Vessel Detecting using Restrict Single Shot Multibox Detector for Intravascular Ultrasounds

Zujie Liu [1], Zuheng Liu[2], Yunfeng Peng [1], and Yanni Guo[1]
[1] School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing,China
[2] State Key Laboratory of Organ Failure Research, Nanfang Hospital, Southern Medical University,Guangzhou,China

Keywords: Intravascular ultrasounds(IVUS), Vessel detecting, Restrict Single Shot MultiBox Detector

Abstract: Intravascular ultrasounds (IVUS) is a technique in scanning coronary artery, which is extensively used in interventional therapy and it can provide valuable clues in detecting coronary plaques. Nevertheless, up to now, most of the image frames of IVUS are manually examined by physicians. In this paper we designed a restrict single shot multibox detector(R-SSD) method to automatically locate the regions of interests, e.g. vessel, for computer-aided IVUS examination, by changing the initial feature extraction network and restricting the range of prior box of original SSD method dedicated for object recognition. The accuracy on locating vessel can achieve 95.4% using the proposed R-SSD.

1 INTRODUCTION

In recent years, the incidence of coronary artery disease is increasing due to various unhealthy lifestyle and aging population throughout world.

Coronary artery disease is an outcome of atherosclerotic, because of vascular stenosis or obstruction, resulting in myocardial ischemia or myocardial infarction (MI). The rupture of atherosclerotic plaques will probably lead to MI, which is a disease with high mortality in clinical practice. Most MI patients need expensive interventional treatment immediately and are probably required to perform IVUS to improve the accuracy and security of intervention operation. Rapid diagnosis and treatment will greatly improve the prognosis and survival rate of MI patients. However, dramatically increased emergency operation and workload will probably lead to inevitable fatigue even for skilled physicians, which will increase the risk of surgery.

To alleviate the repeated medical workloads for physicians on the assessment of coronary angiography, Computer-aided image object detection is now cast a new light on machine aided IVUS image analysis on coronary artery angiography.

Traditional object detection method is usually a brute force algorithm to search the objects using windows with different size sliding from right to left, and from up to down in a image frame, which is low-efficiency.

Some machine learning algorithms such as support vector machine(SVM) and random forest(RF), have been used for binary classification of high risk from low risk vessel(Tadashi et al, 2016) (Sheet et al, 2014). The features inputted to these machine learning algorithm are extracted from IVUS images using statistic methods. By combing with deep learning mechanisms, convolution neural network(CNN) can automatically extract features from images and classify these images (Krizhevsky et al, 2012).

R-CNN(Ross et al, 2014) and Fast-R-CNN (Girshick, 2015) are proposed base on selective search(Uijlings et al, 2013) which combine neighboring pixels as a group by calculating the similarity of each region. The selective searching are based on outside region proposal method, and its processing capacity is still slow. After that, a region proposal network（RPN） is proposed to replace selective searching to be Faster-R-CNN(Ren et al, 2017). However, They, i.e.,R-CNN, Fast-R-CNN and Faster-R-CNN, are two-stage object detection and will spend more time in region proposal. The single shot detection(SSD) method(Liu et al, 2016) is a one-stage object detection and is expected to efficiently solve the region proposal problem.

Liu, Z., Liu, Z., Peng, Y. and Guo, Y.
Vessel Detecting using Restrict Single Shot Multibox Detector for Intravascular Ultrasounds.
DOI: 10.5220/0008096500180023
In *Proceedings of the International Conference on Advances in Computer Technology, Information Science and Communications (CTISC 2019)*, pages 18-23
ISBN: 978-989-758-357-5

In this paper, based on CNN, we design a restrict single shot multibox detector, called R-SSD, which is improved from the existing SSD in(Liu et al,2016) by changing the base network and restrict the range of prior box. Our restrict single shot multibox detector (R-SSD) are appropriate for vessel detecting with high accuracy on locating vessel at 95.4% for IVUS analysis.

2 METHODS AND MATERIAL

2.1 Images Acquisition

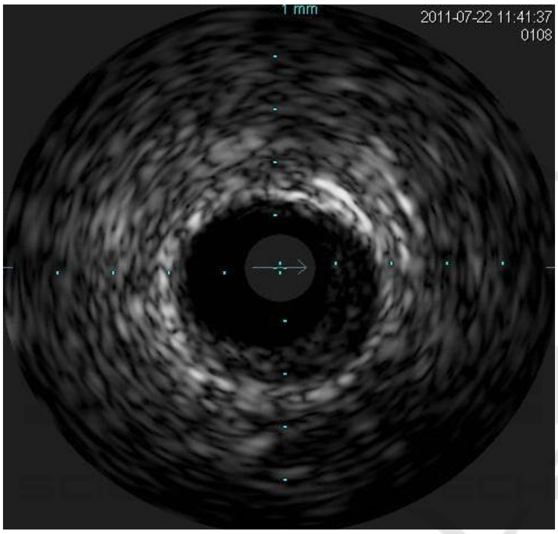

Figure 1: Example of a IVUS image.

Coronary angiogram examination is done by radial artery approach, while a 6-French guiding catheter is used to selectively cannulate the ostium of the target coronary artery. A guiding shot is taken after administering a weight-adjusted dose of unfractionated heparin and nitroglycerin (200ug). Immediately after guidewire advancement, while before balloon predilation, a 20-MHz, 2.9-Fench IVUS-catheter is inserted into the target coronary arteries lesion. Then the IVUS-catheter is automatically pulled back to the coronary ostium at 0.5mm/s using an automated pull-back device. During Pull-back, all IVUS images are recorded and stored.

These process are performed by physicians from Nanfang hospital, Southern medical university. An example IVUS image is illustrated in Figure 1.

2.2 Object Detection Method

In the original SSD method(Liu et al,2016), as shown in Figure 2, images firstly go through a classification network (Howard et al,2017) where its last two full-connection layer are changed into four CNN layers to extract features; and then the Feature Pyramid structures like conv4-3, conv-7(FC7), conv6-2, conv7-2,conv8_2 and conv9_2 are used to generate prior box on different feature maps, so to make classification as well as location regression by using double 3*3 convolution kernel to output 5 value. One is for confidence which generate 2 classification, e.g., vessel and background, and the other one is for localization where each default box generate 4 value. Because SSD method do not have the process of region proposal and use feature pyramid detection method, its speed is high and its accuracy is almost reaching that of Faster-RCNN.

However, the area of vessel in IVUS almost greater than 25% and medical equipment may be portable. So the restrict SSD method (R-SSD) is designed by replacing the base network of original SSD to MobileNet for feature extraction which is lightweight and could be embedded into portable equipment. On the other hand, the range of prior box for original SSD is 0.2-0.9. Because they have small box to detect and consider the area of vessel, we

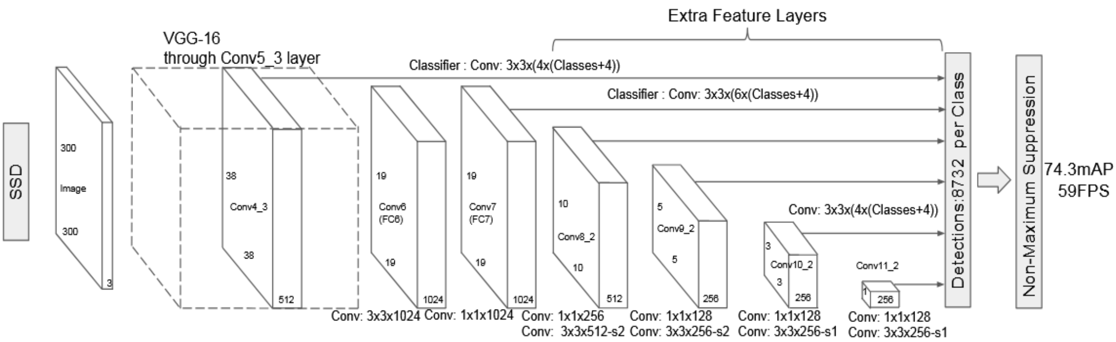

Figure 2: The network of SSD.

restrict the range of 0.4-0.95 for vessel detection in R-SSD.

3 EXPERIMENT

3.1 Date Sampling and Pretreatment

We use OpenCV to sampling the video at 10 frames per second and 4200 images (pic1) are obtained. After sampling, the images are turned into gray scale firstly and then divided into train set (3300 images) and test set (900 images). Furthermore, as shown in Figure 3 and 4, we use Gaussian Filter and Histogram Equalization to pre-process the image and then use *LabelImg* to label the vessel. The vessel ROIs (region of interest) are marked and a set of 4-tuples parameters(*x, y, w, h*) are achieved to describe positions of the vessel, where *x* and *y* denotes the coordinate for the central of the box, *w* and *h* denotes the width and height of the box.

Each labelled image is translated into a *xml* file. After all images are labeled, the *xml* file is saved into a *csv* form. Because Tensorflow is used in this paper to train model, the *csv* form and gray scale image are transferred into *tfrecord* format.

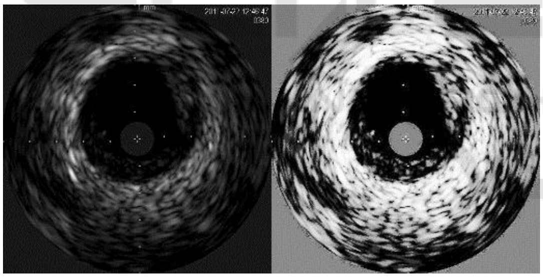

Figure 3: The different between original IVUS image and after preprocess.

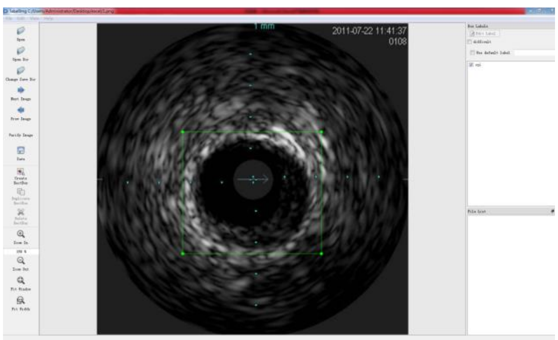

Figure 4: The usage of LabelImg.

3.2 Training

In this paper, the SSD object detection models are trained using *Tensorflow*, via following four-steps: setting the training objective, matching strategy, generating prior box, and training parameters.

Step 1: setting the training objective.

We derive and extend the SSD training objective from the Multibox (Erhan et al, 2014) to handle multiple object categories. The overall objective loss function is a weighted sum of the localization loss (loc) and the confidence loss (conf):

$$L(x,c,l,g) = \frac{1}{N}(L_{conf}(x,c) + \alpha L_{loc}(x,l,g)) \qquad (1)$$

Where N is the number of matched default boxes. If N = 0, we set the loss to 0. The localization loss is a Smooth L1 loss between the predicted box (*l*) and the ground truth box (*g*) parameters. Similar to Faster-RCNN, we regress to offsets for the center (*cx, cy*) of the default bounding box(*d*) and for its width(*w*) and height(*h*).

$$L_{loc}(x,l,g) = \sum_{i \in Pos}^{N} \sum_{m \in \{cx,cy,w,h\}} x_{ij}^{k} smooth_{L1}(l_i^m - \hat{g}_j^m) \qquad (2)$$

The confidence loss is the *softmax* loss over multiple classes confidences(*c*).

$$L_{conf}(x,c) = -\sum_{i \in Pos}^{N} x_{ij}^{p} \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^o) \qquad (3)$$

Where

$$\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(c_i^p)} \qquad (4)$$

and the weight term α is set to 1 by cross validation.

Step 2: matching strategy.

During training, we need to determine which default boxes correspond to a ground truth detection and to train the network accordingly. We therefore select each ground truth box from default boxes that vary over location, aspect ratio, and scale. So that each ground truth box is matched to the default box with the best *jaccard* overlap, e.g. IOU. Unlike that for MultiBox, we here match default boxes to any ground truth with *jaccard* overlap higher than a threshold (0.5) (Liu et al, 2016). This method can simplify the learning problem, by allowing the network to predict high scores for multiple overlapping default boxes rather than requiring it to pick only the one with maximum overlap (Erhan et al, 2014).

Step 3: generating prior box.

Empirically, feature maps from different levels within a network are known to have different receptive field sizes. Fortunately, within the SSD framework, the default boxes do not necessary need to correspond to the actual receptive fields of each layer (He et al, 2015). We design the tiling of default boxes, so that specific feature maps can learn to be responsive to particular scales of the objects (Zhou et al, 2015). Suppose we want to use m feature maps for prediction. The scale of the default boxes for each feature map can be expressed by:

$$S_k = S_{\min} + \frac{S_{\max} - S_{\min}}{m-1}(k-1), k \in [1, m] \quad (5)$$

where $S_{\min}$ is 0.2 and $S_{\max}$ is 0.9 in original SSD. However, the vessel in IVUS would not be too small, so $S_{\min}$ is set to 0.4, $S_{\max}$ is set to 0.9 or 0.95(as shown in Table 1) in our method to promote the performance. We also impose different aspect ratios (a_r) for the default boxes an compute the width($w_k^a = s_k \sqrt{a_r}$) and height($h_k^a = s_k / \sqrt{a_r}$) for each default box.

Table 1: The range of box.

Smin, Smax	0.4-0.9	0.4-0.95
S1	0.4	0.4
S2	0.5	0.51
S3	0.6	0.62
S4	0.7	0.73
S5	0.8	0.84
S6	0.9	0.95

As shown in Figure 5, by combining predictions for all default boxes with different scales and aspect ratios from all locations of feature maps, we have a diverse set of predictions, covering various input object sizes and shapes.

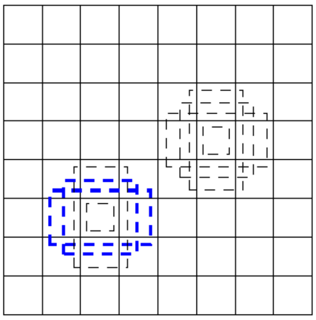

Figure 5: The prior box for each feature map cells.

Step 4: training parameters.

In this paper, we use *GTX1050ti* single GPU for training. The experiments are based on *MobileNet* as feature extract network, which is pre-trained. We fine-tune the resulting model using SGD with initial learning rate 0.004, 0.9 momentum, 0.0005 weight decay. Because the memory of the GPU is only 4G, the batch size is set to 20. Because transfer learning have better performance in deep learning, we use SSD_mobilenet_v1_coco, SSD_mobilenet_v2_coco as initialize models to train our model.

4 EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, we first compare the downtrend curve of lose function to evaluate the performance of training. Then we evaluate the accuracy of our method.

4.1 The Converge of Loss Function

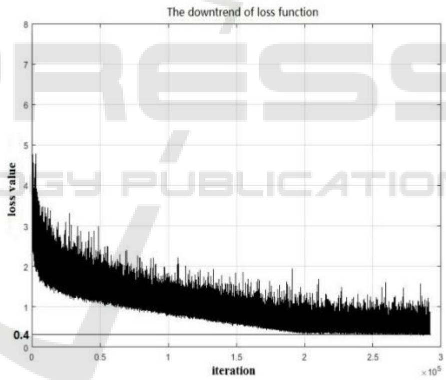

Figure 6: The downtrend of loss function.

During training, The *Tensorflow* is used to compute the loss for every iteration. Figure 6 shows that the R-SSD method can converge at 0.4 with 200000 iteration which is a good result. The final loss value (0.4) means that our SSD method is useful for vessel detection in IVUS. Matlab is used to combine the loss function variation trend of each method to measure their convergence speeds. Figure 7 shows that the converge speed of R-SSD method is more faster than that for original SSD and the best range between 0.4 to 0.95 is achieved. As shown in Figure 8, using MobileNetV2 as the initialize model, can achieve faster converge speed for loss function than that using *MobileNetV1*. Figure 9 shows the performance comparison between the proposed R-

SSD and the original SSD, which shows that our R-SSD is better than the original SSD in training speed.

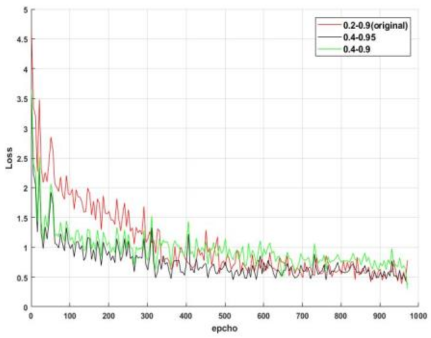

Figure 7: The loss function converge of different prior box shows that restrict range is fast than original range and the range between 0.4 to 0.95 is the best range.

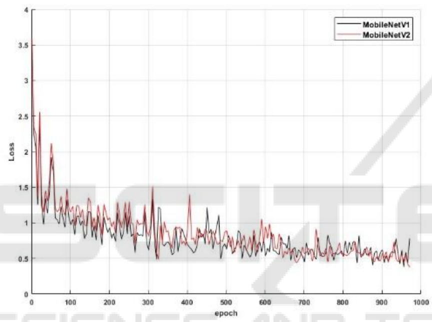

Figure 8: The loss function shows that MobileNetV2 have better converge with likely invert resisdent structure and training speed.

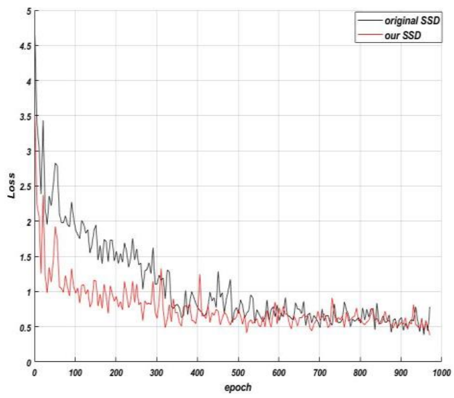

Figure 9: The loss function of our SSD and original SSD (our restict SSD is better in converge speed during training because its default box is more centralized)

4.2 Detection Accuracy

Four metrics including precision, recall, accuracy and F1-score are used to quantify the performance of object detection models. We first analysis the performance with different range of default box. As shown in Table 2, the range between 0.2 to 0.9 is the best while the range between 0.4 to 0.95 is about 0.5%lower and 0.4 to 0.9 is the worst. Although the range between 0.2 to 0.9 has the best performance in accuracy, the range between 0.4 to 0.95 is much better in loss function converge. So we considered the range of default box between 0.4-0.95 has better performance. On the other hand, Table 3 shows that *MobileNetV2*, even it spend more time in a single training, has 0.9% higher in detecting accuracy than that for *MobileNetV1*. Considering the advantage of loss function converge performance in *MobileNetV2*.

Table 2: The performance for different range.

range	P	R	A	F
0.2-0.9	96.3%	96.7%	96.5%	96.5%
0.4-0.9	93.4%	94.4%	93.9%	93.9%
0.4-0.95	**95.6%**	**96.7%**	**96.1%**	**96.1%**

Table 3: The performance for different initialize model.

initialize	P	R	A	T
MobileNetV1	94.8%	95.6%	95.2%	0.893s
MobileNetV2	**95.6%**	**96.7%**	**96.1%**	**0.935s**

Table 4: The performance of R-SSD and original SSD.

method	P	R	A	F
R-SSD	94.6%	95.7%	95.1 %	95.1%
Original SSD	**95.6%**	**96.7%**	**96.1%**	**96.1%**

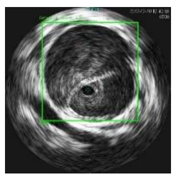

Figure 10: The green rectangle shows the position of the vessel that our object detection model detect.

We can affirm that MobileNetV2 is better than MobileNetV1. We also compare the detecting accuracy of our R-SSD with that of original SSD, shown in Table 4 where our model has 1% higher than original SSD in terms of accuracy. In Figure 10, the vessel detected with our method is labelled in green rectangle.

5 CONCLISIONS

We introduced a restrict-SSD as an object detector for vessel in IVUS, which can restrict the range for default box and change the initialize feature extraction network to *MobileNet*, then to improve training efficiency for models. We compared the training speed and accuracy between the original-SSD and our restrict-SSD, and the result shows that our restrict-SSD outperforms the original-SSD.

The vessel detection will be a good start for the future IVUS image classification. The reliable classification results can do a great help to render IVUS images automatically read by computers.

ACKNOWLEDGEMENTS

This work was supported by the foundation of Guizhou Key Laboratory of Electric Power Big Data ，Guizhou Institute of Technology (2003008002). We also thank Nanfang Hospital providing the IVUS images and helpful comments.

REFERENCES

Tadashi A.,Nobutaka I., Devarshi S., and et al.(2016).A new method for IVUS-based coronary artery disease risk stratification: A link between coronary & carotid ultrasound plaque burdens. *Computer Methods & Programs in Biomedicine.* Volume 124,pages 161-179.

Sheet D., Karamalis A., Eslami A., et al. (2014). Hunting for necrosis in the shadows of intravascular ultrasound. *Computerized Medical Imaging & Graphics*. 38(2):104-112.

Krizhevsky A., Sutskever I., Hinton G E.(2012).ImageNet Classification with Deep Convolutional Neural Networks. In *2012 International Conference on Neural Information Processing Systems(NIPS).*ACM.

Girshick R.(2015). Fast R-CNN. In *2015 IEEE International Conference on Computer Vision (ICCV).* IEEE.

Uijlings J.R.R., Sande K.E.A., Gevers T., Smeulders A.W.M.(2013).Selective Search for Object Recognition. *International Journal of Computer Vision*.104(2):154-171.

Ren S., He K., Girshick R., Sun J.(2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.*IEEE Transactions on Pattern Analysis & Machine Intelligence*.39(6):1137-1149.

Liu W., Anguelov D., Erhan D., et al. (2016).SSD: Single Shot MultiBox Detector. In *European Conference on Computer Vision*. Springer.

Howard A G, Zhu M., Chen B., et al.(2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, https://arxiv.org/abs/1704.04861 [Link].

Erhan D., Szegedy C., Toshev A., Anguelov D. (2014).Scalable Object Detection Using Deep Neural Networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. IEEE.

He K., Zhang X., Ren S., Sun J.(2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 37(9):1904-1916.

Zhou.B.,Khosla A., Lapedriza A., Oliva A., Torrala A.(2015). Object Detectors Emerge in Deep Scene CNNs, In *International Conference on Learning Representations*. https://arxiv.org/pdf/1412.6856.pdf. [Link]