GAN-based data augmentation for the classification of remote sensing multispectral images

Víctor Barreiro[1], Álvaro G. Dieste[2], Dora B. Heras[3], Francisco Argüello[4],

Abstract—
Multispectral images frequently suffer from limited labeled data, which constrains the accuracy of classification. The objective of data augmentation is to improve the performance of machine learning models by artificially increasing the size of the training dataset. This paper introduces mDAGAN, a data augmentation method for the classification of high resolution multispectral remote sensing images. It is an adaptation of DAGAN (Data Augmentation GAN) to multispectral images, a generative adversarial network that consists of a generator and a discriminator. The augmentation capacity of mDAGAN for three different classical supervised classification algorithms has been evaluated over three high resolution multispectral images of vegetation, providing increased classification accuracies.

Keywords— **Data augmentation, GAN, multispectral, classification**

I. Introduction

In the context of machine learning applied to remote sensing multispectral imaging, some challenges arise based on the scarcity of labeled data available in the images. This condition limits the learning capabilities of the machine learning algorithms, in general, and the deep learning ones, in particular. Data augmentation strategies that synthetically generate new training data based on the existing data have been designed to deal with this problem [1], [2], [3], [4]. A notable division exists between classical and machine learning approaches for data augmentation [5]. The classical approach encompasses methods such as homographies, crops, rotations, flips, and noise addition [6]. In contrast, machine learning data augmentation techniques involve constructing a machine learning model to generate synthetic samples. In the case of remote sensing classification problems, the augmentation data techniques need to adequately exploit both the spatial and the spectral information available in the image.

Among the deep learning data augmentation techniques, various types can be identified, mainly, Autoencoders (AE, in special way Variational Autoencoders VAE), and generative adversarial networks (GANs) [5]. Recently, the augmentation techniques that have gained more popularity are those based on GAN models [7], [8], [9].

A GAN model consists basically of two separate networks. First, a generator for building synthetic data. A second network called discriminator distinguishes between real and synthetic samples to teach the generator to produce better real samples. The training process works as a competition of the two networks.

In this work, a GAN-based data augmentation technique for the classification of very high-resolution multispectral images of vegetation, mDAGAN, is proposed. It is based on DAGAN (Data Augmentation GAN) [10]. As DAGAN is designed to operate with RGB images, the most relevant modification proposed in this paper is to adapt the network to operate over multispectral images exploiting all the spectral information. To reduce the computational cost and incorporate more relevant information to the network, the dataset is first segmented using a superpixel algorithm that also analyzes all the spectral bands of the image. A new discriminator is also proposed, a siamese network [11]. The loss functions are also redefined to accelerate the convergence. The resulting modified network features a reduced number of parameters in the discriminator compared to the original network which leads to approximately a 50% reduction in training time.

II. Data augmentation technique

The proposed data augmentation technique is shown in Figure 1. In order to reduce the execution time of the experiments, the input image is first segmented by using the waterpixel segmentation algorithm as described in [7]. The resulting segments are homogeneous regions whose compacity and average size have been selected to adapt to the structures in the image. Each sample is a patch centered in the central pixel of a segment. In each step, the inputs are two patches. The input denoted as G_i corresponds to a synthetic patch generated by the generator network.

As we can see in Figure 1, the first stage of the generator is a U-Net encoder as in the original DAGAN [10], producing features that are concatenated with a $z - noise$ vector. The obtained feature vector is processed by the U-Net decoder, obtaining a new generated patch G_i, the augmented sample of the same class than the real sample used as input. The generator of the DAGAN network [10], originally designed for RGB images, has also been adapted to process

[1]CiTIUS and Dpto. de electrónica y computación USC, e-mail: victorxesus.barreiro.dominguez@usc.es.
[2]CiTIUS and Dpto. de electrónica y computación USC, e-mail: alvaro.goldar.dieste@usc.es.
[3]CiTIUS and Dpto. de electrónica y computación USC, e-mail: dora.blanco@usc.es.
[4]Dpto. de electrónica y computación USC, e-mail: francisco.arguello@usc.es.

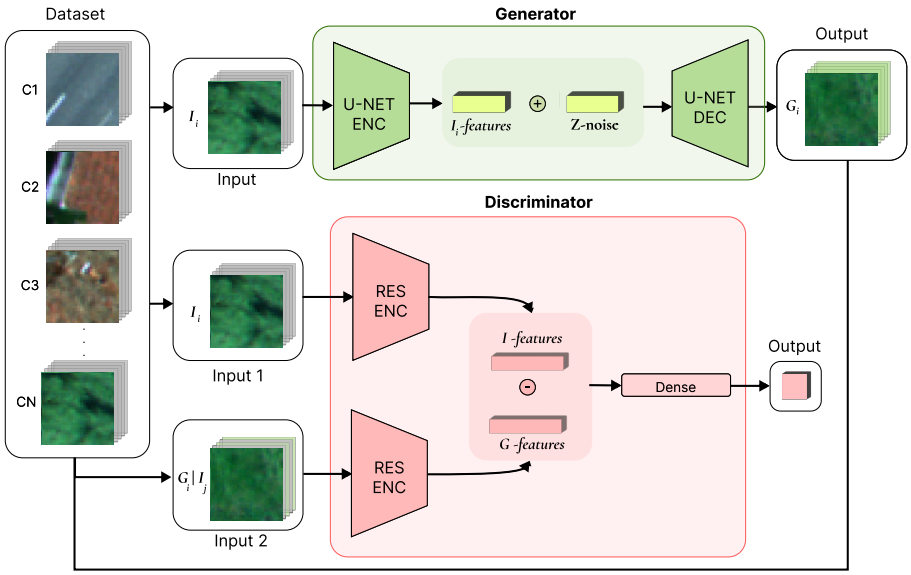

Fig. 1: Architecture of the mDAGAN data augmentation technique.

all the spectral bands available in the multispectral images.

The design of the discriminator in a GAN is very relevant to avoid the so-called *mode collapse* during training, i.e., the loss of discrimination capability of the network. In this case, the discriminator network, also shown in Figure 1, presents with respect to the DAGAN network the substantial modification of including two siamese subnetworks that share the same architecture. It receives two patches of the image as inputs, and each one is processed by one of the siamese networks (RES ENC in figure 1). Figure 2 shows the architecture of the RES ENC module. RES ENC is a residual network with three concatenated blocks of convolutional layers and a dense layer. Each convolutional block has a residual connection with the dense layer. The last layer in each block applies a stride of 2. The difference between the feature vectors produced by the siamese networks is introduced in the dense layer producing a similarity measure represented by a scalar value as output. The discriminator operates in two stages: in the first stage, it receives a pair of real patches from the same class, and in the second stage, the inputs are a pair of patches from the same class, one real and one generated.

The training process of mDAGAN is divided in two stages. During the first stage, critic iterations are performed: the generator network weights are fixed, and the discriminator network weights are trained based on the loss function $Loss_{disc}$ shown in equation 1. During the second stage, the discriminator network weights are fixed and the generator network weights are trained based on its loss function, $Loss_{gen}$ shown in equation 2.

The following loss functions for the discriminator $Looss_{dic}$ and the generator $Looss_{gen}$ are proposed:

$$Loss_{disc} := \sqrt{|disc(I_i, G_i) - disc(I_i, I_j) + gp|}. \quad (1)$$

$$Loss_{gen} := \sqrt{|disc(I_i, I_j) - disc(I_i, G_i)|}. \quad (2)$$

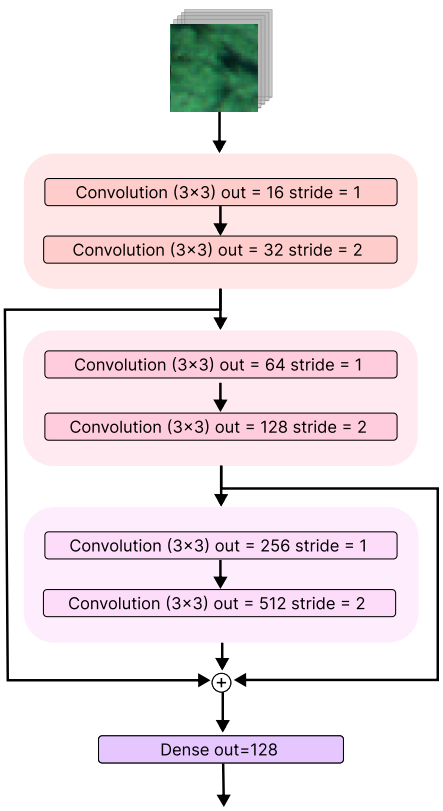

Fig. 2: Residual Encoder (RES ENC in Figure 1).

The higher the disc value for a real and a generated image, $disc(I_i, G_i)$, the higher the value of $Loss_{disc}$. $Loss_{gen}$ has the opposite behaviour. These loss functions avoid possible unbalances between the generation and the discriminator that would cause a mode collapse in the network, and consequently would make it impossible to train the augmentation model.

III. Experimental results

The dataset used for the experiments consists of very high-resolution (10 cm/pixel) multiespectral images of vegetation captured by a MicaSense RedEdge-

MX sensor that features five spectral bands [12]. The distribution of samples in the ten available classes is as indicated in Table I. These classes range from native vegetation to human-made structures such as roads or buildings. It is important to highlight the strong imbalance in number of samples among classes in all datasets. This imbalance introduces a bias in the classification towards majority classes, potentially causing also imbalance in the classification accuracy across the different classes.

Class	Das Mestas River	Eiras Dam	Ermidas Creek
Water	0	3557	890
Bare soil	7076	867	874
Rocks	0	2923	1846
Asphalt	0	340	3597
Concrete	0	182	126
Tiles	0	47	608
Meadows	23 417	5722	13 052
Native trees	2973	13 466	4559
Pines	0	687	1527
Eucalyptus	23 592	74	8984
Total samples:	57 058	27 865	36 063

Table I: Dataset description. The number of samples is the number of labelled samples available. It is reported in number of superpixels after applying segmentation by the waterpixel algorithm.

The experiments were executed in a server using 8 Intel E5-2623 v4 cores, 64 GB of RAM memory, and a Nvidia P40 GPU with 24 GB of memory.

Concerning the configuration of the algorithms for the experiments, the input patches are of 32x32 pixels. All the datasets were segmented using the Waterpixel algorithm [13] by choosing an average size of 400 px/superpixel, allowing a minimum size of 100 px/-superpixel, and employing a compactness factor of 0.5 points. The GAN model is trained during 500 epochs with a batch size of 32. One epoch of training is performed for the generator for each training epoch of the discriminator. The chosen optimizer was the PyTorch implementation of Adam [14]. The learning rate is 0.00001. The β_1 value is 0.1 for the discriminator and 0 for the generator. In the case of β_2 the value is 0.9 for both networks.

Four supervised classification techniques have been selected to test the effectiveness of mDAGAN. These are a linear Support Vector Machine (SVM), a Multilayer Perceptron (MLP), and a Random Forest classifier (RF). The Scikit-Learn [15] implementations of SVM, MLP and RF were used. The MLP was configured as two fully connected layers with 128 and 32 neurons, respectively, and the number of estimators for RF has been set to 1000. For all the experiments, the same 25 samples per class randomly selected used for training the mDAGAN are used for training the classifier.

Table II shows the results offered by the different classification techniques when mDAGAN data augmentation is applied for the high resolution multispectral images considered. The percentage of samples used for training and validation over the total size of each image are indicated below each image name in the first column. For obtaining the classification results, first the mDAGAN is trained, and then it is used for generation of the samples during the training stage of the classification technique. The classification results are expressed in terms of Average Accuracy (AA) and Overall Accuracy (OA) that are in %. Average Accuracy takes into account the imbalance in the number of samples across different classes. Overall accuracy is the proportion of correctly classified samples out of the total samples in the image.

As shown in Table II, two different percentages of synthetic samples were added to the training process for each image (10% and 20%). The results show similarity between the two configurations, suggesting that even a small percentage of augmented samples enhances the accuracy compared to no augmentation (denoted as "None" in the table). The performance of mDAGAN is compared against generating the same percentages of synthetic samples using a simple but standard technique involving randomly selected rotations in multiples of 90 degrees and horizontal flips applied to the samples.

The comparison summary, represented by the average values across all datasets (column "Average" in the table), indicates that the proposed deep learning-based data augmentation yields the best results among the three classification techniques analyzed. While the improvements are modest, the increase in average accuracy (AA) demonstrates that the effect is consistent across different classes of datasets, even when dealing with unbalanced classes.

IV. CONCLUSIONS

In this paper, a data augmentation technique for the classification of remote sensing multispectral images called mDAGAN is proposed. mDAGAN is based on a GAN networks and leverages all the multispectral and spatial information available in the image, accelerating the convergence of the model and reducing its computational cost. The results with three different supervised classification techniques (SVM, RF, and MLP) show a moderate increase of around 1% in classification accuracy for all the high-spatial resolution multispectral images considered regardless of the classification technique applied.

Although these results are promising, further research is required. Firstly, it is necessary to explore more configurations of the GAN network and other parameters involved in the technique, including different sizes of the initial segmentation process. Since the images used in this work are of vegetation, it would also be beneficial to conduct experiments with images from different remote sensing domains and varying spatial and spectral resolutions. Additionally, analyzing the results obtained for deep learning classifiers would be necessary to generalize the findings. Finally, as the network is already optimized to reduce computational costs, the next step would be to investigate the possibility of applying parallel

Image	DA technique	SVM		RF		MLP	
		AA	OA	AA	OA	AA	OA
Das Mestas River 0.4 %	None	71.92	74.25	71.77	72.34	68.91	73.76
	mDAGAN-10 %	71.39	73.01	72.18	72.06	69.48	72.66
	mDAGAN-20 %	71.25	72.54	72.21	72.06	69.22	73.34
	Rotate+Flip 10 %	71.46	74.43	71.72	72.05	68.44	73.36
	Rotate+Flip 20 %	70.97	73.67	71.05	71.86	68.52	72.82
Eiras Dam 1.8 %	None	66.21	64.53	72.34	67.70	64.01	59.26
	mDAGAN-10 %	69.15	64.93	73.62	68.56	65.47	64.05
	mDAGAN-20 %	67.77	64.51	73.42	69.17	65.34	60.79
	Rotate+Flip 10 %	67.49	64.81	73.17	68.09	65.70	61.67
	Rotate+Flip 20 %	66.22	64.57	73.16	67.92	65.72	59.43
Ermidas Creek 1.4 %	None	68.03	58.28	68.34	64.60	65.72	67.77
	mDAGAN-10 %	68.75	59.81	69.81	65.71	66.40	69.89
	mDAGAN-20 %	68.87	66.37	68.61	66.98	66.78	70.73
	Rotate+Flip 10 %	67.96	58.20	68.77	63.73	66.37	68.71
	Rotate+Flip 20 %	68.61	58.90	69.23	65.37	66.20	68.70
Average	mDAGAN	**69.53**	**66.86**	**71.64**	**69.09**	**67.11**	**68.58**
	Rotate+Flip	68.78	65.76	71.18	68.17	66.82	67.45

Table II: Comparative results for different classification techniques. The best results for each technique are in bold. AA and OA in %.

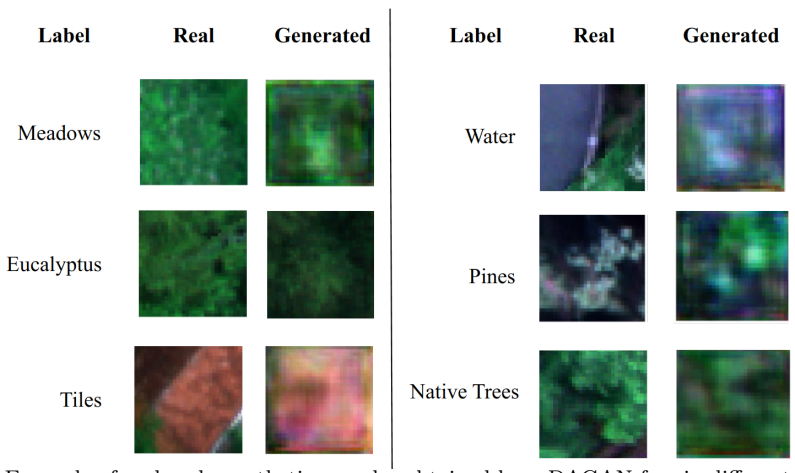

Fig. 3: Example of real and snynthetic samples obtained by mDAGAN for six different classes.

computation to further reduce these costs.

References

[1] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan, "Synthetic data augmentation using gan for improved liver lesion classification," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2018-April, pp. 289–293, 5 2018.

[2] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller, "Data augmentation using synthetic data for time series classification with deep residual networks," 8 2018.

[3] Celso M. de Melo, Antonio Torralba, Leonidas Guibas, James DiCarlo, Rama Chellappa, and Jessica Hodgins, "Next-generation deep learning based on simulators and synthetic data," *Trends in Cognitive Sciences*, vol. 26, pp. 174–187, 2 2022.

[4] Álvaro Acción, Francisco Argüello, and Dora B. Heras, "A new multispectral data augmentation technique based on data imputation," *Remote Sensing 2021, Vol. 13, Page 4875*, vol. 13, pp. 4875, 11 2021.

[5] Connor Shorten and Taghi M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, pp. 1–48, 12 2019.

[6] Álvaro Acción, Francisco Argüello, and Dora B. Heras, "Dual-window superpixel data augmentation for hyperspectral image classification," *Applied Sciences 2020, Vol. 10, Page 8833*, vol. 10, pp. 8833, 12 2020.

[7] Alvaro G. Dieste, Francisco Arguello, and Dora B. Heras, "ResBaGAN: A residual balancing gan with data augmentation for forest mapping," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 6428–6447, 2023.

[8] Jun Hyung Kim and Youngbae Hwang, "Gan-based synthetic data augmentation for infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, 2022.

[9] Vinay Kukreja, Deepak Kumar, Amandeep Kaur, Geetanjali, and Sakshi, "Gan-based synthetic data augmentation for increased cnn performance in vehicle number plate recognition," *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2020*, pp. 1190–1195, 11 2020.

[10] Antreas Antoniou, Amos Storkey, and Harrison Edwards, "Data augmentation generative adversarial networks," *Journal of Physics A: Mathematical and Theoretical*, vol. 44, pp. 1–13, 11 2017.

[11] Sumit Chopra, Raia Hadsell, and Yann LeCun, "Learning a similarity metric discriminatively, with application to face verification," *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. I, pp. 539–546, 2005.

[12] Francisco Argüello, Dora B. Heras, Alberto S. Garea,

and Pablo Quesada-Barriuso, "Watershed monitoring in galicia from UAV multispectral imagery using advanced texture methods," *Remote Sensing 2021, Vol. 13, Page 2687*, vol. 13, pp. 2687, 7 2021.

[13] Francisco Argüello, Dora B. Heras, Alberto S. Garea, and Pablo Quesada-Barriuso, "Watershed monitoring in Galicia from UAV multispectral imagery using advanced texture methods," *Remote Sensing*, vol. 13, no. 14, 2021.

[14] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," 2017.

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.