Pneumonia image detection based on convolutional neural network

Yijin Zhao

Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan, 610065, China

Zyj92730@163.com

Abstract. In recent years, with the continuous increase of various types of data sets and the acceleration of computer hardware, the field of artificial intelligence has achieved unprecedented development. As a core driving technology of artificial intelligence, deep learning has a significant impact on human life. With the emergence of the novel coronavirus and its rapid development into a global new infectious disease, the intelligent medical treatment of pneumonia image detection has become very important. However, due to the particularity and limitation of medical images, medical image segmentation is more challenging than other computer vision tasks. Therefore, it is urgent to carry out research on medical image segmentation. This paper takes convolutional neural network as the research direction, reasonably designs and improves the model, and improves the detection and segmentation performance of pneumonia images. The main research contents are as follows: Aiming at the problems of cumbersome steps and low accuracy in traditional image detection methods and poor detection effect and efficiency in existing deep learning methods, a detection network model based on Mask R-CNN is proposed. The experimental results show that: first, the random weighting algorithm can improve the detection accuracy and training speed of pneumonia; Second, by using ResNet with different layers, it is proved that reducing the number of layers can effectively improve the detection accuracy and training speed of pneumonia.

Keywords: COVID-19, medical image segmentation, deep learning, convolutional neural networks.

1. Introduction

1.1. Research background and significance

In recent years, deep learning has far outperformed traditional methods and machine learning methods in many fields. In the field of medical image processing, medical image detection and segmentation, as the most important means of auxiliary diagnosis and treatment, is related to human health. Therefore, it is urgent to further study the medical image segmentation. Medical CT image plays a key role in the diagnosis of various types of diseases, among which the status of lung CT image is very important for human health. For human lungs, lung infections caused by viruses and bacteria are the most common, with serious cases such as atypical pneumonia in 2003 and novel coronavirus in 2020. CT images of

the lung are necessary to distinguish viral infection from normal influenza, but CT tomography is a three-dimensional imaging with very high resolution and a large amount of data. There are basically hundreds of CT images per patient, which makes the diagnosis inefficient and inaccurate, and also requires a lot of labor time. Therefore, the use of deep learning methods to study the disease diagnosis of lung CT images is an unstoppable trend in the future intelligent medical application.

1.2. Research Status at home and abroad
Medical CT images are more special than natural images. First of all, it is a gray scale map with similar features in each region, which is not easy to be learned by computer. Secondly, it is difficult to obtain and needs to be labeled by professionals. Therefore, there are few existing CT data sets, and the number of images in each kind of data set is also very small, which has certain obstacles to the use of deep learning methods. Therefore, most detection and segmentation models are designed based on natural image datasets, while a few models are designed for medical images.

1.3. Research Status at home and abroad
Although the traditional manual feature extraction method can achieve the purpose of detection and segmentation, but the accuracy is very low and the steps are tedious. Although the deep learning method solves the above problems to a certain extent, there is still room for improvement. For example, the model in literature [1] is complex and requires a lot of time for training. The segmentation accuracy in literature [2] is not high; Literature [3] does not make better use of the underlying characteristics. Aiming at these problems, this paper designs an end-to-end medical image segmentation network to improve the accuracy and reduce the operation time. The main research contents and contributions are as follows: Firstly, this paper analyzes the inaccurate detection problem of MaskRcnn on pneumonia image dataset, and finds that its network structure is unreasonable for medical image detection, and proposes to reduce the number of network layers for feature extraction. Secondly, in view of the problem that MaskRcnn training speed is too slow and the accuracy rate is not good, the idea of literature [4] is combined. In this paper, a random weighted ResNet feature extraction network is proposed to improve the computational efficiency by randomly assigning lightweight weights to the input and jump connections of the network. Finally, the experimental results show that the proposed method realizes the detection of pneumonia images, and compared with the original network, ResNet with random weighting has better improvement in accuracy and speed.

2. Pneumonia image detection based on Mask R-CNN

2.1. Mask R-CNN network structure

2.1.1. Frame detail description. Mask R-CNN improves on the target detection framework Faster Rcnn, and proposes to use Region of interest (Roi) Align instead of ROI-pooling in Faster Rcnn. Since ROI-pooling requires two quantization of the feature map, the rounding operation during quantization will lead to the loss of pixels. However, Roi Align uses bilinear interpolation method to Align pixels, which cancels the quantization operation and makes the segmentation result more accurate. Secondly, an additional branch of full convolutional network is added on the basis of Faster Rcnn to generate segmentation mask. Other parts are basically consistent with Faster Rcnn. The structure of Mask R-CNN network is shown in Figure 1. First, natural images were input, features were extracted through ResNET-101 layer, and RPN and FPN were used to generate proposals, N rectangles were generated for each image. Secondly, they were mapped to the feature map at the last convolutional layer of CNN, and the Roi Align layer was used to generate a fixed-size feature map and complement the pixels. The last branch generates a mask through FCN, and the other branch is divided into two branches again after the average pooling through the full connection layer to complete the task of localization and segmentation.
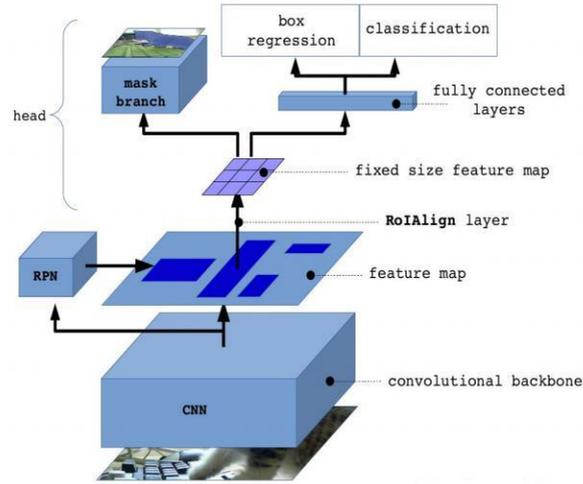
Figure 1. Mask R-CNN network structure.

2.1.2. Evaluation criterion. IOU (Intersection-Over-Precison) and mAP (Mean Average Precision) are two major evaluation indexes in the field of target detection. The target detection results are generally represented by rectangular boxes, so in the evaluation, the overlap degree (IOU) between the network detected rectangular box of the object and the real rectangular box should be considered first, and the average accuracy (mAP) should be considered for the judgment of the results.

2.1.2.1 IOU. The calculation of IOU is called the calculation of intersection ratio, as shown in Figure 2. Assume that yellow area A is the location of the real rectangular box marked manually, and blue area B is the location of the object detected by the neural network. It can be seen that there is A red intersection between the detected area and the real area, so the calculation expression of IOU is as follows:

$$IOU = \frac{S_{A \cap B}}{S_{A \cup B}} \tag{1}$$
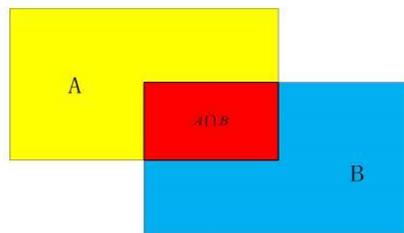
Figure 2. Intersection Over Union of IOU.

2.1.2.2. MAP. MAP is the final index of recognition accuracy, which is mainly calculated by recall and precision. The recall rate and accuracy can be expressed as follows:

$$recall = \frac{TP}{TP + FN} \tag{2}$$

$$precision = \frac{TP}{TP + FP} \tag{3}$$

Among them, True Positives means actually positive sample, and network test result is also positive sample; False Positives means actually positive sample, and network test result is negative sample;

False Positives means actually negative sample, and network test result is positive sample as well. In other words, recall represents the proportion of the correctly identified part in the whole detection result in the whole data set, and precision represents the proportion of the correctly identified part in the detection result in the prediction result of the whole data set.

2.2. The improvement of Mask R-CNN network structure

Firstly, the data is enhanced in the traditional way. Secondly, the improvement on the feature extraction network ResNet consists of two parts: The first is to use ResNet with 18 layers for feature extraction and adjust the corresponding parameters, eliminating the original 101 layers. By comparing the accuracy and speed with ResNet with other different layers, the feasibility of the proposed method for pneumonia CT image detection is verified. Secondly, a random weighting algorithm is proposed, which can help ResNet to eliminate excessive redundant information. By adding the algorithm to ResNet at different levels, it is verified that the random weighting algorithm can improve the accuracy and speed.

2.2.1. Data enhancement. As mentioned in (1), the original network is designed based on large natural image datasets. When it is applied to small data sets, especially medical images, the primary consideration is data enhancement to improve the generalization ability of the model. In this chapter, image mirroring is used for data enhancement, including left and right horizontal mirroring, up and down vertical mirroring and left and right diagonal mirroring. In this way, the original image is increased by 8 times.

2.2.2. Network depth selection. ResNet along with the network was proposed to solve for deep performance degradation problem, but in pneumonia medical image data sets to detect the effect not beautiful, even if the ResNet have characteristics of strong learning ability, but because the original network structures, the 101 layer depth, deep network lead to a few medical image detection model performance and generalization ability. Therefore, the number of original network layers should be reduced to improve the information utilization of all layers. According to literature [5], the FLOPs of 101-layer ResNet is 7.6x109, and the computation of 18th-layer ResNet is only 1.8x109, which is about a quarter of the 101-layer ResNet. Compared with the 101-layer ResNet, the whole network is lighter. The lightweight network is suitable for training a small number of medical images to solve the problem of overfitting. It also reduces the computation amount and saves the training time. Specific parameter Settings of each layer are shown in Table 1. Resnet-18 convolution layer is divided into five stages, and Stage1 to Stage5 contains five convolution kernels with sizes of 7x7x64, 3x3x64, 3x3x128, 3x3x256 and 3x3x512 respectively. The output data from Stage5 are processed through two paths, one of which is classified and border regression through the fully connected layer after average pooling, and the other is output segmentation mask through the fully convolutional kernel.

Table 1. Resnet-18 Layer parameters and model Settings.

Layer Name	Filter Size	max pooling	stride	Output Size
Stage 1	7x7x64	3x3	2	112x112
Stage 2	(3x3x64)x4	/	1	56x56
Stage 3	(3x3x128)x4	/	1	28x28
Stage 4	(3x3x256)x4	/	1	14x14
Stage 5	(3x3x512)x4	/	1	7x7

2.2.3. ResNet random weighted. The idea of stochastic weighting is inspired by batch gradient descent (BGD), stochastic gradient descent (SGD) and Mini-batch gradient descent (MBGD). BGD refers to the updating of weights with all samples at each update, which is fast for simple network models. However, the convergence rate of BGD is very slow for complex models or large models. At present, the tasks processed by deep learning are becoming more and more diverse, and the network model is becoming more and more complex. BGD has been unable to meet the demand for computing speed.

SGD is to use only one sample to approximate all the samples in each update to adjust the weight value. The SGD method converges faster, and although it is not the global optimal solution, its speed can be accepted for some tasks. MBGD approximates all samples with a number of B samples per update, and the value of B can be set according to the network model, the number of samples, and the specific task. In practical tasks, this method is widely used, because the size of B can be adjusted according to the computer hardware configuration, model size and specific task, so MBGD is more flexible and generally accepted. Even though this method does not converge very fast, it can get more local optima. Inspired by the above ideas, in order to solve the problem of Mask R-CNN overfitting and slow training speed in pneumonia examination, a random weighting idea on ResNet was proposed. Figures 3 (a) and 3 (b) show the network details of ResNet and the weighting approach proposed in this chapter, respectively. The green block represents the 3x3 convolution kernel, the red block represents BN layer, and the blue block represents ReLU. The weighting method is as follows:

$$y = F(\omega_1 x) + \omega_2 x \tag{4}$$

$$\omega_1 + \omega_2 = 0.5 \tag{5}$$
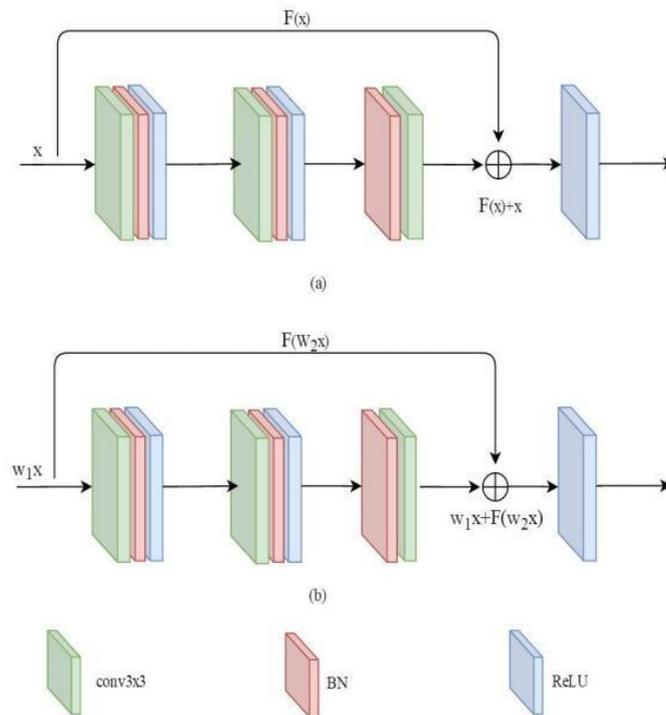
Figure 3. Resnet-18 random weighting method is proposed in this chapter.

Where, ω_1 and ω_2 represent the weighting factor, and ω_1 and ω_2 are in a range of (0,1). In the training process, it is randomly extracted from the range of values to eliminate the redundant information in the network. The lightweight weight can better reduce the redundant parameters and improve the speed of operation. By setting a fixed range, two weighting factors are extracted from the range and involved in network training, which embodies the idea of randomness. In the task of pneumonia image detection based on Mask r-cnn, the range is set in (0,0.5] according to the experience obtained from the actual experiment process, and the constraint conditions are added as shown in formula (5). The purpose of adding constraints and setting the final range is to reduce the error and prevent the results from being local optima in operation.

2.3. Experiment and result analysis

2.3.1. Introduction to Data Set. The dataset consisted of 3000 lung CT images with a resolution of 1920x1080 provided by the hospital. Firstly, the original image and corresponding labels are respectively increased to 24000 by data enhancement. Secondly, they are divided into three data sets, in which the training set accounts for 80%, and the training set and the test set account for 10% respectively.

2.3.2. ResNet and its layers are compared and analyzed. In the experiment, mAP comparison between different ResNet layers and random weighting algorithm (Rw) was carried out on the pneumonia image dataset. The experimental results are shown in Table 2. Firstly, ResNet with different layers in Model 1 in the table is compared. It can be found that with the decrease of network layers, the detection accuracy is gradually improved, from 0.7637 of RESNET-101 to 0.8410 of Resnet-18, and the mAP increases by about 9.1 percentage points. It is proved that shallow network has better detection effect on a few medical image datasets. Secondly, the random weighting algorithm is combined into ResNet for comparison. It can be seen from Model 2 in the table that ResNet+Rw of different layers is improved compared with ResNet of different layers after using the weighting algorithm. Among them, RESNET-18 +Rw is increased from 0.8410 to 0.8515 compared with RESnet-18, which increases by about 1 percentage point. It is proved that reducing the network redundant information is helpful to improve the detection result.

Table 2. ResNet and ResNet random weighting in bladder tumor data.

Model 1	mAP	Model 2	mAP
ResNet-101	0.7234	ResNet-101+Rw	0.7485
ResNet-50	0.8149	ResNet-50+Rw	0.8227
ResNet-18	0.8462	ResNet-18+Rw	0.8593

2.3.3. Comparative analysis of training speed. As the number of network layers is reduced to simplify the model, the training time of the original ResNet decreases greatly with the decrease of network layers. As can be seen from the mAP comparison in the previous section, excessive redundant information in the network will have a negative impact on the performance and operation speed of the model. Therefore, simplifying the feature extraction network can improve the detection and segmentation ability of the model for pneumonia images.

3. Conclusion

In this paper, three improvement measures are proposed to solve the problem that Mask R-CNN based on large natural image detection model is not effective in pneumonia image detection and the operation speed is too slow. Firstly, data enhancement is carried out to solve the problem of too few images in pneumonia dataset. Second, the number of layers of the feature extraction network ResNet in the original Mask R-CNN is reduced to improve the detection effect and operation speed. Thirdly, a random weighting algorithm is proposed to further improve the detection effect and operation speed. Finally, the experiment proves the effectiveness of the improved method in terms of target detection evaluation index, subjective visual effect and training speed.

References
[1]   HE, Kaiming, et al. Mask r-cnn[A]. In: Proceedings of the IEEE international conference on computer vision[C]. 2017:961-2969.
[2]   RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation[A].In: International Conference on Medical image computing and computer-assisted intervention[C]. Springer, Cham, 2015:234-241.

[3]   Gu, Zaiwang, et al. CE-Net: context encoder network for 2D medical image segmentation[J]. IEEE transactions on medical imaging, 2019, 38(10): 2281-2292.

[4]   He, Kaiming, et al. Deep residual learning for image recognition[A]. In: Proceedings of the IEEE conference on computer vision and pattern recognition[C]. 2016:770-778.

[5]   Rüegg, Nadine, et al. Chained Representation Cycling: Learning to Estimate 3D Human Pose and Shape by Cycling Between Representations[J]. arXiv preprint, 2020:2001.01613.