

# Spatial Variance of Fusion and Physics-Aware for Single Image Dehazing

Wan Li<sup>1</sup>, Sha Huang<sup>1,2</sup>, Hangfei Wang<sup>1,2</sup>, Chenyang Chang<sup>1,2</sup>

<sup>1</sup> The School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi, China

<sup>2</sup> Shaanxi Joint Laboratory of Artificial Intelligence, Shaanxi University of Science and Technology Xi'an Shaanxi, 710021, China

**Abstract:** Currently, most image dehazing algorithms overlook the local details of the image and fail to fully exploit different levels of features, resulting in color distortion, decreased contrast, and residual haze in the restored haze-free images. To tackle these issues, this paper proposes a method that dynamically enhances pixels by utilizing the spatial variation of image dark channel prior in images. For regions with high haze density where local information is insufficient, a Transformer is employed to learn the global dependencies of the input features. Conversely, for regions with low haze density where local information is effective, parallel multi-scale attention is used to extract local features. When enhancing each pixel, we dynamically determine the contribution of non-local and local information based on the image features. Furthermore, to better reflect the physical process of haze formation in the image and improve the interpretability of the feature space, a dual-branch physics-aware unit is established. It learns the features related to atmospheric scattering in the image and captures the visual characteristics. In the experiments, a large dataset of dehazing images is used for training and testing, and comparisons are made with other existing dehazing methods. The results demonstrate that this method, which takes into account both local and global information, significantly improves the dehazing performance of the model.

**Keywords:** Image Dehazing; Deep Learning; Spatial Variance Fusion; Physics-aware.

## 1. Introduction

During hazy weather, the air contains numerous floating particles that interact with light during its propagation, causing scattering and resulting in images with low contrast, blurred textures, and edges of objects within the scene. These issues significantly impair the performance of subsequent advanced visual tasks, such as vehicle recognition[1] and scene understanding [2]. Therefore, research on image dehazing is of great importance.

In 1975, McCauley[3] derived the atmospheric scattering model to describe the formation mechanism of hazy images. However, single-image dehazing based on the atmospheric scattering model is an ill-posed problem. Traditional dehazing methods often rely on prior knowledge or constrained equations to estimate the parameters of the physical model [4][5][6][7]. such as He[6] discovered dark channel prior (DCP), the DCP is combined with the haze imaging model to directly estimate the haze thickness and the haze-free image is recovered subsequently. Zhu[7] developed a color attenuation prior, established a linear scene depth model for the hazy image, and supervised learning the model parameters. While these methods have made significant progress, they might fail in some instance.

Different from traditional methods, deep learning-based image dehazing methods obtain haze-free images by directly or indirectly learning mapping functions based on large-scale datasets. Early convolutional neural network-based learning models estimated the parameters in the atmospheric scattering model. DehazeNet[8] and MSCNN[11] are early image dehazing networks that used neural networks to estimate  $t(x)$  and prior-based methods to estimate  $A$ . these methods enhance the accuracy of parameter estimation, yet they can lead to biases in parameter estimation. To avoid complex estimations of the transmission and atmospheric light, end-to-end trainable deep neural networks have been developed.

GridDehazeNet [24] uses a grid-like neural network to obtain the multi-scale features. FFA-Net[10] improved the effect of image dehazing by using a lot of channel Attention and pixel attention. DeHamer [14] combines CNN and Transformer for image dehazing, which can aggregate long-term attention in Transformer and local attention in CNN features. Dehazformer [13] referred to the Swin transformer and modifies the key structure of the Swin transformer to make it more suitable for image dehazing. Although both CNN-based and Transformer-based methods can leverage large effective receptive fields to enhance performance, the multi-scale characteristics of images are always ignored during the dehazing processes, and note that these methods do not consider the physics of the hazing process.

To address these issues, we propose spatial variance of fusion and physics-aware for image dehazing named SVFPA-Net. The spatial variance of fusion leverages Transformer to concentrates on global features and catches regions with significant haze, while parallel multi-scale attention can extract location-dependent local information of the original feature, making it capable of dealing with an uneven hazy distribution effectively. During feature fusion, the pixel-wise dark channel prior dynamically determines the contributions of local and non-local information, effectively addressing the irregularity of haze distribution and differences in haze density, while preserving both the fine details and overall structure of the image. In addition, a physics-aware unit is proposed to predict latent features for atmospheric light and transmission maps. This unit enhances the network's utilization of physically-informed priors during feature extraction and synthesis, leading to more precise feature synthesis of the latent clear image. In brief, the main contributions of our study are as follows:

- We propose SVFPA-Net. It can simultaneously recover texture details and capturing large hazed areas.

while the physics-aware promotes feature transmission and extraction in the feature space.

- During spatial variance of fusion, the dark channel prior is utilized to dynamically determine the contributions of local and non-local information, enhancing pixels in different regions.
- The physics-aware primarily focuses on learning the features related to atmospheric scattering in the image. It then integrates these features with data-driven features.

The remainder of this paper is organized as follows: Section 2 provides the preliminaries of research on single image dehazing. Next, the SVFPA-Net scheme is proposed in Section 3. Experiments on benchmark data sets along with the analysis are presented in Section 4. Section 5 concludes the paper.

## 2. Preliminaries

### 2.1. The Dark Channel Prior

Traditional dehazing algorithms typically rely on the atmospheric scattering model to estimate intermediate parameters such as the transmission map and atmospheric light. The atmospheric scattering model, as shown in equation (1), is widely used in this process. Here,  $I(x)$  represent the hazy image,  $J(x)$  represents the clear image,  $t(x)$  represents the transmission map, and  $A$  represents the constant global atmospheric light. By considering prior assumptions, the parameters of the atmospheric scattering model can be established.

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

The dark channel prior dehazing algorithm is built upon the atmospheric scattering model. The fundamental assumption of the dark channel prior is that, in most outdoor scenes, there exists at least one pixel whose intensity value is close to zero in at least one-color channel, known as the dark channel. This channel represents the minimum intensity in the image and can be used to estimate the level of haze and recover image details. The dark channel is defined by equation (2).

$$J_{\text{dark}}(x) = \min_{y \in \Omega(x)} \left( \min_{c \in \{r, g, b\}} J^c(y) \right) \quad (2)$$

where  $J_{\text{dark}}$  is a color channel of  $J$  and  $\Omega(x)$  is a local patch centered at  $x$ .

### 2.2. Transformer

Swin Transformer[25] partitions tokens into windows and performs self-attention within a window to keep the linear computational cost. It employs the cyclic shift scheme to bridge windows so that adjacent blocks adopt different window partitions. Since then, many follow-ups to Swin Transformer have been proposed. For example, SwinIR[26] is one of the pioneers to employ Swin Transformer in low-level vision tasks, which builds a large residual block consisting of stacked Swin Transformer layers and a subsequent convolutional layer. Uformer[27] uses Swin Transformer blocks to build a U-Net-like network and inserted depth-wise convolution (DWConv) in the feed-forward network (FFN) like LocalViT. However, it has been observed that these methods exhibit poor performance in the context of image dehazing. This can be attributed to their inheritance of the normalization layer, window partitioning scheme, and activation function from the original Swin Transformer. In light of this, Song et al.[13] proposed a novel

image dehazing network called DehazeFormer-Net, which combines the Swin Transformer with U-Net. By improving the normalization layer, activation function, and spatial information aggregation scheme of the Swin Transformer, Dehaze Former aims to make the Transformer more suitable for image dehazing tasks.

## 3. Methodology

Figure 1 shows the overview of our framework. the network comprises a Transformer branch for capturing global contextual information and building global dependencies. The parallel multi-scale attention branch, through stacking multifunctional modules, extracts diverse feature information, complementing the features extracted by the Transformer. The network incorporates a dark channel prior related to haze density; higher haze concentrations increase the weight of Transformer features, focusing more on global structure modeling, while lower concentrations enhance the parallel multi-scale attention features, emphasizing local feature extraction. Finally, to encourage interpretability in the feature space and avoid direct estimation of unknown elements, the paper introduces a physical model that simulates the propagation of light through haze, adjusting and optimizing features.

### Multichannel Feature Extraction

Traditional image dehazing methods often employ convolutional structures to extract local image information. For regions with mild haze, such local information is sufficient to recover the image. However, in areas with heavy haze, local information alone is inadequate for effectively enhancing pixels.

To address the aforementioned issue, this paper dynamically enhances pixels in different regions by acquiring both local and non-local feature information via distinct approaches. For non-local feature extraction, the Dehazeformer module from Dehazeformer-Net[16] is utilized. Local feature acquisition employs a parallel multi-scale attention module, as illustrated in Figure 2. By leveraging parallel dilated convolutions to extract multi-scale features and restore texture detail features, and applying channel and pixel attention to extract shared global information and location-specific local information from the original features.

By utilizing convolutional kernels of varying sizes, multi-scale perception of hazy images is achieved. An input image  $I(x)$  is transformed into shallow features  $x_1$  by a convolution module. we normalize it using BatchNorm via  $B(x_1) = \text{BatchNorm}(x_1)$ . BatchNorm can accelerate network convergence, improve generalization ability and prevent overfitting.

$$\begin{aligned} x_2 &= \text{Conv5}(\text{PWConv}(B(x_1))), \\ x_3 &= \text{Concat} \left( \begin{array}{l} \text{DWDCConv19}(x_2), \\ \text{DWDCConv13}(x_2), \\ \text{DWDCConv7}(x_2) \end{array} \right), \end{aligned} \quad (3)$$

Where PWConv denotes point-wise convolution, DWDCConv19 means which dilated convolution kernel size = 19 and it is  $7 \times 7$  depth-wise dilated convolution with dilation rate 3, DWDCConv13 means which dilated convolution kernel size = 13 and it is  $5 \times 5$  depth-wise dilated convolution with dilation rate 3, DWDCConv7 means which dilated convolution kernel size = 7 and it is  $3 \times 3$  depth-wise dilated convolution

with dilation rate 3. Finally, concat means concatenate feature

in the channel dimension.

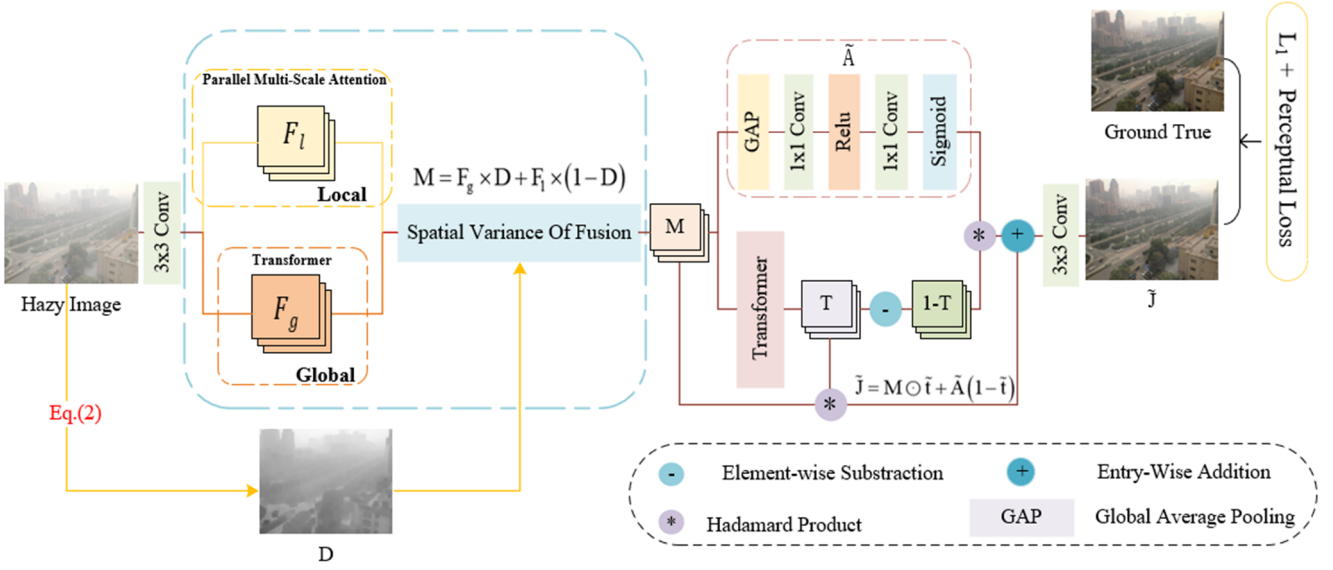


Figure 1. Overall Framework of the SVFPA-Net

$$x_4 = x_1 + \text{PWConv}(\text{RELU}(\text{PWConv}(x_3))), \quad (4)$$

Then we feed  $x_3$  into a multi-layer perceptron that converts the feature dimension of  $x_3$  to be the same as  $I(x)$ . The multi-layer perceptron contains two point-wise convolutions and uses RELU as an activate function. Finally, the output of the multi-layer perceptron is summed with the identity shortcut  $x_1$ . We believe that the multi-layer perceptron can not only combine three different types of features but also play a role in fitting the dehazing features.

While the module contains channel attention and pixel attention. Pixel attention can effectively extract location-dependent informative features, such as different haze distributions.

$$\text{PA}(x_4) = \text{Sigmoid}(\text{PWConv}(\text{RELU}(\text{PWConv}(\text{PWConv}(x_4))))) \quad (5)$$

$$F_p = x_4 \otimes \text{PA}(x_4).$$

Here, we use PWConv-RELU-PWConv to adapt features. The Sigmoid function is utilized to extract global pixel gating features.  $\text{PA}(x)$  serves as a global pixel gating signal for  $x_4$ .

channel attention focuses on inter-channel relationships,

autonomously learning the importance of features across various channels.

$$\text{CA}(x_4) = \text{Sigmoid}(\text{PWConv}(\text{RELU}(\text{PWConv}(\text{GAP}(x_4))))) \quad (6)$$

$$F_c = x_4 \otimes \text{CA}(x_4).$$

We use the global average pooling (GAP), PWConv-RELU-PWConv, and Sigmoid function to extract global channel gating features. And then we use  $\text{CA}(x)$  as a global channel gating signal for  $x_4$ .

We concatenate the diverse attention outcomes along the channel dimension, followed by PWConv-RELU-PWConv to reduce the concatenated feature's channel dimension to match that of the input  $x_4$ . Finally, we perform a residual connection between the PWConv-RELU-PWConv output and the input. This approach extracts effective information features from hazy images by blending cross-channel and spatial information, facilitating the flow of these effective features through the network.

### 3.1. Spatial Variance of Fusion

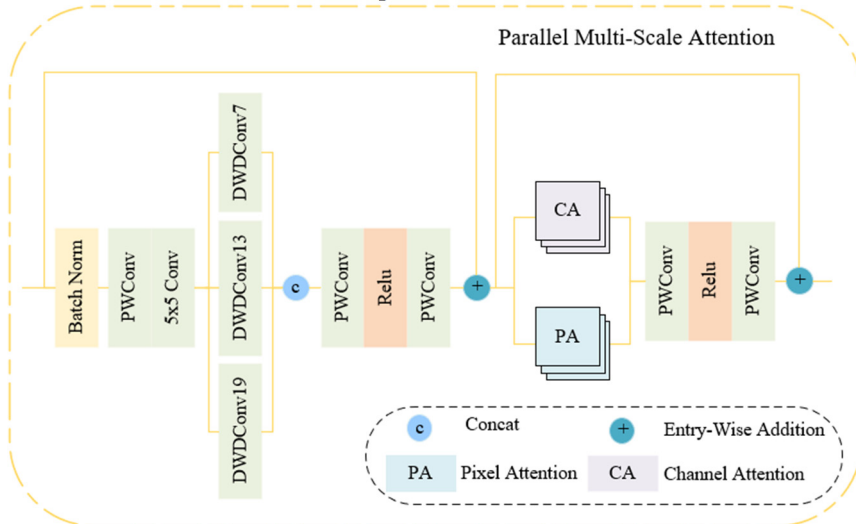


Figure 2. Parallel Multi-Scale Attention

In this paper we transform the input image into a dark

channel prior map, and utilize it as an effective prior for our

spatially-variant feature fusion. As shown in Figure 3, areas with high dark channel values are typically unclear. Therefore, we utilize non-local image information over a larger spatial range for image enhancement. On the other hand, areas with relatively low dark channel values usually have higher visibility and less noise. Therefore, local image information is generally sufficient.

As illustrated in Figure 1, we utilize parallel multi-scale attention to extract local features  $F_l$  from the input image  $I$ , and extract global dependencies  $F_g$  through the Transformer. To adaptively combine these two features, we normalize the dark channel map values to the range of  $[0,1]$ , and use the normalized DCP map  $D$  as the interpolation weight to fuse  $F_l$  and  $F_g$  as follows:

$$M = F_g \times D + F_l \times (1 - D) \quad (7)$$

Since the values in the dark channel map dynamically reveal the density of hazy in different areas of the input image, the fusion can adaptively combine the local and non-local image information to generate  $M$ .

### 3.2. Physics Aware

Traditional methods often estimate unknown parameters directly in the original space, which can introduce errors. To address these issues, this paper proposes a physical model based on the feature space to describe the hazy process. By estimating under the constraints of the physical model, it is possible to effectively obtain more useful feature information for reconstructing a clear image.

To begin, we construct the clear image  $J$  based on the atmospheric scattering model, as shown in Equation 8:

$$\begin{aligned} J(x) &= I(x) \frac{1}{T(x)} + A \left( 1 - \frac{1}{T(x)} \right) \\ &= I(x) \frac{1}{T(x)} + A - A \frac{1}{T(x)} \end{aligned} \quad (8)$$

For simplicity, we extract features via the convolution kernel ( $K$ ). Therefore, Equation 9 can be reformulated as follows:

$$\begin{aligned} K \otimes J &= K \otimes \left( I \odot \frac{1}{T} \right) + \\ &K \otimes A - K \otimes \left( A \odot \frac{1}{T} \right) \end{aligned} \quad (9)$$

Where  $\otimes$  indicates the convolution operator and  $\odot$  denotes the Hadamard product. we can express the algebraic computations using matrix-vector notation as follows:

$$KJ = KDJ + KA - KDA \quad (10)$$

Given that  $K, J, I, A, T$  and  $D$  represent the matrix vector forms of the variables  $K, J, I, A, T$  and  $T/1$  in Equation 7. we can decompose the matrix  $KD$  into a product of two matrices  $QK$ . As  $K, D$  and  $Q$  are all unknown, implementing this decomposition can be indicated as solving an underdetermined system of equations, which can guarantee the existence of  $Q$ . And then, we have

$$KJ = Q(KI) + KA - Q(KA) \quad (11)$$

we denote the approximations associated with the atmospheric light feature  $KA$  by  $\tilde{A}$ , and the approximations associated with the transmission map  $Q$  by  $\tilde{t}$ . Furthermore,  $KI$  and  $KJ$  can be viewed as the extracted features of a hazy image and its corresponding clear image, respectively. we can

calculate the physics-aware features  $\tilde{J}$  by

$$\tilde{J} = M \odot \tilde{t} + \tilde{A}(1 - \tilde{t}) \quad (12)$$

According to Equation (11), we can explicitly construct a dual-branch scattering-aware unit. One branch within the dual-branch scattering-aware unit is dedicated to generating an estimate of the atmospheric light. Given that atmospheric light is typically considered to be uniform, we can employ global average pooling to eliminate unnecessary information in the feature space. The atmospheric light feature estimate can be obtained through the following equation:

$$\tilde{A} = \text{Sigmoid}(\text{Conv}^N(\text{ReLU}(\text{Conv}^N/8(\text{GAP}(M)))))) \quad (13)$$

Where  $\text{Conv}^N(\cdot)$  is the convolutional layer with  $N$  kernels, and  $N$  is set to 64.

$T$  of the atmospheric scattering model represents global features of the image, and the self-attention mechanism of the Transformer model enables it to accurately capture the relationships of different positions within the input sequence, and possesses the ability to handle long-range dependencies. This gives the Transformer model a significant advantage in extracting the global feature  $T$  of the atmospheric scattering model. Therefore, in the lower branch of this paper, we use Dahazeformer to extract  $T$ .

### 3.3. Training Loss

Given the image pair  $I, J$  where  $I$  is hazy image and  $J$  is the corresponding clear image, we let our network predict the dehazing image  $\hat{J}$ . And we use L1 loss and Perceptual Loss to train our model, which can be formulated as:

$$L = L_1 + \lambda L_{\text{vgg}} \quad (14)$$

where  $\lambda$  is hyper parameter to balance L1 loss and Perceptual loss.

## 4. Experiments

In this section, we first show the detail of the proposed method. Then, to fully evaluate the proposed method, we test our method on the RESIDE[23], RESIDE-6K datasets and RS-Haze [47] datasets. The RESIDE datasets contains of RESIDE-IN(ITS), RESIDE-OUT(OTS) and Synthetic Objective Testing task (SOTS). Specifically, we use the ITS (13,990 image pairs) and OTS (313,950 image pairs) to train the models and test them on the indoor set (500 image pairs) and the outdoor set (500 image pairs) of the SOTS, respectively. The RESIDE-6K dataset contains a mix of synthesis images of indoor and outdoor scenes from ITS and OTS. For the RS-Haze dataset, we follow the setup of DehazeFormer [47]. It consists of 54,000 image pairs, of which 51,300 are used for training and the remaining 2,700 for testing. RS-Haze is a remote sensing dehazing image dataset with a more monotonous scene, but its haze is highly non-homogeneous, the exact opposite of RESIDE and RESIDE-6K. All models are trained using their original strategies, and we replicate the best results reported in the previous works.

### Implementation details

We used GeForce RTX 2080Ti GPUs to train our models. During training, images are randomly cropped to  $256 \times 256$  patches. And we set the hyper parameter  $\lambda$  to 0.1. We used the AdamW optimizer to optimize our network with exponential decay rates  $\beta_1$  and  $\beta_2$  equals to 0.9 and 0.999, respectively. We set the initial learning rate to  $2 \times 10^{-4}$ , which gradually decreases from the initial rate to  $2 \times 10^{-6}$  with the

cosine annealing strategy.

### 4.1. Comparison with State-of-the-art Methods

**Quantitative Analysis:** We compared the performance of the proposed method with previous state-of-the-art methods, and the results are shown in Table 1. As demonstrated, in the RESIDE-OUT and RESIDE-IN datasets, our proposed method achieves the best performance on both datasets

compared to other SOTAs, with 37.97dB PSNR and 0.993 SSIM in SOTS-indoor, and 36.61dB PSNR and 0.989 SSIM in SOTS-outdoor. Specifically, our method outperforms the second-best method AECR-Net[21] by a significant margin on SOTS-indoor, 0.6dB PSNR and 0.3 SSIM. Moreover, our method achieves at least 1.43dB PSNR and 0.003 SSIM performance gains on SOTS-outdoor.

**Table 1.** Quantitative comparison of various SOTA methods on three dehazing datasets.

Methods	ITS		OTS		RESIDE-6K	
	SOTS-indoor		SOTS-outdoor		SOTS-mix	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
(TPAMI'10) DCP [17]	16.62	0.818	19.13	0.815	17.88	0.816
(ICCV'17) AOD-Net [18]	20.51	0.816	24.14	0.920	20.27	0.855
(AAAI'20) FFA-Net [19]	36.39	0.989	33.57	0.984	29.96	0.973
(CVPR'21) AECR-Net [21]	37.17	0.990	-	-	28.52	0.964
(CVPR'22) Dehamer [22]	36.36	0.988	34.93	0.987	-	-
Ours	37.97	0.993	36.61	0.989	29.94	0.978

**Qualitative Analysis:** Figure 3 shows the visual results of our method compared to previous state-of-the-art models on RESIDE-6K dataset. Since the prior knowledge is not satisfied, the images recovered by DCP[6] have color distortion. Both AOD-Net[18] and FFA-Net[19] have more haze residue and the distribution of residual haze in the restored image is uneven. while FFA-Net and DeHamer suffer from severe color distortion. Instead, our method generates the most natural restoration that preserves more details and

involves fewer color distortions.

### 4.2. Ablation Study

In this section, we analyze the effectiveness of the different components of the proposed method including non-local extraction of Transformer, Spatial Variance of Fusion and Physics Aware Module. We list the results in Table 2, using the RESIDE-6K dataset.

**Table 2.** Ablation study with different modules and regularizations on RESIDE-6K dataset.

Transformer	Spatial Variance of Fusion	Physics Aware	PSNR	SSIM
×	√	√	27.8	0.963
√	×	√	29.03	0.97
√	√	×	27.28	0.958
√	√	√	29.94	0.978

**Effectiveness of Physics Aware.** The architecture of physics aware is derived from Eq. (1) with a consideration of the physical characteristics of A and T, which introduces a dual-branch interaction for the prediction of both factors. Since the features corresponding to A and T are disentangled by Atmospheric Scattering, the latent structural feature-level information is excavated more accurately. As a result, in Table 2 It can be observed that Atmospheric Scattering remarkably boosts the performance, PSNR improvements of 2.66dB.

**Effectiveness of Spatial Variance of Fusion.** Leveraging the dark channel map as a guiding mechanism in feature fusion, our network is capable of adaptively adjusting the weights of features based on the varying haze density across different regions. This adaptive approach allows for a more effective utilization of the complementary nature of both local and global features. The guiding mechanism enables us to precisely focus our attention on diverse areas, adjusting the contribution of features in response to the changing haze density. As a result, in Table 2 we can see that Dark Channel-Guided Fusion achieves an increase of 0.91dB in PSNR.

## 5. Conclusion

In this paper, we propose a dehazing network that integrates parallel multi-scale attention with a physical model, jointly leveraging both local and non-local information to

dynamically enhance pixels in a spatially variant manner. The dark channel prior is employed to guide feature fusion. To enhance the interpretability of the feature space, we further design a physics aware dual-branch unit based on the physics model. The features produced by the unit are enforced to conform with the hazing process, thus facilitating haze removal. Extensive experiments demonstrate the validity and generality of the proposed method.

## References

- [1] Wei-Ting Chen, I-Hsiang Chen, Chih-Yuan Yeh, Hao Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Sjd1-vehicle: Semi-supervised joint defogging learning for foggy vehicle re-identification. In AAAI, 2022.
- [2] Christos Sakaridis, Dengxin Dai, Simon Hecker, and Luc Van Gool. Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In ECCV, pages 687–704, 2018.
- [3] McCartney, E.J.: Scattering phenomena (book reviews: optics of the atmosphere. scattering by molecules and particles). Science 196, 1084–1085 (1977).
- [4] Berman, Dana, T. Treibitz, and S. Avidan. "Non-local Image Dehazing." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2016.

- [5] Raanan, and Fattal. "Dehazing Using Color-Lines." *ACM Transactions on Graphics (TOG)* 34.1(2014).
- [6] He, Kaiming, et al. "Single Image Haze Removal Using Dark Channel Prior." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 33.12(2011):2341-2353.
- [7] Zhu, Qingsong, J. Mai, and L. Shao. "A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior." *IEEE Transactions on Image Processing* 24.11(2015):3522-3533.
- [8] Li B, Peng X, Wang Z, et al. AOD-Net: All-in-One Dehazing Network [C]//2017 IEEE International Conference on Computer Vision (ICCV).IEEE, 2017.DOI: 10.1109/ ICCV. 2017. 511.
- [9] Zhang H, Patel V M .Densely Connected Pyramid Dehazing Network[J].IEEE, 2018.DOI:10.1109/CVPR.2018.00337.
- [10] Qin X, Wang Z, Bai Y, et al. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing[J]. 2019.DOI: 10.48550/arXiv.1911.07559.
- [11] Dong H, Pan J, Xiang L, et al. Multi-Scale Boosted Dehazing Network with Dense Feature Fusion[J].arXiv, 2020.DOI: 10.1109/CVPR42600.2020.00223.
- [12] Wu H, Qu Y, Lin S, et al. Contrastive Learning for Compact Single Image Dehazing[J]. 2021.DOI:10.48550/ arXiv. 2104.09367.
- [13] Song Y, He Z, Qian H, et al. Vision Transformers for Single Image Dehazing[J].arXiv e-prints, 2022.DOI:10.48550/ arXiv. 2204.03883.
- [14] GUO C L, YAN Q X, ANWAR S, et al. Image dehazing transformer with transmission-aware 3d position embedding [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Jun 18-24, 2022. Washington: IEEE Computer Society, 2022: 5802–5810.
- [15] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [16] Song Y, He Z, Qian H, et al. Vision Transformers for Single Image Dehazing[J].arXiv e-prints, 2022.DOI: 10.48550/ arXiv. 2204.03883.
- [17] He K, Sun J, Tang X. Single Image Haze Removal Using Dark Channel Prior[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011.DOI:10.1109/ CVPRW. 2009. 5206515.
- [18] Li B, Peng X, Wang Z, et al. AOD-Net: All-in-One Dehazing Network[C]//2017 IEEE International Conference on Computer Vision (ICCV).IEEE, 2017.DOI:10.1109/ ICCV. 2017. 511.
- [19] Qin X, Wang Z, Bai Y, et al. FFA-Net: Feature Fusion Attention Network for Single Image Dehazing[J]. 2019. DOI: 10.48550/arXiv.1911.07559.
- [20] Dong H, Pan J, Xiang L, et al. Multi-Scale Boosted Dehazing Network with Dense Feature Fusion[J].arXiv, 2020.DOI: 10.1109/CVPR42600.2020.00223.
- [21] Wu H, Qu Y, Lin S, et al. Contrastive Learning for Compact Single Image Dehazing[J]. 2021.DOI:10.48550/ arXiv. 2104.09367.
- [22] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5812–5820, 2022. 2, 3, 6.
- [23] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Bench-marking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2018.
- [24] Liu X, Ma Y, Shi Z, et al. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV).IEEE, 2020.DOI:10.1109/ICCV.2019.00741.
- [25] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[J]. 2021.DOI: 10.48550/ arXiv.2103.14030.
- [26] Liang J, Cao J, Sun G, et al. SwinIR: Image Restoration Using Swin Transformer [J]. IEEE, 2021.DOI:10.1109/ ICCVW 54120. 2021.00210.
- [27] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. In CVPR, 2022.