Medical Image Analysis 80 (2022) 102487

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels

Chu Han[a,b,c,1], Jiatai Lin[c,d,1], Jinhai Mai[a,c,1], Yi Wang[e], Qingling Zhang[f], Bingchao Zhao[a,c], Xin Chen[g], Xipeng Pan[a,b,c], Zhenwei Shi[a,b,c], Zeyan Xu[a,c], Su Yao[f], Lixu Yan[f], Huan Lin[a,c], Xiaomei Huang[a,c], Changhong Liang[a,c,**], Guoqiang Han[d,**], Zaiyi Liu[a,c,*]

[a] Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, China
[b] Guangdong Cardiovascular Institute, Guangzhou, Guangdong 510080, China
[c] Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, China
[d] The School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China
[e] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
[f] Department of Pathology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, China
[g] Department of Radiology, Guangzhou First People's Hospital, The Second Affiliated Hospital of South China University of Technology, Guangzhou, Guangdong 510180, China

ARTICLE INFO

Article history:
Received 6 November 2021
Revised 7 May 2022
Accepted 20 May 2022
Available online 24 May 2022

Keywords:
Computational pathology
Tissue segmentation
Weakly-supervised learning
Pseudo mask generation

ABSTRACT

Tissue-level semantic segmentation is a vital step in computational pathology. Fully-supervised models have already achieved outstanding performance with dense pixel-level annotations. However, drawing such labels on the giga-pixel whole slide images is extremely expensive and time-consuming. In this paper, we use only patch-level classification labels to achieve tissue semantic segmentation on histopathology images, finally reducing the annotation efforts. We propose a two-step model including a classification and a segmentation phases. In the classification phase, we propose a CAM-based model to generate pseudo masks by patch-level labels. In the segmentation phase, we achieve tissue semantic segmentation by our propose Multi-Layer Pseudo-Supervision. Several technical novelties have been proposed to reduce the information gap between pixel-level and patch-level annotations. As a part of this paper, we introduce a new weakly-supervised semantic segmentation (WSSS) dataset for lung adenocarcinoma (LUAD-HistoSeg). We conduct several experiments to evaluate our proposed model on two datasets. Our proposed model outperforms five state-of-the-art WSSS approaches. Note that we can achieve comparable quantitative and qualitative results with the fully-supervised model, with only around a 2% gap for MIoU and FwIoU. By comparing with manual labeling on a randomly sampled 100 patches dataset, patch-level labeling can greatly reduce the annotation time from hours to minutes. The source code and the released datasets are available at: https://github.com/ChuHan89/WSSS-Tissue.

1. Introduction

Tumor microenvironment (TME), not only plays a vital role in tumor initiation and progression (Hanahan and Weinberg, 2011), but also influences the therapeutic effect and prognosis of cancer patients (Skrede et al., 2020; AbdulJabbar et al., 2020). TME is formed with different types of tissues, including tumor epithelial, tumor-infiltrating lymphocytes (TILs), tumor-associated stroma and etc. They have been proven to be clinically relevant with tu-

* Corresponding author at: Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, China.
E-mail addresses: hanchu@gdph.org.cn (C. Han), liangchanghong@gdph.org.cn (C. Liang), csgqhan@scut.edu.cn (G. Han), liuzaiyi@gdph.org.cn (Z. Liu).
[1] Contributed equally.
** Co-corresponding authors

https://doi.org/10.1016/j.media.2022.102487

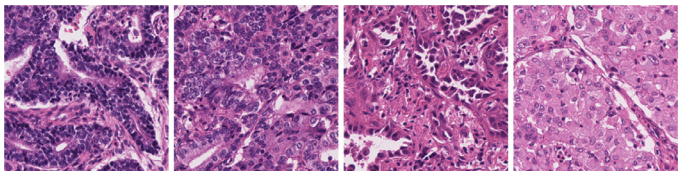

Fig. 1. Demonstration of tumor heterogeneity. Four patches with tumor epithelial from the WSIs of four lung adenocarcinoma patients.

mor progression by previous studies. TILs was considered as prognostic biomarkers in numerous solid tumors, such as lung cancer (Brambilla et al., 2016), breast cancer (Denkert et al., 2018) and colorectal cancer (Kong et al., 2019). While the crosstalk between tumor epithelial and tumor-associated stroma has been associated with tumor progression (Mao et al., 2013; Bremnes et al., 2011). Therefore, it is essential to differentiate and segment different types of tissues for precise quantification of TME.

Conventional approaches perform tissue segmentation by using hand-crafted features, such as textures (Diamond et al., 2004; Sirinukunwattana et al., 2015), morphological features (Anoraganingrum, 1999), color (Tabesh et al., 2007) and etc. Recently, deep learning (LeCun et al., 2015) demonstrates its superiority and shows tremendous success in medical image segmentation tasks (Chen et al., 2016; Zhao et al., 2020; Qaiser et al., 2019). However, collecting dense pixel-level annotations is expensive and labor-intensive, especially for histopathology images. Because of the diversity and complexity, only pathologists or people with the clinical background can handle it. Moreover, due to the heterogeneity and the aggressiveness of tumors, it may exhibit various morphological appearances, as demonstrated in Fig. 1. In the meanwhile, the giga-pixel of the whole slide image (WSI) also increases the difficulty of manual labeling.

Researchers have made attempts to overcome the difficulties of dense annotation acquisition and reduce the annotation efforts, such as active learning (Yang et al., 2017), semi-supervised learning (Liu et al., 2020), learning from sparse annotations (Liang et al., 2018), and weakly-supervised learning (Gao et al., 2020). Class activation mapping (Zhou et al., 2016) is the most common solution for weakly-supervised semantic segmentation (WSSS). The basic idea is to train a classification network and leverage the semantic information from the deeper layers to achieve segmentation. Chan et al. (2019) applied a CAM-based method with a series of post-processing steps for weakly-supervised tissue semantic segmentation. However, CAM-based methods face a great challenge that the classification network tends to differentiate objects by the most discriminative features but the segmentation task aims to find the complete object. The activated regions will gradually shrink and harm the segmentation results. For histopathology images, such contradiction will be amplified because the spatial arrangement of different tissues is relatively random and scatter comparing with the natural images.

In this paper, we present a simple and effective CNN model for patch-level semantic segmentation of histopathology tissues using only patch-level annotations. Pathologists only need to judge the presence or absence of the different tissue categories in the patches instead of carefully drawing the boundaries of tissues on the WSIs, which greatly saves the annotation time. The basic idea of this model is to use patch-level classification labels to automatically generate pixel-level semantic segmentation masks, and then use the generated pseudo masks to train a semantic segmentation model. Our proposed model contains a classification phase and a segmentation phase. In the classification phase, we proposed a CAM-based classification model for pseudo mask generation. To avoid the discriminative region shrinkage problem, we proposed a Progressive-Dropout Attention (PDA) to progressively deactivate the highlighted regions, and push the classification network to differentiate the tissue categories by the non-predominant regions. In the segmentation phase, we train a semantic segmentation model by the pseudo masks generated from multiple layers of the classification network, we called it Multi-Layer Pseudo-Supervision (MLPS). MLPS can provide information from different stages to reduce the information gap between patch-level and pixel-level labels. Due to the long-tail and unbalanced distribution problem, some tissue categories with fewer training samples may not be able to learn a good feature representation from pseudo masks, which could easily lead to false-positive segmentation results. To tackle this problem, we proposed a classification gate mechanism to reduce the false-positive rate for the non-predominant tissue categories. In order to achieve WSI-level semantic segmentation, we provide a "tessellate → segment → tile" solution.

In addition, we introduced a new weakly-supervised tissue semantic segmentation dataset for lung adenocarcinoma (LUAD-HistoSeg), which is the first tissue-level semantic segmentation dataset for LUAD. There are four different types in this dataset, tumor epithelial (TE), tumor-associated stroma (TAS), lymphocyte (LYM) and necrosis (NEC), including 16,678 patches with multi-label binary vectors and 607 patches with pixel-level labels under 10× magnification.

We evaluate our proposed model on two datasets, LUAD-HistoSeg and Breast Cancer Semantic Segmentation (BCSS) (Amgad et al., 2019). Extensive experiments and ablation studies have demonstrated the superiority of our proposed model on semantic segmentation using only patch-level annotations. Comparing with the fully-supervised model, our model shows comparable quantitative and qualitative results with only around a 2% gap for MIoU and FwIoU. The proposed model has been proven to be 10× faster than manual labeling. The main contributions of this paper are summarized as follows:

- We present a tissue semantic segmentation model for histopathology images using only patch-level classification labels, which greatly saves the annotation time for pathologists.
- Multi-layer pseudo-supervision with progressive dropout attention is proposed to reduce the information gap between patch-level and pixel-level labels. And a classification gate mechanism is introduced to reduce the false-positive rate.
- Our proposed model achieves state-of-the-art performance comparing with weakly-supervised semantic segmentation models on two datasets, as well as a comparable performance with fully-supervised baseline.
- The first lung adenocarcinoma (LUAD) dataset is released for weakly-supervised tissue semantic segmentation.

2. Related works

2.1. Histopathology image segmentation

Computational pathology (Srinidhi et al., 2020; Deng et al., 2020) has attracted much attention in recent years with the advance of deep learning techniques. Histopathology image segmentation is the most vital process in computer-aided histopathology image analysis. With the data-driven nature, various segmentation approaches have been carried out and achieved outstanding performance, such as tissue segmentation (van Rijthoven et al., 2021), gland segmentation (Chen et al., 2016; Wen et al., 2021), nuclei segmentation (Zhao et al., 2020; Graham et al., 2019) and etc.

To prepare sufficient manually labeled data for training CNN models, pathologists have to carefully draw pixel-level labels on a giga-pixel whole slide image, which is extremely expensive and time-consuming. Due to the heterogeneity of malignant tumors,

even the same tumor type can show totally different morphological appearances. Therefore, people without clinical backgrounds are not qualified for this job. How to reduce the annotation efforts is still an open problem.

2.2. Reducing annotation efforts for medical image segmentation

Recently, researchers attempt to reduce the costly annotation burden in different technical perspectives, such as active learning, coarse segmentation by patch-level classification, semi-supervised learning and weakly-supervised learning.

2.2.1. Active learning

Active learning (AL) (Settles, 2009; Budd et al., 2021) is a human-in-the-loop learning strategy for alleviating annotation burden. It allows human annotators to revisit and refine the uncertain pseudo-labels generated by machine learning algorithms (Wen et al., 2018), or automatically selects the most informative samples to be labeled next (Yang et al., 2017; Zhou et al., 2021). Mahapatra et al. (2018) proposed to associate GAN with AL to further reduce the annotation effort which saves around 65% annotations for classification and segmentation on a chest XRay dataset. Doyle et al. (2011) proposed AL with a class-balancing strategy to solve the minority class problem, which achieved a higher accuracy for patch-level classification of non-cancer regions in prostate histopathology image compared with random sampling strategy. Belharbi et al. (2021) proposed to incorporate classification with image-level annotations and segmentation with generated pseudo pixel-level labels for both histopathology image segmentation (gland segmentation) and natural image segmentation (bird species). A deeply supervised active learning (DSAL) (Zhao et al., 2021) has been proposed to assign the samples with high uncertainties to strong labelers and the samples with low uncertainties to weak labelers. Shen et al. (2020) considered dissatisfaction, representativeness and diverseness of the samples in AL sampling strategy, for breast cancer segmentation in IHC whole slide images.

2.2.2. Patch-level classification

An alternative solution to alleviate dense pixel-level annotations is to reform the semantic segmentation problem to the patch-level classification problem. And a series of studies have proven the effectiveness of this way by successful diagnostic and survival prediction (Kather et al., 2019b; Kather et al., 2019a; Hou et al., 2016). Several patch-level histopathology classification models have been proposed to avoid densely pixel-level annotation. The most common way is to transfer a classification model trained on ImageNet (Krizhevsky et al., 2012) to the target histopathology domain. Ni et al. (2019) reduced the inference speed by discarding easier-recognized non-malignant regions in the lower layers and letting the higher layer focus on differentiating more complex cancerous regions. Rkaczkowski et al. (2019) proposed an accurate, reliable and active model for histopathology image classification, which segmented whole slide images of colorectal cancer into eight tissue types. Lin et al. (2022) proposed to associate deep learning with broad learning to reduce the annotation efforts. This approach can achieve moderate accuracy even with only 1% training samples. However, patch-level classification can only perform rough segmentation and sacrifices the pixel-level accuracy for a computationally efficient inference time and lower annotation efforts.

2.2.3. Semi-supervised learning

Semi-supervised learning (Blum and Mitchell, 1998) aims to leverage a small set of labeled samples and a large set of unlabeled samples to train the model. To maximize the value of the limited labels, existing works either try to maintain the consistency by competing for the introduced perturbations (Laine and Aila, 2016; Tarvainen and Valpola, 2017) or seek the relationship among different samples (Liu et al., 2019; Battaglia et al., 2018). Self-supervised learning (Zhai et al., 2019; Cheng et al., 2020; Cai et al., 2021; Kim et al., 2021) is a feasible way to learn the visual representation for semi-supervised learning, which can somehow be a complement to the lack of annotations. Specific to medical image segmentation, Xia et al. (2020) proposed uncertainty-aware multi-view co-training for 3D volumetric medical image segmentation. Marini et al. (2021) used a semi-supervised semantic segmentation teacher model to train a semi-weakly supervised student model and achieved prostate histopathology image classification. Li et al. (2020) proposed self-loop uncertainty to generate pseudo labels by a Jigsaw puzzle-solving self-supervised task. Xie et al. (2020) introduced a pair relation network (PR-net) to learn a better image representation by comparing a pair of images in the feature space. Then the well-trained PR-net could be transferred to a gland segmentation network in a semi-supervised learning manner.

2.2.4. Weakly-supervised learning

Besides scarcer annotations with semi-supervised learning, many works focus on using weaker or sparser labels for model training, such as image-level labels (Wang et al., 2020a), point-based annotations (Bearman et al., 2016), bounding boxes (Dai et al., 2015), scribbles (Lin et al., 2016) and etc.

Researchers proposed multiple instance learning models with weak supervision for cancerous region segmentation (Jia et al., 2017; Lerousseau et al., 2020). Lee and Jeong (2020) proposed to use scribble annotations to automatically generate pseudo-labels for microscopic cell segmentation. Qu et al. (2019, 2020) proposed a two-stage weakly supervised learning model with only a small set of point annotations for nuclei segmentation. They first used a semi-supervised model to detect the center points of all the nuclei, and then designed a weakly supervised model for nuclei segmentation. Tokunaga et al. (2020) leveraged the proportion of the tissue subtypes to generate pseudo labels. Zhang et al. (2021) used foreground proportion as the weak labels and then combine FCN and graph convolutional networks (FGNet) for automatic tissue segmentation. Wang et al. (2019) proposed a ScanNet to first train a classification model under the lower resolution and inference with an FCN structure under the higher resolution. With the predicted heatmap of cancerous regions, they can differentiate different types of lung carcinomas. To predict Gleason grades of prostate cancer, Silva-Rodríguez et al. (2021) proposed a weakly supervised semantic segmentation (WSSS) model to distinguish morphological appearances of different Gleason grades at the local level.

2.3. CAM-based methods

In 2016, Zhou et al. (2016) proposed the class activation map (CAM) to visualize the focus of the classification model. To further expand the idea of the original CAM to more CNN architectures, various CAM have been proposed such as Grad-CAM (Selvaraju et al., 2017a) and Grad-CAM++ (Chattopadhay et al., 2018).

Inspired by CAM, a bunch of CAM-based models were proposed for weakly supervised semantic segmentation. Typical CAM-based methods directly generate segmentation masks from the heat maps with different generation strategies (Wei et al., 2018; Chang et al., 2020). Some researchers take the CAM-activated regions as the initial 'seeds' and gradually grow the regions by leveraging the contextual information (Huang et al., 2018; Sun et al., 2020). Recently, researchers applied saliency maps into CAM-based models to introduce more information and help identify the background ar-

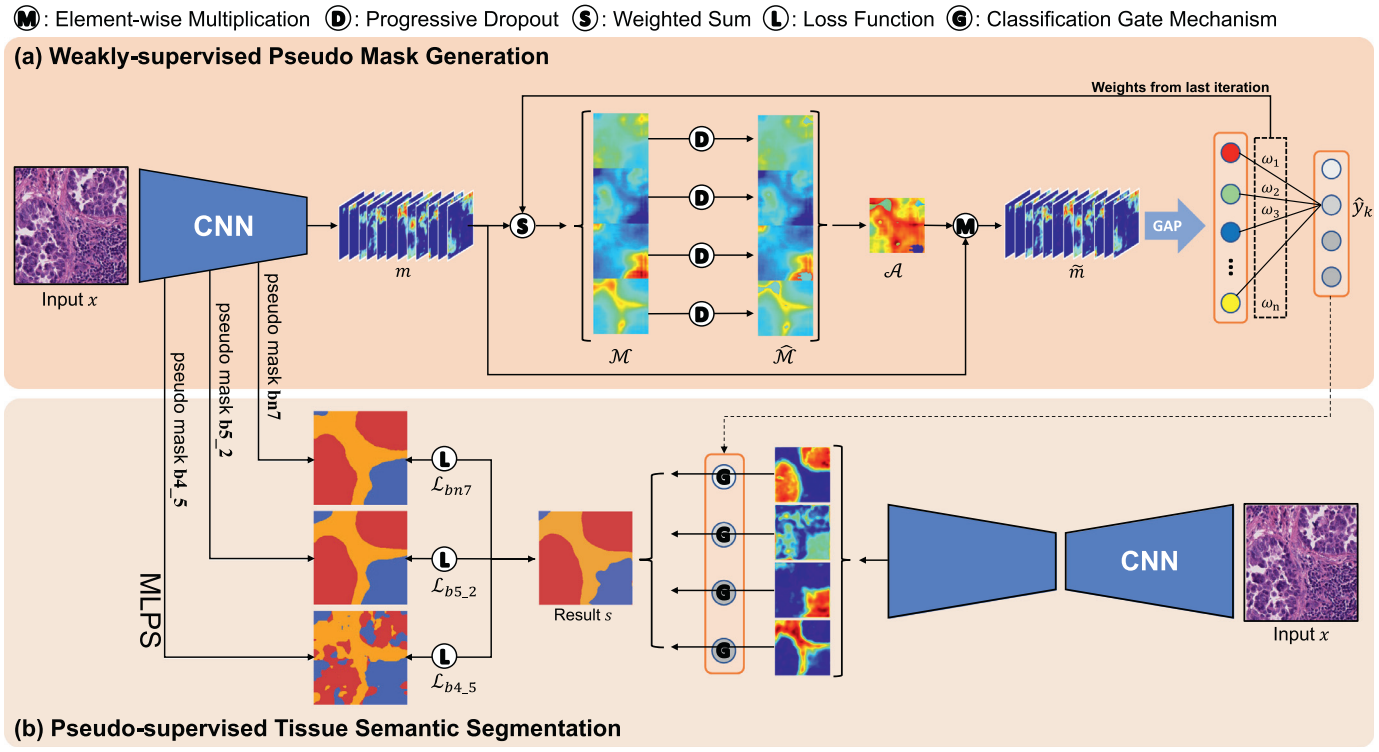

Fig. 2. Weakly supervised tissue semantic segmentation architecture. (a) A weakly-supervised model with Progressive Dropout Attention (PDA) was proposed to generate multi-layer pseudo masks for tissue semantic segmentation. (b) DeepLab V3+ model with a proposed classification gate mechanism was introduced for semantic segmentation, guided by Multi-Layer Pseudo Supervision (MLPS). Theoretically, this model can support semantic segmentation for any number of tissue categories. In this figure, we demonstrate the model using the patch example from our proposed dataset LUAD-HistoSeg with four tissue classes (tumor epithelial, tumor-associated stroma, necrosis and lymphocyte).

eas (Jiang et al., 2021; Lee et al., 2021). Wang et al. (2020b) proposed a self-supervised equivariant attention mechanism to enrich the information of the image-level labels by an affine transform in the siamese network. Kweon et al. (2021) introduced an erasing strategy on the input image to occlude the class-specific regions in order to enrich the information of weak labels.

We noticed that most of the existing CAM-based algorithms were designed for natural images rather than histopathology images. However, natural images have some semantic knowledge that histopathology images do not have. For example, the ground will never be on top of the sky. But in histopathology images, the distribution of different types of tissues is relatively random. Therefore, without such semantic rules, it might be hard to directly apply WSSS methods in natural images to histopathology images. Chan et al. (2019) proposed the only CAM-based method for WSSS for histopathology images. However, it only applied the general Grad-CAM and introduced a series of post-processing steps, which is not precise enough. Hence, in this paper, we aim to design a more effective WSSS method tailored for histopathology images.

3. Methodology

Manual labeling dense pixel-level annotations for histopathology images is extremely difficult and time-consuming. In this paper, we proposed a tissue semantic segmentation model by using only patch-level labels in order to alleviate the annotation efforts. Fig. 2 demonstrates the systematic design of our proposed model. In Section 3.1, we trained a patch-level multi-label classification network with our proposed progressive dropout attention to generate pixel-level pseudo masks. In Section 3.2, we proposed multi-layer pseudo-supervision to train the semantic segmentation model. A classification gate mechanism is proposed to further guide the segmentation results to reduce the false positive rate.

Phase One (classification): Let us denote the given training data in the classification phase as $\mathcal{D}_{cls} = \{(x, y)|x \in \mathcal{X}, y \in \mathcal{Y}\}$, where x is the patch cropped from the whole slide images and y is the multi-label binary vector representing the presence or absence of every tissue category in x. $\hat{y} = f_{cls}(x, \phi_{cls})$ is the multi-label predicted classification result by the classification model f_{cls} with parameters ϕ_{cls}. The objective of this phase is to use only the patch-level label y to generate the dense pixel-level pseudo mask p.

$$f_{cls}(x, y, \phi_{cls}) \rightarrow p \tag{1}$$

Phase Two (segmentation): With the pseudo masks generated in the classification phase, we can form a new training data for the segmentation model as $\mathcal{D}_{seg} = \{(x, p)|x \in \mathcal{X}, p \in \mathcal{P}\}$, where $\mathcal{P}$ is the set of pseudo masks for $\mathcal{X}$. The segmentation model f_{seg} with parameters ϕ_{seg} generates the final semantic segmentation results s.

$$f_{seg}(x, p, \phi_{seg}) \rightarrow s \tag{2}$$

3.1. Weakly-supervised pseudo mask generation

For a giga-pixel whole slide image, defining the presence or absence of the tissue classes in a patch is obviously much easier than carefully drawing pixel-level annotations. So we aim to explore whether patch-level annotations with very limited information are enough for pixel-level semantic segmentation. Zhou et al. (2016) have demonstrated that classification, localization, detection and segmentation tasks share a similar goal. When training a classification model, the feature maps (Class Activation Maps, CAM) deliver the discriminative object location clues which

can be used for object localization and segmentation. Inspired by this, we proposed a novel CAM-based model by first train a classification model. Since the distribution of tissues are somehow random and scatter, it might contain more than one tissue type in one patch. So we define tissue classification as a multi-label classification problem.

3.1.1. Pseudo mask generation

As demonstrated in Fig. 2(a), given an input patch x, we first extract the deep feature maps as follows:

$$f_{cls}(x, \phi_{cls}) \rightarrow m \tag{3}$$

where m denotes the extracted feature maps from the last layer.

In order to provide richer and more comprehensive feature representation, we proposed Progressive Dropout Attention (described in Section 3.1.2) to prevent the classification model from excessively focusing on the most discriminative region.

$$\tilde{m} = \mathcal{A}m \tag{4}$$

where $\mathcal{A}$ is the dropout attention map.

After progressive dropout attention, the probability of the kth tissue class $\hat{y}_k$ can be calculated by a global average pooling and a fully connected layer, the same with Zhou et al. (2016).

$$\hat{y}_k = \sum \omega_k \mathrm{GAP}(\tilde{m}) \tag{5}$$

where $\mathrm{GAP}(\cdot)$ denotes the global average pooling. Multi-label soft margin loss is applied in the classification network.

With a well-trained multi-label classification model, pixel-level pseudo masks p were generated by Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017b) for the next segmentation model.

$$p = \mathrm{Grad-CAM}(f_{cls}(x, \phi_{cls})) \tag{6}$$

3.1.2. Progressive dropout attention

Although the classification model can provide spatial location hints for the segmentation task. But the goals of these two tasks are still different. As the training process goes further, common classification models tend to focus on the most discriminative part/region of the image, while ignoring some insignificant areas. The activated region shrinkage problem will harm the segmentation task. And it will be amplified in the histopathology image of cancers because the spatial arrangement of different tissue types is relatively random comparing with natural images. Moreover, multi-label binary vectors only contain very limited information. There is still a huge information gap from patch-level labels to pixel-level labels. Therefore, how to maximize the value of such sparse annotations in order to close the gap is still an extreme task. To overcome the above two challenges, we proposed a Progressive Dropout Attention (PDA). Let us start with its basic form, Dropout Attention.

Dropout Attention: The idea of the proposed dropout attention is simple and intuitive. We want the neural network to be able to learn as much information as possible from the sparse labels. During the training process, the classification model is not allowed to "make easy money" by only relying on the most discriminative areas. On the contrary, the CNN model has to learn more complete and comprehensive spatial information. Therefore, we deactivate the most significant regions in the class activation maps of all the tissue categories, as demonstrated in Fig. 2(a). Such a strategy will weaken the contribution of the most discriminative regions and force the neural network to perform multi-label classification by non-predominant regions, which can effectively expand the activated regions when extracting deep features. According to this idea, we first generate a class activation map (CAM) for each category by the weighted sum of the feature maps m.

$$\mathcal{M}_k = \sum \omega_k m \tag{7}$$

where $\mathcal{M}_k$ denotes CAM of the kth category.

For each $\mathcal{M}_k$, we set up a dropout cutoff β to deactivate the most highlighted area and refresh CAMs as follows.

$$\hat{\mathcal{M}}_k(i,j) = \begin{cases} \mathcal{M}_k(i,j), & \mathcal{M}_k(i,j) \leq \beta \\ 0, & \mathcal{M}_k(i,j) > \beta \end{cases} \tag{8}$$

where i and j denote the coordinates, $\hat{\mathcal{M}}$ is the CAM with dropout. Note that, β is a relative value which depends on the maximum value of the class activation map.

$$\beta = \mu * \max(\mathcal{M}_k) \tag{9}$$

where μ is the dropout coefficient.

Finally, the dropout attention map $\mathcal{A}$ is the average of all the deactivated CAM.

$$\mathcal{A}(i,j) = \frac{1}{c}\sum_{k=1}^{c} \mathcal{M}_k(i,j) \tag{10}$$

Progressive Dropout Attention: As we mentioned above, when the training process goes further, the activated area will progressively shrink into a smaller area. According to this observation, we proposed a reverse operation based on dropout attention, called Progressive Dropout Attention (PDA). PDA progressively enlarges the deactivated areas to fight against such a shrinking problem. We redesign the original dropout coefficient μ to a progressive dropout coefficient, which is no longer a constant value. The progressive dropout coefficient μ will adaptively decrease when the training epoch increases until μ meets the lower bound l.

$$\mu_t = \begin{cases} \sigma * \mu_{t-1}, & \mu_t > l \\ l, & \mu_t \leq l \end{cases} \tag{11}$$

where t is the ongoing epoch and σ is the decay rate. We set $\sigma = 0.985$ and $l = 0.65$ in practice. The initial μ is set to 1 at the first three epochs for a better initiation of the classification model. After the 3th epoch, we start the dropout and progressively enlarge the dropout area to gradually increase the difficulty of classification.

With progressive dropout attention, the discriminative region shrinkage problem is greatly alleviated and the classification model can learn much richer and wider feature representation, as demonstrated in Fig. 3. The deactivated areas (pointed by black arrows) enlarge with the increasing training epochs, which pushes the model to learn useful information from the surrounding areas, leading to more complete pseudo masks.

3.2. Pseudo-supervised tissue semantic segmentation

In the segmentation phase, we train a semantic segmentation model f_{seg} under the supervision of the pseudo masks p, to get the semantic segmentation result s for the input patch x.

$$s = f_{seg}(x, p, \phi_{seg}) \tag{12}$$

Two specific designs, multi-layer pseudo-supervision and classification gate mechanism, were proposed to further improve the semantic segmentation performance in this phase.

3.2.1. Multi-Layer pseudo-supervision

Due to the information gap between patch-level labels and pixel-level labels, the spatial information learned from the classification network is still incomplete even with progressive dropout attention. To reduce the gap, we have to bring more information to the segmentation model. Since CNN models learn different levels of semantic features at different stages, we generate multi-layer

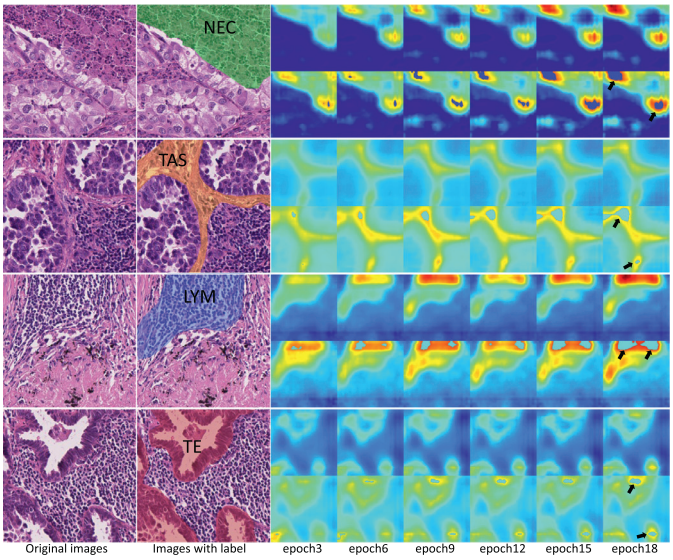

Fig. 3. Examples of progressive dropout attention. We demonstrates the class activation maps $\hat{\mathcal{M}}$ of each tissue category with and without progressive dropout attention (upper: without PDA, lower: with PDA) in different training epochs. For each sample, pathologists were invited to draw the labels for the corresponding category, overlaid on the original images. The deactivated (pointed by black arrows) areas enlarge in the CAMs with PDA when the training epochs increase. From top to bottom are necrosis (NEC), tumor-associated stroma (TAS), lymphocyte (LYM) and tumor epithelial (TE). (Examples are selected from LUAD-HistoSeg).

pseudo masks from three different layers to enrich the information. And then we calculate cross entropy loss between the semantic segmentation results and all the pseudo masks.

$$\mathcal{L}_{seg} = \lambda_1 \mathcal{L}_{b4\_5} + \lambda_2 \mathcal{L}_{b5\_2} + \lambda_3 \mathcal{L}_{bn7} \tag{13}$$

where λ_i is the hyper-parameter. We set $\lambda_1 = 0.2$, $\lambda_2 = 0.2$, $\lambda_3 = 0.6$ in practice. Note that, multi-layer pseudo masks were upsampled to the original image resolution using bilinear interpolation.

3.2.2. Classification gate mechanism

Long tail problem is common for medical data, especially for histopathology images. For those non-predominant tissue categories, like necrosis and lymphocyte, they will be dominated by the predominant tissue categories. It is easier to generate unsatisfactory pseudo masks for the non-predominant categories than the predominant categories, which may increase the false positive rate in the segmentation phase.

To overcome the long tail problem and to reduce the false positive rate for the non-predominant categories, we proposed a classification gate mechanism. In our proposed framework, we observed that the confidence of the classification results is generally higher than the segmentation results on the question of whether a tissue category exists in a patch image, especially for the non-predominant categories. Because the classification model was trained by ground truth labels while the segmentation model was trained by pseudo masks.

Based on this observation, we introduce a gate for each output channel. Let o_k denote the output probability map of the kth tissue category from the segmentation model. For each category k, if the predicted probability $\hat{y}_k$ of the tissue category from the classification model is smaller than a threshold ϵ, it means a low existence rate of this category. Then we will "close the gate" of the probability map o_k by zeroing it.

$$o_k = \begin{cases} 0 * o_k, & \hat{y}_k \leq \epsilon \\ o_k, & \hat{y}_k > \epsilon \end{cases} \tag{14}$$

Then the semantic segmentation result can be obtained by an argmax operation of the probability map o. We set $\epsilon = 0.1$ in practice.

$$s(i, j) = \text{argmax} o(i, j) \tag{15}$$

where (i, j) denotes the coordination.

3.2.3. Semantic segmentation for WSIs

The model we defined above is the patch-level semantic segmentation model. Next, we introduce the way we achieve semantic segmentation for the whole slide images. As demonstrated in Fig. 4, we first cropped patches from a whole slide image with over 50% overlapping region. With the segmentation model, n channels probability maps can be generated for each patch. Then we stitched the probability maps to the WSI-level. For the overlapping regions, we calculated mean of the probabilities of each category at every pixel location. Then we can obtain the semantic segmentation result of the whole slide image by an argmax operation.

3.3. Implementation and training details

In our experiments, all the convolutional neural networks were implemented in PyTorch. The model was trained on an NVIDIA RTX 2080Ti. ResNet38 (Wu et al., 2019) and DeepLab V3+ (Chen et al., 2018) were introduced as the classification and segmentation backbones respectively. In the classification phase, the model was pre-trained on ILSVRC 2012 classification dataset (Ahn and Kwak, 2018). The resolution of the patches is 224×224 and the batch size is set to 20. The number of training epochs is set to 20, 40 for LUAD-HistoSeg and BCSS-WSSS datasets, respectively. All the patches were transformed by random horizontal and vertical flip with the probability 0.5. We set a learning rate of $1e-2$ with a polynomial decay policy. In the segmentation phase, the number of training epochs and the learning rate for both datasets were set 20 and $7e-2$, respectively. There is no restriction of the image resolution in the segmentation phase. Several data augmentation methods were applied, including horizontal and vertical flip, Gaussian blur and normalization.

4. Datasets

We evaluate our proposed model on two tissue semantic segmentation datasets, LUAD-HistoSeg and BCSS-WSSS. LUAD-HistoSeg is the dataset we created for lung adenocarcinoma. BCSS-WSSS is our synthesized WSSS version of BCSS dataset (Amgad et al., 2019), a fully supervised semantic segmentation dataset for breast cancer. Both datasets can be accessed via the Github link in the abstract.

4.1. LUAD-HistoSeg dataset

As a part of this paper, we release a weakly-supervised tissue semantic segmentation dataset for lung adenocarcinoma, named LUAD-HistoSeg, demonstrated in Fig. 5. This dataset aims to use only patch-level annotations to achieve pixel-level semantic segmentation for four tissue categories, tumor epithelial (TE), tumor-associated stroma (TAS), necrosis (NEC) and lymphocyte (LYM).

Dataset Description: In this dataset, 54 patients from the Department of Pathology, Guangdong Provincial People's Hospital with lung adenocarcinoma were chosen. For each patient, three experienced pathologists (at least ten-year working experience) were asked to examine all the pathology sections and select the most representative section for clinical diagnosis. The selected WSIs were randomly split into two sets, a training set with 29 WSIs, and validation and test sets with 25 WSIs. Note that, the validation set and the test set share the same WSIs. For each WSI, we

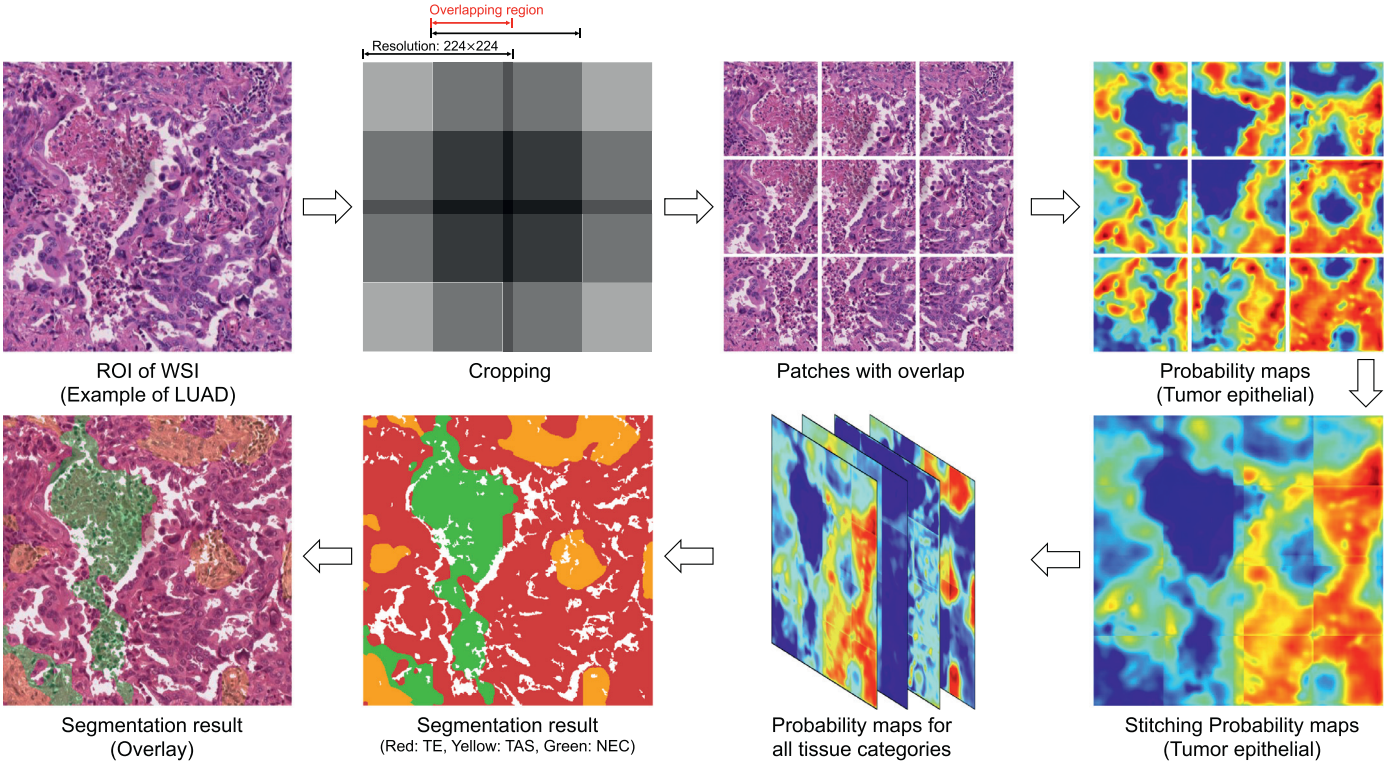

Fig. 4. Semantic segmentation for whole slide images. We only show a very small view of the WSI for simple illustration. We show the probability maps of tumor epithelial as the examples.

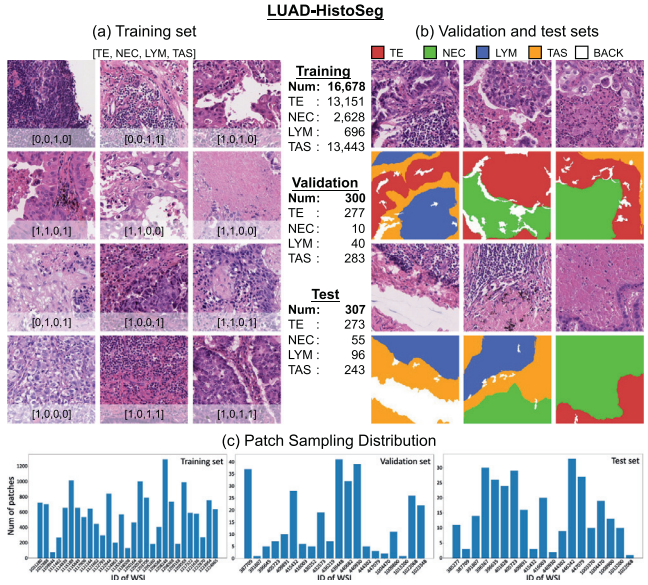

Fig. 5. Examples of the released dataset LUAD-HistoSeg. (a) shows the training set with multi-label binary vectors. (b) demonstrates the validation and test sets with semantic segmentation masks. We define four tissue categories in one tissue patch, including tumor epithelial (TE), tumor-associated stroma (TAS), necrosis (NEC) and lymphocyte (LYM). 'Num' means the number of patches. (c) demonstrates the number of patches sampled from each whole slide image. There is no WSI overlap between the training set and the validation & test sets.

first labeled the tumor bulk and then randomly sampled patches (800 patches per WSI, 224 × 224, 10X magnification) inside the tumor bulk. Next, we dropped the patches with blurry, dirty, large white backgrounds and over-stained problems by a quality control process. We further dropped the ambiguous patches which have classification disagreement among three pathologists. Then we randomly sampled patches from each set, which formed a training set (16,678 patches with patch-level annotations), a validation set (300 patches with pixel-level annotations) and a test set (307 patches with pixel-level annotations). The number of patches sampled from each WSI is shown in Fig. 5(c).

How to Label: We invited five junior clinicians and three experienced pathologists to label all the patches. There are two different kinds of labels, patch-level labels for the training set and pixel-level labels for the validation and test sets. For the training set, annotators have to define whether a specific tissue category is present or absent by a multi-label binary vector, demonstrated in Fig. 5(a). For the validation and test sets, annotators were asked to roughly draw the semantic segmentation masks using Labelme (Wada, 2016) and refine the boundaries using PhotoShop, demonstrated in Fig. 5(b). Junior clinicians were responsible for labeling and pathologists have to finally confirm the labels. The patches were rejected and dropped if there exists ambiguities. Since the lung is mainly composed of the alveolus, there are a lot of white regions randomly distributed in the whole slide image. So we extract these white regions by a color thresholding method. The white backgrounds inside the alveolus were excluded when calculating the performance in all the experiments.

4.2. BCSS-WSSS Dataset

We also evaluate our proposed model on a fully-supervised semantic segmentation dataset, to compare our weakly-supervised approach with the fully-supervised approach in order to observe the potential of our proposed model.

Breast cancer semantic segmentation (BCSS) dataset (Amgad et al., 2019) consists of 151 representative regions of interest (ROIs) from 151 H&E stained whole slide images of breast cancer, which were selected by a study coordinator, a

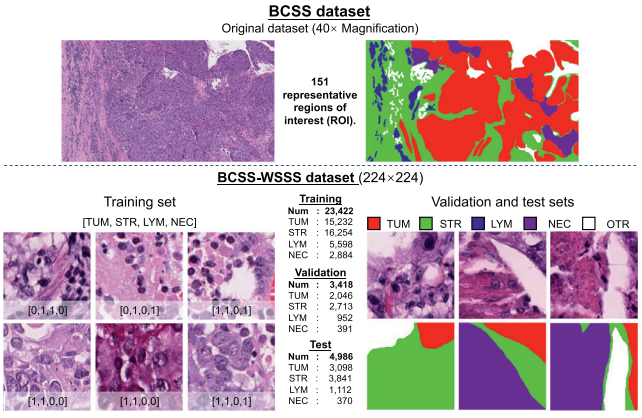

Fig. 6. Examples of BCSS-WSSS dataset. The original BCSS dataset contains 151 large ROIs with pixel-level annotations of five tissue categories, Tumor (TUM), Stroma (STR), Lymphocytic infiltrate (LYM), Necrosis (NEC) and Other (OTR). We generate a synthetic dataset, called BCSS-WSSS, for the weakly-supervised approach. The left-hand side shows the training set with multi-label binary vectors. The right-hand side demonstrates the validation and test sets with semantic segmentation masks. 'Num' means the number of patches.

clinician, and approved by a senior pathologist. The mean size of ROIs is 1.18 mm^2 at 0.25 microns per pixel resolution. As shown in Fig. 6, the original BCSS dataset provides pixel-level annotations for each ROI with 5 classes, including Tumor (TUM), Stroma (STR), Lymphocytic infiltrate (LYM), Necrosis (NEC) and Other (OTR).

In order to perform weakly-supervised semantic segmentation, we randomly cropped patches from the ROIs and used the semantic segmentation masks to generate multi-label binary vector encoding vectors. A total of 31,826 patches were generated and split into a training set (23,422 patches, patch-level annotations), a validation set (3,418 patches, pixel-level annotations), and a test set (4,986 patches, pixel-level annotations), as demonstrated in Fig. 6. We also provide the generated patch-level dataset of BCSS-WSSS via the Github link in the abstract.

5. Experiments

In this section, we conduct several experiments to comprehensively evaluate the capacity of our proposed model on how well it achieves semantic segmentation using only patch-level annotations. Section 5.1 demonstrates the quantitative and qualitative comparisons with state-of-the-art methods. We conduct ablation studies in Section 5.2 to evaluate the effectiveness of our proposed progressive dropout attention, multi-layer pseudo-supervision and classification gate mechanism. Next, we demonstrate the semantic segmentation results of the whole slide images in Section 5.3. We also measure how much labeling time we can save for the pathologists in Section 5.4. We discuss the limitations of the proposed model in Section 5.5.

We evaluate our proposed model by the following metrics, IoU for each category, Mean IoU (MIoU), Frequency weighted IoU (FwIoU) and pixel-level accuracy (ACC).

5.1. Quantitative and qualitative comparisons

Table 1 demonstrates the quantitative comparisons with existing methods. We compare our proposed model with five SOTA CAM-based weakly-supervised semantic segmentation models, one for histopathology images (HistoSegNet Chan et al., 2019) and the other four for natural images (SC-CAM Chang et al., 2020), Grad-CAM++(Chattopadhay et al., 2018), CGNet (Kweon et al., 2021) and OAA (Jiang et al., 2021). (1) HistoSegNet directly applied Grad-CAM with a series of post-processing methods. (2) SC-CAM clusters the

Table 1
Quantitative comparison with existing methods.

Method	LUAD-HistoSeg							BCSS-WSSS						
	TE	NEC	LYM	TAS	FwIoU	MIoU	ACC	TUM	STR	LYM	NEC	FwIoU	MIoU	ACC
HistoSegNet	0.45594	0.36302	0.58283	0.50818	0.48538	0.47749	0.65971	0.33141	0.46457	0.29047	0.01908	0.37191	0.27638	0.56410
SC-CAM	0.68286	0.64284	0.62063	0.61785	0.64743	0.64104	0.78690	0.76788	0.70606	0.58023	0.60073	0.71581	0.66373	0.83427
OAA	0.69557	0.53555	0.67181	0.62905	0.65578	0.63300	0.79251	0.75132	0.68883	**0.61230**	0.60600	0.70469	0.66461	0.82552
Grad-CAM+	0.72897	0.74175	0.67933	0.66018	0.69776	0.70256	0.81967	0.66737	0.62064	0.50077	0.48053	0.62308	0.56733	0.76530
CGNet	0.71853	0.73296	0.69092	0.67262	0.69887	0.70376	0.82219	0.68215	0.61769	0.52240	0.56836	0.63390	0.59765	0.77626
Ours Phase 1	0.75567	0.78079	**0.73694**	0.69690	0.73324	0.74258	0.84508	0.72976	0.68134	0.56191	0.55989	0.68532	0.63323	0.81216
Ours Phase 2	**0.77704**	**0.79321**	0.73406	**0.71980**	**0.75126**	**0.75603**	**0.85701**	**0.78839**	**0.73157**	0.57295	**0.66389**	**0.73745**	**0.68920**	**0.84832**

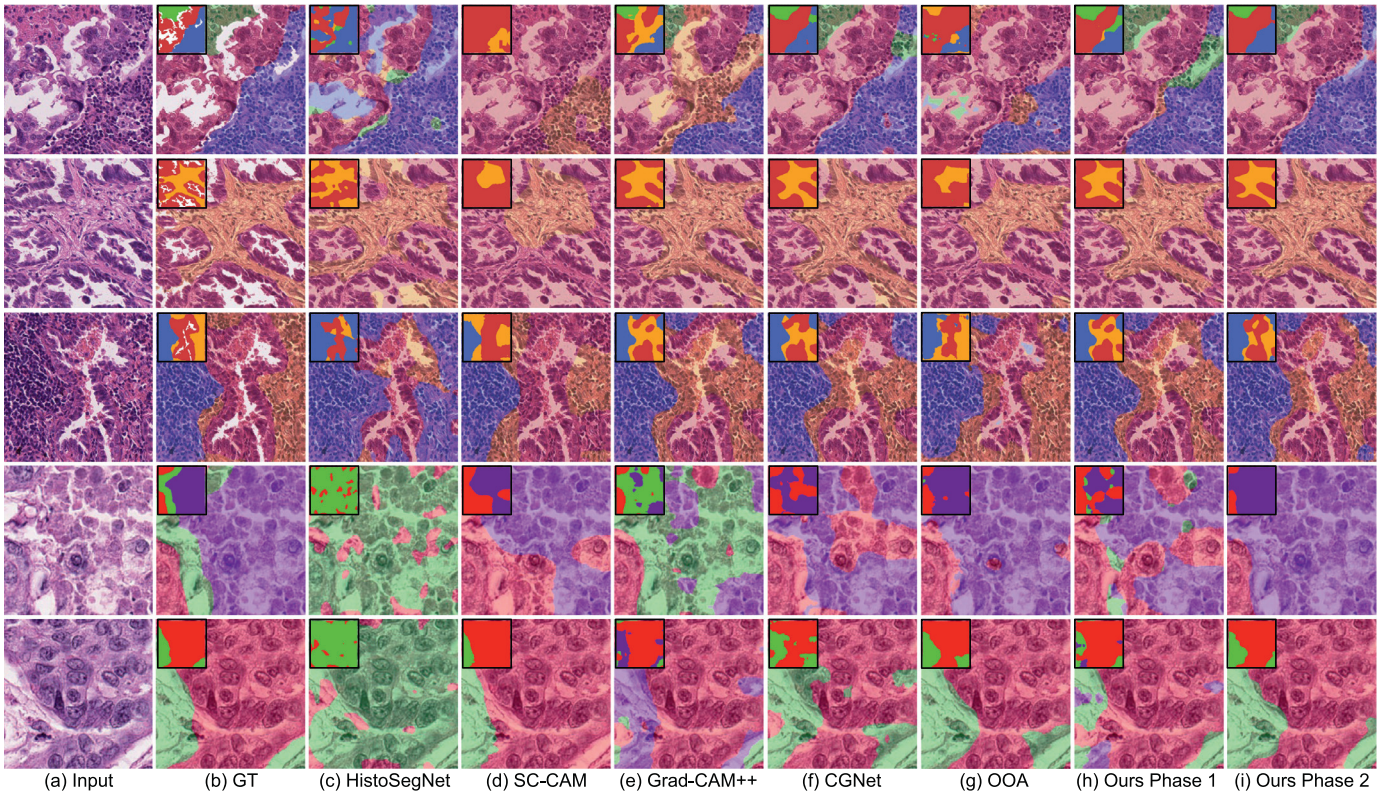

Fig. 7. Qualitative results of patch-level semantic segmentation. Results were overlaid on the input images. The upper three rows are LUAD-HistoSeg. The bottom two rows are BCSS-WSSS. In the top-left corner are the semantic segmentation masks.

image features into several sub-categories. (3) Grad-CAM++ is an extension of Grad-CAM. (4) CGNet introduced an erasing strategy on the input image to occlude the class-specific regions. (5) OAA is a saliency-based method with online attention accumulation. We implemented all the existing methods exactly following the technical details of the original papers or directly run it if the source code is available. "Ours Phase 1" is the classification model trained in phase one. The semantic segmentation results of this model were generated by Grad-CAM from layer *bn*7. "Ours Phase 2" is the semantic segmentation model trained in phase two, which is our final model.

As shown in Table 1, our final model generally outperforms all the existing models on both datasets. In the LUAD-HistoSeg dataset, even the pseudo masks generated from the classification model in phase 1 outperform the existing CAM-based WSSS methods, which justifies the superiority of our proposed progressive dropout attention. After training the segmentation model in phase 2, our model achieves a significant and consistent improvement in all the categories except LYM in LUAD-HistoSeg. Because LYM only occupies around 4% in this dataset, which is extremely imbalanced. The lack of training samples may lead to unstable performance.

Figure 7 demonstrates the qualitative results of different models in both datasets. Our proposed model can generate more precise semantic segmentation results than the existing works. HistoSegNet (Chan et al., 2019) mostly relies on post-processing steps to merge the fragile segments. Therefore, it fails to predict complete and unbroken results. Since the distribution of different tissues is relatively random and scattered, but natural images follow some rules like 'cars mostly appear on the road'. Most of the specific designs for natural images are not so perfectly suitable for histopathology images, thereby generating results with either broken regions or inaccurate boundaries, especially for the non-predominant categories, LYM and NEC. Our proposed progressive dropout attention will deactivate the most discriminative regions and push the neural network to learn more comprehensive features from the entire image. Such design greatly benefits weakly-supervised semantic segmentation in histopathology images. Qualitative results also show that training a segmentation phase of pseudo-supervision is necessary since it can avoid some noisy prediction results in phase one.

Since the original BCSS-WSSS is the tissue semantic segmentation dataset with pixel-level annotations. Therefore, we conducted an additional experiment to evaluate the potential of our proposed model by comparing the proposed pseudo-supervision with fully-supervision. We generated a WSSS dataset from the original BCSS dataset for our proposed model as demonstrated in Section 4.2. To be fair, both fully-supervised and pseudo-supervised models were trained on the same network structure DeepLab V3+ with the same training epochs. Compared with the results generated by the fully-supervised model, shown in Table 2, our proposed pseudo-supervised model demonstrates competitive performance for all the tissue categories, even for the non-predominant ones. The performance gap between the pseudo-supervised model and the fully-supervised model is less than 2%. Figure 8 further demonstrates the qualitative comparisons between the pseudo-supervised model with the fully-supervised model. The semantic segmentation results generated by the pseudo-supervised model show visually no difference from the ones generated by the fully-supervised model. Both two models can generate high concordance semantic segmentation results compared with manual annotations. Unfortunately, when the borders between two tissue categories are not visually clear enough, both two models fail to generate smooth boundaries. It is still a debate whether a smooth and "accuracy" boundary is really meaningful for clinical cancer research. Overall, this experiment proves that only relying on patch-level annotations can also achieve superior semantic segmentation results

Table 2
Quantitative comparison with fully supervision.

	TUM	STR	LYM	NEC	FwIoU	MIoU	ACC
Ours	0.78839	0.73157	0.57295	**0.66389**	0.73745	**0.68920**	0.84832
Fully	**0.81072**	**0.74861**	**0.58680**	0.59873	**0.75310**	0.68622	**0.85760**

Table 3
Quantitative evaluation: ablation studies. (LUAD-HistoSeg).

Phase	PDA	Pseudo Supervision	Class-Gate	TE	NEC	LYM	TAS	FwIoU	MIoU	ACC
(1) Phase 1	-	-	-	0.72862	0.72690	0.71305	0.68952	0.71190	0.71452	0.83111
(2) Phase 1	DA	-	-	0.75191	0.75568	0.72260	0.69435	0.72691	0.73113	0.84087
(3) Phase 1	✓	-	-	0.75567	0.78079	0.73694	0.69690	0.73324	0.74258	0.84508
(4) Phase 2	✓	b_{4_5}	-	0.69942	0.55688	0.70002	0.68347	0.68295	0.65995	0.80978
(5) Phase 2	✓	b_{5_2}	-	0.75831	0.77700	0.67398	0.67792	0.71859	0.72180	0.83454
(6) Phase 2	✓	$bn7$	-	0.77160	0.74853	0.72714	0.70785	0.74028	0.73878	0.84978
(7) Phase 2	✓	$bn7 + b_{4\_5}$	-	0.77061	0.74187	0.71479	0.70553	0.73685	0.73320	0.84747
(8) Phase 2	✓	$bn7 + b_{5\_2}$	-	0.77526	0.75885	**0.74233**	0.71248	0.74634	0.74723	0.85397
(9) Phase 2	✓	Multi-Layer	-	**0.77704**	0.78374	0.73303	0.71724	0.74947	0.75277	0.85586
(10) Phase 2	✓	Multi-Layer	✓	**0.77704**	**0.79321**	0.73406	**0.71980**	**0.75126**	**0.75603**	**0.85701**

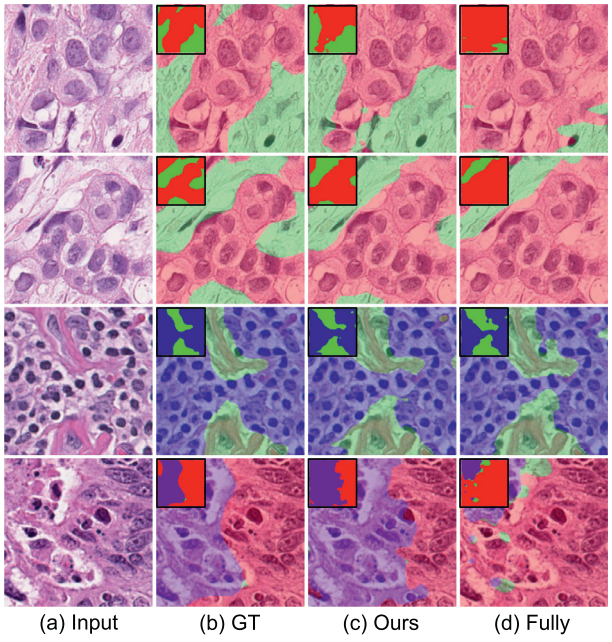

Fig. 8. Comparison with fully-supervision (BCSS). Results were overlaid on the input images. Top-left corner are the semantic segmentation masks.

which is good news for pathologists to reduce the annotation efforts.

5.2. Ablation studies

We conducted a series of ablation studies to quantitatively and qualitatively evaluate the superiority of the novelties, including progressive dropout attention (PDA), multi-layer pseudo-supervision (MLPS) and classification gate mechanism. We compared our final model with several baseline models in LUAD-HistoSeg dataset as follows: (1) Phase 1 alone. (2) Phase 1 with dropout attention (DA) with a constant dropout coefficient $\mu = 0.7$. (3) Phase 1 with progressive dropout attention (PDA). The results of these three models were the pseudo masks p_{bn7} generated by Grad-CAM from layer $bn7$. (4)-(9) Phase 2 trained by different configurations of multi-layer pseudo masks. (10) Our final model. The quantitative results are shown in Table 3. We also selected several representative baseline models (1), (3), (9) and (10) to qualitatively prove the effectiveness of the proposed novelties in Fig. 9.

Table 4
Classification Results with and without PDA (patch-level accuracy).

	LUAD-HistoSeg	BCSS-WSSS
P1 w/o PDA	0.93893	0.90784
P1 w PDA	0.92508	0.90694

5.2.1. Progressive dropout attention

In Table 3, model (2) with DA has already achieved an obvious improvement compared with model (1) in all the tissue categories as well as FwIoU, MIoU and the overall pixel-level accuracy. When equipped with PDA in model (3), the performance continuously improves, especially for the non-predominant categories NEC and LYM. Because deactivating the highlighted areas will push neural networks to learn features from secondary discriminative regions, reducing the information gap between the classification labels and the segmentation labels. But for those non-predominant categories, drastically increasing the difficulty may bring adverse effects. Therefore, progressively increasing the difficulty can smooth the training process, resulting in a better performance improvement. Figure 9(c) & (d) shows the results of model (1) and model (3). In the yellow boxes, we can observe the lymphocyte regions from the model with PDA have higher concordance with ground truth compared with the model without PDA. Although the pseudo masks are still imperfect, we successfully reduce the information gap between image-level labels and pixel-level labels by correcting some false predicted labels.

Besides the improvement of the semantic segmentation performance, we also want to know whether PDA will greatly harm the classification results, which is not our expectation. Table 4 demonstrates the classification results of the classification model after applying PDA. We can find that the overall accuracy only decreases around 1% in LUAD-HistoSeg and less than 0.1% in BCSS-WSSS. We believe that it is worth to trade-off less than 1% classification accuracy for more than 2% semantic segmentation improvement.

5.2.2. Multi-Layer pseudo supervision

Since the information gap between patch-level classification labels and pixel-level segmentation labels is huge. The pseudo masks generated from patch-level annotations are no doubt incomplete and imperfect. It is the reason why we proposed MLPS to provide as much information as possible from different layers of the classification model. To evaluate the effectiveness of MLPS, we compare the proposed MLPS model with the model trained by different con-

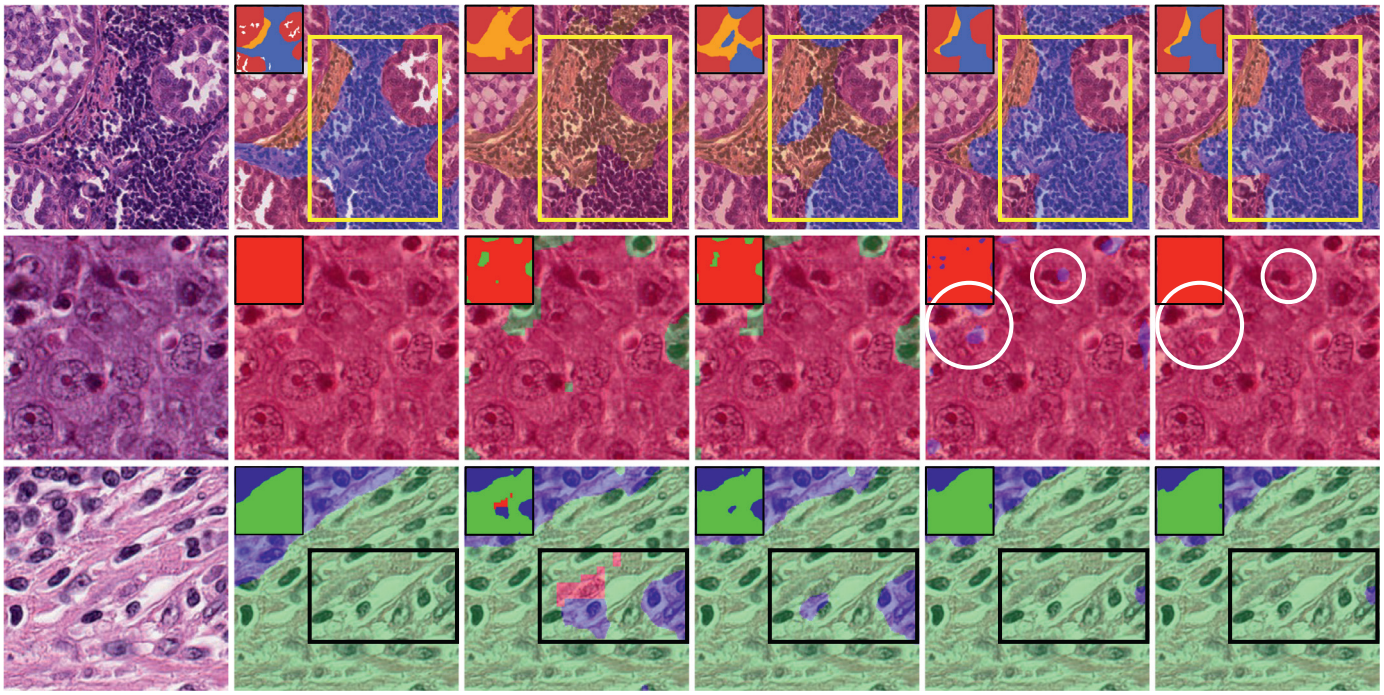

Fig. 9. Qualitative results of ablation studies. The first row is from LUAD-HistoSeg. The next two rows are from BCSS-WSSS. We directly overlaid the results onto the original images. Top-left corner are the semantic segmentation masks.

figurations of the pseudo masks, demonstrated in Table 3. Among the models trained by the pseudo masks from a single layer (4)–(6), model (6) trained by *bn*7 shows the best performance because it is the closest layer to the inference with the finest semantic information. The models trained by the pseudo masks from two different layers (7) and (8) show better quantitative results than all the single-layer models. The model trained by all three layers (9) outperforms the other baseline models. For the long tail classes with fewer training samples, like LYM, noises introduced by the shallow layer b_{4_5} might bring negative effects to the models. Nevertheless, multi-layer pseudo supervision with three layers reduces the information gap between patch-level and pixel-level annotations and achieves the best semantic segmentation performance. Figure 9 demonstrates the results with and without MLPS. In the yellow and black boxes, results generated by model (7) with MLPS are more complete and achieve higher concordance with ground truth. Experimental results prove that introducing multi-layer pseudo masks can provide more information than the pseudo mark from a single layer. And the incorrect noisy pseudo labels can also be regarded as the regularization method to avoid overfitting.

5.2.3. Classification gate mechanism

The classification gate mechanism is proposed to reduce the false-positive rate for the non-predominant tissue categories. Model (8) and Model (7) in Table 3 demonstrate the models with and without classification gate mechanism in LUAD-HistoSeg, respectively. For the predominant categories, tumor epithelial (TE) and tumor-associated stroma (TAS), classification gate mechanism gets a very slight improvement because the predominant ones occupy more than 60% of the samples. The segmentation model can learn a better feature representation of them, which results in a lower false-positive rate. For the non-predominant one necrosis (NEC) in LUAD-HistoSeg, classification gate mechanism improves the IoU by more than 1%. Figure 9(f) & (e) demonstrates the results with and without the gate. In the white circles, false-positive results have been successfully corrected by the classification gate mechanism.

5.3. Qualitative results of WSIs

In Fig. 10, we also demonstrate the semantic segmentation results of two whole slide images with lung adenocarcinoma and breast cancer, respectively. The way we generate WSI-level semantic segmentation is shown in Section 3.2.3. Since BCSS-WSSS was originally introduced for fully-supervised semantic segmentation, we can compare our results with manual annotations. In both lung adenocarcinoma and breast cancer WSIs, our proposed model can generate visually pleasing results. We can find that the predominant categories such as tumor epithelial and tumor-associated stroma have high concordance compared with the ground truth labels, while the non-predominant categories necrosis and lymphocyte have lower concordance but are still visually pleasing.

When zooming in the whole slide images (highlighted in black and blue boxes), some “imperfect” results can be found such as the unsmooth region boundaries and some very small isolated regions. The reason why we double quote “imperfect” is that it is hard to decide whether such results are inaccurate or not. For example, the ROI in the yellow circle, there are some small stroma regions inside the lymphocytic infiltrate region. Globally speaking, they should be categorized as the lymphocytic infiltrate regions but they have the same morphological appearance with stroma. Furthermore, the borders between different tissue types are commonly ambiguous, especially for the tumor invasive regions. It is still a debate whether a smooth and “accuracy” boundary is really meaningful for clinical cancer research.

5.4. How can we reduce annotation efforts?

We also conducted an experiment to quantitatively evaluate the reduced efforts of manual annotation by applying our proposed

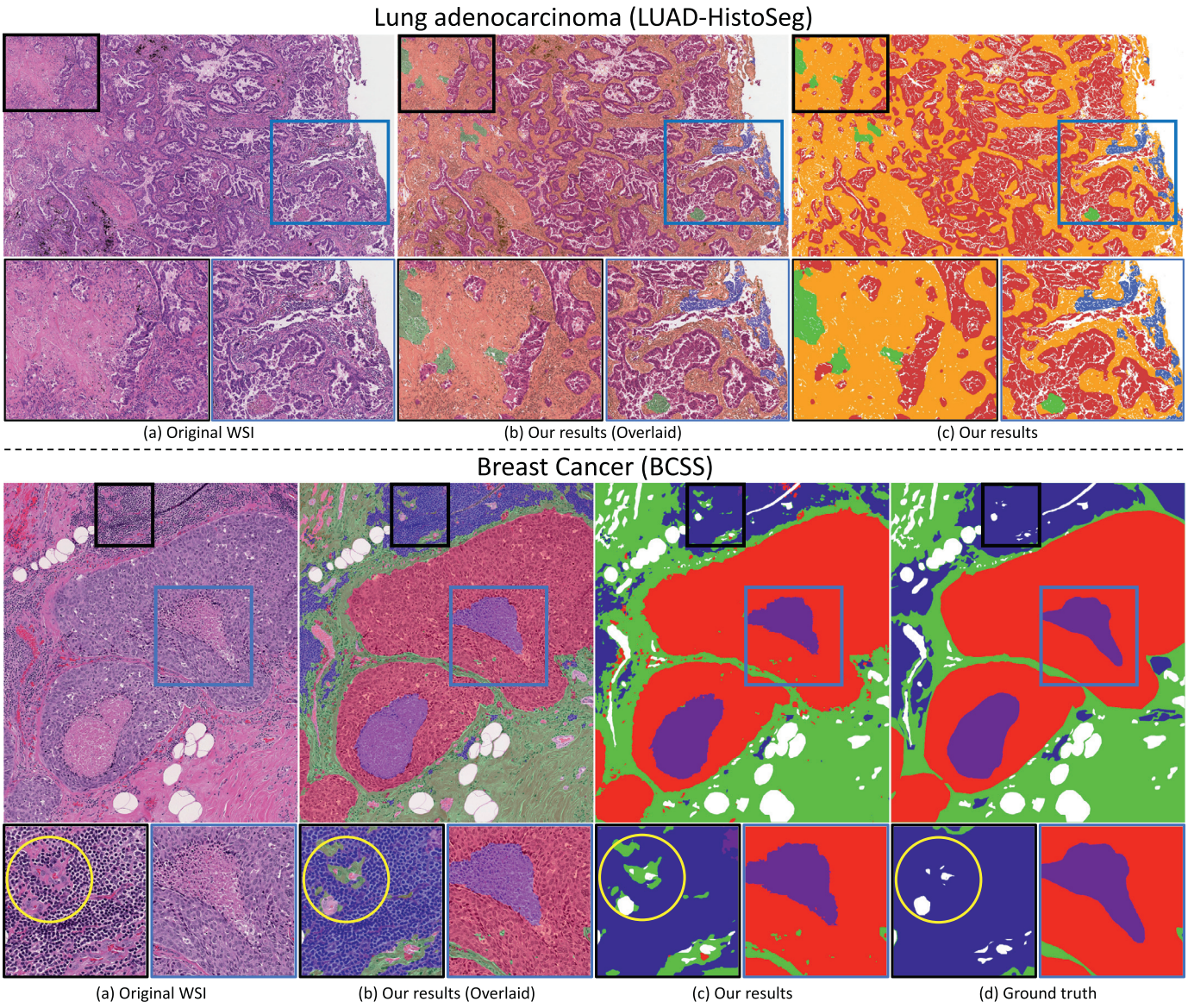

Fig. 10. Semantic segmentation results of the whole slide image in lung adenocarcinoma and breast cancer. We show two zoom in regions in the black and blue boxes. Yellow circle shows the ambiguous region. Since the resolution of the WSI is huge, we only demonstrated a small part it.

Table 5
Comparisons of the timing statistics for patch-level annotations and pixel-level annotations.

Pathologists	Patch-level (minutes)	Pixel-level (minutes)
1	6.8	177.5
2	5.2	209.3
3	7	231.6

model. We randomly selected 100 patches (224 × 224) from the LUAD-HistoSeg dataset. Three junior pathologists were invited to join this test. Pathologists were first asked to label patch-level annotations by our developed tiny tool. For each category, there are two buttons, '✓' and '×', to decide whether a tissue category is present or absent. Next, pathologists were asked to use Labelme (Wada, 2016) to draw pixel-level annotations. There is no doubt that answering "Yes or No" questions is more efficient as shown in Table 5. All three pathologists only spent less than 10 min to finish 100 patch-level annotations while average around 200 min for pixel-level annotations. We also observe that pathologists often struggled and spent much time refining the boundaries when doing pixel-level annotations, while patch-level annotations can avoid this. Besides time efficiency, patch-level annotations also have higher consistency than pixel-level annotations. We measure the consensus score by dividing the sum of agreeing labels by the total number of labels. The consensus scores of patch-level and pixel-level annotations are 92.25% and 85.64%.

5.5. Limitations

There are still some limitations of our proposed model. It has achieved outstanding performance for the predominant tissue categories. But for the non-predominant ones, lack of enough training samples is always the greatest barrier towards precise segmentation results. Collecting more training samples for these categories may alleviate this problem. Second, as discussed in Section 5.3 (Yellow circle in Fig. 10, our model recognize some small stroma regions inside the lymphocytic infiltrate region. Because the model only considers the morphological features within the receptive field, which may introduce a lot of isolated regions

Table 6
Further thinking and exploration on the data augmentation.

LUAD-HistoSeg									
Phase	Cutout	PDA	TE	NEC	LYM	TAS	FwIoU	MIoU	ACC
(1) Phase 1	-	-	0.72862	0.72690	0.71305	0.68952	0.71190	0.71452	0.83111
(2) Phase 1	✓	-	0.73994	0.73287	0.73273	0.69881	0.72319	0.72609	0.83888
(3) Phase 1	-	✓	0.75567	**0.78079**	0.73694	0.69690	0.73324	0.74258	0.84508
(4) Phase 1	✓	✓	**0.75885**	0.75357	**0.75443**	**0.70510**	**0.73794**	**0.74304**	**0.84862**
BCSS									
Phase	Cutout	PDA	TE	NEC	LYM	STR	FwIoU	MIoU	ACC
(1) Phase 1	-	-	0.72387	0.67170	0.56124	0.57892	0.67934	0.63392	0.80808
(2) Phase 1	✓	-	0.72531	0.67474	0.55350	**0.61505**	0.68198	0.62215	0.80954
(3) Phase 1	-	✓	0.72976	0.68134	**0.56191**	0.55989	0.68532	0.63323	0.81216
(4) Phase 1	✓	✓	**0.73008**	**0.68232**	0.56019	0.59475	**0.68713**	**0.64184**	**0.81334**

inside a large region. Actually in clinical practice, pathologists define a tissue category by not only observing the morphological appearances locally but also considering a large surrounding area of the microenvironment globally. Introducing a global-local design may be a solution to solve this problem and we will keep on discovering it in future works.

6. Conclusion

In this paper, we proposed a tissue-level semantic segmentation model for cancer histopathology images. The major contribution of this model is to replace pixel-level annotations with patch-level annotations, which is significant progress for pathologists to reduce their annotation efforts. Our proposed model achieves competitive performance with the fully-supervised model, which means that pathologists only need to define the presence or absence of the tissue categories in a patch instead of carefully drawing the labels. In methodology, we proposed several technical novelties to minimize the information gap between patch-level and pixel-level annotations, and achieved outstanding semantic segmentation performance. Based on the proposed approach, there are still a lot of worthy technical directions to be explored, such as (1) applying some outstanding data augmentation techniques to further improve the model robustness. (2) Associating classification and segmentation by a multi-task learning strategy to learn more representative and correlated features for both tasks. (3) Introducing interactions of the feature maps between the two tasks to enhance both tasks with information exchange. (4) Designing cross-task attention to improve each other. (5) Considering noise-correction mechanism for the pseudo masks.

In Table 6, we explore one of them by applying a data augmentation technique Cutout (DeVries and Taylor, 2017), which randomly blinds an area of the input image in each iteration. In this exploration, we directly compare Cutout with PDA by the pseudo masks generated from the classification in the phase 1. There are four models, (1) baseline alone, (2) baseline with Cutout, (3) baseline with PDA, (4) baseline with Cutout and PDA. We first compare Cutout with PDA. Quantitative results show that Cutout can slightly improve the precision of the pseudo masks generated by the CAM, but it is still less effective than PDA (in both datasets). It proves that adaptively deactivating the most discriminative areas in the feature maps is more effective than randomly blinding some areas in the input images. Considering these two techniques have no conflict, we combine them together in model (4). We can observe that associating Cutout with PDA achieves the best quantitative results among four models. Since PDA is designed to learn more comprehensive feature representations and Cutout aims to improve the model robustness. Combining these two techniques can further improve the generated pseudo masks.

To contribute to the research fields of computational pathology and cancer research, we also introduce a new weakly-supervised semantic segmentation dataset for lung adenocarcinoma, LUAD-HistoSeg. This is the first tissue-level semantic segmentation dataset for lung cancer. By applying our proposed model, we also keep on generating more tissue-level semantic segmentation datasets for different cancer types. More senior pathologists will be invited to join this project for labels verification. Hopefully, these datasets will be released soon.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project is supported by the Key-Area Research and Development Program of Guangdong Province (No. 2021B0101420006), the National Science Fund for Distinguished Young Scholars (No. 81925023), the National Natural Science Foundation of China (No. 82072090, 81771912 and 82071892), the National Science Foundation for Young Scientists of China (No. 62102103, 62002082 and 82102034), Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (No. 2022B1212010011), The Guangdong Natural Science Foundation (No. 2017A030312008), Guangdong Provincial Key Laboratory of Cyber-Physical Systems (No. 2020B1212060069) and National&Local Joint Engineering Research Center of Intelligent Manufacturing Cyber-Physical Systems. We would like to thank all the labelers and pathologists for their efforts on the annotations (Bingbing Li, Yuan Zhang, Huihui Wang, Chao Zhu, Yuchen Song, Huasheng Yao, Yumeng Wang, Ke Zhao, Mengyi Dong, Jia Huang, Yingyi Wang).

References

AbdulJabbar, K., Raza, S.E.A., Rosenthal, R., Jamal-Hanjani, M., Veeriah, S., Akarca, A., Lund, T., Moore, D.A., Salgado, R., Al Bakir, M., et al., 2020. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. Nat. Med. 26 (7), 1054–1062.

Ahn, J., Kwak, S., 2018. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Amgad, M., Elfandy, H., Hussein, H., Atteya, L.A., Elsebaie, M.A.T., Abo Elnasr, L.S., Sakr, R.A., Salem, H.S.E., Ismail, A.F., Saad, A.M., Ahmed, J., Elsebaie, M.A.T., Rahman, M., Ruhban, I.A., Elgazar, N.M., Alagha, Y., Osman, M.H., Alhusseiny, A.M., Khalaf, M.M., Younes, A.-A.F., Abdulkarim, A., Younes, D.M., Gadallah, A.M., Elkashash, A.M., Fala, S.Y., Zaki, B.M., Beezley, J., Chittajallu, D.R., Manthey, D., Gutman, D.A., Cooper, L.A.D., 2019. Structured crowdsourcing enables convolutional segmentation of histology images. Bioinformatics 35 (18), 3461–3467.

Anoraganingrum, D., 1999. Cell segmentation with median filter and mathematical morphology operation. In: Proceedings 10th International Conference on Image Analysis and Processing. IEEE, pp. 1043–1046.

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al., 2018.

Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261.
Bearman, A., Russakovsky, O., Ferrari, V., Fei-Fei, L., 2016. Whats the point: semantic segmentation with point supervision. In: European Conference on Computer Vision. Springer, pp. 549–565.
Belharbi, S., Ben Ayed, I., McCaffrey, L., Granger, E., 2021. Deep active learning for joint classification & segmentation with weak annotator. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3338–3347.
Blum, A., Mitchell, T., 1998. Combining labeled and unlabeled data with co-training. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, pp. 92–100.
Brambilla, E., Le Teuff, G., Marguet, S., Lantuejoul, S., Dunant, A., Graziano, S., Pirker, R., Douillard, J.-Y., Le Chevalier, T., Filipits, M., et al., 2016. Prognostic effect of tumor lymphocytic infiltration in resectable non–small-cell lung cancer. J. Clin. Oncol. 34 (11), 1223.
Bremnes, R.M., Dønnem, T., Al-Saad, S., Al-Shibli, K., Andersen, S., Sirera, R., Camps, C., Marinez, I., Busund, L.-T., 2011. The role of tumor stroma in cancer progression and prognosis: emphasis on carcinoma-associated fibroblasts and non-small cell lung cancer. J. Thorac. Oncol. 6 (1), 209–217.
Budd, S., Robinson, E.C., Kainz, B., 2021. A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med. Image Anal. 102062.
Cai, Z., Ravichandran, A., Maji, S., Fowlkes, C., Tu, Z., Soatto, S., 2021. Exponential moving average normalization for self-supervised and semi-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 194–203.
Chan, L., Hosseini, M.S., Rowsell, C., Plataniotis, K.N., Damaskinos, S., 2019. HistoSegNet: semantic segmentation of histological tissue type in whole slide images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10662–10671.
Chang, Y.-T., Wang, Q., Hung, W.-C., Piramuthu, R., Tsai, Y.-H., Yang, M.-H., 2020. Weakly-supervised semantic segmentation via sub-category exploration. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N., 2018. Grad–CAM++: generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 839–847.
Chen, H., Qi, X., Yu, L., Heng, P.-A., 2016. DCAN: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2487–2496.
Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.
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., 2020. Self-similarity student for partial label histopathology image segmentation. arXiv preprint arXiv:2007.09610.
Dai, J., He, K., Sun, J., 2015. BoxSup: exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1635–1643.
Deng, S., Zhang, X., Yan, W., Chang, I.C., Xu, Y., 2020. Deep learning in digital pathology image analysis: a survey. Front. Med. (6).
Denkert, C., von Minckwitz, G., Darb-Esfahani, S., Lederer, B., Heppner, B.I., Weber, K.E., Budczies, J., Huober, J., Klauschen, F., Furlanetto, J., et al., 2018. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncol. 19 (1), 40–50.
DeVries, T., Taylor, G. W., 2017. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552.
Diamond, J., Anderson, N.H., Bartels, P.H., Montironi, R., Hamilton, P.W., 2004. The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia. Hum. Pathol. 35 (9), 1121–1131.
Doyle, S., Monaco, J., Feldman, M., Tomaszewski, J., Madabhushi, A., 2011. An active learning based classification strategy for the minority class problem: application to histopathology annotation. BMC Bioinf. 12 (1), 424.
Gao, Z., Puttapirat, P., Shi, J., Li, C., 2020. Renal cell carcinoma detection and subtyping with minimal point-based annotation in whole-slide images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 439–448.
Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. HoVer-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 101563.
Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144 (5), 646–674.
Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H., 2016. Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2424–2433.
Huang, Z., Wang, X., Wang, J., Liu, W., Wang, J., 2018. Weakly-supervised semantic segmentation network with deep seeded region growing. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7014–7023.
Jia, Z., Huang, X., Eric, I., Chang, C., Xu, Y., 2017. Constrained deep weak supervision for histopathology image segmentation. IEEE Trans. Med. Imaging 36 (11), 2376–2388.
Jiang, P.-T., Han, L.-H., Hou, Q., Cheng, M.-M., Wei, Y., 2021. Online attention accumulation for weakly supervised semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell..
Kather, J.N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.-A., Gaiser, T., Marx, A., Valous, N.A., Ferber, D., et al., 2019. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med. 16 (1), e1002730.
Kather, J.N., Pearson, A.T., Halama, N., Jäger, D., Krause, J., Loosen, S.H., Marx, A., Boor, P., Tacke, F., Neumann, U.P., Grabsch, H.I., Yoshikawa, T., Brenner, H., Chang-Claude, J., Hoffmeister, M., Trautwein, C., Luedde, T., 2019. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat. Med. 25 (7), 1054–1056.
Kim, B., Choo, J., Kwon, Y.-D., Joe, S., Min, S., Gwon, Y., 2021. SelfMatch: combining contrastive self-supervision and consistency for semi-supervised learning. arXiv preprint arXiv:2101.06480.
Kong, J.C., Guerra, G.R., Pham, T., Mitchell, C., Lynch, A.C., Warrier, S.K., Ramsay, R.G., Heriot, A.G., 2019. Prognostic impact of tumor-infiltrating lymphocytes in primary and metastatic colorectal cancer: a systematic review and meta-analysis. Dis. Colon Rectum 62 (4), 498–508.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105.
Kweon, H., Yoon, S.-H., Kim, H., Park, D., Yoon, K.-J., 2021. Unlocking the potential of ordinary classifier: class-specific adversarial erasing framework for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6994–7003.
Laine, S., Aila, T., 2016. Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Lee, H., Jeong, W.-K., 2020. Scribble2Label: scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 14–23.
Lee, S., Lee, M., Lee, J., Shim, H., 2021. Railroad is not a train: saliency as pseudo-pixel supervision for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5495–5505.
Lerousseau, M., Vakalopoulou, M., Classe, M., Adam, J., Battistella, E., Carré, A., Estienne, T., Henry, T., Deutsch, E., Paragios, N., 2020. Weakly supervised multiple instance learning histopathological tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 470–479.
Li, Y., Chen, J., Xie, X., Ma, K., Zheng, Y., 2020. Self-loop uncertainty: a novel pseudo-label for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 614–623.
Liang, Q., Nan, Y., Coppola, G., Zou, K., Sun, W., Zhang, D., Wang, Y., Yu, G., 2018. Weakly supervised biomedical image segmentation by reiterative learning. IEEE J. Biomed. Health Inform. 23 (3), 1205–1214.
Lin, D., Dai, J., Jia, J., He, K., Sun, J., 2016. ScribbleSup: scribble-supervised convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3159–3167.
Lin, J., Han, G., Pan, X., Liu, Z., Chen, H., Li, D., Jia, X., Shi, Z., Wang, Z., Cui, Y., Li, H., Liang, C., Liang, L., Wang, Y., Han, C., 2022. PDBL: improving histopathological tissue classification with plug-and-play pyramidal deep-broad learning. IEEE Trans. Med. Imaging doi:10.1109/TMI.2022.3161787. 1–1
Liu, Q., Yu, L., Luo, L., Dou, Q., Heng, P.A., 2020. Semi-supervised medical image classification with relation-driven self-ensembling model. IEEE Trans. Med. Imaging 39 (11), 3429–3440.
Liu, Y., Cao, J., Li, B., Yuan, C., Hu, W., Li, Y., Duan, Y., 2019. Knowledge distillation via instance relationship graph. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7096–7104.
Mahapatra, D., Bozorgtabar, B., Thiran, J.-P., Reyes, M., 2018. Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 580–588.
Mao, Y., Keller, E.T., Garfield, D.H., Shen, K., Wang, J., 2013. Stromal cells in tumor micro environment and breast cancer. Cancer Metastasis Rev. 32 (1–2), 303–315.
Marini, N., Otálora, S., Müller, H., Atzori, M., 2021. Semi-supervised training of deep convolutional neural networks with heterogeneous data and few local annotations: an experiment on prostate histopathology image classification. Med. Image Anal. 73, 102165.
Ni, H., Liu, H., Wang, K., Wang, X., Zhou, X., Qian, Y., 2019. Wsi-Net: branch-based and hierarchy-aware network for segmentation and classification of breast histopathological whole-slide images. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 36–44.
Qaiser, T., Tsang, Y.-W., Taniyama, D., Sakamoto, N., Nakane, K., Epstein, D., Rajpoot, N., 2019. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. Med. Image Anal. 55, 1–14.
Qu, H., Wu, P., Huang, Q., Yi, J., Riedlinger, G.M., De, S., Metaxas, D.N., 2019. Weakly supervised deep nuclei segmentation using points annotation in histopathology images. In: International Conference on Medical Imaging with Deep Learning. PMLR, pp. 390–400.
Qu, H., Wu, P., Huang, Q., Yi, J., Yan, Z., Li, K., Riedlinger, G.M., De, S., Zhang, S., Metaxas, D.N., 2020. Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. IEEE Trans. Med. Imaging.

Rkaczkowski, L., Mozejko, M., Zambonelli, J., Szczurek, E., 2019. Ara: accurate, reliable and active histopathological image classification framework with Bayesian deep learning. Sci. Rep. 9 (1), 1–12.
Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad–CAM: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618–626.
Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.
Settles, B., 2009. Active learning literature survey.
Shen, H., Tian, K., Dong, P., Zhang, J., Yan, K., Che, S., Yao, J., Luo, P., Han, X., 2020. Deep active learning for breast cancer segmentation on immunohistochemistry images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 509–518.
Silva-Rodríguez, J., Colomer, A., Naranjo, V., 2021. WegleNet: a weakly-supervised convolutional neural network for the semantic segmentation of Gleason grades in prostate histology images. Comput. Med. Imaging Graph. 88, 101846.
Sirinukunwattana, K., Snead, D.R., Rajpoot, N.M., 2015. A novel texture descriptor for detection of glandular structures in colon histology images. In: Medical Imaging 2015: Digital Pathology, Vol. 9420. International Society for Optics and Photonics, p. 94200S.
Skrede, O.-J., De Raedt, S., Kleppe, A., Hveem, T.S., Liestøl, K., Maddison, J., Askautrud, H.A., Pradhan, M., Nesheim, J.A., Albregtsen, F., et al., 2020. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet 395 (10221), 350–360.
Srinidhi, C.L., Ciga, O., Martel, A.L., 2020. Deep neural network models for computational histopathology: a survey. Med. Image Anal. 101813.
Sun, G., Wang, W., Dai, J., Gool, L. V., 2020. Mining cross-image semantics for weakly supervised semantic segmentation. arXiv preprint arXiv:2007.01947.
Tabesh, A., Teverovskiy, M., Pang, H.-Y., Kumar, V.P., Verbel, D., Kotsianti, A., Saidi, O., 2007. Multifeature prostate cancer diagnosis and Gleason grading of histological images. IEEE Trans Med Imaging 26 (10), 1366–1378.
Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, pp. 1195–1204.
Tokunaga, H., Iwana, B.K., Teramoto, Y., Yoshizawa, A., Bise, R., 2020. Negative pseudo labeling using class proportion for semantic segmentation in pathology. In: European Conference on Computer Vision. Springer, pp. 430–446.
van Rijthoven, M., Balkenhol, M., Siliņa, K., van der Laak, J., Ciompi, F., 2021. HookNet: multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. Med. Image Anal. 68, 101890.
Wada, K., 2016. labelme: Image polygonal annotation with Python. https://github.com/wkentaro/labelme.
Wang, X., Chen, H., Gan, C., Lin, H., Dou, Q., Tsougenis, E., Huang, Q., Cai, M., Heng, P.-A., 2019. Weakly supervised deep learning for whole slide lung cancer image analysis. IEEE Trans. Cybern. 50 (9), 3950–3962.
Wang, Y., Zhang, J., Kan, M., Shan, S., Chen, X., 2020. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12275–12284.
Wang, Y., Zhang, J., Kan, M., Shan, S., Chen, X., 2020. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J., Huang, T.S., 2018. Revisiting dilated convolution: a simple approach for weakly- and semi-supervised semantic segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7268–7277.
Wen, S., Kurc, T.M., Hou, L., Saltz, J.H., Gupta, R.R., Batiste, R., Zhao, T., Nguyen, V., Samaras, D., Zhu, W., 2018. Comparison of different classifiers with active learning to support quality control in nucleus segmentation in pathology images. In: AMIA Summits Translational Science Proceedings, 2018, p. 227.
Wen, Z., Feng, R., Liu, J., Li, Y., Ying, S., 2021. GCSBA-Net: gabor-based and cascade squeeze bi-attention network for gland segmentation. IEEE J. Biomed. Health Inform. 25 (4), 1185–1196.
Wu, Z., Shen, C., Van Den Hengel, A., 2019. Wider or deeper: revisiting the resnet model for visual recognition. Pattern Recognit. 90, 119–133.
Xia, Y., Yang, D., Yu, Z., Liu, F., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H., 2020. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Med. Image Anal. 65, 101766.
Xie, Y., Zhang, J., Liao, Z., Verjans, J., Shen, C., Xia, Y., 2020. Pairwise relation learning for semi-supervised gland segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 417–427.
Yang, L., Zhang, Y., Chen, J., Zhang, S., Chen, D.Z., 2017. Suggestive annotation: a deep active learning framework for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 399–407.
Zhai, X., Oliver, A., Kolesnikov, A., Beyer, L., 2019. S4l: self-supervised semi-supervised learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1476–1485.
Zhang, J., Hua, Z., Yan, K., Tian, K., Yao, J., Liu, E., Liu, M., Han, X., 2021. Joint fully convolutional and graph convolutional networks for weakly-supervised segmentation of pathology images. Med. Image Anal. 73, 102183.
Zhao, B., Chen, X., Li, Z., Yu, Z., Yao, S., Yan, L., Wang, Y., Liu, Z., Liang, C., Han, C., 2020. Triple U-net: hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. Med. Image Anal. 65, 101786.
Zhao, Z., Zeng, Z., Xu, K., Chen, C., Guan, C., 2021. DSAL: deeply supervised active learning from strong and weak labelers for biomedical image segmentation. IEEE J. Biomed. Health Inform..
Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929.
Zhou, Z., Shin, J.Y., Gurudu, S.R., Gotway, M.B., Liang, J., 2021. Active, continual fine tuning of convolutional neural networks for reducing annotation efforts. Med. Image Anal. 71, 101997.