SAR TO OPTICAL IMAGE SYNTHESIS FOR CLOUD REMOVAL WITH GENERATIVE ADVERSARIAL NETWORKS

J. D. Bermudez[1]*, P. N. Happ[1], D. A. B. Oliveira[2], R. Q. Feitosa[1,3]

[1] Pontifical Catholic University of Rio de Janeiro, Brazil - (bermudez, patrick, raul)@ele.puc-rio.br
[2] IBM Research - darioaugusto@gmail.com
[3] Rio de Janeiro State University, Brazil

Commission I, WG I/6

KEY WORDS: Cloud Removal, Conditional Generative Adversarial Networks, Deep Learning, Multispectral Images, SAR

ABSTRACT:

Optical imagery is often affected by the presence of clouds. Aiming to reduce their effects, different reconstruction techniques have been proposed in the last years. A common alternative is to extract data from active sensors, like Synthetic Aperture Radar (SAR), because they are almost independent on the atmospheric conditions and solar illumination. On the other hand, SAR images are more complex to interpret than optical images requiring particular handling. Recently, Conditional Generative Adversarial Networks (cGANs) have been widely used in different image generation tasks presenting state-of-the-art results. One application of cGANs is learning a nonlinear mapping function from two images of different domains. In this work, we combine the fact that SAR images are hardly affected by clouds with the ability of cGANS for image translation in order to map optical images from SAR ones so as to recover regions that are covered by clouds. Experimental results indicate that the proposed solution achieves better classification accuracy than SAR based classification.

1. INTRODUCTION

With the launch of more satellites, with higher spatial resolution and lower revisiting time, remote sensing data became a cost-effective solution for many applications such as agricultural mapping, urban planning, disaster management, weather forecasting, etc. However, most of these applications can be affected by the presence of clouds in optical imagery from passive sensors, especially in tropics and temperate regions, where there is ten to twenty percent more cloud coverage than in the subtropics and the polar regions (Rossow, 2011).

An alternative to the cloud coverage problem is the usage of images from active sensors, like Synthetic Aperture Radar (SAR), which almost do not depend on the atmospheric conditions neither on the solar illumination (Li et al., 2017). However, the information captured by them is less descriptive and more complex to interpret than in optical images. Thus, reconstruction techniques have been proposed and used in an attempt to reduce the effect of clouds in optical imagery. However, there is still no method able to completely solve this problem.

Cloud removal techniques can be categorized into monotemporal and multitemporal based (Xu et al., 2016). Monotemporal-based techniques use the multispectral bands' information from the affected image in order to recover the regions covered by clouds, while multitemporal-based ones use information from other co-registered images acquired at different dates. For monotemporal-based, image filtering approaches are the most used techniques. They are based on the fact that clouds are majorly composed by spectral low-frequency components and then, in theory, they can be removed via a high-pass filtering (Shen et al., 2014). 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 only employed to remove thin clouds.

On the contrary, multitemporal approaches have the capacity of dealing with both thin and thick clouds. They use information from other cloud-free images of the same location in order to reconstruct the regions contaminated by clouds. The most simple 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 (Cheng et al., 2014). Later, a post-processing step is needed to reduce 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 sequence of images of the same region in order to build a time series model which can be used to infer pixels covered by clouds (Gómez-Chova et al., 2017). The main problem in this approach is to acquire enough cloud-free images in the different epochs.

Generative Adversarial Networks (GANs) were firstly introduced in (Goodfellow et al., 2014) and have been widely investigated since then by the computer vision community. More recently, conditional Generative Adversarial Networks (cGANs) (Mirza and Osindero, 2014) have been broadly used in different image generation tasks, such as image inpainting (Pathak et al., 2016), image manipulation (Zhang et al., 2017), and image translation (Isola et al., 2017). For image translation, for instance, a cGAN learns a nonlinear mapping function capable to transform an image from one domain to a version of the same image in another domain. Based on that, (Enomoto et al., 2017) employs a cGAN to recover visible light RGB images from multispectral images. Basically, by using a set of multispectral set of cloud-free ground truth images, it is created an associated set of images covered by clouds through a cloud synthesizing algorithm. Then, a

*Corresponding author

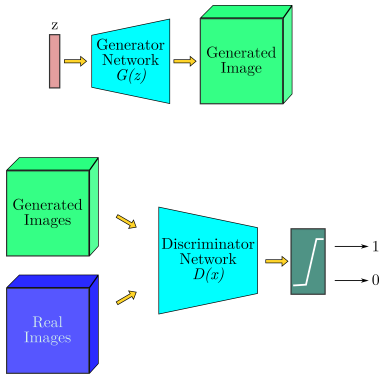


Figure 1. GANs training procedure. The Generator learns a function G that map from a random noise vector z to an output image. The Discriminator learn to classify between real and fake images.

cGAN model is trained to map from images covered by clouds to the corresponding cloud-free ground truth images. However, this approach presents problems in dealing with thick clouds and also with white objects, which are in appearance similar to clouds. Additionally, in the referred work, the analysis of results is done subjectively without considering any performance metric.

Similar to (Enomoto et al., 2017), we want to to explore the ability of cGANs to reconstruct images covered by clouds, but following a different approach. Specifically, we take advantage of the fact that SAR images are almost not affected by clouds to learn, via a cGANs model, a nonlinear mapping function that maps SAR images to optical cloud-free images.

The objective of this paper is to explore the capability of cGANs to map multispectral cloud-free optical images from co-registered SAR images. Our method can be used to remove both thin and thick clouds since it depends only on SAR data, which almost is not affected by clouds. Additionally, our method is not restricted to outputs just visible as RGB images, it also has the capability to reconstruct other spectral bands, which are essential in many remote sensing applications. Finally, we present both a subjective and an objective analysis of the quality of generated images. The first is done via visual inspections and the second by using a Random Forest (RF) classifier.

The remainder of this paper is organized as follows. Section 2 explains the fundamentals of GANs. Section 3 introduces the methodology for cloud removal proposed in this work. Section 4 presents the datasets used in our experiments, the features extracted from them and the experimental protocol. Section 5 shows and discusses the results obtained in our experiments. Finally, Section 6 summarizes the conclusions drawn from our results and future works.

2. FUNDAMENTALS

2.1 Generative Adversarial Networks

GANs are generative models composed by two networks: a generator (G) that outputs synthesized images y, and a discriminator (D) that determines if an input image is synthesized or real one. Both networks are trained in a two-players adversarial scheme, as it can be seen in Figure 1: while G tries to learn how to produce realistic images to fool D, D tries to correctly discriminate between synthesized and real images. Formally, given any data distribution $p_{data}(x)$, the generator G learns a distribution

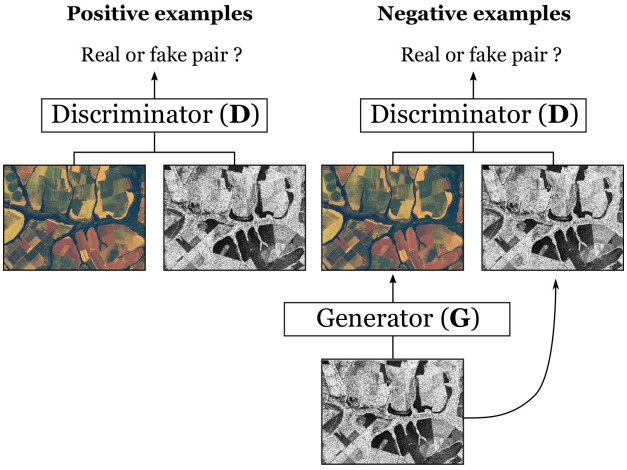


Figure 2. cGANs training procedure. The Discriminator learns to classify between real and fake pairs of images. The Generator learns a mapping function that takes as input a real image and outputs a realistic synthetic image from other domain. Illustration inspired from (Isola et al., 2017).

$p_{model}(w)$ such that the discriminator can hardly distinguish between samples coming from $p_{data}(x)$ and $p_{model}(w)$.

Generally, $p_{model}(w)$ is a complex distribution, so sampling from it is 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 $p_{model}(w)$.

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

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

where $L_{GAN}(G, D)$ is the GAN objective function defined by,

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

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

The solution of Equation 1 is obtained by training the generator G and discriminator D alternately. The discriminator is trained with images produced by the last trained generator and with real images. Similarly, the outcome of the last trained discriminator is used to train the generator. At the end of several training cycles, the generator is supposedly capable of producing images that the discriminator is not able to distinguish from real ones.

2.2 Conditional Generative Adversarial Networks (cGANs)

Conditional GANs, introduced by (Mirza and Osindero, 2014), are an extension of the GANs concept. Basically, in conditional GANs the input to the discriminator consists of samples from two domains (x and y), and the generator synthesizes samples from one of those domains (say y). The loss function for conditional GANs is expressed by Equation 3.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-1-5-2018 | © Authors 2018. CC BY 4.0 License.

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

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

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

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

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

Similar to (Isola et al., 2017), we remove the dependency of the random noise vector z from the cGANS' objective function by applying the dropout regularization on several layers of the generator at both training and test time.

cGANs hold many similarities with the original 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 takes as input a pair of images and learns to correctly identify if they are a real-real or a real-fake pair, while the generator learns to generate synthetic images capable of fooling the discriminator, as described in Figure 2. In cGANs, 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 paper, we use it in the context of cloud removal.

3. METHODOLOGY

This section describes the proposed methodology for cloud removal in optical satellite images. Basically, our method uses SAR data and cGANS to infer the spectral bands' information of optical images partially covered by clouds. We take advantage of the fact that SAR images are almost completely independent of atmospheric conditions and solar illumination to reconstruct cloudless optical images from them. To do that, we train a cGAN model to learn a nonlinear mapping function which takes as input a SAR image and get the corresponding cloudless optical image as output.

Figure 3 summarizes the proposed methodology. In this graph, part of the optical images (gray color circles) represents the area covered by clouds while the rest of the image is supposed to be cloud-free. Given a SAR/optical image pair, with close acquisitions dates, the method follows the pipeline described next,

1. *Identify cloud-free regions*: via visual observation or by using a cloud detection algorithm, the cloud-free regions are selected in order to drive the cGAN to learn to generate only cloud-free images.

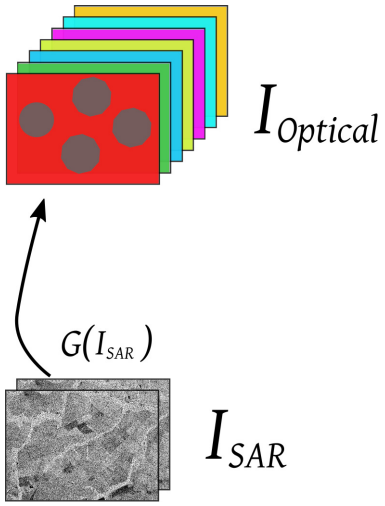


Figure 3. Proposed methodology for cloud removal in optical satellite images. A cGANs is trained to learn nonlinear a mapping function G that map from a SAR image to an optical image. Gray circles represent the regions covered by clouds.

2. *Extract SAR/optical pair of patches*: a set of SAR/optical pairs of patches is extracted from the cloud-free regions. This process can be carried out in two ways: by extracting the patches randomly over the interested region or by using the windows sliding procedure with a fixed stride size. In this paper we used the random approach.

3. *Train the cGAN model*: employing the set of patches extracted from the cloud-free region, the cGAN is trained until convergence.

4. *Generate the cloudless optical image*: once the model has been trained, the generator is used to synthesize the optical cloudless image by taking as input the corresponding SAR data. Because the cGAN model is trained using patches, a mosaic is created from the generated patches to produce the whole optical image. Similar to (Arkadiusz et al., 2017), we adopt the sliding window approach with overlap. This allows building a smoother mosaic by removing weaker predictions on image patch boundaries, where spatial context is generally missing.

In this method, the cloud-free region plays an important role during the training process of the cGANs model because it allows learning the relationship between the SAR data and the correspondent cloud-free optical data. Additionally, it is desirable that the cloud-free region be a representative sample of most of the classes present on the area covered by clouds in order to learn a nonlinear mapping function that can capture all the data variability present on the target image. For instance, if there are classes on the area covered by clouds, which are not present on the cloud-free region, the nonlinear mapping function may not be able to output the correct information, because they were not part of the learning process.

It is also important to note that the difference between the acquisition dates of SAR and optical images should be as short as possible. This is important to reduce the impact of possible changes of classes or even appearance 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.

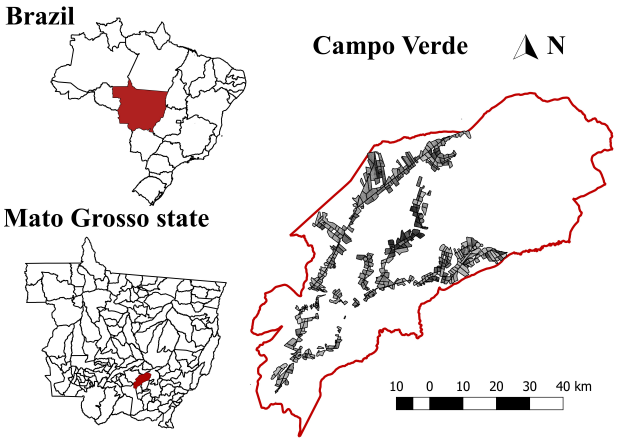


Figure 4. Study area: Campo Verde, Mato Grosso state, Brazil.

4. EXPERIMENTS

The objective of the experiments reported henceforth is to evaluate the ability of the proposed method in removing clouds from optical images by learning a mapping function from each pairs of SAR/Optical images. In short, we present a visual comparison of the original images with the images produced by our method and a numerical result based on the accuracy of a crop classification problem.

4.1 Dataset

The dataset used in our experiments comprises a sequence of 9 co-registered Landsat 8 OLI optical images and 14 co-registered Sentinel-1A SAR images dual polarized (VH and VV), taken between October 2015 and July 2016, from the municipality of Campo Verde in Mato Grosso state, Brazil (Sanches et al., 2018) (see Figure 4). Each image covers an extension of approximately 4782 km^2 with 30m spatial resolution for Landsat images and 10m for the Sentinel-1A. The main crops found in this area are Soybean, Maize and Cotton. Also, there are some minor crops such as Beans and Sorghum. Millet, Brachiaria and Crotalaria were considered as a single class named non-commercial crops (NCC). Other classes present in the dataset are Pasture, Eucalyptus, Soil, Turfgrass and Cerrado. Figure 5 shows the class occurrence per image in the dataset. Observe that the number of crops per image changes along the whole image sequence due to the different phenological cycles of each culture.

4.2 Feature Extraction

For the optical images, we used a feature vector containing the pixel spectral information from bands 1 to 7. For the SAR images we computed features based on the Gray Level Co-occurence Matrix (GLCM). Specifically, four features were computed for the VV and VH bands from 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.

4.3 Networks Architectures

Both Generator and Discriminator network architectures are detailed in the following. We first adapted the network architectures proposed in (Isola et al., 2017) to be capable to work with multispectral optical images, as well as with two-channel (VV and VH) SAR images. In particular, we adopted for the Generator the U-Net (Ronneberger et al., 2015) architecture consisting of 8 convolution layers for encoding and 8 deconvolution layers for decoding, whereas the Discriminator consists of 4 convolutional layers followed by an output sigmoid layer for classification. Both Generator and Discriminator input use 2x2 stride convolution, 5x5 size kernels, ReLU activation functions and Batch Normalization during the training phase. Additionally, the Generator also uses Dropout regularization for each layer of the decoding architecture.

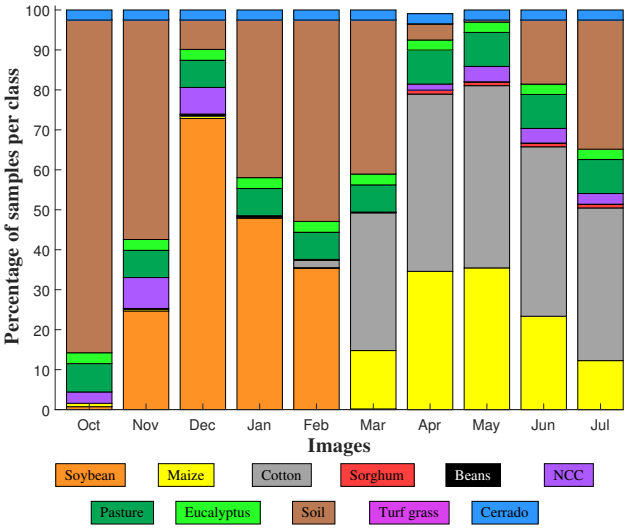


Figure 5. Class occurrences per image in Campo Verde dataset.

4.4 Experimental Protocol

From Campo Verde dataset we selected four SAR/optical pairs of images based on the acquisition date, i.e., each selected pair SAR/optical was acquired approximately at the same date and for each was learned a non-linear mapping function via cGANs. Table 1 lists the four pairs of images chosen by acquisition dates: MAR, JUN, JUL1 and JUL2. To assess the capacity of the

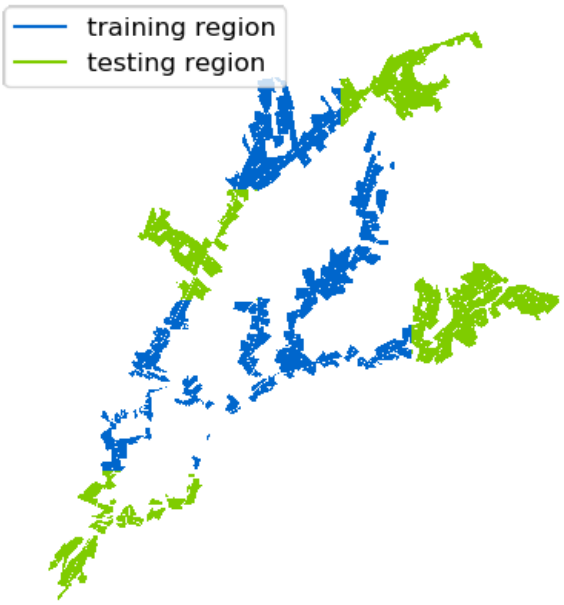


Figure 6. Distribution of training and testing regions used on experiments.

cGANs to map from SAR to optical images, we first defined the regions used for the GAN training phase and for its posterior evaluation using the classifier. The cGANs models were trained approximately with 2000 patches of 256 × 256 pixels extracted from the training region. Figure 6 shows the distribution of these regions. Here, the training region represents the cloud-free area while the evaluation region simulates the areas covered by clouds. Note that regions with no crop information were excluded in order to specialize the cGAN model to map just crop information. In addition, since our SAR and optical images have different spatial resolutions, we downsampled the SAR image to the optical image resolution.

Epochs	SAR	Optical
MAR	21/03/2016	18/03/2016
JUN	08/05/2016	05/05/2016
JUL1	07/07/2016	08/07/2016
JUL2	31/07/2016	24/07/2016

Table 1. SAR/Optical image pairs selected in the experiments.

Next, the generated images were evaluated visually and numerically. The last one was done by classifying the crop areas using a Random Forest classifier (RF). Then, results were compared with the corresponding results obtained by a classification upon ”real” optical and SAR images classifications. Two classifications scenarios were considered: monotemporal and multitemporal. For multitemporal image classification, we followed the traditional image stacking approach, where the descriptor of each pixel is formed by concatenating the features of all epochs at the same pixel location.

For classification, the number of training samples was balanced by replicating samples of less abundant classes. In particular, 30,000 samples per class were selected.

5. RESULTS

Figure 11 shows snips of SAR images and RGB image compositions of real and generated optical images from the simulated cloud regions of the JUL1 epoch. It can be observed the similarity between real and generated images in terms of spectral information as well as in terms of the geometry of objects present in the image. For instance, the structure of rivers and tiles of crops is preserved in most of the cases. However, the generated images are not perfect: it can be noted that in some regions the generated images do not match with the corresponding real one. This is more notorious between the snips of Figure 11f and Figure 11i, where clear differences in spectral information can be observed.

The difficulty to interpret SAR data in comparison to optical images is evident in Figure 7. So, in the next step we compared the classification accuracy for SAR, as well as the generated and real images, using a supervised approach base on a Random Forest classifier.

Results for mono-temporal image classification in terms of Overall Accuracy (OA) and Average Accuracy (AA) are summarized in Figure 7 and Figure 8, respectively. From left to right, each bar in a group corresponds to real optical, generated optical, and SAR image classifications for each evaluated epoch in the experiments (see Table 1). As expected, the results recorded in the experiments using only real optical images were superior to the results obtained upon images provided by the generator. Nevertheless,

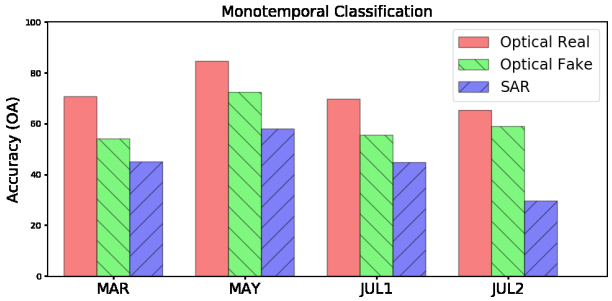


Figure 7. Result for monotemporal image classification in term of OA.

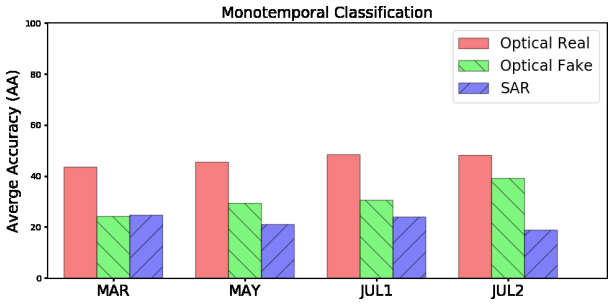


Figure 8. Result for monotemporal image classification in term of AA.

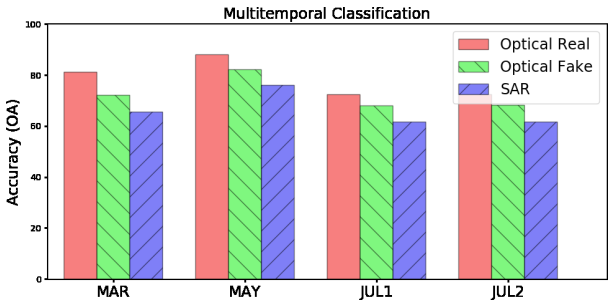


Figure 9. Result for multitemporal image classification in term of OA.

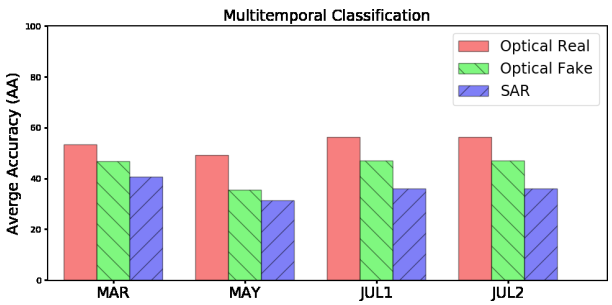
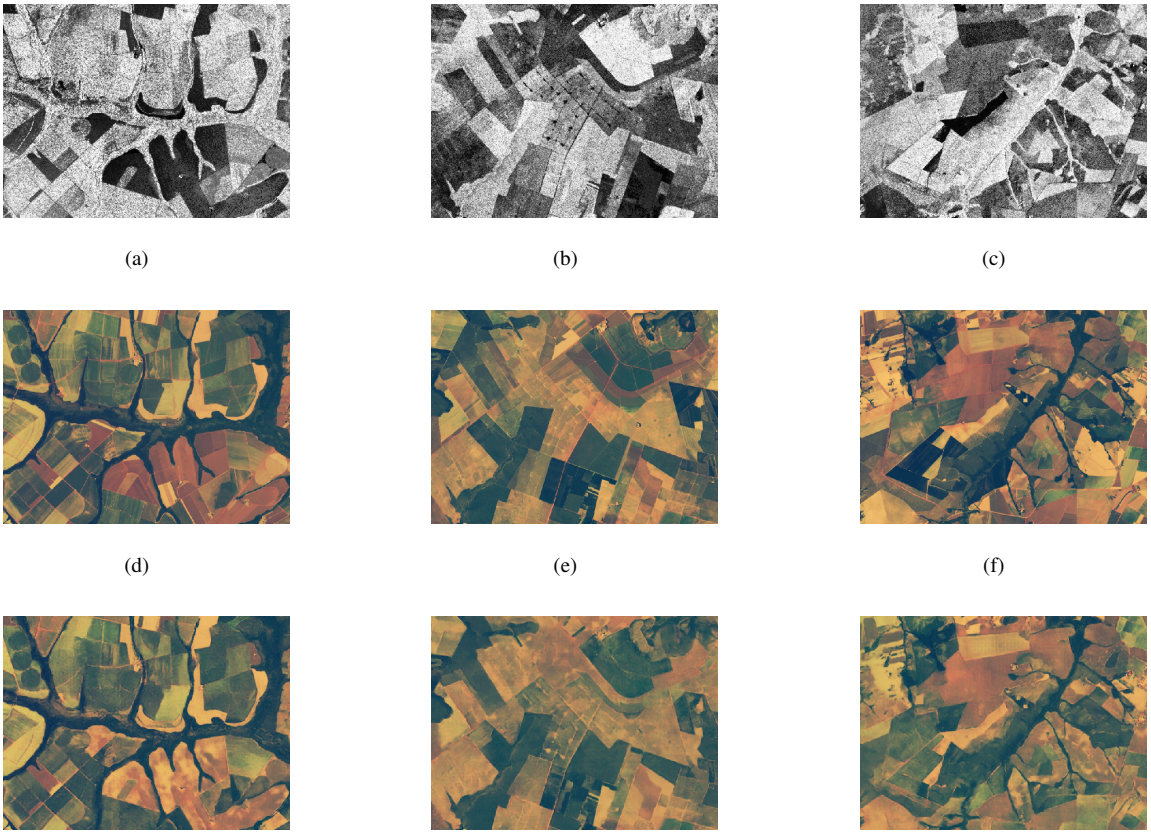


Figure 10. Result for multitemporal image classification in term of AA.

the results of classifying synthesized images were consistently superior to the accuracy obtained in the classification of corresponding SAR images: up to 20% and 29% better in terms of OA and AA, respectively. These results indicate that the proposed method is able to generate optical images, which can be more accurately classified than the corresponding SAR images. Thus,

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-1-5-2018 | © Authors 2018. CC BY 4.0 License.

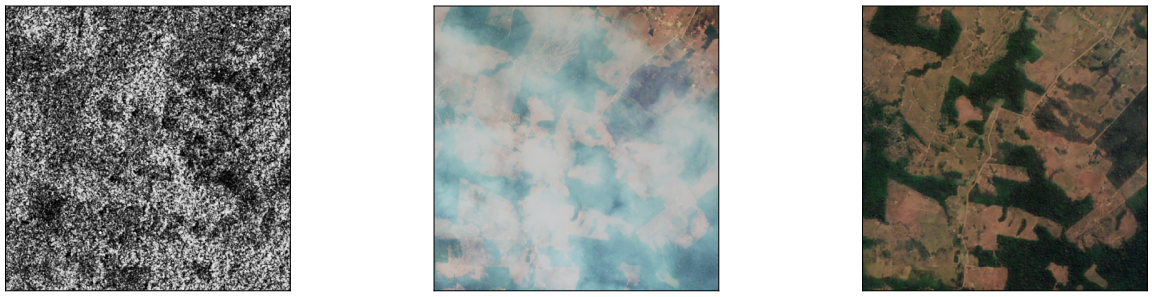


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 11. Snips of SAR and RGB image compositions of corresponding real and generated optical images from JUL1. (a), (b) and (c) are SAR images, f(d), (e) and (f) are the real images whereas (g), (h) and (g) are the corresponding generated images.



(a) (b) (c)

Figure 12. Snips of RGB image compositions of corresponding real and generated optical images from an optical image covered by clouds. (a) is a SAR image, (b) is an optical image and (c) is the corresponding generated image.

it can be an alternative to the common SAR based classification solution, when optical images are partially affected by clouds.

The results of experiments involving multitemporal data are shown in Figure 9 and Figure 10. Similar to the mono-temporal results, the classification on SAR data presented the lowest performance followed by the classification on synthetic and on the real optical image.

In terms of OA and AA, the results for synthetic optical image outperformed those for the SAR images in up to 10% and 6%. With respect to the classification of the real optical images the results upon the synthetic counterparts were inferior in up to 6% and 11%. Similar to the conclusion drawn from the experiments for mono-temporal classification, these results show that images synthesized by the cGAN generator can be used for multitemporal image analysis when optical images are partially covered by clouds.

Finally, Figure 12 shows an example of a real scenario where there is an image covered partially by clouds. Snips of the SAR image and the RGB image compositions of the real and the generated optical image are presented. It can be noted that the proposed method was able to generate the cloud-free optical image, retaining the geometric of the objects as well as their spectral response.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-IV-1-5-2018 | © Authors 2018. CC BY 4.0 License.

6. CONCLUSIONS

In this work, we proposed and assessed the use of conditional generative adversarial networks (cGANs) to remove clouds from optical images. We trained a cGAN upon cloud-free regions of optical images along with the corresponding SAR data. Then, we used the cGAN generator having the SAR data as input to generate a synthetic optical image to recover the optical data covered by clouds.

The experiments in a crop recognition application showed that the classification results obtained on the generated images consistently outperformed similar experiment conducted upon the corresponding SAR image. This corroborates our hypothesis that the proposed method can be used to replace the data that is covered by clouds. Thus, this analysis showed that optical images synthesized from SAR data with the use of cGANs can be used as an alternative for dealing with cloud covering.

Future works include exploring the inclusion of multitemporal information as well as multisensors information. Furthermore, we want to evaluate the proposed methodology in other remote sensing applications and compare it against others techniques for cloud removal.

ACKNOWLEDGEMENTS

The authors acknowledge the funding provided by CAPES, CNPq and FINEP.

REFERENCES

Arkadiusz, N., Michal, R., Adam, J., Michal, T., Konrad, C., Maksymilian, S., Kamil, K. and Piotr, M., 2017. Deep learning for satellite imagery via image segmentation. https://blog.deepsense.ai/deep-learning-for-satellite-imagery-via-image-segmentation/.

Cheng, Q., Shen, H., Zhang, L., Yuan, Q. and Zeng, C., 2014. 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, pp. 54–68.

Enomoto, K., Sakurada, K., Wang, W., Fukui, H., Matsuoka, M., Nakamura, R. and Kawaguchi, N., 2017. Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets. *arXiv preprint arXiv:1710.04835*.

Gómez-Chova, L., Amorós-López, J., Mateo-García, G., Muñoz-Marí, J. and Camps-Valls, G., 2017. Cloud masking and removal in remote sensing image time series. *Journal of Applied Remote Sensing* 11(1), pp. 015005.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In: *Advances in neural information processing systems*, pp. 2672–2680.

Isola, P., Zhu, J.-Y., Zhou, T. and Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134.

Li, Y., Li, W. and Shen, C., 2017. 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), pp. 1870–1882.

Mirza, M. and Osindero, S., 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. and Efros, A. A., 2016. Context encoders: Feature learning by inpainting. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2536–2544.

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241.

Rossow, W. B., 2011. International satellite cloud climatology project.

Sanches, I., Feitosa, R. Q., Diaz, P. M. A., Soares, M. D., Luiz, A. J., Schultz, B. and Maurano, L. E., 2018. Campo verde database: Seeking to improve agricultural remote sensing of tropical areas. *IEEE Geoscience and Remote Sensing Letters* 15(3), pp. 369–373.

Shen, H., Li, H., Qian, Y., Zhang, L. and Yuan, Q., 2014. An effective thin cloud removal procedure for visible remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing* 96, pp. 224–235.

Xu, M., Jia, X., Pickering, M. and Plaza, A. J., 2016. Cloud removal based on sparse representation via multitemporal dictionary learning. *IEEE Transactions on Geoscience and Remote Sensing* 54(5), pp. 2998–3006.

Zhang, Z., Song, Y. and Qi, H., 2017. Age progression/regression by conditional adversarial autoencoder. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2.