Team member name: **Yijie Hao， Xinkai Shen**

School: **Rabun Gap-Nacoochee School / Phillips Academy Andover**

City: **Nacoochee Drive, Rabun Gap, Georgia / Main street, Andover, Massachusetts**

Country: **USA**

Instructor: **Chenxi Huang**

Thesis: **Automatic Detection and Classification of Lesions in the Alimentary Tract in Endoscopic Images Based on Deep Learning**

Automatic Detection and Classification of Lesions in the Alimentary Tract in Endoscopic Images Based on Deep Learning

Yijie Hao[1] Xinkai Shen[2]

1: Rabun Gap-Nacoochee School, 339 Nacoochee Drive, Rabun Gap, Georgia 30568, USA

2: Phillips Academy Andover, 180 Main street, Andover, MA 01810, USA

Abstract:

The endoscope technology today plays crucial roles in the coming up of extensive medical diagnoses of a patient's digestive condition, as it provides for the doctors a precise detection of their patients' endosomatic situation. The real time diagnosis of these endoscopes images, however, stands still as a subject yet to be developed. This following essay will attempt to realize the possibility of computer assistance in the diagnosis of the patient, in computing and labeling the possible syndromes that correlates to the irregularities as shown by the endoscope images, and narrowing these possibilities down by the input of the doctor's prior knowledge about the patient. This essay intends to utilize the deep learning models to realize the automatic detection and distinguishing system for multiple lesions in the alimentary tract, in advance to the detection and distinguishing system for only singular. Yolo is a deep neural network system, and the utilizing of it, as a polyp detecting system, avoids the necessity of pre-processing and positioning of the images, and thus eliminates a certain level of difficulty in the experiment. We have, in this experiment, utilized the neural network system Yolo to realize an end-to-end polyp and esophagitis target-detection system, used the mAP(mean Average Precision), a common scoring system in areas of target-detection, to quantify and review the performance of the prior mentioned polyp detection, and deployed IoU(Intersection over Union) to quantify the accuracy of the final target presented by the algorithm. We have, in finality, received a high stability, fine robustness and mAP, IoU status. We believe that this mechanism could potentially play vital roles in the clinical analysis of alimentary tract disease diagnosis, in more precisely diagnosing the information given by the endoscopic images, making more at ease the doctor's analysis of the patient's condition in accordance to the images, thus providing the doctors additional reference information for the clinical purposes, and finally lowering the rate of misdiagnosis.

Keywords： polyps， computed assisted diagnosis， endoscopic， Yolo

Table of contents

1.INTRODUCTION

The alimentary tract is a major organ by which the human body absorbs nutrients and energy.[1] If any lesions were to take place in the alimentary tract, the physical wellness of the person would be immediately subjected to serious threats. Studies have shown that pressure, contaminated environment and unhealthy eating habits could all lead to gastronomical diseases. In recent years, as the general populace has become ever more strained by the everyday pressure they face at work and at home, the rates and incidences of gastronomical maladies, particularly in terms of esophageal cancer, gastric cancer, and intestinal cancer manifested a general rising trend, creating concerns on a scale scarcely seen before. We are now facing a global increase of 2.8 million cases of gastronomical diseases yearly, resulting in 1.8 million cases of death. Gastrointestinal mucous membrane disease and cancer of digestive tract, indisputably the two most common component of digestive tract diseases, have become most accountable for the increasing number of infirmities.[2] Luckily for our stressed society, however, it has been shown by studies that early detection and diagnosis of these diseases can greatly reduce the possibility of gastrointestinal neoplasia and mortality. The traditional methods of detection for irregularities in the alimentary tract includes the digestive tract test, real-time ultrasound imaging and so on, which are but auxiliary diagnosis methods that cannot determine precisely the pathological irregularities in the tract. [3]

The endoscope (See Image 1-2) is a traditional malady-detecting tool used for the detection of digestive tract illnesses. The gastroscope, colonoscope and esophagoscope can detect the stomach, large intestine, parts of the small intestine and the esophagus for maladies. The medical endoscope has a leading advantage over other means of pathology detection because of its ability to enter the patient’s body and observe directly the performance of his organs.

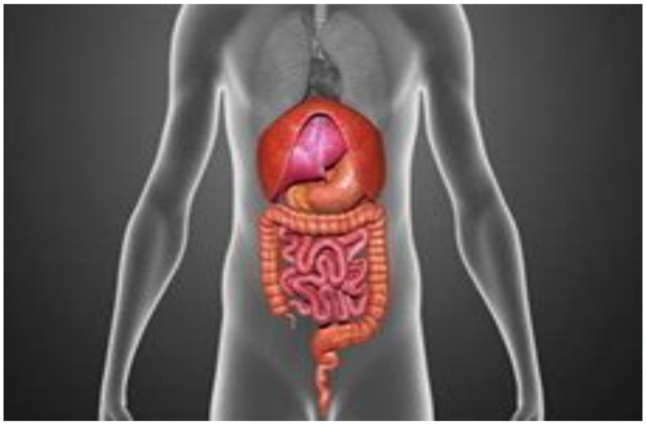


Figure 1-1 The digestive tract

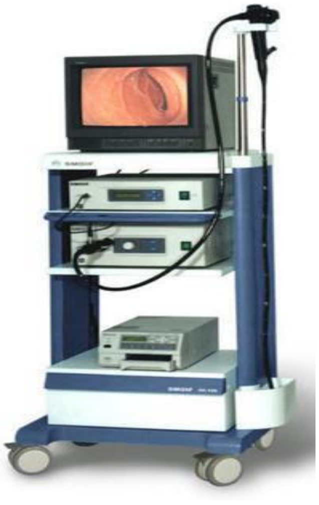


Figure 1-2 Endoscope

The endoscope test [4] has now become one of the standard tests for intestinal tract malady detection. Despite its advantages, however, it makes necessary the involvement of trained professionals in the process of classing, testing and labeling the irregularities from a large amount of images generated by the endoscope, making the diagnosis rather time and energy consuming and yet often erroneous. Leaving these detections manually processed has become increasingly insufficient over these years of continuously rising cases of gastronomical diseases, urging the application of computer aided automation, an important branch of bio-technology developed in the past decade, to be developed in this area.

Currently, the clinical judgement of whether lesions have occurred in the digestive tract is based primarily on two indexes- one being the signs of polyps and the other being the signs of flat lesions such as ulcers and hemorrhage. The polyps, being rather characteristic in an imagological sense, are rather thoroughly studied and can be diagnosed in ways based on Virtual colonoscopy and imagological characterizing based on texture properties. Wang P and his group [5] have, in their prior experiment, used the LBP mode and log-likelihood ratio to extract the variegation texture characteristic of colonoscopy images, and utilized three different modes of artificial deep neural network to categorize regular and irregular colon tissues. Zhao [6] and his fellow have extracted the color characteristic, LBP texture characteristic and shape characteristic of capsule endoscopy images and, with the help of LDA categorizer, invented a categorizing/ searching system of relevant frame pictures based on these characteristics. D.E.Maroulisa [7] and his fellow have proposed a detection system about particularly the gray images using the Wavelet Transformation to extract textural and geometrical characteristics and a neural system to detect colonic polyp; it is notable that this method proved a 95% success rate. Stavros A. [8] and his fellows has utilized the wavelet decomposition to detect tumors in endoscopic images, extract the chromatic wavelet covariance characteristics, and utilized the Linear Discriminant Analysis to distinguish the relevant lesions; which was notably rather precise in detection of colonic pancreatic polyps. There is, however, still not an effective imagological method for the detecting of flat lesions, leaving the detection mainly done based on the doctor's thorough analysis of the color and texture of the tissues and mucous membranes as observed by his raw eye. The fellows of Coimbra M T [9] did attempt to use the visual MPEG-7 descriptor to distinguish flat-lesions such as hemorrhage and ulcers, which did achieve a certain level of effectiveness, but had a particularly high omission ratio.

We have utilized the computer aided diagnosis to extract the characteristics of the relative lesion areas by dividing the lesions into three categories: polyps, dyed and lifted polyps, and esophagitis, hoping, in finality, to combine this with the doctor's priori knowledge to aid the gastroenterology doctors to discover suspicious areas in the digestive tract. We find this method meaningful in lowering the omission ratio, spotting lesions at an early stage and finally curing them, as this method has fully utilized the appropriate digital image processing technology to analyze information involved in medical images, performed qualitative and quantitative analysis of relevant tissues to aid the doctor in his diagnosis of the patient. The Deep learning system[10], in comparison to the traditional machine learning technique[11], eliminates the manual extraction of characteristics, and rather identifies multiple layers of characteristics from the original input data; the system can automatically learn to extract layered characteristic information, which is beneficial for categorization and detection. With the current development in the AI field and digital graphics technology, traditional machine learning can no longer satisfy the need created by big data to process an enormous amount of information, and the Deep Learning technique, on the other hand, makes possible the precise target detection of endoscope images. The research goal of this experiment, in conclusion, is to propose a deep learning based computer assisted digestive endoscope image diagnosis system that is used to categorize digestive endoscope images and ultimately automatically detect the lesions such as polyps[12], dyed and lifted polyps, and ulcers [13]presented in these images.

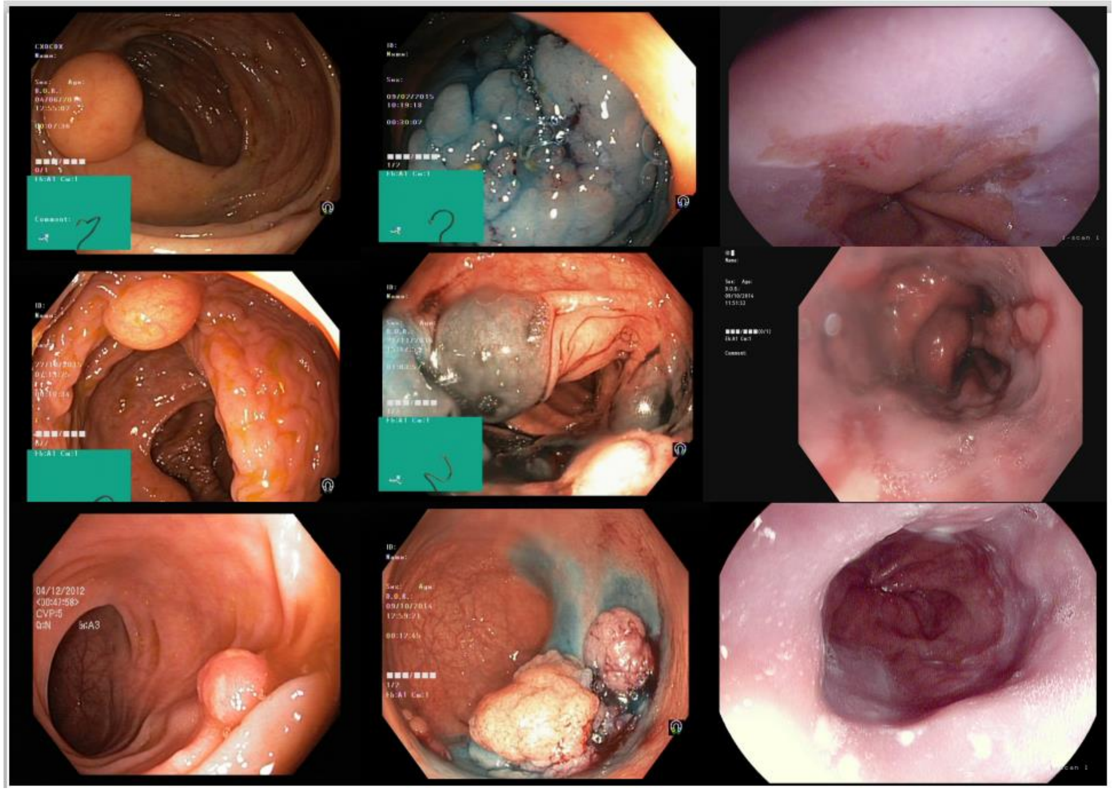


Figure1-3 Polyps, Dyed and lifted polyps , Esophagitis

2. METHODS

Yolo is an end to end deep learning framework purposed for target detection. It was designed to accomplish the mission of target detection without pre-processing the images in forms of extracting relative characteristics within images, locating desired items, dividing images into segments and so forth. As a deep learning neural network, it is designed to handle the input of colored images, output target or the target class and its coordinate at the inference stage. (NEED WORK) The Yolo network is made up of 24 convolution layers and 2 fully connected layers. The convolution layers are used to extract characteristics, while the fully connected layers are used for locating and classing. It uses as an inference the framework of GoogleNet, but did not make use of GoogleNet's inception modules, and rather substituted the inception modules with the 11th and 33th convolution layers, minimizing the level of complication involved in the entire model.

Speciically, the Yolo's CNN network will divide the input images into s*s unit cells, with each cell only running prognosis over one item, and each cell will take in charge and detect the target whose center befalls in it. Each unit cell will run prognosis over a total number of B bounding boxes and check the confidence score about the B bounding boxes. The "condfidence score" as mentioned embodies two aspects, one being the possibility of the bounding box containing the wanted target, and the second being the accuracy of the bounding boxes. We will also run prognosis over a total number of C conditional possibilities (possibility of each target class). For example, for the VOC data group, Yolo applies a 7*7(S=7) unit, with each unit detecting 2(B=2) bounding boxes and 20 (C=20) classes. Each bounding box contains 5 elements (x,y,w,h,conf), each being the coordinates and confidence score of the boundary boxes. We will unify the coordinate positions into a ratio between 0-1. The x and y element as mentioned earlier are the offset amount of the current boundary boxes. Each boundary box will have 20 conditional possibilities, this conditional possibilities standing for the possibility of the detected polyp falling in one of the 20 classes (each class in each unit will generate one odds), leading the Yolo's final prognosis matrix being:

(S,S,B*5+C)=(7,7,2*5+20)=(7,7,30). The image 2-1 below will present the Yolo detection sequence.

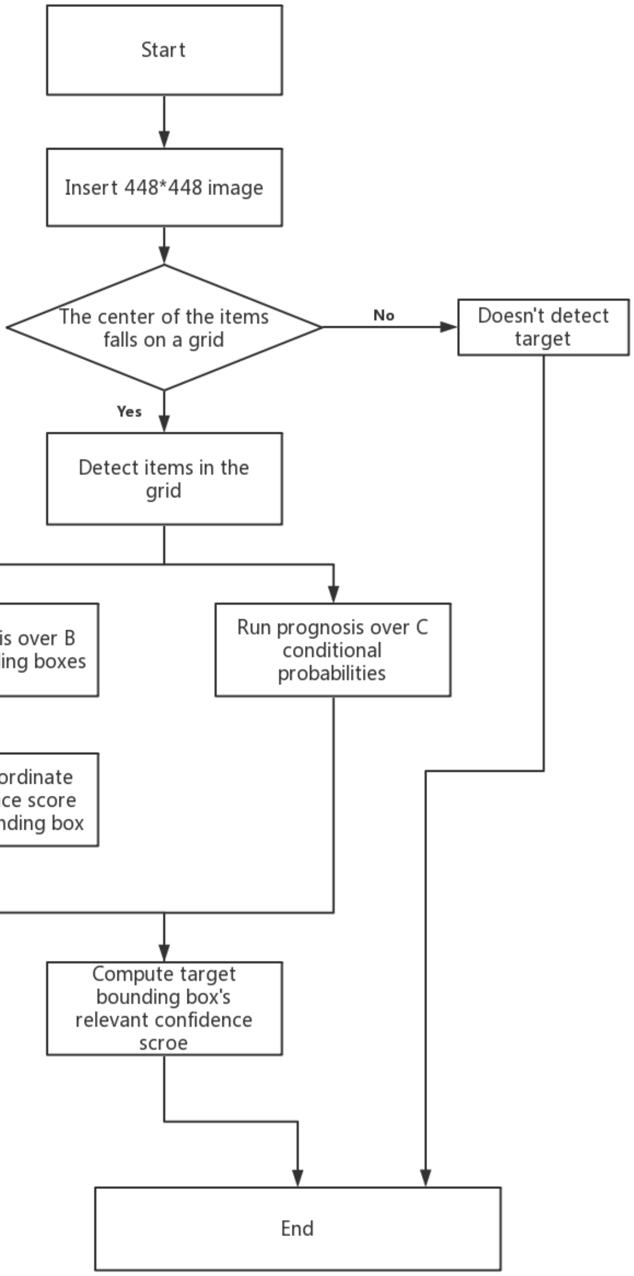

Start
Insert 448*448 image
The center of the items falls on a grid
No
Doesn't detect target
Yes
Detect items in the grid
Run prognosis over B target brounding boxes
Run prognosis over C conditional probabilities
Compute coordinate and confidence score of target bounding box
Compute target bounding box's relevant confidence scroe
End

Figure 2-1 Yolo detection process

Each bounding box's classing confidence score has the computing formula as following: Confidence Score= Box confidence score* Conditional probabilities of classes. The network simultaneously measures and runs prognosis over the confidence score of the boundary boxes.

Each of Yolo's units can run prognosis over multiple boundary boxes (notably, VOC can only run prognosis over 2 boundary boxes) To account for the loss of the positive samples, we only demand one of the boundary boxes to account for the object. For his, we choose the one having the highest IOU status with the ground truth. This strategy will cause the particularity of the bounding boxes' predictions, or, in other words, it will cause each prediction being more accurate in size and ratio.

Yolo takes the quadratic variance between the pronosised boundary boxes and real boundary boxes as the function accounting loss. Before training the Deep Learning framework, We run, for Yolo, a pre-training on ImageNet. The classing model of this pretraining adopted only the first 20 convolution layers, while adding an average-pool layer and fully connected layer. After the pre-training, 4 randomly initiated convolution layers and 2 fully connected layers were added. Because the polyp detection usually requires images with better resolution, we expanded the network's input from 224*224 to 448*448. The Yolo network adopted leaky ReLU, dropout, and data augmentation to augment its learning rate. The first epoch went from 0.001 to 0.01, with 0.01 lasting 75 epochs, 0.001 lasting 30 epochs and 0.0001 lasting 30 epochs.

We have, through the pretraining, discovered a certain flaws to the YoLov1 network. Because the output are fully connected layers, the resolutions of all images had to be unanimous. It was also rather spacially bounded, with each box running prognosis over only 1 target. While each box had B bounding boxes, we only chose the bounding box with the highest IOU to run the prognosis, so while there were B*5 coordinates, there were only C class probabilities, leaving the output being (B*5+C) instead of B*(5+C). This made the analysis for small items, in terms of prognosis and adaptation of length-wide ratio, not particularly ideal.

The Yolo algorithm has seen the polyp detection as a matter of regression, to which it treated with the average quadratic variance function. It has, however, assigned different weights to different parts. First, it differed locating errors from classing errors. For locating errors, or in other words, the errors of bounding boxes' prognosis over coordinates, it assigns a greater weight of 5. It then assigns confidence scores to bounding boxes that does not contain the target item and the ones that does, assigning a weight of 0.5 to the former and 1 to the latter. It then adopts the average quadratic variance function that treats equally all grounding boxes however different in size, while the smaller grounding boxes are in reality more sensitive to locating errors than the larger grounding boxes. For ensuring of treating of the smaller boxes' sensitivity, we have changed the program of prognosis about the bounding boxes' width and height to a prognosis about the root of the width and height. In the system, each unit tries for a forecasting of multiple grounding boxes, while it only corresponds to one class. This made it such that in the pre-training, while there is a target item in the unit, only the grounding box that has the largest IOU with the ground truth was assigned to detect this target; since the other grounding boxes can not identify the item. This setting made it such that the grounding boxes that corresponds to a unit are more professionalized, and can be used for targets of different size and different length-height ratios, which improves the model's efficiency. One might wonder what happens when multiple targets are in one unit, the truth is that the Yolo algorithm will only choose one of the targets for training purposes, which happens to be a flaw of the algorithm. It is worth mentioning also that for grounding boxes whose correlating target does not exist, the loss function only accounts for the confidence score, while locating errors are considered incalculable in the system. In this algorithm, classinging error are only taken into account when a target is actually obtained by a unit, and is otherwise also incalculable.

It is necessary to differentiate the classing training and polyp detecting training in the training of Yolo. Rather alike the human's cognitive process, it can only detect the target polyps once it learns to categorize. The training adopted the first 20 convolution layers, while adding an average-pool layer and fully connected layer, and then was pretrained on ImageNet. 4 convolution layers and 2 fully connected layers were additionally added to the algorithm before we used the PASCALVOC data group to obtain all the parameters of Yolo networks. In fact the first 20 convolution layers were used for classing while the latter 4 convolution layer, as long with the 2 fully connected layers were used for polyp detection for the classed items. With these 2 trainings, the first solving the classing issues and the second fine-tuning and solving the polyp detection issues, we have rather easily solved the loss function's unwillingness to come down.

At the training stage, when the loss function stops descending in value or other issues start to manifest themselves, we can examine the deviated terms one by one to check which term seems to be problematic. For example, in terms when the Yolo function is running the polyp detection training and the loss function does not converge, we check the convergence of each deviated terms and try to solve the issue term by terms.

For solving Yolo's inaccurate locating and low recall ratio, Yolov2 has accomplished some improvements in addition to the system of Yolov1.

A. Batch Normalization

We unified the inputs at every level of the network in Yolov2. In the training of the Deep Neural Network, in terms when the distribution of the input training data changes, the network will have to learn the different distributions in the iteration processes, which lowers the training rate. We added a BN layer to each convolution layer, which should improve the network's rate of convergence. The experiment proves the mAP showing a 2 percent raise after adding the BN levels.

B. Higher resolutions

In the pretraining, Yolo has used only 224*224 images, and input the images at a 448*448 resolution when it was time for polyp detection. This made it such that the system has to adapt the change of resolution in the process between classing and polyp detecting. In the Yolov2, we split the pretraining into two steps, first inputting the 224*224 resolution images, and running fine-tuning for 10 epoch under 448*448 resolution to make smoother the adjustments when the system is at the polyp detecting stage. The experiment proves the mAP showing a 4 percent raise after applying this technique.

C. Lead into Anchor Boxes

Yolo predicts bounding boxes coordinates directly through the full connection layer at the top. Yolov2 referenced anchor boxes from Faster r-cnn (see figure 2-2). In order to prevent the algorithm from aimlessly guessing the size of the target frame, Faster r-cnn generates anchor boxes with a certain ratio of length and width for the reduction in the amount of computation. Yolov2 does not directly predict the bounding boxes' coordinates, but their offset with respect to anchor boxes. By predicting relative offsets instead of directly predicting coordinates simplifies the learning problem and makes the network easier to learn.

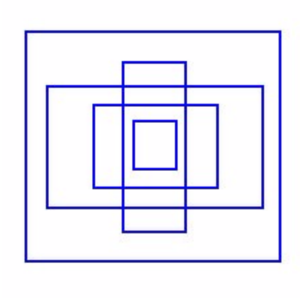


Figure 2-2 Anchor Boxes

D. Cluster produces Anchor Boxes

In the Faster r-cnn, anchor boxes are artificially selected according to experience. Yolov2 wants to select anchor boxes of appropriate size to help the network prediction's accuracy. K-means is utilized to cluster appropriate anchor boxes in the bounding boxes of the training set. In addition, standard k-means with Euclidean distance will cause greater errors in larger bounding boxes. The purpose of adding anchor boxes is to obtain better IOU, therefore IOU is used for the distance calculation:

D (box, centroid) = 1 - IOU (box, centroid)

Anchor boxes selected by clustering are greatly different from those selected artificially, which makes the learning ability of the model stronger.

E. Multi-scale training

The full connection layer in Yolo network makes the input can only be fixed size images. But in Yolov2, the full connection layer is removed, left only convolution layer and pooling layer, therefore make the multi-scale training possible. Since in Yolov2 network the multiplier of down-sampling is 32, therefore the size of the input image is adjusted according to the multiple of 32. The minimum size is 320×320 while the maximum size is 608×608, makes the network can adapt the input of different scales.

Figure 2-3 shows the network structure of Yolov2. Instead of deepening or widening the network to improve the model, Yolov2 simplifies the network. Yolov2 proposed a new classification model darknet-19 (as shown in table 2-1), including 19 convolution layers and 5 maximum pooling layers, which is less than the 24-layer convolution layer of Yolo. It takes 8.52 billion calculations for Yolo to complete a forward process, while only 5.58 billion calculations for Yolov2 to process an image, thus reducing part of the computation.

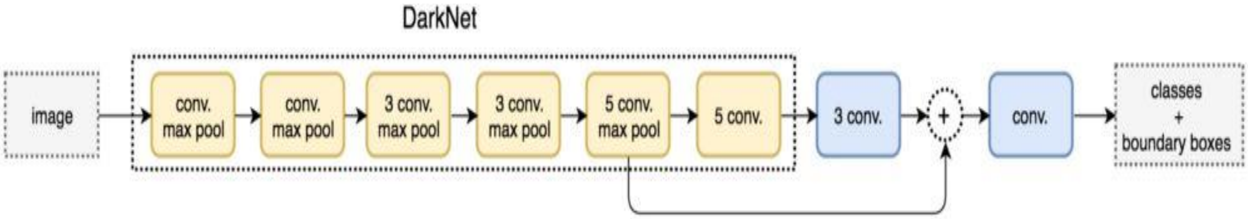
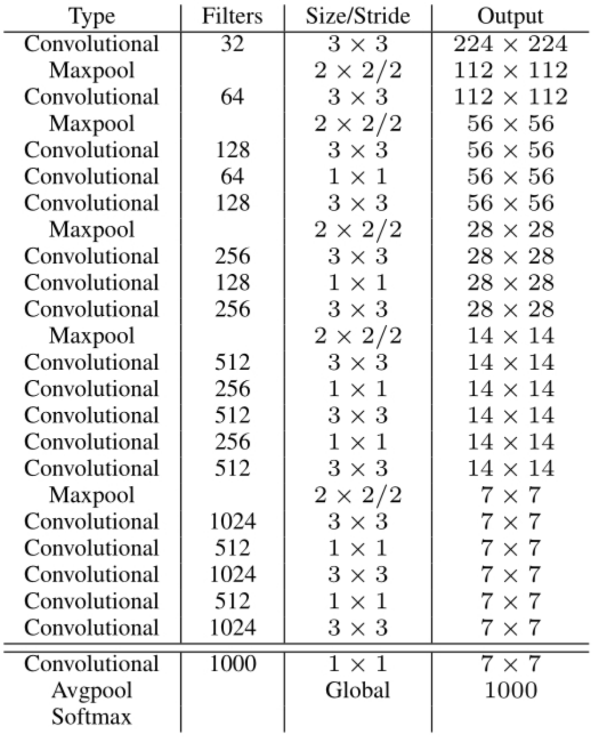


Figure2- 3 The framework of Yolo 2



Type	Filters	Size/Stride	Output
Convolutional	32	3×3	224×224
Maxpool		$2 \times 2/2$	112×112
Convolutional	64	3×3	112×112
Maxpool		$2 \times 2/2$	56×56
Convolutional	128	3×3	56×56
Convolutional	64	1×1	56×56
Convolutional	128	3×3	56×56
Maxpool		$2 \times 2/2$	28×28
Convolutional	256	3×3	28×28
Convolutional	128	1×1	28×28
Convolutional	256	3×3	28×28
Maxpool		$2 \times 2/2$	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Maxpool		$2 \times 2/2$	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	1000	1×1	7×7
Avgpool		Global	1000
Softmax			

Table 2-1 Darknet-19

Figure 2-4 shows the network structure of Yolov3. Compared to Yolov2, the major improvements in v3 are the use of residual models and the adoption of the FPN architecture. The feature extractor of Yolov3 is a residual model. As it contains 53 convolution layers, it is called darknet-53. From the perspective of network structure, compared with the use of residual units for darknet-19 network, it can be built deeper.

Another point is the use of FPN architecture (Feature Pyramid Networks for Object Detection) to achieve multi-scale polyp detection. Yolov3 USES feature maps of three scales (when input is 416*416) : 13*13, 26*26, 52*52. The structure of Yolov3 network on VOC data set is shown in figure 15, where the red part is polyp detection results of each scale's feature map. Yolov3 USES 3 prior boxes for each position, so k-means is used to obtain 9 prior boxes and divide them into three scale feature graphs. Similar to SSD, feature graphs with larger scale use smaller prior boxes.

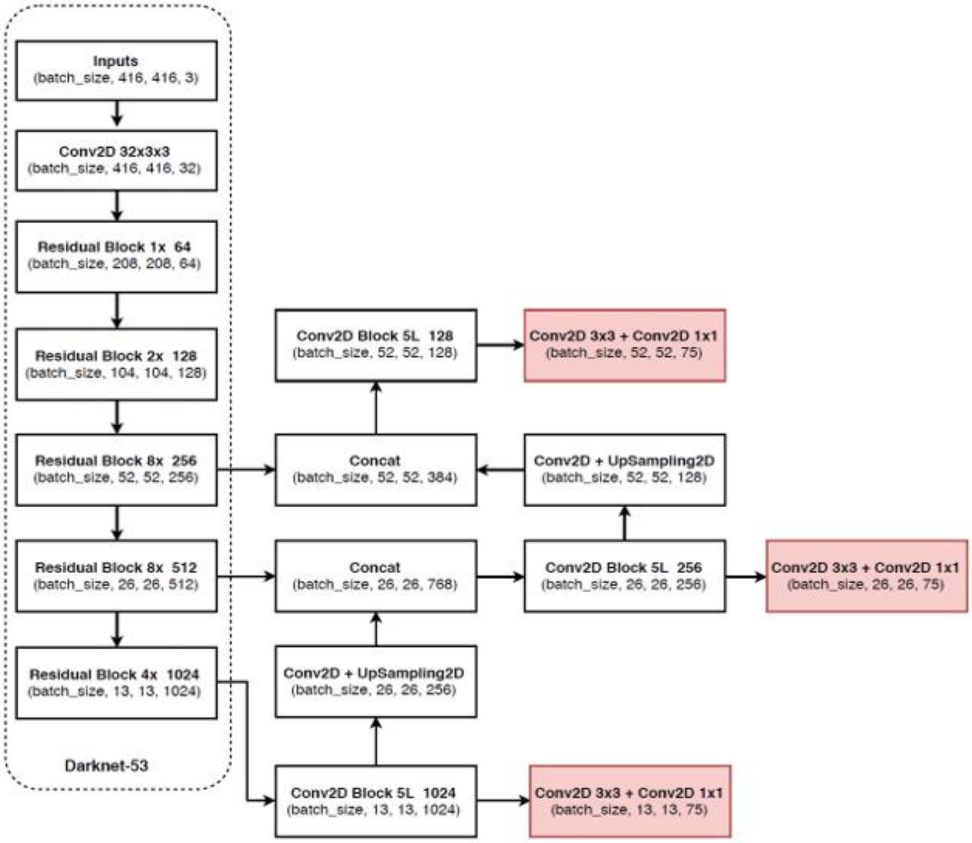


Figure2-4 The framework of Yolo 3

3. RESULTS AND ANALYSIS

A. data set

This paper is mainly about the studies on polyp target detection, so we adopted Kvasir data set. This is a data set that contains images of the gastrointestinal tract. The data set, containing two types of images associated with endoscopic polypectomy, are sequenced and annotated by experienced endoscopists. Therefore, Kvasir plays an important role in the study of computer aided detection of diseases.

We prepared a total of 3000 sample pictures, including 1000 pictures of polyps, stained polyps and esophagitis. For these 3000 pictures, we manually annotated the pretreatment, thus to get a better training effect in the follow-up experiments.

B. Evaluation criteria

MAP (mean Average Precision) was used to quantitatively evaluate the detection performance of polyp targets. MAP is a common measure in the field of target detection. Recall is not used as the evaluation standard, is mainly due to its failure in measuring the position of each object when it comes in evaluating the classification and positioning of the model in the target detection problem that differentiate on detection targets is shared with each image. IoU is the intersection of detection results and actual target borders divided by the union between them, which shows the coincidence degree between the target boxes predicted by the algorithm and the boxes marked in the original picture. The higher the IoU is, the more accurate the target positions detected by the algorithm are. The mathematical formula is formula (3-1).

$$\text{IoU} = \frac{area\ \ (predict)\ \cap area\ \ (truth)}{area\ \ (predict)\ \cup area\ \ (truth)} \qquad (3\text{-}1)$$

An artificial threshold is required for IoU to determine whether the target detection is correctly detected. This threshold is usually set as 0.5. If IoU > and 0.5, we can believe the detection is correct, and vice versa. According to the comparison between the calculated IoU and the set IoU threshold, the correct detection times of each class in each image can be denoted as T times. TP is the real class. Formula (2-2) is used to calculate the accuracy of the model for category C.

$$\text{Precision}_C = \frac{N(TP)_C}{N(TotalObject)_C} \qquad (3\text{-}2)$$

For a given category C, there are N (TotalImages) images in the verification set, and the average precision of this class is calculated, that is, the average of the predicted precision value of category C in all images. The calculation formula is (3-3).

$$AveragePrecision_C = \frac{\sum Precision_C}{N(TotalImages)_C} \qquad (3\text{-}3)$$

There are N (Classes) in the whole training set. IoU, precision and average precision are calculated according to (3-1), (3-2) and (3-3) for each category. The entire model is then measured by the calculated average accuracy of all categories. Here, we take the average of all class average precision values (i.e., mAP) to measure the performance of a model, and the formula is expressed as (3-4).

$$\text{meanAveragePresicion} = \frac{\sum AveragePrecision_C}{N(Classes)} \qquad (3\text{-}4)$$

C. result analysis

In order to evaluate the performance of the method described in this paper in polyp target data set, we first trained by a class of common polyp. Using the best network Yolo3 in the simplest case, we have achieved good results, with the mAP of single polyp reaching 0.754. The training results are shown in figure (3.1).

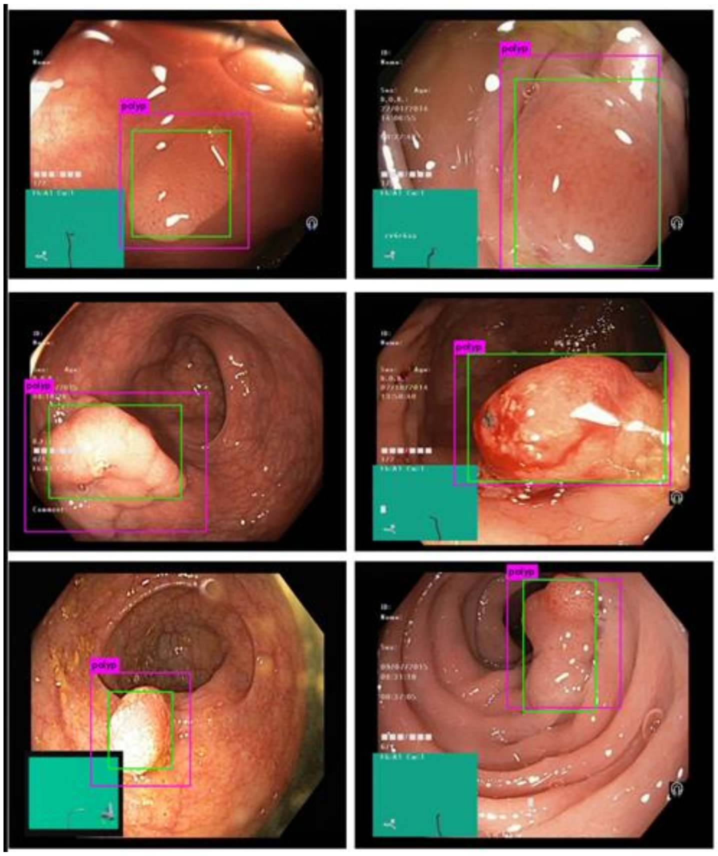


Figure 3-1 Single polyp detection, the green box is the labeling box and the purple box is the prediction box

This inspired us to train on the classification of polyps, stained polyps and esophagitis. Figure (b) is a predictive graph for multiple types of detection.

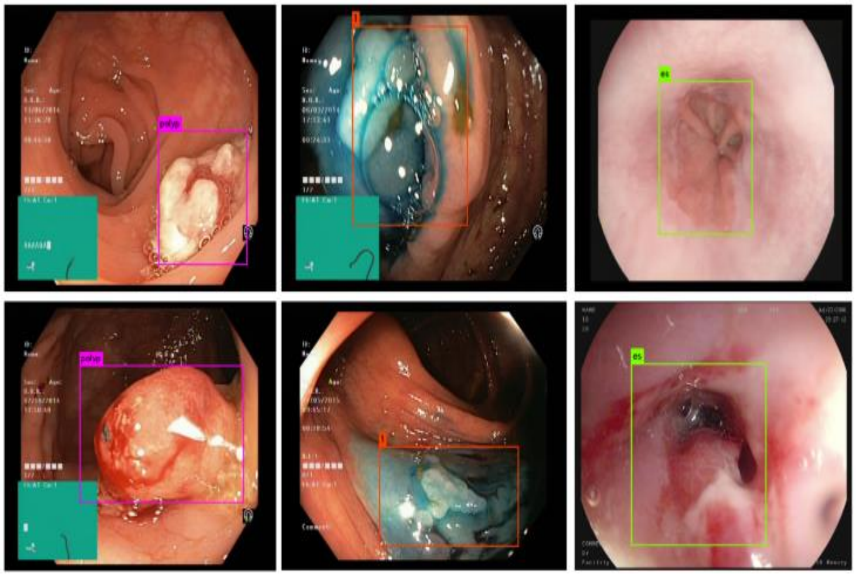


Figure 3-2 The predictive map, with each column from left to right representing polyps, Dyed and lifted polyps, and esophagitis

On the basis of Yolo1, Yolo2 implemented batch standardization to improve speed and mAP, adopted a higher-resolution network, introduced Anchor Boxes to simplify learning, and clustered Anchor Boxes to enhance model learning ability. Multi-scale training enables the network to adapt the input of different scales. Yolo2 also simplifies the Internet. On the basis of Yolo2, Yolo3 improved the residual model to build deeper, and adopted FPN architecture to realize multi-scale detection. Therefore, the structure of Yolo3 is much more complicated than Yolo2.

In the considering of the above situations, we starts to think that the complex structure such as Yolo3 may not be required in the training medical data set. If a simple structure such as Yolo2 can achieve a good training effect, then the workload will be greatly reduced. Considering such a situation, when training polyps, stained polyps and esophagitis, we trained with Yolo2 and Yolo3 pairs respectively and visualized loss curves, iou curves and recall curves. Figure (3-1) is the average loss attenuation curve of Yolo2's loss function. Figure (3-2) is the average loss attenuation curve of Yolo3's loss function. Figure (3-3) is the growth curve of the average precision rate of Yolo2. Figure (3-4) is the growth curve of average precision rate of Yolo3 in Region 82 with IoU=0.5 as the threshold. Figure (3-5) is the average precision growth curve of Yolo3 in Region 82 with IoU=0.75 as the threshold. Figure (3-6) is the growth curve of average precision rate of Yolo3 in Region 94 with IoU=0.5 as the threshold. Figure (3-7) is the average precision growth curve of Yolo3 in Region 94 with IoU=0.75 as the threshold. Figure (3-8) is the average IoU growth curve of Yolo2.

Layer, so you can do multi-scale training. In Yolov2 network, the multiplier of down-sampling is 32, so the size of the input image is adjusted according to the multiple of 32. The minimum size is 320×320, and the maximum size is 608×608, so that the network can adapt to input of different scales.

By comparing figures (3-3) and (3-4), it can be seen that the average loss of Yolo2 and Yolo3 decreases gradually with the increase of training set data. However, by comparing the y-coordinate, we can clearly see that the average loss of Yolo3 decreases faster than that of Yolo2. With the same amount of training data, the average loss is much smaller, therefore the minimum average loss can be much smaller. By comparing chart (3-5), (3-6), (3-7), (3-8), (3-9), it is easy to find out that with the increasion of the training set data, lesion area average precision increased gradually to 1.0. Yolo2 will need to reach a large batch of training set data to reach recall = 1, but Yolo3 in smaller batches on the training set of data can achieve the same effect. By comparing figure (3-6) with figure (3-7), figure (3-8) and figure (3-9), Yolo3 can achieve a better average check in Region 82 with the same IoU. By comparing figures (3-6) and (3-7), (3-8) and (3-9), Yolo3 can achieve better average checking with IoU=0.5 in the case of Region 82 or Region 94. As can be seen from figure (3-10), with the increase of training set data on Yolo2, the average intersection ratio of focal area gradually increases and finally tends to be stable.

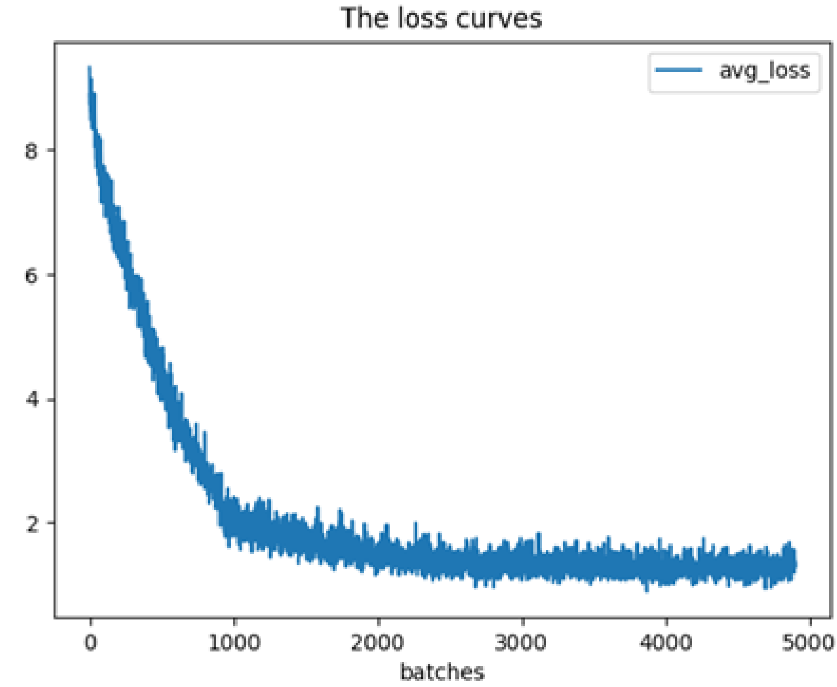


Figure3-3 The average loss attenuation curve of Yolo2's loss function

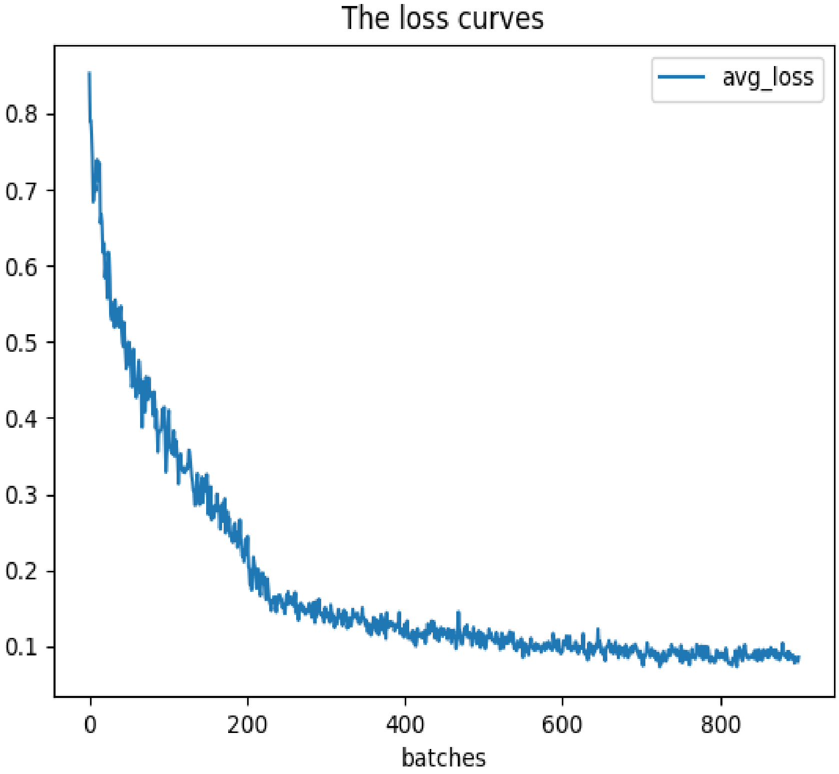


Figure3-4 The average loss attenuation curve of Yolo3's loss function

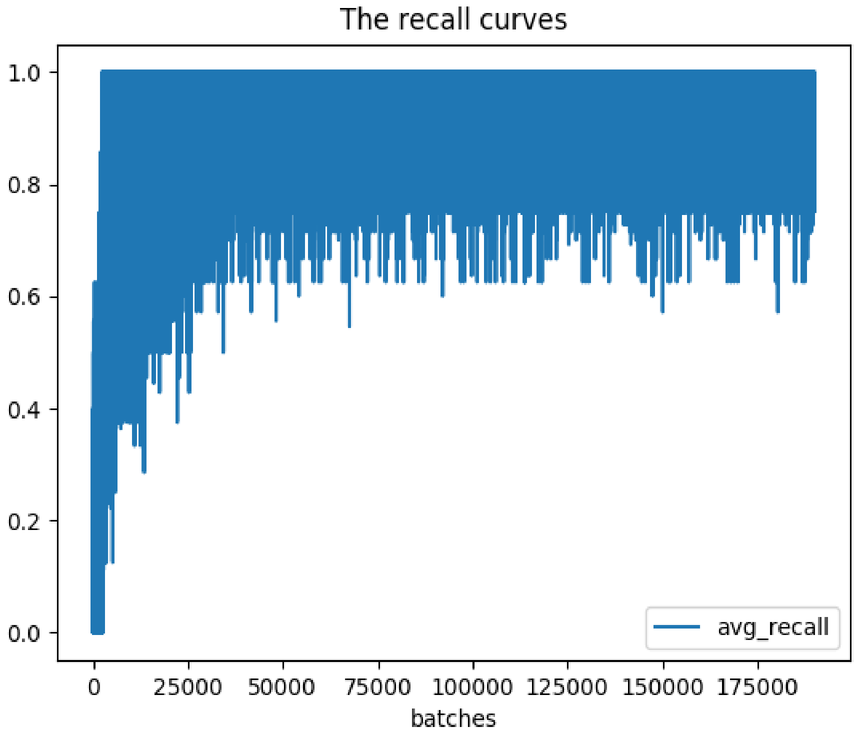


Figure3-5 The average accuracy growth curve of Yolo2

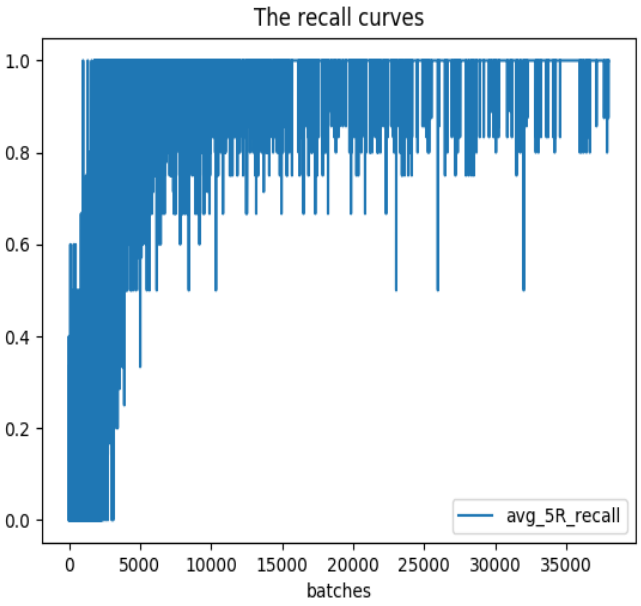


Figure3-6 The average alignment rate growth curve of Yolo3 in Region 82 with IoU=0.5 as the threshold

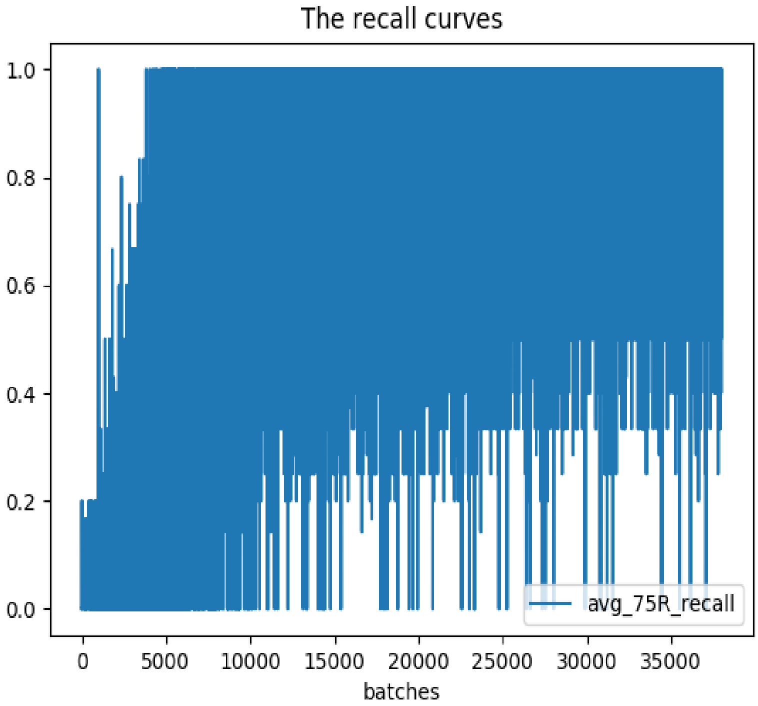


Figure3-7 The average alignment rate growth curve of Yolo3 in Region 82 with IoU=0.75 as the threshold

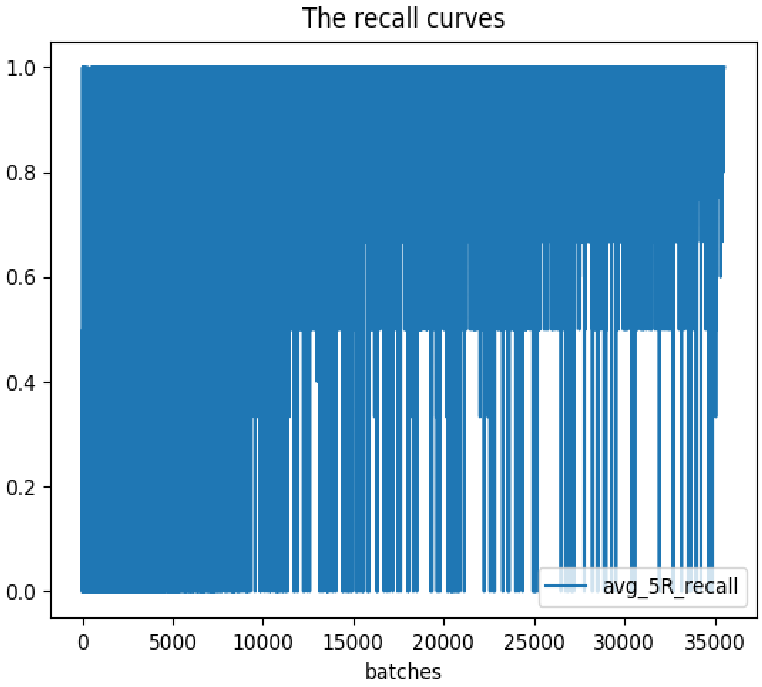


Figure3-8 The average alignment rate growth curve of Yolo3 in Region 94 with IoU=0.5 as the threshold

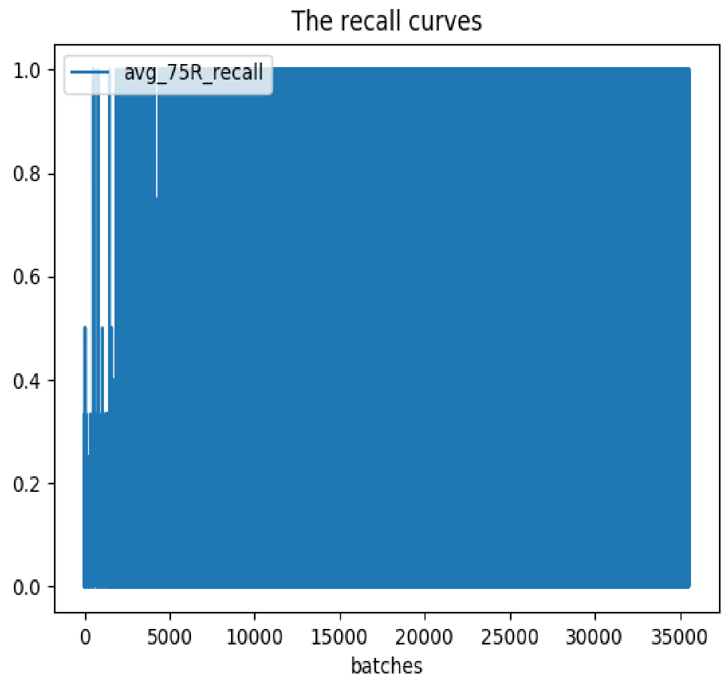


Figure3-9 The average alignment rate growth curve of Yolo3 in Region 94 with IoU=0.75 as the threshold

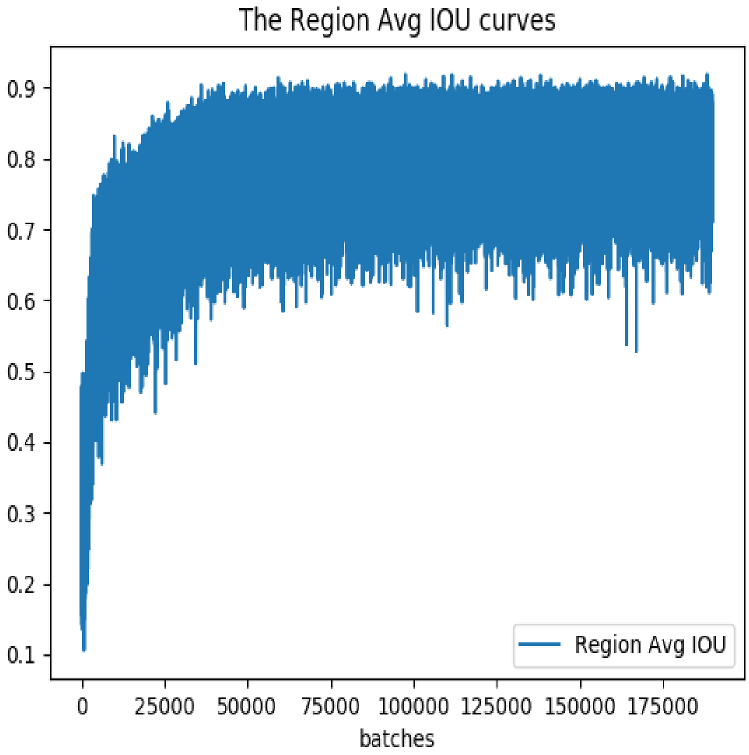


Figure 3-10 The average IoU growth curve for Yolo2

According to the above experimental results, although Yolo3 has a more complex structure, the robustness of the test results is better than that of Yolo2, and Yolo3 only needs to use a smaller batch of training data when achieving the same training results. Therefore, Yolo3 will still be adopted when training medical data sets.

By adopting Yolo3 on three classification problem, polyp ap reached 0.712, dyeing polyps ap reached 0.614, esophagitis ap reached 0.834, mAP = (0.712 + 0.614 + 0.834) / 3 = 0.720. Compared to the mAP = 0.754 reached on the single classification by Yolo3, the mAP on three classification problem is very close to the single classification's mAP. Therefore, our algorithm model of polyps, dyeing polyps, the classification of esophagitis image positioning is proved to be effective with a strong robustness.

4. CONCLUSION

Computer assisting diagnosis has a high application value in the field of gastrointestinal diseases prevention and treatment. In this article, through the utilization of YOLO target detection based on deep learning framework in the discovery on medical endoscope single classification polyp detection and polyp, dyeing polyps and esophagitis three classification and exploration on the automatic recognition and detection, we found that this algorithm has a quicker speed, higher precision and better robustness compared to the traditional manual detection way. Therefore, a new theory in digestive tract disease prevention and control is presented. According to the image processed by this algorithm, doctors can analyze and diagnose the pathological picture of medical endoscope better, thus the misdiagnosis rate of doctors cane reduced, and the accuracy and efficiency of clinical diagnosis can be improved, which has great practical significance and good practical application prospects.

5.REFERENCE

[1] Ji Xiaorong，Gastrointestinal pathology [M]. Beijing：People's military medical publishing house. 2010: 21-40

[2] Timerbulatov V M, Urazbakhtin I M, Shv T, et al. [Pathogenesis, treatment, and prevention of erosive-ulcerative lesions in the mucous membrane of the upper digestive tract].[J]. 2011(1):29-35.

[3] Farraye F A, Odze R D, Eaden J, et al. AGA Technical Review on the Diagnosis and Management of Colorectal Neoplasia in Inflammatory Bowel Disease[J]. Gastroenterology, 2010, 138(2):746-774.e4.

[4] Glassman C I. Endoscopic appendectomy.[J]. Gastrointestinal Endoscopy, 2008, 15(02):59-64.

[5] Wang P, Krishnan S M, Kugean C, et al. Classification of Endoscopic Image Based on Texture and Neural Network[C]// International Conference of the IEEE Engineering in Medicine & Biology Society. 2001.

[6] Qian Z, Max Q.-H. Meng, Li B. WCE video clips segmentation based on abnormality[C]// IEEE International Conference on Robotics & Biomimetics. 2011.

[7] Maroulis D E, Iakovidis D K, Karkanis S A, et al. CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames.[J]. Computer Methods & Programs in Biomedicine, 2003, 70(2):151-166.

[8] Karkanis S A, Iakovidis D K, Maroulis D E, et al. Computer-aided tumor detection in endoscopic video using color wavelet features[J]. IEEE Transactions on Information Technology in Biomedicine A Publication of the IEEE Engineering in Medicine & Biology Society, 2003, 7(3):141.

[9] Coimbra M T, Cunha J P S. MPEG-7 Visual Descriptors—Contributions for Automated Feature Extraction in Capsule Endoscopy[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2006, 16(5):628-637.

[10] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis.[J]. Medical Image Analysis, 2017, 42(9):60-88.

[11] Bruijne M D. Machine learning approaches in medical image analysis: From detection to diagnosis[J]. Medical Image Analysis, 2016, 33:S1361841516301098.

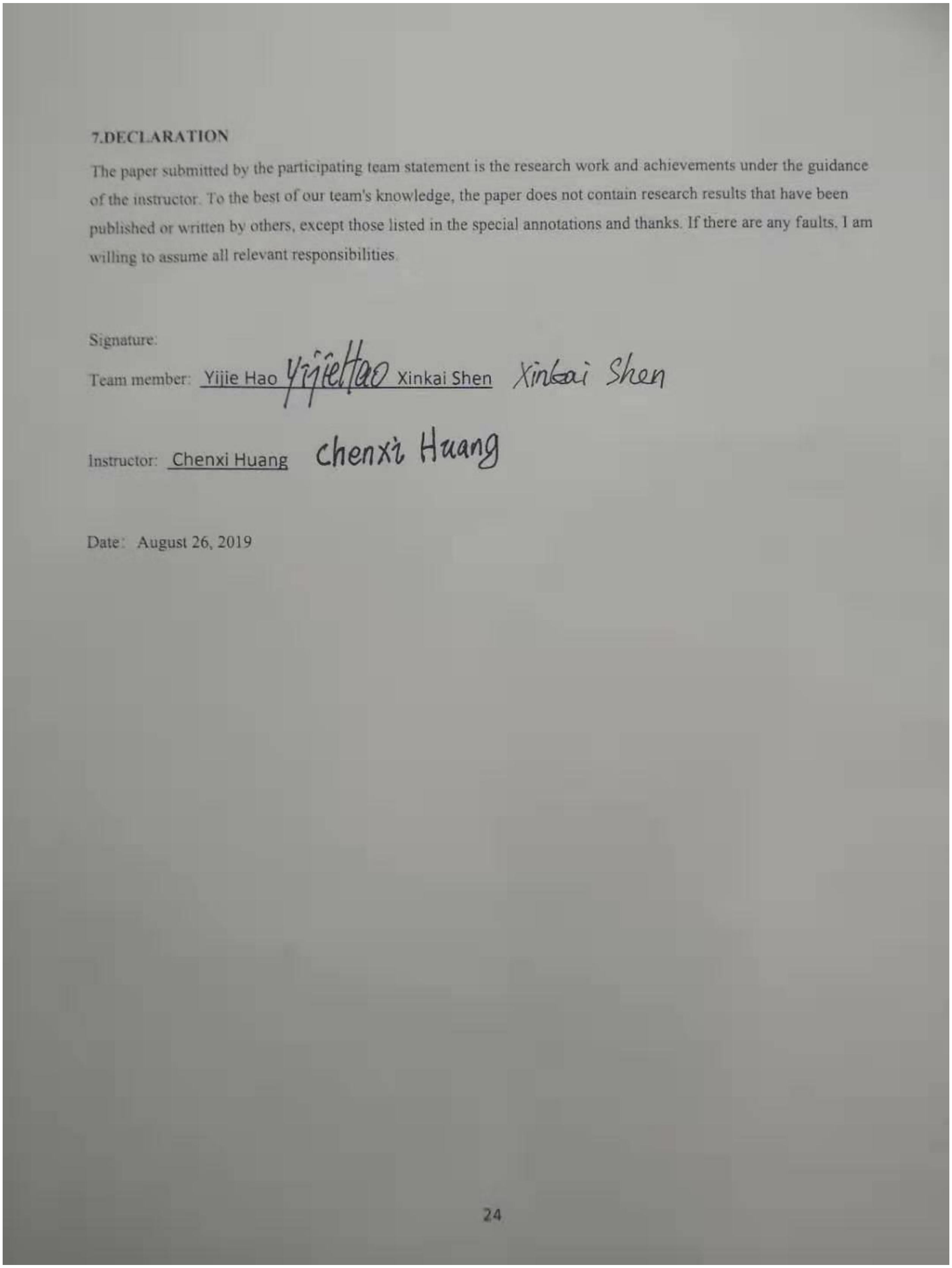
[12] Carmack S W, Genta R M, Graham D Y, et al. Management of gastric polyps: a pathology-based guide for gastroenterologists.[J]. Nature Reviews Gastroenterology & Hepatology, 2009, 6(6):331.

[13] Cheng D C, Ting W C, Chen Y F, et al. AUTOMATIC DETECTION OF COLORECTAL POLYPS IN STATIC IMAGES[J]. Biomedical Engineering Applications Basis & Communications, 2011, 23(05):1100276-.

6.ACKNOWLEDGMENT

In the process of writing this article, I have been greatly helped and supported by Mr. Huang Chenxi. Huang Chenxi has done a lot of guidance and help in the conception of the paper, the establishment of the mathematical model, the explanation of mathematical knowledge, the preparation of experimental materials and images, which has greatly benefited us. Yijie Hao and Xinkai Shen have done a lot of work in feature extraction and programming of neural network image after systematically understanding and learning biomedical knowledge. Together, they completed the study and research of Yolo-based machine vision programming. In this paper, 4000 pictures of gastrointestinal endoscopy were prepared, a large number of pictures were marked and their training works were programmed. This part of the work was completed by Yijie Hao and Xinkai Shen. In the second half of the study, Yijie Hao and Xinkai Shen worked together to complete the whole paper.

After hard work and long study, the research work has obtained this research result. We have a deep understanding of the arduous and long time of scientific research. Fortunately, we have a strong interest in knowledge and a good instructor. Thank you all.



7.DECLARATION

The paper submitted by the participating team statement is the research work and achievements under the guidance of the instructor. To the best of our team's knowledge, the paper does not contain research results that have been published or written by others, except those listed in the special annotations and thanks. If there are any faults, I am willing to assume all relevant responsibilities.

Signature:

Team member: Yijie Hao YijieHao Xinkai Shen Xinkai Shen

Instructor: Chenxi Huang Chenxi Huang

Date: August 26, 2019