Proceedings of the 6th International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/2026.33682

An Improved ResNet-Based U-Net Model for Small Blood Vessel Segmentation

Zhen Du

School of The First Clinical Medical College, Shenyang Medical College, Shenyang, China
2287695985@qq.com

Abstract. Accurate segmentation of cerebral blood vessels remains challenging due to the slender morphology, low contrast, and susceptibility to structural discontinuities of small vessels in cerebral vascular images. To address this issue, this paper proposes an improved ResNet-based U-Net model. By introducing residual connections into the classic U-Net encoder-decoder architecture, the model enhances feature reuse and gradient propagation, thereby improving the recognition of fine-grained vascular structures. Experimental results on the DIAS cerebral DSA dataset demonstrate that the proposed method outperforms the traditional U-Net in terms of both regional overlap and topological continuity, effectively improving the completeness and continuity of small vessel segmentation, and providing a feasible technical pathway for clinical cerebrovascular analysis.

Keywords: Vessel image, Res-Net, U-Net, image segmentation, cerebrovascular analysis

1. Introduction

Cerebrovascular diseases are some of the most widespread diseases across the world, and they lead to very high rates of illness and death [1, 2]. Early diagnosis and accurate treatment are very important to help patients recover better. In clinical work, digital subtraction angiography (DSA) and computed tomography angiography (CTA) are important imaging methods. They are used to check the shape of brain blood vessels, find damaged areas and make surgery plans. Correctly dividing brain blood vessel structures from these images is a key step in brain blood vessel image analysis. The correctness of division results directly affects how reliably we find damaged areas. It also influences the effect of later 3D reconstruction, quantitative analysis and surgery guidance. Especially for clinical surgery guidance, doctors need clear, continuous and complete 3D blood vessel structures. These structures help them know exactly the spatial relation between lesions and nearby blood vessels. If division results have breaks, missing parts or wrong divisions, it will badly affect the making of surgery plans. So, improving the correctness of brain blood vessel division—especially better ability to find small vessel branches—has great practical use and value in clinical work.

In recent years, deep learning methods have developed greatly in medical image division. They have gradually become the main research direction in this field. Among these methods, U-Net works very well in many medical image division tasks. This is because it has a symmetric encoder-decoder structure and skip connection mechanism [3]. Its core idea is to mix two kinds of features. The

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-8818/2026.33682

encoder picks up low-level spatial detail features, and the decoder restores high-level semantic features. This way, it can sort each pixel correctly and keep the position information of the target. Thanks to this structural advantage, U-Net can deal with problems well. These problems include complex target shapes and unclear edges in medical images. So it has become a classic basic model for brain blood vessel division tasks. However, even though U-Net works well in medical image division, it still has some limits in brain blood vessel division. First, its basic convolution modules can only show features in a limited way. When it deals with small, thin and widely spread vessel branches, it often misses them, causes breaks and fails to keep continuous structure. Second, brain blood vessel images often have complex backgrounds, much noise and low local contrast. These things need the model to find edges and restore details better. Third, when the network becomes deeper, traditional U-Net may have trouble in passing gradients and learning deep features. This stops the whole division effect from getting better.

The difficulties of brain blood vessel division mainly show in the following aspects. First, brain blood vessels have very complex anatomical shapes. Their whole structure has many branches, is thin and has many bends. Vessels of different levels have very different sizes and shapes. This makes it harder to build a unified model. Second, small vessels only take up a few pixel areas. They have features like low contrast, unclear edges and easy influence by noise. So they are more likely to be missed during division. At the same time, the gray difference between brain blood vessels and nearby tissues is often small. This brings a big challenge for the model to tell vessel edges from background areas. All these factors show that brain blood vessel division needs the model to do two things well. It needs to have good ability to understand the whole semantic meaning. It also needs to keep local detail information fully, especially the ability to sense small vessel structures and show their continuity.

To solve the above problems, this study adds residual structures to the U-Net framework. It puts forward an improved U-Net model based on ResNet [4]. Residual connections were first put forward by ResNet. Its core idea is to send input information directly through identity mapping and skip-layer connections. This well eases the gradient degradation problem in deep network training and makes better use of features. Adding residual structures to U-Net is expected to improve feature extraction ability and training stability. It still keeps the advantages of the whole encoder-decoder framework. So it can catch the fine structural features of small vessels more effectively. It also makes division results more continuous and complete.

2. Previous works

Before deep learning was widely used in medical image segmentation, cerebral vessel segmentation tasks mainly used traditional image processing and model-based methods. Typical methods included watershed algorithms, level set methods, graph cut algorithms, and region growing methods. These methods gave some technical support for extracting cerebral vascular structures in early research. But they worked poorly in complex medical imaging situations.

The watershed algorithm divides image regions by copying the "flooding" process in landforms [5]. It has clear rules, easy implementation and fast calculation speed. But this method is easily influenced by noise. In areas with dense cerebral vessels or big gray level changes, it often makes too many small segments. This breaks the whole continuity of vascular structures. Also, it cannot model high-level semantic information. So it cannot tell vessels from similar background tissues well.

Level set methods find target edges by changing implicit curves. They work well for main vessel trunks with regular shapes and smooth edges. But cerebral vessel segmentation has many small

DOI: 10.54254/2753-8818/2026.33682

vessels. These small vessels are thin, curved and have complex branches. The curve change of level set methods often stops at local best points. This makes segmentation results discontinuous or even broken. It is hard to keep distal small vessel structures complete. Besides, this method depends a lot on initial contour choice, so its stability is not good.

Graph cut algorithms get the best image segmentation by building graph models and reducing energy functions. They have good global consistency in theory. But their segmentation effect depends greatly on initial seed points. In real clinical data, vessel shapes are very different for each patient. So it is hard to keep seed points stable and usable for all cases. At the same time, graph cut methods need much calculation for high-resolution 3D medical images. This limits their speed when dealing with large-scale data.

Region growing methods start with preselected seed points. They expand regions step by step based on similarity rules [6]. They are easy to use and easy to understand. But these methods are easily affected by seed points and growth thresholds. In areas where vessels divide, cross or have similar gray levels with nearby tissues, they often expand wrongly. This leads to wrong segmentation. Also, they lack global structure rules. So it is hard to keep the complete shape of vascular structures.

In general, traditional methods have several common problems. First, they only use low-level gray, edge or texture information to show features. They cannot model complex semantic features. So they work badly for the complex shape of cerebral vessels. Second, these methods need man-made rules or early assumptions. They cannot fit different patient anatomies and imaging conditions well. So they cannot be used widely for different cases. Third, they cannot balance continuity and detection ability for small vessel branches. Noise often makes vessels broken, missing or divided wrongly. Also, it is hard to keep complete and connected shapes for vessel branches and crossings. These problems are more serious in medical images with low contrast and much noise.

These limitations make deep learning methods suitable for cerebral vessel segmentation. Compared with traditional methods, deep learning models can learn multi-level features through full-process training. They mix low-level space details and high-level semantic information well. This makes segmentation more correct and stable. Among them, U-Net and its improved forms are the main methods in medical image segmentation. But current U-Net-based models still have problems in cerebral vessel segmentation. For example, they cannot extract small vessel features well, cannot tell edges well in complex backgrounds, and may have gradient problems when the network becomes deeper.

3. Dataset and preprocessing

This study uses the DIAS cerebral angiography image dataset. It includes cerebral DSA image sequences from 60 cases, and there are 321 two-dimensional slice images in total [7]. Experienced doctors marked all images accurately by every pixel. The marked areas are the cerebral vascular structures. This offers a dependable reference standard for supervised learning. We divided the dataset into training, validation and test sets with standard experiment rules. This makes the model evaluation fair and repeatable. To be specific, the training set has 30 cases (s0–s29) with 151 image slices. The validation set has 10 cases (s30–s39) with 55 image slices. The test set has 20 cases (s40–s59) with 115 image slices. You can see the detailed grouping in Table 1. All images are grayscale pictures. Their resolution is 800 × 800, and pixel values are from 0 to 255. We should note that many slices from the same case use the same annotation. This way of organizing data matches the link between two-dimensional slices and three-dimensional vascular structures in cerebral vessel segmentation tasks. It also makes the cross-slice consistency model a bit more difficult to build.

DOI: 10.54254/2753-8818/2026.33682

Table 1. Data partitioning

Dataset	Number of Cases	Number of Image Slices	Number of Annotations
Training Set	30	151	30
Validation Set	10	55	10

The specific partitioning is shown in Table 1. During the preprocessing stage, to improve the stability of model training and accelerate convergence, Z-score standardization was performed on the image sequences of each case. Specifically, for all slices of the same case, the pixel-wise mean μ and standard deviationσ were first calculated, and then each slice was normalized according to the following formula:

$$I_{norm} = \frac{(I - \mu)}{(\sigma + \varepsilon)} \quad (1)$$

where ε is a tiny constant (set to 1×10−8) to prevent division by zero. After standardization, the pixel distribution of the images was adjusted to have a mean of 0 and standard deviation of 1, thereby reducing intensity distribution differences among different cases and improving the model's generalization ability. For annotation images, the original pixel values were divided by 255 to map the background and vessel regions to binary labels of 0 and 1, respectively.

Medical image data is limited, and cerebral vascular structures are complex and changeable. So we used many online data augmentation methods in the training stage to make the model more robust and improve its generalization ability. All augmentation steps were used randomly in each training round. They include these points: First, we cut fixed-size small image pieces (such as 256 × 256) from the original 800 × 800 images at random. This fits the input size of the model and reduces calculation cost. Second, we flipped images horizontally with a 50% chance and vertically with a 50% chance. Besides, we rotated images by 90°, 180° or 270° with a 25% chance. All the augmentation steps above used the same random seed. They worked on original images and their corresponding annotation images at the same time. This keeps the consistency of spatial position and avoids new annotation mistakes.

In the validation and test stages, we wanted to avoid information loss caused by cropping and test the whole images completely. So we used a sliding window method to process original images in small pieces. The window size was the same as that in the training stage. We controlled the stride with setting parameters to cover the whole image fully. We mixed prediction results in overlapping areas to get more stable segmentation outputs.

In practice, the data preprocessing steps mainly include image reading, normalization and data augmentation modules. For each case, the program read all its DSA images in slice index order to build the corresponding image sequence. Then it calculated the global average value and standard deviation of the sequence. It standardized each slice by subtracting the average value and dividing by the sum of the standard deviation and a small constant. The standardized image sequences were used for later model training. This improves numerical stability and helps the model converge faster.

DOI: 10.54254/2753-8818/2026.33682

4. Model the overall architecture is summarized in Table 2

Table 2. Model architecture

Layer	Output Size	Number of Residual Blocks	Parameter Count
Input	64×64×1	-	-
Encoder Block 1	32×32×64	1	0.12M
Encoder Block 2	16×16×128	1	0.47M
Encoder Block 3	8×8×256	1	1.88M
Encoder Block 4	4×4×512	1	7.52M
Center	4×4×1024	1	15.04M
Decoder Block 1	8×8×512	1	7.52M
Decoder Block 2	16×16×256	1	1.88M
Decoder Block 3	32×32×128	1	0.47M
Decoder Block 4	64×64×64	1	0.12M
Output	64×64×1	-	-
Total	-	-	19.4M

This study uses an improved U-Net built on ResNet as the basic model for cerebral vessel image segmentation. We add residual links to the classic U-Net structure. This helps the model extract features better and makes deep network training more stable. So it can raise the accuracy of cerebral vessel segmentation, especially when identifying small blood vessel branches. The model has a symmetric encoder-decoder structure as a whole. The encoder includes four downsampling stages. The numbers of feature channels are 64, 128, 256 and 512 in order, and the number reaches 1024 at the deepest layer. Each downsampling stage is made up of a residual convolution block and a downsampling module. The residual convolution block has two 3×3 convolution layers. Each layer is followed by batch normalization, Dropout and LeakyReLU activation functions. It uses identity mapping to add input features to convolution outputs and form a residual link. This design lets the network learn residual information effectively. It eases the gradient degradation problem when the network becomes deeper, and it also helps the network reuse features better. The downsampling module uses 2×2 convolution with stride 2. It reduces the spatial size and doubles the number of channels at the same time. In this way, the receptive field becomes larger gradually, and the model can extract higher-level semantic features. The deepest part of the encoder is the center module (bottleneck). It is also made of a residual convolution block. It is used to extract global semantic information and provide high-level feature representations for the decoding stage. The decoder has four upsampling stages. Each stage first uses transposed convolution. It doubles the size of the feature map and cuts the number of channels in half. Then it connects with the feature map from the corresponding encoder layer through skip links. This mixes low-level spatial detail information with high-level semantic features. The connected feature map goes through a residual convolution block for further feature refinement. So the spatial resolution of the image is restored step by step. Finally, a 1×1 convolution layer maps the feature map to the target class space. It outputs a segmentation result that has the same size as the input.

At the model design stage, the cerebral vessel segmentation task faces key challenges. They include the complex shape of small vessels, their twisted paths and low contrast with surrounding tissues. The model needs to have two good abilities. It needs to model global semantic information

DOI: 10.54254/2753-8818/2026.33682

well and keep local details. Standard U-Net eases the loss of spatial information to some extent through skip connections. But its basic convolution blocks have limited ability to express information. The network may have gradient degradation when it is made deeper. So it cannot recognize fine-grained vascular structures well. To solve this problem, this study uses residual structures to build a U-Net based on ResNet. Identity mapping helps information pass across layers. The network can stay stable in training when it is deeper. It also improves gradient flow and feature reuse. So it can capture the weak feature responses of small vessels more effectively. This works well in complex backgrounds and low-contrast situations. It makes segmentation results more continuous and complete. Besides, the LeakyReLU activation function keeps negative axis information. It lowers the risk of neuron inactivation. It helps improve the location accuracy of vessel boundaries in low-contrast areas. The model uses the AdamW optimizer to update parameters [8]. Compared with traditional Adam, AdamW controls overfitting better by separating weight decay. The initial learning rate is set to 5×10^{-4}. A cosine annealing strategy is used to make it drop gradually to 5×10^{-6}. This helps the model converge more stably. The weight decay coefficient is 0.01. The batch size is 64. The training lasts for 20 epochs. We add an early stopping mechanism. It stops training early if validation set metrics do not improve for many continuous rounds (set to 100). The loss function combines Dice Loss and cross-entropy loss with equal weights. Dice Loss focuses on regional overlap. It makes the structure more continuous. Cross-entropy loss ensures the accuracy of pixel-level classification. Both weights are set to 1.

To improve the model's generalization ability, we use many data augmentation methods in the training stage. They include random cropping to 64×64 pixels, horizontal and vertical flipping and random 90° rotations. These methods simulate different clinical imaging conditions. At the same time, input images are processed with Z-score standardization when being loaded. This improves data distribution and helps the model converge faster. The training process supports distributed data parallelism. It raises training efficiency in multi-GPU environments. In this U-Net structure based on ResNet, the encoder extracts multi-scale semantic features through step-by-step downsampling. The decoder restores spatial resolution gradually through upsampling and skip connections. Residual units are added at each stage. They improve gradient propagation and the ability to express features. So this structure balances deep semantic modeling and fine-grained feature recovery. It provides effective support for improving the performance of cerebral vessel segmentation.

5. Results

Table 3. Quantitative evaluation results on the DIAS test set

Metric	Mean ± Standard Deviation
DSC	0.7529 ± 0.0576
IoU	0.6071 ± 0.0747
Sensitivity	0.7403 ± 0.1131
Specificity	0.9821 ± 0.0156
Accuracy	0.9628 ± 0.0150
AUC	0.9799 ± 0.0102
clDice	0.6679 ± 0.0924

According to the quantitative evaluation results in Table 3, the ResNet-based U-Net model proposed in this paper demonstrates good overall segmentation performance and stability on the

DOI: 10.54254/2753-8818/2026.33682

DIAS test set. The model got a Dice Similarity Coefficient (DSC) of 0.7529 ± 0.0576. This shows the predicted results and real labels have a large overlap. Also, the Intersection over Union (IoU) reached 0.6071 ± 0.0747. This further proves that pixel-by-pixel segmentation is effective. For Sensitivity, the model reached 0.7403 ± 0.1131. This means the model can find brain blood vessel areas well. It can clearly spot blood vessel structures, including some small blood vessel branches. Also, Specificity reached 0.9821 ± 0.0156 and Accuracy reached 0.9628 ± 0.0150. These data show the model is good at telling background areas apart. It has a low wrong positive rate and its overall classification results are reliable. Also, the model's AUC value was 0.9799 ± 0.0102, nearly 1. This means the model works very well at different judgment points and has strong stability. It is worth noting that the clDice metric reached 0.6679 ± 0.0924. This metric checks the consistency and connection of blood vessel shape structures. The results show the model is good at keeping blood vessel structures continuous. But it still needs to be improved compared with area overlap measures like DSC. This also shows that cutting out small blood vessel branches is still hard when the background is complicated.

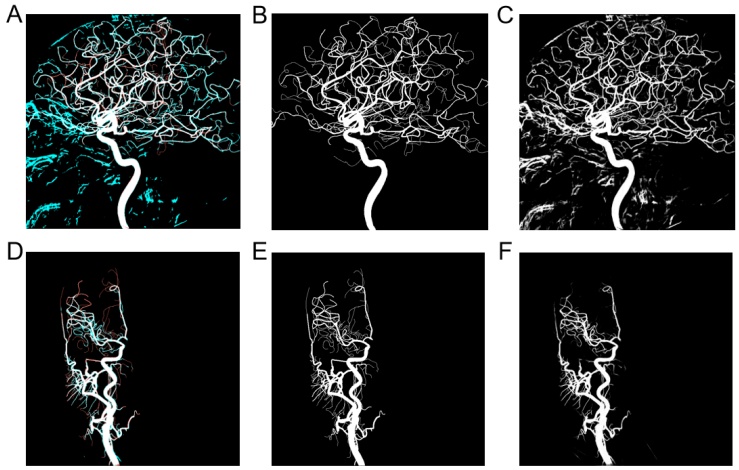

Figure 1. Example of Res-UNet cerebral vessel segmentation results

The segmentation results shown in the figures indicate that the ResNet-based U-Net model can clearly extract the main trunks and major branches of cerebral vessels, with continuous contours and good background noise suppression. The overall segmentation effect is favorable. However, in regions such as vessel terminals, tiny and weak branches, and areas overlapping with tissues, there are still a few instances of breakage, false positives, or false negatives, indicating room for improvement in segmentation accuracy. Overall, the model exhibits strong robustness and accuracy in the cerebral vessel segmentation task, making it suitable for extracting major vascular structures in complex environments.

6. Discussion and conclusion

This study focused on the difficulties of finding small blood vessels and keeping structure continuous in cerebral vessel segmentation. It provided an improved model that is built on ResNet-based U-Net. Residual structures are added to the classic U-Net structure. This model well eases the gradient degradation problem when training deep networks. It also keeps the strong points of the encoder-decoder framework. Besides, it improves the ability to reuse features. So it can model fine-grained blood vessel structures in a better way. Test results show that this method works better than

DOI: 10.54254/2753-8818/2026.33682

traditional U-Net in finding small blood vessel branches. It greatly improves segmentation continuity and boundary clearness. This proves that residual mechanisms are effective and useful for low-contrast medical image segmentation tasks.

Although this study has made some progress, the model we put forward still has some limits. First, the model may still have errors in boundary position in areas with extremely low contrast or strong noise. This means we can still improve its robustness in extreme imaging conditions. Second, the segmentation results sometimes have breaks or stick together in blood vessel bifurcation or crossing areas. This shows that the current model is not good enough at modeling complex topological structures. It is not good at showing local details and global structure together. Third, the model is easily affected by the quality of training data. Its performance may change when labels are not exact or samples are not evenly spread. This limits the model's ability to work well on real clinical data.

To solve these problems, future research can be improved in many ways. First, we can add attention mechanisms or Transformer structures that can model the whole image [9]. Their self-attention mechanisms can catch long-distance connections. This helps keep the continuity and topological consistency of blood vessel structures. Second, we can explore mixed models that combine convolutions and attention mechanisms. These models can balance getting local details and modeling global meanings. So they can improve the whole representation ability and stability [10]. Besides, we can use multi-modal imaging data such as MRI, CTA and DSA. We can mix their information to help the model recognize blood vessel structures better under different imaging conditions. At the same time, we can add blood vessel topological limits. For example, we can use post-processing methods based on graph structure or structure-aware loss functions. This can further improve the structural completeness of segmentation results. For data, we can try weakly supervised or semi-supervised learning ways. These can reduce the need for high-quality pixel-level labels. So the model can be used more widely in real applications. For engineering, we can use ways to make the model lighter such as knowledge distillation, pruning and quantization. These can reduce calculation cost. They make it easier to deploy and use the model in clinical settings.

For clinical use, improving the accuracy of cerebral vessel segmentation is very important. It is crucial for preoperative checks and intraoperative navigation of diseases such as stenosis, aneurysms and arteriovenous malformations. Especially, showing small blood vessel structures completely plays a key role. It helps assess collateral circulation ability and plan surgical paths. Besides, the method put forward in this study can be used widely. It can be extended to other tasks such as retinal vessel segmentation, tumor microvessel identification and coronary artery segmentation.

In short, the ResNet-based U-Net model put forward in this paper effectively improves two things. It enhances the ability to find small blood vessels and keep structure continuous in cerebral vessel segmentation tasks. This proves that residual structures have application value for these problems. In the future, we can add global modeling mechanisms, multi-modal information fusion and structural constraint strategies. We hope to further improve the model's performance. It can serve clinical practice better. It can also push the development of precision neurosurgery and personalized medicine.

References

[1] F. Taher, A. Mahmoud, A. Shalaby, and A. El-Baz, "A Review on the Cerebrovascular Segmentation Methods, " in 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Dec. 2018, pp. 359–364. doi: 10.1109/ISSPIT.2018.8642756.

[2] R. Mehanna and J. Jankovic, "Movement disorders in cerebrovascular disease, " Lancet Neurol., vol. 12, no. 6, pp. 597–608, Jun. 2013, doi: 10.1016/S1474-4422(13)70057-7.

DOI: 10.54254/2753-8818/2026.33682

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation, " in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition, " Dec. 10, 2015, arXiv: arXiv: 1512.03385. doi: 10.48550/arXiv.1512.03385.
[5] A. S. Kornilov and I. V. Safonov, "An Overview of Watershed Algorithm Implementations in Open Source Libraries, " J. Imaging, vol. 4, no. 10, p. 123, Oct. 2018, doi: 10.3390/jimaging4100123.
[6] A. Sakhaei and Z. Ghanbari, "Segmentation of digital images with region growing algorithm, " IOP Publishing, 2024. doi: 10.1088/978-0-7503-5924-5ch4.
[7] W. Liu et al., "DIAS: A dataset and benchmark for intracranial artery segmentation in DSA sequences, " Med. Image Anal., vol. 97, p. 103247, Oct. 2024, doi: 10.1016/j.media.2024.103247.
[8] P. Zhou, X. Xie, Z. Lin, and S. Yan, "Towards Understanding Convergence and Generalization of AdamW, " IEEE Trans. Pattern Anal. Mach. Intell., vol. 46, no. 9, pp. 6486–6493, Sep. 2024, doi: 10.1109/TPAMI.2024.3382294.
[9] S. Madan, M. Lentzen, J. Brandt, D. Rueckert, M. Hofmann-Apitius, and H. Fröhlich, "Transformer models in biomedicine, " BMC Med. Inform. Decis. Mak., vol. 24, no. 1, p. 214, Jul. 2024, doi: 10.1186/s12911-024-02600-5.
[10] S. Mekruksavanich and A. Jitpattanakul, "Hybrid convolution neural network with channel attention mechanism for sensor-based human activity recognition, " Sci. Rep., vol. 13, no. 1, p. 12067, Jul. 2023, doi: 10.1038/s41598-023-39080-y.