Synchronous Inhibition and Activation for Weakly Supervised Semantic Segmentation of Pathology Images

Jiansong Fan[1], Yicheng Di[1], Jiayu Bao[1], Lihua Li[2], and Xiang Pan[1](✉)

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

[2] Institute of Biomedical Engineering and Instrumentation, Hangzhou Dianzi University, Hangzhou, 310018, China

xiangpan@jiangnan.edu.cn

Abstract. Tissue-level semantic segmentation is crucial in digital pathology workflow. However, since dense pixel-level annotation of gigapixel pathology images is expensive and time-consuming, Weakly Supervised Semantic Segmentation (WSSS) methods have gradually attracted attention. The WSSS methods using image-level labels usually rely on Class Activation Map to generate pseudo labels, which have difficulty capturing complete object regions and may incorrectly activate regions with weak semantic relevance of pathology images. In this work, we propose SIA-WSSS, a weakly supervised semantic segmentation model for pathology images that synchronous inhibition and activation. Specifically, we first extract pathology images class and patch tokens using a VisionTransformer (ViT) and construct a Regularized Focus Mechanism (RFM). The RFM implicitly regularizes class-patch interactions through graph learning, ensuring that class tokens can dynamically compress patch information and inhibit irrelevant backgrounds. Next, we introduce a Discriminative Activation Module to contrast the class tokens of fine-grained regions and global objects to capture the unique features of each class and activate the foreground region. Moreover, we design a Regional Self-modulation Module synchronizing each region's activation and inhibition information to generate segmentation results with finer structures. Experimental results on the LUAD-HistoSeg and BCSS-WSSS datasets demonstrate that the proposed SIA-WSSS significantly outperforms state-of-the-art WSSS methods. The code is available at https://github.com/Jsf826/SIA-WSSS.

Keywords: Weakly Supervised Segmentation · Synchronous · Inhibition and Activation · Pathology Images.

1 Introduction

Pathology images are significant for the clinical diagnosis and prognosis of diseases. With the rapid development of artificial intelligence technology in the past decade, the automatic analysis of pathology images has achieved performance

comparable to that of human pathologists in some tasks [8, 17]. However, most methods are based on supervised learning, and their performance mainly depends on many training samples with detailed annotations. These annotations usually require experienced pathologists, are expensive to obtain, and are prone to human errors. In particular, pixel-level annotations on billions of pixels of pathology images are unacceptable.

Compared with supervised and unsupervised learning, Weakly Supervised Semantic Segmentation (WSSS) methods only need coarse-grained annotations to perform semantic segmentation automatically [14,16]. According to the degree of coarse-grained labels, the annotations of WSSS can be divided into image annotations [13,15], bounding box annotations [9,18], and point annotations [25]. In this work, our motivation is to segment regions at the pixel level in pathology images and rely only on image-level labels.

Generally, the pipeline of WSSS can be divided into three steps. It first trains a classification network to generate class activation maps (CAMs) with image-level labels [11, 26]. The CAM is then refined into pseudo-labels, which are further used to provide dense supervision to retrain the segmentation model [1, 23]. However, these methods tend to distinguish objects by the most discriminative features, resulting in the activated area gradually shrinking and failing to identify the complete object area, significantly weakening the performance of WSSS. Recently, Vision Transformer (ViT) [2] is well-known for establishing long-range dependencies and has been widely adopted in WSSS. Benefiting from the self-attention mechanism, several studies have shown that attention maps between class and patch tokens can reliably highlight objects and generate accurate localization maps (LAMs) [4, 12].

However, ViT tends to aggregate global semantics in low-information patches to capture long-range dependencies. It causes irrelevant patches to be frequently associated with class tokens during attention, resulting in many weakly semantically related regions being misactivated and severely damaging the quality of LAMs [10, 20].

In this work, we propose a novel framework named SIA-WSSS for weakly supervised segmentation of pathology images. It comprehensively considers activation and inhibition information in weakly supervised segmentation of pathology images and synchronizes them when generating segmentation results. Specifically, our SIA-WSSS first extracts class and patch tokens using ViT. Subsequently, we design a Regularized Focus Mechanism (RFM) to implicitly regularize class-patch interactions through graph representation learning, ensuring that the model can dynamically compress patch information and inhibit irrelevant background. The RFM represents the class-patch tokens as a directed graph, dynamically updates the graph node knowledge and aggregates independent inhibition information to the class tokens. Next, we introduce a Discriminative Activation Module (DAM), which compares the class tokens of fine-grained regions and global objects in pathology images to activate the foreground region fully. Finally, we design a Region Self-modulation Module (RSM) to synchronize the uncertainty region inhibition information and the foreground activation in-

formation to generate more refined segmentation results. We use two datasets, BCSS-WSSS [11] and LUAD-HistoSeg [3], to verify the effectiveness of our SIA-WSSS in the weakly supervised pathology image segmentation task. In summary, the main contributions of this work are as follows:

•We propose a weakly supervised semantic segmentation model with synchronous inhibition and activation information, which can use image-level labels to finely segment pathology images.

•We design a Regularized Focus Mechanism, which ensures that class tokens can compress patch information and inhibit irrelevant background through implicit regularization dynamic updates of class-patch directed graphs.

•We introduce a Discriminative Activation Module to contrast the class tokens of fine-grained regions and global objects to capture the unique features of each class and activate foreground regions.

•We present a Regional Self-modulation Module that synchronously inhibition and activation to generate segmentation results with delicate structures.

2 Methodology

The detailed structure of our proposed SIA-WSSS is shown in Fig. 1. The SIA-WSSS includes the ViT Encoder, Regularized Focus Mechanism (RFM), Discriminative Activation Module (DAM), and Region Self-modulation Module (RSM). Given an input image X and classification label Y, we first use ViT to generate patch tokens P and class tokens T and send them to RFM. The RFM condenses the relevant patch semantics into a directed graph and generates a feature map I with inhibition information by regularizing the attention among class-patch tokens. Subsequently, we fed the input image X and the cropped X with shared ViT to obtain the input fine-grained and coarse-grained class tokens of DAM. Subsequently, DAM can obtain the contrast activation map A by consistently learning fine-grained and coarse-grained class tokens. Finally, RSM combines the contrast activation map A and the inhibition information map I to generate a segmentation result with a finer structure. The following sections provide a detailed description of these steps.

2.1 Regularized Focus Mechanism

The SIA-WSSS adopts ViT as the backbone to extract pathology image features (Fig. 1a). Given a pathology image X and a classification label Y, the ViT can generate patch tokens P and class tokens T. Subsequently, the Regularized Focus Mechanism constructs class-patch tokens as a novel directed graph structure to suppress the false activations between class-patch tokens. As shown in Fig. 1b, we first use a linear projector to transform patch tokens P into heads and tails, where heads simulate the correlation with patches and tails represent the contribution of patches to heads. In addition, we customize the edge embedding e_{ij} for P and select candidate neighbours to model the relationship between heads h_i and tails t_j. The same operation also happens for class tokens T. After

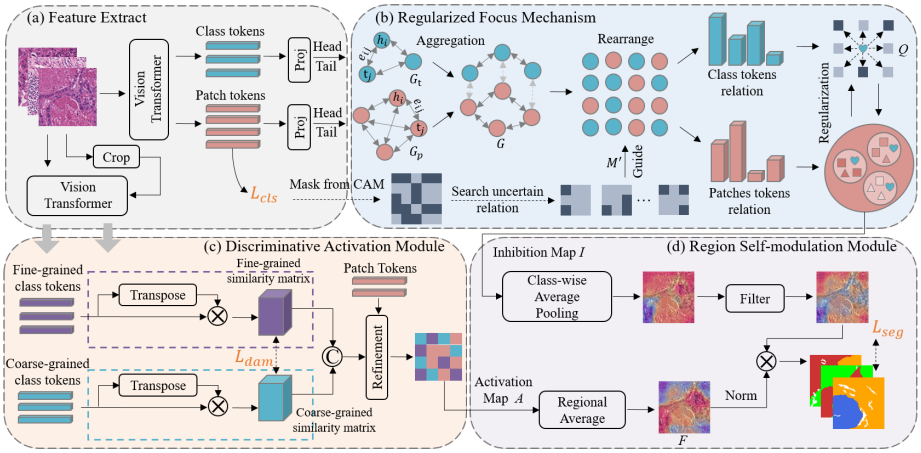

Fig. 1. The overview of the proposed SIA-WSSS. (a) Feature extraction; (b) Regularized Focus Mechanism; (c) Discriminative Activation Module; (d) Region Self-modulation Module.

that, we can get the patch token graph $G_p = \{V_p, R_p, Z_p, E_p\}$ and the class token graph $G_t = \{V_t, R_t, Z_t, E_t\}$, where V is the node corresponding to the class and patch token, Z represents the head-tail, E represents the edge, and R represents the directed information on the directed edge. Based on the constructed patch token graph G_p and class token graph G_t, we further aggregate them into a dynamic directed graph $G = \{G_p, G_t\}$. Then, we use the dynamic directed graph G to generate the graph regularized class token Q. The generation strategy here is expressed as follows:

$$Q = \alpha_1 \left(w_1 \left(h_i + a_i \right) \right) + \alpha_2 \left(w_2 \left(a_i \odot h_i \right) \right), \tag{1}$$

where α_1/α_2 represents LeakyReLU, W_1/W_2 is the projection matrix, and $a_i = softmax(t_j^T \tanh(h_i + e_{ij}))$ is the weighting factor used to quantify the knowledge weight between nodes during the aggregation process.

Moreover, we further utilize CAM as a prior to search for the uncertainty region M and adjust the class-patch representation in a learnable manner. As shown in Fig. 1a, we utilize L_{cls} to maintain the classification head to generate CAM by projecting the classification matrix onto the patch tokens P. Then, we adopt a threshold-based search strategy to refine the CAM into foreground ($threshold > \lambda_1$), background ($threshold < \lambda_2$), and uncertainty regions (otherwise), the threshold $0 < \lambda_1 < \lambda_2 < 1$. In order to prevent irrelevant regions from being erroneously activated, we extract the uncertain regions from M as the inhibition confidence relationship map M' to guide the class-patch representation. Specifically, for the patch token p_{ij} in P, we calculate the cosine similarity with the class token q and use the class index on M' as the relationship supervision. If the class token $q \in Q$ has the same class index as the pixel (i, j) on M',

the correlation between q and p_{ij} is considered to be negatively correlated and suppressed. Finally, we can obtain the inhibition information map I with the inhibition information of the uncertain region.

2.2 Discriminative Activation Module

Using RFM resolves the issue of incorrect activation in weak confidence regions. However, the defect of failing to identify complete object regions is still unavoidable. Therefore, we design a Discriminative Activation Module(DAM) to promote the representation consistency between fine-grained objects and coarse-grained objects in pathology images, which can further enforce the activation of complete object regions in CAM. As shown in Fig. 1c, we randomly crop local images from the global pathology image X and input the shared encoder ViT to obtain fine-grained class tokens and coarse-grained class tokens. Since the class tokens in ViT capture the information of semantic objects [24], fine-grained class tokens and coarse-grained class tokens aggregate the information of global coarse-grained and local fine-grained objects, respectively. Subsequently, minimizing the difference between fine-grained and coarse-grained class tokens can make the representation of the entire object region more consistent. Specifically, the fine-grained class tokens and the coarse-grained class tokens first pass through the projection heads P_f and P_c, respectively, which consist of a linear layer and an L_2 normalization layer. Assuming A_f represents the projection of the Coarse-grained class tokens and A_c represents the projection of the fine-grained class tokens, the goal of DAM is to minimize the difference between A_f and A_c. Here, we use the InfoNCE loss [19] as the target:

$$\mathcal{L}_{dam} = \frac{1}{N} \sum_{A_c} \log \frac{e^{(A_f^\top A_c/\tau)}}{e^{(A_f^\top A_c/\tau)} + \epsilon}, \tag{2}$$

where N represents the number of A_c, τ is the temperature factor, and ϵ is a small positive value. Finally, a more complete contrast activation map A of the object area can be obtained.

2.3 Region Self-modulation Module

We can obtain the inhibition information map I and the contrast activation map A through the above-mentioned Regularized Focus Mechanism and Discriminative Activation Module. In order to give the segmentation result a finer spatial structure, we propose the Region Self-modulation Module to synchronize the inhibition and activation information for further refinement. For the contrast activation map A, we first average its pixels to obtain the regional consistent activation feature representation $F = \frac{ReLU(A)}{\max(A)}$. Then, we apply the Norm operation, including ReLU and maximum normalization, to generate the region-consistent activation map A'.

In addition, we propose a reliable filtering strategy to erase low-confidence regions and keep the activation of high-confidence regions. We first use Class-wise

Average Pooling to extract the inhibition feature representation I' from the inhibition information map I. Then, the object threshold filters out the regions with relatively low attention values. Subsequently, the erosion operation is applied to reduce the weak-activation regions further and obtain the high-confidence activation map I'_h. Finally, the refined segmentation result can be obtained by combining the confidence activation map I'_h and the region-consistent activation map A':

$$Result = \max(A, I'_h \otimes A'). \tag{3}$$

In summary, the total loss of the SIA-WSSS during training can be defined as:

$$L = L_{cls} + \alpha_1 L_{dam} + \alpha_2 L_{seg}. \tag{4}$$

3 Experiments

Dataset: To verify the effectiveness of our proposed SIA-WSSS, we evaluated the weakly supervised pathology image segmentation method on two datasets: BCSS-WSSS [11] and LUAD-HistoSeg [3]. The LUAD-HistoSeg dataset contains 31,826 224×224 pathology images, covering four tissue categories: tumor epithelium (TE), tumor-associated stroma (TAS), necrosis (NEC), and lymphocytes (LYM). The BCSS-WSSS dataset contains 17,286 224×224 pathology images, providing pixel-level annotations of five categories for each ROI, namely tumor (TUM), stroma (STR), lymphocyte infiltration (LYM), necrosis (NEC), and others (OTR).

Table 1. Quantitative comparison with state-of-the-art methods on the BCSS-WSSS and LUAD-HistoSeg datasets. Net denotes the backbone for WSSS methods. mP: mPrecision; mR: mRecall.

Model	Net	BCSS-WSSS				LUAD-HistoSeg			
		mDice	mIoU	mP	mR	mDice	mIoU	mP	mR
HistoSegNet [5]	ResNet34	0.505	0.276	0.582	0.571	0.641	0.478	0.654	0.662
SC-CAM [6]	ResNet50	0.729	0.663	0.742	0.733	0.715	0.641	0.756	0.761
WSSS-Tissue [11]	ResNet101	0.767	0.697	0.786	0.758	0.818	0.756	0.826	0.822
SSC [7]	ResNet101	0.755	0.654	0.727	0.727	0.772	0.731	0.783	0.748
Mctformer+ [22]	ViT-B	0.772	0.707	0.790	0.766	0.825	0.763	0.830	0.825
DuPL [21]	ViT-B	0.781	0.715	0.799	0.782	0.823	0.769	0.838	0.831
Our Model	ViT-B	**0.799**	**0.734**	**0.818**	**0.825**	**0.855**	**0.785**	**0.872**	**0.859**

Competing Methods and Evaluation Metrics: To comprehensively evaluate the proposed method, we compare SIA-WSSS with multiple weakly supervised segmentation methods, including HistoSegNet [5], SC-CAM [6], SSC [7],

WSSS-Tissue [11], Mctformer+ [22], and DuPL [21]. All methods are evaluated using four metrics: mDice, mIoU, mRecall and mPrecision.

Implementation Details: All of our experiments were performed on an NVIDIA GeForce RTX A6000 GPU with 48GB memory and repeated 5 times to calculate the average. The model training uses the Adam optimizer, the weight decay is set to $5e-4$, and the learning rate is fixed to $1e-4$. The batch size is 16, and the λ_1 and λ_2 in RFM are 0.3 and 0.5. The weight parameters α_1 and α_2 of the loss function are set to 0.5, respectively.

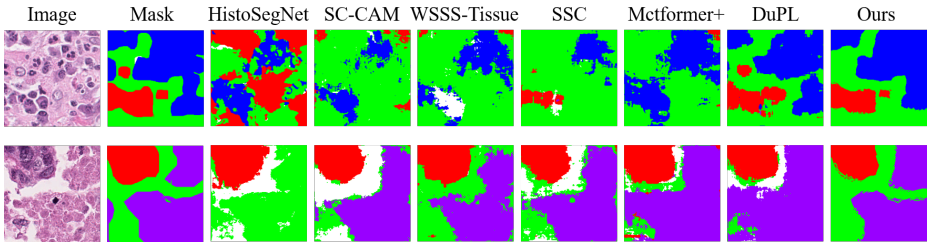

Fig. 2. Visualize the prediction results of our proposed SIA-WSS and the state-of-the-art methods on BCSS-WSSS.

Comparison with SOTA Methods: Table 1 compares the proposed SIA-WSSS quantitatively with recent state-of-the-art methods. Experimental results show that our SIA-WSSS surpasses other state-of-the-art models. We attribute this to the Regularized Focus Mechanism to suppress uncertain regions and the Discriminative Activation Module to activate complete foreground object regions. Specifically, our method produces higher accuracy than weakly supervised segmentation models such as SC-CAM and HistoSegNet (Table 1). Moreover, our SIA-WSSS produces more region-consistent segmentation results (Fig. 2). In addition, our SIA-WSSS significantly outperforms other ViT-based methods in terms of mDice and mIoU, which we attribute to the Region Self-modulation Module's synchronous activation and inhibition information to produce more refined segmentation results. Fig. 3 shows the CAMs generated by the proposed SIA-WSSS on LUAD-HistoSeg, which confirms our model's weakly supervised segmentation capability. In summary, our SIA-WSSS can produce accurate tissue region segmentation results with only image-level labels of pathology images.

Ablation Study: To explore the effectiveness of each component of SIA-WSSS, we conduct ablation studies to select the best settings. Table 2 shows the effects of introducing different components on the segmentation performance (including the Regularized Focus Mechanism (RFM), Discriminative Activation Module

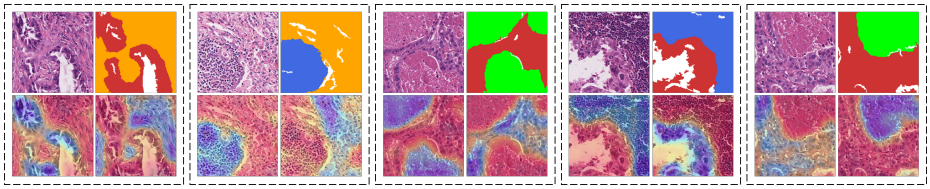

Fig. 3. Visualize the activation map of our SIA-WSS on LUAD-HistoSeg.

(DAM), and Region Self-modulation Module (RSM)). When the inhibition information with uncertain regions generated based on RFM is introduced (model b), the improvement of mDice can be observed, which proves that enhancing the inhibition information in the weak supervision process is beneficial. Model c shows that adding DAM improves the accuracy of segmentation, which proves the necessity of consistent learning of fine-grained and coarse-grained class tokens. In addition, after RSM (model f) is introduced, the model's performance reaches the best because RSM synchronous inhibition and activation information are conducive to refining segmentation results. The results above demonstrate that the SIA-WSSS, which integrates activation and inhibition information, significantly outperforms the baseline model relying solely on activation data.

Table 2. Ablation analysis of different components in the proposed SIA-WSSS on LUAD-HistoSeg.

Model name	Baseline	RFM	DAM	RSM	mDice	mIoU	mPrecision	mRecall
a	✓				0.655	0.495	0.673	0.681
b	✓	✓			0.798	0.740	0.804	0.756
c	✓		✓		0.816	0.752	0.817	0.820
d	✓	✓	✓		0.834	0.771	0.843	0.835
f (Ours)	✓	✓	✓	✓	**0.855**	**0.785**	**0.872**	**0.859**

4 Conclusion

This paper proposes a novel weakly supervised learning method (SIA-WSSS) for pathology images using only image-level labels. The SIA-WSSS explores the inhibition information of uncertain regions and the activation information of foreground regions of objects through the Regularized Focus Mechanism (RFM) and Discriminative Activation Module (DAM). The RFM captures inhibition information from images by implicitly regularizing class-patch directed graphs. At the same time, the DAM exploits the consistency of coarse-grained and fine-grained features of pathology images to activate the complete foreground region fully. In addition, we use the Region Self-modulation Module to coordinate inhibition and activation information better. Experiments show that our SIA-WSSS

can potentially become an effective means of annotating pathology images in clinical applications.

Acknowledgments. This work is supported in part by the National Natural Science Foundation of China under grants W2411054, U21A20521 and 62271178, the Postgraduate Research & Practice Innovation Program of Jiangsu Province KYCX23_2524, National Foreign Expert Project of China under Grant G2023144009L, Zhejiang Provincial Natural Science Foundation of China (LR23F010002), Wuxi Health Commission Precision Medicine Project (J202106), Jiangsu Provincial Six Talent Peaks Project (YY-124), and Major Projects of Wuxi Health Commission (Z202324).

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

References

1. Ahn, J., Kwak, S.: Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4981–4990 (2018)
2. Alexey, D.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv: 2010.11929 (2020)
3. Amgad, M., Elfandy, H., Hussein, H., Atteya, L.A., Elsebaie, M.A., Abo Elnasr, L.S., Sakr, R.A., Salem, H.S., Ismail, A.F., Saad, A.M., et al.: Structured crowdsourcing enables convolutional segmentation of histology images. Bioinformatics **35**(18), 3461–3467 (2019)
4. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9650–9660 (2021)
5. Chan, L., Hosseini, M.S., Rowsell, C., Plataniotis, K.N., Damaskinos, S.: Histosegnet: Semantic segmentation of histological tissue type in whole slide images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10662–10671 (2019)
6. Chang, Y.T., Wang, Q., Hung, W.C., Piramuthu, R., Tsai, Y.H., Yang, M.H.: Weakly-supervised semantic segmentation via sub-category exploration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8991–9000 (2020)
7. Chen, T., Yao, Y., Huang, X., Li, Z., Nie, L., Tang, J.: Spatial structure constraints for weakly supervised semantic segmentation. IEEE Transactions on Image Processing (2024)
8. Cheng, H.T., Yeh, C.F., Kuo, P.C., Wei, A., Liu, K.C., Ko, M.C., Chao, K.H., Peng, Y.C., Liu, T.L.: Self-similarity student for partial label histopathology image segmentation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16. pp. 117–132. Springer (2020)
9. Dai, J., He, K., Sun, J.: Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision. pp. 1635–1643 (2015)
10. Darcet, T., Oquab, M., Mairal, J., Bojanowski, P.: Vision transformers need registers. arXiv preprint arXiv:2309.16588 (2023)

11. Han, C., Lin, J., Mai, J., Wang, Y., Zhang, Q., Zhao, B., Chen, X., Pan, X., Shi, Z., Xu, Z., et al.: Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. Medical Image Analysis **80**, 102487 (2022)
12. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
13. Huang, H., Kang, H., Liu, S., Salvado, O., Rakotoarivelo, T., Wang, D., Liu, T.: Paddles: Phase-amplitude spectrum disentangled early stopping for learning with noisy labels. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16719–16730 (2023)
14. Jia, Z., Huang, X., Eric, I., Chang, C., Xu, Y.: Constrained deep weak supervision for histopathology image segmentation. IEEE transactions on medical imaging **36**(11), 2376–2388 (2017)
15. Li, J., Jie, Z., Wang, X., Wei, X., Ma, L.: Expansion and shrinkage of localization for weakly-supervised semantic segmentation. Advances in Neural Information Processing Systems **35**, 16037–16051 (2022)
16. Li, K., Qian, Z., Han, Y., Eric, I., Chang, C., Wei, B., Lai, M., Liao, J., Fan, Y., Xu, Y.: Weakly supervised histopathology image segmentation with self-attention. Medical Image Analysis **86**, 102791 (2023)
17. Skrede, O.J., De Raedt, S., Kleppe, A., Hveem, T.S., Liestøl, K., Maddison, J., Askautrud, H.A., Pradhan, M., Nesheim, J.A., Albregtsen, F., et al.: Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. The Lancet **395**(10221), 350–360 (2020)
18. Song, C., Huang, Y., Ouyang, W., Wang, L.: Box-driven class-wise region masking and filling rate guided loss for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3136–3145 (2019)
19. Sun, G., Wang, W., Dai, J., Van Gool, L.: Mining cross-image semantics for weakly supervised semantic segmentation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. pp. 347–365. Springer (2020)
20. Sun, M., Chen, X., Kolter, J.Z., Liu, Z.: Massive activations in large language models. arXiv preprint arXiv:2402.17762 (2024)
21. Wu, Y., Ye, X., Yang, K., Li, J., Li, X.: Dupl: Dual student with trustworthy progressive learning for robust weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3534–3543 (2024)
22. Xu, L., Bennamoun, M., Boussaid, F., Laga, H., Ouyang, W., Xu, D.: Mctformer+: Multi-class token transformer for weakly supervised semantic segmentation. IEEE transactions on pattern analysis and machine intelligence (2024)
23. Yang, Z., Meng, Y., Fu, K., Wang, S., Song, Z.: Tackling ambiguity from perspective of uncertainty inference and affinity diversification for weakly supervised semantic segmentation. arXiv preprint arXiv:2404.08195 (2024)
24. Zhang, B., Xiao, J., Wei, Y., Sun, M., Huang, K.: Reliability does matter: An end-to-end weakly supervised semantic segmentation approach. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12765–12772 (2020)
25. Zhao, T., Yin, Z.: Weakly supervised cell segmentation by point annotation. IEEE Transactions on Medical Imaging **40**(10), 2736–2747 (2020)

26. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)