

# Dual Residual Networks Leveraging the Potential of Paired Operations for Image Restoration

Xing Liu<sup>†</sup> Masanori Suganuma<sup>†‡</sup> Zhun Sun<sup>‡</sup> Takayuki Okatani<sup>†‡</sup>

<sup>†</sup>Graduate School of Information Sciences, Tohoku University <sup>‡</sup>RIKEN Center for AIP

{ryu,suganuma,zhun,okalani}@vision.is.tohoku.ac.jp

## Abstract

In this paper, we study design of deep neural networks for tasks of image restoration. We propose a novel style of residual connections dubbed “dual residual connection”, which exploits the potential of paired operations, e.g., up- and down-sampling or convolution with large- and small-size kernels. We design a modular block implementing this connection style; it is equipped with two containers to which arbitrary paired operations are inserted. Adopting the “unraveled” view of the residual networks proposed by Veit *et al.*, we point out that a stack of the proposed modular blocks allows the first operation in a block interact with the second operation in any subsequent blocks. Specifying the two operations in each of the stacked blocks, we build a complete network for each individual task of image restoration. We experimentally evaluate the proposed approach on five image restoration tasks using nine datasets. The results show that the proposed networks with properly chosen paired operations outperform previous methods on almost all of the tasks and datasets.

## 1. Introduction

The task of restoring the original image from its degraded version, or image restoration, has been studied for a long time in the fields of image processing and computer vision. As in many other tasks of computer vision, the employment of deep convolutional networks have made significant progress. In this study, aiming at further improvements, we pursue better architectural design of networks, particularly the design that can be shared across different tasks of image restoration. In this study, we pay attention to the effectiveness of paired operations on various image processing tasks. In [11], it is shown that a CNN iteratively performing a pair of up-sampling and down-sampling contributes to performance improvement for image-superresolution. In [37], the authors employ evolutionary computation to search for a better design of con-

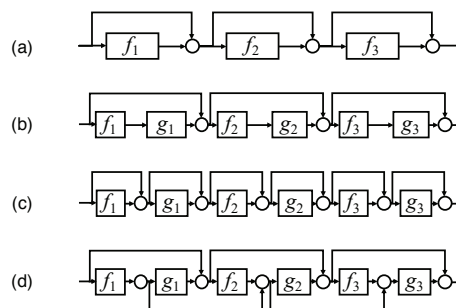


Figure 1: Different construction of residual networks with a single or double basic modules. The proposed “dual residual connection” is (d).

volutional autoencoders for several tasks of image restoration, showing that network structures repeatedly performing a pair of convolutions with a large- and small-size kernels (e.g., a sequence of conv. layers with kernel size 3, 1, 3, 1, 5, 3, and 1) perform well for image denoising. In this paper, we will show further examples for other image restoration tasks. Assuming the effectiveness of such repetitive paired operations, we wish to implement them in deep networks to exploit their potential. We are specifically interested in how to integrate them with the structure of residual networks. The basic structure of residual networks is shown in Fig. 1(a), which have become an indispensable component for the design of modern deep neural networks. There have been several explanations for the effectiveness of the residual networks. A widely accepted one is the “unraveled” view proposed by Veit *et al.* [41]: a sequential connection of  $n$  residual blocks is regarded as an ensemble of many sub-networks corresponding to its implicit  $2^n$  paths. A network of three residual blocks with modules  $f_1$ ,  $f_2$ , and  $f_3$ , shown in Fig. 1(a), has  $(2^3 =) 8$  implicit paths from the input to output, i.e.,  $f_1 \rightarrow f_2 \rightarrow f_3$ ,  $f_1 \rightarrow f_2$ ,  $f_1 \rightarrow f_3$ ,  $f_2 \rightarrow f_3$ ,  $f_1$ ,  $f_2$ ,  $f_3$ , and 1. Veit *et al.* also showed that each block works as a computational unit that can be attached/detached to/from the main network with minimum performance loss. Considering such a property of residual networks, how should we use resid-

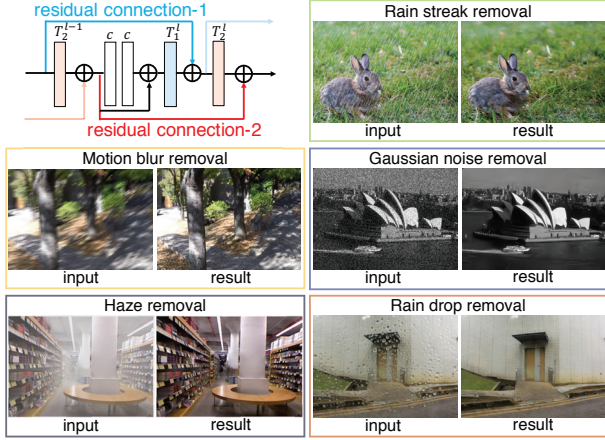


Figure 2: Upper-left: the structure of a unit block having the proposed dual residual connections;  $T_1^l$  and  $T_2^l$  are the containers for two paired operations;  $c$  denotes a convolutional layer. Other panels: five image restoration tasks considered in this paper.

ual connections for paired operations? Denoting the paired operations by  $f$  and  $g$ , the most basic construction will be to treat  $(f_i, g_i)$  as a unit module, as shown in Fig. 1(b). In this connection style,  $f_i$  and  $g_i$  are always paired for any  $i$  in the possible paths. In this paper, we consider another connection style shown in Fig. 1(d), dubbed “dual residual connection”. This style enables to pair  $f_i$  and  $g_j$  for any  $i$  and  $j$  such that  $i \leq j$ . In the example of Fig. 1(d), all the combinations of the two operations,  $(f_1, g_1)$ ,  $(f_2, g_2)$ ,  $(f_3, g_3)$ ,  $(f_1, g_2)$ ,  $(f_1, g_3)$ , and  $(f_2, g_3)$ , emerge in the possible paths. We conjecture that this increased number of potential interactions between  $\{f_i\}$  and  $\{g_j\}$  will contribute to improve performance for image restoration tasks. Note that it is guaranteed that  $f$  and  $g$  are always paired in the possible paths. This is not the case with other connection styles such as the one depicted in Fig. 1(c). We call the building block for implementing the proposed dual residual connections *Dual Residual Block* (DuRB); see Fig. 2. We examine its effectiveness on five image restoration tasks shown in Fig. 2 using nine datasets. DuRB is a generic structure that has two containers for the paired operations, and the users choose two operations for them. For each task, we specify the paired operations of DuRBs as well as the entire network. Our experimental results show that our networks outperform the state-of-the-art methods in these tasks, which supports the effectiveness of our approach.

## 2. Related Work

**Gaussian noise removal** Application of neural networks to noise removal has a long history [1, 18, 43, 53, 54]. Mao *et al.* [26] proposed REDNet, which consists of multiple con-

volutional and de-convolutional layers with symmetric skip connections over them. Tai *et al.* [39] proposed MemNet with local memory blocks and global dense connections, showing that it performs better than REDNet. However, Suganuma *et al.* [37] showed that standard convolutional autoencoders with repetitive pairs of convolutional layers with large- and small-size kernels outperform them by a good margin, which are found by architectural search based on evolutionary computation.

**Motion blur removal** This task has a long history of research. Early works [2, 7, 45, 46] attempt to simultaneously estimate both blur kernels and sharp images. Recently, CNN-based methods [9, 20, 28, 38, 42] achieve good performance for this task. Nah *et al.* [28] proposed a coarse-to-fine approach along with a modified residual block [14]. Kupyn *et al.* [20] proposed an approach based on Generative Adversarial Network (GAN) [10]. New datasets were created in [28] and [20].

**Haze removal** Many studies assume the following model of haze:  $I(x) = J(x)t(x) + A(x)(1 - t(x))$ , where  $I$  denotes a hazy scene image,  $J$  is the true scene radiance (the clear image),  $t$  is a transmission map,  $A$  is global atmospheric light. The task is then to estimate  $A$ ,  $t$ , and thus  $J(x)$  from the input  $I(x)$  [4, 12, 27, 48, 51]. Recently, Zhang *et al.* [51] proposed a method that uses CNNs to jointly estimate  $t$  and  $A$ , which outperforms previous approaches by a large margin. Ren *et al.* [32] and Li *et al.* [24] proposed method to directly estimate  $J(x)$  without explicitly estimating  $t$  and  $A$ . Yang *et al.* [48] proposed a method that integrates CNNs to classical prior-based method.

**Raindrop detection and removal** Various approaches [19, 21, 34, 47, 50] have been proposed to tackle this problem in the literature. Kurihata *et al.* [21] proposed to detect raindrops with raindrop-templates learned using PCA. Ramensh [19] proposed a method based on K-Means clustering and median filtering to estimate clear images. Recently, Qian *et al.* [30] proposed a hybrid network consisting of a convolutional-LSTM for localizing raindrops and a CNN for generating clear images, which is trained in a GAN framework.

**Rain-streak removal** Fu *et al.* [8] use “guided image filtering” [13] to extract high-frequency components of an image, and use it to train a CNN for rain-streak removal. Zhang *et al.* [52] proposed to jointly estimate rain density and de-raining result to alleviate the non-uniform rain density problem. Li *et al.* [25] regards a heavy rainy image as a clear image added by an accumulation of multiple rain-streak layers and proposed a RNN-based method to restore the clear image. Li *et al.* [23] proposed a non-locally enhanced version of DenseBlock [16] for this task, their network outperforms previous approaches by a good margin.

Table 1: Performance of the three connection types of Fig. 1(b)-(c). ‘-’s indicate infeasible applications.

	Gaussian noise	Real noise	Motion blur	Haze	Raindrop	Rain-streak
(b)	24.92 / 0.6632	36.76 / 0.9620	29.46 / 0.9035	31.20 / 0.9803	24.70 / 0.8104	32.85 / 0.9214
(c)	24.85 / 0.6568	36.81 / 0.9627	-/-	-/-	25.12 / 0.8151	33.13 / 0.9222
(d)	<b>25.05 / 0.6755</b>	<b>36.84 / 0.9635</b>	<b>29.90 / 0.9100</b>	<b>32.60 / 0.9827</b>	<b>25.32 / 0.8173</b>	<b>33.21 / 0.9251</b>

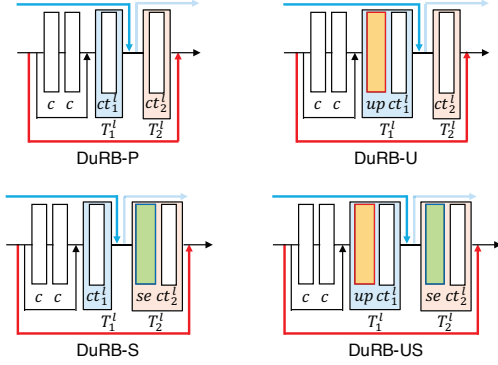


Figure 3: Four different implementations of the DuRB;  $c$  is a convolutional layer with  $3 \times 3$  kernels;  $ct_1^l$  and  $ct_2^l$  are convolutional layers, each with kernels of a specified size and dilation rate;  $up$  is up-sampling (we implemented it using PixelShuffle [36]);  $se$  is SE-ResNet Module [15] that is in fact a channel-wise attention mechanism.

### 3. Dual Residual Blocks

The basic structure of the proposed Dual Residual Block (DuRB) is shown in the upper-left corner of Fig. 2, in which we use  $c$  to denote a convolutional layer (with  $3 \times 3$  kernels) and  $T_1^l$  and  $T_2^l$  to denote the containers for the paired first and second operations, respectively, in the  $l^{th}$  DuRB in a network. Normalization layers (such as batch normalization [17] or instance normalization [40]) and ReLU [29] layers can be incorporated when it is necessary. We design DuRBs for each individual task, or equivalently choose the two operations to be inserted into the containers  $T_1^l$  and  $T_2^l$ . We will use four different designs of DuRBs, DuRB-P, DuRB-U, DuRB-S, and DuRB-US, which are shown in Fig. 3. The specified operations for  $[T_1^l, T_2^l]$  are [conv., conv.] for DuRB-P, [up-sampling+conv., down-sampling (by conv. with stride=2)] for DuRB-U, [conv., channel-wise attention<sup>1</sup>+conv.] for DuRB-S, and [up-sampling+conv., channel-wise attention+down-sampling] for DuRB-US, respectively. We will use DuRB-P for noise removal and raindrop removal, DuRB-U for motion blur removal, DuRB-S for rain-streak and raindrop removal, and DuRB-US for haze removal.

Before proceeding to further discussions, we present here experimental results that show the superiority of the proposed dual residual connection to other connec-

<sup>1</sup>It is implemented using the SE-ResNet Module [15].

tion styles shown in Fig. 1(b) and (c). In the experiments, three networks build on the three base structures (b), (c), and (d) of Fig. 1 were evaluated on the five tasks. For Gaussian&real-world noise removal, motion blur removal, haze removal, raindrop and rain-streak removal, we use DuRB-P, DuRB-U, DuRB-US, DuRB-S&DuRB-P and DuRB-S to construct the base structures. Number of blocks and all the operations in the three structures as well as other experimental configurations are fixed in each comparison. The datasets for the six comparisons are BSD-grayscale, Real-World Noisy Image Dataset, GoPro Dataset, Dehaze Dataset, RainDrop Dataset and DID-MDN Data. Table 1 shows their performance. Note that ‘-’ in the table indicate that the connection cannot be applied to DuRB-U and DuRB-US due to the difference in size between the output of  $f$  and the input to  $g$ . It can be seen that the proposed structure (d) performs the best for all the tests.

### 4. Five Image Restoration Tasks

In this section, we describe how the proposed DuRBs can be applied to multiple image restoration tasks, noise removal, motion blur removal, haze removal, raindrop removal and rain-streak removal.

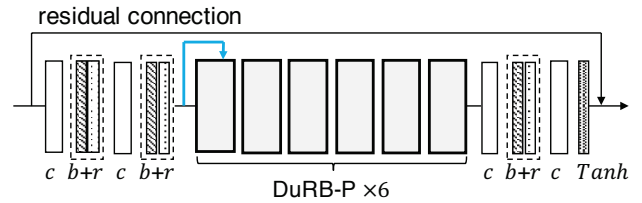


Figure 4: DuRN-P: dual residual network with DuRB-P’s [conv. w/ a large kernel and conv. w/ a small kernel] for Gaussian noise removal.  $b+r$  is a batch normalization layer followed by a ReLU layer; and  $Tanh$  denotes hyperbolic tangent function.

#### 4.1. Noise Removal

**Network Design** We design the entire network as shown in Fig. 4. It consists of an input block, the stack of six DuRBs, and an output block, additionally with an outermost residual connection from the input to output. The layers  $c$ ,  $b+r$  and  $Tanh$  in the input and output blocks are convolutional layer (with  $3 \times 3$  kernels, stride = 1), batch normalization layer



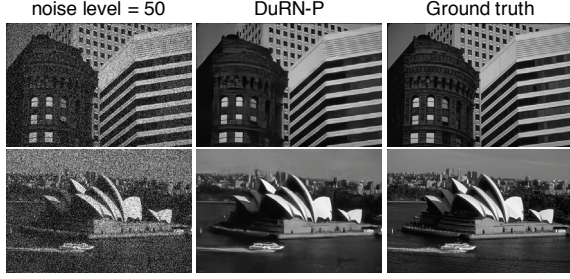


Figure 5: Some examples of the results by the proposed DuRN-P for additive Gaussian noise removal. Sharp images can be restored from heavy noises ( $\sigma = 50$ ).

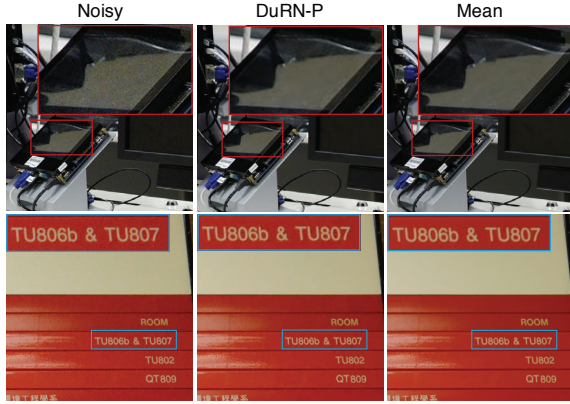


Figure 6: Examples of noise removal by the proposed DuRN-P for images from Real-World Noisy Image Dataset. The results are sometimes even better than the mean image (used as the ground truth); see the artifact around the letters in the bottom.

followed by a ReLU layer, and hyperbolic tangent function layer, respectively.

We employ DuRB-P (i.e., the design in which each of the two operations is single convolution; see Fig. 3) for DuRBs in the network. Inspired by the networks discovered by neural architectural search for noise removal in [37], we choose for  $T_1$  and  $T_2$  convolution with large- and small-size receptive fields. We also choose the kernel size and dilation rate for each DuRB so that the receptive field of convolution in each DuRB grows its size with  $l$ . More details are given in the supplementary material. We set the number of channels to 32 for all the layers. We call the entire network DuRN-P. For this task, we employed  $l_2$  loss for training the DuRN-P.

**Results: Additive Gaussian Noise Removal** We tested the proposed network on the task of removing additive Gaussian noise of three levels (30, 50, 70) from a grayscale noisy image. Following the same experimental protocols used by previous studies, we trained and tested the proposed DuRN-P using the training and test subsets (300 and 200 grayscale images) of the BSD-grayscale dataset.

Table 2: Results for additive Gaussian noise removal on BSD200-grayscale and noise levels (30, 50, 70). The numbers are PSNR/SSIM.

	30	50	70
REDNet [26]	27.95 / 0.8019	25.75 / 0.7167	24.37 / 0.6551
MemNet [39]	28.04 / 0.8053	25.86 / 0.7202	24.53 / 0.6608
E-CAE [37]	28.23 / 0.8047	26.17 / 0.7255	24.83 / 0.6636
<b>DuRN-P (ours)</b>	<b>28.50 / 0.8156</b>	<b>26.36 / 0.7350</b>	<b>25.05 / 0.6755</b>

Table 3: Results on the Real-World Noisy Image Dataset [44]. The results were measured by PSNR/SSIM. The last row shows the number of parameters for each CNN.

	REDNet [26]	MemNet [39]	E-CAE [37]	<b>DuRN (ours)</b>
PSNR/SSIM	35.56 / 0.9475	- / -	35.45 / 0.9492	<b>36.83 / 0.9635</b>
# of param.	$4.1 \times 10^6$	$2.9 \times 10^6$	$1.1 \times 10^6$	$8.2 \times 10^5$

More details of the experiments are provided in the supplementary material. We show the quantitative results in Table 2 and qualitative results in Fig. 5. It is observed from Table 2 that the proposed network outperforms the previous methods for all three noise levels.

**Results: Real-World Noise Removal** We also tested the DuRN-P on the Real-World Noisy Image Dataset [44], which consists of 40 pairs of an instance image (a photograph taken by a CMOS camera) and the mean image (mean of multiple shots of the same scene taken by the CMOS camera). We removed all the batch normalization layers from the DuRN-P for this experiment, as the real-world noise captured in this dataset do not vary greatly. The details of the experiments are given in the supplementary material. The quantitative results of three previous methods and our method are shown in Table 3. We used the authors’ code to evaluate the three previous methods. (As the MemNet failed to produce a competitive result, we left the cell empty for it in the table.) It is seen that our method achieves the best result despite the smaller number of parameters. Examples of output images are shown in Fig. 6. We can observe that the proposed DuRN-P has cleaned noises well. It is noteworthy that the DuRN-P sometimes provides better images than the “ground truth” mean image; see the bottom example in Fig. 6.

## 4.2. Motion Blur Removal

The task is to restore a sharp image from its motion blurred version without knowing the latent blur kernels (i.e., the “blind-deblurring” problem).

**Network Design** Previous works such as [42] reported that the employment of up- and down-sampling operations is effective for this task. Following this finding, we employ up-sampling and down-sampling for the paired operation. We call this as DuRB-U; see Fig. 3. We use PixelShuffle [36] for implementing up-sampling. For the entire network design, following many previous works [20, 42, 51, 55], we

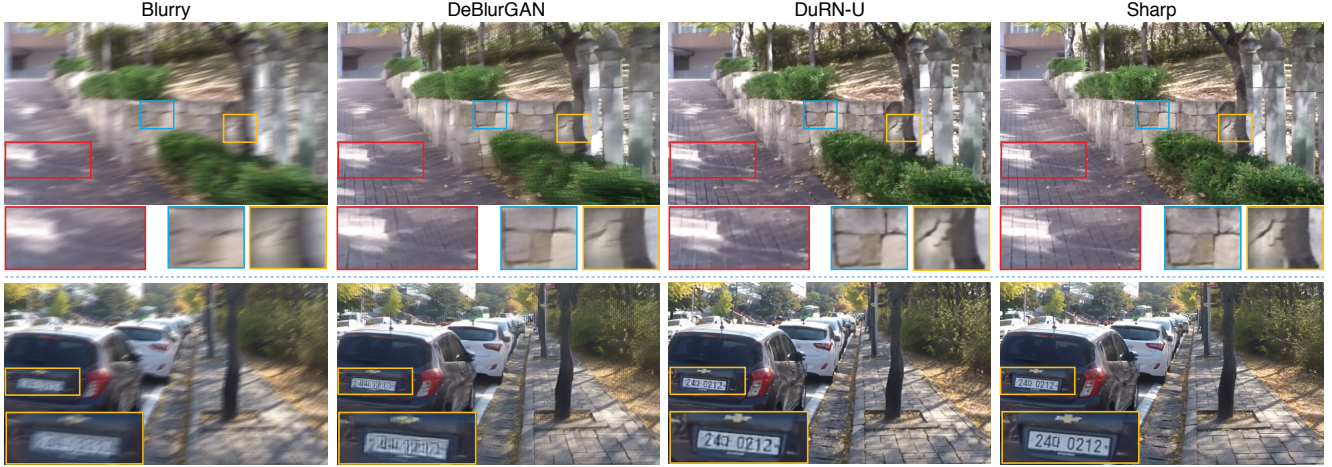


Figure 7: Examples of motion blur removal on GoPro-test dataset.

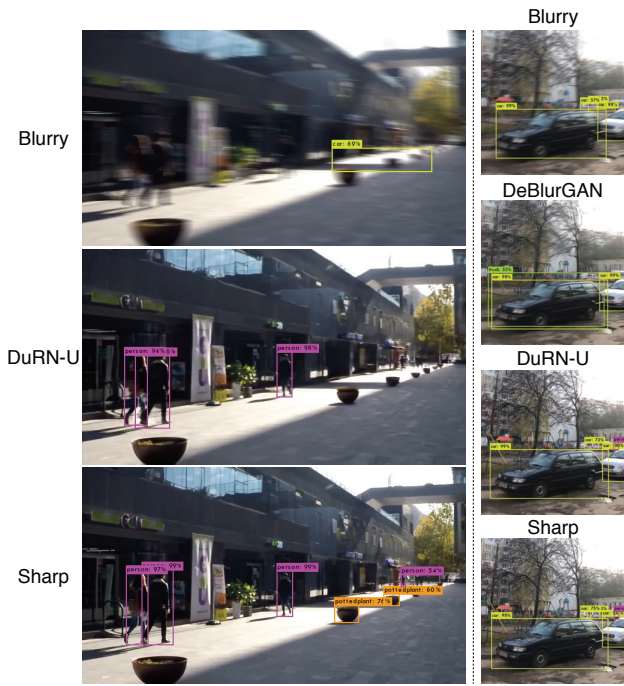


Figure 8: Examples of object detection from original blurred images and their deblurred versions.

choose a symmetric encoder-decoder network; see Fig. 9. The network consists of the initial block, which down-scales the input image by 4:1 down-sampling with two convolution operations ( $c$ ) with stride = 2, and instance normalization + ReLU ( $n + r$ ), and six repetitions of DuRB-U's, and the final block which up-scales the output of the last DuRB-U by applications of 1:2 up-sampling ( $up$ ) to the original size. We call this network DuRN-U. For this task, we employed a weighted sum of SSIM and  $l_1$  loss for training the DuRN-U. The details are given in the supp. material.

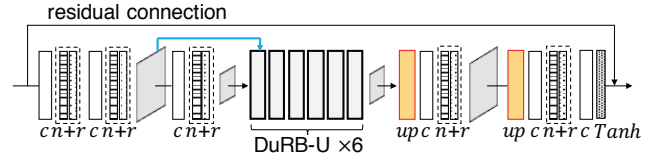


Figure 9: DuRN-U: Dual Residual Network with DuRB-U's (up- and down-sampling) for motion blur removal.  $n + r$  denotes an instance normalization layer followed by a ReLU layer.

Table 4: Results of motion blur removal for the GoPro-test dataset.

GoPro-test	
Sun <i>et al.</i> [38]	24.6 / 0.84
Nah <i>et al.</i> [28]	28.3 / 0.92
Xu <i>et al.</i> [46]	25.1 / 0.89
DeBlurGAN [20]	27.2 / <b>0.95</b>
<b>DuRN-U (ours)</b>	<b>29.9 / 0.91</b>

Table 5: Accuracy of object detection from deblurred images obtained by DeBlurGAN [20] and the proposed DuRN-U on Car Dataset.

	Blurred	DeBlurGAN [20]	<b>DuRN-U (ours)</b>
mAP (%)	16.54	26.17	<b>31.15</b>

**Results: GoPro Dataset** We tested the proposed DuRN-U on the GoPro-test dataset [28] and compared its results with the state-of-the-art DeblurGAN<sup>2</sup> [20]. The GoPro dataset consists of 2,013 and 1,111 non-overlapped training (GoPro-train) and test (GoPro-test) pairs of blurred and sharp images. We show quantitative results in the Table 4. DeblurGAN yields outstanding SSIM number, whereas the

<sup>2</sup>The DeblurGAN refers the “DeblurGAN-wild” introduced in the original paper [20].



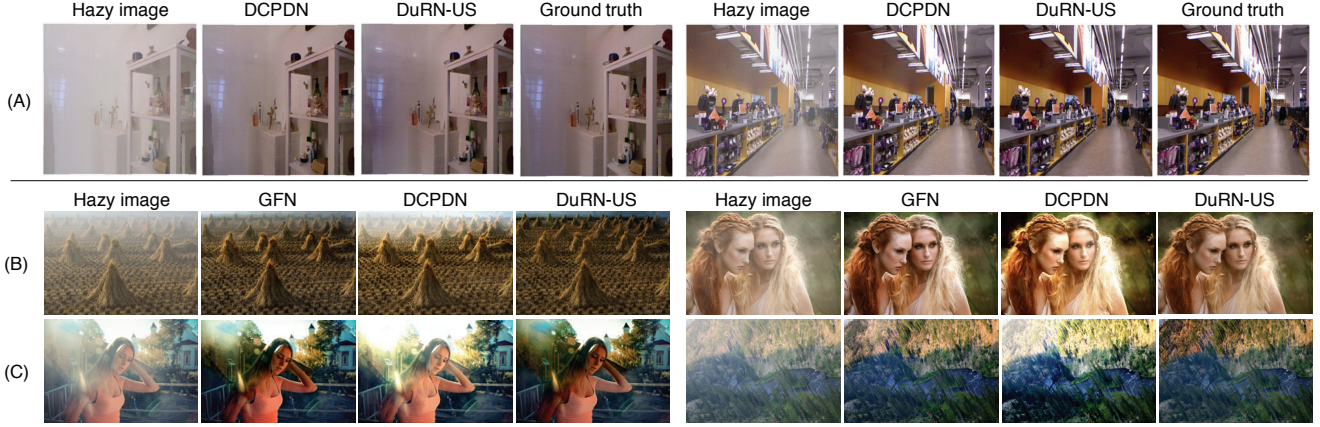


Figure 10: Examples of de-hazing results obtained by DuRN-US and others on (A) synthesized images, (B) real images and (C) light hazy images.

proposed DuRN-U is the best in terms of PSNR. Examples of deblurred images are shown in Fig. 7. It is observed that the details such as cracks on a stone-fence or numbers written on the car plate are restored well enough to be recognized.

**Results: Object Detection from Deblurred Images** In [20], the authors evaluated their deblurring method (DeBlurGAN) by applying an object detector to the deblurred images obtained by their method. Following the same procedure and data (Car Dataset), we evaluate our DuRN-U that is trained on the GoPro-train dataset. The Car Dataset contains 1,151 pairs of blurred and sharp images of cars. We employ YOLO v3 [31] trained using the Pascal VOC [6] for the object detector. The detection results obtained for the sharp image by the same YOLO v3 detector are utilized as the ground truths used for evaluation. Table 5 shows quantitative results (measured by mAP), from which it is seen that the proposed DuRN-U outperforms the state-of-the-art DeBlurGAN. Figure 8 shows examples of detection results on the GoPro-test dataset and Car Dataset. It is observed that DuRN-U can recover details to a certain extent that improves accuracy of detection.

### 4.3. Haze Removal

**Network Design** In contrast with previous studies where a CNN is used to explicitly estimate a transmission map that models the effects of haze, we pursue a different strategy, which is to implicitly estimate a transmission map using an attention mechanism. Our model estimates the de-hazed image from an input image in an end-to-end fashion. We design DuRB’s for this task by employing up-sampling (*up*) implemented using PixelShuffle [36] with a convolutional layer ( $ct_1^l$ ) in  $T_1^l$  and channel-wise attention (*se*) implemented using SE-ResNet module [15] with a conv. layer ( $ct_2^l$ ) in  $T_2^l$ . More details are given in the supplementary

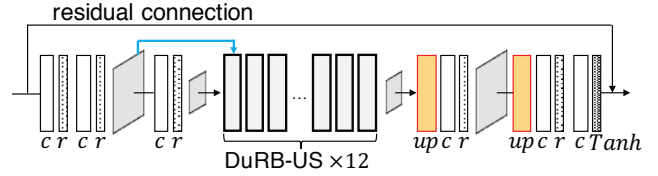


Figure 11: DuRN-US: dual residual network with DuRB-US’s (up- and down-sampling and channel-wise attention (SE-ResNet Module)) for haze removal.

Table 6: Results for haze removal on Dehaze-TestA dataset and RESIDE-SOTS dataset.

Dehaze-TestA		RESIDE-SOTS	
He <i>et al.</i> [12]	0.8642	Berman <i>et al.</i> [3]	17.27 / 0.75
Zhu <i>et al.</i> [56]	0.8567	Ren <i>et al.</i> [32]	17.57 / 0.81
Berman <i>et al.</i> [3]	0.7959	Cai <i>et al.</i> [5]	21.14 / 0.85
Li <i>et al.</i> [22]	0.8842	Li <i>et al.</i> [22]	19.06 / 0.85
Zhang <i>et al.</i> [51]	0.9560	Ren <i>et al.</i> [33]	22.30 / 0.88
<b>DuRN-US (ours)</b>	<b>0.9827</b>	<b>DuRN-US (ours)</b>	<b>32.12 / 0.98</b>

material. The entire network (named DuRN-US) has an encoder-decoder structure similar to the DuRN-U designed for motion blur removal, as shown in Fig. 11. We stack 12 DuRB-US’s in the middle of the network; the number of channels is 64 for all the layers. In the supplementary material, we demonstrate how our network estimates a transmission map inside its attention mechanisms. For this task, we employed a weighted sum of SSIM and  $l_1$  loss for training the DuRN-US.

**Results** In order to evaluate the proposed DuRN-US, we trained and tested it on two datasets, the Dehaze Dataset and the RESIDE dataset. The training and test (Dehaze-TestA) subsets in the Dehaze Dataset consist of 4,000 and 400 non-overlapped samples of indoor scenes, respectively. RESIDE contains a training subset of 13,990 samples of indoor scenes and a few test subsets. Following [33], we used



Figure 12: Examples of raindrop removal along with internal activation maps of DuRN-S-P. The “Attention map” and “Residual map” are the outputs of the Attentive-Net and the last  $Tanh$  layer shown in Fig. 13; they are normalized for better visibility.

a subset SOTS (Synthetic Objective Testing Set) that contains 500 indoor scene samples for evaluation. It should be noted that the state-of-the-art method on the Dehaze Dataset, DCPDN [51], is trained using i) hazy images, ii) ground truth images, iii) ground truth global atmosphere light, iv) ground truth transmission maps; additionally, its weights are initialized by those of DenseNet [16] pretrained on the ImageNet [35]. The proposed DuRN-US is trained only using i) and ii). Table 6 shows results on Dehaze-TestA and RESIDE-SOTS datasets, respectively.

Figure 10 shows examples of the results obtained by the proposed network and others for the same input images. In sub-figure (A), we show results for two synthesized images produced by the DCPDN (the second best approach in terms of SSIM and PSNR) and our DuRN-US. It is observed that DuRN-US yields better results for these two images. In sub-figure (B), we show results for two real-world hazy images<sup>3</sup> produced by two state-of-the-art methods, GFN [33] and DCPDN [51], and by ours. It can be observed that our network yields the most realistic dehazed images. It is noteworthy that our DuRN-US can properly deal with strong ambient light (sunshine coming behind the girl). See the example in the left-bottom of Fig. 10.

#### 4.4. Raindrop removal

**Network Design** The task can naturally be divided into two stages, that of identifying the regions of raindrops and that of recovering the pixels of the identified regions. The second stage is similar to image inpainting and may not be difficult, as there are a lot of successful methods for image inpainting. Then, the major issue is with the first stage. Following this two-stage approach, the state-of-the-art method [30] uses an attentive-recurrent network to produce an attention map that conveys information about raindrops; then, the attention map along with the input image

<sup>3</sup>The images are available from <https://github.com/rwenqi/GFN-dehazing>

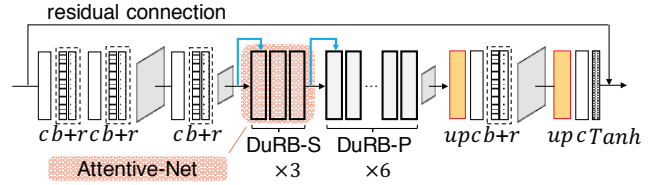


Figure 13: DuRN-S-P: Hybrid dual residual network with DuRB-S’s and DuRB-P’s for raindrop removal.

Table 7: Quantitative result comparison on RainDrop Dataset [30].

	Qian <i>et al.</i> [30]	DuRN-S-P (ours)
TestSetA	<b>31.51</b> / 0.9213	31.24 / <b>0.9259</b>
TestSetB	24.92 / 0.8090	<b>25.32</b> / <b>0.8173</b>

are fed to a convolutional encoder-decoder network to estimate the ground truth image. It also employs adversarial training with a discriminator to make the generated images realistic.

We show our DuRBs are powerful enough to perform these two-stage computations in a standard feedforward network, if we use properly designed DuRBs in proper positions in the entire network. To be specific, we choose the encoder-decoder structure for the entire network, and in its bottleneck part, we set three DuRB-S’s followed by six DuRB-P’s. For  $ct_1^l$  in the three DuRB-S’s, we use convolution with a  $3 \times 3$  kernel with decreasing dilation rates, 12, 8, and 6, in the forward direction, aiming to localize raindrops in a coarse-to-fine manner in the three DuRB-S’s in a row. For the six DuRB-P’s, we employ the same strategy as in noise removal etc., which is to apply a series of convolution with an increasing receptive field size in the forward direction. We call the entire network DuRN-S-P. For this task, we employed a weighted sum of SSIM and  $l_1$  loss for training the DuRN-S-P.





Figure 14: Examples of rain-streak removal obtained by four methods including ours (DuRN-S).

**Results** We trained and evaluated the DuRN-S-P on the RainDrop Dataset. It contains 861 training samples and 58/249 test samples called TestSetA/TestSetB. TestSetA is a subset of TestSetB, and is considered to have better alignment<sup>4</sup> than TestSetB. Table 7 shows the results. It is seen that our method outperforms the state-of-the-art method for three out of four combinations of two test sets and two evaluation metrics. It is noteworthy that our method does not use a recurrent network or adversarial training. Figure 12 shows some examples of the results obtained by our method and the method of [30]. It is seen that the results of our method are visually comparable to the method of [30]. The “Attention map” and “Residual map” of Fig. 12 are the over-channel summation of the output of Attentive-Net and the output of the last *Tanh* layer, respectively; see Fig. 13.

#### 4.5. Rain-streak Removal

**Network Design** It is shown in [23] that the mechanism that selectively weighs feature maps using global information works effectively for this task. Borrowing this idea, we employ a channel-wise attention mechanism to perform similar feature weighting. The overall design of the network for this task is similar to the DuRN-P designed for Gaussian noise removal. A difference is that we use DuRB-S instead of DuRB-P to use the attention mechanism. The details are given in the supplementary material. For this task, we employed a weighted sum of SSIM and  $l_1$  loss for training the network.

**Results** We tested the proposed network (DuRN-S) on two benchmark datasets, the DDN-Data, which consists of 9,100 training pairs and 4,900 test pairs of rainy and clear images, and the DID-MDN Data, which consists of 12,000 training pairs and 1,200 test pairs. Table 8 shows the results. Those for the previous methods except RESCAN [25] are imported from [23]. It is seen that the proposed network achieves the best performance. Examples of the output images are provided in Fig. 14.

<sup>4</sup><https://github.com/rui1996/DeRaindrop>

Table 8: Results on two de-raining datasets.

	DDN Data	DID-MDN Data
DDN [8]	28.24 / 0.8654	23.53 / 0.7057
JORDER [49]	28.72 / 0.8740	30.35 / 0.8763
DID-MDN [52]	26.17 / 0.8409	28.30 / 0.8707
RESCAN [25]	-/-	32.48 / 0.9096
NLEDN [23]	29.79 / 0.8976	33.16 / 0.9192
<b>DuRN-S (ours)</b>	<b>31.30 / 0.9194</b>	<b>33.21 / 0.9251</b>

## 5. Summary and Discussions

We have proposed a style of residual connection, dubbed “dual residual connection”, aiming to exploit the potential of paired operations for image restoration tasks. We have shown the design of a modular block (DuRB) that implements this connection style, which has two containers for the paired operations such that the user can insert any arbitrary operations to them. We have also shown choices of the two operations in the block as well as the entire networks (DuRN) containing a stack of the blocks for five different image restoration tasks. The experimental results obtained using nine datasets show that the proposed approach consistently works better than previous methods.

## Acknowledgement

This work was partly supported by JSPS KAKENHI Grant Number JP15H05919, JST CREST Grant Number JPMJCR14D1, Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (Infrastructure Maintenance, Renovation and Management), and the IMPACT Program Tough Robotics Challenge of the Council for Science, Technology, and Innovation (Cabinet Office, Government of Japan).

## References

- [1] Forest Agostinelli, Michael R Anderson, and Honglak Lee. Adaptive multi-column deep neural networks with application to robust image denoising. In *Proc. International Conference on Neural Information Processing Systems*, 2013.



- [2] S. Derin Babacan, Rafael Molina, Minh N. Do, and Aggelos K. Katsaggelos. Bayesian blind deconvolution with general sparse image priors. In *Proc. European Conference on Computer Vision*, 2012.
- [3] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2016.
- [4] Dana Berman, Tali Treibitz, and Shai Avidan. Air-light estimation using haze-lines. In *Proc. International Conference on Computational Photography*, 2017.
- [5] 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*, 2016.
- [6] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 2015.
- [7] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph. In *Proc. ACM SIGGRAPH*, 2006.
- [8] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2017.
- [9] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian D. Reid, Chunhua Shen, Anton van den Hengel, and Qinfeng Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *Proc. European Conference on Computer Vision*, 2017.
- [10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proc. International Conference on Neural Information Processing Systems*, 2014.
- [11] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2018.
- [12] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.
- [13] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2016.
- [15] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2018.
- [16] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2017.
- [17] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. International Conference on Machine Learning*, 2015.
- [18] Viren Jain and Sebastian Seung. Natural image denoising with convolutional networks. In *Proc. International Conference on Neural Information Processing Systems*, 2009.
- [19] M. Ramesh Kanthan and S. Naganandini Sujatha. Rain drop detection and removal using k-means clustering. In *Proc. International Conference on Computational Intelligence and Computing Research*, 2015.
- [20] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2018.
- [21] Hiroyuki Kurihata, Tatsuro S Takahashi, Ichiro Ide, Y. Mekada, Hiroshi Murase, Yukimasa Tamatsu, and Takayuki Miyahara. Rainy weather recognition from in-vehicle camera images for driver assistance. In *Proc. Intelligent Vehicles Symposium*, 2005.
- [22] Boyi Li, Xiulian Peng, Zhangyang Wang, Ji-Zheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proc. International Conference on Computer Vision*, 2017.
- [23] 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 *Proc. ACM International Conference on Multimedia*, 2018.
- [24] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2018.
- [25] Xia Li, Jianlong Wu, Zhouchen Lin, Hong W. Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proc. European Conference on Computer Vision*, 2018.
- [26] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Proc. International Conference on Neural Information Processing Systems*, 2016.
- [27] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *Proc. International Conference on Computer Vision*, 2013.
- [28] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2017.
- [29] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proc. International Conference on Machine Learning*, 2015.
- [30] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for rain-drop removal from a single image. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2018.
- [31] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv:1804.02767*, 2018.

- [32] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *Proc. European Conference on Computer Vision*, 2016.
- [33] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2018.
- [34] M. Roser and A. Geiger. Video-based raindrop detection for improved image registration. In *Proc. International Conference on Computer Vision Workshops*, 2009.
- [35] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Fei-Fei Li. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 2015.
- [36] Wenzhe Shi, Jose Caballero, Ferenc Huszar, 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 *Proc. Conference on Computer Vision and Pattern Recognition*, 2016.
- [37] Masanori Suganuma, Mete Ozay, and Takayuki Okatani. Exploiting the potential of standard convolutional autoencoders for image restoration by evolutionary search. In *Proc. International Conference on Machine Learning*, 2018.
- [38] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2015.
- [39] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proc. International Conference on Computer Vision*, 2017.
- [40] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv:1607.08022*, 2016.
- [41] Andreas Veit, Michael J. Wilber, and Serge Belongie. Residual networks behave like ensembles of relatively shallow networks. In *Proc. International Conference on Neural Information Processing Systems*, 2016.
- [42] Patrick Wieschollek, Michael Hirsch, Bernhard Schölkopf, and Hendrik P. A. Lensch. Learning blind motion deblurring. In *Proc. International Conference on Computer Vision*, 2017.
- [43] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *Proc. International Conference on Neural Information Processing Systems*, 2012.
- [44] Jun Xu, Hui Li, Zhetong Liang, David Zhang, and Lei Zhang. Real-world noisy image denoising: A new benchmark. *arXiv:1804.02603*, 2018.
- [45] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *Proc. European Conference on Computer Vision*, 2010.
- [46] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2013.
- [47] Atsushi Yamashita, Yuu Tanaka, and Toru Kaneko. Removal of adherent waterdrops from images acquired with stereo camera. In *Proc. International Conference on Intelligent Robots and Systems*, 2005.
- [48] Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *Proc. European Conference on Computer Vision*, 2018.
- [49] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Joint rain detection and removal from a single image. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2017.
- [50] Shaodi You, Robby T. Tan, Rei Kawakami, Yasuhiro Mukaigawa, and Katsushi Ikeuchi. Adherent raindrop modeling, detection and removal in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.
- [51] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2018.
- [52] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proc. Conference on Computer Vision and Pattern Recognition*, 2018.
- [53] 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*, 2017.
- [54] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for CNN based image denoising. *IEEE Transactions on Image Processing*, 2018.
- [55] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *Proc. International Conference on Computer Vision*, 2017.
- [56] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 2015.