MTMU: Multi-domain Transformation based Mamba-UNet designed for unruptured intracranial aneurysm segmentation

Bing Li[1], Nian Liu[1], Jianbin Bai[2], Jianfeng Xu[3], Yi Tang[4] and Yan Liu[1*]

Abstract

The management of Unruptured Intracranial aneurysm (UIA) depends on the shape parameters assessment of lesions, which requires target segmentation. However, the segmentation of UIA is a challenging task due to the small volume of the lesions and the indistinct boundary between the lesion and the parent arteries. To relieve these issues, this article proposes a multi-domain transformation-based Mamba-UNet (MTMU) for UIA segmentation. The model employs a U-shaped segmentation architecture, equipped with the feature encoder consisting of a set of Mamba and Flip (MF) blocks. It endows the model with the capability of long-range dependency perceiving while balancing computational cost. Fourier Transform (FT) based connection allows for the enhancement of edge information in feature maps, thereby mitigating the difficulties in feature extraction caused by the small size of the target and the limited number of foreground pixels. Additionally, a sub task providing target geometry constrain (GC) is utilized to constrain the model training, aiming at splitting aneurysm dome from its parent artery accurately. Extensive experiments have been conducted to demonstrate the superior performance of the proposed method compared to other competitive medical segmentation methods. The results prove that the proposed method have great clinical application prospects.

Keywords Intracranial aneurysm segmentation, Mamba unit, Fourier transform, Geometry constrain

Introduction

Intracranial aneurysm (IA) is generated by the abnormal dilation of blood vessels, which has a relatively high incidence and an onset occult. Once rupturing, it will result in subarachnoid hemorrhage (SAH), a kind of serious malady with high mortality [1, 2]. Therefore, the management of unruptured intracranial aneurysm (UIA) is

important, including risk evaluation as well as treatment planning based on lesion appearance parameters assessment. In current clinical practice, the measurement of shape parameters is manual work depending on the experience of experts, let alone its high inter- and intra-observer variability and time consumption. With the help of semantic segmentation, measuring the appearance parameters could become objective and reproducible as well as fast to meet clinical demand [3]. Thus, automatic delineation of lesions is critical for UIA management.

Currently, there are two categories of UIA segmentation methods. The first category is traditional rule-based 2D/3D shape analysis. The main idea of which is to use image processing algorithms like region-growing and level set models to identify abnormal regions from the radiology scans [4, 5]. These methods perform

*Correspondence:
Yan Liu
liuyan77@scu.edu.cn
[1] College of Electrical Engineering, Sichuan University, Chengdu 610065, China
[2] Department of Neurosurgery 3201 Hospital, Hanzhong 723000, China
[3] Neurosurgery of the Third People's Hospital of Mianyang/Sichuan Mental Health Center, Mianyang 621000, China
[4] North Sichuan Medical College, Nanchong 637000, China

Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 2 of 16

segmentations mainly based on the image to be processed itself therefore they usually work well on typical targets and need target wise parameter tuning in order to guarantee performance. To be specific, the volume size or shape of aneurysm lesions are supposed to meet the requirement of methods. These constrains limit their clinical applications. The other category of methods is neural network based ones, which has become increasingly popular alongside the development of deep learning. The main idea of them is to model the segmentation task as a pixel-wise classification problem and train a neural network model to solve it. Methods based on convolutional neural network (CNN) have been proven to be effective in detecting UIAs, namely the famous semantic segmentation architectures UNet and its variants [6]. However, the CNN based methods usually pay attention to relatively small regions due to limited receptive fields, resulting in incomplete contours of delineated UIAs. Obtaining long-range dependency is one of the key factors for success of image segmentation, specifically for medical targets like vessels since they are widely distributed in human body and supposed to have similar image features. Limited by local receptive field, traditional neural network like CNN will lead to the inadequate feature extraction, resulting in suboptimal contour segmentation results [7]. Recent studies have integrated transformer modules into segmentation frameworks to enhance the network's ability to capture global information through self-attention mechanisms. These methods show great potential in applying on UIA segmentation [8]. Although demonstrating remarkable performance in medical segmentation tasks, it incurs a high computational cost, with computational complexity quadratic to the input sequence length, which makes it face considerable limitations in clinical use.

Briefly speaking, the automatic delineation of UIA has always been a focal point of scientific research and clinical work, yet it remains poorly resolved. The primary obstacle encountered in crafting UIA segmentation algorithms lies in the following aspects. First, most UIA lesions are relatively small, IA larger than 10 mm typically requires attention due to the risk of rupture which escalates with size, making UIA segmentation a challenging instance of small object semantic segmentation [9, 10]. Accordingly, the lesion contains a limited count of pixels on radiology scans, which has increased the difficulty of feature extraction on pure spatial domain. Second, due to the pathogenesis of UIA, it is challenging to separate it from the parent vessel on tomographic scans. An aneurysm is the result of arterial wall damage or degeneration. Both aneurysm and arterial are filled with blood internally. In another words, similar features exhibited on images limit the algorithm's capability to differentiate the aneurysm dome from the vessel. Last but not least, escalating complexity of segmentation models, while enhancing segmentation performance, has also led to a substantial increase in hardware and software costs, which severely constrains the model's clinical utility. To overcome the aforementioned issues, there is a need for a method that can aggregate non-purely spatial features while achieving an optimal balance between efficiency and performance in IA segmentation task.

In this paper, we propose a novel network called MTMU, which is based on a multi-domain transformation and consists of a U shape encoder-decoder architecture enhanced by Mamba blocks for the segmentation of intracranial aneurysm. Inspired by the fact that the original Mamba block could balance the global feature aggregation capability and computational workload well, we explore the use of it to equip the feature extractor of our segmentation model thereby constructing a new block called Mamba & Flip (MF) block. In addition, starting from the fact that the high-frequency components of the whole image usually contain the edge information, in another words, boundaries of UIAs, we propose to embed the Fourier Transform (FT) into traditional skip connection to construct FT connection. The FT connection enriches the high-frequency information transferring across different network levels, allowing the boundaries to receive more attention during decoding phase. Meanwhile, it works in frequency domain whose effectiveness is not compromised by the limited number of target pixels. In order to address the challenge of accurately segmenting aneurysms from parent vessels, we introduce an extra geometry constrain as subtask (GC subtask) to regularize the learning of network, which utilizes target-wise geometric prior constrains as a compensation of pure pixel-wise classification. The main contributions of this study are as follows:

1. We design a novel network called MTMU for UIA segmentation, in which MF block is proposed to construct the feature encoder, thereby incorporating the ability to extract long-distance dependencies while also balancing segmentation performance and computational workload.
2. FT connection is proposed to provide feature aggregating capability in non-spatial domain, which allows the performance of the network getting rid of the influence of the limited number pixels in small UIA lesions.
3. GC subtask is introduced to the main segmentation task to help enhance the accuracy of the delineation of boundaries between the UIA dome and the parent artery, this enhancement is achieved by using geometric priors.

Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 3 of 16

Related work

Semantic segmentation of UIAs

UIAs segmentation is aimed to accurately extract UIA domes from complex surroundings, with the current mainstream approach being neural network-based segmentation. CNN-based semantic segmentation models have been widely and effectively used in the delineation of various medical targets, including UNet and FCN as well as their variants. Podgorsak et al. claimed a CNN model based on VGG −16 to automatically segment aneurysms from digital subtraction angiography (DSA) images, yielding competitive performance compared with manual delineations [11]. Bizjak et al. proposed a two-stage method consisting of multilayer neural networks to work well for segmentation of aneurysms with various sizes and shapes, and it could work effectively even though the configuration of surrounding vessels is complex [12]. Park et al. designed a model named HeadXNet constructing 3-dimensional CNNs into an encoder-decoder structure for extracting the intracranial aneurysms from computer tomography angiography (CTA) scans, which has been proved to be able to improve the ability of clinical doctors to correctly identify aneurysms in challenging imaging contexts [13]. As time goes by, the improvement in model performance is proved to be limited by the finite size of the receptive field. Transformer is introduced to be the basic component of semantic segmentation networks to model global long-range capability, yielding better segmentation accuracy. For example, UNETR, a model designed for 3D medical segmentation, could effectively capture global multi-scale information by utilizing transformers as encoders across different scales [14]. Yawu-Zhao developed ConTNet consisting of two encoders, a CNN based one and a transformer based one respectively [15]. Thanks to the combination of local and global information, ConTNet achieved impressive performance in IA segmentation task. However, the transformer-based segmentation models usually face challenges of balancing performance and computational cost. In addition, it pays attention to the entire image space which weakens its sensitivity of small targets, like IA. We propose to construct our segmentation model aiming at superior delineation performance and computational efficiency.

Mamba in semantic segmentation

Recently, Mamba, a newly coming neural network component originating from state space sequence models, has emerged as a powerful long-sequence modeling approach [16]. The linear computational complexity of Mamba guarantees the high efficiency of the training and inference processes. It has demonstrated performance that surpasses Transformers in both natural language processing and computer vision domains [17]. Researchers have preliminarily tried the Mamba block in the task of medical semantic segmentation, achieving good performance. The innovative architecture of Mamba-UNet, which integrates the U-Net with Mamba, consists of encoder and decoder components that are purely composed of visual Mamba elements. This novel structure has demonstrated effective segmentation performance across various datasets [18]. U-Mamba introduces a hybrid architecture of CNN-Mamba, combing the local feature extraction capability of CNN with the long-range dependency capture capability of Mamba [19]. It proves to be effective in delineating boundary of cells task and abdomen task respectively. Meanwhile, Vision Mamba UNet inserts Mamba block in each encoder layer, and fuse the low-level and high-level information extracted by Mamba block, it exhibits competitive performance in many medical datasets [20]. To the best of our knowledge, there is no work that utilizes the Mamba structure for the task of small-object semantic segmentation. We propose to incorporate it into our model to explore its potential in improving the UIA delineation performance and decreasing the computational cost.

Frequency domain feature in semantic segmentation

Fourier transform is a technology which is widely used in signal processing, providing a way to understand feature maps in transformed domain. In neural network-based computer vision tasks, people begin to try involving frequency information to guide the learning of models and further improve the performance. FcaNet [21] proposed frequency channel attention mechanisms to address the deficiency of extracting features in pure image domain, and achieved good performance in object detection tasks. Yang et al. [22] introduced Fourier transform in semantic segmentation task by using it to swap the low frequence spectrum of the synthetic datasets and real dataset, this swap can preserve the low frequence information which is not obvious in image domain. Overall, the motivation for incorporating frequential features is that the spatial features of the target do not possess sufficient discriminative characteristics. In this case, the frequential features work as compensation. Similar concerns have been proposed in our case. Due to the small size of the UIA target, the number of foreground pixels is limited, which increases the difficulty of feature extraction, not to mention providing boundary information. We propose to introduce Fourier transform to enhance the feature aggregating capability in our model from the perspective of frequency domain.

Geometry priors in semantic segmentation

Due to the low contrast and uneven texture of medical images, segmentation accuracy is often compromised.

Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 4 of 16

Some studies have attempted to improve performance by incorporating anatomical priors which contain lesion shapes and locations. For instance, Huang et al. integrated organ anatomical priors into deep learning models using probabilistic atlases as component of the loss function, which enhanced the accuracy of their segmentation framework [23]. Similarly, Zheng et al. proposed Deep Atlas Prior combined segmentation frameworks with probabilistic atlases to aid segmentation, achieving notable accuracy [24]. Additionally, You et al. introduced the shape prior module to explicitly incorporate geometric priors as input, guiding the model for better target segmentation [25]. These studies demonstrate that leveraging geometric priors can enhance segmentation accuracy. According to the fact that point cloud is one of the most widely used geometric characteristic descriptors, we propose to utilize it to as geometric priors to constrain the delineated target shape to improve our segmentation accuracy.

Method

In this section, we describe the proposed MTMU network designed for the automatic segmentation of the aneurysm. Our model is an encoder-decoder network architecture consisting of the newly designed feature aggregation block (MF block), the Fourier Transform based edge enhancing skipping connection (FT connection) and the extra geometry constrain subtask (GC subtask), as illustrated in Fig. 1. The feature extractor MF blocks cascaded in the encoder are based on the basic Mamba unit with improvement of ability to model global information, aiming to abstract the features from similar anatomy structures (arterial vessels) hidden behind the long space distance. The FT connection is used as skipping connection across corresponding levels of encoder and decoder, enhancing the model's capability of paying attention to high-frequency information during decoding phase, and our decoder phase is composed of convolution layers. GC subtask is applied to ensure geometry consistence of the delineated intracranial aneurysm lesions.

Mamba & Flip direction (MF) block

To balance the network receptive field and the computation cost, we propose a feature extracting unit named MF block to benefit the model with capability of capturing long-range dependency while avoiding too much computational burden.

The introduced MF block consists of a set of parallel basic mamba units to perform feature calculation and corresponding flip operations to ensure rotational invariance during the calculation. From our point of view, incorporating features from different directions could help the model perceive correlations among

them, thus make the captured dependency solid. The detailed structure of MF block is showed in Fig. 2.

Let the input feature $F_{l-1} \in R^{C_{l-1} \times D_{l-1} \times W_{l-1} \times H_{l-1}}$ where C_{l-1}, D_{l-1}, W_{l-1}, H_{l-1} indicate the channels, depth, width and height of the $l-1^{th}$ layer feature map F_{l-1}. Due to the requirement of flattening the feature maps before utilizing the original Mamba unit, during which there is a significant risk of losing spatial information. To compensate for the potentially lost spatial information, F_{l-1} has been flipped along specific directions as preprocessing, yielding a set of flipped feature maps as shown in Eq. (1). Meanwhile the simultaneous multi-directional flipping could help improve the rotational invariance capability of the model.

$$FP_{\dim} = Flip_{\dim}(F_{l-1}), \dim \in \{\emptyset, C, DWH, CDWH\} \tag{1}$$

where the subscript dim indicates the specific operational directions of the multi-directional flipping. For example, $\dim = C$ means to flip along the channels and $\dim = DWH$ refers to the sequential flipping along the depth, width and height dimensions respectively. Here, the symbol $\emptyset$ indicates that no flipping operation is performed. In the next step, $FP_{\dim}$ will be flattened, denoting as $f_{\dim} = Flatten(FP_{\dim})$, where $f_{\dim} \in R^{C_{l-1} \times D_{l-1} \times (W_{l-1}H_{l-1})}$. The obtained $f_{\dim}$ in each direction will be fed into the original Mamba unit $Mamba(\cdot)$ to aggregate the internal feature vector from the input F_{l-1}. The resulting feature vector is the mean of the vectors derived from the subpaths. Finally, by deflating the vectors back to the dimensions $F_l \in R^{C_l \times D_l \times W_l \times H_l}$, the output feature map F_l of the lth MF block is obtained. The calculation of F_l is shown in Eq. (2).

$$F_l = unFlatten\left(\frac{1}{4}\sum Mamba(f_{\dim})\right) \tag{2}$$

where un$Flatten(\cdot)$ is the inverse operation of $Flatten(\cdot)$.

As the core functional unit of our MF block, the detailed structure of the Mamba unit $Mamba(\cdot)$ is shown in Fig. 3. After directly feeding the feature vector $f_{\dim}$ into Mamba unit, two learnable linear projection operators are applied to transfer $f_{\dim}$ into representative vectors, denoting as $proj_{w_{ssm}}(f_{\dim})$ and $proj_{w_o}(f_{\dim})$. Here $proj_W(\cdot)$ means projecting the input vector onto the direction W. W_{ssm} indicates the linear projection of State-Aware Feature (SAF) branch and W_0 indicates that of the stable branch. In branch SAF, the representative vector $S_{\dim}$ will be calculated sequentially from the input $f_{\dim}$ through a depth-wise convolutional operator $conv_{dp}(\cdot)$ and a State Space Sequence model (SSM) as shown in Eq. (3).
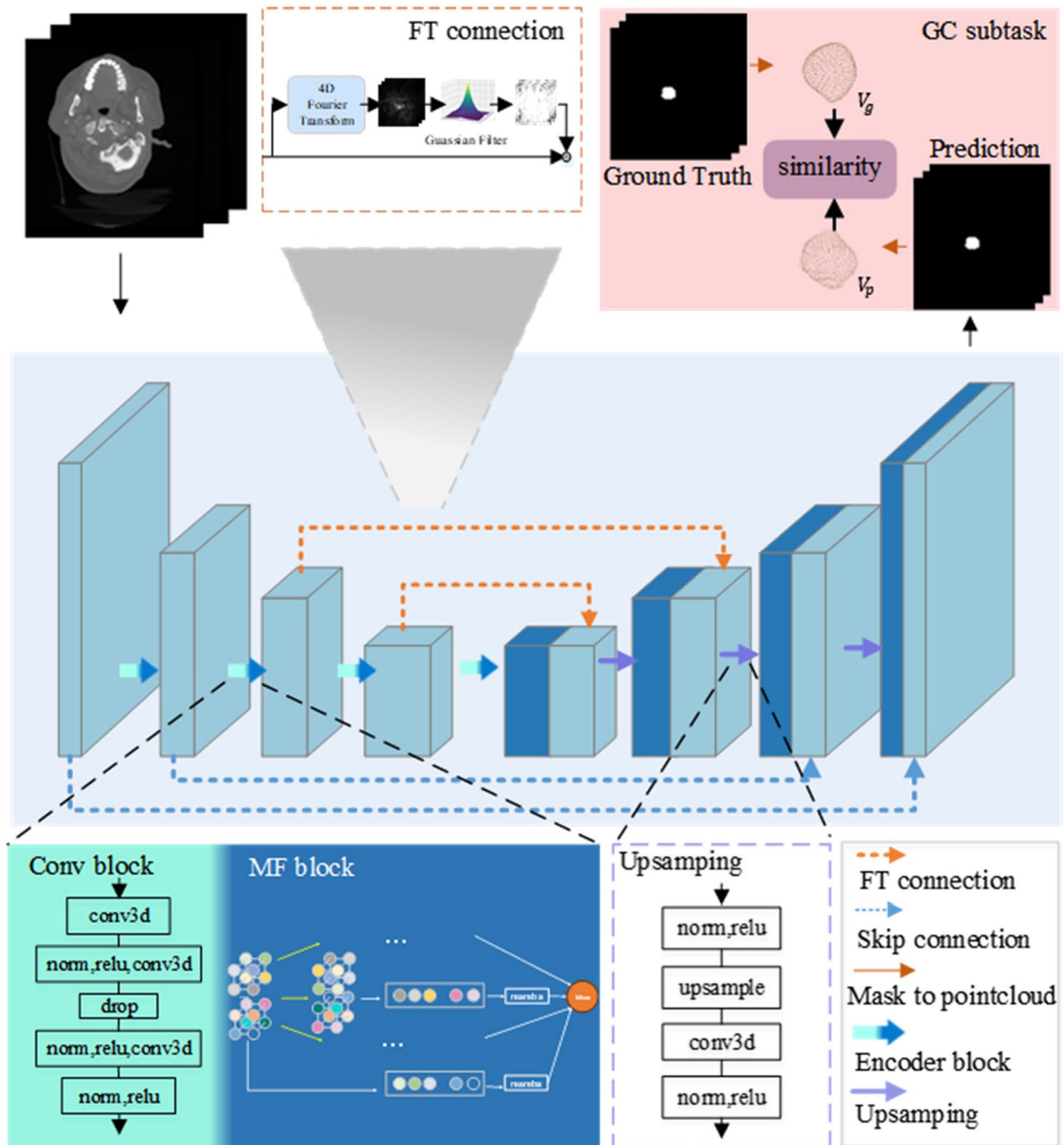
Fig. 1 The pipeline of our segmentation framework

$$S_{\dim} = ssm\left(SiLU\left(conv_{dp}\left(proj_{W_{ssm}}\left(f_{\dim}\right)\right)\right)\right) \quad (3)$$

where $SiLU(x) = x/(1 + e^{-x})$ is the nonlinear activation function. $ssm(\cdot)$ is the SSM whose detail will be talked in the next paragraph. The two branches will be combined together and produced the output of Mamba unit following Eq. (4).

$$Mamba(f_{\dim}) = proj_{W_{conv}}\left(Hadamard\left(S_{\dim}, SiLU\left(proj_{w_o}\left(f_{\dim}\right)\right)\right)\right) \quad (4)$$

where $Hadamard(\cdot, \cdot)$ indicates the Hadamard product of the two inputs and W_{con} is the linear projection located at the output of mamba unit.

The inside of $ssm(\cdot)$ is a discrete state space model that calculates model output y depending on the input
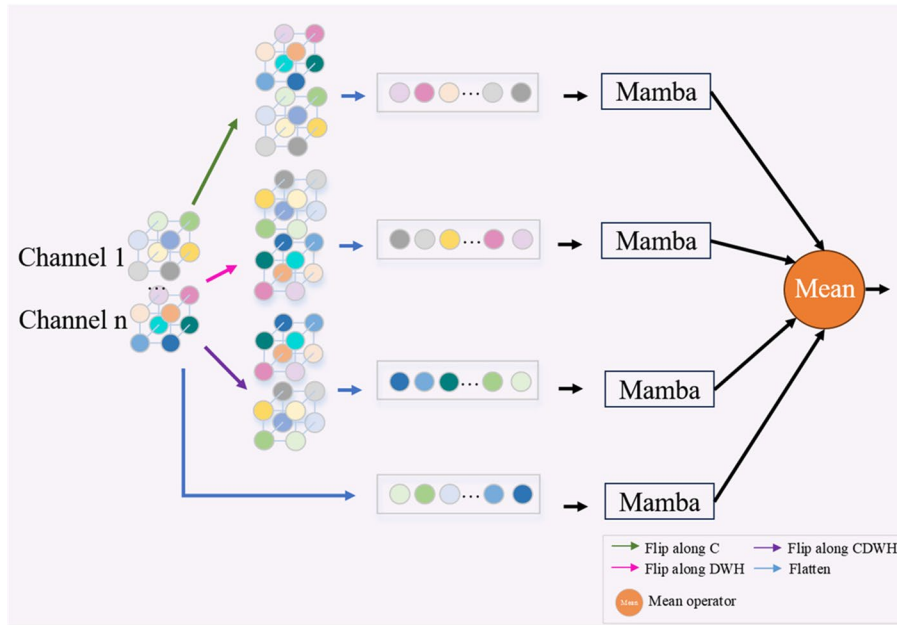
Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 6 of 16

Fig. 2 MF block

vector x and the implicit state vector h in a recursive way. The calculation rules follow the state space equation in Eq. (5).

$$h_k = Ah_{k-1} + Bx_k$$
$$y_k = Ch_k \tag{5}$$

Where A represents the state matrix, B is input matrix, C is the output matrix and subscript k, $k - 1$ mean kth and $(k - 1)th$ element. All of which are learnable parameters. The updating of both A and B will lead to corresponding updating of h_k. Figure 4 illustrates the updating process of h_k.

Fourier transform connection

As target delineation focus the contours which indicate the boundary of foreground and background and usually locate in the position with large intensity gradient, we assume the performance of segmentation would benefit from enhancing the high frequency information during feature extraction. however, the number of contour pixels is limited comparing with the whole image specifically for those tiny targets. it might lead to loss the edge features, during the depth extension of encoder for abstracting semantic information. Consequently, the decoder could not produce satisfied predictions no matter how subtle architecture is designed if not fed by edge corresponding features. To relieve this issue, we design the Fourier transform block and insert it into skipping connection to introduce edge corresponding features to decoder directly, avoiding missing them during the layer-by-layer

feature aggregating. As shown in Fig. 5, the FT block utilize Fourier transform to recognize edge features and enhances the network representation of them.

FT is commonly employed in signal processing, aiding in the oriented analysis of signals according to their frequency components. For the feature map encoded by the lth layer MF block $F_l \in R^{C_l \times D_l \times W_l \times H_l}$, the high frequency components usually denoted as those vary frequently, namely edges and textures, while the low frequency components indicate the smooth region, like the internal of targets. By attenuating low-frequency signals, high-frequency signals can receive more attention, furthermore, edge details can be preserved effectively. To convert the spatial feature map into frequency domain information, we employ an 4-dimensional Discrete Fourier Transform (4-D DFT) to get the frequency map of the features F_l, denoting as $FFT_l = DFT_4(F_l)$, where DFT_4 indicates the 4-D DFT transform.

The high-frequency component $FFT_{l-highpass}$ is obtained by filtering out the low-frequency components from FFT_l. In this paper, we use the high-pass Gaussian filter shown in Eq. (6).

$$FFT_{l-highpass} = 1 - e^{-\frac{FFT_l}{2\sigma^2}} \tag{6}$$

where σ is the cutoff frequency. The high frequency feature map $FFT_{l-highpass}$ is transformed back to the spatial domain feature map $\hat{F} = IDFT_4(FFT_{l-highpass})$, where $IDFT_4$ represents the inverse 4-dimensional Discrete Fourier Transform.
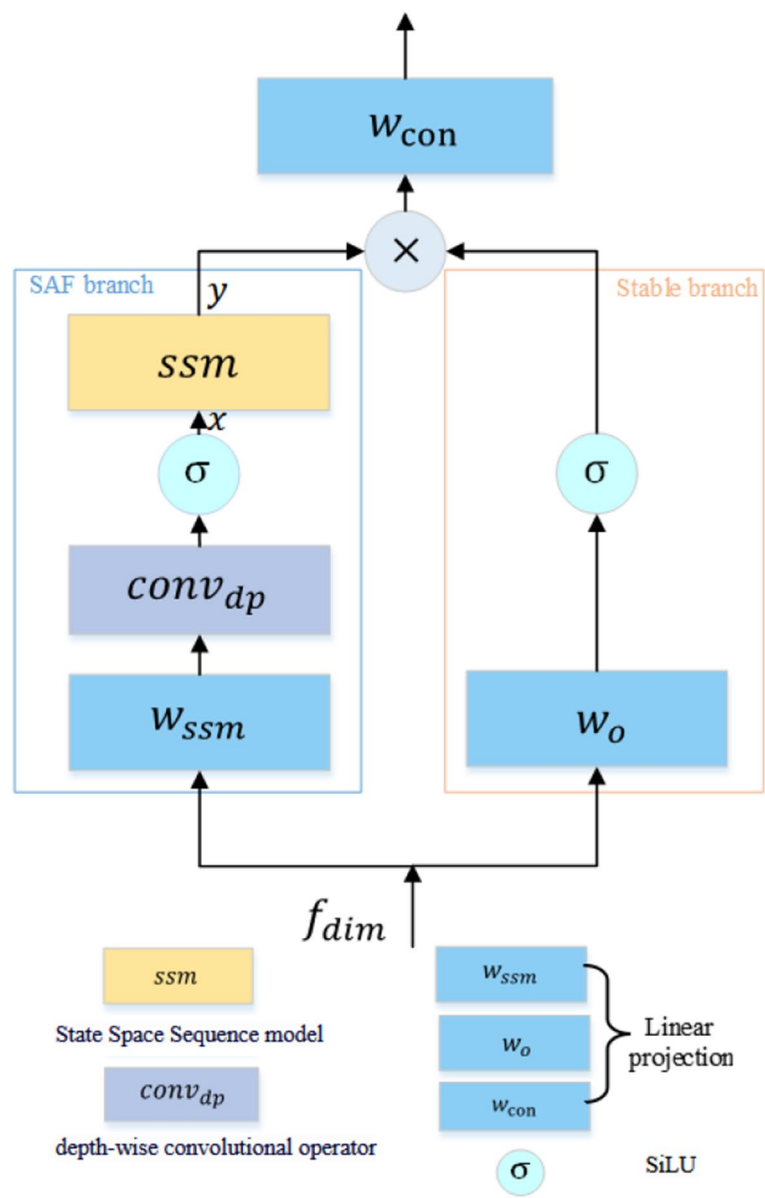
Li *et al. BMC Medical Imaging*   (2025) 25:80
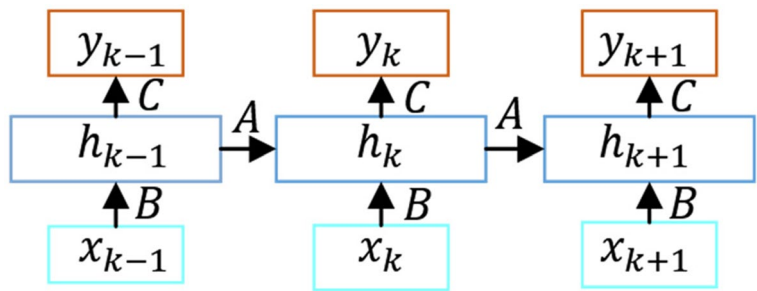
Page 7 of 16

Fig. 3 The detail of Mamba Block

Fig. 4 The recursive updating process of h_k

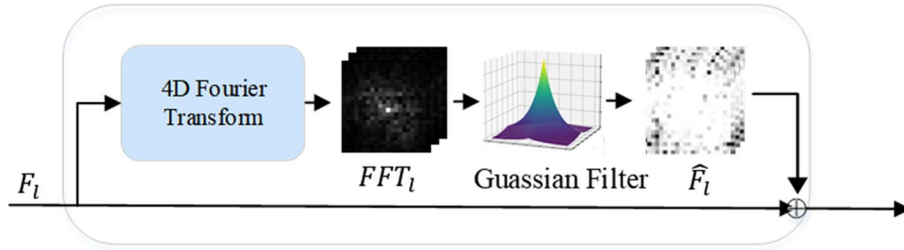Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 8 of 16

Fig. 5 Fourier transform connection

Finally, to fully leverage both spatial and frequency domain features, the original feature map F_l and the high frequency enhanced feature map $\hat{F}_l$ are fused together following Eq. (7).

$$FT(F_l) = \hat{F}_l + F_l \qquad (7)$$

The feature map $FT(F_l)$ is forwarded to the corresponding decoder layer via FT connections as shown in Fig. 5.

GC subtask

With the aforementioned edge features enhancement and feature map enhancement through MF block and FT connection, the UIA can be well segmented from the background via the U shape encoder-decoder neural network solving pixelwise classification problem. However, considering that the UIA is a part of the abnormal expansion of the parent vessels, segmentation remains a challenge when considering accurately extract UIA dome from parent vessels. There is actually not a real border described by the variance of image pixels since blood fills the entire cavity of UIA and parent vessel which significantly has negative impact on the accurate delineation of neck part via pure visual perception method. To overcome this problem, it's important to involve the prior knowledge about lesion geometry provided by experienced doctors into the delineation as a compensation. Therefore, in this work we take the advantage of point clouds in characterizing geometry information and utilize extra subtask to constrain our predicted masks to have similar structures with the manual delineations. Working in a multi-task learning mode, it helps the improvement of the neck region delineation.

Instead of pixelwise calculation, the newly added subtask tries to constrain the target wise similarity. In this case, the marching cubes algorithm, originally described by Lorensen and Cline [26], is applied to convert the predicted segmentation mask P and corresponding ground truth G into target described by a set of point clouds. To be specific, let S_p and S_G the surface mesh generated by the marching cubes algorithm from P and G. The vertices

of S_p and S_G are $V_p = \{v_{p1}, v_{p2}, ..., v_{px}\}, v_{pi} \in R^3$ and $V_G = \{v_{g1}, v_{g2}, ..., v_{gy}\}, v_{gi} \in R^3$, which are the target wise representation of P and G, respectively.

To measure the similarity between the structure defined by V_P and the structure of V_G, the Chamfer distance (CD), which is commonly used in point cloud reconstruction task, is utilized here. Given two set of point clouds, CD measures how close these two shapes are to each other by calculating the average nearest-neighbor distance between points in one set to those in the other set. CD effectively reflects the global similarity between V_P and V_G. The subtask defined in PC block is to align V_P to V_G under measure of CD. In another word, the calculated distance in Eq. (8) should be minimized during learning.

$$d(V_P, V_G) = \frac{1}{N_{V_P}} \sum_{p \in V_P} \min_{g \in V_G} \|p - g\|_2^2 + \frac{1}{N_{V_G}} \sum_{g \in V_G} \min_{p \in V_P} \|g - p\|_2^2 \qquad (8)$$

where N_{V_P} is the number of points in V_P, N_{V_G} is the number of points in V_G.

Therefore, the newly added subtask could be incorporated into our framework and supervise the learning of the neural network model, improving the overall performance via multitask learning.

Experiment

Datasets

In this article, we employ the publicly available dataset from reference [3] to validate our approach. The initially released dataset comprises 1338 CTA cases from six institutions, encompassing a variety of conditions, including ruptured and unruptured lesions, as well as a wide range of lesion sizes. In line with our research objectives, we aim to establish a segmentation model which is capable of delineating UIA lesions to assist in the management of unruptured IAs. Hence, we have excluded 516 cases of ruptured CTA from our study. Similar to reference [3], we exclude the 50 negative cases from our study to mitigate the data imbalance that may arise during the model training process. Since aneurysms with a volume less than 0.1 cc are considered to have a negligible risk of

Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 9 of 16

rupture, and clinically, no proactive medical intervention is typically taken for such minimal lesions, we have used 0.1cc as the threshold to exclude all images with volume sizes below this threshold [9]. As a result, the dataset utilized in our study comprises 284 CTA images with 348 lesions, the average lesion per case is 1.22 ± 0.59.

The images are acquired from 11 devices belonging to 3 manufactures, featuring an in-plane resolution of 512×512 and varying slice thickness from 0.6 mm to 1.0 mm. All images have been resampled to a resolution of 1.0 mm $\times$ 1.0 mm $\times$ 1.0 mm. Some of the CTAs contain the neck or heart regions besides head, we manually review each case and remove the non-head regions in order to exclude the potential interference.

Evaluation metrics

We employ two widely recognized metrics commonly used to assess segmentation performance, including Dice coefficient (*Dice*), 95% Hausdorff Distance (*HD*95). *Dice* is usually used to gauge the overlapping between a segmentation prediction mask P and the corresponding ground truth mask G, whose calculation is following Eq. (9). To accurately measure the accuracy of boundary identification and maintain the overall numerical stability, we use *HD*95 to represent the maximum distance between the predicted segmented region boundary and the real region boundary, as shown in Eq. (10).

$$Dice = \frac{2(P \cap G)}{P \cup G} \tag{9}$$

$$HD(S_P, S_G) = \max\left(h(S_P, S_G), h(S_G, S_P)\right) \tag{10}$$

$$h(S_P, S_G) = \max_{m \in S_P} \left\{ \min_{n \in S_G} \|m - n\| \right\}$$
$$h(S_G, S_P) = \max_{n \in S_G} \left\{ \min_{m \in S_P} \|n - m\| \right\}$$

where S_P and S_G represent the surface of mask volume P and G correspondingly.

Implementation details

The development environment is Ubuntu 20.04, python 3.8 and PyTorch 2.30. The experiments are conducted on workstation with Intel Core i7-12700H and two pieces of NVIDIA GeForce GTX 4090. We utilize online data augmentation during training, applying random intensity scaling and positional rotation to the images through the toolbox MONAI with a random probability [27].

We adopt an experimental protocol similar to AttentionUNet. And the dataset is split into training, validation, and test sets with a ratio of 7:2:1. The proposed model has been trained from scratch with initial weights sampled from a random uniform distribution and the total number of training epoch is 200. The Adam algorithm is employed for optimization. The initial learning rate is set to 1×10^{-3}, whose weight decay and beta are set to 1e-4 and (0.9, 0.99). The entire training process is divided into two phases, the warm-up phase and the refinement phase. During the warm-up phase, the pure dice loss is employed to constrain the model to get a coarse segmentation. Otherwise, the GC subtask which highly depends on the geometrical characteristics won't make sense. Once the model converges to a suboptimal solution after a few epochs, the CD loss given in Eq. (8) is added as compensation. The weights for the CD loss and the Dice loss are set to 0.1 and 0.9, respectively. In this paper, the warm-up phase consists of 180 epochs, while the refinement phase comprises 20 epochs.

Compared methods

We compare our MTMU with 7 State of the Art (SOTA) methods, including SwinUNetr [28], UNetr [14], Seg-ResNet [29], U-Net [30], AttentionUNet [31], DynUNet which is the implementation of nnUNet's [32] architecture in MONAI, GLIANet [3]. Among these methods, U-Net, SegResNet, DynUNet, GLIANet are classic convolution-based methods for medical image segmentation. The UNetr and SwinUNetr have demonstrated outstanding performance in medical image segmentation by leveraging transformer architectures. The GLIANet is the model proposed alongside the dataset we have chosen. For the sake of completeness, we have compared our method against it. From an architectural perspective, the GLIANet employ a two-stage framework which is popular in small target segmentation. It utilizes two serial networks, one for localizing the region of interest and the other for segmenting the target. This approach trades off model parameters for performance. Except for it, the rest competitive methods, as well as our proposed MTMU, are all one-stage framework, which produce segmentation mask via one end-to-end model. This structure is incapable of excluding background interference by focusing on the region of interest explicitly. Therefore, models must address the severe class imbalance issue inherent in small-object segmentation tasks implicitly by themselves. The implementation of the SwinUNetr, UNetr, SegResNet, U-Net, AttentionUNet and DynUNet follow the standard code released in MONAI. The GLIANet algorithm is implemented using the original author's version available on GitHub.[1] To ensure fairness, all models are trained on the same data partitioning.

---

[1] MeteorsHub/GLIA-Net: A segmentation network for intracranial aneurysm on CTA images using pytorch (github.com).

Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 10 of 16

Experiment results

We report the quantitative results in Table 1. From which, it could be seen that our method outperformed others in terms of *Dice* (78.47%) and *HD*95 (7.67mm). It is also noteworthy that the *Dice* score of the transformer-based SwinUNetr is higher than that of UNet and GLI-ANet, demonstrating that capturing global information is beneficial for IAs segmentation. The convolution-based methods, namely U-Net, SegResNet, AttentionU-Net and DynUNet, demonstrate similar performance, which indicates that local features are critical for the pixelwise delineation task. However, the transformer-based approaches, UNetr and SwinUNetr, underperform in our experiments. We think that is because the global attention mechanism requires a large number of training data, for example in the thousands, so that to lead its failure. Even though, our model also incorporates global attention via the MF blocks and yield competitive performance, indicating that extracting global attention for long-term dependencies is helpful, but the structure of the feature extraction module needs to be carefully designed. The superior *Dice* score of our method indicates its strong capability in identifying and segmenting target areas. The two approaches that lead in terms of Dice coefficient are our MTMU model and the AttentionUNet. In terms of the Dice coefficient, the former achieves a slight advantage of 0.1 percentage point over the latter. Nevertheless, compared with AttentionUNet, our model achieves significant improvement in the *HD*95 metric. We also calculated the p-value, and the p-value of all methods are less than 0.05, which means segmentation methods are statistically effective. The experiments have demonstrated that our MTMU model can accurately extract the boundaries of the target, showcasing its excellent capability to preserve geometric morphology. We attribute this success to the high-frequency enhanced FT connection and the geometric priors introduced by the GC subtask.

Table 1 Statistical comparisons with various SOTA methods

Model	Dice(%)	HD95(mm)	p-value
SwinUNetr	77.26 ± 0.16	10.36 ± 13.62	1.28e-20
UNetr	69.97 ± 0.21	16.91 ± 15.14	5.46e-29
SegResNet	77.71 ± 0.13	9.03 ± 11.58	2.6e-2
U-Net	77.41 ± 0.15	9.20 ± 12.23	3.1e-2
AttentionUNet	77.92 ± 0.17	9.28 ± 13.83	1.52e-09
DynUNet	77.92 ± 0.19	9.51 ± 12.99	2.2e-2
GLIANet	74.53 ± 0.22	11.45 ± 13.21	2.9e-2
Ours	$\mathbf{78.47} \pm 0.16$	$\mathbf{7.67} \pm 11.69$	6.35e-11

Figure 6 visually presents the segmentation effects of aneurysms by each competitive model. Figure 6(a) demonstrates five different CTA cases and Fig. 6(b) zooms in the region of interest to observe the details of IA targets, from left to right denoted as case I to case V. Figure 6(c)-(j) show the segmentation contours from SwinUNetr, UNetr, SegResNet, U-Net, AttentionUNet, DynUNet, GLIANet and MTMU(ours); the lower right corner of each image displays the Dice coefficient (%). In the segmentation results of Case I, there is evident interference from surrounding blood vessels that confounds the models UNetr, U-Net, AttentionUNet, and DynNet, leading to over-segmentation. Similar situation was also observed in case IV. Over-segmentation also corroborates the poor HD (Hausdorff Distance) metrics of these models as shown in Table 1. For SegResNet, under-segmentation could be observed in multiple cases, which is detrimental to the accurate measurement of morphological parameters, one of our potential important applications. From the Fig. 6, it could be found that our proposed segmentation method MTMU could effectively delineate the boundaries of small targets while preserving their shape information, even when the textural features in their vicinity are highly similar.

In order to clearly observe the geometric shape of the delineated IAs, the corresponding point cloud and mesh format of IAs are shown in Fig. 7, and the corresponding Dice coefficient (%) is located at the bottom right corner of each image. It could be observed that over-segmentation will produce an erroneous lesion shape in 3D space, thereby affecting the measurement of parameters such as lesion diameter and volume, for example the results of UNetr (Fig. 7(c)). Under-segmentation would lead to over-smooth in 3D shape, for example the result of GLIANet on case V(the last column of Fig. 7(h)). From a geometric perspective, our segmentation results validate the effectiveness of our method in extracting boundary details. Additionally, the shapes we obtained are more closely aligned with the ground truth shapes.

Overall, our method has a good capability in extracting richer boundary information for small IA lesion segmentation and producing more similar geometry structure with the ground truth. We believe that the superior performance of our model is attributed to the proposed MF module designed to capture global information, the FT module that enhances high-frequency edge information, and the GC that integrates geometric prior features. These design elements of the model contribute to enhancing its capability in segmenting small targets, achieving precise boundary and shape capture for these minor objectives.

As aforementioned, the design of the segmentation model requires a balance between model size
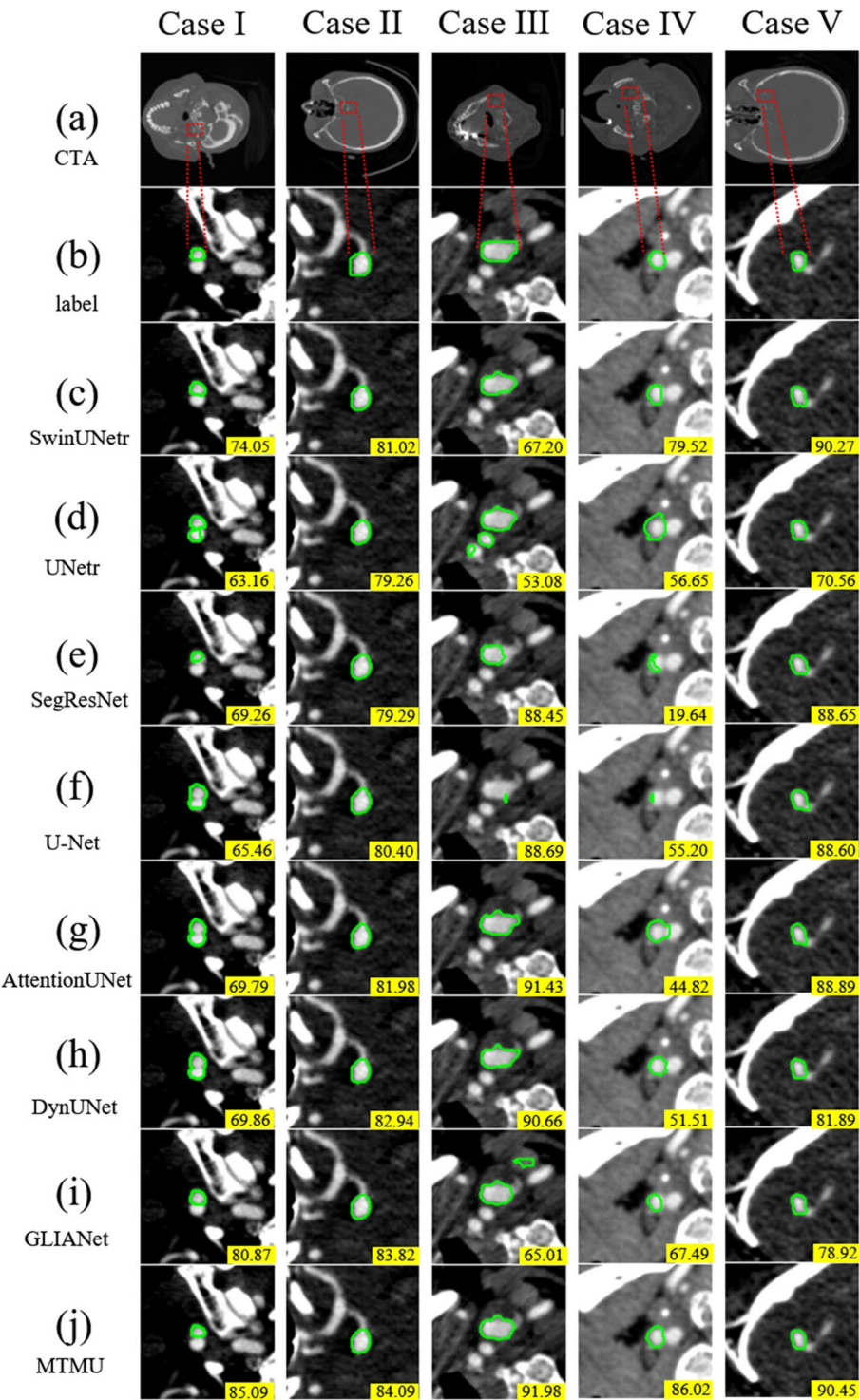
Fig. 6 Visual comparisons with various SOTA methods. Fron top to down are (**a**) original CTAs, (**b**) manual labels, (**c**) SwinUNetr, (**d**) UNetr, (**e**) SegResNet, (**f**) U-Net, (**g**) AttentionUNet, (**h**) DynUNet, (**i**) GLIANet, (**j**) MTMU(ours). The display window is [−138, 302] HU

and performance. The model size will also affect its application prospects to a certain extent, considering the hardware and software support required for its operation. Table 2 details the network parameter counts, the Floating Point Operations (FLOPs), training times and inference times. The parameter count
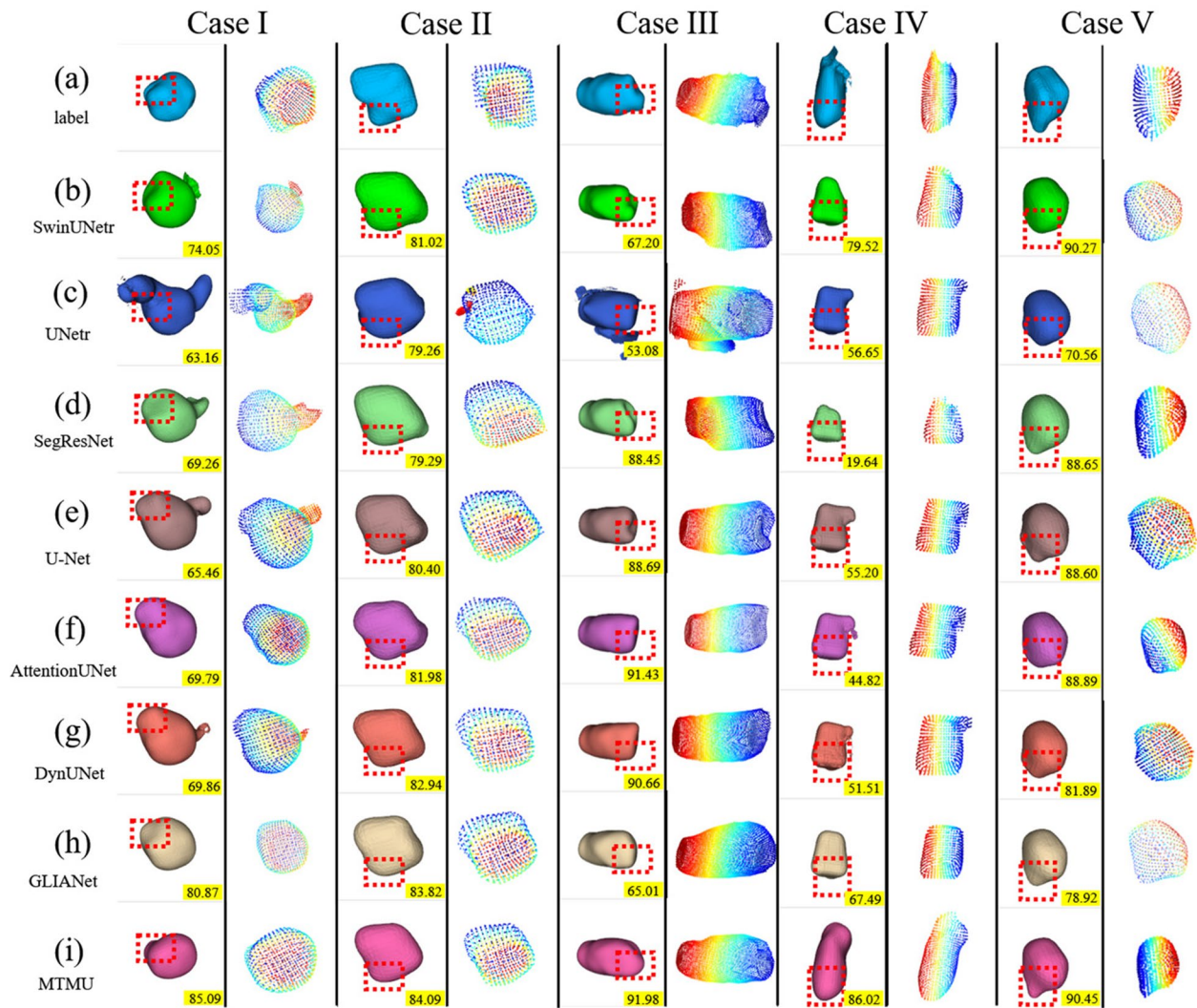
Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 12 of 16

Fig. 7 The geometric structure of the ground truth and the segmentation results. Fron top to down are (**a**) ground truth, (**b**) SwinUNetr, (**c**) UNetr, (**d**) SegResNet, (**e**) U-Net, (**f**) AttentionUNet, (**g**) DynUNet, (**h**) GLIANet, (**i**) MTMU(ours)

Table 2 Statistical comparisons with various SOTA methods on performance metrics

Method	Params(M)	FLOPs(G)	Training Times(s)	Inference Times(ms)
SwinUNetr	61.99	329.29	58.34	53.41
UNetr	101.62	264.66	33.87	36.59
SegResNet	7.16	12.49	8.10	20.26
UNet	1.78	24.27	6.41	20.75
AttentionUNet	5.91	61.77	12.79	20.12
DynUNet	33.09	829.44	22.23	21.45
GLIANet	13.09	814.77	49.74	39.19
Ours	2.10	25.71	31.62	34.25

reflects the model's complexity, while the FLOPs indicate the computational burden. From Table 2, there are only two models have trainable parameters less than 5.0M, UNet and our MTMU. Although our model's parameter count is a little bit larger than that of UNet, our segmentation performance surpassed that of UNet significantly. In terms of FLOPs, the situation is similar. Compared to our approach, AttentionUNet, which is the most competitive in terms of segmentation performance, incurs three times the FLOPs. Our MTMU yields FLOPs of 25.71, which is only slightly lower than UNet and SegRestNet, but our evaluation metrics outperform both of them. Notably, in comparison to the transformer-based SwinUNetr, our model significantly reduces computational cost and complexity. While both

Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 13 of 16

methods are capable of capturing global information, our approach achieves a lower computational load due to the integration of the Mamba unit within the feature extractors. Additionally, we have recorded the training and inference times for each method in Table 2. To measure the training time per epoch, we use a batch size of 4 and record the implementation time for each epoch, calculating the average value. For the inference time per sample, we use a batch size of 4, record the implementation time for each sample, calculating the average value. In the training phase, the proposed method takes longer per epoch compared to the pure CNN based model, we think the convolutional operations are deeply optimized on modern hardware. They make efficient use of GPU parallel computing and memory access, thus offering better performance. Ours is faster than the transformer-based methods UNetr and SwinUNetr. Among the compared models, SwinU-Netr's segmentation performance is closest to that of the proposed method, although it incurs a significant increase in both training and inference time. We think that our method has a lower FLOPs compared to Transformer-based methods, but it significantly reduces the number of parameters. Consequently, it entails less computational load, making it faster than the transformer-based approaches. Among CNN based models, the use of custom operators or network architectures may lead to inconsistencies in the correlation between FLOPs, parameters, training times and inference times. In brief, the proposed MTMU demonstrates superior segmentation performance with relatively low computational load and model complexity.

Ablation Studies

The MTMU is designed with three modules to improve the end-to-end segmentation performance from different perspectives, including MF block, FT connection and GC subtask. To demonstrate their effectiveness, we show the corresponding segmentation results with different combinations in Table 3. Based on the quantitative metrics, it could be concluded that as the three modules are ensembled, the performance of the segmentation model gradually enhances. It proves the effectiveness of our model's design.

Interpretability of the model

Although the data-driven deep learning models are usually treated as a dark box, we try to interpret our models via observing the changes in features obtained during model's convergence. Figure 8 illustrates the effect of the FT connection on the obtained feature maps. We visualize three different IA cases, whose features enhanced by FT connection as well as that

Table 3 Ablation study investigating effectiveness of different modules

#	MF	FT	GC	Dice(%)	HD95(mm)
1	√			78.07 ± 0.16	8.27 ± 12.14
2		√		77.46 ± 0.18	8.28 ± 11.94
3			√	77.51 ± 0.16	10.60 ± 13.36
4	√	√		78.13 ± 0.13	9.26 ± 12.37
5	√		√	78.09 ± 0.16	8.43 ± 12.02
6		√	√	77.69 ± 0.16	11.10 ± 14.13
7	√	√	√	**78.47** ± 0.16	**7.67 ± 11.69**

without FT are shown. For visualization, we plot the red contour around the region with the highest grayscale value, which corresponds to the area of primary attention of the network, as well as surrounding regions with grayscale values above 1000. It is evident that in the feature maps enhanced by the FT connection (Fig. 8), target regions with high feature values are more concentrated. In another word, these feature maps would play a more active role in accurately delineating IAs. Whereas in the feature maps without frequency enhancement, the distribution of high value feature points is more scattered. As a result, it increases the difficulty for the network to focus on useful features. This explains the reason why our method has a stronger capability to converge to the boundaries.

To illustrate the working process of the GC subtask, we also tracked the changes in IA shape during the refinement phase of training. Figure 9 shows that after introducing the GC subtask, the geometric structure of the network prediction gradually approximates that of the ground truth mask (Fig. 9(a)). To facilitate the observation of geometric shape changes, we perform mesh reconstruction on the model output. And the closer the color is to that of the ground truth, the better the segmentation performance is. As shown in Fig. 9(a), the ground truth color is dark blue, the color in (b), (c), and (d) transitions from light blue to dark blue, simulating the process where the model's prediction gradually approaches the ground truth as training progresses. And, in the lower right corner of each figure, the Dice coefficient is displayed. Notably, the right half of the aneurysm structure increasingly protrudes, closely resembling the ground truth. Meanwhile, the creases on the surface of IA lesion were gradually fitted to ground truth. These creases are precisely the regions where geometric features are most typical. It explains the reason that our model could maintain the geometric features of the target during segmentation process. Incorporating geometric priors via the GC subtask helps our model align the structure of the delineated target more closely with the actual target.
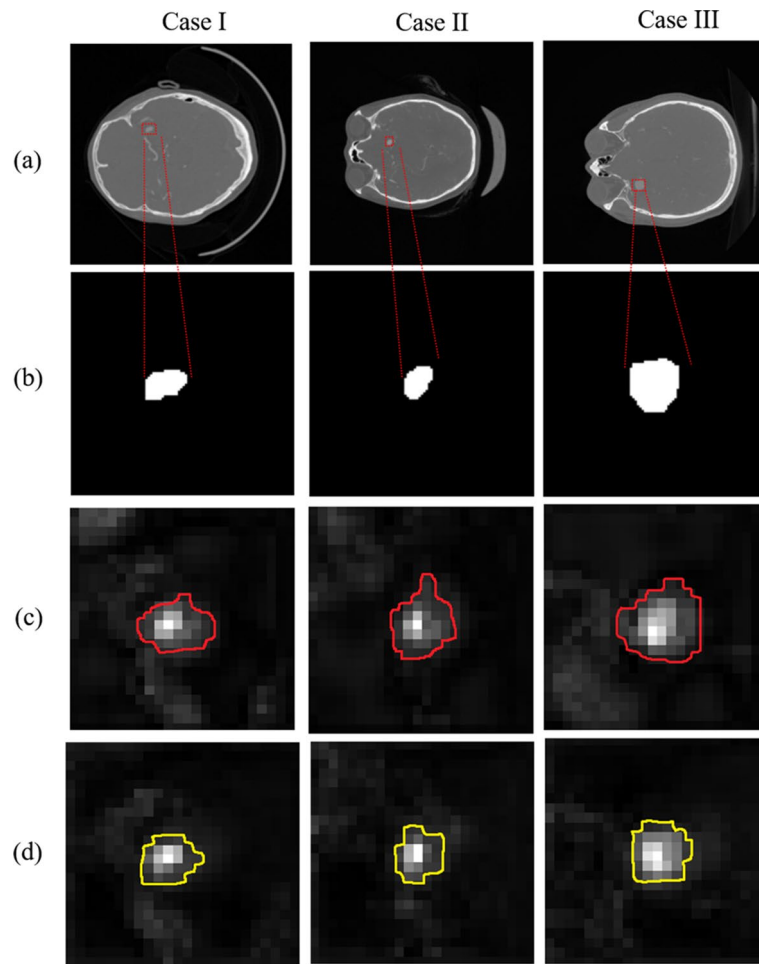
Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 14 of 16

Case I   Case II   Case III

Fig. 8 Illustrates three cases of the feature maps with and without FT enhancement. In this figure, (**b**) represents the ground truth, (**c**) shows the feature map without FT enhancement, and (**d**) displays the feature map with FT enhancement
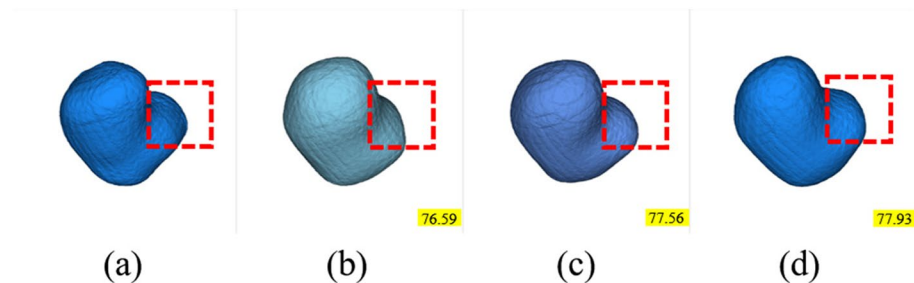

(a)   (b)   (c)   (d)

Fig. 9 In the geometric visualization of segmentation results across different epochs, (**a**) represents the ground truth, while (**b**), (**c**), and (**d**) represent the mesh for epochs 180, 190, and 200, respectively

Discussion

We propose a semantic segmentation model named MTMU for the automatic delineation of intracranial aneurysm lesions from the CTA scans. Comparing with other medical targets, intracranial aneurysm lesions are usually very tiny, which leads an obstacle for the feature extracting. It is because in most cases image features are computed from the foreground pixels, and if the number of available pixels is limited the feature computation will lose accuracy. That's the common problem the small

target delineation algorithms have to face. Another challenge is that the pixel intensities of the IA region are similar to those of the parent vessel region, as both the aneurysm and the vessel are filled with blood. It leads to the difficulty in accurately identifying the IA neck. To solve these issues meanwhile balance the performance and the cost of hardware and software, we employ MF block consisting of multi-orientation Mamba units to construct the feature encoders. It equips the model with the capability in perceiving long-range dependencies while avoiding the introduction of excessive additional computational workload. FT connection which aims to enhance the edge information in feature maps is proposed and used to replace the traditional skip connections of the two deepest layers. Empirically speaking, the deeper the layer is, the less spatial information are kept in feature maps. Therefore, enhancing the edge information of the deep features could significantly improve the segmentation performance. The GC subtask, which focuses on the consistence of the target wise geometric characteristics, is used as compensation to constrain the training of the whole model. With the help of it, the proposed model performs well on the delineation of IA contour, especially on the delineation of neck boundary, so that to yield better segmentation metrics. These contributions are proved to be effective in handling the IA segmentation task.

Extensive experiments have been carried out to demonstrate the superiority of the proposed framework. We compare our model with those popular benchmark semantic segmentation models, namely convolution-based ones and transformer-based ones. From Table 1, our results surpass the rest competitive models on both Dice score(78.47%) and 95% Hausdorff Distance(7.67mm). Taking computational cost into consideration (Table 2), the proposed model balances the cost and the segmentation performance best. It only pays one third resources of the AttentionUNet, which wins the second place in segmentation performance. From the visualization of Fig. 6, it could be observed that our model delineates the contours that are most close to the ground truth, avoiding over- and under-segmentation. To be specific, the Fig. 7 shows the 3D view of the segmentation results of each case. Our method achieves a high degree of consistency in lesion shape compared to manual annotations, especially on the regions with significant geometric variations. From the ablation study (Table 3), the MF block contributes to basic segmentation performance, with five per mille improvement of Dice score. With the help of FT connection and GC subtask, the performance will be further improved, which prove the contributions of this paper are solid.

In brief, our method balances the performance and the computational cost so that to make its clinical application possible. It could be helpful in assessing the IA appearance parameters. For example, with the 3D segmentation masks, the mesh view of IA lesion could be reconstructed (as shown in Fig. 7). The geometric characteristics such as volume size and diameters are easy to be measured, avoiding any subjective bias. Meanwhile, once trained well, the model inference speed is far faster than manual work, which could help accelerate clinical workflow. The model also has advantages on model size and the requirement of hardware and software is easy to meet. Both of these characteristics are beneficial for the potential clinical applications of the model.

Although our method demonstrates the superior performance, its limitations should also be noted. By integrating the MF block, our algorithm gains the ability to capture global information while maintaining low parameters count and achieving higher FLOPs. However, when introducing the GC task, the process of converting volume into point clouds using the marching cubes algorithm and adding labels to the point clouds is performed on the CPU, resulting in slower runtime and significantly affecting the model's training speed. In the future, we plan to accelerate the GC task by offloading the point clouds conversion process to the GPU, thereby improving the overall training efficiency.

Conclusions
In this paper, we proposed the MTMU network to achieve the task of automatically delineating IA lesions. Our network architecture incorporates Mamba-based MF blocks to enhance the capability of capturing global contextual information with low computation cost. Additionally, to increase sensitivity to boundaries, we introduced FT blocks to make our model focus more on high-frequency information. Moreover, GC subtasks were introduced to further constrain the geometric structure of our segmentation results, making them more closely aligned with the ground truth. Experiments show that MTMU network has outstanding performance in IA segmentation.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01611-6.

Supplementary Material 1

Authors' contributions
Study conception and design: B.L., Y.L.; draft manuscript preparation: B.L., Y.L.; All authors reviewed the results and approved the final version of the manuscript.

Li *et al. BMC Medical Imaging*   (2025) 25:80

Page 16 of 16

References

1. Lauric A, Baharoglu MI, Malek AM. Ruptured Status Discrimination Performance of Aspect Ratio, Height/Width, and Bottleneck Factor Is Highly Dependent on Aneurysm Sizing Methodology. Neurosurgery. 2012;71(1):38–46.
2. Juvela S, Poussa K, Porras M. Factors affecting formation and growth of intracranial aneurysms. Stroke. 2001;32(2):485–91.
3. Bo ZH et al. Toward human intervention-free clinical diagnosis of intracranial aneurysm via deep neural network. Patterns. 2021;2(2):100197.
4. Zhang X, Li X, Feng Y. A medical image segmentation algorithm based on bi-directional region growing. Optik. 2015;126(20):2398–404.
5. Manniesing R, Velthuis BK, van Leeuwen MS, van der Schaaf IC, van Laar PJ, Niessen WJ. Level set based cerebral vasculature segmentation and diameter quantification in CT angiography. Medical Image Analysis. 2006;10(2):200–14.
6. Zhang L, Zhang K, Pan H. SUNet++: A Deep Network with Channel Attention for Small-Scale Object Segmentation on 3D Medical Images. Tsinghua Science and Technology. 2023;28(4):628–38.
7. Chen J et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," p.arXiv:2102.04306. https://doi.org/10.48550/arXiv.2102.04306. Available: https://ui.adsabs.harvard.edu/abs/2021arXiv210204306C. Accessed on 01 Feb 2021.
8. Pu Q, Xi Z, Yin S, Zhao Z, Zhao L. Advantages of transformer and its application for medical image segmentation: a survey. BioMed Eng OnLine. 2024;23(1):14.
9. Brown RD, Broderick JP. Unruptured intracranial aneurysms: epidemiology, natural history, management options, and familial screening. Lancet Neurol. 2014;13(4):393–404.
10. Wiebers DO, et al. "Unruptured intracranial aneurysms: natural history, clinical outcome, and risks of surgical and endovascular treatment," (in eng). Lancet. 2003;362(9378):103–10.
11. Podgorsak AR, et al. "Automatic radiomic feature extraction using deep learning for angiographic parametric imaging of intracranial aneurysms," (in eng). J Neurointerv Surg. 2020;12(4):417–21.
12. Bizjak Ž, Likar B, Pernuš F, Špiclin Ž. Vascular surface segmentation for intracranial aneurysm isolation and quantification. p.arXiv:2005.14449. https://doi.org/10.48550/arXiv.2005.14449. Available: https://ui.adsabs.harvard.edu/abs/2020arXiv200514449B. Accessed on 01 May 2020.
13. Park A, et al. Deep learning-assisted diagnosis of cerebral aneurysms using the HeadXNet Model. JAMA Netw Open. 2019;2(6):e195600.
14. Hatamizadeh A et al. UNETR: Transformers for 3D Medical Image Segmentation," p.arXiv:2103.10504. https://doi.org/10.48550/arXiv.2103.10504. Available: https://ui.adsabs.harvard.edu/abs/2021arXiv210310504H. Accessed on 01 Mar 2021.
15. Zhao Y, et al. "ConTNet: cross attention convolution and transformer for aneurysm image segmentation," in. IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2023;2023:3618–25.
16. Gu A, Dao TJ. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. p.arXiv:2312.00752. https://doi.org/10.48550/arXiv.2312.00752. Available: https://ui.adsabs.harvard.edu/abs/2023arXiv231200752G. Accessed on 01 Dec 2023.
17. Zhu L, Liao B, Zhang Q, Wang X, Liu W, Wang XJ. Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. p.arXiv:2401.09417. https://doi.org/10.48550/arXiv.2401.09417. Available: https://ui.adsabs.harvard.edu/abs/2024arXiv240109417Z. Accessed on 01 Jan 2024.
18. Wang Z, Zheng JQ, Zhang Y, Cui G, Li LJ. Mamba-UNet: UNet-Like Pure Visual Mamba for Medical Image Segmentation. p.arXiv:2402.05079. https://doi.org/10.48550/arXiv.2402.05079. Available: https://ui.adsabs.harvard.edu/abs/2024arXiv240205079W. Accessed on 01 Feb 2024.
19. Ma J, Li F, Wang BJ. U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation," p.arXiv:2401.04722. https://doi.org/10.48550/arXiv.2401.04722. Available: https://ui.adsabs.harvard.edu/abs/2024arXiv240104722M. Accessed on 01 Jan 2024.
20. Zhang M, Yu Y, Gu L, Lin T, Tao XJ. VM-UNET-V2 Rethinking Vision Mamba UNet for Medical Image Segmentation. p.arXiv:2403.09157. https://doi.org/10.48550/arXiv.2403.09157. Available: https://ui.adsabs.harvard.edu/abs/2024arXiv240309157Z. Accessed on 01 Mar 2024.
21. Qin Z, Zhang P, Wu F, Li XJ. FcaNet: Frequency Channel Attention Networks. p.arXiv:2012.11879. https://doi.org/10.48550/arXiv.2012.11879. Available: https://ui.adsabs.harvard.edu/abs/2020arXiv201211879Q. Accessed on 01 Dec 2020.
22. Yang Y, Soatto SJ. FDA: Fourier Domain Adaptation for Semantic Segmentation. p.arXiv:2004.05498. https://doi.org/10.48550/arXiv.2004.05498. Available: https://ui.adsabs.harvard.edu/abs/2020arXiv200405498Y. Accessed on 01 Apr 2020.
23. Huang H, et al. Medical Image Segmentation With Deep Atlas Prior. IEEE Trans Med Imaging. 2021;40(12):3519–30.
24. Zheng H et al. Semi-supervised Segmentation of Liver Using Adversarial Learning with Deep Atlas Prior. in Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, Cham, 2019:148–156: Springer International Publishing.
25. You X, He J, Yang J, Gu YJ. Learning with Explicit Shape Priors for Medical Image Segmentation. p.arXiv:2303.17967. https://doi.org/10.48550/arXiv.2303.17967. Available: https://ui.adsabs.harvard.edu/abs/2023arXiv230317967Y. Accessed on 01 Mar 2023.
26. Lorensen WE, Cline HE. Marching cubes: A high resolution 3D surface construction algorithm. J SIGGRAPH Comput Graph. 1987;21(4):163–9.
27. Cardoso MJ et al. MONAI: An open-source framework for deep learning in healthcare. 2022;abs/2211.02701.
28. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. International MICCAI brainlesion workshop. 2021:272–284: Springer.
29. Badrinarayanan V, Kendall A, Cipolla RJ. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. p.arXiv:1511.00561. https://doi.org/10.48550/arXiv.1511.00561. Available: https://ui.adsabs.harvard.edu/abs/2015arXiv151100561B. Accessed on 01 Nov 2015.
30. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger OJ. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. p.arXiv:1606.06650. https://doi.org/10.48550/arXiv.1606.06650. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv160606650C. Accessed on 01 June 2016.
31. Oktay O et al. Attention U-Net: Learning Where to Look for the Pancreas. p.arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999. Available: https://ui.adsabs.harvard.edu/abs/2018arXiv180403999O. Accessed on 01 Apr 2018.
32. Isensee F et al. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. p.arXiv:1809.10486. https://doi.org/10.48550/arXiv.1809.10486. Available: https://ui.adsabs.harvard.edu/abs/2018arXiv180910486I. Accessed on 01 Sept 2018.