Signal, Image and Video Processing (2024) 18:7131–7143
https://doi.org/10.1007/s11760-024-03380-1

ORIGINAL PAPER

Optimized RainDNet: an efficient image deraining method with enhanced perceptual quality

Debesh Kumar Shandilya[1] · Spandan Roy[1] · Navjot Singh[1]

Received: 25 April 2024 / Revised: 24 May 2024 / Accepted: 14 June 2024 / Published online: 27 June 2024

Abstract
RainDNet is an advanced image deraining model that refines the "Multi-Stage Progressive Image Restoration Network" (MPRNet) for superior computational efficiency and perceptual fidelity. RainDNet's innovative architecture employs depthwise separable convolutions instead of MPRNet's traditional ones, reducing model complexity and improving computational efficiency while preserving the feature extraction ability. RainDNet's performance is enhanced by a multi-objective loss function combining perceptual loss for visual quality and Structural Similarity Index Measure (SSIM) loss for structural integrity. Experimental evaluations demonstrate RainDNet's superior performance over MPRNet in terms of Peak Signal-to-Noise Ratio (PSNR), SSIM, and BRISQUE (Blind Referenceless Image Spatial Quality Evaluator) scores across multiple benchmark datasets, underscoring its aptitude for maintaining image fidelity while restoring structural and textural details. Our findings invite further explorations into more efficient architectures for image restoration tasks, contributing significantly to the field of computer vision. Ultimately, RainDNet lays the foundation for future, resource-efficient image restoration models capable of superior performance under diverse real-world scenarios.

Keywords Image restoration · Rain removal · Deep learning · Multi-stage progressive restoration · Depthwise separable convolutions · Perceptual loss

1 Introduction

Restoring images, a process that transforms a degraded picture back to its high-quality state continues to be a substantial obstacle within the realm of computer vision. This degradation could be in the form of noise, haze, blur, or even elements like rain streaks, among other factors. This task poses a high degree of difficulty as the solution space is practically infinite, with countless plausible restorations for any given degraded image. Traditional restoration techniques [1–7] have relied on explicitly defined image priors, handcrafted based on empirical observations. Such an approach, however, is inherently limited due to the difficulty of designing these priors and their lack of generalizability across different scenarios.

With the advancement of machine learning technologies, recent cutting-edge techniques have turned towards the use of convolutional neural networks (CNNs) [8–16]. CNNs offer the ability to implicitly learn more adaptable and general image priors by analyzing the statistical characteristics of natural images across extensive datasets.

The superior yield achieved by CNN-based approaches is mostly due to the meticulous design and integration of numerous network elements and functional segments established for image reinstatement [8, 11, 15, 17–23].

Inspired by the success of the MPRNet architecture [27], we introduce a derivative model, named RainDNet, which further advances the image restoration task, specifically for de-raining. The MPRNet [27] model incorporated an innovative multi-stage design [30–33], contrasting with the traditional single-stage architectures prevalent in low-level vision

Debesh Kumar Shandilya, Spandan Roy and Navjot Singh have contributed equally to this work.

✉ Navjot Singh
navjot@iiita.ac.in

Debesh Kumar Shandilya
rsi2022504@iiita.ac.in

Spandan Roy
spandanroy2@gmail.com

[1] Computer Vision and Biometrics Lab, Department of Information Technology, Indian Institute of Information Technology Allahabad, Jhalwa, Prayagraj, Uttar Pradesh 211015, India

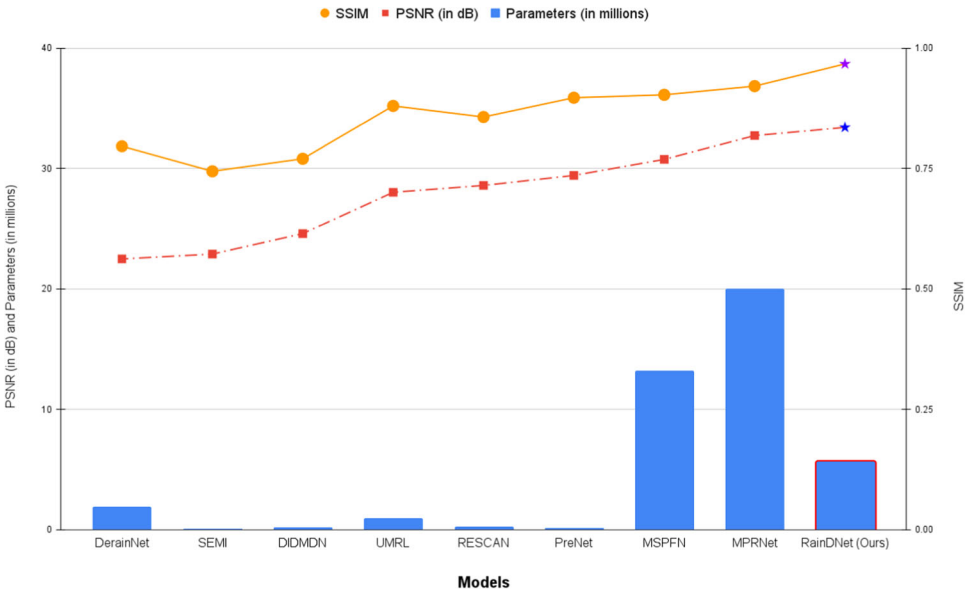


Fig. 1 Image deraining on the Rain100H [18], Rain100L [18], Test100 [24], Test1200 [25] and Test2800 [26] datasets. Despite having a lot fewer parameters than the underlying model, our optimised Multi-stage technique outperforms the original cutting-edge MPRNet [27–29] in terms of PSNR and SSIM, suggesting greater picture restoration quality from a human visual perception standpoint

tasks [30, 34–36]. Our RainDNet model preserves this powerful multi-stage design and introduces further refinements.

To reduce the computational complexity and memory footprint of the network, RainDNet employs depthwise separable convolutions [37–42], a variant of standard convolutions that decouples the learning of spatial and depth-wise features. This change significantly reduces the number of parameters in the model without sacrificing performance. Furthermore, RainDNet enhances the original loss function of MPRNet [27] by integrating perceptual and SSIM losses in addition to the standard L1 and edge losses. These additions promote the preservation of perceptual quality and structural similarity in the restored images, thus further improving the visual quality of the derained outputs.

In this paper, we conduct an extensive evaluation of the proposed RainDNet model. Our comparative (Fig. 1) study shows that RainDNet can mostly achieve better PSNR values than the original MPRNet [27] while significantly surpassing it in terms of SSIM values, thus offering an improved trade-off between accuracy and perceptual quality. Our findings present RainDNet as a new promising approach for deraining tasks, opening up avenues for future improvements in the field of image restoration.

The key contributions presented in this study include:

- We introduce a unique framework, RainDNet, for the restoration of images. This design draws inspiration from the multi-stage structure of MPRNet [27] and incorporates depthwise separable convolutions [37–42] to lessen the computational demand.
- We introduce a modified loss function that incorporates perceptual and SSIM losses in addition to L1 and Edge losses, enhancing the quality of restored images.
- We demonstrate the effectiveness of our model by conducting comprehensive experiments, comparing our model with the cutting-edge MPRNet [27] on multiple datasets. Our results mostly exhibit better PSNR performance, significantly better SSIM results, and better BRISQUE results, thus confirming the efficacy of our approach.

2 Related works

Throughout the last few decades, image-capturing technology has seen a significant transformation. We are transitioning from traditional high-end DSLR cameras towards more compact and user-friendly smartphone cameras. Early restoration approaches were anchored in mathematical and empirical methods like total variation [6, 43], sparse coding [44–46], self-similarity [47, 48], and gradient prior [49, 50]. These methods relied heavily on handcrafted features and were not always generalizable to diverse image degradation scenarios.

Convolutional neural networks (CNNs) in image restoration With the advent of deep learning, the focus shifted towards CNNs. CNN-based restoration methods have outperformed traditional methods [10, 11, 13, 15, 51, 52], providing a more robust and generalizable approach to image restoration. Among CNN-based methods, single-stage approaches currently dominate the field. They often repurpose architectural components developed for high-level vision tasks.

The advent of depthwise separable convolutions Another significant evolution in the field of deep learning is the

introduction of depthwise separable convolutions [37–42], an efficient variant of standard convolutions. This efficient approach has been adopted in a variety of domains, showing great promise in enhancing the performance and efficiency of deep learning models.

Multi-stage approaches In contrast to the prevalent single-stage methods, multi-stage approaches [51, 53–58] aim to tackle the image restoration problem in a more structured manner. These methods progressively restore the clean image by incorporating a lightweight subnetwork at each stage. However, one common practice in these methods that could lead to suboptimal results is the use of identical subnetworks for each stage.

The use of attention mechanisms A more recent innovation in deep learning that has found its way into the domain of image restoration is the attention mechanism [51, 54, 55]. These modules record extensive mutual dependencies along spatial [59] and channel [60] dimensions, allowing for better context-aware processing of features [61].

Introduction of RainDNet In the current work, we introduce a novel variant of the well-established Multi-Stage Progressive Restoration Network (MPRNet [27]) for image deraining. In our proposed model, RainDNet, we replace some of the standard convolutions with depthwise separable convolutions [37–42], yielding a significant reduction in the model's parameter count. Furthermore, we introduce perceptual and structural similarity (SSIM) losses in addition to the conventional L1 and edge losses, contributing to the model's enhanced performance in capturing perceptually important image details. The revamped model, RainDNet, exhibits superior Peak Signal-to-Noise Ratio (PSNR) values and Structural Similarity Index Measure (SSIM) values when compared with the existing MPRNet [27] model. This indicates that our proposed solution while being more computationally efficient, does not compromise the image restoration performance.

3 Gradual multi-stage enhancement

The framework we propose for the restoration of images, depicted in Fig. 2, plays out in three steps to gradually polish the images. As in the predecessor's architecture, the first pair of phases leverage encoder–decoder sub-networks to acquire wide contextual information via extensive receptive fields. Acknowledging that image restoration is an intrinsically position-sensitive operation, the final stage of our model operates on the original input image resolution without any downsampling. This design choice allows the preservation of fine textures and spatial details in the output image, which is critical for the minor boost to PSNR scores and the large improvement in SSIM scores observed in our model.

We weave a supervised attention component between each pair of subsequent stages rather than just stringing together several stages. This component redefines the feature maps from the prior phase prior to feeding them into the next step, which is supervised by ground-truth images. This process optimises the information transfer between phases, which contributes to the improved performance of our model.

In addition to these modifications, we present a cross-stage feature fusion technique. This approach enables intermediate multi-scale context-sensitive traits from prior subnetworks to consolidate intermediate traits from succeeding subnetworks. This intricate interplay among stages and the efficient use of learned features across the network not only slightly improves the PSNR performance than the original model but elevates the SSIM scores significantly, positioning our model as a competent and improved variant of the original MPRNet [27] architecture.

Despite RainDNet having multiple stages, each stage can access the input image directly. As per recent restoration methods [51], we apply a multi-patch hierarchy on the input image and divide it into non-overlapping patches: four for stage 1, two for stage 2, and the original image for the final stage as illustrated in Fig. 2.

At any stage S, we propose the Combined Loss function to handle the task of rain streak removal from images, which is defined as:

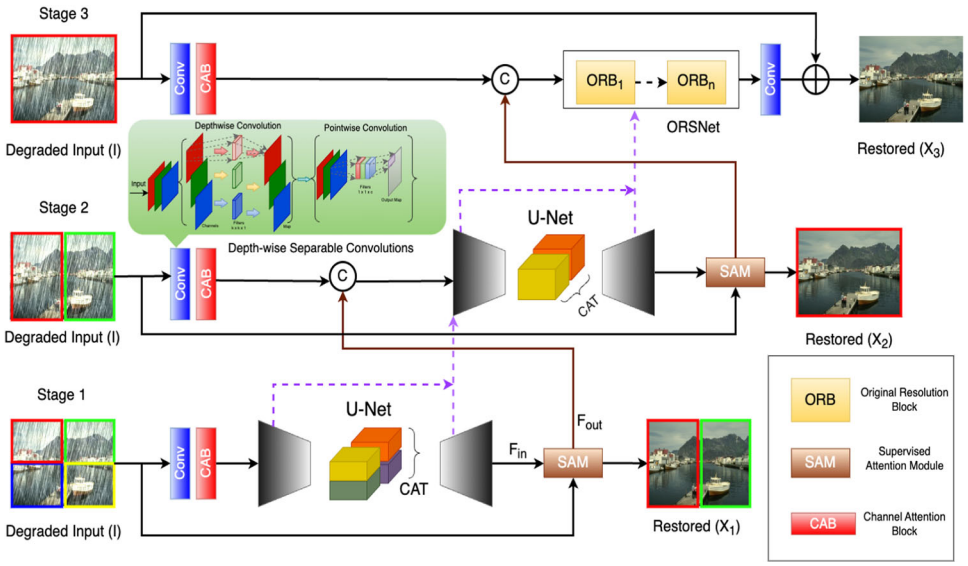
$$\begin{aligned} L_{total} &= \lambda_{L1} \cdot L_{L1} + \lambda_{perc} \cdot L_{perc} \\ &\quad + \lambda_{edge} \cdot L_{edge} + \lambda_{ssim} \cdot L_{ssim} \end{aligned} \tag{1}$$

where, L_{total} is the total loss, L_{L1} is the L1 loss, L_{perc} is the perceptual loss, L_{edge} is the edge loss, and L_{ssim} is the SSIM loss. The coefficients λ_{L1}, λ_{perc}, λ_{edge}, and λ_{ssim} are used to balance these loss components. Each component of the Combined Loss is described as follows:

- L_{L1}: The L1 [62] Loss calculates the absolute difference between the target and output images pixel-wise. It is defined as:

$$L_{L1}(o, t) = \frac{\lambda_{L1}}{N} \sum_{i=1}^{N} |o_i - t_i| \tag{2}$$

  where o and t are the output and target images respectively, and N is the total number of pixels. This loss encourages the model to focus on all discrepancies, big or small, in the restored and target images.
- L_{perc}: The perceptual loss [63] uses a pre-trained VGG16 model to extract feature maps from the restored and target



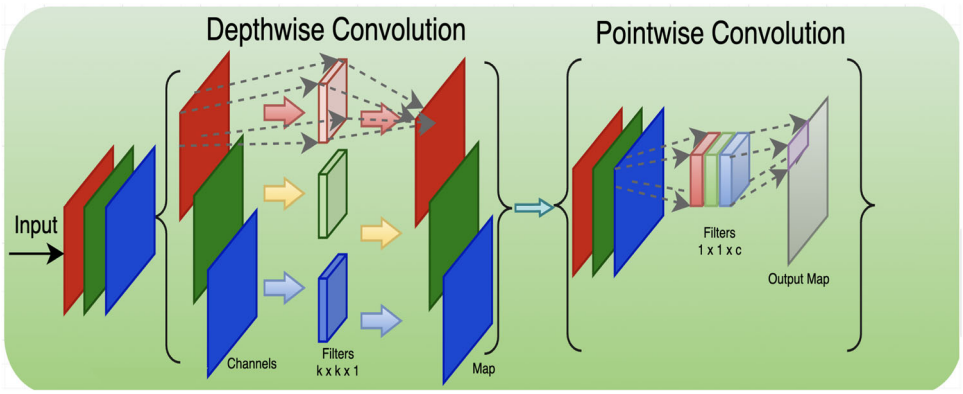


Fig. 2 RainDNet, our proposed architecture for image deraining, employs depthwise separable convolutions [37–42] in certain stages to optimize parameter utilization. The stages and their operations remain consistent with the original MPRNet [27] design, preserving the progressive restoration capability

Fig. 3 Illustration of the depthwise separable convolution [37–42] Operation. This diagram demonstrates the two-step process of depthwise convolution followed by pointwise convolution, showcasing its efficiency in capturing spatial and cross-channel information with significantly fewer parameters compared to standard convolutions. This key alteration in our RainDNet architecture allows for similar restoration performance with a leaner model footprint

images. The L1 loss is then applied to these feature maps, defined as:

$$L_{perc}(o,t) = \frac{\lambda_{perc}}{W \times H \times C} \sum_{w,h,c} |F_o^{w,h,c} - F_t^{w,h,c}| \quad (3)$$

where F_o and F_t are the feature maps of output and target images extracted by the VGG16 model, W, H, and C are the width, height, and number of channels of the feature maps. This loss ensures the model produces a restored image that is not only pixel-wise accurate but also shares similar high-level features (i.e., texture and content) with the target image.

- L_{edge}: The Edge Loss [64] first applies the Sobel filter to the restored and target images to highlight the edges in the images. The L1 loss is then applied to these edge maps. The edge loss is defined as:

$$L_{edge}(o,t) = \frac{\lambda_{edge}}{N} \sum_{i=1}^{N} |E_{o_i} - E_{t_i}| \quad (4)$$

where E_o and E_t are the edge maps of output and target images created using the Sobel operator. This loss encourages the model to pay attention to the edges in the image, which is crucial in maintaining the structure and details of the scene.

- L_{ssim}: The Structural Similarity Index Measure (SSIM) loss [65] is used to ensure that the restored image shares structural similarity with the target image. For each color

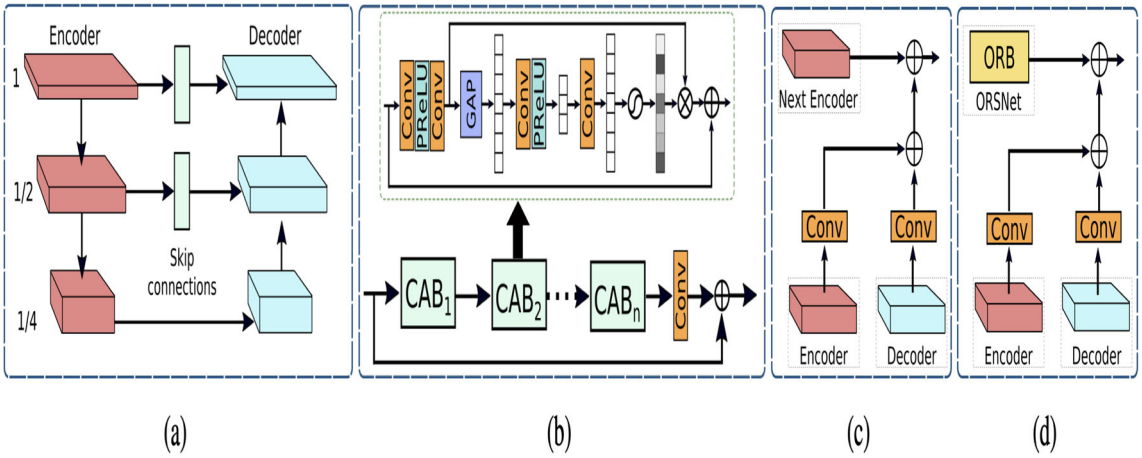


Fig. 4 **a** The encoder–decoder subnetwork, where specific convolution layers are replaced with depthwise separable convolutions [37–42] for a more efficient model. **b** A comprehensive depiction of the modified Original Resolution Block (ORB) in our ORSNet subnetwork is provided. Each ORB consists of several depthwise separable convolutions, in addition to channel attention blocks. The acronym GAP represents Global Average Pooling [66]. **c** Cross Stage Feature Fusion (CSFF) between the first and second phases is demonstrated. **d** Demonstration of CSFF across the second and final phases, highlighting the flow and fusion of features in our RainDNet architecture

channel, the SSIM index is defined as:

$$SSIM_c(o,t) = \frac{(2\mu_{o,c}\mu_{t,c} + c_1)(2\sigma_{o,c,t,c} + c_2)}{(\mu_{o,c}^2 + \mu_{t,c}^2 + c_1)(\sigma_{o,c}^2 + \sigma_{t,c}^2 + c_2)} \tag{5}$$

and the SSIM loss is defined as:

$$L_{ssim}(o,t) = \frac{\lambda_{ssim}}{C}\sum_{c=1}^{C}(1 - SSIM_c(o,t)) \tag{6}$$

where $\mu_{o,c}$ and $\mu_{t,c}$ are the average of o and t for the color channel c, $\sigma_{o,c}^2$ and $\sigma_{t,c}^2$ are the variance of o and t for the color channel c, $\sigma_{o,c,t,c}$ is the covariance of o and t for the color channel c, and c_1 and c_2 are two variables to stabilize the division with weak denominator.

Depth-wise separable convolution block Depth-wise separable convolution blocks [37–42], visualized in Fig. 3, are incorporated for feature extraction and a few other tasks, replacing the standard convolution blocks in the proposed model. This was adapted from the principle of factorizing the standard convolution operation into a depth-wise convolution and a point-wise convolution, which significantly reduces the computational burden without compromising the network's ability to capture complex patterns in the data. Depth-wise separable convolutions exploit the spatial and cross-channel correlations separately, enabling the model to maintain a satisfactory level of representation learning with fewer parameters and computational complexity. They offer the benefits of computational efficiency and parameter reduction, which makes the model lighter, faster, and more suitable for tasks where computational resources are a constraint. Furthermore, the reduced complexity also contributes to alleviating overfitting issues, thus potentially improving the model's performance on unseen data. These advantages make depth-wise separable convolution blocks a preferred choice for our network architecture in the image restoration task.

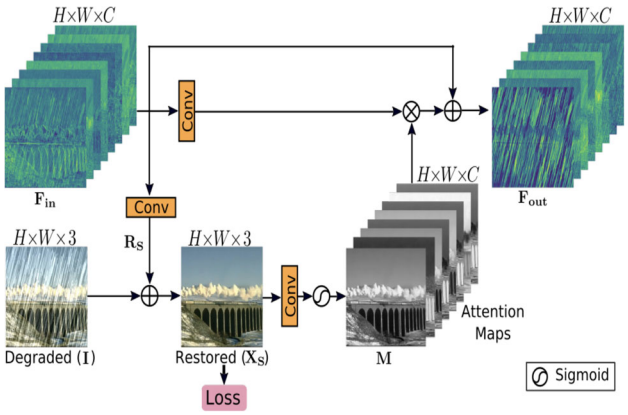


Fig. 5 Modified supervised attention module: this illustration depicts our refined version of the Supervised Attention Module, which emphasizes feature refinement at each stage of the RainDNet architecture

3.1 Processing of complementary features

Modern single-stage CNN models for the restoration of images mostly employ either of the two architectural designs: (1) A framework for encoder–decoders or (2) a singular-scale feature conduit. Encoder–decoder structures [22, 23, 67] begin by converting the input to low-res illustrations and

Table 1 Overview of the image deraining dataset

Tasks	Deraining						
Datasets	Rain14000[26]	Rain1800[18]	Rain800[24]	Rain100H[18]	Rain100L[18]	Rain1200[25]	Rain12[71]
Train samples	11,200	1800	700	0	0	0	12
Test samples	2800	0	100	100	100	1200	0
Testset rename	Test2800	–	Test100	Rain100H	Rain100L	Test1200	–

subsequently employ a reverse mapping process to restore the initial resolution.

Conversely, strategies that operate on a singular-scale feature pipeline are proficient at producing images with precise spatial details [13, 15, 68, 69]. Although these models maintain spatial precision, their outputs often lack semantic richness due to their restricted receptive field. These observations underscore the innate constraints of conventional architectural designs, which can generate either spatially precise or contextually trustworthy outputs, but often grapple to attain both.

Aiming to capitalize on the benefits of both design strategies, we propose a multi-tier framework. In our model, the initial stages utilize encoder–decoder networks, with the last stage operating directly on the original input resolution. Furthermore, we integrate depthwise separable convolutions [37–42] into our framework. This integration considerably lowers the model's complexity, while preserving its competitive performance levels.

Subnetwork with encoder–decoder configuration As depicted in Fig. 4a, our encoder–decoder subnetwork, derived from the standard U-Net [28], includes several adjustments to accommodate our specific requirements. We mostly use channel attention blocks (CABs) [29] for collecting multi-scale traits. The feature maps at the U-Net's skip connections are then processed by the CAB (refer to Fig. 4b). Ultimately, instead of utilising Transposed convolution [70] to increase the spatial scale of each feature within the decoder, we employ bilinear up-sampling complemented by a layer of convolution.

In our proposed architecture, we make significant modifications by employing depthwise separable convolutions [37–42]. This modification allows for more efficient extraction of features at each scale while decreasing the model's complexity. As a result, a balance between performance and computing efficiency has improved, allowing for a more lightweight model with superior performance.

Subnetwork of original resolution We implement a change in the final tier of the architecture to maintain the granular information from the source image in the resultant image. This tweak incorporates a subnetwork that executes in line with the original picture dimensions (refer to Fig. 2). Termed as the original-resolution subnetwork (ORSNet), this module circumvents any downsampling processes and generates features with high resolution, rich in spatial nuances. ORSNet is made up of several original-resolution blocks (ORBs), which in turn incorporate CABs. The structural outline of an ORB can be seen in Fig. 4b.

This novel inclusion of the ORSNet in the final stage ensures that the fine spatial details of the image are preserved, contributing to a more detailed output. The ORB, with its multiple CABs and absence of downsampling, focuses on enhancing the structural similarity of the output image, thereby significantly contributing to the improved SSIM score achieved by our model.

3.2 Supervised attention module

Cutting-edge multi-phase frameworks for restoration of images [51, 58] employ a straightforward strategy wherein each stage produces an image prediction which is subsequently forwarded to the next stage. We present a significant variation in this routine by incorporating a supervised attention module (SAM) between each pair of consecutive stages, contributing towards a substantial improvement in performance.

A structural diagram of the SAM can be seen in Fig. 5, and it lends dual advantages. First, it incorporates ground-truth supervisory cues that aid in the consecutive restoration of images at every tier. Furthermore, we generate attention maps by exploiting locally supervised predictions. These maps are critical in reducing the effect of less useful details at this level, allowing only the most significant features to continue to the subsequent phase.

Notably, this strategy of selectively preserving the most impactful features from one stage to the next directly influences the overall structural similarity of the final result.

SAM functions on the preceding stage's incoming traits, $F_{in} \in \mathbb{R}^{H\times W\times C}$, and produces a residual image, $R_S \in \mathbb{R}^{H\times W\times C}$, using a basic 1×1 convolution. The spatial dimension is denoted by $H \times W$ while the channel count is denoted by C. This residual picture is combined with the impaired input image I to produce the reinstated image, $X_S \in \mathbb{R}^{H\times W\times C}$.

Table 2 Results from image deraining evaluations

Models	Test2800 [26]		Test1200 [25]		Rain100H [18]		Rain100L [18]		Test100 [24]		Average		
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑		SSIM ↑	
DerainNet [74]	24.31	0.861	23.38	0.835	14.92	0.592	27.03	0.884	22.77	0.810	22.48	(71.58%)	0.796	(83.82%)
SEMI [75]	24.43	0.782	26.05	0.822	16.56	0.486	25.03	0.842	22.35	0.788	22.88	(70.24%)	0.744	(87.10%)
DIDMDN [25]	28.13	0.867	29.65	0.901	17.35	0.524	25.23	0.741	22.56	0.818	24.58	(63.81%)	0.770	(85.65%)
UMRL [76]	29.97	0.905	30.55	0.910	26.01	0.832	29.18	0.923	24.41	0.829	28.02	(46.23%)	0.880	(72.49%)
RESCAN [54]	31.29	0.904	30.51	0.882	26.36	0.786	29.80	0.881	25.00	0.835	28.59	(42.58%)	0.857	(76.92%)
PreNet [56]	31.75	0.916	31.36	0.911	26.77	0.858	32.44	0.950	24.81	0.851	29.42	(36.83%)	0.897	(67.96%)
MSPFN [31]	32.82	0.930	32.39	0.916	28.66	0.860	32.40	0.933	27.50	0.876	30.75	(26.37%)	0.903	(65.97%)
MPRNet [27]	33.64	0.938	32.91	0.916	30.41	0.890	36.40	0.938	**30.27**	0.897	32.73	(7.53%)	0.921	(58.22%)
RainDNet (Ours)	**33.67**	**0.972**	**33.09**	**0.957**	**30.86**	**0.954**	**36.58**	**0.983**	30.10	**0.951**	**33.41**	**(0.0%)**	**0.967**	**(0.0%)**

The top-performing and runner-up scores have been emphasised and underlined, correspondingly. The relative error reduction in contrast to the top-performing algorithm is indicated in parentheses for each strategy (refer to Sect. 4. A for the error computation methodology)

Following that, the predicted picture X_S, is provided with explicit supervision using the ground-truth image. Subsequently, Subsequently, attention masks $M \in \mathbb{R}^{H\times W\times C}$ are constructed from the image X_S by employing a 1×1 convolution and sigmoid activation. The masks are put to use to recalibrate the transformed local traits F_{in} (derived after 1×1 convolution), leading to the production of attention-guided features that are subsequently incorporated in the identity mapping path. Ultimately, the output from SAM, an attention-enhanced feature presentation F_{out}, is forwarded to the subsequent stage for additional refinement.

4 Experiments and analysis

We put our proposed technique to the test for a single image restoration task, namely image deraining, across five different datasets.

4.1 Datasets and evaluation protocol

Following the latest research for image deraining, we train our architecture using 13,712 clean rain-image pairs sourced from a diverse set of datasets [18, 24–26, 71], as outlined in Table 1. With this universally developed model, we proceed to evaluations on several testing sets, such as Rain100H [18], Rain100L [18], Test100 [24], Test2800 [26], and Test1200 [25].

We carry out numerical evaluations using the PSNR, SSIM, [72] and BRISQUE [73] metrics. The formula for calculating the Peak Signal-to-Noise Ratio (PSNR) between two images (original and reconstructed) is defined as:

$$PSNR = 20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{7}$$

where MAX_I is the maximum possible pixel value of the image. For an 8-bit grayscale image, the maximum possible pixel value is 255. MSE represents the Mean Squared Error, which measures the average squared differences between the original and the reconstructed images.

The Structural Similarity Index Measure (SSIM) index is a method for comparing similarities between two images (say x and y). The SSIM index is calculated as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

where μ_x is the average of x, μ_y is the average of y, σ_x^2 is the variance of x, σ_y^2 is the variance of y, σ_{xy} is the covariance of x and y, and c_1 and c_2 are two variables to stabilize the division with weak denominator.

Table 3 BRISQUE score results from image deraining evaluations

Models	Test2800 [26]	Test1200 [25]	Rain100H [18]	Rain100L [18]	Test100 [24]
DerainNet [74]	31.32	35.43	44.71	26.04	37.83
SEMI [75]	28.81	32.33	41.90	23.49	35.62
DIDMDN [25]	25.98	30.06	24.92	22.05	31.55
UMRL [76]	23.01	26.61	22.24	16.67	27.97
RESCAN [54]	23.32	27.98	22.23	14.54	23.74
PreNet [56]	23.73	26.68	22.32	15.42	22.21
MSPFN [31]	21.67	23.31	17.78	15.04	19.70
MPRNet [27]	16.61	17.22	17.83	10.21	16.28
RainDNet(ours)	**16.23**	**15.41**	**17.33**	**8.81**	**14.43**

The top-performing and runner-up scores have been emphasised and underlined, correspondingly. A lower BRISQUE score implies better image quality

Table 4 Comparison of trainable parameters

Model	Trainable parameter in millions
DerainNet [74]	1.51
SEMI [75]	0.12
DIDMDN [25]	0.54
UMRL [76]	2.2
RESCAN [54]	0.25
PreNet [56]	0.16
MSPFN [31]	13
MPRNet [27]	20.02
RainDNet(Ours)	5.66

BRISQUE (Blind Referenceless Image Spatial Quality Evaluator) [73] is a no reference image quality score. This uses a pre-trained SVM model to calculate the final score. The pre-trained model takes 5 attributes (MSCN (Mean Subtracted Contrast Normalization) Image and its four shifted versions), which are obtained from the given image. The lesser value of this implies better image quality.

To calculate the MSCN Coefficients, the image intensity I(i,j) at pixel (i, j) is transformed to the luminance $\widehat{I}(i,j)$

$$\widehat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{(\sigma(i,j) + C)} \tag{9}$$

Where i $\in 1, 2, \cdots M$, $j \in 1, 2, \cdots N$ (M and N are height and width respectively). Functions $\mu(i,j) and \sigma(i,j)$ are the local mean field and local variance field, respectively.

We display the relative decrease in error for each approach compared to the top performer by converting PSNR to RMSE ($RMSE \propto \sqrt{10^{-PSNR/10}}$) and SSIM to DSSIM ($DSSIM = (1 - SSIM)/2$).

4.2 Implementation

Our proposed RainDNet model, built to enable end-to-end training, does away with the need for pretraining steps. A distinguishing feature of our model is the application of depthwise separable convolutions [37–42] which efficiently manage computational resources while preserving the ability to learn from a large number of parameters. This approach is especially well-suited for our task and is implemented at various scales of our encoder–decoder network.

To facilitate the extraction of salient features at every scale, we incorporate two Channel Attention Blocks (CABs), utilizing 2×2 max-pooling with a stride of 2 for the downsampling process. The final stage of our model features an Original Resolution Subnetwork (ORSNet), composed of three Original Resolution Blocks (ORBs), each embedded with eight CABs.

In order to customize the network to cater to the specifics of the deraining task, we tweak the network's width by defining the channel count to 40. The training procedure is implemented on patches of size 256×256, using a batch size of 2 per GPU, with two NVIDIA RTX 3090 GPUs, leading to a total of four batches per epoch. This training process lasts for 15k iterations.

In terms of optimization, we deploy the AdamW [77] optimizer with a learning rate set to 3×10^{-4}. Noteworthy is the fact that the AdamW optimizer applies weight decay directly to the weights, bypassing gradient modifications - an approach derived from early neural network methodologies wherein weight decay was achieved through direct weight shrinkage.

4.3 Image deraining results

In accordance with previous studies, specifically reference [31], we utilized the Y channel (from the YCbCr color space) to compute the quality metrics for the image derain-

Fig. 6 Results of image deraining using the RainDNet model. Marking a significant leap in performance, our RainDNet model expertly removes rain streaks and yields images that are not only realistic and devoid of artifacts, but also visually far more similar to the ground truth compared to prior models, particularly its precursor, the MPRNet [27]

ing task. As presented in Table 2, our approach significantly surpasses the current cutting-edge model by mostly yielding superior PSNR and SSIM results across all five datasets. The BRISQUE score of the proposed model and the state-of-the-art models is shown in Table 3. The proposed model outperforms other models in terms of the BRISQUE score over all five datasets. In comparison to the most recent top-performing algorithm, MPRNet [27], our method achieved an average performance enhancement of 0.68 dB across all datasets. Moreover, our model is more efficient, having 3.5x fewer parameters than MPRNet [27], as evident from Table 4, which shows the number of trainable parameters of different models. Fig. 6 provides visual comparisons on challenging images. Our RainDNet demonstrates effectiveness in eradicating rain streaks of different orientations and intensities and generates images that are visually pleasing and closely align with the ground truth. In contrast, other techniques compromise structural content (first row), generate artifacts (second row), and are unsuccessful in completely removing rain streaks (third row).

Table 5 Examination of the individual elements of the proposed RainDNet through an ablation study

#Stage	Components	PSNR
1	U-Nets(baseline)	29.48
1	ORSNet (baseline)	29.30
1	U-Nets(baseline) with Depth. Sep. Conv	29.45
1	ORSNet (baseline) with Depth. Sep. Conv	29.39
2	U-Nets+U-Nets	29.81
2	ORSNet+ORSNet	29.96
2	U-Nets+ORSNet	30.13
2	U-Nets+ORSNet+Depth. Sep. Conv	30.21
3	U-Nets+ORSNet	30.06
3	U-Nets+ORSNet+CSFF+Depth. Sep. Conv	30.37
3	U-Nets+ORSNet+SAM+Depth. Sep. Conv	30.62
3	U-Nets+ORSNet+SAM+CSFF+Depth. Sep. Conv	**30.86**

The best result is shown in bold

4.4 Ablation studies

In this segment, we execute a variety of experiments to comprehend the impact of each element of our RainDNet model. Our examination utilizes the Rain100H [18] dataset, with the deraining models trained on image patches of dimension 256×256. The findings are presented in Table 5.

Number of stages As we escalate the number of stages within our RainDNet, we observe an enhancement in its efficacy, thereby reinforcing the efficiency of our multi-stage framework.

Selection of subnetworks In our model, different types of subnetworks can be utilized in each stage. Hence, we experimented with several alternatives. Our observations reveal that deploying the encoder–decoder subnetwork in the initial stages and the ORSNet in the final stage yields superior outcomes as compared to employing a uniform design across all stages.

SAM, CSFF, and depthwise separable convolutions We also wanted to understand the impact of the Supervised Attention Module (SAM), the Cross Stage Feature Fusion (CSFF) mechanism, and depthwise separable convolutions [37–42] on the performance of our model. When we removed the SAM from our model, there was a significant drop in the PSNR. The same thing happened when we removed the CSFF. Removing both components led to an even bigger drop in performance. The introduction of depthwise separable convolutions, however, resulted in a significant boost to the model's performance, confirming its importance in our architecture.

5 Conclusion

In this research, we present RainDNet, an enhanced multi-stage framework for image restoration. Extending the core principles of MPRNet, our model systematically enhances impaired inputs by embedding supervised attention within every stage. We define key tenets to shape our design, emphasizing the amalgamation of feature processing across several stages, coupled with a flexible exchange of information between stages.

RainDNet brings in stages that are rich in contextual information and spatial precision, working in unison to encode a wide array of features. We have implemented depthwise separable convolutions, thereby improving computational efficiency while preserving the model's effectiveness. To facilitate fruitful cooperation between interconnected stages, we have designed a unique feature integration process across stages, along with a supervised attention module that navigates the exchange of outputs from preceding stages to the ones that follow.

Our advancements yield considerable performance improvements in terms of PSNR, SSIM scores and BRISQUE scores even when compared against the strong baseline of MPRNet. Demonstrated across various benchmark datasets, RainDNet not only exhibits superior restoration capabilities but also showcases a desirable trade-off between model size and efficiency. This advantage makes RainDNet especially fitting for devices with limited resources, without compromising the quality of the restored images.

As we move forward, we envision a promising scope for the continued development and optimization of the RainDNet model. By virtue of its sophisticated design and superior performance, RainDNet has the potential to pioneer new directions in image restoration and even in broader fields of computer vision.

One potential area of future exploration is the application of RainDNet in Advanced Driver-Assistance Systems (ADAS). Given its proficiency in enhancing degraded images, RainDNet could play a pivotal role in improving the accuracy and reliability of such systems, especially under adverse weather conditions. Its ability to accurately eliminate rain and haze from images could significantly improve the visual perception capabilities of ADAS, thus enhancing the safety and efficiency of automated driving.

Furthermore, we anticipate potential modifications in the RainDNet model that could accommodate other types of image degradation, like snow, dust, or fog. Future research might also investigate the application of RainDNet's depthwise separable convolution strategy to other architectures, potentially sparking advancements in computational efficiency across a range of computer vision tasks.

Finally, we recognize the value of continued refinements to our supervised attention module and feature fusion tech-

niques. These enhancements could further boost the model's performance and establish RainDNet as a robust standard in the domain of image restoration. In conclusion, the future seems bright for RainDNet, with numerous avenues for exploration and expansion.

Acknowledgements The authors express their gratitude to the Indian Institute of Information Technology Allahabad (IIIT-A), India, for the obtained financial support in performing this research work. This work is one of the outcomes of the project entitled "Deep Learning based Solutions for Vehicle Detection in Rainy and Foggy Climates under Smart City Environment" with sanction no. IIITA/RO/2022/409 dated 01.12.2022, sponsored by IIIT-A, Ministry of Education, India.

Author Contributions SP and DKS have written the main manuscript including figures and tables. NS have given the idea upon the novelty and reviewed the manuscript.

Funding The work is sponsored by IIIT-A, Ministry of Education, India, for the project entitled "Deep Learning based Solutions for Vehicle Detection in Rainy and Foggy Climates under Smart City Environment" with Sanction No. IIITA/RO/2022/409 dated 01.12.2022.

Data availibility The manuscript has no associated data.

Declarations

Conflict of interest There is no conflict of interest.

References

1. Dong, W., Zhang, L., Shi, G., Wu, X.: Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans. Image Process. **20**(7), 1838–1857 (2011). https://doi.org/10.1109/TIP.2011.2108306
2. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011). https://doi.org/10.1109/TPAMI.2010.168
3. Kim, K.I., Kwon, Y.: Single-image super-resolution using sparse regression and natural image prior. IEEE Trans. Pattern Anal. Mach. Intell. **32**(6), 1127–1133 (2010). https://doi.org/10.1109/TPAMI.2010.25
4. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **12**(7), 629–639 (1990). https://doi.org/10.1109/34.56205
5. Roth, S., Black, M.J.: Fields of experts: a framework for learning image priors. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, pp. 860–8672 (2005). https://doi.org/10.1109/CVPR.2005.160
6. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992). https://doi.org/10.1016/0167-2789(92)90242-F
7. Zhu, S.-C., Mumford, D.: Prior learning and Gibbs reaction–diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **19**, 1236–1250 (1997)
8. Dai, T., Cai, J., Zhang, Y., Xia, S.-T., Zhang, L.: Second-order attention network for single image super-resolution. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11057–11066 (2019). https://doi.org/10.1109/CVPR.2019.01132
9. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690
10. Pan, X., Zhan, X., Dai, B., Lin, D., Loy, C. C., Luo, P. (2021). Exploiting deep generative prior for versatile image restoration and manipulation. IEEE Trans. Pattern Anal. Mach. Intell., 44(11), 7474–7489
11. Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H., Shao, L. (2020). Cycleisp: Real image restoration via improved data synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2696–2705
12. Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H., Shao, L. (2020). Learning enriched features for real image restoration and enhancement. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16, pp. 492–511. Springer International Publishing
13. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017). https://doi.org/10.1109/TIP.2017.2662206
14. Zhang, K., Zuo, W., Gu, S., Zhang, L. (2017). Learning deep CNN denoiser prior for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3929–3938
15. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y. (2018). Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2472–2481
16. Liu, J., Akhtar, N., Mian, A.: Deep reconstruction of 3-d human poses from video. IEEE Trans Artif Intell **4**(03), 497–510 (2023). https://doi.org/10.1109/TAI.2022.3164065
17. Anwar, S., Barnes, N.: Real image denoising with feature attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2019)
18. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1685–1694 (2017). https://doi.org/10.1109/CVPR.2017.183
19. Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration (2019)
20. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 4809–4817 (2017). https://doi.org/10.1109/ICCV.2017.514
21. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C.C., Qiao, Y., Tang, X.: ESRGAN: enhanced super-resolution generative adversarial networks (2018)
22. Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Unprocessing images for learned raw denoising (2018)
23. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (2018)
24. Zhang, H., Sindagi, V., Patel, V. M. (2019). Image de-raining using a conditional generative adversarial network. IEEE Trans. Circuits Syst. Video Technol. 30(11), 3943–3956
25. Zhang, H., & Patel, V. M. (2018). Density-aware single image deraining using a multi-stream dense network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 695–704
26. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)

27. Zamir, S. W., Arora, A., Khan, S., Hayat, M., Khan, F. S., Yang, M. H., Shao, L. (2021). Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14821–14831
28. Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, part III 18, pp. 234–241. Springer International Publishing
29. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 286–301
30. Farha, Y.A., Gall, J.: MS-TCN: Multi-Stage Temporal Convolutional Network for Action Segmentation (2019)
31. Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., Ma, J., Jiang, J. (2020). Multi-scale progressive fusion network for single image deraining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8346–8355
32. Li, S., Farha, Y. A., Liu, Y., Cheng, M. M., Gall, J. (2020). Ms-tcn++: Multi-stage temporal convolutional network for action segmentation. IEEE Trans. pattern Anal. Mach. Intell. 45(6), 6647–6658
33. Li, W., Wang, Z., Yin, B., Peng, Q., Du, Y., Xiao, T., Yu, G., Lu, H., Wei, Y., Sun, J. (2019). Rethinking on multi-stage networks for human pose estimation. arXiv preprint arXiv:1901.00148
34. Chen, Y., Wang, Z., Peng, Y., Zhang, Z., Yu, G., Sun, J. (2018). Cascaded pyramid network for multi-person pose estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7103–7112
35. Cheng, B., Chen, L. C., Wei, Y., Zhu, Y., Huang, Z., Xiong, J., Huang, T.S., Hwu, W.M., Shi, H. (2019). Spgnet: Semantic prediction guidance for scene parsing. In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 5218–5228
36. Newell, A., Yang, K., Deng, J. (2016). Stacked hourglass networks for human pose estimation. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14, pp. 483–499. Springer International Publishing
37. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
38. Ouzar, Y., Djeldjli, D., Bousefsaf, F., Maaoui, C.: X-ippgnet: a novel one stage deep learning architecture based on depthwise separable convolutions for video-based pulse rate estimation. Comput. Biol. Med. **154**, 106592 (2023)
39. Tseng, F.-H., Yeh, K.-H., Kao, F.-Y., Chen, C.-Y.: Mininet: dense squeeze with depthwise separable convolutions for image classification in resource-constrained autonomous systems. ISA Trans. **132**, 120–130 (2023)
40. Hassan, E.: Scene text detection using attention with depthwise separable convolutions. Appl. Sci. **12**(13), 6425 (2022)
41. Kaiser, L., Gomez, A.N., Chollet, F.: Depthwise separable convolutions for neural machine translation. arXiv preprint arXiv:1706.03059 (2017)
42. Guo, J., Li, Y., Lin, W., Chen, Y., Li, J.: Network decoupling: from regular to depthwise separable convolutions. arXiv preprint arXiv:1808.05517 (2018)
43. Chan, T.F., Wong, C.: Total variation blind deconvolution. IEEE Trans. Image Process. **7**(3), 370–375 (1998). https://doi.org/10.1109/83.661187
44. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**(11), 4311–4322 (2006). https://doi.org/10.1109/TSP.2006.881199
45. Luo, Y., Xu, Y., Ji, H.: Removing rain from a single image via discriminative sparse coding. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 3397–3405 (2015). https://doi.org/10.1109/ICCV.2015.388
46. Mairal, J., Elad, M., Sapiro, G.: Sparse representation for color image restoration. IEEE Trans. Image Process. **17**(1), 53–69 (2008). https://doi.org/10.1109/TIP.2007.911828
47. Buades, A., Coll, B., Morel, J.-M.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, pp. 60–652 (2005). https://doi.org/10.1109/CVPR.2005.38
48. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007). https://doi.org/10.1109/TIP.2007.901238
49. Shan, Q., Jia, J., Agarwala, A.: High-quality motion deblurring from a single image. ACM Trans. Graph. **27**(3), 73 (2008). https://doi.org/10.1145/1360612.1360672
50. Xu, L., Zheng, S., Jia, J.: Unnatural l0 sparse representation for natural image deblurring. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1107–1114 (2013). https://doi.org/10.1109/CVPR.2013.147
51. Suin, M., Purohit, K., Rajagopalan, A. N. (2020). Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3606–3615
52. Li, X., Li, X., Li, Z., Xiong, X., Khyam, M., Sun, C.: Robust vehicle detection in high-resolution aerial images with imbalanced data. IEEE Trans. Artif. Intell. **2**(03), 238–250 (2021). https://doi.org/10.1109/TAI.2021.3081057
53. Fu, X., Liang, B., Huang, Y., Ding, X., Paisley, J.W.: Lightweight pyramid networks for image deraining. IEEE Trans. Neural Netw. Learn. Syst. **31**(6), 1794–1807 (2020). https://doi.org/10.1109/TNNLS.2019.2926481
54. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H. (2018). Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 254–269
55. Nah, S., Hyun Kim, T., Mu Lee, K. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3883–3891
56. Ren, D., Zuo, W., Hu, Q., Zhu, P.F., Meng, D.: Progressive image deraining networks: a better and simpler baseline. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3932–3941 (2019)
57. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8174–8182 (2018). https://doi.org/10.1109/CVPR.2018.00853
58. Zhang, H., Dai, Y., Li, H., & Koniusz, P. (2019). Deep stacked hierarchical multi-patch network for image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5978–5986
59. Zhao, H., Zhang, Y., Liu, S., Shi, J., Loy, C.C., Lin, D., Jia, J.: PSANET: point-wise spatial attention network for scene parsing. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
60. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. **42**(8), 2011–2023 (2020). https://doi.org/10.1109/TPAMI.2019.2913372
61. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
62. Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Two deterministic half-quadratic regularization algorithms for com-

puted imaging. In: Proceedings of 1st International Conference on Image Processing, vol. 2, pp. 168–1722 (1994). https://doi.org/10.1109/ICIP.1994.413553
63. Rad, M.S., Bozorgtabar, B., Marti, U.-V., Basler, M., Ekenel, H.K., Thiran, J.-P.: Srobb: targeted perceptual loss for single image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2710–2719 (2019)
64. Seif, G., Androutsos, D.: Edge-based loss function for single image super-resolution. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1468–1472 (2018)
65. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imaging **3**(1), 47–57 (2016)
66. Liu, W., Rabinovich, A., Berg, A. C. (2015). Parsenet: Looking wider to see better. arXiv preprint arXiv:1506.04579
67. Kupyn, O., Martyniuk, T., Wu, J., & Wang, Z. (2019). Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8878–8887
68. Anwar, S., Khan, S., Barnes, N.: A deep journey into super-resolution: a survey. ACM Comput. Surv. **50**, 1–34 (2020)
69. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**, 295–307 (2015). https://doi.org/10.1109/TPAMI.2015.2439281
70. Odena, A., Dumoulin, V., Olah, C.: Deconvolution and checkerboard artifacts. Distill (2016). https://doi.org/10.23915/distill.00003
71. Li, Y., Tan, R.T., Guo, X., Lu, J., Brown, M.S.: Rain streak removal using layer priors. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2736–2744 (2016). https://doi.org/10.1109/CVPR.2016.299
72. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
73. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
74. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.: Clearing the skies: a deep network architecture for single-image rain removal. IEEE Trans. Image Process. **26**(6), 2944–2956 (2017). https://doi.org/10.1109/tip.2017.2691802
75. Wei, W., Meng, D., Zhao, Q., Xu, Z., & Wu, Y. (2019). Semi-supervised transfer learning for image rain removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3877–3886
76. Yasarla, R., & Patel, V. M. (2019). Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8405–8414
77. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)