M.S. THESIS

MC-CNN: Multi-scale Connected Convolutional Neural Network for Single Image Deraining

단일 이미지 내 비제거를 위한 다중스케일 연결 합성곱 신경망

BY

JUNHO LEE

August 2021

Interdisciplinary Program
in Computational Science and Technology
Seoul National University

MC-CNN: Multi-scale Connected Convolutional Neural Network for Single Image Deraining

단일 이미지 내 비제거를 위한 다중스케일 연결 합성곱 신경망

지도교수 강 명 주

이 논문을 이학석사 학위논문으로 제출함

2021년 4월

서울대학교 대학원

협동과정 계산과학전공

이 준 호

이준호의 이학석사 학위 논문을 인준함

2021년 5월

위 원 장: [redacted]

부위원장: [redacted]

위 원: [redacted]

MC-CNN: Multi-scale Connected Convolutional Neural Network for Single Image Deraining

A thesis
submitted in partial fulfillment
of the requirements for the degree of
Master of Science
to the faculty of the Graduate School of
Seoul National University

by

JUNHO LEE

Thesis Director : Professor Myungjoo Kang

Interdisciplinary Program
in Computational Science and Technology
Seoul National University

August 2021

Abstract

In this thesis, we propose an end-to-end multi-scale connected convolutional neural network (MC-CNN) that leverages all scale features to remove rain streaks while recovering detailed information on images. The first key point for recovering details is a multi-scale connection, which connects all scale features of the encoder part to the decoder part to restore the image with as much information as possible. Multi-scale connection considers channel-wise attention to learn which scale features are important in the current process, rather than simply combining the features of each scale. The second key point is a wide regional non-local (WRNL) block. We find that dividing images into wide rectangular patches makes each patch have a more even distribution than the existing method and based on this, we propose a WRNL block. Experimental results on synthetic and real-world datasets demonstrate that MC-CNN quantitatively outperforms existing state-of-the-art models and qualitatively achieves several improvements.

keywords: Convolutional neural network, deraining, deep Learning, image-preprocessing, rain

student number: 2019-28867

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Adverse weather conditions such as rain, haze, and snow, are capable of producing complex visual effects on natural images or videos. In particular, rain streaks, which comprise one of the most commonly occurring phenomena in outdoor imaging, are capable of potentially degrading performance in several computer vision applications. Therefore, it is imperative to develop algorithms that effectively remove rain streaks and restore the pristine background scenes in the context of vision-related tasks.

Over the past few decades, several pieces of research have been dedicated to the removal of rain streaks from captured images. Several traditional deraining methods [2, 1, 4, 13, 19, 17] have been suggested to separate rain streaks from a clean background image based on the physical characteristics or texture appearance pattern of rain streaks. Recently, motivated by the unprecedented success of deep learning in low-level vision, image deraining has made rapid progress with convolutional neural network (CNN)-based methods [7, 8, 14, 15, 16, 22, 23, 24, 29, 32, 36, 37, 42, 45].

Among those, there were several attempts to increase performance with encoder-decoder based structure. For example, to remove fine-grained rain streaks and recover rain-free backgrounds more clearly, Yu *et al.* [42] proposed a two-stage model using encoder-decoder as a coarse deraining stage and a simple network as a fine deraining stage. And Wang *et al.* [29] added a residual learning branch parallel to the encoder part to form a better conditional embedding and eventually generate a much better deraining result in the decoder part. Both methods have achieved notable performance improvements, but there is a limitation that they have improved their performance through additional branches without fully exploiting all the information generated within the U-Net structure.

In order to obtain information on the degraded background from other pixels through spatial attention, Li *et al.* [14] applied non-local block. Because the original non-local block is inefficient with too many computations, Yu *el al.* [42] used a regional non-local operation, which divides images into grids and applies a non-local block to each patch. However, the regional non-local operation was originally designed for the denoising task and therefore never considered the characteristics of rain. Consequently, these methods have difficulties recovering details, especially under extremely adverse weather conditions.

To address the above-mentioned issues, we present a multi-scale connected convolutional neural network (MC-CNN) to carefully remove rain streaks and recover background details leveraging multi-scale features and adaptive non-local operation considering the characteristics of the rain streaks.

Recent deraining papers show slightly complex structures, such as recurrent model [36, 16, 3, 40], multiple inputs model [12], or adding branches in parallel to the main network [29, 6] to achieve better results. However, MC-CNN achieves state-of-the-art with some proposed methods without deviating from the encoder-decoder structure with a single input.

Inspired by [25, 26, 31], we propose a multi-scale connection to efficiently leverage information on various scales in the decoding process. To learn which scale is more important in the decoding process of each scale, multi-scale connection is designed to consider channel attention. Unlike other tasks such as human pose estimation and semantic segmentation, multiple connections without channel attention rather cause performance degradation in the deraining task. After many attempts to optimize multiple connections for the deraining task, we find that considering channel attention is an important point and devise multi-scale connection through it. In Table 4.4 and Figure 4.4, we show that multi-scale connection plays an effective role by comparing the qualitative and quantitative results of models with and without multi-scale connection.

Next, based on a statistical analysis of the distribution of rain pixels in rainy images, we also propose a wide regional non-local (WRNL) block, an adaptive regional non-local block for the deraining task. By analyzing rain pixel distributions over different patch shapes, we find that rain pixel distributions are most uniformly distributed when images are divided into wide rectangular patches (see Fig.3.3). When rain pixels are evenly distributed on each patch, background information is also evenly distributed, leading to overall performance improvement as information-poor patches disappear.

MC-CNN is evaluated on four synthetic and two real-world deraining datasets and compares its performance with those of existing state-of-the-art methods. In summary, the contribution of this thesis may be summarized as follows.

1) We propose multi-scale connection, multiple connections for the deraining task, to ensure that the model utilizes as much information as possible in the decoding process. At each stage of the decoder part, feature information of all the scales in the encoder part is aggregated. By considering channel attention after concatenating all scales of feature, we effectively aggregate different scale characteristics.

2) We propose the WRNL block, which supports the model to effectively restore the background by providing more sufficient rain-free information in each region than the original regional non-local block.

3) We perform experiments on both synthetic and real-world rain datasets and show that the proposed method significantly outperforms existing state-of-the-art methods.

Chapter 2

Related Work

The single image deraining problem begins with the assumption that a rainy image consists of a background layer and a rainy layer. Several traditional training methods based on single images and videos have been proposed. Barnum *et al.* [1] reconstruct rainy images by combining the appearance model with the streak model. The appearance model identifies individual rain streaks and the streak model utilizes the statistical characteristics of rain. Chen and Hsu [4] use the low-rank model to separate the layers in a rainy image. As noted by Yang *et al.* [38], sparse coding is applied during this process to separate the rainy layer from the rainy image [5, 13, 19, 33, 47]. Further, Li *et al.* [2, 17] approach this problem using the Gaussian mixture model.

Because of the remarkable performance exhibited by deep learning-based methods, especially CNN-based ones, the potential use of deep learning in deraining has been extensively researched. Yang *et al.* [37] apply a CNN-based method for the first time and express natural images by adding atmospheric light as a component to rainy images. Fu *et al.* [8] and Fan *et al.* [7] use a single primary network that restores input images using the residual network. Based on the residual network, Li *et al.* [16] attempt to further eliminate overlapping rain streaks by organizing the context aggregate network into multiple stages. Shen *et al.* [24] consider rain streaks to be high-frequency and attempt to remove rain streaks by utilizing DWT. Yang *et al.* [36] divide the deraining process into several stages and reconstruct the image recurrently, beginning with a small portion of the image to eventually obtain the entire image.

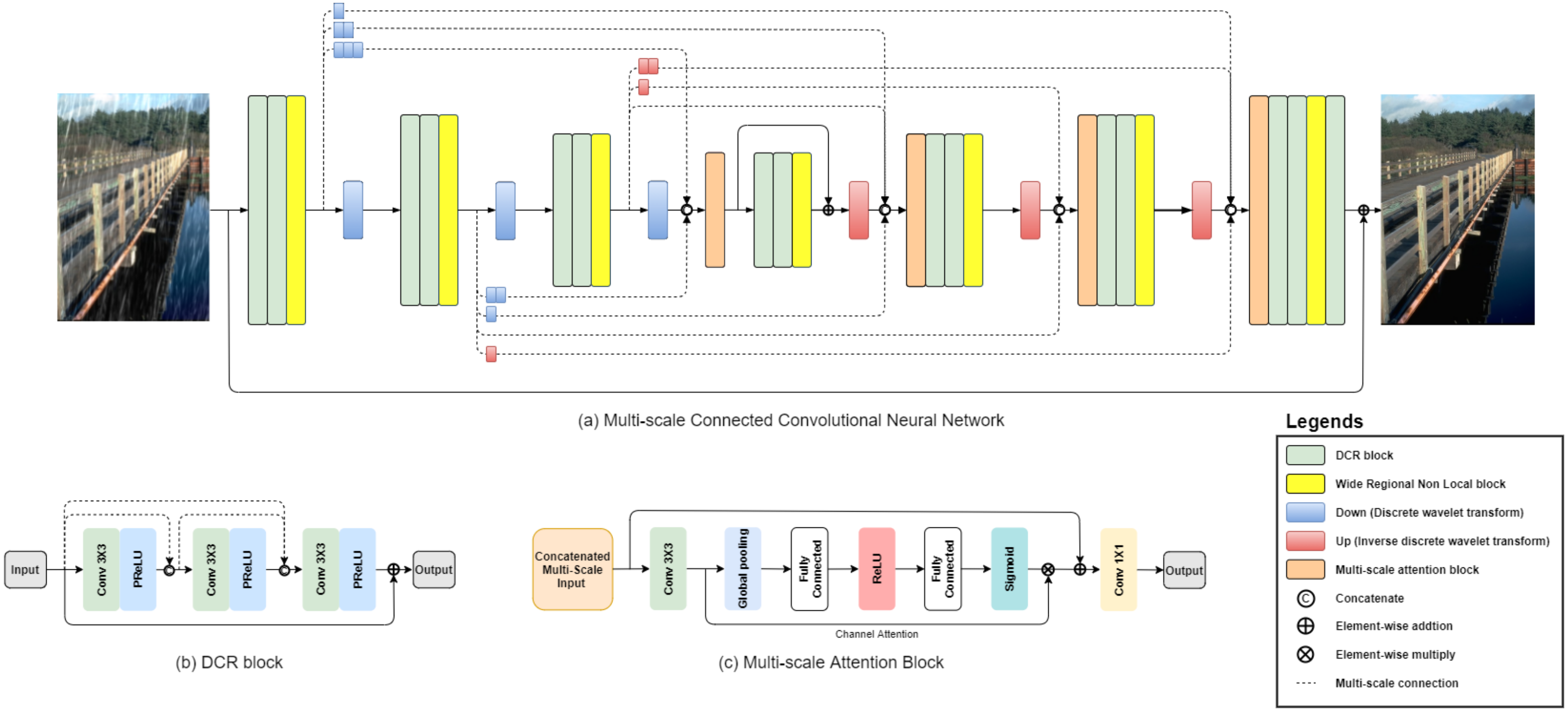
Wang *et al.* [32] capture the spatial contextual information using a four-directional recurrent neural network with the identity matrix initialization model. Ren *et al.* [23] propose progressive ResNet to effectively remove the rain via recursive computation. Yu *et al.* [42] propose GraNet, which is designed to identify rain masks in the coarse stage using a region-aware non-local block. Subsequently, the process uses the rain masks to create the final image using another reconstruction network. To achieve pixel-wise deraining in image recovery, encoder-decoder structures have been used in certain methods. Wang *et al.* [29] propose the residual learning branch as a component of the encoder. Li *et al.* [14] enhance the performance by introducing non-local blocks into the encoder-decoder network. Among the methods that reconstruct the rainy layer to be identical to the background layer, the generative adversarial network is widely used to remove raindrops and rain streaks [15, 22, 45].

Yang *et al.* [39] propose the fractal band learning network based on frequent band recovery. Wang *et al.* [30] propose an interpretable deep network based on a convolutional dictionary network. Jiang *et al.* [12] use the images of various sizes as the input to the model. A multi-scale pyramid structure is used to promote cooperative representation. Deng *et al.* [6] propose two-branch parallel networks, in which one branch performs rain removal and the other branch detail recovery. In [34], newly formulated rain streaks transmission maps, vapor transmission maps, and atmospheric lights are respectively learned by three different networks. Zhang *et al.* [46] propose a paired rain removal network, which exploits both stereo images and semantic information.

Chapter 3

Proposed Network

In this chapter, we describe the overall structure and main components of the proposed MC-CNN. The overview of MC-CNN is depicted in Figure 3.1. In Figure 3.1, let the clustered blocks used within one scale be called stages. The first three stages constitute the encoder part and the other four stages constitute the decoder part. Multi-scale connection connects all output of all encoder parts to all inputs of the decoder. The output of each encoder is concatenated and processed through a multi-scale attention block before entering the input of the decoder. Multi-scale attention blocks serve to change the concatenated feature of all scales to be useful for the model. Each stage of MC-CNN is composed of two densely connected residual (DCR) blocks [20], each of which consists of three convolution layers followed by PReLU [27] (refer toFigure 3.1(b)) and one WRNL block.



(a) Multi-scale Connected Convolutional Neural Network

(b) DCR block

(c) Multi-scale Attention Block

Figure 3.1: Illustration of the proposed network. From (a)-(c): (a) overview of MC-CNN, (b) densely connected residual (DCR) block, and (c) multi-scale attention block.

3.1 Multi-scale Connection

In a typical U-Net-like network, connections exist between features corresponding to the same scale. Although it performs better than the encoder-decoder without connections, they result in missing information in that all features generated in the encoding process are not utilized in the decoding process. Such information loss is undesirable because a single image deraining task is a low-level vision task that attempts to restore each pixel more accurately. Multi-scale connection has been proposed to minimize this loss of information.

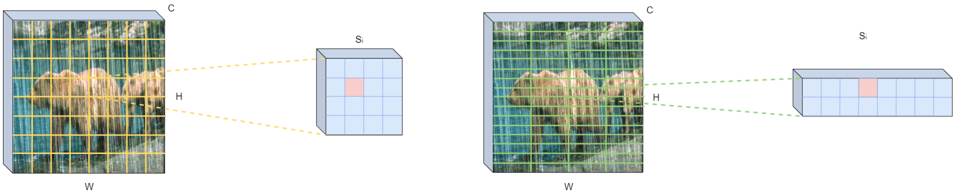
Formally, let E^l_{out} be the output features at level l $(l = 1, 2, 3)$ in the encoder part. At each level l $(l = 1, 2, 3, 4)$ in the decoder part, the input feature D^l_{in} is given as:

$$D^l_{concat} = (\bigoplus_{i=1}^{3} H^l_i(E^i_{out})) \oplus H_{up}(D^{l+1}_{out}), \tag{3.1}$$

$$D^l_{in} = f_{MAB}(D^l_{concat}), \tag{3.2}$$

where $\oplus$ denotes the concatenation operation, $H_{up}(\cdot)$ denotes the up-sampling operation, D^l_{out} denotes the output feature of the decoder part at level l, and $f_{MAB}(\cdot)$ denotes the multi-scale attention block depicted in Figure 3.1(c). $H^l_i(\cdot)$ denotes the sampling operation from level i to l. In other words, H^l_i is the down-sampling by $l - i$ times, identity, and up-sampling by $i - l$ times operations if $l > i$, $l = i$, and $l < i$, respectively. We set $D^5_{in} = 0$ for convenience.

Multi-scale connection is designed to consider channel attention to learn which scale is more important in the decoding process of each scale. In contrast to tasks such as human pose estimation and semantic segmentation, which demonstrated the effectiveness of multiple connections between scales in [25], multiple connections between scales rather cause performance degradation in the deraining tasks. Through several experiments, we find that channel attention is required for multiple connections between scales to be effectively used for the deraining task. To design multi-scale connection to consider channel-wise attention of concatenated multi-scale input, we tried



(a) (b)

Figure 3.2: Examples of patch shapes according to region type. (a) is square patch and (b) is wide rectangular patch. Every pixel in a patch refers to every pixel in the patch.

multi-head attention [28] and squeeze-and-excitation (SE) block [10] respectively, and we adopt SE block with better results. To find the correct correspondence between features at different scales, discrete wavelet transforms (DWT or IWT) is applied.

3.2 Wide Regional Non-Local Block

In this section, we first describe the representation of the WRNL block and then provide an analysis of the effectiveness of the WRNL block based on statistical exploration.

Formally, let denote the input feature to the WRNL block as $X \in \mathbb{R}^{H\times W\times C}$. WRNL block divides X into a $a \times b$ grid of patches $\{X^k\}$, $(k = 1, ..., K = ab)$ where K is the number of patches. The grid division is illustrated in Figure 3.2. The linear embedding processes for X^k to generate the output Z^k are formulated as follows.

$$\Phi(X^k)_i^j = \phi(X_i^k, X_j^k) = \exp\{\theta(X_i^k)\psi(X_j^k)^T\}, \tag{3.3}$$

$$\theta(X_i^k) = X_i^k W_\theta, \psi(X_i^k) = X_i^k W_\psi, G(X)_i^k = X_i^k W_g, \tag{3.4}$$

where X_i^k and X_j^k denote the feature X^k at position i and j, respectively. The learnable weight matrices W_θ, W_ϕ, and W_g have the dimensions of $C \times L$, $C \times L$, and $C \times C$, respectively. In practice, $L = C/2$ is used. The regional non-local operation can be expressed as follows:

$$Z_i^k = \frac{1}{\delta_i(X^k)} \sum\nolimits_{j\in S_i} \Phi(X^k)_i^j \, G(X^k)_i \,, \;\; \forall i \,, \tag{3.5}$$

where $\delta_i(X^k) = \sum_{j \in S_i} \phi(X_i^k, X_j^k)$ denotes the correlation between X_i^k and each X_j^k in S_i, and Z_i^k denotes the output feature Z^k at position i. S_i denotes a set of patch positions. If $a > b$, then the patch is wider than when $a = b$. Therefore, we call the patch a wide rectangular patch, a square patch, and a tall rectangular patch if $a > b$, $a = b$, and $a < b$, respectively. In the WRNL block, we set the $a \times b$ grids to 16 × 4 , 8 × 2, 4 × 1, and 4 × 1 at levels 1, 2, 3, and 4, respectively.

3.2.1 Analysis

Each patch should have sufficient background information in that non-local blocks recover certain pixels based on information from other pixels in the patch. Therefore, if the background information is distributed evenly on each patch, it can be expected that the regional non-local block will restore the image globally well. However, we find that the rain pixels are not evenly distributed between square patches in the images used in the previous deraining research [14, 42]. Since the rain steaks are mostly vertical, wide rectangular patches can be expected to distribute more evenly between patches than square and wide rectangular patches.

To check the distribution of rain pixels in each patch, we analyze 2 synthetic datasets (Rain200L, Rain200H) and one real-world dataset (SPA-data). To match the number of pixels per patch of grid divisions, we divide the height and width of the image into 16 ×4, 8 ×8, 4 × 16 grids, respectively, to create wide, square, and tall rectangular patches (see Figure 3.2). We define pixels as rain if the difference between the pixels in x_{input} and x_{gt} exceeds a certain threshold. The standard deviation between the number of rain pixels in the patches included in each image is depicted in Figure 3.3. Wide rectangular patches are observed to exhibit much smaller average standard deviation values compared to square and tall rectangular patches, which implies an even distribution of rain across all patches. This results in the effective recovery of the image because the usable background information within each patch is also distributed evenly.

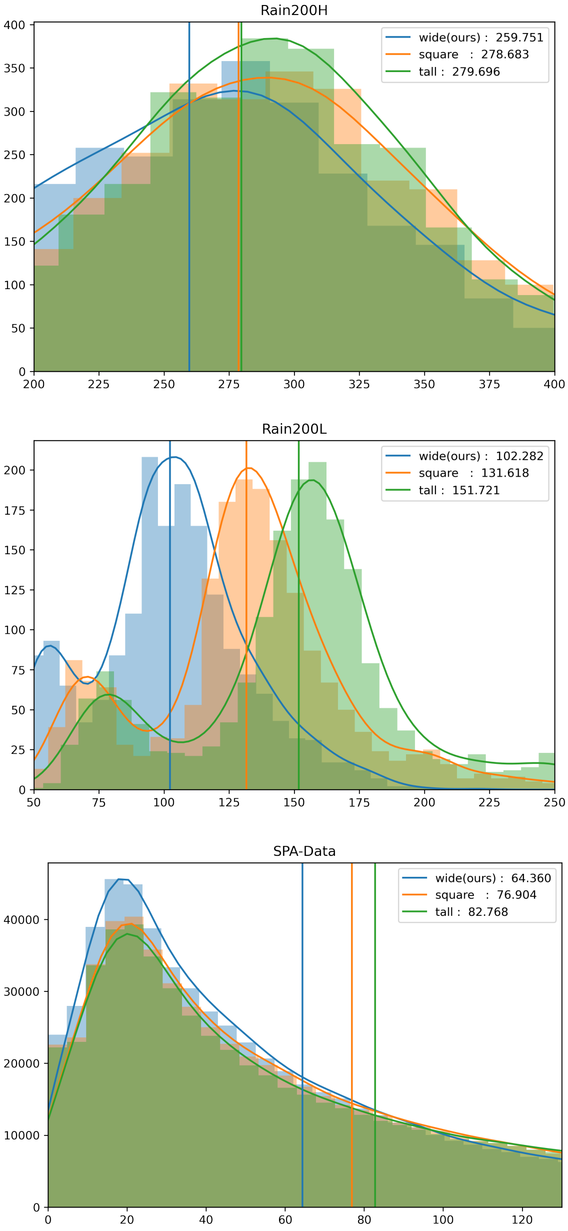


Figure 3.3: Rain pixel distributions across various region types. Rain200H, Rain200L, and SPA-data are analyzed. The x-axis represents the standard deviation between rain pixels per patch in each image. The y-axis represents the number of images. The distribution of the images according to the standard deviation is represented by histograms. We approximate the probability density function of the histogram.

3.3 Discrete Wavelet Transform

MC-CNN use DWT and IWT for down-sampling and up-sampling, respectively. In our implementation, Haar wavelet is adopted, which is simple and widely used method in image processing [9, 18, 21, 24, 36]. In 2D Haar wavelet, four filters, $\mathbf{f}_{LL}$, $\mathbf{f}_{LH}$, $\mathbf{f}_{HL}$ and $\mathbf{f}_{HH}$, are fined as,

$$\mathbf{f}_{LL} = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathbf{f}_{LH} = \frac{1}{4}\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \mathbf{f}_{HL} = \frac{1}{4}\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \mathbf{f}_{HH} = \frac{1}{4}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{3.6}$$

Let LL, LH, HL, and HH be images generated by $\mathbf{f}_{LL}$, $\mathbf{f}_{LH}$, $\mathbf{f}_{HL}$, and $\mathbf{f}_{HH}$ filters respectively. Given that $\mathbf{f}_{LL}$ is the same as average pooling, LL achieves the local translation invariance by reducing the size of the feature map (see Equation 3.6). LH, HL, and HH have edge information. In particular, since LH has vertical edge information, the feature of rain streaks can be effectively obtained.

The IWT operation in the up-sampling process is written as:

$$\begin{aligned} a &= LL + LH + HL + HH, \\ b &= LL - LH + HL - HH, \\ c &= LL + LH - HL - HH, \\ d &= LL - LH - HL + HH, \end{aligned} \tag{3.7}$$

where a, b, c, d are four pixels in every 2×2 block.

3.4 Data Augmentation

In addition to the commonly adopted data augmentation techniques such as random cropping, we opt CutMix [43] augmentation strategy which cuts and pastes ground-truth patches to input images. In visual recognition, CutMix allows the model to use pixels efficiently in training and obtain the regularization effect. Yoo et al. [41] show that this cut-and-paste approach is also useful in low-level vision tasks such as image super-resolution task. To detail CutMix augmentation, let $x_{input}, x_{gt} \in \mathbb{R}^{W \times H \times C}$ be input and ground-truth images. In the deraining task, x_{input} and x_{gt} are rainy images and rain-free images. We perform the cut-and-paste operation as :

$$\tilde{x} = \mathbf{M} \odot x_{input} + (\mathbf{1} - \mathbf{M}) \odot x_{gt}, \tag{3.8}$$

where $\tilde{x}$ denotes the augmented sample, $\mathbf{M} \in \{0, 1\}^{W \times H}$ denote the binary mask indicating where to replace, $\mathbf{1}$ denotes the binary mask filled with ones and $\odot$ denotes the element-wise multiplication. We randomly sample the size of the binary mask $\mathbf{M}$ not more than half of the input image.

3.5 Loss Function

We use L_1, L_2 hybrid loss function because it showed slightly better performance, but our method doesn't appear to be sensitive to loss. Total loss, standard L_1 and L_2 losses are defined as follows.

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_1 + \mathcal{L}_2, \\ \mathcal{L}_1 &= \|x_{gt} - f_{MC-CNN}(x_{input})\|_1, \\ \mathcal{L}_2 &= \|x_{gt} - f_{MC-CNN}(x_{input})\|_2, \end{aligned} \tag{3.9}$$

where x_{input} denotes input rainy image and x_{gt} denotes the corresponding rain-free image and f_{MC-CNN} is a function that denotes the return of the HF-Net output with respect to x_{input}.

Chapter 4

Experiments

In this chapter, we introduce datasets we used, the evaluation method, and the experimental environment, and then demonstrate experimental results. The experimental results are evaluated qualitatively and quantitatively, and the results of other state-of-the-art methods are also compared. Then we conduct ablation studies to verify the main components of our methods introduced in Chapter 3.

4.1 Datasets and Evaluation Metrics

Datasets	Train images	Test images	Data Type	Training Epoch
Rain200L [37]	1,800	200	synthetic	200
Rain200H [37]	1,800	200	synthetic	200
Rain800 [45]	700	100	synthetic	200
Rain1200 [44]	12,000	1,200	synthetic	80
SPA-data [32]	640k	1,000	real-world	3
Yang *et al.* [37]	-	15	real-world	-

Table 4.1: Synthetic and real-world datasets

Four synthetic datasets, *i.e.*, Rain200L [37], Rain200H [37], Rain800 [45], and

Rain1200 [44], and two real-world datasets, *i.e.*, SPA-data [32] and Yang *et al.* [37], are used to evaluate the performance of the proposed method. Details of the datasets are given in Table 4.1 As pointed out by Ren *et al.* [23], certain overlaps of background exist between the training dataset and the test dataset in the Rain100H and Rain100L datasets. Therefore, new test datasets *i.e.*, Rain200H and Rain200L, which do not share the backgrounds with the corresponding training datasets are updated by Yang *et al.* [37]. We strictly evaluate our model using Rain200H and Rain200L.

Because of the unavailability of training images and ground-truth images for test input images in the real-world dataset of Yang *et al.* [37], evaluation is only performed qualitatively on the Yang *et al.* dataset using Rain200H-trained weights.

We use peak single-to-noise ratio (PSNR) [11] and structural similarity index (SSIM) [35] as evaluation metrics to compare the performance of our proposed model with those of other state-of-the-art methods. For equal evaluation, we calculate the PSNR and SSIM in the RGB color space instead of the luminance channel of the YCbCr space.

4.2 Experiment Details

Hardware	**Specification**
CPU	Intel Core i7-9700K
GPU	Titan RTX
Software	**Specification**
OS	Ubuntu 16.04.6 LTS
Python	2.7.12
Pytorch	Version: 1.2.0
CUDA	Version: 10.0

Table 4.2: Training Environment

Details of the training environment are in Table 4.2. Adam optimizer is used for model optimization, and we set the batch size to 4. We used the basic data augmentation methods random cropping, horizontal flipping, and additional advanced data augmentation, CutMix. The patch size for random cropping is set to 256×256. The training epochs are set differently for each dataset and are described in Table 4.1.

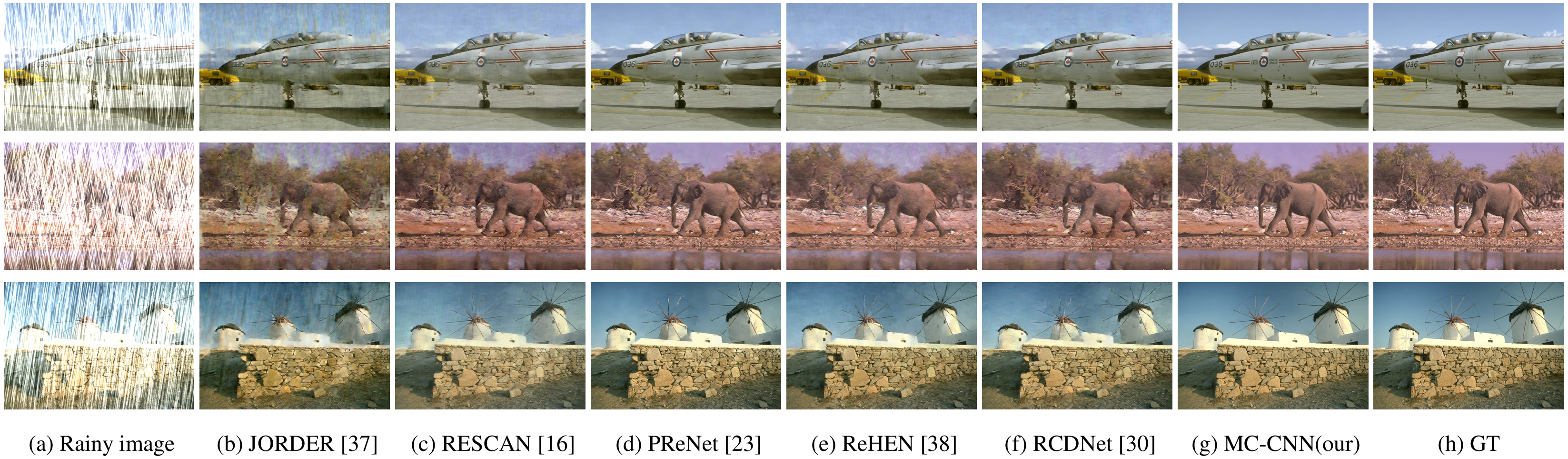
4.3 Results

4.3.1 Synthetic Datasets

The proposed MC-CNN is evaluated on four synthetic datasets [37, 45, 32] and its performance is compared to six state-of-the-art methods [37, 16, 23, 38, 30, 6]. The quantitative results for synthetic datasets are shown in Table 4.3. As can be seen from the data, the proposed MC-CNN achieved a significant improvement over existing state-of-the-art methods for PSNR and SSIM metrics. The original input, ground truth, and qualitative results on the Rain200H are depicted in Figure 4.1. In Figure 4.1, other methods also capture and remove the rain streaks well, but they are lacking by leaving stains or losing detailed background information in the process of removing. The proposed MC-CNN model also does not completely remove and restore all rain streaks, but it significantly improves performance compared to other methods and restores them almost close to ground truth data.

Table 4.3: Average PSNR and SSIM comparison on Rain200H, Rain200L, Rain800, Rain1200, and SPA-data. The highest values are indicated in **bold**. The results confirm that our MC-CNN performs best in quantitative evaluations with PSNR and SSIM metrics

Method	JORDER [37]	RESCAN [16]	PReNet [23]	ReHEN [38]	RCDNet [30]	DRD-Net [6]	MC-CNN
	(CVPR' 2017)	(ECCV' 2018)	(CVPR' 2019)	(MM' 2019)	(CVPR 2020)	(CVPR' 2020)	(ours)
Rain200L [37]	36.95/0.979	36.94/0.980	36.28/0.979	38.57/0.983	35.28/0.971	37.15/0.987	**39.73/0.988**
Rain200H [37]	22.05/0.727	26.62/0.841	27.64/0.884	27.48/0.863	26.18/0.835	28.16/0.920	**30.70/0.922**
Rain800 [45]	22.24/0.776	24.09/0.841	22.83/0.790	26.96/0.854	24.59/0.821	26.32/**0.902**	**28.42**/0.876
Rain1200 [44]	24.32/0.862	32.48/0.910	30.40/0.891	32.64/0.914	33.54/0.913	-	**33.70/0.928**
SPA-data [32]	35.72/0.978	36.99/0.967	35.68/0.942	38.65/0.974	41.47/0.983	-	**46.88/0.991**

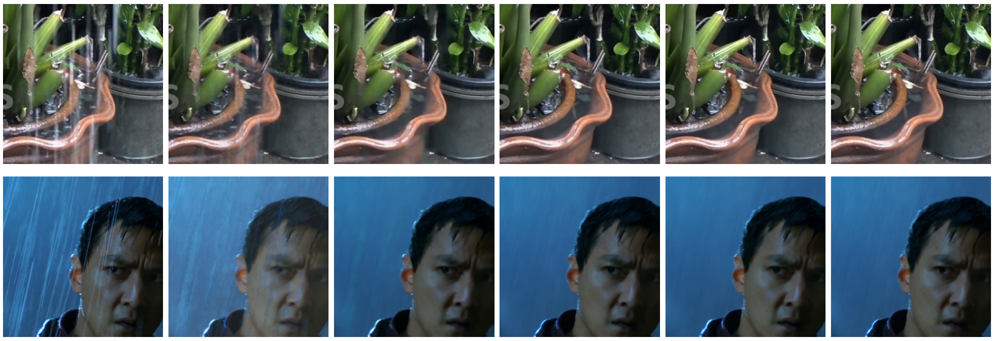


(a) Rainy image (b) JORDER [37] (c) RESCAN [16] (d) PReNet [23] (e) ReHEN [38] (f) RCDNet [30] (g) MC-CNN(our) (h) GT

Figure 4.1: Results images obtained via several state-of-the-art methods on Rain200H dataset for qualitative comparison.

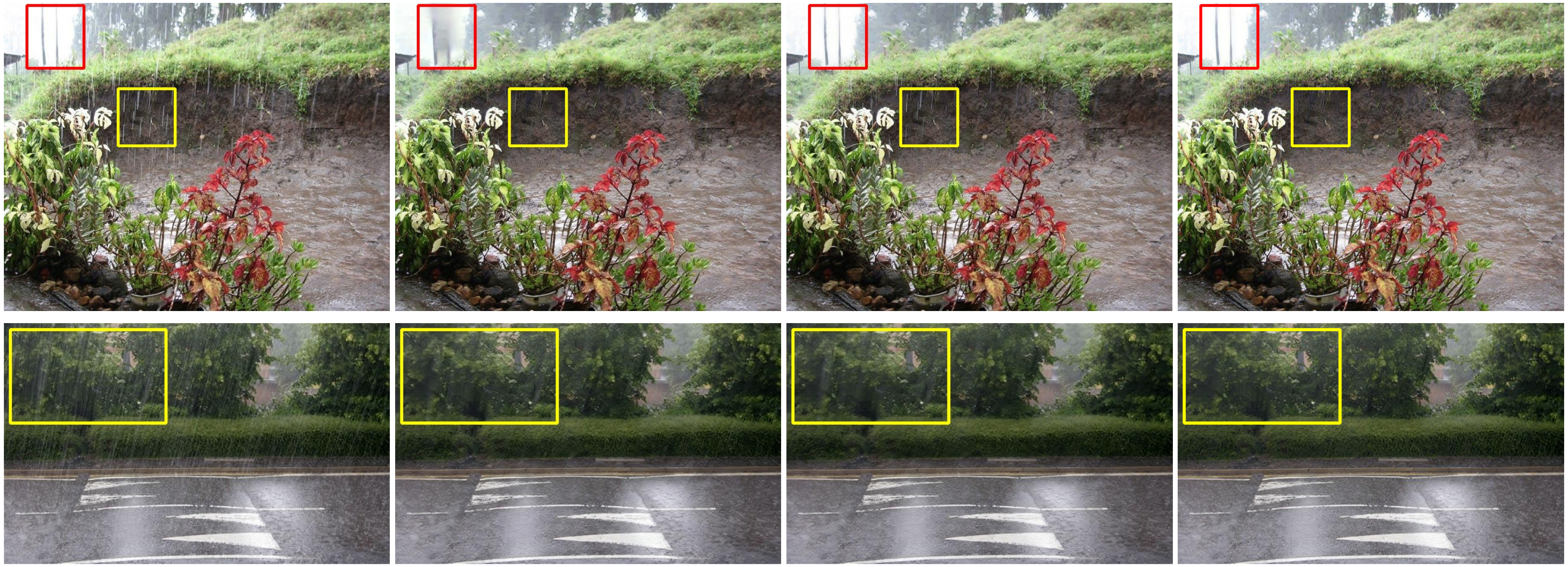
4.3.2 Real-world Datasets

To verify the effectiveness of the model on real-world situations, experiments are also conducted on two real-world datasets [37, 32]. Quantitative evaluation is made only in SPA-data, because only SPA-data has ground truth data among the real-world datasets we experimented on. As shown in Table 4.3, MC-CNN exhibits superior performance with a very large difference quantitatively compared to the rest of the state-of-the-art methods [37, 16, 23, 38, 30] in SPA-data. This shows that our model is not a model that works specifically on synthetic datasets, but rather is more suitable for removing real rain streaks.



(a) Input (b) RESCAN (d) PReNet (e) ReHEN (f) HF-Net (our) (g) GT

Figure 4.2: Results obtained via several different methods on SPA-data [32] images. The outputs of ReHEN [23] and the proposed HF-Net exhibit almost no traces of rain streaks on all two image samples, while the results obtained via other methods [16, 23, 38] exhibit traces of rain streaks. The derained image obtained via the proposed model demonstrates its effectiveness in removing rain streaks that are not even clearly found in the ground truth data.



(a)Rainy image

(b) PReNet [23]

(c) ReHEN [38]

(d) MC-CNN (our)

Figure 4.3: Results images obtained via several different methods on real images. This qualitative comparison compares only three methods with high PSNR values on the Rain200H dataset. For fairness, all methods are trained only with Rain200H training image pairs. Looking closely at the yellow box parts of the two images, we can see that the proposed method removes the rain better than the other methods. Then, looking at the red box, we can see that MC-CNN removes the rain while maintaining the details well.

In order to visually check whether rain streaks are derained well, we conduct a qualitative evaluation in SPA-data and Yang *et al.* dataset. As can be seen in Figure 4.2, we can see that our model removes rain streaks better than other models without blurring or remaining stain. Even SPA-data's ground truth data is not perfect because it made ground truth data with some techniques and manpower, but the background of the second row of Figure 4.2 shows that our results have removed rain stains better than ground truth data. Next, Figure 4.3 is the results for Yang *et al.* dataset, and since Yang *et al.* dataset has no training data, we output the result through a model trained with Rain200H training data. Looking at the boxes shown in the first row of Figure 4.3, PreNet fail to remove rain streaks in the yellow box instead of preserving the background in the red box. Conversely, ReHEN does not preserve the background in the red, although it does remove the rain well in the yellow box. However, our MC-CNN achieves satisfactory results in both boxes. This confirms that MC-CNN also shows robust performance for out-of-domain data.

4.4 Ablation Study

We conduct an ablation study to validate all main components of MC-CNN introduced in Chapter 3. During the ablation study, Rain200H dataset is used as a training and evaluation dataset. Based on the original U-Net structure with DCR block, each component is applied in turn, and the resulting values are shown in Table 4.4. The results are reported as the average of the three experiments. From Table 4.4, we can confirm that each component contributes to improving the model performance.

4.4.1 Multi-scale connection

As mentioned in Chapter 1, we devise a multi-scale connection in the process of designing a model to better recover the details of images through much information. While we already find that multi-scale connection shows quantitative numerical im-

Table 4.4: The results of ablation study on main components of MC-CNN

WRNL	DWT	Multi-scale connection	Cutmix	PSNR	SSIM
				28.91	0.903
✓				29.35	0.909
✓	✓			30.06	0.915
✓	✓	✓		30.24	0.916
✓	✓	✓	✓	30.34	0.916

provements in Table 4.4, we also perform qualitative comparisons to see that multi-scale connections better restore details as intended (see Figure 4.4).

In the first row of Figure 4.4, we can see that multi-scale connection significantly contributes to the clearer restoration of zebra patterns. In the second row, we can also see that multi-scale connection helps to restore the tone and texture of the tree more delicately. Since rain streaks are generally close to white, so many models often make the mistake of lowering the tone of the image. In this respect, multi-scale connection is shown to play a role in improving the fundamental problem of the deraining task.

4.4.2 Region types of non-Local block

To compare the performance of our proposed WRNL block, introduced in Section 3.2, we evaluate the performance of the region types “square”, “tall”, and “wide” in regional non-local blocks. In the experiment, we use the baseline model with the

Table 4.5: The ablation study on region types of Regional Non-Local Blocks

Region Type	PSNR	SSIM
Tall Rectangle	29.78	0.913
Square	29.96	0.914
Wide Rectangle	30.06	0.915

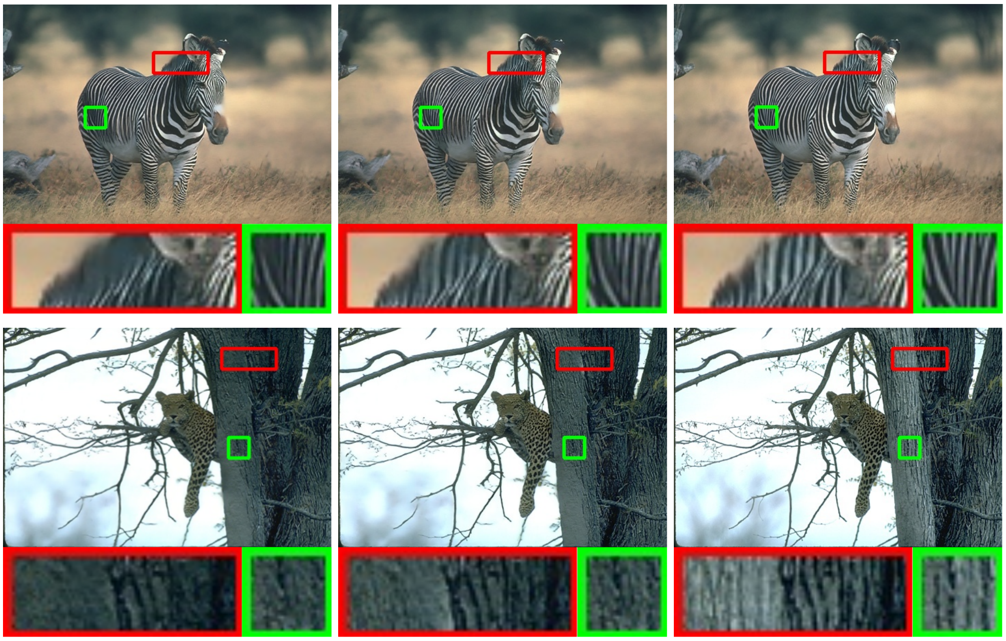


Figure 4.4: The ablation study on multi-scale connection (MSC).

WRNL block and DWT added as a model for comparison. Results presented in Table 4.5 demonstrate that the wide-type regional non-local block achieves the best performance. This result indicates that, as we hypothesize, the even distribution of rain streaks between the regions is advantageous for restoring rain-free background.

Chapter 5

Conclusion

In this study, we proposed the MC-CNN for single image deraining. MC-CNN is a model that attempts to improve detailed restore performance in the deraining process based on encoder-decoder structure in a direction that leverages all existing feature information without additional branches. To maximize the utilization of feature information that we already have, we proposed two methods named multi-scale connection and WRNL block.

Multi-scale connection is proposed to minimize information loss in the encoding-decoding process. However, if multi-scale connection did not consider channel-wise attention of concatenated multi-scale input, performance degradation occurred. So we applied SE block to multi-scale connection to learn which scale is more important in the decoding process of each scale. Through the ablation study, we confirmed that multi-scale connection plays a role in solving the critical problem of the deraining task and improve the performance of the model.

WRNL is proposed based on the assumption that regional non-local block works more effectively when rain pixels between patches are evenly distributed. Through rain pixel distribution analysis, we found that a wide rectangular region provides the evenest distribution to each patch and showed that WRNL improves model performance through experiments in Table 4.5. In several experiments, WRNL showed steady performance improvements.

Finally, MC-CNN achieved state-of-the-art in quantitative comparisons, and also in qualitative comparisons, MC-CNN showed the best recovery of rain-free background details, as well as overcoming the tone-down problem, a chronic problem in deraining models.

However, MC-CNN has limitations that the model, which is a disadvantage of encoder-decoder structure, is heavy, and that it still does not fully restore the details of the image, which should be overcome through future research.

Bibliography

[1] P. C. Barnum, S. Narasimhan, and T. Kanade, *Analysis of rain and snow in frequency space*, International journal of computer vision, 86 (2010), p. 256.

[2] J. Bossu, N. Hautière, and J.-P. Tarel, *Rain or snow detection in image sequences through use of a histogram of orientation of streaks*, International journal of computer vision, 93 (2011), pp. 348–367.

[3] G. Chai, Z. Wang, G. Guo, Y. Chen, Y. Jin, W. Wang, and X. Zhao, *Recurrent attention dense network for single image de-raining*, IEEE Access, 8 (2020), pp. 111278–111288.

[4] Y.-L. Chen and C.-T. Hsu, *A generalized low-rank appearance model for spatio-temporally correlated rain streaks*, in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 1968–1975.

[5] L.-J. Deng, T.-Z. Huang, X.-L. Zhao, and T.-X. Jiang, *A directional global sparse model for single image rain removal*, Applied Mathematical Modelling, 59 (2018), pp. 662–679.

[6] S. Deng, M. Wei, J. Wang, Y. Feng, L. Liang, H. Xie, F. L. Wang, and M. Wang, *Detail-recovery image deraining via context aggregation networks*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14560–14569.

[7] Z. Fan, H. Wu, X. Fu, Y. Huang, and X. Ding, *Residual-guide network for single image deraining*, in Proceedings of the 26th ACM international conference on Multimedia, 2018, pp. 1751–1759.

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

[9] T. Guo, H. Seyed Mousavi, T. Huu Vu, and V. Monga, *Deep wavelet prediction for image super-resolution*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 104–113.

[10] J. Hu, L. Shen, and G. Sun, *Squeeze-and-excitation networks*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.

[11] Q. Huynh-Thu and M. Ghanbari, *Scope of validity of psnr in image/video quality assessment*, Electronics letters, 44 (2008), pp. 800–801.

[12] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang, *Multi-scale progressive fusion network for single image deraining*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8346–8355.

[13] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, *Automatic single-image-based rain streaks removal via image decomposition*, IEEE transactions on image processing, 21 (2011), pp. 1742–1755.

[14] G. Li, X. He, W. Zhang, H. Chang, L. Dong, and L. Lin, *Non-locally enhanced encoder-decoder network for single image de-raining*, in Proceedings of the 26th ACM international conference on Multimedia, 2018, pp. 1056–1064.

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

[16] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, *Recurrent squeeze-and-excitation context aggregation net for single image deraining*, in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 254–269.

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

[18] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, *Multi-level wavelet-cnn for image restoration*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 773–782.

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

[20] B. Park, S. Yu, and J. Jeong, *Densely connected hierarchical network for image denoising*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0.

[21] P. Porwik and A. Lisowska, *The haar-wavelet transform in digital image processing: its status and achievements*, Machine graphics and vision, 13 (2004), pp. 79–98.

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

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

[24] L. Shen, Z. Yue, Q. Chen, F. Feng, and J. Ma, *Deep joint rain and haze removal from a single image*, in 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 2821–2826.

[25] K. Sun, B. Xiao, D. Liu, and J. Wang, *Deep high-resolution representation learning for human pose estimation*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 5693–5703.

[26] M. Tan, R. Pang, and Q. V. Le, *Efficientdet: Scalable and efficient object detection*, arXiv preprint arXiv:1911.09070, (2019).

[27] L. Trottier, P. Gigu, B. Chaib-draa, et al., *Parametric exponential linear unit for deep convolutional neural networks*, in 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2017, pp. 207–214.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention is all you need*, arXiv preprint arXiv:1706.03762, (2017).

[29] G. Wang, C. Sun, and A. Sowmya, *Erl-net: Entangled representation learning for single image de-raining*, in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 5644–5652.

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

[31] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, et al., *Deep high-resolution representation learning for visual recognition*, arXiv preprint arXiv:1908.07919, (2019).

[32] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. W. Lau, *Spatial attentive single-image deraining with a high quality real rain dataset*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 12270–12279.

[33] Y. Wang, S. Liu, C. Chen, and B. Zeng, *A hierarchical approach for rain or snow removing in a single color image*, IEEE Transactions on Image Processing, 26 (2017), pp. 3936–3950.

[34] Y. Wang, Y. Song, C. Ma, and B. Zeng, *Rethinking image deraining via rain streaks and vapors*, arXiv preprint arXiv:2008.00823, (2020).

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

[36] W. Yang, J. Liu, S. Yang, and Z. Guo, *Scale-free single image deraining via visibility-enhanced recurrent wavelet learning*, IEEE Transactions on Image Processing, 28 (2019), pp. 2948–2961.

[37] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, *Deep joint rain detection and removal from a single image*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1357–1366.

[38] W. Yang, R. T. Tan, S. Wang, Y. Fang, and J. Liu, *Single image deraining: From model-based to data-driven and beyond*, arXiv preprint arXiv:1912.07150, (2019).

[39] W. Yang, S. Wang, D. Xu, X. Wang, and J. Liu, *Towards scale-free rain streak removal via self-supervised fractal band learning.*, in AAAI, 2020, pp. 12629–12636.

[40] Y. Yang and H. Lu, *Single image deraining via recurrent hierarchy enhancement network*, in Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1814–1822.

[41] J. Yoo, N. Ahn, and K.-A. Sohn, *Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy*, arXiv preprint arXiv:2004.00448, (2020).

[42] W. Yu, Z. Huang, W. Zhang, L. Feng, and N. Xiao, *Gradual network for single image de-raining*, in Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1795–1804.

[43] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, *Cutmix: Regularization strategy to train strong classifiers with localizable features*, in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 6023–6032.

[44] H. Zhang and V. M. Patel, *Density-aware single image de-raining using a multi-stream dense network*, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 695–704.

[45] H. Zhang, V. Sindagi, and V. M. Patel, *Image de-raining using a conditional generative adversarial network*, IEEE transactions on circuits and systems for video technology, (2019).

[46] K. Zhang, W. Luo, W. Ren, J. Wang, F. Zhao, L. Ma, and H. Li, *Beyond monocular deraining: Stereo image deraining via semantic understanding*, in European Conference on Computer Vision (ECCV), 2020.

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

초 록

본 논문에서는 신경망에서 생성된 모든 스케일의 특징들을 활용하여 이미지의 세부 정보까지 복구할 수 있는 다중스케일 연결 합성곱 신경망(MC-CNN)을 제안한다. 세부 정보 복구를 위한 MC-CNN의 첫 번째 핵심은 다중스케일 연결로, 인코더 부분의 모든 스케일 특징들을 디코더에 연결하여 가능한 많은 정보를 활용하여 이미지를 복구할 수 있도록 하는 것이다. 다중스케일 연결은 단순히 각 스케일의 특징을 합치는 것이 아니라 어느 스케일의 특징이 현재 과정에서 중요한지 배울 수 있도록 채널 어텐션을 고려한다. 두 번째 핵심은 와이드 논로컬 (WRNL) 블록이다. 우리는 넓은 직사각형으로 이미지를 나눌 때 각 패치가 가장 고른 분포를 가진다는 것을 알아냈고, 이를 바탕으로 WRNL을 제안하였다. 합성 및 실제 비 데이터셋으로 진행된 많은 실험 결과들을 통해 MC-CNN이 정량적으로 기존 방법들을 능가하고 정성적으로도 많은 개선이 이루어졌음을 확인하였다.

주요어: 딥러닝, 비제거, 비, 이미지 전처리, 합성곱 신경망
학번: 2019-28867