DCKAN: A Dual-Coordinate KAN Framework for Fibrous Cap Segmentation on Carotid OCT

Tonghua Wan[1,2], Sihan Liu[1,2], Yuxin Cai[1,2], Shengcai Chen[3], Yan Wan[3], Bo Hu[3], and Wu Qiu[1,2⋆]

[1] College of Life Science and Technology, Huazhong Science of Science and Technology, Hubei, China
[2] Advanced Biomedical Imaging Facility, Hubei, China
[3] Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

Abstract. Fibrous cap thickness is a key clinical marker for assessing carotid plaque vulnerability. While intravascular optical coherence tomography (OCT) enables in vivo visualization of fibrous caps, its design for coronary arteries poses challenges in carotid imaging, such as larger vessel size, faster blood flow, limited penetration, and restricted imaging range, leading to incomplete visualization and poor image quality. To address these limitations, we propose a dual-coordinate segmentation framework for carotid OCT fibrous cap segmentation. This framework integrates Cartesian images, which preserve global spatial context, with linear-polar transformed images, effectively representing the annular geometry of fibrous caps. The fusion of dual-coordinate features mitigates incomplete vascular walls and blood artifacts, enhancing segmentation accuracy and robustness. We introduce a Cross-Coordinate Feature Fusion Module (CCFFM) to efficiently integrate features from both coordinate systems and reduce interference from redundant information. Additionally, the Kolmogorov-Arnold Network (KAN) block is incorporated to extract complex nonlinear features while improving model interpretability. Our method achieves state-of-the-art performance on an external carotid OCT dataset, demonstrating the potential of OCT for advancing carotid imaging and improving plaque vulnerability assessment.

Keywords: Intravascular Optical Coherence Tomography · Fibrous Cap · Dual-Coordinate System · KAN · Medical Image Segmentation

1 Introduction

Fibrous cap thickness is a critical determinant of plaque stability, with thinner caps being significantly more prone to rupture than thicker ones [1]. Plaque rupture often triggers thrombosis, emphasizing the importance of accurately assessing fibrous cap morphology and thickness for evaluating plaque vulnerability.

⋆ Corresponding Author: wuqiu@hust.edu.cn(Wu Qiu)

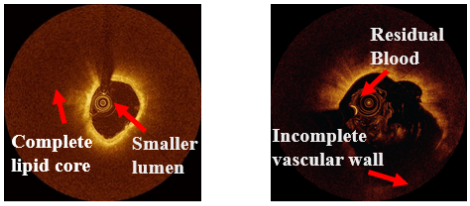

Fig. 1. Visual comparison of Coronary and Carotid Artery OCT Imaging.

Intravascular optical coherence tomography (OCT) is currently the only imaging modality capable of visualizing fibrous cap boundaries and morphology in vivo. However, its application has been primarily limited to coronary arteries, with limited use in carotid arteries, due to the unique anatomical and physiological challenges posed by the latter. As shown in Fig. 1, carotid arteries have a larger lumen diameter and thicker vessel walls compared to coronary arteries. These anatomical differences, coupled with the shallow penetration depth and limited imaging range of intravascular OCT, often result in incomplete visualization of the vessel wall and lipid core, complicating accurate fibrous cap segmentation. Additionally, the higher blood flow in carotid arteries hinders effective flushing of residual blood with contrast agents, causing artifacts that obscure fibrous caps and degrade image quality.

Existing fibrous cap segmentation algorithms are predominantly designed for coronary artery OCT and exhibit significant limitations when applied to carotid artery images. Traditional methods, such as dynamic programming-based approaches [16][15], rely heavily on handcrafted features, which hinder generalization and make the methods vulnerable to noise. Although deep learning-based techniques have shown promise, they typically rely on single-coordinate domain inputs, neglecting the complementary information offered by dual-coordinate representations [7][8]. As a result, these existing algorithms are not directly adaptable to carotid fibrous cap segmentation due to the distinct characteristics and challenges of carotid artery OCT.

This study introduces a dual-coordinate segmentation framework specifically designed for fibrous cap analysis in carotid artery OCT imaging. The framework combines Cartesian and polar coordinate representations through a cross-coordinate integration mechanism, effectively addressing the inherent challenges of vascular tissue characterization in OCT. Our contributions include 1) We establish a dedicated segmentation architecture for fibrous cap quantification in carotid OCT. 2) A novel cross-coordinate feature fusion module aggregates multi-scale spatial information from orthogonal coordinate systems, enhancing boundary delineation accuracy. 3) Integration of Kolmogorov-Arnold Network components strengthens nonlinear feature representation while providing mathematical interpretability, improving segmentation robustness against vascular morphological variations.

2 Methodology

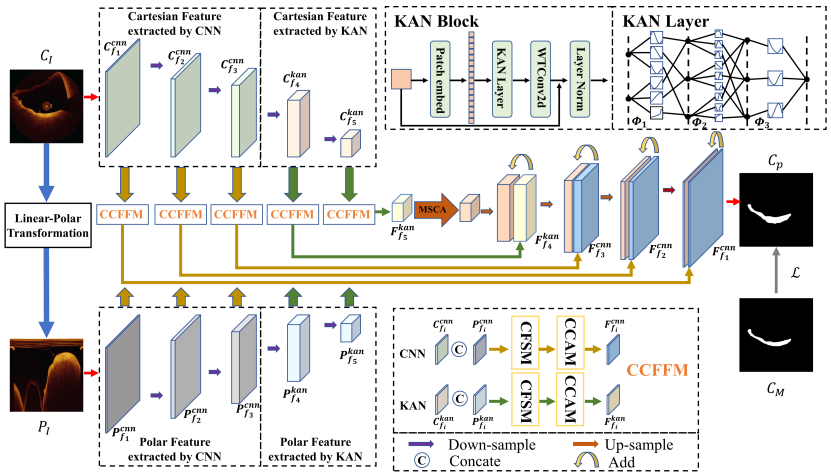

Fig. 2. Overview of our proposed dual-coordinate system feature fusion OCT Segmentation framework

2.1 Overall structure

Given an OCT fibrous cap dataset $(C_I, C_M)_{i=1}^{N}$, where $C_I \in \mathbb{R}^{H\times W\times 3}$ and $C_M \in \mathbb{R}^{H\times W\times 1}$ denote the input images and their corresponding segmentation labels in the Cartesian coordinate system, respectively, the Cartesian images C_I are first transformed into their polar representations $P_I \in \mathbb{R}^{H\times W\times 3}$ using the linear-polar transformation, defined as:

$$\rho = \sqrt{(x-x_c)^2 + (y-y_c)^2}, \quad \theta = \arctan 2\left(\frac{y-y_c}{x-x_c}\right) \tag{1}$$

where (x_c, y_c) represents the center of the Cartesian image. The original Cartesian images C_I and the transformed polar coordinate images P_I are jointly fed into the proposed network.

The network employs a U-shaped architecture comprising a single encoder and a decoder with skip connections. The encoder includes three convolutional blocks followed by two KAN blocks to extract features from different coordinate systems. At each layer, the Cartesian and Polar features are fused using the Cross-Coordinate Feature Fusion Module(CCFFM) to filter and integrate complementary information. The fused features are propagated through skip connections to the corresponding layers in the decoder. Additionally, at the bottleneck layer, a Multi-scale Cross-axis Attention (MSCA)[13] is introduced to

extract and integrate features across multiple scales, enhancing the network's capacity to capture fine-grained and global information. The framework of the proposed method is illustrated in Fig. 2.

2.2 KAN Block

The Kolmogorov-Arnold Network (KAN) [10], inspired by the Kolmogorov-Arnold representation theorem [6], overcomes the limitations of traditional MLPs, such as low parameter efficiency and limited interpretability. By replacing conventional linear weight matrices with learnable, parameterized activation functions, KAN enables a more compact model architecture while maintaining superior performance. Additionally, this approach enhances the model's interpretability, making it particularly suitable for capturing and modeling complex features.

In our framework, the KAN block tokenizes the input feature map $X_L \in \mathbb{R}^{H_L \times W_L \times C_L}$ into $P \times P$ patches, producing a sequence of flattened patches $\{X_L^i \in \mathbb{R}^{P^2 \cdot C_L} | i = 1, \ldots, N\}$, where $N = \frac{H_L \times W_L}{P^2}$ represents the total number of patches. These patches are then projected into a latent D-dimensional embedding space using a trainable linear projection $E \in \mathbb{R}^{(P^2 \cdot C_L) \times D}$, implemented as a convolutional layer with a kernel size of 3. The tokenized embedding is expressed as:

$$Z_0 = [X_L^1 E; X_L^2 E; \cdots; X_L^N E] \tag{2}$$

where $Z_0 \in \mathbb{R}^{N \times D}$ denotes the initial sequence of tokens.

After tokenization, the features are processed through the KAN layer, which consists of three sequential operations, corresponding to a depth of three. To capture both high-frequency details (e.g., edges and textures) and low-frequency components (e.g., shapes and structures), we incorporate wavelet transform convolution [2]. This approach effectively addresses the diverse frequency characteristics of dual-coordinate inputs. At the block level, a residual connection aggregates the input tokens with the processed features to facilitate gradient flow and preserve original information. The output of the k-th KAN block is computed as:

$$Z_k = \mathrm{LN}\left(Z_{k-1} + \mathrm{WTConv2d}(\text{KAN Layer}(Z_{k-1}))\right) \tag{3}$$

where $\mathrm{LN}(\cdot)$ denotes layer normalization. Batch normalization (BN) and ReLU activation are applied after each WTConv2d operation to stabilize training and introduce non-linearity.

2.3 Cross-Coordinate Feature Fusion Module

Although dual-coordinate system images provide rich semantic features, directly integrating features from different coordinate systems presents several challenges. Representational differences between Cartesian and polar domains can lead to spatial misalignment and inconsistencies. The varying distributions of features across coordinate systems may cause imbalanced feature contributions, while redundant information can increase computational complexity and risk overfitting.

Additionally, coordinate transformations can introduce noise, further complicating effective feature fusion.

To address these issues, we propose the Cross-Coordinate Feature Fusion Module (CCFFM), as illustrated in Fig. 3, comprising the Coordinate Feature Selection Module (CFSM) and the Coordinate Channel Attention Module (CCAM). Together, these modules refine and align features, mitigate redundancies, and enhance representation, enabling robust integration of dual-coordinate features.

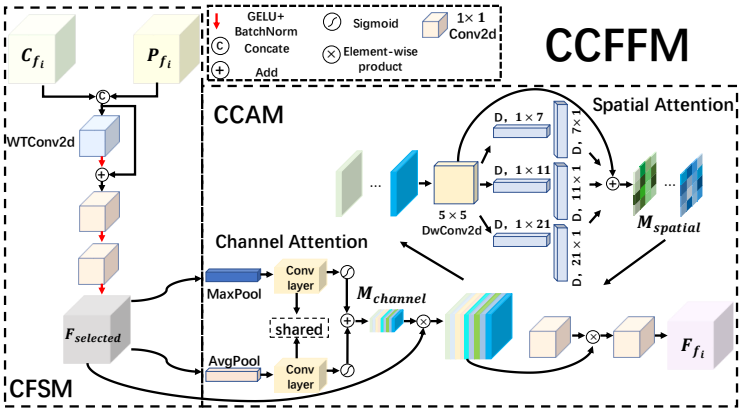

Fig. 3. Cross-Coordinate Feature Fusion Module.

The CFSM serves as the first stage of the Domain Fusion Module and is responsible for filtering and refining the features extracted from the dual-coordinate systems. At the i-th layer, the features $P_{f_i} \in \mathbb{R}^{H_i \times W_i \times C_i}$ from the Polar coordinate system and $C_{f_i} \in \mathbb{R}^{H_i \times W_i \times C_i}$ from the Cartesian coordinate system share identical dimensions. To fully exploit the complementary characteristics of these two representations, the CFSM concatenates the features along the channel dimension, producing a combined feature map $F_i \in \mathbb{R}^{H_i \times W_i \times 2C_i}$.

The concatenated feature map F_i is processed through a wavelet transform convolution (WTConv2d), followed by GELU activation and Batch Normalization to preserve high-frequency details and capture low-frequency components. To ensure consistency and mitigate feature degradation, a residual connection combines the transformed features with the original input:

$$F_{\text{residual}} = \text{BN}(\text{GELU}(\text{WTConv2d}(F_i))) + F_i \tag{4}$$

This residual design enhances the stability and effectiveness of feature refinement, ensuring robust integration of complementary features.

The residual output F_{residual} is refined by two sequential blocks, each consisting of a 1×1 convolution, GELU activation, and Batch Normalization, resulting

in the selected features:

$$F_{\text{selected}} = \text{BN}(\text{GELU}(\text{Conv}_{1\times1}(\text{BN}(\text{GELU}(\text{Conv}_{1\times1}(F_{\text{residual}})))))) \quad (5)$$

The selected features F_{selected} are then passed to the CCAM module for further refinement and attention-based feature fusion. The CCAM enhances feature representation by applying both channel and spatial attention mechanisms. First, max pooling and average pooling operations are applied along the spatial dimensions to generate two feature maps, which are processed through a shared convolutional layer and combined using element-wise addition to produce the channel attention map:

$$M_{\text{channel}} = \sigma\left(\text{Conv}_{\text{shared}}(\text{MaxPool}(F_{\text{selected}}))\right) + \sigma\left(\text{Conv}_{\text{shared}}(\text{AvgPool}(F_{\text{selected}}))\right) \quad (6)$$

where σ denotes the sigmoid activation function. The channel attention map M_{channel} is applied to F_{selected} through element-wise multiplication, resulting in the weighted channel features F_{channel}.

To capture spatial dependencies, the channel-weighted features F_{channel} are first processed through a 5×5 depth-wise convolution to extract local spatial features, resulting in F_{local}. These features are then passed through parallel cascaded depth-wise convolutions with elongated kernels ($1 \times k$ and $k \times 1$, where $k \in \{7, 11, 21\}$), capturing long-range spatial dependencies. The outputs from these convolutions are aggregated with F_{local} to produce the spatial attention map:

$$M_{\text{spatial}} = F_{\text{local}} + \sum_{k\in\{7,11,21\}} \text{DWConv}_{k\times1}\left(\text{DWConv}_{1\times k}(F_{\text{local}})\right) \quad (7)$$

Finally, a 1×1 convolution is applied to fuse the channels, and the spatial attention map is multiplied with the channel-refined features to produce the final fused features:

$$F_{f_i} = \text{Conv}_{1\times1}(\text{Conv}_{1\times1}(M_{\text{spatial}}) \odot F_{\text{channel}}) \quad (8)$$

2.4 Loss function

We design a composite loss function that combines Dice loss, binary cross-entropy (BCE) loss, and clDice loss[14]:

$$\mathcal{L} = \lambda_1 \cdot \mathcal{L}_{\text{Dice}} + \lambda_2 \cdot \mathcal{L}_{\text{BCE}} + \lambda_3 \cdot \mathcal{L}_{\text{clDice}} \quad (9)$$

where λ_1, λ_2 and λ_3 are weighting coefficients balancing the contributions of each term. In particular, the clDice loss is employed to preserve the topological structure of fibrous caps, effectively reducing fragmentation and minimizing spurious regions in the segmentation.

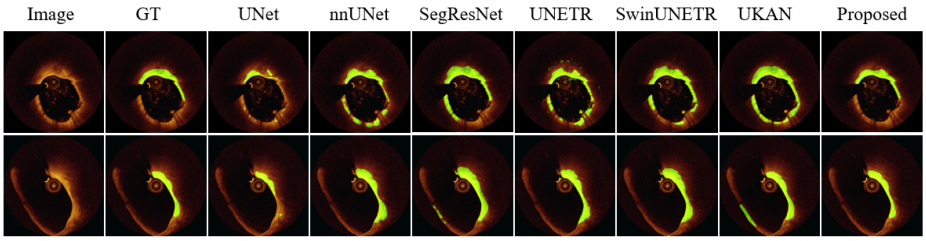

Fig. 4. Visual results of fibrous cap segmentation on Carotid OCT.

3 Experiments and Results

Datasets and Pre-processing: The dataset consists of OCT image pullbacks from 36 patients with fibrous cap lesions, totaling 1,595 cross-sectional images with an original resolution of 704×704 pixels. The data were divided at the patient level into training and validation sets in a 7:3 ratio for model parameter optimization and hyperparameter tuning. An external validation set, comprising 480 cross-sectional images from 24 additional patients, was used to evaluate model performance independently. All annotations were manually delineated by a radiologist with six years of imaging experience and subsequently reviewed by a senior neurologist with ten years of experience to ensure label accuracy. The datasets used for internal and external validation were collected from seven hospitals using two different OCT systems.

Hyper-parameters Settings: The data augmentation techniques employed during training include random rotations, flipping, adjustments to brightness and contrast, as well as shifts in hue, saturation, and value (HSV). The model was trained on a single NVIDIA RTX 3090 GPU with 24 GB of memory, using the Adam optimizer implemented in PyTorch. The initial learning rate was set to 1×10^{-2} for the KAN layers and 1×10^{-4} for the other layers. A weight decay of 1×10^{-4} was applied to all parameters. The learning rate was adjusted according to a Cosine Annealing schedule, with a minimum value of 1×10^{-5}. The batch size was set to 8, and training was conducted for 400 epochs, with early stopping employed to prevent overfitting. Extensive experiments on the validation set determined the optimal loss weights as $\lambda_1 = 0.5$, $\lambda_2 = 0.25$, and $\lambda_3 = 0.5$.

Evaluation Metrics: Three quantitative metrics, *i.e.*, Dice, Intersection over Union (IOU), and Hausdorff Distance at the 95th percentile (HD95), were used for evaluations.

Results: Six state-of-the-art segmentation methods were used as benchmarks to evaluate the effectiveness of our proposed approach, including three CNN-based methods (UNet [12], SegResNet [11], nnUNet [5]), two transformer-based methods (UNETR [4], SwinUNETR [3]), and one KAN-based method (UKAN [9]). Each model was trained using its official settings to ensure optimal performance. The quantitative results presented in Tab. 1 and the qualitative visualizations

shown in Fig. 4 demonstrate the superior performance and robustness of our method in terms of both regional accuracy and boundary alignment.

Table 1. Quantitative results on the external OCT dataset.

Methods	Dice↑	HD95↓	IOU↑	FLOPS(G)	Params(M)
UNet[12]	0.659 ± 0.203	38.10 ± 35.82	0.524 ± 0.214	0.05	0.04
SegResNet[11]	0.755 ± 0.126	43.87 ± 38.86	0.621 ± 0.148	0.99	0.39
nnUNet[5]	0.741 ± 0.143	54.08 ± 43.36	0.607 ± 0.161	25.47	1.83
UNETR[4]	0.721 ± 0.146	52.84 ± 38.90	0.582 ± 0.168	26.30	87.32
SwinUNETR[3]	0.754 ± 0.129	38.03 ± 37.02	0.621 ± 0.151	4.82	6.30
UKAN[9]	0.758 ± 0.126	44.96 ± 39.65	0.626 ± 0.151	1.76	6.36
Proposed	$\mathbf{0.785 \pm 0.114}$	$\mathbf{27.16 \pm 29.99}$	$\mathbf{0.659 \pm 0.137}$	**5.07**	**12.29**

Several ablation studies were performed to evaluate the contributions of the key components in the proposed model, including dual coordinates, CCFFM, and KAN block. The results in Tab. 2 show that using only Cartesian or polar coordinates provides moderate results, while combining them slightly improves segmentation by leveraging complementary spatial and geometric features. The addition of CCFFM greatly enhances performance by effectively integrating multi-scale features across coordinate systems. Incorporating the KAN block further strengthens the framework through improved nonlinear modeling and robustness. The complete framework achieves the best performance.

Table 2. Ablation study results.

Cartesian	Polar	CCFFM	KAN	Dice ↑	HD95 ↓	IOU ↑
✓	×	×	×	0.741 ± 0.159	43.58 ± 41.18	0.610 ± 0.177
✓	×	×	✓	0.749 ± 0.151	38.88 ± 39.82	0.616 ± 0.169
×	✓	×	×	0.748 ± 0.133	43.99 ± 38.29	0.615 ± 0.155
✓	✓	×	×	0.744 ± 0.135	40.73 ± 41.79	0.612 ± 0.155
✓	✓	×	✓	0.755 ± 0.128	34.84 ± 35.13	0.621 ± 0.149
✓	✓	✓	×	0.779 ± 0.126	29.09 ± 26.05	0.653 ± 0.146
✓	✓	✓	✓	$\mathbf{0.785 \pm 0.114}$	$\mathbf{27.16 \pm 29.99}$	$\mathbf{0.659 \pm 0.137}$

4 Discussion and Conclusion

This study presents a novel segmentation algorithm specifically designed for fibrous cap analysis in carotid artery OCT imaging, addressing critical challenges such as incomplete visualization of vascular structures, larger lumen diameters, thicker vessel walls, and residual blood artifacts. By integrating Cartesian and

polar coordinate domains, the proposed dual-coordinate framework effectively captures complementary features to mitigate the effects of incomplete vascular walls and blood artifacts, enhancing segmentation accuracy and robustness. The incorporation of a Kolmogorov-Arnold Network (KAN) block improves interpretability and nonlinear modeling, while the Cross-Coordinate Feature Fusion Module, filters and integrates dual-coordinate features, reducing interference from redundant information. While demonstrating improved segmentation accuracy and clinical interpretability, the method faces limitations in dataset diversity and computational efficiency for real-time applications.

In conclusion, this work highlights the potential of intravascular OCT for carotid artery imaging, presenting a tailored segmentation framework that addresses key anatomical and imaging challenges. The proposed method not only advances the technical capabilities of OCT-based fibrous cap segmentation but also underscores its clinical value in assessing plaque vulnerability, paving the way for broader adoption of OCT in carotid artery diagnostics and treatment planning.

Acknowledgments. The authors are grateful to the High-Performance Computing platform of Huazhong University of Science and Technology and the Supercomputing Platform of Hubei Medical Devices Quality Supervision and Test Institute.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chamié, D., Wang, Z., Bezerra, H., Rollins, A.M., Costa, M.A.: Optical coherence tomography and fibrous cap characterization. Current cardiovascular imaging reports **4**, 276–283 (2011)
2. Finder, S.E., Amoyal, R., Treister, E., Freifeld, O.: Wavelet convolutions for large receptive fields. In: European Conference on Computer Vision. pp. 363–380. Springer (2025)
3. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
4. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
5. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
6. Kolmogorov, A.N.: On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. In: Doklady Akademii Nauk. vol. 114, pp. 953–956. Russian Academy of Sciences (1957)
7. Lee, J., Kim, J.N., Dallan, L.A., Zimin, V.N., Hoori, A., Hassani, N.S., Makhlouf, M.H., Guagliumi, G., Bezerra, H.G., Wilson, D.L.: Deep learning segmentation

of fibrous cap in intravascular optical coherence tomography images. Scientific Reports **14**(1), 4393 (2024)
8. Lee, J., Pereira, G.T., Gharaibeh, Y., Kolluru, C., Zimin, V.N., Dallan, L.A., Kim, J.N., Hoori, A., Al-Kindi, S.G., Guagliumi, G., et al.: Automated analysis of fibrous cap in intravascular optical coherence tomography images of coronary arteries. Scientific Reports **12**(1), 21454 (2022)
9. Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., Chen, Z., Yuan, Y.: U-kan makes strong backbone for medical image segmentation and generation. arXiv preprint arXiv:2406.02918 (2024)
10. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756 (2024)
11. Myronenko, A.: 3d mri brain tumor segmentation using autoencoder regularization. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4. pp. 311–320. Springer (2019)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
13. Shao, H., Zeng, Q., Hou, Q., Yang, J.: Mcanet: Medical image segmentation with multi-scale cross-axis attention. arXiv preprint arXiv:2312.08866 (2023)
14. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P., Bauer, U., Menze, B.H.: cldice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16560–16569 (2021)
15. Wang, Z., Chamie, D., Bezerra, H.G., Yamamoto, H., Kanovsky, J., Wilson, D.L., Costa, M.A., Rollins, A.M.: Volumetric quantification of fibrous caps using intravascular optical coherence tomography. Biomedical optics express **3**(6), 1413–1426 (2012)
16. Zahnd, G., Karanasos, A., Van Soest, G., Regar, E., Niessen, W., Gijsen, F., van Walsum, T.: Quantification of fibrous cap thickness in intracoronary optical coherence tomography with a contour segmentation method based on dynamic programming. International journal of computer assisted radiology and surgery **10**, 1383–1394 (2015)