Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33298

Emerging Deep Learning Strategies for Medical Image Super-Resolution: Trends, Challenges, and Prospects

Minkai Ou

School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China

ouminkai39@gmail.com

Abstract. In the field of computer vision, image super-resolution (SR) plays a crucial role as a class of image processing techniques that boost image resolution, with significant research importance and practical applications in medical imaging. Recently, methods based on deep learning have significantly enhanced the super-resolution processing capabilities of medical images, and high-resolution (HR) medical images have become a necessary condition for accurate clinical diagnosis. This paper aims to elaborate on the innovations and progress of deep learning-driven medical image super-resolution technologies. Besides, it reviews the basic theory of SR, outlines the standard metrics for evaluating the performance of super-resolution, discusses the key challenges in medical image super-resolution reconstruction, and points out the possible future development directions. The results show that the medical image super-resolution technology based on deep learning has evolved from convolutional neural networks (CNN) to more advanced architectures such as Transformer, Mamba and Implicit Neural Representation (INR), but due to limited data and insufficient evaluation metrics, challenges remain. Future work should explore unsupervised learning, lightweight networks, and multi-modal fusion for clinical application.

Keywords: Deep Learning, Super-resolution, Neural Network, Medical Images

1. Introduction

With the relentless evolution of technology, super-resolution is defined as a technique that generates high-resolution (HR) images from their corresponding low-resolution (LR) counterparts. In medical applications, it recovers lost high-frequency details, preserves anatomical consistency, and provides clearer visual information for clinical tasks like diagnosis, treatment planning, and image analysis. Constrained by multiple factors such as the cost of imaging equipment, scanning duration, radiation dose. As a result, they frequently suffer from insufficient spatial resolution, which hinders the clear visualization of tiny lesions and fine tissue structures, directly affecting diagnostic accuracy and the planning of subsequent treatment regimens [1]. Early medical image super-resolution methods were predominantly based on traditional optimization paradigms like interpolation, sparse representation and probabilistic modeling. But these approaches often fail to restore fine textures, balance noise and edges, or overcome the information limits of original images, sometimes producing artifacts [2]. To address these issues, learning-based SR acquires prior knowledge iteratively, generating high-

DOI: 10.54254/2755-2721/2026.GU33298

quality outputs from few LR images, though magnification remains limited [3]. Thus, deep learning, initiated by SRCNN, has further advanced SR research, improving chest CT resolution and boosting MRI quality through models such as GANs [4-6]. As a result, these techniques aid clinicians by enhancing image clarity; yet challenges persist, including preserving structures, protecting sensitive data, and handling limited paired HR-LR datasets. This paper explores the latest advancements in medical image super-resolution technology based on deep learning. Additionally, it focuses on the use and evaluation metrics of current methods, reviews the most advanced network architectures, and outlines the main challenges in this field as well as potential future development directions.

2. Basic theories of image super-resolution

2.1. The super-resolution problem

The SR task aims to recover a HR image from its LR observation by learning the mapping between the two domains. In a typical formulation, the LR image is assumed to be generated from the HR image through a degradation process:

$$I_{LR}=D\left(I_{HR},\delta\right) \tag{1}$$

where D denotes the degradation operation, I_{HR} is the high-resolution input, and δ represents the associated parameters. Since this process generally leads to irreversible information loss, directly computing its inverse is infeasible.

To address this issue, SR methods usually employ a parameterized model to approximate the mapping from LR to HR images. Specifically, a function $f(\bullet)$ is learned to produce an HR estimate from the LR input

$$f\left(I_{LR},\vartheta\right)=D^{-1}\left(I_{HR}\right)=\overline{I_{HR}}\approx I_{HR} \tag{2}$$

where ϑ denotes the model parameters, and $\overline{I_{HR}}$ is the reconstructed image. This mapping is learned in a data-driven manner to ensure both reconstruction fidelity and generalization ability. During training, an objective function is defined to constrain the model output so that it approximates the ground-truth HR image while maintaining model regularity:

$$\hat{\theta}=\underset{\theta}{\operatorname{argmin}} L\left(\overline{I_{HR}}, I_{HR}\right)+\lambda\Phi(\theta) \tag{3}$$

where $L(\bullet)$ measures the reconstruction error, $\lambda\Phi(\theta)$ is the regularization term, and λ balances the trade-off between data fidelity and model complexity.

2.2. Image quality assessment

Image quality refers to the attributes of an image that are significant in terms of visual perception, mainly focusing on human perception. In order to conduct a quantitative assessment of image quality, image quality assessment (IQA) is employed. IQA methods are generally categorized into subjective approaches, which depend on human observers, and objective approaches, which rely on computational models. Although the subjective methods closely align with human visual perception,

DOI: 10.54254/2755-2721/2026.GU33298

they are typically time-consuming, costly, and not suitable for large-scale use. Thus, objective methods have become the predominant approach in both research and practical applications [7].

2.2.1. Peak signal-to-noise ratio for image reconstruction

Peak Signal-to-Noise Ratio (PSNR) is an error-based objective metric widely employed to evaluate the fidelity of reconstructed images in tasks such as restoration, compression, and inpainting. It quantifies the discrepancy between a reconstructed image and its ground-truth reference by relying on the Mean Squared Error (MSE), which measures pixel-wise deviations. The MSE is defined as:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (I(i) - \hat{I}(i))^2, \tag{4}$$

Based on this error measure, PSNR is computed as the logarithmic ratio between the maximum possible signal power and the reconstruction error:

$$PSNR = 10 \cdot lg(\frac{L^2}{MSE}). \tag{5}$$

where L represents the maximum possible pixel intensity value, typically 255 for 8-bit images.

PSNR is favored in practice due to its simplicity and low computational cost. However, it evaluates image quality purely at the pixel level, assigning equal importance to all spatial locations. As a result, it is often insensitive to structural distortions and high-level perceptual differences, leading to potential inconsistencies with human visual judgment in some cases.

2.2.2. Structural similarity for structural fidelity

Structural Similarity Index (SSIM) is a perceptually motivated metric designed to evaluate image quality by comparing structural information between a reconstructed image and its reference [8]. Unlike purely error-based measures, SSIM considers image statistics related to luminance, contrast, and structural correlation, which better aligns with human visual perception. For an image I with N pixels, the brightness and contrast are defined as the average value and standard deviation of the pixel intensities:

$$\mu_I = \frac{1}{N} \sum_{i=1}^{N} I\left(i\right), \tag{6}$$

$$\sigma_I = \left(\frac{1}{N-1} \sum_{i=1}^{N} \left(I\left(i\right) - \mu_I\right)^2\right)^{\frac{1}{2}}, \tag{7}$$

where $I(i)$ represents the intensity of the i -th pixel of image I .

To compare two images, SSIM decomposes similarity into luminance, contrast, and structure components. The luminance comparison is defined as:

$$C_l\left(I, \hat{I}\right) = \frac{2\mu_I\mu_{\hat{I}} + C_1}{\mu_I^2 + \mu_{\hat{I}}^2 + C_1}, \tag{8}$$

$$C_c\left(I, \hat{I}\right) = \frac{2\sigma_I\sigma_{\hat{I}} + C_2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2}, \tag{9}$$

DOI: 10.54254/2755-2721/2026.GU33298

where $C_1=(k_1 l)^2$ and $C_2=(k_2 l)^2$ are small constants introduced to ensure numerical stability, with $k_1 \ll 1$ and $k_2 \ll 1$, and l representing the maximum pixel intensity of the image.

Besides, image structure is represented by normalized pixel intensities, $(I - \mu_I)/\sigma_I$), and structural similarity is quantified through the correlation between the normalized reference and reconstructed images, which corresponds to the correlation coefficient. The structural comparison function $C_S(I,\hat{I})$ is thus given by:

$$\sigma_{I\hat{I}} = \frac{1}{N-1} \sum_{i=1}^{N} \left(I\left(i\right) - \mu_I\right)\left(\hat{I}\left(i\right) - \mu_{\hat{I}}\right), \tag{10}$$

$$C_S\left(I, \hat{I}\right) = \frac{\sigma_{I\hat{I}} + C_3}{\sigma_I \sigma_{\hat{I}} + C_3}, \tag{11}$$

where $\sigma_{I\hat{I}}$ is the covariance between the ground-truth image I and the reconstructed image $\hat{I}$, and C_3 is a constant for stability. Finally, the SSIM index is given by the following formula:

$$\mathrm{SSIM}\left(I, \hat{I}\right) = \frac{(2\mu_I \mu_{\hat{I}} + C_1)(\sigma_{I\hat{I}} + C_2)}{\left(\mu_I^2 + \mu_{\hat{I}}^2 + C_1\right)\left(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2\right)}. \tag{12}$$

Based on human perception of structural information, SSIM overcomes some shortcomings of PSNR. However, SSIM cannot work effectively when images suffer from non-structural distortions like displacement, scaling and rotation. Moreover, SSIM is unstable and may produce inconsistent evaluation results when the reference image is a medical image with low variance or brightness [9].

2.2.3. Mean opinion score for subjective evaluation

The mean opinion score (MOS) serves as a widely adopted subjective metric for evaluating image quality. Participants provide perceptual ratings for images, generally ranging from 1 (poor) to 5 (excellent), and the MOS is obtained by averaging the scores from all evaluators.

Though MOS evaluates image quality according to human perception, it has inherent limitations: parameters cannot be adjusted dynamically, the scale is nonlinear, and results are influenced by viewing distance, display devices, observer vision, and subjective bias. But it remains a reliable IQA method when the number of evaluators is sufficient and the evaluation process is standardized. In practice, some SR models demonstrate low scores on standard objective metrics like PSNR and SSIM, while achieving superior visual perception.

2.2.4. Perceptual image quality evaluation using IFC and VIF

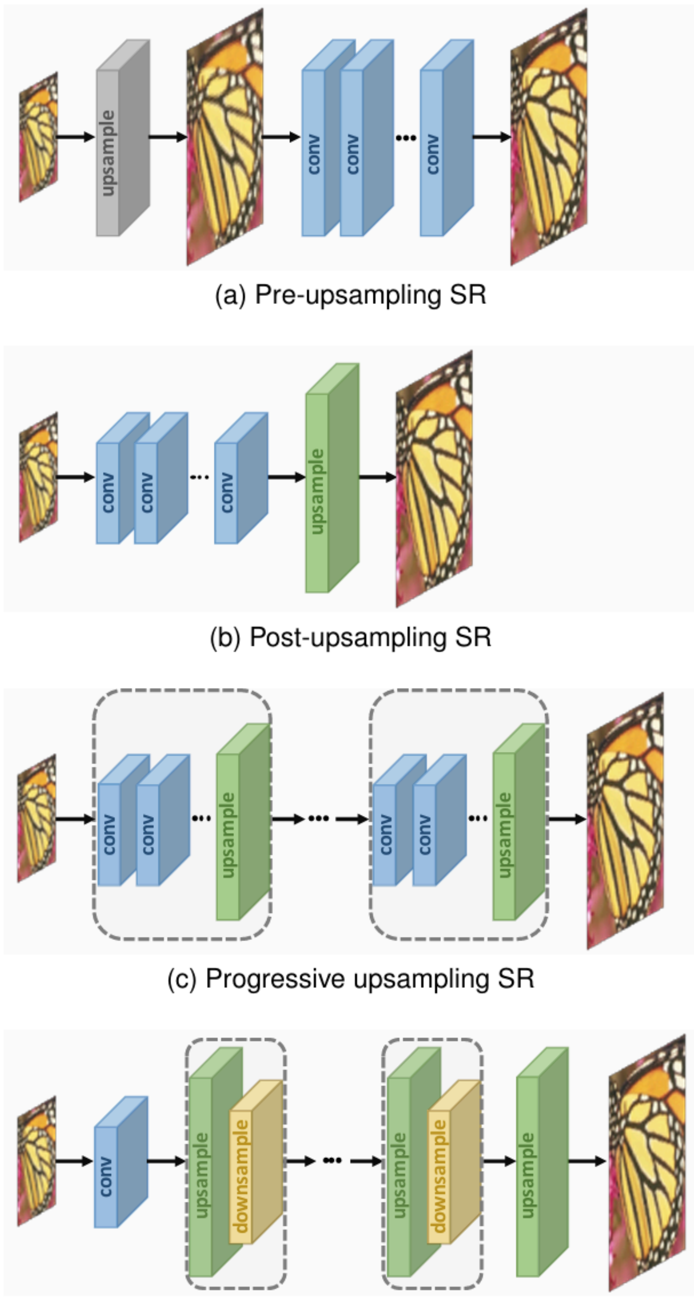
In addition to standard numerical metrics, perceptual measures provide a more human-centered evaluation of super-resolution performance. The Information Fidelity Criterion (IFC) and Visual Information Fidelity (VIF), proposed by Sheikh et al., conceptualize the Human Visual System as a communication channel. These metrics assess the perceptual quality of reconstructed images by quantifying the mutual information shared with the reference image [10,11]. Unlike traditional measures such as Peak Signal-to-Noise Ratio (PSNR) or Structural Similarity Index (SSIM), which focus on pixel-level or structural differences, IFC and VIF better reflect how images are perceived by human observers. However, PSNR and SSIM remain widely used in super-resolution research due to their simplicity and low computational requirements. By combining both conventional and

DOI: 10.54254/2755-2721/2026.GU33298

perceptual metrics, researchers can achieve a more comprehensive evaluation that accounts for both objective fidelity and human visual perception.

2.3. Deep learning frameworks for super-resolution

Image super-resolution constitutes an ill-posed inverse problem, and the upsampling process that generates HR outputs from LR inputs represents the core challenge. According to the upsampling module's placement, SR network frameworks are categorized into four types, as shown in Figure 1.



(a) Pre-upsampling SR

(b) Post-upsampling SR

(c) Progressive upsampling SR

(d) Iterative up-and-down Sampling SR

Figure 1. Deep learning frameworks for image super-resolution

DOI: 10.54254/2755-2721/2026.GU33298

2.3.1. Pre-upsampling SR

The pre-upsampling SR framework first enlarges the LR image to the desired size using traditional interpolation, then reconstructs high-quality details through a neural network, as in SRCNN. And its advantages include low learning difficulty and accepting any image size after interpolation. Its main drawbacks are that classic interpolation methods, like bicubic and cubic spline, can amplify noise and blur images [12,13]. In addition, the model operates in a high-dimensional space, which greatly raises computational complexity as well as memory and time requirements.

2.3.2. Post-upsampling SR

Post-upsampling means that the LR image is first fed into the neural network for feature extraction in a low-dimensional space, and upsampling is implemented using a learnable layer in the last layer of the network. This framework computes in a low-dimensional space, reducing computational cost and model complexity, which makes it widely employed in SR research. However, post-upsampling struggles with multi-scale SR and becomes harder to train for large upsampling factors.

2.3.3. Progressive-upsampling SR

Progressive-upsampling divides the SR model into multiple stages, upsampling the image once per stage to gradually reach the target magnification. Its main advantage is that progressive upsampling reduces learning difficulty and enhances reconstruction. Its drawbacks are that it still cannot handle multi-scale SR, and the multi-stage design increases network complexity and training difficulty. ProSR, proposed by Wang et al., is a typical progressive-upsampling model [14].

2.3.4. Iterative up-and-down sampling SR

Iterative up-and-down sampling employs back-projection to reduce the difference between LR and HR images. It alternates between upsampling and downsampling, refines the reconstruction using back-projection, and adjusts the HR image according to the reconstruction error. PBPN utilizes this strategy, generating the final HR image by progressively combining intermediate HR outputs [15].

3. Application of super-resolution networks in medical images

3.1. Convolutional neural network

Convolutional Neural Networks (CNNs) are early deep learning models widely used in image super-resolution. SRCNN is a representative method that formulates SR as a three-stage pipeline: low-resolution feature extraction, nonlinear feature mapping, and final image reconstruction [2]. This design demonstrates that SR can be learned in an end-to-end manner using CNNs. To further improve reconstruction quality, deeper architectures were later proposed. VDSR (Very Deep Super-Resolution) significantly increases network depth by stacking many convolutional layers with small kernels and introduces residual learning to predict the difference between low-resolution inputs and high-resolution targets [16]. This reformulation simplifies the learning objective and improves optimization efficiency. In addition, VDSR uses skip connections to enhance gradient propagation in deep networks and uses gradient clipping to stabilize training, allowing higher learning rates and faster convergence despite increased depth [16].

DOI: 10.54254/2755-2721/2026.GU33298

3.2. Generative adversarial network

SRGAN, proposed by Ledig et al., introduces GANs into image super-resolution for the first time [17]. It consists of a generator for HR image reconstruction and a discriminator for distinguishing real and synthesized images, trained via adversarial learning with perceptual loss in a minimax framework. In contrast to CNN-based approaches that focus on minimizing pixel-level metrics such as PSNR and SSIM, SRGAN emphasizes perceptual fidelity, generating sharper textures and more visually realistic outputs [17].

Gu et al. developed MedSRGAN, a deep learning framework for super-resolving medical images [18]. Its generator, the Residual Whole Map Attention Network (RWMAN), captures multi-channel features and focuses on clinically relevant regions. A multi-task loss combining adversarial, content, and feature losses guides the generation of accurate and clinically viable high-resolution images. In addition, the MedSRGAN model designs a unique pair discriminator. And this discriminator takes image pairs (LR images and HR/SR images) as its input, in order to learn the pairing information between HR/SR images and LR images, distinguishing HR/SR images from LR images.

3.3. Transformer

Built entirely upon self-attention, the Transformer is a sequence-to-sequence model that replaces recurrent layers in traditional encoder-decoder systems with multi-head self-attention [19]. Deep learning-based SR is an effective approach to enhance image resolution, with most SR models using CNNs as the backbone. CNNs excel in visual tasks, extracting local features via convolution, offering translation invariance, and employing pooling to reduce feature dimensions and prevent overfitting. However, such methods are constrained by the inherent properties of convolution: using the same kernel across different regions may overlook content correlations, and the local nature of convolution hinders modeling of long-range dependencies.

Compared with CNNs, the Transformer overcomes limited receptive fields and difficulty in capturing global context through self-attention, providing a new paradigm for medical image super-resolution. In this context, the SwinIR structure combines the strengths of Transformer and CNN, as shown in Figure 2 [20]. SwinIR, built upon the Swin Transformer, uses a shifted window attention strategy that supports efficient feature representation at full resolution while modeling long-range dependencies [21]. The window shift allows cross-window information exchange without global attention. Its architecture is composed of three main stages: shallow feature extraction, deep feature extraction, and reconstruction. And the shallow module uses convolution to retain low-frequency components, while the deep module is built upon Residual Swin Transformer Blocks (RSTBs), which capture local dependencies and enable cross-feature interaction [20]. Residual connections and feature fusion between shallow and deep representations are then used to generate the final super-resolved output.

DOI: 10.54254/2755-2721/2026.GU33298

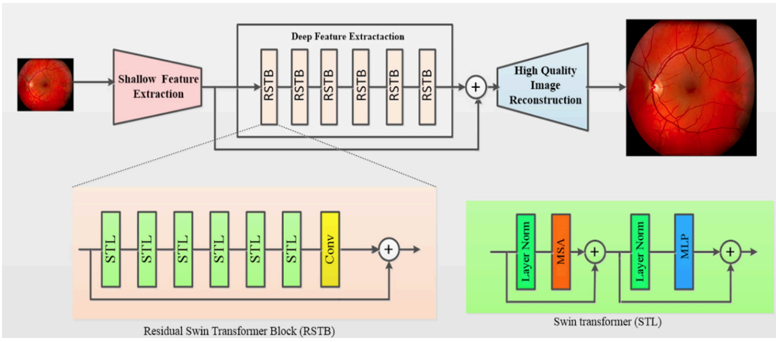


Figure 2. SwinIR transformer architecture for SR image

3.4. Mamba

The Mamba architecture combines the Selective State Space Model (SSM) and gated MLP to efficiently process long sequences, reducing computation and improving training and inference stability [22]. Each Mamba block integrates an SSM layer, convolution, and gated MLP, connected via layer normalization and residual connections, as shown in Figure 3. The selective scan operation allows selective retention of key information, cross-window feature interaction, and state resetting, enhancing long-sequence modeling and multi-sequence processing. Stacking multiple blocks forms a deep network capable of filtering unimportant information while focusing on relevant context.

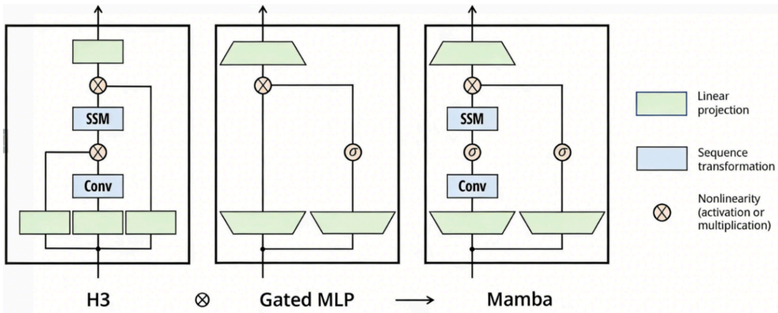


Figure 3. Mamba structure

Originally developed for NLP, Mamba has been adapted for medical imaging in SMamba-UNet [23]. As shown in Figure 4, the ISS2D module expands 2D image features into 1D sequences in four directions, dynamically adjusts the scan range, and enhances both local and global perceptual capabilities.

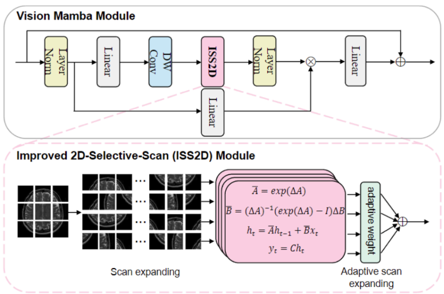


Figure 4. Visual Mamba module framework and improved two-dimensional selective scanning (ISS2D) module

DOI: 10.54254/2755-2721/2026.GU33298

3.5. Implicit neural representation

Implicit neural representation (INR), another advanced SR method, overcomes fixed-magnification limits of traditional upsampling by mapping coordinates to continuous pixel functions, enabling arbitrary-magnification SR reconstruction. It also has stronger adaptability in small-sample data scenarios. Such methods do not rely on large-scale paired training data and can learn continuous image representations according to the intrinsic features of a single low-resolution image, which can better adapt to the actual clinical scenarios with high data annotation costs and limited sample sizes of specific diseases. CycleINR proposed by Fang et al. introduces the Cycle-Consistent Loss (CCL) into the implicit neural representation framework [24]. As shown in Figure 5, through a closed-loop process of generating HR images in the forward path and reconstructing the original LR images from the generated HR images in the backward path, the model ensures the consistency of key features (like noise level) between the generated images and the original images, which effectively alleviates the common problems of over-smoothing and inter-slice discontinuity in medical image super-resolution reconstruction.

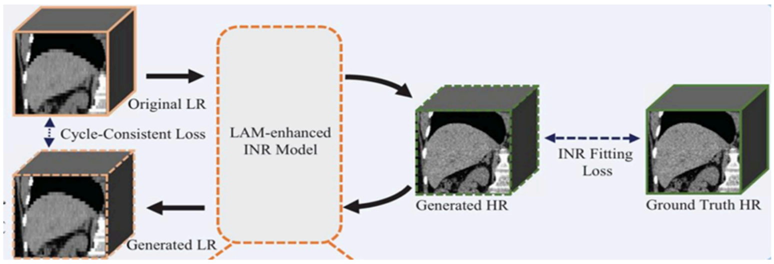


Figure 5. CycleINR architecture

4. Challenges and future development directions

4.1. Current challenges

Most current SR algorithms rely on simple artificial degradation models for training, which poorly reflect real clinical image degradation caused by motion artifacts, system noise, and compression loss. As a result, their reconstruction performance often drops in clinical scenarios. Evaluation metrics such as PSNR and SSIM, commonly used in computer vision, do not fully correspond to clinical utility; for example, high PSNR often causes over-smoothing, and SSIM cannot accurately assess perceptual quality or clinical relevance [8,25].

Therefore, data scarcity and privacy further challenge medical image SR. Supervised methods require paired HR-LR images, but high-quality clinical data are costly, labor-intensive to annotate, and vary across centers and devices. Resource-limited clinical workstations cannot support large SR models due to high parameter counts and computational demands. And lightweight SR approaches reduce computational load but often compromise reconstruction accuracy. Moreover, generalization across centers and devices remains limited. Different scanning protocols and equipment parameters lead to unstable SR performance. Multi-center data collection requires strict regulatory compliance, patient privacy protection, and secure data transfer, hindering large-scale clinical deployment and model generalization.

DOI: 10.54254/2755-2721/2026.GU33298

4.2. Future development directions

Future development in medical image super-resolution is expected to focus on unsupervised and semi-supervised learning, network and learning strategy innovation, and multi-modal information fusion. Transitioning from supervised to unsupervised or semi-supervised methods reduces reliance on large paired LR-HR datasets, lowering data collection and annotation costs. These approaches can directly use real clinical LR images for training and testing, enhancing model generalization in practical scenarios. At the same time, the design of network structures and learning strategies remains crucial for SR performance. Future research should explore universal upsampling methods supporting arbitrary magnification, as well as advanced architectures and techniques, like attention mechanisms, multi-scale learning, network-in-network structures, and information distillation, to improve reconstruction quality. Developing lightweight networks that balance computational cost and accuracy is particularly promising. Additionally, most existing SR methods focus on single-modal images, whereas clinical practice often acquires multiple modalities (CT, MRI, PET) for the same patient. Fusing multi-modal features can compensate for information loss, enhance anatomical and functional details, and improve reconstruction accuracy for complex structures, thus increasing clinical value [1].

5. Conclusion

This paper introduces the basic theory of super-resolution and the commonly used objective and subjective evaluation metrics for assessing the performance of super-resolution. It focuses on the innovations and advancements in the network structure design and performance optimization of five mainstream super-resolution network models (convolutional neural networks, generative adversarial networks, Transformer, Mamba, and INR) in medical image applications. Besides, it examines the key challenges and obstacles in the field of medical CT image super-resolution reconstruction, and summarizes and predicts the future development trends. The results show that deep learning-based super-resolution technology still has vast potential in the field of medical imaging, and there are still many unresolved issues and challenges that need to be addressed. More researchers from the fields of computer vision and medical imaging must carry out interdisciplinary and innovative research work, and combine the characteristics and actual clinical needs of clinical medical images to develop more practical and efficient medical image super-resolution technologies. This will better serve clinical diagnosis and treatment and promote the development of medical imaging technology.

References

[1] Singh, K., Saxena, M., & M. Tech Scholar. (2021). A review on medical image super resolution with application of deep learning. Smart Moves Journal Ijoscience, 7, 25-29.
[2] Hou, H., et al. (2022). CT image quality enhancement via a dual-channel neural network with jointing denoising and super-resolution. Neurocomputing, 492, 343-352.
[3] Baker, S., & Kanade, T. (2002). Limits on super-resolution and how to break them. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(9), 1167-1183.
[4] Dong, C., et al. (2015). Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), 295-307.
[5] Umehara, K., Ota, J., & Ishida, T. (2018). Application of super-resolution convolutional neural network for enhancing image resolution in chest CT. Journal of Digital Imaging, 31(4), 441-450.
[6] Sood, R., et al. (2018). An application of generative adversarial networks for super resolution medical imaging. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE.

[7] Wang, Z., Chen, J., & Hoi, S. C. H. (2020). Deep learning for image super-resolution: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10), 3365-3387.
[8] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600-612.
[9] Pambrun, J.-F., & Noumeir, R. (2015). Limitations of the SSIM quality metric in the context of diagnostic imaging. In 2015 IEEE International Conference on Image Processing (ICIP). IEEE.
[10] Sheikh, H. R., Bovik, A. C., & De Veciana, G. (2005). An information fidelity criterion for image quality assessment using natural scene statistics. IEEE Transactions on Image Processing, 14(12), 2117-2128.
[11] Sheikh, H. R., & Bovik, A. C. (2006). Image information and visual quality. IEEE Transactions on Image Processing, 15(2), 430-444.
[12] Keys, R. (2003). Cubic convolution interpolation for digital image processing. IEEE Transactions on Acoustics, Speech, and Signal Processing, 29(6), 1153-1160.
[13] Hong, S.-H., Wang, L., & Truong, T.-K. (2018). An improved approach to the cubic-spline interpolation. In 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE.
[14] Wang, Y., et al. (2018). A fully progressive approach to single-image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops .
[15] Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In European Conference on Computer Vision. Cham: Springer International Publishing.
[16] 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.
[17] Ledig, C., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
[18] Gu, Y., et al. (2020). MedSRGAN: Medical images super-resolution using generative adversarial networks. Multimedia Tools and Applications, 79(29), 21815-21840.
[19] Vaswani, A., et al. (2017). Attention is all you need. In Advances in Neural Information Processing Systems, 30.
[20] Puttagunta, M., & Subban, R. (2022). Swinir transformer applied for medical image super-resolution. Procedia Computer Science, 204, 907-913.
[21] Liu, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision.
[22] Gu, A., & Dao, T. (2024). Mamba: Linear-time sequence modeling with selective state spaces. In First Conference on Language Modeling.
[23] Ji, Z., et al. (2025). Self-prior guided Mamba-UNet networks for medical image super-resolution. In International Conference on Pattern Recognition. Cham: Springer.
[24] Fang, W., et al. (2024). CycleINR: Cycle implicit neural representation for arbitrary-scale volumetric super-resolution of medical data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
[25] Sajjadi, M. S. M., Scholkopf, B., & Hirsch, M. (2017). Enhancenet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE International Conference on Computer Vision