Non-homogeneous Haze Removal Through a Multiple Attention Module Architecture

Patriot L. Suárez[1(✉)], Dario Carpio[1], and Angel D. Sappa[1,2]

[1] ESPOL Polytechnic University, FIEC, CIDIS, Guayaquil, Ecuador
{plsuarez,dncarpio,asappa}@espol.edu.ec
[2] Computer Vision Center, 08193 Bellaterra, Barcelona, Spain
asappa@cvc.uab.es

Abstract. This paper presents a novel attention based architecture to remove non-homogeneous haze. The proposed model is focused on obtaining the most representative characteristics of the image, at each learning cycle, by means of adaptive attention modules coupled with a residual learning convolutional network. The latter is based on the Res2Net model. The proposed architecture is trained with just a few set of images. Its performance is evaluated on a public benchmark—images from the non-homogeneous haze NTIRE 2021 challenge—and compared with state of the art approaches reaching the best result.

Keywords: Spatial attention · Adaptative · Residual learning · Instance normalization

1 Introduction

In recent years outdoor computer vision applications have increased enormously due to, on the one hand the reduction on the price of cameras on the other hand the technological advancements in processing capabilities (i.e., hardware and deep learning based approaches). This increase on applications can be appreciated in the video surveillance field, where global industry shipments are expected to exceed 100 million units by 2025 [17], remote sensing, driving assistance, just to mention a few. Applications developed for these fields show interesting results in clear scenarios, where sharp and well focused images are captured. However, in spite of the wide acceptance of these applications, their performance drops considerably in poor visibility scenarios. Poor visibility could be due to the lack of light or bad weather conditions (e.g., natural phenomena such as dust, mist, rain, fog or snow). This drop in performance become a challenge and an opportunity to do research on this topics (e.g., high dynamic range cameras for poor lighting conditions, image dehazing/deraining approaches). The current work tackle the dehazing problem, which is focused on removing haze from the given image.

Haze removal is a challenging task due to the complexity of the parameters that can affect the sharpness of the image. During last decades different approaches have been proposed in the literature to the homogeneous haze removal (e.g., [2,3,13,16,19]). The problem become even more complex when the non-homogeneous haze is considered. Actually, most of the approaches in the literature have been proposed by assuming

G. Bebis et al. (Eds.): ISVC 2021, LNCS 13018, pp. 178–190, 2021.
https://doi.org/10.1007/978-3-030-90436-4_14

that the haze is equally distributed all over the image. This assumption is based on the atmospheric scattering model that describes the formation of hazy images. This model has been proposed by McCartney et al. [18] and can be expressed as:

$$\mathbf{H}(p) = \mathbf{t}(p)\mathbf{DH}(p) + (1 - \mathbf{t}(p))\mathbf{A}, \tag{1}$$

where $\mathbf{H}$ is the hazy image, p is the pixel location in the image, $t(p)$ is the medium transmission map, $\mathbf{DH}(p)$ is the dehazed image, and $\mathbf{A}$ is the atmospheric light. The first term is the direct attenuation and the second term is the atmospheric light. According to Eq. 1, $DH(p)$ is recover by:

$$\mathbf{DH}(p) = \frac{\mathbf{H}(\mathbf{p}) - \mathbf{A}}{\mathbf{t}(\mathbf{p})} + \mathbf{A}. \tag{2}$$

However, Eq. 2 has a problem in its definition, because it works with assumptions for $\mathbf{t}(p)$ and $\mathbf{A}$ since they are variables that can take any value depending on the distribution of the haze in the image. Although there are some approaches that relies on this mathematical model (e.g., [3,6,9]), some failure in the haze removal process can be appreciated in these cases. In recent year, with the rise of deep learning based approaches some architectures have been proposed to capture image features and decide which region requires more attention according to the processing criteria. These models that allows to adjust the focus to the different regions of the images are an interesting option to tackle the non-homogeneous haze removal problem. The non-homogeneous haze removal has become a quite active research topic in recent years. As a proof of that it can be mentioned the NTIRE challenge regularly organized since 2017 [1] starting focused on Single Image Super-Resolution, until 2018 when the organizers start several challenges one of them related to single image dehazing and since NTIRE 2020 focus on non-homogeneous image dehazing. This challenge proposes to develop approaches to remove non-homogeneous haze by providing just a small set of images to be used for training. Removing non-homogeneous haze is a difficult problem since the effect of haze in the image causes an uncontrolled scattering of light in different regions of the images. In the current work this challenging problem is addressed by means of a novel architecture that relies on attention modules to detect and focus those regions of the image that require haze removal. In the case of haze removal, the useful information is the regions of the image with the greatest activation values that you want to preserve and locate spatially in the regions of interest on images. The proposed approach uses a feature fusion network proposed by [19], which has been modified to support a spatial attention block, with the aim of adaptively learning the spatial regions of the image that have greater representation and use them to improve the process of haze removal. Additionally, the learning transfer network from [25] has been adapted, including a residual layer of skip connections, to improve the quality of the images, and replacing the batch normalization by instance normalization, to improve image stylization. The manuscript is organized as follows. Related works are presented in Sect. 2. Then, the proposed approach is detailed in Sect. 3. Experimental results are presented in Sect. 4. Finally, conclusions are given in Sect. 5.

2 Related Work

During the last decades different haze removal approaches have been proposed in the literature. Initial approaches were mainly focused on the transmission map (see Eq. 1), while more recent ones are based on the usage of deep learning schemes. One of the approaches tackling the transmission map estimation has been presented by Berman et al. [3]. In this work, the authors propose a non-local prior-based approach. It is based on non-local prior knowledge, which assumes that the colors of a clear image are well projected by hundreds of different colors in the RGB space. These colors are grouped together and, being non-local, can be located throughout the image plane depending on the distance from the RGB sensor. In images with haze varying camera distances are converted into different transmission coefficients. Therefore, each group of colors in the image without haze is called a line in RGB space. With these lines the algorithm calculates the distance map of the image and therefore the image can be obtained without haze. This algorithm requires no training and works with a wide range of images. The algorithm is based on the grouping of the colors present in an image without haze, these groups are converted into lines in RGB space and the authors have called them haze lines, with which the distance map is calculated and it is possible to remove the haze in the images. Another technique, presented in [2], is based on a fusion resulting from a process of applying white balance and contrast enhancement. The method uses the pyramidal Laplacian representation to minimize artifacts. According to the authors, the results are comparable to other similar techniques. On the contrary to previous approaches, Long et al. [16] propose to use a dark channel prior to haze removal. In order to avoid artifacts, the authors use a low-pass Gaussian filter to change the coarse atmospheric estimation. They modified the transmission to prevent color distortion. Experimental results achieve good results showing very fine details in the recovered image. Another method for the haze removal is proposed in [7]; it is based on the minimization of two fused energy functions. As a result of these optimized combined functions, a set of saturation difference maps is obtained that allows to remove the haze from the images. Experimental results show that the image's structures are very well preserved.

In recent years several deep learning based approaches have been proposed; for instance Li et al. [13] presents a semi-supervised approach to image haze removal based on a CNN network with multiple losses. Also, gradient information is used as a prior to facilitate the haze removal process. This approach has been evaluated in real and synthetic images presenting results comparable to other deep learning based approaches Following the line of convolutional networks, Li et al. [12] proposes an adaptive regional network (RATNet) that performs a regular global filtering, and the restoration of lighting and details, separately, which at the end are concatenated to obtain the image without haze. The obtained results are similar to other adaptive learning convolutional but it has not been tested with synthetic images like previous approach. Qin et al. [19] present an interesting approach based on a Feature Attention (FA) module that uses channel and pixel attention to extract the most important characteristics from images and contributes to taking this information proportionally, because the haze is not proportionally distributed in the image. This architecture defines a local residual learning that inhibits the characteristics of the image. It corresponds to the haze from being considered in the learning. The results that have been obtained for the non-homogeneous haze removal process are very efficient and one of the most important in the state of the art. Given these results, in our approach we have decided to include this attention block, but it has been modified to include spatial attention, to improve the removal of the haze, taking into account the most representative characteristics, respecting their spatial location within the image. Another technique presented in the literature [20] is a method also based on deep learning, in this case the authors address the removal of haze in synthetic images. However, they apply a prior bidirectional domain transfer process so that the synthetic images can better generalize in the training process. Then these enhanced images are used as input of two simultaneous haze removal networks with consistency constraint, which are trained and then merged to obtain the resulting clear image. According to the authors, the proposed approached is robust to remove haze in both real and synthetic images. Another method to remove the haze is presented in [5], where a network with two modules is proposed; one of them performs the color correction for each channel to balance the image with haze and the second module to improve the visibility of the images based on attention blocks. The obtained results show that the proposed approach restores hazy images whether these are real or synthetically generated.

3 Proposed Approach

This section presents the approach proposed to perform image haze removal. It is based on the usage of an improved version of a channel adaptive residual learning deep network, which has been initially presented in [26]. According to the authors, the attention blocks allows the network to adaptively learn relevant, versus non-relevant information present in the image, by using the features extracted from each channel of the image. This adaptive process is especially interesting when the image does not have the evenly distributed haze.

3.1 Architecture

As mentioned above, the proposed architecture is based on the combination of a characteristic attention scheme and a stylized transfer learning that allow to eliminate the non-homogeneous haze present in the images. Figure 1 illustrates each of the model components. First, the "Pre-trained general feature extractor" is modeled inspired by [25]. This extractor has been created based on a wide data domain learning which helps during the training with few samples, as is our case, for which the weights of the Res2Net model have been loaded to accelerate the convergence and avoid overfitting, so as not to start learning from scratch. The architecture, also includes multiple skip connections, residual learning and a combination of spatial and channel attention maps to help extract high frequency data. Also, the other component named as "Non-Homogeneous Dehaze Feature Extractor" of the hazy images is modeled, refined by means of attention modules. This component is in charge of the extraction of the most representative characteristics of the hazy images. It is made up of learning groups that contains a stack of attention blocks that focus on extracting spatial descriptors, per channel and per pixel.

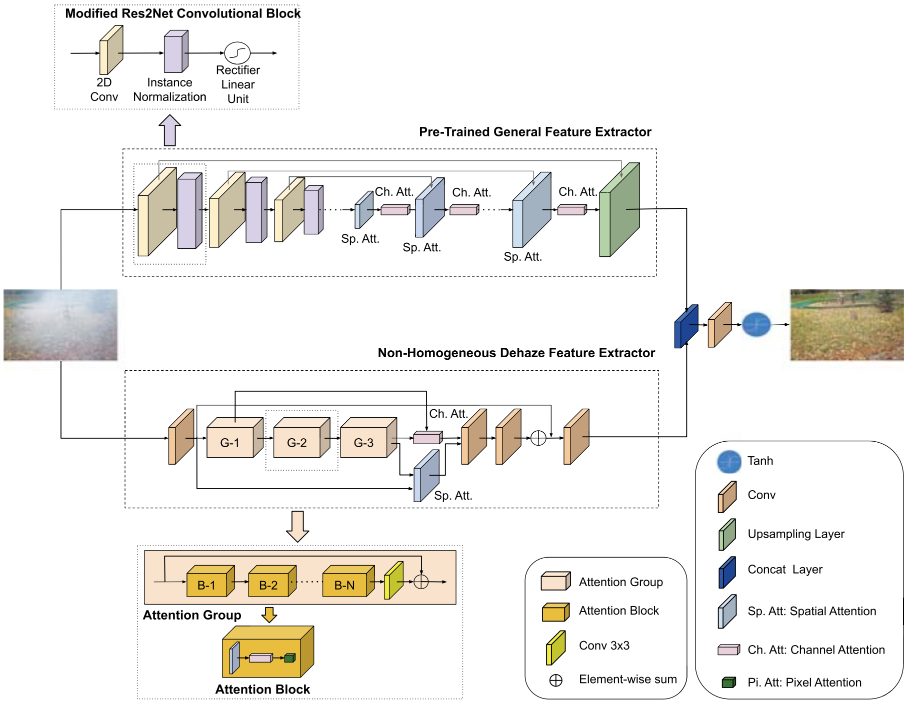

Fig. 1. Proposed dehaze architecture.

The depth of block stacking allows for further refinement of the resulting descriptor characteristics. After the grouping of blocks, the model proposes to apply spatial and channel attention to stylize the information obtained from the stacking of blocks; this is additionally combined with multiple skip connections and residual learning. The final result of the descriptor component of the images with non-homogeneous haze is concatenated with the final result of the global feature extractor. The final result of the descriptor component is concatenated with the last result of the global feature extractor. An hyperbolic tangential activation function has been applied to obtain the reconstruction of the image without haze and be able to estimate the convergence of the model applying gradients and backpropagation.

The proposed model has been designed to work with attention modules combined with residual learning to improve the transfer of characteristics through the learning layers. Regarding the attention modules used in the proposed architecture, the attention modules spatial, channel and pixel are combined to help the haze elimination process inspired by the work presented in [19]. The main contribution of the proposed architecture to improve the haze elimination process is the introduction of the spatial attention module inspired by [24] that allows local focus on the most representative features of the haze image and mapping them with the fundamental truth, to refine the feature descriptor. It is important to emphasize that in the "Non-Homogeneous Dehaze Feature

Extractor" a residual connections have been applied in the model in conjunction with the spatial attention modules to accelerate the convergence of the model and improve the haze elimination process. It is important to highlight that in the learning transfer network based on a Res2Net network, the normalization of the model data has been modified by substituting batch normalization for instances normalization (see Fig. 3), as suggested by Ulyanov et al. [21], in order to improve the stylization of image details and reduce the fading of gradients to speed up model convergence, see Sect. 3.3 for more detail. Additionally, in the general feature extractor of prior knowledge transfer network, a spatial and channel attention layer has been included. Besides, multiple skip connections and residual learning to refine the high-frequency characteristics of the pre-trained model.

This model can be as deep as the non-homogeneous distribution of the haze is complex, which is why it is a parameter of the model. The attention modules applied in the model are explained in detail below.

Attention Modules

Continuing with the evolution of deep learning models, we now have the so-called attention modules that have emerged from some previous works such as [14] and [10], where they were implemented to make the selection of areas of interest for a specific purpose. Nowadays, with the improvements in computer equipment, complex problems can be addressed with deep learning architectures [4, 19]. This technology has evolved in such a way that simple models are now designed to detect relevant features, but with

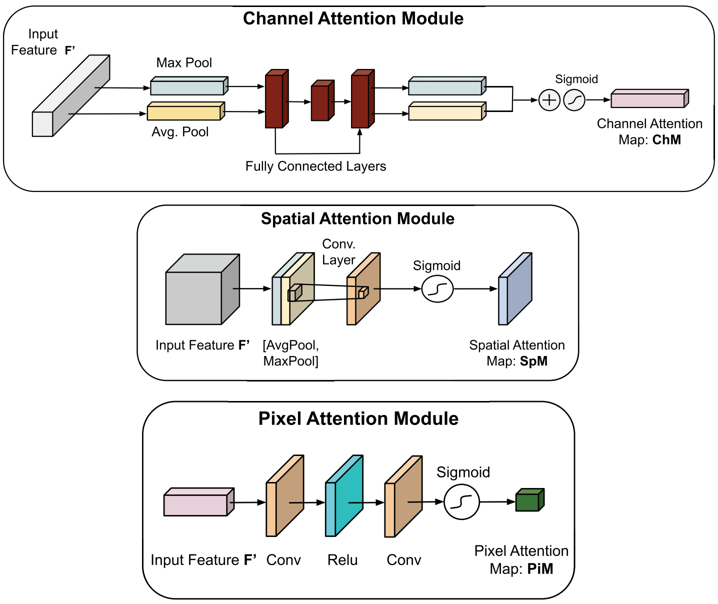

Fig. 2. Attention modules implemented.

multiple levels of learning. These models include attention modules used to generate blocks of attention grouped at various levels so that they can focus their learning on the most representative information and not consider useless low frequency information present in the given image. In the case of haze removal, the useful information are the patterns that best fit in the network activation functions. Our approach uses three levels of attention modules (spatial, channel and pixel) grouped in the learning layers of the proposed model. The main idea is to detect the existing correlation in the characteristic maps of each layer of the model to determine the best attention mechanism that is capable of learning where is the most relevant information finding the mapping between the input and the ground truth using the descending gradient and the back propagation during the learning process. The different attention modules are detailed below.

Spatial Attention: it consists of a convolutional network to generate a map to extract the inter-spatial relationship of characteristics. The spatial attention map (SpM) is characterized by the inter-spatial relationship of image characteristics. In addition, this map focuses on where the most representative part is located, which is complementary to the attention of the channel. This spatial attention is computed by means of average grouping and maximum grouping operations along the channel axis and the outputs are concatenated to generate an efficient characteristic descriptor. A convolution layer is applied to the resulting feature descriptor to generate a spatial attention map using a sigmoid activation function that encodes where to emphasize or suppress. Figure 2 shows the implemented spatial attention module. The spatial attention map can be expressed as:

$$\begin{aligned}\mathbf{SpM}(\mathbf{F}) &= \sigma\left(f^{m\times m}([\mathrm{AvgPool}(\mathbf{F}); \mathrm{MaxPool}(\mathbf{F})])\right)\\ &= \sigma\left(f^{n\times n}\left([\mathbf{F}\mathrm{avg}^{\mathbf{spm}}; \mathbf{F}\mathrm{max}^{\mathbf{spm}}]\right)\right),\end{aligned} \tag{3}$$

where σ corresponds to the sigmoid function and $(f^{n\times n})$ represents a pooling operation with the filter size of $m \times m$

Channel Attention. The channel attention module is a combination of operations performed in a network based on convolutions and pooling operations. A channel attention map is computed by obtaining the relationship of characteristics between channels. Since each map is considered a feature detector, the channel's attention is focused on "what" is representative per channel in a given image used as input to the model. This module of attention per channel works in combination with the module of spatial attention. In our case, we have used channel attention independently in the learning transfer network and in a combined way in the attention block to extract the descriptor that represents the most representative characteristic map. Figure 2 shows the implemented channel attention module. The channel attention map can be formulated as:

$$\mathbf{ChM}(\mathbf{F}) = \sigma(MLP(\mathrm{AvgPool}(\mathbf{F})) + MLP(\mathrm{MaxPool}(\mathbf{F})), \tag{4}$$

where σ corresponds to the sigmoid function.

Pixel Attention. A pixel attention module has been included to maximize the learning of context locations for each pixel. With this information, a pixel map is constructed, which weights correspond to relevant features mapped to contextual information. Figure 2 shows the implemented pixel attention module. This contributes maintain uniformly the removing process despite a non-homogeneous haze. The pixel attention map can be formulated as:

$$\mathbf{PiM}(\mathbf{F}) = \sigma\left(\mathrm{Conv}\left(\delta\left(\mathrm{Conv}\left(F^{*}\right)\right)\right)\right), \tag{5}$$

where σ corresponds to the sigmoid function and δ correspond to RELU function.

3.2 Loss Functions

The loss functions used to optimize the proposed image de-hazing architecture are as follow. Firstly, an L1 loss is considered to minimize the outliers in a better way. This loss function can be formulated as:

$$\mathcal{L}_{l1} = \frac{1}{N}\sum_{j}^{N} 11\ \left(gt_j - f_{\gamma}\left(hz_j\right)\right), \tag{6}$$

where hz_j corresponds to hazy image and gt_j refers to ground truth at pixel j. $f_{\alpha}(gt)$ denotes our model parameterized by α. N is the number of pixels evaluated. In order to address the limitations of the l1 loss function, the usage of a reference-based measure is proposed. One of the reference-based index is the structural similarity index [23], which evaluates images accounting for the fact that the human visual perception system is sensitive to changes in the local structure; the idea behind this loss function is to help the learning model to produce a visually improved image. The structural similarity loss is defined as:

$$\mathcal{L}_{SSIM} = \frac{1}{B}\sum_{i=1}^{B} 1 - SSIM(i), \tag{7}$$

where SSIM(i) is the Structural Similarity Index (see [23] for more details) of each image i of the batch size B.

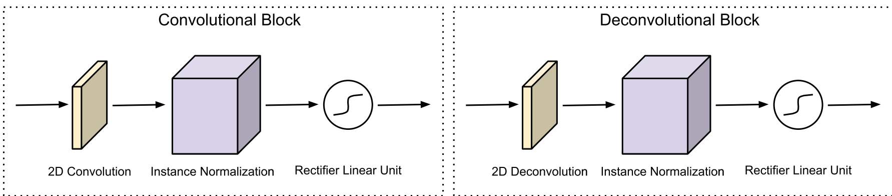

Fig. 3. Instance Normalization applied in Res2Net block.

In addition to the previous terms, the perceptual loss [27] is also applied to extract high-level features to enhanced the restoring image process. We use the Res2Net pre-trained loss network θ. The loss function is described as:

$$\mathcal{L}_{\text{percep}} = \frac{1}{C_l H_l W_l}\left\|\phi_{lj}\left(f_{\theta}(h)\right) - \phi_l(gt)\right\|_2^2, \tag{8}$$

where hz and gt are hazy inputs and ground truth images, respectively. $f_{\theta}(hz)$ is the dehazed images. $\phi_l(gt)$ denotes the feature map with size $C_l \times H_l \times W_l$. The feature restoration loss is the L_2 loss.

3.3 Instance Normalization

Another field of analysis has emerged around the style of an image is evaluated by the statistics of convolutional neural network filters, a renewed interest in the texture generation and image stylization problems in order to obtain qualitative improvement in the generated image, discarding the instance-specific contrast information from an image during style transfer.

Ulyanov et al. [21] introduce a method named *instance normalization* for a better stylization and texture synthesis, that derive entropy loss which improves samples diversity. The instance normalization layer is applied at test time as well as at training time. According to [22] the generator network should discard contrast information in the content image to learn a highly nonlinear contrast normalization function as a combination of such layers. Let $x \in \mathbb{R}^{N \times C \times W \times H}$ an input tensor containing a batch of N images, where C, W and H are the depth, width and high respectively of the image tensor and let x_{tijk} denote its $tijk$-th element of x image tensor, where k and j span spatial dimensions, t is the index of the image in the batch, i is the feature channel (in the case of an RGB image being used as an input, it would represent a color channel), a simple version of instance normalization is defined as:

$$y_{tijk} = \frac{x_{tijk}}{\sum_{l=1}^{W} \sum_{m=1}^{H} x_{tilm}}. \tag{9}$$

A small change in the stylization architecture proposed by [22] is a qualitative improvement in the generated images. The change is limited to swapping batch normalization with instance normalization, and to apply the latter both at training and testing times. The resulting method can be used to train high-performance architectures for real-time image generation.

Table 1. Results from the NH-Haze 2021 validation dataset. The best results are in **bold**, and the second best are underlined.

Methods	NH-Haze 2021	
	PSNR	SSIM
DCP [8]	11.68	0.7090
AOD-Net [11]	13.30	0.4693
GCANet [4]	18.79	0.7729
FFA [19]	20.45	0.8043
TDN [15]	20.23	0.7622
TBD [25]	<u>21.66</u>	<u>0.8430</u>
Ours	**22.42**	**0.8546**

4 Experimental Results

This section starts by describing the datasets; then training details are provided and next the evaluation metrics are introduced; finally, results from the proposed approach

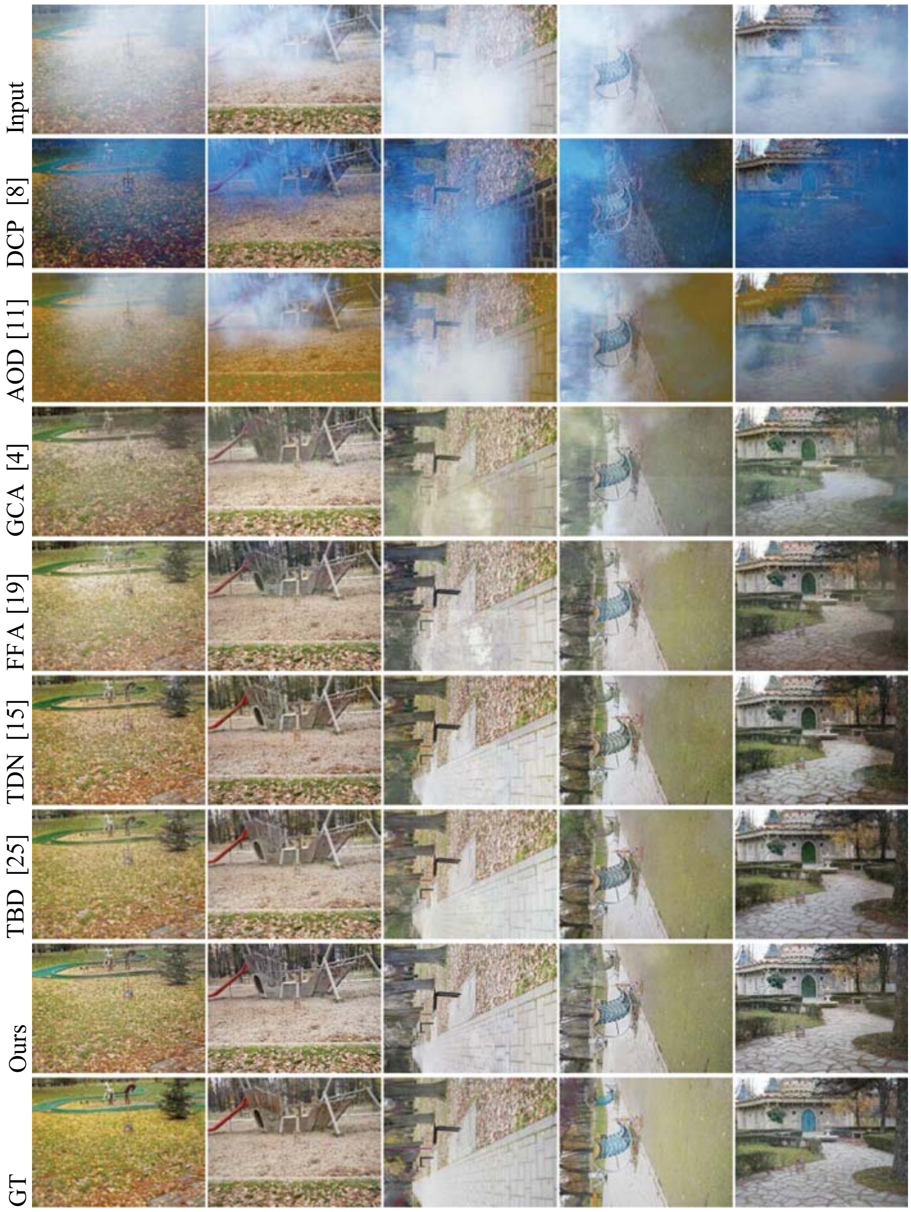

Fig. 4. Experimental results: 1st. row shows the validation images; 2nd. to 7th. row correspond to results from state of the art techniques; 8th. row presents results from the proposed approach; 9th. row depicts ground truth images from NH-Haze 2021 dataset.

are presented and compared with the state-of-the-art dehazing algorithms, both quantitatively and qualitatively.

4.1 Datasets

The proposed architecture has been evaluated by using the NTIRE 2021 challenge dataset, which contains images with non-homogeneous haze. Since the validation images for this challenge are not available yet, a group of 5 pairs of images 21–25 has been selected from the provided set of images and used to compute the metrics to evaluate results. The NTIRE 2021 challenge dataset contains 30 non-homogeneous haze images and their corresponding ground truth. From the group of data to carry out the experiments, 20 pairs of images have been selected for training. Additionally, 5 pairs of images have been selected for testing and 5 images for model validation.

Fig. 5. Results from the proposed approach on the NH-Haze 2021 validation images—note these images have been used as benchmark at the NTIRE 2021 challenge, hence the ground truth are not provided yet and just qualitative results can be presented.

4.2 Training Settings

A data augmentation process has been performed on the training set. It consists on random rotation of 90, 180 and 270 °C; additionally, an horizontal flip has been applied. The hazy-clean pairs of images are randomly cropped in patches of 256×256. Adam optimizer has been used, where β_1 and β_2 take the default values of 0.9 and 0.999, respectively. For quantitative evaluation, we adopt the Peak Signal to Noise Ratio (PSNR) metric and the Structural Similarity Index (SSIM) metric. Two NVIDIA A100 GPU were used during training, with a batch size of 8 and a learning rate of 0.0002. The train process takes about 72 h.

4.3 Comparisons

The performance of the proposed approach has been compared with six state-of-the-art dehazing algorithms, which are: DCP [8], AOD-Net [11], GCA-Net [4], FFA-Net [19], TDN [15] and TBD [25]. Table 1 presents the obtained results from all the approaches. Figure 4, shows the qualitative results obtained with the validation set with all approaches compared with the ground truth images. It can be appreciated that our approach presents better results when compared with all other approaches. Results from the 5 images in the validation set of the NTIRE 2021 dehazing challenge are provided in Fig. 5 for a qualitative evaluation.

5 Conclusions

This paper presents a deep learning network based on three-level attention modules (spatial, channel and pixel) with the aim of extracting the most relevant information to maintain from the image to which the haze elimination is being applied. In particular our contribution is the spatial attention that has been applied. It focuses on the characteristics of the most relevant locally presented in the image and not considering those characteristics that do not contribute to the clarification of the image. Additionally, the learning transfer of the pre-trained Res2Net model has been applied to take the learning of the network and not start from scratch, taking into account that the dataset is very limited. In this way, the Res2Net block has also been modified to work with normalization by instance, in order to facilitate the stylization of the details of the images and accelerate the convergence of the model. Our method presents improved results if we compare it with other methods of the state of the art.

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

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 126–135 (2017)
2. Ancuti, C.O., Ancuti, C.: Single image dehazing by multi-scale fusion. IEEE Trans. Image Process. **22**(8), 3271–3282 (2013)
3. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1674–1682 (2016)
4. Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1375–1383. IEEE (2019)
5. Dudhane, A., Biradar, K.M., Patil, P.W., Hambarde, P., Murala, S.: Varicolored image dehazing. In: proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4564–4573 (2020)
6. Fattal, R.: Dehazing using color-lines. ACM Trans. Graph. (TOG) **34**(1), 1–14 (2014)
7. Galdran, A., Vazquez-Corral, J., Pardo, D., Bertalmio, M.: Fusion-based variational image dehazing. IEEE Sig. Process. Lett. **24**(2), 151–155 (2016)
8. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1956–1963 (2009)
9. Ju, M., Ding, C., Ren, W., Yang, Y., Zhang, D., Guo, Y.J.: Ide: Image dehazing and exposure using an enhanced atmospheric scattering model. IEEE Trans. Image Process. **30**, 2180–2192 (2021)
10. Lee, M., Ban, S.-W.: Incremental knowledge representation based on visual selective attention. In: Ishikawa, M., Doya, K., Miyamoto, H., Yamakawa, T. (eds.) ICONIP 2007. LNCS, vol. 4985, pp. 940–949. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69162-4_98

11. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: AOD-Net: all-in-one dehazing network. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 4780–4788 (2017)
12. Li, H., Wu, Q., Ngan, K.N., Li, H., Meng, F.: Region adaptive two-shot network for single image dehazing. In: 2020 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2020)
13. Li, L., Dong, Y., Ren, W., Pan, J., Gao, C., Sang, N., Yang, M.H.: Semi-supervised image dehazing. IEEE Trans. Image Process. **29**, 2766–2779 (2019)
14. Liew, S.H., Low, Y.F., Lim, K.C., Choo, Y.H., Farghaly, M.R.M.: Performance evaluation of convolutional neural network in classification of EEG signals based on attention task. ARPN J. Eng. Appl. Sci. **13**, 3400–3404 (2006)
15. Liu, J., Wu, H., Xie, Y., Qu, Y., Ma, L.: Trident dehazing network. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1732–1741 (2020)
16. Long, J., Shi, Z., Tang, W., Zhang, C.: Single remote sensing image dehazing. IEEE Geosci. Remote Sens. Lett. **11**(1), 59–63 (2013)
17. Markets and Markets, Inc.: Video Surveillance Market. https://www.marketsandmarkets.com/Market-Reports/video-surveillance-market-645.html
18. McCartney, E.J.: Optics of the Atmosphere: Scattering by Molecules and Particles. Wiley, New York (1976)
19. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: FFA-Net: feature fusion attention network for single image dehazing. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 11908–11915 (2020)
20. Shao, Y., Li, L., Ren, W., Gao, C., Sang, N.: Domain adaptation for image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2808–2817 (2020)
21. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: the missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
22. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6924–6932 (2017)
23. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
24. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
25. Yu, Y., Liu, H., Fu, M., Chen, J., Wang, X., Wang, K.: A two-branch neural network for non-homogeneous dehazing via ensemble learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 193–202 (2021)
26. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 294–310. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_18
27. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)