Pneumonia Detection on Chest X-Rays with Deep Learning

Zhou Yang
George Washington University
zhou_yang@gwu.edu

Fang Jin
George Washington University
fangjin@gwu.edu

ABSTRACT

Automating the process of diagnosing pneumonia from radio-graph is of great importance and practical implication. However, this task is challenging due to three facts. First, training a reliable disease detection model requires a large amount of labeled dataset that is not available before the Chest X-ray 14 dataset. Second, unlike general image classification that intends to extract main patterns (feature maps) from images, the thoracic disease diagnosing task aims at extracting disease information (feature maps) from the background. In this paper, we model the disease detection task using convolutional neural network and customized the specialty of Chest X-ray dataset. The experimental results showed that the proposed model can achieve the good performance. We conclude that: (1).CNN with Transfer learning can classify the X-ray images with an accuracy of around 70.5% based on this ChestX-ray14 dataset, and the transfer learning saves computation resources; (2).Like pneumonia, other thoracic disease detection and localization can also be automated.

1 INTRODUCTION

Automating the process of diagnosing pneumonia by reading chest X-ray images is of practical importance but challenging. Pneumonia refers to lung inflammation (pneumonitis), which affects approximately 450 million people globally (7% of the population) and results in around 4 million deaths every year [4, 7]. Detecting and diagnosing pneumonia is of great importance since it's the first step towards pneumonia treatment. Combined with physical signs and medical tests, a chest x-ray is the best available method for diagnosing pneumonia since it is relatively cheap and safe. However, detecting pneumonia using a chest x-ray is challenging because it requires expertise and experience to correctly interpret the pathology. Also, using chest x-ray alone is even difficult for practicing radiologists to have an accurate diagnosis and interpretation.

The pneumonia diagnosing process can be automated by well trained models. Traditionally, diagnosing and localizing disease patterns heavily relies on well-trained radiologists or medical image professionals. Besides, the diagnosis process is highly repetitive because the task majorly depends on detecting certain patterns on the chest X-rays. That is, a pathology is always associated with certain patterns on chest X-rays, while other pathology is connected with distinct patterns. Those characteristics in chest X-rays diagnosis facilitate automated disease detection since well-trained models are good at detecting repetitive patterns with invariant characteristics. However, limited work has been done to address this critical problem due to several reasons. The fundamental reason is that the available dataset is far from enough to train a reliable model to detect pneumonia from radiograph before the release of NIH ChestX-ray14 dataset. Since NIH ChestX-ray14 includes about 112,120 images from 30,805 different patients, it provides enough labeled images to train a reliable model. Motivated by the availability of dataset and advancement in deep learning, we present

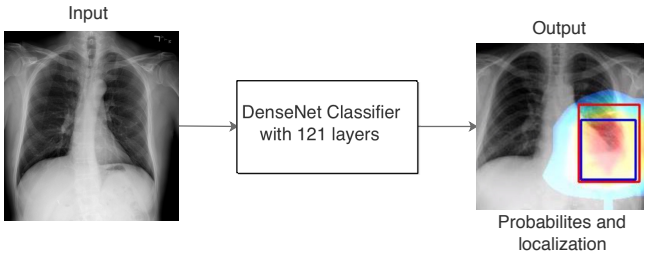

Figure 1: Pneumonia classifier framework. The input is a chest x-ray, and the output is the probability of the presence of pneumonia.

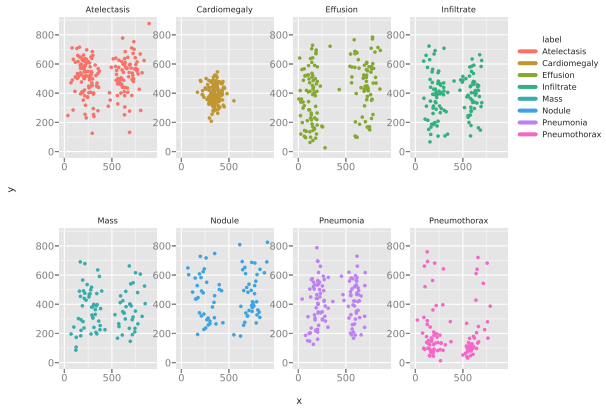

Figure 2: The coordinates distribution of 8 thoracic disease according to the coordinates labeled by radiologists.

a model that automates the pneumonia diagnosis by using image classification method. The framework is shown in Figre 1.

The automation of pneumonia diagnosis is accomplished by image classification. The task in Image Classification is to take an array of pixels that represents a single image and assign a label to it. The input of a classification task consists of a set of N images, each labeled with one of K different classes. The main task in classification is to learn a mapping from images to its label while minimizing the total differences between the mapping results and labels. Our pneumonia classifier is based on a pre-trained 121-layer DenseNet, ingesting chest x-ray images and outputting the probabilities for pneumonia. Also, disease localization is provided by a heatmap that can further be transformed into bounding boxes. The pneumonia classifier is trained on ChestX-ray14 dataset [8], which contains 112,120 chest x-ray images from either posteroanterior (PA) or anteroposterior (AP) view [5]. The Chestx-ray14 dataset covers 14

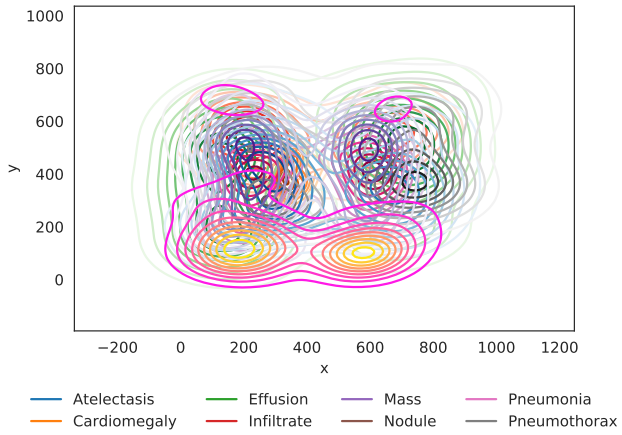

Figure 3: The coordinates distribution of eight thoracic diseases in a radiograph.

diseases. We show the disease distribution in Figure 2 and Figre 3. In this paper, we first show how it works to classify pneumonia, and then modify the model to apply it on eight diseases [1].

Although the experimental result presented in this paper is output from an 8-disease detection classifier, it can be easily modified to detect all the 14 diseases in the ChestX-ray14 dataset.

2 RELATED WORK

Recent advancements in deep learning and the availability of large labeled datasets have enabled disease detection with reliable performance in a wide variety of medical imaging tasks, including thoracic disease detection [8], skin cancer classification [**?**], arrhythmia detection, and hemorrhage identification. Wang et al. proposed the TieNet that combines image features and text embedding extracted from medical report to classify the chest x-ray images [9]. Yao et al. presented how to leverage interdependencies among target labels in predicting 14 pathologic patterns from chest x-rays and establish state-of-the-art results [10]. Rajpurkar et al. proposed the ChesXnet and showed that it can outperform radiologists on thoracic disease detection. Li et al. presented a unified approach that performs disease identification and localization with weak supervision [3]. Wang et al. opened the ChestX-ray14 dataset that makes large scale thoracic disease classification possible [8]

3 PROBLEM FORMULATION

Given pneumonia disease (or any thoracic disease), the pneumonia disease detection task is an image classification problem, where the input is a chest X-ray image X with label $\tilde{Y} \in 0, 1'$. and the output is the probability y denoting the probability of the existence of pneumonia. More generally, we formulated the thoracic disease classification task as follow.

Definition 3.1 (Thoracic Disease Classification Task). *Given a set of thoracic radiographs $X = \{X_1, X_2, x_3, ...\}$, find a mapping $\Psi : R^m \rightarrow Y$, where $Y \in [0, 1]$, such that the task error is minimized. Formally, we have*

$$\arg\min_{\Psi} \sum_{j=1}^{m} ||\Psi(X_j) - Y_j||_2 \tag{1}$$

The mapping Ψ from radiopgrah features to labels is too complex to represent by a single function. Usually, such a complex mapping is implemented by a neural network. In our case, we implement it based on a pre-trained DenseNet that has 121 layers.

For loss function, it is formulated as to minimize the cross entropy as denoted in equation 2 without taking into consideration of the difference of data distribution.

$$L(X, y) = -y \log p(Y = 1|X) - (1 - y) \log p(Y = 0|X) \tag{2}$$

Where y is the radiograph label and $y \in \{0, 1\}$.

Considering the unbalanced distribution among different labels, we include a balancing factor that was introduced by Rajpurkar et al. in ChestXnet [6]. The balancing factor transforms the cross entropy into a weighted cross entropy loss.

$$\begin{aligned} &L_{Weighted-Cross-Entropy}(\psi(X, Y) = \\ &\beta_{Pos} \sum_{y=1} -p(Y = 1|X) + \beta_{Neg} \sum_{y=0} -p(Y = 0|X) \end{aligned} \tag{3}$$

Where $p(Y = i|X) \quad i \in \{0, 1\}$ is the probability that the network assigns to the label i, and β_{Pos} is set to $\frac{|P|+|N|}{|P|}$ while β_{Neg} is set to $\frac{|P|+|N|}{|N|}$, with $|P|$ and $|N|$ denoting the total number positive labels ($y = 1$) and negative labels ($y = 0$) in a batch of image labels respectively.

We relabel each chest x-ray image with a 9-dimension vector $li = [l_{i1}, l_{i2}, ..., l_{iC}]$ in which $l_{ic} \in 0, 1$ and $C = 9$. $l_c = 1$ denotes there is pathology findings while $l_c = 0$ denotes that no pathology is founded. The last element of L represents the label with "No Finding".

4 CLASSIFIER ARCHITECTURE

Transfer Learning is utilized to train the thoracic disease classification model to save time and resources. The modern neural network usually has millions of parameters. Usually, training a complex neural network model requires a huge amount of computing resources such as RAM and CPU/GPU if it is trained. Since the High-Performance Computing Center available could only provide 120 slots with 196 Gigabytes RAM, transfer learning is used to reduce the amount of training resource required and shorten computing time. The reason behind this achievement is that transfer learning could take a piece of a model that has already been trained on a related task and reusing it in a new model

Adam (short for Adaptive Moment Estimation) optimizer is used in model training. In Adam, running averages of both the gradients and the second moments of the gradients are used. Given parameters $w^{(t)}$ and a loss function $L^{(t)}$, where t indexes the current training iteration (indexed at 0, Adam's parameter update is given by [2].

[1] "Atelectasis", "Cardiomegaly","Effusion", "Infiltration", "Mass","Nodule", "Pneumonia","Pneumothorax"

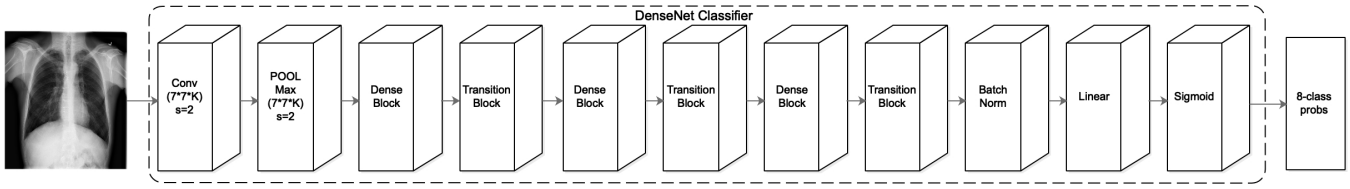

Figure 4: Thoracic disease classification framework.

$$
\begin{aligned}
m_t &= \beta_1 m_{t-1} + (1-\beta_1)\, g_t \\
v_t &= \beta_2 v_{t-1} + (1-\beta_2)\, g_t^2 \\
\hat{m}_t &= \frac{m_t}{1-\beta_1^t} \\
\hat{v}_{t+1} &= \Theta_t - \frac{\alpha}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t
\end{aligned}
\tag{4}
$$

where ϵ is a small scalar used to prevent division by 0, and β_1 and β_2 are the forgetting factors for gradients and second moments of gradients, respectively. Squaring and square-rooting are done elementwise.

Thoracic disease classifier is based on a 121-layer Dense Convolutional Network (DenseNet) [1], which can make the optimization of very deep networks converge since it improve gradient flows. The framework is shown in Figure 4. Each layer has direct access to the gradients from the loss function and the original input signal, leading to an implicit deep supervision [1]. The network structure of thoracic disease classifier is presented in Figure 4. We modify the last two layers to fit into our application by replacing the final fully connected layer with the one that has a single output, and applying a sigmoid activation function to introduce non-linearity.

The weight of thoracic disease classifier is initialized from the pre-trained DenseNet [1]. The input pictures were tiled to 3 channel, resized to 224 and normalized with mean [0.485, 0.456, 0.406], standard deviation [0.229, 0.224, 0.225] on three channels. The network is trained using Adam with standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) starting from 0.0001 learning rate.

4.1 Dataset and Data Process

We utilize the **NIH ChestX-ray14 Dataset** [8] to verify the performance of the proposed thoracic classifier. ChestX-ray14 is the largest publicly available radiograph dataset which has 112,120 chest X-ray images with disease labels collected from 30,805 different patients. The total size of the dataset is 42 GB and each image has 1024 by 1024 pixels. The labels accuracy are expected to be larger than 90% and are suitable for supervised classification and localization of Common Thorax Diseases. There are 15 classes (14 diseases, and one for "No findings"). The 14 diseases are Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural thickening, Cardiomegaly, Nodule Mass, and Hernia [8].

Radiographs with any one of the eight classes *Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax* are extracted from the whole dataset. Also, the radiographs are further processed by data augmentation, including image processing techniques such as Methodology horizontal flipping, shifting the image, rotating the image and center cropping. Moreover, before an image is fed into the network, the pixel value in the image array is divided by 255. This

All experiments are run on the CPU cluster (Quanah) in the High-Performance Computing Center (HPCC), with 200GB RAM and 120 slots.

4.2 Localization

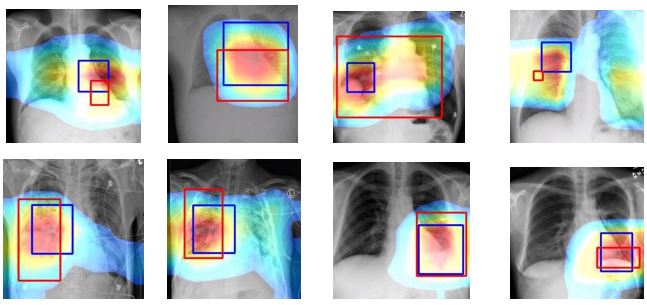

Figure 5: Localization for different diseases.

We applied class-activation maps (CAM, grad-CAM) on models to assist in predicting the bounding boxes [11]. Localization can be easily achieved with the global average layer and the dense layer in the very last part of the model if they exist. By performing a linear combination on the maps before entering GAP, heat-maps are obtained and localization predictions are made according to it. After calculating the heat map, we select global peak value, scaled by a factor of 0.9, as the threshold to select local maximum points with intensity high enough. We apply a maximum filter and a minimum filter to each heat map, and calculate the difference between resulting heat maps to get probable regions with local maximal centroids. Then, for all the local maxima greater than the threshold in the image, we applied dilation on them to accumulate multiple candidate points, and choose the centroid of the accumulated components as the center of a predicted bounding box.

Also, we utilized a fixed size bounding box for each disease, since we have observed that the same disease tends to have a similar box size. The localization results are shown in Figure 5. For each image with certain prediction and its corresponding heat map, we first construct a box covering every local maximal centroid whose boundaries are not lower than a given threshold. The box size is then calculated as the average of all these boxes, for each distinct disease. Thresholds for each class is initially set as 0.88, and decreases by 0.02 on every experiment to get the threshold producing boxes with

Pathology	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax
Wang et al.	0.716	0.807	0.784	0.609	0.706	0.671	0.633	0.806
Yao et al.	0.772	0.904	0.859	0.695	0.792	0.717	0.713	0.841
CheXNet	0.809	0.925	0.864	0.735	0.868	0.780	0.768	0.889
Thoracic Classifier	0.701	0.837	0.778	0.604	0.578	0.718	0.649	0.772

Table 1: Classification results of eight diseases.

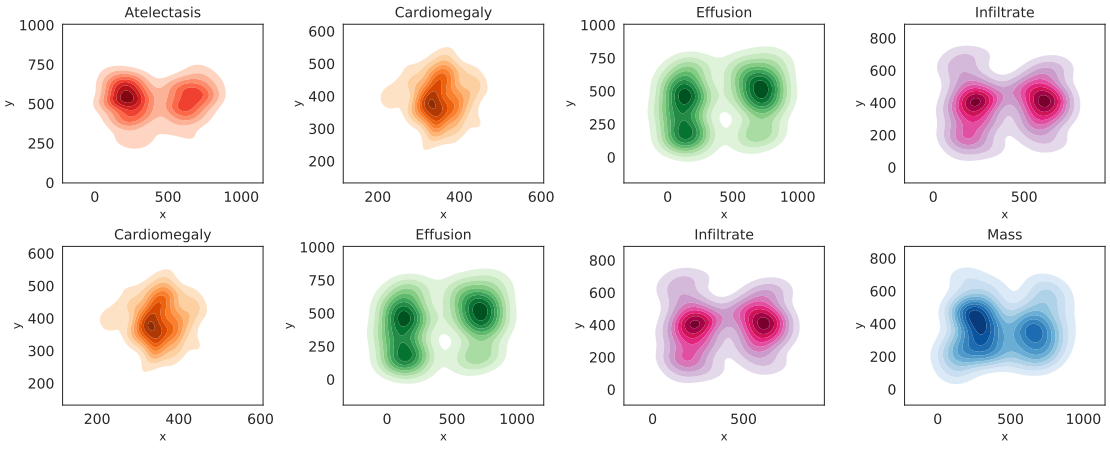

Figure 6: The density distribution of 8 thoracic disease according to the coordinates output by CAM.

a size resembling ones in the validation set the most. And then we applied the mean width and height to each predicted centroid for each distinct disease.

4.3 Using pertained weights vs train from scratch

With ImageNet pre-trained weight	0.705
Without pre-train weigh	0.686

Table 2: Average classification results comparison between models with pre-trained weight and without pre-trained weight.

Table 2 shows two different classification results about whether make use of pre-trained weights. We therefore did some experiments on those settings. As shown in Table 2, model with pre-trained parameters has a better performance.

Wang et al.	(|P|+|N|)/|P|	(|P|+|N|)/|N|	0.705
Rajpurkar et al.	|N|/(|P|+|N|)	|P|/(|P|+|N|)	0.601
without weight	-	-	0.694

Table 3: Comparison of models with different balancing factors.

Also, different weight balancing factors on loss function also influence the classification performance. Table 3 shows the results for different balancing factors.

4.4 Results

We applied the thoracic disease classifier on 8 diseases and presented the result in Table 1. As shown in this table, the proposed classifier could achieve stable performance. For localization, we showed the distrition in Figure 6. The density distribution is consistent with Figure **??**, which depicts the disease coordinate distribution on radiograph produced by the radiologists labeled images.

5 CONCLUSIONS

In this paper, we model the thoracic disease classification task using convolutional neural network and customized the specialty of Chest X-ray dataset. Also, the localization of pathology is privided in heatmap. The experimental results showed that the proposed model

can achieve a good performance. We conclude that: (1).CNN with Transfer learning can classify the X-ray images with an accuracy of around 70.5% based on this dataset; (2).Like pneumonia, other thoracic disease detection and localization can also be automated.

REFERENCES

[1] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.

[2] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[3] Zhe Li, Chong Wang, Mei Han, Yuan Xue, Wei Wei, Li-Jia Li, and Li Fei-Fei. 2018. Thoracic disease identification and localization with limited supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8290–8299.

[4] Rakesh Lodha, Sushil K Kabra, and Ravindra M Pandey. 2013. Antibiotics for community-acquired pneumonia in children. *Cochrane Database of Systematic Reviews* 6 (2013).

[5] Amanda E Nelson, Jordan B Renner, Xiaoyan A Shi, Jack H Shreffler, Todd A Schwartz, and Joanne M Jordan. 2010. Cross-sectional comparison of extended anteroposterior and posteroanterior fixed flexion positioning to assess radiographic osteoarthritis at the knee: The Johnston County Osteoarthritis Project. *Arthritis care & research* 62, 9 (2010), 1342–1345.

[6] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. 2017. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225* (2017).

[7] Olli Ruuskanen, Elina Lahti, Lance C Jennings, and David R Murdoch. 2011. Viral pneumonia. *The Lancet* 377, 9773 (2011), 1264–1275.

[8] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2097–2106.

[9] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, and Ronald M Summers. 2018. Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 9049–9058.

[10] Li Yao, Eric Poblenz, Dmitry Dagunts, Ben Covington, Devon Bernard, and Kevin Lyman. 2017. Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv preprint arXiv:1710.10501* (2017).

[11] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2921–2929.