Deep Learning based Single Image Dehazing

Patricia L. Suárez[1], Angel D. Sappa[1,2], Boris X. Vintimilla[1] and Riad I. Hammoud[3]

[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

[3]BAE Systems, FAST Labs
600 District Avenue, Burlington, MA 01803, USA

plsuarez@espol.edu.ec, sappa@ieee.org, bvintimi@espol.edu.ec, riad.hammoud@baesystems.com

Abstract

This paper proposes a novel approach to remove haze degradations in RGB images using a stacked conditional Generative Adversarial Network (GAN). It employs a triplet of GAN to remove the haze on each color channel independently. A multiple loss functions scheme, applied over a conditional probabilistic model, is proposed. The proposed GAN architecture learns to remove the haze, using as conditioned entrance, the images with haze from which the clear images will be obtained. Such formulation ensures a fast model training convergence and a homogeneous model generalization. Experiments showed that the proposed method generates high-quality clear images. [1]

1. Introduction

The appearance of outdoor images is easily affected by natural phenomena such as fog, dust, rain, snow, etc. This reduces considerably the visibility of the objects in the images, therefore, processes such as feature detection, segmentation or object recognition, among others, will not be able to obtain results that meet the required objectives. The outdoor scenes mostly suffer from low contrast and poor visibility due to adverse atmospheric conditions allowing the particles in the air to disperse the light present in the atmosphere. One of the atmospheric effects that occur is the haze, which is independent of the brightness of the scene and generates effects of attenuation. It is affected by the ambient light at the moment of the acquisition of the image. It is necessary to consider that at a greater distance from the focus of the camera most diffuse the image becomes.

Different approaches on image quality improvements have been made, including those that specialize in removing fog; some focus on working with depth or with multiple views of the same image as presented by [19]. In [9] a technique to improve casual outdoor photographs by combining them with existing georeferenced digital terrain and urban models is proposed. This approach uses a registration process to align a photograph with that model. These methods typically involve multi-step approaches that use depth information for removal of those degradation effects. Most methods for remove haze on images only consider to use hard threshold assumptions or user input to estimate atmospheric light. Artificial lighting or applied adaptive filters [12] are also considered in some methods to remove the haze in the images. However, error estimation of atmospheric light may affects the results of the remove process.

The effect of haze on image quality is as a result of a random scattering of light and hence affects all pixels of the image. In recent years, deep learning has been extensively used in a wide range of fields. In deep learning, Convolutional Neural Networks are found to give the most accurate results in solving real world problems. Among the different networks architectures, Generative Adversarial Networks (GANs) have obtained outstanding results to solve problems like colorization [22], face generation, super-resolution [10], text-to-image synthesis [15], cross-spectral similarity [21].

In the particular problem tackled in this work for remove the haze to obtain a clear RGB representation the usage of a GAN architecture is proposed. In our approach ev-

[1]Approved for public release; unlimited distribution.

ery channel is mapped into a three dimensional space, using an stacked GAN model to speed up convergence. The manuscript is organized as follows. Section 2 introduces works related with the remove haze problem as well as the basics concepts and notation of GAN networks. The proposed approach is detailed in Section 3. Experimental results with a set of real images are presented in Section 4. Discussions on the usage of NIR images with the proposed GAN architecture are provided in Section 5. Finally, conclusions are given in Section 6.

2. Related Work

Image haze removal has been studied for more than two decades (e.g., [17]). Some of the approaches proposed in the literature start from the usage of images from other spectral band to extract certain characteristics that serve to remove the haze. In [1] a non-local haze-lines for remove haze from image is proposed; this method is based on the observation that the number of distinct colors in an image is orders of magnitude smaller than the number of pixels, based on the assumption that an image can be faithfully represented with just a few hundreds of distinct colors. Another model based approach has been presented by [23]; this work proposes a selection of an atmospheric light value that is directly responsible for the color authenticity and contrast of the resulting image. Additionally, they propose a fast transmission estimation algorithm to be more efficient and reduce the process time. Also using a haze model, [4] presents a haze removal technique that uses a fusion-based variational remove haze method, which combine the minimized outputs of two energy functionals to produce a haze-free version. Ju et al. [8] present an improvement by addressing the weaknesses inherent of the atmospheric scattering models; the authors develop a way to remove the haze using an adaptive method for adjusting scene transmission based on scene features. The input image is partitioned into several scenes based on the haze thickness. Then, they obtain the rough scene transmission map by maximizing the contrast in each scene and then remove the haze using the proposed adaptive method. Similarly to the previous work, Fattal et al. [3] propose to estimate the optical transmission in hazy scenes, given a single input image; the scattered light is eliminated to increase scene visibility and recover haze-free scene contrasts. Recently, in [24], a fast algorithm for single remove haze is proposed, it is based on linear transformation, by assuming that a linear relationship exists in the minimum channel between the hazy image and the haze-free image.

Lately, novel image haze removal approaches based on deep learning techniques have been proposed obtaining acceptable results. In [11] a model based on a reformulated atmospheric scattering model is proposed, instead of estimating the transmission matrix and the atmospheric light separately. Ren et.al. [16] presents a multi-scale deep neural network for single-image remove haze by learning the mapping between hazy images and their corresponding transmission maps. The proposed algorithm consists of a coarse-scale net which predicts a holistic transmission map based on the entire image and a fine-scale net that refines results locally. Cai et.al. [2] proposes 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 [27] the authors propose a unified single remove haze GAN network that jointly estimates the transmission map and performs the haze process; the network is trained using synthetic images and a two-terms loss function. The first term of the loss function is a pixel-wise Euclidean distance, while the second term consider perceptual information. In the current work a loss function based on multiple terms is proposed. Additionally, in the GAN architecture a stacking strategy is proposed to speed up the learning process. Furthermore, the proposed network architecture is trained using real images.

There are two basic things that can be done with generative based deep learning models. One is to take a collection of points and infer a function that describes the distribution that generated them. The second is to build a generative model which is to take a machine that observes many samples from a distribution and is able to create more samples from the same distribution. They allow a network to learn to generate data with the same internal structure as other data. It is a framework presented on [6] 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 [14], 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 $V(G, D)$:

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

Generative adversarial nets 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:

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

There are some techniques to improve the effectiveness of generative adversarial networks for semi-supervised learning, according to [18] that proposes some techniques, like feature matching which addresses the instability of GANs establishing a new objective for the generator that prevents it from over-training maximizing the output of the discriminator, requiring to generate data that matches the statistics of the real data. Considering the use of conditional generative networks models, this work propose the usage of an architecture similar to the one presented in [20], but by including the variation of a stacked multiple generator-discriminator networks, inspired on the work presented in [7], which consists in a top-down stack of GANs, each designed to generate lower-level representations conditioned on higher level representations. We propose a stacked 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. This work also proposes to include a multiple loss term for discriminator which make the learning process continuous and differentiable and consequently the times of convergence for the generalization of learning are improved.

3. Proposed Approach

The proposed approach is based on a of generative model, which take a collection of haze patches and form some image representation without the haze. Generative adversarial networks generate the solution, rather than finding a function; based on this principle we propose the usage of a stacked network architecture with a multiple loss to improve the generalization learning model that allows accelerate the diversity obtained in the multiple level of training. 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.

A stacked conditional network based architecture is selected due to: $i)$ the input that is introduced to the network comes from a conditional predefined latent space that optimizes the higher-level features obtained from the generator model; $ii)$ the architecture infer models in a competitive setting until it reaches some level of accuracy; $iii)$ the discriminator network is a perfect loss function for a generative model; $iv)$ it has a fast convergence capability. The network is intended to learn 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 stacking learning process. Additionally, the model has been designed to 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 an haze 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 (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 function ($\mathcal{L}$) 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 :

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

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 near image in each channel of the Stacked Gan Network.

The **intensity loss** is defined as:

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

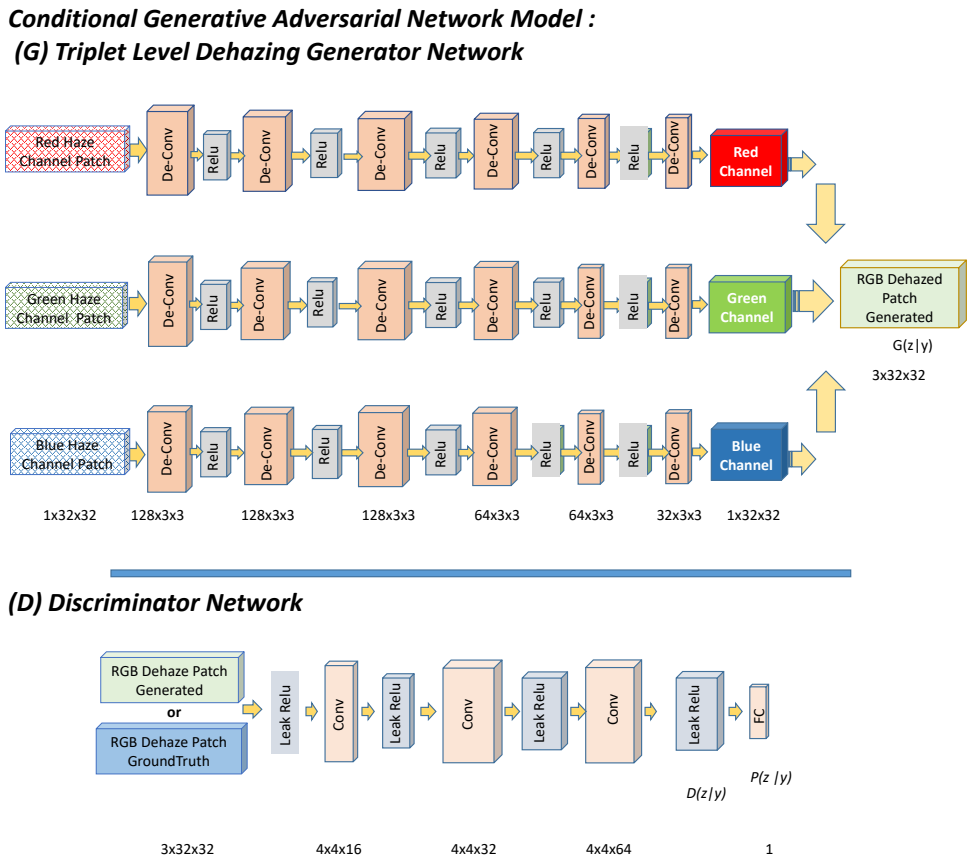

Figure 1. Illustration of the proposed triplet GAN architecture used for image dehazing.

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) [26], 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 as a function of loss is to help the learning model to produce a visually improved image, because 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 [26] 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 [25] was designed to model any image distortion via a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Let $x = \{x_i \mid i= 1,2 \ldots\ldots \mathrm{N}\}$ and y=$\{y_i \mid i= 1,2 \ldots\ldots \mathrm{N}\}$ be the original and the test image signals respectively. The proposed quality index is defined as :

$$Quality = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\,\overline{x}\,\overline{y}}{(\overline{x})^2 + (\overline{y})^2} \cdot \frac{2\,\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}, \tag{6}$$

where the first component of the equation is the correlation coefficient between x and y. The second component measures how close the mean luminance is between x and

y. The third component measures the similarity of the contrasts of the compared images. 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, for which uses a sliding window in a similar way to when a convolution is carried out, in such a way that the whole image is scanned pixel by pixel by moving the sliding window of size BxB through all the rows and columns of the image. 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 :

$$Q = \frac{1}{M}\sum_{j=1}^{M} Q_j, \tag{7}$$

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

$$\mathcal{L}_Q = \frac{1}{M}\sum_{j=1}^{M}(1 - Q_j). \tag{8}$$

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:

$$\mathcal{L}_{final} = 0.50\mathcal{L}_{Adversarial} + 0.2\mathcal{L}_{Intensity} + \\ + 0.15\mathcal{L}_{SSIM} + 0.15\mathcal{L}_Q. \tag{9}$$

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; therefore, the losses with greater fluctuation were assigned a greater proportion of impact on the optimization of the model.

The Stacked GAN network 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 0.00002 for the generator and 0.00004 for the discriminator networks respectively; epsilon = 1e-08; exponential decay rate for the 1st moment momentum 0.3 for discriminator and 0.3 for the generator; weight initializer with a standard deviation of 0.0004582; $l1$ weight regularizer; weight decay 1e-2; leak relu 0.18 and patch's size of 32×32.

The triplet architecture, see Fig. 1, maintains similar structure found in [20]. Basically in the architecture a new layer of learning was added, as well as the depth of the learning layers was increased—the learning model 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 that consists of multiple compositions of affine transformation with element-wise nonlinearity and do not stuck on saturation mode. 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 (λ) 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.

The generator (G) and discriminator (D) are both feed-forward deep neural networks that play a min-max game between one another. The generator takes as input each channel of the haze image and transforms it 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 [14]; thus, the conditional discriminator network is updated as follow:

$$\bigtriangledown_{\theta_d}\frac{1}{m}\sum_{i=1}^{m}[\log D(x^{(i)}) + \log(1 - D(G(y^{(i)}, z^{(i)})))], \tag{10}$$

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:

$$\bigtriangledown_{\theta_g}\frac{1}{m}\sum_{i=1}^{m}\log(1 - D(G(y^{(i)}, z^{(i)}))), \tag{11}$$

where m is the number of samples in each batch, y is the image without haze (RGB) generated by the network and z

Table 1. Angular Errors (AE), Mean Squared Errors (MSE), Structural Similarities (SSIM) and Image Quality Index(Q Index) obtained with the proposed Proposed Stacked Conditional GAN architecture by using different loss functions (SSIM and Q index values, the bigger the better).

Training	AE		MSE		SSIM		Q Index	
	Light Haze	*Dense Haze*	*Urban*	*Dense Haze*	*Light Haze*	*Dense Haze*	*Light Haze*	*Dense Haze*
Proposed Stacked CGAN with $\mathcal{L}_{Adversarial} + \mathcal{L}_{Intensity}$	7.18	7.11	21.96	23.75	0.72	0.69	0.62	0.59
Proposed Stacked CGAN with $\mathcal{L}_{Adversarial} + \mathcal{L}_{SSIM}$	7.12	7.03	20.97	20.74	0.78	0.72	0.64	0.61
Proposed Stacked CGAN 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 Stacked CGAN with $\mathcal{L}_{final}$	5.95	6.12	18.74	19.21	0.84	0.80	0.71	0.68

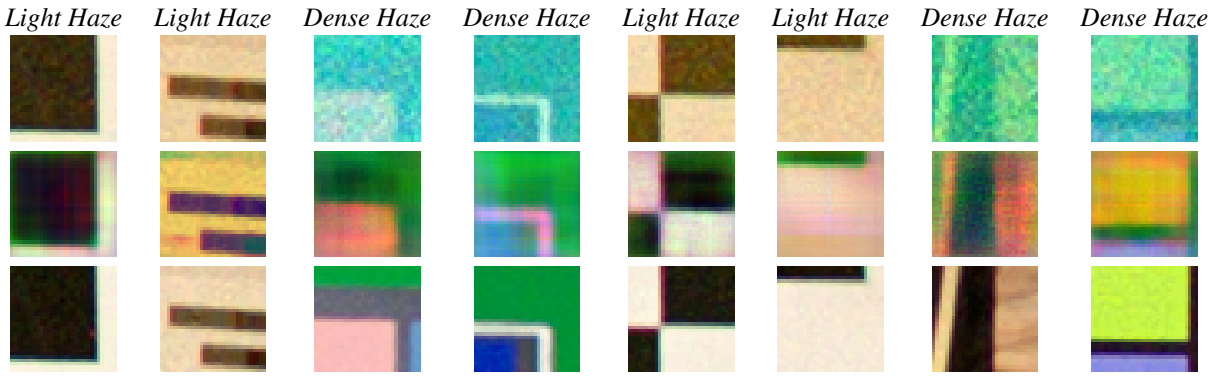

Figure 2. *(1st row)* Haze patches. *(2nd row)* Results from the proposed approach (Loss Function: $\mathcal{L}_{final}$). *(3rd row)* Ground truth images.

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.

4. Experimental Results

The proposed architecture has been evaluated using real hazed images and their corresponding clear RGB representations obtained from [13]. Figure 3 presents four images from this dataset, where ground truth image can be appreciated on (a) while different real hazed images are depicted on (b), (c) and (d). See more details about data set generation in [13]. 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 process took about 60 hours using a 3.2 eight core processor with 16GB of memory with a NVIDIA TITAN XP GPU. Some patches, with the corresponding result obtained with the proposed approach are depicted in Fig. 2; 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, [5]—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. The SSIM provides a measurement of local image quality over space while QIndex models the image distortion relative to the reference image as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. These metrics have a high degree of sensitivity to measure to image degradations, therefore, they are the more appropriate to this type of quantitative evaluation.

With the metrics mentioned above combinations of the different loss functions are evaluated, results are provided in Table 1. It can be appreciated that in all the cases the results obtained with the final loss proposed with Stacked Conditional GAN are better than those obtained with the other

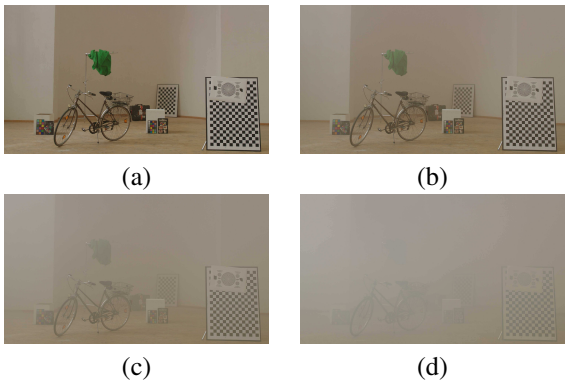

Figure 3. Set of RGB images from an indoor environment: (a) Ground truth image; (b), (c) and (d) Real images with different haze levels.

combination of losses, because are not based solely on the difference of the information of the pixels, they are based on the high-level characteristics of the images for which they are able to reconstruct better the fine details in comparison with the methods trained only by distance value of pixels. In addition, these losses, being perfectly differentiable, allow for a better optimization of the network, thus accelerating the convergence process. Just as illustrations, a few RGB images from *Light Haze* and *Dense Haze* categories, generated with the proposed Stacked GAN network, are depicted in Fig. 2 for qualitative evaluation.

5. Discussion on the usage of NIR images

The approach presented in this paper requires the existence of ground truth data, which is not always possible. In the current work, thanks the authors of [13], we were able to get real images with and without haze for training the proposed network. In order to increase the size of the dataset synthetic haze images, obtained by using an atmospheric scattering model, could be used. The problem with these kind of approaches lies on the selection of scattering model. Another option to tackle the data set drawback is based on the usage of images from other spectral band. In this direction, we have started exploring the possibility of using NIR information to remove haze in RGB images, although this work is still in progress, due to the large amount of time required for algorithm training, we are confident to obtain good results. The approach we are testing now consists of using a GAN network architecture where the generator tries to remove haze, while the discriminator evaluates the obtained images without haze with respect to the corresponding NIR image. Actually, the evaluation in the discriminator is not performed at image level but at image characteristic (we are testing image sharpness). This NIR-RGB GAN based remove haze approach can be trained with data sets such as the one used in [22] for NIR image colorization.

6. Conclusions

This paper tackles the challenging problem of generating clear RGB representations from haze images by using a novel Stacked Conditional Generative Adversarial Network model. 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 similar GAN architecture, but feed with NIR images in the discriminator to overcome this limitation. 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 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] D. Berman, S. Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016. 2

[2] 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*, 25(11):5187–5198, 2016. 2

[3] R. Fattal. Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3):72, 2008. 2

[4] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío. Fusion-based variational image dehazing. *IEEE Signal Processing Letters*, 24(2):151–155, 2017. 2

[5] A. Gijsenij, T. Gevers, and M. P. Lucassen. A perceptual comparison of distance measures for color constancy algo-

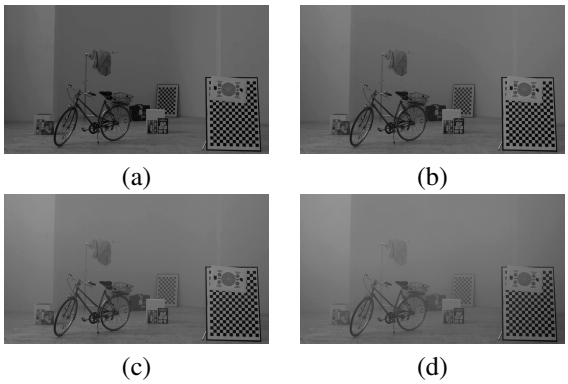

Figure 4. Set of NIR images from the indoor environment presented in 3: (a) Image without haze; (b), (c) and (d) Real images with different haze levels.

rithms. In *European Conference on Computer Vision*, pages 208–221. Springer, 2008. 6

[6] 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*, pages 2672–2680, 2014. 2

[7] X. Huang, Y. Li, O. Poursaeed, J. Hopcroft, and S. Belongie. Stacked generative adversarial networks. *arXiv preprint arXiv:1612.04357*, 2016. 3

[8] M. Ju, D. Zhang, and X. Wang. Single image dehazing via an improved atmospheric scattering model. *The Visual Computer*, 33(12):1613–1625, 2017. 2

[9] 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*, 27(5), 2008. 1

[10] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint*, 2016. 1

[11] 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. 2

[12] H. Lu, Y. Li, S. Nakashima, and S. Serikawa. Single image dehazing through improved atmospheric light estimation. *Multimedia Tools and Applications*, 75(24):17081–17096, 2016. 1

[13] J. Lüthen, J. Wörmann, M. Kleinsteuber, and J. Steurer. A rgb/nir data set for evaluating dehazing algorithms. *Electronic Imaging*, 2017(12):79–87, 2017. 6, 7

[14] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 2, 5

[15] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396*, 2016. 1

[16] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, pages 154–169. Springer, 2016. 2

[17] R. Richter. Atmospheric correction of satellite data with haze removal including a haze/clear transition region. *Computers & Geosciences*, 22(6):675–681, 1996. 2

[18] 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*, pages 2226–2234, 2016. 3

[19] S. Shwartz, E. Namer, and Y. Y. Schechner. Blind haze separation. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1984–1991. IEEE, 2006. 1

[20] P. L. Suárez, A. D. Sappa, and B. X. Vintimilla. Colorizing infrared images through a triplet conditional dcgan architecture. In *19th International Conference on Image Analysis and processing*, 2017. 3, 5

[21] P. L. Suárez, A. D. Sappa, and B. X. Vintimilla. 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*, pages 1–5. IEEE, 2017. 1

[22] 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*, pages 212–217. IEEE, 2017. 1, 7

[23] J.-B. Wang, N. He, L.-L. Zhang, and K. Lu. Single image dehazing with a physical model and dark channel prior. *Neurocomputing*, 149:718–728, 2015. 2

[24] W. Wang, X. Yuan, X. Wu, and Y. Liu. Fast image dehazing method based on linear transformation. *IEEE Transactions on Multimedia*, 19(6):1142–1155, 2017. 2

[25] Z. Wang and C. Alan. Bovik,. *A Universal Image Quality Index, In IEEE Signal Processing Letters*, 9(3):81–84, 2002. 4

[26] 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*, 13(4):600–612, 2004. 4

[27] H. Zhang, V. Sindagi, and V. M. Patel. Joint transmission map estimation and dehazing using deep networks. *arXiv preprint arXiv:1708.00581*, 2017. 2