MVRNet: A Multi-View Refinement Network for Accurate Recognition of Challenging Intracranial Aneurysms in Enhanced 3D CTA Images

Peiying Li[1] , **Yongchang Liu**[2] , **Shikui Tu**[2,*] and **Lei Xu**[2,3,*]
[1]School of Computer Engineering and Science, Shanghai University, Shanghai, China
[2]School of Computer Science, Shanghai Jiao Tong University, Shanghai, China
[3]Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Guangdong, China
lipeiying@shu.edu.cn, {liuyongchang,tushikui,leixu}@sjtu.edu.cn

Abstract

Intracranial aneurysms are life-threatening and require accurate, timely detection. Traditional manual diagnosis by radiologists can be subjective, leading to misdiagnoses, while existing deep learning approaches struggle with small aneurysms or cases complicated by surrounding tissues. In this paper, we present the Multi-View Refinement Network (MVRNet), a framework that delivers clinically meaningful performance gains on the most challenging intracranial aneurysm cases. It directly addresses real-world diagnostic difficulties, particularly for small, obscured, or ambiguously bounded lesions that conventional methods often miss. In particular, we propose a feature enhancement technique to obtain boundary-clear CTA, an informative supplement to the bone-free CTA, to further reduce the impact of noise and enhance the robustness of IA recognition. Moreover, we develop a multi-view encoder with 2D slicing in different directions to mitigate tissue occlusion effects, coupled with a progressively refined decoder that iteratively corrects uncertain predictions, ensuring precise localization and segmentation. Experimental results indicate that MVRNet improves the F1-score by 24% and the IoU by 16% in comparison with the state-of-the-art methods. Remarkably, previous methods struggled with challenging cases, while our method still achieves satisfactory results, demonstrating a $> 2\times$ improvement over prior arts on challenging test sets. Our code has been released in https://github.com/TheResearchWorks/MVRNet.

1 Introduction

Intracranial aneurysms (IA) are defined as localized dilations of the walls of cerebral blood vessels, typically resulting from weaknesses or damage in the vascular structure. These lesions can often exist without causing symptoms until they rupture, leading to hemorrhagic stroke, which can pose a significant threat to life. The prevalence of intracranial aneurysms in the global population is approximately 6% [Rikhtegar *et al.*, 2021]. In the United States, an IA ruptures approximately every 18 minutes, leading to about 30,000 cases of subarachnoid hemorrhage each year, with a mortality rate of up to 40% [Irfan *et al.*, 2023]. Therefore, accurate IA detection and rupture risk prediction are essential. Recent deep learning methods show promise for rupture risk prediction [Li *et al.*, 2023], but precise IA detection and segmentation, especially in difficult cases, remain major challenges.

Computed Tomography Angiography (CTA) is an essential imaging technique for the diagnosis of IAs [Kato *et al.*, 1999], as it can swiftly and safely deliver high-resolution vascular images along with comprehensive angiographic details. Traditionally, CTA images are analyzed and interpreted by experienced radiologists, whose expertise is crucial for accurate diagnoses. However, manual analysis of medical images is a time-consuming and labor-intensive process, and it is susceptible to subjective factors. Moreover, it is very difficult to detect the challenging cases, such as small IAs or those significantly affected by surrounding tissues, which are easily missed or incorrectly diagnosed. Fig. 1 shows examples of IAs. The first row shows EIAs (Easy IAs), which are easily identifiable and have good visibility, with sizes larger than 4 mm. Most previously proposed deep learning models perform well on this kind of data. The second row presents AIAs (Adjacent IAs), which are difficult to discern due to interference from adjacent tissues. The third row presents SIAs (Small IAs), which are smaller than 4 mm in diameter and challenging to detect. Due to the difficulties of AIA and SIA, previous approaches have been largely ineffective, making their clinical application challenging.

In recent years, several deep learning methods have been proposed for the detection and segmentation of IAs. Ueda et al. [Ueda *et al.*, 2019] extracted hundreds of arterial anomaly samples to train a ResNet-18 [He *et al.*, 2016] model. This model outputs a probability for each image patch, and after sorting them from high to low, the final aneurysm candidates are obtained. Dai et al. [Dai *et al.*, 2020] performed preprocessing on raw CTA slices, including the removal of bright pixels and nearby projections, resulting in NP images. They then trained a Faster R-CNN [Ren *et al.*, 2017] model to detect aneurysms from NP images. Both methods are based on 2D images and do not fully utilize 3D information, resulting

*Corresponding author

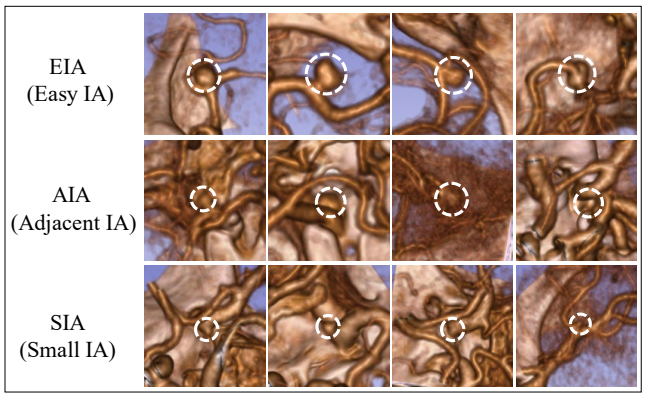

Figure 1: Some examples of IAs, with the dotted white circles indicating the locations.

in certain information loss. In addition, they can only detect the position of IAs, and cannot perform voxel-wise segmentation to obtain the complete 3D shapes of IAs.

Some methods were proposed to perform segmentation of IAs. Shi et al. [Shi *et al.*, 2020] developed a 3D CNN called DAResUNet for the segmentation of IA from bone-free CTA images. The network features an encoder-decoder architecture with residual blocks for stable training, dilated convolutions for an enlarged receptive field, and a dual attention module [Fu *et al.*, 2019] to capture long-range contextual information. Input volumes of size $80 \times 80 \times 80$ were sampled during training. For inference, predictions from uniformly sampled patches were merged. Unfortunately, due to the use of bone-free CTA images as input, some IAs near the bone were incorrectly removed, leading to missed detections. Bo et al. [Bo *et al.*, 2021] emphasized the importance of global information for the recognition of IAs and proposed a two-pathway deep learning framework, called GLIA. It employs a global positioning network to extract global features that capture the distribution of IAs, which are then combined with local image patches for voxel-wise segmentation through a modified U-Net [Çiçek *et al.*, 2016] architecture. However, this method has two key clinical limitations: (1) computational inefficiency (slow speed, high memory usage); (2) reliance on full head scans, failing with partial mid-cranial (orbital-nasal level) data. These limitations may affect its practical utility in routine clinical settings.

Both existing IA methods use generic 3D U-Net architectures, ignoring complex anatomical context of IAs. This architectural limitation leads to suboptimal performance in two clinically significant cases: (1) small intracranial aneurysms (SIAs) and (2) aneurysms adjacent to bony structures or other tissues (AIAs), as defined in our study.

Currently, due to medical ethics reasons, there are very few open CTA datasets for IA detection and segmentation. As a result, most studies rely on proprietary datasets, leading to poor consistency in algorithm design and comparison. Yang et al. [Yang *et al.*, 2020] introduced an open-access 3D intracranial aneurysm dataset called IntrA. This dataset provides 3D point cloud data of healthy vascular segments and aneurysm segments, meticulously annotated by experts. Several studies [Liu *et al.*, 2023; Estrella-Ibarra *et al.*, 2024] have conducted model innovation and performance evaluation on this dataset. For example, Cao et al. [Cao *et al.*, 2025] proposed a dual-branch fusion network called PMMNet for IA classification and segmentation. It utilizes one branch to extract spatial features from 3D point clouds and another to gather pixel features from multi-view images, ultimately fusing these features for improved performance. However, these point cloud-based methods are more suitable for segmenting vascular segments where aneurysms have already been detected, rather than for detecting aneurysms throughout the entire intracranial CTA image. Meanwhile, because point clouds are not suitable for representing SIA and AIA, these methods are also not proficient in segmenting these IA.

To address the limitations of previous methods, we propose Multi-View Refinement Network (MVRNet) based on feature-enhanced 3D CTA images. To mitigate the strong interference caused by bone in IA detection, prior methods utilized bone-free CTA images. However, our experimental observations revealed that the bone removal process can easily lead to the erroneous elimination of IAs that are close to the bone, resulting in severe consequences. Therefore, we propose to perform feature enhancement on the original 3D CTA images to obtain boundary-clear CTAs, making boundaries between IAs and surrounding tissues clearer and facilitating detection and segmentation. In light of poor prediction performance of existing methods for SIA and AIA, we design a multi-view encoder and a progressively-refined decoder. The multi-view encoder extracts fine features from each slice using 2D convolutions from three dimensions and then employs 3D convolutions for global context modeling, enabling model to process information at multiple levels. The progressively-refined decoder gradually refines uncertain areas using masks to achieve more accurate segmentation results.

Our contributions can be summarised as follows:

- We propose a framework called MVRNet that excels at solving the most challenging scenarios in intracranial aneurysm detection and segmentation. The multi-view encoder fuses specific slice features and global structural features to enhance the model's understanding of overall spatial information. The progressively-refined decoder gradually refines uncertain regions, better handling complex backgrounds.
- We introduce a CTA feature enhancement technique to compute boundary-clear CTA, which are then combined with bone-free CTA as inputs for the model. This approach reduces interference from unrelated tissues, making IAs more visible and facilitating the representation learning of IAs.
- MVRNet greatly outperforms the state-of-the-art methods with a substantial performance improvement on the total test set, and its performance on the challenging AIA and SIA test sets exceeds that of prior methods by more than double.

2 Methods

An overview of the proposed MVRNet is shown in Fig. 2. Since U-Net has been well investigated in the literature to

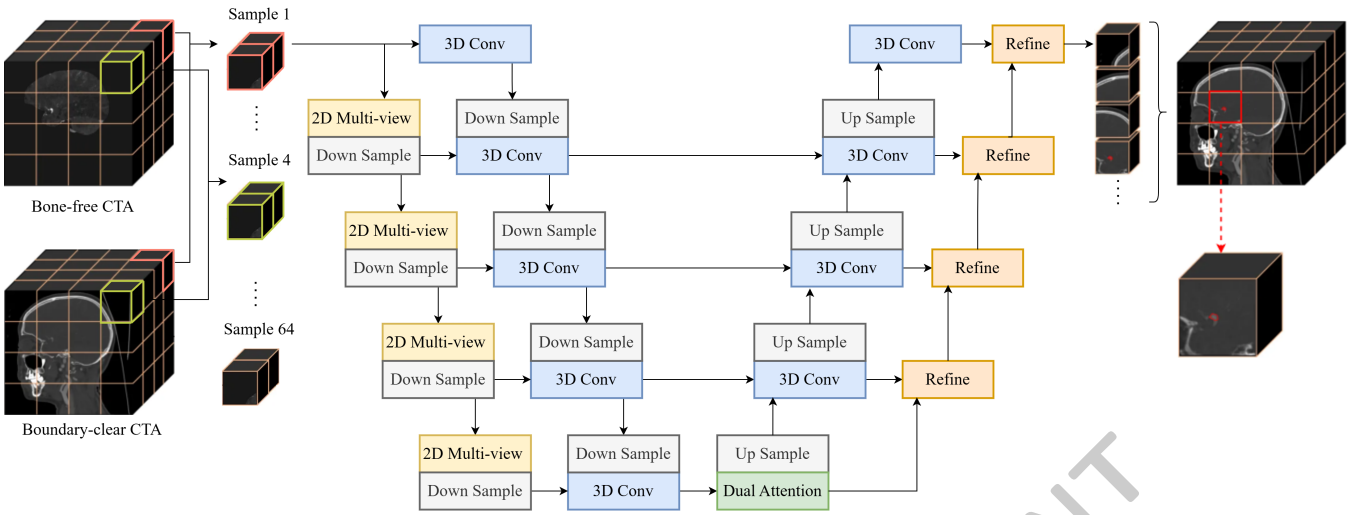

Figure 2: An overview of MVRNet. It has three novel designs, i.e., feature enhancement on the input, multi-view information powered encoder, and progressively-refined decoder.

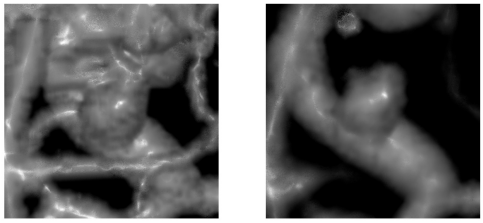

Figure 3: Left: An example of adjacent IA (AIA) that is connected to or occluded by other tissues, making it difficult to identify. Right: After our feature enhancement, it is clearly distinguished from the surrounding complex structures.

have excellent performance on image segmentation, we employ it as the backbone of MVRNet. However, the standard U-Net for 3D data, which usually consists of 3D convolution blocks and down/up-sampling, as indicated by light blue in Fig. 2, does not perform well on segmenting a relatively small, irregularly-shaped IA from the 3D CTA images, and even worse on difficult cases where IA is very small or obstructed by surrounding tissues. We address these issues by three novel modifications. First, a feature enhancement process is designed to obtain boundary-clear CTA. Then, a 2D multi-view block is developed and plugged into the encoder at different levels to augment the feature extraction by slicing the CTA from different directions, which is able to mitigate the occlusion effect from other tissues. Finally, the decoder is progressively refined by an automatically generated mask to guide the detection and segmentation on the uncertain areas.

2.1 Feature enhancement to obtain boundary-clear IA

Some IAs are difficult even for doctors to identify, because they are in contact with bony structures or other tissues. We design a feature enhancement technique that clarifies the boundaries between IA and surrounding tissues, resulting in boundary-clear CTA, as illustrated in Fig. 3. After feature enhancement, the IAs become exposed to the representation learning and thus accurately recognized.

Specifically, the feature enhancement process is consecutively implemented by sharpening, closing operation, and opening operation. The sharpening step detects boundary information, such as bone edges and aneurysm boundaries. By setting the voxel values of the detected boundaries to zero, the aneurysms are separated from the tissues they contact. Next, a closing operation, which involves dilation followed by erosion, fills small holes in the image and smooth the edges, thereby improving the overall image quality. Finally, an opening operation, which consists of erosion followed by dilation, removes small noise and artifacts. Between the erosion and dilation of the opening operation, we introduce a denoising process to eliminate small regions with a volume below a certain threshold, further reducing the impact of noise and thus enhancing the reliability of IA recognition.

The boundary-clear CTA offers dual benefits: (1) effectively mitigating surrounding tissue interference for adjacent IAs (AIAs), and (2) significantly enhancing localization accuracy for small IAs (SIAs). In CTA imaging, the high signal intensity of osseous structures often obscures small lesions. Our boundary-clear CTA addresses this limitation by suppressing bone-induced signal interference, thereby creating a more favorable environment for SIAs identification.

In this work, in addition to the boundary-clear CTA, the bone-free CTA obtained by removing the bone structure from the original CTA is also used as the input of MVRNet. In the related literature, the bone-free CTA are usually taken as the sole input. Our experiments show that the bone removing process may mistakenly eliminate the IAs that are close to the bone. Here, we combine bone-free CTA with boundary-clear CTA as the input x of model, which significantly decreases false negative rates.

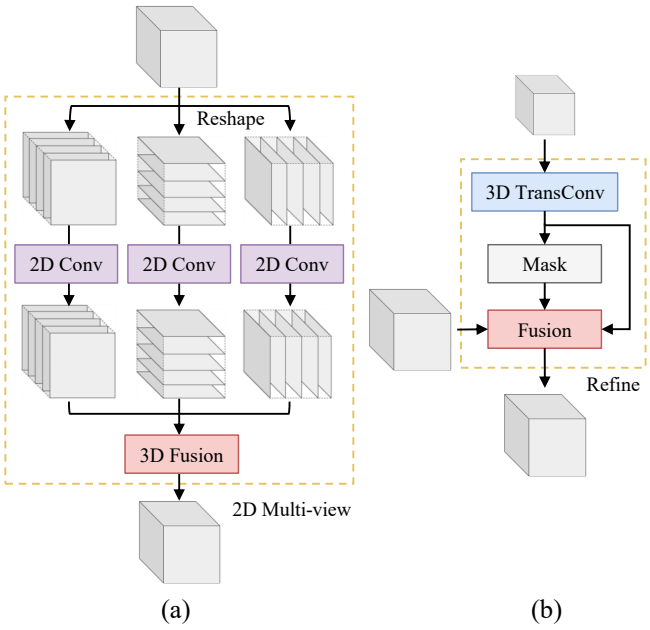

Figure 4: (a) Details of the multi-view block. For a 3D representation, features are extracted using 2D convolution along three orthogonal planes and fused into a new 3D representation by concatenating and applying 3D convolution for integration. (b) Details of the refinement block. A mask is generated from the output of the previous block as guidance to gradually refine the uncertain areas.

2.2 Multi-view encoder

Although the standard encoder module of 3D U-Net is effective in 3D data representation learning, it becomes powerless in IA recognition when the IAs are very small or surrounded by dense blood vessels. The 3D convolution is able to capture the global spatial information, but it may lead to the loss of some local detail information, because it typically treat the entire volume as a whole. This is particularly true for the edges and fine structures of IA, where 3D convolutions may not adequately capture these critical features. Thus, the results on IA detection and segmentation are compromised.

To overcome the above limitations, we develop a multi-view encoder to improve representation learning on fine structures and contextual information of small IAs. As illustrated in Fig. 4 (a), we design a multi-view 2D convolutional architecture that processes 3D CTA images through triplanar slicing. The 2D convolutions are able to effectively capture more local details by focusing on the current slice, and multi-view operation is capable of suppressing the interference from surrounding tissues.

Our methodology innovatively combines 2D and 3D convolutional advantages through a hybrid architecture: (1) Multi-view 2D convolutions first extract local features by processing triple orthogonal planes of 3D space, demonstrating superior sensitivity to fine anatomical details; (2) Subsequent 3D convolutions then perform spatial feature fusion to establish global contextual relationships. Crucially, the multi-view 2D operations not only compensate for 3D convolution's computational inefficiency but, more importantly, overcome its inherent limitations in characterizing complex anatomical structures through cross-planar complementary features.

2.3 Progressively-refined decoder

To further improve the prediction results, we develop a progressive refinement process hierarchically along with the standard decoder blocks, as in Fig. 2. The output from each 3D convolution block, i.e., $\widetilde{g}_i$, $i \in \{4,3,2,1\}$, is fed into a refinement block for further improvement. From bottom to top, it corresponds to g_4 to g_1. Then, we construct the refinement block as in Fig. 4 (b). The main idea is to generate a rough mask for self-guidance while progressively refining uncertain regions, which is inspired by [Yu *et al.*, 2021]. The refinement block is computed as follows:

$$r'_i = \text{3D-TransConv}(r_i), \tag{1}$$

$$\phi_{mask} = \begin{cases} 1, & \text{if } 0 < r'_i < 1, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

$$r_{i-1} = \widetilde{g}_i \times \phi_{mask} + r'_i \times (1 - \phi_{mask}), \tag{3}$$

where r_4 is the output of Dual Attention Block. In Eq. (1), the feature map r_i is received from the previous refinement block, and it is initially upscaled using a 3D transpose convolution to produce r'_i. In Eq. (2), the masking mechanism categorizes features within the range of 0 and 1 as uncertain regions, while the remaining features are designated as certain regions. The mask matrix ϕ_{mask} represents the uncertain areas by a value of 1, and the deterministic areas by 0. Subsequently, as described in Eq. (3), the mask filters the previous features to retain identified regions while also filtering the current features to isolate the uncertain areas for further refinement. It should be noted that the Eq. (2)&(3) are computed in an element-wise manner.

When IAs are small and susceptible to interference from surrounding tissues, their accurate identifications become particularly challenging. The self-guidance mechanism through the refinement block allows the model to focus on uncertain areas that may contain critical features indicative of IA. By progressively refining these regions, the model effectively differentiates between noise and relevant features, thereby improving the detection and segmentation of IAs. This iterative refinement ensures that the model remains adaptive and responsive to the subtle characteristics inherent in the imaging data, ultimately leading to enhanced performance in clinical applications.

3 Experiments

3.1 Implementation details and data description

We collected cerebrovascular CTA data from 887 patients from a hospital. The dataset used in this study includes 787 training cases and 100 testing cases. The location distribution of IAs is given in Tab. 1. Both the training dataset and total test set contain IAs widely distributed in various locations within the cranial cavity. To further evaluate the performance on difficult cases, we constructed two test subsets, one for IAs adjacent to other tissues or bones (AIA subset), and another for very small IAs (SIA Subset). To summarize, we have:

Dataset	Anterior cerebral artery	Middle cerebral artery	Posterior cerebral artery	*Anterior communicating artery*	Posterior communicating artery	*Internal carotid artery*	Basilar artery	Other locations	Total
Train Set	26	497	6	93	104	47	10	4	787
Test Set (Total)	6	50	3	17	10	11	2	1	100
Test Set (AIA Subset)	6	0	3	17	10	11	2	1	50
Test Set (SIA Subset)	2	6	1	7	3	5	1	0	25

Table 1: Statistical information on the location of IAs in the dataset. "AIA" and "SIA" represent specific challenging subsets of the total test set. IAs located in the anterior communicating artery and internal carotid artery are more difficult to detect.

Model	MVB	RB	FE	Total Test Set				AIA Test Set				SIA Test Set			
				IoU	Dice	Prec.	Rec.	IoU	Dice	Prec.	Rec.	IoU	Dice	Prec.	Rec.
Baseline				0.20	0.27	0.21	0.52	0.11	0.14	0.11	0.24	0.03	0.05	0.03	0.21
MVRNet	✓			0.29	0.38	0.34	0.52	0.19	0.24	0.26	0.29	0.05	0.07	0.07	0.18
	✓	✓		0.31	0.40	0.36	0.52	0.21	0.26	0.25	0.33	0.05	0.08	0.10	0.15
	✓	✓	✓	**0.36**	**0.44**	**0.43**	**0.56**	**0.27**	**0.35**	**0.36**	**0.42**	**0.10**	**0.14**	**0.23**	**0.22**

Table 2: Experimental results of MVRNet on the total, AIA, and SIA test sets. The terms "Prec." and "Rec." are abbreviations for "Precision" and "Recall", respectively. The evaluation metrics with higher values indicate better performance.

- **The total test set**: It contains all 100 testing cases.
- **The AIA test set**: There are 50 cases of Adjacent IAs (AIAs), typically located in the anterior communicating artery and internal carotid artery. These IAs, being adjacent to osseous structures or embedded within dense vascular networks, are particularly vulnerable to surrounding tissue interference during detection procedures.
- **The SIA test set**: This set collects 25 cases of Small IAs (SIAs), with sizes smaller than 4 mm.

Both AIA set and SIA set are subsets of total test set, and the IAs of two sets are extremely difficult to detect and segment.

3.2 Ablation Study

To verify the effectiveness of the proposed three key components, i.e., feature enhancement (FE), multi-view block (MVB), and refinement block (RB), we use DAResUNet [Shi *et al.*, 2020] as the baseline and add the three components one by one into the baseline, and test the performances of intermediate models. The input to the baseline model is bone-free CTA images. The results are reported in Tab. 2.

Results on the total test set

The experimental results on the total test set are reported in Tab. 2 (left). After including the multi-view block, the IoU, Dice, and Precision scores are improved by 9%, 11%, and 13%, respectively, due to the detailed feature extraction provided by 2D convolutions on each slice. Further adding the refinement block also yielded improvements, indicating that the refinement block is able to further focus on key features and optimize prediction results based on the multi-view block. Notably, after employing feature enhancement on the CTA data, not only did IoU, Dice, and Precision continue to improve, but the Recall is also increased by 4%, suggesting that feature enhancement reduces the rates of false negative detections. Therefore, based on the high-quality input provided by feature enhancement, and the mutual benefits of multi-view block and refinement block, MVRNet achieved significant improvements compared to the baseline, while only incurring a small increase in the number of parameters, i.e., from 19.03M to 20.62M.

Results on the AIA test set

Since AIAs are often located close to bones or other tissues, it is difficult even for human doctors to identify and segment them from the surroundings. Moreover, they may be erroneously removed during the bone removal process which is usually adopted in the literature. Therefore, addressing this challenge is of great help in the clinical recognition of IAs. Tab. 2 (middle) reports the experimental results on the AIA test set. It is observed that due to its difficulty, the baseline performance declines greatly to a low level, only about half of the metric scores on the total test set. In such a challenging scenario, the multi-view block and refinement block effectively enhance the model's performance, restoring the scores of IoU, Dice, and Precision back to the baseline performance on the total test set. The Recall score is further largely improved by feature enhancement, indicating that feature enhancement plays a critical role in handling challenging cases that are susceptible to interference from surrounding tissues. The substantial improvements in terms of all metrics demonstrate the superiority of MVRNet in difficult scenarios.

Results on SIA test set

Although clinical practice typically prioritizes the treatment of unruptured IAs larger than 7 mm in diameter, small aneurysms may also lead to subarachnoid hemorrhage. The experimental results on the SIA test set are given in Tab. 2 (right). The baseline performance is extremely low, because small IAs are very difficult to detect. The multi-view block and refinement block improve the performance by strength-

Model	Total Test Set				AIA Test Set				SIA Test Set			
	Prec.	Rec.	F1	FP/c	Prec.	Rec.	F1	FP/c	Prec.	Rec.	F1	FP/c
Ueda et al. [Ueda *et al.*, 2019]	0.30	0.43	0.34	1.03	0.13	0.16	0.14	0.98	0.05	0.04	0.04	1.12
Dai et al. [Dai *et al.*, 2020]	0.31	0.46	0.36	1.01	0.12	0.14	0.12	1.10	0.12	0.22	0.15	1.04
GLIA	0.29	0.44	0.32	1.79	0.06	0.10	0.07	2.04	0.05	0.14	0.07	2.08
DAResUNet	0.31	0.58	0.38	1.46	0.14	0.25	0.17	1.48	0.10	0.26	0.14	1.48
MVRNet	**0.58**	**0.71**	**0.62**	**0.43**	**0.48**	**0.53**	**0.49**	**0.40**	**0.34**	**0.42**	**0.36**	**0.44**

Table 3: Detection results on the total, AIA, and SIA test sets. Precision (Prec.), Recall (Rec.), and F1-score (F1) are better when larger, while False Positives per case (FP/c) is better when smaller.

Model	Total Test Set				AIA Test Set				SIA Test Set			
	IoU	Dice	Prec.	Rec.	IoU	Dice	Prec.	Rec.	IoU	Dice	Prec.	Rec.
U-Net [Ueda *et al.*, 2019]	0.12	0.17	0.14	0.38	0.06	0.08	0.07	0.13	0.02	0.03	0.02	0.09
U-Net [Dai *et al.*, 2020]	0.14	0.20	0.15	0.40	0.06	0.08	0.07	0.14	0.03	0.04	0.03	0.19
GLIA	0.22	0.27	0.28	0.30	0.05	0.06	0.08	0.05	0.06	0.07	0.07	0.09
DAResUNet	0.20	0.27	0.21	0.53	0.11	0.14	0.11	0.24	0.03	0.05	0.03	0.21
MVRNet	**0.36**	**0.44**	**0.43**	**0.56**	**0.27**	**0.35**	**0.36**	**0.42**	**0.10**	**0.14**	**0.23**	**0.22**

Table 4: Segmentation results on the total, AIA and SIA test set. IoU, Dice, Precision and Recall are better when larger.

ening the model's ability to extract detailed features from pure 3D convolutions. Meanwhile, feature enhancement once again proves its effectiveness on challenging cases.

Feature map visualization

We visualized the feature maps of a representative case, as shown in Fig. 5, where the bone removal process incorrectly interfered with IA, causing its voxels to fade to nearly imperceptible levels, which led to DAResUNet's failure in detecting it. Surprisingly, MVRNet is still able to recognize the IA to some extent with the same bone-free CTA input. The boundary-clear CTA compensates for the missing information from bone-free CTA and offers more distinguishable features. Therefore, MVRNet achieves excellent results when using feature-enhanced CTA as input.

3.3 Comparison with SOTA Methods in the Field

We conducted comprehensive comparisons between MVRNet and dedicated SOTA (state-of-the-art) models for IA detection or segmentation, benchmarking their performance on both task scenarios. The methods proposed by Ueda et al. and Dai et al. were originally developed for aneurysm detection based on ResNet-18 and Faster R-CNN [Ren *et al.*, 2017], respectively. We trained U-Net models for segmentation based on their detection results, respectively, for comparison on the segmentation task. GLIA, DAResUNet, and our MVRNet can perform both detection and segmentation. The results are reported in Tab. 3&4 in terms of *Precision, Recall, F1 score*, and *FP/case* (False Positives per case).

Detection

The detection results on the total test set in Tab. 3 (left) indicate that MVRNet outperforms the other methods by an obviously large increment, while [Ueda *et al.*, 2019], [Dai *et al.*, 2020], GLIA, and DAResUNet are comparably close. When these methods are applied on the challenging AIA and SIA test sets, their performance drops dramatically in Tab. 3 (middle, right). Our MVRNet is still the best, and maintains the performance at a reasonably good level, which is about 200% to 300% increment against the existing methods in terms of Precision, Recall, F1 scores. This indicates that our method is robust to small IA sizes and skull stripping procedures that may incorrectly resect IAs, and effectively mitigates the impact of complex background structures in CTA images, which usually severely impairs existing methods.

Segmentation

We go further to segment IAs from the 3D CTA images in addition to detecting them. Segmentation requires to obtain the exact locations and shapes of IAs, more difficult than detection. The segmentation results on the total test set are reported in Tab. 4 (left). GLIA and DAResUNet outperforms [Ueda *et al.*, 2019] and [Dai *et al.*, 2020], demonstrating that an end-to-end model for both detection and segmentation tasks is superior to a two-step approach. MVRNet outperforms GLIA and DAResUNet by an over 60% increment in terms of IoU and Dice, by more than half in terms of Precision, while maintaining Recall at an excellent level. When applied on the AIA test set and SIA test set, as in Tab. 4 (middle, right), the IoU, Dice, and Precision of the previous methods mostly dropped below 10%, while MVRNet still performed well.

3.4 Comparisons with 3D Segmentation Baselines

In addition to comparing with IA-specific detection and segmentation methods, we also evaluated against SOTA 3D segmentation baselines, including nnFormer [Zhou *et al.*, 2021] and MedNeXt [Roy *et al.*, 2023], on the complete test set, as presented in Tab. 5. Comprehensive evaluation demonstrated MVRNet's superior performance in both IA detection and segmentation tasks. For detection, MVRNet achieved SOTA

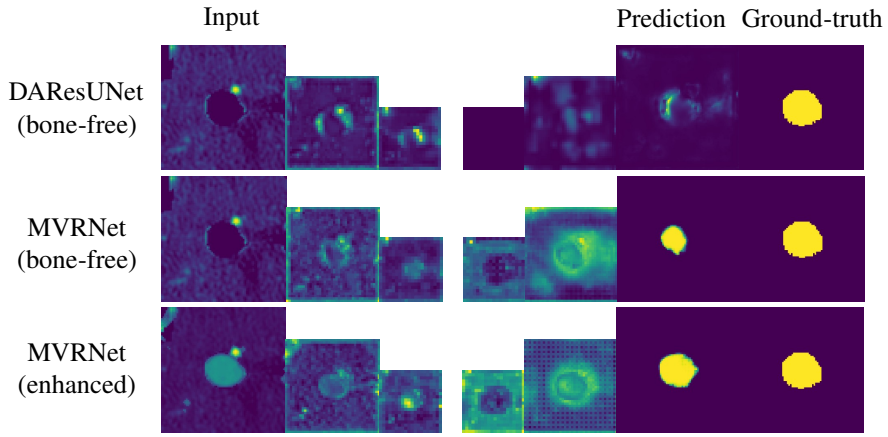

Figure 5: Visualization of feature maps (resolutions $> 10 \times 10$ shown only).

Model	Detection				Segmentation			
	Prec.	Rec.	F1	FP/c	IoU	Dice	Prec.	Rec.
nnFormer	0.14	0.49	0.20	3.88	0.16	0.22	0.17	0.40
MedNeXt	0.18	0.57	0.25	3.54	0.19	0.26	0.21	0.45
MVRNet	**0.58**	**0.71**	**0.62**	**0.43**	**0.36**	**0.45**	**0.43**	**0.56**

Table 5: Comparisons with 3D segmentation baselines on the complete test set for the segmentation task.

Model	Detection				Segmentation			
	Prec.	Rec.	F1	FP/c	IoU	Dice	Prec.	Rec.
GLIA	0.16	0.56	0.19	10.66	0.06	0.10	0.10	0.36
DAResUNet	0.22	0.56	0.26	3.70	0.08	0.12	0.10	0.38
MVRNet	**0.29**	**0.60**	**0.36**	**1.53**	**0.16**	**0.24**	**0.22**	**0.41**

Table 6: Models' performance on the OpenNeuro dataset.

Model	Detection				Segmentation			
	Prec.	Rec.	F1	FP/c	IoU	Dice	Prec.	Rec.
GLIA	0.10	0.32	0.14	6.70	0.04	0.07	0.12	0.07
DAResUNet	0.22	0.53	0.26	3.30	0.08	0.13	0.12	0.36
MVRNet	**0.28**	**0.59**	**0.36**	**2.10**	**0.12**	**0.20**	**0.17**	**0.37**

Table 7: Models' performance on the ADAM dataset.

performance with an F1-score of 0.62 (vs. 0.20-0.25 for baselines) while maintaining a low false positive rate (0.43/case). In segmentation, it significantly outperformed existing methods with IoU of 0.36 (vs. 0.16-0.19 for baselines). The results highlight that conventional 3D segmentation models (nnFormer, MedNeXt) are suboptimal for IA analysis, underscoring the need for specialized architectures like MVRNet.

3.5 Evaluation on Public Dataset

Due to the scarcity of publicly available IA datasets, our study utilized two limited-scale public datasets for experimental validation: OpenNeuro [Di Noto *et al.*, 2023] and ADAM. We conducted comparisons with specialized IA models on these two test sets. The result is presented in Tab. 6&7. Because the global localization network of GLIA is highly dependent on full-head scan data, it exhibits obvious adaptability limitations for datasets that only contain a partial scan range of the mid-cranial region (e.g. OpenNeuro and Adam datasets), resulting in a significant decline in model performance. MVRNet surpasses related IA methods in all indicators on both testsets. These results highlight the effectiveness and generalizability of MVRNet in handling cross-domain public datasets. In particular, the average diameter of IAs in ADAM is about 4 mm, making them clinically challenging small targets. However, our method effectively alleviates the problem of missed detection of small targets.

4 Conclusion

In this paper, we propose MVRNet, a novel framework designed to address the challenges of IA detection and segmentation in 3D CTA image data. The proposed method particularly focuses on the extremely challenging scenarios where the aneurysms are very small, or surrounded by other tissues, or too close to the bone with the risk of being removed with the bone. This work delivers clinically valuable solutions to these real-world diagnostic difficulties, where improved performance directly benefits patient care. MVRNet has three key features to fulfill this challenging goal. First, we present a feature enhancement technique to obtain boundary-clear CTA data, which are fused with the bone-free CTA data as input, to further reduce the impact of noise and enhance the reliability of IA recognition. Second, we design a multi-view encoder by slicing the 3D CTA data from various directions to capture more details around the aneurysms. Moreover, we develop a progressively-refined decoder, which gradually refines uncertain regions to better handle complex backgrounds. Experiments on a collected CTA dataset of 887 patients demonstrate that the proposed MVRNet outperforms the existing SOTA methods by a large increment consistently on both tasks of detection and segmentation, and MVRNet is particularly robust on those extremely difficult cases.

Acknowledgments

This works was supported by Medical-Industrial Crossover Research Fund of Shanghai Jiao Tong University "Jiao Tong University Star" Program (No. YG2025ZD02) and the Fundamental Research Funds for the Central Universities (project number YG2026QNB01).

References

[Bo *et al.*, 2021] Zi-Hao Bo, Hui Qiao, Chong Tian, Yuchen Guo, Wuchao Li, Tiantian Liang, Dongxue Li, Dan Liao, Xianchun Zeng, Leilei Mei, et al. Toward human intervention-free clinical diagnosis of intracranial aneurysm via deep neural network. *Patterns*, 2(2), 2021.

[Cao *et al.*, 2025] Ruifen Cao, Dongwei Zhang, Pijing Wei, Yun Ding, Chunhou Zheng, Dayu Tan, and Chao Zhou. Pmmnet: A dual branch fusion network of point cloud and multi-view for intracranial aneurysm classification and segmentation. *IEEE Journal of Biomedical and Health Informatics*, 29(5):3137–3147, 2025.

[Çiçek *et al.*, 2016] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[Dai *et al.*, 2020] Xilei Dai, Lixiang Huang, Yi Qian, Shuang Xia, Winston Chong, Junjie Liu, Antonio Di Ieva, Xiaoxi Hou, and Chubin Ou. Deep learning for automated cerebral aneurysm detection on computed tomography images. *International Journal of Computer Assisted Radiology and Surgery*, 15:715–723, 2020.

[Di Noto *et al.*, 2023] Tommaso Di Noto, Guillaume Marie, Sebastien Tourbier, Yasser Alemán-Gómez, Oscar Esteban, Guillaume Saliou, Meritxell Bach Cuadra, Patric Hagmann, and Jonas Richiardi. Towards automated brain aneurysm detection in tof-mra: open data, weak labels, and anatomical knowledge. *Neuroinformatics*, 21(1):21–34, 2023.

[Estrella-Ibarra *et al.*, 2024] Luis Felipe Estrella-Ibarra, Alejandro de León-Cuevas, and Saul Tovar-Arriaga. Nested contrastive boundary learning: Point transformer self-attention regularization for 3d intracranial aneurysm segmentation. *Technologies*, 12(3), 2024.

[Fu *et al.*, 2019] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154, 2019.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Irfan *et al.*, 2023] Muhammad Irfan, Khalid Mahmood Malik, Jamil Ahmad, and Ghaus Malik. Strokenet: An automated approach for segmentation and rupture risk prediction of intracranial aneurysm. *Computerized Medical Imaging and Graphics*, 108:102271, 2023.

[Kato *et al.*, 1999] Yoko Kato, Hirotoshi Sano, Kazuhiro Katada, Yuko Ogura, Motoharu Hayakawa, Narimasu Kanaoka, and Tetsuo Kanno. Application of three-dimensional ct angiography (3d-cta) to cerebral aneurysms. *Surgical neurology*, 52(2):113–122, 1999.

[Li *et al.*, 2023] Peiying Li, Yongchang Liu, Jiafeng Zhou, Shikui Tu, Bing Zhao, Jieqing Wan, Yunjun Yang, and Lei Xu. A deep-learning method for the end-to-end prediction of intracranial aneurysm rupture risk. *Patterns*, 4(4), 2023.

[Liu *et al.*, 2023] Yifan Liu, Wuyang Li, Jie Liu, Hui Chen, and Yixuan Yuan. Grab-net: Graph-based boundary-aware network for medical point cloud segmentation. *IEEE Transactions on Medical Imaging*, 42(9):2776–2786, 2023.

[Ren *et al.*, 2017] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

[Rikhtegar *et al.*, 2021] Reza Rikhtegar, Pascal J Mosimann, Jan Rothaupt, Mohammad Mirza-Aghazadeh-Attari, Shahin Hallaj, Mehdi Yousefi, Atefeh Amiri, Ebrahim Farashi, Atefeh Kheyrollahiyan, and Sanam Dolati. Non-coding rnas role in intracranial aneurysm: general principles with focus on inflammation. *Life sciences*, 278:119617, 2021.

[Roy *et al.*, 2023] Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F Jaeger, and Klaus H Maier-Hein. Mednext: transformer-driven scaling of convnets for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 405–415. Springer, 2023.

[Shi *et al.*, 2020] Zhao Shi, Chongchang Miao, U Joseph Schoepf, Rock H Savage, Danielle M Dargis, Chengwei Pan, Xue Chai, Xiuli Li, Shuang Xia, Xin Zhang, Yan Gu, Yonggang Zhang, Bin Hu, Wenda Xu, Changsheng Zhou, Song Luo, Hao Wang, Li Mao, Kongming Liang, Lili Wen, Longjiang Zhou, Yizhou Yu, Guangming Lu, and Longjiang Zhang. A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. *Nature communications*, 11(1):6090, 2020.

[Ueda *et al.*, 2019] Daiju Ueda, Akira Yamamoto, Masataka Nishimori, Taro Shimono, Satoshi Doishita, Akitoshi Shimazaki, Yutaka Katayama, Shinya Fukumoto, Antoine Choppin, Yuki Shimahara, et al. Deep learning for mr angiography: automated detection of cerebral aneurysms. *Radiology*, 290(1):187–194, 2019.

[Yang *et al.*, 2020] Xi Yang, Ding Xia, Taichi Kin, and Takeo Igarashi. Intra: 3d intracranial aneurysm dataset for deep learning. In *Proceedings of the IEEE/CVF Confer-*

ence on Computer Vision and Pattern Recognition, pages 2656–2666, 2020.

[Yu *et al.*, 2021] Qihang Yu, Jianming Zhang, He Zhang, Yilin Wang, Zhe Lin, Ning Xu, Yutong Bai, and Alan Yuille. Mask guided matting via progressive refinement network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1154–1163, 2021.

[Zhou *et al.*, 2021] Hong-Yu Zhou, Jiansen Guo, Yinghao Zhang, Lequan Yu, Liansheng Wang, and Yizhou Yu. nnformer: Interleaved transformer for volumetric segmentation. *arXiv preprint arXiv:2109.03201*, 2021.