**PAPER**

An Improved U-Net Architecture for Image Dehazing

Wenyi GE†, Member, Yi LIN†††, Zhitao WANG††††, Guigui WANG††††, and Shihan TAN††††, Nonmembers

**SUMMARY** In this paper, we present a simple yet powerful deep neural network for natural image dehazing. The proposed method is designed based on U-Net architecture and we made some design changes to make it better. We first use Group Normalization to replace Batch Normalization to solve the problem of insufficient batch size due to hardware limitations. Second, we introduce FReLU activation into the U-Net block, which can achieve capturing complicated visual layouts with regular convolutions. Experimental results on public benchmarks demonstrate the effectiveness of the modified components. On the SOTS Indoor and Outdoor datasets, it obtains PSNR of 32.23 and 31.64 respectively, which are comparable performances with state-of-the-art methods. The code is publicly available online soon.

**key words:** image dehaze, deep learning, fully convolutional networks, U-Net

1. Introduction

Due to the existence of dust, mist, or smoke, the natural light is greatly absorbed and scattered and thereby affects the quality of the image collected under those conditions. In this way, the complicated noise is imposed on the images, which causes quality degradation and image visibility. Those images not only severely hinder human viewing, but also reduce the performance of the following computer vision tasks, such as object detection, classification, and recognition. A typical case is the remoting sensing image, which is easy to be a barrier for computer vision tasks, such as object detection, classification, and recognition. A typical case is the remoting sensing image, which is easy to be a barrier for computer vision tasks, such as object detection, classification, and recognition. A typical case is the remoting sensing image, which is easy to be a barrier for computer vision tasks, such as object detection, classification, and recognition. A typical case is the remoting sensing image, which is easy to be a barrier for computer vision tasks, such as object detection, classification, and recognition.

A high-resolution remote sensing image makes humans to better observe the earth. It has been widely used in various applications such as biophysical estimation, temperature retrieval, multi-specialist architecture, and target detection. However, the images are usually degraded by the turbid medium in the atmosphere during the imaging formation process.

Therefore, image dehazing has always been a hot topic and considered by researchers all over the world. In general, the goal of image dehazing is to restore a clean scene from a hazy image. The atmospheric scattering model has been the classical description for the hazy image generation:

\[
I(x) = t(x)J(x) + (1 - t(x))A
\]

where \(I\) is a hazy image, \(J\) denotes a haze-free image, \(A\) describes the global atmospheric light indicating the intensity of ambient light, \(t\) is the transmission map which describes the portion of the light that can reach the camera without being scattered, and \(x\) represents a pixel or pixel block position. According to the Lambert Beer law, the transmission map \(t\) is given by:

\[
t(x) = e^{-\beta d(x)}
\]

where \(\beta\) is the scattering coefficient of the atmosphere, and \(d(x)\) is the distance between the object and the camera. Thus, we can rewrite the model for the clean image as the output:

\[
J(x) = \frac{1}{t(x)}I(x) - A\frac{1}{t(x)} + A
\]

From the above formulations, we can notice that if we estimate the global atmosphere \(A\) and transmission map \(t\) properly for a captured hazy image, we can restore a clear haze-free image. But the first step is a very challenging problem because both of them are often unknown in real scenarios.

Learning from existing state-of-the-art works, the image dehazing task is achieved by a pixel-to-pixel mapping, in which a deep learning model is designed to build the mapping patterns. In this work, we present a simple yet powerful deep neural network for natural image dehazing. Our method is based on U-Net[1] architecture and we integrate some of the latest design methods into the network to make it better. We first adopt Group Normalization (GN) [2] to replace Batch Normalization (BN) [3] to solve the problem of insufficient batch size. Normally, BN has been established as a very effective component in deep learning. But due to hardware limitations, we have to reduce the batch size in actual operation, which may cause abnormal work of BN. In contrast, GN’s computation is independent of batch size, and its accuracy is stable in a wide range of batch sizes. Second, we introduce FReLU [4] activation into the U-Net block, which can achieve capturing complicated visual layouts with regular convolutions. This new activation function
has been proven effective on multiple vision tasks, we apply it in the image dehazing field and prove whether it is effective in this work. The resulting model is finally trained in an end-to-end manner, i.e., estimate the clean image from a single input hazy image directly. The public benchmarks are applied to validate the proposed approach. Experimental results demonstrate that the proposed approach is able to achieve state-of-the-art performances. In addition, to further validate the proposed approach, some remote sensing images are also regarded as the test cases, and the results are also reported. At last, all of our training and testing code will open source online\(^1\) soon.

We summarize our primary contributions as follows.

- We propose a simple yet powerful deep neural network for natural image dehazing task. Our method is integrated some simple and minor improvements to achieve comparable performances with state-of-the-art methods. On the SOTS Indoor and Outdoor datasets, it obtains PSNR of 32.23 and 31.64 respectively.
- In addition, we also collected an additional remote sensing image dataset to illustrate the effectiveness of our method from the perspective of visual contrast. The results also demonstrate that our method is capable of applying practically in unseen contexts.
- All of the code and dataset in this work will be open source soon.

2. Related Work

The purpose of image dehazing is to eliminate the influence of haze on image quality and increase image visibility. It has been a hot research topic in the field of image processing and computer vision for many decades, which has attracted enormous attention from all over the world. Existing approaches can generally be categorized into two types, as shown below:

**Prior-based approaches:** the core idea of this type of approach is to apply the prior knowledge to recover the lost image information based on the image statistics. The dark channel prior was proposed by He et al. [5] to estimate the transmission map for hazy images. However, applying this prior is hard to cope with the images that are similar to the atmospheric light. The color attenuation prior [6] was proposed to model the depth of the hazy scene by building a linear model. The maximum contrast [7] was proposed and applied to recover image scenes from hazy input. The scattered light was also studied to estimate the optical transmission in hazy scenes by increasing the image visibility and recovering the contrast with haze-free objects. In summary, all the aforementioned approaches showed promising performance on some specific datasets, they still failed to robustly enhance the hazy images with more complex situations, such as unconstrained environment in the wild, which also promotes us to develop a more flexible approach by learning mechanism.

**Deep learning-based approaches:** with the development of the neural network theory and computer hardware, the deep learning techniques have been widely applied in many research fields [8]–[13]. The availability of large synthetic datasets also promoted the application of deep learning-based approaches on image dehazing. The core idea of this type of approach is to learn the high-level transmission patterns by optimizing the deep neural networks on large-scale data pairs, instead of the conventional strategy and handcrafted priors. The convolutional neural network (CNN) was first proposed to achieve the image dehazing in an end-to-end manner [14]. The proposed model generates the medium transmission map by taking a hazy image as input, which is further applied to estimate the image scene based on a pre-defined atmospheric scattering model. Lately, the multi-scale convolutional neural network was proposed to cope with the complex image features [15], which allows the model to learn a more robust transmission map and finally enhances the model performance. The all-in-one dehazing (AOD-Net) [16] is able to recover a clean image from hazy input directly by a light-weight CNN, which optimizes the whole pipeline of image dehazing by a deep learning-based approach. Chen et al. [17] proposed a gated context aggregation network to directly restore the final haze-free image, which adopted the latest smoothed dilation technique to help remove the gridding artifacts caused by the widely used dilated convolution with negligible extra parameters. The disentanglement-based network [18] was also studied to achieve image dehazing task by the adversarial training (without paired supervision). In this framework, the physics model is implemented in the disentanglement model to learn hidden patterns from a hazy image and further generates a clean image. The cyclic perceptual-consistency loss [19] was also proposed to improve the dehazing performance in a generative adversarial network framework. In general, thanks to the powerful ability of neural networks on learning complicated non-linear features, the deep learning-based approaches are now the most popular ones for the research of image dehazing and showed performance superiority over prior-based approaches.

3. Methodology

3.1 Baseline Model

We first recall the architecture of the U-Net [1] network. In pixel-to-pixel mapping field, U-Net is one of the most representative and widely used network. The original U-Net used a network entirely of convolutional layers to perform the task of medical semantic segmentation. The network is symmetric, having an Encoder that extracts spatial features from the image, and a Decoder that constructs the segmentation map from the encoded features. Encoder consists of the repeated application of two 3 × 3 convolutions, each followed by a ReLU activation function and a 2 × 2 max pooling operation with stride 2 for down sampling. Decoder also has the same convolutional block and starts with a 2 × 2 deconv-
Fig. 1 The architecture of our network. Different colored arrows represent different operations. Best viewed in color.

olution for upsampling. Systematically analyzed and proved the importance of long skip connections in U-Net for image segmentation and transformation, which combine deep, semantic, coarse-grained feature maps from the decoder sub-network with shallow, low-level, fine-grained feature maps from the encoder sub-network. The skip connections have proved effective in recovering fine-grained details of the target objects. In this paper, we transplanted this model to the natural image dehazing task and made a series of modifications to make it more suitable for dehazing. We will describe the specific details in the following sections.

3.2 Incremental Details

3.2.1 Normalization Strategy

Batch Normalization (BN) is a very effective network component in the deep learning field, which normalizes the features by the mean and variance computed within a mini-batch. This has been shown by many practices to easy optimization and enables very deep networks to converge. However, in practice, it is required for BN to work with a sufficiently large batch size because a small mini-batch leads to an inaccurate estimation of the batch statistics. Due to hardware limitations, we have to reduce the batch size in actual operation, which may cause abnormal work of BN. For example, in many detection network implements, we need to freeze the update of BN layers cause the batch size is typically quite low on a standard PC.

Recently, Wu et al. [2] presented Group Normalization (GN) as a simple alternative to BN. GN divides the channels into groups and computes within each group the mean and std for normalization. GN’s computation is independent of batch sizes, and its accuracy is stable in a wide range of batch sizes. A family of normalization methods performs the following computation

\[ \hat{x}_i = \frac{1}{\sigma_i} (x_i - \mu_i) \]  

with \( \epsilon \) is a small constant, \( S_i \) is the set of pixels in which the mean and std are computed, and \( m \) is the size of \( S_i \). The different normalization methods are how the \( S_i \) is defined. We use \((N, C, H, W)\) to denote the shape of the input 4D tensor. As shown in Fig. 2, BN computes \( \mu \) and \( \sigma \) along the \((N, H, W)\) axes, and GN perform within \((H, W)\) axes, so it doesn’t exploit the batch dimension and its computation is independent of batch sizes. In this paper, we simply utilize GN to replace all the BN layers and the ablation study results will be discussed in Sect. 4.3.

3.2.2 Activation Function

In the original U-Net model, the Rectified Linear Unit (ReLU) is utilized as the activation function, which is the most widely used scalar activation on various tasks, in the form of \( y = \max(x, 0) \). Use \( \max(\cdot) \) to serve as non-linearity and used a hand-designed zero as the condition. The non-linear transformation acts as a supplement of the linear transformation such as convolution and fully-connected layers.

Recently, Ma et al. [4] presented a conceptually simple but effective funnel activation for vision tasks, called Funnel activation (FReLU), that extends ReLU to a 2D activation by adding a negligible overhead of spatial condition. The form of FReLU is given by

\[ y = \max(x, T(\cdot)) \]  

where \( T(\cdot) \) represents the simple and efficient spatial contextual feature extractor. In [4], authors simply use a \( 3 \times 3 \)
convolutional layer and a Batch Normalization to achieve this operation. This new activation function has been proven effective on multiple vision tasks, like classification, detection, and segmentation. In this paper, we replace all ReLU in the original U-Net model with FReLU and the improvement will be shown later.

3.3 Loss Function

In our implementation, we follow [20] and adopt smoothL1 and perceptual losses to train the proposed network. The smoothL1 loss is a robust L1 loss defined in Fast R-CNN [21]. It provides a quantitative metric of the difference between the clear and dehazed image, which is less sensitive to outliers than other distance losses. The smoothL1 loss is given by

\[
L_{\text{smoothL1}} = \frac{1}{N} \sum_{x=1}^{N} \text{smoothL1}(J(x) - J'(x))
\]

where

\[
\text{smoothL1}(x) = \begin{cases} 
0.5x^2 & \text{if } |x| < 1 \\
|x| - 0.5 & \text{otherwise} 
\end{cases}
\]

Different from the pixel-wise losses, the perceptual loss [19] leverages multi-scale features extracted by a pre-trained convolutional network model to measure high-level perceptual and semantic differences between images. In this paper, the loss makes use of VGG-16 [22] pretrained on ImageNet [23] as the loss network \(\phi\) and extracts the features from the last layer of each of the first three stages. The perceptual loss is defined as

\[
L_{\text{perceptual}} = \sum_{j=1}^{3} \frac{1}{C_j H_j W_j} ||\phi_j(J') - \phi_j(J)||_2^2
\]

where \(\phi_j(J')\), \(\phi_j(J)\), \(j = 1, 2, 3\) denote the aforementioned three VGG-16 [22] feature maps associated with the dehazed image \(J'\) and the clear image \(J\), and \(C_j, H_j, W_j\) specify the dimension of \(\phi_j(J')\) and \(\phi_j(J)\).

Based on the above definitions, we give the multi-task loss \(L\) which is a weighted sum of these losses:

\[
L = L_{\text{smoothL1}} + \lambda L_{\text{perceptual}}
\]

where \(\lambda\) is user-defined constants indicating the relative strength of each component. In our experiment, \(\lambda\) is set to 0.1.

4. Experiments

4.1 Dataset

Theoretically, it is impossible to collect real-world images and their haze-free counterparts at various locations, which poses a challenge to collect a large-scale useful dataset for data-driven dehazing methods. To address this problem, some synthetic datasets have been proposed, in which haze images are generated from clear images based on the atmosphere scattering model via a proper choice of the scattering coefficient \(\beta\) and the atmospheric light intensity \(A\). In this paper, we utilize a large-scale synthetic dataset, named RESIDE [24], to train and test the proposed network. RESIDE contains synthetic hazy images in both indoor and outdoor scenarios. In its Indoor Training Set (ITS), 13,990 hazy indoor images are generated from 1,399 haze-free images with \(\beta \in [0.6, 1.8]\) and \(A \in [0.7, 1.0]\); the depth

| Metric   | Batch Norm | 8  | 16  | 32  | +Swish | +FReLU | Group Norm | 8  | 16  | 32  | +Swish | +FReLU |
|----------|------------|----|-----|-----|--------|--------|------------|----|-----|-----|--------|--------|
| PSNR     |            | 28.54 | 29.19 | 30.06 | 30.21 | 30.56 | 31.01 | 31.19 | 31.25 | 31.15 | **31.64** |
| SSIM     |            | 0.9714 | 0.9754 | 0.9772 | 0.9765 | 0.9779 | 0.9797 | 0.9780 | 0.9781 | 0.9775 | **0.9792** |

(a) PSNR
(b) SSIM

Fig. 3 The visualization of comparison results in Table 1.
| Method          | Indoor PSNR | Indoor SSIM | Outdoor PSNR | Outdoor SSIM |
|----------------|-------------|-------------|--------------|--------------|
| DCP [5]        | 16.61       | 0.8546      | 19.14        | 0.8605       |
| DehazeNet [14] | 19.82       | 0.8209      | 24.75        | 0.9269       |
| MSCNN [15]     | 19.84       | 0.8327      | 22.06        | 0.9078       |
| AOD-Net [16]   | 20.51       | 0.8162      | 24.14        | 0.9198       |
| GFN [30]       | 24.91       | 0.9186      | 28.29        | 0.9621       |
| GCANet [17]    | 30.23       | 0.9800      | -            | -            |
| GridDehazeNet  | 32.16       | **0.9836**  | 30.86        | **0.9819**   |
| Ours           | **32.23**   | 0.9810      | **31.64**    | 0.9792       |

Fig. 4  Qualitative comparisons on outdoor dataset from SOTS for different methods.

maps \(d(x)\) are obtained from the NYU Depth V2 [25] and Middlebury Stereo datasets [26]. After data cleaning, the Outdoor Training Set (OTS) contains a total 296,695 hazy outdoor images, generated from 8,477 clear images with \(\beta \in [0.04, 0.2]\) and \(A \in [0.8, 1.0]\); the depth maps of outdoor images are estimated using the algorithm developed in [27].
We also adopt the Synthetic Objective Testing Set (SOTS) in RESIDE for testing our models, which consists of 500 indoor and 500 outdoor hazy images.

4.2 Training Details

The network is trained end-to-end with Adam [28] optimizer. The training images are all collected from RESIDE and randomly cropped to 240 × 240 pixels. Random flipping is also used for data augmentation. Restricted by the hardware, the batch size is set to 32 and the initial learning rate is set to $10^{-3}$, parameter $\beta_1$ and $\beta_2$ in optimizer take the default values of 0.9 and 0.999. For OTS, the network is trained only for 50k iterations and the learning rate is reduced to $10^{-4}$ after 30k iterations. The proposed method is implemented using PyTorch\(^1\) and all experiments are carried out on a standard PC with Intel i7-6800k and two NVIDIA TITAN RTX GPUs.

4.3 Ablation Study

To investigate the effectiveness of our method, we conduct several ablation studies. Each model is evaluated on SOTS outdoor dataset.

**GN vs BN:** We first investigate the impacts of different normalization methods on performances, including PSNR and SSIM. The baseline model is directly extended from the original U-Net by simply changing the output of the network for the dehazing task. We test the impact of different batch sizes on different normalization methods. As shown in Table 1, we can see that BN is more sensitive to batch size (from 28.54 to 30.06 in PSNR). In addition, with the change of batch size, GN has more stable PSNR and SSIM values than BN. It should be noted that, due to hardware limitations, we did not use a larger batch size to train the model. According to [2], under sufficient batch size, there is no significant difference between GN and BN.

**FReLU vs ReLU vs Swish[29]:** We also test the impact of different activation functions on the model. From Table 1, we can see that after replacing ReLU with FReLU, both BN and GN have a certain improvement. This proves that FReLU has a certain effect on the dehazing model. But compared with the normalization method, this improvement is not very obvious. FReLU is easy to implement and makes little changes to the network, so we add it to get better performance.

4.4 Experimental Results

For fair comparison, we evaluate the proposed method on SOTS for qualitative and quantitative comparisons with other methods including DCP[5], DehazeNet[14], MSCNN[15], AOD-Net[16], GFN[30], GCANet[17] and GridDehazeNet[20]. For convenience, we only adopt PSNR and SSIM as evaluation metrics, which are the two most widely used in previous methods. Table 2 shows the

\(^1\)https://pytorch.org/
quantitative comparisons on indoor and outdoor images from SOTS in terms of average PSNR and SSIM values. As depicted in Table, the result of the proposed method is the comparable performance with a state-of-the-art method [20]. Our method also outperforms other compare methods including DehazeNet, MSCNN, AOD-Net, and GCANet, which are mainly designed for natural image dehazing. This proves the effectiveness of our method.

Figure 4 and Fig. 5 show the qualitative comparisons on outdoor and indoor images from SOTS. Due to the inaccurate estimation of haze thickness, the results of DCP are typically darker than the ground truth. AOD-Net can not remove the haze completely and tends to output low-brightness images. The processing power of GCANet at high-frequency detail information performance such as textures, edges, and the blue sky is always unsatisfactory. From Fig. 4, we can observe that our method is closer to ground truth in all comparison methods, including color, brightness, and details.

We also harvest a remote sensing image dataset from the Internet to evaluate our network. It is worth noting that no images from this dataset are involved in the training phase. The presented results demonstrate that our method is capable of applying practically in unseen contexts. Due to the lack of corresponding clear images, we did not perform quantitative tests on this dataset. As shown in Fig. 6, we can observe that our method has a certain effect on remote sensing images, it can more effectively display landmarks such as airports, roads, lakes, etc. However, there are still some failure cases for some extreme conditions, as shown in Fig. 6 (b). It cant remove the foggy part very well, and there will be unidentified dark color blocks.

5. Conclusions

In this paper, we propose a simple and end-to-end dehazing network for natural and remote sensing images. Experimental results on public benchmarks and real remote sensing images confirm that our method is capable of achieving comparable performance with recent state-of-the-art methods. On the SOTS Indoor and Outdoor datasets, it obtains PSNR of 32.23 and 31.64 respectively. In the future, we are interested in integrating the object detector with our dehazing network to evaluate them on real scene applications, like detecting people and cars in hazy images.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS, vol.9351, pp.234–241, Springer, 2015. (available on arXiv:1505.04597 [cs.CV]).
[2] Y. Wu and K. He, “Group normalization,” Proc. European conference on computer vision (ECCV), pp.3–19, 2018.
[3] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” International Conference on Machine Learning, 2015.
[4] N. Ma, X. Zhang, and J. Sun, “Funnel activation for visual recognition,” arXiv preprint arXiv:2007.11824, 2020.
[5] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp.1956–1963, 2009.
[6] Q. Zhu, J. Mai, and L. Shao, “A fast single image haze removal algorithm using color attenuation prior,” IEEE Trans. Image Process., vol.24, no.11, pp.3522–3533, 2015.
[7] R. Fattal, “Single image dehazing,” Acm Transactions on Graphics, vol.27, no.3, pp.1–9, 2008.
[8] Y. Lin, D. Guo, J. Zhang, Z. Chen, and B. Yang, “A unified framework for multilingual speech recognition in air traffic control systems,” IEEE Trans. Neural Netw. Learn. Syst., 2020.
[9] Y. Lin, L. Li, H. Jing, B. Ran, and D. Sun, “Automated traffic incident detection with a smaller dataset based on generative adversarial networks,” Accident; analysis and prevention, 2020.
[10] Y. Lin, L. Deng, Z. Chen, X. Wu, J. Zhang, and B. Yang, “A real-time atc safety monitoring framework using a deep learning approach,” IEEE Trans. Intell. Transp. Syst., pp.1–10, 2019.
[11] J.-H. Kim, B.-G. Kim, P.P. Roy, and D.-M. Jeong, “Efficient facial expression recognition algorithm based on hierarchical deep neural network structure,” IEEE Access, vol.7, pp.41273–41285, 2019.
[12] M. Chhetri, S. Kumar, P.P. Roy, and B.-G. Kim, “Deep blstm-gru model for monthly rainfall prediction: A case study of simtokha, bhutan,” Remote Sensing, vol.12, no.19, 2020.
[13] D. Jeong, B.-G. Kim, and S.-Y. Dong, “Deep joint spatiotemporal network (djsnt) for efficient facial expression recognition,” Sensors, vol.20, no.7, 2020.
[14] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “Dehazenet: An end-to-end system for single image haze removal,” IEEE Trans. Image Process., vol.25, no.11, pp.5187–5198, 2016.
[15] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.H. Yang, Single Image Dehazing via Multi-scale Convolutional Neural Networks, Springer International Publishing, 2016.
[16] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, “Aod-net: All-in-one dehazing network,” 2017 IEEE International Conference on Computer Vision (ICCV), 2017.
[17] 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,” 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp.1375–1383. IEEE, 2019.
[18] X. Yang, Z. Xu, and J. Luo, “Towards perceptual image dehazing by physics-based disentanglement and adversarial training,” AAAI, 2018.
[19] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” European conference on computer vision, pp.694–711, Springer, 2016.
[20] X. Liu, Y. Ma, Z. Shi, and J. Chen, “Griddehazenet: Attention-based multi-scale network for image dehazing,” ICCV, 2019.
[21] R. Girshick, “Fast R-CNN,” Proc. International Conference on Computer Vision (ICCV), 2015.
[22] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
[23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei, “ImageNet large scale visual recognition challenge,” International journal of computer vision, vol.115, no.3, pp.211–252, 2015.
[24] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, “Benchmarking single-image dehazing and beyond,” IEEE Trans. Image Process., vol.28, no.1, pp.492–505, 2018.
[25] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from rgbd images,” European conference on computer vision, pp.746–760, Springer, 2012.
[26] D. Scharstein and R. Szeliski, “High-accuracy stereo depth maps using structured light,” 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings, pp.I–I, 2003.
[27] Liu, Fayao, Shen, Chunhua, Lin, Guosheng, Reid, and Ian., “Learning depth from single monocular images using deep convolutional neural fields,” IEEE Trans. Pattern Anal. Mach. Intell., 2016.
[28] D.P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.
[29] P. Ramachandran, B. Zoph, and Q.V. Le, “Searching for activation functions,” arXiv preprint arXiv:1710.05941, 2017.
[30] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, “Gated fusion network for single image dehazing,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.3253–3261, 2018.

**Wenyi Ge** is currently an assistant research professor at the Department of Computer Science, Chengdu University of Information Technology. His research interests include graphics and image processing, remote sensing image enhancement.

**Yi Lin** is an assistant professor at the College of Computer Science, Sichuan University. His research interests include intelligent transportation systems and deep learning applications on computer vision and speech processing.

**Zhitao Wang** is a researcher of Beijing Satellite Navigation Center. His research interests include satellite navigation measurement and communication, satellite images and digital maps.

**Guigui Wang** is a Ph.D. student at the College of Computer Science, Sichuan University, whose research interests include the theoretical research and application realization of computer graphics technologies in the flight simulator and sailing simulator.

**Shihan Tan** is a Ph.D. student at the College of Computer Science, Sichuan University. His research interests relate to computer graphics and VR development of real-time rendering systems for flight simulation.