IKAN: Interactive KAN with Modulation Fusion for Medical Image Segmentation

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

[1] College of Life Science and Technology, Huazhong University 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. Interactive segmentation in medical imaging remains challenged by progressive loss of crucial interaction cues (click responsiveness, boundary fidelity) in deep networks. To address this limitation, we propose Interactive Kolmogorov-Arnold Network with adaptive modulation (*IKAN*), a unified framework that synergistically preserves interaction signals through spline-activated basis functions while enabling iterative anatomical refinement. The architecture achieves enhanced diagnostic fidelity by integrating three core components: hierarchical multi-scale feature extraction through Hierarchical Inception and Channel Attention Module (HICAM), dual-branch adaptive probability modulation for backbone/side-feature fusion, and click density-guided prediction sharpening. By dynamically correlating user-provided clicks with multi-modal data patterns, our method resolves ambiguous boundaries in complex clinical scenarios. Evaluated across OCT, BUSI, and AISD datasets, our method demonstrates enhanced segmentation accuracy in complex clinical scenarios, outperforming state-of-the-art approaches through systematic preservation and amplification of diagnostic interaction cues. The code is available online.

Keywords: Medical Image Interactive Segmentation · Probability Modulation · KAN · Carotid Artery OCT

1 Introduction

Interactive segmentation plays a crucial role in medical image analysis, especially in clinical settings where accurate delineation of anatomical structures is essential[18]. Unlike fully automated segmentation, which relies only on image data, interactive segmentation combines human cues (e.g., clicks or scribbles) and coarse segmentation from the previous iteration[19,15,3]. This dual guidance improves segmentation accuracy and allows for iterative refinement based on expert input, increasing clinical reliability.

* S. Liu and T. Wan contributed equally.
Correspondence to Wu Qiu: wuqiu@hust.edu.cn

However, current interactive segmentation methods have limitations. Many approaches simply add or concatenate the image, click map, and previous probability map, assuming each modality contributes equally[14,15]. This naive fusion ignores the distinct roles of each input: user cues provide localized, high-priority signals, while previous segmentation results offer useful context. As the network deepens, these cues can weaken, leading to suboptimal performance. Additionally, the iterative nature of interactive segmentation, where new inputs refine previous results, is often not fully realized. Many methods fail to properly merge new user inputs with historical probability maps, resulting in a refinement process that doesn't fully capture the contributions of the cues [12]. This can undermine the precision and robustness needed for clinical applications.

Large models tailored for medical tasks, such as SAMMed2D[4] and MedSAM[13], have been recently introduced. While these models have achieved notable progress in certain areas, they may not fully adapt to domain-specific knowledge and features, leading to decreased performance in specialized tasks.

This study presents a novel interactive segmentation framework with three key advancements: 1) a KAN framework with fusion modules that leverages multi-source inputs and iterative refinements; 2) a hierarchical multi-scale feature extraction module with channel attention to prioritize informative cues; 3) a dual-branch fusion strategy with adaptive probability modulation, combining backbone and side-branch features while refining pixel values near user clicks for precise guidance.

2 Methodology

KAN's mathematically transparent function composition and enhanced nonlinear modeling capabilities provide inherent interpretability for capturing complex anatomical patterns in medical imaging. Building on UKAN [11], we develop an enhanced backbone through three architectural innovations: wavelet convolutions integrated into KAN blocks enable multi-frequency feature learning while preserving interpretability; a Multi-Scale Cross-Attention (MSCA) module [2] bridges CNN blocks and KAN layers via spatial-channel feature recalibration; and adaptive fusion strategies prioritize error-prone regions through coordinated preprocessing and modulation. The overall architecture and interactive loop of IKAN is depicted in Figure 1.

The framework operates through three sequential phases (Fig. 1): 1) Hierarchical Preprocessing: Input tensor $X = [X_{\text{image}}, X_{\text{coord}}, X_{\text{prev}}]$ undergoes multi-scale feature extraction via HICAM's channel-attentive spatial pyramid. 2) Modulated Feature Extraction: The enhanced UKAN backbone processes features while a parallel ResNet branch derives modulation signals from current clicks and previous probability maps. 3) Adaptive Decoding: Features fused by the attentional feature fusion (AFF) [5] in bottleneck layers are decoded through residual connections preserving spatial fidelity, culminating in 1×1 convolution and softmax classification.

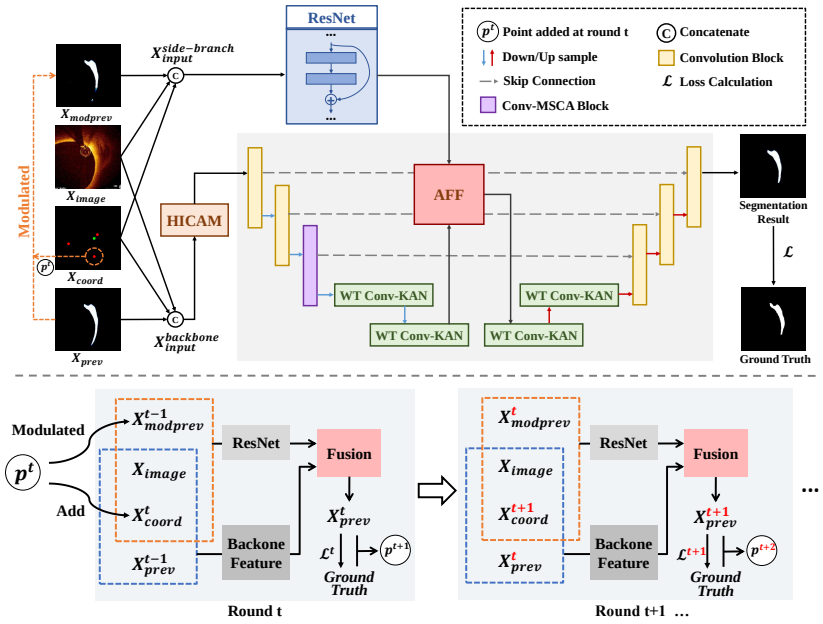

Fig. 1. Network architecture and Interactive Loop of IKAN

The IKAN framework establishes an iterative refinement loop through sequential interaction rounds denoted by t. In each round t, the system modulates the previous probability map X_{prev}^{t-1} with newly added user clicks p^t to generate enhanced guidance X_{modprev}^{t-1}, computes the current probability map X_{prev}^{t} through network processing and propagates X_{prev}^{t} as the initial probability map for round $t+1$.

The framework's closed-loop mechanism drives progressive refinement of probability maps by integrating HICAM's hierarchical context aggregation, MSCA-driven CNN/KAN synergy, and AFF-guided fusion of UKAN features with ResNet-modulated click adjustments—collectively enabling precise segmentation through iterative updates.

Our loss function combines a weighted SoftIoU [8] with iteration-specific scaling iterloss [16] to adaptively balance training objectives, where T denotes the total iterations and W_t follows $W_{t+1} > W_t$ to progressively prioritize fine-grained predictions. Each iteration loss is defined as:

$$\mathcal{L}_t = 1 - \frac{\sum(\hat{y}_t \odot y \odot w_{\text{pixel}})}{\sum(\max(\hat{y}_t, y) \odot w_{\text{pixel}})}, \quad \mathcal{L}_{\text{Total}} = \sum_{t=1}^{T} W_t \mathcal{L}_t \tag{1}$$

with $\hat{y}_t \in [0, 1]$ as the predicted probability tensor at iteration t, $y \in \{-1, 0, 1\}$ as the ground truth (where -1 indicates ignored regions), and $w_{\text{pixel}} = \mathbb{I}(y \neq -1)$ masking valid pixels.

2.1 Hierarchical Inception and Channel Attention Module (HICAM)

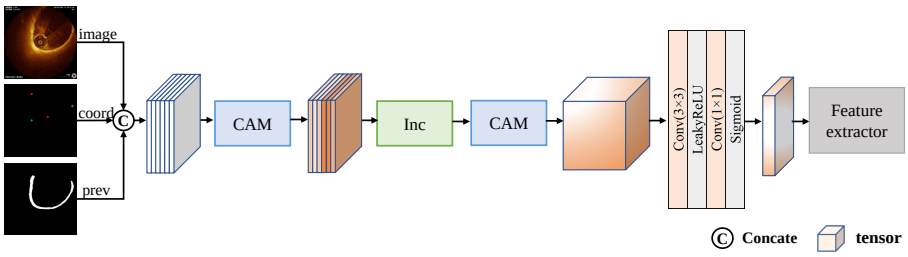

Fig. 2. Structure of HICAM

The Hierarchical Inception and Channel Attention Module (HICAM) processes multi-source inputs through a cascaded architecture to enhance interactive segmentation performance. As illustrated in Fig. 2, the module operates in three sequential stages.

The input sources, including the raw image X_{image}, user interaction coordinates X_{coord}, and prior probability map X_{prev}, are concatenated into tensor $\mathbf{X}$, which then undergoes channel-wise attention modulation (CAM) via squeeze-and-excitation operations adaptively reweighted to emphasize the importance of each input type[7] . The processed features are then processed through a multi-scale inception structure with varying convolutional kernel sizes[17], capturing spatial patterns efficiently. Finally, a second CAM layer refines the multi-scale features obtained from the inception module, enabling the model to better focus on deeper, more informative cues. The complete transformation can be formally expressed as:

$$\mathbf{Z} = \text{CAM}_{\text{final}}(\text{Inc}(\text{CAM}_{\text{init}}(\mathbf{X}))) \tag{2}$$

HICAM enhances feature focus while filtering out irrelevant information, improving segmentation accuracy. Its integration of inception-based multi-scale extraction and adaptive channel weighting makes it a critical component in interactive segmentation networks.

2.2 Enhanced Probability Map Modulation

We improve the modulation mechanism of Lee et al. [10] by addressing two limitations: (1) ineffective correction for large initial prediction errors and (2) over-expanded modulation radius. Our enhanced approach introduces three components:

1) Adaptive Modulation Radius: The radius R is constrained within $[R_{\min}, R_{\max}]$:

$$R = \min\left(R_{\max}, \max\left(R_{\min}, \frac{1}{2}\min_{v \in O}\|u - v\|\right)\right) \tag{3}$$

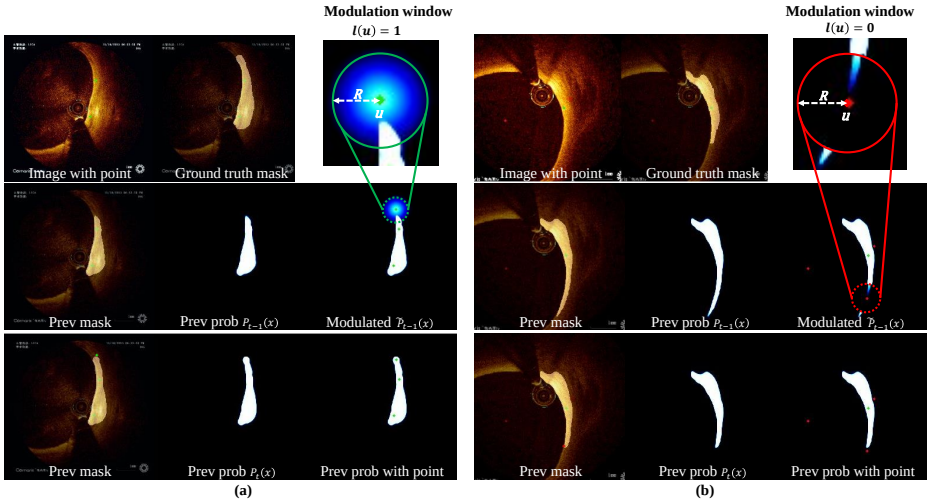

Fig. 3. Visualization of modulation effects for (a) positive clicks (enhancement) and (b) negative clicks (suppression). The color gradient indicates probability values.

where u represents the current click position, $O = \{v_1, v_2, \dots\}$ is the set of all previous clicks of the type opposite to u. This ensures balanced spatial context by considering distances to all relevant opposite-type clicks.

2) Distance-based Modulation Mask: A linear mask $M(x)$ with intensity α:

$$M(x) = \begin{cases} \alpha \left(1 - \frac{\|x-u\|}{R}\right), & \|x-u\| \le R \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where x represents the spatial coordinates of an arbitrary pixel in the image.

3) Refined Probability Update: The modulated probability becomes:

$$\tilde{P}_{t-1}(x) = \begin{cases} P_{t-1}(x)^{1/\gamma} + M(x), & l(u) = 1 \wedge \|x-u\| \le R \\ P_{t-1}(x)^{\gamma} - M(x), & l(u) = 0 \wedge \|x-u\| \le R \\ P_{t-1}(x), & \text{otherwise} \end{cases} \tag{5}$$

where $P_{t-1}(x)$ is the previous prediction probability at location x, $l(u) \in \{0, 1\}$ is the click label (1 for positive/foreground click, 0 for negative/background click), $\gamma > 1$ is the adaptive gamma factor that sharpens probabilities for positive clicks and smooths them for negative clicks.

As shown in Figure 3, this design enforces $P(x) = 1$ at positive click centers and $P(x) = 0$ at negative click centers, with smooth distance-based transitions. The adaptive radius prevents excessive modulation scope while maintaining spatial relationships to existing clicks.

The enhanced modulation mechanism offers three primary advancements compared to the original framework: 1) rigorous boundary enforcement at user-

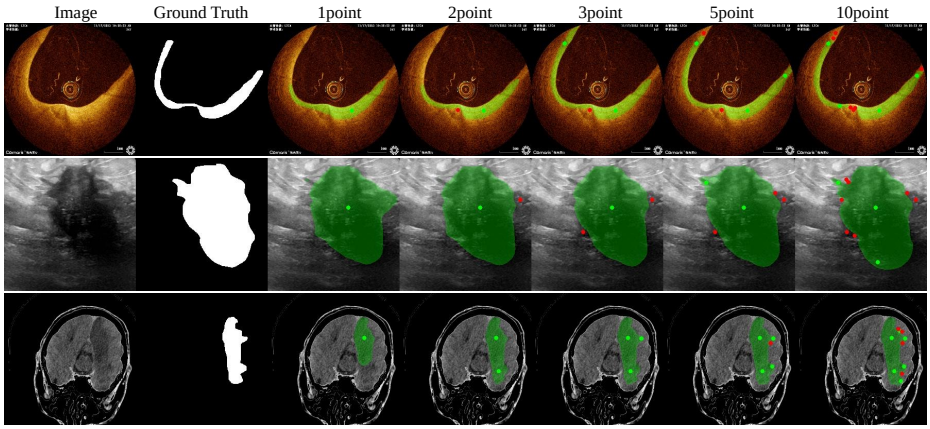

Fig. 4. Results of IKAN, from top to bottom are carotid artery OCT, BUSI, AISD.

specified click locations through precise probability assignment (1.0 for positive clicks, 0.0 for negative clicks), eliminating regional ambiguity; 2) improved probability transition smoothness via linear distance masking, outperforming gamma correction approaches; 3) adaptive radius constraints that prevent over-modulation in sparse interaction scenarios, effectively preserving spatial relationships between local refinements and global context. These optimizations collectively strengthen iterative segmentation robustness, especially when addressing substantial initial prediction inaccuracies or spatially distributed user inputs.

2.3 Dual-Branch Interactive Fusion

User-provided positive/negative clicks iteratively guide prediction refinement in interactive segmentation. Our dual-branch architecture (Figure 1) synergizes global context and localized correction. The backbone branch processes concatenated inputs $X = [X_{\text{image}}, X_{\text{coord}}, X_{\text{prev}}]$ to maintain multi-scale anatomical context. Simultaneously, the ResNet side branch [6] enhances spatial modulation of error-prone regions using $X = [X_{\text{image}}, X_{\text{coord}}, X_{\text{modprev}}]$ where X_{modprev} encodes previous predictions adjusted by current user clicks.

The extracted features F_{backbone} and F_{side} are fused via attentional feature fusion(AFF) [5], where $\mathcal{A}$ generates spatial attention weights α through channel-wise averaging and convolution operations:

$$F_{\text{fused}} = \alpha \cdot F_{\text{backbone}} + (1 - \alpha) \cdot F_{\text{side}}, \quad \alpha = \sigma(\mathcal{A}([F_{\text{backbone}}, F_{\text{side}}])) \tag{6}$$

Our dual-branch fusion architecture synergizes global context preservation with user-guided refinement to enhance interactive segmentation accuracy. The backbone branch processes the original image, coordinate maps, and prior predictions to capture multi-scale contextual features, while the spatially modulated side branch selectively enhances regions marked by user annotations. This

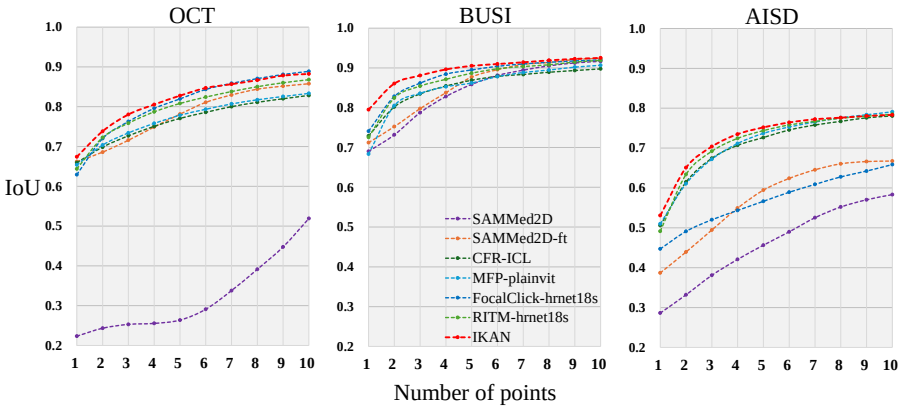

Fig. 5. IoU improvement with increasing number of points.

spatial modulation explicitly prioritizes error-prone regions, thereby improving segmentation precision. An attention-based fusion mechanism dynamically balances these complementary streams—preserving anatomical coherence through global context while incorporating real-time user corrections—ensuring robust performance across diverse interaction patterns. The modular architecture facilitates flexible customization of interaction handling and feature integration strategies.

3 Experiments and Results

Datasets and Pre-processing: The following image sets including our own dataset and publicly available datasets are used to evaluate the performance of our approaches: (1) Private carotid artery OCT dataset (1,142 training and 453 test samples), which consists of OCT image pullbacks from 36 patients with fibrous cap lesions. (2) BUSI breast ultrasound dataset (452 training and 75 test) which collected in 2018 covering female patients aged 25 to 75 years[1]. (3) AISD acute ischemic stroke dataset (2,945 training and 841 test)[9] which contains 397 Non-Contrast-enhanced CT (NCCT) scans of acute ischemic stroke.

Hyper-parameters Settings: The data augmentation techniques employed during training include random resizing, horizontal flipping, padding, random cropping, and adjustments to brightness and contrast. We trained on an NVIDIA RTX 4080 GPU with 16 GB of memory. We use the Adam optimizer with an initial learning rate of 5×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$ and MultiStepLR scheduler reducing rates by $10\times$ at epochs over 220 with early stopping employed to prevent overfitting. The batch size is set to 4, and the input images are resized to 256×256. Extensive experiments on the validation set determined the optimal iteration loss weight $W_t \in \{1, 2, 3\}$ and the modulated radius $[R_{\min}, R_{\max}]$ is [6, 25].

Evaluation Metrics: We adopt rigorous clinical standards using Number of Clicks (NoC) at three thresholds: NoC@80, NoC@85, and NoC@90. Our click-based method is grounded on NoBRS[15], which optimizes image segmentation results through user interactions. Simulated clicks are generated until reaching target IoU or max 20 clicks.

Results: We conducted a comparative evaluation of our approach with several state-of-the-art methods, including RITM-hrnet18s[15], FocalClick-hrnet18s[3], MFP-plainvit[10], CFR-ICL[16], SAMMed2D[4] and SAMMed2D-ft which is obtained by fine-tuning the SAMMed2D on the same datasets. These evaluations were performed across both publicly available datasets (BUSI and AISD) and our own carotid artery OCT dataset. As shown in Table 1, IKAN achieves higher accuracy than all baseline methods. The framework requires minimal user interactions to attain competitive segmentation quality (Figure 5). The ablation studies confirm the critical contributions of both HICAM and adaptive mask fusion modules outperforming the versions with individual components removed with all modules enabled. These components collectively enhance segmentation precision in complex medical imaging scenarios, as evidenced by the performance improvements in Table 2 and visual results in Figure 4.

Table 1. Comparison of segmentation performance

Method	**OCT**			**BUSI**			**AISD**		
NoC@	80	85	90	80	85	90	80	85	90
RITM-hrnet18s[15]	4.73	7.54	12.95	2.41	3.41	6.45	9.70	13.46	17.68
FocalClick-hrnet18s[3]	3.98	6.16	**11.25**	1.92	3.13	6.01	13.4	15.99	18.59
MFP-plainvit[10]	5.59	8.32	13.85	2.52	3.87	7.55	9.36	13.50	17.66
CFR-ICL[16]	5.90	8.62	14.38	2.60	3.61	7.09	9.92	13.93	17.84
SAMMed2D-ft[4]	5.82	9.81	16.55	2.89	3.77	6.75	14.71	17.33	19.27
SAMMed2D[4]	14.99	17.6	19.75	3.41	5.07	7.84	16.11	17.92	19.28
IKAN$^{\text{Proposed}}$	**3.56**	**5.98**	12.35	**1.57**	**2.29**	**4.59**	**9.24**	**13.26**	**17.41**

Table 2. Ablation Study Results

KAN	**HICAM**	**Fusion Modulation**	**AFF**	**NoC@80**	**NoC@85**	**NoC@90**
✓	×	×	×	4.00	6.40	12.39
✓	✓	×	×	3.80	6.14	**11.9**
✓	✓	✓	×	3.83	6.25	12.41
✓	✓	✓	✓	3.79	6.14	12.51
×	✓	✓	✓	4.01	6.43	12.56
✓	✓	✓	✓	**3.56**	**5.98**	12.35

4 Discussion and Conclusion

This study introduces IKAN (Interactive KAN framework), a clinically viable medical image segmentation system built upon an enhanced UKAN backbone. Designed to address complex anatomical segmentation challenges, IKAN achieves high precision and robustness across diverse medical imaging scenarios. The framework integrates multiple innovative components, such as HICAM and an adaptive dual-branch fusion mechanism with dynamic modulation, enabling effective adaptation to various imaging modalities and clinical tasks. Particularly effective in demanding applications like carotid artery segmentation in OCT imaging and demonstrates superior performance in resource-constrained environments where high-quality annotated data are limited. Quantitative evaluations across three benchmark datasets reveal that IKAN enhances input utilization efficiency compared to conventional approaches while maintaining critical interactive cues throughout network processing.

In conclusion, IKAN provides an efficient solution for interactive segmentation, improving performance in clinically challenging, low-annotation scenarios.

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. No competing interests to declare.

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)
2. Chen, J., Wan, L., Zhu, J., Xu, G., Deng, M.: Multi-scale spatial and channel-wise attention for improving object detection in remote sensing imagery. IEEE Geoscience and Remote Sensing Letters **17**(4), 681–685 (2019)
3. Chen, X., Zhao, Z., Zhang, Y., Duan, M., Qi, D., Zhao, H.: Focalclick: Towards practical interactive image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1300–1309 (2022)
4. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: Sam-med2d. arXiv preprint arXiv:2308.16184 (2023)
5. Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., Barnard, K.: Attentional feature fusion. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 3560–3569 (2021)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
8. Huang, Y., Tang, Z., Chen, D., Su, K., Chen, C.: Batching soft iou for training semantic segmentation networks. IEEE Signal Processing Letters **27**, 66–70 (2019)

9. Kongming Liang, Kai Han, X.L.X.C.Y.L.Y.W., Yu, Y.: Symmetry-enhanced attention network for acute ischemic infarct segmentation with non-contrast ct images. In: MICCAI (2021)
10. Lee, C., Lee, S.H., Kim, C.S.: Mfp: Making full use of probability maps for interactive image segmentation. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4051–4059. IEEE (2024)
11. 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)
12. Lin, Z., Zhang, Z., Chen, L.Z., Cheng, M.M., Lu, S.P.: Interactive image segmentation with first click attention. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13339–13348 (2020)
13. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
14. Sofiiuk, K., Petrov, I., Barinova, O., Konushin, A.: F-brs: Rethinking backpropagating refinement for interactive segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
15. Sofiiuk, K., Petrov, I.A., Konushin, A.: Reviving iterative training with mask guidance for interactive segmentation. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 3141–3145. IEEE (2022)
16. Sun, S., Xian, M., Xu, F., Capriotti, L., Yao, T.: Cfr-icl: Cascade-forward refinement with iterative click loss for interactive image segmentation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 5017–5024 (2024)
17. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9 (2015)
18. Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., et al.: Deepigeos: a deep interactive geodesic framework for medical image segmentation. IEEE transactions on pattern analysis and machine intelligence **41**(7), 1559–1572 (2018)
19. Xu, N., Price, B., Cohen, S., Yang, J., Huang, T.S.: Deep interactive object selection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 373–381 (2016)