A Review of Research on Super-resolution Image Reconstruction Based on Deep Learning

Ning An

Sydney Smart Technology College, Northeastern University at Qinhuangda, Liaoning Province, China

202219023@stu.neu.edu.cn

Abstract. Super-resolution image reconstruction (SRIR) endeavors to restore high-resolution (HR) images with enhanced detail from corresponding low-resolution (LR) inputs. With the rapid development of deep learning, integrating deep learning methods provides new solutions for the super-resolution (SR) field. This paper first reviews the background and significance of SR, development process, and the technical value of applying deep learning to SR. Next, SR methods based on deep learning are categorized according to different network types, with a focus on analyzing and comparing the applications of Convolutional Neural Networks (CNNs), Residual Networks (ResNet), Generative Adversarial Networks (GANs), and Diffusion Models in SR. The paper also introduces key evaluation metrics and problem-solving strategies, followed by a performance comparison of mainstream methods on publicly available datasets. Finally, a summary of SR algorithms based on deep learning is provided, along with an outlook on future development trends in the field and explore the possible next research directions in the field of image super-resolution.

Keywords: Image super-resolution, Deep learning, Convolutional neural networks, Generative adversarial networks.

1. Introduction

Super-resolution (SR) seeks to generate high-resolution (HR) images from low-resolution (LR) observations[1]. A challenge of SR lies in its ill-posedness, meaning that a single LR image can correspond to more than one valid HR reconstructions. SR, extensively applied across various domains, finds significant use in fields like medical imaging [2] and remote sensing imaging [3].

Amid the swift advancements in deep learning, this technology has been extensively integrated into a wide range of artificial intelligence tasks, including image classification and object detection, leading to remarkable breakthroughs. Simultaneously, numerous deep learning approaches have been employed to address super-resolution challenges. Both early methodologies utilizing Convolutional Neural Networks (CNNs) and subsequent techniques grounded in Generative Adversarial Networks (GANs) have demonstrated exceptional performance. More recently, diffusion-based super-resolution has shown great potential in enhancing the resolution of LR images. Recently, the problem of artifacts in diffusion-based super-resolution was confronted by Zheng et al. [4] through the introduction of innovative techniques known as Reality-Guided Refinement (RGR) and Self-Adaptive Guidance (SAG). Wang et al. [5] introduced a one-step bi-directional distillation approach, which is specifically designed to

elucidate the deterministic mapping between input noise and the resultant high-resolution image, along with the reverse process, by utilizing a teacher diffusion model in conjunction with their developed deterministic sampling method.

This article presents an extensive review of the latest advancements in deep learning-based super-resolution techniques. The SR methods based on deep learning are categorized according to the learning approaches and network types used. These categories include SR methods based on Convolutional Neural Networks (CNNs), Residual Networks (ResNet), Generative Adversarial Networks (GANs), and diffusion-based SR networks, each of which is introduced and discussed in detail.

2. Introduction to Deep Learning SR Method

2.1. CNN-based Super-Resolution Networks

2.1.1. Super-Resolution Convolutional Neural Network (SRCNN). The Super-Resolution Convolutional Neural Network (SRCNN) [6] is different from traditional interpolation and reconstruction algorithms. It employs a three-layer neural network to extract feature information from images, thereby enabling the learning of the mapping relationship between LR images and HR images in an end-to-end manner. Ultimately, the system, in a significant development, reconstructs the image while simultaneously simulating the traditional super-resolution process. As the pioneering deep learning-based image super-resolution model, SRCNN significantly enhanced the performance of image resolution reconstruction at the time, notwithstanding its relatively simple architecture.

2.1.2. Fast Super-Resolution Convolutional Neural Network (FSRCNN). Fast Super-Resolution Convolutional Neural Network (FSRCNN) [7] is an improved version of SRCNN, designed to enhance both the computational speed and performance of image super-resolution. Compared to SRCNN, FSRCNN incorporates deconvolution layers at the final part of the network, utilizes a deeper network architecture, and employs smaller convolutional kernels. This design enables the model to effectively acquire the mapping from the original LR images to HR images, significantly increasing the restoration speed while maintaining the quality of the recovery.

2.1.3. Efficient Sub-Pixel Convolutional Neural Network (ESPCN). Unlike SRCNN, Efficient Sub-Pixel Convolutional Neural Network (ESPCN) [8] does not compute directly at the HR level. At the heart of this network lies the incorporation of a sub-pixel convolutional layer structure, which is placed at the last layer of the network. This method regulates the quantity of feature maps through convolutional operations and integrates multiple feature maps by rearranging the pixels, achieving different upscaling effects, thereby improving both reconstruction efficiency and quality.

2.1.4. Residual Channel Attention Network (RCAN). The Residual Channel Attention Network (RCAN) [9] achieves significant advancements by introducing the Residual in Residual (RIR) structure, which constructs deep networks to enhance visual recognition performance. This approach allows the model to capture high-frequency information and combine it with low-frequency information to improve resolution. Additionally, RCAN also marks the inaugural introduction of the attention mechanism in LR image tasks, adjusting the allocation of resources to more information-rich areas through channel attention and improving the resolution improvement effect.

2.1.5. Holistic Attention Network (HAN). The Holistic Attention Network (HAN) enhances the generation of clearer and more detailed high-resolution images by introducing a holistic attention mechanism that captures both global and local information at multiple scales [10]. The Spatial Attention Mechanism focuses on the spatial relationships within the image, enhancing spatial features. The Channel Attention Mechanism, similar to RCAN, underscores the significance of features across various channels.

2.2. ResNet-based SR Networks

2.2.1. Very Deep Convolutional Networks for Super-Resolution (VDSR). Very Deep Convolutional Networks for Super-Resolution (VDSR) utilizes an exceptionally deep convolutional network [11]. The increase in network depth leads to a significant enhancement in image clarity. VDSR is characterized by the use of 20 weight layers, which effectively capitalize on contextual information across extensive image areas. However, in extremely deep networks, the speed of convergence emerges as a critical issue during the training process.

2.2.2. Enhanced Deep Residual Networks (EDSR). Unlike typical generative network components, Enhanced Deep Residual Networks (EDSR) removes unnecessary batch normalization operations [12]. This allows EDSR to stack more network layers without more resources. Consequently, each layer of the network can extract more features to enlarge the model size and improve reconstruction quality. Additionally, this method integrates residual scaling by incorporating a constant scaling layer following the convolutional layers in each residual block, which serves to stabilize the training of large models.

2.2.3. Multi-Scale Residual Network (MSRN). The Multi-Scale Residual Network (MSRN) advances the residual block by incorporating convolutional kernels of diverse scales. This enhancement facilitates the adaptive detection of image features across varying scales [13]. The Multi-Scale Residual Blocks employs multi-scale feature fusion to capture image feature information at various scales, thereby significantly reducing computational complexity. The Hierarchical Feature Fusion Structure filters and fuses all feature outputs from above blocks, minimizing unnecessary redundant information while adaptively highlighting useful information.

2.2.4. Deep Back-Projection Networks (DBPN). Deep Back-Projection Networks (DBPN) progressively enhances image resolution through a multi-level forward and backward projection mechanism, thereby better recovering image details [14]. The core of DBPN involves using forward projection to map low-resolution features into high-resolution space, followed by backward projection to remap them back into low-resolution space. This iterative process helps reduce error accumulation and improves reconstruction accuracy. The multiple bidirectional projection operations enable DBPN to excel in detail recovery and edge reconstruction, particularly demonstrating outstanding performance in high-magnification tasks.

2.2.5. Cascading Residual Network (CARN). Cascading Residual Network (CARN) applies a progressive learning approach to cascading residual networks [15]. In the training method for extreme super-resolution scenarios, the model initially generates relatively low-resolution outputs and then incrementally increases the output resolution by adding additional networks to the model. This progressive upsampling of images helps to reduce abrupt size changes in the model, thereby alleviating training instability.

2.3. GAN-based SR Networks

2.3.1. Super-Resolution Generative Adversarial Network (SRGAN). Super-Resolution Generative Adversarial Network (SRGAN) introduces GANs into the image SR field, proposing a network model for image SR reconstruction that utilizes generative adversarial networks [16]. Through adversarial training between the generator and discriminator, the model generates high-resolution images from low-resolution inputs. The generator aims to create visually realistic high-resolution images, while the discriminator evaluates the authenticity of these images.

2.3.2. Enhanced Super-Resolution Generative Adversarial Network (ESRGAN). Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) [17] builds upon the foundation of SRGAN

with several improvements. The paper introduces a network unit called the Residual-in-Residual Dense Block (RRDB), which eliminates the Batch Normalization layers. Additionally, the authors enhance the discriminator's objective in the GAN framework and utilize an activation-based feature composition for the perceptual loss function. These modifications contribute significantly to improving the visual quality of the output images.

2.3.3. Blind Super-Resolution Generative Adversarial Network (BSRGAN). Blind Super-Resolution Generative Adversarial Network (BSRGAN) employs random permutations of various degradation factors, including blurring, downsampling, and noise [18]. This approach enables the trained BSRGAN to exhibit significant performance improvements when handling real-world degraded images.

2.4. Diffusion-based SR Networks

2.4.1. Super-Resolution Diffusion Probabilistic Model (SRDiff). While learning-based SISR methods significantly outperform traditional techniques, approaches focusing on PSNR, GAN-driven methods, and flow-based methods each face issues such as over-smoothing or excessive model size. To address these challenges, the Super-Resolution Diffusion Probabilistic Model (SRDiff) is proposed [19]. This model is the first single-image super-resolution model based on diffusion processes, gradually transforming Gaussian noise into super-resolution images conditioned on low-resolution inputs through a Markov chain, while incorporating residual prediction to accelerate convergence.

2.4.2. Frequency Domain-guided Multiscale Diffusion Model (FDDiff). The Frequency Domain-guided Multiscale Diffusion Model (FDDiff) [20] decomposes the high-frequency information complementary process into more granular steps. FDDiff directs the backward diffusion process to supplement the missing high-frequency details over time steps. Furthermore, a multiscale frequency refinement network is designed, which is capable of processing information across multiple scales, allowing for more precise identification and prediction of high-frequency details in signals or images.

2.4.3. Implicit Diffusion Models for Continuous Super-Resolution (IDM). Implicit Diffusion Models for Continuous Super-Resolution (IDM) [21] integrates implicit neural representations with denoising diffusion models into a unified end-to-end framework. During the decoding process, implicit neural representations are utilized to learn continuous resolution representations. A scalable adjustment mechanism is designed, which includes a LR adjustment network and a scaling factor.

3. Discussion and Analysis

3.1. Evaluation Metrics
Evaluation metrics typically employ established computational formulas to assess the errors between SR images and HR images, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) . As the ratio of the maximum power of a signal to the power of noise in the signal, PSNR measures the quality of the compressed reconstructed image of the input. SSIM is a method which can quantify the similarity between two images. Higher values of PSNR and SSIM indicate better quality of the reconstructed images, reflecting superior performance of the SR methods.

3.2. Datasets
The datasets used include Set5[22], Set14[23], B100[24], Urban100[25], and Manga109[26]. These datasets contain varying numbers of high-resolution images, ranging from 5 to 2,650, covering a wide range of content such as environments, flora and fauna, racing cars, natural landscapes, and people. They effectively and comprehensively test the performance of algorithms. Additionally, other related algorithms have been designed to improve the practical application value of SR in more realistic life scenarios by utilizing image datasets that better reflect real-world conditions for experimentation.

3.3. Analysis of Experimental Results

This study conducts a comparative analysis of the experimental results from traditional methods and deep learning approaches based on different network types, as reported in relevant literature. The public datasets used for the experiments include Set5[22], Set14[23], and BSD100[24]. The performance of several mainstream deep learning-based super-resolution methods is compared, including Bicubic interpolation [27], SRCNN[6], FSRCNN[7], VDSR[11], EDSR[12], RCAN[9], MSRN[13], HAN[10], and DBPN[14], with Bicubic serving as the benchmark method for comparison. Since the Set5, Set14, and BSD100 datasets only provide HR images, simulated LR images need to be generated from the high-resolution images beforehand. In this study, Bicubic interpolation is used to downsample each high-resolution image by factors of 2, 3, and 4, resulting in the corresponding low-resolution images. Subsequently, experiments on super-resolution using various deep learning-based models are conducted on the LR images for 2x, 3x, and 4x upscaling, with the results presented in Tables 1-3.

Table 1. Quantitative comparison with SR methods at ×2 scale

Method	Set5		Set14		BSD100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic[27]	33.66	0.929 9	30.24	0.868 8	29.56	0.843 1
SRCNN[6]	36.66	0.954 2	32.45	0.906 7	31.36	0.887 9
FSRCNN[7]	37.05	0.956 0	32.66	0.909 0	31.53	0.892 0
VDSR[11]	37.53	0.958 7	33.05	0.912 7	31.90	0.896 0
EDSR[12]	38.11	0.960 2	33.92	0.919 5	32.32	0.901 3
RCAN[9]	38.27	0.961 4	34.12	0.921 6	32.41	0.902 7
MSRN[13]	38.08	0.960 5	33.74	0.917 0	32.23	0.900 2
HAN[10]	38.27	0.961 4	34.16	0.921 7	32.41	0.902 7
DBPN[14]	38.09	0.960	33.85	0.919	32.27	0.900

Table 2. Quantitative comparison with SR methods at ×3 scale

Method	Set5		Set14		BSD100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic[27]	30.39	0.868 2	27.55	0.774 2	27.21	0.738 5
SRCNN[6]	32.75	0.909 0	29.30	0.821 5	28.41	0.786 3
FSRCNN[7]	33.18	0.914 0	29.37	0.824 0	28.53	0.791 0
VDSR[11]	33.66	0.921 3	29.78	0.831 8	28.83	0.797 6
EDSR[12]	34.65	0.928 0	30.52	0.846 2	29.25	0.809 3
MSRN[13]	34.38	0.926 2	30.34	0.839 5	29.08	0.804 1
HAN[10]	34.75	0.929 9	30.67	0.848 3	29.32	0.811 0

Table 3. Quantitative comparison with SR methods at ×4 scale

Method	Set5		Set14		BSD100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic[27]	28.42	0.810 4	26.00	0.702 7	25.96	0.667 5
SRCNN[6]	30.48	0.862 8	27.50	0.751 3	26.90	0.710 1
FSRCNN[7]	30.72	0.866 0	27.61	0.755 0	26.98	0.715 0
VDSR[11]	31.35	0.883 8	28.02	0.767 8	27.29	0.725 2
SRGAN[16]	32.05	0.891 0	28.53	0.780 4	27.57	0.735 4
EDSR[12]	32.46	0.896 8	28.80	0.787 6	27.71	0.742 0
MSRN[13]	32.07	0.890 3	28.60	0.775 1	27.52	0.727 3
HAN[10]	32.64	0.900 2	28.90	0.789 0	27.80	0.744 2
DBPN[14]	32.47	0.898	28.82	0.786	27.72	0.740

Deep learning-based methods consistently outperform the traditional Bicubic method across all metrics at ×2, ×3 and ×4 scales, and this advantage becomes more pronounced as the scaling factor increases. While SRCNN, as one of the earliest deep learning SR methods, demonstrated excellent performance at its inception, its efficacy has gradually been surpassed by newer methods in recent years.

Across all scales and datasets, the HAN method performs well in terms of PSNR, and in most cases, its SSIM values also rank among the top. This indicates that HAN possesses strong capabilities in image super-resolution tasks. Simpler models like SRCNN and FSRCNN, while advantageous in computational efficiency, perform worse than more complex deep learning models when handling high-resolution images and intricate scenes. At higher scaling factors, the MDSR method excels in SSIM, particularly on the Set5 dataset. The introduction of attention mechanisms in both the RCAN and HAN methods generally results in outstanding performance across all datasets and scaling factors, suggesting that attention mechanisms can effectively enhance model performance.

As the scale increases, the difficulty of image reconstruction rises. Taking a scaling factor of 4 as an example, the PSNR and SSIM values of nearly all algorithms show a significant decline. This indicates that the scaling factor has a substantial impact on the effectiveness of super-resolution reconstruction. The PSNR and SSIM values for SRCNN and FSRCNN decrease relatively gradually with increasing scaling factors, whereas more complex models like VDSR, EDSR, and RCAN are better at preserving image quality, particularly performing exceptionally well at factors of 3 and 4.

4. Conclusion

As a fundamental task in computer vision, image SR reconstruction plays a crucial role in practical applications such as criminal investigation, medical diagnosis, and smart cities. This paper reviewed recent deep learning-based image SR methods and compared the performance of mainstream approaches based on two metrics: PSNR and SSIM. Although SR technology has achieved substantial research progress, several challenges remain in the following areas:

(1) Model Lightweight Design. As the depth and complexity of models increase, so does their computational load. Designing lightweight network architectures that maintain high-quality high-resolution images while reducing computational demands is an important direction in current research.

(2) Generalization Ability. Most current SR models perform exceptionally well in specific scenarios but struggle to adapt to diverse and complex environments. Therefore, further research is needed on how to maintain strong generalization capabilities across different contexts.

References

[1]   Tsai, R. Y., & Huang, T. S. (1984). Multiframe image restoration and registration. Multiframe image restoration and registration, 1, 317-339.

[2]   Isaac, J. S., & Kulkarni, R. (2015, February). Super resolution techniques for medical image processing. In 2015 International Conference on Technologies for Sustainable Development (ICTSD) (pp. 1-6). IEEE.

[3]   Thornton, M. W., Atkinson, P. M., & Holland, D. A. (2006). Sub‑pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super‑resolution pixel‑swapping. International Journal of Remote Sensing, 27(3), 473-491.

[4]   Zheng, Q., Zheng, L., Guo, Y., Li, Y., Xu, S., Deng, J., & Xu, H. (2024). Self-Adaptive Reality-Guided Diffusion for Artifact-Free Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 25806-25816).

[5]   Wang, Y., Yang, W., Chen, X., Wang, Y., Guo, L., Chau, L. P., ... & Wen, B. (2024). SinSR: diffusion-based image super-resolution in a single step. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 25796-25805).

[6]   Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.

[7]   Dong, C., Loy, C. C., & Tang, X. (2016). 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 (pp. 391-407).

[8]   Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). 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 (pp. 1874-1883).

[9]   Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV) (pp. 286-301).

[10]   Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., ... & Shen, H. (2020). Single image super-resolution via a holistic attention network. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16 (pp. 191-207). Springer International Publishing.

[11]   Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).

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

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

[14]   Haris, M., Shakhnarovich, G., & Ukita, N. (2018). Deep back-projection networks for super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1664-1673).

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

[16]   Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

[17]   Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops (pp. 0-0).

[18]   Zhang, K., Liang, J., Van Gool, L., & Timofte, R. (2021). Designing a practical degradation model for deep blind image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4791-4800).

[19]   Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., ... & Chen, Y. (2022). Srdiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing, 479, 47-59.

[20]   Wang, X., Chai, L., & Chen, J. (2024). Frequency-Domain Refinement with Multiscale Diffusion for Super Resolution. arXiv preprint arXiv:2405.10014.

[21]   Gao, S., Liu, X., Zeng, B., Xu, S., Li, Y., Luo, X., ... & Zhang, B. (2023). Implicit diffusion models for continuous super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10021-10030).

[22]   Bevilacqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding.

[23]   Zeyde, R., Elad, M., & Protter, M. (2012). On single image scale-up using sparse-representations. In Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7 (pp. 711-730).

[24] Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001, July). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings eighth IEEE international conference on computer vision. ICCV 2001 (Vol. 2, pp. 416-423). IEEE.

[25] Huang, J. B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5197-5206).

[26] Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T., & Aizawa, K. (2016, December). Manga109 dataset and creation of metadata. In Proceedings of the 1st international workshop on comics analysis, processing and understanding (pp. 1-5).

[27] Gribbon, K. T., & Bailey, D. G. (2004, January). A novel approach to real-time bilinear interpolation. In Proceedings. DELTA 2004. Second IEEE international workshop on electronic design, test and applications (pp. 126-131). IEEE.