

Frequency and Spatial Dual Guidance for Image Dehazing

Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao*

University of Science and Technology of China
{yuhu520,nszheng,manman,hj0117,zeyuxiao}@mail.ustc.edu.cn,
fzhao956@ustc.edu.cn

Abstract. In this paper, we propose a novel image dehazing framework with frequency and spatial dual guidance. In contrast to most existing deep learning-based image dehazing methods that primarily exploit the spatial information and neglect the distinguished frequency information, we introduce a new perspective to address image dehazing by jointly exploring the information in the frequency and spatial domains. To implement frequency and spatial dual guidance, we delicately develop two core designs: amplitude guided phase module in the frequency domain and global guided local module in the spatial domain. Specifically, the former processes the global frequency information via deep Fourier transform and reconstructs the phase spectrum under the guidance of the amplitude spectrum, while the latter integrates the above global frequency information to facilitate the local feature learning in the spatial domain. Extensive experiments on synthetic and real-world datasets demonstrate that our method outperforms the state-of-the-art approaches both visually and quantitatively. Our code is released publicly at <https://github.com/yuhuUSTC/FSDGN>

Keywords: Image Dehazing, Frequency and Spatial Dual-Guidance, Amplitude and Phase

1 Introduction

Haze is a common atmospheric phenomenon, which is composed of tiny water droplets or ice crystals suspended in the air near the ground. Images captured in hazy environments usually have noticeable visual quality degradation in object appearance and contrast. The goal of image dehazing is to restore a clean scene from a hazy image. The performance of high-level computer vision tasks such as object detection [10,27] and scene understanding [44,43] are considerably influenced by the input images captured in hazy scenes. Thus, throughout the last decade, restoring clear photographs from hazy ones has been a focus of research in the computational photography and vision communities.

Estimating the clean image from a single hazy input is an ill-posed and challenging problem. Conventional approaches rely on the physical scattering

* Corresponding author.

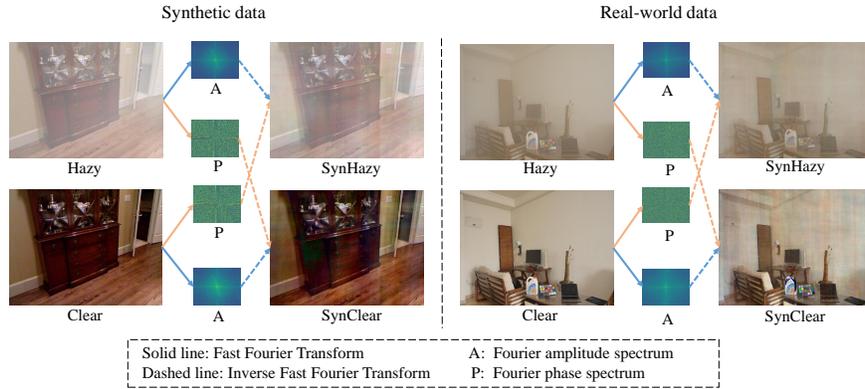


Fig. 1. Visualization on the relationship between the haze degradation and the characteristics of amplitude spectrum and phase spectrum in the frequency domain. We denote the image with clear image amplitude and hazy image phase as **SynClear**, and the image with hazy image amplitude and clear image phase as **SynHazy**.

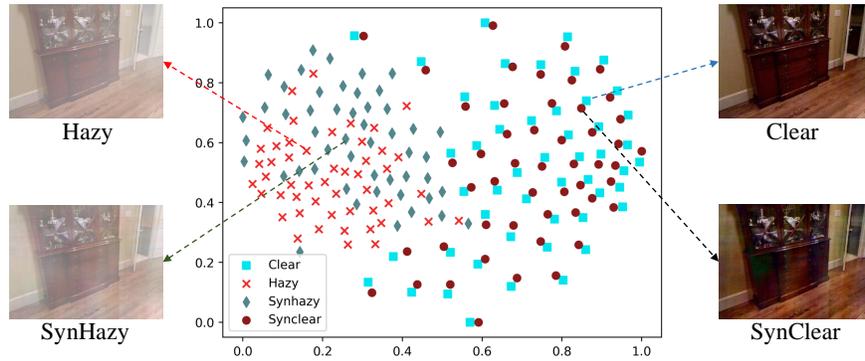


Fig. 2. The t-SNE map of hazy, clear, **SynHazy**, and **SynClear** images. Obviously, clear and **SynClear** images are tightly connected and coupled, indicating more similar distributions. Similarly, hazy and **SynHazy** images are clustered together.

model [35] and regularize the solution space using a variety of crisp image priors [5,19,16,15]. However, these hand-crafted image priors are created based on specific observations, which may not be reliable to model the intrinsic features of images or estimate the transmission map in the physical scattering model.

Inspired by the success of deep learning, numerous deep learning-based approaches [6,26,41,32,37,18,30,13,49] have been developed recently to learn the translation from hazy image to clear image in an end-to-end manner. Although these methods have made remarkable progress in image dehazing tasks, they have a main limitation: they primarily exploit the spatial information and neglect the distinguished frequency information. Compared to spatial domain processing, the difference between hazy and clear image pairs in the frequency domain is physically definite. Thus, finding the correlation between haze degradation and frequency is of great importance for understanding the dehazing problem.

In this paper, we reveal the relationship between the haze degradation and the characteristics of the amplitude and phase spectrums in the frequency domain (see Fig. 1). Specifically, we first transform a spatial domain image to frequency-domain amplitude and phase spectrums by fast Fourier transform, and then exchange the amplitude and phase spectrums of hazy and clear image pairs. Finally, the exchanged spectrums are transformed back to get **SynClear** and **SynHazy** images by inverse fast Fourier transform. From Fig. 1, we can see that: (1) Clear and **SynClear** images have the same amplitude spectrum but different phase spectrums, and they look similar; (2) Clear and **SynHazy** images have the same phase spectrum but different amplitude spectrums, and they look different. This observation leads to the conclusion that: (1) the degradation property induced by haze is mainly manifested in the amplitude spectrum; and (2) the difference between phase spectrums of hazy and clear image pairs is small.

To further explain and testify this conclusion, we show the t-SNE map of 50 groups of hazy, clear, **SynClear**, and **SynHazy** images in Fig. 2. It is apparent that clear and **SynClear** images are clustered together, indicating highly similar distributions. This also applies to hazy and **SynHazy** images.

Based on the above observation and conclusion, we propose a novel Frequency and Spatial Dual-Guidance Network (FSDGN) for single-image dehazing. From a new perspective, we address image dehazing by jointly exploring the information in the frequency and spatial domains. To implement the frequency and spatial dual guidance, we delicately develop two core designs, i.e., Amplitude Guided Phase (AGP) module in the frequency domain and Global Guided Local (GGL) module in the spatial domain. Specifically, the AGP module processes the global frequency information via deep Fourier transform and reconstruct the phase spectrum under the guidance of the amplitude spectrum, while the GGL module integrates the above global frequency information to facilitate the local feature learning in the spatial domain. Thanks to the frequency property observation and our finely constructed modules, our method achieves state-of-the-art (SOTA) performance efficiently. In Fig. 3, several SOTA models are shown in terms of performance, parameters, and FLOPs.

In conclusion, the main contributions of our work are as follows:

- We reveal the correlation between haze degradation and the statistical properties of amplitude and phase spectrums, and integrate the frequency and spatial information for image dehazing.
- We propose a novel Frequency and Spatial Dual-Guidance Network for effectively generating high-quality haze-free images by completely utilizing dual guidance in both the frequency and spatial domains. To the best of our knowledge, we are the first to introduce amplitude and phase spectrums for the image dehazing task.
- We propose the tailor designed GGL and AGP modules for the spatial-domain and frequency-domain guidance, respectively.
- Extensive experiments demonstrate that our method outperforms state-of-the-art approaches with fewer parameters and FLOPs.

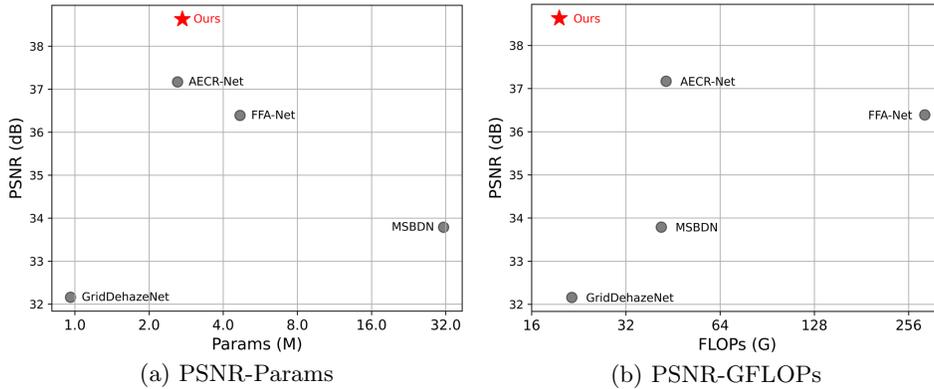


Fig. 3. Trade-off between PSNR, number of parameters, and FLOPs. FLOPs are calculated with an input size 256×256 .

2 Related Work

Single Image Dehazing. In recent years, we have witnessed significant advances in single image dehazing. Existing methods can be roughly categorized into two classes: physical-based methods and deep learning-based methods.

Physical-based methods depend on the physical model [35] and the hand-craft priors from empirical observations, such as dark channel prior [19], color line prior [16], color attenuation prior [57], sparse gradient prior [8], maximum reflectance prior [55], and non-local prior [5]. However, the density of haze can be affected by various factors including temperature, altitude, and humidity, which make the haze formation at individual spatial locations space-variant and non-homogeneous. Therefore, the haze usually cannot be accurately characterized by merely a single transmission map.

Different from the physical-based methods, deep learning-based methods employ convolution neural networks to learn the image prior [6,26,40,54,31,33] or directly learn hazy-to-clear translation [41,32,38,14,12,22,13,37,49,51,47,56,17]. For example, AOD-Net [26] produces the recovered images by reformulating the physical scattering model. MSBDN [13] proposes a boosted decoder to progressively restore the haze-free images. Ye et al. [51] developed perceiving and modeling density for uneven haze distribution to increase its generalization ability on real-world hazy images. AECR-Net [49] introduces the contrastive regularization to exploit both the information of hazy images and clear images as negative and positive samples, respectively. The above techniques have shown outstanding performance on image dehazing. However, they only utilize the information in the spatial domain, which cannot sufficiently model the characteristics of haze degradation. It is necessary to mention that DW-GAN [17] also employs the frequency information, but it works in the wavelet domain and exploits the low-high frequency property. Instead, our method works in the Fourier domain and reveals the relationship between amplitude-phase and haze degradation.

Applications of Fourier Transform. In recent years, some algorithms [11,50,39,46,58,34,7] have been proposed to extract the information from the frequency domain to address different tasks. For instance, DeepRFT [34] applies the convolution operations to the real and imaginary parts of the spectrum in the frequency domain to restore the blurry images. FDIT [7] decomposes the images into low-frequency and high-frequency components to enhance the image generation process. However, existing frequency-based methods neglect to build the relationship between the frequency property and the image degradation. Different from these existing techniques, we further discover the correlation between haze degradation and the characteristics of amplitude and phase spectrums in the frequency domain and tailor design the AGP module to exploit this observation.

3 Method

3.1 Motivation

Our main inspiration comes from observing the relationship between haze degradation and the characteristics of Fourier amplitude and phase spectrums in the frequency domain. As shown in Fig. 1 and analyzed above, we get the conclusion that the degradation property mainly manifests in the amplitude spectrum, while the phase spectrum just has a slight difference between hazy image and corresponding clear image. Moreover, the illumination contrast of an image is represented by the amplitude spectrum, while the texture structure information is represented by the phase spectrum [36,45]. Therefore, our conclusion is also consistent with this theory, for the reason that haze mainly affects the illumination contrast of an image, while the structural information is immune to haze degradation. In addition, the artifacts in **SynClear** image show that slight difference exists between the phase spectrums of clear and hazy image pairs and it is related to the global distribution of haze. Thus, the amplitude spectrum’s learned residual can be utilized as a guide to restore the phase spectrum. According to these observations, we design the Amplitude Guided Phase (AGP) module to deal with amplitude and phase spectrums and exploit the spectrum’s learned residual to guide phase restoration in the frequency domain.

Our second insight is the discrepancy between global and local modeling for image dehazing in the spatial domain. We explain their difference through the limited receptive field of CNN-based network. Receptive field can be defined as the region around a pixel that contributes to the output at that pixel [4,25]. Due to the convolution operator’s intrinsic limitations, the network has a limited receptive field, particularly in the early layers. Consequently, existing CNN-based methods fail to accurately model the long-range dependency of an image and the context for comprehending the haze’s global distribution. Specifically for image dehazing, when the dense haze block gets larger than the receptive field, the pixels that fall into the haze block can’t get enough information to remove the haze. Undoubtedly, understanding the content globally is essential for reconstructing a high-quality clear image from its hazy counterpart. Based on this insight, we

design the Global Guided Local (GGL) module to enable local modeling part with global information.

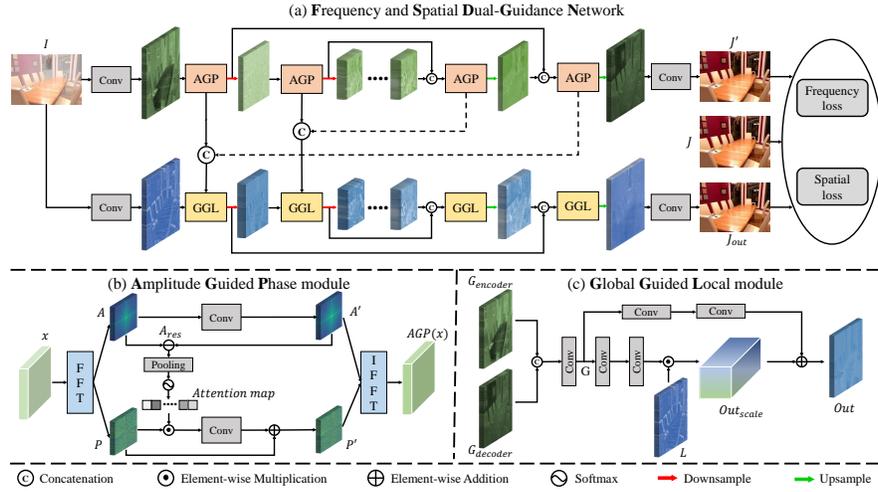


Fig. 4. Overview of the proposed network architecture for image dehazing. (a) The whole framework consists of a frequency branch and a spatial branch with the proposed (b) AGP module in the frequency domain and (c) GGL module in the spatial domain.

3.2 Frequency and Spatial Dual-Guidance Network

The two-branch-designed network has been successfully applied in various image restoration methods [9,53]. Because each branch concentrates on its own information processing procedure, it might extract distinct representations of the same input. Further, if we can use such distinct information wisely and introduce proper guidance between these two branches, comprehensive information from two branches can significantly boost the performance of image dehazing. Based on this idea, we design our two-branch neural network.

Our proposed network is based on the U-Net [42] architecture. As shown in Fig 4, the network includes two branches, a frequency/global branch and a spatial/local branch. For an input hazy image I , frequency branch outputs J' , spatial branch outputs the final dehazed image J_{out} and J is the corresponding ground-truth image. Specifically, for the frequency branch, the feature passes the delicately designed Amplitude Guided Phase (AGP) module. The AGP module performs in the frequency domain, which is global information. Besides, in order to fully exploit the restored global information and introduce global guidance for local feature learning, we introduce Global Guided Local (GGL) module to provide global information for the learning of local features at every stage of the U-Net architecture. Besides the specially designed modules mentioned above, we also use dense connection [23] and skip connection [20] in our network.

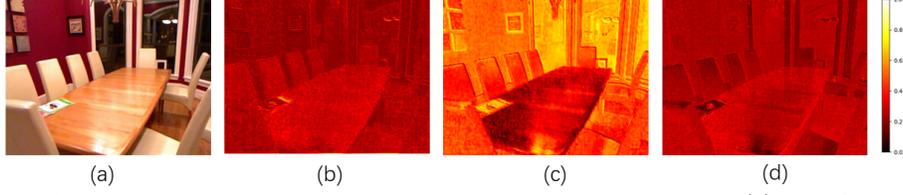


Fig. 5. Features from the same stage of global branch and local branch. (a) Clear image, (b) feature without applying GGL and AGP modules (local feature), (c) feature (b) + AGP module (global feature), and (d) feature (b) + GGL module (local feature with global feature guidance).

3.3 Amplitude Guided Phase Module

Phase conveys more information regarding image structure than amplitude does and is highly immune to noise and contrast distortions [36,45]. Along with this theory, we further find that the degradation caused by haze mainly manifests in the amplitude spectrum, and the phase spectrum is affected slightly. In other words, haze primarily changes the illumination contrast of an image, and the structure of an image is influenced mildly due to reduced visibility of the whole image. Based on our discovery, we propose the Amplitude Guided Phase (AGP) module. Let x and $AGP(x)$ denote the input and output of the AGP module. AGP module first transforms the spatial domain feature x to its frequency domain Fourier transformation $\mathcal{F}(x)$, formulated as follows:

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

The frequency-domain feature $\mathcal{F}(x)$ is denoted as $\mathcal{F}(x) = \mathcal{R}(x) + j\mathcal{I}(x)$, where $\mathcal{R}(x)$ and $\mathcal{I}(x)$ represent the real and imaginary part of $\mathcal{F}(x)$. Then the real and imaginary parts are converted to amplitude and phase spectrums, which can be formulated as:

$$\begin{aligned} \mathcal{A}(x)(u, v) &= [\mathcal{R}^2(x)(u, v) + \mathcal{I}^2(x)(u, v)]^{1/2}, \\ \mathcal{P}(x)(u, v) &= \arctan \left[\frac{\mathcal{I}(x)(u, v)}{\mathcal{R}(x)(u, v)} \right], \end{aligned} \quad (2)$$

where $\mathcal{A}(x)$ is the amplitude spectrum, $\mathcal{P}(x)$ is the phase spectrum. Given that amplitude is severely distorted, we first restore the amplitude using a 1×1 convolution. Then, the residual $\mathcal{A}_{res}(x)$ of the restored amplitude $\mathcal{A}'(x)$ and the raw amplitude $\mathcal{A}(x)$ is expressed as follows:

$$\begin{aligned} \mathcal{A}'(x)(u, v) &= \mathcal{A}(x)(u, v) \otimes k_1, \\ \mathcal{A}_{res}(x)(u, v) &= \mathcal{A}'(x)(u, v) - \mathcal{A}(x)(u, v), \end{aligned} \quad (3)$$

here, \otimes denotes the convolution operator. In this paper, we denote k_s as the convolution filter with kernel size of $s \times s$ pixel for simplicity. Further, we apply the attention map $Atten(x)$ of residual amplitude $\mathcal{A}_{res}(x)$ to compensate for the slight phase change, formulated by:

$$\begin{aligned} Atten(x)(u, v) &= Softmax[GAP(\mathcal{A}_{res}(x)(u, v))], \\ \mathcal{P}'(x)(u, v) &= [Atten(x)(u, v) \odot \mathcal{P}(x)(u, v)] \otimes k_1 + \mathcal{P}(x)(u, v), \end{aligned} \quad (4)$$

where, GAP means the global average pooling and \odot denotes the element-wise product operation. After the recovery in amplitude and phase, we turn them back to real and imaginary parts by:

$$\begin{aligned}\mathcal{R}'(x)(u, v) &= \mathcal{A}'(x)(u, v) \cos \mathcal{P}'(x)(u, v), \\ \mathcal{I}'(x)(u, v) &= \mathcal{A}'(x)(u, v) \sin \mathcal{P}'(x)(u, v).\end{aligned}\quad (5)$$

Finally, we transform the frequency domain feature $\mathcal{F}'(x) = \mathcal{R}'(x) + j\mathcal{I}'(x)$ back to the spatial domain feature $AGP(x)$ by inverse Fourier transform. Besides the frequency transformation, we keep a 3×3 convolution branch in the spatial domain to stabilize the training of the AGP module.

In Fig. 5, we show features produced under different settings. The local branch b generates artifacts and fails to model the global architecture, but with finer details in some regions. Compared with feature b, global feature c models the low-frequency representations (illumination, color and contrast) and generates globally visual pleasing results with the consistent overall structure of the scene. This well proves the validity and importance of our AGP module.

3.4 Global Guided Local Module

In order to gain global context modeling ability, some image restoration methods employ transformer/non-local [29,52]. But the considerable computational complexity of aforementioned global modeling strategy usually hampers its efficient usage. In contrast, we possess a global context modeling ability by exploiting frequency characteristics.

Due to the limited receptive field in the early layers of the local branch, the network fails to capture long-range dependency and has no enough information to remove the haze in local regions. Therefore, we propose Global Guided Local (GGL) module. In the GGL module, we incorporate the features from the same stage in the encoder and decoder of the global branch to guide the local feature in the corresponding encoder stage of the local branch. Similar to SFT [48], we inject the local feature with the scaling and shifting of global feature. Concretely, we first concatenate the two global features $G_{encoder}^n$ and $G_{decoder}^n$ from the n -th stage of the encoder and decoder and perform a simple convolution to get G^n .

$$G^n = Cat(G_{encoder}^n, G_{decoder}^n) \otimes k_3, \quad (6)$$

where, Cat denotes the concatenation operation. Then, we use the global feature G^n to guide the local feature L^n from the n -th stage of the encoder in the spatial branch. We first inject the local feature L^n with the scaling operation of G^n to get Out_{scale}^n as:

$$Out_{scale}^n = (G^n \otimes k_1 \otimes k_1) \odot L^n. \quad (7)$$

Then, with the obtained Out_{scale}^n , We introduce the shifting operation of G^n to get the output of the n -th GGL module as:

$$Out^n = (G^n \otimes k_1 \otimes k_1) + Out_{scale}^n. \quad (8)$$

In this way, our method not only focuses on details by the stack of convolution but also introduces global information to enrich the global context structure.

In Fig. 5, feature (d) is produced by introducing global information guidance to local feature (b). Obviously, after the GGL module, feature (d) is not only well structured but also have fine details and less artifacts. This means that feature (d) combines the strong point of local feature (b) and global feature (c), as we ideally expected.

3.5 Frequency and spatial dual supervision Losses

The training loss of our FSDGN is comprised of both spatial and frequency domain losses. In addition to the spatial domain Charbonnier loss [24] \mathcal{L}_{cha} , the frequency domain loss \mathcal{L}_{fre} consists of \mathcal{L}_{amp} and \mathcal{L}_{pha} for supervision from the ground-truth amplitude and phase spectrums during training.

$$\mathcal{L}_{cha} = \sqrt{(J_{out}(x) - J(x))^2 + \varepsilon} + \sqrt{(J'(x) - J(x))^2 + \varepsilon}, \quad (9)$$

$$\mathcal{L}_{amp} = \frac{2}{UV} \sum_{u=0}^{U/2-1} \sum_{v=0}^{V-1} \left(\left\| |A_{out}|_{u,v} - |A|_{u,v} \right\|_1 + \left\| |A'|_{u,v} - |A|_{u,v} \right\|_1 \right), \quad (10)$$

$$\mathcal{L}_{pha} = \frac{2}{UV} \sum_{u=0}^{U/2-1} \sum_{v=0}^{V-1} \left(\left\| |P_{out}|_{u,v} - |P|_{u,v} \right\|_1 + \left\| |P'|_{u,v} - |P|_{u,v} \right\|_1 \right). \quad (11)$$

Note, in our implementation, $\varepsilon = 1 \times e^{-12}$ and the summation for u is only performed up to $U/2 - 1$, since 50% of all frequency components are redundant. Thus, the total loss L of our network is denoted as:

$$\mathcal{L} = \mathcal{L}_{cha} + \beta (\mathcal{L}_{amp} + \mathcal{L}_{pha}), \quad (12)$$

where β is weight factor and set to 0.1 empirically.

4 Experiments

In this section, we first introduce the datasets and implement details of our experiment. Then, we make a comprehensive comparison with existing methods quantitatively and visually. Experiments on the public synthetic dataset RESIDE [28] and two different type real-word datasets demonstrate the superiority of the proposed FSDGN. Furthermore, extensive ablation studies and statistical analysis are conducted to justify the effectiveness of the core modules of FSDGN.

4.1 Experiment Setup

Datasets. We evaluate the proposed method on synthetic and real-world datasets. For synthetic scenes, we employ RESIDE[28] dataset. The subset Indoor Training Set (ITS) of RESIDE contains a total of 13990 hazy indoor images, generated

from 1399 clear images. The subset Synthetic Objective Testing Set (SOTS) of RESIDE consists of 500 indoor hazy images and 500 outdoor ones. We apply ITS and SOTS indoor as our training and testing sets. In addition, we adopt two real-world datasets: Dense-Haze [1] and NH-HAZE [2], to evaluate the robustness of our method in the real-world scenarios. Dense-Haze consists of dense and homogeneous hazy scenes, whereas NH-HAZE consists of nonhomogeneous hazy scenes. Both of the two datasets consist of 55 paired images.

Implementation Details. Our FSDGN is implemented by PyTorch with an NVIDIA RTX 2080Ti. We use ADAM as the optimizers with $\beta_1 = 0.9$, and $\beta_2 = 0.999$, and the initial learning rate is set to 2×10^{-4} . The learning rate is adjusted by the cosine annealing strategy [21]. In the training stage, we empirically set the total number of iteration to 600k. The batch and patch sizes are set to 16 and 256×256 , respectively.

Table 1. Quantitative comparison with SOTA methods on synthetic and real-world dehazing datasets.

Method	SOTS [28]		Dense-Haze [1]		NH-HAZE [2]		Param (M)	GFLOPs
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
DCP [19]	15.09	0.7649	10.06	0.3856	10.57	0.5196	-	-
DehazeNet [6]	20.64	0.7995	13.84	0.4252	16.62	0.5238	0.01M	-
AOD-Net [26]	19.82	0.8178	13.14	0.4144	15.40	0.5693	0.002M	0.1
GridDehazeNet [32]	32.16	0.9836	13.31	0.3681	13.80	0.5370	0.96M	21.5
FFA-Net [37]	36.39	0.9886	14.39	0.4524	19.87	0.6915	4.68M	288.1
MSBDN [13]	33.79	0.9840	15.37	0.4858	19.23	0.7056	31.35M	41.5
KDDN [22]	34.72	0.9845	14.28	0.4074	17.39	0.5897	5.99M	-
AECR-Net [49]	37.17	0.9901	15.80	0.4660	19.88	0.7173	2.61M	43.0
Ours	38.63	0.9903	16.91	0.5806	19.99	0.7306	2.73M	19.6

4.2 Comparison with State-of-the-art Methods

We compare our FSDGN with the SOTA methods qualitatively and quantitatively, including one prior-based algorithm (DCP [19]) and six deep learning-based methods (DehazeNet [6], AOD-Net [26], GridDehazeNet [32], FFA-Net [37], MSBDN [13] and AECR-Net [49]). The results are produced by using publicly available source codes with recommended parameters. To evaluate the performance of our method, we employ two widely used metrics for quantitative comparisons, including the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity index (SSIM).

Results on Synthetic Dataset. Table 1 compares the quantitative results of different methods on SOTS dataset [27], which indicates our FSDGN achieves the best performance with 38.36dB PSNR and 0.9903 SSIM. To further demonstrate the effectiveness of our method, we also show the visual comparison with other techniques on the typical hazy images sampled from the SOTS dataset in

Fig. 6. Compared with the ground truths, it is evident that the results of DCP, AODNet and DehazeNet not only fail to remove the dense haze but also suffer from severe color distortion (see the table and wall in Fig. 6(b)-(d)). Different from these three techniques, GridDehazeNet [32], FFA-Net [37], MSBDN [13] and AECR-Net [49] perform the hazy-to-clear image translation in an end-to-end manner. Undeniably, they mitigate the color distortion problem and achieve the restored images with higher PSNR and SSIM. However, they cannot completely remove the haze in their results (*e.g.* the red square in Fig. 6(e), the wall in Fig. 6(f) and the gap between chairs in Fig. 6(g)), and produce color shift (see the desktop in Fig. 6(h)). In contrast, our FSDGN generates the highest-fidelity dehazed results that also look perceptually close to the reference ground truths.

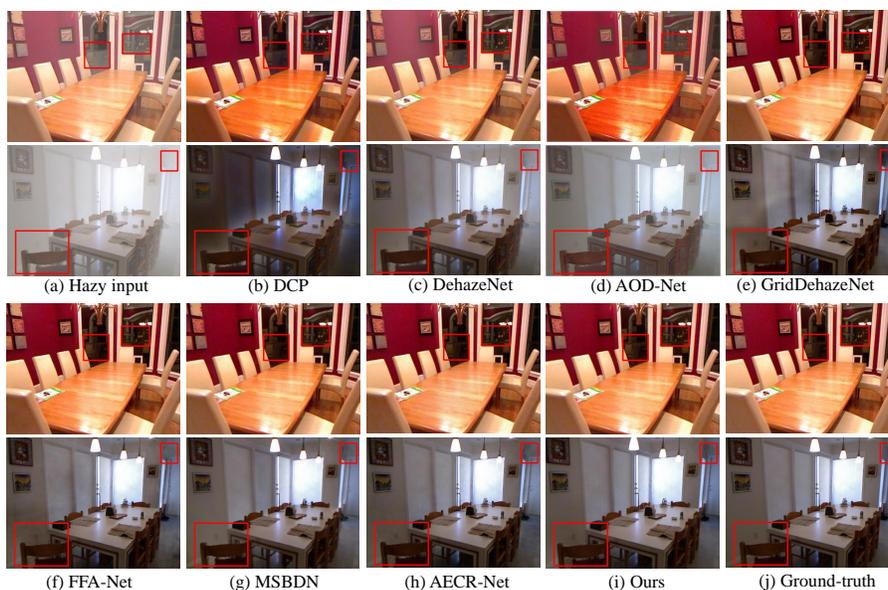


Fig. 6. Comparison of visual results on SOTS [28] dataset. Red boxes indicate the obvious differences. Zoom in for best view.

Results on Real-world Datasets. We further compare our FSDGN with SOTA methods on the two real-world datasets: Dense-Haze [1] and NH-HAZE [2] datasets. Due to the dense and nonhomogeneous distribution of haze in real-world, removing the real-world haze is more complex and challenging. As described in Table 1, the proposed FSDGN achieves the best performance on both datasets, outperforming the second-highest performance AECR-Net [49] with 1.11dB PSNR and 0.1146 SSIM on the Dense-Haze dataset and 0.11dB PSNR and 0.0233 SSIM on the NH-HAZE dataset. Fig. 7 and Fig. 8 illustrate the results of the real-world haze images sampled from the Dense-Haze and NH-HAZE datasets, respectively. The compared methods generate either color distortion or haze-remained results. Concretely, DCP [19], AOD-Net [26], GridDehazeNet [32],

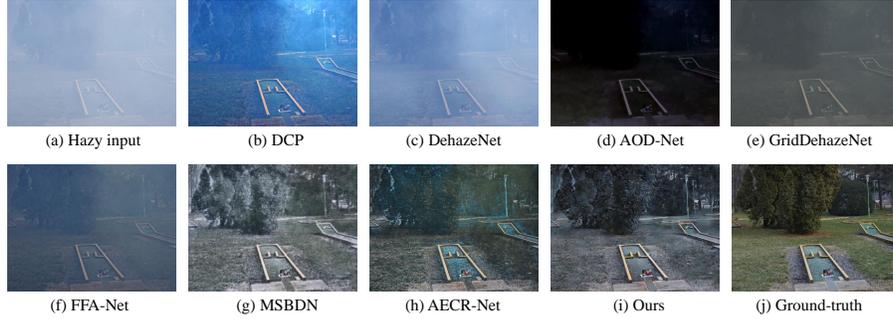


Fig. 7. Comparison of visual results on Dense-Haze [1] dataset. Zoom in for best view.

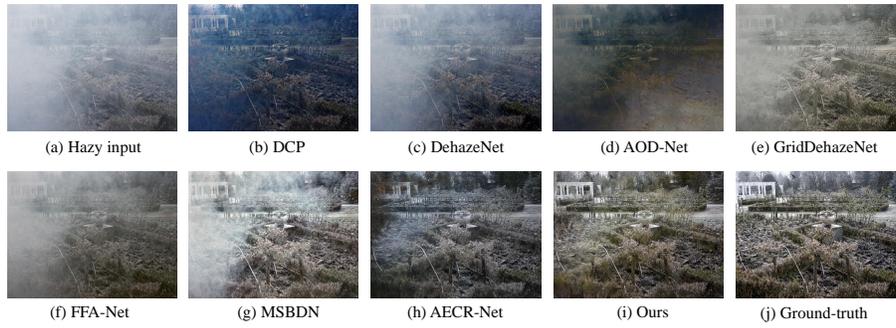


Fig. 8. Comparison of visual results on NH-HAZE [2] dataset. Zoom in for best view.

FFA-Net [37] and AECR-Net [49] produce the serious color deviation and texture loss in the restored images. Besides, apparent thick haze residual is remained in the results of DCP [19], DehazeNet [6], GridDehazeNet [32] and MSBDN [13]. In contrast, our model generates the natural and visually desirable results.

Note that we don't compare the results on Real-world datasets with workshop methods. For the reason that workshop methods achieve high performance at the cost of huge parameters. For example, the model size of DW-GAN [17] is 51.51M, much larger than ours. The champion model iPAL-AtJ on NTIRE 2019 Challenge [3] has a parameter of 46.17M.

Table 2. Ablation study on our FSDGN. AllG* represents the model consisting of two global branches without spatial guidance.

Label	AllL	GGL	AGP	AllG*	\mathcal{L}_{fre}	PSNR (dB)	SSIM	Params (M)
a	✓					37.41	0.9889	2.573
b	✓	✓				38.02	0.9899	2.580
c	✓		✓			38.34	0.9901	2.725
d				✓		37.56	0.9893	2.876
e		✓		✓		38.19	0.9901	2.883
f	✓	✓	✓			38.51	0.9902	2.731
g	✓	✓	✓		✓	38.63	0.9903	2.731

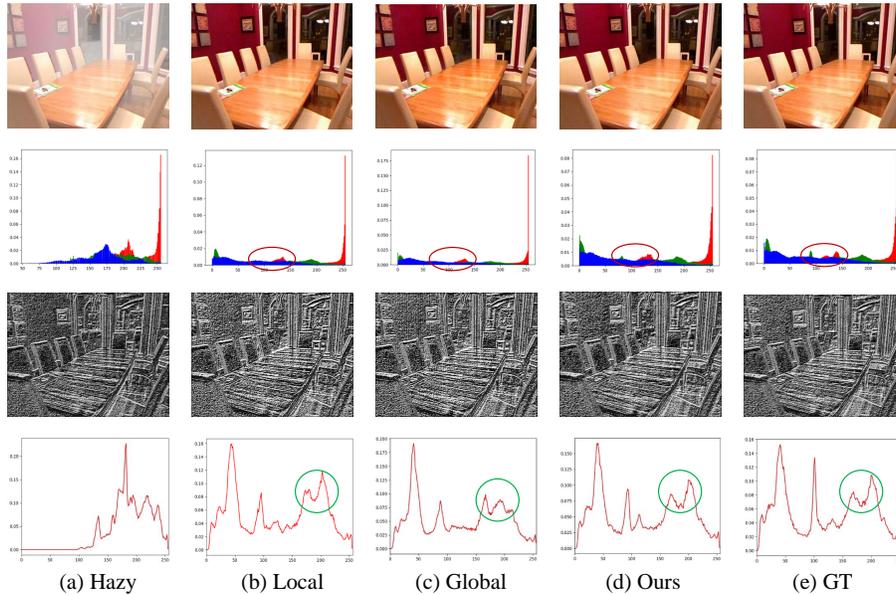


Fig. 9. Visualization of output images, histogram of color images, texture images, and histogram of gray images. (a) Input hazy image, (b) output of the local branch (without spatial guidance), (c) output of the global branch (with additional AGP module compared to the local branch), (d) output of our model (with AGP and GGL modules), and (e) ground-truth clear image.

4.3 Ablation Studies

In this section, we perform several ablation studies to analysis the effectiveness of the proposed method. In order to verify the effect of two global branches, we further introduce the AllG model, which applies the AGP module in both branches and involves no local features. Different models are denoted as follows: (a) **AllL**: The same as our FSDGN except for the AGP and GGL module. Namely, two spatial branches without spatial guidance and frequency loss. (b) **AllL+GGL**: Two spatial branches, with spatial guidance. (c) **AllL+AGP**: A spatial branch and a frequency branch. (d) **AllG**: Two global branches, without spatial guidance. (e) **AllG+GGL**: Two global branches, with spatial guidance. (f) **AllL+GGL+AGP**: A global branch and a local branch, with spatial guidance. (g) **(Ours)AllL+GGL+AGP+ \mathcal{L}_{fre}** : Our final setting in this method. The performance of these models are summarized in Table 2.

Effectiveness of the AGP module. Compared to the AllL model, the AllL+AGP model possesses a significant performance improvement of 0.93 PSNR with negligible parameters increasing. This indicates that the AGP module is an indispensable component of our network, representing the global information and, more importantly, correlating with the haze degradation.

Effectiveness of the GGL module. Compared to the AllL model, the AllL+GGL model improves the performance from 37.41 to 38.02 PSNR. In ad-

dition, the improvement of the model from AllG to AllG+GGL increases the PSNR by 0.63 dB.

More comparisons. AllL and AllG models extract all local information and global information, respectively. Experiments show that combining the global and local information is superior to using just one. Besides, the comparison between model f and g demonstrates that the introduction of the frequency loss \mathcal{L}_{fre} is beneficial for performance improvement.

To further explain and prove the superiority of our proposed modules, we present the statistical distributions in Fig. 9. Specifically, the first row is the hazy image, the output of different models, and the ground truth, respectively. The second row is the histogram of the above images, which represents the illumination and color distribution. The third row describes the textures of the first row produced by the LBP operator, and the bottom row is the histogram of the corresponding gray image, indicating the texture, light, and contrast changes. Undeniably both the local and global branch outputs remove the haze and preserve textures well with the visually pleasing image. However, it is easy to find the discrepancies in their histograms compared with the ground truth. It is evident that our FSDGN achieves a more similar distribution to the ground truth compared with the global or local branch (*e.g.* the red circles in the second row and the green circles in the bottom row of Fig. 9).

5 Conclusion

In this paper, we revisit the haze degradation in the frequency domain via Fourier transform. Based on the frequency analysis, we propose a novel image dehazing framework, termed as Frequency and Spatial Dual-Guidance Network, to explore potentials in the frequency and spatial dual domains. Specifically, we introduce two core designs to equip the proposed network, *i.e.*, the Global Guided Local module and Amplitude Guided Phase module for the spatial-domain and frequency-domain guidance, respectively. Extensive experiments validate that the proposed method achieves state-of-the-art performance on both synthetic and real hazy images in an efficient way.

Acknowledgements. This work was supported by the JKW Research Funds under Grant 20-163-14-LZ-001-004-01 and the University Synergy Innovation Program of Anhui Province under Grant GXXT-2019-025. We acknowledge the support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC. We also thank the technical support from Jie Huang.

References

1. Ancuti, C.O., Ancuti, C., Sbert, M., Timofte, R.: Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. In: Proceedings of the IEEE International Conference on Image Processing. pp. 1014–1018 (2019)
2. Ancuti, C.O., Ancuti, C., Timofte, R.: NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 444–445 (2020)
3. Ancuti, C.O., Ancuti, C., Timofte, R., Van Gool, L., Zhang, L., Yang, M.H.: Ntire 2019 image dehazing challenge report. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 1–13 (2019)
4. Araujo, A., Norris, W., Sim, J.: Computing receptive fields of convolutional neural networks. *Distill* **4**(11), e21 (2019)
5. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1674–1682 (2016)
6. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: DehazeNet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing* **25**(11), 5187–5198 (2016)
7. Cai, M., Zhang, H., Huang, H., Geng, Q., Li, Y., Huang, G.: Frequency domain image translation: More photo-realistic, better identity-preserving. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13930–13940 (2021)
8. Chen, C., Do, M.N., Wang, J.: Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: Proceedings of the European Conference on Computer Vision. pp. 576–591. Springer (2016)
9. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: HINet: Half instance normalization network for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 182–192 (2021)
10. Chen, Y., Li, W., Sakaridis, C., Dai, D., Van Gool, L.: Domain adaptive faster r-cnn for object detection in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3339–3348 (2018)
11. Chi, L., Tian, G., Mu, Y., Xie, L., Tian, Q.: Fast non-local neural networks with spectral residual learning. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 2142–2151 (2019)
12. Deng, Q., Huang, Z., Tsai, C.C., Lin, C.W.: HardGAN: A haze-aware representation distillation gan for single image dehazing. In: Proceedings of the European Conference on Computer Vision. pp. 722–738. Springer (2020)
13. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2157–2167 (2020)
14. Dong, J., Pan, J.: Physics-based feature dehazing networks. In: Proceedings of the European Conference on Computer Vision. pp. 188–204. Springer (2020)
15. Fattal, R.: Single image dehazing. *ACM Transactions on Graphics (TOG)* **27**(3), 1–9 (2008)
16. Fattal, R.: Dehazing using color-lines. *ACM Transactions on Graphics (TOG)* **34**(1), 1–14 (2014)

17. Fu, M., Liu, H., Yu, Y., Chen, J., Wang, K.: DW-GAN: A discrete wavelet transform gan for nonhomogeneous dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 203–212 (2021)
18. Guo, T., Li, X., Cherukuri, V., Monga, V.: Dense scene information estimation network for dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019)
19. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**(12), 2341–2353 (2010)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
21. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of tricks for image classification with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 558–567 (2019)
22. Hong, M., Xie, Y., Li, C., Qu, Y.: Distilling image dehazing with heterogeneous task imitation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3462–3471 (2020)
23. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4700–4708 (2017)
24. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 624–632 (2017)
25. Le, H., Borji, A.: What are the receptive, effective receptive, and projective fields of neurons in convolutional neural networks? arXiv preprint arXiv:1705.07049 (2017)
26. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: AOD-Net: All-in-one dehazing network. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4770–4778 (2017)
27. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: End-to-end united video dehazing and detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
28. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing* **28**(1), 492–505 (2018)
29. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: SwinIR: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1833–1844 (2021)
30. Liu, J., Wu, H., Xie, Y., Qu, Y., Ma, L.: Trident dehazing network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 430–431 (2020)
31. Liu, R., Fan, X., Hou, M., Jiang, Z., Luo, Z., Zhang, L.: Learning aggregated transmission propagation networks for haze removal and beyond. *IEEE Transactions on Neural Networks and Learning Systems* **30**(10), 2973–2986 (2018)
32. Liu, X., Ma, Y., Shi, Z., Chen, J.: GridDehazeNet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7314–7323 (2019)
33. Liu, Y., Pan, J., Ren, J., Su, Z.: Learning deep priors for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2492–2500 (2019)

34. Mao, X., Liu, Y., Shen, W., Li, Q., Wang, Y.: Deep residual Fourier transformation for single image deblurring. arXiv preprint arXiv:2111.11745 (2021)
35. McCartney, E.J.: Optics of the atmosphere: scattering by molecules and particles. New York (1976)
36. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. Proceedings of the IEEE **69**(5), 529–541 (1981)
37. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: FFA-Net: Feature fusion attention network for single image dehazing. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 11908–11915 (2020)
38. Qu, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced pix2pix dehazing network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8160–8168 (2019)
39. Rao, Y., Zhao, W., Zhu, Z., Lu, J., Zhou, J.: Global filter networks for image classification. Advances in Neural Information Processing Systems **34** (2021)
40. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.H.: Single image dehazing via multi-scale convolutional neural networks. In: Proceedings of the European Conference on Computer Vision. pp. 154–169. Springer (2016)
41. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.H.: Gated fusion network for single image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3253–3261 (2018)
42. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
43. Sakaridis, C., Dai, D., Hecker, S., Van Gool, L.: Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In: Proceedings of the European Conference on Computer Vision. pp. 687–704 (2018)
44. Sakaridis, C., Dai, D., Van Gool, L.: Semantic foggy scene understanding with synthetic data. International Journal of Computer Vision **126**(9), 973–992 (2018)
45. Skarbnik, N., Zeevi, Y.Y., Sagiv, C.: The importance of phase in image processing. Technion-Israel Institute of Technology, Faculty of Electrical Engineering (2009)
46. Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.: Resolution-robust large mask inpainting with Fourier convolutions. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2149–2159 (2022)
47. Wang, C., Shen, H.Z., Fan, F., Shao, M.W., Yang, C.S., Luo, J.C., Deng, L.J.: EAA-Net: A novel edge assisted attention network for single image dehazing. Knowledge-Based Systems **228**, 107279 (2021)
48. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 606–615 (2018)
49. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10551–10560 (2021)
50. Yang, Y., Soatto, S.: FDA: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4085–4095 (2020)
51. Ye, T., Jiang, M., Zhang, Y., Chen, L., Chen, E., Chen, P., Lu, Z.: Perceiving and modeling density is all you need for image dehazing. arXiv preprint arXiv:2111.09733 (2021)

52. Yi, P., Wang, Z., Jiang, K., Jiang, J., Ma, J.: Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3106–3115 (2019)
53. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14821–14831 (2021)
54. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3194–3203 (2018)
55. Zhang, J., Cao, Y., Fang, S., Kang, Y., Wen Chen, C.: Fast haze removal for nighttime image using maximum reflectance prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7418–7426 (2017)
56. Zheng, Z., Ren, W., Cao, X., Hu, X., Wang, T., Song, F., Jia, X.: Ultra-high-definition image dehazing via multi-guided bilateral learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16180–16189. IEEE (2021)
57. Zhu, Q., Mai, J., Shao, L.: Single image dehazing using color attenuation prior. In: BMVC. Citeseer (2014)
58. Zou, W., Jiang, M., Zhang, Y., Chen, L., Lu, Z., Wu, Y.: SDWNet: A straight dilated network with wavelet transformation for image deblurring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1895–1904 (2021)