

MFID-Net: Multi-scaled feature-fused image dehazing via dynamic weights[☆]

Yifan Liu ^{a,c}, Jincai Chen ^{a,b,c,*}, Ping Lu ^{a,b,c}, Chuanbo Zhu ^{a,c}, Yugen Jian ^{a,c}, Chao Sun ^{a,c}, Han Liang ^{a,c}

^a Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China

^b School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China

^c Key Laboratory of Information Storage System, Engineering Research Center of Data Storage Systems and Technology, Ministry of Education of China, Huazhong University of Science and Technology, Wuhan, 430074, People's Republic of China

ARTICLE INFO

MSC:

68U10
68T45
68T05
68T10

Keywords:

Image dehazing
Feature fusion
Dynamic convolution
Dynamic activation

ABSTRACT

Methods based on convolutional neural networks have achieved excellent performance in the image dehazing task. Unfortunately, most of the dehazing methods that exist suffer from loss of detail in the convolution and activation operations and failure to consider the effects of superimposing different intensities of haze, such as under-exposed and over-exposed images. To address these issues, we propose a dynamic dehazing convolution (DDC) based on attentional weight calculation and dynamic weight fusion and a dynamic dehazing activation (DDA) based on the input global context encoding function to address the problem of detail loss. And we propose a multi-scaled feature-fused image dehazing network (MFID-Net) based on DDC and DDA to address the effects of haze superposition. We also design a loss function based on the physical model with dynamic weights. Extensive experimental results demonstrate that the proposed MFID-Net performs favorably against the state-of-the-art algorithms on the hazy dataset while improving further on hazy images with large differences in haze concentration, and producing satisfactory dehazing results. The code is available at <https://github.com/awhitewhale/MFID-Net>.

1. Introduction

In daytime, light scattering occurs in foggy, hazy, and other environments. This scattering phenomenon can cause low image contrast and loss of detail problems [1]. In reality, the hazy environment is complex and changeable. The large color difference contrast of the environment leads to the non-uniform scattering of light with different wavelengths scattered by haze. Non-uniform scattering can lead to the failure of some image dehazing methods based on the premise of uniform scattering or relying on hand-crafted priors. This failure will lead to a reduction in the accuracy of subsequent vision processes such as object detection [2].

Traditional image dehazing methods rely on the atmospheric scattering model [3], which can be formulated as:

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

where $I(x)$ is the hazy image to be processed, $J(x)$ is the haze-free image to be recovered, $t(x)$ is the transmission map and A is the global atmospheric light value. The methods based on the atmospheric scattering model mainly consider the effect of light degradation on the image in the existing atmospheric scattering model [4,5]. However,

for images with high haze concentration, there are usually some noise and color-cast phenomena in the dehazed results [6]. This is caused by an inaccurate assessment of the transmittance in the atmospheric scattering model. Furthermore, a hazy image is not considered the fact that when the haze concentration is high, scattering caused by dust particles in the air can introduce some implied noise into the image during the image processing [7]. This results in the image details being restored after dehazing while the noise is amplified, making the dehazed results subjectively poor [8]. The use of convolutional neural networks (CNNs) for single image dehazing is also emerging as a result of the significant advances made in CNN for tasks such as object detection and semantic segmentation [9]. CNN-based dehazing methods are partly based on the atmospheric scattering model and use deep learning methods to estimate the intermediate parameters in the atmospheric scattering model. This restores a clear, haze-free image and achieves the goal of image dehazing [6,10,11]. The other part does not rely on the physical model but uses the powerful learning capabilities of deep learning to learn the mapping relationships between hazy and clear images. Single image dehazing is performed directly end-to-end [12–15]. However, static CNN is not well implemented for dehazing in the

[☆] This paper was recommended for publication by Jing Wu.

* Corresponding author.

E-mail address: jchen@hust.edu.cn (J. Chen).

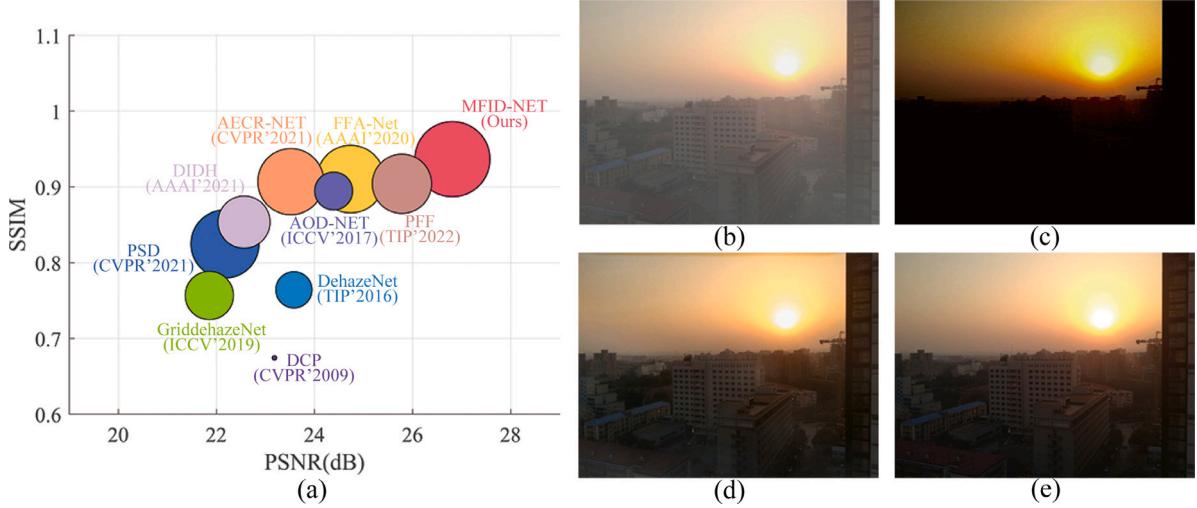


Fig. 1. (a) Comparison of the proposed method MFID-Net with the state-of-the-art dehazing methods on the I/OST datasets and the under-exposed and over-exposed images (UOI). (c) PSD, (d) PFF, and (e) MFID-Net (ours) are dehazing examples of hazy images with large differences in haze concentration(b). MFID-Net can match the performance of the other methods on the dehazing dataset but has further performance improvements on the UOI.

complex real world. This happens because of the tendency for excessive dark and bright contrast in the real world with non-uniform scattering. If the image is color overspilled or has a lot of underexposure and overexposure, using a static CNN can make for a poor image. For example, as shown in Fig. 1(a). The horizontal and vertical coordinates are the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) value of the dehazing method on the under-exposed and over-exposed images (UOI) dataset, respectively. The size of the bubble is the PSNR value of the method on the I/OST dataset [16]. The color of each bubble corresponds to its method of the same color. The dehazing effect of GriddehazeNet [17], DCP [18], DehazeNet [4], and PSD (Fig. 1(c)) [19] on the UOI almost failed. In the meantime, AOD-NET [20], AECCR-NET [21], FFA-Net [22], PFF (Fig. 1(d)) [22], and DIDH [23] on the under-exposed and over-exposed images show a greater degradation in performance.

The features extracted for different haze density regions of a single haze image are also different [24]. Excessive differences between different haze density regions can lead to loss of details during the convolution and activation operations. Most of the existing methods compensate for this deficiency by stacking convolution kernels to expand the receptive field [4,17]. However, this compensation is still limited, causing the problems of over-dehazing and introducing new noise. In addition, both physical model-based and end-to-end dehazing methods suffer from inefficient long-range feature propagation, and these methods ignore the effects of the haze superposition of different intensities.

To address these problems, we propose the multi-scaled feature-fused image dehazing network (MFID-Net). Specifically, to solve the problem of detail loss in convolution and activation operations, we propose the dynamic dehazing convolution (DDC) based on attentional weight calculation and dynamic weight fusion and the dynamic dehazing activation (DDA) based on the input global context encoding function. To solve the problem of the impact caused by the superposition of different-intensity haze, we propose the multi-scaled feature-fused architecture, which can decompose haze of different intensities into different features.

MFID-Net decomposes the input haze image into three scaled feature routes, namely, the feature distillation route for visible haze image estimation, the feature distillation route for adaptive potential haze-free image estimation with three-color channels, and the dynamic feature distillation route for potential haze image estimation. In the first route, the input image is downsampled and upsampled several times to extract

the high-level visible haze image features. However, traditional upsampling and downsampling cause the loss of image details, especially haze details. We solve this problem in the second route and the third route. In the second route, the input image is decomposed in three-color channels. Each channel image is passed through a potential haze-free image estimation module constructed by including a DDC and a DDA. This adaptive feature largely compensates for the loss of details caused by the upsampling and downsampling processes. In the third route, the input image is passed through a feature extraction module similar to that in the first route, but instead of a pooling operation at the end, the information loss due to pooling is weakened by DDC and DDA. And the estimated features of the potential haze image are output. After that, bilinear interpolation is applied between the visible haze features and the haze-free features of the three-color channels separately to predict the strong haze features. The strong haze features and potential haze features of the three channels are fused with the feature fusion machine to output the summed feature maps. Finally, the original haze image is constrained by the summed feature map, and its separation loss is calculated to obtain the haze-free image after dehazing. In addition, we propose a dynamic weight loss function based on the physics model of the atmospheric scattering principle to improve the efficiency of backpropagation in regions with different haze densities.

Thanks to the design of three scale feature routes, MFID-Net can not only dehaze well under daytime but also achieve excellent performance on under-exposed and over-exposed hazy images. Our contributions are summarized as follows:

- We propose the multi-scaled feature-fused image dehazing network called MFID-Net, which extracts information from different color channels and separates multi-scale haze concentrations to efficiently generate high-quality haze-free images from the input hazy images.

- Our proposed MFID-Net uses proposed dynamic dehazing convolution based on attentional weight calculation and dynamic weight fusion and proposed dynamic dehazing activation based on the input global context encoding function to improve image dehazing with excessive differences in light scattering degree.

- MFID-Net achieves the state-of-the-art performance on dehazing datasets (ITS, OST, SOTS [16], I-HAZE [25], and O-HAZE [26]). And MFID-Net has even better performance on under-exposed and over-exposed hazy images.

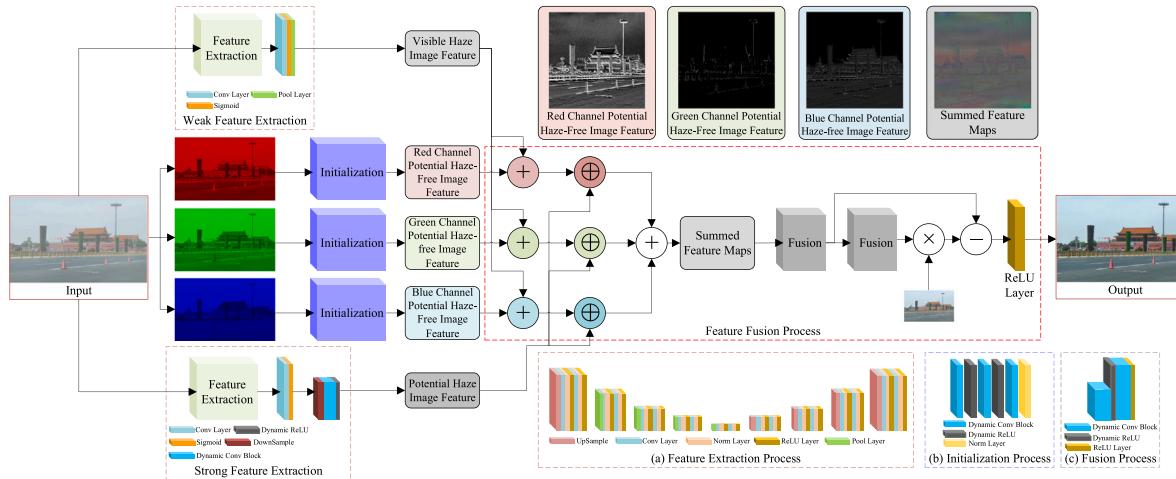


Fig. 2. The architecture of MFID-Net. \oplus is the bilinear interpolation operator mark. The overall framework consists of three routes, the feature distillation route for visible haze image estimation, the feature distillation route for adaptive potential haze-free image estimation with three-color channels, and the dynamic feature distillation route for potential haze image estimation. Three routes that aim to recover clear images from haze are extracted from each of the three different scales, and then multiple feature fusions are performed on each of them to promote better convergence. More details will be presented in Section 3.1.

2. Related work

2.1. Single image dehazing

In recent years, many researchers have been using deep learning to tackle and solve problems in the field of computer vision due to its powerful feature learning capabilities [27,28]. Wu et al. [21] proposed a dehazing network with autoencoder and contrastive regularization named AEGR-NET. This dehazing network uses a contrastive regularization method based on contrast learning, using information from hazy and clear images as samples. Liu et al. [29] proposed indirect domain shift, a network that adds explicit constraints to the deep CNN model to guide the recovery process. In contrast to direct learning, this mechanism shifts and narrows the candidate regions for estimating the output through multiple confidence neighborhoods. Chen et al. [19] proposed principled synthetic-to-real dehazing (PSD). The dehazing network starts with a backbone of dehazing models pretrained with synthetic data and fine-tunes the models in an unsupervised manner using real hazy images. The network allows most existing dehazing models to be used as its backbone. The combination of multiple physical priors significantly improves the dehazing effect. Yi et al. [30] proposed the two-step dehazing network, a two-step structure with intra-domain and inter-domain adaptive constraints, which subdivides the distribution in the synthetic domain into subsets and mines the optimal subset by loss-based supervision. Zhang et al. [31] argued that a hazed map should be divided into a high-frequency part and a low-frequency part. Based on this, they proposed a dual-path recurrent network (DPRN). Two branches of the DPRN can benefit each other by alternately recovering the essential content and image details, thus alleviating the color distortion problem during the dehazing process. However, [19,21,29–31] all fail to consider the effect of non-uniform scattering of different wavelengths of light by haze scattering. Qin et al. [22] proposed a feature fusion attention network (FFA-Net) for dehazing. FFA-Net consists of a novel feature attention module, a block structure consisting of local residual learning and feature attention, and a feature fusion structure attention-based at different levels. Chang et al. [32] proposed a dual-branch network named adaptive dehazing network, which uses a density-aware data enhancement method that generates synthetic haze samples based on the haze density level in the target domain. Bai et al. [33] proposed a deep pre-dehazer with a progressive feature fusion module (PFF). The PFF uses a deep pre-dehazer, a progressive feature fusion module, and an image restoration module to help with haze removal. Zhao et al. [34] proposed a hybrid

local-global vision transformer, a dehazing network consisting of a shared encoder and three decoders for reflection prediction, colorings prediction, and haze-free image generation. However, these methods do not consider large chromatic contrasts in real environments, excessively dark and bright contrasts, color spillover, and large amounts of underexposure and overexposure. These harsh environments will result in a less robust dehazing network in the above-mentioned light non-uniformity scattering situations.

2.2. Dynamic convolution

Dynamic convolution is a type of convolution in which the convolution parameters are adaptively adjusted according to the input image. Dynamic convolution is a special kind of adaptive filter. Wu et al. [35] and Wang et al. [36] proposed neural network layers that perform bootstrap filtering and non-local averaging within the CNN, respectively. Su et al. [37] proposed pixel-adaptive convolution, which predicts different kernels at different pixel locations and adjusts the standard spatially invariant convolution filter at each pixel by multiplying each pixel by a spatially varying filter. Chen et al. [38] proposed a dynamic convolution that dynamically aggregates multiple parallel convolution kernels according to their input dependencies. This dynamic convolution increases model complexity without increasing the depth or width of the network. However, we note that these dynamic convolutions are mainly used in image processing areas such as target recognition and semantic segmentation. Dynamic convolution has been used sparingly in the direction of image dehazing.

2.3. Dynamic activation function

Rectified linear unit (ReLU) is an important milestone in deep learning, simple and powerful, and can greatly improve the performance of neural networks, such as ResNet [39], MobileNet [40], and ShuffleNet [41]. There are also many improved versions of ReLU [42], such as Leaky ReLU [43] and PReLU [44]. And the final parameters of these improved and original versions are fixed, which may lose image details. To solve this problem, dynamic activation functions that adjust the ReLU parameters according to the input features are proposed. Chen et al. [45] proposed DY-ReLU, which is a dynamic segmented linear function with input-dependent parameters that increases neither the depth nor the width of the network but can effectively increase the model capability. Ma et al. [46] proposed ACON, which can adaptively choose whether to activate neurons or not. However, ACON only flexibly selects whether a neuron is activated or not but cannot fundamentally solve the problem of detail loss.

2.4. Multi-scale feature fusion

CNN extracts features of the target by sampling layer by layer [47–49]. High layer network has a strong ability to represent semantic information but a weak ability to represent geometric information. The lower layer network has a smaller perceptual field and a strong ability to characterize geometric detail information but a weak ability to characterize semantic information. Therefore, feature fusion of different scales of features extracted from high-level and low-level networks is helpful for haze feature extraction. Zhang et al. [47] proposed a dehazing network combining multi-scale hierarchical feature fusion and hybrid convolutional attention. The hybrid convolutional attention mechanism proposed by this network can reduce feature redundancy and learn compact and effective internal representations. It can further help the model estimate images with sharper texture details and more vivid colors. Dong et al. [48] proposed a multi-scale boosted dehazing network (MSBDN). MSBDN is based on the U-Net architecture and uses a dense feature fusion module to simultaneously compensate for the missing spatial information in high-resolution features. However, MSBDN only utilizes different resolution scales and does not take into account the physical laws of haze. Wang et al. [49] proposed a multi-scale feature fusion dehazing network with dense connection (MSF²DN). MSF²DN skips thin haze and clear regions by focusing most of the network's weight on regions that are difficult to dehaze. The network extracts haze features by fusing features from both the encoder and decoder stages. However, MSF²DN only considers the above two scales and does not consider the fusion of different color channel scales and visible and potential haze feature scales.

3. Proposed method

In this section, we first describe the architecture of MFID-Net in detail. The architecture of MFID-Net is shown in Fig. 2. Then, we propose the dynamic dehazing convolution (DDC) and the dynamic dehazing activation (DDA) function. Finally, we design a loss function based on the physical model with dynamic weights.

3.1. Network architecture

For recovering hazy images with large differences in haze concentration, we propose the multi-scaled feature-fused image dehazing network architecture (MFID-Net). The architecture of MFID-Net is shown in Fig. 2. Hazy image is generated through the three routes mentioned in Section 1, namely the visible haze image feature, the potential haze-free image feature for the three-color channels, and the potential haze image feature.

First, the first and third routes use a feature extraction block to represent the input hazy image as a mixture of visible haze and potential haze features. The feature extraction block passes through one up-dimensioning operation and four down-dimensioning operations. The up-dimensioning operation includes an upsampling operation, a convolution filter, a batch normalization layer, and a ReLU layer. The four down-dimensioning operations replace the previous up-dimensioning operation with a max-pooling operation to aggregate the features. To reconstruct the haze-free image, four more up-dimensioning operations are used. In the first route, weaker extraction is made at different scales along with the convolution operation. The extracted weak feature is therefore passed through a convolution filter, a sigmoid function layer, and a max-pooling layer to generate a visible haze image feature. In cases where the difference in haze concentration is too large, the area affected by haze may extend beyond the receptive field of the convolution kernel. Therefore, the receptive field needs to be dynamically adjusted according to the distribution of haze. The initialization operation consists of two consecutive DDCs and DDAs, one DDC, and one batch normalization layer to generate the three-channel potential

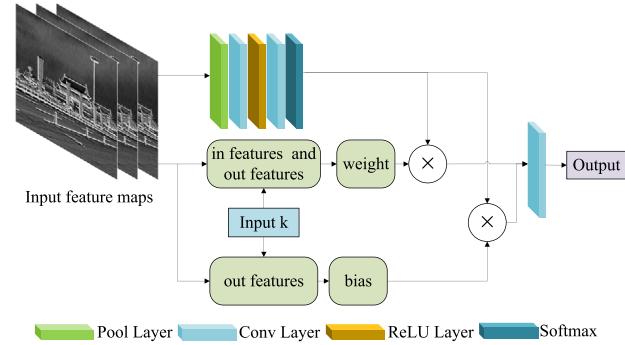


Fig. 3. The architecture of the proposed DDC.

haze-free image feature. The architectural details of DDC and DDA are described in Sections 3.2 and 3.3 respectively.

The visible haze image feature is then summed with each of the three-channel potential haze-free image features. The summed three-channel features are then bilinearly interpolated with the potential haze-free image feature. To maintain the boundary properties of the objects in the image, the summed feature map is then passed through two fusion blocks, consisting of a DDC, a DDA, a DDC, and a ReLU layer. The fused feature FF_1 generated by the first fusion block, the fused feature FF_2 generated by the second fusion block, and the original hazy image \mathbb{I} are then subjected to the following operations to produce the final hazy-free image \mathbb{O} :

$$\mathbb{O} = PReLU(FF_2 \times \mathbb{I} - FF_1 + 1), \quad (2)$$

where $PReLU(\cdot)$ is the activation operation of the parametric rectified linear unit (PReLU).

3.2. Dynamic dehazing convolution

To effectively address the effects of non-uniform scattering of different wavelengths of light by haze scattering, a dynamic dehazing convolution (DDC) based on attentional weight calculation and dynamic weight fusion has been proposed. Fig. 3 shows the architecture of the DDC. Different from dynamic convolution, DDC dynamically aggregates multiple convolution kernels based on the attention input for each different haze density region separately. For each channel H_c , it is passed through an average pooling layer, a convolution layer, an activation function layer, a convolution layer, and a softmax layer. This series of operations is called the direct feature extraction part. In addition, the DDC has a dynamic feature extraction part. The dynamic feature extraction part is determined by the dynamic parameter k controlling the dimensionality and the input and output features determining the weight, and the output feature determining the bias. The directly extracted feature is fused with the weight and bias, respectively, and the final output is processed by one convolution to control the output dimension.

In the direct feature extraction part, the H_c is first fed into a two-dimensional adaptive averaging pool and the output is an $H \times H$ matrix. This matrix is passed through a convolution process as:

$$O_{f1}(N_b, C_o) = \sum_{k=0}^{H_c-1} \mathbb{W}(H_c \times \mathbb{R} + 1, K) \otimes \mathbb{I}(N_b, K), \quad (3)$$

where $O_{f1}(\cdot, \cdot)$ is the output of the first convolution, $\mathbb{W}(\cdot, \cdot)$ denotes the weight, $\mathbb{I}(\cdot, \cdot)$ denotes the input, N_b is the batch size, C_o is the number of output channels, \mathbb{R} controls the ratio of the number of input channels to the number of output channels, K is a positive integer similar to that in deep convolution, and \otimes is a valid two-dimensional cross-correlation operator. The batch size is treated as a dimensional variable for group convolution. Because the weights of the group convolution

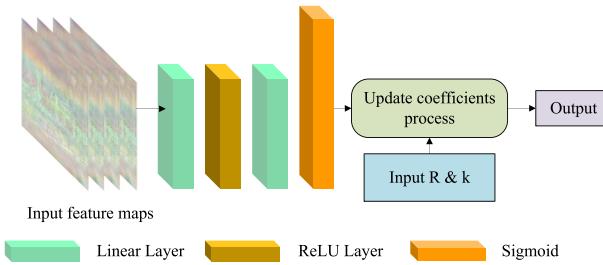


Fig. 4. The architecture of the proposed DDA.

are different, the weights of the DDC are also different. The output of the above process, $O_{f1}(N_b, C_o)$, is then fed to the DDA and processed by convolution:

$$O_{f2}(N_b, C_o) = \sum_{k=0}^{H_c \times R} \mathbb{W}(K, 1) \otimes \mathbb{I}(N_b, 1) + \mathbb{B}(K), \quad (4)$$

where $O_{f2}(\cdot, \cdot)$ is the output of the second convolution and $\mathbb{B}(\cdot, \cdot)$ denotes the bias. The above process yields the output O_{f2} of the dehazing part of the DDC module.

In the direct mapping part, H_c and O_{f2} are first reconstructed into a dimension C_{c+f2} for group convolution. The weights of the DDC \mathbb{W}_{Dy} are determined by the product of H_c and O_{f2} , generating N_b convolution parameters:

$$\mathbb{W}_{Dy} = \sum_{k=0}^{H_c \times R} H_c \otimes O_{f2}(N_b, C_o). \quad (5)$$

The bias of the dynamic dehazing convolution \mathbb{B}_{Dy} is determined by the product of O_{f2} and $\mathbb{B}(K)$:

$$\mathbb{B}_{Dy} = \sum_{k=0}^{H_c \times R} O_{f2}(N_b, C_o) \otimes \mathbb{B}(K). \quad (6)$$

Finally, the convolution operation will be performed with the above dynamic weight and dynamic bias as C_{c+f2} of the weights and bias.

3.3. Dynamic dehazing activation

To solve the problem of noise being amplified while details are restored after dehazing, we design a dynamic dehazing activation (DDA) instead of a conventional activation layer to amplify important features while suppressing noise. Fig. 4 shows the architecture of the DDA.

The parameters of the DDA are determined by the input image. First, the feature maps of RGB three-channel inputs are given for the input image H . For each channel H_c (where $c = r, g, b$) a linear transformation is applied. The input and output features are $I_F \times (I_F // R)$, with R being the dimension to be controlled, $//$ being the integer division sign. The output features λ are subjected to ReLU and then a linear transformation is applied to obtain the second output features μ . The number of output features μ is $2k$, where k is the dynamic parameter of the control dimension and $k = I_F / (2R)$. Next, the feature map extracted by the above process is constructed as a tensor Δ of $I_F \times 2k$. Then the feature parameter λ and μ are constructed for the tensor Δ . λ and μ are calculated as:

$$\lambda = \alpha \times k + \beta \times k, \quad (7)$$

$$\mu = \alpha + \beta \times (2k - 1), \quad (8)$$

where α and β are constants, used to control the range of λ and μ . To update the weight coefficients and prevent the backward transfer of the multiplication after the derivation of the function, the second output feature is fed to the update coefficients process stage based on the input global context encoding function. Specifically, the second output

feature μ is passed through the sigmoid function so that its interval lies at $(-1, 1)$ and has a mean value of 0. This DDC method has a pixel-level or area-level feature processing effect. In our example, the final output of the dynamic activation, Y_R , can be expressed as:

$$Y_R = \max(\Delta \times \lambda + \mu, \Delta \times \mu + \lambda). \quad (9)$$

That is, for a normal hazy image, the difference between the values of $\Delta \times \lambda + \mu$ and $\Delta \times \mu + \lambda$ is not significant. However, for an over-exposed (or under-exposed) image, the former value will be much larger (or smaller) than the latter, and the DDA will optimize the haze feature in the over-exposed (or under-exposed) image to ensure balance. More specifically, Eq. (9) can be interpreted as when a region of the haze feature map appears to have a much larger number of pixels in the bright region than in the dark region, then Y_R is a linear transformation of Δ with λ as the weight and μ as the bias. Conversely, when a region of the haze feature map appears to have far fewer pixels in the bright region than in the dark region, then Y_R is a linear transformation of Δ with μ as the weight and λ as the bias.

3.4. Loss function with dynamic weights

To optimize the loss function to increase the efficiency of backpropagation in different haze density regions, the L1 parametric loss was suitably improved. Firstly, for images of size $n \times m$, the L1 parametric loss \mathcal{L}_1 is calculated as:

$$\mathcal{L}_1(i, j) = \frac{1}{nm} \sum_{j=0}^{m-1} \sum_{i=0}^{n-1} [J(i, j) - I(i, j)]^2, \quad (10)$$

where J is the dehazed image, I is the hazy image, and (i, j) are the coordinates of the pixels. To pixelate the hazy image based on the atmospheric light scattering model, we can rewrite Eq. (1) as follows:

$$I(i, j) = J(i, j)t(i, j) + A(1 - t(i, j)). \quad (11)$$

Based on Eqs. (10) and (11), we propose a loss function \mathcal{L}_p with pixel-based:

$$\mathcal{L}_p(i, j) = \frac{1}{nm} \sum_{j=0}^{m-1} \sum_{i=0}^{n-1} [(J(i, j) - A) \times (1 - t(i, j))]^2, \quad (12)$$

where $t(i, j)$ is defined as:

$$t(i, j) = e^{-\eta d(i, j)}, \quad (13)$$

where η is the scattering coefficient of the atmosphere and $d(i, j)$ is the distance between the object corresponding to that pixel and the camera.

In addition, to better preserve the concentration characteristics of haze in different regions with the color distribution of the image, we used the heavy haze multi-scale structural similarity (HHMS-SSIM) proposed by Chang et al. [32] which is defined as:

$$\mathcal{L}_H(i, j) = \mathcal{L}_1(I(i, j), J(i, j)) + \gamma \mathcal{L}_M(I(i, j), J(i, j)), \quad (14)$$

where \mathcal{L}_H is the loss function of HHMS-SSIM and γ is the assigned weight. $\mathcal{L}_M(I(i, j), J(i, j))$ is calculated as:

$$\mathcal{L}_M(i, j) = 1 - l_M(i, j) \times \prod_{j=1}^M cs_j(i, j), \quad (15)$$

where l_M and cs_j are the luminance and structural similarity to scales M and j , respectively. M denotes the total number of scales.

HHMS-SSIM focuses on removing the more severe haze images. And dense haze is more likely to form clumps compared to thin haze images. HHMS-SSIM focuses on block-by-block rather than pixel-by-pixel calculations. Therefore HHMS-SSIM does not capture well the normalized L1 norm error between a point $I(i, j)$ of the haze image and the corresponding point $J(i, j)$ of the ground truth image. We consider that since each pixel has a different L1 norm error, its loss function when back-propagating is also different. Therefore, we propose a loss

function \mathcal{L}_{Dy} based on the dynamic weights of the pixels. \mathcal{L}_{Dy} is calculated as:

$$\mathcal{L}_{Dy} = \gamma_{Dy} \times \mathcal{L}_p(i, j) + \gamma_M \times \mathcal{L}_M(i, j), \quad (16)$$

where γ_{Dy} and γ_M are the loss weights of \mathcal{L}_p and \mathcal{L}_M , respectively. \mathcal{L}_{Dy} changes according to the difference between the normalized L1 norm error of the haze and ground truth images at different pixel points, alleviating the problem of HHMS-SSIM focusing on block-by-block calculations.

4. Experiments

To train and fairly evaluate the proposed network as well as to compare methods, we selected the RESIDE [16] dataset with different haze densities and distributions and the NTIRE2018-Dehazing challenge dataset [50]. RESIDE is a dataset widely used for the evaluation of dehazing networks with different data sources and image contents highlighted in three datasets: Indoor Training Set (ITS), Outdoor Training Set (OTS), and Synthetic Objective Testing Set (SOTS). NTIRE2018-Dehazing challenge dataset includes an indoor dataset (I-HAZE [25]) and an outdoor dataset (OHAZE [26]). We selected ITS & OTS (I/OST), SOTS, and I-HAZE & O-HAZE (I/O-HAZE) as training datasets. These datasets have been used as evaluation datasets by some of the most advanced dehazing methods [51]. We perform an evaluation of the MFID-Net dehazing effect by using hazy dataset images as objects. All results are compared with nine state-of-the-art (SOTA) dehazing methods: DCP [18], DehazeNet [4], AOD-NET [20], GriddehazeNet [17], FFA-Net [22], DIDH [23], AECCR-NET [21], PSD [19], and PFF [33]. In addition, we perform ablation experiments to demonstrate the effectiveness of each of our modules.

4.1. Implementation details

MFID-Net was implemented by PyTorch 1.9.0 and Ubuntu 20.04 using GTX TITAN GPUs. To compare MFID-Net more fairly with the nine advanced dehazing methods mentioned above, we used different settings and parameter sizes for the different models. All images were randomly cropped to 512×512 pixels and normalized to pixel values from -1 to 1 . For MFID-Net, we used the Adam optimizer to perform the optimization. The initial learning rate was set to 10^{-4} . The coefficients $\beta_1 = 0.9$ and $\beta_2 = 0.999$ are used to calculate the running average of the gradient and its square. 300 epochs were trained and the decay rate was 0.75 per 10 epochs. $\alpha = 1.0$, $\beta = 0.5$, $A \in [0.7, 1.0]$, and $\beta_t \in [0.6, 1.8]$ in the dynamic activation function, both of which parameters are determined by the specific values given in the dataset. The weighting weights were set to $\gamma_{Dy} = 0.05$ and $\gamma_M = 10^{-3}$.

4.2. Evaluation and results

We use three widely used quantitative comparison metrics, the peak signal-to-noise ratio (PSNR), the structural similarity (SSIM) [52], and the fog aware density evaluator (FADE). All the methods of deep learning involved in the control group were fine-tuned on the above dataset. Tables 1, 2, and 3 summarize the PSNR, SSIM, and FADE scores of MFID-Net and the nine control methods on the I/OST, SOTS, and I/O-HAZE datasets, respectively. Table 4 summarizes the number of parameters and the average running time spent on dehazing 100 random 512×512 hazy images for MFID-Net and other dehazing methods. The 100 random images were randomly selected from the test set of I/OST, SOTS, and I/O-HAZE datasets. The running time is only the processing time of the GPU and does not take into account I/O operations. Bolded numbers indicate that the method is ranked first in performance. Underlined numbers indicate the second-ranked performance.

Our MFID-Net achieved PSNR of 37.6081 dB, SSIM of 0.9806, and 0.2766 of FADE on the I/OST test set, PSNR of 35.5436 dB, SSIM

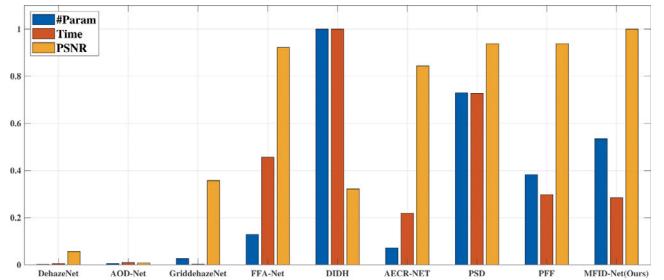


Fig. 5. Comparison bar chart of #Param, Time, and PSNR for 9 methods other than DCP. All data are normalized. It can be found that our proposed MFID-Net has the largest PSNR value with smaller number of parameters and running time.

of 0.9745, and 0.2931 of FADE on the SOTS test set, and PSNR of 23.6426 dB, SSIM of 0.8843, and 0.3189 of FADE on the I/O-HAZE test set. On the above test set, the average PSNR of MFID-Net is 32.2648 dB which outperforms PFF by up to 0.7915 dB, the average SSIM of MFID-Net is 0.9485 which outperforms PFF by up to 0.0077, and the average FADE of MFID-Net is 0.2962 which outperforms AECCR-NET by down to 0.0297. We compare the dehazing results of the above ten dehazing methods visually using output images on the I/OST, SOTS, and I/O-HAZE datasets, as shown in Figs. 6, 7, and 8 respectively. The numbers after the model name used in each figure indicate the PSNR and SSIM values in comparison with ground truth and the FADE scores.

For the indoor environments, we note that DCP (Figs. 6(c) and 7(c)) and DehazeNet (Figs. 6(d) and 8(c)) do not dehaze well. AOD-Net (Figs. 6(c), 7(c), and 8(c)), GriddehazeNet (Figs. 6(f) and 7(f)), DIDH (Fig. 6(h)), AECCR-Net (Fig. 6(i)) and PSD (Fig. 6(j)) remove most of the haze, but a small amount of clump haze remains. FFA-Net (Figs. 6(g), 7(g), and 8(g)), on the other hand, generates a darker image. PFF (Figs. 6(k), 7(k), and 8(k)) removes the haze well but introduces new noise. MFID-Net, however, does a better job of removing the haze and producing true colors, as shown in Figs. 6(l), 7(l), and 8(l).

For outdoor environments, DCP (Figs. 6(o), 7(o), and 8(o)) and DehazeNet (Figs. 6(p), 7(p), and 8(p)) have only limited dehazing effect. AOD-Net (Figs. 6(q) and 8(q)), DIDH (Figs. 6(t), 7(t), and 8(t)) and AECCR-Net (Fig. 6(q)) have some improvement in PSNR and SSIM values, but their dehazing effect is hardly satisfactory. GriddehazeNet (Figs. 6(r), 7(r), and 8(r)) and PSD (Figs. 7(v) and 8(v)) introduce more noise. FFA-Net (Figs. 7(s) and 8(s)) and PFF (Figs. 6(w) and 7(w)) have better dehazing effect, but their performance is poor on some images (Fig. 8(w)). MFID-Net is able to achieve the target of dehazing tasks in various environments, as shown in Figs. 6(x), 7(x), and 8(x), thanks to the dynamic weighting design of MFID-Net.

For the number of parameters and running time, our proposed MFID-Net is not superior, but it is not too bad either. The results are shown in Table 4. Fig. 5 shows the comparison bar chart of #Param, Time, and PSNR for our proposed MFID-Net and other than DCP. The shorter the #Param bar (blue) and the Time bar (orange), the lighter the model is. The higher the PSNR bar (yellow), the better the model performs. To better compare the gaps, all data are normalized. DCP is not a deep learning method, so there is no parameter. Early dehazing methods, DehazeNet, AOD-Net, and GriddehazeNet, have low parameters and are fast. FFA-Net, AECCR-NET, PFF, and our MFID-Net all have complex models with multiple sampling operations, which inevitably increase the computational cost. DIDH and PSD are built on multiple tandem networks, which leads to their high number of parameters.

We also note that MFID-Net has further excellence in performance on some hazy images. These hazy images are over-exposed or under-exposed. To further demonstrate the dynamic dehazing effect of MFID-Net, 11,473 under-exposed and over-exposed images (UOI) were selected from the test set of the above dataset for further comparison.

Table 1

Quantitative evaluation of our proposed method MFID-Net on the I/OST dataset, compared with the SOTA methods.

Methods	DCP	DehazeNet	AOD-Net	GriddehazeNet	FFA-Net	DIDH	AECR-NET	PSD	PFF	MFID-Net(Ours)
PSNR	14.0113	19.1207	24.5892	25.1380	36.4640	20.0812	35.1903	36.4370	<u>37.0431</u>	37.6081
SSIM	0.3996	0.8444	0.4519	0.8713	0.9845	0.8834	0.9603	0.9743	<u>0.9789</u>	<u>0.9806</u>
FADE	0.8516	0.4471	0.4220	0.4743	0.3429	0.8218	0.3429	<u>0.2945</u>	0.3341	0.2766

Table 2

Quantitative evaluation of our proposed method MFID-Net on the SOTS datasets, compared with the SOTA methods.

Methods	DCP	DehazeNet	AOD-Net	GriddehazeNet	FFA-Net	DIDH	AECR-NET	PSD	PFF	MFID-Net(Ours)
PSNR	15.4439	21.1410	17.1504	24.6334	34.8850	28.3028	35.1757	35.3538	<u>35.5436</u>	35.5436
SSIM	0.3996	0.8444	0.4519	0.8713	<u>0.9745</u>	0.8834	0.9603	0.9743	<u>0.9735</u>	0.9806
FADE	0.7713	0.4032	0.3776	0.3803	0.3270	0.5359	<u>0.3047</u>	0.3142	0.3526	0.2931

Table 3

Quantitative evaluation of our proposed method MFID-Net on the I/O-HAZE datasets, compared with the SOTA methods.

Methods	DCP	DehazeNet	AOD-Net	GriddehazeNet	FFA-Net	DIDH	AECR-NET	PSD	PFF	MFID-Net(Ours)
PSNR	15.0785	20.4167	16.7885	22.4308	22.4472	22.4651	20.4450	<u>22.6268</u>	21.8330	23.6426
SSIM	0.6933	0.7194	0.8434	0.8172	0.8362	0.8236	0.7974	0.8224	<u>0.8699</u>	0.8843
FADE	0.5192	0.6500	0.4469	0.3777	0.3792	0.3840	<u>0.3301</u>	0.4035	0.3951	0.3189

Table 4The number of parameters and the average running time spent on dehazing 100 random 512×512 hazy images for MFID-Net and other dehazing methods.

Methods	DCP	DehazeNet	AOD-Net	GriddehazeNet	FFA-Net	DIDH	AECR-NET	PSD	PFF	MFID-Net(Ours)
#Param	–	<u>0.01M</u>	0.002M	0.96M	4.68M	36.21M	2.61M	26.43M	13.87M	19.36M
Time	152.5 s	37.21 m s	<u>20.63 m s</u>	<u>30.75 m s</u>	1.54 s	3.35 s	750 m s	2.44 s	1.01 s	970 m s

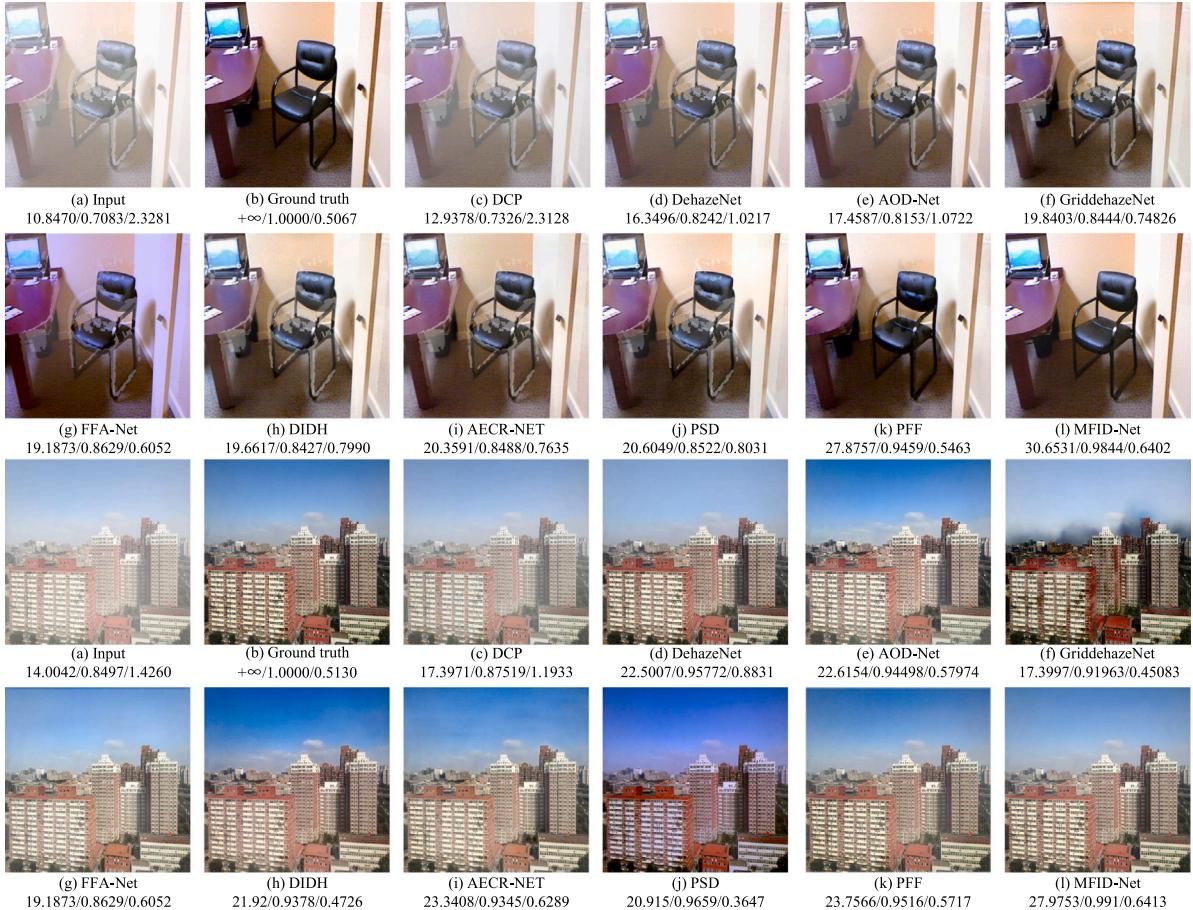


Fig. 6. Dehazing visualization results on the I/OST dataset. Our proposed method, MFID-Net, is better at removing haze, generating better visual quality, and restoring original colors than the state-of-the-art methods.

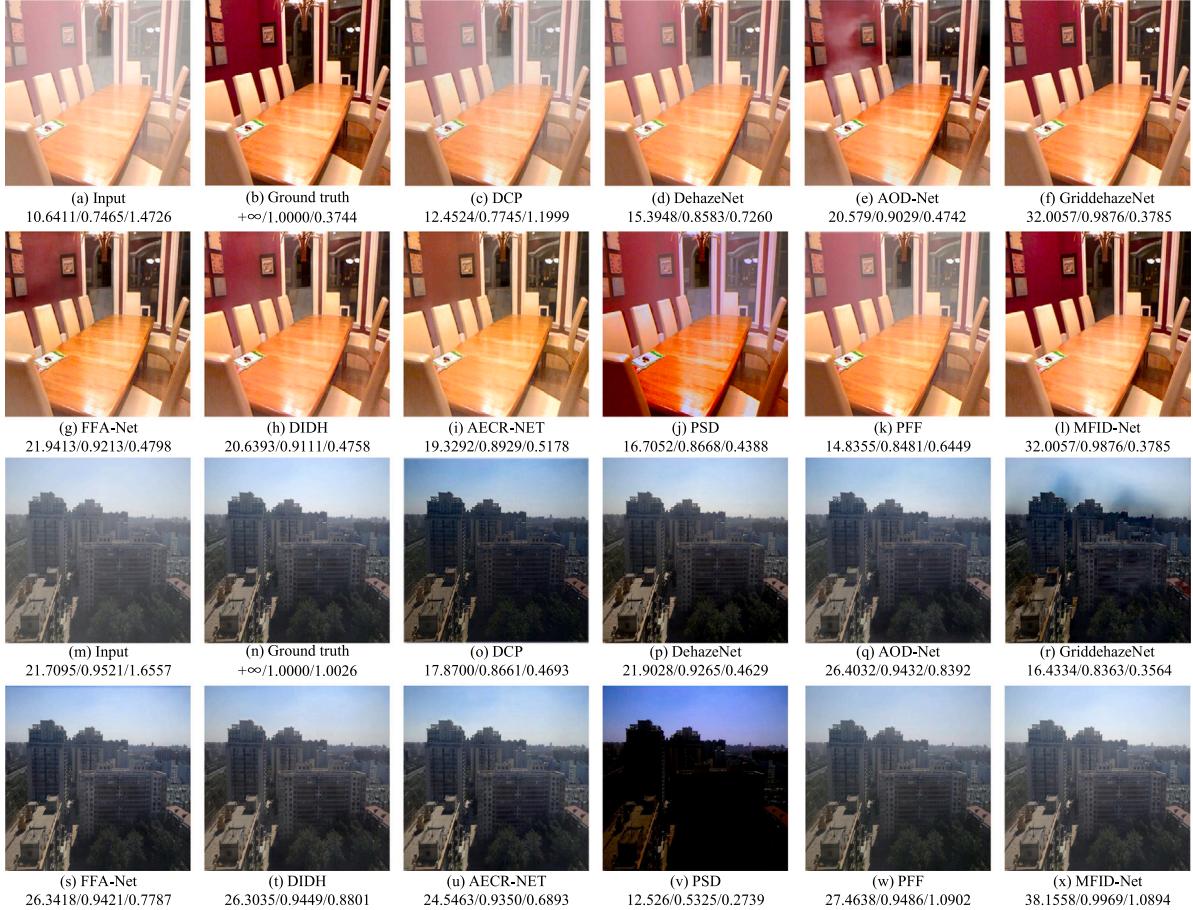


Fig. 7. Dehazing visualization results on the SOTS dataset. Our result is visually close to the ground truth.

Table 5

Quantitative evaluation of our proposed method MFID-Net on the UOI, compared with the SOTA methods.

Methods	DCP	DehazeNet	AOD-Net	GriddehazeNet	FFA-Net	DIDH	AECCR-NET	PSD	PFF	MFID-Net(Ours)
PSNR	23.1832	23.5790	<u>24.3903</u>	13.8547	22.3642	22.6531	23.5168	<u>22.1739</u>	20.6014	26.8114
SSIM	0.6744	0.7642	<u>0.8946</u>	0.6567	0.8974	0.8536	<u>0.9069</u>	0.3247	0.8163	0.9366
FADE	2.1058	0.8102	1.2767	0.6922	0.8657	1.5548	0.8489	<u>0.4822</u>	1.5548	0.4102

The image in UOI is defined as an image in which the number of pixels with a value of more than 240 and less than 20 is greater than 10% of the total number of pixels in the image. We performed a further comparison on the UOI, and Table 5 shows the scores of MFID-Net and the nine control methods on the UOI. Bolded numbers indicate that the method is ranked first in performance. Underlined numbers indicate the second-ranked performance. Our MFID-Net achieves a PSNR of 26.8114 dB, SSIM of 0.9366, and FADE of 0.4102 on the UOI dataset. The PSNR of MFID-Net is 2.4211 dB higher than the second place AOD-Net whose PSNR is 24.3903 dB. And the SSIM of MFID-Net is 0.0297 higher than the second place AECCR-NET whose SSIM is 0.9069. We also use Fig. 9(a) as an example. A more visual comparison of the above methods is shown in Fig. 9. For the images in the UOI, some of the dehazing methods fail, such as PSD, DIDH, FFA-Net, GriddehazeNet, and DCP. AECCR-NET, AOD-NET, and DehazeNet, although partially improved in PSNR or SSIM, do not restore the original colors of the under-exposed parts very well (e.g. Fig. 9(i), (e), (d), and (k) for the blue and red roofs in the partially enlarged section), nor can they suppress further overexposure in the over-exposed section after methodic processing (e.g. the sun and the sky near the sun in Fig. 9(d), (f), (i), and (k)). The dynamic characteristics of MFID-Net enable it to well adapt to the interference caused by over-exposed and under-exposed parts. The MFID-Net in Fig. 9(l) produces a dehazing result that is closer to the ground truth haze-free image in Fig. 9(b).

4.3. Ablation experiments

To demonstrate the effectiveness of the proposed method MFID-Net, we conducted ablation studies on the I/OST dataset to analyze the impact of different modules, including the DDC, DDA, and dynamic loss function. We designed the following ablation studies: (1) **base (BS)**: baseline without DDC, DDA, and dynamic loss function; (2) **base + dynamic dehazing convolution (BS+DDC)**: replace the normal convolution layer in the baseline with the DDC. (3) **base + dynamic dehazing convolution + dynamic dehazing activation (BS+DDC+DDA)**: replace the ordinary convolution layer and activation in the baseline with the DDC and DDA. (4) **base + dynamic dehazing convolution + dynamic dehazing activation + dynamic loss function (BS+DDC+DDA+DL)**: MFID-Net. Ablation studies are performed on the I/OST dataset and their performance is shown in Table 3. The PSNR, SSIM, and loss for each ablation experiment in the first 200 epochs are shown in Fig. 10. Although the performance of the dynamic loss function seems to be modest compared to the other parts, as shown in Fig. 10(a)(b). But the dynamic loss function speeds up the convergence speed during training, as shown in Fig. 10(c). Faster convergence means that the accuracy requirement can be achieved faster using the same computational resources. This is significant for training on low-performance computational resources. We find that our

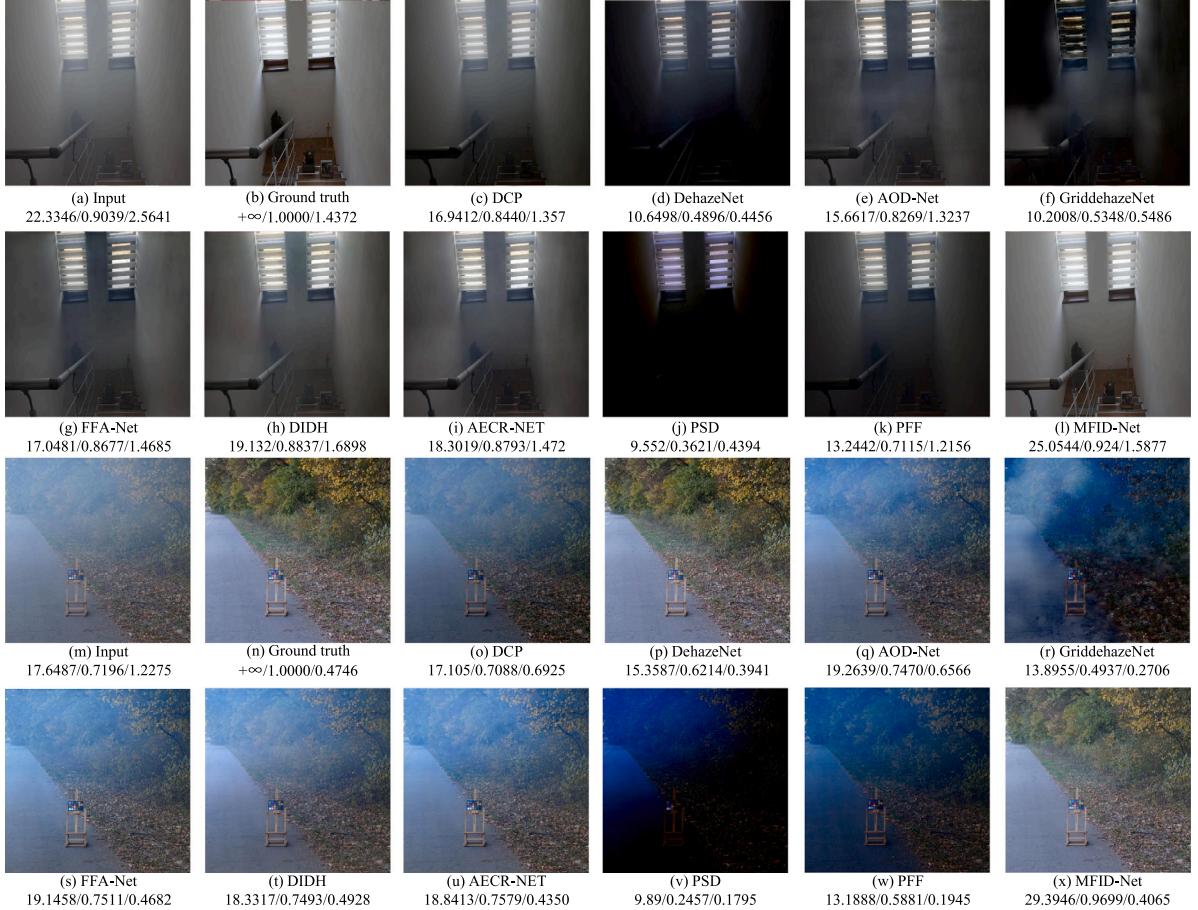


Fig. 8. Dehazing visualization results on the I/O-HAZE dataset. Our result is visually close to the ground truth.



Fig. 9. Dehazed results on the UOI dataset. The dynamic nature of MFID-Net enables better handling of under-exposed and over-exposed hazy images.

proposed DDC, DDA, and dynamic loss function can help MFID-Net to dehaze better, where the DDC can increase the PSNR of the base on the I/O-HAZE dataset by 3.4424 dB, and the DDA can increase the SSIM by 0.0092. The visualization results of ablation experiments are shown in Fig. 11. This result demonstrates that the DDC, DDA, and dynamic loss function can improve the performance of the network to different degrees (see Table 6).

5. Conclusions

We propose the multi-scaled feature-fused image dehazing network called MFID-Net, which extracts information from different color channels and separates multi-scale haze concentrations to efficiently generate high-quality haze-free images from the input hazy images. The key to our method is the use of dynamic dehazing convolution based on

Table 6
Ablation studies on MFID-Net.

Methods	PSNR	SSIM
BS	32.9309	0.9676
BS+DDC	36.3733 (3.4424↑)	0.9709 (0.0033↑)
BS+DDC+DDA	37.6032 (1.2299↑)	0.9801 (0.0092↑)
BS+DDC+DDA+DL (MFID-Net)	37.6081 (0.0049↑)	0.9806 (0.0005↑)

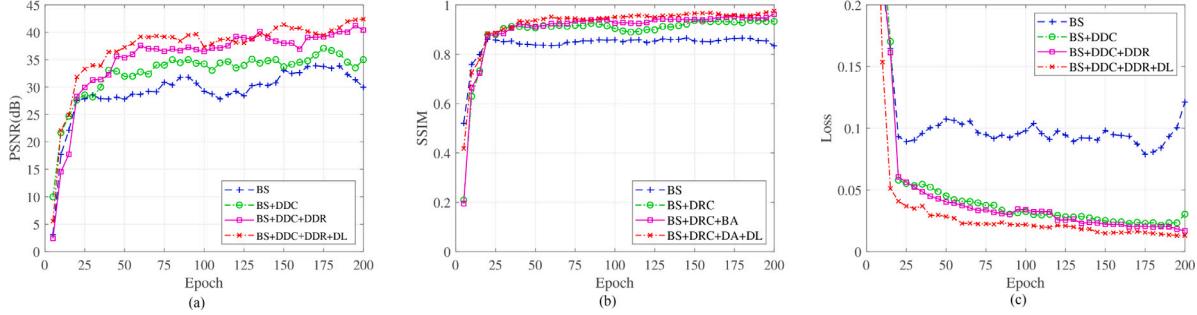


Fig. 10. The PSNR (a), SSIM (b), and loss (c) for each ablation experiment in the first 200 epochs.

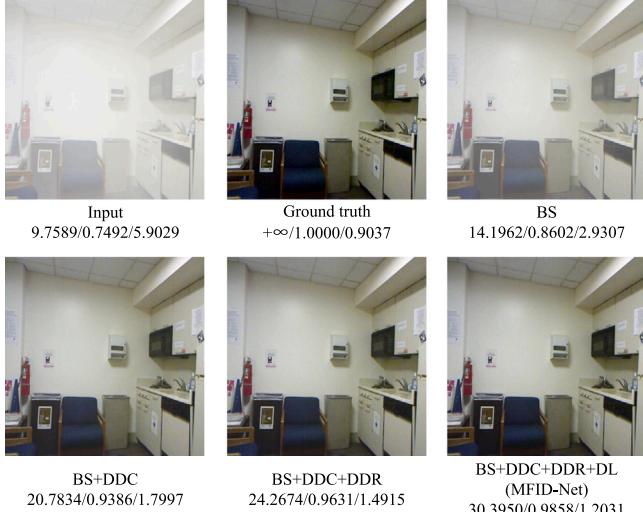


Fig. 11. The visualization results of ablation experiments.

attentional weight calculation and dynamic weight fusion to self-adapt its input parameters according to the haze concentration on different pixels. The dynamics and adaptability of the network are further enhanced by the dynamic dehazing activation function based on the input global context encoding function. On the I/OST, SOTS, and I/O-HAZE datasets, the average PSNR of MFID-Net is 32.2648 dB, which is 0.1361 dB higher than the SOTA methods. The average SSIM of MFID-Net is 0.9716, which is 0.7915 higher than the SOTA methods. MFID-Net has excellent dehazing ability on under-exposed and over-exposed images. The PSNR of MFID-Net on the UOI is 26.8114 dB, which is 3.2681 dB higher than the SOTA methods. Qualitative and quantitative results show that MFID-Net can demonstrate superior performance over the SOTA methods while improving further on under-exposed and over-exposed hazy images, and producing satisfactory dehazing results. However, although MFID-Net performs well on daytime hazy images and under- and over-exposed hazy images, MFID-Net does not dehaze well at nighttime because MFID-Net focuses on considering the dynamic transformation of light. In addition, the computational resources and dehazing runtime used by MFID-Net are not ideal. How to use dynamic weights to dehaze in nighttime environment and how to optimize the efficiency of MFID-Net will be our future research goals.

Funding

This work was supported in part by the project of National Natural Science Foundation of China under Grant No. 62272178.

CRediT authorship contribution statement

Yifan Liu: Wrote the main manuscript text, Designed the deraining method, Improved the performance of the model and designed the ablation experiments, Performed the experiments, Reviewed the manuscript. **Jincai Chen:** Improved the performance of the model and designed the ablation experiments, Reviewed the manuscript. **Ping Lu:** Improved the performance of the model and designed the ablation experiments, Reviewed the manuscript. **Chuanbo Zhu:** Wrote the main manuscript text, Reviewed the manuscript. **Yugen Jian:** Wrote the main manuscript text, Reviewed the manuscript. **Chao Sun:** Designed the deraining method, Reviewed the manuscript. **Han Liang:** Designed the deraining method, Reviewed the manuscript.

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

I have shared the data in the manuscript

References

- [1] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluiu, Radu Timofte, NTIRE 2021 nonhomogeneous dehazing challenge report, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 627–646.
- [2] Zhangkai Lyu, Andrew Peng, Qingwei Wang, Dandan Ding, An efficient learning-based method for underwater image enhancement, Displays 74 (2022) 102174.
- [3] Earl J. McCartney, Optics of the atmosphere: scattering by molecules and particles, New York (1976).
- [4] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, Dacheng Tao, Dehazenet: An end-to-end system for single image haze removal, IEEE Trans. Image Process. 25 (11) (2016) 5187–5198.
- [5] Jeya Maria Jose Valanarasu, Rajeev Yasarla, Vishal M. Patel, TransWeather: Transformer-based restoration of images degraded by adverse weather conditions, 2021, arXiv preprint [arXiv:2111.14813](https://arxiv.org/abs/2111.14813).

- [6] Chenggang Dai, Mingxing Lin, Xiaojian Wu, Dong Zhang, Single hazy image restoration using robust atmospheric scattering model, *Signal Process.* 166 (2020) 107257.
- [7] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [8] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko, End-to-end object detection with transformers, in: European Conference on Computer Vision, Springer, 2020, pp. 213–229.
- [9] Hayat Ullah, Khan Muhammad, Muhammad Irfan, Saeed Anwar, Muhammad Sajjad, Ali Sharif Imran, Victor Hugo C de Albuquerque, Light-DehazeNet: A novel lightweight CNN architecture for single image dehazing, *IEEE Trans. Image Process.* 30 (2021) 8968–8982.
- [10] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, Dacheng 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.
- [11] Qili Chen, Junfang Fan, Wenbai Chen, An improved image enhancement framework based on multiple attention mechanism, *Displays* 70 (2021) 102091.
- [12] Mingye Ju, Can Ding, Wenqi Ren, Yi Yang, Dengyin Zhang, Y Jay Guo, Ide: Image dehazing and exposure using an enhanced atmospheric scattering model, *IEEE Trans. Image Process.* 30 (2021) 2180–2192.
- [13] Yanyun Qu, Yizi Chen, Jingying Huang, Yuan Xie, Enhanced pix2pix dehazing network, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8160–8168.
- [14] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, Ming-Hsuan Yang, Single image dehazing via multi-scale convolutional neural networks, in: European Conference on Computer Vision, Springer, 2016, pp. 154–169.
- [15] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, Zhangyang Wang, Enlightengan: Deep light enhancement without paired supervision, *IEEE Trans. Image Process.* 30 (2021) 2340–2349.
- [16] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, Zhangyang Wang, Benchmarking single-image dehazing and beyond, *IEEE Trans. Image Process.* 28 (1) (2018) 492–505.
- [17] Xiaohong Liu, Yongrui Ma, Zhihao Shi, Jun 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.
- [18] Kaiming He, Jian Sun, Xiaowu Tang, Single image haze removal using dark channel prior, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (12) (2010) 2341–2353.
- [19] Zeyuan Chen, Yangchao Wang, Yang Yang, Dong Liu, PSD: Principled synthetically-real dehazing guided by physical priors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7180–7189.
- [20] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, Dan Feng, An all-in-one network for dehazing and beyond, 2017, arXiv preprint [arXiv:1707.06543](https://arxiv.org/abs/1707.06543).
- [21] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, Lizhuang 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.
- [22] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, Huizhu 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.
- [23] Pranjay Shyam, Kuk-Jin Yoon, Kyung-Soo Kim, Towards domain invariant single image dehazing, 2021, arXiv preprint [arXiv:2101.10449](https://arxiv.org/abs/2101.10449).
- [24] Hao Zhou, Ze Zhao, Hailing Xiong, Yun Liu, A unified weighted variational model for simultaneously haze removal and noise suppression of hazy images, *Displays* 72 (2022) 102137.
- [25] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, Christophe De Vleeschouwer, I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images, in: International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, 2018, pp. 620–631.
- [26] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, Christophe 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.
- [27] Wei-Ting Chen, Hao-Yu Fang, Jian-Jiun Ding, Cheng-Che Tsai, Sy-Yen Kuo, JSTASR: Joint size and transparency-aware snow removal algorithm based on modified partial convolution and veiling effect removal, in: European Conference on Computer Vision, Springer, 2020, pp. 754–770.
- [28] Pengyue Li, Jiandong Tian, Yandong Tang, Guolin Wang, Chengdong Wu, Deep retinex network for single image dehazing, *IEEE Trans. Image Process.* 30 (2020) 1100–1115.
- [29] Huan Liu, Chen Wang, Jun Chen, Indirect domain shift for single image dehazing, 2021, arXiv preprint [arXiv:2102.03268](https://arxiv.org/abs/2102.03268).
- [30] Xin Yi, Bo Ma, Yulin Zhang, Longyao Liu, JiaHao Wu, Two-step image dehazing with intra-domain and inter-domain adaptation, 2021, arXiv preprint [arXiv:2102.03501](https://arxiv.org/abs/2102.03501).
- [31] Xiaoqin Zhang, Runhua Jiang, Tao Wang, Wenhan Luo, Single image dehazing via dual-path recurrent network, *IEEE Trans. Image Process.* 30 (2021) 5211–5222.
- [32] Chia-Ming Chang, Chang-Sung Sung, Tsung-Nan Lin, DAMix: Density-aware data augmentation for unsupervised domain adaptation on single image dehazing, 2021, arXiv preprint [arXiv:2109.12544](https://arxiv.org/abs/2109.12544).
- [33] Haoran Bai, Jinshan Pan, Xinguang Xiang, Jinhui Tang, Self-guided image dehazing using progressive feature fusion, *IEEE Trans. Image Process.* 31 (2022) 1217–1229.
- [34] Dong Zhao, Jia Li, Hongyu Li, Long Xu, Hybrid local-global transformer for image dehazing, 2021, arXiv preprint [arXiv:2109.07100](https://arxiv.org/abs/2109.07100).
- [35] Huihai Wu, Shuai Zheng, Junge Zhang, Kaiqi Huang, Fast end-to-end trainable guided filter, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1838–1847.
- [36] Xiaolong Wang, Ross Girshick, Abhinav Gupta, Kaiming He, Non-local neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.
- [37] Hang Su, Varun Jampani, Deqing Sun, Orazio Gallo, Erik Learned-Miller, Jan Kautz, Pixel-adaptive convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11166–11175.
- [38] Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, Zicheng Liu, Dynamic convolution: Attention over convolution kernels, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11030–11039.
- [39] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [40] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017, arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861).
- [41] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, Jian Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6848–6856.
- [42] Xavier Glorot, Antoine Bordes, Yoshua Bengio, Deep sparse rectifier neural networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, in: JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.
- [43] Andrew L. Maas, Awni Y. Hannun, Andrew Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. ICML, Volume 30, No. 1, Citeseer, 2013, p. 3.
- [44] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1026–1034.
- [45] Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, Zicheng Liu, Dynamic relu, in: European Conference on Computer Vision, Springer, 2020, pp. 351–367.
- [46] Ningning Ma, Xiangyu Zhang, Ming Liu, Jian Sun, Activate or not: Learning customized activation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 8032–8042.
- [47] Xiaoqin Zhang, Jinxin Wang, Tao Wang, Runhua Jiang, Hierarchical feature fusion with mixed convolution attention for single image dehazing, *IEEE Trans. Circuits Syst. Video Technol.* 32 (2) (2022) 510–522.
- [48] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, Ming-Hsuan Yang, Multi-scale boosted dehazing network with dense feature fusion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020.
- [49] Guangfa Wang, Xiaokang Yu, MSF\$^2\$DN: Multi scale feature fusion dehazing network with dense connection, in: Proceedings of the Asian Conference on Computer Vision, ACCV, 2022, pp. 2950–2966.
- [50] Cosmin Ancuti, Codruta O. Ancuti, Radu Timofte, Ntire 2018 challenge on image dehazing: Methods and results, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 891–901.
- [51] Ruoteng Li, Robby T. Tan, Loong-Fah Cheong, All in one bad weather removal using architectural search, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3175–3185.
- [52] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.