EVALUATION OF SAR TO OPTICAL IMAGE TRANSLATION USING CONDITIONAL GENERATIVE ADVERSARIAL NETWORK FOR CLOUD REMOVAL IN A CROP DATASET

L. E. Christovam [1], M. H. Shimabukuro [1,2], M. L. B. Trindade Galo [1,3], E. Honkavaara [4]

[1]Graduate Program in Cartographic Sciences, São Paulo State University, Brazil – (luiz.christovam, milton.h.shimabukuro, trindade.galo)@unesp.br
[2] Dept. of Mathematics and Computer Science, São Paulo State University, Brazil
[3]Dept. of Cartography, São Paulo State University, Brazil
[4]Dept. of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute in National Land Survey of Finland, Finland – eija.honkavaara@nls.fi

KEY WORDS: cGAN, Sar-to-Optical, Remote Sensing, Image-to-Image, Image Translation, Pix2Pix, Synthetic Images, Sentinel-2

ABSTRACT:

Most methods developed to map crop fields with high-quality are based on optical image time-series. However, often accuracy of these approaches is deteriorated due to clouds and cloud shadows, which can decrease the availably of optical data required to represent crop phenological stages. In this sense, the objective of this study was to implement and evaluate the conditional Generative Adversarial Network (cGAN) that has been indicated as a potential tool to address the cloud and cloud shadow removal; we also compared it with the Witthaker Smother (WS), which is a well-known data cleaning algorithm. The dataset used to train and assess the methods was the Luis Eduardo Magalhães benchmark for tropical agricultural remote sensing applications. We selected one MSI/Sentinel-2 and C-SAR/Sentinel-1 image pair taken in days as close as possible. A total of 5000 image pair patches were generated to train the cGAN model, which was used to derive synthetic optical pixels for a testing area. Visual analysis, spectral behaviour comparison, and classification were used to evaluate and compare the pixels generated with the cGAN and WS against the pixel values from the real image. The cGAN provided consistent pixel values for most crop types in comparison to the real pixel values and outperformed the WS significantly. The results indicated that the cGAN has potential to fill cloud and cloud shadow gaps in optical image time-series.

1. INTRODUCTION

Food demand is increasing rapidly and therefore sustainable food production is currently a worldwide concern. Tillman et al. (2011) predicted that by 2050 the world agricultural production should double to meet the food supply needs. In this context, stakeholders require regular and high-quality agricultural statistics to improve crop yield.

The remote sensing community has made efforts to develop methods to respond this demand. The majority of existing approaches utilize optical image time-series to account for the crop phenological stages. However, noises, particularly clouds and cloud shadows in optical data can deteriorate the accuracy of land cover mapping. Lunetta et al. (2006) pointed out the requirement of data cleaning pre-processing to remove the uncertainty and provide a better land cover mapping using time series. Shao et al. (2016) outlined several smoothing algorithms that can be used to minimize noise in optical time series. The authors evaluated several algorithms and reported that Witthaker Smother (WS) provided the best distinction between classes of interest.

The Generative Adversarial Network (GAN) is a novel approach for cloud removal. It comprises two deep neural networks competing against each other, a Generative model (G) and a Discriminative model (D). G learns to map an input data into output and D discriminates if its input is real or fake. Another version of GAN is the Conditional GAN (cGAN) which has its output conditioned in some input data. Bermudez et al., (2018) and (2019) addressed cloud removal with cGAN using SAR images to conditionate the delivery of optical data. Both approaches have been used in previous studies, however, to the best of our knowledge, up to now there has not been evaluation concerning which one delivers the best synthetic optical data. Our aim in this work was to perform an evaluation on the quality of pixel values delivered by the cGAN and WS in a crop dataset. This paper is organized as follows: an overview of GAN is presented in section 2; in section 3 are presented the dataset, pre-processing, and experimental protocol; results and discussion are given in section 4; and the conclusions in section 5.

2. GENERATIVE ADVERSARIAL NETWORK

GAN was first introduced by Goodfellow et al. (2014) and it has two deep neural networks, a Generative model (G) and a Discriminative model (D). G is a mapping function that given any data distribution $p_{data}(x)$, learns to map a random noise vector z, to produce an output y. The distribution of the output data, $p_{model}(x)$, must be as close as possible to $p_{data}(x)$. D is a function that discriminates if its input is real or fake, therefore if the input comes from $p_{data}(x)$ or $p_{model}(x)$.

The main goal of a GAN is to improve G until D cannot tell if generated images are real or fake. To this end the models are trained in an adversarial manner in a zero-sum game, trying to find the optimal mapping function, which is presented in Equation 1.

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

where $\mathcal{L}_{GAN}(G,D)$ is the objective function presented in Equation 2.

https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-823-2021

$$\mathcal{L}_{GAN}(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_{(z)}}[\log(1 - D(G(z)))] \quad (2)$$

where z is the random noise vector; and $p_z(z)$ is the probability distribution of the noise data.

The backpropagation algorithm is used in the training. First, D is trained with real and generated samples considering random initial weights, and G also with random weights but fixed (not trainable). After, D is set as not trainable and the parameters of G are updated to minimize the loss from D, considering now that the output of the G model is set as “real samples”. Training a GAN includes minimizing the Jensen-Shannon (JS) divergence (Equation 3) between the data distributions $p_{data}(x)$ or $p_{model}(x)$ as stated before.

$$JS\big(p_{data}(x), p_{model}(x)\big) = KL(p_{data}(x)||p_m) + KL(p_{model}(x)||p_m) \quad (3)$$

where p_m is $(p_{data}(x) + p_{model}(x))/2$. This divergence is symmetrical and always defined since it is chosen $\mu = p_m$; KL stands for Kullback-Leibler (KL) divergence that is defined in Equation 4.

$$KL(p_{data}(x)||p_{model}(x)) = \int \log\left(\frac{p_{data}(x)}{p_{model}(x)}\right) p_{data}(x) \quad (4)$$

Mirza and Osindero (2014) presented the cGAN which is an extension of GAN. The main difference is that the generative and discriminative models have their outputs conditioned on extra information, like an observed image x. So, the G function learns to map an observed image x and a random noise vector z, to produce an output y. The objective function of GAN conditioned in extra information (x) is expressed in Equation 5.

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + \mathbb{E}_{x\sim p_{(x)}, z\sim p_{(z)}}[\log(1 - D(x, G(x,z))] \quad (5)$$

3. MATERIAL AND METHODS

3.1 Dataset and Pre-processing

To compare the performance of cGAN and WS, we carried out experiments using the Luís Eduardo Magalhães (LEM) benchmark for tropical agricultural remote sensing application (Sanches et al., 2018). The LEM database embraces optical and SAR image time series, having 54 MSI/Sentinel-2, 24 OLI/Landsat-8, and 30 C-SAR/Sentinel-1 images. The optical images were provided in surface reflectance and the SAR images in sigma nought (σ^0) in dB scale.

Each pixel of the C-SAR/Sentinel-1 images are represented by a vector compromising the polarizations VV and VH. For the MSI/Sentinel-2 images, each pixel is represented by a vector with values for Blue, Green, Red, NIR, Red-Edge1 (RE1), Red-Edge2 (RE2), Red-Edge3 (RE3), SWIR1, and SWIR2 bands. The pixel values of both optical and SAR images were normalized to lie between -1 and 1. The optical images were normalized using linear mapping, whereas the SAR images were normalized as presented in Enomoto et al. (2018).

Due to the different spatial resolutions of the images, the C-SAR/Sentinel-1 images were resampled to a spatial resolution of 20 meters using the nearest neighbor algorithm. This resolution was chosen because it is the predominant resolution of the MSI/Sentinel-2 bands and because it is lower than the resolution of the C-SAR/Sentinel-1 images provided in the LEM dataset (10 meters).

Besides the images, the LEM database also has a ground truth with 794 polygons representing crop fields, each polygon has as attributes a sequence of monthly land use (type of agricultural crop) for 2017/2018 harvests (summer and winter). The classes in this database are beans, cerrado (Brazilian savanna), coffee, conversion area, cotton, crotalaria, eucalyptus, grass, hay, corn, corn+crotalaria, millet, non-commercial crops, unidentified, pasture, sorghum, soybean, uncultivated soil, and wheat.

To perform the experiments, the ground truth was split into training and testing areas. The training area represents polygons free of cloud and cloud shadows, while the testing area simulates crop fields covered by cloud and cloud shadows. The boundaries of Luís Eduardo Magalhães municipality, as well as the crop fields of the ground truth, split into training and testing areas are depicted in Figure 1.

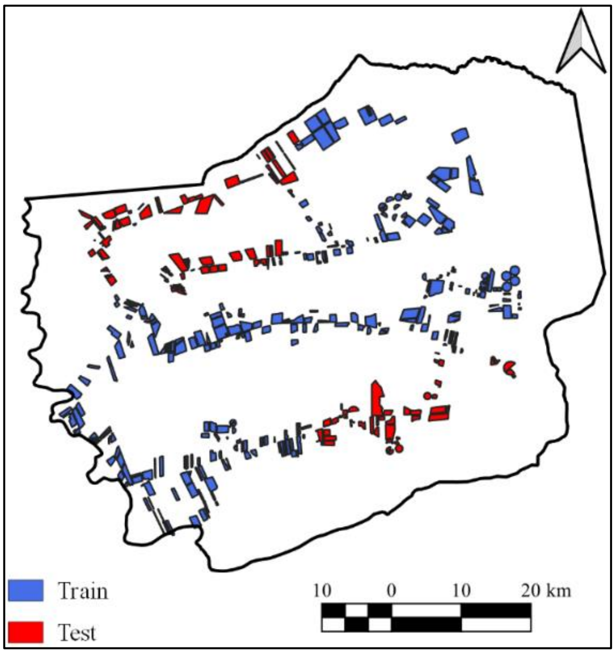

Figure 1. Luís Eduardo Magalhães municipality boundaries with ground truth polygons

3.2 Experimental Protocol

We selected one cloud free MSI/Sentinel-2 image on the dataset to perform the evaluation process, the image chosen was taken on April 30th, 2018. To perform the image translation, we selected the C-SAR/Sentinel-1 image collected on May 2nd, 2018, which was close date to the date of the optical image. A total of 5000 patches with 256x256 pixels were randomly extracted over the crop fields and were balanced regarding the class of the central pixel. The architecture of the cGAN (Generator and Discriminator) used is the same as pix2pix GAN presented by Isola et al. (2017).

Since the WS algorithm works over time series, we built a data cube for each band using every MSI/Sentinel-2 image available on the dataset. Each cube was smoothed using the WS algorithm with a λ equal to 10, where a larger λ gives a smoother result.

The evaluation was carried out first with visual analysis, comparing patches gotten over the real, the cGAN, and the WS

https://doi.org/10.5194/isprs-archives-XLIII-B3-2021-823-2021 | © Author(s) 2021. CC BY 4.0 License.

images. Following, the spectral behavior for various classes was evaluated by comparing the real image against the synthetic pixels delivered by cGAN and WS. Finally, the three images were classified using the Random Forest (RF) algorithm. The classification model was trained using data from the training area using the real image pixels; the model was then used to classify the test area for the real and both synthetic images. The RF version used is available in the python *scikit-learn* package (Pedregosa et al., 2011). The RF was set with 10000 trees, depth equal to 12, the *criterion* was the entropy, and *class_weight* was set as balanced; other parameters were left as default. To evaluate the classification performance, the Kappa coefficient and F_1-score were computed. The Kappa coefficient is used for a general evaluation of the classification, while the F1-score is the harmonic mean between the precision and recall for each class, for the whole model it represents how well the model classifies each class.

4. RESULTS AND DISCUSSION

Figure 2 shows false color composites of image patches from three different locations. In the first column are shown patches from the real optical image (a, e, i), in the second column are presented C-SAR patches (b, f, j), and in the third (c, g, k) and fourth (d, h, l) columns are shown patches for the synthetic optical images generated by cGAN and WS, respectively. Comparing the patches (a) and (e), which have clouds and cloud shadows with the cGAN and WS patches, (c, g) and (d, h), respectively, it is possible to notice that the cGAN approach was capable to generate synthetic pixels that visually fit in the places that had noise. However, it is also possible to observe that the areas not affected by noises in the real images were more blurred in the cGAN patches. In the WS patches, the dark clouds and cloud shadows diminished, however, new noises, looking like thin clouds, appeared in the patches, which is probably related to the presence of clouds in previous and subsequent images in the LEM image time series.

Regarding the spectral behaviour, it is shown in Figure 3 six different charts for the land use classes (a) maize; (b) uncultivated soil; (c) cerrado (Brazilian savannah); (d) cotton; (e) eucalyptus; and (f) pasture. In each chart, there are boxplots of normalized pixel values for every spectral band of the three different images evaluated, the real image is depicted in gray, cGAN in light blue, and WS in dark blue, inside each boxplot the dot and line represent the mean and median, respectively. It is possible to notice that the mean values for almost every cGAN band are closer to the real image than the WS image is, the clear exceptions are the RE2 and NIR bands for maize (Figure 3a) and the RE2 and RE3 bands for cerrado (Figure 3c) where the mean for the WS pixels are closer to the real image than the cGAN.

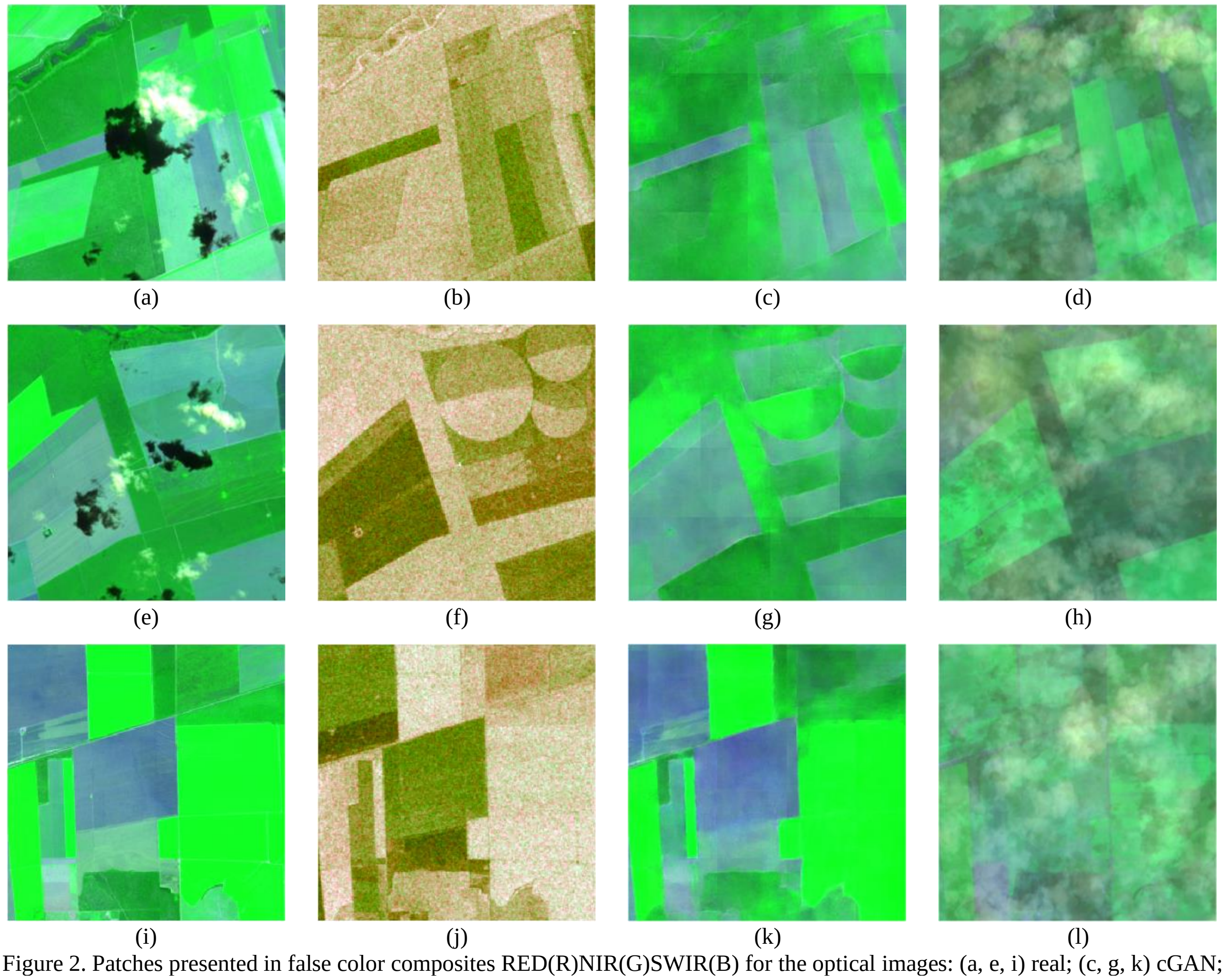

Figure 2. Patches presented in false color composites RED(R)NIR(G)SWIR(B) for the optical images: (a, e, i) real; (c, g, k) cGAN; (d, h, l) WS. Patches presented in false color composites HH(R)VV(G)HH+VV(B) for the C-SAR images (b, f, j)

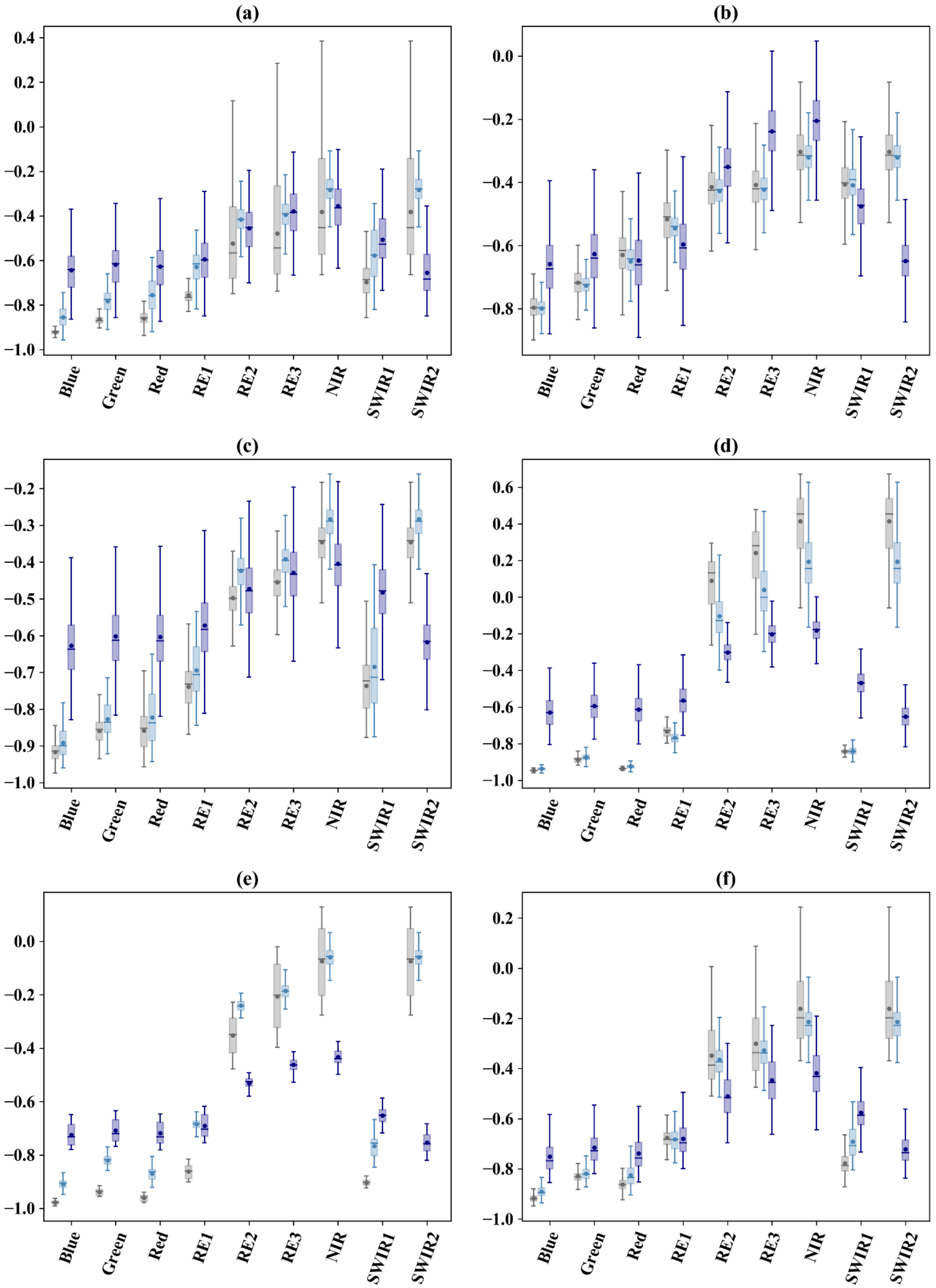

Figure 3. Spectral behaviour depicted using boxplots of the normalized pixel values for the real image (gray) and synthetic pixel values for cGAN (light blue) and WS (dark blue) images. Each chart corresponds to one land use class (a) maize; (b) uncultivated soil; (c) cerrado; (d) cotton; (e) eucalyptus; and (f) pasture. Inside the boxplots the line indicates the pixels median, while the dot represents the mean.

The results in Figure 3 show that for the most of classes the bands in the visible spectrum (Blue, Green, and Red) from the real and cGAN images are similar, especially the classes cotton (Figure 3d) and pasture (Figure 3f) which match almost perfectly. Considering the visible spectral range, the only class where the boxplots are quite different is the eucalyptus (Figure 3e), although this difference decreases in other spectral bands. Even so, the eucalyptus class has the worst performance for both synthetic images, since for most bands the boxplots do not overlap with the boxplots for the real images. Considering the variability, it is possible to observe for most of the classes that the cGAN again outperformed the WS since the range between the minimum and maximum in the boxplots for the cGAN were closer to the range in the real image boxplots than in the WS image boxplots. Although large differences can be seen in some charts, like for the bands RE2, RE3, NIR, and SWIR2 in the class maize (Figure3a), yet this is not an issue since all generated pixels for the bands mentioned lie inside the same range that the real pixels for this class lie. Regarding the variability, a real issue can be seen in almost every chart for the WS and in certain specific cases for the cGAN, for instance, in the visible spectral interval for maize (Figure 3a), since the variability of the synthetic pixels are higher than the variability in the real image, meaning that the generated pixel values do not lie in the spectral interval of this class and which can lead to further misclassification.

Regarding the classification results, the Kappa coefficient was 0.645 for the real image, 0.418 for the cGAN image, and 0.019 for the WS image. The F_1-score was 0.756 for the real image, 0.570 for the cGAN, and 0.288 for the WS image. These results are graphically shown in Figure 4.

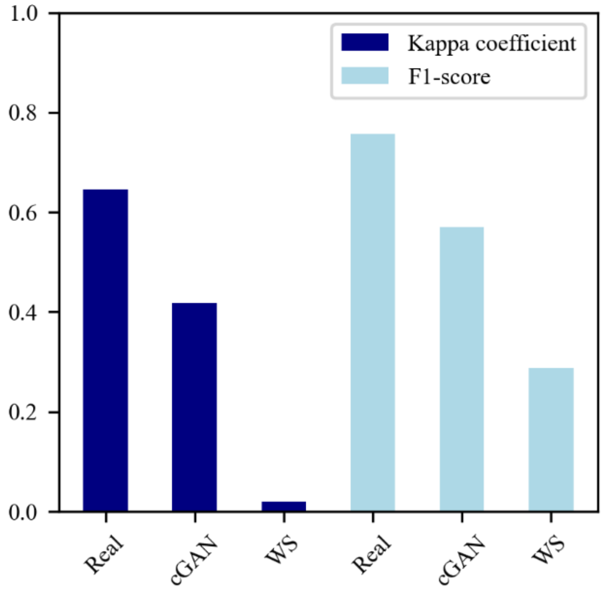

Figure 4. Metrics used for evaluation of the classification: Kappa coefficient and F_1-score

The classification metrics shown in Figure 4, confirm the visual analysis and the spectral behavior evaluation commented above where the cGAN also outperformed by far the WS image.

5. CONCLUSION

Our study compared the cGAN against the WS algorithm for delivering synthetic optical pixels for one image from the LEM benchmark and dataset. We have carried out a visual analysis, spectral behaviour comparison, and classification, which showed the cGAN outperformed the WS. In visual analysis, the capability of cGAN for removing clouds and cloud shadows was shown, however, the cGAN image patches were blurred when compared to the real image patches. Regarding the spectral behaviour, the cGAN delivered pixel values consistent with the real image for most of the classes. Furthermore, cGAN outperformed the WS in the classification analysis, but was worse than the real image. It is important to highlight that these are initial results and that there is room for improvement. A more comprehensive evaluation considering more image dates is necessary. In this data set, the WS results may have been deteriorated by cloud and cloud shadows in previous or further images in the dataset. In future studies we aim to improve the classification results using the cGAN, adopting the multitemporal approach as presented by Bermudez et al. (2019) and using a loss function that enforces the spectral behaviour for the generated pixels to be closer to the real ones.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from the Coordination for the Improvement of Higher-Level Personnel – CAPES (88887.473380/2020-00 and 88882.433956/2019-01), Academy of Finland (grant number 335612), and the anonymous paper reviewers for their constructive comments.

REFERENCES

Bermudez, J. D. et al., 2018. SAR to optical image synthesis for cloud removal with generative adversarial networks. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *4*(1).

Bermudez, J. D. et al., 2019 Synthesis of multispectral optical images from sar/optical multitemporal data using conditional generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters,* v. 16, n. 8, p. 1220-1224, 2019.

Enomoto, K., Sakurada, K., Wang, W., Kawaguchi, N., Matsuoka, M. and Nakamura, R., 2018, July. Image translation between SAR and optical imagery with generative adversarial nets. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1752-1755). IEEE.

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).

Lunetta, R. S. et al. 2006. Land-cover change detection using multi-temporal MODIS NDVI data. *Remote sensing of environment,* v. 105, n. 2, p. 142-154, 2006.

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), pp.2825-2830.

Sanches, I. D. et al, 2018. LEM benchmark database for tropical agricultural remote sensing application. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*(1).

Shao, Y. et al. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. *Remote Sensing of Environment*, v. 174, p. 258-265, 2016.

Tilman, D. et al. 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, *108*(50), pp.20260-20264