



An attention residual u-net with differential preprocessing and geometric postprocessing: Learning how to segment vasculature including intracranial aneurysms



Nan Mu ^{a,b}, Zonghan Lyu ^{a,b}, Mostafa Rezaeitaleshmahalleh ^{a,b}, Jinshan Tang ^c, Jingfeng Jiang ^{a,b,*}

^a Biomedical Engineering, Michigan Technological University, Houghton, MI United States

^b Center for Biocomputing and Digital Health, Health Research Institute and Institute of Computing and Cybernetics, Michigan Technological University, Houghton, Michigan, United States

^c Department of Health Administration and Policy, George Mason University, Fairfax, Virginia, United States

ARTICLE INFO

Keywords:

3D segmentation
Intracranial aneurysm
Depth-aware attention gate
Multiscale supervision
Conditional random field

ABSTRACT

Objective: Intracranial aneurysms (IA) are lethal, with high morbidity and mortality rates. Reliable, rapid, and accurate segmentation of IAs and their adjacent vasculature from medical imaging data is important to improve the clinical management of patients with IAs. However, due to the blurred boundaries and complex structure of IAs and overlapping with brain tissue or other cerebral arteries, image segmentation of IAs remains challenging. This study aimed to develop an attention residual U-Net (ARU-Net) architecture with differential preprocessing and geometric postprocessing for automatic segmentation of IAs and their adjacent arteries in conjunction with 3D rotational angiography (3DRA) images.

Methods: The proposed ARU-Net followed the classic U-Net framework with the following key enhancements. First, we preprocessed the 3DRA images based on boundary enhancement to capture more contour information and enhance the presence of small vessels. Second, we introduced the long skip connections of the attention gate at each layer of the fully convolutional decoder-encoder structure to emphasize the field of view (FOV) for IAs. Third, residual-based short skip connections were also embedded in each layer to implement in-depth supervision to help the network converge. Fourth, we devised a multiscale supervision strategy for independent prediction at different levels of the decoding path, integrating multiscale semantic information to facilitate the segmentation of small vessels. Fifth, the 3D conditional random field (3DCRF) and 3D connected component optimization (3DCCO) were exploited as postprocessing to optimize the segmentation results.

Results: Comprehensive experimental assessments validated the effectiveness of our ARU-Net. The proposed ARU-Net model achieved comparable or superior performance to the state-of-the-art methods through quantitative and qualitative evaluations. Notably, we found that ARU-Net improved the identification of arteries connecting to an IA, including small arteries that were hard to recognize by other methods. Consequently, IA geometries segmented by the proposed ARU-Net model yielded superior performance during subsequent computational hemodynamic studies (also known as "patient-specific" computational fluid dynamics [CFD] simulations). Furthermore, in an ablation study, the five key enhancements mentioned above were confirmed.

Conclusions: The proposed ARU-Net model can automatically segment the IAs in 3DRA images with relatively high accuracy and potentially has significant value for clinical computational hemodynamic analysis.

1. Introduction

Intracranial aneurysm (IA) is a cerebrovascular disease caused by the thinning and weakening of the arterial walls. There is a small (0.1% per

year) probability that an IA may rupture; the consequence of IA rupture is dire, with a high likelihood of significant morbidity and mortality. However, most IAs do not rupture and remain asymptomatic. As a result, how to manage unruptured IAs is still controversial. Since the

* Corresponding author at: M & M 309, 1400 Townsend Drive, Houghton, MI 49931, United States.

E-mail address: jjiang1@mtu.edu (J. Jiang).

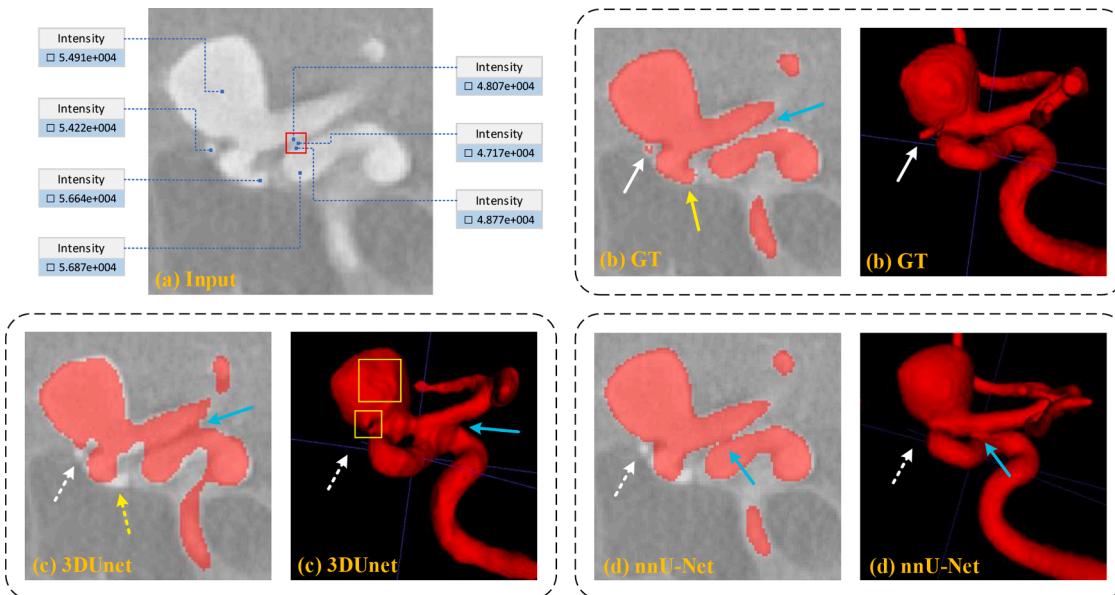


Fig. 1. A visual representation of shortcomings of two convolution-based models: (a) input 3D image from a coronal view. (b) corresponding ground truth from the same coronal view in conjunction with a 3D surface rendering. (c) and (d) are segmentation results of 3D U-Net (Çiçek et al., 2016) and nnU-Net (Isensee et al., 2021).

prevalence of unruptured IAs in the adult population is between 3 and 7%, clinical management of a large number of unruptured IAs is an unmet need, motivating continued technological developments that would accurately characterize IAs based on their risk of rupture. If successful, only IAs with a high risk of rupture would be treated immediately, while routine imaging follow-ups would be used to manage low-risk IAs conservatively.

In the last two decades, considerable research efforts have been devoted to computerized analytics of unruptured IAs (Saqr et al., 2020). Image-based computational pipelines (Antiga et al., 2008; Cebral et al., 2011) mainly consist of two major steps: 1) the segmentation of IAs, including adjacent cerebral vessels using "patient-specific" 3D images, and 2) the subsequent computational analysis, e.g., computational hemodynamics (also known as "patient-specific" *computational fluid dynamics* [CFD] simulation) (Meng et al., 2014), morphological analysis (Dhar et al., 2008), virtual intervention, and treatment planning (Damiano et al., 2020). In particular, computational hemodynamics requires an anatomically accurate vasculature model with the following characteristics: 1) water-tight, 2) includes all essential arteries, and 3) exhibits correct vessel connections. Typically, the state-of-the-art vascular segmentation methods (Xiang et al., 2016) require substantial manual interactions and are not ideal in the clinical workflow due to their high resource cost, inaccuracy, and subjectivity. More recently, automatic segmentation models (Ronneberger et al., 2015) based on deep learning have emerged as a promising alternative to traditional model-based manual segmentation methods. Despite their superior performance (Yang et al., 2021), these deep learning models have largely not been tested for computational hemodynamics applications. Our initial experiments identified three severe drawbacks of some deep-learning image segmentation methods, as demonstrated in a showcase example in Fig. 1. First, patch-based deep-learning segmentation networks operate on many small divided patches with a limited field of view (FOV), resulting in global and contextual information loss. This shortcoming may lead to an incomplete aneurysm structure (see the yellow arrow in Fig. 1(c)) and the omission of smaller vessels (see the white arrows in Figs. 1(c) and 1(d)). Second, feature maps of different (spatial) resolutions/scales generated by the encoder-decoder network have large semantic gaps. Hence, prediction derived mostly from the deepest feature layer with weak semantic information will inevitably result in a loss of details. As a result, a complex IA sac may not be

accurately extracted, and small vessels are often omitted, as shown in Fig. 1(c) (see the two marked square areas indicating a dented IA sac and a missing small artery, respectively). Third, recall that adjacent pixels share considerable similarities and very tight spacing between neighboring vessels (see the square area marked in Fig. 1(a)). When convolutions are applied to a receptive region belonging to different vessels but containing similar intensity values, spatial context post-convolutions may further amplify the neighborhood dependencies, causing unwanted adhesions/connections between neighboring vessels (see the blue arrows in Figs. 1(c) and 1(d)). Consequently, this problem is intrinsic to many convolution-based segmentation methods. Our study is the first to explore mitigation of this fundamental limitation in applications tailored for computational hemodynamics.

During computational hemodynamic modeling of an IA, proximal and distal vessels connected to the IA and the IA itself must be accurately extracted to reproduce physiological blood flow patterns in and around the IA. Referring back to Fig. 1, if a small artery around an IA is not included or an unwanted vessel-to-vessel connection is made, the gross hemodynamic pattern in and around the IA will not reflect the actual physiological flow. In another scenario, when a physiological flow rate (e.g., 300 ml/min for internal carotid artery [ICA]) is prescribed to the segmented ICA as a velocity boundary condition, how accurately the ICA cross-sectional area is determined impacts the outcome of hemodynamic modeling.

In this paper, to address each of the above-said challenges, we proposed an *attention-based residual U-Net* (ARU-Net) framework with differential preprocessing and geometric postprocessing for automatic segmentation of IAs and their adjacent arteries. Our adoption of the "attention mechanism" is inspired by the human visual attention mechanism (Kastner and Ungerleider, 2000) because human beings often selectively concentrate their limited processing resources only on what is important. Studies have demonstrated that computational modeling of visual attention can quickly locate the most salient objects in a cluttered visual scene (Itti and Koch, 2001), thereby improving the effectiveness of visual processing. Thus, in this study, we aimed to equip the network with such "attention" properties to automatically learn in the presence of clutter by focusing on the object structures and ignoring the irrelevant parts without additional supervision.

Additionally, benefiting from the residual learning structure (He et al., 2016), which uses shortcut connection and element-wise addition

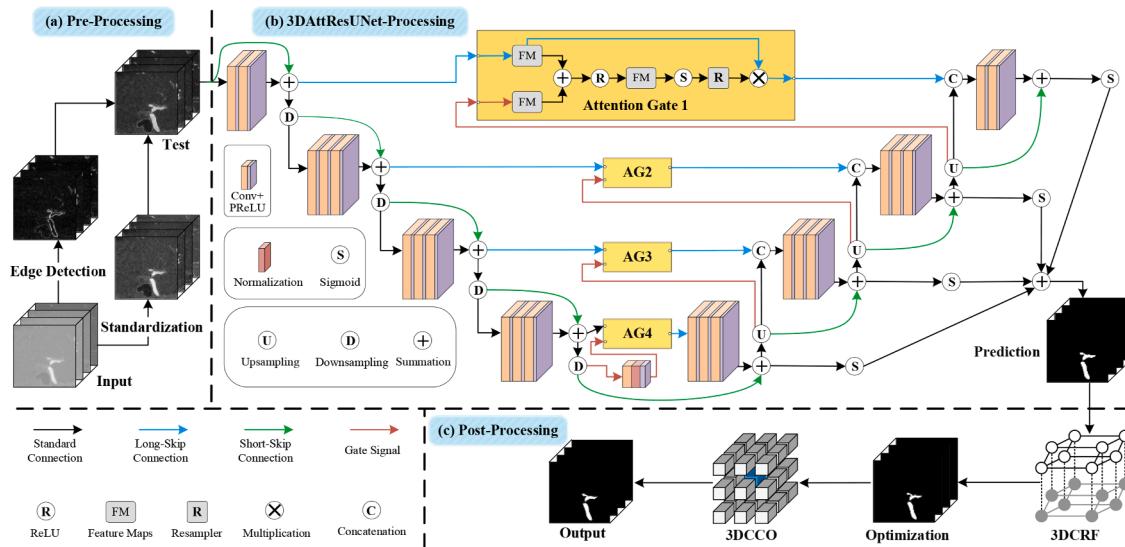


Fig. 2. An overview of the proposed 3D image segmentation process. In the testing phase, (a) preprocessing (for boundary detection and normalization) of an input image is done first. Then, the preprocessed image data are fed-forward into the (b) 3D AttResUNet and are refined using (c) postprocessing strategies by executing 3D conditional random field (3DCRF) and 3D connected component optimization (3DCCO) in a sequence.

to perform residual mapping to solve the degradation problem, we utilize long and short skip connections to preserve richer spatial information by adopting this "residual" strategy.

A graphic overview of the proposed model can be found in Fig. 2. We first performed boundary enhancement (edge detection and standardization) as preprocessing (see Fig. 2(a)) on the original *3D rotational angiography* (3DRA) images to improve (imaging) contrast around the vessel contours, thereby preserving small (~1 mm diameter or 2~3 pixel-sized) vessel information to propagate downstream into the deep-learning network. Next, we optimized the feature maps generated in the encoding paths by exploiting the depth feature maps in the decoding paths (see Fig. 2(b)); each path contains four pyramid layers of different resolutions for encoding and decoding. Specifically, we used the gate signals of multiple attention modules to target IAs and their adjacent small vessels. Thus, the proposed ARU-Net's learning focused on the targeted IAs and their proximity. It is also worth noting that, for efficient training, residual connections were leveraged to avoid exponential gradient decay in deep networks while enabling the model to aggregate semantic information at different scales. Then, we designed the multiscale features produced at different levels in the decoding path as feature pyramids and made predictions independently at each pyramid level to integrate the multiscale semantic information to facilitate the segmentation of small vessels. Lastly, we exploited the *3D conditional random field* (3DCRF) and *3D connected component optimization* (3DCCO) as postprocessing (see Fig. 2(c)) to remove the unwanted connections between neighboring arteries and optimize the segmentation components.

Extensive experiments (including comparative experiments with established deep-learning (Çiçek et al., 2016; Isensee et al., 2021; Kamnitsas et al., 2017; Bhalerao and Thakur, 2019; Müller and Kramer, 2021; J.M.J. Valanarasu et al., 2021) and traditional image segmentation methods (Piccinelli et al., 2009; Jirik et al., 2013), ablation experiments, and CFD experiments) confirmed that the proposed ARU-Net model quantitatively and qualitatively outperforms other methods in terms of aneurysm accuracy and vascular integrity. In short, the contributions and innovations of this work are as follows:

- 1) We proposed a deep encoder-decoder architecture with preprocessing and postprocessing for IA segmentation.
- 2) We constructed the multiscale supervision-based ARU-Net by adopting the depth-aware attention gate module to guide the deep

network to learn more accurate IA structures and improve the processing of features concerning small vessels.

- 3) We exploited the superiority of 3DCRF to eliminate the unwanted adhesions/connections between adjacent arteries and an IA and its arteries in the segmentation results.
- 4) We conducted comprehensive experiments and performed "patient-specific" CFD modeling to demonstrate the superior performance of the proposed ARU-Net model. It is worth noting that the applications of deep-learning-based image segmentation for "patient-specific" CFD simulations have rarely been reported.

2. Related works

The rupture of IAs can lead to high mortality. Clinical management of IAs has been focused on (1) treating ruptured IAs to stop bleeding immediately and (2) selecting some unruptured IAs with a high risk of rupture to avoid the worst outcomes. As discussed in Section 1 above, accurate segmentation of IAs is a critical step in supporting the clinical decision for efficiently managing patients with IAs. Given the high prevalence of (unruptured) IAs in the adult population, this topic is of great significance and impacts our society.

Generally speaking, the methods for IA segmentation are mainly divided into three categories: traditional image segmentation methods (Sen et al., 2014), machine learning-based methods (Flanders, 2019), and deep learning-based methods. Traditional methods rely heavily on prior knowledge, and machine learning methods often require complex feature engineering, all of which tend to be inferior to deep learning methods in terms of capability and accuracy.

Due to the excellent feature learning ability, deep learning technology has been explored for IA segmentation. Nakao et al. (Nakao et al., 2018) exploited a *computer-assisted detection* (CAD) system for IA detection on *magnetic resonance angiography* (MRA) images, which was based on the *deep convolutional neural network* (CNN) and the *maximum intensity projection* (MIP) method. Their model had satisfactory accuracy in detecting aneurysms with larger diameters but was less sensitive for relatively small vascular aneurysms. Stember et al. (Stember et al., 2019) developed a CNN to detect IAs based on the MRA automatically. The limitation of their work is that the training was performed on 2D MIP images, and the loss of spatial information resulted in an incomplete representation of the vasculature. Jin et al. (Jin et al., 2020) proposed an end-to-end spatiotemporal deep neural network for IA segmentation,

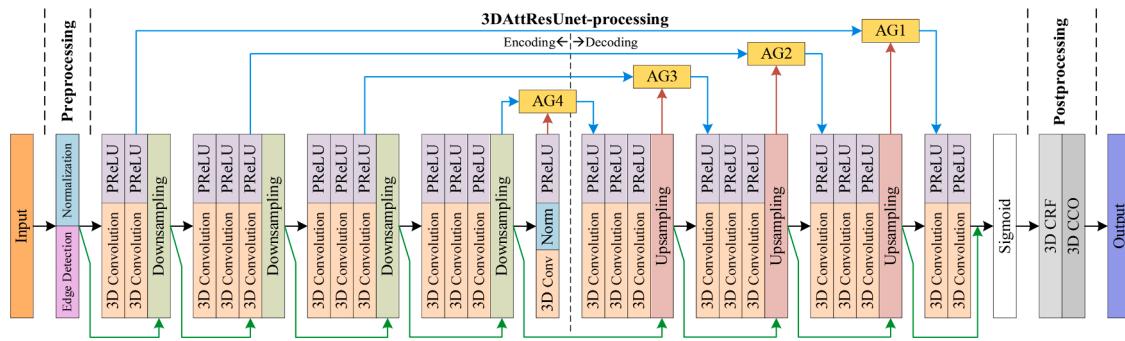


Fig. 3. A detailed processing flowchart of the proposed ARU-Net model.

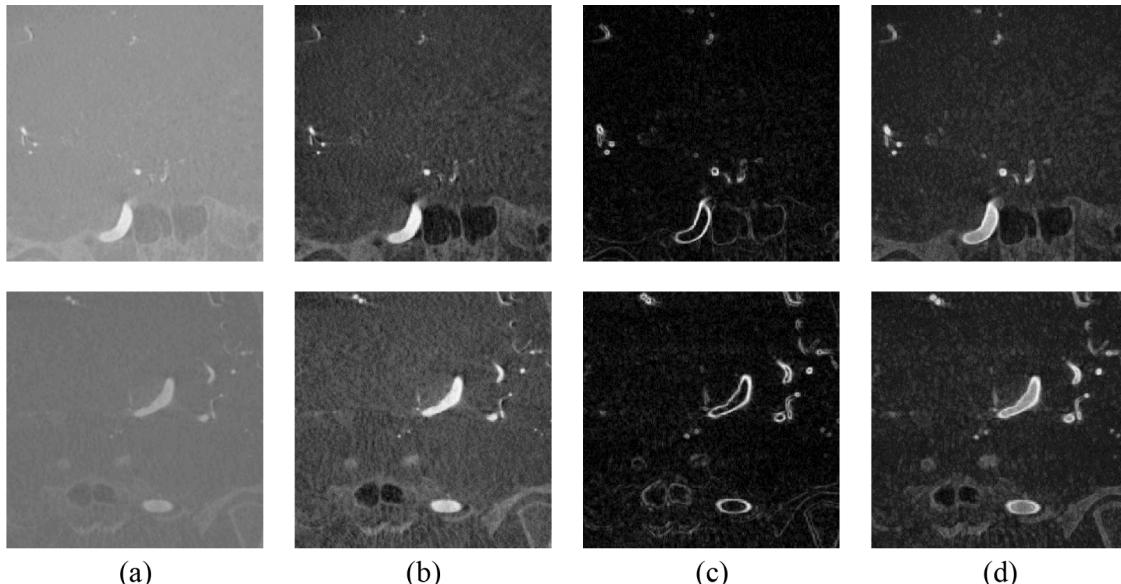


Fig. 4. Representative examples of the results by performing preprocessing on 3D NIFTI images: (a) Original 3D NIFTI images, (b) post-standardization, (c) post-boundary enhancement, (d) standardized + boundary enhanced images.

which incorporated the 2D and time *digital subtraction angiography* (DSA) information. Their model considered only the segmentation of 2D images, making it difficult to distinguish the small aneurysms from the vascular accumulations. Patel et al. (Patel et al., 2020) implemented a DeepMedic (Kamnitsas et al., 2017) based architecture for IA segmentation from DSA sequences and compared its performance with the 3D U-Net (Çiçek et al., 2016) model. Their models are cumbersome and need to be retrained to make predictions if the resolution of input images changes. Shahzad et al. (Shahzad et al., 2020) employed a deep learning model to segment aneurysms on *computed tomography angiography* (CTA) automatically. The disadvantage of their model was that the performance on CTA images was obtained from one clinical scanner and the generalization of their method was not assessed. Cheng et al. (Cheng et al., 2021) performed IA segmentation by exploring the U-shaped deep neural network structure to fuse spatial information together from 3DRA images. The shortcoming of their method is the poor segmentation accuracy for small-sized IAs (e.g., < 3 mm in height or width). Shao et al. (Shao et al., 2022) presented an unsupervised IA detection method based on 3D point cloud data. Their model has advantages on limited labeled data but is not robust to other data types.

It is important to note that none of the methods mentioned above were evaluated for computational hemodynamics applications with IAs. In general, most of the existing deep learning-based IA segmentation methods face the problems of insufficient utilization of 3D spatial information and suboptimal detection of small targets. Moreover, another

problem identified is the unwanted adhesions/connections between two adjacent vessels or the IA to vessel.

To this end, the primary objective of this paper is to address these challenges (i.e., identification of small vessels, avoiding unwanted vessel/IA connections, accurate representation of irregularly shaped IAs) by introducing the proposed innovative deep-learning variant ARU-Net. Strategically, we introduced depth-aware attention gates and multiscale supervision to fully capture and fuse spatial feature information. Theoretically, the idea of the attention mechanism is to generate a context dependency, allowing more adaptive weight assignments to input features. In other words, it automatically highlights salient regions conditioned on the current task while suppressing irrelevant counterparts. As a result, our deep-learning algorithms can make predictions more contextualized. Traditional hard attention (Mnih et al., 2014) typically uses iterative local region proposal-based techniques: its sampling is often non-differentiable, and the parameter update relies on reinforcement learning. Thus, hard attention makes the training of models difficult. To enable long-distance interactions, global attention mechanisms (e.g., augmented attention (Bello et al., 2019) and dual attention (Fu et al., 2019)) have been designed to capture long-range dependencies. But they are more computationally expensive and require more memory storage; e.g., adding augmented attention (Bello et al., 2019) to ResNet50 resulted in an extra 24.3 million parameters, and the inference and training time increased by 29% and 25%, respectively. More recently, the self-attention-based transformer

Table 1

Specification of the proposed 3D ARU-Net architecture.

Path	Operation Layer		Kernel Size	Stride	Padding	Output Size
Encoder	Input image		-	-	-	$1 \times 48 \times 256 \times 256$
	Convolution Layer (Two Times)	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$16 \times 48 \times 256 \times 256$ $16 \times 48 \times 256 \times 256$
	Downsampling Layer	3D Convolution PReLU	$2 \times 2 \times 2$ -	2 -	- -	$32 \times 24 \times 128 \times 128$ $32 \times 24 \times 128 \times 128$
	Convolution Layer (Three Times)	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$32 \times 24 \times 128 \times 128$ $32 \times 24 \times 128 \times 128$
	Downsampling Layer	3D Convolution PReLU	$2 \times 2 \times 2$ -	2 -	- -	$64 \times 12 \times 64 \times 64$ $64 \times 12 \times 64 \times 64$
	Convolution Layer (Three Times)	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$64 \times 12 \times 64 \times 64$ $64 \times 12 \times 64 \times 64$
	Downsampling Layer	3D Convolution PReLU	$2 \times 2 \times 2$ -	2 -	- -	$128 \times 6 \times 32 \times 32$ $128 \times 6 \times 32 \times 32$
	Convolution Layer (Three Times)	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$128 \times 6 \times 32 \times 32$ $128 \times 6 \times 32 \times 32$
	Downsampling Layer	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$256 \times 6 \times 32 \times 32$ $256 \times 6 \times 32 \times 32$
	Attention Layer 4		-	-	-	$128 \times 6 \times 32 \times 32$
Decoder	Attention Layer 3		-	-	-	$64 \times 12 \times 64 \times 64$
	Attention Layer 2		-	-	-	$32 \times 24 \times 128 \times 128$
	Attention Layer 1		-	-	-	$16 \times 48 \times 256 \times 256$
	Convolution Layer (Three Times)	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$256 \times 6 \times 32 \times 32$ $256 \times 6 \times 32 \times 32$
	Upsampling Layer	3D Transposed Convolution PReLU	$2 \times 2 \times 2$ -	2 -	- -	$128 \times 12 \times 64 \times 64$ $128 \times 12 \times 64 \times 64$
	Convolution Layer (Three Times)	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$128 \times 12 \times 64 \times 64$ $128 \times 12 \times 64 \times 64$
	Upsampling Layer	3D Transposed Convolution PReLU	$2 \times 2 \times 2$ -	2 -	- -	$64 \times 24 \times 128 \times 128$ $64 \times 24 \times 128 \times 128$
	Convolution Layer (Three Times)	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$64 \times 24 \times 128 \times 128$ $64 \times 24 \times 128 \times 128$
	Upsampling Layer	3D Transposed Convolution PReLU	$2 \times 2 \times 2$ -	2 -	- -	$32 \times 48 \times 256 \times 256$ $32 \times 48 \times 256 \times 256$
	Convolution Layer (Two Times)	3D Convolution PReLU	$3 \times 3 \times 3$ -	1 -	1 -	$32 \times 48 \times 256 \times 256$ $32 \times 48 \times 256 \times 256$
	Output Layer	3D Convolution Sigmoid	$1 \times 1 \times 1$ -	1 -	- -	$32 \times 48 \times 256 \times 256$ $1 \times 48 \times 256 \times 256$

architectures (e.g., vision transformer (Dosovitskiy et al., 2021), swin transformer (Liu et al., 2021), and medical transformer (J.M.J. Valanarasu et al., 2021)) have been proposed to identify the complex dependencies between elements of each input feature. The known limitations of these self-attention-based transformers are their inability to capture local knowledge, difficulty in characterizing position information, vanishing top layer gradients, etc.

In contrast, the proposed depth-aware attention gate mechanism is a new idea. Our depth-aware attention gates first employ the depth feature in the decoding path as a gating signal to optimize the feature generated in the encoding path. Then, the attention map can be further exploited to guide the decoding of features in later layers. This process ensures that the target is highlighted, and at the same time, the background is suppressed during information propagation. The advantages of this process are three-fold: 1) grid-based gating allows attention coefficients to focus on specific local areas, preserving small vessel information; 2) dense label prediction maintains a large amount of detailed knowledge and location information; and 3) computational overhead is low because of our simple network structure. As a result, the IA morphological differences and small vessel branches can be accurately depicted for subsequent clinical decision-making. At the same time, the unwanted adhesion/connections can be effectively eliminated by the proposed 3DCRF postprocessing.

3. The proposed segmentation model

The proposed 3D image segmentation pipeline is illustrated in Fig. 2 as a three-step process: preprocessing, 3DAttResUnet-processing, and postprocessing. For each image in the 3D *Neuroimaging Informatics Technology Initiative* (NIfTI) format, boundary detection and

standardization are utilized as preprocessing to enhance (image) contrast of vascular contours and reduce the background noise. Then, the proposed 3D ARU-Net with multiscale prediction is initially utilized to segment the 3D NIfTI image. Finally, the fully connected 3DCRF and 3DCCO are exploited as the postprocessing step to refine the segmentation outcomes. The detailed workflow for the proposed three-step processing is presented in Fig. 3.

3.1. Preprocessing

3DRA data are in the Hounsfield scale, and the (image) dynamic range is not suitable for image segmentation. We first standardized the original grayscale values of each image in a 3D NIfTI sequence and compressed the image dynamic range to 11 bits (i.e., 0 to 511). The standardization often allows us to better visualize the vasculature of interest while suppressing the image noise. To preserve the information of small vessels during the forward and backward propagation of network training, we also utilized the Sobel filter, an edge detector, to enhance the contrast of the vessel boundaries. The final preprocessed image is a weighted summation of the normalized and edge-enhanced images. Fig. 4 shows the effectiveness of the above-said preprocessing in a representative example.

3.2. 3DAttResUNet-Processing

The proposed 3D ARU-Net architecture aims to segment the IAs and the associated parent vascular structures accurately. The proposed network architecture allows us to better depict small vessels within the proximity of IAs. As shown in Fig. 3, similar to the standard U-Net (Ronneberger et al., 2015) segmentation network, our network consists

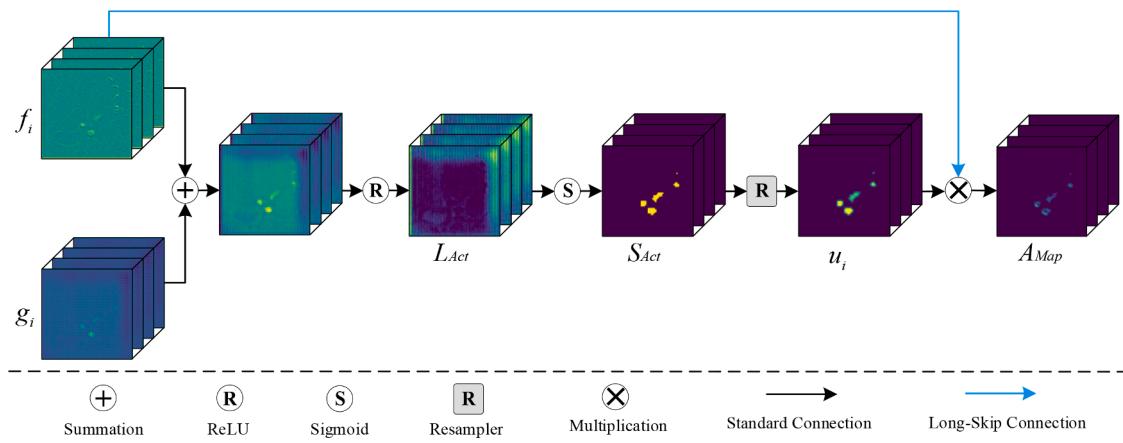


Fig. 5. A visualization schematic diagram showing the structure of the attention gate.

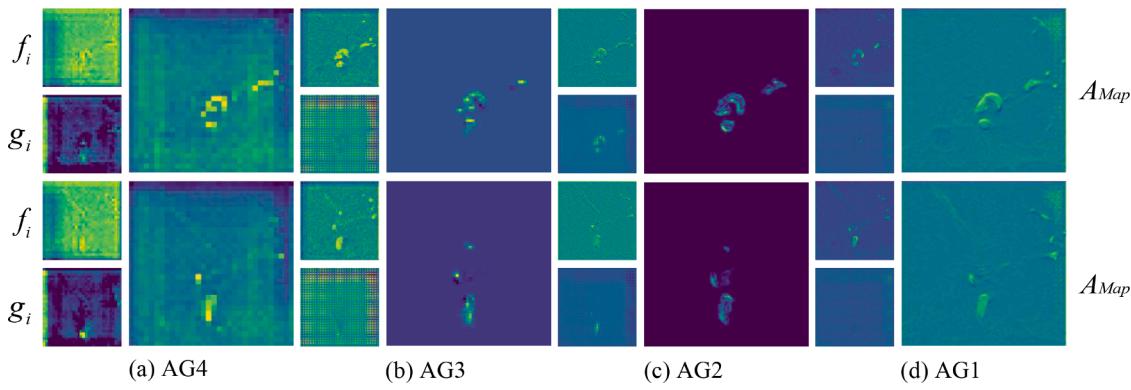


Fig. 6. Resulting examples $AMap$ obtained by performing AGs on the feature maps f_i and the gate signals g_i at four different spatial scales.

of encoding and decoding paths; each path has four Pyramid layers of different resolutions, with independent predictions made from all layers. Each layer contains two or three 3D convolutions in both decoding and encoding paths, each followed by a *parametric rectified linear unit* (PReLU) (He et al., 2015). Mathematically, $PReLU(x) = \max(0, x) + a \times \min(0, x)$, which can improve the model accuracy at a negligible extra computational cost. The proposed 3D ARU-Net was trained by using cropped patches of size $48 \times 256 \times 256$ with a sliding window running over the input 3D NIfTI image of size $256 \times 256 \times 256$ to reduce memory requirements. A detailed description of our 3D ARU-Net can be found in Table 1.

To improve convergence, our network not only utilizes the long skip connections of U-Net to propagate the discriminative information of layers of equal resolution from the encoding path to the decoding path but also incorporates the short skip connections of ResNet (He et al., 2016) in each layer to ensure the provision of essential high-resolution feature information for the following layer. This configuration can significantly mitigate the vanishing gradient problem during the back-propagation and effectively aggregate the semantic information of different feature scales.

Depth-Aware Attention Gate Module: Generally speaking, as the number of layers increases in an encoding-decoding structure, the network gradually has a richer representation of possible features at different spatial scales and orientations. However, the segmentation of small objects requires special considerations for two reasons. On the one hand, due to cascaded convolutions, spatial details are lost at these high-level (i.e., coarse spatial resolution) feature maps, resulting in false-negative detection of small objects. On the other hand, reducing false-positive detections for small objects exhibiting large shape variability is difficult at the coarse resolution (Oktay et al., 2018). *Attention gates*

(AGs) are adopted and incorporated into long skip connections to preserve features relevant to small vessels. The AGs enable the proposed ARU-Net model to emphasize relevant spatial information from feature maps at multiple scales and subsequently propagate it to the decoding path, as illustrated in Fig. 5. Specifically, the depth feature maps in the decoding path are used as the gate signals (denoted g_i at each pixel i to determine the focus region) to optimize the feature maps (defined as f_i) generated during the coding process. This process suppresses the feature responses of the irrelevant background regions (i.e., through selective activation, as shown in Fig. 5). The gate signal of AG at the network layer with the lowest resolution is obtained by performing convolution, normalization, and activation operations on the feature map output by the encoder. It is worth noting that a lattice-like structure (see g_i in Fig. 5) will appear at the edge of the feature map during the upsampling process of the decoder (also known as "checkerboard artifacts"). However, the activation operation in our AG module can largely eliminate these background artifacts (see $SAct$ in Fig. 5). After implementing the AG module in all four pyramid layers, the lattice-like structure gradually weakened until it fully disappeared. It thus effectively highlights the valuable feature critical for the IA segmentation task.

By linearly mapping the concatenated features f_i and g_i to an intermediate space, the intermediate activation maps (denoted as $LAct$ and $SAct$, respectively) are calculated as:

$$LAct = \psi(\sigma_1(w_f f_i + w_g g_i + b_\psi)) + b_\psi, \quad (1)$$

$$SAct = \sigma_2(LAct(f_i, g_i)), \quad (2)$$

where σ_1 and σ_2 represent the ReLU and Sigmoid activation functions, respectively. ψ , w_f , and w_g are the linear transformations, which are computed by using $1 \times 1 \times 1$ convolutions on the channels of the input

tensor. b_g and b_ψ are the bias terms. In Fig. 5, $LAct$ and $SAct$ are two nonlinear layers that selectively activate attention coefficients in a sequence. Then, the attention coefficients $SAct$ are resampled with trilinear interpolation. The final output (denoted as $AMap$) of AG is the multiplication of the feature maps f_i and the upsampled maps (denoted as u_i), as follows:

$$AMap = f_i \times u_i. \quad (3)$$

This study used four AGs to deal with shallow and deep features. The results of AG processing at four different (spatial) scales are shown in Fig. 6. Usually, the attention maps have large and small values in the target vessel and background regions, respectively. Thus, as shown in

$$T(\alpha, \beta) = \frac{\sum_{i=1}^N p_i(x=1)q_i(x=1)}{\sum_{i=1}^N p_i(x=1)q_i(x=1) + \alpha \sum_{i=1}^N p_i(x=1)q_i(x=0) + \beta \sum_{i=1}^N p_i(x=0)q_i(x=1)}, \quad (4)$$

Fig. 6. AG processing can suppress background noise. Collectively, the AG can improve the accuracy of IA segmentation through more "directed" information dissemination at different (spatial) scales.

Multiscale Supervision Strategy: We designed a multiscale supervision strategy to improve the segmentation performance on small vessels and morphological details of irregularly shaped IAs. The proposed 3D ARU-Net first learns the feature information of different scales independently. Then, the semantic feature information of small vessels at low and high levels is fused through a top-down pathway with lateral connections.

Our network configuration is different from the traditional U-Net architecture at the implementation level. Instead of adopting a single high-resolution feature map and making predictions based on that high-resolution map in the U-Net model, the proposed network utilizes the multiscale features generated at different layers in the decoding stage as a feature pyramid. In other words, our network makes predictions at each pyramid level independently. Specifically, the sigmoid function is used to make predictions layer by layer, and the prediction results of all layers are summed together to provide the final segmentation probability. Since the feature maps of different spatial resolutions generated by the network have significant semantic gaps due to different depths, the prediction made by a single high-resolution feature with weak semantics will inevitably lead to the information loss of small vessels. In contrast, the proposed multiscale supervision strategy can synthesize low-resolution features with strong semantics and high-resolution features with weak semantics to combine the coarse-level and fine-level dense predictions, thereby enhancing the representation capacity for small vessel segmentation and creating a more accurate depiction of irregularly shaped IAs.

The proposed multiscale supervision is inspired by the well-known feature pyramid networks (FPNs, e.g., (Lin et al., 2017; Seferbekov et al., 2018; Kong et al., 2018)). FPNs' main advantages include 1) distributing objects of various sizes to different feature layers for accurate detection of multiscale objects and 2) enriching semantic and location information of objects by aggregating information received from different decoding layers. However, drawbacks of the FPN structure are also known: 1) the nearest-neighbor algorithm is used for upsampling and leads to ineffective propagation of high-level semantic information; 2) mapping of high-level features in a top-down pathway often causes the loss of some semantic information; and 3) multiple downsampling and upsampling may cause inaccurate feature/information fusion and propagation. To address these drawbacks, we adopt the following strategies: 1) upsampling based on trilinear interpolation to reduce image quality loss with a slightly high computational cost; 2) using short skip connections in each layer to aggregate the

features before and after convolutions to fully retain semantic information; and 3) securing more accurate information/feature fusion by directly passing attention-based long skip connections from the finest level (bottom-up maps) to the coarsest level (top-down maps).

Tversky Loss Function: To train the proposed 3D ARU-Net model, a Tversky loss function (Salehi et al., 2017) was utilized to address the data imbalance issue in image segmentation. In the presence of small vascular structures, the Tversky loss function acquires a relatively better trade-off between precision and recall. In other words, weighting false negatives (FNs) more than false positives (FPs) in training highly imbalanced data is beneficial to highlight the critical small vessels. Formally, the Tversky loss function reads:

where p_i and q_i denote the set of predicted and ground truth binary labels, respectively, within which the pixel value x is 1 for vascular regions and 0 for non-vascular regions. The hyperparameters α and β tune the penalty magnitudes of FPs and FN, respectively. By placing more emphasis on FN, a larger β gives more weight to recall instead of precision. It is generally understood that a higher β in the generalized loss function during training results in higher generalization and improves performance on imbalanced data. Such a strategy effectively shifts the focus of prediction to lower false-negatives and promotes recall, thus, to a certain extent, avoiding missed detection of small vessels.

To sum up, compared with the traditional U-Net structure, the proposed 3D ARU-Net has three advantages: 1) the residual connections deepen the network to learn 3D features, improving the model's convergence; 2) the depth-aware attention gate module helps the network to focus on more local-perceptual features, thereby enhancing the network's ability to learn the structural information of IAs; and 3) combined with multiscale supervision, this network improves the aggregation of small ROI features in the cascade convolution, thus facilitating the segmentation of small vessels and irregularly-shaped IAs.

3.3. Postprocessing

After the prediction by the proposed 3D ARU-Net model, the initial segmentation map was generated from the test image. Although the generated result has a relatively complete vascular structure, some issues remain. First, since neighboring pixels share a large amount of similar spatial receptive fields, the soft segmentation maps obtained by the network tend to be smooth, as the boundary pixels can be identified as either foreground or background. Hence, modeling neighboring target regions can also lead to unwanted adhesion between adjacent vessels or between an IA and its adjacent vessels because the very tight spacing between vessels and their surroundings (see Fig. 1) can be easily predicted as vessel regions. Second, the proposed model considers the identification of individual pixels only and does not fully exploit the contextual and relevant information of adjacent regions in 2D images and 3D volumes, resulting in incomplete or discontinuous vascular structures (e.g., missing small vessels or inaccurate depiction of IA morphology) due to insufficient knowledge of neighborhood dependencies. Third, due to spatial noise in the input image and local minima in training, some spurious outputs appear, with small holes or isolated regions in the prediction results. We first adopted a fully connected 3D conditional random field (3DCRF) (Krähenbühl and Koltun, 2011) to eradicate the first two issues above-listed. Subsequently, 3D connected component optimization (3DCCO) was another postprocessing step to remove noise over the initial segmentation (i.e., the third issue).

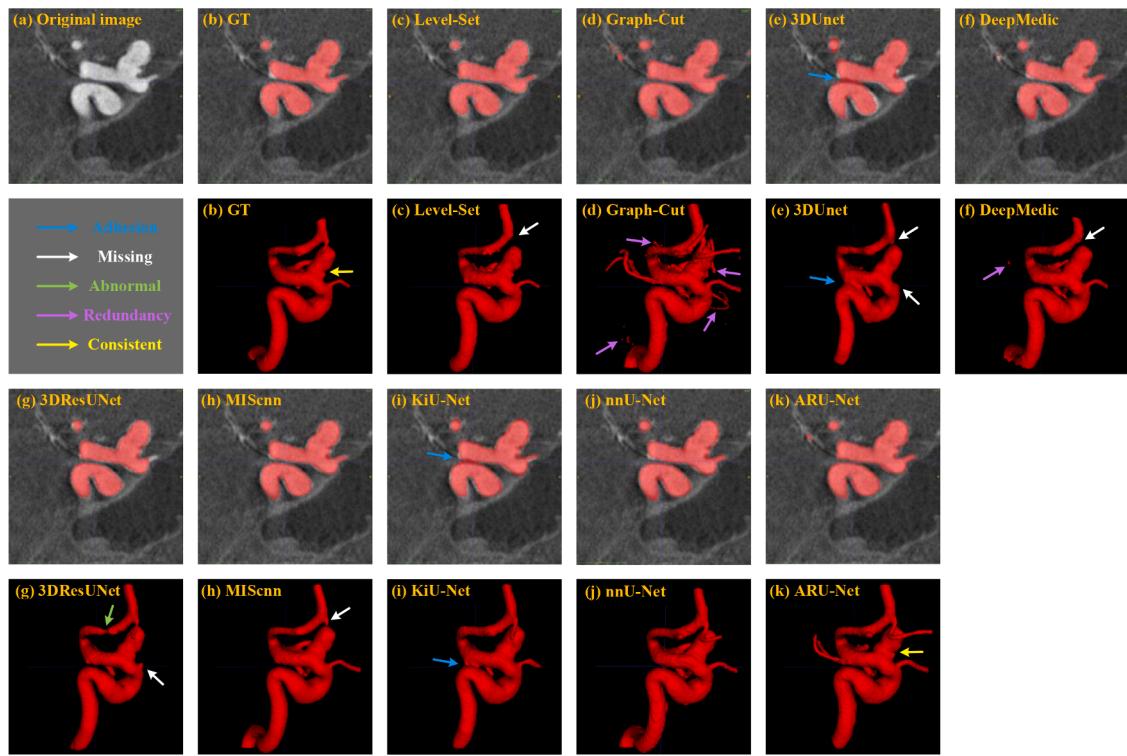


Fig. 7. Visual comparisons of segmentation results generated by different models. (a) Coronal plane of the original 3D NIfTI images. (b) Ground truth (manual annotation). (c-j) Visualization of results from the eight segmentation models. (k) Segmentation results of the proposed ARU-Net model. The coronal and 3D views are provided for each result.

A framework based on 3DCRF was used to regularize the initially segmented binary mask; our goal is to make our final segmentation more spatially coherent. Letting I and p_i denote the 3D input image and predicted segmentation labels, respectively, the Gibbs energy in the 3DCRF model (X, I) can be calculated by:

$$E(X=x|I) = \sum_{i=1}^N \varphi_u(p_i|I) + \sum_{i=1, j=1, i \neq j}^N \varphi_p(p_i, p_j|I), \quad (5)$$

where i and j denote the number of pixels in the random field X and test image I , respectively. The unary potential $\varphi_u(p_i|I) = -\log P(p_i|I)$ is derived from the probability output $P(p_i|I)$ of the proposed 3D ARU-Net. In Eqn. (5), the pairwise potential $\varphi_p(p_i, p_j|I)$ is defined as follows:

$$\varphi_p(p_i, p_j) = \mu(p_i, p_j) \left[\omega_1 \exp \left(-\frac{|s_i - s_j|^2}{2\eta_1^2} \right) + \omega_2 \exp \left(-\frac{|s_i - s_j|^2}{2\eta_2^2} - \frac{|I_i - I_j|^2}{2\eta_3^2} \right) \right], \quad (6)$$

where μ denotes the label compatibility function derived from the Potts model $\mu(p_i, p_j) = [1; p_i \neq p_j]$, which provides the compatibility between different pairs of adjacent pixels with dissimilar labels. ω_1 and ω_2 denote the linear combination weights to tune the pairwise terms. $|s_i - s_j|$ denotes the spatial distance between pixels i and j , while $|I_i - I_j|$ denotes their intensity difference in the input image. The hyperparameters η_1 , η_2 , and η_3 adjust the degree of proximity and similarity.

The optimal/final labels $x^* = \operatorname{argmin} E(X=x|I)$ of the segmentation map can be obtained by minimizing Eqn. (5) using a mean-field approximation algorithm (Krähenbühl and Koltun, 2011). By encouraging spatial consistency to optimize the pixel labels in adjacent spaces, 3DCRF can eliminate adhesion between neighboring vessels or between an IA and its parent artery to a certain extent.

A typical problem after 3DCRF processing is the presence of some

extraneous disconnected small vessel regions. Formally, we name the segmentation result as post-3DCRF which is for processing a binary 3D volume in which vessel labels are ones and the rest are zeros. Starting from an arbitrary vessel label (hereafter referred to as the selected first vessel label) on the binary 3D volume, 3DCCO recursively searched all neighboring vessel labels and added them to a group associated with the selected first vessel label. This process is also known as region-growing and would not stop until all vessel labels connected to the selected first vessel label were found. If there were still vessel labels with no group association, another arbitrary vessel label would be initialized, and the region-growing process would start again. After all vessel labels had been processed, this process resulted in several connected vessel groups. The group with the largest number of vessel labels would be selected as the final segmentation result. Finally, morphological operations (e.g., dilation and erosion) were also used to eliminate holes in the vascular area.

4. Experimental evaluations and discussion

In this section, quantitative and qualitative experimental results were provided to verify the effectiveness of the proposed model, including experimental setup, comparative experiments, ablation experiments, and CFD experiments.

4.1. Experimental setup

Experimental Dataset: A dataset of 23 3DRA images from different patients with one IA was used. Data were extracted from a publicly available database hosted at Emory University (<http://ecm2.mathcs.emory.edu/aneuriskweb/about>). Images were manually annotated and verified with clinician annotated geometries available in the above-said public database to provide the training and testing samples. Each 3DRA image contains 256 slices, each of which has a dimension of 256 × 256. In our experiments, 65% 3DRA images (15 cases) of the dataset were

Table 2

Quantitative performance comparisons (mean \pm 95% confidence) of different segmentation models using the six metrics. The up arrow \uparrow indicates that the larger the value, the better the performance. The down arrow \downarrow means the opposite. The best results are highlighted in **bold**.

	DSC \uparrow	RVE \downarrow	SE \uparrow	SP \uparrow	HD95 (mm) \downarrow	ASSD (mm) \downarrow
Level-Sets	0.836 ± 0.034	0.175 ± 0.069	0.787 ± 0.070	0.998 ± 0.001	24.950 ± 11.216	2.599 ± 0.779
Graph-Cuts	0.830 ± 0.028	0.150 ± 0.093	0.846 ± 0.072	0.996 ± 0.002	43.173 ± 11.232	4.497 ± 0.796
3DUNet	0.756 ± 0.065	0.259 ± 0.126	0.678 ± 0.106	0.998 ± 0.001	46.650 ± 17.126	6.409 ± 3.150
DeepMedic	0.888 ± 0.028	0.128 ± 0.077	0.895 ± 0.059	0.998 ± 0.001	15.841 ± 7.502	1.692 ± 0.509
3DResUNet	0.800 ± 0.085	0.235 ± 0.132	0.716 ± 0.111	0.999 ± 0.001	47.751 ± 18.308	5.093 ± 2.332
MIScnn	0.877 ± 0.041	0.164 ± 0.103	0.903 ± 0.077	0.997 ± 0.002	29.314 ± 16.810	3.367 ± 2.591
KiU-Net	0.796 ± 0.064	0.187 ± 0.118	0.772 ± 0.112	0.997 ± 0.002	43.766 ± 13.857	4.943 ± 2.029
nnU-Net	0.866 ± 0.038	0.186 ± 0.121	0.912 ± 0.073	0.996 ± 0.002	29.408 ± 15.720	3.316 ± 1.832
ARU-Net	0.868 ± 0.038	0.118 ± 0.076	0.853 ± 0.066	0.998 ± 0.001	27.599 ± 16.998	2.696 ± 1.502

selected for training, and the remaining 35% (8 cases) were exploited for testing. To fully validate the comprehensive performance of the proposed segmentation model, both the training and testing cases covered a range of complex situations, e.g., aneurysms of minor secondary vessels, non-uniformly mixed contrast, tortuous vasculature, and narrow spacing between the aneurysm and adjacent parent artery.

Implementation Details: The proposed ARU-Net segmentation model was implemented under the PyTorch framework and tested on a computer workstation with an Intel Core i9-10900k CPU (3.7 GHz; 128 GB RAM) and an NVIDIA GeForce RTX3090 GPU (24GB GPU RAM). The initial learning rate was set to 1×10^{-4} and was decreased with a supervised decay coefficient of 0.33. An Adam optimizer was selected to train the proposed model with a batch size of 3 due to the limitation of CUDA memory. The training took about six hours to complete 2000 epochs from a random initialization of weights. The total number of the network parameters was approximately 9.59 M. After training, the average time for the proposed model to generate a 3D segmentation map from each $256 \times 256 \times 256$ tested NIfTI image was about 3.277 s.

Comparative Models: The proposed ARU-Net framework was compared with two state-of-the-art traditional segmentation models and six deep learning-based models, including *Vascular Modelling ToolKit* (VMTK) Level-Sets (Piccinelli et al., 2009), Graph-Cuts (Jirik et al., 2013), 3DUNet (Çiçek et al., 2016), DeepMedic (Kamnitsas et al., 2017), 3DResUNet (Bhalerao and Thakur, 2019), MIScnn (Müller and Kramer, 2021), KiU-Net (J.M.J. Valanarasu et al., 2021), and nnU-Net (Isensee et al., 2021). To be fair, all these models were trained and tested on the same data with the same environment configuration (i.e., identical computer workstation and operating system).

Evaluation Metrics: To evaluate the performance of the proposed ARU-Net model against the other eight models, six segmentation metrics were exploited, including *dice similarity coefficient* (DSC), *relative volume error* (RVE), *sensitivity* (SE), *specificity* (SP), the *95th percentile Hausdorff distance* (HD95), and *average symmetric surface distance* (ASSD). In particular, DSC measures the similarity between the predicted segmentation map and reference ground truth; RVE is used to quantify the "water balance" error and to indicate whether a model is poorly representative; SE mainly measures the missed detection rate; SP measures the false detection rate; HD95 is a generalized assessment of distance between two point sets, quantifying the maximum segmentation error; ASSD determines the average difference between the surface of the predicted map and ground truth. Letting TP , TN , FP , and FN denote the rate of true-positive, true-negative, false-positive, and false-negative,

respectively, the above evaluation metrics are defined by:

$$DSC = 2TP / (2TP + FP + FN), \quad (7)$$

$$RVE = |FN - FP| / (TP + FN), \quad (8)$$

$$SE = TP / (TP + FN), \quad (9)$$

$$SP = TN / (TN + FP), \quad (10)$$

$$HD95 = p_{95\%}(\min_{y \in G} \|x - y\|^2 \cup \min_{x \in P} \|x - y\|^2), \quad (11)$$

$$ASSD = (\sum_{x \in P} \min_{y \in G} \|x - y\|^2) / 2|P| + (\sum_{y \in G} \min_{x \in P} \|y - x\|^2) / 2|G|, \quad (12)$$

where P and G denote the predicted map and the ground truth, respectively. TP and TN are calculated as the sets of correctly classified IA voxels and non-IA voxels, respectively; FP and FN are the sets of falsely detected IA voxels and non-IA voxels, respectively. Both HD95 and ASSD are given in mm, and a lower value means better segmentation.

4.2. Comparative experiments

Qualitative Results: To show the subjective performance comparisons between the segmentation results and the ground truths, the visualization of all cases was performed. One representative example is presented in Fig. 7. Thanks to the depth-aware attention gate and multiscale supervision configuration, the proposed ARU-Net model not only obtained a highly accurate IA region but also successfully detected more small vessels. As shown in Fig. 7, in contrast to our model, almost all the comparative models have missed the small vessels, as indicated by the white arrows. In addition, the results of 3DUNet and KiU-Net models showed severe adhesions of adjacent vessels, as indicated by the blue arrows. The results of the 3DResUNet model showed a sudden decrease in vessel diameter (see Fig. 7(g)), as indicated by the green arrows. Furthermore, the results of Graph-Cuts and DeepMedic models had some cluttered false positive areas, as indicated by the purple arrows in Figs. 7(d) and 7(f). Compared with GT, our model has two main advantages; one is the accurate delineation of small vessels that we omitted in the annotation stage (see the ground truth in (b)), and the other is the highly consistent IA shape. One small "wrinkle" on the IA dome is clearly visible (indicated by the yellow arrows in Figs. 7(b) and 7(k)).

Quantitative Results: The segmentation performance of various models in terms of the six metrics on the intracranial aneurysm testing dataset are listed in Table 2. As can be seen in Table 2, the DeepMedic model yields the best performance in the DSC, HD95, and ASSD metrics, while our model achieves only the third-best on these metrics. In addition, our model also performs the fourth-best on the sensitivity metric, with a small difference (0.059) from the best result of the nnU-Net model. In terms of specificity, the 3DResUNet model achieves the highest score 0.999, while the other models are only slightly lower (less than 0.003) than it. Traditional methods, Level-Sets and Graph-Cuts, can outperform partial deep models in some metrics, but these models rely too much on manual intervention. The Level-Sets model requires the selection of two specified thresholds as the initial level sets and requires placing two seeds on the image. The Graph-Cuts model needs a human user to determine the foreground and background seeds. Note that the proposed ARU-Net model obtains the best result on the RVE score. The difference between our predicted segmentation map and the ground truth is relatively minimal compared with other models.

It is interesting to state that the proposed ARU-Net model is capable of segmenting cases with small vessels. However, please note that the performance of the proposed ARU-Net model is not the best on the traditional metrics listed in Table 2 partially because our predictions contain a large number of small vessels; one example is shown in Fig. 7

Table 3

Objective performance comparisons (mean \pm 95% confidence) using different loss functions of the proposed ARU-Net. The dash (-) means that the segmentation map cannot be obtained to calculate the metric.

	DSC↑	RVE↓	SE↑	SP↑	HD95(mm)↓	ASSD(mm)↓
BCE/Hybrid Loss	-	-	-	-	-	-
ELDice Loss	0.277 \pm 0.160	8.047 \pm 14.214	0.303 \pm 0.222	0.875 \pm 0.245	104.191 \pm 21.392	35.766 \pm 10.081
Dice Loss	0.716 \pm 0.123	0.375 \pm 0.153	0.598 \pm 0.138	0.999\pm0.000	49.320 \pm 19.807	8.906 \pm 6.757
Jaccard Loss	0.801 \pm 0.065	0.217 \pm 0.117	0.741 \pm 0.108	0.998 \pm 0.001	49.367 \pm 17.039	5.882 \pm 2.522
SS Loss	0.832 \pm 0.045	0.141 \pm 0.091	0.815 \pm 0.080	0.997 \pm 0.002	38.462 \pm 17.371	4.114 \pm 1.927
Tversky Loss	0.850\pm0.039	0.105\pm0.091	0.853\pm0.059	0.997 \pm 0.002	33.779\pm16.033	3.541\pm1.499

Table 4

Objective performance comparisons using different variants of the proposed ARU-Net model. The best results are highlighted in **bold**.

	DSC↑	RVE↓	SE↑	SP↑	HD95(mm)↓	ASSD(mm)↓
(1) Backbone without MS	0.5913	0.3139	0.5756	0.9988	153.2808	26.8980
(2) Backbone	0.6233	0.8843	0.8465	0.9961	163.7476	38.3391
(3) Backbone+Pre	0.6673	0.6690	0.8483	0.9970	158.2298	32.2356
(4) Backbone+Att	0.7371	0.4791	0.8718	0.9978	135.3749	24.5004
(5) Backbone+Pre+Att	0.7514	0.3962	0.8715	0.9981	133.9563	19.1065
(6) Backbone+Pre+Att+CCO	0.7892	0.2714	0.8623	0.9987	61.0244	7.4198
(7) Backbone+Pre+Att+CRF	0.8265	0.1485	0.8592	0.9991	57.0811	6.856
(8) Backbone+Pre+Att+Post (ARU-Net)	0.8681	0.1184	0.8533	0.9978	27.5992	2.6960

Table 5

Quantitative performance comparison of our multiscale supervision with traditional FPNs.

	DSC↑	RVE↓	SE↑	SP↑	HD95 (mm)↓	ASSD (mm)↓
FPN	0.7650	0.4723	0.9249	0.9902	43.3859	4.8937
ResFPN	0.7638	0.4255	0.9046	0.9908	40.2781	4.7031
HPU- Net	0.7785	0.3455	0.8793	0.9924	28.3825	4.2050
ARU- Net	0.8681	0.1184	0.8533	0.9978	27.5992	2.6960

(k). The rationale is explained as follows: due to the requirement of CFD modeling, we omitted some small vessels in the manual labeling process. As a result, the small vessels we predicted should be true positives but become false negatives. Since the ground truths we used for training and testing did not contain these small vessels, our prediction results include a large number of small vessels, which increases the gap with ground truth. That explains why our DSC and other criteria scores are slightly lower than those of other models. Furthermore, we attest that the traditional metrics shown in Table 2 may not be appropriate for assessing how well CFD models can be created. It is important to note that CFD simulations of the predicted vascular structures yield important biomarkers that can be used to predict the prognosis of IAs (Sunderland et al., 2022). Thus, the following sub-section describes how different image segmentation methods influence CFD outcomes.

To verify the effectiveness of the Tversky loss we adopted, we also compared the prediction results trained with six other loss functions, including *binary cross-entropy* (BCE) loss, hybrid loss consisting of dice loss and BCE loss, *exponential logarithmic* dice (ELDice) loss (Wong et al., 2018), dice loss (Milletari et al., 2016), Jaccard loss (Yuan et al., 2017), and *sensitivity-specificity* (SS) loss (Brosch et al., 2015). The effects of different loss functions on the performance of our ARU-Net are reported in Table 3. We found that the Tversky loss was more competitive than all other loss functions, achieving the best scores on all evaluation metrics except SP. The Jaccard and SS loss functions achieved comparable performance because they could handle unbalanced classes to a certain extent. The BCE and Hybrid losses could not converge during the training process and thus failed to get the prediction results. Overall, the Tversky loss function contributed to the improved segmentation results for imbalanced samples, guiding our ARU-Net model to learn more informative IA features in the framework of the proposed multiscale

supervision.

4.3. Ablation experiments

To evaluate the effectiveness of different configurations of the proposed ARU-Net model, ablation experiments were performed on eight different variants of our model, including: 1) the basic encoder-decoder architecture with residual structure and single supervised prediction (Backbone without MS); 2) adding multiscale supervision strategy (Backbone); 3) adding preprocessing (Backbone+Pre); 4) adding attention gates (Backbone+Att); 5) adding both preprocessing and attention gates (Backbone+Pre+Att); 6) adding 3D connected component optimization (Backbone+Pre+Att+CCO); 7) adding 3D conditional random field (Backbone+Pre+Att+CRF); 8) the complete structure with preprocessing, attention gates, and postprocessing (Backbone+Pre+Att+Post). The quantitative performance comparisons of these various variants are listed in Table 4.

As observed in Table 4, our ARU-Net outperformed other baselines and achieved optimal results on most evaluation metrics and was inferior only to the first-ranked baseline (4) and baseline (7) in SE and SP, respectively. The backbone without multiscale supervision (see Table 4 (1)) had the worst performance and achieved the worst scores in almost all the evaluation metrics, suggesting that supervised learning without the proposed multi-supervision strategy was indeed inadequate. The addition of the proposed multi-supervision strategy improved these evaluation metrics; e.g., the DSC score was improved from 0.5913 to 0.6233 (see (2) vs. (1) in Table 4). We also compared the performance of our multiscale supervision strategy with three traditional FPNs, i.e., FPN (Lin et al., 2017), ResFPN (Seferbekov et al., 2018), and hybrid pyramid U-Net (HPU-Net) (Kong et al., 2018). The quantitative performance comparison is presented in Table 5, and our results show that the FPN obtained the best SE score; however, our model achieved the best results on all other evaluation metrics, where DSC was 0.8681, considerably higher than the second-ranked HPU-Net, and RVE was 0.1184, considerably lower than the second-ranked HPU-Net. Our results in Table 5 indicate that our multiscale supervision was significantly better than the FPN structures, and the segmentation results were closer to GT.

By comparison, preprocessing the test images can improve the performance of the model to some extent, as shown in (3) vs. (2) and (5) vs. (4) in Table 4. The DSC scores increased by 0.0440 and 0.0143, the volume errors (see RVE) decreased by 0.2153 and 0.0829, and the surface distances (see ASSD) decreased by 6.1035 mm and 5.3939 mm,

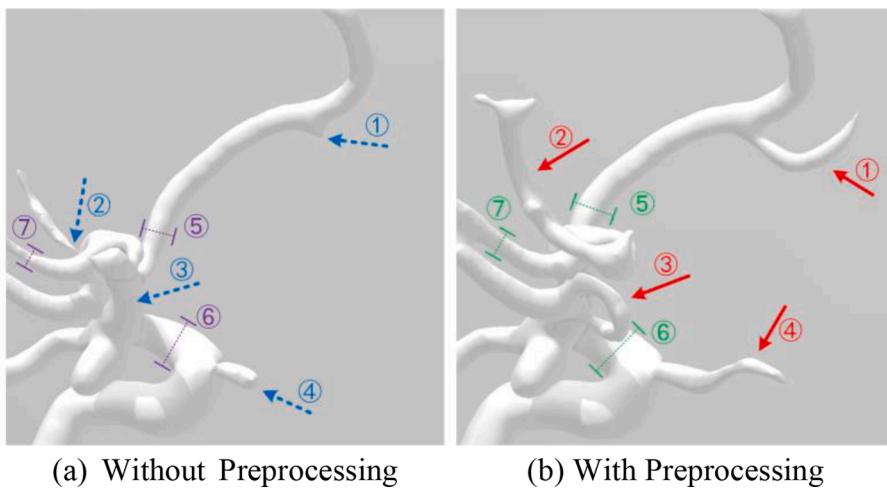


Fig. 8. Subjective performance comparisons of the predicted segmentation results by adding the preprocessing. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

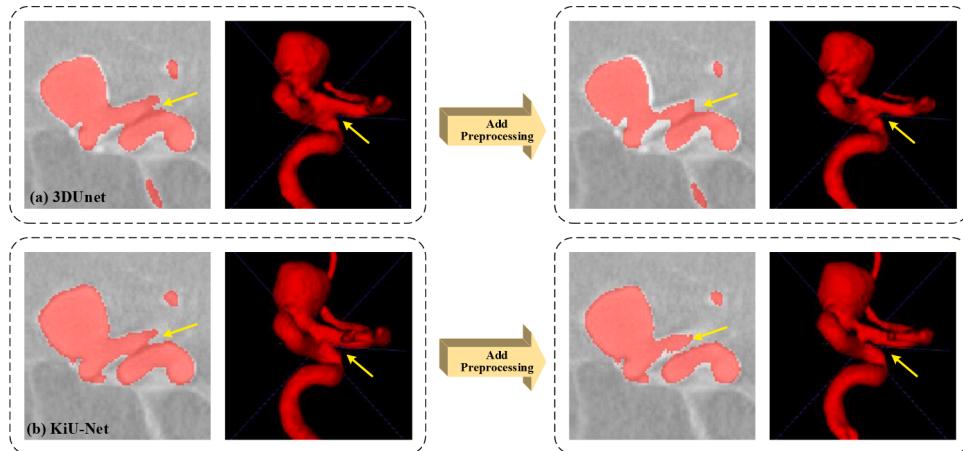


Fig. 9. Visual comparisons of 3DUnet and KiU-Net predictions by adding preprocessing. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6

Objective performance comparisons of 3DUnet and KiU-Net predictions by adding preprocessing.

	DSC↑	RVE↓	SE↑	SP↑	HD95(mm)↓	ASSD(mm)↓
3DUNet	0.7559	0.2593	0.6777	0.9982	46.6503	6.4091
3DUNet + preprocessing	0.7307	0.3036	0.6273	0.9988	54.5144	7.5026
KiU-Net	0.7963	0.1872	0.7720	0.9971	43.7665	4.9428
KiU-Net + preprocessing	0.7549	0.2582	0.7134	0.9978	51.1659	8.3863

respectively. The advantages of preprocessing can also be found after the visualization of intermediate results. In an example shown in Fig. 8, the model with preprocessing can segment more small vessels (as indicated by serial no. 1–4) and effectively avoid the shrinkage of parent arteries (as indicated by serial no. 5–7). In particular, the average time of our model for preprocessing was only 0.860 s per 3D image ($256 \times 256 \times 256$).

Additionally, to verify how our preprocessing plays a role in eliminating the unwanted adhesion of adjacent vessels, we compared the segmentation results of adding the same preprocessing to 3DUnet and KiU-Net, and the qualitative comparison is shown in Fig. 9. After adding preprocessing, the original large-area adhesions were reduced substantially (indicated by the yellow arrows in Fig. 9) because our boundary enhancement operation increased the intensity values of edge pixels and decreased the similarity of adjacent vessel pixels. However, the

adhesions still existed, suggesting that preprocessing alone could not solve the vessel adhesion issue.

We added preprocessing to 3DUnet and KiU-Net and quantitatively compared the performance before and after adding preprocessing. Results are shown in Table 6. Interestingly, the performance of 3DUnet and KiU-Net on six evaluation metrics degraded after adding preprocessing, and both were inferior to those of our ARU-Net. We infer that this result may stem from the inability of these two models to transmit edge signals in network learning. Although preprocessing is beneficial to obtain more edge information, an attention structure is also required to ensure that such valuable information is retained.

Furthermore, our ablation experiments show that evaluation metrics were significantly improved by incorporating the attention gate in the backbone. In Table 4, there are also two pairs of comparable ablations (see (4) vs. (2) and (5) vs. (3)) that can prove the effectiveness of our

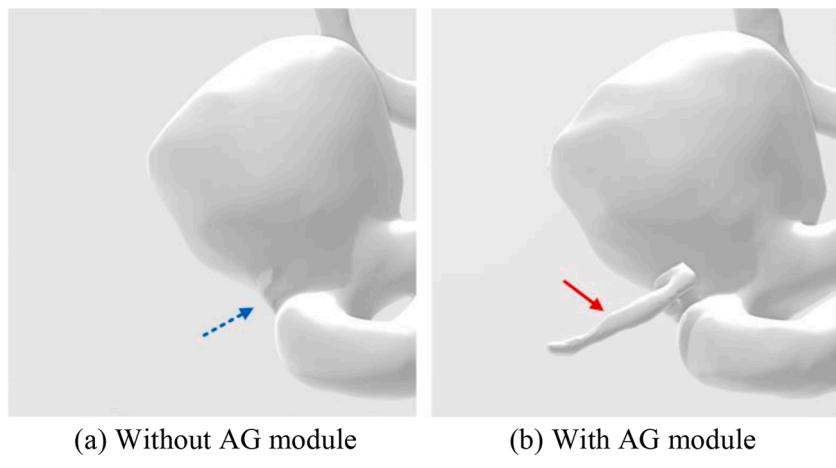


Fig. 10. Subjective performance comparisons of the predicted segmentation results by adding attention gate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

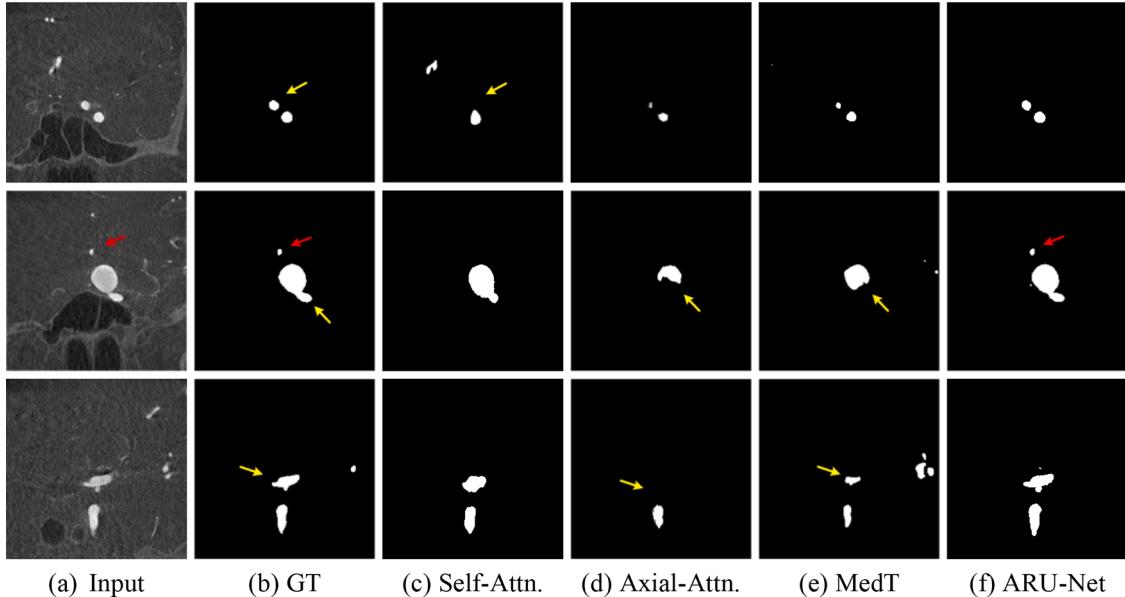


Fig. 11. Subjective performance comparisons between our ARU-Net and the SOTA attention models.

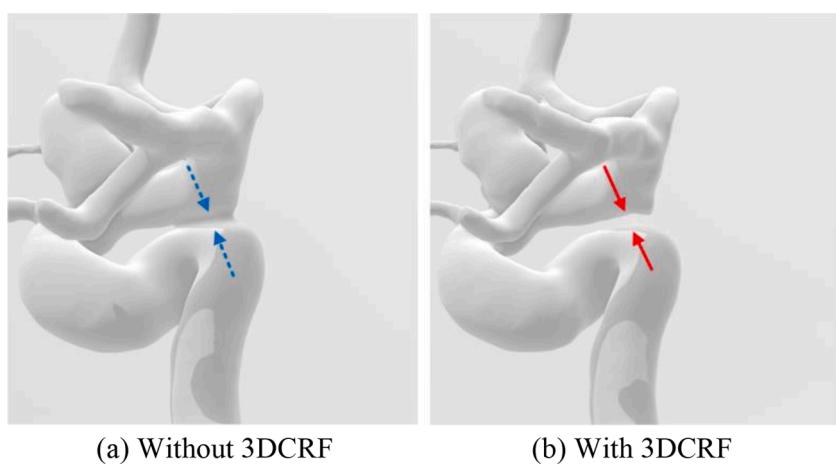


Fig. 12. Subjective performance comparisons of the predicted segmentation results by adding 3DCRF. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

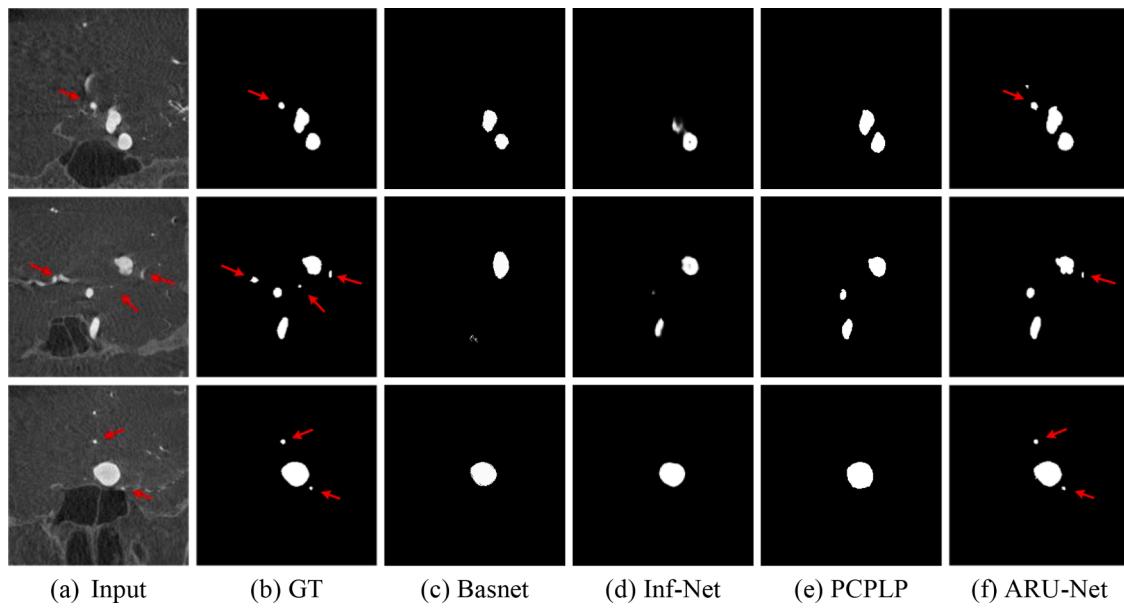


Fig. 13. Qualitative performance comparisons of our ARU-Net and the boundary-aware segmentation models.

Table 7

The PCC values between various automatic segmentation and manual segmentation results with respect to nine hemodynamic parameters and 11 geometric variables. The best results are shown in red color. The asterisk * means that a particular model's results were obtained after removing the failed CFD simulation cases.

	Level-Sets*	Graph-Cuts	DeepMedic	ResUNet*	MIScnn	KiU-Net*	nnU-Net	ARU-Net
Systole STAWSS	0.9913	0.9370	0.9717	0.9191	0.9303	0.9075	0.9892	0.9772
Systole WSSMin	0.9773	0.9974	0.9688	0.6994	0.8471	0.9744	0.9753	0.9828
Systole WSSMax	0.7373	0.4522	0.7842	0.5932	0.7587	0.6616	0.7826	0.8888
Mean OSI	0.2281	0.7064	0.8260	0.4896	0.2680	0.2583	0.7522	0.9111
Std OSI	-0.0538	0.4596	0.7618	0.1724	0.2428	0.5493	0.6332	0.6065
TA LSA 2	0.9709	0.9324	0.8771	0.9722	0.8897	0.3931	0.9737	0.9033
TA LSA Std 2	-0.2551	0.1326	0.8768	0.9743	0.8620	0.8831	0.9793	0.9691
Systole TADVO	0.5256	0.5297	0.1820	0.9939	0.5575	0.5661	0.9905	0.9914
Systole DVOSTd	0.8513	0.8837	0.0532	0.9527	0.8823	-0.4629	0.9692	0.9784
Mean PCC	0.5525	0.6701	0.7002	0.7519	0.6932	0.5256	0.8939	0.9121
±95% confidence	± 0.3097	± 0.1925	± 0.2218	± 0.1873	± 0.1773	± 0.2887	± 0.0879	± 0.0791
Aneurysm Volume	0.9996	0.9973	0.9996	0.9992	0.9995	0.9998	0.9974	0.9995
Aneurysm Height	0.9851	0.9972	0.9965	0.9904	0.9916	0.9912	0.9730	0.9538
Sac Max Width	0.9913	0.9874	0.9971	0.9726	0.9965	0.9939	0.9801	0.9855
Size Ratio Height	0.9193	0.9801	0.9775	0.9760	0.9861	0.9712	0.9660	0.8950
Size Ratio Width	0.9840	0.9714	0.9880	0.9898	0.9676	0.9875	0.9682	0.9760
Aspect Ratio Star	0.9605	0.9889	0.9760	0.9332	0.9497	0.9742	0.9538	0.8801
Vessel Diameter	-0.0783	0.9469	0.9318	0.8306	0.9104	0.8693	0.8907	0.8594
Ostium Minimum	0.9528	0.9446	0.8902	0.9339	0.8564	0.9487	0.9169	0.8120
Ostium Maximum	0.9810	0.9547	0.9752	0.8914	0.9741	0.9430	0.9895	0.9730
Aneurysm Area	0.9997	0.9983	0.9991	0.9997	0.9988	0.9997	0.9972	0.9993
Ostium Area	0.9863	0.9752	0.9890	0.9404	0.9606	0.9706	0.9829	0.9710
Mean PCC	0.8801	0.9765	0.9745	0.9507	0.9628	0.9681	0.9651	0.9368
±95% confidence	± 0.1884	± 0.0119	± 0.0201	± 0.0311	± 0.0261	± 0.0224	± 0.0198	± 0.0379

attention gate module, i.e., ablations (4) and (5) with attention-based long skip connections versus ablations (2) and (3) using traditional long skip connections of U-Net, respectively. Comparing (4) to (2) in Table 4, after adding the attention gates, the DSC score increased by 0.1138, RVE decreased by 0.4052, and HD95 and ASSD decreased by 28.3727 mm and 13.8387 mm, respectively. Similarly, comparing (5) to (3), we found that the DSC score increased by 0.0841, RVE decreased by 0.2728, and HD95 and ASSD decreased by 24.2735 mm and 13.1291 mm, respectively.

Although traditional long skip connections had a positive effect on propagating spatial information from the encoding stage, the transmitted information is redundant; i.e., the target and background knowledge were simultaneously compensated to the decoding stage, which cannot provide an effective reference for subsequent layers. However, our attention-based long connections were capable of

highlighting targets and suppressing backgrounds during information propagation. Overall, the attention gates in the proposed model can help the local segmentation network learn global saliency information to improve the structural accuracy of IAs. This is extremely important for the current patch-based segmentation network, which can process only one small local patch split from the global image at a time, losing the global structural information of an IA and its attached arteries. As can be seen from Fig. 10, the model containing the attention gate can generate aneurysms with a more accurate structure, avoiding the loss of some key arteries, e.g., the ophthalmic artery growing on the aneurysm indicated by the red arrow, which has an important impact on the CFD simulation.

To further verify the superiority of our depth-aware attention gate module, we compared our ARU-Net with the state-of-the-art (SOTA) attention structures, including self-attention (Vaswani et al., 2017), axial attention (Wang et al., 2020), and medical transformer (MedT) (J.

Table 8

Comparison of P-values for hemodynamic and geometric parameters obtained from different segmentation models. The best results are shown in red color. The asterisk * means that a particular model's results were obtained after removing the failed CFD simulation cases.

	Level-Sets*	Graph-Cuts	DeepMedic	ResUNet*	MIScnn	KiU-Net*	nnU-Net	ARU-Net
Systole STAWSS	0.0318	0.0343	0.4338	0.1046	0.4757	0.0744	0.3249	0.0960
Systole WSSMin	0.2411	0.0841	0.3706	0.4920	0.3564	0.1936	0.3209	0.1184
Systole WSSMax	0.1556	0.1441	0.1655	0.4294	0.3464	0.0163	0.0604	0.0253
Mean OSI	0.3314	0.2779	0.2264	0.3177	0.2521	0.2274	0.0244	0.0400
Std OSI	0.3692	0.4195	0.2066	0.3757	0.2964	0.1832	0.0352	0.0898
TA LSA 2	0.0309	0.0100	0.1029	0.0379	0.1856	0.2690	0.0556	0.1538
TA LSA Std 2	0.3173	0.0248	0.2704	0.1375	0.1426	0.0844	0.1828	0.0251
Systole TADVO	0.2254	0.1806	0.2133	0.1074	0.2304	0.1597	0.1223	0.3582
Systole DVOSTd	0.0543	0.0962	0.3350	0.2675	0.0058	0.2492	0.0365	0.1709
Mean P-value	0.1952	0.1413	0.2583	0.2522	0.2546	0.1619	0.1292	0.1197
±95% confidence	± 0.0870	± 0.0880	± 0.0685	± 0.1060	± 0.0892	± 0.0564	± 0.0789	± 0.0679
Aneurysm Volume	0.0144	0.0361	0.0375	0.0537	0.1168	0.0516	0.3938	0.1373
Aneurysm Height	0.2491	0.3734	0.0774	0.1026	0.4734	0.0989	0.2487	0.4463
Sac Max Width	0.2639	0.3051	0.4380	0.2656	0.3748	0.0526	0.3558	0.3217
Size Ratio Height	0.2483	0.1206	0.1020	0.3152	0.4627	0.1062	0.1176	0.3600
Size Ratio Width	0.0771	0.1331	0.0771	0.0336	0.1789	0.0055	0.2944	0.0878
Aspect Ratio Star	0.0803	0.0054	0.1950	0.4155	0.3281	0.0457	0.0646	0.2363
Vessel Diameter	0.1791	0.0013	0.0967	0.0475	0.1385	0.3917	0.4441	0.1467
Ostium Minimum	0.0153	0.0048	0.1364	0.1378	0.2887	0.1347	0.1956	0.2163
Ostium Maximum	0.0008	0.0031	0.0107	0.0134	0.0126	0.3604	0.4673	0.0526
Aneurysm Area	0.0029	0.0301	0.0507	0.0299	0.2599	0.0453	0.4564	0.0344
Ostium Area	0.0039	0.0012	0.0098	0.0227	0.0557	0.1204	0.1567	0.0804
Mean P-value	0.1032	0.0922	0.1119	0.1307	0.2446	0.1285	0.2905	0.1927
±95% confidence	± 0.0650	± 0.0778	± 0.0715	± 0.0819	± 0.0930	± 0.0759	± 0.0850	± 0.0804

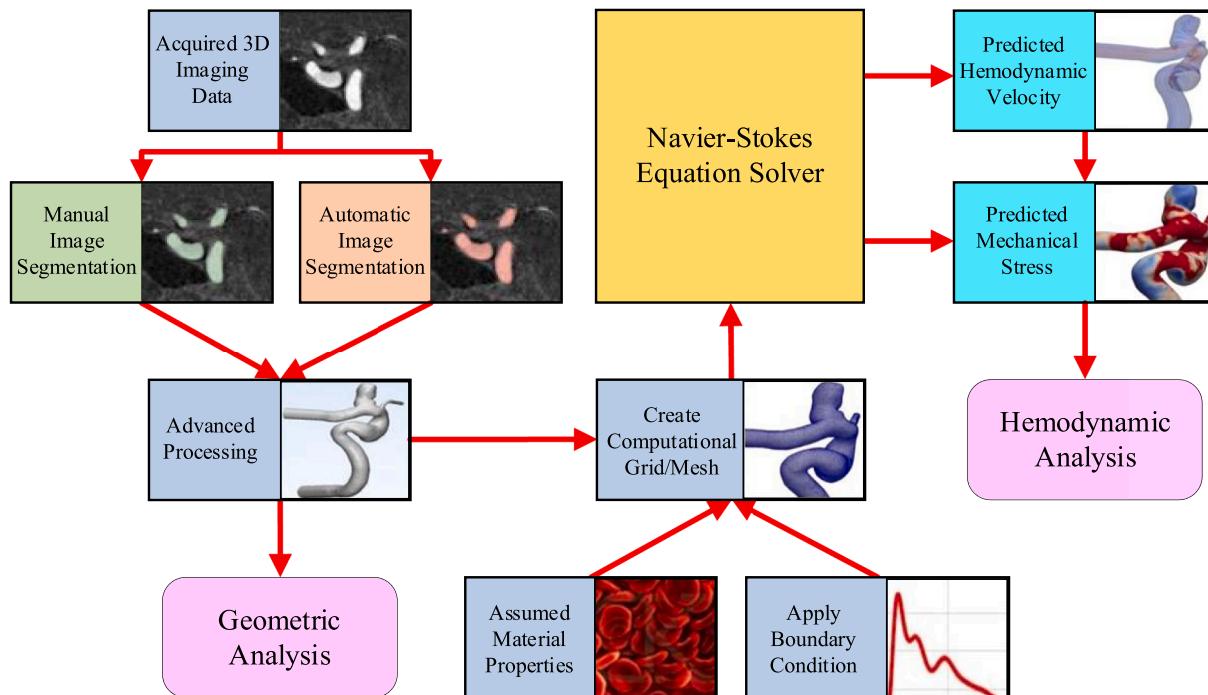


Fig. A.1. An overview of the workflow for generating the hemodynamic and geometric parameters.

(M.J. Valanarasu et al., 2021). The subjective performance comparisons are presented in Fig. 11. It shows that the IA regions detected by these SOTA methods were incomplete (indicated by the yellow arrows). At the same time, the IAs predicted by our ARU-Net were much closer to GT, and our model could preserve small vessels (indicated by the red arrows).

In addition, Table 4 indicates that adding CCO (see (6) vs. (5)) and CRF (see (7) vs. (5)), respectively, can also improve the performance of all six evaluation metrics, which can optimize the structure of the segmented region. By analyzing Table 4, after adding CCO and CRF operations, DSC scores increased by 0.0378 and 0.0751, RVE scores

decreased by 0.1248 and 0.2477, and HD95 scores decreased by 72.9319 mm and 76.8752 mm, respectively. In addition, by adding both CCO and CRF (see (8) vs. (5)), the performance of most evaluation metrics has been greatly improved, with DSC, RVE, HD95, and ASSD achieving the best results among all eight variants. As shown in Fig. 12, adding 3DCRF can effectively eliminate the unwanted connections between the internal carotid arteries. The average time of our post-processing for 10 iterations of CRF was 21.518 s per 3D image.

We also compared the performance of our ARU-Net with the boundary-aware segmentation models, e.g., boundary-aware network (Hatamizadeh et al., 2019), Basnet (Qin et al., 2019), Inf-Net (Fan et al.,

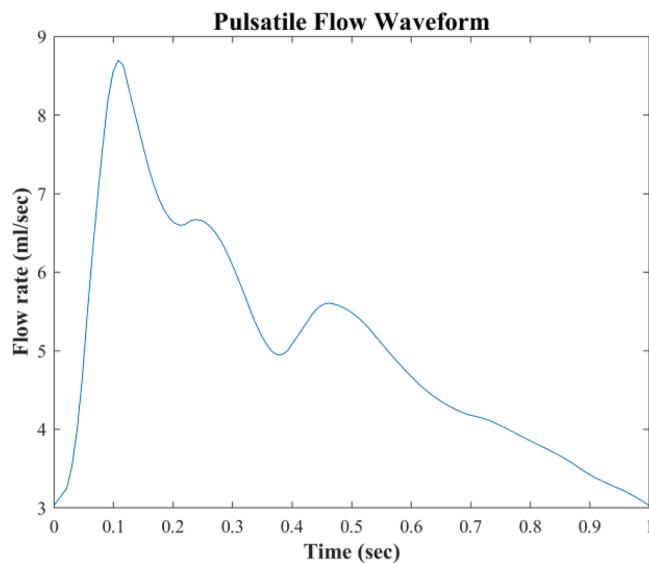


Fig. A.2. The inlet waveform of volumetric flow rate used in this research.

Table A.1
An overview of the hemodynamic and geometric parameters.

Types	Parameters	Descriptions
Hemodynamic parameters	Systole	Spatially and temporally averaged wall shear stress during systole
	STAWSS	Wall shear stress minimum during systole
	Systole	Wall shear stress maximum during systole
	WSSMin	Spatially averaged oscillatory shear index
	WSSMax	One standard deviation of oscillatory shear index
	Mean OSI	Time-averaged low shear area less than 2 Pa
	Std OSI	One standard deviation of time-averaged low shear area less than 2 Pa
	TA LSA 2	Time-average degree of overlap between flow vortex cores during systole
	TA LSA Std 2	One standard deviation of time-average degree of overlap between flow vortex cores during systole
	Systole	One standard deviation of time-average degree of overlap between flow vortex cores during systole
	TADVO	Volume of the aneurysm
	Systole	Height of the aneurysm
Geometric parameters	DVOStd	Maximum width of aneurysm sac
	Width	Size ratio between aneurysm height and parent vessel diameter
	Size Ratio	Size ratio between aneurysm width and parent vessel diameter
	Height	Aspect ratio of the intracranial aneurysm
	Size Ratio	Aspect ratio of the intracranial aneurysm
	Width	Star
	Aspect Ratio	Diameter of the parental vessel connected to the aneurysm
	Vessel	Diameter
	Diameter	Ostium
	Ostium	Minimum
	Minimum	Ostium
	Ostium	Maximum
	Maximum	Aneurysm
	Aneurysm	Area
	Area	Ostium Area

2020), and PCPLP (Mu et al., 2021). These models typically utilized boundary loss to supervise the outputs of the mainstream and edge streams to jointly learn semantics and boundaries, whereas our preprocessing was implemented through a boundary augmentation operation that did not require any training operations. Coupled with the

attention gate mechanism, our preprocessing allows the ARU-Net to preserve and emphasize the edge details of small vessels and IAs to provide rich discriminative information for subsequent predictions. The subjective performance comparisons are presented in Fig. 13, which shows that our ARU-Net outperformed the boundary-aware segmentation models in segmenting small vessels (indicated by the red arrows). Moreover, the average test time of Basnet, Inf-Net, and PCPLP for a 256×256 image was 0.079, 0.047, and 5.533 s, respectively, while ours was 0.013 s, which was more efficient.

4.4. CFD experiments

Besides the subjective and objective performance comparisons of various segmentation models, we also explored the clinical utility of the proposed ARU-Net model. In particular, we performed CFD simulations. "Patient-specific" CFD simulations are well documented in the literature (Sunderland et al., 2021; Sunderland et al., 2016), and thus only essential details are provided in the Appendix for the sake of completeness.

Upon completion of CFD simulations, in-house Python scripts derived from the Visualization ToolKit (Kitware Inc., NY, USA) were used to compute the hemodynamic and geometric parameters of the IAs for CFD models created by nine image segmentation methods summarized in Sub-Section 4.1.

The Pearson correlation coefficient (PCC) was calculated to compare the consistency between model segmentation and human expert segmentation results in terms of nine hemodynamic parameters and 11 geometric variables. Those parameters can be found in the Appendix as well. The PCC values of different models compared with the manual segmentation results on 20 parameters are presented in Table 7. In the execution, each of the Level-Sets, ResUNet, and KiU-Net models contained one case that could not be correctly segmented and led to failed CFD simulations. For traditional methods (Level-Sets and Graph-Cuts), the segmentation results required extensive postprocessing to be successfully simulated by CFD. In addition, since the frameworks of 3DUNet and 3DResUNet are similar and the performance of 3DUNet is inferior to that of 3DResUNet, we did not conduct CFD modeling for 3DUNet.

As shown in Table 7, the larger the PCC value is, the closer the simulation value of the automatic segmentation result is, compared with the manual segmentation result. For the proposed ARU-Net model, the PCC values of all the hemodynamic and geometric parameters except Std OSI are higher than 80%, and 14 of them are higher than 90%. Furthermore, our model had the highest average PCC value (0.9121) on hemodynamic parameters, which was higher than the second-ranked nnU-Net model (0.8939). That is, the CFD simulations of our predicted segmentation results were strongly positively correlated with the human-annotated measurements. Meanwhile, the lowest confidence score (0.0791) also confirmed no statistical difference between our segmentation and manual segmentation. For geometric variables, the mean PCC value of our model is not the best, but there is only a small difference (0.0397) from the best Graph-Cuts model.

Furthermore, the P-values of the nine hemodynamic parameters and 11 geometric variables of the automatic and manual segmentation results were also calculated and are summarized in Table 8. Our results' mean P-values (0.1197) were the best for hemodynamic parameters. The lower 95% confidence (0.0679) also confirmed the stability and accuracy of our results. For geometric variables, the mean P-values of our ARU-Net were significantly better than those of nnUNet. The traditional models, Level-sets and Graph-Cuts, had excellent P-values on some parameters. However, manual postprocessing after image segmentation by these two traditional methods was required for CFD simulations. In addition, as we mentioned earlier, the hemodynamic and geometric results of Level-Sets, ResUNet, and KiU-Net were obtained after removing the CFD simulation failure cases (one failed case for each model). Overall, the segmentation results obtained by the proposed ARU-Net were helpful for the clinical interpretation of IAs. Our model

can help clinicians improve their diagnostic accuracy and save diagnostic time.

5. Conclusion

In this paper, we proposed an attention-based multiscale supervision fully convolutional encoder-decoder network with differential preprocessing and geometric postprocessing for the segmentation of IAs from "patient-specific" 3DRA images. Our work demonstrated that depth-aware attention gates could effectively emphasize important targets on 3DRA images, thereby improving the accurate segmentation of smaller but critical vessels by enhancing the FOV of the structure. The multiscale supervision strategy also helped the network integrate multi-level spatial and semantic information to increase the sensitivity of smaller aneurysms and vessels. The 3DCRF and 3DCCO were introduced to achieve better segmentation results, as they have the ability to eliminate unwanted connections and optimize connected components. Extensive experiments demonstrated that the proposed ARU-Net is effective and can achieve performance better than or comparable to state-of-the-art segmentation methods. Meanwhile, the CFD model derived from the predicted mask showed good clinical application prospects.

Recall that the objectives of our proposed image segmentation method were to 1) include small arteries, 2) maintain accurate representations of IAs, and 3) eliminate unwanted vessel/aneurysm connections. We innovatively used and augmented available arsenals in the literature to reach our objectives. This work is an excellent example of how deep-learning methods can be devised to solve a domain-specific problem.

Although our ARU-Net achieved remarkable performance for automatic segmentation of IAs, especially for detecting small critical vessels and avoiding unwanted adhesions, future research will focus on using a cascaded network structure for better training. Additionally, since our experimental results showed that traditional evaluation metrics (e.g., DISK score) could not adequately verify the performance of segmentation models, we will develop more domain-specific metrics for quantitative analysis. This is also an important task, as researchers in the healthcare space are increasingly interested in the integration of deep-learning-based methods for advanced modeling. We feel our interdisciplinary group is in an excellent position to work with others to accomplish this goal.

Appendix A. A Description of Computational Hemodynamic Analysis

A.1. Process for Acquiring Hemodynamic and Geometric Parameters

The overall workflow of obtaining hemodynamic and geometric parameters for CFD simulations of the manual/automatic segmentation is shown in Fig. A.1.

A.2. Navier-Stokes Solver and Boundary conditions

To computationally simulate blood flow in and around an IA as shown in Fig. A.1, 3D Navier-Stokes equations (i.e., Equation A.1) were solved using a finite-volume-based CFD solver (Fluent, ANSYS Inc., PA, USA):

$$\rho \frac{\partial \vec{v}}{\partial t} + \rho(\vec{v} \cdot \vec{\nabla})\vec{v} = -\vec{\nabla}p + \mu \nabla^2 \vec{v}, \\ \vec{\nabla} \cdot \vec{v} = 0, \quad (A.1)$$

where ρ denotes the blood density, p denotes the pressure, \vec{v} denotes the three-dimensional velocity vector, and μ denotes the viscosity.

The vessel walls were assumed rigid with a no-slip boundary condition, and blood was modeled as incompressible and Newtonian. The dynamic viscosity and the mass density of blood were set to 0.004 kg/m·s and 1050 kg/m³, respectively. As shown in Fig. A.2, a pulsatile flow waveform measured using magnetic resonance flow imaging (Gwilliam et al., 2009) was applied as the inlet boundary condition.

The zero-pressure condition was used for all outlets. Four cardiac cycles were simulated at 2000 steps per period (0.0005 s/timestep). By selecting

Author declaration template

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from jjiang1@mtu.edu.

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.

Data availability

Data will be made available on request.

Acknowledgments

The study is supported by a grant from the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (R01-EB029570A1). We also want to thank anonymous reviewers for their constructive comments that have made our presentation and testing complete, improving the quality of our work.

2000 time steps, Courant numbers were below 1 for the above-said waveform. Time-resolved hemodynamic data (20 time points) were saved for the last cardiac cycle for further processing. The systolic or peak Reynolds numbers were approximately between 350 and 550 for the ICAs, similar to those reported by others. We verified that our mesh density (1.5 million computing cells for a typical CFD model) was appropriate.

A.3. An Overview of the Hemodynamic and Geometric Parameters

The nine hemodynamic parameters and 11 geometric variables utilized in the manuscript are presented in Table A.1.

References

- Saqr, K.M., Rashad, S., Tupin, S., Niizuma, K., Hassan, T., Tominaga, T., Ohta, M., 2020. What does computational fluid dynamics tell us about intracranial aneurysms? A meta-analysis and critical review. *J. Cereb. Blood Flow Metab.* 40 (5), 1021–1039.
- Antiga, L., Piccinelli, M., Botti, L., Ene-Iordache, B., Remuzzi, A., Steinman, D.A., 2008. An image-based modeling framework for patient-specific computational hemodynamics. *Med. Biol. Eng. Comput.* Vol. 46 (11), 1097–1112.
- Cebral, J.R., Mut, F., Weir, J., Putman, C., 2011. Quantitative characterization of the hemodynamic environment in ruptured and unruptured brain aneurysms. *Am. J. Neuroradiol.* 32 (1), 145–151.
- Meng, H., Tutino, V.M., Xiang, J., Siddiqui, A., 2014. High WSS or low WSS? Complex interactions of hemodynamics with intracranial aneurysm initiation, growth, and rupture: toward a unifying hypothesis. *Am. J. Neuroradiol.* 35 (7), 1254–1262.
- Dhar, S., Tremmel, M., Mocco, J., Kim, M., Yamamoto, J., Siddiqui, A.H., Hopkins, L.N., Meng, H., 2008. Morphology parameters for intracranial aneurysm rupture risk assessment. *Neurosurgery* 63 (2), 185–197.
- Damiano, R.J., Tutino, V.M., Paliwal, N., Patel, T.R., Waqas, M., Levy, E.I., Davies, J.M., Siddiqui, A.H., Meng, H., 2020. Aneurysm characteristics, coil packing, and post-coiling hemodynamics affect long-term treatment outcome. *J. Neurointerv. Surg.* 12 (7), 706–713.
- Xiang, J., Antiga, L., Varble, N., Snyder, K.V., Levy, E.I., Siddiqui, A.H., Meng, H., 2016. AView: an image-based clinical computational tool for intracranial aneurysm flow visualization and clinical management. *Ann. Biomed. Eng.* 44 (4), 1085–1096.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. *Med. Image Comput. Comp.-Assist. Intervent.* 234–241.
- Yang, J., Xie, M., Hu, C., Alwallid, O., Xu, Y., Liu, J., Jin, T., Li, C., Tu, D., Liu, X., Zhang, C., Li, C., Long, X., 2021. Deep learning for detecting cerebral aneurysms with CT angiography. *Radiology* 298 (1), 155–163.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 424–432.
- Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18 (2), 203–211.
- Kastner, S., Ungerleider, L.G., 2000. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23 (1), 315–341.
- Itti, L., Koch, C., 2001. Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2 (3), 194–203.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78.
- Bhalerao, M., Thakur, S., 2019. Brain tumor segmentation based on 3D residual U-Net. In: International MICCAI Brainlesions Workshop, pp. 218–225.
- Müller, D., Kramer, F., 2021. MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. *BMC Med. Imag.* 21 (1), 1–11.
- Valanarasu, J.M.J., Sindagi, V.A., Hacihamoglu, I., Patel, V.M., 2021a. KiU-Net: overcomplete convolutional architectures for biomedical image and volumetric segmentation. *IEEE Trans. Med. Imag.* 1–13.
- Piccinelli, M., Veneziani, A., Steinman, D.A., Remuzzi, A., Antiga, L., 2009. A framework for geometric analysis of vascular structures: application to cerebral aneurysms. *IEEE Trans. Med. Imag.* 28 (8), 1141–1155.
- Jirik, M., Lukes, V., Svobodova, M., Zelezny, M., 2013. Image segmentation in medical imaging via graph-cuts. In: 11th International Conference on Pattern Recognition and Image Analysis: New Information Technologies, pp. 201–204.
- Sen, Y., Qian, Y., Avolio, A., Morgan, M., 2014. Image segmentation methods for intracranial aneurysm haemodynamic research. *J. Biomech.* 47 (5), 1014–1019.
- Flanders, A.E., 2019. Machine learning detection of intracranial aneurysms—Will it play in Peoria? *Radiology* 290 (1), 195–197.
- Nakao, T., Hanaoka, S., Nomura, Y., Sato, I., Nemoto, M., Miki, S., Maeda, E., Yoshikawa, T., Hayashi, N., Abe, O., 2018. Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography. *J. Magnet. Reson. Imag.* 47 (4), 948–953.
- Stember, J.N., Chang, P., Stember, D.M., Liu, M., Grinband, J., Filippi, C.G., Meyers, P., Jambawalikar, S., 2019. Convolutional neural networks for the detection and measurement of cerebral aneurysms on magnetic resonance angiography. *J. Digit. Imag.* 32, 808–815.
- Jin, H., Geng, J., Yin, Y., Hu, M., Yang, G., Xiang, S., Zhai, X., Ji, Z., Fan, X., Hu, P., He, C., Qin, L., Zhang, H., 2020. Fully automated intracranial aneurysm detection and segmentation from digital subtraction angiography series using an end-to-end spatiotemporal deep neural network. *J. Neurointerv. Surg.* 12 (10), 1023–1027.
- Patel, T.R., Paliwal, N., Jaiswal, P., Waqas, M., Mokin, M., Siddiqui, A.H., Meng, H., Rai, R., Tutino, V., 2020. Multi-resolution CNN for brain vessel segmentation from cerebrovascular images of intracranial aneurysm: a comparison of U-Net and DeepMedic. *Comp.-Aid. Diagn.* 2020, 113142W.
- Shahzad, R., Pennig, L., Goertz, L., Thiele, F., Kabbasch, C., Schlamann, M., Krischek, B., Maintz, D., Perkuhn, M., Borggrefe, J., 2020. Fully automated detection and segmentation of intracranial aneurysms in subarachnoid hemorrhage on CTA using deep learning. *Sci. Rep.* 10 (1), 1–12.
- Cheng, M., Xiao, N., Yuan, H., Wang, K., 2021. Automatic intracranial aneurysm segmentation based on spatial information fusion feature from 3D-RA using U-Net. In: IEEE International Conference on Mechatronics and Automation, pp. 236–241.
- D. Shao, X. Lu, and X. Liu, “3D intracranial aneurysm classification and segmentation via unsupervised Dual-branch learning,” arXiv preprint arXiv:2201.02198, pp. 1–10, 2022.
- Mnih, V., Heess, N., Graves, A., kavukcuoglu, k., 2014. Recurrent models of visual attention. In: Advances in Neural Information Processing Systems, pp. 1–9.
- Bello, I., Zoph, B., Vaswani, A., Shlens, J., Le, Q.V., 2019. Attention augmented convolutional networks. In: IEEE International Conference on Computer Vision, pp. 3286–3295.
- Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3146–3154.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An image is worth 16x16 words: transformers for image recognition at scale. In: International Conference on Learning Representations, pp. 1–22.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: hierarchical vision transformer using shifted windows. In: IEEE International Conference on Computer Vision, pp. 10012–10022.
- Valanarasu, J.M.J., Oza, P., Hacihamoglu, I., Patel, V.M., 2021b. Medical transformer: gated axial-attention for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention, pp. 36–46.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: IEEE International Conference on Computer Vision, pp. 1026–1034.
- Oktay, J., Schlemper, J.L., Folgoc, M., Lee, M., Heinrich, K., Misawa, K., Mori, S., McDonagh, N.Y., Hamerla, B., Kainz, B., Glocker, and D. Rueckert, “Attention u-net: learning where to look for the pancreas,” arXiv preprint arXiv:1804.03999, pp. 1–10, 2018.
- Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125.
- Seferbekov, S., Iglovikov, V., Buslaev, A., Shvets, A., 2018. Feature pyramid network for multi-class land segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 272–275.
- Kong, X., Sun, G., Wu, Q., Liu, J., Lin, F., 2018. Hybrid pyramid u-net model for brain tumor segmentation. In: International Conference on Intelligent Information Processing, pp. 346–355.
- Salehi, S.S.M., Erdogmus, D., Gholipour, A., 2017. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: International Workshop on Machine Learning in Medical Imaging, pp. 379–387.
- Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected CRFs with gaussian edge potentials. In: Advances in Neural Information Processing Systems, pp. 109–117.
- Sunderland, K., Jiang, J., Zhao, F., 2022. Disturbed flow's impact on cellular changes indicative of vascular aneurysm initiation, expansion, and rupture: a pathological and methodological review. *J. Cell. Physiol.* 237 (1), 278–300.
- Wong, K.C.L., Moradi, M., Tang, H., Syeda-Mahmood, T., 2018. 3D segmentation with exponential logarithmic loss for highly unbalanced object sizes. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 612–619.
- Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision, pp. 565–571.
- Yuan, Y., Chao, M., Lo, Y.-C., 2017. Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Trans. Med. Imag.* 36 (9), 1876–1886.
- Brosch, T., Yoo, Y., Tang, L.Y.W., Li, D.K.B., Traboulsi, A., Tam, R., 2015. Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 3–11.

- Hatamizadeh, A., Terzopoulos, D., Myronenko, A., 2019. End-to-end boundary aware networks for medical image segmentation. In: International Workshop on Machine Learning in Medical Imaging, pp. 187–194.
- Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M., 2019. Basnet: boundary-aware salient object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 7479–7489.
- Fan, D.-P., Zhou, T., Ji, G.-P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L., 2020. Inf-net: automatic covid-19 lung infection segmentation from ct images. *IEEE Trans. Med. Imag.* 39 (8), 2626–2637.
- Mu, N., Wang, H., Zhang, Y., Jiang, J., Tang, J., 2021. Progressive global perception and local polishing network for lung infection segmentation of COVID-19 CT images. *Pattern Recognit.* 120, 1–12.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: 31st Conference on Neural Information Processing Systems, pp. 1–11.
- Wang, H., Zhu, Y., Green, B., Adam, H., Yuille, A., Chen, L.-C., 2020. Axial-deeplab: stand-alone axial-attention for panoptic segmentation. In: European Conference on Computer Vision, pp. 108–126.
- Sunderland, K., Wang, M., Pandey, A.S., Gemmete, J., Huang, Q., Goudge, A., Jiang, J., 2021. Quantitative analysis of flow vortices: differentiation of unruptured and ruptured medium-sized middle cerebral artery aneurysms. *Acta Neurochir (Wien)* 163, 2339–2349.
- Sunderland, K., Haferman, C., Chintalapani, G., Jiang, J., 2016. Vortex analysis of intraneurismal flow in cerebral aneurysms. *Comput. Math. Methods Med.* 2016 (1–17).
- Gwilliam, M.N., Hoggard, N., Capener, D., Singh, P., Marzo, A., Verma, P.K., Wilkinson, I.D., 2009. MR derived volumetric flow rate waveforms at locations within the common carotid, internal carotid, and basilar arteries. *J. Cereb. Blood Flow Metabol.* 29 (12), 1975–1982.