Journal of Computers Vol. 36 No. 3, June 2025, pp. 105-113
https://doi.org/10.63367/19911599202506360300 8

Research on Lightweight Super-Resolution Reconstruction Algorithm Based on Progressive Feature Fusion Convolutional

Wen Qu[1], Jianxin Li[2*], Chaoying Wang[2], Jiajun Li[2], and Hongxing Huang[2]

[1] School of Information Engineering, Gannan University of Science and Technology, Ganzhou, 341000, China
QUWEN2020@126.com

[2] School of Electronic Information, Dongguan Polytechnic, Dongguan, 523808, China
279149042@qq.com, wangcy@163.com, 3345340890@qq.com, 2960632949@qq.com

Received 6 July 2024; Revised 6 August 2024; Accepted 6 August 2024

Abstract: This paper proposes a lightweight super-resolution reconstruction algorithm based on a progressive feature fusion convolutional network. The algorithm improves image resolution by incorporating a progressive attention block and a multi-scale pixel attention mechanism to reduce computational complexity. The progressive attention block uses multiple parallel paths to gradually aggregate multi-scale features, achieving a larger receptive field with fewer parameters. The multi-scale pixel attention mechanism builds an efficient upsampling module, ensuring effective utilization of the attention mechanism while maintaining lightweight characteristics. Additionally, the paper introduces a cosine annealing-based training strategy to enhance model expressiveness and image restoration quality without altering the model structure. The experimental results show that the algorithm proposed in this paper performs better than current mainstream algorithms in terms of running speed and efficiency balance.

Keywords: image superresolution, light weight, attention mechanism, convolutional neural network

1 Introduction

Image super-resolution (SR) enhances low-resolution (LR) images to high-resolution (HR) while preserving visual quality, benefiting applications like target detection [1], medical imaging [2], video surveillance, satellite remote sensing, and face recognition.

Deep learning has significantly advanced SR, starting with Dong et al. [3] who introduced SRCNN. Shao et al. [4] furthered this with VDSR in 2023, using 20 layers of residual blocks to improve training. This approach led to a focus on deeper networks with refined structures and residual learning. While models like RCAN [5] and RDN [6] improve peak signal-to-noise ratio (PSNR), they often face challenges with parameter size and efficiency, which can lead to overfitting and limit application on low-cost devices.

Recent research has turned to lightweight models to balance size and performance. Initial efforts, including FSRCNN [7] and Lap SRN [8], encountered precision issues. Advances such as information distillation and neural architecture search (NAS) strategies, exemplified by ESRN [9], have sought to address these. Newer models like MAFFSRN [10] use multi-attention mechanisms, while Lattice Net [11] features a butterfly structure to enhance generalization. A2F-M [12] achieves high efficiency by projecting features into a shared space. Despite progress, there remains room for improvement in developing more efficient and lightweight SR models. This paper proposes a new SR algorithm with a lightweight network structure based on progressive feature fusion to better balance model size and reconstruction quality.

* Corresponding Author

2 Research on Lightweight Superresolution Based on Progressive Feature Fusion

2.1 Progressive Feature Fusion Network

To improve efficiency and develop a lightweight image super-resolution model, this paper introduces the Improve Multi-scale Fine-grained Model (IMFM), as shown in Fig. 1. The model features a Progressive Attention Block (PAB) for fine-grained feature extraction and a Multi Scale Weight Allocation Module (MSWAM) that uses hierarchical residuals to expand the network's receptive field without adding convolution layers.

The improved Coordinate Attention Module (CAM) focuses on high-frequency positional information, preserving key details while filtering out irrelevant data. The Upsampling Module based on Multi-scale Weight Allocation (MWAUM) combines MSWAM with bilinear interpolation to capture key feature information and reduce parameters while maintaining performance.

A cosine function-based learning rate adjustment is used for optimal model training. PAB is essential for progressive feature aggregation, and MSWAM enhances feature extraction in both PAB and MWAUM.

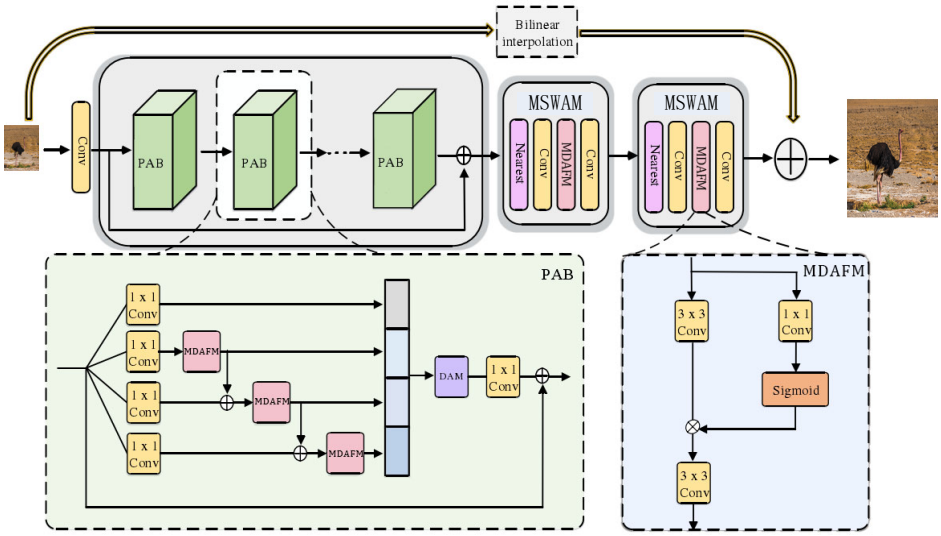

Fig. 1. Schematic diagram of progressive feature fusion network

2.2 Overall Network Structure

The proposed lightweight Progressive Feature Fusion Network comprises four components: initial convolution, progressive attention modules, feature upsampling module, and global residual connection. Here, 3×3 convolution extracts rough features from blurry images, and then inputs the features into 3 channels to convert them into a high-dimensional feature space:

$$F_0 = Conv_{3\times3}(I_{LR}) \tag{1}$$

The input image with 3 channels produces extracted features, denoted as I_{LR}. To maintain a lightweight design, the network uses a convolution layer with 40 channels, significantly reducing parameter size. A series of Progressive Attention Blocks (PABs) are then stacked to progressively enhance feature representation:

$$F_k = PAB_k(F_{k-1}),\ k = 1,2...,n \tag{2}$$

In this network, F_k and F_{k-1} denote the input and output features of the *k-th* Progressive Attention Block (PAB). After extracting fine features through multiple PABs, the upsampling module increases the spatial size of the feature map:

$$I_{up} = H_{MPAU}(F_n + F_0) \tag{3}$$

In the network, I_{up} represents the image after upsampling, and H_{MWAUM} is the MSWAM module used for feature upsampling. As shown in Fig. 1, IMFM incorporates the Multi Scale Weight Allocation Module (MSWAM) into the upsampling process to enhance convolution representation. The MSWAM module, which includes bilinear interpolation and MSWAM between two convolution layers, ensures a compact model size due to its parameterless neural network interpolation. Following classical lightweight super-resolution methods [13-18], IMFM generates the final super-resolution (*SR*) image using a global residual connection:

$$I_{SR} = I_{up} + H_{bi}(I_{LR}) \tag{4}$$

In the network, H_{bi} denotes bilinear interpolation, which resizes I_{LR} to match the dimensions of I_{SR}. This result is combined with I_{UP} to produce the final SR image. The IMFM is optimized using the mean absolute error loss function, also known as L_{MAR} loss, defined as follows:

$$L_{MAR} = \frac{1}{N}\sum_{i=1}^{N} \| I_{HR} - I_{SR} \| \tag{5}$$

In the network, N denotes the number of paired image blocks in a training batch, with I_{HR} representing the target high-resolution image. Meanwhile, I_{SR} denotes the generated SR image.

2.3 Progressive Attention Block

The Progressive Attention Block (PAB) uses multiple parallel, interconnected paths to split the feature map into *s* subsets along the channel dimension. This paper employs 1×1 convolution instead of direct splitting, as shown by Liu et al. [18], for its superior nonlinear transformation and efficiency with GPU acceleration. During this process, the feature map will be reduced to a sample space of the same size. However, there are fewer channels (m, where m<s). While the feature x_i from the first path is directly connected to an intermediate layer, other paths pass through the Multi-scale Pixel Attention Module (MSWAM) to generate features y_i, represented as F^i_{MPAM}. These parallel paths are combined similarly to residuals, adding features x_i to F^{i-1}_{MPAM} before processing in the next path. The processing flow of the PAB module is thus:

$$y_i = \begin{cases} x_i & i=1 \\ F^i_{MPAM}(x_i) & i=2 \\ F^i_{MPAM}(x_i + y_{i-1}) & i>2 \end{cases} \tag{6}$$

By allowing feature y_i on path i to receive all previous feature maps $\{y_k, k < i\}$, the output features of convolutional layers have a larger receptive space, creating multi-scale feature combinations. This enhances fine-grained feature extraction. Compared to traditional *SR* networks, the PAB gradually converges features, achieving competitive performance with lower operational complexity.

2.4 Coordinate Attention Module

Coordinate Attention generates spatially aware feature maps by aggregating features from two directions on the feature map plane and performing different convolution operations. Traditional super-resolution networks based on attention mechanisms use channel and spatial attention, but they often ignore location information, which is crucial for creating spatial selective attention feature maps [19]. This paper enhances coordinate attention by removing normalization layers to reduce contrast distortion in color distribution.

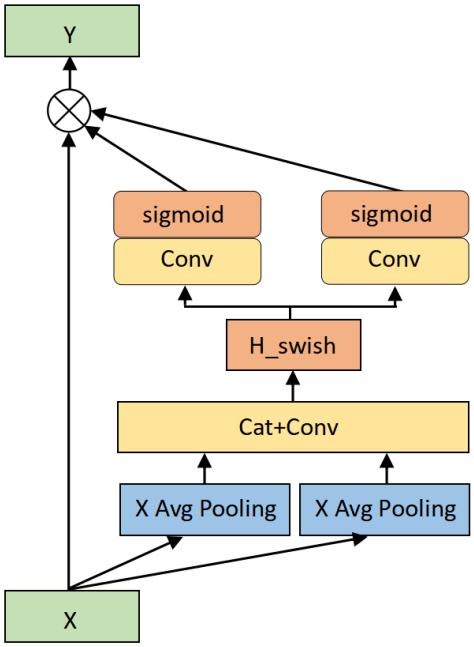

Fig. 2. Coordinate attention

As shown in Fig. 2, the input feature X, after passing through a 1×1 convolution, has a shape of {*nf*, 64, 64}, where *nf* represents the number of channels, and 64 indicates the height and width. Adaptive average pooling (Avg Pool2d) is applied both horizontally and vertically to the input feature map. This process smooths the noise in the LR image by averaging feature values, reducing dimensions, and retaining essential information. The pooling kernels (H, 1) and (1, W) encode each channel along the horizontal and vertical axes, respectively. For an input X with height h, the output Y in channel c is represented as follows.

$$Y_c^h(h) = \frac{1}{w}\sum_{0\ll i<W} X_c(h,i) \tag{7}$$

Assuming that the width of X is set to w, the output Y of channel c can be calculated by formula (8):

$$Y_c^w(w) = \frac{1}{H}\sum_{0\ll i<H} X_c(j,w) \tag{8}$$

The adjustment layer generates two position sensitive feature maps Y^h and Y^w, each of which retains the same number of channels as the input X. The shape of {*nf*, 1, 64} and Y^w, respectively, reflect their sensitivity to horizontal and vertical positions. Then combine these two feature maps together and align them to capture the spatial dependencies of the two axes. After the stacking process, perform 1×1 convolution operation on the combined feature map to obtain the final output feature map f. This convolution further refines the combined information, resulting in a more cohesive representation of input features.

$$f = \delta(conv_{1\times1}(cat(Y^h, Y^w)) \tag{9}$$

In formula 9, $cat(\cdot)$ represents the superposition operation, and *H_Sigmoid* represents the S-shaped growth curve δ. The input is divided into two tensors f^h and f^w along the spatial dimension. Then, by combining X with the channel, f^h and f^w are converted into variant tensors, and the calculation formula is as follows:

$$g^h = \sigma(conv_{1\times1}(f^h)) \tag{10}$$

$$g^h = \sigma(conv_{1\times1}(f^w)) \tag{11}$$

In equations 10 and 11, σ represents the S-shaped growth curve. Combine the weight distribution of the input sequence and then perform weighted summation to obtain the final output Y. The formula is as follows:

$$Y_c(i,j) = X_c(i,j) * g_c^h(i) * g_c^w(j) \tag{12}$$

2.5 Training Strategy

The incremental gradient descent method can iteratively update weights and bias terms, and combined with optimization methods such as Adam [20], deep neural networks perform well in computer vision and language tasks. SGD is a mainstream model optimization method, but it may get stuck in local minima during training, especially for lightweight models. To mitigate this, this paper employs a hot restart mechanism [21], periodically resetting the learning rate to a high value and then gradually decreasing it. The learning rate is attenuated using the cosine annealing function, as follows:

$$\eta_{cur} = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})(1 + \cos(\frac{T_{cur}}{RI}\pi)) \tag{13}$$

Where η_{cur} represents the current learning rate, T_{cur} the number of traversals since the last restart, $\eta_{\min}$ and $\eta_{\max}$ the range of learning rates, and RI the restart interval. After RI traversals, an SGD restart is triggered. When T_{cur} = 0, the cosine function outputs 1, making $\eta_{cur} = \eta_{\max}$. Once $T_{cur} = RI$, the cosine function outputs -1, so $\eta_{cur} = \eta_{\min}$.

3 Experimental Results and Analysis

3.1 Experimental Environment

The experimental environment configuration of this article is shown in Table 1 below.

Table 1. Experimental environment configuration

CPU	RAM	GPU	OS	Language	DL framework
5 GHz	64 GB	NVIDIA RTX4090 24G	Ubuntu 18.04	Python 3.7	Python-based framework

3.2 Datasets and Parameter Settings

In the experiment, 1200 images from reference [22] were used as the dataset, with 900 for training and 300 for testing. Use bicubic interpolation and scaling factors (2 ×, 3 ×, and 4 ×) to downsample high-resolution images. The standard test set includes Set5, Set14, B100, and Urban100. Evaluate model performance by calculating the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) on the Y (luminance) channel of the YCbCr color space.

3.3 Analysis of Experimental Results

Model Analysis.

(1) Ablation experiment

Extensive ablation experiments were conducted to evaluate the effectiveness of PAB, MSWAM, and CAM components in IMFM. The basic model lacking these components uses a 3x3 convolutional sequence to maintain the same channel count. As shown in Table 2, the second and third rows are a comparison made on the basis that adding PAB can improve the PSNR of four datasets, reduce model parameters by 63.7K, and reduce multiplica-

tion operations by 3.5G. This indicates that PAB improves performance while reducing parameters. Including MSWAM, only an additional 21.8K parameters are needed to further improve PSNR, verifying that pixel level attention can improve performance with lower computational complexity. Finally, integrating CAM, a lightweight module with coordinated attention, can enable the entire model to achieve optimal performance compared to other models.

Table 2. Comparison of the effects of models trained with different learning rate attenuation schemes

	CARN	MAFFSRN	IMDN	A2F-M	IMFM
step_lr	33.12	33.18	33.1	33.24	33.3
cos_lr	33.21	33.25	33.2	33.31	33.41
gain	0.09	0.07	0.1	0.06	0.11

(2) Learning rate attenuation scheme

The training strategy significantly affects the final performance of the model, and the learning rate is a key factor. Usually, SR models use a progressive decay scheme. However, this article adopts a cosine annealing scheme with thermal restart. Fig. 3 shows two learning rate decay methods: the blue line represents the traditional gradual descent, while the red line represents the proposed cosine annealing function. In the step_1r scheme, the learning rate is halved every 200 stages. In contrast, the proposed dynamic adjustment learning rate scheme resets the learning rate to its initial value every 250 cycles. This article uses the above two schemes to train the proposed network, and the convergence curve during training is shown in Fig. 4. Training the network using these schemes shows that step_1r can achieve a smoother convergence curve, but the learning rate becomes too small to improve performance in the later stages. Although the convergence curve fluctuates greatly, the proposed cos_1r scheme achieves higher performance.

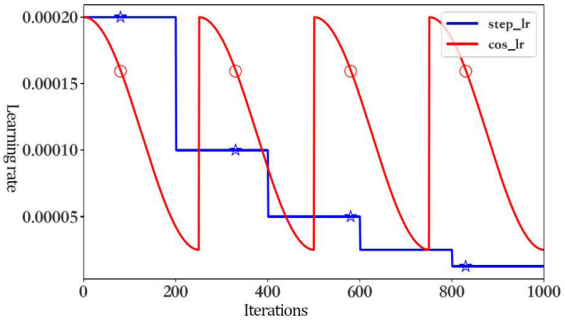

Fig. 3. Schematic diagram of two attenuation modes with different learning rates

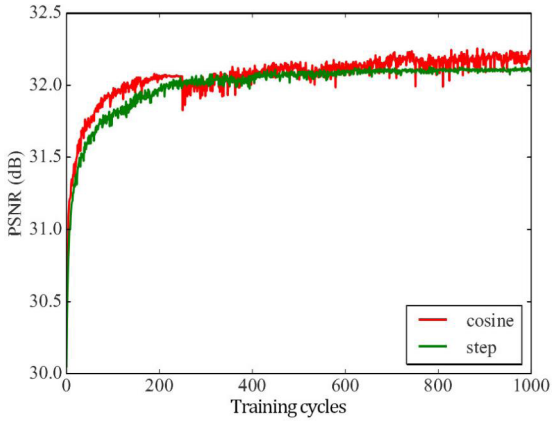

Fig. 4. Schematic diagram of two attenuation modes with different learning rates

(3) Up sampling module

The effectiveness of the proposed MWAUM was demonstrated by comparing it with commonly used upsampling modules such as sub-pixel layers [23] and deconvolution layers. Additionally, two models were designed to replace MWAUM with these substitutes for feature upsampling. Table 3 shows that the proposed MWAUM achieved better results with fewer parameters, indicating its contribution to building an efficient and lightweight model.

Table 3. Compare different upsampling modules on the 4× scale of the Set5 dataset

Model	Number of parameters	Running time (s)	PSNR
Deconvolution layer	576.169	0.3413	26.21
Sub pixel layer	594.011	0.3324	26.23
MWAUM	596.362	0.3437	26.25

Run Time and Model Complexity. The proposed IMFM was compared with other methods in terms of runtime and model complexity. The detailed results in Table 4 include model parameters, GPU consumption, and runtime on GPU and CPU.

Table 4. The comparison of model complexity is conducted on a 4× scale using the Urban100 dataset

Model name	Number of parameters	GPU consumption (M)	PSNR	
			CPU	GPU
EDSR [24]	44080K	8806	140.005	0.212
RDN [25]	11120K	15881	71.091	0.152
WSR [15]	1080K	6635	41.018	0.081
MAFFSRN [14]	980K	4410	29.113	0.072
IMFM	670 K	5109	21.225	0.053

Table 4 compares the proposed IMFM with non-lightweight models like EDSR and RDN, which have over 10M parameters. IMFM demonstrates significant advantages, having the fewest parameters and the fastest running speed, being nearly five times faster than EDSR. The results indicate that IMFM achieves a better balance between parameter size and operational complexity compared to leading methods like MAFFSRN and WSR, highlighting its strong practical application potential.

Display and Analysis of Subjective Effects. Fig. 5 and Fig. 6 demonstrate the reconstruction results of the images "253027.jpg" from the BSD100 dataset and "img2012" from the Urban100 dataset, both at 4x magnification. For "253027.jpg," the proposed algorithm excels in restoring edges and texture details, achieving a more realistic result compared to other methods. Similarly, for "img2012," while other algorithms introduce errors in texture direction and a checkerboard effect, the proposed algorithm accurately reconstructs the texture details, producing an image that closely resembles the original.

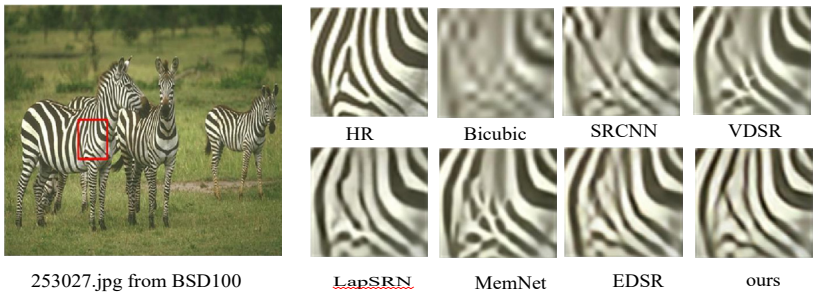

Fig. 5. Results of different algorithms for the "253027.jpg" in BSD100 dataset under 4× magnification

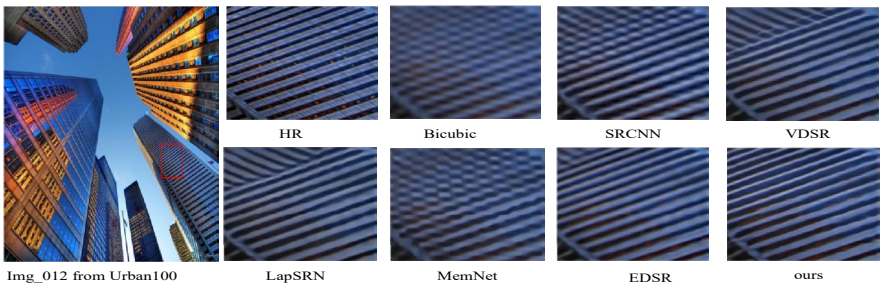

Fig. 6. Results of different algorithms for the "img_012" in Urban100 dataset under 4× magnification

4 Conclusion

This article proposes a new progressive attention block (PAB) as a backbone network component to extract useful information; Designed a multidimensional attention fusion module that significantly improves the perceptual visual perception of the entire network without adding additional convolutional layers, avoiding blind deepening of the network; Meanwhile, a Coordinate Attention Module (CAM) was designed to generate low complexity attention expressions. The ablation experiment shows that the components in the network are smaller than those in the normal structural model, with lower computational costs, ensuring high accuracy of image SR. This article abandons the traditional learning rate decay strategy and proposes a new training method based on cosine annealing learning scheme, which can significantly tap the potential of the model without changing the network structure. The experimental results show that compared with this algorithm, this method reduces the number of parameters, performs better, and achieves a good balance between the accuracy of objective evaluation indicators and operational efficiency; In terms of visual quality, this method can better reconstruct texture details.

Future research will focus on collaborative optimization of algorithms, data, and hardware.

(1) Algorithm Optimization: Develop lightweight models (e.g., dynamic pruning and knowledge distillation) and cross-modal fusion technologies, leveraging unsupervised learning to address complex degradation in real-world scenarios.

(2) Data Enhancement: Construct realistic degradation datasets using multi-sensor systems and enable simulation-to-reality migration through meta-learning.

(3) Hardware Innovation: Design edge-device-specific acceleration architectures to reduce power consumption and support real-time processing.

Acknowledgement

This paper was supported by the Dongguan Science and Technology Ombudsman Project in 2022 (No. 20221800500732); 2023 Dongguan Social Development Technology Project (No. 2023pz08); Science and Technology Research Project in Department of Education of Jiangxi Province (No. GJJ2203611); Special Fund for Guangdong Province's Science and Technology Innovation Strategy (No. pdjh2024b663), and the Dongguan Science and Technology Ombudsman Project (No. 20231800500282), and the Dongguan Science and Technology of Social Development Program (No. 20231800903592 & 20221800905842), Special fund for electronic information engineering technology specialty group of national double high program of Dongguang Polytechnic (No. ZXD202303, NO. ZXB202203, NO. ZXD202204, NO. ZXC202201), the Special for key fields of colleges and universities in Guangdong Province under Grant (NO. 2024ZDZX1091 & NO. 2024ZDZX1092 & NO. 2024ZDZX1093).

References

[1] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-Based Convolutional Networks for Accurate Object Detection and Segmentation, IEEE Transactions on Pattern Analysis & Machine Intelligence 38(1)(2016) 142-158.

[2] J. Ma, J. Yu, S. Liu, L. Chen, X. Li, J. Feng, Z. Chen, S. Zeng, X. Liu, S. Cheng, PathSRGAN: Multi-Supervised Super-Resolution for Cytopathological Images Using Generative Adversarial Network, IEEE Transactions on Medical Imaging 39(9)(2020) 2920-2930.

[3] C. Dong, C.C. Loy, K.M. He, X.O. Tang, Image Super-Resolution Using Deep Convolutional Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 38(2)(2016) 295-307.

[4] G.F. Shao, Q. Sun, Y.L. Gao, Q.Y. Zhu, F.Q. Gao, J.F. Zhang, Sub-Pixel Convolutional Neural Network for Image Super-Resolution Reconstruction, Electronics 12(17)(2023) 3572.

[5] G.Y. Yang, Y.M. Wang, C. Yi, Z.Q. Wang, A New super-resolution restoration method with Generated Adversarial Network for underground video images in coal mines, Journal of Physics Conference Series 2031(1)(2021) 012011.

[6] P. Purkait, B. Chanda, Super Resolution Image Reconstruction Through Bregman Iteration Using Morphologic Regularization, IEEE Transactions on Image Processing 21(9)(2012) 4029-4039.

[7] W-S Lai, J-B. Huang, N. Ahuja, M-H. Yang, Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[8] D.H. Song, C. Xu, X. Jia, Y.Y. Chen, C.J. Xu, Y.H. Wang, Efficient Residual Dense Block Search for Image Super-Resolution, in: Proc. of the AAAI Conference on Artificial Intelligence, 2020.

[9] S. Karthick, N. Muthukumaran, Deep RegNet-150 Architecture for Single Image Super Resolution of Real-time Unpaired Image Data, Applied Soft Computing 162(2024) 111837.

[10] X.T. Luo, Y. Xie, Y.L. Zhang, Y.Y. Qu, C.H. Li, Y. Fu, LatticeNet: Towards Lightweight Image Super-Resolution with Lattice Block, in: Proc. Computer Vision-ECCV2020, 2020.

[11] Q.P. Chen, W.S. Wen, J.H. Qin, GlobalSR: Global context network for single image super-resolution via deformable convolution attention and fast Fourier convolution, Neural Networks 180(2024) 106686.

[12] J.H. Qin, R.M. Zhang, Lightweight single image super-resolution with attentive residual refinement network, Neurocomputing 500(2022) 846-855.

[13] J.H. Qin, Y.J. Huang, W.S. Wen, Multi-scale feature fusion residual network for Single Image Super-Resolution, Neurocomputing 379(2020) 334-342.

[14] A. Muqeet, J. Hwang, S. Yang, J.H. Kang, Y.G. Kim, S.H. Bae, Multi-attention Based Ultra Lightweight Image Super-Resolution, in: Proc. of Computer Vision - ECCV 2020 Workshops, 2020.

[15] H.R. Zhang, Z. Jin, X.J. Tan, X.Y. Li, Towards Lighter and Faster: Learning Wavelets Progressively for Image Super-Resolution, in: Proc. of the 28th ACM International Conference on Multimedia (MM'20), 2020.

[16] S.I. Fernando, J.T. Martineau, R.J. Hobson, T.N. Vu, B. Baker, B.D. Mueller, R. Menon, E.M. Jorgensen, J.M. Gerton, Simultaneous Spectral Differentiation of Multiple Fluorophores in Super-resolution Imaging Using a Glass Phase Plate, Optics express 31(20)(2023) 33565-33581.

[17] C. Qu, X.Y. Chen, Q.H. Xu, J. Han, Frequency-Aware Degradation Modeling for Real-World Thermal Image Super-Resolution, Entropy 26(3)(2024) 209.

[18] Y.H. Liu, Z.Z. Chu, L.F. Wei, A Channel Contrastive Attention-Based Local-Nonlocal Mutual Block on Super-Resolution, Ieice Transactions on Information and Systems E107-D(9)(2024) 1219-1227.

[19] I. Loshchilov, F. Hutter, SGDR: Stochastic Gradient Descent with Warm Restarts, in: Proc. 5th International Conference on Learning Representations, 2017.

[20] Z.J. Gao, L. Shen, Z.K. Song, H. Yan, JOA-GAN: An improved single-image super-resolution network for remote sensing based on GAN, IET Image Processing 18(12)(2024) 3530-3544.

[21] H. Lin, J.J. Yang, Image super-resolution reconstruction based on implicit image functions, IET Image Processing 18(10)(2024) 2690-2701.

[22] E. Agustsson, R. Timofte, NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.

[23] W.Z. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[24] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced Deep Residual Networks for Single Image Super-Resolution, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.

[25] Y.L. Zhang, Y.P. Tian, Y. Kong, B.N. Zhong, Y. Fu, Residual Dense Network for Image Restoration, IEEE Transactions on Pattern Analysis and Machine Intelligence 43(7)(2021) 2480-2495.