Image Rain Removal and Illumination Enhancement Done in One Go

Yecong Wan[a,2], Yuanshuo Cheng[a,2], Mingwen Shao[a,*] and Jordi Gonzàlez[b]

[a]School of computer science and technology, China University of Petroleum (East China), QingDao, 266580, China
[b]Computer Vision Center, Univ. Autònoma de Barcelona, Barcelona, 08193, Spain

ARTICLE INFO

Keywords:
Image rain removal
Low-light image enhancement
Spatially-adaptive network
Contrastive learning

ABSTRACT

Rain removal plays an important role in the restoration of degraded images. Recently, CNN-based methods have achieved remarkable success. However, these approaches neglect that the appearance of rain is often accompanied by low light conditions in real-world rainy weather, which will further degrade the image quality, thereby hindering the restoration of the rain image. Therefore, it is very indispensable to jointly remove the rain and enhance illumination for real-world rain image restoration. To this end, in this paper we proposed a novel spatially-adaptive network, namely SANet, which can remove the rain and enhance illumination in one go with the guidance of degradation mask. Meanwhile, to take full use of negative samples, a contrastive loss is proposed to further improve the visual effect of the restored image. In addition, we present a new synthetic dataset, named DarkRain, to boost the development of rain image restoration algorithms in practical scenarios. DarkRain not only contains different degrees of rain, but also considers different lighting conditions, and more realistically simulates real-world rainfall scenarios. SANet is extensively evaluated on the proposed benchmark and achieves new state-of-the-art results. Moreover, after a simple transformation, our method outperforms existing the state-of-the-art methods in both rain removal and low-light image enhancement.

1. Introduction

Rain is a common weather phenomenon in our daily life, which may seriously affect the visual effect of the image. Images captured in rainy conditions usually suffer terrible visibility, low contrast, and insufficient lighting. These degenerations would severely affect the performance of outdoor vision systems. Hence, restoring photos taken in rainy weather is of great significance for outdoor vision applications, including but not limited to autonomous driving [29], surveillance systems [8].

In the last decade, various image deraining methods have been proposed to restore the image degraded by rain, ranging from model-driven [27, 56, 40] to data-driven [31, 38, 37, 39, 55], from supervised [22, 33, 34] to semi-supervised [46], and from synthetic dataset [49, 8, 52] to real-world dataset [43]. Recently, deep learning-based rain removal methods have achieved promising performance. Although these methods remove rain well in the image, they ignore the fact that real rainy days are often accompanied by low light and dark environments, which further degrade the image quality. Existing image deraining methods only consider the rain removal and cannot completely restore degraded images. Thus, in order to completely restore the quality of the rain image, it is essential to restore the brightness of the image while removing the rain.

One possible solution to this problem is to cascade rain removal followed by illumination enhancement or illumination enhancement followed by rain removal, but that will cause errors to accumulate and yields suboptimal results (see Sec. 4.3 for details). On the contrary, if a degraded image is restored in a uniform network, the restored image suffers from severe blurring because there are two different forms of degradation in the image, and the severity of degradation in different regions of the image is not unique.

Figure 1: Three example results of our proposed SANet on Rain12 [23] dataset. Left: Rainy images. Middle: Corresponding rain-free images. Right: Our results. Our SANet achieves more satisfying visual effect, while deraining alone cannot completely restore the image. **Best viewed at screen!**

To address the aforementioned issues, we propose a unified framework, dubbed Spatially-Adaptive Network (SANet), to remove the rain and enhance the illumination jointly. As illustrated in Fig. 1, our SANet not only fully remove the rain, but also adjust the lighting of a rainy image. In addition, it avoids the cumulative effect of errors and obtains higher-quality images.

*Corresponding author
yecongwan@gmail.com (Y. Wan); cys1294414023@gmail.com (Y. Cheng); smw278@126.com (M. Shao); Jordi.Gonzalez@uab.cat (J. Gonzàlez)
ORCID(s):
[2]These authors contributed equally to this work

Specifically, SANet contains four subnetworks, Rain Removal Network (RRNet), Illumination Estimation Network (IENet), Degradation Localization Network (DLNet) and Detail Recovery Network (DRNet) respectively. DLNet gathers information from the entire image to estimate a binary mask (localizing "hard" and "easy" regions) which is used to guide the image restoration of different degraded areas in the image. To make full use of the features extracted from the previous phase of the network we propose a Spatially-Adaptive Block (SABlock) to implement progressive feature enhancement. That is, the extracted degradation localization feature is transmitted to the RRNet by means of SABlock to help the rain removal. Similarly, the extracted features of the rain-free image are transmitted to the IENet by means of SABlock to help illumination enhancement. Considering that the encoder-decoder structure is unreliable in preserving spatial details, we employ a detail recovery network to restore the desired fine texture. DRNet consists of multiple Detail Recovery Blocks (DRBlocks), which have larger reception fields and make full use of original and generated images for better detail recovery. Moreover, a contrastive loss is proposed to enhance the visual effect of images, which ensures that the restored image is pulled to closer to the groundtruth image and pushed to far away from the degraded images in the representation space. To efficiently train the proposed network, we manage to synthesize a large scale dataset termed DarkRain, which consists of 5075 image pairs with different rain-density and luminance. To the best of our knowledge, it is the first synthetic dataset that takes lighting conditions into account in a simulated rainfall environment.

Overall, our contributions are summarized as follows:

- We propose a novel end-to-end unified framework, named SANet, to simultaneously remove rain and enhance illumination. With the guidance of DLNet, RRNet and IENet can treat different degraded areas of the image differently and restore the degraded image more efficiently.
- To take full advantage of negative samples, a contrastive loss is designed to obtain more vivid results.
- A new synthetic dataset is synthesized to train our network. Although the network is trained on synthetic dataset, we observe that it generalizes very well to real-world rainy images, because the dataset more realistically simulates the rainfall environment.
- Apart from achieving state-of-the-art performance on our proposed benchmark, our method can be transferred to only rain removal or low-light image enhancement after a simple replacement, and outperforms existing methods.

2. Related work

2.1. Rain removal

Rain removal methods can be categorized into two groups: video rain removal [3, 25, 24] and single image rain removal [6, 31, 28]. Comparatively, removing rain from an individual image is more challenging due to the lack of time-series information [7, 24]. Previous model-based single image deraining methods can not remove the rain in complex scenes due to the limited linear-mapping transformation [23, 16]. Recently, data-driven methods [48, 2, 21, 10, 12] have demonstrated their superiority in single image rain removal. For instance, Fu *et al.* [8] decomposed a rain image into a low-frequency structure layer and a high-frequency detail layer and then removed rain only in the high-frequency layer. Yang *et al.* [49] designed a deep recurrent dilated network to jointly detect and remove rain streaks. By enforcing an additional constraint, Zhang *et al.* [52] introduced a conditional generative adversarial networks (CGAN) to get better visual quality. Later, a recurrent network [22] is designed to progressively remove rain streaks from the image. Inspired by direction-aware attention mechanism [13], Wang *et al.* [43] utilize spatial attention to better remove rain with diverse appearances. More recently, Deng *et al.* [5] introduced two parallel sub-networks that synergize to remove rain and recover lost details caused by deraining. Jiang *et al.* [17] constructed a multi-scale pyramid structure to remove rain by using multi-scale rain representation information. Wang *et al.* [40] designed a fully interpretable network structure which can automatically extract rain kernels and proximal operators, faithfully characterizing the features of both rain and clean background layers. In order to balance spatial details and high-level contextualized information, Zamir *et al.* [51] proposed a multi-stage architecture, which can break down the overall recovery process into more manageable steps, and remove rain progressively.

Nonetheless, above rain-removal methods only remove the rain and do not take the illumination enhancement into account.

2.2. Low-light image enhancement

Data-driven methods have achieved remarkable success in low-light image enhancement [42, 47, 41]. Specifically, Wei *et al.* [45] decomposed a low-light image into reflectance and illumination, and conducted lightness enhancement on illumination. Further, Zhang *et al.* [54] connected the feature-level information of illumination and reflectance in the decomposition step. Considering that image enhancement can be achieved by an intuitive and simple nonlinear curve mapping, Guo *et al.* [9, 20] proposed a lightweight deep network to estimate pixel-wise and high-order curves for dynamic range adjustment of low-light image. Through elaborately designed discriminators and loss functions, Jiang *et al.* [18] introduced an unsupervised GAN-based method that can be trained without low/normal-light image pairs. Upon Retinex rule, Liu *et al.* [26] established models to characterize the intrinsic underexposed structure of low-light images and unroll their optimization processes to brighten the low-light images.

Retinex model has been demonstrated to be effective for low-light image enhancement. Similar to that of [42, 26],

in this paper, we also enhance the illumination of low-light rainy image based on Retinex theory.

3. Proposed method

In order to remove rain and enhance illumination in one go, we propose a novel entanglement network. In this section, we first introduce the formulation of rain and illumination in Sec. 3.1. Then, we present our spatially-adaptive network architecture in Sec. 3.2. Moreover, we introduce our elaborately designed loss function in Sec. 3.3.

3.1. Formulation of rain and illumination

Rain. A rain image S can be modeled as the addition of a clean background scene B and accumulated rain layer R:

$$S = B + R. \tag{1}$$

Based on this formulation, we can get the clean background scene B by subtracting the rain layer R from the degraded rain image S.

Illumination. The classic Retinex theory [30] assumes that the observed image S can be decomposed into reflectance and illumination:

$$S = F \odot I, \tag{2}$$

where F represents the captured unlit observation, I represents the lightness condition and $\odot$ represents element-wise multiplication. Therefore, We can enhance the lighting by adjusting I.

Rain and illumination. In real-world rainy weather scenarios, rain and low-light may co-occur during outdoor image capture. A low-light rain image can be modeled as:

$$S = F \odot I + R. \tag{3}$$

Following the formulation, we can obtain the enhanced rain-free image O by $O = (S - R) \odot M$, where M represents the illumination map. The illumination map can be easily obtained because it has a relatively simple form with known priors [42].

3.2. Framework architecture

The architecture of our proposed SANet is shown in Fig. 2. Instead of simply cascading rain removal and illumination enhancement, we proposed an entangled structure to remove rain and enhance illumination in a feature-enhanced manner. We designed three U-Net-like [32] networks, RRNet, IENet and DLNet. DLNet is a lightweight network with dense convolution block [14]. Given a degraded image, DLNet produces a single channel mask and is trained using binary cross-entropy loss to match the GT binary mask. To obtain the groundtruth mask, we use the absolute difference between degraded image and restored image, and threshold to obtain a binary mask, classifying pixels into "hard" (value 1) or "easy" (value 0).

The knowledge from intermediate features of DLNet improves RRNet's training. Similarly, the knowledge from intermediate features of RRNet imporves IENet' ability of feature extracting from regions containing both rain streaks and low light degradation. While the mask itself lends adaptiveness to the rain removal and illumination enhancement process. To this end, we propose a novel spatially-adaptive block to achieve progressive feature enhancement. The detail of SABlock is illustrated in Fig. 3 (a).

Two input feature maps are first concatenated together in channel dimension. After the concat operation, the channel weight is obtained by passing feature to one global average pooling layer, two 1×1 convolution layers and one leaky ReLU layer. A simple sigmoid is used to map the output to between 0 and 1. Next, the channel-weighted output is obtained by multiplying the input features and weights. Then a mask guided non-local module (MGN) is followed to enhance feature.

As illustrated in Fig. 3(b), MGN first generates $\mathbf{Q}$, $\mathbf{K}$,$\mathbf{V} \in \mathbb{R}^{C\times H\times W}$ with three 3×3 deformable convolution [4] from the incoming feature $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$, where $H \times W$ denotes the spatial dimension and C is the number of channels. Specially, for $\mathbf{Q}$ and $\mathbf{K}$ the input feature are first multiplied with 1-mask and mask. The meaning of this is to obtain the correspondence between "hard" and "easy" regions, and to use the features of the "easy" regions to repair the pixels of the "hard" regions. Then, $\mathbf{Q}$ and the transposed $\mathbf{K}$ do matrix multiplication to obtain the attentiona map $\mathbf{A}$ of size $\mathbb{R}^{C\times H\times H}$. Next, $\mathbf{A}$ is used to recalibrate $\mathbf{V}$ with matrix multiplication. A deformable convolution is also followed to further enhance the processed feature. Finally, we add the attention-guided $\mathbf{V}$ to the input feature $\mathbf{X}$ to obtain the output of MGN. Overall, the MGN process is defined as:

$$MGN(\mathbf{X}) = DConv(SoftMax(\mathbf{Q}\mathbf{K}^T)\mathbf{V}) + \mathbf{X}, \tag{4}$$

$$\mathbf{Q} = DConv((1 - mask) \cdot \mathbf{X}), \tag{5}$$

$$\mathbf{K} = DConv(mask \cdot \mathbf{X}), \tag{6}$$

$$\mathbf{V} = DConv(\mathbf{X}), \tag{7}$$

where $DConv$ represents 3×3 deformable convolution.

According to Retinex theory [30] it can be known that rain is often contained in the high-frequency information of the image, so we input the high-frequency details obtained through the Laplace transform [19] into the RRNet together with the original image. On the contrary, illumination is often contained in the low-frequency information of the image, so we input the low-frequency global information obtained by one minus Laplace transform into IENet together with the original image. We use bilinear sample instead of convolution and pooling for upsampling and downsampling. This helps reduce unexpected noise and artifacts in the output image.

In order to preserve fine details after rain removal and illumination enhancement, we introduce the DRNet which consists of multiple DRBlocks. The schematic of DRBlock is illustrated in Fig. 3 (b). Since enlarging receptive field can improve remote information acquisition capabilities, dilated convolution [50] is used to extract richer contextual information.

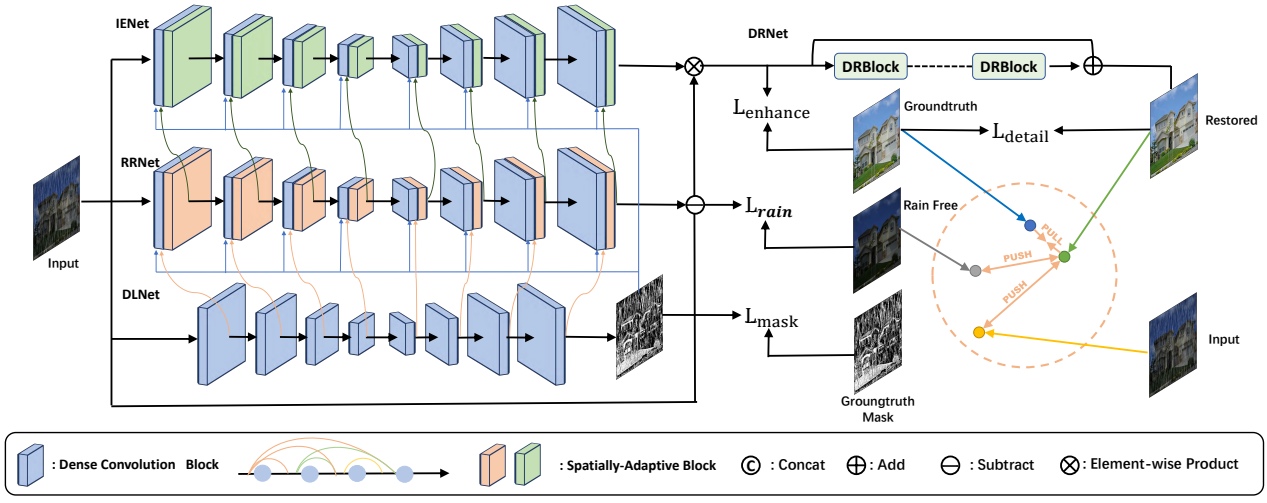

Figure 2: Overview of our proposed SANet. SANet consists of four subnetworks, RRNet, IENet, LNet and DRNet respectively. We jointly minimize the L1 loss and constrative loss to better pull the restored image to the groundtruth and push the restored image to the degraded images.

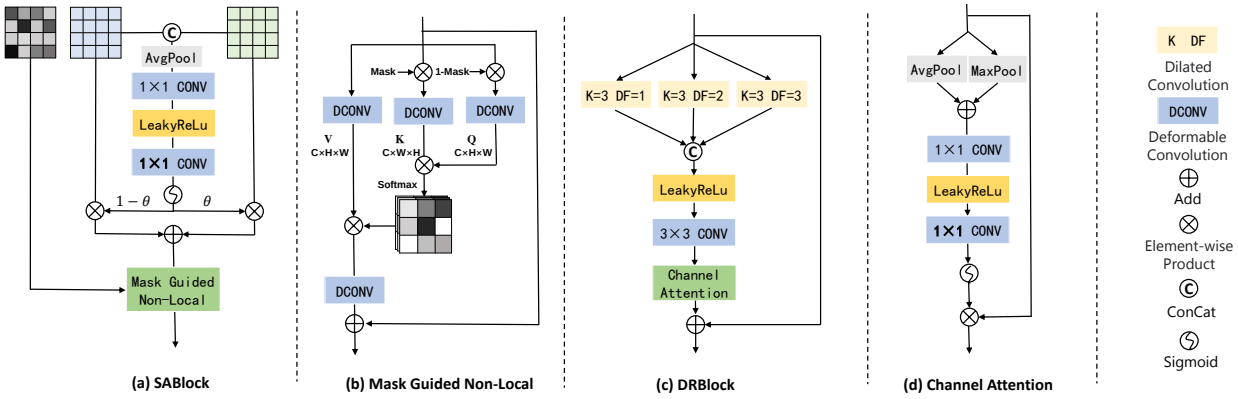

Figure 3: Illustration of SABlock, Mask Guided Non-Local used in SABlock, DRBlock and Channel Attention used in DRBlock.

3.3. Loss function

Inspired by contrastive learning, which aims to learn a representation to pull positive pairs and push apart negative pairs, we propose a new contrastive loss to further improve the quality of restored images. Therefore, there are two aspects to consider here: one is to construct positive and negative sample pairs, the other is to find the representation of these pairs for contrast. In our methods, the positive pair and negative pair is generated by the group of a groundtruth image R and its restored image $\hat{R}$ by the SANet, and the group of $\hat{R}$, low-light rainy image R_{lr} and, low-light rain-free image R_l, respectively. For the representation, we select the output feature from the pre-trained model G, e.g. VGG-16 [36]. Thus, the contrastive loss can be reformulated as:

$$L_c = -\mathbb{E}_X[log\frac{exp(G(R)\cdot G(\hat{R}))}{exp(G(R_{lr})\cdot G(\hat{R})) + exp(G(R_l)\cdot G(\hat{R}))}]. \tag{8}$$

The contrastive loss plays a role of opposing forces pulling the restored image $\hat{R}$ to its groundtruth image R and pushing $\hat{R}$ to its negative samples R^{lr} and R^{l}.

Our SANet consists of four sub-networks, in order to let them perform their duties, we need to design specific constraints for each network. For a input low-light rainy image R_{lr}, P_{dl}, P_{rr}, P_{ie}, R^{dr} are the predictions of DLNet, RRNet, IENet and DRNet respectively. The image after rain removal $Y_r = R_{lr} - P_{rr}$, the enhanced image $Y_e = Y_r \odot P_{ie}$ and the final restored image $\hat{R} = Y_e + P_{dr}$. Therefore, the

overall restoration loss function can be further formulated as:

$$L = L_{mask} + L_{rain} + L_{enhance} + L_{detail} + L_c, \quad (9)$$

$$L_{mask} = \frac{1}{N}\sum_{i=1}^{N} -[M_i \cdot log(P_{dli}) + (1 - M_i) \cdot log(1 - P_{dli})], \quad (10)$$

$$L_{rain} = \| Y_r - R_l \|, \quad (11)$$

$$L_{enhance} = \| Y_e - R \|, \quad (12)$$

$$L_{detail} = \| \hat{R} - R \|, \quad (13)$$

where M represents the groundtruth mask. L_{mask} is used to constrain the binary mask predicted by DLNet, L_{rain} is used to constrain RRNet to get rain-free images, $L_{enhance}$ is used to constrain IENet to get light-enhanced images, and L_{detail} is used to constrain DRNet to get detail-refined images. The above loss can be trained via an optimizer (e.g. Adam) in an end-to-end manner. Different from the perceptual loss uses only positive samples, we also adopt low-light rainy image and rain free low-light image as negatives to constrain the solution space.

4. Experiments

4.1. Implementation details

We employ DRNet that contains 4 DRBlocks and the threshold used to obtain groundtruth degratation mask is set to 180. All experiments are implemented with PyTorch framework on a computer equipped with an NVIDIA RTX 3090 GPU. The input image is randomly cropped to 256 × 256. We train our SANet for 300 epochs using Adam as the optimizer with a batch size of 6. The initial learning rate is set to 0.0001 and decreases uniformly to 0 after 150 epochs.

For the rain removal task, we replace IENet with RRNet to progressively remove the rain and recover details, similarly, for the low-light image enhancement task, we replace RRNet with IENet to progressively enhance the illumination and recover details. All their experimental settings are the same as SANet.

4.2. Dataset and evaluation metrics

DarkRain. In order to train the proposed SANet, we manage to synthesize a new rain image restoration dataset, named DarkRain, including 5075 image pairs with various rain-density and lighting conditions. Our dataset is based on SICE [1], which includes 589 multi-exposure sequences. We first carefully select low-light images in accordance with the real rainfall environment from each sequence. Then we scale the smallest side of the picture to 500 with a fixed aspect ratio. Each low-light image was used to generate 3 rainy images with different streak orientations and magnitudes by following the guidelines mentioned in [7] using Photoshop. Because SICE itself has a corresponding normal light label for each low-light image, so as you can see in Fig. 4, after adding rain, each pair of our dataset contains three pictures, rainy image, rain-free image and corresponding normal-light rain-free image. Finally, we randomly select 400 sequences total of 3432 images for training and the rest for testing.

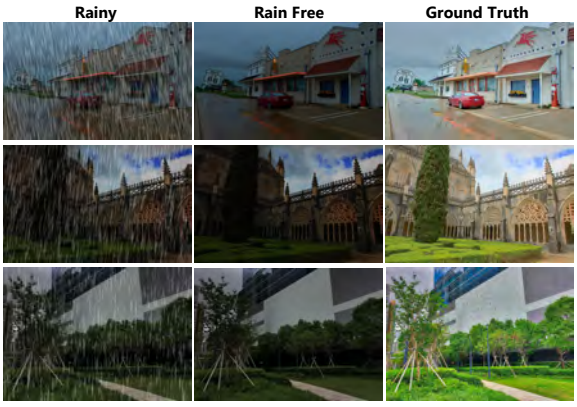

Figure 4: Example images in our dataset. **Left**: Rainy images. **Middle**: Rain-free images. **Right**: Corresponding normal lighting rain-free images.

Rain removal. Rain800 [52] is a widely used synthetic dataset for image deraining, which contains 800 rain/clean image pairs. 700 of them are in the training set and the rest are in the testing set. We also use real-world rain images collected by Wang *et al.* [43] and synthetic rain images collected by Li *et al.* [23] for visual comparison.

Low-light image enhancement. LOL dataset, collected by Wan *et al.* [45], contains 500 low/normal-light image pairs. 485 of them are used for training and 15 for testing.

Evaluation metrics. The following measures are used to evaluate the performance of different methods: Peak Signal to Noise Ratio (PSNR) [15], Structural Similarity Index (SSIM) [44], Feature Similarity Index Measure (FSIM) [53] and Visual Information Fidelity (VIF) [35].

4.3. Comparison results on DarkRain

We conduct extensive experiments on DarkRain dataset to evaluate the performance of the proposed SANet. Because the current methods are individually designed to be used for rain removal or illumination enhancement, and cannot achieve two tasks at the same time, we use the combining method to simulate the benchmark we proposed. To be precise, we selected two SOTA methods on the rain removal task, MPRNet [51] and RCDNet [40], and two SOTA methods on the low-light image enhancement task, KinD [54] and EnlightenGAN [18], to achieve rain removal and illumination enhancement through the two-group sum method (rain removal followed by illumination enhancement or illumination enhancement followed by rain removal). Each pair of DarkRain contains a rainy image, rain-free image, and the corresponding normal-light rain-free image, so these methods can be trained on our dataset. It is important to point out that for a fair comparison we use a fully supervised approach to train EnlightenGAN. We produce a total of eight combinations, Comb I : KinD→RCDNet, Comb II : KinD→MPRNet, Comb III : EnlightenGAN→RCDNet,

Table 1
Quantitative results (PSNR, SSIM, FSIM, VIF) of combining methods and ours on DarkRain dataset. Comb I : KinD [54]→RCDNet [40], Comb II : KinD [54]→MPRNet [51], Comb III : EnlightenGAN [18]→RCDNet [40], Comb IV : EnlightenGAN [18]→MPRNet [51], Comb V : RCDNet [40]→KinD [54], Comb VI : MPRNet [51]→KinD [54], Comb VII : RCDNet [40]→EnlightenGAN [18], Comb VIII : MPRNet [51]→EnlightenGAN [18]. Best and second best scores are highlighted and underlined

Methods	Comb I	Comb II	Comb III	Comb IV	Comb V	Comb VI	Comb VII	Comb VIII	Ours
PSNR	17.69	17.75	17.84	17.99	17.82	17.83	17.96	<u>18.07</u>	**18.36**
SSIM	0.6749	0.6764	0.6779	<u>0.6807</u>	0.6801	0.6776	0.6803	0.6795	**0.6879**
FSIM	0.7998	0.8046	0.8051	0.8072	0.8036	0.8059	0.8087	<u>0.8095</u>	**0.8193**
VIF	0.1961	0.1982	0.1997	0.1998	0.1989	0.2005	0.1990	<u>0.2014</u>	**0.2115**

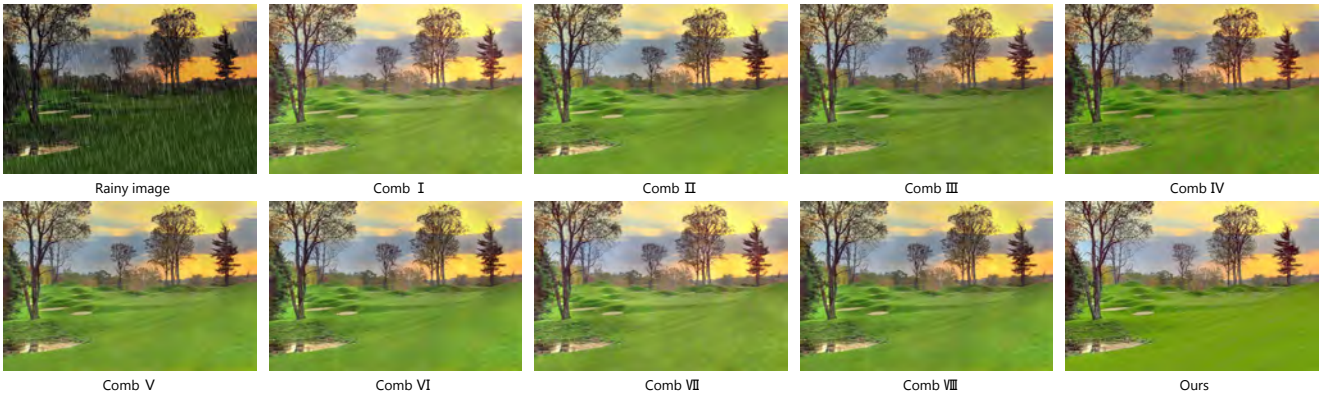

Figure 5: Visual comparison of joint rain removal and illumination enhancement on DarkRain dataset. Comb I : KinD [54]→RCDNet [40], Comb II : KinD [54]→MPRNet [51], Comb III : EnlightenGAN [18]→RCDNet [40], Comb IV : EnlightenGAN [18]→MPRNet [51], Comb V : RCDNet [40]→KinD [54], Comb VI : MPRNet [51]→KinD [54], Comb VII : RCDNet [40]→EnlightenGAN [18], Comb VIII : MPRNet [51]→EnlightenGAN [18].

Comb IV : EnlightenGAN→MPRNet, Comb V : RCDNet→KinD, Comb VI : MPRNet→KinD, Comb VII : RCDNet→EnlightenGAN, Comb VIII : MPRNet→EnlightenGAN. The quantitative comparison results are shown in Table 1. Our SANet obtains the highest values of PSNR, SSIM, FSIM, and VIF on DarkRain dataset and far exceeds other combining methods. Especially PSNR, SANet is 0.29dB higher than the best combining method.

We also visually evaluate their performance in Fig. 5 and Fig. 6. As shown in these results, our SANet can effectively

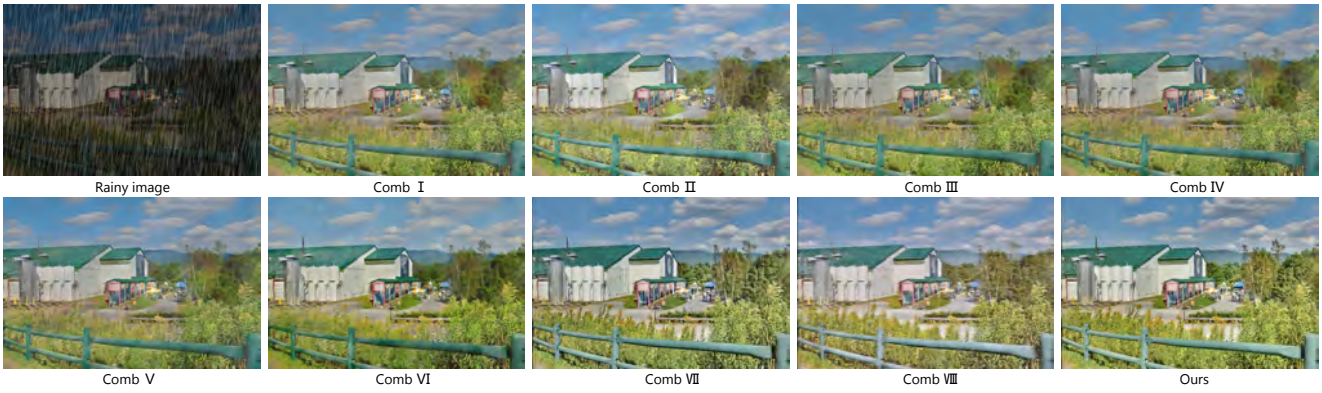

Figure 6: Visual comparison of joint rain removal and illumination enhancement on DarkRain dataset. Comb I : KinD [54]→RCDNet [40], Comb II : KinD [54]→MPRNet [51], Comb III : EnlightenGAN [18]→RCDNet [40], Comb IV : EnlightenGAN [18]→MPRNet [51], Comb V : RCDNet [40]→KinD [54], Comb VI : MPRNet [51]→KinD [54], Comb VII : RCDNet [40]→EnlightenGAN [18], Comb VIII : MPRNet [51]→EnlightenGAN [18].

(a) Rainy image (b) EnlightenGAN⇨MPRNet (c) MPRNet⇨EnlightenGAN (d) Ours (e) Ground Truth

Figure 7: Visual comparison of joint rain removal and illumination enhancement of three examples in DarkRain dataset. (a) Rainy image, (b) EnlightenGAN [18]→MPRNet [51], (c) MPRNet [51]→EnlightenGAN [18], (d) Our SANet, and (e) Ground Truth. The images generated by our SANet are natural, artifact-free and visually closer to the ground-truth.

(a) Rainy image (b) EnlightenGAN⇨MPRNet (c) MPRNet⇨EnlightenGAN (d) Ours (e) Ground Truth

Figure 8: Visual comparison of joint rain removal and illumination enhancement of three examples in DarkRain dataset. (a) Rainy image, (b) EnlightenGAN [18]→MPRNet [51], (c) MPRNet [51]→EnlightenGAN [18], (d) Our SANet, and (e) Ground Truth.

remove the rain and adjust the lighting to the normal range, while the combining methods cause underexposure or overexposure, and leave a lot of rain streaks visible to the naked eye, failing to restore degraded rain image well.

In addition, our SANet can effectively avoid error accumulation process, which can be demonstrated in Fig. 8 and Fig. 7. The color shift caused by the first rain removal will be further amplified when the illumination is enhanced, resulting in more obvious artifacts. The first illumination enhancement will make the stripes more obvious and increase the difficulty of removing the rain and. at the same time, the lack of background information can not restore the light well. By contrast, our method can effectively deal with this problem.

Table 2
Quantitative comparisons with the state-of-the-art methods on Rain800 datasets. Best and second best scores are highlighted and underlined.

Dataset	Rain800		Rain12	
Method	PSNR	SSIM	PSNR	SSIM
DetailsNet [8]	21.16	0.7320	23.06	0.7749
GMM [23]	22.27	0.7413	24.83	0.8121
JORDER [49]	22.24	0.7763	25.12	0.8215
RESCAN [22]	24.09	0.8410	26.74	0.8843
ID-CGAN[52]	24.34	0.8430	-	-
SPANet [43]	22.65	0.8781	26.08	0.9045
DAF-Net [12]	25.27	0.8895	-	-
DRD-Net [5]	26.32	0.9018	27.97	0.9247
MSPFN [17]	26.74	0.9058	28.27	0.9271
RCDNet [40]	26.89	0.9102	28.31	0.9306
MPRNet [51]	27.01	0.9131	28.39	0.9297
Ours	**27.33**	**0.9175**	28.60	0.9381

4.4. Comparison with the State-of-the-Arts

In order to prove the effectiveness and generalization ability of our network, we conducted separate experiments on rain removal and low-light image enhancement tasks. For fair comparisons with state-of-the-art methods, we use the same training and testing sets of existing datasets.

Comparison with rain removal algorithms. We compare our method with other deraining methods on Rain800 dataset, including DetailsNet [8], GMM [23], JORDER [49], RESCAN [22], ID-CGAN[52], SPANet [43], DAF-Net [12], DRD-Net [5], MSPFN [17], RCDNet [40], and MPRNet [51]. Quantitative results are shown in Table 2. One can see that our method achieves the best performance, surpassing MPRNet by 0.32dB and 0.21dB on Rain800 and Rain12, respectively, in terms of PSNR.

Meanwhile, we proved the generalization ability of our method in real-world rain image restoration and other rain removal dataset. Visual comparison of our deraining method and SANet with other state-of-the-art deraining methods are presented in Fig. 9 and Fig. 10. Our deraining method exhibits impressive restoration performance on all scenarios, generating rich and realistic image textures while removing rain streaks. For other deraining methods, they still leave some visible rain streaks. In addition, our SANet can not only effectively remove the rain in the picture, but also adjust the brightness to obtain a more satisfactory visual effect, and more comprehensively restore the degraded image captured in the rainy weather.

Comparison with low-light image enhancement algorithms. In Table 3, we compare our method on LOL dataset with several existing low-light image enhancement methods quantitatively including LIME [11], Retinex-Net [45], Zero-DCE [9], Zero-DCE++ [20], RUAS [26], KinD [54] and EnlightenGAN [18]. Obviously, our method obtains a competitive PSNR and SSIM scores and outperforms the state-of-the-art method EnlightenGAN [18] by 0.55dB on PSNR. We provide additional visual comparison of our method with the state-of-the-art. Results on the LOL dataset [45] are shown in Fig. 11. Intuitively, our results are more realistic, with more detail recovered.

Table 3
Quantitative comparisons with the state-of-the-art methods on LOL datasets. Best and second best scores are highlighted and underlined, where * represents training in a supervised manner.

Method	PSNR	SSIM
LIME [11]	16.76	0.5644
Retinex-Net[45]	16.77	0.5594
Zero-DCE [9]	15.47	0.6343
Zero-DCE++ [20]	16.69	0.6325
RUAS[26]	16.39	0.6518
KinD [54]	20.87	0.8022
EnlightenGAN [18]	18.86	0.8122
EnlightenGAN* [18]	21.04	0.8462
Ours	**21 .69**	**0.8663**

Table 4
Ablation study of the threshold, where T represent the value of threshold.

T	120	140	160	180	200	220
PSNR	18.09	18.17	18.25	18.36	18.31	18.20

Table 5
Effect of SABlock on our proposed benchmark. Among all variants, using guidance to both main line feature and secondary line feature achieves better performance.

Method	PSNR	SSIM
$SABlock_{w/oM}$	18.11	0.6814
$SABlock_{wML}$	18.27	0.6834
$SABlock_{wSL}$	18.23	0.6820
Ours	18.36	0.6879

4.5. Ablation study

Threshold to get binary mask.

We evaluate the proposed spatially-adaptive method with different thresholds on our proposed data set. The PSNR results are shown in Table 4. The highest PSNR is obtained when T = 180, and when T decreases or increases, both bring about a decrease in performance. This is due to the fact that when T is too small, the vast majority of the image is judged as hard-to-repair regions and we lack enough features to repair degraded regions, accordingly, when T is too large some hard-to-repair regions are judged as easy-to-repair regions and their suboptimal features are passed to the repair regions, leading to a drop in performance.

Effect of SABlock. To demonstrate the effectiveness of our SABlock, we compare our method with 3 variants which vary on whether the degradation mask-based guidance used

(a) Rainy image (b) SPANet (c) DRD-Net (d) MPRNet (e) Ours (Deraining) (f) Ours (SANet)

Figure 9: Visual comparison of our methods with the state-of-the-art deraining methods on real world rain images. (a) Rainy image, (b) SPANet [43], (c) DRD-Net [5], (d) MPRNet [51], (e) Our deraining method, and (f) Our SANet. For better observation, we improved the contrast of the red box enlarged area. Zooming-in the figure will provide a better look at the restoration.

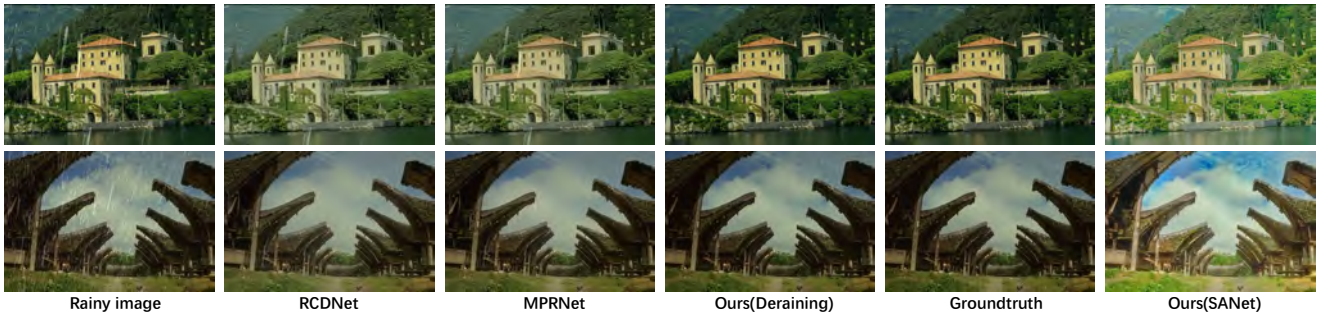

Figure 10: Visual comparison of our methods with the state-of-the-art deraining methods on Rain12 datasets. Zooming-in the figure will provide a better look at the restoration.

and how to use it: (i) $SABlock_{w/oM}$ denotes without exploiting the guidance. (ii) $SABlock_{wML}$ denotes using the guidance only to main line feature. (iii) $SABlock_{wSL}$ denotes using the guidance only to secondary line feature. (iv) *Ours* denotes using the guidance to both main line feature and secondary line feature.

Table 5 lists the quantitative comparison among the variants. From the table, directly incorporating the guidance into the main line feature or secondary line feature improves the performance, indicating that the degradation mask-based guidance is beneficial to image restoration. With the guidance to both main line feature and secondary line feature, our method achieves the best performance, clearly demonstrating the effectiveness of our SABlock.

5. Discussion

To demonstrate that our proposed can benefit vision based applications, we employ Google Vision API to evaluate the restoration results, as shown in Fig. 12. Compared to the rain free image, the number of detections in our result is more, indicating that SANet allows for more adequate restoration of degraded image and effectively improve the performance of detection.

Table 6
The time complexity (in seconds) of different methods with size 256×256 on GPU.

GMM [23]	JORDER [49]	RESCAN [22]
14.582	0.031	0.011
DRD-Net [5]	RCDNet [40]	Ours
0.023	0.023	0.026

5.1. Running time

Table 6 reports the running time of our SANet and other deraining methods. From the results, it can be seen that SANet is only a little slower than the current state-of-the-art rain removal algorithms such as DRD-Net [5] and RCDNet [40]. Notably, Compared with other deraining

Figure 11: Low-light image enhancement results on LOL [45] dataset

Figure 12: The restoration results are tested on the Google Vision API. Left: object recognition result in the rain image. Middle: object recognition result in the rain free image. Right: object recognition result in the image restored using our method.

methods, SANet not only remove rain well but also adjust the lighting to get more profound visual impression.

6. Conclusion

In this paper, we propose a novel spatially-adaptive network, namely SANet, to remove rain and enhance illumination in a unified framework. Since the image contains two different kinds of degradation and the severity of degradation varies in different regions, our proposed method uses features from the less degraded regions to repair the pixels in the severely degraded regions under the guidance of an adaptive degradation mask. Meanwhile, to use negative samples to improve the quality of the restored image we propose contrastive loss to ensure that the restored image is pulled to closer to the groundtruth image and pushed to far away from the degraded images in the representation space. Moreover, we present a new synthetic dataset DarkRain to better restore real-world rain images. Extensive experiments demonstrate that SANet is very effective for rain image restoration, and has a robust generalization ability. In addition, the results also show great superiority of our proposed algorithm over other top-performing deraining and low-light image enhancement methods. We believe our proposed benchmark would significantly advance the research of real-world rain image restoration task in the future.

Acknowledgments

This work was supported by the grants from the National Natural Science Foundation of China (Nos.61673396), the Spanish Ministry of Economy and Competitiveness (MINECO) and the European Regional Development Fund (ERDF) under Project PID2020-120311RB-I00/AEI/10.13039/501100011033.

References

[1] Cai, J., Gu, S., Zhang, L., 2018. Learning a deep single image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing 27, 2049–2062.

[2] Chen, C., Xiong, Z., Tian, X., Zha, Z.J., Wu, F., 2019. Real-world image denoising with deep boosting. IEEE transactions on pattern analysis and machine intelligence 42, 3071–3087.

[3] Chen, J., Tan, C.H., Hou, J., Chau, L.P., Li, H., 2018. Robust video content alignment and compensation for rain removal in a cnn framework, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6286–6295.

[4] Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y., 2017. Deformable convolutional networks, in: Proceedings of the IEEE international conference on computer vision, pp. 764–773.

[5] Deng, S., Wei, M., Wang, J., Feng, Y., Liang, L., Xie, H., Wang, F.L., Wang, M., 2020. Detail-recovery image deraining via context aggregation networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 14560–14569.

[6] Eigen, D., Krishnan, D., Fergus, R., 2013. Restoring an image taken through a window covered with dirt or rain, in: Proceedings of the IEEE international conference on computer vision, pp. 633–640.

[7] Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J., 2017a. Clearing the skies: A deep network architecture for single-image rain removal. IEEE Transactions on Image Processing 26, 2944–2956.

[8] Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J., 2017b. Removing rain from single images via a deep detail network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3855–3863.

[9] Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Runmin, C., 2020a. Zero-reference deep curve estimation for low-light image enhancement. CVPR .

[10] Guo, Q., Sun, J., Juefei-Xu, F., Ma, L., Xie, X., Feng, W., Liu, Y., 2020b. Efficientderain: Learning pixel-wise dilation filtering for high-efficiency single-image deraining. arXiv preprint arXiv:2009.09238 .

[11] Guo, X., Li, Y., Ling, H., 2016. Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on image processing 26, 982–993.

[12] Hu, X., Fu, C.W., Zhu, L., Heng, P.A., 2019. Depth-attentional features for single-image rain removal, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8022–8031.

[13] Hu, X., Zhu, L., Fu, C.W., Qin, J., Heng, P.A., 2018. Direction-aware spatial context features for shadow detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7454–7462.

[14] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.

[15] Huynh-Thu, Q., Ghanbari, M., 2008. Scope of validity of psnr in image/video quality assessment. Electronics letters 44, 800–801.

[16] Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., Ma, J., Jiang, J., 2020a. Multi-scale progressive fusion network for single image deraining, in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[17] Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., Ma, J., Jiang, J., 2020b. Multi-scale progressive fusion network for single image deraining, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 8346–8355.

[18] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z., 2021. Enlightengan: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing 30, 2340–2349.

[19] Kamgar-Parsi, B., Rosenfeld, A., 1999. Optimally isotropic laplacian operator. IEEE Transactions on Image Processing 8, 1467–1472.

[20] Li, C., Guo, C.G., Loy, C.C., 2021. Learning to enhance low-light image via zero-reference deep curve estimation, in: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[21] Li, R., Cheong, L.F., Tan, R.T., 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1633–1642.

[22] Li, X., Wu, J., Lin, Z., Liu, H., Zha, H., 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining, in: European Conference on Computer Vision, Springer. pp. 262–277.

[23] Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S., 2016. Rain streak removal using layer priors, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2736–2744.

[24] Liu, J., Yang, W., Yang, S., Guo, Z., 2018a. D3r-net: Dynamic routing residue recurrent network for video rain removal. IEEE Transactions on Image Processing 28, 699–712.

[25] Liu, J., Yang, W., Yang, S., Guo, Z., 2018b. Erase or fill? deep joint recurrent rain removal and reconstruction in videos, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3233–3242.

[26] Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z., 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10561–10570.

[27] Luo, Y., Xu, Y., Ji, H., 2015. Removing rain from a single image via discriminative sparse coding, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 3397–3405.

[28] Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J., 2018. Attentive generative adversarial network for raindrop removal from a single image, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2482–2491.

[29] Quan, R., Yu, X., Liang, Y., Yang, Y., 2021. Removing raindrops and rain streaks in one go, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9147–9156.

[30] Rahman, Z.u., Jobson, D.J., Woodell, G.A., 2004. Retinex processing for automatic image enhancement. Journal of Electronic imaging 13, 100–110.

[31] Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D., 2019. Progressive image deraining networks: A better and simpler baseline, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3937–3946.

[32] Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

[33] Shao, M., Li, L., Wang, H., Meng, D., 2021a. Selective generative adversarial network for raindrop removal from a single image. Neurocomputing 426, 265–273.

[34] Shao, M.W., Li, L., Meng, D.Y., Zuo, W.M., 2021b. Uncertainty guided multi-scale attention network for raindrop removal from a single image. IEEE Transactions on Image Processing 30, 4828–4839.

[35] Sheikh, H.R., Bovik, A.C., 2006. Image information and visual quality. IEEE Transactions on image processing 15, 430–444.

[36] Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

[37] Sun, G., Shao, H., Cattani, C., 2022. A priori-guided multi-layer rain-aware network for single image deraining. Knowledge-Based Systems 235, 107613.

[38] Wang, C., Zhu, H., Fan, W., Wu, X.M., Chen, J., 2022. Single image rain removal using recurrent scale-guide networks. Neurocomputing 467, 242–255.

[39] Wang, H., Wu, Y., Xie, Q., Zhao, Q., Liang, Y., Zhang, S., Meng, D., 2021. Structural residual learning for single image rain removal. Knowledge-Based Systems 213, 106595.

[40] Wang, H., Xie, Q., Zhao, Q., Meng, D., 2020a. A model-driven deep neural network for single image rain removal, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3103–3112.

[41] Wang, L.W., Liu, Z.S., Siu, W.C., Lun, D.P., 2020b. Lightening network for low-light image enhancement. IEEE Transactions on Image Processing 29, 7984–7996.

[42] Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J., 2019a. Underexposed photo enhancement using deep illumination estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6849–6857.

[43] Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W., 2019b. Spatial attentive single-image deraining with a high quality real rain dataset, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[44] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13, 600–612.

[45] Wei, C., Wang, W., Yang, W., Liu, J., 2018. Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560 .

[46] Wei, W., Meng, D., Zhao, Q., Xu, Z., Wu, Y., 2019. Semi-supervised transfer learning for image rain removal, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3877–3886.

[47] Xu, K., Yang, X., Yin, B., Lau, R.W., 2020. Learning to restore low-light images via decomposition-and-enhancement, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2281–2290.

[48] Yang, W., Tan, R.T., Feng, J., Guo, Z., Yan, S., Liu, J., 2019. Joint rain detection and removal from a single image with contextualized deep networks. IEEE transactions on pattern analysis and machine intelligence 42, 1377–1393.

[49] Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S., 2017. Deep joint rain detection and removal from a single image, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1357–1366.

[50] Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 .

[51] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L., 2021. Multi-stage progressive image restoration, in: CVPR.

[52] Zhang, H., Sindagi, V., Patel, V.M., 2019a. Image de-raining using a conditional generative adversarial network. IEEE transactions on circuits and systems for video technology 30, 3943–3956.

[53] Zhang, L., Zhang, L., Mou, X., Zhang, D., 2011. Fsim: A feature similarity index for image quality assessment. IEEE transactions on Image Processing 20, 2378–2386.

[54] Zhang, Y., Zhang, J., Guo, X., 2019b. Kindling the darkness: A practical low-light image enhancer, in: Proceedings of the 27th ACM international conference on multimedia, pp. 1632–1640.

[55] Zhang, Y., Zhang, J., Huang, B., Fang, Z., 2021. Single-image deraining via a recurrent memory unit network. Knowledge-Based Systems 218, 106832.

[56] Zhu, L., Fu, C.W., Lischinski, D., Heng, P.A., 2017. Joint bi-layer optimization for single-image rain streak removal, in: Proceedings of the IEEE international conference on computer vision, pp. 2526–2534.