CYCLE GENERATIVE ADVERSARIAL NETWORK: TOWARDS A LOW-COST VEGETATION INDEX ESTIMATION

Patricia L. Suárez [1], *Angel D. Sappa*[1,2], *Boris X. Vintimilla*[1]

[1]Escuela Superior Politécnica del Litoral, ESPOL,
Facultad de Ingeniería en Electricidad y Computación,CIDIS, Campus
Gustavo Galindo, Guayaquil, Ecuador; {plsuarez, asappa,boris.vintimilla}@espol.edu.ec.

[2] Computer Vision Center, Edifici O, Campus UAB, 08193, Bellaterra, Barcelona, Spain;
asappa@cvc.uab.es

ABSTRACT

This paper presents a novel unsupervised approach to estimate the Normalized Difference Vegetation Index (NDVI). The NDVI is obtained as the ratio between information from the visible and near infrared spectral bands; in the current work, the NDVI is estimated just from an image of the visible spectrum through a Cyclic Generative Adversarial Network (CyclicGAN). This unsupervised architecture learns to estimate the NDVI index by means of an image translation between the red channel of a given RGB image and the NDVI unpaired index's image. The translation is obtained by means of a ResNET architecture and a multiple loss functions. Experimental results obtained with this unsupervised scheme show the validity of the implemented model. Additionally, comparisons with the state of the art approaches are provided showing improvements with the proposed approach.

Index Terms— CyclicGAN, NDVI, near infrared spectra, instance normalization

1. INTRODUCTION

Computer vision is a technology that, combined with machine learning and remote sensing, allows computers to understand and estimate the quantity, quality, and condition of crops. These estimations can be made based on the intensity of radiation reflected by certain bands of the electromagnetic spectrum. Nowadays, the usage of unmanned aerial vehicles, which can incorporate sensors sensitive to near infrared (NIR), in addition to the visible, are considered to simultaneously acquire images of the same scene in different spectra. With such cross-spectral information, more efficient solutions can be implemented to help farmers with their crops to apply more efficient growth methods, increase yields and profits, [1] [2]. Whereby, the agricultural industry has adopted the use of these new technologies to automate all activities related to improve yields productivity, decrease rising labor costs and prepare farmers to be ready to face increasingly aggressive globalized competition.

A vegetation index is a single value that quantifies vegetation biomass and/or plant health for each pixel in a remote sensing image. Among the different indexes proposed in the literature, the NDVI is the most widely used [3]; This vegetation index is often used to monitor drought, forecast agricultural production, assist in forecasting fire zones and desert offensive maps [4]. This index is calculated as the ratio between the difference and sum of the reflectance in NIR and red regions:

$$NVDI = \frac{R_{\text{NIR}} - R_{\text{RED}}}{R_{\text{NIR}} + R_{\text{RED}}}, \tag{1}$$

where R_{NIR} is the reflectance of NIR radiation and R_{RED} is the reflectance of visible red radiation.

This index defines values from -1.0 to 1.0, basically representing greens, where negative values are mainly formed from clouds, water and snow, and values close to zero are primarily formed from rocks and bare soil. Very small values (0.1 or less) of the NDVI function correspond to empty areas of rocks, sand, or snow. Moderate values (from 0.2 to 0.3) represent shrubs and meadows, while large values (from 0.6 to 0.8) indicate temperate and tropical forests [5], [6].

Although NDVI has been largely used in the agricultural sector, its main limitation is related to the need of having two cameras, one in the visible and one in the NIR spectral bands. This requirement becomes a drawback if we consider these cameras have to be on-board a UAV or satellite—i.e., power consumption of the cameras and weight of the whole system. Hence, trying to tackle this drawback, this work proposes the NDVI estimation just by using images from the visible spectrum (the red channel). In the proposed approach, an unsupervised learning model with a set of unpaired images is used as an input, one from the visible spectrum and the other corresponds to an NDVI image. The manuscript is organized as

Thanks to XYZ agency for funding.

follows. The proposed approach is detailed in Section 2. Experimental results with a set of real images are presented in Section 3. Finally, the conclusions are given in Section 4.

2. PROPOSED APPROACH

The architecture used in the current work is based on the one presented on [7], a previous work that presents an unpaired image to image translation, through a cycle generative adversarial network (CyclicGAN). This type of network allows domain style transfer, which is a convenient method for image-to-image translation because it is not necessary to have a set of input images that capture the scene at the same time and place from different spectra. Before presenting the proposed approach, a brief description of CyclicGAN is given.

2.1. Cyclic Generative Adversarial Networks

Image-to-image translation is the process of transforming an image from one domain to another, where the goal is to learn the mapping between an input and an output image. This task has been generally performed by using a training set of aligned image pairs. However, for many tasks, paired training data are not available, and to prepare them often takes a lot of work from experts to obtain thousands of paired image datasets, especially with complex image translations [7]. CyclicGAN is an architecture to address this problem because it learns to perform image translations without explicit pairs of images. In our case, we use a translation of unpaired images. Thus, the goal is to learn a mapping G: X $\rightarrow Y$ such that the distribution of images from $G(X)$ is indistinguishable from the distribution of Y using an adversarial loss, [8]. Because this mapping is highly under-constrained, it is necessary an inverse mapping F: $Y \rightarrow X$ and introduce a cycle consistency loss to enforce $F(G(X)) \approx X$ (and vice versa).

2.2. Instance Normalization

GANs are a framework in which two networks compete with each other, with the objective of obtaining that the training process between the generator G and discriminator D finds an equilibrium—the Nash equilibrium [9], [8]. Much of the recent work on GANs is focused on developing techniques to stabilize training. To improve the stability during the GAN's training phase, one kind of analysis has emerged around the style of an image evaluated by the statistics of convolutional neural network filters, a renewed interest in the texture generation, and image stylization problems in order to obtain qualitative improvement in the generated image.

Ulyanov et al. [10] show that it is possible to train a generator network $G(s, z)$ that can apply to a given input image s the style of another $s0$. They introduce a method named ***instance normalization*** for a better stylization and texture synthesis, which derive entropy loss that improves sample diversity. According to [10], the generator network should discard contrast information in the content image to learn a highly nonlinear contrast normalization function as a combination of such layers. Let $s \in \mathbb{R}^{NCWH}$ be an input tensor containing a batch of N images, where C, W and H are the depth, width and high respectively of the image tensor and let s_{tijk} denote the $tijk$-th element of s image tensor, where k and j span spatial dimensions, t is the index of the image in the batch and i is the feature channel (in the case of a RGB image being used as an input, it would represent a color channel). Thus, a simple version of instance normalization is defined as:

$$y_{tijk} = \frac{s_{tijk}}{\sum_{l=1}^{W} \sum_{m=1}^{H} s_{tilm}}. \qquad (2)$$

2.3. Proposed Architecture

Our work focuses on the estimation of the NDVI vegetation index using the red channel of RGB images. The proposed model can learn to translate the images between the red channel to their corresponding NDVI indexes, using an unpair dataset. This allows us to transform from RGB to NDVI. For simplicity, NDVI indexes are represented as image values so hereinafter the terms NDVI indexes and NDVI images will be indistinctly used. Another advantage of the proposed network to generate the synthetic NDVI images is that, undoubtedly, the costs of agricultural solutions may decrease, since there would be no need to acquire sensors sensitive to NIR spectra together with all the costs associated with the synchronized image acquisition and registration.

Our architecture uses a modified residual block from (ResNET) [11] to perform the image transformation from one spectrum to vegetation index. In order to avoid the vanishing gradient problem, we define the residual function using $F(x)$ = $H(x)$ — x, which can be reframed into $H(x)$ = $F(x) + x$, where $F(x)$ and x represents the stacked non-linear layers and the identity function (input=output) respectively. With this architecture design, it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping. Additionally, we introduce instance normalization, to the original residual block to improve the quality of the NDVI image obtained by the network.

Figure 1 depicts the CyclicGAN model proposed in the current work. It is composed of two generators (G, F) and two discriminators (Dx, Dy). In order to generate a synthetic NDVI vegetation index, the architecture takes advantage of the joint of cycle consistency and least square losses [12] in addition to the usual discriminator and generator losses. The experimental results have shown that these loss functions demand that the model maintains textural information of the visible (corresponding red channel) and NDVI images and generate uniform synthetic outputs. According with [7] the objective of a CyclicGAN is to learn mapping functions between two domains X and Y given training samples ${x_i}_{i=1}^{N} \in$

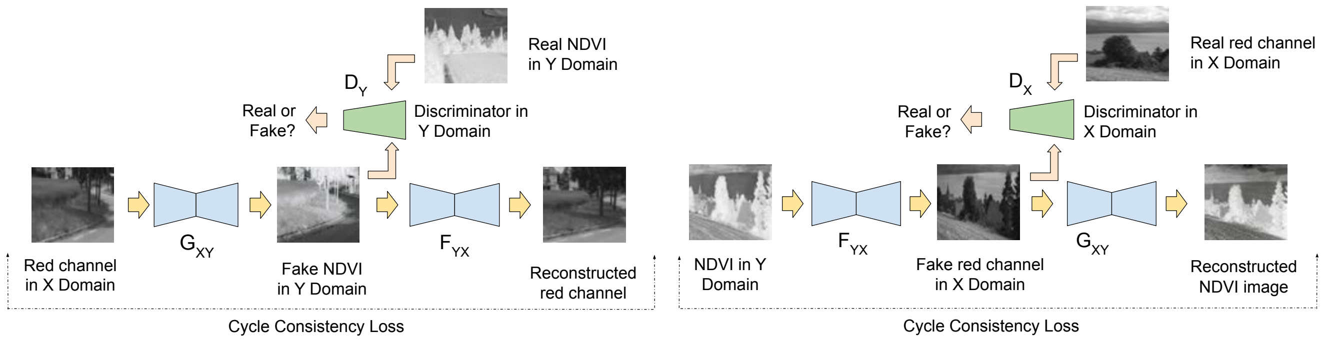

Fig. 1. Cycle generative adversarial model: F:$Y(NDVI) \rightarrow$ *X(red channel)* and its discriminator D_x and G: *X(red channel)* $\rightarrow Y(NDVI)$ and its discriminator D_y.

X and $y_{i}{}_{i=1}^{N} \in Y$. The loss functions applied to the model are detailed below.

2.4. Loss Functions

The adversarial losses, according to [8], are applied to both mapping functions. For the mapping function $G : X \rightarrow Y$ and its discriminator D_y and vice versa. Also, according to [7], for each image x from domain X, the image translation cycle should be able to bring x back to the original image, i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ and vice versa, calling this forward cycle consistency. To solve the problem of disappearing gradients, we have implemented a least squares GAN loss function (LSGAN), which is briefly detailed below.

Least squares GAN's loss

Generative Adversarial Networks [8] have demonstrated impressive performance for unsupervised learning tasks. GANs do not require any approximation and can be trained end-to-end through the differentiable networks. Regular GANs adopt the sigmoid cross-entropy loss function for the discriminator [12]. This loss generates samples that are closer to real data, in our case the synthetic NDVI image. The experiments performed with this loss instead of negative log-likelihood shown better results and replaced the adversarial losses presented by [8]. The least square losses are defined as :

$$\mathcal{L}_{LSGAN}(G_{D_y}) = \mathbb{E}_y \sim_{p \text{ data}(y)}[(D_Y(y)-1)^2] + \mathbb{E}_x \sim_{p \text{ data}(x)}[D_Y(G(x))^2], \quad (3)$$

and

$$\mathcal{L}_{LSGAN}(F_{D_x}) = \mathbb{E}_x \sim_{p \text{ data}(x)}[(D_X(x)-1)^2] + \mathbb{E}_y \sim_{p \text{ data}(y)}[(D_X(F(y))^2)]. \quad (4)$$

For this unsupervised approach, the standard CYCLE-GAN $\mathcal{L}_{CYCLE}$: (cycle-consistent loss) and $\mathcal{L}_{LSGAN}$: (least square loss), have been implemented, both with their corresponding weights distributions for the multiple loss functions. For the first loss, the weighted sum of the individual loss function terms designed to obtain the best results is defined as:

$$\mathcal{L}_{RED_{CY-GAN}} = 0.38\,\mathcal{L}_{GAN} + 0.62\,\mathcal{L}_{CYCLE} \quad (5)$$

The second loss evaluated in this approach is the LSGAN loss, where the weighted sum of the individual loss function terms is defined as:

$$\mathcal{L}_{RED_{CY-LSGAN}} = 0.65\,\mathcal{L}_{LSGAN} + 0.35\,\mathcal{L}_{CYCLE} \quad (6)$$

The combination of the weights associated with each loss function is focused on improving the quality of the images for human perception and at the same time, they are used as regularization terms that determine which loss function is the most significant in the optimization of the model.

3. RESULTS AND DISCUSSIONS

3.1. Datasets for training and testing

The proposed approach has been evaluated using the red channel of RGB images and their corresponding NVDI vegetation indexes from [15]. From the aforementioned data set the *country*, *mountain* and *field* categories have been considered for evaluating the performance of the proposed approach. This dataset consists of 477 registered images categorized in 9 groups captured in visible and near infrared spectrum. For each image, the NDVI index is computed from the red channel, together with the corresponding NIR image. This couple of images (i.e., NDVI index and red channel) are used unpaired during the training stage.

Training	RMSE			SSIM		
	country	*field*	*mountain*	*country*	*field*	*mountain*
Results from [13]	3.53	3.70	–	0.94	0.91	–
Results from [7]	3.46	3.53	3.82	0.93	0.90	0.88
Results from [14]	3.39	3.56	3.81	0.94	0.92	0.89
NDVI estimation with $\mathcal{L}_{RED_{CY-GAN}}$	3.37	3.50	3.72	0.94	0.92	0.90
NDVI estimation with $\mathcal{L}_{RED_{CY-LSGAN}}$	3.15	3.11	3.20	0.94	0.92	0.91

Table 1: Average Root Mean Square Errors (RMSE) and Structural Similarities (SSIM) obtained from the estimated NDVI vegetation index and the real one computed from eq. (1) (SSIM the bigger the better). Note NDVI values are scaled up to a range of [0-255] since they are depicted as images as shown in Fig. 2.

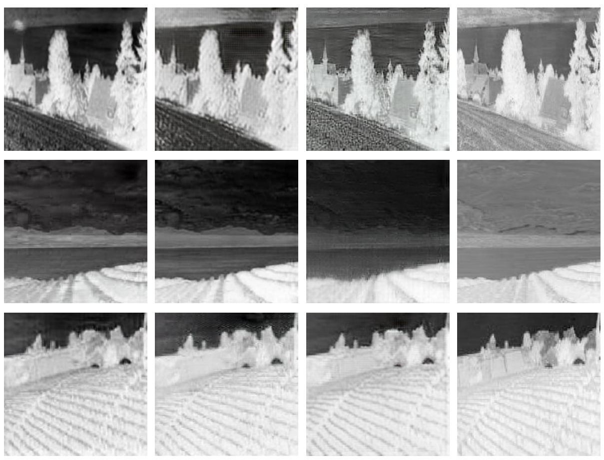

Fig. 2. Images of NDVI vegetation indexes obtained with the proposed CycleGAN: (*1st col*) NDVI estimated with [7]; (*2nd col*) NDVI estimated by the proposed CyclicGAN; (*3rd col*) NDVI estimated by the proposed CyclicGAN with LSGAN; (*4th col*) Ground truth NDVI vegetation index. Images from [15], *mountain* category.

3.2. Evaluation Metrics

Digital images resulting from an artificial intelligence process, such as deep neural networks, are subject to a wide variety of distortions. In an image-based technique, image quality is a prime criterion. Commonly, for a good image quality evaluation, complete reference metrics are applied, like Mean Square Error, one of the most used image quality metrics. Also, a perceptual metric that measures image quality level, Structured Similarity Indexing Method [16], has been developed to compare the structural and feature similarity measures between restored and original objects. For our approach, we use RMSE and SSIM metrics, with which we are able to obtain consistent and better results.

3.3. Experimental Results

Results from the Cyclic Generative Adversarial Network (synthetic NDVI images) are presented in Figure 2, Where the illustrations of the estimation of the vegetation indices obtained are shown. Quantitative evaluations are presented in Table 1. In this table average root mean square errors of the NDVI values, scaled up to a range of [0-255], and structural similarity index metric computed over the validation set are depicted, when different combinations of the proposed loss functions are considered. We implement the least square loss to accelerate and maintain a stable training process. Additionally, in this Table, results from Conditional GAN ([13] and [14]) and standard CyclicGAN [7] are presented. It can be appreciated that in all the cases the results obtained with the least square loss in the proposed CyclicGAN are better than those obtained with the other approaches. To increase the cyclic loss effect over the network we use $L1$ (λ). The network has been trained using Stochastic AdamOptimazer. The image dataset is normalized in a (-1,1) range.

4. CONCLUSIONS

This paper proposes a novel architecture to estimate the NDVI vegetation index using a Cycle-Consistent Adversarial Network, in order to avoid the dependence on NIR sensors. This novel approach tackles the challenging problem of synthesizing NDVI images from a single channel (red) of an RGB representation. Experimental results have shown that the NDVI images estimated with our approach are better than those obtained with the standard CyclicGAN model. The quantitative values presented also show better results than previous approaches for NDVI index estimation.

Acknowledgements

This work has been partially supported by the ESPOL Polytechnic University; the Spanish Government under Project TIN2017-89723-P; and the "CERCA Programme / Generalitat de Catalunya". The authors gratefully acknowledge the support of the CYTED Network: "Ibero-American Thematic Network on ICT Applications for Smart Cities" (REF-518RT0559) and the NVIDIA Corporation with the donation of the Titan Xp GPU used for this research..

5. REFERENCES

[1] Salvatore Filippo Di Gennaro, Fulvia Rizza, Franz Werner Badeck, Andrea Berton, Stefano Delbono, Beniamino Gioli, Piero Toscano, Alessandro Zaldei, and Alessandro Matese, "UAV-based high-throughput phenotyping to discriminate barley vigour with visible and near-infrared vegetation indices," *International journal of remote sensing*, vol. 39, no. 15-16, pp. 5330–5344, 2018.

[2] M Fernanda Dreccer, Gemma Molero, Carolina Rivera-Amado, Carus John-Bejai, and Zoe Wilson, "Yielding to the image: How phenotyping reproductive growth can assist crop improvement and production," *Plant science*, vol. 282, pp. 73–82, 2019.

[3] J_W Rouse Jr, RH Haas, JA Schell, and DW Deering, "Monitoring vegetation systems in the great plains with erts," *NTRS, NASA Technical Reports Server*, 1974.

[4] Sergii Skakun, Christopher O Justice, Eric Vermote, and Jean-Claude Roger, "Transitioning from modis to viirs: an analysis of inter-consistency of ndvi data sets for agricultural monitoring," *International journal of remote sensing*, vol. 39, no. 4, pp. 971–992, 2018.

[5] Toby N Carlson and David A Ripley, "On the relation between ndvi, fractional vegetation cover, and leaf area index," *Remote sensing of Environment*, vol. 62, no. 3, pp. 241–252, 1997.

[6] Amanda Heemann Junges, Denise Cybis Fontana, and Cristian Scalvi Lampugnani, "Relationship between the normalized difference vegetation index and leaf area in vineyards," *Bragantia*, vol. 78, no. 2, pp. 297–305, 2019.

[7] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[9] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen, "Improved techniques for training gans," in *Advances in neural information processing systems*, 2016, pp. 2234–2242.

[10] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6924–6932.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[12] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2794–2802.

[13] Patricia L Suárez, Angel D Sappa, and Boris X Vintimilla, "Vegetation index estimation from monospectral images," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 353–362.

[14] Patricia L. Suarez, Angel D. Sappa, Boris X. Vintimilla, and Riad I. Hammoud, "Image vegetation index through a cycle generative adversarial network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019, pp. 301–310.

[15] Matthew Brown and Sabine Süsstrunk, "Multi-spectral SIFT for scene category recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 177–184.

[16] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.