Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ29490

Super-resolution Medical Deep Learning in Brain, Lung and Teeth Regions

Ruochen Xing

College of Computing & Data Science, Nanyang Technological University, Singapore, Singapore
XING0065@e.ntu.edu.sg

Abstract. Image super-resolution technology is an important technique that breaks through the hardware limitations of existing medical imaging equipment, enhances image resolution and signal-to-noise ratio, and provides key support for precise disease diagnosis and clinical treatment decisions. This article aims to systematically review the latest research progress and application value of image super-resolution technology in the three key medical fields of brain, lung and dentistry. Through in-depth analysis of relevant domestic and foreign literature, the mainstream super-resolution algorithms based on deep learning models such as convolutional neural networks, generative adversarial networks, and Transformer were sorted out and compared. The technical paths and adaptability of these methods on different modalities of medical images such as MRI, CT, and X-rays were evaluated with emphasis. The research results show that networks such as CNN, GAN, and ESRGAN have developed to varying degrees in different fields, which can provide certain reference value for the future technological development of this field and related technologies in other fields.

Keywords: Super-resolution, Medical Image Processing, Deep Learning, Image Reconstruction

1. Introduction

In today's medical field, image analysis has become an indispensable part. Its role is not only reflected in clinical diagnosis but also in providing great assistance for the research of various difficult and complicated diseases. In image analysis, the clarity of the image can be said to be the most important factor. Therefore, making every effort to improve the clarity of the image has become a time-honored topic. Image super-resolution is a technique that restores high-resolution images through low-resolution images or image sequences, which provides great assistance in achieving the goal of improving image clarity.

During the development of image super-resolution technology, the following stages have been gone through: 1) Traditional algorithm stage: Traditional algorithms include interpolation methods and reconstruction methods [1, 2]. These algorithms do not rely on big data training and are mainly implemented based on artificially designed models and mathematical principles. However, neither of these two algorithms can provide excellent output results, cannot restore high-frequency details, is computationally complex, and relies on manual prior accuracy. 2) Deep learning-based methods: This type of method is currently a research point hotspot and mainstream. By training neural

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ29490

networks to learn complex mapping relationships from low definition to high definition from large amounts of data, its effect far exceeds that of traditional methods. At present, all algorithms based on deep learning are supervised learning with paired data. Convolutional Neural Networks (CNN) needs to be used in super-resolution algorithms, and then upsampling is performed to achieve the goal [3]. Predefined upsampling and end upsampling are two upsampling methods. The combination of predefined upsampling and traditional algorithms, using bicubic interpolation to enlarge low-resolution (LR) images and then send them to the network for detail restoration, faces the disadvantages of slow speed and large computational load. Most of the computations for end upsampling are carried out in a low-resolution space, and finally sub-pixel convolution or transposed convolution is used, which overcomes the disadvantage of slow speed [4]. On this basis, the image super-resolution algorithm continuously evolves its network structure, achieving further clarification of the image through three methods: iterating deeper networks, residual learning, or modifying the attention mechanism [5-7]. Meanwhile, to address the texturing distortion issue of generated images, an algorithm based on generative adversarial networks was proposed, which not only ensures high precision but also offers more realistic textures and better visual effects in the generated high-definition images [8].

The aim of this paper is to provide a comprehensive review of different super-resolution models used in various medical fields. It will mainly present from three aspects: brain super-resolution, lung super-resolution, and dental super-resolution. Comparisons will be made in the temporal dimension based on the development of model usage before and after time or different models adopted within the same period, as well as in the spatial dimension within different fields. In the second part, “method”, the differences in the ways each model analyzes datasets within each medical field will be elaborated in detail. In the third part, discussion, the difficulties, challenges and future improvement directions faced when applying super-resolution networks in the medical field will be summarized and described. A full summary and description will be provided in the conclusion section at the end.

2. Method

2.1. Brain super-resolution

Brain imaging generally uses Magnetic Resonance Imaging (MRI) to show the condition of the brain. MRI has a good contrast and can display the soft tissue structure within the brain [9]. The hospital will use dedicated scanning equipment to scan the patient’s head.

Nandal et al. used the BraTS dataset in their study, slicing 3D MRI volumes into 2D images and generating LR images by downsampling at 4x [10]. The pre-trained Real-ESRAN model was first applied to the brain MRI image super-resolution task. In transfer learning, part of the underlying network was frozen, and the Adam optimizer was used to fine-tuning the high-level structure, enhancing the processing ability for complex degradation situations [11]. The generator in the model still uses the same Residual-in-Residual Dense Block (RRDB) structure as ESRGAN, ensuring that the network can greatly enhance its representativeness through multi-layer residuals and dense connections. The key improvement of Real-ESRGAN over ESRGAN lies in the discriminator. The original ESRGAN uses a VGG-style discriminator that only outputs a true/false probability value, providing global, image-level feedback. Real-ESRGAN uses a U-Net architecture with jump connections, which can not only output a global authenticity judgment, but also provide pixel-by-pixel and local gradient feedback [12]. This improvement enables Real-ESRGAN not only to

DOI: 10.54254/2755-2721/2026.TJ29490

maintain high perceptual quality but also to effectively suppress artifacts, making it suitable for the high requirements of medical images for details and edges.

Zhou et al., in their research, based on the concept of ESRGAN, modified some architectures, and finally used the MRBT-SR-GAN neural network to analyze and process the images [13]. After collecting T2-FLAIR images of 237 glioma patients, all the data were manually labeled with 76 examples to form a segmented dataset [14]. First, train a PSNR-oriented model mainly based on L1 loss as the generator and training. Then, on this basis, introducing adversarial loss and perceptive loss based on U-Net to formally train MRBT-SR-GAN. The MRBT-SR-GAN not only uses the U-Net encoder as the sensing loss, but also its generator adopts a uniquely designed dual-path generator with unique high-resolution paths and low-resolution paths. High-resolution paths maintain the overall structure and edge fidelity of the image, while low-resolution paths focus on restoring rich detailed textures. The two are fused in the convolutional part at the tail to jointly generate high-fidelity and detailed super-resolution results. This network has significantly enhanced the accuracy and clinical practicality of brain MRI super-resolution.

2.2. Lung super-resolution

Kim et al. used the National Lung Screening Trial (NLST) public dataset to select 28628 chest CT images from hundreds of thousands of CT sequences as the training set [15, 16]. This project selects ESRGAN as the basic framework, but it has undergone a three-dimensional transformation, replacing all 2D convolutional layers with 3D convolutional layers to process data. In addition, the VoxelShuffle-SliceGeneration(SG) module has been innovatively designed to replace the original 2D PixelShuffle module in PyTorch [17]. This module can convert the information of the channel dimension to the spatial depth dimension, thereby achieving an improvement in resolution along the Z-axis. After the dataset is imported, the axial images are reformatted as coronal planes and cropped into groups of 16 volume blocks for input into the network. A two-stage training strategy is adopted. Firstly, a PSNR-oriented model is pre-trained using L1 loss. Then, the perceptual loss and adversarial loss based on VGG are introduced for fine training. Finally, the coronal plane super-resolution results output by the network are re-sliced back to the axial plane. In clinical verification, this not only conducted qualitative and quantitative evaluations of image quality, but also significantly improved classification consistency.

Chen et al. utilized the publicly available COVID-CT dataset and proposed a dual-path residual information distillation network for super-resolution of lung CT images to achieve super-resolution of lung images [18, 19]. The dual-path structure can handle low-frequency and high-frequency information separately, avoiding frequency confusion and improving the reconstruction quality. Low-resolution CT images are inverted and sent to the low-frequency path and the high-frequency path after input. The contour information of the low-frequency path is extracted through shallow convolution. In addition, the residual information distillation module has been designed to extract clearer high-frequency features through multi-level information distillation. High-frequency paths extract detailed features through RIDM and gradient identity paths. RIDM is composed of multiple residual blocks and information distillation fast cascades, gradually refining high-frequency information. Finally, the image gradient residual training method is introduced. The Sobel operator is utilized to extract edge information, guide high-frequency path learning, accelerate convergence and enhance detail recovery. The output features of the two paths are fused and upsampled through sub-pixel convolution to generate a high-resolution image. This network achieves superior performance to methods such as SRCNN [20], ESPCN [21], VDSR [22] and IMDN [23] on the

DOI: 10.54254/2755-2721/2026.TJ29490

COVID-CT dataset. Meanwhile, it has a small number of parameters and a fast reasoning speed, and is suitable for medical image-assisted diagnosis.

2.3. Teeth super-resolution

Hartvani et al. employed two CNN architectures to enhance the resolution of dental cone-beam CT(CBCT) images [24]. For the effects from 17 isolated teeth, they respectively used Subpixel CNN and U-net for super-resolution reconstruction and compared them with the traditional SR method based on reconstruction [25]. Subpixel CNN adopted a six-layer convolutional structure. The last layer outputs a four-channel feature map and achieves 2x upsampling through the Retilling operation. U-net adopts an encoder-decoder structure, consisting of four downsampling and upsampling steps [26]. Each layer contains convolution, batch normalization, and Leaky-ReLU activation functions. In this process, a background removal preprocessing strategy was also proposed to effectively suppress the influence of noise structure differences between CBCT and μCT on the super-resolution effect.

Mohammad-Rahimi et al. applied a variety of new network applications to dental panoramic X-ray films based on the above two super-resolution networks [27]. All the datasets are derived from over 800 anonymous panoramic X-ray images from a private radiology center in Tehran, Iran. They are divided into 70% training, 15%validation, and 15% testing segments, and are downsampled 4x to simulate low-resolution images, with the original images serving as high-resolution images. This is the first systematic comparison of the performance of multiple advanced SR models on panoramic X-ray films, including traditional CNNS, generative adversarial networks [28], medical image segmentation networks, Transformer architectures (SwinIR) and frequency-domain texture estimators. SwinIR is based on the Swin Transformer and includes shallow feature extraction, deep Transformer block residual learning and reconstruction modules, which capture the global image by using the self-attention mechanism [29]. LTE jointly trains the frequency-domain dominant frequency estimator and the depth SR architecture, characterizes textures in 2D Fourier space, continuously reconstructs details through implicit functions, and emphasizes the recovery of high-frequency information [30]. This study provides a reference for model selection in dental imaging super-resolution.

3. Discussion

3.1. Challenge

Even though image super-resolution technology in the medical field has made considerable progress, it still faces many difficulties and challenges on the whole.

3.1.1. The difference between actual image data and real data

The acquisition of medical imaging data is highly challenging. The ideal scenario of “low-definition and high-definition paired data collected from the same part and at the same time” is almost non-existent in reality. What is more common in clinical practice is a large amount of isolated, unpaired data collected from different devices and under different protocols. For instance, it is difficult to require patients to repeatedly take over CT scans, especially high-definition scans with high radiation doses, to build a training set. This makes mainstream supervised learning algorithms(such as EDSR and RCAN [7]) hard to apply directly due to the lack of pairing labels.

DOI: 10.54254/2755-2721/2026.TJ29490

3.1.2. The contradiction between computational efficiency and clinical real-time requirements

The PACS system in the hospital generates terabytes of image data every day, which radiologists need to quickly access and diagnose. A super-resolution algorithm that takes several seconds or even minutes to process a single image cannot be integrated into the clinical workflow. Especially the latest diffusion models (such as SR3 [31]), although the generation quality is astonishing, the computational cost of their iterative denoising process is huge, which directly conflicts with the "urgent" demand. In addition, the processing of three-dimensional medical volume data (such as the entire sequence of MRI and CT) will further increase the computational burden.

3.1.3. The trade-off between texture authenticity and diagnostic safety

In the natural image super-minute, the goal of GAN and diffusion models is to generate visually realistic and textured images, even allowing for the "creation" of reasonable details. However, in the medical field, any "creation" that cannot be clearly verified may be well-known. If the details generated by the algorithm are confused with the real pathology, it may lead to misdiagnosis. This is the fundamental contradiction between "authenticity" and "security". An image with clearer texture but with a smile artifact introduced may have negative clinical value.

3.2. Future prospects

To solve the above problems, the following aspects can be used as references.

3.2.1. Address data differences through both algorithm and data aspects

At the algorithmic level, vigorously develop unsupervised and weakly supervised learning frameworks. Recurrent consistent generative adversarial networks such as CycleGAN are among the most promising solutions at present [32]. It only requires images from two domains (such as the low-definition CT image set and the high-definition CT image set), without pairing data, to learn the mapping relationship between them. Future research can focus on improving the anatomical structure fidelity of the generated results and preventing the generation of "hallucination" artifacts.

At the data level, simulation degradation technology is adopted. By precisely modeling the point spread function (PSF) [33], noise model and downsampling kernel of the medical imaging system. Realistic low-definition images can be artificially generated from existing high-definition medical images, thereby "creating" a large amount of paired data for training supervisory models. The key lies in the accuracy of the degradation model. The key lies in the accuracy of the degradation model, and it is necessary to collaborate with physical engineers to ensure that the distribution of the simulated data is consistent with that of the real clinical low-definition data.

3.2.2. Address the timeliness of algorithms through two aspects: hardware deployment and algorithms

At the hardware and deployment level, dedicated hardware acceleration and model optimization are adopted. By leveraging inference acceleration libraries such as NVIDIA GPUs and TensorPix to perform quantization, layer fusion, and graph optimization on trained models, the inference speed can be significantly enhanced. Meanwhile, explore the deployment of lightweight models on edge computing devices to achieve real-time bedside super-resolution.

DOI: 10.54254/2755-2721/2026.TJ29490

At the algorithm level, design a more efficient network architecture. The research direction should tend towards training small models with large ones, removing redundant network connections, and adjusting the computational load according to the complexity of image regions, etc. For instance, simplify the background area and conduct full force calculations on the key lesion areas to balance the overall efficiency and effect.

3.2.3. Address the security contradiction through two aspects: algorithm design and verification evaluation

At the algorithm design level, priority should be given to developing algorithms that prioritize fidelity. In the medical field, models based on L1/L2 loss functions such as EDST may produce overly smooth images, but their outputs are more deterministic and have lower risks. If GAN is used, the constraints of its discriminator must be strengthened, for instance, by introducing a weighted combination of perceptual loss and adversarial loss, with fidelity loss as the main focus, to ensure that structural accuracy is superior to texture richness.

At the verification and evaluation level, establish a clinical relevance assessment system that goes beyond PSNR/SSIM. The output of the super-resolution algorithms must be blindly evaluated by radiologists. The key indicators should include: whether the pathological detection rate has significantly improved, whether the measurement error of the anatomical structure is within the clinically acceptable range, and whether there are misleading artifacts. Algorithms cannot be “black boxes”, and their uncertainties need to be quantified and informed.

4. Conclusion

This study mainly studied the development of image super-resolution technology in different medical fields to make medical images clearer, and made a comprehensive comparison from both the temporal longitudinal and domain transverse aspects. After studying various super-resolution network models and related factors such as activation functions and loss functions, the research results show that most of the image super-resolution technologies in brain and lung medicine are based on secondary or multiple optimizations of the ESRGAN model, which dental medicine still uses CNN architecture. In the future, the further study plans to continuously monitor the development of image super-resolution technology in various medical fields, and focus on whether these improvements can solve the problems that image super-resolution technology is currently facing in the medical field.

References

[1] Siu, W. C., & Hung, K. W. (2012). Review of image interpolation and super-resolution. In Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (pp. 1–10). IEEE.
[2] Qiu, D., Cheng, Y., & Wang, X. (2023). Medical image super-resolution reconstruction algorithms based on deep learning: A survey. Computer Methods and Programs in Biomedicine, 238, 107590.
[3] Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In European Conference on Computer Vision (pp. 391–407). Cham: Springer International Publishing.
[4] Michelini, P. N., Lu, Y., & Jiang, X. (2022). Edge-SR: Super-resolution for the masses. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1078–1087).
[5] Hitawala, S., Li, Y., Wang, X., et al. (2018). Image super-resolution using VDSR-ResNeXt and SRCGAN. arXiv preprint arXiv: 1810.05731.
[6] Kuriakose, B. M. (2023). EDSR: Empowering super-resolution algorithms with high-quality DIV2K images. Intelligent Decision Technologies, 17(4), 1249–1263.

DOI: 10.54254/2755-2721/2026.TJ29490

[7] Lin, Z., Garg, P., Banerjee, A., et al. (2022). Revisiting RCAN: Improved training for image super-resolution. arXiv preprint arXiv: 2201.11279.
[8] Wang, X., Yu, K., Wu, S., et al. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops (pp. 0–0).
[9] Rudie, J. D., Gleason, T., Barkovich, M. J., et al. (2022). Clinical assessment of deep learning–based super-resolution for 3D volumetric brain MRI. Radiology: Artificial Intelligence, 4(2), e210059.
[10] Nandal, P., Pahal, S., Khanna, A., & Pinheiro, P. R. (2024). Super-resolution of medical images using Real ESRGAN. IEEE Access, 12, 176155–176170. https: //doi.org/10.1109/ACCESS.2024.3497002
[11] Zhang, Z. (2018). Improved Adam optimizer for deep neural networks. In Proceedings of the IEEE/ACM 26th International Symposium on Quality of Service (IWQoS) (pp. 1–2).
[12] Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) (pp. 1905–1914).
[13] Zhou, Z., Ma, A., & Feng, Q. (2022). Super-resolution of brain tumor MRI images based on deep learning. Journal of Applied Clinical Medical Physics, 23(11), e13758.
[14] Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.
[15] Kim, D. O. (2025). CT image quality enhancement via deep-learning-enabled super-resolution (Doctoral dissertation, Graduate School of Seoul National University).
[16] NLST. (205). The Cancer Imaging Archive (TCIA). Retrieved from https: //www.cancerimagingarchive.net/collection/nlst/
[17] PyTorch. (2025). PixelShuffle — PyTorch 2.5 documentation. Retrieved from https: //pytorch.org/docs/stable/generated/torch.nn.PixelShuffle.html
[18] Chen, Y., Zheng, Q., & Chen, J. (2022). Double paths network with residual information distillation for improving lung CT image super resolution. Biomedical Signal Processing and Control, 73, 103412.
[19] Graviti. (2020). A CT Scan Dataset about COVID-19. Retrieved from https: //www.graviti.cn/open-datasets/COVID_CT
[20] Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In European Conference on Computer Vision (pp. 184–199). Springer.
[21] Shi, W., Caballero, J., Huszár, F., et al. (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 (CVPR) (pp. 1874–1883).
[22] Kim, J., Kwon Lee, J., & Mu Lee, K. (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).
[23] Hui, Z., Gao, X., Yang, Y., et al. (2019). Lightweight image super-resolution with information multi-distillation network.
[24] Hatvani, J., Horváth, A., Michetti, J., Basarab, A., Kouamé, D., & Gyöngy, M. (2019). Deep learning-based super-resolution applied to dental computed tomography. IEEE Transactions on Radiation and Plasma Medical Sciences, 3(2), 120–128. https: //doi.org/10.1109/TRPMS.2018.2827239
[25] Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In Proceedings of the 13th European Conference on Computer Vision (ECCV) (pp. 184–199).
[26] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Conference on MICCAI (pp. 234–241).
[27] Mohammad-Rahimi, H., Vinayahalingam, S., Mahmoudinia, E., Soltani, P., Bergé, S. J., Krois, J., & Schwendicke, F. (2023). Super-resolution of dental panoramic radiographs using deep learning: A pilot study. Diagnostics, 13, 996. https: //doi.org/10.3390/diagnostics13050996
[28] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., 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 (pp. 4681–4690).
[29] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1833–1844).
[30] Lee, J., & Jin, K. H. (2022). Local texture estimator for implicit representation function. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1929–1938).
[31] Saharia, C., Ho, J., Chan, W., et al. (2022). Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4), 4713–4726.

DOI: 10.54254/2755-2721/2026.TJ29490

[32] Liu, H., Liu, J., Hou, S., et al. (2023). Perception consistency ultrasound image super-resolution via self-supervised CycleGAN. Neural Computing and Applications, 35(17), 12331–12341.
[33] Joshi, N., Szeliski, R., & Kriegman, D. J. (2008). PSF estimation using sharp edge prediction. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1–8). IEEE.