A COMPARISON OF CLOUD REMOVAL METHODS FOR DEFORESTATION MONITORING IN AMAZON RAINFOREST

J. A. C. Martinez[1]*, M. X. O. Adarme[1], J. N. Turnes[2], G. A. O. P. Costa[3], C. A. De Almeida[4], R. Q. Feitosa[1]

[1] Pontifical Catholic University of Rio de Janeiro, Dept. of Electrical Engineering, Rio de Janeiro, Brazil
jchamorro@aluno.puc-rio.br, mortega@aluno.puc-rio.br, raul@ele.puc-rio.br
[2] University of Waterloo, Dept. System Design Engineering, Waterloo, Canada - jnoaturn@uwaterloo.ca
[3] Rio de Janeiro State University, Dept. Mathematic and Statistic Rio de Janeiro, Brazil - gilson.costa@ime.uerj.br
[4] National Institute for Space Research, São José dos Campos, Brazil - claudio.almeida@inpe.br

Commission III, WG III/7

KEY WORDS: Cloud Removal, Optical imagery, SAR-optical Data fusion, Deep learning, Deforestation.

ABSTRACT:

Deforestation in tropical rainforests is a major source of carbon dioxide emissions, an important driver of climate change. For decades, the Brazilian government has maintained monitoring programs for deforestation detection in the Brazilian Legal Amazon area based on remotely sensed optical images in a protocol that involves considerable efforts of visual interpretation. However, the Amazon region is covered with clouds for most of the year, and deforestation assessment can rely only on images acquired in the dry season when cloud-free images are more likely to capture. One possibility to lessen that restriction and enable deforestation detection throughout the year is to synthesize cloud-free optical images from corresponding SAR images, which are only marginally influenced by atmospheric conditions. This work compares a set of such image synthesis methods, considering deforestation detection in the Amazon forest as the target application. Specifically, we evaluate three deep learning methods for cloud removal in Sentinel-2 images: a conditional Generative Adversarial Network (cGAN) based on the *pix2pix* architecture; an extension of that method, which uses atrous convolutions (*Atrous cGAN*) to enhance fine image details; and a non-generative method (*DSen2-CR*) based on residual networks. In the evaluation, we assess both the quality of the generated images and the accuracy obtained when performing deforestation detection from those images. We further compare those methods with an image aggregation tool available in Google Earth Engine (*GEE Tool*), which creates cloud-free mosaics from sequences of images acquired at nearby dates. In this study, we considered two sites in the Brazilian Amazon, characterized by distinct vegetation and deforestation patterns. In terms of the quality metrics and classification accuracy, the *Atrous cGAN* was the best performing deep learning method. The *GEE Tool* outperformed all those methods when dealing with images from the dry season but turned out to be the poorest performing method in the wet season.

1. INTRODUCTION

Deforestation is one of the most significant sources of CO_2 emissions to Earth's atmosphere. With 5.5 million km^2, the Amazon rainforest is the largest tropical forest in the world. In the past decade, the forest has become, for the first time a source rather than a sink of CO_2. According to (Gatti et al., 2021), the emissions currently amount to a billion tons of carbon dioxide per year, most of which are due to forest fires deliberately set to clear land for subsistence agriculture, cattle farming and soy plantations.

In the past decade, the Amazon forest has become, for the first time a source rather than a sink of CO_2. According to (Gatti et al., 2021), the emissions currently amount to a billion tons of carbon dioxide per year, most of which are due to forest fires deliberately set to clear land for subsistence agriculture, cattle farming and soy plantations.

The Brazilian government maintains the Monitoring Program for the Amazon and Other Biomes (PAMZ+) (de Almeida et al., 2021), which includes the Brazilian Amazon Rainforest Monitoring Program by Satellite (PRODES) (INPE, 2021). PRODES has produced reports on newly deforested areas each year since 1988. The reports produced by PRODES present official figures and typically result in on-the-spot checks that involve the mobilization of substantial human and material resources. Therefore, high accuracy of the related classification processes is required, thus demanding considerable effort to audit the outcome of automatic interpretation tools visually.

Furthermore, climate conditions restrict the deforestation monitoring frequency as the Amazon region is covered with clouds for most of the year. For that reason, the PRODES program relies on anniversary images captured in the dry season (from late May to late August), when cloud-free images are more likely to be obtained. Interpretation of SAR data could be an alternative outside the dry season, as they are marginally affected by atmospheric conditions. However, SAR images are more challenging to interpret, visually or automatically.

The recent literature reports the efforts of several researchers to conceive methods able to synthesize cloud-free optical images from SAR images. The present work aims at evaluating a set of such methods, having deforestation detection in the Amazon forest as the target application. Specifically, we compare three deep learning-based methods for cloud removal in Sentinel-2 images:

- a conditional Generative Adversarial Network (cGAN) based on the *pix2pix* model that synthesizes a cloud-free

* Corresponding author

version of a cloudy Sentinel-2 image, using as inputs the original optical image and a Sentinel-1 image acquired at a close date, (Bermudez et al., 2019);

- a similar cGAN method based on a different architecture, i.e, the *Atrous cGAN* (Turnes et al., 2020), and
- a non-generative method, called *DSen2-CR* (Meraner et al., 2020), which also takes as input a cloudy Sentinel-2 image and a corregistered contemporary Sentinel-1 image to generate a cloud-free Sentinel-2 image version.

Additionally, we assess the performance of a Google Earth Engine tool (*GEE tool*), introduced in (Schmitt et al., 2019), which essentially aggregates cloud-free Sentinel-2 images of user-defined areas captured within a selected time range.

Finally, besides evaluating the quality of the images produced with those methods, we employed the generated images in the deforestation detection task, considering two sites in the Brazilian Amazon, which are characterized by distinct vegetation and deforestation patterns. As baselines, we used real optical, cloud-free Sentinel-2 images, and plain SAR Sentinel-1 images for the same task.

The contribution of this work is twofold:

- an evaluation of state-of-the-art methods for cloud removal in optical images of the Amazon rainforest; and
- an evaluation of the cloud-free images produced by those methods as inputs for an automatic deep-learning-based deforestation detection network.

The remainder of this work is organized as follows: Section 2 presents the selected cloud removal methods, namely, *pix2pix*, *Atrous cGAN* and *DSen2-CR*; and the image aggregation (*GEE tool*) method. Section 3 presents the experiments carried out, the study areas, and the configuration of each method, as well as the metrics used in the evaluation. Section 4 presents a both a qualitative analysis of the synthesized images, and a quantitative analysis based on image similarity metrics and classification performance.

2. METHODS

For all but the (*GEE tool*) method, a pair of co-registered Sentinel-1 (S1) and cloudy Sentinel-2 (S2 Cloudy) images is used as input, as well as a target cloud-free Sentinel-2 image (S2 Clear) from a nearby date. As the end objective is to use the synthesized cloud-free images for deforestation detection, and as that application requires images from different epochs (T_0 and T_1), we synthesize one image for each epoch, using corespondent image triplets: S1, S2 Cloudy and S2 Clear.

2.1 pix2pix

This method performs pixel to pixel mapping using a conditional GAN (cGAN) model (Isola et al., 2017). Its basic adversarial training scheme is presented in Figure 1.

The input of the generator network (G) is a patch extracted from the concatenation of a co-registered Sentinel-1 (S1) image and a Sentinel-2 cloudy (S2 Cloudy) image. The input is forwarded through a U-Net based generator, which tries to produce a S2 cloud-free synthesized image patch. During training, patches of the real (S2 Clear) and synthesized image patches are presented to a discriminator (D) alternately, together with the corresponding (conditioning) patches of the input image pair (S1 and S2 Cloudy). In the adversarial training scheme, G tries to produce synthetic outcomes which are similar to the real (S2 Clear) image patches, with the aim of fooling the discriminator D, i.e., making it incorrectly classify them as real. At the same time, D is trained to distinguish between real and synthesized image patches, considering the corresponding conditioning image patches (from S1 and S2 Cloudy). When training finishes, it is expected that G is able to properly map the inputs (S1 and S2 Cloudy) to the target (S2 Clear) image.

Further detains about the architecture of the networks that comprise the method evaluated in this work can be found in (Bermudez et al., 2019).

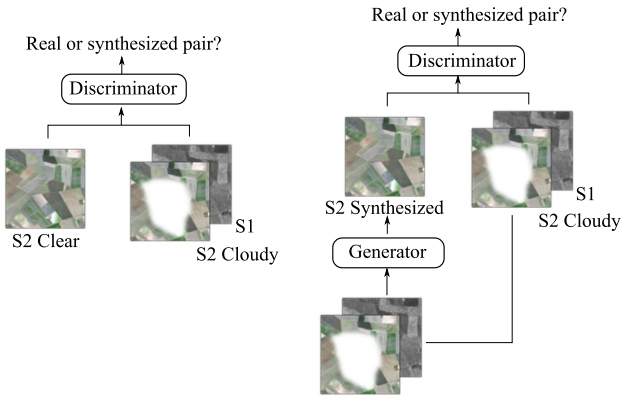

Figure 1. cGAN adversarial training. *pix2pix* uses U-Net as generator, while *Atrous cGAN* uses Deeplabv3 as generator.

2.2 Atrous cGAN

The *Atrous cGAN* method basically extends the *pix2pix* method by adding atrous convolutions to capture and properly generate fine details in the reconstruction outcome, thus producing less noisy images. Originally, the *Atrous cGAN* method was employed to translate S1 images into LANDSAT-8 images (Turnes et al., 2020). In this work, we adapted it to synthesize S2 cloud-free images from the co-registered S1 and S2 cloudy images.

The underlying deep learning model considers multiple receptive fields, both in the generator G and the discriminator D architectures. The generator includes a backbone based on the Deeplabv3 (Chen et al., 2017) architecture, using several dilated convolutions in parallel, and a very deep residual configuration. Additionally, the discriminator uses parallel dilated convolutions to grab multiscale information, being able to better discriminate real and synthetic samples. Further detains about the architecture of the networks can be found in (Turnes et al., 2020).

2.3 DSen2-CR

The input of this method is the concatenation of S1 and S2 Cloudy images. The resulting tensor is forwarded through a sequence of residual blocks, which resulting representation is added to the original S2 Cloudy image. Thus, the collection of residual blocks learns the corrections that need to be made to the S2 Cloudy image in order to produce the desired cloud-free synthetic image. During training, the generated images are

compared with the corresponding cloud-free target image (S2 Clear) in the loss function. The method's basic architectural scheme is presented in Figure 2. Details about the architecture of the network can be found in (Meraner et al., 2020).

As the input (S2 Cloudy) and target (S2 Clear) optical images may present undesirable spectral differences due to their temporal distance, using S2 Clear as target might result in synthesizing errors. To mitigate this problem, *DSen2-CR* incorporates a cloud-regularized loss function. First, a cloud mask is precomputed using the input S2 Cloudy image. During training, the L1 loss is calculated with S2 Clear as target only in the cloud-covered areas; in the cloud-free areas S2 Cloudy original pixels are used as target in the L1 loss.

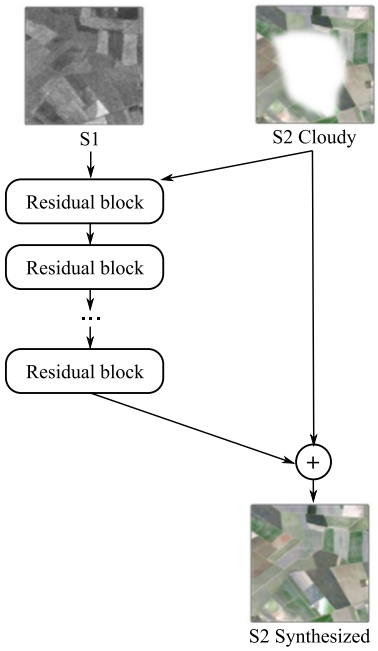

Figure 2. *DSen2-CR* basic architecture.

2.4 Google Earth Engine (GEE tool)

GEE tool is a cloud computing-based platform that (in most cases) allows aggregating cloud-free Sentinel-2 multispectral images (Schmitt et al., 2019). Contrary to the aforementioned DL-based methods, the GEE tool does not rely on statistical or machine learning algorithms, it rather uses posterior information calculated from Sentinel-2 images acquired in a user defined time period.

The method's workflow is composed of three main stages: (1) the *Query Module* loads images from the image collection; (2) the *Quality Score Module* produces pixel-wise cloud scores; and the (3) *Image Merging Module* creates cloud-free images by mosaicking the less cloudy pixels from the image collection, according to (2).

It is important to observe that the tool's outcome may still contain cloud covered pixels, depending on the time period specified by the user.

3. EXPERIMENTS

3.1 Study Areas

The selected methods were evaluated in two regions of the Brazilian Legal Amazon. The datasets comprise co-registered S1 and S2 (clear and cloudy) images. The first study site is located in the Pará (PA) State, Brazil, covering an area that corresponds to 17730×9200 pixels. The second site is located in the Mato Grosso (MT) State, Brazil, covering an area that corresponds to 16795×10420 pixels. Figure 3 shows the location of the two study areas. For the first site (PA), we considered the deforestation that occurred between 2018 and 2019; for the second site (MT), we considered the deforestation that occurred between 2019 and 2020.

For the *pix2pix* and *Atrous cGAN* methods, we excluded the Sentinel-2 bands with 60m resolution. For DSen2-CR, we used all Sentinel-2 bands. In all cases, we resampled some of the bands, standardizing all of them to 10m resolution. Considering that two dates T_0 and T_1 are required to perform deforestation detection, we performed the synthesis on both dates for each study area. Specific acquisition dates of Sentinel-1 and Sentinel-2 images are presented in Table 1.

3.2 Experimental Protocol

For the DL-based adversarial image synthesis methods, i.e., *pix2pix* and *Atrous cGAN* we used the Adam optimizer starting with a 2e-4 learning rate for both the generator and the discriminator networks. Following (Turnes et al., 2020), for both methods, we employed a learning rate decay policy to reach stable models. Following (Meraner et al., 2020), we used Nadam optimizer with 7e-5 learning rate for the *DSen2-CR* method.

We used input patches of size 256×256 for *Pix2pix* and *Atrous cGAN*, and 128×128 for *DSen2-CR*. For the adversarial approaches, we used the loss function described in (Turnes et al., 2020), with $\lambda = 100$. In all cases, we trained the networks for 60 epochs.

To evaluate the quality of the images generated through the aforementioned methods, we compared them with the corresponding S2 Clear images for the different sites and epochs. Accordingly, we measured the following similarity metrics: peak signal-to-noise ratio (PSNR); spectral angle mapper (SAM); and structural similarity index (SSIM). The first two metrics assess spectral similarity, while the latter is a perceptual-based metric that evaluates the structure of objects in the visual scenes. The higher the PSNR and SSIM and the lower the SAM, the better the generated image.

To complete the analysis, we used the generated images for semantic segmentation, employed for deforestation detection. Following (Ortega et al., 2021), we used a ResUNet network to perform semantic segmentation, taking as input a tensor formed by stacking the corresponding T_0 and T_1 images. We trained the network on the synthesized images and compared the results with those obtained using real cloud-free S2 images from a nearby dates, and using S1 image pairs. The network's input patch size was 128×128 pixels. Further details about the network architecture can be found at (Ortega et al., 2021).

For training the semantic segmentation network, we used the Adam optimizer with a 0.001 learning rate. Due to the high class unbalance, we used weighted categorical cross-entropy as loss function, with a vector of weights $[0.1, 0.9, 0]$ for forest, deforestation, previous-deforestation classes. We used a batch size of 32, and an the early stopping strategy, halting training after 10 epochs without performance improvement over the validation set.

Images	Para (PA)		Mato Grosso (MT)	
	T_0 2018	T_1 2019	T_0 2019	T_1 2020
PRODES Ref.	21 July	24 July	[02,18] August	04 August
S1	[19,26] July	[21,26] July	[02,09] August	[03,08] August
S2 (Clear)	[21,26] July	[21,26] July	[02,05] August	[04,11] August
S2 (Cloudy)	11 June	06 July	[26,29] August	[15,18] September
GEE Tool	01 to 15 July	01 to 10 July	15 to 30 September	10 to 30 September

Table 1. Acquisition dates of Sentinel-1 and Sentinel-2 images of the selected study areas. Observe that the days between brackets indicate that the corresponding images are mosaics of two Sentinel scenes, assemble to cover the whole study areas' extents.

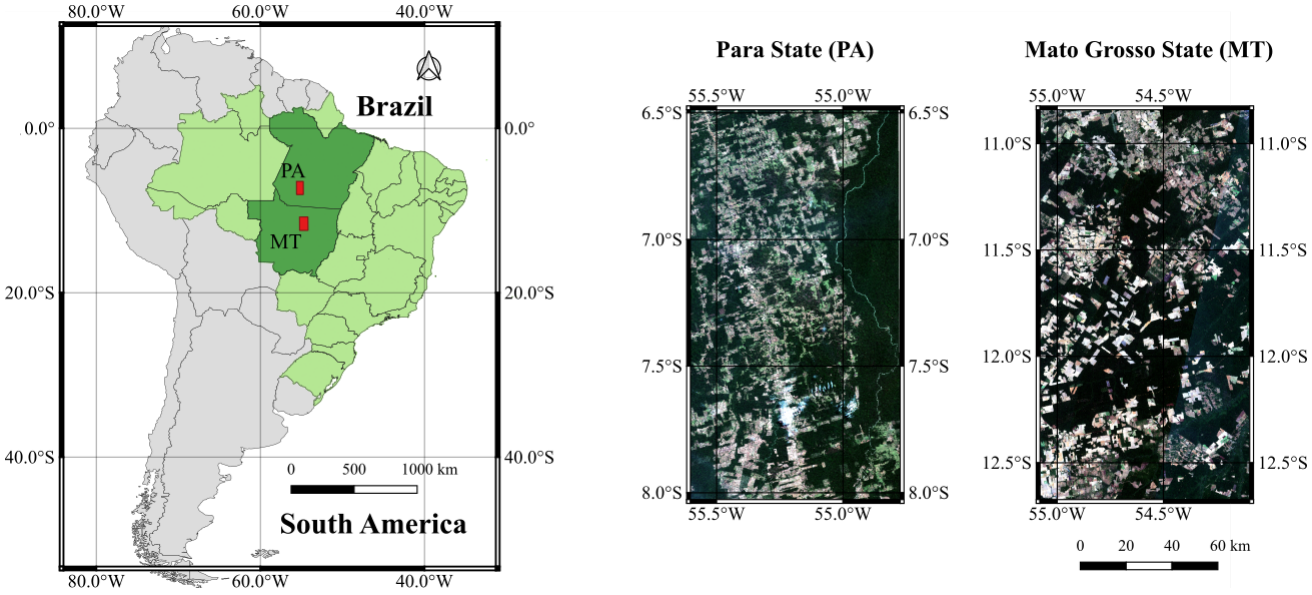

Figure 3. Geographical location of the study areas, and RGB composition of the corresponding Sentinel-2 images acquired at T_0.

For training and testing, we used the same ground truth polygons, obtained from the the PRODES database. We used approximately 40% of the sites' areas for training, 10% for validation, and 50% for testing. We present classification results in terms of the F1 score for the deforestation class. The classification results correspond to the average values obtained from three experiment runs. In all cases, we used code provided by the original authors. In *Atrous cGAN*, we deactivated dropout and increased the overlap at inference. Because of this, we provide the network's modified code. [1]

4. RESULTS

4.1 Para Site (PA)

Figure 4 presents visual results for all methods (RGB composition). From left to right: clear and cloudy S2 images of the Para region captured at close dates of T_1, as indicated below the image. The next four images illustrate the results obtained with each evaluated method (*pix2pix*, *Atrous cGAN*, *DSen2-CR*, and *GEE tool*), whereby the period given for the *GEE-tool* search was 10 days.

Observing the images in Figure 4, it is possible to notice the superiority of *Atrous cGAN* over the other deep learning-based methods, presenting the sharpest synthesis with colors very similar to the S2 Clear image. This result was confirmed in all experiment runs. Nevertheless, the best visual results were obtained with the *GEE tool*. It should be noted that such good performance delivered by *GEE tool* refers to a favourable setup. Notice that the search period, in this case, fell in the centre of the dry season when the probability of images with cloud-free areas is maximum. Figure 5 shows *GEE tool* results for the same area for a search period of 1 and 3 months, respectively, in the wet season. Clearly, such results are inferior to those produced by the methods evaluated in this study (see Figure 4). In short, *GEE tool* produced high-quality results in the dry season, but its performance tends to drop significantly outside this range.

Table 2 reports the performance of each method in terms of SAM, PSNR, and SSIM. The metrics' values confirm the conclusions drawn from the visual results. *Atrous cGAN* produced the best results in terms of the spectral similarity metrics PSNR and SAM. However, the *GEE tool* delivered the highest SSIM. That was an expected result, because the latter method does not synthesize new images, but instead pieces together a mosaic from images of a same area, which allows it to better preserve structural similarity.

Area	Method	PSNR	SAM	SSIM
PA	*Pix2pix*	28.8	0.11	0.81
	Atrous cGAN	**31.9**	**0.06**	0.85
	DSen2-CR	31.7	0.07	0.87
	GEE tool	29.9	0.09	**0.91**

Table 2. Similarity metrics for the images generated for the PA site. The values correspond to averages, considering T_0 and T_1 synthetic images - best results in bold face.

We further assessed the performance of the methods as a kind of representation learner for an automatic deforestation detector.

[1] Code available at https://github.com/DiMorten/ForstCARe-Clouds–tf1

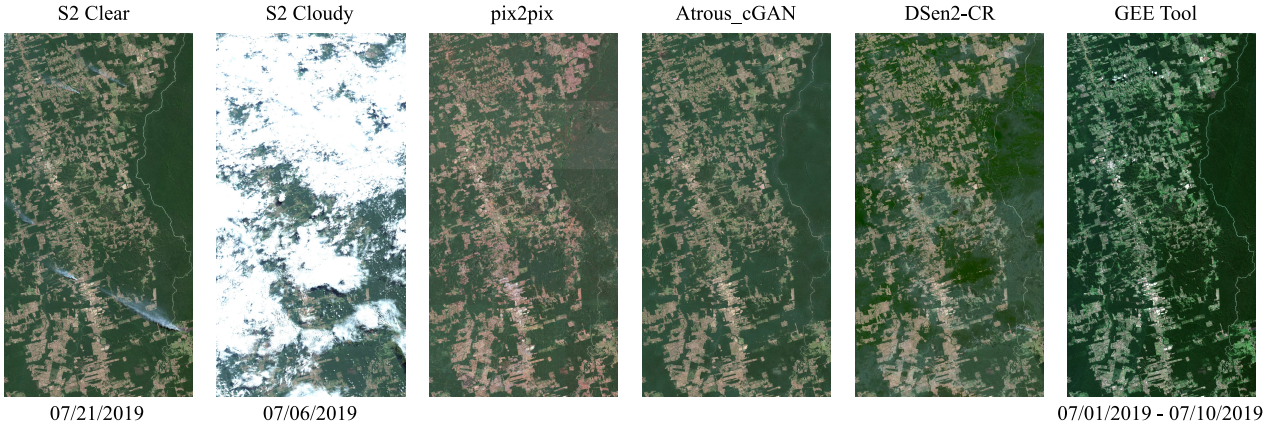

Figure 4. Sample visual results of the images generated with the methods being compared for the PA site. From left to right: S2 Clear (reference image), S2 Cloudy (cloudy image), results of the *pix2pix*, *Atrous cGAN*, *DSen2-CR* and *GEE tool* methods.

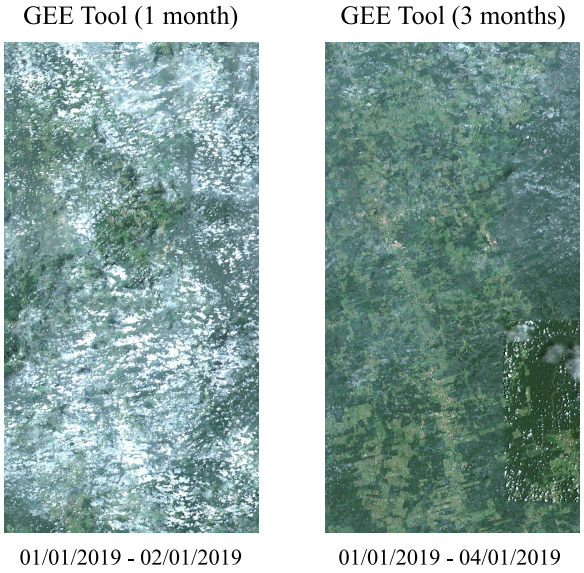

Figure 5. Sample visual results obtained with the *GEE tool* for different search periods during the wet season in the PA area.

In that respect, Figure 6 presents classification results obtained using real and synthesized images as inputs, in terms of averaged F1 score. As expected, classification using real images as input produced the best results: the highest F1 score was obtained using a pair of S2 Clear images. The second best result was obtained when classifying multidate pairs of S1 images, with only a 13% drop in accuracy in relation to the best score.

In general, the classification outcomes obtained with synthesized images produced worse scores as compared to aforementioned ones, with the best performing method being the *GEE tool*, which resulted in a 13% drop compared to the S1 alternative. The best performing method among the deep learning-based ones was the *Atrous cGAN*, which outperformed its counterparts by a large margin, of up to 20%.

4.2 Mato Grosso Site (MT)

By and large, the results for the MT site were similar to the ones obtained for PA. Figure 7 presents the images generated with the alternative methods for the MT site (T_1 date). In this case, the best visual results among the deep learning methods were produced by the *Atrous cGAN* and *pix2pix* methods. Although *DSen2-CR* produced similar results, the synthesized image shows a clear difference in the reconstruction of cloudy and cloud-free regions from the original S2 Cloudy image.

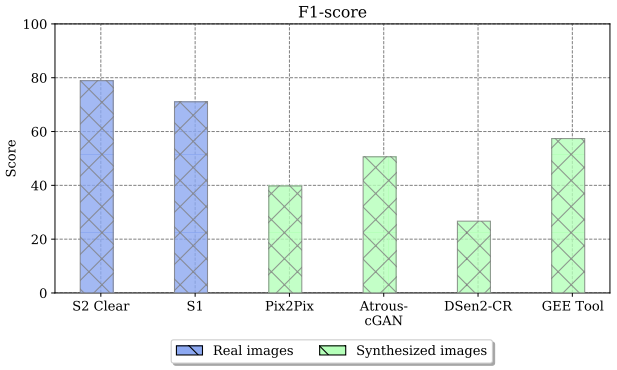

Figure 6. Classification results in terms of F1 score for the deforestation class in the PA site.

In this case, however, the *GEE tool* generated an image that still has cloud covered areas. As in the PA site, *GEE tool* also presented low quality results during the wet season, failing to produce a cloud-free image (Figure 9).

Table 3 presents similarity metrics values computed for the generated images. The results are again consistent with the qualitative analysis. The *Atrous cGAN* method is associated with the best spectral similarity metrics, while the *GEE tool* delivered the highest SSIM value.

Area	Method	PSNR	SAM	SSIM
MT	*Pix2pix*	30.5	0.07	0.85
	Atrous cGAN	**31.0**	**0.06**	0.87
	DSen2-CR	25.2	0.23	0.85
	GEE tool	29.8	0.08	**0.92**

Table 3. Similarity metrics for the images generated for the MT site. The values correspond to averages, considering T_0 and T_1 synthetic images - best results in bold face.

Figure 8 presents deforestation detection results in terms of F1 score. Similarly to what was observed for the PA site, the best results were obtained with the real images (S2 Clear and S1). None of the cloud removal methods produced better classification results than the ones obtained with the S1 images. The *pix2pix* and *Atrous cGAN* methods produced the best results

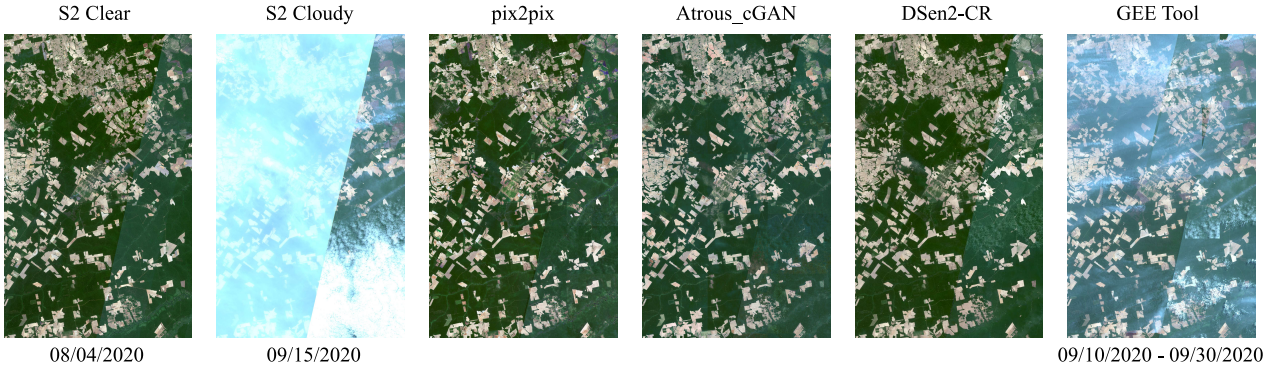

Figure 7. Sample visual results of the images generated with the methods being compared for the MT site. From left to right: S2 Clear (reference image), S2 Cloudy (cloudy image), results of the *pix2pix*, *Atrous cGAN*, *DSen2-CR* and *GEE tool* methods.

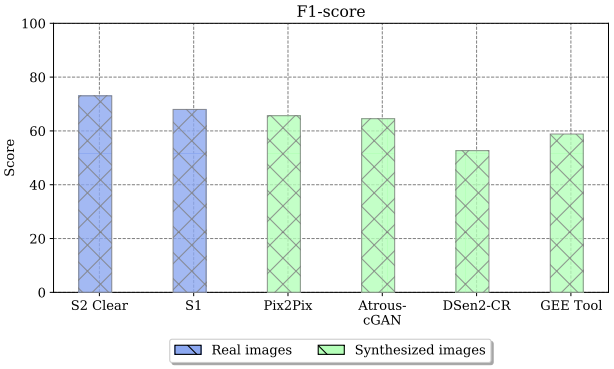

Figure 8. Classification results in terms of F1 score for the deforestation class in the MT site.

among the synthesizing approaches with small drops of approximately 3% compared to the S1 images. In this case, the deep learning methods outperformed *GEE Tool* by up to 6.7% in terms of F1 score.

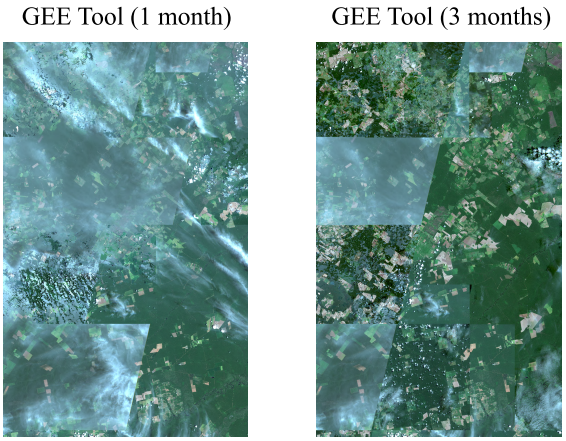

Figure 9. Sample visual results of *GEE tool* for different search periods during the wet season in the MT area.

5. CONCLUSIONS

This work compared methods for cloud removal from optical and SAR images for deforestation detection. Qualitatively, the best results were obtained with the *GEE tool* and the *Atrous cGAN*. Although the former method produced high quality results for the dry season, its performance significantly decreased during the wet season.

In terms of deforestation detection, classification accuracy was highest when optical cloud-free images were used, as expected. The next best result was obtained using Sentinel-1 images, which indicates its usefulness for deforestation detection in the Amazon region, where images are usually covered with large amounts of clouds for most of the year. Although synthesized cloud-free images produced satisfactory visual results and good similarity metrics values, the classification performances obtained when using those images were inferior to those obtained with the Sentinel-1 baseline.

Although the *GEE Tool* has outperformed the deep-learning-based methods for the dry season in the PA site, its performance tends to decrease for other periods of the year. Indeed, it delivered the worst performance under all metrics among the methods investigated in the middle of the wet season.

ACKNOWLEDGEMENTS

The authors acknowledge the funding provided by CNPq and CAPES.

REFERENCES

Bermudez, J. D., Happ, P. N., Feitosa, R. Q., Oliveira, D. A., 2019. Synthesis of multispectral optical images from SAR/optical multitemporal data using conditional generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 16(8), 1220–1224.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

de Almeida, C. A., Maurano, L. E. P., de Morisson Valeriano, D., Camara, G., Vinhas, L., Gomes, A. R., Monteiro, A. M. V., de Almeida Souza, A. A., Rennó, C. D., Silva, D. E. et al., 2021. Methodology for Forest Monitoring used in PRODES and DETER Projects. *CEP*, 12, 010.

Gatti, L. V., Basso, L. S., Miller, J. B., Gloor, M., Gatti Domingues, L., Cassol, H. L., Tejada, G., Aragão, L. E., Nobre, C., Peters, W. et al., 2021. Amazonia as a carbon source linked to deforestation and climate change. *Nature*, 595(7867), 388–393.

INPE, 2021. National institute for space research. general coordination of earth observation. monitoring program of the amazon and other biomes. deforestation - legal amazon -. http://terrabrasilis.dpi.inpe.br.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Meraner, A., Ebel, P., Zhu, X. X., Schmitt, M., 2020. Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166, 333 - 346.

Ortega, M. X., Feitosa, R. Q., Bermudez, J. D., Happ, P. N., De Almeida, C. A., 2021. Comparison of optical and sar data for deforestation mapping in the amazon rainforest with fully convolutional networks. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, 3769–3772.

Schmitt, M., Hughes, L., Qiu, C., Zhu, X. X., 2019. Aggregating cloud-free Sentinel-2 images with Google earth engine. *PIA19: Photogrammetric Image Analysis*, 145–152.

Turnes, J. N., Castro, J. D. B., Torres, D. L., Vega, P. J. S., Feitosa, R. Q., Happ, P. N., 2020. Atrous cGAN for SAR to Optical Image Translation. *IEEE Geoscience and Remote Sensing Letters*.