High-Resolution Single Image Dehazing using Encoder-Decoder Architecture

Simone Bianco Luigi Celona Flavio Piccoli Raimondo Schettini
Department of Informatics, Systems and Communication
University of Milano-Bicocca
viale Sarca, 336, Milano, Italy
{simone.bianco, luigi.celona, flavio.piccoli, schettini}@disco.unimib.it

(a) Real-world hazy image

(b) Image corrected by the proposed system (HR-Dehazer)

Abstract

In this work we propose HR-Dehazer, a novel and accurate method for image dehazing. An encoder-decoder neural network is trained to learn a direct mapping between a hazy image and its respective clear version. We designed a special loss that forces the network to keep into account the semantics of the input image and to promote consistency among local structures. In addition, this loss makes the system more invariant to scale changes. Quantitative results on the recently released Dense-Haze dataset introduced for the NTIRE2019-Dehazing Challenge demonstrates the effectiveness of the proposed method. Furthermore, qualitative results on real data show that the described solution generalizes well to different never-seen scenarios.

1. Introduction

Outdoor images are often affected by visibility degradation due to bad weather conditions. In particular, hazy and foggy weather generates tiny water droplets or ice crystals suspended in the air. This aerosol filters the sunlight causing color shifting, contrast loss, saturation attenuation and loss in sharpness. This phenomenon is heavily connected to the distance of the subjects in the scene [6] and is spatially varying in such a way that is not predictable. In extreme cases, parts of the image are totally lost. The process of removal of those effects in favor of a clearer version of the input image is called *dehazing* and it is a very challenging and important task. It finds many applications in computer vision such as video surveillance [9], smartphone cameras [37], remote sensing and visual navigation [22] and nowadays, with the development of autonomous driving vehicles, dehazing has gained even more attention [31, 34]. In fact, haze heavily affects the performance of a classification or a semantic segmentation system, and thus a pre-processing that recovers back the depicted subjects is highly desirable.

The model that best describes haze has been proposed by Mc Cartney *et al.* [24]. It is called *atmospheric scattering model* and combines the scene radiance of the haze-free input with the global atmospheric illumination, blended together with a transmission coefficient that models the haze effects. Specifically, it is:

$$\mathbf{I}(\mathrm{x}) = \mathbf{J}(\mathrm{x})t(\mathrm{x}) + \mathbf{A}(\mathrm{x})(1 - t(\mathrm{x})) \tag{1}$$

where x is the 2D pixel spatial location, $\mathbf{I}(\mathrm{x})$ is the observed haze image, $\mathbf{J}(\mathrm{x})$ is the haze-free scene radiance to be recovered, while $\mathbf{A}(\mathrm{x})$ and $t(\mathrm{x})$ are two critical parameters. The first, $\mathbf{A}(\mathrm{x})$, denotes the global atmospheric illumina-

tion, and $t(\mathrm{x})$ is the transmission matrix defined as:

$$t(\mathrm{x}) = e^{-\beta d(\mathrm{x})}, \tag{2}$$

where β is the scattering coefficient of the atmosphere, and $d(\mathrm{x})$ is the distance between the object in the scene projected at the spatial location x and the camera. In Equations 1 and 2 the bold variables represent full color three-channel images, while transmission $t(x)$ is a single-channel image that is replicated along channel dimension; with abuse of notation the multiplications are intended channel-by-channel and pixel by pixel. From Equations 1 and 2, it is possible to observe that dehazing is an ill-posed problem, in fact multiple solutions can be found for a single input hazy image. The state of the art is studied by several methods that tackle the problem under different points of view. Depending on the adopted features and the hypothesis at the basis of the dehazing methods, there are two different types of restoration: multi-image and single-image. Algorithms using several images exploit different inter-image differences in order to remove the haze. For example Schechner *et al.* [32] use two images of the same target acquired with different polarization. Nayar *et al.* [27] and Narasimhan *et al.* [25, 26] exploit the dynamics of the hazing phenomenon and acquire the target in different moments. Although this can lead to very good results, in practice it is not applicable in time-constrained scenarios. Kopf *et al.* [18] use multiple images in combination with 3D registered models. Despite their ability in removing haze, multi-image algorithms are in practices not very useful because they require special equipment or take too much time.

The first group of single-image methods follows a physical model. Fattal *et al.* [11] try to model the transmission (see Eq. 2). This method is quite slow and thus not suitable in real time applications. In Cheng *et al.* [8], semantic information is used to infer color priors. Semantic context is used also to estimate the ambient illumination. Cai *et al.* [7] and Ren *et al.* [29] estimate the medium transmission map through the use of a convolutional neural network (CNN). The former recovers the image using an atmospheric scattering model, while the latter employs a linear combination of the estimated transmission map together with the estimated global atmospheric light. Among the physical methods there is a subgroup of methods that exploit the fading of haze in function of the distance. Different techniques are used to estimate the fading. Hautiere *et al.* [14] use depth for road scenes. Berman *et al.* [6] estimate haze lines to define the evolution direction of the haze.

A second approach to dehazing is given by methods that try to maximize the contrast. Tan *et al.* [34] propose a technique based on Markov random fields optimized through a cost function that quantitatively measures the image contrast in association with the number of edges. He *et al.* [15] use the *dark channel prior* to improve the contrast. Dark channel prior is based on statistics of haze-free outdoor images. The hypothesis held in this paper states that in local regions which do not cover the sky, it is very often that some pixels (the *dark pixels*) have very low intensity in at least one RGB component. The value of the dark pixel is directly affected by the airlight, therefore it directly provides an accurate estimation of the haze's transmission. This strong assumption however, fails when the scene object has a color similar to the airlight over a large local region and no shadow is cast on the object. Similarly, Golts *et al.* [13] exploit the dark channel prior to train a convolutional neural network in an unsupervised manner. Differently from He *et al.* [15], the dark channel prior used in their work is an image statistical property indicating that in small patches belonging to clear images, the darkest pixel across all color channels is very dark and thus, close to zero. The concept of dark channel prior is used in the loss module, which is composed also by a matting term and a data term.

Another way to approach dehazing consists in learning a direct mapping between the hazy image and its corresponding clear, haze-free, image. In this case, the distance between local or global statistics among the pair images is used as a training loss. In Galgran *et al.* [12], Retinex [19] is used as a guidance training loss. This module however is very slow and leads to poor results. Zhang *et al.* [39] use perceptual loss over the luminance channel of the YCbCr color space. A deeper analysis done by Wang *et al.* [35] in fact, shows that haze affects more the luminance than the chrominance. Yang *et al.* [38] extend this concept and adapt the perceptual loss to the domain under analysis by learning the weights of the neural network composing the perceptual loss through adversarial learning.

To assess the quality of the dehazing methods, in the recent years many benchmarks of increasing difficulty have been released to push researches improving their methods. In particular, in 2018 two datasets for image dehazing have been made available. Ancuti et al., in fact, proposed an indoor [2] and an outdoor [5] dataset composed by real RGB images to which is applied a synthetic fog. This benchmarks have been used for the two tracks of the NTIRE18 Image Dehazing Challenge [1].

We strongly believe that a dehazing system based on neural networks trained with low- and high-level feature matching can dramatically improve the performance of a dehazing system and therefore we used this approach. Results scored over several datasets show that our method outperforms the state of the art. In short, our contribution is:

- an end-to-end trainable model for single-image dehazing that learns a direct mapping among hazy and haze-free images;

- a training loss that forces the neural network in charge of restoring the input image to learn the semantic of the

Channel	MSE ($\times 10^{-2}$)
R	9.15
G	11.17
B	15.52
Y	8.01
Cb	0.23
Cr	0.11

Table 1. Mean square error (MSE) of every channel for RGB and YCbCr color spaces on the training set images of Dense-Haze dataset [4].

input image and to promote local consistency among nearby structures as well as visual consistency;

- a method able to generalize well on synthetic as well as on real-world hazy images.

2. Proposed method

In this section the proposed solution for single image dehazing, called HR-Dehazer, is described. First of all an analysis of the influence of haze on different color spaces is conducted. Then, the proposed encoder-decoder architecture is described and finally the loss functions and the training procedure are presented.

2.1. Color space

Atmospheric illumination in hazy images has a different impact on each channel in RGB and YCbCr color spaces. The hazy effect is mainly concentrated on the luminance (Y) channel of the YCbCr color space. This hypothesis is here quantitatively demonstrated by estimating the mean square error (MSE) between the hazy image and its corresponding haze-free image for each channel both in the RGB and the YCbCr color space. Table 1 reports MSE of each channel on the training set images of the Dense-Haze dataset. From the results reported we can observe that among RGB channels is approximately the same, while the one for the luminance channel is much higher than the one for blue and red chroma components. Therefore, to allow the model to independently improve luminance and chroma channels thus preserving the color balance, the proposed model will process a YCbCr version of the input image.

2.2. Architecture

Figure 1 shows the proposed encoder-decoder architecture used for mapping the hazy image into a reconstructed haze-free image, which is inspired by the U-Net architecture [30]. This model has a symmetric shape, *i.e.* the encoder and the decoder contract and expand in a similar way, thus allowing the preservation of boundary information (spatial information). The proposed architecture contains concatenated skip connections between feature maps of the downsampling path and feature maps of the upsampling path. Skip connections allow features reuse and implicit deep supervision which guarantees better gradient propagation across the network. In order to reduce the generation of artifacts and gain sample independence, we respectively replace Convolution Transpose layers with Pixel Shuffle layers [33] and remove all Batch Normalization layers.

2.3. Loss function

We optimize the proposed model by minimizing the linear combination of mean square error (MSE) loss, $\mathcal{L}_E$, and perceptual loss, $\mathcal{L}_P$, as shown below:

$$\mathcal{L} = \mathcal{L}_E + \mathcal{L}_P. \tag{3}$$

$\mathcal{L}_E$ measures the error between pixel values of the target haze-free image, I_t^{YCbCr}, and the corresponding reconstructed hazy image, $G(I^{YCbCr}, \Theta)$, in the YCbCr color space and is defined as follows:

$$\mathcal{L}_E = \left\| G(I^{YCbCr}, \Theta), I_t^{YCbCr} \right\|_2, \tag{4}$$

where G is the proposed encoder-decoder network and Θ the corresponding parameters.

The perceptual loss $\mathcal{L}_P$ [17] computes the error between features extracted from both haze-free and hazy images with a pre-trained CNN. Formally, given a pair of images, I^{RGB} and I_t^{RGB}, where the first is the hazy image and the second is the corresponding haze-free image, the perceptual loss is defined as:

$$\mathcal{L}_P = \sum_{l \in L} w_l \left\| F_l(I^{RGB}, \Theta) - F_l(I_t^{RGB}, \Theta) \right\|_2, \tag{5}$$

where F_l is the output feature map of a specific layer l, and w_l is the weight used for combining losses for several layers. In our work, we use features at layers $L = \{\text{RELU11}, \text{RELU21}, \text{RELU31}\}$ of a VGG-19 model trained on ImageNet dataset.

2.4. Multi-scale training

To allow the network to take into account the semantics of the entire input image and also to impose coherence among local structures, we propose a new multi-scale training procedure. During training, a batch consists of a pair of images: hazy and corresponding haze-free images. Multi-scale training is carried out by resizing batch images while preserving their aspect ratio. For each batch, images are resized by randomly sampling the dimension of the short side from the following sizes: 512, 1024, or 2048 pixels. Applied data augmentation includes horizontal flipping, random rotation in the range $\pm 45°$, perspective transform, and affine transform.

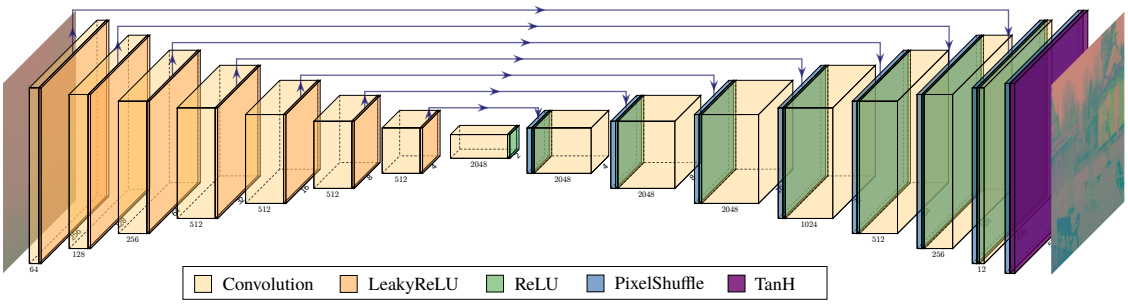


Figure 1. The HR-Dehazer architecture is an encoder-decoder network inspired by U-Net [30]. It consists of Convolution and LeakyReLU layers in the downsampling part, while upsampling part includes PixelShuffle, Convolution and ReLU layers.

As described in the previous section, we optimize the proposed model by minimizing the linear combination of mean square error (MSE) and perceptual loss [17] as illustrated in Figure 2. The MSE is computed on data at the randomly sampled scale in YCbCr color space, while the perceptual loss is estimated on RGB data at two different scales: the randomly sampled one, and 224×224 pixels (the same used for training the feature extraction model on the ImageNet data).

2.5. Inference phase

At inference time, given an input RGB hazy image of any size, the proposed solution: (1) pre-processes the image by converting it from the RGB to the YCbCR color space; (2) performs a forward pass through the encoder-decoder architecture using the pre-processed image as input; (3) transforms the reconstructed image from the YCbCr color space to the RGB color space.

3. Experiments and results

In this section, we describe the datasets along with the settings used for experiments. Then, quantitative and qualitative results achieved by the HR-Dehazer are compared with state-of-the-art methods.

3.1. Training data

To allow the proposed model to be able to reconstruct in a meaningful way a vast range of semantic concepts, we have considered the use of two datasets for training. The first dataset is created by the authors for training from scratch the proposed model and is obtained by synthesizing a pair of haze-free/hazy images. While the second is the Dense-Haze dataset, which is the dataset used for the NTIRE2019-Dehazing Challenge [3] and is used for fine-tuning the model.

PASCAL VOC2012: The PASCAL VOC2012 dataset [10] consists of 17125 haze-free images containing both outdoor and indoor scenes depicting an instance of 20 different categories. We select this dataset for training from scratch the proposed model. Hazy images are synthesized starting from haze-free images by using the image augmentation toolbox.[1] This toolbox generates synthetic samples exploiting an approximation of the atmospheric scattering model (Eq. 1). Figure 3 illustrates samples obtained by using the aforementioned augmentation. We randomly split the dataset into 17000 training images, and 125 validation images.

Dense-Haze: The Dense-Haze dataset [4] is the most recent dataset introduced in the NTIRE2019-Dehazing Challenge [3] for validating single-image dehazing algorithms. It contains images collected using a setup that included professional fog generators and a professional camera setup. It consists of 55 both indoor and outdoor high resolution (1600×1200 pixels) pairs of images affected by dense haze and their corresponding ground truth divided as follows: 45 training images, 5 validation images, and 5 test images.

3.2. Parameter settings

We implement the proposed method in Python 3.6 using the PyTorch package [28]. The proposed model is trained on a workstation equipped with an Intel I7-7700 CPU @ 3.60GHz, 16GB DDR4 RAM 2400MHz, NVIDIA Titan X Pascal GPU with 3840 CUDA cores. All the convolutional weights are initialized with the method proposed in [16], while all the biases are set equal to zero.

We train the model from scratch by using Adam optimizer with a fixed learning rate of 0.0001, a batch-size equal to 1, and a momentum term of 0.5 for a total of 30 epochs. Hyper-parameters for the fine-tuning of the model are the same as the training ones apart from the learning rate which corresponds to 1e-5 and the number of epochs equal to 150.

3.3. Comparison with state-of-the-art methods

In this section, we demonstrate the effectiveness of the proposed approach by estimating results on both paired synthetic images and unpaired real-world images. We compare

[1] https://imgaug.readthedocs.io/en/latest/

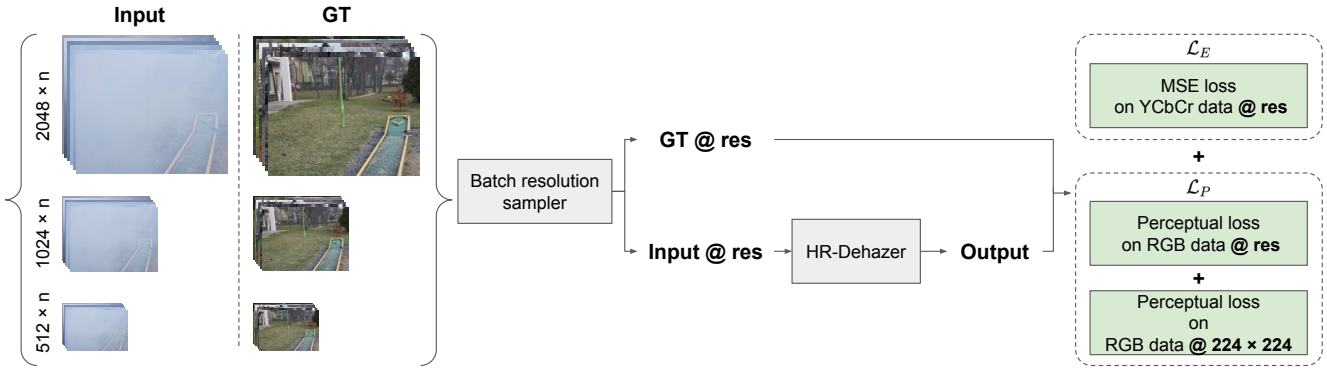


Figure 2. During training the images of each batch are randomly resized (short side randomly chosen among 512, 1024, and 2048 pixels). Hazy images are then processed using the HR-Dehazer. Finally, the loss is computed by linearly combining MSE at the randomly sampled resolution and perceptual loss at two resolutions (input as well as 224 × 224 pixels).

Figure 3. Examples of haze-free training and corresponding augmented images.

our results with other four recent methods in the state-of-the-art: DehazeNet by Cai *et al.* [7], Ren *et al.* [29], AOD-Net by Li *et al.* [20], and Zhang *et al.* [39]. We used the original code provided by the authors for comparing all the methods.

Evaluation on paired synthetic images The evaluation of dehazing algorithms on paired synthetic images is carried out by comparing the restored hazy images with the ground-truth which consists of a version of the image without haze. Commonly employed metrics are the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity (SSIM) [36] index. Figure 4 and Table 2 show the results of the comparison on the Dense-Haze validation set. It can be observed that the proposed method significantly outperforms the other considered approaches. From our personal standpoint, this is mainly justified by the nature of data used for training these methods. In fact, Dense-Haze training images present a very intense and high haze phenomenon, while datasets used for training state-of-the-art contains images with sparse haze. Table 3 reports the comparison on the test set of the O-Haze dataset [5] used for the NTIRE2018 Image Dehazing Challenge [1]. The proposed method trained on Dense-Haze data achieves comparable performance respect to Zhang *et al.* [39], which has been trained on O-Haze data.

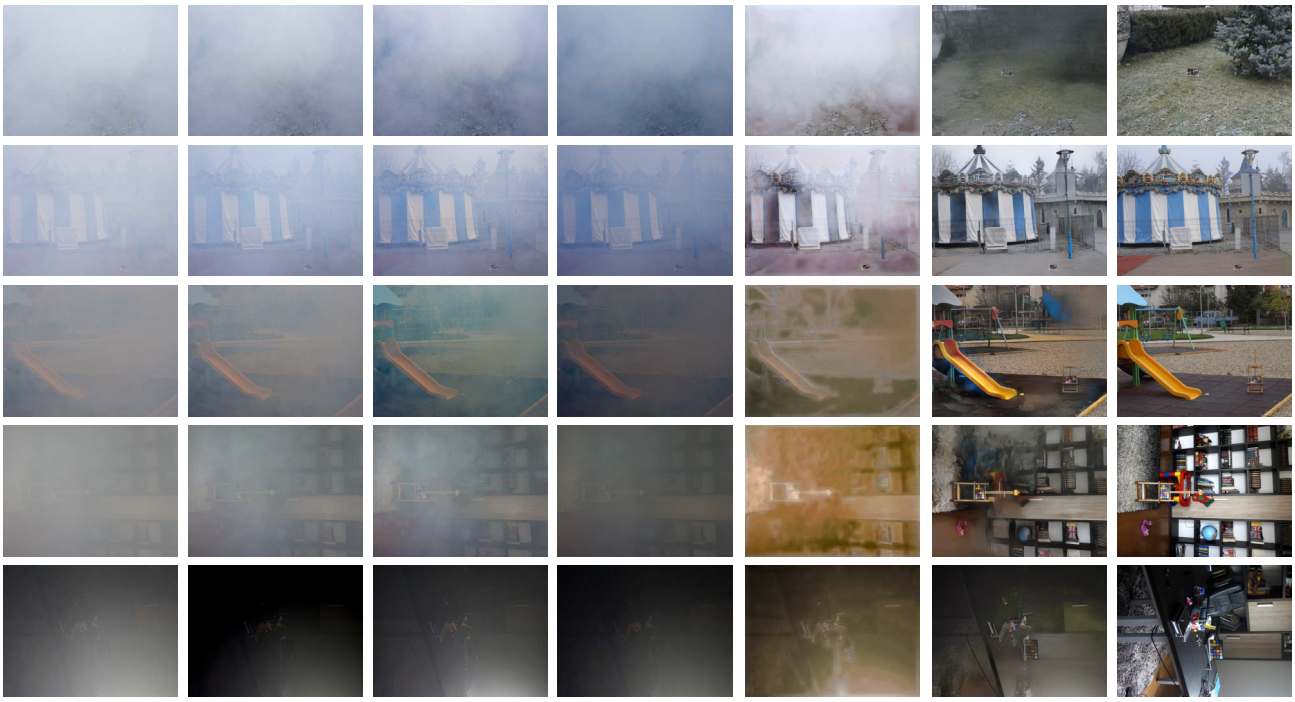
Method	PSNR	SSIM
DehazeNet [7]	10.87	0.3556
Ren *et al.* [29]	12.61	0.4196
AOD-Net [20]	12.79	0.4230
Zhang *et al.* [39]	12.04	0.4284
HR-Dehazer	**16.47**	**0.5181**

Table 2. Quantitative comparison with state-of-the-art methods on the Dense-Haze dataset [4]. For both the measures higher is better.

Method	PSNR	SSIM
DehazeNet [7]	16.65	0.6397
Ren *et al.* [29]	17.53	0.6773
AOD-Net [20]	15.46	0.6076
Zhang *et al.* [39]	24.03	0.7750
HR-Dehazer	**21.46**	**0.6912**

Table 3. Quantitative comparison with state-of-the-art methods on the test set of the O-Haze dataset [5] released for the NTIRE 2018 Image Dehazing Challenge - Track 2: Outdoor.

Evaluation on unpaired real-world images The HR-Dehazer is also evaluated on unpaired real-world images without any adaptation to the new domain. We considered images from two standard datasets, namely the IVC Waterloo dataset [23] and the HSTS subset of RESIDE dataset [21]. Figure 5 and Figure 6 show restoration results on real-world images belonging to both the datasets. It can be noticed that methods in the state of the art are generally able to reduce the haze, but at the same time they compromise image properties: for example, [7] and [20] darken some regions, while [29] distorts colors also of haze-free regions. The HR-Dehazer instead is able to recover almost



Input DehazeNet [7] Ren et al. [29] AOD-Net [20] Zhang et al. [39] Ours Ground-truth

Figure 4. Qualitative comparison of results on images from the Dense-Haze validation set [4].

completely the structure of images affected by dense haze without introducing artifacts. For example, it can be observed in the last row of Figure 6 that the color of the water in the fountain is preserved and simultaneously that the building with the arc in the upper part of the image is recovered using realistic colors. Finally, we want to highlight the effectiveness of the proposed solution by focusing on the clarity of the text in the image on the third row in Figure 6 and on the increased depth of view of the image of the highway (fourth row in Figure 5): cars are visible after the application of the proposed approach.

4. NTIRE2019-Dehazing Challenge

The HR-Dehazer was initially proposed in order to participate in the NTIRE2019-Dehazing Challenge. The quantitative performance results in terms of PSNR and SSIM on the test set of the Dense-Haze dataset, provided by the organizers [3], are reported in Table 4. As can be seen from the table, the proposed method places in the second half of the rank respect to the PSNR score, instead it obtains a SSIM of 0.05 smaller than the best method.

Team	User	PSNR	SSIM
iPAL-AtJ	moonriverLucy	20.26	0.657
iPAL-COLOR	DH-IRCNN_123_CEDH	19.92	0.653
MT.MaxClear	venkat2	19.47	0.652
BMIPL-UNIST-DW-1	Sprite+Ours	18.84	0.633
xddqm	Untitled Folder	18.52	0.640
ECNU	emmm+dpn_best	17.83	0.617
MOMOCV	meshpop	17.18	0.564
BMIPL-UNIST-DW-2	BMIPL-PDW+Hazing	16.86	0.610
BOE-IOT-AIBD	BOE_DH_Submission3	16.78	0.612
MAHA@IIT	akshay.aad16	16.47	0.548
FastNet	tzofi+submission	16.37	0.569
Our	**IVL+submission16**	**16.19**	**0.601**
ecsuiplab1	san_santra+up_4	16.15	0.564
IPCV IITM	maitreya_ipcv+final	16.13	0.595
shh	sunhee+res	16.05	0.562
ecsuiplab2	ranjanisi+ranjan	15.97	0.539
Alex_SDU	wang_cheng	15.94	0.557
hcilab	hcilab+final	15.12	0.580
IMag	dxllx+t88	14.93	0.555
XZSYS	ChuanshengWang	14.34	0.491
Vintage	jptarel+simple1	14.02	0.529

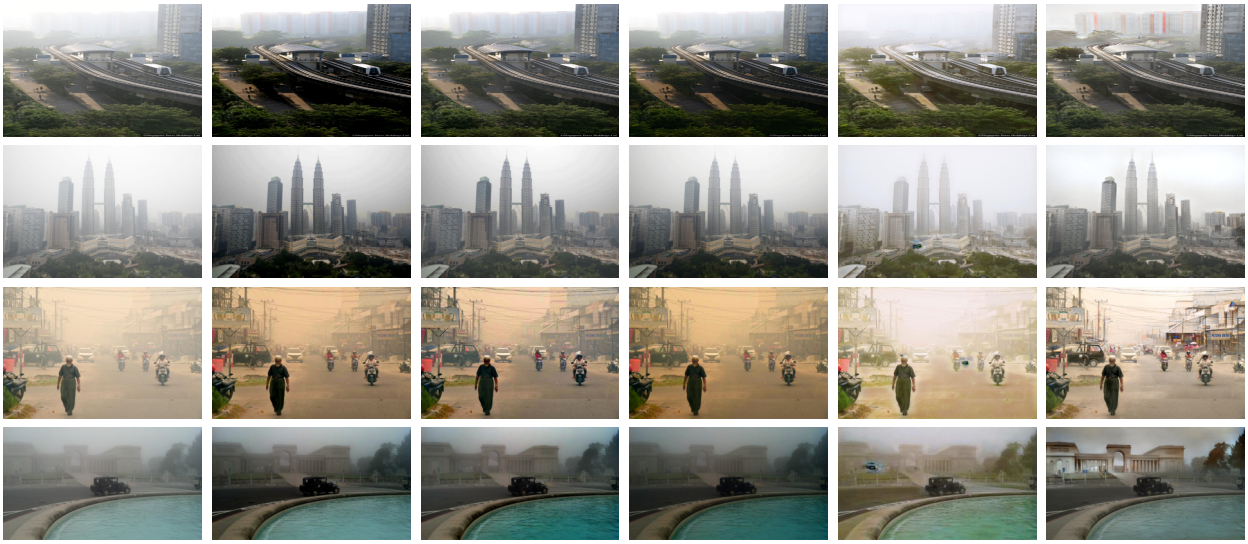
Table 4. Quantitative comparison in terms of PSNR and SSIM for the NTIRE-2019 Dehazing Challenge.

5. Conclusion

In this paper we proposed HR-Dehazer, a method for single image dehazing that learns a direct mapping between a hazy input image and the corresponding haze-free version. The proposed method treats differently the luminance information and the chroma components given that it processes a YCbCr version of the input hazy image. It consists of a fully end-to-end trainable encoder-decoder architecture. The training schema forces the dehazing network to consider the semantics of the input images and it has been de-

Input DehazeNet [7] Ren et al. [29] AOD-Net [20] Zhang et al. [39] Ours

Figure 5. Qualitative comparison of results on real-world images from the IVC Waterloo dataset [23].



Input DehazeNet [7] Ren et al. [29] AOD-Net [20] Zhang et al. [39] Ours

Figure 6. Qualitative comparison of results on real-world images from the HSTS subset of the RESIDE dataset [21].

signed so that it is scale-invariant. Specifically, during training, input images are rescaled. It has been shown that the proposed method outperforms current state-of-the-art methods over several datasets. Although it is a very successful method, there are still other aspects that need to be investigated. For example, the metric used to evaluate the genuineness of the results may promote more blurry results in favor of sharper results. Finally, a subjective study should be conducted to evaluate the perceptual quality of the obtained results.

Acknowledgment

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research. This work was supported by TEINVEIN, CUP: E96D17000110009 – Call "Accordi per la Ricerca e l'Innovazione", co-funded by POR FESR 2014-2020.

References

[1] Cosmin Ancuti, Codruta O Ancuti, and Radu Timofte. Ntire 2018 challenge on image dehazing: Methods and results.

In *Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W)*, pages 891–901. IEEE, 2018. 2, 5

[2] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 620–631. Springer, 2018. 2

[3] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte et al. Ntire 2019 challenge on image dehazing: Methods and results. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W)*. IEEE, 2019. 4, 6

[4] Codruta O. Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. *arXiv preprint arXiv:1904.02904*, 2019. 3, 4, 5, 6

[5] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W)*, pages 754–762, 2018. 2, 5

[6] Dana Berman, Tali Treibitz, and Shai Avidan. Single image dehazing using haze-lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 1, 2

[7] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 2, 5, 6, 7

[8] Ziang Cheng, Shaodi You, Viorela Ila, and Hongdong Li. Semantic single-image dehazing. *arXiv preprint arXiv:1804.05624*, 2018. 2

[9] Serhan Coşar, Giuseppe Donatiello, Vania Bogorny, Carolina Garate, Luis Otavio Alvares, and François Brémond. Toward abnormal trajectory and event detection in video surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3):683–695, 2017. 1

[10] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010. 4

[11] Raanan Fattal. Single image dehazing. *ACM Transactions On Graphics (TOG)*, 27(3):72, 2008. 2

[12] Adrian Galdran, Aitor Alvarez-Gila, Alessandro Bria, Javier Vazquez-Corral, and Marcelo Bertalmío. On the duality between retinex and image dehazing. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8212–8221. IEEE, 2018. 2

[13] Alona Golts, Daniel Freedman, and Michael Elad. Unsupervised single image dehazing using dark channel prior loss. *arXiv preprint arXiv:1812.07051*, 2018. 2

[14] Nicolas Hautière, Jean-Philippe Tarel, and Didier Aubert. Towards fog-free in-vehicle vision systems through contrast restoration. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2007. 2

[15] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1956–1963, June 2009. 2

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015. 4

[17] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 694–711. Springer, 2016. 3, 4

[18] Johannes Kopf, Boris Neubert, Billy Chen, Michael Cohen, Daniel Cohen-Or, Oliver Deussen, Matt Uyttendaele, and Dani Lischinski. Deep photo: Model-based photograph enhancement and viewing. *ACM Transactions on Graphics (TOG)*, 27(5), 2008. 2

[19] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971. 2

[20] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *International Conference on Computer Vision (ICCV)*, volume 1, page 7. IEEE, 2017. 5, 6, 7

[21] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 5, 7

[22] Yan Lu and Dezhen Song. Visual navigation using heterogeneous landmarks and unsupervised geometric constraints. *IEEE Transactions on Robotics*, 31(3):736–749, 2015. 1

[23] Kede Ma, Wentao Liu, and Zhou Wang. Perceptual evaluation of single image dehazing algorithms. In *International Conference on Image Processing (ICIP)*, pages 3600–3604. IEEE, 2015. 5, 7

[24] Earl J McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976. 1

[25] Srinivasa G Narasimhan and Shree K Nayar. Chromatic framework for vision in bad weather. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, page 1598. IEEE, 2000. 2

[26] Srinivasa G Narasimhan and Shree K Nayar. Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):713–724, 2003. 2

[27] Shree K Nayar and Srinivasa G Narasimhan. Vision in bad weather. In *International Conference on Computer Vision (ICCV)*, page 820. IEEE, 1999. 2

[28] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 4

[29] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision (ECCV)*, pages 154–169. Springer, 2016. 2, 5, 6, 7

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image com-*

puting and computer-assisted intervention, pages 234–241. Springer, 2015. 3, 4

[31] Ugo Rosolia, Stijn De Bruyne, and Andrew G Alleyne. Autonomous vehicle control: A nonconvex approach for obstacle avoidance. *IEEE Transactions on Control Systems Technology*, 25(2):469–484, 2017. 1

[32] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Polarization-based vision through haze. *Applied optics*, 42(3):511–525, 2003. 2

[33] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1874–1883. IEEE, 2016. 3

[34] Robby T Tan. Visibility in bad weather from a single image. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. 1, 2

[35] Anna Wang, Wenhui Wang, Jinglu Liu, and Nanhui Gu. Aipnet: Image-to-image single image dehazing with atmospheric illumination prior. *IEEE Transactions on Image Processing*, 28(1):381–393, 2019. 2

[36] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[37] Fengying Xie, Mengyun Shi, Zhenwei Shi, Jihao Yin, and Danpei Zhao. Multilevel cloud detection in remote sensing images based on deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(8):3631–3640, 2017. 1

[38] Xitong Yang, Zheng Xu, and Jiebo Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 2

[39] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Multi-scale single image dehazing using perceptual pyramid deep network. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPR-W)*, pages 902–911, 2018. 2, 5, 6, 7