# Knowledge Transfer Dehazing Network for NonHomogeneous Dehazing

Haiyan Wu[1] Jing Liu[1] Yuan Xie[1]* Yanyun Qu[2] Lizhuang Ma[1]
[1]School of Computer Science and Technology, East China Normal University, Shanghai, China
[2]School of Information Science and Engineering, Xiamen University, Fujian, China
704289013@qq.com, 51174500035@stu.ecnu.edu.cn,
xieyuan8589@foxmail.com, yyqu@xmu.edu.cn, lzma@cs.ecnu.edu.cn

## Abstract

*Single image dehazing is an ill-posed problem that has recently drawn important attention. It is a challenging image process task, especially in nonhomogeneous scene. However, the existing dehazing methods are commonly designed to handle homogeneous haze which is easily violated in practice, due to the unknown haze distribution of real world. In this paper, we propose a knowledge transfer method that utilizes abundant clear images to train a teacher network to provide strong and robust image prior. The derived architecture is referred to as the Knowledge Transform Dehaze Network (KTDN), which consists of the teacher network and the dehazing network with identical architecture. Through the supervision between intermediate features, the dehazing network is encouraged to imitate the teacher network. In addition, we use attention mechanism to combine channel attention with pixel attention to capture effective information, and employ an enhancing module to refine detail textures. Extensive experimental results on synthetic and real scene datasets demonstrates that the proposed method outperforms the state-of-the-arts in both quantitative and qualitative evaluations. The KTDN ranks* $\mathbf{2^{nd}}$ *in NTIRE-2020 NonHomogeneous Dehazing Challenge [4, 5].*

## 1. Introduction

Single image dehazing is an ill-posed problem that has recently drawn important attention. The digital image will be degraded in hazy scenes that typically characterised by color and texture distortion.

Many dehazing methods [12, 9, 14, 22, 10, 19, 10] have been proposed to solve this problem and improve visibility of the hazy image. Some successful dehazing methods are based on the physical scattering model [17] which formulate as

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

where $I$ is the input hazy images, $J$ is the restored images, $t$ and $A$ represent the transmission map and the global atmospheric light respectively. However, as the transmission map and the global atmospheric light are difficult to estimate, the visibility of results is unpleasing, and the inaccurate estimation of the transmission map and the global atmospheric light may cause a cumulative error. Although some end-to-end methods [19, 10, 18] have been proposed, most of previous works hypothesis the haze is homogeneous while easily violated in practice due to the complex haze distribution of real world. What's more, a lot of information lost when training model, which exacerbates degradation of the dehazing results. The hazy image and haze-free corresponding scenes contain the same visual content, most methods make full use of input information (hazy images). However, the pair training images are difficult to collect.

Fortunately, it is more easily to collect the clear images. Could the clear images information be mined as the prior knowledge to help train the dehazing model? To solve the problems, we propose a knowledge transfer method that utilizes abundant clear images to train a teacher network which can learn strong and robust prior. We supervise the intermediate features and use the feature similarity to encourage the dehazing network imitates the teacher network. The prior knowledge transfered to the dehaizng network by intermediate feature map.

Inspired by knowledge distillation [13], which transferring knowledge from the teacher network to the student network, we designed a dual network that consists of the teacher network and the dehazing network. We train the teacher network for image to assist in training the dehazing network via providing the prior knowledge. In detail, we use haze-free images to train the teacher network, and then transfer the teacher network's knowledge to dehazing network in intermediate feature map by feature level loss. The architectures of networks are identical and both based on

*Corresponding author

encoder-decoder structure. In addition, we use pre-trained Res2Net [11] as encoder to extract detail information of hazy images, and add skip connection to preserve information. Moreover, in order to process nonhomogeneous dehazing task, we use attention mechanism that combine channel attention with pixel attention to let network pay more attention to effective information such as textures, colors and thick haze regions. Finally, we employ an enhancing module to refine the results. Based on fidelity (PSNR, SSIM) and perceptual (LPIPS [23], PI [7], MOS) quality results, the KTDN ranks $2^{nd}$ in NTIRE-2020 NonHomogeneous Dehazing Challenge [4, 5].

In summary, the contributions of our work are as follows:

1. We propose a knowledge transfer method for image dehazing with a dual network. The teacher network learned the distributions of clear images via image reconstruction task, and has ability to provide favorable prior knowledge that can be utilized to assist the dehazing network to retore clear images from hazy images.

2. We employ the feature attention module (FAM). The feature weights are adaptively learned from the FAM, thus the important features were given more weight. It provides additional flexibility in dealing with nonhomogeneous haze.

3. We employ a multi-scale enhancing module to fuse global context information to refine the result and expand the representational ability of network.

## 2. Related work

In this section, we briefly review the related methods of the single image dehazing task and introduce the knowledge distillation.

**Prior-based dehazing.** Prior-based methods solve dehaizng problem based on statistical prior, which need design hand-crafted features of natural images. There are some simple but powerful priors such as dark channel prior [12], color attenuation prior [26] and non-local prior [6]. The dark channel prior [12] proposed for the estimation of the transmission map. The color attenuation prior [26] models the scene depth of the hazy images via a linear model to estimate depth. The non-local prior [6] hypothesis the colors of a haze-free image are well approximated by a few hundred distinct colors, which forms tight clusters in RGB space. Although these priors used widely, priors may violated in practice because the haze distribution of real world is always complex and affected by other factors.

**Learning-based dehazing.** In view of the success of deep learning, there are more and more learning-based methods proposed. DehazeNet [9] is a dehazing model based on CNN. Input hazy images, and then output transmission map, which is subsequently used to recover a haze-free image via atmospheric scattering model. AOD-Net [14] jointly estimates the transmittance and atmospheric light, then directly generates a clean image. GFN [20] adopts a fusion-based strategy which derives three inputs from an original hazy image, and uses a multi-scale structure to refine the result. DCPDN [22] is a edge-preserving densely connected encoder-decoder structure with multi-level pyramid pooling module for estimating the transmission map. EPDN [19] views dehaing task as image-to-image translation task, and embedded by a generative adversarial network, which is followed by a enhancer. GCANet [10] adopts the latest smoothed dilation technique to help remove the grid artifacts caused by the widely used dilated convolution. These methods have made a series of success, but highly dependent on datasets and can't handle all cases. What's more, most of previous works hypothesis the haze is homogeneous while easily violated in practice due to the complex haze distribution of real world, thus the performance drop largely in dense haze scene and nonhomogeneous scene.

**Knowledge distillation.** Knowledge distillation [13] transfers knowledge from one deep learning model (the teacher) to another (the student). It has been applied to image classification, image segmentation, object detection and other tasks. [13] introduced the idea of knowledge distillation between large, cumbersome models into smaller, faster models without losing too much generalization power. While now knowledge distillation is implemented to knowledge transfer between two deep models, our work inspired by it but applys it in different way that we let the teacher and the student handle different tasks.

## 3. Method

### 3.1. Architecture

In this section, we outline the architecture and elaborate the three key components: knowledge transfer, attention module, and enhacing module.

**Knowledge Transfer.** We propose a dual network, which contains a teacher network and a dehazing network as shown in Figure 1. The teacher network trained by clear images for image reconstruction to help train the dehazing network. If the teacher network can restore images precisely, we consider it learned the distributions of clear images, and has ability to provide favorable prior knowledge that can be utilized to assist the dehazing network to retore clear images from hazy images. The architectures of networks are identical, which based on encoder-decoder structure, but handle different tasks. The teacher network's inputs are clear images (ground truth), while the dehazing network's inputs are hazy images. In order to enhance the encoder's performance, we use Res2Net [11] as encoder, which is originally trained for image classification.

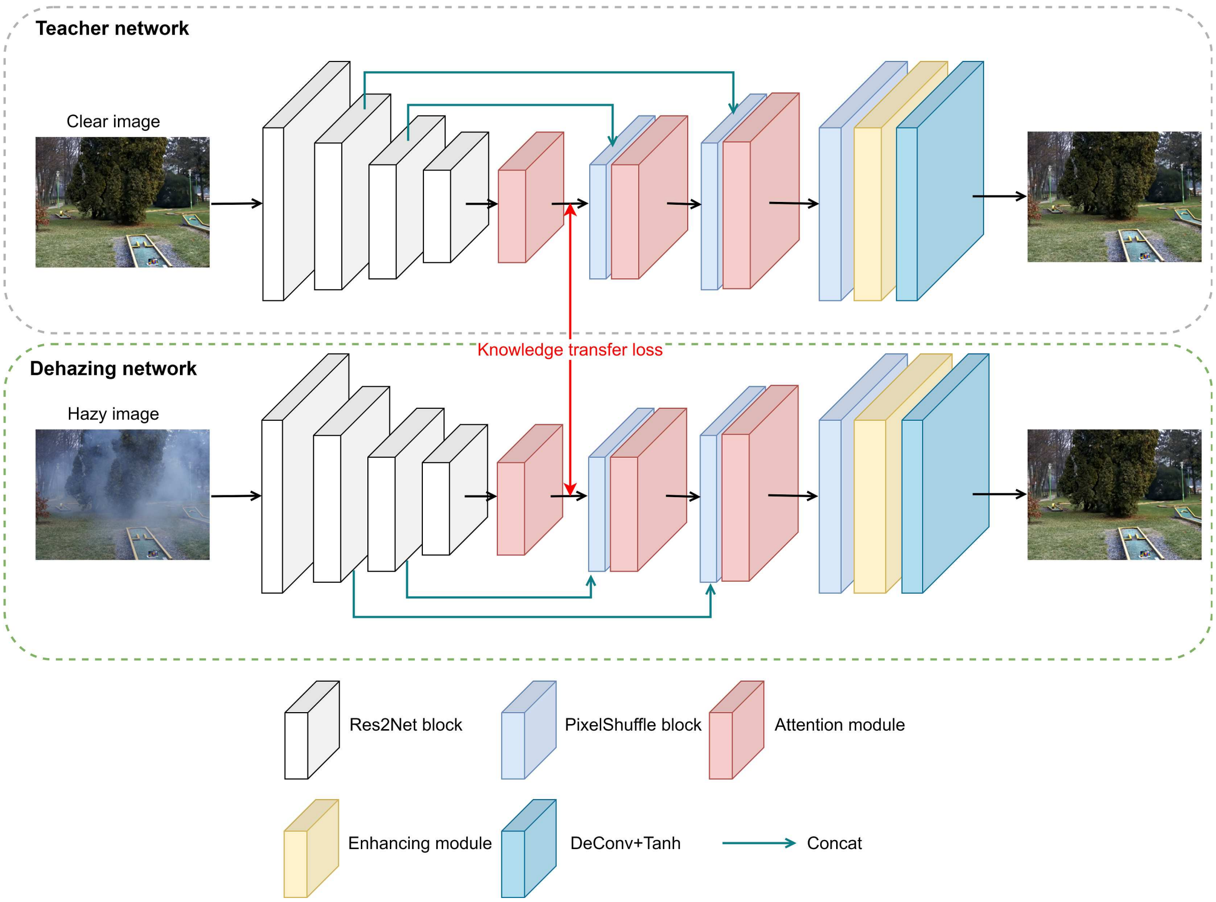


Figure 1. The architecture of the proposed KTDN.

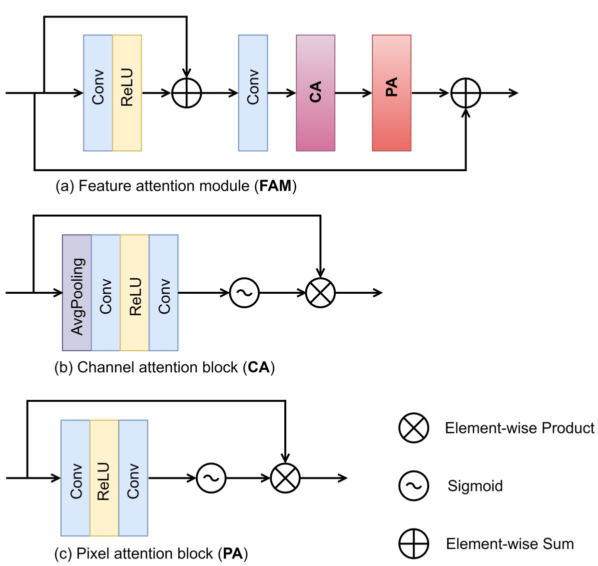


Figure 2. The feature attention module.

Note that the Res2Net we used is removed the full connection layer, only 16x downsampled and initialized with pretrained parameters from [11]. The encoder's output is a feature map, which obtains representative features, we transfer the teacher network's knowledge to dehazing network in intermediate feature map by L1 loss. The decoder consists of upsample module, attention module, and enhancing module. The upsample module is PixelShuffle [21] layer, which can mitigate the grid artifacts. While the attention module and enhancing module's specific details can be seen in the following. What's more, in order to preserve more details infomation, we also add skip connections between encoder and decoder in 8x, 4x layers.

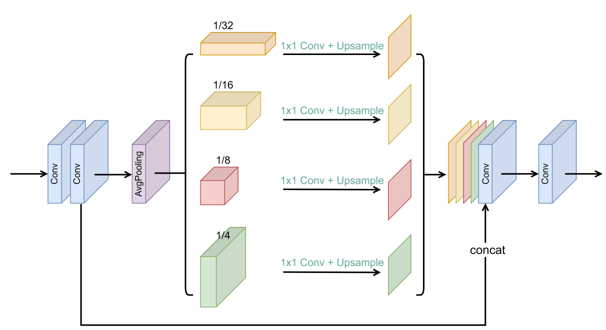


Figure 3. The enhancing module.

**Attention module.** In nonhomogeneous haze scene, haze distribution is uneven on the different image pixels. Inspired by [18], we add the feature attention module (**FAM**), which contains two attention blocks with skip connection to conduct residual learning. As shown in Figure 2, FAM consists of the channel attention block (CA) and the pixel attention block (PA). The feature map first passes through CA, then fed to PA. The channel attention block consists of an average pooling layer, two convolutions, a ReLU activation, and a sigmoid function subsequently, which achieves a linear transformation to output weights respectively for each channel. The weights will be used to feature maps by element-wise product. Similarly, the pixel attention block has two convolutions, a ReLU activation, and a sigmoid function but without average pooling layer. Finally, we add skip connection to preserve more details information and pass it into deep layers. The feature weights are adaptively learned from the FAM, thus the important features were given more weight, we let network pay more attention to effective information such as textures, colors and thick haze regions, it provides additional flexibility in dealing with nonhomogeneous haze.

**Enhancing module.** In order to expand the representational ability of network, we introduce an enhancing module (**EM**) before the last convolution layer. As shown in Figure 3, we use two convolutions extract feature map firstly, then the pyramid pooling [25] was used to integrate the details of features from multi-scale layers, which obtains global context information by learning on different recepive fields. In detail, there are two 3x3 convolution layers, and an average pooling layer to downsample the output of convolution layers by factors of 4x, 8x, 16x, 32x to build a four-scale pyramid. The 1x1 convolution followed each scale layer, and then we upsample four outputs. After that, we concatenate feature maps before and after the pyramid pooling. Finally, we add a 3x3 convolution subsequently to align feature map. Different from [19], we only use one EM, then a convolution and a Tanh activation followed. Our network can learn more context information based on different receptive fileds, so that it can refine the results.

### 3.2. Loss function

We utilize three loss functions, there are the reconstruction loss $L_{rec}$, the laplace loss $L_{lap}$, and the knowledge transfer loss $L_{kt}$, as Eq. (2).

$$L_{total} = \alpha L_{rec} + \beta L_{lap} + \lambda L_{kt}, \tag{2}$$

**Reconstruction loss.** We use $L1$ loss to train network, which demonstrated by [24] that training with L1 loss achieved a better performance than L2 loss in terms of PSNR and SSIM metrics in many image restoration tasks.

$$L_{rec} = |I_{gt} - D(I_{haze})|_1, \tag{3}$$

$I_{gt}$ denotes ground truth, $I_{haze}$ is input, and $D(\cdot)$ stands for dehazing network.

**Laplace loss.** In order to preserve low-frequency content such as color information, we use Laplacian pyramid Lap1 loss [8]:

$$L_{lap}(x, x') = \sum_{j}^{2^{2j}} |L^j(x) - L^j(x')|_1, \tag{4}$$

Where $L^j(x)$ is $j$-th level of the Laplacian pyramid representation of $x$ [16]. The $L_{lap}$ weights the details at fine scales more heavily, thus it can avoid blurry reconstructions of natural images.

**Knowledge transfer loss.** The teacher network can extract intermediate features of the clear image (ground truth) by training with clear images, which contain abundant knowleage for reconstruction. Hence, transfering this knowledge may be assist in training the dehazing network. We define the following feature matching objective function for the knowledge transfer loss:

$$L_{kt} = |T(I_{gt}) - D(I_{haze})|_1, \tag{5}$$

Where $T(\cdot)$ is the teacher network, $D(\cdot)$ is dehazing network. Note that $L_{kt}$ is feature level loss, while $L_{rec}$ is image level loss. In detail, we constrain the feature maps of the 1th FAM's output of the teacher network and the corresponding feature maps of the dehazing network. In other words, $T(I_{gt})$ denotes the feature maps of the 1th FAM's output of the teacher network, $T(I_{gt})$ denotes the feature maps of the 1th FAM's output of the dehazing network.

## 4. Experiments

In this section, we describe the datasets we used for training and testing along with some training details firstly. Then, we evaluate the dehazing results of our proposed method qualitatively and quantitatively, and compare with some other state-of-the-art methods. In NTIRE-2020 Non-Homogeneous Dehazing Challenge [4, 5], our method ranks $\boldsymbol{2^{nd}}$ based on fidelity (PSNR, SSIM) and perceptual (LPIPS [23], PI [7], MOS) quality results.

### 4.1. Datasets

**RESIDE.** RESIDE [15] is a large-scale hazy image dataset that widely used as a benchmark. It consists of both indoor and outdoor hazy images. There are five subsets totally: Indoor Training Set (ITS), Outdoor Training Set (OTS), Synthetic Objec- tive Testing Set (SOTS), Real World task-driven Testing Tet (RTTS), and Hybrid Subjective Testing Set (HSTS). Among the five subsets, ITS, OTS, SOTS are synthetic datasets, RTTS is the real-world dataset, both synthetic datas and real-word hazy datas are involved in HSTS. We use ITS as training set, which consists of

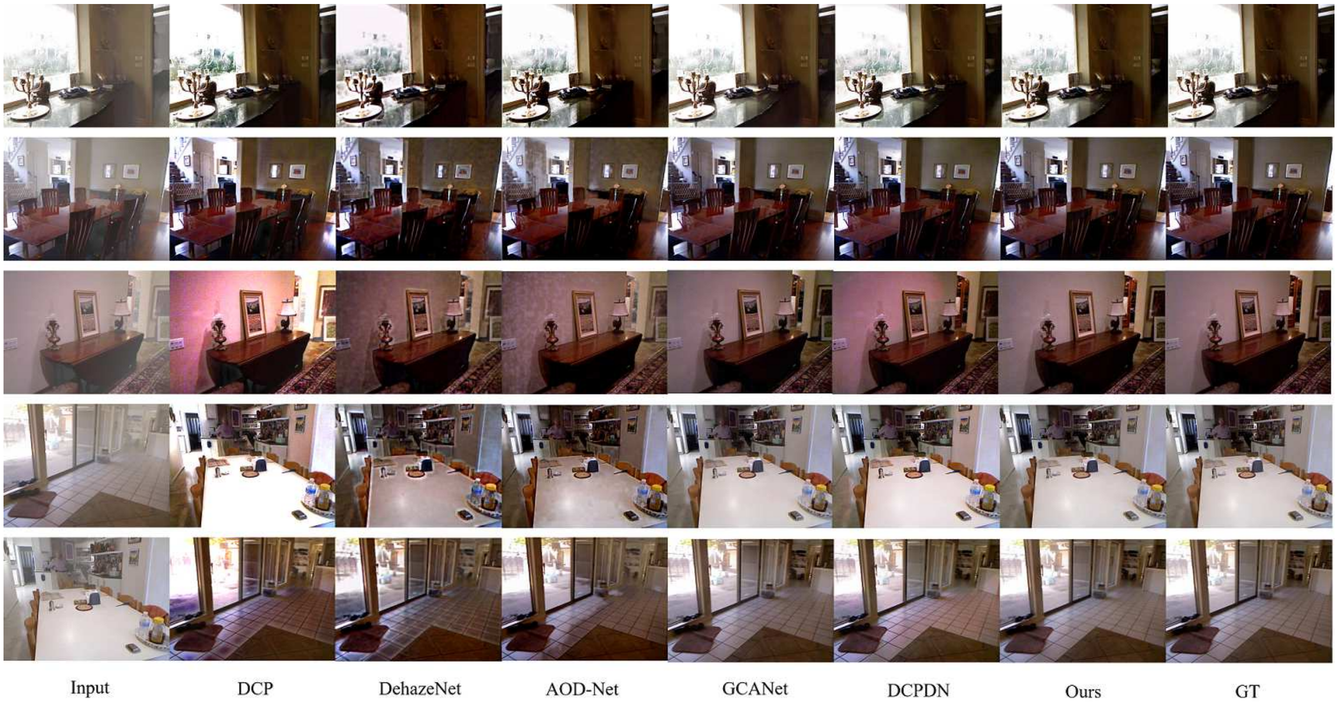


Figure 4. Quantitative comparisons of the state-of-the-art dehazing methods and our method on SOTS indoor.

Table 1. Quantitative comparisons of the state-of-the-art dehazing methods and our method on SOTS indoor.

| Methods | DCP [12] | DehazeNet [9] | AOD-Net [14] | DCPDN [22] | GCANet [10] | **Ours** |
|---|---|---|---|---|---|---|
| PSNR | 15.09 | 20.64 | 19.82 | 28.13 | 30.23 | 30.59 |
| SSIM | 0.7649 | 0.7995 | 0.8178 | 0.9551 | 0.9800 | 0.9531 |

Table 2. Quantitative comparisons of the state-of-the-art dehazing methods and our method on Dense HAZE.

| Methods | DCP [12] | DehazeNet [9] | AOD-Net [14] | DCPDN [22] | GCANet [10] | **Ours** |
|---|---|---|---|---|---|---|
| PSNR | 10.06 | 13.84 | 13.14 | 14.48 | 10.51 | 15.25 |
| SSIM | 0.3856 | 0.4252 | 0.4144 | 0.4870 | 0.3612 | 0.5206 |

Table 3. Quantitative comparisons of the state-of-the-art dehazing methods and our method on NH-HAZE.

| Methods | DCP [12] | DehazeNet [9] | AOD-Net [14] | DCPDN [22] | GCANet [10] | **Ours** |
|---|---|---|---|---|---|---|
| PSNR | 10.57 | 16.99 | 15.4 | 22.73 | 14.27 | 21.44 |
| SSIM | 0.5196 | 0.5471 | 0.5693 | 0.7351 | 0.5850 | 0.7354 |

13990 synthetic images. The testing set we use is SOTS, which includes 500 indoor and 500 outdoor images.

**NH-HAZE.** NH-HAZE [2, 3] is the NTIRE2020 challenge dataset on single image dehazing task. It contains 45 training data, 5 validation data and 5 test data. The resolution of all images is 1600x1200. Different from other datasets, the hazy images are uneven with nonhomogeneous haze. In our work, we use 40 images of training data as training set and use the remaining 5 images as testing set.

**Dense HAZE.** Dense HAZE [1] is the NTIRE2019 challenge dataset on single image dehazing task, which characterized by dense and homogeneous hazy scenes. It contains 45 training data, 5 validation data and 5 test data. The hazy scenes have been recorded by introducing real haze, generated by professional haze machines. The hazy and haze-free corresponding scenes contain the same visual content captured under the same illumination parameters. The resolution of all images is 1600x1200. In our work, we use 45 training data as training set, and use 5 validation data as testing set.

### 4.2. Experimental setting

**Training details.** During the training, we augment the training dataset with randomly rotated by 90, 180, 270 degrees and horizontal flip. In particular, we use Dense HAZE [1] as extra data for training NH-HAZE [2, 3] (the

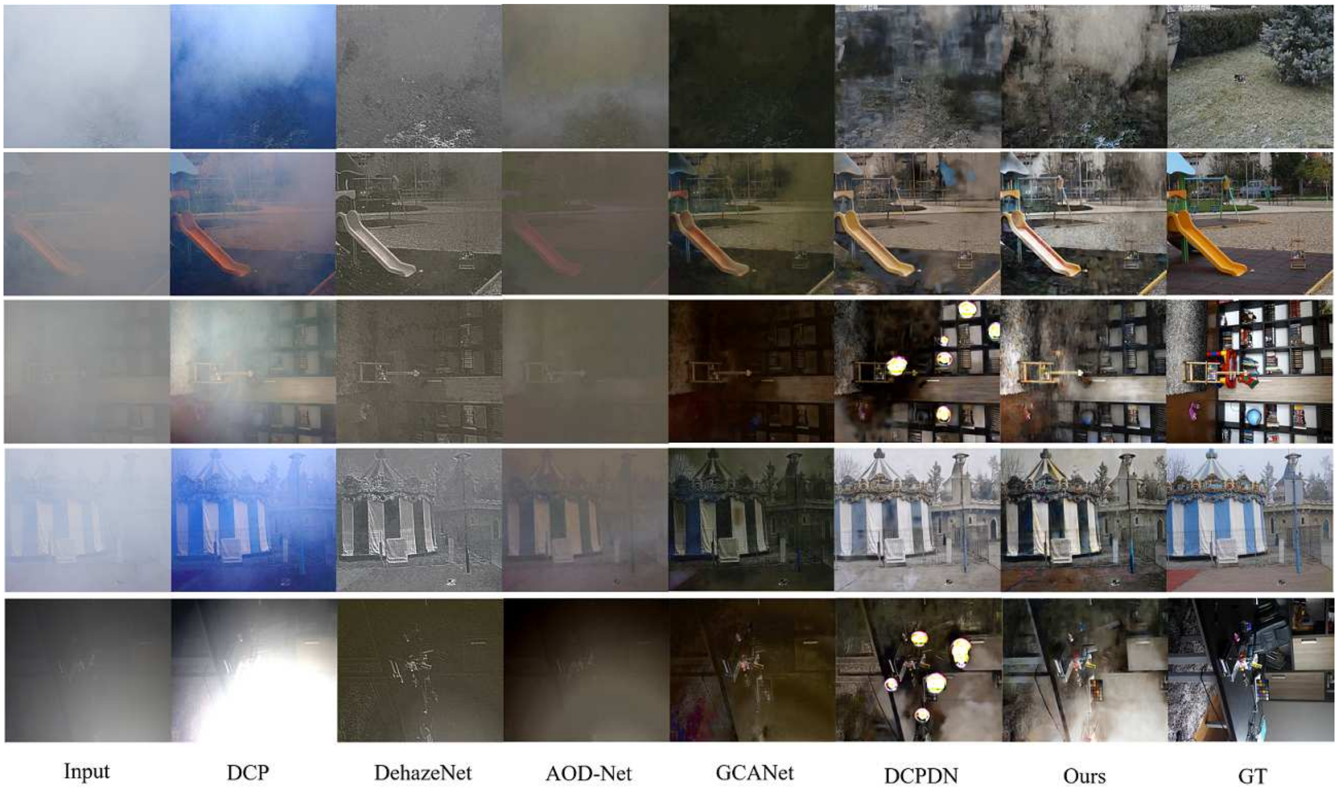


Figure 5. Quantitative comparisons of the state-of-the-art dehazing methods and our method on Dense HAZE.

Table 4. Ablation study settings and comparison of variants with different components on NH-HAZE [2, 3].

| Methods | $l_1$ | $l_{lap}$ | $l_{kt}$ | PSNR ↑ | SSIM ↑ | LPIPS ↓ |
|---|---|---|---|---|---|---|
| base | √ | - | - | 20.89 | 0.7243 | 0.4123 |
| base + FAM | √ | - | - | 21.03 | 0.7241 | 0.3656 |
| base + FAM + EM | √ | - | - | 21.31 | 0.7033 | 0.3487 |
| base + FAM + EM | √ | √ | - | 21.29 | 0.7247 | **0.3364** |
| base + FAM + EM (**Ours**) | √ | √ | √ | **21.44** | **0.7354** | 0.3394 |

NTIRE2020 challenge images) to improve performance in challenge, note that we do not adopt any extra data in this paper. We extracted the patches (256x256) from original images as input images. Besides, we adopt Adam optimizer, where $\beta_1$ and $\beta_2$ take the default values of 0.9 and 0.999, respectively. The initial learning rate is 0.0001. The hyper-parameter of loss function is set as $\alpha$ = 1, $\beta$ = 0.3, $\lambda$ = 1. We implement our model with PyTorch and 1 RTX 2080ti GPU.

**Quality measures.** To evaluate the performance of our method, we adopt three metrics: the Peak Signal to Noise Ratio (PSNR), the Structural Similarity index (SSIM), and LPIPS [23], which are often used as criteria for evaluating image quality in dehazing task.

### 4.3. Comparisons with state-of-the-art methods

We compare the proposed method with one representative prior-based method: DCP [12] , and four advanced learning-based methods: DehazeNet [9], AOD-Net [14] DCPDN [22] and GCANet [10].

**Results on synthesis dataset.** The result of our method and other comparison methods on synthesis dataset shown in Table 1 and Figure 4. As shown in Table 1, our method achieves the best performance on PSNR, which surpasses the second place 0.36 dB in indoor testing. As shown in Figure 4, our network can magically remove the haze without lose important information such as color, high-frequency textures and edges. We can observe that other representative dehazing methods both have drawback. DCP [12] loses import information with serious color distortion. DehazeNet [9] generates the blurry edges and remains noise. AOD-Net [14] remains some hazy regions evidently and also has noise. The color information cannot be estimated correctly by DCPDN [22]. GCANet [10] generate pleasing results.

**Results on a real-world dataset.** We also evaluate the proposed model on a real-world dataset, which the hazy


Figure 6. Quantitative comparisons of the state-of-the-art dehazing methods and our method on NTIRE-2020 Challenge: NH-HAZE [2, 3].

scenes have been recorded by real haze, generated by professional haze machines. As shown in Table 2, our method achieves the best performance on both PSNR and SSIM, which surpasses the second place 0.77 dB and 0.0336 dB respectively. From Figure 5, it is observed that the dense haze of input images severe destroyed details, textures, edges and colors. DCP [12] suffers color distortion where the colors of results are bluer than real sense and priors violated in some scenes. DehazeNet [9] can not estimate the color information, and always with noise. The results of AOD-Net [14] is darker than ground truth, which remains the dense haze and loses details. GCANet [10] can remove more dense haze in areas with large color differences, but distorts the color, especially in areas with similar colors. It has no ability to estimate correct details in areas with similar colors, and the results are darker. Moreover, We can see that DCPDN [22] can remove the dense haze in some regions, regretfully, it generated severe artifacts that poor visual impressions. Our method remains some haze in dehazing images, but obtains a relatively pleasing visual effect.

**NTIRE-2020 NonHomogeneous Dehazing Challenge.** Different from other datasets, the hazy images of NTIRE-2020 NonHomogeneous Dehazing Challenge (NH-HAZE [2, 3]) characterized by nonhomogeneous hazy scenes. As shown in Table 3, most state-of-the-art methods' performances drop largely due to the nonhomogeneous haze, since most previous methods assume the haze is homogeneous. Our method can restore more accurate information from the nonhomogeneous haze images, and generate more pleasing results. From Figure 6, it can be obviously seen that our method generates better results. DCP [12] suffers from severe color distortion because of their underlying prior assumptions, and loses details. DehazeNet [9] not only distorts the color, but also introduces the noise. AOD-Net [14] and GCANet [10] can't remove haze in dense haze regions. Although DCPDN [22] remove haze successful in some regions, it generated severe artifacts. Our method alleviates the problem of nonhomogeneous haze to some extent, by paying more attention to thick haze regions.

## 4.4. Ablation study

In order to intentionally analyse and demonstrate the effectiveness of the different components of the architecture, we conduct an ablation study by considering the combination of four factors: attention module, enhancing module, laplace loss and knowledge transfer loss. The ablation experiments as following: 1) **base**: only use Res2Net as encoder, and the decoder only has upsample blocks; 2) **base + FAM**: the decoder add three feature attention modules (FAM), which followed 16x layer, 8x layer and 4x layer; 3) **base + FAM + EM**: the decoder add three feature attention modules (FAM) and one enhancing module. In detail, we use NH-HAZE [2, 3] as training set and testing set, contains 1-40 images and 41-45 images respectively.

As shown in Table 4, every factor we consider plays an important role in the network performance. The knowledge transfer significantly improve performance both on PSNR and SSIM, since the knowledge transfer loss makes the dehazing network imitates the teacher network to reconstruct clear images. We can see that the prior knowledge provided from the teacher network is significant. In addition, we can observe FAM and EM help network extract more detail information that both raise score of PSNR. Comparing three loss, it can be obviously seen that $l_{lap}$ and $l_{kt}$ are efficient.

## 5. Conclusion

In this paper, we propose a knowledge transfer method that utilizes abundant clear images to train the teacher network which can learn strong and robust prior. We supervise the intermediate features and use the feature similarity to encourage the dehazing network imitates the teacher network. Besides, we introduce an attention mechanism that combine channel attention with pixel attention to let network pay more attention to effective information, an enhancing module to refine results, and the powerful loss which consists of L1 loss, laplace loss and knowledge transfer loss. Comparing with other dehazing methods, our results achieve satisfactory PSNR and SSIM values as well as visual effect. Finally, we conduct ablation experiments to demonstrate the effectiveness of the various compositions of the network.

## 6. Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (61972157, 61902129, 61772524, 61876161, 61701235, 61373077, 61602482), the Science and Technology Commission of Shanghai Municipality Program (18D1205903), the Shanghai Pujiang Talent Program (19PJ1403100), the Science and Technology Commission of Pudong (PKJ2018Y46), the Beijing Municipal Natural Science Foundation (4182067); and partly by the Fundamental Research Funds for the Central Universities associated with Shanghai Key Laboratory of Trustworthy Computing.

## References

[1] Codruta O. Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. In *IEEE International Conference on Image Processing (ICIP)*, IEEE ICIP 2019, 2019.

[2] C. O. Ancuti, C. Ancuti, and R. Timofte. NH-HAZE: An image dehazing benchmark with nonhomogeneous hazy and haze-free images. *IEEE CVPR, NTIRE Workshop*, 2020.

[3] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. NH-HAZE: An image dehazing benchmark with nonhomogeneous hazy and haze-free images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[4] C. O. Ancuti, C. Ancuti, Florin-Alexandru Vasluianu, and R. Timofte et al. Ntire 2020 challenge on nonhomogeneous dehazing. *IEEE CVPR, NTIRE Workshop*, 2020.

[5] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. Ntire 2020 challenge on nonhomogeneous dehazing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[6] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016.

[7] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[8] Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. Optimizing the latent space of generative networks. *arXiv preprint arXiv:1707.05776*, 2017.

[9] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[10] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1375–1383. IEEE, 2019.

[11] Shanghua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip HS Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[12] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[13] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[14] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017.

[15] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019.

[16] Haibin Ling and Kazunori Okada. Diffusion distance for histogram comparison. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 246–253. IEEE, 2006.

[17] William Edgar Knowles Middleton. Vision through the atmosphere. In *Geophysik II/Geophysics II*, pages 254–287. Springer, 1957.

[18] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. *arXiv preprint arXiv:1911.07559*, 2019.

[19] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8160–8168, 2019.

[20] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018.

[21] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[22] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2018.

[23] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.

[24] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016.

[25] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[26] Qingsong Zhu, Jiaming Mai, and Ling Shao. Single image dehazing using color attenuation prior. In *BMVC*. Citeseer, 2014.