

A Weakly Supervised Segmentation Network Embedding Cross-scale Attention Guidance and Noise-sensitive Constraint for Detecting Tertiary Lymphoid Structures of Pancreatic Tumors

Bingxue Wang, Liwen Zou, Jun Chen, Yingying Cao, Zhenghua Cai, Yudong Qiu, Liang Mao, Zhongqiu Wang, Jingya Chen, Luying Gui and Xiaoping Yang

Abstract—The presence of tertiary lymphoid structures (TLSs) on pancreatic pathological images is an important prognostic indicator of pancreatic tumors. Therefore, TLSs detection on pancreatic pathological images plays a crucial role in diagnosis and treatment for patients with pancreatic tumors. However, fully supervised detection algorithms based on deep learning usually require a large number of manual annotations, which is time-consuming and labor-intensive. In this paper, we aim to detect the TLSs in a manner of few-shot learning by proposing a weakly supervised segmentation network. We firstly obtain the lymphocyte density maps by combining a pretrained model for nuclei segmentation and a domain adversarial network for lymphocyte nuclei recognition. Then, we establish a cross-scale attention guidance mechanism by jointly learning the coarse-scale features from the original histopathology images and fine-scale features from our designed lymphocyte density attention. A noise-sensitive constraint is introduced by an embedding signed distance function loss in the training procedure to reduce tiny prediction errors. Experimental results on two collected datasets demonstrate that our proposed method significantly outperforms the state-of-the-art segmentation-based algorithms in terms of TLSs detection accuracy. Additionally, we apply our method to study the congruent relationship between the density of TLSs and peripancreatic vascular invasion and obtain some clinically statistical results.

Index Terms—Tertiary lymphoid structures, Pancreatic tumor, Weakly supervised segmentation, Cross-scale attention, Noise-sensitive constraint.

I. INTRODUCTION

HISTOPATHOLOGY images are often the gold standard for disease detection, diagnosis, and prognostic analysis. However, analysis for such images is a very challenging

task because of their large sizes and numerous elements. In recent years, the development of deep neural networks has led to many breakthroughs in automatic histopathology image classification and segmentation [1], [2]. These methods depend highly on extensive training and accurate pixel-level labels, which are particularly labor-intensive and time-consuming due to the huge sizes of pathological images. Actually, only a trained pathologist can distinguish precisely among the elements. Therefore, acquiring accurate annotations on pathological images is more complicated than annotating other medical images. The development of weakly supervised and unsupervised learning methods has become an inevitable trend in pathological image processing [3], [4].

In pathology research, the immune micro-environment is a topic of great concern. The main research target of the immune micro-environment is lymphocytes. Their particular aggregation clusters, known as tertiary lymphoid structures (TLSs), have received significant attention in recent years. Actually TLSs are discrete structured tissues of infiltrating immune cells, or in other words, more organized aggregation structures of lymphocytes. In some studies, TLSs are considered as any lymphoid aggregate similar to secondary lymphoid organs in non-lymphoid structures [5]. Several researches [6], [7] concentrated on the presence, composition, and location of TLSs in health and disease tissues and organs. The high density of TLSs found in many cancers is associated with prolonged survival of patients, such as colorectal cancers [8], [9] and breast cancers [10].

Previous studies have generally considered pancreatic tumors as cold tumors and lack of immune response [11], [12]. In contrast, some studies in recent years have found that the immune micro-environment of pancreatic tumors is highly heterogeneous, which largely contributes to its lack of mechanical clarity to date. One of the critical elements of its heterogeneity is a large number of immune cell subsets [13]. This means accurately segmenting lymphocytes and classifying clusters are crucial for studying pancreatic tumors. In the pathologist's opinion, lymphocytes in a TLS have a particular specific aggregation pattern, such as having a distinctive high density. Although a TLS does not have an envelope but is still clearly delimited from the surrounding tissues. However, it is challenging to identify all TLSs on a pathological image because of the considerable size variants of TLSs and their ir-

Bingxue Wang and Luying Gui are with School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing, 210094, P. R. China.

Liwen Zou and Xiaoping Yang are with Department of Mathematics, Nanjing University, Nanjing, 210093, P. R. China.

Jun Chen is with Department of Pathology, Nanjing Drum Tower Hospital, Nanjing, 210008, P. R. China.

Yingying Cao, Zhongqiu Wang and Jingya Chen are with Department of Radiology, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, 210000, P. R. China.

Zhenghua Cai is with Medical School, Nanjing University, Nanjing, 210007, P. R. China.

Yudong Qiu and Liang Mao are with Department of General Surgery, Nanjing Drum Tower Hospital, Nanjing, 210008, P. R. China.

Corresponding authors: Xiaoping Yang (e-mail: xpyang@nju.edu.cn) and Luying Gui (e-mail: ly.gui@njust.edu.cn)

regular distributions on the histopathology images. In addition, backgrounds with different colors and textures may interfere with the recognition of TLSs by humans. Also, an essential feature of the pancreas lacking immune response determines that there are fewer lymphocytes and TLSs in the pancreas, and it is more difficult to obtain a large number of training samples.

In order to accurately detect TLSs on a small-scale dataset containing whole slide images (WSIs) of hematoxylin and eosin (H&E) stained histopathology, we propose a weakly supervised segmentation network embedding cross-scale attention guidance and noise-sensitive constraint in this work. The proposed method is mainly divided into three parts: (1) all lymphocyte nuclei are segmented and identified from the unlabeled pancreatic pathology images by combining a pretrained nuclei segmentation model and adversarial learning. The corresponding lymphocyte density maps are constructed; (2) based on the aggregation characteristics of TLSs, we convert the lymphocyte density map in WSI to the attention guidance and establish a cross-scale attention network to learn the TLS features from different scales; (3) considering that the location of a TLS is critical information, we introduce a noise-sensitive constraint by a signed distance function to train a segmentation network with weak bounding box annotation. It is designed to make the network explicitly learn the TLSs' location distribution. To verify the effectiveness of our method in few-shot learning, the proposed method is trained on a small-scale dataset and evaluate on another large-scale dataset. The experimental results show that our method can accurately detect TLSs, and outperforms the state-of-the-art (SOTA) segmentation-based algorithms in terms of detection accuracy. As a clinical application, we use the density of TLSs to predict peripancreatic vascular invasion based on our proposed method.

The major contributions of this work are listed as follows.

- To the best of our knowledge, it is the first work to realize the few-shot detection task of TLSs on pancreatic histopathology images by embedding cross-scale attention guidance and noise-sensitive constraint into a weakly supervised segmentation network, which can be applied to the TLS analysis for other tumors.
- We construct the density map of lymphocytes which reflects the distribution characteristics of TLSs on pathological images and propose a cross-scale attention mechanism by jointly learning the TLS features from the coarse-scale pathological images and the fine-scale lymphocyte density maps.
- A noise-sensitive constraint with a signed distance function loss is introduced for training the TLS segmentation network with weak bounding box annotations, which helps to explicitly learn the TLS distribution and avoid enormous performance drops caused by tiny predicted errors.
- Experimental results show that our method significantly outperforms the state-of-the-art (SOTA) segmentation-based algorithms in terms of the TLS detection accuracy on pancreatic pathological images. And we validate that the TLS density is significantly related to the peripancre-

atic vascular invasion by our proposed method based on the clinical data acquired from two independent institutions.

II. METHOD

We propose a weakly supervised segmentation network for the few-shot detection of TLSs on pancreatic pathological images by embedding cross-scale attention guidance and noise-sensitive constraint to improve the TLS segmentation performance. The pipeline of our proposed method is shown in Figure 1. There are three core modules in our proposed model: (1) Lymphocyte nuclei segmentation and classification by combining nuclei segmentation baseline model and domain adversarial learning; (2) cross-scale attention guidance mechanism by jointly learning the TLS features from the coarse-scale H&E image and the fine-scale lymphocyte density map; (3) a noise-sensitive constraint by embedding a sign distance function loss for training the segmentation network with weak bounding box annotations.

When the proposed model is fed with an H&E WSI from a patient with pancreatic tumor, there are two processing flows for extracting cross-scale features of the TLS targets. The original H&E image is fed into a pretrained baseline model for nuclei segmentation and a domain adversarial network for lymphocyte nuclei classification. Then, we construct the lymphocyte density map as the fine-scale feature on each $d \times d$ tile image based on the previous predictions. Meanwhile, in order to learn the generic global context information, we get the coarse-scale feature from the original image by $d \times d$ mean pooling operation to make it the same size as the lymphocyte density map. Additionally, we utilize a reverse operation on the lymphocyte density map to make it be an attention map, termed as the lymphocyte density attention (LDA), which is compatible with the original H&E image in terms of the intensity distribution. Then, a cross-scale attention guidance mechanism based on a U-shape backbone with four-channel inputs is proposed to process the above cross-scale images. It learns the macroscopic features from the coarse-scale H&E image and microscopic features from the fine-scale lymphocyte density attention. Additionally, the proposed segmentation model is trained with bounding box annotations, which is weakly supervised for the segmentation task. Furthermore, a noise-sensitive constraint with a signed distance function loss (SDF) is used in the training procedure to explicitly learn the TLS distribution and avoid huge performance drops caused by tiny predicted errors.

A. Segmentation and classification for lymphocyte nuclei

The immune response in pancreatic tumors is usually not strong, and the distribution of immune cells is relatively sparse. In contrast, we note that the lymphocyte nuclei' morphological characteristics are independent of the organs in which they are located. The lymphocyte nuclei on pathology images of other organs should be morphologically identical to the pancreatic lymphocyte nuclei, although these images have different backgrounds. Therefore, we considered transferring the lymphocyte nuclei from a public dataset with lymphocyte

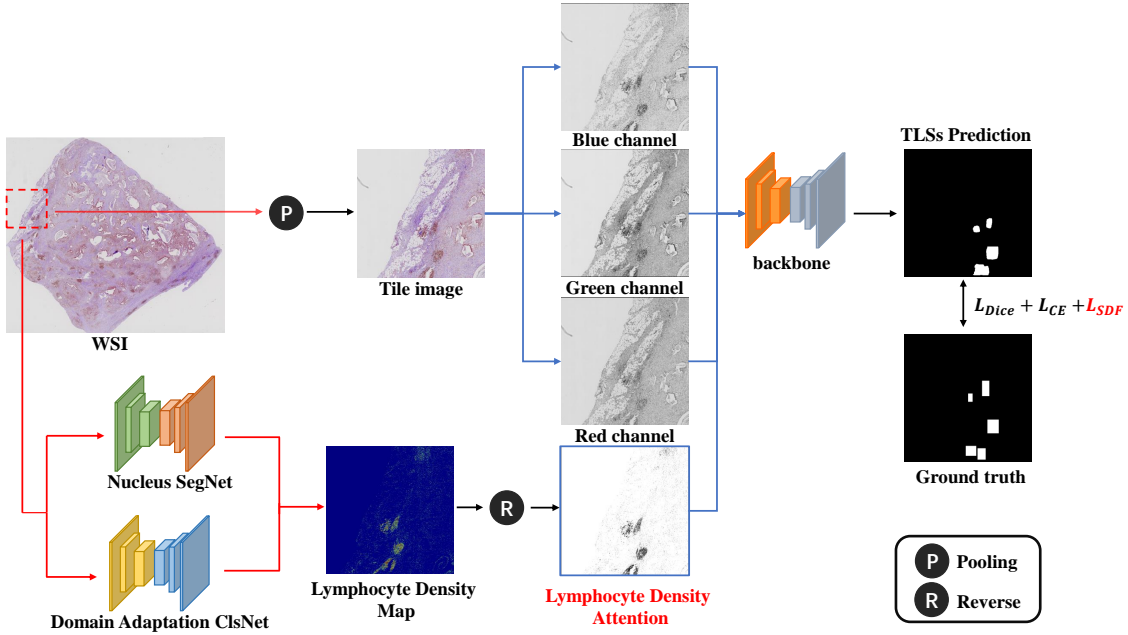


Fig. 1. Pipeline of our proposed segmentation network embedding cross-scale attention guidance and noise-sensitive constraint for weakly supervised TLSs segmentation of pancreatic cancer.

nuclei annotations of other organs to our pancreatic pathological dataset without lymphocyte nuclei annotations.

To this end, we introduce the combination of a pretrained baseline model for nuclei segmentation and a domain adversarial network for lymphocyte nuclei classification shown in Figure 2. HoVer-Net is a robust baseline model for nuclei segmentation and has demonstrated state-of-the-art performance on various nuclei segmentation tasks [14]. There are three branches in the HoVer-Net architecture: the segmentation branch for getting coarse prediction, the structure branch for solving the nuclei overlapping, and the classification branch for nuclei recognition. We use the segmentation and structure branches from the pretrained HoVer-Net in the PanNuke dataset [15] to segment the nuclei for our task. As for the classification, we train a domain adversarial neural network (DANN) [16] to get better performance on lymphocyte nuclei recognition. A domain adversarial training strategy is used in the classification network. Figure 3 demonstrates our training strategy for domain adversarial learning. We crop 239503 lymphocyte nuclei images and 265241 non-lymphocyte nuclei images from the PanNuke dataset with nuclei annotations. Then we obtain 121954 nuclei images with unknown categories (lymphocyte or non-lymphocyte nuclei) based on the segmentation results from the pretrained HoVer-Net. We choose the former as the source domain and the latter as the target domain to train the domain adversarial network for lymphocyte nuclei recognition. ResNet18 [17] is used as the feature extractor, and two loss terms are calculated as follows.

$$L_{cls} = \sum_{i=1}^M y_i \log \frac{1}{G_y(G_f(x_i))}, \quad (1)$$

$$L_{adv} = \sum_{i=1}^M d_i \log \frac{1}{G_d(G_f(x_i))} + (1 - d_i) \log \frac{1}{G_d(G_f(x_i))}, \quad (2)$$

where G_f , G_y , and G_d denote the feature extractor, image classifier, and domain discriminator, respectively. M denotes the number of training samples. x_i , y_i , and d_i denote the image input, nucleus category annotation, and domain label of sample i , respectively. The overall loss of DANN is defined by

$$L_{DANN} = L_{cls} + L_{adv} \quad (3)$$

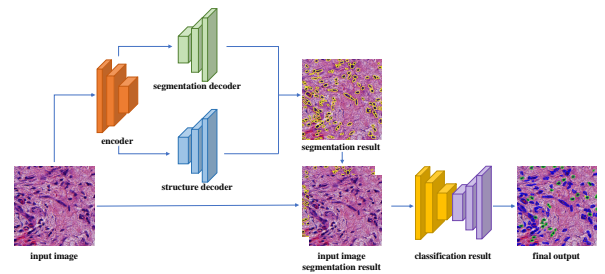


Fig. 2. Structure of the designed combination of the nuclei segmentation baseline model and domain adversarial network for recognizing lymphocyte nuclei.

B. Cross-scale attention guidance mechanism

TLSs are the particular forms of lymphocyte aggregations, and it is natural to look for TLSs where lymphocytes are relatively aggregated. Therefore, we calculate the lymphocyte

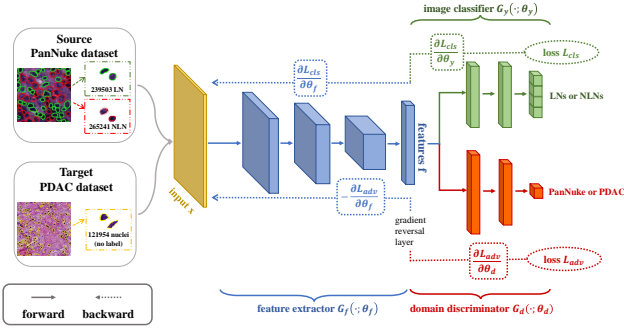


Fig. 3. The domain adversarial neural network training strategy based on the annotated lymphocyte nuclei dataset from PanNuke and our nuclei dataset obtained from the pretrained HoVer-Net model. LN and NLN denote lymphocyte nuclei and non-lymphocyte nuclei, respectively.

density map based on the detected lymphocyte nuclei and convert it to be attention map to guide the detection of TLSs. To obtain the lymphocyte density map on a WSI, we count the number N_{ij} of predicted lymphocyte nuclei in each non-overlapping $d \times d$ patch at location (i, j) . Each pixel occupies $25 \times 25 \mu\text{m}^2$. Then we get a gray-scale lymphocyte density map by the following formulation:

$$D_{ij} = 255 \times \frac{N_{ij} - N_{min}}{N_{max} - N_{min}}, \quad (4)$$

where N_{max} and N_{min} are the maximum and minimum of the predicted number of lymphocyte nuclei among the patches. Figure 4 shows the lymphocyte density map calculated for a WSI from our collected dataset. We split the three channels of the original H&E images into blue, green, and red channels shown in Figure 1. It can be found that the target TLS regions show low intensity in each color channel. We add a reverse operation to make the predicted lymphocyte density map compatible with the original H&E image to generate our desired attention map. Each lymphocyte density attention A_{ij} at location (i, j) is calculated as $A_{ij} = 255 - D_{ij}$. Then a cross-scale attention guidance network is established on the U-shape backbone architecture. There are four channel inputs of the backbone to jointly learn the coarse-scale features from the original H&E images and the fine-scale features from the calculated lymphocyte density attention.

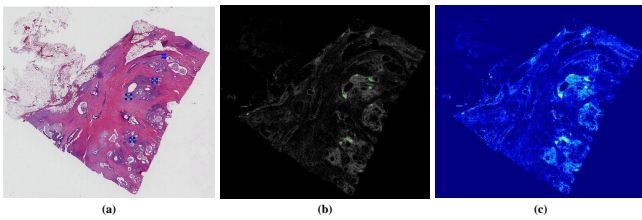


Fig. 4. The visualization of the calculated lymphocyte density map of a WSI from our collected dataset. (a) WSI with TLS annotations (shown in blue boxes); (b) Gray-scale image of the calculated lymphocyte nuclei density map, high density regions are denoted by green contours; (c) Heatmap of the calculated lymphocyte nuclei density.

C. SDF loss for noise-sensitive constraint

Considering that accurate positioning usually has high priority than finding the boundary of a TLS for clinical issues. Therefore, we train the proposed network with the bounding box annotations from the experts, which is weakly supervised for a segmentation task. Generally, it is difficult for a semantic segmentation network to detect small objects by training with general segmentation loss, such as Dice loss and cross-entropy (CE) loss, because these small regions almost have no impact on the overall loss. However, minor prediction errors can lead to misdiagnosis. To solve this problem, we introduce a noise-sensitive constraint by embedding a signed distance function (SDF) loss into the overall loss function for the training procedure. The signed distance function of a predicted binary mask Y can be calculated as follows:

$$SDF(x, Y) = \begin{cases} \min_{y \in \partial Y} -d(x, y), & x \in Y_{in} \\ 0, & x \in \partial Y \\ \min_{y \in \partial Y} d(x, y), & x \in Y_{out}, \end{cases} \quad (5)$$

where $d(x, y)$ is the Euclidian distance between x and y , and Y_{in} , Y_{out} and ∂Y denote the inside, the outside and the boundary of the object, respectively. Then we get the signed distance loss by calculating the mean square error (MSE) between the SDFs of segmentation and ground truth:

$$L_{SDF}(Y, \hat{Y}) = \frac{\sum_{i=1}^N (SDF(x_i, Y) - SDF(x_i, \hat{Y}))^2}{N}. \quad (6)$$

where N is the number of pixels in an input image. As shown in Figure 5, small segmentation errors significantly impact the SDF distribution. Besides, the traditional Dice and CE losses are also used in the training procedure. They can be calculated as follows:

$$L_{Dice} = -\frac{2 \sum_{i=1}^N s_i g_i}{\sum_{i=1}^N s_i^2 + \sum_{i=1}^N g_i^2}, \quad (7)$$

$$L_{CE} = -\frac{\sum_{i=1}^N g_i \ln p_i}{N}, \quad (8)$$

where s_i and g_i denote the predicted segmentation and the ground truth of pixel i , respectively. p_i denotes the softmax output of s_i . Therefore, the overall loss function of our proposed TLSs segmentation network can be formulated as follows:

$$L = L_{Dice} + L_{CE} + L_{SDF}. \quad (9)$$

III. EXPERIMENTS

A. Dataset

In this work, we evaluate our proposed method on two datasets collected from Nanjing Drum Tower Hospital (NDTH) and Jiangsu Province Hospital of Chinese Medicine (JHCM), respectively. The NDTH dataset consists of 38 WSIs from 12 surgical pathology-confirmed PDAC patients. The JHCM dataset is composed of 57 WSIs from 41 PDAC patients. To verify the generalization and the performance

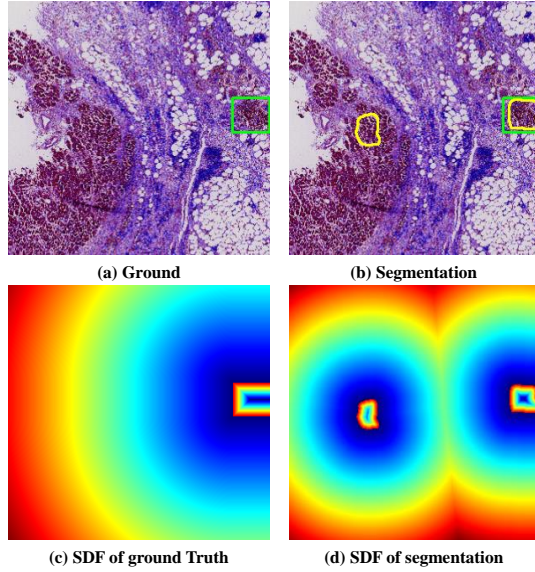


Fig. 5. Visual comparison between the SDF of ground truth and SDF of segmentation. (a) and (b) are the ground truth (green rectangle) and the segmentation (yellow contours) of TLSs. (c) and (d) are the corresponding signed distance function representation. It can be found that tiny segmentation errors can significantly affect the SDF distribution.

in few-shot learning of our proposed method, we train and validate our model on the smaller-scale NDTH dataset and test on the larger-scale JHCM dataset. This study was approved by the Ethics Committee of Nanjing Drum Tower Hospital and Jiangsu Province Hospital of Chinese Medicine.

The TLSs on the WSIs are annotated with bounding boxes by two experienced pathologists. The NDTH dataset is divided into four equal folds. One fold containing 9 WSIs is fixed as the validation set, and the other three folds containing 29 WSIs are adopted for training throughout the experiments. The JHCM dataset is only used for testing (i.e., zero-shot evaluation).

B. Evaluation metrics

We employ the precision, recall and F_β score to measure the detection accuracy in our experiments, which can be calculated as follows:

$$Precision = \frac{TP}{TP + FP}, \quad (10)$$

$$Recall = \frac{TP}{TP + FN}, \quad (11)$$

$$F_\beta = (1 + \beta^2) \times \frac{Precision \times Recall}{\beta^2 \times Precision + Recall}, \quad (12)$$

where TP , FP and FN denote the true positive, false positive and false negative TLS predictions, respectively. Generally, F_1 score (i.e., $\beta = 1$ in Eq. 15) is used to be the harmonic mean of both the precision and recall. However, because the false negative predictions generally bring greater harm than the false positive predictions in clinical issues, which means the recall

metric is more important than the precision metric. Therefore, we also use the F_2 score (i.e., $\beta = 2$ in Eq. 15) to evaluate the detection performance with greater weight on recall.

Moreover, because of the difference between the TLS annotations by different experts, especially in the regions where it is difficult to distinguish whether there is a single TLS or aggregated TLSs. Fig. 6 shows two examples to describe the annotation difference.

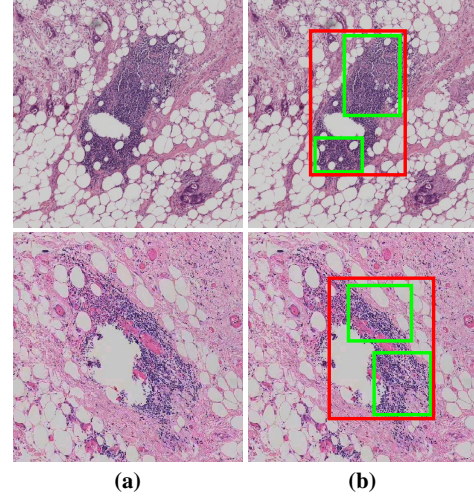


Fig. 6. Two examples to illustrate the annotation difference by different experts. (a) Image; (b) The TLS annotations by two experts which are denoted by red and green rectangles, respectively.

Therefore, we propose some new evaluation metrics to coordinate the annotation difference between the experts. Without loss of generality, since the precision describes how many positive cases are true among all the model predictions, we introduce the TPS here instead of TP to represent the TLS number of segmentation results that overlap with ground truths. Similarly, the recall describes how many annotated boxes are truly predicted, we introduce the TPB instead of TP to represent the number of annotated bounding boxes that overlap with the segmentation results. It can be seen that the TPS and TPB are generalizations of the general TP without the one-on-one constraint. Then, we define the corresponding segmentation precision (SP) and box recall (BR) as follows:

$$SP = \frac{TPS}{TPS + FP}, \quad (13)$$

$$BR = \frac{TPB}{TPB + FN}. \quad (14)$$

Similarly, the general F_β score (GF_β) can be calculated as follows:

$$GF_\beta = (1 + \beta^2) \times \frac{SP \times BR}{\beta^2 \times SP + BR}. \quad (15)$$

Fig. 7 illustrates the insight of the proposed metrics.

C. Ablation study

In this section, we present the ablation study to verify the effectiveness of our proposed LDA and SDF in our method.

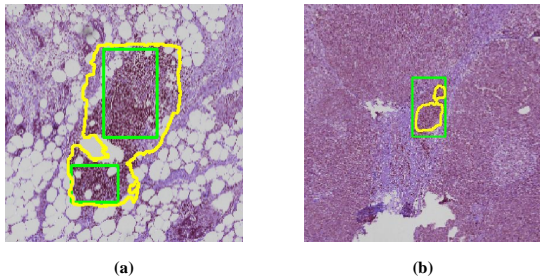


Fig. 7. Two examples to illustrate the proposed evaluation metrics. (a) When one predicted segmentation denoted by yellow contour covers two annotated boxes denoted by green rectangles, there are $TPS = 1$ and $TPB = 2$. (b) When one annotated box denoted by green rectangle is covered by two predicted segmentation denoted by yellow contours, there are $TPS = 2$ and $TPB = 1$.

We show the ablation study results in Table I to investigate the individual impact of the proposed LDA and SDF module.

1) *Effectiveness of LDA*: Without using LDA, the performance significantly drops on our collected datasets, leading to a decrease of 1.60%-3.56% in precision, 2.24%-5.42% in the recall, 3.16%-3.42% in F_1 score, and 2.70%-4.60% in F_2 score. For the introduced metrics, there is a performance decrease of 3.47%-6.11% in SP, 2.15%-7.23% in BR, 4.08%-5.32% in GF_1 score, and 3.01%-6.46% in GF_2 score without the LDA. Figure 8 shows the visual comparison of the ablation study for the proposed LDA. It can be found that the false positive predictions with low attention and false negative predictions with high attention can be efficiently captured.

2) *Effectiveness of SDF*: Without using the SDF loss, the performance also drops on our collected datasets, leading to a decrease of 2.39%-5.48% in precision, 1.85%-3.45% in F_1 score, and 1.27%-1.48% in F_2 score. For the introduced metrics, there is a performance decrease of 2.41%-3.88% in SP, 1.73%-2.41% in GF_1 score, and 0.51%-2.41% in GF_2 score without the SDF loss. It should be pointed that the SDF loss leads to an improvement of 1.21% in recall on the NDTH dataset while it leads to a decrease of 0.84% in recall on the JHCM dataset, which indicates that the SDF loss improves the overall performance of the detection model by mainly reducing the tiny false positive predictions instead of false negative predictions with relatively large sizes. Figure 9 presents the visual comparison of the ablation study for the SDF loss. We can see that tiny segmentation errors can cause huge losses in terms of the signed distance function map, and the SDF loss can significantly reduce the tiny false positive predictions.

3) *Effectiveness of LDA and SDF*: Additionally, it can be observed that without using LDA and SDF losses leads to the worst performance in the recall, F_1 score, F_2 score BR, GF_1 score, GF_2 score, indicating that both the proposed LDA and SDF losses are essential for performance improvements. It should be pointed that the baseline model (the proposed method without LDA and SDF loss) achieves the best performance in segmentation precision on NDTH dataset, while the proposed method outperforms the baseline model by 0.29% in the precision and achieves the best performance in all other metrics.

D. Comparison with other state-of-the-art segmentation methods

In this section, we compare our proposed method with other state-of-the-art (SOTA) methods for TLS detection. The UNet [18], Deeplab v3+ [19], and nnUNet [20] are included in our experiments for their excellent performance in medical image segmentation. All these methods are trained on the NDTH training set and evaluated on the NDTH validation set and the JHCM dataset. Quantitative comparison results are presented on Table II.

Experimental results in Table II demonstrate that our proposed method significantly outperforms these SOTA methods, and achieves the best performance in recall, F_1 score, F_2 score, SP, GF_1 score and GF_2 score. We observe that the generic UNet [18] and deeplab v3+ [19] always have conservative predictions by minimizing the false positive predictions as much as possible, despite causing many false negative predictions. Therefore, they have a higher precision and SP than ours, but leading to 18.07%-44.36%, 6.58%-21.01%, 13.68%-35.25%, 20.48%-47.35%, 9.25%-29.10% and 16.28%-40.68% decrease in recall, F_1 score, F_2 score, GF_1 score and GF_2 score, respectively comparing with ours. Figure 10 shows the visual comparison of TLS detection results between the above methods and ours.

E. Application for studying the relationship between TLS density and peripancreatic vascular invasion

For patients with pancreatic tumors accompanied by peripancreatic vascular invasion, especially venous invasion, whether the peripancreatic vessel is invaded determines the direction of treatment, and also affect the survival time and prognosis of patients [21]–[23]. In this section, we apply our proposed TLS detection method to study the relationship between TLS density and peripancreatic vascular invasion.

We collect the clinical information from 12 patients in NDTH dataset and 40 patients (there is a patient whose clinical information is not available) in JHCM dataset to carry out our statistical experiments. There are 19 patients without peripancreatic vascular invasion and 33 patients with peripancreatic vascular invasion. Therefore, we divide the 52 samples into two groups: the no-invasion group and the invasion group. Then we apply our proposed method to detect the TLSs on their WSIs, and calculate the TLS density of the patients in the two groups, respectively. The TLS density is calculated as the ratio of the TLS number to the WSI's area.

The Shapiro-Wilk (S-W) test [24] is conducted to check whether the TLS densities of the 52 patients obey the normal distribution. Based on the S-W test results of $p < 0.05$ and $p < 0.001$ on the invasion and no-invasion group, respectively (i.e., they do not obey the normal distribution), we use the Mann-Whitney U test [25] to observe the correlation between the TLS density and the peripancreatic vascular invasion. We get the result of $p = 0.03$ for the Mann-Whitney U test, which indicates that the TLS density is indeed related to the peripancreatic vascular invasion.

For comparison, we also study on the correlation between the lymphocyte density and the peripancreatic vascular inva-

TABLE I

ABLATION STUDY RESULTS OF THE PROPOSED METHOD ON TWO COLLECTED DATASETS. EXPERIMENTAL RESULTS ON THE NDTH VALIDATION SET AND THE JHCM DATASET ARE PRESENTED. W/O MEANS REMOVING THE CORRESPONDING MODULE. P, R, F1 AND F2 DENOTE THE PRECISION, RECALL, F_1 SCORE AND F_2 SCORE, RESPECTIVELY. SP, BR, GF1 AND GF2 REPRESENT THE INTRODUCED SEGMENTATION PRECISION, BOX RECALL, GF_1 SCORE AND GF_2 SCORE, RESPECTIVELY.

Datasets	Methods	P	R	F1	F2	SP	BR	GF1	GF2
NDTH	Proposed	78.21	84.34	81.16	83.04	84.80	87.95	86.34	87.30
	Proposed w/o L_{SDF}	75.82	83.13	79.31	81.56	82.39	85.54	83.93	84.89
	Proposed w/o LDA	76.61	78.92	77.74	78.44	81.33	80.72	81.02	80.84
	Proposed w/o LDA, L_{SDF}	77.92	72.29	75.00	73.35	86.21	74.70	80.04	76.75
JHCM	Proposed	62.25	83.13	71.19	77.90	76.15	87.70	81.52	85.12
	Proposed w/o L_{SDF}	56.77	83.97	67.74	76.63	72.87	88.16	79.79	84.61
	Proposed w/o LDA	58.69	80.89	68.03	75.20	70.74	85.55	77.44	82.11
	Proposed w/o LDA, L_{SDF}	59.98	77.17	62.12	70.35	69.89	81.36	75.19	78.77

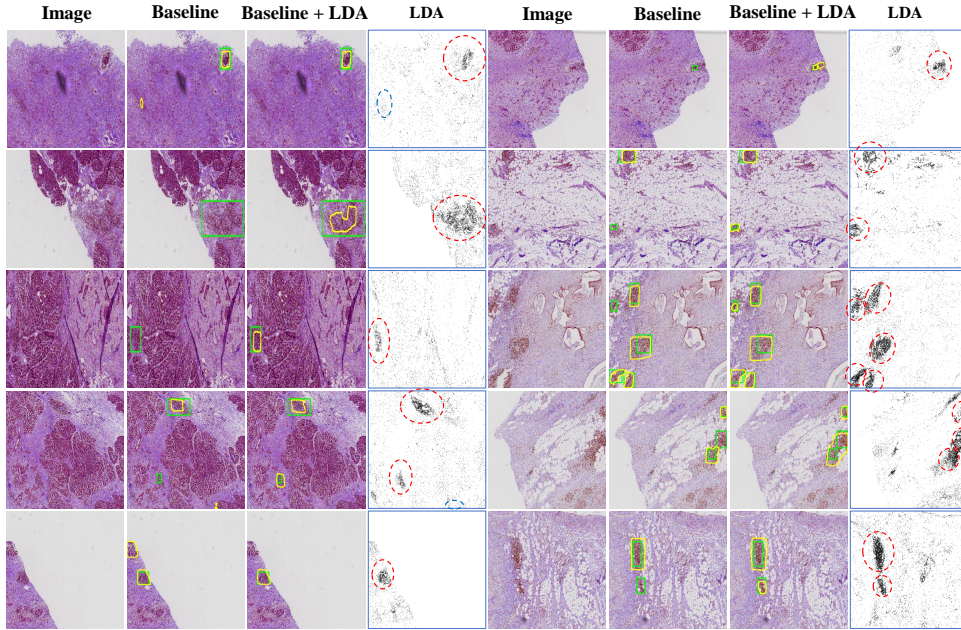


Fig. 8. Visual comparison of the ablation study for the proposed LDA. The baseline model is the proposed method without LDA and SDF loss. Regions with high and low attention on the calculated LDAs are signed with red and blue dashed circles, respectively. It can be found that the calculated LDA can significantly improve the TLS detection performance.

TABLE II

QUANTITATIVE COMPARISON RESULTS WITH OTHER SOTA METHODS FOR TLSS DETECTION ON TWO COLLECTED DATASET. EXPERIMENTAL RESULTS ON THE NDTH VALIDATION SET AND THE JHCM DATASET ARE PRESENTED. P, R, F1 AND F2 DENOTE THE PRECISION, RECALL, F_1 SCORE AND F_2 SCORE, RESPECTIVELY. SP, BR, GF1 AND GF2 REPRESENT THE INTRODUCED SEGMENTATION PRECISION, BOX RECALL, GF_1 SCORE AND GF_2 SCORE, RESPECTIVELY.

Datasets	Methods	P	R	F1	F2	SP	BR	GF1	GF2
NDTH	UNet [18]	85.27	66.27	74.58	69.36	89.92	67.47	77.09	71.02
	DeepLab v3+ [19]	84.68	63.25	72.41	66.62	85.48	63.86	73.10	67.26
	nnUNet [20]	77.92	72.29	75.00	73.35	86.21	74.70	80.04	76.75
	Proposed	78.21	84.34	81.16	83.04	84.80	87.95	86.34	87.30
JHCM	UNet [18]	76.57	45.39	56.99	49.11	82.88	49.02	61.60	53.38
	DeepLab v3+ [19]	71.11	38.77	50.18	42.65	74.78	40.35	52.42	44.44
	nnUNet [20]	59.98	77.17	62.12	70.35	69.89	81.36	75.19	78.77
	Proposed	62.25	83.13	71.19	77.90	76.15	87.70	81.52	85.12

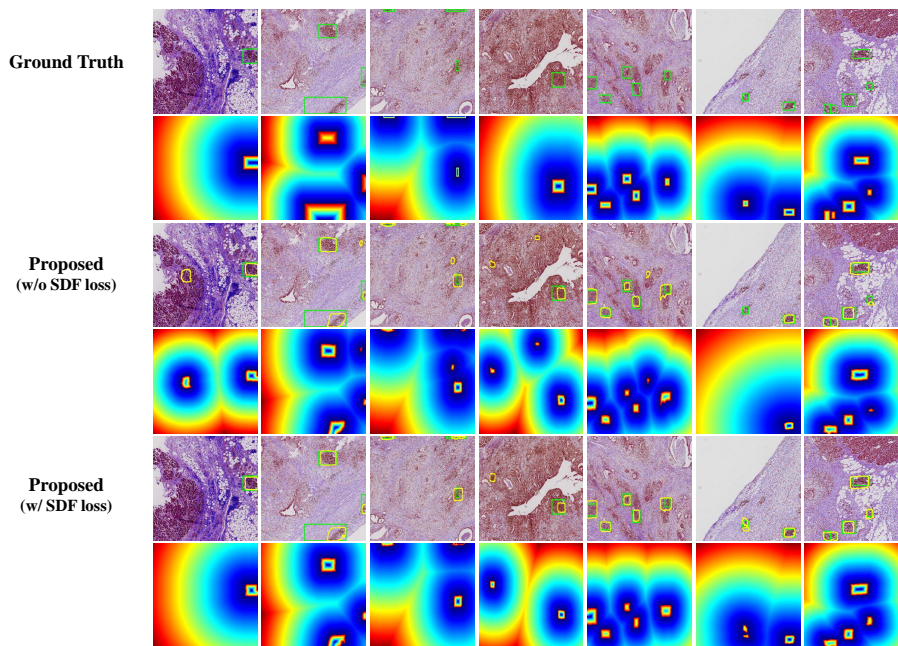


Fig. 9. Visual comparison of the ablation study for the SDF loss. Segmentation and ground truth are denoted by green and yellow contours in pathological images, respectively. The SDF maps are shown under the corresponding images.

sion. The same statistical testing experiments are conducted for the lymphocyte density. We get the S-W test results of $p < 0.01$ on the invasion group (i.e., they do not obey the normal distribution). We use the Mann-Whitney U test to observe the correlation between the TLS density and the peripancreatic vascular invasion. The we get the result of $p = 0.1$ for the Mann-Whitney U test, which indicates that the lymphocyte density has no significant relationship to the peripancreatic vascular invasion. Figure 11 presents the distributions of the TLS density and lymphocyte density for the two groups.

IV. DISCUSSIONS

In this paper, we develop a novel weakly-supervised segmentation framework for TLS detection. Since TLSs are aggregates of lymphocytes, we utilize the density map of lymphocyte nuclei as an attention guidance in our approach. Although the segmentation of individual lymphocyte nucleus may not affect the TLS detection, the overall accuracy of lymphocyte nuclei segmentation results may still impact the detection results. Therefore, we also show the lymphocyte nuclei segmentation results in this section. As described in Section II-A, we have used a domain-adversarial approach to segment lymphocyte nuclei on pancreatic pathology images.

We show the segmentation and classification results of our designed combination of the baseline model for nuclei segmentation and the domain adversarial network for lymphocyte nuclei recognition in Table III. The segmentation accuracy, sensitivity, and specificity for the nucleus achieve 86.20%, 76.29%, and 80.94%, respectively. The classification accuracy, sensitivity, and specificity for lymphocyte and non-lymphocyte nuclei achieve 88.87%, 94.79%, and 34.41%, respectively. Figure 12 presents two visual results for the segmentation and classification, respectively.

TABLE III
SEGMENTATION AND CLASSIFICATION RESULTS FOR LYMPHOCYTE NUCLEUS ON OUR TEST DATASET.

Methods	Accuracy	Sensitivity	Specificity
HoVerNet	86.20	76.29	80.94
DANN	88.87	94.79	34.41

We would like to highlight that we have opted for a simple U-shape segmentation network to detect TLSs rather than a complex object detection framework. This decision is based on the fact that detection frameworks usually require a larger amount of annotated data and parameters to be trained compared to the U-shape segmentation network. Therefore, in the context of few-shot learning, the detection framework does not notably outperform the U-shape network. It is worth noting that the designed combination of the baseline model for nuclei segmentation and domain adversarial network for lymphocyte nuclei classification can still be improved, especially in terms of lymphocyte nuclei recognition specificity.

The proposed LDA improves the TLS segmentation performance by multi-channel fusion. However, the proposed LDA can also be embedded in the U-shape network architecture or used as an extra loss function for training, which can be studied in our subsequent work. Additionally, we set the patch size d for coarse-scale pooling and fine-scale density calculation as 32 based on medical experts' experience.

Our proposed LDA shows an improvement in TLS segmentation performance through multi-channel fusion. However, we intend to investigate further the potential benefits of embedding the proposed LDA in the U-shape network architecture or using it as an additional loss function during training in

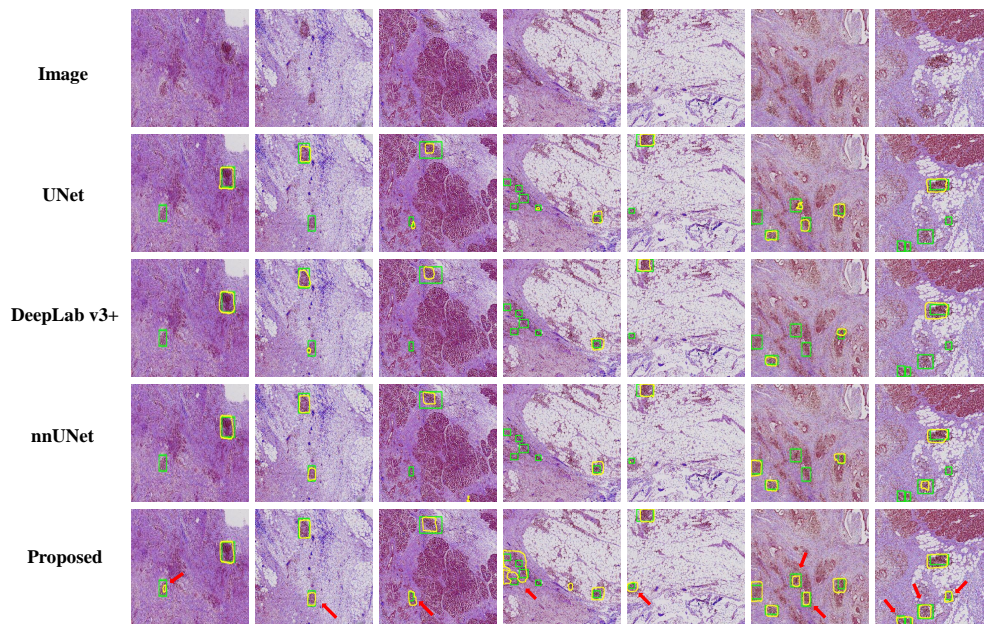


Fig. 10. Visual comparison results between the SOTA segmentation methods and our proposed method. The green and yellow denote the ground truth and predictions, respectively. Improved predictions are signed by red arrows.

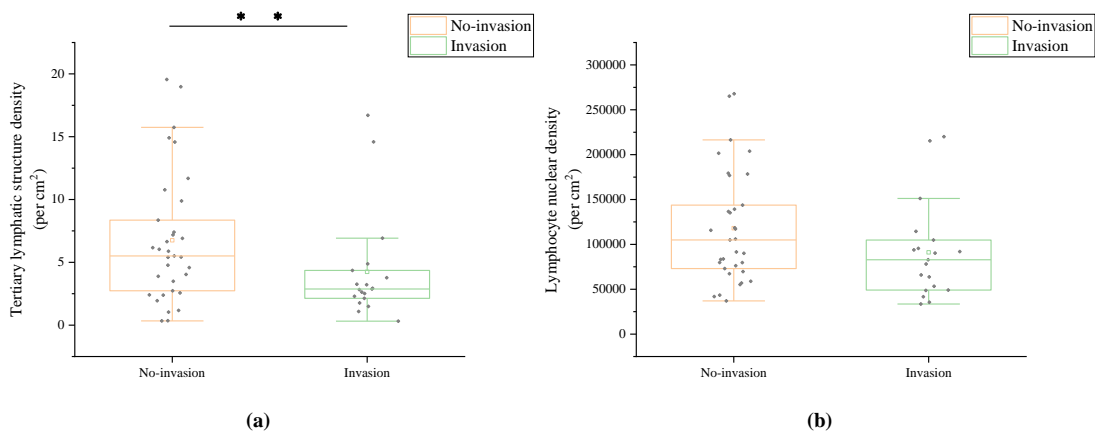


Fig. 11. The distribution of (a) TLS density and (b) lymphocyte density for two groups (the no-invasion group and the invasion group). ** indicates $p < 0.05$ using the Mann-Whitney U test.

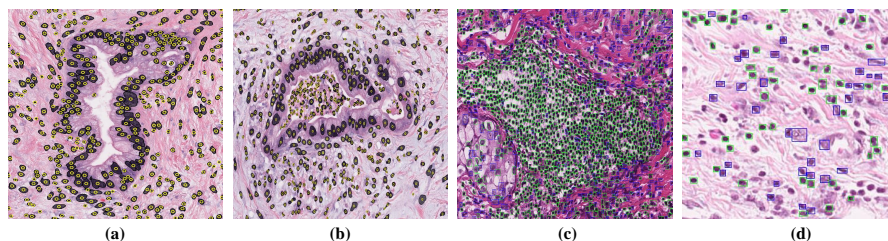


Fig. 12. Segmentation and classification results of the designed combination of baseline model for nuclei segmentation and domain adversarial network for lymphocyte nuclei recognition. (a)-(b): Segmentation results. The predicted nuclei are shown in dark masks whose centers are marked by yellow circles. (c)-(d): Classification results. The lymphocyte and non-lymphocyte nuclei are shown in green and blue boxes, respectively.

future work. Furthermore, we have chosen a patch size d of 32 for coarse-scale pooling and fine-scale density calculation based on medical experts' recommendations.

Regarding the detected TLSs, we have examined the relationship between TLSs and vascular invasion. Through a study of two central patients, we found that TLSs have a stronger association with vascular invasion than individual lymphocytes, highlighting TLSs as an important component of the tumor immune microenvironment. With the expansion of sample size and the subdivision of lymphocyte categories, we will conduct more detailed and in-depth investigations of the immune microenvironment.

V. CONCLUSIONS

In this paper, we propose a novel weakly supervised segmentation network embedding cross-scale attention guidance and noise-sensitive constraint for TLS detection. We firstly obtain the segmentation and classification results of the lymphocyte nuclei by combining a pretrained baseline model for nuclei segmentation and a domain adversarial network for lymphocyte nuclei recognition. Then, we establish a cross-scale attention guidance network by jointly learning the coarse-scale features from the original H&E images and fine-scale features from our calculated lymphocyte density attention. A noise-sensitive constraint is introduced by embedding the signed distance function loss in the training procedure to reduce tiny segmentation errors. Experimental results on two collected datasets demonstrate that our proposed algorithm outperforms the state-of-the-art segmentation-based methods for TLS detection. Additionally, we apply our method to validate that the TLS density is significantly related to the peripancreatic vascular invasion based on the clinical data acquired from two independent institutions. Our proposed approach can be applied to the TLS analysis for the tumors in other organs.

ACKNOWLEDGMENT

This work is supported by China's Ministry of Science and Technology(No. SQ2020YFA0713800), the National Natural Science Foundation of China(No. 11971229, 12001273), and the Fundamental Research Funds for the Central Universities(No.30922010904).

REFERENCES

- [1] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," *Medical Image Analysis*, vol. 33, pp. 170–175, 2016.
- [2] A. L. Kiemen, A. M. Braxton, M. P. Grahn, K. S. Han, J. M. Babu, R. Reichel, A. C. Jiang, B. Kim, J. Hsu, F. Amoa, *et al.*, "Coda: quantitative 3d reconstruction of large tissues at cellular resolution," *Nature Methods*, pp. 1–10, 2022.
- [3] J. Zhang, Z. Hua, K. Yan, K. Tian, J. Yao, E. Liu, M. Liu, and X. Han, "Joint fully convolutional and graph convolutional networks for weakly-supervised segmentation of pathology images," *Medical Image Analysis*, vol. 73, p. 102183, 2021.
- [4] G. Xu, Z. Song, Z. Sun, C. Ku, Z. Yang, C. Liu, S. Wang, J. Ma, and W. Xu, "Camel: A weakly supervised learning framework for histopathology image segmentation," in *Proceedings of the IEEE/CVF International Conference on computer vision*, pp. 10682–10691, 2019.
- [5] L. Munoz-Erazo, J. L. Rhodes, V. C. Marion, and R. A. Kemp, "Tertiary lymphoid structures in cancer—considerations for patient prognosis," *Cellular & Molecular Immunology*, vol. 17, no. 6, pp. 570–575, 2020.
- [6] C. Sautès-Fridman, F. Petitprez, J. Calderaro, and W. H. Fridman, "Tertiary lymphoid structures in the era of cancer immunotherapy," *Nature Reviews Cancer*, vol. 19, no. 6, pp. 307–325, 2019.
- [7] T. N. Schumacher and D. S. Thommen, "Tertiary lymphoid structures in cancer," *Science*, vol. 375, no. 6576, p. eabf9419, 2022.
- [8] G. Trajkovski, L. Ognjenovic, Z. Karadzov, G. Jota, D. Hadzi-Manchev, O. Kostovski, G. Volceviski, V. Trajkovska, D. Nikolova, L. Spasevska, *et al.*, "Tertiary lymphoid structures in colorectal cancers and their prognostic value," *Open Access Macedonian Journal of Medical Sciences*, vol. 6, no. 10, p. 1824, 2018.
- [9] F. Bergomas, F. Grizzi, A. Doni, S. Pesce, L. Laghi, P. Allavena, A. Mantovani, and F. Marchesi, "Tertiary intratumor lymphoid tissue in colo-rectal cancer," *Cancers*, vol. 4, no. 1, pp. 1–10, 2011.
- [10] A. Heindl, I. Sestak, K. Naidoo, J. Cuzick, M. Dowsett, and Y. Yuan, "Relevance of spatial heterogeneity of immune infiltration for predicting risk of recurrence after endocrine therapy of ER+ breast cancer," *JNCI: Journal of the National Cancer Institute*, vol. 110, no. 2, pp. 166–175, 2018.
- [11] K. Mortezaee, "Enriched cancer stem cells, dense stroma, and cold immunity: Interrelated events in pancreatic cancer," *Journal of Biochemical and Molecular Toxicology*, vol. 35, no. 4, p. e22708, 2021.
- [12] A. Redding and E. Grabocka, "A splendid new beginning at the end of a 40-year quest: The first krasg12d inhibitor in pancreatic cancer," *Cancer Discovery*, vol. 13, no. 2, pp. 260–262, 2023.
- [13] S. J. Rubin, R. S. Sojwal, J. Gubatan, and S. Rogalla, "The tumor immune microenvironment in pancreatic ductal adenocarcinoma: Neither hot nor cold," *Cancers*, vol. 14, no. 17, p. 4236, 2022.
- [14] S. Graham, Q. D. Vu, E. Shan, A. Azam, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical Image Analysis*, vol. 58, p. 101563, 2019.
- [15] J. Gamper, N. A. Koohbanani, K. Benes, S. Graham, M. Jahanifar, S. A. Khurram, A. Azam, K. Hewitt, and N. Rajpoot, "Pannuke dataset extension, insights and baselines," *arXiv preprint arXiv:2003.10778*, 2020.
- [16] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [18] O. Ronneberger, P. Fischer, and T. Brox, "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.
- [19] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.
- [20] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, *et al.*, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [21] E. M. Loyer, C. L. David, R. A. Dubrow, D. B. Evans, and C. Charnsangavej, "Vascular involvement in pancreatic adenocarcinoma: reassessment by thin-section ct," *Abdom Imaging*, vol. 21, no. 3, pp. 202–6, 1996.
- [22] M. Klaus, A. Mohr, H. Von Tengg-Kobligh, H. Friess, R. Singer, P. Seidensticker, H. U. Kauczor, G. M. Richter, G. W. Kauffmann, and L. Grenacher, "A new invasion score for determining the resectability of pancreatic carcinomas with contrast-enhanced multidetector computed tomography," *Pancreatology*, vol. 8, no. 2, pp. 204–210, 2008.
- [23] K. Teramura, T. Noji, T. Nakamura, T. Asano, and S. Hirano, "Pre-operative diagnosis of portal vein invasion in pancreatic head cancer: appropriate indications for concomitant portal vein resection," *Journal of Hepato Biliary Pancreatic Sciences*, vol. 23, no. 10, pp. 643–649, 2016.
- [24] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality," *Biometrika*, no. 3–4, pp. 591–611, 1965.
- [25] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.