Automatic CT Lesion Detection Based on Feature Pyramid Inference with Multi-scale Response

Yangyang Tang[1], Zhe Liu[1](✉), Yuqing Song[1], Kai Han[1], Jun Su[1], Wenqiang Wang[1], Fan Hu[2], and Jiawen Zhang[3]

[1] School of Computer Science and Communications Engineering, Jiangsu University, Zhenjiang 212013, China
1000004088@ujs.edu.cn

[2] School of Financial Management, College of Mobile Telecommunications Chongqing, University of Posts and Telecom, Chongqing 401520, China

[3] Huashan Hospital, Fudan University, Department of Radiology, 12 Middle Wulumuqi Road, Shanghai 200040, China

Abstract. In modern medical applications, computed tomography image is used as one of the most useful tools for diagnosis and localization of lesions. It can provide patients with precise information about the location and size of their tumor lesions. Traditional medical diagnosis is not only very time consuming but also not very accurate. Nowadays, the automatic detection of lesions on computed tomography has become a research area of great interest, and researchers aim to use computer-aided diagnosis to assist in clinical medical diagnosis. However, for current detection algorithms, the accuracy of automatic lesion detection is still low, especially for small lesions. In this paper, to improve the accuracy of detection of small lesions, we propose a Multi-Scale Response Module (MSR) that incorporates global attention into Feature Pyramid Network (FPN) build on backbone. At each pyramid level, the proposed Aggregated Dilation Block (ADB) is used to capture the variations in the fine-grained scales. The response of the network to small lesion features is then reinforced by the Global Attention Block (GAB). We build a Feature Pyramid Network (FPN) based on the highly responsive output of the MSR module, with each layer of the FPN fusing high semantic information from low resolution layers. The experimental results show that our method has a higher detection accuracy with mAP value of 58.4 and a high sensitivity compared to the state-of-the-art methods.

Keywords: Deep lesion detection · Attention mechanism · Dilated convolution · RCNN

1 Introduction

According to the most recent statistics, we found that the diseases with the highest mortality rates in the world are malignant tumors, heart disease, pneumonia, and cerebrovascular diseases. In addition, the number of deaths caused by malignant tumors is

X. Sun et al. (Eds.): ICAIS 2021, LNCS 12736, pp. 167–179, 2021.
https://doi.org/10.1007/978-3-030-78609-0_15

increasing each year [1]. Therefore, accurate diagnosis is of great significance to the discovery and treatment of tumor disease. Diagnostic imaging is the most common method used to detect cancer, especially the rapid development of medical imaging technology in recent years has made it possible to obtain high-resolution CT and MRI datasets. At the same time, it also makes it possible to train high-resolution CT images to detect lesions.

In the diagnosis based on CT images, clinicians judge the presence of tumor based on their subjective medical knowledge and relevant laboratory report data, which is largely dependent on the experience of the doctor. However, with the rapid advancement of medical information systems in recent years, doctors now can easily access the medical image data of patients to facilitate the diagnosis of the disease, the massive task of analyzing lesions on computed tomography (CT) and magnetic resonance images (MRI) images is onerous for the physician. Therefore, we propose to automate the detection of tumors in computed tomography images by computer, aiming at alleviating the burden on doctors and expecting to achieve higher accuracy of analysis.

Currently, many laboratories and scholars in the field of computer vision and image recognition are attempting to introduce object detection techniques in the medical field. For example, [2–8] showcase some of the explorations that researchers have made in the field of medicine in recent years. However, the size of tumors varies greatly within a certain range. For example, in DeepLesion dataset, the size of the lesion area varies from 0.21 mm to 342.5 mm. As for the treatment of such large differences in spatial scale changes, the above methods failed to deal well with the low precision of detection of small lesions. Especially for small lesion regions detection, which is an important evidence to detect early lesions in the body, it is necessary to improve the detection accuracy for small lesions.

In this paper, we propose a fine-grained lesion detection method with a novel multi-scale global attention mechanism. For the samples of small lesions in dataset (DeepLesion), we improve the network structure based on the two-stage network of Faster RCNN and use ResNet101 network as backbone. First of all, we use dilation and erosion operations in a mathematical morphology approach to process the input CT images, which makes the diseased areas more visible. In order to fuse more semantic information when constructing the feature pyramid, we continue to deepen the network based on ResNet101 until the resolution of the feature map is 8×8. After up-sampling, the output of each layer residue block is fused with the high-resolution topographic map to preserve as much spatial and semantic information as possible at different scales. We then use multi-scale response (MSR) to facilitate lesion detection across fine-grained scales. Given a feature map with a specific resolution, the aggregated dilation block (ADB) in MSR is based on the split-transform merge principle, using the regional correlations in each pyramid feature generation block. As the dilation rate increases, new feature maps are generated from a wide range of contexts. Aggregated dilation blocks further increase the receptive fields of the top-down paths in the feature pyramid. The channel and spatial attention modules in the MSR then focus on the different lesion responses in feature maps. Finally, we re-sample the high-response features with the upper feature map and fuse it with the upper map again, which is then fed into the RPN to obtain the coordinates of the prediction score and the bounding box of the output category. The experiments show that the network using the MSR module has a significant improvement in accuracy compared to the original two-stage network.

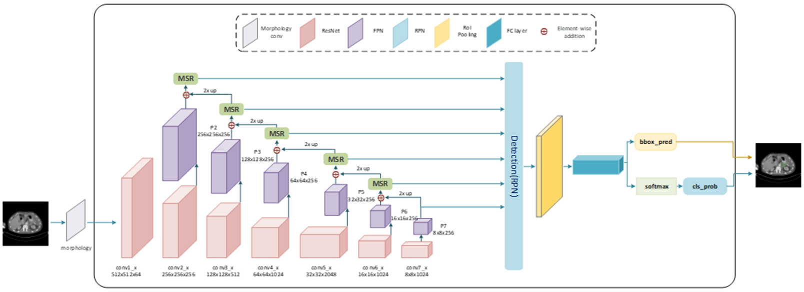

Fig. 1. The framework of the proposed lesion detection method

The main contributions of this work can be summarized as follows.

1) An Aggregated Dilation Block (ADB) is proposed. The block alleviates the shortcomings of the low-resolution feature layer in the network due to the large receptive field and the small target features are not obvious.
2) The Global Attention Block (GAB) is designed to reduce the influence of background noise and highlight the target features, which is effective for detecting obscure objects at different scales.

2 Proposed Method

The structure of framework we proposed is illustrated in Fig. 1. Firstly, we selected different morphological operations (dilation and erosion) for tumors with different colors in each organ as pre-processing, and then these feature-enhanced maps were fed into the network. We extracted features from the ResNet backbone (P2-P7) (Sect. 2.2). The extracted features are further processed by ADB (Sect. 2.3) and GAB (Sect. 2.4) in the MSR module. The output of the MSR is subsequently fed into the RPN while being up-sampled and summed with the output of the previous layer of high-resolution feature maps to create a feature pyramid network constructed from the output of the MSR module. Finally, the network undergoes Softmax loss (detection of classification probability) and Smooth L1 loss (detection of frame regression) to train classification probability and bounding box regression.

2.1 Mathematical Morphology Operations

Mathematical morphology image processing refers to a series of image processing techniques to enhance image shape features. The basic idea of morphology is feature enhancement by using a special structure element to measure or extract the corresponding shape or feature when given an input image. Common morphological processing methods include: erosion, dilation, opening and closing. Morphological operations rely on the correlation of pixel values rather than their absolute values, so they are very suitable for binary image processing. By constructing the corresponding mathematical morphology structure element suitable for DeepLesion dataset, the maximum or minimum area can be effectively found in the image, which can reduce some noises in CT scanning and make the lesions more obvious. For the DeepLesion dataset, we could roughly divide it into two categories, dark background with light color lesion area or the opposite. We selected the CT images of representative Lung and Liver organs, which respectively represented the images of the previous two background types. Through the experiment of morphological processing method as shown in Fig. 2, the images obtained after different morphological operations for light and dark tumors are shown. We can see that for CT images containing light color tumors, dilation operation can magnify the lesion area and facilitate the network to detect small lesions. For the CT images containing dark tumors, we can see that the tumor portion of the erosion manipulated images not only becomes larger, but also retains a lot of texture information. The formulas of dilation and erosion operations are shown as follows:

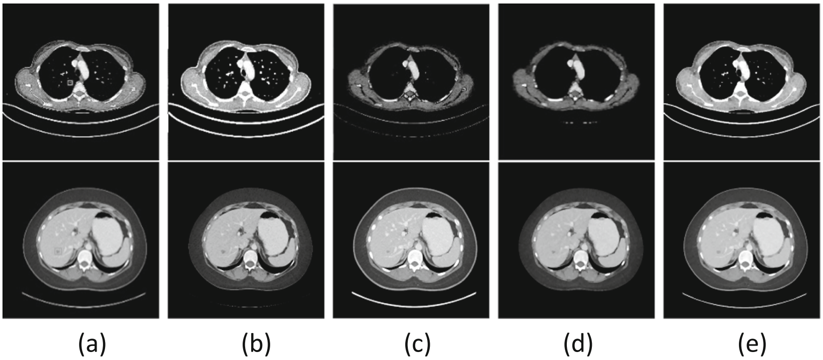

Fig. 2. The output of different colors tumors after morphological operations. Organs with representative light and dark colors of lesions in the dataset are shown. Where, (a)-(e) represent the labeled image, erosion, dilation, opening and closing respectively. First row is lung CT image and second row is liver CT image.

$$A \oplus B = \{x, y | (B)_{xy} \cap A \neq \varnothing\} \tag{1}$$

$$A \ominus B = \{x, y | (B)_{xy} \subseteq A\} \tag{2}$$

Where, $\oplus$ and $\ominus$ represent dilation and erosion respectively.

2.2 Feature Extraction Network

Faster-RCNN [9] is an object detection algorithm proposed by Kaiming He in 2015. Based on Fast-RCNN [10], this algorithm puts forward the RPN region proposal box

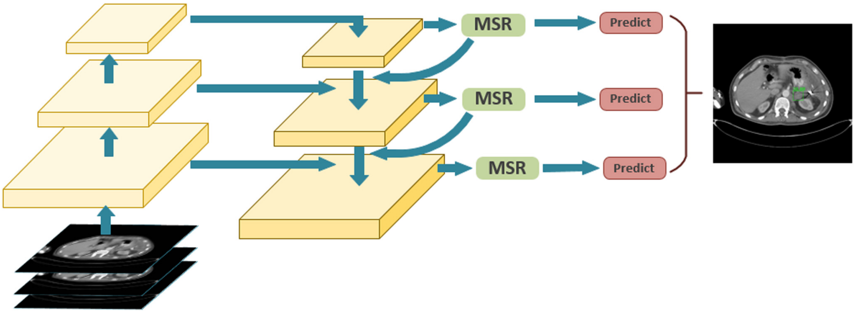

Fig. 3. The structure of FPN built based on the proposed MSR module

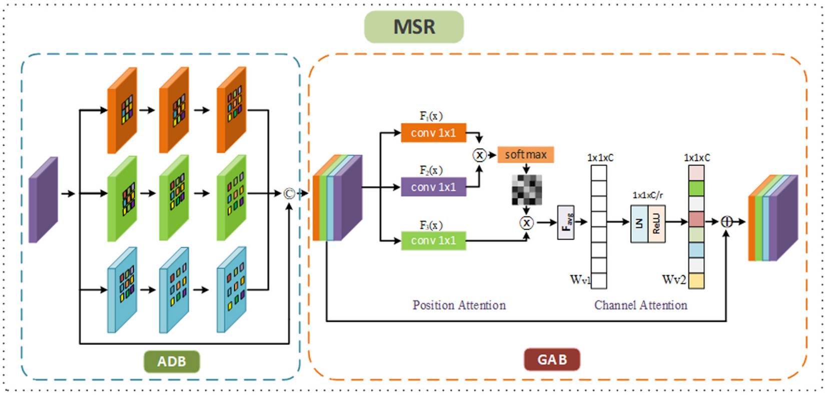

Fig. 4. The structure of MSR module. The detailed architecture of the Multi-Scale Response (MSR) module consists of two parts: Aggregated dilated block (ADB) in the blue dashed box on the left and Global attention Block (GAB) in the orange dashed box on the right. (Color figur online)

generation algorithm, which greatly improves object detection speed. The detection part is divided into four steps: The first step is to input the whole image into CNN for feature extraction. The second step is to generate the anchor box by RPN. In the third step, the RoI pooling layer makes each RoI generate a feature map with a fixed size. In the fourth step: Softmax Loss and Smooth L1 Loss are used for classification and Bounding box regression respectively.

We first modified the backbone of the model, the original backbone of Faster RCNN is VGG16 and we replaced it with ResNet101 [11]. The ResNet results on the ImageNet dataset [12] show that the performance of the residual structure is significantly better than the traditional convolution framework. Due to the down-sampling effect in traditional convolutional neural networks, small objects cannot acquire obvious features.

To address this issue. First of all, we deepened the backbone to continue to enhance the network's ability to extract deeper semantic information. conv2_ ×, conv3_ ×, conv4_ ×, conv5_ ×, conv6_ ×, and conv7_ × blocks are used to build the feature pyramid. P2, P3, P4, P5, P6 and P7 correspond to conv2_ × –conv7_ ×. We convolved the corresponding bottom-up feature maps by 1 × 1 kernels to reduce the number of channels and fuse the deep feature [13]. After processing by the MSR module, the corresponding bottom-up feature map is added to the corresponding bottom-up feature map by up-sampling. The feature map of P2–P4 resolution in Backbone can help the network find and locate small lesions. With the deepening of the network, the small objects information will dismiss because of the convolution and pooling operations. There are two purposes for adding conv6_ × and conv7_ ×. The first point is to improve the feature extraction capability of the backbone. The second point is to fuse with the upper layer feature map after up-sampling when constructing FPN, see Fig. 3, which can bring more deep semantic information into the higher resolution feature map in the upper layer.

2.3 Aggregated Dilation Block

In the process of generating feature pyramids based on ResNet residual blocks, the imbalanced problem between spatial and semantic information will emerge. To address this issue, we build a feature pyramid network constructed by multiple scale output of Res-block in the top-down pathway. We introduced dilated convolution in the ADB module by using multi-branch structure to adapt to the receptive field of feature maps with multi-scales through different dilation rates. In each parallel dilated convolution branch, the feature map is enhanced by the cascade convolution kernels with different dilation rates. After the convolution of each layer, the output is nonlinearized by the activation function to prevent gradient explosion and can bring more differential representations for feature transformation. Weighted combinations in the multi-branch dilation convolution process can eliminate to some extent the noise left behind in low-resolution images. Then we concatenate the output features of each branch with the original image to become an aggregated feature map. The feature map output by the ADB module has a larger receptive field.

In ADB module, $f \in R^{W \times D}$ is used to describe the architecture of ADB, where W and D are the width and depth of ADB respectively. The dilation rate of specific layer in ADB is expressed as f_{ij}, where $\imath = 1, 2, \ldots, W$ and $\jmath = 1, 2, \ldots, D$ represent the index of width and depth respectively. The aggregated dilated operation is shown as follows:

$$\mathcal{F}(x) = \sum\nolimits_{\imath=1}^{W} \mathcal{T}_{\imath}(x|f_{i1}, f_{i2}, \ldots, f_{iD}) \tag{3}$$

where $T_i(x)$ represents the cascade-transformation.

As shown in Fig. 4, the parallel structure branch inside the ADB module is connected in series with convolution kernels with different dilation rates, and the output multi-scale feature map restores more detailed spatial information of the instance. It also provides more long-range context information for the construction of the feature pyramid. The receptive field of each layer is expressed as follows:

$$\mathcal{A}_{i0} = 1 \tag{4}$$

$$k_{ij} = k_{ij} + \left(k_{ij} - 1\right) \times \left(f_{ij} - 1\right) \tag{5}$$

$$\mathcal{A}_{ij} = r_{i,j-1} \times k_{ij} - \left(k_{ij} - 1\right) \times (\mathcal{A}_{i,j} - \prod_{k=1}^{j-1} s_k) \tag{6}$$

Where, $\mathcal{A}_{ij}$ denotes the receptive field, k_{ij} denotes the kernel size and s_k represents the stride.

From the formula, we can see that the size of the receptive field extracted by the convolution kernel with different dilation rates is also different. Usually in the feature extraction network, that is, in the backbone, the dilation convolution helps us identify the large object from the enlarged receptive field [14]. However, we added dilation convolution follows the output of the feature pyramid, expecting to provide more context spatial information to improve the detection accuracy of small lesions.

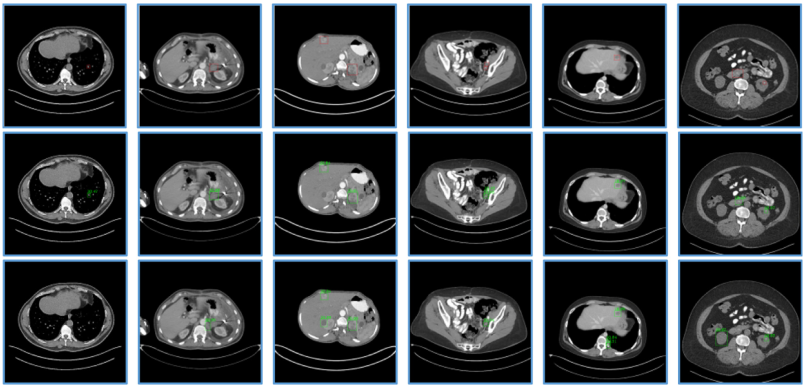

Fig. 5. Lesion detection results for sample CT images of various methods. Each row from top to bottom represents the label image, our proposed method, Faster R-CNN respectively

2.4 Global Attention Block

Inspired by the current popular attention mechanism, we propose a Global Attention Block (GAB). When the network's backbone extracts information from a large number of feature maps, the GAB allows the network to pay more attention to some vital information and improving the accuracy of detection. Each attention block consists of two parts: a spatial attention block and a channel attention block.

Object detection needs to be extremely sensitive to changes in spatial location, so we proposed spatial attention block which uses a self-attentive mechanism to model remote dependencies in order to enhance the network's global understanding of the visual scene. In addition, inspired by SENet [15], we proposed a channel attention block that aims to focus on the feature information we need.

We converted the input feature map x into three paths F_1, F_2 and F_3, where $F_1(x) = W_1x$ and $F_2(x) = W_2x$. Firstly, obtaining the attention map of the long-range correlation between each position in the feature map throughS_{ij}, where$S_{ij} = F_1(x_i)^T \otimes F_2(x_j)$.$S_{ij}$ is transformed into $A_{i,j}$ by *softmax*, where, $A_{ij} = softmax(s_{ij})$ represents the relationship between the position of i and the position of j in the feature map, and then A_{ij} and F_3 are multiplied to query the response relationship between pixels on the feature map.

$$z_i = \sum_{j=1}^{N} A_{i,j} \otimes F3\left(x_j\right), F_3(x) = W_3 x \tag{7}$$

Where i represent the index of query position. $N = H \times W$ and N represents number of feature locations. $\otimes$ denotes matrix multiplication.

After the spatial attention block, we compress the global information into channels through global average pooling, the main difference between the SE block and the block we proposed is the fusion module, which reflects the different goals of the two blocks. The SE module uses re-adjustment to re-calibrate the importance of the channel, but it does not fully simulate the long-range correlation. We captured the long-range correlation by using addition to aggregate the global context to all positions. The detailed architecture of the global attention block (GAB) is formulated as follows:

$$y_i = x_i + W_{v2} ReLU(LN(W_{v1} z)) \tag{8}$$

Where $W_{v1} \in R^{\frac{C}{r} \times C}$, $W_{v2} \in R^{C \times \frac{C}{r}}$. In order to obtain the lightweight attribute of the channel attention block, we reduced the parameters of the module from C to C/r. Where r is the bottleneck ratio, setting r too large will lose feature information and vice versa will consume a lot of computation, so we need to strike a balance between two costs, we found that when r = 4, the model performs best.

3 Experiments and Results

3.1 Datasets

The DeepLesion is the largest open dataset of multi-category, lesion-level labeled clinical medical CT images ever published by an NIH Clinical Center. By training deep neural networks on this dataset, it will be possible to obtain a large-scale universal lesion detector that can more accurately and automatically measure the size of all lesions in the patient's body, allowing initial assessment of cancer system-wide. The dataset contains 32,735 labeled lesion instances from 4,427 independent, anonymous patients. The dataset covers a wide range of lesions involving the lung, liver, mediastinum (mainly lymph nodes), kidney, pelvis, bone, abdomen and other soft tissues. We used 70% samples of the dataset for training, 15% for validation, and 15% for testing.

3.2 Training Schedule

We set training learning rate to 0.008 and training momentum to 0.9; after 10,000 iterations, the training weights decayed to 0.001; the training batch is 128, the mini_batch is 2; and the learning process is 12 epochs. The initialization weights for P1–P5 are from the ImageNet pre-trained model, and for the deepened network part.

Table 1. mAP and AP of each lesion type on the official split set of DeepLesion.

Methods	Total	BN	AB	ME	LV	LU	KD	ST	PV
Baseline	0.484	0.524	0.512	0.549	0.549	0.582	0.419	0.436	0.368
Baseline + (Res101)	0.512	0.526	0.515	0.553	0.557	0.591	0.436	0.447	0.414
Baseline + + ADB	0.545	0.527	0.521	0.557	0.569	0.604	0.487	0.452	0.465
Baseline + + MSR (ours)	**0.584**	**0.531**	**0.534**	**0.563**	**0.573**	**0.659**	**0.532**	**0.471**	**0.533**

Table 2. An ablation study with various configurations of the proposed modules. Lesion detection sensitivity is reported at different false positive (FP) rates on the DeepLesion test set.

Method	FPs per image				
	0.5	1	2	4	8
Baseline	0.560	0.677	0.763	0.832	0.867
Baseline + (Res101)	0.592	0.693	0.788	0.843	0.871
Baseline + + ADB	0.615	0.719	0.804	0.851	0.887
Baseline + + MSR (ours)	**0.658**	**0.752**	**0.827**	**0.881**	**0.891**

We initialize the parameters randomly. We resized the input image to 512 × 512 size. The optimization algorithm we used is stochastic gradient descent (SGD). The training time of our detector is about 60 h, compared to one-stage detector we have drawbacks in training cost and test speed, but speed is not very important in medical image detection scenario, compared to the need to accurately detect the focal area.

3.3 Hardware and Software Setup

Experiments were conducted on a Workstation with IntelCore i7, 2.7 GHz CPU, 8 GB RAM under Ubuntu 18.4, and a NVIDIA GTX 2080 video processing card with 11 GB memory. Faster RCNN was deployed in pytorch 1.5.1 framework and based on python3.7, cuda 10.1 and cudnn 7.6.3.

3.4 Evaluation

We quoted two evaluation metrics in our subsequent ablation experiments and comparisons. In our experiments, the object detection accuracy was measured by mean Average Precision (mAP) when IoUthres = 0.5. Another evaluation metric is the average sensitivity values at different false positives rates (FROC) of the whole test set.

3.5 Results

Ablation Study

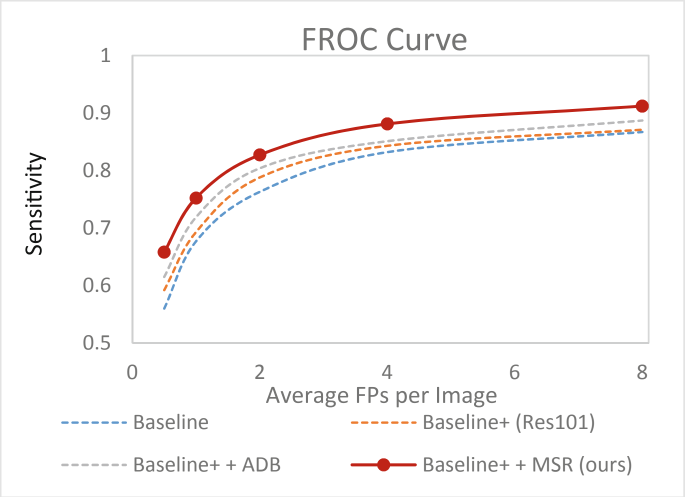

Fig. 6. Ablation FROC Curve on test set of official split test set of DeepLesion.

Table 3. mAP and AP of each lesion type of various methods on the official split test set of DeepLesion.

Methods	Total	BN	AB	ME	LV	LU	KD	ST	PV
YOLOv3 [16]	0.468	0.371	0.372	0.587	0.524	0.562	0.373	0.403	0.435
RetinaNet [17]	0.510	0.539	0.430	0.555	0.524	0.612	0.424	0.455	0.421
3DCE, 3 slices [5]	0.506	0.434	0.424	0.522	0.543	0.633	0.426	0.421	0.423
3DCE, 9 slices [5]	0.544	0.492	0.468	**0.577**	0.564	**0.663**	0.480	0.441	0.470
Ours	**0.584**	**0.531**	**0.534**	0.563	**0.573**	0.659	**0.532**	**0.471**	**0.533**

The proposed network consists of three main components: Faster R-CNN, ADB and GAB. In Table 1 and Table 2, Baseline denotes the original Faster R-CNN model, and Baseline + (Res101) denotes the method for deepening the backbone mentioned in Sect. 2.2. To assess the validity of each module, we performed ablation studies on the DeepLesion dataset. From Table 1, we can see that the original model was improved by our proposed method and the mAP values were improved to 10%. Based on the coarse lesion types provided by DeepLesion for each CT slice, we calculated the AP for each lesion type. Besides, the table shows that the AP values were increased by different magnitudes for different sites. The metrics assessed in Table 2 are the average sensitivity

Table 4. Comparison of the proposed method with state-of-the-art methods on the DeepLesion test set. Lesion detection sensitivity values are reported at different false positive (FP) rates.

Method	FPs per image				
	0.5	1	2	4	8
RetinaNet [17]	0.458	0.542	0.625	0.698	0.742
YOLOv3 [16]	0.520	0.626	0.719	0.795	0.843
Mask R-CNN [18]	0.398	0.527	0.656	0.777	0.852
ULDor [4]	0.529	0.648	0.748	0.844	0.861
3DCE, 3, slices [5]	0.569	0.673	0.756	0.816	0.858
3DCE, 9, slices [5]	0.593	0.707	0.791	0.843	0.878
3DCE, 27, slices [5]	0.625	0.733	0.807	0.857	**0.891**
Ours	**0.658**	**0.752**	**0.827**	**0.881**	**0.891**

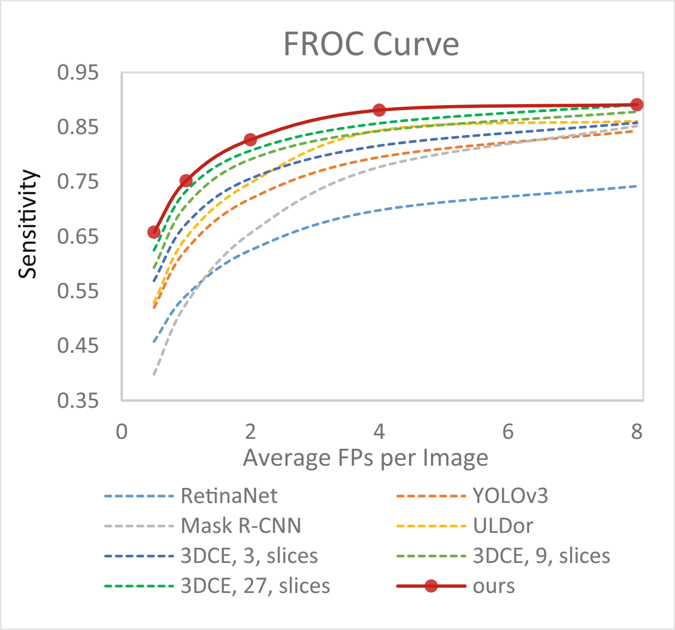

Fig. 7. FROC Curves of various methods.

values for the entire test set at different false positives rates. The comparison between different configurations shows that the proposed method achieves the highest sensitivity at different false positives rates. We plotted the FROC curves to make the results more intuitive, see Fig. 6. The detection results are also shown in Fig. 5.

Comparisons with State-of-the-Art. We compared our model with state-of-the-art methods. As can be seen from Table 2 and Table 4 we have chosen YOLOv3 [16], Reti-naNet [17], Mask R-CNN [18] and 3DCE [5] Yan et al. selected different numbers of slices and sent them to the detector. The detector generated 3D context information from different numbers of slices to help the detector make the final lesion prediction.From Table 3, we can see that the mAP value of our proposed detector is higher than other advanced detectors. Although it is slightly lower than 3DCE on the medi-astinum and lung organs, higher than other detectors on all other organs.

Table 4 shows the results of the evaluation, which indicates that our method is superior to the existing methods. We can see that the higher the number of CT slices that the detector selects as input for 3DCE, the higher sensitivity the detector has, due to the fact that the number of CT slices can provide more contextual information to the detector. However, in this paper, we obtained better results by selecting only a single CT slice as input. To make the comparison more straightforward, we plotted the FROC curves comparing the experimental results of several methods in Fig. 7.

In terms of the overall comparison results with state-of-the-art methods, our proposed method is superior in detection accuracy and sensitivity.

4 Conclusion

We propose a fine-grained lesion detection method with a novel multi-scale global attention mechanism to enhance the detection of lesions on feature maps of different sizes. In different scale convolution levels of the detection network, we augment the detector's awareness of feature map scale variation by ADB. ADB provides finer size estimates of the feature map to capture the response to scale under different receptive fields. To effectively choose meaningful responses, we propose the GAB attention module, where the results of ablation experiments on the DeepLesion dataset demonstrate the effectiveness of our proposed method for detection at different scales.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (61976106, 61572239), Zhenjiang Key Deprogram "Fire Early Warning Technology Based on Multimodal Data Analysis" (SH2020011) Jiangsu Emergency Management Science and Technology Project "Research on Very Early Warning of Fire Based on Multi-modal Data Analysis and Multi-Intelligent Body Technology" (YJGL-TG-2020–8).

References

1. Contact Us|National Cancer Center Japan. https://www.ncc.go.jp/en/contact/index.html. Accessed 16 Sep. 16 2020
2. Lee, S.G., Bae, J.S., Kim, H., Kim, J.H., Yoon, S.: Liver lesion detection from weakly-labeled multi-phase CT volumes with a grouped single shot multibox detector. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 693–701. Springer, Cham, September 2018. https://doi.org/10.1007/978-3-030-00934-2_77

3. Yan, K., Wang, X., Lu, L., Summers, R.M.: Deeplesion: automated deep mining, categorization and detection of significant radiology image findings using large-scale clinical lesion annotations (2017). arXiv preprint arXiv:1710.01766
4. Tang, Y.B., Yan, K., Tang, Y.X., Liu, J., Xiao, J., Summers, R.M.: ULDor: a universal lesion detector for CT scans with pseudo masks and hard negative example mining. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 833–836. IEEE, April 2019
5. Yan, K., Bagheri, M., Summers, R.M.: 3D context enhanced region-based convolutional neural network for end-to-end lesion detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 511–519. Springer, Cham, September 2018. https://doi.org/10.1007/978-3-030-00928-1_58
6. Chiao, J.Y., Chen, K.Y., Liao, K.Y.K., Hsieh, P.H., Zhang, G., Huang, T.C.: Detection and classification the breast tumors using mask R-CNN on sonograms. Medicine **98**(19) (2019)
7. Ding, J., Li, A., Hu, Z., Wang, L.: Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 559–567. Springer, Cham, September 2017. https://doi.org/10.1007/978-3-319-66179-7_64
8. Dou, Q., Chen, H., Yu, L., Qin, J., Heng, P.A.: Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. IEEE Trans. Biomed. Eng. **64**(7), 1558–1567 (2016)
9. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
10. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
12. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE, June 2009
13. Zhang, X., Lu, W., Li, F., et al.: Deep feature fusion model for sentence semantic matching. Comput. Mater. Continua **61**, 601–616 (2019)
14. Li, Z., Peng, C., Yu, G., Zhang, X., Deng, Y., Sun, J.: Detnet: a backbone network for object detection (2018). arXiv preprint arXiv:1804.06215.
15. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of THE IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
16. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement (2018). arXiv preprint arXiv:1804.02767.
17. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
18. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)