MCAD-Net: Multi-scale Coordinate Attention Dense Network for Single Image Deraining

Pengpeng Li, Jiyu Jin(✉), Guiyue Jin, Jiaqi Shi, and Lei Fan

Dalian Polytechnic University, Dalian 116034, China
{jiyu.jin,guiyue.jin,fanlei}@dlpu.edu.cn

Abstract. Single image rain removal is an urgent and challenging task. Images acquired under natural conditions are often affected by rain, which leads to a serious decline in the visual quality of images and hinders some practical applications. Therefore, the research of image rain removal has attracted much attention. However, both the model based method and the deep learning based method can not adapt to the spatial and channel changes of rain feature information. In order to solve these problems, this paper proposes an end-to-end Multi-scale Coordinate Attention Dense Network (MCAD-Net) for single image deraining. MCAD-Net can accurately identify and characterize rain streaks and remove them, while preserving image details. To better solve the problem, the Multi-scale Coordinate Attention Block (MCAB) is introduced into the MCAD-Net to improve the ability of feature extraction and representation of rain streaks. MCAB first uses different convolution kernels to extract and fuse multi-scale rain streaks features, and then uses the coordination attention module with adaptation module to recognize rain streaks in different spaces and channels. A large number of experiments have been carried out on several commonly used synthetic datasets and real datasets. The quantitative and qualitative results show that the proposed method is superior to the recent state-of-the-art methods in improving the performance of image rain removal and preserving image details.

Keywords: Deraining · Coordinate attention · Multi-scale

1 Introduction

The images captured from the outdoor vision system are often affected by rain, resulting in a serious degradation of the visual quality of the captured images. Generally, raindrops and rain streaks in the vicinity tend to block or distort the content of the background scene, while raindrops and rain streaks in the distance tend to produce atmospheric texture effects [1–4]. Especially in heavy rain, rain fog, rain streaks and rain particles in the air are superimposed on the background to form a veil-like visual degradation, which greatly reduces the contrast and visibility of the scene. Therefore, rain removal

Fund Project: Scientific Research Project of the Education Department of Liaoning Province (LJKZ0518, LJKZ0519, LJKZ0515).

H. Gao et al. (Eds.): ChinaCom 2021, LNICST 433, pp. 405–421, 2022.
https://doi.org/10.1007/978-3-030-99200-2_31

has become a necessary preprocessing step for subsequent tasks, such as object tracking [5], scene analysis [6], personnel reidentification [7], and road condition detection of automatic driving to further improve its performance. As an important research topic, image rain removal has attracted widespread attention in the field of computer vision and pattern recognition in recent years [3, 8–11]. In many practical application scenarios, there is an urgent need to restore rain-free images in rainy days.

In recent years, with the much application of artificial intelligence in the field of computer vision, especially deep learning has achieved extraordinary results in image processing. The use of deep learning and neural networks to solve the problem of image enhancement and restoration in harsh environments has become research hotspots. The single-image rain removal method has gradually transitioned from model-driven to data-driven [12]. Traditional methods based on model-driven include filtering-based methods and prior-based methods. The method based on filtering is to use physical filtering to restore the rainless image [13–16]. The prior-based method considers single image rain removal as an optimization problem, which includes sparse prior [17] Gaussian Mixture Model (GMM) [10] and low-rank representation [18]. Compared with the model-based method, the data-driven method regards the removal of a single image as a process of learning a nonlinear function [19]. Motivated by the success of deep learning in recent years, researchers have begun to use convolutional neural networks (CNN) or Generative Adversarial Networks (GAN) to model mapping functions [20]. Both the CNN-based design method [21–23] or the GANs-based design method have achieved better results than traditional methods.

Although the methods mentioned above have achieved good results in many application scenarios, there are still many limitations. Because image rain removal is a complex process, most rain removal models based on filtering and prior methods are not enough to cover some important factors in real rainfall images, such as rain streaks. As for the method based on deep learning, it ignores the internal mechanism of rain, making it easy to fall into the process of over-adapting training. Many existing algorithms are limited to some image blocks or limited receptive fields, and can not obtain image feature connection between large regions, resulting in usually unable to restore structure and details. But this kind of information that is ignored due to the limited receptive field has been proved to be very helpful to the image to remove the rain.

To address the above-mentioned issues, the paper present a Multi-scale Coordinate Attention Dense Network called MCAD-Net. The proposed MCAD-Net is based on DenseNet. The DenseNet has several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters. Multi-scale Coordinate Attention Block (MCAB) is introduced to better utilize multi-scale information and feature attention for improving the rain feature representation capability. Combing the features of different scales and layers, multi-scale manner is an efficient way to capture various rain streak components especially in the heavy rainy conditions. Recent studies on single image deraining network design have demonstrated the remarkable effectiveness of channel attention (e.g., the Squeeze-and-Excitation Attention) for lifting model performance, but they generally neglect the positional information, which is important for generating spatially selective attention maps. Therefore, coordinate attention module is

involved in the MCAB. In this way, long-range dependencies can be captured along one spatial direction and meanwhile precise positional information can be preserved along the other spatial direction. The coordinate attention module helps the proposed network to adjust the three-color channels respectively and identify the rainy region properly. We evaluate the proposed network on the public competitive benchmark synthetic and real-world datasets and the results significantly outperform the current outstanding methods on most of the deraining tasks.

In summary, the contributions of this work may be summarized as follows:

1) We propose a MCAD-Net to address the single image deraining problem, which can effectively remove the rain streaks while well preserve the image details. The modified DenseNet is applied to boost the model performance via multi-level features reuse and maximum information flow between layers. It can alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, while fully utilizing the features of different layers to restore the details.
2) To our knowledge, the Multi-scale Coordinate Attention Block (MCAB) is first constructed to improve the representation of rain streaks. This hierarchical structure uses different convolution kernels to generate features of different scales that contain rich hierarchical information. In addition, the inherent correlation of multi-scale features can be used to gain a deeper understanding of the image layout and improve the performance of feature extraction to a large extent. Then, by introducing the coordinate attention module added to the adaptation module, the color channel and spatial location information are used to extract features better.
3) We perform experiments on both synthetic and real-world rain datasets (4 synthetic and 2 real-world datasets). Our proposed network outperforms the state-of-the-art methods in visually and quantitatively comparisons. Furthermore, ablation research is provided to verify the rationality and necessity of the important modules involved in our network.

2 Related Work

In this section, some image deraining methods are reviewed. Compared with the rain removal in video, single image rain removal is more challenging because of less available information. Therefore, more and more researchers pay attention to the algorithm design of single image rain removal in recent years. At present, the existing single image rain removal methods can be divided into three categories: filtering-based methods, prior-based methods and deep learning-based methods. The method based on filtering and prior is also called model-based method.

Model Based Methods. Xu et al. [13] proposed a single image rain removal algorithm with guided filtering kernel [14]. In short, it first uses the chromaticity characteristics of rain streaks to obtain a rain-free image with lower accuracy, and then filters the rain image to obtain a rain-free image with higher accuracy. Ding et al. [24] designed a guided L0 smoothing filter to obtain a rain-free image to improve the performance of a single image to remove rain.

In recent years, the Maximum A Posteriori (MAP) has been widely used in the method of removing rain from a single image [25, 26], which can be mathematically described as:

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

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. $p(\mathrm{B},\mathrm{R}|\mathrm{O})$ is the posterior probability and $p(\mathrm{O}|\mathrm{B},\mathrm{R})$ is the likelihood function. $\Omega := \{\mathrm{B},\mathrm{R}|0 \le \mathrm{B}_i, \mathrm{R}_i \le \mathrm{O}_i, \forall i \in [1, M \times N]\}$ is the solution space.

Various methods have been proposed for designing the forms of all terms involved in (1). Fu et al. [27] used morphological component analysis (MCA) to describe image removal as an image decomposition problem. First, bilateral filtering is used to divide the rainy image into two parts: low-frequency component and high-frequency component, and then the low-frequency component and the non-rain component are combined to obtain the rain removal result. Recently, Gu et al. [11] proposed a joint convolution analysis and synthesis (JCAS) sparse representation model, using analytical sparse representation (ASR) to approximate the large-scale structure of the image, and using synthetic sparse representation (SSR) to describe the image Fine texture. The complementarity of ASR and SSR enables JCAS to effectively extract the image texture layer without excessively smoothing the background layer.

Deep Learning Based Methods. Single image rain removal based on deep learning started in 2017. Yang et al. [3] 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. [28] designed an Attention Generation Network whose basic idea is to inject visual information. Since the deep residual network (ResNet) [29] has achieved the greatest success in the field of deep learning, Fu et al. [2] further proposed a deep detail network (DDN) to achieve better Rain 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 [30, 31]. Li et al. [32] 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. [33] presented a simple and effective progressive recurrent deraining network (PReNet). Chen et al. [34] present a Multi-scale Hourglass Hierarchical Fusion Network (MH2F-Net) in end-to-end manner, this network accurately obtain rain trace features through multi-scale extraction, hierarchical extraction and information fusion. However, most of the existing single-image rain-removing networks have not well noticed the internal connection of rain streaks at different scales. And although they have tried to introduce the attention module in the rain-removing network, whether it is channel attention or spatial attention, they have ignored the position information, which is important for generating spatially selective attention map.

3 Proposed Method

In this section, the MCAD-Net is described in Sect. 3.1 and the Multi-scale Coordinate Attention Block (MCAB) is introduced in Sect. 3.2. The proposed loss function is described in Sect. 3.3.

3.1 Proposed MCAD-Net

This paper proposes an MCAD network based on the DenseNet, which is an end-to-end network that can input any rain image for training. The overall architecture of the MCAD network is shown in Fig. 1. The network effectively achieves the maximum reuse of the rain image features, thereby reducing the number of parameters and improving the efficiency of training and testing. In order to speed up the training process, global skip connections are introduced between different MCAB modules. These skip connections can not only help the back-propagation gradient to update the parameters, but also directly spread the lossless information through the entire network, so it is very helpful for the estimation of the final rain removal image. In addition, because different rain traces have different directions and shapes, MCAB is introduced into the network, which uses multi-scale and the latest feature attention module to obtain various rain trace characteristics and structure information. The details will be described in the following section.

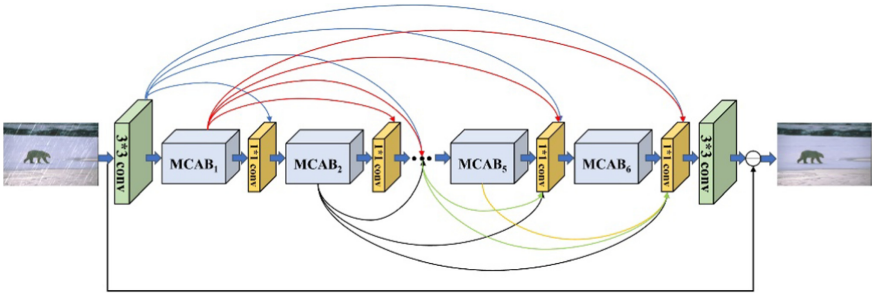

Fig. 1. The overall architecture of proposed MCAD-Net for image deraining. MCAB is shown in Fig. 3. The goal of the MCAD-Net is to recover the corresponding rainless image from the rainy image.

3.2 Design of Multi-scale Coordinate Attention Block (MCAB)

Multi-scale Feature. Multi-scale feature acquisition methods effectively combine image features at different scales, and are currently widely used to acquire useful information about objects and their vicinity. In order to further improve the ability of network representation, inter-layer multi-scale information fusion is applied in MCAB, which realizes the information fusion between features of different scales. This structure also ensures that the input information can be propagated through all parameter layers, so that the characteristic information of the original image can be better learned.

Coordinate Attention. It is well known that a color image is composed of RGB three channels, and the rain density of a rainy image is different on each channel in actual conditions. Most traditional rain removal methods often ignored this situation, until the channel attention mechanism was proposed, which can effectively obtain the important characteristics of rain on different channels. For example, in the attention mechanism network SENet, channel characteristic information is obtained through global average pooling (GAP). In addition, the study also found that the distribution of rain streaks in space is also uneven, so the spatial attention mechanism proposed later is also very important for removing rain from images.

However, whether it is channel attention or spatial attention mechanism, they usually ignore the specific position information, and the position information is very important for generating spatial selective attention feature maps. Therefore, the coordinated attention mechanism proposed by Qi et al. [35] is introduced in MACB to solve the problem of rain removal better, enhance the ability of the network to extract characteristic information, and improve network performance and accuracy. The coordinated attention mechanism is different from the previous channel attention mechanism. It does not convert multiple feature vectors into a single feature vector through 2D global pooling, but decomposes the channel attention into two 1D feature encoding processes, thus achieve the effect of gathering feature information along the X and Y spatial directions respectively. Then, the two feature maps with specific direction information are coded into two attention maps, and each attention map captures the input feature information along a spatial direction. Therefore, the location information can be saved in the generated attention graph. In order to make the coordinated attention mechanism applicable to the training of any rain image, the design proposed in this paper adds an adaptation module to the original coordinated attention mechanism to adjust the dimension of the training output weight, which is called Coordinate Attention-Block as shown in the Fig. 2.

Under the guidance of the above ideas, we propose MCAB and use it to learn the rain streaks information in rainy images more comprehensively and effectively, as shown in Fig. 3.

The MCAB can be described in detail with mathematical formulas. Referring to Fig. 3, the input feature image of MCAB is set as F_{in}, which first passes through the convolutional layers with the convolution kernel sizes of 1×1, 3×3, and 5×5, and the output is expressed as follows:

$$F_a^{1\times1} = Conv_{1\times1}\left(F_{in}; \theta_a^{1\times1}\right) \tag{2}$$

$$F_a^{3\times3} = Conv_{3\times3}\left(F_{in}; \theta_a^{3\times3}\right) \tag{3}$$

$$F_a^{5\times5} = Conv_{5\times5}\left(F_{in}; \theta_a^{5\times5}\right) \tag{4}$$

where $F_{\mathrm{a}}^{n\times n}$ presents the first layer output of multi-scale convolution with the convolution size of $n \times n$, $Conv_{n\times n}(\cdot)$ presents convolution operation, and $\theta_{\mathrm{a}}^{n\times n}$ means the hyperparameter formed by the first multi-scale convolutional layer with the convolution

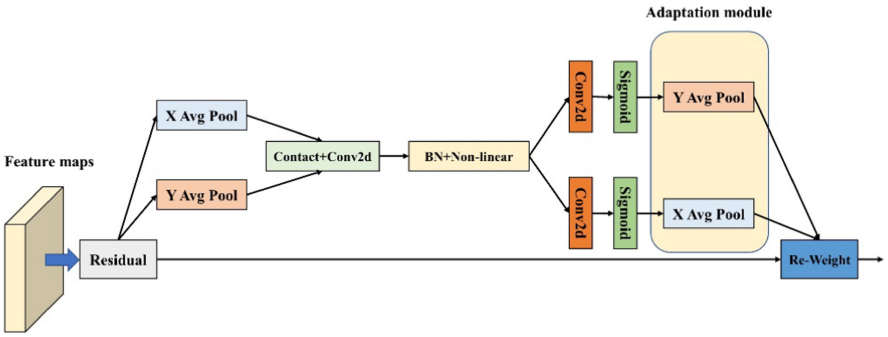

Fig. 2. The overall architecture of our proposed Coordinate Attention-Block for image deraining. Compared with the original coordinated attention, an adaptation module is added. The goal of Coordinate Attention-Block is to capture cross-channel information, while also capturing direction perception and position-sensitive information, which helps the overall network model to more accurately locate and identify more important objects.

kernel size of $n \times n$. The image features can be further extracted by using the convolution kernel size to be 1×1, 3×3, and 5×5

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

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

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

where $F_{\mathrm{b}}^{n\times n}$ presents the output of the second layer of multi-scale convolution with size $n\times n$, $Conv_{n\times n}(\cdot)$ 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:

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

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

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

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

$$F_{\text{out}} = ca\left(\left(Conv_{3\times3}\left(Conv_{1\times1}\left(Cat\left(F_c^{1\times1}, F_c^{3\times3}, F_c^{5\times5}\right); \eta_1\right); \eta_2\right); \eta_3\right); \eta_4\right) + F_{in} \tag{11}$$

where F_{out} denotes the output of the MCAB, $ca(\cdot)$ indicate the coordinated attention mechanism module, respectively, and $\{\eta_1; \eta_2; \eta_3; \eta_4\}$ indicates the hyperparameters of the MCAB output.

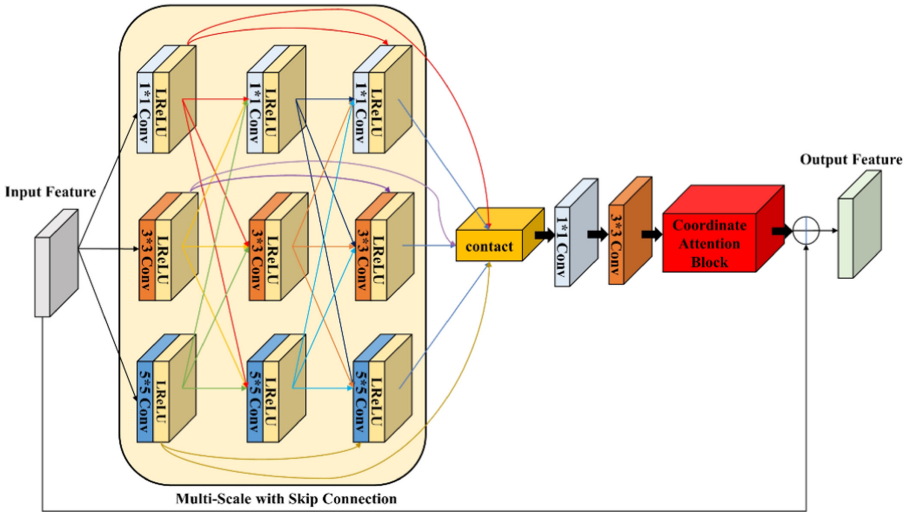

Fig. 3. The overall architecture of our proposed Multi-scale Coordinate Attention Block (MCAB). It is mainly composed of a multi-scale feature acquisition module and a coordinated attention module. The coordinated attention mechanism has been shown in Fig. 2. This architecture enables the network to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

3.3 Loss Function

The commonly used loss function in image processing research is MSE loss function, 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), 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 SSIM 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.

MSE Loss. A pixel-wise MSE loss can be defined as follows

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

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.

SSIM Loss. SSIM is an important indicator to measure the structural similarity between two images [36], with the equation as follows:

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

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:

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

Total Loss. 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 introduced. Then some details of the experimental environment and settings are described. In order to prove that the proposed MCAD-Net has a good effect on the image rain removal problem, we have carried out a quantitative and qualitative evaluation of the proposed method on the synthetic dataset and the real dataset, and compared the results with the recent state-of-the-art methods. At the same time, complete ablation studies are conducted to prove the importance of each component in the proposed network.

4.1 Experiment Details

As shown in Fig. 1, in order to achieve the best image removal effect, the number of MCAB in MCAD-Net is set to 6. The reason will be explained in Sect. 4.6 Ablation Experiment. For all the datasets, we randomly crop 64×64 patch from each input image. The training workstation is configured as Ubuntu 18.04, memory 16G, NVIDIA P102 GPU (10G). During training, the batch size is set to 32 and use the Adam optimizer to train a total of 100 epochs. The initial learning rate is set to 1×10^{-3}, and the learning rate is reduced by half every 25 epochs.

4.2 Datasets and Evaluation Metrics

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 Table 1.

Table 1. Synthetic and real-world datasets.

Datasets	Training set	Testing set	Type
Rain100L	200	100	Synthetic
Rain100H	1800	100	Synthetic
Rain800	700	100	Synthetic
Rain1400	12600	1400	Synthetic
MPID	–	185	Real-world
Li et al.	–	34	Real-world

As shown in Table 1, experiments are conducted on the proposed MCAD-Net on the four synthetic datasets Rain100L [3], Rain100H [3], Rain800 [37] and Rain1400 [2]. The four datasets include rain streaks of various sizes, shapes, and directions. Among them, Rain100L is a light rain dataset which contains only one kind of rain streaks, which is composed of 200 training image pairs and 100 test image pairs. The Rain100H dataset contains 5 rain streaks in different directions, consisting of 1800 training image pairs and 100 test image pairs. Rain800 consists of 700 training image pairs and 100 test image pairs. Rain1400 contains 14 rain streaks of different directions and sizes, from which 12600 image pairs with rain are selected as training data, and the other 1400 image pairs are used for testing. The real-world dataset is very important for evaluating the performance of image rain removal, so we further conducted experiments on two real-world datasets: One of them is the MPID dataset proposed by Li et al., and the other is also proposed by Li et al. in 2019. They are composed of 185 and 34 real pictures of rain [38].

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.

4.3 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 MCAD-Net proposed in this paper on

the datasets compared with some recent mainstream advanced methods: GCANet [39], LPNet [40], RESCAN [22], JORDER [3], JORDER-R [3], JORDER-E [3], DID-MDN [41], SPANet [42], ReHEN [43], PReNet [34], RCDNet [44], MPRNet [45].

Table 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 MCAD-Net has better robustness and versatility.

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.

Methods	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	24.80/0.814	–	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
JORDER-R	36.71/0.980	23.45/0.749	26.73/0.859	–
JORDER-E	37.20/0.979	24.54/0.802	27.10/0.863	–
DID-MDN	28.27/0.857	17.39/0.612	21.89/0.795	27.99/0.869
SPANet	35.33/0.970	25.11/0.833	24.37/0.861	28.57/0.891
ReHEN	–	27.97/0.864	26.96/0.854	30.63/0.918
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	–
MPRNet	37.10/0.975	26.31/0.846	–	–
Ours	**37.33/0.982**	**28.51/0.901**	**27.54/0.865**	**30.92/0.927**

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, Rain100H and Rain100L images are provided for visual comparison. We selected a detail in the image and enlarged it. By observing the enlarged local area, although 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 pattern can be effectively removed while preserving the background details on the synthetic datasets.

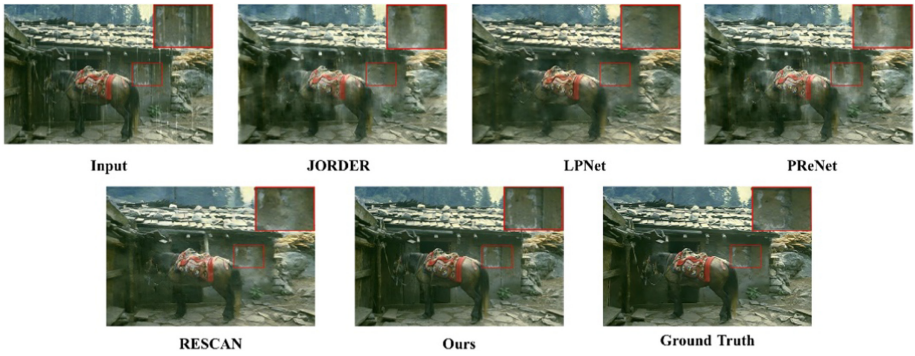

Fig. 4. Visual quality comparisons of all competing methods on synthetic dataset (Rain100H)

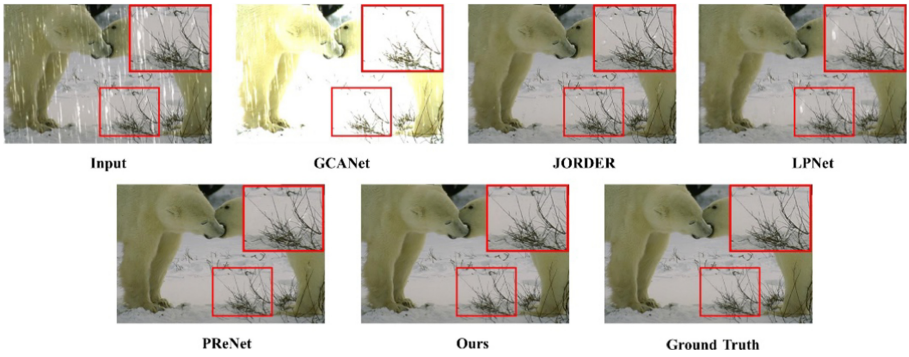

Fig. 5. Visual quality comparisons of all competing methods on synthetic dataset (Rain100L)

4.4 Results on Real-World Datasets

In order to evaluate the 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 Sect. 4.2 for further experimental evaluation. In order to compare the fairness, all methods use the weight of the pretraining 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 three 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 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.

4.5 Ablation Study

In order to prove the validity and rationality of the structure configuration and parameter setting in MACD-Net proposed in this paper, ablation experimental studies are

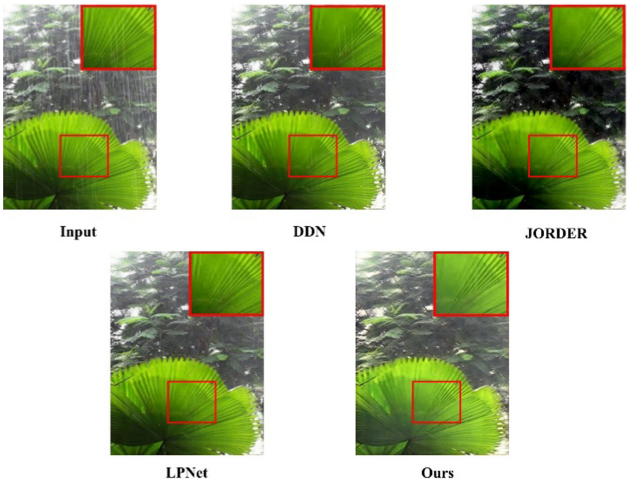

Fig. 6. Visual quality comparisons of all competing methods on Real-world dataset

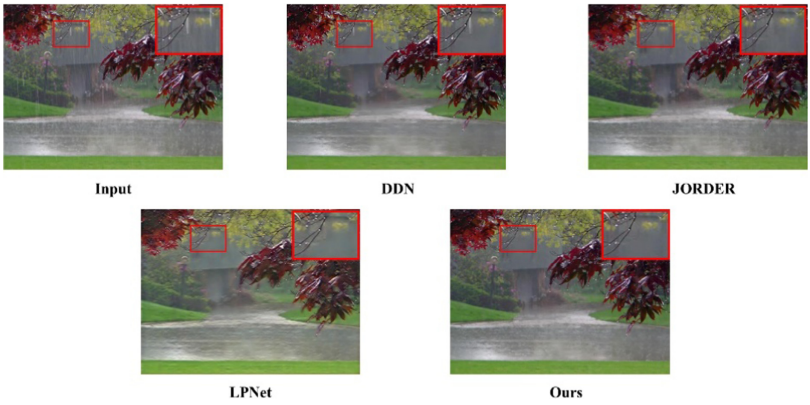

Fig. 7. Visual quality comparisons of all competing methods on Real-world dataset

conducted. All the studies involved use the Rain100L dataset and are guaranteed to be carried out in the same environment.

Ablation study on Multi-scale Coordinate Attention Block Numbers. In order to study the influence of different numbers of MCAB in MCAD-Net on network image rain removal. Experiments are conducted with Rain100L as the reference dataset, where the number of MCAB ranges from 2 to 10, and the experimental results are shown in Table 3.

It can be seen from Table 3 that in the ablation experiment of Multi-Scale Coordinate Attention Block Numbers, the PSNR and SSIM values can be improved as the number of MCAB is increased. When $n = 6$, the evaluation value reaches the highest, and

Table 3. Comparison of the test results of PSNR and SSIM on different MCAB numbers

MCAB no.	N = 2	N = 4	N = 6 (default)	N = 8	N = 10
PSNR	35.76	36.91	**37.33**	37.10	36.81
SSIM	0.977	0.980	**0.982**	0.981	0.979

then it begins to decline. Therefore, considering the balance between effectiveness and efficiency, we set the number of MCAB in the proposed network to 6 by default.

Ablation study on Coordinate Attention Modules. As introduced in Sect. 3.2, the coordinated attention module is introduced in MCAB. In order to further verify the effectiveness of the coordinated attention module, the basic multi-scale fusion network compared with the improved network with the attention module. As shown in Table 4, the coordinated attention module can bring improvements in PSNR and SSIM.

Table 4. Comparison of the test results of PSNR and SSIM on Baseline and Attention module

Methods	Baseline	Baseline + Coordinate Attention
PSNR	36.48	**37.33**
SSIM	0.979	**0.982**

Ablation Study on Loss Function. As mentioned in Sect. 3.3, this paper refers to the MSE and SSIM loss functions and proposes the Total loss function. In order to further study the influence of the MSE and SSIM loss functions on the network's image rain removal effect, we also did ablation experiments for comparison. As shown in Table 5, the quantitative evaluation of different loss functions under the same environment and parameter configuration. It can be found that when the MSE and SSIM loss functions are used alone, the MSE loss function makes the PSNR value higher, while the SSIM loss function focuses on the attack and structural similarity of the picture, which leads to a higher SSIM value. Therefore, comprehensively considering the quantitative combination of the MSE and SSIM loss functions, the Total loss function has achieved better results.

Table 5. Comparison of test results between PSNR and SSIM on different Loss Functions

Loss	MSE loss	SSIM loss	MSE+SSIM (Total loss)
PSNR	36.20	35.13	**37.33**
SSIM	0.968	0.9758	**0.982**

5 Conclusion

In this paper, a Multi-scale Coordinate Attention Dense Network (MCAD-Net) is proposed to remove rain from a single image. MCAD-Net is based on DenseNet and uses multiple skip connections to achieve feature reuse and full dissemination of feature information. In order to better identify and characterize the feature information of rain streaks, multi scale feature acquisition and coordinated attention module are introduced to form a Multi-scale Coordinated Attention Block (MCAB). First, get rain streaks features of different scales through convolution kernels of different sizes and fuse them. Then the attention module is used to further explore the rain streaks features in different channels and spaces. The rationality of module introduction and parameter setting is proved by ablation experiments. Finally, the proposed method is tested under common synthetic datasets and real datasets, and compared with the latest advanced methods, both showed better performance. In addition, the method proposed in this paper also better restores the details of the rainless image. In the future, we plan to extend our research work to video deraining under autonomous driving. For example, adding an additional module to our network to capture time information.

References

1. Yang, W., Tan, R.T., Feng, J., Guo, Z., Yan, S., Liu, J.: Joint rain detection and removal from a single image with contextualized deep networks. IEEE Trans. Pattern Anal. Mach. Intell. **42**, 1377–1393 (2020)
2. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: 2017 IEEE Conference on ComputerVision and Pattern Recognition (CVPR), pp. 1715–1723 (2017)
3. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detectionand removal from a single image. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1685–1694 (2017)
4. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scaleimage recognition. In: ICLR 2015: International Conference on Learning Representations 2015 (2015)
5. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. **5**, 564–575 (2003)
6. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 1254–1259 (1998)
7. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of localfeatures. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2360–2367 (2010)
8. Zhang, X., Li, H., Qi, Y., Leow, W., Ng, T.: Rain removal in video by combiningtemporal and chromatic properties. In: 2006 IEEE International Conference on Multimedia and Expo, pp. 461–464 (2006)
9. Luo, Y., Xu, Y., Ji, H.: Removing rain from a single image via discriminative sparsecoding. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 3397–3405 (2015)
10. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layerpriors. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2736–2744 (2016)

11. Gu, S., Meng, D., Zuo, W., Zhang, L.: Joint convolutional analysis and synthesissparse representation for single image layer separation. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 1717–1725 (2017)
12. Li, S., et al.: Single image deraining: a comprehensivebenchmark analysis. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3838–3847 (2019)
13. Xu, J., Zhao, W., Liu, P., Tang, X.: Removing rain and snow in asingle image using guided filter. In: IEEE International Conference on Computer Science and Automation Engineering (CSAE), vol. 2, pp. 304–307 (2012)
14. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1397–1409 (2012)
15. Xu, J., Zhao, W., Liu, P., Tang, X.: An improved guidance image based method to remove rain and snow in a single image. Comput. Inf. Sci. **5**(3), 49 (2012)
16. Zheng, X., Liao, Y., Guo, W., Fu, X., Ding, X.: Single-image-based rain and snow removal using multi-guided filter. In: Lee, M., Hirose, A., Hou, Z.-G., Kil, R.M. (eds.) ICONIP 2013. LNCS, vol. 8228, pp. 258–265. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-42051-1_33
17. Zhang, H., Patel, V.M.: Convolutional sparse and low-rank coding-based rainstreak removal. In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1259–1267 (2017)
18. Chen, Y.L., Hsu, C.T.: A generalized low-rank appearance model for spatio temporally corelated rain streaks. In: 2013 IEEE International Conference onComputer Vision, pp. 1968–1975 (2013)
19. Yang, W., Tan, R.T., Wang, S., Fang, Y., Liu, J.: Single image deraining: from model-based to data-driven and beyond. IEEE Trans. Pattern Anal. Mach. Intell. **43**, 4059–4077 (2020)
20. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in NeuralInformation Processing Systems, vol. 27, pp. 2672–2680 (2014)
21. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks:a better and simpler baseline. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3937–3946 (2019)
22. 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 (ECCV), pp. 262–277 (2018)
23. Yang, W., Liu, J., Yang, S., Guo, Z.: Scale-free single image deraining via visibility enhanced recurrent wavelet learning. IEEE Trans. Image Process. **28**, 2948–2961 (2019)
24. Mu, P., Chen, J., Liu, R., Fan, X., Luo, Z.: Learning bilevel layerpriors for single image rain streaks removal. IEEE Signal Process. Lett. **26**(2), 307–311 (2019)
25. Mu, P., Chen, J., Liu, R., Fan, X., Luo, Z.: Learning bilevel layer priors for single image rain streaks removal. IEEE Signal Process. Lett. **26**(2), 307–311 (2019)
26. Gauvain, J.-L., Lee, C.-H.: Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. IEEE Trans. Speech Audio Process. **2**(2), 291–298 (1994)
27. Fu, Y.-H., Kang, L.-W., Lin, C.-W., Hsu, C.-T.: Single-frame-basedrain removal via image decomposition. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2011)
28. Qian, R., Tan, R.T., Yang, W., Su, J., Liu, J.: Attentive generativeadversarial network for raindrop removal from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2482–2491 (2018)
29. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for imagerecognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, pp. 1453–1456 (2016)

30. Pan, J., et al.: Learning dual convolutional neural networks for low-level vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3070–3079 (2018)
31. Wang, Y.-T., Zhao, X.-L., Jiang, T.-X., Deng, L.-J., Chang, Y., Huang, T.-Z.: Rain streak removal for single image via kernel guided CNN. arXiv preprint arXiv:1808.08545 (2018)
32. Li, G., He, X., Zhang, W., Chang, H., Dong, L., Lin, L.: Non-locally enhanced encoder-decoder network for single image de-raining. arXiv preprint arXiv:1808.01491 (2018)
33. Ren, D., Zuo, W., Hu, Q., Zhu, P., Meng, D.: Progressive image deraining networks: a better and simpler baseline. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3937–3946 (2019)
34. Chen, X., Huang, Y., Xu, L.: Multi-scale hourglass hierarchical fusion network for single image deraining. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2021)
35. Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design (2021)
36. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)
37. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. IEEE Trans. Circuits Syst. Video Technol. **30**, 3943–3956 (2017)
38. Wang, H., Li, M., Wu, Y., Zhao, Q., Meng, D.: A survey on rain removal from video and single image. arXiv preprint arXiv:1909.08326 (2019)
39. Chen, D., et al.: Gatedcontext aggregation network for image dehazing and deraining. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1375–1383 (2019)
40. Fu, X., Liang, B., Huang, Y., Ding, X., Paisley, J.: Lightweight pyramid networks for image deraining. IEEE Trans. Neural Netw. **31**, 1794–1807 (2020)
41. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 695–704 (2018)
42. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentivesingle-image deraining with a high quality real rain dataset. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12270–12279 (2019)
43. Yang, Y., Lu, H.: Single image deraining via recurrent hierarchy enhancement network. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 1814–1822. ACM (2019)
44. Wang, H., Xie, Q., Zhao, Q., Meng, D.: 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 (2020)
45. Mehri, A., Ardakani, P., Sappa, A.: MPRNet: multi-path residual network for lightweight image super resolution (2020)