Deep Scale-space Mining Network for Single Image Deraining

Pengpeng Li [1], Jiyu Jin [1*], Guiyue Jin [1], Lei Fan [1], Xiao Gao [1], Tianyu Song [1], Xiang Chen [2]
[1] School of Information Science and Engineering, Dalian Polytechnic University
[2] College of Electronic and Information Engineering, Shenyang Aerospace University
1310951990@qq.com, jiyu.jin@dlpu.edu.cn

Abstract

Images captured by outdoor vision systems can often be affected by rain weather, resulting in severe degradation of the visual quality of the captured images. Therefore, image deraining has attracted attention as urgent and challenging research. Many current data-driven approaches achieve better performance but are limited in recovering image details. This is because these methods do not fully mine the correlation of scale-space, which are beneficial for rain removal. In this paper, we design an end-to-end Deep Scale-space Mining Network (DSM-Net) for single image deraining to solve these problems. The proposed network with multi-scale extraction, concurrent attention distillation, and hierarchical information fusion accurately captures scale-space features and learns richer information for better deraining. For better feature extraction, a Multi-scale Attention Block (MAB) is introduced to obtain multi-scale rain streak features by different dilated convolutions. Besides,a Concurrent Attention Distillation Block (CADB) is developed which combined channel attention and subspace attention to calibrate the image features obtained from multi-scale acquisition and hierarchical learning, then eliminate redundant features. Importantly, the overall architecture of DSM-Net is inspired by the HourglassNet and DenseNet, which progressively explores and fuses local and global features at different scales in a hierarchical manner instead of direct concatenation. Extensive experiments on synthetic and real datasets show that the proposed DSM-Net outperforms recent state-of-the-art deraining algorithms in terms of both performance and preservation of image details.

1. Introduction

Rain streaks cause significant blurring and visual quality degradation because they vary in size, direction, and density. In particular, the veil-like visual degradation formed by rain streaks superimposed on the background together with raindrops in the air reduces the contrast and visibility of the scene [5, 28, 37]. Therefore, single image deraining has become a necessary pre-processing step in many practical application scenarios, such as scene analysis [15], recognition [10], object tracking [24], intelligent security, and road condition detection under autonomous driving. With the extraordinary results of deep learning in image processing, solving single image deraining has become a research hot-spot [9,20,40,42]. Compared with video deraining, single image deraining is more challenging because less correlation information is available. Therefore, in recent years, more researchers have focused on designing algorithms for single image deraining, which has gradually transitioned from model-driven to data-driven [19].

Traditional model-driven methods include filter-based methods and priori-based methods: filter-based methods use physical filtering to recover clean images [4], while priori-based methods consider single image deraining as an optimization problem, which typically includes sparse prior [40], Gaussian Mixture Model (GMM) [20] and low-rank representation [3]. However, the model-driven approach based on the model can only filter out the noise that obeys a specific distribution (such as Gaussian noise). The physical model also has some limitations and does not sufficiently cover some essential factors in real images with rain, such as rain streaks of different sizes, directions, and densities. Therefore the recovery effect is limited.

Compared with the model-driven-based methods, the data-driven methods treat single image deraining as a process of learning a nonlinear function [39]. In recent years, driven by deep learning techniques, researchers have started to use Convolutional Neural Networks (CNNs) [38], Generative Adversarial Networks (GANs) [18, 41], and semi/unsupervised learning methods to solve single image deraining [34]. One of the CNN-based research methods was first proposed by Yang et al. [38], which built a joint rain detection and removal network (JORDER) focusing on removing overlapping rain streaks under heavy rain. This method achieved impressive results under heavy rain con-

*Corresponding author : Jiyu Jin
Fund Project: Scientific Research Project of the Education Department of Liaoning Province (LJKZ0518, LJKZ0519)

ditions. With the popularity of CNNs, more CNN-based methods have been proposed [2, 23, 32, 35]. GAN-based research methods started later, which was introduced to capture some features in severe weather that cannot be modeled and synthesized to reduce the gap between the generated results and real clean images. Recently, to further improve the recovery performance on real images with rain, semi-supervised and unsupervised learning methods have been proposed, which learn features directly from real rain data to improve generality and scalability. Although the above methods have achieved good results in many application scenarios, they still have many limitations. Due to the difficulty of single image deraining, how to make full use of the shallow and deep convolution features of the depth model to explore the scale-space features is of great importance for deraining. In addition, many existing algorithms make few attempts to eliminate redundant feature information resulting in usually not better structure restoration and detail preservation while their receptive fields are limited and cannot deal with extremely rainy conditions. Therefore, it is necessary to explore the scale-space feature correlation from a global and local perspective.

We propose a Deep Scale-Space Mining Network (DSM-Net) for single image deraining that combines multi-scale feature extraction, concurrent attention feature distillation, and hierarchical feature information fusion. Specifically, the entire network first uses average pooling to achieve a multi-scale hierarchical parallel structure, and then uses the integrated densely connected Multi-scale Attention Blocks (MAB) to extract rich detailed features. The proposed Concurrent Attention Distillation Block (CADB) is embedded in the multi-scale attention block and the cross-layer fusion to recalibrate the features obtained in and between layers respectively. To maximize the use of multi-scale features from different sources, intra-layer and inter-layer fusion are achieved through dense connection and down sampling, respectively.

In summary, our major contributions are as follows:

- We propose a DSM-Net to explore and aggregate scale-space correlations for the specific image deraining task, with a novel hierarchical mining architecture to effectively learn richer feature representations.
- We propose a MAB and a CADB. The MAB uses dilated convolutions of different sizes to extract features from different scales. The CADB combines channel attention and subspace attention mechanisms to recalculate the rain streaks features map in order to reduce useless features and retain space and background information.
- We perform experiments on both synthetic and real-world rain datasets (4 synthetic and 2 real-world datasets). In both visual and quantitative comparisons, our propose network surpasses state-of-the-art approaches. In addition, ablation research is presented to validate the rationale and necessity of the critical modules included in our network.

2. Related Work

In this section, some image deraining methods are reviewed. At the moment, single image rain removal methods are classified into three types: filtering-based methods, priori-based methods, and deep learning-based methods. The model-based method is another name for the filtering and prior method. Following is a brief overview of the most relevant deraining technologies.

2.1. Model Based Methods

Xu et al. [16] proposed a single image rain removal algorithm that includes a guided filtering kernel. In short, it first uses the chromaticity characteristics of rain streaks to produce a rain-free image with lower accuracy, and then filters the rain image to produce a rain-free image with higher accuracy. Ding et al. [4] created a rain-free image using a guided L0 smoothing filter to improve the performance of a single image. In recent years, the Maximum A Posteriori (MAP) [8, 21] method of removing rain from a single image has been widely used, which can be mathematically described as:

$$\max_{\mathrm{B,R}\in\Omega} p(\mathrm{B,R} \mid \mathrm{O}) \propto p(\mathrm{O} \mid \mathrm{B,R}) \cdot p(\mathrm{B}\mid) \cdot (\mathrm{R}), \quad (1)$$

where $\mathrm{O} \in \mathbb{R}^{h\times w}, \mathrm{B} \in \mathbb{R}^{h\times w}$, and $\mathrm{R} \in \mathbb{R}^{h\times w}$ denote the observed rainy image, rain free image, and rain streaks, respectively. $p(\mathrm{B,R} \mid \mathrm{O})$ is the posterior probability and $p(\mathrm{O} \mid \mathrm{B,R})$ is the likelihood function. $\Omega := \{\mathrm{B,R} \mid 0 \leq \mathrm{B}_i, \mathrm{R}_i \leq \mathrm{O}_i, \forall i \in [1, M \times N]\}$ is the solution space.

Fu et al. [7] described image removal as an image decomposition problem using morphological component analysis (MCA). First, bilateral filtering is used to divide the rainy image into two parts: low-frequency component and high-frequency component. The low-frequency component and non-rain component are then combined to obtain the rain removal result. Gu et al. [9] recently proposed a sparse representation model based on joint convolution analysis and synthesis (JCAS), which uses analytical sparse representation (ASR) to approximate the image's large-scale structure and synthetic sparse representation (SSR) to describe the image's Fine texture. JCAS can effectively extract the image texture layer without overly smoothing the background layer due to the complementarity of ASR and SSR.

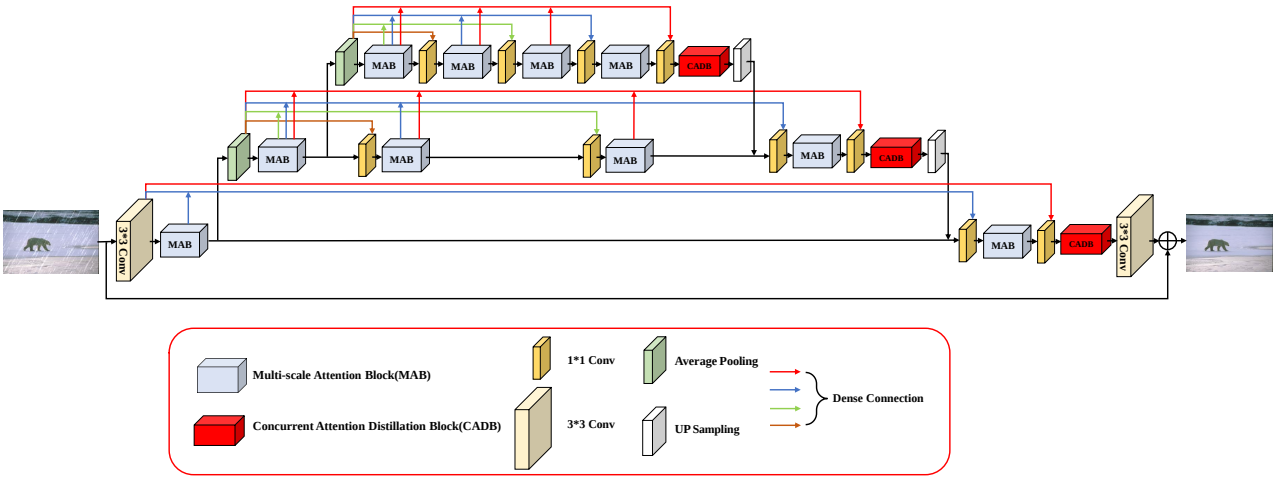

Figure 1. The overall architecture of proposed DSM-Net for image deraining. CADB and MAB are shown in Figure 2 and Figure 3 respectively. The goal of the DSM-Net is to recover the corresponding rain-free image from the rainy image.

2.2. Deep Learning Based Methods

Single image rain removal based on deep learning started in 2017. Yang et al. [38] constructed a joint rain detection and removal network (JORDER) to focus on the removal of overlapping rain streaks under heavy rain. The network can better monitor the rain and locate the rain through prediction. This method has achieved impressive results under heavy rain conditions, but it may delete some texture details by mistake. Qian et al. [25] designed an attention generation network whose basic idea is to inject visual information. Since the deep residual network (ResNet) [12] has achieved the greatest success in the field of deep learning, Fu et al. [5] further proposed a deep detail network (DDN) to achieve better deraining effect. Existing deep learning methods usually treat the network as an end-to-end mapping module, instead of studying the rationality of removing rain streaks [23, 32]. Li et al. [17] proposed a non-local enhancement encoder-decoder network, which can effectively learn more abstract features, so as to achieve more accurate image removal while retaining image details.

In order to alleviate the problem that the deep network structure is difficult to reproduce, Ren et al. [26] presented a simple and effective progressive recurrent deraining network (PReNet). Lightweight pyramid networks (LPNet) [6] pursued a light-weighted pyramid to eliminate rain, resulting in a network that was simple and comprised fewer parameters. However, most of the existing single image deraining networks have not well noticed the internal connection of rain streaks at different scales. RESCAN [36] employed the dilated convolution method to obtain contextual information, and used a recurrent neural network to remodel the rain features. GCANet [1] adopted smooth dilated convolution instead of dilated convolution, and fused high-level and low-level features to improve the recovery effect. SPANet [31] created a recurrent network to capture spatial contextual information from local to global scales. [34]calculateed the residual difference between the input image and the derained image used a semi-supervised learning method. RCDNet [30] represented the rain feature with a convolution dictionary and simplified the network with proximal gradient descent technology. Chen et al. [2] presented a multi-scale hourglass hierarchical fusion network (MH2F-Net) in end-to-end manner, this network accurately obtained rain trace features through multi-scale extraction, hierarchical extraction and information fusion. However, existing multi-scale deraining methods do not fully exploit the relevance of scale-space. Inspired by hourglass networks and dense networks, this paper designs a hierarchically structured feature mining framework to efficiently learn richer features for better deraining.

3. Proposed Method

In this section, the design of the overall DSM-Net architecture will be described. The key modules included in the designed network are described in the subsections, as well as a description of loss function.

3.1. The framework of DSM-Net

This paper proposes an effective DSM-Net based on the DenseNet [13] and HourglassNet [22], which is an end-to-end network that can input any rainy image for training. As is shown in Fig. 1 , the overall architecture of DSM-Net mainly consists of MAB and CADB for feature extraction and distillation, respectively.

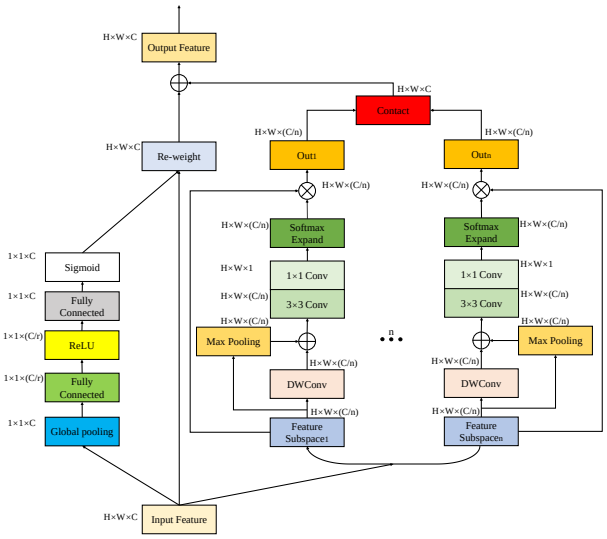

Figure 2. The overall architecture of proposed Concurrent Attention Distillation Block for image deraining.

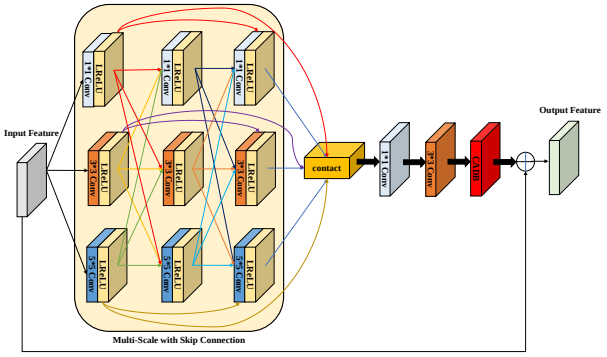

Figure 3. The overall architecture of our proposed Multi-scale Attention Block (MAB). It is mainly composed of a multi-scale feature acquisition module and a CADB. The CADB has been shown in Figure 2. This architecture enables the network to better explore and reorganize features in different scales.

3.2. Concurrent Attention Distillation Block

In terms of the single image rain removal problem, the key is how to better acquire rain features and characterize them for removal. Although a deeper network facilitates layer-by-layer extraction of rain features, as the depth of the network increases, the ability for characterizing with features will gradually weaken with the process of transmission, and a large amount of redundant feature information will be generated.

Therefore how to solve these problems will directly affect the quality of the image after enhancement. In this paper, we adopt a simple distillation structure with attention mechanism as shown in Fig. 2. The CADB is placed at the cross-layer fusion and also embedded in each MAB to solve the information loss in the process of multi-scale feature acquisition and transmission and fusion. The core of CADB is the recalibration of feature information using concurrent channel and subspace attention mechanisms to achieve feature distillation. It is useful for mining scale space feature information.

Concurrent Channel and Subspace Attention Mechanism Inspired by the success of visual attention mechanisms, we implement a concurrent structure by combining the channel attention module [35] and the subspace attention module [27] to eliminate a large number of aimless features and extract more useful hierarchical features. The concurrent channel and subspace attention mechanisms investigate useful spatial and channel components and allow only features containing useful information to proceed.

3.3. Multi-scale Attention Block

Multi-scale feature acquisition methods effectively combine image features at different scales and are now widely used to collect useful information about objects and their surroundings. In order to improve the ability of network representation even further, MAB employs inter-layer multi-scale information fusion, which realizes information fusion between features of different scales. This structure also ensures that the input information is propagated through all parameter layers, allowing the original image's characteristic information to be better learned.

Under the guidance of the above ideas, we propose MAB and use it to learn the different scale space feature information in rainy images more comprehensively and effectively, as shown in Fig. 3. The MAB can be described in detail with mathematical formulas. Referring to Fig. 3, the input feature image of MAB is set as F_{in}, which first passes through the convolutional layers with the convolution kernel sizes of $1\times1, 3\times3$, and 5×5, and the output is expressed as follows:

$$F_a^{1\times1} = \text{Conv}_{1\times1}\left(F_{in}; \theta_a^{1\times1}\right), \tag{2}$$

$$F_a^{3\times3} = \text{Conv}_{3\times3}\left(F_{in}; \theta_a^{3\times3}\right), \tag{3}$$

$$F_a^{5\times5} = \text{Conv}_{5\times5}\left(F_{in}; \theta_a^{5\times5}\right), \tag{4}$$

where $F_{\text{a}}^{n\times n}$ presents the first layer output of multi-scale convolution with the convolution size of $n\times n$, $\text{Conv}_{n\times n}(\cdot)$ presents convolution operation, and $\theta_{\text{a}}^{n\times n}$ means the hyperparameter formed by the first multi-scale convolutional layer with the convolution kernel size of $n\times n$. The image features can be further extracted by using the convolution kernel size to be $1\times1, 3\times3$, and 5×5

$$F_b^{1\times1} = \text{Conv}_{1\times1}\left(\left(F_a^{1\times1} + F_a^{3\times3} + F_a^{5\times5}\right); \theta_b^{1\times1}\right), \tag{5}$$

$$F_b^{3\times3} = \mathrm{Conv}_{3\times3}\left(\left(F_a^{1\times1} + F_a^{3\times3} + F_a^{5\times5}\right); \theta_b^{3\times3}\right), \tag{6}$$

$$F_b^{5\times5} = \mathrm{Conv}_{5\times5}\left(\left(F_a^{1\times1} + F_a^{3\times3} + F_a^{5\times5}\right); \theta_b^{5\times5}\right), \tag{7}$$

where $F_{\mathrm{b}}^{n\times n}$ presents the output of the second layer of multi-scale convolution with size $n \times n$, $\mathrm{Conv}_{n\times n}(\cdot)$ presents a convolution of size $n \times n$, and $\theta_{\mathrm{b}}^{n\times n}$ means the hyperparameter formed by the second multi-scale convolutional layer with a size of $n \times n$. Similarly, we can express the output of the multi-scale third layer as follows:

$$F_c^{1\times1} = \left(\left(\mathrm{Conv}_{1\times1}\left(F_b^{1\times1} + F_b^{3\times3} + F_b^{5\times5} + F_a^{1\times1}\right) + F_a^{1\times1}\right); \theta_c^{1\times1}\right), \tag{8}$$

$$F_c^{3\times3} = \left(\left(\mathrm{Conv}_{3\times3}\left(F_b^{1\times1} + F_b^{3\times3} + F_b^{5\times5} + F_a^{3\times3}\right) + F_a^{3\times3}\right); \theta_c^{3\times3}\right), \tag{9}$$

$$F_c^{5\times5} = \left(\left(\mathrm{Conv}_{5\times5}\left(F_b^{1\times1} + F_b^{3\times3} + F_b^{5\times5} + F_a^{5\times5}\right) + F_a^{5\times5}\right); \theta_c^{5\times5}\right). \tag{10}$$

As shown in Fig.3, MAB realizes multi-scale information fusion through convolutional layers with the convolution kernel sizes of 1×1 and 3×3, and finally introduces a CADB to improve feature fusion. We can express the final output of MAB as follows:

$$F_{\text{out}} = \mathrm{CADB}\left(\left(\mathrm{Conv}_{3\times3}\left(\mathrm{Conv}_{1\times1}\left(\mathrm{Cat}\left(F_C^{1\times1}, F_C^{3\times3}, F_C^{5\times5}\right); \eta_1\right); \eta_2\right); \eta_3\right); \eta_4\right) + F_{\text{in}}, \tag{11}$$

where F_{out} denotes the output of the MAB, CADB$(\cdot)$ indicate the Concurrent Attention Distillation Block, respectively, and $\{\eta_1; \eta_2; \eta_3; \eta_4\}$ indicates the hyperparameters of the MAB output.

3.4. Loss Function

The commonly used loss function in image processing research is MSE loss function [11], because it has better results in most application scenarios. However, the disadvantages of the MSE loss function are also obvious. When the application task involves image quality evaluation, it is assumed that the influence of noise is independent of the local features of the image. This is contrary to the human visual system (HVS) [14], so the correlation between MSE as a loss function and image quality is poor. In order to solve the above shortcomings, we combine MSE and structural similarity index (SSIM) loss [33] to propose a loss function, so as to achieve a balance between the effect of image rain removal and image quality evaluation. In this paper, the rain model we refer to is as follows:

$$I = B + R. \tag{12}$$

We can obtain the rain-free background by subtracting rain streaks R from the rainy image I The MSE loss and SSIM loss can be formulated as:

$$L_{MSE} = \frac{1}{HWC}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{C}\left\|\hat{B}_{i,j,k} - B_{i,j,k}\right\|^2, \tag{13}$$

Table 1. Synthetic and real-world datasets.

Datasets	Rain100L	Rain100H	Rain800	Rain1400	MPID	Li et al.
Training Set	200	1800	700	12600	-	-
Testing Set	100	100	100	1400	185	34

where H, W and C represent height, width, and number of channels respectively. $\hat{B}$ and B denote the restored rain-free image and the groundtruth, respectively.

$$\mathrm{SSIM}(\hat{B}, B) = \frac{2\mu_{\hat{B}}^{\mu_B} + C_1}{\mu_{\hat{B}}^2 + \mu_B^2 + C_1} \cdot \frac{2\sigma_{\hat{B}}\sigma_B + C_2}{\sigma_{\hat{B}}^2 + \sigma_B^2 + C_2}, \tag{14}$$

where μ_x, σ_x^2 are the mean and the variance value of the image: x. The covariance of two images is σ_{xy}, C_1 and C_2 are constants value used to maintain equation stability. The value range of SSIM is from 0 to 1 . In the image rain removal problem, the larger the value obtained by SSIM in the interval means that the recovered rain-free image is closer to the real image. Therefore, the loss function based on SSIM can be defined as:

$$L_{SSIM} = 1 - \mathrm{SSIM}(\hat{B}, B). \tag{15}$$

The total loss is defined by combing the MSE loss and the SSIM loss as follows:

$$L = L_{MSE} + \lambda L_{SSIM}, \tag{16}$$

where λ is a hyperparameter that balances the weight between MSE loss and SSIM loss. By properly setting λ, the similarity of each pixel can be ensured while maintaining the global structure. This helps to get a better rain image.

4. Experiments

In this section, the dataset used in the experiment is described, and some details of the experimental environment and settings are described. In order to prove that the proposed DSM-Net has a good effect on the image rain removal problem, we performed a quantitative and qualitative evaluation of the proposed method on synthetic and real datasets, comparing the results to recent state-of-the-art methods. Simultaneously, complete ablation studies are carried out to demonstrate the significance of each component in the proposed network.

4.1. Experimental Settings

Datasets Setup. Four synthetic datasets and two real-world datasets will be used to evaluate the performance of the proposed method. The composition of the datasets is shown in Tab. 1. Experiments on the proposed DSM-Net are carried out on the four synthetic datasets Rain100L [38], Rain100H [38], Rain800 [41], and Rain1400 [5]. Rain streaks of various sizes, shapes, and directions are included in the four

Table 2. The quantitative results in the table are evaluated based on the PSNR and SSIM average results of the synthetic benchmark datasets (Rain100L, Rain100H, Rain800 and Rain1400), and the best results are shown in bold.

Datasets	Rain100L(PSNR/SSIM)	Rain100H(PSNR/SSIM)	Rain800(PSNR/SSIM)	Rain1400(PSNR/SSIM)
Rainy	26.91/0.838	13.35/0.388	21.16/0.652	25.24/0.810
GCANet	31.70/0.932	24.10/0.814	-	27.84/0.841
LPNet	33.39/0.958	24.39/0.820	25.26/0.781	22.03/0.800
RESCAN	36.12/0.970	27.88/0.816	24.09/0.841	29.88/0.905
DDN	–	24.95/0.781	22.16/0.732	27.61/0.901
JORDER	36.55/0.974	22.79/0.697	26.24/0.850	27.55/0.853
SPANet	35.33/0.970	25.11/0.833	24.37/0.861	28.57/0.891
PReNet	37.11/0.971	28.06/0.888	22.83/0.790	30.73/0.920
RCDNet	35.28/0.971	26.18/0.835	24.59/0.821	–
Ours	**38.27/0.982**	**28.62/0.902**	**27.76/0.871**	**30.93/0.929**

datasets. Rain100L is a light rain dataset that contains only one type of rain streak and is made up of 200 training image pairs and 100 test image pairs. The Rain100H dataset includes 5 rain streaks in different directions, as well as 1800 training and 100 test image pairs. Rain800 is made up of 700 training and 100 test image pairs. Rain1400 contains 14 rain streaks of varying sizes and directions, from which 12600 image pairs with rain are chosen as training data and the remaining 1400 image pairs are used for testing. Because the real-world dataset is critical for assessing the performance of image rain removal, we conducted additional experiments on two real-world datasets: one is the MPID dataset proposed by Li et al., and the other is also proposed by Li et al. [29] in 2019. They are made up of 185 and 34 real rain pictures, respectively.

Evaluation Metrics. The performance of image rain removal methods is usually evaluated according peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). The higher the PSNR value, the better the performance of recovering the rainless image from the rainy image. The SSIM value means the similarity of two different images to each other, and the value range is 0 to 1. When the SSIM is closer to 1, the rain removal performance is good. Since it is a basic fact that there is no completely clean image in the real world, which makes it impossible to quantitatively analyze the rain removal effect, we will intuitively evaluate the performance on the real world datasets from the visual effect.

Implementation Details. Fig. 1 shows the overall structure and parameter settings of the designed DSM-Net. For better feature extraction, the number of MAB is set to 10. During the training process, the loss weight λ is set to 0.2, and data is supplemented by randomly cropping 64×64 patch pairs with horizontal flipping. Using the Adam optimization, the parameters are as follows: initial learning rate is 0.001 and batch size is 32, where $\beta 1$ and $\beta 2$ have default values of 0.9 and 0.999, respectively. To improve performance, we train our model with 200 epochs for the Rain100L/H dataset and 100 epochs for the Rain800/1400 dataset. PyTorch is used for all training and testing on a workstation with an NVIDIA Geforce RTX 3080Ti GPU (12G).

4.2. Experimental Results

Results on Synthetic Datasets. In this section, a large number of experiments are conducted on the synthetic datasets Rain100L, Rain100H, Rain800 and Rain1400, which are commonly used in the problem of image rain removal. The results of the DSM-Net proposed in this paper on the datasets compared with some recent mainstream advanced methods: GCANet [1], LPNet [6], RESCAN [36], DDN [5], JORDER [38], SPANet [31], PReNet [26], RCDNet [30]. Tab. 2 shows the quantitative results of the proposed method on the four synthetic datasets. It can be seen that the method proposed in this paper has improved PSNR and SSIM value compared with the advanced method of reference. It shows that DSM-Net has better robustness and versatility.

In addition to the quantitative evaluation of the rain removal effect of the images, some images are provided for intuitive comparison. As shown in Fig. 4 and Fig. 5, Rain100L and Rain100H images are provided for visual comparison. We selected some details in the image and enlarged it. By observing the enlarged local area, although GCANet, JORDER, LPNet, PReNet and RESCAN have removed a lot of rain patterns, they will all cause different degrees of background blur and there are certain shortcomings in the preservation of the image background details. Compared with Ground Truth, the results obtained by the method in this paper have achieved good results. Therefore, by comparing the method proposed in this paper, the rain streaks can be effectively removed while preserving the background details on the synthetic datasets.

Results on Real-world Datasets. In order to evaluate the

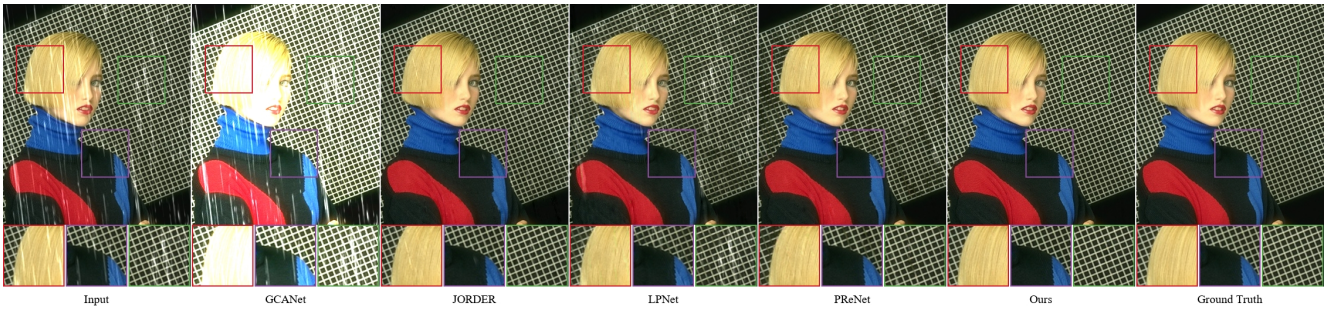

Figure 4. Deraining performance comparison on synthetic dataset (Rain100L).

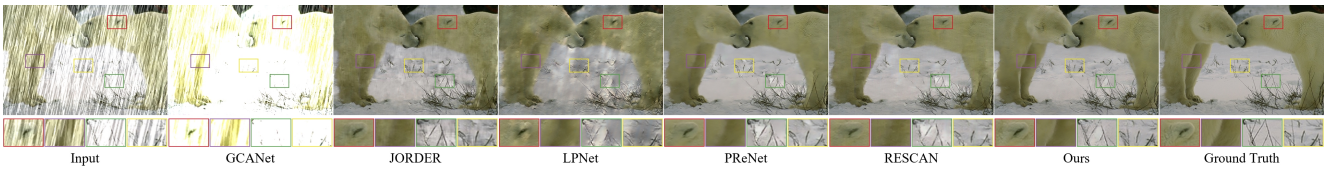

Figure 5. Deraining performance comparison on synthetic dataset (Rain100H).

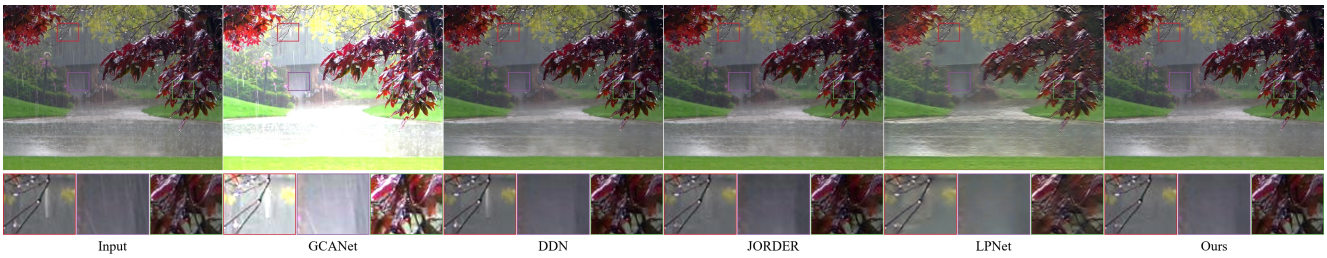

Figure 6. Deraining performance comparison on Real-world dataset.

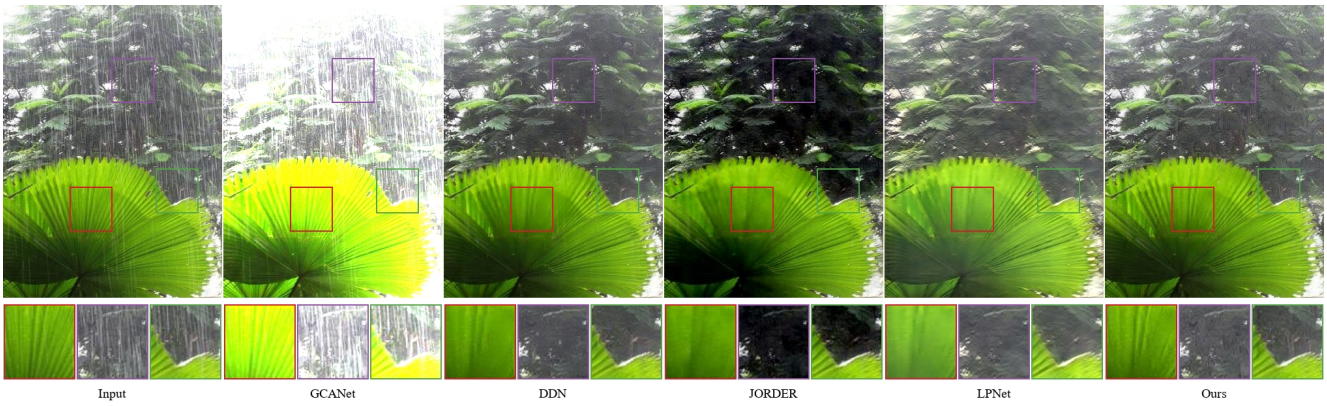

Figure 7. Deraining performance comparison on Real-world dataset.

effectiveness of the proposed method in practical application, the proposed method is compared with the reference method on two real-world rainy datasets mentioned in Section 4.1 for further experimental evaluation. In order to compare the fairness, all methods use the weight of the pre-training model obtained from Rain100H dataset to remove the rain streaks from the real rain dataset. As shown in Fig. 6 and Fig. 7, compared with the most advanced methods, the proposed method produces a more natural and pleasant rain removal image. Specifically, from the enlarged local details, it can be seen that GCANet and DDN can not completely remove the rain streaks in most cases, while JORDER and LPNet blur the details of the rain removal results more. The method in this paper can remove the rain streaks in the real world rain image more effectively and retain more texture details, through mining scale-space feature information.

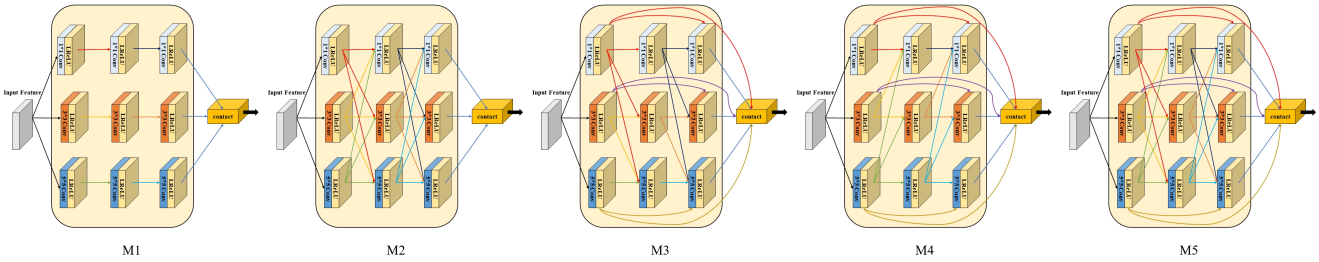

Figure 8. The five kinds of blocks with different operations based MAB.Their abbreviations are as follows: M1: The MAB stripped of all connections. M2: The MAB that does not have a Dense-Fusion connection. M3: The MAB that does not have an Across-Up connection. M4: MAB without an Across-Down connection. M5: The MAB that we proposed.

4.3. Ablation Studies

In order to prove the validity and rationality of the structure configuration and parameter setting in DSM-Net proposed in this paper, ablation experimental studies are conducted. All the studies involved use the Rain100L dataset and are guaranteed to be carried out in the same environment.

Analysis of the proposed CADB. In this paper, we propose CADB which uses a combination of channel attention mechanism and subspace attention mechanism.To explore the structure on the effect of network deraining, We analyze the network designment that consists of Channel Attention mechanism (CA), Sptial Attention menchanism (SP), Subspace Attention mechanism (SA) The results are illustrated in Tab. 3.

Table 3. Ablation study on analysis of the proposed CADB.

Framework	CA	SP	SA	CA+SP	SP+SA	CA+SA
PSNR/SSIM	37.36/0.976	37.90/0.980	37.08/0.978	38.21/0.981	37.86/0.981	**38.27/0.982**

Analysis Number of subspaces for CADB. To investigate the effects of number on distillation capacity, we run deraining experiments with varying numbers of subspaces to the CADB. In particular, the number of subspaces is set to $n \in \{4, 8, 16\}$, and the corresponding PSNR/SSIM results are shown in Tab. 4. As shown, increasing blocks can produce higher PSNR/SSIM values, resulting in better extractive performance. The PSNR improvement appears to be limited after n=8, despite the massive calculated cost. As a result, we choose n=8 as the default parameter.

Table 4. Ablation study on number of subspaces for CADB.

Metric	n= 4	n=8(default)	n=16
PSNR/SSIM	38.08/0.9825	**38.27/0.9829**	38.11/0.9826

Analysis of the proposed MAB. It is meaningful to analyze the different connection operations of MAB. Fig. 8 shows other different modules. The results are shown in Tab. 5. Compared with other moudules, we can see that our proposed default module obtains the best result.

Table 5. Results of different operations of MAB on Rain100L. The best results are marked in bold.

Experiments	M1	M2	M3	M4	M5
Across-Down		✓	✓		✓
Across-up		✓		✓	✓
Dense-Fusion connection			✓	✓	✓
PSNR/SSIM	36.15/0.9757	36.40/0.9760	36.65/0.9770	36.44/0.9765	**38.27/0.9829**

5. Conclusion

In this paper, we propose a deep scale-space mining network (DSM-Net) to deal with single image deraining. A noval multi-scale attention structure is being developed to extract local and global features at multiple scales. In particular, a concurrent attentive distillation block is used first to recalibrate the hierarchical features by utilizing the channel attention module and the subspace attention module feature responses. Furthermore, an advanced Hierarchical feature fusion strategy is introduced to achieve comprehensive feature aggregation, so that features from various sources are progressively discriminated and fused to improve deraining performance. On both synthetic and real-world rainy datasets, quantitative and visual results show that our developed model outperforms other comparing deraining approaches.

References

[1] D. Chen, M. He, Q. Fan, J. Liao, and G. Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019. 3, 6

[2] Xiang Chen, Yufeng Huang, and Lei Xu. Multi-scale hourglass hierarchical fusion network for single image deraining. In *2021 IEEE/CVF Conference on Computer Vision and*

Pattern Recognition Workshops (CVPRW), pages 872–879, 2021. 2, 3

[3] Y. L. Chen and C. T. Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *IEEE International Conference on Computer Vision*, 2013. 1

[4] Xinghao Ding, Liqin Chen, Xianhui Zheng, Yue Huang, and Delu Zeng. Single image rain and snow removal via guided l0 smoothing filter. *Multimedia Tools and Applications*, 75(5):2697–2712, 2016. 1, 2

[5] X. Fu, J. Huang, D. Zeng, H. Yue, and J. Paisley. Removing rain from single images via a deep detail network. In *IEEE Conference on Computer Vision Pattern Recognition*, 2017. 1, 3, 5, 6

[6] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley. Lightweight pyramid networks for image deraining. *IEEE Transactions on Neural Networks and Learning Systems*, 31(6):1794–1807, 2020. 3, 6

[7] Y. H. Fu, L. W. Kang, C. W. Lin, and C. T. Hsu. Single-frame-based rain removal via image decomposition. In *IEEE International Conference on Acoustics*, 2014. 2

[8] J. L. Gauvain and C. H. Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Transction on Speech Audio Processing*, 2(2):291–298, 1994. 2

[9] S. Gu, D. Meng, W. Zuo, and Z. Lei. Joint convolutional analysis and synthesis sparse representation for single image layer separation. In *IEEE International Conference on Computer Vision*, 2017. 1, 2

[10] Liu. Haihua, Tang Na, Shu. Qiling, and Zhang. Wensheng. Computational model based on neural network of visual cortex for human action recognition. *IEEE Transactions on Neural Networks Learning Systems*, 2017. 1

[11] Hang, Zhao, Orazio, Gallo, Iuri, Frosio, Jan, and Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017. 5

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 3

[13] G. Huang, Z. Liu, Vdm Laurens, and K. Q. Weinberger. Densely connected convolutional networks. *IEEE Computer Society*, 2016. 3

[14] K. Q. Huang, Q. Wang, W. U. Zhen-Yang, and H. U. Xue-Long. Multi-scale color image enhancement algorithm based on human visual system (hvs). *Journal of Image and Graphics*, 2003. 5

[15] L. Itti. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1998. 1

[16] X. Jing, Z. Wei, L. Peng, and X. Tang. Removing rain and snow in a single image using guided filter. In *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, 2012. 2

[17] Guanbin Li, Xiang He, Wei Zhang, Huiyou Chang, and Liang Lin. Non-locally enhanced encoder-decoder network for single image de-raining. *ACM*, 2018. 3

[18] R. Li, L. F. Cheong, and R. T. Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[19] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K. Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. Single image deraining: A comprehensive benchmark analysis. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3833–3842, 2019. 1

[20] Y. Li. Rain streak removal using layer priors. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[21] R. Liu, X. Fan, Z. Luo, P. Mu, and J. Chen. Learning bilevel layer priors for single image rain streaks removal. *IEEE Signal Processing Letters*, 2019. 2

[22] A. Newell, K. Yang, and D. Jia. Stacked hourglass networks for human pose estimation. *Springer International Publishing*, 2016. 3

[23] J. Pan, S. Liu, D. Sun, J. Zhang, Y. Liu, J. Ren, Z. Li, J. Tang, H. Lu, and Y. W. Tai. Learning dual convolutional neural networks for low-level vision. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 2, 3

[24] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. *IEEE Transactions on Intelligent Transportation Systems*, pages 1993–2016, 2017. 1

[25] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu. Attentive generative adversarial network for raindrop removal from a single image. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[26] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A better and simpler baseline. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3932–3941, 2019. 3, 6

[27] Rajat Saini, Nandan Kumar Jha, Bedanta Das, Sparsh Mittal, and C. Krishna Moha. Ulsam: Ultra-lightweight subspace attention module for compact convolutional neural networks. In *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1616–1625, 2020. 4

[28] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer Science*, 2014. 1

[29] H. Wang, M. Li, Y. Wu, Q. Zhao, and D. Meng. A survey on rain removal from video and single image. *arXiv preprint arXiv.1909.08326*, 2019. 6

[30] H. Wang, Q. Xie, Q. Zhao, and D. Meng. A model-driven deep neural network for single image rain removal. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3, 6

[31] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attention single-image deraining with a high quality real rain dataset. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12262–12271, 2019. 3, 6

[32] Ye-Tao Wang, Xi-Le Zhao, Tai-Xiang Jiang, Liang-Jian Deng, Yi Chang, and Ting-Zhu Huang. Rain streaks removal for single image via kernel-guided convolutional neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3664–3676, 2021. 2, 3

[33] Z. Wang. Image quality assessment : From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004. 5

[34] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu. Semi-supervised transfer learning for image rain removal. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 3

[35] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. *arXiv preprint 1807.06521*, 2018. 2, 4

[36] L. Xia, J. Wu, Z. Lin, L. Hong, and H. Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV 2018 : Computer Vision – ECCV 2018*, 2018. 3, 6

[37] Wenhan Yang, Robby T. Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6):1377–1393, 2020. 1

[38] W. Yang, R. T. Tan, J. Feng, J. Liu, and S. Yan. Deep joint rain detection and removal from a single image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 3, 5, 6

[39] W. Yang, R. T. Tan, S. Wang, Y. Fang, and J. Liu. Single image deraining: From model-based to data-driven and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2020. 1

[40] L. Yu, X. Yong, and J. Hui. Removing rain from a single image via discriminative sparse coding. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015. 1

[41] H. Zhang, V. Sindagi, and V. M. Patel. Image deraining using a conditional generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017. 1, 5

[42] X. Zhang, H. Li, Y. Qi, W. K. Leow, and T. K. Ng. Rain removal in video by combining temporal and chromatic properties. In *IEEE*, 2006. 1