

LiU-Net: Ischemic Stroke Lesion Segmentation Based on Improved KiU-Net

Yingwei Li, Xiaoxia Zhang, Luzhou Liu

Abstract—Earlier and more accurate diagnosis of ischemic stroke is crucial in enhancing the therapeutic outcome for patients. CT technology currently stands as the most rapid diagnostic modality in clinical medicine. Due to the diverse and complex shape of ischemic stroke lesions, accurate segmentation remains a challenging task for automated diagnosis systems. This paper, proposes an ischemic stroke lesion segmentation network, LiU-Net. It based on KiU-Net, which improves network performance and is more suitable for practical lesion segmentation applications. Firstly, KiU-Net combines the undercomplete network U-Net and the overcomplete network Kite-Net. It can simultaneously learn both image detail features and global structural features. Secondly, LiU-Net combines the axial self-attention module with KiU-Net. The introduction of attention can make the network achieve both segmentation accuracy and efficiency. In addition, to improve the flexibility of axial self-attention, a gate factor is introduced within the module to encode information about spatial structure of image. Finally, to address the issue of gradient vanishing, we incorporated residual connection into the network to bolster the feature maps at each depth level and facilitate effective cross depth feature integration. Since there are few publicly available datasets of CT images of ischemic stroke in medical images. We applied to Longcheng District People's Hospital, and processed the obtained images to form a dataset of ischemic stroke. The experimental results shown, LiU-Net is more accurate in segmenting different shapes of ischemic stroke lesions. Compared with KiU-Net, LiU-Net improves the Dice, Acc, and mIoU metrics by 2.44%, 3.4%, and 3.89% respectively. Therefore, LiU-Net is highly suitable for ischemic stroke lesion segmentation, and effectively assist computers in this task.

Index Terms—Ischemic Stroke, Medical Image Segmentation, KiU-Net, Residual Connection, Axial Self-Attention Machine.

I. INTRODUCTION

Ischemic stroke is a cerebrovascular disease [1], and it is also one of the most common death and disability diseases today. It can occur at any age, but primarily affects individuals aged 50-70 [2]. The hallmark symptom of ischemic stroke is the abrupt onset of focal neurological impairment, such as dysphasia, hemianopsia, and sensory loss. These symptoms can potentially evolve into chronic

conditions like dementia and hemiplegia.

In clinical practice, the severity of stroke is determined based on the location, size, and density of the infarction. Stroke has no obvious warning signs, and its progression of the disease can be rapid. Within hours of onset, the patient's body mechanism can be seriously affected. This can lead to hypoxia in the brain, thrombosis in the carotid or internal carotid artery, and widespread cerebral infarction. At present, the first six hours after onset are considered the 'golden hour' [3]. During this period, both short-term and long-term treatment are very ideal, and the patient has the possibility of complete recovery. The timeliness of diagnosis is crucial for treatment, requiring rapid localization and quantification of the lesion after onset. This process can be influenced by subjective factors such as doctor's experience, resulting in diagnostic errors, or even missing the optimal opportunity for treating ischemic stroke patients. Therefore, how to save the time and energy of doctors while observing ischemic stroke lesions in CT, and is still an urgent computer-aided diagnosis method in clinical practice [4-6].

In the past, there have been many methods for the segmentation of stroke lesions. For example, texture-based feature extraction algorithms are widely used in the segmentation of brain lesions in medical imaging [7]. Another approach is to use the Gray Level Co-occurrence Matrix [8] (Gray Level Co-occurrence Matrix, GLCM) to extract the features of the image, and then classifying these features using neural network models. In recent years, many stroke lesion segmentation methods based on random forest [9-11], which have yielded favorable outcomes. But they rely heavily on manual feature extraction and have more complex steps. However, most of these methods can only analyze and apply to ischemic stroke MRI images and low-level image features, lacking robustness for CT images with low signal-to-noise ratio and artifacts. U-Net [12] is a classic 2D segmentation network, which is used in many medical image lesion segmentations, but it still lacks attention to image features. Some scholars have improved U-Net deficiency, and proposed KiU-Net [13], which further captures edges and minute anatomical structures that are often overlooked by other approaches.

Although there have been improvements in image segmentation networks for ischemic stroke in recent years, their segmentation accuracy still requires more in-depth research and improvement. In this paper, we presented LiU-Net network for segmentation of ischemic stroke. Compared to previous networks, LiU-Net network differs from them in the following aspects. First, LiU-Net network synthesizes the strengths of both undercomplete U-Net and overcomplete network Kite-Net, which can learn image detail features and global structural features at the same time.

Manuscript received July 25, 2023; revised December 2, 2023.

Y. W. Li is a postgraduate student of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China (e-mail: 1294999811@qq.com).

X. X. Zhang is a Professor of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, 11041, China (corresponding author, phone: 86-0412-5929812; e-mail: aszhangxx@163.com).

L. Z. Liu is a postgraduate student of School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China (e-mail: liu_luzhou@qq.com).

Second, the axial self-attentive module is embedded into KiU-Net network, which endows the network with a powerful image spatial structure encoding ability, and enables the network to better learn the characteristics of ischemic stroke lesions. In addition, we introduced gate factors into the axial self-attention module, enabling the network to acquire more accurate feature information during upsampling and downsampling. Finally, during the training process, as the network structure is getting deeper and deeper, the gradient of KiU-Net network disappears. To heighten the feature maps at different depths of the network, we introduced residual connection modules. This approach allowed for the integration of features at various levels, resulting in effective fusion. The parallel network architecture with these changes is the proposed LiU-Net. To validate the efficacy of the improved network in ischemic stroke segmentation, we applied to the Longcheng District People's Hospital of Chaoyang City, Liaoning Province. We obtained ischemic stroke slice samples under the guidance of professional doctors with 30 years of experience to form an ischemic stroke dataset, and annotated by MITK Workbench software. The finding of the experiment indicate that the performance of LiU-Net is better than that of the original U-Net, KiU-Net, and some recent new networks. It has achieved remarkable results in the segmentation small anatomical landmarks and fuzzy noise boundaries.

II. MATERIALS AND METHOD

A. U-Net Segmentation Network

U-Net is an FCN-based segmentation network proposed by Ronneberger et al. [14] in 2015 for the ISBI Challenge in 2015. U-Net increases the number of channels by upsampling, which allows it to propagate contextual information to higher resolution. By cleverly combining high-level and low-level information through residual connections, and the deep abstract information at the decoding layer can better utilize the shallow information transmitted by the encoding layer, making the image segmentation effect better [15].

U-Net consists of two parts, the left side is the contraction path, and the right side is the expansion path. The purpose of

the contraction path is to capture content, while the role of the expansion path is to aid in precise positioning. As shown in Fig. 1, its expansion path is symmetrical to the contraction path, forming a U-shaped segment, hence the name U-Net.

Its emergence to some extent solves the problems of Convolutional Neural Network (CNN [16]) that require a large amount of labeled data for training, and the high cost of data labeling for medical images. By establishing different scale feature fusion channels between the symmetrical encoder and decoder, the network can better capture both global and local features of images. It is also suitable for medical image segmentation tasks with limited data annotation.

The convolution structure used in the standard U-Net network structure is unified into 3x3 convolution kernels, followed by four pooling layers, with a total of five scales used for dimensionality reduction. The total number of feature channels in U-Net can reach thousands, resulting in a large number of training parameters. During training, the encoding and decoding paths need to extract deep features repeatedly. Due to the abstraction and low-resolution characteristics of deep features, the difficulty of training increases, and even the training is unstable and insufficient.

Compared with U-Net, the current U-Net3+ [17] has limitation such as excessive parameter quantity and redundancy, which may have some drawbacks for small sample and small target ischemic stroke lesion segmentation.

B. Multi-branch Network KiU-Net

U-Net is an outstanding deep learning model, and has been widely used in medical image segmentation in recent years. However, in previous studies, its detection results have decreased in terms of reference value when detecting smaller anatomical structures with fuzzy noise boundaries. Vishal M. Patel et al. [18] proposed an overcomplete network architecture called Kite-Net (Ki-Net) [19] to solve this problem. In terms of spatial representation, it can project data into higher dimensions. Because the shape of Ki-Net is similar to that of a kite, it is called Kite Net. The knowledge acquired by Ki-Net enables to capture better shape contours and edges in comparison to the entire network. At the same

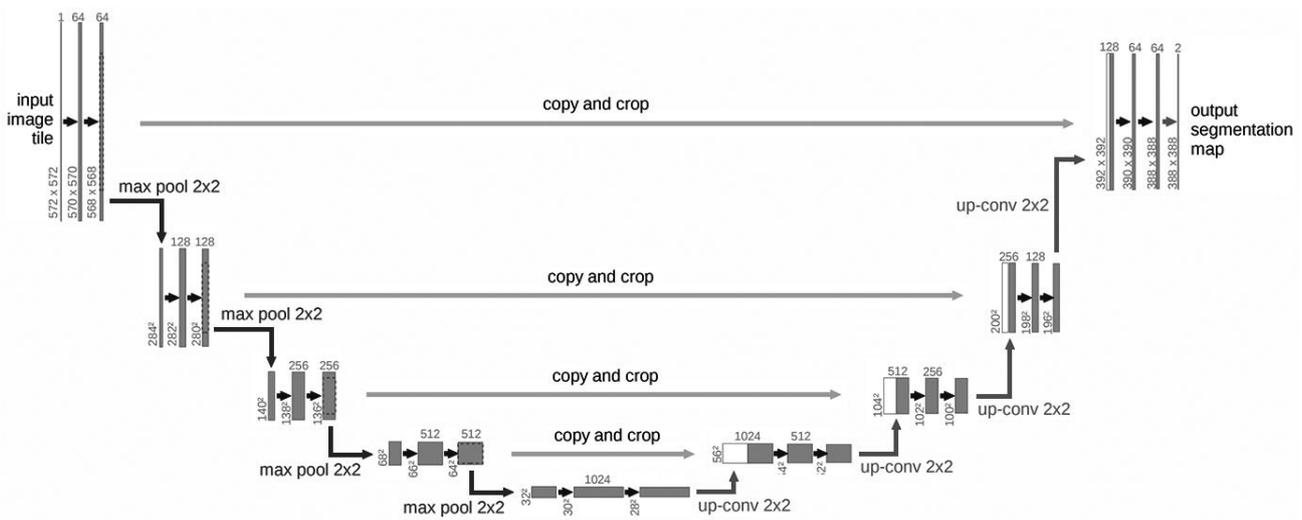


Fig. 1. U-Net network structure.

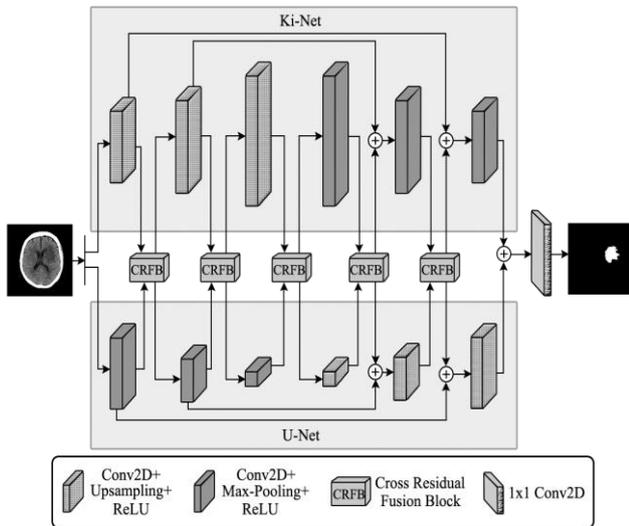


Fig. 2. Overview of KiU-Net architecture.

time, Vishal M. Patel et al. used a new cross-scale fusion strategy to effectively combine the advantages of Ki-Net and U-Net, and proposed a new architecture KiU-Net, as shown in Fig. 2.

KiU-Net is a parallel network architecture that employs parallel processing, dividing the input image into 2 branches for simultaneous analysis. One branch is Ki-Net, and the other is U-Net. The former is suitable for capturing the edge details of low-level features, while the latter is suitable for capturing high-level features of the entire image. Additionally, to more effectively combine features at various level, Vishal M. Patel et al. proposed Cross Residual Fusion Block (CRFB) [20].

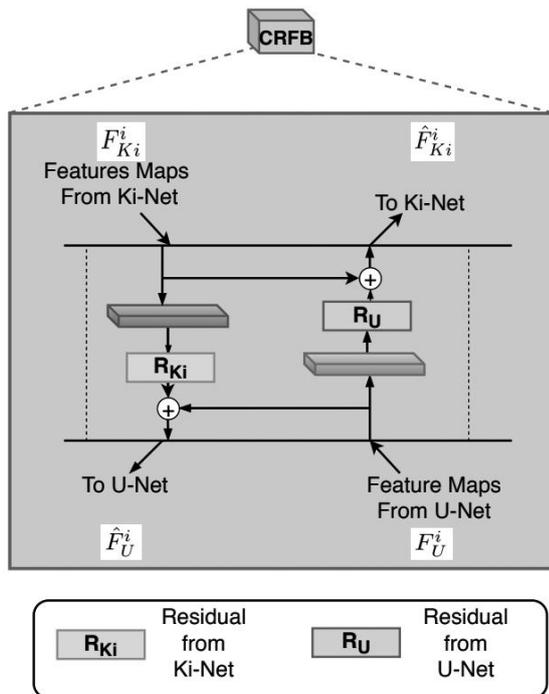


Fig. 3. Cross residual fusion block architecture.

CRFB use two branches to extracts complementary features, and sends them separately through each brach, as

shown in Fig.3. Feature maps from two networks F_U^i (U-Net) and $F_{K_i}^i$ (Ki-Net) are combined, it first estimates cross residual feature R_U^i and $R_{K_i}^i$ through a set of convolutional layers. Then, these crossed residual features are added to the original features F_U^i and $F_{K_i}^i$, resulting in complementary features \hat{F}_U^i and $\hat{F}_{K_i}^i$, that is, $\hat{F}_U^i = F_U^i + R_U^i$ and $\hat{F}_{K_i}^i = F_{K_i}^i + R_{K_i}^i$. Finally, the decoders are added to both branches, forwarded through a 1x1 conv layer, to generate the final segmentation mask.

KiU-Net combines the characteristics of undercomplete and overcomplete networks, better capturing image information that U-Net's encoder-decoder architecture. This helps achieve accurate segmentation, and obtain better overall performance. It has achieved significant improvement in segmenting small anatomical structures and blurred noise boundaries.

III. THE PROPOSED METHODS

A. Overall Structure of the Network

In order to enhance the performance of KiU-Net and prevent gradient vanishing, we have introduced residual connections [21] in the branch U-Net of KiU-Net in this paper, naming this network as RKiU-Net. Moreover, we have introduced a gate factor to improve axial self-attention mechanism, forming a new attention module IASA. Finally, IASA attention module is introduced into RKiU-Net network to form a new segmentation network, LiU-Net, as shown in Fig.4. Compared with KiU-Net, our proposed LiU-Net has improved the accuracy and efficiency of global feature extraction.

LiU-Net is composed of two parts: the overcomplete architecture Ki-Net, and the U-Net architecture that introduces the improved axial self-attention mechanism. Ki-Net is better at capturing the edge details of lesion features, while U-Net with the improved axial self-attention mechanism is responsible for capturing the overall features of lesions. The axial self-attention encoding channel is mainly responsible for improving the location correlation features between pixels in the image. The introduction of residual connections can avoid gradient vanishing caused by network deepening.

In both branches of LiU-Net, we have 3-layer conv blocks in encoder and decoder. Each encoder in KiU-Net branch consists of a single 2D conv. After the 2D conv, bilinear interpolation with a scaling factor of 2 is added, followed by applying ReLU for nonlinear activation on this encoder. Likewise, each decoder is also composed of a single 2D conv. Instead of bilinear interpolation, a max pooling layer with a pooling coefficient of 2 is added after 2D conv. In addition, in the U-Net branch, an "encoder-decoder" structure is used. The image passes through the encoding convolution module to extract low-level features, and then enters two downsampling blocks in sequence to reduce the spatial size and obtain advanced features.

After each downsampling block, the number of channels is doubled. Then, the downsampling end is sent to the improved axial self-attention module to aggregate global information

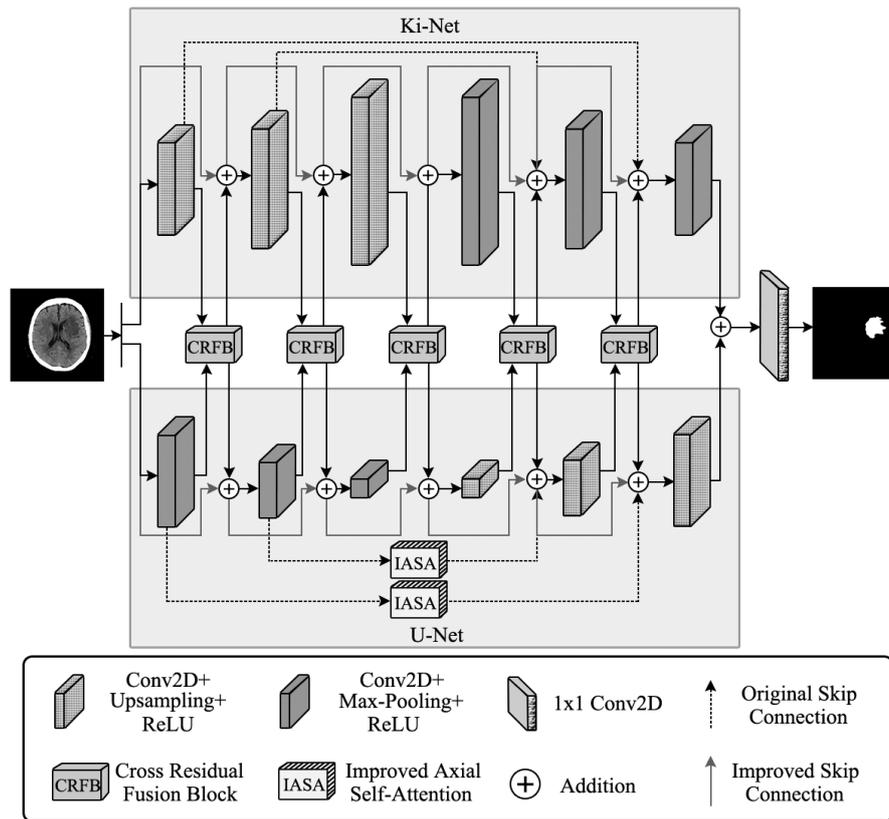


Fig. 4. LiU-Net network structure.

and generate the output of the encoder. Correspondingly, the decoder uses three upsampling blocks to restore its corresponding feature images. In each level of the encoder and decoder of both branches of LiU-Net, residual connections are incorporated into each conv block. The architectural details of LiU-Net are illustrated in Fig.4, where residual connections are represented by solid line arrows.

To adapt the segmentation task of ischemic lesions with different positions, shapes, and sizes for LiU-Net, we enhance the attention of the segmentation network to position encoding information. By adding an improved axial self-attention mechanism in the long connections of the branch U-Nets, it has a specific structure to process encoded position information, pays more attention to locations with rich feature information, and improves the segmentation performance of the network.

B. Introducing KiU-Net with Residual Connection

Residual connections are mainly used in the training of deep neural networks, especially when the number of layers in the network is large. The basic idea is to use multi-layer convolution to fit a residual mapping $F(x)$, which is easier to optimize than directly learning an approximate identity mapping. Assuming the input variable is x , residual connections enable the direct transmission of the input variable x to the output through a "shortcut connection", serving as the initial outcome, and the output result is:

$$H(x) = F(x) + x \tag{1}$$

When $F(x)=0$, $H(x)=x$, which is also known as the identity mapping. In this case, the learning objective of the residual is no longer a complete output, but the difference between the target value $H(x)$ and x , namely:

$$F(x) = H(x) - x \tag{2}$$

The way of "shortcut connection" can skip one or multiple layers and perform an identity mapping. This method can avoid problems such as gradient vanishing caused by network deepening, making the network better optimized.

KiU-Net may not achieve idea results during training due to the small size of lesions and limited experimental data. Therefore, this paper introduces residual connections in the convolutional layer, which call RKiU-Net as shown in Fig.5.

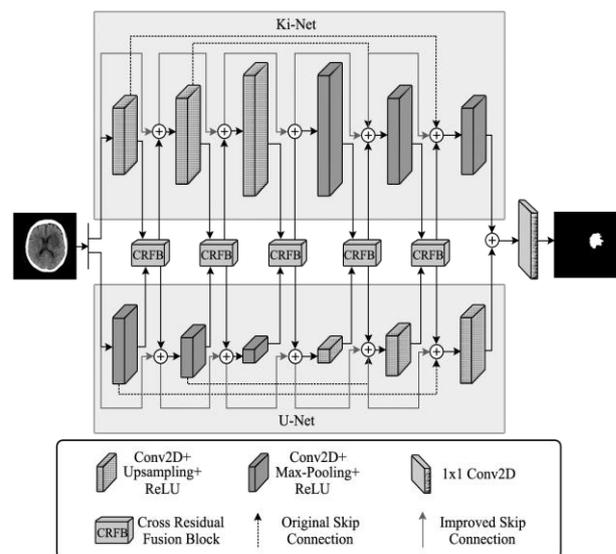


Fig. 5. RKiU-Net network structure.

First, residual connections can solve the problem of gradient vanishing. When segmenting small lesions in stroke

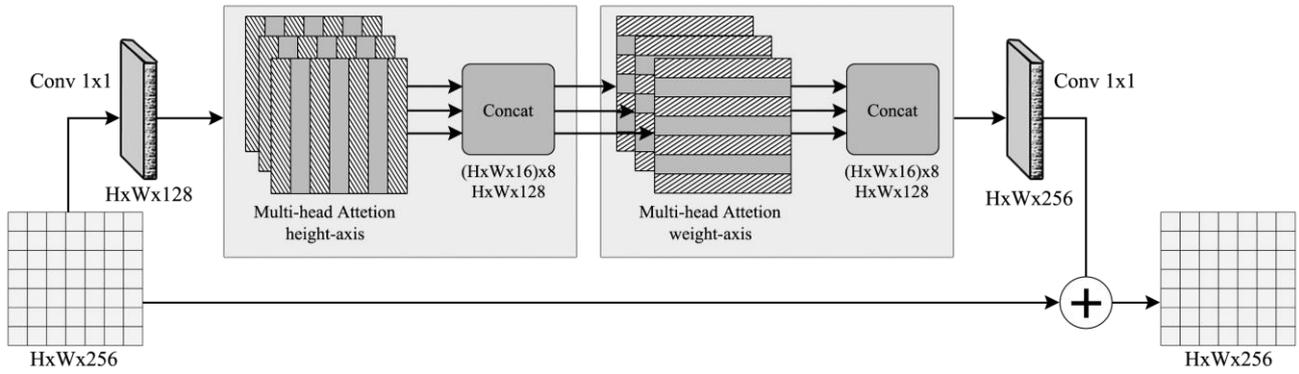


Fig. 6. Axial self-attention module.

patients, as the depth of neural networks increases, the propagation of gradients is easily affected by gradient vanishing, making training difficult. Residual connections can provide a shortcut through direct connecting across layers, thus mitigating gradient vanishing. Second, due to the ability of residual connections to provide direct connections across layers, it can help models learn features better, further improving model performance. Finally, residual connections can accelerate model training and reduce training difficulty, making it easier for models to converge and reducing the consumption of computing resources.

C. Improved Axial Self-Attention IASA

Self-attention mechanism [22] is usually used as a module to enhance the output of CNN, and has been successfully applied in many computers vision tasks. When the input feature map is complex, self-attention mechanism has a high computational complexity, thus limiting its application in some scenarios. In addition, self-attention mechanism does not contain any location information, when calculating non-local context dependency, but location information is very critical in vision tasks. To overcome these two limitations and retain the advantages of the self-attention mechanism, Jonathan Ho [23] et al. proposed a module called axial self-attention, which deconstructs the two-dimensional self-attention into two One-dimensional self-attention, the schematic diagram of the structure is shown in Fig.6.

However, since the axial attention mechanism needs to calculate the relative position codes of each position, and other positions, the fixed-position code method is used. The network has problems such as high computational complexity, and insufficient flexibility to adapt to different positions of lesions.

To improve the shortcomings of the network and enhance its efficiency, we introduce gate factors into the axial self-attention mechanism [24], forming improved axial self-attention (IASA), which a structure shown in Fig. 7. By improving computational efficiency and flexibility, we aim to enhance the performance of the model while also enabling it to have strong ability to encode spatial structural information of images.

For a given input feature map $x \in R^{h \times w \times d_{in}}$, where h is the height, w is the width, and d_{in} is the number of channels, the expression for the self-attention mechanism along the x-axis with position encoding is as follows:

$$y_o = \sum_{l \in \omega, x \in m(o)} \text{softmax}_l \left(q_o^T k_l + G_Q q_o^T r_{l-o}^q + G_K k_l^T r_{l-o}^k \right) \left(v_l + G_V r_{l-o}^v \right) \quad (3)$$

Where w represents the local grid area of pixels. The $q_o = W_Q x_o$ represents Queries. The $k_o = W_K x_o$ represents Keys. The $v_o = W_V x_o$ is the linear projection of all $x_o (v_o \in w)$. W_Q , W_K , and W_V are the transformation matrices for these three projections. The Softmax_l represents a softmax activation function applied at position $l = (a, b) \in w$ in l . The $w_{l \times m}(o)$ is the x-axis area where position $o = (i, j)$ is located. The learnable vectors $r_{p-o}^q \in R^{d_q}$, $r_{l-o}^k \in R^{d_k}$, and $r_{l-o}^v \in R^{d_v}$ are added relative positional encodings. The inner product $q_o^T r_{p-o}^q$ and $k_l^T r_{l-o}^k$ represent key and query dependencies on relative positional encodings.

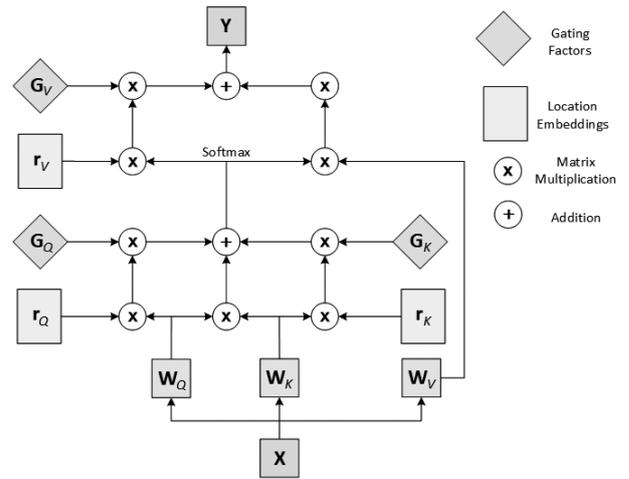


Fig. 7. IASA attention structure.

In expression (3), the input feature map x is transformed through W_Q , W_K , and W_V to generate q_o^T , k_l^T , and v_l . The corresponding structural diagram in Fig.7. corresponds to expression (3). " \otimes " represents multi-dimensional matrix multiplication. " \oplus " represents element-by-element addition of multidimensional matrices. r_Q , r_K , and r_V represent learnable relative positional encodings, and y is the output feature map. softmax means use softmax activation function for the last dimension of multidimensional matrix. r_Q , r_K , r_V are learnable relative position codes, y is the output feature map. The green diamond box contains G_Q , G_K , $G_V \in R$, which are the learnable gate factors introduced in this paper based on the position-sensitive axial attention. Together, they control the influence of learning relative position encodings on encoding global context information. If the relative positional encoding is learned more accurately, the

corresponding gate weight will be higher; conversely, if it is less accurate, it will be lower. This enables more accurate learning of relative positional encoding information on small-scale medical image datasets. Similarly, the calculation method for the y-axis is similar.

IV. EXPERIMENT

A. Datasets Information

In the experiment, the dataset for segmentation ischemic strokes includes 245 sets of CT images. Each 3D brain CT images contains 24 to 32 two-dimensional axial slices. All slices were shuffled, totaling 6862 two-dimensional CT image slices. Among them, there are a total of 2314 sample of two-dimensional CT images with ischemic areas, 2314 labeled sampled, and a resolution of 512x512. To save computing resources and facilitate network training, the resolution of each two-dimensional image slice in the dataset was adjusted to 128x128. Finally, the training set, validation set, and test set were split into 3 parts in ratio of 6:2:2.

In this paper, the dataset used was annotated on the x-axis of CT image slice using the brush mode of the MITK Workbench software [25] to mark lesions present in ischemic stroke lesions as a reference standard for segmentation. The images of ischemic stroke lesions have issues such as blurred edges and diffuseness that require guidance and verification by professional neuroradiologists to ensure accuracy of labeling. Through the MITK Workbench software, we can obtain a binary gold standard image in NIFTI format from the original CT image. In order to better process it, we also convert the DICOM format data of the original CT into NIFTI. We record the category of each patient (1 means ischemic stroke patients, 0 means non-ischemic stroke patients) and the corresponding CT data file in the form of XML file.

B. Evaluation Indicators

In order to evaluate the performance of the LiU-Net, this paper adopts Dice coefficient (*Dic*), mean Intersection over Union (*mIoU*), precision (*Pre*), recall (*Rec*), and accuracy (*Acc*) as evaluation metrics. In the experiments, ischemic stroke can be categorized as true negative/positive or false negative/positive. True Positives (*TP*) indicates that both the predicted result and the actual result are ischemic strokes. True Negatives (*TN*) indicates that the predicted result is not ischemic stroke while the actual result is. False Positive (*FP*) indicates that the predicted result is ischemic stroke while the actual result is not. False Negative (*FN*) indicates that both the predicted result and the actual are not ischemic strokes. Precision and recall, which are often used to evaluate the robustness of binary classification models. In (4) and (5), *X* represents the predicted result value, and *Y* represents the real label value.

Dice is used to evaluate the overall accuracy of the network, which can provide a more intuitive analysis of the network's strengths and weaknesses. Dice is shown in (4):

$$Dic = 2 \frac{|X \cap Y|}{|X| + |Y|} \tag{4}$$

mIOU is a widely used segmentation metric for calculating the ratio of intersection and fusion between predicted and true segmentation. It can be defined as follows:

$$mIoU = \frac{X \cap Y}{X \cup Y} \tag{5}$$

Precision is a measure of the ability of the classifier to correctly identify samples. It represents the proportion of correctly predicted samples out of all positively identified samples, and is calculated as (6) :

$$Pre = \frac{TP}{TP + FP} \tag{6}$$

Recall is the ratio of the number of Positive samples correctly predicted by the classification network to all true Positive samples, as shown in (7):

$$Rec = \frac{TP}{TP + FN} \tag{7}$$

Accuracy is the proportion of pixels correctly assigned to the objective region or background in the whole image, as shown in (8):

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \tag{8}$$

C. Experimental Environment Settings

In this paper, LiU-Net was used to segment lesions in the ischemic stroke dataset to prove the effectiveness of this network in dividing lesions in ischemic stroke. In addition, to prevent overfitting, the use of an early-stop training strategy was implemented to avoid overfitting of the model. If the network's performance fails to enhance after an additional 10 training epochs, the network training is terminated. The experimental environment used in this experiment is based on the Pytorch framework and Python 3.6. The experiments were conducted in a computer with Intel core i5-12600f CPU, 16.0 GB RAM, and NVIDIA GeForce RTX 2080Ti GPU. Adam optimization method is used to optimize parameters, and the learning rate is set to 1×10^{-4} , with a batch size of 10, and the training is performed for 100 iterations.

D. Experimental Results

U-Net is considered to be the baseline for image segmentation tasks, and KiU-Net is an extension of U-Net and Ki-Net architectures based on parallel networks. Therefore, we compare LiU-Net with KiU-Net to demonstrate its effectiveness on the stroke dataset. To verify the contributions of residual connections and improved axial attention self-attention (IASA) to ischemic stroke

TABLE I
RESULTS OF ABLATION EXPERIMENTS ON THE ISCHEMIC STROKE DATASET

Method	Residual Connection	IASA	Acc	Dic	mIoU	Pre	Rec
KiU-Net			0.9189	0.8901	0.8312	0.8814	0.8992
Network 1	√		0.9245	0.8943	0.8446	0.8831	0.9058
Network 2		√	0.9389	0.9004	0.8564	0.8876	0.9136
Network 3	√	√	0.9501	0.9118	0.8635	0.8962	0.9279

segmentation, we conducted ablation experiments, network comparison experiments, and comparative experiments with different network models facing varying degree of complexity in lesions segmentation.

We first conducted ablation experiments, detailed results are shown in Table I. In the network model, the encoder features include a lot of local information, while the decoder includes more semantic information. Combining these two types of features will cause redundant feature information between local and semantic information. This paper introduces residual connections to alleviate this issue. Comparing with evaluation metrics of KiU-Net and Network 1, all indicators have improved. Similarly, IASA improved the high computational complexity and inflexibility in adapting to different positions of lesions, enhanced useful features and suppressed weak features to improve representation ability.

When comparing KiU-Net with Network 2, using IASA improved the model's attention to relevant features. *ACC* increases by 2.18%, *Dic* increases by 1.16%, *mIoU* increases by 3.03%, *Pre* increases by 0.7%, and *Rec* increases by 1.6%. In Network 3, applying residual connections and IASA to KiU-Net yielded even more significant improvements compared to adding only one of these modules individually. Specifically, *ACC* increases by 3.4%, *Dic* increases by 2.44%, *mIoU* increases by 3.89%, *Pre* increases by 1.68%, and *Rec* increases by 3.19%.

The experimental results show that through the improved feature extraction method, LiU-Net increases the attention to relevant features and obtains the distinct characteristics of different receptive fields. It effectively improves the

performance of ischemic stroke lesion segmentation, making the segmented boundary smoother and closer to the Ground Truth.

To better understand the segmentation effect of each network in the ablation experiment, the segmentation results of the baseline and the network model after introducing different modules on the dataset are shown in Fig.8, where Fig.8 (a) is the Ground Truth. By comparing the results, before improving the network, the segmentation of the lesion edges was relatively rough and the segmentation of details was poor, resulting in insufficient segmentation, as shown in the result of Fig.8 (b). After adding the residual connections to the baseline, the edge segmentation became closer to the identified lesion, as shown in Fig.8 (c). However, for the case of lesion edges, insufficient segmentation still occurred, which was far from the Ground Truth and could not meet the needs of practical segmentation tasks. The result in Fig.8 (d) was obtained by introducing both residual connections and IASA into the network model, making it pay more attention to visually relevant areas and suppressing irrelevant areas, thereby reducing the probability of mis-segmentation. It can be seen that the result in Fig.8 (d) is closer to the manual segmentation result and has more accurate segmentation of lesion boundary details. This indicates that LiU-Net has more accurate segmentation results and meets the practical needs of applications.

Through ablation experiments, it is also confirmed that the modules introduced in this paper can improve the segmentation validity of network model and have a positive effect on ischemic stroke segmentation.

In addition, in order to further verify the effectiveness of

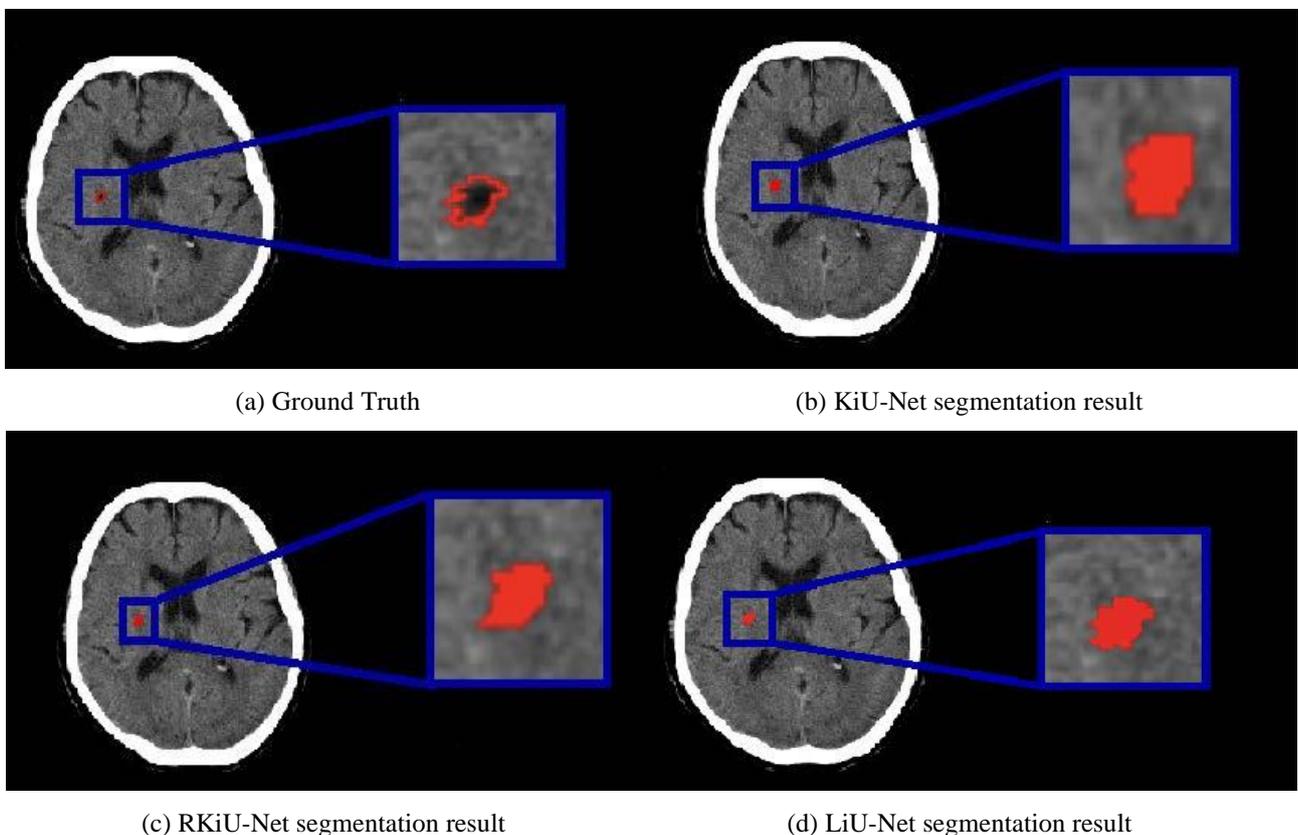


Fig. 8. Comparison chart of segmentation results.

LiU-Net and evaluate its performance, we compared the results of LiU-Net with those of recent state-of-the-art semantic segmentation algorithms on the ischemic stroke dataset, including U-Net, AU-Net, U-Net3+, and KiU-Net. The test results of LiU-Net and other network models on the ischemic stroke dataset are shown in Table II.

TABLE II
EVALUATION OF SEGMENTATION RESULTS FOR DIFFERENT NETWORKS

Method	Acc	Dic	mIoU	Pre	Rec
U-Net	0.8560	0.7528	0.6988	0.7365	0.7500
AU-Net	0.8755	0.7984	0.7343	0.7623	0.8125
U-Net3+	0.9165	0.8433	0.7904	0.7963	0.8630
KiU-Net	0.9397	0.8901	0.8312	0.8465	0.9225
LiU-Net	0.9501	0.9118	0.8635	0.8876	0.9279

From the results on the ischemic stroke dataset, it can be seen that LiU-Net outperforms other network models in most segmentation performance metrics. The advantages in *Dice* and *mIoU*, two core evaluation metrics, are more significant. LiU-Net has a *Dice* that is 2.44% higher than KiU-Net and an *mIoU* that is 3.89% higher. It terms of *Acc* and *Rec*, LiU-Net also shows improvements over KiU-Net, with *Acc* being 1.11% higher and *Rec* being 0.59% higher. The *Pre* result of LiU-Net is 0.8876, which is 4.97% higher than KiU-Net3+. All evaluation metrics indicate superiority over other network models.

Finally, in order to visually observe the segmentation effect, we divided the experimental data of ischemic stroke

into simple lesions and complex lesions for training, and compared the output results of the selected network models with Ground Truth. The effectiveness of different network models in segmenting simple and complex lesions is shown in Fig.9 and Fig.10, while Table III and Table V correspond to the experimental results of simple and complex lesions respectively.

TABLE III
EXPERIMENTAL RESULTS OF DIFFERENT NETWORKS FOR SEGMENTING SIMPLE LESIONS

Method	Acc	Dic	mIoU	Pre	F1 score
U-Net	0.8825	0.8282	0.7462	0.8082	0.8282
AU-Net	0.8997	0.8483	0.7689	0.8247	0.8483
U-Net3+	0.9278	0.8654	0.8128	0.8404	0.8654
KiU-Net	0.9488	0.8918	0.8401	0.8696	0.8918
LiU-Net	0.9621	0.9263	0.8689	0.9122	0.9263

In Fig.9, the lesions have clear edges. The segmentation results have similar contours to manual segmentation. When the lesion area is relatively complex or there are multiple lesions, as shown in Fig.10. UNet3+ and KiU-Net have segmentation results that are highly similar to Ground Truth, while U-Net and AU-Net have more segmentation error areas, resulting in larger differences between the segmentation results and Ground Truth, as shown in columns (a) and (b). These network models have issues of over-segmentation and under-segmentation.

Although U-Net3+ and KiU-Net perform better than the above two models, when faced with images with blurred

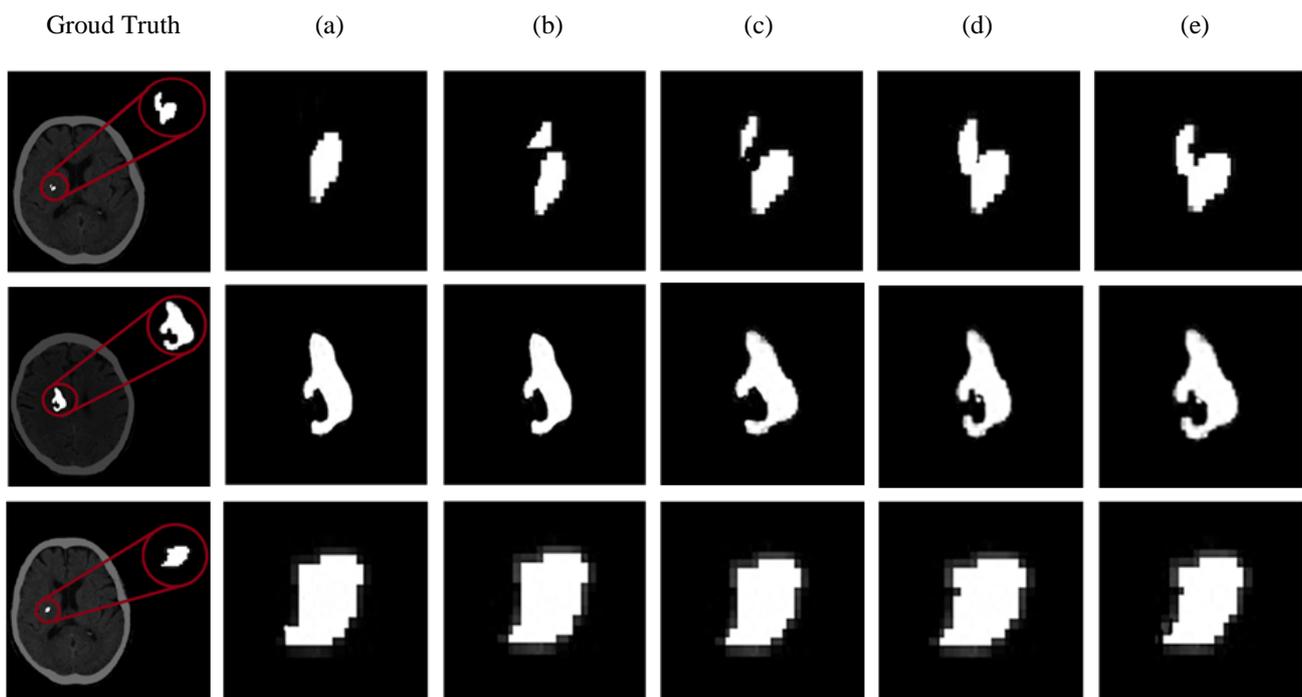


Fig. 9. Comparison of segmentation results for simple lesions. (a) segmentation result from U-Net; (b) segmentation result from AU-Net; (c) segmentation result from U-Net3+; (d) segmentation result from KiU-Net; (e) segmentation result from LiU-Net.

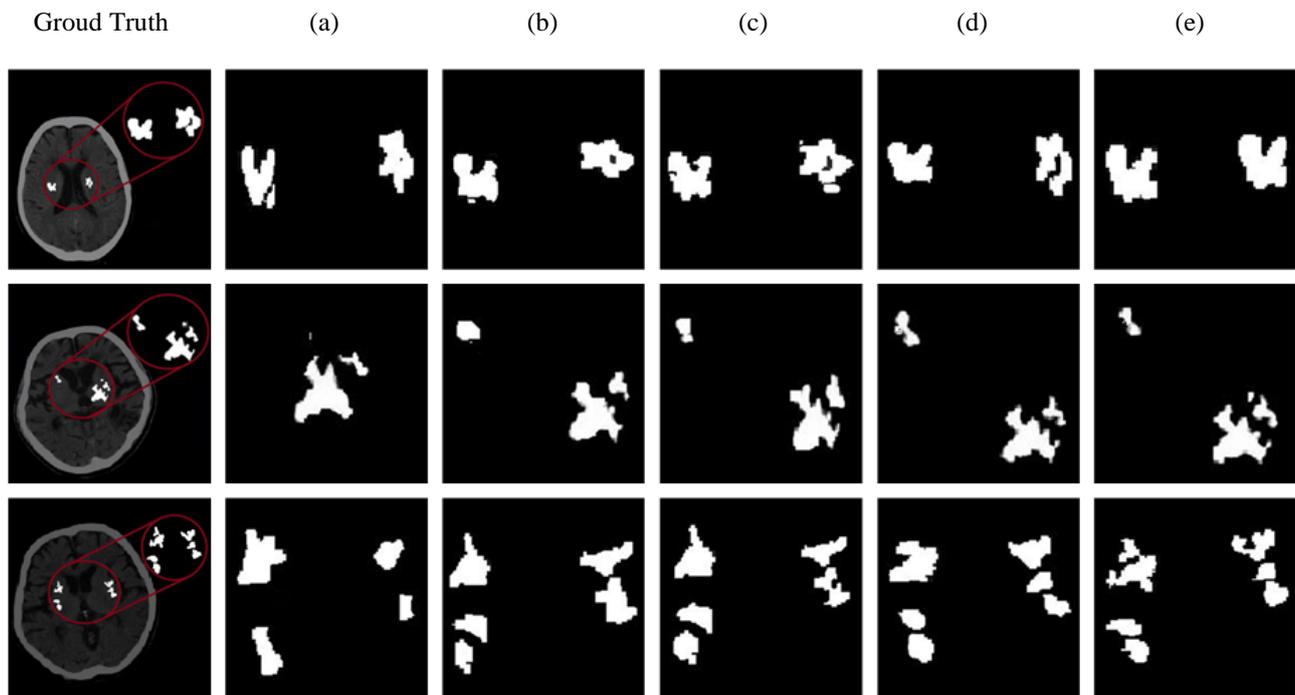


Fig. 10. Segmentation results comparison of complex lesions. (a) segmentation result from U-Net; (b) segmentation result from AU-Net; (c) segmentation result from U-Net3+; (d) segmentation result from KiU-Net; (e) segmentation result from LiU-Net.

edges, U-Net3+ does not have sufficiently smooth edges segments, and KiU-Net does not segment irregular edge

TABLE V
EXPERIMENTAL RESULTS OF DIFFERENT NETWORKS FOR
SEGMENTING COMPLEX

Method	Acc	Dic	mIoU	Pre	F1 score
U-Net	0.8522	0.7689	0.6879	0.7496	0.7489
AU-Net	0.8699	0.7822	0.7328	0.7588	0.7822
U-Net3+	0.9122	0.8233	0.7826	0.7969	0.8233
KiU-Net	0.9184	0.8469	0.8102	0.8204	0.8469
LiU-Net	0.9422	0.9002	0.8469	0.8768	0.9002

details as well as LiU-Net does. As shown in column (e), LiU-Net can well preserve detailed features of the lesion edge. It not only reduces the classification errors of noise pixels, but also provides segmentation results closer to Ground Truth for images, with tiny lesion areas and complex lesion edges.

V. CONCLUSION

In this paper, a new segmentation network model called LiU-Net is proposed for the complex features of ischemic stroke lesions, such as blurred edges and varied shapes. LiU-Net introduces residual connections and an improved axial self-attention (IASA). By incorporating residual connections, the problem of gradient disappearance caused by deepening the network can be avoided, and the training speed can be accelerated while reducing the difficulty of training. The attention mechanism enables U-Net to focus on and utilize low-level features extracted at different levels, and

then fuse them with high-level features. This allows the segmentation network to pay attention to lesion regions and non-lesion regions, boundary features, and feature channel information, thereby fully learning the characteristics. Therefore, LiU-Net can better learn features and further improve the performance of the model, making it more accurate in segmentation while addressing optimization difficulties.

Public datasets on ischemic stroke are difficult to obtain, so we built our own datasets for training, testing, and validation. This dataset includes not only simple cases but also challenging ones. It contains small size lesion images with unclear boundaries, as well as many interferences such as noise and other parts of cells that can affect the segmentation results. For some complex lesion images, the boundaries of the focus are very blurry and difficult to distinguish, and the shape, size, structure, and location of the lesions vary greatly. These factors make this dataset the most challenging one.

Both for simple lesions and those with challenges, LiU-Net outperforms the state-of-the-art network models. Through extensive experiments, it has been shown that there are still many shortcomings when using U-Net and KiU-Net for segmentation. When using LiU-Net, there are improvements in *Dice*, *Acc*, and *mIoU* metrics compared to U-Net of 15.90%, 9.14%, and 16.47%, respectively. Compared to KiU-Net, there are improvements of 2.44%, 3.40%, and 3.89% in these metrics respectively. LiU-Net achieves higher segmentation accuracy, with results closer to Ground Truth. U-Net and KiU-Net may completely lose the segmented object. In contrast, LiU-Net is more reliable and robust. It can detect blurred boundaries and avoid noise interference. Even in challenging cases, LiU-Net demonstrates stronger ability to capture details.

In conclusion, the proposed LiU-Net in this paper has

advantages such as high segmentation accuracy, fast speed, and no need for manual intervention. A specific brain stroke automatic segmentation network model is designed for patients with ischemic stroke. Based on this network model, computer-aided diagnosis systems have certain reference significance in the medical field, which can further assist doctors in objectively diagnosing, evaluating lesions, and planning treatments.

REFERENCES

- [1] Kassebaum N J, Bertozzi-Villa A, Coggeshall M S, et al. Global, regional, and national levels and causes of maternal mortality during 1990-2013: a systematic analysis for the global burden of disease study 2013[J]. *The Lancet*, 2014, 384(9947): 980-1004.
- [2] Wang W, Wang D, Liu H, et al. Trend of declining stroke mortality in China: reasons and analysis[J]. *Stroke and Vascular Neurology*, 2017, 2(3): 132-139.
- [3] Maier O, Menze B H, Vonder G J, et al. ISLES2015 — a public evaluation benchmark for ischemic stroke lesions segmentation from multispectral MR [J]. *Medical Image Analysis*, 2017, 35: 250 – 269.
- [4] Mullins M E, Schaefer P W, Sorensen A G, et al. CT and conventional and diffusion-weighted MR imaging in acute stroke: study in 691 patients at presentation to the emergency department[J]. *Radiology*, 2002, 224(2): 353-360.
- [5] Frdric J, Clarisse F, MICHEL G, et al. An alternating projection-image domains algorithm for spectral CT[C]//Proceedings. IEEE International Symposium on Biomedical Imaging, United states, 2020: 187-190.
- [6] Geng M F, Tian Z F, Jiang Z, et al. PMS-GAN: Parallel multi-stream generative adversarial network for multi-material decomposition in spectral computed tomography[J]. *IEEE Transactions on Medical Imaging*, 2021, 40(2): 571584.
- [7] Zhi C H, Shi Y R, Dang U L, et al. Application of feature extraction combined with machine learning in pathological classification and diagnosis of lung adenocarcinoma and lung squamous cell carcinoma [C] // Proceedings of the 2022 Chinese Medical Equipment 37.08(2022): 42-45+69.
- [8] Karthikeyan S M, Ezhilarasi M. Automatic stroke lesion segmentation from diffusion weighted MRI images [J]. *International Journal of Engineering Research & Technology*, 2016, 7 (2) : 111 – 115.
- [9] Shivahare B D, Gupta S K. Efficient COVID-19 CT scan image segmentation by automatic clustering algorithm[J]. *Journal of healthcare Engineering*, 2022, 2022.
- [10] Biratu E S, Schwenker F, Ayano Y M, et al. A survey of brain tumor segmentation and classification algorithms[J]. *Journal of Imaging*, 2021, 7(9): 179.
- [11] Le F L, Nori A V, Ancha S, et al. Lifted auto-context forests for brain tumor segmentation [C] // Proceedings of the 2016 International Workshop on Brain lesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, LNCS 10154 . Berlin: Springer, 2016: 171 – 183.
- [12] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[C] *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2015: 234-241.
- [13] Valanarasu J M J, Sindagi V A, Hacıhaliloğlu I, et al. Kiu-net: Towards accurate segmentation of biomedical images using over-complete representations[C]//*Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*. Springer International Publishing, 2020: 363-373.
- [14] Yang Z, Li S. Dual-path network for liver and tumor segmentation in CT images using Swin Transformer encoding approach[J]. *Current medical imaging*, 2023, 19(10): 1114-1123.
- [15] Che H, Chen S, Chen H. Image Quality-aware Diagnosis via Meta-knowledge Co-embedding[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 19819-19829.
- [16] Zhou R, Hu S, Ma B, et al. Automatic Segmentation of MRI of Brain Tumor Using Deep Convolutional Network[J]. *BioMed Research International*, 2022, 2022.
- [17] Huang H, Lin L, Tong R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]//*ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2020: 1055-1059.
- [18] Valanarasu J M J, Sindagi V A, Hacıhaliloğlu I, et al. Kiu-net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation[J]. *IEEE Transactions on Medical Imaging*, 2021, 41(4): 965-976.
- [19] Tummala B M, Barpanda S S. Curriculum learning based overcomplete U-Net for liver tumor segmentation from computed tomography images[J]. *Bulletin of Electrical Engineering and Informatics*, 2023, 12(3): 1620-1629.
- [20] Jabeen K, Khan M A, Alhaisoni M, et al. Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion[J]. *Sensors*, 2022, 22(3): 807.
- [21] Cheng J, Tian S, Yu L, et al. ResGANet: Residual group attention network for medical image classification and segmentation[J]. *Medical Image Analysis*, 2022, 76: 102313.
- [22] Xiang T, Zhang Y, Lu Y, et al. SQUID: Deep Feature In-Painting for Unsupervised Anomaly Detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023: 23890-23901.
- [23] Ho J, Kalchbrenner N, Weissenborn D, et al. Axial attention in multidimensional transformers[J]. *arXiv preprint arXiv:1912.12180*, 2019.
- [24] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: Learning where to look for the pancreas[C]//*1st Conference on Medical Imaging with Deep Learning (MIDL2018)*. Amsterdam: MIDL, 2018:4861068.
- [25] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2020:13001-13008.
- [26] Li R, Li M, Li J, et al. Connection sensitive attention U-NET for accurate retinal vessel segmentation[J]. *arXiv preprint arXiv: 1903.05558*, 2019.