Gated Fusion Network for Single Image Dehazing

Wenqi Ren[1]*, Lin Ma[2], Jiawei Zhang[3], Jinshan Pan[4], Xiaochun Cao[1,5]†, Wei Liu[2], and Ming-Hsuan Yang[6]
[1]State Key Laboratory of Information Security (SKLOIS), IIE, CAS
[2]Tencent AI Lab [3]City University of Hong Kong [4]Nanjing University of Science and Technology
[5]School of Cyber Security, University of Chinese Academy of Sciences
[6]Electrical Engineering and Computer Science, University of California, Merced
https://sites.google.com/site/renwenqi888/research/dehazing/gfn

Abstract

In this paper, we propose an efficient algorithm to directly restore a clear image from a hazy input. The proposed algorithm hinges on an end-to-end trainable neural network that consists of an encoder and a decoder. The encoder is exploited to capture the context of the derived input images, while the decoder is employed to estimate the contribution of each input to the final dehazed result using the learned representations attributed to the encoder. The constructed network adopts a novel fusion-based strategy which derives three inputs from an original hazy image by applying White Balance (WB), Contrast Enhancing (CE), and Gamma Correction (GC). We compute pixel-wise confidence maps based on the appearance differences between these different inputs to blend the information of the derived inputs and preserve the regions with pleasant visibility. The final dehazed image is yielded by gating the important features of the derived inputs. To train the network, we introduce a multi-scale approach such that the halo artifacts can be avoided. Extensive experimental results on both synthetic and real-world images demonstrate that the proposed algorithm performs favorably against the state-of-the-art algorithms.

1. Introduction

The single image dehazing problem [9, 45] aims to estimate the unknown clean image given a hazy or foggy image. This is a classical image processing problem, which has received active research efforts in the vision communities since various high-level scene understanding tasks [19, 29, 32, 40] require the image dehazing to recover the clear scene. Early approaches focus on developing hand-crafted features based on the statistics of clear images, such as dark channel prior [9] and local max contrast [2, 33]. To avoid hand-crafted priors, recent work [4, 14, 28, 41] automatically learns haze relevant features by convolutional neural networks (CNNs). In the dehazing literature, the hazing process is usually modeled as,

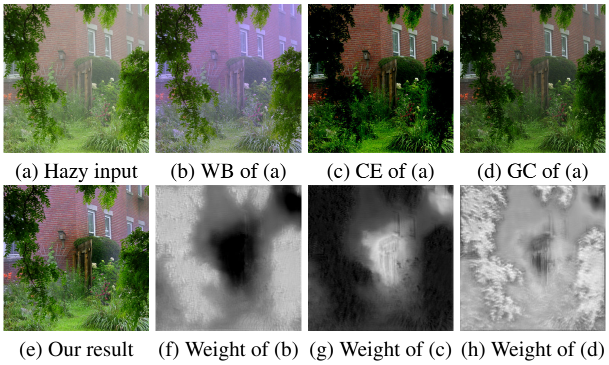


Figure 1. Image dehazing result. We exploit a gated fusion network for single image deblurring. (a) Hazy input. (b)-(d) are the derived inputs. (f)-(h) are learned confidence maps for (b), (c) and (d), respectively. (e) Our result.

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

where $\mathbf{I}(x)$ and $\mathbf{J}(x)$ are the observed hazy image and the haze-free scene radiance, $\mathbf{A}$ is the global atmospheric light, and $t(x)$ is the scene transmission describing the portion of light that is not scattered and reaches the camera sensors. In practice, transmission and atmospheric light are unknown. Thus, most dehazing methods try to estimate the transmission $t(x)$ and the atmospheric light $\mathbf{A}$, given a hazy image.

Estimating transmission from a hazy image is a severely ill-posed problem. Some approaches try to use visual cues to capture deterministic and statistical properties of hazy images [3, 6, 8, 30]. However, these transmission approximations are inaccurate, especially in the cases of the scenes where the colors of objects are inherently similar to those of atmospheric lights. Note that such an erroneous transmission estimation directly affects the quality

*Part of this work was done while Wenqi Ren was with Tencent AI Lab as a Visiting Scholar.

†Corresponding author.

of the recovered image, resulting in undesired haze artifacts. Instead of using hand-crafted visual cues, CNN-based methods [4, 28] are proposed to estimate the transmissions. However, these methods still follow the conventional dehazing methods in estimating atmospheric lights to recover clear images. Thus, if the transmissions are not estimated well, they will interfere the following atmospheric light estimation and thereby lead to low-quality results.

To address the above issues, we propose a novel end-to-end trainable neural network that does not explicitly estimate the transmission and atmospheric light. Thus, the artifacts arising from transmission estimation errors can be avoided in the final restored results. The proposed neural network is built on a fusion strategy which aims to seamlessly blend several input images by preserving only the specific features of the composite output image.

There are two major factors in hazy images that need to be dealt with. The first one is the color cast introduced by the atmospheric light. The second one is the lack of visibility due to attenuation. Therefore, *we tackle these two problems by deriving three inputs from the original image with the aim of recovering the visibility of the scene in at least one of them.* The first input ensures a natural rendition (Figure 1(b)) of the output by eliminating chromatic casts caused by the atmospheric light. The second contrast enhanced input yields a better global visibility, but mainly in the thick hazy regions (*e.g*., the rear wall in Figure 1(c)). However, the contrast enhanced images are too dark in the light hazy regions. Hence, to recover the light hazy regions, we find that the gamma corrected images restore information of the light hazy regions well (*e.g*., the front lawn in Figure 1(d)). Consequently, the three derived inputs are gated by three confidence maps (Figure 1(f)-(g)), which aim to preserve the regions with good visibility.

The contributions of this work are three-fold. First, we propose a deep end-to-end trainable neutral network that restores clear images without assuming any restrictions on scene transmission and atmospheric light. Second, we demonstrate the utility and effectiveness of a gated fusion network for single image dehazing by leveraging the derived inputs from an original hazy image. Finally, we train the proposed model with a multi-scale approach to eliminate the halo artifacts that hurt image dehazing. We show that the proposed dehazing model performs favorably against the state-of-the-arts.

2. Related Work

There mainly exist three kinds of methods for image dehazing: multi-image based methods, hand-crafted priors based methods, and data-driven methods.

Multi-image aggregation. Early methods often require multiple images to deal with the dehazing problem [23, 13, 36]. Kopf *et al*. [13] used an approximated 3D model of the scene for dehazing. Different polarized filters were used in [36] to capture multiple images of the same scene, and then degrees of polarization were used for haze removal. Narasimhan and Nayar [23] also used the differences between multiple images for estimating the haze properties.

All these methods make the same assumption of using multiple images in the same scene. However, there only exists one image for a specific scene in most cases.

Hand-crafted priors based methods. Different image priors have been explored for single image dehazing in previous methods [16]. Tan *et al.* [33] enhanced the visibility of hazy images by maximizing the contrast. The dehazed results of this method often present color distortions since this method is not physically valid. He *et al*. [9] presented a dark channel prior (DCP) for outdoor images, which asserts that the local minimum of the dark channel of a haze-free image is close to zero. The DCP has been shown effective for image dehazing, and a number of methods improve [9] in terms of efficiency [35] or quality [24]. Fattal [7] discovered that pixels of image patches typically exhibit a one-dimensional distribution, and used it to recover the scene transmission. However, this approach cannot guarantee a correct classification of patches. Recently, Berman *et al*. [3] observed that colors of a haze-free image can be well approximated by a few hundred distinct colors, and then proposed a dehazing algorithm based on this prior.

Another line of research tries to make use of a fusion principle to restore hazy images in [1, 5]. However, these methods need complex blending based on luminance, chromatic and saliency maps. In contrast, we introduce a gated fusion based single image dehazing technique that blends only the derived three input images.

All of the above approaches strongly rely on the accuracy of the assumed image priors, so may perform poorly when the assumed priors are insufficient to describe real-world images. As a result, these approaches tend to introduce undesirable artifacts such as color distortions.

Data-driven methods. Tang *et al*. [34] combined four types of haze-relevant features with Random Forest to estimate the transmission. Zhu *et al*. [46] created a linear model for modeling the scene depth of the hazy image under a color attenuation prior, and learned the parameters of the model in a supervised manner. However, these methods are still developed based on hand-crafted features.

Recently, CNNs have also been used for image recovering problems [4, 14, 39, 42, 43, 44]. Cai *et al*. [4] proposed a DehazeNet and a BReLU layer to estimate the transmissions from hazy inputs. In [28], a coarse-scale network was first used to learn the mapping between hazy inputs and their transmissions, and then a fine-scale network was exploited to refine the transmission. One problem of these CNNs based methods [4, 27, 28] is that all these methods require an accurate transmission and atmospheric light

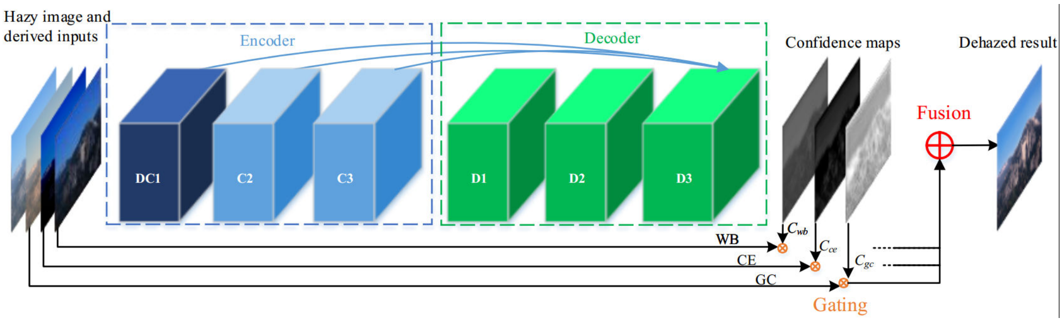


Figure 2. The coarsest level network of GFN. The network contains layers of symmetric encoder and decoder. To retrieve more contextual information, we use Dilation Convolution (DC) to enlarge the receptive field in the convolutional layers in the encoder block. Skip shortcuts are connected from the convolutional feature maps to the deconvolutional feature maps. Three enhanced versions are derived from the input hazy image. Then, these three inputs are weighted by the three confidence maps learned by our network, respectively.

estimation step for restoring the clear image. Although the recent AOD-Net [14] bypasses the estimation step, this method still needs to compute a newly introduced variable $\mathbf{K}(x)$ which integrates both transmission $t(x)$ and atmospheric light $\mathbf{A}$. Therefore, AOD-Net still falls into a physical model in (1).

Different from these CNNs based approaches, our proposed network is built on the principle of image fusion, and is learned to produce the sharp image directly without estimating transmission and atmospheric light. The main idea of image fusion is to combine several images into a single one, retaining only the most significant features. This idea has been successfully used in a number of applications such as image editing [25] and video super-resolution [18].

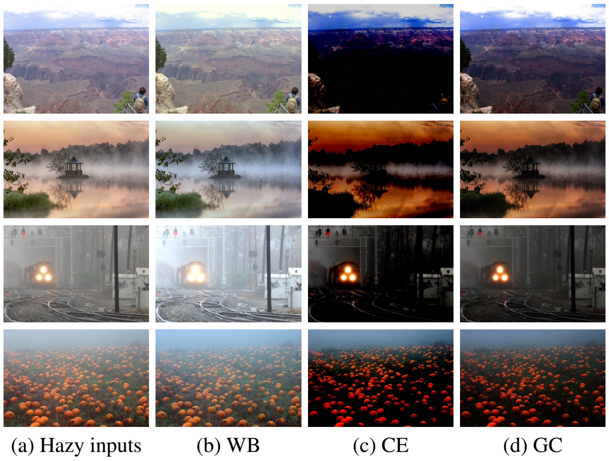

(a) Hazy inputs (b) WB (c) CE (d) GC

Figure 3. We derive three enhanced versions from an input hazy image. These three derived inputs contain different important visual cues of the input hazy image.

3. Gated Fusion Network

This section presents the details of our gated fusion network that employs an original hazy image and three derived images as inputs. We refer to this network as *Gated Fusion Network*, or *GFN*, as shown in Figure 2. The central idea is to learn the *confidence maps* to combine several input images into a single one by keeping only the most significant features of them. Obviously, the choice of inputs and weights is application-dependent. By learning the confidence map for each input, we demonstrate that our fusion based method is able to dehaze images effectively.

3.1. Derived Inputs

We derive several inputs based on the following observations. The first one is that the colors in hazy images often change due to the influence of the atmospheric light. The second one is the lack of visibility in distant regions due to scattering and attenuation phenomena. Based on these observations, we generate three inputs that recover the color and visibility of the entire image from the original hazy image. We first estimate the White Balanced (WB) image $\mathbf{I}_{wb}$ of the hazy input $\mathbf{I}$ to recover the latent color of the scene. Then we extract visible information including the Contrast Enhanced (CE) image $\mathbf{I}_{ce}$ and the Gamma Corrected (GC) image $\mathbf{I}_{gc}$ to yield a better global visibility.

White balanced input. Our first input is a white balanced image which aims to eliminate chromatic casts caused by the atmospheric color. In the past decades, a number of white balancing approaches [11] have been proposed. In this paper, we use the gray world assumption [26] based technique. Despite its simplicity, this low-level approach has shown to yield comparable results to those of more complex white balance methods [17]. The gray world assumption is that given an image with a sufficient quantity of color variations, the average value of the Red, Green and Blue components of the image should average out to a common gray value. This assumption is in generally valid in any given real-world scene since the variations in colors are random and independent. It would be safe to say that given a large number of samples, the average should tend to converge to the mean value, which is gray. White balancing algorithms can make use of this gray world assumption by forcing images to have a uniform average gray value for the

R, G, and B channels. For example, if an image is shot under a hazy weather condition, the captured image will have an atmospheric light $\mathbf{A}$ cast over the entire image. The effect of this atmospheric light cast disturbs the gray world assumption of the original image. By imposing the assumption on the captured image, we would be able to remove the atmospheric light cast and re-acquire the colors of our original scene. Figure 3(b) demonstrates such an effect.

Although white balancing could discard the color shifting caused by the atmospheric light, the results still present low contrast. To enhance the contrast, we introduce the following two derived inputs.

Contrast enhanced input. Inspired by the previous dehazing approaches [1] and [5], our second input is a contrast enhanced image of the original hazy input. Ancuti and Ancuti [1] derived a contrast enhanced image by subtracting the average luminance value $\tilde{I}$ of the entire image $\mathbf{I}$ from the hazy input, and then using a factor μ to linearly increase the luminance in the recovered hazy regions as follows:

$$\mathbf{I}_{ce} = \mu\big(\mathbf{I} - \tilde{I}\big), \tag{2}$$

where $\mu = 2(0.5 + \tilde{I})$. Although $\tilde{I}$ is a good indicator of image brightness, there is a problem in this input, especially in denser haze regions. The main reason is that the negative values of $(\mathbf{I} - \tilde{I})$ may dominate the contrast enhanced input as $\tilde{I}$ increases. As shown in Figure 3(c), the dark image regions tend to be black after contrast enhancing.

Gamma corrected input. To overcome the dark limitation in $\mathbf{I}_{ce}$, we create another type of contrast enhanced image using gamma correction:

$$\mathbf{I}_{gc} = \alpha \mathbf{I}^{\gamma}. \tag{3}$$

Gamma correction is a nonlinear operation which is used to encode ($\gamma < 1$) and decode ($\gamma > 1$) luminance or tristimulus values in image content, In this paper, we use $\alpha = 1$ and a decoding gamma correction $\gamma = 2.5$. We find that using these parameters achieves satisfactory results, as shown in Figure 3(d). The derived inputs by decoding gamma correction effectively remove the severe dark aspects of $\mathbf{I}_{ce}$ and enhance the visibility of the original image $\mathbf{I}$.

3.2. Network Architecture

We use an encoder-decoder network, which has been shown to produce good results for a number of generative tasks such as image denoising [20], image harmonization [37], time-lapse video generation [38]. In particular, we choose a variation of the residual encoder-decoder network model for image dehazing. We use skip connections between encoder and decoder halves of the network, where features from the encoder side are concatenated to be fed to the decoder. This significantly accelerates the convergence [20] and helps generate a much clear dehazed image.

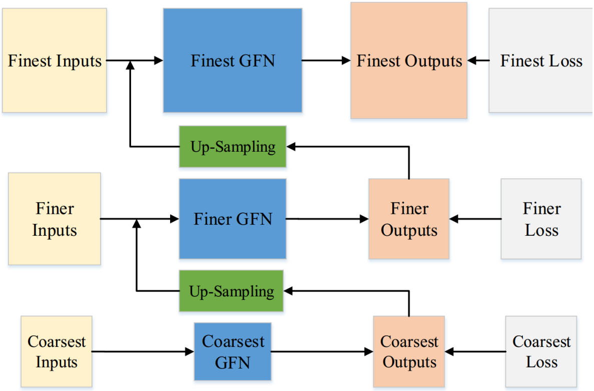


Figure 4. Multi-scale GFN structure.

We perform an early fusion by concatenating the original hazy image and three derived inputs in the input layer. The network is of a multi-scale style in order to prevent halo artifacts, which will be discussed in more details in Section 3.3. We show a diagram of GFN in Figure 2. Note that we only show the coarsest level network of GFN in Figure 2. To leverage more context without losing local details, we use dilation network to enlarge the receptive field in the convolutional layers. Rectification layers are added after each convolutional or deconvolutional layer. The convolutional layers act as a feature extractor, which preserve the primary information of scene colors in the input layer, meanwhile eliminating the unimportant colors from the inputs. The deconvolutional layers are then combined to recover the weight maps of three derived inputs. In other words, the outputs of the deconvolutional layers are the *confidence maps* of the derived input images $\mathbf{I}_{wb}$, $\mathbf{I}_{ce}$ and $\mathbf{I}_{gc}$.

We use 3 convolutional blocks and 3 deconvolutional blocks with stride 1 in each scale. Each layer is of the same type: 32 filters of the size $3 \times 3 \times 32$ except the first and last layers. The first layer operates on the input image with kernel size 5×5, and the last layer is used for confidence map reconstruction. In this work, we demonstrate that explicitly modeling confidence maps has several advantages. These are discussed later in Section 5.2. Once the confidence maps for the derived inputs are predicted, they are multiplied by the three derived inputs to give the final dehazed image in each scale:

$$J = C_{wb} \circ \mathbf{I}_{wb} + C_{ce} \circ \mathbf{I}_{ce} + C_{gc} \circ \mathbf{I}_{gc}, \tag{4}$$

where $\circ$ denotes element-wise multiplication, and C_{wb}, C_{ce}, and C_{gc} are the *confidence maps* for gating $\mathbf{I}_{wb}$, $\mathbf{I}_{ce}$, and $\mathbf{I}_{gc}$, respectively.

3.3. The multi-Scale Refinement

The network described in the previous subsection is subject to halo artifacts, particularly for strong transitions within the confidence maps [1, 5]. Hence, we perform estimation by varying the image resolution in a coarse-to-fine manner to prevent halo artifacts. The multi-scale approach is motivated by the fact that the human visual system is sensitive to local changes (*e.g.*, edges) over a wide range of scales. As a merit, the multi-scale approach provides a convenient way to incorporate local image details over varying resolutions.

Figure 4 shows the proposed multi-scale fusion network, in which the coarsest level network is shown in Figure 2. Finer level networks basically have the same structure as the coarsest network. However, the first convolutional layer takes the sharp image from a previous stage as well as its own hazy image and derived inputs, in a concatenated form. Each input size is twice the size of its coarser scale network. There is an up-sampling layer before the next stage. At the finest scale, the original high-resolution image is restored.

The multi-scale approach desires that each scale output is a clear image of the corresponding scale. Thus, we train our network so that all intermediate dehazed images should form a pyramid of the sharp image. The MSE criterion is applied to every level of the pyramid. In specific, given a collection of N training pairs $\mathbf{I}_i$ and $\mathbf{J}_i$, where $\mathbf{I}_i$ is a hazy image and $\mathbf{J}_i$ is the clean version as the ground truth, the loss function at the k-th scale is defined as follows:

$$\mathcal{L}_{cont}(\Theta, k) = \frac{1}{N}\sum_{i=1}^{N} \|\mathcal{F}(\mathbf{I}_{i,k}, \Theta, k) - \mathbf{J}_{i,k}\|^2, \qquad (5)$$

where Θ keeps the weights of the convolutional and deconvolutional kernels.

3.4. Adversarial Loss

Recently, generative adversarial networks (GANs) are reported to generate sharp realistic images [22]. Therefore, we follow the architecture introduced in [22], and build a discriminator to take the output of the finest scale or the ground-truth sharp image as input. The adversarial loss is defined as follows:

$$\begin{aligned}\mathcal{L}_{adv} = &\mathop{\mathbb{E}}_{\mathbf{J}\backsim p_{\text{clear}}(\mathbf{J})}\big[\log D(\mathbf{J})\big] \\ &+ \mathop{\mathbb{E}}_{\mathbf{I}\backsim p_{\text{hazy}}(\mathbf{I})}\big[\log\big(1 - D(\mathcal{F}(\mathbf{I}))\big)\big],\end{aligned} \qquad (6)$$

where $\mathcal{F}$ is our multi-scale network in Figure 4, and D is the discriminator. Finally, by combining the multi-scale content loss and adversarial loss, our final loss function is

$$\mathcal{L}_{total} = \mathcal{L}_{cont} + 0.001\mathcal{L}_{adv}. \qquad (7)$$

Through optimizing the network parameters, we train the model in the combination of two losses, multi-scale content loss (5) and adversarial loss (6).

4. Experimental Results

We quantitatively evaluate the proposed algorithm on both synthetic dataset and real-world hazy photographs, with comparisons to the state-of-the-art methods in terms of accuracy and visual effect. The implementation code can be found at our project website.

4.1. Implementation Details

In our network, patch size is set as 128 × 128. We use ADAM [12] optimizer with a batch size 10 for training. The initial learning rate is 0.0001 and we decrease the learning rate by 0.75 every 10,000 iterations. For all the results reported in the paper, we train the network for 240,000 iterations, which takes about 35 hours on an Nvidia K80 GPU. Default values of β_1 and β_2 are used, which are 0.9 and 0.999, respectively, and we set weight decay to 0.00001. Since our approach dehazes images in a single forward pass, it is computationally very efficient. Using a NVidia K80 GPU, we can process a 640 × 480 image within 0.3s.

4.2. Training Data

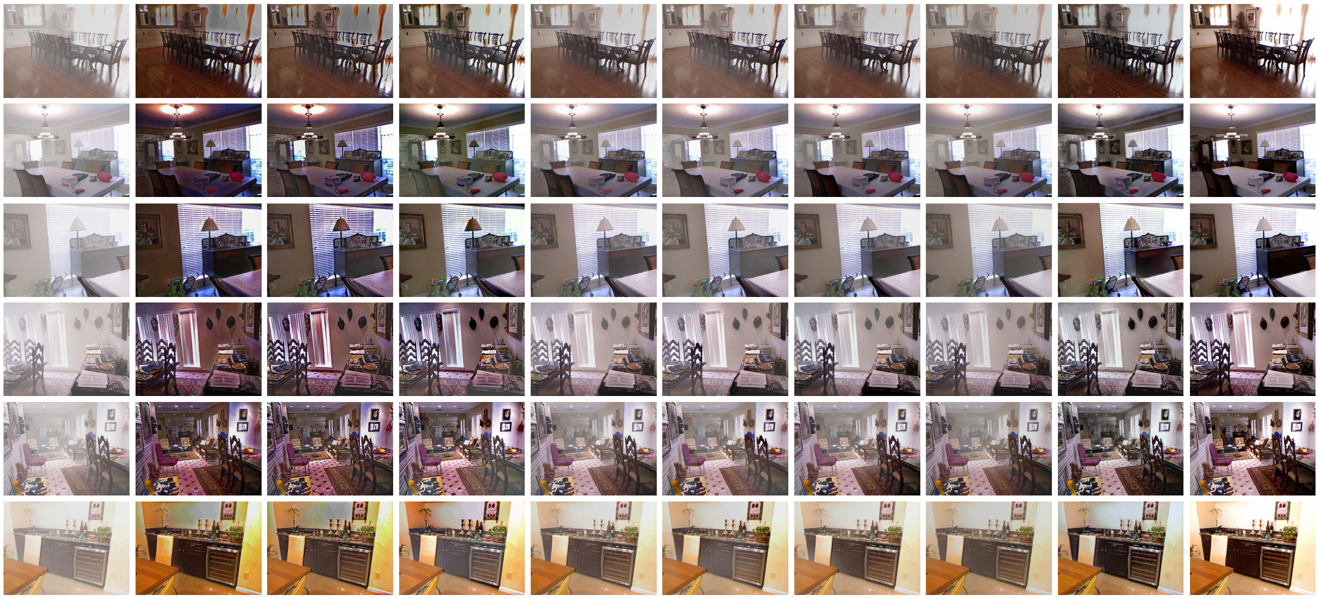
Generating realistic training data is a major challenge for tasks where ground truth data cannot be easily collected. For training our neural network, we adopt the NYU2 dataset [31] and the synthetic method in [28] to synthesize the training data. We use 1400 clean images and the corresponding labeled depth maps from the NYU Depth dataset [31] to construct the training set. Given a clear image $\mathbf{J}$, a random atmospheric light $\mathbf{A} \in (0.8, 1.0)$ and the ground truth depth d, we use $t(x) = e^{-\beta d(x)}$ to synthesize transmission first, then generate hazy image using the physical model (1). For scattering coefficient β, we randomly select it from 0.5 to 1.5 as suggested in [28]. We use 7 different β for each clean image, so that we can synthesize different haze concentration images for each input image. In addition, 1% Gaussian noise is added to each hazy input to increase the robustness of the trained network.

4.3. Quantitative Evaluation on Synthetic Dataset

For quantitative evaluation, we use the remaining 49 clean images in the label data except the 1400 training images from the NYU2 dataset [31] to synthetic hazy images with known depth map d as like in [28]. We evaluate these methods by two criteria: Structure Similarity (SSIM) and Peak Signal to Noise Ratio (PSNR). In this section, we compare the proposed algorithm with the following seven methods on the synthesized datasets.

Priors based methods [10, 21, 3]. We use three prior based methods for comparisons. The first one is the DCP proposed by He *et al.* [9, 10]. This is a commonly used baseline approach in most dehazing papers. The second is Boundary Constrained Context Regularization (BCCR) proposed by Meng *et al.* [21] and the third is the Non-local Image Dehazing (NLD) algorithm in [3].

Learning based methods [46, 4, 28, 14]. We also use four learning based methods for comparisons. The first



(a) Hazy inputs (b) DCP (c) BCCR (d) NLD (e) CAP (f) MSCNN (g) DehazeNet (h) AOD-Net (i) GFN (j) Ground truths

Figure 5. Dehazed results on the synthetic dataset. Dehazed results generated by the priors based methods [10, 21, 3] have some color distortions in some regions. The learning based methods [46, 28, 4, 14] tend to underestimate haze concentration so that the dehazed results have some remaining hazes. In contrast, the dehazed results by our method are close to the ground-truth images.

Table 1. Average PSNR and SSIM values of dehazed results on the synthetic dataset.

	PSNR/SSIM								
	DCP [10]	BCCR [21]	NLD [3]	CAP [46]	MSCNN [28]	DehazeNet [4]	AOD-Net [14]	GFN (G)	GFN ($G + D$)
Light	18.74/0.77	17.72/0.76	18.61/0.71	21.92/0.83	22.34/0.82	**24.87**/0.84	22.64/**0.85**	24.78/**0.85**	24.60/0.83
Medium	18.68/0.77	17.54/0.75	18.47/0.70	21.40/0.82	21.21/0.80	23.37/0.83	21.33/**0.84**	**23.68/0.84**	23.55/**0.84**
Heavy	18.67/0.77	17.43/0.75	18.21/0.70	20.21/0.80	20.51/0.79	21.98 /0.82	20.24/0.81	**22.32/0.83**	22.75/0.82
Random	18.58/0.77	17.35/0.75	18.28/0.71	19.99/0.78	20.01/0.78	20.97/0.80	19.36/0.78	**22.41**/0.81	22.20/**0.82**

one learns a linear model based on Color Attenuation Prior (CAP). The second and third are CNNs based methods of DehazeNet [4] and MSCNN [28]. These methods implement image dehazing by learning the map between hazy inputs and their transmission based on convolutional neural networks. The last AOD-Net [14] is also a CNNs based method, but integrates the transmission and atmospheric light into a new variable.

Figure 5 shows some dehazed images by different methods. Since we directly restore the final dehazed image without transmission estimation in our algorithm, we only compare the final dehazed results with other methods. The priors based image dehazing methods [10, 21, 3] overestimate the haze thickness, so the dehazed results tend to be darker than the ground truth images and contain color distortions in some regions, *e.g.*, the desks in the second row and the wall in the last row in Figure 5(b)-(d). We note that the dehazed results by CAP [46], DehazeNet [4], MSCNN [28] and AOD-Net [14] methods are similar as shown in Figure 5(e)-(h). Although the dehazed results by CAP, DehazeNet, MSCNN and AOD-Net are closer to ground truth than the results by [10, 21, 3], there are still some remaining haze as shown in Figure 5(e)-(h).

In contrast, the dehazed results generated by our approach in Figure 5(i) are close to the ground truth haze-free images in Figure 5(j). Overall, the dehazed results by the proposed algorithm have higher visual quality and fewer color distortions. The qualitative results are also reflected by the quantitative PSNR and SSIM metrics in Table 1.

In addition, to further test the dehazing effect on different haze concentration, we use three scattering coefficient $\beta = 0.8, 1.0$ and 1.2 to synthesize three haze concentration on the 49 testing images, respectively. As shown in Table 1, our method without adversarial loss performs favorably against the state-of-the-art image dehazing methods [10, 21, 3, 46, 28, 4, 14] on all of these haze concentrations. However, if we use adversarial loss, the network can still recover better dehazed results than without adding adversarial loss in terms of SSIM in some cases. Although the SSIM values by [14] are close to ours in some cases, the PSNR generated by our method are higher than [14] by up to 2dB, especially for heavy haze concentration images.

RESIDE dataset. Recently, a dehazing benchmark is proposed in [15], which is an extended version of our data in Table 1. We further evaluate our method on the RESIDE dataset in Table 2. As shown, our method performs favor-

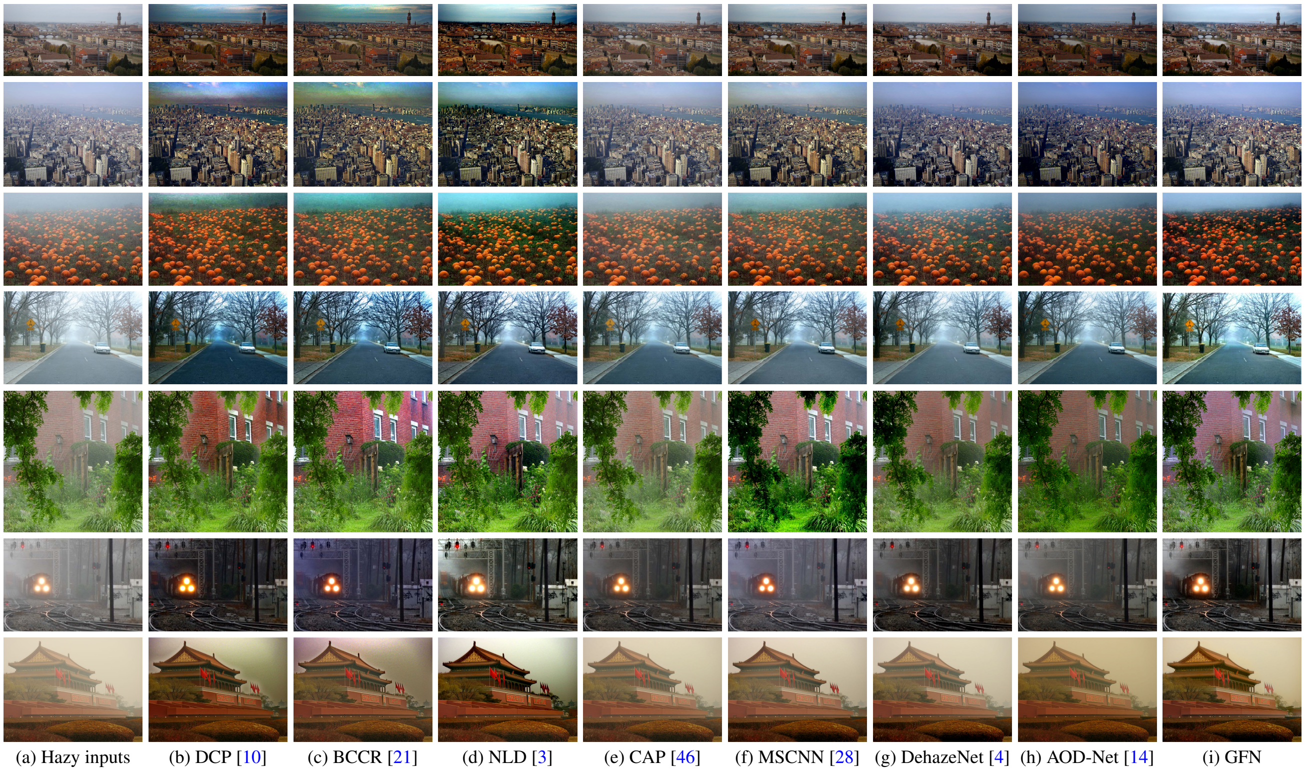


Figure 6. Qualitative comparison of different methods on real-world images. Best viewed on high-resolution display.

Table 2. Average PSNR/SSIM of dehazed results on the SOTS dataset from **RESIDE**.

NLD [3]	MSCNN [28]	DehazeNet [4]	AOD-Net [14]	GFN
17.27/0.75	17.57/0.81	21.14/0.85	19.06/0.85	**22.30/0.88**

ably against other competitors [3, 4, 14, 28] in this dataset.

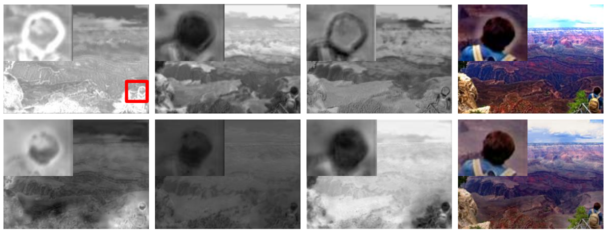
4.4. Evaluation on Real Images

To further evaluate the proposed method, we use the real image dataset in Fattal [7] and compare with different state-of-the-art methods. Figure 6 shows the qualitative comparison of results with the seven state-of-the-art dehazing algorithms [9, 21, 3, 28, 4, 14] on challenging real-world images. Figure 6(a) shows the hazy images to be dehazed. Figure 6(b)-(h) shows the results of DCP [9], BCCR [21], NLD [3], CAP [46], MSCNN [28], DehazeNet [4] and AOD-Net [14], respectively. The results generated by the proposed algorithm are given in Figure 6(i). As shown in Figure 6(b)-(d), most of the haze is removed by DCP, BCCR and NLD methods, and the details of the scenes and objects are well restored. However, the results significantly suffer from over-enhancement (for instance, the sky region of the first and second images are much darker than it should be as shown in Figure 6(b)-(d), and there are some color distortions in the second and last images in Figure 6(c) and (d)). This is because these algorithms are based on hand-crafted priors which have an inherent problem of overestimating the transmission as discussed in [9, 46]. The results of CAP [46] do not have the over-estimation problem and maintain the original colors of the objects as shown in Figure 6(e). But have some remaining haze in the dehazed results. For example, the third image. The dehazed results by MSCNN [28] and DehazeNet [4] have a similar problem as [46] tends to have some remaining haze. Especially the last image in Figure 6(f) and the first image in Figure 6(g). The method of AOD-Net [14] generates relatively clear results, but the images in first three rows are still dark than ours, while the results in last two rows still have some remaining haze as shown in Figure 6(h). In contrast, the dehazed results by our method are clear and the details of the scenes are enhanced moderately.

5. Analysis and Discussions

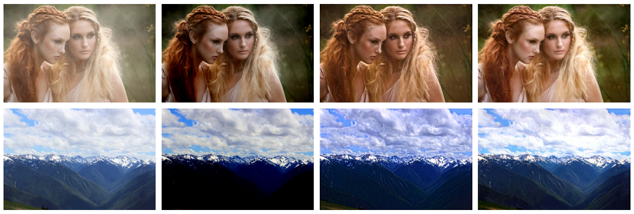
5.1. Effectiveness of Multi-Scale Network

In this section we analyze how the multi-scale network helps refine dehazed results. The recovered images from coarser-scale network provide additional information in the finer-scale net, which can greatly improve the final dehazed results. We show the dehazed results generated by only using the finest-scale and the proposed multi-scale networks in Figure 7. Figure 7 shows that dehazed results and corresponding confidence maps. The first row is the dehazed results by only using the finest scale network and the second row is the results by the proposed multi-scale approach. As



(a) Maps of $\mathbf{I}_{wb}$ (b) Maps of $\mathbf{I}_{ce}$ (c) Maps of $\mathbf{I}_{gc}$ (d) GFN

Figure 7. Effectiveness of the proposed multi-scale approach. The first and second rows are the results by single and multi-scale networks, respectively. The zoomed-in regions are shown in the left-top corner in each image.



(a) Hazy inputs (b) Without gating (c) Without fusion (d) GFN

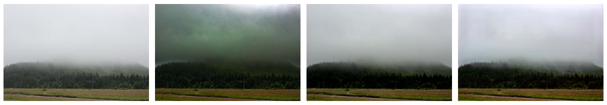
Figure 8. Effectiveness of the gated fusion network.

shown in the first row in Figure 7(a) and (c), there are obvious halo around the *head of the person* in the confidence maps, so the final dehazed result in the first row Figure 7(d) has the halo artifacts. In contrast, the dehazed results generated by the proposed multi-scale approach has a more clean edge as shown in the second row in Figure 7(d).

5.2. Effectiveness of Gating Strategy

Image fusion is a method to blend several images into a single one by retaining only the most useful features. To blend effectively the information of the derived inputs, we filter their important information by computing corresponding confidence maps. Consequently, in our gated fusion network, the derived inputs are gated by three pixel-wise confidence maps that aim to preserve the regions with good visibility. Our fusion network has two advantages: the first one is that it can reduce patch-based artifacts (*e.g*. dark channel prior [9]) by single pixel operations, and the other one is that it can eliminate the influence caused by transmission and atmospheric light estimation.

To show the effectiveness of fusion network, we also train an end-to-end network without fusion process. This network has the same architecture as DFN except the input is hazy image and output is dehazed result without confidence maps learning. In addition, we also conduct a experiment based on equivalent fusion strategy, *i.e*., all the three derived inputs are weighted equally using $1/3$. Figure 8 shows visual comparisons of on two real-world examples with different settings. In these examples, the approach without gating generates very dark images in Figure 8(b), and the method without fusion strategy generates results with color distortion and dark regions as shown in Figure 8(c). In contrast, our results recover most scene details and maintain the original colors.



(a) Hazy input (b) DCP [9] (b) DehazeNet [4] (d) GFN

Figure 9. A failure case for a thick foggy image.

5.3. Limitations

The proposed DFN performs well in general natural images. However, as the previous methods [28, 4], a limitation of our method is that the DFN cannot handle corrupted images with very large fog as shown in Figure 9. As heavy haze seriously interferes the atmospheric light (which is not a constant), the hazy model does not hold for such examples. Figure 9(d) shows an example where the proposed method does not generate a clear image. Future work will consider this problem with haze-free reference retrieval based on an effective deep neural network model.

6. Conclusions

In this paper, we addressed the single image dehazing problem via a multi-scale gated fusion network (GFN), a fusion based encoder-decoder architecture, by learning confidence maps for derived inputs. Compared with previous methods which impose restrictions on scene transmission and atmospheric light, our proposed GFN is easy to implement and reproduce since the proposed approach does not rely on the estimations of transmission and atmospheric light. In the approach, we first applied white balance method to recover the scene color, and then generated two contrast enhanced images for better visibility. Third, we carried out the GFN to estimate the confidence map for each derived input. Finally, we used the confidence maps and derived inputs to render the final dehazed result. The experimental results on synthetic and real-world images demonstrate the effectiveness of the proposed approach.

Acknowledgments. This work is supported in part by National Key Research and Development Plan (No.2016YFB0800603), National Natural Science Foundation of China (No.U1636214, 61733007), Beijing Natural Science Foundation (No.4172068). W. Ren and M.-H. Yang are supported by NSF CAREER (No. 1149783) and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR).

References

[1] C. O. Ancuti and C. Ancuti. Single image dehazing by multi-scale fusion. *TIP*, 22(8):3271–3282, 2013. 2, 4

[2] L. Bao, Y. Song, Q. Yang, and N. Ahuja. An edge-preserving filtering framework for visibility restoration. In *ICPR*, pages 384–387, 2012. 1

[3] D. Berman, S. Avidan, et al. Non-local image dehazing. In *CVPR*, 2016. 1, 2, 5, 6, 7

[4] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *TIP*, 25(11):5187–5198, 2016. 1, 2, 5, 6, 7, 8

[5] L. K. Choi, J. You, and A. C. Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *TIP*, 24(11):3888–3901, 2015. 2, 4

[6] R. Fattal. Single image dehazing. In *SIGGRAPH*, 2008. 1

[7] R. Fattal. Dehazing using color-lines. *TOG*, 34(1):13, 2014. 2, 7

[8] N. Hautière, J.-P. Tarel, and D. Aubert. Towards fog-free in-vehicle vision systems through contrast restoration. In *CVPR*, 2007. 1

[9] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. In *CVPR*, 2009. 1, 2, 5, 7, 8

[10] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *TPAMI*, 33(12):2341–2353, 2011. 5, 6, 7

[11] R. Kawakami, H. Zhao, R. T. Tan, and K. Ikeuchi. Camera spectral sensitivity and white balance estimation from sky images. *IJCV*, 105(3):187–204, 2013. 3

[12] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[13] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski. Deep photo: Model-based photograph enhancement and viewing. In *SIGGRAPH Asia*, 2008. 2

[14] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *ICCV*, 2017. 1, 2, 3, 5, 6, 7

[15] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Reside: A benchmark for single image dehazing. *arXiv preprint arXiv:1712.04143*, 2017. 6

[16] Y. Li, R. T. Tan, and M. S. Brown. Nighttime haze removal with glow and multiple light colors. In *ICCV*, 2015. 2

[17] Y. Li, S. You, M. S. Brown, and R. T. Tan. Haze visibility enhancement: A survey and quantitative benchmarking. *arXiv preprint arXiv:1607.06235*, 2016. 3

[18] D. Liu, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, and T. Huang. Robust video super-resolution with learned temporal dynamics. In *ICCV*, 2017. 3

[19] S. Liu, Y. Sun, D. Zhu, G. Ren, Y. Chen, J. Feng, and J. Han. Cross-domain human parsing via adversarial feature and label adaptation. In *AAAI*, 2018. 1

[20] X. Mao, C. Shen, and Y.-B. Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *NIPS*, 2016. 4

[21] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *ICCV*, 2013. 5, 6, 7

[22] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 5

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

[24] K. Nishino, L. Kratz, and S. Lombardi. Bayesian defogging. *IJCV*, 98(3):263–278, 2012. 2

[25] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *TOG*, 22(3):313–318, 2003. 3

[26] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *Computer graphics and applications*, 21(5):34–41, 2001. 3

[27] W. Ren and X. Cao. Deep video dehazing. In *PCM*, 2017. 2

[28] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, 2016. 1, 2, 5, 6, 7, 8

[29] C. Sakaridis, D. Dai, and L. Van Gool. Semantic foggy scene understanding with synthetic data. *arXiv preprint arXiv:1708.07819*, 2017. 1

[30] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar. Instant dehazing of images using polarization. In *CVPR*, 2001. 1, 2

[31] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from RGBD images. In *ECCV*, 2012. 5

[32] Y. Song, L. Bao, and Q. Yang. Real-time video decolorization using bilateral filtering. In *WACV*, 2014. 1

[33] R. T. Tan. Visibility in bad weather from a single image. In *CVPR*, 2008. 1, 2

[34] K. Tang, J. Yang, and J. Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *CVPR*, 2014. 2

[35] J.-P. Tarel and N. Hautiere. Fast visibility restoration from a single color or gray level image. In *ICCV*, 2009. 2

[36] T. Treibitz and Y. Y. Schechner. Polarization: Beneficial for visibility enhancement? In *CVPR*, 2009. 2

[37] Y.-H. Tsai, X. Shen, Z. Lin, K. Sunkavalli, X. Lu, and M.-H. Yang. Deep image harmonization. *arXiv preprint arXiv:1703.00069*, 2017. 4

[38] W. Xiong, W. Luo, L. Ma, W. Liu, and J. Luo. Learning to generate time-lapse videos using multi-stage dynamic generative adversarial networks. In *CVPR*, 2018. 4

[39] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M.-H. Yang. Learning to super-resolve blurry face and text images. In *ICCV*, 2017. 2

[40] Y. Yuan, X. Liang, X. Wang, D.-Y. Yeung, and A. Gupta. Temporal dynamic graph lstm for action-driven video object detection. In *ICCV*, 2017. 1

[41] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. In *CVPR*, 2018. 1

[42] H. Zhang and V. M. Patel. Density-aware single image deraining using a multi-stream dense network. In *CVPR*, 2018. 2

[43] H. Zhang, V. Sindagi, and V. M. Patel. Image de-raining using a conditional generative adversarial network. *arXiv preprint arXiv:1701.05957*, 2017. 2

[44] J. Zhang, J. Pan, W.-S. Lai, R. W. Lau, and M.-H. Yang. Learning fully convolutional networks for iterative non-blind deconvolution. In *CVPR*, 2016. 2

[45] X.-S. Zhang, S.-B. Gao, C.-Y. Li, and Y.-J. Li. A retina inspired model for enhancing visibility of hazy images. *Frontiers in computational neuroscience*, 9, 2015. 1

[46] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *TIP*, 24(11):3522–3533, 2015. 2, 5, 6, 7