UDCP: Uncertainty Driven Context-prompted nnUNet for Tertiary Lymphoid Structure Segmentation in H&E Images

1st Meng Wang
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
wangmeng_dx@lntu.edu.cn

2nd Yuxin Liang
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
2833391387@qq.com

3rd Yarong Feng
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
18093490172@163.com

4th Yongwei Tang
School of Electronic and Information Engineering
Liaoning Technical University
Xingcheng City, 125105, P. R. China
2458672708@qq.com

5th Tianyu Shi*
School of information Science and Engineering
Shenyang Ligong University
Shenyang city, 110159, P.R.China
ty.shi@sylu.edu.cn

6th Chao Lv*
dept. name of organization (of Affiliation)
China Medical University
Shenyang city, 110159, P.R.China
clu@cmu.edu.cn

***Abstract*—In real pathological scenarios, tertiary lymphoid structures (TLS) exhibit varying sizes, diverse morphologies, complex cellular compositions, and unclear boundaries. Traditional deep convolutional neural networks, such as nnU-Net, struggle to fully utilize contextual information in TLS segmentation tasks, often leading to missed and false detections. To address this issue, this paper proposes an uncertainty-driven contextual cueing nnU-Net based on a clinically collected TLS pathological image dataset. The method first trains an nnU-Net baseline model using pathological images. Then, it concatenates the foreground probability map output by the model with the original image and inputs it back into nnU-Net, explicitly incorporating high-level contextual information and constructing a two-stage auto-context framework. On this basis, to address potential uncertainties in the probability map, this paper further designs two parallel contextual information purification strategies: a region-level probability gating method based on an adaptive threshold of Otsu; and a pixel-level uncertainty modeling method constructed based on predicted probabilities. These two contextual cueing masks are fused to generate an uncertainty-driven cue. Experimental results show that on a TLS independent test set strictly divided by patients, the proposed method can effectively improve the segmentation performance of the baseline model without requiring additional annotations or complex structural modifications.**

Keywords—Pathological image segmentation; tertiary lymphoid structure; Otsu threshold; automatic context; uncertainty learning

I. INTRODUCTION

Tumor-associated tertiary lymphoid structures (TLS) are lymph node-like structures composed of B cells, T cells, dendritic cells, and other cells, which are closely related to the prognosis and immunotherapy benefit of patients with various solid tumors [1-3]. Accurately depicting the presence and spatial distribution of TLS on conventional hematoxylin and eosin (H&E) pathological sections is of great significance for the quantitative analysis of the tumor microenvironment. In recent years, deep learning methods have begun to be used for the automatic detection and quantification of TLS [4-6], but existing research has mostly focused on the presence and density estimation at the whole-section scale.For TLS regions with small volume, blurred boundaries, and complex backgrounds within the sections, research on fine segmentation is still limited.

In the field of medical image segmentation, U-Net and its variants have become mainstream baseline models [7-12]. The classic U-Net integrates multi-scale features through an encoder-decoder structure and skip connections [7]; Attention U-Net introduces an attention mechanism to suppress background interference [8]; the MONAI framework provides a unified implementation and training tool for medical segmentation tasks [9]. In addition, recent U-Net-based methods have further improved segmentation performance by enhancing decoder feature interaction and multiscale feature fusion. For example, HD-RDS-UNet leverages spatial-temporal correlation between decoder feature maps to improve lymphoma segmentation [10], while an improved U-Net with multiscale fusion has been reported for retinal vessel segmentation [11]. Another representative U-Net-style segmentation variant is also widely adopted in medical imaging [12]. Furthermore, nnU-Net achieves competitive performance on multiple public benchmarks by automatically configuring network structures and training strategies [13].

However, single-stage, single-channel segmentation frameworks still face challenges in TLS segmentation tasks. On the one hand, TLS accounts for a very small proportion of the slice, resulting in a highly imbalanced foreground-background ratio[14-15]. On the other hand, the boundary between TLS and surrounding inflammatory areas is blurred, making it prone to false positive predictions. To address this, the Auto-Context cascaded framework introduces the prediction from the previous stage as contextual information to gradually refine the segmentation results[16]. Simultaneously, uncertainty learning is employed to characterize the model's prediction confidence in space, but it relies heavily on multiple inferences, leading to significant computational overhead [17-18]. Recently, prompt-based and foundation-model-driven medical segmentation methods have further highlighted the importance of explicitly injecting auxiliary guidance into the segmentation pipeline. For example, TP-DRSeg introduced explicit text prompts to customize SAM for diabetic retinopathy lesion segmentation, demonstrating that external priors can improve the discrimination of subtle lesion structures[19]. DAPSAM further showed that domain-adaptive prompt generation can

*Corresponding author.

enhance the generalization ability of SAM-based medical segmentation models under distribution shifts [20]. More recently, Med-SA and DB-SAM adapted SAM to medical image segmentation through lightweight adapters and dual-branch architectural designs, respectively, confirming the effectiveness of prompt-conditioned or prompt-related guidance in improving medical segmentation performance[21-22]. Nevertheless, these methods are mainly developed for foundation-model adaptation or general medical segmentation, and their applicability to small, sparse, and boundary-ambiguous TLS segmentation in routine H&E pathology images remains underexplored. This motivates us to develop a lightweight uncertainty-driven context prompting framework built on nnU-Net rather than a large foundation model.To tackle these issues, this paper proposes a refined TLS segmentation method based on a self-built TLS pathological image dataset within the nnU-Net framework, combining two-stage contextual cues with probabilistic uncertainty modeling. The main contributions include:

1) Construct a two-stage contextual cueing process, utilizing the TLS probability map generated in the first stage as the foreground prior for the second stage, and cascading it with the original image for input into the network;

2) Propose two complementary strategies for probability map purification, including a high-confidence foreground gating based on the Otsu [23] threshold and a pixel-level uncertainty modeling based on probability margin, to highlight boundaries and difficult regions;

3) During the inference stage, Multi-level context aggregation balances false positive suppression and boundary refinement, significantly improving the Dice and IoU performance of TLS segmentation on an independent test set.

II. METHOD

A. Overall Framework and Two-Stage Context Design

To introduce efficient contextual information while maintaining the original advantages of nnU-Net, this paper adopts an easily implementable two-stage automatic context cueing framework (auto context nnUNet, AC-nnUNet). The overall process of AC-nnUNet is illustrated in Figure [Fig.1]. It expands only at the input channel level, without making any modifications to the network structure, loss function, or training strategy of nnU-Net. This makes it simple and efficient.

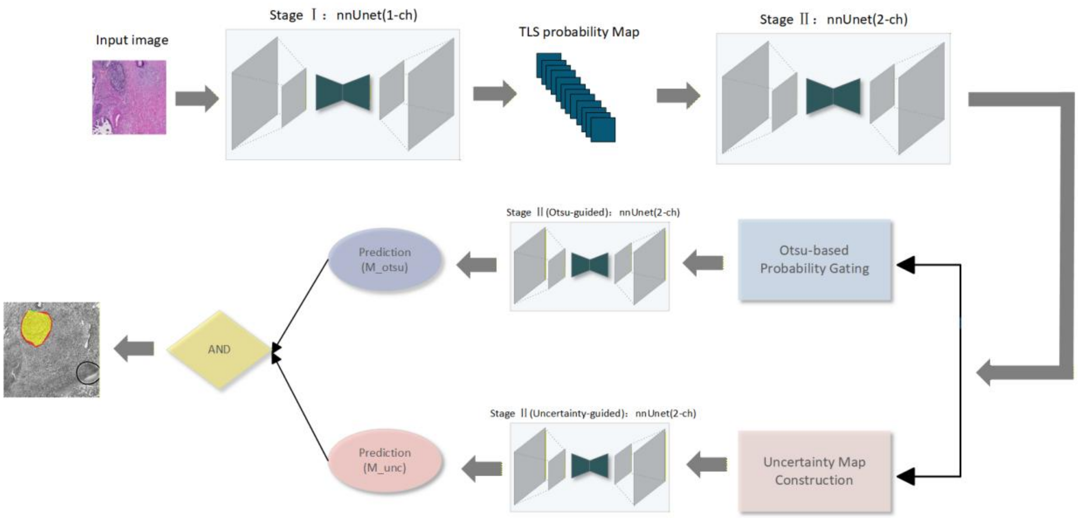

Fig. 1. Overall pipeline of UDCP. Stage-I nnU-Net outputs a TLS probability map P, which is transformed into an Otsu-gated foreground prompt P' and an uncertainty prompt U. Two stage-II nnU-Net branches predict M_{otsu} and M_{unc}, and the final mask is obtained by AND fusion.

In the first stage of AC-nnUNet, the original TLS pathological images are used to train a standard nnU-Net model, To mitigate stain-induced distribution shifts, stain normalization is commonly adopted in computational pathology[24].obtaining a pixel-level TLS foreground prediction probability map. The prediction probability at pixel position x is denoted as

$$P(x) \in [0,1] \tag{1}$$

Here, $P(x)$ represents the confidence level of the first-stage model in determining that this pixel belongs to the TLS foreground. This probability map can be regarded as the model's preliminary high-level semantic description of the spatial distribution of TLS.

In the second stage, context cues are constructed based on the probability map output from the first stage and concatenated with the original image in the channel dimension as input, enabling the network to perceive spatial prior information early in feature extraction. Specifically, from the perspectives of high-confidence foreground regions and potential boundary hard cases, we construct foreground context based on Otsu gating and uncertainty context based on probability margin, respectively. In the inference stage, the two prediction results are fused at the mask level to obtain the final segmentation result.

This framework follows the classic auto-context idea, guiding model learning solely through input-level prompts, without requiring additional annotations or complex cascaded structures, making it easy to integrate into existing nnU-Net workflows[16].

B. Otsu-Based Foreground Context Gating

Although the probability map generated in the first stage provides valuable spatial priors, it may still contain noise responses caused by staining differences, tissue structural similarities, or model misjudgments. To address this, this paper first employs the Otsu adaptive threshold method to perform region-level gating on the TLS foreground probability map, in order to construct a more stable foreground context.

The Otsu method[23] automatically estimates the optimal threshold t by analyzing the global histogram of the probability map, without the need for manual parameter setting, thus achieving maximum distinction between foreground and background in the sense of inter-class variance. Compared to the fixed threshold strategy, the adaptive threshold can dynamically adjust the threshold position according to changes in the probability distribution shape in different slices. After obtaining the optimal threshold t, this paper performs a gating operation on the probability map in the first stage, retaining only the high-confidence foreground regions:

$$P'(x) = \begin{cases} P(x),\ P(x) \geq \tau \\ 0,\ otherwise \end{cases} \quad (2)$$

Here, $P'(x)$ represents the probability map after Otsu gating. This operation explicitly suppresses low-confidence responses while maintaining probability continuity, weakening scattered false high responses in the background region, and making contextual cues more compact and stable in space.

Subsequently, the gated probability map $P'(x)$ is incorporated into a new dataset as the context piror and inputted into nnU-Net along with the original pathological images for the second stage of training. This results in a context branch primarily composed of high-confidence foreground, which provides stable and conservative hints for the TLS region.

C. Uncertainty-Aware Context Modeling and Mask-Level Fusion

Merely emphasizing high-confidence foreground regions is still insufficient to fully depict the complex morphology of TLS. Especially in regions with blurred boundaries or similar organizational structures, the model often exhibits high uncertainty. To address this, this paper further constructs pixel-level uncertainty contexts from the perspective of discriminative difficulty modeling, and performs mask-level fusion with the Otsu branch during the inference stage.

1) Margin-based Uncertainty Map

Unlike confidence scores that only reflect how strongly the model favors a specific category, uncertainty is used to characterize the ambiguity of foreground/background discrimination at different spatial locations. Existing uncertainty estimation methods are commonly based on approximate Bayesian inference or ensemble learning, such as Monte Carlo dropout and deep ensembles [17],[18],[25],[26]. However, these methods usually require repeated stochastic inference or multiple models, which may introduce additional computational overhead.

In our framework, the uncertainty map is directly derived from the probability output of the stage-I nnU-Net. When the predicted probability approaches 0 or 1, the model is relatively confident at that location. In contrast, when the probability approaches 0.5, foreground and background become equally plausible, and the corresponding uncertainty should be maximized.

Based on this observation, we construct a lightweight pixel-wise uncertainty measure based on the prediction margin:

$$U(x)=1-|2P(x)-1| \quad (3)$$

A larger value indicates a higher degree of uncertainty at pixel x. This function reaches its maximum at the decision boundary, where the predicted probability equals 0.5, and gradually decreases toward 0 as the probability approaches either 0 or 1. Therefore, it provides a symmetric and computationally efficient way to quantify how far the prediction deviates from the decision boundary.

In implementation, we write the uncertainty map into the new dataset as the second-channel context prior, which is concatenated with the original pathological image and then fed into nnU-Net, forming an uncertainty-guided contextual branch. By explicitly highlighting ambiguous boundary regions and hard examples, this branch encourages the network to focus more on difficult TLS patterns during training, thereby improving its ability to delineate complex TLS structures.

2) *Multi-level context aggregation*

In the inference stage, this paper obtains binary prediction masks, denoted as M_{otsu} and M_{unc}, from the Otsu gated context branch and the uncertainty-based context branch, respectively. To integrate the two complementary types of contextual information and suppress isolated false positives, a mask-level fusion strategy using logical AND is adopted to construct the final prediction result:

$$M_{final}=M_{otsu}\cap M_{unc} \quad (4)$$

This strategy only includes corresponding pixels in the final segmentation region when both branches are simultaneously determined to be TLS foreground. Since the Otsu branch tends to favor highly confident compact foregrounds, while the uncertainty branch emphasizes boundaries and complex local structures, the two branches exhibit significant complementarity in nature. Through Multi-level context aggregation , this paper maintains the coherence of the foreground region while effectively suppressing scattered false positives in the background, thereby enhancing the consistency and robustness of the final segmentation results across the entire slide scale.

III. Experimental Setup

A. Dataset Description and Preprocessing

The experiments in this paper are based on a self-built TLS pathological image dataset. All images are derived from real clinical pathological sections and annotated at the pixel level by experienced pathologists, with the annotation target being tertiary lymphoid structures (TLS) regions.

To avoid potential data leakage issues between different slides of the same patient, this paper adopts a strict patient-level data partitioning strategy. The entire dataset comprises 52 patients, with slides from 46 patients used for model training and slides from 6 patients serving as an independent test set. Correspondingly, the training set contains approximately 4,800 slides, and the test set contains approximately 900 slides. All experimental results are

This work was supported by the National Natural Science Foundation of China (Grant No. 82572327), Science and Technology Plan Joint Program of Liaoning Province of China (Natural Science Foundation - General Program, 2024-MSLH-596)、Science and Technology Program of Liaoning Province of China (Provincial Doctoral Research Startup Fund Program, 2024- BS-258) and National Natural Science Foundation of China (Grant No. 62506243).

reported on this independent test set.

In the data preprocessing stage, all slices are processed according to the default workflow of nnU-Net, including size normalization and intensity standardization. This paper adopts a two-dimensional slice-level segmentation setting, where each slice is treated as an independent sample input to the model for training and inference.

B. Evaluation Metrics

To quantitatively evaluate the performance of the model in the task of TLS pathological image segmentation, this paper adopts the $Dice$ similarity coefficient as the primary evaluation metric[27-28]. The $Dice$ coefficient can effectively measure the degree of overlap between the predicted segmentation results and the manually annotated regions. Furthermore, it exhibits good stability in segmentation tasks where the foreground region accounts for a small proportion and the class distribution is highly imbalanced. Therefore, it is widely used in pathological image segmentation research.

In addition, this paper also reports Intersection over Union (IoU) as a supplementary evaluation metric. IoU evaluates segmentation results from the perspective of region intersection-over-union ratio, which is highly correlated with the $Dice$ metric but has certain differences in numerical scale and interpretation method. The combined use of $Dice$ and IoU can reflect the segmentation quality of the model for TLS regions from different perspectives, thus providing a more comprehensive and intuitive performance evaluation.

For ease of notation in the following text, let G denote the set of manually labeled TLS foreground pixels, and P denote the set of foreground pixels predicted by the model. The $Dice$ coefficient and IoU can be formally defined as follows:

$$Dice=\frac{2|P\cap G|}{|P|+|G|} \tag{5}$$

$$IoU=\frac{|P\cap G|}{|P\cup G|} \tag{6}$$

Equivalently, if TP, FP, and FN represent the number of true positive, false positive, and false negative pixels, respectively, then:

$$Dice=\frac{2TP}{2TP+FP+FN} \tag{7}$$

$$IoU=\frac{TP}{TP+FP+FN} \tag{8}$$

Among them, $Dice$ focuses more on the overall degree of overlap, while IoU characterizes the consistency between prediction and annotation from the perspective of intersection over union. The joint use of the two metrics helps evaluate the quality of TLS segmentation from different scales.

C. Foreground-Aware Context Construction and Refinement

All experiments were conducted under the standard 2D nnU-Net setting, and no architectural modifications were introduced to the backbone. The same patient-level split was adopted for all methods, with approximately 3800 patches for training, about 990 patches for internal validation, and 978 patches from 6 independent patients for final testing.

In the proposed UDCP framework, the probability maps generated by the stage-I model were processed offline to construct the second-channel context prompts for stage-II training. Specifically, the Otsu branch used the gated probability map P'(x), whereas the uncertainty branch used the margin-based uncertainty map U(x)=1-|2P(x)-1|. These context maps were written as fixed second-channel inputs and did not participate in backpropagation.

All models were optimized using a compound Dice and Cross-Entropy loss. During inference, the probability outputs were binarized using a threshold of 0.5, and the final segmentation mask was obtained via a mask-level logical AND fusion of the Otsu and uncertainty branches.

IV. Experimental Results

A. Comparison with Baseline Models

In this section, the proposed method is compared with multiple commonly used segmentation models through experimental contrast to verify its overall performance in the TLS pathological image segmentation task. These comparative methods include U-Net and Attention U-Net. For fairness, all comparative models use single-channel pathological images as input and adopt the same patient-level data partitioning strategy as the method proposed in this paper.

The quantitative experimental results are presented in Table I. It can be observed that U-Net, Attention U-Net, and MONAI U-Net achieved Dice/IoU scores of 0.749/0.612, 0.752/0.618, and 0.740/0.604, respectively, on the independent test set; the Dice/IoU score for the single-channel nnU-Net baseline was 0.745/0.610. In comparison, our proposed method improved the Dice score to 0.769 and the IoU score to 0.639 on the same test set, indicating that introducing context cueing and context purification mechanisms can stably enhance TLS segmentation performance without altering the network structure.

TABLE I. Quantitative comparison with baseline models on the TLS test set.

Method	Dice	IoU
U-Net	0.749	0.612
Attention U-Net	0.752	0.618
MONAI U-Net	0.740	0.604
nnU-Net(1-ch)	0.745	0.610
Ours	**0.769**	**0.639**

In addition to the aforementioned quantitative results, this paper further presents the segmentation visualization results of various methods on three representative TLS cases, as shown in Figure 2. It can be observed that for TLS regions with small scales, blurred boundaries, and intermingled with surrounding tissues, U-Net, Attention U-Net, and MONAI U-Net all exhibit issues such as missed detections, incomplete boundaries, or scattered false positives to varying degrees. In contrast, the method proposed in this paper maintains fewer false positives while achieving more coherent TLS segmentation results that closely conform to the labeled contours, better capturing the fine TLS structures and their true morphology.

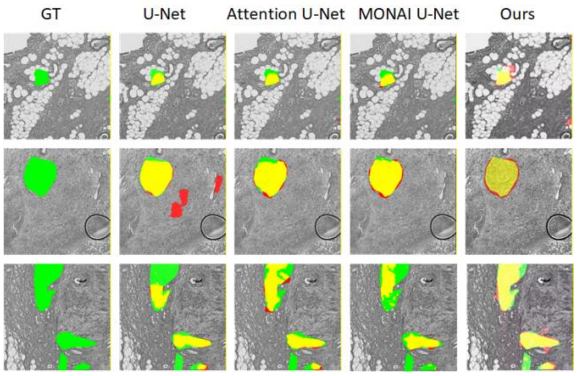

Fig. 2. Qualitative comparison of TLS segmentation results on representative cases.From left to right: ground truth (GT), U-Net, Attention U-Net, MONAI U-Net, and the proposed method (Ours). Ground-truth TLS regions are shown in green, predicted TLS regions are shown in red, and overlapping areas between prediction and ground truth are highlighted in yellow. Compared with baseline models, the proposed method produces more accurate TLS delineation with fewer false positives and more complete boundaries, particularly for small or irregular TLS structures.

It should be noted that the comparative experiments in this study are designed to validate the effectiveness of the proposed method against representative 2D segmentation baselines and within the nnU-Net framework, rather than to provide an exhaustive benchmark over all recently proposed segmentation architectures. In particular, our goal is to verify whether uncertainty-driven context prompting can serve as an effective plug-in strategy for a strong nnU-Net baseline in TLS segmentation.

B. Ablation Study

To analyze the contribution of each component module in the proposed method to performance improvement, systematic ablation experiments were conducted within the nnU-Net framework. The results are presented in Table II. All experiments employed the same data partitioning and training configuration, with adjustments only made to the input channel format and context modeling strategy.

Firstly, the Dice coefficient of the single-channel nnU-Net baseline model on the independent test set is 0.745. By introducing a two-stage context cueing mechanism, where the TLS foreground probability map predicted in the first stage is used as the input for the context piror , the model performance is improved to 0.750, indicating that probability priors can provide effective spatial cues for TLS localization.

Based on this, Otsu gating and uncertainty context modeling based on probability were introduced respectively. Experimental results showed that Otsu gating could improve the Dice coefficient to 0.764, while uncertainty context modeling achieved 0.762, indicating that both of them contributed to enhancing the model's discriminative ability towards the TLS region.

Furthermore, in the post-processing stage, a Multi-level context aggregation strategy was employed to integrate the two prediction results, resulting in the model achieving the highest performance on the independent test set, with a Dice coefficient of 0.769. The results indicate that different context purification strategies are complementary in nature, and their joint use can enhance overall segmentation stability and accuracy while suppressing false positives.

TABLE II. ABLATION STUDY OF THE PROPOSED CONTEXT-AWARE FRAMEWORK.

Design Variant	Context Prior	Dice
nnU-Net baseline	—	0.745
+Context(raw prob)	P	0.750
+Otsu gating	P′	0.764
+Uncertainty	U	0.762
+AND fusion	P′ +U	**0.769**

V. DISCUSSION

TLS has become an important immunological marker and is closely related to the prognosis of patients and their response to immunotherapy[1-3],[29]. Furthermore, the pathological images of TLS exhibit characteristics such as small target scale, blurred boundaries, and complex background noise,making it difficult for traditional single-stage segmentation frameworks to strike a balance between recall and false positives. In this paper, we propose a foreground-aware two-stage contextual cueing method based on the nnU-Net baseline framework. Without modifying the network structure and training strategy, we generate a TLS probability map in the first stage and concatenate it as an additional channel with the original pathological image in the second stage. This enables the network to explicitly perceive the spatial prior knowledge of TLS, thereby enhancing its modeling ability for small target regions.

Based on this, this paper designs two complementary context purification strategies at the probability map level: Otsu adaptive gating focuses on extracting stable context from high-confidence foreground regions, while uncertainty measurement explicitly marks prediction boundaries and easily confused regions, guiding the network to perform more refined learning at positions with high discrimination difficulty. The two strategies constrain context information from two dimensions: region-level confidence and pixel-level discrimination difficulty, avoiding noise interference that may be introduced by directly using the original probability map.

The experimental results demonstrate that, compared wit h various two-dimensional segmentation models and the sing le-channel nnU-Net baseline, the proposed method improves Dice from 0.745 to 0.769 on the independent test set of the se lf-built TLS dataset, and achieves consistent gains in metrics such as IoU. Ablation experiments further verify the effectiv eness of each module, with the Multi-level context aggregati on strategy performing optimally in terms of accuracy and r obustness. Since this method is only extended in a "plug-and-play" manner at the input layer and probability map post-pro cessing stage, it has advantages such as simple implementati on, controllable computational overhead, and good versatility. In the future, it can be extended to other pathological and me dical imagesegmentation scenarios.Recent studies have explo red deep learning for automated TLS detection/characterizati on on histopathology, yet robustness under stain variability a nd hard boundaries remains challenging[4-6],[30].

A limitation of the current study is that recent transforme r-based or prompt-based segmentation models were not inclu ded in the comparative experiments. This is mainly because t he present work focuses on evaluating the plug-in effectivene ss of the proposed uncertainty-driven context prompting strat egy within the nnU-Net framework, rather than re-benchmar king all recent segmentation backbones. Extending the comp arison to newer transformer-based and prompt-based method s will be an important direction for future work.

Code and Data Availability

The source code of the proposed method will be released upon acceptance. Due to patient privacy and institutional rest rictions, the complete TLS pathology dataset cannot be fully made public. However, the data split protocol, preprocessing pipeline, and implementation details will be provided to supp ort reproducibility. A de-identified subset will be released if permitted by the corresponding ethics and data governance p olicies.

References

[1] T. N. Schumacher and D. S. Thommen, "Tertiary lymphoid structures in cancer," Science, vol. 375, no. 6576, Art. no. eabf9419, 2022.

[2] B.A, Reddy S M, Gao J, et al. B cells and tertiary lymphoid structures promote immunotherapy response[J]. Nature, 2020, 577(7791): 549-555.

[3] F, De Reyniés A, Keung E Z, et al. B cells are associated with survival and immunotherapy response in sarcoma[J]. Nature, 2020, 577(7791): 556-560.

[4] Z. Chen et al., "Deep learning on tertiary lymphoid structures in hematoxylin – eosin predicts cancer prognosis and immunotherapy response," npj Precision Oncology, vol. 8, no. 1, p. 73, 2024.

[5] Z. Li et al., "Development and validation of a machine learning model for detection and classification of tertiary lymphoid structures in gastrointestinal cancers," JAMA Network Open, vol. 6, no. 3, p. e2252553, 2023.

[6] M. van Rijthoven et al., "Multi-resolution deep learning robustly characterises tertiary lymphoid structures across cancers, " Communications Medicine, vol. 4, no. 1, p. 5, 2024.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234–241.

[8] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," arXiv:1804.03999, 2018.

[9] M. J. Cardoso et al., "MONAI: An open-source framework for deep learning in healthcare," arXiv:2211.02701, 2022.

[10] M. Wang, H. Jiang, T. Shi, and Y. -d. Yao, "HD-RDS-UNet: Leveraging Spatial-Temporal Correlation Between the Decoder Feature Maps for Lymphoma Segmentation," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 3, pp. 1116 – 1127, 2022.

[11] J. Ni, J. Shi, Q. Zhan, Z. Zhang, and Y. Gu, "An improved U-Net model with multiscale fusion for retinal vessel segmentation," Intelligent Robotics, vol. 5, no. 3, pp. 679–694, 2025.

[12] Zhou et al., "UNet++: A Nested U-Net Architecture for Medical Image Segmentation", IEEE TMI 2018

[13] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203–211, 2021.

[14] Li Y, Wu X, Li C, et al. A hierarchical conditional random field-based attention mechanism approach for gastric histopathology image classification[J]. Applied Intelligence, 2022, 52(9): 9717-9738.

[15] Kumar N, Verma R, Sharma S, et al. A dataset and a technique for generalized nuclear segmentation for computational pathology[J]. IEEE transactions on medical imaging, 2017, 36(7): 1550-1560.

[16] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks," in Proc. CVPR, 2010, pp. 1–8.

[17] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in Proc. NIPS, 2017, pp. 5574–5584.

[18] G. Wang et al., "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," Medical Image Analysis, vol. 68, p. 101855, 2021.

[19] W. Li, X. Xiong, P. Xia, L. Ju, and A. Wang, "TP-DRSeg: Improving diabetic retinopathy lesion segmentation with explicit text-prompts assisted SAM," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), 2024, doi: 10.1007/978-3-031-72111-3_70.

[20] Z. Wei, H. Hu, Y. Wang, and D. Shen, "Prompting Segment Anything Model with domain-adaptive prototype for generalizable medical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), 2024, doi: 10.1007/978-3-031-72111-3_50.

[21] J. Wu, Z. Wang, M. Hong, W. Ji, H. Fu, Y. Xu, M. Xu, and Y. Jin, "Medical SAM adapter: Adapting segment anything model for medical image segmentation," Med. Image Anal., vol. 102, p. 103547, 2025, doi: 10.1016/j.media.2025.103547.

[22] C. Qin, J. Cao, H. Fu, F. S. Khan, and R. M. Anwer, "DB-SAM: Delving into high quality universal medical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), 2024, doi: 10.1007/978-3-031-72390-2_47.

[23] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[24] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in Proc. IEEE Int. Symp. Biomed. Imaging (ISBI), 2009, pp. 1107–1110.

[25] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Proc. Int. Conf. Mach. Learn. (ICML), 2016, pp. 1050–1059.

[26] A. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 6402–6413.

[27] L. R. Dice, "Measures of the amount of ecologic association between species," Ecology, vol. 26, no. 3, pp. 297–302, 1945.

[28] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et du Jura," Bull. Soc. Vaudoise Sci. Nat., vol. 37, pp. 547–579, 1901.

[29] C. Sautès-Fridman et al., "Tertiary lymphoid structures in the era of cancer immunotherapy," Nat. Rev. Cancer, vol. 19, no. 6, pp. 307–325, 2019.

[30] P. Barmpoutis et al., "Automated detection and quantification of tertiary lymphoid structures in colorectal cancer using deep learning," PLOS ONE, 2021.