Review Article

Deep Learning for Enhancing High-resolution BOLD-fMRI: A Narrative Review of Super-resolution, Segmentation, and Registration Methods

Yanong Li[1,2#], Yawei Liu[3,4#], Zewen Zhang[5], Tao Wan[5*] and Hailong Liu[1,6,7,8*]

[1]*Department of Radiation Oncology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China;* [2]*Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China;* [3]*School of Basic Medicine, Capital Medical University, Beijing, China;* [4]*Beijing Luhe Hospital Affiliated to Capital Medical University, Beijing, China;* [5]*School of Biological Science and Medical Engineering (BSME), Beihang University, Beijing, China;* [6]*Beijing Neurosurgical Institute, Beijing, China;* [7]*China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing, China;* [8]*Chinese Institute for Medical Research, Beijing, China*

Abstract

Blood oxygen level-dependent (BOLD) functional magnetic resonance imaging (fMRI) is essential for non-invasively investigating brain function. However, conventional fMRI methods are limited by low spatial and temporal resolution. This narrative review evaluates recent advancements in deep learning techniques for high-resolution BOLD-fMRI reconstruction, focusing on super-resolution, segmentation, and image registration. A comprehensive literature search was conducted across PubMed, IEEE, Scopus, and Web of Science databases for the period 2000–2023. Studies employing deep learning methods, including convolutional neural networks, transformer-based models, and generative adversarial networks for super-resolution, segmentation, and registration of BOLD-fMRI, were included. Deep learning approaches demonstrated significant improvements in spatial resolution, segmentation accuracy, and registration robustness. Convolutional neural network-based models, particularly generative adversarial networks, notably improved image reconstruction quality and detail preservation. Preliminary studies targeting specific brain regions such as the cerebellum and hippocampus showed promise; however, systematic evaluations across broader brain areas and large-scale clinical validations remain limited. While deep learning techniques have led to substantial advancements in high-resolution BOLD-fMRI reconstruction, future research should focus on standardized protocols, multi-center validation, and improving computational efficiency and model generalization to enhance clinical utility.

Introduction

Blood oxygen level-dependent (BOLD) functional magnetic resonance imaging (fMRI) has revolutionized neuroscience research by providing a non-invasive method to visualize and quantify brain activity associated with various cognitive, motor, and emotional functions.[1] However, traditional fMRI methods face critical limitations in spatial and temporal resolution, which restrict their capability to detect subtle and detailed neural activation patterns necessary for precise diagnosis and scientific investigation.[1,2]

Recent advances in artificial intelligence, particularly deep learning, have shown promise in overcoming these limitations. Techniques involving convolutional neural networks (CNNs), Transformer-based architectures, and generative adversarial networks have significantly enhanced fMRI capabilities in terms of spatial resolution, automated segmentation accuracy, and the robustness of image registration processes.[1,3] These methods enhance the visualization of neural activity and provide detailed insights into brain structures and functions, facilitating more accurate interpretations in both clinical and research contexts.[4]

Although deep learning techniques have been widely applied across various neuroimaging domains, research specifically evaluating their effectiveness in individual brain regions remains com-

paratively limited.[5] Some studies have implemented deep learning approaches for segmentation and super-resolution reconstruction in specific brain regions.[6–8] However, systematic evaluations across broader brain regions and large-scale clinical applications remain relatively scarce.

This narrative review aimed to comprehensively summarize and critically assess recent advancements in deep learning applied to high-resolution fMRI reconstruction, with a particular focus on image super-resolution, segmentation, and registration techniques. Additionally, it discusses the key challenges impeding clinical implementation of these technologies, including computational costs, data heterogeneity, and limited generalizability across diverse patient populations and imaging settings. Addressing these challenges through standardized protocols and extensive multi-center validation studies will be essential for translating deep learning methods from research environments into routine clinical practice.

Literature search strategy

A comprehensive literature search was conducted across PubMed, IEEE, Scopus, and Web of Science to identify relevant articles related to brain segmentation, super-resolution reconstruction, and image registration. Search terms included "cerebellar segmentation", "BOLD-fMRI", "super-resolution reconstruction", "deep learning", and "image registration". Only articles published in English between 2000 and 2023 were included. Studies that were case reports, reviews, or not focused on deep learning techniques were excluded. A narrative synthesis approach was used to summarize the findings across the included studies. The data were categorized into three main areas: (1) segmentation of cerebellar tissues, (2) super-resolution reconstruction of BOLD-fMRI, and (3) registration algorithms. Where appropriate, findings were compared across studies, and key trends and knowledge gaps were highlighted.

Current landscape and challenges in deep learning for BOLD-fMRI

The three-dimensional segmentation algorithm for cerebellar tissues in this study encompasses cerebellar segmentation for both structural and functional magnetic resonance imaging (MRI), super-resolution reconstruction, and BOLD sequence registration. Accordingly, we analyze current research in three key areas: segmentation, super-resolution, and registration.

Status of research on image segmentation

In recent years, deep learning has achieved remarkable progress in medical image segmentation, leading to the development of several highly effective models. One pioneering model, U-Net, introduced by Ronneberger *et al*.[9] in 2015, employs an encoder-decoder architecture with skip connections, enabling the integration of multi-resolution features. U-Net has been extensively applied to tasks such as cell nucleus segmentation, organ segmentation, and lesion detection. Subsequently, Milletari *et al*.[10] introduced V-Net in 2016, a model specifically designed for 3D segmentation. V-Net is capable of processing volumetric data, making it particularly suitable for applications involving computerized tomography (CT) and MRI scans. Attention U-Net enhances the standard U-Net architecture by incorporating attention mechanisms,[11] which allow the model to focus on relevant regions within an image, thereby improving accuracy in complex backgrounds. This model is especially effective for fine-grained segmentation tasks, such as

tumor delineation. The DeepLab series,[12] including DeepLabv3 and DeepLabv3+,[13] represents a significant advancement in high-resolution image segmentation. These models employ dilated convolutions and atrous spatial pyramid pooling to capture multi-scale contextual information, enhancing segmentation performance. nnU-Net,[14] proposed by Isensee *et al*.,[14] is a self-adapting U-Net framework that automatically configures the network architecture and training strategies based on the dataset characteristics. This model has demonstrated exceptional adaptability and performance across a wide range of medical image segmentation tasks. Çiçek *et al*.[15] extended the U-Net architecture to 3D U-Net, which is specifically designed for 3D segmentation tasks involving organs such as the brain, lungs, and liver. This model leverages 3D convolutions to effectively capture volumetric information. Figure 1 presents sketches of the deep neural networks mentioned.

Transformer-based models have also made significant contributions to medical image segmentation. Swin-Unet combines the strengths of the Swin Transformer and U-Net, utilizing hierarchical attention mechanisms to achieve multi-scale feature fusion and significantly improve segmentation performance.[16] TransUNet integrates Transformer modules into the U-Net framework, effectively handling complex image structures by capturing long-range dependencies.[17] MedT (Medical Transformer) further exemplifies the potential of Transformer-based architectures in medical image segmentation.[18] By capturing long-range dependencies, MedT significantly enhances segmentation accuracy. Finally, SegResNet combines residual networks with U-Net, employing residual connections to improve model depth and performance.[19] This model is particularly well-suited for tasks requiring high-detail processing, such as tissue and lesion segmentation.

In summary, these deep learning models have demonstrated exceptional performance in various medical image segmentation tasks through continuous optimization and innovation. Their advancements have significantly propelled the field of medical image processing, highlighting the transformative potential of deep learning in medical applications.

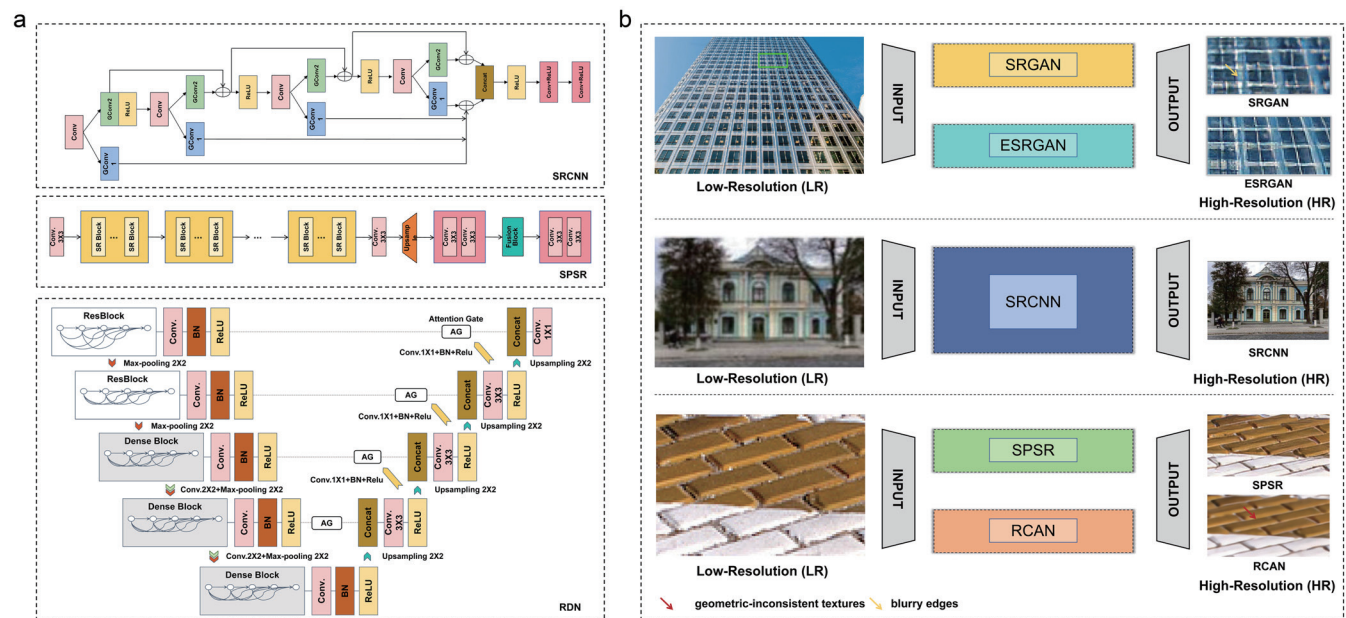#### Status of super-resolution research

Image super-resolution reconstruction, a method of reconstructing low-resolution images into high-resolution images, is one of the key technologies used to improve the resolution of real-world images and videos in computer vision tasks. Image super-resolution reconstruction has been widely applied in the real world, including in hyperspectral imaging,[20] medical image processing,[21] and facial recognition.[22] Apart from improving image resolution, image super-resolution reconstruction also assists, to a certain extent, with other tasks related to computer vision.[23] Due to the inadaptability of the image super-resolution problems, where multiple high-resolution images can correspond to one low-resolution image, the task of image super-resolution is quite challenging in the image reconstruction process.[24]

In recent years, with the application of convolutional neural networks in image super-resolution research, from the super-resolution convolutional neural network,[25] based on traditional convolutional neural networks, to the super-resolution generative adversarial network, based on deep residual generative adversarial networks,[26] various image super-resolution methods based on deep convolutional neural networks, relying on different network architecture designs and training strategies, have developed rapidly to improve the performance of image super-resolution reconstruction tasks (Fig. 2).

Super-resolution of medical images refers to the acquisition
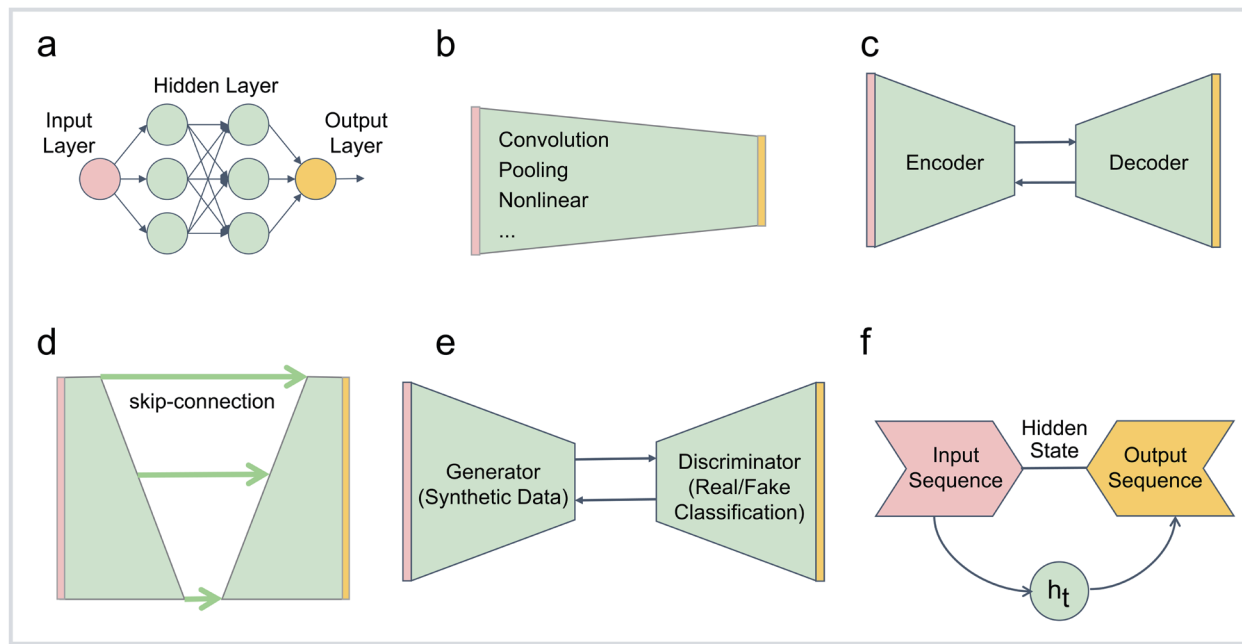
Fig. 1. Overview of deep learning architectures used in magnetic resonance imaging (MRI) super-resolution, segmentation, and registration. (a) A standard convolutional neural network (CNN) architecture commonly used for MRI image enhancement. (b) A generative adversarial network (GAN)-based approach for MRI super-resolution reconstruction, where the generator synthesizes high-resolution images from low-resolution inputs and the discriminator ensures realism. (c) A Transformer-based architecture, leveraging self-attention mechanisms to capture long-range spatial dependencies in MRI scans. (d) A U-Net structure for functional magnetic resonance imaging (fMRI) segmentation, demonstrating skip connections that help retain spatial information. (e) A deep learning-based image registration framework that aligns fMRI scans across different acquisition conditions. (f) A comparative analysis of different deep learning models in terms of structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR), showcasing performance improvements over traditional methods.

Fig. 2. Super-resolution research section. (a) Simplified network architecture of the super-resolution convolutional neural network (SRCNN), structure-preserved super-resolution (SPSR), and residual dense network (RDN). (b) Comparison of visual effects via different super-resolution algorithm outputs. The peak signal-to-noise ratio (PSNR)/structural similarity index measure (SSIM) values tested on the 'Urban100' dataset were 24.397/0.9381 for enhanced super-resolution generative adversarial networks (SRGAN), 24.360/0.9453 for European Space GaN (ESGAN), and 26.24/0.7989 for extended super-resolution convolutional neural network (ESRCNN) (×3). Residual channel attention networks (RCAN) generates straight but blurry edges for the bricks, while SPSR methods better preserve gradients and structures. AG, attention gate; BN, batch normalization.
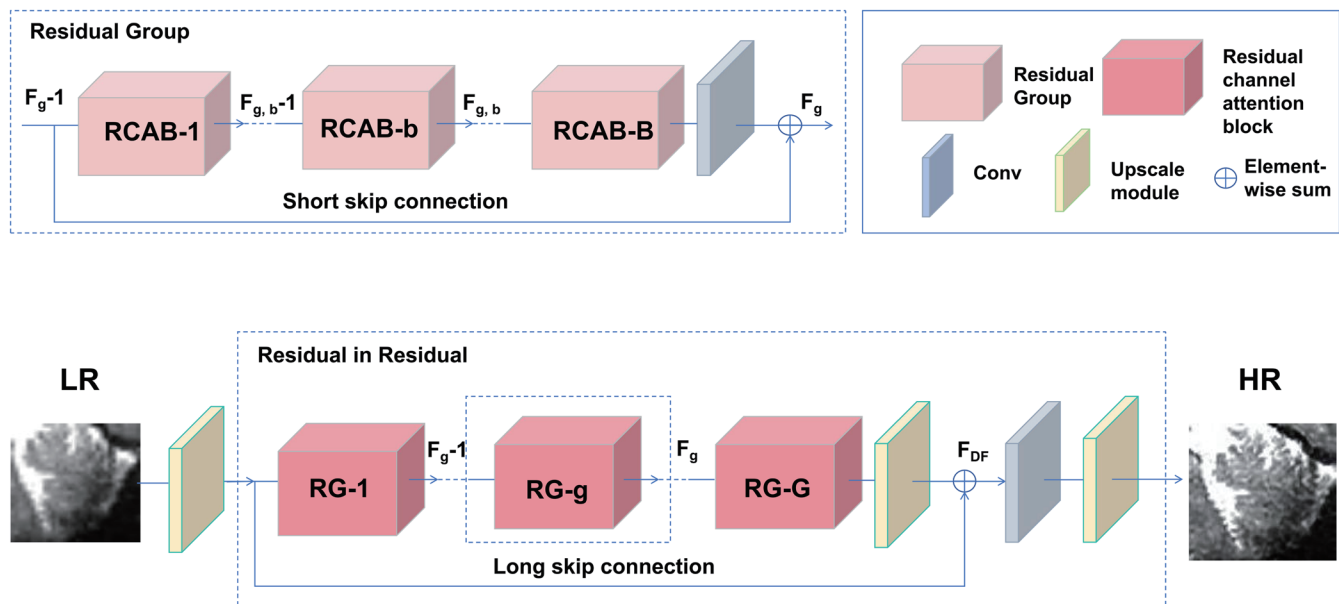
Fig. 3. The network architecture of the residual channel attention network (RCAN). Low-resolution images (LR) are input, and feature maps are obtained through 3×3 convolution. After passing through a residual in residual (RIR) module, followed by upsampling and a 3×3 convolution layer, high-resolution images (HR) are obtained. RCAB, residual channel attention network; RG, residual group.

of low-resolution medical images from various medical imaging devices and the restoration of high-resolution medical images with rich details and clear textures using deep convolutional neural networks. This is conducive to clinical diagnosis, image segmentation,[27,28] image registration,[29] image fusion,[30,31] and three-dimensional visualization of images in medical research. In the process of medical magnetic resonance (MR) image acquisition, various factors, such as imaging equipment, imaging techniques, external interference, and chessboard artifacts in boundary models and human tissue images, lead to low-resolution images and interfere with the accuracy of clinical diagnosis and subsequent medical research. Therefore, it is of great significance to use deep learning and other approaches to restore MR images with clear tissue boundaries and rich details. With the rapid development of deep learning and computer hardware, computer vision tasks have attracted increasing attention from the academic community, and researchers have begun to explore the application of CNNs to such tasks. Since Tian *et al*.[32] initially applied deep convolutional neural networks to image super-resolution, the quality of image super-resolution reconstruction algorithms based on these networks has significantly improved.

Moreover, Chao *et al*.[33] proposed an enhanced deep super-resolution network for single images, based on deep residual networks, to address the issue of shallow convolutional neural networks not being able to fully extract contextual feature information from images. By removing batch normalization from residual modules, they were able to stack more convolutional layers using the same computational resources, allowing the network to learn more contextual feature information. Given that the quality of reconstructed images can be improved by increasing the depth of convolutional neural networks in image super-resolution reconstruction task, Yang *et al*. successively proposed a deeply-recursive low- and high-frequency fusing network and a precise image super-resolution method based on the depth convolutional neural network, the very-deep super-resolution network.[34,35] While further improv-

ing algorithm performance by increasing network depth, various studies on image super-resolution based on convolutional neural networks have encountered issues such as vanishing and exploding gradients during training. Cui *et al*.[36] applied residual learning to computer vision tasks and proposed the image processing network ResLT. Building upon research on residual network for image processing, Ledig *et al*.[26] proposed the super-resolution residual network, which utilizes the concept of residual learning. This approach avoids the loss of contextual information during image propagation through the network, addressing the gradient vanishing and exploding problems caused by increased network depth. Moreover, the super-resolution residual network shows improvement in preserving details in reconstructed images.

The application of attention mechanisms in image super-resolution enhances the accuracy of reconstructed images to a certain extent. Zhang *et al*.[37] developed the residual dense network by combining residual learning and further proposed an image super-resolution network. Deep residual channel attention networks,[38] based on the channel attention mechanism, are illustrated in Figure 3. Du *et al*.[39] achieved a reduction in network parameter count and an improvement in MR image super-resolution reconstruction quality by using depth-wise separable convolutions instead of traditional convolutional layers. However, the aforementioned image super-resolution reconstruction algorithms based on convolutional neural networks prioritize higher objective performance metrics but overlook perceptual image quality, resulting in problems such as artifacts and blurriness in the reconstructed super-resolution images.

Compared with conventional super-resolution algorithms based on interpolation theory, image super-resolution algorithms based on convolutional neural networks show remarkable enhancements in network performance and image reconstruction effects.[40] However, due to the intrinsic limitations of image super-resolution, convolutional neural networks often encounter challenges such as checkerboard artifacts and loss of image details when reconstructing super-resolution images with high upscaling factors.[41] This is

mainly because, in convolutional neural networks, when the stride of the convolutional kernel is not equal to 0, it introduces interference in computer vision research based on image super-resolution tasks. Li *et al*.[42] introduced an image super-resolution algorithm based on generative adversarial networks to solve the image super-resolution problem, thus further promoting the development of image super-resolution research. In order to solve the problem of image blur caused by detail loss in medical image super-resolution models based on traditional convolutional neural networks, Inspired by Tran *et al*.,[43] who introduced the disentangled representation learning-generative adversarial network (DR-GAN) for pose-invariant face recognition, subsequent studies have extended similar adversarial and disentanglement principles to image super-resolution tasks. These approaches aim to mitigate blurred edges and detail loss typically observed in CNN-based SR methods. Building upon the residual concept, Zhao *et al*.[44] further proposed the laplacian pyramid generative adversarial network based on dense residual blocks. This approach effectively addresses blurriness and size inconsistency in reconstructing medical images using existing image super-resolution algorithms. Wang *et al*.[45] proposed the enhanced super-resolution generative adversarial network (ESRGAN) by introducing the residual-in-residual dense block on top of residual blocks in the network model, which significantly improves the performance of generative adversarial network-based image super-resolution algorithms. By incorporating perceptual loss, adversarial loss, and a relative discriminator, the discriminator network assesses the relative authenticity of reconstructed images compared to traditional discriminator networks.[46] This assessment guides the generator network to produce more realistic images through parameter updates during training. Shang *et al*.[47] introduced the receptive field block ESRGAN, based on the enhanced generative adversarial network super-resolution algorithm, which features receptive field modules. These modules, with different sizes of receptive fields, enable the network to extract richer image detail features, thus enhancing the quality of reconstructed images. While existing image super-resolution methods based on generative adversarial networks have improved the overall visual quality of images in practical applications, they often introduce unnatural artifacts when reconstructing image details. Geets *et al*.[48] proposed a method based on the statistical dependencies of image gradients and edges at different resolutions. Sun *et al*.[49] presented a method based on gradient contours representing image gradients and gradient field transformations. Yan *et al*.[50] introduced an image super-resolution algorithm based on gradient contour sharpness to improve the clarity of super-resolved images. In these methods, the statistical dependency relationship is modeled by estimating parameters related to high-resolution edges based on parameters learned from low-resolution images.[50] Ma *et al*.[51] proposed a structure-preserved super-resolution algorithm based on gradient guidance. This approach employs second-order gradient constraints in deep generative adversarial networks to provide better structural guidance for image reconstruction, effectively addressing issues such as structural distortion in reconstructed images.[51] Compared with super-resolution reconstruction algorithms using convolutional neural networks, medical image super-resolution reconstruction algorithms based on deep generative adversarial networks achieve higher accuracy in restoring edge details and texture information in reconstructed images, making the visual effects of reconstructed images more suitable for clinical diagnostic needs. However, ordinary convolutional neural networks exhibit translational invariance in convolutional kernels, causing the loss of shallow and local features in images as the network depth

increases, resulting in blurred edges and potential checkerboard artifacts in reconstructed images. In contrast, image super-resolution reconstruction algorithms based on deep generative adversarial networks update generator and discriminator network parameters through backpropagation, guiding the generated sample values toward more realistic values.

Status of image registration algorithm research

Currently, image registration algorithms can be divided into non-learning-based methods and learning-based methods. Traditional non-learning-based registration methods are mainly feature-based registration algorithms. The following provides a detailed description of their research status.[52] Feature-based registration algorithms first extract features from the reference image and the floating image, generally including feature points, image edges, image structures, and statistical features. Then, through a matching strategy, they establish correspondences between features and calculate the deformation parameters of the image pairs through feature matching. Specifically, feature-based image registration algorithms involve the following steps[53]:

Feature extraction

Feature extraction is a pivotal task in the image registration process. It can be either manual or automatic, depending on the image complexity. Features such as closed boundary regions, textures, edges, points, lines, statistical features, and more advanced structures and semantic descriptions can serve as distinctive characteristics. These features must be easily identifiable and invariant to ensure that both the reference and floating images share sufficient common features. Robust algorithms are required for feature detection to extract as many features as possible from image pairs, irrespective of structural deformations.

Feature matching

The goal of this step is to establish precise correspondences between features, creating a matching method between the features of the reference and floating images. Various feature descriptors and similarity measures are employed to facilitate accurate feature correspondence. Feature descriptor designs should ensure the accurate reflection of global or local image characteristics, even in the presence of noise.

Transformation model estimation

Registration transformation models encompass rigid and non-rigid transformations. The selection of a transformation model depends on the image acquisition process and prior knowledge of expected image deformations. To align the reference and floating images, the deformation parameters of the transformation model must be estimated using feature correspondences.

Image resampling

Resampling of the floating image is performed using the estimated optimal deformation parameters. Following the transformation of image coordinates, the resulting position coordinates are typically non-integer values. Thus, interpolation operations are commonly employed for image resampling to address this issue.

In the context of medical image registration, Al-Khafaji *et al*.[54] proposed the scale-invariant feature transform (SIFT) algorithm, which has been widely applied. SIFT, an early algorithm for keypoint detection, ensures invariance to translation, rotation, and other transformations. It requires extracting many point features, resulting in high computational complexity. To accelerate SIFT

feature computation, Bay *et al*.[55] applied an improved algorithm, Speeded-Up Robust Features (SURF). SURF is more stable and computationally efficient than SIFT. Apart from SIFT and SURF, various other feature description operators have been utilized in image registration,[56] such as Harris corners.[57] Shen *et al*.[58] proposed the hierarchical attribute matching mechanism for elastic registration algorithm, which extracts a set of geometric moment invariants for each image point. Experimental results have demonstrated its effectiveness in registering brain images with significant anatomical differences. However, the hierarchical attribute matching mechanism for elastic registration algorithm requires pre-segmentation of brain tissues, which poses a challenge and limits its applicability. To overcome this limitation, Papamarkos *et al*.[59] proposed an approach that achieves gray-level reduction through the combined utilization of both the image's gray levels and additional local spatial features. These histogram-based attribute vectors exhibit rotation invariance and have been successfully applied to register various data, including brain MRI and diffusion tensor imaging. Nonetheless, registration results are significantly impacted by the accuracy of feature extraction. Inaccurate feature extraction can lead to substantial registration errors. Therefore, research on these algorithms primarily focuses on feature design.

With the rapid advancement of deep learning in computer vision and other fields, there has been an abundance of registration algorithms based on deep learning, with CNN playing a significant role in medical image registration. Early deep learning registration methods primarily focused on using deep learning to extract features from reference and floating images,[60] or to learn similarity metrics for image pairs.[61] These learned features and similarity metrics were then integrated into traditional registration frameworks to significantly improve registration effectiveness. Yoo *et al*.[62,63] utilized a stacked convolutional autoencoder to extract features from pairs of 3D brain MR images, followed by optimizing the normalized cross correlation between the two sets of features using gradient descent. The experiment indicated that in single-modality registration, the feature descriptors extracted by deep learning might not surpass manually defined descriptors, but they could be used to obtain complementary information. Additionally, drawing on the registration experience of CT images, Zhu *et al*.[64] used a CNN to estimate the registration error of chest CT-MRI images and employed the learned registration error as a similarity metric for subsequent registration (Fig. 4).[65]

The advantage of deep learning in grayscale-based registration is particularly evident in multi-modal scenarios, where designing effective multi-modal similarity metrics is challenging. Andrade *et al*.[66] designed a stacked denoising autoencoder to learn a similarity metric for CT and MR images to achieve rigid registration. The multi-modal similarity metric learned by the model outperformed traditional NMI. These methods break the limitations of manually designed prior knowledge, effectively improving registration performance while retaining the iterative nature of traditional registration. However, these deep learning methods have not fundamentally solved the problem of slow registration speed due to iterative optimization. Therefore, more and more researchers are focusing on directly estimating deformation parameters using convolutional neural networks (ConvNets). Sankareswaran *et al*.[67] used ConvNets to learn rigid transformation parameters, showing significant advantages in registration accuracy and real-time performance compared to grayscale-based methods. Huang *et al*.[68] trained ConvNets to directly estimate the displacement vector field of image pairs, achieving the same accuracy as traditional registration methods. Yan *et al*.[69] proposed the adversarial image registra-

tion framework, inspired by generative adversarial networks, for rigid registration of 3D MR and transrectal ultrasound (TRUS) image pairs. The generator estimates the deformation parameters of the image pair, while the discriminator distinguishes between real and predicted deformation images. The network is trained using an adversarial supervision strategy. These methods demonstrate good registration performance but require labeled training data. Typically, traditional registration methods are used to obtain deformation parameters or synthetic supervised training data is created using random deformation parameters. It can be seen that the performance of such supervised methods largely depends on the reliability of the labels, which has driven the development of semi-supervised registration models. Haskins *et al*.[70] used label similarity to train a CNN model for MR-TRUS image registration. In their initial registration scheme, two network models were used to train global affine transformation parameters and local dense deformation fields. The result of global registration was used as input for local registration, achieving coarse-to-fine registration. However, to further improve the practicality of the model, in subsequent work, they combined the two parts of the network into an end-to-end framework, achieving end-to-end registration using CNNs. In another work, Saldanha *et al*.[71] introduced a loss function based on label similarity and image grayscale similarity metrics on the basis of double supervision and weak supervision using breast phantoms, ultimately achieving deformable registration of 2D MR images, using both segmentation overlap distance and edge-based normalized gradient field distance to construct the loss function.

Blendowski *et al*.[72] introduced an integrated multimodal registration method guided by a shape encoder-decoder network. First, a segmentation network is trained with anatomically labeled data, and then an energy-based iterative optimization method is used to estimate the deformation parameters between image pairs. This method relies on intermediate segmentation results but can simplify the registration of CT images and MR images in cases of large deformations.[72] Similarly, Fu *et al*.[73] used a Laplacian pyramid network with anatomical label supervision to overcome large structural differences between image pairs and used data augmentation to mitigate overfitting. These semi-supervised methods reduce the model's reliance on labeled data but are still largely influenced by real labels. Therefore, many researchers are focusing on the study of unsupervised registration models. Especially since the emergence of spatial transformer networks, a large number of spatial transformer network-based image registration models have emerged.[74] de Vos *et al*.[75] proposed the unsupervised deformable registration model DIRNet, which first applies a ConvNet regressor to 2D control points and then uses cubic B-splines as a spatial transformer to output the displacement vector field of image pairs, followed by a resampler to achieve deformation of floating images. Similarly, Ji *et al*.[76] developed an unsupervised end-to-end brain MRI image registration framework ADMIR (affine and deformable medical image registration), which includes affine registration and non-linear registration. When the sizes of the reference image and the floating image vary, pre-registration is usually required, while ADMIR can complete end-to-end registration, effectively saving registration time. However, this method cannot adapt to images of arbitrary sizes. When using this model for registration, the image size needs to be consistent with the size of the model training set. Balakrishnan *et al*.[77] built the VoxelMorph model, achieving nonlinear registration of brain MRI images, with results superior to the SyN algorithm in terms of Dice score. Although VoxelMorph can accurately estimate dense vector fields of image pairs, later work has shown that the model performs poor-
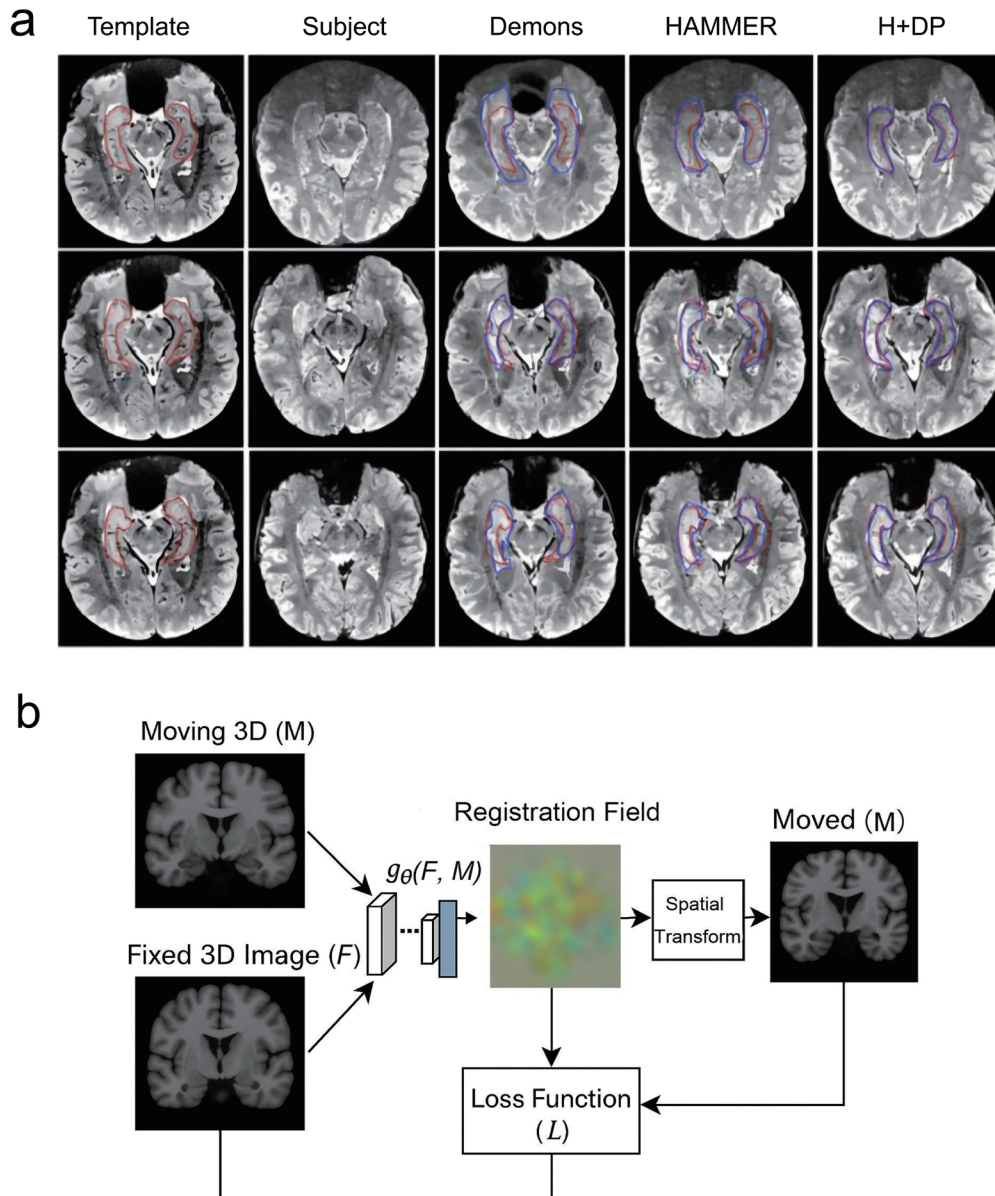
Fig. 4. Image registration section. (a) Registration results on 7.0-T magnetic resonance (MR) brain images by Demons, hierarchical attribute matching mechanism for elastic registration (HAMMER), and H+DP.[65] (b) Deep learning-based image registration workflow.

ly for heart CT data. Zhao *et al.*[78] used VoxelMorph as the base network and proposed a recursive cascade registration network, which can reduce the number of network parameters and improve the registration speed through weight sharing during the testing phase. However, this method has difficulty maintaining smooth deformation fields during the recursive process. The works are all based on high-dimensional image space for network design. These algorithms based on deep unsupervised deformation parameter estimation do not require labeled data, reducing the requirements for data, but unsupervised registration requires selecting appropriate image similarity metrics as optimization targets. These similarity metrics are mostly based on global grayscale metrics, performing well in overall structural registration, but it is hard to accurately estimate local deformations. Additionally, the grayscale and texture

information of multi-modal medical images differ greatly. After extracting image features based on deep convolution, how to select appropriate features from significantly different features to quantify the similarity between reference images and floating images has become a future challenge in multi-modal image registration.

Application prospect

Deep learning has great potential in reconstructing high-resolution BOLD images from low-resolution ones.[79] By learning and analyzing many low-resolution BOLD images and their corresponding high-resolution images, deep learning can automatically extract features and build models to achieve the reconstruction from low-resolution to high-resolution.

Improving spatial resolution

Deep learning can infer the missing high-frequency information by learning the association between low-resolution and high-resolution BOLD images, thus improving the spatial resolution of the images. This will be conducive to more accurately locating and analyzing areas of brain activity.

Ameliorating quantitative analysis of brain activity

High-resolution BOLD images provide more detailed and precise information about brain activity and aid researchers in better understanding and analyzing brain function, while deep learning improves the quantitative analysis of brain activity and supplies more accurate results by reconstructing high-resolution BOLD images.

Enhancing brain network connectivity analysis

Brain network connectivity is one method for studying brain function and diseases. By reconstructing high-resolution BOLD images, deep learning provides more accurate and reliable brain network connectivity analysis, helping to reveal the structure and function of brain networks.

Assisting clinical diagnosis

High-resolution BOLD images can furnish more detailed information about brain structure and function, thus having a great impact on clinical diagnosis. By reconstructing high-resolution BOLD images, deep learning can improve accuracy and efficiency in diagnosis and assist in treatment decisions for brain diseases.

The limitations of high-resolution BOLD images

Despite advancements in deep learning for high-resolution BOLD-fMRI analysis, several interconnected limitations hinder clinical translation. Technical barriers include high computational costs tied to GPU/TPU dependencies and performance variability across heterogeneous scanner environments (e.g., 1.5T vs. 3T systems, protocol differences), which compromise model generalizability. Data constraints further exacerbate these challenges, as acquiring large-scale, annotated, and demographically diverse datasets remains hindered by privacy regulations, ethical concerns, and inconsistent imaging standards, risking biased performance in underrepresented populations. Model trustworthiness is undermined by the "black-box" nature of many architectures, limiting interpretability crucial for clinical adoption in neurological decision-making, while reliance on retrospective validation raises questions about real-world reliability. Finally, clinical integration faces infrastructural mismatches (e.g., incompatible software/hardware), regulatory complexities, and workflow disruptions, necessitating not only algorithmic improvements but also systemic collaboration across technical, clinical, and policy domains to align innovations with practical healthcare needs.

Conclusions

This review underscores the transformative potential of deep learning in enhancing high-resolution BOLD-fMRI reconstruction through super-resolution, segmentation, and registration methods. The integration of convolutional neural networks, transformer-based models, and generative adversarial networks has demonstrated significant improvements in image quality, accuracy, and clinical applicability. These advancements facilitate detailed analysis of brain function, improve diagnostic accuracy, and enable more precise clinical interventions. Nonetheless, practical implementation remains challenged by high computational costs, limited generalizability, and data heterogeneity. Future research efforts should prioritize the development of standardized protocols, extensive multi-center validations, and strategies to enhance model interpretability and computational efficiency, ultimately bridging the gap between technical innovation and clinical practice.

Conflict of interest

None.

Author contributions

Study conception, framework design (YLi), drafting of the manuscript, initial literature search, data extraction (YLiu), content related to super-resolution techniques, critical insights (ZZ), substantial input on image registration methods (TW), supervision, review of the manuscript, guidance and final approval of the manuscript (HL), editing (TW, HL), and revision of the manuscript (ZZ, TW). All authors have reviewed and approved the final version and publication of the manuscript.

References

[1] Yablonskiy DA, Sukstanskii AL, He X. Blood oxygenation level-dependent (BOLD)-based techniques for the quantification of brain hemodynamic and metabolic properties - theoretical models and experimental approaches. NMR Biomed 2013;26(8):963–986. doi:10.1002/nbm.2839, PMID:22927123.

[2] Bollmann S, Barth M. New acquisition techniques and their prospects for the achievable resolution of fMRI. Prog Neurobiol 2021;207:101936. doi:10.1016/j.pneurobio.2020.101936, PMID:33130229.

[3] Celard P, Iglesias EL, Sorribes-Fdez JM, Romero R, Vieira AS, Borrajo L. A survey on deep learning applied to medical images: from simple artificial neural networks to generative models. Neural Comput Appl 2023;35(3):2291–2323. doi:10.1007/s00521-022-07953-4, PMID:36373133.

[4] Nazir N, Sarwar A, Saini BS. Recent developments in denoising medical images using deep learning: An overview of models, techniques, and challenges. Micron 2024;180:103615. doi:10.1016/j.micron.2024.103615, PMID:38471391.

[5] Zhang L, Wang M, Liu M, Zhang D. A Survey on Deep Learning for Neuroimaging-Based Brain Disorder Analysis. Front Neurosci 2020;14:779. doi:10.3389/fnins.2020.00779, PMID:33117114.

[6] Conte GM, Weston AD, Vogelsang DC, Philbrick KA, Cai JC, Barbera M, *et al*. Generative Adversarial Networks to Synthesize Missing T1 and FLAIR MRI Sequences for Use in a Multisequence Brain Tumor Segmentation Model. Radiology 2021;299(2):313–323. doi:10.1148/

radiol.2021203786, PMID:33687284.

[7] Zhang H, Yang X, Cui Y, Wang Q, Zhao J, Li D. A novel GAN-based three-axis mutually supervised super-resolution reconstruction method for rectal cancer MR image. Comput Methods Programs Biomed 2024;257:108426. doi:10.1016/j.cmpb.2024.108426, PMID:39368440.

[8] Zhou Z, Ma A, Feng Q, Wang R, Cheng L, Chen X, *et al*. Super-resolution of brain tumor MRI images based on deep learning. J Appl Clin Med Phys 2022;23(11):e13758. doi:10.1002/acm2.13758, PMID:36107021.

[9] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds). Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer; 2015:234–241. doi:10.1007/978-3-319-24574-4_28.

[10] Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV); 2016 Oct 25-28; Stanford (CA), USA. Piscataway (NJ): IEEE; 2016. p. 565-571. doi:10.1109/3DV.2016.79.

[11] Rajamani KT, Rani P, Siebert H, ElagiriRamalingam R, Heinrich MP. Attention-augmented U-Net (AA-U-Net) for semantic segmentation. Signal Image Video Process 2023;17(4):981–989. doi:10.1007/s11760-022-02302-3, PMID:35910403.

[12] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell 2018;40(4):834–848. doi:10.1109/TPAMI.2017.2699184, PMID:28463186.

[13] Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds). Computer Vision – ECCV 2018. Cham: Springer; 2018:833–851. doi:10.1007/978-3-030-01234-2_49.

[14] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18(2):203–211. doi:10.1038/s41592-020-01008-z, PMID:33288961.

[15] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W (eds). Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Cham: Springer; 2016:424–432. doi:10.1007/978-3-319-46723-8_4.

[16] Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, *et al*. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky L, Michaeli T, Nishino K (eds). Computer Vision – ECCV 2022 Workshops. ECCV 2022. Cham: Springer; 2023:205–218. doi:10.1007/978-3-031-25066-8_9.

[17] Al Qurri A, Almekkawy M. Improved UNet with Attention for Medical Image Segmentation. Sensors (Basel) 2023;23(20):8589. doi:10.3390/s23208589, PMID:37896682.

[18] Zhao W, Su X, Guo Y, Li H, Basnet S, Chen J, *et al*. Deep learning based ultrasonic visualization of distal humeral cartilage for image-guided therapy: a pilot validation study. Quant Imaging Med Surg 2023;13(8):5306–5320. doi:10.21037/qims-23-9, PMID:37581069.

[19] Yang T, Zhu G, Cai L, Yeo JH, Mao Y, Yang J. A benchmark study of convolutional neural networks in fully automatic segmentation of aortic root. Front Bioeng Biotechnol 2023;11:1171868. doi:10.3389/fbioe.2023.1171868, PMID:37397959.

[20] Fang L, Zhuo H, Li S. Super-resolution of hyperspectral image via superpixel-based sparse representation. Neurocomputing 2018;273:171–177. doi:10.1016/j.neucom.2017.08.019.

[21] Reichenbach JR, Haacke EM. High-resolution BOLD venographic imaging: a window into brain function. NMR Biomed 2001;14(7-8):453–467. doi:10.1002/nbm.722, PMID:11746938.

[22] Zou WW, Yuen PC. Very low resolution face recognition problem. IEEE Trans Image Process 2012;21(1):327–340. doi:10.1109/TIP.2011.2162423, PMID:21775262.

[23] Feng R, Lei B, Wang W, Chen T, Chen J, Chen DZ, *et al*. SSN: A Stair-Shape Network for Real-Time Polyp Segmentation in Colonoscopy Images. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020 April 03-07; Iowa City (IA), USA. Piscataway (NJ): IEEE; 2020. p. 225-229. doi:10.1109/ISBI45749.2020.9098492.

[24] Lin Z, Shum HY. Fundamental limits of reconstruction-based super-resolution algorithms under local translation. IEEE Trans Pattern Anal Mach Intell 2004;26(1):83–97. doi:10.1109/tpami.2004.1261081, PMID:15382688.

[25] Elsaid NMH, Wu YC. Super-Resolution Diffusion Tensor Imaging using SRCNN: A Feasibility Study(). Annu Int Conf IEEE Eng Med Biol Soc 2019;2019:2830–2834. doi:10.1109/EMBC.2019.8857125, PMID:31946482.

[26] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, *et al*. ROSE: Multi-level super-resolution-oriented semantic embedding for 3D microvasculature segmentation from low-resolution images. Neurocomputing 2024;599:128038. doi:10.1109/CVPR.2017.19.

[27] Wang Y, Zhu H, Li H, Yan G, Buch G, Wang Y, *et al*. ROSE: Multi-level super-resolution-oriented semantic embedding for 3D microvasculature segmentation from low-resolution images. Neurocomputing 2024;599:128038. doi:10.1016/j.neucom.2024.128038.

[28] de Farias EC, di Noia C, Han C, Sala E, Castelli M, Rundo L. Impact of GAN-based lesion-focused medical image super-resolution on the robustness of radiomic features. Sci Rep 2021;11(1):21361. doi:10.1038/s41598-021-00898-z, PMID:34725417.

[29] Favorskaya MN, Jain LC, Nishchhal. Pancreatic Cancer Classification Using Multimodal Imaging. In: Lim CP, Vaidya A, Jain N, Mahorkar U, Jain LC (eds). Advances in Intelligent Disease Diagnosis and Treatment. Cham: Springer; 2024:13–34. doi:10.1007/978-3-031-65640-8_2.

[30] Du J, Li W, Lu K, Xiao B. An overview of multi-modal medical image fusion. Neurocomputing 2016;215:3–20. doi:10.1016/j.neucom.2015.07.160.

[31] Das M, Gupta D, Radeva P, Bakde AM. NSST domain CT-MR neurological image fusion using optimised biologically inspired neural network. IET Image Process 2021;14(16):4291–4305. doi:10.1049/iet-ipr.2020.0219.

[32] Tian C, Yuan Y, Zhang S, Lin CW, Zuo W, Zhang D. Image super-resolution with an enhanced group convolutional neural network. Neural Netw 2022;153:373–385. doi:10.1016/j.neunet.2022.06.009, PMID:35779445.

[33] Chao HS, Wu YH, Siana L, Chen YM. Generating High-Resolution CT Slices from Two Image Series Using Deep-Learning-Based Resolution Enhancement Methods. Diagnostics (Basel) 2022;12(11):2725. doi:10.3390/diagnostics12112725, PMID:36359568.

[34] Yang C, Lu G. Deeply Recursive Low- and High-Frequency Fusing Networks for Single Image Super-Resolution. Sensors (Basel) 2020;20(24):7268. doi:10.3390/s20247268, PMID:33352901.

[35] Hwang JJ, Jung YH, Cho BH, Heo MS. Very deep super-resolution for efficient cone-beam computed tomographic image restoration. Imaging Sci Dent 2020;50(4):331–337. doi:10.5624/isd.2020.50.4.331, PMID:33409142.

[36] Cui J, Liu S, Tian Z, Zhong Z, Jia J. ResLT: Residual Learning for Long-Tailed Recognition. IEEE Trans Pattern Anal Mach Intell 2023;45(3):3695–3706. doi:10.1109/TPAMI.2022.3174892, PMID:35560104.

[37] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018 June 18-23; Salt Lake City (UT), USA. Piscataway (NJ): IEEE; 2018. p. 2472-2481. doi:10.1109/CVPR.2018.00262.

[38] Qi Y, Gu J, Li W, Tian Z, Zhang Y, Geng J. Pulmonary nodule image super-resolution using multi-scale deep residual channel attention network with joint optimization. J Supercomput 2020;76:1005–1019. doi:10.1007/s11227-019-03066-3.

[39] Du J, He Z, Wang L, Gholipour A, Zhou Z, Chen D, *et al*. Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network. Neurocomputing 2020;392:209–220. doi:10.1016/j.neucom.2018.10.102.

[40] Park S, Gach HM, Kim S, Lee SJ, Motai Y. Autoencoder-Inspired Convolutional Network-Based Super-Resolution Method in MRI. IEEE J Transl Eng Health Med 2021;9:1800113. doi:10.1109/JTEHM.2021.3076152, PMID:34168920.

[41] Sugawara Y, Shiota S, Kiya H. Checkerboard artifacts free convolutional neural networks. APSIPA Trans Signal Inf Process 2019;8(1):e9.

doi:10.1017/ATSIP.2019.2.

[42] Li C, Xu K, Zhu J, Liu J, Zhang B. Triple Generative Adversarial Networks. IEEE Trans Pattern Anal Mach Intell 2022;44(12):9629–9640. doi:10.1109/TPAMI.2021.3127558, PMID:34767502.

[43] Tran L, Yin X, Liu X. Representation Learning by Rotating Your Faces. IEEE Trans Pattern Anal Mach Intell 2019;41(12):3007–3021. doi:10.1109/TPAMI.2018.2868350, PMID:30183620.

[44] Zhao M, Liu X, Liu H, Wong KKL. Super-resolution of cardiac magnetic resonance images using Laplacian Pyramid based on Generative Adversarial Networks. Comput Med Imaging Graph 2020;80:101698. doi:10.1016/j.compmedimag.2020.101698, PMID:31935666.

[45] Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, *et al*. Esrgan: Enhanced super-resolution generative adversarial networks. In: Leal-Taixé L, Roth S (eds). Computer Vision – ECCV 2018 Workshops. Cham: Springer; 2018. doi:10.1007/978-3-030-11021-5_5.

[46] Wang C, Xu C, Wanga C, Tao D. Perceptual Adversarial Networks for Image-to-Image Transformation. IEEE Trans Image Process 2018;27(8):4066–4079. doi:10.1109/TIP.2018.2836316, PMID:29993743.

[47] Shang T, Dai Q, Zhu S, Yang T, Guo Y. Perceptual extreme super-resolution network with receptive field block. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); June 14-19 2020; Seattle (WA), USA. Piscataway (NJ): IEEE; 2020. p. 1778-1787. doi:10.1109/CVPRW50498.2020.00228.2020.

[48] Geets X, Lee JA, Bol A, Lonneux M, Grégoire V. A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur J Nucl Med Mol Imaging 2007;34(9):1427–1438. doi:10.1007/s00259-006-0363-4, PMID:17431616.

[49] Sun J, Sun J, Xu Z, Shum HY. Gradient profile prior and its applications in image super-resolution and enhancement. IEEE Trans Image Process 2011;20(6):1529–1542. doi:10.1109/TIP.2010.2095871, PMID:21118774.

[50] Yan Q, Xu Y, Yang X, Nguyen TQ. Single image superresolution based on gradient profile sharpness. IEEE Trans Image Process 2015;24(10):3187–3202. doi:10.1109/TIP.2015.2414877, PMID:25807567.

[51] Ma C, Rao Y, Lu J, Zhou J. Structure-Preserving Image Super-Resolution. IEEE Trans Pattern Anal Mach Intell 2022;44(11):7898–7911. doi:10.1109/TPAMI.2021.3114428, PMID:34550879.

[52] Wang J, Yang X. Non-Deep-Learning-Based Medical Image Synthesis Methods. In: Yang X (ed). Medical Image Synthesis. 1st ed. Boca Raton: CRC Press; 2023:3–13. doi:10.1201/9781003243458-2.

[53] Wang HJ, Lee CY, Lai JH, Chang YC, Chen CM. Image registration method using representative feature detection and iterative coherent spatial mapping for infrared medical images with flat regions. Sci Rep 2022;12(1):7932. doi:10.1038/s41598-022-11379-2, PMID:35562370.

[54] Al-Khafaji SL, Zhou J, Zia A, Liew AW. Spectral-Spatial Scale Invariant Feature Transform for Hyperspectral Images. IEEE Trans Image Process 2018;27(2):837–850. doi:10.1109/TIP.2017.2749145, PMID:28880176.

[55] Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (SURF). Comput Vis Image Underst 2008;110(3):346–359. doi:10.1016/j.cviu.2007.09.014.

[56] Zhang W, Zhao Y. An improved SIFT algorithm for registration between SAR and optical images. Sci Rep 2023;13(1):6346. doi:10.1038/s41598-023-33532-1, PMID:37072451.

[57] Pei Y, Wu H, Yu J, Cai G. Effective image registration based on improved harris corner detection. In: 2010 International Conference on Information, Networking and Automation (ICINA); 2010 Oct 18-19; Kunming, China. Piscataway (NJ): IEEE; 2010. p. V1-93-V1-96. doi:10.1109/ICINA.2010.5636429.

[58] Shen D, Davatzikos C. HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans Med Imaging 2002;21(11):1421–1439. doi:10.1109/TMI.2002.803111, PMID:12575879.

[59] Papamarkos N, Atsalakis A. Gray-level reduction using local spatial features. Computer Vision and Image Understanding 2000;78(3):336–350. doi:10.1006/cviu.2000.0838.

[60] Ofverstedt J, Lindblad J, Sladoje N. Fast and Robust Symmetric Image Registration Based on Distances Combining Intensity and Spatial Information. IEEE Trans Image Process 2019. doi:10.1109/TIP.2019.2899947, PMID:30794174.

[61] Siebert H, Hansen L, Heinrich MP. Learning a Metric for Multimodal Medical Image Registration without Supervision Based on Cycle Constraints. Sensors (Basel) 2022;22(3):1107. doi:10.3390/s22031107, PMID:35161851.

[62] Yoo I, Hildebrand DGC, Tobin WF, Lee WCA, Jeong WK. ssEMnet: Serial-Section Electron Microscopy Image Registration Using a Spatial Transformer Network with Learned Features. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, *et al* (eds). Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2017. Cham: Springer; 2017:249–257. doi:10.1007/978-3-319-67558-9_29.

[63] Yoo IW. Deformable image registration using convolutional neural networks for connectomics [Dissertation]. Ulsan: Ulsan National Institute of Science and Technology (UNIST), Graduate School; 2018.

[64] Zhu F, Wang S, Li D, Li Q. Similarity attention-based CNN for robust 3D medical image registration. Biomed Signal Process Control 2023;81:104403. doi:10.1016/j.bspc.2022.104403.

[65] Wu G, Kim M, Wang Q, Munsell BC, Shen D. Scalable High-Performance Image Registration Framework by Unsupervised Deep Feature Representations Learning. IEEE Trans Biomed Eng 2016;63(7):1505–1516. doi:10.1109/TBME.2015.2496253, PMID:26552069.

[66] Andrade N, Faria FA, Cappabianco FAM. A Practical Review on Medical Image Registration: From Rigid to Deep Learning Based Approaches. In: 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI); 2019 January 17; Parana, Brazil. Piscataway (NJ): IEEE; 2018. p. 463-470. doi:10.1109/SIBGRAPI.2018.00066.

[67] Sankareswaran SP, Krishnan M. Unsupervised End-to-End Brain Tumor Magnetic Resonance Image Registration Using RBCNN: Rigid Transformation, B-Spline Transformation and Convolutional Neural Network. Curr Med Imaging 2022;18(4):387–397. doi:10.2174/1573405617666210806125526, PMID:34365954.

[68] Huang M, Ren G, Zhang S, Zheng Q, Niu H. An Unsupervised 3D Image Registration Network for Brain MRI Deformable Registration. Comput Math Methods Med 2022;2022:9246378. doi:10.1155/2022/9246378, PMID:36226240.

[69] Yan P, Xu S, Rastinehad AR, Wood BJ. Adversarial image registration with application for MR and TRUS image fusion. In: Shi Y, Suk HI, Liu M (eds). Machine Learning in Medical Imaging. MLMI 2018. Cham: Springer; 2018:197–204. doi:10.1007/978-3-030-00919-9_23.

[70] Haskins G, Kruecker J, Kruger U, Xu S, Pinto PA, Wood BJ, *et al*. Learning deep similarity metric for 3D MR-TRUS image registration. Int J Comput Assist Radiol Surg 2019;14(3):417–425. doi:10.1007/s11548-018-1875-7, PMID:30382457.

[71] Saldanha OL, Zhu J, Müller-Franzes G, Carrero ZI, Payne NR, Escudero Sánchez L, *et al*. Swarm learning with weak supervision enables automatic breast cancer detection in magnetic resonance imaging. Commun Med (Lond) 2025;5(1):38. doi:10.1038/s43856-024-00722-5, PMID:39915630.

[72] Blendowski M, Bouteldja N, Heinrich MP. Multimodal 3D medical image registration guided by shape encoder-decoder networks. Int J Comput Assist Radiol Surg 2020;15(2):269–276. doi:10.1007/s11548-019-02089-8, PMID:31741286.

[73] Fu J, Li W, Du J, Xiao B. Multimodal medical image fusion via laplacian pyramid and convolutional neural network reconstruction with local gradient energy strategy. Comput Biol Med 2020;126:104048. doi:10.1016/j.compbiomed.2020.104048, PMID:33068809.

[74] Upendra RR, Simon R, Linte CA. Joint Deep Learning Framework for Image Registration and Segmentation of Late Gadolinium Enhanced MRI and Cine Cardiac MRI. Proc SPIE Int Soc Opt Eng 2021;11598:115980F. doi:10.1117/12.2581386, PMID:34079155.

[75] de Vos BD, Berendsen FF, Viergever MA, Staring M, Išgum I. End-to-End Unsupervised Deformable Image Registration with a Convolutional Neural Network. In: Cardoso MJ, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares JMRS, Moradi M, *et al* (eds). Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer; 2017:204–212. doi:10.1007/978-3-319-67558-9_24.

[76] Ji W, Yang F. Affine medical image registration with fusion feature mapping in local and global. Phys Med Biol 2024;69(5):055029.

doi:10.1088/1361-6560/ad2717, PMID:38324893.

[77] Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Trans Med Imaging 2019. doi:10.1109/TMI.2019.2897538, PMID:307 16034.

[78] Zhao S, Dong Y, Chang E, Xu Y. Recursive Cascaded Networks for Unsupervised Medical Image Registration. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27-Nov 02; Seoul, Korea (South). Piscataway (NJ): IEEE; 2019. p. 10599-10609. doi:10.1109/ICCV.2019.01070.

[79] Yan Y, Dahmani L, Ren J, Shen L, Peng X, Wang R, *et al*. Reconstructing lost BOLD signal in individual participants using deep machine learning. Nat Commun 2020;11(1):5046. doi:10.1038/s41467-020-18823-9, PMID:33028816.