Enjoying Information Dividend: Gaze Track-based Medical Weakly Supervised Segmentation

Zhisong Wang[1,2⋆], Yiwen Ye[1,2⋆], Ziyang Chen[1,2], and Yong Xia[1,2,3(✉)]

[1] Ningbo Institute of Northwestern Polytechnical University, Ningbo 315048, China
[2] National Engineering Laboratory for Integrated Aero-Space-Ground-Ocean Big Data Application Technology, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[3] Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, China
yxia@nwpu.edu.cn

Abstract. Weakly supervised semantic segmentation (WSSS) in medical imaging struggles with effectively using sparse annotations. One promising direction for WSSS leverages gaze annotations, captured via eye trackers that record regions of interest during diagnostic procedures. However, existing gaze-based methods, such as GazeMedSeg, do not fully exploit the rich information embedded in gaze data. In this paper, we propose GradTrack, a framework that utilizes physicians' gaze track, including fixation points, durations, and temporal order, to enhance WSSS performance. GradTrack comprises two key components: (1) the Gaze Track Map Generation module for creating hierarchical attention maps, and (2) the Track Attention module for integrating attention features, which collaboratively enable progressive feature refinement through multi-level gaze supervision during the decoding process. Experiments on the Kvasir-SEG and NCI-ISBI datasets demonstrate that our GradTrack consistently outperforms existing gaze-based methods, achieving Dice score improvements of 3.21% and 2.61%, respectively. Moreover, GradTrack significantly narrows the performance gap with fully supervised models, such as nnUNet.

Keywords: Eye-tracking · Gaze Supervision · Segmentation.

1 Introduction

Fully supervised medical image segmentation methods require pixel-wise annotations, which are time-consuming and labor-intensive. As a result, weakly supervised semantic segmentation (WSSS), which leverages sparse annotations, has emerged as a promising alternative to significantly reduce annotation costs. Such sparse annotations include image-level labels [25, 19, 26], bounding boxes [6, 20], point annotations [5], and scribbles [4, 14, 16, 23]. Recently, advancements in human-computer interaction technologies [7] have enabled physicians to use eye

⋆ Z. Wang and Y. Ye contributed equally. Corresponding author: Y. Xia.

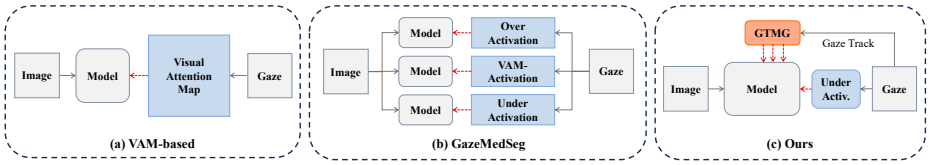

Fig. 1. Three frameworks of gaze-based WSSS. (a) VAM-based: Directly using VAM as supervision. (b) GazeMedSeg: Applying multi thresholds to VAM to generate over- and under-activation maps for joint supervision. (c) GradTrack: Using multiple gaze track attention maps to provide stronger information for overly under-activation maps. "Activ.": Abbreviation of activation.

trackers to scan regions of interest (ROIs) during diagnostic procedures. These eye trackers provide gaze-based annotations [18], marking fixation points along with their respective durations and start timestamps, offering an efficient alternative to traditional annotations without additional effort, thus further reducing annotation costs.

Existing gaze-based methods focus on utilizing gaze annotations as auxiliary priors to enhance segmentation [11], classification [22], and detection tasks [12]. For segmentation[3], a common approach is to convert gaze data into a visual attention map (VAM) (see Fig. 1(a)) by associating the duration of fixation points with their level of importance, typically through the convolution of an isotropic Gaussian kernel over the fixation points. Regions around fixation points with longer durations are given higher weights. The heatmap can also be refined using dense conditional random fields (D-CRF) [13]. However, the VAM-based approach often introduces noisy and redundant information, leading to performance degradation. To our knowledge, GazeMedSeg [27] is the first gaze-based WSSS method in medical image segmentation. It applies multiple thresholds to VAM refined by D-CRF to generate over- or under-activated hard masks and trains multiple expert models to learn from them. While this method helps mitigate noise and redundancy, the limited information derived from gaze annotations still constrains GazeMedSeg's effectiveness. We argue that **both fixation points and their durations are valuable, but the order of these fixation points, inferred from their start timestamps, must also be considered**. During diagnostic imaging, physicians dynamically shift their attention across different ROIs, progressively refining their conclusions. These gaze patterns inherently capture the diagnostic reasoning process, encompassing initial analysis, intermediate cognitive processing, and final decision-making [1, 8].

Motivated by these insights, we propose GradTrack, a **Grad**ual supervision framework based on gaze **Track**s for medical image segmentation. In contrast to GazeMedSeg, we use a more conservative thresholding strategy to achieve more reliable results, even at the cost of increased under-segmentation. Additionally, we introduce the gaze track map generation (GTMG) module and track attention (TA) module to guide the model in progressively refining its segmentation predictions. The GradTrack framework is shown in Fig. 1(c). Specifically, the

GTMG module generates track maps based on fixation points and their start timestamps. By truncating the gaze sequence from the last fixation point, we create multiple track attention maps, which are then transformed into soft attention maps via distance-based calculations. These soft track maps are treated as supervision signals to be fed into the TA modules located within the decoder to guide the model. This approach enables the model to predict track attention maps and integrate them into the decoding process, providing strong prior information for segmentation. Unlike the physician's diagnostic workflow, we argue that noise has a greater impact in the early decoding process. Therefore, we reverse the diagnostic procedure, gradually incorporating more uncertain information until a complete track map is constructed. To evaluate GradTrack, we conduct extensive experiments on two public datasets with different modalities: the Kvasir-SEG dataset [10] for polyp segmentation in endoscopic images, and the NCI-ISBI dataset [2] for prostate segmentation in T2-weighted MRI scans.

Our contributions are as follows. (1) We conduct an in-depth analysis of prior works, highlighting the under-utilization of gaze annotations, particularly neglecting the temporal information. (2) We design the GTMG module to generate multiple track attention maps, which are served as supervision signals to provide guidance. We also devise the TA module to subsequently integrate segmentation predictions. (3) We present a comprehensive evaluation of GradTrack, comparing it with multiple existing WSSS methods, demonstrating the superiority of GradTrack over other methods.

2 Methods

2.1 Overview

GradTrack adopts a U-Net model [17] as the backbone, consisting of an encoder, a decoder, and a segmentation head. For a gaze annotation, GradTrack first generates multiple gaze track attention maps by using the GTMG module. Then, GradTrack uses the TA module to learn to predict these maps using the outputs of the decoder blocks as inputs. The predictions obtained by the TA module are added into the decoding process for providing strong prior information for segmentation. The pipeline of our GradTrack was illustrated in Fig. 2. We now delve into the details of each part.

2.2 Gaze Track Map Generation Module

Generating gaze track attention maps aims to simulate the diagnosis of physicians. As shown in Fig. 2 (c), we first generate the gaze track T according to fixation points and their start timestamps. Then, we reversely truncate T with 50%, 75%, and 100% ratios, creating $T^{50\%}$, $T^{75\%}$, and $T^{100\%}$, respectively. Then, we compute the Euclidean distance from each pixel p to the nearest point t on the gaze track T^r with ratio r, resulting in a distance map D^r as follows:

$$D^r(p) = \min_{t \in T^r} \|p - t\|_2. \tag{1}$$

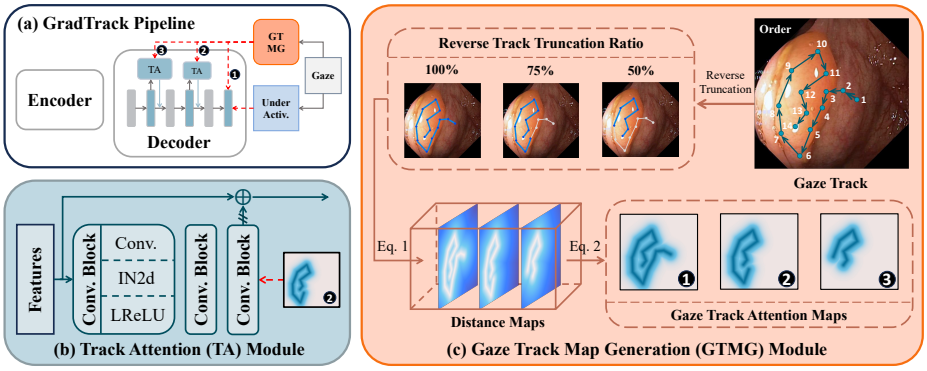

Fig. 2. Overview of the proposed GradTrack. (a) The training pipeline of GradTrack, where the red dashed lines denote the supervision information. (b) Track Attention (TA) Module: Composed of three convolutional blocks, it enhances the model's prior information by truncating the learned GTMG information and fusing it with the main features. (c) Gaze Track Map Generation (GTMG) Module: Generates gaze attention maps by applying different reverse truncation ratios to the gaze track and using a distance-based exponential decay function to enrich the supervision information. "Activ.": Abbreviation of activation.

Subsequently, we utilize an exponential decay function to convert each D^r into the gaze track attention map G^r, which can be is defined as:

$$G^r = e^{-D^r/\beta} \cdot \mathbb{I}(e^{-D^r/\beta} > \tau), \tag{2}$$

where β controls the decay rate, τ represents the field of view threshold, and $\mathbb{I}(\cdot)$ is an indicator function. It ensures that only regions with sufficient confidence are retained in G^r.

2.3 Track Attention Module

To introduce gaze track priors into the model, we design the TA module. The TA modules are added into the decoder blocks. The output features F^h of the 2nd and 4th decoder blocks are separately processed through a TA module to generate a prediction P^h, while the 6th (last) block directly generates P^h. Each P^h consists of two channels, denoting background prediction (P^h_{bg}) and foreground prediction (P^h_{fg}), where h denoting the block number. The gaze track attention maps generated by the GTMG module are used to supervise the prediction by:

$$\mathcal{L}^{r,h}_{GTMG} = -\Big[(1 + G^r_{re}) \cdot \mathbb{I}(G^r_{re} > 0) \cdot \log(P^h_{fg}) + \mathbb{I}(G^r_{re} = 0) \cdot \log(P^h_{bg})\Big], \tag{3}$$

where G^r_{re} denotes the bilinearly interpolated version of G^r, resized to match the spatial resolution of P^h.

For $h = 2/4$, we extend the channel dimension of P_{fg}^{h} from 1 to the one of F^{h}. F^{h} is further enhanced by P_{fg}^{h} through element-wise fusion, formulated as:

$$\hat{F}^{h} = F^{h} \oplus |P_{fg}^{h}|, \tag{4}$$

where $\oplus$ means element-wise adding, and $|\cdot|$ is the stop-gradient operation.

2.4 Optimization of GradTrack

For the under-activation hard pseudo ground truths, generated by applying a more conservative threshold (*i.e.*, 0.7) to $\text{VAM}_{\text{D-CRF}}$ to preserve a smaller portion of the activation information. In contrast, GazeMedSeg employs thresholds of 0.3 and 0.2 for the Kvasir-SEG and NCI-ISBI datasets, respectively.

For the basic overly under-activation hard label supervision, we employ the sum of Dice loss and cross-entropy loss as the supervision mechanism, denoted as $\mathcal{L}_{\text{VAM}}$. The overall loss function integrates all individual losses into a unified optimization objective, formulated as:

$$\mathcal{L} = \mathcal{L}_{GTMG}^{100\%,6} + \lambda_1 \cdot \mathcal{L}_{GTMG}^{50\%,2} + \lambda_1 \cdot \mathcal{L}_{GTMG}^{75\%,4} + \lambda_2 \cdot \mathcal{L}_{VAM}, \tag{5}$$

where λ_1 and λ_2 are balancing coefficients that control the contribution of each term. This loss function ensures multi-level optimization, enabling the model to effectively learn from both gaze track information and visual attention mechanism.

3 Experiments and Results

3.1 Datasets and Evaluation Metric

Datasets. We evaluated the performance of our proposed GradTrack model on two publicly available medical imaging datasets: Kvasir-SEG for polyp segmentation and NCI-ISBI for T2-weighted MRI prostate segmentation. The Kvasir-SEG dataset comprises 900 training images and 100 test images, while the NCI-ISBI dataset includes 789 training images and 117 test images. The training annotations were meticulously generated by annotators with professional training in eye-tracking technology, ensuring high-quality gaze-based supervision for weakly supervised learning.

Evaluation Metric. The Dice similarity coefficient (Dice) is used to evaluate segmentation performance by measuring the overlap between the predicted segmentation and the ground truth.

Table 1. Performance comparison of two fully supervised, ten WSSS methods, and our GradTrack on the Kvasir-SEG and NCI-ISBI datasets across three trial seeds. The best results except for fully supervised methods are highlighted in **bold**. "Std.": Abbreviation of standard deviation.

Method	Supervision	Kvasir-SEG		NCI-ISBI	
		Mean	Std.	Mean	Std.
U-Net [17]	Full	82.12	1.11	80.58	0.48
nnU-Net [9]	Full	88.41	0.47	82.43	0.18
PointSup [5]	Point	73.05	1.64	73.46	4.71
AGMM [24]	Point	75.57	0.84	73.86	1.26
AGMM [24]	Scribble	67.23	1.02	72.70	1.03
USTM [15]	Scribble	66.31	0.93	56.89	1.23
DMPLS [16]	Scribble	69.23	0.32	57.44	0.56
BoxInst [20]	Box	65.72	2.97	73.78	1.15
BoxTeacher [6]	Box	73.33	1.30	75.60	1.15
VAM	Gaze	72.21	0.71	73.21	0.47
$\text{VAM}_{\text{D-CRF}}$ [13]	Gaze	73.12	0.60	73.86	1.91
GazeMedSeg [27]	Gaze	77.80	1.02	77.64	0.57
GradTrack (Ours)	Gaze	**81.01**	0.66	**80.25**	0.40

3.2 Implementation Details

We employ nnU-Net [9] as the backbone architecture, optimized using Stochastic Gradient Descent (SGD) with a batch size of 8. The initial learning rate l_0 was set to 0.01 and decayed according to the polynomial rule [9] $l_t = l_0 \times (1 - t/T)^{0.9}$, where t is the current epoch and T is the maximum epoch, which was set to 100. All input images are preprocessed by resizing to 224×224 resolution. The balancing coefficients λ_1 and λ_2 are both set to 0.5, empirically. The decay rate β is set to 10 for the Kvasir-SEG dataset and 15 for the NCI-ISBI dataset. The field of view threshold τ is set to 0.25 for both datasets.

3.3 Results

Comparison with state-of-the-art WSSS methods. We evaluate our model against multiple state-of-the-art WSSS methods, including point-supervised approaches (PointSup [5], and AGMM [4]), scribble-supervised methods (USTM [15], AGMM [4], and DMPLS [16]), bounding box-supervised methods (Boxinst [20], and BoxTeacher [6]), and gaze-supervised methods (VAM, $\text{VAM}_{\text{D-CRF}}$ [13], and GazeMedSeg [27]). The scribble annotations in AGMM adopt the style proposed by [21], while DMPLS and USTM follow the annotation style introduced by [16]. For gaze-based methods, VAM and $\text{VAM}_{\text{D-CRF}}$ normalize the VAM maps by treating regions with values below 0.5 as background in both the Kvasir-SEG and NCI-ISBI datasets, utilizing these normalized soft labels for supervision. Notably, the results of PointSup, AGMM, Boxinst, BoxTeacher, and GazeMedSeg are inherited from [27].

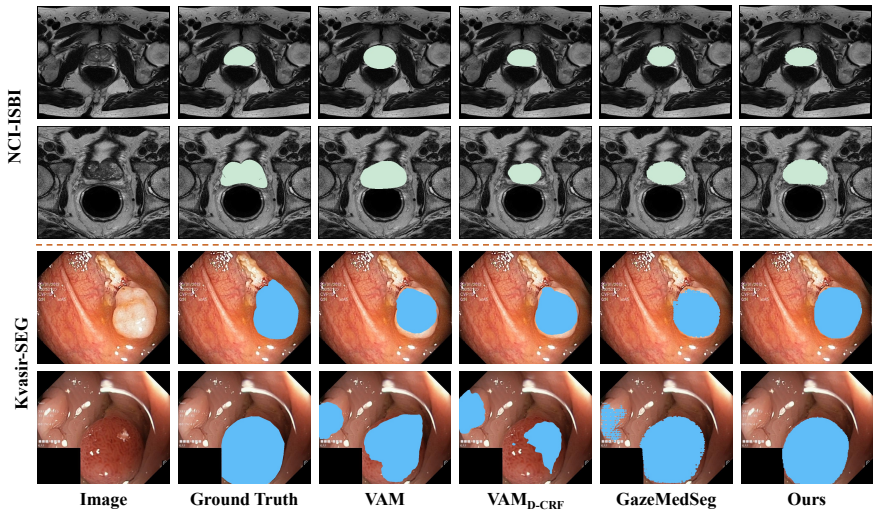

Fig. 3. Visualization of segmentation results obtained using gaze-based methods, including VAM, VAM$_{\text{D-CRF}}$, GazeMedSeg, and GradTrack, on the NCI-ISBI and Kvasir-SEG datasets.

Table 2. Performance ($mean_{std}$) of our GradTrack and four variants on the Kvasir-SEG dataset across three trial seeds. The best result is highlighted in **bold**. "Enc.": Abbreviation of encoder. "Dec.": Abbreviation of decoder.

Method	Truncation		TA Location		
	w/o	Sequential	Enc. + Dec.	Enc.	Dec. (Ours)
Dice	$80.28_{0.02}$	$79.16_{0.19}$	$80.09_{0.53}$	$79.96_{0.30}$	$\mathbf{81.01}_{0.66}$

The results shown in Table 1 reveal that our GradTrack outperforms all gaze-based supervision methods, achieving the highest performance across both datasets. Specifically, it improves the Dice score by 3.21% on Kvasir-SEG and 2.61% on NCI-ISBI compared to the best comparing gaze-based approach, *i.e.*, GazeMedSeg. Our GradTrack exhibits remarkably competitive performance compared to fully supervised methods, achieving results only 1.11% and 0.33% lower than the fully supervised U-Net on the Kvasir-SEG and NCI-ISBI datasets, respectively. This demonstrates the effectiveness of our GradTrack in narrowing the gap between WSSS methods and fully supervised methods.

For qualitative analysis, we select two images from each dataset to compare the segmentation results with other gaze-based supervision methods, alongside the ground truth. The segmentation comparison is shown in Fig. 3. These results demonstrate that our segmentation results are the closest to the ground truth. Notably, in the second row of the Kvasir-SEG dataset, it can be observed that other methods exhibit overconfidence in segmenting polyps, whereas our model does not produce any erroneous predictions.

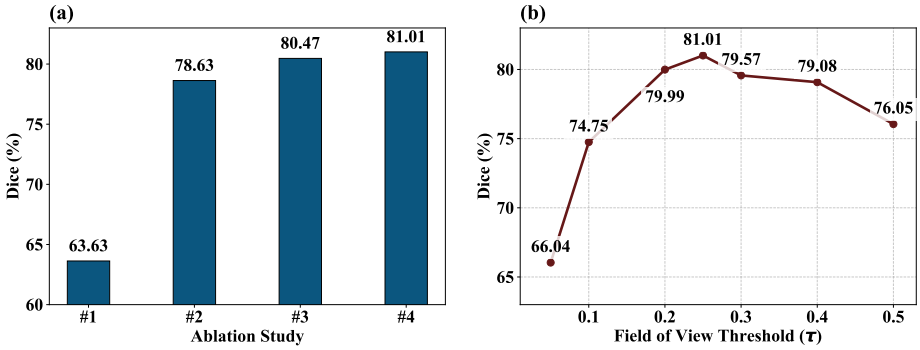

Fig. 4. Ablation study and hyper-parameter τ discussion on the Kvasir-SEG dataset. (a) Analyzing the contribution of each component within our GradTrack. (b) Performance of our GradTrack with various values of τ.

Ablation study We conducted an ablation study to assess the effectiveness of each component. The results are presented in Fig. 4(a), where #1 refers to using only the under-activated VAM as supervision, #2 indicates adding the track attention maps generated by the GTMG module for supervision at the final network layer, #3 extends this by using the TA module, and #4 represents our full design, with two reverse-truncated track weighted maps guiding the TA modules to further enhance the decoder's ability. It reveals that (1) using only the under-activated VAM leads to a poor performance; (2) introducing both GTMG and TA modules can improve the segmentation; (3) the best performance is achieved when they are jointly used (*i.e.*, our GradTrack).

In Table 2, we further analyze the impact from two aspects: track attention map truncation and the placement of the TA module, resulting in four variants. The results demonstrate that (1) deeper track information is more beneficial for guiding the model's learning; (2) more accurate information at deeper network layers is more effective than redundant information; and (3) training the model to learn supervisory information in encoder may have a negative effect on the feature extraction ability of the model.

Sensitivity to field of view thresholds. For the track attention maps, the coverage of each track determines the upper limit of the information the model can absorb. By analyzing the field of view range, we present in Figure 4 the model's performance under different field of view thresholds, including 0.05, 0.1, 0.2, 0.25, 0.3, 0.4, and 0.5. When the threshold is set to 0.25, the model absorbs information most effectively. A threshold as small as 0.05 introduces excessive redundancy, disrupting the model's perception, while a threshold too large, such as 0.5, exposes the model to insufficient data, hindering its ability to learn useful information.

4 Conclusion

We propose GradTrack, a gaze-based weakly supervised segmentation model, consisting of GTMG and TA modules. The GTMG module converts gaze data into the gaze track attention map using reverse truncation and a distance-based exponential decay function, and the TA module provides guidance to the decoder under the multi-level supervision. Extensive experiments on the Kvasir-SEG and NCI-ISBI datasets demonstrate superior performance of our GradTrack, which can significantly reduce pixel-level annotation requirements and decrease the gap with fully supervised methods.

Acknowledgments. This work was supported in part by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang, China, under Grant 2025C01201(SD2), in part by the National Natural Science Foundation of China under Grants 62171377 and 92470101, in part by the Ningbo Clinical Research Center for Medical Imaging under Grant 2021L003 (Open Project 2022LYKFZD06), in part by Shenzhen Science and Technology Program under Grants JCYJ20220530161616036, and in part by the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under Grants CX2024016 and CX2025020.

Disclosure of Interests. The authors declare no relevant competing interests.

References

1. Bisogni, C., Nappi, M., Tortora, G., Del Bimbo, A.: Gaze analysis: A survey on its applications. Image and Vision Computing **144**, 104961 (2024)
2. Bloch, 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. The Cancer Imaging Archive (2015). https://doi.org/http://doi.org/10.7937/K9/TCIA.2015.zF0vlOPv
3. Chen, J., Ma, B., Cui, H., Xia, Y.: Think twice before selection: Federated evidential active learning for medical image analysis with domain shifts. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11439–11449 (2024)
4. 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 (2025)
5. 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)
6. 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)
7. Ghosh, S., Dhall, A., Hayat, M., Knibbe, J., Ji, Q.: Automatic gaze analysis: A survey of deep learning based approaches. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(1), 61–84 (2023)

8. 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 (2024)
9. 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)
10. 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: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26. pp. 451–462. Springer (2020)
11. Jiang, H., Gao, M., Liu, Z., Tang, C., Zhang, X., Jiang, S., Yuan, W., Liu, J.: Glanceseg: Real-time microaneurysm lesion segmentation with gaze-map-guided foundation model for early detection of diabetic retinopathy. IEEE Journal of Biomedical and Health Informatics (2024)
12. Kong, Y., Wang, S., Cai, J., Zhao, Z., Shen, Z., Li, Y., Fei, M., Wang, Q.: Gaze-detr: Using expert gaze to reduce false positives in vulvovaginal candidiasis screening. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 133–143. Springer (2024)
13. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. Advances in neural information processing systems **24** (2011)
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 (2024)
15. Liu, X., Yuan, Q., Gao, Y., He, K., Wang, S., Tang, X., Tang, J., Shen, D.: Weakly supervised segmentation of covid19 infection with scribble annotation on ct images. Pattern recognition **122**, 108341 (2022)
16. Luo, X., Hu, M., Liao, W., Zhai, S., Song, T., Wang, G., Zhang, S.: Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 528–538. Springer (2022)
17. 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)
18. Song, Y., Wang, X., Yao, J., Liu, W., Zhang, J., Xu, X.: Vitgaze: gaze following with interaction features in vision transformers. Visual Intelligence **2**(1), 1–15 (2024)
19. Tang, F., Xu, Z., Qu, Z., Feng, W., Jiang, X., Ge, Z.: Hunting attributes: Context prototype-aware learning for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3324–3334 (2024)
20. 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. 5443–5452 (2021)
21. Valvano, G., Leo, A., Tsaftaris, S.A.: Learning to segment from scribbles using multi-scale adversarial attention gates. IEEE Transactions on Medical Imaging **40**(8), 1990–2001 (2021)
22. Wang, B., Pan, H., Aboah, A., Zhang, Z., Keles, E., Torigian, D., Turkbey, B., Krupinski, E., Udupa, J., Bagci, U.: Gazegnn: A gaze-guided graph neural network for chest x-ray classification. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2194–2203 (2024)

23. Wang, Z., Ye, Y., Chen, Z., Shu, M., Xia, Y.: From few to more: Scribble-based medical image segmentation via masked context modeling and continuous pseudo labels. arXiv preprint arXiv:2408.12814 (2024)
24. 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)
25. Zhang, S., Zhang, J., Xie, Y., Xia, Y.: Tpro: Text-prompting-based weakly supervised histopathology tissue segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 109–118. Springer (2023)
26. Zhao, X., Tang, F., Wang, X., Xiao, J.: Sfc: Shared feature calibration in weakly supervised semantic segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 7525–7533 (2024)
27. 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: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 530–540. Springer (2024)