

Research on detection method of lumbar disc herniation based on one-stage object detection

Zehua He, Zimin Wang*, Xia Li, Yue Zhou, Tingqiang Guan, Xin Guo
School of Computer Science and Information Security, Guilin University of Electronic Technology,
Guilin, Guangxi, China

ABSTRACT

Today, medical and health care is playing an increasingly important role in people's lives, the number of medical images increases rapidly, and the traditional manual diagnosis can't meet the increasing demand of clinical diagnosis. It is the core of this research to realize the localization and classification of lesions by training a neural network, and to assist doctors in clinical diagnosis. Based on the one-stage object detection network FCOS, this paper improves the detection accuracy to a level comparable to the two-stage detection network while taking into account the detection speed, and achieves a mAP of 0.664 on the MRI test set, which is higher than most existing networks. Overall, the proposed network has good performance in clinical assistance, and can be competent for some clinical application tasks of real-time detection.

Keywords: Deep learning, object detection, computer-aided diagnosis system, lumbar disc herniation, magnetic resonance imaging

1. INTRODUCTION

Low back pain (LBP) has become a hackneyed spinal surgery question in today's era of overtime. According to a survey¹, approximately 70% to 85% of adults are affected by lumbar spine disorders at some point in their lives. In China, lumbar disc herniation (LDH) patients account for the majority of patients with lumbar disease.

Conventional medical image processing still relies on the man-made fetching method, but with the explosive increase of medical view data, the disadvantages of the artificial reading method are steadily revealed. Computer-aided diagnosis (CAD) means comprehensive imaging, medical image processing technology and other potential biological means, through computer analysis and detection, the lesions can be found and the diagnostic accuracy can be improved. The first attempt of CAD system was in the 1960s².

Along with evolution of computer vision and picture processing techniques, the use of deep learning methods to process medical images and assist physicians in clinical diagnosis has become a popular means³. Many studies have emerged to analyze medical images more accurately, some of which focus on specific deformations or injuries in spine images^{4, 5}, others aim at automatically detect vertebrae^{6, 7}. Zhao et al.⁸ performed detection of spine MRI based on the category consistent self-calibration detection framework; Han et al.⁹ applied generative adversarial network (GAN) to semantic segmentation of the spine; Wang et al.¹⁰ realized automatic vertebral localization and identification in CT by training a key point localization model and introducing an anatomical constraint optimization module; Zhang et al.¹¹ raised a serial conditional strengthening learning network that innovatively models the top-down spatial correlation between vertebrae as a continuous dynamic interaction process, and thus conducting global focused detection and segmentation of each vertebra.

2. OUR APPROACH

2.1 Structure of network

At present, the more advanced one-stage object detection network is the FCOS network proposed by Tian et al.¹², it has a faster detection speed than the two-stage detection network and a similar detection accuracy, which is the basis of this paper.

* worthyman@guet.edu.cn

FCOS network achieves proposal-free and anchor-free by many clever designs (FPN, center-ness, scale limitation of feature point regression in each layer, etc.). At the same time, it avoids complex IOU calculation and the matching between anchor and GT Bounding box during training. The network architecture of this study can be seen in Figure 1, which adopts feature pyramid Network (FPN) and three-branch head detection network.

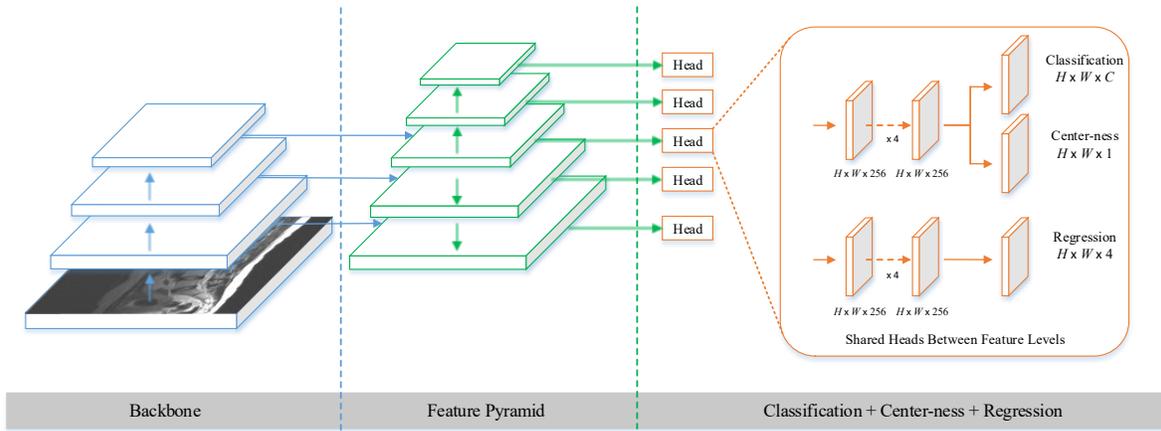


Figure 1. Structure of the network.

2.2 Self-calibrated convolutions

In order to further optimize the network parameters on the foundation of the FCOS network and improve network accuracy, this paper draws on the SC-Net suggested by Liu et al.¹³. The network mainly uses a self-calibrating convolutional SCConv, whose structure is shown in Figure 2.

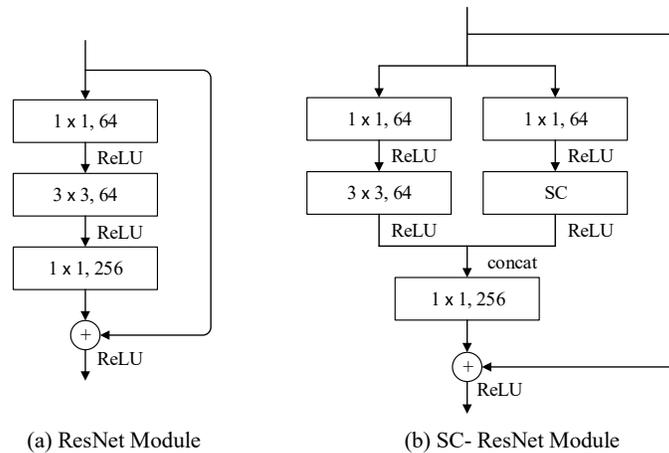


Figure 2. Structure of the SCConv module.

The design of SCConv is simple and versatile, and can easily enhance the performance of standard convolution layers without introducing additional parameters and complexity. The architecture of the SC module is shown in Figure 3.

2.3 Squeeze-and-excitation blocks

To better exploit the dynamic relationship between feature channels, this paper introduces the SE-Net proposed by Hu et al.¹⁴. SE-Net is very simple in construction and apt to instruct, there is no need to introduce new functions or layers, with good performance in the aspect of parameter complexity and network structure. The network structure is shown in Figure 4. Figure 5 reveals the structure of embedding the SE module into the Res-Net module.

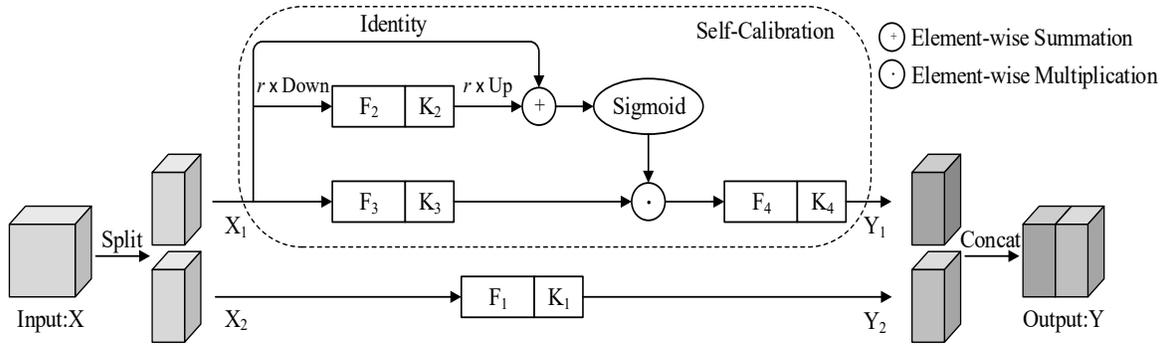


Figure 3. Schematic illustration of the self-calibrated convolutions.

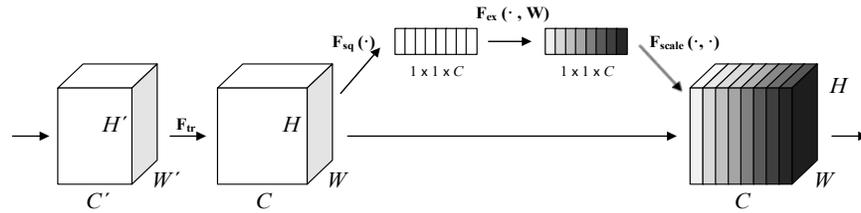


Figure 4. Squeeze-and-excitation block.

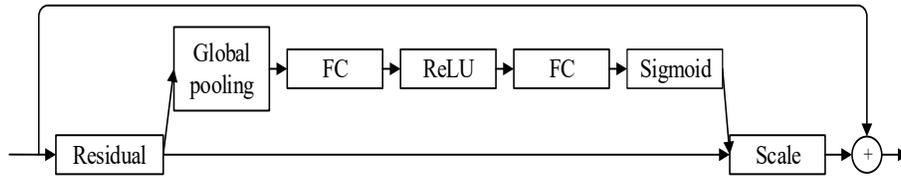


Figure 5. The architecture of the SE-ResNet module.

In this paper, we use SE-ResNet to focus on the channel relationship of the network, Squeeze operation to establish the dependency relationship between channels, and Excitation operation to recalibrate the features. The combination of the two emphasizes the useful features and suppresses the useless features, which can effectively improve the model performance and increase the accuracy rate.

2.4 Soft-NMS

In our study, we use Soft-NMS¹⁵, which mainly solves the problem of excessive deletion of boxes by NMS, Soft-NMS has learned the lessons of NMS, during the execution of the algorithm, instead of simply deleting the detection box whose IOU is above the threshold, lower its score. The algorithm process is the same as NMS, but the function operation is used for the original confidence score, and the goal is to reduce the confidence score. Soft-NMS is expressed as follows.

$$s_i = \begin{cases} s_i, & iou(M, b_i) < N_t \\ s_i(1 - iou(M, b_i)), & iou(M, b_i) \geq N_t \end{cases} \quad (1)$$

2.5. Improved res module

Based on the improvement of the traditional Res-Net network, the SC module and SE module are introduced to modify the residual module in the traditional Res-Net network. The improved module architecture is shown in Figure 6.

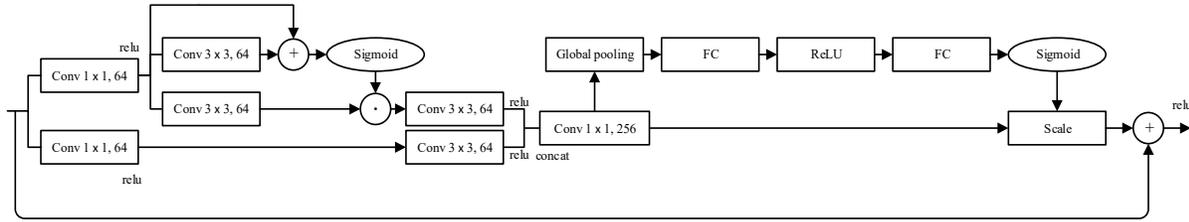


Figure 6. The architecture of the improved Res-Net module.

2.6 Loss function

Three parts constitute the loss function of this network, Focal loss for classification loss, IOU loss for regression loss and center-ness loss for BCE. The SoftMax is discarded in the classification loss, and the sigmoid function is used for each channel (each channel represents a category) of the classification output by the head, and then the Focal loss is used. IOU Loss only performs regression calculation for those meaningful feature points. The specific loss function is as below.

$$L(\{p_{x,y}\}, \{t_{x,y}\}) = \frac{1}{N_{pos}} \sum_{x,y} L_{cls}(p_{x,y}, c_{x,y}^*) + \frac{\lambda}{N_{pos}} \sum_{x,y} \mathbb{1}_{\{c_{x,y}^* > 0\}} L_{reg}(t_{x,y}, t_{x,y}^*) \quad (2)$$

3. EXPERIMENTS

3.1 Implementation details

The experimental environment of this paper is 20.04.3-Ubuntu system, the processor model is Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40GHz, the graphics card model is Nvidia Quadro M4000 8GB, the python version used is 3.6.13. We use pytorch to build the deep learning framework.

In this paper, we use a non-public human lumbar disc MRI-T2 image dataset collected from the Internet with 470 images, which is divided into train dataset (376 images), val dataset (47 images) and test dataset (47 images) by 8:1:1. The target categories in the dataset are divided into normal and herniated. The training process is optimized using stochastic gradient descent (SGD) with an initial learning rate of 0.005 and momentum of 0.9.

3.2 Results on MRI

The experiment was carried out according to the five-fold cross-validation, take the mean value of five experimental results as the final experimental result. Partial detection results of the model on MRI dataset are shown in Figure 7.

Laboratory result of some model on MRI test set are listed in Table 1. The network model proposed in this paper has few parameters, and its detection speed is faster and the detection accuracy is the highest among all models. Compared with the generic Faster R-CNN and Retina-Net, it has a large performance improvement, which fully reflects the practical application value of the model. Moreover, our network has no complex network structure and has great advantages in the deployment of the network model.

3.3 Ablation study

For verifying the availability of the modules introduced in our network, we conducted ablation experiments on the main modules to check on their role in the network. The results of ablation experiments are shown in Table 2. We can see that the module we introduced improves the performance of the network.

Furthermore, we performed an ablation experiment on the SE modules to evaluate the effect of their alignment position when integrating them into existing frameworks. Figure 8 shows the structure of these variants, and Table 3 shows the property of these variants. As can be seen from the laboratory results, the performance of SE-PRE block, SE-Identity block, and Standard SE block is similar, while the use of the SE-POST block results in a degradation in capability. The test shows that the property improvement brought by the SE block is robust to its position change if it is used before branch aggregation.

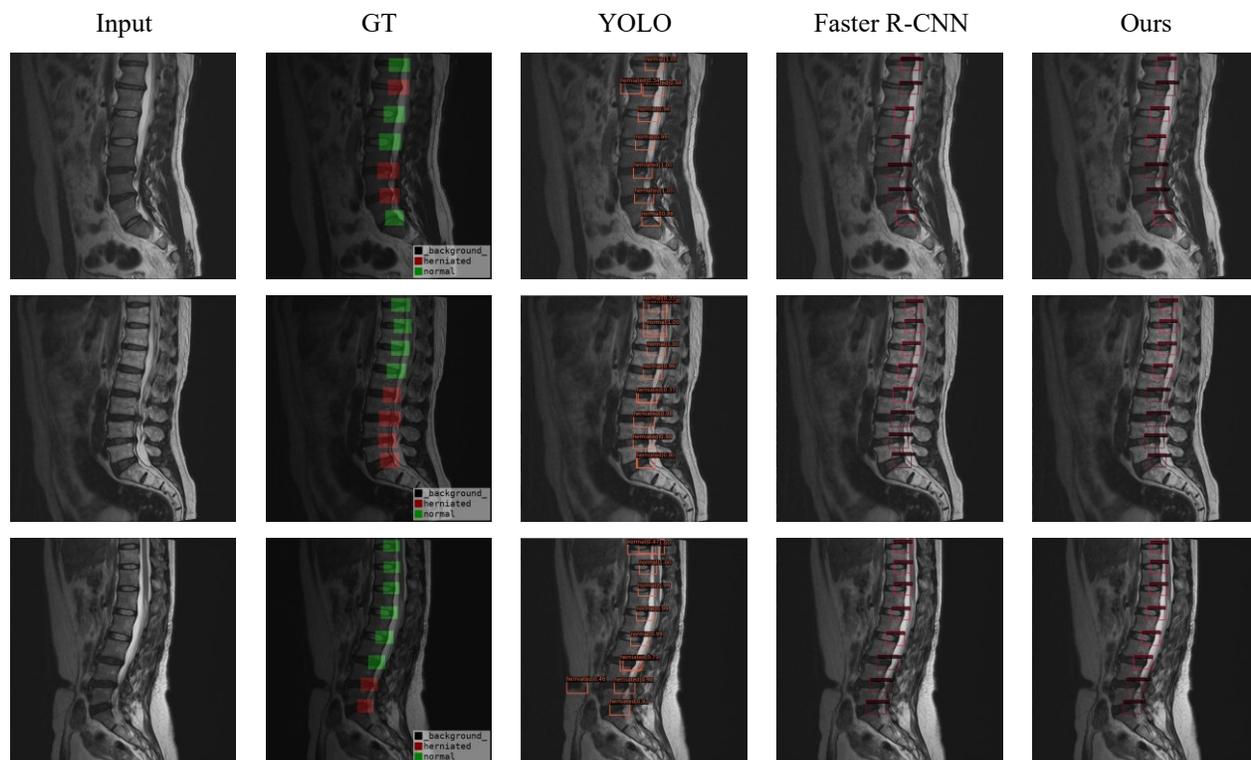


Figure 7. Some detection results on spine MRI dataset.

Table 1. Comparisons on spine MRI dataset.

Model	Backbone	mAP-1	mAP-2	mAP-3	mAP-4	mAP-5	mAP
Faster R-CNN	Res-50-FPN	0.640	0.651	0.649	0.650	0.645	0.647
Retina-Net	Res-50-FPN	0.648	0.641	0.645	0.639	0.646	0.644
YOLOv2 ¹⁶	DarkNet-19	0.635	0.633	0.629	0.630	0.629	0.631
SSD	Res-50-SSD	0.641	0.639	0.635	0.640	0.640	0.639
Ours	Res-50-FPN	0.665	0.662	0.666	0.662	0.665	0.664

Table 2. Ablation study on spine MRI dataset.

	Faster R-CNN	Retina-Net	Ours
-	0.647	0.647	0.655
SC	0.649	0.649	0.658
SE	0.650	0.650	0.660
SC+SE	0.654	0.652	0.664

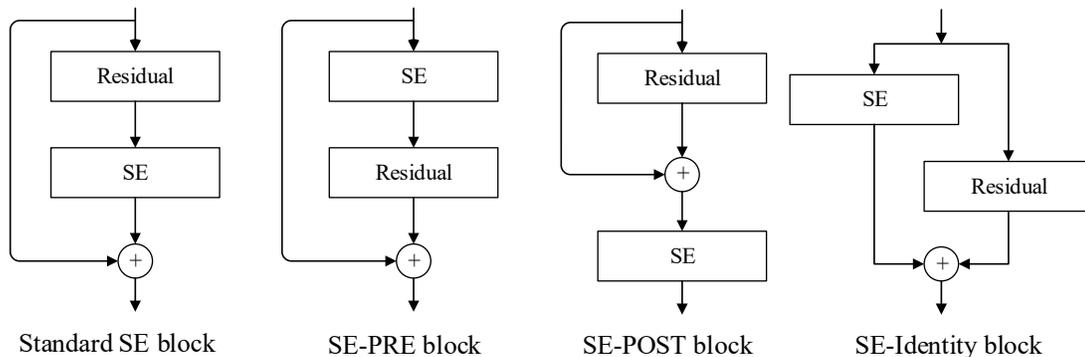


Figure 8. SE block with different structures.

Table 3. Effect of different SE block.

Design	mAP
SE	0.659
SE-PRE	0.660
SE-POST	0.647
SE-Identity	0.656

4. CONCLUSIONS AND FUTURE WORK

We propose an anchor-free and proposal-free one-stage detector that incorporates the SCConv module and the SE attention mechanism, the network property is greatly improved. Laboratory results on spine MRI show that it outperforms currently fashionable anchor-based single-stage detectors, comprises Retina-Net, YOLO, and SSD, and the design complexity is much lower. Not only that, we also envision to combine the FCOS network with Faster R-CNN, in the cause of further enhance the detection speed of our network without increasing additional parameters and operations to satisfy the demands of doctors in the aspect of detection speed and design a more perfect clinical auxiliary diagnosis system.

REFERENCES

- [1] Ohtori, S., Inoue, G., Orita, S., et al., "No acceleration of intervertebral disc degeneration after a single injection of bupivacaine in young age group with follow-up of 5 years," *Asian Spine Journal*, 7(3), 212(2013).
- [2] Lodwick, G. S., Keats, T. E. and Dorst, J. P., "The coding of roentgen images for computer analysis as applied to lung cancer," *Radiology*, 81(2), 185-200(1963).
- [3] Zhang, X., "New concept of the development of modern medicine: Make full use of the internet, large data, and artificial intelligence," *Chinese Journal of Lung Cancer*, 21(3), 141-2(2018).
- [4] Anitha, H. and Prabhu, G. K., "Identification of apical vertebra for grading of idiopathic scoliosis using image processing," *Journal of Digital Imaging*, 25(1), 155-61(2012).
- [5] Kumar, S., Nayak, K. P. and Hareesha, K. S., "Improving visibility of stereo-radiographic spine reconstruction with geometric inferences," *Journal of Digital Imaging*, 29(2), 226-34(2016).
- [6] Kumar, V. P. D. and Thomas, T., "Automatic estimation of orientation and position of spine in digitized X-rays using mathematical morphology," *Journal of Digital Imaging*, 18(3), 234-41(2005).
- [7] Benjelloun, M. and Mahmoudi, S., "Spine localization in X-ray images using interest point detection," *Journal of Digital Imaging*, 22(3), 309-18(2009).
- [8] Zhao, S., Wu, X., Chen, B., et al., "Automatic vertebrae recognition from arbitrary spine MRI images by a category—Consistent self-calibration detection framework," *Medical Image Analysis*, 67, 101826(2021).
- [9] Han, Z., Wei, B., Mercado, A., et al., "Spine-GAN: Semantic segmentation of multiple spinal structures," *Medical Image Analysis*, 50, 23-35(2018).

- [10] Wang, F., Zheng, K., Lu, L., et al., "Automatic vertebra localization and identification in CT by spine rectification and anatomically-constrained optimization," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 5280-8(2021).
- [11] Zhang, D., Chen, B. and Li, S., "Sequential conditional reinforcement learning for simultaneous vertebral body detection and segmentation with modeling the spine anatomy," Medical Image Analysis, 67, 101861(2021).
- [12] Tian, Z., Shen, C., Chen, H., et al., "Fcos: Fully convolutional one-stage object detection," Proc. of the IEEE/CVF Inter. Conf. on Computer Vision, 9627-36(2019).
- [13] Liu, J., Hou, Q., Cheng, M., et al., "Improving convolutional networks with self-calibrated convolutions," Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 10096-105(2020).
- [14] Hu, J., Shen, L. and Sun, G., "Squeeze-and-excitation networks," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 7132-41(2018).
- [15] Bodla, N., Singh, B., Chellappa, R., et al., "Soft-NMS--improving object detection with one line of code," Proc. of the IEEE Inter. Conf. on Computer Vision, 5561-9(2017).
- [16] Redmon, J. AND Farhadi, A., "YOLO9000: Better, faster, stronger," Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 7263-71(2017).