Dense '123' Color Enhancement Dehazing Network

Tiantong Guo, Venkateswararao Cherukuri, Vishal Monga
The Pennsylvania State University, The Department of Electrical Engineering, University Park, PA, USA
tiantong@ieee.org, vmc5164@psu.edu, vmonga@engr.psu.edu

Abstract

Single image dehazing has gained much attention recently. A typical learning based approach uses example hazy and clean image pairs to train a mapping between the two. Of the learning based methods, those based on deep neural networks have shown to deliver state of the art performance. An important aspect of recovered image quality is the color information, which is severely compromised when the image is corrupted by very dense haze. While many different network architectures have been developed for recovering dehazed images, an explicit attention to recovering individual color channels with a design that ensures their quality has been missing. Our proposed work, focuses on this issue by developing a novel network structure that comprises of: a common DenseNet based feature encoder whose output branches into three distinct DensetNet based decoders to yield estimates of the R, G and B color channels of the image. A subsequent refinement block further enhances the final synthesized RGB/color image by joint processing of these color channels. Inspired by its structure, we call our approach the One-To-Three Color Enhancement Dehazing (123-CEDH) network. To ensure the recovery of physically meaningful and high quality color channels, the main network loss function is further regularized by a multi-scale structural similarity index term as well as a term that enhances color contrast. Experiments reveal that 123-CEDH has the ability to recover color information at early training stages (i.e. in the first few epochs) vs. other highly competitive methods. Validation on the benchmark datasets of the NTIRE'19 and NTIRE'18 dehazing challenges reveals the 123-CEDH to be one of the Top-3 methods based on results released in the NTIRE'19 competition.

1. Introduction

Photos captured by mobile devices has spiked up recently due to the availability of a fast and convenient photography experience [1]. On many occasions, the captured image may be hazy and lead to scene erosions when taken in unfavorable weather or environmental conditions. Because sophisticated optics is difficult to deploy on mobile devices such as cellphone cameras, a software or algorithmic processing of the hazy images is desired. Other application areas with similar constraints are fast emerging in autonomous driving and navigation.

Haze is caused by floating partials in the atmosphere which can scatter or absorb light. Hazing adversely affects not only captured image quality but also subsequent computer vision tasks such as detection and recognition [2] of objects in the scene. The effect of hazing may be mathematically characterized using the classical haze model [3]

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

where $\mathbf{I}$ is the observed hazy image, $\mathbf{J}$ is the true scene radiance, $\mathbf{A}$ is the ambient light intensity, $\mathbf{t}$ is the transmission map. Transmission map is a distance-dependent factor that affects the fraction of light that reaches the camera sensor. The transmission map can be expressed as $\mathbf{t} = e^{-\beta \mathbf{d}}$, where β represents the attenuation coefficient of the atmosphere and $\mathbf{d}$ is the scene depth. Most existing single image dehazing methods attempt to recover the clean image or scene radiance $\mathbf{J}$ based on the observed hazy image $\mathbf{I}$ via estimating $\mathbf{t}$ which is a well-known ill-posed problem.

Numerous studies have been done in the past to enhance the quality of hazy images [4, 5, 6, 7, 8, 9, 10, 11, 12] and they can be broadly categorized into multi-image dehazing and single image dehazing techniques. Multi-image dehazing methods rely on capture of the same underlying scene under different environmental conditions [13, 14]. The benefit of these approaches is that they do not require any explicit learning or precise knowledge of haze model parameters. But multiple captures of the same scene with the desired environmental diversity are rarely available. This has led to the emergence and popularity of single image dehazing [15, 16, 17]. Most existing single image dehazing methods attempt to recover $\mathbf{J}$ via the estimation of $\mathbf{t}$ by applying dark channel and other suitable priors [18, 19, 20, 21, 22, 23, 24, 25].

Deep learning methods have quickly supplanted the state of the art in image dehazing. A typical approach involves training of a deep neural network that requires example hazy and haze-free image pairs to learn a non-linear map-

Figure 1: A ground-truth/hazy image pair from NTIRE'19.

ping between them [26] or a mapping between the hazy image and physical parameters of the haze model [27]. This learned mapping is applied to a new image during the test phase to obtain a clean image. Owing to the difficulty of obtaining a reasonable number of real-world $\mathbf{I}$ and $\mathbf{J}$ pairs, often synthetic image pairs are designed for training deep networks [28]. A detailed review of literature related to the dehazing methods using deep learning frameworks is presented in Sec. 2.

Motivation and Contributions: In many real-world examples of hazy images, the haze is much denser than is observed in the widely used synthetic datasets. One such example is the NTIRE19 dataset illustrated in Fig. 1, wherein the images are covered by a very dense haze such that the scene of interest is almost completely obscured visually. The visual distortions include severe color loss, poor contrast and loss of structural information. Our proposed work addresses these challenges by developing a novel network structure that comprises of: a common DenseNet based feature encoder whose output branches into three distinct DensetNet based decoders to yield estimates of the R, G and B color channels of the image. A subsequent refinement block further enhances the final synthesized RGB/color image by joint processing of these color channels. Inspired by its structure, we call our approach the One-To-Three Color Enhancement Dehazing (123-CEDH) network. To ensure the recovery of physically meaningful and high quality color channels, the main network loss function is further regularized by a multi-scale structural similarity index term as well as a term that enhances color contrast. Finally, for stability, we employ a novel 2-stage training process in which we train an encoder with a single decoder followed by using this pre-trained encoder with the proposed 3-decoder architecture. The proposed 123-CEDH ranked in the Top 3 methods in the NTIRE'19 competition based on results released [29]. More detailed results and comparisons against state of the art dehazing methods are reported in Section 5.4 [1].

2. Related Work

Deep learning based method has been studied to solve the single image dehazing problem. Commonly, end-to-end Convolutional Neural Networks (CNN) are employed to learn a non-linear mapping between the input and the desired output. [30] is among the first methods that uses a deep learning model to generate the haze-free image from a single observation wherein a CNN is employed to estimate the transmission map $\mathbf{t}$ and then $\mathbf{A}$ is obtained based on the estimated $\mathbf{t}$. Following the footsteps of [30], [31] used a multi-scale CNN to further enhance the estimation of the transmission map $\mathbf{t}$ and $\mathbf{A}$. [2] learned the $\mathbf{t}$ and $\mathbf{A}$ jointly as one single parameter using a CNN.

Following the success of GANs [32] in synthesizing realistic images, in [26], the authors proposed an end-to-end dehazing method DCPDN to jointly learn the transmission map, atmospheric light and combine them to recover the dehazed image. The end-to-end learning is achieved by directly embedding the atmospheric scattering model into the network, thereby ensuring that the proposed method can strictly follow the physics driven scattering model for dehazing. Furthermore, to incorporate the mutual structure information between the estimated transmission map and the dehazed result, they also proposed a joint-discriminator based on GAN to decide whether the corresponding dehazed image and the estimated transmission map are real or fake. In [27], the authors present a multi-scale image dehazing method using Perceptual Pyramid Deep network based on the recently popular dense and residual blocks. This method involves an encoder-decoder structure with a pyramid pooling module in the decoder to incorporate contextual information of the scene into the network.

2.1. Color Information Orientated Dehazing

While the aforementioned methods focus on estimating the scenery information which follows a physical haze models, the limitation of these methods is clear when the amount of the haze presented in the image is much denser and the haze does not obey the underlying physical model. Moreover, at certain locations in the dense-haze image, original scenery information is not preserved and hence no meaningful observation can be made for estimation. This phenomena becomes more evident in the experimental results section, as we show that most of the existing state-of-the-art methods fail to recover a meaningful dehazed image given a dense-haze image as the input. As the scenery is heavily polluted with the air-light haze, the existing methods fails to recover the color information. Although [33] addressed this color distortion issue using a multi-stage CNN, their method fails to recover clean images with dense haze from real-world image datasets. Therefore, to recover true color information from the dense-haze images, we propose a method that employs a unique one-to-three network structure which is presented in detail in the following section.

3. Proposed 123-CEDH Network

Our proposed deep network consists of two major components: one encoder and three decoders. The encoder

[1] Code is available at the project page: http://signal.ee.psu.edu/research/CEDH.html

serves as a general feature extractor and the decoders are trained to recover the three color channels based on the features obtained from the encoder. Following the success of dense networks in various imaging applications that include dehazing [26], we build our encoder and decoders using dense blocks. The idea of sharing an encoder with multiple decoders has been shown to be powerful for several vision problems such as denosing, surface normalization, unsupervised 2D-segmentation, etc. [34]. A shared encoder makes sure that the extracted general features encompass geometric as well as color information from all the color channels, and then each customized decoder can recover the color in each of the individual R, G and B channels accurately. Further, to exploit inter-channel information, a multi-scale refinement block is employed as a (learnable) post processing operation. We then employ a compound loss function which consists of 4 different loss terms with each term designed to offer a complementary benefit.

3.1. Network Structure

The building blocks in 123-CEDH are:
1) Encoder: the encoder is constructed based on the Densely Connected Network (DCN) [35].
2) Decoder: the decoder has a similar structure as the encoder with more batch normalization layers.
3) Refinement blocks as suggested by [26] are used for merging the inter-color channel information.

As shown in Table 1, the encoder blocks from 'Base.0' to 'Dense.4' are initialized from [35] which is originally trained for image classification tasks. These pre-trained blocks serves as the feature extractor in the image classification network which has the ability to obtain useful features for other vision tasks. Other dense blocks in [35] are not utilized in the later half of the DCN as the features are mapped into a lower-dimension feature space for classification. To alleviate this issue for dehazing, we append the encoder with newly added blocks 'Trans.4' and 'Res.4' which enlarges the features generated by 'Dense.4'. This practice preserves more spatial information in the encoder which is then utilized by the decoder for further enhancements.

Table 2 gives the detailed structure of the decoder. The decoder is build to interpolate the extracted features from the encoder. As shown in the table, the decoder contains 4 'Trans' blocks, which stands for transformation block. The transformation block essentially takes the refined image/features into a reorder and enlargement process. The reordering process is accomplished by a 1×1 convolutional layer, and the enlargement is done by the upsampling layer. We further added new residual blocks [36] between two successive dense blocks to enhance the high-frequency information which leads to better details in the recovered image. More batch normalization layers are added to the dense blocks to normalize the training data so that the manifold of the network parameter will be more smooth and the network will have better training stability [37]. The bottom row of Table 2 details the structure of the refinement blocks as suggested by [27, 26]. These blocks first use average pooling layer at spatial size of 32×32, 16×16, 8×8, and 4×4 to extract local average information. Then the 1×1 convolutional layer reorganizes the inter-color channel information which is followed by an upsample layer to enlarge the image into the desired spatial size. These enlarged locally reorganized images are then appended together and passed trough the final refinement layer which uses a 3×3 convolutional layer to eliminate the blocking artifacts. This refinement practice would allow the image information to be merged and retouched at different scales.

With the aforementioned building blocks, the proposed 123-CEDH network structure is described in Table 3. The encoder in 123-CEDH is constructed as described in Table 1 and the Decoder.R, Decoder.G, and Decoder.B are constructed as mentioned in Table 2. R, G, and B represents the red, green, and blue channels in the RGB color space. The features generated by the encoder is densely connected to the decoder. As shown in Table 2, 'Dense.5' utilizes the concatenated outputs from 'Res.4' and 'Trans.2' to have better information flow as suggested in [35]. Similar connection is used in 'Dense.6' as well. The outputs of the Decoder.R,.G,.B are concatenated together and sent trough the refinement block which is described as Refine.10 to Refine.13 blocks in Table 2. The output layer uses a 3×3 conv. layer to further eliminate the blocking artifacts which might be generated by the refinement blocks and outputs the three color channel dehazed image. The representative diagram of the 123-CEDH is shown in Fig. 2.

3.2. Loss Function

Given the hazy image $\mathbf{I}$ and the ground-truth haze-free image $\mathbf{J}$, we intend to learn network parameters by minimizing a loss function that consists of 4 different components each of which is used for a specific purpose. The individual loss terms are described as follows:

Reconstruction Loss: This is a standard ℓ_2 loss function that is commonly used for regression problems and is mathematically defined as $\mathcal{L}_{\ell_2} = \|f(\mathbf{I}) - \mathbf{J}\|_2^2$ where the $f(\cdot)$ denotes the non-linear mapping function of 123-CEDH.

Perceptual Loss: The perceptual loss is used to control the overall image content agreement with the ground-truth image. To achieve this, the high-level features from the pre-trained VGG network are used. It is given by: $\mathcal{L}_{\text{vgg}} = \sum_{i=1}^{3} \|g_i(f(\mathbf{I})) - g_i(\mathbf{J})\|_2^2$ where the $g_i(\cdot)$ represents the features obtained from the pre-trained VGG-16 at layer ReLui_i, $i = 1, 2, 3$.

Multi-Scale Structural Similarity Index (MS-SSIM): Another important aspect of obtaining a clean image is to preserve its structure. Given the dense nature of the haze, no

Table 1: 123-CEDH Encoder Structure

	Base.0	Dense.1	Trans.1	Dense.2	Trans.2
Input	input patch/image	Base.0	Dense.1	Trans.1	Dense.2
Structure	$\begin{bmatrix} 7\times 7 \text{ conv.} \\ 3\times 3 \text{ max-pool} \end{bmatrix}$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ 2\times 2 \text{ avg-pool} \end{bmatrix}$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 12$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ 2\times 2 \text{ avg-pool} \end{bmatrix}$
Output	$64\times 64\times 64$	$64\times 64\times 256$	$32\times 32\times 128$	$32\times 32\times 512$	$16\times 16\times 256$
	Dense.3	Trans.3	Dense.4	Trans.4	Res.4
Input	Trans.2	Dense.3	Trans.3	Dense.4	Trans.4
Structure	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 24$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ 2\times 2 \text{ avg-pool} \end{bmatrix}$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 12$	$\begin{bmatrix} 3\times 3 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 2$
Output	$16\times 16\times 1024$	$8\times 8\times 512$	$8\times 8\times 768$	$16\times 16\times 128$	$16\times 16\times 128$

Table 2: 123-CEDH Decoder Structure

	Dense.5	Trans.5	Res.5	Dense.6	Trans.6	Res.6
Input	[Res.4, Trans.2]	Dense.5	Trans.5	[Trans.1, Res.5]	Dense.6	Trans.6
Structure	$\begin{bmatrix} \text{batch norm} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 7$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ \text{upsample 2} \end{bmatrix}$	$\begin{bmatrix} 3\times 3 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 2$	$\begin{bmatrix} \text{batch norm} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 7$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ \text{upsample 2} \end{bmatrix}$	$\begin{bmatrix} 3\times 3 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 2$
Output	$16\times 16\times 640$	$32\times 32\times 128$	$32\times 32\times 128$	$32\times 32\times 384$	$64\times 64\times 64$	$64\times 64\times 64$
	Dense.7	Trans.7	Res.7	Dense.8	Trans.8	Res.8
Input	Res.6	Dense.7	Trans.7	Res.7	Dense.8	Trans.8
Structure	$\begin{bmatrix} \text{batch norm} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 7$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ \text{upsample 2} \end{bmatrix}$	$\begin{bmatrix} 3\times 3 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 2$	$\begin{bmatrix} \text{batch norm} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 7$	$\begin{bmatrix} 1\times 1 \text{ conv.} \\ \text{upsample 2} \end{bmatrix}$	$\begin{bmatrix} 3\times 3 \text{ conv.} \\ 3\times 3 \text{ conv.} \end{bmatrix} \times 2$
Output	$64\times 64\times 128$	$128\times 128\times 32$	$128\times 128\times 32$	$128\times 128\times 64$	$256\times 256\times 16$	$256\times 256\times 16$
	Refine.9	Refine.10	Refine.11	Refine.12	Refine.13	Output.14
Input	[Input, Res.8]	Refine.9	Refine.9	Refine.9	Refine.9	[Refine.9.10.11.12.13]
Structure	3×3 conv.	$\begin{bmatrix} 32\times 32 \text{ avg-pool} \\ 1\times 1 \text{ conv.} \\ \text{upsample} \end{bmatrix}$	$\begin{bmatrix} 16\times 16 \text{ avg-pool} \\ 1\times 1 \text{ conv.} \\ \text{upsample} \end{bmatrix}$	$\begin{bmatrix} 8\times 8 \text{ avg-pool} \\ 1\times 1 \text{ conv.} \\ \text{upsample} \end{bmatrix}$	$\begin{bmatrix} 4\times 4 \text{ avg-pool} \\ 1\times 1 \text{ conv.} \\ \text{upsample} \end{bmatrix}$	3×3 conv.
Output	$256\times 256\times 20$	$256\times 256\times 1$	$256\times 256\times 1$	$256\times 256\times 1$	$256\times 256\times 1$	$256\times 256\times 1$

Table 3: 123-CEDH Structure

	Encoder	Decoder.R	Decoder.G	Decoder.B	Refine	Output
Input	Input	Encoder	Encoder	Encoder	Decoder.[R,G,B]	Trans.6
Structure	As in Table 1	As in Table 2	As in Table 2	As in Table 2	As in Refine.10-13	As in Output14 [2]

structure is preserved in the input images. Hence, to retain the structure, we employ a MS-SSIM loss function [38, 39]. It is given by: $\mathcal{L}_{\text{ms-ssim}}(x,y) = 1 - \text{MS-SSIM}(x,y)$ where $\text{MS-SSIM}(x,y) = l_M(x,y)\prod cs_M(x,y)$, x is a pixel in $f(\mathbf{I})$, y is a pixel in $\mathbf{J}$, and M is the total number of dyadic pyramid levels of image decomposition and $l(x,y)$, $cs(x,y)$ are defined as: $l(x,y) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_{y^2}+C_1}$, $cs(x,y) = \frac{2\sigma_x\sigma_y+C_2}{\sigma_x+\sigma_y+C_2}$ where μ and σ are the mean and standard deviations of a patch surrounding a given pixel.

Contrast Enhancement Loss: to improve the color contrast of the output images generated by the three decoders, we maximize the variance of each individual color channel as described in [40] and is given by $\mathcal{L}_{\text{ce}} = \sqrt{\frac{1}{N}\sum_{i\in\{R,G,B\}}\sum_{x=1}^{N}\left(f(\mathbf{I})_x^i - \bar{f}(\mathbf{I})^i\right)^2}$ where x denotes the pixel index of the image and the total number of pixels is denoted by N. $\bar{f}(\mathbf{I})$ denotes the average pixel value of the output image $f(\mathbf{I})$.

The **overall loss function** is defined as:

$$\mathcal{L} = \mathcal{L}_{\ell_2} + \alpha\mathcal{L}_{\text{vgg}} - \beta\mathcal{L}_{\text{ce}} + \gamma\mathcal{L}_{\text{ms-ssim}} \tag{2}$$

where α, β, and γ are positive regularization constants. Note that the desired output dehazed image should have its contrast enhanced, thus the term $\mathcal{L}_{\text{ce}}$ needs to be maximized, hence a negative sign before it. Further, all the loss terms described above are differentiable and hence can be incorporated into a deep learning framework.

[2]The 3×3 conv. layer now generates 3 channel outputs.

4. Dataset, Training, and Test Procedure

4.1. Datasets

To train 123-CEDH, we primarily use the NTIRE2019-Dehaze dataset [41]. The images were collected using a setup that included professional fog generators and a professional camera setup, so as to capture the same scene with and without haze. The training data consists from 45 hazy images (both indoor and outdoor) and their corresponding ground truth images.

Further, to obtain a network that is more diverse, we also utilized the NTIRE2018-Dehaze dataset [42]. Comparing to NTIRE2018 dataset, the haze is much denser in the NTIRE2019 dataset. We developed a synthetic method to thicken the haze present in NTIRE2018 training images to mimic the haze level in NTIRE2019. We generated the synthetic dense-haze image following the practice demonstrated in [30]. The synthetic dense-haze image is generated using the inverse haze model $\mathbf{I}_{syn} = \mathbf{I}\cdot\mathbf{t} + (1-\mathbf{t})\cdot\mathbf{A}$ where $\mathbf{I}_{syn}$ is the synthetic dense-haze image, $\mathbf{I}$ is the

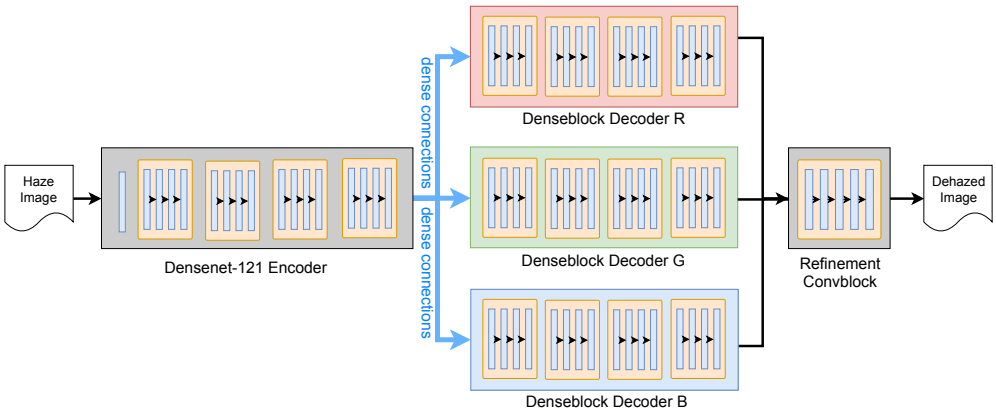


Figure 2: The proposed 123-CEDH network structure. There is one DenseNet based encoder to extract general image features, and three decoders to recover 3 color channels. The final refinement block exploits the inter-color channel information for further enhancement. A detailed architecture of both the encoder and decoders is listed in Table 1 and 2. The overall connection strategy is listed in Table 3

(a) Haze free image (b) NTIRE18 hazy image (c) Synthetic denser haze

Figure 3: NTIRE18 image *36-outdoor.png*(upper) and *27-indoor.png*, the synthetic dense-haze image adds more haze based on the NTIRE18 hazy image.

NTIRE2018 hazy image. $\mathbf{A}$ is selected to be a fixed value to reduce the uncertainty in variable learning. For indoor images, we set $\mathbf{A} = 0.6$, and for outdoor images we set $\mathbf{A} = [0.80, 0.81, 0.86]$[3] to mimic a blueish atmosphere light which is estimated by using method described in [22]. The $\mathbf{t}$ is selected uniformly between 0.01 and 0.3 to add denser haze as smaller $\mathbf{t}$ yields more atmosphere information. These values of $\mathbf{t}$ are selected by a cross-validation method during the development phase. Fig. 3 demonstrates the synthetic dense-haze images. To obtain a sizable amount of training set, we extract patches of size 512×512 from the training dataset. To further diversify the training samples, the following data augmentation techniques are used: **1**) horizontal flip, rotation by $90°$, $180°$, and $270°$; **2**) scale to 0.7, 0.8, and 0.9 of the original image size.

For the inference, we feed the complete hazy image $\mathbf{I}$ at once to obtain the dehazed image $f(\mathbf{I})$.

4.2. Training

Learning the complete network at once is challenging and can result in training instability. Hence, we employed a two stage training strategy:

Stage 1 - Pre-training of Encoder: We first pre-train the encoder by combining it with only a single decoder. The network architecture for this stage is same as described in Tables 1 and 2. In this stage, we used the NTIRE19 and synthetic dense-haze images from NTIRE18 to train the encoder. We train the network for 80 epochs in this stage.

Stage 2 - color enhancement training: In this stage we use the encoder trained from stage 1 and combine it with 3 new decoders to enhance the color information, as shown in Table 3. For the first 50 epochs we use the same training data as in Stage 1, and for the later 70 epochs we only use the training data from NTIRE19.

α, β, and γ are set to 0.3, 0.1, and 0.01 respectively by cross-validation [43]. During the training, Adam optimizer [44] is used with initial learning rate of 1×10^{-3}. The learning rate is degraded to its 70% every 35 epochs. The learning rate is reset when the training enters stage 2.

4.3. Optional Post-processing

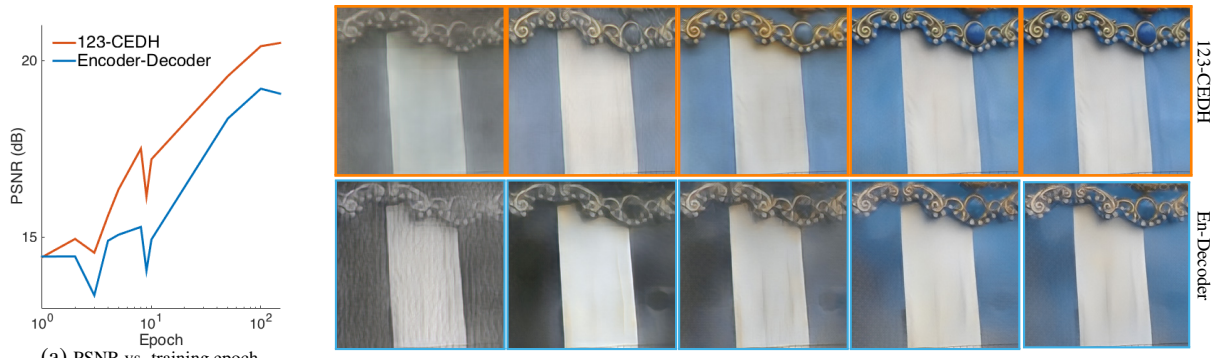
We also used the IRCNN [45] denosier with $\sigma = 15$ to further improve the results visually. IRCNN method combines the benefits of model based techniques and learning based techniques for image restoration applications. For this dehazing problem, we use the pre-trained CNN denoiser and incorporate it as a post processing unit after the output obtained by our proposed 123-CEDH framework. We use 123-CEDH+ to indicate that the post processing is used. Detailed results of 123-CEDH and 123-CEDH+ is listed in Section 5.3.

5. Experimental Results

5.1. Color Enhancement

The key merit of the proposed network is that it can enhance the color information at an early training stage. As we discovered during the training, the network usually learns the luminance information at the first few epochs, and then gradually adds color information stage by stage. By using the proposed '123' structure, the network starts to recover color information while simultaneously learning the lumi-

[3] $\mathbf{A}$ is set to 0.80, 0.81, and 0.86 for RGB channels, respectively.



(a) PSNR vs. training epoch (b) Ep. 1 (c) Ep. 5 (d) Ep. 10 (e) Ep. 100 (f) Ep. 150

Figure 4: The PSNR vs. training epoch plot and the validation results at different epoch from proposed method and the Encoder-Decoder structure. As it shown in Fig. 4b to 4f, the proposed 123-CEDH generates better color details at early training stage than the Encoder-Decoder structure which only uses one decoder to recover all the color channels.

Table 4: The validation PSNR and SSIM score using different loss and training configurations.

Loss Terms			Stages		PSNR	SSIM
$\mathcal{L}_{vgg}$	$\mathcal{L}_{ssim}$	$\mathcal{L}_{ce}$	1	2		
✓	✓	✓	✓		15.48	0.46
✓	✓	✓		✓	16.24	0.48
✓			✓	✓	16.81	0.50
✓	✓		✓	✓	16.98	0.51
✓		✓	✓	✓	16.94	0.50
	✓	✓	✓	✓	16.92	0.51
✓	✓	✓	✓	✓	17.10	0.52

nance component from the early epochs.

Fig. 4 illustrates a sample of the proposed method comparing to dense CNN dehazing networks. The Encoder-Decoder methods marked by blue line is essentially a similar structure to [26] which is the winner in NTIRE18 dehazing contest. As can be observed, the proposed 123-CEDH is indeed generating more vivid color earlier than the existing encoder-decoder structure.

5.2. Ablation Study

The effects of different configurations of the proposed 123-CEDH are investigated in this section. Table 4 reports the results of our proposed method with different combinations of individual loss terms described in Section 3.2. First, the performance difference between 1 decoder and 3 decoders architecture is close to 0.75db (15.48 to 16.24) which is significant and thus validating our proposed network architecture. Second, the benefit of two-stage training process is readily apparent as we observe significant amount of performance gains. Further, training with the $\mathcal{L}_{ms\text{-}ssim}$ improves the SSIM score during the test and combining the VGG perceptual loss term $\mathcal{L}_{vgg}$ with the contrast enhancement loss $\mathcal{L}_{ce}$ improves both the PSNR and SSIM score, and produces visually pleasing images (see Fig. 4).

5.3. Comparison with State-of-the-art Methods

In this section, detailed comparisons w.r.t the state-of-the-art methods on real-world benchmark data sets I-HAZE [46], and O-HAZE [47, 42] are reported.

State-of-the-art Methods The state-of-the-art methods included in the comparisons are: CVPR'09 [48, 19], TIP'15 [49], ECCV'16 [31], TIP'16 [30], CVPR'16 [23], ICCV'17 [2], CVPR'18 [27], and CVPRW'18 [26].

Evaluation Datasets The comparisons are conducted on the I-HAZE (indoor-haze) and O-HAZE (outdoor-haze) validation dataset [42]. Each of the dataset contains 5 pairs of haze and haze-free image pairs. Detailed acquisition methods of these real-world haze/haze-free image pairs are discussed in [42].

Fig. 5 and 6 demonstrate the state-of-the-art methods comparing with 123-CEDH over NTIRE2018 indoor and outdoor validation datasets. As can be observed, 123-CEDH generated visually pleasing results compared to the most recent developed methods. As shown in Table 5 and 6, 123-CEDH outperforms other state-of-the-art methods. If the post processing procedure described in Section 4.3 is used, the final dehazed images generated by 123-CEDH+ have the highest scores.

5.4. NTIRE-2019 Dehazing Challenge

For the newly published NTIRE2019-Dehaze dataset, the haze presented in the images are much denser than normal images in the literature. As shown in Fig. 7, the state-of-the-art methods' performances degraded heavily due to the amount of haze covering essentially all the scenery information. As we discussed that 123-CEDH can enhance the color information using special designed network structure and the compound losses, the dehazed images generated by 123-CEDH are more visually pleasing. We compute the quantitative performance on the NTIRE2019 validation set since the ground-truth images are made available [41]. As shown in Table 7, 123-CEDH outperforms other state-of-the-art methods. If the post processing procedure described in Section 4.3 is used, the final dehazed images generated by 123-CEDH+ have the highest scores.

Figure 5: The visual results of NTIRE2018-indoor validation dataset.

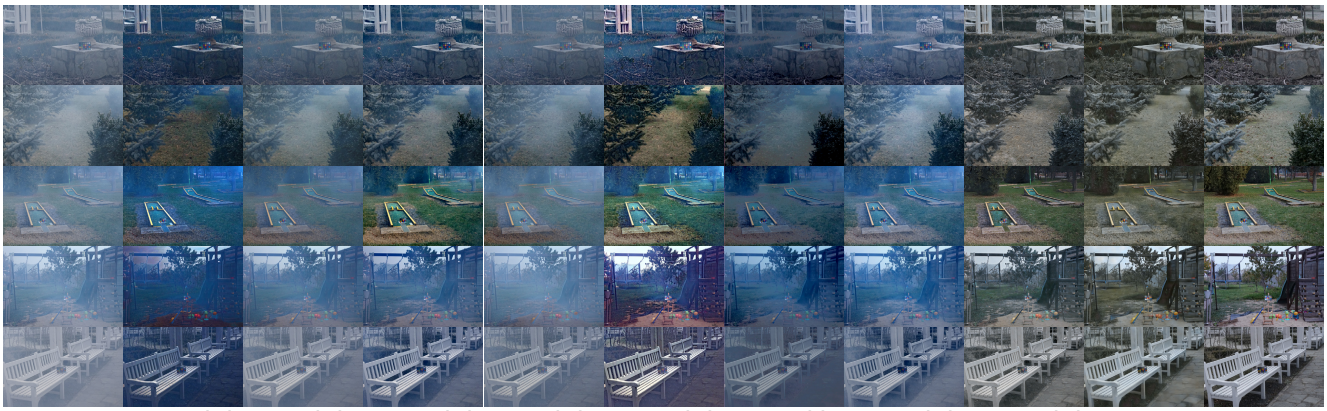


Figure 6: The visual results of NTIRE2018-outdoor validation dataset.

Figure 7: The visual results of NTIRE2019 validation dataset.

In Table 8 we include the top-6 methods from the contest. It can be observed that 123-CEDH and 123-CEDH+ are among the top performing methods in the NTIRE2019-Dehazing Challenges.

6. Conclusion

In this work, we develop a novel network structure that consists of one encoder and three decoders to yield estimates of the R, G and B color channels of the image. We further added a subsequent refinement block further enhances the final synthesized RGB/color image by joint processing of these color channels. The One-To-Three Color Enhancement Dehazing (123-CEDH) network ensures the recovery of physically meaningful and high quality color channels with regularized loss function by a multi-scale structural similarity index term as well as a term that enhances color contrast. 123-CEDH has the ability to recover

Table 5: The PSNR/SSIM of different methods over NTIRE2018-indoor validation dataset.

method	26.*png*	27.*png*	28.*png*	29.*png*	30.*png*	avg.
CVPR09 [48]	8.2706/0.3545	12.8863/0.2014	12.7162/0.5485	12.1518/0.5411	13.4688/0.2753	11.8988/0.3842
TIP15 [49]	13.1816/0.6581	16.6858/0.3952	11.5135/0.5590	17.1496/0.7803	15.7567/0.3215	14.8574/0.5428
TIP16 [31]	10.1699/0.5498	14.5147/0.3094	13.3890/0.6349	11.9041/0.5369	15.5312/0.3412	13.1018/0.4744
CVPR16 [30]	12.4147/0.4800	14.7990/0.3639	13.2925/0.5489	14.6639/0.5296	13.9293/0.4057	13.8199/0.4656
ICCV17 [2]	10.8313/0.6185	16.8387/0.3943	12.7391/0.4692	15.3688/0.8054	17.2741/0.3095	14.6104/0.5194
CVPR18 [27]	15.3106/0.6283	16.0856/0.3512	9.8470/0.5540	22.2085/0.8013	15.4517/0.1977	15.7807/0.5065
CVPRW18 [26]	14.2680/0.6778	20.8952/0.7533	18.4479/0.6983	20.5845/0.8154	16.4299/0.5445	18.1251/0.6978
123-CEDH	22.1640/0.8921	23.8423/0.8600	19.5356/0.8281	23.3367/0.9144	19.7775/0.8085	21.7312/0.8606
123-CEDH+	22.2046/0.9088	23.9359/0.8767	19.5797/0.8455	23.3583/0.9309	19.8056/0.8284	**21.7768/0.8781**

Table 6: The PSNR/SSIM of different methods over NTIRE2018-outdoor validation dataset.

method	36.*png*	37.*png*	38.*png*	39.*png*	40.*png*	avg.
CVPR09 [48]	18.1820/0.4474	16.0912/0.4983	14.1227/0.0835	12.8787/0.3575	14.2106/0.3864	15.0970/0.3546
TIP15 [49]	17.4660/0.4976	16.1686/0.4533	15.1391/0.1796	14.7964/0.4131	16.3732/0.5683	15.9887/0.4224
TIP16 [31]	16.5891/0.4862	15.7593/0.4334	13.2500/0.1890	12.7816/0.3935	16.5339/0.5597	14.9828/0.4123
CVPR16 [30]	16.9236/0.4267	14.9854/0.4776	15.5448/0.3390	17.6496/0.4751	17.0424/0.5350	16.4292/0.4507
ICCV17 [2]	17.0951/0.4516	16.4676/0.3886	16.1153/0.1194	15.0439/0.3388	15.9477/0.5043	16.1339/0.3606
CVPR18 [27]	17.1374/0.4385	15.2847/0.4173	14.6555/0.1143	15.2353/0.3530	17.7805/0.5198	16.0187/0.3686
CVPRW18 [26]	24.6703/0.7288	22.4079/0.6551	23.7469/0.7199	21.9055/0.6296	22.2878/0.6822	23.0037/0.6831
123-CEDH	24.8369/0.7982	23.8385/0.7341	24.8175/0.7910	22.4016/0.7562	25.9663/0.8005	24.3722/0.7760
123-CEDH+	24.8714/0.8102	23.8462/0.7323	24.9090/0.7970	22.4053/0.7626	25.9656/0.8039	**24.3995/0.7812**

Table 7: The PSNR/SSIM of different methods over NTIRE2019 **validation** dataset.

method	46.*png*	47.*png*	48.*png*	49.*png*	50.*png*	avg.
CVPR09 [48]	10.2792/-0.0030	14.7181/0.3669	14.1058/0.2821	10.4352/0.2112	11.0434/0.3764	12.1163/0.2467
TIP15 [49]	9.3752/0.0187	16.1040/0.4152	13.4858/0.2282	10.9173/0.2735	11.2999/0.2600	12.2364/0.2391
TIP16 [31]	8.1189/0.0432	12.8189/0.3496	12.3575/0.2172	10.5432/0.2403	10.5051/0.2119	10.8687/0.2124
CVPR16 [30]	9.5131/0.0130	13.9106/0.3744	12.9015/0.2796	10.1643/0.2208	11.6160/0.3645	11.6211/0.2505
ICCV17 [2]	10.1190/-0.0160	14.1507/0.2835	12.1667/0.1804	8.4235/0.2104	12.8605/0.3342	11.5441/0.1985
CVPR18 [27]	8.0126/0.0288	13.0654/0.3498	13.5508/0.1945	11.0016/0.2758	14.0006/0.4475	11.9262/0.2593
CVPRW18 [26]	8.8668/0.2272	15.1229/0.3635	15.6723/0.4603	11.8703/0.3592	15.0124/0.4305	13.3090/0.3681
123-CEDH	15.2626/0.2604	23.0794/0.7480	18.4653/0.5383	13.5508/0.4768	15.1661/0.5589	17.1048/0.5165
123-CEDH+	15.2992/0.2744	23.1300/0.7567	18.5113/0.5650	13.5577/0.4863	15.1646/0.5905	**17.1326/0.5346**

Table 8: The average PSNR/SSIM of top methods over NTIRE2019 **test** dataset.

team	contest method	PSNR	SSIM
ours	123-CEDH+	**19.923**	**0.653**
	123-CEDH	19.882	0.633
other teams	method1	19.469	0.652
	method2	18.521	0.640
	method3	18.387	0.630

color information at early training stages vs. other competitive methods. Validation on the real-world datasets of NTIRE'18 and NTIRE'19 dehazing challenges reveals 123-CEDH to be one of the most competitive methods.

References

[1] L. Gye, "Picture this: The impact of mobile camera phones on personal photographic practices," *Continuum*, vol. 21, no. 2, pp. 279–288, 2007.

[2] B. Li et al., "Aod-net: All-in-one dehazing network," in *Proc. IEEE Conf. on Comp. Vis.*, 2017.

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

[4] C. O. Ancuti, C. Ancuti, and C. D. Vleeschouwer, "Effective local airlight estimation for image dehazing," in *Proc. IEEE Conf. on Image Proc.*, 2018, pp. 2850–2854.

[5] C. O. Ancuti et al., "Color transfer for underwater dehazing and depth estimation," in *Proc. IEEE Conf. on Image Proc.*, 2017, pp. 695–699.

[6] C. O. Ancuti et al., "Locally adaptive color correction for underwater image dehazing and matching," in *Proc. IEEE Conf. on Image Proc.*, 2017, pp. 1–9.

[7] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion authors," *IEEE Trans. on Image Proc.*, vol. 22, no. 8, pp. 3271–3282, 2013.

[8] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Effective single image dehazing by fusion," in *Proc. IEEE Conf. on Image Proc.*, 2010, pp. 3541–3544.

[9] C. Ancuti et al., "Night-time dehazing by fusion," in *Proc. IEEE Conf. on Image Proc.*, 2016, pp. 2256–2260.

[10] W. Ren et al., "Gated fusion network for single image dehazing," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2018, pp. 3253–3261.

[11] C. Chen, M. N. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *Proc. IEEE European Conf. on Comp. Vision*. Springer, 2016, pp. 576–591.

[12] R. Li et al., "Single image dehazing via conditional generative adversarial network," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2018, pp. 8202–8211.

[13] S. G. Narasimhan and S. K. Nayar, "Chromatic framework for vision in bad weather," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2000, vol. 1, pp. 598–605.

[14] Z. Li et al., "Simultaneous video defogging and stereo reconstruction," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2015, pp. 4988–4997.

[15] J. P. Oakley and B. L. Satherley, "Improving image quality in poor visibility conditions using a physical model for contrast degradation," *IEEE Trans. on Image Proc.*, vol. 7, no. 2, pp. 167–179, 1998.

[16] S. G. Narasimhan and S. K. Nayar, "Interactive (de) weathering of an image using physical models," in *Proc. IEEE Workshop Color and Photometric Methods in Comp. Vision*. France, 2003, vol. 6, p. 1.

[17] N. Hautière et al., "Towards fog-free in-vehicle vision systems through contrast restoration," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2007, pp. 1–8.

[18] R. Fattal, "Single image dehazing," *ACM Trans. on Graphics*, vol. 27, no. 3, pp. 72, 2008.

[19] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. on Patt. Analysis and Machine Int.*, vol. 33, no. 12, pp. 2341–2353, 2011.

[20] C. O. Ancuti, C. Ancuti, and P. Bekaert, "Effective single image dehazing by fusion," in *Proc. IEEE Conf. on Image Proc.*, 2010, pp. 3541–3544.

[21] K. Tang et al., "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2014.

[22] R. Fattal, "Dehazing using color-lines," *ACM Trans. on Graphics*, vol. 34, no. 1, pp. 13, 2014.

[23] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2016, pp. 1674–1682.

[24] C. O. Ancuti, C. Ancuti, and C. De Vleeschouwer, "Effective local airlight estimation for image dehazing," in *Proc. IEEE Conf. on Image Proc.*, 2018, pp. 2850–2854.

[25] C. O. Ancuti et al., "A fast semi-inverse approach to detect and remove the haze from a single image," in *Proc. IEEE Asian Conf. on Comp. Vision*, 2010, pp. 501–514.

[26] H. Zhang, V. Sindagi, and V. M. Patel, "Multi-scale single image dehazing using perceptual pyramid deep network," in *Proc. IEEE Conf. Workshop on Comp. Vis. Patt. Recog.*, 2018, pp. 902–911.

[27] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2018, pp. 3194–3203.

[28] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-hazy: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. IEEE Conf. on Image Proc.*, 2016, pp. 2226–2230.

[29] C. O. Ancuti, C. Ancuti, and R. T. et al., "Ntire 2019 challenge on image dehazing: Methods and results," in *Proc. IEEE Conf. Workshop on Comp. Vis. Patt. Recog.*, 2019.

[30] B. Cai et al., "Dehazenet: An end-to-end system for single image haze removal," *IEEE Trans. on Image Proc.*, vol. 25, no. 11, pp. 5187–5198, 2016.

[31] W. Ren et al., "Single image dehazing via multi-scale convolutional neural networks," in *Proc. IEEE European Conf. on Comp. Vision*. Springer, 2016, pp. 154–169.

[32] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Advances in Neural Information Proc. Systems*, 2014, pp. 2672–2680.

[33] A. Dudhane and S. Murala, "Cˆ2MSNet: A novel approach for single image haze removal," in *Winter Conf. on Applications of Computer Vision*, 2018, pp. 1397–1404.

[34] A. R. Zamir et al., "Taskonomy: Disentangling task transfer learning," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2018.

[35] G. Huang et al., "Densely connected convolutional networks," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2017, pp. 4700–4708.

[36] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. on Comp. Vis. Patt. Recog.*, 2016, pp. 46–54.

[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[38] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, 2003, vol. 2, pp. 1398–1402.

[39] H. Zhao et al., "Loss functions for image restoration with neural networks," *IEEE Trans. on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[40] J.-H. Kim et al., "Single image dehazing based on contrast enhancement," in *Proc. IEEE Int. on Conf. Acoustics, Speech, and Signal Proc.*, 2011, pp. 1273–1276.

[41] C. O. Ancuti et al., "Dense haze: A benchmark for image dehazing with dense-haze and haze-free images," *arXiv preprint arXiv:1904.02904*, 2019.

[42] C. Ancuti, C. O. Ancuti, and R. Timofte, "Ntire 2018 challenge on image dehazing: Methods and results," in *Proc. IEEE Conf. Workshop on Comp. Vis. Patt. Recog.*, 2018, pp. 891–901.

[43] V. Monga, *Handbook of Convex Optimization Methods in Imaging Science*, Springer, 2017.

[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[45] K. Zhang et al., "Learning deep cnn denoiser prior for image restoration," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2017, pp. 3929–3938.

[46] C. Ancuti et al., "I-haze: a dehazing benchmark with real hazy and haze-free indoor images," in *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2018, pp. 620–631.

[47] C. O. Ancuti et al., "O-haze: a dehazing benchmark with real hazy and haze-free outdoor images," in *Proc. IEEE Conf. Workshop on Comp. Vis. Patt. Recog.*, 2018, pp. 54–62.

[48] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Proc. IEEE Conf. on Comp. Vision Patt. Recog.*, 2009.

[49] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. on Image Proc.*, vol. 24, no. 11, pp. 3522–3533, 2015.