Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 17, No. 2, 2025

A Review of Deep Learning-Based Image Super-Resolution Reconstruction Methods

Wenqiang Xi [1, *], Zairila Juria Zainal Abidin[2], Cheng Peng [3], Tadiwa Elisha Nyamasvisva [4]

[1] Center for Postgraduate Studies, Infrastructure University Kuala Lumpur, 43000, Malaysia
[2] Faculty of Architecture and Built Environment, Infrastructure University Kuala Lumpur, 43000, Malaysia
[3] Faculty of Physics and Electrical Engineering, Weinan Normal University, 714000, China
[4] Department of Computing, Faculty of Engineering Science and Technology, Infrastructure University Kuala Lumpur, 43000, Malaysia

[*] **Corresponding author**: Wenqiang Xi (Email: 213923103@s.iukl.edu.my)

Abstract: Image Super-Resolution (SR) technology aims to reconstruct High-Resolution (HR) images from Low-Resolution (LR) images, holding significant application value in fields such as medical imaging analysis, satellite remote sensing, video enhancement, and security surveillance. In recent years, deep learning methods have significantly advanced the development of image super-resolution technology due to their powerful feature extraction capabilities. This paper systematically reviews the current research status of Single Image Super-Resolution (SISR) technology, focusing on three mainstream deep learning frameworks: Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Transformers, and summarizes their latest research progress. Firstly, the paper introduces the fundamental principles of traditional super-resolution methods and their limitations in complex scenarios. Secondly, it provides a detailed analysis of the network architectures, optimization strategies, and performance advantages of various deep learning-based super-resolution models. Finally, the paper discusses the challenges currently faced by deep learning-based super-resolution technology and outlines potential future research directions.

Keywords: Image Super-Resolution; Deep Learning; Convolutional Neural Networks; Generative Adversarial Networks; Transformer.

1. Introduction

Computer vision is a significant research direction in the field of deep learning and has garnered extensive attention in recent years. Among its various tasks, Single Image Super-Resolution (SISR), as a fundamental problem in computer vision, has become a research hotspot. Image resolution is typically measured in Pixels Per Inch (PPI), which denotes the number of pixels contained within a unit length. Higher resolution implies a denser distribution of pixels, thereby providing richer image details and superior visual quality. SISR, also referred to as image upscaling, upsampling, or interpolation, primarily aims to effectively increase pixel density during the process of image enlargement while restoring lost high-frequency details as much as possible. For numerous practical applications, such as medical imaging, remote sensing image analysis, and video enhancement, the recovery of high-quality texture details plays a crucial role in improving image discernibility and information representation.

Image super-resolution reconstruction is inherently an ill-posed problem, meaning that for a given low-resolution input, there may exist multiple corresponding high-resolution solutions. To address this issue, traditional methods often rely on introducing reliable prior information to constrain the solution space. Based on this approach, researchers have proposed various classical super-resolution algorithms, primarily including gradient-based prior methods, statistical learning-based methods, and sparse representation-based methods, among others.

2. Traditional Image Super-Resolution Methods

Traditional super-resolution methods can be broadly categorized into three main classes: interpolation-based methods, reconstruction-based methods, and learning-based methods.

Interpolation-based methods (e.g., bilinear interpolation, bicubic interpolation, Lanczos interpolation) estimate missing pixel values using mathematical interpolation functions. While these methods are computationally efficient, they often struggle to effectively restore high-frequency details

Reconstruction-based methods (e.g., edge-preserving reconstruction, regularization-based optimization) leverage prior information about the image to solve the high-resolution image through constrained optimization problems. However, these methods are computationally expensive and prone to artifacts in complex scenes.

Learning-based methods (e.g., sparse representation, dictionary learning) improve reconstruction quality to some extent by learning the mapping relationship between LR and HR images from large datasets. Nevertheless, their performance is still limited by the representational capacity of the features.

In summary, while traditional methods laid the foundation for image super-resolution, their reliance on handcrafted features and limited representational power constrained their effectiveness, particularly in complex scenarios. The advent of deep learning has since revolutionized the field by enabling data-driven, end-to-end learning of intricate mappings between LR and HR images.

2.1. Interpolation-Based Methods

Interpolation-based methods primarily rely on image interpolation techniques (also known as image scaling). The core idea is to use interpolation functions or kernels to estimate unknown pixel values based on known pixel information. These methods calculate the intensity values of target pixels through weighted computations of known pixels, thereby reconstructing high-resolution images in the spatial domain. Finally, image restoration techniques are applied to optimize the reconstruction results, reducing noise and mitigating blurring effects to enhance image clarity and visual quality.

Common interpolation algorithms include nearest-neighbor interpolation, bilinear interpolation, and bicubic interpolation. Nearest-neighbor interpolation assigns the grayscale value of the target pixel to that of its closest neighboring pixel. While computationally efficient, this method may introduce jagged edges. Bilinear interpolation, on the other hand, utilizes four adjacent pixels of the target pixel to perform linear interpolation in both horizontal and vertical directions, thereby improving smoothness to some extent. In contrast, bicubic interpolation extends the interpolation range further by using 16 neighboring pixels around the target pixel for cubic interpolation calculations. This approach preserves image details while reducing blurring, resulting in higher reconstruction quality.

Despite the advantages of simplicity and computational efficiency, interpolation-based methods exhibit significant limitations in super-resolution tasks. Particularly in regions with abrupt pixel changes, such as edges and textures, these methods struggle to accurately restore high-frequency details, often leading to artifacts like jagged edges, blocky effects, and blurring, which degrade visual quality. Consequently, the super-resolution reconstruction performance of interpolation-based methods is limited in practical applications, making them insufficient for meeting the demands of high-quality reconstruction.

2.2. Reconstruction-Based Methods

Reconstruction-based methods primarily rely on probability theory and set theory. The core idea is to combine low-resolution (LR) images with prior knowledge to construct an optimization model for recovering high-resolution (HR) images. These methods first compute local or global prior information of the image to establish a mapping relationship between LR and HR. Subsequently, data constraints are constructed to model the LR image, and regularization strategies are employed to enhance image prior information. Finally, optimization algorithms are used to solve for the HR image. Depending on the approach, reconstruction methods can be categorized into frequency-domain methods and spatial-domain methods.

In early research, first proposed the concept of frequency-domain-based super-resolution reconstruction, which utilizes multiple frames of images in the Fourier transform domain to recover additional high-frequency information, thereby enhancing image resolution. However, this method did not fully account for motion blur, diffusion effects, and noise interference in images, making it suitable only for ideal noise-free environments and limited in complex noise conditions. To improve reconstruction quality, subsequent studies proposed various denoising algorithms, such as recursive least squares, discrete cosine transform (DCT), and wavelet transform, which effectively enhanced the visual quality of super-resolution images.

Spatial-domain methods primarily simulate spatial factors affecting image quality, such as optical blur and motion blur, to better meet the practical requirements of super-resolution (SR) reconstruction. Compared to frequency-domain methods, spatial-domain methods can more flexibly handle complex imaging degradation processes, enabling higher-quality image reconstruction across different scenarios. Common spatial-domain SR methods include non-uniform sampling interpolation, maximum a posteriori (MAP) estimation, iterative back-projection (IBP), projection onto convex sets (POCS), and hybrid algorithms.

These methods leverage image prior information to varying degrees and employ optimization strategies to reduce noise and blurring effects, thereby improving super-resolution reconstruction. Among them, MAP methods optimize the LR-to-HR mapping based on probabilistic models, IBP methods iteratively correct errors through back-projection, and POCS methods solve for the HR image using constrained optimization. With further research, many hybrid methods have been developed to combine the strengths of different algorithms, enhancing reconstruction accuracy and adaptability.

2.3. Shallow Learning-Based Methods

Shallow learning-based super-resolution methods primarily achieve image reconstruction by learning the mapping relationship between low-resolution (LR) and high-resolution (HR) images. These methods typically utilize large-scale training samples to construct a transformation model from LR to HR and apply the learned mapping to input LR images to predict the corresponding SR images.

Shallow learning-based super-resolution methods can be categorized into several representative approaches, including machine learning-based methods, manifold learning-based methods, example-based learning methods, and sparse coding-based methods.

These shallow learning-based approaches have laid the groundwork for more advanced deep learning techniques, which have since revolutionized the field of image super-resolution by enabling more complex and accurate mappings between LR and HR images.

3. Convolutional Neural Network (CNN)-Based Methods

Convolutional Neural Networks (CNNs) are a type of feedforward neural network trained using gradient descent algorithms. Their unique local connectivity and weight-sharing mechanisms endow them with significant advantages in image feature learning and representation, making them widely applicable in the field of image processing. CNN-based single-image super-resolution (SISR) reconstruction methods leverage the structural characteristics of CNNs to maximize the Peak Signal-to-Noise Ratio (PSNR) as the optimization objective. These methods aim to enhance the detail authenticity and fidelity of reconstructed images, as evidenced by significant improvements in objective evaluation metrics such as PSNR and Structural Similarity Index (SSIM).

3.1. Shallow Convolutional Neural Network Methods

In 2014, Dong et al. pioneered the application of deep learning to single-image super-resolution (SISR) tasks by proposing the first CNN-based SISR network model, SRCNN (Super-Resolution Convolutional Neural Network). Compared to traditional super-resolution methods, SRCNN achieved significant improvements in both reconstruction quality and computational speed. However, due to its pre-upsampling approach where the input image is upsampled using bicubic interpolation before being fed into the network, the model faced limitations in computational complexity and training convergence speed. Additionally, the relatively simple architecture of SRCNN hindered its ability to fully leverage contextual information from the image, thereby limiting its reconstruction capabilities.

To address these issues, Dong et al. introduced an improved version, FSRCNN (Fast SRCNN), in 2016. While retaining the core features of SRCNN, FSRCNN incorporated a post-upsampling mechanism. The network consists of a feature extraction layer, a shrinking layer, a nonlinear mapping layer, an expanding layer, and a deconvolution layer. The key improvement lies in replacing the pre-upsampling framework with a post-upsampling framework, where high-resolution reconstruction is achieved through a deconvolution layer at the end of the network. This modification reduces computational complexity and enhances efficiency. Figure 1 shows the difference between SRCNN and FSRCNN models.

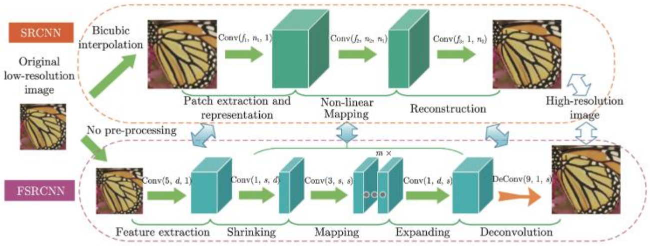


Figure 1. Comparison of FSRCNN and SRCNN network structure.

In the same year, Shi et al. proposed another efficient SR network model, ESPCN (Efficient Sub-Pixel Convolutional Network). Similar to FSRCNN, ESPCN adopts a post-upsampling strategy. However, instead of relying on deconvolution operations, ESPCN utilizes a sub-pixel convolutional layer for upsampling, enabling more efficient resolution enhancement.

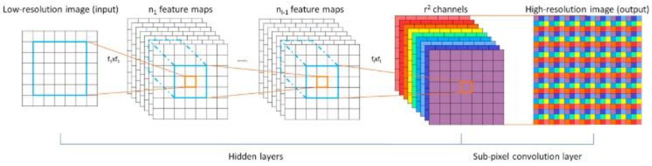


Figure 2. The Efficient Sub-Pixel Convolutional Neural Network (ESPCN).

Since sub-pixel convolution directly learns the mapping from low-resolution to high-resolution images and reduces computational redundancy, ESPCN outperforms traditional CNN architectures in both speed and performance. This approach provided new insights for the design of subsequent super-resolution networks.

3.2. Residual Network-Based Methods

To enhance the performance of network models and improve image feature extraction capabilities, the most straightforward approach is to increase the depth or width of the network, thereby expanding the number of network parameters. However, such simple network expansion can easily lead to issues like gradient vanishing, gradient explosion, and network degradation. To address these challenges, He et al. proposed the Residual Network (ResNet), As shown in Figure 3. which introduces shortcut connections or skip connections into plain networks, transforming them into residual networks. This effectively mitigates gradient and degradation problems in deep networks.

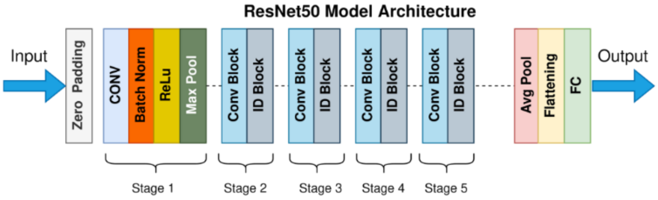


Figure 3. The architecture of ResNet50.

Inspired by the VGG-net deep convolutional neural network from the ImageNet classification competition, Kim et al. first applied ResNet to the image super-resolution task in 2016, proposing the Very Deep CNN for SR (VDSR) model with 20 weight layers, As shown in Figure 4. Considering the high similarity of low-frequency information between low-resolution and high-resolution images, VDSR employs a residual learning strategy, focusing on learning the residual high-frequency information between the two, thereby significantly reducing training time and improving training efficiency. In the same year, Mao et al. proposed the Residual Encoder-Decoder Network (RED-Net), which is not only suitable for image super-resolution but also capable of handling other image restoration tasks like image denoising. In 2017, Lim et al. optimized SRResNet , proposing the Enhanced Deep SR (EDSR) network. The innovation of EDSR lies in removing the Batch Normalization (BN) layers from SRResNet, further enhancing network performance.

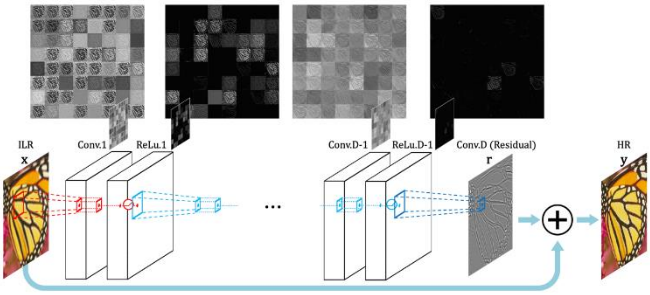


Figure 4. Very Deep Convolutional Networks for image Super Resolution (VDSR).

In 2018, Li et al. proposed a Multi-Scale Residual Network (MSRN) that achieves multi-scale super-resolution tasks with a single model without relying on any training tricks. MSRN combines ResNet with multi-scale convolutional kernels to extract local features at different scales and fuses them with global features, thereby fully utilizing the feature information of low-resolution images to generate high-quality reconstructed images.

In 2021, Lan et al. pointed out that most CNN-based network models fail to fully utilize underlying features, limiting network performance. To address this, they proposed two novel network models: the Cascading Residual Network (CRN). As shown in Figure 5, which includes multiple local shared groups to promote feature fusion and gradient propagation, thereby more efficiently extracting image features; and the Enhanced Residual Network (ERN), which captures long-range spatial features through a dual global path structure, achieving more powerful feature representation capabilities. Through structural optimization, CRN and ERN achieve performance comparable to or even superior to EDSR with significantly reduced parameter counts.

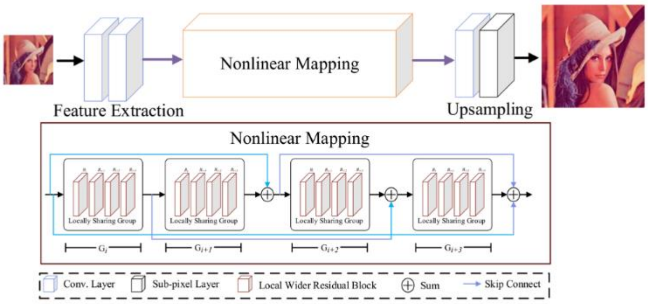
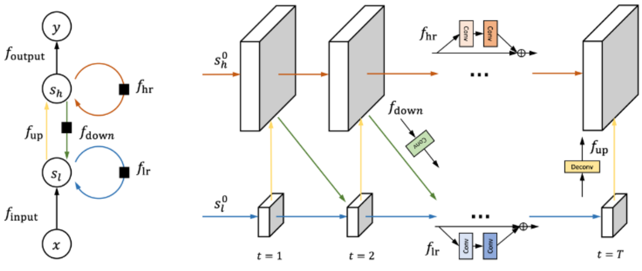


Figure 5. Architecture of the CRN.

3.3. Recurrent Neural Network (RNN)-Based Methods

To effectively control the number of parameters in deep networks and avoid issues such as overfitting caused by increasing network depth, Kim et al. were the first to introduce Recurrent Neural Networks (RNNs) into the field of image super-resolution. In 2016, they proposed the Deeply-Recursive Convolutional Network (DRCN), which incorporates the idea of residual learning. This network includes up to 16 recursive layers and enhances network performance without significantly increasing the number of parameters by leveraging recursive learning in partial convolutional layers. Subsequently, Tai et al. further integrated the advantages of ResNet and RNN, proposing the Deep Recursive Residual Network (DRRN). DRRN constructs a deep network structure with up to 52 convolutional layers, combining local and global residual learning in a multi-path mode and a multi-weight recursive learning mechanism. This approach effectively controls the number of parameters while improving model performance and enhancing network stability.



(a) DSRN

(b) Unfolding DSRN T times

Figure 6. (a) The recurrent representation of the proposed DSRN. (b) The unrolled DSRN.

Han et al. pointed out that many deep super-resolution network structures can be viewed as finite unfoldings of single-state RNNs with different recursive functions. Based on this perspective, they proposed the Dual-State Recurrent Network (DSRN). As shown in Figure 6, further expanding the application of recursive networks in super-resolution tasks. In 2019, Li et al. also utilized feedback mechanisms to propose the Super-Resolution Feedback Network (SRFBN). This network refines high-order information into low-order representations step by step and gradually generates the final high-resolution image, significantly improving network performance while reducing the number of parameters.

3.4. Lightweight Network-Based Methods

Lightweight networks aim to enhance computational efficiency while maintaining or improving network performance by designing compact network architectures or introducing lightweight strategies. These networks provide crucial support for the practical application of super-resolution (SR) algorithms. In 2018, Ahn et al. proposed an efficient Cascading Residual Network (CARN) and its lightweight version, CARN-M (CARN-Mobile), laying the foundation for lightweight SR algorithm design. In the same year, Hui et al. introduced lightweight strategies such as group convolution and information distillation blocks, proposing the Information Distillation Network (IDN). As shown in Figure 7. IDN consists of feature extraction blocks, information distillation blocks, and reconstruction blocks, significantly reducing network runtime. In 2019, Hui et al. further improved the information distillation blocks, designing Information Multi-Distillation Blocks and constructing the lightweight Information Multi-Distillation Network (IMDN). IMDN addresses the SR problem for arbitrary scale factors through an adaptive cropping strategy, further enhancing the network's practicality.

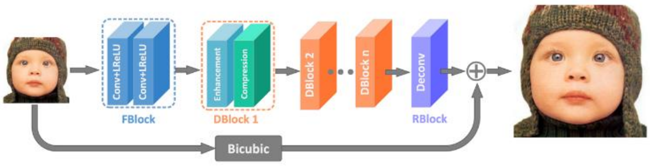


Figure 7. Architecture of the Deep Information Distillation Network.

In 2020, Liu et al. optimized IMDN by proposing Feature Distillation Connections, which operate similarly to channel splitting strategies, and constructed the Residual Feature Distillation Network (RFDN). RFDN achieves more lightweight and flexible SR reconstruction under the influence of information distillation mechanisms and won the championship in the AIM 2020 Efficient Super-Resolution Challenge. In 2021, Wang et al. proposed the Sparse Mask Super-Resolution Network (SMSR), which reduces redundant computations by studying image sparsity, significantly improving the network's inference efficiency.

4. Generative Adversarial Networks

The Generative Adversarial Network (GAN), introduced by Goodfellow et al. in 2014, presents a novel framework for evaluating generative models through an adversarial process. This framework comprises two core modules: the Generator (G) and the Discriminator (D). The Generator is responsible for learning the data distribution and generating new samples, while the Discriminator is tasked with distinguishing whether input data originates from real training data or the Generator's output. Through an adversarial game, the Generator and Discriminator are iteratively optimized until the Discriminator struggles to differentiate between generated and real data, at which point training is considered complete.

In 2017, Ledig et al. pioneered the application of GANs in the realm of image super-resolution, introducing the Super-Resolution Generative Adversarial Network (SRGAN) model. SRGAN consists of a generation network and an adversarial network: the generation network reconstructs high-resolution (HR) images from low-resolution (LR) images, while the discrimination network is responsible for determining whether the input image is a reconstructed HR image or an original HR image. Both are continuously optimized through iterative training until the discrimination network is unable to distinguish between reconstructed and real images. It is shown in Figure 8.

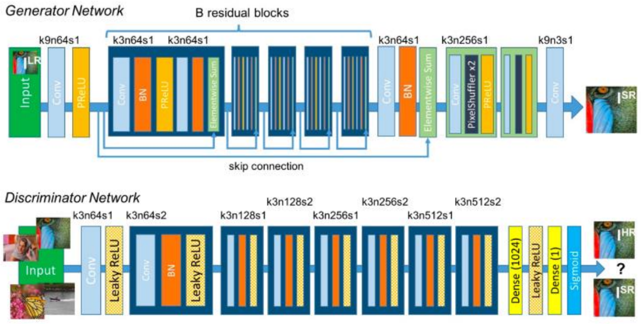


Figure 8. Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

To generate SR images with realistic textures and natural visual effects, Sajjadi et al. proposed EnhanceNet, which incorporates perceptual loss and automatic texture synthesis techniques. The EnhanceNet's generation network employs a feedforward fully convolutional neural network for texture synthesis. Through the combination of GAN and perceptual loss, EnhanceNet achieves the generation of realistic textures at high magnification, while the discrimination network follows a conventional design.

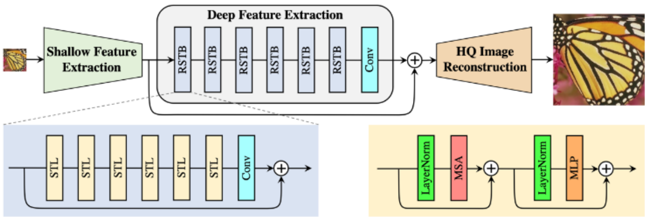
Inspired by SRGAN, Wang et al. proposed ESRGAN (Enhanced SRGAN) in 2018, which improved SRGAN in terms of network architecture and loss functions, further enhancing the quality of reconstructed images. Zhang et al. pointed out that many perceptual quality metrics, which are highly correlated with human subjective evaluations (no-reference objective evaluation metrics), are often non-differentiable and cannot be directly used as loss functions to optimize network models. To address this, they proposed the RankSRGAN model, which consists of a standard SRGAN and a Ranker. By introducing a ranking learning mechanism, RankSRGAN optimizes perceptual quality.

5. Transformer-Based Single Image Super-Resolution Reconstruction Methods

The Transformer, proposed by Vaswani et al. at Google in 2017, is a network architecture initially designed for Natural Language Processing (NLP) tasks. Originally developed to address sequence modeling problems in NLP, the Transformer has gradually been introduced into the field of computer vision with the successive proposals of models such as ViT (Vision Transformer), DETR (Detection Transformer), and ViViT (Video Vision Transformer). These models have demonstrated superior performance over traditional CNNs, Non-Maximum Suppression (NMS), and 3D convolutions in various tasks.

Transformer-based methods primarily leverage the self-attention mechanism to enhance the representation capability of image features. Benefiting from the Transformer's powerful modeling capabilities and flexible network structure, these methods have achieved superior reconstruction results and objective evaluation metrics in image super-resolution (SR) tasks compared to traditional CNNs. Currently, Transformer-based SR methods are mainly divided into two categories: one is a pure Transformer network model based entirely on the Transformer architecture; the other is a hybrid Transformer network model combining Transformer and CNN.

In 2021, Chen et al. first applied the Transformer architecture to the field of computer vision and proposed a pre-trained model IPT (Image Processing Transformer) for handling multiple low-level vision tasks such as super-resolution, denoising, and deraining. In the same year, Liang et al. proposed a network model SwinIR (Image Restoration Using Swin Transformer) for image restoration based on the structure of Swin Transformer. As shown in Figure 9.SwinIR belongs to the hybrid Transformer model, and its network structure consists of three modules: shallow feature extraction, deep feature extraction, and image reconstruction. This model combines the ability of CNN to process large-scale factor images and the advantages of Transformer in modeling long-distance dependencies, and performs well in image super-resolution, denoising, and JPEG compression artifact reduction tasks.



(a) Residual Swin Transformer Block (RSTB)

(b) Swin Transformer Layer (STL)

Figure 9. The architecture of the SwinIR for image restoration.

In addition, Lu et al. pointed out that Vision Transformer has problems such as high computational cost and large GPU memory occupation, which limits the depth design of the network. To address this, they proposed an efficient lightweight network ESRT (Efficient SR Transformer), which explores the feasibility of using Transformer for super-resolution tasks under resource-constrained conditions. As shown in Figure 10.

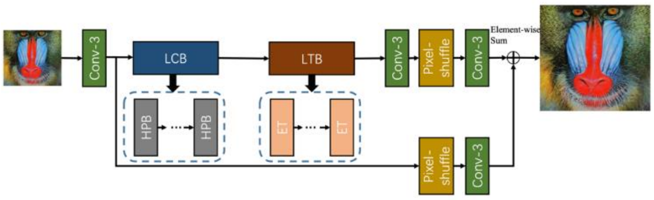


Figure 10. The architecture of the proposed Efficient Super-Resolution Transformer. LCB, LTB, HPB, and ET stand for the Lightweight CNN Backbone, the Lightweight Transformer Backbone, high preserving block, and efficient Transformers.

ESRT optimizes computational efficiency and memory usage, providing new ideas for the design of lightweight Transformer models. In 2023, Baghel et al. proposed the SRTransGAN (Image Super-Resolution using Transformer based Generative Adversarial Network) model. This method utilizes a Transformer-based generative adversarial network for image super-resolution reconstruction.

6. Conclusion

This paper systematically reviews the research progress and latest achievements in single-image super-resolution (SISR) reconstruction technology based on deep learning. With the continuous advancement of technology, SISR reconstruction has evolved from traditional methods aimed at maximizing Peak Signal-to-Noise Ratio (PSNR) to perception-driven deep learning models, and further to innovative breakthroughs centered on Transformer architectures. This evolutionary process has not only driven significant progress in image super-resolution reconstruction algorithms but also provided new research directions for the optimization and innovation of subsequent network architectures, significantly promoting technological development in this field. Currently, SISR technology has achieved remarkable results at the theoretical research level. Future research will gradually shift toward the application deployment of algorithms in practical scenarios and their transformation into industrial products.

Acknowledgment

This work was supported in part by the Faculty of Engineering Science and Technology, Infrastructure University Kuala Lumpur (IUKL). This research was also supported by the Education Department of Shaanxi Province, China, under Project No. 24JK0437, titled "Research on Super-Resolution Technology for Complex Industrial X-ray Images," and by the 2024 Qin Chuang Yuan Scientific Research Special Project of Shaanxi Province under Project No. 2024QCY-YJ17, titled "Research on Deep Learning-Based Super-Resolution Technology for Industrial X-ray Images." We thank Dr. Tadiwa Elisha Nyamasvisva, the doctoral advisor, for his valuable guidance and continuous support throughout this research.

References

[1] X. Yu et al., “Towards efficient and scale-robust ultra-high-definition image demoiréing,” in European Conference on Computer Vision, Springer, 2022, pp. 646–662.

[2] K. I. Kim and Y. Kwon, “Single-image super-resolution using sparse regression and natural image prior,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 6, pp. 1127–1133, 2010.

[3] W. Jenkins, B. Mather, and D. Munson, “Nearest neighbor and generalized inverse distance interpolation for Fourier domain image reconstruction,” in ICASSP’85. IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE, 1985, pp. 1069–1072.

[4] T. Blu, P. Thévenaz, and M. Unser, “Linear interpolation revitalized,” IEEE Trans. Image Process., vol. 13, no. 5, pp. 710–719, 2004.

[5] C. Lin, M. Sheu, H. Chiang, C. Liaw, and Z. Wu, “The efficient VLSI design of BI-CUBIC convolution interpolation for digital image processing,” in 2008 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2008, pp. 480–483.

[6] H. Prashanth, H. Shashidhara, and K. B. Murthy, “Image scaling comparison using universal image quality index,” in 2009 international conference on advances in computing, control, and telecommunication technologies, IEEE, 2009, pp. 859–863.

[7] K. T. Gribbon and D. G. Bailey, “A novel approach to real-time bilinear interpolation,” in Proceedings. DELTA 2004. Second IEEE international workshop on electronic design, test and applications, IEEE, 2004, pp. 126–131.

[8] R. Keys, “Cubic convolution interpolation for digital image processing,” IEEE Trans. Acoust. Speech Signal Process., vol. 29, no. 6, pp. 1153–1160, 1981.

[9] H. Xie, K. Xie, and H. Yang, “Research progress of image super-resolution methods,” Comput. Eng. Appl., vol. 56, no. 19, pp. 34–41, 2020.

[10] R. Y. Tsai and T. S. Huang, “Multiframe image restoration and registration,” Multiframe Image Restor. Regist., vol. 1, pp. 317–339, 1984.

[11] L. Ye and C. Zou, “A Survey of Image Super-Resolution Reconstruction Based on Deep Learning,” in INTERNATIONAL CONFERENCE ON WIRELESS COMMUNICATIONS, NETWORKING AND APPLICATIONS, Springer, 2022, pp. 584–594.

[12] S. Dai, M. Han, Y. Wu, and Y. Gong, “Bilateral back-projection for single image super resolution,” in 2007 IEEE International Conference on Multimedia and Expo, IEEE, 2007, pp. 1039–1042.

[13] H. Chang, D.-Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., IEEE, 2004, p. I–I.

[14] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[15] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, 2015.

[16] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14, Springer, 2016, pp. 391–407.

[17] W. Shi et al., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1874–1883.

[18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[19] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1646–1654.

[20] X.-J. Mao, “Image restoration using convolutional auto-encoders with symmetric skip connections,” ArXiv Prepr. ArXiv160608921, 2016.

[21] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 136–144.

[22] C. Ledig et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4681–4690.

[23] J. Li, F. Fang, K. Mei, and G. Zhang, “Multi-scale residual network for image super-resolution,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 517–532.

[24] R. Lan et al., “Cascading and enhanced residual networks for accurate single-image super-resolution,” IEEE Trans. Cybern., vol. 51, no. 1, pp. 115–125, 2020.

[25] J. Kim, J. K. Lee, and K. M. Lee, “Deeply-recursive convolutional network for image super-resolution,” in

Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1637–1645.

[26] Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3147–3155.

[27] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, and T. S. Huang, “Image Super-Resolution via Dual-State Recurrent Networks,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 1654–1663. doi: 10.1109/CVPR.2018.00178.

[28] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, “Feedback Network for Image Super-Resolution,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA: IEEE, Jun. 2019, pp. 3862–3871. doi: 10.1109/CVPR.2019.00399.

[29] N. Ahn, B. Kang, and K.-A. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 252–268.

[30] Z. Hui, X. Wang, and X. Gao, “Fast and Accurate Single Image Super-Resolution via Information Distillation Network,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 723–731. doi: 10.1109/CVPR.2018.00082.

[31] Z. Hui, X. Gao, Y. Yang, and X. Wang, “Lightweight Image Super-Resolution with Information Multi-distillation Network,” in Proceedings of the 27th ACM International Conference on Multimedia, Nice France: ACM, Oct. 2019, pp. 2024–2032. doi: 10.1145/3343031.3351084.

[32] J. Liu, J. Tang, and G. Wu, “Residual feature distillation network for lightweight image super-resolution,” in Computer vision–ECCV 2020 workshops: Glasgow, UK, August 23–28, 2020, proceedings, part III 16, Springer, 2020, pp. 41–55.

[33] L. Wang et al., “Exploring Sparsity in Image Super-Resolution for Efficient Inference,” in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA: IEEE, Jun. 2021, pp. 4915–4924. doi: 10.1109/CVPR46437.2021.00488.

[34] I. Goodfellow et al., “Generative adversarial nets,” Adv. Neural Inf. Process. Syst., vol. 27, 2014.

[35] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, “EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis,” in 2017 IEEE International Conference on Computer Vision (ICCV), Venice: IEEE, Oct. 2017, pp. 4501–4510. doi: 10.1109/ICCV.2017.481.

[36] X. Wang et al., “Esrgan: Enhanced super-resolution generative adversarial networks,” in Proceedings of the European conference on computer vision (ECCV) workshops, 2018, pp. 0–0.

[37] W. Zhang, Y. Liu, C. Dong, and Y. Qiao, “RankSRGAN: Generative Adversarial Networks With Ranker for Image Super-Resolution,” in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South): IEEE, Oct. 2019, pp. 3096–3105. doi: 10.1109/ICCV.2019.00319.

[38] A. Vaswani et al., “Attention is all you need,” Adv. Neural Inf. Process. Syst., vol. 30, 2017.

[39] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” ArXiv Prepr. ArXiv201011929, 2020.

[40] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-End Object Detection with Transformers,” May 28, 2020, arXiv: arXiv:2005.12872. Accessed: Jul. 26, 2023. [Online]. Available: http://arxiv.org/abs/2005.12872

[41] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lucic, and C. Schmid, “ViViT: A Video Vision Transformer,” in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada: IEEE, Oct. 2021, pp. 6816–6826. doi: 10.1109/ICCV48922.2021.00676.

[42] J. Bi, Z. Zhu, and Q. Meng, “Transformer in computer vision,” in 2021 IEEE International conference on computer science, electronic information engineering and intelligent control technology (CEI), IEEE, 2021, pp. 178–188.

[43] H. Chen et al., “Pre-Trained Image Processing Transformer,” in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA: IEEE, Jun. 2021, pp. 12294–12305. doi: 10.1109/CVPR46437.2021.01212.

[44] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, “Swinir: Image restoration using swin transformer,” in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 1833–1844.

[45] Z. Liu et al., “Swin transformer: Hierarchical vision transformer using shifted windows,” in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 10012–10022.

[46] Z. Lu, H. Liu, J. Li, and L. Zhang, “Efficient transformer for single image super-resolution,” ArXiv Prepr. ArXiv210811084, vol. 2, 2021.

[47] N. Baghel, S. R. Dubey, and S. K. Singh, “SRTransGAN: Image Super-Resolution using Transformer based Generative Adversarial Network,” ArXiv Prepr. ArXiv231201999, 2023.