Weakly supervised gland instance segmentation based on point labeling

Da-Han Wang*, Haili Ye*, Jianmin Li*, Si Chen*, Chenyan Zhu, Shunzhi Zhu*
*Fujian Key Laboratory of Pattern Recognition and Image Understanding, School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China
2 Motic (Xiamen) Medical Diagnostic Systems Co. Ltd. Xiamen, China

***Abstract*—Colon cancer is a common digestive tract cancer occurring in the colon. How to effectively segment the glands in the pathological image of the colon is a difficult problem. As well as distinguishing the glands from the background, it is necessary to distinguish the boundaries between the different instances of the glands. Common instance segmentation methods require pixel-level instance annotation, which takes a lot of time and resources. The pixel-level weakly supervised instance segmentation method cannot provide sufficient supervision, which makes the quality of gland segmentation poor. In this paper, we propose a weakly supervised gland segmentation method based on point labeling. We trained a glandular point detection model to predict high confidence points in gland images using the supervision information of point labeling. Then, the high-confidence point-assisted training instance segmentation model is used to implement the instance segmentation of the glands. We tested our method on the 2015 MICCAI Gland Challenge dataset, and the experimental results show that our method can effectively segment the instance of the Gland, and its performance is even better than that part of the method by training using the fully supervised.**

Keywords—Weak supervision; gland instance segmentation; style; histology image analysis; Medical image analysis

I. INTRODUCTION

Colon cancer is a common digestive tract cancer occurring in the colon, and it is one of the most common cancers with the highest incidence rate of cancer [1-5]. Under the microscope, tumor cells and tissues show certain structural characteristics different from normal cells and tissues, also known as histopathological characteristics. Doctors grade cancer according to the histopathological characteristics, and then determine the cancer situation and treatment plan of patients. It is a key step for pathologists to segment gland cases accurately from pathological images to quantitatively analyze the malignant degree of adenocarcinoma for further diagnosis. However, manual segmentation of gland cases from pathological images is a very time-consuming work. The histopathological images were stained with hematoxylin and eosin (H & E), which made it more difficult to segment the gland cases. Pathological whole section scanning equipment can digitize pathological images. Computer-Aided Diagnosis (CAD) of digital pathological images is one of the research hotspots in the field of medical image analysis [6-8]. Nevertheless, manual annotation of gland instances is time consuming given the large size of a histology image and a large number of glands in an image. Not only that, the malignant and benign glands show different image features, which makes the traditional segmentation methods have great limitations. Therefore, accurate and automatic methods for gland instance segmentation are in great demand.

At present, with the development of deep learning and computer vision, the technology of pathological image intelligent analysis is constantly improving. In recent years, most methods use semantic segmentation to realize gland segmentation [16-19], but these methods can't separate the overlapping glands. Therefore, more and more researchers use the instance segmentation method to deal with the problem of gland segmentation [20-23]. (As shown in Figure 1, segmentation of gland image (Fig 1.a) into gland instance mask (Fig 1.b)). The current image instance segmentation model [11-14] mainly uses a multi task composite structure, which divides the image instance segmentation problem into two steps: 1) gland detection which separates glands from the background and 2) boundary refined segmentation so as to identify each gland individually. This method uses pixel level gland instance annotation for training, which requires manual annotation of gland instance mask and bounding box parameters. At the same time, there are high requirements for the professional quality of taggers. For the annotation of gland instance boundary, there are some errors in the process of annotation because of the complexity of gland boundary contour.

Weak supervised instance segmentation algorithm is an effective method to reduce the workload of data annotation in instance segmentation, but the current weak supervised instance segmentation algorithm needs to use a high-precision classification model for gradient back propagation to get the calorific value location between instances [24-28]. Most of the time, gland image contains both the gland and the background, which makes training classification network meaningless. This is because classification model cannot effectively distinguish the difference between gland instance and background. These limitations cause the current weak supervised instance segmentation algorithm cannot achieve the segmentation of glands.

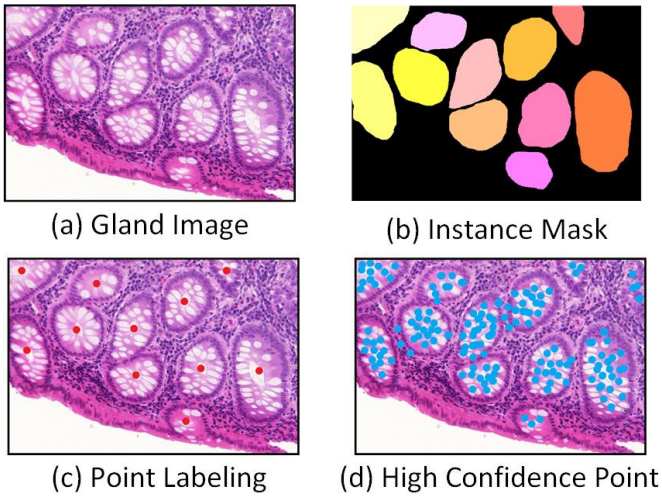

Fig. 1. *Gland instance segmentation related data, a. Gland image, b. Instance mask, c. Point labeling, d. High confidence point.*

Recently, the Amy's [34] uses point annotation to segment instances in a weakly supervised manner, that is, only one reference point is labeled for each instance in the image, which can effectively reduce the cost of image annotation, and the effect is not bad. Inspired by this paper, we propose a weak supervised gland instance segmentation method based on point annotation. We train gland point detection model by point labeling (Figure 1. C), generate high confidence point (Figure 1. D), and use high confidence point to assist in training gland instance segmentation model. We reuse the feature extraction part of the point detection model to locate the calorific value of the gland instance in the image, which solves the problem that the gland image cannot train the classification model to locate the calorific value. The experimental results show that this method achieves the segmentation of gland instance under the weak supervision of only using point annotation, and the performance of the method is better than that of partial full supervision gland segmentation method. The main contributions of this paper are as follows:

1) Based on the point annotation, the weak supervised instance segmentation of gland image is realized.

2) By training gland point detection model with point annotation, the problem that gland image cannot be pre trained to classify model for calorific value location is solved.

3) The segmentation model of gland instance is trained by high confidence points.

4) The method proposed in this paper has high scalability and can be applied to other small datasets except glands.

II. Related Works

A. Instance Segmentation

Instance segmentation is the combination of object detection and semantic segmentation. It is necessary to segment each instance from pixel level while locating and classifying the instances in the image through object detection. In 2017, He et al [11] proposed a two-stage instance segmentation model. In the first stage, the proposed region is generated by scanning the image to locate the object. The second stage classifies the proposed area and generates a boundary frame and mask. Based on the accurate location and recognition of the object, this method can segment the instance area, which has high accuracy. Compared with the two-stage model, Bolya et al [15] proposed a one-stage instance segmentation model, which divides the instance segmentation into two parallel sub tasks: One branch segmented the original mask of the instance; the other branch predicts the mask coefficient corresponding to each instance. In this method, the original method of instance segmentation is changed from serial to parallel, which improves the calculation efficiency.

Comparatively, Gland image are more complex and changeable than scene image due to many overlaps between the glands and low differences in gland boundary and background area. As a result, Common instance segmentation models are often under performed. Chen et al [19], proposed a new depth profile sensing network. In order to separate the adjacent glands, two branches are used to up-sample and fuse the images feature of different levels, and the location information and contour mask of the segmented glands are output respectively. The network structure consists of two parts: the down-sampling path and the up-sampling path, which effectively combines the context information of features. Xu et al [18] proposed a multi-channel depth neural network to extract gland area, boundary and location information. The final segmentation result is obtained by fusing the results of different channels. However, most of the current gland instance segmentation methods need pixel level annotation information for full supervision training.

B. Weak supervised instance segmentation

A key aspect of the Instance segmentation models is their dependency on large amount of manual annotation data. For Gland instance segmentation, a critical question is fine annotation of gland instances is very resource-consuming. Weak supervised semantic segmentation technology [29-32], which only relies on lightweight annotation data such as image category labels, is also becoming a hot topic in academic research. Zhou et al [31], Using the image-level category annotation supervision information, the peak value of the category response graph is explored and mapped to the object instance by back propagation, and the peak value response graph is obtained to segment the instance mask. Most of the current weak supervised instance segmentation methods [33-35] use classification network to locate the calorific value, but in the gland image, the gland and non-gland regions basically exist at the same time. Therefore, it is impossible to pre train the classification network with better accuracy, which makes the current weak supervised instance segmentation algorithm not applicable to the gland instance segmentation.

In addition, Amys[34] introduce point labeling to achieve instance segmentation of scene image. By contrast, point annotation can effectively reduce the cost of annotation and

realize instance segmentation. Furthermore, compared with image level annotation, point annotation can help the model locate the instance location better and achieve higher accuracy. The accuracy of semantic segmentation under weak supervision is greatly improved without increasing the cost of annotation. However, this method requires that the categories in the image are quite different, while the category information in the gland image is relatively single and this method is not applicable in the instance segmentation problem. Therefore, we propose a weak supervised gland instance segmentation method based on point annotation, which can effectively segment gland instances and reduce the cost of data annotation.

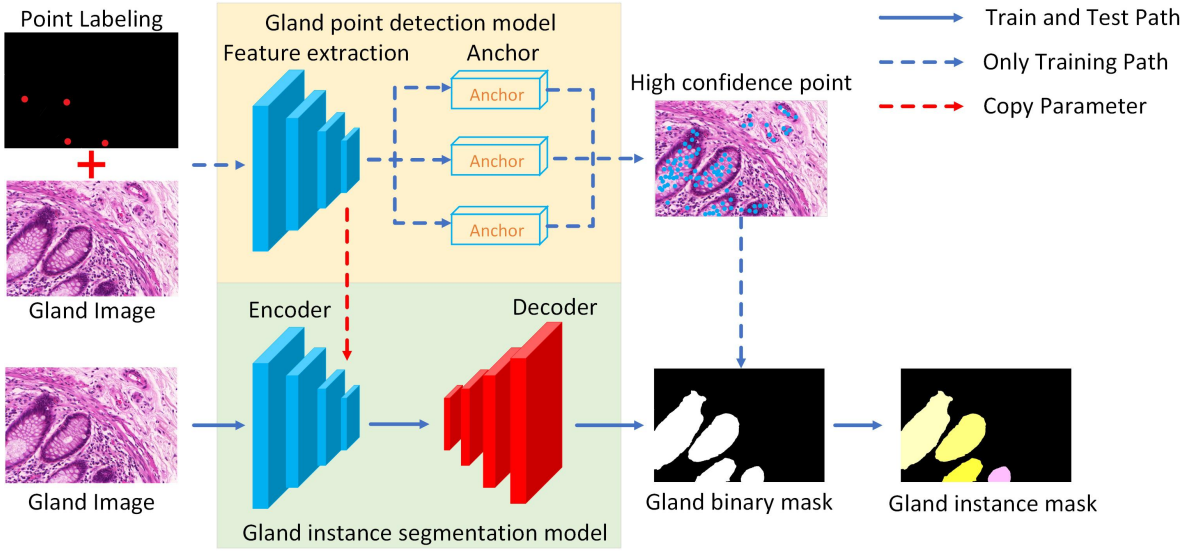

Fig. 2. An overview of the method for segmentation of weakly supervised gland cases based on point labeling.

III. METHODS

A. System Overview

Fig. 2 shows the overview of our method. We proposed a new Weak supervision method to segment the gland instance in gland images only using point labeling. Due to the point labeling contains less supervision information than the pixel-level labeling, the normal instance segmentation model cannot segment the gland area well. On the other hand, the general weak supervised instance segmentation model usually needs to rely on the pre-trained classification model for calorific value localization. As shown in Fig 1.a, the normal glandular image contains the glandular and the background. The classification model recognizes objects by learning their feature differences. When the glands and the background are always present at the same time, the classifier cannot distinguish the characteristic differences between them very well. For this, our method provides efficient solutions to segment the gland instance using point labeling by Weak supervision. Our method was divided into two parts: gland point detection model (Fig. 2 orange box) and gland instance segmentation model (Fig. 2 green box).

In the training steps, we first trained the gland point detection model. In this model, gland image is the model input and the high confidence points is the model output. The output of the point detection model is constrained by the weak supervision information in the point labeling. Gland image feature extraction was performed using Point detection model Feature Extraction part. The feature extraction part has multiple feature extraction layers. In the second half of the gland point detection model, we use a similar anchor location method to predict the location of high confidence points. The meaning of a high confidence point can be defined as a point of high probability in the glandular region. We hope that the high confidence points predicted by the gland point detection model will fall within the gland instance and cover as many glandular boundaries as possible. The high confidence point has two main roles: 1) supporting training gland instance segmentation model and 2) Help feature extraction section learn to extract gland image features. To achieve this goal, we made use of point labeling contains supervisory information to guide anchor points to predict the location of high confidence points. For the structural design description of the point detection model and the process of anchor prediction with high confidence points, please refer to section III-B.

The second part of our method is the gland instance segmentation model. In this gland instance segmentation model, gland image is the model input and the gland binary mask is the model output. We consider the feature extraction part of the point detection to be a general feature of the gland. Therefore, we use the idea of transfer learning for reference and reuse the feature extraction part as the encoder of gland instance segmentation model. The red dashed line in Fig. 2 illustrates this process. We copy the feature extraction part as an encoder for the segmentation model of the gland instance. This has two main roles: 1) It helps the instance segmentation model to locate the calorific value of the gland instance and 2） The pretraining parameters of the encoder can effectively accelerate the model fitting. The segmentation of glandular instances was

achieved by using the predicted high confidence point assisted training model. We transform the gland instance segmentation problem into a simpler semantic segmentation problem, which is realized by using a concise binary semantic segmentation model instead of relying on a complex two-stage model. The final gland instance mask is obtained by dividing each closed connected region in the gland binary mask. For the structural design description of the gland instance segmentation model, please refer to section III-C. After training the model, we only keep the gland instance segmentation model in the prediction process. The point detection model main effect was auxiliary training gland instance segmentation model. Fig. 2 the blue solid arrows represent the paths that need to be used during both training and testing, while the dotted blue arrows represent the paths that are only used during training.

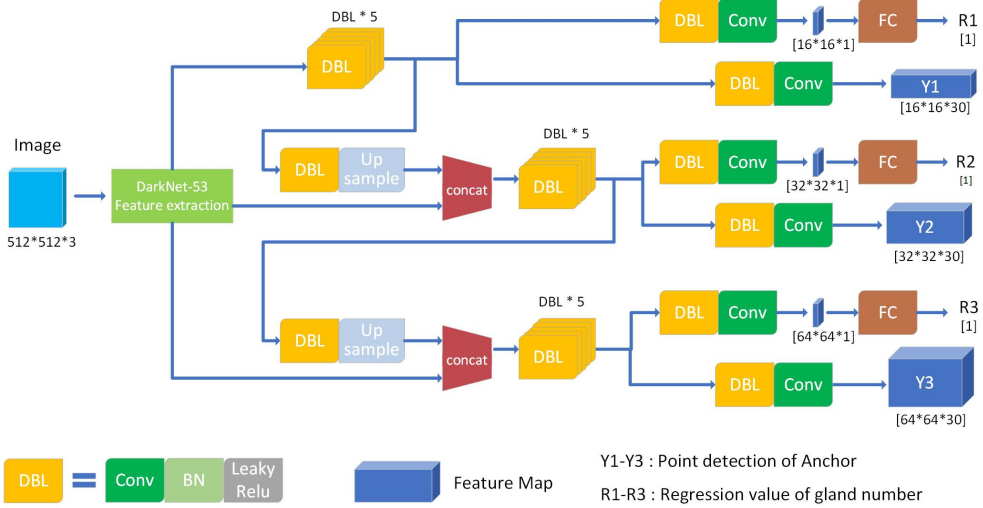

Fig. 3. Gland Point detection model design overview

B. *Gland Point Detection Model*

The gland point detection model detecting the high confidence points for the gland images. These high confidence points represent the peak points in the image for the possible gland areas. e g. On the key point detection of human body, the posture and movement can be analyzed by detecting and locating the key points of human body. The gland point detection model can be used to preliminarily locate the area that belongs to the gland area of the image. We designed the gland point detection model by referring to the one-stage detection structure of YOLOv3[15], (as shown in Fig. 3) . Firstly, Gland image were feature extraction using feature extraction part. Then, the predicted coordinate and confidence of the gland high confidence points are obtained by prediction of Anchor mechanism.

In the gland point detection model, we used Darknet-53 as the backbone of the model to extract and encode semantic features in the image. DarkNet adopts a residual connection method similar to the structure of ResNet[37]. With the number of layers increases, the neural network shows a degradation problem, performance the deep level of the network is not as good as the shallow level of the network. The residual structure in Darknet makes the network have a strong ability of identity mapping, thus expanding the depth of the network and improving the performance of the network. In the feature extraction part, we down-sampling the image features for 5 times, during which the length and width were compressed to 1/2 of the original length and width. In order to reduce the loss of feature information in the process of sampling, we replace the pooling layer in the feature extraction part with the convolution layer with stride size of 2 and filter size of 2.

Multiple gland with different sizes and different shape may appear simultaneously in a gland image. So, gland detection model needs detection and location of glands from different receptive fields. The shallow feature can be used to distinguish the simple target, the deep feature can be used to distinguish the complex target. Therefore, instance segmentation of glands necessary to synthesize the semantic information of multi-scale features and detect glands of different sizes. We adopt the structure of feature pyramid networks (FPN) [36] to extract multi-scale feature information for fusion, so as to improve the accuracy of target detection. In our detection layer, the top-level features are fused by the upper sampling and the lower features, so that the model can combine the context information of the features. We used the DBL feature extraction module as the basic component of the detection part. DBL is a combination of a convolution layer, a batch normalization layer and a Leaky ReLu layer (as shown in Fig. 3).

We regressed to the coordinate of the high confidence points on the feature maps sampled at a factor of 1/8、1/16、1/32, respectively. By improving the anchor mechanism, the detection layer can predict the coordinates and confidence of high confidence points. Before using the anchor mechanism for high-confidence point prediction, it is necessary to change the grid size of feature map into N*N

(each small cell is called grid cell). In this paper, the grid size of feature map is changed to 16,32,64 feature map squares (as shown in Fig. 3). Then, a convolution layer is used to convert the feature maps to obtain the anchor feature maps. This step is to convert the dimension of the feature maps. The depth of the anchor feature maps is 3*M, and M represents the number of predicted high confidence points for each cell. e.g., When M is equal to 10, each cell predicts the coordinate of 10 high confidence points, so the depth of the anchor feature map is 3*10=30. Each predicted high confidence point has three attribute values, which describe the central coordinates and confidence degree of each high confidence points. In this paper, we predicted 10 high confidence points for each cell. The process of obtaining the coordinate of high confidence points through grid output:

$$p_x = \sigma(t_x) + c_x \ , \ p_y = \sigma(t_y) + c_y \quad (1)$$

Where p_x and p_y are the central coordinates x and y of the predicted high confidence points, t_x and t_y are the offset of the predicted anchor, and c_x and c_y are the upper-left coordinates of the grid. (Fig. 3, y1 -y3 are the anchor detection results at three scales). if the predicted offset values t_x and t_y are greater than 1, the center coordinate value will beyond the grid range, but the anchor mechanism is to predict the offset of high confidence points in the corresponding cell. To address this problem, the offset is compressed into a range of 0 to 1 by the sigmod function. This effectively keeps the offset in the corresponding cell.

In order to improve the gland point detection model description of glands instance relationship performance, we design a parallel to the anchor mechanism of the regression branch. The branch uses a convolution will figure characteristics of the deep compression is 1 and conversion the feature vector, use full connection layer within the feature vector regression forecast on glands instance number (Fig. 2, r1 - r3 is gland instance number prediction results). Gland point detection model prediction high confidence points through the detection branch y1-y3, and prediction the number of gland instances in the image by regression branch r1-r3. In the prediction process, we set a threshold value τ to filter the high confidence points of the model output. The points confidence is above the threshold value τ are reserved. points with confidence above the threshold were set to be high confidence point as gland detection model output.

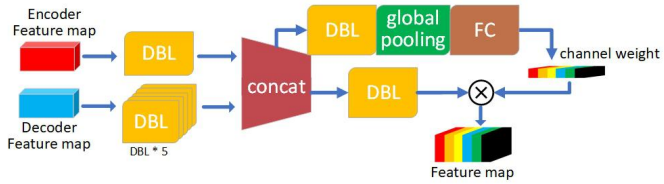

Fig. 4. Gland instance segmentation model decoding unit overview

C. Gland instance segmentation model

The gland instance segmentation module divides the gland region in the image into different classes/regions, which is a pixel level binary classification problem. The networks take as input an image of size W × H and output a W × H × 2 score map where 2 is the set of Gland or non-gland areas the CNN was trained to recognize. At test time, the score map is converted to per-pixel predictions of size W ×H by either simply taking the maximally scoring class at each pixel.

Here we refer the encoder-decoder structure by U-Net [10] to design our gland instance segmentation model. We used the feature extraction part of the trained gland point detection model as the encoder of the gland instance segmentation model (Fig.2). When the gland point detection model is trained with point labeling information, the feature extraction part is restricted by the supervision information in the point labeling. So, the feature extraction part can extract glandular features more effectively. Therefore, when training the gland instance segmentation model, reusing the coding structure and initializing the parameters can effectively improve the convergence rate of the gland segmentation model and reduce the probability of overfitting. This procedure compensates for the lack of pre-training models for migration training and calorific value localization of gland instance segmentation under weak supervision. The decoder of gland instance segmentation model is designed by referring to the decoding structure of U-Net. In the decoder, we use the number of up-sampling relative to the feature extraction part. In order to restore the scale of the feature graph, the feature graph was up-sampling for 5 times. The structure of the decoding unit is shown in Fig. 4, which integrates the feature maps of different scales and depths, combines the rough segmentation of low resolution with the fine segmentation of high resolution, and obtains a good segmentation result. The deep information can be understood as the low-resolution information after several times of sampling, which can provide the contextual semantic information of the segmentation target in the whole image, and can be understood as the characteristics of the relationship between the target and its environment. Shallow information is high-resolution information passed directly from encoder to the same height decoder via the concatenate operation.

In addition to the lateral connection between the depth information and the shallow information, we also introduce an attention mechanism to redistribute the attention weights between the channels of the feature map. First use of global pooling layer get the direction of the channel characteristics of the image global features. Then use the full connection learn the weights of relationships between each channel, redistribute the weight proportion between the channel. The channel weights were multiplied original feature maps according channel sequence. Essentially, the attention mechanism does attention or blocking operations on the channel dimension. This attention mechanism allows the model to focus more on the channel features with the most information and suppress those that are not important.

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors).*

D. Loss function

1) Gland point detection model loss function: In the point detection model, we predicted the anchor with high confidence point in the grid feature graph at three different scales. For the prediction loss calculation of high confidence point coordinates, it is similar to the detection loss of bounding box:

$$L_D = \frac{1}{3}\sum_{n=3}^{N}\sum_{i=0}^{M^2}\sum_{j=0}^{P}\frac{1}{p}\mu_{i,j}[|x_i - \hat{x}_i| + |y_i - \hat{y}_i|] \quad (2)$$

Where N represents the number of scales detected, M^2 represents the grid size, P represents the predicted number of high confidence points in each grid, x_i and y_i are the coordinate values of grid prediction, $\hat{x}_i$ and $\hat{y}_i$ are the coordinate values of ground truth output from the grid. ω represents the validity of the anchor prediction high confidence point coordinates in the grid, with values of 0 and 1. $\mu_{i,j}$ represents the validity of the j anchor prediction in the i grid. If there are high confidence points in the grid, the predicted high confidence point coordinates of all anchors in the cell are valid. The value of $\mu_{i,j}$ of all anchors in the grid is 1, and the total loss is the average of all predicted losses.

In the gland number regression branch, we transformed the problem of gland detection into the regression problem of gland number calculation to improve the experimental efficiency of labeling internal supervision information. We used the log-cosh loss function to evaluate the regression loss of the number of glands in this branch:

$$L_R = \frac{1}{3}\sum_{i=3}^{N}\log(\cosh(r_i - \hat{r}_i)) \quad (3)$$

Where N represents the number of scales detected, r_i represents the number of glands predicted by the regression branch, and $\hat{r}_i$ represents the actual number of glands. The log-cosh loss is a loss function that is applied to regression problems and is smoother than L2. It's calculated as the logarithm of the hyperbolic cosine of the prediction error. Log-cosh Loss is not as sensitive to the abnormal points in the data as the square error Loss, and the second derivative of Log-cosh is differentiable. The gland point detection model loss was a combination of equations 2 and 3.

2) Gland instance segmentation model loss function: In the gland instance segmentation model, we obtain the binary instance segmentation mask of the gland instance by encoding and decoding. The loss design for segmentation of gland instances L_S consists of the following two parts:

$$L_S = L_{S-G} + L_{S-C} \quad (4)$$

$$L_{S-G} = \frac{1}{W \times H}\sum_{i}^{W\times H}[\hat{o}_i \log(o_i)) + (1 - \hat{o}_i)\log(1 - o_i)] \quad (5)$$

$$L_{S-C} = \frac{1}{T}\sum_{t}^{T}\log(s_i) \quad (6)$$

Where W and H represent the size of the image, o_i is the segmentation mask of the gland instance output of the model, and $\hat{o}_i$ is the grad-cam calorific value output of the model. T represents the number of high confidence points predicted, and s_i is the output of the instance segmentation model of the location of high confidence points. This loss function encourages the model to divide the high confidence points into gland regions. We used the binary cross entropy function in pixel direction to evaluate the consistency between the calorific value map and the segmentation results. For high confidence points we also used the cross-entropy function to encourage the model to classify high confidence points as gland regions.

IV. Experiments

We empirically demonstrate the efficiency of our method of segmentation of weakly supervised gland instances using only points labeling. We compare our method with the current popular methods of full pixel supervision and weak point annotation supervision. Our conclusion is that under the condition of only using point labeling, the gland instance segmentation model with high confidence point training can effectively segment the gland instance and improve the mobility and applicability of the weak supervision algorithm.

A. Dataset

We evaluated Our method on the 2015 MICCAI Gland Challenge dataset [16]. The data set contains 165 gland images, of which 85 are the training set and 80 are the test set (including 60 in part A and 20 in part B).The training set contained 769 instances of glands and the test set 761 instances of glands (666 in part A and 95 in part B). Each gland image in the data set provides pixel level labeling of the gland instance. In addition, we manually annotated the 85 images in the training set with point labeling, and each gland instance contained a point label.

B. Evaluation criteria

1) Evaluation of detection gland points: We define that if the distance between all the high confidence points and the callout points is less than the preset threshold value D, the callout points are successfully detected. We use Accuracy to assess the accuracy of spot detection, to calculate the percentage of detected points labeling in the total number of points labeling.

2) Evaluation of Segmentation gland instances: We used the following three metrics to evaluate the model's instance segmentation performance: $F1\ score$, $Dice_{object}$ and H_{object}. The first criterion for evaluation reflets the accuracy of gland detection which is called F1score. The segmented gland object of True Positive (TP) is the object that shares more than 50% areas with the ground truth. Otherwise, the segmented area will be determined as False Positive (FP). Objects of ground truth without corresponding prediction are considered as False Negative (FN).

$$F1\ \text{score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

Dice is the second criterion for evaluating the performance of segmentation. Dice index of the whole image is

$$Dice(G,S) = \frac{2|G \cap S|}{|G| \cup |S|} \quad (8)$$

of which G represents the ground truth and S is the segmented result. However, it is not able to differentiate instances of same class. As a result, object-level dice score is employed to evaluate the segmentation result. The definition is as follows:

$$Dice_{object}(G,S) = \frac{1}{2}\left[\sum_{i=1}^{n_S} \omega_i D(G_i,S_i) + \sum_{i=1}^{n_S} \widetilde{\omega}_i D(\widetilde{G}_i,\widetilde{S}_i)\right] \quad (10)$$

$$\omega_i = \frac{|S_i|}{\sum_{j=1}^{n_S} |S_i|}, \quad (11)$$

$$\omega_i = \frac{|\widetilde{G}_i|}{\sum_{j=1}^{n_G} |\widetilde{G}_i|}. \quad (12)$$

n_S and n_G are the number of instances in the segmented result and ground truth.

Shape similarity reflects the performance on morphology likelihood which plays a significant role in gland segmentation. Hausdorff distance is exploited to evaluate the shape similarity. To assess glands respectively, the index of Hausdorff distance deforms from the original formation:

$$H(G,S) = \max\left\{ \sup_{x\in G} \inf_{y\in S} \|x-y\|, \sup_{y\in G} \inf_{x\in S} \|x-y\| \right\}, \quad (13)$$

to the object-level formation:

$$H_{object}(S,G) = \frac{1}{2}\left[\sum_{i=1}^{n_S} \omega_i H(G_i,S_i) + \sum_{i=1}^{n_G} \widetilde{\omega}_i H(\widetilde{G}_i,\widetilde{S}_i)\right]. \quad (14)$$

Similar to object-level dice index n_S and n_G represents instances of segmented objects and ground truth. The definition of ω_i and $\widetilde{\omega}_i$ in $H_{object}(S,G)$ is the same as that in $Dice_{object}(G,S)$, With reference to equations 11 and 12.

C. *The Effects of τ to Segmentation of gland instances and Detection of High confidence point*

In the point detection model, we use a threshold τ to filter the output of the point detection model. For output of point detection model, we applied a threshold τ to filter low confidence points and ignored points less than the threshold. The point where the confidence is higher than the threshold τ is the high confidence point. As for the optimal value of threshold τ, the relevant experimental results are shown in table 1 and table 2. In table 1, We compared the performance of models with different thresholds τ under different sizes of evaluation D. In table 2, we compared the performance of the gland instance segmentation model using different thresholds τ.

TABLE I. THE PERFORMANCE OF DETECTION OF HIGH COORDINATE POINTS WITH DIFFERENT T

D	τ	*Acc*	
		Part A	*Part B*
10	0.25	0.948	0.968
10	0.5	**0.941**	**0.968**
10	0.75	0.851	0.926
5	0.25	0.903	0.947
5	0.5	**0.936**	**0.968**
5	0.75	0.870	0.842

Through the experimental results, we can find that when the threshold is too large or too small, the instance segmentation results of the glands are decreased. When the value of τ is 0.5, the segmentation effect of gland instance is best. Figure 5 is an example of a gland point detection result. The red points are the manually labeled points, and the blue points are the high confidence points predicted by the model. As can be seen in Fig 5, when the threshold τ is too small, some glands in the image are not detected. The gland instance segmentation model mistakenly segments some glands into background. When the threshold τ is too large, a large number of high coordinate points in the background region. As a result, part of the background was segmented into glands. Therefore, it is very important to choose the right τ value. When most of the high confidence points are located in the gland region, it can effectively improve the localization for the gland instances boundary and the ability to distinguish the background between the instances.

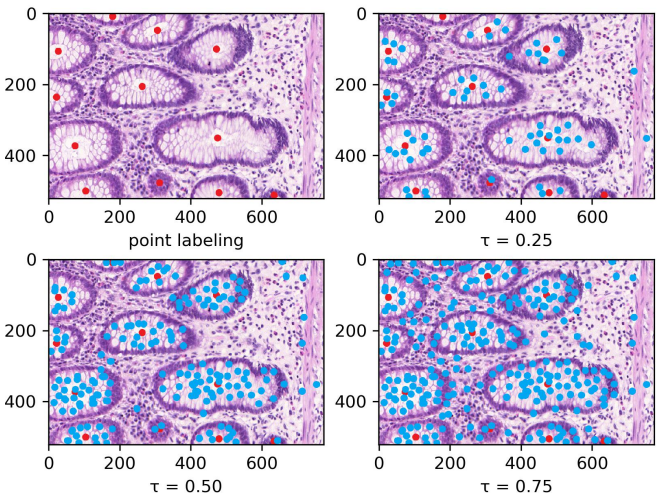

Fig. 5. Sample gland point detection model output

TABLE II. THE PERFORMANCE OF SEGMENTATION OF GLAND INSTANCES WITH DIFFERENT T

τ	$F1\ score$		$Dice_{object}$		H_{object}	
	Part A	***Part B***	***Part A***	***Part B***	***Part A***	***Part B***
0.25	0.744	0.673	0.657	0.516	108.873	219.453
0.5	**0.817**	**0.749**	**0.826**	**0.754**	**76.768**	**164.638**
0.75	0.782	0.621	0.802	0.697	95.237	180.439

TABLE III. EXPERIMENTAL RESULTS OF STRUCTURE ABLATION WHEN THE THRESHOLD T IS 0.5

Copy-parameter	Points-detection	Regression-branch	$F1\ score$		$Dice_{object}$		H_{object}	
			Part A	***Part B***	***Part A***	***Part B***	***Part A***	***Part B***
			0.378	0.261	0.357	0.284	247.864	291.341
	✓	✓	0.744	0.606	0.781	0.657	124.706	190.447
✓		✓	0.452	0.318	0.437	0.433	195.433	216.244
✓	✓		0.777	0.686	0.801	0.673	92.260	188.835
✓	✓	✓	**0.817**	**0.749**	**0.826**	**0.754**	**76.768**	**164.638**

TABLE IV. GLAND INSTANCES SEGMENTATION RESULTS

	$F1\ score$		$Dice_{object}$		H_{object}	
Model	***Part A***	***Part B***	***Part A***	***Part B***	***Part A***	***Part B***
DSE[22]	0.926	0.862	0.927	0.871	31.209	80.509
MILD-Net[20]	0.914	0.844	0.913	0.836	41.540	105.890
SPL-Net[21]	0.924	0.844	0.902	0.840	49.881	106.075
DMCN[17]	0.893	0.843	0.908	0.833	44.129	116.821
DCAN[19]	0.912	0.716	0.897	0.781	57.418	160.347
MPCNN[16]	0.891	0.703	0.882	0.786	57.413	145.575
DCAI [37]						
FCN[9]	0.788	0.764	0.813	0.796	95.054	146.247
CVML	0.652	0.541	0.644	0.654	155.433	176.244
LIB	0.777	0.306	0.781	0.617	112.706	190.447
SDS[38]	0.545	0.322	0.647	0.495	116.833	229.853
Ours	**0.817**	**0.749**	**0.826**	**0.754**	**76.768**	**164.638**

D. Structure ablation experiment

For completeness, we tested the Ablation Experiment of the model structure. The experimental results are shown in Table 3. We tested the experimental controls of Regression-branch, Points- detection, and Copy-parameter separately. Among them, the significance of Regression-branch is whether the point detector add Regression-branch. If Regression-branch is turned off, the point detection model will only output the points detection results. The point detection option means whether to use the gland point detection model to predict high confidence points and assist in training the gland instance segmentation model. If Points-detection is turned off, high confidence points are not added to assist in training the gland instance segmentation model. Finally, the meaning of the Copy-parameter whether to copy the parameters of the feature extraction part of the gland point detection model to the encoder of the gland instance segmentation model as initialization. We performed the following ablation experiments when the threshold τ was 0.5.

We compared the effects of three structures on model performance. From the results in Table 3, it can be seen that whether the high confidence points predicted by the point detection model are added has the greatest impact on the model performance. From the experimental results, we can draw the following conclusions:1) When the point detection is turned off, Regression-branch improvements to the model are further reduced.2) Both Regression-branch and Copy-parameter can effectively improve the quality of gland instance segmentation, but the premise is to perform point detection. 3)When all three structures are missing, the severe lack of supervision information makes the model perform worst. When all structures exist simultaneously, the performance of the model is optimal. The experimental results further prove that the method of assisted training

through point detection can effectively improve the performance of weakly supervised gland instance segmentation.

E. Segmentation results of gland instances

We compared our method with the current best fully supervised gland instance segmentation algorithm, and the results are shown in table 4. In this paper, the method of instance segmentation of weak supervised glands based on point detection is more effective than some full-supervised method using pixel level labeling. The high confidence points in the image of the gland were predicted to assist the training of the gland instance segmentation model. Because this method uses the method of binary segmentation to distinguish the pixels in the gland image into gland region and non-gland region, there is no constraint on the difference between instances. At present, the best method [34] uses the complex structure of detector model and segmentation model, which has higher requirements for data annotation. However, the method in this paper can achieve accurate segmentation of gland instances only by using point labeling, and the accuracy is due to the partial full supervision model.

According to the experimental results, the high confidence points generated by the point detection model can effectively locate the gland, so our method performs well in $F1\,score$ .Under the appropriate threshold τ , the gland instance segmentation model can effectively divide the gland boundary, so our method also has good effect on $Dice_{object}$ and H_{object} . Exemplar segmentation results are shown in Fig. 6. In the example, we can intuitively see the influence of threshold τ on the instance segmentation of gland. With the increase of threshold τ, the high confidence point can be improved the coordinating of boundary of gland. When threshold τ is too large, many backgrounds will be misclassified. When threshold τ is too small, the number of high confidence points at the boundary of the gland is less, which makes the model unable to locate the boundary of the gland well. Therefore, it is very important to choose the right threshold τ. Although in most cases the model can distinguish the border of the glands. But when the interval between the glands is too small, two or more glands will stick together. In addition, when the gland is too small, the model identifies the gland as the background. Although in most cases the model can distinguish the border of the glands. But when the interval between the glands is too small, two or more glands will stick together. To the best of our knowledge we are the first to segment a weakly supervised instance of the gland using point labeling, so more work can be done in this direction.

V. CONCLUSIONS

In this paper, we propose a method for instance segmentation of weak supervised glands based on point detection. The gland point detection model was trained to predict high confidence points in the image using only point notation. These high confidence points were used to assist the training of gland instance segmentation model to achieve the segmentation of gland instances.

Fig. 6. In the example of weak supervision of gland instance segmentation based on point labeling, the first column is the original picture, the second column is the ground truth, the third column is the effect when τ is 0.25, the fourth column is the effect when τ is 0.5, and the fifth column is the effect when τ is 0.75

VI. REFERENCES

[1] Simon Cross, Sa Betmouni, Julian L Burton, A. Dube, Kenneth M. Feeley, Miles R. Holbrook, Robert J. Landers, Phillip B. Lumb, Timothy J. Stephenson:What Levels of Agreement Can Be Expected Between Histopathologists Assigning Cases to Discrete Nominal Categories? A Study of the Diagnosis of Hyperplastic and Adenomatous Colorectal Polyps.Modern Pathology, 13(9), 941-944

[2] K Komuta, K Batts, Jose Jessurun, Dale Snover, J Garcia-Aguilar, D Rothenberger, R Madoff:Interobserver variability in the pathological assessment of malignant colorectal polyps.British Journal of Surgery 91(11): 1479-1484

[3] Thomas R Fanshawe, Andrew G Lynch, Ian O Ellis, Andrew R Green, Rudolf Hanka:Assessing Agreement between Multiple Raters with Missing Rating Information, Applied to Breast Cancer Tumour Grading.Plos One: e2925-3(8)

[4] Hao Fu, Guoping Qiu, Jie Shu, Mohammad Ilyas: A Novel Polar Space Random Field Model for the Detection of Glandular Structures. IEEE Trans. Med. Imaging 33(3): 764-776 (2014)

[5] Matthew Fleming, Sreelakshmi Ravula, Sergei F Tatishchev, Hanlin Wang:Colorectal carcinoma: Pathologic aspects.ournal of gastrointestinal oncology.3(3):153-73

[6] Ahmed Fakhry, Tao Zeng, Shuiwang Ji: Residual Deconvolutional Networks for Brain Electron Microscopy Image Segmentation. IEEE Trans. Med. Imaging 36(2): 447-456 (2017)

[7] Pericles S. Giannaris, Zainab Al-Taie, Mikhail Kovalenko, Richard Hammer, Mihail Popescu, Dmitriy Shin: Informatics Framework to Identify Consistent Diagnostic Techniques. BIBM 2019: 1481-1486

[8] Anant Madabhushi, George Lee: Image analysis and machine learning in digital pathology: Challenges and opportunities. Medical Image Anal. 33: 170-175 (2016)

[9] Jonathan Long, Evan Shelhamer, Trevor Darrell: Fully convolutional networks for semantic segmentation. CVPR 2015: 3431-3440

[10] Olaf Ronneberger, Philipp Fischer, Thomas Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI (3) 2015: 234-241

[11] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross B. Girshick: Mask R-CNN. ICCV 2017: 2980-2988

[12] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, Chen Change Loy, Dahua Lin: Hybrid Task Cascade for Instance Segmentation. CVPR 2019: 4974-4983

[13] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang: UNet++: A Nested U-Net Architecture for Medical Image Segmentation. DLMIA/ML-CDS@MICCAI 2018: 3-11

[14] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, Hanqing Lu: Dual Attention Network for Scene Segmentation. CVPR 2019: 3146-3154

[15] Daniel Bolya, Chong Zhou, Fanyi Xiao, Yong Jae Lee: YOLACT: Real-Time Instance Segmentation. ICCV 2019: 9156-9165

[16] Korsuk Sirinukunwattana, Josien P. W. Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J. Matuszewski, Elia Bruni, Urko Sanchez, Anton Böhm, Olaf Ronneberger, Bassem Ben Cheikh, Daniel Racoceanu, Philipp Kainz, Michael Pfeiffer, Martin Urschler, David R. J. Snead, Nasir M. Rajpoot: Gland segmentation in colon histology images: The glas challenge contest. Medical Image Anal. 35: 489-502 (2017)

[17] Yan Xu, Yang Li, Yipei Wang, Mingyuan Liu, Yubo Fan, Maode Lai, Eric I-Chao Chang:Gland Instance Segmentation Using Deep Multichannel Neural Networks. IEEE Trans. Biomed. Engineering 64(12): 2901-2912 (2017)

[18] Yan Xu, Yang Li, Mingyuan Liu, Yipei Wang, Maode Lai, Eric I-Chao Chang: Gland Instance Segmentation by Deep Multichannel Side Supervision. MICCAI (2) 2016: 496-504

[19] Hao Chen, Xiaojuan Qi, Lequan Yu, Pheng-Ann Heng.DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation.(CVPR 2016)

[20] Simon Graham, Hao Chen, Jevgenij Gamper, Qi Dou, Pheng-Ann Heng, David R. J. Snead, Yee-Wah Tsang, Nasir M. Rajpoot: MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. Medical Image Anal. 52: 199-211 (2019)

[21] Zengqiang Yan, Xin Yang, Kwang-Ting (Tim) Cheng: A Deep Model with Shape-Preserving Loss for Gland Instance Segmentation. MICCAI (2) 2018: 138-146

[22] Yutong Xie, Hao Lu, Jianpeng Zhang, Chunhua Shen, Yong Xia: Deep Segmentation-Emendation Model for Gland Instance Segmentation. MICCAI (1) 2019: 469-477

[23] Hui Qu, Zhennan Yan, Gregory M. Riedlinger, Subhajyoti De, Dimitris N. Metaxas: Improving Nuclei/Gland Instance Segmentation in Histopathology Images by Full Resolution Neural Network and Spatial Constrained Loss. MICCAI (1) 2019: 378-386

[24] Jiwoon Ahn, Sunghyun Cho, Suha Kwak: Weakly Supervised Learning of Instance Segmentation With Inter-Pixel Relations. CVPR 2019: 2209-2218

[25] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, Thomas S. Huang: Revisiting Dilated Convolution: A Simple Approach for Weakly- and Semi-Supervised Semantic Segmentation. CVPR 2018: 7268-7277

[26] Yunhang Shen, Rongrong Ji, Yan Wang, Yongjian Wu, Liujuan Cao: Cyclic Guidance for Weakly Supervised Joint Detection and Segmentation. CVPR 2019: 697-707

[27] Wataru Shimoda, Keiji Yanai: Self-Supervised Difference Detection for Weakly-Supervised Semantic Segmentation. ICCV 2019: 5207-5216

[28] Zeng Yu, Yun-Zhi Zhuge, Huchuan Lu, Lihe Zhang: Joint Learning of Saliency Detection and Weakly Supervised Semantic Segmentation. ICCV 2019: 7222-7232

[29] Xiang Wang, Huimin Ma, Shaodi You: Deep clustering for weakly-supervised semantic segmentation in autonomous driving scenes. Neurocomputing 381: 20-28 (2020)

[30] Xiang Wang, Sifei Liu, Huimin Ma, Ming-Hsuan Yang: Weakly-Supervised Semantic Segmentation by Iterative Affinity Learning. CoRR abs/2002.08098 (2020)

[31] Yanzhao Zhou, Yi Zhu, Qixiang Ye, Qiang Qiu, Jianbin Jiao: Weakly Supervised Instance Segmentation Using Class Peak Response. CVPR 2018: 3791-3800

[32] Yi Zhu, Yanzhao Zhou, Huijuan Xu, Qixiang Ye, David S. Doermann, Jianbin Jiao: Learning Instance Activation Maps for Weakly Supervised Instance Segmentation. CVPR 2019: 3116-3125

[33] Weifeng Ge, Sheng Guo, Weilin Huang, Matthew R. Scott: Label-PEnet: Sequential Label Propagation and Enhancement Networks forWeakly Supervised Instance Segmentation. CoRR abs/1910.02624 (2019)

[34] Amy L. Bearman, Olga Russakovsky, Vittorio Ferrari, Fei-Fei Li: What's the Point: Semantic Segmentation with Point Supervision. ECCV (7) 2016: 549-565

[35] Joseph Redmon, Ali Farhadi: YOLOv3: An Incremental Improvement. CoRR abs/1804.02767 (2018)

[36] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, Serge J. Belongie: Feature Pyramid Networks for Object Detection. CVPR 2017: 936-944

[37] Liye Mei, Xiaopeng Guo, Xin Huang, Yueyun Weng, Sheng Liu, Cheng Lei: Dense Contour-Imbalance Aware framework for Colon Gland Instance Segmentation. Biomed. Signal Process. Control. 60: 101988 (2020)

[38] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep Residual Learning for Image Recognition. CVPR 2016: 770-778

[39] Bharath Hariharan, Pablo Andrés Arbeláez, Ross B. Girshick, Jitendra Malik: Simultaneous Detection and Segmentation. ECCV (7) 2014: 297-312