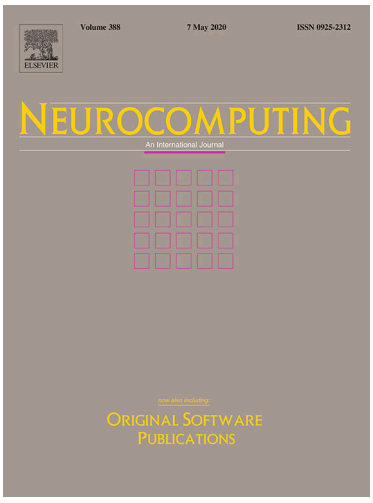
Journal Pre-proofs



NLDN: Non-local Dehazing Network for Dense Haze Removal

Shengdong Zhang, Fazhi He, Wenqi Ren

PII: S0925-2312(20)31012-2
DOI: https://doi.org/10.1016/j.neucom.2020.06.041
Reference: NEUCOM 22458

To appear in: *Neurocomputing*

Received Date: 20 January 2019
Accepted Date: 9 June 2020

Please cite this article as: S. Zhang, F. He, W. Ren, NLDN: Non-local Dehazing Network for Dense Haze Removal, *Neurocomputing* (2020), doi: https://doi.org/10.1016/j.neucom.2020.06.041

NLDN: Non-local Dehazing Network for Dense Haze Removal

Shengdong Zhang[a,b], Fazhi He[a,*], Wenqi Ren[b]

[a]*School of Computer Science, Wuhan University, Wuhan, China*
[b]*State Key Laboratory of Information Security (SKLOIS), IIE, CAS*

Abstract

Single image dehazing is one of the most challenging and important tasks in computer vision and image processing. In this paper, we propose a Non-local Dehazing Network (NLDN), which learns the mapping between hazy images and haze-free images. Our network architecture consists three components: the first is full point-wise convolutional part, which extracts Non-local statistical regularities; the second is feature combination part, which learns the spatial relation of statistical regularities; the third is reconstruction part, which recovers the haze-free image by the features extracted from the second part. By using these three components, we obtain a high quality dehazing result. Experimental results show that our method performs favorably against other state-of-the-art methods on both synthetic dataset and real-world images.

Keywords: non-local dehazing, deep learning, image restoration.

1. Introduction

Hazy image often shows low contrast and poor visibility since the light reflected from scene objects is attenuated in the air and further blended with the atmospheric light scattered by some particles before it reaches the camera. Image dehazing has received a lot of attention in recent years due to its wide application in media systems. All the systems assume the inputs are clean, so removing haze from input will boost the performance of media systems.

However, single image dehazing is massively ill-posed since the degradation depends on the unknown depth which varies at different positions. To solve this problem, many prior methods provide a solution by including more information such as multiple images of the same scene or depth information [1, 2, 3, 4].

Single image dehazing has made dramatically progress in past decades. We broadly divide these methods into two groups: prior-based methods and learning-based method. Prior-based methods solve the problem by utilizing visual cues or statistical properties of haze-free images [8, 9, 10, 11, 12, 13]. He

*Corresponding author

(a) input (b) Cai [5] (c) Li [6] (d) Ren [7] (e) Our result

Figure 1: Visual comparison on real-world dense hazy image. The result of Cai *et al.*'s method [5] exists color distortion. The results generated by Li *et al.* [6] and Ren *et al.* [7] still have most residual haze. In contrast, our algorithm performs well in dense hazy situation, the result looks more pleasant.

et al. [14] propose a dark channel prior based on the static of haze-free images that most local outdoor clear image patches contain some pixels whose intensity is very low in at least one color channel. Meng *et al.* [10] improve the dark channel prior using a more general case-boundary constraint. In [15], Fattal propose a novel method based on the prior that pixels in a local patch of haze-free images form a line in RGB color space. Berman *et al.* [12] propose a haze-line prior based on the fact that a clear image can be represented by a few hundred distinct colors. Some methods are designed for special cases [16]. All the aforementioned methods can generate a high quality image when the prior holds. However, these methods cannot generate a plausible result when the prior breaks.

Recently, learning-based methods have achieved high performance by using machine learning, convolutional neural networks and large-scale synthetic training datasets [17, 18, 19, 20, 21]. In [22], Tang *et al.* seek a best combination of local maximum contrast, dark channel prior, hue disparity and local maximum saturation using forest regression to estimate transmission map. A multi-scale fusion scheme [23] for single image dehazing is proposed. Zhu *et al.* [11] propose a linear model to create the relation between the scene depth of the hazy image under color attenuation prior and solve the model with a supervised manner. Convolutional neural networks [24, 25] also have been applied in single image dehazing. In [5], Cai *et al.* propose a novel CNN-based method to estimate the relation between the hazy image and transmission map. Ren *et al.* [26] propose a multi-scale CNN to solve the problem of transmission map estimating. Inspired by recently End-to-End deep learning approaches for image restoration and enhancement [27, 28, 29], four fully End-to-End methods [6, 7, 25, 30] are proposed. In [6], Li *et al.* propose an End-to-End method for single image dehazing via introducing a new variable. Ren *et al.* [7] introduce a network fusing the three preprocessed images to generate the final haze-free image. Zhang *et al.* [30] propose a network which can jointly optimize transmission map, atmospheric light and also image dehazing. Although these methods have achieved End-to-End dehazing, we find that they can't deal heavy hazy image well as shown in Figure 1.

To address above limitations, we propose a novel Fully Convolutional Dehazing Networks to exploit global context information in single image haze removal.

We first use point-wise convolutions to extract non-local features, we call this part of network as independent extractor and the output as non-local features. Second, we exploit the relation in non-local features using dilated convolution. Finally, we reconstruct the haze-free image from the fusion features. The contributions of this paper can be summarized as follows:

- We present a novel fully convolutional neural network for single image dehazing, refer to as NLDN, which optimizes the end-to-end pipeline from hazy images to clean ones without pre- and post- processing or intermediate variables estimation step.
- We use point-wise convolution to extract an intermediate representation from hazy image, which can reduce the input variation of model and make the process of train easy to converge.
- In order to further improve the discriminative of non-local feature, we propose a Non-local loss, which reduces the intra-color variations.
- We conduct extensive experiments to quantitatively and qualitatively compare our method with the state-of-the-art single image dehazing algorithms and demonstrate the effectiveness of the proposed method.

2. Related work

2.1. Atmospheric Scattering Model

The density of haze depends on the distance between the scene and the camera, which can be expressed as [8]:

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

where $I(x)$ denotes the observed hazy image, $J(x)$ is the corresponding haze-free scene radiance to be recovered, A represents the global atmospheric light, and $t(x)$ is the transmission map expressed as:

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

where β is the scattering coefficient of the atmosphere, and $d(x)$ denotes the distance between the object and the camera. We can infer the corresponding clear image by inversing Eq. 1

$$J(x) = \frac{1}{t(x)} I(x) - A\frac{1}{t(x)} + A. \tag{3}$$

Given the atmospheric scattering model, the key point is solving the atmospheric light and transmission map from the hazy image.

2.2. Separately Optimize Approaches

Based on the model 1, numerous methods [31, 32, 8, 9, 10, 33, 15, 22, 11, 12, 5, 26, 34, 35, 36, 37, 38] try to solve the single image dehazing task via estimating atmospheric light and transmission map separately, then get the final dehazing result via model 3.

By comparing hazy and clear images, Tan [31] finds that a haze-free image always has high contrast. Based on this observation, Tan removes haze by maximizing the contrast. However, this method always generates color distortion in the dehazing result. Fattal proposes a method [32] based on the albedo which is a constant vector, the transmission and the clean image are assumed to be independent in a local patch. Although this method produces a natural dehazing result, it would loss effect in the dense haze regions. Based on the statistical property of clear image patches, He *et al.* propose a new prior [14] that in a local patch at least one pixel has a low pixel intensity. However, this prior is not always valid and the dehazing result tends to over estimate the density of haze. In [15], Fattal proposes a novel dehazing method based on the observation that pixels in a local clean patch will form a line in RGB color space. Zhu *et al.* propose a color attenuation prior [11] and learn a linear function to predict the depth from a hazy image.

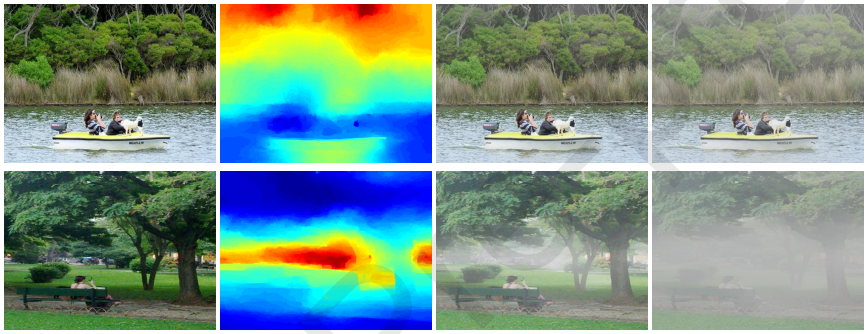
Recently, deep learning based methods have been proposed to solve this problem. Cai *et al.* propose a CNN-based method [5] which learns effective features for transmission map estimating. Ren *et al.* propose a multi-scale network [26] for effective transmission map estimating.

Despite the remarkable progress, the aforementioned methods are limited by the same reason of estimating transmission map and atmospheric light in a separated manner. Therefore, these approaches often are less effective especially for the dense haze images.

2.3. Jointly Optimize Methods

Traditional methods adopt the separate fashion of estimating atmospheric light and transmission map for single image dehazing independently. Recently, some End-to-End methods are proposed for single image dehazing [6, 7, 25, 30, 39] based on CNNs. By fusing the transmission map and atmospheric light into a new variable, Li *et al.* propose an End-to-End method for dehazing [6]. Zhang *et al.* propose a novel dehazing method [25] based on the dilation convolutional neural network and skip connection. A Densely Connected Pyramid Dehazing Network (DCPDN) is proposed by Zhang *et al.* [30], which jointly optimizes the atmospheric light, transmission map and dehazing all together. In [7]], Ren *et al.* use a gated network by fusing the White Balance , Contrast Enhancing and Gamma Correction of the input hazy image. All of these methods try to recover clear images using the End-to-End fashion, However, these methods still cannot handle heavy hazy image as shown in Figure 1. We find that Zhang *et al.*'s method still has to estimate atmospheric light, transmission map and dehazing result. Further more, this method needs resize, crop and different pre- and post-processing to make it accept different size. Ren *et al.*'s method still has to derive

three inputs from an original hazy image by applying White Balance, Contrast Enhancing and Gamma Correction. LAP-Net [40] is a level-aware network for dehazing, which recovers the clean image under the guidance of haze level. GridDehazeNet [41] is an attention-based dehazing network, which does not depend on the atmosphere scattering model. EPDN [42] models the dehazing problem as an image-to-image translation problem. FFA-Net [43] designs a feature attention for dehazing. FD-GAN [44] employs generative adversarial networks with fusion-discriminator for single image dehazing. Guo *et al.* [45] propose a fusion based dehazing network. In contrast, we propose a fully End-to-End method in this paper by using independent and recombination of features extracted by point-wise convolution, which does not need any pre- and post-processing.



(a) Haze-free (b) Depth (c) Light (d) Heavy

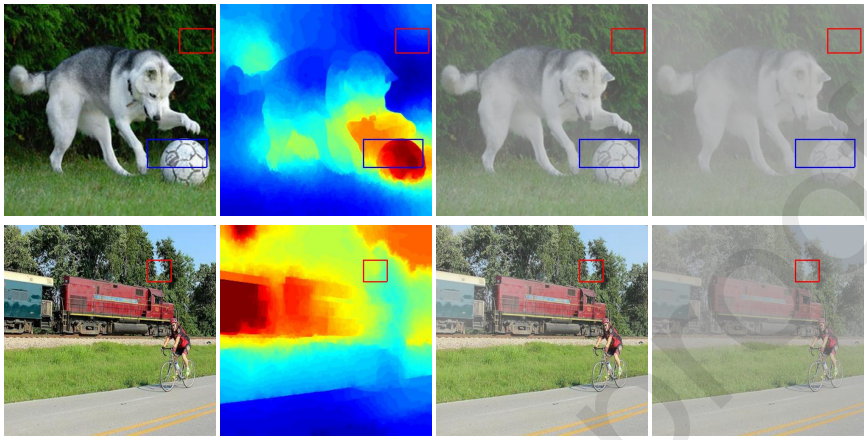
Figure 2: Some examples of our training data preparation. (a) Haze-free image. (b) The corresponding depth. (c) the corresponding light hazy image. (d) the corresponding heavy hazy image.

3. Proposed method

In this section, we will describe the training data preparation method firstly, which motives us to design a robust method for dehazing with noise label. Secondly, we describe our method, including the motivation, network architecture, loss function as well as network training. We show our network architecture in Figure 4, which includes independent extract sub-network, feature recombination sub-network and clean image reconstruct sub-network.

3.1. Training Data Preparation

CNN benefits from large-scale training data, which leads high performance of learning-based methods [46, 28]. However, it is hard to collect a large-scale data to train a model for image dehazing. Cai *et al.* propose a method for preparation of training data based on two assumptions: 1) the context of image patch has no relation with transmission map; 2) transmission map of a local image patch is constant. Therefore, Cai *et al.* assume that an individual image



(a) Haze-free (b) Depth (c) Light (d) Heavy

Figure 3: Problem of training data preparation. (a) Haze-free image. (b) The corresponding depth. (c) the corresponding light hazy image. (d) the corresponding heavy hazy image. We find that the depth top logic has been changed in first row. We also find that the sky near tree tend to have similar depth with the tree in second row.

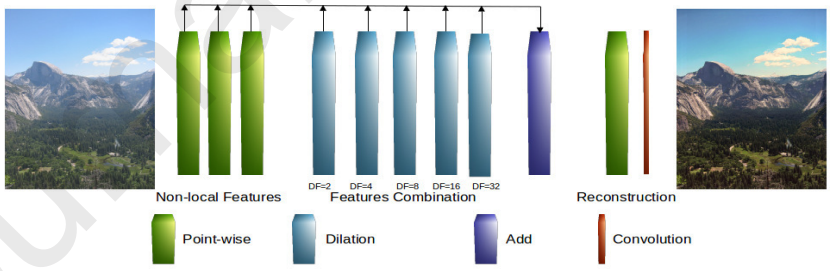


Figure 4: The framework of the proposed NLDN. The first three layers are used to extract non-local features. In order to eliminate the ambiguity of clean pixels, we impose the spatial information via dilation convolution layers. Then all the features from the first part and second part are collected and added to formulate the input of reconstruction subnetwork. The output of third layer is treated as non-local features, which is optimized by non-local loss.

patch could have an arbitrary transmission map. Based these assumptions and model 2, Cai *et al.* use a random atmospheric light A and a random transmission t to synthesize a hazy image patch when given a haze-free patch. Although this method generates a large-scale training dataset, this synthesized data is inaccurate since the training data do not consider the relationship between the transmission map and the depth.

Recently, Ren *et al.* [26] use the NYU-Depth dataset to synthesize a training dataset, which is more accurate according to scene depth. However, this method cannot generalize to outdoor images well since the training dataset is consisted of indoor images, especially for the hazy images with large scene depth.

To overcome this problem, we propose a method to generate an outdoor training dataset based the estimated depth by Liu *et al.*'s method [47]. We follow the pipeline proposed by Ren *et al.*[26]. First, we collect haze-free images from the SYSU-Scene Dataset,which is a parsing database including elaborately annotated objects. Second, we estimate the depth map for each image in our collected dataset using [47]. Finally, we generate a transmission map using model 2 and randomly choose an atmospheric light $A = [k, k, k]$, where $k \in [0.7, 1]$ to synthesize a hazy image.

Given a haze-free image $J(x)$ in our dataset and the corresponding depth map estimated by [47], we generate a hazy image using the model 1. To generate a high quality dataset, we choose the $\beta \in [0, 0.4]$. We do not use big $\beta \in [0.4, \infty]$, which will generate very small transmission. Therefore, we have 12,000 hazy images and transmission maps (1,200 images $\times$ 10 medium extinction coefficients β) in the training set.

We show some examples of the synthetic training image in Figure 2. As shown, both the light hazy and heavy hazy images look realistic. Although our method could synthesize decent training data for our algorithm, there are some problems need to be solved in the training stage since the imperfection of the depth estimation method [47]. Figure 3 shows some problems of [47]. The first one is that the result by [47] cannot keep the ordinal depth. As shown, the football should be farther than the dog and the background. The second problem is lost of detailed edge information as shown in red area in second row. Therefore, we propose a point-wise sub-network in Section 3.3 to increase the independence of pixels in hazy image to overcome the aforementioned problems.

3.2. Motivation

As illustrated in Section 3.1, our training data have some inaccurate regions, which will affects the performance of deep learning-based methods. To overcome the degradations caused by inaccurate depth estimation, we propose a point-wise sub-network and non-local loss to make the pixels belonged to same haze-line shares similar features and increase the flexibility of the network. The goal of the point-wise sub network is to ignore the influence of spatial information. The goal of non-local loss is to make the pixels belonged to same haze-line cluster to features center. In other words, the point-wise sub-network is to extract an intermediate representation for hazy pixels, which reduces the variation of input for feature combination subnetwork.

Haze-line has been proved effect for dehazing, which assumes colors of a haze-free image are well approximated by a few hundred distinct colors, that form tight clusters in RGB space. However, in the presence of haze each color cluster in the clear image becomes a line in RGB space, which is termed as haze-line. We can express the relation between hazy and clear pixel as follows:

$$H = C * t + A * (1 - t), \tag{4}$$

where the C denotes clear pixel, the t represents transmission and the A denotes the air-light. We can see that the hazy pixels will form a line in RGB space, We can express the line as:

$$C = (H - A)/t + A \tag{5}$$

Based on the Eq.(5), we can see that the hazy and clear pixel can formulate a haze-line, which can be well represented by a clear pixel. we use 1×1 filter to ignore the spatial relation between hazy pixels, and make same clear pixels share similar features:

$$O = w * I + b, \tag{6}$$

where w denotes filter size 1×1, b represents bias, and we call it as point-wise convolution. In order to achieve this goal, we propose non-local loss to make the features of same color cluster to the center represented by color. Conventional deep learning methods [48] use kennel size of 3×3 for the first layer, which will consider the relation of neighbor pixels in an input image. However, inaccurate depth values between neighbor pixels result in degradation of dehazing performance. To reduce the influence of the inaccurate depth estimation, we use point-wise convolution, which makes our model only consider the dependent depth values at the first layer. Then, we use a feature combination layer to learn the high-level information from a larger region, *i.e.*, the output of the independent extractor sub-network is the input of feature recombination sub-network.

The estimation of Eq. 6 is pixel-based, without imposing spatial coherency. Based on the Eq. 4, we can see that some hazy colors can be projected to several haze-line. In order to eliminate this ambiguity, we design a feature combination subnetwork. Since the feature combination sub-network considers the relation between neighbor pixel in feature space, large contextual information will help the network to eliminate the ambiguity of colors. Existing deep learning methods [48] use pooling layer to increase respective field of the network. However, it has been proved that spatial resolution of the input would lose after pooling layers [49]. Although progressive pooling achieves successful performance in semantic segment, the loss of spatial information may affect natural images understanding and significantly interfere other tasks that involve spatially detailed image understanding. To overcome this problem, we seek to use dilation convolution layer to increase the respective field of the network. Dilation convolution layers with different dilation rates have been proved effectively captures multi-scale information for image processing [49]. However, we find that as the

dilation rate becomes larger, less weights can be activated. We demonstrate this effect in Figure 5. An extreme case is that the dilation rate value is equal to the feature map size, the convolution filter with size of $3 * 3$. Instead of taking the whole image feature, the dilation convolution degrades to a simple $1 * 1$ filter which only considers the center pixel of the image. Therefore, we use small dilation rate and consider the relationship between neighboring pixels in feature space.

(a) (b[])

Figure 5: The respective field of large dilation rate. The left image shows that the convolution filter can capture the image structure well. When the dilation rate becomes large out of image size, we find its activated weights becomes less as shown in right image.

3.3. Network Architecture

Based on the motivation in Section 3.2, we propose a novel end-to-end deep learning model for single image dehazing, which is shown in Figure 4 with eleven layers. Our model consists of convolution and Leaky ReLU operations, which makes our model more unified than traditional dehazing model. We use point-wise convolution operation in our model, which has two different functions. The first one used in independent feature extracting layers is to eliminate the influence of inaccurate depth values. The second one used in reconstructing layer is to fuse the high-level features from feature combination layers. The large receptive field of filters makes the model captures more contextual information. However, existing models increase the receptive field via pooling or increasing the layers. We find that pooling operation will break the unify of learning model and increase the layers and result in high computation cost. In contrast, we use dilation convolution layer to increase the respective field for our proposed model. Shallow layers have more detailed information while the deeper ones have high-level or structure information, we use dense skip connection to collect the effective feature from all output of dilation convolution.

In addition, we propose a multi-scale dilation sub-network to further improve the dehazing effect. Tang *et al.* have proven that the multi-scale features are

effective for single image dehazing [22]. They compute multi-scale features for an input image at different spatial scales. In [5], Cai *et al.* propose a multi-scale sub-network via parallel convolutional operations, the kernel size is among 3×3, 5×5 and 7×7, and the filter number is the same for all three scales. Ren *et al.* [26] propose a multi-scale dehazing network to generate a haze-free image, which first generated a coarse-scale transmission map based on the entire image and later refined it locally. Inspired by these successes of multi-scale for dehazing, we apply multi-scale dilation sub-network to improve the dehazing quality. First, we extract features from the independent features using five dilation layers with rate equal to 2, 4, 8, 16 and 32, respectively. Then we collect all outputs of shadow convolution layers, then feed it to the reconstructing sub-network.

In the reconstructing sub-network, we consider how to utilize the outputs from the second sub-network. First, we extract a features from the output of second sub-network using point-wise convolution. Secondly, we predict the final dehazing result using convolution with kernel size of $3*3$.

3.4. Non-local Loss

In order to improve the non-local features extraction, we propose a non-local loss, which can be used to reduce the variation of intra-color features. Thus, the key is to define the number of pixel colors in clean image. However, the color space contains 256^3 colors. To reduce the number of colors needed to process, we first quantize each color channel to have 51 different values, which reduces the number of colors to 51^3. The output NF of third convolution layer is defined as non-local features and its dimension is $h*w*c$ and $nf_i(x,y)$ takes as input the coordinates (x,y) of NF, and outputs the vector of non-local feature, i is determined by the color kind in clean image.

The non-local loss can be formulated as:

$$\mathcal{L}_{\mathcal{N}} = \sum_{i=1}^{n} ||nf_i(x,y) - C_i||^2 , \tag{7}$$

the C_i denotes the center of i-th color cluster of non-local features, which is updated as the deep features changed, nf represents the features for a clean pixel in features. This formulation effectively reduces the intra-color variations, which makes our model easy to converge and recover a colorful dehazing result. We define the update equation of C_i are computed as:

$$C_i = C_i + \alpha\Delta C_i \tag{8}$$

where we set $\alpha = 0.95$. we define ΔC_i as:

$$\Delta C_i = \frac{\sum_{i=1}^{m} \delta(J(x,y)=i)(nf_i(x,y) - C_i)}{1 + \sum_{i=1}^{m} \delta(J(x,y)=i)} \tag{9}$$

where $\delta(condition) = 1$ if the condition is satisfied, otherwise $\delta(condition) = 0$.

It should be pointed out that similar colors in hazy image need have a similar appearance in dehazed result. The non-local loss provides this similarity in

features spaces, which also can be seen as features supervision learning. Pixels have similar radiance colors come from objects, located over the entire image. It's hard to consider while image context information by only employing network architecture, we provide a non-local loss that collects entire image context information based on the color similarity.

3.5. Loss Function

In this subsection, we describe our loss function in detail. Training a model with the Euclidean loss (L_2 loss) often blurs the dehazing result. In order to overcome this problem, we use L_1 loss to train our network which sharpens the final results. We define $L1$ loss as follows:

$$\mathcal{L}_{\mathcal{E}}(\tilde{J}, J) = \left|\left|\tilde{J} - J\right|\right|, \tag{10}$$

where $\tilde{J}$ denotes the predicted result, J represents the ground truth haze-free image.

To further improve the dehazing quality, we propose a new loss based on image gradient. Xu *et al.* propose a deep edge-aware filter using image gradients [50], which shows the effectiveness of image gradients in low-level problems. We define the novel gradient loss (GL), which can be combined with the $L1$ loss to boost the dehazing quality. We define our gradient loss function as following:

$$\begin{aligned}\mathcal{L}_{\mathcal{G}}(\tilde{J}, J) = \sum_{i,j} &\left|\left|\tilde{J}_{i,j-1} - \tilde{J}_{i,j}\right| - |J_{i,j-1} - J_{i,j}|\right| + \\ &\left|\left|\tilde{J}_{i-1,j} - \tilde{J}_{i,j}\right| - |J_{i-1,j} - J_{i,j}|\right|\end{aligned} \tag{11}$$

Therefore, our overall loss function is:

$$\mathcal{L}(\tilde{J}, J) = \mathcal{L}_{\mathcal{E}}(\tilde{J}, J) + \lambda_1 \mathcal{L}_{\mathcal{G}}(\tilde{J}, J) + \lambda_2 \mathcal{L}_{\mathcal{N}}, \tag{12}$$

where we use λ_1 and λ_2 to control the importance of the gradient and non-local loss.

3.6. Implementation Details

We generate a training dataset using the method proposed in section 3.1. It has been shown that identity initializer [49] is better than Gaussian random variables for dilation layers. We initialize the weights of dilation layers using identity initializer. In addition, we initialize the weights of the left layers with Gaussian random variables. It has been shown that *LeakyReLU* neuron is more effective than the *ReLU*, we utilize *LeakyReLU* neuron as activation function. We use batch size of 1 and patch size of 500×500. The λ_1 and λ_2 are 1.0 and 0.01, respectively. During training, we use ADAM as the optimization solver and learning rate 0.0001 to train our network. Our NLDN can be converged around 56 epochs as the solid black line shown in Figure 6. As shown in Figure 6, we can see that NLDN although has higher loss, the generalization ability of NLDN is better than the other network architecture since NLDN can deal with inaccurate areas in synthetic dataset well.

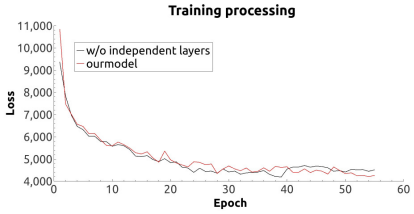
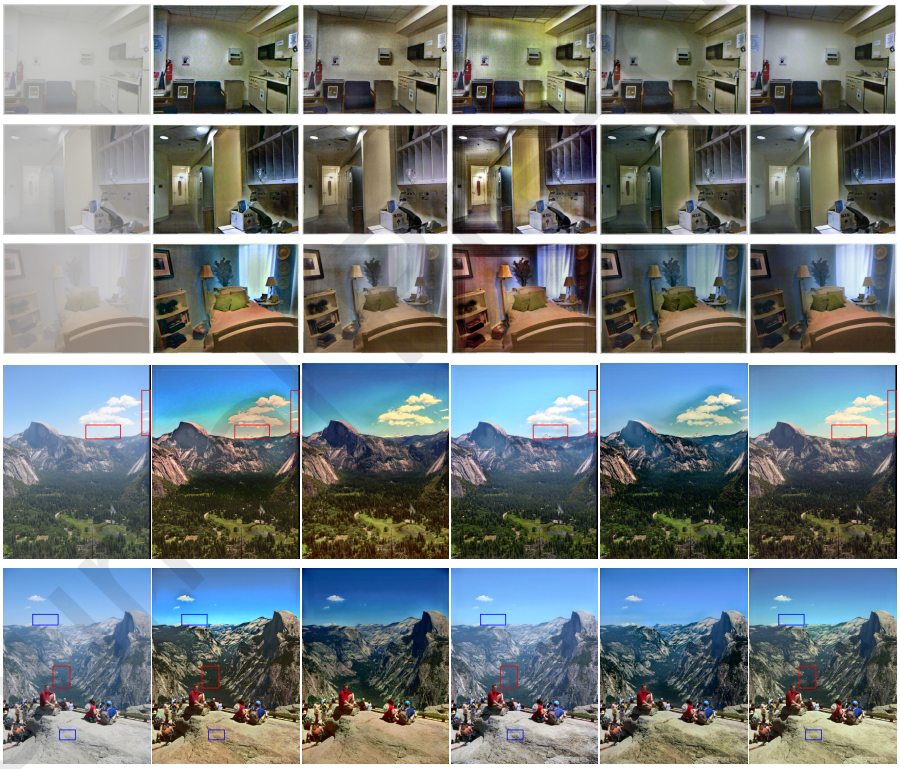


Figure 6: The training loss with different configurations.



(a) (b) (c) (d) (e) (f)

Figure 7: Visual comparisons using different network configurations on the hazy images. (a) Input hazy image. (b) Results obtained by model trained with Euclidean loss. (c) Results obtained by model trained with $L1$ loss. (d) Results obtained by model trained w/o point-wise module. (e) Results obtained by model trained w/o Non-local loss module. (f) Results obtained by model trained with full model.

Table 1: Quantitative comparisons using different network configurations on the HIFA dataset.

	w/o independent	Euclidean loss	w/o non-local loss	Ours
PSNR	17.89	17.49	19.00	20.20
SSIM	0.734	0.731	0.803	0.827

4. Analysis and Discussion

To demonstrate the improvement obtained by the proposed network, we perform an ablation study involving the following three experiments: 1) Independent extractor layers, 2) Gradient Loss, 3) Non-local Loss. We generate a testing dataset using the method in Section 3.1 with different outdoor images. We show the comparison results in Table 1 and use PSNR and SSIM as metrics to show the performance of our network.

4.1. Effectiveness of Independent Extractor Layers

In this subsection, we show the improvement using independent extractor layers. To show the improvement, we design two models. The first model is the first three convolution kernels are of $1 * 1$ size and the other is the first three convolution kernels are of $3*3$ size. Due to the existing of point-wise convolution layers, our network extracts a feature which services as a bridge between hazy image and haze-free image. The point-wise convolution layers make our model ignore the relation between pixels in first three layers, which results in our model more flexible.

However, the first three $3 * 3$ convolutional kernels will consider the relation between the adjacent pixels, the noise in training data will affect the performance of the proposed model. As shown in Table 1, our model with independent layers has high performance in terms of PSNR and SSIM. We also compare the model configurations according to visual results in Figure 7. As shown, the results without the independent extractor layers often have some color distortions. Such as the red boxes in the fourth row in Figure 7. In contrast, our model generates visual pleased results without color distortions.

In Figure 6, we show the training losses with and without the independent extractor layers. We can see that these two models have similar loss at the beginning of the training process. However, the model with the independent extractor layers has the lower loss in the ending of the training process. As can be seen in Figure 7(d), the results by the model without the independent extractor layers have some blurry artifacts around the boundaries. Instead, the model with the independent extractor layers often shows sharper and clearer results.

4.2. Effectiveness of the Gradient Loss

In this subsection, we analyze the effectiveness of the gradient loss function. We use the same network architecture with different training losses (Euclidean and gradient losses) to show the performance of the proposed network. As shown in Table 1, our model using the gradient loss (w/o non-local loss) achieves performance gain of 1.5dB in terms of PSNR than the model using Euclidean

Figure 8: Visual comparison on synthetic hazy images.

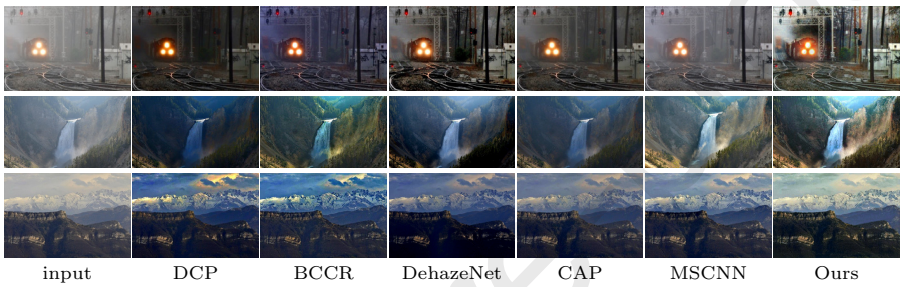


Figure 9: Visual comparison on real-world images.

loss. We also find that the visual results using Euclidean loss have some blur artifacts as shown in the red boxes of Figure 7(b) on the "Yosemite" image. With the help of gradient and L_1 losses, our model keeps the details well. For indoor heavy hazy image, we can see that gradient and L_1 losses also perform well than Euclidean loss. The results of using the gradient and L_1 losses often present more details.

4.3. Effectiveness of the Non-local Loss

In this subsection, we analyze the effectiveness of the non-local loss, which reduces the artifacts shown in Figure 7(e). The goal of the non-local loss is to reduce the inter-color variation. We first train the network without using the non-local loss. When training with 40 epochs, we set the weight of non-local loss as 0.01. As shown in Table 1,our model trained with non-local loss has high performance. Figure 7 also shows that the results with non-local loss look more pleased.

5. Experimental Results

To demonstrate the effectiveness of the proposed method, we conduct various experiments on synthetic datasets and a natural real-world dataset that contain a variety of hazy conditions, with comparisons to the state-of-the-art methods in terms of accuracy and visual effect.

(a) input (b) AOD-Net (c) FEED-Net (d) DCPDN (e) GFN (f) Ours

Figure 10: Visual comparison on real-world images.

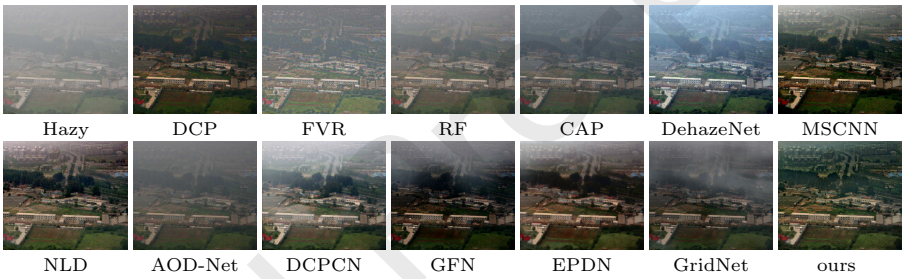


Figure 11: Qualitative evaluations on the real hazy images. The proposed method generates much clearer images with clearer structures and characters.

5.1. Synthetic Image Dehazing

In order to compare with state-of-the-art methods, we synthesize a test dataset based on model (1). During synthesis, we choose a fixed atmospheric light whose RGB channels are equal to 0.784, and choose three scattering coefficients $\beta \in \{0.06, 0.3, 0.54\}$ to generate their corresponding transmission map, then use atmospheric light, clear image and transmission map to synthesize the hazy image. We choose all images in the NYU-D dataset to generate the test hazy dataset. Hence, there are in total 4347 testing images, denoted as **HIFA** dataset. Since the dataset is synthesized, the ground truth haze-free images are available, which enables us to evaluate the performance qualitatively as well as quantitatively using $SSIM$ and $PSNR$. It is worth noting that our test dataset contains light , middle and heavy haze images, which can evaluate dehazing methods comprehensively.

First, we compare our method with two steps state-of-the-art methods on **HIFA** dataset. For indoor hazy image dataset, we set the leaving haze rate

as 1.0 for all traditional methods. We generate the corresponding results using code provided by authors and set parameters according the papers. We use the method proposed in [34] to get the atmospheric light for method [12] and [10] for left methods. We show quantitative comparison results in Table 2. We can see that our method significantly outperforms the current state-of- the-art methods.

Table 2: Quantitative comparison on the HIFA dataset w/o the ground truth atmospheric lights A.

	BCCR [10]	NLD [12]	DehazeNet [5]	MSCNN [26]	Ours
PSNR	11.46	15.54	16.81	17.68	20.20
SSIM	0.527	0.713	0.745	0.788	0.827

In order to show the high performance of our method, we exclude the influence of atmospheric lights, and only consider the transmission map estimating accurate. We set the atmospheric lights to ground truth for all methods and generate the dehazing results. The $SSIM$ and $PSNR$ results are shown in table 3, we find that our method gets the highest score.

Table 3: Quantitative comparison on the HIFA dataset with the ground truth atmospheric lights A.

	BCCR [10]	NLD [12]	DehazeNet [5]	MSCNN [26]	Ours
PSNR	18.70	17.38	19.06	18.64	20.20
SSIM	0.794	0.764	0.789	0.800	0.827

Second, we compare with End-to-End methods. For Zhang *et al.*'s method we resize the hazy image and corresponding ground truth haze-free image to 516×516, then we generate the corresponding results using the code provided by author. For Ren et al's method we generate the result using the code provide by author. The $SSIM$ and $PSNR$ results are shown in Table 4. We can see that our method generates a much higher $PSNR$ metrics than other state-of-the-art methods. It worth noting that our model is trained on outdoor hazy image dataset, which result in that our performance only slight better than Ren *et al.*'s model which is trained on indoor dataset. It also worth noting that our **HIFA** dataset contains training images of Ren *et al.*'s model.

Table 4: Quantitative comparison on the HIFA dataset.

	DCPDN [30]	GFN [7]	AOD-Net [6]	Ours
PSNR	15.93	19.97	18.64	20.20
SSIM	0.736	0.820	0.747	0.827

In Figure 8, we show some dehazing results by state-of-the-art methods. Since our method directly recovers the haze-free image from single input hazy image, we only compare the final output results with other state-of-the-art methods. The prior based method [10] tends to over estimate the density of haze in image and darken the final dehazing results. The learning-based methods [26, 5] tend to leave a lot haze in final restore results. Method of Ren *et al.* [7] tends to have white block in dehazing results. In contrast, the dehazed results generated by our methods shown in Figure 8(g) are close to the ground truth haze-free

images in Figure 8(h). The final results generated by our method have higher visual quality and fewer color distortions in general.

5.2. Natural Image Dehazing

To demonstrate the generalization ability of the proposed method, we conduct a system comparison with state-of-the-art methods on natural hazy images.

First, we compare our method with separate optimize methods on several real-world hazy images provided by previous works. In Figure 9, we list the dehazing results and compare with five state-of-the-art methods [8, 10, 11, 26, 5]. As revealed in Figure 9, He *et al.*'s method's results tend to be darker than normal, which can be observed in first and second rows. Ren *et al.*'s and Zhu et al's method tend to under estimating the density of haze and leave haze in final restore results (shown on the second row). We also find that Cai *et al.*'s method losses some details and shows low contrast (shown in forth row). In contrast, Our method can overcome these problems well. Our method is able to generate photo realistic colors while better removing haze. This can be seen by comparing the first and the second row results. From the first row image, we can see that methods [8, 10, 11, 26] tend to leave haze in the dehazing result, Cai *et al.*'s method alleviates this problem. However, the detail of tree is losing. Our method can generate a haze-free dehazing result and keep the detail well. From the fifth row image, we can see that [8] and [10] generate some color distortion in final dehazing result. Cai *et al.*'s method tends to leave haze in result and loss some detail in far mountain. Methods [11, 26] can't deal the region of village well and can't recover the detail well. Our result can overcome the mentioned problem well and generate a high quality result.

Second, we compare our method with jointly optimization methods on heavy hazy images. Four heavy haze images are used to show the high performance of our method by comparing with four End-to-End dehazing methods [6, 7, 25, 30]. The dehazed results are shown in Figure 10. Method of Li *et al.*(shown on the first and second row) tends to under estimate the density of haze and hence leave haze in final result(notice the grass in first row and the tree in second row). It can be observed from these results that outputs from methods of S Zhang *et al.* [25] and Ren *et al.* suffer from color distortions and halo especially in lake area of first row image(please see the rectangle in first row). Although method of H Zhang [30] can generate a more haze-free result than Li *et al.*'s method without undesirable artifacts. Compare with our results, the results of H Zhang [30] tend to leave haze in the results (please see the blue rectangle in first row, you only can see one shadow). From the third row, we can see that S Zhang et al's method also exists some griding artifacts (we show it in lightcyan rectangle). We also notice that the texture in blue rectangle of [6, 7, 25, 30] tend to be blur in third row. In contrast, our method is able to yield better dehazing with visually pleased results without any undesirable artifacts in the final output images.

Third, we compare our method with more state-of-the-art dehazing methods [14, 9, 22, 11, 5, 26, 12, 6, 30, 42, 41] on a challenging image. The results obtained by dehazing methods are shown in Figure 11. As shown in Figure 11,

the result of DCP [14] tends to be two dark due to the over dehazing. The result of FVR [9] tends to show color distortion. The results of RF [22], CAP [11] DehazeNet [5] and AOD-Net [42] tend to retain haze. The result of DCPDN [30] tends to lose detail in some areas. The result of GFN [42] tends to show color distortion. EPDN [7] cannot deal far area well. GridDehazeNet [41] (Termed as GridNet in Figure 11), MSCNN [26] and NLD [12] tend to leave haze in far areas. In contrast, our method can generate a visual pleased dehazed result.

Finally, we retrain the AOD-Net and FFA-Net on the proposed dataset. We test the performance of AOD-Net and FFA-Net on HIFA dataset. The performance results of these two methods are listed in Table 5. We also show an example in Figure 12. As shown in Figure 12, the result of FFA-Net tends to retain haze, the result of AOD-Net tends to an unnatural appearance. In contrast, our method removes haze well and shows a natural appearance.

(a) Hazy input (b) AOD-Net [6] (c) FFA-Net [43] [43] (d) ours

Figure 12: Qualitative evaluations on the real hazy images. The proposed method generates much clearer images with clearer structures and characters.

Table 5: Quantitative comparison on the HIFA dataset.

	AOD-Net	FFA-Net	Ours
PSNR	18.54	19.60	20.20
SSIM	0.750	0.760	0.827

5.3. Running Time Comparison

Our light-weight deep learning model NLDN results to faster dehazing. We select 50 images from HIFA dataset to test all methods on the same machine (Intel(R) Core(TM) i5-6300HQ CPU@2.3GH and 8GB memory). The per-image average running time of all methods are shown in Table 6. The result shows the high efficiency of our model.

6. Conclusions

In this paper, we present a fully End-to-End learning-based dehazing method. Our method uses point-wise convolution to extract better intermediate representation, then uses dilation convolution to seek a better combination of intermediate representation, and finally we use this combination to reconstruct the haze-free result. Compared with prior methods which impose assumption on scene transmission map and atmospheric light, our proposed NLDN method is

Table 6: Average running time on the HIFA dataset.

Method	DCP	BCCR	NLD	MSCNN	DehazeNet	GFN	Li	DCPDN	EPDN	FD-GAN	Our
Language	Matlab						Python				
Platform				MatConvNet	Caffe		pytorch				TF
CPU (s)	25.08	3.52	8.41	2.45	2.56	19.03	1.23	3.24	3.51	4.02	1.84
GPU (s)						0.49	0.16	0.36	0.39	0.42	0.26

more easy to implement and reproduce. Experiments evaluated on synthetic and real datasets shows that the proposed method is able to generate significantly better and more visual pleased results as compared to the recent state-of-the-art methods.

7. Acknowledgements

This work is supported by National Key Research and Develop Program of China, Grant No. 2017YFB0503004

References

[1] Y. Y. Schechner, S. G. Narasimhan, S. K. Nayar, Instant dehazing of images using polarization, in: CVPR, 2001.

[2] S. G. Narasimhan, S. K. Nayar, Contrast restoration of weather degraded images, TPAMI 25 (6) (2003) 713–724.

[3] S. Shwartz, E. Namer, Y. Y. Schechner, Blind haze separation, in: 2006 IEEE on Computer Vision and Pattern Recognition, Vol. 2, IEEE, 2006, pp. 1984–1991.

[4] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, D. Lischinski, Deep photo: Model-based photograph enhancement and viewing, in: TOG, Vol. 27, 2008, p. 116.

[5] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: An end-to-end system for single image haze removal, TIP 25 (11) (2016) 5187–5198.

[6] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, An all-in-one network for dehazing and beyond, in: ICCV, 2017.

[7] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, M.-H. Yang, Gated fusion network for single image dehazing, in: CVPR, 2018.

[8] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, in: CVPR, 2009.

[9] J.-P. Tarel, N. Hautiere, Fast visibility restoration from a single color or gray level image, in: ICCV, 2009.

[10] G. Meng, Y. Wang, J. Duan, S. Xiang, C. Pan, Efficient image dehazing with boundary constraint and contextual regularization, in: ICCV, 2013.

[11] Q. Zhu, J. Mai, L. Shao, A fast single image haze removal algorithm using color attenuation prior, TIP 24 (11) (2015) 3522–3533.

[12] D. Berman, S. Avidan, et al., Non-local image dehazing, in: CVPR, 2016.

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

[14] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, TPAMI 33 (12) (2011) 2341–2353.

[15] R. Fattal, Dehazing using color-lines, TOG 34 (1) (2014) 13.

[16] W. Wang, X. Yuan, X. Wu, Y. Liu, Dehazing for images with large sky region, Neurocomputing 238 (2017) 365–376.

[17] W. Ren, J. Pan, H. Zhang, X. Cao, M.-H. Yang, Single image dehazing via multi-scale convolutional neural networks with holistic edges, International Journal of Computer Vision (2019) 1–20.

[18] W. Ren, J. Zhang, X. Xu, L. Ma, X. Cao, G. Meng, W. Liu, Deep video dehazing with semantic segmentation, IEEE Transactions on Image Processing 28 (4) (2018) 1895–1908.

[19] W. Ren, J. Zhang, L. Ma, J. Pan, X. Cao, W. Zuo, W. Liu, M.-H. Yang, Deep non-blind deconvolution via generalized low-rank approximation, in: Advances in Neural Information Processing Systems, 2018, pp. 297–307.

[20] Y. Yan, W. Ren, X. Cao, Recolored image detection via a deep discriminative model, IEEE Transactions on Information Forensics and Security 14 (1) (2018) 5–17.

[21] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, M.-H. Yang, Low-light image enhancement via a deep hybrid network, IEEE Transactions on Image Processing 28 (9) (2019) 4364–4375.

[22] K. Tang, J. Yang, J. Wang, Investigating haze-relevant features in a learning framework for image dehazing, in: CVPR, 2014, pp. 2995–3000.

[23] L.-Y. He, J.-Z. Zhao, D.-Y. Bi, Effective haze removal under mixed domain and retract neighborhood, Neurocomputing 293 (2018) 29–40.

[24] W. Ren, X. Cao, Deep video dehazing, in: Pacific-Rim Conference on Multimedia, 2017.

[25] S. Zhang, W. Ren, J. Yao, Feed-net: Fully end-to-end dehazing, in: ICME, 2018.

[26] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, M.-H. Yang, Single image dehazing via multi-scale convolutional neural networks, in: ECCV, 2016.

[27] J. Xie, L. Xu, E. Chen, Image denoising and inpainting with deep neural networks, in: NIPS, 2012, pp. 341–349.

[28] C. Dong, C. C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, TPAMI 38 (2) (2016) 295–307.

[29] C. J. Schuler, M. Hirsch, S. Harmeling, B. Schölkopf, Learning to deblur, TPAMI 38 (7) (2016) 1439–1451.

[30] H. Zhang, V. M. Patel, Densely connected pyramid dehazing network, in: CVPR, 2018.

[31] R. T. Tan, Visibility in bad weather from a single image, in: CVPR, 2008.

[32] R. Fattal, Single image dehazing, TOG 27 (3) (2008) 72.

[33] M. Sulami, I. Glatzer, R. Fattal, M. Werman, Automatic recovery of the atmospheric light in hazy images, in: ICCP, 2014.

[34] D. Berman, T. Treibitz, S. Avidan, Air-light estimation using haze-lines, in: ICCP, 2017.

[35] S. Zhang, F. He, J. Yao, Single image dehazing using deep convolution neural networks, in: Pacific Rim Conference on Multimedia, Springer, 2017, pp. 315–325.

[36] S. Zhang, J. Yao, Single image dehazing using fixed points and nearest-neighbor regularization, in: Asian Conference on Computer Vision, 2016, pp. 18–33.

[37] S. Zhang, F. He, W. Ren, J. Yao, Joint learning of image detail and transmission map for single image dehazing, The Visual Computer (12) (2018) 1–12.

[38] S. Zhang, J. Yao, E. B. Garcia, Single image dehazing via image generating, in: Pacific-Rim Symposium on Image and Video Technology, Springer, 2017, pp. 123–136.

[39] S. Zhang, F. He, Drcdn: learning deep residual convolutional dehazing networks, The Visual Computer (2019) 1–12.

[40] Y. Liu, J. Pan, J. Ren, Z. Su, Learning deep priors for image dehazing, in: ICCV, 2019.

[41] X. Liu, Y. Ma, Z. Shi, J. Chen, Griddehazenet: Attention-based multi-scale network for image dehazing, in: ICCV, 2019.

[42] Y. Qu, Y. Chen, J. Huang, Y. Xie, Enhanced pix2pix dehazing network, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[43] X. Qin, Z. Wang, Y. Bai, X. Xie, H. Xie, Ffa-net: Feature fusion attention network for single image dehazing, in: AAAI, 2020.

[44] Y. Dong, Y. Liu, H. Zhang, S. Chen, Y. Qiao, Fd-gan: Generative adversarial networks with fusion-discriminator for single image dehazing, in: AAAI, 2020.

[45] F. Guo, X. Zhao, J. Tang, H. Peng, L. Liu, B. Zou, Single image dehazing based on fusion strategy, Neurocomputing 378 (2020) 9–23.

[46] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: NIPS, 2012.

[47] F. Liu, C. Shen, G. Lin, I. Reid, Learning depth from single monocular images using deep convolutional neural fields, TPAMI 38 (10) (2016) 2024–2039.

[48] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: ICLR, 2015.

[49] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122.

[50] L. Xu, J. Ren, Q. Yan, R. Liao, J. Jia, Deep edge-aware filters, in: International Conference on Machine Learning, 2015, pp. 1669–1678.