their central part, which can be subsequently merged to produce a smoother mosaic.

Note that, the cGAN must be trained on a set of cloud-free patches that represents the distribution of missing data. It is, therefore, important that the cloud-free samples encompass most of (ideally all) the classes that may be present on the area covered by clouds. Otherwise, the cGAN will not be able to capture all data variability present on the target image during its training phase. In other words, if there are classes on the cloudy area that are not represented on the samples collected over the cloud-free region, the nonlinear mapping function may not be able to synthesize plausible data for the cloud covered regions.

It is similarly important that the difference between the acquisition dates of SAR and optical image pairs taken as referring to the same date should be as short as possible. This is important to reduce the impact of possible changes of classes distribution or even appearance of changes in a class, like seasonal variations in crops, for instance. So, depending on the application, this time difference can be a crucial factor to achieve a quality result.

In addition, two aspects must be observed for the *multitemporal* method. First, the training samples collected from the optical image O_b must be cloud-free. Second, those samples must have been generated by the same sensor, which acquired the cloudy optical image O_a in order to model just temporal relationships. Although the method is not restricted to the use of the same optical sensor-based technology of the cloudy image, in the explored scenario the complexity of the model is expected to increase if another sensors are used.

5
EXPERIMENTAL ANALYSIS

This chapter reports the set of experiments conducted to assess the capability of the proposed methods for synthesizing multispectral optical images from SAR/Optical multitemporal data. Most of the works on cloud removal assessed the performance in terms of similarity metric between the synthesized data and a reference. In this work, in addition to these similarity metrics, we also evaluated the quality of the synthesized images in terms of semantic image segmentation performance, since it is the focus of our application. Additionally, we present a visual comparison analysis between the original image and its corresponding synthesized ones.

5.1
Datasets

Two different scenarios were selected for evaluating experimentally the performance of the proposed methods. The first one corresponds to a crop recognition application and the second one to the wildfire detection. The two datasets are described in the following.

5.1.1
Campo Verde

This dataset [64] refers to Campo Verde municipality in the state of Mato Grosso, Brazil, which has an extension of approximately 4782 km^2 (see Figure 5.1). Four co-registered images were taken from this locality: two Landsat 8 OLI and two Sentinel-1A SAR scenes with 30 m and 10 m spatial resolution, respectively. Table 5.1 summarizes the acquisition dates of the corresponding images.

Table 5.1: Acquisition dates for *Campo Verde* dataset.

Image	Sensor	Acquisition Date
O_a	Landsat 8 OLI	05 May 2016
O_b	Landsat 8 OLI	24 May 2017
S_a	Sentinel 1A	08 May 2016
S_b	Sentinel 1A	20 May 2017

Note that the images at (t_b) were acquired approximately one year after the target date (t_a). Therefore, those images refer to the same cropping season and are expected to contain approximately identical crops with differences in spatial distribution.

Nine land cover classes were considered for this work. Four of them are related to crop types (maize, cotton, sorghum and noncommercial crops), and five are related to non crop classes (pasture, eucalyptus, soil, turfgrass, and Cerrado). Table 5.1 indicates the class occurrences.

Table 5.2: Class occurrences for *Campo Verde* dataset.

Class	**%**	**# pixels**
Maize	35.79	243,097
Cotton	45.33	307,883
Sorghum	0.94	6,404
NCC	3.87	26,328
Pasture	8.58	58,308
Eucalyptus	2.53	17,250
Soil	0.36	2,457
Turfgrass	0.02	108
Cerrado	2.55	17,349

5.1.2 Rio Branco

This dataset is from Rio Branco municipality, located in the state of Acre, Brazil with an extension of 8,836 km^2 [17] (See Figure 5.2). Similar to *Campo Verde*, this dataset comprises four images, two Sentinel-2A, and two Sentinel-1A SAR, all of them with 10 m of spatial resolution. The acquisition dates are indicated in Table 5.3.

Table 5.3: Acquisition dates for *Rio Branco* dataset.

Image	**Sensor**	**Acquisition Date**
O_a	Sentinel-2A	25 Aug 2016
O_b	Sentinel-2A	31 Jul 2017
S_a	Sentinel 1A	09 Sep 2016
S_b	Sentinel 1A	31 Jul 2017

The occurrence of wildfires in the region is driven by deforestation and extractivism, which generate ignition sources mainly in areas close to the forest [65]. Table 5.4 summarizes the distribution of wildfire and non-wildfire samples in the study region. Main land cover classes include burned areas, forest, agricultural use, areas without vegetation and water bodies. The

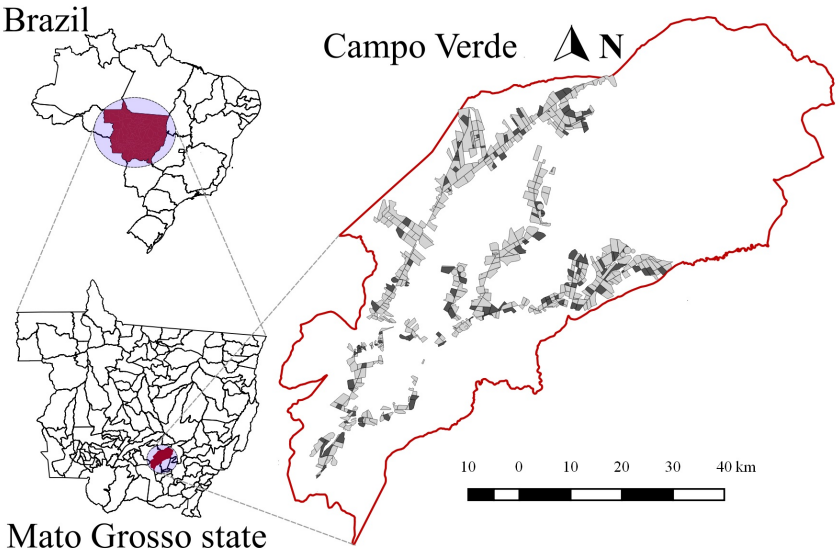

Figure 5.1: Study area: Campo Verde, Mato Grosso state, Brazil.

experiments performed using this dataset were only focused on detection of the burned areas.

Table 5.4: Class occurrences for *Rio Branco* dataset.

Class	**%**	**# pixels**
Wildfire	0.90	869,211
Non-wildfire	99.0	95.194,563

5.2 Evaluation Metrics

In this thesis we adopted the *Overall Accuracy* and F_1-*score* performance metric of the semantic segmentation outcome. In addition, we used the *Root Mean Square Error*, *Peak Signal-to-Noise Ratio* and the *Spectral Angle Mapper* as similarity metrics for comparing the synthesized images with their correspondent references. The definition of each metric is given next,

– Overall Accuracy (OA): The OA indicates the percent of samples correctly classified by the model, being 100% a perfect classification.

$$OA = \frac{number\,of\,correctly\,predicted\,samples}{total\,of\,samples\,to\,predict} \tag{5-1}$$

– F_1-*score*: F_1-*score* can be interpreted as the harmonic mean between *Precision* and *Recall*. The F_1-*score* reaches its best value at 100 and worst score at 0. The relative contribution of *Precision* and *Recall* to the F_1-*score* are equal [66].

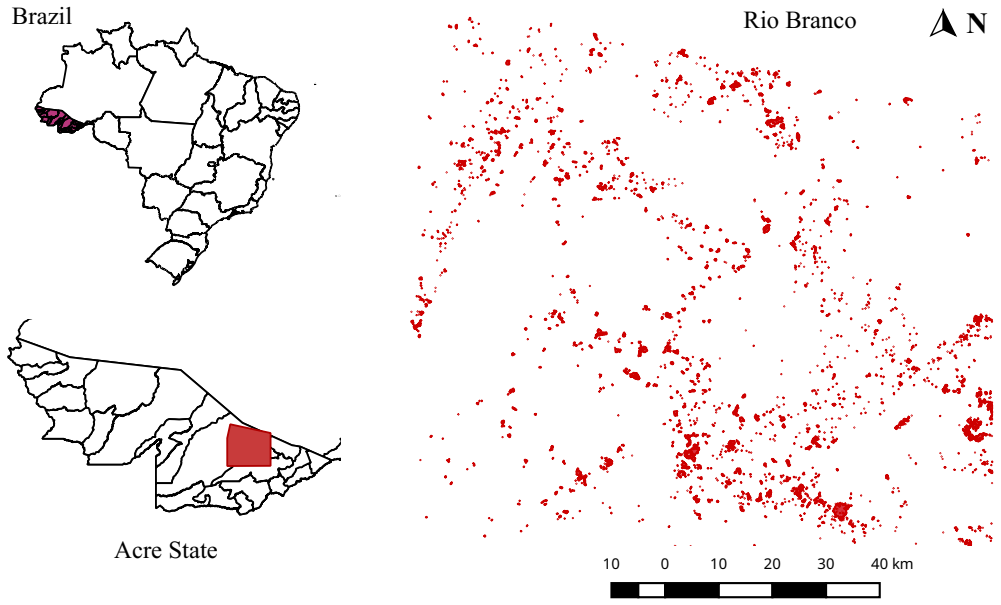

Figure 5.2: Study area: Rio Branco, Acre state, Brazil. Wildfire samples are represented in red.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{5-2}$$

where $Precision$ and $Recall$ are defined as follow,

$$Precision = \frac{tp}{tp + fp} \tag{5-3}$$

and,

$$Recall = \frac{tp}{tp + fn} \tag{5-4}$$

being tp the number of true positives, fp the number of false positives and fn the number of false negatives.

For *Campo Verde* dataset, which represents a multiclass problem, we report the average F_1-*score*. It consists of computing the precision and recall of all the classes, calculates the F_1-*score* per class, and then calculate the average of each measure.

– Root Mean Square Error (RMSE): Given a $m \times n$ image I and a synthetic version $\hat{I}$, the $RMSE$ is defined by,

$$RMSE = \sqrt{\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - \hat{I}(i,j)]^2} \tag{5-5}$$

Lower values of RMSE indicate high similarity between compared images. The maximum score that the RMSE reaches depends on the image data type. In this work, all images were codified as 16-bit signed integers. Thus, the maximum value is $65,535$ for two complete different pixels.

– Peak Signal-to-Noise Ratio (PSNR): The PSNR (in *db*) is defined as,

$$PSNR = 20 * \log_{10}\left(\frac{I_{MAX}}{RMSE}\right) \tag{5-6}$$

where I_{MAX} is the maximum possible pixel value of an image. Contrary to the *RMSE* metric, high values of *PSNR* are associated with good quality of the synthesized image.

– Spectral Angle Mapper (SAM): The SAM measures the spectral angle between the spectral signatures of the pixels from two images [48].

$$\theta(x, y) = cos^{-1}\left(\frac{\sum_{i=1}^{nb} x_i y_i}{(\sum_{i=1}^{nb} x_i^2)^{\frac{1}{2}} * (\sum_{i=1}^{nb} y_i^2)^{\frac{1}{2}}}\right) \tag{5-7}$$

where nb represents the number of spectral bands of the image, and x and y refers to pixel from image I and $\hat{I}$ at particular location. Like *RMSE*, lower values of *SAM* are related to a high similarity between images.

5.3 Network Architectures

The Generator and Discriminator network architectures used in this work are based on the architectures proposed in [52]. In particular, we modified these architectures to be able to work with multispectral optical images and multiresolution sensors. More specifically, we took as base a Tensorflow [67] implementation of the *pix2pix* [52] framework[1] and adapted it for our framework[2]. We adopted a specific architecture for each method and for each dataset since the characteristics of their input and output images are different. The configuration of these networks are described in details in Table 5.5 and Table 5.6 for *Campo Verde* and *Rio Branco* dataset, respectively. In these tables, symbols C, B, R and D denote convolution (C), batch normalization (B), ReLU (R) and dropout (D) for each layer. The numbers of filters, filter dimensions and convolution strides are indicated in this sequence in parentheses. All filters are square, and the stride is equal in the horizontal and vertical directions. Similar to [52], we remove the dependency of the random noise vector z from the cGAN's objective function by applying the dropout regularization on several layers of the Generator at both training and test time.

For *Campo Verde*, the input patches have 256×256 pixels for Landsat images and 768×768 for the corresponding SAR data. In order to deal with the difference in the spatial resolution, we included a 2D convolutional layer

[1] https://github.com/yenchenlin/pix2pix-tensorflow
[2] https://github.com/bermudezjose/SAR2Optical-using-cGANS

Table 5.5: Network Architectures for *Campo Verde*.

Encoder	Decoder	Discriminator
CBR(4, 5, 3)*	CBRD(512, 5, 2)	CBR(4, 5, 3)*
CR(64, 5, 2)**	CBRD(512, 5, 2)	CR(64, 5, 2)**
CBR(128, 5, 2)	CBRD(512, 5, 2)	CBR(128, 5, 2)
CBR(256, 5, 2)	CBRD(512, 5, 2)	CBR(256, 5, 2)
CBR(512, 5, 2)	CBRD(256, 5, 2)	CBR(512, 5, 2)
CBR(512, 5, 2)	CBRD(128, 5, 2)	*sigmoid*(.)
CBR(512, 5, 2)	CBRD(64, 5, 2)	
CBR(512, 5, 2)	C(7, 5, 2)	
CBR(512, 5, 2)	*tanh* (.)	

*The input is the concatenation of S_a and S_b patches. **The input is the concatenation of the output of the prior layer and the O_b patches.

Table 5.6: Network Architectures for *Rio Branco*.

Encoder	Decoder	Discriminator
CR(64, 5, 2)*	CBRD(512, 5, 2)	CR(64, 5, 2)*
CBR(128, 5, 2)	CBRD(512, 5, 2)	CBR(128, 5, 2)
CBR(256, 5, 2)	CBRD(512, 5, 2)	CBR(256, 5, 2)
CBR(512, 5, 2)	CBRD(256, 5, 2)	CBR(512, 5, 2)
CBR(512, 5, 2)	CBRD(128, 5, 2)	*sigmoid*(.)
CBR(512, 5, 2)	CBRD(64, 5, 2)	
CBR(512, 5, 2)	C(4, 5, 2)	
	tanh (.)	

*The input is the concatenation of S_a, S_b and O_b patches.

in the cGAN network to map the SAR patches to the resolution of the optical data. This process is illustrated in Figure 5.3 and Figure 5.4, respectively, for the Generator and Discriminator architectures for the *multitemporal* method. Similar scheme was followed for *monotemporal* approach. We preferred to introduce the 2D convolutional layer instead of downsampling the SAR patches via a traditional image interpolation function. In this way, we rely on the capacity of the convolutional layer to learn an interpolation function tailored to our application, which is supposed to result into less information loss than traditional interpolation techniques [68].

Downsampling was not necessary for *Rio Branco*, since the optical and SAR images have the same spatial resolution. For this dataset, we worked with smaller patches of 128×128 pixels seeking to balance the proportion of wildfires pixels per patch. This way, we tried to avoid that the cGANs learn only the majority classes and disregard the wildfires.

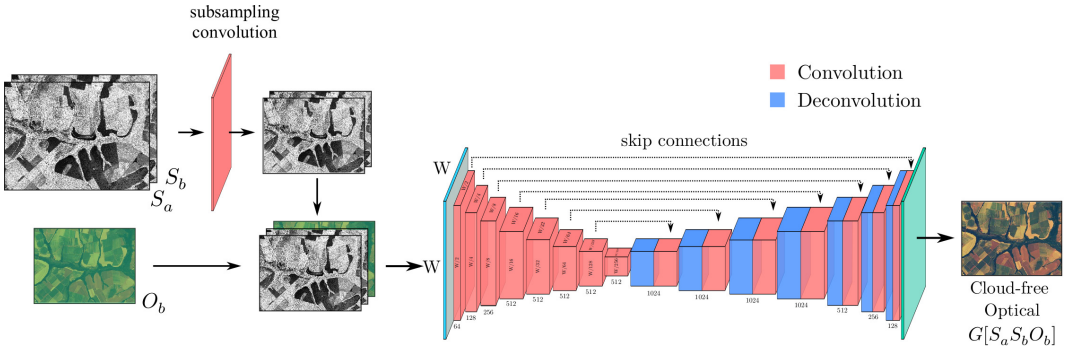

Figure 5.3: Generator Network architecture for the *monotemporal* approach used for *Campo Verde* dataset.

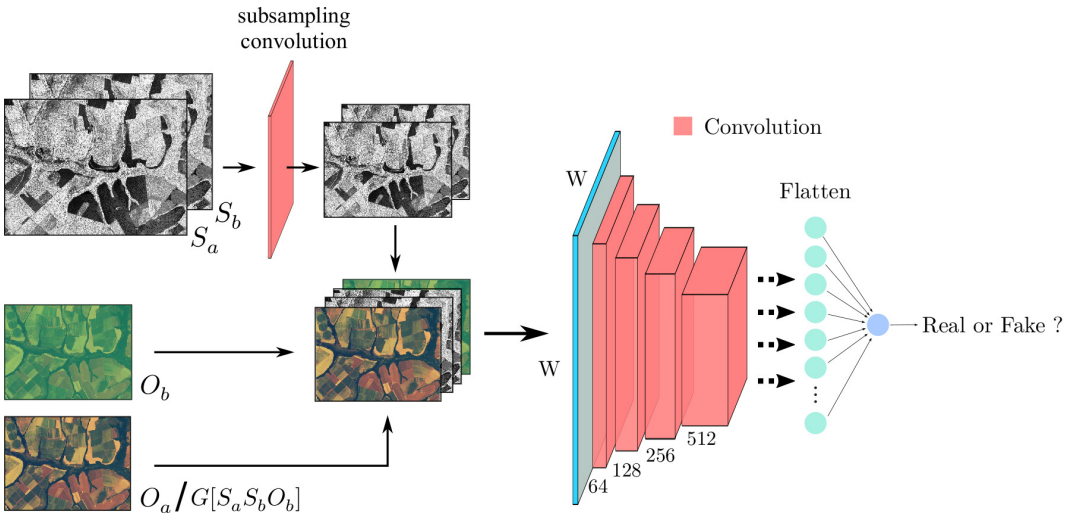

Figure 5.4: Discriminator Network architecture for the *multitemporal* approach used for *Campo Verde* dataset.

5.4 Experimental Protocol

For the experimental analysis we selected two pairs of co-registered cloud-free optical/SAR images from both datasets, acquired on the dates mentioned in Sections 5.1.1 and 5.1.2. The images acquired in 2016 were associated with the target date (t_a), whereas the images from 2017 were associated with the other date (t_b).

We split the imaged areas into two spatially disjointed sets, as shown in Figure 5.5 and Figure 5.6 for *Campo Verde* and *Rio Branco* datasets, respectively. We used all regions of the first set (blue color) for simulating the regions with unavailable optical data at t_a. These regions are referred as *cloudy* hereinafter. Similarly, we selected all regions of the second set (green color) to represent the regions in which optical data were available at both dates. These

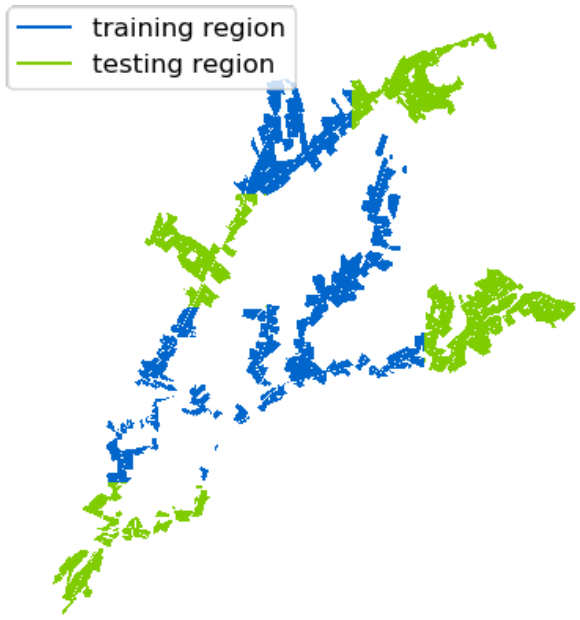

Figure 5.5: Distribution of training (blue) and testing (green) regions used on experiments for *Campo Verde* dataset.

regions are named *clear* hereafter. The criteria for selecting the *clear* and *cloudy* regions consisted of guaranteeing the representation of all classes in both regions. For *Campo Verde*, the *cloudy* and *clear* sets contained 327,248 and 1,571,515 pixels, respectively. For the *Rio Branco* dataset, 26,184,068 and 69,860,629 pixels were included in the *cloudy* and *clear* sets, respectively. The cGANs were trained upon the *clear* regions, while methods' performance were evaluated on *cloudy* regions.

For *Campo Verde*, we only considered patches around agricultural areas in order to specialize the cGANs model to cropland. Data outside the agricultural area were not considered to train the cGANs model due to could present information not related to the problem that might degrade the cGANs model for agricultural mapping. For the experiments on *Rio Branco*, we selected training samples in order to balance the number of samples per class. In particular, we followed a stratified sampling procedure consisting of extracting more patches from areas with presence of wildfires.

Training patches were extracted from the *clear* regions following the sliding window strategy. Specifically, 4,000 SAR-optical patches were cropped for each dataset. These samples were augmented by applying random cropping and horizontal and vertical flip transformations to the first set of patches. In all experiments, the networks were trained for 100 epochs using the Adam [69] optimizer with default parameters: learning rate=0.001, β_1=0.9 and β_2=0.999.

The performance of the methods was assessed in terms of semantic segmentation metrics and similarity metrics. The pixel-wise accuracy of the

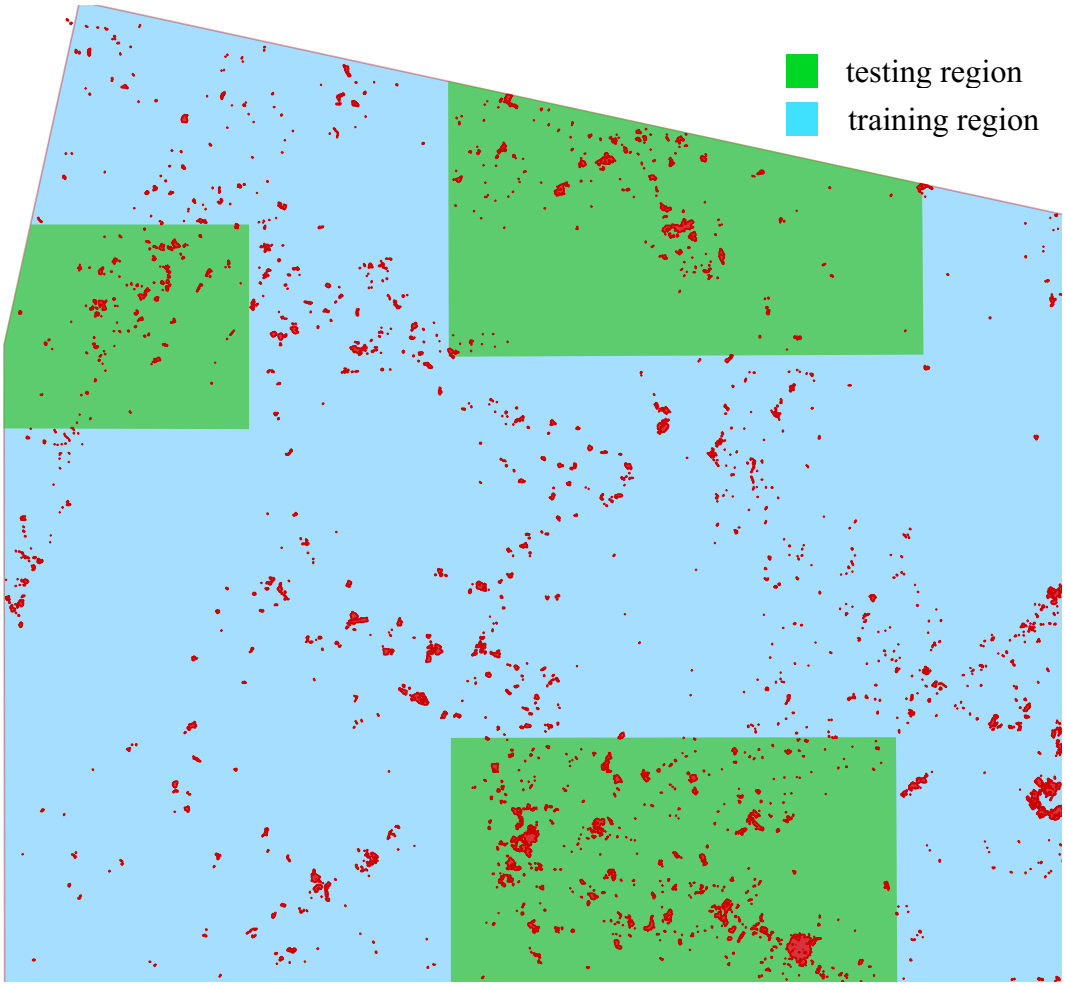

Figure 5.6: Distribution of *clear*(blue) and simulated *cloudy* (green) regions used on experiments for *Rio Branco* dataset. Wildfire samples are represented in red.

semantic segmentation were computed upon the *cloudy* regions of the SAR or real/synthetic optical images at t_a. The results in all cases were obtained by a Random Forest (RF) [70] classifier, which used as pixel-wise feature descriptor, the vector comprising bands 1 to 7 for Landsat images and the Red, Green, Blue and NIR bands for Sentinel 2 images. For the SAR images, we computed features based on the Gray Level Co-occurrence Matrix (GLCM). Specifically, we used the VV and VH polarizations for computing the GLCM correlation, homogeneity, mean and variance in four directions (0, 45, 90 and 135 degrees) using 7×7 windows. Then, each SAR pixel was represented by a feature vector of dimensionality 32.

For each evaluated image, an RF model was trained upon the *clear* region using the corresponding pixel descriptor of the assessed image. In particular, for *Campo Verde* dataset the RFs were trained on approximately 25% of the *clear* pixels . For *Rio Branco*, all available wildfire samples on *clear* regions,

and the same number of samples of the non-wildfire complementary class were used for training the RF classifiers.

We selected the RF as the algorithm for performing the semantic segmentation task due to its performance in classification, speed, and insensitivity to overfitting [71]. Indeed, this classifier has become popular within the RS community because of these characteristics.

As similarity metrics, we computed the *root mean square error*, the *spectral angle mapper* and the *Peak Signal-to-Noise Ratio*, as defined in Section 5.2. These metrics were computed only on *cloudy* regions, i.e., we evaluated only the testing set.

5.5 Results

5.5.1 Semantic Segmentation

The bar graphs illustrated in Figure 5.7 and Figure 5.8 summarize the results of the experiments carried out in *Campo Verde* and *Rio Branco* datasets, respectively. Each figure shows the classification performance achieved by the RF classifier over the *cloudy* pixels for the evaluated images in terms of *Overall Accuracy* (OA) and average F_1-*score*.

From left to right, the first bar group (O_a) refers to the classification of the *cloudy* pixels of the optical image acquired at t_a assuming that this data is available. Recall that the real optical data over *cloudy* regions were available in our data set. These results were regarded as the best image that could be possibly synthesized. In other words, these bars represent the upper bound for the classification accuracy. Note that the performance of RF classifier in this image O_a was superior to the other assessed images.

The second bar group (S_a) relates to classification performance of the SAR data at t_a using GLCM-based features. We decided to use GLCM features because they are by far the most commonly used SAR descriptor data in RS applications. In this work, GLCM-based features served as a baseline. Notice that, for *Campo Verde* the classification of SAR data at t_a presented the lowest accuracy while for *Rio Branco* dataset, the second lowest one.

The third bar group from the left (O_b) corresponds to the classification of the optical image acquired at t_b over *cloudy* regions, considering the reference at t_a. The inclusion of these results in our reports was motivated by the following rationale. Replacing the cloudy optical image (O_a) by a cloud-free optical image (O_b) of the same area and the same sensor, but at a different acquisition date,

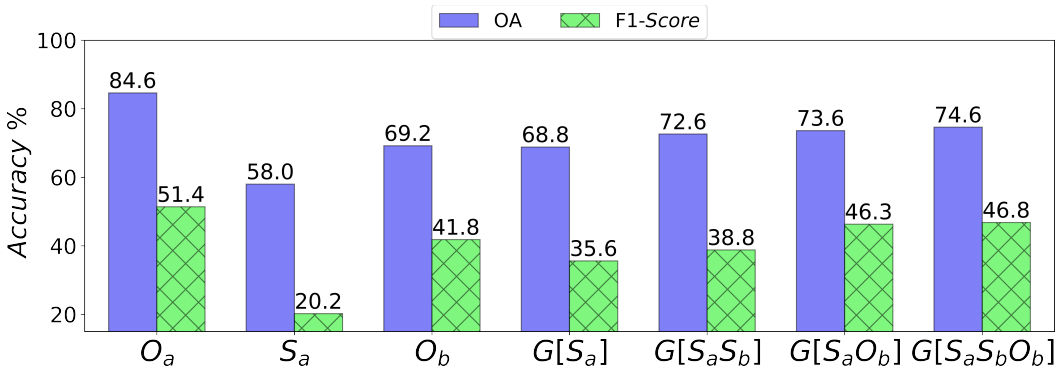

Figure 5.7: Result for *Campo Verde* in terms of OA and Average F_1-*score*.

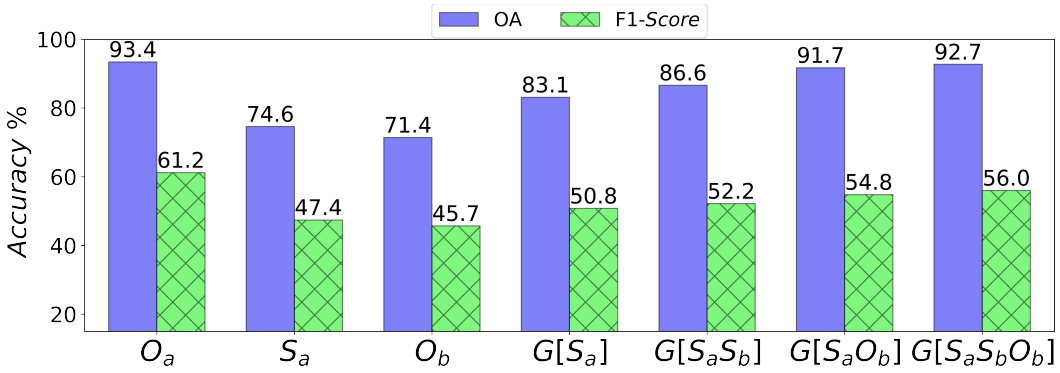

Figure 5.8: Result for *Rio Branco* in terms of OA and Average F_1-*score*.

can be an acceptable solution for some applications. Note that this variant would be plausible in areas with low temporal dynamic. So, we carried out this experiment to verify whether this approach would make sense for our datasets, i.e., if the images at t_a and t_b are similar. The Figures show that our results clearly refuted this hypothesis. Observe that the accuracies for image O_b in both datasets were significantly lower than those obtained for O_a. Nonetheless, this variant surpassed its SAR counterpart at $t_a(S_a)$, in *Campo Verde* and was the poorest variant considered in our experiments on *Rio Branco* dataset.

The next four bar groups indicate the performance obtained on the semantic segmentation of optical images synthesized by the cGAN Generator using different conditioning data. The fourth group from left, ($G[S_a]$), refers to the classification of the synthesized optical images at t_a produced by the *monotemporal* variant (as in [5]) from the correspondent SAR data at the same date (t_a). This model performed better than the S_a variant by approximately 10.8% and 8.5% in terms of OA for *Campo Verde* and *Rio Branco*, respectively. These results are consistent with the accuracies reported in [5]. The *monotemporal* approach was outperformed by the O_b variant in terms of OA and F_1-*score* in *Campo Verde*. In *Rio Branco* dataset these

variants performed similarly. It is important to emphasize that this variant does not create any information that is not already contained in the SAR data. The results just show that the method was able to extract more discriminative representations than the GLCM-based features.

The three rightmost bar groups refer to the *multitemporal* variants of our method. In these experiments, the conditioning data comprised the SAR data at t_a combined with SAR at t_b, optical at t_b and both. These variants are denoted in Figure 5.7 and Figure 5.8 by $G[S_aS_b]$, $G[S_aO_b]$ and $G[S_aS_bO_b]$, respectively. All of them outperformed their *monotemporal* counterpart ($G[S_a]$) in terms of OA, with accuracy gains between 3.8% and 5.8% for $Campo\ Verde$ and between 3.5% and 9.0% for *Rio Branco*.

The $G[S_aS_b]$ and $G[S_aO_b]$ bars in both plots also reveal that the performance gain from including data from another date (t_b) to condition the cGAN was greater for the optical image than for the SAR data, especially for the experiments on *Rio Branco* dataset. This result is not surprising, because data of the same optical domain was added as conditioning variable. Here, the cGAN is focused into modeling the temporal relationship between the cloudy and cloud-free image.

The right-most bars in both plots ($G[S_aS_bO_b]$) present the best accuracies among the synthesis approaches. They correspond to forming the conditioning data with both optical and SAR data from t_b in addition to the SAR data at t_a. This variant was the closest one to the ideal performance, represented by the leftmost bars for both datasets. Indeed, ($G[S_aS_bO_b]$) was 10% lower for *Campo Verde* and only 0.7% lower for the *Rio branco* dataset in terms of OA. As for the F_1-*scores* the difference was approximately 5% for both datasets.

The F_1-*scores* were much lower than OA in both plots for all variants. The reason lies in the high class imbalance in the test data, particularly for *Campo Verde*. However, the plots show a similar relative profile for OA and F_1-*score*.

5.5.2 Visual Analysis

Figure 5.9 shows snips of the evaluated images (the original images and the images synthesized by each of the tested variants), and the corresponding classification maps form *Campo Verde* dataset. The first and third columns contain the evaluated images, and the second and fourth columns show the corresponding classification maps delivered by the RF classifier. For the SAR image, S_a, VH polarization is presented and for the optical images, a RGB true color composition.

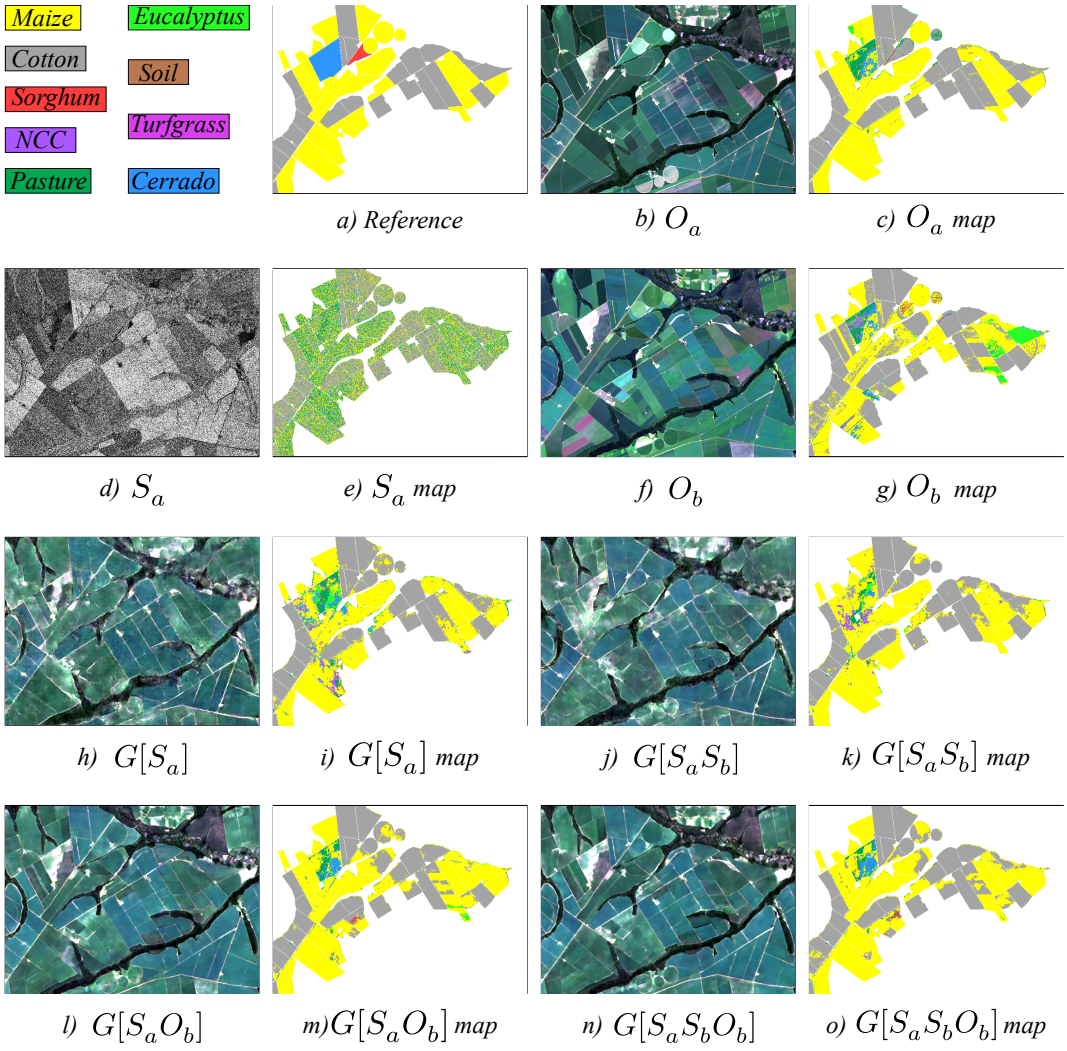

Figure 5.9: Snips of the evaluated images (the original images and the images synthesized by each of the tested variants), and the corresponding classification maps delivered by the RF classifier over the *cloudy* pixels for *Campo Verde* dataset. The RF was trained upon each of these images. The snips of optical images correspond to the RGB composition band. The contrast was adjusted for better visualization.

A comparison with the reference shows that, except for the SAR image ($G[S_a]$), the RF classifier managed to correctly classify most of the *maize* and *cotton* samples, which represent the majority classes. In contrast, the *Sorghum* samples were almost totally misclassified, even for the real optical image O_a. Recall that this class represents only 0.02% of labeled samples. Also, it can be perceived confusion between the *Cerrado*, *Pasture* and *NCC* samples, and also between *maize* and *cotton*.

Notice that some parcels of image O_b belonging to classes maize and cotton were classified as Eucalyptus. This error might have been caused by

the temporal displacement between the image acquisition date and the date the reference data refers to.

All classification maps presented the *salt-and-pepper* effect, typical of pixel-wise classification approaches, especially in the map produced from SAR, (Figure 5.9-e -S_a). This effect was particularly significant in the results obtained by the *monotemporal* (Figure 5.9-i - $G[S_a]$) and *multitemporal* (Figure 5.9-k - $G[S_aS_b]$) approaches, and declined for the variants (Figure 5.9-m - $G[S_aO_b]$) and (Figure 5.9-o - $G[S_aS_bO_b]$).

The last two rows show the improvements brought by the inclusion of more conditional variables in the cGANs scheme. In fact, the *salt-and-pepper* effect became nearly imperceptible in some parcels, e.g., in parcels belonging to *maize* and *cotton* classes in image Figure 5.9-o ($G[S_aS_bO_b]$).

Figure 5.9 shows how close were the synthesized images to the corresponding real ones in terms of the spectral information and the geometry of the objects. For example, the structure of rivers and crop parcels were preserved in most cases. Indeed, the synthesized images were not perfect as some regions did not match their correspondent in the real image.

It is notorious that the *multitemporal* approaches produced images closer to the real ones than the *monotemporal* approach. These results also support the hypothesis that the inclusion of conditioning variables that convey multitemporal relations into the cGAN framework helps to improve the synthetic images.

Figure 5.10 shows snips of the results from a particular location of a *cloudy* region of *Rio Branco* dataset produced by all tested variants. The first and third columns show the evaluated images in NIRGB composition, whereas the second and fourth columns show their correspondent classification maps. For this dataset we didn't use the RGB composition because it is easier to identify visually the wildfire samples from the NIRGB composition. As in the results for *Campo Verde* experiments, we present here a snip of the VH polarization band together with its prediction map.

Comparing the snip of the NIRGB composition of image O_a (Figure 5.10-b) with the corresponding reference (Figure 5.10-a), we notice a set of pixels spread over the image that are not labeled as wildfires, but visually present a similar spectral response to wildfires. In fact, the classification map of image O_a (Figure 5.10-c) shows that the RF classifier assigned many of these pixels to the wildfire class. This scenario illustrates the complexity of the problem.

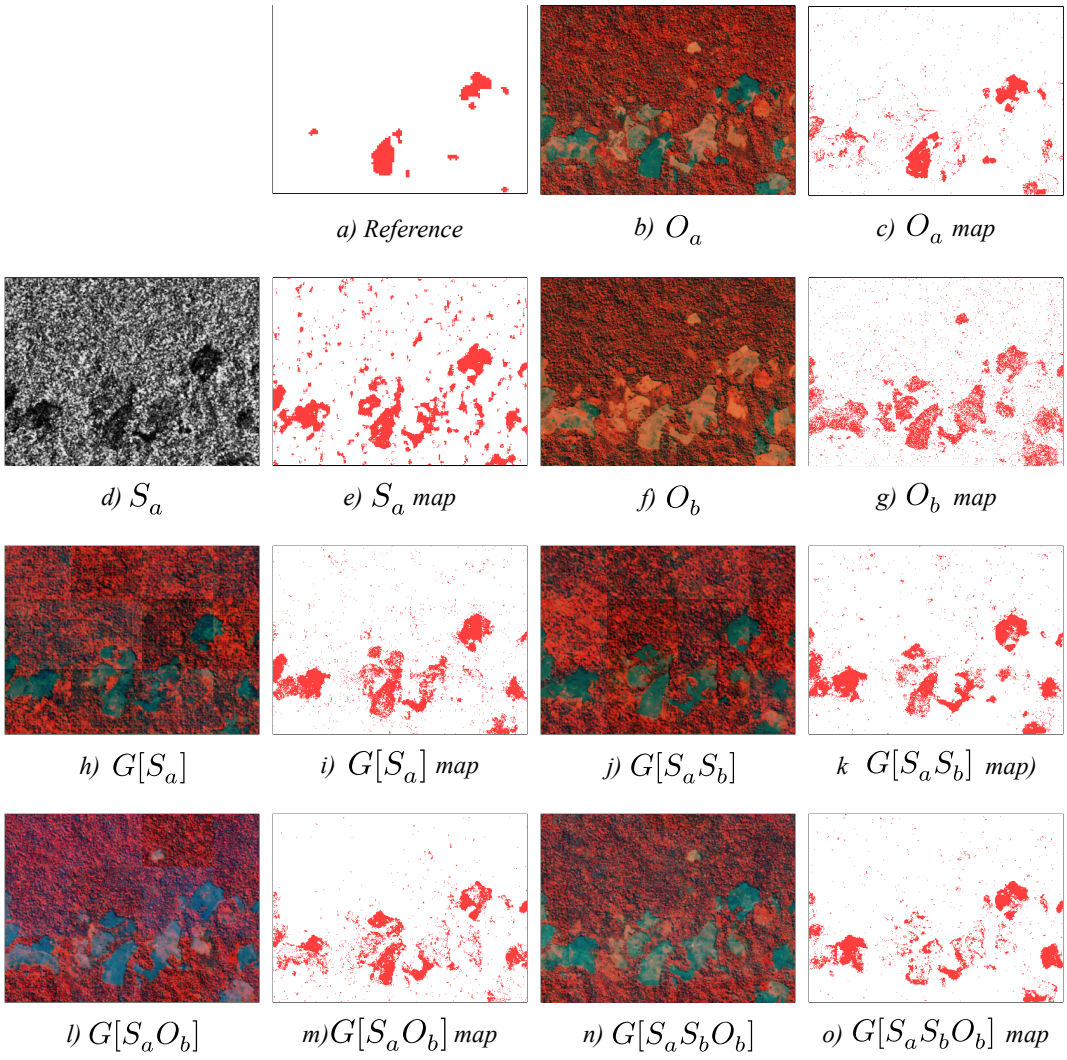

Figure 5.10: Snips of the evaluated images (the original images and the images synthesized by each of the tested variants), and the corresponding classification maps delivered by the RF classifier over the *cloudy* pixels for *Rio Branco* dataset. The RF was trained upon each of these images. The snips of optical images correspond to the RGB composition band. The contrast was adjusted for better visualization.

Figure 5.10-*d* and Figure 5.10-*e* show the SAR image and its associated classification map, respectively. The difficulty to interpret SAR data in comparison to optical images is even more evident in this dataset. Therefore, the images synthesized by the proposed methods can also help to support in the visual inspection of this data.

The classification of the SAR image (Figure 5.10-*e*) presented the worst classification performance among all evaluated images considering the *salt-and-pepper* effect. It is most pronounced in this image, following the same pattern already observed in the results for *Campo Verde* dataset.

As for image O_b, the changes in spectral appearance relative to O_a are substantial. These difference in appearance explain the poor classification performance of this image, which was the worst among other evaluated alternatives for this dataset (see Figure 5.8). Particularly, it can be seen that from t_a to t_b most of the affected areas started recovering from wildfire. Also, it can be distinguished the emergence of possible new damaged regions, which augment even more the complexity of the problem. In fact, the classifier predicted both the recovered and the possible new damaged zones as wildfire.

The snips in the last two rows refer to images synthesized by cGANs models. Among the NIRGB compositions, the *monotemporal* variant (Figure 5.10h) presented the worst image quality. Similar to what occurred in *Campo Verde*, this image is blurred and the boundaries of patches, after mosaicking them, are easily recognizable. Furthermore, on some areas that started recovering, the cGANs generated images with similar spectral appearance of wildfire spots. In consequence, these regions were also classified as wildfires.

Contrarily, the shortcomings discerned in the *monotemporal* image (Figure 5.10-h - $G[S_a]$) were less pronounced in the results produced by the *multitemporal* variants, especially in the images synthesized by the cGANs models that included the image O_b (Figure 5.10$-l$ and Figure 5.10$-n$) as conditioning variable. In those cases, the optical image (O_b) of the other date improved considerably the quality of the synthesized output. For instance, the cGANs were able to discern between wildfires and recovered areas, where the *monotemporal* model assigned both to wildfire. In the same way, the *salt-and-pepper* effect diminished in the generated classification maps.

5.5.3 Similarity Metrics

Table 5.7 and Table 5.8 summarize the results of computing the similarity metrics over the cloudy pixels, between the target (O_a) and synthesized optical images ($G[S_a]$, $G[S_aS_bO_b]$, $G[S_aO_b]$ and $G[S_aS_b]$) for *Campo Verde* and *Rio Branco* datasets, respectively. For completeness, O_b was also included in the table as a baseline, i.e., to verify if using a cloud-free image from other date would be enough. The similarity was measured in terms of the following metrics: *Root Mean Square Error* (RMSE), *Spectral Angle Mapper* (SAM), and *Peak Signal-to-Noise Ratio* (PSNR) as presented in Section 5.2. Recall that the most similar image has low values of RMSE and SAM, and high values of PSNR.

For both datasets, the similarity metrics improved as more conditioning variables have been used to synthesize the output image, in particular when

Table 5.7: Similarity metrics for *Campo Verde* dataset.

Metrics	O_b	$G[S_a]$	$G[S_aS_b]$	$G[S_aO_b]$	$G[S_aS_bO_b]$
RMSE	9217	1424	581	488	**471**
SAM	16,32	5,67	2,22	1,84	**1,78**
PSNR(dB)	16,35	29,69	35,83	36,73	**37,11**

Table 5.8: Similarity metrics for *Rio Branco* dataset.

Metrics	O_b	$G[S_a]$	$G[S_aS_b]$	$G[S_aO_b]$	$G[S_aS_bO_b]$
RMSE	5374	1122	993	849	**737**
SAM	9.68	6,30	5,61	5,68	**4,72**
PSNR(dB)	19.30	31,97	32,61	34,43	**35,08**

the optical image of the other date (O_b) was included. In contrast, the real optical image acquired at date t_b (O_b) achieved the worst results.

Indeed, the difference in performance between (O_b) and $G[S_aS_bO_b]$ was remarkable for both datasets. Particularly, the results for image $G[S_aS_bO_b]$ were 19.56 and 7.29 times lower in terms of RMSE, 9.16 and 2.1 times lower in terms of SAM, and 21dB and 15.8dB higher in terms of PSNR, when compared to image (O_b) for *Campo Verde* and *Rio Branco* dataset, respectively.

Concerning the images synthesized by the cGANs, the *monotemporal* ($G[S_a]$) approach presented the worst similarity values, followed, in increasing order, by the *multitemporal* variants, $G[S_aS_b]$, $G[S_aO_b]$ and $G[S_aS_bO_b]$. Here, the images $G[S_aS_bO_b]$) outperformed their counterparts $G[S_a]$ in up to 3 times in terms of RMSE and SAM metrics, and up to 7dB in terms of PSNR.

Although the images synthesized using all three conditioning data ($G[S_aS_bO_b]$) presented the best similarity values, the results of $G[S_aO_b]$ variants were not far behind. These results are compatible with those reported in Figure 5.7 and Figure 5.8, where similar patterns in the OA and F_1-*score* values were observed.

It is worth emphasizing the substantial differences between images O_b and those synthesized by the cGANs models. This supports the conclusion that in our experiments the cGANs did not merely copy the information of O_b, when this image was used as conditioning data.

In comparison with the results reported in Figure 5.7 and Figure 5.8, where $G[S_a]$ performed better than O_b for *Rio Branco* and similar for *Campo Verde*, in terms of the similarity metrics, the values of images $G[S_a]$ were consistently better than that of images O_b for both datasets. These results show that the *monotemporal* approach is still a good alternative to deal with the problem of missing data in optical imagery when just SAR data at the same date is available.

It is also important to stress the consistency among the results in Tables 5.7 and 5.8 and in the corresponding Figures 5.7 and Figure 5.8 for all evaluated variants. In summary, these results consistently support our working hypothesis that the more conditioning variables, the more realistic are the optical images synthesized by cGANs.

Figure 5.11 and Figure 5.12 show the heatmaps of RMSE and SAM metrics from the same image locations of the snips of Figure 5.9 and Figure 5.10, respectively. The heatmaps indicate for each pixel site how different are the synthesized and the reference spectral values. A darker color is associated with similar spectral information in terms of the computed metric, while brighter colors represent higher dissimilarity. To facilitate the analysis, the figures also shows the correspondent target images (O_a) and their references.

The heatmaps show that image O_b was the most discrepant one in relation to the target. It exhibited the highest *RMSE* and *SAM* values. In *Campo Verde*, *RMSE* was almost constant at a high value for all sites. In terms of SAM, the major variations took place over rivers and some specific crop fields. For *Rio Branco*, the major differences for both metrics were located over regions affected by wildfires. These results are consistent with Figure 5.10-f, where regions more recently affected by wildfire can be distinguished from recovering regions.

With regards to the synthesized images, the *monotemporal* approach (Figure 5.11-f and Figure 5.12-f) presented the highest *RMSE* and *SAM* values for both datasets. Particularly, this behavior was more evident for *Campo Verde* dataset than *Rio Branco*, where the differences among synthesized images are not so large as for *Campo Verde*. These results are consistent with the values reported in Table 5.7 and Table 5.8.

Interestingly, some of the synthesized images of *Campo Verde* are patchy: some synthetic patches are closer to the correspondent reference than others. This behavior is possibly related to the spectral variability over the crop phenological stages. In this case, the cGAN has possibly generated samples from the same class, but in a different phenological stage.

Regarding the *Rio Branco* dataset, the largest deviation of the synthesized images from the reference occurred over wildfire samples. This pattern is more apparent in the SAM heatmaps, where some pixels present high scores for this metric.

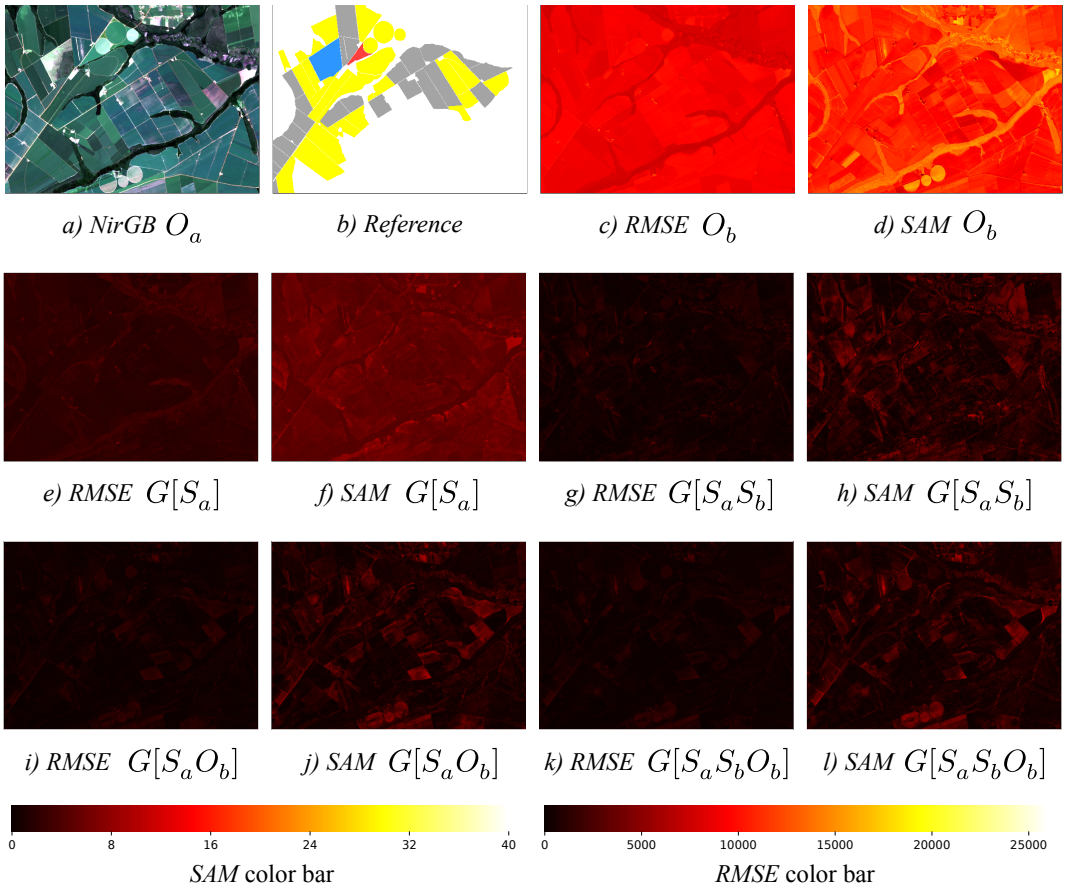

Figure 5.11: Snip of heatmaps of RMSE and SAM metrics from the same image locations of the snips of Figure 5.9 for *Campo Verde* dataset.

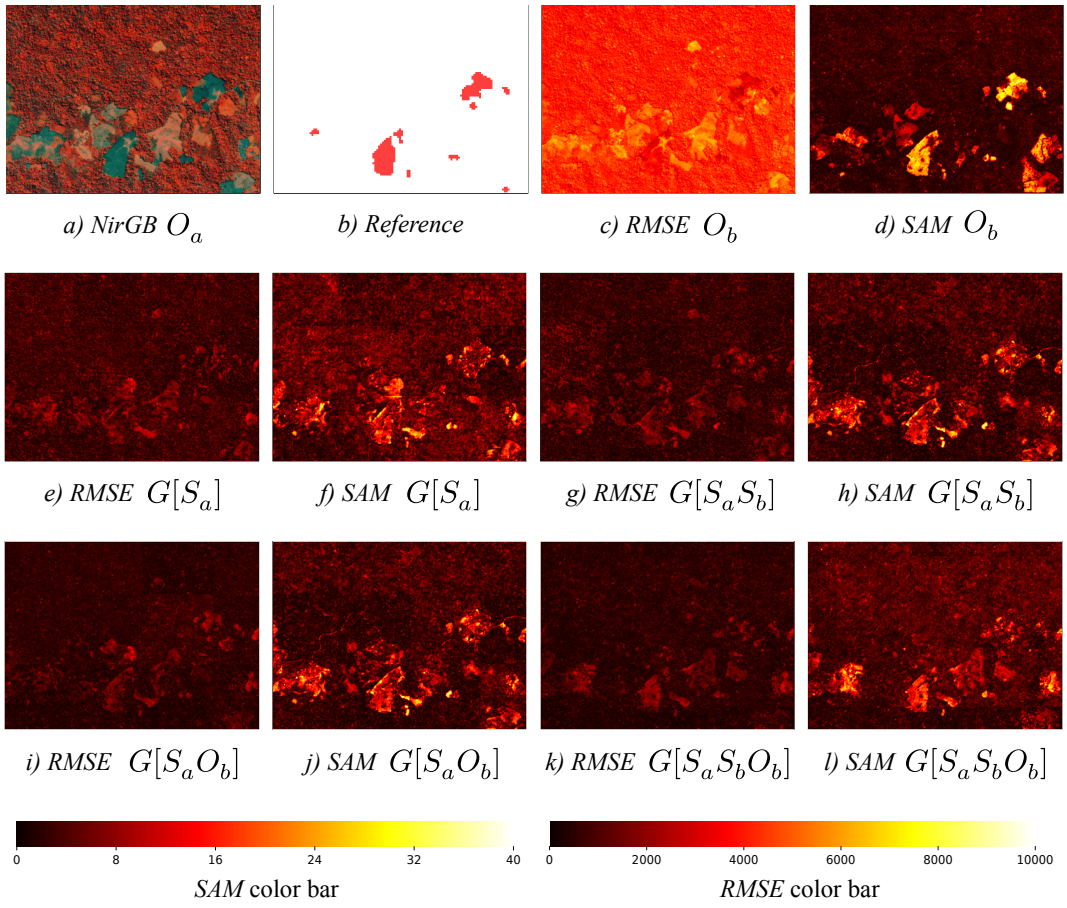

Figure 5.12: Snip of heatmaps of RMSE and SAM metrics from the same image locations of the snips of Figure 5.10 for *Rio Branco* dataset.

6
CONCLUSIONS

In this work, we proposed a framework based on cGANs to synthesize Remote Sensing optical data from multisensor, multitemporal and multiresolution data. Particularly, the framework is able to generate missing optical data due to thick clouds that totally block the spectral information. In fact, this approach can be applied to any other problem that causes data missing or corrupting in some regions of an image.

The proposal was evaluated in terms of classification as well as in terms of spectral similarity of the generated images in relation to an optical ground truth. The experiments were carried out on two datasets from two municipalities in Brazil. The first dataset focus on crop recognition, while the second one on wildfire detection.

The experimental analysis showed that the accuracy of a pixel-wise classification conducted on the optical images generated by the proposed strategy was close to the results achieved on the reference cloud-free images for both applications.

The experiments confirmed the working hypothesis that the quality of synthesized images improved as more data was added to condition the cGAN operation. In particular, information from the same optical sensor at another date improved significantly the classification performance.

It is important to emphasize that the method does not create any new information that is not already embedded in the conditioning data. It just learns more discriminative representations than conventional features.

The conclusions drawn from the similarity analysis were consistent with the results observed in the experiments for image classification. In addition, a visual inspection leads to the same conclusions.

The obtained results open up multiple possibilities for future work. We believe that even better results might be achieved by fine-tuning the networks' architectures and their hyperparameters. In particular, semisupervised techniques for cGAN training may also improve the quality of synthesized images. Above all, we believe that the choice of additional conditioning variables for the cGAN design for different applications constitutes a promising research direction.

Although the proposed method was initially structured for the context of cloud removal, it can easily be extended to other problems of missing data. For instance, it can be used for filling corrupted data commonly present in Time Series Cube Data applications [72]. Future works can be addressed to explore the capability of the proposed method in this scenario. Finally, another possible application to explore is the recovering missing data in Landsat 7 imagery affected by acquisition errors in the SLC sensor.

Bibliography

1 ENOMOTO, K.; SAKURADA, K.; WANG, W.; FUKUI, H.; MATSUOKA, M.; NAKAMURA, R. ; KAWAGUCHI, N.. **Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets**. arXiv preprint arXiv:1710.04835, 2017.

2 SINGH, P.; KOMODAKIS, N.. **Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks**. In: IGARSS 2018-2018 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, p. 1772–1775. IEEE, 2018.

3 GROHNFELDI, C.; SCHMITT, M. ; ZHU, X.. **A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images**. In: IGARSS 2018-2018 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, p. 1726–1729. IEEE, 2018.

4 ENOMOTO, K.; SAKURADA, K.; WANG, W.; KAWAGUCHI, N.; MATSUOKA, M. ; NAKAMURA, R.. **Image translation between sar and optical imagery with generative adversarial nets**. In: IGARSS 2018-2018 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, p. 1752–1755. IEEE, 2018.

5 BERMUDEZ, J.; HAPP, P.; OLIVEIRA, D. ; FEITOSA, R.. **Sar to optical image synthesis for cloud removal with generative adversarial networks.** ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4(1), 2018.

6 ACHANCCARAY DIAZ, P.; QUEIROZ FEITOSA, R. ; DEL ARCO SANCHES, I.. **Crop recognition in tropical regions based on spatio-temporal conditional random fields from multi-temporal and multi-resolution sequences of remote sensing images**. 2019.

7 TEMPFLI, K.; HUURNEMAN, G.; BAKKER, W.; JANSSEN, L. L.; FERINGA, W.; GIESKE, A.; GRABMAIER, K.; HECKER, C.; HORN, J.; KERLE, N. ; OTHERS. **Principles of remote sensing: an introductory textbook**. ITC, 2009.

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

9 SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I. ; SALAKHUTDINOV, R.. **Dropout: a simple way to prevent neural networks from overfitting**. The Journal of Machine Learning Research, 15(1):1929–1958, 2014.

10 MATESE, A.; TOSCANO, P.; DI GENNARO, S. F.; GENESIO, L.; VACCARI, F. P.; PRIMICERIO, J.; BELLI, C.; ZALDEI, A.; BIANCONI, R. ; GIOLI, B.. **Intercomparison of uav, aircraft and satellite remote sensing platforms for precision viticulture**. Remote Sensing, 7(3):2971–2990, 2015.

11 JABAR, A.; SADIQ, A.; SULONG, G. ; GEORGE, L. E.. **Survey on gap filling algorithms in landsat 7 etm+ images.** Journal of Theoretical & Applied Information Technology, 63(1), 2014.

12 LI, X.; SHEN, H.; ZHANG, L.; ZHANG, H. ; YUAN, Q.. **Dead pixel completion of aqua modis band 6 using a robust m-estimator multiregression**. IEEE Geoscience and Remote Sensing Letters, 11(4):768–772, 2014.

13 SCARAMUZZA, P.; BARSI, J.. **Landsat 7 scan line corrector-off gap-filled product development**. In: PROC. PECORA, p. 23–27, 2005.

14 ROSSOW, W. B.. **International satellite cloud climatology project**. 2011.

15 HERMELINGMEIER, C.. **The competitive firm and the role of information about uncertain factor prices**. Economic Modelling, 27(2):547–552, 2010.

16 EBERHARDT, I.; SCHULTZ, B.; RIZZI, R.; SANCHES, I.; FORMAGGIO, A.; ATZBERGER, C.; MELLO, M.; IMMITZER, M.; TRABAQUINI, K.; FOSCHIERA, W. ; OTHERS. **Cloud cover assessment for operational crop monitoring systems in tropical areas**. Remote Sensing, 8(3):219, 2016.

17 PENHA, T. V.; GARCIA, L. M. F. ; SEHN, T. K.. **Detecção de áreas queimadas utilizando imagens multi-sensores de média resolução espacial, técnicas de geobia e mineração de dados na amazônia brasileira**. National Institute for Space Research, 2018.

18 ANDERSON, L. O.; ARAGÃO, L. E.; GLOOR, M.; ARAI, E.; ADAMI, M.; SAATCHI, S. S.; MALHI, Y.; SHIMABUKURO, Y. E.; BARLOW, J.; BERENGUER, E. ; OTHERS. **Disentangling the contribution of multiple land covers to fire-mediated carbon emissions in amazonia during the 2010 drought**. Global biogeochemical cycles, 29(10):1739–1753, 2015.

19 BARLOW, J.; PERES, C. A.; LAGAN, B. O. ; HAUGAASEN, T.. **Large tree mortality and the decline of forest biomass following amazonian wildfires**. Ecology letters, 6(1):6–8, 2003.

20 BARLOW, J.; PERES, C. A.. **Fire-mediated dieback and compositional cascade in an amazonian forest**. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1498):1787–1794, 2008.

21 PIVELLO, V. R.. **The use of fire in the cerrado and amazonian rainforests of brazil: past and present**. Fire ecology, 7(1):24–39, 2011.

22 SMITH, L. T.; ARAGAO, L. E.; SABEL, C. E. ; NAKAYA, T.. **Drought impacts on children's respiratory health in the brazilian amazon**. Scientific reports, 4:3726, 2014.

23 LI, Y.; LI, W. ; SHEN, C.. **Removal of optically thick clouds from high-resolution satellite imagery using dictionary group learning and interdictionary nonlocal joint sparse coding**. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(5):1870–1882, 2017.

24 LI, X.; WANG, L.; CHENG, Q.; WU, P.; GAN, W. ; FANG, L.. **Cloud removal in remote sensing images using nonnegative matrix factorization and error correction**. ISPRS Journal of Photogrammetry and Remote Sensing, 148:103–113, 2019.

25 CHENG, Q.; SHEN, H.; ZHANG, L.; YUAN, Q. ; ZENG, C.. **Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal mrf model**. ISPRS Journal of Photogrammetry and Remote Sensing, 92:54–68, 2014.

26 HOAN, N. T.; TATEISHI, R.. **Cloud removal of optical image using sar data for alos applications. experimenting on simulated alos data**. Journal of The Remote Sensing Society of Japan, 29(2):410–417, 2009.

27 XU, M.; JIA, X. ; PICKERING, M.. **Automatic cloud removal for landsat 8 oli images using cirrus band**. In: 2014 IEEE GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, p. 2511–2514. IEEE, 2014.

28 ALOM, M. Z.; TAHA, T. M.; YAKOPCIC, C.; WESTBERG, S.; SIDIKE, P.; NASRIN, M. S.; HASAN, M.; VAN ESSEN, B. C.; AWWAL, A. A. ; ASARI, V. K.. **A state-of-the-art survey on deep learning theory and architectures**. Electronics, 8(3):292, 2019.

29 BALL, J. E.; ANDERSON, D. T. ; CHAN, C. S.. **Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community**. Journal of Applied Remote Sensing, 11(4):042609, 2017.

30 GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A. ; BENGIO, Y.. **Generative adversarial nets**. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, p. 2672–2680, 2014.

31 PARK, T.; LIU, M.-Y.; WANG, T.-C. ; ZHU, J.-Y.. **Semantic image synthesis with spatially-adaptive normalization**. arXiv preprint arXiv:1903.07291, 2019.

32 ZHU, J.-Y.; PARK, T.; ISOLA, P. ; EFROS, A. A.. **Unpaired image-to-image translation using cycle-consistent adversarial networkss**. In: COMPUTER VISION (ICCV), 2017 IEEE INTERNATIONAL CONFERENCE ON, 2017.

33 HE, C.; XIONG, D.; ZHANG, Q. ; LIAO, M.. **Parallel connected generative adversarial network with quadratic operation for sar image generation and application for classification**. Sensors, 19(4):871, 2019.

34 SINGH, P.; KOMODAKIS, N.. **Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks**. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2018 IEEE INTERNATIONAL. IEEE, 2018.

35 GROHNFELDT, C.; SCHMITT, M. ; ZHU, X.. **A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images**. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2018 IEEE INTERNATIONAL. IEEE, 2018.

36 ENOMOTO, K.; SAKURADA, K.; WANG, W.; KAWAGUCHI; MATSUOKA, M. ; NAKAMURA, R.. **Image translation between sar and optical imagery with generative adversarial nets**. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2018 IEEE INTERNATIONAL. IEEE, 2018.

37 LEY, A.; DHONDT, O.; VALADE, S.; HAENSCH, R. ; HELLWICH, O.. **Exploiting gan-based sar to optical image transcoding for improved classification via deep learning**. In: EUSAR 2018; 12TH EUROPEAN CONFERENCE ON SYNTHETIC APERTURE RADAR, p. 1–6. VDE, 2018.

38 GAO, Y.; WANG, Y. ; LV, H.. **Extendibility of a thin-cloud removal algorithm to hi-resolution visible bands of sentinel-2 data**. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2018 IEEE INTERNATIONAL. IEEE, 2018.

39 SHEN, H.; LI, H.; QIAN, Y.; ZHANG, L. ; YUAN, Q.. **An effective thin cloud removal procedure for visible remote sensing images**. ISPRS Journal of Photogrammetry and Remote Sensing, 96:224–235, 2014.

40 LIU, J.; WANG, X.; CHEN, M.; LIU, S.; ZHOU, X.; SHAO, Z. ; LIU, P.. **Thin cloud removal from single satellite images**. Optics express, 22(1):618–632, 2014.

41 SOPHIA, D. L.; LALITHA, K. ; CHANDAR, J. P.. **Reconstruction of cloud contaminated remote sensing images using inpainting strategy**. International Journal of Electronics Communication and Computer Technology, 3(3):407–411, 2013.

42 GUILLEMOT, C.; LE MEUR, O.. **Image inpainting: Overview and recent advances**. IEEE signal processing magazine, 31(1):127–144, 2014.

43 LIN, C.-H.; TSAI, P.-H.; LAI, K.-H. ; CHEN, J.-Y.. **Cloud removal from multitemporal satellite images using information cloning**. IEEE transactions on geoscience and remote sensing, 51(1):232–241, 2013.

44 GÓMEZ-CHOVA, L.; AMORÓS-LÓPEZ, J.; MATEO-GARCÍA, G.; MUÑOZ-MARÍ, J. ; CAMPS-VALLS, G.. **Cloud masking and removal in remote sensing image time series**. Journal of Applied Remote Sensing, 11(1):015005, 2017.

45 VUOLO, F.; NG, W.-T. ; ATZBERGER, C.. **Smoothing and gap-filling of high resolution multi-spectral time series: Example of landsat**

data. International journal of applied earth observation and geoinformation, 57:202–213, 2017.

46 HUANG, Y.; LIU, H.; YU, B.; WU, J.; KANG, E. L.; XU, M.; WANG, S.; KLEIN, A. ; CHEN, Y.. **Improving modis snow products with a hmrf-based spatio-temporal modeling technique in the upper rio grande basin**. Remote Sensing of Environment, 204:568–582, 2018.

47 YAN, L.; ROY, D.. **Large-area gap filling of landsat reflectance time series by spectral-angle-mapper based spatio-temporal similarity (samsts)**. Remote Sensing, 10(4):609, 2018.

48 KRUSE, F. A.; LEFKOFF, A.; BOARDMAN, J.; HEIDEBRECHT, K.; SHAPIRO, A.; BARLOON, P. ; GOETZ, A.. **The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data**. Remote sensing of environment, 44(2-3):145–163, 1993.

49 MIRZA, M.; OSINDERO, S.. **Conditional generative adversarial nets**. arXiv preprint arXiv:1411.1784, 2014.

50 PATHAK, D.; KRAHENBUHL, P.; DONAHUE, J.; DARRELL, T. ; EFROS, A. A.. **Context encoders: Feature learning by inpainting**. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, p. 2536–2544, 2016.

51 ZHANG, Z.; SONG, Y. ; QI, H.. **Age progression/regression by conditional adversarial autoencoder**. In: THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), volumen 2, 2017.

52 ISOLA, P.; ZHU, J.-Y.; ZHOU, T. ; EFROS, A. A.. **Image-to-image translation with conditional adversarial networks**. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, p. 1125–1134, 2017.

53 TOTH, C.; JÓŹKÓW, G.. **Remote sensing platforms and sensors: A survey**. ISPRS Journal of Photogrammetry and Remote Sensing, 115:22–36, 2016.

54 EASTMAN, J.. **Introduction to remote sensing and image processing**. Idrisi for Windows User's Guide. Cap, 3, 2001.

55 LIPTÁK, I.; ERDÉLYI, J.; KYRINOVIČ, P. ; KOPÁČIK, A.. **Monitoring of bridge dynamics by radar interferometry**. Geoinformatics FCE CTU, 12:10–15, 2014.

56 ERTEN, E.; LOPEZ-SANCHEZ, J. M.; YUZUGULLU, O. ; HAJNSEK, I.. **Retrieval of agricultural crop height from space: A comparison of sar techniques**. Remote Sensing of Environment, 187:130–144, 2016.

57 GOODFELLOW, I.; BENGIO, Y. ; COURVILLE, A.. **Deep Learning**. MIT Press, 2016. `http://www.deeplearningbook.org`.

58 IOFFE, S.; SZEGEDY, C.. **Batch normalization: Accelerating deep network training by reducing internal covariate shift**. arXiv preprint arXiv:1502.03167, 2015.

59 GOODFELLOW, I.. **Nips 2016 tutorial: Generative adversarial networks**. arXiv preprint arXiv:1701.00160, 2016.

60 JEBARA, T.; PENTLAND, A. P.. **Discriminative, generative and imitative learning**. PhD thesis, PhD thesis, Media laboratory, MIT, 2001.

61 BERMUDEZ, J. D.; HAPP, P. N.; FEITOSA, R. Q. ; OLIVEIRA, D. A.. **Synthesis of multispectral optical images from sar/optical multitemporal data using conditional generative adversarial networks**. IEEE Geoscience and Remote Sensing Letters, 2019.

62 ZHU, Z.; WOODCOCK, C. E.. **Object-based cloud and cloud shadow detection in landsat imagery**. Remote sensing of environment, 118:83–94, 2012.

63 LOUIS, J.; DEBAECKER, V.; PFLUG, B.; MAIN-KNORN, M.; BIENIARZ, J.; MUELLER-WILM, U.; CADAU, E. ; GASCON, F.. **Sentinel-2 sen2cor: L2a processor for users**. In: PROCEEDINGS OF THE LIVING PLANET SYMPOSIUM, PRAGUE, CZECH REPUBLIC, p. 9–13, 2016.

64 SANCHES, I.; FEITOSA, R. Q.; DIAZ, P. M. A.; SOARES, M. D.; LUIZ, A. J.; SCHULTZ, B. ; MAURANO, L. E.. **Campo verde database: Seeking to improve agricultural remote sensing of tropical areas**. IEEE Geoscience and Remote Sensing Letters, 15(3):369–373, 2018.

65 BORMA, L. D. S.; NOBRE, C. A.. **Secas na Amazônia: causas e consequências**. Oficina de Textos, 2016.

66 SASAKI, Y.; OTHERS. **The truth of the f-measure**. Teach Tutor mater, 1(5):1–5, 2007.

67 ABADI, M.; BARHAM, P.; CHEN, J.; CHEN, Z.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; IRVING, G.; ISARD, M. ; OTHERS. **Tensorflow: A system for large-scale machine learning**. In: 12TH {USENIX} SYMPOSIUM ON OPERATING SYSTEMS DESIGN AND IMPLEMENTATION ({OSDI} 16), p. 265–283, 2016.

68 ODENA, A.; DUMOULIN, V. ; OLAH, C.. **Deconvolution and checkerboard artifacts**. Distill, 1(10):e3, 2016.

69 KINGMA, D. P.; BA, J.. **Adam: A method for stochastic optimization**. arXiv preprint arXiv:1412.6980, 2014.

70 LIAW, A.; WIENER, M. ; OTHERS. **Classification and regression by randomforest**. R news, 2(3):18–22, 2002.

71 BELGIU, M.; DRĂGUŢ, L.. **Random forest in remote sensing: A review of applications and future directions**. ISPRS Journal of Photogrammetry and Remote Sensing, 114:24–31, 2016.

72 LEWIS, A.; LYMBURNER, L.; PURSS, M. B.; BROOKE, B.; EVANS, B.; IP, A.; DEKKER, A. G.; IRONS, J. R.; MINCHIN, S.; MUELLER, N. ; OTHERS. **Rapid, high-resolution detection of environmental change over continental scales from satellite data–the earth observation data cube**. International Journal of Digital Earth, 9(1):106–111, 2016.