Nighttime Image Dehazing Based on Multi-Scale Gated Fusion Network

Bo Zhao¹,²,*, Han Wu¹, Zhiyang Ma², Huini Fu², Wenqi Ren³ and Guizhong Liu¹

¹ School of Information and Communication Engineering, Xi’an Jiaotong University, Xi’an 710049, China
² China North Vehicle Research Institute, Beijing 100072, China
³ School of Cyber Science and Technology, Sun Yat-sen University, Shenzhen Campus, Shenzhen 528406, China
* Correspondence: whu_c2003@163.com

Abstract: In this paper, we propose an efficient algorithm to directly restore a clear image from a hazy input, which can be adapted for nighttime image dehazing. The proposed algorithm hinges on a trainable neural network realized in an encoder–decoder architecture. 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 input 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 clear image is generated by gating the important features of the derived inputs. To train the network, we introduce a multi-scale approach to avoid the halo artifacts. Extensive experimental results on both synthetic and real-world images demonstrate that the proposed algorithm performs favorably against the state-of-the-art dehazing for nighttime images.

Keywords: night image dehazing; encoder–decoder architecture; image fusion; multi-scale network

1. Introduction

The single-image dehazing [1,2] aims to estimate the unknown clean scene given a hazy or foggy image. This is a classical image processing problem, which has received active research efforts in the computer vision communities [3]. Early dehazing methods focus on exploiting hand-crafted features based on the statistics of clean images, such as dark channel prior [1] and local max contrast [4]. To avoid hand-crafted priors, recent work [5–7] automatically learns haze-relevant features using convolutional neural networks (CNNs). In the dehazing literature, under the assumption of spatially invariant atmospheric light, the hazing process is usually modeled as [1],

\[ I(x) = J(x)t(x) + A(1 - t(x)), \]

where J(x) and I(x) denote the haze-free scene radiance and the observed hazy image, 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. To recover the clear scene from a hazy input, most dehazing methods try to estimate the transmission t(x) and the atmospheric light A, given a hazy image.

Estimating transmission from hazy images is a severely ill-posed problem. Some approaches try to use visual cues to capture statistical properties of hazy images [8,9]. However, these transmission approximations are inaccurate, especially for 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 of the dehazed image, resulting in undesired haze artifacts. Instead of using hand-crafted features, CNN-based
approaches [5,7] are proposed to estimate the transmissions. However, these methods still follow the conventional dehazing methods in estimating atmospheric lights to recover clean images. Thus, if the transmission maps are not estimated well, they will interfere with the following airlight estimation and thereby lead to low-quality dehazed results.

In addition, even the state-of-the-art deep learning based methods need to compute the atmospheric light [5,7,10] or reformulated variables which are dependent on the atmospheric light [6,11]. These approaches suffer from important limitations on nighttime hazy scenes. This is mainly due to the multiple light sources that cause a strongly non-uniform illumination of the scene. However, we note that there are a few works to address nighttime dehazing.

To address the above issues, we propose a novel trainable neural network that does not explicitly estimate the transmission and atmospheric light. Thus, the artifacts arising from transmission and airlight estimation errors can be alleviated 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. We derive several inputs based on two major factors in nighttime hazy images that need to be dealt with. The first one is the color cast introduced by the environmental 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 degraded image with the aim of recovering the visibility of the scene in at least one of them. The first input ensures a natural rendition (second column of Figure 1) of the output by eliminating chromatic casts caused by the atmospheric or environmental light. The second contrast-enhanced input generates a better holistic appearance but mainly in the thick hazy regions. 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. Consequently, the three derived inputs are gated by three confidence maps (fifth, sixth, and seventh columns of Figure 1), which aim to preserve the regions with good visibility. In addition, we propose to use the normalization (NM) of nighttime hazy images to provide detailed scene information by substituting gamma correction.

This paper is an extension of our preliminary version [12], which concentrates on daytime dehazing. In this paper, we first improve the network architecture (Section 3.2) and then adapt our network to work effectively on nighttime hazy scenes (Section 4). The contributions of this paper are summarized as follows:

• We propose a deep trainable neural network that restores clear images without assuming restrictions on scene transmission and atmospheric light.
• We demonstrate the effectiveness of a gated fusion network for single nighttime image dehazing by leveraging the derived inputs from an original input.
• We train the proposed model with a multi-scale approach to eliminate the halo artifacts that hurt image recovering.
• We show that the proposed algorithm can effectively process nighttime hazy images which are not well handled by most dehazing methods. We show that the proposed method performs favorably against the state-of-the-arts.
2. Related Work

2.1. Day-Time Image Dehazing

Tang et al. [13] combined four types of haze-relevant features with Random Forest to estimate the transmission. Zhu et al. [14] introduced a linear model and learned the parameters of the model in a supervised manner under a color attenuation prior. However, these methods are still developed based on hand-crafted features.

Recently, CNNs have also been used for haze removal and related problems [15–18]. Cai et al. [5] proposed a DehazeNet and a BReLU layer to estimate the transmissions from hazy inputs. In [7], 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. Zhang and Patel [10] proposed a densely connected encoder–decoder structure for joint estimating the transmission map and atmospheric light. Yang and Sun [11] build a deep architecture incorporating the prior learning for single image dehazing. In the recent level-aware progressive network (LAP-Net) model, an image is restored by fusing the results at various haze levels at different stages. However, one problem of these CNN-based methods [5,7] is that all these models require accurate transmission and atmospheric light estimation steps to restore clear images. Although the AOD-Net [6] method bypasses the estimation step, this approach still needs to compute an additional variable \( K(x) \) which integrates both transmission \( t(x) \) and atmospheric light \( A \). Thus, the AOD-Net falls as one of the physics models as described in (1) that encounters issues with ill-posed problems. To alleviate these problems, several end-to-end networks [19–22] have recently been proposed to directly filter the input image.

Different from these CNN-based approaches, our proposed network is built on the principle of image fusion, and it is trained 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 used in a number of applications such as image editing [23] and video super-resolution [24].

2.2. Nighttime Dehazing

Different from common image dehazing, nighttime hazy images often include visible man-made light sources with varying colors and non-uniform illumination [25]. These light sources may introduce noticeable amounts of glow that are not present in haze images taken in the daytime, which makes the estimation of atmospheric light inaccurate and causes some sharp images prior to becoming invalid. However, in recent years, the community has paid relatively less research attention to the nighttime haze removal problem.

Pei and Lee [26] estimate the ambient illumination and the haze thickness by transferring the hazy input into a grayish one; then, they recover the dehazed result using the refined DCP by a bilateral filter in local contrast correction. Zhang et al. [27] build a new imaging model for nighttime conditions; then, they remove the haze by using the DCP along with estimating the point-wise environmental light. Based on the proposed physics model, they estimate the ambient illumination and transmission by combining a maximum reflectance prior (MRP) [28]. However, MRP shares the common limitations of most statistical prior-based methods. When the scene objects are inherent with a solely distinct color, the maximum reflectance prior becomes invalid in nighttime scenes. In [29], Li et al. also introduce a nighttime haze model that is a linear combination of the direct transmission, airlight and glow. Using the physics model, the authors first reduce the effect of the glow and then recover the final dehazed result. Nevertheless, this approach tends to generate some halo artifacts in the dehazed results. Ancuti et al. [30] assume that the brightest pixels of local patches filtered by a minimal operator can capture the properties of atmospheric light, and they use the multi-scale fusion approach to obtain a visibility-enhanced image.
Similar to [25,30], we also propose a multi-scale fusion network for nighttime dehazing. Differently, without any tedious estimation of contrast, saturation, saliency, and airlight, we directly predict the weight maps for each derived input by the trainable network.

3. Multi-Scale Gated Fusion Network Architecture

This section presents the details of our multi-scale gated fusion network that employs an original degraded image and three derived images as inputs. We refer to this network as multi-scale GFN, or MSGFN, 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.

Figure 2. The architecture of the proposed multi-scale GFN, which takes a hazy image pyramid and the corresponding three enhanced versions as the input and outputs a latent image pyramid. These three derived inputs are weighted by the three confidence maps in each scale learned by our network, and the full-resolution output is the final dehazed result. The network contains layers of symmetric encoders and decoders. Skip shortcuts are connected from the convolutional feature maps to the deconvolutional feature maps.

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 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 $I_{wb}$ of the hazy input $I$ to recover the latent color of the scene. Then, we extract visible information including the contrast enhanced (CE) $I_{ce}$ and the gamma corrected (GC) $I_{gc}$ to generate better holistic quality.
**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 [31,32] have been proposed. In this paper, we use the gray world assumption [33] based technique. Despite its simplicity, this low-level approach has shown to generate comparable results to those of more complex white balance methods [3]. 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 general 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 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 3b 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.** Similar to prior dehazing methods [34,35], our second input is a contrast-enhanced image of the original hazy input. Ancuti [34] derived a contrast-enhanced image by subtracting the average luminance value \( \bar{I} \) of the entire image \( I \) from the hazy input and then using a factor \( \mu \) to linearly increase the luminance in the recovered hazy regions as follows:

\[
I_{ce} = \mu (I - \bar{I}),
\]

where \( \mu = 2(0.5 + \bar{I}) \). Although \( \bar{I} \) 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 \( (I - \bar{I}) \) may dominate the contrast-enhanced input as \( \bar{I} \) increases. As shown in Figure 3c, the dark image regions tend to be black after contrast enhancing.

![Figure 3](image)

**Figure 3.** We derive three enhanced versions from nighttime hazy images. These derived inputs contain different important visual cues of the input hazy images. (a) Inputs; (b) WB; (c) CE; (d) NM.

### 3.2. Network Architecture

Only using one scale is subject to halo artifacts in the dehazed results, particularly for strong transitions within the confidence maps [34,35]. 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.
The proposed multi-scale GFN is shown in Figure 2. Finer level networks basically have the same structure as the coarsest network. However, the first convolutional layer takes the dehazed output 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. As shown in Figure 2, there is an up-sampling layer to resize the coarser output before the next stage. At the finest scale, the original full-resolution image is recovered.

We use an encoder–decoder network in each scale, which has been shown to produce good results for a number of generative tasks. In particular, we choose a variation of the residual encoder–decoder block 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 and helps generate a much clear dehazed image. In addition, we improve encoder–decoder modules by using residual blocks [36] after each convolution layer. We use shared weights in each scale, which operates in a way similar to using data multiple times [37] (i.e., data augmentation regarding scales) and reduces the number of parameters need to be learned.

We perform an early fusion by concatenating the original hazy image and three derived inputs in the input layer. Rectification layers are added after each convolutional or deconvolutional layer. The convolutional layers act as a feature extractor, which preserves 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 $I_{wb}$, $I_{ce}$ and $I_{gc}$.

We use three down-convolutional blocks and three deconvolutional blocks in each scale. The stride for down-convolution layer is two, which down-samples feature maps to half size and doubles the channel of the previous layer. Each of the following ResBlocks contains two convolution layers. Each convolutional layer is of the same kernel size of $3 \times 3$ except the first layer. The first layer operates on the input image with kernel size of $5 \times 5$. In this work, we demonstrate that explicitly modeling confidence maps has several advantages. These are discussed later in Section 7.1. Once the confidence maps for the derived inputs are predicted, we fuse different inputs using the proposed gating method as illustrated in Figure 2,

$$J^k = \text{Gating}(I^k_{wb}, I^k_{ce}, I^k_{gc}),$$

where $J^k$ is the gated result at scale $k$. The gating function is defined by

$$\text{Gating}(x, y, z) = C_x \circ x + C_y \circ y + C_z \circ z,$$

where $\circ$ denotes element-wise multiplication, and $C_\cdot$ is the confidence map for the input.

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 particular, given a collection of $N$ training pairs $I_i$ and $J_i$, where $I_i$ is a hazy image and $J_i$ is the clean version as the ground truth, the loss function at the $k$-th scale is defined as follows:

$$\mathcal{L}(\Theta, k) = \frac{1}{N} \sum_{i=1}^{N} \left\| F(I_{ik}, \Theta, k) - J_{ik} \right\|_2^2, \quad k \in \{1, 2, 3\},$$

where $\Theta$ keeps the weights of the convolutional and deconvolutional kernels.

### 4. Nighttime Image Dehazing

Since nighttime scenes usually have artificial light sources that generate a glow effect in hazy images, most state-of-the-art dehazing methods based on (1) suffer from significant limitations on nighttime hazy scenes. Although several physics-based models [28,29] are developed to relax those strict constraints in (1) (e.g., homogeneous atmosphere illumina-
tion, unique extinction coefficient), a straightforward extension of common hazy image modeling to nighttime scenes cannot always hold in real cases. This is why our approach does not resort to an explicit inversion of the nighttime light propagation model in [28,29].

Fusion Process of Nighttime Dehazing

In this paper, we demonstrate that the proposed MSGFN can also effectively enhance nighttime hazy images. We employ the strategy described in Figure 2 to remove haze in nighttime images. For the derived inputs, we also use WB and CE to process a color correction step and visibility enhancement, respectively. However, there is another problem in nighttime hazy images that needs to be dealt with, i.e., non-uniform illumination caused by multiple light sources in the low-light environment. Therefore, we derive a third input, normalization (NM) of the nighttime hazy image, to obtain an illumination-balanced result and enhance the finest details in the nighttime scene.

The NM operation is obtained by linearly stretching all the pixel values in order to fit them into the interval [0, 1]. In this case, we achieve a better illumination result by contrast stretching the range of intensities of the hazy input. The main advantage of this operation is that we do not require any parameter to be tuned, and therefore, without information loss in the derived input. As shown in Figure 3d, the NM operation shifts and scales all the color pixel intensities of the input so that the pixel values cover the entire available dynamic range and obtain a balanced illumination.

Similar to the dehazing approach described in Section 3, we use the proposed MSGFN to predict three confidence maps for the derived inputs to ensure that regions of high contrast or high saliency will receive greater weights in the gated fusion process:

\[ J^k = \text{Gating}(I_{wb}^k, I_{ce}^k, I_{nm}^k), \]  

where \( I_{nm}^k \) is the normalized version of the nighttime hazy input at scale \( k \).

5. Nighttime Dehazing Results

We evaluate the proposed algorithm with nighttime configuration on real-world night hazy scenes, with comparisons to the state-of-the-art methods in terms of visual effect.

5.1. Training Data

Owing to the difficulty in obtaining realistic nighttime training data, we adopt the similar strategy as the daytime methods [38] to synthesize nighttime hazy scenes. Specifically, we select 4500 clear nighttime scenes in the KAIST dataset [39] and use the method proposed in [40] to estimate depth maps, which has been demonstrated to be effective for nighttime scene depth estimation. Then, we synthesize 4500 nighttime hazy images according to (1). Note that although some nighttime hazy imaging models are proposed [28,29] to account for artificial light sources, we found our synthesized nighttime hazy images based on (1) look natural as shown in Figure 4, since the proposed model in [28,29] is a generalization of (1) when the illumination is assumed to be a constant.

5.2. Quantitative Evaluation

For quantitative performance evaluation, we construct a new dataset of synthesized nighttime hazy images. We select 100 clear nighttime scenes (different from those that were used for training) from the KAIST dataset [39] to synthesize 500 hazy images (using different scattering coefficients to synthesize different haze concentrations). Figure 5 shows some dehazed images by the evaluated methods. The nighttime dehazing methods of MRP [28] and GMLC [29] generate the results with significant color distortions. The dehazed images by the deep learning approaches of MSCNN [7], GCAN [19], and GDN [41] still contain significant haze residuals. In contrast, our algorithm restores these images well. Overall, the dehazed results by the proposed algorithm are of higher visual quality and with fewer color distortions. The visual results in Figure 5 match the quantitative results shown in Table 1.
Figure 4. The proposed method for synthesizing nighttime hazy images. The first row shows original clear night scenes from [39], and the second row shows the synthesizing hazy images.

Figure 5. Dehazed results on synthetic nighttime images. The results by learning-based methods of MSCNN [7], GCAN [19], and GDN [41] have some remaining haze, while the nighttime dehazing methods of MRP [28] and GMLC [29] tend to generate some color distortions. In contrast, the dehazed results by our algorithm are close to the ground-truth images. (a) Hazy inputs; (b) DCP [42]; (c) MRP [28]; (d) GMLC [29]; (e) MSCNN [7]; (f) GCAN [19]; (g) GDN [41]; (h) Our results; (i) Ground truth.

Table 1. Average PSNR/SSIM of dehazed results by state-of-the-art dehazing methods on nighttime hazy images.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13.70/0.6063</td>
<td>24.94/0.902</td>
<td>17.45/0.7113</td>
<td>14.49/0.6936</td>
<td>19.18/0.8133</td>
<td>21.03/0.8916</td>
<td>30.92/0.9492</td>
</tr>
</tbody>
</table>

5.3. Qualitative Evaluation

To demonstrate that the proposed method generalizes well in real-world nighttime hazy scenes, we use real-world hazy images for experiments against the state-of-the-art dehazing algorithms designed for nighttime scenes, i.e., Maximum Reflectance Prior (MRP) [28] as well as Glow and Multiple Light Colors (GMLC) [29], and daytime scenarios, i.e., DCP [42], MSCNN [7], GCAN [19], and GDN [41].

Figure 6b,c show the results by the recent nighttime dehazing methods, i.e., MRP [28] and GMLC [29]. The MRP method [28] tends to darken the hazy inputs in some regions. For example, the road regions of the first image are much darker than those obtained by other methods. In addition, the GMLC model [29] generates some artifacts in sky regions, e.g., the first and third images in Figure 6e. Figure 6d–g demonstrate the limitations of the daytime dehazing approaches, i.e., DCP [1], MSCNN [7], GCAN [19], and GDN [41] when applied to nighttime hazy inputs. Both the prior-based [1] and CNN-based [7,19,41] methods cannot recover colors well, and they only slightly remove the haze in these night scenes. In contrast, our algorithm generates dehazed results with clearer and sharper details and without artifacts in the sky regions as shown in Figure 6h.
6. Further Experiments

6.1. Comparison on O-Haze

In the main paper, we evaluate the proposed algorithm on all the 45 hazy images from the O-HAZE dataset [43] against the state-of-the-art methods. In this supplementary material, we retrain the proposed MSGFN using the same 40 training data as in the NTIRE 2018 challenge [2] and compare it with the winning methods in [2] on the five test images. As shown in Table 2, our proposed method performs favorably against the winning methods in the NTIRE 2018 challenge [2] and achieves the highest SSIM score.

Table 2. Average PSNR/SSIM of dehazed results on the 5 test images in the O-Haze [43] dataset. Although our algorithm ranks third in terms of PSNR, our method achieves the highest SSIM score.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BJTU</td>
<td>24.598</td>
<td>0.777</td>
</tr>
<tr>
<td>2</td>
<td>KAIST-VICLAB [22]</td>
<td>24.232</td>
<td>0.687</td>
</tr>
<tr>
<td></td>
<td>Ours (MSGFN)</td>
<td>24.054</td>
<td>0.787</td>
</tr>
<tr>
<td>3</td>
<td>Scarlet Knights [21]</td>
<td>24.029</td>
<td>0.775</td>
</tr>
<tr>
<td>4</td>
<td>FKS</td>
<td>23.877</td>
<td>0.775</td>
</tr>
<tr>
<td>5</td>
<td>Dq-hisfriends</td>
<td>23.207</td>
<td>0.770</td>
</tr>
<tr>
<td>6</td>
<td>Ranjanisi [44]</td>
<td>23.180</td>
<td>0.705</td>
</tr>
<tr>
<td>7</td>
<td>Mt.Phoenix</td>
<td>23.124</td>
<td>0.755</td>
</tr>
<tr>
<td>8</td>
<td>Ranjanisi [44]</td>
<td>22.997</td>
<td>0.701</td>
</tr>
<tr>
<td>9</td>
<td>KAIST-VICLAB [45]</td>
<td>22.705</td>
<td>0.707</td>
</tr>
<tr>
<td>10</td>
<td>Mt.Phoenix</td>
<td>22.080</td>
<td>0.731</td>
</tr>
<tr>
<td>11</td>
<td>IVLab</td>
<td>21.750</td>
<td>0.717</td>
</tr>
<tr>
<td>12</td>
<td>CLEAR</td>
<td>20.291</td>
<td>0.683</td>
</tr>
<tr>
<td>13</td>
<td>CLFStudio</td>
<td>20.230</td>
<td>0.722</td>
</tr>
<tr>
<td>14</td>
<td>SiMiT-Lab [46]</td>
<td>19.628</td>
<td>0.674</td>
</tr>
<tr>
<td>15</td>
<td>AHappyFace</td>
<td>18.494</td>
<td>0.669</td>
</tr>
<tr>
<td>16</td>
<td>ASELSAN</td>
<td>18.123</td>
<td>0.675</td>
</tr>
<tr>
<td>17</td>
<td>Dehazing-by-retinex</td>
<td>17.547</td>
<td>0.652</td>
</tr>
<tr>
<td>18</td>
<td>IMCL</td>
<td>16.527</td>
<td>0.616</td>
</tr>
</tbody>
</table>

|               | baseline (hazy images) | 15.784 | 0.634 |

Figure 6. Qualitative comparison of different methods on real-world images. (a) Hazy inputs; (b) MRP [28]; (c) GMLC [29]; (d) DCP [42]; (e) MSCNN [7]; (f) GCAN [19]; (g) GDN [41]; (h) Our results.
6.2. Mixed Training Strategy

To demonstrate the robustness of the proposed MSGFN on different training strategies, we train an additional network with all three datasets (daytime, nighttime, and underwater datasets) together. We refer to this network as “all-in-one” and refer to the original network in the main paper as “separate”.

As shown in Table 3, the proposed model performs better on the daytime (SOTS and O-Haze) and nighttime datasets with the “separate” training strategy. Meanwhile, the performance on the underwater dataset becomes better with the “all-in-one” training strategy. Since the underwater inputs in the UIEB dataset are real-world images, the main reason may be that more types of training data benefit real-world image reconstruction.

Table 3. Comparison of MSGFN using different training strategies (“separate” vs. “all-in-one”).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Separate</th>
<th>All-in-One</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daytime (SOTS)</td>
<td>25.37/0.93</td>
<td>23.19/0.94</td>
</tr>
<tr>
<td>Daytime (O-Haze)</td>
<td>21.21/0.76</td>
<td>19.05/0.74</td>
</tr>
<tr>
<td>Nighttime</td>
<td>30.92/0.95</td>
<td>22.69/0.86</td>
</tr>
<tr>
<td>Underwater (UIEB)</td>
<td>17.61/0.86</td>
<td>21.99/0.91</td>
</tr>
</tbody>
</table>

7. Analysis and Discussions

7.1. Effectiveness of Fusion Strategy

Image fusion is a method to blend several images into a single one by retaining only the most useful features. To effectively blend 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 [1]) 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 a fusion process for the dehazing task. This network has the same architecture as MSGFN except the input is hazy image and output is dehazed result without confidence maps learning at each scale. In addition, we also conduct an experiment based on an equivalent fusion strategy, i.e., all the three derived inputs are weighted equally using 1/3. Figure 7 shows visual comparisons of on two real-world examples with different settings. In these examples, the approach without gating generates dark images in Figure 7b, and the method with an equivalent fusion strategy generates results with color distortion and dark regions as shown in Figure 7c. In contrast, our results contain most scene details and maintain the original colors which demonstrate the effectiveness of the learned confidence maps.

7.2. Effectiveness of Derived Inputs

We can design different inputs for different enhancement tasks. In practice, it is difficult to entirely remove the haze effects of hazy images by an enhancing approach. Therefore, the input generation process seeks to recover sharp regions in at least one of the derived inputs as analyzed in Section 3. They complement each other nicely to help dehazing by the gated fusion network as shown in Table 4.

Although we do not claim that these are the optimal inputs, our experiments show that the three derived inputs are the minimum inputs. Using two or fewer of them will not generate better results in the proposed network (Table 4) for nighttime image dehazing. In the future work, we will explore more effective derived inputs or directly learn the derived inputs in the fusion network. The network parameters comparison can be found in Table 5.
Figure 7. Effectiveness of the gated fusion network. (a) Hazy inputs; (b) w/o fusion; (c) Equivalent fusion; (d) MSGFN.

Table 4. Average PSNR/SSIM using different derived inputs. The method only using the original image means that we directly learn the mapping from degraded images to the clear ones.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Dehazing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>WB</td>
</tr>
<tr>
<td>√</td>
<td>×</td>
</tr>
<tr>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>√</td>
<td>×</td>
</tr>
<tr>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Table 5. Comparison of MSGFN and state-of-the-art dehazing approaches with respect to parameters.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOD-Net [6]</td>
<td>1.83 × 10^3</td>
<td>1.83 × 10^3</td>
</tr>
<tr>
<td>MSCNN [7]</td>
<td>8.01 × 10^3</td>
<td>8.01 × 10^3</td>
</tr>
<tr>
<td>Domain adaption [48]</td>
<td>2.27 × 10^5</td>
<td>2.27 × 10^5</td>
</tr>
<tr>
<td>PMS-Net [17]</td>
<td>2.44 × 10^5</td>
<td>2.44 × 10^5</td>
</tr>
<tr>
<td>GCAN [19]</td>
<td>7.03 × 10^5</td>
<td>7.03 × 10^5</td>
</tr>
<tr>
<td>EPDN [18]</td>
<td>1.74 × 10^7</td>
<td>1.74 × 10^7</td>
</tr>
<tr>
<td>DCPDN [10]</td>
<td>6.69 × 10^7</td>
<td>6.69 × 10^7</td>
</tr>
<tr>
<td>CGAN [16]</td>
<td>1.23 × 10^8</td>
<td>1.23 × 10^8</td>
</tr>
<tr>
<td>Ours</td>
<td>5.15 × 10^5</td>
<td>5.15 × 10^5</td>
</tr>
</tbody>
</table>

8. Conclusions

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

**Author Contributions:** Supervision, W.R. and G.L.; Validation, Z.M.; Visualization, H.F.; Writing—original draft, B.Z.; Writing—review & editing, H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

23. Pérez, P.; Gangnet, M.; Blake, A. Poisson image editing. ACM Trans. Graph. 2003, 22, 313–318. [CrossRef]