

SRKTDN: Applying Super Resolution Method to Dehazing Task

Tianyi Chen¹, Jiahui Fu², Wentao Jiang², Chen Gao², Si Liu^{2*}

¹Shenyuan Honors College, Beihang University, Beijing, China

²School of Computer Science and Engineering, Beihang University, Beijing, China

18374165@buaa.edu.cn, 17373294@buaa.edu.cn, jiangwentao@buaa.edu.cn,

gaochen@iie.ac.cn, liusi@buaa.edu.cn

Abstract

Nonhomogeneous haze removal is a challenging problem, which does not follow the physical scattering model of haze. Numerous existing methods focus on homogeneous haze removal by generating transmission map of the image, which is not suitable for nonhomogeneous dehazing tasks. Some methods use end-to-end model but are also designed for homogeneous haze. Inspired by Knowledge Transfer Dehazing Network and Trident Dehazing Network, we propose a model with super-resolution method and knowledge transfer method. Our model consists of a teacher network, a dehaze network and a super-resolution network. The teacher network provides the dehaze network with reliable prior, the dehaze network focuses primarily on haze removal, and the super-resolution network is used to capture details in the hazy image. Ablation study shows that the super-resolution network has significant benefit to image quality. And comparison shows that our model outperforms previous state-of-the-art methods in terms of perceptual quality on NTIRE2021 NonHomogeneous Dehazing Challenge dataset, and also performs well on other datasets.

1. Introduction

Haze is a common natural phenomenon that causes image perceptual quality decrease. Haze can block objects in the image, results in image color distortion and visibility decrease. Thus restoring hazy images has been a challenging ill-posed problem drawing great attention.

Deep models have made significant progresses in vision tasks [13, 14, 21, 16, 15, 22] recently. Since image dehazing has been put forward as a challenging problem, a number of learning-based image dehazing methods [9, 30, 23, 36, 10, 26, 28, 11, 31, 25, 34] have been proposed and outperformed non-learning methods [12, 32, 18, 7]. A



Figure 1. The left is the output generated by Knowledge Transfer Dehazing Network. The right is groundtruth. It can be seen that KTDN gets a blurry result.

great number of these methods, including some learning-based methods, are based on physical scattering model of haze [27], which can be represented by the following equation:

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

$$t(x) = \exp(-\beta d(x))$$

where I refers to the hazy image, J refers to the original image, t refers to the transmission map, A is the global atmospheric light, and $d(x)$ is the distance between object and camera. As for the nonhomogeneous dehazing task, A is often represented as $A(x)$ to describe haze with fluctuating density. However, the estimation of transmission map $t(x)$ and atmospheric light map $A(x)$ are difficult to implement, and the solution of $J(x)$ may accumulate error. Thus results of models using physical scattering model on non-homogeneous dehazing tasks are often not as satisfactory as is expected. Meanwhile, various learning-based methods without dependence of the physical scattering model has been proposed recently. However, some of these end-to-end methods, like GCANet [10], FFANet [28], EPDN [29] and MSBDN-DFD [11], are not designed for nonhomogeneous dehazing, while other methods like BPPNet [31] suffer from color distortion, which needs further work.

In NTIRE 2020 NonHomogeneous Dehazing Challenge [5], team ECNU_KT proposed a knowledge distillation [20] model, Knowledge Transfer Dehazing Network (aka. KTDN) [34], and won 2nd prize. KTDN trains a teacher

*Corresponding author

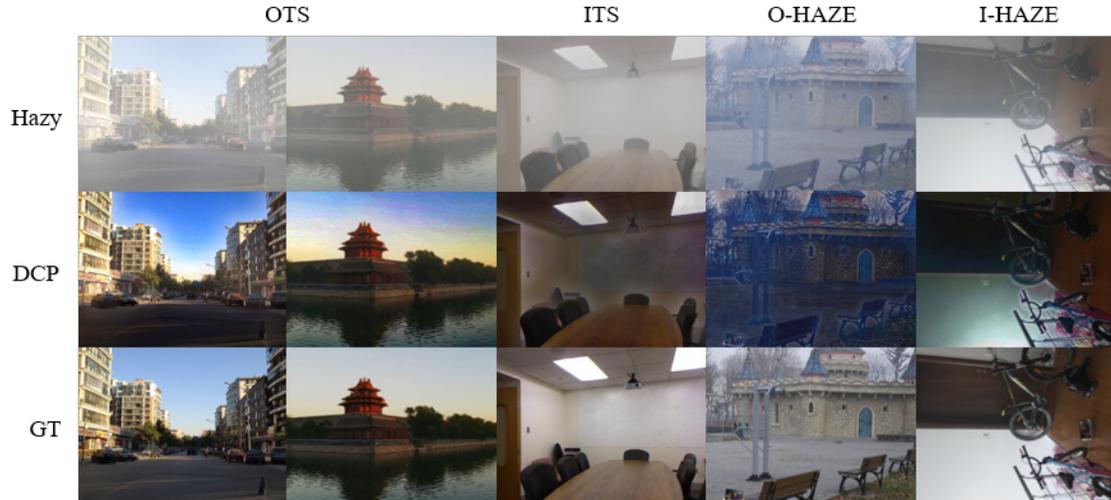


Figure 2. The results of DCP on RESIDE OTS, RESIDE ITS, O-HAZE and I-HAZE datasets. As is shown in the figure, DCP generates outputs with incorrect color when background color is bright, or objects’ brightness is similar to the background.

network firstly. While training the dehaze network, KTDN uses teacher network to produce downsampled high-level feature map, and uses L1 loss between the feature map of teacher network and the one of dehaze network to let dehaze network imitate teacher network. Although KTDN achieved high PSNR and SSIM and 2nd MOS ranking in the challenge, we find that KTDN is not capable of capturing details in the input images, and the output is relatively more blurry than groundtruth, which can be seen in Figure 1.

For this reason, inspired by [25], we infer that a network used for capturing high-frequency details may prevent the result from blurry. Thus we propose Super Resolution Knowledge Transfer Dehazing Network (SRKTDN). Our main contribution is to connect a super-resolution network to the dehaze network to refine the result, and to prove its effectiveness. The super-resolution network runs in parallel with the dehaze network, and output of two networks are concatenated to go through a tail block. Ablation study shows that, evaluated by SSIM, the super-resolution network greatly enhances the result, avoiding the result to be blurry. Our method got 20.6192 on PSNR (ranked 4) and 0.8349 on SSIM (ranked 7) in final test of NTIRE2021 Non-Homogeneous Dehazing Challenge [6].

2. Related Work

In this section, we take a look at former methods for single image dehazing, and introduce knowledge distillation. Former methods can be classified into three groups: prior-based methods, physical-model-based learning methods and physical-model-independent learning methods.

2.1. Prior-based methods

Early single image dehazing methods are usually prior-based. Prior-based methods, based on various prior knowledge or assumptions, use interpretable constraints to restore the image [12, 32, 18, 7, 1], mainly by restoring transmission map $t(x)$. Fattal [12] proposed a method using surface shading information to estimate transmission map. Tan et al. [32] removed single image haze by maximizing contrast and smoothing airlight using Markov matrix. He et al. [18] proposed a method using dark channel prior to estimate transmission map. Berman et al. [7] estimated transmission map by finding non-local haze-lines in hazy images. Recently, Ancuti et al. [1] proposed a novel fusion-based method to enhance both day-time and night-time hazy scenes and outperformed several previous state-of-the-art models, but the method is designed for homogeneous haze. Although these prior-based methods worked well in regular homogeneous hazy scenes, the results are easily violated when priors and assumptions are not suitable in some situations. Taking dark channel prior (DCP) as an example, we can see in Figure 2 that DCP generates relatively reliable result on RESIDE OTS dataset [24], but fails to generate correct output for RESIDE ITS, O-HAZE [3] and I-HAZE [4]. And these methods are also not suitable for nonhomogeneous dehazing task.

2.2. Physical-model-based learning methods

A great amount of learning-based dehazing methods rely on physical model [9, 30, 23, 36]. Setting $A = 1$, Cai et al. proposed DehazeNet [9] to estimate transmission map and use physical scattering model to calculate hazy-free image J . Ren et al.[30] proposed Multi-scale CNN with a coarse network and a fine network to estimate and refine $t(x)$.

AOD-Net [23] used a K-estimation module and a clean image generation module to estimate $K(x)$, which joins estimation of $t(x)$ and A together, and calculate hazy-free output. These methods usually consider atmosphere light A as a constant, which is appropriate for homogeneous haze removal, but not for nonhomogeneous dehazing tasks. Meanwhile, although some methods like DCPDN [36] regarded A as a function, most of them are also based on assumption that atmospheric light map is homogeneous.

2.3. Physical-model-independent learning methods

Recently, a number of learning-based methods independent of physical model are established [10, 26, 28, 11, 31, 25, 34]. GCANet [10] is an end-to-end network that uses smoothed dilated convolution and a gate fusion sub-network to fuse features from different levels. GridDehazeNet [26] proposed a novel backbone module with attention-based multi-scale estimation. FFA-Net [28] proposed channel attention block and pixel attention block to provide model flexibility. MSBDN-DFF [11] applied boosting algorithm to single image dehazing using a designed SOS boosted module. However, these methods are also based on assumption that haze is homogeneous. Meanwhile, though BPPNet [31] produced reliable results on nonhomogeneous haze dataset, NH-HAZE [2], it suffered from color distortion.

In NTIRE2020 NonHomogeneous Dehazing Challenge, Trident Dehazing Network (TDN) [25] and Knowledge Transfer Dehazing Network (KTDN) [34] won the 1st and 2nd prize respectively. These two methods also took no consideration of physical methods.

3. Method

In this section, we introduce the network architecture of the proposed Super Resolution Knowledge Transfer Dehazing Network (SRKTDN), and the loss functions we used for model training.

3.1. Network Architecture

We propose a dual network, designed based on Knowledge Transfer Dehazing Network. As is shown in Figure 3, the teacher network and dehaze network share identical network structure after the encoder, and the output of dehaze network is concatenated with the output of super-resolution network. The teacher network is used to generate high-level feature map produced by the encoder. We train the teacher network with groundtruth of the dataset to let the teacher learn the distribution of the high-level feature map required to recover the real groundtruth image. Then we use knowledge transfer loss to let the output of dehaze network’s encoder resemble the produced feature map. As the teacher network is capable of restoring the groundtruth, the dehaze

network should also be capable if the outputs of encoders are identical, due to similar network structure.

Encoder. We use ResNet18 as encoder of teacher network, and Res2Net101 for dehaze network. The ResNet18 and Res2Net101 encoders are pretrained on ImageNet, provided by [19] and [17] respectively. While ResNet18 is already capable of feature extraction for image restoring task with little CUDA memory cost, Res2Net101 has greater ability capable of learning dehazing knowledge. We remove the last layers of encoders and reserve the rest for only 16x downsample.

We use knowledge transfer loss to transfer the knowledge from the ResNet18 encoder of teacher network to the Res2Net101 encoder of dehaze network. Using L1 loss to restrict the output of dehazing encoder, the dehazing encoder can output feature map similar to the teacher, which is used for the decoder to restore haze-free image.

Decoder. As is shown in Figure 2, we use identical network architecture of decoder to KTDN.

The attention module in the model contains a channel attention block and a pixel attention block proposed in [28]. The input feature map first passes through a channel attention block (CA), then a pixel attention block (PA). The channel attention block consists of an average pooling layer, a 1x1 convolution layer, a ReLU layer, another 1x1 convolution layer and a Sigmoid layer subsequently, providing refinement with identical weight for each channel. The pixel attention block is similar to channel attention block, but without pooling layer, and has only 1 output channel, providing pixel-wise refinement. Both CA and PA calculate product of produced weights and the original input to produce the refined feature map, in order to refine important aspects such as color and thick haze, as important information gets a higher weight. The attention module has two convolution layer and a ReLU layer with a skip connection before attention blocks, and another skip connection through the whole module to preserve features from former modules.

The model uses PixelShuffle layers to upsample, which is lightweight and does not cause checkerboard artifact, unlike transposed convolution. There are skip connections between the encoder and the decoder on x4 and x8 downsampled feature maps. After upsampled, x4 and x8 downsampled feature maps are concatenated and sent to attention modules.

At tail of teacher network and dehaze network is a PSPNet [38] module. The PSPNet module combines features of different levels together to refine local area with global features. Thus the PSPNet module is used to learn context information of different receptive fields to enhance the result.

Super Resolution Network. Inspired by TDN, we use a super-resolution network to restore high gradient image details. The network architecture of super-resolution net-

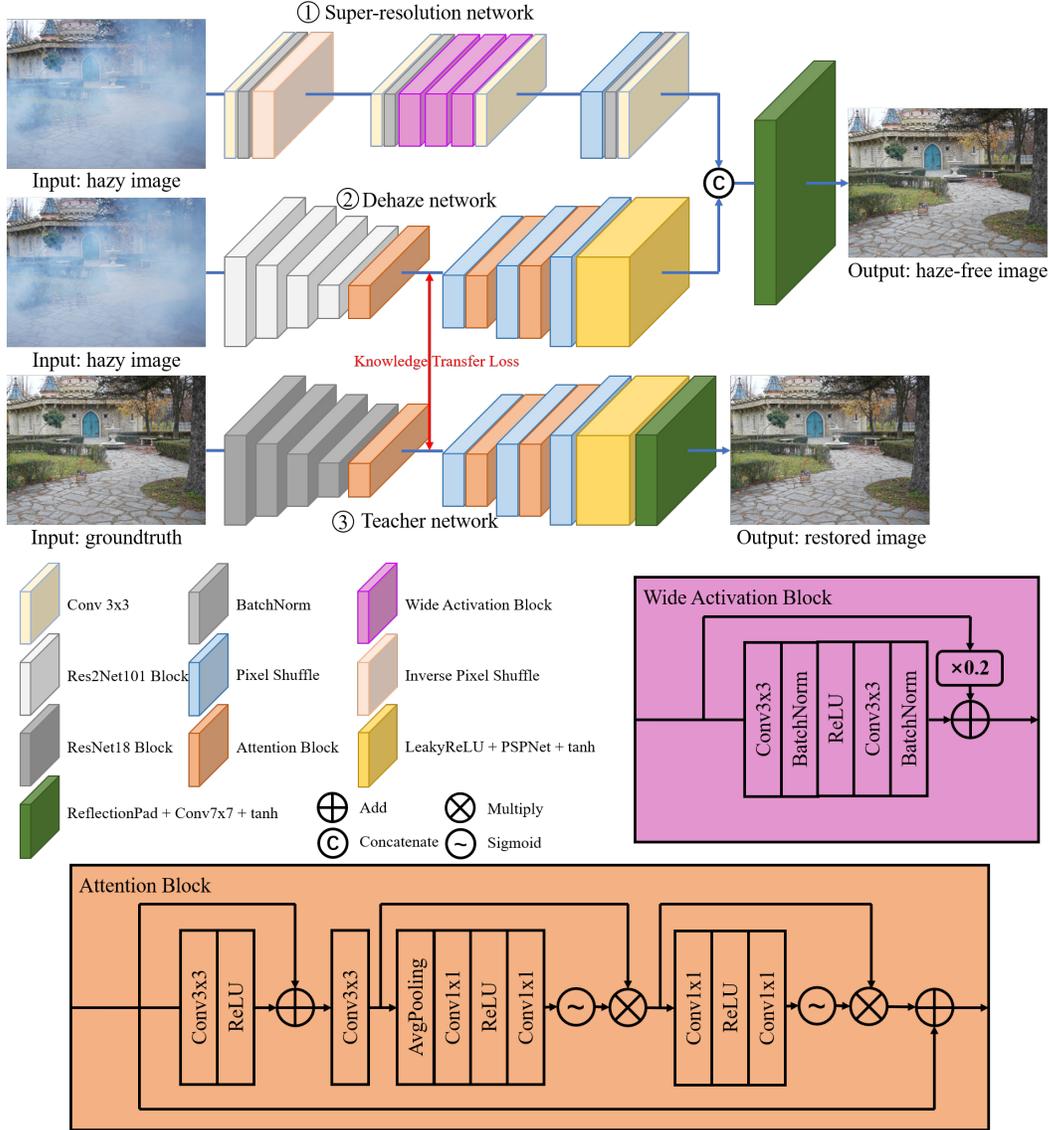


Figure 3. The upper is the architecture of SRKTDN. The teacher network uses knowledge transfer loss to restrict intermediate output of the dehaze network. The output of super resolution network and dehaze network are concatenated and pass through a tail block. There are skip connections between the encoder and the decoder on x_4 and x_8 downsampled feature maps. The lower is structure of blocks used in SRKTDN.

work is the same as Detail Refinement Net in TDN. The network uses an inverse pixel shuffle layer after a 3×3 convolution and batch normalization to x_4 downsample the feature map, and uses a 3×3 convolution, a batch normalization layer, three wide activation block [35], and a 3×3 convolution to capture details. Finally, the network uses a pixel shuffle layer to upsample, followed by a batch normalization layer and a 3×3 convolution. The wide activation block uses residual scaling inspired by [33] to prevent training instability, shown in Figure 2. The super-resolution can refine the result to be clearer on details.

The outputs of dehaze network and super-resolution net-

work go through a tail block to integrate features. The tail block contains a reflection padding layer, a 7×7 convolution layer and a tanh layer.

3.2. Objective Function

We implement the loss function with four portions, each with a specific reason: L1 loss, Laplacian loss, Lab-color-space L2 loss and Knowledge Transfer loss.

L1 loss. L1 loss is one of the most commonly used loss functions in various tasks. L1 loss is calculated by the following formula, where I and J refer to the hazy image and

groundtruth, and $M(\cdot)$ stands for the main network.

$$L_1 = |J - M(I)|_1$$

Laplacian loss. Laplacian loss uses Laplacian pyramid representation of the image and calculates L1 loss for 5 levels [8]. $L^j(\cdot)$ in the following formula is the j -th level of the Laplacian pyramid representation. Laplacian loss focuses on edge of the image and prevent the output from blurry to some extent.

$$L_{lap} = \sum_{j=1}^5 2^{2j} |L^j(J) - L^j(M(I))|$$

Lab-color-space L2 loss. L2 loss of Lab color space is used to refine color of the output image. Different from L1 loss, L2 loss pay more attention to pixels that have a relatively high deviation from the groundtruth. Besides, unlike RGB color space, Lab color space is designed to resemble human vision. $\text{Lab}(\cdot)$ in the following formula refer to the RGB-to-Lab transformation.

$$L_{Lab} = |\text{Lab}(J) - \text{Lab}(M(I))|_2$$

Knowledge Transfer loss. Identical to the method proposed in [34], Knowledge Transfer loss is L1 loss between feature map of dehaze network and the one of teacher network. Knowledge Transfer loss helps the Res2Net101 encoder to imitate the teacher’s output, hence learning information of haze removal. In the following formula, I' and J' refer to the output feature map of dehaze network’s first attention module and teacher network’s first attention module respectively.

$$L_{KT} = |J' - I'|_1$$

The total loss is calculated using the following formula.

$$L = \lambda_{L1} \times L_1 + \lambda_{lap} \times L_{lap} + \lambda_{lab} \times L_{Lab} + \lambda_{KT} \times L_{KT},$$

where λ_{L1} , λ_{lap} , λ_{Lab} and λ_{KT} are coefficients of each loss function.

4. Experiments

In this section, we clarify training details and the datasets we used for model training and testing, evaluate and compare our results with other models, and show our ablation study results.

4.1. Training Details

During training, the images of train dataset are randomly cropped into 256×256 fragments, and augmented with random rotation and flip, and the fragments are packed into batches sized 15. The optimizer is Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$. The learning rate is

10^{-4} and is decreased by 0.025 every 4 epochs. We train 400 epochs with 100 iterations for each epoch, and because the dataset and batch we used is small in size, we train 100 epochs in train mode to improve generalization performance, and switch the model to evaluate mode after 100 epochs to stabilize batch normalization layers. The coefficients of loss functions are $\lambda_{L1} = 1$, $\lambda_{lap} = 0.3$, $\lambda_{Lab} = 0.5$ and $\lambda_{KT} = 1$. The teacher network and the main network shares identical training method, and we use parameters pretrained by ImageNet to initialize the encoders.

We implement the model by PyTorch 1.7.1 on Ubuntu. The model is trained with three RTX2080Ti and the training time is about 8 hours. We augment the input with 90° clockwise rotation, horizontal and vertical flip to get 8 inputs during evaluation and test, and the result is average of output images. The runtime for one image on one RTX2080Ti is 1.770s on average.

4.2. Datasets

During training and testing, we used NTIRE 2021 Non-Homogeneous Dehazing Challenge dataset and part of NH-HAZE dataset [2] as extra data.

The NTIRE 2021 NonHomogeneous Dehazing Challenge dataset contains 25 pairs of training images, 5 validation images and 5 test images. The resolution of each image is 1600×1200 . The haze in the hazy images are nonhomogeneous, which indicates the density of haze is not evenly distributed. In the competition, we used 20 pairs of training images for training and preserve 5 pairs for validation.

NH-HAZE is the nonhomogeneous haze dataset used in NTIRE 2020 NonHomogeneous Dehazing Challenge [5], which contains 45 pairs of training images, 5 pairs of validation images and 5 pairs of test images. The resolution of each image is 1600×1200 . Among the 55 pairs of hazy and haze-free images, we chose 10 pairs whose image brightness resemble the challenge dataset: the 6th, 8th, 9th, 15th, 23th, 24th, 25th, 40th, 46th and 48th as extra dataset for model training, in order to augment color and haze density distribution variety of dataset.

4.3. Evaluation Metrics

For evaluation, we used Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as metrics to evaluate image quality. PSNR and SSIM are often used for image quality evaluation as criteria in similar tasks. Besides, to evaluate perceptual quality, we also introduced LPIPS [37] as criterion.

4.4. Ablation Study

To prove the effectiveness of super-resolution network, we implemented a series of experiments as ablation study. We analysed the effectiveness of knowledge transfer loss

Table 1. Results of previous state-of-the-art methods and ours on O-HAZE, I-HAZE and NTIRE2021 Challenge dataset

model	O-HAZE			I-HAZE			NTIRE2021 Challenge			Runtime
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	
DCP	14.69	0.5779	0.3294	11.53	0.5912	0.3062	11.10	0.5337	0.4208	0.419
AODNet	18.08	0.6514	0.2903	14.95	0.7274	0.2024	15.20	0.6413	0.3103	0.054
FFANet	21.62	0.7381	0.2552	14.33	0.7541	0.2155	20.99	0.8020	0.1825	0.870
MSBDN-DFF	23.72	0.7524	0.2111	19.85	0.8047	0.1510	20.11	0.8004	0.1909	0.437
TDN	23.97	0.7763	0.2327	20.40	0.8484	0.1667	21.24	0.7882	0.1740	0.279×8
Ours(SRKTDN)	24.83	0.7808	0.2407	19.09	0.8397	0.1700	20.13	0.8034	0.1664	0.221×8

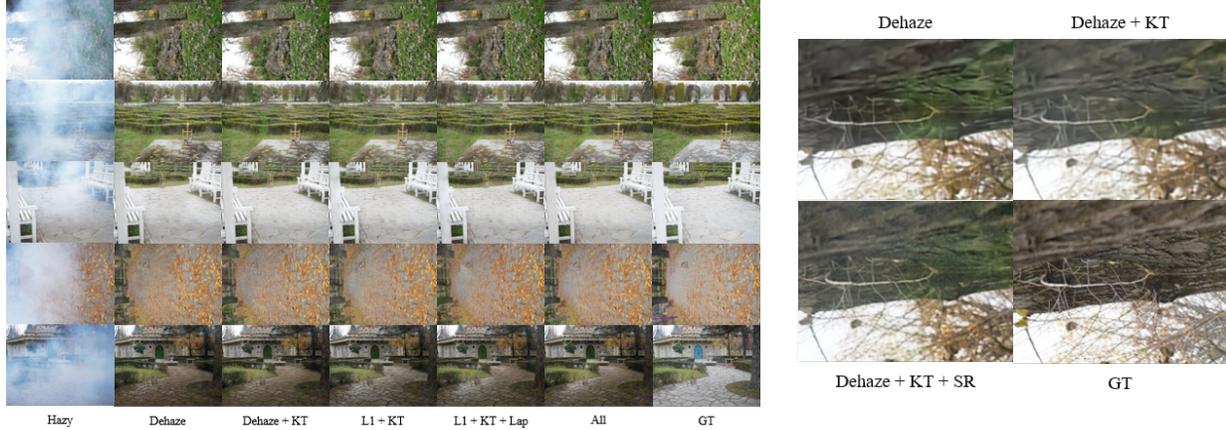


Figure 4. The outputs of experiments for ablation study. *All* is the result with all parts of network and loss function. On the right are the zoomed-in results of ablation study on super-resolution network, where *Dehaze* is the result without knowledge transfer loss and super-resolution network, *Dehaze + KT* is the result with knowledge transfer loss but without super-resolution network, and *Dehaze + KT + SR* is the result with both.

and super-resolution network by following experiments: train without knowledge transfer loss and super-resolution network; train with knowledge transfer loss but without super-resolution network; train with both knowledge transfer loss and super-resolution network. And to evaluate the effectiveness of loss functions we used, we implemented some other experiments: train without Laplacian loss and Lab-color-space L2 loss, train with Laplacian loss but without Lab-color-space L2 loss, and train with both.

Table 2. The results of ablation study. *Dehaze* is to train only dehaze network, *Dehaze + KT* is to train dehaze network with knowledge transfer loss, and *Dehaze + KT + SR* is to train dehaze network and super-resolution network with knowledge transfer loss. *L1 + KT* is to train with only L1 loss and Knowledge Transfer loss, *L1 + KT + Lap* is to train with Laplacian loss in addition, and *L1 + KT + Lap + Lab* is to train with all loss functions we used.

method	PSNR	SSIM	LPIPS
Dehaze	18.82	0.7631	0.2540
Dehaze + KT	19.35	0.7519	0.2892
Dehaze + KT + SR	20.13	0.8034	0.1664
L1 + KT	20.01	0.7993	0.1634
L1 + KT + Lap	19.70	0.8011	0.1651
L1 + KT + Lap + Lab	20.13	0.8034	0.1664

As is shown in Table 2, it’s obvious that super-resolution network have tremendous positive effect on SSIM, mainly because of capturing high-frequency information; and the improvement from training without knowledge transfer loss to training with it is not as obvious. Compared to those two factors, the effectiveness of other loss functions are relatively low. By examining outputs of these experiments, we can find the outputs are almost identical from a low-level perspective, but it’s worth notice that results without super-resolution network have lost high-frequency information. As is shown in Figure 4, the results of using only dehaze network and using dehaze network and teacher network are apparently blurry, while the results of training dehaze network, super-resolution network and teacher network succeed in detail reservation.

4.5. Comparison with other methods

We have done a number of experiments on previous state-of-the-art methods to make comparisons between those methods and SRKTDN. We trained learning-based models on corresponding datasets for a more accurate comparison. The models we used are DCP [18], AOD-Net [23], FFA-Net [28], MSBDN-DFF [11] and TDN [25].

O-HAZE dataset. O-HAZE [3] is an outdoor real scene

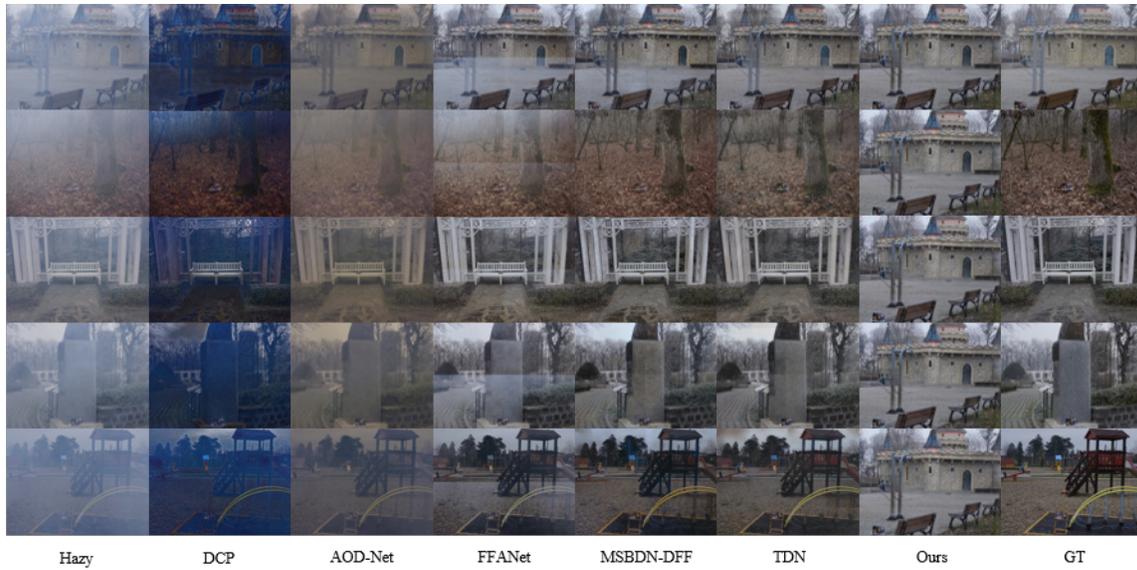


Figure 5. The results on O-HAZE dataset.

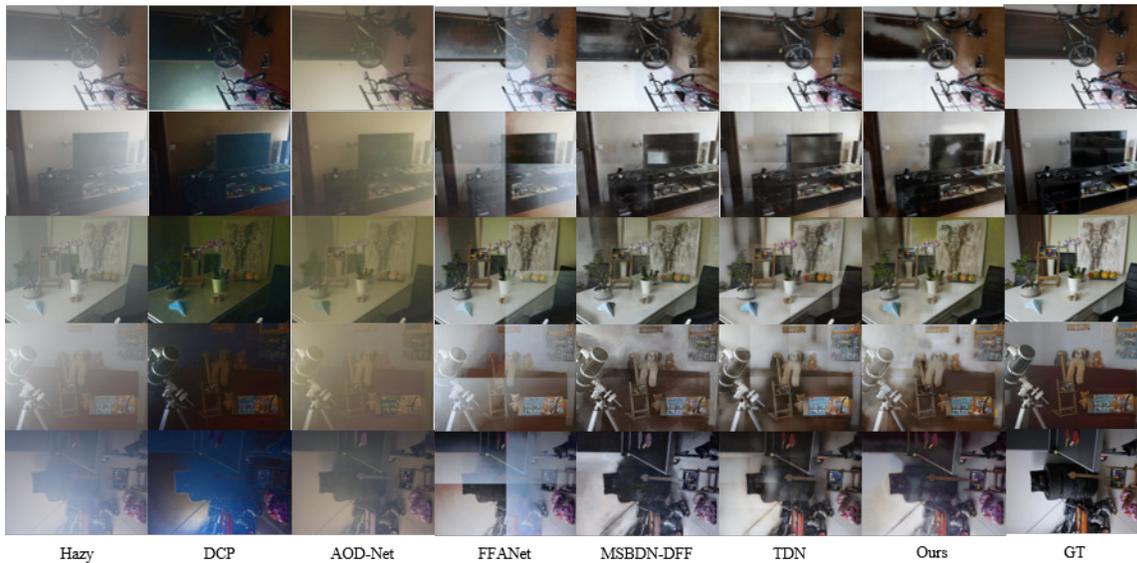


Figure 6. The results on I-HAZE dataset.

haze dataset, which was used in NTIRE 2018 Dehazing Challenge and contains 45 pairs of hazy and haze-free images of various size. Among these 45 pairs of images, 5 pairs are used as validation dataset and 5 pairs as test dataset, and the rest is train dataset. We used train dataset and validation dataset to train models, and test dataset to validate. During test, because the images are too large, we cropped the input image in some experiments to prevent running out of memory.

I-HAZE dataset. Same as O-HAZE, I-HAZE [4] is an indoor real scene haze dataset, which was used in NTIRE 2018 Dehazing Challenge and contains 30 pairs of hazy

and haze-free images of various size. Among these 30 pairs of images, 5 pairs are used as validation dataset and 5 pairs as test dataset, and the rest is train dataset. We used train dataset and validation dataset to train models, and test dataset to validate. During test, because the images are too large, we cropped the input image in some experiments.

NTIRE2021 NonHomogeneous Dehazing Challenge dataset. The competition dataset contains 25 hazy images for training, 5 images for validation and 5 images for testing. Since groundtruth of validation dataset and test dataset are not accessible, we picked 5 images among training dataset as validation dataset, and used the rest to train

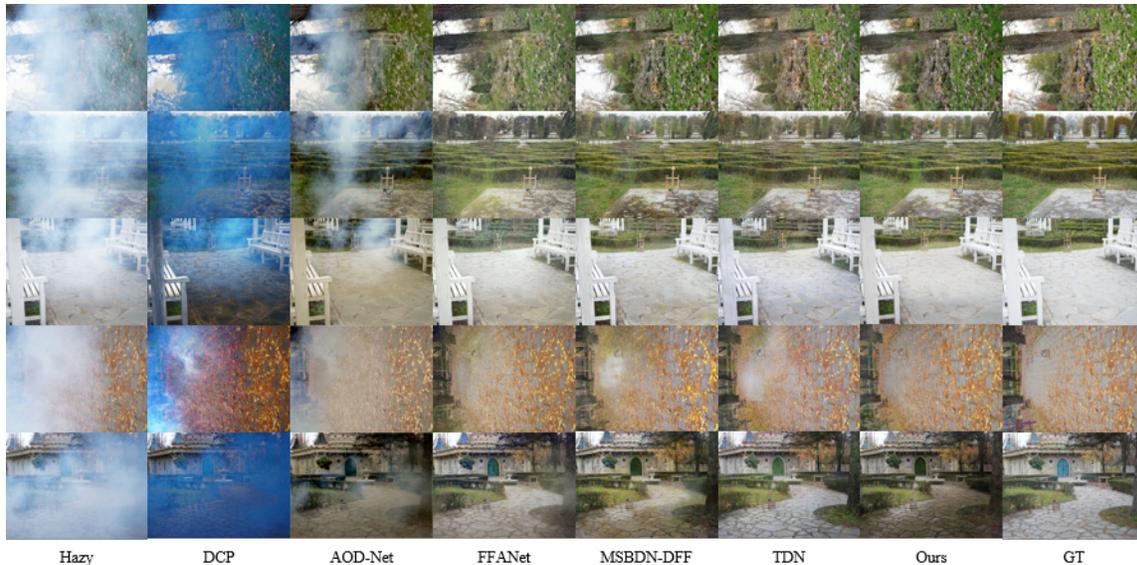


Figure 7. The results on NTIRE2021 NonHomogeneous Dehazing Challenge dataset.

models.

Results. The results are shown in Table 1. On real outdoor scene dataset O-HAZE, our method outperformed all listed methods on PSNR and SSIM, and has a relatively high performance on LPIPS among these methods. On real indoor scene dataset I-HAZE, our method’s performance is only lower than Trident Dehazing Network on PSNR and SSIM. Since I-HAZE has more objects in images and has a greater color variety than O-HAZE, dehaze on I-HAZE dataset is more difficult. As is shown in Figure 5 and Figure 6, we can find that colors in results of DCP are distorted greatly. AOD-Net failed to remove haze on both O-HAZE and I-HAZE dataset. The results of FFANet remains hazy in some parts of cropped images, showing that FFANet fails to dehaze evenly for each pixel in an image. Although MSBDN-DFE got the best performance on LPIPS, color distortion at monochromatic areas in output images can be seen. Our method still got reasonable results on I-HAZE.

The results on NTIRE2021 Challenge dataset shows that while our PSNR is lower than TDN and FFANet, our SSIM and LPIPS outperformed all other methods. As is shown in Figure 7, the 5th result of us suffered from color distortion in some area, but the rest 4 results highly resemble groundtruth. Meanwhile, MSBDN-DFE has a color distortion much severer than us, AOD-Net failed to remove non-homogeneous haze, mainly due to its dependence on physical model, and DCP got a wrong color. Although FFANet got a high PSNR, haze of the 5th image is not successfully removed. The comparison shows that our method is highly competitive due to preserving the most details and outperforms previous state-of-the-art models on perceptual quality.

We also recorded runtime for one 1600x1200 image of each method on one RTX2080Ti during our experiments. Table 1 shows that although our runtime is relatively low for one image, the total time is relatively high because we augmented the input with flip and rotation. Besides, TDN has a higher runtime than us, and FFANet got an unreasonably high runtime per image. Although AODNet has the shortest runtime, it is not capable of nonhomogeneous dehazing tasks.

5. Conclusion

In this paper, we proposed a model, Super Resolution Knowledge Transfer Dehazing Network (SRKTDN), which uses knowledge distillation method to transfer knowledge from teacher network to dehaze network, and uses a super-resolution network to refine details. By ablation study, we proved the super-resolution network produce significant positive effect on the result, and the method reached a higher PSNR and SSIM value and a lower LPIPS than the implementation without this network. We made a comparison between our method and various previous state-of-the-art solutions and drew a conclusion that our model outperformed those models on perceptual quality on nonhomogeneous dehazing task, and is also capable of homogeneous haze removal with a high performance.

References

- [1] C. Ancuti, C. O. Ancuti, C. De Vleeschouwer, and A. C. Bovik. Day and night-time dehazing by local airlight estimation. *IEEE Transactions on Image Processing*, 29:6264–6275, 2020. 2

- [2] C. O. Ancuti, C. Ancuti, and R. Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1798–1805, 2020. 3, 5
- [3] 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 *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 867–8678, 2018. 2, 6
- [4] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images, 2018. 2, 7
- [5] C. O. Ancuti, C. Ancuti, F. Vasluianu, R. Timofte, et al. Ntire 2020 challenge on nonhomogeneous dehazing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2029–2044, 2020. 1, 5
- [6] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. Ntire 2021 nonhomogeneous dehazing challenge report. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021. 2
- [7] D. Berman, T. Treibitz, and S. Avidan. Non-local image dehazing. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1674–1682, 2016. 1, 2
- [8] Piotr Bojanowski, Armand Joulin, David Lopez-Paz, and Arthur Szlam. Optimizing the latent space of generative networks, 2019. 5
- [9] 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*, 25(11):5187–5198, 2016. 1, 2
- [10] 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)*, pages 1375–1383, 2019. 1, 3
- [11] 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 *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2154–2164, 2020. 1, 3, 6
- [12] Raanan Fattal. Single image dehazing. *ACM Trans. Graph.*, 27(3):1–9, Aug. 2008. 1, 2
- [13] Chen Gao, Jinyu Chen, Si Liu, Luting Wang, Qiong Zhang, and Qi Wu. Room-and-object aware knowledge reasoning for remote embodied referring expression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 1
- [14] Chen Gao, Yunpeng Chen, Si Liu, Zhenxiong Tan, and Shuicheng Yan. Adversarialnas: Adversarial neural architecture search for gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5680–5689, 2020. 1
- [15] Chen Gao, Si Liu, Ran He, Shuicheng Yan, and Bo Li. Recapture as you want. In *arXiv preprint arXiv:2006.01435*, 2020. 1
- [16] Chen Gao, Si Liu, Defa Zhu, Quan Liu, Jie Cao, Haoqian He, Ran He, and Shuicheng Yan. Interactgan: Learning to generate human-object interaction. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 165–173, 2020. 1
- [17] S. H. Gao, M. M. Cheng, K. Zhao, X. Y. Zhang, M. H. Yang, and P. Torr. Res2net: A new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2):652–662, 2021. 3
- [18] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011. 1, 2, 6
- [19] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 3
- [20] Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*, 2015. 1
- [21] Wentao Jiang, Si Liu, Chen Gao, Jie Cao, Ran He, Jiashi Feng, and Shuicheng Yan. Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5194–5202, 2020. 1
- [22] Wentao Jiang, Si Liu, Chen Gao, Ran He, Bo Li, and Shuicheng Yan. Beautify as you like. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 4542–4544, 2020. 1
- [23] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4780–4788, 2017. 1, 2, 3, 6
- [24] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single image dehazing and beyond, 2019. 2
- [25] J. Liu, H. Wu, Y. Xie, Y. Qu, and L. Ma. Trident dehazing network. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1732–1741, 2020. 1, 2, 3, 6
- [26] X. Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7313–7322, 2019. 1, 3
- [27] W. E. K. Middleton. *Vision through the Atmosphere*, pages 254–287. Springer Berlin Heidelberg, Berlin, Heidelberg, 1957. 1
- [28] Xu Qin, Zhilin Wang, Yuanhao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing, 2019. 1, 3, 6
- [29] Y. Qu, Y. Chen, J. Huang, and Y. Xie. Enhanced pix2pix dehazing network. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8152–8160, 2019. 1
- [30] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In Bastian Leibe, Jiri

- Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 154–169, Cham, 2016. Springer International Publishing. 1, 2
- [31] Ayush Singh, Ajay Bhawe, and Dilip K. Prasad. Single image dehazing for a variety of haze scenarios using back projected pyramid network, 2020. 1, 3
- [32] R. T. Tan. Visibility in bad weather from a single image. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. 1, 2
- [33] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. Esrgan: Enhanced super-resolution generative adversarial networks, 2018. 4
- [34] H. Wu, J. Liu, Y. Xie, Y. Qu, and L. Ma. Knowledge transfer dehazing network for nonhomogeneous dehazing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1975–1983, 2020. 1, 3, 5
- [35] Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Zhaowen Wang, Xinchao Wang, and Thomas Huang. Wide activation for efficient and accurate image super-resolution, 2018. 4
- [36] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3194–3203, 2018. 1, 2, 3
- [37] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric, 2018. 5
- [38] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, 2017. 3