

SAC-Net: Learning with weak and noisy labels in histopathology image segmentation

Author:
Guo, R; Xie, K; Pagnucco, M; Song, Y

Publication details:
Medical Image Analysis
v. 86
Medium: Print-Electronic
1361-8415 (ISSN); 1361-8423 (ISSN)

Publication Date:
2023-05-01

Publisher DOI:
https://doi.org/10.1016/j.media.2023.102790

License:
https://creativecommons.org/licenses/by-nc-nd/4.0/
Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/106756 in https://unsworks.unsw.edu.au on 2026-02-27

Medical Image Analysis (2022)

Contents lists available at ScienceDirect

Medical Image Analysis

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

SAC-Net: Learning with Weak and Noisy Labels in Histopathology Image Segmentation

Ruoyu Guo[a,1], Kunzi Xie[a,1], Maurice Pagnucco[a], Yang Song[a,*]

[a]*School of Computer Science and Engineering, University of New South Wales, Australia*

ARTICLE INFO

Article history:
Received 16 Dec 2021
Received in final form 16 Dec 2021

Communicated by S. Sarkar

Keywords: Nuclei segmentation, Weakly supervised learning, Noise labels, Point annotations

ABSTRACT

Deep convolutional neural networks have been highly effective in segmentation tasks. However, segmentation becomes more difficult when training images include many complex instances to segment, such as the task of nuclei segmentation in histopathology images. Weakly supervised learning can reduce the need for large-scale, high-quality ground truth annotations by involving non-expert annotators or algorithms to generate supervision information for segmentation. However, there is still a significant performance gap between weakly supervised learning and fully supervised learning approaches. In this work, we propose a weakly-supervised nuclei segmentation method in a two-stage training manner that only requires annotation of the nuclear centroids. First, we generate boundary and superpixel-based masks as pseudo ground truth labels to train our SAC-Net, which is a segmentation network enhanced by a constraint network and an attention network to effectively address the problems caused by noisy labels. Then, we refine the pseudo labels at the pixel level based on Confident Learning to train the network again. Our method shows highly competitive performance of cell nuclei segmentation in histopathology images on three public datasets. Code will be available at: `https://github.com/RuoyuGuo/MaskGA_Net`

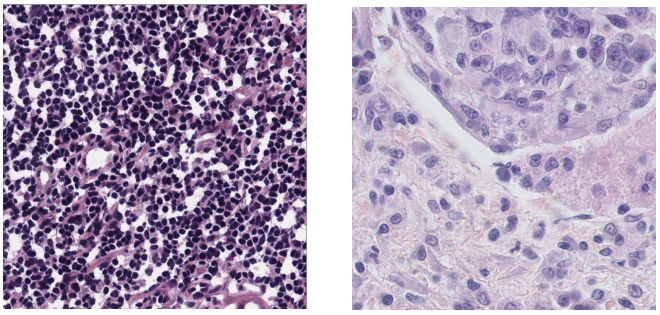
© 2022 Elsevier B. V. All rights reserved.

1. Introduction

Automated nuclei segmentation in microscopy images is critical in biomedical research and clinical practice, including nuclear morphology analysis, cancer diagnosis and grading (Liang et al., 2022). The morphological features of nuclei, such as the number, shape and size of cells, can be extracted after segmentation for quantitative analysis of diseases. Nuclei segmentation in histopathology images is however a challenging task due to the complex cellular structure. As shown in Figure 1, there are typically hundreds or thousands of cells in one image, and the separation between foreground and background can be quite obscure (Kumar et al., 2017; Naylor et al., 2019; Graham et al., 2019b; Vu et al., 2019). Besides, cells often touch each other and have different appearances even from the same organ.

With the recent success in deep learning, many complicated computer vision tasks have achieved significant breakthroughs. Similarly, some studies have shown that deep learning-based methods can effectively improve the performance of nuclei segmentation in histopathology images. In particular, fully supervised histopathology image segmentation algorithms have achieved excellent performance (Chen et al., 2017; Graham et al., 2019a). Some methods have incorporated boundary detection (Kumar et al., 2017; Zhou et al., 2019) and recurrent feature extraction (Xiang et al., 2020) to further improve the performance of nuclei segmentation. However, these studies are based on fully-supervised learning that requires a large amount of high-quality annotations to achieve high performance. Due to the prior knowledge required for annotating such images accurately, it is very challenging to acquire large-scale datasets to train fully-supervised nuclei segmentation models.

*Corresponding author:
e-mail: `yang.song1@unsw.edu.au` (Yang Song)
[1]These authors contributed equally to this work.



(a) MoNuSeg (b) TNBC

Fig. 1. Samples of histopathology images from the MoNuSeg and TNBC datasets.

In recent years, weakly-supervised learning methods have received increasing attention. For segmentation problems, weakly-supervised learning only needs a small number of coarse annotations, such as point (Laradji et al., 2018; Selvaraju et al., 2017), scribble (Zhang et al., 2020), bounding box (Oh et al., 2021) or image labels (Li et al., 2018) in comparison to fully-supervised learning that requires pixel-wise annotations. While these approaches can significantly reduce the annotation burden when collecting datasets, it is not easy to apply these methods to nuclei segmentation since nuclei segmentation tasks require dense annotation (Koohbanani et al., 2020).

There have been limited studies on weakly-supervised cell or nuclei segmentation (Yoo et al., 2019; Qu et al., 2019a; Tian et al., 2020; Dong et al., 2020; Chamanzar and Nie, 2020; Nishimura et al., 2019; Lee and Jeong, 2020). Most of these approaches directly train the segmentation network with few annotations (e.g., point annotations) and apply some techniques such as conditional random fields (CRF) to refine the network, or generate pseudo contour labels to train the network. However, without precise pixel-wise ground truth annotations, it is difficult to train the deep learning models to achieve accurate segmentation. Furthermore, the pseudo labels around the contours of the nuclei are often highly noisy and hence introduce mislabelling during model training. In our research, we adopt point annotations to develop a weakly supervised method for nuclei segmentation in histopathology images. We choose to work with point annotations because they are easier to obtain and most existing weakly-supervised nuclei segmentation approaches (Yoo et al., 2019; Qu et al., 2019a; Tian et al., 2020; Qu et al., 2020a,b) have been developed using point annotations.

In this paper, we present a triple network, namely SAC-Net, with a two-stage label refinement framework to train a weakly-supervised nuclei segmentation model on histopathology images. SAC-Net aims to reduce the impact of the noisy label in weakly supervised learning for cell segmentation. We first generate pseudo labels comprising centroids and Voronoi boundaries, and superpixel labels to provide supervision information for training the networks. The proposed SAC-Net comprises a feature aggregation segmentation network, an attention network and a constraint network with a pseudo label refinement method. More specifically, the segmentation network consists of an encoder-decoder with multi-view aggregation. Two feature maps are generated by the constraint and attention network to assist the segmentation network to extract the correct label information from pseudo labels for improving the final segmentation results. The shallow constraint network extracts low-level feature maps and combines the high-level features from the segmentation network. This extracted information can be emphasised during the inference process in the segmentation network to reduce the impact of false positive labels from superpixels. This combined representation can keep more details of the image information. On the other hand, the attention network restricts the segmentation network from learning the wrong information from the superpixel labels. The integration of attention network and constraint network helps alleviate false and missed detections through two-level attention maps.

In addition, to further tackle the noise in pseudo labels, we adopt the Confident Learning (CL) (Northcutt et al., 2019) algorithm in our weakly-supervised segmentation model, which can further refine the boundary of the segmented nuclei. Our method achieves competitive results with a relatively lightweight segmentation model when compared to recently proposed weakly-supervised or fully-supervised nuclei segmentation methods on three public datasets. Our contributions can be summarised as follows:

- We propose a weakly-supervised learning framework, which can further refine cell boundaries for nuclei segmentation based on point annotations. The CL-based label correction process dramatically improves the segmentation result, which, for the first time, has been utilised for weakly-supervised nuclei segmentation.
- We introduce a SAC-Net, which includes a novel triple network for nuclei segmentation, consisting of a segmentation network, an attention network and a constraint network. We improve our backbone such that it can handle the complex shapes of nuclei by adding deformable convolutional layers and incorporating spatial attentions. The constraint network is introduced to better preserve the low-level texture features in order to tackle the over-fitting issue of deep network architectures on small datasets as well as the problem caused by imprecise labels by superpixels. It can significantly reduce the impact of noisy labels without introducing extra computational cost at test time.
- We incorporate the attention network and constraint network to further address the difficulty of training due to noisy pseudo labels. The attention network aims to solve the problem of data-dependent denoising. The constraint network tackles the over-fitting issue of deep network architectures on small datasets as well as the problem caused by imprecise labels.
- We also conducted experimental evaluations in a highly noisy setup different from previous work, as a way to better simulate clinical practice.

To this end, this paper is an extended version of our recent paper published in MICCAI 2021 (Guo et al., 2021). In this study, besides adding the constraint network, we also improve

the backbone of the segmentation and attention networks by integrating deformable operations and attention modules. We have also conducted more comprehensive experimental evaluations using three public datasets MoNuSeg (Kumar et al., 2017), TNBC (Naylor et al., 2019) and CPM-17 (Vu et al., 2019).

2. Related Work

2.1. Cell and Nuclei Segmentation

There are many studies on network architecture design to improve segmentation results. CNN3 (Kumar et al., 2017) extends the binary classification problem of nuclei segmentation to classification of three classes by adding the nucleus boundary class. CIA-Net (Zhou et al., 2019) achieves excellent performance by leveraging a bi-decoder architecture with additional supervision of contours. Similarly, a two-stage learning method is proposed in (Kang et al., 2019) to improve the accuracy by including another network to learn a third boundary class from segmentation outputs. Hover-Net (Graham et al., 2019c) is proposed to address the intra-class variability problem for different kinds of cells and crowded cells to enhance the boundary delineation as well. Rather than using the boundary information, BiO-Net (Xiang et al., 2020) applies skip connections in both the forward and backward directions to enhance semantic feature learning. To avoid the loss of local details after the downsampling operation, FullNet (Qu et al., 2019b) employs dilated convolution layers in each convolutional layer.

Furthermore, a two-stage fully convolutional neural network is designed by Kablan et al. (2020) for cell segmentation. A CU-Net is proposed by Akbaş and Kozubek (2020), which is a condensed network based on U-Net (Ronneberger et al., 2015). Alternatively, a U-Net and Mask R-CNN (He et al., 2017) ensembled method (Vuola et al., 2019) with colour transfer (Reinhard et al., 2001) as preprocessing is proposed. To solve the problem of inconsistent colours in the dataset, a triple U-Net structure is designed (Zhao et al., 2020), which does not require colour normalisation. The network gradually merges and learns features from different fields or tasks through a feature aggregation module. In addition, due to the small datasets, an image synthesis network is introduced (Mahmood et al., 2020), which utilises CycleGAN (Zhu et al., 2017) to learn a mapping from masks to pathology images. However, these methods are designed for fully-supervised learning which cannot be directly applied to point annotations. Also, high-quality supervision that can highlight boundary information often increases the network complexity.

2.2. Weakly Supervised Cell and Nuclei Segmentation

For weakly-supervised cell segmentation, the Voronoi boundary (Aurenhammer, 1991) plays a crucial role in this area since it can generate negative labels with only point annotations. Segmentation methods often train segmentation networks with the Voronoi boundary and richer annotations generated from points, then apply methods for refinement. For example, the method of (Qu et al., 2019a) produces pseudo labels for nuclei segmentation from clustering and Voronoi diagrams by using an improved U-Net. With only sparse point annotations for weakly supervised learning, Chen et al. (2020) obtained good segmentation results with better generalisability by using a novel end-to-end weakly supervised learning framework named WESUP. PseudoEdgeNet (Yoo et al., 2019) equips another shallow CNN network to recognise the contour of nuclei with the Sobel filtered segmentation output as supervision. In addition, an instance loss is designed to help the neural network learn a continuous peak-shaped map from point annotations (Dong et al., 2020). A multi-task learning method (Chamanzar and Nie, 2020) that combines three different losses and learns from one of them at each iteration also shows performance gains.

Moreover, a self-supervised iterative refinement framework is proposed for weakly supervised nuclei segmentation (Tian et al., 2020). Besides point annotations, other methods have been developed based on scribble labels with an iterative learning method (Lee and Jeong, 2020), and such labels provide more supervision information than point annotations. Another method involves annotations of mixed labels combining point and full annotations (Qu et al., 2020a). In addition, in another method (Qu et al., 2020b), uncertainty maps generated from Bayesian deep learning tools are utilised to select the representative nuclei for mask annotations automatically. Although these methods achieve relatively good segmentation performance with weak labels, the iterative training processes increase the method complexity and there is often still a high level of noise in pseudo labels which affects the model training.

2.3. Weakly Supervised Segmentation for General Objects

For weakly supervised learning, point annotations are also widely used in general object segmentation, such as (Maninis et al., 2018). Weakly-supervised segmentation with bounding box labels has also been designed (Zhao et al., 2018). Another weakly-supervised nuclei segmentation method (Hsu et al., 2019) adopts a multi-instance learning technique for training with tight bounding box annotations.

Furthermore, some studies propose new losses to improve performance. Constrained-CNN losses (CCL) (Kervadec et al., 2019) is a general loss for weakly-supervised segmentation. This loss limits the object within a range based on prior knowledge, like the size of an object or the presence of a particular disease. A localisation-based counting loss (LC) with four terms is proposed to encourage the network to learn from different levels (Laradji et al., 2018). In addition, to help the network find discriminative regions, a consistency-based (CB) loss (Laradji et al., 2021; Jeong et al., 2019) is introduced that encourages output predictions to be consistent with spatial transformations of the input images. However, CCL requires additional annotations, whereas LC and CB cannot constrain precise object shapes since they only learn the approximate locations of objects.

3. Methods

As shown in Fig. 2, at Stage I, SAC-Net is trained using the pseudo ground truth labels that are generated based on superpixel and Voronoi boundary information. At Stage II, the CL

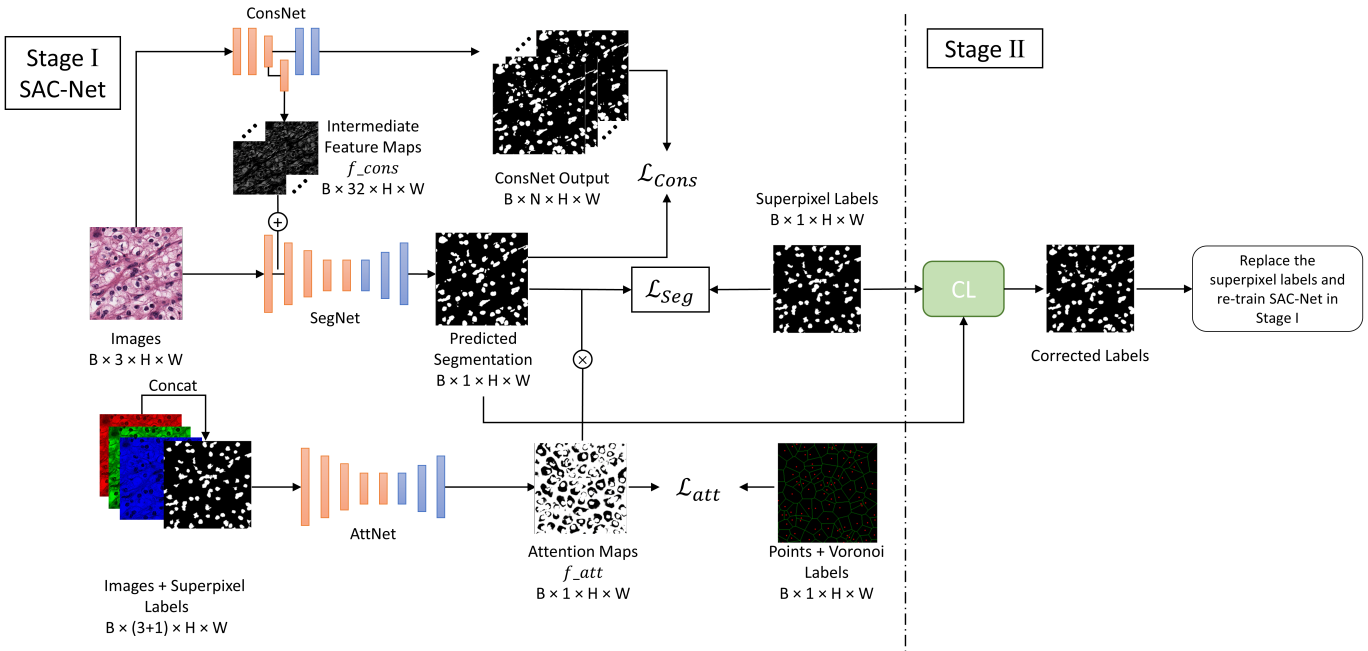


Fig. 2. Overview of our proposed weakly-supervised nuclei segmentation method, where Stage I contains the designed SAC-Net and Stage II applies CL for label refinement and retraining. + and × represent for element-wise addition and multiplication operations.

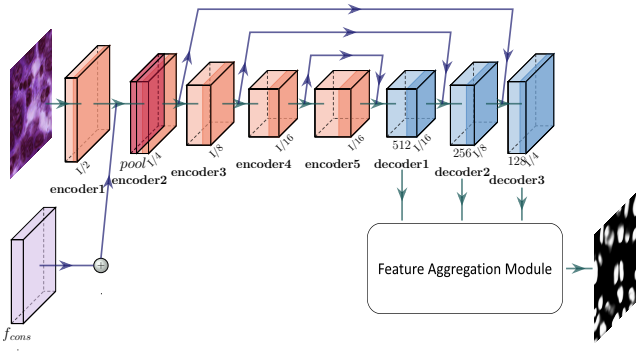


Fig. 3. Feature aggregation network architecture. + represents for element-wise addition operation.

algorithm is adopted to refine labels, and we replace the superpixel labels with the revised labels and conduct the same training process again as Stage I. SAC-Net contains three sub-networks and, at test time, only the segmentation network (SegNet) is used.

3.1. Pseudo Labels

While the datasets contain full pixel-wise annotations, to simulate a weakly-supervised learning scenario, we generate point annotations by calculating the centroid of each nucleus. This supervision is a binary mask that for each nucleus, only the centroid pixel of it is marked as 1 and other pixels are marked as 0. Subsequently, since training a network with few annotations is incapable of producing good results, we need to provide pixel-wise annotations based on point annotations. To do this, two types of pseudo ground truth labels, y_v and y_s, are generated, which are **Points+Voronoi Labels** and **Superpixel Labels**, as shown in Fig. 2.

We first generate the pseudo labels y_v to provide location and boundary information. Specifically, the Euclidean distance transform is applied based on the centroid points to produce the Voronoi boundaries. The Voronoi boundary partitions the image into Voronoi cells and the pixels within each Voronoi cell are closer to the point annotation in that cell than any other point annotations. Similar to previous approaches (Yoo et al., 2019; Qu et al., 2019a; Tian et al., 2020; Dong et al., 2020; Chamanzar and Nie, 2020; Qu et al., 2020a), we assume that nuclei only appear within each Voronoi cell. On the other hand, existing methods typically use Voronoi boundaries as negative labels and point annotations as positive labels to train the segmentation network with partial binary cross-entropy loss, ignoring the pixels without pseudo labels. In our method, the Voronoi boundaries and points are considered true labels of nuclei and background in superpixels, whereas the remaining pixels that are neither the Voronoi boundaries nor points might be mislabeled in superpixels. We made this assumption because points are always accurate nuclei pixels (annotated by human) and most of the Voronoi boundaries lie reliably within background pixels. With these annotations, we can train the attention network with our attention loss (Section 3.2.2).

In practice, points and Voronoi labels are incapable of providing enough pixel-level information for training an accurate segmentation network, which may lead to poor convergence of networks. In addition, we find the segmentation performance does not improve much with the CL component alone. The CL component demonstrates an effect in denoising the probability output of the segmentation but shows little improvement in the actual segmentation results. To achieve more accurate segmen-

tation results, our first technique is to generate a second type of pseudo labels y_s based on superpixels. The SLIC (Achanta et al., 2012) algorithm is utilised to divide the images into superpixels. We argue that superpixels with points inside them have a high probability of overlapping with true nuclei labels. Therefore, such superpixels are considered true nuclei locations and assigned positive labels. The other superpixels are considered background and given negative labels. In this way, all pixels are annotated with positive or negative pseudo labels.

3.2. SAC-Net

As shown in Fig. 2, the whole SAC-Net framework contains three networks: the segmentation network (SegNet), attention network (AttNet) and constraint network (ConsNet). ConsNet and SegNet both take the original image as input, while AttNet feeds the concatenation of original images and corresponding superpixel labels. To highlight the nuclear areas, we extract the intermediate feature maps f_{cons} from ConsNet and merge them into SegNet. AttNet generates the attention maps f_{att} and corrects the nuclei segmentation from SegNet. At test time, only SegNet is used for inference.

3.2.1. Segmentation Network

To better learn the complex features of nuclei, we design a feature aggregation segmentation network, SegNet, containing an encoder-decoder architecture with an additional feature aggregation module, as shown in Fig. 3. The original images are used as the inputs and the superpixel labels y_s are used as the supervision for training the segmentation network. The segmentation network aims to adapt to the complexity of nuclei and obtain the initial segmentation results.

·**Network structure.** ResNet-50 (He et al., 2016) is used in the encoder, while the last four blocks are connected to the decoder to reduce the computation caused by the spatial dimension. Besides, the feature map f_{cons} from ConsNet (Section 3.2.3) is merged with the feature maps generated by **encoder1**. We use element-wise addition to aggregate them and pass the results into the following layers. We also utilise deformable convolution (Dai et al., 2017) in each encoder layer, to compensate for the limitations of fixed geometry transformation in the convolutional building module. In deformable convolution, an offset is added to each sampling point in the convolution operation and the offsets are learned through an additional convolution layer, so that the network can better exploit spatial information for feature representation. Furthermore, a dilation strategy (Yu and Koltun, 2015) is used in the last encoder block and the dilation rates are 2 with an output stride of 1/16. This method can preserve image resolution and increase the receptive-field size at the same time.

Traditional decoder layers concatenate feature maps from two levels followed by deconvolutional layers to recover the feature resolution. However, it might preserve redundant features from previous layers and increase the computational complexity. Thus, we design a more effective decoder block, as shown in Fig. 4. We first apply a 1×1 convolutional layer to the output of the previous layer as a transformation. After upsampling to the same shape of feature maps from lower encoders,

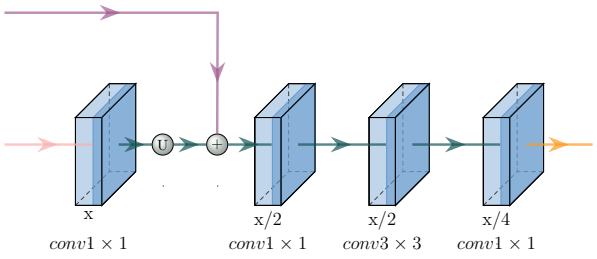


Fig. 4. Decoder block structure. The pink arrow denotes previous layers, the purple arrow is a skip connection, and the orange arrow represents the output of the decoder block. U represents for upsample. + represents for adding the features.

an element-wise addition is utilised to merge them. This is followed by 1×1, 3×3, and 1×1 convolutional layers to learn a weighted sum of feature maps and downsample them.

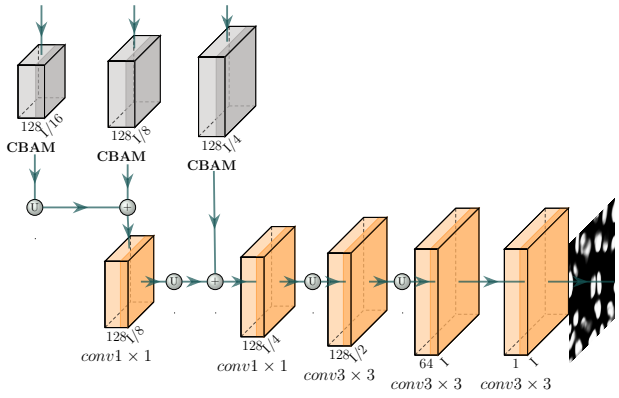


Fig. 5. Feature aggregation module. U represents for upsample. + represents for adding the features.

In addition, since the feature maps from each decoder block have different scales and not all levels of features contribute equally to the final output, we attach an aggregation module to the decoder blocks to produce the segmentation output. The aggregation module has a combined channel and spatial attention layer. The inspiration comes from the convolution block attention module (CBAM of (Woo et al., 2018)), which can highlight important information from the output of each decoder block for better aggregation. As illustrated in Fig. 5, the feature maps from three decoder blocks are passed through three corresponding CBAM blocks. The outputs from CBAM are merged by element-wise addition and fed through an upsampling operation to obtain the final predicted probability maps. CBAM helps refine the intermediate features by sequencing channel and spatial attention of each convolution block with 1D channel attention map and a 2D spatial attention map multiplication (Woo et al., 2018). In detail, the channel sub-module utilises both max-pooling and average-pooling outputs, which are then forwarded to a shared network to produce the channel attention map. The shared network is composed of a multi-layer perceptron (MLP) with one hidden layer. The output features from

MLP are merged by using element-wise addition. The spatial sub-module utilises similar outputs pooled along the channel axis and forwards them to a convolution layer to produce a spatial attention map. CBAM sequentially applies channel and spatial attention sub-modules and efficiently helps the information flow within the network by learning which information to emphasise or suppress.

Loss function. To train the SegNet, we choose binary cross-entropy loss as $\mathcal{L}_{seg}$ to optimise it:

$$\mathcal{L}_{seg} = \mathcal{L}_{bce}(f(I) \odot f_{att}, y_s) \tag{1}$$

where f, I and $\odot$ represent the SegNet, input image and element-wise multiplication, respectively. Here, each vector of the SegNet prediction is element-wise multiplied by the corresponding vector f_{att} from the output of the attention network (Section 3.2.2). The attention network aims to guide the segmentation of SegNet during the training process to achieve more precise segmentation result. More details of the attention network are described in the next section.

3.2.2. Attention Network

Our superpixel labels can capture rich contour information in the image; however, there is a high degree of label noise in such labels. When there is noise in ground truth labels due to imprecise manual annotations, existing approaches have utilised data re-weighting to improve the network training (Zhu et al., 2019; Xue et al., 2019; Wang et al., 2020; Mirikharaji et al., 2019). Though such methods have solved the noisy label issue to a certain extent, they require a fair amount of clean labels for supervision and have not been adapted into weakly-supervised segmentation problems, which exhibit a much higher degree of noise in pseudo labels.

In our approach, instead of explicitly exploiting clean labels from pseudo labels, we design an attention network (AttNet), which can be considered a data-dependent denoising method to address noisy labels. The attention feature map generated by the attention network can help identify the true labels in y_s and further guide the segmentation network to focus on these regions. More specifically, AttNet produces a binary mask, which separates the central nuclei areas and background from the noisy label regions around the nuclei boundaries. This binary mask is then element-wise multiplied with the segmentation output of SegNet to focus the training of SegNet on the more confident regions.

Network structure. As illustrated in Fig. 2, we first concatenate the input image and its corresponding superpixel labels to form a 4-channel input to AttNet. The superpixel channel provides additional structural information in the image, particularly highlighting the contours. Training of AttNet is supervised by the points and Voronoi labels, which indicate the true nuclei and background areas. The output of AttNet thus identifies the regions that are more confidently classified as nuclei or background, masking the areas around nuclei boundaries as noisy regions. In other words, AttNet effectively assigns pixel-wise weights to the output of SegNet for refining the segmentation learning.

The attention network employs the same architecture as the segmentation network but with a smaller ResNet-18 as the encoder. ResNet-18 is used here only to extract features that aim to simplify the aggregation module. We observed that using a deeper ResNet, such as ResNet-50 or ResNet-101, would not improve the segmentation performance but require more computational time during training.

Loss function. To train AttNet, we compute partial pixel-wise $\mathcal{L}_1$ loss between the AttNet output f_{att} and the Voronoi ground truth y_v:

$$f_{att} = g\left(c\left(I, y_s\right)\right) \tag{2}$$

$$\mathcal{L}_{att}\left(f_{att}, y_v\right) = \frac{\sum_{y \in y_v^+} |1 - f_{att}|_1}{|y_v^+|} + \alpha \cdot \frac{\sum_{y \in y_v^-} |1 - f_{att}|_1}{|y_v^-|} \tag{3}$$

where g, y_v^+, y_v^-, $|y_v^+|$, and $|y_v^-|$ denote attention network outputs, positive labels, negative labels, and the total numbers of positive labels and negative labels of I in y_v. $c(I, y_s)$ is the concatenation operation. α is a constant to balance the positive and negative labels. Note this loss ignores all pixels that are not annotated in y_v. A more intuitive view can be seen in Fig. 2. The attention feature map f_{att} is a binary mask where the white regions denote the background and a small central part of nuclei. The black regions surrounding the nuclei indicate the noise in y_s. When training SegNet, the loss in these black regions will be ignored, as defined in Eq. (1). In this way, the training of SegNet can focus on the regions with superpixel labels that are more likely to be accurate.

3.2.3. Constraint Network

While the attention network helps address the label noise in superpixels, the supervision is still not as accurate as full pixel-wise annotations. Therefore, to further enhance segmentation performance, we design a shallow constraint network, ConsNet, which can run in parallel with SegNet. ConsNet is a shallow CNN which can extract low-level information such as edges at a low level. These edges may be inaccurate but can provide structural constraints to train SegNet in addition to the superpixel labels. The feature map from ConsNet is also combined into SegNet effectively as an attention map to highlight the foreground areas.

ConsNet helps learn lower-level features, which are useful to correct the noise introduced by superpixel labels and further enhance the segmentation performance. Specifically, ConsNet extracts the features using a shallow network with three convolution layers which can capture the low-level image features such as edges. These low-level features are incorporated to constrain SegNet by element-wise addition, which can help maintain the edge information and then correct the segmentation problems around cell boundaries caused by SegNet. This operation can be regarded as a weighted summation and is useful for noise removal and feature enhancement.

Network structure. As illustrated in Fig. 6, in our design, inputs to ConsNet are the original images. To emphasise details of the input images, we first utilise two convolutional layers (**conv1** and **conv2**) to learn local details. Then, in the main

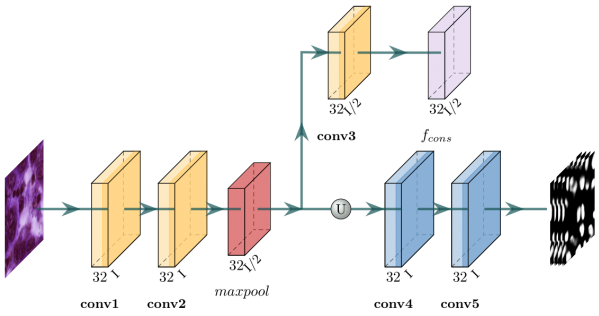


Fig. 6. Constraint network architecture. U represents for upsample. + represents for adding the features.

branch, we obtain the feature maps through an upsampling operation and another two convolutional layers (**conv4** and **conv5**) to get the final output result of ConsNet. In addition, another branch utilises the max-pooling layer (maxpool) to produce the feature maps with another convolutional layer (**conv3**) to generate f_{cons}. This intermediate feature map f_{cons} is merged into SegNet by an element-wise addition operation. To guarantee that ConsNet can generate feature maps at a different level from the other two networks, we use the output of SegNet $f(I)$ as supervision. Moreover, to avoid generating redundant feature maps, **conv1**-**4** in ConsNet are formed with a 3×3 kernel with only 32 channels followed by the BatchNorm layer and ReLU activation, while **conv5** has only a 3×3 kernel with N channels. Consequently, these shallow network layers can keep the low-level features of the original images to integrate with the high-level feature for predicting the segmentation results.

Loss function. We train ConsNet with our $\mathcal{L}_{cons}$ loss and use the output $f(I)$ from SegNet as ground truth:

$$\mathcal{L}_{cons} = \frac{\sum_{n=1}^{N} |f(I) - h(I)_n|_1}{N} \tag{4}$$

where $h(I)_n$ is the nth channel of the ConsNet output and $|...|_1$ is the common $\mathcal{L}_1$ loss. Unlike SegNet, which uses binary cross-entropy as the loss function, we utilise $\mathcal{L}_1$ as the loss function in ConsNet. This is because $\mathcal{L}_1$ loss is less sensitive to outliers, while the penalty of cross-entropy loss is higher than $\mathcal{L}_1$. Here, the outliers are hard samples, which correspond to noisy labels in our study. Thus, $\mathcal{L}_{cons}$ can help reduce the effects of noisy labels.

3.2.4. Total Loss Function

To train SegNet, AttNet and ConsNet simultaneously, we define the total loss $\mathcal{L}_{total}$ as a linear combination of $\mathcal{L}_{seg}$, $\mathcal{L}_{att}$ and $\mathcal{L}_{cons}$:

$$\mathcal{L}_{total} = \beta \cdot \mathcal{L}_{seg} + \gamma \cdot \mathcal{L}_{att} + \delta \cdot \mathcal{L}_{cons} \tag{5}$$

where β, γ and δ are the hyper-parameters to indicate the weights of $\mathcal{L}_{seg}$, $\mathcal{L}_{att}$ and $\mathcal{L}_{cons}$.

3.3. Pseudo Ground Truth Revision

After segmentation through SAC-Net, the segmentation results are still imprecise due to the supervision by pseudo labels. Thus, we further apply the CL (Northcutt et al., 2019) method to refine the superpixel labels y_s. Then the refined superpixel labels y_{re} are used to train the overall SAC-Net again. This process is defined in our model as Stage II. Specifically, assume that image data $\mathbf{X} = \{(x_0, y_{s0}), ..., (x_n, y_{sn})\}$ are labelled as class a. However, if the data is predicted to belong to another class b and its predicted probability is higher than a threshold, we consider that the label contains noise.

To identify the label error, CL first estimates the joint distribution $\mathbf{Q}_{y_s,y^*}$ between the superpixel labels y_s and true labels y^*. Specifically, we first build $\mathbf{C}_{y_s,y^*}$, which is an $|m| \times |m|$ matrix and m is the set of unique classes ({negative, positive} in our method and $|m| = 2$). Each entry $\mathbf{C}_{y_s=i,y^*=j}$[i][j] represents the total number of pixels with $y_s = i$ and having $\hat{p}(y^{=}j; x) > t_j$. Here, $\hat{p}(y^{=}j; x)$ is the predicted probability of a pixel x belonging to class j, y is the network output and t_j is a threshold. Formally, the definition of $\mathbf{C}_{y_s,y^*}$ is:

$$\begin{aligned}
\mathbf{C}_{y_s=i,y^*=j} =& |\mathbf{Z}_{y_s=i,y^*=j}|, where, \\
\mathbf{Z}_{y_s=i,y^*=j} =& \{z \in \mathbf{X} : y_{sz} = i, \hat{p}(y = j; z) > t_j, \\
& j = \underset{k}{\operatorname{argmax}}\, \hat{p}(y = k; z)\}
\end{aligned}$$

The choice of j in $\mathbf{Z}_{y_s=i,y^*=j}$[i][j] is the class that has the highest predicted probability among all classes of that pixel, and threshold t_j is the averaged predicted probability of all pixels belonging to class j and having $y_s = j$. Given $\mathbf{C}_{y_s,y^*}$, we estimate $\mathbf{Q}_{y_s,y^*}$ as a normalised version of $\mathbf{C}_{y_s,y^*}$:

$$\mathbf{Q}_{y_s,y^*} = \frac{\frac{\mathbf{C}_{y_s=i,y^*=j}}{\sum_{j \in m} \mathbf{C}_{y_s=i,y^*=j}} \cdot |\mathbf{X}_{y_s=i}|}{\sum_{i \in m, j \in m} \frac{\mathbf{C}_{y_s=i,y^*=j}}{\sum_{j \in m} \mathbf{C}_{y_s=i,y^*=j}} \cdot |\mathbf{X}_{y_s=i}|} \tag{6}$$

The mislabelled pixels in y_s are selected from each off-diagonal entry in $\mathbf{Q}_{y_s,y^*}$. Rather than selecting all pixels, we only consider those with a large margin. Specifically, we first compute the margin $\hat{p}(y = j; x) - \hat{p}(y = i; x)$ of each pixel in $\mathbf{Q}_{y_s=i,y^*=j,i \neq j}$, and the top $n \cdot \mathbf{Q}_{y_s=i,y^*=j}$ (n is the total number of pixels in an image) pixels of this difference between $\hat{p}(y = j; x)$ and $\hat{p}(y = i; x)$ are selected as mislabelled pixels in each entry. The revised ground truth y_{re} is then generated by flipping these mislabelled pixels in y_s, as our segmentation problem targets binary labels.

4. Experiments

This section presents the description of datasets and our experimental results and discussion.

4.1. Datasets and Evaluation Metrics

Our proposed method is evaluated on three public datasets of histopathology images for cell nuclei segmentation, including MoNuSeg (Kumar et al., 2017), TNBC (Naylor et al., 2019), and CPM-17 (Vu et al., 2019). These datasets all provide pixel-wise ground truth annotations, and hence the point annotations can be generated automatically and evaluation of segmentation performance on full masks is possible. MoNuSeg contains 30 images with a total of $21,623$ nuclei from 7 different organs, and each image is of $1,000 \times 1,000$ pixels. TNBC contains 50

Datasets	Number of image tiles	Number of nuclei	Number of organs	Size
MoNuSeg	30	21623	7	1000*1000
TNBC	50	4022	1	512*512
CPM-17	32	7570	4	500*500 to 600*600

Table 1. Summary of datasets.

Supervision	Method	MoNuSeg		TNBC	
		IoU	Dice	IoU	Dice
Scribble annotations	(Lee and Jeong, 2020)	0.6408	0.5811	-	-
Point annotations	(Kumar et al., 2017)	0.5710(±0.02)	-	0.5504(±0.04)	-
	(Qu et al., 2019a)	0.5789(±0.06)	0.7320(±0.05)	0.5420(±0.04)	0.7008(±0.04)
	(Yoo et al., 2019)	0.6136(±0.04)	-	0.6038(±0.03)	-
	(Tian et al., 2020)	0.6239(±0.03)	0.7638(±0.02)	**0.6393**(±**0.03**)	0.7510(±0.04)
	(Guo et al., 2021)	0.6401(±0.03)	0.7662(±0.04)	0.6234(±0.02)	0.7582(±0.03)
	Ours	**0.6515**(±**0.04**)	**0.7942**(±**0.04**)	0.6367(±0.03)	**0.7764**(±**0.03**)
Pixel-wise annotations	(Yoo et al., 2019)	0.6522(±0.03)	-	0.6619(±0.04)	-
	(Tian et al., 2020)	0.6494(±0.04)	0.7859(±0.02)	**0.6950**(±**0.03**)	**0.8022** ± (**0.03**)
	(Lee and Jeong, 2020)	0.7014	0.6677	-	-
	(Xiang et al., 2020)	**0.704**	**0.825**	0.651	0.780
	(Guo et al., 2021)	0.6723(±0.05)	0.7775(±0.04)	0.6537(±0.03)	0.7718(±0.04)
	Ours	0.6833(±0.03)	0.8096(±0.02)	0.6642(±0.04)	0.7957(±0.03)

Table 2. Segmentation results compared with other state-of-the-art approaches.

images with a total of 4,022 nuclei from 11 different patients with an image resolution of 512×512 pixels. CPM-17 contains 32 images with 7,570 nuclei and each image has 500×500 pixels.

We implemented the models using Pytorch, and the training was conducted on NVIDIA Tesla P100 GPUs. We used the Adam optimiser with a learning rate of 0.001 and a weight decay of 0.0005. The learning rate is halved when $\mathcal{L}_{total}$ on the validation set does not decrease for 4 epochs. The input images are cropped to patches of 512×512 pixels for MoNoSeg and input images of CPM-17 are resized to 512×512 pixels. The training process includes 60 epochs and each epoch trains a batch set of 2. Data augmentations include Gaussian blur, adjustment of hue and saturation, affine transformation, horizontal and vertical flips. Bilinear upsampling is utilised for the upsampling operation. The hyperparameters α, β, γ and δ are set to 0.4, 0.8, 1 and 0.4 respectively.

Overall, the confident leaning implements only one round to fix pseudo-labels and we only retrain Stage1 once. Specifically, the network is trained in an end-to-end segmentation process in the first stage. Then, we use confident learning to refine the superpixels. In the second stage, we used the same network to train the network with the refined superpixels and produce the segmentation results.

We use two common evaluation metrics in all experiments: Intersection over Union (IoU) and pixel-level Dice coefficient (Yoo et al., 2019; Tian et al., 2020). The metrics are defined as follows:

$$IoU = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{TP} + \mathrm{FN}} \tag{7}$$

$$Dice = \frac{2\mathrm{TP}}{\mathrm{FP} + 2\mathrm{TP} + \mathrm{FN}} \tag{8}$$

where TP, FP and FN represent the amount of true positive predictions (true nuclear pixels), false positive predictions and false negative predictions. Note that similarly to existing studies (Woo et al., 2018; Yoo et al., 2019; Natarajan et al., 2020; Liang et al., 2022), our method is designed for semantic segmentation, and hence we use IoU and Dice for evaluation. We follow (Yoo et al., 2019) to conduct 10-fold cross-validation. Specifically, each dataset is divided into 10 equal folds, which are split into training, validation and test sets with an 8:1:1 ratio. The splitting is repeated 10 times and we report the mean scores on the test set. The validation and test sets have no intersection across iterations.

4.2. Comparison with Existing Methods

Comparison with state-of-the-art approaches is conducted on the MoNuSeg and TNBC datasets because existing weakly-supervised nuclei segmentation studies have commonly used these datasets. As shown in Table 2, we compared with five methods that conducted weakly-supervised nuclei segmentation based on point annotations, one based on scribble annotations and another five fully-supervised methods to evaluate the quantitative results.

The settings of different approaches are similar. In (Lee and Jeong, 2020), (Kumar et al., 2017), (Qu et al., 2019a), (Yoo et al., 2019), (Tian et al., 2020) and (Guo et al., 2021), ten-fold cross validation has been employed. However, the training, validation and testing split ratios in (Lee and Jeong, 2020), (Tian

Point location	Method	MoNuSeg(r=0)		TNBC(r=0)		CPM-17(r=0)	
		IoU	Dice	IoU	Dice	IoU	Dice
Centroid (r=0)	Seg	0.5580(±0.07)	0.6921(±0.04)	0.5329(±0.05)	0.65782(±0.03)	0.4636(±0.07)	0.6143(±0.07)
	Seg+CL	0.5919(±0.06)	0.7507(±0.05)	0.5585(±0.04)	0.6986(±0.04)	0.5663(±0.04)	0.7141(±0.04)
	Seg+Att	0.6222(±0.05)	0.7515(±0.02)	0.5788(±0.04)	0.7208(±0.03)	0.626(±0.05)	0.7657(±0.04)
	Seg+Att+CL	0.6401(±0.03)	0.7662(±0.04)	0.6234(±0.02)	0.7582(±0.03)	0.6761(±0.04)	0.7929(±0.03)
	SAC-Net w/o DConv and SA	0.6388(±0.03)	0.7846(±0.02)	0.6039(±0.04)	0.7475(±0.04)	0.6748(±0.05)	0.8008(±0.04)
	SAC-Net	**0.6515(±0.04)**	**0.7942(±0.04)**	**0.6367(±0.03)**	**0.7764(±0.03)**	**0.6910(±0.02)**	**0.8150(±0.01)**
	(Yoo et al., 2019)	0.6173(±0.06)	0.7586(±0.05)	0.5907(±0.08)	0.7454(±0.05)	0.6511(±0.06)	0.7813(±0.05)
Point location	Method	MoNuSeg(r=5)		TNBC(r=3)		CPM-17(r=3)	
		IoU	Dice	IoU	Dice	IoU	Dice
Shift (r=3/5)	Seg	0.5437(±0.09)	0.6920(±0.03)	0.5196(±0.08)	0.6563(±0.05)	0.4459(±0.04)	0.6061(±0.04)
	Seg+CL	0.5897(±0.08)	0.7359(±0.03)	0.5531(±0.03)	0.7096(±0.05)	0.5462(±0.06)	0.6948(±0.05)
	Seg+Att	0.6296(±0.07)	0.7486(±0.01)	0.5776(±0.04)	0.7130(±0.05)	0.6194(±0.04)	0.7603(±0.03)
	Seg+Att+CL	0.6324(±0.04)	0.7551(±0.02)	0.6153(±0.04)	0.7338(±0.05)	0.6621(±0.03)	0.7935(±0.02)
	SAC-Net w/o DConv and SA	0.6237(±0.03)	0.7543(±0.02)	0.6008(±0.03)	0.7444(±0.03)	0.6501(±0.05)	0.7838(±0.03)
	SAC-Net	**0.6369(±0.06)**	**0.7673(±0.02)**	**0.6197(±0.03)**	**0.7639(±0.02)**	**0.6653(±0.05)**	**0.7945(±0.04)**
	(Yoo et al., 2019)	0.6011(±0.08)	0.7607(±0.03)	0.5912(±0.05)	0.7236(±0.04)	0.6110(±0.06)	0.7478(±0.05)

Table 3. Ablation study of our framework components and comparison of different models. Here $r = 0$ means point annotations based on centroids and $r = 3/5$ indicates noisy point annotations with random shift.

Methods	MoNuSeg		TNBC		CPM-17	
	IoU	Dice	IoU	Dice	IoU	Dice
SAC-Net w/o DConv and SA	0.6388(±0.03)	0.7846(±0.02)	0.6039(±0.04)	0.7475(±0.04)	0.6748(±0.05)	0.8008(±0.04)
SAC-Net w/o DConv	0.6465(±0.03)	0.7921(±0.03)	0.6154(±0.05)	0.7582(±0.04)	0.6826(±0.06)	0.8085(±0.05)
SAC-Net w/o SA	0.6476(±0.04)	0.7920(±0.03)	0.6242(±0.03)	0.7643(±0.03)	0.6860(±0.04)	0.8101(±0.03)
SAC-Net	**0.6515(±0.04)**	**0.7942(±0.04)**	**0.6367(±0.03)**	**0.7764(±0.03)**	**0.6910(±0.02)**	**0.8150(±0.01)**

Table 4. Ablation study of different components (DConv: deformable convolution; SA: spatial attention) in SegNet and AttNet.

et al., 2020) and (Qu et al., 2019a) are slightly different from ours. It should be noted that results from (Yoo et al., 2019) are obtained by running the provided code, while results for (Kumar et al., 2017) are obtained from (Yoo et al., 2019) since (Kumar et al., 2017) was not originally designed for weakly-supervised segmentation. Results for (Qu et al., 2019a) are extracted from (Tian et al., 2020) since the latter includes experiments on both MoNuSeg and TNBC.

Table 2 shows the IoU and Dice results of segmentation on MoNuSeg and TNBC by the various approaches. We first compare our method with weakly-supervised learning methods. Qu et al. (2019a) use the Voronoi diagram and k-means clustering to generate pseudo labels. However, it adopts the simple U-Net structure for segmentation, which results in poorer results than other methods. Yoo et al. (2019) combines edge information in their network to enhance the results of nuclei segmentation. The edge information was obtained by applying the Sobel filtering on segmentation results, whereas in our model, we introduce the ConsNet sub-network to incorporate low-level edge information into the segmentation network rather than focusing on boundaries. Tian et al. (2020) supervised model training by iterative optimisation of the generated distance map. The main issue with this approach is that it requires more iterative optimisation. When there is significant noise in the data, it is difficult to converge to better performance. Our model utilises AttNet to reduce the noisy label influence when training the segmentation network and the CL refinement module can reduce noisy labels in Stage II to enhance our performance. Our previous model (Guo et al., 2021) has a similar two-stage design, which however does not include ConsNet and the network architectures for SegNet and AttNet were not optimised. In addition, we compare with a scribble annotation-based approach (Lee and Jeong, 2020), which was applied to MoNuSeg. This method achieved a better IoU evaluation result but a much lower Dice score than our method and other weakly-supervised approaches. We argue this is because scribble labels are better suited to circle-shaped cell segmentation, but the cells have different shapes which is hard for scribble annotations to achieve better performance on the pixel-wise Dice score.

We also conducted fully supervised learning using our model with pixel-wise annotations and compare with other fully-supervised segmentation models. We train our model by replacing the superpixel labels with original pixel-wise labels for SegNet without Stage II. The highest score on MoNuSeg is more than 1.5% of IoU and 2% of Dice in Xiang et al. (2020) compared to our results. Xiang et al. (2020) utilises paired forward and backward skip connections between the encoder and decoder to further enhance feature learning, while our network has not incorporated such optimisations. On MoNuSeg, our IoU result is lower, but our Dice score is higher than Lee and Jeong (2020). They only use the U-Net with the ResNet-34 for the encoder, which is suitable for fully supervising the small dataset MoNuSeg. In contrast, our approach contains a more complex structure that may cause lower IoU when working on a small dataset. However, our network is competitive with noisy detection around the cell boundaries, which increases the Dice re-

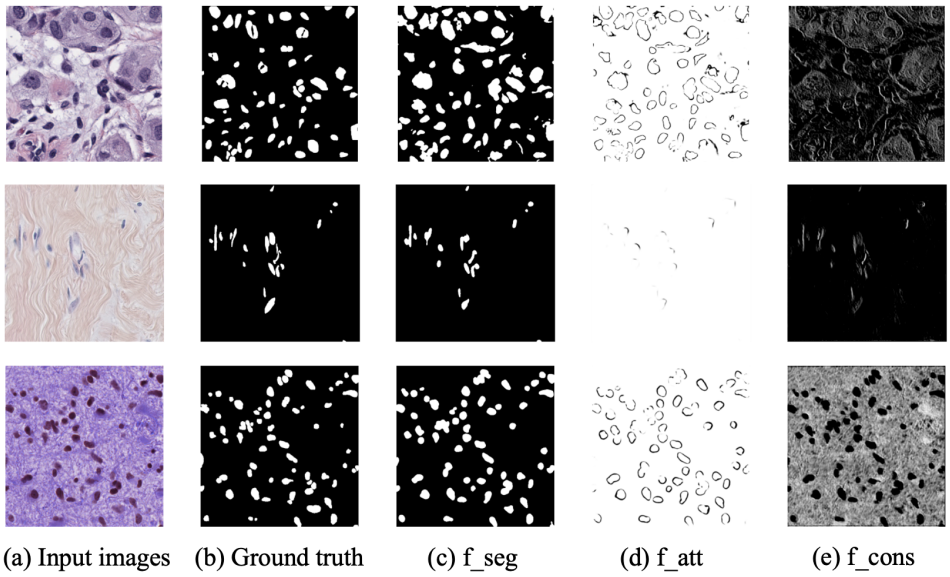


Fig. 7. Visualisation of input image, ground truth, f_{seg}, f_{att}, f_{cons} on MoNuSeg, TNBC and CPM-17 datasets.

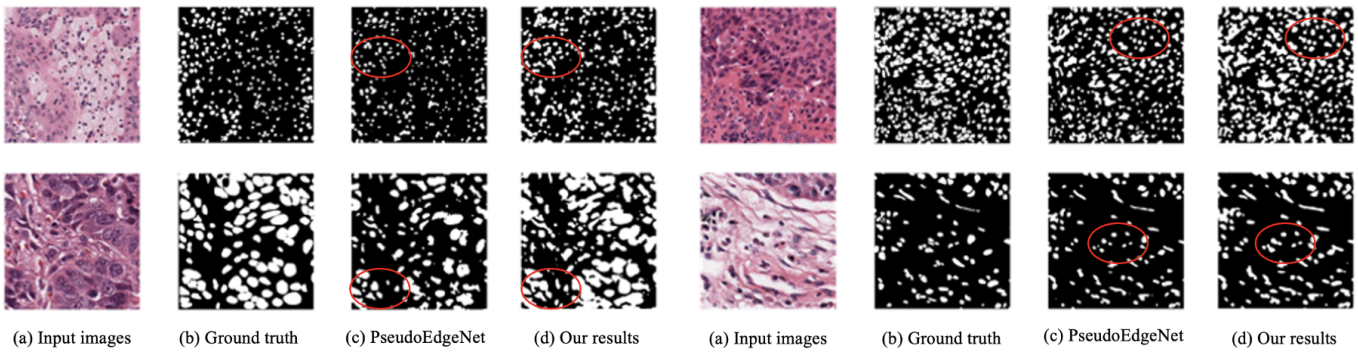


Fig. 8. Visualisation of input images, ground truth, outputs from PseudoEdgeNet and our final outputs on MoNuSeg. The red circles highlight some regions where our results are better than PseudoEdgeNet.

sults compared to Lee and Jeong (2020). On TNBC, Tian et al. (2020) achieves higher performance than our method; however, our method obtained a very close Dice score on TNBC and much improved results on MoNuSeg. We argue that the MoNuSeg dataset contains many overlapping cells, making it difficult for fully convolutional networks to delineate the contours, whereas our network includes a more complex architecture to better analyse cluttering cells. However, the TNBC contains fewer cells and overlapping regions, and hence it does not well demonstrate the advantage of our network model.

In summary, nuclei segmentation is complex since the datasets typically contain highly heterogeneous images obtained from multiple organs and the cells are crowded with unclear boundaries from the background. Our method addresses the problem of pseudo labels around the contours containing a high degree of noise by SAC-Net. In addition, our approach further corrects mislabelling during model training using the CL module. Consequently, our proposed architecture yielded superior results compared to most state-of-the-art weakly supervised learning methods.

4.3. Ablation Studies

To further evaluate our method, we performed the following ablation experiments: (i) we evaluate the generalisability of our method by using one more dataset, CPM-17; (ii) considering that it might not be easy to annotate the centroids of nuclei as the ground truth precisely, we simulate noisy point supervision; and, (iii) we evaluate the effectiveness of different components including SegNet, AttNet, ConsNet and the CL refinement module. In particular, the noisy point annotations are obtained by randomly shifting the centroid points in a circle within a radius $r = 5$ for MoNuSeg and $r = 3$ for datasets TNBC and CPM-17, with the radius selected based on the image sizes. A larger radius would sometimes render the centre point outside of the cell nuclei. This experiment is meant to demonstrate that, when there is an offset in the centre point, the effect does not much affect the segmentation results.

Table 3 shows the results of the ablation studies. We also include comparison with PseudoEdgeNet (Yoo et al., 2019), since we could use the provided code to run the experiments. It can be seen that by introducing the various components (AttNet, ConsNet and CL), various performance improvements have been

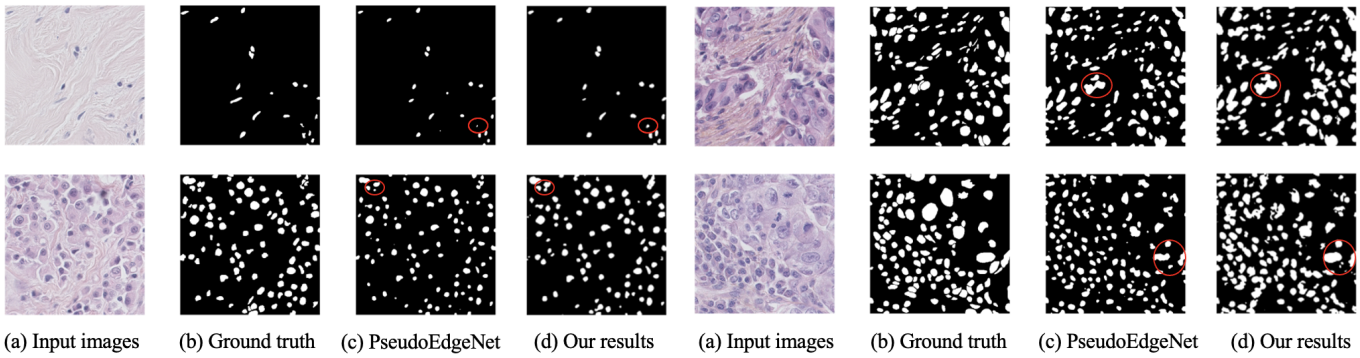


Fig. 9. Visualisation of input images, ground truth, outputs from PseudoEdgeNet and our final outputs on TNBC. The red circles highlight some regions where our results are better than PseudoEdgeNet.

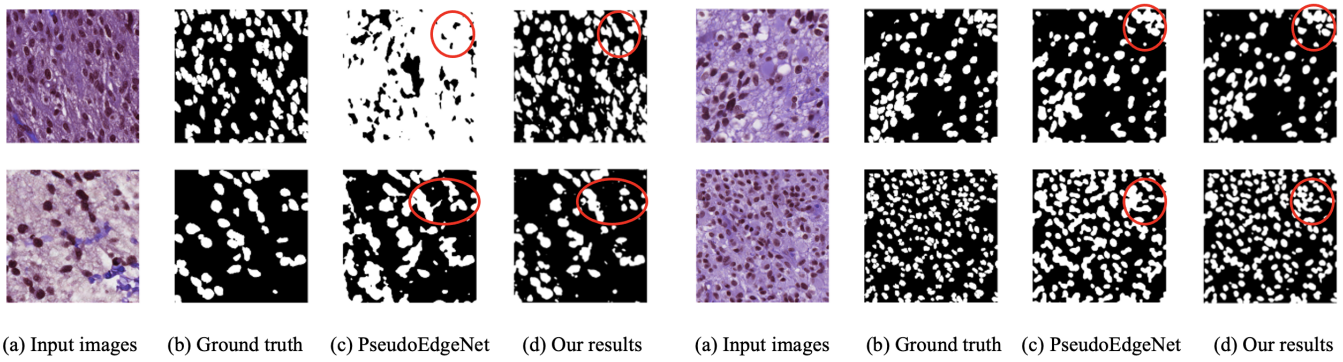


Fig. 10. Visualisation of input images, ground truth, outputs from PseudoEdgeNet and our final outputs on CPM-17. The red circles highlight some regions where our results are better than PseudoEdgeNet.

achieved and the results are consistent on all three datasets. Our method also shows good robustness with a small performance drop when noisy point annotations are used. In addition, both our SAC-Net model and the model without ConsNet (Seg+Att+CL) outperforms PseudoEdgeNet, indicating the advantage of incorporating label refinement via the attention mechanism and CL. The performance of SAC-Net without deformable convolution (DConv) and spatial attention (SA) is higher than Seg+Att in our experiments. The network structure difference between Seg+Att and SAC-Net w/o DConv and SA is only a ConsNet. This demonstrates that even without DConv and SA, the additional ConsNet can introduce additional improvement over Seg+Att, while SegNet gained improvement on nuclei detection after fusing with the intermediate feature maps. On the other hand, during our experiments, we find that the results before and after applying CL on SAC-Net without DConv and SA do not change a lot and the refined superpixels do not truly reflect the ground truth of nuclei. The performance after applying CL is improved after we incorporate DConv and SA into the network structure. We suggest this is because the DConv and SA highly enrich the meaningful spatial features in the output probability maps that are used as input of CL to refine superpixel labels.

In addition, the SAC-Net without Deformable convolution and special attention is compared with SegNet, SegNet with CL, SegNet with AttNet and SegNet, and AttNet with CL. Also,

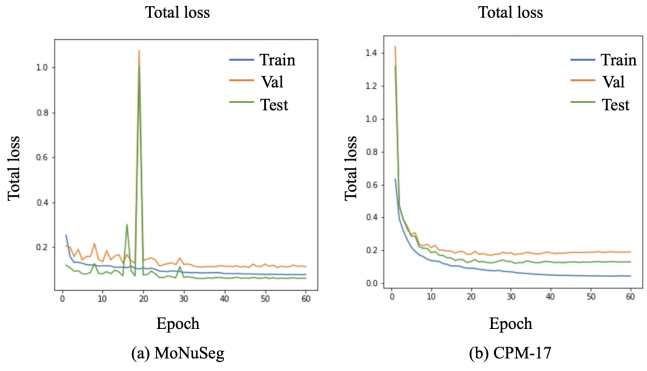


Fig. 11. Loss converge curves in MoNuSeg and CPM-17.

it compared with the SAC-Net. We can see from Table 3 that the SAC-Net without Dconv and SP are competitive with SegNet, SegNet + CL or SegNet + AttNet. Since the CL function can refine superpixel, the performance is slightly lower than SegNet, AttNet and CL but very close.

Our experimental results can demonstrate that the retraining process can converge since the additional of CL brings additional performance improvement. Fig. 11 shows that the loss during training becomes stable after 30 epochs. We can see that the total loss is higher and uncertain before the 20 epochs. Between 20 and 30 epochs, it still fluctuates a little while tending

to stabilise. Thus, we suggest that our two stages training procedure is able to converge.

We also compare the results of the different blocks in SAC-Net, including the deformable convolution layers and spatial attention blocks. As mentioned in Section 3.2.1, the network structure contains deformable convolution layers in each encoder layer and spatial attention blocks in decoder blocks. Table 4 shows that SAC-Net with both deformable convolution layer and spatial attention blocks outperforms other configurations with single components in SAC-Net. It proves that adopting deformable convolution layers and spatial attention blocks can improve our framework to some extent.

For qualitative comparison, Fig. 7 shows the segmentation results on example images from the three datasets using SAC-Net. It shows the original images, ground truth, the f_{seg}, the f_{att} and f_{cons} in order. f_{seg} is the final segmentation output. f_{att} contains the noisy pixels of the black regions, whereas the white regions denote the background and a small central part of nuclei. f_{cons} is the low-level feature of input images, visualising the edge information extraction, which is more accurate when using the shallow convolution layers. In Figs. 8, 9 and 10, we compare our results of final outputs with outputs from PseudoEdgeNet (Yoo et al., 2019) using its provided code. It can be seen that PseudoEdgeNet has more issues when the cells are overlapping and crowded. Our SAC-Net can better handle such cases by extracting the low-level features to delineate the actual cell shapes but also keep the edge information of nuclei during the network training.

5. Conclusions

In this paper, we present a weakly supervised method, SAC-Net, to segment cell nuclei in histopathology images using point annotations. The designed network adopts two stages: first, segmentation networks combined with an attention mechanism and low-level feature constraints are trained based on point/Voronoi and superpixel labels; and, the second stage applies the confident learning algorithm to refine the segmentation outputs. In our design, we consider the pseudo point/Voronoi and superpixel labels as noisy labels and incorporate various ways to address this issue: an attention network in the first stage to focus the training on more confident regions, a constraint network for highlighting the low-level structural information and confident learning in the second stage to explicitly revise the labels. Our experimental results on three histopathology image datasets demonstrate that our method achieved the best results compared to other weakly supervised methods. In further work, we will implement a better way to generate the centroid of the points from the nucleus, making it not rely on the traditional mathematics method. Moreover, the findings from the shifted centroids still existed the deviation of the results. Thus, it is suggested that existing weakly supervised segmentation methods in histopathology images need further improvement in order to work with non-centroid points. Therefore, the researchers do not need the labelled images to obtain the centroid points to predict the segmentation results. The acquisitive points from cells are still a difficult challenge for further work, and we will work on this in the following studies.

We will investigate how to improve the segmentation performance when the annotated points are not precisely at the nuclear centroids. Based on this, we will further investigate a solution to generate the point annotations from unlabelled images, making the method unsupervised without relying on simulation of weakly-supervised learning from pixel-wise ground truth annotations. The structure of our model might change to unsupervised learning or semi-supervised learning for detecting the centroid points automatically. In addition, we may use some popular structure like Transformers to improve our network.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 34, 2274–2282.

Akbaş, C.E., Kozubek, M., 2020. Condensed U-Net (Cu-Net): An improved U-Net architecture for cell segmentation powered by 4x4 max-pooling layers, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 446–450.

Aurenhammer, F., 1991. Voronoi diagrams—a survey of a fundamental geometric data structure. ACM Computing Surveys (CSUR) 23, 345–405.

Chamanzar, A., Nie, Y., 2020. Weakly supervised multi-task learning for cell detection and segmentation, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 513–516.

Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.A., 2017. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Medical image analysis 36, 135–146.

Chen, Z., Chen, Z., Liu, J., Zheng, Q., Zhu, Y., Zuo, Y., Wang, Z., Guan, X., Wang, Y., Li, Y., 2020. Weakly supervised histopathology image segmentation with sparse point annotations. IEEE Journal of Biomedical and Health Informatics 25, 1673–1685.

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., Wei, Y., 2017. Deformable convolutional networks, in: Proceedings of the IEEE international conference on computer vision, pp. 764–773.

Dong, M., Liu, D., Xiong, Z., Chen, X., Zhang, Y., Zha, Z.J., Bi, G., Wu, F., 2020. Towards neuron segmentation from macaque brain images: A weakly supervised approach, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 194–203.

Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.A., Snead, D., Tsang, Y.W., Rajpoot, N., 2019a. MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. Medical image analysis 52, 199–211.

Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019b. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis 58, 101563.

Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019c. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis 58, 101563.

Guo, R., Pagnucco, M., Song, Y., 2021. Learning with noise: Mask-guided attention model for weakly supervised nuclei segmentation, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 461–470.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN, in: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Hsu, C.C., Hsu, K.J., Tsai, C.C., Lin, Y.Y., Chuang, Y.Y., 2019. Weakly supervised instance segmentation using the bounding box tightness prior. Advances in Neural Information Processing Systems 32, 6586–6597.

Jeong, J., Lee, S., Kim, J., Kwak, N., 2019. Consistency-based semi-supervised learning for object detection. Advances in neural information processing systems 32, 10759–10768.

Kablan, E.B., Dogan, H., Ercin, M.E., Ersoz, S., Ekinci, M., 2020. An ensemble of fine-tuned fully convolutional neural networks for pleural effusion cell nuclei segmentation. Computers & Electrical Engineering 81, 106533.

Kang, Q., Lao, Q., Fevens, T., 2019. Nuclei segmentation in histopathological images using two-stage learning, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 703–711.

Kervadec, H., Dolz, J., Tang, M., Granger, E., Boykov, Y., Ben Ayed, I., 2019. Constrained-cnn losses for weakly supervised segmentation. Medical Image Analysis 54, 88–99.

Koohbanani, N.A., Jahanifar, M., Tajadin, N.Z., Rajpoot, N., 2020. Nuclick: A deep learning framework for interactive segmentation of microscopic images. Medical Image Analysis 65, 101771.

Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A., 2017. A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Transactions on Medical Imaging (TMI) 36, 1550–1560.

Laradji, I., Rodriguez, P., Manas, O., Lensink, K., Law, M., Kurzman, L., Parker, W., Vazquez, D., Nowrouzezahrai, D., 2021. A weakly supervised consistency-based learning method for covid-19 segmentation in CT images, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 2453–2462.

Laradji, I., Rostamzadeh, N., Pinheiro, P., Vázquez, D., Schmidt, M., 2018. Where are the blobs: Counting by localization with point supervision. European Conference on Computer Vision (ECCV) 11206, 560–576.

Lee, H., Jeong, W.K., 2020. Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency, in: Medical Image Computing and Computer Assisted Intervention (MICCAI). Springer International Publishing, pp. 14–23.

Li, K., Wu, Z., Peng, K.C., Ernst, J., Fu, Y., 2018. Tell me where to look: Guided attention inference network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Liang, H., Cheng, Z., Zhong, H., Qu, A., Chen, L., 2022. A region-based convolutional network for nuclei detection and segmentation in microscopy images. Biomedical Signal Processing and Control 71, 103276.

Mahmood, F., Borders, D., Chen, R.J., Mckay, G.N., Salimian, K.J., Baras, A., Durr, N.J., 2020. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. IEEE transactions on medical imaging 39, 3257–3267.

Maninis, K.K., Caelles, S., Pont-Tuset, J., Van Gool, L., 2018. Deep extreme cut: From extreme points to object segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 616–625.

Mirikharaji, Z., Yan, Y., Hamarneh, G., 2019. Learning to segment skin lesions from noisy annotations, in: Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data, pp. 207–215.

Natarajan, V.A., Kumar, M.S., Patan, R., Kallam, S., Mohamed, M.Y.N., 2020. Segmentation of nuclei in histopathology images using fully convolutional deep neural architecture, in: 2020 International Conference on computing and information technology (ICCIT-1441), IEEE. pp. 1–7.

Naylor, P., Laé, M., Reyal, F., Walter, T., 2019. Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE Transactions on Medical Imaging (TMI) 38, 448–459.

Nishimura, K., Ker, D.F.E., Bise, R., 2019. Weakly supervised cell instance segmentation by propagating from detection response, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 649–657.

Northcutt, C.G., Jiang, L., Chuang, I.L., 2019. Confident learning: Estimating uncertainty in dataset labels. `arXiv:1911.00068`. arXiv preprint arXiv:1911.00068.

Oh, Y., Kim, B., Ham, B., 2021. Background-aware pooling and noise-aware loss for weakly-supervised semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6913–6922.

Qu, H., Wu, P., Huang, Q., Yi, J., Riedlinger, G.M., De, S., Metaxas, D.N., 2019a. Weakly supervised deep nuclei segmentation using points annotation in histopathology images. Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning , 390–400.

Qu, H., Wu, P., Huang, Q., Yi, J., Yan, Z., Li, K., Riedlinger, G.M., De, S., Zhang, S., Metaxas, D.N., 2020a. Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. IEEE transactions on medical imaging 39, 3655–3666.

Qu, H., Yan, Z., Riedlinger, G.M., De, S., Metaxas, D.N., 2019b. Improving nuclei/gland instance segmentation in histopathology images by full resolution neural network and spatial constrained loss , 378–386.

Qu, H., Yi, J., Huang, Q., Wu, P., Metaxas, D., 2020b. Nuclei segmentation using mixed points and masks selected from uncertainty, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 973–976.

Reinhard, E., Adhikhmin, M., Gooch, B., Shirley, P., 2001. Color transfer between images. IEEE Computer Graphics and Applications 21, 34–41.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

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.

Tian, K., Zhang, J., Shen, H., Yan, K., Dong, P., Yao, J., Che, S., Luo, P., Han, X., 2020. Weakly-supervised nucleus segmentation based on point annotations: A coarse-to-fine self-stimulated learning strategy, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 299–308.

Vu, Q.D., Graham, S., Kurc, T., To, M.N.N., Shaban, M., Qaiser, T., Koohbanani, N.A., Khurram, S.A., Kalpathy-Cramer, J., Zhao, T., et al., 2019. Methods for segmentation and classification of digital microscopy tissue images. Frontiers in bioengineering and biotechnology 7, 53.

Vuola, A.O., Akram, S.U., Kannala, J., 2019. Mask-RCNN and U-Net ensembled for nuclei segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 208–212.

Wang, J., Zhou, S., Fang, C., Wang, L., Wang, J., 2020. Meta corrupted pixels mining for medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 335–345.

Woo, S., Park, J., Lee, J., Kweon, I.S., 2018. CBAM: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), pp. 3–19.

Xiang, T., Zhang, C., Liu, D., Song, Y., Huang, H., Cai, W., 2020. BiO-Net: Learning recurrent bi-directional connections for encoder-decoder architecture, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 74–84.

Xue, C., Dou, Q., Shi, X., Chen, H., Heng, P., 2019. Robust learning at noisy labeled medical images: Applied to skin lesion classification, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1280–1283.

Yoo, I., Yoo, D., Paeng, K., 2019. Pseudoedgenet: Nuclei segmentation only with point annotations. Medical Image Computing and Computer Assisted Intervention – MICCAI 11764, 731–739.

Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122 .

Zhang, J., Yu, X., Li, A., Song, P., Liu, B., Dai, Y., 2020. Weakly-supervised salient object detection via scribble annotations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

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. Medical Image Analysis 65, 101786.

Zhao, X., Liang, S., Wei, Y., 2018. Pseudo mask augmented object detection, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4061–4070.

Zhou, Y., Onder, O., Dou, Q., Tsougenis, E., Chen, H., Heng, P., 2019. CIA-Net: Robust nuclei instance segmentation with contour-aware information aggregation, in: Information Processing in Medical Imaging (IPMI), pp. 682–693.

Zhu, H., Shi, J., Wu, J., 2019. Pick-and-learn: Automatic quality evaluation for noisy-labeled image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 576–584.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networkss, in: Computer Vision (ICCV), 2017 IEEE International Conference on.