Applied Mathematics, Modeling and Computer Simulation
C. Chen (Ed.)

doi:10.3233/ATDE220087

Image Segmentation of Lesions in Wireless Capsule Endoscopy Based on Mask R-CNN

Nianfeng LI[a], Nan ZHAO[b,1], Zhiguo XIAO[a] and Jia LU[b]
[a] *College of Computer Science and Technology, Changchun University, China*
[b] *Graduate School of Changchun University, Changchun University, China*

Abstract. The wireless capsule endoscope can comprehensively inspect the digestive tract. It has the advantages of safety, painlessness, and no postoperative reaction, but it has some disadvantages. When the most critical issue is, the entire inspection process is very time-consuming. Using image segmentation technology, the digestive tract lesion tissue and surrounding areas can be locally enhanced, and the Region of Interest (ROI) can be segmented, which is helpful for the doctor to read the image. Based on the Mask R-CNN network, this paper proposes a wireless capsule endoscope lesion image segmentation method. Experimental results show that this method can achieve effective segmentation of the capsule lens lesion image.

Keywords. Wireless Capsule Endoscope, Image Segmentation, Deep Learning.

1. Introduction

China is a country with a high incidence of gastrointestinal cancer. Taking esophageal cancer as an example, there are about 246,000 new cases and 188,000 deaths every year in China [1]. Wireless Capsule endoscopy is a capsule-shaped endoscope, which is a medical instrument used to examine the human intestines. The capsule endoscope can enter the human body, take all-round photographs of the digestive tract and transmit the photographed information to the external information collection device. Traditionally, after collecting the information of the patient's digestive tract, the doctor needs to classify, analyze and diagnose the information (video or image) collected by the wireless capsule endoscope. This process requires manual participation in the whole process, which has a long cycle and high cost. It also has high requirements for the professional knowledge and experience of the reading doctor. The computer vision technology of deep learning can realize the intelligent detection of digestive tract lesions, thereby assisting doctors in the diagnosis of digestive tract lesions, improving the efficiency of image reading, and shortening the detection cycle.

Image segmentation is an important research field of computer vision. Using image segmentation technology, the digestive tract lesion tissue and surrounding areas can be locally enhanced, and the ROI can be segmented, which is beneficial for the

[1] Corresponding Author, Nan ZHAO, Graduate School of Changchun University, Changchun University, China; E-mail: 200701193@mails.ccu.edu.cn.

doctor to read the image. So far, in the field of medical image segmentation, researchers have proposed many excellent methods. According to different technical methods, they can be divided into traditional image segmentation methods and image segmentation methods based on deep learning methods.

Traditional image segmentation methods mainly include threshold segmentation method, region growing method and graph cut method. The core of the threshold segmentation method is to select the best threshold, and the image can be binarized and segmented through the best threshold. However, the threshold segmentation method does not make good use of the spatial information between pixels, and is easily affected by noise in the image. The core of the region growing method is to calculate the gray value of the image, and merge the subregions of similar pixels to form the largest region. The region growing method can ensure the spatial continuity of the segmented image, but this process requires manual participation and is also susceptible to noise in the image. The core of graph cut method is to use Markov random field (probabilistic undirected graph model) to realize image segmentation. The graph cut method has better robustness and can realize more complex image segmentation, but the space complexity and time complexity are higher, so it is often used in combination with other traditional image segmentation methods.

Image segmentation methods based on deep learning methods mainly include image segmentation methods based on Fully Convolutional Networks (FCN), image segmentation methods based on U-Net network, image segmentation based on U-Net++ network and Mask R-CNN (Mask Region-based Convolutional Neural Network) network image segmentation and other methods. FCN is an encoding-decoding network structure model, which only contains a convolutional layer and a pooling layer. It has no size requirements for the image input to the network, and can output images of the same size after inference and learning. FCN achieves pixel-level image segmentation, but its pixel-by-pixel segmentation mode makes it ignore context information during segmentation. The structure of U-Net and U-Net++ is similar to the FCN structure, and it also only contains the convolutional layer and the pooling layer. They are named because their network structure is similar to the letter U. The network structure of U-Net mainly includes up-sampling, down-sampling, and skip connection parts. The up-sampling part is used to extract the global features of the picture, and the down-sampling part has higher accuracy, which can improve the accuracy of image segmentation. The jump connection part improves the accuracy of segmentation by sharing deep and low-level information. But U-Net has poor performance in object boundary segmentation. U-Net++ changed the direct series method of the jump part to the dense connection method, which reduced the difference between the features of the upper and lower sampling stages, and further improved the segmentation accuracy. However, the network structure is complex and the number of parameters is large. Mask R-CNN is based on three target detection frameworks: R-CNN (Region-based Convolutional Neural Network) , Fast R-CNN (Fast Region-based Convolutional Neural Network) [2-4], Faster R-CNN Developed from above. R-CNN was proposed in 2014. It first uses Selective Search algorithm [4] to extract Region Proposals, then uses pre-trained CNN to extract features, and finally uses SVM (Support Vector Machine) classify recommendations for each area. Fast R-CNN abandons the use of SVM for classification on the basis of R-CNN, but retains the selective search algorithm and realizes efficient end-to-end training. At the same time, the ROI Pool module is used to extract the ROI feature vector, and finally two fully connected layers are used for object classification and bounding box coordinate

regression. However, the use of an independent search and selection algorithm makes the algorithm inefficient in the inference stage. Therefore, Faster R-CNN replaces Fast R-CNN with a selection search algorithm with a regional recommendation network (Region Proposal Network, RPN), which integrates the generation of regional recommendations into the architecture and improves efficiency. However, the above three convolutional neural networks based on region detection can only complete the task of target detection. In order to complete the target segmentation task at the same time, Mask R-CNN adds a prediction segmentation branch on the basis of Faster R-CNN, replacing ROI Pool with a more accurate ROI Align module; and adding the FCN branch after ROI Align to get the target The binary mask of the instance, thus realizing the image segmentation at the instance level.

In order to better segment the digestive tract pictures, this paper applies the deep convolutional neural network based on Mask R-CNN to the digestive tract lesion image segmentation. Experiments show that the method in this paper has good performance in the digestive tract lesion image segmentation. Performance, can realize the segmentation of 7 types of lesions in the digestive tract. Figure 1 is an example segmentation picture of the digestive tract image picture.

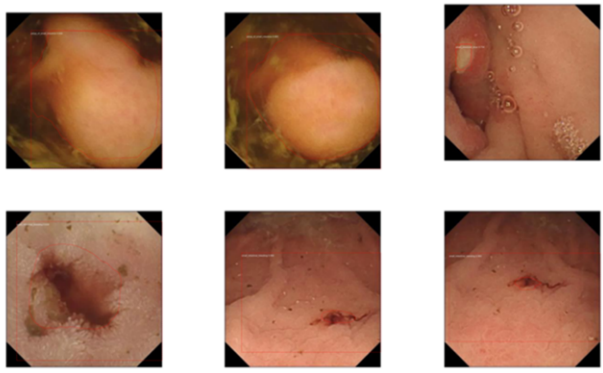

Figure 1. Example segmentation picture of digestive tract image picture

2. Related Work

2.1. Traditional Image Segmentation Method

As mentioned above, traditional image segmentation methods mainly include threshold segmentation method, region growing method and graph cut method. The representative method of threshold segmentation is based on the gray-scale image segmentation OTSU method, which calculates the optimal threshold by maximizing the between-class variance. Jan Hendrik Moltz [5] et al. proposed a semi-automatic algorithm for segmenting liver metastases. The algorithm consists of two parts. First, based on the gray value analysis of the given ROI, the coarse segmentation is generated by the adaptive threshold method. Secondly, the structures adjacent to the tumor are removed through model-based morphological processing. This algorithm successfully segmented all 10 tumors in the 2008 MICCAI3D liver tumor segmentation challenge. The representative method of the regional growth method is the watershed algorithm, which simulates the altitude based on the gray value of the image by simulating

geological landforms. The local minimum value of the gray value is the valley bottom and the local maximum value is the valley peak. The representative method of the graph cut method is the graphcut method. This algorithm divides a number of unconnected subgraphs by creating a weighted graph, removing edges with smaller weights.

In summary, traditional image segmentation methods are mostly based on the gray value of pixels to classify pixels. Although they have achieved certain effects in the field of image segmentation, they do not use the context information between pixels, so they are susceptible to noise in the image. Traditional image segmentation methods are outdated, or have been integrated into mainstream image segmentation methods based on deep learning.

2.2. Image Segmentation Method Based on Deep Learning

With the explosive growth of computer computing speed and data, deep learning methods have been widely used in many fields. As mentioned above, image segmentation methods based on deep learning mainly include image segmentation methods based on FCN, image segmentation methods based on U-Net network, image segmentation based on U-Net++ network, and image segmentation based on Mask R-CNN network, etc. method. Essentially, FCN, U-Net and U-Net++ are all neural networks based on full convolution. Wenxuan Xue [6] et al. proposed a fundus retinal blood vessel segmentation algorithm based on a full convolutional network. The convolutional layer in the original U-Net network up-sampling was changed to the Inception module to increase the feature vector in the up- and down-sampling process. The multi-scale information and shape structure balance the depth and width of the network at the same time, and complete the segmentation of the fundus retinal blood vessel image. Mohsen Hajabdollahi [7] et al. proposed a method for segmenting the bleeding area of the Wireless capsule endoscope image. This method is based on the multi-layer perception (MLP) structure to select the appropriate color channel and classify it. The MLP structure is quantified so that the implementation does not require multiplication, and the average DICE score reaches 0.8403. Mask R-CNN is widely used in the field of image segmentation. Researchers mostly improve Mask R-CNN, mainly including the following aspects: Yang Zheng [8] et al., based on the Mask R-CNN network, proposed a cervical cell image segmentation method, in the network feature pyramid network (FPN)) Is transformed into DFPN by adding hole convolution to reduce the loss of image information and improve the accuracy of segmentation. Tao Feng [9] and others improved the Mask R-CNN network and proposed a chromosome image segmentation framework, Mask Oriented R-CNN, which introduces orientation information to segment chromosome images. The framework adds a branch of directed bounding box regression on the basis of Mask R-CNN; proposes a new intersection-union ratio (IoU) metric-angle-weighted intersection-union ratio (AwIoU); realizes a directed convolutional path structure , By copying the mask branch path and selecting the training path according to the direction information of the instance to reduce the interference in the mask prediction. Yaodong Wang [10] et al. modified the mask branch of the Mask RCNN network to improve the accuracy of pedestrian re-recognition and the effect of instance segmentation. Liu Fang [11] et al., based on the Mask-RCNN algorithm, used triple loss function and rotation angle regression technology to optimize and segment Oracle images. In the training data set, the accuracy of Oracle character recall is 82%, and the accuracy of detection and

recognition can reach 95%. Tianquan Wu [12] et al. proposed a corrosion detection method based on the improved Mask R-CNN model for the problem of corrosion detection of power equipment in substations. This method uses the residual network Resnet101 as the basic network of the model, and uses the improved Non-Maximum Suppression (NMS) algorithm to improve the detection accuracy of semantic segmentation. Experimental results show that the accuracy and recall of the improved model are better than the original Mask R-CNN network.

In summary, the image segmentation methods based on deep learning methods are very rich, of which Mask R-CNN is the most applied field, which reflects the strength of the Mask R-CNN network on the one hand. However, up to now, no researchers have applied Mask R-CNN network to the segmentation of the Wireless capsule endoscope image lesion image. Therefore, this paper applies the Mask R-CNN network to the Wireless capsule endoscope image lesion image segmentation to explore its feasibility.

2.3. Study on The Detection And Segmentation of Lesions in Wireless Capsule Endoscope Images

2.3.1. Image Detection Method of Wireless Capsule Endoscopy

Wireless capsule endoscope image lesion detection is very hot, and the current research method is mainly based on deep learning target detection technology to detect small intestinal lesions. Haya Alaskar [13] detects ulcers in Wireless capsule endoscope images based on convolutional neural networks, and extensively evaluates AlexNet and GoogLeNet networks. The evaluation results show that CNN has superior performance in Wireless capsule endoscope lesion detection and greatly exceeds A traditional machine learning method. Tomonori Aoki [14] et al., a deep convolutional neural network (CNN) system based on a single-lens multi-box detector to detect erosion and ulcers in Wireless capsule endoscope images. The trained model only takes 233 seconds to detect 10440 images. The average accuracy rate of erosion and ulcer is 90.8%. The above two methods are used to detect a small number of digestive tract lesions, and do not realize multi-category lesion detection.

At present, research on the detection of types of lesions based on the digestive tract is still relatively rare. For example, Xiao Z[15] and other YOLOv3 detection network improved, according to the characteristics of small intestine lesions, a method of labeling and feature detection was adopted, and the detection process was optimized by analyzing the WCE image of the small intestine and comparing experiments. At the same time, the original basic features of the YOLO v3 detection network are retained, and the improved detection network is further optimized and effectively verified, and finally the detection of 27 types of lesions is realized.

In summary, the convolutional neural network has superior performance in capsule lens lesion detection, but the current research mainly focuses on single-type or small-type detection, which cannot well meet the needs of hospitals.

2.3.2. Image Segmentation Method of Wireless Capsule Endoscopy

At this stage, the image segmentation method of digestive tract lesions is still relatively rare. For example, Qian Zhao [16] et al. proposed an adaptive non-parametric key point detection method based on multi-feature extraction and fusion, which realized the

effective segmentation of WCE video clips without losing the key information of the original video recording. Pedro M. Vieira [17], etc., proposed a method to automatically detect vasodilatation, which relies on the automatic selection of the region of interest, through the use of the image segmentation module based on the maximum posterior probability (MAP) method to select, which also A new accelerated version of Expectation Maximization (EM) algorithm is proposed. Using Markov random field and weighted boundary function, the spatial context information is modeled in the prior probability density function. Experimental results show that the method achieves a sensitivity and specificity value of more than 96%. Xiao Jia [18] et al. proposed an automatic segmentation method for blood regions in WCE images based on deep learning. First, according to the statistical characteristics of the probability of the color space histogram, the bleeding samples were divided into active and inactive categories. Then, for each subgroup, the blood area is highlighted through a Fully Convolutional Network (FCN). The experimental results on the clinical WCE data set prove the effectiveness of this method.

In summary, the current research on the segmentation of the capsule lens lesion image mainly focuses on the segmentation of a single or a few types of lesion images, and there is no research on multiple types of lesion images. Therefore, this paper tries to apply Mask R-CNN to the segmentation of multi-class lesion image with Wireless capsule endoscope.

3. Experimental Procedure

3.1. Dataset and Dataset Processing

Our experimental data is obtained from the Kvasir-Capsule[19] data set, and the statistical data of some selected data sets are shown in table 1. Keep 10% of the lesion pictures as a validation set for hyperparameter tuning and early stopping, and use the remaining data sets for training and test partitions at a ratio of 3:1. The names and numbers of different lesions are shown in table 1.

Table 1. Name and number of digestive tract lesions

Lesion name	Number of pictures
Duodenal ulcer	10
Reflux esophagitis	10
Colic ulcer	10
Small intestinal bleeding	10
Polyp of small intestine	10
Small intestinal ulcer	10
Intestinal erosion	10

Before model training, the data set needs to be processed. First, we use the LabelMe[2] tool to annotate the Wireless capsule endoscope image. Then, organize the data into a form recognized by the model and create a new folder train, which contains four subfolders. The following contents are stored in the subfolders, see table 2.

[2] Deep Learning Image Annotation Tool, https://github.com/CSAILVision/LabelMeAnnotationTool

Table 2. Training set content

folder	content
cv2_mask	json_to_dataset generates the label file in PNG format in the folder
json	Json file generated by LabelMe
pic	Folder generated by json_to_dataset
LabelMe_json	Original image after size standardization

3.2. Model Architecture

As shown in figure 2, the Mask R-CNN network structure consists of two parts, one is the backbone, and the main function is to extract the features of the picture. It uses a kind of FPN network, which can improve the accuracy of detection to a certain extent; the other part is the head, which is used for classification, box regression and mask prediction of each ROI.

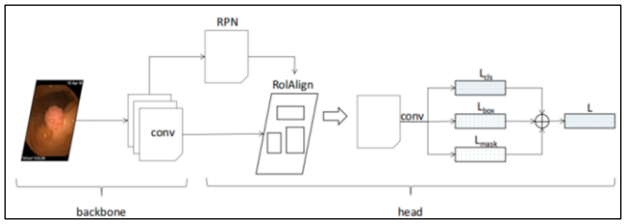

Figure 2. Mask R-CNN network structure diagram

4. Result

In order to effectively verify the experimental results, this experiment uses Google's tensorflow for evaluation. A PC configured with Inter(R) Xeon(R) CPU E5-1603 v4 @ 2.8Hz 32GB RAM and GPU 1660Ti is used to train and test network performance.

The loss function of Mask R-CNN is: $L=L_{cls}+L_{box}+L_{mask}$. This article uses the TensorBoard tool to visualize the output of each loss function. As shown in figure 3, for each loss curve.

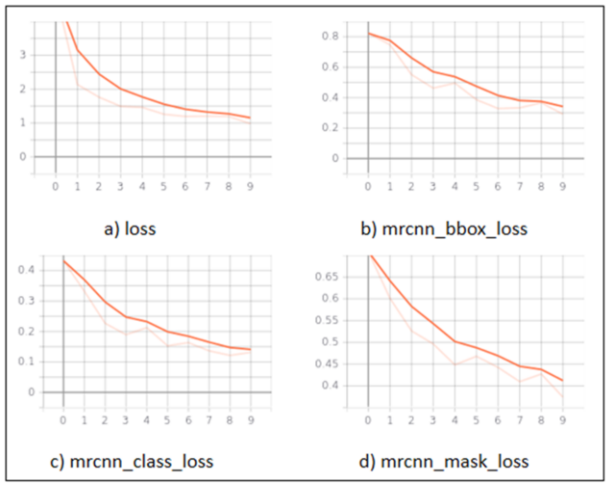

Figure 3. Loss, bbox_loss, class_loss and mask_loss curve

5. Conclusion

This paper innovatively applies the Mask R-CNN network to wireless capsule endoscope image segmentation, and collects 7 types of different digestive tract lesions for different digestive tract lesions by manually making a data set. The experimental results show that the Mask R-CNN network can effectively segment the Wireless capsule endoscope image, so as to achieve local enhancement of the digestive tract lesion tissue and surrounding area, which is beneficial to doctors to read the image.

6. Acknowledgement

The work was supported by the project plan of science and technology development center of the Ministry of Education (No. 2020hyb03002) and in part by the Jilin Science and Technology Development Plan Project (Project No. 20210201083GX) and in part by Jilin Provincial Department of Education Plan Project (Project No. JJKH20210632KJ).

References

[1] https://share.gmw.cn/health/2021-04/17/content_34766494.htm.
[2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
[3] Uijlings J R R, Van De Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154-171.
[4] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
[5] Moltz J H, Bornemann L, Dicken V, et al. Segmentation of liver metastases in CT scans by adaptive thresholding and morphological processing[C]//MICCAI workshop. 2008, 41(43): 195.
[6] Xue Wenxuan,Liu Jianxia.An improved U-Net method for fundus blood vessel segmentation[J].Electronic Design Engineering,2021,29(20):165-168.
[7] Hajabdollahi M, Esfandiarpoor R, Soroushmehr S M, et al. Segmentation of bleeding regions in wireless capsule endoscopy images an approach for inside capsule video summarization[J]. arXiv preprint arXiv:1802.07788, 2018.
[8] Zheng Yang,Liang Guangming,Liu Renren.Segmentation of Cervical Cell Image Based on Mask R-CNN[J].Computer Times,2020(10):68-72.
[9] Feng Tao,Chen Bin,Zhang Yuefei.Chromosome image segmentation framework based on improved Mask R-CNN[J].Journal of Computer Applications,2020,40(11):3332-3339.
[10] Wang Yaodong. Research on Pedestrian Re-identification Based on Mask RCNN Neural Network[D]. Xi'an University of Science and Technology, 2020.
[11] Liu Fang,Li Huabiao,Ma Jin,Yan Sheng,Jin Peiran.Automatic detection and recognition of Oracle rubbings based on Mask-RCNN[J/OL].Data Analysis and Knowledge Discovery: 1-12[2021-10-30] .http://kns.cnki.net/kcms/detail/10.1478.G2.20210906.1332.002.html
[12] Wu Tianquan,Dai Meisheng,Yang Gang,Chen Weijie,Chen Guiping,Zhang Pengju,Gou Xiantai,Zhou Weichao.Corrosion Recognition of Power Equipment Based on Improved Mask-RCNN Model[J].Electric Power Information and Communication Technology,2021,19(04):25-30.
[13] Alaskar H, Hussain A, Al-Aseem N, et al. Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images[J]. Sensors, 2019, 19(6): 1265.
[14] Aoki T, Yamada A, Aoyama K, et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network[J]. Gastrointestinal endoscopy, 2019, 89(2): 357-363. e2.
[15] Xiao Z, Feng L N. A Studyon Wireless Capsule Endoscopy for Small Intestinal Lesions Detection Basedon Deep Learning Target Detection [J]. IEEE Access, 2020, 8: 159017-159026.

[16] Zhao Q, Meng M Q H. An abnormality based WCE video segmentation strategy[C]//2010 IEEE International Conference on Automation and Logistics. IEEE, 2010: 565-570.
[17] Vieira P M, Silva C P, Costa D, et al. Automatic segmentation and detection of small bowel angioectasias in WCE images[J]. Annals of biomedical engineering, 2019, 47(6): 1446-1462.
[18] Jia X, Meng M Q H. A study on automated segmentation of blood regions in wireless capsule endoscopy images using fully convolutional networks[C]//2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017: 179-182.
[19] Smedsrud P H, Thambawita V, Hicks S A, et al. Kvasir-Capsule, a video capsule endoscopy dataset[J]. Scientific Data, 2021, 8(1): 1-10.