A Semi-supervised Nighttime Dehazing Baseline with Spatial-Frequency Aware and Realistic Brightness Constraint

Xiaofeng Cong[1] Jie Gui[1]* Jing Zhang[2] Junming Hou[1] Hao Shen[3]
[1]Southeast University [2]University of Sydney [3]Hefei University of Technology
cxf_svip@163.com, {guijie,junming_hou}@seu.edu.cn, {jingzhang.cv,haoshenhs}@gmail.com

Abstract

Existing research based on deep learning has extensively explored the problem of daytime image dehazing. However, few studies have considered the characteristics of nighttime hazy scenes. There are two distinctions between nighttime and daytime haze. First, there may be multiple active colored light sources with lower illumination intensity in nighttime scenes, which may cause haze, glow and noise with localized, coupled and frequency inconsistent characteristics. Second, due to the domain discrepancy between simulated and real-world data, unrealistic brightness may occur when applying a dehazing model trained on simulated data to real-world data. To address the above two issues, we propose a semi-supervised model for real-world nighttime dehazing. First, the spatial attention and frequency spectrum filtering are implemented as a spatial-frequency domain information interaction module to handle the first issue. Second, a pseudo-label-based retraining strategy and a local window-based brightness loss for semi-supervised training process is designed to suppress haze and glow while achieving realistic brightness. Experiments on public benchmarks validate the effectiveness of the proposed method and its superiority over state-of-the-art methods. The source code and Supplementary Materials are placed in the https://github.com/Xiaofeng-life/SFSNiD.

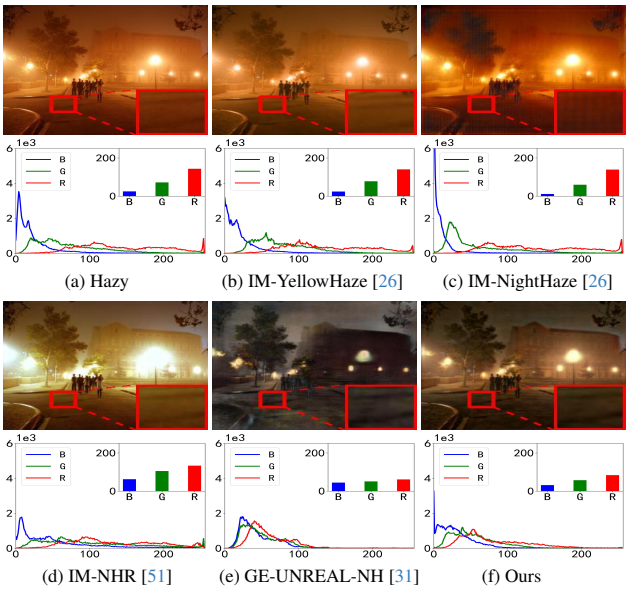


Figure 1. Visualization of real-world dehazed images, where the "IM-" and "GE-" denote the dehazed results obtained by training on imaging model (IM) and game engine (GE) simulated datasets, respectively. The curve figure represents the pixel histogram, where the x and y coordinates represent the pixel values and corresponding numbers, respectively. The x and y coordinates of the bar figure represent the color channel and the corresponding average pixel value, respectively.

1. Introduction

Nighttime and daytime images may contain hazy effects, which may cause their quality to be degraded [7, 15, 39, 50]. Therefore, two valuable research fields are proposed, which are daytime single image dehazing (DaSID) [2, 38, 54] and nighttime single image dehazing (NiSID) [14, 19, 31], respectively. Compared with the daytime hazy image, the imaging of the nighttime hazy image is more complex [28, 49]. Currently, NiSID is still a challenging problem.

Existing research on DaSID [20, 25, 27, 34, 43, 48, 52, 53] have achieved impressive performance. Various effective DaSID algorithms have been proposed and verified on benchmark daytime datasets [21]. However, these DaSID algorithms are designed for the properties of daytime hazy and haze-free images, without taking into account the characteristics of nighttime hazy and haze-free images.

Currently, NiSID research is divided into two types, namely non-deep learning-based NiSID and deep learning-based NiSID. On the one hand, the prior hypotheses and statistical laws are explored [50, 51]. The maximum reflectance prior to estimate the varying ambient illumination is proposed by [50]. The illumination estimation, color

*Corresponding author

correction and image prior are integrated by [49]. On the other hand, the deep learning-based architectures are designed for the NiSID task [14, 31]. Liu et al. [31] combine the dark channel and bright channel prior with the Transformer mechanism [32] into an end-to-end training flow. The gradient-adaptive convolution and glow pair synthesis are designed by Jin et al. [14]. Existing learning-based algorithms have achieved remarkable performance on synthetic datasets. However, these methods still lack consideration of the characteristics of nighttime hazy images.

During the day, the main source of imaging light is sunlight [7]. The formation of the daytime hazy image can be described by the atmospheric scattering model [7] as

$$I(a) = J(a)t(a) + A(a)(1 - t(a)), \quad (1)$$

where $I(a)$, $J(a)$, $t(a)$ and $A(a)$ denote the hazy image, clear image, transmission map and global atmospheric light, respectively. The a means the pixel location. Meanwhile, a widely used physical model [16, 18] in the NiSID task is

$$I(a) = J(a)t(a) + A(a)(1 - t(a)) + L_s(a) * \varkappa(a), \quad (2)$$

where $L_s(a)$ and $\varkappa(a)$ denote the light sources and atmospheric point spread function. As shown in Eq. 1 and Eq. 2, the main distinction between daytime and nighttime haze imaging is light sources [1, 4, 24, 29, 30, 41, 46], which we consider to be the main source of the difficulty. Specifically, two outstanding issues are considered as follows.

- **Localized, Coupled and Frequency Inconsistent**: As shown in Figure 1, multiple active light sources may exist simultaneously. Therefore, the distortion of nighttime images, namely the haze that is mainly generated by suspended particles and liquid water droplets, the glow that is mainly produced by active light sources and the noise that is mainly caused by low intensity, is usually *localized*. Meanwhile, these types of distortions are mixed throughout the image, which is *coupled*. Furthermore, the haze and glow will cause the loss of high-frequency signals, while the noise belongs to high-frequency disturbance signals [22] that needs to be eliminated. This means that these distortions have *inconsistent frequency characteristics*. In a word, a challenging issue is *how to simultaneously handle distortions with localized, coupled and frequency inconsistent characteristics.*
- **Unrealistic Brightness Intensity**: Nighttime hazy datasets based on real-world images synthesized by imaging model (IM) are difficult to simulate multiple active light sources, while nighttime hazy datasets based on game engine (GE) cannot perfectly reproduce the harmonious brightness of real-world nighttime scenes. As we observed in Figure 1, the dehazed images obtained under IM datasets still suffer from the glow and haze that caused by multiple light sources, but the overall brightness is realistic. The dehazed images obtained under GE dataset show less haze and glow, but the scene brightness is unrealistic. In a word, an unsolved problem faced by data-driven algorithms is *how to suppress haze and glow while achieving realistic brightness.*

Therefore, we propose a semi-supervised dehazing framework that can be used for the real-world NiSID task. Firstly, the local attention [32] is adopted to learn the inductive bias in the spatial domain to suppress local distortions. A frequency spectrum dynamic filtering strategy is designed to handle distortions with inconsistent frequency characteristics. Considering the coupled of these distortions, the spatial and frequency information are integrated as a bidomain interaction module for feature extraction and image reconstruction. Secondly, aiming at suppressing distortions while achieving realistic brightness. The simulation data provided by the game engine is utilized to generate pseudo labels that can suppress haze and glow for retraining process. Then, real-world hazy images are adopted as brightness-realistic signals for the realistic brightness constraint. Overall, the main contributions of this paper are as follows.

- We propose a spatial and frequency domain aware semi-supervised nighttime dehazing network (SFSNiD). SFSNiD can remove nighttime haze that is accompanied by glow and noise. The experimental results on synthetic and real-world datasets show that the proposed method can achieve impressive performance.
- We design a spatial and frequency domain information interaction (SFII) module to *simultaneously handle the haze, glow and noise with localized, coupled and frequency inconsistent characteristics*. The multi-channel amplitude and phase spectrums are dynamically filtered and aggregated. The spatial and frequency domain features are integrated by local attention.
- A retraining strategy and a local window-based brightness loss for semi-supervised training process are designed to *suppress haze and glow while achieving realistic brightness*. The retraining strategy is based on pseudo labels. The hazy image is divided into non-overlapping windows for the calculation of local brightness map to provide realistic brightness supervision.

2. Related Work

2.1. Daytime Dehazing

A variety of effective dehazing algorithms for DaSID have been proposed. An ultra-high resolution dehazing method based on bilateral gird is proposed by 4KDehazing [54]. AECRNet [43] introduces the contrastive learning to the dehazing process. The prior information and visual attention mechanism are utilized in DeHamer [9]. DF [38] designs an encoder-decoder architecture which totally based on multi-head self-attention [32]. MITNet [37] combines the mutual information-driven constraint and adaptive triple interaction

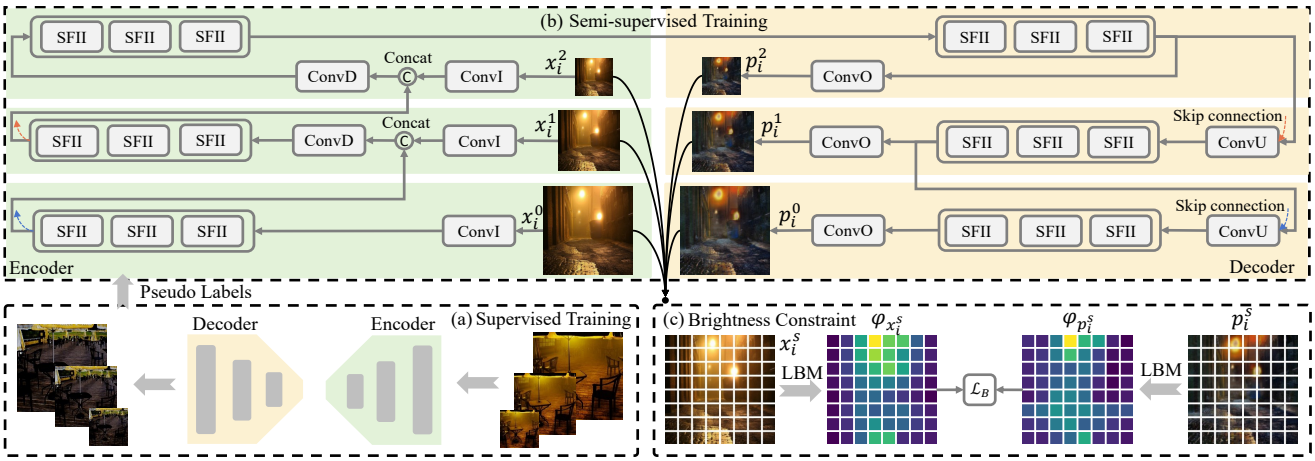


Figure 2. The overall pipeline of the proposed SFSNiD.

strategy into a supervised training process. Although these DaSID algorithms have achieve impressive performance, they are not designed for the characteristics of nighttime hazy images, which may cause them to have certain limitations on the NiSID task [31].

2.2. Nighttime Dehazing

Compared with DaSID, NiSID has received fewer attention. On the one hand, the prior hypotheses and statistical laws are utilized in the non-deep learning-based NiSID methods [50, 51]. A maximum reflectance prior is proposed by MRP [50], which providing a way to estimate the varying ambient illumination. An optimal-scale fusion-based method is designed by OSFD [51], which utilizes a parameter estimation dehazing flow. On the other hand, the data-driven strategies [23, 40] are adopted in the deep learning-based NiSID methods [14, 31, 45]. NightHazeFormer [31] combines the visual transformer and prior knowledge (dark channel and bright channel) into an end-to-end enhancement process. GAC [14] utilizes the angular point spread function to reduce the glow effect in nighttime scenes. Yan et al. [45] propose a strategy which decomposes the image into scene texture information and scene structure information. According to recent research, deep learning-based NiSID algorithms can achieve relatively better quantitative performance according to sufficient synthetic data. However, the haze, glow, and noise with localized, coupled and frequency inconsistent characteristics are not fully considered by these deep learning-based NiSID algorithms.

3. Methods

The hazy domain and haze-free domain are marked as X and Y, respectively. The synthesized hazy and haze-free image datasets are denoted $\mathcal{D}_X$ and $\mathcal{D}_Y$, which contain N images, respectively. The real-world hazy image and haze-free datasets are denoted as $\mathcal{R}_X$ and $\mathcal{R}_Y$, which include M images, respectively. The convolution operation is denoted as $C_t^k(\cdot)$, where the superscript k and subscript t denote the kernel size and stride, respectively. The $\varpi(\cdot)$, $\sigma(\cdot)$, $\delta(\cdot)$ and $sf(\cdot)$ denote the global average pooling, LeakyReLU, sigmoid and softmax operations, respectively. The input hazy images and predicted dehazed images at three scales are marked $x_i^s \in \mathcal{D}_X$ and p_i^s respectively, where $s \in \{0, 1, 2\}$ and i denotes the i-th example. The size of x_i^0, x_i^1 and x_i^2 are $H \times W \times C$, $\frac{H}{2} \times \frac{W}{2} \times C$ and $\frac{H}{4} \times \frac{W}{4} \times C$, respectively. The H, W and C denote the height, width and number of channels, respectively. The size of p_i^s remains the same as x_i^s. The network at scale s is denoted as $\Psi^s(\cdot)$.

3.1. Network Structure

The multi-scale structure [3] of the SFSNiD is shown in Figure 2. Two kinds of modules are included in the proposed network, namely (i) spatial and frequency information interaction (SFII) model, (ii) convolution input (ConvI), convolution output (ConvO), convolution downsampling (ConvD), and convolution upsampling (ConvU). The ConvI projects the image into the feature space, while ConvO does the opposite. ConvD reduces the length and width of the feature map by half, while ConvU does the opposite.

3.2. Spatial and Frequency Information Interaction

Preliminary. For a feature map $z \in \mathbb{R}^{\widetilde{H} \times \widetilde{W} \times \widetilde{C}}$, where $\widetilde{H}$, $\widetilde{W}$ and $\widetilde{C}$ denote the height, width and number of channels, respectively. We first project each of its channel $z_{\widetilde{c}}$ to the frequency domain by the Fourier [10] transformation $\mathcal{F}$ as

$$\mathcal{F}(z_{\widetilde{c}})(u, v) = \sum_{h=0}^{\widetilde{H}-1} \sum_{w=0}^{\widetilde{W}-1} z_{\widetilde{c}}(h, w) e^{-j2\pi(\frac{h}{\widetilde{H}}u + \frac{w}{\widetilde{W}}v)}, \quad (3)$$

where (h, w) and (u, v) represent the coordinates in the spatial and frequency domain, respectively. The $\widetilde{c} \in$

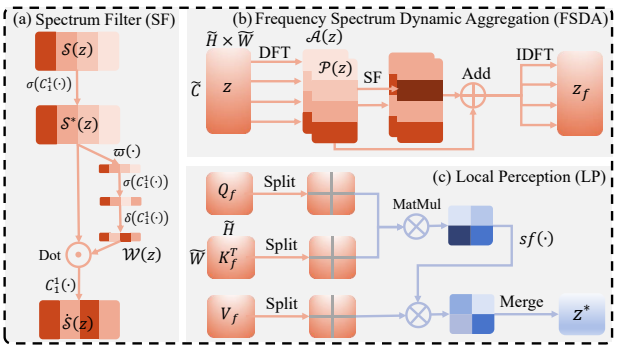


Figure 3. The sub-modules of the proposed SFII.

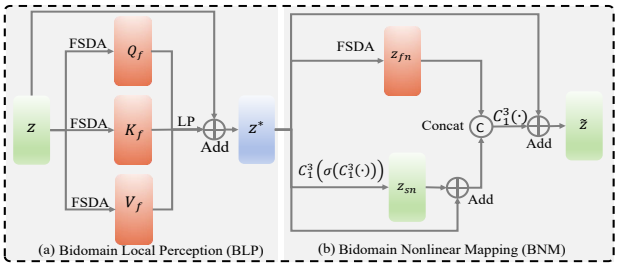


Figure 4. The overall architecture of the proposed SFII.

$\{0, 1, ..., \widetilde{C}\}$ denotes the channel index. Correspondingly, the $\mathcal{F}^{-1}$ is defined as the inverse Fourier transformation [55]. Then, the real part $\mathcal{R}(z_{\widetilde{c}})(u,v)$ and imaginary part $\mathcal{I}(z_{\widetilde{c}})(u,v)$ can be obtained by $\mathcal{F}(z_{\widetilde{c}})(u,v)$. The amplitude spectrum $\mathcal{A}(z_{\widetilde{c}})(u,v)$ and phase spectrum $\mathcal{P}(z_{\widetilde{c}})(u,v)$ of $\mathcal{F}(z_{\widetilde{c}})(u,v)$ on the single channel can be obtained by

$$\mathcal{A}(z_{\widetilde{c}})(u,v) = \sqrt{\mathcal{R}^2(z_{\widetilde{c}})(u,v) + \mathcal{I}^2(z_{\widetilde{c}})(u,v)}, \quad (4)$$

$$\mathcal{P}(z_{\widetilde{c}})(u,v) = \arctan[\frac{\mathcal{I}(z_{\widetilde{c}})(u,v)}{\mathcal{R}(z_{\widetilde{c}})(u,v)}]. \quad (5)$$

The full channel amplitude spectrum $\mathcal{A}(z)(u,v) \in \mathbb{R}^{\widetilde{H}\times\widetilde{W}\times\widetilde{C}}$ and phase spectrum $\mathcal{P}(z)(u,v) \in \mathbb{R}^{\widetilde{H}\times\widetilde{W}\times\widetilde{C}}$ can be obtained by applying the Eq. 3, Eq. 4 and Eq. 5 on each channel of z.

Frequency Spectrum Dynamic Aggregation (FSDA). The haze, glow and noise with inconsistent frequency characteristics can be processed in the frequency domain by dynamic spectrum filter. The amplitude spectrum and phase spectrum of different channels are aggregated by the pointwise convolution as

$$\mathcal{S}^*(z)(u,v) = \sigma(C_1^1(\mathcal{S}(z)(u,v))), \quad (6)$$

where $S(z)(u,v) \in \{\mathcal{A}(z)(u,v), \mathcal{P}(z)(u,v)\}$. To perform channel aggregation of spectral information, the channel weight [12] map $\mathcal{W}$ are calculated as

$$\mathcal{W}(z)(u,v) = \delta(C_1^1(\sigma(C_1^1(\varpi(S^*(z)(u,v)))))), \quad (7)$$

where $\mathcal{W}(z)(u,v) \in \mathbb{R}^{1\times1\times\widetilde{C}}$. Then the channel weight map is applied to the frequency spectrum as

$$\dot{\mathcal{S}}(z)(u,v) = C_1^1(\mathcal{W}(z)(u,v) \cdot \mathcal{S}^*(z)(u,v)), \quad (8)$$

where the spectrum filter (SF) of $\dot{S}(z)(u,v)$ is shown in Figure 3-(a). The filtering operation is performed by the residual connection, the filtered component is obtained by

$$\widetilde{\mathcal{S}}(z)(u,v) = \dot{\mathcal{S}}(z)(u,v) + \mathcal{S}(z)(u,v). \quad (9)$$

The filtered $\widetilde{\mathcal{A}}(z)(u,v)$ and $\widetilde{\mathcal{P}}(z)(u,v)$ can be obtained based on the processing flow from $\mathcal{S}(z)(u,v)$ to $\widetilde{S}(z)(u,v)$. Then, the real and imaginary parts are obtained by

$$\widetilde{\mathcal{R}}(z)(u,v) = \widetilde{\mathcal{A}}(z)(u,v) \cdot \cos\widetilde{\mathcal{P}}(z)(u,v), \quad (10)$$

$$\widetilde{\mathcal{I}}(z)(u,v) = \widetilde{\mathcal{A}}(z)(u,v) \cdot \sin\widetilde{\mathcal{P}}(z)(u,v). \quad (11)$$

After dynamic parameter learning in the frequency domain, we remap the feature map to the spatial domain as

$$z_f = \mathcal{F}^{-1}(\widetilde{\mathcal{R}}(z)(u,v), \widetilde{\mathcal{I}}(z)(u,v)), \quad (12)$$

where $z_f \in \mathbb{R}^{\widetilde{H}\times\widetilde{W}\times\widetilde{C}}$. The Fourier transformation and inverse Fourier transformation can be implemented using DFT and IDFT algorithms [6, 11, 56]. Here, we define the calculation from Eq. 3 to Eq. 12 as frequency spectrum dynamic aggregation (FSDA), which represent the processing flow from z to z_f that is shown in Figure 3-(b). For convenience, the FSDA is denoted as $\mathcal{FS}(\cdot)$.

Frequency Domain Projection (FDP). To deal with distortions in the frequency domain, we first introduce frequency domain interactions before computing local inductive bias. For the input feature map $z \in \mathbb{R}^{\widetilde{H}\times\widetilde{W}\times\widetilde{C}}$, it is processed by the layer normalization operation ($LN(\cdot)$) [32] to obtain the normalized feature $z_l = LN(z)$. Then, the normalized feature z_l is projected into Q_f (query), K_f (key) and V_f (value) by the projection in the frequency domain as

$$Q_f = \mathcal{FS}_Q(z_l), K_f = \mathcal{FS}_K(z_l), V_f = \mathcal{FS}_V(z_l), \quad (13)$$

where the $\mathcal{FS}_Q(\cdot)$, $\mathcal{FS}_K(\cdot)$ and $\mathcal{FS}_V(\cdot)$ denote three independent projection operation with learnable parameters, respectively. The generation process of the Q_f, K_f and V_f is denoted as the frequency domain projection (FDP).

Bidomain Local Perception (BLP). After obtaining the features Q_f, K_f and V_f which consider the information in frequency domain, we perform spatial domain learning on the features from a local perspective. The self-attention [32] with local perception (LP) that is shown in Figure 3-(c) is computed within 8×8 non-overlapping windows as

$$\mathcal{AT}(Q_f, K_f, V_f) = sf(\frac{Q_f \otimes K_f^T}{\sqrt{d}} + B) \otimes V_f, \quad (14)$$

where d and B denote the dimensionality and position bias, respectively. The $\otimes$ denotes the matrix multiplication (MatMul). Information is transferred by the residual connection

$$z^* = \mathcal{AT}(Q_f, K_f, V_f) + z, \tag{15}$$

where the calculation from z to z^* is marked as bidomain local perception (BLP), which is shown in Figure 4-(a).

Bidomain Nonlinear Mapping (BNM). The computation of window attention does not provide nonlinear representation capabilities. Therefore, we use the frequency and spatial domain interaction module to learn nonlinear mapping. The FSDA is used to provide the frequency domain information. Besides, a residual block which consists of $C_1^3(\sigma(C_1^3(\cdot)))$ is used to provide the spatial interaction. The immediate feature z^* is fed into the frequency nonlinear mapping branch and spatial nonlinear mapping branch, as

$$z_{fn} = \mathcal{FS}_A(z^*), \tag{16}$$

$$z_{sn} = C_1^3(\sigma(C_1^3(z^*))), \tag{17}$$

where the subscript A in $\mathcal{FS}_A(\cdot)$ means the frequency interaction performed after the attention operation. Then frequency domain and spatial domain features are fused as the final nonlinear mapping output by

$$\widetilde{z} = C_1^3([z_{fn}, z_{sn} + z^*]) + z^*, \tag{18}$$

where the $[\cdot,\cdot]$ denotes the channel concatenation. The calculation from z^* to $\widetilde{z}$ is marked as the bidomain nonlinear mapping (BNM), which is shown in Figure 4-(b).

Spatial and Frequency Information Interaction (SFII). As shown in Figure 4, the calculation process from z to $\widetilde{z}$ is called spatial and frequency information interaction (SFII). The proposed SFII aggregates spatial domain information and frequency domain information from a local perspective.

3.3. Spatial and Frequency Loss

The supervised loss consists of two parts, namely the pixel-by-pixel loss in geometric space and the frequency domain loss obtained by Fourier transform [3]. By sampling $x_i^s \in \mathcal{D}_X$ and $y_i^s \in \mathcal{D}_Y$, the losses calculated at three scales are

$$\mathcal{L}_G = \sum_{s=0}^{2} \lambda_g \cdot \sum_{i=0}^{N-1} ||\Psi^s(x_i^s) - y_i^s||_1, \tag{19}$$

$$\mathcal{L}_F = \sum_{s=0}^{2} \lambda_f \cdot \sum_{i=0}^{N-1} ||\mathcal{F}(\Psi^s(x_i^s)) - \mathcal{F}(y_i^s)||_1, \tag{20}$$

where λ_g and λ_f denote weight factors.

3.4. Retraining and Realistic Brightness Loss

Pseudo-label Fusion Retraining. There are inherent domain discrepancy between synthetic hazy images and real-world hazy images. Therefore, we adopt a retraining strategy which utilizes pesudo labels. Pseudo labels $\mathcal{R}_Y^P$ are obtained based on the model trained on synthetic datasets. We put the original synthetic dataset $\{\mathcal{D}_X, \mathcal{D}_Y\}$ and the pseudo-labeled dataset $\{\mathcal{R}_X, \mathcal{R}_Y^P\}$ into the network simultaneously for retraining. Supervised losses Eq. 19 and Eq. 20 are used in the retraining process at three scales.

Prior Brightness Constraint. We conduct a quantitative statistics on the brightness of nighttime hazy and clear images provided by [14]. The brightness intensity corresponding to $x_i^0 \in \mathcal{R}_X$ and $y_i^0 \in \mathcal{R}_Y$ are $\mu(x_i^0)$ and $\mu(y_i^0)$, respectively, where $\mu(\cdot)$ denote the average pixel value across three channels. We randomly select $\hat{M} = \frac{M}{2}$ images from the dataset multiple times, and we get

$$\sum_{i=0}^{\hat{M}-1} \mu(y_i^0) < \sum_{i=0}^{\hat{M}-1} \mu(x_i^0). \tag{21}$$

Therefore, we assume the brightness of the dehazed image p_i^s should be lower than that of the x_i^s. This assumption is consistent with the imaging model Eq. 2.

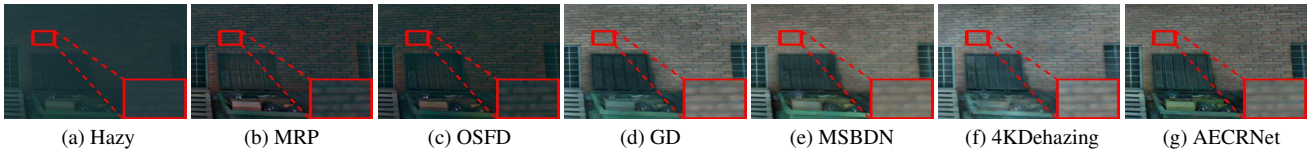
Local Brightness Map (LBM). We divide the image into non-overlapping local windows. The width and height of each square window is denoted as γ^s, where $s \in \{0, 1, 2\}$. The value in local brightness map (LBM) $\varphi_{x_i^s}$ that corresponding to x_i^s is obtained by

$$\varphi_{x_i^s}(\hat{h}, \hat{w}) = \frac{1}{3(\gamma^s)^2} \sum_{c=0}^{2} \sum_{h=\hat{h}\cdot\gamma^s}^{(\hat{h}+1)\cdot\gamma^s} \sum_{w=\hat{w}\cdot\gamma^s}^{(\hat{w}+1)\cdot\gamma^s} x_i^s(h, w, c), \tag{22}$$

where $(\hat{h}, \hat{w})$ and (h, w) denote the pixel index of $\varphi_{x_i^s}$ and x_i^s, respectively. Meanwhile, the local brightness map $\varphi_{p_i^s}$ corresponding to p_i^s is defined in the same way. As shown in Figure 2-(c), the locations with high brightness may be active light sources or objects close to the light source, while the locations with low brightness may be objects and backgrounds far away from the light source.

Realistic Brightness Loss. The brightness of hazy images is approximately globally realistic, so it can be used to supervise the brightness of dehazed images. As we observed in Eq. 21, the brightness of the dehazed image should be lower than that of the hazy image. Meanwhile, in order to ensure the relative numerical relationship between areas with high brightness and low brightness before and after dehazing, we use a power function with monotonically increasing properties to process the $\varphi_{x_i^s}(\hat{h}, \hat{w})$, as

$$\widetilde{\varphi}_{x_i^s}(\hat{h}, \hat{w}) = (\varphi_{x_i^s}(\hat{h}, \hat{w}))^{\kappa}, \tag{23}$$



(a) Hazy (b) MRP (c) OSFD (d) GD (e) MSBDN (f) 4KDehazing (g) AECRNet

(h) DeHamer (i) FSDGN (j) DF (k) MITNet (l) Fourmer (m) Ours (n) Label

Figure 5. Visual results on synthetic dataset [31].

Table 1. Quantitative results on datasets that generated by imaging model.

Methods	NHR		NHM		NHCL		NHCM		NHCD		NightHaze		YellowHaze	
	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑
MRP (CVPR 2017)	0.776	19.848	0.666	15.993	0.747	22.497	0.693	20.494	0.624	17.651	0.295	12.138	0.249	13.473
GD (ICCV 2019)	0.969	30.107	0.861	20.689	0.973	36.506	0.958	34.448	0.932	31.509	0.832	25.324	0.915	27.410
OSFD (ACMMM 2020)	0.808	21.028	0.722	18.491	0.786	22.329	0.739	20.929	0.672	18.501	0.304	13.387	0.259	14.775
MSBDN (CVPR2020)	0.970	31.335	0.818	20.514	0.965	35.963	0.938	32.848	0.903	30.475	0.950	33.156	0.921	29.834
4KDehazing (CVPR2021)	0.950	28.613	0.830	20.429	0.967	35.006	0.958	35.162	0.912	30.048	0.850	26.562	0.861	25.835
AECRNet (CVPR 2021)	0.915	24.864	0.817	19.420	0.951	33.183	0.943	33.498	0.890	28.742	0.946	32.344	0.937	29.417
DeHamer (CVPR 2022)	0.966	31.017	0.823	23.095	0.966	36.038	0.944	33.908	0.915	31.389	0.954	33.432	0.931	30.334
FSDGN (ECCV 2022)	0.975	32.072	0.874	21.415	0.972	36.432	0.952	33.723	0.922	31.559	0.948	33.521	0.955	**33.062**
DF (TIP 2023)	0.969	31.644	0.896	23.207	0.975	37.383	0.960	35.038	0.934	32.079	0.931	31.489	0.948	32.244
MITNet (ACMMM 2023)	0.974	31.969	0.859	20.884	0.969	35.794	0.945	32.849	0.916	30.628	0.946	34.114	0.932	31.186
Fourmer (ICML 2023)	0.969	31.660	0.862	21.423	0.963	35.714	0.943	33.201	0.928	32.103	0.949	33.419	0.958	31.978
Ours	**0.978**	**33.180**	**0.905**	**23.705**	**0.979**	**38.146**	**0.968**	**36.146**	**0.951**	**34.001**	**0.968**	**35.527**	**0.965**	32.981

where $\kappa \geq 1$ is the brightness intensity coefficient. The realistic brightness constraint within one single window is

$$\mathcal{L}_B^{p_i^s}(\hat{h},\hat{w}) = (\varphi_{p_i^s}(\hat{h},\hat{w}) - \xi \cdot \widetilde{\varphi}_{x_i^s}(\hat{h},\hat{w}))^2, \tag{24}$$

where ξ is a hyperparameter. The realistic brightness loss calculated over all windows is

$$\mathcal{L}_B = \sum_{s=0}^{2} \frac{\lambda_b}{\hat{N}\hat{W}^s\hat{H}^s} \cdot \sum_{i=0}^{\hat{N}-1} \sum_{\hat{h}=0}^{\hat{H}^s-1} \sum_{\hat{w}=0}^{\hat{W}^s-1} \mathcal{L}_B^{p_i^s}(\hat{h},\hat{w}), \tag{25}$$

where $\hat{W}^s = W^s/\gamma^s$, $\hat{H}^s = H^s/\gamma^s$. And $\hat{N} = N + M$. The λ_b denotes the weights of scale loss of $\mathcal{L}_B$.

3.5. Total Loss

The overall loss is a combination of supervised and semi-supervised losses, which is

$$\mathcal{L} = \mathcal{L}_G + \alpha\mathcal{L}_F + \beta\mathcal{L}_B, \tag{26}$$

where α and β are the weights of the frequency domain loss and the realistic brightness loss, respectively.

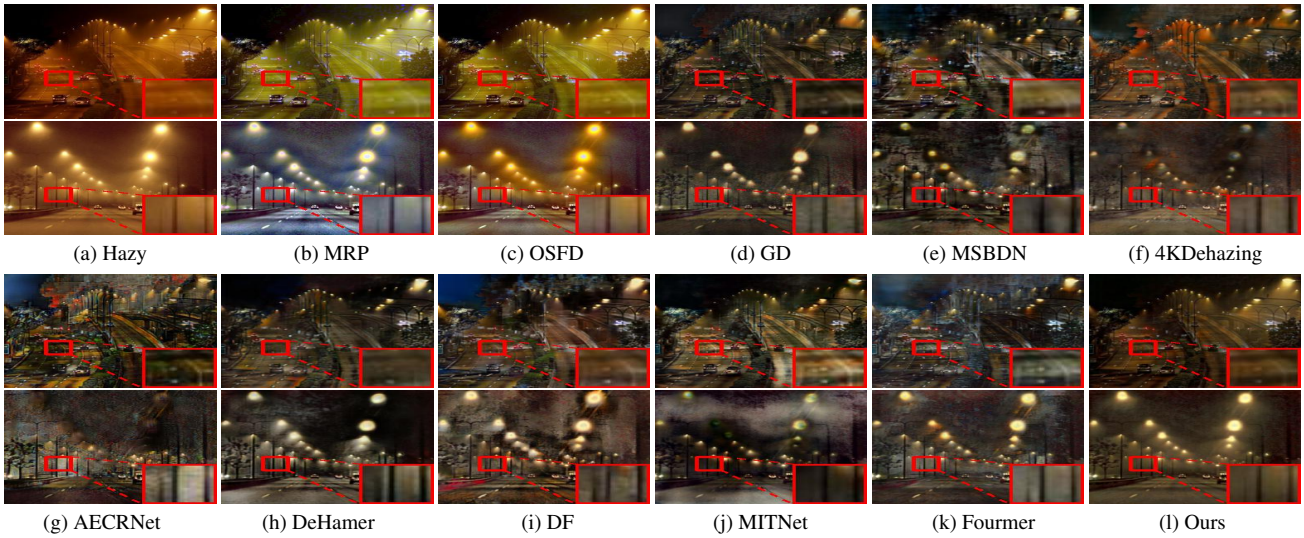
4. Experiments

4.1. Experiment Setting

Datasets. To comprehensively compare the performance of different algorithms, we conducted experiments on both synthetic and real-world datasets. The synthetic datasets include GTA5 [45], UNREAL-NH [31], {NHR, NHM, HNCL, NHCM, NHCD} [51] and {NightHaze, YellowHaze} [26]. The real-world nighttime haze (RWNH) is provided by [14]. Since the brightness level of the ground-truth label in the UNREAL-NH is close to daytime, we adjust the brightness of the hazy image and corresponding label to the level of the nighttime low-light image by the Gamma correction [33] for the evaluation of the RWNH.

Comparison Methods and Evaluation Metrics. MRP [50], GD [27], OSFD [51], MSBDN [5]. 4KDehazing [54], AECRNet [43], DeHamer [9], FSDGN [47], DF [38], MITNet [37] and Fourmer [55] are used as comparisons. PSNR [22, 35, 36] and SSIM [8, 42] are used to evaluate the performance on labeled datasets. BRISQUE [44] and MUSIQ [13, 17] are computed to evaluate the performance on unlabeled dataset. The ↑ represents a larger value, a higher quality, while ↓ represents a larger value, a lower quality.

Implementation Details. The batch size is chosen as 4. The image size is set to $256 \times 256 \times 3$. The learning rate is initialized to 0.0001 and linearly decays by a factor of 0.95 every 10 epochs. The Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$) is used. The λ_g, λ_f and λ_b are all set to 1. The α and β are set to 0.1 and 20, respectively. The window size γ^s are set to 16, 8 and 4, where $s \in \{0, 1, 2\}$, respectively. The coefficient ξ and κ is set to 1 and 1.3, respectively. The



(a) Hazy (b) MRP (c) OSFD (d) GD (e) MSBDN (f) 4KDehazing

(g) AECRNet (h) DeHamer (i) DF (j) MITNet (k) Fourmer (l) Ours

Figure 6. Visual results on real-world hazy images [14].

(a) Hazy (b) $\kappa = 1.0$ (c) $\kappa = 1.3$ (d) $\kappa = 1.5$ (e) $\kappa = 1.8$ (f) $\kappa = 2.0$ (g) $\kappa = 2.3$ (h) $\kappa = 2.5$ (i) $\kappa = 2.8$ (j) $\kappa = 3.0$

Figure 7. Dehazed images obtained under different κ.

proposed model is implemented by PyTorch and trained on the single NVIDIA RTX 4090 platform.

4.2. Comparison with State-of-the-art Algorithms

Evaluation on Synthetic Datasets. Table 1 and Table 2 show the quantitative dehazing results obtained by state-of-the-art methods. Figure 5 shows the corresponding visual results. The quantitative and visual results demonstrate that the proposed methods achieve an overall better performance than state-of-the-art algorithms.

Evaluation on Real-world Datasets. Table 2 shows the evaluation results of real-world dehazed images. It is worth pointing out that existing research [7] proposes that the reliability of no-reference metrics in the dehazing task is lower than that of full-reference metrics. Figure 6 shows that the details of the dehazed results obtained by our method are visually better. Meanwhile, the brightness of the dehazed images obtained by most comparison algorithms is obvious unrealistic, while the brightness of the dehazed images obtained by our algorithm is approximately globally realistic.

Overall Evaluation. According to the quantitative and visual results on synthetic and real-world datasets, the proposed SFSNiD achieves overall better performance. More results are placed at Supplementary Materials.

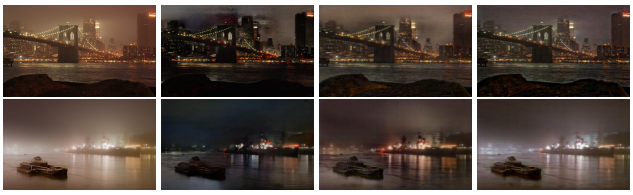
4.3. Ablation Study and Discussions

Spatial and Frequency Information Interaction. The spatial and frequency information interaction (SFII) modules and naive convolution module are used in the proposed SFSNiD. In order to prove the usefulness of the FDP, LP and BNM that contained in the SFII, ablation experiments for different sub-blocks are performed. The ablation experiment on the proposed SFII includes (i) removing the FDP, (ii) removing the LP, (iii) removing the frequency domain processing in BNM, and (iv) removing the spatial domain process in BNM. These four settings are denoted $R1$, $R2$, $R3$ and $R4$, respectively. Table 3 shows the ablation results under different settings on the UNREAL-NH [31]. The quantitative results demonstrate that the FDP, LP and BNM all have a positive effect on the dehazing performance. Since we must control the size of the paper, visualizations of the amplitude and phase spectrums are placed in Supplementary Materials.

Hierarchical Training and Frequency Domain Loss. The training process of the proposed SFSNiD takes a hierarchical strategy by using differ scales $s \in \{0, 1, 2\}$. Two

Table 2. Quantitative results on datasets generated by game engine (GTA5 and UNREAL-NH) and the real-world dataset (RWNH).

Methods	GTA5		UNREAL-NH		RWNH	
	SSIM↑	PSNR↑	SSIM↑	PSNR↑	BRISQUE↓	MUSIQ↑
MRP	0.662	19.460	0.467	10.039	**19.418**	41.194
GD	0.900	30.090	0.767	21.202	31.359	33.433
OSFD	0.711	21.461	0.443	9.169	20.860	**41.779**
MSBDN	0.909	32.029	0.827	25.680	38.910	29.968
4KDehazing	0.903	30.314	0.774	23.087	34.965	33.536
AECRNet	0.888	26.846	0.731	21.566	27.084	37.034
DeHamer	0.928	32.597	0.740	22.441	42.269	26.788
FSDGN	0.923	32.642	0.702	21.736	32.216	35.200
DF	0.918	32.856	0.770	23.017	33.678	31.663
MITNet	0.899	31.118	0.766	21.860	35.404	31.768
Fourmer	0.917	31.926	0.772	22.799	35.850	31.367
Ours	**0.935**	**33.708**	**0.862**	**25.907**	30.975	32.120



(a) Hazy (b) Pseudo Label (c) Retraining (d) Retraining + $\mathcal{L}_B$

Figure 8. Visual results under different training strategies.

ablation studies are adopted, which are denoted as (i) $S1$: $s \in \{0\}$, and (ii) $S2$: $s \in \{0, 1\}$. Meanwhile, in our experimental setup, the spatial domain loss L_G and the frequency domain loss L_F are applied simultaneously. To verify the effectiveness of frequency domain loss, the setting when L_F is not used is denoted as $S3$ ($s \in \{0, 1, 2\}$). Table 4 shows the ablation results under the three different settings. The quantitative results demonstrate two main conclusions. First, the hierarchical training strategy can improve the dehazing performance. Second, the loss in the frequency domain is crucial as it improves the SSIM from 0.816 to 0.862.

Retraining Strategy and Realistic Brightness Loss. To verify the effectiveness of the retraining strategy and the realistic brightness loss $\mathcal{L}_B$, the visual effects are shown in Figure 8. As shown in Figure 8-(b), the texture of the pseudo-labels is blurred due to the domain discrepancy between the synthetic and real-world data. The dehazed images obtained after retraining has unrealistic brightness as shown in Figure 8-(c). It can be seen that the best effect occurs when the retraining strategy and $\mathcal{L}_B$ are used simultaneously as shown in Figure 8-(d). The BRISQUE (↓) and MUSIQ (↑) obtained for the three settings (b), (c) and (d) in Figure 8 are $\{33.316, 30.432\}$, $\{34.210, 32.373\}$ and $\{30.975, 32.120\}$, respectively. Taking a comprehensive look at the visual and quantitative evaluation results, our proposed strategy is effective.

Brightness intensity coefficient κ in $\mathcal{L}_B$. In order to demonstrate the effectiveness of κ on the real-world dehazing task, we manually set κ to different values. The dehazed images and average pixel value when κ takes different values are shown in Figure 7 and Figure 9, respectively. There are two conclusions that can be drawn. First, as κ increases, the brightness of the dehazed image continues to decrease, which proves that κ can control the brightness of the dehazed image. Second, when κ equals 1.3, the average pixel value (0.225) of dehazed images is close to the average pixel value real-world nighttime clear images (0.217) [14]. Therefore, we set κ to 1.3 as the final setting.

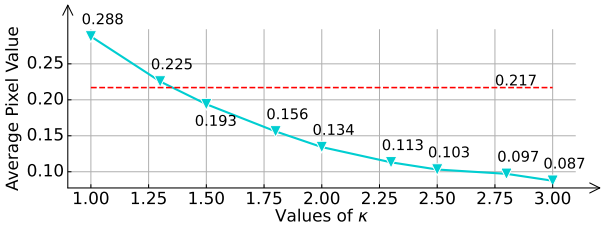


Figure 9. The average pixel value obtained under different κ. The horizontal dashed line represents the average pixel value of real-world nighttime clear images [14].

Table 3. Ablation study on the SFII.

Settings	$R1$	$R2$	$R3$	$R4$	Ours
SSIM	0.848	0.858	0.851	0.845	**0.862**
PSNR	25.353	25.808	25.642	24.301	**25.907**

Table 4. Ablation study on the scale loss and frequency loss.

Settings	$S1$	$S2$	$S3$	Ours
SSIM	0.854	0.851	0.816	**0.862**
PSNR	25.601	25.134	24.464	**25.907**

5. Conclusion

In this paper, a semi-supervised nighttime image dehazing baseline SFSNiD is proposed for real-world nighttime dehazing. A spatial and frequency domain information interaction module is proposed to handle the haze, glow, and noise with localized, coupled and frequency inconsistent characteristics. A retraining strategy and a local window-based brightness loss for semi-supervised training process are designed to suppress haze and glow while achieving realistic brightness. Experiments on public benchmarks validate the effectiveness of the proposed method and its superiority over state-of-the-art methods.

Acknowledgment. This work was supported in part by the grant of the National Science Foundation of China under Grant 62172090; Start-up Research Fund of Southeast University under Grant RF1028623097; CAAI-Huawei MindSpore Open Fund. We thank the Big Data Computing Center of Southeast University for providing the facility support on the numerical calculations in this paper.

References

[1] Cosmin Ancuti, Codruta O Ancuti, Christophe De Vleeschouwer, and Alan C Bovik. Day and night-time dehazing by local airlight estimation. *IEEE Transactions on Image Processing*, 29:6264–6275, 2020. 2

[2] Xiaofeng Cong, Jie Gui, Kai-Chao Miao, Jun Zhang, Bing Wang, and Peng Chen. Discrete haze level dehazing network. In *ACM International Conference on Multimedia*, pages 1828–1836, 2020. 1

[3] Yuning Cui, Yi Tao, Zhenshan Bing, Wenqi Ren, Xinwei Gao, Xiaochun Cao, Kai Huang, and Alois Knoll. Selective frequency network for image restoration. In *The Eleventh International Conference on Learning Representations*, 2022. 3, 5

[4] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological nighttime flare removal dataset. *Advances in Neural Information Processing Systems*, 35:3926–3937, 2022. 2

[5] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2157–2167, 2020. 6

[6] Matteo Frigo and Steven G Johnson. Fftw: An adaptive software architecture for the fft. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1381–1384, 1998. 4

[7] Jie Gui, Xiaofeng Cong, Yuan Cao, Wenqi Ren, Jun Zhang, Jing Zhang, Jiuxin Cao, and Dacheng Tao. A comprehensive survey and taxonomy on single image dehazing based on deep learning. *ACM Computing Surveys*, 2023. 1, 2, 7

[8] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 6

[9] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5812–5820, 2022. 2, 6

[10] Xin Guo, Xueyang Fu, Man Zhou, Zhen Huang, Jialun Peng, and Zheng-Jun Zha. Exploring fourier prior for single image rain removal. In *International Joint Conferences on Artificial Intelligence*, pages 935–941, 2022. 3

[11] Junming Hou, Qi Cao, Ran Ran, Che Liu, Junling Li, and Liang-jian Deng. Bidomain modeling paradigm for pansharpening. In *ACM International Conference on Multimedia*, pages 347–357, 2023. 4

[12] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018. 4

[13] Shirui Huang, Keyan Wang, Huan Liu, Jun Chen, and Yunsong Li. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 18145–18155, 2023. 6

[14] Yeying Jin, Beibei Lin, Wending Yan, Wei Ye, Yuan Yuan, and Robby T Tan. Enhancing visibility in nighttime haze images using guided apsf and gradient adaptive convolution. In *ACM International Conference on Multimedia*, 2023. 1, 2, 3, 5, 6, 7, 8

[15] Mingye Ju, Can Ding, Charles A Guo, Wenqi Ren, and Dacheng Tao. Idrlp: Image dehazing using region line prior. *IEEE Transactions on Image Processing*, 30:9043–9057, 2021. 1

[16] Mingye Ju, Can Ding, Wenqi Ren, Yi Yang, Dengyin Zhang, and Y Jay Guo. Ide: Image dehazing and exposure using an enhanced atmospheric scattering model. *IEEE Transactions on Image Processing*, 30:2180–2192, 2021. 2

[17] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *IEEE International Conference on Computer Vision*, pages 5148–5157, 2021. 6

[18] Beomhyuk Koo and Gyeonghwan Kim. Nighttime haze removal with glow decomposition using gan. In *Pattern Recognition: 5th Asian Conference*, pages 807–820, 2020. 2

[19] Shiba Kuanar, Dwarikanath Mahapatra, Monalisa Bilas, and KR Rao. Multi-path dilated convolution network for haze and glow removal in nighttime images. *The Visual Computer*, pages 1–14, 2022. 1

[20] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *IEEE International Conference on Computer Vision*, pages 4770–4778, 2017. 1

[21] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018. 1

[22] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv preprint arXiv:2302.11831*, 2023. 2, 6

[23] Kun Li, Dan Guo, and Meng Wang. Proposal-free video grounding with contextual pyramid network. In *AAAI Conference on Artificial Intelligence*, pages 1902–1910, 2021. 3

[24] Yu Li, Robby T Tan, and Michael S Brown. Nighttime haze removal with glow and multiple light colors. In *IEEE International Conference on Computer Vision*, pages 226–234, 2015. 2

[25] Yudong Liang, Bin Wang, Wangmeng Zuo, Jiaying Liu, and Wenqi Ren. Self-supervised learning and adaptation for single image dehazing. In *International Joint Conference on Artificial Intelligence*, pages 1–15, 2022. 1

[26] Yinghong Liao, Zhuo Su, Xiangguo Liang, and Bin Qiu. Hdp-net: Haze density prediction network for nighttime dehazing. In *Pacific Rim Conference on Multimedia*, pages 469–480, 2018. 1, 6

[27] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *IEEE International Conference on Computer Vision*, pages 7314–7323, 2019. 1, 6

[28] Yun Liu, Anzhi Wang, Hao Zhou, and Pengfei Jia. Single nighttime image dehazing based on image decomposition. *Signal Processing*, 183:107986, 2021. 1

[29] Yun Liu, Zhongsheng Yan, Jinge Tan, and Yuche Li. Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(4):1643–1657, 2022. 2

[30] Yun Liu, Zhongsheng Yan, Aimin Wu, and Tian Ye. Nighttime image dehazing based on variational decomposition model. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 640–649, 2022. 2

[31] Yun Liu, Zhongsheng Yan, Sixiang Chen, Tian Ye, Wenqi Ren, and Erkang Chen. Nighthazeformer: Single nighttime haze removal using prior query transformer. In *ACM International Conference on Multimedia*, 2023. 1, 2, 3, 6, 7

[32] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *IEEE International Conference on Computer Vision*, pages 10012–10022, 2021. 2, 4

[33] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Ming-Hsuan Yang. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 28(9):4364–4375, 2019. 6

[34] Wenqi Ren, Jinshan Pan, Hua Zhang, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks with holistic edges. *International Journal of Computer Vision*, 128:240–259, 2020. 1

[35] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2808–2817, 2020. 6

[36] Hao Shen, Zhong-Qiu Zhao, and Wandi Zhang. Adaptive dynamic filtering network for image denoising. In *AAAI Conference on Artificial Intelligence*, pages 2227–2235, 2023. 6

[37] Hao Shen, Zhong-Qiu Zhao, Yulun Zhang, and Zhao Zhang. Mutual information-driven triple interaction network for efficient image dehazing. In *ACM International Conference on Multimedia*, pages 7–16, 2023. 2, 6

[38] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE TIP*, 32:1927–1941, 2023. 1, 2, 6

[39] Shangquan Sun, Wenqi Ren, and Tao Wang. Rethinking image restoration for object detection. *Advances in Neural Information Processing Systems*, 35:4461–4474, 2022. 1

[40] Fei Wang, Dan Guo, and Kun Li. Eulermormer: Robust eulerian motion magnification via dynamic filtering within transformer. *arXiv preprint arXiv:2312.04152*, 2023. 3

[41] Wenhui Wang, Anna Wang, and Chen Liu. Variational single nighttime image haze removal with a gray haze-line prior. *IEEE Transactions on Image Processing*, 31:1349–1363, 2022. 2

[42] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

[43] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 1, 2, 6

[44] Rui-Qi Wu, Zheng-Peng Duan, Chun-Le Guo, Zhi Chai, and Chongyi Li. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 22282–22291, 2023. 6

[45] Wending Yan, Robby T Tan, and Dengxin Dai. Nighttime defogging using high-low frequency decomposition and grayscale-color networks. In *European Conference on Computer Vision*, pages 473–488, 2020. 3, 6

[46] Minmin Yang, Jianchang Liu, and Zhengguo Li. Superpixel-based single nighttime image haze removal. *IEEE Transactions on Multimedia*, 20(11):3008–3018, 2018. 2

[47] Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao. Frequency and spatial dual guidance for image dehazing. In *European Conference on Computer Vision*, pages 181–198, 2022. 6

[48] Jing Zhang and Dacheng Tao. Famed-net: A fast and accurate multi-scale end-to-end dehazing network. *IEEE Transactions on Image Processing*, 29:72–84, 2019. 1

[49] Jing Zhang, Yang Cao, and Zengfu Wang. Nighttime haze removal based on a new imaging model. In *IEEE International Conference on Image Processing*, pages 4557–4561, 2014. 1, 2

[50] Jing Zhang, Yang Cao, Shuai Fang, Yu Kang, and Chang Wen Chen. Fast haze removal for nighttime image using maximum reflectance prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7418–7426, 2017. 1, 3, 6

[51] Jing Zhang, Yang Cao, Zheng-Jun Zha, and Dacheng Tao. Nighttime dehazing with a synthetic benchmark. In *ACM International Conference on Multimedia*, pages 2355–2363, 2020. 1, 3, 6

[52] Jingang Zhang, Wenqi Ren, Shengdong Zhang, He Zhang, Yunfeng Nie, Zhe Xue, and Xiaochun Cao. Hierarchical density-aware dehazing network. *IEEE Transactions on Cybernetics*, 52(10):11187–11199, 2021. 1

[53] Shengdong Zhang, Wenqi Ren, Xin Tan, Zhi-Jie Wang, Yong Liu, Jingang Zhang, Xiaoqin Zhang, and Xiaochun Cao. Semantic-aware dehazing network with adaptive feature fusion. *IEEE Transactions on Cybernetics*, 53(1):454–467, 2021. 1

[54] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 16180–16189, 2021. 1, 2, 6

[55] Man Zhou, Jie Huang, Chun-Le Guo, and Chongyi Li. Fourmer: an efficient global modeling paradigm for image restoration. In *International Conference on Machine Learning*, pages 42589–42601, 2023. 4, 6

[56] Man Zhou, Keyu Yan, Xueyang Fu, Aiping Liu, and Chengjun Xie. Pan-guided band-aware multi-spectral feature enhancement for pan-sharpening. *IEEE Transactions on Computational Imaging*, 9:238–249, 2023. 4