# Cross-spectral image dehaze through a dense stacked conditional GAN based approach

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 remove haze from RGB images using a near infrared images based on a dense stacked conditional Generative Adversarial Network (CGAN). The architecture of the deep network implemented receives, besides the images with haze, its corresponding image in the near infrared spectrum, which serve to accelerate the learning process of the details of the characteristics of the images. The model uses a triplet layer that allows the independence learning of each channel of the visible spectrum image to remove the haze on each color channel separately. A multiple loss function scheme is proposed, which ensures balanced learning between the colors and the structure of the images. Experimental results have shown that the proposed method effectively removes the haze from the images. Additionally, the proposed approach is compared with a state of the art approach showing better results.**

***Index Terms*—Infrared imaging, Dense, Stacked CGAN, Cross-spectral, Convolutional networks**

## I. Introduction

The images can be seriously affected by different causes, one of the most common are the natural phenomena that occur, such as fog, dust, rain, snow, etc. This considerably reduces the visibility of the objects in the images, thus affecting the understanding of the same. Therefore, processes such as the detection, segmentation or the recognition of objects, among others, will not be able to obtain results that meet the required objectives. Outdoor scenes usually suffer mainly from low contrast and poor visibility due to the adverse weather conditions that cause airborne particles to scatter the light present in the atmosphere. One of the atmospheric effects that occur is the mist, which is independent of the brightness of the scene and generates attenuation effects. It is affected by ambient light at the time of image acquisition. It is necessary to consider that at a greater distance from the focus of the more diffuse camera the image becomes.

Improving the quality of images has been one of the problems that computer vision has sought to solve, several approaches have been proposed, especially aimed at removing climatic effects such as haze; some traditional techniques were focused for the elimination of the haze presented on images using the characteristics present on them. In [1] a method based on generic regularity in natural images is presented where the pixels of small image patches usually exhibit a 1D distribution in the RGB color space, known as color lines. This method derives a local training model that explains the color lines in the context of fuzzy scenes and uses it to recover scene transmission based on the displacement of the lines from the origin. In [2] a novel system is proposed to explore, improve and manipulate casual outdoor photographs, combining them with georeferenced digital terrain and existing urban models. A simple interactive registration process is used to align a photograph with that model. These methods generally involve multi-step approaches that use depth information to eliminate these degradation effects. Most methods to eliminate haze in the images only consider the use of assumptions of hard thresholds or user input to estimate atmospheric light.

In recent years, deep learning through convolutional neural networks has been widely used in a wide range of fields. In deep learning, these networks are found to give the most accurate results in solving real-world problems. Among the different network architectures, Generative Adversarial Networks (GANs) have obtained excellent results to solve problems such as [3] colorization, face generation, cross-spectral similarity [4], single image dehazing [5] or NVDI vegetation index generation [6]. Some of these approaches have used NIR images to improve the results obtained by the networks. In a previous work [5] we propose to remove the haze from a image implementing a CGAN network working with only RGB images. In the current work we tackles the dehaze problem proposing the usage of cross-spectral images (RGB+NIR) to enhance the removal process implemented by our stacked dense CGAN network.

The NIR spectrum is independent of the brightness and color of the targets, including non-visible illumination requirements. Images from the NIR spectral band have surface reflection which is material dependent. 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 colors. Nowadays, although the additional information that can be obtained from the images of the infrared spectrum, the people interested in the visual analysis of the information prefer that images are able to be perceptible to the human eye, this means that they are in RGB representation, because it is easier for people to understand the scene to be analyzed due to the familiarity of the shapes and colors of the objects, facilitating decision making.

In our approach, each channel is mapped in a three-dimensional space, using a stacked dense CGAN model to accelerate convergence, we propose a dense model to to improve the accuracy and efficiency of training.The manuscript is organized as follows. Section II presents works related to the haze removal problem, as well as the basic concepts and notation of GAN networks. The proposed approach is detailed in section III. The experimental results with a set of real images are presented in section IV. Finally, the conclusions are given in section V.

## II. Related Work

Different image haze removal techniques have been proposed in last decade, some of them based on image attributes, transmission map, air light conditions, atmospheric scattering model, etc. (e.g., [7]). In [8] the authors proposed an enhanced detail and dehaze technique for haze removal based on modified channel prior scheme and combine the dehazed image with a non-sky detail layer using enhanced method in order to improve the image details. After that, the recovered image contrast will be enhanced based on a histogram equalization approach. Also using a haze model, [9] proposes an improved contrast enhanced restoration. This technique is based on a quadtree subdivision searching method, which the sky area of multi-channel polarization image are extracted automatically, and the atmospheric light and degree of polarization is calculated; then the scene depth information of image is calculated based on contrast enhancement method. Finally, the atmospheric intensity is thinly restored by guided filter, and the degraded image is restored. In [10] a haze removal technique that uses a fusion-based variational method is presented, which combine the minimized outputs of two energy functionals to produce a haze-free version.

The authors in [11] present a detailed survey and experimental analysis on DCP-based methods will explain the effectiveness of the individual step of the dehazing process and will facilitate development of advanced dehazing algorithms. Another model based approach has been presented in [12], which proposes an algorithm based on an image filtering, dark channel prior, estimations of atmospheric light to obtain an unhazed image and finally improving the local contrast. Another method proposed in [13], consists of a combined algorithm based on both dark channel prior and histogram optimization, which can make the image contrast stretching, so the impact of the haze on the image can be weakened. If the obtained dehazing image cannot meet the minimum quality required, the dark channel prior can be used to estimate the haze intensity.

Lately, novel image haze removal approaches based on deep learning techniques have been proposed obtaining acceptable results. In [14] a model based on a reformulated atmospheric scattering model is proposed, instead of estimating the transmission matrix and the atmospheric light separately. Cai et.al. [15] propose a trainable end-to-end system called DehazeNet, for medium transmission estimation. DehazeNet takes a hazy image as input, and outputs its medium transmission map that is subsequently used to recover a haze-free image via atmospheric scattering model. More recently the Generative Adversarial Network (GAN) framework has been used obtaining appealing results. In [16] the authors propose a unified single removal haze GAN network that jointly estimates the transmission map and performs the haze process; the network is trained using synthetic images and a two-term loss function. Additionally, in the GAN architecture a stacking strategy is proposed to speed up the learning process. Furthermore, the proposed network is trained using real images. Also in [17] the authors proposes a method for combining dark channel prior (DCP) and bright channel prior (BCP) for single image dehazing. The proposed technique achieves airlight approximations by implementing numerical proximity of atmospheric light, which use the average value of the DCP and BCP. A previous work presented in [5] proposed an novel stacked conditional GAN to removal the haze on RGB image, and also proposes a multiple loss to accelerate the learning process at training time, this approach have obtained good results. Based on this work, our current work proposed a cross-spectral dense CGAN to achieve better accuracy and reduce the training time.

Generative based deep learning models can be used to take a collection of points and infer a function that describes the distribution that generated them. The generator model after the training could create samples of the distribution that you just learned, this allow the network to learn to generate data with the same internal structure as other data. It is a framework presented on [18] for estimating generative models via an adversarial process, in which simultaneously two models are trained: a generative model $G$ that captures the data distribution, and 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. This framework corresponds to a minimax two-player game. In the space of arbitrary functions $G$ and $D$, a unique solution exists, with $G$ recovering the training data distribution and $D$ equal to 1/2 everywhere. According to [19], to learn the generators distribution $p_g$ over data $\boldsymbol{x}$, the generator builds a mapping function from a prior noise distribution $p_z$ to a data space $G(z;\theta_g)$. 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 $V(G,D)$:

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

GANs networks can be extended to a conditional model if both the generator and discriminator are conditioned on some extra information $y$. 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:

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

Recently applications of GANs have shown that they can produce excellent samples. According to [20], training GANs networks requires finding a Nash equilibrium of a non-convex game with continuous, highdimensional parameters. GANs are typically trained using gradient descent techniques that are designed to find a low value of a cost function, rather than to find the Nash equilibrium of a game. When used to seek for a Nash equilibrium, these algorithms may fail to converge. In that work, they introduce several techniques intended to encourage convergence of the GANs game, motivated by a heuristic understanding of the non-convergence problem. They lead to improved semi-supervised learning peformance and improved sample generation.

Considering the use of conditional generative networks models, this work propose the usage of an architecture similar to the one presented in [5], but by including a densely connected layers architecture according to [21] to perform shorter connections between the layers to merge the features and make it more deeper, accurate, and efficient to train. Also the model include a stacked network inspired on the work presented in [22], which consists in a top-down stack of GANs, each designed to generate lower-level representations conditioned on higher level representations. In the current work we propose a dense stacked conditional learning process of the generator-discriminator to accelerate the convergence of the network, this stacking strategy allows accelerating the learning process to generate a clear image representation from those affected by haze. The current work also proposes include the corresponding NIR image to increase the effectiveness of the haze removal process, and a multiple loss term for discriminator, which makes the learning process continuous and differentiable and consequently the times of convergence for the generalization of learning are improved.

## III. Proposed Approach

The proposed approach is based on a cross-spectral generative adversarial dense network stacked at several levels to accelerate the training process. This model, unlike the one proposed in [5], receives as an additional input the image with haze, its corresponding in the near infrared spectrum with the objective of obtaining an image with greater clarity in the details of the images. The GANs generate the image without haze starting from the images of both RGB and NIR concatenated spectra, this architecture with the stacked scheme uses a multiple loss to learn more efficiently and to improve the convergence of the model, which allows to accelerate the obtained diversity and to generalize the learning model. A $l1$ regularization term has been added at every layer of the generator network in order to prevent the coefficients to fit so perfectly to overfit and to introduce more robustness to the generalization of the model; additionally, it helps reducing the time to reach a well trained network. $l1$ helps perform feature selection in sparse feature spaces, this help to know which features are helpful and which are redundant.

The network is designed to learn how to generate new images without haze from an conditional latent distribution. In our case, the generator network has been modified to use feature hierarchical representation; we use three levels of dense stacking conditional learning process. Additionally, the model has been designed to receive cross-spectral concatenated images as an input and use a multiple loss function. In order to optimize the model generalization, the GAN framework is reformulated for a conditional generative image modeling tuple. In other words, the generative model $G(z;\theta_g)$ is trained from a haze and an infrared concatenated image and contrary to the original GAN model formulation, the random noise $z$ is not used; with the assumption that the randomness has already been preserved by the conditioning variables provided by the images with haze, in order to produce a clear RGB image. The discriminative model $D(z;\theta_d)$ is trained to assign the correct label to the generated clear RGB image, according to the provided original color image, which is used as a ground truth. Variables $(\theta_g)$ and $(\theta_d)$ represent the weighting values for the generative and discriminative networks.

Our proposed approach introduce a dense connection between layers on the architecture, according to [21] we propose a network that implements shorter connections generally at the beginning and the end of the learning layers in the model, this give to the network the capacity to train more rapidly using less layers. Use dense connections have several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters. Applying this kind of models of connectivity between the layers achieved as a direct consequence of the input concatenation of RGB and NIR image at any level of the learning layers, permits that all the feature maps learned by any of the dense net layers can be accessed by all subsequent layers. This encourages feature reuse throughout the network, and leads to more compact models.

In addition, similar to [5], multiple loss functions ($\mathcal{L}$) have been implemented, which was conformed by the combination of the adversarial loss plus the intensity loss (MSE), the structural loss (SSIM) and the image quality loss (IQ). This combined loss function has been defined to avoid the usage of only a pixel-wise loss 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 below.

The **adversarial loss** is designed to minimize the cross-entropy to improve the texture loss :

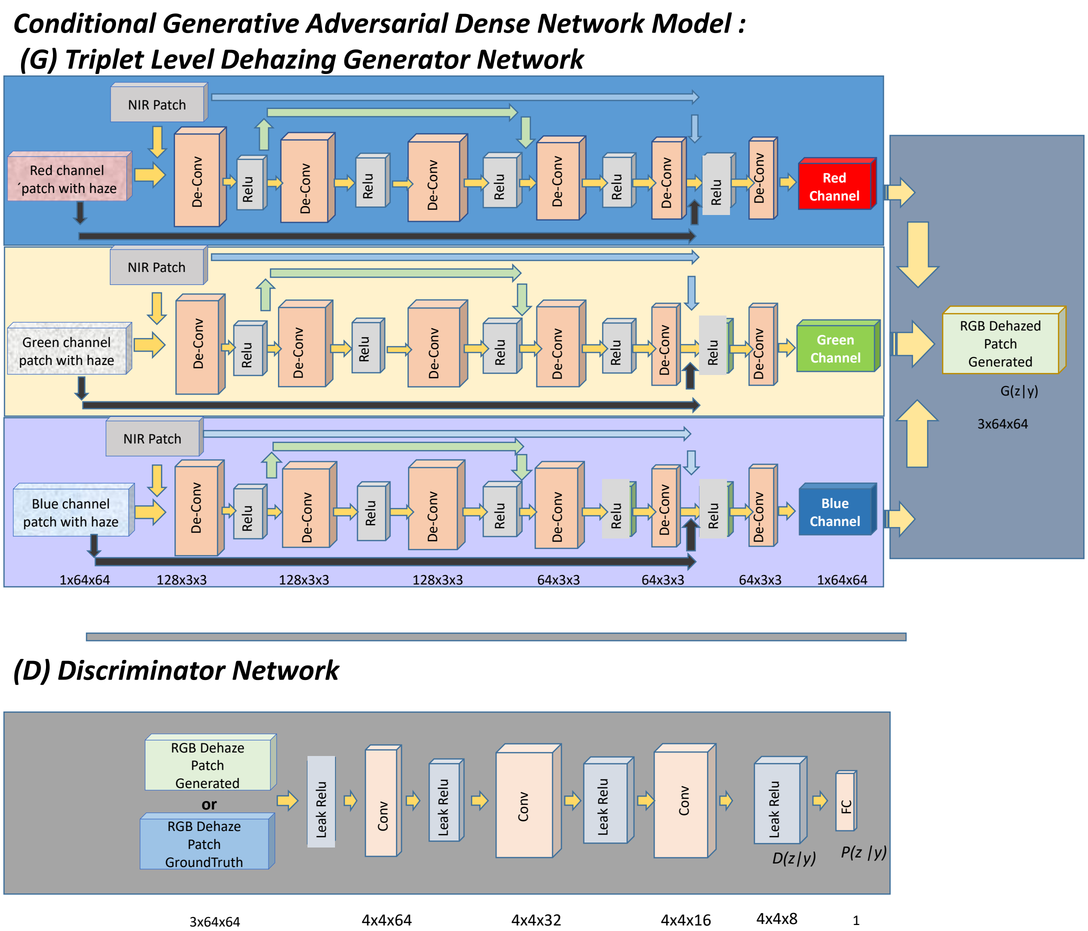


Fig. 1. Illustration of the proposed triplet cross-spectral dense CGAN architecture used for image dehazing.

$$\mathcal{L}_{Adversarial} = -\sum_{i} logD(G_w(I_{z|y}), (I_{x|y}), \tag{3}$$

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 haze and near infrared image feeded in each channel of the Stacked Gan Network.

The **intensity loss** is defined as:

$$\mathcal{L}_{Intensity} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}(RGBe_{i,j} - RGBg_{i,j})^2, \tag{4}$$

where $RGBe_{i,j}$ is the estimated RGB representation and $RGBg_{i,j}$ is the ground-truth RGB image. This loss measures the difference in intensity of the pixels between the images without considering texture and content comparisons. 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) [23], which evaluates images accounting for the fact that the human visual perception system is sensitive to changes in local structure; the purpose of using this index defines the structural information in an image as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast. The **structural loss** for a pixel $p$ is defined as:

$$\mathcal{L}_{SSIM} = \frac{1}{NM}\sum_{p=1}^{P} 1 - SSIM(p), \tag{5}$$

where $SSIM(p)$ is the Structural Similarity Index (see [23] for more details) centered in pixel $p$ of the patch $P$.

Another loss function that proposes this work is based on the universal image quality index, the method proposed by [24] was designed to model any image distortion via a combination of three factors: loss of correlation, luminance distortion, and contrast distortion.

The main reason to use this quality index as a loss function is its strong ability to measure the structural distortions existing in the images with haze. It is important to bear in mind that because the signals of the images are non-stationary it is preferable to evaluate the quality of the images by measuring their statistical characteristics in a local way and then combine them all together in a single measurement of image quality. If there are a total of M steps, at the $j$-th step the local quality index $Q_j$ is computed, then the overall quality index is given by :

$$Q = \frac{1}{M}\sum_{j=1}^{M} Q_j, \tag{6}$$

Hence, we can formulate the **quality loss** function as:

$$\mathcal{L}_Q = \frac{1}{M}\sum_{j=1}^{M}(1 - Q_j). \tag{7}$$

The **final loss function** ($\mathcal{L}$) used in this work is the accumulative weighted sum of the individual adversarial, intensity, structural and quality loss functions:

$$\begin{aligned}\mathcal{L}_{final} = 0.40\mathcal{L}_{Adversarial} + 0.25\mathcal{L}_{Intensity} + \\ +0.20\mathcal{L}_{SSIM} + 0.15\mathcal{L}_{Q}.\end{aligned} \tag{8}$$

The proportion assigned to each loss has been defined based on the variability of the values obtained by each of the losses during the training process.

The dense cross-spectral stacked CGAN network proposed has been trained using Stochastic AdamOptimazer since it is well suited for problems that are large in terms of data and/or parameters, very appropriate for non-stationary objectives and for problems with very noisy/or sparse gradients. Also the Hyper-parameters have intuitive interpretation and typically require less tuning, prevents overfitting and leads to convergence faster. Furthermore, it is computationally efficient, has little memory requirements, is invariant to diagonal rescaling of the gradients. The image dataset was normalized in a (-1,1) range. The following hyper-parameters were used during the training process: learning rate 0.00004 for the generator and 0.00003 for the discriminator networks respectively; epsilon = 1e-08; exponential decay rate for the 1st moment momentum 0.4 for discriminator and 0.3 for the generator; weight initializer with a standard deviation of 0.04582; $l1$ weight regularizer; weight decay 1e-2; leak relu 0.21 and patch's size of 64×64.

The triplet architecture, see Fig. 1, maintains similar structure presented in [5]. Basically in the architecture a layer of learning was suppressed, as well as the depth of the learning layers was decreased because of the concatenation of the NIR with the haze image. The learning architecture is conformed by convolutional, de-convolutional, relu, leak-relu, fully connected and activation function tanh and sigmoid for generator and discriminator networks respectively. Additionally, every layer of the model uses batch normalization for training any type of mapping to prevent underfitting. It is very important to maintain the spatial information in the generator model, there is not pooling and drop-out layers and only the stride of 1 is used to avoid downsize the image shape. To prevent overfitting we have added a $l1$ regularization term ($\lambda$) in the generator model, this regularization has the particularity that the weights matrix end up using only a small subset of their most important inputs and become quite resistant to noise in the inputs. Additionally the architecture includes dense model implemented by at the first and bottom layers in the model to increase the generalization and obtain more optimization of the learning process.

The generator ($G$) and discriminator ($D$) are both feedforward deep neural networks that play a min-max game between one another. The generator takes as input on each channel the hazy and NIR image and it is transformed into the form of the data we are interested in imitating, in our case a RGB clear image. The discriminator takes as an input a set of data, either real image ($z$) or generated image ($G(z)$), and produces a probability of that data being real ($P(z)$). The discriminator is optimized in order to increase the likelihood of giving a high probability to the real data (the ground truth given image) and a low probability to the fake generated data (wrongly clarified haze image), as introduced in [19]; thus, the dense conditional discriminator network is updated as follow:

$$\bigtriangledown_{\theta_d}\frac{1}{m}\sum_{i=1}^{m}[\log D(x^{(i)}) + \log(1 - D(G(y^{(i)}, z^{(i)})))], \tag{9}$$

where $m$ is the number of patches in each batch, $x$ is the ground truth image, $y$ is the image without haze (RGB) generated by the network and $z$ is the random Gaussian sampled noise. The weights of the discriminator network ($D$) are updated by ascending its stochastic gradient. On the other hand, the generator is then optimized in order to increase the probability of the generated data being highly rated, it is updated as follow:

$$\bigtriangledown_{\theta_g}\frac{1}{m}\sum_{i=1}^{m}\log(1 - D(G(y^{(i)}, z^{(i)}))), \tag{10}$$

where $m$ is the number of samples in each batch, $y$ is the image without haze (RGB) generated by the network and $z$ is the random Gaussian sampled noise. Like in the previous case, the weights of the generator network ($G$) are updated by descending its stochastic gradient.

## IV. Experiments Results

### A. Results and comparisons

The proposed architecture has been evaluated using real hazed images and their corresponding clear RGB and NIR infrared representations obtained from [25]. Figure 2 presents a set of images from this dataset. From all these images 85000 pairs of patches of (32×32 pixels) have been cropped both, in the hazed images as well as in the corresponding clear RGB images. Additionally, 8500 pairs of patches have been also generated for validation. On average, every training

(a) (b) (c)

Fig. 2. Set of images from the dataset obtained from [25]: $(a)$ NIR image, $(b)$ hazy image and $(c)$ Groundtruth image.

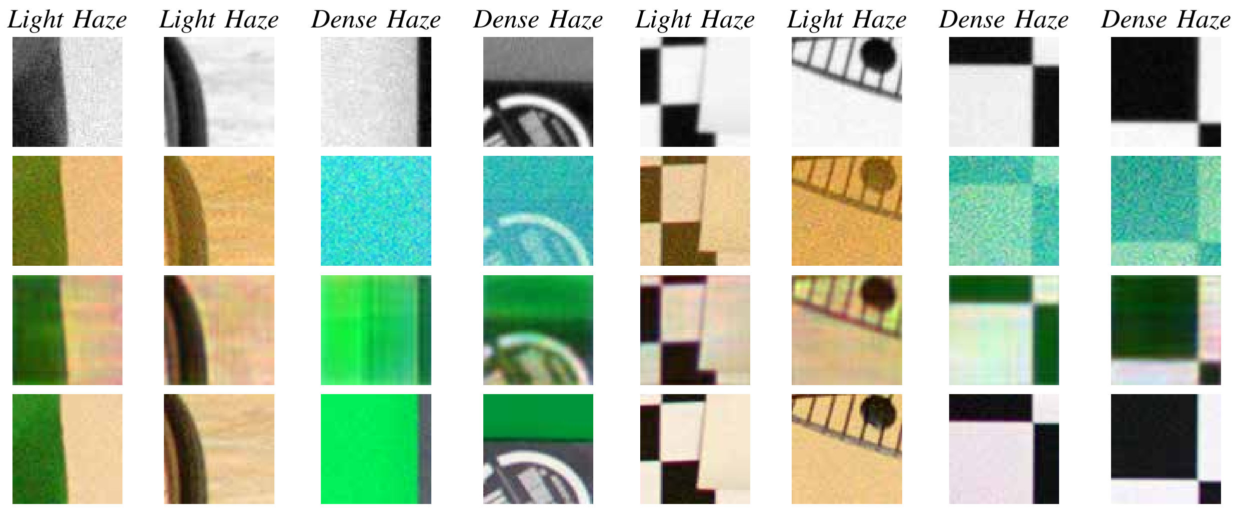


Fig. 3. *(1st row)* NIR patches. *(2nd row)* Haze patches. *(3rd row)* Results from the proposed approach dense CGAN (Loss Function: $\mathcal{L}_{final}$). *(4th row)* Ground truth images.

process took about 60 hours using a 3.2 eight core processor with 16GB of memory with a NVIDIA TITAN V GPU. Some patches, with the corresponding result obtained with the proposed approach are depicted in Fig. 3; just for making easier the evaluation of results from the proposed approach patches have been split up into *Light Haze* and *Dense Haze*.

The quantitative evaluation consists of measuring several metrics with the results obtained with the proposed Stacked GAN approach when different combinations of the proposed loss functions where considered; one of the metrics consists of measuring at every pixel the angular error (AE) between the obtained result ($RGBo_{i,j}$) and the corresponding ground truth value ($RGBg_{i,j}$). AE is included since this measure is quite similar to the human visual perception system, [26]—AE is probably the most widely used performance measure in color constancy research. Additionally, the Mean Squared Error (MSE), the Quality Index (QIndex) and the Structural Similarity (SSIM) metrics are also considered in this quantitative evaluation. On the contrary to AE and MSE, which can be considered as pixel level evaluation metrics, the SSIM and QIndex are methods for evaluating the perceived quality of the results. With the metrics mentioned above combinations of the different loss functions are evaluated, results are provided in Table I. It can be appreciated that in all the cases the results obtained with the final loss proposed with dense Stacked Conditional GAN are better than those obtained with the approach presented in [5]. In addition, these losses, being perfectly differentiable, allow for a better optimization of the network, thus accelerating the convergence process. Just as illustrations, few RGB images from *Light Haze* and *Dense Haze* categories, generated with the proposed Stacked GAN network, are depicted in Fig. 3 for qualitative evaluation.

## V. Conclusion

This paper tackles the challenging problem of generating clear RGB representations from hazed and their corresponding NIR images by using a novel dense stacked cross-spectral conditional generative adversarial network. Results have shown that in most of the cases the network is able to obtain reliable clear RGB representations. As mentioned in the discussion section, this approach has as a limitation the need of having ground truth images without haze for training, as future work, actually, as work in progress we have proposed the usage of a cycle GAN architecture, but feed it with RGB hazed and their corresponding NIR image in the generator to speed up the generalization. Future work will also consider other loss functions to improve the training process.

## Acknowledgment


This work has been partially supported by: the ESPOL project PRAIM (FIEC-09-2015); the Spanish Government

TABLE I

ANGULAR ERRORS (AE), MEAN SQUARED ERRORS (MSE), STRUCTURAL SIMILARITIES (SSIM) AND IMAGE QUALITY INDEX(Q INDEX) OBTAINED WITH THE PROPOSED CROSS-SPECTRAL DENSE STACKED CONDITIONAL GAN ARCHITECTURE BY USING MULTIPLE LOSS FUNCTIONS (SSIM AND Q INDEX VALUES, THE BIGGER THE BETTER) AND THE APPROACH IN [5].

| Training | AE | | MSE | | SSIM | | Q Index | |
|---|---|---|---|---|---|---|---|---|
| | *Light Haze* | *Dense Haze* | *Urban* | *Dense Haze* | *Light Haze* | *Dense Haze* | *Light Haze* | *Dense Haze* |
| *Stacked CGAN presented in [5] with* $\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity}$ | 7.18 | 7.11 | 21.96 | 23.75 | 0.72 | 0.69 | 0.62 | 0.59 |
| *Proposed cross-spectral dense stacked CGAN with* $\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity}$ | 6.46 | 6.11 | 19.86 | 21.74 | 0.80 | 0.73 | 0.68 | 0.64 |
| *Stacked CGAN presented in [5] with* $\mathcal{L}_{Adversarial} + \mathcal{L}_{SSIM}$ | 7.12 | 7.03 | 20.97 | 20.74 | 0.78 | 0.72 | 0.64 | 0.61 |
| *Proposed cross-spectral dense stacked CGAN with* $\mathcal{L}_{Adversarial} + \mathcal{L}_{SSIM}$ | 6.28 | 5.97 | 19.19 | 20.33 | 0.86 | 0.84 | 0.76 | 0.69 |
| *Stacked CGAN presented in [5] with* $\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity} + \mathcal{L}_{SSIM}$ | 6.32 | 6.24 | 19.65 | 20.08 | 0.80 | 0.77 | 0.68 | 0.66 |
| *Proposed cross-spectral dense stacked CGAN with* $\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity} + \mathcal{L}_{SSIM}$ | 6.13 | 5.93 | 18.47 | 19.85 | 0.89 | 0.81 | 0.78 | 0.71 |
| *Stacked CGAN presented in [5] with* $\mathcal{L}_{final}$ | 5.95 | 6.12 | 18.74 | 19.21 | 0.84 | 0.80 | 0.71 | 0.68 |
| *Proposed cross-spectral dense stacked CGAN with* $\mathcal{L}_{final}$ | 5.10 | 5.65 | 17.92 | 18.74 | 0.92 | 0.86 | 0.82 | 0.77 |

under Projects TIN2014-56919-C3-2-R and 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). The authors would also like to thank NVIDIA for GPU donations.

## REFERENCES

[1] S. Shwartz, E. Namer, and Y. Y. Schechner, "Blind haze separation," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2. IEEE, 2006, pp. 1984–1991.

[2] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski, "Deep photo: model-based photograph enhancement and viewing," *ACM transactions on graphics*, vol. 27, no. 5, 2008.

[3] P. L. Suárez, A. D. Sappa, and B. X. Vintimilla, "Infrared image colorization based on a triplet dcgan architecture," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017, pp. 212–217.

[4] ——, "Cross-spectral image patch similarity using convolutional neural network," in *Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM), 2017 IEEE International Workshop of*. IEEE, 2017, pp. 1–5.

[5] P. L. Suárez, A. D. Sappa, B. X. Vintimilla, and R. I. Hammoud, "Deep learning based single image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1169–1176.

[6] P. L. Suarez, A. D. Sappa, and B. X. Vintimilla, "Learning image vegetation index through a conditional generative adversarial network," in *Ecuador Technical Chapters Meeting (ETCM), 2017 IEEE*. IEEE, 2017, pp. 1–6.

[7] R. Richter, "Atmospheric correction of satellite data with haze removal including a haze/clear transition region," *Computers & Geosciences*, vol. 22, no. 6, pp. 675–681, 1996.

[8] D. N. L. Yao, A. Bade, N. M. Suaib, and H. A. B. Sulaiman, "Digital image enhancement using enhanced detail and dehaze technique (dde)," *Advanced Science Letters*, vol. 24, no. 3, pp. 1559–1561, 2018.

[9] X. Li and Q. Huang, "Polarization filtering for automatic image dehazing based on contrast enhancement," in *Communication Software and Networks (ICCSN), 2017 IEEE 9th International Conference on*. IEEE, 2017, pp. 1266–1271.

[10] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío, "Fusion-based variational image dehazing," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 151–155, 2017.

[11] S. Lee, S. Yun, J.-H. Nam, C. S. Won, and S.-W. Jung, "A review on dark channel prior based image dehazing algorithms," *EURASIP Journal on Image and Video Processing*, vol. 2016, no. 1, p. 4, 2016.

[12] H. Mahdi, N. EL ABBADI, and H. RUSTUM, "Single image de-hazing through improved dark channel prior and atmospheric light estimation." *Journal of Theoretical & Applied Information Technology*, vol. 95, no. 15, 2017.

[13] J. Yang, B. Jiang, Z. Lv, and N. Jiang, "A real-time image dehazing method considering dark channel and statistics features," *Journal of Real-Time Image Processing*, vol. 13, no. 3, pp. 479–490, 2017.

[14] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "An all-in-one network for dehazing and beyond," *arXiv preprint arXiv:1707.06543*, 2017.

[15] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[16] H. Zhang, V. Sindagi, and V. M. Patel, "Joint transmission map estimation and dehazing using deep networks," *arXiv preprint arXiv:1708.00581*, 2017.

[17] J. Jackson, O. Ariyo, K. Acheampong, M. Boakye, E. Frimpong, E. Ashalley, and Y. Rao, "Hybrid single image dehazing with bright channel and dark channel priors," in *Image, Vision and Computing (ICIVC), 2017 2nd International Conference on*. IEEE, 2017, pp. 381–385.

[18] 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.

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

[20] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Advances in Neural Information Processing Systems*, 2016, pp. 2226–2234.

[21] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

[22] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie, "Stacked generative adversarial networks," *arXiv preprint arXiv:1612.04357*, 2016.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[24] Z. Wang and C. Alan, "Bovik,," *A Universal Image Quality Index, In IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[25] J. Lüthen, J. Wörmann, M. Kleinsteuber, and J. Steurer, "A rgb/nir data set for evaluating dehazing algorithms," *Electronic Imaging*, vol. 2017, no. 12, pp. 79–87, 2017.

[26] A. Gijsenij, T. Gevers, and M. P. Lucassen, "A perceptual comparison of distance measures for color constancy algorithms," in *European Conference on Computer Vision*. Springer, 2008, pp. 208–221.