JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 1

LFD-Net: Lightweight Feature-interaction Dehazing Network for Real-time Vision Tasks

Yizhu Jin, Jiaxing Chen, Feng Tian and Kun Hu *Member,* *IEEE*

***Abstract***—Single image dehazing is a challenging low-level vision task, which is required to be both effective and efficient for downstream real-time applications like autonomous navigation, video surveillance and remote sensing. Existing methods usually suffer from high computation cost with densely connected resid- ual modules. They might also struggle with color consistency and the maintenance of visual quality. To tackle these problems, we design a lightweight architecture to extract, fuse and weight multi-level features with the assistance of physics-based Atmo- sphere Scattering Model (ASM). Our proposed LFD-Net demon- strates strong interpretability by exploiting Gated Fusion module and attention mechanism to make interaction between multi-level representation. The evaluation on outdoor SOTS dataset reaches an average Frequency Per Second (FPS) of 54.41, nearly 8 times faster than seven most popular SOTA methods with equivalent metrics. It also improves the performance of object detection in terms of mean Average Precision when IoU = 0.5 (mAP@0.5) based on YOLOv5 by 4.73% on DAIR-V2X, ensuring practica- bility and adaptability for real-time vision tasks. Our codes are available at **https://github.com/RacerK/LFD-Net**.

***Index*** ***Terms***—Single Image Dehazing, real-time application, model compression, interpretablity.

I . INTRODUCTION

N

OWADYAS, diverse vision tasks including object de- tection, segmentation, inpainting and compression, are having progressive results. However, their performance is limited confronted with noised, distorted or degraded image inputs under greasy weather conditions. This is because cam- eras absorb in decaying and scattered light rays when small particles like water vapor, sand, dust and smoke suspend in the air, which causes a decrease in information entropy [1].

Haze is a prevalent phenomenon that occurs in many parts of the world, and it has a significant impact on real-time ap- plications. In the field of remote sensing, haze can obscure the details of the earth’s surface, affecting the accuracy of various applications such as land cover mapping, change detection, and environmental monitoring. This is particularly relevant for low altitude Unmanned Aerial Vehicles (UAVs), which are often

Manuscript received April 19, 2021; revised August 16, 2021. This work was supported in part by the Institute of Artificial Intelligence, the Youth talent support program of Beihang University under Grant KLSMNR-202208, in part by the National Innovation Center of Intelligent and Connected Vehicles, and in part by the Key Laboratory of Land Satellite Remote Sensing Application, Ministry of Natural Resources of the People’s Republic of China under Grant YWF-22-L-1277. *(Corresponding* *author:* *Kun* *Hu.)*

Kun Hu is with the Institute of Artificial Intelligence, Beihang, China. Yizhu Jin is with the School of Automation Science and Electrical Engineering, Beihang, China (e-mail: kunhu@buaa.edu.cn, 19374316@buaa.edu.cn).

Jiaxing Chen and Feng Tian are with the National Innovation Center of In- telligent and Connected Vehicles, Beijing, China (e-mail: chenjiaxing@china- icv.cn, fengtianreal95@gmail.com).

(Yizhu Jin and Jiaxing Chen contributed as co-first authors.)

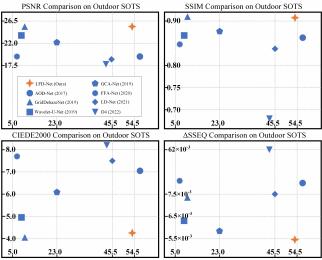


Fig. 1. Comparison Metrics on Outdoor SOTS, in terms of PSNR, SSIM, CIEDE2000, ∆SSEQ (vertical-axis) and FPS (horizontal-axis).

used in agriculture, forestry, and urban planning. The presence of haze can significantly reduce the effectiveness of these applications by limiting the visibility and detail of the captured images. Similarly, satellite imagery, which is widely used for environmental monitoring and disaster response, also suffers from the impact of haze. The haze in satellite imagery can significantly reduce the accuracy of data interpretation, leading to incorrect conclusions and ineffective decisions [2], [3]. To address the issue of haze in remote sensing and navigation, researchers have developed various methods to dehaze the images, including the use of image priors, physical models and deep learning techniques. These methods aim to remove the effects of haze from the images to improve their clarity and detail, thereby enhancing the accuracy of high-level perception tasks.

The earliest dehazing methods are mainly based on prior knowledge. For instance, in [4], Dark Channel Prior makes an approximation that haze-effected pixels have at least one rela- tively low intensity value among RGB channels. In [5], a semi- physical guided-filter based approach is adopted to refine the coarse haze thickness map to restore textural information. In [6], depth estimation and image segmentation are incorporated with Dark Channel Prior to generate the final transmittance. Generally, these methods are subject to empirical or statistical regularities, which limits the application scenarios to some extent.

Furthermore, as a physic-based model, ASM is also ex-

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 2

tensively researched and introduced in dehazing methods. It is physically-grounded for an unrestricted access to various image scenes through the estimation of global atmosphere light and transmission map. For instance, in [7], an end-to-end DehazeNet combines dark channel, maximum contract, color attenuation as well as hue disparity prior to compute the trans- mission map and assigns a default value to atmosphere light. In [8], a Haze Density Prediction Network is designed for a more accurate approximation of atmosphere light to better fit for nighttime occasions. In [9], a multi-decoder framework is presented to handle multiple bad weather restoration, with rain veiling effect embedded into the conventional ASM. In [10], a differential guided layer is embedded with the backbone and substituted to the physical scattering equation.

In addition to traditional dehazing methods, deep learning techniques have been widely adopted in dehazing research. Among these techniques, Convolutional Neural Networks (CNNs) have been the most extensively studied architecture for deep-learning-based dehazing methods. In our method, we also use CNNs as the underlying architecture. In the second section, we will discuss the common challenges and techniques in developing CNN-based dehazing methods. Besides CNNs, Generative Adversarial Networks (GANs) have also been applied in image dehazing. For example, in [11], an integrated multi-task algorithm is proposed to jointly remove raindrop and haze from both sky and non-sky regions. In [12], the generator is designed to capture uneven foggy features in sequential and parallel manners to obtain haze-free images.

In [13], an inverse-reverse module is utilized to correlate diverse image styles in adversarial training. In [14], unpaired clean and hazy images are utilized for training, and attention- guided generators produce attention masks, which are fused with generation outputs to enhance image quality. These GAN- based methods perform well on synthetic image data, but they may suffer from a loss of authenticity when applied to natural hazy images. Therefore, our method, which is CNN-based, may be more appropriate for dehazing natural hazy images.

In deep-learning-based techniques, there is a persistent chal- lenge of the trade-off between high performance on specific datasets and generalization to diverse real-world applications. It is crucial to consider both performance and generalization when developing and evaluating potential solutions in im- age restoration. Although current methods may demonstrate promising results under certain conditions, their lack of effi- ciency and generalization capabilities render them less suitable for real-time and practical applications. Due to the urgent need for a context-awareness system supporting quick response confronted with extreme situations, the study of lightweight dehazing methods has become increasingly crucial. AOD-Net [15] strives to concatenate multi-level features in different patterns, which is the baseline of other lightweight dehazing models. Both FAOD-Net [16] and GADO-Net [17] introduce depth-wise, point-wise convolution to reduce the amount of parameters and aggregate context information in pyramid pool- ing module. FAMED-Net [18] applies cascaded and densely connected point-wise convolutional and pooling layers at three scales. LD-Net [19] concatenates convolutional layers that have a semantic gap rather than combining adjacent layers.

Besides, it leverages a Color Visible Restoration module to improve the color consistency.

Our proposed Lightweight Feature-interaction Dehazing Network (LFD-Net) utilizes convolutional layers of different kernel sizes as a sequence to extract multi-level features.

The feature-interaction process is coherently addressed, which takes in, redistributes and reassigns weights to the extracted features. Each component of our network performs its own functions, but also makes interactions effectively and effi- ciently as a whole. Overall, our main contributions are three- fold:

• Our method adopts the ASM to jointly approximate atmosphere light and transmission map to increase the authenticity of restored images as well as the inference efficiency. It incorporates the convolutional operations into more specialized modules while maintaining the conciseness.

• Our proposed method is designed to provide interpretabil- ity by assigning distinct tasks to each module, as demon- strated by the results of our visualization and abla- tion experiments. The feature-interaction process relies heavily on element-wise multiplication, which has been shown to enhance the performance of pure convolutional operations.

• Our proposed method has been extensively validated across various scenarios to demonstrate its stability, prac- ticality, and generalization for real-world vision tasks, achieving an excellent trade-off between accuracy and efficiency. Our approach effectively addresses common challenges such as halo effect, gridding artifact, and color inconsistency, resulting in significant improvements in the performance of object detection tasks.

II . RELATED WORK

Previous research has shown that deep learning techniques have achieved superior results in dehazing tasks. Our approach considers image dehazing as an image restoration task that utilizes deep learning techniques, with emphasis on the feature extraction and feature utilization processes, as discussed in the first two parts of this section. Moreover, the evaluation metrics used to assess the performance of dehazing methods are crucial and should consider multiple dimensions, as described in the third part of this section.

*A.* *Feature* *Extraction*

One of the key challenges in image reconstruction is the extraction of multi-level or multi-scale features, which can be facilitated by the use of a symmetric encoder-decoder structure. The U-Net architecture, originally designed for effective extraction of context information at different scales or levels [20], has been widely used as a backbone in various reconstruction tasks. In [21], the Strengthen-Operate-Subtract boosting strategy is incorporated into the decoder, and a dense feature fusion module utilizing a back-projection feed- back scheme is leveraged to compensate the missing spatial information from high-resolution features. In [22], the U- Net architecture is modified to incorporate discrete wavelet

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 3

transform and inverse discrete wavelet transform in place of conventional down-sampling and up-sampling. In [23], hybrid convolution is applied in the U-Net encoder, which combines standard convolution with dilated convolution, to expand the receptive field and extract image features with more detail.

As opposed to a fixed backbone like U-Net, some methods utilize more flexible structures with multiple paths to diversify color information or conduct various tasks. For instance, in [24], image dehazing and depth estimation are addressed simultaneously in a framework with four decoders sharing information from the same encoder. In [25], a multi-color space encoder that incorporates RGB, LAB, and HSV is applied to extract representative features in separate paths. In [26], quadruple color-cue is integrated into a multi-look architecture with multi-weighted training loss for autonomous vehicular application.

*B.* *Feature* *Utilization*

Another major challenge in image reconstruction tasks is the

effective utilization of extracted features, which has prompted the exploration of various feature fusion strategies and at- tention mechanisms. For instance, in [27], a novel attention- based multi-scale estimation module is implemented in the backbone on a grid network to alleviate the bottleneck issue encountered in conventional multi-scale approaches. In [28], a block structure integrated with Channel-wise Attention (CA), Pixel-wise Attention (PA) is stacked to form a group structure, which is progressively triple-stacked and concatenated to feed

into another CA-PA attention mechanism for feature fusion.

In [29], a multi-level fusion module is presented to integrate low-level and high-level representations. Besides, the Residual Mixed-convolution Attention Module is developed to guide the network to focus on significant features in the learning process. In [30], the feature fusion method progressively aggregates the features of hazy image and generated reference image to remove the useless features.

Moreover, self-attention mechanism, proposed in Trans- former, has also been practiced in dehazing methods. For instance, a Transformer-based channel attention module and a spatial attention module are combined to form an attention module that enhances channel and spatial features [31]. Long- range dependencies of image information can be effectively extracted through Transformer blocks in image dehazing [32]. Recently, it has been revealed in [33] that self-attention mech- anism inherently functions as a two-order feature interaction. Based on this, gated convolution has been developed as an alternative method to achieve an equivalent effect to self- attention, while reducing the computation cost.

*C.* *Quality* *Evaluation*

Existing methods usually focus on high performance quan- tified by metrics in terms of peak-signal-to-noise-ratio (PSNR) and structure similarity index (SSIM). To be specific, PSNR measures the ratio between the maximum possible value of a pixel and the power of corrupting noise that affects the restoration fidelity. As opposed to directly estimating absolute

error, SSIM reveals inter-dependencies within pixels by lumi- nance masking and contrast masking between spatially-close image pairs. Besides the commonly used metrics PSNR and SSIM, CIE2000 Delta E formula (CIEDE2000) and Spatial- Spectral Entropy-based Quality (SSEQ) are also introduced in our comparison metrics. CIEDE2000 is used for quantifying the visual difference between two colors. It takes into account the chromaticity and luminance of the colors being compared, as well as the surrounding colors and the viewing conditions [34]. SSEQ is calculated by separating the image into its spatial and spectral components, calculating and combining the entropy of each component [35]. Taken halo effects into consideration, in our comparison, the average CIEDE2000 of each pixel is calculated in image pairs and the average absolute value of relative error on SSEQ, namely ∆SSEQ is provided.

III . PROPOSED METHOD

*A.* *Preliminaries*

The conventional ASM can be reformulated to jointly estimate the global atmosphere light A and the transmission map t, resulting in a reduction of parameters [15]:

I(θ) = J(θ) × t(θ) + A(1 − t(θ)), (1)

where A is treated as a constant, t ∈ (0, 1] denotes the pixel- wise transmittance of light, θ represents the pixel coordinate of an H×W image of height H and width W, with I and J being the hazy input and haze-free output, respectively. Therefore, our desired value, the haze-free approximation, can be written as:

J(θ) =  + A. (2)

This model is also adopted in our proposed method due to its contribution to the efficiency of real-time vision tasks. To encapsulate these two factors into one variable, the formula of

the reformulated ASM is as follows:

J(θ) = K(θ) × I(θ) − K(θ) + b, (3)

where K(θ) represents the new incorporated variable, which can be derived as:

K(θ) =  × (I(θ) − A) + (A − b) (4)

I(θ) − 1 .

To be specific, K is the intermediate evaluation parameter of the network. Our ultimate goal is to generate a separate K value for each input channel, usually in terms of RGB. That is, K in size 3 × H × W is substituted into equation (3) at the end of the network, with a most commonly used default value b = 1.

*B.* *Network* *Design*

The proposed LFD-Net differs from heavyweight frame- works or other lightweight frameworks due to its concise method of feature extraction and interaction, as depicted in Fig. 1.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 4

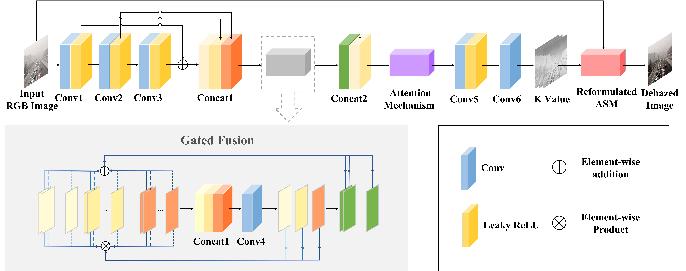


Fig. 2. Architecture of the Lightweight Feature-interaction Dehazing Network. The reformulated ASM generates a clear output by substituting the evaluated K value. The network primarily consists of convolutional layers and concatenation layers, with the use of element-wise product in the Gated Fusion module and attention mechanism.

In CNNs, convolution kernels of varying sizes are used to extract features at different levels of abstraction. Specifically, smaller-sized kernels are effective at capturing local features, while larger-sized kernels are better suited to capturing fea- tures with a larger receptive field, which can be thought of as more global features. The most commonly used kernel size is 3 × 3. However, it is not effective enough to stack convolutional layers with this typical kernel size in lightweight models. In [19], different kernel sizes ranging from 1 × 1, 3 × 3, 5 × 5 to 7 × 7 are utilized alternatively. Besides, concatenation layers are introduced at three different stages to combine the low- with high-level features, which compensates for the loss of information from initial layers as the network proceeds deeper. Therefore, the formation of convolutional and concatenation layers are crucial and can be flexibly designed to meet with specific needs. In our proposed method, we benefit from this design concept. Different from existing methods, we further simplify the formation of convolutional layers in feature extracting process. We also introduce feature- interaction strategies as opposed to merely using convolutional

and concatenation layers, including the Gated Fusion Module and Attention Mechanism.

To be specific, in feature extraction, a sequence of convo- lutional layers with ascending kernel sizes is implemented, ranging from 3 × 3, 5 × 5 to 7 × 7, namely *Conv* *1*, *Conv* *2* and *Conv* *3*. A residual connection is utilized between *Conv* *1* and *Conv* *3* to refine the characteristic representations between low- and high-level features.

A concatenation layer, namely *Concat* *1*, is applied to com- bine multi-level features from the extraction process. These

features are then fed into the Gated Fusion module to make

spatial interactions, which includes a convolutional operation, namely *Conv* *4*. The output features are passed to the second concatenation layer, namely *Concat* *2*, which progressively integrates features extracted in *Conv* *3* layer. This is because higher-level information is always more global, thus being

distributed to lower levels in Gated Fusion module while mak-

ing feature interactions. This information is also indispensable for image restoration, especially for the following attention mechanism, which makes it necessary to involve the *Concat* *2* layer. The attention mechanism adaptively learns channel-wise and pixel-wise weights to enhance conducive features. After that, all features are fed into the high-resolution stage, which consists of two convolutional layers, namely *Conv* *5*, *Conv* *6*, respectively. The details of the proposed method are illustrated

in Table I.

TABLE I. Details of the LFD-Net architecture

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kernel Size Stride Padding Channel (In / Out) | | | | |
| *Conv* *1*  *Conv* *2*  *Conv* *3*  *Concat* *1*  *Conv* *4*  *Concat* *2*  *Conv* *5*  *Conv* *6* | 3  5  7 | 1  1  1 | 1  2  2 | 3 / 32  32 / 32  32 / 32 |
| *Conv* *1*, *Conv* *2*, *Conv* *1* *+* *Conv* *3*  3 1 1 96 / 3 *Conv* *3*, The Output of Gated Fusion module  64 / 16 16 / 3  3  1  1  1  1  9 | | | |

*C.* *Gated* *Fusion* *Module*

Our proposed LFD-Net replaces densely connected resid- ual blocks with effective feature-interaction based strategies. Gated Fusion module aims at making two-order interactions between multi-level features. This idea is demonstrated in Transformer-based architecture through two successive pixel- wise product in terms of K and V [36]. While Transformers are effective, they are often too computationally expensive for use in low-level pretreatment tasks. CNNs that aim to achieve equivalent results as Transformers often expand the flexibility of convolutional operations, like adding dynamic weights to improve the modeling power of convolution [33], [37], [38]. Similar techniques have been practiced in image dehazing methods [39], [40] but are still in need of further exploration and interpretation.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 5

Our proposed method also takes advantage of pixel-wise multiplication by directly implementing it to successive feature levels, the concatenation layer *Concat* *1* that combines the sequence of convolutional layers *Conv* *1*, *Conv* *2* and *Conv* *3*. To illustrate, these features are denoted as F1 , F2 and F3 respectively. Additionally, these three convolutional operations are referred to as C1 , C2 and C3 , and the *i*-th feature map of the output layer as Gi . The process of Gated Fusion module can be expressed mathematically as follows:

3

Gi = 工 Ck (F) ⊗ Fk,i

k=1

3

= 工 Ck (Fk,i ⊕ 工 Fk,j ) ⊗ Fk,i ([5)](#_bookmark1)

k=1 j i

3

= 工 Ck (Fk,i ) ⊗ Fk,i +工Ck (Fk,j ) ⊗ Fk,j ,

k=1 j i

where Fk,i is the original *i*-th feature map of the *k*-th group. As shown in equation (5), the input of Gated Fusion module consists of three levels. The number of output feature maps reduces the input by one-third, equal to the number of feature maps in each level of the input. Gated Fusion module enhances the features within a feature map with neighboring pixels and introduces interactions by dynamically assigning weights to other feature maps through pixel-wise multiplication. This reinforces the ability of convolution to retain and utilize multi- level features in an intensive and expansive manner.

*D.* *Attention* *Mechanism*

According to equation (5), the Gated Fusion module adap- tively enhances and interacts with multi-level features. How- ever, in instances where haze is not uniformly distributed, it is still challenging to accurately evaluate the haze region and thickness, which might lead to degradation in the quality of the image due to the presence of fancy shades or black spots. This is a common problem faced by many methods. Attention mechanisms, which have been designed to focus on distinctive parts when processing large amounts of information [41], can be utilized to address this issue in the image dehazing task. Specifically, channel-wise attention selects the feature levels that are particularly significant to features related to the haze region, while pixel-wise attention refines the selected haze region. In [28], attention mechanism [42] is integrated into a block structure, which is stacked in feature extraction process. Before the high-resolution stage, attention mechanism is uti- lized as an individual module to finalize feature weights. This approach might reserve a large space for weight adjustment, but also add to computation cost. In our proposed method, attention mechanism is only applied once prior to the high- resolution stage.

The adopted attention mechanism is composed of channel- wise attention and pixel-wise attention, as depicted in Fig. 3, serving as a compensation to the Gated Fusion module. All of the convolution operations used in the attention mechanism have a kernel size of 1×1, similar to a Multi-Layer Perceptron

(MLP) architecture, with global average pooling and channel- wise mixing [43]. In this mechanism, element-wise product is also used in place of absolute convolutional operations to increase the flexibility and decrease the computation complex- ity.

In detail, channel-wise attention (CA) first assigns weights to each channel by a global average pooling. The average pooling value of the *c*-th feature map, namely Mc , can be formulated as follows:

*W* *H*

 工工Mc,i,j .

Mc =

(6)

i=1 j=1

Then, two successive convolutional layers with activation layers are utilized as linear transformation to obtain a one- dimensional weight vector that element-wisely multiplies the *c*-th feature map as follows:

M = σ(C6(C(Mc ))) ⊗ Mc , (7)

where C and C are the two convolutional layers respectively, with 6(·) and σ(·) be the corresponding activation function.

Similarly, pixel-wise attention (PA) transforms the output feature maps of channel-wise attention M∗ on a pixel scale with the output namely M。derived as follows:

M。= σ(Cσ(C (M∗ ))) ⊗ M∗ , (8)

where C and C are the two convolutional operations respec- tively with σ(·) be the shared activation function.

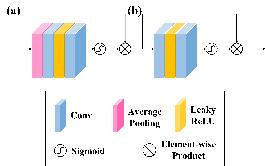


Fig. 3. The structure of Attention Mechanism. (a), (b) stand for channel-wise attention (CA) and pixel-wise attention (PA) separately.

Unlike the Gated Fusion module, which reduces the number of channels by one-third, the attention mechanism maintains an equal number of input and output channels. This suggests that the attention mechanism is able to effectively preserve the feature representation through channel-wise interaction, leading to fine-tuning of pixel-wise features with relatively low computation cost. In comparison to the approach presented in [19], which utilizes 1 × 1 convolutional layers at the begin- ning and end of the network, our method incorporates fully connected layers into the attention mechanism with element- wise product to further enhance the power of convolutional operations.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 6

*E.* *Loss* *Function*

While a combination of L1 loss, L2 loss, SSIM, or per- ceptual loss as loss functions has been shown to achieve good performance in previous works [11], [44]– [46], our experiments on LFD-Net indicate that the most widely used L2 loss, also known as Mean Squared Error (MSE), is the most

suitable loss function for LFD-Net. The L2 loss is defined as

follows:

*W* *H*

L =  工工(Is,t − Js,t )2 , (9)

s=1 t=1

where I is the input hazy image and J is the haze-free output. The intermediate value being approximated is K, which is not a direct output and therefore introduces a natural discrepancy with the output from VGG. As a result, it is not practical to utilize perceptual loss. Furthermore, the small number of parameters in the proposed method minimizes the risk of overfitting, so there is no need to introduce regularization terms. Experimental results also confirm that involving L1 loss is counterproductive.

IV. EXPERIMENTS

*A.* *Dataset*

To enhance the scientific validity and persuasiveness of our research, we utilize several widely recognized datasets for the training and evaluation of our model, including the Realistic Single Image Dehazing (RESIDE) dataset [47], the real hazy and haze-free outdoor images dataset (O-HAZE) [48], the Aerial Image dataset (AID) [49], the Remote sensing Image Cloud rEmoving (RICE) dataset [50], DAIR-V2X [51] and VisDrone2019 [52].

RESIDE is a novel benchmark for single-image de-noising consisting of synthetic and real-world hazy images. To adapt to outdoor perception vision tasks, we train LFD-Net on the Outdoor Training Set (OTS) from RESIDE, which is composed of 72,135 hazy images with atmospheric lighting ranging from 0.8 to 1.0 and scattering parameters from 0.02 to 0.4. The majority of the hazy images generated by OTS do not exhibit a halo effect, which is a strong advantage as a training set. The test set of RESIDE consists of the Synthetic Objective Testing Set (SOTS) and the Hybrid Subjective Testing Set (HSTS), which we utilize to evaluate the performance of our model using 492 synthetic outdoor images from SOTS and 10 real-world hazy outdoor images from HSTS.

To further assess the generalization of our proposed method, we also conduct evaluations on O-HAZE, which consists of 45 real outdoor scenes recorded over an extended period of time, spanning more than eight weeks under both cloudy and sunny conditions. We also randomly select hazy images from the internet to supplement our test results. Additionally, to verify the effectiveness of LFD-Net in the domain of remote sensing, we fine-tune the pretrained model on ordinary outdoor scenarios using the Aerial Image dataset (AID) [49], a large- scale dataset for aerial scene classification containing 10,000 images. We utilize the first part of Remote sensing Image Cloud rEmoving dataset (RICE) [50], comprising 500 pairs

of hazy and clear images from various regions of the world under diverse conditions for test.

To validate the effectiveness of our method for real-time vision tasks, we evaluate its performance on the DAIR- V2X dataset [51] for ordinary outdoor scenarios and the VisDrone2019 dataset [52] for remote sensing scenarios. The DAIR-V2X dataset is the first large-scale, multi-modality, multi-view dataset for Vehicle-Infrastructure Cooperative Au- tonomous Driving, consisting of 71,254 LiDAR frames and camera frames from real-world scenarios. The VisDrone2019 dataset consists of 8,599 images captured by drone platforms in various locations at different heights, with over 540,000 annotations in ten categories.

TABLE II. Average Comparison of Metrics on SOTS for 492 JPG Images

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | PSNR↑ | SSIM↑ | CIEDE↓ | ∆SSEQ | ↓ | FPS↑ |
| AOD-Net | 19.45 | 0.8593 | 7.12 | 0.0080 |  | 56.85 |
| GridDehazeNet | 25 07 | 0.9108 | 4.02 | 0.0074 |  | 10.04 |
| Wavelet U Net | 23 73 | 0 8661 | 4 93 | 0 0064 |  | 8 28 |
| GCA-Net | 22.13 | 0.8766 | 6.19 | 0.0058 |  | 23.01 |
| FFA-Net | 19.36 | 0.8472 | 7.70 | 0.0131 |  | 6.91 |
| LD-Net | 18 62 | 0 8380 | 7 52 | 0 0075 |  | 45 44 |
| D4 | 18 09 | 0 6668 | 8 36 | 0 0624 |  | 44 79 |
| LFD-Net | 25.12 | 0.9087 | 4.24 | 0.0054 |  | 54.41 |
| TABLE III. Average Comparison of Metrics on O-HAZE for 45 JPG Images | | | | | | |
|  | PSNR↑ | SSIM↑ | CIEDE↓ | ∆SSEQ ↓ | | FPS↑ |
| AOD-Net GridDehazeNet Wavelet-U-Net | 15.06  16.68  15.87 | 0.5412 0.6361 0.5058 | 17.53  13.54  14.93 | 0.0146  0.0055  0.0041 | | 13.93  1.09  8.66 |
| GCA-Net | 17.24 | 0.6523 | 13.81 | 0.0077 | | 4.92 |
| FFA-Net | 14 62 | 0 5881 | 14 72 | 0 0067 | | 1 52 |
| LD-Net | 14.72 | 0.5583 | 16.74 | 0.0057 | | 10.62 |
| D4 | 11.51 | 0.2564 | 18.58 | 0.1612 | | 2.70 |
| LFD-Net | 17.67 | 0.6532 | 11.80 | 0.0041 | | 14.65 |

TABLE IV. Average Comparison of Metrics on RICE1 for 500 PNG Images

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | PSNR↑ | SSIM↑ | CIEDE↓ | ∆SSEQ | ↓ | FPS↑ |
| AOD-Net | 14.80 | 0.6578 | 16.73 | 0.0508 |  | 58.62 |
| GridDehazeNet | 19 14 | 0 8351 | 11 44 | 0 0021 |  | 5 08 |
| GCA Net | 18 35 | 0 7237 | 15 25 | 0 0075 |  | 23 61 |
| FFA-Net | 19.92 | 0.8117 | 10.39 | 0.0029 |  | 7.06 |
| MSBDN | 19.77 | 0.8477 | 10.65 | 0.0022 |  | 21.38 |
| D4 | 19 29 | 0 8258 | 12 21 | 0 0202 |  | 15 59 |
| DehazeFormer | 19 95 | 0.8615 | 10.82 | 0.0018 |  | 4.29 |
| LFD-Net | 30.88 | 0.9420 | 3.32 | 0.0008 |  | 45.51 |

*B.* *Experiment* *Results*

We faithfully reproduce 7 state-of-the-art methods for or-

dinary outdoor scenarios, including AOD-Net [15], GridDe- hazeNet [27], Wavelet-U-Net [22], GCA-Net [40], FFA-Net [28], LD-Net [19] and D4 [53]. All the experiments are conducted on a PC with an R9-5900HX CPU (E5-1650) and an NVIDIA RTX-3080 GPU. The quantitative comparison results on the outdoor SOTS and O-HAZE datasets can be found in Tables II and III, respectively. The visual comparison results from the outdoor SOTS and O-HAZE datasets are

shown in Fig. 4 and Fig. 5. Furthermore, we also conduct experiments using real-world hazy images with no reference both from HSTS and images randomly selected on the Internet, as depicted in Fig. 6 and Fig. 7.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 7



Hazy Image AOD-Net [15] GridDehazeNet [27] Wavelet-U-Net [22] GCA-Net [40]



FFA-Net [28] LD-Net [19] D4 [53] Ours Ground Truth



Hazy Image AOD-Net [15] GridDehazeNet [27] Wavelet-U-Net [22] GCA-Net [40]



FFA-Net [28] LD-Net [19] D4 [53] Ours Ground Truth



Hazy Image AOD-Net [15] GridDehazeNet [27] Wavelet-U-Net [22] GCA-Net [40]



FFA-Net [28] LD-Net [19] D4 [53] Ours Ground Truth



Hazy Image AOD-Net [15] GridDehazeNet [27] Wavelet-U-Net [22] GCA-Net [40]



FFA-Net [28] LD-Net [19] D4 [53] Ours Ground Truth

Fig. 4. Visual Comparison on Outdoor SOTS. We compare our methods with AOD-Net [15], GridDehazeNet [27], Wavelet-U-Net [22], GCA-Net [40], FFA-Net [28], LD-Net [19] and D4 [53]. Our proposed method exhibits adaptability to diverse scenarios and possesses a noteworthy level of generalization.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 8



Hazy Image AOD-Net [15] GridDehazeNet [27] Wavelet-U-Net [22] GCA-Net [40]



FFA-Net [28] LD-Net [19] D4 [53] Ours Ground Truth



Hazy Image AOD-Net [15] GridDehazeNet [27] Wavelet-U-Net [22] GCA-Net [40]



FFA-Net [28] LD-Net [19] D4 [53] Ours Ground Truth



Hazy Image AOD-Net [15] GridDehazeNet [27] Wavelet-U-Net [22] GCA-Net [40]



FFA-Net [28] LD-Net [19] D4 [53] Ours Ground Truth



Hazy Image AOD-Net [15] GridDehazeNet [27] Wavelet-U-Net [22] GCA-Net [40]



FFA-Net [28] LD-Net [19] D4 [53] Ours Ground Truth

Fig. 5. Visual Comparison Results on O-HAZE. We compare our methods with AOD-Net [15], GridDehazeNet [27], Wavelet-U-Net [22], GCA-Net [40], FFA-Net [28], LD-Net [19] and D4 [53]. AOD-Net and LD-Net produce relatively dark in visual quality. GCA-Net performs well on irregular haze but suffers from inconsistency in color blocks. Our proposed method exhibits adaptability to diverse scenarios and possesses a noteworthy level of generalization.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 9



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 6. Visual Comparison Results on Real-world HSTS. (a) Hazy image, (b) AOD-Net [15], (c) GridDehazeNet [27], (d) Wavelet-U-Net [22], (e) GCA-Net [40], (f) FFA-Net [28], (g) LD-Net [19], (h) D4 [53] and (i) Ours. Our proposed method exhibits adaptability to diverse scenarios and possesses a noteworthy level of generalization.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 7. Visual Comparison Results on Randomly Selected Real-world Images. (a) Hazy image, (b) AOD-Net [15], (c) GridDehazeNet [27], (d) Wavelet-U-Net [22], (e) GCA-Net [40], (f) FFA-Net [28], (g) LD-Net [19], (h) D4 [53] and (i) Ours. Our proposed method exhibits adaptability to diverse scenarios and possesses a noteworthy level of generalization.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 10



AOD-Net [15] GridDehazeNet [27] GCA-Net [40] FFA-Net [28]

Hazy Image



D4 [53] DehazeFormer [55] Ours Ground Truth

MSBDN [54]



AOD-Net [15] GridDehazeNet [27] GCA-Net [40] FFA-Net [28]

Hazy Image



D4 [53] DehazeFormer [55] Ours Ground Truth

MSBDN [54]



AOD-Net [15] GridDehazeNet [27] GCA-Net [40] FFA-Net [28]

Hazy Image



D4 [53] DehazeFormer [55] Ours Ground Truth

MSBDN [54]



AOD-Net [15] GridDehazeNet [27] GCA-Net [40] FFA-Net [28]

Hazy Image



MSBDN [54] D4 [53] DehazeFormer [55] Ours Ground Truth

Fig. 8. Visual Comparison Results on O-HAZE. We compare our methods with AOD-Net [15], GridDehazeNet [27], Wavelet-U-Net [22], GCA-Net [40], FFA-Net [28], LD-Net [19] and D4 [53]. AOD-Net and LD-Net produce relatively dark in visual quality. GCA-Net performs well on irregular haze but suffers from inconsistency in color blocks. Our proposed method exhibits adaptability to diverse scenarios and possesses a noteworthy level of generalization.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 11



The 1st Output Map of *Conv* *4* The 2nd Output Map of *Conv* *4* The 3rd Output Map of *Conv* *4*

(a) The Output of Convolutional Operation in Gated Fusion Module



The 1st Level of *Concat* *1* The 2nd Level of *Concat* *1* The 3rd Level of *Concat* *1* Output of Gated Fusion

(b) The 15th Feature Map of Each Level of *Concat* *1* and the Output of Gated Fusion Module



The 1st Level of *Concat* *1* The 2nd Level of *Concat* *1* The 3rd Level of *Concat* *1* Output of Gated Fusion

(c) The 30th Feature Map of Each Level of *Concat* *1* and the Output of Gated Fusion Module



The 1st Level of *Concat* *1* The 2nd Level of *Concat* *1* The 3rd Level of *Concat* *1* Output of Gated Fusion

(d) The 31st Feature Map of Each Level of *Concat* *1* and the Output of Gated Fusion Module

Fig. 9. Visualization Results of the changes in layers before and after the Gated Fusion Module: (a) represents the three output feature maps of the convolutional operation *Conv* *4* incorporated into Gated Fusion module, (b)-(d) stand for the changes in the 15th, 30th and 31st feature map of the layers respectively. (b) shows that the contrast of the image is enhanced, resulting in distant objects becoming more distinct. (c) and (d) show more abstract feature representations, which are significantly shifted compared to the input. Specifically, (c) emphasizes the outline of substances, while (d) highlights the blocks within substances.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 12



Ordinary Condition Simulated Hazy Condition Dehazing Condition

(a) Comparison of Object Detection Results Under Different Conditions from DAIR-V2X



Ordinary Hazy Dehazing Ordinary Hazy Dehazing

(b) Sub-scene 1 (c) Sub-scene 2



Ordinary Hazy Dehazing Ordinary Hazy Dehazing (d) Sub-scene 3 (e) Sub-scene 4

Fig. 10. Reference Object Detection Results: (a) comparison of object detection results under ordinary, simulated hazy, and dehazed conditions, (b)-(e) detailed sub-scenes of detection results. (b) and (c) show an improvement in the detection rate. (d) corrects the error of mistaking a roadblock for a car in the hazy condition. (e) shows the detection of another car compared to the ground-truth clear image.

In the remote sensing domain, as far as we are aware, the pretrained models of dehazing methods are not publicly available. However, we also reproduce 7 state-of-the-art meth- ods using default outdoor weights, including AOD-Net [15], GridDehazeNet [27], GCA-Net [40], FFA-Net [28], MSBDN [54], D4 [53] and DehazeFormer [55]. As expected, AOD-Net has limited performance due to its small number of parameters,

while the other methods show similar performance before being fine-tuned. We make our pretrained model open publicly for further comparison.

It can be observed that a large majority of networks might suffer from obvious inconsistency within color blocks or misrepresent original information, as reflected in terms of CIEDE2000 and ∆SSEQ. Specifically, other lightweight methods such as AOD-Net [15] and LD-Net [19] produce relatively dark visual quality, with colors appearing darker than in the original images, making it difficult to distinguish objects. In the synthetic SOTS dataset where haze is more uniformly distributed, GCA-Net [40] encounters severe color shift occasionally. While in realistic scenarios like O-HAZE dataset, where the haze is thick and irregular, GCA-Net [40] provides remarkable dehazing effect, but still suffers from slight halo effects. FFA-Net [28] performs well on specific datasets, but is not generalizable enough for shifted domains, lacking dehazing capability. It can be concluded that the

incorporation of modules like attention mechanism prevents the image from getting uniformly dehazed without region dis- crepancy, compared to networks with absolute convolutional and concatenation layers. However, the stack of sophisticated modules might also lead to overfitting to some extent, not flexible enough for real-world vision tasks.

*C.* *Ablation* *Study*

The experiment results confirm us that our LFD-Net is effective and efficient for real-time applications. Since it has a different principle compared to that of other methods, we conduct a series of ablation study to ensure that each compo- nent of the network is indispensable. The detailed experiment conditions and the corresponding metrics tested on the outdoor

SOTS are listed in Table V.

Inspired by [15] and [19], we add a second concatenation layer, referred to as *Concat* *2*, to our method. In Case 1, we omit *Concat* *2* and observe a slight loss of detailed texture information due to the reduced high-level information.

In Cases 2, 3, and 4, we investigate the importance of the Gated Fusion module or attention mechanism in our model.

These cases demonstrate that these two sub-networks work to- gether to facilitate feature interaction. Specifically, the removal of the attention mechanism leads to the occasional appearance of black spots on the images, significantly lowering the overall

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 13



Ordinary Condition Simulated Hazy Condition Dehazing Condition

(a) Comparison of Object Detection Results Under Different Conditions from VisDrone2019



Ordinary Hazy Dehazing Ordinary Hazy Dehazing

(b) Sub-scene 1 (c) Sub-scene 2



Ordinary Hazy Dehazing Ordinary Hazy Dehazing (d) Sub-scene 3 (e) Sub-scene 4

Fig. 11. Reference Remote Sensing Object Detection Results: (a) comparison of remote sensing object detection results under ordinary, simulated hazy, and dehazed conditions, (b)-(e) detailed sub-scenes of detection results, in which the detection rate for pedestrians is enhanced to a large extent. In particular, (b)

and (e) highlight instances of pedestrians that are not visible in the ordinary conditions but are detected after dehazing, similar to the results from DAIR-V2X.

TABLE V. Ablation Experiment of LFD-Net on Outdoor SOTS Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Concat* *2* | Gated Fusion Module | Attention Mechanism PSNR↑ | SSIM↑ | CIEDE↓ | ∆SSEQ ↓ |
| Case 1 |  | 24.35 | 0.9062 | 5.09 | 0.0055 |
| Case 2 |  | 23.33 | 0.8910 | 5.56 | 0.0066 |
| Case 3 |  | 21.62 | 0.8642 | 6.65 | 0.0116 |
| Case 4 |  | 23.17 | 0.8878 | 5.54 | 0.0076 |
| Default |  | 25.12 | 0.9087 | 4.24 | 0.0054 |

performance. In comparison to other lightweight methods, our method partially addresses this issue. Additionally, the Gated Fusion module is a crucial component in enhancing the dehazing capability, acting as a bridge between the multi-level feature extraction process, which ends at the first concatenation layer *Concat* *1*, and the attention mechanism, which begins at the second concatenation layer *Concat* *2*. When both

the attention mechanism and the Gated Fusion module are

involved, the detailed information in the images is further refined, making it more authentic and faithful to the original information. This structure helps to preserve and interact with multi-level information to improve the overall image quality.

*D.* *Visualization* *Results*

We have visualized the intermediate feature maps before and after the Gated Fusion Module, as depicted in Fig. 9. As shown in (a), the incorporated convolutional layer combines features of three levels from *Conv* *1*, *Conv* *2* and *Conv* *1* *+* *Conv* *3* to

generate three distinctive feature maps. They distinguish from each other in terms of the focus on close or distant objects and the lightness or contrast of pixels.

In Fig. 9 (b)-(d), we demonstrate the changes in specific feature maps after the Gated Fusion module. Fig. 9 (b) shows that the contrast of the image is enhanced with the hierarchi- cal information, resulting in distant objects becoming more distinct. Fig. 9 (c) and Fig. 9 (d) show more abstract feature representations, which are significantly shifted compared to the input features. Specifically, Fig. 9 (c) emphasizes the outline of substances, while Fig. 9 (d) highlights the blocks within substances.

Gated Fusion module reallocates the distributed feature

representations of the multi-level layers through feature- interaction strategies. The feature extraction process is com- pressed to three successive convolutional layers, for which we make compensation through intra-level enhancement and inter- level combination.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 14

*E.* *Application* *for* *Object* *Detection* *Task*

Haze, as a severe weather condition, can significantly reduce the effectiveness of a surveillance system. For instance, the object detection results in automatic driving application suffer from the hazy environments, which might be put into a risky situation due to the degradation of the image quality. Therefore, a pretreatment procedure of image enhancement before performing those tasks is of great significance. As far as we know, there is no dataset with built-in synthetic fog images for object detection. In our experiment, we randomly select

100 images from each dataset (DAIR-V2X and VisDrone2019) and produce their synthetic hazy versions. We use the default outdoor pretrained weight for the former, while the fine-tuned remote sensing pretrained weight for the latter. Both object detection processes are based on YOLOv5. Our experiment results show that the mean Average Precision when IoU = 0.5 (mAP@0.5) of the dehazed condition improves by 4.73% compared to the hazy condition in DAIR-V2X, while by 0.81% in VisDrone2019.

Furthermore, overall detection result of a particular scene is shown in Fig. 10 (a), while Fig. 10 (b)-(e) illustrate the most representative perspectives of the dehazing effect. In Fig. 10 (b) and (c), it can be seen that dehazing improves the detection rate. In Fig. 10 (d), the roadblock is mistakenly identified as a car in the hazy condition, but the dehazing method is able to correct this error. In Fig. 10 (e), another car instance is shown before and after dehazing the synthetic hazed image.

In Fig. 11 (a), we show the overall remote sensing object detection results from the perspective of a drone in a particular scene. Compared to the driving perspective from DAIR-V2X, the drone faces more complex screen layouts and smaller objects, which pose challenges for both dehazing and object detection methods. Fig. 11 (b)-(e) illustrate the difficulties object detection methods encounter when detecting smaller pedestrian instances, especially in hazy conditions. However, dehazing methods can partially address this issue and enhance the detection rate of small objects like pedestrians, as shown in Fig. 11 (c) and (d). In Fig. 10 (b) and (e), two additional pedestrian instances are detected after dehazing compared to the original conditions, similar to that in Fig. 10 (e). This suggests that haze can have unpredictable effects on normal conditions, but our method can find a better solution com- pared to the ground truth in representing high-level semantic information to some extent.

V. CONCLUSION

In this paper, we present a novel end-to-end image dehazing model called LFD-Net. As a pretreatment to downstream vision tasks, it not only ensures the effectiveness and efficiency required for real-time applications, but also outperforms state- of-the-art methods in terms of region-balance and color- fidelity. By designing this framework, we demonstrate the potential of CNN-based network through making two-order spatial interaction. Specifically, we show that the capabilities of these networks can be enhanced not only by adding more complex modules to create deeper networks, but also by effectively combining individual and natural feature extrac- tion, fusion, and attention with feature interaction strategies,

particularly in the field of image super-resolution. Our testing on various scenarios also shows that reducing the number of parameters in a network can help mitigate overfitting, which might be conducive for future network design.

REFERENCES

[1] A. Makarau, R. Richter, R. Mller, and P. Reinartz,“Haze detection and

removal in remotely sensed multispectral imagery,”*IEEE* *Transactions* *on* *Geoscience* *and* *Remote* *Sensing*, vol. 52, no. 9, pp. 5895–5905, 2014.

[2] Y. Zheng, J. Su, S. Zhang, M. Tao, and L. Wang, “Dehaze-aggan:

Unpaired remote sensing image dehazing using enhanced attention-guide generative adversarial networks,”*IEEE* *Transactions* *on* *Geoscience* *and* *Remote* *Sensing*, vol. 60, pp. 1– 13, 2022.

[3] Y. Han, M. Yin, P. Duan, and P. Ghamisi, “Edge-preserving filtering-

based dehazing for remote sensing images,”*IEEE* *Geoscience* *and* *Remote* *Sensing* *Letters*, vol. 19, pp. 1–5, 2021.

[4] K. He, J. Sun, and X. Tang, “Single image haze removal using dark

channel prior,”*IEEE* *transactions* *on* *pattern* *analysis* *and* *machine* *intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[5] F. Liu, Y. Lv, B. Li, S. Gao, and Y. Qin,“A semiphysical approach

of haze removal for landsat image,”*IEEE* *Journal* *of* *Selected* *Topics* *in* *Applied* *Earth* *Observations* *and* *Remote* *Sensing*, vol. 14, pp. 7410– 7421, 2021.

[6] B. Xie, J. Yang, J. Shen, and Z. Lv,“Image defogging method combining

light field depth estimation and dark channel,”in *2021* *IEEE* *Interna-* *tional* *Conference* *on* *Consumer* *Electronics* *and* *Computer* *Engineering*

*(ICCECE)*. IEEE, 2021, pp. 745–749.

[7] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “Dehazenet: An end-to-end

system for single image haze removal,”*IEEE* *Transactions* *on* *Image* *Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[8] Y. Liao, Z. Su, X. Liang, and B. Qiu, “Hdp-net: Haze density prediction

network for nighttime dehazing,”in *Pacific* *Rim* *Conference* *on* *Multi-* *media*. Springer, 2018, pp. 469–480.

[9] R. Li, R. T. Tan, and L.-F. 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.

[10] L. Jiao, C. Hu, L. Huo, and P. Tang, “Guided-pix2pix: End-to-end

inference and refinement network for image dehazing,”*IEEE* *Journal* *of* *Selected* *Topics* *in* *Applied* *Earth* *Observations* *and* *Remote* *Sensing*, vol. 14, pp. 3052–3069, 2021.

[11] Y. Guo, J. Chen, X. Ren, A. Wang, and W. Wang, “Joint raindrop

and haze removal from a single image,”*IEEE* *Transactions* *on* *Image* *Processing*, vol. 29, pp. 9508–9519, 2020.

[12] Y. Li, Y. Liu, Q. Yan, and K. Zhang, “Deep dehazing network with latent

ensembling architecture and adversarial learning,”*IEEE* *Transactions* *on* *Image* *Processing*, vol. 30, pp. 1354– 1368, 2020.

[13] F. Ding, K. Yu, Z. Gu, X. Li, and Y. Shi, “Perceptual enhancement for

autonomous vehicles: restoring visually degraded images for context prediction via adversarial training,”*IEEE* *Transactions* *on* *Intelligent* *Transportation* *Systems*, 2021.

[14] H. Tang, H. Liu, D. Xu, P. Torr, and N. Sebe,“Attentiongan: Unpaired

image-to-image translation using attention-guided generative adversarial networks,”*IEEE* *Transactions* *on* *Neural* *Networks* *and* *Learning* *Sys-* *tems*, 2019.

[15] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng,“Aod-net: All-in-one

dehazing network,”in *Proceedings* *of* *the* *IEEE* *international* *conference* *on* *computer* *vision*, 2017, pp. 4770–4778.

[16] W. Qian, C. Zhou, and D. Zhang,“Faod-net: A fast aod-net for dehazing

single image,”*Mathematical* *Problems* *in* *Engineering*, vol. 2020, pp. 1– 11, 2020.

[17] S. Chen, J. Cheng, and Z. Huang, “Gado-net: An improved aod-net

single image dehazing algorithm,”in *2021* *3rd* *International* *Academic* *Exchange* *Conference* *on* *Science* *and* *Technology* *Innovation* *(IAECST)*,

2021, pp. 640–646.

[18] J. Zhang and D. Tao, “Famed-net: A fast and accurate multi-scale end-

to-end dehazing network,”*IEEE*, no. 1, 2020.

[19] H. Ullah, K. Muhammad, M. Irfan, S. Anwar, M. Sajjad, A. S. Imran,

and V. H. C. de Albuquerque, “Light-dehazenet: a novel lightweight cnn architecture for single image dehazing,”*IEEE* *Transactions* *on* *Image* *Processing*, vol. 30, pp. 8968–8982, 2021.

[20] O. Ronneberger, P. Fischer, and T. Brox,“U-net: Convolutional networks

for biomedical image segmentation,”in *International* *Conference* *on* *Medical* *image* *computing* *and* *computer-assisted* *intervention*. Springer, 2015, pp. 234–241.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 15

[21] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and 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] H.-H. Yang and Y. Fu, “Wavelet u-net and the chromatic adaptation

transform for single image dehazing,”in *2019* *IEEE* *International* *Conference* *on* *Image* *Processing* *(ICIP)*. IEEE, 2019, pp. 2736–2740.

[23] T. Feng, C. Wang, X. Chen, H. Fan, K. Zeng, and Z. Li, “Urnet: A u-net

based residual network for image dehazing,”*Applied* *Soft* *Computing*, vol. 102, p. 106884, 2021.

[24] B.-U. Lee, K. Lee, J. Oh, and I. S. Kweon, “Cnn-based simultaneous

dehazing and depth estimation,”in *2020* *IEEE* *International* *Conference* *on* *Robotics* *and* *Automation* *(ICRA)*. IEEE, 2020, pp. 9722–9728.

[25] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, “Underwater

image enhancement via medium transmission-guided multi-color space embedding,”*IEEE* *Transactions* *on* *Image* *Processing*, vol. 30, pp. 4985– 5000, 2021.

[26] A. Mehra, M. Mandal, P. Narang, and V. Chamola, “Reviewnet: A

fast and resource optimized network for enabling safe autonomous driving in hazy weather conditions,”*IEEE* *Transactions* *on* *Intelligent* *Transportation* *Systems*, vol. 22, no. 7, pp. 4256–4266, 2020.

[27] X. Liu, Y. Ma, Z. Shi, and 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.

[28] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. 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.

11 908– 11 915.

[29] X. Zhang, T. Wang, W. Luo, and P. Huang, “Multi-level fusion and

attention-guided cnn for image dehazing,”*IEEE* *Transactions* *on* *Circuits* *and* *Systems* *for* *Video* *Technology*, vol. 31, no. 11, pp. 4162–4173, 2020.

[30] H. Bai, J. Pan, X. Xiang, and J. Tang, “Self-guided image dehazing using

progressive feature fusion,”*IEEE* *Transactions* *on* *Image* *Processing*, vol. 31, pp. 1217– 1229, 2022.

[31] G. Gao, J. Cao, C. Bao, Q. Hao, A. Ma, and G. Li,“A novel transformer-

based attention network for image dehazing,”*Sensors*, vol. 22, no. 9, p. 3428, 2022.

[32] Y. Yang, H. Zhang, X. Wu, and X. Liang, “Mstfdn: Multi-scale trans-

former fusion dehazing network,”*Applied* *Intelligence*, pp. 1– 12, 2022.

[33] Y. Rao, W. Zhao, Y. Tang, J. Zhou, S.-N. Lim, and J. Lu, “Hornet: Effi-

cient high-order spatial interactions with recursive gated convolutions,” *arXiv* *preprint* *arXiv:2207.14284*, 2022.

[34] M. R. Luo, G. Cui, and B. Rigg, “The development of the cie 2000

colour-difference formula: Ciede2000,”*Color* *Research* *&* *Application*, vol. 26, no. 5, pp. 340–350, 2001.

[35] L. Liu, B. Liu, H. Huang, and A. C. Bovik,“No-reference image quality

assessment based on spatial and spectral entropies,”*Signal* *processing:* *Image* *communication*, vol. 29, no. 8, pp. 856–863, 2014.

[36] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez,

Ł . Kaiser, and I. Polosukhin,“Attention is all you need,”*Advances* *in* *neural* *information* *processing* *systems*, vol. 30, 2017.

[37] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,”in

*Proceedings* *of* *the* *IEEE* *conference* *on* *computer* *vision* *and* *pattern* *recognition*, 2018, pp. 7132–7141.

[38] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, “Dynamic

convolution: Attention over convolution kernels,”in *Proceedings* *of* *the* *IEEE/CVF* *Conference* *on* *Computer* *Vision* *and* *Pattern* *Recognition*, 2020, pp. 11 030– 11 039.

[39] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang,

“Gated fusion network for single image dehazing,”in *Proceedings* *of* *the* *IEEE* *conference* *on* *computer* *vision* *and* *pattern* *recognition*, 2018, pp. 3253–3261.

[40] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and

G. Hua, “Gated context aggregation network for image dehazing and deraining,”in *2019* *IEEE* *winter* *conference* *on* *applications* *of* *computer* *vision* *(WACV)*. IEEE, 2019, pp. 1375– 1383.

[41] Z. Niu, G. Zhong, and H. Yu,“A review on the attention mechanism of

deep learning,”*Neurocomputing*, vol. 452, pp. 48–62, 2021.

[42] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “Cbam: Convolutional

block attention module,”in *Proceedings* *of* *the* *European* *conference* *on* *computer* *vision* *(ECCV)*, 2018, pp. 3– 19.

[43] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai,

T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit,

M. Lucic, and A. Dosovitskiy, “Mlp-mixer: An all-mlp architecture for vision,”in *Advances* *in* *Neural* *Information* *Processing* *Systems*,

M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W.

Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 24 261– 24 272. [Online]. Available: https://proceedings.neurips.cc/paper/2021/ file/cba0a4ee5ccd02fda0fe3f9a3e7b89fe-Paper.pdf

[44] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, “Loss functions for image

restoration with neural networks,”*IEEE* *Transactions* *on* *computational* *imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[45] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep

residual networks for single image super-resolution,”in *Proceedings* *of* *the* *IEEE* *conference* *on* *computer* *vision* *and* *pattern* *recognition* *workshops*, 2017, pp. 136– 144.

[46] M. S. Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. K. Ekenel, and

J.-P. Thiran, “Srobb: Targeted perceptual loss for single image super- resolution,”in *Proceedings* *of* *the* *IEEE/CVF* *International* *Conference* *on* *Computer* *Vision*, 2019, pp. 2710–2719.

[47] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang,

“Benchmarking single-image dehazing and beyond,”*IEEE* *Transactions* *on* *Image* *Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[48] C. O. Ancuti, C. Ancuti, R. Timofte, and 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.

[49] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu,

“Aid: A benchmark data set for performance evaluation of aerial scene classification,”*IEEE* *Transactions* *on* *Geoscience* *and* *Remote* *Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[50] D. Lin, G. Xu, X. Wang, Y. Wang, X. Sun, and K. Fu,“A remote sensing

image dataset for cloud removal,”*arXiv* *preprint* *arXiv:1901.00600*,

2019.

[51] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li,

X. Hu, J. Yuan *et* *al.*, “Dair-v2x: A large-scale dataset for vehicle- infrastructure cooperative 3d object detection,”in *Proceedings* *of* *the* *IEEE/CVF* *Conference* *on* *Computer* *Vision* *and* *Pattern* *Recognition*, 2022, pp. 21 361–21 370.

[52] D. Du, P. Zhu, L. Wen, X. Bian, H. Lin, Q. Hu, T. Peng, J. Zheng,

X. Wang, Y. Zhang *et* *al.*, “Visdrone-det2019: The vision meets drone object detection in image challenge results,”in *Proceedings* *of* *the* *IEEE/CVF* *international* *conference* *on* *computer* *vision* *workshops*, 2019, pp. 0–0.

[53] Y. Yang, C. Wang, R. Liu, L. Zhang, X. Guo, and D. Tao, “Self-

augmented unpaired image dehazing via density and depth decomposi- tion,”in *Proceedings* *of* *the* *IEEE/CVF* *Conference* *on* *Computer* *Vision* *and* *Pattern* *Recognition*, 2022, pp. 2037–2046.

[54] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and 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.

[55] Y. Song, Z. He, H. Qian, and X. Du, “Vision transformers for single

image dehazing,”*arXiv* *preprint* *arXiv:2204.03883*, 2022.

Yizhu Jin is currently a final year undergraduate stu-

dent in the School of Automation Science and Elec- trical Engineering from Beihang University, Beijing, China. Jin’s research interests lie in the intersection

between valuable application and general methods with high interpretability in Computer Vision and Artificial Intelligence.

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2021 16

Jiaxing Chen received M.S. degree in Electrical and

Computer Engineering from University of Illinois,

Chicago, USA, in 2021 and worked as an Algorithm

engineer in National Innovation Center of Intelligent

and Connected Vehicles in 2022. He is currently a

researcher with Tsinghua University, Beijing, China.

Chen’s research interests include Computer Vision

with Deep Learning, Machine Learning, Multimodal

Fusion, and Trajectory Prediction.

Feng Tian received her B.S. degree in engineer-

ing in 2018. She is currently studying in the Col-

lege of Traffic and Logistics Engineering, Xinjiang

Agricultural University, Xinjiang, China, pursuing

a Postgraduate degree in transportation engineering.

Tian’s research interests include Machine Learning,

Data Mining, Intelligent Transportation, and Target

Detection.

Kun Hu received the B.Sc. degree in Remote Sens-

ing Science and Technology from Wuhan University,

Wuhan, China, in 2010, the M.Sc. and Ph.D. de-

gree in Photogrammetry and Remote Sensing from

Wuhan University in 2012 and 2016, respectively.

He worked as an assistant professor with the In-

stitute of Electronics, Chinese Academy of Sciences,

Beijing, China from 2012 to 2017, as a post-doctor

with the Ohio State University, Ohio, USA, from

2017 to 2019, and as an associate professor with the

Aerospace Information Research Institute, Chinese

Academy of Sciences, Beijing, China from 2019 to 2022.

He is currently an associate professor with the Institute of Artificial Intelli-

gence, Beihang University, Beijing, China. His research interests focus on the

accurate processing and intelligent application of multi-source remote sensing

data, such as camera calibration, image production, information fusion, target

detection, clarification, 3D reconstruction and quality evaluation.