



Deep learning-based automated scan plane positioning for brain magnetic resonance imaging

Gaojie Zhu^{1,2}, Xiongjie Shen², Zhiguo Sun², Zhongjie Xiao², Junjie Zhong², Zhe Yin², Shengxiang Li³, Hua Guo¹

¹Center for Biomedical Imaging Research, School of Biomedical Engineering, Tsinghua University, Beijing, China; ²Anke High-tech Co., Ltd., Shenzhen, China; ³Department of Medical Imaging, Hengyang Central Hospital, Hengyang, China

Contributions: (I) Conception and design: G Zhu, X Shen; (II) Administrative support: G Zhu, H Guo; (III) Provision of study materials or patients: Z Xiao, S Li; (IV) Collection and assembly of data: Z Sun, Z Xiao, S Li; (V) Data analysis and interpretation: G Zhu, X Shen, Z Yin; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Hua Guo, PhD. Center for Biomedical Imaging Research, School of Biomedical Engineering, Tsinghua University, No. 30 Shuangqing Road, Beijing 100084, China. Email: huaguo@tsinghua.edu.cn.

Background: Manual planning of scans in clinical magnetic resonance imaging (MRI) exhibits poor accuracy, lacks consistency, and is time-consuming. Meanwhile, classical automated scan plane positioning methods that rely on certain assumptions are not accurate or stable enough, and are computationally inefficient for practical application scenarios. This study aims to develop and evaluate an effective, reliable, and accurate deep learning-based framework that incorporates prior physical knowledge for automatic head scan plane positioning in MRI.

Methods: A deep learning-based end-to-end automated scan plane positioning framework has been developed for MRI head scans. Our model takes a three-dimensional (3D) pre-scan image input, utilizing a cascaded 3D convolutional neural network to detect anatomical landmarks from coarse to fine. And then, with the determined landmarks, accurate scan plane localization can be achieved. A multi-scale spatial information fusion module was employed to aggregate high- and low-resolution features, combined with physically meaningful point regression loss (PRL) function and direction regression loss (DRL) function. Meanwhile, we simulate complex clinical scenarios to design data augmentation strategies.

Results: Our proposed approach shows good performance on a clinically wide range of 229 MRI head scans, with a point-to-point absolute error (PAE) of 0.872 mm, a point-to-point relative error (PRE) of 0.10%, and an average angular error (AAE) of 0.502°, 0.381°, and 0.675° for the sagittal, transverse, and coronal planes, respectively.

Conclusions: The proposed deep learning-based automated scan plane positioning shows high efficiency, accuracy and robustness when evaluated on varied clinical head MRI scans with differences in positioning, contrast, noise levels and pathologies.

Keywords: Deep learning; magnetic resonance imaging (MRI); automated scan plane positioning; head scan; domain knowledge

Submitted Dec 13, 2023. Accepted for publication Apr 10, 2024. Published online Apr 30, 2024.

doi: 10.21037/qims-23-1740

View this article at: <https://dx.doi.org/10.21037/qims-23-1740>

Introduction

Magnetic resonance imaging (MRI) is an advanced imaging technique based on the principle of nuclear magnetic resonance (NMR), which acquires images by using signals from the spin resonance of atomic nuclei in human tissues (1). Several advantages, such as non-invasiveness, high-resolution and multimodality, which eliminates the need for radiation (2,3), have led to the widespread use of MRI in brain imaging for oncology, cerebrovascular disease, neurological disorders, inflammation and functional MRI. Spatial encoding in MRI is accomplished with gradient waveforms (4), enabling hassle-free arbitrary orientation scans comprising the imaging acquisition, without necessitating movement of the subject or instrument. However, this prescription task requires scanning technicians to accurately set the orientation and position of the scan plane prior to the imaging scan. The precision and consistency of the scan plane positioning has a critical impact on the resultant image quality and associated clinical diagnostic value.

MRI scan plane positioning involves analyzing and identifying the region of interest using pre-scan images. The pre-scan images can be either two-dimensional (2D) or three-dimensional (3D). At the outset of an MRI scan, it is customary to conduct a brief and low-resolution pre-scan of a patient. Subsequently, technicians utilize the pre-scan images to finalize the scan plane positioning by determining the positioning and orientation of the scan planes among other parameters, in light of medical knowledge and scan objectives. *Figure 1* demonstrates the procedure for planning a common MRI head scan using the 3D pre-scan images. As can be seen from the pre-scan images, the patient's position may not always align with the intended diagnostic target. Upon manual configuration by the technician, the vital anatomical regions are precisely captured in the scan. A standard scan in clinical practice involves configuring various planes of target areas, which places significant demands on the anatomical proficiency of technicians. Moreover, manual positioning is laborious and ineffective. In addition, high accuracy of repeat scans of the same patient is critical for comparing anatomical evolution before and after clinical therapy. Furthermore, in the context of technical and clinical research, different institutions and researchers need to ensure reliable comparisons and analyses across different study samples, which requires a high level of consistency in scan positioning. Finally, as intelligent image analysis advances, uniform and highly

consistent scan data has a significant impact on the training and use of deep neural networks (5).

Despite various attempts by researchers to develop and study automated scan plane positioning methods in the head, knee and heart (6-8), there remain numerous challenges to implementing widely applicable automated scan plane positioning methods in a clinical setting. These challenges are threefold. First, various anatomical components of the human body possess distinct anatomical structures and morphological attributes. Such variations introduce more intricate image contents in the scanned images, thereby significantly increasing the complexity of the algorithms for identifying and localizing anatomical landmarks. Second, the individual differences of patients are another key challenge for automated scan plane positioning. The anatomical structure and tissue characteristics of each patient may differ. At the same time, local tissue aberrations, tumors and other abnormalities can significantly affect the overall shape of the target area. Consequently, the robustness and accuracy of automated scan plane positioning algorithms must account for individual patient variations. The automated scan plane positioning algorithms must be able to recognize and adapt to anatomical variations in different patients to ensure accuracy and consistency of localization (9). Finally, even the same anatomical region may have a variety of different application needs for localization in the clinic. Within the head, various applications such as ophthalmic imaging and pituitary imaging may be necessary in addition to regular head scans. These specialized applications involve distinct requirements for scan plane positioning. The automated scan plane positioning algorithms, therefore, must be capable of adjusting to different positioning needs, and flexibly modifying the scanning range and direction to satisfy diverse clinical application requirements.

The task of automated scan plane positioning in medical imaging is generally defined as a landmark recognition problem. By detecting the relevant landmarks in the target anatomical region, the most suitable scan position and orientation for patient scanning can then be determined (10). Classical scan plane positioning methods are primarily based on target detection and image alignment, where the key idea is to compare the pre-scan image with pre-prepared standard anatomical templates and to obtain the localization parameters through an alignment algorithm. van der Kouwe *et al.* (6) determined the probability distributions of the different tissue constituents of the brain by obtaining a 3D pre-scan. Then, they aligned

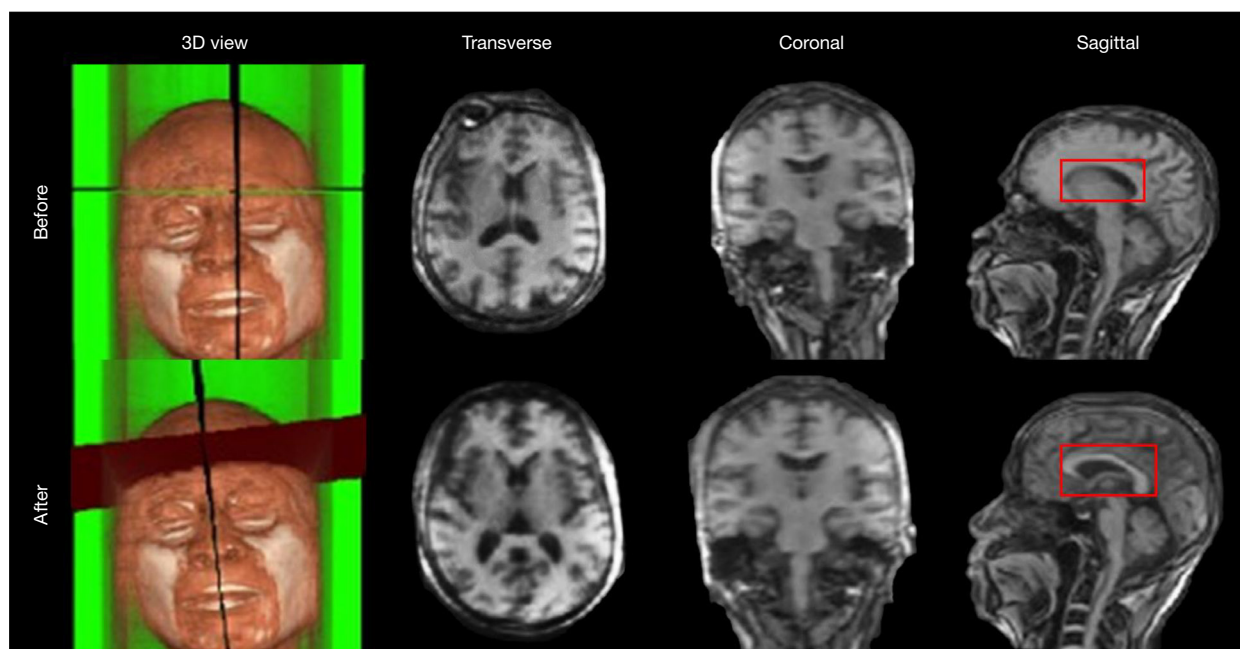


Figure 1 A typical head MRI scan plane positioning process. Based on the 3D pre-scan images (before), the transverse, coronal, and sagittal planes orientation is set by default in the middle of the VOI. Nonetheless, the patient's head is typically not optimally positioned in the VOI center, leading to deviations of the scanning planes from the accurate location. With proper manual scan plane positioning (after), the scan plane can be set with precision. As can be seen in the red box, the corpus callosum is shown to be complete and accurate in the sagittal plane. MRI, magnetic resonance imaging; 3D, three-dimensional; VOI, volume of interest.

these probability distributions with pre-existing brain templates, resulting in the calculation of brain localization parameters. Sharp *et al.* (11) proposed a semi-automatic localization method. Instead of using whole-brain data for the alignment process, the approach selects a series of landmark structures from pre-scan and estimates the automatic locating parameters from the spatial locations of these brain structures. The technique demonstrates high accuracy and stability. However, it necessitates manual extraction of landmark brain structures and demands high anatomical expertise, thus limiting its applicability for efficient clinical scan usage. Lu and colleagues (12) employed a comparable approach to (11) for localizing the cardiac long axis, with an improved localization time of less than 10 seconds for a single patient. Unlike conventional image matching, Nitta *et al.* (13) proposed an a priori knowledge-based approach for detecting anatomical features. They employed a machine learning model to identify landmarks from pre-scan images, which in turn provided the localization planes. This approach has superior practicality and stability compared to template matching. It also surpasses templated matching method by not relying on

template images and adjusting to patient tissue deformation. Nevertheless, the classifier itself can only extract limited spatial contextual information of the landmarks. Moreover, an imbalance in the ratio of positive and negative samples contributes to significant difficulties for the classifier to deal with complicated scenarios. Zhan *et al.* (14) proposed a multi-layer machine learning-based approach that uses redundant information for the automatic localization of magnetic resonance (MR) slice position of the knee. The technique includes the duplicated and stratified anatomical structure data in the machine learning-driven training method via a pre-defined base function and the positioning approach demonstrates remarkable robustness in various experimental scenes.

Recently, deep learning techniques such as convolutional neural networks have made significant progress in the fields of computer vision, speech signal processing, and text processing (15). Medical image processing methods that utilize deep learning techniques have demonstrated exceptional performance in several tasks including medical image segmentation, quantitative estimation of image features, disease risk diagnosis, automatic localization, and

so on (16). Blansit *et al.* (17) presented a technique that employed the U-Net network for landmarks heat map regression to identify the mitral valve, the apical and short-axis planes of the heart in MRI. The mitral, tricuspid, aortic and pulmonary valves were then detected to generate the desired imaging plane. To address the problems posed by densely distributed, obstructed and transformed landmarks, Payer *et al.* (18) introduced a spatial configuration module to the heat map regression network. This module enhances the constraints on the spatial relationships of landmarks, leading to more accurate inference of their locations. Wang and colleagues (19) used a deep metric learning model based on convolutional neural networks to perform an in-depth analysis of cardiac MRI for left ventricle. Recognizing the effectiveness of recurrent convolutional neural networks in extracting sequential features and capturing structural information in image context, van Zon *et al.* (20) unified them to identify mitral and right ventricular landmarks in cardiac MRI. Le *et al.* (21) presented a technique that employs a 3D convolutional neural network for automatically acquiring scanning planes of the heart. The approach uses a 3D convolutional network to process 3D cardiac images, resulting in improved localization and processing efficiency compared to 2D techniques.

In this study, we propose and evaluate an effective, reliable and accurate framework for automatic head scan plane positioning based on state-of-art deep learning landmark detection technique that incorporates a prior physical information into neural networks. The contributions of this paper are as follows. First, we propose a clinical application scheme utilizing 3D pre-scan acquisitions and fusion of medical anatomical knowledge. The 3D image-based scheme efficiently and comprehensively covers the 3D volume of interest, which benefits the consistent identification of anatomical landmarks. It also tackles the issues of missing and damaged data that arise from 2D acquisitions in real-life scenarios. Second, a two-tiered, end-to-end 3D cascaded convolutional network framework is presented to identify the anatomical landmarks and plane directions in a step-by-step manner. Third, we have incorporated a multi-scale spatial information merging module into the network structure, which adaptively merges low-resolution semantic features with high-resolution positional features. Fourth, we propose a loss function for automated scan plane positioning that combines point regression loss (PRL) and direction regression loss (DRL), given the physical background of the application scenarios. Finally, the article tackles the need

for clinical generalization by replicating a range of intricate scenarios observed in the clinic to supplement the training data and improve the neural network's generalization capacity.

Methods

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional review board of Hengyang Central Hospital, China (the registration number of ethics board: No. 17. Dated June 30, 2023), and informed consent was taken from all the patients and volunteers. The paper presents an end-to-end deep learning framework that employs a 3D pre-scan image as input and generates a heat map indicating the location of five clinically predetermined landmarks in the head scan. Our framework incorporates physics knowledge in the neural network structure design, loss function definition and training data augmentation to enhance the model's generalization ability in the context of head scanning application. In the following section, the general structure of the network will be presented initially, followed by the detailed components of the framework individually.

Network architecture

The automated scan plane positioning framework proposed in this paper is the 3D cascade feature pyramid U-Net (3D CFP-U-Net), as presented in *Figure 2*. The structure of the 3D CFP-U-Net is derived from the typical 3D U-Net (22). It has been improved in three key areas: network architecture, feature extraction module and cascade structure. The 3D CFP-U-Net comprises two sets of spaced 3D-U-Net and feature pyramid modules (FPMs) that connect in cascade. In the first stage, "3D-U-net + FPM", the input 3D image roughly locates the landmarks and passes the output feature map to the second stage "3D-U-net + FPM", which confirms the landmarks more accurately. The FPM enhances the network's ability to perceive various levels of image information. The model outputs the coordinates for predetermined landmarks.

The backbone architecture

3D U-Net was developed from U-Net, which shares a similar infrastructure except that all computational operations are extended from 2 to 3 dimensions. This facilitates more efficient processing of voxel data with

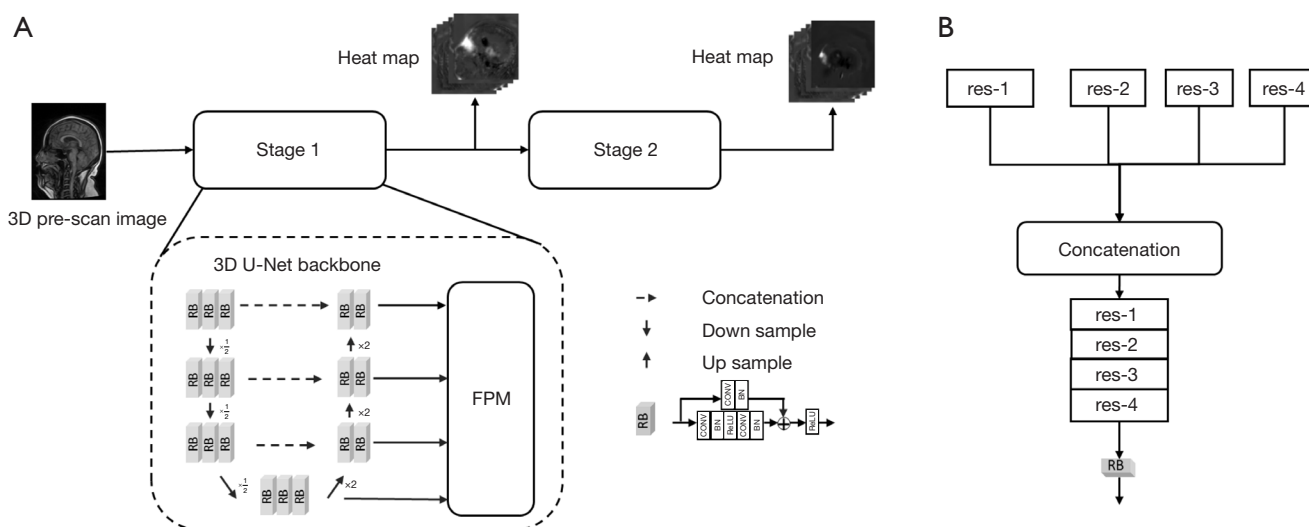


Figure 2 Overview of the proposed 3D CFP-U-Net for automatic scan plane positioning. (A) The general framework of 3D CFP-U-Net which comprises two sets of spaced 3D-U-Net and feature pyramid modules that connect in cascade. Stage 1 and stage 2 share the same network structure. (B) Feature pyramid module, four feature maps with varying resolutions (res-1 to res-4) are combined to provide both identification precision and efficiency. 3D, three-dimensional; RB, residual block; FPM, feature pyramid module; CONV, convolution; BN, Batch normalization; ReLU, rectified linear unit; CFP, cascade feature pyramid.

a 3D structure, such as MRI images. In this paper, we modified the classical 3D U-net to improve its applicability for landmark detection in MRI by implementing several optimization strategies. Firstly, inspired from residual networks (23,24), we construct a more profound 3D U-Net backbone with enhanced feature extraction ability by incorporating additional residual convolutional kernels. As a 3D U-Net backbone in *Figure 2A*, each encoder layer holds three residual blocks with two convolution kernels each, and each decoder layer holds two residual blocks, thereby increasing the depth of a solitary 3D U-Net module by 4–6 times the original depth. Second, the 3D U-Net backbone is designed with an asymmetric structure whereby the encoder comprises more convolutional kernels than the decoder. This design choice enhances the model's feature extraction capability and effectively limits the number of parameters in the model (25). Finally, it is necessary to reduce the number of convolutional kernel channels while increasing network depth to mitigate the risk of overfitting. This is because the number of feature map channels grows exponentially as the feature map resolution decreases during the encoding process. The convolutional kernel base of the 3D U-Net backbone is set to 12 in order to control over the model's parameters, which is approximately 3–4 times fewer than that of the classical 3D U-Net.

FPM

The FPM integrates local and global semantic information to enhance multiscale image representation. It has extensive applications in bolstering the precision and robustness of medical image analysis conducted through computer vision (26). Pre-scan images of MRI at low resolution have a complex and variable background region, with various regions and anatomical structures having similar grayscales and textures. Accurate identification of landmark locations necessitates reliance on fine local information. Therefore, this paper proposes integrating the FPM into the overall network structure, with *Figure 2B* illustrating its basic structure. The paper's FPM upscale four feature maps with distinctive resolutions to a standard size, concatenating them in the channel direction to create a feature pyramid. Following this, multiscale feature fusion is achieved with the aid of two residual blocks. In this module of feature pyramid, four feature maps with varying resolutions are combined. The high-resolution, shallow feature maps capture fine-grained details and local spatial information, which are particularly useful for accurately identifying anatomical landmarks. Meanwhile, the low-resolution feature maps encapsulate global contextual information and larger-scale patterns within the image, providing a broader understanding of the overall brain anatomy. At the same

time, FPM model enables feature fusion, allowing the model to use both high-resolution image detail and contextual information from low-resolution images, resulting in a comprehensive and robust feature representation.

Cascade multi-stage framework

Initially, simple single-stage architecture networks (27) were commonly used for deep learning based landmark detection. However, the emergence of multi-stage network architectures has led to successful applications to a variety of tasks. For instance, Zhong *et al.* (28) proposed a two-stage attention-oriented deep regression model for landmark detection in head and neck X-ray images, which achieved improved detection performance without increasing model complexity. Andermatt *et al.* (29) introduced a two-stage multi-dimensional gated recurrent unit network to localize head and neck MRI in the cerebral pontine cerebellar sulcus. This network achieved better localization errors compared to manual methods. Li *et al.* (30) provided a thorough analysis of cascade structures, presenting a multi-stage cascade network framework for human pose recognition tasks. This framework resulted in significant localization accuracy improvements over single-stage networks. As shown in *Figure 2A*, the 3D CFP-U-Net proposed in this study utilizes a two-stage 3D U-Net cascade composition, employing a coarse-to-fine design approach. In the first stage, the network coarsely defines the location of the landmarks. Subsequently, the feature maps produced in the first stage are fed into the second stage network to precisely determine the coordinates of the target landmarks.

Integrated physical loss function

The suggested loss function in this study comprises two main components: the PRL, typically used in landmark identification problems, and the DRL, which is based on the physical meaning of the scan plane positioning task. Calculated consecutively in two stages of the 3D CFP-U-Net model, the final loss function includes four loss factors that are combined and weighted for supervised training of the network. The general loss function is computed as follows:

$$L_{CFP} = \sum_{i=1}^2 \alpha_i (w_1 L_{PRL} + w_2 L_{DRL}) \quad [1]$$

In which, L_{PRL} and L_{DRL} are the PRL and DRL, correspondingly. The weights for the above two losses

w_1 and w_2 are both arranged as “1” in the experiment. α_i represents the loss weights of each stage in the cascade network, and the first and second stages of the network have empirical weights of 1/3 and 2/3, respectively (29,30).

The PRL denotes the discrepancy between the landmark position predicted by the neural network model and the labelled position. The calculation of the PRL is commonly used in computer vision by using the heat map. The following equation shows the calculation of the PRL:

$$L_{PRL} = -\sum_{n=1}^N (1 - Z_n)^2 * \log(Z_n) \quad [2]$$

Here, N represents the count of landmarks and Z_n signifies the network model's confidence level in a particular landmark. The aim of optimizing the PRL is to increase the overall confidence of the network in all landmark regions, ultimately converging Z_n to 1.

In MRI scan plane positioning tasks, it is important to note that precise spatial positioning of landmarks does not necessarily align with plane orientation for scanning purposes. Given the topological relationship between landmarks in space and the scanning plane's orientation, we have introduced plane normal vectors carrying physical meanings to calculate the DRL. In particular, the DRL aims to optimize the normal vector directions of the sagittal, transverse, and coronal planes predicted by the model. This is mathematically expressed as:

$$L_{DRL} = \sum_{d=1}^3 (1 - \vec{v}_0^d \cdot \vec{v}_m^d) \quad [3]$$

Where \vec{v}_0^d represents the labelled direction of a certain vector, \vec{v}_m^d indicates the predicted direction of that vector by the model, and “.” denotes the point formation. The loss is minimized to zero when the predicted normal vector direction aligns with the true normal vector direction.

Data augmentation with domain knowledge

In this paper, a carefully designed data augmentation approach as shown in *Figure 3* is implemented to improve the adaptability of the model to different scenarios encountered in the clinical setting, taking into account the characteristics of head planning. The first augmentation technique involves a stochastic 3D rotation which mimics the various tilt angles that may occur during scanning. This experiment sets the probability of random 3D rotation to 0.5, with the three axes' rotation angles consisting of random values within $[-20^\circ, 20^\circ]$. The second technique is

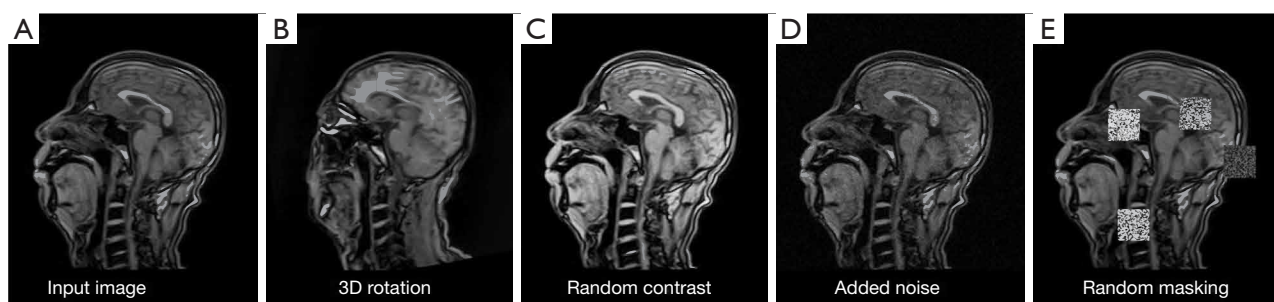


Figure 3 Data augmentation with domain knowledge. (A) Input image; (B) effect of stochastic 3D rotation; (C) effect of random contrast variation; (D) effect of adding random Gaussian noise; and (E) effect of random masking. 3D, three-dimensional.

random contrast variation. Random contrast is employed to replicate situations where an image is influenced by the physiological composition, instrument status, magnetic field strength, and other factors which impact the contrast. In this paper, such simulations are realized using a random gamma transform where $s = c * I^\gamma$. Here, I is the input image, γ is the contrast adjustment factor, c is the scaling factor, and S represents the output image. During the experiment, the value of c was set to 1, while γ was set randomly within the range of [0.7, 1.3]. A gamma transform was applied with a probability of 0.5.

Gaussian noise was also added randomly. When operating MRI scanners, the quality of the image may degrade due to interference from the system's temperature, random thermal motion of electronic components, and other electronic noises. These factors can affect the model's recognition capabilities. To improve the noise immunity of the model, Gaussian noise is randomly added to the training samples with a probability of 0.2, a mean of 0, and a variance of 0.02. The probability of 0.2 is a moderate value to ensure sufficient noise injection during training without affecting each sample. A zero mean was used so that the noise was symmetrically distributed and unbiased, and 0.02 is a low variance that provided relatively subtle noise to avoid obscuring key image features critical to landmark localization.

Random masking involves randomly replacing pixel values in certain regions of an image with the noise mask during the training process. Its purpose is to simulate specific clinical scenarios, such as tumor occupancy, organ lesions, and missing images. The impact of random masking is illustrated in *Figure 3E*. In our experiment, we mask out four random equal-sized regions in the original image using four square noise blocks with side lengths of 40 pixels. The probability of random masking is set to 0.2. We employ

a uniform distribution $U(0, 1)$ for the brightness of the square noise blocks.

Evaluation metric

Two fundamental quantitative metrics in detecting landmarks are the point-to-point absolute error (PAE) and the point-to-point relative error (PRE). According to their respective definitions, the PAE and PRE are as follows:

$$PAE = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\| \quad [4]$$

$$PRE = PAE / L \quad [5]$$

Here, N symbolizes the total number of target landmarks, x_i signifies the labeled coordinates of i -th landmark, \hat{x}_i represents the model-predicted coordinates of the i th landmark, and L signifies the overall size of the image.

The determination of the orientation of the target plane is the most essential aspect of the automatic scan plane positioning task. Thus, this paper suggests using the average angular error (AAE) to assess the degree of deviation of the targeting plane. The AAE is described as:

$$AAE = \frac{1}{3} \sum_{i=1}^3 \frac{180}{\pi} \arccos(v_i \cdot \hat{v}_i) \quad [6]$$

Where, v_i represents the labelling direction of sagittal, coronal and transverse planes, and \hat{v}_i represents the predicted direction of the corresponding plane by the model.

Experiments

2D image acquisitions in traditional auto scan plane

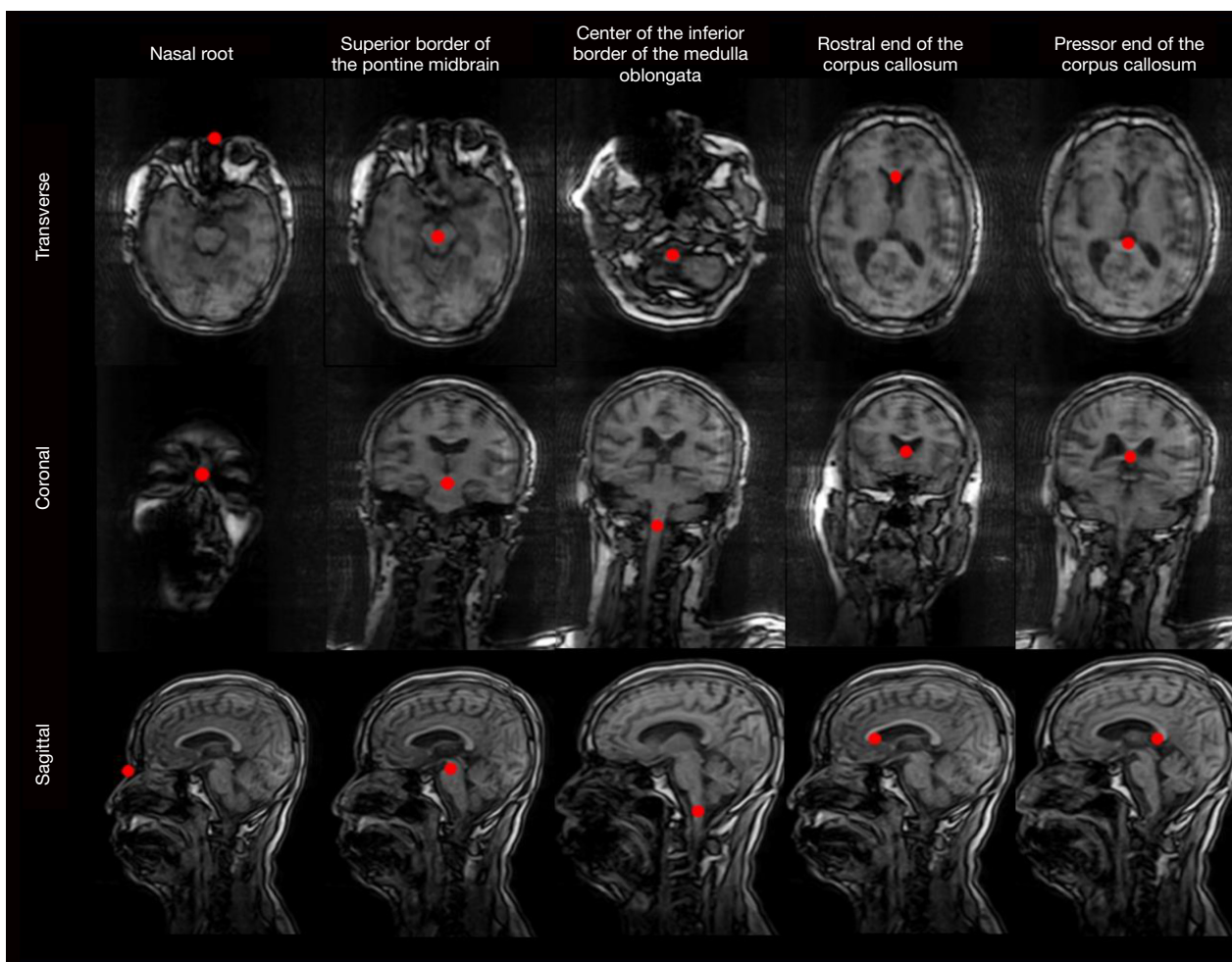


Figure 4 All five anatomical landmarks position from the view of transverse, coronal and sagittal, respectively. Five anatomical landmarks with clinical significance are chosen as: nasal root, superior border of the pontine midbrain, center of the inferior border of the medulla oblongata, rostral end of the corpus callosum, and the proximal end of the corpus callosum.

positioning pose a significant challenge to the accuracy and stability of landmark detection. One of the reasons for this is that 2D images often fail to provide complete coverage of the region of interest, leading to the loss of crucial anatomical information. Additionally, 2D images are unable to convey information on target landmarks from 3D space, which significantly constrains the performance and stability of landmark detection algorithms. In this study, we used the Turbo Field Echo 3D (TFE3D) sequence to acquire data and generate 3D images, and then used a 3D network to process and identify these images. For all images presented in this paper, we performed the TFE3D sequence scanning configured with a 16-ch head neck coil, with an image spatial resolution of $1.875 \times 1.875 \times 2 \text{ mm}^3$, matrix size is

$160 \times 160 \times 120$, repetition time (TR) = 4.8 ms, echo time (TE) = 2.2 ms, slice number = 120, slice thickness = 2.0 mm, field of view (FOV) = $300 \times 300 \text{ mm}^2$.

Landmarks for the automatic scan plane positioning task must support accurate scanning plane orientation and be clinically significant. Our selection of five anatomical landmarks was based on clinical relevance and reliable identification. We chose landmarks that are clearly visible on MRI scans, distinct within the anatomy to avoid confusion, and minimally affected by movement or change between patients and scans. These criteria (localizability, visibility, specificity, and stability) ensure accurate and consistent positioning of the scan plane across diverse clinical applications, as shown in *Figure 4*: nasal root, superior

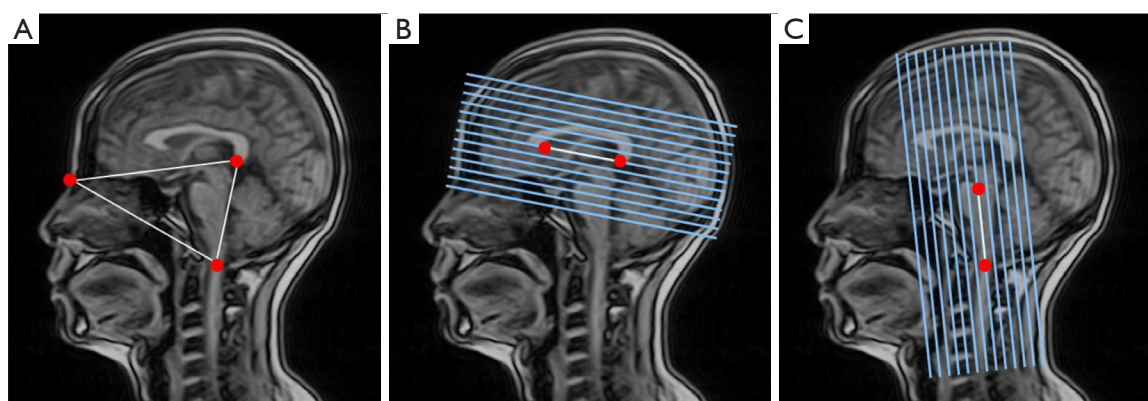


Figure 5 Sagittal, transverse, and coronal planes determination based on five anatomical landmarks. (A) The sagittal plane is determined using three landmarks: the nasal root, pressor end of the corpus callosum, and center of the inferior border of the medulla oblongata. (B) The transverse plane is defined as the plane that crosses the pressor end and rostral end of the corpus callosum, and is perpendicular to the sagittal plane. (C) The coronal plane is defined as perpendicular to the sagittal plane and should cross the landmarks of the center of the inferior border of the medulla oblongata and the superior border of the pontine midbrain.

border of the pontine midbrain, center of the inferior border of the medulla oblongata, rostral end of the corpus callosum, and the pressor end of the corpus callosum. All of these specific and stable landmarks are easily identifiable within MRI images due to their distinct characteristics. Planning for accurate scans along the sagittal, transverse, and coronal planes can be achieved by combining these aforementioned landmarks in an orderly fashion, as shown in *Figure 5*. Specifically, the sagittal plane is defined by three landmarks: the nasal root, the pressor end of the corpus callosum, and the center of the inferior border of the medulla oblongata (*Figure 5A*). Additionally, the transverse plane is defined as perpendicular to the sagittal plane and crossing the pressor and rostral ends of the corpus callosum. The coronal plane is defined as perpendicular to the sagittal plane and should cross the landmarks of the center of the inferior border of the medulla oblongata and the superior border of the pontine midbrain.

A total of 559 brain images were collected from volunteers at Hengyang Central Hospital and of which 553 were selected with acceptable image quality in this experiment after data screening. Clinicians manually labelled the test dataset using 3D Slicer software (31), of which 312 were assigned to the training set, 12 to the validation set and 229 to the test set. The data for the training, validation, and test sets were randomly sampled from the entire dataset to ensure the generalizability of the network model's performance. The dataset included participants with an average age of 42.3 years (range,

19.2–63.7 years), with a gender distribution of 33.8% female and 66.2% male. Notably, 92.4% of the participants were healthy controls, while the remaining 7.6% were patients. The data were obtained using a SuperMark 1.5T MRI system (Anke High-Tech Co., Ltd., Shenzhen, China), and all volunteers provided written informed consent. We have executed our 3D CFP-U-Net model within the PyTorch deep learning framework, utilizing four NVIDIA A4000 graphics cards with 16G video memory. The batch size was 12, and the learning was set at a rate of 0.001. Employing a regular term coefficient of 0.0005, 200 epochs of training were conducted in approximately 12 hours.

Results

Evaluation of 3D CFP-U-Net performance

The performance of 3D CFP-U-Net was assessed on 229 samples with an average prediction time of 0.2 seconds per sample. The results are presented in *Table 1*. The test data revealed a PAE of 0.872 mm, PRE of 0.10%, and an AAE of 0.502°, 0.381°, and 0.675° in sagittal, transverse, and coronal planes, respectively. To provide a comprehensive assessment of the accuracy of orientation across all test images, we calculate the proportion of samples where the predicted AAEs in all three dimensions were simultaneously less than a specific threshold. In the test data, we found that 100% of AAEs were $\leq 3^\circ$, and 92% were $\leq 2^\circ$.

Figure 6 demonstrates the model's accuracy in predicting

Table 1 Average 3D CFP-U-Net test results on 229 samples

Model	PAE, mm	PRE (%)	AAE (SAG), °	AAE (TRA), °	AAE (COR), °
3D CFP-U-Net	0.872	0.10	0.502	0.381	0.675

3D, three-dimensional; CFP, cascade feature pyramid; PAE, point-to-point absolute error; PRE, point-to-point relative error; AAE, average angular error; SAG, sagittal; TRA, transverse; COR, coronal.

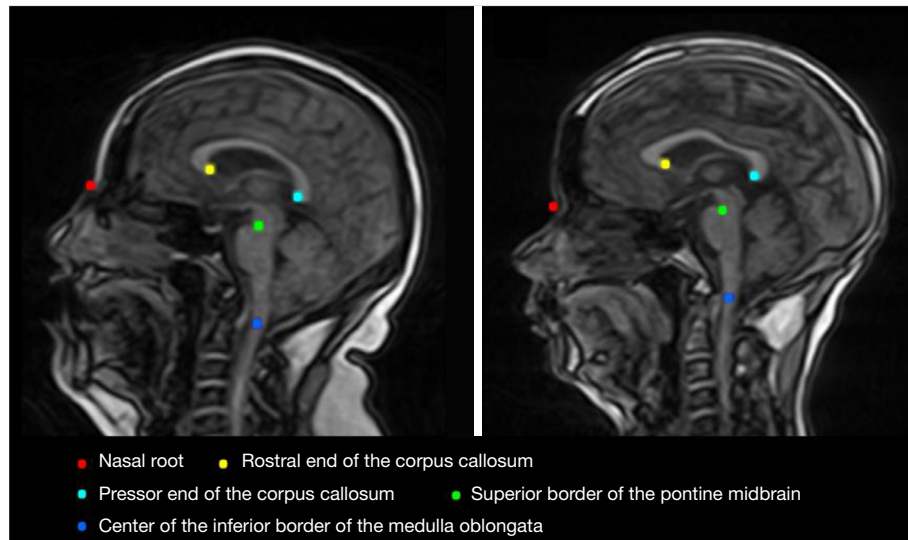


Figure 6 Sagittal view of the detection results for five anatomical landmarks in two samples using the 3D CFP-U-Net model. Points with different colors are used to indicate five landmarks defined in the experiment process. 3D, three-dimensional; CFP, cascade feature pyramid.

anatomical landmarks in two typical clinical samples. The figure shows that the model accurately detected the locations of five anatomical landmarks, including the nasal root (red dot), the point at the upper edge of the pontine midbrain (green dot), the point at the center of the lower edge of the medulla oblongata (blue dot), the point at the rostral end of the corpus callosum (yellow dot) and the point at the end of the corpus callosum compression section (light green dot).

Based on the predicted landmark locations, the sagittal, transverse, and coronal imaging planes were estimated, with the depiction of a typical sample outcome in *Figure 7*. As shown in the figure before automatic scan plane positioning, there is a noticeable tilt angle in the pre-scan image of the subject within VOI owing to the incline of the head (*Figure 7A*). If the transverse, coronal and sagittal plane were simply set in the center of VOI, significant skewing would result (*Figure 7B-7D*). Accurate scanning planning can be accomplished by utilizing the anatomical landmarks anticipated by the 3D CFP-U-Net model (for instance, *Figure 7F-7H*).

Effect of network structure and physical loss function

For this study, we assess the effects of enhancing the network model and loss function. We will compare four combinations of network structures and loss functions, specifically: (I) 3D U-Net, a fundamental 3D U-Net network structure with mean square error (MSE) as the loss function; (II) 3D U-Net + PRL, a basic 3D U-Net network structure with PRL as the loss function; (III) 3D CFP-U-Net + PRL, a 3D CFP-U-Net and use PRL as the loss function; (IV) 3D CFP-U-Net + PRL + DRL, a 3D CFP-U-Net and use point regression, and directional regression as the combined loss function. *Table 2* shows the quantitative performance evaluation results for these four models. The table shows that the 3D CFP-U-Net + PRL + DRL proposed in this paper has the best performance in all quantitative metrics, including a PAE error of 0.886 mm, a PRE of 0.11%, and AAEs of 0.521°, 0.384°, and 0.681° were measured in the sagittal, transverse, and coronal planes, respectively. In comparison to the 3D U-Net, the 3D U-Net + PRL model demonstrated varied levels of

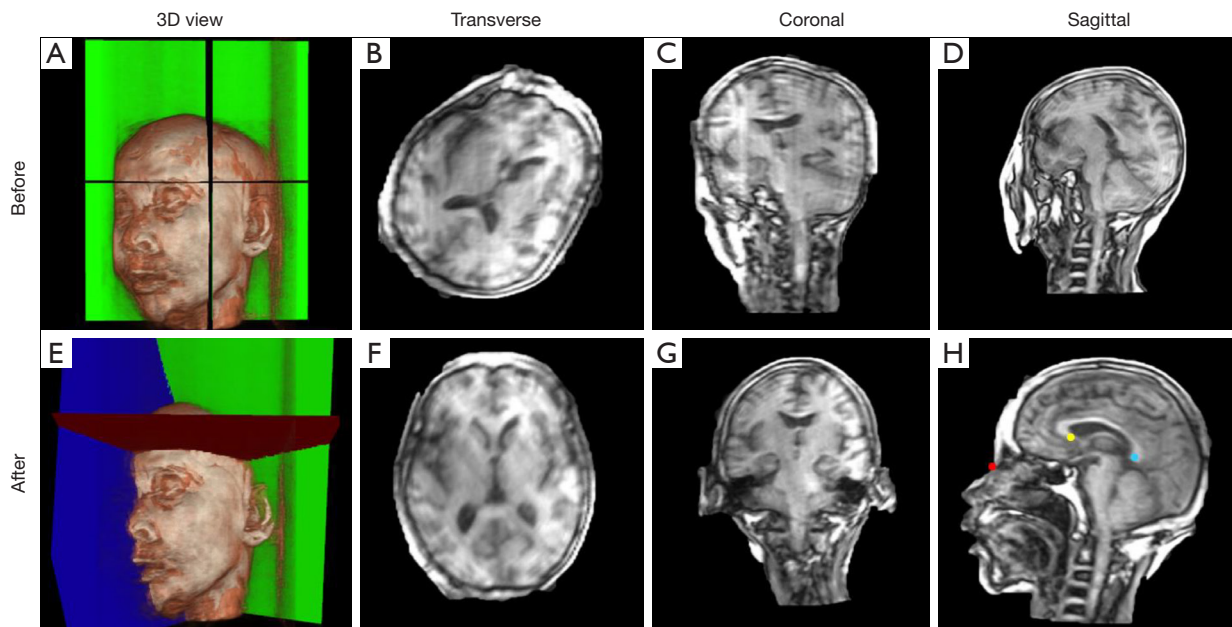


Figure 7 A typical sample scan plane positioning outcome based on the prediction of the 3D CFP-U-Net model. (A-D) The 3D view, transverse, coronal and sagittal images of the volunteer before automatic scan planning; (E-H) the 3D view, transverse, coronal and sagittal images of the volunteer after 3D CFP-U-Net model scan planning. Before automatic scan plane positioning, there was a noticeable tilting angle in the subject's 3D view. However, after automatic scan plane positioning, an accurate scan position can be placed in each plane. Specifically, landmarks such as the nasal root (red dot), the point at the rostral end of the corpus callosum (yellow dot) and the point at the end of the corpus callosum compression section (light green dot) are quite prominent in the sagittal plane, which provides strong evidence for the intuitive proof of prediction accuracy. 3D, three-dimensional; CFP, cascade feature pyramid.

Table 2 Average quantitative performance evaluation results of four models on 229 test samples

Models	PAE, mm	PRE (%)	AAE (SAG), °	AAE (TRA)	AAE (COR)
3D U-Net	3.665	0.70	1.405	2.042	2.192
3D U-Net + PRL	2.413	0.46	0.912	1.623	1.832
3D CFP-U-Net + PRL	1.125	0.25	0.603	0.525	0.732
3D CFP-U-Net + PRL + DRL	0.886	0.11	0.521	0.384	0.681

PRL, point regression loss; PAE, point-to-point absolute error; PRE, point-to-point relative error; AAE, average angular error; SAG, sagittal; TRA, transverse; COR, coronal; 3D, three-dimensional; CFP, cascade feature pyramid; DRL, direction regression loss.

improvement in all quantitative assessment metrics. The model showed an increase in improvement with the PAE reducing from 3.668 to 2.413 mm and the PRE reducing from 0.70% to 0.46%. This suggests that the PRL loss function provides a greater advantage than MSE in locating landmarks. In comparison, the 3D CFP-U-Net + PRL model showed varying degrees of improvement in all quantitative evaluation metrics when compared to 3D U-Net + PRL, with PAE decreasing from 2.413 to 1.125 mm and PRE decreasing from 0.46% to 0.25%. The

network structure of 3D CFP-U-Net, including FPM and cascade structure, suggests superior learning capability in comparison to that of 3D U-Net. Finally, when compared to 3D CFP-U-Net + PRL, the ultimate model, 3D CFP-U-Net + PRL + DRL, exhibits various levels of improvement across all quantitative evaluation metrics. Specifically, there is an improvement in PAE from 1.125 to 0.886 mm and PRE from 0.25% to 0.11%, strongly suggesting that the directional loss has significantly enhanced the model's orientation determination capability.

Table 3 Average quantitative performance evaluation results of 3D CFP-U-Net model under three different data augmentation strategies

Models	PAE, mm	PRE (%)	AAE (SAG), °	AAE (TRA), °	AAE (COR), °
3D CFP-U-Net-0	0.886	0.11	0.521	0.384	0.681
3D CFP-U-Net-1	0.875	0.10	0.509	0.382	0.679
3D CFP-U-Net-2	0.872	0.10	0.502	0.381	0.675

3D CFP-U-Net-0, model without data augmentation; 3D CFP-U-Net-1, model with random 3D rotation and random contrast transformation data augmentation; 3D CFP-U-Net-2, model with random 3D rotation, random contrast transformation, random Gaussian noise, and random masking data augmentation. 3D, three-dimensional; CFP, cascade feature pyramid; PAE, point-to-point absolute error; PRE, point-to-point relative error; AAE, average angular error; SAG, sagittal; TRA, transverse; COR, coronal.

Effect of data augmentation with domain knowledge

This experiment investigates how different data augmentation techniques can improve the network model's adaptability to complex clinical images. During the training phase, we implemented three distinct data augmentation schemes based on the 3D CFP-U-Net model. The 3D CFP-U-Net-0 model did not use data augmentation. For the 3D CFP-U-Net-1 model, two types of data augmentation were added, specifically random 3D rotation and random contrast transformation. The 3D CFP-U-Net-2 model employed four data augmentation techniques during training, including random 3D rotation, random contrast transformation, random Gaussian noise, and random masking. *Table 3* displays the quantitative performance evaluation outcomes for all three data augmentation schemes. From the table, it can be seen that: (I) in comparison with 3D CFP-U-Net-0, 3D CFP-U-Net-1 reduces the PAE from 0.886 to 0.875 mm, the PRE from 0.11% to 0.10%, and the AAE in the sagittal, transverse, and coronal planes decreases from 0.521°, 0.384°, and 0.681°, respectively, to 0.509°, 0.382°, and 0.679°; (II) compared to 3D CFP-U-Net-1, the quantitative evaluation indices of 3D CFP-U-Net-2 show further improvement. The PAE has decreased from 0.875 to 0.872 mm, and the AAE in sagittal, transverse, and coronal planes has also reduced from 0.509°, 0.382°, and 0.679° to 0.502°, 0.381°, and 0.675°, respectively. The experiments conducted in this paper demonstrate that the multiple data augmentation strategies applied have improved the model's ability to generalize.

To provide a more intuitive demonstration of the impact of data augmentation on automatic scan plane positioning performance, two representative samples featuring abnormal clinic images were specifically chosen for detailed analysis. The impact of scan plane localization on the concerning images is depicted in *Figure 8*. In sample A, the pre-

scan image displays a wide range of areas with low signal intensity (red arrow in *Figure 8A*), potentially a result of clinical scanning or patient movement. Moreover, the signal corresponding to the corpus callosum pressure section at the critical anatomical location is largely absent, presenting a significant challenge for the landmark identification-based automatic localization algorithm. The 3D localization maps indicate that the 3D CFP-U-Net-0 model and 3D CFP-U-Net-1 model exhibit notable bias in predicting such samples. In contrast, the 3D CFP-U-Net-2 model, which utilizes a more robust augmentation strategy, can achieve accurate scan plane localization (red box in *Figure 8A*). Significant regions of tumor occupied abnormal sample B in the patient's pre-scan images, which resulted in deviations from normal for the location of key anatomical points in the patient's images. In the results, model 3D CFP-U-Net-0 was shown to be incorrect in its prediction without data augmentation, and models 3D CFP-U-Net-1 and 3D CFP-U-Net-2 with data augmentation achieved more accurate scan plane localization.

Discussion

The proposed 3D CFP-U-Net is an end-to-end network architecture that can be applied in diverse conditions. In this framework, a 3D pre-scan image serves as input and localized labelling of predefined landmarks is the output during the training phase. To expand the application of this framework beyond the head region, we can substitute the input pre-scan image and the output labelled landmark localization for the target anatomy during the training stage. To reduce scanning time, the 3D pre-scan data presented in this paper is at a lower resolution which may present some challenges for finer identification and localization. In future, we will integrate fast imaging techniques into 3D pre-scanning to obtain higher resolution pre-scan images while

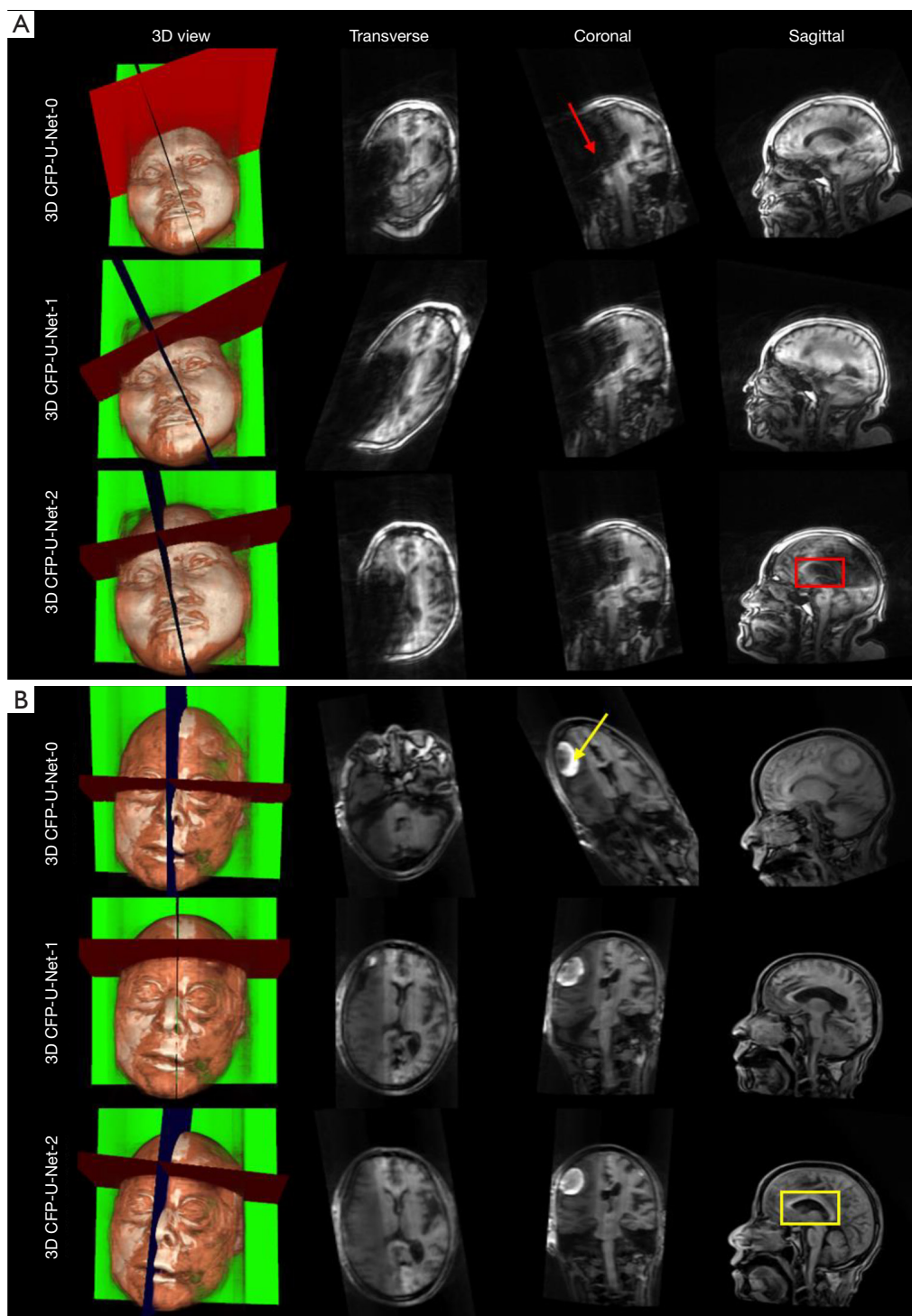


Figure 8 The performance of data augmentation strategy for 3D CFP-U-Net model under two clinically abnormal samples (A,B). The red and yellow arrows indicate the regions of critical abnormalities, respectively. Compared to 3D CFP-U-Net-0 without data augmentation, 3D CFP-U-Net-1 and 3D CFP-U-Net-2 show better identification and plane localization performance in abnormal samples (red and yellow boxes). 3D CFP-U-Net-0: model without data augmentation; 3D CFP-U-Net-1: model with random 3D rotation and random contrast transformation data augmentation; 3D CFP-U-Net-2: model with random 3D rotation, random contrast transformation, random Gaussian noise, and random masking data augmentation. 3D, three-dimensional; CFP, cascade feature pyramid.

improving scanning efficiency. The proposed physically-based directional regression loss function is pivotal in advancing the model's localization accuracy. Optimizing model parameters, such as weight values in the loss function and parameters used in data augmentation, could potentially improve the model's performance further. It is observed that the normal vectors of the scanning plane represent a topological association based on landmarks, and extending the constraints imposed on the model by this association has the potential to improve its effectiveness (32).

Conclusions

In this paper, we address the inefficiency, low accuracy and poor generalization ability of traditional scan plane positioning algorithms for MRI, and propose a deep learning automatic scan plane positioning solution based on a 3D data scheme by designing a specified data acquisition strategy, data augmentation method, network architecture and optimized loss function. The approach utilizes TFE3D sequences to promptly obtain 3D pre-scanning localization images for subsequent identification of anatomical landmarks and intelligent computation of plane orientation. Our proposal is a two-stage end-to-end 3D cascaded convolutional network framework, called 3D CFP-U-Net, which localizes the positions of five key anatomical landmarks and achieves a coarse-to-fine result. Our approach yields satisfactory scan plane positioning outcomes on 229 test samples, with PAE and PRE reaching 0.872 mm and 0.10%, respectively, evidencing the efficacy of the cascade framework and the multi-scale spatial information fusion module. In comparison, a semi-automatic scan plane positioning method for cardiac imaging (12) resulted in higher mean distance errors of 4.96 mm. Additionally, Zhan *et al.*'s machine learning approach (14) for femur cartilage positioning produced a translation error of 1.53 mm, larger than our method's 0.872 mm. We propose loss functions PRL and DRL with a physical meaning in automatic scan plane positioning, and verify the performance of PRL and DRL through comparative experiments. Considering the diverse clinical application scenarios, we create simulations of numerous complex situations encountered in the clinic to expand the training dataset. We then integrate domain knowledge into the neural network using the augmented training data to enhance the neural network's generalization ability. Data augmentation proves to be effective in improving the performance indices of the model, as demonstrated by the results of 229 test samples. Its

improved performance in clinical anomaly samples provides further evidence of the potential of data augmentation in enhancing model robustness.

Acknowledgments

We would like to acknowledge Mulan Li (Anke High-tech Co., Ltd., Shenzhen, China) and Yuling Jiang (Anke High-tech Co., Ltd., Shenzhen, China) for their valuable contributions to the discussions on experimental protocols and clinical applications.

Funding: None.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-1740/coif>). H.G. serves as an unpaid editorial board member of *Quantitative Imaging in Medicine and Surgery*. G.Z., X.S., Z.S., Z.X., J.Z. and Z.Y. report that they are employees of Anke High-tech Co., Ltd. The other author has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional review board of Hengyang Central Hospital, China (the registration number of ethics board: No. 17. Dated June 30, 2023), and informed consent was taken from all the patients and volunteers.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Wright GA. Magnetic resonance imaging. *IEEE Signal*

- Process Mag 1997;14:56-66.
2. Symms M, Jäger HR, Schmierer K, Yousry TA. A review of structural magnetic resonance neuroimaging. *J Neurol Neurosurg Psychiatry* 2004;75:1235-44.
 3. Hennig J, Speck O, Koch MA, Weiller C. Functional magnetic resonance imaging: a review of methodological aspects and clinical applications. *J Magn Reson Imaging* 2003;18:1-15.
 4. Lauterbur PC, Kramer DM, House WV Jr, Chen CN. Zeugmatographic high resolution nuclear magnetic resonance spectroscopy. Images of chemical inhomogeneity within macroscopic objects. *Journal of the American Chemical Society* 1975;97:6866-8.
 5. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging* 2000;19:143-50.
 6. van der Kouwe AJ, Benner T, Fischl B, Schmitt F, Salat DH, Harder M, Sorensen AG, Dale AM. On-line automatic slice positioning for brain MR imaging. *Neuroimage* 2005;27:222-30.
 7. Lecouvet FE, Claus J, Schmitz P, Denolin V, Bos C, Vande Berg BC. Clinical evaluation of automated scan prescription of knee MR images. *J Magn Reson Imaging* 2009;29:141-5.
 8. Mäkelä T, Clarysse P, Sipilä O, Pauna N, Pham QC, Katila T, Magnin IE. A review of cardiac image registration methods. *IEEE Trans Med Imaging* 2002;21:1011-21.
 9. Mahmoud A, Awad NA, Alsubaie NM, Ansarullah SI, Alqahtani MS, Abbas M, Usman M, Soufiene BO, Saber A. Advanced Deep Learning Approaches for Accurate Brain Tumor Classification in Medical Imaging. *Symmetry* 2023;15:571.
 10. Probst T, Maninis KK, Chhatkuli A, Ourak M, Poorten EBV, Van Gool L. Automatic Tool Landmark Detection for Stereo Vision in Robot-Assisted Retinal Surgery. *IEEE Robotics and Automation Letters* 2017;3:612-9.
 11. Sharp GC, Kollipara S, Madden T, Jiang SB, Rosenthal SJ. Anatomic feature-based registration for patient set-up in head and neck cancer radiotherapy. *Phys Med Biol* 2005;50:4667-79.
 12. Lu X, Jolly MP, Georgescu B, Haye C, Speier P, Schmidt M, Bi X, Kroeker R, Comaniciu D, Kellman P, Mueller E, Guehring J. Automatic view planning for cardiac MRI acquisition. *Med Image Comput Comput Assist Interv* 2011;14:479-86.
 13. Nitta S, Takeguchi T, Matsumoto N, Kuhara S, Yokoyama K, Ishimura R, Nitatori T. Automatic slice alignment method for cardiac magnetic resonance imaging. *MAGMA* 2013;26:451-61.
 14. Zhan Y, Dewan M, Harder M, Krishnan A, Zhou XS. Robust automatic knee MR slice positioning through redundant and hierarchical anatomy detection. *IEEE Trans Med Imaging* 2011;30:2087-100.
 15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
 16. Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ, Kadoury S, Tang A. Deep Learning: A Primer for Radiologists. *Radiographics* 2017;37:2113-31.
 17. Blansit K, Retson T, Masutani E, Bahrami N, Hsiao A. Deep Learning-based Prescription of Cardiac MRI Planes. *Radiol Artif Intell* 2019;1:e180069.
 18. Payer C, Štern D, Bischof H, Urschler M. Integrating spatial configuration into heatmap regression based CNNs for landmark localization. *Med Image Anal* 2019;54:207-19.
 19. Wang X, Zhai S, Niu Y. Left ventricle landmark localization and identification in cardiac MRI by deep metric learning-assisted CNN regression. *Neurocomputing* 2020;399:153-70.
 20. van Zon M, Veta M, Li S, editors. Automatic cardiac landmark localization by a recurrent neural network. *Medical Imaging: Image Processing*; 2019. doi: 10.1117/12.2512048.
 21. Le M, Lieman-Sifry J, Lau F, Sall S, Hsiao A, Golden D, editors. Computationally Efficient Cardiac Views Projection Using 3D Convolutional Neural Networks. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer International Publishing; 2017.
 22. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, editors. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*; 2016.
 23. Zhang D, Wang J, Noble JH, Dawant BM. HeadLocNet: Deep convolutional neural networks for accurate classification and multi-landmark localization of head CTs. *Med Image Anal* 2020;61:101659.
 24. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2015:770-8.
 25. Rosas-Gonzalez S, Birgui-Sekou T, Hidane M, Zemmoura I, Tauber C. Asymmetric Ensemble of Asymmetric U-Net Models for Brain Tumor Segmentation With Uncertainty Estimation. *Front Neurol* 2021;12:609646.
 26. Lin TY, Dollár P, Girshick RB, He K, Hariharan B,

- Belongie SJ. Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017:936-44.
27. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence; San Francisco, CA, USA: AAAI Press; 2017:4278-84.
 28. Zhong Z, Li J, Zhang Z, Jiao Z, Gao X, editors. An Attention-Guided Deep Regression Model for Landmark Detection in Cephalograms. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019; Cham: Springer International Publishing; 2019.
 29. Andermatt S, Pezold S, Amann M, Cattin PC. Multi-dimensional Gated Recurrent Units for Automated Anatomical Landmark Localization. arXiv 2017. doi: 10.48550/arXiv.1708.02766.
 30. Li W, Wang Z, Yin B, Peng Q, Du Y, Xiao T, Yu G, Lu H, Wei Y, Sun J. Rethinking on Multi-Stage Networks for Human Pose Estimation. arXiv 2019. doi: 10.48550/arXiv.1901.00148.
 31. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S, Kikinis R. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* 2012;30:1323-41.
 32. Kipf T, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. arXiv 2016. doi: 10.48550/arXiv.1609.02907.

Cite this article as: Zhu G, Shen X, Sun Z, Xiao Z, Zhong J, Yin Z, Li S, Guo H. Deep learning-based automated scan plane positioning for brain magnetic resonance imaging. *Quant Imaging Med Surg* 2024;14(6):4015-4030. doi: 10.21037/qims-23-1740