IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 April 2024, accepted 19 May 2024, date of publication 31 May 2024, date of current version 7 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3407750

RESEARCH ARTICLE

Layer Decomposition Learning Based on Discriminative Feature Group Split With Bottom-Up Intergroup Feature Fusion for Single Image Deraining

YUNSEON JANG[1], DUC-TAI LE[1], CHANG-HWAN SON[2], AND HYUNSEUNG CHOO[1], (Member, IEEE)

[1]Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea
[2]Department of Software Science and Engineering, Kunsan National University, Gunsan 54150, South Korea

Corresponding authors: Chang-Hwan Son (cson@kunsan.ac.kr) and Hyunseung Choo (choo@skku.edu)

This work was supported by Institute for Information Communication Technology Planning and Evaluation (IITP) Grant funded by the Korean Government [Ministry of Science and Information and Communication Technology (ICT)]; in part by the ICT Creative Consilience Program under Grant IITP-2024-2020-0-01821 (40%); in part by the Artificial Intelligence Graduate School Program, Sungkyunkwan University, under Grant 2019-0-00421 (10%); in part by the Artificial Intelligence Innovation Hub under Grant RS-2021-II212068 (40%); and in part by the Brain Korea 21 Program for Leading Universities and Students (BK21 FOUR) Project (10%).

ABSTRACT Rain streaks impede image feature extraction, hindering the performance of computer vision algorithms such as pedestrian and lane detection in adverse weather conditions. Image deraining is crucial for enhancing reliability of such algorithms. However, detail and texture information of objects in background areas are often lost during the deraining process due to their structural similarity with rain streaks. To remove rain streaks effectively while preserving image details, we propose a novel layer decomposition learning network (LDLNet) to separate rain streaks and object details in rainy images. LDLNet consists of two parts: the discriminative group feature split (DGFS) and the group feature merging (GFM). DGFS utilizes sparse residual attention modules (SRAM) to capture spatial contextual features of rainy images, enhancing the network's ability to understand the complex relationships between rain streaks and object details. In addition, DGFS employs the bottom-up intergroup feature fusion (BIFF) approach to aggregate multi-scale context information from continuous SRAMs, facilitating the decomposition of rainy images into discriminative feature groups. Subsequently, GFM integrates these feature groups by concatenating them, preserving the interdependent characteristics of clean backgrounds and rain layers. Experimental results reveal that the proposed approach achieves superior rain removal and detail preservation in both synthetic datasets and real-world rainy images compared to the state-of-the-art rain removal models.

INDEX TERMS Computer vision, deep learning, image deraining, image detail maintenance.

I. INTRODUCTION

The visual quality of images is degraded when captured in rainy weather. As shown in Fig. 1(a), rain streaks obscure the view and prevent the detection of details. As these rain streaks hinder visual feature extraction, they adversely affect the performance of computer vision algorithms including object detection and image recognition. For outdoor vision systems such as visual surveillance and autonomous vehicles, it is essential to remove rain streaks and preserve object details from input rain images. Rain removal, also referred to as image deraining, aims to remove rain streaks from input rain images and produce high-quality images with details (e.g., edges and texture patterns).

The problem is nontrivial as rain streaks appear in various directions and shapes in a rain image and objects

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

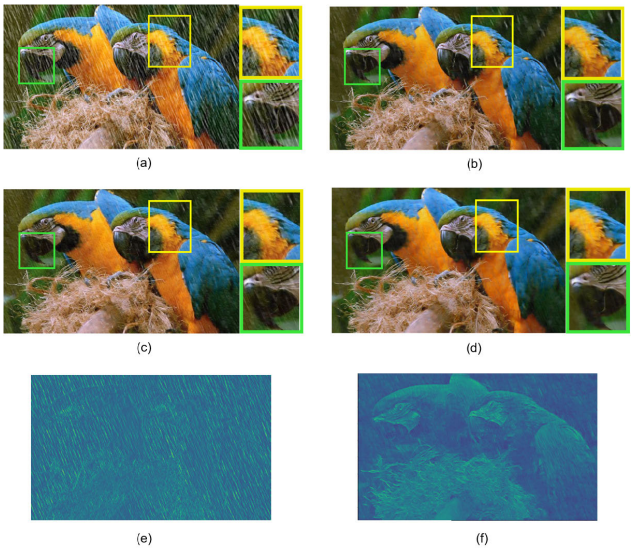


FIGURE 1. Single image deraining visualizations with rain streak and object detail. From (a)-(f): (a) Rain image, (b) RESCAN [21], (c) DRD-Net [5], (d) Our proposed LDLNet, (e) Appearance of rain streaks generated by the proposed method, and (f) Appearance of object details generated by the proposed method.

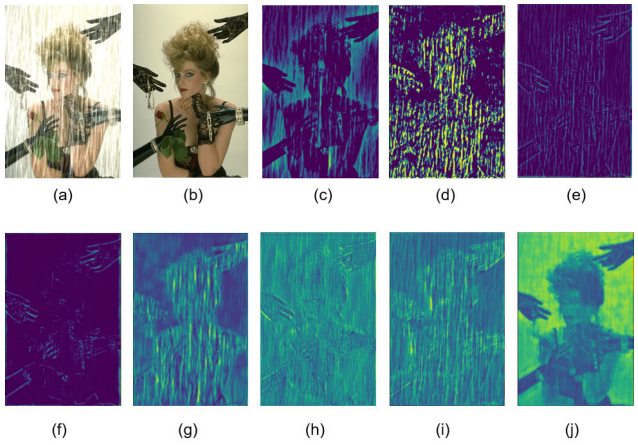


FIGURE 2. Feature map visualizations of rain streak and detail layers from different models. (a) Input rain image, (b) Ground truth, (c) Detail layer from DDN [8], (d) Negative residual from DDN [8], (e) 5th level of Laplacian pyramids from LPNet [9], (f) 5th level of reconstruction layer from LPNet [9], (g) Rain streak feature from DRD-Net [5], (h) Detail repair feature from DRD-Net [5], (i) Feature group of rain streaks from LDLNet, and (j) Feature group of object details from LDLNet.

affected can disappear or get blurred during the rain streak removal. In contrast with rain removal in a video, where temporal information leveraged from adjacent frames of the video helps detect rain streaks [2], [11], [38], single-image deraining is relatively more difficult, thus, it has attracted considerable research attention recently. Rain streaks in a single image were traditionally removed using bilateral filter [20], low rank [3], sparse coding [23], and Gaussian mixing model [22]. Methods based on deep learning have been recently deployed for the task [7], [8], [10], [32], [35]. Although the state-of-the-art (SOTA) methods achieve notable results as in Figs. 1(b) and (c), there is still room for improvement as rain removal and detail preservation have not been solved simultaneously.

For perfect rain removal, it is crucial to separate features of rain streaks and object details from input images. To achieve this, we designed a novel layer decomposition learning network (LDLNet) to split the features of rain images into groups. The proposed network generates different image appearances, which contain discriminative and complementary visual features of rain streaks and object details, as shown in Figs. 1(e) and (f), respectively. Such a feature decomposition learning eventually results in superior performances in rain removal and image detail preservation, as shown in Fig. 1(d).

A. DRAWBACKS OF CONVENTIONAL FEATURE DECOMPOSITION MODELS

Several rain-removal models have been developed to separate image features into different layers. The deep detail network (DDN) [8] employs a guided filter to decompose the input image in Fig. 2(a) into a low-frequency part and a high-frequency one. The high-frequency part indicates the detail layer of the input image, as shown in Fig. 2(c). The detail layer then becomes an input of the stacked layers in DDN to estimate the negative residual which contains rain streak information. However, the rain streaks are still unclear in the negative residual, as shown in Fig. 2(d). In contrast, lightweight pyramid network (LPNet) [9] extracts multi-scale features using a pyramid-based image decomposition technique. The model produces a multi-resolution Laplacian pyramid of the input image and generates a rain-free image using a reconstruction layer. However, the object details and rain streaks are not completely separated in the Laplacian pyramid, as in Fig. 2(e), and the details of the woman's face are not visible from the image appearance of the reconstruction layer as in Fig. 2(f); therefore, LPNet incurs detail loss in rain-free images. To supplement the information lost, DRD-Net [5] employs a detail repair network after removing the rain streaks using a rain residual network. However, the edges of objects remain in the rain streak layer, as shown in Fig. 2(g), and the edges and rain streaks are not well separated, as shown in 2(h). The proposed LDLNet overcomes these limitations by decomposing discriminative feature groups of rain images into semantic and discriminative layers. This discriminative layer decomposition results in a clear separation of detail and rain streak layers, as shown in Fig. 2(i) and (j), respectively.

B. THE PROPOSED DISCRIMINATIVE FEATURE GROUP DECOMPOSITION APPROACH

The proposed LDLNet comprises two main components, discriminative group feature split (DGFS) and group feature merging (GFM) parts. DGFS aims to decompose rainy images into discriminative feature groups, while GFM merges them to facilitate image deraining. Three discriminative feature groups are designed based on the belief that image deraining involves decomposing a rainy image into a clean

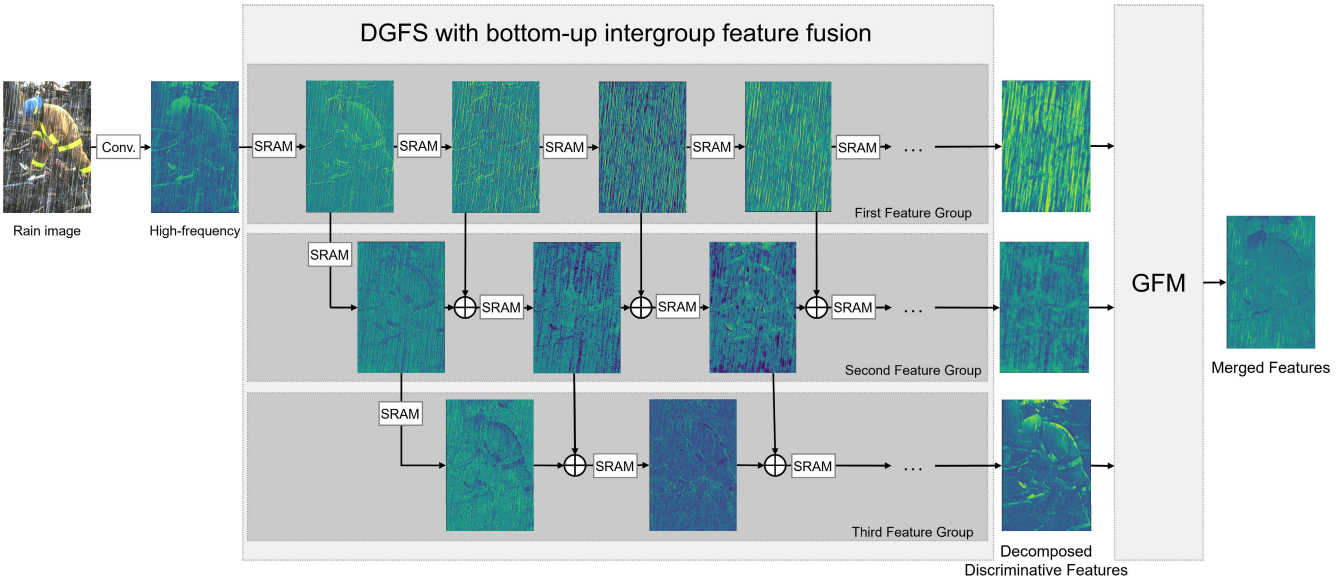


FIGURE 3. A conceptual framework of LDLNet. In the framework, DGFS decomposes rainy images into discriminative feature groups and GFM merges them while maintaining interdependent characteristics of rain layer and clean background for an image deraining.

background and a rain layer, each containing discriminative feature groups. As shown in Fig. 3, each row corresponds to a feature group, wherein the first group predicts rain streaks from high-frequency features initialized by convolution operations, the second group preserves intermediate features, and the third one indicates object details by accumulating complementary features through a bottom-up intergroup feature fusion (BIFF) approach. The approach is employed based on the idea that the clean background and the rain layers are not entirely independent due to the presence of objects sharing similar colors and shapes with rain, and therefore their feature groups should be jointly learned for better performance. Particularly, the BIFF approach complementarily transfers feature information among sparse residual attention modules (SRAMs) from low to high levels to fuse the discriminative feature groups of a rainy image. This approach allows for better utilization of feature information, resulting in superior performance in single image deraining.

In summary, the contributions of the paper are as follows:

- Introduction of LDLNet, a novel single-image deraining framework employing DGFS and GFM components. The proposed framework leverages layer decomposition learning to effectively remove rain streaks while preserving image details.
- Design of SRAM to capture spatial contextual features of rainy images, enhancing the network's ability to understand the complex relationships between rain streaks and object details.
- Implementation of the BIFF approach in DGFS to aggregate multi-scale context information from continuous SRAMs, facilitating the decomposition of rainy images into discriminative feature groups.
- Integration of GFM to concatenate feature groups from DGFS, thereby maintaining interdependent characteristics of clean background and rain layer.
- Comprehensive performance evaluation of LDLNet against state-of-the-art deraining models using different datasets, demonstrating its superior performance on both synthetic and real-world rainy images.

II. RELATED WORK

A. DEEP-LEARNING-BASED METHODS

Deep-learning-based methods have been adopted to remove rain streaks, achieving significant results. DerainNet [7] uses CNN-based non-linear mapping to transform a rainy image into a clean one. The results were improved using a deep detail network (DDN) [8] with a combination of ResNet [12] and a global skip connection. However, those approaches are not practical because they do not deal with various directions, scales, and amounts of rain in the training datasets. To this end, the multi-stream dense network (DID-MDN) [35] determines rain information using a residual-aware rain-density classifier. Joint rain detection and removal (JORDER) [32] solves the lack of local contextual information without loss of local detail by a dilated factor. The dual graph convolutional network (DualGCN) [10] uses local spatial patterns and global contextual information to effectively remove rain. Furthermore, the multi-stage progressive image restoration network (MPRNet) [34] recovers the degraded inputs through information exchange between different stages of a multi-stage architecture. The progressive coupled network (PCNet) [18] extracts the hierarchical features of multi-scale rain streaks and separates the rain-free content and rain streaks progressively. Recently, a general U-shaped Transformer (Uformer) [33] has been proposed leveraging the

Transformer block to capture global dependencies through window-based self-attention. Uformer employs U-shaped encoder-decoder networks with skip connections to capture multi-scale information hierarchically, making it suitable for various image restoration tasks.

B. BATCH NORMALIZATION REMOVAL

Batch normalization (BN) [16] reduces the internal coverage shift by standardizing data distribution. It is widely adopted in deep-learning-based models because employing BN in the training process stabilizes the model faster than the one without BN. However, as BN limits the distribution of values, it adversely affects rain streak removal because the properties of rain streaks vary according to their intensity, direction, brightness, and shape. The lightweight pyramid of networks (LPNet) [9] decomposes the rainy image using the Laplacian pyramid and predicts the clean image using the Gaussian pyramid without BN. The recurrent SE context aggregation network (RESCAN) [21] is designed to deeply remove rain streaks through continuously deployed networks. Also, BN is not used in RESCAN to remove rain streaks with various directions, colors, and shapes. Furthermore, excluding BN from the training process reduces the required computing resources, and thus makes the process faster.

C. ATTENTION MECHANISM

The efficiency of attention mechanisms in several computer vision tasks including image classification [13], [43], segmentation [6], [14], detection [31], denoising [1], [29], super-resolution [4], [39], deblurring [27], [28], and restoration [34] has been proven. Accordingly, recent deraining methods employ attention to improve their performance. The multi-scale progressive fusion network (MSPFN) [17] obtains rain streak information in different scales by combining an attention mechanism with the multi-scale pyramid structure. The detail-recovery image deraining network (DRD-Net) [5] utilizes two sub-networks for image deraining and preservation of image detail. In DRD-Net, the rain residual network effectively removes rain using an attention mechanism, and then the detail repair network recovers image details using spatial information obtained by a dilated convolution.

III. RAIN FORMULATION

A simple rainy image including rain with the same direction can be expressed by the following equation:

$$O = B + R \tag{1}$$

where B is the image with a clean background and R is the rain streak layer. Image deraining can be seen as a problem that decomposes the rainy image O into the above two elements. However, real-world rain streaks have irregular distributions. If there is no wind, the rain falls perpendicularly to the ground. Otherwise, it falls diagonally. Fast wind speed causes the rain to scatter over a wider area, and this phenomenon leads the rain to have an inconstant distribution, for which the equation is as follows:

$$O = B + \sum_{i=1}^{S} R_i \tag{2}$$

where R_i is i-th rain streak layer among S different ones.

If the rain is heavy, the problem becomes more complicated because rain appearance is shaped by not only individual rain streaks but also the accumulation and dispersion of rain streaks in the atmosphere. Rain accumulation results in visual effects similar to mist or fog. Owing to this phenomenon, visibility decreases significantly. Since there are various cases in real rain environments, the synthetic rainy images are expressed as follows:

$$O = \alpha(B + \sum_{i=1}^{S} R_i) + (1 - \alpha)A, \qquad 0 \le \alpha \le 1 \tag{3}$$

where A is global atmospheric light and α is atmospheric transmission [5], [32].

IV. PROPOSED METHODS

Our single image deraining method aims to remove rain streaks and maintain image details through layer decomposition learning with BIFF approach. In this section, we explain our proposed network and loss function.

A. OVERALL STRUCTURE

A typical phenomenon when removing rain is that parts of objects in the input image become blurred and disappear when the patterns of rain streaks and object details are similar. To simultaneously achieve rain removal and image detail maintenance, we aim to decompose feature groups with different appearances in a rain image using a single network rather than network separation as in DRD-Net [5]. The proposed LDLNet consists of two parts: DGFS and GFM, as shown in Fig. 4(a). The first two convolutional layers are interpreted as embedding to transform the rain image into high-frequency feature maps. A novel SRAM is developed to acquire spatial contextual information from high-frequency feature maps based on dilated convolution and squeeze-and-excitation attention. Such SRAMs are arranged continuously in three rows of DGFS. The unique aspect of LDLNet is the BIFF-based feature fusion in DGFS, where each SRAM aggregates a feature group from not only the previous SRAM in the same row but also the SRAM in the lower layer. This layer decomposition learning facilitates the representation of distinct feature groups for rain streaks and object details, even when they share similar colors and shapes. As a result, LDLNet achieves effective rain removal while preserving image details.

In the GFM part, three feature groups of DGFS are concatenated to accumulate feature information. Such concatenation ensures that the distinct characteristics of each feature group are preserved, thereby facilitating the generation

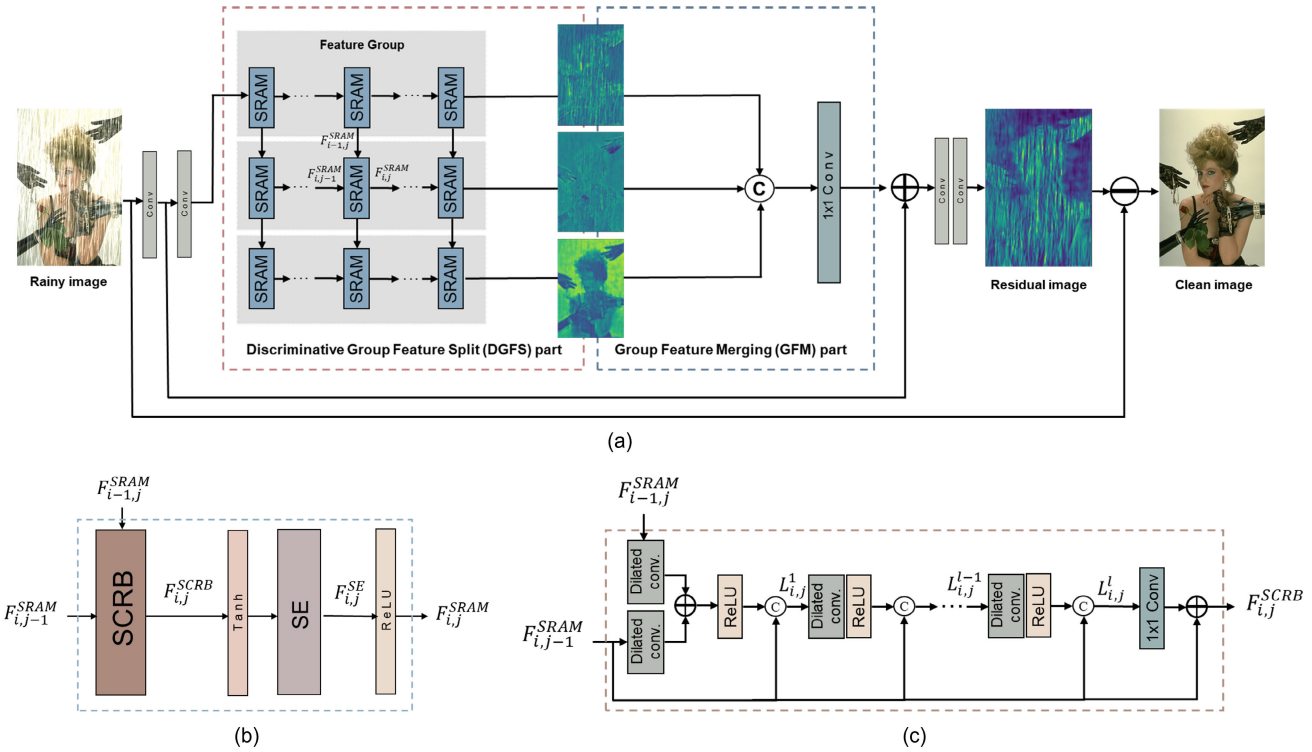


FIGURE 4. Layer decomposition learning network (LDLNet). From (a)-(c): (a) Overall architecture, (b) Sparse residual attention module (SRAM), (c) Sparse connected residual block (SCRB). In LDLNet, DGFS part generates discriminative and complementary feature groups by fusing features from low to high levels based on BIFF approach, and GFM part aggregates the feature groups for rain removal.

of a high-quality residual image for rain streak removal. A 1×1 convolution is used to reduce feature dimensions. Afterward, residual learning is introduced into our model to lower complexity. The last two convolutional layers act as decoders to recover the RGB channel of the color image. Eventually, LDLNet produces a clean image by subtracting the simple and sparse residual image from the rain image. Moreover, a novel loss function, which also contributes to quality improvement of the generated rain-free images, is developed in the proposed network.

B. SPARSE RESIDUAL ATTENTION MODULE (SRAM)

Fig. 4(b) shows the structure of SRAM with a sparse connected residual block (SCRB) and a squeeze-and-excitation (SE) operation [13]. The proposed SCRB is simplified from the residual dense block (RDB) [40] which hierarchically extracts features and then connects the generated feature maps in all layers by a densely connected convolution layer. Although RDB utilizes sufficiently local features and improves the flow of information, it suffers from a large number of training parameters due to the dense connections, i.e. a network with L layers requires $L(L+1)/2$ connections. To reduce the feature duplication caused by redundant connections in RDB [41], we design SCRB which consists of minimal connections for initial information preservation.

1) CONTEXTUAL INFORMATION EXTRACTION WITH SCRB

The proposed SCRB, located at the beginning of SRAM j ($j \geq 1$) in feature group i ($i \geq 1$), receives two inputs $F^{\text{SRAM}}_{i,j-1}$ and $F^{\text{SRAM}}_{i-1,j}$, which represent the outputs of the previous SRAM in the same feature group and the same order SRAM in the lower feature group, respectively. The input features are fused through convolution layers to form discriminative and complementary feature groups, as shown in Fig. 4(c). The output of the first convolution layer in the SCRB is:

$$L^1_{i,j} = \text{ReLU}(\text{Conv}^d_{3\times3}(F^{\text{SRAM}}_{i,j-1}) + \text{Conv}^d_{3\times3}(F^{\text{SRAM}}_{i-1,j})) \| F^{\text{SRAM}}_{i,j-1} \quad (4)$$

where ReLU denotes the rectified linear unit function, $\text{Conv}^d_{3\times3}$ is a dilated convolution layer with 3×3 kernel size, and $\|$ denotes the concatenation function. Employing dilation enables SRAM to obtain a large receptive field to increase the size of the area exposed to the convolution kernel. The contextual information provided by such a large receptive field is essential for rain removal and image detail recovery, as visually verified in Section V-C.

Noting that $F^{\text{SRAM}}_{1,0}$ is the high-frequency features of the input rainy image after two convolution layers. $F^{\text{SRAM}}_{i,0} = F^{\text{SRAM}}_{0,j} = \mathbf{0}$ ($\forall i > 1, \forall j \geq 1$) indicates that the first SRAM in each feature group has only one input from the lower one, except the first feature group. To minimize the number of connections, the proposed block does not connect the outputs of all convolution layers. It concatenates only the output of the previous SRAM in the same feature group with the output of

each layer. Let l denote the number of convolution layers in each block, the output of layer k ($1 < k \leq l$) is formulated as follows:

$$L_{i,j}^{k} = \mathrm{ReLU}(\mathrm{Conv}_{3\times3}^{d}(L_{i,j}^{k-1}))\|F_{i,j-1}^{\mathrm{SRAM}} \tag{5}$$

Lastly, SCRB deploys a 1×1 convolution layer to control the output information. The previous SRAM of the same feature group is also added to preserve the information flow. Consequently, SCRB of SRAM j in feature group i produces the output as follows:

$$F_{i,j}^{\mathrm{SCRB}} = \mathrm{Conv}_{1\times1}(L_{i,j}^{l}) + F_{i,j-1}^{\mathrm{SRAM}} \tag{6}$$

where $\mathrm{Conv}_{1\times1}$ represents the convolution with the kernel size of 1×1. Experimental results reveal that SCRB with minimum connections to prevent feature duplication performs superior to RDB.

2) FEATURE BOOSTING WITH SE OPERATION

After applying SCRB to extract contextual information, SRAM focuses on more important features through SE operation which improves model performance without significantly increasing computation costs. In the operation, the feature channel with a relatively large amount of contextual information is reinforced to derive only important features by modeling the correlation between different feature channels. SE operation in SRAM j of feature group i is mathematically expressed as follows:

$$F_{i,j}^{\mathrm{SE}} = \mathrm{Sigmoid}(\mathrm{FC}(\mathrm{ReLU}(\mathrm{FC}(\mathrm{GAP}(\mathrm{Tanh}(F_{i,j}^{\mathrm{SCRB}})))))\times\mathrm{Tanh}(F_{i,j}^{\mathrm{SCRB}}) \tag{7}$$

where Sigmoid and Tanh denote sigmoid and hyperbolic tangent functions, respectively. FC is a fully connected layer and GAP is a global average pooling one. The SE operation enables SRAM to extract spatial contextual information. Eventually, the output of SRAM j in feature group i is described as follows:

$$F_{i,j}^{\mathrm{SRAM}} = \mathrm{ReLU}(F_{i,j}^{\mathrm{SE}}) \tag{8}$$

C. DISCRIMINATIVE GROUP FEATURE SPLIT AND GROUP FEATURE MERGING PARTS

LDLNet preserves object details, which tend to fade or disappear when the rain is removed from rainy images, by generating discriminative and complementary visual feature groups based on a bottom-up intergroup feature fusion approach. In the DGFS part, SRAMs are arranged in three groups to fuse spatial contextual information. The discriminative feature information generated by SRAMs of the lower feature group is forwarded to the ones of the upper feature group as complementary features, as detailed in Section IV-B. The output of the DGFS is formulated as follows:

$$F^{\mathrm{DGFS}} = \begin{cases} F_{1,m}^{\mathrm{SRAM}} \\ F_{2,m}^{\mathrm{SRAM}} \\ F_{3,m}^{\mathrm{SRAM}} \end{cases} \tag{9}$$

where m denotes the number of SRAMs in each group.

Recall that SRAM employs dilated convolutions to efficiently mitigate resolution loss without increasing the kernel size, which leads to a proportional growth in the number of parameters. The dilation rates vary depending on the location of SRAMs, as shown in Table 1. Particularly, $\mathrm{SRAM}_{i,1}$, $\mathrm{SRAM}_{i,2}$, $\mathrm{SRAM}_{i,6}$, and $\mathrm{SRAM}_{i,7}$ share the same dilation scale, while $\mathrm{SRAM}_{i,3}$, $\mathrm{SRAM}_{i,4}$, and $\mathrm{SRAM}_{i,5}$ use the higher ones. The receptive field obtained by such dilated convolutions increases gradually through the bottom-up intergroup feature fusion approach. This accumulation of feature information enables DGFS to decompose the rainy image into three feature groups with different visual appearances. Particularly, the lowest feature group predicts low-level features, i.e., rain streaks, and the highest group captures meaningful object details, as shown in Fig. 4(a). The rainy image is decomposed into three feature groups naturally during the forward pass because the underlying operations primarily involve stacked convolutional layers.

In the GFM part, the three output feature groups of DFGS are merged to preserve image information. The three output feature groups are merged in GFM part to preserve image information. Particularly, GFM concatenates features from the last SRAM of each feature group and deploys a 1×1 convolution layer to control the output information. The output of GFM part is formulated as follows:

$$F^{\mathrm{GFM}} = \mathrm{Conv}_{1\times1}(F_{1,m}^{\mathrm{SRAM}}\|F_{2,m}^{\mathrm{SRAM}}\|F_{3,m}^{\mathrm{SRAM}}) \tag{10}$$

Noting that BN is not used in training as it adversely affects the robustness of the deraining model by limiting the variety of rain distribution [5], [21]. The performance of the proposed model is influenced by the dilation, the number of SRAMs, and also the connections between SRAMs within DGFS. These aspects are discussed in more detail in Section V-C.

D. LOSS FUNCTION

We use the Euclidean distance as the training loss to train the model and formulate it as follows:

$$\mathrm{Loss}_{\mathrm{euc}} = \sqrt{\sum_{i=1}^{N}(P_i - B_i)^2 + \epsilon} \tag{11}$$

P_i is the predicted clean image from our LDLNet and B_i is the ground truth, and N is the number of training data samples for each iteration. We add ϵ to prevent the training loss from becoming zero. However, the rain makes the detail of the background blurred which tends to result in blurred reconstruction. To improve the image detail maintenance, we use the Laplacian operator [19], which is widely used when detecting object edges in image processing. We additionally propose the edge loss between the predicted clean image and ground truth using the Laplacian operator, which is defined as

$$\mathrm{Loss}_{\mathrm{edge}} = \sqrt{\sum_{i=1}^{N}(\mathrm{Lap}(P_i) - \mathrm{Lap}(B_i))^2 + \epsilon} \tag{12}$$

TABLE 1. The list of dilation scales of SRAM in each feature group (i ranges from 1 to 3).

SRAM	$\text{SRAM}_{i,1}$	$\text{SRAM}_{i,2}$	$\text{SRAM}_{i,3}$	$\text{SRAM}_{i,4}$	$\text{SRAM}_{i,5}$	$\text{SRAM}_{i,6}$	$\text{SRAM}_{i,7}$
Convolution	3×3	3×3	3×3	3×3	3×3	3×3	3×3
Dilation	$1\ (2^0)$	$1\ (2^0)$	$3\ (2^1+1)$	$5\ (2^2+1)$	$9\ (2^3+1)$	$1\ (2^0)$	$1\ (2^0)$

where $\text{Lap}(P_i)$ and $\text{Lap}(B_i)$ are edge maps through the Laplacian operator. Structural similarity index measure (SSIM) loss is usually used in the reconstruction process, therefore we include SSIM in training loss. It compares the local region of pixels in the predicted image with the ground truth. The similarity of each pixel and the global structure are preserved simultaneously.

$$\text{Loss}_{\text{ssim}} = 1 - \sum_{i=1}^{N} \text{SSIM}(P_i, B_i) \tag{13}$$

Since the maximum value of SSIM is 1, we define the output by subtracting the SSIM value between two images as SSIM loss. We minimize the three loss functions and the overall loss function used to train LDLNet is formulated as follows:

$$\text{Loss}_{\text{total}} = \text{Loss}_{\text{euc}} + \text{Loss}_{\text{edge}} + \text{Loss}_{\text{ssim}} \tag{14}$$

It is important to note that the loss function is calculated with three subsets of parameters corresponding to the three feature groups. During backpropagation, parameters in each group are supervised by the corresponding gradient from the loss, as the gradient of other parameters is zero. Consequently, the parameters in each group update accordingly to minimize the overall loss, leading to effective learning and feature extraction across all three groups.

V. PERFORMANCE EVALUATION

A. EXPERIMENT SETTINGS

Since it is difficult to obtain a pair consisting of a rainy image and a clean one in the real world, we used multiple synthetic datasets for training. The first group of datasets consists of images with synthesized rain streaks. From the Rain2600 [35], we randomly selected 300 heavy rainy images, 300 medium rainy images, and 100 light rainy images for training. Then, a total of 1200 images were randomly picked from the dataset to form our first testing set, which is Test1200. We divided Rain800 [36] into two sets, 700 images for training and 100 images for testing, named Test100. Similarly, we picked 100 and 12 images in Rain100H and Rain12 [32] for training, respectively. The testing sets Test100H, Test1000, and Test1400 are 100, 1000, and 1400 synthesized rainy images of Rain100H [32], Rain1000 [35], and Rain1400 [8], respectively. Detailed descriptions of the datasets used are shown in Table 2.

The second group of datasets was designed to simulate the common occurrences of rain-streaks and raindrops in real-world rainy scenarios. We used the RainDS [25] dataset for this purpose, as described in Table 3. For training, we randomly selected 1600 images from the dataset,

TABLE 2. Datasets of images with synthesized rain-streaks.

Dataset	Training Samples	Testing Samples	Testset Name
Rain2600 [35]	700	1,200	Test1200
Rain800 [36]	700	100	Test100
Rain100H [32]	100	100	Test100H
Rain12 [22]	12	0	-
Rain1000 [35]	0	1,000	Test1000
Rain1400 [8]	0	1,400	Test1400
Total Count	1,512	3,800	-

TABLE 3. Dataset of images with synthesized rain-streaks and raindrops.

RainDS [25]	Training Samples	Testing Samples	Testset Name
Rain-streak only	300	200	RS200
Raindrop only	300	200	RD200
Both	1,000	200	RDS200
Total Count	1,600	600	-

including 300 images with only rain-streaks, 300 images with only raindrops, and 1000 images with both raindrops and rain-streaks. To evaluate the performance of our method on the specific rain types, we randomly formed three testing sets of RS200, RD200, and RDS200 from 200 rainy images each, including only synthesized rain-streaks, only synthesized raindrops, and both raindrops and rain-streaks, respectively. We further experimented on real-world images downloaded from the internet to evaluate the generalization ability.

We set the number of training epochs to 200. Each epoch contained 161 iterations. To optimize the model, we used the Adam optimizer and set the batch size to 10. The initial learning rate was 0.01, which was divided by 2 after every 15 epochs. From 120 epochs, the learning rate was fixed. Each training image was randomly cropped to make 64×64 patches, on which flip and rotation were optionally applied. We ran all experiments in the same environment with the Nvidia RTX A6000 GPU.

Our performance evaluation employs four commonly used metrics to measure image quality: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Visual Information Fidelity (VIF), and Feature Similarity (FSIM). PSNR evaluates the quality of derained images by comparing pixel-wise differences between them and ground truth images [15]. SSIM measures the similarity between derained images and ground truth images in terms of luminance, contrast, and structure [30]. VIF [26] and FSIM [37] assess the preservation of visual information and image features, respectively. These metrics offer complementary insights into different aspects of image quality, enabling a comprehensive evaluation of our proposed method's performance compared to state-of-the-art (SOTA) techniques. However, the quantitative evaluation is only available on synthesized rainy

TABLE 4. Quantitative comparisons on synthesized rain-streak datasets. (Best results: red, second-best: blue).

Methods		Rainy images	DerainNet [7]	DDN [8]	DID-MDN [35]	LPNet [9]	RESCAN [21]	DRD-Net [5]	DualGCN [10]	LDLNet
Test1200	PSNR	22.32	22.37	27.06	23.83	24.75	31.23	27.71	26.24	29.08
	SSIM	0.6961	0.7887	0.8316	0.7261	0.8133	0.8820	0.8295	0.8204	0.8684
	VIF	0.3306	0.3860	0.4169	0.2783	0.4216	0.4801	0.4110	0.4214	0.4744
	FSIM	0.6673	0.6752	0.6973	0.6181	0.6839	0.7262	0.6893	0.6807	0.7176
Test100	PSNR	21.23	21.79	22.25	21.38	22.86	25.73	23.44	23.13	25.18
	SSIM	0.6492	0.7882	0.7721	0.7110	0.8082	0.8217	0.7867	0.7953	0.8407
	VIF	0.2419	0.3129	0.3479	0.2524	0.3483	0.3827	0.3381	0.3414	0.3954
	FSIM	0.6377	0.6553	0.6745	0.5979	0.6650	0.6863	0.6585	0.6549	0.6884
Test100H	PSNR	12.23	13.65	19.02	19.51	18.15	25.30	22.15	20.03	24.44
	SSIM	0.3546	0.6013	0.6017	0.5805	0.6550	0.7861	0.6882	0.6481	0.7800
	VIF	0.0787	0.1568	0.1585	0.1339	0.1746	0.2873	0.1984	0.1773	0.2665
	FSIM	0.5335	0.5601	0.5744	0.5264	0.5781	0.6394	0.5853	0.5664	0.6294
Test1000	PSNR	19.51	20.42	24.97	22.71	23.02	24.51	24.32	23.63	23.65
	SSIM	0.6481	0.7542	0.7930	0.6720	0.7879	0.7935	0.7817	0.7752	0.8108
	VIF	0.2519	0.3247	0.3518	0.2465	0.3625	0.3743	0.3484	0.3483	0.4091
	FSIM	0.6417	0.6510	0.6700	0.6000	0.6634	0.6798	0.6656	0.6581	0.6889
Test1400	PSNR	23.92	24.26	26.98	21.56	26.37	27.53	26.77	26.72	26.43
	SSIM	0.7636	0.8244	0.8551	0.6612	0.8504	0.8593	0.8416	0.8456	0.8615
	VIF	0.3545	0.3801	0.4237	0.2971	0.4228	0.4502	0.4141	0.4268	0.4781
	FSIM	0.7188	0.7044	0.7322	0.5831	0.7204	0.7430	0.7215	0.7154	0.7451

TABLE 5. Quantitative comparisons on the synthesized rain-streak and raindrop dataset. (Best results: red, second-best: blue).

Methods		Rainy images	LPNet [9]	RESCAN [21]	DRD-Net [5]	DualGCN [10]	MPRNet [34]	PCNet [18]	Uformer [33]	LDLNet
RS200	PSNR	21.15	26.78	28.14	25.15	34.73	30.76	29.36	30.15	37.11
	SSIM	0.6501	0.8718	0.8949	0.8170	0.9569	0.8853	0.8754	0.9273	0.9713
	VIF	0.2654	0.4224	0.5674	0.3977	0.6469	0.4816	0.4553	0.5837	0.7069
	FSIM	0.6658	0.6990	0.7304	0.6782	0.8110	0.7320	0.7020	0.7363	0.8348
RD200	PSNR	20.52	22.99	28.14	25.15	30.78	26.45	27.21	26.84	33.26
	SSIM	0.8250	0.8268	0.8949	0.8170	0.9469	0.8968	0.9001	0.9178	0.9665
	VIF	0.4692	0.4644	0.5674	0.3977	0.7280	0.5832	0.5595	0.62.41	0.8077
	FSIM	0.7636	0.6990	0.7304	0.6782	0.8647	0.7663	0.7487	0.7371	0.8778
RDS200	PSNR	17.93	24.05	26.62	24.11	29.74	24.43	25.48	26.33	31.68
	SSIM	0.5501	0.8098	0.8101	0.7404	0.9145	0.7937	0.8132	0.8685	0.9393
	VIF	0.1692	0.3305	0.3654	0.2834	0.5445	0.3396	0.3541	0.4632	0.6140
	FSIM	0.5621	0.6387	0.6559	0.6248	0.7627	0.6459	0.6434	0.6871	0.7942

datasets, as ground truth does not exist for real rainy images. We evaluated the real-world dataset from a visual perspective.

B. EXPERIMENT RESULTS AND DISCUSSION

We compared the proposed model with the most notable single-image deraining models. All related models were re-implemented from their published source codes and retrained with the same training datasets and the environments described in the related papers. The comparison models are as follows:

- Deep CNN based method (DerainNet) [7]
- Deep detail network (DDN) [8]
- Density-aware deraining network (DID-MDN) [35]
- Lightweight pyramid of networks (LPNet) [9]
- Recurrent squeeze-and-excitation context aggregation network (RESCAN) [21]
- Detail-recovery image deraining network (DRD-Net) [5]
- Dual graph convolutional networks (DualGCN) [10]
- Multi-stage progressive image restoration network (MPRNet) [34]
- Progressive coupled network (PCNet) [18]
- U-shaped Transformer (Uformer) [33]

The performance of LDLNet is quantitatively compared to SOTA methods on the two groups of synthesized rainy image datasets. Table 4 shows the performance of the methods on the first group of datasets, which includes only synthesized rain streaks. The superior performances of the proposed model on Test1000 and Test1400, in which the testing images are different from training ones, reveal that LDLNet generalizes the rain streak features and works well with new kinds of images. In the case of Test1200, Test100, and Test100H, the proposed model approaches the performance of RESCAN when the testing and training images are similar. Table 5 shows the performance of the models on the second group of datasets, which include both synthesized rain-streaks and raindrops. LDLNet outperforms all SOTA models in all testing sets of RS200, RD200, and RDS200.

We also evaluate the qualitative performance of the deraining methods on the two groups of synthetic datasets. On the synthesized rain-streak datasets, the proposed LDLNet removes the rain streaks better than other models, as shown in the first three rows of Fig. 5. The synthetic rainy images in the fourth and fifth rows of the figure imitate the phenomenon of dark clouds, which block the sun-sight and make the images darker in rainy circumstances. Other

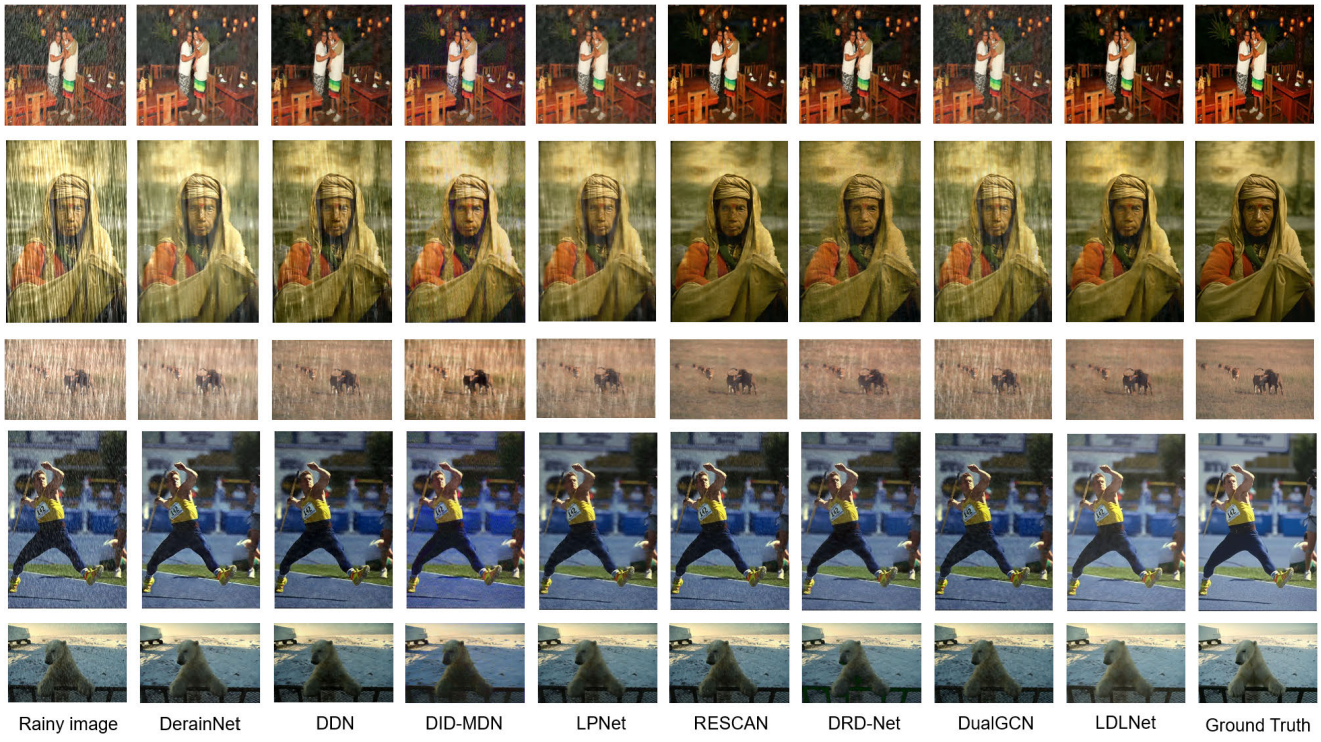


FIGURE 5. Qualitative comparisons of rain removal on synthesized rain-streak datasets.

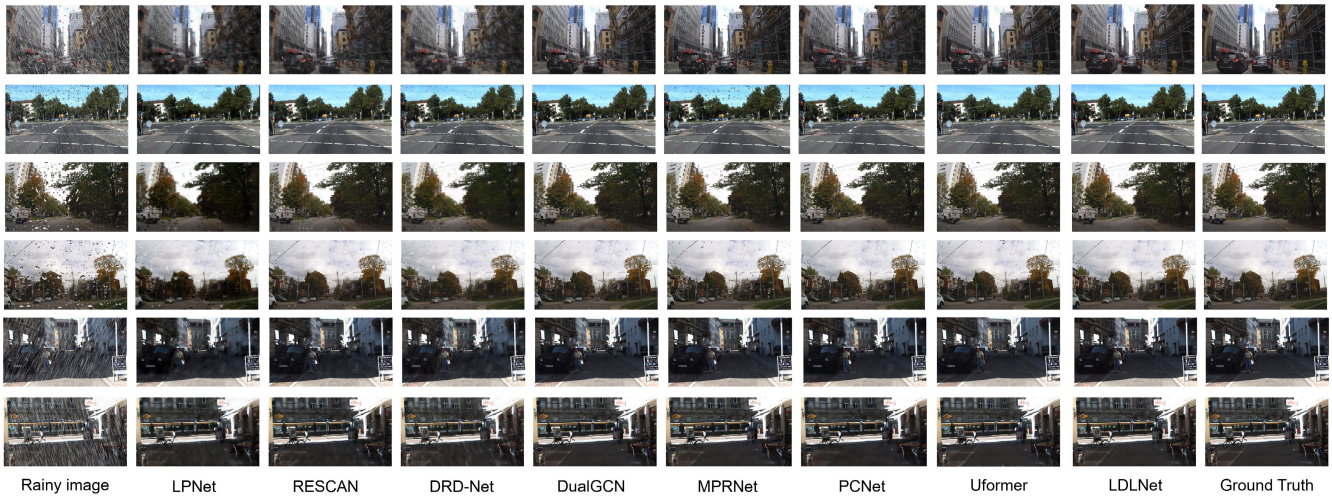


FIGURE 6. Qualitative comparisons of rain removal on the synthesized rain-streak and raindrop dataset.

methods removed the rain well, but could not predict a proper brightness in their outputs. Our de-rained images are closer to the ground truth by satisfying both rain removal and brightness improvement.

Due to the varying sizes, shapes, and intensities of synthesized raindrops and rain streaks in the RainDS dataset, rain removal becomes more challenging. Fig. 6 shows that LDLNet effectively removes the rain from the images, whether they have raindrops, rain streaks, or both. The state-of-the-art models perform well when tested using images with only synthesized rain streaks, as seen in the first two rows of the figure. However, when tested using images with synthesized raindrops, the compared models suffered from removing raindrops completely, as demonstrated in the last four rows of the figure. Their outputs contain raindrops, while LDLNet produces rain-free images.

Rain removal is difficult because of not only rain variations, but also the existence of objects that have similar colors

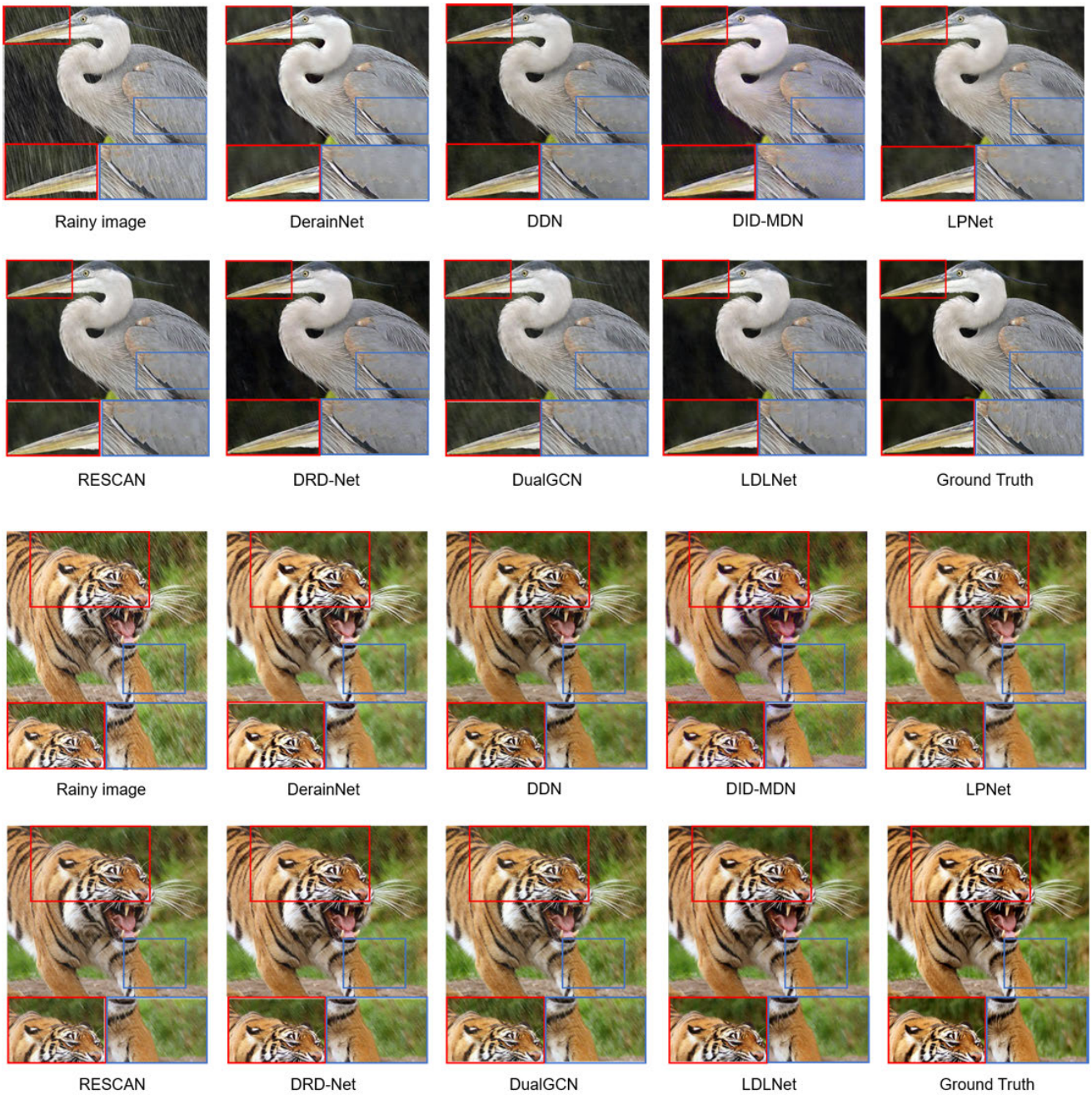


FIGURE 7. Comparisons of object details preservation after rain removal on synthesized rain-streak datasets.

and shapes as rain streaks or raindrops. Fig. 7 demonstrates the object preservation abilities of the deraining models using the first group of datasets. In the figure, the line of the feathers in the bird's wings is similar in color and shape to the rain streaks. It results in either rain and feathers being removed together, or rain cannot be completely removed to maintain the line of the feathers. Thanks to decomposition learning, LDLNet maintains more feather details than the other methods. Both cases demonstrated in the figure reveal that the proposed model neatly removes the rain and outputs images closer to the ground truth. Parts of the pictures were enlarged at the bottom of each image to confirm the details of the output images. Besides rain streak removal, it is certified that LDLNet also produces proper brightness and resolves the detail loss problem in its outputs.

In the case of the second group of datasets, which includes both synthesized rain-streaks and raindrops, three samples with different types of rain are shown in Fig. 8 to illustrate the object detail preservation after rain removal. For the first sample image with the street sign and traffic sign, as it

FIGURE 8. Comparisons of object details preservation after rain removal on the synthesized rain-streak and raindrop dataset.

FIGURE 9. Qualitative comparisons of rain removal on real-world rainy images.

contains only synthesized rain streaks, the deraining results are consistent with those of previous experiments shown in Fig. 7. LDLNet outputs a rain-free image and restores the street name and traffic signs more similar to the ground truth. When raindrops appear in the two other samples, even with or without rain streaks, the proposed model outperforms the other models in removing the rain and preserving the details of the underlying objects. For example, in the second case with only raindrops, LDLNet effectively removes the rain and restores the crosswalk details, while some of the other models struggle to remove the raindrops completely, leaving some artifacts in the output images. In the last sample with both rain streaks and raindrops, LDLNet again outperforms the other models in removing the rain and preserving the details of the underlying objects, such as the windows in the buildings.

To show the practicality of the proposed network, we evaluated its performance on real-world rainy images using the trained model with the first group of datasets. Three samples are illustrated in Fig. 9 for a quality comparison with recent rain removal methods. In the first sample with trees, there are some rain streaks remaining in the derained images using the other methods, but not in the one of LDLNet. We use the second and third samples to show that the proposed model preserves object details even if the objects are similar to the rain in the images. White letters in the second sample are blurred in all derained images of other methods as their color is similar to the rain. In the same manner, the white zipper attached to the woman's outer clothing and the string of the hat are faded in the deraining results due to their shape and color. In contrast, LDLNet only removes the rain streaks and maintains the object details in both cases. The experiments prove that our model produces realistic and reliable derained images. Overall, the qualitative evaluation results align with those of the quantitative evaluation in making the proposed model a more robust and versatile deraining model compared to other models.

The complexity and inference time of LDLNet are compared with SOTA models in Table 6. The models are sorted in ascending order based on their publication times, which span from 2018 to 2022. The table highlights that the proposed LDLNet has a lower number of parameters compared to some of the compared models, such as DRD-Net and Uformer, while still achieving superior performance in terms of rain removal and object detail preservation. This demonstrates the effectiveness of the decomposition learning strategy employed in LDLNet, which allows for better utilization of parameters in single image deraining.

C. ABLATION STUDY

We performed various ablation studies on the Test100 dataset to verify the efficiency of the proposed LDLNet. The performance of SCRB in DGFS part is first evaluated in a comparison with RDB. Then, the contribution of GSM part is verified. Finally, we studied the effects of SRAM module by changing not only the number of modules but also the connections between them.

TABLE 6. Complexity and inference time comparison of deraining models.

	Number of parameters	Inference time
RESCAN [21]	149,823	1.213
DRD-Net [5]	5,230,214	5.406
DualGCN [10]	2,731,071	3.829
MPRNet [34]	3,637,249	2.386
PCNet [18]	627,563	1.272
Uformer [33]	5,290,000	2.675
LDLNet	3,957,987	5.333

TABLE 7. Ablation study on different combinations in DGFS and GFM.

DGFS			GFM	Metrics		
RDB	SCRB	SE	Concatenation	PSNR	SSIM	VIF
✓			✓	23.57	0.8071	0.3401
✓		✓	✓	24.80	0.8333	0.3782
	✓		✓	23.64	0.8149	0.3501
	✓	✓	✓	25.18	0.8407	0.3954
	✓	✓		24.86	0.8273	0.3757

TABLE 8. Ablation study on different numbers of SRAMs.

Number of SRAMs	Metrics		
	PSNR	SSIM	VIF
9 (3×3)	24.44	0.8308	0.3777
15 (3×5)	24.52	0.8370	0.3831
21 (3×7)	25.18	0.8407	0.3954

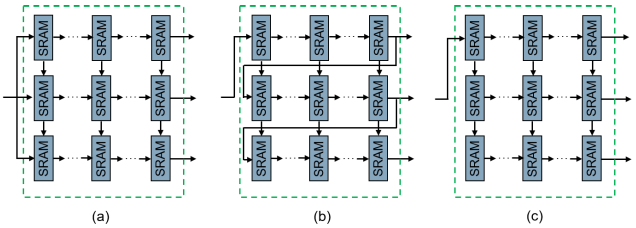


FIGURE 10. Three different connection types used in DGFS: (a) C1 type (b) C2 type (c) C3 type.

1) ABLATION STUDY ON DGFS AND GFM

RDB uses dense connections, which induces overlapping features. To prevent this, SCRB with pruned connection is used in DGFS. Table 7 shows that SCRB with pruned connection has better performance than RDB. Combining SE, which strengthens the correlation between feature channels, with SCRB rather than RDB improves performance. The concatenated connection of GFM preserves the information of DGFS and its efficiency is demonstrated in the table.

2) ABLATION STUDY ON SRAM

The experiment was conducted on the premise that the same number of SRAMs exist in each feature group because connections are consecutively placed between SRAMs located in the same column in the DGFS. Ablation studies proceed only when there are 3, 5, and 7 SRAMs in each feature group. For example, if there are 3 SRAMs, a total of 9 SRAMs exist in DGFS. Table 8 shows that the performance is improved as the number of SRAMs increases.

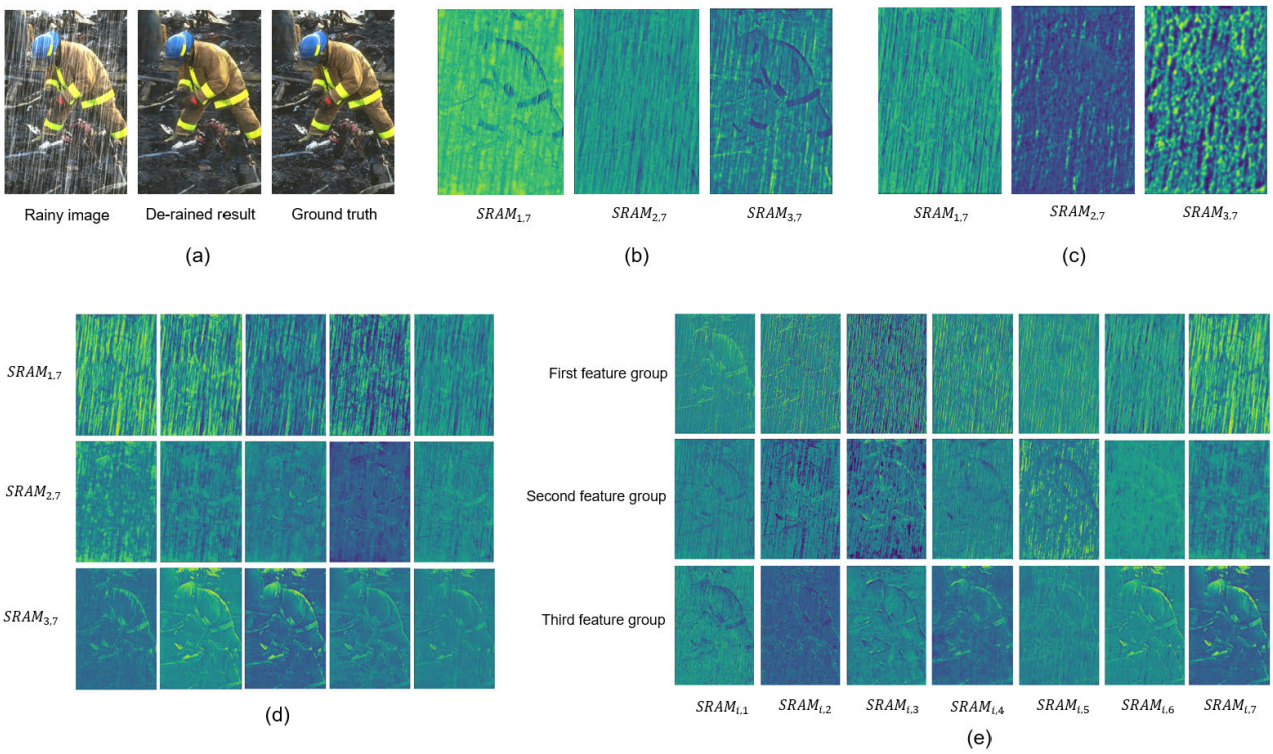


FIGURE 11. Feature map visualization: (a) Sample images (b) Feature maps using C1 type (c) Feature maps using C2 type (d) Feature maps using C3 type (e) Feature maps of all SRAMs in DGFS when C3 is used.

3) ABLATION STUDY ON CONNECTION BETWEEN SRAMS

The initial input and output of each module are concatenated as in Attentive GAN [24], and features in which raindrops and detail coexist are extracted in depth when as they pass through the modules. However, we chose decomposition rather than coexistence to discover details hidden by the rain. For verification, we used three types of connections. Fig. 10(a) shows the C1 type, where the initial input of DGFS is inputted in each feature group. The architecture of DGFS with C2 type contains the connections between the final SRAM of the previous feature group and the first SRAM of the next one as in Fig. 10(b). Finally, the C3 type containing the connection as in LDLNet is presented in Fig. 10(c).

Feature map of the sample image from the Test100 dataset, shown in Fig. 11(a), is visualized in this ablation study. When C1 and C2 are used, the final outputs of each feature group are shown in Fig. 11(b) and (c), respectively. Observably, the feature layers are not well separated with either C1 or C2. The proposed model with C3 creates better separation. Each row in Fig. 11(d) visualizes five feature maps from the final outputs of the three feature groups with C3 type. The feature map of the first group contains rain streak features, while the third feature group focuses on object details. Fig. 11(e) illustrates the feature maps visualizing the output of all SRAMs in the DGFS. By employing the BIFF approach, DGFS decomposes the discriminative and complementary features from the blended ones in the initial SRAM of each feature group. The feature map visualization indicates that using C3 clearly separates discriminative features, which are rain streak and object detail features. Table 9 quantitatively reports that C3 rather than C1 and C2 is the effective connection in terms of performance.

TABLE 9. Ablation study on different connection types.

Connection types			Metrics		
C1	C2	C3	PSNR	SSIM	VIF
✓			24.94	0.8389	0.3854
	✓		24.36	0.8225	0.3599
		✓	25.18	0.8407	0.3954

VI. CONCLUSION

We introduced LDLNet for single-image deraining. The proposed framework decomposes rainy images into discriminative and complementary feature groups representing rain streaks and object details. This decomposition learning operates in the DGFS part using the novel BIFF approach, which aggregates multi-scale context information from connected SRAMs. The decomposed features are preserved through concatenation in the GFM part, contributing to superior quantitative performances of LDLNet in rain removal and image detail preservation. Comparative experiments on both synthesized and real datasets of rainy images demonstrate that our proposed method outperforms the state-of-the-art methods in terms of image quality.

While LDLNet produces superior results in real-world rainy images, there is still room for generality improvement since it is trained primarily with synthetic datasets. Additionally, it can extend to domains like stereo and light field images, which feature dual views and abundant 3D structure with texture information, respectively. In these domains, leveraging the inherent characteristics of stereo images and light field images could enhance rain removal performance [42]. In future work, we plan to employ contrastive learning with real-world rainy images to enhance the robustness of the network and explore the possibility of applying the proposed method to various image modalities.

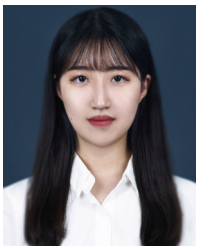
ACKNOWLEDGMENT

(Yunseon Jang and Duc-Tai Le are co-first authors.)

REFERENCES

[1] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3155–3164.
[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," *Int. J. Comput. Vis.*, vol. 93, no. 3, pp. 348–367, Jul. 2011.
[3] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1968–1975.
[4] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11057–11066.
[5] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14548–14557.
[6] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3141–3149.
[7] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2944–2956, Jun. 2017.
[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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1715–1723.
[9] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 6, pp. 1794–1807, Jun. 2020.
[10] X. Fu, Q. Qi, Z. J. Zha, Y. Zhu, and X. Ding, "Rain streak removal via dual graph convolutional network," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2021, pp. 1–9.
[11] K. Garg and S. K. Nayar, "Detection and removal of rain from videos," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2004, pp. 1–12.
[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[13] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[14] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: Criss-cross attention for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 603–612.
[15] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, 2008.
[16] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[17] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8343–8352.
[18] K. Jiang, Z. Wang, P. Yi, C. Chen, Z. Wang, X. Wang, J. Jiang, and C.-W. Lin, "Rain-free and residue hand-in-hand: A progressive coupled network for real-time image deraining," *IEEE Trans. Image Process.*, vol. 30, pp. 7404–7418, 2021.
[19] B. Kamgar-Parsi and A. Rosenfeld, "Optimally isotropic Laplacian operator," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1467–1472, Oct. 1999.
[20] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742–1755, Apr. 2012.
[21] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeezeand-excitation context aggregation net for single image deraining," in *Proc. Eur. Conf. Comput. Vision(ECCV)*, Jul. 2018, pp. 254–269.
[22] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2736–2744.
[23] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3397–3405.
[24] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2482–2491.
[25] R. Quan, X. Yu, Y. Liang, and Y. Yang, "Removing raindrops and rain streaks in one go," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9143–9152.
[26] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[27] M. Suin, K. Purohit, and A. Rajagopalan, "Spatially-attentive patchhierarchical network for adaptive motion deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Aug. 2020, pp. 3606–3615.
[28] F.-J. Tsai, Y.-T. Peng, Y.-Y. Lin, C.-C. Tsai, and C.-W. Lin, "Banet: A blur-aware attention networks for dynamic scene deblurring," *IEEE Trans. Image Process.*, vol. 31, pp. 6789–6799, Oct. 2022, doi: 10.1109/TIP.2022.3216216.
[29] Y. Wang, X. Song, and K. Chen, "Channel and space attention neural network for image denoising," *IEEE Signal Process. Lett.*, vol. 28, pp. 424–428, 2021.
[30] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
[31] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.
[32] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1685–1694.
[33] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general U-shaped transformer for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17683–17693.
[34] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14816–14826.
[35] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 695–704.
[36] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, Nov. 2020.
[37] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[38] X. Zhang, H. Li, Y. Qi, W. Leow, and T. Ng, "Rain removal in video by combining temporal and chromatic properties," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2006, pp. 461–464.
[39] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image superresolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[40] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2021.
[41] Y. Zhang, Z. Zheng, and R. Hu, "Super resolution using segmentation-prior self-attention generative adversarial network," 2020, *arXiv:2003.03489*.
[42] T. Yan, M. Li, B. Li, Y. Yang, and R. W. H. Lau, "Rain removal from light field images with 4D convolution and multi-scale Gaussian process," *IEEE Trans. Image Process.*, vol. 32, pp. 921–936, 2023.
[43] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5514415, doi: 10.1109/TGRS.2023.3284671.



YUNSEON JANG received the B.S. degree in cyber security from Ajou University, South Korea, in 2019, and the M.S. degree in electrical and computer engineering from Sungkyunkwan University (SKKU), South Korea, in 2022.

She specializes in image processing, with a particular focus on image denoising. She has authored several papers on this topic and received the Best Paper Award from the Korea Society of Internet Information (KSII), in 2021. Her current research interests include deep learning, computer vision, and image processing.

DUC-TAI LE received the M.S. degree in computer science from the University of Science, Vietnam National University Ho Chi Minh City, Vietnam, in 2010, and the Ph.D. degree in computer engineering from Sungkyunkwan University, South Korea, in 2016.

He was a Postdoctoral Researcher with Convergence Research Institute, Sungkyunkwan University, from 2016 to 2019. From September 2019 to February 2024, he was a Research Professor with the College of Computing in Informatics, Sungkyunkwan University. Since March 2024, he has been with the Department of Electrical and Computer Engineering, Sungkyunkwan University, as a Research Professor. His current research interests include wireless ad hoc and sensor networks, software defined networks, intelligent networking, and medical image processing.

Dr. Le is a Special Issue Editor of *Electronics* (MDPI) and an Area Editor of *KSII Transactions on Internet and Information Systems*.

CHANG-HWAN SON received the B.S., M.S., and Ph.D. degrees in electronic engineering from Kyungpook National University, Daegu, South Korea, in 2002, 2004, and August 2008, respectively.

During the Ph.D. program, he was selected as a National Scholarship Student, in 2006. After the Ph.D. study, he was with Samsung Electronics, Suwon, South Korea. As a Senior Researcher, he developed color processing algorithms in digital printers. From 2015 to 2017, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, Ryerson University, Canada, where he received full funding from his supervisor Prof. Xiao-Ping Zhang. Since 2017, he has been with the Department of Software Science and Engineering, Kunsan National University, South Korea, where he is currently an Associate Professor and the Director of the Vision and Learning Laboratory. His current research interests include image processing, computer vision, deep learning, machine learning, multispectral processing, and color imaging.

Dr. Son was/is an Associate Editor of IEEE Access and *Journal of Imaging Science and Technology*.

HYUNSEUNG CHOO (Member, IEEE) is a Professor with the College of Computing and Informatics, Sungkyunkwan University (SKKU), South Korea. He is also the National Project Director of the ICT Creative Consilience Program (2020–2029), supported by the Ministry of Science and ICT (MIST); and the Operation Chair of the SDN/NFV Industrial Forum with Korea Association of Network Industries. He has held several notable positions, including the Director of the Intelligent HCI Convergence Research Center, Ministry of Knowledge Economy, from 2005 to 2013; the Priority Research Centers Program, Ministry of Education, Science, and Technology, from 2010 to 2019; and the Grand ICT Research Center, Ministry of Science, ICT and Future Planning, from 2015 to 2022. He also served as a Technical Adviser for Next-Generation Interaction with the Research and Development Center, Samsung Electronics. His research interests include network softwarization, intelligent mobile and edge computing, and medical image processing. He has published over 350 papers in international journals and major conferences, and holds 30 U.S. patents and 243 Korean patents in these fields.

He is a member of ACM and IEICE. His research achievements have been recognized among the Top 100 National Research and Development Excellence Achievements by the Ministry of Education, in 2005 and 2010, and by MIST, in 2019. He has received several awards, including the Excellence Awards and Commendations from MIST, and the Bronze Best Paper Award from Samsung, in 2023. He is actively involved in several academic societies, serving as the Director for KIISE, KSII, KPIS, and HCI Korea. He served as the Editor-in-Chief for *Journal of Internet Computing and Services*; and an Editor for *ACM Transactions on Internet Technology*, *Journal of Communications and Networks*, and *The Journal of Supercomputing*. Since 2010, he has been the Founding Editor of *KSII Transactions on Internet and Information Systems*.