



Residual Contextual Hourglass Network for Single-Image Deraining

Weina Zhou¹ · Linhui Ye¹ · Xiu Wang¹

Accepted: 9 January 2024
© The Author(s) 2024

Abstract

Rain streaks could blur and distort images, significantly impacting further image processing. Single-image deraining is a hotspot and has practical application value, while most existing methods still have problems such as residual rain streaks and inadequate recovery of detail textures. To address these issues, we propose a Residual Contextual Hourglass Network (RCHNet) for single-image deraining, which could adapt to remove rain streaks in complex environments. Firstly, a contextual distillation block is presented to obtain local and global features across different scales. Further, residual downsampling block and residual upsampling block are used to maintain the residual nature of our architecture and better restore the details of the image. Finally, a dual attention mechanism is introduced to compensate for the spatial information lost by the downsampling. Extensive experiments on five synthetic datasets and a real-world dataset demonstrate that our proposed RCHNet outperforms existing state-of-the-art deraining approaches. On average across all synthetic datasets, the PSNR score of RCHNet is as high as 33.31 dB.

Keywords Single-image deraining · Hourglass network · Residual block · Attention mechanism

1 Introduction

Images collected in rainy weather are easily affected by rain refraction and occlusion, which seriously affect subsequent processing such as object detection [1, 2], semantic segmentation [3] and target tracking [4]. Image deraining is an important preprocessing for them, especially in complex maritime scenarios. It could benefit marine object detection [5, 6] a lot. Therefore, an efficient deraining method has an urgent requirement for many applications as a crucial visual preprocessing.

✉ Weina Zhou
wnzhou@shmtu.edu.cn

Linhui Ye
1870190134@qq.com

¹ College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China

At present, deraining methods could be divided into video-based deraining [7, 8] and single-image deraining (SIDR). And SIDR is more popular in the research field for it could be utilized as an auxiliary method when video-based deraining is disabled in a rapidly changed background. However, removing rain streaks from a single image is also more challenging since it lacks any prior information such as the temporal and spatial correlation of rain streaks in a video.

In the researches of SIDR, some model-based deraining algorithms [9–11] always build models to simulate the relationship between rain and the background, then optimize the model to complete the deraining task. Their deraining effect is constantly constrained by the complicated environment, thus the model exhibits poor generalizability. Nowadays, driven by a large amount of training data and the excellent fitting ability of convolutional neural networks [12–15], deep learning has become the research trend of the current deraining task. Fu et al. [16] propose a method for deraining by reducing the topological range (DDN). Li et al. [17] present a recursive context expansion network to extract rain information (RESCAN). Ren et al. [18] propose a recurrent structure to synthesize the background progressively (PreNet). Jiang et al. [19] propose a multi-scale progressive fusion deraining network (MSPFN), and non-local fusion module [20] and attention fusion module [21] to further explore the local–global features in the progress network. Chen et al. [22] and Frants et al. [23] both introduce a multi-stage multiscale neural network to boost the performance of image restoration networks. In recent days, researchers consider deraining tasks from different perspective, transformer-based methods [24, 25] have achieved great success in capture long-range dependencies. Works [26–28] adopted progressive coupled network, unpaired adversarial framework and a direction aware wavelet network for SID tasks in different perspectives. These deraining networks improve network performance by adopting a multi-branch structure or increasing depth. The architecture is too complicated, making the network harder to train. Additionally, they did not fully utilize pixel and channel information, which usually resulted in problems such as residual rain streaks and inadequate recovery of texture details.

To address these issues, we propose a novel single-image deraining network called Residual Contextual Hourglass Network (RCHNet). RCHNet uses an end-to-end network structure with parameter sharing to reduce the total number of parameters and simplify the network complexity. Since rain streaks show apparent self-similarity in different scales, our framework can acquire complementary rain streaks feature by the residual up-down sampling structure. The hourglass network enhances the receptive field of the network through a multi-level up-down sampling structure, which improves the perception of global information and accurately predicts the critical details of each key point. Moreover, we proposed residual downsampling block (RDB) and residual upsampling block (RUB) to maintain residual features and dual attention mechanism (DAM) to fully utilize pixel and channel information. Therefore, RCHNet can quickly extract detailed information to achieve rain removal in complex rainy environments. The main contributions of our work are as follows:

1. A contextual distillation block (CDB) is proposed to extract local features across different scales and capture rich contextual detail information.
2. The proposed dual attention mechanism (DAM) integrates multiple receptive fields to compensate for spatial and channel information lost in sampling during the coding stage with the parallel spatial and channel dual attention.
3. Residual downsampling block (RDB) and residual upsampling block (RUB) are used in our hourglass framework to perform the up-down sample operation, which maintain residual features in the structure and gain deeper image feature.

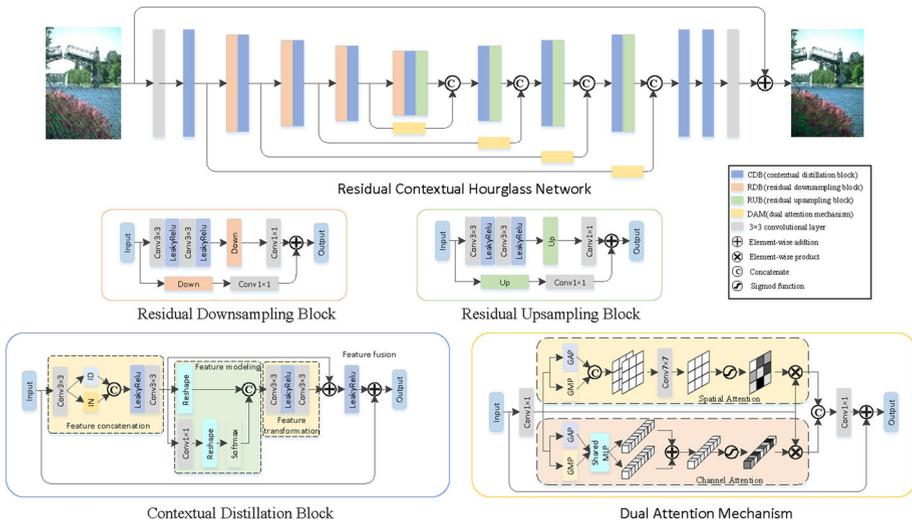


Fig. 1 Full overview of the proposed Residual Contextual Hourglass Network structure

- Extensive experiments on five synthetic datasets and a real-world dataset demonstrate that our proposed RCHNet outperforms existing state-of-the-art deraining approaches. The score of average PSNR by RCHNet on all synthetic datasets is as high as 33.31dB.

2 Methods

The hourglass structure is a relatively basic module in convolutional neural networks, including bottom-up and top-down branches. It is similar to an encoder-decoder structure. The proposed Residual Contextual Hourglass Network (RCHNet) is composed of four important modules: residual downsampling block (RDB), residual upsampling block (RUB), contextual distillation block (CDB), and dual attention mechanism (DAM). The overall scheme of the RCHNet can be found in Fig. 1 Bottom-up: When the resolution is high, it utilizes the large receptive field CDB to capture background image information before using the RDB to scale down the feature map’s resolution. Top-down: After achieving the lowest resolution, the network employs CDB much more effectively to collect deep information. Moreover, the scale of the feature map is expanded layer by layer using the RUB operation, and cross-scale feature fusion is accomplished using the DAM. The final network output is obtained by using a 3 x 3 convolution after getting the same resolution as the input of network.

2.1 Contextual Distillation Block

Motivated by the advances of recent deraining methods [17, 22], we introduce a contextual distillation block (CDB) as the basic structure of RCHNet to extract features at each scale. Its architecture is illustrated in Fig. 1 To suppress useless features and extract useful information from feature tensors, CDB consists of four parts: (1) Feature concatenation: Perform instance normalization (IN) [29] on half of the input feature $X_{in} \in \mathbb{R}^{H \times W \times C}$ while keeping context information (CI) by the other half of it, then contact them and output the concate-

nation feature $X_{fc} \in \mathbb{R}^{H \times W \times C}$. This methodology effectively retains crucial details and contextual information within the image. Here, $H \times W$ is the spatial dimension and C is the number of channels. (2) Feature modeling: Reshaping X_{fc} to obtain the global feature $X_{fm} \in \mathbb{R}^{1 \times 1 \times C}$, facilitating the comprehensive capture of global rain texture characteristics. (3) Feature transformation: Capturing the passed inter-channel dependencies and yielding novel attention feature $X_{ft} \in \mathbb{R}^{1 \times 1 \times C}$. This enhancement significantly boosts the network's ability to identify rainfall-related patterns across channels. (4) Feature fusion: Element-wise additive aggregation of the transformed features X_{ft} onto the input features X_{in} , and yield features $X_{cdb} \in \mathbb{R}^{H \times W \times C}$. Feature fusion combines features from different stages, achieving more precise rain removal while preserving image details and structural information. By combining these four components, RCHNet could extract key rain streak features and address the complexity of deraining more effectively. The output of CDB can be expressed as:

$$\begin{cases} X_{fc} = f^{3 \times 3}(\sigma_1(\text{cat}[IN(f^{3 \times 3}(X_{in}))_{mid}, (f^{3 \times 3}(X_{in}))_{mid}]))) \\ X_{fm} = \text{cat}(rs(X_{fc}), \sigma_2(rs(f^{1 \times 1}(X_{fc})))) \\ X_{ft} = X_{fc} + f^{3 \times 3}(\sigma_1(f^{3 \times 3}(X_{fm}))) \\ X_{cdb} = X_{in} + \sigma_1(X_{ft}) \end{cases} \quad (1)$$

where $f^{3 \times 3}(\cdot)$ is the convolution layer with convolution kernels of size 3×3 . $IN(\cdot)$ denotes the instance normalization. σ_1 indicates a non-linearity activation function termed *LeakyRelu* with a negative slope of 0.2. $f^{1 \times 1}(\cdot)$ is the convolution layer with convolution kernels of size 1×1 . $rs(\cdot)$ denotes the reshaping operation. $\text{cat}(\cdot)$ is a concatenating operation. σ_2 indicates a Softmax function.

2.2 Residual Downsampling and Upsampling Block

The hourglass network presents a robust framework for capturing diverse raindrop information across multiple scales. However, the upsampling and downsampling operations within its layers are the consequential loss of nuanced rain streaks and intricate texture details within the images. In order to uphold the residual characteristics integral to the RCHNet architecture, this study introduces two key components: the Residual Downsampling Block (RDB) and the Residual Upsampling Block (RUB). Both are designed to harness the efficacy of residual connections, thereby enhancing the preservation of fine image details. The RDB serves to mitigate the challenge of gradient vanishing as the network progresses in depth, ensuring improved information flow throughout deeper layers. Conversely, the RUB specializes in recovering and retaining high-frequency fine details during the upsampling process. As shown in Fig. 1, RDB first consists of two 3×3 convolutional layers, and each convolutional layer is followed by a leaky rectified linear unit (*LeakyRelu*). Then it utilizes anti-aliasing downsampling [30] and 1×1 convolution to get the output features, and finally connects to the same downsampling features as the input layer. Given an input feature $X_{in} \in \mathbb{R}^{H \times W \times C}$, the output feature of RDB is $X_{rdb} \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 2C}$. Similarly, bilinear upsampling is used in RUB as shown in Fig. 1. Given an input feature $X_{in} \in \mathbb{R}^{H \times W \times C}$, the output feature of RUB is $X_{rub} \in \mathbb{R}^{2H \times 2W \times \frac{C}{2}}$.

2.3 Dual Attention Mechanism

It is crucial to exploit the image background features and keep them to the end of the model for deraining. However, as the depth of the network increases, some useful feature information is

Table 1 Datasets description

Datasets	Training samples	Testing samples	Testset rename
Rain800 [35]	700	100	Test100
Rain200H [36]	0	100	Rain100H
Rain200L [36]	0	100	Rain100L
Rain14000 [16]	11, 200	2800	Test2800
Rain1200 [37]	0	1200	Test1200
Rain1800 [36]	1800	0	Rain1800
Total count	13, 700	4300	–

easily lost in the sampling process. And it also affects the ability of spatial representation in the transmission process, resulting in a large number of redundant features. To address these issues, we propose a dual attention mechanism (DAM) between the same-scale encoder and decoder to integrate multiple receptive fields. Its structure is shown in Fig. 1. DAM compensates for the loss of spatial information during sampling and transmission at the bottom-up branch. The spatial attention and channel attention mechanisms [31] are utilized to achieve feature recalibration. Specifically, spatial attention (SA) [32] captures the inter-spatial dependencies of different features by employing global average and max pooling operations. Channel attention (CA) [33, 34] applies global average, max pooling, and multi-layer perceptron to calculate channel attention features more efficiently. And it can utilize feature context information and mine the relationship between hierarchical features to achieve feature distillation. Given an input feature $X_{in} \in \mathbb{R}^{H \times W \times C}$, its output is described as:

$$\begin{cases} X_{fr} = \text{cat}(X_{sa} \otimes X_{ca} \otimes f^{1 \times 1}(X_{in}), X_{ca} \otimes X_{sa} \otimes f^{1 \times 1}(X_{in})) \\ X_{dam} = X_{in} + f^{1 \times 1}(X_{fr}) \end{cases} \quad (2)$$

where $X_{sa} \in \mathbb{R}^{H \times W \times 1}$ is the output of spatial attention, $X_{ca} \in \mathbb{R}^{1 \times 1 \times C}$ is the output of channel attention, $X_{fr} \in \mathbb{R}^{H \times W \times C}$ denotes the output of feature recalibration. $f^{1 \times 1}(\cdot)$ is the convolution layer with convolution kernels of size 1×1 . $\text{cat}(\cdot)$ is a concatenating operation. \otimes indicates element-wise product.

3 Experimental Results

3.1 Datasets and Metrics

In our experiments, the detailed descriptions of datasets for training and testing are listed in Table 1. A total of 13,700 clean/rain image pairs [19] from multiple datasets are used for training. Testing 4300 labeled reference samples and 147 real-world images (DDN-SIRR [38]). We employ peak signal-to-noise ratio (PSNR) [39] and structural similarity index (SSIM) [40] as the evaluation index on synthetic datasets, when compared with several state-of-the-art deraining methods. Due to the difficulties of collecting corresponding ground truth for real-world images, the performance is evaluated via visual comparisons.

Table 2 Quantitative results evaluated on five synthetic datasets

Methods	Test100 [35] PSNR/SSIM (%)	Rain100H [36] PSNR/SSIM (%)	Rain100L [36] PSNR/SSIM (%)	Test2800 [16] PSNR/SSIM (%)	Test1200 [37] PSNR/SSIM (%)
RESCAN [17]	25.09/81.7	25.86/79.3	30.82/85.1	31.27/90.4	30.20/84.7
PreNet [18]	25.15/83.6	25.92/85.0	31.65/94.5	31.63/91.5	31.12/90.8
MSPFN [19]	26.53/86.6	27.75/85.2	31.50/92.8	32.79/92.9	32.14/91.3
HINet [22]	29.22/88.4	30.78/89.5	36.37/96.5	33.91/93.9	32.80/91.6
QSAM-Net [23]	29.855/89.7	30.46/89.2	37.35/97.2	33.87/93.9	32.95/91.5
RCHNet (ours)	30.74/91.0	30.96/89.7	37.61/97.3	33.92/94.2	33.31/92.6

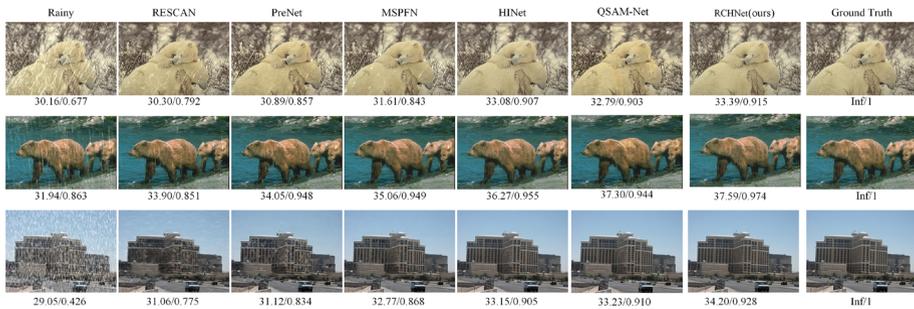


Fig. 2 Visual comparisons of state-of-the-art deraining methods on synthetic datasets

Table 3 Comparison result of model parameters and average inference time

Methods	RESCAN	PreNet	MSPFN	HINET	QSAM-Net	RCHNet
Par.(M)	0.15	0.169	13.35	88.67	22.29	56.92
Times(S)	0.562	0.167	0.519	0.548	0.362	0.298

3.2 Implementation Details

In the training process, our network is implemented using the PyTorch deep learning framework and trained on two Tesla V100 GPUs. The learning rate is initially set to 2×10^{-4} by using Adam optimizer and then reduced to 1×10^{-7} using the cosine annealing technique. The network is trained on 256×256 patches with a batch size of 8 for 5×10^5 iterations. Image flipping and rotation are used at random for data augmentation [22].

Loss Function: PSNR [39] loss is verified in many literatures on deraining tasks, thus it is also employed in RCHNet to construct our objective function. It is formulated as follows:

$$Loss = -PSNR((O + I), B) \quad (3)$$

where O represents the output derained image, I is the input image, and B is the ground-truth image.

3.3 Results on Synthetic Datasets

Our method are compared with five state-of-the-art methods on synthetic datasets (Test100 [35], Rain100H [36], Rain100L [36], Test2800 [16], and Test1200 [37]), and the test results are shown in Table 2. The bold entries represent the best results in the table. All methods are trained and tested on the same datasets and experimental environment. On all five datasets, our method achieves the highest PSNR and SSIM scores. Especially, the test results on datasets Rain100H [36], Rain100L [36], and Test1200 [37] have been significantly improved, which shows that our model has better generalization ability. Furthermore, a comprehensive comparison about parameter numbers and reference time is shown in Table 3. Compared to the recent best method QSAM-Net [23], our model parameters is larger, but our detection time is competitive compared to most of the state-of-the-art algorithms.

We randomly select the test results, and compare the deraining performance of different methods from the visual level. As shown in Fig. 2, RCHNet effectively removes rain streaks,

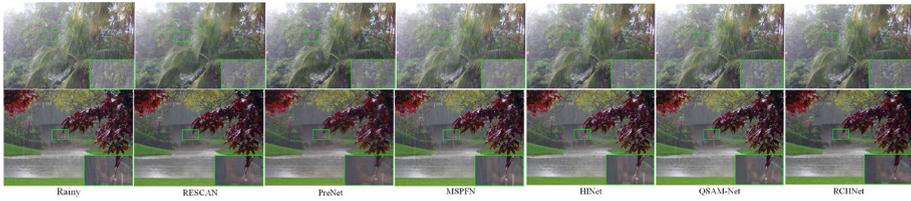


Fig. 3 Visual comparisons of state-of-the-art deraining methods on real datasets

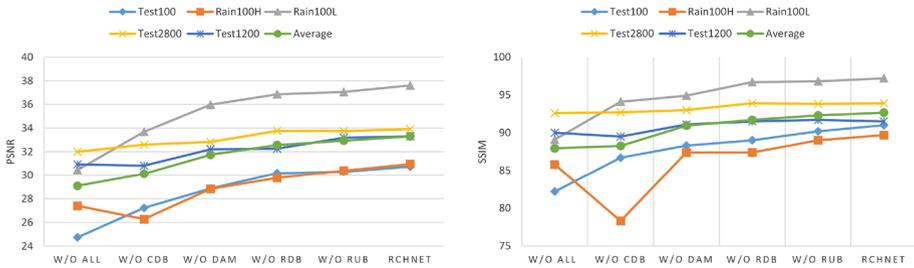


Fig. 4 Effectiveness of different components on five datasets

and its result is closer to the ground truth in terms of details and colors. In comparison, other methods have problems such as blurred details and texture distortion. Especially, RESCAN [17] and PreNet [18] still have the residue of rain marks.

3.4 Results on Real-World Images

To more strongly demonstrate the adaptability of the proposed method, test results on real rainy images from the dataset DDN-SIRR [38] are randomly selected to show by Fig. 3. The results of RESCAN [17] and MSPFN [19] have residual rain streaks, while RCHNet obtains better results. Moreover, compared to PreNet [18], HINet [22] and QSAM-Net [23], which have texture distortion and blurred details in magnifying boxes, RCHNet preserve details much successfully.

3.5 Ablation Study

To verify the effectiveness of the proposed different components in the architecture, we design some variants of the model to perform ablation studies, which are illustrated as follows: (1) *w/o* CDB: RCHNet directly uses the residual block [41] without CDB. (2) *w/o* DAM: RCHNet without DAM. (3) *w/o* RDB: RCHNet adopts maximum pooling for downsampling without RDB. (4) *w/o* RUB: RCHNet employs transposed convolution for upsampling without RUB. (5) *w/o* all: RCHNet without all components above.

Based on the results presented in Table 4 and Fig. 4, all the variant models showed a drop in the average of the PSNR and SSIM index in five datasets compared to the original RCHNet. Meanwhile, our RCHNet achieved a significant improvements of 4.2 dB on average values across all datasets compared to the baseline. It suggests that the original RCHNet outperforms its variants in terms of deraining. Furthermore, the absence of CDB resulted in the most

Table 4 Effectiveness of Different Components on five dataset

Modules	Test100 [35] PSNR/SSIM (%)	Rain100H [36] PSNR/SSIM (%)	Rain100L [36] PSNR/SSIM (%)	Test2800 [16] PSNR/SSIM (%)	Test1200 [37] PSNR/SSIM (%)
<i>w/o</i> CDB	27.23/86.7	26.28/78.3	33.69/94.1	32.60/92.7	30.82/89.5
<i>w/o</i> DAM	28.88/88.3	29.84/87.4	35.98/94.9	32.83/93.0	32.21/91.1
<i>w/o</i> RDB	30.18/89.0	29.78/87.4	36.86/96.7	33.77/93.9	32.26/91.5
<i>w/o</i> RUB	30.30/90.2	30.39/89.0	37.05/96.8	33.76/93.8	33.19/91.7
<i>w/o</i> all	24.74/82.2	27.4/85.8	30.46/89.1	32.00/92.6	30.93/90.0
RCHNet	30.74/91.0	30.96/89.7	37.61/97.3	33.92/94.2	33.31/92.6

apparent drop in performance, demonstrating that CDB plays the most crucial role in our network.

4 Conclusion

In this paper, we propose a novel network named RCHNet for single-image deraining to extract and integrate rain features of different scales. Specifically, it could adapt to the irregular distribution of rain patterns in space and color, and realize rain removal in complex environments. Experimental results show that our method achieves the highest PSNR and SSIM scores while obtaining strong generalizability at the same time. The visual results also demonstrated that our model could better restore the details and color information of the image. However, our algorithm still has potential for further improvement in terms of the model parameters. By pruning our model to get a more effective and lightweight model, we can improve the efficiency and performance of the algorithm. We expect our model could benefit to solve more deraining tasks in vision applications in the future.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 61404083) and Key Projects of the National Natural Science Foundation of China (Grant No. 52331012).

Author Contributions WZ: Conceptualization, Methodology, Resources, Supervision, Writing—review & editing, Project administration. LY: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—original draft, Visualization. XW: Methodology, Software, Validation, Writing—original draft.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. 2961–2969
2. Zhou W-N, Zhou Y, Zeng X-Y (2022) An attention nested u-structure suitable for salient ship detection in complex maritime environment. *IEICE Trans Inf Syst E105-D(6)*:1164–1171
3. Chen Y-C, Lin Y-Y, Yang M-H, Huang J-B (2020) Show, match and segment: joint weakly supervised learning of semantic matching and object co-segmentation. *IEEE Trans Pattern Anal Mach Intell* 43(10):3632–3647
4. Li Y, Li M, Li Z, Xiao C, Li H (2023) Correction to: Efrnet: efficient feature reuse network for real-time semantic segmentation. *Neural Process Lett* 55(1):873–873
5. Xu X, Jinjia Fu (2021) Scale-aware feature pyramid architecture for marine object detection. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-020-05217-7>
6. Xu F, Wang H, Sun X, Fu X (2022) Refined marine object detector with attention-based spatial pyramid pooling networks and bidirectional feature fusion strategy. *Neural Comput Appl* 34:14881–14894
7. Chen J, Tan C-H, Hou J, Chau L-P, Li H (2018) Robust video content alignment and compensation for rain removal in a CNN framework. 6286–6295

8. Su R, Zhang L, Zhang Y, Xu F, Lu K, Tong N, Li F (2023) Complex scene video frames alignment and multi-frame fusion deraining with deep neural network. *Neural Comput Appl* 35(7):5369–5380
9. Li Y, Tan RT, Guo X, Lu J, Brown MS (2016) Rain streak removal using layer priors. 2736–2744
10. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior. *IEEE Trans Pattern Anal Mach Intell* 33(12):2341–2353
11. Celik T, Tjahjadi T (2011) Contextual and variational contrast enhancement. *IEEE Trans Image Process* 20(12):3431–3441
12. Dong J, Cong Y, Sun G, Zhong B, Xu X (2020) What can be transferred: unsupervised domain adaptation for endoscopic lesions segmentation. 4023–4032
13. Ren W, Tian J, Wang Q, Tang Y (2020) Dually connected deraining net using pixel-wise attention. *IEEE Signal Process Lett* 27:316–320
14. Zhou W-N, Chen K (2022) A lightweight hand gesture recognition in complex backgrounds. *Displays* 74:102226
15. Zhou W-N, Huang X-X, Zeng X-Y (2022) Obstacle detection for unmanned surface vehicles by fusion refinement network. *IEICE Trans Inf Syst* (8)
16. Fu X, Huang J, Zeng D, Huang Y, Ding X, Paisley J (2017) Removing rain from single images via a deep detail network. 3855–3863
17. Li X, Wu J, Lin Z, Liu H, Zha H (2018) Recurrent squeeze-and-excitation context aggregation net for single image deraining. 254–269
18. Ren D, Zuo W, Hu Q, Zhu P, Meng D (2019) Progressive image deraining networks: a better and simpler baseline. 3937–3946
19. 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: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)
20. Jiang K, Wang Z, Yi P, Chen C, Wang G, Han Z, Jiang J, Xiong Z (2023) Multi-scale hybrid fusion network for single image deraining. *IEEE Trans Neural Netw Learn Syst* 34(7):3594–3608
21. Jiang K, Wang Z, Yi P, Chen C, Han Z, Lu T, Huang B, Jiang J (2021) Decomposition makes better rain removal: An improved attention-guided deraining network. *IEEE Trans Circuits Syst Video Technol* 31(10):3981–3995
22. Chen L, Lu X, Zhang J, Chu X, Chen C (2021) Hinet: half instance normalization network for image restoration. 182–192
23. Frants V, Aгаian S, Panetta K (2023) Qsam-net: rain streak removal by quaternion neural network with self-attention module. *IEEE Trans Multimed*. <https://doi.org/10.1109/TMM.2023.3271829>
24. Chen X, Li H, Li M, Pan J (2023) Learning a sparse transformer network for effective image deraining
25. Jiang K, Wang Z, Chen C, Wang Z, Cui L, Lin C-W (2022) Magic ELF: image deraining meets association learning and transformer
26. Jiang K, Wang Z, Yi P, Chen C, Wang Z, Wang X, Jiang J, Lin C-W (2021) Rain-free and residue hand-in-hand: a progressive coupled network for real-time image deraining. *IEEE Trans Image Process* 30:7404–7418
27. Chen X, Pan J, Jiang K, Li Y, Huang Y, Kong C, Dai L, Fan Z (2022) Unpaired deep image deraining using dual contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2017–2026
28. Jiang K, Liu W, Wang Z, Zhong X, Jiang J, Lin C-W (2023) Dawn: direction-aware attention wavelet network for image deraining. In: Proceedings of the 31st ACM international conference on multimedia, pp 7065–7074
29. Huang X, Belongie S (2017) Arbitrary style transfer in real-time with adaptive instance normalization. 1501–1510
30. Zhang R (2019) Making convolutional networks shift-invariant again, PMLR. pp 7324–7334
31. Guo M-H, Xu T-X, Liu J-J, Liu Z-N, Jiang P-T, Mu T-J, Zhang S-H, Martin RR, Cheng M-M, Hu S-M (2022) Attention mechanisms in computer vision: a survey. *Computat Vis Media* 8:1–38
32. Woo S, Park J, Lee J-Y, Kweon IS (2018) Cbam: convolutional block attention module. 3–19
33. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. 7132–7141
34. Wang C, Fan W, Zhu H, Su Z (2020) Single image deraining via nonlocal squeeze-and-excitation enhancing network. *Appl Intell* 50(9):2932–2944
35. Zhang H, Sindagi V, Patel VM (2019) Image de-raining using a conditional generative adversarial network. *IEEE Trans Circuits Syst Video Technol* 30(11):3943–3956
36. Yang W, Tan RT, Feng J, Liu J, Guo Z, Yan S (2017) Deep joint rain detection and removal from a single image. 1357–1366
37. Wei W, Meng D, Zhao Q, Xu Z, Wu Y (2019) Semi-supervised transfer learning for image rain removal. 3877–3886

38. Wang T, Yang X, Xu K, Chen S, Zhang Q, Lau RW (2019) Spatial attentive single-image deraining with a high quality real rain dataset. 12270–12279
39. Huynh-Thu Q, Ghanbari M (2008) Scope of validity of PSNR in image/video quality assessment. *Electron Lett* 44(13):800–801
40. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
41. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. 770–778

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.