Liver Lesion Detection: Image Preprocessing for Faster R-CNN

Keyword: image cropping, data augmentation, feature pyramid, Faster R-CNN

Zheng Chuanyang(17251311), Gong Yikai(17251567)

Introduction

Background

The liver is the largest organ in humans' bodies, which carries lots of functions [1]. Nowadays, its health issues are getting increasingly more attention. As a group of abnormal cells [2], the liver lesion is a signal of the liver healthy problem, and it may lead to a more serious problem such as tumors and cancer. Early detection and early treatment are the keys to patient recovery. In this article, using Faster R-CNN as a baseline, we are designing an end-to-end, efficient and accurate algorithm to detect liver lesions, which improve the performance 0.6% at least.

Preliminary Studies

Related work

Object Detection. Faster R-CNN is one of the most efficient methods for object detection and classification since 2015. It introduces a Region Proposal Network (RPN) which enables almost cost-free region proposals [3]. Faster R-CNN is designed for RGB images, which needs three channels.

Feature Pyramid. Feature Pyramid [5], or multi-scale, the method tries to use different scales information, compared traditional one-scale model. The main method of feature pyramid is using different kernel sizes. Different kernel sizes give different receptive fields and give different feature sizes. Feature Pyramid uses different kernel sizes to get different feature sizes and computer combined results by giving these features different weights. Compared to the original simple feature pyramid method[5], several research papers are discussing a new feature pyramid method [4][6][7].

Motivation

As mentioned before, the input of Faster R-CNN needs three channels. However, in this task, only grey images are given. That is to say, we have only one channel. There are three ways to solve this problem. First, input the same data into three channels by duplicate the gray images. Second, modify the Faster R-CNN code to make it compatible with one channel condition. Third, preprocess grey image by image convolution. The first and second ways make the remain channels underutilized. And the third method only uses some of the images to generate the next one, which may lead to data loss.

Preliminary experiments.

The first one is the baseline. No image pre-cropping, data augmentation or feature pyramid. And same information was inputted into three channels. Then faster r-CNN was adopted to train. The accuracy is 58.0%. It takes four hours to train, and it takes one minute to test.

In the second experiment, we just add simple convolution methods to test whether our direction is correct. **Our simple model (not the final model)** is the following.

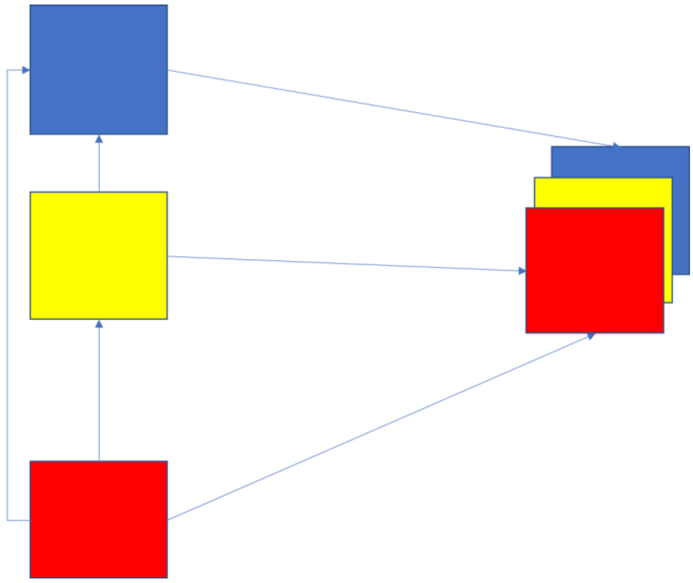

Figure 1. Input is a red image. To get yellow image (1 channel), convolute red image (1 channel) with kernel size=3, stride=1, padding=1. To get a blue image (1 channel), convolute red image (1 channel), convolute red image and yellow image with kernel size=3, stride=1,padding=1. Then, stack red image, yellow image, and blue image

Method	IOU Accuracy
Baseline(duplicate 2 other images)	58%
Baseline(original image+lapalican method processed image+max pool image)	54.3%
Simple Model	58.4%

Table 1. Comparison with baseline and simple model

This result shows that our method may improve accuracy, so further study is to research for an advanced way to generate the three channels.

Project Objective

To find the best way to detect liver lesions. Meanwhile to explore the performance of the feature pyramid. And to observe how other preprocessing methods influence the result.

Algorithm and Rationale

Image pre-cropping

The goal is to cut a rectangle of the abdominal section. This algorithm repeats four times from four directions. First, determine a dark threshold (e.g. 10) and stopping threshold (e.g. 95%). For example, scan every column from the left-hand side to the right-hand side, if in one column, the ratio of dark pixels (value<10) is lower than 95%, stop scanning and record the column. Repeat from four directions so that we can get a bounding box. The choice of dark threshold and stopping threshold need to optimize according to experiments. **After the pre-cropping image, the image will be resized to 3200*3200.** With the simple method, the detection field is expanded, which may help to improve detection accuracy.

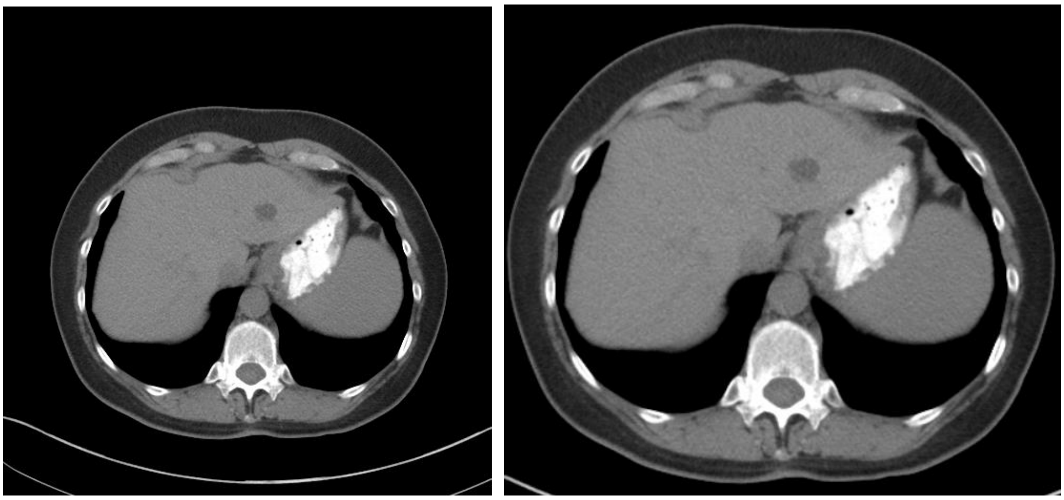

Figure 2. Left image: Original image. Right image: image after pre-cropping.

Data augmentation

Data augmentation is conducted in four ways: 90-degree spin, axisymmetric method, centrosymmetric method and erase. Two problems need further research. The first one is 90-degree spin will change the original shape of the lesion, what influence will the change make? The second one is how to solve the problem that new lesion edge may differ from the background after 90-degree spin, axisymmetric method and centrosymmetric method.

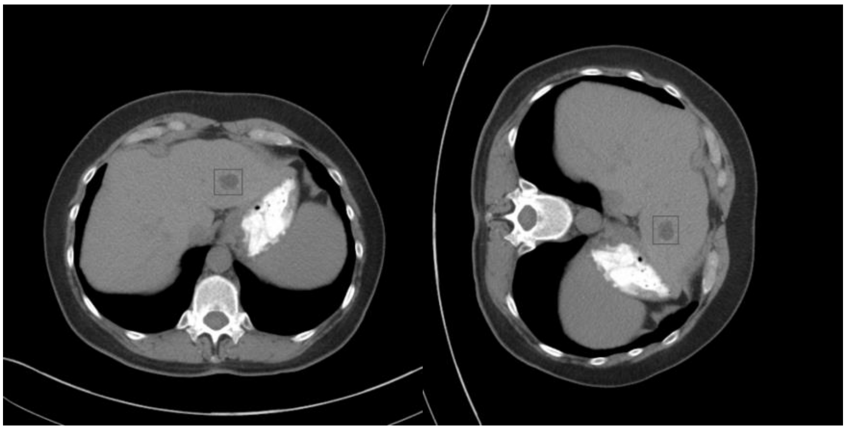

Figure 3. Left image: original image. Right image: rotated 90-degree image

Feature Pyramid.

We mainly use the feature pyramid method to get a three-channel image from a grey image. Faster R-CNN takes the output of the feature pyramid. Today's feature pyramid only considers from top to down or from down to top. Just like the following.

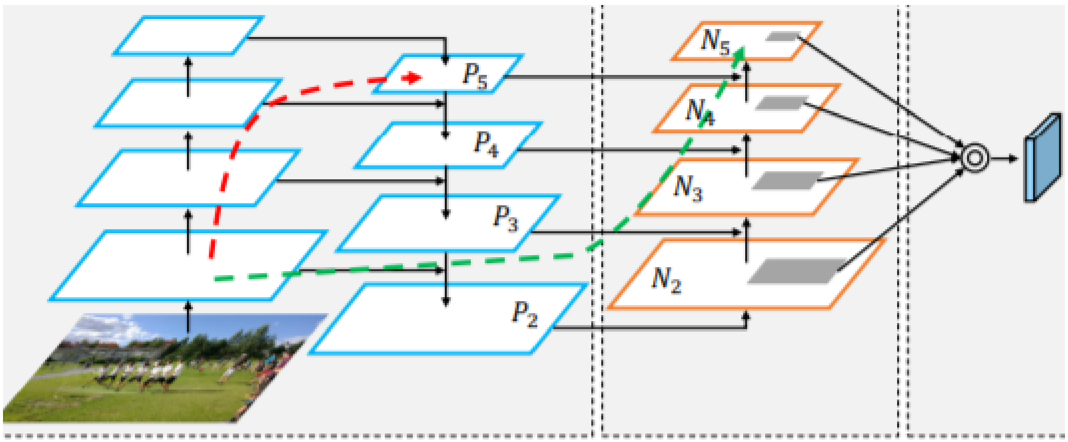

Figure 4. Feature pyramid.[6]

We all know that after convolution, some information may be lost. Therefore, according to ResidualNet[8] and DenseNet[9] idea, we use the following method for the feature pyramid.

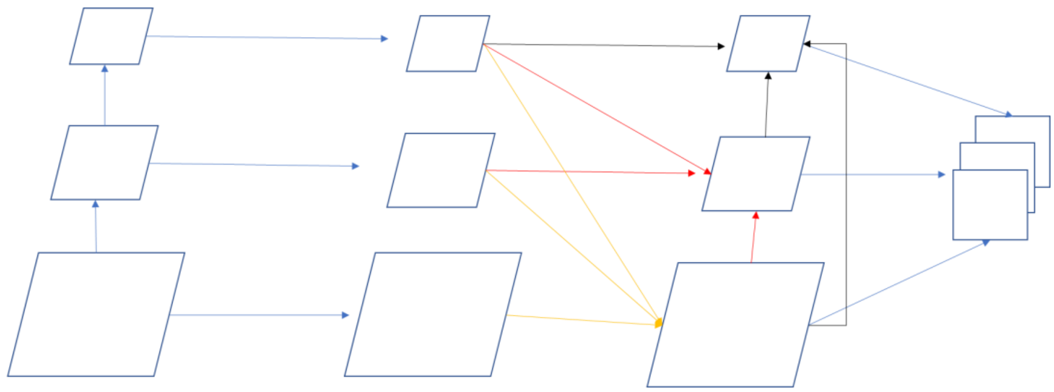

Figure 5. Yellow Line: Create The biggest image. Red Line: Create a Medium image. Black Line: Create the smallest image.

Theoretically, we believe that the feature pyramid method will be better than another pyramid method. We not only consider top-down or button-up method feature pyramid, but also use the residual and dense idea. Therefore, we have strong confidence that our model will improve object detection IOU accuracy. And we will conduct the related experiments as soon as possible. For each model, we will run 10 times and calculate the average value as a result.

Final Model

The following is a proposed model, which combines all methods.

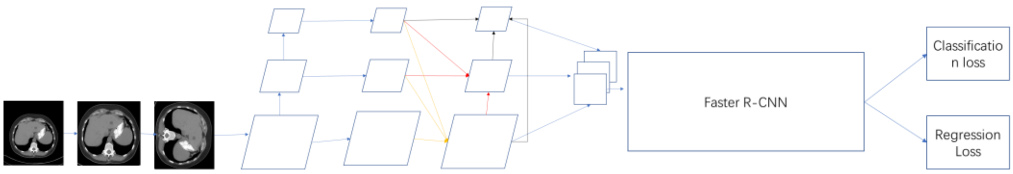

Figure 6. Proposed Model. First, using pre-cropping to crop the image. Then, using data augmentation, such as rotate the image. After that, using the feature pyramid method. Finally, the image will go to Faster R-CNN[3] to train the model.

The above is our proposed model. We use pre-cropping, data augmentation, and feature pyramid sequentially. After that, Faster R-CNN will be trained by the processed images. The loss is classification Loss and regression Loss.

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*).$$

Figure 7. classification Loss and regression Loss defined in Faster R-CNN[3]. p^* and t^* are ground truth, and p_i and t_i are predicted. N_{cls} and N_{reg} are coefficients defined to balance the weight of

classification loss and regression loss.

L_{close} is cross-entropy loss, and L_{reg} is smooth L1 loss defined in Fast R-CNN[10]

In the model, we mainly change the image pre-processing part, including pre-cropping, data augmentation, and feature pyramid. For the Faster R-CNN part, we hardly change the model structure. We will do the corresponding experiments as soon as possible to verify that our model is better than the original Faster R-CNN. To get a high confident experiment result, for our proposed model and original Faster model, we will run five times and calculate average IOU accuracy as a result.

References:

[1]. Lee, S. Y., Kim, H. J., & Choi, D. (2015). Cell sources, liver support systems, and liver tissue engineering: alternatives to liver transplantation. International journal of stem cells, 8(1), 36.

[2]. https://www.webmd.com/hepatitis/liver-lesions#1

[3]. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[4]. Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017

[5]. Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.

[6]. Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition. 2018: 8759-8768.

[7]. Pang J, Chen K, Shi J, et al. Libra R-CNN: Towards Balanced Learning for Object Detection[J]. arXiv preprint arXiv:1904.02701, 2019

[8]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778).

[9]. Gao H., Liu, Z., Weinberger, K. Q., & Maaten, L. van der. (2016). Densely Connected Convolutional Networks. ArXiv Preprint ArXiv:1608.06993.

[10]. Girshick, R.B. (2015). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1440-1448.