DOI: 10.53469/jtpes.2023.03(12).06

Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN

Bo Liu1, Xinyu Zhao2 ,Hao Hu3,Qunwei Lin4,Jiaxin Huang5

[1] Software Engineering,Zhejiang University ,Hangzhou China

[2]Information Studies,Trine University,Phoenix, USA

[3]Software Engineering,Zhejiang University ,Hangzhou China

[4]Information Studies,Trine University,Phoenix, USA

[5]Information Studies,Trine University,Phoenix, USA

Abstract: ***Esophageal cancer is a common malignant tumor in daily life, which seriously affects human health. Esophageal cancer in China. The incidence rate is among the highest in the world, and there are a large number of new cases of esophageal cancer every year. At present, the diagnosis of esophageal cancer is mainly based on the use of electronic gastroscopy, which reflects the observed situation on the fluorescent screen and conducts detection through the fluorescent screen. With the increasing number of patients, the pressure and intensity of the doctor's work are getting greater and greater[1]. From the perspective of molecular level, the occurrence and development of esophageal cancer, like other cancers, are related to the activation of proto-oncogenes and the inhibition of anti-apoptosis genes. CBAM Faster R-CNN is proposed to solve the problems such as the esophageal region is not obvious and the background region occupies a large proportion in the feature map obtained by the backbone network of Faster R-CNN in barium meal imaging. CBAM is added to the convolutional attention module CBAM on the basis of the original Faster R-CNN model. To enhance the saliency of the features of the esophageal region in the feature map. CBAM Faster R-CNN model was used to train the training set after data enhancement, and Recall, Precision and AP values were used for evaluation and analysis.***

Keywords: Esophageal Cancer; Anti-Apoptotic Gene; Faster RCNN; Object Detection.

1. INTRODUCTION

At present, according to relevant data, the prevalence rate of esophageal cancer in China is in the forefront of the world. Domestic and foreign experts have also done related research on esophageal cancer. A few years ago, researchers proposed one Automatic recognition and classification system of early esophageal cancer cells based on PC, this system uses traditional methods to preprocess, segment, and extract features of esophageal cancer cells to obtain the relevant information of cancer transformation. The doctor combines this information to make a diagnosis of the patient. In recent years, the rise of machine learning, artificial intelligence, and computer vision have been further developed. Computer vision is mainly the use of computer simulation human image related processing, to obtain valuable information in the picture. Computer vision has been widely used in medical image processing, industrial robots, image monitoring, unmanned driving and many other fields, and the effect is also remarkable [2].

As a basic field of computer vision application, object detection mainly combines object segmentation and object recognition, and its recognition accuracy, recognition efficiency and positioning accuracy are the main performance indicators of the whole system. In recent years, with the annual PASCAL VOC Challenge held, more and more teams have participated. Every year, the participating teams will put forward some advanced algorithms or propose improvements on existing algorithms, and it is because of their efforts that object detection has been rapidly developed. With the rise of machine learning and deep learning, deep convolutional neural networks have been widely used in image classification and object detection. In this paper, the object detection algorithm based on deep learning is studied, and the Faster RCNN algorithm is systematically studied.

2. RELATED WORK

SOMMEN F V D et al. proposed Computer Aided Diagnosis (CAD) for esophageal cancer detection. Local color and texture features are calculated using the esophageal endoscopic images based on Gabor filtering, and the features are classified by a trained Support Vector Machine (SVM) classifier. The recall and accuracy rates were 95% and 75%, respectively. SCHOON EJ et al. proposed a CAD system for detecting early esophageal endoscopy images, which used specific textures, color filters, and machine learning techniques to achieve a sensitivity and specificity of 0.86 in 100 esophageal endoscopy images of 44 patients with esophageal cancer And 0.87. The CAD system for esophageal cancer mentioned in the above literature mainly focuses on the study of esophageal cancer cell images and esophageal endoscopy images. At present, there is no relevant study on the esophageal cancer detection system based on barium meal imaging at home and abroad [3]. The meal imaging refers to the development process of the stomach and duodenum of the patient after swallowing pasted barium sulfate through the esophagus with steel agent. The lock-in imaging can well show the lesion site, mucosal changes and the length of the cancer. The lesion can be located and the date can be determined. In the actual clinical study, X-ray meal and CT examination have their own advantages, and the combination of the two is conducive to the diagnosis and clinical treatment of esophageal cancer.

In recent years, the rise of machine learning and artificial intelligence has promoted the further development of computer vision. Achieve a level of performance comparable to or exceeding that of humans in tasks such as object classification, detection and segmentation. At present, computer vision has been maturely applied in medical image processing, document analysis port, military field, unmanned driving and other hot fields[4-5]. In 2014, GIRSHICK R et al. proposed the Convolutional Neural Network based on Region of Interest (ROI) and convolutional neural network. The algorithm is the first masterpiece of object detection algorithm based on deep learning, and it has also inspired more excellent scholars' research interest in object detection algorithm. R-CNN first selects 2000 candidate regions containing suspicious targets from the input image, and then sends these 2000 candidate regions into the CNN model to obtain feature vectors. These feature vectors are used as the inputs of SVM and border regression model, and the probability of ROI belonging to a certain type of object is obtained through SVM. The coordinate position of the target to be detected is obtained through the border regression model. Although R-CNN implements the object detection algorithm based on deep learning, in the process of ROI feature vector extraction, Parameter sharing of convolutional layer cannot be realized[6]. In 2015, GIRSHICK R, inspired by SPPNET, proposed the ROI-based deep learning object detection algorithm Fast R-CNN IV, which for the first time proposed the ROI pooling layer. The convolution layer used for ROI feature vector extraction realizes parameter sharing, and classification and regression tasks are added to the network at the same time. But its disadvantage is that it still uses Selective Search algorithm in the extraction of ROI, which is run on the CPU. In 2016, Ren Shaoqing et al[7-8]. adopted RegionProposal Network (RPN) to extract ROI, and proposed the target detection algorithm Faster R-CNMing. There are three basic steps for target detection: The ROI extracted from original images, ROI feature extraction, ROI classification and border regression are all unified into the same deep network framework, which truly realizes the end-to-end learning of target detection tasks. Therefore, the Faster R-CNN, once proposed, has obvious advantages over other networks in medical image detection[9].

In this paper, cases of esophageal cancer patients are used as experimental data set, and the Faster R-CNN network is used for training and testing. In order to further improve the detection performance, "Convolutional Block Attention Module," convolutional block attention module, The evaluation results of CBAM integrated into the backbone network of Faster R-CNN show that the proposed method is significantly better than the original network of Faster R-CNN.

3. METHODOLOGY

3.1 Model Architecture

In essence, esophageal cancer detection algorithm is mainly divided into two steps, the first step is to extract ROI, the second step is to extract ROI target classification and border regression. The esophageal cancer lesion area detection algorithm CBAM Faster R-CNN used in this paper is formed by introducing CBAM module on the basis of Faster R-CNN network. The flow of the algorithm is shown in Figure 1. First, the food-lock contrast image is input to the backbone (VGG16) network introduced into the CBAM module to obtain a series of feature maps, and then the RPN network generates candidate region Proposals. The feature vector corresponding to these Proposals is obtained from the feature map. Finally, these vectors are input to the ROI pooling layer to complete target classification and border regression operation. The result is the candidate region detection box and the probability that the detection box is the target class.

Network and ROI Pooing layer. CBAM backbone network is composed of a lightweight convolutional neural network attention module, namely CBAM, convolutional layer and pooling layer. CBAM combines the Channel attention mechanism and the Spatial attention mechanism. By introducing the attention mechanism, the output information can be more focused on the more critical and useful information, and the interference of irrelevant information can be reduced or even eliminated, so as to improve the accuracy of the model. The channel attention module is shown in Figure 2. The input of the channel attention module is the feature graph F, whose height is H, width is w, and the number of channels in the feature graph is C. First, the average and maximum pooling operations based on the height and width of F are respectively adopted for F.

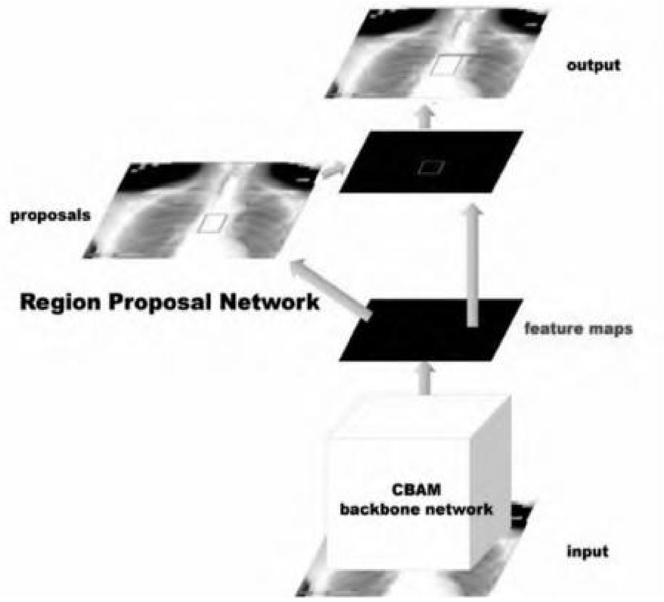

Figure 1: Network structure of CBAM Faster R-CNN

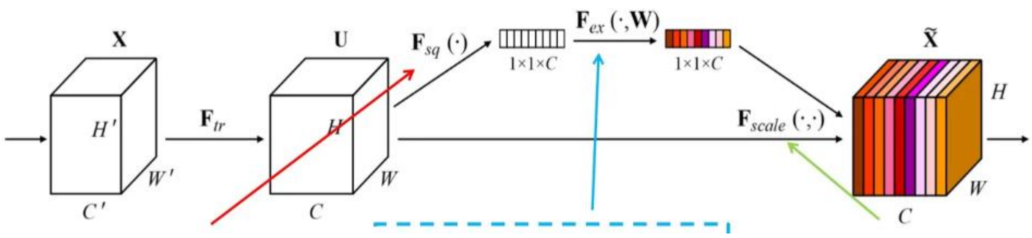

Figure 2: Channel attention module

AvgPoolw and MaxPool are obtained respectively, and then AvgPool and MaxPool are input into the fully connected layer with shared parameters of the two layers. The number of output units of the first fully connected layer is the quotient of the number of channels in the feature graph and ratio (ratio=8 is set in this paper). The number of output units of the second fully connected layer is the number of channel in the feature graph. Then, the average pooling output AvgPool obtained after the full connection layer and the maximum pooling output MaxPool are summed and the Sigmoid activation function is used to obtain the channel attention diagram M.

3.2 DataSet Introduction

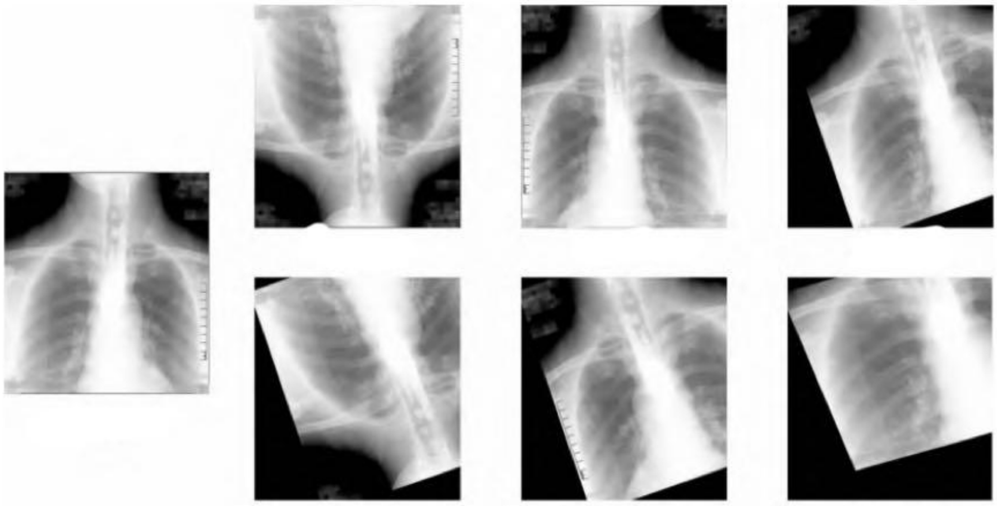

Figure 3: Data augmentation of hard examples

In order to enhance the detection ability of the model for samples with high or low brightness value of the target region, affine transformation is first implemented for the small number of difficult samples with high or low brightness value of the focus region in the training set. The difficult samples after affine transformation are combined with the original training set to form a new training set, and then the new training set is enhanced by horizontal flipping and vertical flipping.

In this experiment, a total of 1166 images from 40 cases in the test set were evaluated. For each image, the rectangle box of the test result and its gold standard rectangle box were calculated by the Intersection over Union (IOU). The definition of IOU is as follows:

$$IOU = \frac{Area(a) \cap Area(b)}{Area(a) \cup Area(b)}, \tag{1}$$

Where Area(a) represents the Area of the gold standard rectangular box and area (b) represents the area of the predicted rectangular box. If the IOU ratio of the two is greater than or equal to the TP threshold, the model is considered to successfully predict the lesion area of the image, FP (true positive number) plus one; Otherwise, the lesion area of the image cannot be predicted, FP(number of false positives) plus one. And calculate according to TP, FN, PF. Recall, Precision, AP. Recall is also called Recall rate, which indicates how many samples of esophageal cancer patients have been correctly predicted among all samples of esophageal cancer patients. The higher the recall value, the lower the missed detection rate of the model. Precision, also known as Precision rate, indicates how many samples of esophageal cancer patients are predicted correctly among all predicted results. A higher precision value also indicates a lower false detection rate of the model. AP is the area under the PR curve enclosed by the two dimensions of sitting mark Recall and sitting Precision.

The greater the value of AP, the better the comprehensive performance of the detection model. Recall, Precision is calculated as follows:

$$Recall = \frac{TP}{TP + FN}, \tag{2}$$

$$Precision = \frac{TP}{TP + FP}. \tag{3}$$

3.3 Results

The so-called CBAM Faster R-CNN model is implemented on the Windows platform using TensorFlow1.13. In the training process, the momentum value, initial learning rate, weight decay and maximum number of iterations are set to 0.9,0.001,0.0005 and 70,000 respectively. In order to comprehensively evaluate the performance of the algorithm, the experiments before and after the model improvement were trained on the training samples with enhanced data and tested on the same test samples[10]. Compared with the original FasterR-CNN algorithm, the Recall, Precision and AP indexes of the improved model are significantly improved. At present, there is no relevant research on esophageal cancer detection system based on barium meal images at home and abroad, and the algorithm in this paper is mainly compared with the original FasterR-CNN algorithm. Figure 7 shows the detection results of the esophageal cancer detection algorithm proposed in this paper on two barium meal images of esophageal cancer. Among them, the blue line represents the esophageal cancer lesion outline drawn by the radiologist circle, that is, the gold standard area, while the green line represents the detection results obtained by the detection algorithm. From the detection results, it can be seen that the algorithm in this paper is closer to the gold standard of physicians.

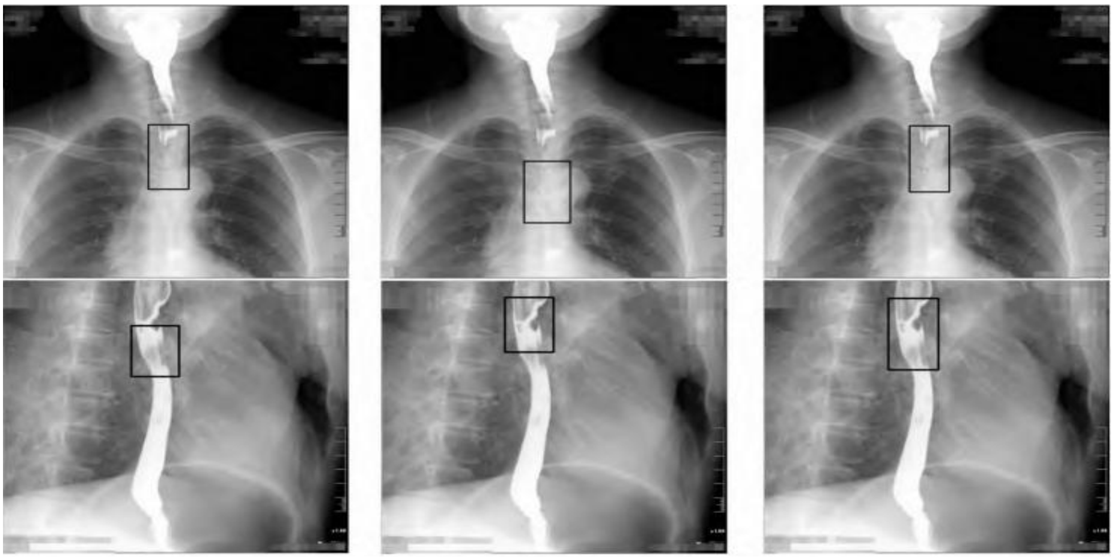

Figure 4: Esophageal cancer image detection results

4. CONCLUSION

With the rise of machine learning, artificial intelligence, and deep learning, Ross B. Gurshick et al., 2014 an object detection algorithm based on deep learning (RCNN) is proposed, which makes a breakthrough in object detection and triggers a research upsurge of object detection based on deep learning.In order to improve the detection ability of esophageal cancer lesions, CBAM module is added to the backbone network of the original Faster R-CNN model in this paper, and the difficult samples in the training set are enhanced with different multifold data compared with ordinary samples, so as to improve the detection ability of the model for esophageal cancer lesions with high or low brightness value. The experimental results show that after data enhancement, the threshold value of IOU is less than 0. At 5, the AP value of the model is increased by 1. 26% ~ 5. 61%. After the improvement of the model, the AP value of the model is increased by 0. 1% ~ 5. 0%. The above experimental results show that the detection efficiency of esophageal cancer lesions by CBAM Faster R-CNN designed in this paper is significantly improved compared with the Faster R-CNN algorithm[11-12].

Because the overall training process of RCNN algorithm is Faster, the stochastic gradient descent method is used for training. In order to share the convolutional network calculation among regions of interest, the stochastic gradient descent is created in small batches and layers. In the fine-tuning stage of specific samples, the ratio of positive and negative samples in each random gradient small batch sampling is set to 1:3 for random selection to ensure that the proportion of positive samples reaches 25%. In this paper, the online difficult sample mining mechanism is introduced into the Faster RCNN algorithm, so that the Faster RCNN algorithm can calculate the loss of the extracted border of the region of interest in the training process. Then, according to the calculated loss value, the first k regions of interest are selected as positive and negative samples for training, so as to remove this proportion parameter[13-23].

In conclusion, a large number of experimental results were compared and analyzed on the esophageal cancer data set to verify the detection results on the esophageal cancer data set after the introduction of online difficult sample mining (OHEM) by Faster RCNN algorithm, indicating that the combination of the two can improve the average precision mean mAP to a certain extent. From the recognition results, it is found that the[17-18].

REFERENCES

[1] CIRSHICK R,DONAHUE J.DARRELL T et al. Richfeature hierachies for accurate object detection andsemantic segmentation [C] //IEEE. Computer Visionand Pattern Recognition. Columbus: IEEE,2014: 580587.

[2] LAZEBNIK S,SCHMID C,PONCE J. Beyond bags ofSpatial pyramid matching for recognizingfeatures'natural scene categories [C] //IEEE. Computer SocietyConferenceComputerVision andPatternonRecognition. New York: IEEE,2006: 2169-2178.

[3] CIRSHICK R. Fast R-CNN [C] //IEEE. InternationalSantiago: IEEE Conference on Computer Vision.2015: 1440-1448.

[4] Chang Che, Bo Liu, Shulin Li, Jiaxin Huang, and Hao Hu. Deep learning for precise robot position prediction in logistics. Journal of Theory and Practice of Engineering Science, 3(10):36–41, 2023.DOI: 10.1021/acs.jctc.3c00031.

[5] Hao Hu, Shulin Li, Jiaxin Huang, Bo Liu, and Change Che. Casting product image data for quality inspection with xception and data augmentation. Journal of Theory and Practice of Engineering Science, 3(10):42–46, 2023. https://doi.org/10.53469/jtpes.2023.03(10).06.

[6] REN S O,HE K M,CIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with regionproposal networks L] . IEEE Transactions on PatternAnalysis and Machine Intelligence , 2017, 39 (6) :1137-1149.

[7] WOO S,PARK J,LEE J Y, et a. CBAM:Convolutional block attention module [C] //Springer.European Conference on Computer Vision. Munish:Springer,2018: 3-19.

[8] Chang Che, Qunwei Lin, Xinyu Zhao, Jiaxin Huang, and Liqiang Yu. 2023. Enhancing Multimodal Understanding with CLIP-Based Image-to-Text Transformation. In Proceedings of the 2023 6th International Conference on Big Data Technologies (ICBDT '23). Association for Computing Machinery, New York, NY, USA, 414–418. https://doi.org/10.1145/3627377.3627442.

[9] Lin, Q., Che, C., Hu, H., Zhao, X., & Li, S. (2023). A Comprehensive Study on Early Alzheimer's Disease Detection through Advanced Machine Learning Techniques on MRI Data. Academic Journal of Science and Technology, 8(1), 281–285.DOI: 10.1111/jgs.18617.

[10] Che, C., Hu, H., Zhao, X., Li, S., & Lin, Q. (2023). Advancing Cancer Document Classification with R andom Forest. Academic Journal of Science and Technology, 8(1), 278–280. https://doi.org/10.54097/ajst.v8i1.14333.

[11] Tianbo, Song, Hu Weijun, Cai Jiangfeng, Liu Weijia, Yuan Quan, and He Kun. "Bio-inspired Swarm Intelligence: a Flocking Project With Group Object Recognition." In 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 834-837. IEEE, 2023.DOI: 10.1109/mce.2022.3206678.

[12] VerySIMONYAN K,ZISSERMANA .deepconvolutional networks for large scale image recognitionC] //IEEE. Computer Vision and Pattern Recognition.Columbus: IEEE,2014: 1-15

[13] LITJENS C,KOOI T,BEJNORDI B E,et al. A surveyon deep learning in medical image analysis l] . Medicallmage Analysis, 2017, 42: 60 88.

[14] Yao, J., Zou, Y., Du, S., Wu, H., & Yuan, B. Progress in the Application of Artificial Intelligence in Ultrasound Diagnosis of Breast Cancer. DOI:https://api.semanticscholar.org/Corpus.

[15] XIE H,FANG S,ZHA Z J, et al. Convolutiona.ACMattention networks for scene text recognitionTransactions on Multimedia ComputingCommunicationsand Applications , 2019 , 15(1s) : 1-17.

[16] Li L, Xu C, Wu W, et al. Zero-resource knowledge-grounded dialogue generation[J]. Advances in Neural Information Processing Systems, 2020, 33: 8475-8485. DOI: https://doi.org/10.48550/arXiv.2008.12918.

[17] SOMMEN FV D,ZINGER S,SCHOON E J, et al.Supportive automaticesophagealannotation of earlycancer using localgaborand color[J]featuresNeurocomputing ,2014,144(20) : 92-106.

[18] SCHOON E J,SOMMEN F V D,ZINCER S, et al.Computer aided delineation of early neoplasia in barrettsesophagus using high definition endoscopic images llGastrointestinal E.

[19] Zhang, Q., Cai, G., Cai, M., Qian, J., & Song, T. (2023). Deep Learning Model Aids Breast Cancer Detection. Frontiers in Computing and Intelligent Systems, 6(1), 99-102.

[20] Xu, J., Pan, L., Zeng, Q., Sun, W., & Wan, W. (2023). Based on TPUGRAPHS Predicting Model Runtimes Using Graph Neural Networks. Frontiers in Computing and Intelligent Systems, 6(1), 66-69.

[21] Liu, B., Yu, L., Che, C., Lin, Q., Hu, H., & Zhao, X. (2023). Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms. arXiv preprint arXiv:2312.12872.

[22] Miao, T., Zepeng, S., Xingnan, W., Kuo, W., & Yuxiang, L. (2023). The Application of Artificial Intelligence in Medical Diagnostics: A New Frontier. Academic Journal of Science and Technology, 8(2), 57-61.
[23] Hong, Z., Yan, L., Jize, X., Yixu, W., Yuxiang, L. (2023). Improvement of Deep Learning Model for Gastrointestinal Tract Segmentation Surgery. Frontiers in Computing and Intelligent Systems, 6(1), 103-106.