Vegetation Index Estimation from Monospectral Images

Patricia L. Suárez[1](✉), Angel D. Sappa[1,2], and Boris X. Vintimilla[1]

[1] Facultad de Ingeniería en Electricidad y Computación, CIDIS,
Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo,
09-01-5863, Guayaquil, Ecuador
{plsuarez,asappa,boris.vintimilla}@espol.edu.ec
[2] Computer Vision Center, Edifici O, Campus UAB, 08193
Bellaterra, Barcelona, Spain

Abstract. This paper proposes a novel approach to estimate Normalized Difference Vegetation Index (NDVI) from just the red channel of a RGB image. The NDVI index is defined as the ratio of the difference of the red and infrared radiances over their sum. In other words, information from the red channel of a RGB image and the corresponding infrared spectral band are required for its computation. In the current work the NDVI index is estimated just from the red channel by training a Conditional Generative Adversarial Network (CGAN). The architecture proposed for the generative network consists of a single level structure, which combines at the final layer results from convolutional operations together with the given red channel with Gaussian noise to enhance details, resulting in a sharp NDVI image. Then, the discriminative model estimates the probability that the NDVI generated index came from the training dataset, rather than the index automatically generated. Experimental results with a large set of real images are provided showing that a Conditional GAN single level model represents an acceptable approach to estimate NDVI index.

1 Introduction

Computer vision applications can be found in almost every domain, including topics such as medical imaging, gaming, video surveillance, multimedia, industrial applications, remote sensing, just to mention a few. In most of the cases these applications are based on images obtained from a cameras working at the visible spectrum. There are some cases, in particular in medical imaging and remote sensing, where cross-spectral and multispectral images are considered. The appealing factor of using images from different spectral bands lies on the one hand on the possibility to obtain information that cannot be seen at the visible spectrum; on the other hand, on the combined use of information that can be considered to generate some kind of high level reasoning; for instance in remote sensing the combined usage of images from different spectral bands is considered to generate vegetation indexes. These vegetation indexes are used

A. Campilho et al. (Eds.): ICIAR 2018, LNCS 10882, pp. 353–362, 2018.
https://doi.org/10.1007/978-3-319-93000-8_40

to determine the health and strength of vegetation and their definitions involve several factors, like soil reflectance, atmosphere, vegetation density, etc.

Among the different indexes proposed in the literature, the Normalized Difference Vegetation Index (NDVI) is the most widely used [1]; in general, it is used to determine the condition, developmental stages and biomass of cultivated plants and to forecasts their yields. The values of this index go from -1 to 1, with the value zero representing the approximate where the absence of vegetation begins. Negative values represent non-vegetated surfaces. This index is calculated as the ratio between the difference and sum of the reflectance in NIR and red regions:

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

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

Although interesting, cross/multi-spectral solutions need the set up of more than one camera. For instance, in the case of NDVI, an image from the visible and an image from the NIR spectra are required. In other words, we need two cameras, acquiring images at the same time of the same scene, in order to be able to compute the values of Eq. (1). It should be noticed that before computing Eq. (1) images need to be accurately registered—i.e. the information should be referred to the same reference system. Unfortunately, since images from different spectra are considered their may look different, so the problem is how to find the same set of points in both spectra [2] to be used as references. Recently, deep learning based approaches have been proposed to overcome this drawback and find correspondences in cross-spectral domains [3,4]. Once these points are obtained we can proceed by registering the images in a single reference system.

Cross/multi-spectral approaches provide unique solutions to different complex problems, however, as mentioned above, different preprocessing stages need to be performed before computing these solutions; hence, in the current work we wonder whether it is possible to obtain the same result but just using information from a single spectral band. Actually, a similar philosophy has been recently presented in [5] where vegetation index is estimated based on a learning approach from a single near infrared spectral band image. Although interesting results have been obtained, the weakness point of that approach lies on the need of having NIR images, which are not that much common like visible spectrum images. In the current work we propose to explore the possibility to estimate NDVI vegetation index using the red channel from the visible spectrum. The index is estimated from a learning based approach, where a Conditional Generative Adversarial Network (CGAN) is trained with a large data set. The CGAN architecture used in the current work is similar to the one presented in [6], but including a conditional red channel image at the final layer of the learning model to improve the details of the generated NDVI vegetation index. Additionally, a more elaborated loss function is proposed to preserve details of the given image.

The rest of the paper is organized as follows. Section 2 introduces the Generative Adversarial Network formulation. Then, Sect. 3 presents the architecture

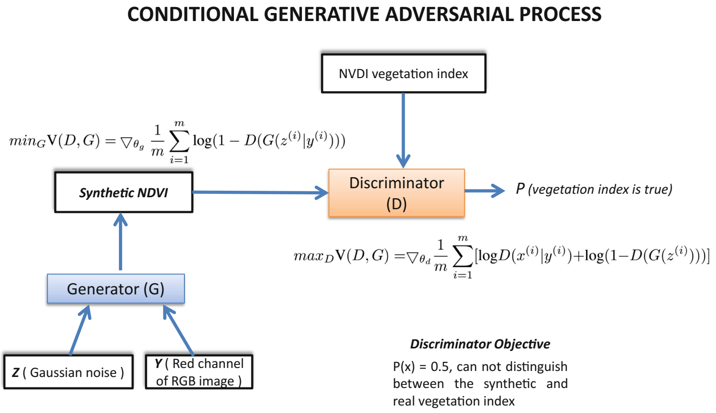
CONDITIONAL GENERATIVE ADVERSARIAL PROCESS
NVDI vegetation index
$min_G \mathrm{V}(D,G) = \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^{(i)}|y^{(i)})))$
Synthetic NDVI
Discriminator (D)
P (vegetation index is true)
$max_D \mathrm{V}(D,G) = \nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} [\log D(x^{(i)}|y^{(i)}) + \log(1 - D(G(z^{(i)})))]$
Generator (G)
Z (Gaussian noise)
Y (Red channel of RGB image)
Discriminator Objective
P(x) = 0.5, can not distinguish between the synthetic and real vegetation index

Fig. 1. Conditional Generative Adversarial process implemented on the current work to estimate NDVI vegetation index.

proposed in the current work, detailing the design, proposed loss functions and training with cross-spectral datasets. Section 4 depicts the experimental results and finally, conclusion are presented in Sect. 5.

2 Generative Adversarial Networks

Generative Adversarial Networks (GANs) are powerful and flexible tools quite useful in several computer vision problems; one of their most common applications is image generation. In the GAN framework [7], generative models are estimated via an adversarial process, in which simultaneously two models are trained: (i) a generative model G that captures the data distribution, and (ii) a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. In this architecture it is possible to apply certain conditions to improve the learning process. According to [8], to learn the generator's distribution p_g over data $\boldsymbol{x}$, the generator builds a mapping function from a prior noise distribution $p_z(z)$ to a data space $G(z;\theta_g)$ and the discriminator, $D(x;\theta_d)$, outputs a single scalar representing the probability that x came from training data rather than p_g. G and D are both trained simultaneously, the parameters for G are adjusted to minimize $log(1 - D(G(z)))$ and for D to minimize $logD(x)$ with a value function $V(G,D)$:

$$\frac{min}{G}\frac{max}{D} V(D,G) = \mathbb{E}_x{\sim}_{p\,\mathrm{data}_{(x)}}[logD(x)] + \\ \mathbb{E}_z \sim_{p\,\mathrm{data}_{(z)}} [log(1 - D(G(z)))]. \tag{2}$$

Generative adversarial networks can be extended to a conditional model if both the generator and discriminator are conditioned on some extra information y. This information could be any kind of auxiliary information, such as class labels or data from other modalities. We can perform the conditioning by feeding y into both discriminator and generator as additional input layer. The objective function of a two-player minimax game would be as:

$$\frac{min}{G}\frac{max}{D}V(D,G) = \mathbb{E}_x \sim_{p\,\mathrm{data}_{(x)}} [logD(x|y)] + \\ \mathbb{E}_z \sim_{p\,\mathrm{data}_{(z)}} [log(1 - D(G(z|y)))]. \tag{3}$$

In the current work a novel Conditional GAN model is proposed for vegetation index estimation from just the red channel of a RGB image; it is inspired on both the GAN network architecture presented in [9] for NIR colorization and on the triplet model proposed by [5] for learning vegetation indexes using NIR images. Actually, it is an adaptation of the architectures mentioned above, which consists of reducing the number of layers and removing the internal number of levels of learning architecture (FLAT or single). Another difference with previous approaches lies on the proposed loss function, which do not take into account only intensity level information but also it considers image structure information.

3 Proposed Approach

This section presents the approach proposed for NDVI index vegetation estimation. As mentioned above, it uses a similar architecture to the one presented in [5], where a conditional adversarial generative learning network has been proposed. A traditional scheme of layers in a deep network is considered. In the current work the usage of a Conditional GAN model is evaluated with a Flat scheme, this GAN's model has been used because it presented good performance to solve problems like colorization, dehazing, enhancement, object recognition, etc. Based on the results that have been obtained on this type of problems, where improvements in accuracy and performance have been obtained, we propose the usage of a learning model that allows the mapping of a vegetation index based on a single channel of RBG images (the red channel). The model will receive as an input a patch corresponding to red channel of a RGB image. Gaussian noise is added to each patch of the learning architecture to increase the variability in the learning process of the generation index patches, increasing the time of the convergence and generalization. A $l1$ regularization term has been added on each layer of the model in order to prevent the coefficients to overfit, which make the network learns small weights to minimize the loss, maximizes the distribution of model outputs, and improve the generalization capability of the model. Figure 1 depicts the Conditional GAN process proposed in the current work.

As mentioned above, in our case, the generator network has been implemented using a single level of layers (FLAT). Figure 2 presents an illustration of the GAN network used in this research. In all the cases, at the output of the generator network the vegetation index is obtained. This vegetation index will be validated by the discriminative network, which will evaluate the probability that the generated image (vegetation index in grayscale), is similar to the real

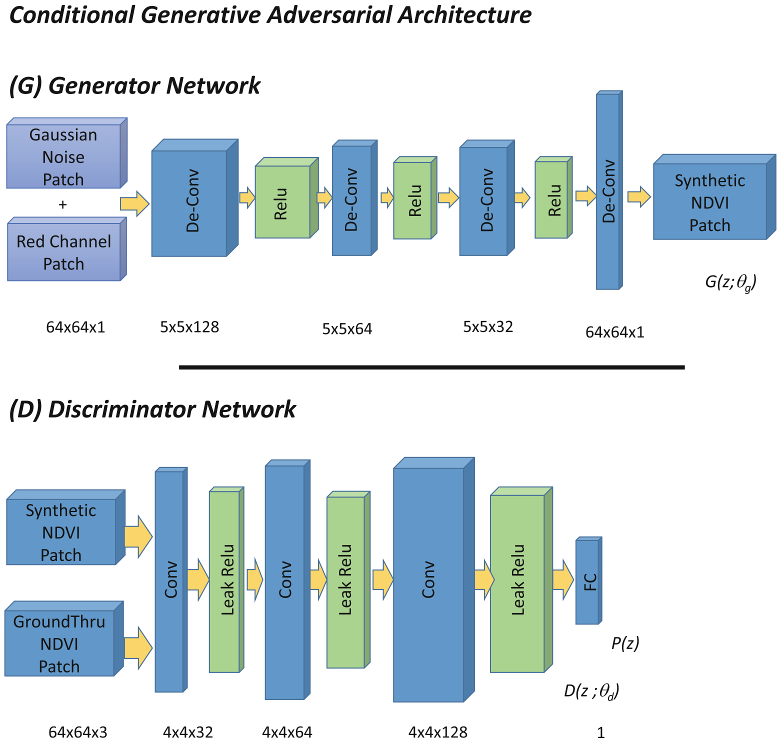

Fig. 2. GAN architecture for NDVI Vegetation Index estimation; A single level layer model (FLAT) evaluated as Generator Network; on bottom the Discriminator Network.

one used as a ground truth. Additionally, in the generator model, in order to obtain a better image representation, the CGAN framework is reformulated for a conditional generative image modeling tuple. In other words, the generative model $G(z;\theta_g)$ is trained from a red channel of a RGB image plus Gaussian noise, in order to produce a NDVI vegetation index image; additionally, a discriminative model $D(z;\theta_d)$ is trained to assign the correct label to the generated NDVI image, according to the provided real NDVI image, which is used as a ground truth. Variables (θ_g) and (θ_d) represent the weighting values for the generative and discriminative networks.

The model has been defined with a multi-term loss $(\mathcal{L})$ conformed by the combination of the Adversarial loss plus the Intensity loss (MSE) and the Structural loss (SSIM). This combined loss has been defined to avoid the usage of only a pixel-wise loss (PL) to measure the mismatch between a generated image and its corresponding ground-truth image. This multi-term loss function is better designed to human perceptual criteria of image quality, which is detailed next. The Adversarial loss is designed to minimize the cross-entropy to improve the texture loss :

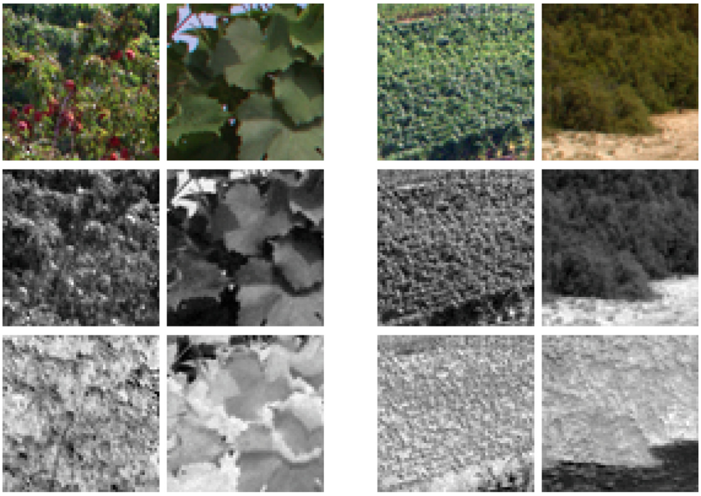

Fig. 3. Pairs of patches (64 × 64) from *country* category (two-left columns) and *field* category (two-right columns) [10]: (*top*) RGB image; (*middle*) Red channel of the given RGB image; (*bottom*) NDVI vegetation index computed from RGB images and the corresponding NIR images.

$$\mathcal{L}_{Adversarial} = -\sum_{i} logD(G_w(I_{z|y}), (I_{x|y}), \tag{4}$$

where D and G_w are the discriminator and generator of the real $I_{x|y}$ and generated $I_{z|y}$ images conditioned by the red channel of the RGB of the GAN network.

The Intensity loss is defined as:

$$\mathcal{L}_{Intensity} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(NDVIe_{i,j} - NDVIg_{i,j})^2, \tag{5}$$

where $NDVIe_{i,j}$ is the vegetation index estimated by the network and $NDVIg_{i,j}$ is the ground-truth vegetation index and $N \times M$ is the size of the patches. This loss measures the difference in intensity of the pixels between the images without considering texture and content comparisons. Additionally, this loss penalizes larger errors, but is more tolerant to small errors, without considering the specific structure in the image.

To address the limitations of the simple Intensity loss function, the usage of a reference-based measure is proposed. One of the reference-based index is the Structural Similarity Index (SSIM) [11], which evaluates images accounting for the fact that the human visual perception system is sensitive to changes in local structure; the idea behind this loss function is to help the learning model to produce a visually improved image. The Structural loss for a pixel p is defined as:

$$\mathcal{L}_{SSIM} = \frac{1}{NM}\sum_{p=1}^{P} 1 - SSIM(p), \tag{6}$$

where SSIM(p) is the Structural Similarity Index (see [11] for more details) centered in pixel p of the patch P.

The Final loss ($\mathcal{L}$) used in this work is the accumulative sum of the individual Adversarial, Intensity and Structural loss functions:

$$\mathcal{L}_{Final} = \mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity} + \mathcal{L}_{SSIM} \tag{7}$$

4 Experimental Results

The proposed approach has been evaluated using the red channel of RGB images and their corresponding NDVI vegetation index (ground truth), computed from Eq. (1) using NIR and red channel images; this cross-spectral data set came from [10]. The *country* and *field* categories have been considered for evaluating the performance of the proposed approach, examples of this dataset are presented in Fig. 3. This dataset consists of 477 registered images categorized in 9 groups captured in RGB (visible spectrum) and NIR (Near Infrared spectrum). The *country* category contains 52 pairs of images of (1024 × 680 pixels), while the *field* category contains 51 pairs of images of (1024 × 680 pixels). In order to train our network to generate vegetation index from each of these categories 380.000 pairs of patches of (64 × 64 pixels) have been cropped both, in the RGB images as well as in the corresponding NDVI images. Additionally, 3800 pairs of patches, per category, of (64 × 64 pixels) have been also generated for validation. It should be noted that images are correctly registered, so that a pixel-to-pixel correspondence is guaranteed.

Table 1. Root Mean Squared Errors (RMSE) and Structural Similarities (SSIM) obtained with the proposed GAN architecture by using different loss functions (SSIM the bigger the better).

Training	RMSE		SSIM	
	Country	*Field*	*Country*	*Field*
GAN with $\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity}$	3.93	4.12	0.86	0.83
GAN with $\mathcal{L}_{Adversarial} + \mathcal{L}_{SSIM}$	3.81	3.96	0.91	0.89
GAN with $\mathcal{L}_{Final}$	3.53	3.70	0.94	0.91

The Conditional Generative Adversarial network evaluated in the current work is a Flat (single level of learning layer) for NDVI vegetation index estimation. It has been trained using a 3.4 four core processor with 16GB of memory with a NVIDIA Titan XP GPU. Qualitative results are presented in Figs. 4 and 5. Figure 4 shows NDVI vegetation index images from the *country* category generated with the proposed Flat GAN network. Additionally, Fig. 5 shows NDVI vegetation index images from the *field* category generated with the proposed Flat GAN network. Quantitative evaluations for the different loss functions have been

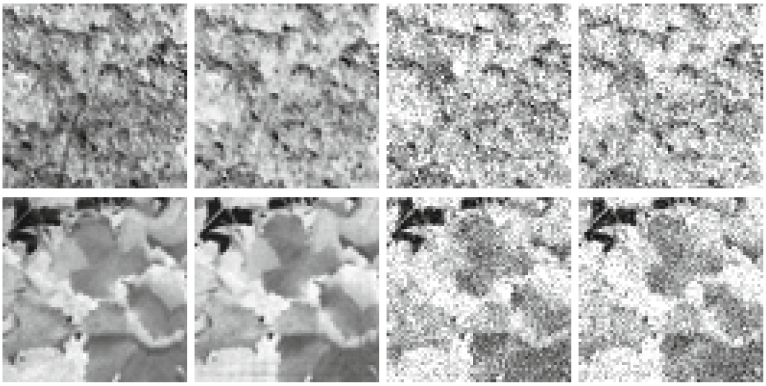

Fig. 4. ($1st.Col$) Ground truth NDVI index from the ***Country category***. ($2nd.-4th.Col$) NDVI index obtained with the proposed GAN architecture with different loss functions: $\mathcal{L}_{Final}$, $\mathcal{L}_{Adversarial} + \mathcal{L}_{SSIM}$ and $\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity}$.

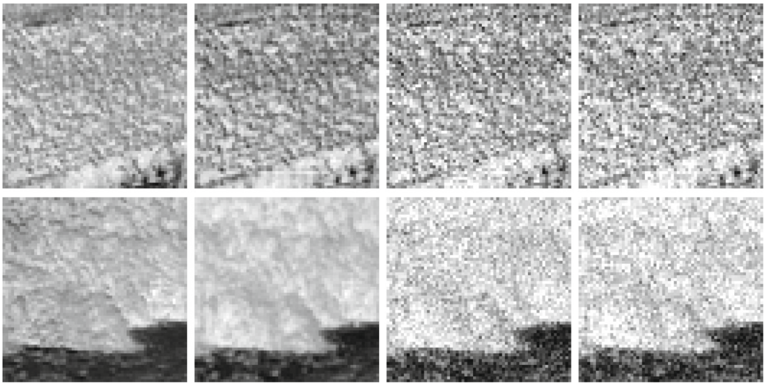

Fig. 5. ($1st.Col$) Ground truth NDVI index from the ***Field category***. ($2nd.-4th.Col$) NDVI index obtained with the proposed GAN architecture with different loss functions: $\mathcal{L}_{final}$, $\mathcal{L}_{Adversarial} + \mathcal{L}_{SSIM}$ and $\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity}$.

obtained and provided below. Up to our humble knowledge there are not previous work on similar technique to estimate vegetation index using only the red channel of RGB images. Hence, the only way to evaluate results is by comparing the Root Mean Square Error (RMSE) of each approach. The RMSE measures the distance between the estimated NDVI with respect to the ground truth, which is the standard deviation of the residuals. Residuals are measures of how different are the images compared from each other.

The results obtained with the multi-term loss approach show that the Structural Similarity metric contributes to improve the texture of the estimated NDVI vegetation index. Furthermore, the Intensity level loss function, which measure the Mean Square Error between the estimated value and the corresponding ground truth, helps to evaluate the estimation.

Table 1 presents the average Mean Square Errors (MSE) and the Structural Similarity metric (SSMI) obtained with the the single level architecture when different loss functions ($\mathcal{L}_{Adversarial} + \mathcal{L}_{SSIM}$), ($\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity}$) and ($\mathcal{L}_{Final}$) are evaluated in the two categories used as case studies. It can be appreciated that the results obtained with the $\mathcal{L}_{Final}$ loss function reaches the best result. The results obtained show that the more elaborated the loss function is, the better results will be obtained, since the network will be more capable to learn complex scenes at a faster convergence. Having in mind that the NDVI indexes resulting from the learning process are represented as images in the range of $[0, 255]$, the results presented in Table 1 show that the average deviation of the estimated values is 1.4%. Additionally, looking at the SSIM metric, which is a perception-based model that considers image degradation as perceived change in structural information, we can observe that on average, in both categories, results are above 0.9. This value means that obtained results highly pixels inter-dependencies. These dependencies carry important information about the structure of the objects in the visual scene. This metric combined with MSE allows us to confirm that the NDVI index obtained with the proposed results is a valid approach.

5 Conclusion

This paper tackles the challenging problem of NDVI vegetation index estimation by using a novel Conditional Generative Adversarial Network model. The novelty of the proposed approach lies on the usage of just a single spectral band (the red channel of RGB images). The architecture proposed for the generative network consists of a single level structure, which combines at the final layer results from convolutional operations together with the given red channel, resulting in a sharp NVDI image. Then, the discriminative model estimates the probability that the NDVI generated index came from the training dataset, rather than the index automatically generated. Different loss functions are evaluated trying to help the learning model to produce a visually improved image. The proposed loss function takes into account both intensity level information together with image structure information. Experimental results with a large set of outdoor images shows the validity of the proposed approach to estimate NDVI index from monospectral images. As a future work the possibility to obtain the NDVI from all the channels will be considered.

Acknowledgment. This work has been partially supported by: the ESPOL project PRAIM (FIEC-09-2015); the Spanish Government under Projects TIN2014-56919-C3-2-R and TIN2017-89723-P; and the "CERCA Programme/Generalitat de Catalunya". The authors would like to thank NVIDIA for GPU donations.

References

1. Rouse Jr., J., Haas, R., Schell, J., Deering, D.: Monitoring vegetation systems in the great plains with erts (1974)
2. Ricaurte, P., Chilán, C., Aguilera-Carrasco, C.A., Vintimilla, B.X., Sappa, A.D.: Feature point descriptors: infrared and visible spectra. Sensors **14**, 3690–3701 (2014)
3. Aguilera, C.A., Aguilera, F.J., Sappa, A.D., Aguilera, C., Toledo, R.: Learning cross-spectral similarity measures with deep convolutional neural networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, p. 9. IEEE (2016)
4. Suárez, P.L., Sappa, A.D., Vintimilla, B.X.: Cross-spectral image patch similarity using convolutional neural network. In: 2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM), pp. 1–5. IEEE (2017)
5. Suárez, P.L., Sappa, A.D., Vintimilla, B.X.: Learning image vegetation index through a conditional generative adversarial network. In: 2nd Ecuador Technical Chapters Meeting (2017)
6. Suárez, P.L., Sappa, A.D., Vintimilla, B.X.: Learning to colorize infrared images. In: De la Prieta, F., Vale, Z., Antunes, L., Pinto, T., Campbell, A.T., Julián, V., Neves, A.J.R., Moreno, M.N. (eds.) PAAMS 2017. AISC, vol. 619, pp. 164–172. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-61578-3_16
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
8. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
9. Suárez, P.L., Sappa, A.D., Vintimilla, B.X.: Infrared image colorization based on a triplet DCGAN architecture. In: Computer Vision and Pattern Recognition (2017)
10. Brown, M., Süsstrunk, S.: Multi-spectral SIFT for scene category recognition. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 177–184. IEEE (2011)
11. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)