

Article

Multi-Scale Aggregation Residual Channel Attention Fusion Network for Single Image Deraining

Jyun-Guo Wang ^{1,*} and Cheng-Shiuan Wu ²¹ The Department of Medical Informatics, Tzu Chi University, Hualien County 97004, Taiwan² 701 Zhongyang Rd., Sec. 3, Hualien County 97004, Taiwan

* Correspondence: jgwang@gms.tcu.edu.tw

Abstract: Images captured on rainy days are prone to rain streaking on various scales. These images taken on a rainy day will be disturbed by rain streaks of varying degrees, resulting in degradation of image quality. This study sought to eliminate rain streaks from images using a two-stage network architecture involving progressive multi-scale recovery and aggregation. The proposed multi-scale aggregation residual channel attention fusion network (MARCAFNet) uses kernels of various scales to recover details at various levels of granularity to enhance the robustness of the model to streaks of various sizes, densities, and shapes. When applied to benchmark datasets, the proposed method outperformed other state-of-the-art schemes in the restoration of image details without distorting the image structure.

Keywords: multi-scale aggregation residual channel attention fusion network (MARCAFNet); derain



Citation: Wang, J.-G.; Wu, C.-S. Multi-Scale Aggregation Residual Channel Attention Fusion Network for Single Image Deraining. *Appl. Sci.* **2023**, *13*, 2709. <https://doi.org/10.3390/app13042709>

Academic Editor: Antonio Fernández-Caballero

Received: 5 January 2023

Revised: 10 February 2023

Accepted: 17 February 2023

Published: 20 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In outdoor photography, the quality of images is often susceptible to the influence of different weather conditions. Rainy weather is common and can influence high-level tasks in computer vision, such as object detection [1] and the operation of intelligent vehicles [2]. To improve performance in these tasks, algorithms for removing artifacts and restoring background information in distorted images must be developed. However, developing such algorithms is challenging because of the uncertainty of the background information in distorted images.

Over the past few decades, several studies have attempted to address the aforementioned problems, and an extensive range of techniques have been proposed for rain removal in images. For example, rain removal from a single image can be realized using various wave filters such as bilateral filters [3] and guided filters [4] to decompose the rainy image into high-frequency and low-frequency parts and then recover a rain-free image by combining the feature selection of the low-frequency and high-frequency parts. Recently, various methods have been proposed for rain streak removal through theoretical analyses of the physical properties of the rain and background layers. The representative methods that involve this approach include those using layer priors based on the Gaussian mixture model (GMM) [5], discriminative sparse coding (DSC) [6], joint convolutional analysis (JCAS), and synthesis sparse representation (SSR) [7]. However, images with complex rain shapes and background scenes cannot be flexibly processed using techniques based on prior analytical models.

The recent success of deep learning methods in computer vision has encouraged researchers to explore approaches to rain removal in images. Instead of manually inputting information regarding the properties and features of rain-affected images, deep-learning-based methods leverage sophisticated model architectures, data-rich training samples, and non-linear representation modeling. The representative approaches have been RESCAN [7], PReNet [8], LPNet [9], MSPFN [10], MPRNet [11], VRGNet [12], PCNet [13], and RAiA-Net [14]. Deep-learning-based methods are more robust and have quick development

cycles compared with traditional-algorithm-based methods [3–6]. However, deep learning models are hampered by the vanishing and exploding gradient problem because of their deep architecture. A few weaknesses and areas of improvement based on recent studies are listed as follows: the receptive field size of a convolutional kernel is essential for modeling visual patterns. However, using a sizeable convolutional kernel throughout the model is not feasible and incurs a large computational and memory expense. A few studies on computer vision tasks [15,16] have demonstrated that multi-scale feature encoding could effectively help extract features on different scales. After our analysis, none of these methods [8–14] used multi-scale kernels for rain removal, resulting in poor rain removal performance. Furthermore, the loss function in several methods [7,13] is based on pixel-level differences; thus, the trained models are not focused on structural representation.

The overall objective of this dissertation is to develop a novel deep learning method to solve image deraining. The organization of each section in this dissertation are as follows. In Section 2, the related work of rain removal is briefly introduced, including on video deraining methods and single image deraining methods. A multi-scale aggregation residual channel attention fusion network (MARCAFNNet) is proposed in Section 3. The experimental results are demonstrated in Section 4. Finally, brief concluding remarks and prospects for future works are presented in Section 5.

To address the aforementioned problems, this thesis proposes a deep learning architecture that accounts for the different scales of rain streaks and aids the recovery of the affected images to obtain images with satisfactory quality. The proposed model involves the recovery of the input image using three consecutive coding structure blocks (CS blocks) with different dilation rates. Thereafter, the feature maps of the submodules are blended, and the recovered image is generated through the scale blend block (SB block) and simple mathematical operations.

The contributions of this study are as follows:

1. A U-Net-based [17] encoder–decoder submodule called a CS block is proposed. Unlike a typical U-Net, the feature maps of the encoding steps are passed through a densely connected network [18] before being merged in the corresponding decoding steps. Three CS blocks are arranged consecutively in ascending order of dilation rate.
2. To effectively utilize the feature maps of the CS block, a cross-stage feature fusion block (CSFF block) is proposed for providing information to subsequent CS blocks.
3. A scale-blending operation is proposed and implemented in the model.
4. To balance the tradeoff between pixel-level accuracy and structural accuracy, the model is trained using the mean square error (MSE) loss and structural similarity (SSIM) loss [19] functions.

2. Related Work

This section examines the relevant studies on deraining. The existing studies can be divided into two categories of video deraining and single image deraining.

2.1. Video Deraining Methods

Garg [20] proposed a photometric model to detect candidate pixels affected by rain in each frame of a video feed by modeling the intensity of raindrops among frames. The pixels affected by rain were detected and removed by updating the intensity to the mean of the sum with the subsequent frame. However, Garg’s method is unable to detect rain streaks that are severely defocused or against a bright background. Zhang [21] analyzed rain streaks in video images in terms of temporal characteristics (i.e., streaks did not affect image pixels throughout the entire video sequence) and chromatic characteristics (i.e., changes in the R, G, and B channels were roughly the same). Neither of these methods perform well under heavy rain conditions. Chen [22] focused on the fact that most rain streaks share similar repeating patterns (e.g., similar direction). Kim [23] proposed a novel algorithm that creates a map of rain streaks via temporal correlation for classification using a support vector machine and removal using the low-rank matrix completion technique.

Li et al. [24] used the prior structure of rain streaks in conjunction with a multi-scale convolutional sparse coding model to eliminate streaks based on local patterns and multi-scale features. Liu et al. [25] proposed a hybrid model using a recurrent network and reconstruction network for binary classification, deraining, and background restoration. Using a dynamic rain generator, Yue et al. [26] sought to fit rain layers via statistical modeling with the aim of encoding the physical (spatial) structure and (temporal) variations in rain streaks to facilitate the removal of streaks using a semi-supervised modular approach. Although these methods are effective at removing rain from the video, they depend on the analysis of temporal variations, which makes the model unable to solve the problem of removing rain from a single image. For the solution of the single image derain issues, some experts [3–6,27,28] have begun to consider removing rain streaks from a single image.

2.2. Single Image Deraining Methods

Image processing and machine learning were the early methods used for the removal of rain from a single image [3–6,27,28]. Kang et al. [3] used bilateral filters to decompose high and low frequencies, and then decomposed the high-frequency parts into rain and non-rain components. Finally, they used a sparse coding dictionary to learn and recover the rain-affected areas. Huang et al. [27] proposed an unsupervised model using a priori knowledge of images to distinguish rain streaks from other high-frequency details. Luo et al. [6] employed discriminative encoding in learning a dictionary for use in separating the background and rain streaks using the coefficient vector sparsity of rain streaks. Li et al. [5] explored the use of hybrid model GMMs, including GMMs of the background layer learned from natural images. The scale and directions of rain streaks are learned from images to enable removal from the background layer and rain layer. Wang et al. [4] used guided filters to decompose input images into high-frequency and low-frequency parts to facilitate the extraction of image details from high-frequency components and subsequent rain removal. Zhu et al. [28] decomposed a single input image into a rain-free background layer and a rain streak layer. They then analyzed the local gradient of the rain streak layer as prior knowledge for rain streak removal. These methods proved effective in the rain streak removal and for reducing the memory requirements for deraining. However, with the development of technology, image resolution quality is a growing concern and further advances are required to enhance image quality.

In recent years, with the rapid development of deep learning, a number of researchers have reported that deep learning methods outperform image processing and machine learning methods [7–14,29–36]. Eigen et al. [29] proposed a deep-learning-based correlation module that used photos taken through a window as training data to guide the mapping of noisy image patches against clean image patches for the removal of static rain streaks and dirt artifacts. Fu et al. [30,31] used a convolution neural network to map the relationship between rain streak images and rain-free images for rain removal via residual learning and image parsing. Li et al. [7] combined a convolutional network and a recurrent neural network to enhance rain removal performance. The method obtained image features using an extended convolutional network and retained useful information from previous stages via a recurrent neural network to further improve rain removal performance. Hu et al. [32] analyzed the visual effects of rain on images according to scene depth and then used attention mechanisms to learn deep features for use in rain streak removal in an end-to-end neural network. Wang et al. [33] built a large dataset of rain/no rain pairs and developed a semi-supervised model to derive local-to-global spatial attention features for rain streak removal. Ren et al. [8] proposed the ResNet model that removed rain streaks from pictures by reusing shallow ResNet modules in conjunction with feature stacking between modules of different phases for rain removal. Fu et al. [9] proposed a lightweight pyramid module (LPNet) using the Gauss-Laplacian decomposition technique, which greatly reduced the parameter problem in deraining images. Jiang et al. [10] proposed an end-to-end multi-scale progressive fusion network that used recurrent computation to capture global details, explore information in spatial dimensions, and cooperate with attention mechanisms to

facilitate image fusion on various scales. Zhang et al. [34] proposed a conditional generative adversarial network, which produced global and local image information through the generator and used the discriminator to identify true and false. Chen et al. [35] proposed the half-instance normalization block with two subnetworks for training to improve image recovery. Wang et al. [36] proposed an effective pyramid feature decoupling network (i.e., PFDN) for single image deraining, which could accomplish image deraining and detail recovery with the corresponding features. Zamir et al. [11] proposed a multi-stage encoder-decoder network architecture that used supervised attention to pre-weight local features and lateral connections with the aim of preventing information loss. Wang et al. [12] built a full Bayesian generative model for rainy images where the rain layer is parameterized as a generator with the input being some latent variables representing the physical structural rain factors. To solve this model, they employed the variational inference framework to approximate the expected statistical distribution of a rainy image in a data-driven manner. Jiang et al. [13] proposed a progressive coupled network (PCNet) to well separate rain streaks while preserving rain-free details. To this end, they investigated the blending correlations between them and particularly devised a novel coupled representation module (CRM) to learn the joint features and the blending correlations. Yin et al. [14] proposed a multi-stage deep neural network for image deraining, which was equipped with the refined attention in attention (RAiA) module. Specifically, RAiA extends the original attention in attention by exploiting the joint dependencies of channel and spatial information to generate the dynamic allocated weights. Although the aforementioned methods solve some problems, the problems of blurring and smoothing of image details resulting in an inability to distinguish between rain streaks and texture features remain unsolved in the images. In short, the details of texture are sometimes blurry and not clearly visible in the images. The effect of removing rain streaks is not good. To develop a robust streak removal and improved detail blurring method, the present study proposes a multi-scale aggregation residual channel attention fusion network for single image deraining.

3. Multi-Scale Aggregation Residual Channel Attention Fusion Network

This section introduces the overall architecture of MARCAFNNet and describes each submodule in detail.

3.1. Architecture of Model

Figure 1 presents an overview of the proposed MARCAFNNet, which uses a convolutional layer to extract features followed by three consecutive coding structure (CS) blocks of different dilation sizes [37] to recover the input image. The three blocks are aligned in ascending order according to their dilation rate. Features of various scales are captured by adopting dilation rates of 2 and 3. When a kernel of size 3×3 is coupled with a dilation rate of 2, the block obtains a receptive field of 5×5 with the same parameter size as 3×3 . By including a dilated convolutional layer, it is possible to obtain a larger receptive field using fewer resources. By adjusting the size of the receptive field among the three CS blocks, it is possible to learn and recover details on different scales. The recovered feature map of each block is annotated as F_s (dilation rate = 1), F_M (dilation rate = 2), and F_L (dilation rate = 3), where F_s obtains the micro-features because of the smaller dilation rate, F_M obtains the structural feature of the image because of the larger dilation rate, and F_L obtains the rain streak feature of the largest area of the image because the dilation rate is the largest. To increase the reusability of feature maps and prevent the vanishing of gradients within individual CS blocks, each encoding-decoding step is connected using densely connected convolutional layers. Features collected within each block are shared with the subsequent blocks through the CSFF block. Thereby, the next CS block can obtain the features of the previous CS block to facilitate learning via feature fusion. We use a combination of Conv2d-BN with LeakyReLU [38] for modeling using a residual channel attention block, which is a modified version of the Squeeze-and-Excitation (SE) block [39]. F_s , F_M , and F_L are subsequently fused via channel concatenation and to the SB block, which

is responsible for modeling the correlation between F_s , F_M , and F_L . The output is split into three chunks and annotated individually as A_s , A_M , and A_L . A_s , A_M , and A_L record the output from F_s , F_M , and F_L , respectively. These blended feature maps are then multiplied with their corresponding features to generate the outputs $\tilde{F}_s$, $\tilde{F}_M$, and $\tilde{F}_L$. Finally, they are fused via channel-wise summation and the result is activated using the sigmoid function to obtain the recovered image $\tilde{B}$.

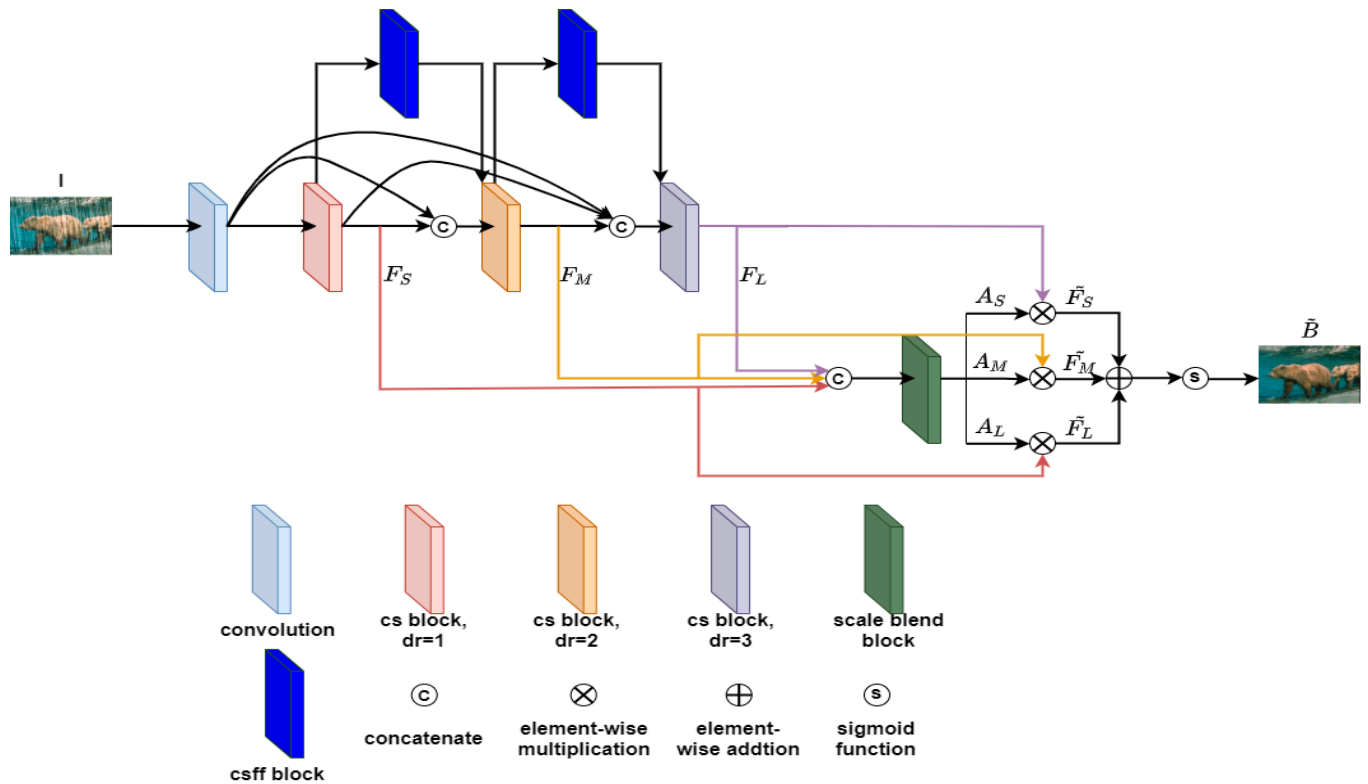


Figure 1. Outline of proposed MARCAFNet.

The proposed MARCAFNet was tested using two images with heavy and light rain streaking, the results of which are shown in Figure 2. In both scenarios, three sets of results are obtained based on a 3×3 convolution with various dilation rates, as follows: F_s (dilation rate = 1), F_M (dilation rate = 2), and F_L (dilation rate = 3). A receptive field of 3×3 allowed for the capture of details on the micro-scale. A receptive field of 5×5 allowed for the capture of structural features. A receptive field of 7×7 focused the capture of information related to features in larger areas, such as those associated with rain streaks.

3.1.1. Coding Structure Block

First, a modified module was developed in accordance with the encoder–decoder architecture of U-Net in which the encoder and decoder were scaled symmetrically. There are three similar structures, and the encoded rain background feature maps F_s , F_M , and F_L with different scales of information are obtained using dilation convolution. In the following, CS blocks are obtained using a dilation rate of 2 to illustrate the architecture, as shown in Figure 3. In the encoder, a convolution layer is first used to convert the number of channels of the input image and increase the non-linear adjustment of the features. The features obtained at each scale are first multiplied by image features from the previous submodule, whereupon a residual channel attention (RCA) block is used to enhance the channel features, before implementing a maximum pooling layer and reducing the image scale by half. The maximum pooling layer sets the sampling rate to 2. In the decoder, up-sampling is applied from the bottom to the top, such that the length and width of the picture are doubled. This module is meant to enhance the representation of image content

in order to obtain detailed information related to features in the rain-streaked image, based on the concept of residual networks [40], SE block, and combined shallow and deep features in dense connection.

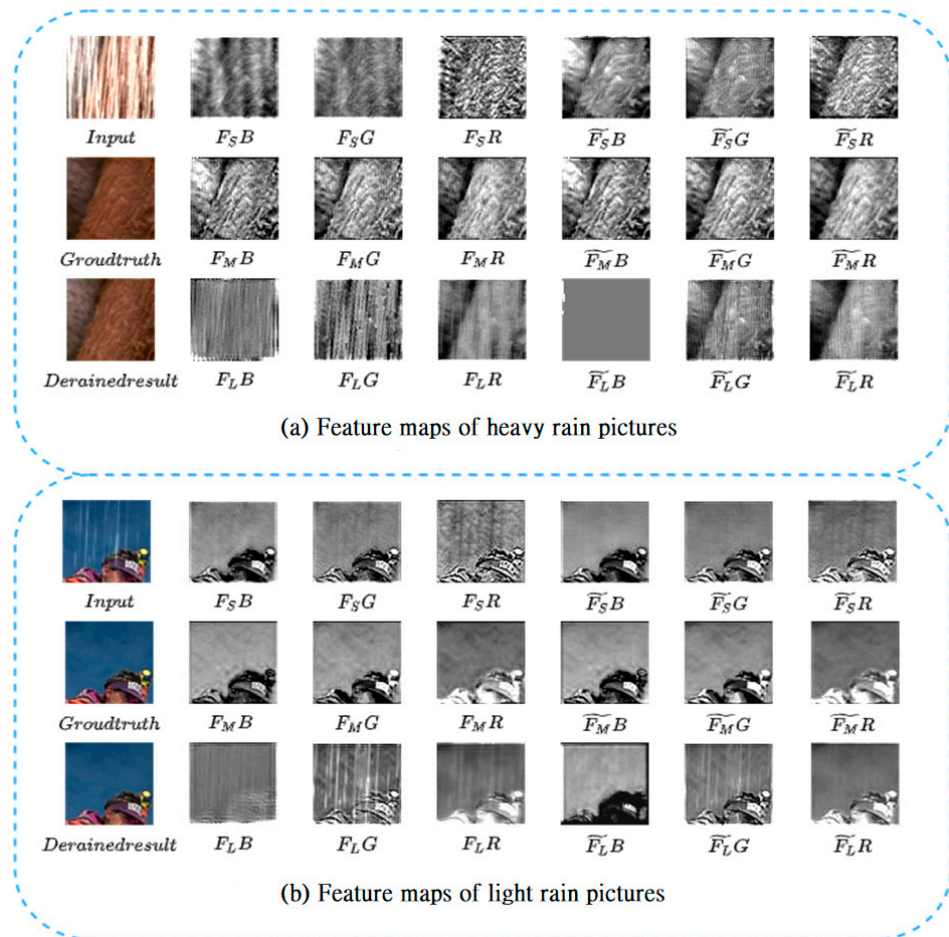


Figure 2. Visualization of the components in our proposed MARCAFNet.

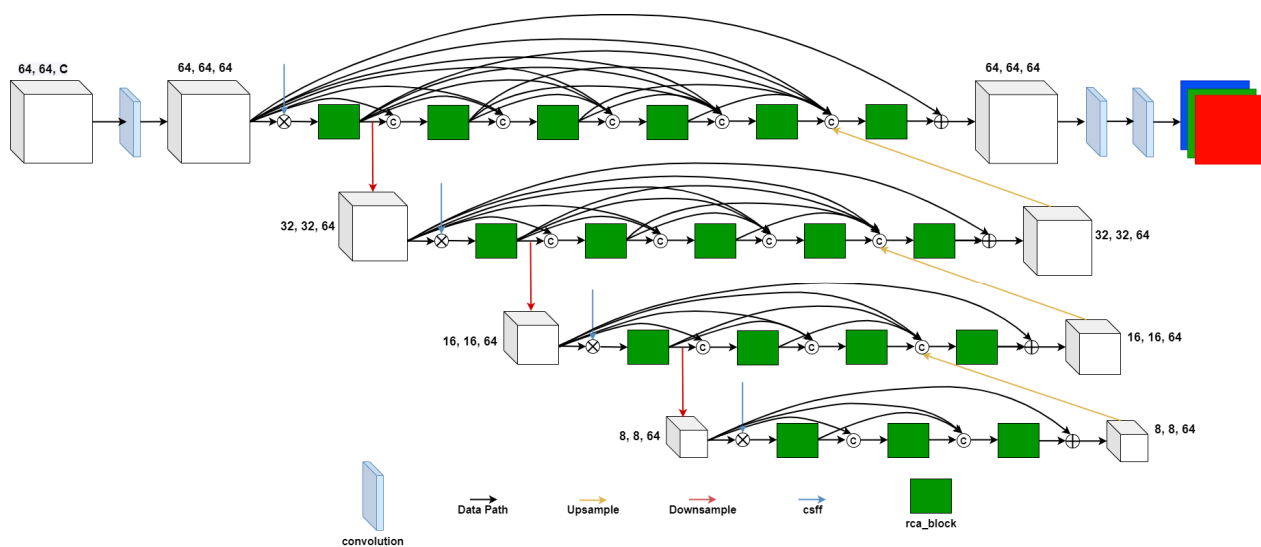


Figure 3. Architecture of proposed CS block.

3.1.2. Residual Channel Attention Block

In the module in Figure 4, the strategy mechanism proposed an RCA block to improve the performance of channel attention. The RCA block is a submodule that was developed by combining an SE module and the ResNet module. The RCA block adds features after the second convolutional layer to alleviate the problem of gradient descent and thereby obtain additional features. The feature map obtained from the second convolutional layer is then used to obtain the global features of each channel using the global average pooling layer based on the concept of SE.

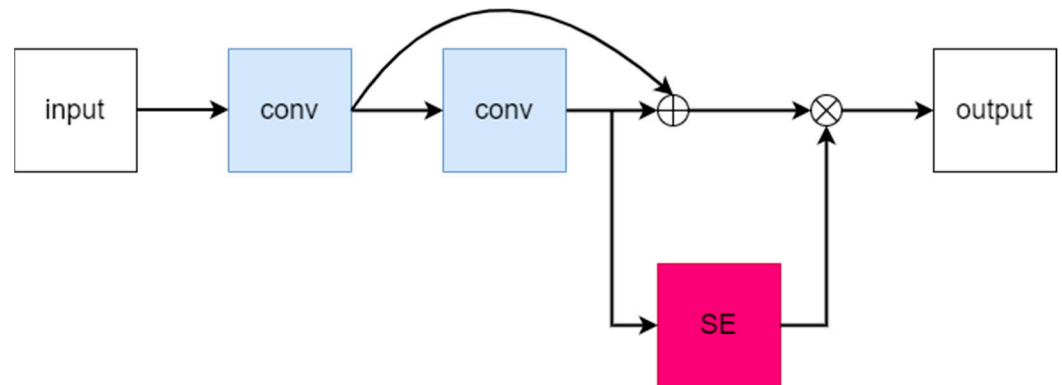


Figure 4. Architecture of proposed RCA block.

3.1.3. Cross Stage Feature Fusion Block

The module in Figure 5 performs channel-wise concatenation between each encoding–decoding pair in a CS block, and then multiplies its result through the RCA-block-enhanced channel attention with the result of the corresponding step in the subsequent CS block. Essentially, this module is used for cross-stage feature sharing.

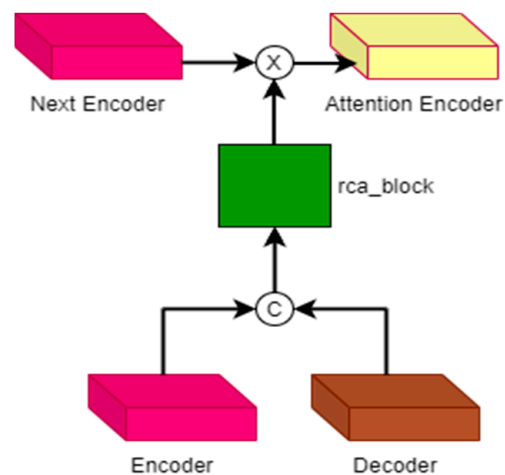


Figure 5. Architecture of proposed CSFF block.

3.1.4. Scale Blend Block

The module in Figure 6 aggregates features obtained from each CS block and makes the channel attention features enhanced by a stack of three RCA blocks to produce attention maps for $\tilde{F}_s$, $\tilde{F}_M$, and $\tilde{F}_L$, which are multiplied back to the original F_s , F_M , and F_L to finally obtain the derain image. The channel sizes of the input and output are nine.

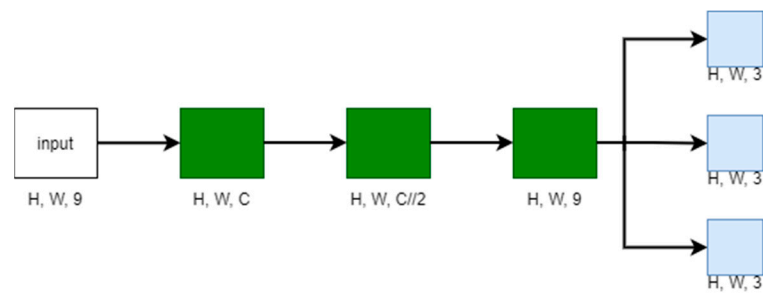


Figure 6. Architecture of proposed SB block.

3.2. Loss Function

The loss function is calculated using the sum of SSIM loss and mean square error (MSE). Brightness, contrast, and structure formulations were used in SSIM to calculate the structural loss. The MSE algorithm was used to calculate the difference between pixel points in the two images in order to derive loss at the pixel level. These algorithms enable the proposed MARCAFNet to obtain structural- and pixel-level information and then balance the loss from different aspects to obtain better results. The SSIM loss is defined as follows:

$$L_{SSIM} = 1 - SSIM(B, \tilde{B}) \quad (1)$$

where B and $\tilde{B}$, respectively, refer to the ground truth and predicted recovered image. In addition to structural loss, we also use MSE loss to facilitate the recovery of more realistic scenes. MSE loss is defined as follows:

$$L_{MSE} = \|B, \tilde{B}\|^2 \quad (2)$$

Finally, the overall loss used to train MARCAFNet is derived as follows:

$$L_{Total} = L_{SSIM} + L_{MSE} \quad (3)$$

This equation is minimized to balance the different loss terms and train the proposed MARCAFNet architecture.

4. Experiment Results and Analysis

The experiments and datasets (synthetic and real-world images) used in this study are herein outlined. The proposed scheme is then compared with state-of-the-art methods, including RESCAN [7], PReNet [8], LPNet [9], MSPFN [10], MPRNet [11], VRGNet [12], PCNet [13], and RAiA-Net [14].

4.1. Datasets

To enable a fair comparison, this study used the same datasets as those used in related studies. Our training set was a combination of data from Rain14000 [31], Rain1800 [41], Rain800 [41], and Rain12 [41]. Rain14000 comprises 1000 original images and 14 distorted versions of each of these original images. Each version presents rain streaks at different angles. Rain14000 comprises 11,200 pairs of training samples with a 1:1 ratio of images of heavy and light rain images. Rain14000 includes landscape images, flower images, traffic images, building images, and people images. Rain1800 consists of a sample of images showing heavy rain. Rain1800 includes landscape images, traffic images, building images, and animal images, etc. Rain800 consists of 700 pairs of training samples showing heavy and light rain streaks. Rain800 includes landscape images, flower images, building images, animal images, and people images, etc. Testing was performed using Rain100H (100 samples of heavy rain) [41] and Rain100L (100 samples of light rain) [41] separately, and the proposed MARCAFNet was compared with the state-of-the-art methods for the use of these testing datasets. The other comparison methods are also the same. The purpose

of this test method is to demonstrate that the model method has good performance even when the test data. Table 1 lists the quantity of samples and the distribution of training and testing data in each dataset.

Table 1. Description of datasets used for synthesis rain images.

Datasets	Rain14000 [31]	Rain1800 [41]	Rain100h [41]	Rain100l [41]	Rain800 [41]	Rain12 [41]
Train samples	11,200	1800	0	0	700	12
Test samples	-	0	100	100	-	0
Test rename	-	-	Rain100h	Rain100l	-	-

4.2. Training Details

Image patches (64×64 pixels) were randomly sampled from each rain and no-rain image and dataset as inputs and outputs. The proposed MARCAFNet was trained using the loss function displayed in (3). The loss function was optimized using the Adam optimizer [42]. The activation function was LeakyReLU, with an alpha of 0.01. The initial learning rate was 0.001, with a decay factor of 0.1 at intervals of 20 epochs. The models were trained for up to 30 epochs. MARCAFNet was implemented using Tensor Flow 2.4.1 and trained using a single NVIDIA RTX 3080 GPU.

4.3. Quantitative Analysis

SSIM and PSNR [43] were used as metric functions for quantitative analysis. To maintain consistency with previous research, this study computed the SSIM and PSNR of the Y-channel (YCbCr color space). As summarized in Table 2, our test results were compared with those for the state-of-the-art methods RESCAN [7], PReNet [8], LPNet [9], MSPFN [10], MPRNet [11], VRGNet [12], PCNet [13], and RAiANet [14]. Our proposed method used the average of five cross-validations as the PSNR and SSIM results. This ensures that the module output was stable and accurate. Overall, MARCAFNet resulted in the most favorable PSNR and SSIM for the Rain100H and Rain100L datasets among [7–14]. For the heavy rain synthetic dataset Rain100H, the proposed MARCAFNet improved a PSNR and SSIM of 2.2 dB and 0.01, respectively, outperforming the current state-of-the-art methods. For the light rain synthetic dataset Rain100L, the proposed MARCAFNet again outperformed the current state-of-the-art methods with a PSNR and SSIM of added 0.6 dB and 0.002, respectively. These results indicate that MARCAFNet can capture both light and heavy rain features in images and can complete the deraining task to obtain clean rain-free images. Notably, the proposed MARCAFNet, using a combination of receptive fields on various scales, can process images with various degrees of complexity, ensuring its robustness across datasets.

Table 2. PSNR and SSIM quantitative comparison of the proposed MARCAFNet with state-of-the-art methods.

Methods	Rain100H		Rain100L	
	PSNR	SSIM	PSNR	SSIM
RESCAN [7]	28.86	0.8646	33.12	0.9534
PReNet [8]	26.83	0.8582	32.53	0.9501
LPNet [9]	23.77	0.8226	25.63	0.8983
MSPFN [10]	28.23	0.8496	32.13	0.9263
MPRNet [11]	30.41	0.8895	36.40	0.9646
VRGNet [12]	30.06	0.8855	36.87	0.9743
PCNet [13]	28.28	0.8699	34.19	0.9526
RAiANet [14]	30.42	0.8941	36.79	0.9673
MARCAFNet (our)	32.62	0.9046	37.47	0.9766

4.4. Qualitative Analysis

As depicted in Figures 7 and 8, the images recovered using the proposed MARCAFNet were compared side-by-side with those derived using other state-of-the-art methods. The blue and red bounding boxes are presented as enlargements in the region of the image. The zoomed-in area is used to clearly present the difference between the proposed method which outperformed other state-of-the-art methods. Figures 7 and 8, respectively, present the results obtained using Rain100H and Rain100L. Figure 9 presents the results obtained using a real-world dataset shared by [14]. In these tests, the other state-of-the-art methods exhibited some weaknesses. The edge details of the trees are blurred in the red boxes in the first row of Figure 7b,e,h,i. In Figure 7, the rain streaks have not been removed and are visible in the red boxes in the first row of (c), (d), (f), and (g). The details of the girl's hair have been blurred, as is visible in the blue and red boxes in the second row of Figure 7b–i. The proposed MARCAFNet improved detail blurring and removed rain streaks, as indicated by Figure 7j. The bucket hat details are blurry and not clearly visible in the red and blue boxes in the first row of Figure 8b–i. In Figure 8, the rain streaks have not been removed, as shown in the red and blue boxes in the second row of (c), (e), (f), (h), and (i). Detail blurring and rain streak removal were superior when the proposed MARCAFNet was employed, as depicted in Figure 8j. In short, the blurring and smoothing of image details resulted from an inability to distinguish between rain streaks and texture features in the images.

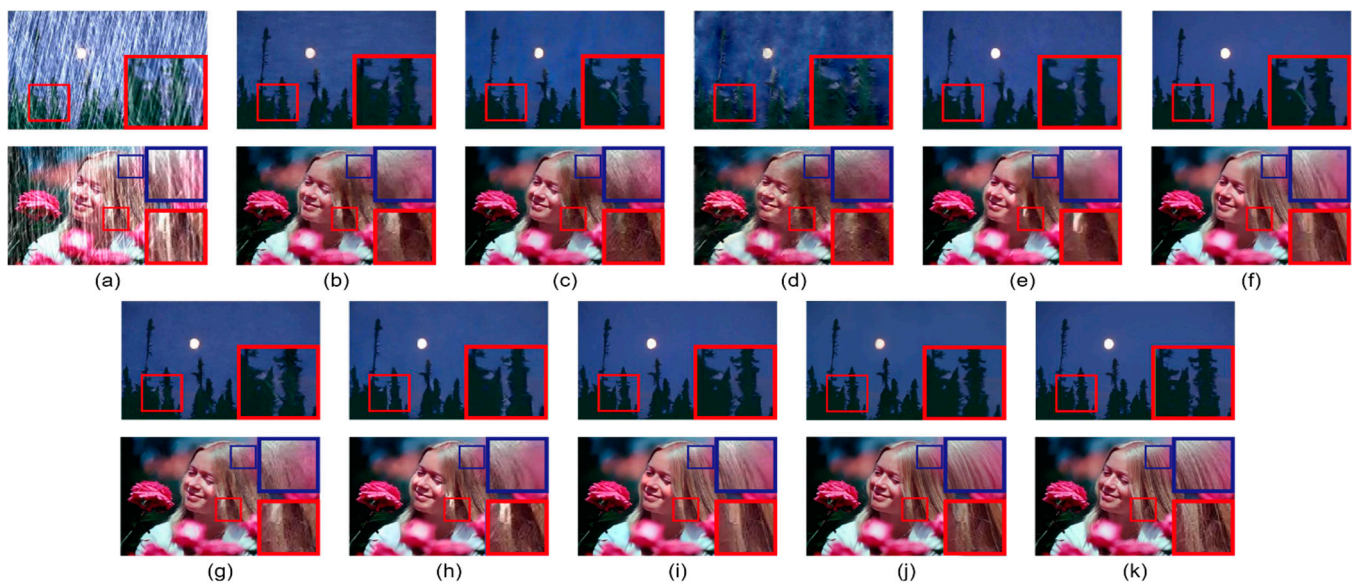


Figure 7. Deraining task performance of the proposed MARCAFNet and state-of-the-art methods when using the Rain100H dataset. (a) Rainy image, (b) RESCAN, (c) PReNet, (d) LPNet, (e) MSPFN, (f) MPRNet, (g) VRGNet, (h) PCNet, (i) RAiANet, and (j) MARCAFNet, and (k) ground truth.

The source of the deraining result image depicted in Figure 9 is the real-world image shared by [15]. In Figure 9, the rain streak has not been removed on the clothes and is still visible in the blue boxes in the first row for (b)–(h). The folds of the pants have been blurred in the red boxes in the first row for (b)–(h). The intervals of the tiles are blurred, as visible in the red and blue boxes in the two rows of (b)–(h). Detail blurring and rain streak removal were improved when the proposed MARCAFNet was used, as illustrated in Figure 9i.

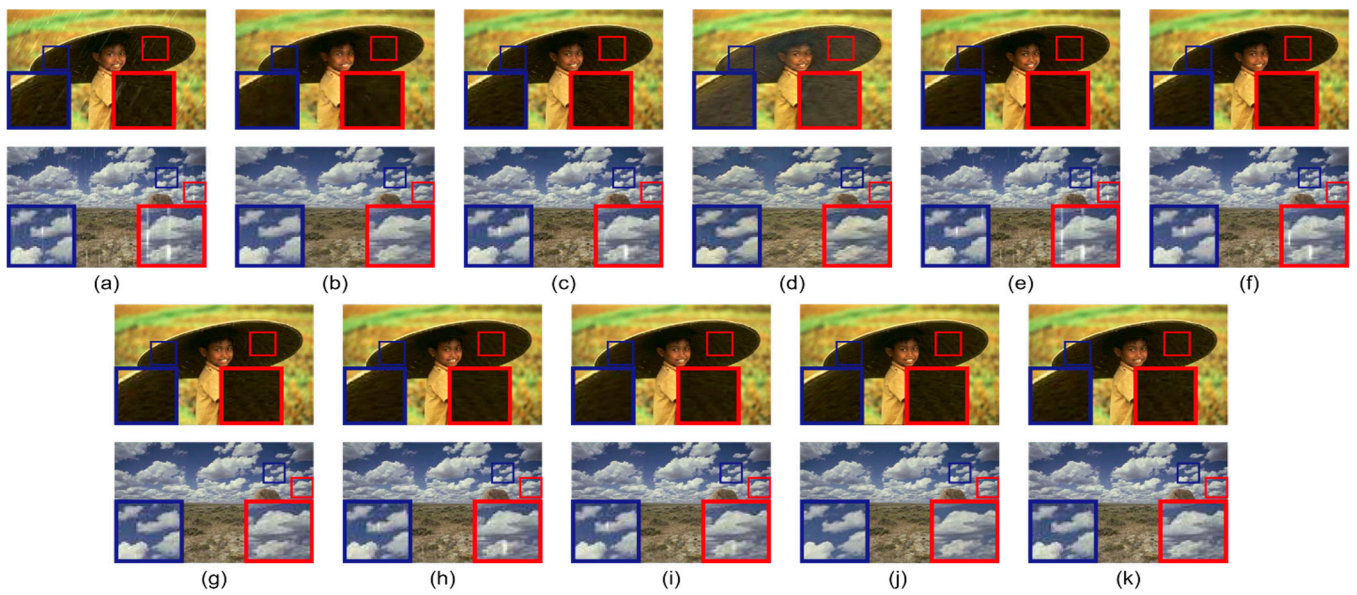


Figure 8. Deraining task performance of the proposed MARCAFFNet and state-of-the-art methods when using the Rain100L dataset. (a) Rainy image, (b) RESCAN, (c) PReNet, (d) LPNet, (e) MSPFN, (f) MPRNet, (g) VRGNet, (h) PCNet, (i) RAiANet, (j) MARCAFFNet, and (k) ground truth.

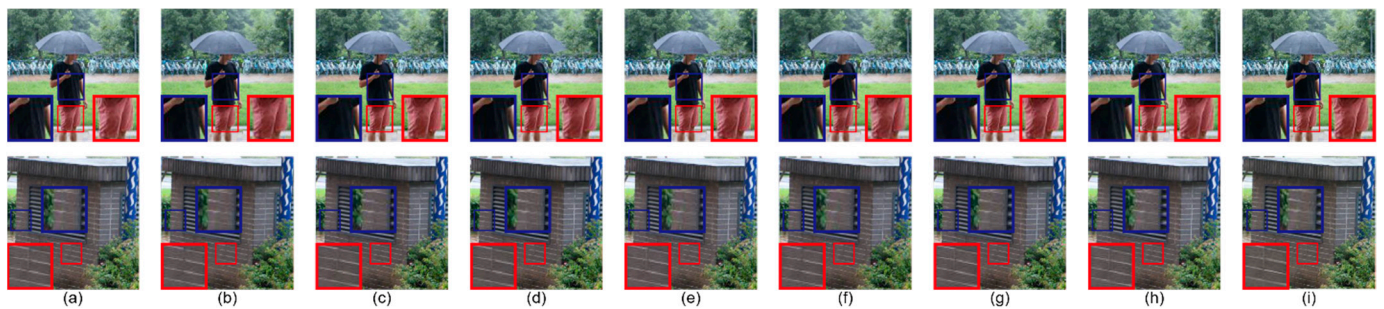


Figure 9. Deraining task performance of the proposed MARCAFFNet and state-of-the-art methods when using the real-world dataset. (a) Rainy image and results, (b) RESCAN, (c) PReNet, (d) MSPFN, (e) MPRNet, (f) VRGNet, (g) PCNet, (h) RAiANet, and (i) MARCAFFNet.

4.5. Ablation Study

To demonstrate the reliability of the proposed MARCAFFNet for various configurations, ablation studies were performed using CSFF, SB, and CS blocks with various dilation rates.

4.5.1. Effectiveness of Sub-Modules

The performance of the proposed scheme was assessed through four experiments using various configurations of the CSFF and SB blocks (see Table 3). The four experiments had the following setups: M1 (without CSFF and SB blocks), M2 (without CSFF block), M3 (without SB block), and M4 (with CSFF and SB blocks). For the models without the CSFF block, cross-stage multiplication was removed. For models without the SB block, the recovered result was summed across the three scales. Overall, the SB block had a more pronounced effect on performance by increasing the adaptivity of F_s , F_M , and F_L .

Table 3. Average ablation results for various combinations of the proposed sub-models when applied to various synthetic datasets.

Model	M1	M2	M3	M4
CSFF	✗	✗	✓	✓
SB	✗	✓	✗	✓
PSNR	27.87	33.65	27.94	35.04
SSIM	0.8118	0.9403	0.8131	0.9406

4.5.2. Effectiveness of Dilation Rate

The performance of the proposed scheme was assessed through four experiments using various dilation rate configurations to assess the effectiveness of the convolution kernels of various scales (see Table 4). The four experiments had the following setups: the dilation rates used were 1, 2, and 3 in the first, second, and third columns, respectively. In the fourth columns, the dilation rates used were 1, 2, and 3 in ascending order. The data presented in Table 4 indicate that when all three modules had a dilation rate of 1, the SSIM values were higher than those for dilation rates of 2 and 3. This was because SSIM focused on the pixel-to-pixel relationship in the calculation. When all three modules had a dilation rate of 1, the pixels of the whole image were not ignored, which resulted in a higher SSIM. When all three modules had a dilation rate of 2, the PSNR was higher than when the modules had a dilation rate of 1. This was because PSNR calculates the value of a single pixel; therefore, with a dilation rate of 2, the current pixel can be supplemented with a wider range of features in the image, increasing the module's robustness for use across different environmental images. Both the SSIM and PSNR were lowest when a dilation rate of 3 was applied. This was because the insufficient size of the training image generated dilation factor gridding effects [16], resulting in module underperformance in this experiment. MARCAFFNet combines the modules of three scales to simultaneously obtain the features of each pixel on small-scale learning images and obtain the features of a wider range of pixels on large-scale learning images, thereby generating favorable results.

Table 4. Average PSNR and SSIM of proposed MARCAFFNet when using CS blocks with various dilation rates in two synthetic datasets (Rain100H and Rain 100L).

Dilation Rate	1,1,1	2,2,2	3,3,3	1,2,3
PSNR	33.99	34.46	32.70	35.04
SSIM	0.9403	0.9329	0.9259	0.9406

5. Conclusions

The proposed MARCAFFNet architecture uses kernels of different dilation rates for the recovery of image features and the elimination of rain streaks. The proposed model applies the RCA block to the CS block in accordance with the encoders–decoders of U-Net. The RCA block is meant to enhance the representation of image content in order to obtain detailed information related to features in the rain-streaked image, thereby enabling the CS block to obtain global features as well as attention features. The proposed CSFF block connects encoder–decoders of the same scale to pass channel features to the kernel with larger receptive fields to enhance the image detail information. The consecutive connection of three CS blocks enables the capturing of textural features, contour features, and rain streak features, which are then aggregated into a new feature map by the SB block. Finally, the image is recovered through adaptive summation. The proposed MARCAFFNet outperformed the state-of-the-art methods, and the present findings revealed that deraining a single image requires the integration of global and local loss to preserve the structure of the image as well as the local features.

6. Future Works

The quantitative experiment of the proposed MARCAFNNet did not achieve the highest value of the evaluation standard. Currently, we have found that the recovery of rain streaks in some directions of the test data is not very good by using manual statistical analysis. We plan to add datasets and other methods to enhance the model's feature extraction capability and further improve its performance. For example, increasing the direction datasets more and preprocessing the pictures could increase the capturing of useful features. Furthermore, in the qualitative experiment, some details were different from the ground truth without rain, implying that the structural- and pixel-level feature extraction of the model could be improved. Therefore, in the future, we plan to add other methods to enhance the model's feature extraction capability and further improve its performance. For example, preprocessing the pictures could increase the capturing of useful features and increase the stages of image deraining, thereby improving the model's deraining performance. Furthermore, the extended experiments indicated that rain removal would be improved if the size of the training images was closer to the model's architecture.

Author Contributions: Conceptualization, J.-G.W. and C.-S.W.; methodology, J.-G.W. and C.-S.W.; software, C.-S.W.; validation, C.-S.W.; formal analysis, J.-G.W. and C.-S.W.; investigation, J.-G.W. and C.-S.W.; resources, J.-G.W.; data curation, J.-G.W.; writing—original draft preparation, J.-G.W.; writing—review and editing, J.-G.W.; visualization, J.-G.W.; supervision, J.-G.W.; project administration, J.-G.W.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available and can be obtained. (Web link to datasets: <https://xueyangfu.github.io/projects/cvpr2017.html> [31], accessed on 1 December 2021). (Web link to datasets: <https://github.com/nnUyi/DerainZoo/blob/master/DerainDatasets.md> [41], accessed on 1 December 2021).

Conflicts of Interest: The authors declare no conflict of interest. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

References

1. Sun, Z.; Bebis, G.; Miller, R. On-road vehicle detection using Gabor filters and support vector machines. In Proceedings of the International Conference on Digital Signal Processing, Santorini, Greece, 1–3 July 2002; pp. 1019–1022.
2. Janai, J.; Guney, F.; Behl, A.; Geiger, A. Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art. *Found. Trends Comput. Graph. Vis.* **2020**, *12*, 1–308. [\[CrossRef\]](#)
3. Kang, L.W.; Lin, C.W.; Fu, Y.H. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.* **2011**, *21*, 1742–1755. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wang, Y.; Liu, S.; Chen, C.; Zeng, B. A hierarchical approach for rain or rain or snow removing in a single color image. *IEEE Trans. Image Process.* **2017**, *26*, 3936–3950. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Li, Y.; Tan, R.T.; Guo, X.; Lu, J.; Brown, M.S. Rain streak removal using layer priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2736–2744.
6. Yu, L.; Yong, X.; Hui, J. Removing rain from a single image via discriminative sparse coding. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 3397–3405.
7. Li, X.; Wu, J.; Lin, Z.; Liu, H.; Zha, H. Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 254–269.
8. Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; Meng, D. Progressive image deraining networks: A better and simpler baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 20–25 June 2019; pp. 3937–3946.
9. Fu, X.; Liang, B.; Huang, Y.; Ding, X.; Paisley, J. Lightweight pyramid networks for image deraining. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 1794–1807. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; Jiang, J. Multi-Scale Progressive Fusion Network for Single Image Deraining. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8346–8355.

11. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-Stage Progressive Image Restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 14816–14826.
12. Wang, H.; Yue, Z.; Xie, Q.; Zhao, Q.; Zheng, Y.; Meng, D. From rain generation to rain. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 14791–14801.
13. Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Wang, Z.; Wang, X.; Jiang, J.; Lin, C. Rain-Free and Residue Hand-in-Hand: A Progressive Coupled Network for Real-Time Image Deraining. *IEEE Trans. Image Process.* **2021**, *30*, 7404–7418. [[CrossRef](#)] [[PubMed](#)]
14. Yin, H.; Deng, H. RAiA-Net: A Multi-Stage Network with Refined Attention in Attention Module for Single Image Deraining. *IEEE Signal Process. Lett.* **2022**, *29*, 747–751. [[CrossRef](#)]
15. Korus, P.; Huang, J. Multi-scale analysis strategies in PRNU-based tampering localization. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 809–824. [[CrossRef](#)]
16. Wang, P.; Chen, P.; Yuan, Y.; Liu, D.; Huang, Z.; Hou, X.; Cottrell, G. Understanding Convolution for Semantic Segmentation. In Proceedings of the IEEE Winter Conference, Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1451–1460.
17. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
18. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
19. Zhou, W.; Alan Conrad, B.; Hamid Rahim, S.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612.
20. Garg, K.; Nayar, S.K. Detection and removal of rain from videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; pp. 528–535.
21. Zhang, X.; Li, H.; Qi, Y.; Leow, W.K.; Ng, T.K. Rain removal in video by combining temporal and chromatic properties. In Proceedings of the IEEE International Conference on Multimedia and Expo, Hilton Toronto, ON, Canada, 9–12 July 2006; pp. 461–464.
22. Chen, Y.L.; Hsu, C.T. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 1968–1975.
23. Kim, J.H.; Sim, J.Y.; Kim, C.S. Video deraining and desnowing using temporal correlation and low-rank matrix completion. *IEEE Trans. Image Process.* **2015**, *24*, 2658–2670. [[CrossRef](#)] [[PubMed](#)]
24. Li, M.; Xie, Q.; Zhao, Q.; Wei, W.; Gu, S.; Tao, J.; Meng, D. Video rain streak removal by Multiscale Convolutional sparse coding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6644–6653.
25. Liu, J.; Yang, W.; Yang, S.; Guo, Z. Erase or fill? Deep joint recurrent rain removal and reconstruction in videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3233–3242.
26. Yue, Z.; Xie, J.; Zhao, Q.; Meng, D. Semi-Supervised Video Deraining with Dynamical Rain Generator. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 642–652.
27. Huang, D.A.; Kang, L.W.; Wang, Y.C.; Lin, C.W. Self-learning based image decomposition with applications to single image denoising. *IEEE Transaction. Multimed.* **2014**, *16*, 83–93. [[CrossRef](#)]
28. Zhu, L.; Fu, C.W.; Lischinski, D.; Heng, P.A. Joint bi-layer optimization for single-image rain streak removal Deep-learning. In Proceedings of the International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2526–2534.
29. Eigen, D.; Krishnan, D.; Fergus, R. Restoring an image taken through a window covered with dirt or rain. In Proceedings of the International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 633–640.
30. Fu, X.; Huang, J.; Ding, X.; Liao, Y.; Paisley, J. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* **2017**, *26*, 2944–2956. [[CrossRef](#)] [[PubMed](#)]
31. Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; Paisley, J. Removing rain from single images via a deep detail network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3855–3863.
32. Hu, X.; Fu, C.W.; Zhu, L.; Heng, P.A. Depth-attentional features for single-image rain removal. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 20–25 June 2019; pp. 8022–8031.
33. Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; Lau, R.W. Spatial attentive single-image deraining with a high quality real rain dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 20–25 June 2019; pp. 12262–12271.
34. Zhang, H.; Sindagi, V.; Patel, V.M. Image de-raining using a conditional generative adversarial network. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 3943–3956. [[CrossRef](#)]
35. Chen, L.; Lu, X.; Zhang, J.; Chu, X.; Chen, C. HINet: Half Instance Normalization Network for Image Restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Nashville, TN, USA, 19–25 June 2021; pp. 182–192.
36. Wang, Q.; Sun, G.; Dong, J.; Zhang, Y. PFDN: Pyramid Feature Decoupling Network for Single Image Deraining. *IEEE Trans. Image Process.* **2022**, *31*, 7091–7101. [[CrossRef](#)] [[PubMed](#)]

37. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016; pp. 1–13.
38. Maas, A.L.; Hannum, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1–6.
39. Hu, J.L.; Albanie, S.; Sun, G.; Vedaldi, A. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
40. He, K.; Sun, J.; Tang, X. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
41. Yang, W.; Tan, R.T.; Feng, J.; Liu, J.; Guo, Z.; Yan, S. Deep joint rain detection and removal from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1357–1366.
42. Kingma, D.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–9.
43. Huynh, Q.T.; Ghanbari, M. Scope of validity of psnr in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.