Structure Interaction Dehazing Network Combined with YCbCr Color Space for Real-World Image Dehazing

Fengnian Zhao 107552301703@STU.XJU.EDU.CN and **Kai Lv*** LVKAI@XJU.EDU.CN
School of Software, Xinjiang University, Xinjiang, China
**Corresponding author*

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Dehazing in the RGB space often causes artifacts and detail blurring due to the difficulty in separating luminance and color. Traditional encoder-decoder models also suffer from semantic discontinuity and insufficient multi-scale feature interaction. To address these issues, we propose a novel method called the Structure Interaction Dehazing Network (SIDN), which leverages the advantages of the YCbCr color space in separating color and luminance to guide RGB feature extraction. SIDN consists of two core components: the Dual-space Feature Branch (DFB) and the Cross-Feature Block (CFB). The DFB integrates YCbCr features through the Phase Fusion Module (PFM) and Density-aware Feature Extraction Block (DFEB), enhancing texture recovery and guiding RGB feature reconstruction. The CFB improves multi-scale feature interaction and semantic alignment through an enhanced cross-layer mechanism. Experimental results show that SIDN achieves a 1.25dB improvement in PSNR and SSIM metrics on real-world datasets compared to previous methods, and also outperforms the latest methods in terms of FADE and NIQE metrics.

Keywords: Transformer; Image Dehazing; Frequency Domain; YCbCr Color Space

1. Introduction

Image dehazing is to recover a clear haze-free image from a hazy and blurred one. The formation of hazy images can typically be modeled using the Atmospheric Scattering Model:

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

where $I(x)$ and $J(x)$ represent the foggy image and clean image, and $A(x)$ and $t(x)$ represent the global atmospheric light and the transmission map, respectively. The transmission map $t(x) = e^{-\beta d(x)}$ is determined by the scattering coefficient β and the scene depth $d(x)$. Early image dehazing methods evolved from CNN-based techniques to utilizing deep learning for estimating physical parameters, aiming to recover clear images. With the application of Transformers in image processing, Transformer-based approaches have facilitated the development of end-to-end dehazing models by directly converting hazy images to clear ones through supervised learning. However, previous methods face limitations in real-world applications: Firstly, uneven haze distribution obscures texture details and complicates separating luminance from chrominance in RGB space, making local detail recovery challenging and often leaving residual haze and artifacts. Secondly, obtaining large-scale paired images (hazy/clear) with consistent backgrounds is difficult, prompting reliance on synthetic datasets for training, which limits real-world performance.

In previous research, the chrominance components in YCbCr are less affected by haze, preserving color and detail better. We propose a Structure Interaction Dehazing Network using YCbCr. It

employs an asymmetric interactive encoder-decoder structure, sharing encoders for both RGB and YCbCr spaces. We design a Dual-space Feature Branch (DFB) to enhance structure and refine textures. DFB has two modules: Phase Fusion Module (PFM) and Density-aware Feature Extraction Block (DFEB). PFM fuses RGB and YCbCr phase spectra in the frequency domain to reconstruct clearer textures. DFEB extracts haze density from dark channel images, guiding RGB feature reconstruction. For better network learning, we introduce the Cross-Feature Block (CFB) to replace skip connections, allowing high-level semantic layers to refine low-level feature extraction, improving overall network performance.

Our contributions are summarized as follows:

- We propose a novel method called the Structure Interaction Dehazing Network (SIDN), which effectively leverages the clear textures in the YCbCr color space, less affected by haze, to guide the dehazing process of real-world images.
- We designed the Dual-space Feature Branch (DFB) to integrate frequency and spatial relationships, boosting RGB space representation. Within the DFB, the Density-aware Feature Extraction Block (DFEB) uses the dark channel prior to enhance haze perception and real-world adaptability.
- To reduce semantic loss and enhance feature interaction, we designed cross-feature blocks (CFBs). This module facilitates multi-scale feature fusion, improves the network's ability to capture relevant information, and enhances detail alignment in restored images.

2. Related Work

Image dehazing mainly involves prior-based methods and deep learning-based approaches. This chapter will provide an overview of the research progress in these two areas as well as the existing color models utilized for image dehazing.

2.1. Prior-based dehazing methods

Prior-based dehazing techniques primarily rely on the atmospheric scattering model, utilizing manually designed priors or statistical learning to estimate physical parameters. For instance, the DCP (He et al., 2009) method estimates scene depth to compute transmission and reflection maps, while RefineDNet (Zhao et al., 2021) calculates transmission rates by analyzing the relationship between color and brightness in hazy images. However, these methods heavily depends on the accuracy of parameter estimation, which may lead to suboptimal results in complex real-world environments.

Our method employs deep learning to automatically extract features, eliminating the need for complex hand-crafted prior rules, thereby enhancing flexibility and efficiency in handling diverse haze conditions.

2.2. Deep learning-based dehazing methods

End-to-end dehazing based on deep learning can directly restore clear images from blurry inputs. Some studies focus on real-world unpaired image dehazing, such as the D4 (Yang et al., 2022) models based on the CycleGAN framework. However, these GAN-based methods may generate artifacts due to insufficient constraints. An alternative approach involves leveraging prior knowledge for

model optimization, such as the PSD (Chen et al., 2021) algorithm proposed, which uses prior loss functions and unsupervised fine-tuning to enhance dehazing in real-world settings. Nonetheless, methods relying on manually defined priors still struggle to overcome their inherent limitations.

In this paper, we leverage the YCbCr color space's advantage over RGB in preserving image texture features, guiding the model to learn features that are not influenced by haze, thereby achieving more effective dehazing.

2.3. Color model in Image Dehazing

Currently, several research approaches focus on image dehazing in the HSV, YCbCr, and YUV color spaces. AIPNet (Wang et al., 2019) enhances the contrast of the Y channel using a multi-scale network to effectively remove haze. However, this method relies on a single assumption and shows limitations when dealing with complex scenes. In contrast, MPCNet (Lyu et al., 2024) improves model robustness by incorporating color priors from the HSV and YCbCr color spaces for correction. These methods are prone to introducing color errors due to their reliance on ASM parameter estimation and color conversion processes. Recently, it has been proposed in SGDN (Fang et al., 2024) to use the properties of the YCbCr color space to guide the reconstruction of haze-independent features, so as to improve the accuracy of dehazing results. However, increasing haze concentration significantly reduces the color and contrast of distant objects. Relying only on depth estimation to correct for atmospheric scattering limits the model's real-world adaptability. This enables targeted dehazing for areas with different haze levels, effectively restoring color, contrast, and detail in the image.

In our approach, we use the dark channel prior to assess local haze density, preventing over-processing that introduces noise or artifacts. This enables targeted dehazing for areas with different haze levels, effectively restoring color, contrast, and detail in the image.

3. Methodology

This section introduces our proposed SIDN framework, which consists of the Dual-space Feature Branch and the Cross-Feature Block. We will provide a detailed description from these two core components.

3.1. Overall framework

The overall framework of SGDN, as shown in Figure 1 (a), employs an interactive asymmetric encoder-decoder structure. First, the hazy image $I(x) \in R^{C\times W\times H}$ undergoes linear transformation and dark channel calculation to obtain the YCbCr feature map $Y(x) \in R^{C\times W\times H}$ and the dark channel image $I_{\text{dark}}(x) \in R^{1\times W\times H}$. Subsequently, the same encoder branch is used to extract features from different color spaces. These features, along with the dark channel image, are then fed into the Dual-space Feature Branch (DFB). Here, the Phase Fusion Module (PFM) and Density-aware Feature Extraction Block (DFEB) work together to generate rich, refined textures. The visualization of the feature maps of the RGB and YCbCr color spaces is shown in Figure 2. Finally, the Cross-Feature Block (CFB) integrates global and local features to enhance multi-scale information interaction, thereby improving the overall visual quality and color representation of the image.

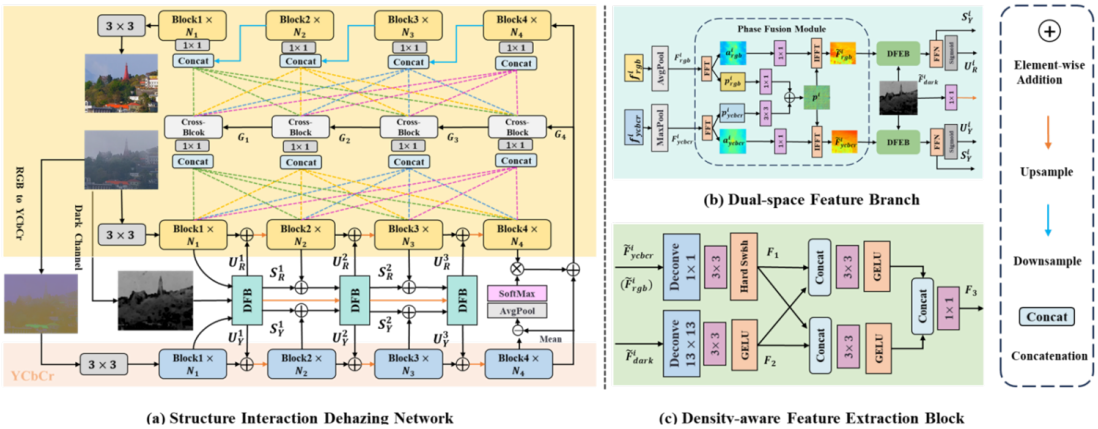

Figure 1: (a) Shows the overall framework of SIDN. (b) Introduces the DFB structure. (c) Presents the design of CFB.

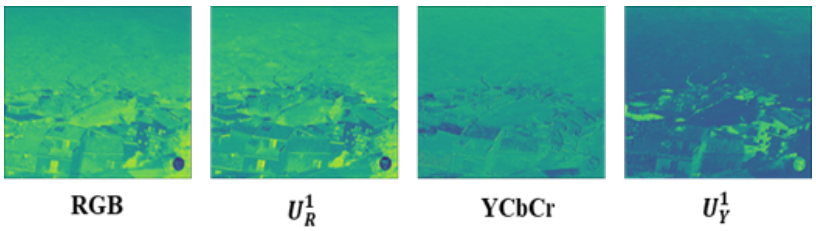

Figure 2: Visualization of DFB inputs and outputs. U_R^1 captures rich details but struggles to separate luminance and chrominance. U_Y^1 distinguishes them more clearly.

3.2. Dual-space Feature Branch

In the RGB color space, rich structural and textural details can be obtained, but haze makes it difficult to distinguish luminance from chromatic components. Conversely, the YCbCr color space easily separates luminance and chrominance. The Dual-space Feature Branch exploits both spaces, enhancing detail and color capture while improving structural integrity and textural refinement.

To balance global and local detail extraction, as shown in Figure 1(b), we apply Adaptive Average Pooling to RGB features $f_{rgb}^i \epsilon R^{C\times W\times H}$ and Max Pooling to YCbCr features $f_{ycbcr\epsilon R}^i C\times W\times H$. This reduces noise while preserving details and color. Specifically, RGB features are smoothed for global structure, and YCbCr features are enhanced for detail saliency and robustness to changes. This process can be defined as follows:

$$F_{rgb}^i, F_{ycbcr}^i = \text{AvgPool}\left(f_{rgb}^i\right), \text{MaxPool}\left(f_{ycbcr}^i\right) \tag{2}$$

3.2.1. PHASE FUSION MODULE

Studies have shown that phase spectra capture image structure better than amplitude spectra, with minimal difference between sharp and blurry images in phase spectra. The PFM applies a Fast

Fourier Transform to convert spatial features into the frequency domain, then separates the phase from the magnitude components.

$$\begin{aligned} a^i_{rgb}, p^i_{rgb} &= FFT\left(F^i_{rgb}\right) \\ a^i_{ycbcr}, p^i_{ycbcr} &= FFT\left(F^i_{ycbcr}\right) \end{aligned} \tag{3}$$

where a^i_{rgb} and a^i_{ycbcr} denote the phase spectra, while p^i_{rgb} and p^i_{ycbcr} represent the magnitude component. $FFT(\cdot)$ represent the Fourier transform and inverse Fourier transform operations. Subsequently, convolution operations are used to recover amplitude information and extract detailed structures. Two phase components are then linearly combined to form the mixed phase spectrum p^i . Finally, using p^i and the magnitude spectrum, an inverse Fourier transform reconstructs the spatial features. This process can be expressed as:

$$\begin{aligned} p^i &= \mathrm{Conv}_{1\times1}\left(p^i_{ycbcr}\right) + \mathrm{Conv}_{3\times3}\left(p^i_{rgb}\right) \\ \tilde{F}^i_{rgb} &= \mathrm{IFFT}\left(\mathrm{Conv}_{1\times1}\left(p^i_{rgb}\right), a^i_{rgb}\right) \\ \tilde{F}^i_{ycbcr} &= \mathrm{IFFT}\left(\mathrm{Conv}_{1\times1}\left(p^i_{ycbcr}\right), a^i_{ycbcr}\right) \end{aligned} \tag{4}$$

where $\tilde{F}^i_{rgb}$ and $\tilde{F}^i_{ycbcr}$ are outputs of the DFB. $IFFT(\cdot)$ represents the inverse Fast Fourier Transform. $\mathrm{Conv}_{1\times1}(\cdot)$ denote convolution operations.

3.2.2. DENSITY-AWARE FEATURE EXTRACTION BLOCK

Figure 1(c) depicts the DFEB architecture, using a dual-branch design. One branch extracts structural features from PFM outputs using dynamic convolution layers, followed by enhancement via 1×1 convolutions and the HardSwish function to produce feature representation F_1. The other branch processes dark channel images $\tilde{F}^i_{\text{dark}}$, employing dilated convolutions with large kernels for capturing small-scale features, then refines these through 1×1 convolutions with GELU activation to obtain F_2.

To optimize feature utilization across both branches, the architecture merges representations F_1 and F_2. It further processes this combined data through 3×3 convolutions with GELU activation to refine structural features and 1×1 convolutions with GELU activation for dark channel features. This integration, followed by dimensionality reduction, reconstructs the ultimate fog feature F_3. This allows later network stages to leverage haze features at multiple scales, enhancing dehazing with precise and comprehensive guidance. This process can be expressed as:

$$\begin{aligned} F_1 &= \mathrm{Hard}\left(\mathrm{Conv}_{3\times3}\left(\mathrm{Deconv}_{1\times1}\left(\tilde{F}^i_{rgb}\right)\right)\right), \\ F_2 &= \mathrm{GELU}\left(\mathrm{Conv}_{3\times3}\left(\mathrm{Deconv}_{13\times13}\left(\tilde{F}^i_{dark}\right)\right)\right), \\ F'_1 &= \mathrm{GELU}\left(\mathrm{Conv}_{3\times3}\left(F_1\right)\right), \\ F'_2 &= \mathrm{GELU}\left(\mathrm{Conv}_{3\times3}\left(F_2\right)\right), \\ F_3 &= \mathrm{Concat}\left[\hat{F}_1, \hat{F}_2\right], \end{aligned} \tag{5}$$

where $\mathrm{Hard}(\cdot)$ and $\mathrm{Deconv}_{i\times i}(\cdot)$ represent the HardSwish function and dynamic convolution layers, respectively. GELU($\cdot$) and Concat $[\cdot]$ represent activation functions and aggregation operations, respectively. $\mathrm{Conv}_{3\times3}(\cdot)$ represent the convolution operation.

3.3. Cross-Feature Block

The encoder-decoder architecture effectively captures multi-scale information but struggles with inefficient multi-level feature interaction via skip connections, causing semantic inconsistencies. We introduce a CFB. replaces skip connections, integrating multi-scale features to enrich semantic information.

This enhances the model's efficiency in utilizing various scale features and improves semantic consistency and integrity.

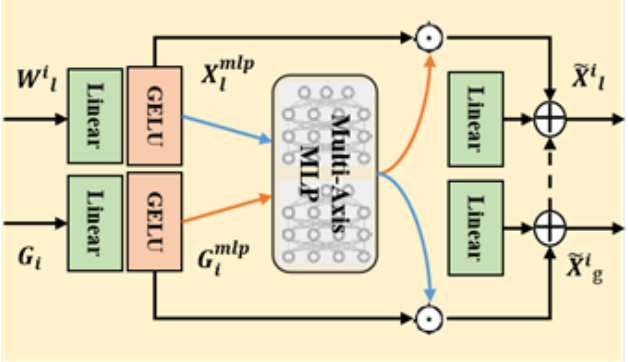

Figure 3: The structure of Cross-Feature Block (CFB).

In Figure 3, CFB employs a dual-stream input: a matrix $W^i{}_l$ formed from local features X^i of multiple encoder layers and global feature G_i from previous CFB output. These features undergo channel aggregation and size adjustment, followed by dimension adjustment using a 1×1 convolution layer. Subsequently, $W^i{}_l$ and G_i are processed through a linear layer followed by the GELU activation function and then fed into a Multi-Axis MLP (Tu et al., 2022) for computation. This generates multi-scale features $\tilde{X}_l^i$ that contain local semantic information and a semantically rich global feature $\tilde{X}_g^i$. The computation process can be summarized as follows:

$$\begin{aligned} X^i{}_l &= \text{Smaple}\left(X^i\right), i = 1, 2, 3, 4, \\ W^i{}_l &= \text{Conv}_{1\times1}\left(\text{Concat}\left[X^1{}_l, X^2{}_l, X^3{}_l, X^4{}_l\right]\right), \\ X_i^* &= \text{GeLu}\left(\text{Linear}\left(W^i{}_l\right)\right), \\ G_i^* &= \text{GeLu}\left(\text{Linear}\left(G_i\right)\right), \\ \tilde{X}^i{}_l = X_i^* \odot MLP\left(G_i^*\right), \tilde{X}^i{}_g &= G_i^* \odot MLP\left(X_i^*\right), \end{aligned} \tag{6}$$

where Linear$(\cdot)$ and MLP$(\cdot)$ represent the processing through a linear layer and Multi-Axis MLP, respectively. Additionally, $\odot$ andSmaple $(\cdot)$ denote element-wise multiplication and sampling operations, respectively.

4. Experimental Setup and Result Analysis

4.1. Datasets and Evaluation Metrics

We conducted qualitative and quantitative analyses on three real-world smoke/haze datasets. To evaluate the dehazing performance, we used PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) to measure the differences between the dehazed images and the original

haze-free images. Additionally, we utilized non-reference metrics, specifically FADE and NIQE, to evaluate the clarity and level of distortion in the dehazed images.

RW2AH dataset: RW2AH (Fang et al., 2024) is a real-world dataset that encompasses a wide range of climates and scenes captured from diverse geographical locations. It consists of 1,406 image pairs for training and 352 pairs for testing, with images featuring different resolutions. The precise alignment of background details between the clear and hazy images in the RW2AH dataset makes it an ideal choice for supervised learning in practical dehazing applications.

Real-world datasets: Real-world (Fang et al., 2024) comprises the I-HAZE, O-Haze, NH-Haze, and NH-Haze-v2 subsets, featuring both indoor and outdoor environments with images exhibiting both uniform and non-uniform haze patterns. The dataset contains 155 paired images, of which 147 pairs are utilized for training and 8 pairs are reserved for validation.

RTTS Dataset: RTTS is a component of the RESIDE (Li et al., 2019) dataset, consisting of 4,322 real-world hazy images primarily focusing on traffic and driving scenarios. Due to the absence of corresponding clear images, RTTS serves exclusively as a test set for evaluating model performance.

4.2. Comparison with State-of-the-Art Dehazing Methods

To validate the effectiveness of SIDN in real-world haze, we compared it with several representative SOTA methods. For models with publicly available weights, we directly tested them on three real-world datasets. Otherwise, we retrained the models using default parameters. Table 1 shows the quantitative comparison results for these three real-world hazy datasets.

Table 1: The quantitative study on real dehazing datasets. Best metrics are highlighted in bold, and the second-best results are underlined.

Methods	Venue	Real-world				RW2AH				RTTS		Overhead	
		PSNR	SSIM	FADE	NIQE	PSNR	SSIM	FADE	NIQE	FADE	NIQE	Params	FLOPS
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
D4	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**
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
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
RIDCP	CVPR'23	19.86	0.618	0.3932	3.7959	19.87	0.627	0.5147	5.3657	0.4796	**5.0843**	28.72M	188.65G
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
SGDN	AAAI'25	**23.41**	**0.790**	**0.3042**	3.4365	22.26	0.668	0.4001	5.0080	0.4611	5.2114	13.32M	53.40G
Ours	-	23.26	0.776	0.3257	**3.4275**	**22.38**	**0.679**	**0.3827**	**4.9893**	**0.4597**	5.1367	16.96 M	68.71 G

Comparison on the RW2AH dataset: Figure 4 compares various dehazing methods on the RW2AH dataset. Our method not only effectively removes the haze, but also maintains the natural contrast of the image. In contrast, PSD is unable to completely remove haze and causes color distortion. RIDCP (Wu et al., 2023) still show a noticeable residual haze in their dehazing images. DehazeFormer (Song et al., 2022) and DCMPNet (Zhang et al., 2024) can effectively eliminate smoke, but they fall short in preserving image detail and texture. In addition, the quantitative comparison in Table 1 shows the superiority of our method in terms of evaluation indicators. Only slightly inferior to the latest SGDN on the FADE indicator.

Comparison on the Real-World dataset: According to Table 1, SIDN outperforms other state-of-the-art methods in terms of performance. Specifically, it has a 0.12dB increase in PSNR, a 0.011dB increase in SSIM, a 0.018 decrease in FADE, and a 0.01 decrease in NIQE compared

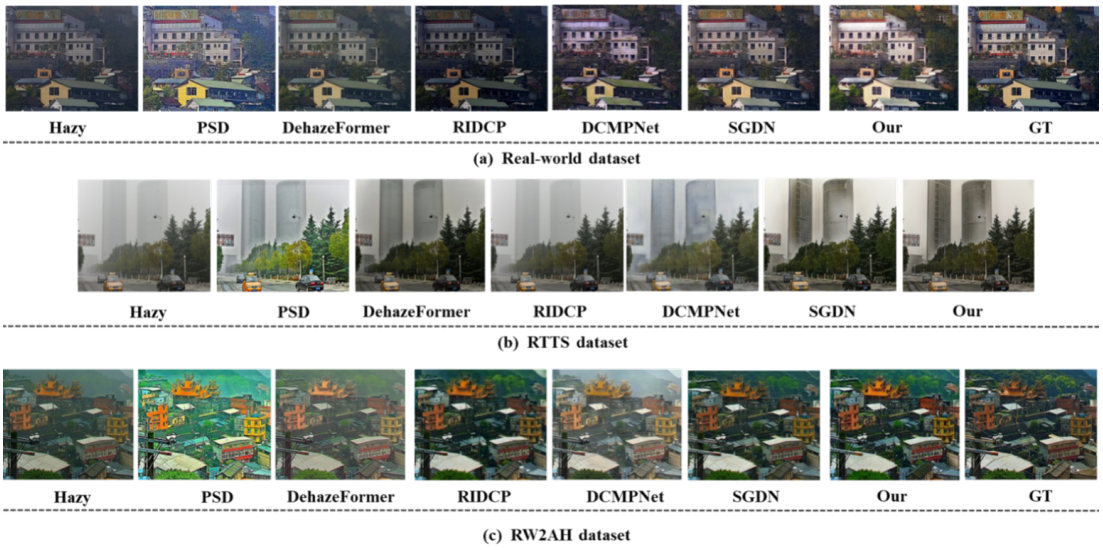

Figure 4: The visualization comparison results on the Real-world, RW2AH, and RTTS datasets.

to SGDN. Visual comparisons in Figure 4 show that methods relying on the RGB color space, such as PSD, DehazeFormer, and RIDCP, perform poorly in high-density haze scenarios. Although DCMPNet performs relatively well, it still shows slight color deviations. In contrast, SIDN utilizes the characteristics of the YCbCr color space and haze density awareness to generate nearly haze-free clear images, achieving results comparable to SGDN in terms of evaluation metrics.

Comparison on the RTTS dataset: As shown in Table 1, SIDN performs very well on the FADE metric, demonstrating its strong dehazing capabilities. Although it is slightly lower than RIDCP and SGDN on NIQE indicators, satisfactory results have been achieved. As shown in Figure 4, our method is significantly superior to other methods under heavy haze conditions.

4.3. Ablation study

In this section, to evaluate the contribution of the core components DFB and CFB to the overall performance of SIDN, we conduct ablation experiments on the RW2AH dataset.

Effect of core component in DFEB: To evaluate the performance impact of each component in DFB, YCbCr and RGB branches were introduced into the baseline model, followed by the incremental addition of DFB subcomponents. Table 3 shows that DFEB achieves the most significant improvement in PSNR, with a gain of 2.15 dB. Although PFM provides a lower gain than DFEB and only boosts 1.26dB, it plays a positive role in improving the visual quality of the image, as shown in Figure 6.

Effect of DFB and CFB: Table 2 demonstrates that the baseline model, built on YCbCr and RGB branches, achieves performance improvements of 2.67 dB and 1.61 dB when DFB and CFB are added separately. The results show that DFB contributes significantly higher gains by effectively enhancing image reconstruction quality. Although the improvement of CFB on the PSNR metric is marginal, with only a 1.61dB increase, as a skip connection mechanism, it effectively mitigates information loss during feature propagation. As shown in Figure 5, the feature maps output by the

encoder and decoder layers exhibit strong alignment of semantic information across hierarchical levels, indicating that information is better preserved during inter-level transmission.

Effect of different color models: We designed comparative experiments in various color spaces, with results shown in Table 4. Models relying solely on the RGB color space exhibit limited performance. Combining RGB and HSV improves PSNR by 0.76dB. As shown in Figure 7, the hue and saturation channels in the HSV color space lack edge information, making it difficult to effectively enhance RGB features. The YUV color space struggles to preserve color information and texture details. In contrast, the model combining RGB and YCbCr color spaces performs better, outperforming models using any single color space or other combinations.

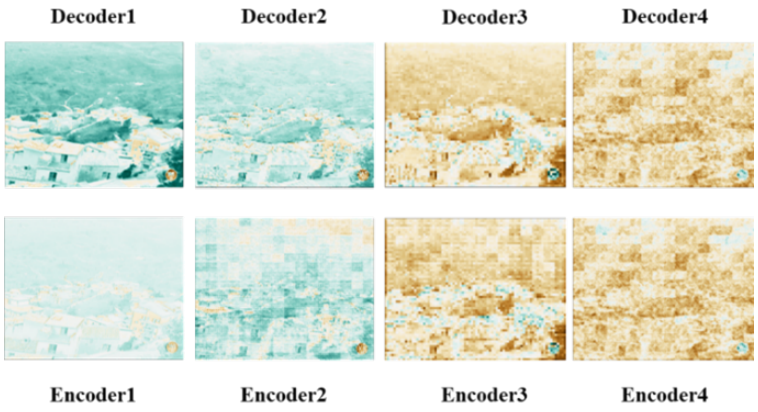

Figure 5: The feature maps output by the encoder layer and the feature maps output by the decoder.

Figure 6: Ablation study of PFM and DFEB. Under the effect of PFM, the contrast and color saturation of the image are enhanced, and DFEB removes local haze more effectively.

5. Conclusions

This paper presents the Structure Interaction Dehazing Network (SIDN), which enhances dehazing by integrating RGB and YCbCr color spaces. SIDN uses haze-resistant texture from YCbCr to guide the process. The Dual-space Feature Branch (DFB) combines YCbCr details with RGB for better representation. Inside the DFB, the Density-aware Feature Extraction Block (DFEB) utilizes the dark channel prior to improve haze detection and real-world adaptability. The Cross-Feature Block

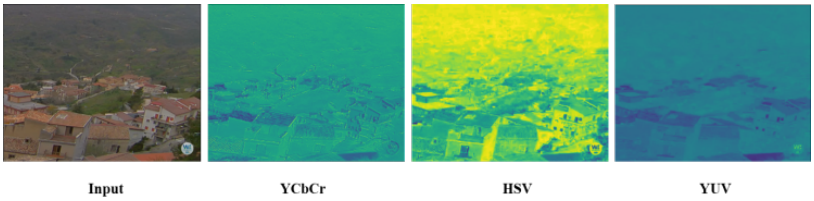

Figure 7: Visualization of the Hue and Saturation Channels in the HSV Color Space, the YUV Color Space, and the YCbCr Color Space.

Table 2: Ablation study of our proposed module.

Ablations study for DFB and CFB	RW2AH		Real-word		Overall	
	PSNR	SSIM	PSNR	SSIM	Params	FLOPS
Baseline	19.08	0.458	20.13	0.604	10.30M	19.14G
Baseline+DFB	21.75	0.646	22.63	0.671	13.11M	53.75M
Baseline+CFB	20.69	0.663	21.36	0.605	16.17M	67.35 G
Baseline+CFB+DFB	**22.38**	**0.679**	**23.26**	**0.776**	16.96 M	68.71 G

Table 3: Ablation study of PFM and dfeb on the RW2AH dataset

Ablations study for the core component of DFB	RW2AH		Overall	
	PSNR	SSIM	Param	FLOPS
Baseline	19.08	0.458	10.30M	19.14G
Baseline +PFM	20.32	0.504	10.62M	20.34G
Baseline +DFEB	21.23	0.518	12.78M	47.40G
Baseline+DFB	**21.75**	**0.646**	13.11M	53.75M

Table 4: Ablation study of PFM and dfeb on the RW2AH dataset

Ablations study for color models	RW2AH	
	PSNR	SSIM
RGB	19.08	0.458
YCbCr	20.32	0.514
RGB+HSV	21.03	0.562
RGB+ YUV	21.69	0.638
RGB+ YCbCr	**22.38**	**0.679**

(CFB) promotes multi-scale feature fusion, reducing semantic loss and detail errors in restored images. Experimental results demonstrate that, compared to previous methods, SIDN achieves significant improvements in PSNR and SSIM metrics on real-world datasets, while also surpassing the latest methods in terms of FADE and NIQE evaluations.

However, our method struggles to extract meaningful features when handling real images with small but dense fog patches that cause sharp visibility drops. Consequently, the model may occasionally introduce artifacts in the dehazing results, particularly under extreme dense fog conditions.

Future work should focus on improving the model's adaptability to varying fog conditions and reducing artifact generation.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 52275003, in part by the National Key Research and Development Program of China under Grant 2023YFB4704000; Tianshan Talent Training Program (NO. 2023TSYCLJ0052)

References

Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7176–7185, 2021. doi: 10.1109/CVPR46437.2021.00710.

Wenxuan Fang, Jankai Fan, and et al. Guided real image dehazing using ycbcr color space. In *AAAI Conference on Artificial Intelligence*, 2024.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963, 2009. doi: 10.1109/CVPR.2009.5206515.

Boyi Li, Wenqi Ren, Dengpan Fu, and et al. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. doi: 10.1109/TIP.2018.2867951.

Zhiyu Lyu, Yan Chen, and Yimin Hou. Mcpnet: Multi-space color correction and features prior fusion for single-image dehazing in non-homogeneous haze scenarios. *Pattern Recognition*, 150: 110290, 2024. doi: https://doi.org/10.1016/j.patcog.2024.110290.

Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2022.

Zhengzhong Tu, Hossein Talebi, Han Zhang, and et al. Maxim: Multi-axis mlp for image processing. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5759–5770, 2022. doi: 10.1109/CVPR52688.2022.00568.

Anna Wang, Wenhui Wang, Jinglu Liu, and Nanhui Gu. Aipnet: Image-to-image single image dehazing with atmospheric illumination prior. *IEEE Transactions on Image Processing*, 28(1): 381–393, 2019. doi: 10.1109/TIP.2018.2868567.

Rui-Qi Wu, Zheng-Peng Duan, and et al. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22282–22291, 2023. doi: 10.1109/CVPR52729.2023.02134.

Yang Yang, Chaoyue Wang, and et al. Self-augmented unpaired image dehazing via density and depth decomposition. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2027–2036, 2022. doi: 10.1109/CVPR52688.2022.00208.

Yafei Zhang, Shen Zhou, and Huafeng Li. Depth information assisted collaborative mutual promotion network for single image dehazing. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2846–2855, 2024. doi: 10.1109/CVPR52733.2024.00275.

Shiyu Zhao, Lin Zhang, Ying Shen, and Yicong Zhou. Refinednet: A weakly supervised refinement framework for single image dehazing. *IEEE Transactions on Image Processing*, 30:3391–3404, 2021. doi: 10.1109/TIP.2021.3060873.