

Pulmonary inflammation region detection algorithms based on deep learning: a review

Tianqi Fang^{1, †}, Xuanyu He^{2, *, †}, Lizhe Xu^{3, †}

¹ School of Information Science and Engineering, Harbin Institute of Technology, Weihai 264200, China

² School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100000, China

³ School of Artificial Intelligence and Data Sciences, Hebei University of Technology, Tianjin 300000, China

[†]These authors contributed equally.

* Corresponding Author Email: 19721005@bjtu.edu.cn

Abstract. With the popularity and development of object detection in deep learning, it is used in more and more industries, including the field of medical science. This paper summarizes the target detection algorithms for the pneumonia region. Firstly, this paper briefly introduced the existing detection methods of target detection and summarized the main pneumonia datasets and image preprocessing methods. Then we focused on the framework composition and detection effect of the main model. Finally, through experimental analysis, we proposed to apply some new models to the detection of pneumonia and used some improvements and techniques to improve the detection effect.

Keywords: Target detection, Pneumonia detection, Deep learning.

1. Introduction

Pneumonia has become an increasingly serious threat in recent years and therefore accurate diagnosis of pneumonia is one of the most important areas of modern medical science. The use of X-rays to diagnose disease is of great importance in the diagnosis of lung disease. Pneumonia usually appears as increased opacity in the lungs on X-ray images and there are some other problems, such as pulmonary edema and pleural effusion appearing as opaque areas on X-ray images. All these problems interfere with the diagnosis of pneumonia, so the correct interpretation of information is always a major challenge for doctors. It requires an experienced radiologist to make a diagnostic confirmation. It would lack human resources and be difficult to guarantee accuracy if we only rely on medical practitioners to determine whether a patient has a confirmed diagnosis based on an empirical judgment of the images.

The diagnosis of pneumonia by the way of science and technology is an inevitable development in the modern medical field. With the rapid development of deep learning, using deep learning algorithms to diagnose medical images has great progress. For the detection of pneumonia, the use of deep learning algorithms in the target detection algorithm can greatly improve the efficiency and accuracy of diagnosing pneumonia.

Currently, target detection algorithms can be divided into two main categories: One-Stage and Two-Stage target detection algorithms.

Two-Stage target detection algorithms generally use algorithms (selective search or region suggestion networks, etc.) to extract candidate frames from images, and then perform secondary correction on the candidate frames to obtain detection results. The algorithms include RCNN, SPP-Net, Faster-RCNN, Faster RCNN, feature pyramid networks (FPN), and Mask-RCNN. They have a complex detection process, and although the algorithms have high accuracy, the detection speed is slow and cannot meet the needs of industrial real-time applications.

The One-Stage target detection algorithms differ from the Two-Stage target detection algorithms in that the former does not have a candidate region recommendation stage, and the training process is relatively simple, allowing the target category to be determined and the location detection frame to be obtained in one stage. Nowadays, new algorithms such as NAS-FPN, EfficientDet, and YOLOF have emerged. The YOLO series of One-Stage target detection is popular in the industry because it improves the accuracy of the algorithm. By improving the underlying network, adding FPN structures, using more powerful data enhancement strategies, and adding new loss functions we can realize the purpose of improving the accuracy while maintaining the speed advantage of the algorithms.

There are several relatively large publicly available lung image datasets for scholars to study. NIH Chest X-rays^[1]: This public dataset is provided by the National Institutes of Health (NIH). There are 112120 X-ray lung images and 14 kinds of chest diseases, including pneumonia. The method of Natural Language Processing (NLP) is used to label the pictures on the dataset and the accuracy rate is more than 90%. Chest X-ray Images^[2]: This dataset is from the University of California. It collected 5863 chest radiographs of the lungs of children aged from 1 to 5 years from Guangzhou Women's and children's Medical Center in 2018, which are divided into two categories: normal and pneumonia. Using the method of transfer learning makes the accuracy up to 92.80% accuracy. The Pneumonia Recognition Competition public dataset^[3]: This dataset is provided in the medical image pneumonia recognition competition sponsored by RSNA and Kaggle company in September 2018. Each set of chest radiographs also provides the boundary of the patient's lung lesions box value and target value. In addition, the dataset is divided into pneumonia and non-pneumonia.

2. Current pre-processing methods

The initial dataset image may have some defects, such as inconsistent image size, a small number of samples, and low brightness. Therefore, there is a pretreatment link before training to improve these problems.

Rahman et al. adjusted the size of the image according to the type of network model. After that, they used data enhancement on the dataset by rotating the image 315 degrees clockwise and enlarging it by 10%. Finally, the image was translated by 10% in the horizontal and vertical directions respectively^[4]. This series of methods can effectively improve the quality of the dataset and prevent overfitting.

To prevent overfitting, Chouhan et al. added some noise to the dataset to significantly improve the generalization ability of the dataset. At the same time, it could also be used as an extension of the dataset. After that, they unified the image format to $224 \times 224 \times 3$. Then the dataset image was enhanced by three techniques: random horizontal flipping, random size clipping, and adjusting image intensity^[5].

Heidari et al. proposed a scheme to add an image preprocessing algorithm to identify and delete the aperture region of the image. Specifically, the algorithm detects the pixel values of the maximum and minimum images and then divides the original image into binary images. Then bilateral filter and histogram equalization technology were applied to detect different connection components with higher intensity (brightness) in the image. From all detected connection areas, select the largest area at the bottom of the image, define it like the aperture area and remove it^[6]. Using this method, the pattern and characteristics of infection can be better improved.

Ferreira et al. developed an algorithm based on u-net CNN to cut the chest from the X-ray image. Firstly, the algorithm uses the pre-trained U-shaped net weight and transfer learning method to segment the lung from CXR to create a binary mask. It then creates a region of interest (ROI) from the extreme points on the lung mask and generates a bounding box to finally crop the chest and remote areas that are not important for pneumonia detection^[7]. This method can also significantly improve the detection results.

In Khatri, Archit, et al.'s method, they cropped the image to have only lungs. Then they rotated the X-ray image slightly to make the lungs vertically symmetrical; For some X-ray images that are

too bright or too low, the formula is used for normalization calculation to reduce the error of subsequent steps^[8].

Table 1. Comparison of various pneumonia preprocessing techniques

Reference number	Processing Technique	Type of data processed	Strength
[4]	Image normalization+ Rotation+Scaling+ Translation	Chest X-ray	improve the training accuracy by increasing the number of samples
[5]	Adding noise+ Random horizontal flip + Random crop size + Images with different intensities	Chest X-ray	Improve the generalization of datasets
[6]	add an algorithm to identify and remove the diaphragm region depicted on the image	Chest X-ray (COVID-19)	Prevent the influence of the diaphragm region on training results
[7]	remove anatomical regions from the chest X-ray to improve the training ability	Chest X-ray	Reduce interference in irrelevant areas
[8]	Image size unification + intensity normalization	Chest X-ray	Unify the image brightness features and eliminate the influence of background light

3. Target detection algorithm

3.1. Existing results

We look for the common target detection algorithms and list them as follows. It can be seen from the table that the mAP of these models does not exceed 30%, but the highest precision reaches 0.968 when the threshold = 0.6.

Table 2. Target detection algorithm and dataset

Reference	Author	Database	Method	Result
[9]	Z. -Y. Yang and Q. Zhao	Kaggle RSNA	Mask-RCNN Resnet50	Threshold=0.93 Precision=0.691 MS=0.362
[9]	Z. -Y. Yang and Q. Zhao	Kaggle RSNA	Mask-RCNN ResNet101	Threshold=0.97 Precision=0.841 MS=0.363
[9]	Z. -Y. Yang and Q. Zhao	Kaggle RSNA	Faster-RCNN ResNet101	Threshold=0.6 Precision=0.968 MS=0.311
[10]	A. Nayyar, R. Jain and Y. Upadhyay	Kaggle RSNA	Mask-RCNN	IOU=0.155
[10]	A. Nayyar, R. Jain and Y. Upadhyay	Kaggle RSNA	CNN Segment	IOU=0.116
[10]	A. Nayyar, R. Jain and Y. Upadhyay	Kaggle RSNA	Yolo V3	IOU=0.141
[10]	A. Nayyar, R. Jain and Y. Upadhyay	Kaggle RSNA	Unet	IOU=0.094
[10]	A. Nayyar, R. Jain and Y. Upadhyay	Kaggle RSNA	Resnet Segmentation	IOU=0.109

[11]	A. F. Al Mubarak, J. A. M. Dominique and A. H. Thias	Kaggle RSNA	Mask-RCNN	Accuracy=0.856 Sensitivity=0.5152 Specificity=0.9483
[12]	N. Darapaneni et al	Kaggle RSNA	Mask-RCNN	mAP=0.89
[13]	H. Ko, H. Ha, H. Cho, K. Seo and J. Lee	Subset of NIH dataset	Mask 150	mAP=0.15035
[13]	H. Ko, H. Ha, H. Cho, K. Seo and J. Lee	Subset of NIH dataset	Mask 160	mAP=0.16211
[13]	H. Ko, H. Ha, H. Cho, K. Seo and J. Lee	Subset of NIH dataset	RN 178	mAP=0.17813
[13]	H. Ko, H. Ha, H. Cho, K. Seo and J. Lee	Subset of NIH dataset	RN 184	mAP=0.18563
[13]	H. Ko, H. Ha, H. Cho, K. Seo and J. Lee	Subset of NIH dataset	RN 201	mAP=0.20147
[14]	Z. Yu	Kaggle RSNA	U-Efficient	mAP=0.18142

3.2. Two-Stage target detection algorithm

Generally speaking, the work of this kind of network is first generating a proposal region, then identifying the category of this region. R-CNN network is the original two-stage target detection algorithm. As shown in table 3, Mask-RCNN and Faster-RCNN are both two-stage. These networks first extract the region of interest(ROI) from the image, then use CNN network to extract the features of images through convolution calculation. Finally, these networks use SVM classifier to decide the categories of the detected object. The advantage of two-stage networks is high detecting accuracy.

As for Faster-RCNN network, it uses RPN(Region Proposal Network) to generate the region of interest. The image will first be extracted by the features by convolution layers, this structure is called Extractor. Then RPN(shown in Figure 2) will use these features for classification and regression. The anchors will be generated by AnchorTargetCreator. RPN only identifies background and objects to generate region proposals. Finally, features, ROIs, and ground truth box will be sent to ROIHead(shown in Figure 3) to generate the label and location of target objects.

An effective improvement of Faster-RCNN is changing the Extractor from VGG16 to ResNet50. ResNet50 can solve the problem of gradient descent disappearing. Two kinds of Bottlenecks used in ResNet50 are shown below.

Mask-RCNN is an extended network of Faster-RCNN. It adds a new layer, which is parallel to the existing bounding box and classification, to predict the separation mask on each ROI. Compared with Faster-RCNN, it adds ROIAlign and Fully Convolutional Network. Mask-RCNN contains two kinds of predictions, they are classification prediction and mask prediction. Classification prediction is the same as that in Faster-RCNN, which predicts the locations and labels. Mask prediction is used to do pixel-level segmentation according to the binary mask.

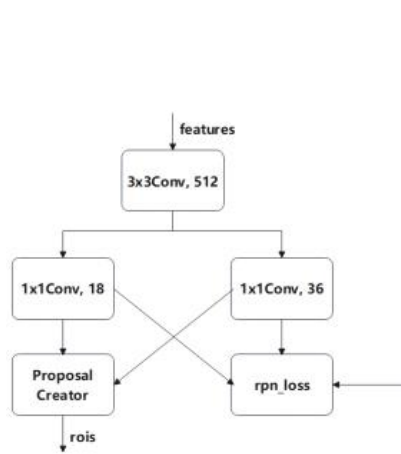


Figure 1. RPN and AnchorCreator

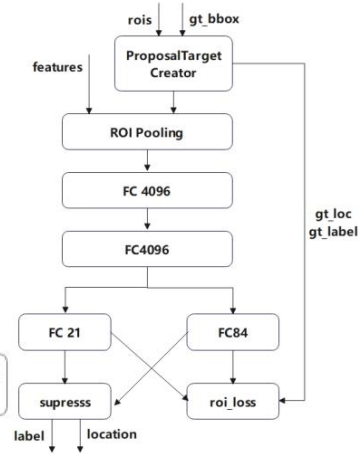


Figure 2. ROIHead

3.3. One-Stage target detection algorithm

Compared with the two-stage target detection algorithm, the one-stage network does not need the proposal region. The network will uniformly sample at different positions of the picture, use CNN to extract features, and directly do classification and regression based on these features. It is just like RPN structure complete positioning and classification at the same time. This kind of network has a fast running speed.

A typical example is Yolo. Yolo has a fast running speed and convenient optimization mode. It transforms the problem of target detection into an end-to-end regression problem. Firstly, it convolutes the whole image and divides it into $S \times S$ grids. Then it predicts the probability of each category for the bounding boxes in each grid, and finally uses the maximum suppression of the loss function to obtain the final prediction result. After upgrading several versions, YOLOv5 has a quite different network from YOLOv1. YOLOv5 uses image enhancement technology and CSP structure. It also combines the anchor technology of Faster-RCNN, uses a multi-scale training method, and changes the loss function to CIOU_Loss (shown below).

$$CIOU_Loss = 1 - CIOU = 1 - \left(IOU - \frac{Distance_c^2}{Distance_c^2} - \frac{v^2}{(1 - IOU) + v} \right) \quad (1)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p} \right)^2 \quad (2)$$

Another one-stage target detection network is SSD (Single Shot MultiBox Detector). It also transforms detection to regression and combines the anchor technology in Faster-RCNN. SSD is based on VGG16, and adds Pyramidal Feature Hierarchy detection method. SSD has a speed close to Yolo and the detection accuracy close to Faster-RCNN. However, the prior box needs to be set manually, which leads to the case that the rich experience of personnel is needed in the commissioning process.

3.4. Analysis

From the results above, it can be seen that most pneumonia detection technologies today use Mask-RCNN network. From the results, mask RCNN is indeed more suitable for the detection of pneumonia than Faster-RCNN. However, if the extractor is changed from VGG16 to ResNet50, Faster-RCNN can also have a performance as good as that of Mask-RCNN. At the same time, YOLOv5x, as a very popular network, although its accuracy is not as good as the previous two networks, its light volume and high detection speed will become its unique advantages. In addition, Resnet Segmentation, Unet and other object detection algorithm also have their advantages and disadvantages.

It is worth noting that the accuracy of pneumonia detection can be improved not only by modifying the network structure. Image preprocessing is also a method worth trying. People can improve the

detection accuracy and robustness of images through doing image de marginalization, threshold processing, and so on, to get better network performance.

4. Method Validation

To prove the feasibility of the target detection networks mentioned above, we used Faster-RCNN-VGG16, Yolov5x, and Faster-RCNN-ResNet50 to verify them on RSNA datasets. These networks are trained with RTX 2080ti under the TensorFlow framework. Nearly 30,000 chest radiographs from the competition dataset RSNA were trained, of which over 9,000 were radiographs with pneumonia, and sometimes a single radiograph with pneumonia can contain multiple lesions. The underlying form of the dataset is a .csv file and the format is not the same as the commonly used dataset format. We, therefore, wrote our script to extract the 9000+ X-rays with pneumonia lesions and calculate their coordinates to get a working dataset.

The mAP, Precision of Recall of these three networks are shown in table 4. We first use Faster-RCNN-VGG16 network. As can be seen from the table, the mAP, precision, and recall are relatively high. The good performance proves the suitability of this network for pneumonia detection. Then we use Yolov5x. At present, most of the authors use Yolov3 network. Compared with Yolov3, Yolov5x has many updates and optimizations, so we use Yolov5x to verify its feasibility. It can be seen from the results that Yolov5x has better performance than Yolov3. Finally, we use Faster-RCNN-ResNet50. Resnet50 solves the problems of gradient descent disappearance and difficult optimization of VGG16. It can also be seen from the performance that ResNet50 is more suitable to be used as a feature extraction structure of Faster-RCNN.

Table 3. Results of three networks

	mAP	Recall	Precision
Faster-RCNN-VGG16	52.26%	54.81%	73.84%
YOLOv5x	53.37%	55.77%	75.62%
FasterRCNN-ResNet50	54.53%	58.95%	78.97%

From the experimental results, it can be found that Faster-RCNN and Yolov5x are all suitable to be used in this dataset, RSNA. It means that not only traditional target detection algorithms, but also new research achievements, such as Yolov5, and improved networks, such as Faster-RCNN-VGG16, can also be used for pneumonia X-ray detection. In addition, these target detection networks can not only be used on one or two datasets, but also perform well on different datasets.

5. Conclusion

This paper summarizes and analyzes the existing data sets, image preprocessing technology, and image detection technology. This paper validates three networks Faster-RCNN-VGG16, Yolov5x, and Faster-RCNN-Resnet50 by using the RSNA dataset. The results show that both one-stage and two-stage target detection networks can be well applied to pneumonia detection. Even if different data sets are used for training, these target detection networks will all have good performance. At the same time, image preprocessing technology can also effectively improve the accuracy of pneumonia detection.

References

- [1] NIH Chest X-ray. 2020-01-05. <http://kaggle.com/nih-chest-xrays/data>.
- [2] Chest X-Ray Images (Pneumonia). 2020-01-05. <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- [3] RSNA Pneumonia Detection Challenge. 2020-01-05. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge> .

- [4] Rahman, Tawsifur, et al. "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray." *Applied Sciences* 10.9 (2020): 3233.
- [5] Chouhan, Vikash, et al. "A novel transfer learning based approach for pneumonia detection in chest X-ray images." *Applied Sciences* 10.2 (2020): 559.
- [6] Heidari, Morteza, et al. "Detecting COVID-19 infected pneumonia from x-ray images using a deep learning model with image preprocessing algorithm." *Medical Imaging 2021: Computer-Aided Diagnosis*. Vol. 11597. International Society for Optics and Photonics, 2021.
- [7] Ferreira, José Raniery, et al. "Multi-view ensemble convolutional neural network to improve classification of pneumonia in low contrast chest x-ray images." 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020.
- [8] Khatri, Archit, et al. "Pneumonia identification in chest X-ray images using EMD." *Trends in Communication, Cloud, and Big Data* (2020): 87-98.
- [9] Z. -Y. Yang and Q. Zhao, "A Multiple Deep Learner Approach for X-Ray Image-Based Pneumonia Detection," 2020 International Conference on Machine Learning and Cybernetics (ICMLC), 2020, pp. 70-75, doi: 10.1109/ICMLC51923.2020.9469043.
- [10] A. Nayyar, R. Jain and Y. Upadhyay, "Object detection based approach for Automatic detection of Pneumonia," 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), 2020, pp. 1-6, doi: 10.1109/icABCD49160.2020.9183876.
- [11] A. F. Al Mubarak, J. A. M. Dominique and A. H. Thias, "Pneumonia Detection with Deep Convolutional Architecture," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT), 2019, pp. 486-489, doi: 10.1109/ICAIIIT.2019.8834476.
- [12] N. Darapaneni et al., "COVID 19 Severity of Pneumonia Analysis Using Chest X Rays," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020, pp. 381-386, doi: 10.1109/ICIIS51140.2020.9342702.
- [13] H. Ko, H. Ha, H. Cho, K. Seo and J. Lee, "Pneumonia Detection with Weighted Voting Ensemble of CNN Models," 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2019, pp. 306-310, doi: 10.1109/ICAIBD.2019.8837042.
- [14] Z. Yu, "Pneumonia Detection with U-EfficientNet," 2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC), 2021, pp. 591-594, doi: 10.1109/DSC53577.2021.00093.