Knowledge-Based Systems 266 (2023) 110410

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Physical-priors-guided DehazeFormer

Hao Zhou [a,b,c], Zekai Chen [c], Yun Liu [d], Yongpan Sheng [a], Wenqi Ren [e], Hailing Xiong [f,*]

[a] College of Computer and Information Science, Southwest University, Chongqing 400715, China
[b] School of Big Data and Artificial Intelligence, Chizhou University, Chizhou 247000, China
[c] School of Big Data and Intelligent Engineering, Chongqing College of International Business and Economics, Chongqing, China
[d] College of Artificial Intelligence, Southwest University, Chongqing 400715, China
[e] School of Cyber Science and Technology, Sun Yat-sen University, ShenZhen 510006, China
[f] College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China

ARTICLE INFO

Article history:
Received 14 November 2022
Received in revised form 14 February 2023
Accepted 15 February 2023
Available online 17 February 2023

Keywords:
Image dehazing
Dark channel prior
Swin Transformer
Deep learning
Self-supervised learning

ABSTRACT

Single-image dehazing is a challenging task in several machine-vision applications. Methods based on physical models and prior knowledge fail under certain conditions, resulting in defects such as color distortion. Transformer-based methods have a strong representation ability owing to their self-attention mechanism that can effectively obtain global information. However, this approach is computationally expensive, and its weak inductive bias capability increases the risk of overfitting on small-sample datasets. To address these problems, in this study, we propose a novel DehazeFormer guided by physical priors, named SwinTD-Net, which is trained according to supervised and self-supervised learning, and combines the advantages of physical priors and transformers. The proposed DehazeFormer learns features guided by physical priors, which improves the generalization ability of the network and enables it to achieve good restoration effects on both synthetic and real-world hazy images. In addition, we propose a more appropriate prior input to better use physical priors, and we design a multi-scale dark-light enhancement algorithm for image restoration post-processing, which can improve the visual perception quality for human observers while performing some local enhancements. Extensive experiments illustrate that the proposed method outperforms state-of-the-art methods. The code and pre-trained models are available to academics so that they can reproduce our results and test them (https://github.com/hocking-cloud/SwinTD_Net).

1. Introduction

Images collected on hazy days suffer from reduced contrast and color degradation, which seriously affect the functionality of outdoor machine vision systems, such as object detection and video surveillance. Therefore, image dehazing has been extensively researched in the past decade [1–8].

Existing methods can be divided into image-enhancement-based, prior-based, and learning-based approaches. The first method does not consider the degradation mechanism and only processes the image from the spatial or frequency domain or performs color correction. Common examples include histogram equalization [9], Retinex algorithm [10], and homomorphic filtering [11]. These methods may lead to color distortion, incomplete dehazing, and other problems.

Prior-based methods combine prior knowledge regarding the image and transform the task into a process of estimating the model parameters. The atmospheric scattering model (ASM) [12–15] is used to represent the imaging process in hazy weather, as follows:

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

where $\mathbf{x}$ represents the position of a pixel and $J(\mathbf{x})$ denotes hazy-free scene radiation. $I(\mathbf{x})$ denotes the hazy image, A is the global atmospheric light, $t(\mathbf{x})$ is the transmission map, which is defined as follows:

$$t(\mathbf{x}) = e^{-\beta d(\mathbf{x})} \tag{2}$$

where β is the atmospheric scattering coefficient and $d(\mathbf{x})$ represents the distance from the scene to the imaging system. Because only $I(\mathbf{x})$ is known in Eq. (1), $J(\mathbf{x})$ cannot be directly calculated. Therefore, several methods based on prior knowledge have been proposed to estimate the unknown parameters in Eq. (1) to recover $J(\mathbf{x})$, for example, dark channel prior (DCP) [2], non-local prior [16], and color-line prior [17]. Dynamic stochastic resonance has been widely used for image dehazing [18,19]. This method uses a standard algorithm with satisfactory results. These methods use the statistical features of the images to traverse $t(\mathbf{x})$. These methods reflect the underlying characteristics of images

[*] Corresponding author.
E-mail address: xionghl@swu.edu.cn (H.L. Xiong).

https://doi.org/10.1016/j.knosys.2023.110410
0950-7051/© 2023 Elsevier B.V. All rights reserved.

and are highly effective. However, these priors are not fully consistent with hazy environments, leading to inaccurate parameter estimates, and resulting in unwanted artifacts, such as halos and color blockings.

Deep-learning-based methods restore haze-free images by training neural networks on large-scale datasets. These methods either estimate the parameters A and $t(\mathbf{x})$ or directly obtain $J(\mathbf{x})$, e.g., DehazeNet [4], AodNet [5], MSCNN [20], MSBDNNet [21], Griddehazenet [22], EAA-Net [23], DBDN-SCAC [24], LKD-Net [25], PM-Net [26], DID-Net [27], and PA-DC-MSPCNN [28]. These methods can directly learn the mapping from $I(\mathbf{x})$ to $J(\mathbf{x})$, which ensures the realness of image restoration and reduces artifacts and color distortion. Most deep-learning-based dehazing models adopt supervised training methods that require synthetic hazy image datasets. However, neural network models trained with synthetic datasets are prone to overfitting, and the subsequent restoration effect on real-world hazy images is not ideal.

Prior knowledge-based methods are relatively good at restoring image visibility but are prone to creating unwanted artifacts, whereas deep learning-based methods are relatively good at restoring image realness but are prone to overfitting. Therefore, we propose a novel single-image dehazing framework that combines prior knowledge and deep learning frameworks. We combine the advantages of the two methods based on DCP and deep learning to further improve the dehazing quality. First, we pre-train the dehazing network via supervised learning and then fine-tune the network via self-supervised learning guided by physical priors to achieve excellent dehazing performance on both synthetic and real-world hazy images. In addition, we propose a more appropriate prior input for using ASM and DCP more effectively; that is, we believe that ASM and DCP are more suitable for RAW images in the image signal processing (ISP) pipeline [29,30] rather than red-green-blue (RGB) space images. In the ISP pipeline, gamma correction changes the relationship between the light intensity of the original scene radiation in the haze image through nonlinear transformation, affecting the neural network's extraction capability of haze features. Therefore, the RGB space image obtained through the ISP pipeline is not conducive to the image dehazing task, thus, we restore a RGB image to a RAW image for the dehazing operation.

There exist studies that aim to combine physical priors with deep learning for dehazing tasks, such as RefineDNet [31], and PSD [32]. The dehazing effect of RefineDNet is worthy of recognition but the obtained restored image is prone to noise artifacts. The images recovered by PSD are prone to color distortion and remaining haze, whereas our method achieves an excellent dehazing effect and can effectively reduce the number of artifacts. The proposed dehazing framework is divided into two stages. In stage 1, we estimate $t(\mathbf{x})$ and $J(\mathbf{x})$ separately through supervised training using two parallel branches based on the residual Swin Transformer blocks [33]. In stage 2, we fine-tune the network in a self-supervised training manner. The second stage combines ASM and DCP to constrain the network feature learning with prior knowledge, which enables the network to be trained under the physical mechanism and improves the dehazing performance for real-world scenarios. In stage 2, A is obtained by DCP, and the network reconstructs the hazy image through ASM using $t(\mathbf{x})$ and $J(\mathbf{x})$ obtained in stage 1 and conducts self-supervised training using the hazy input as the label. The main structure of the proposed dehazing framework is inspired by the SwinIR block [34], which consists of a residual module [35] and Swin Transformer layers (STL) [33]. In addition, we simplified the structure of SwinIR and reduced the number of residual Swin Transformer blocks, which significantly improved the inference speed of the model. We refer to the proposed dehazing network as the SwinTD-Net (Swin Transformer and DCP Network). Through a two-stage strategy, SwinTD-Net divides image dehazing into two tasks: the transmission map and clear images are estimated by supervised learning, and real-world dehazing performance improvement is based on self-supervised learning with physical priors.

Therefore, the proposed SwinTD-Net combines the advantages of prior-and deep learning-based methods. In addition, to enhance the details of the restored images, particularly for the dark areas of the image, we design a multi-scale dark-light enhancement algorithm for post-processing operations.

The contributions of this study can be summarized as follows:

- In this study, we propose a novel two-stage single-image dehazing framework. The first stage uses two parallel Swin Transformer-based branches to estimate the transmission map and clear image. The second stage fine-tunes the network by reconstructing hazy images guided by physical priors, enabling the network to learn physical mechanisms and reduce overfitting, thereby improving its dehazing performance in real-world applications. The first and second stages use supervised and self-supervised learning, respectively, for training.
- We propose a more appropriate prior input for better use of the ASM and DCP. Gamma correction in the ISP pipeline causes the RGB image to change the linear relationship of the light intensity in the scene radiance. Therefore, we convert a RGB image to a RAW image using inverse gamma correction to restore the linear relationship of light intensity to better fit the ASM and DCP.
- We design a multi-scale dark-light enhancement algorithm (MSDLEA) for image dehazing post-processing, which can improve the quality of images from the perspective of human visual perception, especially by improving the brightness of dark areas and highlighting their details. This is a general image enhancement algorithm.

2. Related research

We briefly reviewed related research, namely, prior-based methods, deep learning-based methods, and Vision Transformer.

2.1. Prior-based methods

Because $J(\mathbf{x})$ stated in Eq. (1) cannot be solved directly; many scholars have proposed methods based on prior knowledge. Fattal [1] redefined the atmospheric transport model and increased the surface shadow variable, and based on the premise that the surface shadow and the transmission map are statistically irrelevant. Then, ASM was calculated and analyzed to achieve dehazing. Based on the stability of atmospheric light in a local area in the atmosphere, Tan [36] assumed that A is a constant for constructing the edge intensity cost function and obtain the optimal solution. He et al. [2] proposed a DCP to reckon the initial $t(\mathbf{x})$, and achieved very effective results. DCP is a well-known prior knowledge framework; accordingly, many dehazing methods are based on DCP for improvement. Tarel and Hautiere [37] defined an atmospheric veil, which tends to infinity in its local pixel range and can be approximated as a constant. Moreover, the atmospheric veil can be estimated using median filtering to restore a clear image. Meng et al. [3] added boundary constraints based on DCP and used weighted variational regularization to refine $t(\mathbf{x})$. Zhu et al. [38] proposed a color attenuation prior that learns a linear function to predict the depth of hazy images. Berman et al. [16] proposed a non-local prior that uses the K-means clustering algorithm to form different clusters of RGB values of a haze-free image. Berman et al. found that hazy lines were formed under the influence of haze, and then $t(\mathbf{x})$

of each pixel was estimated based on the hazy lines. Regarding the failure of the DCP in the sky area, Sahu et al. [39] proposed a new color model to calculate $t(\mathbf{x})$ and $J(\mathbf{x})$. Although prior-based dehazing methods are effective, the performance of these methods is limited because hand-crafted priors are insufficient for all cases and may result in unwanted artifacts.

2.2. Deep learning-based methods

Deep learning-based methods [40] are used in various application scenarios, such as natural language processing (NLP), recommender systems (RS), computer vision (CV), etc [41–52]. Several scholars have proposed methods based on deep learning for image dehazing. Most of these methods use convolutional neural networks (CNN) to construct a dehazing network. Cai et al. [4] proposed DehazeNet, which estimates the mapping relationship between a hazy image and transmission map by learning the features, and then obtains $t(\mathbf{x})$ according to the features of the hazy input image. The main contribution of DehazeNet is the proposal of an end-to-end dehazing network for the first time. Ren et al. [20] proposed a multi-scale deep neural network model (Multi-Scale CNN, MSCNN) that uses a coarse-scale network to estimate the overall $t(\mathbf{x})$, and then uses a fine-scale network for local refinement. Li et al. [5] combined $t(\mathbf{x})$ and A in the ASM into one parameter and then combined it with a CNN to build a lightweight network (All-in-One Network, AOD-Net) that is fast and efficient. AOD-Net can also be embedded in other networks for advanced computer vision tasks. Qin et al. [53] proposed a feature fusion attention network (FFA-Net), which added a feature attention mechanism so that more attention was paid to useful information in hazy images and representation ability of the CNN was improved. Zhang et al. [54] proposed an end-to-end deep residual convolutional dehazing network, which first restored the image and then refined it. Wu et al. [55] proposed a regularization method based on contrastive learning, taking clear and hazy images as positive and negative samples, respectively, and training a dehazing network. Zhao et al. proposed RefineDNet [31], which added a DCP to the network. Although the effect of this method was satisfactory, it produced noise artifacts. Chen et al. proposed PSD [32], combined DCP, bright channel prior (BCP), and histogram equalization in a network to establish a loss function committee to guide network learning. This idea is worthy of affirmation; however, the experimental results show that the images restored by this method exhibit an incomplete dehazing effect and color distortion. Sahu et al. [56] studied a dual-channel deep neural network (DCD-Net) to estimate $t(\mathbf{x})$ and calculate A to restore $J(\mathbf{x})$, combined with Internet of Things (IoT) applications in real-time systems, such as transportation and monitoring industries. Some scholars have also proposed methods that use unpaired hazy and clear image datasets to train their own designed neural network architectures to achieve dehazing, such as CycleDehaze [57], DisentGAN [58], YOLY [59], and D^4 [60]. Deep learning-based methods perform well on synthetic hazy images but their real-world performance is not entirely satisfactory.

2.3. Vision transformer

Recently, Transformers [61] have attracted the attention of many scholars; accordingly, many studies have been conducted based on Transformers, including NLP and CV, such as Vision Transformer (ViT) [62]. Many new studies based on ViT have been conducted for tasks such as object detection and semantic segmentation. Wang et al. proposed a Pyramid Vision Transformer (PVT) [63] which introduces a pyramid structure into the ViT, making it accessible to various downstream tasks while achieving outstanding results. Liu et al. proposed a Swin Transformer [33] that incorporates an attention mechanism from local to global and greatly reduces the computational complexity. Carion et al. proposed a Detection Transformer (DETR) [64] that replaced the traditional complex object detection step with a Transformer. The DETR simplifies the detection process and effectively eliminates many hand-designed components. Tang et al. [65] proposed a dynamic Transformer-based infrared and visible-light image fusion method. ViT has also been widely used in the medical field. For example, Zhu et al. [66] proposed to use a Swin Transformer to extract semantic features, which improved the accuracy of brain tumor segmentation. In the field of low-level vision, Chen et al. proposed an image processing Transformer (IPT) [67] to allow ViT to specifically solve low-level vision tasks. Liang et al. [34] proposed the so-called SwinIR for image restoration. SwinIR mainly includes shallow and deep feature extraction, and high-quality image reconstruction, which greatly reduces the number of network parameters while improving performance. For image dehazing, a small number of scholars have also proposed Transformer-based dehazing networks. HyLoG-ViT [68] learns image decomposition and image dehazing through a hybrid local–global Transformer structure. Transweather [69] is an end-to-end network based on the Transformer framework and can be used for image restoration, including image dehazing under various weather conditions. Song et al. [70] improved several key designs of the Swin Transformer, such as the normalization layer and activation function. Aiming to address the feature inconsistency between the CNN and Transformer, DeHamer [71] learns a modulation matrix to modulate the features of the CNN and implements image dehazing using a transmission-aware 3D position embedding module. However, these methods rarely consider the incorporation of prior knowledge, resulting in trained networks that rely heavily on synthetic hazy image datasets and are prone to overfitting. When encountering new synthetic hazy image datasets or real-world hazy images, these trained networks may not perform well owing to differences in the haze distribution, transmission map distribution, and other characteristics of the images. Therefore, our method incorporates prior knowledge into the network, allowing the network to learn features under the guidance of physical priors to improve generalization performance.

3. The proposed method

First, we describe the so-called more appropriate prior input. Then, we introduce the overall architecture and loss function of the proposed SwinTD-Net. Finally, we introduce the designed MSDLEA algorithm.

3.1. Prior input

In the past decade years, single-image dehazing algorithms based on ASM and DCP have mostly operated on RGB images. However, the images collected by electronic devices such as digital cameras are converted from the original RAW domain images to RGB domain images after a series of processing through the ISP pipeline [29,30]. The gamma correction in the ISP pipeline is a nonlinear operation that changes the linear relationship of the light intensity of the scene radiance, as shown in Fig. 1. This affects the use of physical priors and also the extraction of haze features via neural networks. Therefore, we restored the RGB image to the RAW image before gamma correction as the input for the ASM and DCP, which is more in line with the mechanisms of atmospheric scattering [14]. Numerically, a RAW domain image

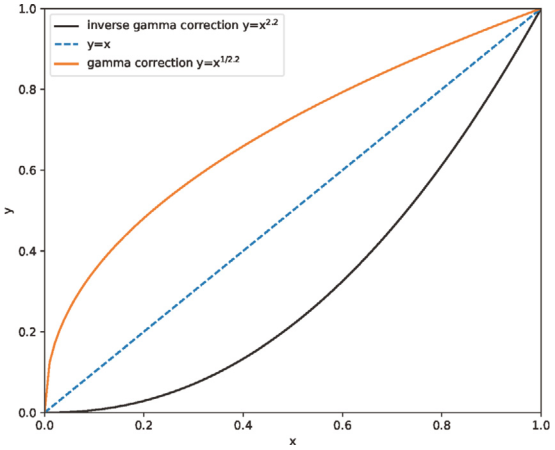


Fig. 1. The Curve Diagram of Gamma Correction and Inverse Gamma Correction Function. The abscissa x represents the normalized pixel value of one input image pixel, and ordinate y represents the corresponding pixel value of the output image. The black curve represents the inverse gamma correction function, while the orange curve represents the gamma correction function.

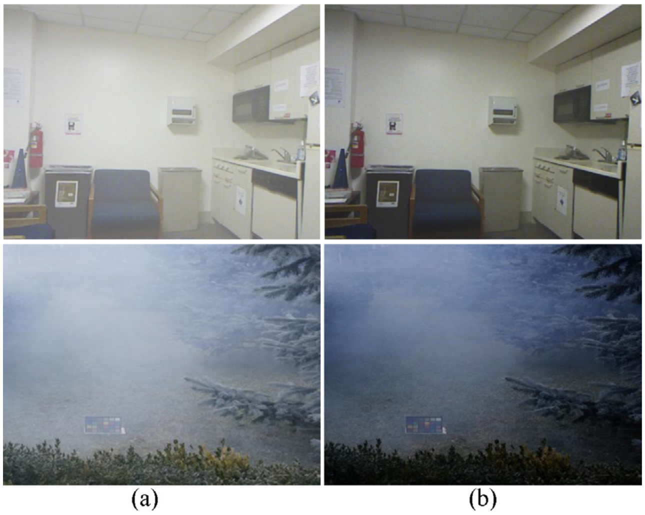


Fig. 2. (a) Hazy image, (b) Hazy image after inverse gamma correction.

can be restored by performing an inverse gamma correction that is expressed as follows:

$$I_{RAW}(\mathbf{x}) = I(\mathbf{x})^{\frac{1}{\gamma}} \tag{3}$$

where $I(\mathbf{x})$ represents the RGB space image collected by electronic equipment in a hazy environment and γ is the gamma correction value in the ISP pipeline, whose value is 1/2.2 [72], which is used in all our experiments. $I_{RAW}(\mathbf{x})$ represents the RAW image, which restores the linear relationship of the light intensity of the scene radiance, as shown in Fig. 2. In the proposed SwinTD-Net, we first preprocessed RGB images to RAW images; subsequently, the resulting clear scene radiance was converted to RGB images by gamma correction.

Although this is a simple modification, we experimentally determined that the dehazing effect was significantly improved. According to atmospheric scattering theory [14], we believe that ASM and DCP are more suitable and efficient in the RAW domain. In addition, from a visual point of view, the overall pixel value of the image is high owing to the existence of haze, which may reduce the attention of the neural network to haze. Our proposed inverse gamma transform can reduce the overall brightness of the image and increase the focus on haze. Furthermore, artifacts often appear in image dehazing. The final gamma correction of our method amplifies artifacts, which also causes the neural network to pay more attention to and eliminate artifacts.

3.2. Overall architecture of the SwinTD-Net

The dehazing task was divided into two stages, as shown in Fig. 3, that is, $t(\mathbf{x})$ and clear images are estimated by supervised learning, while real-world dehazing performance improvement is obtained based on self-supervised learning. Methods based on prior knowledge can achieve better dehazing effects but introduce some artifacts. Methods based on deep learning can produce visually better results but may not dehaze completely and are prone to overfitting, resulting in poor generalization. Therefore, we combined the advantages of both methods.

In **Stage 1**, a hazy RGB image is converted to a RAW image $I_{RAW}(\mathbf{x})$ using Eq. (3). The transmission map $T(\mathbf{x})$ and RAW clear image $J_{RAW}(\mathbf{x})$ depicted in Fig. 3 are then estimated using two parallel simplified SwinIR [34] branches. The Swin Transformer can extract the global features of the image via the self-attention mechanism. Moreover, the sliding window operation can limit the attention calculation within a window, which can introduce the locality of the CNN convolution operation and reduce the amount of calculation. Therefore, the Swin Transformer is adopted as the main component of our network. Similarly, in the dehazing task, the Swin Transformer can extract the global features of hazy images to better restore the clear image and transmission map.

Simplified SwinIR branch. On the one hand, we want to obtain the clear image $J_{RAW}(\mathbf{x})$ directly from the hazy input in an end-to-end manner. On the other hand, according to ASM, a very important parameter in image restoration is $T(\mathbf{x})$, in order to incorporate the physics prior to guiding the network learning, we obtain $T(\mathbf{x})$ through another SwinIR branch, and then reconstruct the hazy image according to ASM. Then, the network is then fine-tuned using self-supervised learning to improve its generalization ability. Therefore, we adopt two parallel SwinIR architectures [34]. To reduce the number of model parameters, we simplify the original SwinIR architecture. The original SwinIR model contains six RSTB modules for image restoration. In our image-dehazing task, we reduced the number of RSTB modules to two while ensuring satisfactory results. In addition, we expanded the number of attention heads in the STL to eight and expanded the dimension of patch embedding to 360. The simplified SwinIR is more suitable for image-dehazing tasks, and provides value for research on lightweight dehazing networks. The two simplified SwinIR branch architectures are shown in Fig. 4. Among them, shallow feature extraction is a 3 × 3 convolutional layer, as well as the Conv behind RSTB. High-quality (HQ) image reconstruction involves aggregating the convolutional shallow and deep features after the Swin Transformer; the reconstruction function also uses a 3 × 3 convolutional layer in our dehazing task. The shallow features mainly contain low-frequency information, whereas the deep features contain high-frequency information, which is finally aggregated through the reconstruction module [34]. The structures of the RSTB and Swin Transformer layer (STL) are shown in Fig. 5. The convolutional layers in RSTB ensure that the Swin Transformer has inductive bias and provide the basis for the final feature aggregation. To improve training stability and convergence, we also added a residual connection to SwinIR branch 2. Under our experimental machine and conditions, we measured the inference speed of the simplified SwinIR and the original SwinIR. Each time we input an image (resize to 128 × 128), we calculated the average value 100 times and repeated the operation ten times. After testing, the original SwinIR obtained 5.9

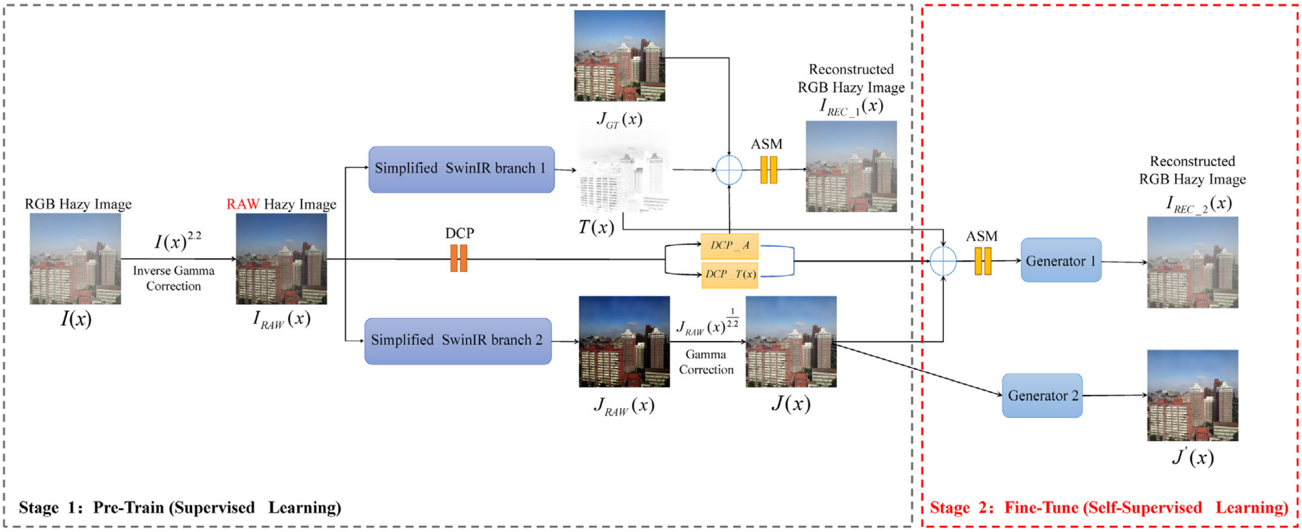
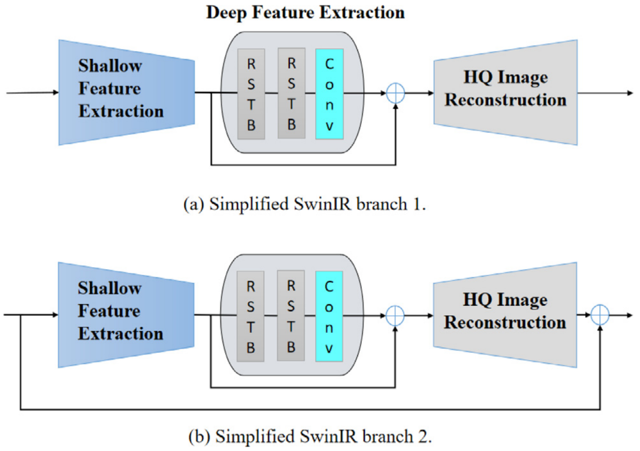

RGB Hazy Image
$I(x)$
$I(x)^{2.2}$
Inverse Gamma Correction
RAW Hazy Image
$I_{RAW}(x)$
Simplified SwinIR branch 1
Simplified SwinIR branch 2
DCP
$J_{GT}(x)$
$T(x)$
ASM
Reconstructed RGB Hazy Image
$I_{REC\_1}(x)$
DCP_A
DCP_T(x)
$J_{RAW}(x)$
$J_{RAW}(x)^{\frac{1}{2.2}}$
Gamma Correction
$J(x)$
ASM
Generator 1
Generator 2
Reconstructed RGB Hazy Image
$I_{REC\_2}(x)$
$J'(x)$
Stage 1: Pre-Train (Supervised Learning)
Stage 2: Fine-Tune (Self-Supervised Learning)

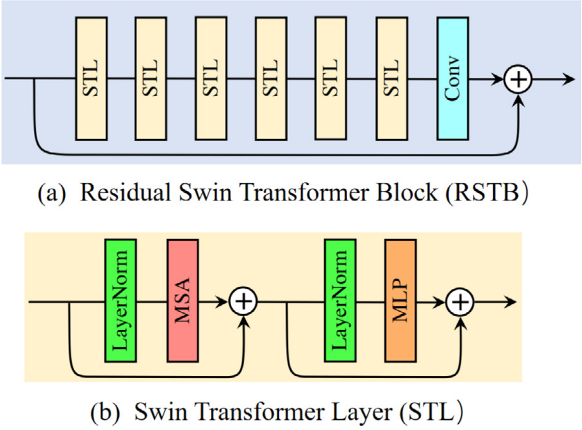
Fig. 3. The overall architecture of the SwinTD-Net. The black dotted box represents Stage 1, that is, pre-training the network through supervised learning. The red dotted box represents Stage 2, that is, fine-tuning the network through self-supervised learning.



(a) Simplified SwinIR branch 1.

(b) Simplified SwinIR branch 2.

Fig. 4. Simplified SwinIR branches 1 and 2. Both branches include shallow feature extraction, deep feature extraction, and image reconstruction modules. Branch 2 has one more residual connection compared to branch 1.



(a) Residual Swin Transformer Block (RSTB)

(b) Swin Transformer Layer (STL)

Fig. 5. RSTB and STL [34]. RSTB consists of six STL modules and a 3 × 3 convolution, plus a residual connection. STL includes LayerNorm, multi-head self-attention (MSA), and multi-layer perceptron (MLP).

FPS, while our simplified SwinIR obtained 7.6 FPS. Accordingly, the inference speed was greatly improved.

Loss function of Stage 1. The SwinIR branch 2 is used to estimate the scene radiation in the RAW domain (*i.e.* $J_{RAW}(\mathbf{x})$ in Fig. 3), and then restore it to the RGB space image $J(\mathbf{x})$ (*i.e.* $J_{RAW}(\mathbf{x})^{\frac{1}{2.2}}$ in Fig. 3) using a gamma correction. We used Charbonnier [73] and SSIM losses [74,75]. The Charbonnier loss function can better handle outliers in network training and improve network robustness [76], while the SSIM loss function can allow the network to improve the image restoration quality from the perspective of human subjective perception. The parameters of SwinIR branch 2 are optimized by minimizing the following loss function:

$$\begin{aligned}\zeta_{SwinIR2} &= Charbonnier\_loss(J_{RAW}(\mathbf{x})^{\gamma}, J_{GT}(\mathbf{x})) \\ &\quad + SSIM\_loss(J_{RAW}(\mathbf{x})^{\gamma}, J_{GT}(\mathbf{x}))\end{aligned} \tag{4}$$

where $J_{GT}(\mathbf{x})$ is the corresponding ground truth (GT). Charbonnier loss can be expressed as:

$$\begin{aligned}&Charbonnier\_loss(J_{RAW}(\mathbf{x})^{\gamma}, J_{GT}(\mathbf{x})) \\ &= \sqrt{\|J_{RAW}(\mathbf{x})^{\gamma} - J_{GT}(\mathbf{x})\|^2} + eps\end{aligned} \tag{5}$$

where *eps* is a constant that takes a value 10^{-7}. SSIM loss is stated as follows:

$$\begin{aligned}&SSIM\_loss(J_{RAW}(\mathbf{x})^{\gamma}, J_{GT}(\mathbf{x})) \\ &= 1 - \frac{1}{Num}\sum_{x=1}^{Num}(SSIM(J_{RAW}(\mathbf{x})^{\gamma}, J_{GT}(\mathbf{x})))\end{aligned} \tag{6}$$

where *Num* is the total number of pixels, and $SSIM(J_{RAW}(\mathbf{x})^{\gamma}, J_{GT}(\mathbf{x}))$ can be expressed as:

$$\begin{aligned}&SSIM(J_{RAW}(\mathbf{x})^{\gamma}, J_{GT}(\mathbf{x})) \\ &= \frac{2\mu_{J_{RAW}(\mathbf{x})^{\gamma}}\mu_{J_{GT}(\mathbf{x})} + C_1}{\mu^2_{J_{RAW}(\mathbf{x})^{\gamma}} + \mu^2_{J_{GT}(\mathbf{x})} + C_1} \cdot \frac{2\sigma_{J_{RAW}(\mathbf{x})^{\gamma}J_{GT}(\mathbf{x})} + C_2}{\sigma^2_{J_{RAW}(\mathbf{x})^{\gamma}} + \sigma^2_{J_{GT}(\mathbf{x})} + C_2}\end{aligned} \tag{7}$$

where μ is the mean and σ is the standard deviation. $\sigma_{J_{RAW}(\mathbf{x})^{\gamma}J_{GT}(\mathbf{x})}$ is the covariance of the two. Both C_1 and C_2 are set to 0.01 in our experiments.

SwinIR branch 1 estimates the transmission map $T(\mathbf{x})$. Because the transmission map of hazy outdoor images contains little ground truth, we use $J_{GT}(\mathbf{x})$ to train SwinIR branch 1 by reconstructing hazy images $I_{REC\_1}(\mathbf{x})$. According to the ASM, the reconstructed hazy image $I_{REC\_1}(\mathbf{x})$ is stated as follows:

$$I_{REC\_1}(\mathbf{x}) = J_{GT}(\mathbf{x}) \cdot T(\mathbf{x}) + DCP\_A \cdot (1 - T(\mathbf{x})) \tag{8}$$

where DCP_A is calculated using DCP. According to DCP theory [2], in most non-sky localized regions of the image, at least one color channel has very low-intensity values that is close to zero, which can be expressed as follows:

$$J^{dark}(\mathbf{x}) = \underset{c\in\{r,g,b\}}{min}\left(\underset{y\in\Omega(\mathbf{x})}{min}\left(J^c(y)\right)\right), J^{dark}(\mathbf{x}) \to 0 \tag{9}$$

where J^c is the color channel of J, $\Omega(\mathbf{x})$ is the local area centered at $\mathbf{x}$, and J^{dark} is the dark channel of J. According to DCP, we first retrieve the dark channel of $I_{RAW}(\mathbf{x})$, then select the 0.1% brightest pixels of the dark channel and select the corresponding highest intensity value from $I_{RAW}(\mathbf{x})$ as the atmospheric light DCP_A.

The parameters of SwinIR branch 1 are optimized by minimizing the following loss function :

$$\begin{aligned}\zeta_{SwinIR1} = {} & Charbonnier\_loss(I_{REC\_1}(\mathbf{x}), I(\mathbf{x})) \\ & + SSIM\_loss(I_{REC\_1}(\mathbf{x}), I(\mathbf{x}))\end{aligned} \tag{10}$$

Stage 2 enhances the dehazing performance of the network in the real world. In this stage, we only use hazy images without a GT to fine-tune the network in a self-supervised learning manner. This method is highly beneficial to real-world hazy images, as it is difficult to obtain a clear GT. We also combine the DCP and ASM in this stage to allow the network to further learn the physical prior. However, two problems were encountered during the experiments. First, we found that the restored clear image obtained by the fine-tuned network did not improve significantly when self-supervised learning was performed using A (estimated by DCP), $T(\mathbf{x})$ (estimated by SwinIR branch 1), and $J_{RAW}(\mathbf{x})^{\gamma}$ (estimated by SwinIR branch 2) to reconstruct the hazy image through ASM. Second, we found that the proposed network model led to dark patches in local regions of the recovered image in the synthetic hazy dataset, which may be caused by the uniform haze distribution of synthetic hazy images. For the first problem, we believe that the process of reconstructing hazy images based on the ASM is not sufficiently accurate. In fact, this is relatively easy to understand because the images we mainly input at this stage are hazy real-world images. These images have different haze concentrations and non-uniform haze distributions; however, ASM is an idealized modeling framework, which is not fully suitable for complex hazy scenes. Therefore, after reconstructing the hazy image based on the ASM, we used a generator (Generator 1 in Fig. 3) to further fit the imaging process to compensate for the deficiencies of ASM and DCP. For the second problem, we separately trained another generator (Generator 2 in Fig. 3) as an image enhancer to further enhance the recovery of scenes in dark patches in the image. The two generators we used were relatively simple generators in CycleGan [77], mainly composed of nine residual blocks [35]. Generator 2 has one more skip-connect than Generator 1, as shown in Fig. 6.

Loss function of Stage 2. ASM, DCP, and Generator 1 in this stage all participate in self-supervised learning to fine-tune the network parameters used in Stage 1. The reconstructed hazy image $I_{REC\_2}(\mathbf{x})$ in this stage is stated as follows:

$$\begin{aligned}I_{REC\_2}(\mathbf{x}) = {} & Generator\_1(J_{RAW}(\mathbf{x})^{\gamma} \cdot T(\mathbf{x}) \\ & + DCP\_A \cdot (1 - T(\mathbf{x})))\end{aligned} \tag{11}$$

We fine-tune the network parameters by minimizing the following loss function:

$$\begin{aligned}\zeta_{Stage\_2} = {} & Charbonnier\_loss(I_{REC\_2}(\mathbf{x}), I(\mathbf{x})) \\ & + SSIM\_loss(I_{REC\_2}(\mathbf{x}), I(\mathbf{x})) \\ & + Charbonnier\_loss(DCP\_T(\mathbf{x}), T(\mathbf{x}))\end{aligned} \tag{12}$$

where $DCP\_T(\mathbf{x})$ is the transmission map estimated using DCP. According to the ASM and DCP (Eqs. (1) and (9)), $DCP\_T(\mathbf{x})$ can be derived as follows:

$$DCP\_T(\mathbf{x}) = 1 - \omega \underset{c\in\{r,g,b\}}{min}\left(\underset{y\in\Omega(\mathbf{x})}{min}\left(\frac{I^c_{RAW}(y)}{DCP\_A^c}\right)\right) \tag{13}$$

where ω is a parameter that ensures the existence of a small amount of haze, which is usually set to 0.95. In this loss function, we impose a DCP constraint on $T(\mathbf{x})$ so that the $T(\mathbf{x})$ estimated by SwinIR branch 1 is more in line with the physical prior. Through experiments, we found that if there is no prior constraint imposed on $T(\mathbf{x})$, the fine-tuning process will be very unstable and difficult to converge. For Generator 2, we optimize the generator network parameters by minimizing the following loss function:

$$\begin{aligned}\zeta_{Generator\_2} = {} & Charbonnier\_loss(J'(\mathbf{x}), J_{GT}(\mathbf{x})) \\ & + SSIM\_loss(J'(\mathbf{x}), J_{GT}(\mathbf{x}))\end{aligned} \tag{14}$$

where $J'(\mathbf{x})$ is a clear image obtained by $J(\mathbf{x})$ using Generator 2. Generator 2 is trained only on synthetic datasets and was not part of the self-supervised learning of the network.

In particular, we did not add hyperparameter weights to all loss functions, so as to verify the real performance of our network architecture without manually tuning the hyperparameters to obtain satisfactory results.

Through the abovementioned detailed introduction to our network architecture, we skillfully combine several different concepts, methods, techniques, and components, including deep learning, physical model, and prior knowledge, Swin Transformer, supervised learning and self-supervised learning. Experiments have proved that this can be well used for image dehazing tasks. The pseudocode of the SwinTD-Net training scheme is presented in Algorithm 1.

Algorithm 1: Training Scheme of SwinTD-Net

```
Input:
  Hazy images $I(\mathbf{x})$
  Ground truth images $J_{GT}(\mathbf{x})$
Initialize: Learning rate and network parameters
  (1) Convert $I(\mathbf{x})$ to the RAW hazy image $I_{RAW}(\mathbf{x})$ through
      inverse gamma correction.
  (2) Pre-training the network via supervised learning in
      Stage 1, where SwinIR branch 1 and SwinIR branch 2 are
      trained using Eqs. (10) and (4), respectively.
  (3) Convert $J_{RAW}(\mathbf{x})$ estimated by SwinIR branch 2 to RGB
      image $J(\mathbf{x})$ through gamma correction.
  (4) Fine-tune the network via self-supervised learning
      using Eq. (12) in Stage 2. For the synthetic dataset, train
      Generator 2 for image enhancement using Eq. (14) (note
      that it is not required for the real-world dataset).
Output: Trained network parameters
```

3.3. Multi-scale dark-light enhancement algorithm (MSDLEA)

The proposed MSDLEA is used to improve the visual quality perceived by the human eye while performing image detail enhancement, particularly in darker areas in the image. The design of the proposed MSDLEA includes the following requirements.

(1) *The brightness area of the image remains basically unchanged.*

(2) *The brightness of dark areas of the image is increased.*

(3) *The image transitions naturally, and the relative relationship between the brightness and darkness of the image cannot be changed.*

(4) *The dark area of the image cannot be improved infinitely, and good layering and contrast should be retained.*

(5) *Image color distortion is slight, and hue is maintained.*

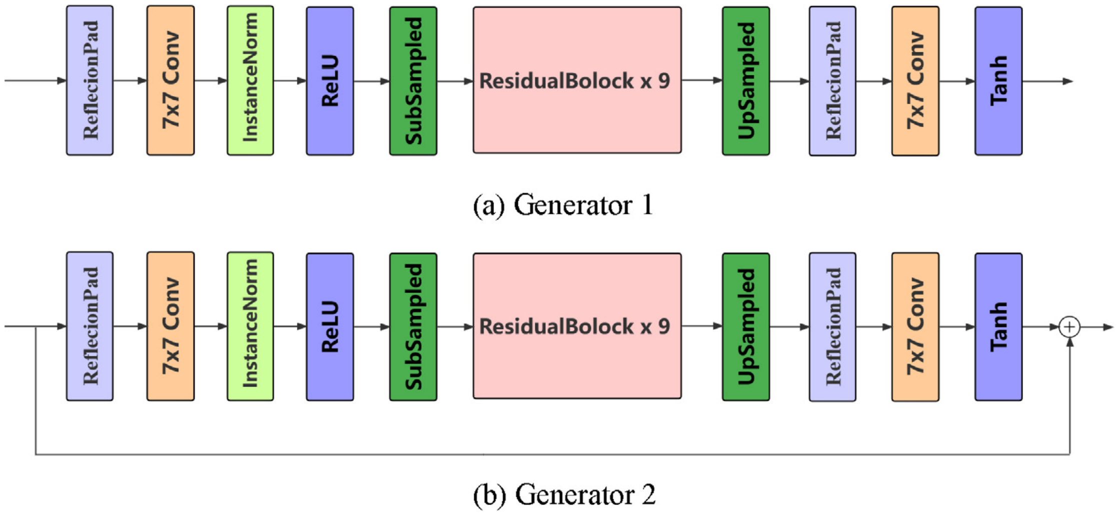


Fig. 6. Generator 1 and Generator 2. The two generators are mainly composed of ReflectionPad, 7*7 convolution, InstanceNorm, Relu, subsampled, 9 residual modules, upsampled, Tanh and other layers. Generator 2 has one more skip connection than Generator 1.

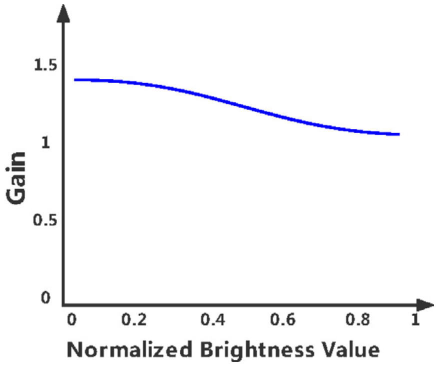


Fig. 7. Schematic diagram of brightness gain function curve. The abscissa represents the normalized brightness value, while the ordinate represents the brightness value after gain.

(6) *The contrast of details inside the dark areas of the image should be increased.*

To satisfy the above requirements, we chose to convert the image to the HSV space and operate on the V channel to preserve its original color as much as possible. Then, we used mean filters of different scales to filter the images of the V channel to obtain different relatively smooth background brightness maps. Next, through the designed Brightness-Gain function, multiple background brightness maps were generated to obtain a multi-scale brightness gain map and fused to obtain the final brightness gain map. Then, we processed the brightness gain map through Weighted Guided Image Filtering (WGIF) [78] to maximize image edge preservation and suppress artifacts. Finally, the RGB channels of the input image were multiplied by the brightness gain map and then fused to obtain the obtained image. A schematic of the brightness gain-function curve is shown in Fig. 7. The designed Brightness-Gain function $BG(x)$ can be expressed as follows:

$$maxGain = \frac{1}{1 + e^{\alpha \cdot (0-\beta)}} \tag{15}$$

$$minGain = \frac{1}{1 + e^{\alpha \cdot (L-\beta)}} \tag{16}$$

$$bg(x) = \frac{1}{1 + e^{(x-\beta)}} \tag{17}$$

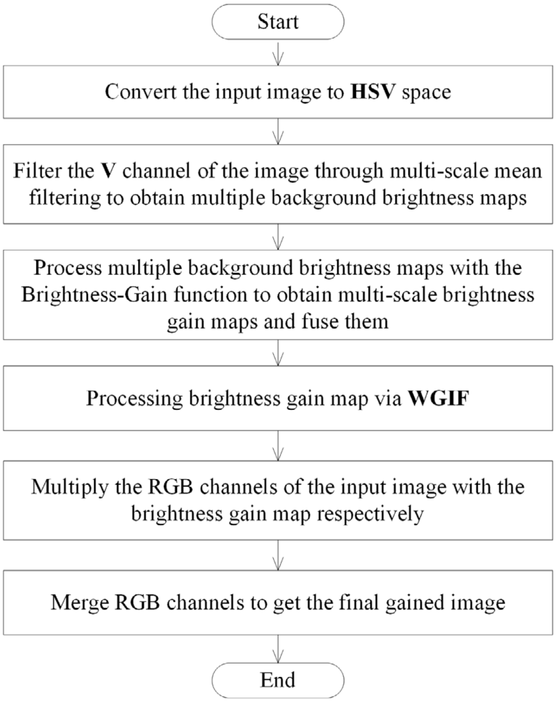

Start
Convert the input image to **HSV** space
Filter the **V** channel of the image through multi-scale mean filtering to obtain multiple background brightness maps
Process multiple background brightness maps with the Brightness-Gain function to obtain multi-scale brightness gain maps and fuse them
Processing brightness gain map via **WGIF**
Multiply the RGB channels of the input image with the brightness gain map respectively
Merge RGB channels to get the final gained image
End

Fig. 8. Flowchart of the MSDLEA.

$$BG(x) = 1 + C \times \frac{bg(x) - minGain}{maxGain - minGain} \tag{18}$$

where α, β and C are all hyperparameters that can be adjusted as needed. As shown in Fig. 7, α determines the steepness of the curve, β determines the position of the turning point of the curve, and C controls the curve's maximum gain. $maxGain$ is the original maximum gain, L is the brightness value corresponding to the gain of 1, $minGain$ is the original gain corresponding to the brightness L, $bg(x)$ is the original Brightness-Gain function that changes with brightness x, $BG(x)$ is the Brightness-Gain function after gain offset and amplitude change. A flowchart of the MSDLEA is shown in Fig. 8. A visual presentation of the intermediate results of each step of the MSDLEA is shown in Fig. 9.

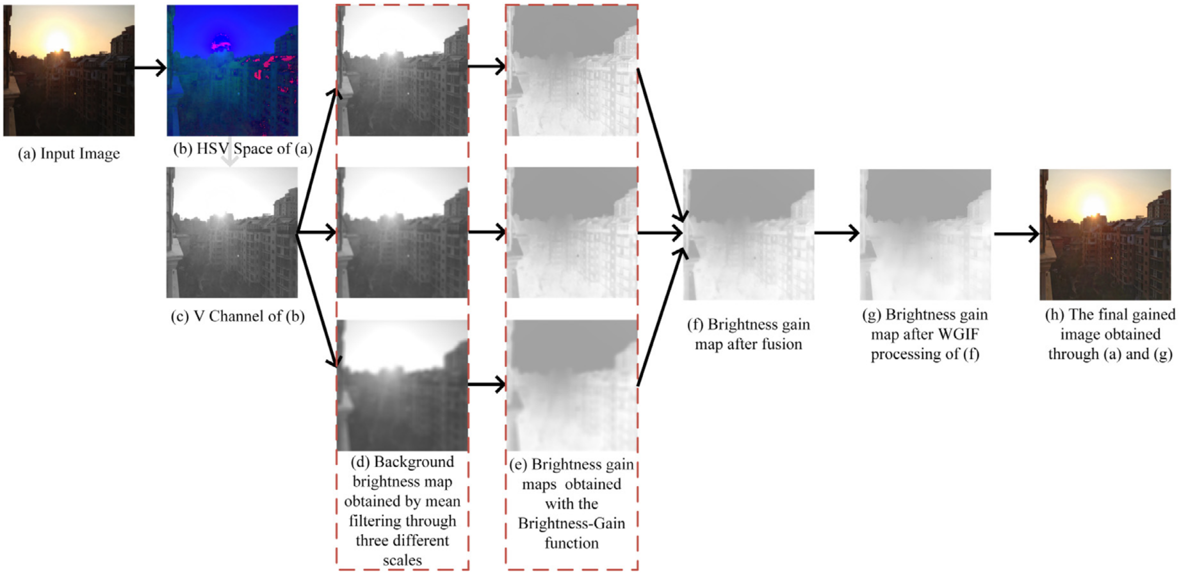

(a) Input Image
(b) HSV Space of (a)
(c) V Channel of (b)
(d) Background brightness map obtained by mean filtering through three different scales
(e) Brightness gain maps obtained with the Brightness-Gain function
(f) Brightness gain map after fusion
(g) Brightness gain map after WGIF processing of (f)
(h) The final gained image obtained through (a) and (g)

Fig. 9. Visual presentation of intermediate results for each step of MSDLEA. (a) is the input image, (b) is the HSV space image of (a), (c) is the V channel image of (b), (d) is the background brightness map after three-scale mean filtering of (c), (e) is the brightness gain map after Brightness-Gain function processing of (d), (f) is the brightness gain map after fusion of (e), (g) is the brightness gain map after WGIF processing of (f), (h) is the final gained image processed through (a) and (g).

MSDLEA is a general image enhancement algorithm that can be used for post-processing operations in various application scenarios such as image deraining, image denoising, and medical image enhancement.

4. Experiments

We present the datasets used and detailed experimental details and analysis, especially the comparison with other competing methods, as well as ablation studies and discussions.

4.1. Datasets and training details

Synthetic Dataset. The RESIDE dataset [79] was selected as the training dataset. It contains the indoor synthetic datasets ITS, outdoor synthetic datasets OTS, and synthetic test datasets SOTS. Note that only ITS contains the transmission map ground truth, whereas OTS does not; therefore, our network does not use the transmission map ground truth.

Real-world dataset. We selected the O-HAZE [80], NH-HAZE [81,82] and Dense-HAZE datasets [83,84] for the experiments.

Training details. Our experimental environment was Pytorch 1.10.2, with four NVIDIA GeForce GTX 1080Ti GPUs. We used the AdamW [85] optimizer to optimize the SwinTD-Net.

(1) *Indoor Training Set:* In Stage 1, we used all images of ITS for training. We scaled the pixel value of the image to between 0 and 1 with a batch size of 32 and patch size of 64 × 64. Therefore, the data size of the network input was [32, 64, 64, 3]. The two branches of SwinIR are independently trained using Eqs. (4) and (10), with a learning rate of 0.00002, for 1000 epochs. In Stage 2, we first train Generator 1. However, when we train directly with self-supervised learning, it is difficult for the network to converge. Therefore, Generator 1 is also pre-trained. We freeze the parameters of the two SwinIR (*i.e.*, the learning rate of the network is 0), and set the data size of the network input to [16, 64, 64, 3] (because now two SwinIR branches are forward propagating, the batch size is halved). Then we randomly select 1280 images (*i.e.*, train 80 batches) to pre-train Generator 1 by $J_{RAW}(\mathbf{x})^{\gamma}$, $T(\mathbf{x})$ and *DCP_A* via Eq. (12). After pre-training Generator 1, the network has preliminary parameters, then we randomly select 4160 images (*i.e.*, train 260 batches) to use self-supervised learning to fine-tune the two SwinIR branches and Generator 1 through Eq. (12). Subsequently, the learning rates of SwinIR branch 1, SwinIR branch 2, and Generator 1 are set to 0.00002, 0.000002, and 0.0002, respectively. Finally, we train the image enhancer (*i.e.*, Generator 2), using the clean image $J_{RAW}(\mathbf{x})^{\gamma}$ recovered by SwinIR2, providing the input Generator 2 for 200 training epochs with a patch size of 32 and learning rate of 0.00002. Note that Generator 2 is only used for synthetic hazy images and not for real-world hazy images. For the synthetic dataset, we selected the top five percent darker images for training Generator 2 to enhance the details in the darker regions of the image. For the real-world dataset, we did not train Generator 2 because of too few samples (for example, the NH-Haze has only 55 images, and 5 percent of them have only two images, which is not enough to train the network).

(2) *Indoor Evaluation:* We used 500 indoor image pairs obtained from SOTS for testing, including quantitative and qualitative evaluations. For a quantitative evaluation, we used the PSNR and SSIM indicators [74] widely used by scholars.

(3) *Outdoor Training Set:* We use all images from OTS for training with the same training method and settings as the indoor training.

(4) *Outdoor Evaluation:* We use 500 outdoor image pairs from SOTS for testing, including quantitative and qualitative evaluations.

(5) *Training and evaluation on real-world hazy images:* Image dehazing primarily involves achieving outdoor real-world image dehazing. To experimentally verify the performance of the proposed algorithm on real hazy images, we selected the O-HAZE [80], NH-HAZE [81,82] and Dense-HAZE datasets [83,84]. The haze in these datasets was generated by professional haze machines in outdoor scenes. Among them, the O-HAZE contains 45 pairs of different haze scene images and corresponding clear images, NH-HAZE contains 55 pairs of different outdoor scene images with non-homogeneous haze and corresponding clear images, and Dense-HAZE publishes 55 pairs of dense hazy images and corresponding clear images of different real-world outdoor scenes. Because these datasets had fewer images, we adopted a cross-training and testing approach. We retrained our network on NH-HAZE and tested it on O-HAZE. Similarly, we retrained our network on the O-HAZE and then tested it on the NH-HAZE and

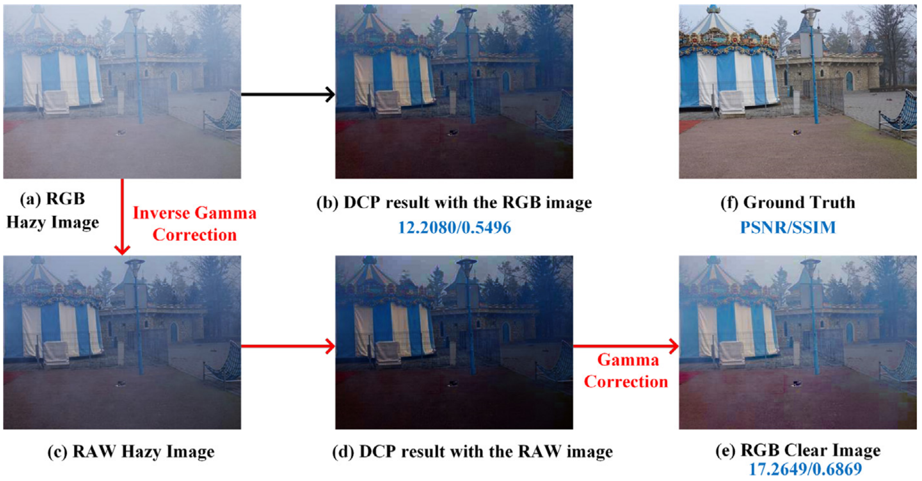


Fig. 10. Comparison of visual effects after DCP dehazing using the RGB hazy image and the RAW hazy image respectively, where (a) is a RGB hazy image, (b) is to use (a) for dehazing directly, (c) is to convert (a) into a RAW image through inverse gamma correction, (d) is the dehazing result of (c), and (e) is to perform gamma correction on (d) to obtain the RGB image, (f) is ground truth. The blue numbers are the PNSR and SSIM scores of the two methods, respectively.

Table 1
The average quantitative evaluation results for prior input. The bolded number represents the best metric score.

Methods	SOTS-Outdoor		O-Haze		NH-Haze		Dense-Haze	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP with the RGB Image	14.80	0.8029	14.42	0.5028	11.57	0.4183	12.03	0.4033
DCP with the RAW Image	**18.59**	**0.8590**	**17.38**	**0.6285**	**13.14**	**0.4918**	**12.46**	**0.4447**

Dense-HAZE. Therefore, for real-world hazy images, the network was trained using a small sample of data. Although we did not train Generator 2, our network performed well.

Comparison with state-of-the-art (SOTA) methods. We separately selected prior-based methods and deep learning-based methods for comparative experiments, including **DCP** [2], **DehazeNet** [4], **AODNet** [5], **GridDehazeNet** [22], **MSBDN** [21], **FFANet** [53], **RefineDNet** [31], **PSD** [32], $\boldsymbol{D^4}$ [60]. For quantitative evaluation, we employ PSNR and SSIM metrics [74] for comparative experiments. To ensure the fairness of the experiments, we re-tested all datasets used for experimental comparison, including visual results and quantitative evaluations, using publicly available codes and models for these methods.

4.2. Experiments for prior input

To verify the proposed more appropriate prior input for more effectively using ASM and DCP, we conducted comparative experiments. Because DCP is based on ASM, we used DCP for the experiments. First, we used the RGB hazy image as input so that it could be restored through DCP; then used the inverse gamma-corrected RAW hazy image as the input to restore the image through DCP to obtain the RAW clear image. Subsequently, we performed gamma correction to obtain the RGB clear image. We randomly selected a hazy image from the O-haze dataset and visually observed the differences during the process of restoring a clear image. Fig. 10 depicts the result of the DCP restoration for the RGP input is dark, which loses a significant amount of color and detail. However, converting a RGB input into a RAW image through inverse gamma correction and then using DCP to achieve dehazing and converting result into a RGB clear image can restore more image color and detail. The PSNR and SSIM scores are presented in Fig. 10, which also demonstrates the effectiveness of the RAW image input. In addition, to perform a quantitative experimental evaluation in outdoor scenes, we selected SOTS-Outdoor, O-HAZE, NH-HAZE, and Dense-HAZE datasets; the results are reported in Table 1. Observe from Table 1 that using DCP with RAW image inputs can obtain higher PNSR and SSIM scores, which strongly illustrates the rationality of the proposed approach that takes RAW images as prior inputs. Therefore, we recommend that, in the ISP pipeline of a digital camera, the dehazing operation should be applied on the RAW image before gamma correction.

4.3. Experiments on synthetic hazy images

We conduct comparative experiments to verify the performance of the proposed network framework through qualitative and quantitative evaluations, respectively. We selected the test datasets SOTS-Indoor, SOTS-Outdoor from RESIDE dataset [79], and Frida foggy road image database (FRIDA2) [86] to conduct comparative experiments.

SOTS-Indoor Experiment Results. Fig. 11 depicts a comparison of the visual results obtained with different dehazing methods. Observe from Fig. 11 that the images recovered by DCP are darker, whereas the images recovered by DehazeNet, AODNet, PSD, and D^4 retain residual haze. The images recovered by GridDehazeNet is distorted, such as the obvious black area can be seen in the partially enlarged part in Image 2 and Image 3. The images recovered by the MSBDN have insufficient contrast and insufficient detail recovery, such as the partially enlarged part in Images 1 and 3. The FFANet method is slightly overly dehazing, and the partially enlarged part in Image 1 excessively increases the brightness of the dark area, whereas the partially enlarged part in Image 3 insufficiently restores color. The images restored by RefineDNet demonstrate obvious noise, as can be seen from the enlarged part of Image 3, and obvious color distortion, as shown in the table in Image 1 with obvious deviation. In addition, observe from the partially enlarged part in Image 3 that the D^4 method may also produce overexposure. The images recovered by our method are closest to GT in terms of color and detail, and the PSNR and SSIM scores further reflect the effectiveness of our method.

SOTS-Outdoor Experiment Results. Fig. 12 compares the visual results obtained by different methods. From Fig. 12, observe that the images recovered by the DCP are dark and heavily distorted in

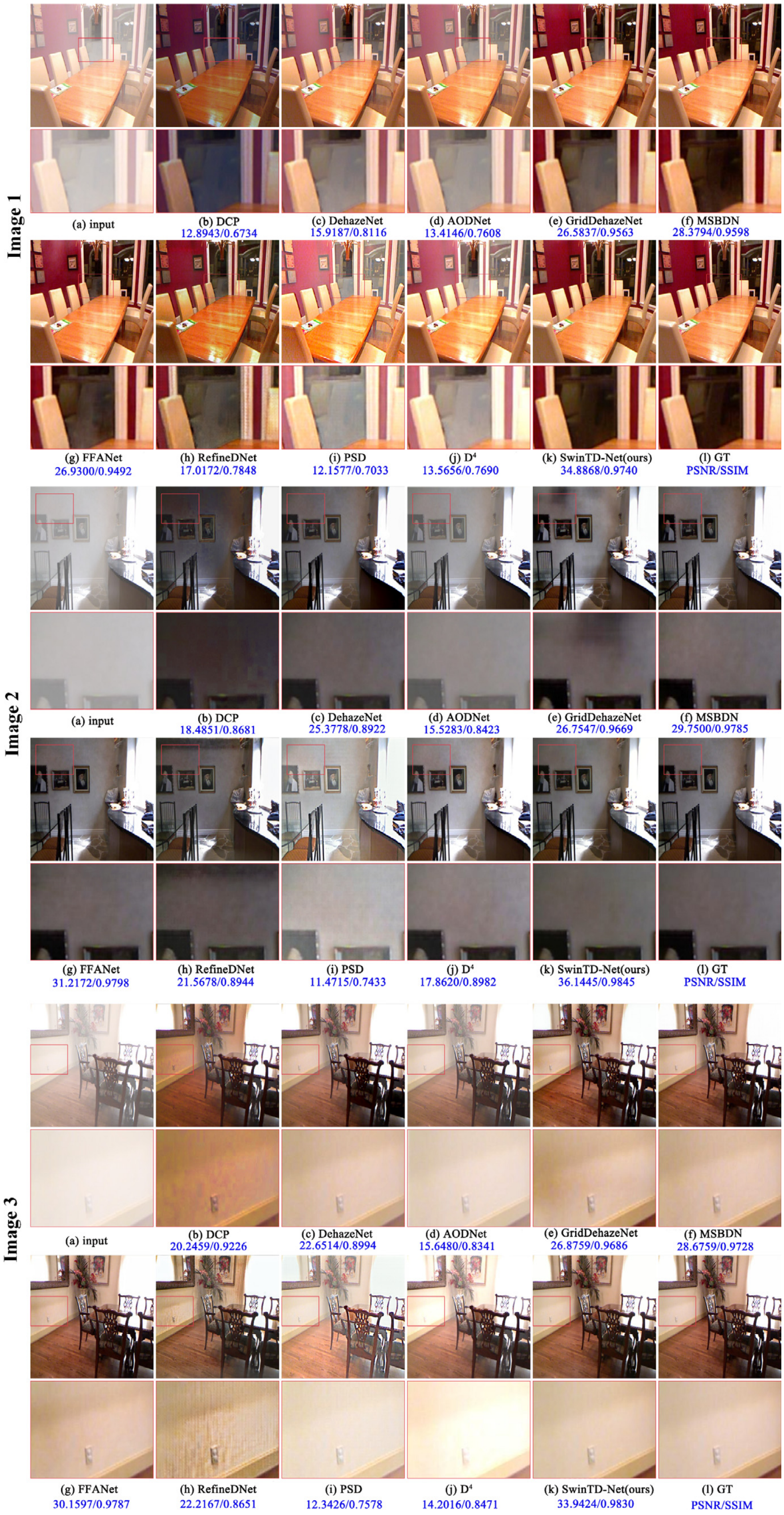


Fig. 11. Comparison of visual effects of different image dehazing methods on SOTS-Indoor hazy images. Blue text indicates PSNR/SSIM values. (a) is the hazy input image, from (b) to (j) are the results recovered by the SOTA methods, (k) is the result recovered by our method, and (l) is the ground truth image.

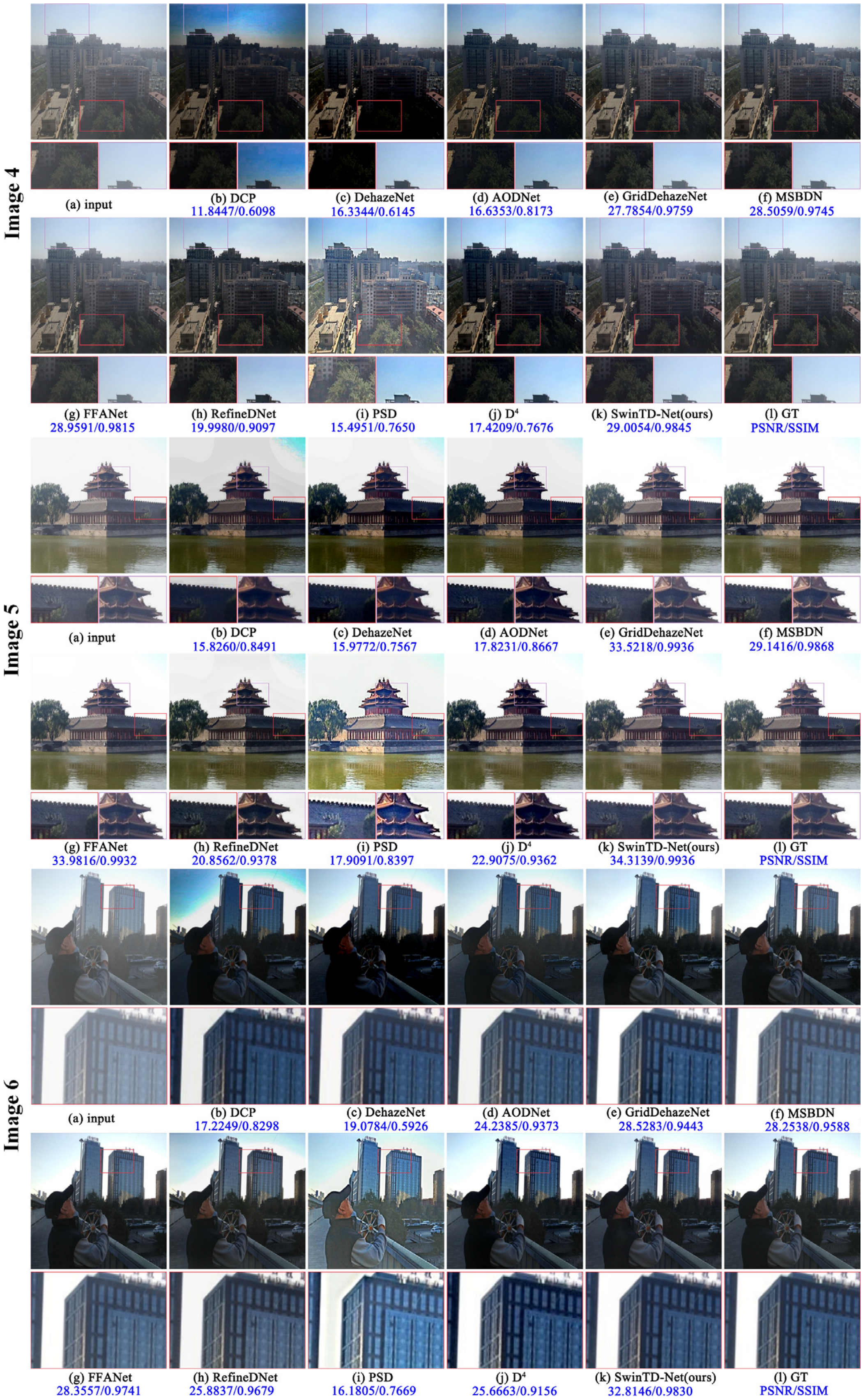


Fig. 12. Comparison of visual effects of different image dehazing methods on SOTS-Outdoor hazy images. Blue text indicates PSNR/SSIM values. (a) is the hazy input image, from (b) to (j) are the results recovered by the SOTA methods, (k) is the result recovered by our method, and (l) is the ground truth image.

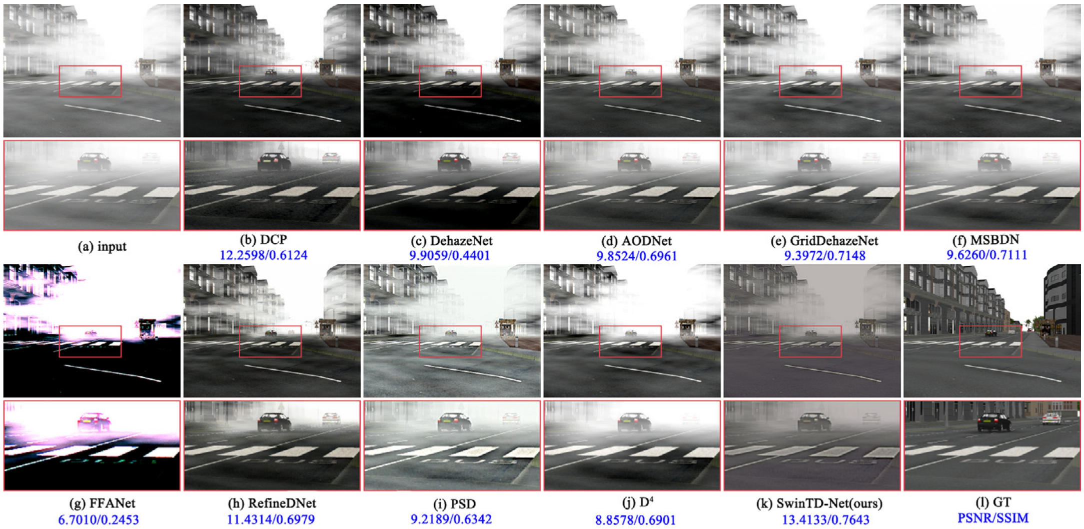

(a) input
(b) DCP 12.2598/0.6124
(c) DehazeNet 9.9059/0.4401
(d) AODNet 9.8524/0.6961
(e) GridDehazeNet 9.3972/0.7148
(f) MSBDN 9.6260/0.7111
(g) FFANet 6.7010/0.2453
(h) RefineDNet 11.4314/0.6979
(i) PSD 9.2189/0.6342
(j) D^4 8.8578/0.6901
(k) SwinTD-Net(ours) 13.4133/0.7643
(l) GT PSNR/SSIM

Fig. 13. Comparison of visual effects of different methods on an image from FRIDA2 dataset. Blue text indicates PSNR/SSIM values. (a) is the hazy input image, from (b) to (j) are the results recovered by the SOTA methods, (k) is the result recovered by our method, and (l) is the ground truth image.

Table 2
Quantitative evaluation on synthetic datasets. The bolded number represents the best metric score, and the underlined number represents the second-best metric score.

Methods	SOTS-Indoor		SOTS-Outdoor		FRIDA2	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP(CVPR 2009) [2]	18.42	0.8259	14.80	0.8029	12.61	0.6777
DehazeNet (TIP 2016) [4]	20.84	0.8122	18.69	0.7426	10.99	0.5440
AODNet (ICCV 2017) [5]	18.73	0.8268	19.64	0.8920	11.52	0.7454
GridDehazeNet (ICCV 2019) [22]	29.86	0.9742	30.34	0.9797	11.34	0.7752
MSBDN (CVPR 2020) [21]	31.98	**0.9796**	**33.49**	**0.9835**	11.40	0.7618
FFANet (AAAI 2020) [53]	32.07	0.9794	30.12	0.9721	8.400	0.3714
RefineDNet (TIP 2021) [31]	20.32	0.8364	20.77	0.9093	13.00	0.7544
PSD (CVPR 2021) [32]	12.46	0.6802	15.31	0.7621	10.95	0.6979
D4 (CVPR 2022) [60]	19.66	0.8563	25.06	0.9396	10.53	0.7022
SwinTD-Net (Ours)	**36.79**	0.9676	32.07	0.9811	**13.04**	**0.7933**

the sky region. The DehazeNet and AODNet methods lead to over-saturation, for example, the buildings in Image 4. GridDehazeNet and FFANet can cause color distortions, such as in the roof area in Image 5. MSBDN also exhibits color distortion, as evident in the partially enlarged area in Image 6. The images recovered by RefineDNet, PSD, and D^4 exhibit severe chromatic aberrations, such as the sky area in Image 4. PSD also contains the over-explosion phenomenon, and the brightness of the restored image is high. The images recovered by our method are closest to the GT, and the PSNR and SSIM scores further reflect the effectiveness of our method.

FRIDA2 Experiment Results. To experimentally verify the performance of the proposed method on foggy images, we selected heterogeneous road foggy images from the FRIDA2 dataset. [86] for experimental comparison. Fig. 13 shows a visual comparison of the foggy road image. Observe that DCP, DehazeNet, and RefineDNet perform better with foggy images; the details of buildings in the distance are restored more but the ground remains too dark. The recovery results of AODNet, GridDehazeNet, MSBDN, PSD, and D4 retain more residual fog. The results recovered by FFANet are severely distorted, which illustrates the limitation of the method on foggy images. The defogging ability of our method is not as good as that of the DCP; nonetheless, the distortion is reduced as illustrated by the PNSR and SSIM scores.

The quantitative evaluation results of all images taken from the SOTS and FRIDA2 datasets are reported in Table 2. For the SOTS-Indoor dataset, the PSNR score of our method is the highest, whereas the SSIM score is only 0.012 away from the highest value. For the SOTS-Outdoor dataset, our method achieves the second-best PNSR and SSIM scores, which are extremely close to the top score. Our method scores slightly lower than MSBDN, which may be because MSBDN is a supervised learning method. Specifically, MSBDN is trained with supervised learning on a synthetic dataset and achieves high scores on synthetic hazy images. However, because it does not incorporate physical priors, its dehazing effect on real-world scenes is worse than that of our method. Subsequent experiments illustrate this point. For the FRIDA2 dataset, our method still achieves the highest score but its advantage is not obvious compared to the second-best approach, which underlines that the performance of our method on foggy images needs to be further improved.

4.4. Experiments on real-world hazy images

Image dehazing comprises a preprocessing step for advanced computer vision tasks undertaken in real hazy environments; thus, we further validate the performance of the proposed method on real-world hazy images. As shown in Figs. 14, 15 and 16, we sampled 2 images from the O-HAZE, NH-HAZE and Dense-HAZE datasets for visual effect comparison, respectively. From Fig. 14, observe that the images recovered by DCP and AODNet have less haze but are too dim. The image brightness recovered by DehazeNet and PSD methods is higher but the dehazing effect is insufficient. The images recovered by methods such as GridDehazeNet, MSBDN, FFANet, RefineDNet, and D^4 methods all exhibit residual haze. However, the haze residue on the images obtained our method is the least, and the obtained result is the closest to GT. Similarly, from Fig. 15, observe that

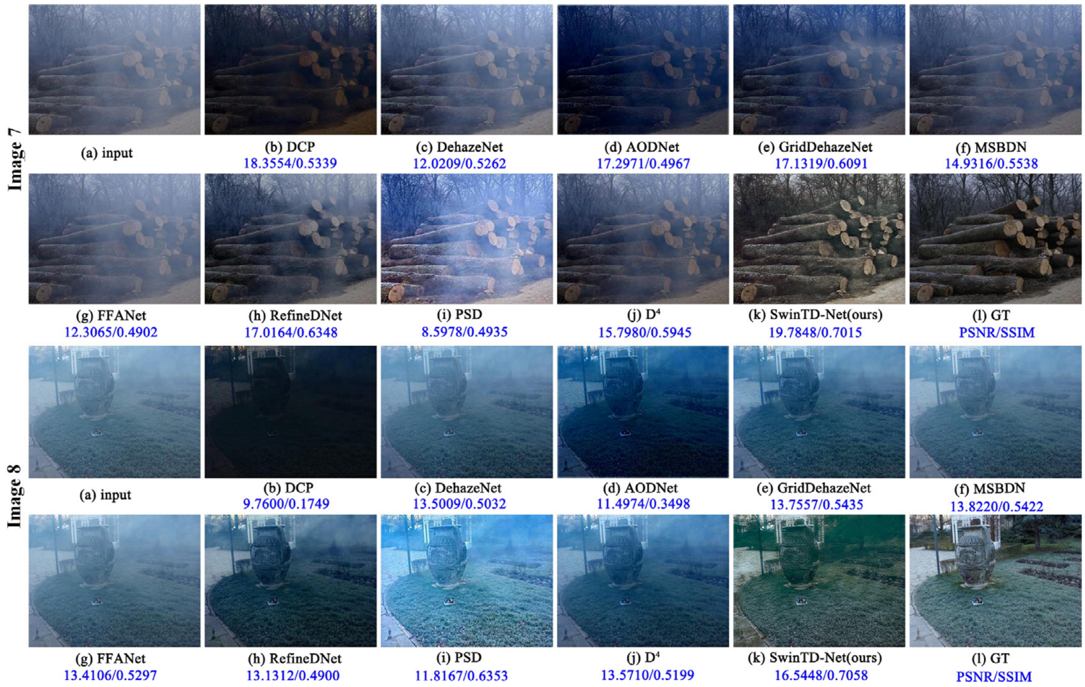


Image 7
(a) input
(b) DCP 18.3554/0.5339
(c) DehazeNet 12.0209/0.5262
(d) AODNet 17.2971/0.4967
(e) GridDehazeNet 17.1319/0.6091
(f) MSBDN 14.9316/0.5538
(g) FFANet 12.3065/0.4902
(h) RefineDNet 17.0164/0.6348
(i) PSD 8.5978/0.4935
(j) D⁴ 15.7980/0.5945
(k) SwinTD-Net(ours) 19.7848/0.7015
(l) GT PSNR/SSIM
Image 8
(a) input
(b) DCP 9.7600/0.1749
(c) DehazeNet 13.5009/0.5032
(d) AODNet 11.4974/0.3498
(e) GridDehazeNet 13.7557/0.5435
(f) MSBDN 13.8220/0.5422
(g) FFANet 13.4106/0.5297
(h) RefineDNet 13.1312/0.4900
(i) PSD 11.8167/0.6353
(j) D⁴ 13.5710/0.5199
(k) SwinTD-Net(ours) 16.5448/0.7058
(l) GT PSNR/SSIM

Fig. 14. Comparison of visual effects of different image dehazing methods on the O-HAZE dataset. Blue text indicates PSNR/SSIM values. (a) is the hazy input image, from (b) to (j) are the results recovered by the SOTA methods, (k) is the result recovered by our method, and (l) is the ground truth image.

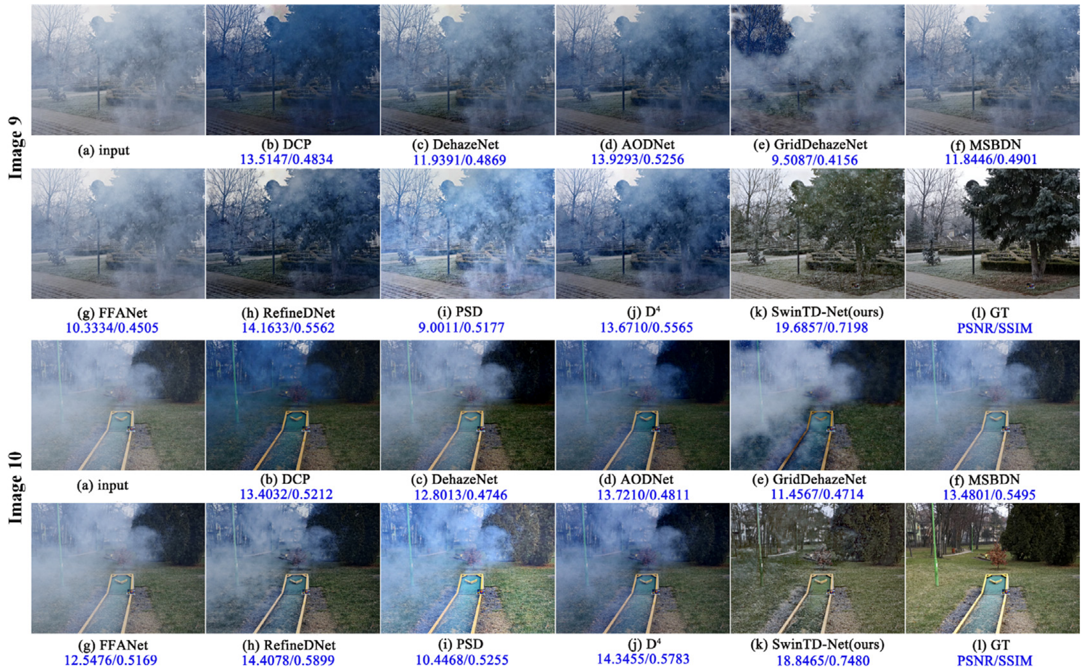


Image 9
(a) input
(b) DCP 13.5147/0.4834
(c) DehazeNet 11.9391/0.4869
(d) AODNet 13.9293/0.5256
(e) GridDehazeNet 9.5087/0.4156
(f) MSBDN 11.8446/0.4901
(g) FFANet 10.3334/0.4505
(h) RefineDNet 14.1633/0.5562
(i) PSD 9.0011/0.5177
(j) D⁴ 13.6710/0.5565
(k) SwinTD-Net(ours) 19.6857/0.7198
(l) GT PSNR/SSIM
Image 10
(a) input
(b) DCP 13.4032/0.5212
(c) DehazeNet 12.8013/0.4746
(d) AODNet 13.7210/0.4811
(e) GridDehazeNet 11.4567/0.4714
(f) MSBDN 13.4801/0.5495
(g) FFANet 12.5476/0.5169
(h) RefineDNet 14.4078/0.5899
(i) PSD 10.4468/0.5255
(j) D⁴ 14.3455/0.5783
(k) SwinTD-Net(ours) 18.8465/0.7480
(l) GT PSNR/SSIM

Fig. 15. Comparison of visual effects of different image dehazing methods on the NH-HAZE dataset. Blue text indicates PSNR/SSIM values. (a) is the hazy input image, from (b) to (j) are the results recovered by the SOTA methods, (k) is the result recovered by our method, and (l) is the ground truth image.

only the image recovered by our method is closest to GT, and the other methods cannot effectively eliminate haze. In Fig. 16, owing to the high haze concentration, it is difficult for the human eye to depict the real scene in the haze, and various advanced dehazing methods cannot restore high-quality clear images well. However, our method remains the best for image dehazing, returning the results closest to GT. Furthermore, we found that the learning-based methods GridDehazeNet, FFANet, and MSBDN attaining high PSNR and SSIM scores on the synthetic datasets basically fail on real-world hazy images, especially on the NH-HAZE and Dense-HAZE dataset. Our method also obtains the highest metric score in Figs. 14, 15 and 16. Furthermore, from Image 12, observe that, except for our method, the images recovered by DCP and RefineDNet methods can retain more scene content and less haze than other methods, which we believe is due to these two methods both utilizing physical mechanisms (*i.e.*, dark channel prior and atmospheric scattering model). Therefore, we trust that it is necessary to study image dehazing algorithms guided by physical mechanisms, especially for densely hazy scenes.

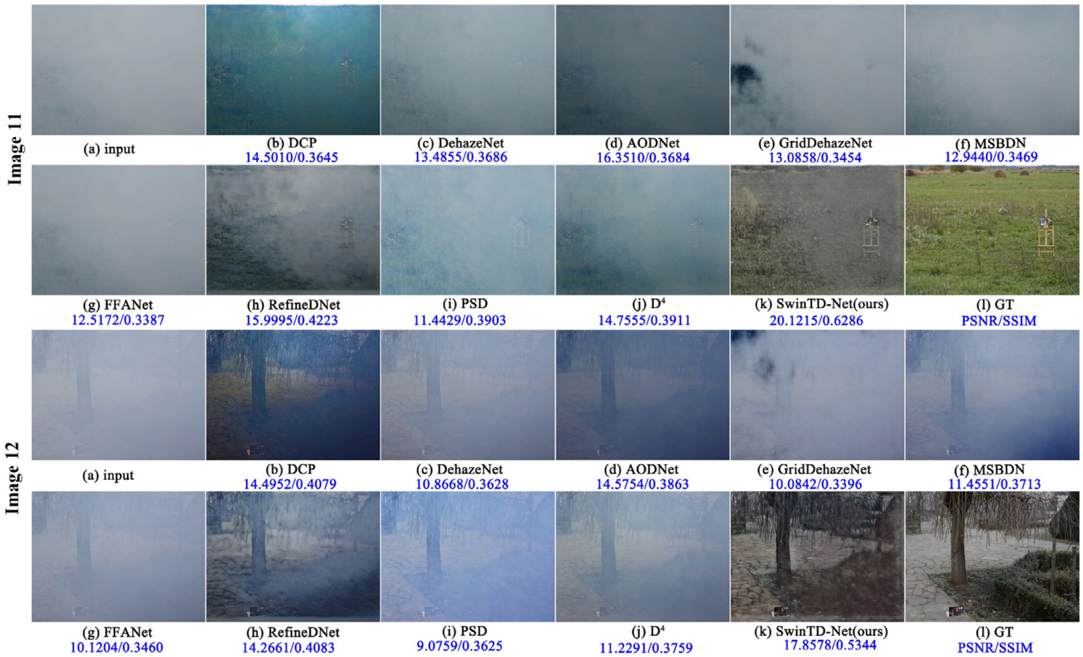


Fig. 16. Comparison of visual effects of different image dehazing methods on the Dense-HAZE dataset. Blue text indicates PSNR/SSIM values. (a) is the hazy input image, from (b) to (j) are the results recovered by the SOTA methods, (k) is the result recovered by our method, and (l) is the ground truth image.

Table 3
Quantitative evaluation on the real-world hazy dataset. The bolded number represents the best metric score, and the underlined number represents the second-best metric score.

Methods	O-HAZE		NH-HAZE		Dense-HAZE	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DCP(CVPR 2009) [2]	14.42	0.5028	11.57	0.4183	12.03	0.4033
DehazeNet (TIP 2016) [4]	15.48	0.6016	11.85	0.4435	11.04	0.4334
AODNet (ICCV 2017) [5]	15.09	0.5433	11.87	0.4246	12.69	0.4591
GridDehazeNet (ICCV 2019) [22]	15.96	0.6395	10.53	0.3874	10.07	0.3986
MSBDN (CVPR 2020) [21]	16.32	0.6366	12.45	0.4903	11.02	0.4533
FFANet (AAAI 2020) [53]	14.68	0.5990	11.93	0.4708	10.35	0.4412
RefineDNet (TIP 2021) [31]	17.32	0.6699	12.79	0.5146	12.91	0.4584
PSD (CVPR 2021) [32]	11.39	0.6095	10.32	0.5274	09.73	0.4345
D4 (CVPR 2022) [60]	16.75	0.6578	12.66	0.5072	11.49	0.4536
SwinTD-Net (Ours)	**18.57**	**0.7064**	**15.92**	**0.6789**	**14.62**	**0.4821**

The quantitative evaluation results of all images taken from the three real-world hazy datasets are presented in Table 3. Our method achieves the highest metric score on these real-world hazy datasets and significantly outperforms the second-best method in terms of PSNR score (on the NH-HAZE dataset, 3.13 points more than the second-best method). Both qualitative and quantitative experimental comparison results conducted based on real-world scenarios illustrate the excellent performance of the proposed method. However, our method suffers from limitations when applied to real-world hazy images. The subjective experimental results presented in Image 7 – Image 10 indicate that the sharp lines in the images processed by the proposed method are relatively clear but clumpy haze appears on the surfaces of lawns and trees, which indicates to a weaker performance compared with the other methods. This may be because our network pays more attention to the overall image dehazing and does not pay attention to the processing of local patches of the image. Moreover, the proposed method suffers from color distortion and loss of detail in real-world scenarios. For example, in Image 10, the texture details of trees far away are significantly lost, and the tree contour cannot be distinguished. In dense hazy scenes, although our method achieves the best results, the color distortion remains significant, as shown in Image 11. In real-world image-dehazing tasks, our method may suffer from these problems. This may be because there are few real images for training and subsequently it is difficult to fully restore the image details. These shortcomings and deficiencies are future research directions.

4.5. Analysis of model parameters and computational cost

We compared the number of model parameters and the computational cost of various methods, as reported in Table 4. Notice that the number of parameters of our model is lower than that of MSBDN, RefineDNet, and PSD; however, the floating-point operations per second (FLOPs) of our model are very large. This is because of using two SwinIR branches; the STL attention heads of each branch are expanded to eight, and the self-attention mechanism has a high computational cost. The expensive computational cost is a shortcoming of our method, and we will continue to study how to reduce the computational cost and improve the dehazing performance of the model in future research.

4.6. Ablation study

First, we validated the effectiveness of the proposed simplified SwinIR method. Using our network architecture , we conducted comparative experiments with the original SwinIR [34] and tested

Table 4
Comparison of the number of the parameter (# Param), and floating point operations (FLOPs) of various dehazing methods. Note: FLOPs are measured on 256 × 256 size images.

Methods	DehazeNet [4]	AODNet [5]	GridDehazeNet [22]	MSBDN [21]	FFANet [53]	RefineDNet [31]	PSD [32]	D4 [60]	SwinTD-Net (Ours)
# Param (M)	0.01	0.002	0.956	31.35	4.456	65.8	33.11	10.7	22.24
FLOPs (GMac)	0.513	0.115	21.49	41.54	288.3	75.41	182.5	2.246	1457

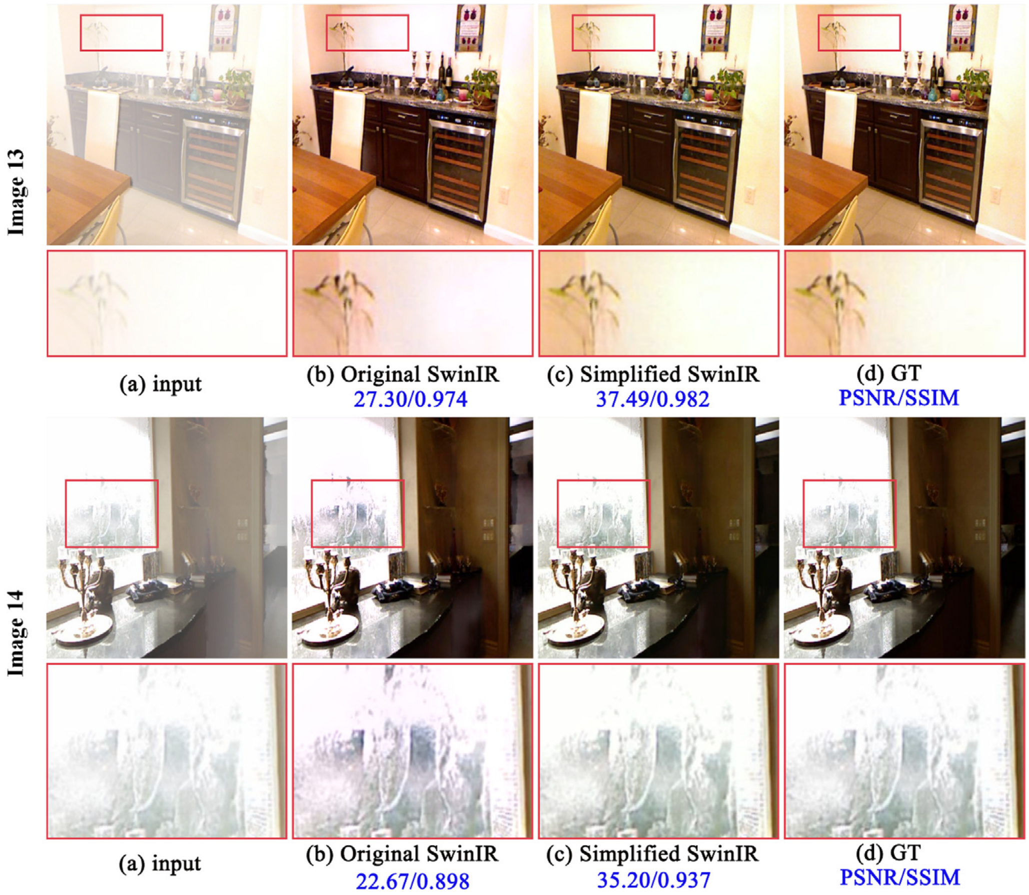


Fig. 17. Comparison of visual results of a SwinIR ablation study on images sampled from SOTS-Indoor. Blue text indicates PSNR/SSIM values. (a) is the hazy input, (b) is the result of using the original SwinIR recovery in our network framework, (c) is the result of using our simplified SwinIR recovery in our network framework, (d) is the ground truth.

Table 5
SwinIR ablation study for SOTS-Indoor dataset. The table shows the PNSR and SSIM scores obtained by testing the SOTS-Indoor dataset with the original SwinIR's SwinTD-Net, and the PNSR and SSIM scores obtained by testing the SOTS-Indoor dataset with our simplified SwinIR's SwinTD-Net.

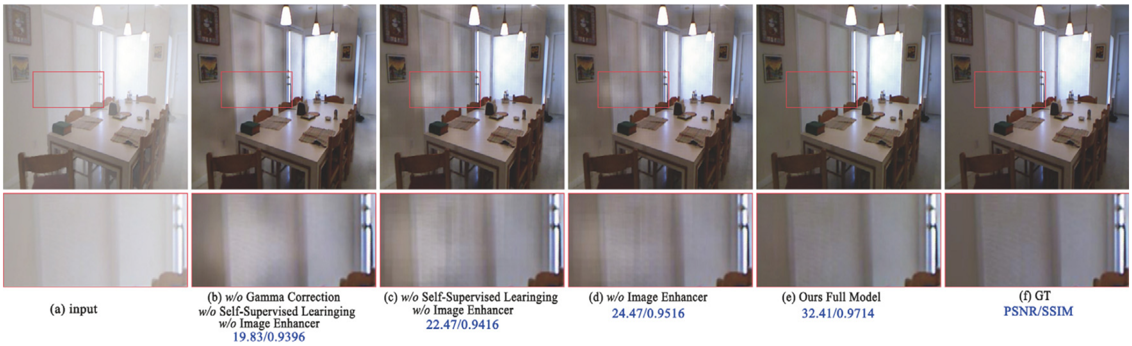
Metrics	SOTS-Indoor	
	PSNR	SSIM
SwinTD-Net with original SwinIR	24.55	0.9601
SwinTD-Net with simplified SwinIR	**36.79**	**0.9679**

them on the SOTS-Indoor dataset. From Fig. 17,we observe that the restoration results using the original SwinIR exhibit color distortion, whereas the restoration results using our simplified SwinIR are closer to GT and attain higher PSNR and SSIM scores. Table 5 reports that the performance of SwinTD-Net with our simplified SwinIR has been greatly improved from a quantitative evaluation standpoint. In particular, the PSNR score has been greatly improved, which strongly verifies the effectiveness of the proposed simplified SwinIR framework.

Then, we investigated the effect of applying gamma correction (and inverse gamma correction) in Stage 1 and self-supervised reconstruction learning and image enhancer (*i.e.*, Generator 2) in Stage 2 on the dehazing performance through the following ablation models:

- *w/o* Gamma Correction: removing the gamma correction and inverse gamma correction in Stage 1;
- *w/o* Self-Supervised Learning: removing the self-supervised reconstruction training in Stage 2; that is, the network is not fine-tuned with physical priors such as DCP and ASM.
- *w/o* Image Enhancer: removing the Generator 2 in Stage 2;

We performed the abovementioned ablation experiments on the SOTS-Indoor dataset. From Fig. 18, observe that the restoration results without gamma correction, self-supervised learning, and image enhancer have some black-and-white blocks. After adding gamma correction, the image becomes smoother. After adding self-supervised learning, the color transition of the image becomes more natural. The result after adding the image enhancer is closer to GT. The PSNR and SSIM scores of the image also illustrate the effectiveness of the proposed module. The quantitative evaluation results of the ablation study conducted using all the images in SOTS-Indoor dataset are reported in Table 6. Notice that, increasing the number of each module can improve



(a) input

(b) *w/o* Gamma Correction *w/o* Self-Supervised Learinging *w/o* Image Enhancer 19.83/0.9396

(c) *w/o* Self-Supervised Learinging *w/o* Image Enhancer 22.47/0.9416

(d) *w/o* Image Enhancer 24.47/0.9516

(e) Ours Full Model 32.41/0.9714

(f) GT PSNR/SSIM

Fig. 18. Comparison of visual results of ablation studies on an image sampled from SOTS-Indoor. Blue text indicates PSNR/SSIM values. (a) is the hazy input, (b) is the recovery result after removing Gamma correction, self-supervised learning and image enhancer in our network framework, (c) is the recovery result after removing self-supervised learning and image enhancer in our network framework, (d) is the recovery result after removing image enhancer in our network framework, (e) is the recovery result of our full network model, (f) is the ground truth.

Table 6
Ablation study for SOTS-Indoor dataset. The table shows the PNSR and SSIM scores for the full model, and the PNSR and SSIM scores after removing some components in the network framework.

Metrics	SOTS-Indoor	
	PSNR	SSIM
w/o Gamma Correction *w/o* Self-Supervised Learning *w/o* Image Enhancer	28.89	**0.9783**
w/o Self-Supervised Learning *w/o* Image Enhancer	30.69	0.9749
w/o Image Enhancer	32.05	0.9676
Ours Full Model	**36.79**	0.9676

Table 7
Ablation study for the real-world hazy dataset. The table shows the PNSR and SSIM score for the full model, and the PNSR and SSIM score after removing some components in the network framework.

Metrics	O-HAZE		Dense-HAZE	
	PSNR	SSIM	PSNR	SSIM
w/o Gamma Correction *w/o* Self-Supervised Learning	16.94	0.6734	14.12	0.4717
w/o Self-Supervised Learning	18.28	0.6864	14.50	0.4782
Ours Full Model	**18.57**	**0.7064**	**14.62**	**0.4821**

the PSNR score but the reduce the SSIM score slightly. Specifically, after adding self-supervised learning, the SSIM score drops approximately to 0.01. We believe that this is mainly because of the synthetic haze. SSIM performs quality evaluation based on structural information [74], whereas the haze in the synthetic dataset is uniform and evenly covers the structural information of the scene. However, the self-supervised learning framework we proposed is to reconstruct the real hazy scene using the physical model, where the real haze is basically not uniform and structural information of the real hazy image is relatively obvious. Therefore, the proposed model is more suitable for real-world scenarios. In addition, the Image Enhancer improves the PSNR score while keeping the SSIM unchanged. This is because the Image Enhancer in the network is a module that enhances scene recovery of dark patches in images. This module has a low impact on the scene structural information; therefore, the SSIM score may remain unchanged. However, enhancing the dark patches of the image can improve the PSNR score. To validate our idea, we performed quantitative ablation studies on hazy real-world image datasets. As reported in Table 7, both gamma correction and self-supervised learning improve the PSNR and SSIM scores, which illustrates the effectiveness of both methods.

4.7. MSDLEA experiments

Because DCP causes a localized darker effect in the image, a similar situation may occur in the method based on DCP, resulting in unclear local image details. The proposed MSDLEA improves the local details from the perspective of the human eye, especially the dark areas in the image, so that the human eye can obtain more image scene information. As shown in Fig. 19, the darker regions in the haze-free images recovered by our method are clearly enhanced by MSDLEA; subsequently, the human observer can perceive more detailed information. MSDLEA is a general image enhancement algorithm that can be applied to various scenarios. We conducted experiments in the underwater image enhancement [87,88] and medical image enhancement fields. Fig. 20 depicts the enhancement results of MSDLEA on the UIEB Dataset [89], which underlines that the information on the dark area of the underwater image has been significantly enhanced. Fig. 21 shows the results of MSDLEA in medical images, where the enhanced results attain more local details, which may be more conducive to medical pathological diagnoses. Note that all the qualitative and quantitative evaluations in our experiments are conducted with respect to images without MSDBLEA enhancement.

5. Conclusions and future work

In this study, we propose a novel single-image dehazing framework. Owing to the characteristic differences between synthetic and real-world hazy images, such as haze distribution and haze concentration, networks that rely heavily on synthetic datasets are prone to overfitting. Therefore, we integrate prior knowledge and deep learning methods effectively, utilizing the physical model, dark channel prior, and Swin Transformer with global feature extraction capabilities so that the network can learn under the guidance of physical priors, which improves the generalization ability of the network. In addition, we propose a more suitable prior input that can be better suited to ASM and DCP. Furthermore, we design an image enhancement algorithm, MSDLEA, that can be effectively used for post-processing operations in image dehazing and other application scenarios to enhance details in the dark parts of images. Extensive comparative experiments illustrate the superiority of the proposed method, particularly in real-world scenarios; training on small-sample datasets also yields good results. Our research validates the necessity of deep learning methods guided by physical priors, particularly in real-world image dehazing tasks.

We will further improve prior knowledge method, aiming to address the shortcomings of DCP and ASM, and study more

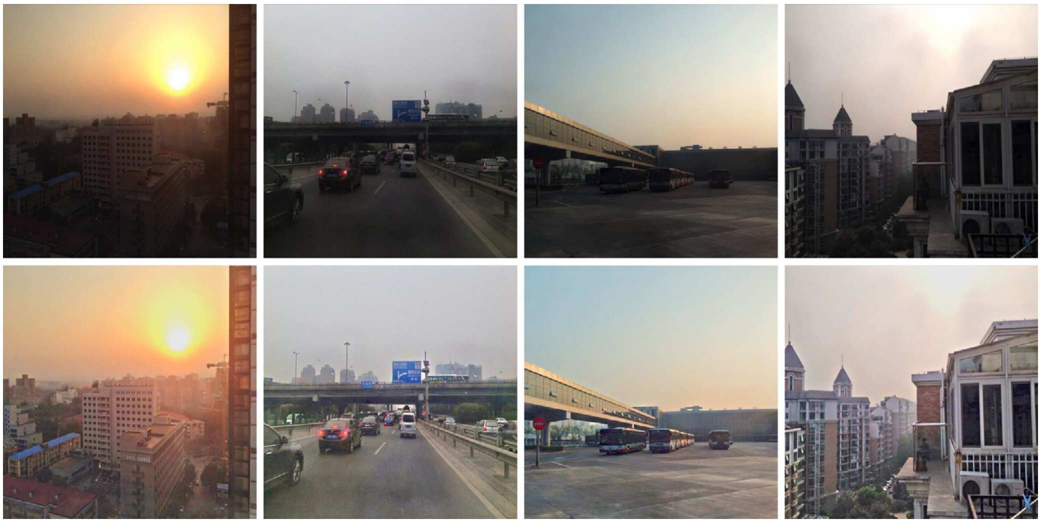


Fig. 19. MSDLEA experimental results. The first row is the haze-free image obtained by our method, and the second row is the MSDLEA enhancement result.

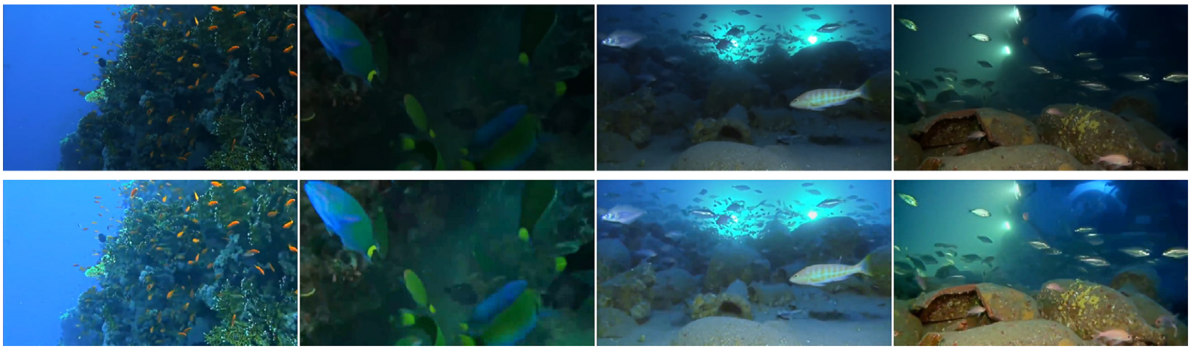


Fig. 20. MSDLEA experimental results for underwater images. The first row is the underwater images, and the second row is the MSDLEA enhancement result.

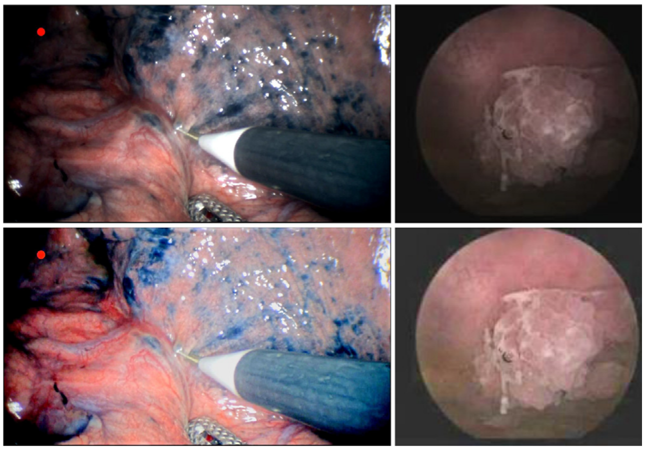


Fig. 21. Results of MSDLEA experiments conducted with medical images. The first row presents the image acquired by the medical device, while the second row presents the MSDLEA enhancement result.

robust physical priors to guide the network to learn features that are more suitable for real-world scenarios. In addition, we hope to further simplify the SwinIR structure, improve the inference speed, and propose a lightweight physics-prior-guided dehazing network.

CRediT authorship contribution statement

Hao Zhou: Conceptualization, Methodology, Software, Validation, Writing. **Zekai Chen:** Software, Data curation. **Yun Liu:** Supervision. **Yongpan Sheng:** Writing – review & editing. **Wenqi Ren:** Writing – review & editing. **Hailing Xiong:** Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have shared the link to the code and model in our paper.

Acknowledgments

This work was supported by the National Fund Cultivation Project of Chizhou University, China (CZ2021GP05), the Science and Technology Research Program of Chongqing Municipal Education Commission, China (Grant No. KJQN202102002), the Fundamental Research Funds for the Central Universities of China (SWU2009107), the Open Fund of Anhui Engineering Laboratory for Intelligent Applications and Security of Industrial Internet, Anhui University of Technology, China (IASII21-06), and the Excellent Scientific Research and Innovation Team of Anhui Colleges, China (2022AH010098). We are grateful for the help of Mr. Zhang Zhuangzhuang (Teemsun(Beijing)Technology Co., Ltd.) in our work.

References

[1] R. Fattal, Single image dehazing, ACM Trans. Graph. 27 (3) (2008) 1–9.

[2] K.M. He, J. Sun, X.O. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2010) 2341–2353.

[3] G.F. Meng, Y. Wang, J.Y. Duan, S.M. Xiang, C.H. Pan, Efficient image dehazing with boundary constraint and contextual regularization, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 617–624.

[4] B. Cai, X.M. Xu, K. Jia, C.M. Qing, D.C. Tao, DehazeNet: An end-to-end system for single image haze removal, IEEE Trans. Image Process. 25 (11) (2016) 5187–5198.

[5] B. Li, X. Peng, Z.Y. Wang, D. Xu, AOD-net: All-in-one dehazing network, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4770–4778.

[6] H. Zhang, V.M. Patel, Densely connected pyramid dehazing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3194–3203.

[7] Y.Y. Qu, Y.Z. Chen, J.Y. Huang, Y. Xie, Enhanced PIX2PIX dehazing network, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vol. 2019-June, 2019, pp. 8160–8168.

[8] H. Zhou, Z.Z. Zhang, Y. Liu, M.Y. Xuan, W.W. Jiang, H.L. Xiong, Single image dehazing algorithm based on modified dark channel prior, IEICE Trans. Inf. Syst. 104 (10) (2021) 1758–1761.

[9] W. Kim, J. You, J. Jeong, Contrast enhancement using histogram equalization based on logarithmic mapping, Opt. Eng. 51 (6) (2012) 067002.

[10] H. Li, W.H. Xie, X.G. Wang, S.S. Liu, Y.Y. Gai, L. Yang, GPU implementation of multi-scale retinex image enhancement algorithm, in: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications, AICCSA, 2016, pp. 1–5.

[11] J.A. Stark, Adaptive image contrast enhancement using generalizations of histogram equalization, IEEE Trans. Image Process. 9 (5) (2000) 889–896.

[12] W.E.K. Middleton, V. Twersky, Vision through the atmosphere, in: Geophysik II/Geophysics II, Springer, 1957, pp. 254–287.

[13] E.J. McCartney, Optics of the atmosphere: scattering by molecules and particles, New York (1976).

[14] S.G. Narasimhan, S.K. Nayar, Vision and the atmosphere, Int. J. Comput. Vis. 48 (3) (2002) 233–254.

[15] G. Sahu, A. Seal, D. Bhattacharjee, M. Nasipuri, P. Brida, O. Krejcar, Trends and prospects of techniques for haze removal from degraded images: A survey, IEEE Trans. Emerg. Top. Comput. Intell. 6 (4) (2022) 762–782.

[16] D. Berman, T. Treibitz, S. Avidan, Non-local image dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1674–1682.

[17] R. Fattal, Dehazing using color-lines, ACM Trans. Graph. 34 (1) (2014) 1–14.

[18] A. Kumar, R.K. Jha, N.K. Nishchal, A multi-exposure fusion framework for contrast enhancement of hazy images employing dynamic stochastic resonance, J. Vis. Commun. Image Represent. 81 (2021) 103376.

[19] A. Kumar, R.K. Jha, N.K. Nishchal, Joint Gamma correction and multi-resolution fusion scheme for enhancing haze degraded images, Opt. Eng. 60 (6) (2021) 063103.

[20] W.Q. Ren, J.S.. Pan, H. Zhang, X.C. Cao, M.H. Yang, Single image dehazing via multi-scale convolutional neural networks with holistic edges, Int. J. Comput. Vis. 128 (1) (2020) 240–259.

[21] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, M.-H. Yang, Multi-scale boosted dehazing network with dense feature fusion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2157–2167.

[22] X.H. Liu, Y.R. Ma, Z.H. Shi, J. Chen, Griddehazenet: Attention-based multi-scale network for image dehazing, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7314–7323.

[23] C. Wang, H.Z. Shen, F. Fan, M.W. Shao, C.S. Yang, J.C. Luo, L.J. Deng, EAA-net: A novel edge assisted attention network for single image dehazing, Knowl.-Based Syst. 228 (2021) 107279.

[24] C.Y. Zheng, J. Zhang, J.N. Hwang, B. Huang, Double-branch dehazing network based on self-calibrated attentional convolution, Knowl.-Based Syst. 240 (2022) 108148.

[25] P.J. Luo, G.Q. Xiao, X.B. Gao, S. Wu, LKD-net: Large kernel convolution network for single image dehazing, 2022, arXiv preprint arXiv:2209.01788.

[26] Y. Liu, L. Wan, H. Fu, J. Qin, L. Zhu, Phase-based memory network for video dehazing, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 5427–5435.

[27] Y. Liu, L. Zhu, S. Pei, H. Fu, J. Qin, Q. Zhang, L. Wan, W. Feng, From synthetic to real: Image dehazing collaborating with unlabeled real data, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 50–58.

[28] G. Sahu, A. Seal, D. Bhattacharjee, R. Frischer, O. Krejcar, A novel parameter adaptive dual channel MSPCNN based single image dehazing for intelligent transportation systems, IEEE Trans. Intell. Transp. Syst. (2022) 1–21.

[29] M. Buckler, S. Jayasuriya, A. Sampson, Reconfiguring the imaging pipeline for computer vision, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 975–984.

[30] H.S. Park, Architectural analysis of a baseline isp pipeline, in: Theory and Applications of Smart Cameras, Springer, 2016, pp. 21–45.

[31] S.Y. Zhao, L. Zhang, Y. Shen, Y.C. Zhou, RefineDNet: A weakly supervised refinement framework for single image dehazing, IEEE Trans. Image Process. 30 (2021) 3391–3404.

[32] Z.Y. Chen, Y.C. Wang, Y. Yang, D. Liu, PSD: Principled synthetic-to-real dehazing guided by physical priors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7180–7189.

[33] Z. Liu, Y.T. Lin, Y. Cao, H. Hu, Y.X. Wei, Z. Zhang, S. Lin, B.N. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.

[34] J.Y. Liang, J.Z. Cao, G.L. Sun, K. Zhang, L. Van Gool, R. Timofte, Swinir: Image restoration using swin transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.

[35] K.M. He, X.Y. Zhang, S.Q. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[36] R.T. Tan, Visibility in bad weather from a single image, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.

[37] J.P. Tarel, N. Hautiere, Fast visibility restoration from a single color or gray level image, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 2201–2208.

[38] Q.S. Zhu, J.M. Mai, L. Shao, A fast single image haze removal algorithm using color attenuation prior, IEEE Trans. Image Process. 24 (11) (2015) 3522–3533.

[39] G. Sahu, A. Seal, O. Krejcar, A. Yazidi, Single image dehazing using a new color channel, J. Vis. Commun. Image Represent. 74 (2021) 103008.

[40] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[41] W. Liao, Q. Zhang, B. Yuan, G.Q. Zhang, J. Lu, Heterogeneous multidomain recommender system through adversarial learning, IEEE Trans. Neural Netw. Learn. Syst. (2022) 1–13.

[42] Y. Zhang, M.J. Wu, W. Miao, L. Huang, J. Lu, Bi-layer network analytics: A methodology for characterizing emerging general-purpose technologies, J. Informetr. 15 (4) (2021) 101202.

[43] J.Y. Liao, J.J. Zhou, Y.L. Song, B.L. Liu, J. Lu, D.Y. Jin, Optical fingerprint classification of single upconversion nanoparticles by deep learning, J. Phys. Chem. Lett. 12 (41) (2021) 10242–10248.

[44] W.J. Chang, Q. Zhang, C. Fu, W.Y. Liu, G.Q. Zhang, J. Lu, A cross-domain recommender system through information transfer for medical diagnosis, Decis. Support Syst. 143 (2021) 113489.

[45] Q. Zhang, J. Lu, Y.C. Jin, Artificial intelligence in recommender systems, Complex Intell. Syst. 7 (1) (2021) 439–457.

[46] F. Liu, G.Q. Zhang, J. Lu, Multisource heterogeneous unsupervised domain adaptation via fuzzy relation neural networks, IEEE Trans. Fuzzy Syst. 29 (11) (2020) 3308–3322.

[47] H. Yu, J. Lu, G.Q. Zhang, Topology learning-based fuzzy random neural network for streaming data regression, IEEE Trans. Fuzzy Syst. 30 (2) (2022) 412–425.

[48] R.X. Liang, Q. Zhang, J. Lu, G.Q. Zhang, J.Q. Wang, A cross-domain group recommender system with a generalized aggregation strategy, in: Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference, FLINS 2020, World Scientific, 2020, pp. 455–462.

[49] L. Zhong, Z. Fang, F. Liu, B. Yuan, G.Q. Zhang, J. Lu, Bridging the theoretical bound and deep algorithms for open set domain adaptation, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–15.

[50] R.P. Yin, K. Li, G.Q. Zhang, J. Lu, A deeper graph neural network for recommender systems, Knowl.-Based Syst. 185 (2019) 105020.

[51] J. Lu, A.J. Liu, F. Dong, F. Gu, J. Gama, G.Q. Zhang, Learning under concept drift: A review, IEEE Trans. Knowl. Data Eng. 31 (12) (2018) 2346–2363.

[52] T.Z. Si, F.Z. He, Z. Zhang, Y.S. Duan, Hybrid contrastive learning for unsupervised person re-identification, IEEE Trans. Multimed. (2022).

[53] X. Qin, Z.L. Wang, Y.C. Bai, X.D. Xie, H.Z. Jia, FFA-net: Feature fusion attention network for single image dehazing, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 07, 2020, pp. 11908–11915.

[54] S.D. Zhang, F.Z. He, DRCDN: learning deep residual convolutional dehazing networks, Vis. Comput. 36 (9) (2020) 1797–1808.

[55] H.Y. Wu, Y.Y. Qu, S.H. Lin, J. Zhou, R. Qiao, Z.Z. Zhang, Y. Xie, L.Z. Ma, Contrastive learning for compact single image dehazing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10551–10560.

[56] G. Sahu, A. Seal, A. Yazidi, O. Krejcar, A dual-channel dehaze-net for single image dehazing in visual internet of things using PYNQ-Z2 board, IEEE Trans. Autom. Sci. Eng. (2022) 1–15.

[57] D. Engin, A. Genç, H. Kemal Ekenel, Cycle-dehaze: Enhanced cyclegan for single image dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 825–833.

[58] X.T. Yang, Z. Xu, J.B. Luo, Towards perceptual image dehazing by physics-based disentanglement and adversarial training, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, No. 1, 2018.

[59] B.Y. Li, Y.B. Gou, S.H. Gu, J.Z. Liu, J.T.Y. Zhou, X. Peng, You only look yourself: Unsupervised and untrained single image dehazing neural network, Int. J. Comput. Vis. 129 (5) (2021) 1754–1767.

[60] Y. Yang, C.Y. Wang, R.S. Liu, L. Zhang, X.J. Guo, D.C. Tao, Self-augmented unpaired image dehazing via density and depth decomposition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2037–2046.

[61] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017.

[62] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X.H. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16 × 16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.

[63] W.H. Wang, E. Xie, X. Li, D.P. Fan, K.T. Song, D. Liang, T. Lu, P. Luo, L. Shao, Pyramid vision transformer: A versatile backbone for dense prediction without convolutions, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 568–578.

[64] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: European Conference on Computer Vision, Springer, 2020, pp. 213–229.

[65] W. Tang, F.Z. He, Y. Liu, YDTR: infrared and visible image fusion via y-shape dynamic transformer, IEEE Trans. Multimed. (2022).

[66] Z.Q. Zhu, X.Y. He, G.Q. Qi, Y.Y. Li, B. Cong, Y. Liu, Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI, Inf. Fusion 91 (2023) 376–387.

[67] H. Chen, Y.H. Wang, T.Y. Guo, C. Xu, Y.P. Deng, et al., Pre-trained image processing transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 12299–12310.

[68] D. Zhao, J. Li, H.Y. Li, L. Xu, Hybrid local-global transformer for image dehazing, 2021, arXiv preprint arXiv:2109.07100.

[69] J.M.J. Valanarasu, R. Yasarla, V.M. Patel, Transweather: Transformer-based restoration of images degraded by adverse weather conditions, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2353–2363.

[70] Y.D. Song, Z.Q. He, H. Qian, X. Du, Vision transformers for single image dehazing, 2022, arXiv preprint arXiv:2204.03883.

[71] C.L. Guo, Q.X. Yan, S. Anwar, R.M. Cong, W.Q. Ren, C.Y. Li, Image dehazing transformer with transmission-aware 3D position embedding, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5812–5820.

[72] Y.Z. Wang, H.B. Huang, Q. Xu, J.M. Liu, Y.Q. Liu, J. Wang, Practical deep raw image denoising on mobile devices, in: European Conference on Computer Vision, Springer, 2020, pp. 1–16.

[73] P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, Two deterministic half-quadratic regularization algorithms for computed imaging, in: Proceedings of 1st International Conference on Image Processing, Vol. 2, IEEE, 1994, pp. 168–172.

[74] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[75] C. Li, J. Guo, C. Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer, IEEE Signal Process. Lett. 25 (3) (2018) 323–327.

[76] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Fast and accurate image super-resolution with deep laplacian pyramid networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (11) (2018) 2599–2613.

[77] J.Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.

[78] Z.G. Li, J.H. Zheng, Z.J. Zhu, W. Yao, S.Q. Wu, Weighted guided image filtering, IEEE Trans. Image Process. 24 (1) (2014) 120–129.

[79] B.Y. Li, W.Q. Ren, D.P. Fu, D.C. Tao, D. Feng, W.J. Zeng, Z.Y. Wang, Reside: A benchmark for single image dehazing, IEEE Trans. Image Process. 28 (1) (2018) 492–505.

[80] C.O. Ancuti, C. Ancuti, R. Timofte, C. De Vleeschouwer, 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, 2018, pp. 754–762.

[81] C.O. Ancuti, C. Ancuti, R. Timofte, 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, 2020, pp. 444–445.

[82] C.O. Ancuti, C. Ancuti, R. Timofte, et al., NTIRE 2020 challenge on NonHomogeneous dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE CVPR 2020, 2020.

[83] C.O. Ancuti, C. Ancuti, M. Sbert, R. Timofte, Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images, in: 2019 IEEE International Conference on Image Processing, ICIP, IEEE, 2019, pp. 1014–1018.

[84] C.O. Ancuti, C. Ancuti, R. Timofte, L.V. Gool, L. Zhang, M.H. Yang, NTIRE 2019 image dehazing challenge report, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE CVPR 2019, 2019.

[85] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, 2017, arXiv preprint arXiv:1711.05101.

[86] J.P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, D. Gruyer, Vision enhancement in homogeneous and heterogeneous fog, IEEE Intell. Transp. Syst. Mag. 4 (2) (2012) 6–20.

[87] Y. Wang, H. Liu, L.P. Chau, Single underwater image restoration using adaptive attenuation-curve prior, IEEE Trans. Circuits Syst. I. Regul. Pap. 65 (3) (2017) 992–1002.

[88] L.T. Peng, C.L. Zhu, L.H. Bian, U-shape transformer for underwater image enhancement, 2021, arXiv preprint arXiv:2111.11843.

[89] C.Y. Li, C.L. Guo, W.Q. Ren, R.M. Cong, J.H. Hou, S. Kwong, D.C. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Trans. Image Process. 29 (2019) 4376–4389.