Detection of spinal fracture lesions based on Improved Yolov3

Gang Sha1, a, Junsheng Wu2, b and Bin Yu3, c

1School of computer science, Northwestern Polytechnical University, Xi’an 710072, Shaanxi, Xian, China
2School of software and microelectronics, Northwestern Polytechnical University, Xi’an 710072, Shaanxi, Xian, China
3School of computer science and technology, Xi’dian University, Xi’an 710071, Shaanxi, Xian, China

a shagang@mail.nwpu.edu.cn, bwujunsheng@nwpu.edu.cn, cyui-107@foxmail.com

Abstract: Yolo[1]algorithm has a good detection effect in target detection. Because of its high detection accuracy and fast detection speed, it is widely used in practice. Because of the problem that the complexity of spine CT images, the irregular shape of vertebral boundary, which needs doctors’ prior knowledge and clinical experience to determine lesions location in CT images, so it can not meet the clinical real-time needs. In this paper, We use deep learning to process the CT images of spine, and to detect and locate lesion of (cervical fracture, cfracture), (thoracic fracture, tfracture), (lumbar fracture, lfracture) by the improved YOLOv3. Through using lesions bounding box dimensional cluster, multiscale transformation of input CT images, and change NMS to MAX value to improve detection efficiency and accuracy. The experiment shows the results are more accurate, and mAP (mean average precision) of detection algorithm is 73.63%, detection rate is 0.027 seconds per detection, and IOU((Intersection-over-Union) is 75.9, which can basically meet the clinical real-time needs.

Keyword: Deep learning Yolov3 Detection and location

1. Introduction

At present, spondylosis has become a common and high incidence, the traditional way to view CT images by doctors is checking CT images one by one, and judging there are lesions whether or not by experience, so it takes about 10 minutes to check a set of CT images, and need to watch and confirm repeatedly. By using deep learning to view CT images, the initial screen checking can be completed quickly, and computer can observe the whole slice completely without omission, and the diagnosis results can be kept completely objective.

There are two kinds detection algorithm base on deep learning: (1) target detection algorithm based on Region recommended, such as RCNN(Region with CNN features)[2], Fast-RCNN(Fast-Regions with CNN features)[3], Faster-RCNN[2] target detection algorithm based on regression, such as YOLO(You Only Look Once)[4], SSD(SingleShot Multibox Detector)[5].

Yolov3 refers to the network structure of Yolo and SSD, and designs a new classified network darknet-53 as the basic network model. Before Yolov3, most of the target detection frameworks used VGG-16[2] as feature extraction network. The spinal fracture lesion detection algorithm we used in this paper belongs to the field of target detection based on Region recommended.

1 Spinal fracture lesion detection system based on Yolov3
Figure 1 is the overall framework of spinal fracture lesion detection system in this paper. First, we select appropriate spine CT images from hospital to label spinal fracture lesions and make CT images into VOC 2007 format for training and testing; Second, the input image is resized by model and then divided into S×S grids. If detection lesions is located in a grid, the corresponding grid is responsible for identifying the target lesions. In the process of lesionst recognition, it is necessary to predict the location of the lesions bounding box and the category of the lesions. We assume that there are C classes of lesions categories in the image, in every grid, so the probability P of generating B lesions bounding box and C lesions category is $P(\text{class}, \text{obj})_{i=1, \ldots, C}$. Each bounding box contains five values: x, y, w, h and score. The (x, y) coordinates represent the center of the bounding box. w and h are predicted width and height relative to the image. Score represents the IOU between the predicted box and the actual bounding box, we definite score as formula (1):

$$score = P(\text{obj}) \times IOU$$  \hspace{1cm} (1)$$

Where: P (obj) represents the probability that there is a lesion in the predicted bounding box. If there is a target lesion, P (object) = 1, otherwise P (OB object) = 0; IOU is the intersection of the predicted lesion bounding box and the real lesion bounding box, reflecting the accuracy of the predicted lesion bounding box. Then we definite category confidence score as formula (2):

$$\text{Class}_{i, \text{score}} = score \times P(\text{class}, \text{obj})_{i=1, \ldots, C} = P(\text{obj}) \times IOU \times P(\text{class}, \text{obj})_{i=1, \ldots, C}$$  \hspace{1cm} (2)$$

$\text{Class}_{i, \text{score}}$ reflects the probability of i category lesion in the related grid and accuracy of the predicted lesion bounding box. Then we compare $\text{Class}_{i, \text{score}}$ with the specified threshold value, which in this paper is 0.5. If it is greater than 0.5, the corresponding predicted bounding box is reserved, otherwise, the corresponding predicted bounding box is discarded. Finally, we filter the remaining predicted lesions bounding box by NMS (Non-Maximum Suppression), and get predicting results.
2. Improvement of Yolov3

2.1 Lesions bounding box dimensional cluster
The anchor in yolov3 refers to the initial candidate box with a certain size and aspect ratio. The anchor will have an impact on each index detected. Through dimensional clustering, we can analyze and understand the statistical law of the target box in the data set. The number of clusters is the same as that of the anchor, which is represented by K. We cluster the dimensions into 4 categories, so the anchor in Yolov3 is 4.

2.2 Multiscale transformation of input CT images
Yolov3 can change the size of the input image. It uses the method of multi-scale input to train the detection network. During the training process, the input size of the model is changed every 10 epoches, so the model has robustness for different size images, which is shown in formula (3):

\[ S = 32 \times (7 + n) \]  

(3)

Where, S is the size of input CT image, n is a random number in 0 to 12.

2.3 change NMS to MAX
As we mentioned above, we can get predicting lesions from formula (1) by using NMS to filter remaining predicted lesions bounding box, but in this paper, we change NMS to maximum operation, which is the maximum value among several groups of comprehensive scores that are greater than the threshold value, and uniquely determine the lesions bounding box location and its prediction category, the unique location of lesion bounding box and its prediction category is expressed as formula (4):

\[ (i, P_{\text{pred}}) = \max[P(\text{obj})_i \times IOU], P(\text{obj})_i \times IOU \geq T_{\text{thresh}} \]  

(4)

Where, \( P_{\text{pred}} \) is the predicting location of lesions bounding box, i is the category, \( T_{\text{thresh}} \) is the threshold value, 0.5 in this paper. Through the maximum operation, only one lesion bounding box can be detected in the detection area, so as to improve the accuracy of detection.

![Training flow for lesions detection by Improved Yolov3](image)

**Figure 2.** Training flow for lesions detection by Improved Yolov3
3 Experiment and Analysis

3.1. Experimental data
The experimental data in this paper are accumulated clinical data from Xijing Hospital, the Second Affiliated Hospital of Shaanxi University of traditional Chinese medicine and other hospitals. We use LabelImg to mark the location, size and category of the lesion in each CT image, and classify as (cervical fracture, cfracture), (thoracic fracture, tfracture), (lumbar fracture, lfracture) according to lesion category. Then make lesion data to VOC 2007 format for training and testing.

We get 40 cases spinal fracture clinical data from hospital, which including 5134 CT images, it is impossible to include all kinds of spinal fracture lesions. So we add data by data augmentation which rotating the image 180 degrees, and flinging left and right, up and down. After data enhancement, the data set of spinal focus was expanded to 10268 CT images.

We randomly generates the training set and testing set according to the ratio of 4:1, In training verification set, 80% are used as training set, and the remaining 20% are used as verification set. So 4108 CT images are selected for training sets and 1026 CT images are test set. Figure 2 is the training flow from original CT images to ultimately predicted lesions by improved Yolov2.

3.2. Experimental environment
We use CPU: Intel Core i7-6700 @ 3.40GHz * 8, memory 32G, graphics card GTX1070, 8G, 128G + 1t hard disk, operating system is Ubuntu 16.04, Caffe, Cuda8.0 and Cudnn6.0 to train and test data. The super parameters are as follow: learning rate:0.0001, momentum: 0.9, decay: 0.0005, batch size:8, subdivisions:4, steps: 30000,40000, max_batches: 50200.

3.3. Experiment
3.3.1 Data augmentation and comparison. Table 1 is the comparison of training data is expanded from 4108 images to 8216 by using 6 model(Yolov3_288, Yolov3_416, Yolov3_480, Yolov3_512, Yolov3_544, Yolov3_608). As we can see in Table 1, with increasing size of input CT images, AP is increasing and IOU is also increasing. Therefore, expansion of the training set is conducive to the network to fully learn the characteristics and improve the performance. Our later experiments are based on data set augmentation.

| Data sets | Model    | AP(%) | IOU  |
|-----------|----------|-------|------|
| 4108      | Yolov3_288 | 73.13 | 63.9 |
|           | Yolov3_416 | 73.25 | 65.1 |
|           | Yolov3_480 | 75.90 | 68.1 |
|           | Yolov3_512 | 76.10 | 70.5 |
|           | Yolov3_544 | 78.10 | 71.8 |
|           | Yolov3_608 | 81.20 | 73.1 |
| 8216      | Yolov3_288 | 73.10 | 65.2 |
|           | Yolov3_416 | 73.50 | 67.1 |
|           | Yolov3_480 | 75.11 | 70.3 |
|           | Yolov3_512 | 78.23 | 71.5 |
|           | Yolov3_544 | 81.51 | 73.1 |
|           | Yolov3_608 | 83.60 | 75.3 |
3.3.2 Comparison of different input CT image sizes

Table 2. Different training model

| Model         | Lesion detection | AP(%) | mAP(%) | IOU  | Time(s) |
|---------------|------------------|-------|--------|------|----------|
| Yolov3_288    | cfracture        | 68.01 | 69.27  | 65.2 | 0.02425  |
|               | tfracture        | 69.50 |        |      |          |
|               | lfracture        | 70.30 |        |      |          |
| Yolov3_416    | cfracture        | 69.25 | 70.43  | 67.1 | 0.02450  |
|               | tfracture        | 70.50 |        |      |          |
|               | lfracture        | 71.55 |        |      |          |
| Yolov3_480    | cfracture        | 70.50 | 71.30  | 70.3 | 0.02675  |
|               | tfracture        | 71.95 |        |      |          |
|               | lfracture        | 71.45 |        |      |          |
| Yolov3_512    | cfracture        | 69.50 | 70.30  | 72.5 | 0.02775  |
|               | tfracture        | 69.95 |        |      |          |
|               | lfracture        | 71.45 |        |      |          |
| Yolov3_544    | cfracture        | 71.50 | 73.63  | 75.9 | 0.02895  |
|               | tfracture        | 73.95 |        |      |          |
|               | lfracture        | 75.45 |        |      |          |

We change input CT images from 288 to 608 by using 6 different model to train, it can be seen from Table 2 after increasing CT images input size, AP, mAP, and IOU is increasing, but we need more time to detect lesions. So changing CT images input size can effectively improve the effect of target detection at the cost of more time.

3.3.3 Average IOU and Loss curve of different models

![512_avg_iou](image1.png) ![512_avg_loss](image2.png) ![608_avg_iou](image3.png) ![608_avg_loss](image4.png)

Figure 3. avg_iou and avg_loss
We only show avg_iou and avg_loss of Yolov3_512 and Yolov3_608 in Figure 3, because of the length of this paper, we can see that avg_iou is more near to 1 and avg_loss curve changes to smooth with input size of CT images increase from 512 to 608.

3.3.4. Comparison of Different Algorithms. We compare our algorithm to Faster-RCNN and YOLOv3 as shown in Table 3. YOLOv3 takes short time to detect lesion but has low accuracy, and Faster-RCNN detection velocity is low, we improve the accuracy and detection velocity meanwhile by using lesions bounding box dimensional cluster and multiscale transformation of input CT images.

| Detection network | mAP(%) | Time(s) |
|-------------------|--------|---------|
| Faster-RCNN       | 71.56  | 0.041   |
| Yolov3            | 72.62  | 0.025   |
| Ours              | 75.30  | 0.027   |

4. Conclusion
In this paper we studies the detection of spinal fracture lesion based on improved Yolov3, expands the data set through data augmentation, solves the problem of insufficient experimental medical data, and according to characteristics of spinal CT data, proposes to use lesions bounding box dimensional cluster and multiscale transformation of input CT images, and change NMS to MAX value to improve accuracy and efficiency of spinal fracture lesions detection, and finally gets mAP 73.63%, IOU 75.9. The detection time of a single CT image is 0.027 seconds, which can basically meet the clinical needs of the hospital.

Reference
[1] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2015: 779-788.
[2] ZEILE R M, DFE R GUS R Visualizing and understanding convolutional network [DB/OL] [2018-10-22]. https://arxiv.org/pdf/1311.2901.pdf.
[3] DAI J, LI Y, HE K, et al. R-FCN: object detection via region-based fully convolutional networks [DB/OL]. [2018-10-22]. https://arxiv.org/pdf
[4] Matthew D. Zeiler and Rob FergusD. Fleet et al. (Eds.): ECCV 2014, Part I, LNCS 8689, pp. 818–833, 2014.
[5] SIMONYAN K, ZISSE R MAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. [2018-10-22]. https://arxiv.org/pdf/1409.1556. Pdf