Learning Image Vegetation Index through a Conditional Generative Adversarial Network

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, 09-01-5863, Guayaquil, Ecuador

[2]Computer Vision Center, Edifici O, Campus UAB,
08193, Bellaterra, Barcelona, Spain

{plsuarez, asappa, boris.vintimilla}@espol.edu.ec

***Abstract*—This paper proposes a novel approach to generate Normalized Difference Vegetation Index (NDVI) from just a near infrared (NIR) image. NDVI values are obtained by using images from the visible and infrared spectral bands. The proposed approach is based on the usage of a Conditional Generative Adversarial Network (CGAN) architecture model. In the first stage it learns how to generate the NDVI index from the given input image. Three different architectures are evaluated, flat, siamese and triplet models. In the evaluated models, the final layer of the architecture considers the infrared image to enhance the details, resulting in a sharp NVDI image. Then, in the second stage, a discriminative model is used to estimate the probability that the generated index cames from the training dataset, rather than the index automatically generated. In the experiments phase the three generative adversarial models were tested with the objective of determining which one generates NDVI values with the greatest similarity to the ones numerically calculated from the usage of visible and infrared images (ground truth). Experimental results with a large set of real images are provided showing that triplet model is the best one that reaches the best performance.**

***Index Terms*—Image vegetation index, Convolutional neural networks, Generative adversarial networks, Cross-spectral imaging.**

I. Introduction

Computer vision tackles problems related with object detection and recognition, texture classification, action recognition, segmentation, tracking, data retrieval, image alignment, just to mention a few. In general, computer vision solutions are based on representing the given image using some global or local image properties, and then comparing them using some similarity measure [1]. Recently, deep learning based approaches are obtaining remarkable results in computer vision applications, as well as in a large number of fields. For instance, learning visual similarities has been recently presented with success working on images in the mono-spectral or in cross-spectral domain [2], [3].

Geographic Information Systems (GIS) and remote-sensing technology allow following of changes to the earths surface on larger spatial and temporal scales than are possible through ground census techniques. Remotely sensed data are an interpretation of various spectral signals that reach a sensor after interacting with objects on the earths surface. These interpretations can reveal many physical characteristics of objects present in the scene, including surface elevation, temperature, and various aspects of the vegetation and land cover. One of the branches of the environmental resources management is the study of agricultural crops and vegetation cover, which are commonly the principal focus of remote sensing investigations. The obtained information is used for monitoring and evaluating the earths vegetative cover. One of the ways to obtain this kind of information is by means of the usage of vegetation indexes. Vegetation indexes are used to determine the health and strength of vegetation and their definitions involve several factors, like soil reflectance, atmosphere, vegetation density, etc. with the aim to obtain those formulas that get more reliable information about vegetation based on remotely sensed values. The usual form of a vegetation index (VI) is a ratio of reflectance measured in two bands, or their algebraic combination. Spectral ranges (bands) to be used in VIs calculation are selected depending on the spectral properties of plants. In the area of applications and research in satellite remote sensing, over forty different vegetation indexes have been developed during the last two decades, like RVI, NRVI, TVI, CTVI, etc.; it can be observed that many scientists have developed indexes related to their specific field of research [4]. The most commonly used index is the Normalized Difference Vegetation Index (NDVI), proposed by Rouse et al. [5]; in general, it is used to determine the condition, developmental stages and biomass of cultivated plants and to forecasts their yields. The scale of this index goes from -1 to 1, with the value zero representing the approximate where the absence of vegetation begins. Negative values represent non-vegetated surfaces. According to [5] 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.

978-1-5386-3894-1/17/$31.00 ©2017 IEEE

The NIR spectral band is the closest in wavelength to the radiation detectable by the human eye; hence, NIR images share several properties with visible images. The interest of using NIR images is related with their capability to segment images according to the object's material. Surface reflection in the NIR spectral band is material dependent, for instance, most pigments used for material colorization are somewhat transparent to NIR. This means that the difference in the NIR intensities is not only due to the particular color of the material, but also to the absorption and reflectance of dyes.

In this context, the current paper tackles the NVDI vegetation index generation automatically from just a near infrared (NIR) image after a learning process using Conditional Generative Adversarial Network (CGAN). Different applications could take advantage of this contribution—another type of vegetation index can be inferred from any spectral band using the same learning process.

Based on the usage of Conditional Generative Adversarial Networks (CGANs), we propose to use the architecture presented in [6], but including a triplet learning model and a conditional NIR image at the final layers of the learning model to improve the details of the generated NDVI vegetation index. The rest of the paper is organized as follows. Section II describes the most recent work on deep learning based remote sensing for vegetation index estimation. Section III presents the adapted Conditional Generative Adversarial Network architecture proposed in this work, detailing the design and training with cross-spectral datasets. Section IV depicts the experimental results and finally, conclusion are presented in section V.

II. Related Work

Recently, the improvements in the acquisition process that allow to obtain images of higher resolution along with the computational intelligence using models of deep learning has been able to enhance the intelligent interpretation of the images coming from huge datasets. In [7] a semantic segmentation algorithm to process earth observation data using multi-modal and multi-scale deep networks has been presented. The approach is able to generate dense scene labels, using an encoder-decoder architecture. A similar approach has been proposed to robust semantic scene understanding of unstructured environments to support robots operating in the real world [8]. Several inherent natural factors such as shadows, glare, vegetation and snow make the semantic scene understanding problem highly challenging. In [8] the usage of multi-spectral and multimodal images helps to increase the robustness of segmentation in real-world outdoor environments. The authors introduce early and late fusion architectures for dense pixel-wise segmentation from RGB, Near-Infrared (NIR) channels, and depth data. There is another paper that addresses the problem of automated detection of harmful algal blooms (HABs) via analysis of image data of inland water bodies. These image data are acquired using a variety of smartphones and communicated via popular OSM platforms such as Facebook, Twitter and Instagram accounts for the wide variations in imaging parameters and ambient environmental parameters, see [9]. In this applications a deep learning approach is used to extract image features and classify them for the purpose of HAB detection. Another approach has been presented by [10]; this work exploits a pipeline that includes two different Convolutional Neural Networks (CNNs). These CNNs are applied to the input RGB+NIR images in order to extract the pixels that represent projections of 3D points that belong to green vegetation. Then, a deeper CNN is then used to classify the extracted pixels between the crop and weed classes. The important contribution of this work is the novel unsupervised dataset summarization algorithm that automatically selects from a large dataset the most informative subsets that better describe the original one. This fact permits to streamline and speed-up the manual dataset labeling process, against an extremely time consuming, while preserving good classification performances.

Recently, Generative Adversarial Network based learning techniques have been used obtaining appealing results; actually, in most of the cases they are among the best options, (e.g., see [11]. This (GAN) networks are becoming the dominant tool to tackle most of computer vision problems. GANs are powerful and flexible tools, one of their most common applications is image generation. In the GAN framework [12], 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 [13], to learn the generators distribution p_g over data $\boldsymbol{x}$, the generator builds a mapping function from a prior noise distribution ${}_pz(z)$ 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 $V(G,D)$:

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

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:

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

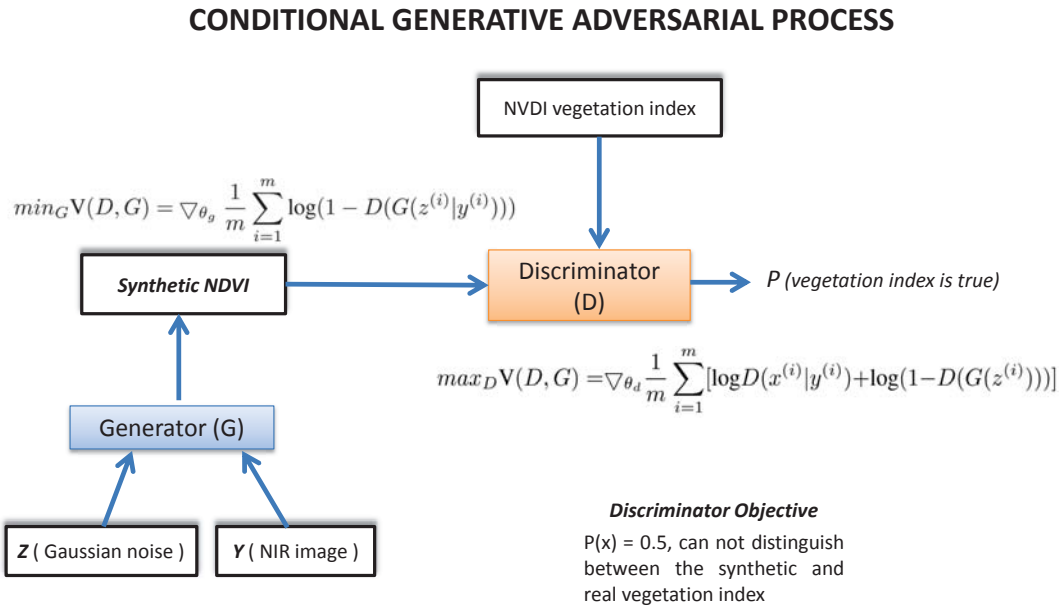

Fig. 1. Conditional Generative Adversarial process implemented on the current work to estimate NDVI Vegetation Index.

In the current work a novel Conditional GAN model [14] is proposed for vegetation index estimation; it is inspired on both the GAN's network architecture presented in [15] for NIR colorization and on the triplet model proposed by [6] for learning color channels from NIR images. Actually, it is an adaptation of the architectures mentioned above, which consists of reducing the number of layers, and removing the internal noise layer, preserving the use of the NIR image added at the final learning process to improve the image details of the generated vegetation index.

III. Proposed Approach

This section presents the approach proposed for NDVI index vegetation estimation. As mentioned above, it uses a similar architecture like the one proposed on a recent works for NIR colorization [14], where the usage of a conditional adversarial generative learning network has been proposed. A traditional scheme of layers in a deep network is used. In the current work the usage of a Conditional GAN model is evaluated in three different schemes: Flat, Siamese and Triplet. These models have presented good performance to solve problems like colorization, segmentation, classification, similarity learning, object recognition, etc. Based on the results that have been obtained on this type of solutions, where improvements in accuracy and performance have been obtained, we propose the usage of a learning model that allows the mapping representation of a vegetation index based on cross-spectral images. Therefore, the model will receive as input a near infrared patch (NIR), with a Gaussian noise added in each element of the learning model to generate the necessary variability of the vegetation index patches, to be able to generalize the learning process. A $l1$ regularization term has been added on a single layer in order to prevent the coefficients to fit so perfectly to overfit, which can improve the generalization capability of the model. Figure 1 depicts the Conditional GAN model proposed in the current work.

A Conditional triplet GAN network based architecture is selected due to several reasons: i) the learning process is conditioned on NIR images from the source domain; ii) its fast generalization capability; iii) the capacity of the generator model to easily serve as a density model of the training data; and $iiii$) sampling is quickly and efficient. The network is intended to learn to generate new samples from an unknown probability distribution. As mentioned above, in our case, the generator network has been implemented in three different schemes: Flat, Siamese and Triplet, which are evaluated in the experimental result section. Figure 2 presents an illustration of the GAN network with the three generator schemes. 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 one that used as 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 near infrared 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.

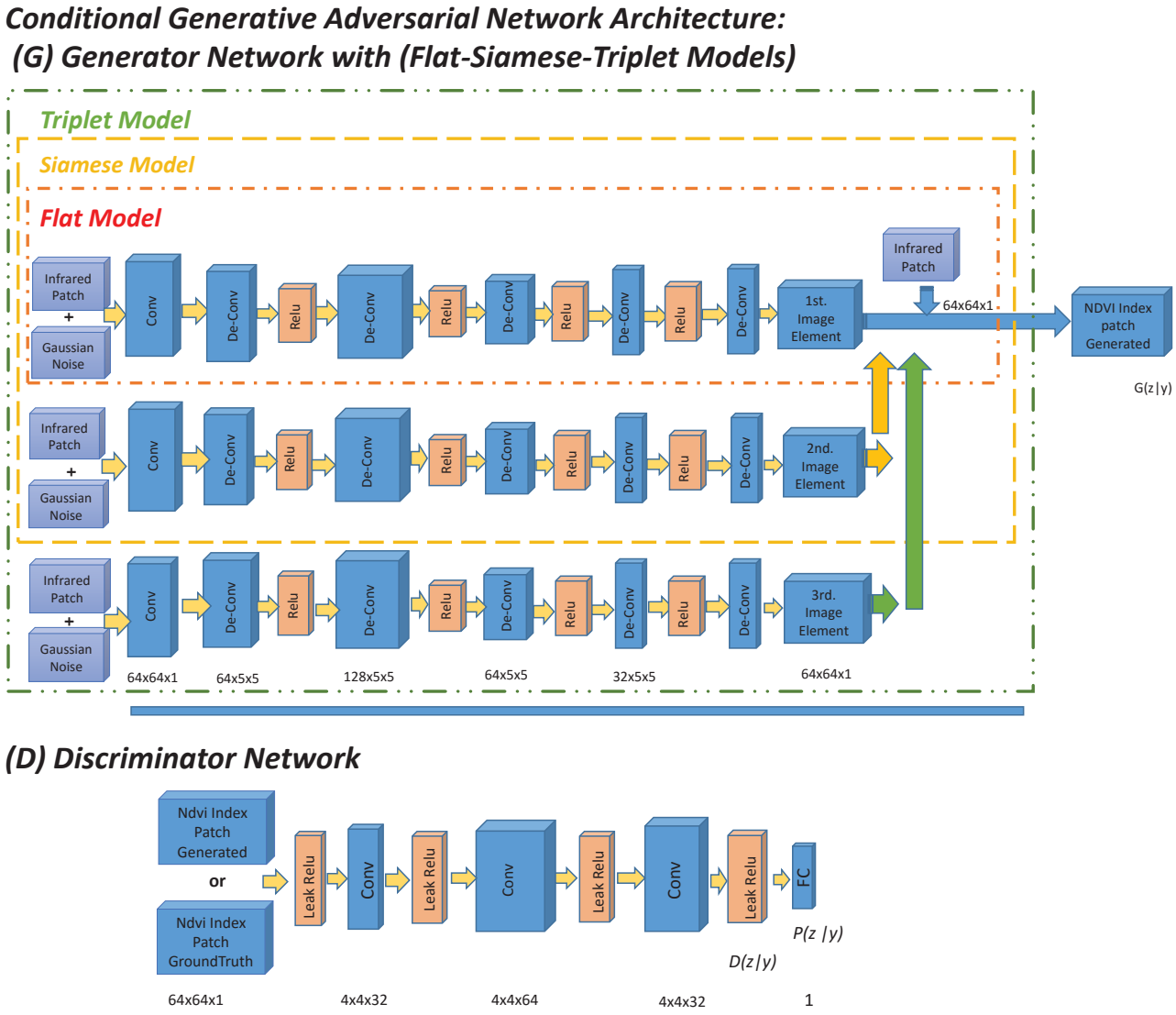

Fig. 2. GAN architecture for NDVI Vegetation Index estimation; on top the three models (Flat, Siamese and Triple) evaluated as Generator Networks; on bottom the Discriminator Network.

IV. Experimental Results

The proposed approach has been evaluated using NIR images and their corresponding NVDI vegetation index, obtained from the equation presented above, in which the RGB was used; this cross-spectral data set came from [16]. The *country* and *flied* 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* contains 51 pairs of images of (1024×680 pixels). In order to train our network to generate vegetation index from each of these categories 280.000 pairs of patches of (64×64 pixels) have been cropped both, in the NIR images as well as in the corresponding NVDI images. Additionally, 2800 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.

The three Conditional Generative Adversarial networks evaluated in the current work (Generator: Flat, Siamese and Triplet) for NDVI vegetation index estimation have been trained using a 3.2 eight core processor with 16GB of memory with a NVIDIA GeForce GTX970 GPU. Qualitative results are presented in Fig. 4 and Fig. 5. Figure 4 shows NDVI vegetation index images from the *country* category generated with the flat, siamese and triplet proposed GAN network. Additionally, Fig. 5 shows NDVI vegetation index images from the *field* category generated with the flat, siamese and triplet proposed GAN network. Quantitative evaluations for the different architectures have been obtained and provided below. Up to our humble knowledge there are not previous work on

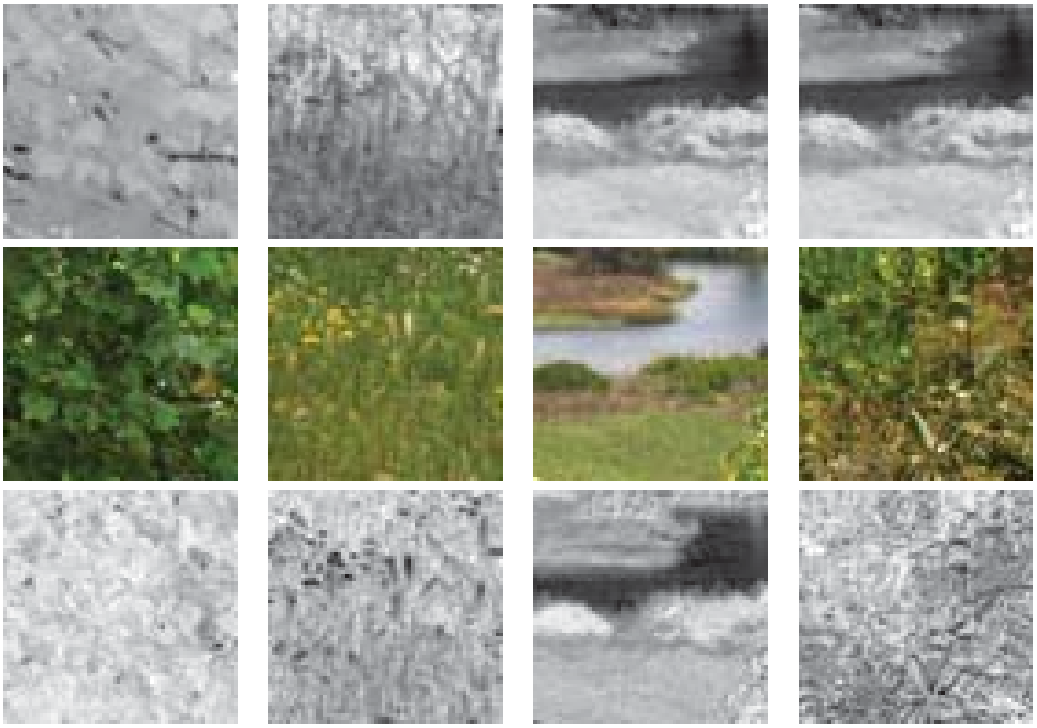

Fig. 3. Pairs of images (1024×680 pixels) from [16]; *country* category (the two-left columns) and *field* category (the two-right columns): (top) NIR images; ($middle$) RBG images; ($bottom$) NDVI vegetation index computed from the NIR and RGB images.

TABLE I
ROOT MEAN SQUARED ERRORS AND STANDARD DEVIATION OBTAINED WITH THE FLAT, SIAMESE AND TRIPLET CONDITIONAL GAN ARCHITECTURES.

Training	RMSE		$STD\sigma_s$	
	country	*field*	*country*	*field*
Flat Network	19.35	20.71	0.68	0.74
Siamese Network	17.47	18.04	0.58	0.62
Triplet Network	12.33	12.65	0.31	0.39

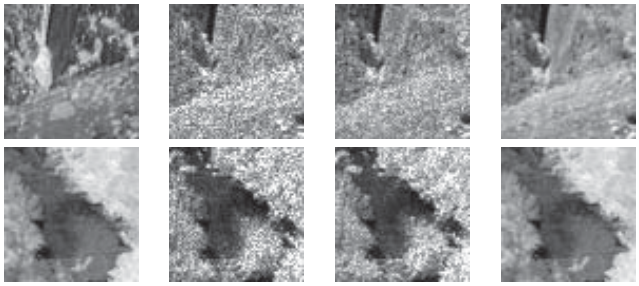

Fig. 4. ($1st.Col$) NDVI index as ground truth images from the ***Country category***. ($2nd.Col$) NDVI index results from the Flat GAN Network ($3rd.Col$) NDVI index obtained with the Siamese GAN network) ($4th.Col$) NDVI index obtained with the Triplet GAN network.

similar technique to estimate vegetation index using only a single spectral band (NIR in our case). Hence, the only way to evaluate results is by comparing the Root Mean Square Error (RMSE) of each approach. The RMSE measures the similarity between the estimated NDVI with respect to the ground truth, which is the standard deviation of the residuals. Residuals are a measure of how distant are the images compared from each other.

Estimated NDVI vegetation index are referred to as ($NDVI_{EST_{i,j}}$) while the corresponding ground truth NDVI vegetation index, numerically computed from the given data sets, are referred to as ($NDVI_{GT_{i,j}}$). The quantitative evaluation consists of measuring at every image, the root mean square error between the estimated value and the corresponding ground truth. Additionally, the standard deviation value is computed to verify the dispersion of the errors previously computed, and to determine whether there are or not big outliers in the experimental results.

Table I presents the average root mean square errors (RSME) and the standard deviation obtained with the three architectures evaluated in the proposed work for the two categories. It can be appreciated that the Triplet model reaches the best result; both the RMSE and the standard deviation of the Triplet GAN are better that those obtained with Flat o Siamese approaches previously explained. The obtained results show that as much levels as better results, since the network will be more capable to learn complex scenes.

V. CONCLUSION

This paper tackles the challenging problem of NDVI vegetation index estimation by using a novel Conditional Generative

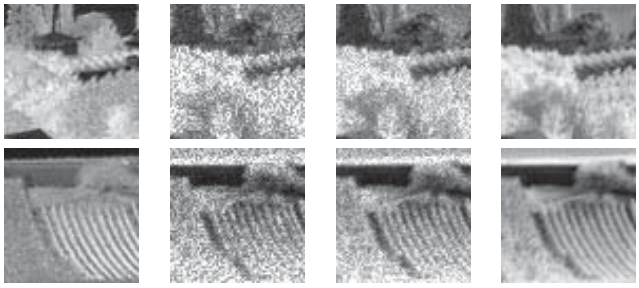

Fig. 5. ($1st.Col$) NDVI index as Ground truth images from the ***Field category***. ($2nd.Col$) NDVI index results from the Flat GAN Network($3rd.Col$) NDVI index obtained with the Siamese GAN network) ($4th.Col$) NDVI index obtained with the Triplet GAN network.

Adversarial Network model. The novelty of the proposed approach lies on the usage of just a single spectral band (NIR in our case). In the work three different schemes are evaluated (Flat, Siamese and Triplet GAN) Results have shown that in most of the cases the network is able to obtain a reliable Normalized Difference Vegetation Index representation from the given NIR image. This technique is so novel, that comparisons with a previous approach are not possible; so that different variants of generative adversarial convolutional networks had to be used to verify that the best results were obtained with the Triplet GAN network version conditioned to the NIR image. Future work will be focused on evaluating others network architectures, like variational auto-encoders, recurrent networks, cycle-consistent adversarial networks, which have shown appealing results in recent works. Additionally, this technique could be used to generate for any other type of vegetation index that are so necessary at present to improve the yield of the agricultural products controlling the biomass of the vegetables species.

ACKNOWLEDGMENT

This work has been partially supported by the ESPOL under projects PRAIM and KISHWAR; by the Spanish Government under Project TIN2014-56919-C3-2-R; and by the "CERCA Programme / Generalitat de Catalunya".

REFERENCES

[1] P. Ricaurte, C. Chilán, C. A. Aguilera-Carrasco, B. X. Vintimilla, and A. D. Sappa, "Feature point descriptors: Infrared and visible spectra," *Sensors*, vol. 14, no. 2, pp. 3690–3701, 2014.

[2] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4353–4361.

[3] C. A. Aguilera, F. J. Aguilera, A. D. Sappa, C. Aguilera, and R. Toledo, "Learning cross-spectral similarity measures with deep convolutional neural networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. IEEE, Jun 2016, p. 9.

[4] A. Bannari, D. Morin, F. Bonn, and A. Huete, "A review of vegetation indices," *Remote sensing reviews*, vol. 13, no. 1-2, pp. 95–120, 1995.

[5] J. Rouse Jr, R. Haas, J. Schell, and D. Deering, "Monitoring vegetation systems in the great plains with erts," 1974.

[6] P. L. Suarez, A. D. Sappa, and B. X. Vintimilla, "Learning to colorize infrared images," in *15th International Conference on Practical Applications of Agents and Multi-Agent Systems*, 2017.

[7] N. Audebert, B. L. Saux, and S. Lefèvre, "Semantic segmentation of earth observation data using multimodal and multi-scale deep networks," *arXiv preprint arXiv:1609.06846*, 2016.

[8] A. Valada, G. Oliveira, T. Brox, and W. Burgard, "Towards robust semantic segmentation using deep fusion," in *Robotics: Science and Systems (RSS 2016) Workshop, Are the Sceptics Right? Limits and Potentials of Deep Learning in Robotics*, 2016.

[9] A. C. Kumar and S. M. Bhandarkar, "A deep learning paradigm for detection of harmful algal blooms," in *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*. IEEE, 2017, pp. 743–751.

[10] C. Potena, D. Nardi, and A. Pretto, "Fast and accurate crop and weed identification with summarized train sets for precision agriculture," in *International Conference on Intelligent Autonomous Systems*. Springer, 2016, pp. 105–121.

[11] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," *arXiv preprint arXiv:1701.04862*, 2017.

[12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

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

[14] P. L. Suarez, A. D. Sappa, and B. X. Vintimilla, "Colorizing infrared images through a triplet conditional dcgan architecture," in *19th International Conference on Imaage Analysis and processing*, 2017.

[15] ——, "Infrared image colorization based on a triplet dcgan architecture," in *Computer Vision and Pattern Recognition*, 2017.

[16] M. Brown and S. Süsstrunk, "Multi-spectral SIFT for scene category recognition," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 177–184.