ARTICLE

Joint Rain Streaks & Haze Removal Network for Object Detection

Ragini Thatikonda[1], Prakash Kodali[1,*], Ramalingaswamy Cheruku[2] and Eswaramoorthy K.V[3]

[1]Department of Electronics and Communication Engineering, National Institute of Technology, Warangal, 506004, India

[2]Department of Computer Science and Engineering, National Institute of Technology, Warangal, 506004, India

[3]Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Design and Manufacturing, Kurnool, 518008, India

*Corresponding Author: Prakash Kodali. Email: kprakash@nitw.ac.in

ABSTRACT

In the realm of low-level vision tasks, such as image deraining and dehazing, restoring images distorted by adverse weather conditions remains a significant challenge. The emergence of abundant computational resources has driven the dominance of deep Convolutional Neural Networks (CNNs), supplanting traditional methods reliant on prior knowledge. However, the evolution of CNN architectures has tended towards increasing complexity, utilizing intricate structures to enhance performance, often at the expense of computational efficiency. In response, we propose the Selective Kernel Dense Residual M-shaped Network (SKDRMNet), a flexible solution adept at balancing computational efficiency with network accuracy. A key innovation is the incorporation of an M-shaped hierarchical structure, derived from the U-Net framework as M-Network (M-Net), within which the Selective Kernel Dense Residual Module (SDRM) is introduced to reinforce multi-scale semantic feature maps. Our methodology employs two sampling techniques-bilinear and pixel unshuffled and utilizes a multi-scale feature fusion approach to distil more robust spatial feature map information. During the reconstruction phase, feature maps of varying resolutions are seamlessly integrated, and the extracted features are effectively merged using the Selective Kernel Fusion Module (SKFM). Empirical results demonstrate the comprehensive superiority of SKDRMNet across both synthetic and real rain and haze datasets.

KEYWORDS

Image deraining; Selective Dense Residual Module (SDRM); Selective Kernel Fusion Module (SKFM); Selective Kernel Dense Residual M-Shaped Network (SKDRMNet)

Nomenclature

CV	Computer Vision
CNN	Convolutional Neural Networks
GPU	Graphics Processing Unit
RDM	Residual Dense Module
RRM	Residual Resize Module
SDRM	Selective Dense Residual Module

SKFM   Selective Kernel Fusion Module
SKDRMNet   Selective Kernel Dense Residual M-shaped Network
PSNR   Peak Signal to Noise Ratio
SSIM   Structural Similarity Index
LPIPS   Learned Perceptual Image Patch Similarity
FSIM   Feature Similarity Index Metric
AMBE   Absolute Mean Brightness Error
BRISQUE   Blind/Referenceless Image Spatial Quality Evaluator
NIQE   Natural Image Quality Evaluator

1 Introduction

Image restoration, a foundational task in low-level Computer Vision (CV), has attracted significant research interest in recent years. In practical outdoor scenarios, environmental factors such as rain, raindrops, and haze inevitably degrade the quality of captured images. This degradation has a ripple effect on the accuracy of subsequent high-level CV tasks like object detection [1–3]. Thus, our primary objective is to develop a network capable of producing noise-free images from degraded inputs. This network serves as a critical pre-processing stage for various downstream high-level CV applications. The image restoration process targets the removal of noise pixels from images captured by highway or airport cameras on rainy or hazy days. These images are represented as 'O', consisting of clean background ('B') and noise patterns ('N'), where O = B + N. While many restoration networks effectively eliminate noise, some encounter difficulties in extracting rain or haze features. Moreover, prolonged transmission processes can lead to the loss of crucial features, thereby impacting overall performance.

In the realm of single image deraining and dehazing networks, three primary categories have emerged: Filter-based, prior-knowledge-based, and CNN-based networks. Filter-based networks, demonstrated in references [4–6], exploit the low and high-frequency characteristics of image data to eliminate noise. However, they often lack discrimination, applying the same operation uniformly to both rain streaks (or haze) and intricate image features, potentially leading to loss of detail. Networks falling within the prior-knowledge category, as exemplified by references [7–10], leverage sparsity, randomness, and low-rank properties inherent in the data during image restoration. While effective in removing rain streaks or haze, these networks may inadvertently introduce colour distortion in the restored image. The third category harnesses CNNs, showcased in architectures [11–15], to address these challenges with data-driven approaches, leveraging the network's ability to learn intricate patterns and features from large datasets.

CNN-based networks, often heralded as state-of-the-art (SOTA) solutions, frequently incorporate a variety of sophisticated structures aimed at enhancing performance in deraining or dehazing tasks. These structures encompass a range of techniques, including residual learning [16,17] attention mechanisms [18,19] dense connections [20,21] channel attention modules [22,23] residual dense modules [24] and hierarchical architecture styles [25–27]. The integration of these advanced CNN-based structures signifies a prevailing trend toward harnessing complex neural network architectures to bolster noise removal and image restoration capabilities.

Existing networks predominantly concentrate on either image deraining or dehazing tasks and commonly grapple with issues such as limited diverse receptive fields and weak supervision signals. Moreover, SOTA approaches frequently confront challenges like sluggish convergence, excessive parameterization, high computational costs, and prolonged inference times. These difficulties arise

from the necessity of a substantial number of parameters, resulting in intricate and complex network structures. Furthermore, striking a balanced trade-off between noise removal and preservation of essential image details remains a critical challenge.

This study aims to achieve a harmonious balance between computational efficiency and accuracy within our network, while also addressing multiple low-level vision tasks. Our approach begins by adapting the Residual Dense Module (RDM) into the Selective Dense Residual Module (SDRM) and introducing a selective residual M-hierarchy architecture. Additionally, we incorporate two distinct sampling techniques, bilinear and pixel un-shuffled [16] along with a multi-scale feature fusion strategy. At the critical reconstruction stage, we depart from traditional concatenation methods, opting instead for the Selective Kernel Fusion Module (SKFM) [27] to seamlessly combine the extracted feature maps at multi-scales. This strategic adjustment is specifically designed to efficiently yield refined features, thereby contributing to a more effective trade-off between computational efficiency and the network's overall accuracy.

This work presents several significant contributions:

1. We introduce the Selective Dense Residual Module (SDRM), a highly efficient feature extraction module tailored specifically for low-level CV tasks.
2. A novel M-Net hierarchy architecture is proposed to effectively address the challenges posed by heavy rain streaks, raindrops, and haze in noisy images.
3. Through extensive experimentation, we demonstrate that our SKDRMNet achieves comparable performance in both quantitative and qualitative assessments. Notably, it operates with reduced inference time and requires fewer parameters compared to other SOTA networks.
4. The superiority of our network is further validated across various low-level haze removal tasks, as well as high-level CV tasks such as object detection. This showcases the versatility and enhanced performance of our approach in diverse scenarios.

2 Related Work

This section offers an overview of the research endeavours undertaken in the realm of image deraining and dehazing over the past decade. While video-deraining networks [28–30] demonstrate promising results in deraining tasks; they often encounter efficiency challenges when faced with heavy rain streaks or haze, primarily due to the absence of temporal information between adjacent frames. This limitation has prompted a notable shift in focus among researchers towards the domain of image restoration.

2.1 Single Image Deraining Networks

Researchers have extensively explored various deep learning methodologies to tackle the removal of both light and heavy rain streaks. The classification of single image deraining networks hinges on whether the network derives all the rain model information, resulting in two distinct categories: White-box architectures and black-box architectures.

2.1.1 White-Box Architectures

A network tasked with deriving all the rainfall model information falls under the category of white-box network architecture [20,21]. However, it is important to note that these networks typically necessitate either ground truth data or prior knowledge for effective operation. In response to this, Zhang et al. [21] proposed a density-aware densely connected CNN. Their approach considers the

shape and distinct scale properties of rain streaks, integrating this knowledge into a multi-stream deep CNN tailored specifically for rain streak removal. Additionally, to address the challenging nature of the image deraining problem, a novel network called the Conditional General Adversarial Network (CGAN) was introduced [31]. Furthermore, in [32], Wang et al. developed a rain information model by accounting for the alignment structures of vapours and rain streaks.

In an effort to bridge the divide between synthetic and real-world image deraining challenges, a dual-stage neural network was introduced [33,34]. This innovative approach aimed to mitigate the disparities between synthetic and real images by integrating physical properties of the environment, such as atmospheric light, rain streaks, and transmission, into the image deraining process. However, it is recognized that while many of these networks demonstrate proficiency in removing rain streaks with specific shapes, they may encounter difficulties when confronted with rain streaks of varying shapes.

2.1.2 Black-Box Architectures

In contrast to white-box architectures, black-box architectures [35–37] eschew explicit learning of the entire rain model information, instead harnessing the black-box capabilities of deep CNNs during training for robust fitting. Li et al. [7] introduced the Deep Detail Network (DDN), drawing inspiration from ResNet [16,38] to effectively remove rain streaks using deep learning techniques. Their approach involved designing a deep detail CNN tailored to reconstruct noise-free images, showcasing the potential of black-box architectures in addressing image deraining challenges.

Lee et al. [39] introduced the K-NN Image Transformer (KiT), pioneering a novel attention mechanism for image restoration. These advancements exemplify innovative strategies within deraining networks, effectively tackling various challenges in rain removal and image restoration. In response to the limitations of convolution models in capturing long-range dependencies and removing complex rainy artifacts, an efficient transformer-based network was proposed [40]. This approach introduces locality and hierarchy into the network architecture, resulting in superior deraining outcomes, albeit at the expense of consuming more parameters to generate a clean image. Addressing the relative independence between background and rain streak features often overlooked by existing SOTA networks in the feature domain; the Pyramid Feature Decoupling Network (PFDN) [41] was introduced. This network aims to remove noise from rain images while preserving detailed background image information. Additionally, Liu et al. [42] presented the Context-Aware Recurrent Multi-stage Network (ReCMN) for removing heavy rain streaks and producing noise-free images. Leveraging a multi-stage strategy, this network effectively models contextual relationships among features. However, it should be noted that this network requires more runtime to generate a derained image.

2.2 Single Image Dehazing Networks

The single-image dehazing (SID) task presents a significant challenge as it is inherently ill-posed, primarily due to the limited information available to establish the haze model. Compounding this challenge is the non-linear nature of the haze model, which amplifies sensitivity to noise and adds further complexity to the problem. In response to these inherent difficulties, researchers have proposed various strategies aimed at addressing these challenges and improving the efficacy of dehazing algorithms over time. Various approaches have been introduced by researchers to tackle the challenge of SID. For instance, a GridNet [22], based on CNNs, was proposed to remove artifacts from hazy images. In another study, an attention-based CNN network was developed to directly restore haze-free images by integrating feature-attention modules, a local residual learning mechanism, and attention-based feature fusion mechanisms [43]. Additionally, Wu et al. [44] employed a contrastive

learning-based auto-encoder framework to leverage both clear and hazy images, aiming to reduce memory storage and enhance the network's performance. However, it is worth noting that this network may struggle to remove dense haze from certain hazy images.

In their work detailed in [45], researchers combined transformer and CNN architectures for image dehazing, aiming to address the inconsistency between the two by employing a learning modulation technique to modulate CNN features. Another approach, proposed by Wu et al. [46], introduced a real dehazing network utilizing codebook priors rather than relying solely on a physical scattering model. In a separate study [47], a recurrent structure-based CNN dehazing network was designed with the goal of restoring haze-free images. Additionally, Wang et al. [48] presented a dual-stage image dehazing network that also takes into account low-frequency haze characteristics. To enhance texture details and edges in the image, they devised a frequency stream-based CNN attention octave network. However, it should be noted that this approach overlooks the low similarity between noise and noise-free haze components. In a recent study [49], a novel recurrent feedback structure was introduced to enhance dehazing performance. This approach leverages deep feature dependencies across various stages, aiming to effectively mitigate haze effects. Additionally, a pioneering end-to-end self-regularization triple unfolding framework was proposed [50]. This framework strategically capitalizes on the distinct characteristics of different haze and clear components. By doing so, it ensures that the extracted features possess superior generalization capabilities.

Our approach is designed to address multiple low-level CV tasks, ultimately aiming to improve the detection accuracy of subsequent high-level vision tasks. Prioritizing simplicity and efficiency, our network endeavours to strike a balance between performance and computational resource utilization. By doing so, we aim to enhance the effectiveness and generalization capability of low-vision tasks.

3 Proposed Methodology

This section focuses primarily on the proposed SKDRMNet architecture, providing a detailed explanation of the various components utilized in our network.

3.1 Proposed Architecture

Fig. 1 depicts the proposed SKDRMNet architecture, designed for the highly effective restoration of noise-free images. The process initiates with the input rainy image $Y \in R^{HxWxC}$, where a 3×3 weight-sharing convolution extracts low-level features, followed by bilinear down sampling of the original input. Selective Dense Residual Modules (SDRM) then extract high-level semantic feature maps from each layer, incorporating the pixel-unshuffled technique for down-sampling at the end of SDRM, resulting in multi-scale feature maps.

To enhance the capture of semantic information across cascaded features, our network integrates two down-sampling variants: Bilinear and pixel-unshuffled methods. Shallow features extracted from the previous stage, obtained through bilinear down-sampling, are subsequently concatenated with feature maps in line with the U-Net process. This strategic integration ensures the preservation and incorporation of valuable information at different scales throughout the network architecture. Subsequently, in the reconstruction phase aimed at generating a noise-free image, multi-scale feature maps are seamlessly integrated using the SKFM. A 3×3 convolution is applied, yielding a residual image $R \in R^{HxWx3}$. The final restored image is obtained through the equation $\hat{Y} = Y + R$. Importantly, the post-convolution maintains a fixed number of 72 input channels, contributing to the network's efficiency and overall effectiveness in image restoration.
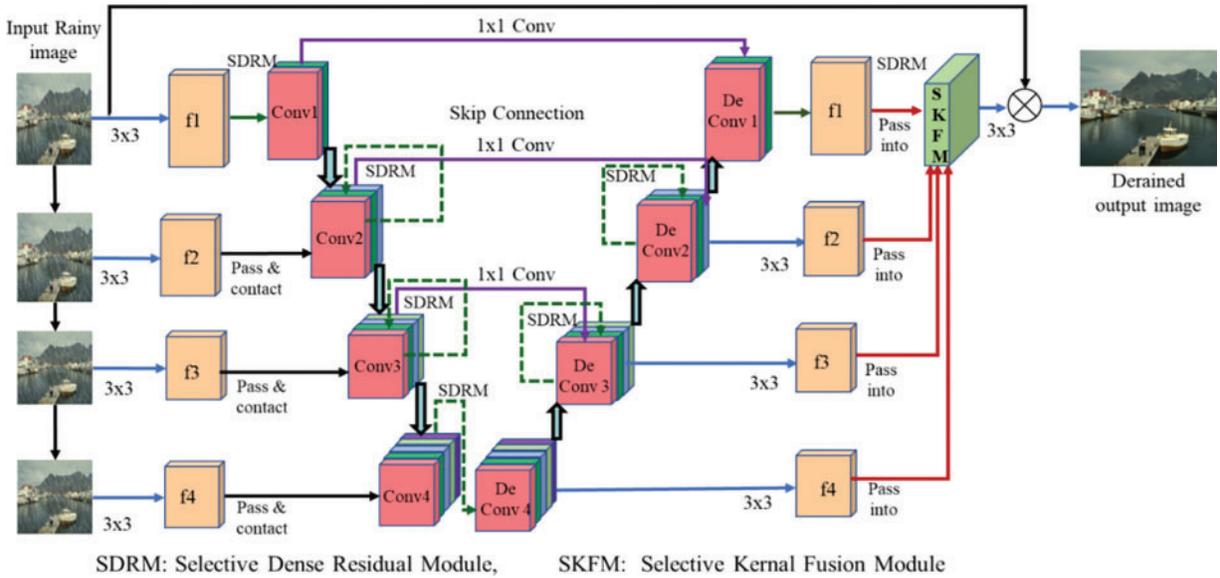
Figure 1: Selective kernel dense residual M-shaped network (SKDRMNet) architecture

Our SKDRMNet is optimized in an end-to-end manner, and for all low-level CV tasks, we utilize the Charbonnier loss function to restore noise-free images. The formulation is as follows:

$$L_{char} = \sqrt{\left\| \hat{Y} - Y \right\|^{2.} + \varepsilon^2} \qquad (1)$$

where $\hat{Y}$, $Y \in R^{H \times W \times C \times B}$ denotes restored images and ground-truth images. H, W, C & B represents height of image, width of image, feature channels count and batch size during network training. In Eq. (1) the constant ε was empirically set to 0.001. Now we provide in-detail information of various sub-components used in our network.

We introduce Residual Resize Module (RRM) to perform bilinear down-sampling and pixel-unshuffled operations while maintaining the residual nature in our network. To resize the input images $Y \in R^{H \times W \times C}$ (Y is the same as Eq. (1)); we apply bilinear down-sampling. Likewise, to resize shallow features following convolution, we employ the pixel-unshuffled technique. This ensures that while the size of the original image is downscaled by two, the dimensions of input and output feature channels remain consistent. An observation reveals that the lower layer feature map encompasses both pixel-unshuffled and bilinear features. However, it is noted that the quality of bilinear features is comparatively inferior to that of pixel-unshuffled features, leading to an imbalance in the feature map. To rectify this imbalance, these feature maps undergo processing through the SDRM, which amplifies the weights of bilinear features. This adjustment effectively mitigates the unbalanced feature map problem, thereby enhancing the overall quality and balance of the feature representation.

3.1.1 M-Network (M-Net)

Traditional U-Net architectures encounter challenges in effectively enriching contextual feature map information across varying image resolutions and addressing spatial loss induced by down-sampling in encoder stages. Given that low-level CV tasks are inherently pixel-wise vision tasks, the

utilization of max-pooling for down-sampling in encoder stages proves inappropriate as it results in a significant loss of spatial feature map information. To overcome these limitations, M-Net architecture offers a solution by integrating additional gatepost feature paths in both encoder and decoder stages. Originally devised for medical image segmentation [51], M-Net has since been successfully adapted for image denoising [25,26], super-resolution [52] demonstrating its efficacy in mitigating these drawbacks and enhancing the performance of low-level CV tasks.

To enhance M-Net for image restoration, we introduced two improvements and incorporated them into our SKDRMNet. Firstly, we aimed to increase the diversity of multi-scale cascading features. In the original implementation [51], both gatepost and U-Net paths utilized 2×2 max pooling, resulting in combined feature maps. In SKDRMNet, we adopted bilinear down-sampling in the gatepost path and pixel un-shuffle down-sampling in the U-Net path, thereby enhancing diversity in the cascaded features. Secondly, we addressed the challenge of summarizing information in decoders, particularly concerning high-dimensional cascading features in shallow layers. While the original M-Net [51] employed techniques such as changing input image size, batch normalization, and reducing input feature map dimensions, these methods often increased parameters and computational complexity. To mitigate these drawbacks, SKDRMNet utilizes the SKFM [27] to aggregate weighted features instead of concatenating each feature map separately during the reconstruction process. These design modifications contribute to the generation of clean images, enhancing generalization capability across different environmental conditions and ultimately improving the detection accuracy of object detectors [3]. This adaptation renders M-Net suitable for image restoration, offering improved efficiency and performance.

3.1.2 Selective Dense Residual Module (SDRM)

In our network, the improved Residual Dense Module (RDM) [52] has been enhanced and is now referred to as the Selective Dense Residual Module (SDRM). The architectures of both the SKFM and SDRM are depicted in Figs. 2a and 2b, respectively. The SDRM incorporates dynamic adjustments through two key operations: Fuse and select. The fuse operator functions by combining information from dual streams, thereby generating global feature descriptors. Subsequently, the select operator utilizes these descriptors to recalibrate feature maps, followed by their aggregation. These operations play a crucial role in enhancing the adaptability and effectiveness of the SDRM in capturing and integrating multi-scale feature information for improved image restoration performance.

In our SDRM structure, we use 3 residual modules and each residual module takes two input features $f_{res}, f_{ms} \in R^{H \times W \times C}$ representing residual and mainstream features, respectively. One pathway employs a conventional 3×3 convolution kernel to extract features, while the other pathway employs varying kernel sizes (such as 5×5 and 7×7) to acquire a larger receptive field. Initially, element-wise addition is used to combine two multi-scale features, followed by Global Average Pooling (GAP) across spatial dimensions to compute channel-wise statistics $s \in R^{1 \times 1 \times C}$. Next, a channel downscaling convolution is applied to generate a compact feature map representation $z \in R^{1 \times 1 \times r}$, where $r = 0.125$ C. Finally, the feature vector (z) passes through two parallel channel up-scaling convolution layers, providing corresponding feature descriptors $v_1, v_2 \in R^{1 \times 1 \times C}$. Select: The select operator applies the Softmax function to v_1 and v_2 feature descriptors, yielding attention activations s_1 and s_2 that are used to adaptively recalibrate multi-scale channel weighted feature maps $\hat{f}_{res}, \hat{f}_{ms}$, respectively. The overall feature recalibration and aggregation of every residual module is defined as $f_0 = s_1 . \hat{f}_{res} + s_2 . \hat{f}_{ms}$. After three residual modules, a 3×3 convolution is applied, and a long-skip connection along with a 1×1 convolution is inserted between the input and output to reconstruct a noise-free image.
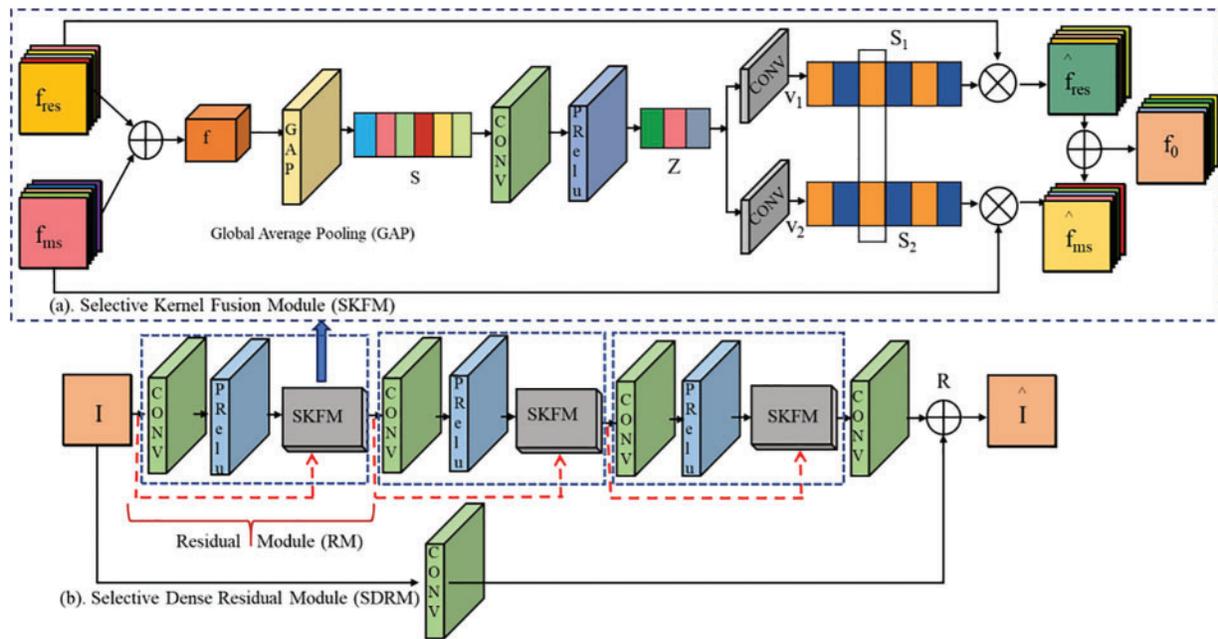
Figure 2: (a) Selective Kernel Fusion Module (SKFM) (b) Selective Dense Residual Module (SDRM)

4 Experimental Results

This section offers a comprehensive overview of several synthetic and real-world benchmark datasets utilized for various low-level CV tasks. Additionally, we provide detailed implementation procedures and present experimental results for both deraining and dehazing.

4.1 Datasets

In the real world, capturing both noisy and noise-free image pairs simultaneously is not feasible. Therefore, we trained and tested SKDRMNet using synthetic rain, raindrop, and haze benchmark datasets. Table 1 summarizes the synthetic and real-world datasets utilized in our network, specifically tailored for rain, raindrop and haze tasks. We also verified the effectiveness of SKDRMNet on other low-level vision tasks such as image dehazing.

Table 1: Summary of used datasets in our network

Datasets	Train images	Test images	Renamed test set
Rain12 [7]	12	0	NC
Rain14000 [17]	11200	2800	Test2800
Rain100L [20]	0	100	Rain100L
Rain100H [20]	0	100	Rain100H
Rain1800 [20]	1800	0	NC
Rain1200 [30]	0	1200	Test1200
Rain800 [31]	700	100	Test100
RealRain200*	0	200	RR200

(Continued)

Table 1 (continued)

Datasets	Train images	Test images	Renamed test set
RIS [53]	0	2400	NC
RESIDE [54]	5000	1000	RESIDE6K
Dense-Haze [55]	50	5	NC
NH-HAZE [56]	45	5	NC

Note: *- we collected real-rain images from internet.

4.2 Implementation Details

The proposed network underwent training on Google Colab Pro+ with a single Tesla V100 GPU, utilizing PyTorch 1.8.0 for implementation. To bolster the robustness of the network, various data augmentation techniques such as vertical, rotation, and horizontal flips were applied. Training commenced from scratch in an end-to-end manner, employing hyperparameters including a patch size of 256×256, Adam optimizer, batch size of 8, weight decay set at 0.0005, momentum at 0.9, and a variable learning rate ranging from 0.0002 to 0.00001 and trained for 200 epochs.

4.3 Image Deraining

4.3.1 Quantitative Evaluation

We conducted thorough experiments to systematically evaluate the deraining performance of our network across multiple benchmark synthetic rain datasets. Our experimental analysis involved comparing the results obtained from our network with those of 13 mainstream SOTA networks, including RESCAN [11], MPRNet [12], DIDMDN [21], SEMI [24], PReNet [37] among others. To quantitatively assess deraining performance, we computed the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) scores on the recovered images, with higher PSNR and SSIM scores indicating superior performance. Table 2 provides a comprehensive overview of the quantitative results obtained by various existing SOTA networks on standard synthetic rain datasets. The comparative analysis reveals that our network consistently achieved improved PSNR and SSIM scores when compared to current mainstream networks, highlighting its effectiveness in image deraining task.

Table 2: Comparative analysis of mainstream SOTA networks on synthetic rain datasets

Datasets	Test2800 [17]	Rain100L [20]	Rain100H [20]	Test1200 [30]	Test100 [31]	Avg. PSNR/
Networks/Metrics	PSNR/SSIM↑	PSNR/SSIM↑	PSNR/SSIM↑	PSNR/SSIM↑	PSNR/SSIM↑	Avg. SSIM↑
RESCAN [11]	31.29/0.90	29.80/0.88	26.36/0.78	30.51/0.88	25.01/0.83	28.59/0.85
MPRNet [12]	31.64/0.93	36.69/0.97	27.65/0.87	31.73/0.91	27.86/0.85	30.98/0.90
SAPNet [13]	31.18/0.93	34.77/0.97	29.46/0.89	32.46/0.91	29.13/0.88	31.60/0.91
DGUNet [14]	31.68/0.93	37.42/0.96	30.40/0.89	33.11/0.92	30.32/0.90	32.98/0.92
DIDMDN [21]	28.13/0.86	25.23/0.74	17.35/0.52	29.65/0.90	22.56/0.82	24.58/0.77
SEMI [24]	24.38/0.73	22.25/0.84	18.08/0.57	23.91/0.71	20.72/0.68	21.87/0.70
ECNet [34]	31.42/0.93	33.42/0.95	27.91/0.86	30.05/0.90	27.55/0.88	30.27/0.90
MSPNet [35]	31.82/0.93	32.44/0.95	28.66/0.86	32.39/0.91	27.50/0.87	30.76/0.74
PReNet [37]	31.75/0.91	32.44/0.95	26.77/0.85	31.36/0.91	24.81/0.85	29.42/0.89
KiT [39]	31.25/0.94	36.56/0.97	30.46/0.89	32.81/0.92	30.26/0.90	32.79/0.92

(Continued)

Table 2 (continued)

Datasets	Test2800 [17]	Rain100L [20]	Rain100H [20]	Test1200 [30]	Test100 [31]	Avg. PSNR/
Networks/Metrics	PSNR/SSIM↑	PSNR/SSIM↑	PSNR/SSIM↑	PSNR/SSIM↑	PSNR/SSIM↑	Avg. SSIM↑
IDT [40]	31.38/0.93	37.01/0.97	29.95/0.90	31.38/0.91	29.69/0.90	32.28/0.92
PFDN [41]	32.63/0.93	37.10/0.98	30.13/0.89	33.74/0.93	31.88/0.91	33.09/0.92
ReCMN [42]	32.46/0.93	37.64/0.98	30.51/0.89	32.85/0.93	31.97/0.91	33.08/0.93
Proposed	**32.88/0.94**	**39.63/0.98**	**30.56/0.90**	**33.78/0.93**	**32.36/0.92**	**33.84/0.93**

4.3.2 Qualitative Evaluation

Fig. 3 visually demonstrates the efficacy of our network in effectively removing heavy rain streaks, showcasing its robustness and generalizability in image deraining tasks. The images utilized for this visual comparison were sourced from the Rain100H [20] synthetic dataset. In Fig. 3, specific regions within the testing images are selected and magnified to provide a closer inspection of the deraining results. This visual representation allows for a clear and intuitive assessment of the performance of our network in mitigating the adverse effects of heavy rain streaks, thereby reinforcing the quantitative findings presented in the previous section.
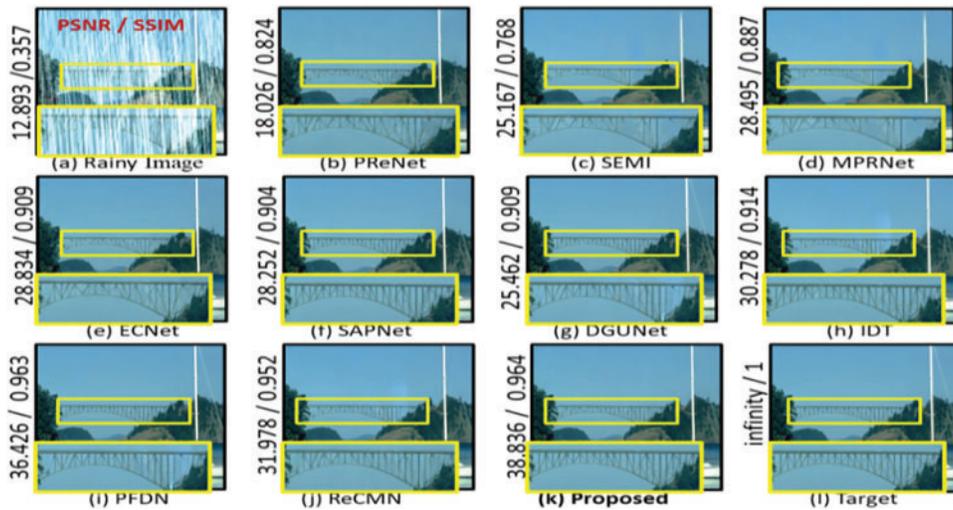

Figure 3: Visual results of existing SOTA networks and our network on Rain100H dataset

Upon closer scrutiny of the magnified local area, it becomes evident that PReNet [37] demonstrates limited effectiveness in generating a rain-free image, despite achieving a low PSNR score on the Rain100H dataset. Conversely, networks such as MPRNet [12], SAPNet [13], DGUNet [14], ECNet [34] and PFDN [41] exhibit successful removal of heavy rain streaks, albeit with potential drawbacks such as introducing varying degrees of blur in the background and deficiencies in retaining background details in the restored image. Our network surpasses other mainstream networks in terms of both PSNR and SSIM scores, particularly excelling in completely eliminating complex heavy rain streaks. Additional visual results obtained from comparisons on the Rain100L [20] dataset with existing SOTA networks are presented in Fig. 4. This visual analysis offers further insight into the performance of our

network relative to other competing methods, reinforcing its efficacy in tackling challenging deraining tasks.
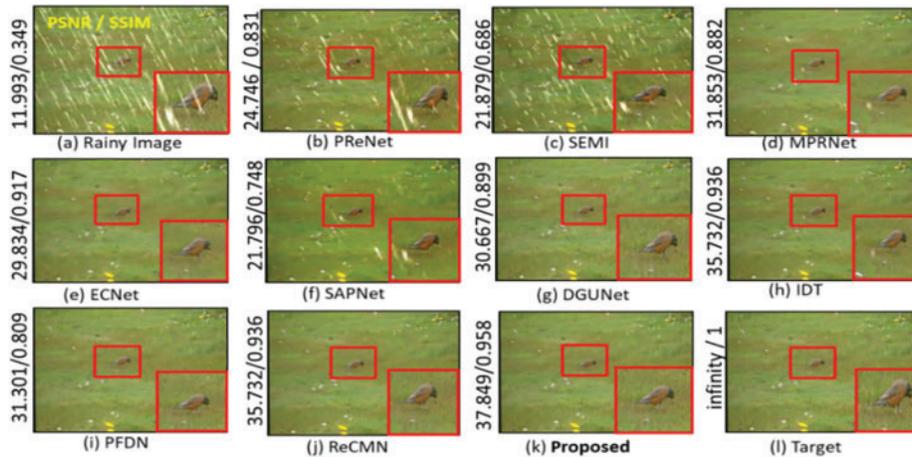

Figure 4: Visual results of existing SOTA networks and our network on Rain100L dataset

Fig. 5 illustrates our network's capability to generate more natural-looking images when compared to other existing mainstream networks on real-world rain datasets. For this purpose, we collected 200 real rain images from the internet and designated them as RealRain200. Upon closer examination of zoomed-in local details, it becomes apparent that other SOTA networks struggle to produce clean images, resulting in partial texture distortion and blurred backgrounds. Even MPRNet [12], despite its prominence, generates images with noticeable blurring and colour distortion. Through visual comparisons on real-world rain datasets conducted in this study, our network demonstrates its superiority in recovering rain streaks while retaining more texture details. Furthermore, Table 3 provides a comparison of SOTA networks with other non-reference evaluation metrics such as BRISQUE and NIQUE on the RealRain200 dataset, offering further insights into the performance of our network in real-world rain image deraining task.

Table 3: Comparison of SOTA networks on RealRain200 dataset

Networks/ Metrics	Rainy images	RESCAN [11]	MPRNet [12]	SAPNet [13]	ECNet [34]	PReNet [37]	IDT [40]	PFDN [41]	ReCMN [42]	Proposed
BRISQUE↓	28.43	28.29	28.31	27.43	27.69	26.82	27.78	27.64	28.16	**26.04**
NIQE↓	5.12	4.67	4.77	4.92	4.74	4.69	4.81	4.83	4.78	**4.65**

4.4 Image Dehazing

4.4.1 Quantitative Evaluation

To assess the effectiveness of our network, we conducted a comprehensive performance comparison with 10 mainstream SOTA dehazing networks, employing both quantitative and qualitative evaluations. Tables 4 and 5 present the qualitative comparison of our network with these counterparts on hazy images, assessing their performance across three key metrics: PSNR, SSIM (distortion metrics), Feature Similarity Index Metric (FSIM), Learned Perceptual Image Patch Similarity (LPIPS), and Absolute Mean Brightness Error (AMBE) (perceptual metrics).
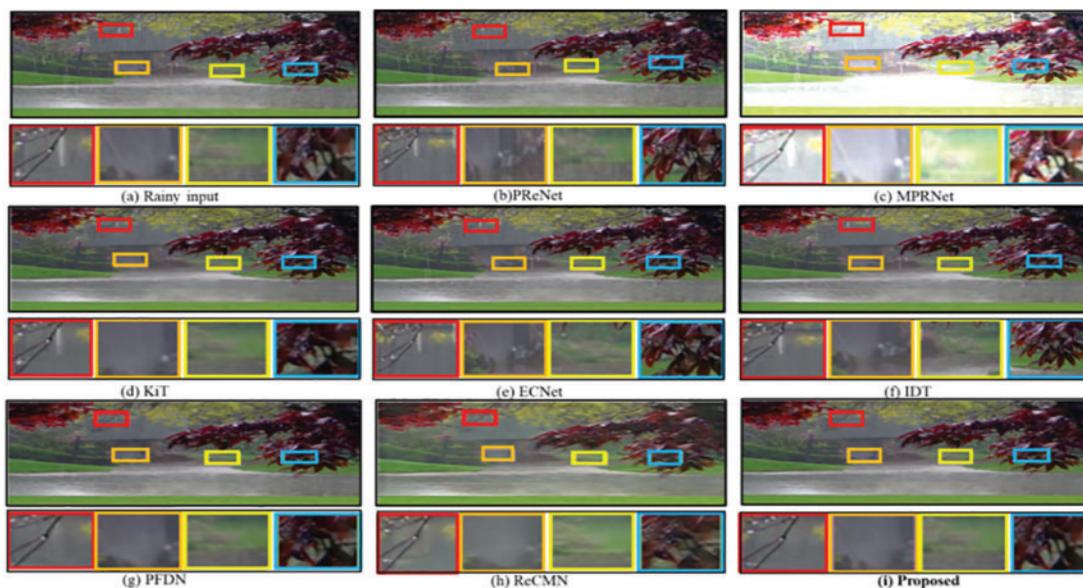
Figure 5: Visual deraining results of RealRain200 dataset and comparison with other networks

Table 4: Comparison of qualitative results of several existing SOTA dehazing networks

| Dataset	RESIDE6K [54]				
Networks/Metrics	PSNR↑	SSIM↑	FSIM↑	LPIPS↓	AMBE↓
GridNet [22]	17.524	0.849	0.873	0.623	30.89
FFA-Net [23]	19.182	0.864	0.912	0.487	21.46
AECR-Net [44]	20.237	0.885	0.951	0.464	19.42
IDHT [45]	21.949	0.906	0.948	0.451	16.13
FSDGN [46]	21.745	0.898	0.957	0.446	12.24
RIDCP [47]	23.816	0.901	0.849	0.478	14.57
FCDSNet [48]	23.351	0.904	0.896	0.347	8.46
FSADNet [49]	24.073	0.923	0.921	0.392	10.89
TSURNet [50]	23.426	0.917	0.957	0.305	7.83
Proposed	25.294	0.926	0.966	0.238	5.42

The results indicate that our network consistently outperforms existing SOTA networks in terms of both distortion and perceptual metrics on RESIDE6K [54], Dense-Haze [55], and NH-HAZE [56] test datasets. These experimental findings underscore the efficacy of SKDRMNet in effectively handling these challenging datasets, reaffirming its superiority in dehazing tasks.

Table 5: Comparison of qualitative results of several existing SOTA dehazing networks

Dataset	Dense-Haze [55]			NH-HAZE [56]		
Networks/Metrics	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
GridNet [22]	13.24	0.359	0.512	13.64	0.526	0.303
FFA-Net [23]	14.27	0.451	0.496	19.75	0.681	0.364
AECR-Net [44]	15.73	0.458	0.527	19.77	0.702	0.269
IDHT [45]	16.57	0.546	0.469	20.18	0.673	0.225
FSDGN [46]	16.82	0.578	0.425	19.84	0.705	0.221
RIDCP [47]	8.16	0.406	0.542	12.12	0.497	0.346
FCDSNet [48]	17.24	0.743	0.404	18.97	0.695	0.214
FSADNet [49]	15.79	0.567	0.426	18.63	0.639	0.232
TSURNet [50]	17.86	0.797	0.376	20.85	0.717	0.183
Proposed	**19.13**	**0.93**	**0.315**	**22.17**	**0.741**	0.164

4.4.2 Qualitative Evaluation

Visual comparisons between our proposed network and several mainstream SOTA networks were conducted on the Dense-Haze [55] and NH-HAZE [56] datasets, yielding insightful observations. In Figs. 6 and 7, it becomes evident that many existing mainstream networks encounter challenges in effectively removing non-homogeneous and dense haze. These techniques often suffer from significant information loss in image contents and colour fidelity. In contrast, our network demonstrates superior performance by effectively eliminating nearly all haze, preserving intricate background details in the restored image, and generating higher fidelity dehazed results. Furthermore, our network consistently produces outcomes that closely resemble the ground truth image, showcasing its ability to faithfully reconstruct the dehazed images. These visual comparisons serve to underscore the effectiveness and reliability of our proposed network in addressing complex haze removal tasks.
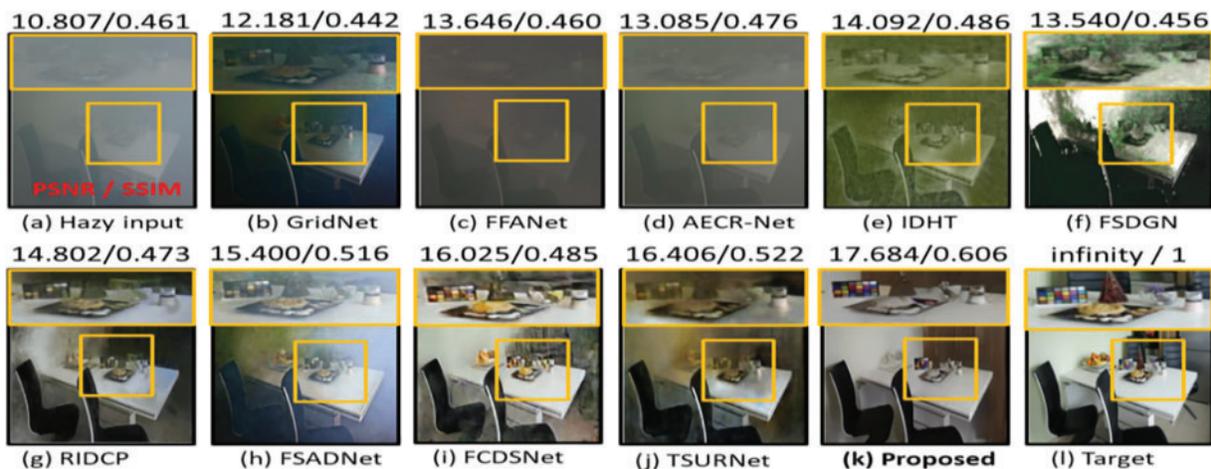
Figure 6: Visual results of several mainstream SOTA networks on Dense-Haze dataset
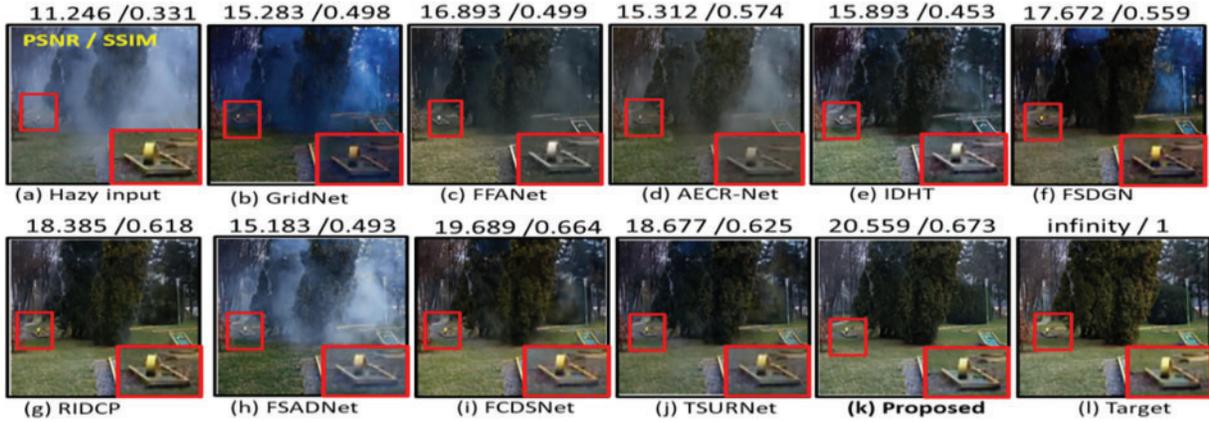
Figure 7: Visual results of several mainstream SOTA networks on NH-HAZE dataset

4.5 Resource Efficient

In our comparison with the latest SOTA networks on the Rain100L [20] dataset, our network emerges as the top performer, achieving a superior PSNR score while demanding fewer computational resources. This efficiency is evidenced by factors such as reduced Floating Point Operations per Second (FLOPS), parameters, and run-time requirements. Table 6 provides a detailed computational complexity comparison with various SOTA networks. By outperforming these networks in terms of PSNR while maintaining lower computational demands, our network demonstrates its effectiveness in image deraining tasks without compromising efficiency. This highlights the practical utility and superiority of our proposed approach in real-world applications where computational resources are often constrained.

Table 6: Comparison of computational complexity in terms of PSNR, FLOPS, parameters, and runtime (milliseconds) of several networks on 512×512 images

Networks/Metrics	PSNR	FLOPS (G)	Parameters	Runtime (ms)
MPRNet [12]	36.69	572.77	36,37,303	128.66
SAPNet [13]	34.77	**36.25**	27,49,214	200.04
ECNet [34]	33.42	212.59	8,93,56,24	271
MSPNet [35]	32.44	708.44	35,64,216	212.84
IDT [40]	37.01	61.8	16,44,68,59	283
PFDN [41]	37.1	246.3	24,46,72,36	212
ReCMN [42]	37.64	209.7	13,68,43,84	104
Proposed	**39.63**	185.26	**3,66,527**	**70.19**

Our network imposes a computational burden of 185.26 GFLOPs, learns 0.366 million parameters, and produces a derained image within a swift 70.19 milliseconds (ms). Compared to PFDN [41], SKDRMNet boasts 0.75 fewer FLOPs. Additionally, our network achieves a notable +1.99 dB higher PSNR score than ReCMN [42], while demanding approximately 37 times fewer parameters and running ~1.5 times faster. Notably, our network exhibits lower runtimes compared to existing

SOTA networks, performing approximately three times faster than PFDN [41], four times faster than IDT [40], and ~2.8 times faster than SAPNet [13]. Furthermore, our network excels in eliminating unnecessary noise, yielding rain-free images that closely resemble the ground truth. This outcome underscores the efficiency of our network, which demonstrates the least computational complexity among existing SOTA networks while delivering superior deraining performance.

4.6 Application in High-Level CV Task

Single image deraining stands as a pivotal low-level CV task aimed at improving the precision of subsequent high-level CV tasks. When autonomous vehicles capture images under adverse weather conditions, such as heavy rain or haze, they often struggle to accurately locate and detect objects amidst these challenging environments. Factors like fluctuating lighting conditions, variations in viewpoint, scale discrepancies of objects, and inherent noise in captured images collectively diminish the accuracy of existing object detectors. As a result, addressing image deraining becomes imperative for enhancing the overall effectiveness of object detection systems in complex scenarios.

In our final evaluation, we gauge the performance of SKDRMNet concerning a high-level vision task, specifically object detection, employing SSAD [1] and YOLOv8 [2] as our chosen object detectors. The detection accuracy (%mAP@0.5) of state-of-the-art (SOTA) image deraining networks, when paired with SSAD [1] and YOLOv8 [2] on the RIS [53] dataset, is presented in Table 7. The individual detection accuracies of SSAD and YOLOv8 stand at 27.84% and 31.42%, respectively, indicating suboptimal performance due to the presence of rainy artifacts in the testing images. However, when combined with SKDRMNet, the detection accuracy improves to 28.92% for SSAD and 32.56% for YOLOv8. This enhancement of 1.08% and 1.14%, respectively, signifies the effectiveness of SKDRMNet in facilitating more accurate object detection. Consequently, SKDRMNet serves as a valuable pre-processing stage for existing object detectors, significantly enhancing their capability to detect objects accurately amidst adverse weather conditions.

Table 7: Detection results (mAP@0.5) of several SOTA networks on RIS [53] dataset

Networks	Rainy images	RESCAN [11]	MPRNet [12]	SAPNet [13]	ECNet [34]	PReNet [37]	IDT [40]	PFDN [41]	ReCMN [42]	Proposed
SSAD [1]	27.84	23.92	24.51	25.36	24.67	24.18	26.6	27.03	27.14	**28.92**
YOLOv8 [2]	31.42	26.73	28.13	28.9	28.51	28.06	28.9	30.83	31.07	**32.56**

4.7 Limitations

While our network outperforms existing SOTA networks, it faces limitations in scenarios where images exhibit a combination of rain and haze, low-light conditions with rain, and dense haze. These scenarios pose challenges as our network lacks the capability to effectively address such complex environmental conditions. Particularly, it struggles to adequately remove haze from real rainy images, leading to derained outputs that remain unclear. This limitation stems from the absence of training data containing a mixture of rain and haze, underscoring a significant constraint of our current network architecture. Addressing this challenge requires exploring avenues for collecting real-world datasets that encompass both rain and haze, facilitating the development of more resilient models capable of handling diverse environmental conditions.

4.8 Ablation Study

The comprehensive evaluation of each component within our network architecture is conducted through meticulous ablation studies, focusing on the Rain100L dataset [20]. The robustness of our network is meticulously assessed and presented in Table 8. Notably, the significance of skip connections is highlighted by the results, demonstrating a notable decrease in the PSNR score upon their removal. Skip connections play a pivotal role in facilitating rapid convergence and mitigating larger training errors, thereby contributing significantly to the network's performance. Moreover, the incorporation of the SKFM at the higher end proves beneficial, facilitating effective information transfer among parallel convolution streams and leading to enhanced performance outcomes. Furthermore, the replacement of the U-Net architecture with our refined M-Net structure yields a substantial improvement in the PSNR score, demonstrating a notable increase from 34.39 to 35.48 dB. This enhancement underscores the efficacy of integrating our modified M-Net architecture into the network design, further validating its effectiveness.

Table 8: Impact of each component on our network

Skip connections	N	Y	N	N	Y	Y	Y	Y
Bilinear down-sampling	Y	N	Y	Y	N	N	N	Y
SKFM intermediate	Y	N	N	Y	Y	Y	N	N
SKFM final	N	Y	Y	N	N	N	Y	Y
SDRM	N	Y	N	Y	N	N	Y	Y
Pixel unshuffled	N	N	Y	N	Y	Y	Y	Y
M-Net	N	Y	Y	N	Y	Y	Y	Y
Residual resize module	Y	N	Y	Y	Y	Y	Y	Y
PSNR (dB)	34.39	35.48	36.25	37.06	37.69	38.14	38.57	**39.63**

In our comparative analysis presented in Table 9, the efficacy of our SKFM is evident, showcasing superior performance when compared to other methods. Notably, SKFM achieves this superior performance while utilizing ~6× fewer parameters compared to concatenation operations, highlighting its efficiency and resource-effectiveness. This demonstrates the robustness and practicality of SKFM as a critical component within our network.

Table 9: Feature aggregation strategy. SKFM uses ~6× lesser parameters than concatenation

Method	Summation	Concatenation	SKFM
PSNR	37.42	38.49	**39.63**
Parameters	0	12,188	**2048**

5 Conclusion

In this study, we introduce SKDRMNet as a versatile solution that not only achieves SOTA results in image deraining and dehazing but also addresses the crucial balance between computational efficiency and network accuracy. At the core of SKDRMNet lies the M-shaped network architecture,

a pivotal component that significantly enhances feature enrichment by integrating multi-resolutions, surpassing traditional RDM's in efficacy. Our comprehensive experiments underscore the network's proficiency across various low-level vision tasks, highlighting its potential to make significant contributions to the broader field of CV. Nevertheless, challenges persist, particularly in optimizing PSNR and SSIM scores for images with substantial rain. Real-world scenarios, such as images with dense rain or a combination of rain and haze, continue to present on-going challenges that necessitate further exploration and refinement. Looking ahead, the exploration of additional low-level vision tasks, such as desnowing and deblurring, remains an area of interest for future research endeavours.

Author Contributions: Study conception and design: Ragini Thatikonda, Prakash Kodali; data collection, analysis, and interpretation of results: Ragini Thatikonda, Prakash Kodali, Ramalingaswamy Cheruku; draft manuscript preparation: Ramalingaswamy Cheruku, Eswaramoorthy K.V. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data openly available in a public repository.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] B. M. Chintakindi and M. F. Hashmi, "SSAD: Single-shot multi-scale attentive detector for autonomous driving," *IETE Tech. Rev.*, vol. 40, no. 6, pp. 793–805, 2023. doi: 10.1080/02564602.2023.2176932.

[2] G. Jocher, A. Chaurasia, and J. Qiu, "Yolo by ultralytics (version 8.0.190)," GitHub, 2023. Accessed: Apr. 1, 2024. [Online]. Available: https://github.com/ultralytics/ultralytics.

[3] C. B. Murthy, M. F. Hashmi, M. Ghulam, and S. A. AlQahtani, "YOLOv2PD: An efficient pedestrian detection algorithm using improved YOLOv2 model," *Comput. Mater. Contin.*, vol. 69, no. 3, pp. 3015–3031, 2021. doi: 10.32604/cmc.2021.018781.

[4] X. Ding, L. Chen, X. Zheng, Y. Huang, and D. Zeng, "Single image rain and snow removal via guided L0 smoothing filter," *Multimed. Tools Appl.*, vol. 75, no. 5, pp. 2697–2712, 2016. doi: 10.1007/s11042-015-2657-7.

[5] J. Xu, W. Zhao, P. Liu, and X. Tang, "Removing rain and snow in a single image using guided filter," presented at the 2012 IEEE Int. Conf. Comp. Sci. Autom. Eng. (CSAE), Zhangjiajie, China, May 25–27, 2012.

[6] X. Zheng, Y. Liao, W. Guo, X. Fu, and X. Ding, "Single-image-based rain and snow removal using multi-guided filter," presented at the 2013 Neural Inform. Process. 20th Int. Conf. (ICONIP), Daegu, Korea, Nov. 3–7, 2013.

[7] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," presented at the 2016 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Las Vegas, NV, USA, Jun. 27–30, 2016.

[8] Y. L. Chen and C. T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," presented at the 2013 Proc. IEEE Int. Conf. Comp. Vis. (ICCV), Washington DC, USA, Dec. 1–8, 2013.

[9] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," presented at the 2018 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Salt Lake City, UT, USA, Jun. 18–23, 2018.

[10] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, 2010.

[11] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," presented at the 2018 Proc. Euro. Conf. Comp. Vis. (ECCV), Munich, Germany, Sep. 8–14, 2018.

[12] S. W. Zamir *et al.*, "Multi-stage progressive image restoration," presented at the 2021 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Nashville, TN, USA, Jun. 20–25, 2021.

[13] S. Zheng, C. Lu, Y. Wu, and G. Gupta, "SAPNet: Segmentation-aware progressive network for perceptual contrastive deraining," presented at the 2022 Proc. IEEE Win. Conf. App's. Comp. Vis., Waikoloa, HI, USA, Jan. 4–8, 2022.

[14] C. Mou, Q. Wang, and J. Zhang, "Deep generalized unfolding networks for image restoration," presented at the 2022 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), New Orleans, LA, USA, Jun. 18–24, 2022.

[15] S. Cao, L. Liu, L. Zhao, Y. Xu, J. Xu and X. Zhang, "Deep feature interactive aggregation network for single image deraining," *IEEE Access*, vol. 10, pp. 103872–103879, 2022. doi: 10.1109/ACCESS.2022.3210190.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," presented at the 2016 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Las Vegas, NV, USA, Jun. 27–30, 2016.

[17] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," presented at the 2017 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Honolulu, Hawaii, USA, Jul. 21–26, 2017.

[18] N. Jiang, W. Chen, L. Lin, and T. Zhao, "Single image rain removal via multi-module deep grid network," *Comput. Vis. Image Underst.*, vol. 202, no. 2–3, pp. 103106, 2021. doi: 10.1016/j.cviu.2020.103106.

[19] R. Yasarla, J. M. J. Valanarasu, and V. M. Patel, "Exploring overcomplete representations for single image deraining using cnns," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 2, pp. 229–239, 2020. doi: 10.1109/JSTSP.2020.3039393.

[20] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo and S. Yan, "Deep joint rain detection and removal from a single image," presented at the 2017 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Honolulu, Hawaii, USA, Jul. 21–26, 2017.

[21] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," presented at the 2018 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Salt Lake City, UT, USA, Jun. 18–23, 2018.

[22] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," presented at the 2020 Proc. AAAI Conf. Artif. Intel., New York, USA, Apr. 3–5, 2020.

[23] N. Jiang, J. Luo, J. Lin, W. Chen, and T. Zhao, "Lightweight semi-supervised network for single image rain removal," *Pattern Recognit.*, vol. 137, no. 9, pp. 109277, 2023. doi: 10.1016/j.patcog.2022.109277.

[24] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu, "Semi-supervised transfer learning for image rain removal," presented at the 2019 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Long Beach, CA, USA, Jun. 15–20, 2019.

[25] V. S. Adiga and J. Sivaswamy, "FPD-M-net: Fingerprint image denoising and in painting using M-net based convolutional neural networks," *Inpainting Denoising Challenges*, vol. 8, pp. 51–61, 2019.

[26] C. M. Fan, T. J. Liu, K. H. Liu, and C. H. Chiu, "Selective residual M-net for real image denoising," presented at the 2022 30th Euro. Signal Process. Conf. (EUSIPCO), Belgrade, Serbia, Aug. 29–Sep. 2, 2022, pp. 469–473.

[27] S. W. Zamir *et al.*, "Learning enriched features for real image restoration and enhancement," presented at the 2020 16th Euro. Conf. Comp. Vis. (ECCV), Glasgow, UK, Aug. 23–28, 2020.

[28] J. H. Kim, J. Y. Sim, and C. S. Kim, "Video deraining and desnowing using temporal correlation and low-rank matrix completion," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2658–2670, 2015. doi: 10.1109/TIP.2015.2428933.

[29] K. Garg and S. K. Nayar, "Detection and removal of rain from videos," presented at the 2004 Proc. IEEE Comp. Society Conf. Comp. Vis. Pattern Rec. (CVPR), Washington DC, USA, Jun. 27– Jul. 02, 2004.

[30] J. Liu, W. Yang, S. Yang, and Z. Guo, "D3R-Net: Dynamic routing residue recurrent network for video rain removal," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 699–712, 2018. doi: 10.1109/TIP.2018.2869722.

[31] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 30, no. 11, pp. 3943–3956, 2019. doi: 10.1109/TCSVT.2019.2920407.

[32] Y. Wang, Y. Song, C. Ma, and B. Zeng, "Rethinking image deraining via rain streaks and vapors," presented at the 2020 Proc. 16th Euro. Conf. Comp. Vis. (ECCV), Glasgow, UK, Aug. 23–28, 2020.

[33] R. Li, L. F. Cheong, and R. T. Tan, "Heavy rain image restoration: Integrating physics model and conditional adversarial learning," presented at the 2019 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Long Beach, CA, USA, Jun. 15–20, 2019.

[34] Y. Li, Y. Monno, and M. Okutomi, "Single image deraining network with rain embedding consistency and layered lstm," presented at the 2022 Proc. IEEE Winter Conf. Apps Comp. Vis. (CACV), Waikoloa, HI, USA, Jan. 3–8, 2022.

[35] K. Jiang *et al.*, "Multi-scale progressive fusion network for single image deraining," presented at the 2020 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Seattle, WA, USA, Jun. 13–19, 2020.

[36] 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, 2017. doi: 10.1109/TIP.2017.2691802.

[37] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," presented at the 2019 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Long Beach, CA, USA, Jun. 15–20, 2019.

[38] C. Chen and H. Li, "Robust representation learning with feedback for single image deraining," presented at the 2021 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Nashville, TN, USA, Jun. 20–25, 2021.

[39] H. Lee, H. Choi, K. Sohn, and D. Min, "KNN local attention for image restoration," presented at the 2022 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), New Orleans, LA, USA, Jun. 18–24, 2022.

[40] J. Xiao, X. Fu, A. Liu, F. Wu, and Z. J. Zha, "Image deraining transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 12978–95, 2022.

[41] Q. Wang, G. Sun, J. Dong, and Y. Zhang, "PFDN: Pyramid feature decoupling network for single image deraining," *IEEE Trans. Image Process.*, vol. 31, pp. 7091–7101, 2022. doi: 10.1109/TIP.2022.3219227.

[42] Y. Liu *et al.*, "Recurrent context-aware multi-stage network for single image deraining," *Comput. Vis. Image Underst.*, vol. 227, no. 6, pp. 103612, 2023. doi: 10.1016/j.cviu.2022.103612.

[43] W. Yang, R. T. Tan, S. Wang, Y. Fang, and J. Liu, "Single image deraining: From model-based to data-driven and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4059–4077, 2020. doi: 10.1109/TPAMI.2020.2995190.

[44] H. Wu *et al.*, "Contrastive learning for compact single image dehazing," presented at the 2021 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Nashville, TN, USA, Jun. 20–25, 2021.

[45] C. L. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, "Image dehazing transformer with transmission-aware 3d position embedding," presented at the 2022 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), New Orleans, LA, USA, Jun. 18–24, 2022.

[46] R. Q. Wu, Z. P. Duan, C. L. Guo, Z. Chai, and C. Li, "RIDCP: Revitalizing real image dehazing via high-quality codebook priors," presented at the 2023 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Vancouver, BC, Canada, Jun. 17–24, 2023.

[47] H. Yu, N. Zheng, M. Zhou, J. Huang, Z. Xiao and F. Zhao, "Frequency and spatial dual guidance for image dehazing," presented at the 2022 Euro. Conf. Comp. Vis. (ECCV), Tel Aviv, Israel, Oct. 23–27, 2022.

[48] M. Wang, L. Liao, D. Huang Z. Fan, J. Zhuang and W. Zhang, "Frequency and content dual stream network for image dehazing," *Image Vis. Comput.*, vol. 139, no. 9, pp. 104820, 2023. doi: 10.1016/j.imavis.2023.104820.

[49] Y. Zhou, Z. Chen, P. Li, H. Song, C. L. P. Chen and B. Sheng, "FSAD-Net: Feedback spatial attention dehazing network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 10, pp. 7719–7733, 2022. doi: 10.1109/TNNLS.2022.3146004.

[50] X. Song *et al.*, "TUSR-Net: Triple unfolding single image dehazing with self-regularization and dual feature to pixel attention," *IEEE Trans. Image Process.*, vol. 32, pp. 1231–1244, 2023.

[51] R. Mehta and J. Sivaswamy, "M-net: A convolutional neural network for deep brain structure segmentation," presented at the 2017 IEEE 14th Int. Sym. Biomed. Imag. (ISBI), Melbourne, Australia, Apr. 18–21, 2017.

[52] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," presented at the 2018 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Salt Lake City, UT, USA, Jun. 18–23, 2018.

[53] S. Li *et al.*, "Single image deraining: A comprehensive benchmark analysis," presented at the 2019 Proc. IEEE Conf. Comp. Vis. Pattern Rec. (CVPR), Long Beach, CA, USA, Jun. 15–20, 2019.

[54] B. Li *et al.*, "Benchmarking single image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, 2018. doi: 10.1109/TIP.2018.2867951.

[55] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-Haze: A benchmark for image dehazing with Dense-Haze and haze-free images," presented at the 2019 IEEE Int. Conf. Image Pro. (ICIP), Taipei, Taiwan, Sep. 22–25, 2019.

[56] C. O. Ancuti, C. Ancuti, and R. Timofte, "NH-HAZE: An image dehazing benchmark with non-homogeneous hazy and haze-free images," presented at the 2020 Proc. EEE/CVF Conf. Com. Vis. Pattern Rec. Workshops (CVPRW), Seattle, WA, USA, Jun. 14–19, 2020.