Graph-based Neighbor-Aware Network for Gaze-Supervised Medical Image Segmentation

Shaoxuan Wu[1], Jingkun Chen[2], Zhuo Jin[1], Peilin Zhang[1], Zhizezhang Gao[1], Jun Feng[1(✉)], Xiao Zhang[1(✉)], and Dinggang Shen[3,4,5]

[1] School of Information Science and Technology, Northwest University, Xi'an, China
{xiaozhang,fengjun}@nwu.edu.cn
[2] Department of Engineering Science, University of Oxford, Oxford, UK
[3] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
[4] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China
[5] Shanghai Clinical Research and Trial Center, Shanghai, China

Abstract. Creating fully annotated labels for medical image segmentation is time-consuming and expensive, underscoring the need for efficient labeling schemes to alleviate the workload. Eye tracking presents a cost-effective solution, seamlessly integrating into radiologists' workflows while offering task-relevant eye gaze supervision. However, due to the inaccuracy and ambiguity of gaze, it may introduce erroneous supervision and hinder the model's ability to learn robust features. To address these challenges, we propose the graph-based neighbor-aware network (GNAN). The network constructs a graph structure from the image, separating different categories of nodes by simulating the attention distribution during the diagnostic process, to learn image segmentation based on the radiologist's gaze information. The GNAN comprises neighbor-aware pseudo supervision (NAP) and graph contrastive decoupling (GCD). NAP utilizes the neighbor features of graph nodes to infer pseudo-labels for uncertain regions, effectively compensating for the inaccuracy in gaze supervision and further refining the supervisory signal. GCD decouples the graph structure by maximizing the inter-class node feature differences to distinguish between different categories, thereby improving segmentation performance. Experimental results on the public dataset demonstrate that GNAN outperforms state-of-the-art methods. Our code is available at https://github.com/IPMI-NWU/GNAN.

Keywords: Eye-tracking · Graph Neural Network · Medical Image Segmentation.

1 Introduction

Deep learning models have achieved impressive performance in medical image segmentation tasks [15]. However, achieving competitive accuracy and robust generalization typically requires a complete, large annotation [18]. Since manual annotation of medical images requires the specialized expertise of clinical professionals, the process is time-consuming and labor-intensive, making it costly

and posing a significant barrier to the widespread clinical adoption of these technologies [29]. To mitigate this issue, weakly supervised learning methods have gained popularity in medical image segmentation. Current strategies often rely on bounding boxes [22], points [26], or scribbles [5] to provide sparse supervision. However, these methods typically require additional human effort, disrupt clinical workflows, and increase the burden on radiologists [24].

Eye gaze, as a form of human-computer interaction data, can be automatically captured using an eye tracker and seamlessly integrated into the daily workflow of radiologists [1]. It provides insights into areas of focus during the diagnostic process, facilitating the generation of dense annotations relevant to the task. Compared to existing sparse annotation methods, gaze data offers a more effective and practical approach to annotation [17].

Leveraging the rich expert knowledge embedded in gaze data, several studies have attempted to use it for medical image segmentation [4,20,10]. Xie et al. [25] used gaze heatmaps to correct network errors, guiding the network and improving segmentation accuracy. Wang et al. [21] proposed a cross-attention transformer that incorporates gaze heatmaps into the model. While these methods treat gaze as auxiliary information, they are still constrained by their reliance on full annotation. Zhong et al. [30] binarize gaze heatmaps at different thresholds to generate multiple pseudo-masks, integrating multi-level human attention to enhance discrimination. However, existing methods typically treat gaze as a simple distinction between foreground and background, while its inherent inaccuracy leads to erroneous supervision and segmentation bias. Additionally, erroneous supervision further results in biased target locations and structures, thereby negatively impacting segmentation performance.

To address these challenges, we propose a novel graph-based neighbor-aware network (GNAN), which divides gaze into background, foreground, and uncertain regions. The uncertain regions represent ambiguous attention areas that are considered noisy and unreliable. The GNAN constructs a graph from the image and separates different categories of nodes by simulating visual attention distribution during the diagnostic process. GNAN features two special designs: 1) Neighbor-aware pseudo supervision (NAP) utilizes the neighbor feature of graph nodes to infer pseudo-labels for uncertain regions, thereby enhancing the supervision constraint. 2) Graph contrastive decoupling (GCD), which decouples the graph structure by maximizing the inter-class node feature differences to improve segmentation performance. The experiments on the public dataset demonstrate that GNAN outperforms state-of-the-art methods.

2 Method

2.1 GNAN Architecture

As illustrated in Fig.1, the graph-based neighbor-aware network (GNAN) partitions the image into multiple patches, which are subsequently flattened into vectors to form nodes for graph construction. GNAN learns attention distribution during the diagnostic process, categorizing nodes within the graph structure

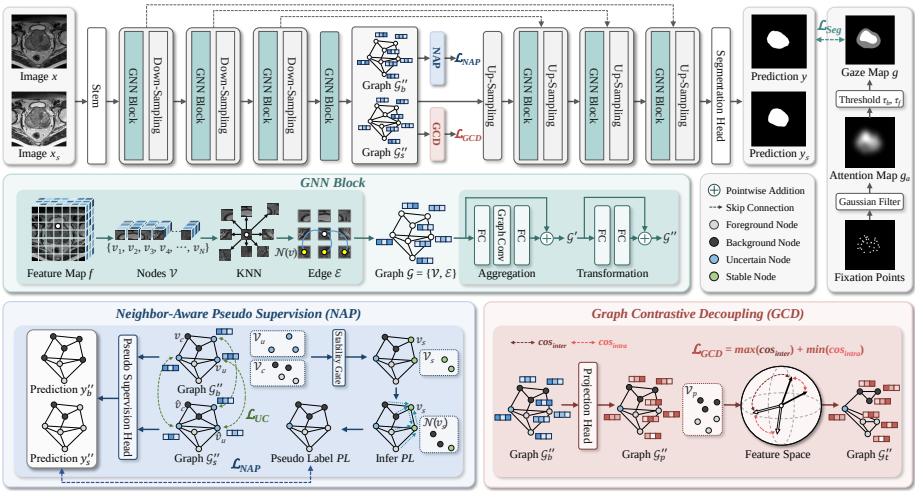

Fig. 1. Illustration of GNAN, dividing gaze into foreground, background, and uncertain regions. GNAN consists of GNN blocks for graph construction and message passing. It includes neighbor-aware pseudo supervision for pseudo-label estimation and enhanced supervision, and also graph contrastive decoupling for better separating categories.

and identifying image segmentation patterns based on the radiologist's gaze information. The GNAN architecture consists of GNN blocks and sampling layers. Given an input image x, an augmented version x_s with brightness-contrast enhancement and added random noise is first downsampled by a factor of four through the stem layer. Then it passes through multiple GNN blocks, followed by the sampling layer, skip connection, and a segmentation head to generate the predicted maps y and y_s.

The GNN block in the GNAN extracts features by constructing graph from the image, and performs aggregation and transformation operations [9]. The feature map $f \in \mathbb{R}^{H\times W\times C}$ is divided into a set of nodes $\mathcal{V} = \{v_1, v_2, \cdots, v_N\}$, where each node corresponds to a pixel. For every node $v \in \mathbb{R}^C$, the KNN algorithm is applied to identify neighboring nodes $\mathcal{N}(v)$, and edges e are formed between the nodes. The final graph is represented as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{E}$ denotes the set of edges. During message passing and feature extraction, $\mathcal{G}$ undergoes aggregation and transformation operations, as follows:

$$\mathcal{G}' = \underbrace{FC_2(GC(FC_1(\mathcal{G}))) + \mathcal{G}}_{Aggregation}, \quad \mathcal{G}'' = \underbrace{FC_4(FC_3(\mathcal{G}')) + \mathcal{G}'}_{Transformation}, \tag{1}$$

where $FC(\cdot)$ represents fully connected layers, and $GC(\cdot)$ denotes max-relative graph convolutions [13].

To incorporate gaze, the raw gaze data is first processed for extracting fixation points, and then a 2D Gaussian filter is used to generate the attention map g_a. The background threshold and foreground threshold, τ_b and τ_f, are applied

to g_a to generate the gaze map g into three regions: background ($g_a \leq \tau_b$), uncertain region ($\tau_b < g_a < \tau_f$), and foreground ($g_a \geq \tau_f$), where the uncertain region is considered noisy and unreliable. The segmentation loss is defined as:

$$\mathcal{L}_{Seg} = \mathcal{L}_{pce}(y, g) + \mathcal{L}_{pce}(y_s, g), \tag{2}$$

$$\mathcal{L}_{pce}(y, g) = -\frac{1}{|\Omega_{coor}|} \sum_{i \in \Omega_{coor}} g_i \, log(y_i) + (1 - g_i) \, log(1 - y_i), \tag{3}$$

where $\mathcal{L}_{pce}$ represents the partial cross-entropy loss and Ω_{coor} denotes the set of pixel coordinates within the certain region of the gaze map g.

2.2 Neighbor-Aware Pseudo Supervision

Inaccurate or ambiguous gaze can lead to erroneous supervision, negatively impacting performance. To mitigate this issue, neighbor-aware pseudo supervision (NAP) utilizes neighboring node information from uncertain regions to infer their pseudo-labels, thereby refining supervision for areas with uncertain gaze data. For the graph $\mathcal{G}_b''$ and $\mathcal{G}_s''$ constructed at the bottom-most GNN block of the network, which correspond to the input image x and the perturbed image x_s, respectively, the nodes are divided into two sets: 1) certain nodes, $\mathcal{V}_c$ (including foreground and background), and 2) uncertain nodes, $\mathcal{V}_u$, based on their locations in the gaze map g. NAP then applies a stability gate to partition the uncertain node set $\mathcal{V}_u$ into stable nodes $\mathcal{V}_s$, defined as:

$$\mathcal{V}_s = \{v_u \in \mathcal{V}_u \mid cos(v_u, \hat{v}_u) > \tau_s\}, \tag{4}$$

where v_u denotes an uncertain node in $\mathcal{G}_b''$, and $\hat{v}_u$ represents the corresponding uncertain node at the same position in $\mathcal{G}_s''$. The function $cos(\cdot)$ denotes cosine similarity, i.e., $cos(v_1, v_2) = \frac{v_1 \cdot v_2}{||v_1||_2 \cdot ||v_2||_2}$. The threshold τ_s is determined by the average cosine similarity of all certain nodes, given by:

$$\tau_s = \frac{1}{|\mathcal{V}_c|} \sum\nolimits_{v_c \in \mathcal{V}_c} cos(v_c, \hat{v}_c), \tag{5}$$

where $\hat{v}_c$ denotes the node corresponding to v_c at the same position in $\mathcal{G}_s''$. For each stable node v_s, its pseudo-label is estimated based on the neighboring nodes $\mathcal{N}(v_s)$ as:

$$label(v_s) = \mathbb{I}\Big(\big(\frac{1}{|\mathcal{N}(v_s)|} \sum_{v \in \mathcal{N}(v_s)} cos(v, v_s) \cdot label(v)\big) > 0\Big), \tag{6}$$

where $label(\cdot)$ represents the node's class, $label(v) = 1$ if v is foreground, -1 if background, and 0 if uncertain. The indicator function $\mathbb{I}(\cdot)$ evaluates to 1 if the condition is satisfied, and otherwise -1. After estimating the labels for all stable nodes, the pseudo-label PL is generated. The NAP loss is then defined as:

$$\mathcal{L}_{NAP} = \mathcal{L}_{pce}(y_b'', PL) + \mathcal{L}_{pce}(y_s'', PL), \tag{7}$$

where y_b'' and y_s'' are the predictions for each node in $\mathcal{G}_b''$ and $\mathcal{G}_s''$, respectively, obtained from the pseudo supervision head.

To further strengthen the supervision of the uncertain regions, feature consistency regularization is enforced. For uncertain nodes v_u and their augmented versions $\hat{v}_u$, uncertain consistency (UC) is applied to implicitly ensure that the model adheres to the smoothness and enhanced stability and reliability, which can be formulated as:

$$\mathcal{L}_{UC} = 1 - \frac{1}{|\mathcal{V}_u|} \sum_{v_u \in \mathcal{V}_u} cos(v_u, \hat{v}_u). \tag{8}$$

2.3 Graph Contrastive Decoupling

Incomplete gaze annotations lead to erroneous supervision, hindering the model's ability to learn discriminative inter-class features. To mitigate this issue, we propose graph contrastive decoupling (GCD), which enhances inter-class feature discrimination in gaze maps and decouples features from distinct categories within the graph structure. GCD first applies a projection head [6] to map the node features of the graph $\mathcal{G}_b''$ into a new space, producing $\mathcal{G}_p''$. Subsequently, GCD maximizes the inter-class while minimizing the intra-class feature discrimination, as formulated below:

$$\mathcal{L}_{GCD} = -\frac{1}{|\mathcal{V}_p|} \sum_{v \in \mathcal{V}_p} log(\frac{\sum_{v_+ \in \mathcal{V}_{v+}} exp(cos(v, v_+)/\tau)}{\sum_{v' \in \mathcal{V}_p} exp(cos(v, v')/\tau)}), \tag{9}$$

where $\mathcal{V}_p$ denotes the set of certain and stable nodes in the graph $\mathcal{G}_p''$, and $\mathcal{V}_{v+}$ refers to the set of nodes that belong to the same class as v. The temperature coefficient τ controls the sharpness of the output probability distribution. The goal of $\mathcal{L}_{GCD}$ is to decouple different-category nodes in $\mathcal{G}_p''$, producing $\mathcal{G}_t''$, which increases the model's ability to learn discriminative inter-class features. The NAP and BAP modules are used exclusively during training; at inference time, only the network is employed to segment the original image x.

During the upsampling stage, after each GNN block, the strategies of NAP, UC, and GCD are applied to the graph structure. The final optimization objective for GNAN is formulated below:

$$\mathcal{L} = \mathcal{L}_{Seg} + \lambda(\mathcal{L}_{UC} + \mathcal{L}_{NAP} + \mathcal{L}_{GCD}). \tag{10}$$

3 Experiments and Results

3.1 Dataset and Evaluation Metrics

GNAN was evaluated on the GazeMedSeg dataset [30], which includes KvasirSEG [12] and NCI-ISBI [2]. The KvasirSEG dataset contains 900 training and 100 test images for polyp segmentation in gastrointestinal images. The NCI-ISBI dataset is used for prostate segmentation from T2-weighted MRI images, containing

Table 1. Comparison with different methods for five annotation types. Bold indicates the best results among the weakly supervised methods and AT denotes the annotation time corresponding to each annotation type.

Method	Sup.	NCI-ISBI	KvasirSEG	
		Dice(%)↑	Dice(%)↑	AT↓
▲ UNet [16]	Full	80.58±0.48	82.12±1.11	18.7 hrs
▲ nnUNet [11]	Full	81.54±0.45	85.37±0.48	18.7 hrs
♣ BoxInst [19]	Box	73.78±1.15	65.72±2.97	3.1 hrs
♣ BoxTeacher [8]	Box	75.60±1.15	73.33±1.30	3.1 hrs
♦ PointSup [7]	Point	73.46±4.71	73.05±1.64	4.8 hrs
♦ AGMM [23]	Point	73.86±1.26	75.57±0.84	4.8 hrs
♠ AGMM [23]	Scribble	72.70±1.03	67.23±1.02	2.6 hrs
♠ CycleMix [27]	Scribble	73.41±1.09	76.43±0.65	2.6 hrs
♠ ShapePU [28]	Scribble	73.06±1.18	77.26±0.73	2.6 hrs
♠ ScribFormer [14]	Scribble	74.31±1.29	75.69±0.48	2.6 hrs
■ UNet [16]	Gaze	74.75±1.58	73.74±0.94	2.2 hrs
■ TransUNet [3]	Gaze	75.46±1.20	70.38±0.86	2.2 hrs
■ nnUNet [11]	Gaze	77.20±1.03	74.42±0.92	2.2 hrs
■ GazeMedSeg [30]	Gaze	77.64±0.57	77.80±1.02	2.2 hrs
■ **Ours**	Gaze	**80.33**±0.24	**79.32**±0.39	2.2 hrs

789 training images and 117 test images. Segmentation performance across various methods was evaluated using the Dice coefficient. All experiments were conducted on an NVIDIA 3080Ti GPU (12GB) using PyTorch. The Adam optimizer was employed for training, with a learning rate of 7×10^{-5} and a batch size of 8. The training process ran for 100 epochs. The background and foreground thresholds, τ_b and τ_f, were set to 0.3 and 0.6, respectively. In Eq.4, Eq.9 and Eq.10, the stability threshold τ_s, the temperature coefficient τ and the balance coefficient λ are set to 0.981, 0.1 and 0.5, respectively.

3.2 Comparison with State-of-the-Art Methods

Quantitative Results. The quantitative results are summarized in Table.1, where five different supervision methods are evaluated: full annotation (black triangle ▲), bounding box annotation (orange club ♣), point annotation (green diamond ♦), scribble annotation (blue spade ♠), and gaze annotation (red square ■). All methods are expressed as the mean and standard deviation of three different seed runs. On the NCI-ISBI dataset, GNAN achieves state-of-the-art performance among weakly supervised methods, with a Dice score of 80.33%, an improvement of 2.69% over the previous best-performing method GazeMedSeg (80.33% vs. 77.64%), and also demonstrates comparable performance to fully supervised methods. Our method also achieves the highest Dice score of 79.32% on the KvasirSEG dataset among weakly supervised approaches. The results

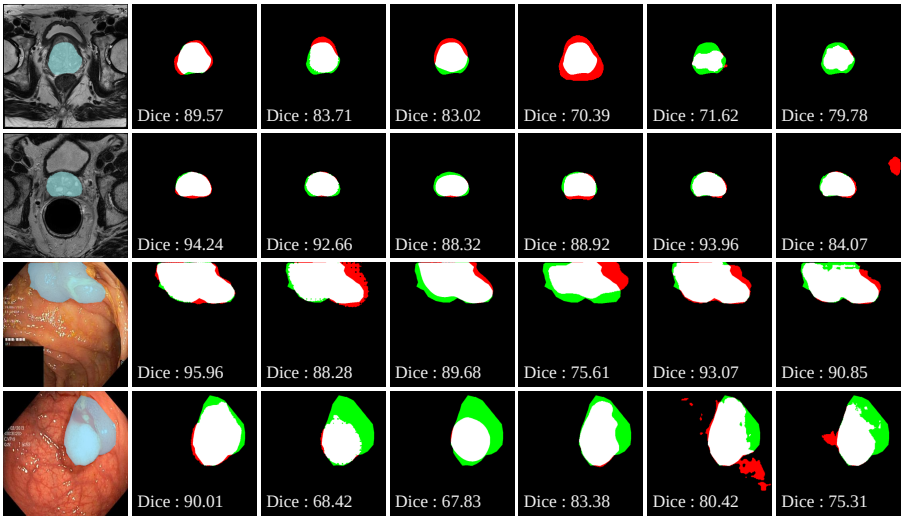

Fig. 2. Qualitative comparison of GNAN with other state-of-the-art methods. Over-segmented areas are marked in red, and under-segmented areas are marked in green.

Table 2. Ablation studies with respect to different settings, to evaluate the contribution of different components. Bold denotes the best Dice scores.

$\mathcal{L}_{Seg}$	$\mathcal{L}_{UC}$	$\mathcal{L}_{NAP}$	$\mathcal{L}_{GCD}$	**NCI-ISBI**	**KvasirSEG**
✓				77.49±0.67	76.42±0.59
✓	✓			78.29±0.60	77.09±0.46
✓		✓		78.53±0.64	78.03±0.42
✓	✓	✓		79.11±0.55	78.54±0.71
✓		✓	✓	79.40±0.55	78.62±0.56
✓	✓	✓	✓	**80.33**±0.24	**79.32**±0.39

highlight the effectiveness of GNAN in leveraging imperfect gaze by synthesizing PL from neighbors to enhance supervision while learning discriminative features.
Qualitative Visualization. Fig.2 presents qualitative results comparing GNAN with other state-of-the-art methods under weak supervision across both datasets. Red regions indicate over-segmentation, while green regions highlight under-segmentation. Compared to different methods, our approach demonstrates reduced segmentation biases and superior performance. The improvement can be attributed to GNAN's ability to better handle gaze inaccuracy through NAP and GCD, which allows for more precise feature learning.

3.3 Ablation Study

Contribution of Different Components. Table.2 presents an ablation study to evaluate the contribution of each module in GNAN. The results reveal that

Fig. 3. Visualization of the pseudo-labels generated by GNAN for uncertain regions using neighborhood information.

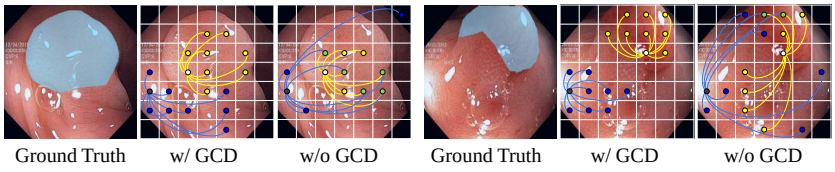

Fig. 4. Visualization of the graph structure constructed by GNAN. By introducing GCD, the graph structure effectively separates the foreground and background, enhancing the robustness of the features.

incorporating the $\mathcal{L}_{UC}$ and $\mathcal{L}_{NAP}$ modules leads to significant performance improvements, indicating that the uncertainty consistency regularization and the strategy of introducing additional supervision via reliable PL effectively enhance the model's performance. Furthermore, adding the $\mathcal{L}_{GCD}$ yields the highest Dice score, confirming the effectiveness of GCD in decoupling the graph structure and facilitating better segmentation performance.

Pseudo-Label Generation with NAP. Fig.3 illustrates the pseudo-label (PL) generated by the NAP module during training, following the final GNN block in GNAN. The results clearly show that NAP effectively estimates the labels for uncertain regions, and the generated PL is closely aligned with the ground truth. This demonstrates that NAP provides more reliable supervisory signals for training uncertain regions, mitigating the potential errors associated with supervision during binary mask generation. By leveraging neighboring node features within the graph, NAP compensates for the inaccuracy associated with gaze, improving the quality of the supervisory signal.

Decoupling the Graph Structure with GCD. Fig.4 showcases the success of the GCD module in decoupling the graph structure. White and black nodes represent foreground and background center nodes, respectively. Without GCD, incomplete gaze supervision results in feature instability, causing unwanted coupling between nodes (green nodes). This interference negatively affects the feature aggregation process, leading to background regions influencing foreground regions, which in turn reduces segmentation accuracy. After introducing GCD, the graph structure effectively decouples features of different categories, improving feature robustness and enhancing segmentation performance.

4 Conclusion

We have proposed a novel graph-based neighbor-aware network (GNAN) to simulate attention distribution during the diagnostic process for separating various categories of nodes within the graph, which enables image segmentation based on radiologists' gaze information. GNAN integrates neighbor-aware pseudo supervision (NAP) and graph contrastive decoupling (GCD). In particular, NAP utilizes neighboring features of graph nodes to infer pseudo-labels for uncertain regions, effectively mitigating gaze supervision inaccuracy and enhancing supervision constraint. GCD decouples the graph structure by maximizing inter-class node feature differences, thereby distinguishing between categories and enhancing segmentation performance. Experiments on the public dataset demonstrate better performance of GNAN over state-of-the-art methods, by maintaining high-quality segmentation even with limited annotation resources.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (No. 62403380), and the Shaanxi Province Postdoctoral Science Foundation (No.2024BSHSDZZ042).

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

References

1. Bhattacharya, M., Jain, S., Prasanna, P.: RadioTransformer: A cascaded global-focal transformer for visual attention–guided disease classification. In: European Conference on Computer Vision. pp. 679–698. Springer (2022)
2. Bloch, B.N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enquobahrie, A., Jaffe, C., Clarke, L., Farahani, K.: NCI-ISBI 2013 challenge: Automated segmentation of prostate structures (2013), http://doi.org/10.7937/K9/TCIA.2015.zF0vlOPv
3. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
4. Chen, J., Duan, H., Zhang, X., Gao, B., Tan, T., Grau, V., Han, J.: From gaze to insight: Bridging human visual attention and vision language model explanation for weakly-supervised medical image segmentation. arXiv preprint arXiv:2504.11368 (2025)
5. Chen, J., Huang, W., Zhang, J., Debattista, K., Han, J.: Addressing inconsistent labeling with cross image matching for scribble-based medical image segmentation. IEEE Transactions on Image Processing **34**, 842–853 (2025)
6. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 1597–1607. PMLR (2020)
7. Cheng, B., Parkhi, O., Kirillov, A.: Pointly-supervised instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2617–2626 (2022)

8. Cheng, T., Wang, X., Chen, S., Zhang, Q., Liu, W.: Boxteacher: Exploring high-quality pseudo labels for weakly supervised instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3145–3154 (2023)
9. Han, K., Wang, Y., Guo, J., Tang, Y., Wu, E.: Vision GNN: An image is worth graph of nodes. Advances in Neural Information Processing Systems **35**, 8291–8303 (2022)
10. Ibragimov, B., Mello-Thoms, C.: The use of machine learning in eye tracking studies in medical imaging: A review. IEEE Journal of Biomedical and Health Informatics **28**(6), 3597–3612 (2024)
11. 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)
12. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., de Lange, T., Johansen, D., Johansen, H.D.: Kvasir-SEG: A segmented polyp dataset. In: MultiMedia Modeling. pp. 451–462. Springer (2020)
13. Li, G., Muller, M., Thabet, A., Ghanem, B.: DeepGCNs: Can GCNs go as deep as CNNs? In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9267–9276 (2019)
14. Li, Z., Zheng, Y., Shan, D., Yang, S., Li, Q., Wang, B., Zhang, Y., Hong, Q., Shen, D.: Scribformer: Transformer makes CNN work better for scribble-based medical image segmentation. IEEE Transactions on Medical Imaging **43**(6), 2254–2265 (2024)
15. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
17. Saab, K., Hooper, S.M., Sohoni, N.S., Parmar, J., Pogatchnik, B., Wu, S., Dunnmon, J.A., Zhang, H.R., Rubin, D., Ré, C.: Observational supervision for medical image classification using gaze data. In: Medical Image Computing and Computer Assisted Intervention. pp. 603–614. Springer (2021)
18. Shen, W., Peng, Z., Wang, X., Wang, H., Cen, J., Jiang, D., Xie, L., Yang, X., Tian, Q.: A survey on label-efficient deep image segmentation: Bridging the gap between weak supervision and dense prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(8), 9284–9305 (2023)
19. Tian, Z., Shen, C., Wang, X., Chen, H.: BoxInst: High-performance instance segmentation with box annotations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5439–5448 (2021)
20. Wang, B., Aboah, A., Zhang, Z., Pan, H., Bagci, U.: GazeSAM: Interactive image segmentation with eye gaze and segment anything model. In: Proceedings of The 2nd Gaze Meets ML workshop. vol. 226, pp. 254–265. PMLR (2024)
21. Wang, C., Zhang, D., Ge, R.: Eye-guided dual-path network for multi-organ segmentation of abdomen. In: Medical Image Computing and Computer Assisted Intervention. pp. 23–32. Springer (2023)
22. Wei, J., Hu, Y., Cui, S., Zhou, S.K., Li, Z.: WeakPolyp: You only look bounding box for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention. pp. 757–766. Springer (2023)
23. Wu, L., Zhong, Z., Fang, L., He, X., Liu, Q., Ma, J., Chen, H.: Sparsely annotated semantic segmentation with adaptive gaussian mixtures. In: Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15454–15464 (2023)
24. Wu, S., Zhang, X., Wang, B., Jin, Z., Li, H., Feng, J.: Gaze-Directed Vision GNN for mitigating shortcut learning in medical image. In: Medical Image Computing and Computer Assisted Intervention. pp. 514–524. Springer (2024)
25. Xie, J., Zhang, Q., Cui, Z., Ma, C., Zhou, Y., Wang, W., Shen, D.: Integrating eye tracking with grouped fusion networks for semantic segmentation on mammogram images. IEEE Transactions on Medical Imaging **44**(2), 868–879 (2025)
26. Zhai, S., Wang, G., Luo, X., Yue, Q., Li, K., Zhang, S.: PA-Seg: Learning from point annotations for 3D medical image segmentation using contextual regularization and cross knowledge distillation. IEEE Transactions on Medical Imaging **42**(8), 2235–2246 (2023)
27. Zhang, K., Zhuang, X.: CycleMix: A holistic strategy for medical image segmentation from scribble supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11656–11665 (2022)
28. Zhang, K., Zhuang, X.: ShapePU: A new PU learning framework regularized by global consistency for scribble supervised cardiac segmentation. In: Medical Image Computing and Computer Assisted Intervention. pp. 162–172. Springer (2022)
29. Zhang, X., Sun, K., Wu, D., Xiong, X., Liu, J., Yao, L., Li, S., Wang, Y., Feng, J., Shen, D.: An anatomy- and topology-preserving framework for coronary artery segmentation. IEEE Transactions on Medical Imaging **43**(2), 723–733 (2024)
30. Zhong, Y., Tang, C., Yang, Y., Qi, R., Zhou, K., Gong, Y., Heng, P.A., Hsiao, J.H., Dou, Q.: Weakly-supervised medical image segmentation with gaze annotations. In: Medical Image Computing and Computer Assisted Intervention. pp. 530–540. Springer (2024)