

Guided Real Image Dehazing Using YCbCr Color Space

Wenxuan Fang¹, Junkai Fan¹, Yu Zheng¹, Jiangwei Weng¹, Ying Tai², Jun Li^{1*}

¹School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

²School of Intelligence Science and Technology, Nanjing University, Suzhou, China.

{wenxuan_fang, junkai.fan, yuzheng, wengjiangwei, junli}@njust.edu.cn, yingtai@nju.edu.cn

Abstract

Image dehazing, particularly with learning-based methods, has gained significant attention due to its importance in real-world applications. However, relying solely on the RGB color space often fall short, frequently leaving residual haze. This arises from two main issues: the difficulty in obtaining clear textural features from hazy RGB images and the complexity of acquiring real haze/clean image pairs outside controlled environments like smoke-filled scenes. To address these issues, we first propose a novel **Structure Guided Dehazing Network (SGDN)** that leverages the superior structural properties of YCbCr features over RGB. It comprises two key modules: **Bi-Color Guidance Bridge (BGB)** and **Color Enhancement Module (CEM)**. BGB integrates a phase integration module and an interactive attention module, utilizing the rich texture features of the YCbCr space to guide the RGB space, thereby recovering clearer features in both frequency and spatial domains. To maintain tonal consistency, CEM further enhances the color perception of RGB features by aggregating YCbCr channel information. Furthermore, for effective supervised learning, we introduce a **Real-World Well-Aligned Haze dataset**, which includes a diverse range of scenes from various geographical regions and climate conditions. Experimental results demonstrate that our method surpasses existing state-of-the-art methods across multiple real-world smoke/haze datasets.

Introduction

Image dehazing aims to remove the blurring effect caused by atmospheric particles such as haze and smoke in images and improve image quality, which has important application value in object detection (Zhao et al. 2024), image segmentation (Kirillov et al. 2023) and object tracking (Hu et al. 2023, 2024). The formation of foggy images can usually be modeled using an Atmospheric Scattering Model (ASM):

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

where $I(x)$ and $J(x)$ represent the observed foggy image and clean image, and $A(x)$ and $t(x)$ are the global atmospheric light and transmission map, respectively. The transmission map $t(x) = e^{-\beta d(x)}$ is determined by the scattering coefficient β and the scene depth $d(x)$. Many dehazing

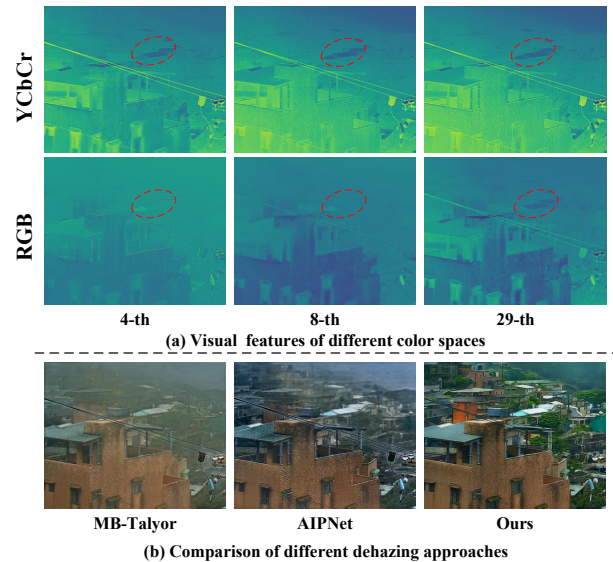


Figure 1: Visual comparison of different color spaces: (a) RGB features degrade, blurring textures, while YCbCr is less affected by fog and shows clearer textures. (b) RGB models (e.g., MB-Taylor) leave residual haze, YCbCr models (e.g., AIPNet) distort colors. Our approach removes heavy fog while preserving color accuracy.

methods (Li et al. 2017, 2021) use deep networks to estimate physical parameters and apply Eq. (1) to restore clean images. Additionally, some methods (Zhang et al. 2023; Chen, He, and Lu 2024) focus on learning a mapping model from hazy to clean images in a supervised manner, offering an end-to-end dehazing solution. Despite their promising results on synthetic haze datasets, they face two major challenges in real-world haze scenes.

Firstly, most mapping models based on the RGB color space perform suboptimally in real-world foggy conditions because fog blurs textures within RGB features, as shown in Fig. 1(a). For instance, MB-Taylor (Qiu et al. 2023) (Fig. 1(b)) attempts to extract features and remove degradation in the RGB space but still leaves significant residual haze. However, we have observed that YCbCr features are less affected by fog and provide clearer image textures (Fig. 1(a)).

*Corresponding authors

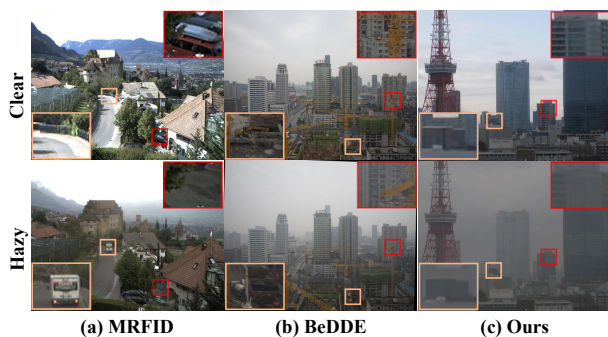


Figure 2: Examples from the MRFID and BEDDE datasets show noticeable background differences between reference and haze images due to varying shooting angles and long time intervals. In contrast, our RW^2AH dataset achieves excellent alignment in physical space.

This is because the chrominance components of YCbCr are less responsive to the neutral colors impacted by haze, allowing YCbCr to preserve more color and details in hazy conditions (Fang et al. 2019). Despite their advantages over RGB, converting YCbCr to RGB can introduce rounding errors that cause severe color distortion, as seen with AIPNet (Wang et al. 2018) in Fig. 1(b). This observation motivates us to use YCbCr features to guide the recovery of clearer structures in RGB mapping models.

Secondly, most existing datasets are unsuitable for fully supervised training (see Fig. 2) due to background inconsistencies (Liu et al. 2020b) and spatial misalignments between hazy/clean images (Zhao et al. 2020). Capturing well-aligned hazy/clean image pairs in real-world conditions is particularly challenging, as uncontrolled environments make it difficult to photograph images with consistent backgrounds within a short timeframe at the same location. Consequently, most dehazing methods are trained on synthetic datasets, which have significant domain gaps from real-world scenes, leading to suboptimal performance in practical applications.

To this end, we propose a novel image dehazing method under RGB and YCbCr color spaces for real haze removal, named the **Structure Guided Dehazing Network (SGDN)**. Specifically, our SGDN employs an asymmetric encoder-decoder structure that shares a single encoder for feature extraction from both RGB and YCbCr color spaces. First, we propose a **Bi-Color Guidance Bridge (BGB)** that enhances structural integrity and refine texture details to improve image quality. It includes two key designs: a **Phase Integration Module (PIM)** and an **Interactive Attention Module (IAM)**. The former guides RGB features to reconstruct clearer textures by the phase spectrum of YCbCr features in the frequency domain, and the latter further fuses the discriminative features of YCbCr space in the spatial domain. Subsequently, we propose a **Color Enhancement Module (CEM)** that aggregates channel information from YCbCr features to further enhance the color perception of RGB features. Finally, we collect a large-scale Real-World Well-Aligned

Haze (RW^2AH) dataset that includes multiple scenes. Notably, it can be used for supervised network training and serves as a fair benchmark for evaluating dehazing methods. Our contributions are summarized as follows:

- We propose a novel dehazing method, called SGDN, which exploits the clear texture of the YCbCr to guide the RGB features for real-world image dehazing.
- Bi-color Guidance Bridge is proposed to guide the recovery of RGB features from the frequency and spatial domains. Besides, we design the Color Enhancement Module (CEM) to further enhance the color perception capabilities of RGB features.
- We introduce RW^2AH , the first well-aligned real-world dehazing dataset, which contains 1,758 real-world image pairs of corresponding clear and hazy images. While these image pairs may not be perfectly aligned, they are sufficiently aligned for use in supervised learning. This dataset sets a new benchmark for evaluating dehazing methods in real-world image dehazing tasks.

Related Works

Image Dehazing

Image dehazing can be regarded as an inverse problem, where the goal is to recover the clean image from a hazy image. Here, we primarily review two types of methods:

Prior-based dehazing methods (He, Sun, and Tang 2010; Bui and Kim 2017; Fattal 2014; Li et al. 2021) mainly rely on ASM and estimate physical parameters through manual priors or statistical learning methods. For example, the Color Attenuation Prior (Zhu, Mai, and Shao 2015) estimated the transmittance by analyzing the relationship between color and brightness in foggy images and establishing a linear model. Although these methods were effective in some scenarios, they rely on the accuracy of estimating physical parameters, which may produce disappointing results in some more complex real-world scenarios.

Learning-based dehazing methods (Chen et al. 2022; Zheng et al. 2023; Qiu et al. 2023; Zhang, Zhou, and Li 2024; Cui, Ren, and Knoll 2024; Chen, He, and Lu 2024) were proposed to directly estimate haze-free images from hazy inputs. Recently, some works (Fan et al. 2023, 2024, 2025) have been designed for real-world unpaired dehazing models. Some typical examples are built on the CycleGAN (Zhu et al. 2017) framework, such as DAD (Shao et al. 2020), CycleDehaze (Engin, Genç, and Kemal Ekenel 2018) and D^4 (Yang et al. 2022). However, the generative approach of GAN networks often lacks strong constraints, leading to the creation of artifacts that interfere with training. Another approaches (Chen et al. 2021; Zhao et al. 2021) were to introduce prior knowledge to guide the model’s optimization direction. For instance, PSD (Chen et al. 2021) designed a series of prior losses and fine-tuned real-world dehazing in an unsupervised manner. However, these methods cannot avoid the inherent flaws of handcrafted priors. In this paper, we leverage the advantages of dual color spaces to guide the model to learn haze-independent features, achieving superior dehazing results.

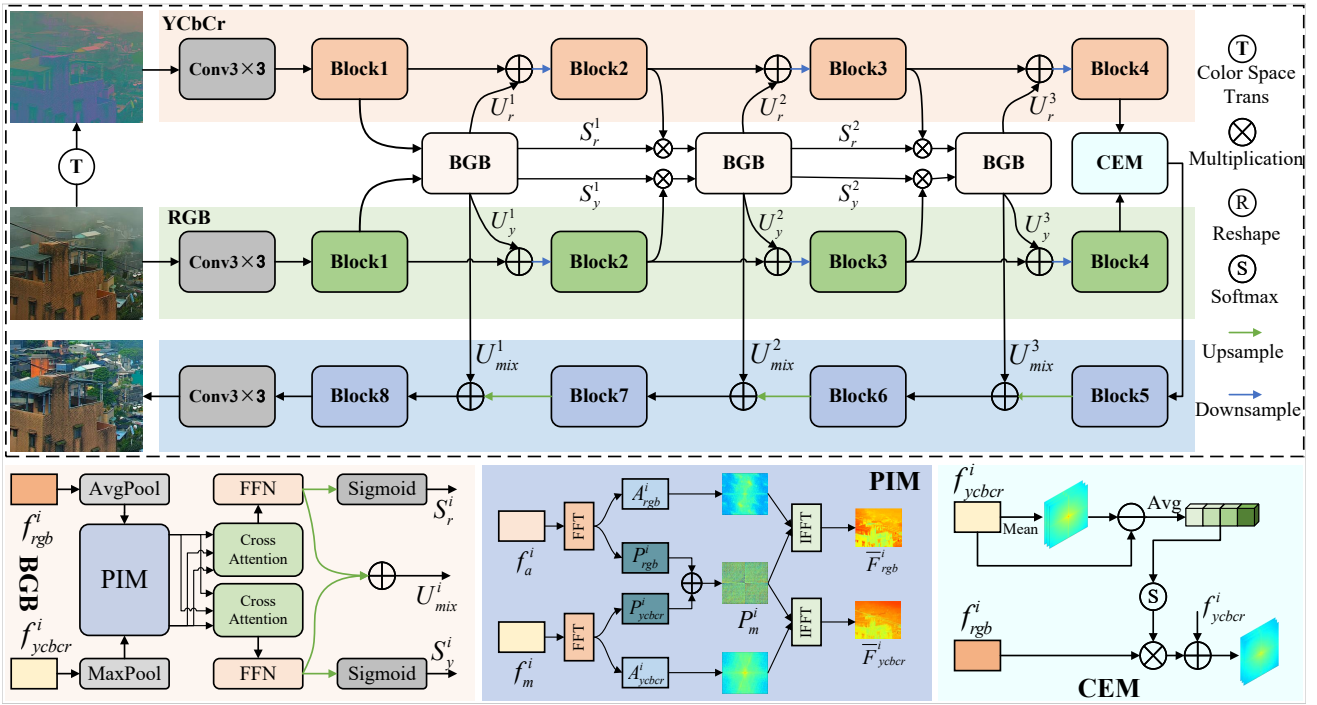


Figure 3: The overall pipeline of our SGDN. It includes the proposed Bi-Color Guidance Bridge (BGB) and Color Enhancement Module (CEM). BGB promotes RGB features to produce clearer textures through YCbCr color space in both frequency and spatial domain, while CEM significantly enhances the visual contrast of the images.

Color Model in Image Dehazing

In addition to the commonly used RGB color model, some studies (Dudhane and Murala 2019; Wang et al. 2018; Lyu, Chen, and Hou 2024; Fang et al. 2019) focus on using other color models for haze removal, such as HSV, YCbCr and YUV. For example, AIPNet (Wang et al. 2018) employed a multi-scale network to enhance the contrast of the Y channel, achieving effective haze removal. However, these methods often fail when dealing with complex structures due to their reliance on a single assumption. To introduce robust constraints, MPCNet (Lyu, Chen, and Hou 2024) leveraged HSV and YCbCr color spaces, proposing a color correction network and a prior-based dehazing network to improve the removal of non-uniform smoke. Although these methods have achieved excellent performance on training data, they require the assistance of ASM and color conversion processes, which inevitably introduce color errors.

Unlike previous methods, we reveal that YCbCr exhibits clearer features and colors under haze compared to RGB. To avoid conversion errors, we apply YCbCr to guide the reconstruction of haze-independent features in RGB, leading to more effective dehazing.

Methodology

The overall framework of our SGDN is illustrated in Fig. 3, our SGDN is an asymmetric encoder-decoder structure consisting of Large Kernel Attention Blocks (Li et al. 2023). Given a haze image $I(x) \in \mathcal{R}^{C \times H \times W}$, we first convert

it from RGB to YCbCr space using a linear transformation defined by standard conversion matrices (Wang et al. 2018), to obtain $Y(x) \in \mathcal{R}^{C \times H \times W}$. Subsequently, we capture the features of different color spaces through the shared encoders of the two branches. Next, we input features from different color spaces into our Bi-color Guidance Bridge, so that RGB features generate the rich refine textures by the YCbCr features. Finally, we employ the rich chromaticity information of the YCbCr to enhance the color perception of the RGB, thereby improving the overall visual effect and color expression of the image.

Bi-color Guidance Bridge

In the RGB color space, capturing features with poor background contrast is challenging, but it excels in representing overall image structure and coarse textures. Conversely, the YCbCr color space separates luminance and chrominance, enhancing attention to image details and color information. Therefore, we propose the BGB to enhance structural integrity and refine texture details by leveraging the inherent advantages of the YCbCr color space.

Our BGB consist of the a Phase Integration Module (PIM) and an Interaction Attention Module (IAM). As shown in Fig. 3, given RGB feature $f_{rgb}^i \in \mathcal{R}^{C \times H \times W}$, we perform Avgpool operation to smooth out the noise effects, thereby retaining the global structure in a balanced manner. For the YCbCr feature $f_{yecbr}^i \in \mathcal{R}^{C \times H \times W}$, we apply Maxpool operation to ensure that image details and colors have higher activation values, making them more robust to minor varia-

tions in position and shape. So, they are defined as:

$$f_a^i, f_m^i = \text{Avgpool}(f_{\text{rgb}}^i), \text{Maxpool}(f_{\text{ycbcr}}^i). \quad (2)$$

Phase Integration Module. Follow the (Yu et al. 2022), clear and hazy images exhibit only minor differences in their phase spectrum. Furthermore, phase spectrum conveys more structural details of the image compared to amplitude spectrum, and it is more robust to contrast distortion and noise. Therefore, we first apply the Fast Fourier Transform (FFT) to convert the features into the frequency domain, and then we can separate the phase and amplitude components:

$$\mathcal{A}_{\text{rgb}}^i, \mathcal{P}_{\text{rgb}}^i = \mathcal{S}(\mathbf{F}(f_a^i)), \quad \mathcal{A}_{\text{ycbcr}}^i, \mathcal{P}_{\text{ycbcr}}^i = \mathcal{S}(\mathbf{F}(f_m^i)), \quad (3)$$

where $\mathcal{A}_{\text{rgb}}^i$ and $\mathcal{P}_{\text{rgb}}^i$ denote the amplitude and phase spectrum of the RGB feature, respectively. $\mathcal{A}_{\text{ycbcr}}^i$ and $\mathcal{P}_{\text{ycbcr}}^i$ represent the corresponding features in the YCbCr space. $\mathcal{S}(\cdot)$ and $\mathbf{F}(\cdot)$ refer to the decoupling and FFT operations.

Considering that amplitude spectrum can suffer from distortion, we employ a 1×1 convolutional layer to restore the amplitude, and a 3×3 convolutional layer to extract more detailed structure information. Next, $\mathcal{P}_{\text{rgb}}^i$ and $\mathcal{P}_{\text{ycbcr}}^i$ is combined linearly to introduce additional structural details, and obtain a blended phase spectrum \mathcal{P}_m^i . Finally, we perform the inverse Fourier transform using the \mathcal{P}_m^i and the corresponding amplitude spectrum to reconstruct the features in the spatial domain. This process can be expressed as:

$$\begin{aligned} \mathcal{P}_m^i &= C_{3 \times 3}(\mathcal{P}_{\text{ycbcr}}^i) + C_{3 \times 3}(\mathcal{P}_{\text{rgb}}^i), \\ \bar{F}_{\text{rgb}}^i &= \bar{\mathbf{F}}(C_{1 \times 1}(\mathcal{A}_{\text{rgb}}^i), \mathcal{P}_m^i), \\ \bar{F}_{\text{ycbcr}}^i &= \bar{\mathbf{F}}(C_{1 \times 1}(\mathcal{A}_{\text{ycbcr}}^i), \mathcal{P}_m^i), \end{aligned} \quad (4)$$

where \bar{F}_{ycbcr}^i and \bar{F}_{rgb}^i are the output of PIM, and $\bar{\mathbf{F}}(\cdot)$ is the inverse Fourier transform.

Interaction Attention Module. Afterwards, IAM is presented to learn important regions of Ycber features in the spatial domain. Specifically, IAM utilizes a Cross-Attention (CA) mechanism and a Feed-Forward Network (FFN). For both RGB and YCbCr features, the original feature serves as the query, while the key and value are derived from the features of the other color space. This can be defined as:

$$\begin{aligned} Q_r, K_r, V_r &= \text{Linear}(\bar{F}_{\text{rgb}}^i), \\ Q_y, K_y, V_y &= \text{Linear}(\bar{F}_{\text{ycbcr}}^i), \\ \tilde{f}_{\text{rgb}}^i &= \mathcal{F}_n(\mathcal{F}_c(Q_r, K_y, V_y)), \\ \tilde{f}_{\text{ycbcr}}^i &= \mathcal{F}_n(\mathcal{F}_c(Q_y, K_r, V_r)), \end{aligned} \quad (5)$$

where $\text{Linear}(\cdot)$ represents the linear layer and \mathcal{F}_n is the FFN, and \mathcal{F}_c is the cross attention.

To further enhance the information flow within the pipeline, we first upsample \tilde{f}_{rgb}^i and $\tilde{f}_{\text{ycbcr}}^i$ to obtain U_r^i and U_y^i . Next, we compress them to the range (0,1) to obtain S_r^i and S_y^i , and propagate them effectively to the next stage through matrix multiplication. This ensures the maximization of both the global structure and image details during the fusion process. Finally, we perform element-wise addition

operation on U_r^i and U_y^i to obtain U_{mix}^i as the residual of the decoder. It can be represented by the following formula:

$$\begin{aligned} U_r^i, U_y^i &= \text{Up}(\tilde{f}_{\text{rgb}}^i), \text{Up}(\tilde{f}_{\text{ycbcr}}^i), \\ f_{\text{rgb}}^{i+1}, f_{\text{ycbcr}}^{i+1} &= \mathbf{Sig}(U_r^i) \otimes f_{\text{rgb}}^{i+1}, \mathbf{Sig}(U_y^i) \otimes f_{\text{ycbcr}}^{i+1}, \\ U_{mix}^i &= U_r^i + U_y^i, \end{aligned} \quad (6)$$

where $\mathbf{Sig}(\cdot)$ represents the sigmoid operation and \otimes denotes element-wise multiplication.

Color Enhancement Module

In the YCbCr color space, the chrominance components have a constant or zero response to neutral colors directly affected by haze, while the luminance is proportional to the haze density. This allows YCbCr to retain more colors and image details in hazy environments. Therefore, we propose a Color Enhancement Module (CEM) that leverages the rich color information in the YCbCr color space to enhance the RGB features. Formally, let f_{rgb}^i and f_{ycbcr}^i denote the inputs of the CEM. We further remove global illumination and haze effects from f_{ycbcr}^i to emphasize local variations and color. We achieve this by averaging across channels at each position and subtracting this mean, centering the pixel values symmetrically around zero:

$$\mathbf{A}_m = f_{\text{ycbcr}}^i - \mathcal{M}(f_{\text{ycbcr}}^i), \quad (7)$$

where \mathbf{A}_m is local concentration features, and $\mathcal{M}(\cdot)$ is the mean normalization. We then transform \mathbf{A}_m into distribution intensity \mathbf{v}_c using a project network with global average pooling and softmax. Finally, \mathbf{v}_c modulates f_{rgb}^i to enhance local RGB feature perception while preserving the gradient in f_{ycbcr}^i . This can be defined as:

$$\mathbf{v}_c = \text{Softmax}(\text{Avg}(\mathbf{A}_m)), \quad \mathbf{D}_o = \mathbf{v}_c \otimes f_{\text{rgb}}^i + f_{\text{ycbcr}}^i, \quad (8)$$

where \mathbf{D}_o denotes the final output of CEM.

Training Loss

Follow the (Cui et al. 2023), we employ multi-scale loss to train our SGDN. The overall loss can be expressed as:

$$\mathcal{L}_{all} = \sum_{s=1}^3 [\eta \mathcal{L}_{\ell_1}^s + \theta \mathcal{L}_{ssim}^s + \lambda \mathcal{L}_{\text{fft}}^s], \quad (9)$$

where \mathcal{L}_{ℓ_1} represents ℓ_1 loss, \mathcal{L}_{ssim} is the structural similarity loss (Wang et al. 2004), and \mathcal{L}_{fft} means Fourier loss (Liu et al. 2020a); s is the input and output of different sizes, with values of 1, 0.5, and 0.25. η, θ and λ represent the weight coefficient of the corresponding loss function, which are set to 1.0, 0.5, and 0.1, respectively. In addition, \mathcal{L}_{ssim} can better preserve the details and color information of the image in cases of poor alignment, making the resulting image look more natural.

Real-World Well-Aligned Haze Dataset

To address the lack of paired real-world haze datasets for fully supervised learning, we create a Real-world Well-Aligned Haze (RW²AH) dataset comprising 1,758 pairs. The collection process and statistical analysis are as follows:

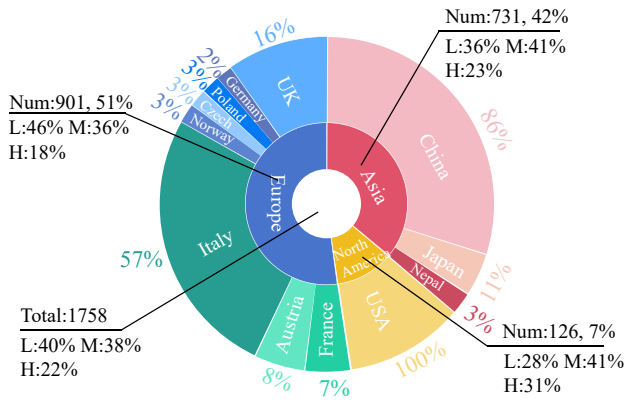


Figure 4: Geographic and haze distribution of our RW²AH.

- 1) **Source Selection:** We use stationary webcams from YouTube to capture hazy and clean images, spanning diverse environments such as landscapes, vegetation, buildings, and mountains.
- 2) **Geographical Coverage:** The dataset includes webcams from twelve countries across Asia, Europe, and America, capturing various climate conditions and haze formations.
- 3) **Temporal Pairing:** Hazy and clean images are recorded on the same day, ensuring they are captured from the same webcam, with intervals ranging from 1 to several hours.
- 4) **Haze Density:** We categorize the captured haze images into three density levels: light haze (40%), moderate haze (38%), and heavy haze (22%), as shown in Fig. 4.
- 5) **Exclusion of Dynamic Objects:** To maintain scene consistency, we exclude images with uncontrollable objects (e.g., vehicles, pedestrians) and those where more than two-thirds of the image is sky.
- 6) **Cropping:** To ensure background alignment, we apply artificial cropping, keeping the most similar image pairs and minimizing discrepancies from slight camera shifts or lighting changes.

Compared to the BeDDE (208 pairs) (Zhao et al. 2020) and the MRFID (800 pairs) (Liu et al. 2020b) datasets, our RW²AH (1,758 pairs) offers a larger volume of data and greater scene diversity. Notably, existing supervised models trained on our dataset using only ℓ_1 loss achieve excellent performance. As a result, our RW²AH can serve as a new benchmark for evaluating real-world dehazing methods using standard metrics like PSNR and SSIM.

Experiments

Datasets and Metrics

We conduct qualitative and quantitative studies on three real-world smoke/haze datasets. Here, we employ four metrics to evaluate the dehazing effectiveness and image quality. These include the commonly used reference quality assessment metrics, PSNR and SSIM. Additionally, we employ the no-reference image evaluation metrics, FADE (Choi, You, and

Bovik 2015) and NIQE (Mittal, Soundararajan, and Bovik 2012), to measure the haze density and image quality.

Our RW²AH dataset is a real-world haze dataset collected from global online webcams, encompassing various climate conditions and diverse scenes from different geographic locations. The dataset includes 1,406 pairs of images with different resolutions for training and 352 pairs of images for evaluation. It features well-aligned backgrounds, making it suitable for effective supervised learning.

Real-World Smoke (RWS) Dataset is a collection of I-HAZE (Ancuti, Ancuti, and Timofte 2018), O-Haze (Ancuti et al. 2018), and NH-Haze (Ancuti, Ancuti, and Timofte 2020) datasets. It contains 155 pairs of homogeneous and non-homogeneous indoor and outdoor smoke images. Following (Fan et al. 2023), we select 147 images pairs for training and 8 images pairs for evaluation.

Real-world Task-Driven Testing Set (RTTS) is a subset of the RESIDE (Li et al. 2018) dataset, comprising 4,322 outdoor real-world haze images. Since it lacks corresponding ground truth, it is only used for testing purposes.

Comparison with State-of-the-arts

To validate the effectiveness of our SGDN in real haze/smoke scenarios, we compare it with several representative SOTA methods. For fairness, we use their default settings and fine-tune them on our RW²AH dataset for optimal performance. Table 1 reports the quantitative comparison results on three real-world haze/smoke datasets.

Results on RWS. As depicted in Table 1, our SGDN outperforms all SOTA methods. Specifically, it surpasses DCMPNet (Zhang, Zhou, and Li 2024) by approximately 10.99% in PSNR, 2.99% in SSIM, 18.66% in FADE, and 9.27% in NIQE. Additionally, Fig. 5 displays the visual results of each method on real smoke. PSD (Chen et al. 2021), MB-Taylor (Qiu et al. 2023), DehazeFormer (Song et al. 2023), and RIDCP (Wu et al. 2023) typically struggle with dense smoke, as they all learn smoke mapping in the RGB color space, which makes it difficult to capture useful textures among blurry features. DCMPNet, with robust depth estimation, performs better in dense smoke but still shows minor color deviations. In contrast, our SGDN uses YCbCr texture details to guide RGB dehazing, producing results closer to the clear ground truth.

Results on our RW²AH. Fig. 6 illustrates the dehazing comparison of various methods on our RW²AH dataset. Our method outperforms the others in dehazing effectiveness. Specifically, PSD significantly enhances image contrast but fails to fully eliminate haze. MB-Taylor and RIDCP still retain a considerable amount of haze in their results. Although DehazeFormer and DCMPNet remove haze effectively, they struggle to preserve image details and textures. In contrast, our method achieves better dehazing while maintaining the natural contrast. Additionally, our method demonstrates stronger performance in the quantitative comparisons reported in Table 1.

Results on RTTS. From Table 1, our SGDN achieves the highest FADE score, indicating its robust dehazing capability. In terms of NIQE, SGDN delivers a competitive result,

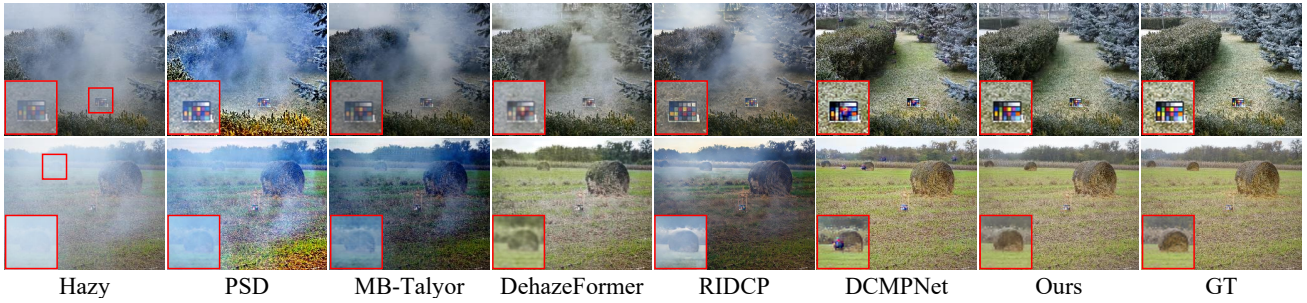


Figure 5: Visual comparison results on the real-world smoke dataset. Zoom in for a better view.

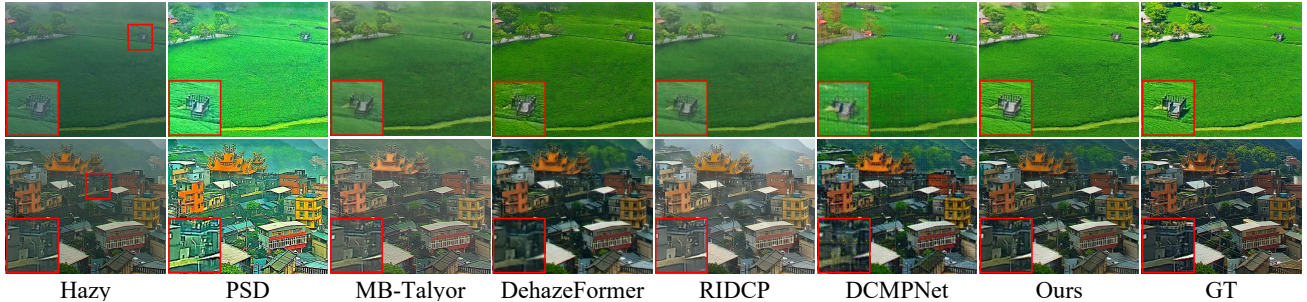


Figure 6: Visual comparison results on the our RW^2AH dataset. Zoom in for a better view.

Settings	Methods	Venue	Real-world Smoke				RW^2AH (Ours)				RTTS		Overhead	
			PSNR \uparrow	SSIM \uparrow	FADE \downarrow	NIQE \downarrow	PSNR \uparrow	SSIM \uparrow	FADE \downarrow	NIQE \downarrow	FADE \downarrow	NIQE \downarrow	Params	FLOPS
Unpaired	DCP	CVPR'09	15.01	0.392	0.4961	4.1303	15.56	0.479	0.5716	5.8557	0.7865	6.6401	-	-
	RefineDNet	TIP'21	17.44	0.525	0.6042	5.1060	16.62	0.517	0.4908	5.8190	0.6149	6.9708	65.78M	276.81G
	CDD-GAN	ECCV'22	12.16	0.250	0.3174	5.3888	17.78	0.601	<u>0.4111</u>	5.7625	0.7470	6.6245	29.27M	279.44G
	D ⁴	CVPR'22	18.21	0.644	0.6907	5.0909	15.68	0.519	0.6791	7.8726	0.7444	5.7960	10.7M	2.25G
Paired	PSD	CVPR'21	12.86	0.440	1.1304	4.4485	16.95	0.448	0.6204	5.9418	0.6115	6.0018	6.21M	143.91G
	YOLY	IJCV'21	14.29	0.409	0.9502	4.9007	15.40	0.451	0.4827	6.0521	0.8663	7.0794	1.24M	303.90G
	FSDGN	ECCV'22	19.43	0.671	0.9904	5.7410	18.57	0.430	0.8973	6.4823	0.8034	6.7189	<u>2.73M</u>	<u>19.6G</u>
	DehazeFormer	TIP'23	18.64	0.688	0.4580	4.1083	20.36	0.612	0.6904	5.8067	0.6151	5.8130	25.44M	139.85G
	C ² PNet	CVPR'23	17.86	0.361	0.8795	6.1102	17.48	0.472	0.6154	7.0129	0.6086	5.7815	7.17M	460.95G
	RIDCP	CVPR'23	19.86	0.618	0.3932	3.7959	19.87	<u>0.627</u>	0.5147	5.3657	<u>0.4796</u>	5.0843	28.72M	188.65G
	MB-Taylor	ICCV'23	16.49	0.616	0.6911	6.1694	19.49	0.609	0.7979	5.9100	0.5680	6.9914	7.43M	44.05G
	NSDNet	arXiv'23	19.92	<u>0.780</u>	<u>0.3031</u>	3.7884	<u>21.39</u>	0.619	0.4576	<u>5.3337</u>	0.4821	5.3629	11.38M	56.86G
	DAE-Net	TIP'24	<u>22.59</u>	0.778	0.5524	3.9841	21.14	0.574	0.8042	5.4397	0.7297	<u>5.1751</u>	3.65M	32.23G
	OKNet	AAAI'24	21.93	0.750	0.3379	<u>3.5567</u>	21.28	<u>0.627</u>	0.7124	6.0040	0.7512	6.0741	4.72M	39.67G
	DCMPNet	CVPR'24	21.01	0.767	0.3740	3.8295	20.13	0.587	0.4726	5.6546	0.6257	8.4573	7.16M	62.89G
	Ours	-	23.41	0.790	0.3042	3.4365	22.26	0.668	0.4001	5.0080	0.4611	5.2114	13.32M	53.40G

Table 1: Quantitative study on real-world smoke/haze datasets, \downarrow indicates that lower values are better, while \uparrow indicates that higher values are better. The best results are highlighted in bold, and the second-best results are underlined.

slightly trailing the top-performing RIDCP, while maintaining strong overall visual quality. Additionally, Fig. 7 shows that most methods struggle to effectively remove haze, particularly in dense fog conditions, while our approach significantly outperforms the others.

Ablation Study

Our proposed SGDN primarily incorporates the YCbCr color model, BGB, and CEM. To evaluate their contributions to the overall performance, we conduct a series of studies on our RW^2AH and real-world smoke datasets.

Effect of different color models. To verify the beneficial impact of the different color models, we conduct comparative experiments, with results presented in Table 2. The findings show that models using only a single color space are limited in performance. In contrast, models combining RGB and HSV color spaces outperform those using either space alone. This is because the hue and saturation channels in HSV lack edge information (Dudhane and Murala 2019), making it difficult to effectively enhance RGB features. In comparison, our SGDN recovers clean features with the help of YCbCr, achieving the best overall performance. Addi-

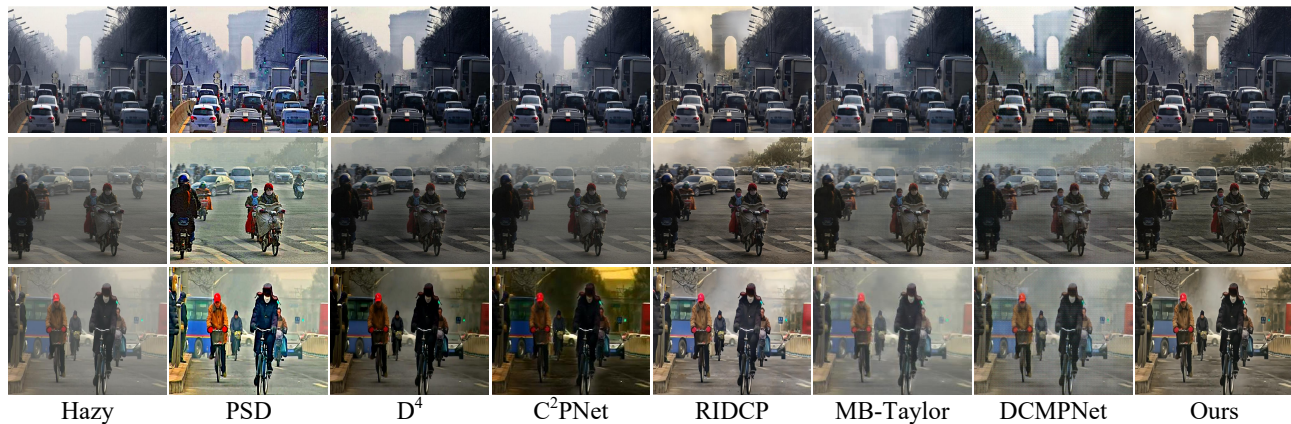


Figure 7: Visual comparison results on the RTTS dataset.

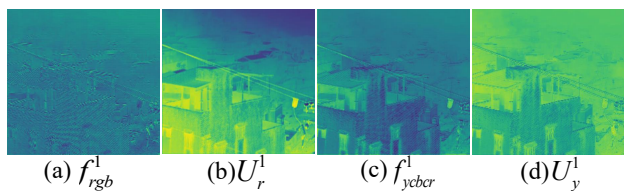


Figure 8: Visualization of BGB inputs (f^1_{rgb} and f^1_{ycbcr}) and outputs (U^1_r and U^1_y).

Ablations for color space	Real-world Smoke		Our RW ² AH			
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	FADE \downarrow	NIQE \downarrow
Only RGB	19.36	0.593	18.44	0.417	0.679	7.190
Only YCbCr	19.48	0.607	19.02	0.472	0.640	6.971
RGB+HSV	22.46	0.703	21.17	0.593	0.556	6.748
RGB+YUV	23.04	0.736	21.49	0.608	0.489	5.878
Ours	23.41	0.790	22.26	0.668	0.400	5.008

Table 2: Quantitative study of different color models on our RW²AH dataset and real-world smoke dataset.

tionally, we visualize the input and output of BGB in Fig. 8. The input f^1_{rgb} of BGB produces more detailed texture in U^1_r when guided by f^1_{ycbcr} . Meanwhile, U^1_y offers a more comprehensive global response after BGB processing. **Effect of BGB and CEM.** Table 3 reports the ablation results for BGB and CEM. We exclude BGB and CEM, and form a baseline by adding the YCbCr and RGB branches. The results show that BGB and CEM bring gains of 2.73 dB and 2.26 dB on our RW²AH dataset, respectively. The gain from BGB is more significant as it promotes clearer RGB features over Ycbr features. While the gain from CEM is smaller, it significantly improves visual quality. In Fig. 9, the model using CEM can restore better colors, bringing the images closer to the ground truth, thus validating its effectiveness. Finally, when both BGB and CEM are combined, our SGDNet achieves its best performance.

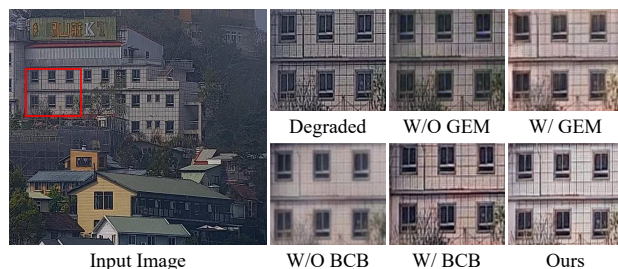


Figure 9: Ablation study of CEM. Our CEM can significantly improve the visual contrast of the image.

Ablations for BGB and CEM	Real-world Smoke		Our RW ² AH			
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	FADE \downarrow	NIQE \downarrow
Baseline	20.17	0.603	19.10	0.468	0.613	7.003
+BGB	22.75	0.686	21.83	0.641	0.497	6.077
+CEM	21.46	0.627	21.36	0.605	0.607	6.917
+BGB+CEM	23.41	0.790	22.26	0.668	0.400	5.008

Table 3: Ablation study of our proposed module.

Conclusion

In this paper, we propose the **Structure Guided Dehazing Network (SGDN)**, a novel framework specifically designed for the real-world dehazing. It simultaneously extracts features from both RGB and YCbCr color spaces to better collaborate in removing haze. To guide RGB features to produce clearer textures, we design the **Bi-Color Guidance Bridge (BGB)**. Given the more pronounced color information in the YCbCr color space, we propose the **Color Enhancement Module (CEM)** to enhance the natural color perception in the RGB color space. Due to the effectiveness of our designs, our SGDNet achieves optimal performance on several real-world haze datasets. Finally, to advance the development of real-world dehazing task, we construct a new trainable real-world well-aligned (RW²AH) dataset, which includes a diverse range of geographical regions and climate conditions worldwide, ensuring the dataset's diversity.

Acknowledgements

This work was supported by the National Science Fund of China under Grant Nos. 62072242, U24A20330 and 62361166670.

References

- Ancuti, C.; Ancuti, C. O.; and Timofte, R. 2018. Ntire 2018 challenge on image dehazing: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 891–901.
- Ancuti, C. O.; Ancuti, C.; and Timofte, R. 2020. 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*, 444–445.
- Ancuti, C. O.; Ancuti, C.; Timofte, R.; and De Vleeschouwer, C. 2018. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 754–762.
- Bui, T. M.; and Kim, W. 2017. Single image dehazing using color ellipsoid prior. *IEEE Transactions on Image Processing*, 27(2): 999–1009.
- Chen, X.; Fan, Z.; Li, P.; Dai, L.; Kong, C.; Zheng, Z.; Huang, Y.; and Li, Y. 2022. Unpaired deep image dehazing using contrastive disentanglement learning. In *European conference on computer vision*, 632–648. Springer.
- Chen, Z.; He, Z.; and Lu, Z.-M. 2024. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Transactions on Image Processing*.
- Chen, Z.; Wang, Y.; Yang, Y.; and Liu, D. 2021. PSD: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7180–7189.
- Choi, L. K.; You, J.; and Bovik, A. C. 2015. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11): 3888–3901.
- Cui, Y.; Ren, W.; Cao, X.; and Knoll, A. 2023. Image restoration via frequency selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Cui, Y.; Ren, W.; and Knoll, A. 2024. Omni-Kernel Network for Image Restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 1426–1434.
- Dudhane, A.; and Murala, S. 2019. RYF-Net: Deep fusion network for single image haze removal. *IEEE Transactions on Image Processing*, 29: 628–640.
- Engin, D.; Genç, A.; and Kemal Ekenel, H. 2018. Cycle-dehaze: Enhanced cycleGAN for single image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 825–833.
- Fan, J.; Guo, F.; Qian, J.; Li, X.; Li, J.; and Yang, J. 2023. Non-aligned supervision for real image dehazing. *arXiv preprint arXiv:2303.04940*.
- Fan, J.; Wang, K.; Yan, Z.; Chen, X.; Gao, S.; Li, J.; and Yang, J. 2025. Depth-Centric Dehazing and Depth-Estimation from Real-World Hazy Driving Video. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Fan, J.; Weng, J.; Wang, K.; Yang, Y.; Qian, J.; Li, J.; and Yang, J. 2024. Driving-Video Dehazing with Non-Aligned Regularization for Safety Assistance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 26109–26119.
- Fang, F.; Wang, T.; Wang, Y.; Zeng, T.; and Zhang, G. 2019. Variational single image dehazing for enhanced visualization. *IEEE Transactions on Multimedia*, 22(10): 2537–2550.
- Fattal, R. 2014. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1): 1–14.
- He, K.; Sun, J.; and Tang, X. 2010. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12): 2341–2353.
- Hu, X.; Zhong, B.; Liang, Q.; Zhang, S.; Li, N.; and Li, X. 2024. Towards Modalities Correlation for RGB-T Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*.
- Hu, X.; Zhong, B.; Liang, Q.; Zhang, S.; Li, N.; Li, X.; and Ji, R. 2023. Transformer tracking via frequency fusion. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(2): 1020–1031.
- Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026.
- Li, B.; Gou, Y.; Gu, S.; Liu, J. Z.; Zhou, J. T.; and Peng, X. 2021. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129: 1754–1767.
- Li, B.; Peng, X.; Wang, Z.; Xu, J.; and Feng, D. 2017. Aodnet: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, 4770–4778.
- Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; and Wang, Z. 2018. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1): 492–505.
- Li, Y.; Hou, Q.; Zheng, Z.; Cheng, M.-M.; Yang, J.; and Li, X. 2023. Large selective kernel network for remote sensing object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 16794–16805.
- Liu, J.; Wu, H.; Xie, Y.; Qu, Y.; and Ma, L. 2020a. Trident dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 430–431.
- Liu, W.; Zhou, F.; Lu, T.; Duan, J.; and Qiu, G. 2020b. Image defogging quality assessment: Real-world database and method. *IEEE Transactions on image processing*, 30: 176–190.

- Lyu, Z.; Chen, Y.; and Hou, Y. 2024. MCPNet: Multi-space color correction and features prior fusion for single-image dehazing in non-homogeneous haze scenarios. *Pattern Recognition*, 150: 110290.
- Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3): 209–212.
- Qiu, Y.; Zhang, K.; Wang, C.; Luo, W.; Li, H.; and Jin, Z. 2023. MB-TaylorFormer: Multi-branch efficient transformer expanded by Taylor formula for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12802–12813.
- Shao, Y.; Li, L.; Ren, W.; Gao, C.; and Sang, N. 2020. Domain adaptation for image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2808–2817.
- Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32: 1927–1941.
- Wang, A.; Wang, W.; Liu, J.; and Gu, N. 2018. AIPNet: Image-to-image single image dehazing with atmospheric illumination prior. *IEEE Transactions on Image Processing*, 28(1): 381–393.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.
- Wu, R.-Q.; Duan, Z.-P.; Guo, C.-L.; Chai, Z.; and Li, C. 2023. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22282–22291.
- Yang, Y.; Wang, C.; Liu, R.; Zhang, L.; Guo, X.; and Tao, D. 2022. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2037–2046.
- Yu, H.; Zheng, N.; Zhou, M.; Huang, J.; Xiao, Z.; and Zhao, F. 2022. Frequency and spatial dual guidance for image dehazing. In *European Conference on Computer Vision*, 181–198. Springer.
- Zhang, G.; Fang, W.; Zheng, Y.; and Wang, R. 2023. SDBAD-Net: A spatial dual-branch attention dehazing network based on meta-former paradigm. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(1): 60–70.
- Zhang, Y.; Zhou, S.; and Li, H. 2024. Depth Information Assisted Collaborative Mutual Promotion Network for Single Image Dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2846–2855.
- Zhao, S.; Zhang, L.; Huang, S.; Shen, Y.; and Zhao, S. 2020. Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines. *IEEE Transactions on Image Processing*, 29: 6947–6962.
- Zhao, S.; Zhang, L.; Shen, Y.; and Zhou, Y. 2021. RefineD-Net: A weakly supervised refinement framework for single image dehazing. *IEEE Transactions on Image Processing*, 30: 3391–3404.
- Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; and Chen, J. 2024. Detsr beat yolos on real-time object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16965–16974.
- Zheng, Y.; Zhan, J.; He, S.; Dong, J.; and Du, Y. 2023. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5785–5794.
- Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.
- Zhu, Q.; Mai, J.; and Shao, L. 2015. A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing*, 24(11): 3522–3533.