Malignancy Suspicious Region Guided Deep Neural Networks for Gastric Ulcer Classification

Xiaoyan Zheng[1], Zeng Zeng[†2], Can Ma[1], Qing Chang[3], Ziyuan Zhao[2] and Xulei Yang[2]

***Abstract*— Malignant transformation of gastric ulcer can result in gastric cancer, hence an accurate gastric ulcer classification method is of vital importance. Despite marvelous progress has been achieved in recent years, there are still many challenges in diagnosis of gastric ulcer. In this paper, we propose a mechanism to mimic gastroenterologist's behaviours based on deep learning techniques, by integrating the segmented malignancy suspicious masks with gastroscopic images for gastric ulcer classification, which instructs the model to focus on the area where symptoms occur for gastric ulcer diagnosis. Specifically, a U-Net-type deep neural network is built to segment the suspicious pathological regions from gastroscopic images, then the segmented regions are treated as an attention channel of gastroscopic images for the gastric ulcer classification by a ResNet-type deep neural network. Experiments on a real gastroscopic dataset with 900+ patient cases demonstrate that our proposed approach achieves much better performance for gastric ulcer diagnosis, compared with standard method with only gastroscopic images.**

I. INTRODUCTION

Although gastric ulcer is a common disease, malignant gastric ulcer can lead to gastric cancer, which is one of the top five frequently diagnosed cancers in the world [1], [2]. At present, gastroscope is the most effective and direct way to detect malignant gastric ulcers. Therefore, it is of great significance to classify benign and malignant gastric ulcers accurately based on gastroscopic images.

Generally, gastroenterologists usually diagnose gastroscopic images based on personal experience without any auxiliary tools. However, the digestive endoscopy center has a large number of gastroscopic data every day, gastroenterologists' workload is very heavy. Moreover, the endoscopic imaging of early gastric cancer and ulcerative advanced gastric cancer can easily be confused with benign gastric ulcers, which may cause missed diagnosis or diagnostic errors. Thus, an automatic way is highly desirable for gastric ulcer diagnosis.

In recent years, artificial intelligence related technologies have been widely used in the biomedical field [3], [4], [5], [6], [7], [8], [9]. Among them, medical image analysis methods based on deep learning have also shown superior performance and better robustness [10], [11]. Great progress has also been made in the analysis and processing of gastroscopic images by deep learning. Coimbra et al.[12] performed a good classification of bleeding, polyps and ulcers in the gastrointestinal tracts by extracting the color, texture and shape features of images. Pan *et al*.[13] used a probabilistic neural network to detect the bleeding areas on these endoscopic images. Georgakopoulos *et al*.[14] proposed a novel weakly-supervised learning method based on CNN architecture to realize the detection of inflammatory gastrointestinal lesions in wireless capsule endoscopic videos. Whereas, accurate diagnosis of gastric ulcer based on gastroscopic images still remains a challenge.

† Corresponding Author. [1]School of Microelectronics, Shanghai University, Shanghai, China, 200444. Email: {zxy20000421, oohmc}@shu.edu.cn. [2]Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632. Email: {zengz, zhaoz, yang_xulei}@i2r.a-star.edu.sg. [3]Clinical Research Center, Jiading District Central Hospital Affiliated Shanghai University of Medicine & Health Sciences, Shanghai, 201800, China.

Since gastroscopic images usually have problems such as low imaging resolution and trouble in pathological areas recognition, we adopt an attention mechanism [15] to enable the neural network to learn the main features of gastroscopic images more effectively. We first train a deep neural network based on U-Net [16] to segment the malignancy suspicious regions within the gastroscopic images, then integrate the segmented masks into the gastroscopic images for gastric ulcer classification by using a deep neural network based on ResNet [17]. Compared with the standard method using only gastroscopic images, our proposed approach achieves much better accuracy on a real gastroscopic dataset with 900+ patient cases.

II. RELATED WORK

A. Classification of Gastric Ulcers

At present, the classification methods of benign and malignant gastric ulcer mainly rely on analysis of the morphological characteristics of gastroscopic images, or take part of gastric ulcer tissue for slicing, and finally confirm whether it is malignant gastric ulcer through pathological methods [18]. Malignant gastric ulcers are generally irregular in shape and redden on the mucosa surface under gastroscope, usually accompanied by erosion, bleeding and other pathological features [18] . However, due to the complex and changeable situation between benign gastric ulcer lesions and malignant gastric ulcer lesions, as well as cancer and precancerous lesions, there is no precise diagnostic standard now. Therefore, in the vast majority of cases, the diagnosis of malignant gastric ulcer is a subjective judgment relative to the characteristics of normal gastroscopic images. A more systematic and accurate computer-aided diagnosis is required to improve the accuracy of the classification of benign and malignant gastric ulcers under gastroscope.

B. Researches of Gastric Ulcers

With the widespread application of deep learning in the medical field [3], [4], [19], there are many methods have been

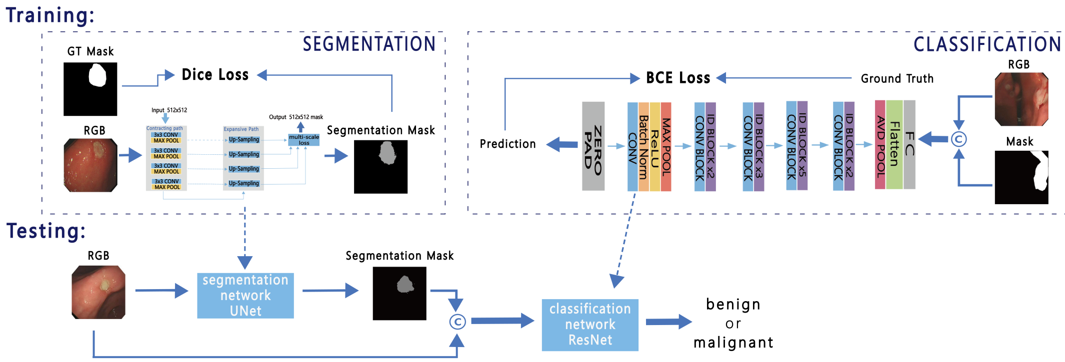

Fig. 1. The overall flowchart of our proposed approach

proposed for the diagnosis of malignant gastric ulcer. Huang et al.[20] trained and classified gastric ulcer images based on the RseNet50 of the Fastai framework and the VGG16 [21] model based on the Keras framework, and used binary classification model to show great ability in finding gastric ulcer lesions and distinguishing benign and malignant ulcers. Similarly, Sun et al.[22] utilized VGGnet, ResNet V2 [23] and R-FCN [24] to classify benign and malignant gastric ulcers on esophagogastroduodenoscopy images respectively. By changing the fully connected layer of the network, it is directly connected to two output nodes corresponding to the benign and malignant type labels, they all showed performance comparable to the endoscopic diagnosis made by doctors. Wang et al.[24] migrated the collected gastroscopic images to the trained VGG16 and ResNet34 networks through transfer learning, which showed a superb accuracy of gastric cancer image recognition.

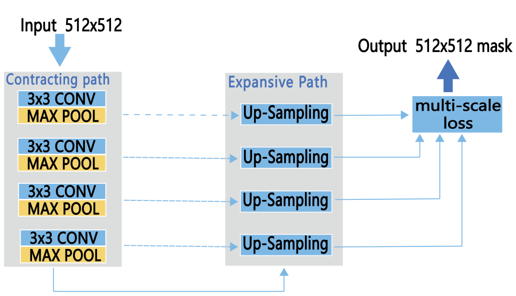

Fig. 2. Our segmentation model – U-Net with multi-scale

III. Methodology

In this section, we first show the overall diagram of the proposed approach, then discuss the segmentation network and classification network respectively.

A. Overall Flowchart

The overall flowchart of our work is shown in Figure 1, which consists of the following main steps:

1) Guided by gastroenterologist, we first label the suspicious pathological regions in the gastroscopic images. The labeled regions are served as the ground truth for the deep segmentation model.
2) We then train a variant of U-Net to perform the segmentation on gastroscopic images. Through segmentation, we obtain masks containing malignancy suspicious regions, which are served as the attention channel for the classification model.
3) We finally propose a special classification model–ResNet with an attention channel as input. The addition of this attention channel allows ResNet to pay more attention to the feature learning of pathological regions in gastroscopic images.

B. Segmentation–U-Net

The U-Net proposed by Ronneberger et al.[16] was originally aiming to solve the problems of medical image segmentation. Medical image segmentation often faces the issues of 1) small amount of samples and 2) samples from multimodal. While U-Net combines low-resolution and high-resolution information, it performs well in medical image segmentation. The structure of U-Net includes two parts: contracting path and expanding path. Among them, the contracting path is used for feature extraction to obtain context information, and the expanding path adopts up-sampling to obtain accurate positioning. The specific U-Net structure is shown in the Figure 2. The left half (contracting path) consists of four 3×3 convolution kernels (no padding convolution), each followed by a ReLU function for activation and a 2×2 max pooling operation for down-sampling. In the right half (expanding path), each step includes a up-sampling layer for the feature map up-sampling. In this study, we propose a modified U-Net architecture with multi-scale loss function [25] to perform suspicious pathological areas segmentation from gastroscopic images. The final output loss of U-Net is a combination of different individual loss resolutions which tend to extract more meaningful features by minimizing the loss at each individual resolution scale.

C. Classification–ResNet

Recently, Convolutional Neural Networks (CNN) [26] have achieved great success in the field of medical imaging [3], [4], [27], especially the classification of medical

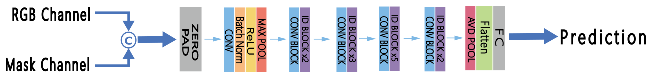

Fig. 3. Our proposed ResNet structure diagram.

images [28], [10], [29]. Among them, Deep Residual Network (ResNet) is created to solve convolution neural network gradient disappearance and degradation problems with the increase of depth and won the first place in the classification task of ImageNet [30] competition in 2015. ResNet have been successfully applied on classification problems in computer vision. The main contribution of ResNet is to propose the concept of residual block. The emergence of residual network allows neural network to be greatly deepened to achieve high accuracy. In our study, we employ ResNet50 network, containing 50 layers, to learn more complicated feature representations from gastroscopic images. The network structure is shown in Figure 3. We additionally concatenate an attention channel–output of U-Net segmentation to the input layer of the model. In other words, the input of our gastric ulcer classification task for training and testing includes not only raw gastroscopic images, but also semantic segmentation images of suspicious pathological regions obtained by U-Net segmentation model. As a result, the input of the proposed ResNet changes from traditional ”RGB 3-channels” to ”RGB+mask 4-channels” to improve model learning efficiency and classification accuracy.

IV. Experiments

A. Dataset

The endoscopic images data used in this paper comes from the Digestive Endoscopy Center of Jiading District Central Hospital Affiliated Shanghai University of Medicine & Health Sciences. All endoscopic images contained in this dataset are collected by experienced gastroenterologist. The dataset consists of 954 gastroscopic pictures, including 586 benign gastric ulcer pictures, 368 malignant gastric ulcer pictures. Two image samples (one is benign, the other is malignant) from the dataset are shown in Figure 4.

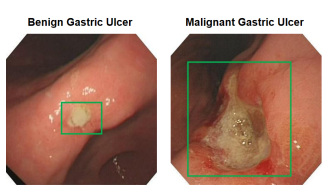

Fig. 4. The suspicious regions (green boxes): under normal circumstances, the pathological areas of benign gastric ulcer are small, while those of malignant gastric ulcer are large and accompanied by erosion and bleeding.

The resolution size of these images is 512×512, the bit depth of picture pixels is 24, the horizontal resolution and vertical resolution are 96dpi, and the image format is stored as JPG. The classification of gastric ulcer in each gastroscopic image has been marked by the attending gastroenterologist in the Department of Gastroenterology. Specially, we processed each image - using LabelMe to label suspicious pathological regions of gastric ulcers.

B. Experimental Settings

In this study, we use the pytorch environment built by GPU to conduct experiments. Each image in the dataset has a uniform size of 512×512. We divide the dataset into a training set and a testing set, *i.e.*, 801 images for training, and the remaining 153 for testing. Considering the small quantity of images in the dataset, we perform extensive data augmentation in both segmentation and classification models.

1) Segmentation Results: The segmentation results by the proposed multi-scale U-Net model on the gastroscopic images are promising. Most of the segmented masks match the ground truth well. Figure 5 shows the segmentation results of two selected images by the U-Net segmentation model. We can see that our segmented masks (middle) are close to the ground truth (right). That indicates that the proposed segmentation model successfully find the suspicious pathological regions of gastric ulcers.

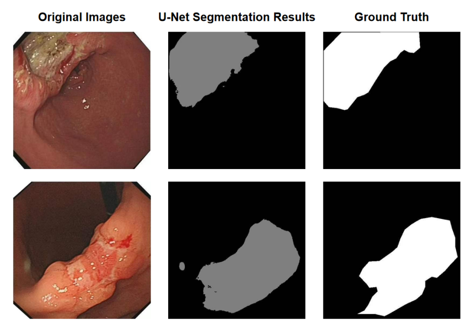

Fig. 5. The segmented masks (middle) obtained by U-Net are close to ground truth (right)

2) Classification Results: Through the classification results of the model, we can get the following four types of indicators:

TP: True Positive; the number of malignant gastric ulcer images that are correctly classified;

FP: False Positive; the number of malignant gastric ulcer images that are incorrectly classified;

TN: True Negative; the number benign gastric ulcer images that are correctly classified;

FN: False Negative; the number of benign gastric ulcer images that are incorrectly classified.

Through the above four classification indicators, we propose three model evaluation metrics: Sensitivity, Specificity and Accuracy as follows:

$$Sensitivity = TP/(TP + FN), \tag{1}$$

$$Specificity = TN/(TN + FP), \tag{2}$$

$$Accuracy = (TP + TN)/(TP + FN + TN + FP). \tag{3}$$

The higher values of accuracy, sensitivity, and specificity, the better performance of the classification model.

TABLE I

PERFORMANCE COMPARISONS BETWEEN OUR PROPOSED APPROACH WITH MASK ATTENTION AND STANDARD APPROACH.

Methods	Sensitivity	Specificity	Accuracy
our proposed method	74.6	87.9	82.6
standard ResNet	47.5	85.7	69.8

To show the effectiveness of adding an attention channel to input on the classification performance, we compare the performance of two different ResNet models: our proposed method with attention channel and standard ResNet without attention channel. Both models are trained from scratch under the same setting. The comparison results of the two models on the test set are shown in Table 1. From the table, we can see that the overall classification performance of the model is commendably improved by adding attention channel. Especially, the sensitivity value of our proposed approach is much higher than that of standard approach, which indicates that the classification accuracy of malignant gastric ulcer has been significantly improved. Nevertheless, by virtue of our insufficient amount of gastroscopic images, there is still room for improvement in classification performance.

V. CONCLUSION

We propose a novel ResNet-based gastroscopic classification model with an attention channel, which guides the model to focus on the suspicious pathological region in gastroscopic images. By integrating the mask segmented from U-Net segmentation model, ResNet takes "RGB+mask 4-channels" input to improve model efficiency and classification accuracy. Experiments show that our proposed method has superior performance in terms of classification accuracy of benign and malignant gastric ulcer.

ACKNOWLEDGMENT

The authors would be grateful to Jiading District Central Hospital Affiliated Shanghai University of Medicine & Health Sciences for the support on dataset and annotations.

REFERENCES

[1] Christopher P Wild, Bernard W Stewart, and C Wild, *World cancer report 2014*, World Health Organization Geneva, Switzerland, 2014.

[2] Lindsey A Torre, Rebecca L Siegel, Elizabeth M Ward, and Ahmedin Jemal, "Global cancer incidence and mortality rates and trends—an update," *Cancer Epidemiology and Prevention Biomarkers*, 2016.

[3] Dinggang Shen, Guorong Wu, and Heung-Il Suk, "Deep learning in medical image analysis," *Annu Rev Biomed Eng*, 2017.

[4] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez, "A survey on deep learning in medical image analysis," *Med Image Anal*, 2017.

[5] Xiaodong Qu, Qingtian Mei, Peiyan Liu, and Timothy Hickey, "Using eeg to distinguish between writing and typing for the same cognitive task," in *Int Conf on Brain Function Assessment in Learning*, 2020.

[6] Xiaodong Qu, Peiyan Liu, Zhaonan Li, and Timothy Hickey, "Multi-class time continuity voting for eeg classification," in *Int Conf on Brain Function Assessment in Learning*, 2020.

[7] Jiapan Gu, Ziyuan Zhao, Zeng Zeng, Yuzhe Wang, Zhengyiren Qiu, Bharadwaj Veeravalli, Brian Kim Poh Goh, Glenn Kunnath Bonney, Krishnakumar Madhavan, Chan Wan Ying, et al., "Multi-phase cross-modal learning for noninvasive gene mutation prediction in hepatocellular carcinoma," in *EMBC*. IEEE, 2020.

[8] Peisheng Qian, Ziyuan Zhao, Haobing Liu, Yingcai Wang, Yu Peng, Sheng Hu, Jing Zhang, Yue Deng, and Zeng Zeng, "Multi-target deep learning for algal detection and classification," in *EMBC*. IEEE, 2020.

[9] Ziyuan Zhao, Kaixin Xu, Shumeng Li, Zeng Zeng, and Cuntai Guan, "Mt-uda: Towards unsupervised cross-modality medical image segmentation with limited source labels," in *Med Image Comput Comput Assist Interv (MICCAI)*. Springer, 2021.

[10] Ziyuan Zhao, Kartik Chopra, Zeng Zeng, and Xiaoli Li, "Sea-net: Squeeze-and-excitation attention net for diabetic retinopathy grading," in *ICIP*. IEEE, 2020.

[11] Ziyuan Zhao, Zeng Zeng, Kaixin Xu, Cen Chen, and Cuntai Guan, "Dsal: Deeply supervised active learning from strong and weak labelers for biomedical image segmentation," *IEEE J Biomed Health Inform*, 2021.

[12] M Coimbra, Paulo Campos, and JP Silva Cunha, "Topographic segmentation and transit time estimation for endoscopic capsule exams," in *ICASSP*. IEEE, 2006.

[13] Guobing Pan, Guozheng Yan, Xiangling Qiu, and Jiehao Cui, "Bleeding detection in wireless capsule endoscopy based on probabilistic neural network," *J Med Syst*, 2011.

[14] Spiros V Georgakopoulos, Dimitris K Iakovidis, Michael Vasilakakis, Vassilis P Plagianakos, and Anastasios Koulaouzidis, "Weakly-supervised convolutional learning for detection of inflammatory gastrointestinal lesions," in *IST*. IEEE, 2016.

[15] Volodymyr Mnih, Nicolas Heess, Alex Graves, et al., "Recurrent models of visual attention," *Adv Neural Inf Process Syst*, 2014.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *IEEE CVPR*, 2016.

[18] Qinhua Li, "Clinical diagnosis and treatment analysis of different types of digestive ulceration," *Chinese Tradit. Chin. Med.*, 2010.

[19] Kaixin Xu, Ziyuan Zhao, Jiapan Gu, Zeng Zeng, Chan Wan Ying, Lim Kheng Choon, Thng Choon Hua, and Pierce KH Chow, "Multi-instance multi-label learning for gene mutation prediction in hepatocellular carcinoma," in *EMBC*. IEEE, 2020.

[20] Li HUANG, Yanxia LI, Lianlian WU, Shan HU, Yiyun CHEN, Jun ZHANG, Ping AN, and Honggang YU, "Artificial intelligence-assisted diagnosis system of benign and malignant gastric ulcer based on deep learning," *Chinese Journal of Digestive Endoscopy*, 2020.

[21] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[22] Jee Young Sun, Sang Won Lee, Mun Cheon Kang, Seung Wook Kim, Seung Young Kim, and Sung Jea Ko, "A novel gastric ulcer differentiation system using convolutional neural networks," in *CBMS*. IEEE, 2018.

[23] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *AAAI*, 2017.

[24] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Adv Neural Inf Process Syst*, 2016.

[25] Shuang Leng, Xulei Yang, Xiaodan Zhao, Zeng Zeng, Yi Su, Angela S Koh, David Sim, Ju Le Tan, Ru San Tan, and Liang Zhong, "Computational platform based on deep learning for segmenting ventricular endocardium in long-axis cardiac mr imaging," in *EMBC*. IEEE, 2018.

[26] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998.

[27] Ziyuan Zhao, Xiaoyan Yang, Bharadwaj Veeravalli, and Zeng Zeng, "Deeply supervised active learning for finger bones segmentation," in *EMBC*. IEEE, 2020.

[28] Ziyuan Zhao, Kerui Zhang, Xuejie Hao, Jing Tian, Matthew Chin Heng Chua, Li Chen, and Xin Xu, "Bira-net: Bilinear attention net for diabetic retinopathy grading," in *ICIP*. IEEE, 2019.

[29] Peisheng Qian, Ziyuan Zhao, Cong Chen, Zeng Zeng, and Xiaoli Li, "Two eyes are better than one: Exploiting binocular correlation for diabetic retinopathy severity grading," in *EMBC*. IEEE, 2021.

[30] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al., "Imagenet large scale visual recognition challenge," *Int J Comput Vis*, 2015.