Task Adaptive Network for Image Restoration with Combined Degradation Factors

Jingyuan Zhou
Sun Yat-sen University
Guangzhou, China
zhoujy53@mail2.sysu.edu.cn

Chaktou Leong
Wuhan University
Wuhan, China
ctleong18@whu.edu.cn

Minyi Lin
Sun Yat-sen University
Guangzhou, China
linmy23@mail2.sysu.edu.cn

Wantong Liao
Sun Yat-sen University
Guangzhou, China
liaowt6@mail2.sysu.edu.cn

Congduan Li
Sun Yat-sen University
Guangzhou, China
licongd@mail.sysu.edu.cn

Abstract

Existing methods have achieved excellent performance on image restoration, but most of them are designed for one type of degradation. However, the weather is complex in the real world. So networks designed for single tasks are usually difficult to apply. Therefore, we propose a task-adaptive attention module to enable the network to restore images with multiple degradation factors. The task-adaptive attention module mainly includes three parts: Task-Adaptive sub-network, Task Channel Attention, and Task Operation Attention. To evaluate the model, we construct a mixed degradation factors dataset that combines three degradation factors of rain, haze, and raindrop. The experimental results show that our method not only better restores images with mixed degradation factors, but also show competitive results compared to the state-of-the-art models of each task.

1. Introduction

There have been many studies on image restoration, which aims to restore degraded images to a certain extent. Most of the traditional image restoration methods are based on defined prior methods or physical models. For example, dark channel prior algorithms[8] for image dehazing, Wiener filtering[26] for image denoising and sparse coding[20] for image deraining. Recently, learning-based methods are widely applied in various image restoration tasks and have become state-of-the-art models in almost all sub-tasks. Including image deraining[19, 11, 17, 6], dehazing[1, 15, 24, 3, 4, 21, 7], deblur[12, 29, 22, 12], denoising[26, 37, 14], low-light enhancement[33, 34, 2, 13], etc. These previous studies have proved that deep learning can be more superior and flexible than those traditional methods in image restoration.

Although the existing deep learning models have achieved good results in each sub-task of each image restoration, the generalization ability of these models have a lot of room for improvement. In the real world, the weather environment is complex and changeable so that these models designed for a single task cannot be applied in the real world. At present, there are few studies on image restoration with mixed degradation factors or multiple scenes image restoration[35, 27, 18]. As in actual autonomous driving or surveillance environments, the main factor of image degradation is the weather condition, so we specifically selects three types of weather conditions that usually accompany in reality: rain, hazy, and adherent raindrops to create a dataset which mixes three degradation factors for training and evaluation. In addition, we mix the single degradation factor datasets of three tasks(rain, hazy, and adherent raindrops) for comparison experiments with the state-of-the-art models. Although it is usually difficult to design models for removing multiple degradation factors at the same time, there is a potential relation among these tasks, that is, they all aim to remove some noise from the images. It also means that the model designed for a certain task can usually achieve the desired effect after training on the dataset of another task. In addition, according to the lottery ticket hypothesis of the neural network[5], a model with a large number of parameters usually has a sub-model with a relatively small number of parameters but similar performance. The parameters are pruned to reduce the number of parameters but achieve similar performance. According to this assumption, a model for a certain image restoration task can also find a subspace with similar performance in its parameter space. Based on the above-mentioned assumption, we introduce a task-adaptive attention mechanism to guide the model to allocate parameter subspaces for different degradation factors or tasks.

At the same time, since the previous work did not use the prior information of image degradation intensity or category, there is still a lot of room for improvement in image restoration with mixed degradation factors. Therefore, in this paper, we propose the Task Adaptive Attention Module to supervisedly adapt to the mixed degradation factors because supervised attention can better allocate parameter subspace for the model. The Task Adaptive Attention Module contains three sub-modules, namely Task Adaptor subnet, Task Operation Attention, and Task Channel Attention. The role of Task Adaptor sub-net is to predict the intensity vector of the mixed degradation factor, which is served as the prior information of Task Operation Attention and Task Channel Attention. Task Operation Attention assigns corresponding weights to a series of Operation Blocks to adapt to different tasks, which is similar to supervised selection of multiple Operation Blocks. Task Channel Attention performs channels selection and removes the noisy channels.

Our contribution can be summarized as follows:

- We propose a mixed degradation factors dataset, which is more in line with the severe weather conditions that may be encountered on the vehicle camera. According to the knowledge of authors, this is the first dataset containing multiple degradation factors under complex weather.
- We propose the Task-Adaptive Attention Module, which show good performance on mixed degradation factors image restoration and multi-scene image restoration.
- Our network can be trained in an end-to-end manner and has achieved competitive results on both mixed degradation factors dataset and mixture of different bad weather datasets.

2. Related Work

In this paragraph, we will briefly review some researches on leanrning-based image restoration and the attention mechanism in computer vision

2.1. Deep Learning in Image Restoration

Convolutional neural networks have been widely used in various computer vision tasks and have achieved exciting results in the field of image restoration. Most of the recent works of single image dehazing are to generate the residual map of the haze. [24] proposed the residual block and the attention mechanism for feature fusion, which has achieved state-of-the-art result in the indoor dehazing dataset. [3] introduced a threshold fusion sub-network combined with the smooth dilated convolution, which has achieved good results on both rain removal and haze removal. Recently, there have been many GAN-based image dehazing work. [4] introduced cycleGAN to restore a hazy image. [19] proposed a progressive single image deraining network, which proved that the application of progressive threshold fusion network can make a great improvement in image deraining. [36] proposed a multi-step progressive image restoration network, which has achieved state-of-the-art results in various image restoration tasks. The GAN network combined with visual attention has made a great progress on deraindrop[23] as well.

At present, there are only a few works aimed at image restoration with combined degradation factoirs. [35] introduced reinforcement learning to enable effective image restoration in a variety of degradation levels. [27] integrates attention into image restoration of mixed distortions. The operation is selected based on the dynamic weights of attention so that a single model can remove raindrops, blur, noise, and JEPG at the same time. However, the above-mentioned scene restoration is less relevant in actual automatic driving, monitoring, face recognition and other applications and the performance still has a lot of room for improvement.

2.2. Attention Mechanisms for Computer Vision

Recently, there have been many studys on the attention mechanism for computer vision. The attention mechanism searchs for the relevant feature to learn the weight distribution and then uses the learned parameters to weight the corresponding feature map. [10] can find the relationship among the channels of feature map and adaptively suppress the channels that are not important to the current task. [32] proposed a model that combines the channel attention module and the spatial attention module, which can better extract the weights of the spatial domain and the channel domain. [30] proposed a non-local method to search for the global relationship.

The attention mechanism also has many applications in the field of image restoration. Non-local was introduced to image restoration by [38] for the first time. Operation-wise attention[28] was proposed to restore images with mixed degradation factors. [24] combined the CBMA attention module and feature fusion ,which achieved state-of-the-art performance on the image dehazing. The Task Adaptive Attention Module that we propose is basically similar to the above methods, but it targeted for multiple mixed degradation factors. We proposed a supervised attention mechanism to adaptively remove various degradation factors at the same time.

3. Methods

3.1. Architecture Design

Our network structure is shown in Figure 1. Firstly, the input image is down-sampled by the encoder and the dimension of feature map is expanded. The feature map extracted by the encoder is will be fed into two paths. One path is called the task adaptor, which generates the normalized intensity values corresponding to each degradation factors. The intensity values served as the prior information of Task Attention Group and Task Attention convolution. Another path is the main network, consisting of Task Attention groups and gate sub-net. Multiple task attention groups extract the feature of different degradation. The Gate sub-network was firstly introduced by [3], which performs weighted fusion on the outputs of all the Task Attention groups. After up-sampling, the predicted degradation residual is added to the input degraded image through skip-connection to obtain a clean image.
In particular, we added an additional layer of first-order differential feature map with the size of $H \times W \times 1$ to the input dimensions. According to [31], edge feature map can help training convergence.

3.2. Task Adaptor

The task adaptor is a simplified version of MobileNet v3[9]. Its role is to predict the values of each degradation factor. N indicates the number of degradation factors.

$$\text{TaskSelection } = \text{ classifier } (\text{input}) \in \mathrm{R}^{1 \times \mathrm{N}} \tag{1}$$

TaskSelection is used as the prior information of Task Attention Block so that the model gives attention weights of different operation blocks and channels according to the intensity of different degradation factors.

3.3. Task Attention Block

As shown in Figure 2, the Task Attention Block have a task operation-wise attention and a task channel-wise attention. The commonly used operation-wise attention and channel-wise attention were both self-supervised, using the information of the feature map to generate a weight vector. Our idea is to use the intensity information of each degradation obtained by the Task Adaptor to supervisedly learn the weight vector. The purpose of our method is to better divide parameters into multiple sub-tasks and improve the efficiency of the model.

3.3.1 Task-Operation Attention

The Task-Operation Attention module $\mathcal{F}_l$ is composed of weights $\mathbf{W}_1 \in \mathbb{R}^{T \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{\mathcal{O} \times T}$. $\sigma(\cdot)$ represents the activation function ReLU. T indicates the number of degradatrion tasks and O indicates the number of operation blocks.

$$\mathcal{F}_l(\mathbf{x}) = \mathbf{W}_2 \sigma\left(\mathbf{W}_1 \mathbf{x}\right) \tag{2}$$

Task Adaptor predicts the value vector $\mathbf{A}$ corresponding to each task. The weight vector of each operation block is obtained through $\mathcal{F}_l$. O_i denotes the output of the i-th operation block. The feature map weighted by Operation-wise Attention denotes X_o, then we have

$$X_o = \sum_i \mathcal{F}_i \times O_i \tag{3}$$

To illustrate how Task-Operation Attention works, we visualize the mean value of attention weights of different bad weather tasks in Figure 3. Each row and column indicates one of the attention layers and one of the operations employed in each layer. We can observe that the attention weights differently depending on different tasks to a certain extent, it indicates that different tasks own and share the operation blocks.

3.3.2 Task-Channel Attention

Similarly, the Task-Channel Attention module is composed of weight $\mathbf{W}_3 \in R^{T \times C}$, and bias $\mathbf{B} \in \mathbf{R}^C$. $\mathrm{z} \in \mathrm{R}^{\mathrm{C} \times \mathrm{H} \times \mathrm{W}}$ denotes the input feature map. The output feature map is calculated by the following formula:

$$\text{attention } = \text{Sigmoid } \left(\mathrm{W}_3 \text{ TaskSelection } + \mathrm{B}_3\right) \in \mathrm{R}^{\mathrm{C} \times 1 \times 1} \tag{4}$$

$$\text{output } = \text{ attention } * \mathrm{z} \in \mathrm{R}^{\mathrm{C} \times \mathrm{H} \times \mathrm{W}} \tag{5}$$

The mean value of channel attention weights are shown in Figure 4. It can be seen that the channel weights differently depending on the degradation types. It shows that our method does select channels to adapt different degradation factors.

3.3.3 Operation Block

Operation Block is composed of smooth dilate convolution and skip-connection. Each Task Attention group contains multiple Task Attention Blocks. The dilation in each Group is set to 2, 3, 4, 1, respectively. The filter size of all convolution is $H \times W \times C$ and the number of C is set to 120. As shown in the Figure. 2, all operation blocks are performed in parallel.

3.4. Loss Function

Loss function is a key point for multi-task learning. As for the image restoration, we combine smooth L1 and SSIM as the loss function. The gradient of Smooth L1 will decrease as the loss becomes small. Differential coefficient of smooth L1 remains 1 when the loss is rather large. So it

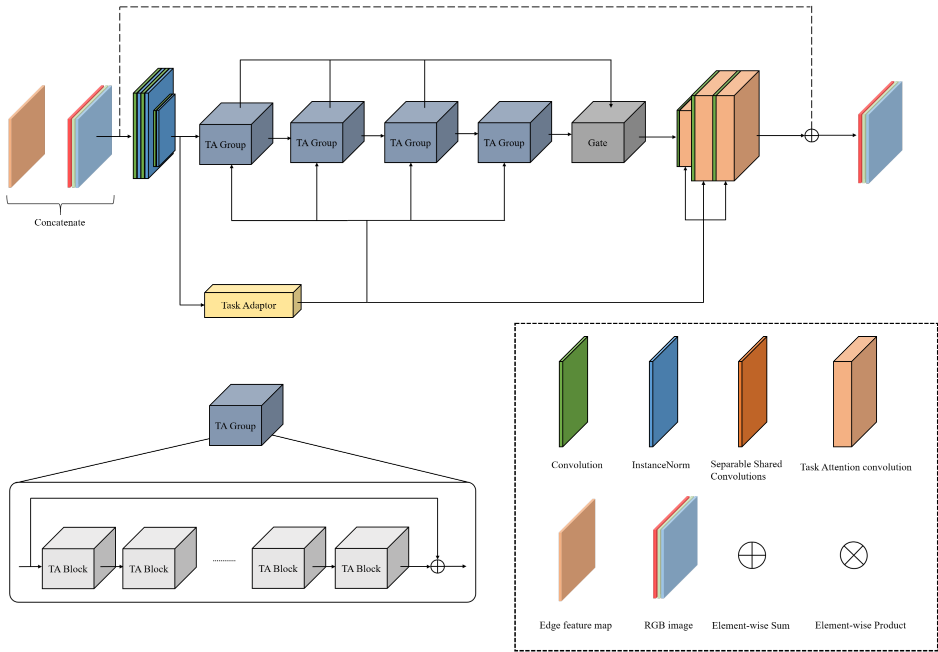

Figure 1. The architecture of proposed network

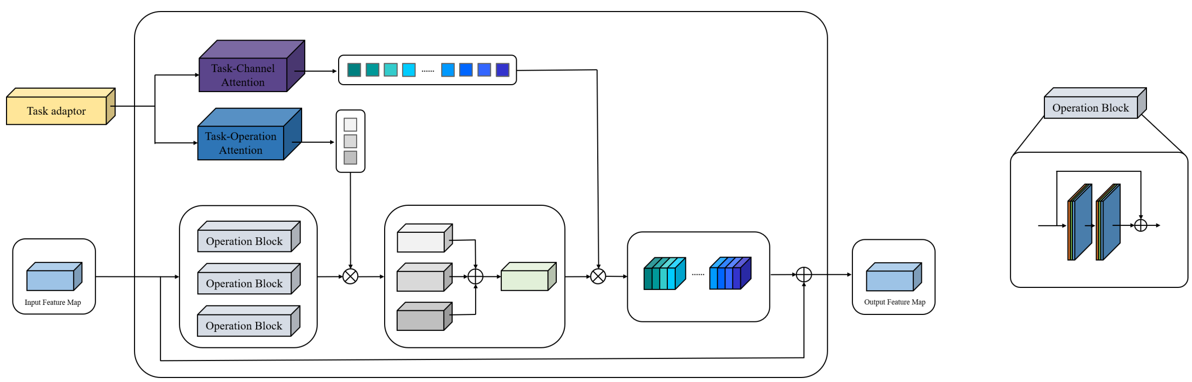

Figure 2. The architecture of Task Attention Block

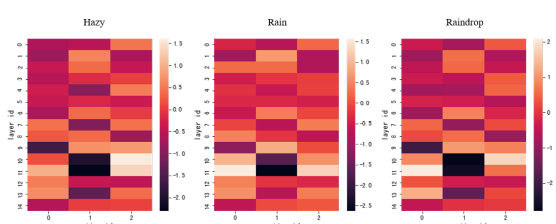

Figure 3. Task-Operation Attention Map

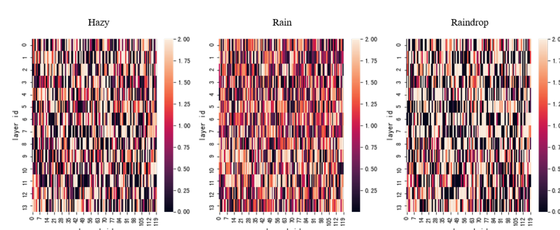

Figure 4. Task-Channel Attention Map

will be more stable at the beginning of training than L2 and have better convergence than L1. J indicates the input image. The outputs of our network include the restored image I and the predicted classification Y.

$$\hat{I}, \hat{Y} = Network(J) \tag{6}$$

$$\text{smooth\_}L_1(I, \hat{I}) = \begin{cases} 0.5(I - \hat{I})^2 & \text{if } |I - \hat{I}| < 1 \\ |I - \hat{I}| - 0.5 & \text{otherwise} \end{cases} \tag{7}$$

SSIM pays more attention to the whole brightness and contrast ratio of the image instead of the pixels. Therefore,

weighting these two loss functions as the loss of the image restoration can evaluate the image-wise and pixel-wise performance better.
For task adaptor, we use L2 as the loss function. Considering that the learning rate is reduced in the later stage of training and the classifier is basically stable, we add the coefficient of cosine decay to it to make it smoothly decrease.

$$L_2(Y,\hat{Y}) = (Y-\hat{Y})^2 \tag{8}$$

The loss function L of our network is as below. λ_1, λ_2 and λ_3 indicate the coefficients.

$$\begin{aligned} L = \lambda_1 Smooth\_L_1(I,\hat{I}) + \lambda_2(1-SSIM(I,\hat{I})) \\ +cosdecay(\lambda_3, epochs)L_2(Y,\hat{Y}) \end{aligned} \tag{9}$$

4. Experiment

4.1. Dataset

According to our knowledge, there is no publicly available dataset with various degradation intensity labels. However, the degradation intensity label can help the model better adapt to different degradation factors, so we synthesized a new dataset, which includes 2061 image pairs and each image pair has a corresponding degradation factor intensity label. The 2061 clean images with depth maps are collected by [16]. The generation process of our dataset is shown in Figure 5. We add the weather degradation factors(haze, rain and adherent raindrop) from far to near. For each image, the degradation levels of haze, rain and raindrop are randomly chose from uniform distribution in the range [60,150], [250,300] and [5,25]. Moreover, for a more objective comparison, we train our model on different bad weather dataset, including image deraining: ”DDN” dataset[6], image dehazing dataset: ”RESIDE V0”[16] and raindrop dataset[23]. The ”DDN” dataset contains 12600 training samples and 1400 testing samples.”Reside V0” contains 13990 training samples. We use SOTS indoor as the testset, with 500 indoor test samples. The raindrop dataset contains 1119 raindrop and ground truth pairs, of which 58 real-world images are selected as the test images. Considering the imbalance of size among datasets, we oversample the raindrop dataset and amplify all the images by rotation, random clipping, and other methods. After processing, the number of samples in each epoch is 38644.

4.2. Training detail

Our model can be trained end-to-end on the mixed degradation factors dataset and different bad weather dataset. The initial learning rate is set to 0.0003 and the number of epochs is 40. We use cosine decay to gradually reduce the learning rate. The formula is as follows. We use AdamW as the optimizer and weight decay is set to 2×10^{-4} .

Table 1. Results of mixed degradation factors test set in ablation study

TOA	TCA	PSNR	SSIM
✓	✓	**29.62**	**0.9528**
✓	✗	<u>29.30</u>	<u>0.9509</u>
✗	✓	28.12	0.9410
✗	✗	27.77	0.9386

Table 2. Results of test methods on ”DDN” dataset

Metrics	DDN[6]	NLEDN[17]	MSPFN[11]	Ours
PSNR	28.24	29.79	**32.82**	<u>30.68</u>
SSIM	0.8654	0.8976	**0.9302**	<u>0.9248</u>

Table 3. Results of test methods on RESIDE V0 dataset

Metrics	AOD[15]	PFFN[21]	MSBDN[7]	Ours
PSNR	19.67	24.78	**33.79**	<u>32.34</u>
SSIM	0.8065	0.8923	**0.9842**	<u>0.9737</u>

Table 4. Results of test methods on Raindrop dataset

Metrics	AttentGAN[23]	Quen *et al*. [25]	Ours
PSNR	**31.57**	<u>31.44</u>	29.41
SSIM	0.9023	**0.9263**	<u>0.9248</u>

4.3. Ablation study

To show the effectiveness of the Task Attention module of our network, we conducted an ablation study in mixed degradation factors dataset, which is shown in Table 1. We have verified the effectiveness of Task Operation Attention(TOA) and Task Channel Attention(TCA) respectively. As can be seenthe network with both Task Operation Attention and Task Channel Attention has better performance than the others. The visualization of the results are shown in Figure 6.

4.4. Comparison with State-of-the-art Methods

We take some state-of-the-art methods as the baseline for comparison, including rain removal : DDN[6], NLEDN[17], MSPFN[11]; hazy removal: AODNet[15], PFFN[21], MSBDN[7]; raindrop removal: AttentGAN[23], Quen *et al*. [25]. It is important to note that the compared networks are trained on the dataset of single tasks and our network is trained on the different bad weather dataset which merges all the above-mentioned datasets.

4.4.1 Qualitative Results

PSNR(Peak Signal to Noise Ratio) and SSIM(Structural Similarity) are recognized image quality evaluation standards. We use them as the metrics to compare our network with some state-of-the-art methods in various tasks. The results are shown in Table 2, Table 3 and Table 4. Even

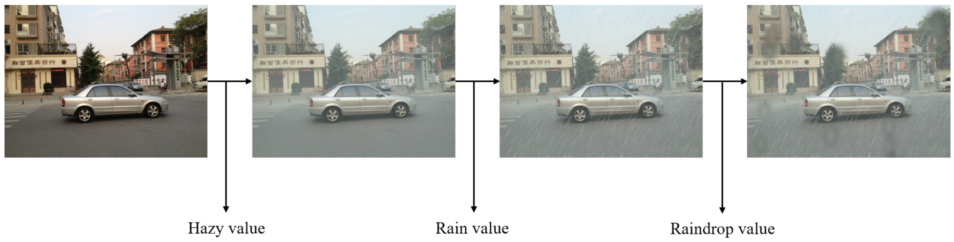

Figure 5. The generation process of mixed degradation factors dataset

Figure 6. Visualization of results in ablation study

Figure 7. Rain removal results of our method compared with state-of-the-art rain removal methods.

though our network is designed for removing multiple degradation factors, it still shows competitive results in PSNR and SSIM of the three tasks, which are close to the best results of the state-of-the-art methods.

4.4.2 Quantitative Results

Figure 7, Figure 8 and Figure 9. show the output of different methods respectively. We can see that our restoration results have achieved competitive results in color restoration, and preserved the details well. In Figure 7, the girl's face: It can be seen that the texture of the skin and lip is well preserved rather than becoming smooth and the athlete: The number on the vest still remains clear and well-recognized. In Figure 9, We can see that the edges of the object and building haven't turned blunt, but still sharp enough to be distinguished.

5. Conclusion

In this paper, we propose an end-to-end mixed degradation factors image restoration network and complex weather dataset. The key of our network to adapt combined distortions is the Task Adaptive Attention Module, which includes Task Adaptor, Task Operation Attention and Task Channel Attention. The above-mentioned method can adaptively generate corresponding weights for different degradation factors. Results in complex weather dataset show that our network can restore mixed bad weather degradation factors images robustly. At the same time, competitive experimental results are obtained on the mixed dataset of rain, haze and raindrop removal compared with state-of-the-art methods.

Acknowledgement

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Key Project 2019B1515120032, in part by the Na-

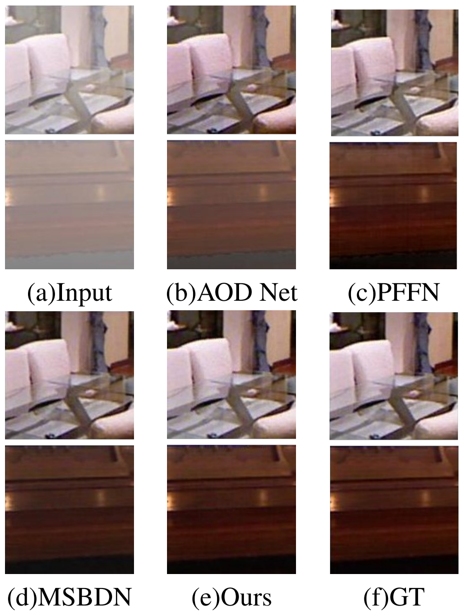

Figure 8. Hazy removal results of our method compared with state-of-the-art hazy removal methods.

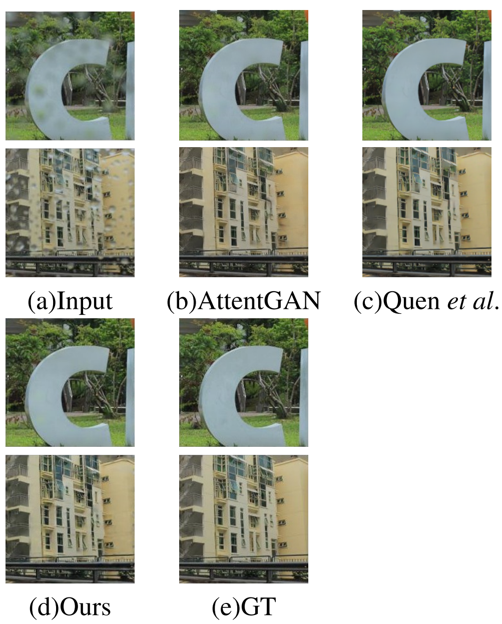

Figure 9. 7.Raindrop removal results of our method compared with state-of-the-art adherent Raindrop removal methods.

tional Science Foundation of China (NSFC) with grant no. 61901534, in part by the Science, Technology and Innovation Commission of Shenzhen Municipality with grant no. JCYJ20210324120002007 and JCYJ20190807155617099. Corresponding author: Congduan Li.

References

[1] 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.

[2] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.

[3] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, L. Zhang, Dongdong 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)*, pages 1375–1383, 2019.

[4] Deniz Engin, Anil Genç, and H. K. Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 938–9388, 2018.

[5] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv: Learning*, 2019.

[6] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1715–1723, 2017.

[7] Dong Hang, Pan Jinshan, Hu Zhe, Lei Xiang, Zhang Xinyi, Wang Fei, and Yang Ming-Hsuan. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, 2020.

[8] 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, 2011.

[9] Andrew G. Howard, M. Sandler, Grace Chu, Liang-Chieh Chen, B. Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324, 2019.

[10] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[11] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8343–8352, 2020.

[12] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8183–8192, 2018.

[13] Mohit Lamba, Atul Balaji, and Kaushik Mitra. Towards fast and light-weight restoration of dark images. *arXiv preprint arXiv:2011.14133*, 2020.

[14] Stamatios Lefkimmiatis. Universal denoising networks : A novel cnn architecture for image denoising. In *2018*

IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3204–3213, 2018.

[15] 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.

[16] 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, 2018.

[17] Guanbin Li, Xiang He, Wei Zhang, Huiyou Chang, Le Dong, and Liang Lin. Non-locally enhanced encoder-decoder network for single image de-raining. In Susanne Boll, Kyoung Mu Lee, Jiebo Luo, Wenwu Zhu, Hyeran Byun, Chang Wen Chen, Rainer Lienhart, and Tao Mei, editors, *2018 ACM Multimedia Conference on Multimedia Conference, MM 2018, Seoul, Republic of Korea, October 22-26, 2018*, pages 1056–1064. ACM, 2018.

[18] Ruoteng Li, Robby T. Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3172–3182, 2020.

[19] X. Li, Jianlong Wu, Zhouchen Lin, H. Liu, and H. Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*, 2018.

[20] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3397–3405, 2015.

[21] Kangfu Mei, Aiwen Jiang, Juncheng Li, and Mingwen Wang. Progressive feature fusion network for realistic image dehazing. In *Asian Conference on Computer Vision (ACCV)*, 2018.

[22] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 257–265, 2017.

[23] Rui Qian, R. Tan, W. Yang, J. Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2482–2491, 2018.

[24] X. Qin, Z. Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI*, 2020.

[25] Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. Deep learning for seeing through window with raindrops. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2463–2471, 2019.

[26] Peng-Lang Shui. Image denoising algorithm via doubly local wiener filtering with directional windows in wavelet domain. *IEEE Signal Processing Letters*, 12(10):681–684, 2005.

[27] Masanori Suganuma, X. Liu, and Takayuki Okatani. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9031–9040, 2019.

[28] Masanori Suganuma, Xing Liu, and Takayuki Okatani. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[29] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 769–777, 2015.

[30] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7794–7803, 2018.

[31] Zhengyang Wang and Shuiwang Ji. Smoothed dilated convolutions for improved dense prediction. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, Jul 2018.

[32] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 3–19, Cham, 2018. Springer International Publishing.

[33] Ke Xu, Xin Yang, Baocai Yin, and Rynson W.H. Lau. Learning to restore low-light images via decomposition-and-enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2278–2287, 2020.

[34] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3060–3069, 2020.

[35] K. Yu, Chao Dong, L. Lin, and Chen Change Loy. Crafting a toolchain for image restoration by deep reinforcement learning. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2443–2452, 2018.

[36] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, F. Khan, Ming-Hsuan Yang, and L. Shao. Multi-stage progressive image restoration. *ArXiv*, abs/2102.02808, 2021.

[37] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[38] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082*, 2019.