# Attention-Guided Convolutional Neural Network for Detecting Pneumonia on Chest X-Rays

Bingchuan Li, Guixia Kang, Kai Cheng and Ningbo Zhang

*Abstract*— **Pneumonia is a common infectious disease in the world. Its main diagnostic method is chest X-ray (CXR) examination. However, the high visual similarity between a large number of pathologies in CXR makes the interpretation and differentiation of pneumonia a challenge. In this paper, we propose an improved convolutional neural network (CNN) model for pneumonia detection. In order to guide the CNN to focus on disease-specific attended region, the pneumonia area of image is erased and marked as a non-pneumonia sample. In addition, transfer learning is used to segment the interest region of lungs to suppress background interference. The experimental results show that the proposed method is superior to the state-of-the-art object detection model in terms of accuracy and false positive rate.**

## I. Introduction

Pneumonia is a major cause of global morbidity and mortality. In the United States, pneumonia accounts for over 500,000 visits to emergency departments [1] and over 50,000 deaths in 2015, keeping the ailment on the list of top 10 causes of death in the country [2].

Chest X-ray (CXR) is the most suitable imaging modality to diagnose pneumonia. Usually, pneumonia manifests as an area or areas of increased opacity [3] on CXR. However, the imaging reviews on CXR are complicated because a number of factors, such as positioning of the patient and depth of inspiration, can alter the appearance of CXR [4]. In addition, some other conditions of the lungs can also affect the analysis and diagnosis of pneumonia, such as infiltration, mass, volume loss, or post-radiation and surgical changes [5], [6]. These conditions have greatly increased the difficulty for clinicians to read and analyze CXR images.

Recently, deep learning has played an increasingly important role in the automatic analysis and clinical diagnosis of medical images. In particular, some methods based on convolutional neural networks (CNN) have been successfully applied to classify diseases, locate abnormal regions or segment lesions in CXR images. For instance, Rajpurkar, *et al.* [7] developed an end-to-end classification model based on CNN. This model takes a CXR image as input, and outputs 14 pathological probabilities, including predictions of pneumonia. Although this model achieved good predictive results in some of chest diseases, it is relatively low in the classification accuracy of pneumonia. Further, Sedai, *et al.* [8] combined weak supervised learning algorithms and CNN to locate the disease region in CXR image. This model only uses the disease category labels to intend to avoid the difficulty in obtaining pixel level annotation. However, the detection accuracy of this model is unsatisfactory and needs to be improved.

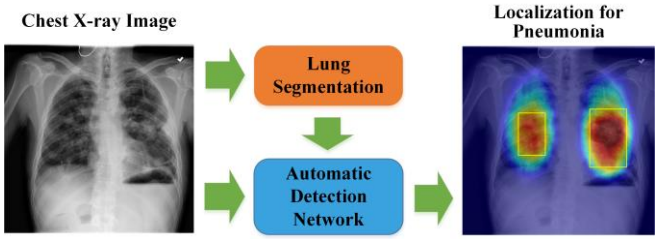


Figure 1. Overview of our method for pneumonia diagnosis. The network reads chest X-ray images and produces predictions of pneumonia regions. The visualization result is generated by rendering the final output tensor as heatmap and overlapping on the original image.

As the Radiological Society of North America (RSNA) pneumonia detection challenge [9] discloses relevant datasets, some advanced object detection models, such as CoupleNet [10], RetinaNet [11], and Mask R-CNN [12], have been applied to this task. However, unlike the detection of targets with salient features in natural images, the visual similarity between different pathological characteristics in CXR images makes the differentiation and interpretation of pneumonia more difficult. In addition, the clavicle, complex lung structure and fine texture in chest X-rays show white opacity in CXR, which can cause a lot of noise during the detection of pneumonia. Therefore, effectively suppressing background interference and reducing the false positive rate predicted by the model remains a challenge. Based on this fact, we propose an improved squeeze-and-excitation network (SENet) architecture [13], which delivers the outstanding performance in suppressing useless features. In order to maximize the attention of the network to the characteristics of pneumonia, we erase the pneumonia area in the CXR image as a non-pneumonia sample for adversarial learning. Furthermore, the transfer learning method is utilized to segment the region of interest (ROI) in the lungs to get rid of unnecessary information in the background. An overview of our approach is shown in Fig. 1.

The rest of the paper is organized as follows: Section II clarifies the specific details of our proposed method. The specific experimental procedures and evaluation of experimental results are described in Section III. Finally, section IV summarizes the work and forecast the future research direction.

*The research is supported by the National Natural Science Foundation of China (No.61471064), and National Science and Technology Major Project of China (No. 2017ZX03001022).

Bingchuan Li, Guixia Kang, Kai Cheng and Ningbo Zhang are with the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China (corresponding author's e-mail: gxkang@bupt.edu.cn).

Bingchuan Li, Guixia Kang are with the Wuxi BUPT Sensory Technology and Industry Institute CO.LTD, Wuxi, 214135, China (e-mail: bcli.wta@gmail.com)

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

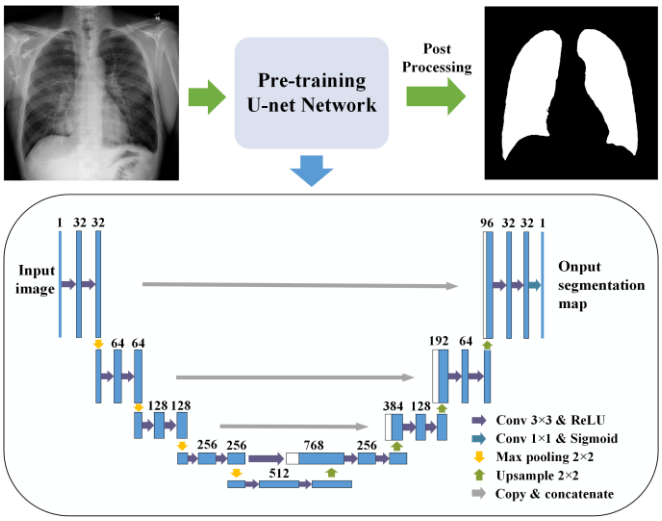


Figure 2. The architecture of lungs ROI segmentation. The original image is subjected to pre-trained U-net model and post-processed to obtain segmentation map. The inside of the rounded rectangle in the image is an U-net model architecture diagram.

## II. METHODOLOGY

In this section, the proposed algorithm for pneumonia detection contains three main parts: CXR images preprocessing, lungs ROI segmentation with transfer learning, and design of automatic detection model for pneumonia based on CNN.

### *A. CXR Image Preprocessing*

In general, the performance of CXR images varies greatly among individuals or under different conditions, so that the model is difficult to distinguish between pneumonia characteristics and other noise. So we design adversarial samples to make the model more focused on the target being detected. The specific approach is to replace the pixels of pneumonia area by the average pixels of the whole training image, which is represented as erasing the object region in the image with a single pixel value. The erased image is relabeled as a non-pneumonia sample and fed into the model for training. Since the erased image and the original image are identical in the non-pneumonia region, it can increase the model's attention to the characteristics of the pneumonia area in the training process and reduce the model's misjudgment of noise. In addition, deep learning often requires more training data to improve the robust performance of the network. Thus we enlarge the training dataset by data augmentation, including rotation, translation, flipping and scaling.

### *B. Lungs region of interest segmentation*

Segmentation of the ROI in the lungs can effectively suppress background interference in CXR images. Generally, automatic image segmentation requires a large number of accurate ground truth masks as the label for supervised learning of the model, which is not available on our pneumonia detection dataset. So we migrate the pre-trained model on other datasets for fine-tuning to extract the lungs ROI.

The U-net [14] model has been proved to be effective in medical image segmentation tasks. For example, Rashid, *et al.* [15] have used this model to segment lungs ROI from CXR images, achieving 97.7% and 97.1% segmentation accuracy on the Montgomery [16] and JSRT [17] datasets, respectively. In this work, we use their approach to pre-train the U-net model on these two datasets and generate lung ROI masks on our dataset. After that, post-processing is utilized to improve segmentation performance, including using conditional random fields [18], setting thresholds to remove small connected domains, and removing segmented images that are out of bounds. Segmented images and original images are combined as the training dataset for pneumonia detection model. It is worth noting that the segmentation performance of the automatic segmentation model may behave differently on diverse CXR images. A few CXR images with poor image quality cannot achieve good segmentation results, which are unfavorable for training process of the subsequent pneumonia detection model. Thus, poor segmentation images are discarded during the strict post processing phase. The architecture of lungs ROI segmentation model is shown in Fig. 2.

### *C. Convolutional Neural Network Model Design*

In this paper, SENet design is used to improve the full convolutional neural network [19] architecture. As shown in Fig. 3 (b) and (c), the architecture consists of two parts. In the first part, SE-ResNet34 [20] architecture is introduced as backbone to extract features. As can be seen in Fig. 3 (a), the side branch of SE-ResNet modules can automatically learn weights and generates the importance of each feature mapping channel. Multiplying the generated weights by the local feature channel enables the network to selectively enhance the beneficial feature channel and suppress the useless feature channel, thereby implementing feature channel adaptive calibration. Simultaneously, the use of the residual block [21] enables the deep gradient to be smoothly transmitted to the shallow layer during the training process in the case where the number of network layers is deepened, so that the entire network is effectively trained. After feature extraction, we use consecutive deconvolution processes instead of the fully connected layer of SE-ResNet to map the extracted high-level features to the size of the input image. In addition, we designed a 1×1 convolution to fuse features of different hierarchical features, which are shown in the

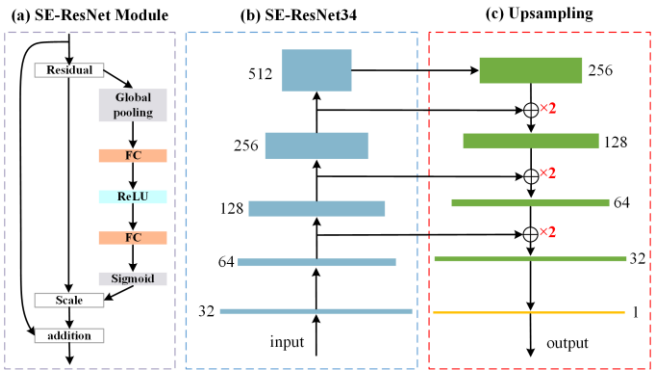


Figure 3. (a) SE-ResNet module structure, where FC refers to fully connected layer, ReLU and Sigmoid represent activation functions. (b) and (c) constitute the architecture of the detection model, where (b) represents feature extraction using stacked SE-ResNet modules, and (c) represents the output of the feature map through the upsampling process. The horizontal arrow in the figure indicates the result after 1×1 convolution process.

horizontal arrows in the Fig. 3. Finally, the feature maps of multiple channels are integrated into a single channel for output and activated by Sigmoid function. Each pixel of the output image represents the probability of having pneumonia, and the pneumonia area in the image is obtained by setting an appropriate threshold.

## III. Experiments and Results

In this section, we introduced the experimental data and implementation details. After that, we evaluated the results of our model and the winner's model in the RSNA pneumonia detection challenge. Since the methods in the top rankings in the competition mostly use RetinaNet [11] and Mask R-CNN [12] as their basic detection models, we constructed these two models and implemented the algorithm details of the winners open source. The visualized pneumonia detection results of our method were shown at the end of the section.

### A. Dataset

In this research, CXR images are obtained from the (RSNA) pneumonia detection challenge. There are 8,964 pneumonia-labeled CXR images, and every pneumonia area of each image is marked with a bounding box. In addition, the remaining 20,025 non-pneumonia CXR images can be available. Considering that the 3000 CXR images used as test sets in the competition are susceptible to overfitting, it is impossible to accurately judge the performance of the model. Thus, we conduct a 5-fold cross-validation to evaluate all models.

### B. Implementation Details

According to the methodology described in Section II, the data is processed first before being sent to our designed network training. The raw data is processed through the following three steps: (a) Resize the original single-channel CXR image from 1024×1024 to 512×512 pixels for faster training. (b) Replicate and expand the image into 3 channels, and perform preprocess for each channel to obtain a random image enhancement effect. (c) Replace one of the channels with segmented image when segmented image is not discarded. Besides, the label of the training data needs to be made. Since the bounding box is the pneumonia region, we generate a zero matrix of the original image size and mask the bounding box region to get the ground truth (GT) label.

After that, Pytorch [22] framework is used to build our CNN detection model. Binary cross entropy [23] is used as loss function for model training and validation, and stochastic gradient descent algorithm (SGD) is used to update the weight of the model to convergence. We set the initial learning rate to 0.001, the momentum to be 0.9 and the weight decay to be 0.0001. When the value of loss function in validation set no longer falls within 5 epochs during training process, the learning rate will be decrease by 5 times. Training process will not stop until the learning rate decays more than 100 times. The model is trained on an NVIDIA 1080Ti GPU and executed approximately 40 epochs.

### C. Experimental Results

In general, when the model is reliable, the higher the confidence of the predicted bounding box, the more likely it is to hit the real pneumonia area. The corresponding threshold and other hyperparameters are adjusted according to the performance settings of the verification set to get final detection result. The visual results predicted by our model are shown in Fig. 4.

TABLE I. The Precision and Recall under varying T(IoU)

| Methods | Indicator | T(IoU) | | | | |
|---|---|---|---|---|---|---|
| | | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** |
| Mask R-CNN [12] | precision | 0.532 | 0.453 | 0.365 | 0.314 | 0.287 |
| | recall | 0.759 | 0.715 | 0.697 | 0.643 | 0.626 |
| Retinanet [11] | precision | 0.578 | 0.499 | 0.420 | 0.365 | 0.313 |
| | recall | 0.818 | 0.774 | 0.753 | 0.726 | 0.695 |
| Our method | precision | 0.611 | 0.565 | 0.493 | 0.431 | 0.389 |
| | recall | 0.835 | 0.816 | 0.798 | 0.770 | 0.758 |

TABLE II. The average accuracy and FPR results

| Methods | Accuracy | FPR |
|---|---|---|
| Mask R-CNN[12] | 0.183 | 0.291 |
| RetinaNet[11] | 0.225 | 0.243 |
| Proposed method in [24] | 0.231 | 0.227 |
| Our method | **0.262** | **0.194** |

In order to more fully measure the performance of the model, we not only used the official evaluation indicators to calculate the accuracy and the false positive rate (FPR) under varying intersection over union (IoU) thresholds, but also calculate the precision and recall rate at a single IoU threshold. We simply write IoU threshold as T(IoU). The IoU of a set of predicted bounding boxes and GT bounding boxes is calculated as follows:

$$IoU(A, B) = \frac{A \cap B}{A \cup B}, \tag{1}$$

where A and B denote the area enclosed by the predicted bounding box and GT bounding box, respectively. We set five different T(IoU) from 0.3 to 0.7 to find out whether the target has been hit or missed. For example, T(IoU)>0.5 means a predicted bounding box is considered to hit the real pneumonia regions if its IoU with a ground truth bounding box is greater than 0.5. For each threshold, true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) are obtained by comparing predicted results to all GT objects. Thus, the *precision* and *recall* can be calculated at varying T(IoU) points, which can be written as

$$precision = TP / (TP + FP), \tag{2}$$

$$recall = TP / (TP + FN). \tag{3}$$

Generally, the precision reflects the accuracy of the model, and the recall reflects whether the model can find all the correct samples. As can be seen from Tab. I, the evaluated models have a low precision and a high recall, which indicates that there are larger samples for model error detection and fewer leaks. More specifically, T(IoU) is set from 0.4 to 0.75 with a step of 0.05 according to the official metrics, and the average accuracy and FPR are calculated under this set of thresholds. The calculation formulas are shown in equation (4) and (5).

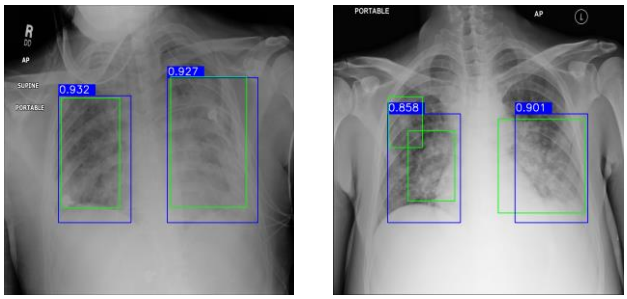


Figure 4. Visual display of pneumonia dectection. The green and blue boxes in the figure are bounding box markers for ground truth and prediction, respectively. The numbers in the figure represent the confidence of the predicted bounding box, which is the average of the probability of predicting pneumonia per pixel in a closed region.

$$Accuracy = \frac{1}{|thresholds|}\sum_{t}\frac{TP(t)}{TP(t)+FP(t)+FN(t)}, \quad (4)$$

$$FPR = \frac{1}{|thresholds|}\sum_{t}\frac{FP(t)}{FP(t)+TN(t)}. \quad (5)$$

Among them, accuracy represents the overall consideration of precision and recall rate, which directly reflects the overall level of the ratio of misdetection and missed detection. It can be seen in Tab. II that our method has better detection performance in terms of the accuracy and false positive rate. This result is due to the guiding model's attention to the pneumonia area, which enables the model to more fully learn the difference between pneumonia and noise, thus effectively suppress the detection of false positive samples.

## IV. Conclusion

In this paper, we analyze the characterization of disease in CXR images and propose a method based on attention-guided CNN for detecting pneumonia. In this method, we use SE-ResNet as a backbone to design a fully convolutional neural network model for end-to-end output detection objects. In particular, constructing adversarial samples by erasing the pneumonia area and segmenting the lungs ROI by utilizing can increase the model's attention to the pneumonia area and suppress other noise. From experimental results, it can be seen that the end-to-end detection model can easily be misled to predict more false positive samples due to the diversity and complexity of CXR images. On the contrary, the proposed method enables the network to learn the features of pneumonia in a targeted manner, and the discriminative power of false positive samples is significantly improved. In the next step, we will integrate the proposed method into a unified framework. Future research directions will be to improve the model's ability to identify false positive samples through more advanced scene interpretations and more adequate validation labels.

## Acknowledgment

The dataset we used comes from the Radiological Society of North America (RSNA) pneumonia detection challenge. Thanks to the RSNA, the US National Institutes of Health, the Society of Thoracic Radiology, and MD.ai to develop this dataset.

## References

[1] Rui P, Kang K. National Ambulatory Medical Care Survey: 2015 Emergency Department Summary Tables. Table 27.

[2] Deaths: Final Data for 2015. Supplemental Tables. Tables I-21, I-22.

[3] Franquet, Tomás. "Imaging of community-acquired pneumonia." *Journal of thoracic imaging* 33.5 (2018): 282-294.

[4] Kelly, Barry. "The Chest Radiograph." *Ulster Medical Journal*81.3(2012):143-8.

[5] Wang, Xiaosong, et al. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017.

[6] Scipione, Roberto, et al. "Pulmonary Imaging Findings After Surgery, Chemotherapy and Radiotherapy." *Diagnostic Imaging for Thoracic Surgery*. Springer, Cham, 2018. 343-358.

[7] Rajpurkar, Pranav, et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." *arXiv preprint arXiv:1711.05225* (2017).

[8] Sedai, Suman, et al. "Deep multiscale convolutional feature learning for weakly supervised localization of chest pathologies in X-ray images." *International Workshop on Machine Learning in Medical Imaging*. Springer, Cham, 2018.

[9] RSNA Pneumonia Detection Challenge. Radiological Society of North America. *https://www.kaggle.com/c/rsna-pneumonia-detection-challenge*, 2018.

[10] Zhu, Yousong, et al. "Couplenet: Coupling global structure with local parts for object detection." *Proc. of Int'l Conf. on Computer Vision (ICCV)*. Vol. 2. 2017.

[11] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *IEEE transactions on pattern analysis and machine intelligence* (2018).

[12] He, Kaiming, et al. "Mask R-CNN." *IEEE Transactions on Pattern Analysis & Machine Intelligence* PP.99(2017):1-1.

[13] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[14] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *arXiv preprint arXiv:1709.01507* 7 (2017).

[15] Ronneberger, Olaf, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2015:234-241.

[16] Rashid, Rabia, Muhammad Usman Akram, and Taimur Hassan. "Fully Convolutional Neural Network for Lungs Segmentation from Chest X-Rays." *International Conference Image Analysis and Recognition*. Springer, Cham, 2018.

[17] Jaeger, Stefan, et al. "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases." *Quantitative imaging in medicine and surgery* 4.6 (2014): 475.

[18] Shiraishi, Junji, et al. "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules." *American Journal of Roentgenology* 174.1 (2000): 71-74.

[19] Lafferty, John, Andrew McCallum, and Fernando CN Pereira. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data." (2001).

[20] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition(CVPR)*. 2015.

[21] He, Kaiming, et al. "Identity mappings in deep residual networks." *European conference on computer vision*. Springer, Cham, 2016.

[22] Paszke, Adam, et al. "Automatic differentiation in pytorch." (2017).

[23] Kroese, Dirk P., Sergey Porotsky, and Reuven Y. Rubinstein. "The cross-entropy method for continuous multi-extremal optimization." *Methodology and Computing in Applied Probability* 8.3 (2006): 383-407.

[24] Team, The DeepRadiology. "Pneumonia Detection in Chest Radiographs." *arXiv preprint arXiv:1811.08939* (2018).