Applied Intelligence (2026) 56:124
https://doi.org/10.1007/s10489-026-07121-4

SFJD-Net: spatial-frequency domain joint feature enhancement with differential learning for brain stroke segmentation

Xin Huang[1] · Yu Zhu[1] · YaTong Liu[1] · Yuhao Zhang[2,3,4]

Received: 12 August 2025 / Accepted: 19 January 2026

Abstract

Brain stroke is a major cause of disability and death, and accurate lesion segmentation is essential for early diagnosis and treatment planning. Although CT and MRI provide critical diagnostic information, the large variations in lesion appearance and the noise introduced by manual annotations make precise segmentation challenging. To address these issues, we propose SFJD-Net, a novel Stroke lesion segmentation network that leverages joint spatial-frequency domain feature enhancement and differential learning. SFJD-Net consists of three core modules: Multi-Branch Convolution Attention Encoder (MBCAE), Spatial-Frequency domain Joint feature Enhancement (SFJE), and Differential Learning Decoder (DLD). Compared with the traditional U-Net architecture, SFJD-Net introduces shallow edge information into deep semantic features to enhance texture and boundary representation. The MBCAE module adaptively captures multi-scale lesion features to enrich representations. The SFJE module enhances feature representations from both the spatial and frequency domains, which integrates positional cues and structural details to guide the network in focusing more accurately on target regions. The DLD module uses the reconstructed convolution kernel to record the differences between encoder and decoder, and integrates it into the decoding process through convolution operation, which reduces the semantic gaps and the probability of misjudgment of decoder. Extensive experiments on the published Ischemic Stroke Lesion Segmentation (ISLES) 2022 and 2018 datasets demonstrate that our method achieves state-of-the-art performance. In addition, SFJD-Net is successfully migrated to the pancreas cancer segmentation task of the MSD Cancer dataset, which fully proves that the network has a certain generalization ability.

Keywords Medical image segmentation · Brain stroke · Encoder · Decoder

1 Introduction

The brain, as the central controller of the human body, is responsible for regulating various physiological and psychological functions[1–3]. Its physiological activities and pathological changes are closely related to human life. Stroke is one of the leading causes of death and disability worldwide, posing a significant threat to human health[4–6]. With the help of automatic computer segmentation technology, doctors can effectively locate the region of brain stroke lesions, which has very important application significance[7]. Figure 1 shows the annotation of brain stroke from the public ISLES 2022 Dataset, where the red area represents the lesions, and it can be seen that the volume occupied by the stroke region is very small, which puts higher requirements on the performance of the segmentation algorithm.

U-shaped networks have become the mainstream framework for medical image segmentation. Owing to their

✉ Yu Zhu
zhuyu@ecust.edu.cn

✉ Yuhao Zhang
zhang.yuhao@zs-hospital.sh.cn

1 School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

2 Department of Neurology, Zhongshan Hospital, Fudan University, Shanghai 200032, China

3 National Clinical Research Center for Interventional Medicine, Shanghai 200032, China

4 Shanghai Clinical Research Center for Interventional Medicine, Shanghai 200032, China

Published online: 03 March 2026

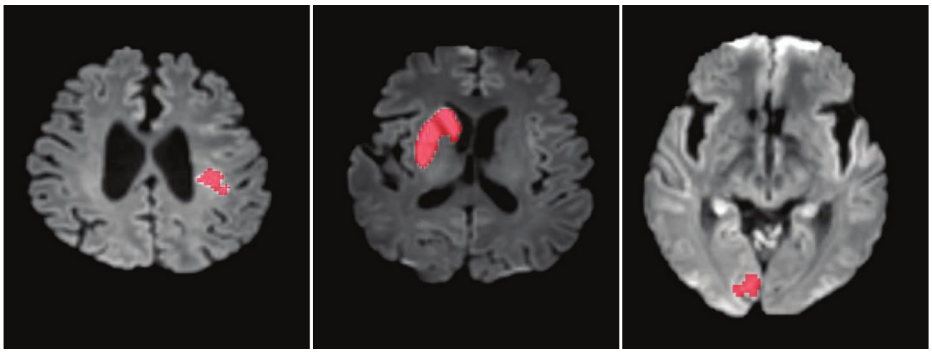


Fig. 1 The annotation of brain stroke from the public ISLES 2022 Dataset for different cases

encoder-decoder architecture and skip connections, they have been widely applied to various medical segmentation tasks, including brain stroke segmentation. However, the inherent semantic gap between shallow and deep features often introduces redundancy and noise through skip connections, which leads to boundary ambiguity or miss small lesions during decoding [8–11]. Although existing studies have attempted to alleviate this issue by improving skip connections or optimizing decoder structures, such as incorporating structural dependencies, enhancing multi-scale features, and introducing context modeling modules [12–14], these methods have achieved impressive results in general medical image segmentation tasks. However, they still struggle to deliver stable and accurate performance for brain stroke segmentation due to the high variability, low contrast, and blurry boundaries of stroke lesions. Therefore, it is necessary to design a specialized network architecture tailored to stroke segmentation to improve lesion delineation accuracy.

In recent years, some scholars in the field of medical image segmentation have taken into account the role of signals in the frequency domain and made certain breakthroughs [15]. For example, Nam et al. [16] captured the contour information of a target by combining its multi-frequency features, and successfully completed various medical image segmentation tasks. Chen et al. [17] designed a novel Multi-scale frequency domain filter to separate high-frequency details from low-frequency structural components, and combined with multi-scale dilated convolution, effectively improved the sensitivity of the model to the shape and boundary of lesions in medical images. The core advantage of these approach combining frequency-domain signals lies in its ability to separate structural components and detailed information of features, which enhances the model's perception of lesion morphology and boundaries. This offers a novel perspective for addressing the core challenges in stroke segmentation.

Motivated by the aforementioned methodologies, we propose a new stroke segmentation network called SFJD-Net, which is specifically designed to mitigate the semantic discrepancy issue in U-shaped architectures and enhance the fine-grained segmentation of stroke lesions. SFJD-Net consists of five layers. It covers two novel strategies: the Spatial-Frequency domain joint feature enhancement and differential learning. In the frequency domain, the amplitude multiplication and phase addition of deep and shallow layer features are used to enhance the signal of the same frequency point. In the spatial domain, a lightweight axial attention mechanism is used to pay attention to the position relationship between the segmented target and the surrounding tissue, so as to obtain a more comprehensive feature expression. In the process of decoding, the convolution kernel describing the difference of codecs is reconstructed to alleviate the ambiguity of skip connection.

The main contributions of this work can be summarized as follows:

(1) A multi-branch convolutional attention encoder (MBCAE) is designed, which uses a lightweight convolutional attention mechanism to optimize the receptive field of the network, and learns the scale information under different dilated rates to weight the encoded features, so as to obtain a richer feature expression.
(2) A novel feature enhancement module called SFJE is developed. With the help of fourier transform and axial attention mechanism, signals in the spatial domain and frequency domain are processed respectively to realize the effect of feature enhancement, making it easier for the network to pay attention to the information of key regions.
(3) A novel differential learning decoder (DLD) is designed, reconstructing a convolution kernel according to the difference between encoder and decoder, then integrating it into the decoding features by convolution operation to decrease the semantic gap between encoder and decoder.

The remainder of this paper is organized as follows: Section 2 briefly reviews the related work. Section 3 describes the architecture of the proposed SFJD-Net in detail. Section 4 presents the experimental settings and results. Finally, discussions and conclusions are provided in Sections 5 and 6, respectively.

2 Related work

2.1 Brain stroke segmentation

With the continuous development of medical imaging technology, multimodal medical images such as computed tomography (CT), magnetic resonance imaging (MRI), and CT perfusion (CTP) have been widely used in the clinical diagnosis of stroke[18]. In recent years, researchers have proposed a variety of automatic segmentation methods for stroke based on multimodal inputs to improve the accuracy and robustness of lesion identification.

Zhou et al.[19] proposed D-UNet, which innovatively fuses 2D and 3D convolutional structures at the encoding stage to effectively improve the segmentation of stroke lesions. Liu et al.[20] designed the Pool-UNet, which combines the Poolformer with a convolutional neural network and introduces the DSE-ResNet module to model the spatial channel correlation. Liu et al.[21] proposed the HCSNet, which is designed to precisely identify and segment small stroke lesions in MRI scans. Kuang et al.[22] designed a hybrid CNN-Transformer model to enhance the feature representation of stroke lesions through feature interaction and cross-modal mechanisms.

In addition, Shi et al.[23] designed C2MA-Net, which constructs cross-modal multi-group attention modules to enhance the information interaction between different modalities to improve the segmentation performance of stroke lesions. Liu et al.[24] proposed a dual-branch dense CNN with multi-kernel and dropout strategies for stroke MRI segmentation. Zhang et al.[25] proposed MDANet, which integrates a difference-aware module with a graph convolution mechanism to enhance the modeling capability of multimodal differences. de Vries et al.[26] proposed a symmetry-aware spatio-temporal segmentation model that can directly segment strokes from CT perfusion raw data and achieves good performance.

2.2 Deep-learning-based mainstream segmentation models

With the continuous advancement of deep learning, medical image segmentation techniques have achieved significant improvements. The mainstream methods can be broadly categorized into three classes: convolutional neural network (CNN), Transformer, and Mamba. In addition, with the development of foundation vision models, segmentation approaches based on SAM, CLIP, and KAN have also attracted increasing attention.

CNNs is built on local receptive field modeling and extract local image features through convolutional kernels. It have been widely applied to various computer vision tasks and have achieved excellent performance[27]. U-Net[28], a milestone model, introduces an encoder-decoder architecture with skip connections. This design effectively fuses deep semantic features with shallow detail features. Building on U-Net, researchers have developed various improved models, such as Attention U-Net[29], ResUNet[30], and UNet++[31], which further enhance cross-scale feature interaction and feature representation.

In order to overcome the problem of limited receptive field in CNN networks, Transformer[32] uses the attention mechanism to effectively model the global dependency between features. Transformer was initially developed in the fields of natural language processing and later introduced into many image processing tasks. Swin-Unet[33] is the first U-shaped segmentation network based on pure Transformer, consisting of encoder, bottleneck layer, skip connection and decoder. Researchers Cao et al.[33] conducted a large number of experiments on multi-organ and heart segmentation tasks, confirming its good segmentation performance and robustness. To fully combine the local features and global features of the input image for decoding prediction, many researchers combined CNN and Transformer to design a series of excellent segmentation models such as TransUNet[34] and TransFuse[35].

In 2023, Albert Gu et al.[36] designed Mamba architecture with the same long-distance modeling capability as the State Space Model[37], becoming a strong competitor of Transformer. After that, some scholars introduced the mamba model into the field of medical image segmentation. Ma et al.[38] designed the U-mamba network based on a hybrid CNN-SSM module to build a comprehensive capture of useful features, and demonstrated impressive performance in a variety of biomedical image segmentation tasks.

In recent years, emerging methods represented by SAM[39], KAN[40], and CLIP[41] have introduced new approaches to medical image segmentation. Methods based on SAM typically adapt the general segmentation model to medical imaging scenarios by designing task-specific prompts, such as points, bounding boxes, or coarse masks. This enables efficient segmentation under zero-shot or few-shot conditions, significantly reducing manual annotation costs and enhancing model transferability. Methods based on CLIP leverage the visual-language alignment mechanism, which incorporates textual semantic information as prior knowledge into the segmentation process. Through cross-modal feature fusion, they improve the semantic discriminability of lesion regions and the accuracy of boundary localization. Meanwhile, methods based on KAN employ Kolmogorov-Arnold Networks to replace or complement traditional CNN and Transformer structures, better modeling the complex nonlinear relationships in medical images and thereby enhancing feature representation and segmentation performance.

Different from the above methods, the algorithm proposed in this paper starts from the spatial domain and the frequency domain respectively. It interacts fully according to the different concerns of the deep and shallow layers of information, so as to obtain more accurate feature expression and achieve better segmentation effect.

3 Method

In this part, an overview of the proposed SFJD-Net is presented in Section 3.1. The key modules of SFJD-Net: the Multi-Branch Convolutional Attention Encoder (MBCAE) module, the Spatial-Frequency domain joint feature enhancement (SFJE) module and the Differential Learning Decoder (DLD) are introduced in Sections 3.2, 3.3 and 3.4 respectively. Finally, the loss function is described in Section 3.5.

3.1 Overview

The framework of the brain stroke segmentation network SFJD-Net is shown in Fig. 2. SFJD-Net consists of five layers. Considering the shallow feature contains more details in the coding process, we added a new branch specially for extracting the shallow texture details, in order to assist the U-shaped network in judging the target region. MBCAE includes a lightweight convolutional attention mechanism to extract higher-quality encoding feature representations, setting different dilate rates for sensing scale changes during coding. Specifically, the features from the first layer after downsampling are extracted by the Sobel edge operator to obtain reliable edge features, and the edge features are fused with the deep semantic information extracted from MBCAE4 by the SFJE module. We mainly use the downsampling operation to adjust the size of the two features to maintain consistency. For the SFJE module, in the frequency domain, the amplitudes of the deep and shallow layers are multiplied and the phases are added to enhance the energy of the target region at the same frequency. In the frequency, the spatial position relationship of width and height is established by using axial attention. Finally, the fused feature processed by the SFJE module is sent to the decoder. In order to make the decoder make full use of the encoding features and reduce the semantic gap between the encoder and the decoder, an independent convolution kernel is designed through the DLD module. It dynamically learns the difference between the encoder and the decoder, and the difference is integrated into the decoding process to reduce noise interference and obtain more stable and effective prediction results. The Sobel edge extraction feature part belongs to a parallel branch of the second layer coding stage, and CBR represents a series of operations including 3×3 convolution, Batch normalization and ReLU activation.

3.2 Multi-branch Convolutional Attention Encoder

The structure of the Multi-branch Convolutional Attention Encoder (MBCAE) is shown in Fig. 3. Due to the large differences in the morphology of stroke lesions at different scales, the MBCAE module adopts multi-branch dilated convolution operation to introduce rich scale information. To further improve the efficiency of the multi-branch convolution process, a more lightweight deep separable calculation mode is adopted for the dilated convolution.

The module extracts local detail features initially through 3×3 convolution and then obtains multi-scale

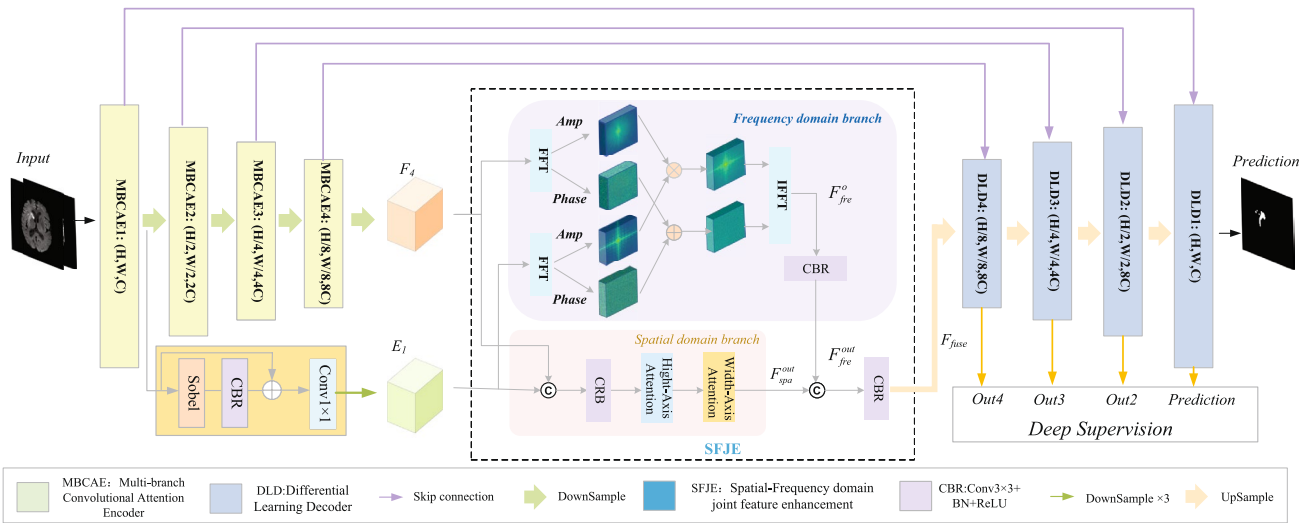


Fig. 2 The Network architecture of the proposed SFJD-Net

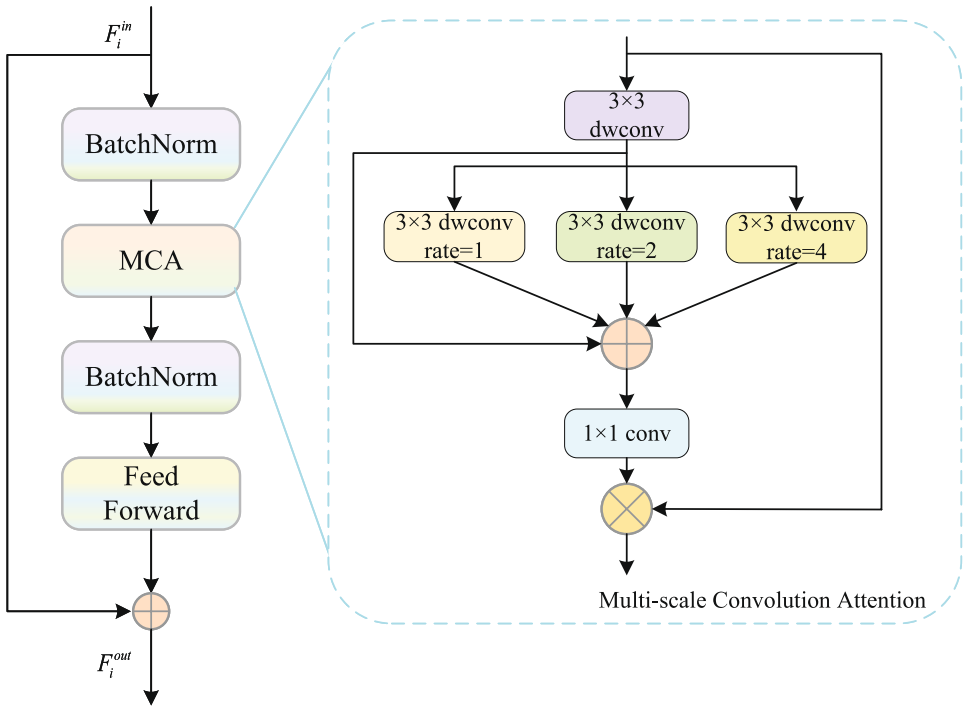


Fig. 3 The architecture of Multi-branch Convolutional Attention Encoder

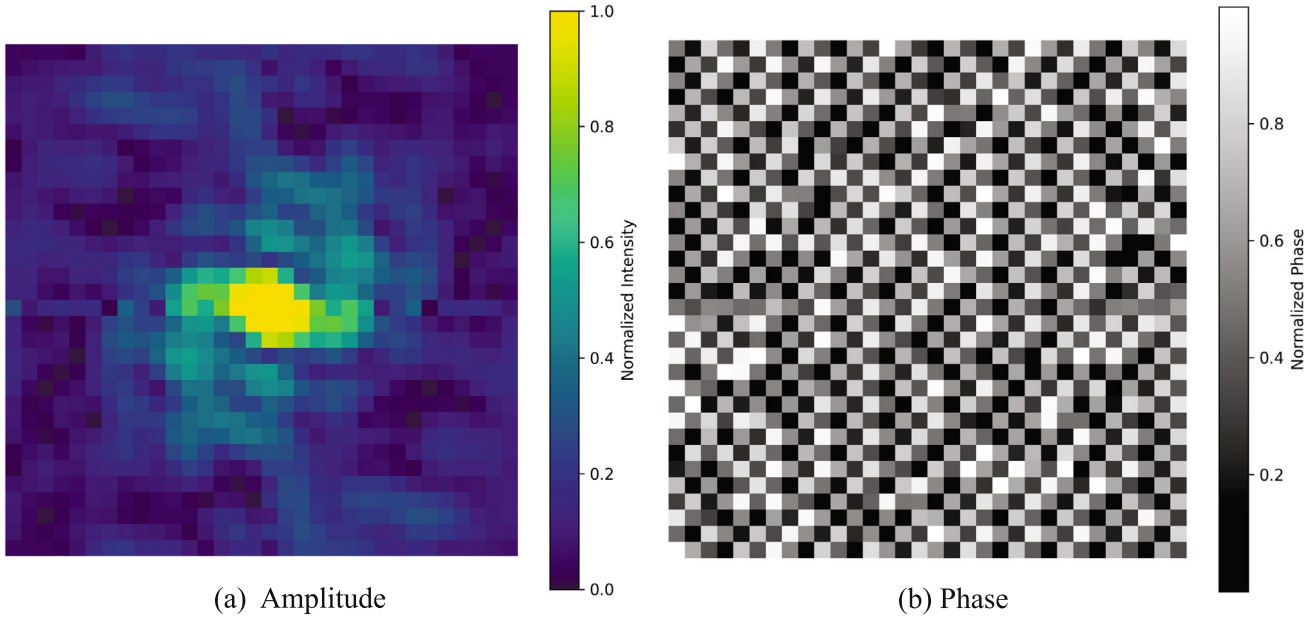
features through three dilated convolution branches with dilation rates of 1, 2, and 4, respectively. After that, the outputs of the three branches are added element by element to aggregate scale-aware information. To further model the interaction among multi-scale channels, a 1×1 convolution is applied to compress and fuse the aggregated features. This operation produces a single-channel attention weight map, where each pixel reflects the relative importance of the corresponding encoder feature based on the learned multi-scale context. Finally, this attention map is multiplied with the encoder features to adaptively emphasize stroke-relevant regions. The main formula is provided in (1):

$$F_i^{out} = F_i^{in} + FFN\left(BN\left(MCA\left(F_i^{in}\right)\right)\right) \tag{1}$$

Where F_i^{in} represents the input feature of the i layer corresponding to the encoder, F_i^{out} is the output feature after MBCAE module processing, BN corresponds to the batch normalization operation, and MCA corresponds to the multi-scale convolution attention operation. FFN denotes Feed-Forward Neural Network, which is mainly composed of convolution operation and ReLU function to further enhance the nonlinear feature expression capability of the MBCAE module. In the multi-branch structure, the propagation path of the gradient is more complex. In order to prevent the problem of gradient disappearing or gradient explosion, the MBCAE module introduces batch normalized operation before and after the MCA layer.

3.3 Spatial-frequency domain joint feature enhancement module

The detailed information on neural networks from the shallow layer is rich How to effectively combine the shallow detail information with the deep semantic information can improve the accuracy of feature expression. To extract shallow edge features, we employ the Sobel operator. Compared with the Canny or Laplacian operators, the Sobel operator achieves a good balance between noise smoothing and edge preservation while maintaining computational efficiency. In order to effectively inject shallow edge features extracted by Sobel operator into deep semantic information, SFJD-network considers both the spatial domain and frequency domain to enhance the network's ability to perceive important information. The structure of the SFJE (Spatial-Frequency domain joint feature enhancement) module is shown in Fig. 2. The top half of the SFJE module in Fig. 2 is the frequency domain processing process, and the bottom half is the spatial domain processing process. For the frequency domain, we first obtain the amplitude and phase of the shallow feature graph E_1 and deep feature graph F_4 by Fourier transform (FFT). Correspondingly, the amplitude and phase maps of the feature are visualized in the Fig. 4. Then the amplitudes of **the two** are multiplied to enhance the energy of the signal at the same frequency, and the phase correction is carried out by phase addition. In this way, the SFJE module directly promotes the complementarity of semantic information and target boundary. Finally, the enhanced features in



(a) Amplitude (b) Phase

Fig. 4 The Visualization of amplitude and phase

the frequency domain are restored to the spatial domain by inverse Fourier transform (IFFT), and the enhanced features are output by convolution operation. The main formulas involved in the frequency domain are shown in (2) :

$$F_{fre}^{o} = IFFT\left(Amp\left(F_4^{Amp} \otimes E_1^{Amp}\right), Ph\left(F_4^{ph} \oplus E_1^{ph}\right)\right) \quad (2)$$

Where F_{fre}^{o} represents the fusion result of the deep semantic feature F_4 and the shallow edge feature E_1, Amp and Ph represent the amplitude spectrum and phase spectrum of the feature map respectively; $\otimes$ is the element-by-element multiplication operation, $\oplus$ is the element-by-element addition operation.

For the spatial domain, we use an efficient Axis Attention mechanism to aggregate complex spatial detail features from the two main directions of height and width. Axial attention reduces the squared computation complexity of the self-attention mechanism in traditional transformer and effectively expands the network receptive field to capture global information in images. After concatenating the deep feature map with the shallow feature map along the channel dimension, the spatial relationship is built by conducting self-attention mechanism along the height direction of the image, and then the processing is continued along the width direction. In the spatial domain, the axial attention mechanism can be used to connect the context dependence, which can better assist the model in distinguishing the semantics of edge and background. The main formulas involved in the spatial domain are shown in (3) :

$$F_{spa}^{out} = WidAtten\left(HigAtten\left(CBR\left(E_1, F_4\right)\right)\right) \quad (3)$$

Among them, F_{spa}^{out} represents the output features after spatial domain processing, *WidAtten* and *HigAtten* represent axial attention operations along the image width direction and height direction, respectively, and CBR represents 3×3 convolution, batch normalization and ReLU activation operations. Finally, the features enhanced in the spatial domain and frequency domain are concatenated, and the number of channels is adjusted by CBR operation to obtain F_{fuse}.

3.4 Differential learning decoder

There is an unavoidable semantic gap between encoder and decoder, and the decoder's understanding performance of encoder information is closely related to the final prediction result. In the past, the operation of U-type network usually combines the codec information directly by concat operator or addition, ignoring the difference between the two. In this paper, a novel differential learning decoder is proposed to solve this problem, and a $K \times K$ convolution kernel is designed to dynamically map the differential features obtained by subtraction operation. The structure of Differential Learning Decoder (DLD) is shown in Fig. 5.

Firstly, the deep decoder feature map F_D is adjusted the channel number and size to be consistent with the shallow encoder feature map F_l by convolution and upsampling operations. Then the two features are subtracted element by element, and the difference information between the encoder and decoder is initially extracted by 3×3 convolution, and the channel number is adjusted to $K \times K$ size. Softmax kernel normalization is performed on the differential feature

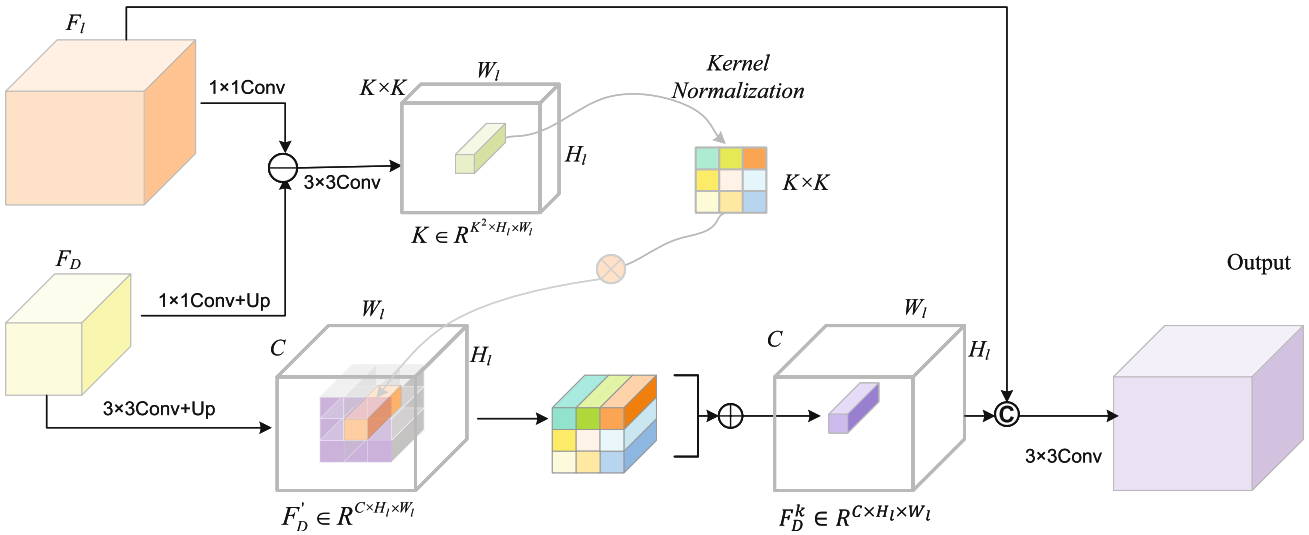


Fig. 5 The architecture of the Differential Learning Decoder module

map along the channel dimension to obtain the weight representation reflecting the difference between the encoder and the decoder. Then the channel dimension of the differential weight map is expanded in spatial dimension to generate a convolution kernel, which reflects the distribution difference in codec information. Finally, the decoder feature map is convolved with the kernel, so that the semantic difference between encoder and decoder can be made up in the decoding process. After the decoder features complete differential learning, then concatenate with the encoder map and output. The main formulas involved are as follows:

$$K = Softmax\left(Conv_{3\times3}\left(Conv_{1\times1}(F_l) - \mathrm{Up}\left(Conv_{1\times1}(F_D)\right)\right)\right) \tag{4}$$

$$F_D^K[i,j] = \sum_{x=-\frac{k}{2}}^{\frac{k}{2}} \sum_{y=-\frac{k}{2}}^{\frac{k}{2}} K[x,y] \cdot F_D^{'}[i+x, j+y] \tag{5}$$

Among them,F_l and F_D represent the encoder feature and the decoder feature respectively. $Conv_{1\times1}$ denotes 1×1 convolution, which is mainly used to adjust the consistency of the channel number of the feature map between different layers, and Up is an up-sampling operation to maintain the consistency of the scale. *Softmax* operation normalized the differential feature map to get the difference distribution convolution kernel $K \in R^{k\times k}$. Carrying out the convolution operation using this kernel with decoder feature $F_D^{'}$ to obtain F_D^K, where $[i, j]$ represents the position of pixels in the corresponding feature map.

3.5 Loss function

In this paper, we employ binary cross-entropy loss and dice loss as the hybrid loss to optimize network training. Binary cross-entropy loss is widely applied in binary classification tasks and dice loss is used for the class imbalanced problems in segmentation tasks. They can be formulated as:

$$\begin{cases} L_{bce} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_l) + (1-y_i)\log(1-\hat{y}_l)\right] \\ L_{dice} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_l}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_l} \end{cases} \tag{6}$$

Where $\hat{y}_l$ represents the predicted value of the network, y_i is the value of corresponding ground truth, and N is the number of pixels. Therefore, the hybrid loss can be defined as:

$$L_{seg} = L_{bce} + L_{dice} \tag{7}$$

To fully optimize the training of parameters across different network depths and enhance overall robustness, the deep supervision strategy is adopted to refine the hybrid loss. The coefficient of the loss is set to 0.1, 0.3, 0.6, and 1 according to the influence effect of each layer on the final output.

4 Experimental results

4.1 Datasets

The performance of the proposed segmentation network SFJD-Net is validated on the ISLES 2022[42] and the ISLES 2018[43] dataset. In order to further verify the network generalization performance, we conducted a generalization experiment on the MSD[44] pancreas cancer dataset.

ISLES 2022 brain stroke dataset The ISLES 2022 dataset consists of 250 MRI sequences from three independent

medical centers. In order to obtain more clear input information, we use ADC mode and DWI mode as input. According to the official dataset partitioning method, we use a five-fold cross-validation approach and averaged the results.

ISLES 2018 brain stroke dataset The ISLES 2018 dataset consists of 94 sets of CT perfusion images from 63 acute stroke patients, and each set of cases includes four modalities: CBF (cerebral blood flow), CBV (cerebral blood volume), MTT (mean time to passage), and Tmax (maximum residual time). To obtain richer input information, we use all four modalities as model inputs. We use a five-fold cross-validation approach and averaged the results.

MSD pancreas cancer dataset The MSD Pancreas Cancer dataset from the Medical Segmentation Decathlon (MSD) challenge comprises 281 abdominal contrast-enhanced CT scans, each annotated with pancreas and pancreatic tumor labels. The scan resolutions range from $512 \times 512 \times 37$ to $512 \times 512 \times 751$, with a slice thickness of 2.5 mm.

To evaluate the performance of the proposed method, we adopt a fourfold cross-validation strategy. Specifically, the dataset is divided into four subsets (i.e., #1–#70, #71–#140, #141–#210, and #211–#281). Each subset is used once as the test set, while the remaining three serve as the training set. The final performance is reported by averaging the results across all four folds.

4.2 Implementation details

The proposed network was implemented using the PyTorch framework and trained on a single NVIDIA GeForce RTX 3090 GPU with 24GB of memory. We employed stochastic gradient descent (SGD) as the optimizer, with an initial learning rate of 1×10^{-4}. The batch size was set to 4, and the network was trained for 250 epochs. During data preprocessing, all input images from the datasets were normalized to the range [0,1]. To enhance model robustness, we applied various data augmentation strategies, including random flipping, rotation (-90° to 90°), and scaling (0.8 to 1.2).

4.3 Evaluation metrics

To evaluate the performance of the proposed SFJD-Net, five metrics are used: Dice Similarity Coefficient (DSC), Average Symmetric Surface Distance (ASD), Hausdorff Distance (HD), Precision, and Recall. The details are as follows:

$$DSC = \frac{2 \times |f_{pre} \cap f_{GT}|}{(|f_{pre}| + |f_{GT}|)} \tag{8}$$

f_{pre} represents the prediction result of the model and f_{GT} denotes the real mask information. DSC measures the overlap between the predicted result and the ground truth. The closer its value is to 1, the more similar it is.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{10}$$

TN, FP, FN, and TP represent true negatives, false positives, false negatives, and true positives, respectively. Sensitivity and specificity are used to measure the true positive rate and false negative rate, respectively.

$$HD = \max\left\{ \max_{x \in f_{GT}} \min_{y \in f_{pre}} d\{x, y\}, \max_{x \in f_{GT}} \min_{y \in f_{pre}} d\{y, x\} \right\} \tag{11}$$

$$ASD = \frac{1}{2}\left\{ \text{mean} \min_{x \in f_{GT}, y \in f_{pre}} d\{x, y\} + \text{mean} \min_{x \in f_{GT}, y \in f_{pre}} d\{y, x\} \right\} \tag{12}$$

x and y represent the voxels of the ground truth and the prediction results, respectively, and $d\{x, y\}$ denotes the Euclidean distance between x and y. HD is used to measure the maximum distance between the boundaries of the predicted segmentation and the ground truth region. ASD is used to measure the average distance between the predicted segmentation boundary and the ground truth boundary.

4.4 Segmentation results on the ISLES 2022 dataset

To fully evaluate the performance of SFJD-Net, it was compared with several mainstream medical image segmentation networks (Attention U-Net, Swin-Unet, TransUNet, TransFuse, U-Mamba) and the brain stroke segmentation network CLCI-Net at the same time. Table 1 shows the comparison results for different models on the ISLES 2022 dataset. As can be seen from Table 1, due to the difficulty in segmenting stroke lesions, the DSC coefficients of the network segmentation are all below 90.00%, and SFJD-Net achieves the optimal value of 87.02%. It shows that SFJD-Net combines the shallow texture details and deep semantics effectively through the SFJE module, thus reducing the probability of misjudgment in the segmentation process. In addition, the two distance-based metrics HD and ASD obtain the minimum values of 1.28 mm and 2.95 mm, respectively, indicating that SFJD-Net achieves the highest integrity and accuracy in the segmentation of the brain stroke region edge. Moreover, SFJE-Net achieves the minimum variance in all indicators, which strongly proves that SFJE-Net has strong robustness.

As shown in Fig. 6, the violin diagram of five-fold cross-verification on the ISLES 2022 data set shows that the

Table 1 The results (measured by the DSC, Sensitivity, Specificity, ASD and HD) of Brain Stroke segmentation on the ISLES 2022 dataset. "-" denotes that the corresponding results are not provided in the literature. ↑ means the higher the better and ↓ represents the opposite. Optimal results (described by mean ± std) are shown in bold

Method	DSC(%)↑	Sensitivity(%)↑	Specificity(%)↑	ASD(mm)↓	HD(mm)↓
Attention U-Net[29]	84.91±1.60	85.94±2.13	95.02±2.26	1.73±0.21	3.54±0.16
Swin-Unet[33]	84.62±3.87	85.21±2.59	93.47±3.02	1.75±0.14	3.59±0.20
CLCI-Net[14]	85.04±2.67	85.42±1.94	95.25±1.84	1.46±0.08	3.22±0.17
TransUNet[34]	85.78±1.94	86.41±1.98	95.24±1.86	1.35±0.07	3.15±0.15
TransFuse[35]	85.40±1.79	85.78±2.04	95.11±2.13	1.40±0.07	3.17±0.15
nnUNet[45]	86.19±1.45	85.89±2.10	95.43±1.54	1.26±0.12	3.09±0.11
U-Mamba[38]	85.92±1.58	86.37±1.87	95.38±1.75	1.33±0.06	3.04±0.09
SFJD-Net (Ours)	**87.02±1.52**	**87.93±1.46**	**95.62±1.37**	**1.28±0.06**	**2.95±0.09**

distribution of DSC accuracy still maintains a certain consistency among different folds, further demonstrating the robustness of the model. At the same time, we can observe from Fig. 6 that the DSC results of each fold are above 85.00%, further demonstrating the effectiveness of SFJE-Net for stroke segmentation. In Fig. 7, the segmentation sample results were selected for display. The dataset entered into the network covers both ADC and DWI modes, and the shaded area in the input represents the lesion region. From Fig. 7, although there are large differences in the morphology and location of focal areas between different samples, SFJD-Net can still achieve high segmentation accuracy, indicating that SFJD-Net can be successfully migrated to other focal segmentation tasks.

As shown in Fig. 8, we randomly selected a case of brain stroke from the public ISLES 2022 dataset to carry out a visual comparison. The first image denotes an image with label. It can be intuitively seen that compared with other excellent models, the output of the model proposed in this paper is closer to the shape of the label, which is consistent with the precision performance of DSC in Table 1. From Fig. 8, it can also be intuitively seen that SFJE-Net reduces the false positive area of stroke as much as possible compared with other excellent models. In the process of decoding, the differential learning decoder is used to fully bridge the semantic gap between the encoder and decoder, so that SFJD-Net has a relatively excellent performance.

4.5 Segmentation results on the ISLES 2018 dataset

To further evaluate the robustness and consistency of SFJD-Net on ischemic stroke segmentation, additional experiments were conducted on the ISLES 2018 dataset. We compared SFJD-Net with several mainstream segmentation networks (U-Net++, Swin-Unet, Attention U-Net), as well as stroke-specific segmentation models MDANet, PoolNet and PerfU-net. Table 2 shows the comparison results of different models on the ISLES 2018 dataset. As seen in Table 2, due to the small amount of data in the stroke 2018 dataset and the difficulty of lesion segmentation, the overall DSC values are below 70.00%. Nevertheless, SFJD-Net attains the highest DSC of 69.66%. Furthermore, it achieves the lowest HD and ASD with values of 2.86 mm and 2.70 mm, respectively, demonstrating superior performance in boundary segmentation. This indicates that the proposed SFJD module fuses shallow and deep features in both the

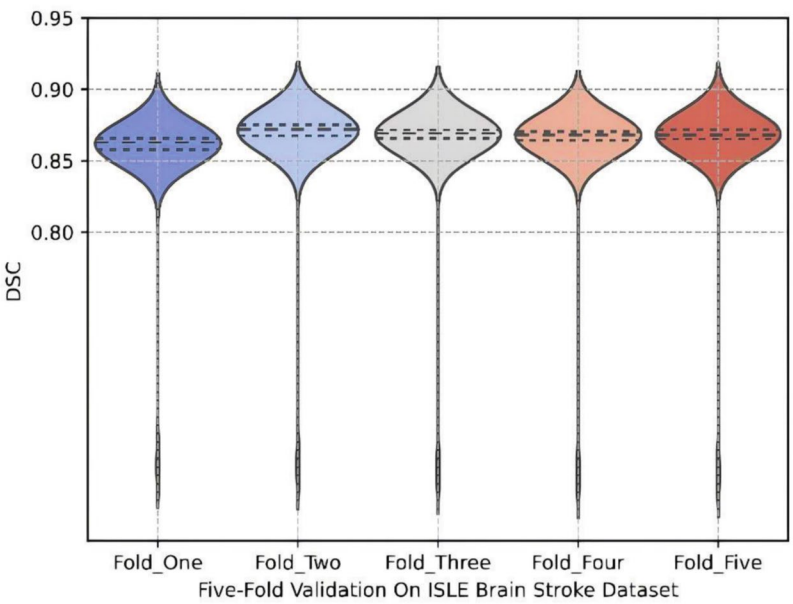


Fig. 6 Violin Plot of SFJD-Net on the Five-fold Cross-validation of the ISLES 2022 dataset

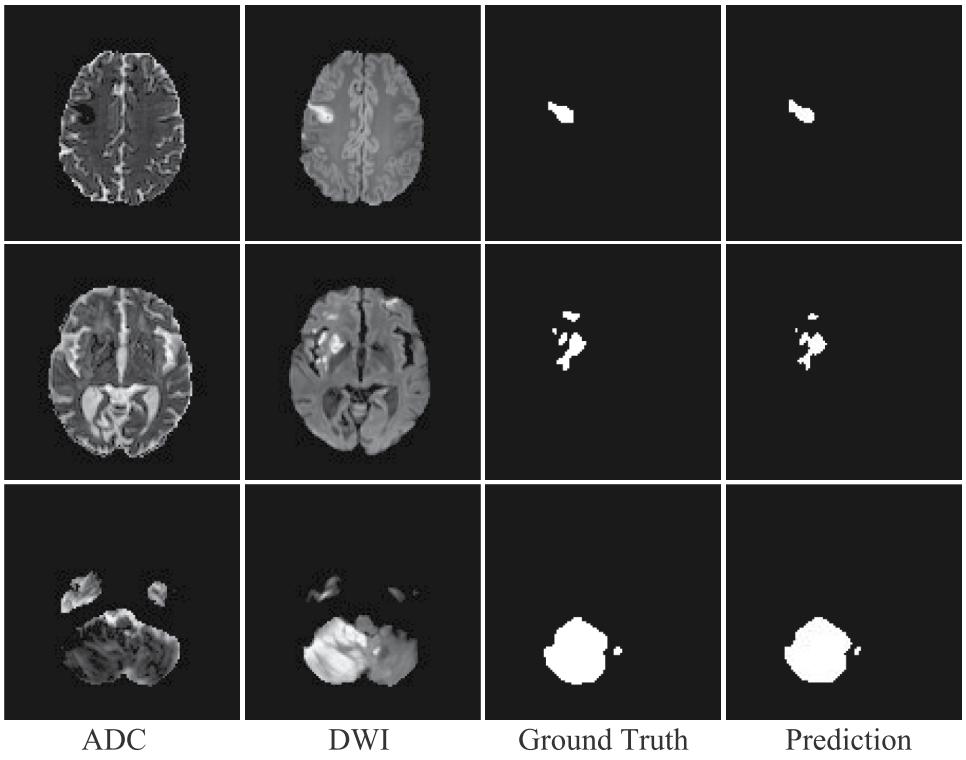


Fig. 7 The segmentation results of SFJD-Net for different cases on the ISLES 2022 dataset

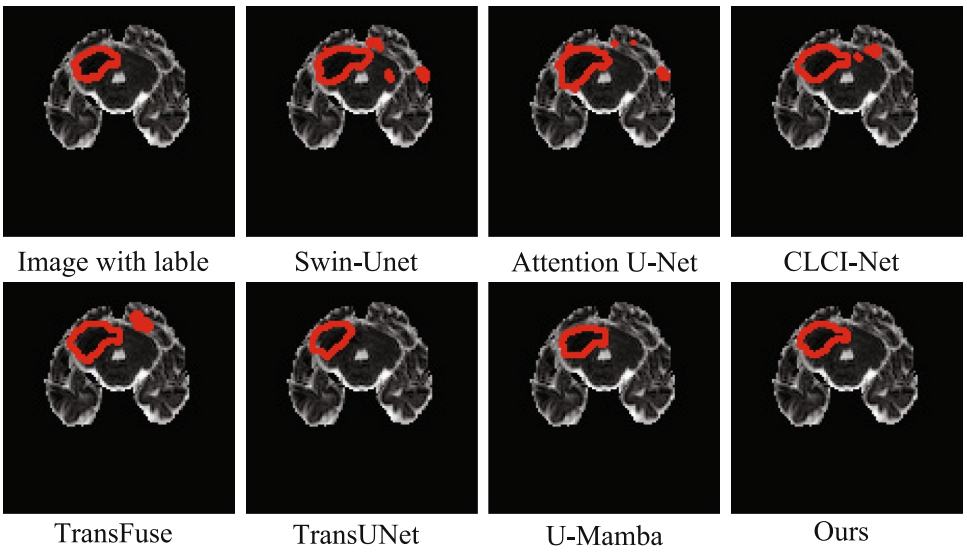


Fig. 8 Visual comparison of SFJD-Net and Previous Excellent Segmentation Models on the ISLES 2022 dataset

Table 2 The results (measured by the DSC, Sensitivity, Specificity, ASD and HD) of Brain Stroke segmentation on the ISLES 2018 dataset. "-" denotes that the corresponding results are not provided in the literature. ↑ means the higher the better and ↓ represents the opposite. Optimal results (described by mean ± std) are shown in bold

Method	DSC(%)↑	Sensitivity(%)↑	Specificity(%)↑	ASD(mm)↓	HD(mm)↓
Attention U-Net[29]	65.28±2.45	52.52±5.71	99.19±0.08	3.28±0.24	2.93±0.05
Swin-Unet[33]	60.10±2.94	43.59±3.20	99.10±0.11	3.96±0.92	3.15±0.07
Unet++[31]	58.86±3.37	48.83±6.70	99.05±0.14	4.58±0.86	3.24±0.08
CLCI-Net[14]	67.98±2.70	53.00±7.47	99.02±0.16	3.00±0.43	3.02±0.06
TransUNet[34]	68.29±1.36	49.82±5.15	**99.41±0.18**	2.72±0.34	2.93±0.05
TransFuse[35]	64.95±3.86	45.34±5.05	98.09±0.16	5.59±0.40	4.23±0.13
SFJD-Net (Ours)	**69.66±3.17**	**57.52±4.33**	99.12±0.17	**2.70±0.34**	**2.86±0.05**

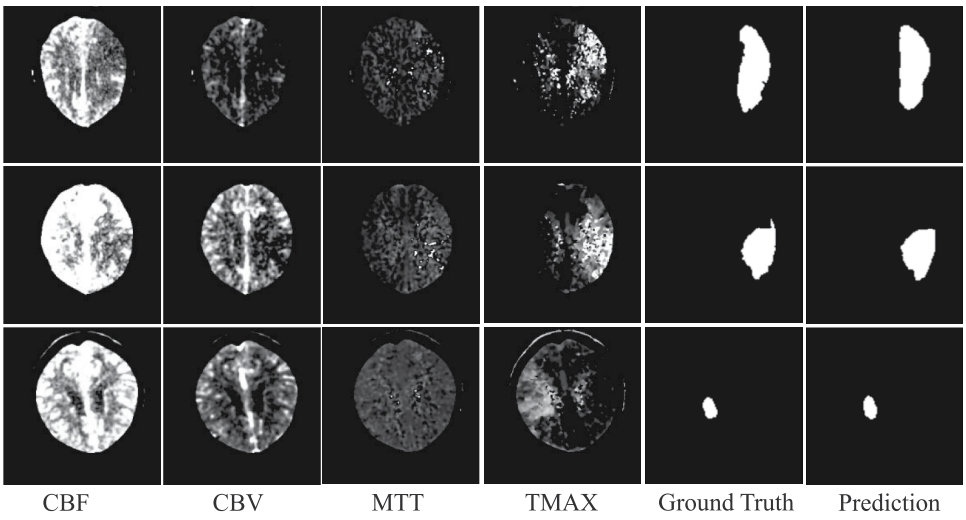


Fig. 9 The segmentation results of SFJD-Net on representative cases on the ISLES 2018 dataset

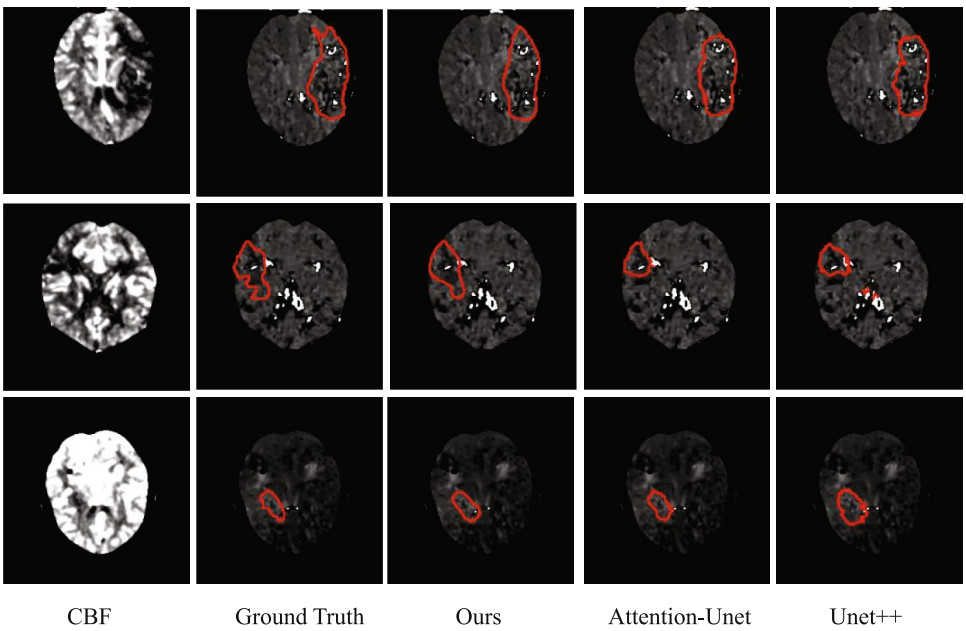


Fig. 10 Visual comparison of SFJD-Net with existing segmentation models on the ISLES 2018 dataset

spatial and frequency domains. It allows the deep features to be supplemented with the rich edge information from the shallow layers, which helps the model learn clearer lesion boundaries.

As shown in Fig. 9, some representative segmentation samples were selected for demonstration. The input data are the four modalities in the ISLES 2018 dataset: CBF, CBV, MTT, and TMAX. As can be seen from Fig. 10, SFJD-Net demonstrates high segmentation accuracy, both in the segmentation task of larger lesions and in the recognition of tiny lesions. The results validate the effectiveness and robustness of SFJD-Net in complex stroke image segmentation tasks.

Figure 10 shows the visualized comparison results of SFJD-Net with other models on the ISLES 2018 dataset, which contains three representative cases of ischemic stroke. As can be seen from Fig. 10, the segmentation results of SFJD-Net are closer to the real labels in terms of shape, whether for larger lesions, tiny lesions, or lesions with more complex edge morphology. This demonstrates that the MBCAE module proposed in this paper can introduce rich scale information through multiple dilated convolutions, which captures features of lesions of varying sizes during the encoder stage. Compared to other models, SFJD-Net exhibits superior segmentation performance in critical regions, fully validating the effectiveness of the proposed method.

4.6 Segmentation results on the MSD pancreas cancer dataset

In order to further verify the generalization performance of SFJD-Net, generalization experiments were carried out on the published pancreas cancer dataset. Similarly, SFJE-Net was compared with several mainstream segmentation networks as well as the pancreas cancer segmentation network Multi-scale U-Net. Table 3 shows the comparison results for different models on the MSD pancreas cancer dataset. The dataset was divided according to the published partitioning method, and finally the excellent DSC value of 77.26% was obtained by SFJE-Net. The minimum distance index value proved that SFJE-Net could capture the edge of pancreas lesions more accurately. As can be seen from Table 3, the pancreas cancer segmentation task is more challenging than the stroke segmentation task, but SFJE-Net still maintains the optimal index, which fully proves the effectiveness and stability of SFJE-Net.

Figure 11 shows the comparison of DSC coefficients between SFJD-Net and previous excellent segmentation models on the experimental results of the MSD pancreas cancer dataset. This figure also integrates box diagram, violin diagram and scatter diagram, among which the box median line of SFJD-Net is ahead of other models, and the density of data points can be intuitively seen. This fully demonstrates that the distribution of SFJD-Net experimental data is better than that of other networks, and the value of DSC accuracy is mostly stable above 75.00%.

Figure 12 shows the comparison between SFJD-Net and other models on the MSD pancreatic cancer dataset. In order to more conveniently observe the differentiation of the segmentation results, we show the local view of the pancreatic lesion in Fig. 12 (a) and the global CT view in Fig. 12 (b). The first image of each group (a) and (b) represents the real label, and the blue part is the pancreatic cancer area. It can be clearly seen that the pancreatic lesion

Table 3 The results (measured by the DSC, Sensitivity, Specificity, ASD and HD) of pancreas cancer segmentation on the MSD dataset. "-" denotes that the corresponding results are not provided in the literature. ↑ means the higher the better and ↓ represents the opposite. Optimal results (described by mean ± std) are shown in bold

Method	DSC(%)↑	Sensitivity(%)↑	Specificity(%)↑	ASD(mm)↓	HD(mm)↓
Attention U-Net[29]	65.77±3.04	80.19±3.86	84.02±4.06	2.06±0.14	3.21±0.12
Swin-Unet[33]	63.52±4.53	78.24±4.91	82.36±4.85	2.19±0.19	3.35±0.17
Multi-scale U-Net[13]	65.85±3.06	80.58±4.09	83.55±3.53	2.03±0.15	3.12±0.13
TransUNet[34]	73.26±3.02	84.81±4.15	87.82±3.46	1.38±0.14	2.54±0.12
TransFuse[35]	70.18±3.18	82.36±4.19	85.94±4.09	1.47±0.15	2.67±0.14
U-Mamba[38]	74.58±2.69	85.17±3.94	87.61±3.22	1.28±0.13	2.40±0.11
SFJD-Net (Ours)	**77.26±2.05**	**88.93±3.24**	**89.72±2.36**	**1.05±0.09**	**2.14±0.08**

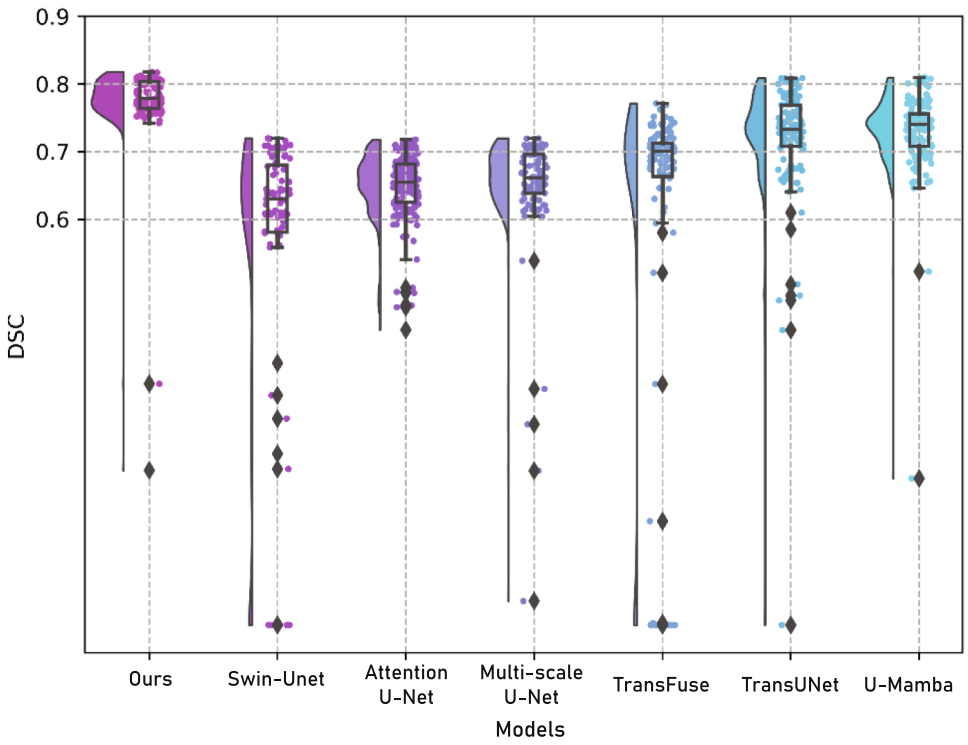


Fig. 11 Raincloud plot Comparing DSC of SFJD-Net with Previous Excellent Segmentation Models on the MSD Cancer Segmentation Dataset

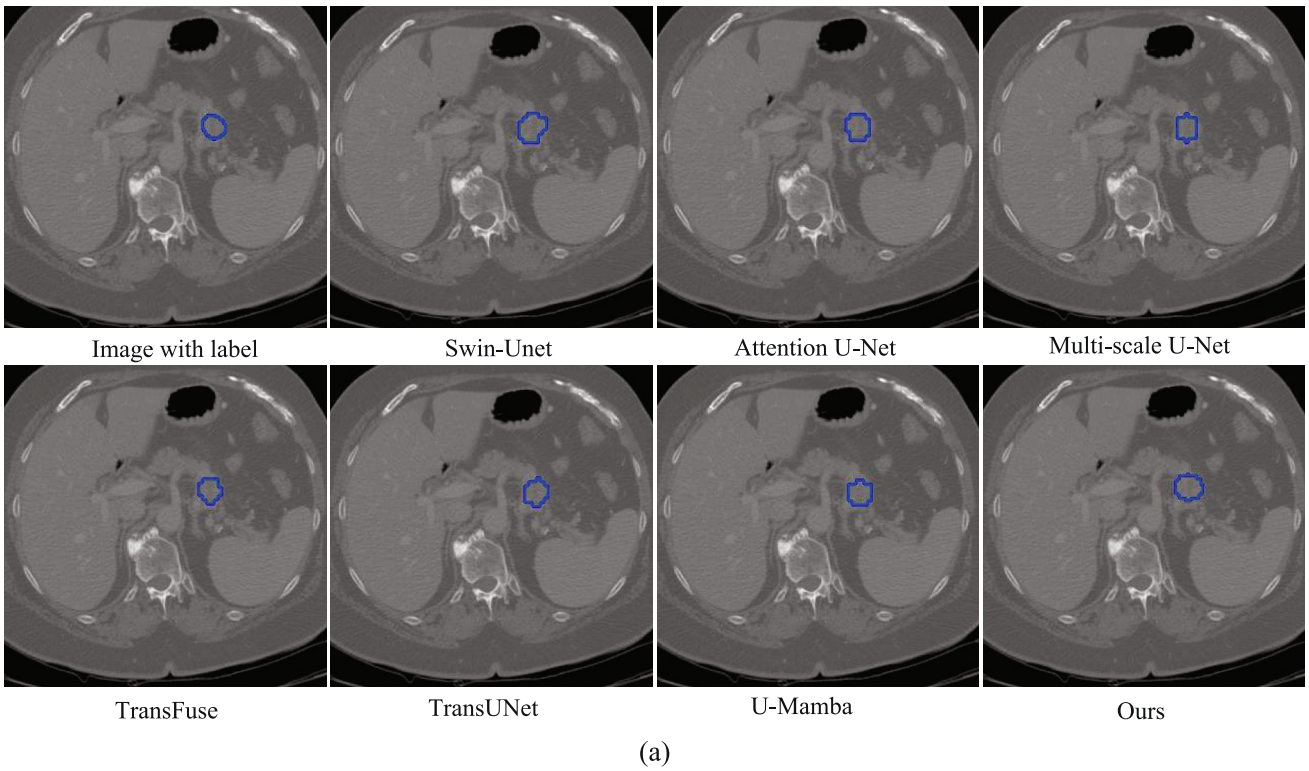


Fig. 12 Visual Comparison of SFJD-Net and Previous Excellent Segmentation Models on the MSD Cancer Segmentation Dataset. (**a**) Local view of the pancreas lesion region (**b**) Global view of the whole CT image

region of most patient samples is small and changeable. Compared with the dedicated pancreatic lesion segmentation network and several other mainstream segmentation networks, SFJD-Net can more completely segment the scope of the pancreatic lesion region. The contour of the lesion predicted by SFJD-Net is smoother and closer to the real label. This proves that SFJD-Net effectively uses the multi-scale convolution attention mechanism to extract more accurate pancreatic cancer features in the coding stage, and makes up the semantic gap caused by simple skip connections as much as possible through differential learning strategies in the decoding process, reaching a higher segmentation level.

4.7 Abaltion study

To obtain an in-depth analysis of the three main innovative modules in SFJD-Net (MBCAE; SFJE; DLD), we performed an ablation experiment on the ISLES 2022 dataset. The experimental results are shown in Table 4, where Baseline represents the initial U-Net network, Baseline+MBCAE represents the use of multi-scale convolution attention encoder on the basis of U-Net, Baseline+MBCAE+SFJE represents the continuous addition of the SFJE module. SFJD-Net represents the integration of the three modules.

Performance analysis of MBCAE module As shown in Table 4, after adding MBCAE, the DSC coefficient increases by about 0.58%, indicating that for the 3×3 convolutions in the original U-Net, the MBCAE module effectively helps the network to obtain higher quality feature expression through the scale variation rule in medical images.

Performance analysis of SFJE module As shown in Table 4, adding the SFJE module significantly improves the DSC. Compared with the baseline, the DSC increases by 1.38%, indicating that SFJE effectively integrates shallow features into deep representations. Furthermore,when Compared with the baseline+MBCAE model, incorporating SFJE further increases the DSC by approximately 1.01%, which indicates that SFJE module effectively utilizes the high-quality features extracted from MBCAE multi-scale convolution attention encoder. The feature enhancement is successfully realized in the spatial domain and frequency domain. As shown in Fig. 13, the leftmost column represents the real label of the brain stroke lesion area. The evolution trend of the stroke lesion boundary

Table 4 The ablation results of Brain Stroke segmentation on the ISLES 2022 dataset."-" denotes that the corresponding results are not provided in the literature. ↑ means the higher the better and ↓ represents thea opposite. Optimal results (described by mean ± std) are shown in bold

Method	DSC(%)↑	Sensitivity(%)↑	Specificity(%)↑	ASD(mm)↓	HD(mm)↓
Baseline (U-Net)	84.69±2.04	83.65±2.16	93.28±2.10	1.74±0.07	3.56±0.11
Baseline+SFJE	86.07±1.89	85.91±1.36	95.44±1.84	1.50±0.11	3.23±0.06
Baseline+MBCAE	85.27±1.76	85.74±1.69	95.50±1.72	1.44±0.07	3.19±0.10
Baseline+MBCAE+SFJE	86.28±1.55	86.20±1.67	95.55±1.50	1.30±0.06	3.02±0.09
SFJD-Net (Ours)	**87.02±1.52**	**87.93±1.46**	**95.62±1.37**	**1.28±0.06**	**2.95±0.09**

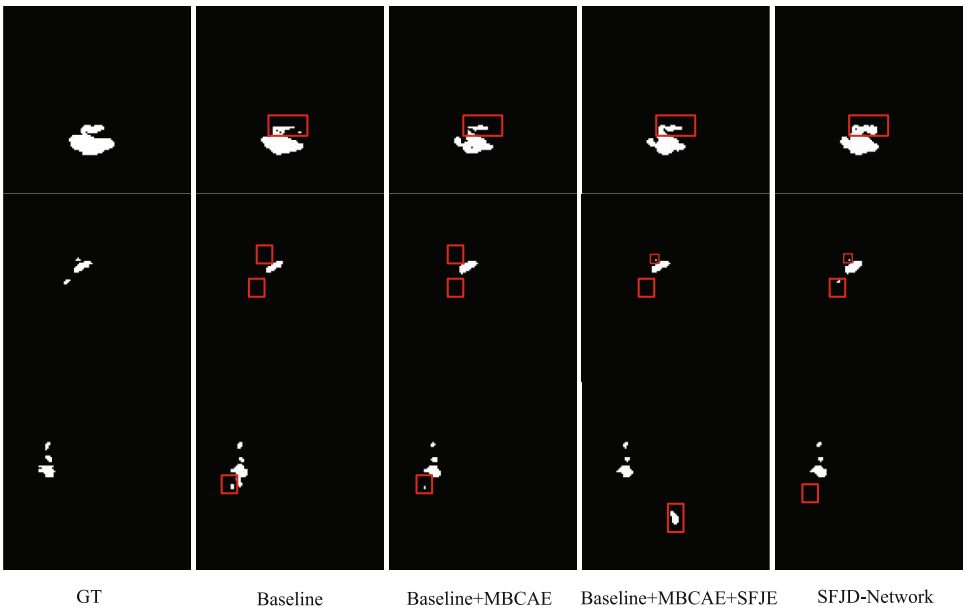


Fig. 13 Visual Comparison of ablation experiments for SFJD-Net on the ISLES 2022 dataset. The leftmost column denotes the ground truth

Table 5 Comparison study on computational efficiency and accuracy. Optimal results are shown in bold

Method	FLOPs(G)↓	Params(M)↓	Inference Time(ms)↓	Memory(M)↓	DSC(%)↑
Attention UNet[29]	14.21	**8.44**	**12.43**	3274	65.28±2.45
CLCI[14]	17.41	19.67	13.03	6584	67.98±2.70
UNet++[31]	33.87	9.05	14.60	**1906**	58.86±3.37
TransUNet[34]	32.67	106.95	18.90	4672	68.29±1.36
SwinUNet[33]	**7.73**	27.15	15.45	2806	60.10±2.94
SFJD-Net (Ours)	12.88	14.02	15.73	3858	**69.66±3.17**

predicted by Baseline+MBCAE+SFJE is closer to the real label than the Baseline.

Performance analysis of DLD module SFJD-Net is obtained after adding differential learning decoder, and the DSC increases by 0.74% compared with the former. The DLD module dynamically complements the differences between each codec layer by a constructed convolutional kernel and avoids noise interference while aligning the codec semantics, thus guiding the network to focus on the target region.

As shown in Fig. 13, SFJD-Net achieves the optimal indicator in the brain stroke segmentation task. For stroke lesion targets with a small proportion of the region, SFJD-Net reduces the false negative and false positive regions as much as possible for other variants.

4.8 Model complexity

We compared SFJD-Net with several existing methods in terms of computational efficiency and segmentation accuracy, with the results summarized in Table 5. Attention UNet and UNet++ have relatively small parameter counts and memory usage, but their segmentation accuracy is limited. CLCI, specifically designed for stroke segmentation, has a parameter size comparable to SFJD-Net and slightly faster inference speed. Its segmentation performance is still inferior to SFJD-Net. Moreover, SwinUNet shows lower segmentation accuracy, while TransUNet achieves acceptable accuracy but with an excessively large parameter count, making it less practical for deployment. In contrast, SFJD-Net achieves the highest segmentation Dice (69.66%) while maintaining a modest parameter size (14.02M) and reasonable inference time (15.73 ms). Overall, SFJD-Net strikes a favorable balance between segmentation accuracy, computational efficiency, and model complexity, making it suitable for real-world applications where both precision and efficiency are essential.

5 Discussions

Although our proposed SFJD-Net shows excellent performance in the stroke segmentation task, and also achieves good results in the pancreas segmentation task, reflecting the generalization ability of the model, there is still some room for improvement. Currently, our method uses fully supervised learning, which requires pixel-level labeling. However, in the field of medical image segmentation, pixel-level labeling of the target is very time- and energy-consuming, and it is also difficult to obtain high-quality segmentation maps. Therefore, in our future work, we will consider using semi-supervised learning for optimization to reduce the dependence on a large amount of labeled data.

In addition, although we have fused multimodal image information in the design to effectively improve the model's understanding of the lesion region, the current modal fusion approach is still relatively simple, and the potential complementary relationship between modalities has not yet been fully explored. In our future work, we will explore the effective use of different modal features to enhance the model's ability to perceive complex lesions and segmentation accuracy.

6 Conclusion

In this paper, a novel SFJD-Net network for brain stroke segmentation is proposed. In this network, an additional branch is added to the U-shaped network to fully extract the shallow texture details by using the edge operator, and a feature enhancement module that combines the space and frequency domain is designed to skillfully integrate the details into the deep semantics. In the encoding stage, the multi-scale convolution attention mechanism is used to pay attention to the scale change law of lesions in medical images, so that the network can obtain more reliable feature expression. In addition, a decoder based on differential learning strategy

is designed, and the corresponding convolution kernel is obtained by reconstructing the difference between encoder and decoder, which effectively alleviates the ambiguity between codecs and eliminates noise interference, so that the decoding process can pay more attention to the real shape of the segmentation target. In order to verify the segmentation performance of SFJD-Net, a series of comparative experiments were carried out on the ISLES 2022 and the ISLES 2018 dataset. The segmentation accuracy was significantly better than the previous stroke segmentation model and mainstream deep neural networks. The other experiment was conducted on the MSD pancreas cancer dataset to explore the generalization of SFJD-Net. The experimental results show that SFJD-Net has certain robustness and generalization. In addition, the ablation experiments for the three modules (MBCAE; SFJE; DLD) are conducted to analyze their contribution to the overall performance. Finally, the shortcomings of SFJD-Net and the direction of future efforts are discussed in order to achieve a better segmentation level.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant 62476088, the Exploratory Device R&D Projects of National Clinical Research Center for Interventional Medicine (NO. 2021-002), and the Key Clinical Research Projects of National Clinical Research Center for Interventional Medicine (NO. 2019-003).

Author Contributions **Xin Huang:** Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review and editing. **Yu Zhu:** Methodology, Resources, Supervision, Project administration, Funding acquisition, Writing - review and editing. **YaTong Liu:** Conceptualization, Software, Writing - review & editing. **Yuhao Zhang:** Resources, Data curation, Funding acquisition.

Data Availability The original data used is a public dataset.

Declarations

Declaration of Competing Interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Approval This work does not involve experimental procedures with human subjects or animals.

References

1. Raza R, Bajwa UI, Mehmood Y, Anwar MW, Jamal MH (2023) dresu-net: 3d deep residual u-net based brain tumor segmentation from multimodal mri. Biomed Signal Process Control 79:103861. https://doi.org/10.1016/j.bspc.2022.103861
2. Deepa S, Janet J, Sumathi S, Ananth J (2023) Hybrid optimization algorithm enabled deep learning approach brain tumor segmentation and classification using mri. J Digit Imaging 36(3):847–868. https://doi.org/10.1007/s10278-022-00752-2
3. Matos B, Publicover SJ, Castro LFC, Esteves PJ, Fardilha M (2023) Brain and testis: more alike than previously thought? Adv Surgical Med Specialt 425–445. https://doi.org/10.1098/rsob.200322
4. Wei Y-C, Huang W-Y, Jian C-Y, Hsu C-CH, Hsu C-C, Lin C-P, Cheng C-T, Chen Y-L, Wei H-Y, Chen K-F (2022) Semantic segmentation guided detector for segmentation, classification, and lesion mapping of acute ischemic stroke in mri images. NeuroImage: Clinical 35:103044. https://doi.org/10.1016/j.nicl.2022.103044
5. Rebouças EdS, Marques RC, Braga AM, Oliveira SA, De Albuquerque VHC, Rebouças Filho PP (2019) New level set approach based on parzen estimation for stroke segmentation in skull ct images. Soft Computing 23(19):9265–9286. https://doi.org/10.1007/s00500-018-3491-4
6. Kousar T, Rahim MSM, Iqbal S, Yousaf F, Sanaullah M (2025) Applications of deep learning algorithms in ischemic stroke detection, segmentation, and classification. Artif Intell Rev 58(5):1–48. https://doi.org/10.1007/s10462-025-11119-8
7. Chavva IR, Crawford AL, Mazurek MH, Yuen MM, Prabhat AM, Payabvash S, Sze G, Falcone GJ, Matouk CC, Havenon A et al (2022) Deep learning applications for acute stroke management. Ann Neurol 92(4):574–587. https://doi.org/10.1002/ana.26435
8. Xie B, Cao J, Xie J, Khan FS, Pang Y (2024) Sed: A simple encoder-decoder for open-vocabulary semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3426–3436. https://doi.org/10.1109/CVPR52733.2024.00329
9. Li Y, Wang Z, Yin L, Zhu Z, Qi G, Liu Y (2023) X-net: a dual encoding-decoding method in medical image segmentation. Vis Comput 39(6):2223–2233. https://doi.org/10.1007/s00371-021-02328-7
10. Hoorali F, Khosravi H, Moradi B (2022) Irunet for medical image segmentation. Expert Syst Appl 191:116399. https://doi.org/10.1016/j.eswa.2021.116399
11. Shen N, Wang Z, Li J, Gao H, Lu W, Hu P, Feng L (2023) Multi-organ segmentation network for abdominal ct images based on spatial attention and deformable convolution. Expert Syst Appl 211:118625. https://doi.org/10.1016/j.eswa.2022.118625
12. Rahman MM, Marculescu R (2024) G-cascade: Efficient cascaded graph convolutional decoding for 2d medical image segmentation. In: 2024 IEEE/CVF Winter conference on applications of computer vision (WACV), pp 7713–7722. https://doi.org/10.1109/WACV57701.2024.00755
13. Qiu D, Ju J, Ren S, Zhang T, Tu H, Tan X, Xie F (2024) A deep learning-based cascade algorithm for pancreatic tumor segmentation. Front Oncol 14:1328146. https://doi.org/10.3389/fonc.2024.1328146
14. Yang H, Huang W, Qi K, Li C, Liu X, Wang M, Zheng H, Wang S (2019) Clci-net: Cross-level fusion and context inference networks for lesion segmentation of chronic stroke. In: International conference on medical image computing and computer-assisted intervention, pp 266–274. Springer https://doi.org/10.1007/978-3-030-32248-9_30
15. Qiong L, Chaofan L, Jinnan T, Liping C, Jianxiang S (2025) Medical image segmentation based on frequency domain decomposition svd linear attention. Sci Rep 15(1):2833. https://doi.org/10.1038/s41598-025-86315-1
16. Nam J-H, Syazwany NS, Kim SJ, Lee S-C (2024) Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention. In: 2024 IEEE/CVF Conference on computer vision and pattern recognition (CVPR), pp 11480–11491. https://doi.org/10.1109/CVPR52733.2024.01091
17. Chen Y, Zhang X, Peng L, He Y, Sun F, Sun H (2024) Medical image segmentation network based on multi-scale frequency domain filter. Neural Netw 175:106280. https://doi.org/10.1016/j.neunet.2024.106280
18. Luo J, Dai P, He Z, Huang Z, Liao S, Liu K (2024) Deep learning models for ischemic stroke lesion segmentation in medical

images: A survey. Comput Biol Med 175:108509. https://doi.org/10.1016/j.compbiomed.2024.108509
19. Zhou Y, Huang W, Dong P, Xia Y, Wang S (2019) D-unet: a dimension-fusion u shape network for chronic stroke lesion segmentation. IEEE/ACM Trans Comput Biol Bioinf 18(3):940–950. https://doi.org/10.1109/TCBB.2019.2939522
20. Liu R, Pu W, Zou Y, Jiang L, Ye, Z (2022) Pool-unet: Ischemic stroke segmentation from ct perfusion scans using poolformer unet. In: 2022 6th Asian conference on artificial intelligence technology (ACAIT), pp 1–6. IEEE. https://doi.org/10.1109/ACAIT56212.2022.10137834
21. Liu L, Chang J, Liu Z, Zhang P, Xu X, Shang H (2023) Hybrid contextual semantic network for accurate segmentation and detection of small-size stroke lesions from mri. IEEE J Biomed Health Inform 27(8):4062–4073. https://doi.org/10.1109/JBHI.2023.3273771
22. Kuang H, Wang Y, Liu J, Wang J, Cao Q, Hu B, Qiu W, Wang J (2024) Hybrid cnn-transformer network with circular feature interaction for acute ischemic stroke lesion segmentation on non-contrast ct scans. IEEE Trans Med Imaging 43(6):2303–2316. https://doi.org/10.1109/TMI.2024.3362879
23. Shi T, Jiang H, Zheng B (2021) C 2 ma-net: Cross-modal cross-attention network for acute ischemic stroke lesion segmentation based on ct perfusion scans. IEEE Trans Biomed Eng 69(1):108–118. https://doi.org/10.1109/TBME.2021.3087612
24. Liu L, Wu F-X, Wang J (2019) Efficient multi-kernel dcnn with pixel dropout for stroke mri segmentation. Neurocomputing 350:117–127. https://doi.org/10.1016/j.neucom.2019.03.049
25. Zhang K, Zhu Y, Li H, Zeng Z, Liu Y, Zhang Y (2024) Mdanet: Multimodal difference aware network for brain stroke segmentation. Biomed Signal Process Control 95:106383. https://doi.org/10.1016/j.bspc.2024.106383
26. Vries L, Emmer BJ, Majoie CB, Marquering HA, Gavves E (2023) Perfu-net: Baseline infarct estimation from ct perfusion source data for acute ischemic stroke. Med Image Anal 85:102749. https://doi.org/10.1016/j.media.2023.102749
27. Zhao X, Wang L, Zhang Y, Han X, Deveci M, Parmar M (2024) A review of convolutional neural networks in computer vision. Artif Intell Rev 57(4):99. https://doi.org/10.1007/s10462-024-10721-6
28. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, pp 234–241. Springer. https://doi.org/10.1007/978-3-319-24574-4_28
29. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B et al (2018) Attention u-net: Learning where to look for the pancreas. arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999
30. Diakogiannis FI, Waldner F, Caccetta P, Wu C (2020) Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J Photogramm Remote Sens 162:94–114. https://doi.org/10.1016/j.isprsjprs.2020.01.013
31. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J (2018) Unet++: A nested u-net architecture for medical image segmentation. In: International workshop on deep learning in medical image analysis, pp 3–11. Springer. https://doi.org/10.1007/978-3-030-00889-5_1
32. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst 30. arxiv:1706.03762
33. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, Wang M (2022) Swin-unet: Unet-like pure transformer for medical image segmentation. In: European conference on computer vision, pp 205–218. Springer. https://doi.org/10.1007/978-3-031-25066-8_9
34. Chen J, Lu Y, Yu Q, Luo X, Adeli E, Wang Y, Lu L, Yuille AL, Zhou Y (2021) Transunet: Transformers make strong encoders for medical image segmentation. arXiv:2102.04306. https://doi.org/10.48550/arXiv.2102.04306
35. Zhang Y, Liu H, Hu Q (2021) Transfuse: Fusing transformers and cnns for medical image segmentation. In: International conference on medical image computing and computer-assisted intervention, pp 14–24. Springer. https://doi.org/10.1007/978-3-030-87193-2_2
36. Gu A, Dao T (2023) Mamba: Linear-time sequence modeling with selective state spaces. arXiv:2312.00752. https://doi.org/10.48550/arXiv.2312.00752
37. Basar T (2001) A new approach to linear filtering and prediction problems. https://doi.org/10.1109/9780470544334.ch9
38. Ma J, Li F, Wang B (2024) U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv:2401.04722. https://doi.org/10.48550/arXiv.2401.04722
39. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y et al (2023) Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 4015–4026. https://doi.org/10.1109/ICCV51070.2023.00371
40. Essahraui S, Lamaakal I, Makkaoui KE, Ouahbi I, Bouami MF, Maleh Y (2025) Kolmogorov—arnold networks: Overview of architectures and use cases. In: 2025 International conference on circuit, systems and communication (ICCSC), pp 1–6. https://doi.org/10.1109/ICCSC66714.2025.11135248
41. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021) Learning transferable visual models from natural language supervision. arxiv:2103.00020
42. Hernandez Petzsche MR, De La Rosa E, Hanning U, Wiest R, Valenzuela W, Reyes M, Meyer M, Liew S-L, Kofler F, Ezhov I et al (2022) Isles 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. Scientific data 9(1):762. https://doi.org/10.1038/s41597-022-01875-5
43. Maier O, Menze BH, Gablentz J, Häni L, Heinrich MP, Liebrand M, Winzeck S, Basit A, Bentley P, Chen L et al (2017) Isles 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral mri. Med Image Anal 35:250–269. https://doi.org/10.1016/j.media.2016.07.009
44. Simpson AL, Antonelli M, Bakas S, Bilello M, Farahani K, Van Ginneken B, Kopp-Schneider A, Landman BA, Litjens G, Menze B et al (2019) A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv:1902.09063. https://doi.org/10.48550/arXiv.1902.09063
45. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH (2021) nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18(2):203–211. https://doi.org/10.1038/s41592-020-01008-z