Positive-Unlabeled Learning with Field of View Consistency for Histology Image Segmentation

Xiaoqi Jia[1,3], Chong Fu[⋆1,2], Jiaxin Hou[3,4], and Wenjian Qin[3]

[1] School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China
fuchong@mail.neu.edu.cn
[2] Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Northeastern University, Shenyang 110819, China
[3] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[4] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China

Abstract. Histology image annotation is costly and time-consuming. Utilizing Positive and Unlabeled (PU) data for model training offers a more resource-efficient alternative. However, previous methods for PU learning suffer from the noise arising from label assignment to unlabeled data. We observe that predictions on noisy data lack consistency under data augmentation. In this paper, we present Field of View (FoV) consistency regularization for PU segmentation in histology images, which effectively reduces the noise influence by promoting consistent predictions across varying FoVs. Using only 20% of positive labels on the Glas Dataset, our approach outperforms previous methods, achieving a Dice score of 90.69%—almost reaching the fully supervised result of 93.30%. Source code is available at: https://github.com/lzaya/PU_with_FoV.

Keywords: Histology Image Segmentation, PU Learning, FoV Consistency Regularization

1 Introduction

The advancement of digital pathology in clinical diagnostics has led to an increasing demand for histology image analysis. Deep learning-based segmentation algorithms, such as fully convolutional networks, have achieved remarkable accuracy and efficiency in histology image analysis, fostering progress in disease diagnosis, treatment, and research [18]. However, model training depends on extensive fully annotated data, and the high costs associated with medical image labeling pose challenges in this field. Positive-Unlabeled (PU) data, a weakly labeled dataset type, can help alleviate annotation efforts in histology image segmentation scenarios. It consists of a subset of positive examples (i.e., instances of the class of interest) and unlabeled examples (i.e., instances not labeled as

⋆ Corresponding author

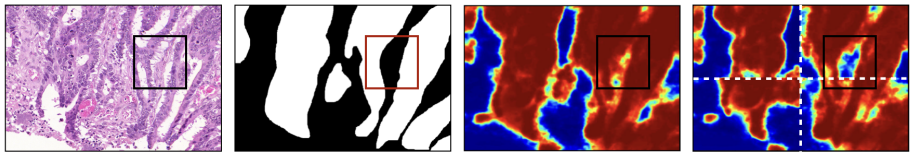

Fig. 1. Comparison of prediction results with and without data augmentation, using FoV disturbance as the augmentation method. (a) The original input image. (b) The ground truth. (c) The prediction without data augmentation. (d) The prediction with data augmentation. The inconsistency between (c) and (d) indicates areas where the model fails to accurately predict.

either positive or negative). Due to the large size and heterogeneity of histology images, identifying specific structures and abnormal tissues may be prone to omissions [3], leading to a situation where only a portion of positive examples is labeled, and the problem can be addressed as a PU learning problem.

Two main approaches exist for handling positive and labeled data [1]. The first identifies reliable positive or negative instances within unlabeled data to expand the labeled set. The second treats all unlabeled data as weighted negative samples. Both methods, however, cannot prevent introducing noise, such as misidentification in the first approach or positive cases within unlabeled data in the second. This noise can lead to performance decline when the model overfits to these inaccurate data points [19].

We observe that predictions of noisy data points are inconsistent under data augmentation, as illustrated in Figure 1. To address this issue, we apply consistency regularization to PU learning, a strategy commonly employed in weakly-supervised learning [2,9,12,22]. Moreover, we consider the importance of Field of View (FoV) in histology image analysis. Small FoV patches provide cellular-level details, while large FoV images offer more global information. Many multi-magnification studies [8,17] combine pixel-aligned feature maps across different FoVs to enrich representation. The underlying principle of these studies is that semantic information should be consistent under different FoVs. For example, when zoomed in (i.e., small FoV) or out (i.e., large FoV), the same object retains the same essential information. However, this aspect is not explicitly addressed in previous work. Therefore, we introduce FoV consistency regularization to enhance the performance of PU learning for histology image semantic segmentation. Specifically, we divide the entire input image (i.e., large FoV) into several patches (i.e., small FoV) and regularize the prediction of the whole input and the reassembled prediction of these patches. To evaluate our method, we conduct experiments on the Glas dataset and consider a setting where the ratio of unlabeled/all positives is controllable. We average metrics over the last 10 epochs to assess our method since a clean validation set is unavailable in PU learning. Full labels are not accessible during training but are available for per-

formance verification. Our results demonstrate that consistency regularization can mitigate noise impact. For example, when the ratio of unlabeled/all positives is 80%, adding consistency loss to nnPU [10] can improve the Dice score by 1.70% to 6.58%, and our FoV consistency regularization outperforms other popular regularization methods based on data augmentation.

Our contributions are summarized as follows: **1)** First, we empirically demonstrate that consistency regularization can reduce the inevitable noise impact in PU learning. **2)** Second, We introduce a new regularization method called FoV consistency to improve the performance of PU learning in histology image segmentation, leveraging the semantic invariance across different FoVs of the same input. **3)** Finally, on the Glas dataset, even with only 20% positive labels, our method achieves a competitive result with a Dice score of 90.69%, approaching the fully supervised result of 93.30%.

2 Related Work

2.1 PU Learning

PU learning aims to train a binary classifier using a dataset containing only positive and unlabeled (PU) samples, without labeled negatives. The primary challenge in PU learning is handling the unlabeled data. Two main approaches address this issue. The first approach, known as the two-step technique [7,14], seeks to expand the labeled set by identifying unlabeled data points likely to be negative or positive and using them to train the model. However, incorrect identification in this approach can cause the model to overfit on noisy data points, leading to performance degradation [19,13]. The second approach [5,10,21] treats all unlabeled data as negative samples and accounts for the presence of noise (i.e., positive data) within them. Nonetheless, this approach is also vulnerable to noisy data and may result in suboptimal performance.

2.2 Consistency Regularization

Consistency regularization enforces consistent predictive results under various disturbances, improving the network's generalization capability. It can be applied as a form of supervision without requiring additional manual annotations. Previous studies have demonstrated the effectiveness of consistency regularization in scenarios with limited labeled data, such as weakly-supervised learning. For example, [12] enforces output consistency across multi-scales to achieve more accurate predictions under noisy labels. [22] utilizes cutout consistency to penalize inconsistent segmentation results. Recently, [9] proposes Puzzle-CAM to identify the most integrated pseudo-labels in weakly-supervised semantic segmentation by encouraging consistency between features from separate local patches and the entire image. In contrast, our FoV consistency regularization technique aims to prevent PU learning models from overfitting to noise that arises when assigning labels to unlabeled data.

3 Method

3.1 Review of PU Segmentation

Our method focuses on Positive-Unlabeled learning for binary segmentation tasks (PU segmentation). In a PU dataset $D = (x_i, s_i, y_i)_{i=1}^{n}$, $x_i \in R^{H\times W\times 3}$ represents the i-th image instance, $s_i \in \{0,1\}^{H\times W}$ indicates the positive pixels selected for labeling in x_i, and $y_i \in \{0,1\}^{H\times W}$ is the true label, which is unavailable. In the binary mask s_i, s_i^j represents the labeling status of a pixel p_j in x_i. Specifically, $s_i^j = 1$ indicates p_j is a positive pixel, while $s_i^j = 0$ signifies p_j is an unlabeled pixel that could belong to either class.

PU segmentation aims to learn a mapping function $f(\cdot;\theta)$ from PU data by minimizing empirical risk $\min_\theta R(f) = E_D[L(f(x;\theta), y)]$, where L is a loss function. Since only part of the positive examples is known, computing empirical risk is challenging. One solution is to relabel the unlabeled data, thereby constructing negative sample sets D_n and positive sample sets D_p. Subsequently, we minimize the approximated empirical risk on the relabeled data, given by:

$$\min_\theta E_{D_n}(L(f(x;\theta), 0)) + E_{D_p}(L(f(x;\theta), 1)). \tag{1}$$

However, this approach overlooks noise in D_n and D_p due to incorrect relabeling, potentially causing overfitting and reduced performance. We note that predictions on noisy data points often lack consistency when subjected to data augmentation. To address this, we add a consistency regularization to the optimization function. Consequently, our empirical risk can be expressed as:

$$\min_\theta E_{D_n}(L(f(x;\theta), 0)) + E_{D_p}(L(f(x;\theta), 1)) + \lambda\Omega_D(x;\theta), \tag{2}$$

where $\Omega_D(x;\theta)$ is the regularization term, and λ is a weighting factor controlling regularization strength. By enforcing consistency on predictions for noisy data points, we enhance the model's robustness and generalization performance.

3.2 Field of View Consistency

The concept of Field of View (FoV) consistency is inspired by multi-magnification research in histology image analysis. Whole Slide Images (WSIs) are typically too large for direct GPU processing and must be divided into smaller patches for training Convolutional Neural Networks (CNNs). Basic patch-based methods utilize only a single FoV of the input image, whereas multi-magnification approaches integrate representations from multiple FoVs. This process mirrors the way pathologists examine WSIs by zooming in and out to study tissues at various magnifications. They observe details of individual cells at smaller FoVs with high magnification and their surroundings at larger FoVs with low magnification. By incorporating representations from different FoVs, multi-magnification studies enhance input features and yield improved results. The underlying principle is that semantic information remains invariant across various FoVs. In this paper,

we explicitly regularize this invariance by designing a consistency regularization term. This loss can be flexibly applied to histology images and other domains that necessitate multi-scale processing.

To implement this consistency regularization term, we divide the input images into smaller patches and feed each patch into the network separately. Then, we merge the results back together to obtain the reassembled prediction at its original size. In our experiment, we divide the image x_i into 4 non-overlapping patches $\{x_i^k\}_{k=1}^4, x_i^k \in R^{\frac{H}{2} \times \frac{W}{2} \times 3}$. The consistency loss can be formulated as:

$$\Omega(x;\theta) = \sum_{i=0}^{n} \|f(x_i;\theta) - merge\{f(x_i^k;\theta)\}_{k=1}^4\|_1, \tag{3}$$

where $f(x_i)$ is the prediction of the full image and $f(x_i^k)$ is the prediction of the j-th patch. The merge operation combines the predictions of the four patches to reconstruct the full image prediction, as shown in (d) of Figure 1. The consistency loss encourages the predictions of both patches and the full image to be consistent with each other. This helps to ensure that the semantic information is invariant across different FoVs.

3.3 A Simple Two-step Method

To assess the effectiveness of the FoV consistency regularization, we design a simple two-step method for PU learning, which we refer to as probability thresholding.

In the first step, our objective is to augment the labeled dataset by choosing reliable negative and positive samples based on the model's output. A probability thresholding strategy is employed to ensure the reliability of unlabeled data. Specifically, we classify unlabeled examples as reliable negatives or positives according to their sigmoid probabilities. Examples with probabilities less than 0.5 are considered reliable negatives, while those greater than 0.8 are deemed reliable positives. This simple and intuitive step selects reliable examples for the second step.

During the second step, we employ the relabeled data and labeled positive data to refine the model using the sigmoid activation function and binary cross-entropy loss function.

4 Experiment

4.1 Experimental Setup

Dataset: We perform PU segmentation experiments on the Gland Segmentation in Colon Histology Images (GlaS) Dataset [15], containing 165 images from 16 H&E stained T3/T4 colorectal adenocarcinoma histological sections. The dataset is split into 85 training images (37 benign and 48 malignant cases) and 80 test images (37 benign and 43 malignant cases), following the protocol of [15]. We manually corrupt the original GlaS dataset labels to create a PU dataset by

randomly assigning connected positive annotation components to the unlabeled class. We also evaluate our method's performance by varying the ratio (η) of unlabeled/all positives to 20%, 50%, and 80%.

Baseline: We use three PU learning methods as baselines, including uPU [5], nnPU [10], and probability thresholding. Both uPU and nnPU treat unlabeled examples as negative and assume the positive class prior probability π_p is known. In the GlaS dataset, $\pi_p = 0.45$.

We compare FoV with three data augmentation techniques for consistency regularization, including:

- **Scale**: Adjust the input image size by downscaling by 0.5 or upscaling by 1.5.
- **Cutout** [4]: Randomly erase a square region of each input image during training, with an area not exceeding 256×256 pixels.
- **Rotation**: Rotate the input image by angles of $\gamma \cdot 90$, with $\gamma \in 1, 2, 3$.

Evaluation Metrics: We evaluate segmentation performance using Intersection over Union (IoU) and Dice coefficient (Dice), both averaged over the last 10 epochs, following the evaluation methods in [16]. Prior research [6] has shown that training with corrupted labels can cause unstable training, leading to considerable performance fluctuations. Thus, we calculate the average of the segmentation metrics across the final 10 epochs for a more reliable performance estimate.

Implementation Details: In our experiments, we preprocess each image by resizing it to 512×512 pixels. We use the state-of-the-art transformer-based segmentation algorithm Segformer [20] as our backbone model and initialize it with pre-trained weights from the CityScapes dataset. The model is trained using the AdamW optimizer, with a learning rate of 1e-4. Following [11], we set λ as a Gaussian ramp-up curve. Each experiment has a batch size of 8 and runs for 100 epochs. We conduct our experiments on a single NVIDIA Quadro A8000 GPU using the PyTorch library.

4.2 Segmentation performance

The results of our experimental comparisons are presented in Table 1, which demonstrate the performance of three PU learning methods as well as the upper bound (i.e., the model trained with clean labels). We conduct a comprehensive comparison between our method and other popular consistency regularization methods based on disturbances. Our results indicate that integrating consistency regularization leads to improved performance relative to the baseline. Moreover, the results suggest that the FoV consistency regularization outperforms other methods, particularly when the ratio η of unlabeled/all positives is high. For instance, when η is 80%, our FoV method achieves a Dice score of 90.69%, which closely approaches the upper bound result of 93.32%. To further illustrate the superiority of the FoV method, we provide representative segmentation results in Figure 2 using the probability thresholding baseline. These results confirm the superior qualitative performance of the FoV consistency regularization.

Baseline method	Consistency method	η-80%		η-50%		η-20%	
		Dice (%)	IoU (%)	Dice (%)	IoU (%)	Dice (%)	IoU (%)
uPU (known π_p)	N/A	43.37 ± 8.24	32.47 ± 7.17	74.12 ± 2.30	61.37 ± 2.96	89.78 ± 1.27	82.49 ± 1.87
	Scale	42.55 ± 9.23	34.47 ± 8.23	85.15 ± 1.75	76.08 ± 2.25	92.00 ± 0.40	85.94 ± 0.61
	Cutout	39.94 ± 6.23	29.74 ± 4.97	75.88 ± 2.76	65.34 ± 3.27	**92.32 ± 0.79**	**86.50 ± 1.19**
	Rotation	**54.46 ± 10.84**	**44.27 ± 10.68**	**85.65 ± 2.69**	**76.69 ± 3.46**	91.92 ± 0.77	85.83 ± 1.20
	FoV	42.06 ± 6.62	33.59 ± 5.81	80.70 ± 4.25	71.48 ± 4.81	92.04 ± 0.82	86.15 ± 1.20
nnPU (known π_p)	N/A	82.94 ± 3.12	72.52 ± 3.82	87.62 ± 1.78	78.75 ± 2.59	85.80 ± 1.00	75.77 ± 1.47
	Scale	86.96 ± 2.43	78.40 ± 2.92	90.32 ± 0.44	82.89 ± 0.74	86.40 ± 1.06	76.53 ± 1.64
	Cutout	84.64 ± 1.70	75.91 ± 1.97	**90.97 ± 1.35**	**84.09 ± 1.93**	**89.27 ± 0.40**	**81.15 ± 0.64**
	Rotation	88.29 ± 0.74	79.87 ± 1.15	88.04 ± 0.51	80.23 ± 0.84	86.52 ± 0.30	76.73 ± 0.47
	FoV	**89.52 ± 0.67**	**81.81 ± 1.07**	90.38 ± 0.52	83.02 ± 0.84	87.84 ± 0.56	78.81 ± 0.90
Probability thresholding (unknown π_p)	N/A	44.24 ± 9.95	33.92 ± 8.59	87.84 ± 2.20	79.66 ± 3.05	91.15 ± 0.54	84.45 ± 0.84
	Scale	89.55 ± 0.60	81.98 ± 0.90	90.61 ± 0.56	83.44 ± 0.85	89.95 ± 0.46	82.30 ± 0.77
	Cutout	87.17 ± 0.99	79.24 ± 1.42	**92.56 ± 0.21**	**86.61 ± 0.36**	**91.87 ± 0.18**	**85.46 ± 0.31**
	Rotation	90.26 ± 0.98	83.15 ± 1.30	91.55 ± 0.32	84.97 ± 0.53	88.94 ± 0.26	81.74 ± 0.43
	FoV	**90.69 ± 0.64**	**83.86 ± 0.96**	89.44 ± 0.36	82.64 ± 0.61	91.29 ± 0.53	84.58 ± 0.72
Upper bound		Dice = 93.32 ± 0.17 %; IoU = 88.03 ± 0.26 %					

Table 1. Segmentation performance of PU learning methods with different consistency regularization techniques for different η. The best result is indicated with **bold text**, while the second-best result is underlined. N/A denotes Not Applicable, indicating that the baseline method is employed without any consistency regularization technique.

Additionally, Figure 3 shows the test Dice score curves for each regularization method across all epochs. Although the numerical differences among the methods are relatively small, the curve graphs reveal that our method offers more stable performance. The consistent performance of our method throughout the training process emphasizes the effectiveness of FoV consistency regularization in handling noise and achieving robust, accurate segmentation results.

However, we emphasize that the performance of PU segmentation is strongly influenced by the choice of consistency regularization technique. While the FoV method outperforms other techniques in some cases, there are scenarios in which Cutout or Rotation yield better results, such as when η is 50%. This suggests the importance of selecting the most appropriate consistency regularization technique, tailored to the specific dataset and task at hand.

5 Conclusion

In this paper, our study investigate the effectiveness of consistency regularization for PU segmentation in histology images. Our experiments demonstrate that the choice of consistency regularization technique strongly influences the performance of PU segmentation, and the most suitable method should be tailored to the specific dataset and task. Additionally, our proposed Field of View (FoV) consistency regularization achieve the best results in scenarios with high ratio of unlabeled positives. The results also demonstrate the stability of our method's performance, indicating its ability to handle noisy data and achieve robust and accurate segmentation results. Our findings provide insights into the selection of appropriate consistency regularization techniques for PU learning in histology

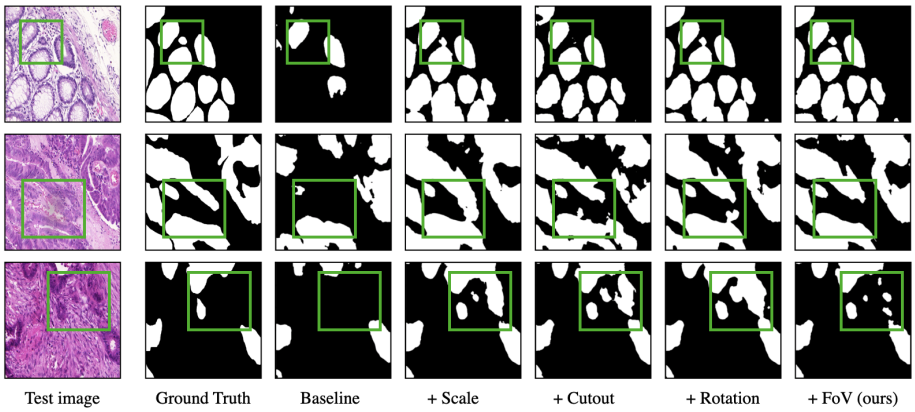

Fig. 2. Comparison of segmentation results when the ratio η of unlabeled positives is 80%.

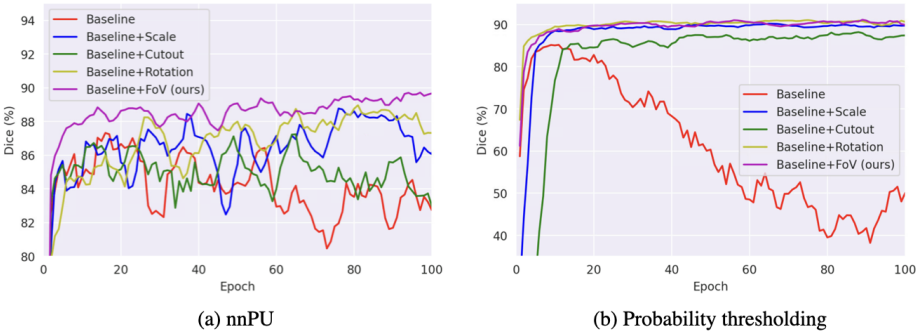

Fig. 3. Test dice score curves for each method across all epochs when the ratio η of unlabeled positives is 80%. The curves demonstrate the stability of each method's performance over time.

image segmentation, which could have significant implications for improving the performance of segmentation models in real-world clinical applications.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No.62271475).

References

1. Bekker, J., Davis, J.: Learning from positive and unlabeled data: A survey. Machine Learning 109, 719–760 (2020)

2. Chen, X., Chen, W., Chen, T., Yuan, Y., Gong, C., Chen, K., Wang, Z.: Self-pu: Self boosted and calibrated positive-unlabeled training. In: International Conference on Machine Learning. pp. 1510–1519. PMLR (2020)
3. 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)
4. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
5. Du Plessis, M.C., Niu, G., Sugiyama, M.: Analysis of learning from positive and unlabeled data. Advances in neural information processing systems 27 (2014)
6. Guo, X., Yuan, Y.: Joint class-affinity loss correction for robust medical image segmentation with noisy labels. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV. pp. 588–598. Springer (2022)
7. He, F., Liu, T., Webb, G.I., Tao, D.: Instance-dependent pu learning by bayesian optimal relabeling. arXiv preprint arXiv:1808.02180 (2018)
8. Ho, D.J., Yarlagadda, D.V., D'Alfonso, T.M., Hanna, M.G., Grabenstetter, A., Ntiamoah, P., Brogi, E., Tan, L.K., Fuchs, T.J.: Deep multi-magnification networks for multi-class breast cancer image segmentation. Computerized Medical Imaging and Graphics 88, 101866 (2021)
9. Jo, S., Yu, I.J.: Puzzle-cam: Improved localization via matching partial and full features. In: 2021 IEEE International Conference on Image Processing (ICIP). pp. 639–643. IEEE (2021)
10. Kiryo, R., Niu, G., Du Plessis, M.C., Sugiyama, M.: Positive-unlabeled learning with non-negative risk estimator. Advances in neural information processing systems 30 (2017)
11. Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.A.: Transformation-consistent self-ensembling model for semisupervised medical image segmentation. IEEE Transactions on Neural Networks and Learning Systems 32(2), 523–534 (2020)
12. Liu, S., Liu, K., Zhu, W., Shen, Y., Fernandez-Granda, C.: Adaptive early-learning correction for segmentation from noisy annotations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2606–2616 (2022)
13. Liu, S., Niles-Weed, J., Razavian, N., Fernandez-Granda, C.: Early-learning regularization prevents memorization of noisy labels. Advances in neural information processing systems 33, 20331–20342 (2020)
14. Nguyen, N.V., Rigaud, C., Revel, A., Burie, J.C.: A learning approach with incomplete pixel-level labels for deep neural networks. Neural Networks 130, 111–125 (2020)
15. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al.: Gland segmentation in colon histology images: The glas challenge contest. Medical image analysis 35, 489–502 (2017)
16. Tan, C., Xia, J., Wu, L., Li, S.Z.: Co-learning: Learning from noisy labels with self-supervision. In: Proceedings of the 29th ACM International Conference on Multimedia. pp. 1405–1413 (2021)

17. Van Rijthoven, M., Balkenhol, M., Siliņa, K., Van Der Laak, J., Ciompi, F.: Hooknet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. Medical image analysis 68, 101890 (2021)
18. Wang, S., Yang, D.M., Rong, R., Zhan, X., Xiao, G.: Pathology image analysis using segmentation deep learning algorithms. The American journal of pathology 189(9), 1686–1698 (2019)
19. Wang, Z., Jiang, J., Long, G.: Positive unlabeled learning by semi-supervised learning. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 2976–2980. IEEE (2022)
20. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems 34, 12077–12090 (2021)
21. Xu, Y., Gong, M., Chen, J., Chen, Z., Batmanghelich, K.: 3d-boxsup: Positive-unlabeled learning of brain tumor segmentation networks from 3d bounding boxes. Frontiers in Neuroscience 14, 350 (2020)
22. 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–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VIII. pp. 162–172. Springer (2022)