

VIETNAM NATIONAL UNIVERSITY – HO CHI MINH CITY
THE INTERNATIONAL UNIVERSITY
BIOMEDICAL ENGINEERING DEPARTMENT



Mimics state of the art deep learning to diagnose common thorax disease

by

Dat Quang Tran

A pre-thesis submitted to the Biomedical Engineering Department in partial fulfillment of the
requirements for the degree of Engineer

HoChiMinh city, Vietnam

June 2018



**Mimics state of the art deep learning to diagnose
common thorax disease**

APPROVED BY: Advisor

APPROVED BY: Committee

Pham Thi Thu Hien, PhD



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my advisor Dr. Pham Thi Thu Hien for the continuous support of my study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my undergraduate study.

My sincere thanks also go to Tuan Hue Thi and his team in EyeQ, who provided me an opportunity to join their team and gave access to the laboratory and research facilities. Without their precious support it would not be possible to conduct this research.

Last but not the least, I would like to thank my family for supporting me spiritually throughout this pre-thesis.



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

TABLE CONTENT

Table of Contents

<i>LIST OF TABLES.....</i>	5
<i>LIST OF FIGURES.....</i>	6
<i>Abstract.....</i>	8
<i>Chapter 1: Introduction</i>	9
1.1 Problem statement	9
1.2 Organization of Report.....	9
<i>Chapter 2: Literature Review.....</i>	11
<i>Chapter 3: Methodology</i>	12
3.1 Dataset.....	12
3.2 Loss function	14
3.3 CNN architecture.....	15
3.4 Pooling layer	18
3.5 Training setting.....	18
3.6 Heatmap Localization.....	19
<i>Chapter 4: Result</i>	21
4.1 Experimental Result	21
4.2 Compare to current SOTA.....	26
4.3 CAM Heat map.....	29
<i>Chapter 5: Conclusion</i>	31
<i>Reference.....</i>	32



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

LIST OF TABLES

Table 1: Evaluation of different model on ChestX-ray 14 dataset, the performance is reported using AUC score.....	27
---	----



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

LIST OF FIGURES

Figure 1: From left to right: low quality image, not chest X-ray, inverted color, children.	12
Figure 2: Left: original image, Right: image processed with histogram.	13
Figure 3: The circular diagram shows the proportions of images with multi-labels in each of 14 pathology classes and.....	14
Figure 4: MNI ST hand-writing number classification [14] using the traditional convolution architecture.....	16
Figure 5: the training error (left) and test error (right) on CIFAR-10 using 20-layer and 56-layer traditional network. The performance degraded as the network get deeper.	16
Figure 6: the “residual” connection.	17
Figure 7: block of convolution layer with result concatenated	17
Figure 8: The overall flowchart of CNN model for disease classification and localization.	20
Figure 9: The Loss of different learning rate over 100 epochs. The solid line represents training phase and dash line represents validation phase.....	21
Figure 10: The AUC of different learning rate over 100 epochs. The solid line represents training phase and dash line represents validation phase.....	22
Figure 11: The Loss of different batch size over 100 epochs. The solid line represents training phase and dash line represents validation phase.....	23



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

Figure 12: The AUC of different batch size over 100 epochs. The solid line represents training phase and dash line represents validation phase.....	24
Figure 13: The progress of Loss over different epoch size. The solid line represents training phase and dash line represents validation phase.....	25
Figure 14: The progress of AUC over different epoch size. The solid line represents training phase and dash line represents validation phase.....	26
Figure 15: Patient with cardiomegaly (enlarged heart).....	29
Figure 16: Patient with right pleural effusion (fluid in pleural space).....	29
Figure 17: Patient with right side Pneumothorax (collapsed lung).	30



Abstract

Lung diseases are some of the most common medical condition in the world. In the U.K alone, there is somebody dies from lung disease every 5 minutes and about 10,000 people are newly diagnosed with lung disease every week.

Chest X-ray is one of the most common medical imaging and can be used to diagnose a wide range of lung related disease. In this proposed study we implement the basis CAD (Computer Aided Diagnostic) to predict the probabilities of 14 different lung diseases (i.e. Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening and Hernia). The model is build using over 112,000 images from ChestX-ray 14 dataset, currently the largest publicly available dataset on the subject.

In this study, by using the DenseNet model, we have successfully trained a Deep Learning network that can predict the probabilities of each 14 lung diseases with average Area Under ROC curve of 0.841. We also produce a localization heatmap on the lung X-ray that can greatly assist the doctor to make better decision.

Keywords: ChestX-ray 14, Deep learning, DenseNet, Area Under ROC, CAM Heatmap.



Chapter 1: Introduction

1.1 Problem statement

Many lung diseases like pneumonia, effusion are among the major cause of death in many countries. In Vietnam, 10% of patient who has lung disease is diagnosed of pneumonia. Among many methods to diagnose lung diseases such as flow cytometer, mucus examination, chest Xray stand out as an effective and inexpensive tool accurately points out many respiratory diseases. In Vietnam today there are many lung imaging study along with radiological reports that are generated daily in both public and private sector. This raise to the question if we can develop a computational system that can help doctor to diagnose more accurately.

Alongside with chest Xray, biomedical engineers around the world have developed many CAD system to assist doctor in many medical related tasks. An example of such system that have been deployed in real world situation is a CAD system that can detect tiny lesion from an image by 12sigma. A team from Google has successfully developed a microscopic software that can detect cancer cell and segment them out from the tissue sample in real time.

1.2 Organization of Report

The remaining of this report is divided into the following sections:



Vietnam National Universities – HCMC International University Biomedical Engineering Department

Chapter 2: Related Works will show technique researcher has used, from wavelet transform to the currently breakthrough Deep Learning

Chapter 3: Methodology will show how the problems is tackle using the Deep Learning framework

Chapter 4: Results will show and compare my result with other related papers

Chapter 5: Conclusion give summation of the project and some further development ideas



Chapter 2: Literature Review

Prior to the current advanced technique of deep learning, researchers have many different approaches to implement the Computer Aid Diagnostics system to detect different diseases from Chest X-ray (CXR) images. [1] used Haar wavelet transform to extract lung feature image and weighted nearest neighbors based on Euclidean distance to measure similarity between images. In [2] used Fuzzy C-means clustering for feature extraction and classification.

After the winning of AlexNet [3] on ImageNet in 2012, the majority of research on medical imaging has shifted toward deep learning approach: [4] used deep learning to detect diabetic retinopathy, [5] address the problem of skin cancer classification.

Automated diagnosis from CXR also draw a lot of attention from research community. [6] used Recurrent Neural Network (RNN) to automatically annotate CXR image using publicly available OpenI dataset [7]. Using the same dataset, [8] combined several Convolution Neural Network (CNN) using ensemble methods on localization and classification task. [9] ChestX-ray 14, a largest CXR publicly available on the Internet, and also provide a baseline result using pretrain model from ImageNet. [10] uses a novel DenseNet architecture [11] and outperforms other models on ChestX-ray 14, which they call CheXNet. Later on, [12] build a model that exploits the dependencies between labels. Although [12] can't beat CheXNet in term of performance metrics, but their method proposes an interesting direction for other researchers to explore.



Chapter 3: Methodology

3.1 Dataset

[9] ChestX-ray 14 dataset consists of 112,120 frontal-view X-ray images of 30,805 unique patients. The images are mined from NIH's Picture Archiving and Communication System. Each image is labelled with up to 14 different thorax pathology using extracted features from radiology report. The dataset is then split into training (78468 images, 21528 patients), validation (11219 images, 3090 patients) and testing (22433 images, 6187 patients) with no patients overlap between each set.

The images in the dataset has several problems:

- Low image quality
- Inverted color
- Chest X-ray of children
- Not chest X-ray



Figure 1: From left to right: low quality image, not chest X-ray, inverted color, children.



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

All of these images are removed.

Before input into the model, each image is going through the following steps to make the model more robust to change in rotation, scaling and illumination:

- Follow the practice in [13], we apply histogram equalization to make relevant information to standout. This image processing steps use python PIL image library
- Resize to 256x256
- Randomly crop the image and resize to 224x224, this step will introduce scaling noise
- Add color jitter, this step will introduce illumination noise
- Normalize the image using ImageNet mean and standard derivative

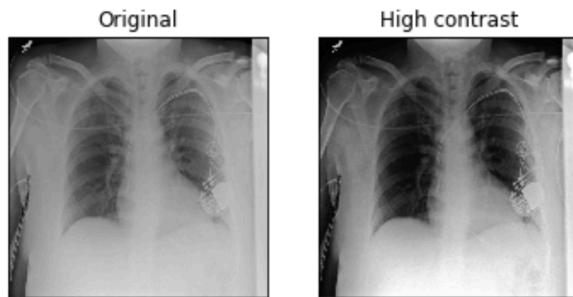


Figure 2: Left: original image, right: image processed with histogram equalization.

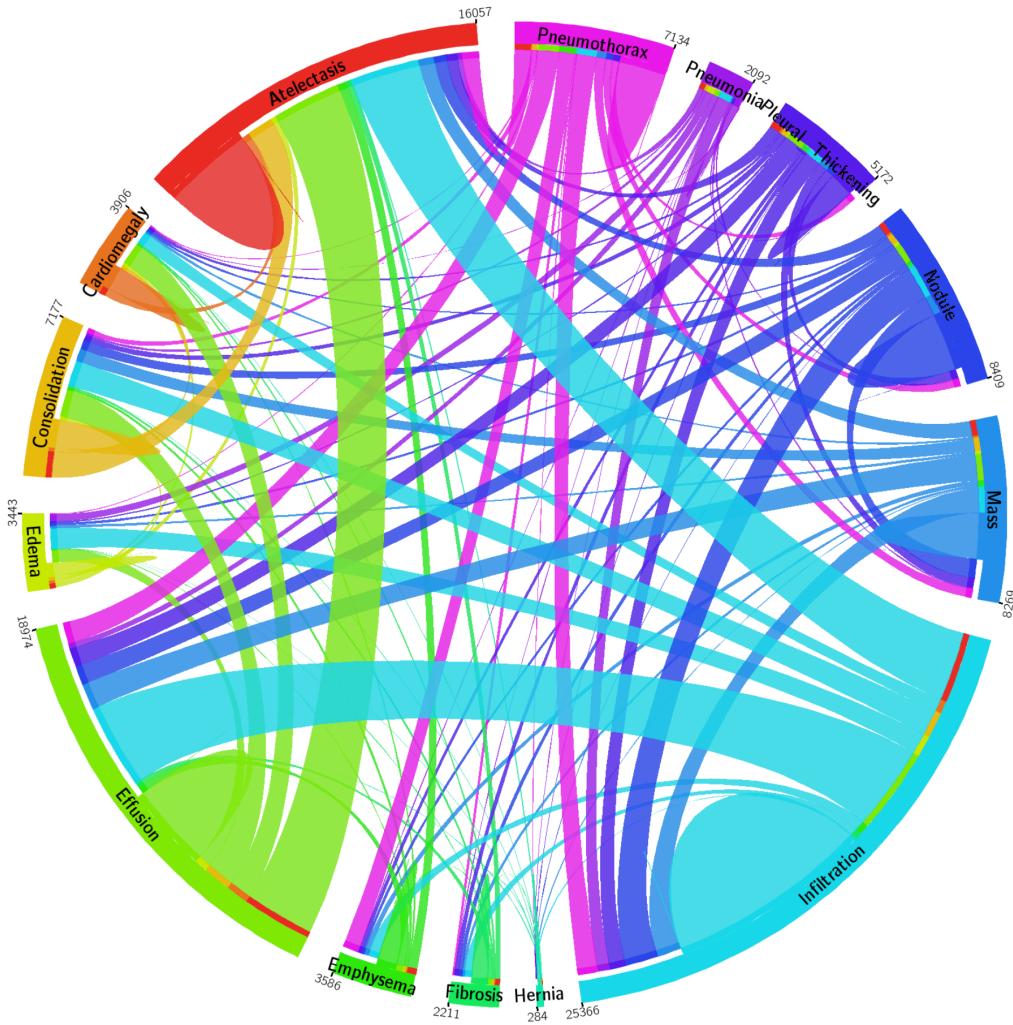


Figure 3: The circular diagram shows the proportions of images with multi-labels in each of 14 pathology classes and the labels' co-occurrence statistics.

3.2 Loss function

To represent the multi-label, we choose a 14-dimension label vector $y = [y_1, y_2, \dots, y_c, \dots, y_{14}]$ where $y_c = 1$ indicates the represent of pathology c and $y_c = 0$ otherwise, $C = 14$ is all possible pathology in the dataset. This means an all-zero vector $y = [0, 0, \dots, 0]$ represent “Normal” status



Vietnam National Universities – HCMC International University Biomedical Engineering Department

(no pathology in the scope of 14 disease is found in this sample). This setting transits the multi-label classification problem into 14 independent binary classification problems.

Using the above definition, the loss function is equal to the total binary cross entropy loss for each possible pathology. And since the image labels are rather spare (i.e., more ‘0’s than ‘1’s), the positive/negative balancing factor β_P, β_N is introduced to enforce the learning of positive labels.

Equation 1

$$L(p, y) = \beta_P \sum_{c=1, y_c=1}^{c=14} -\ln(p) + \beta_N \sum_{nc=1, y_c=0}^{c=14} -\ln(1-p)$$

Where:

p: The probability of $y=1$, output of the sigmoid function.

$\beta_P = \frac{|P|+|N|}{|P|}$ is positive balancing factor, $\beta_N = \frac{|P|+|N|}{|N|}$ is negative balancing factor where $|P|$ and $|N|$ is the total number of ‘1’s and ‘0’s in a batch of image labels.

3.3 CNN architecture

In a traditional convolution neural network (CNN), the input image usually go through a sequences of convolution and pooling to squeeze the width and height dimension white increase the feature depth. The purpose of these stacked convolution and pooling layer is to learn the representation of the image, or features. The learned featured will be then the input to the fully connected layers for classification. In many early architectures, people have tried to stack more and more convolution



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

and pooling layers together, hence the term Deep Learning, with a assumption that deeper network will learn better representation. However, in practice this design suffers from vanishing gradient problem and the performance degraded as the network get deeper

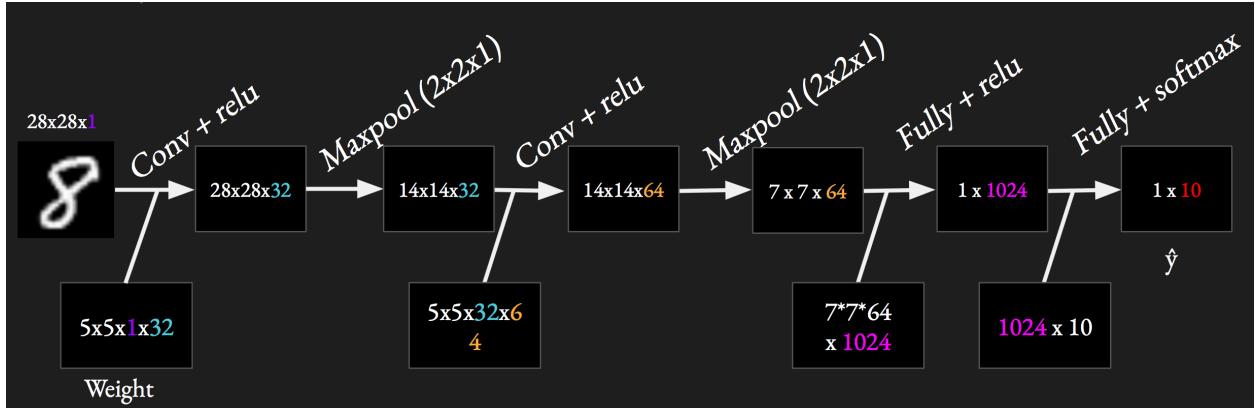


Figure 4: MNI ST hand-writing number classification [14] using the traditional convolution architecture.

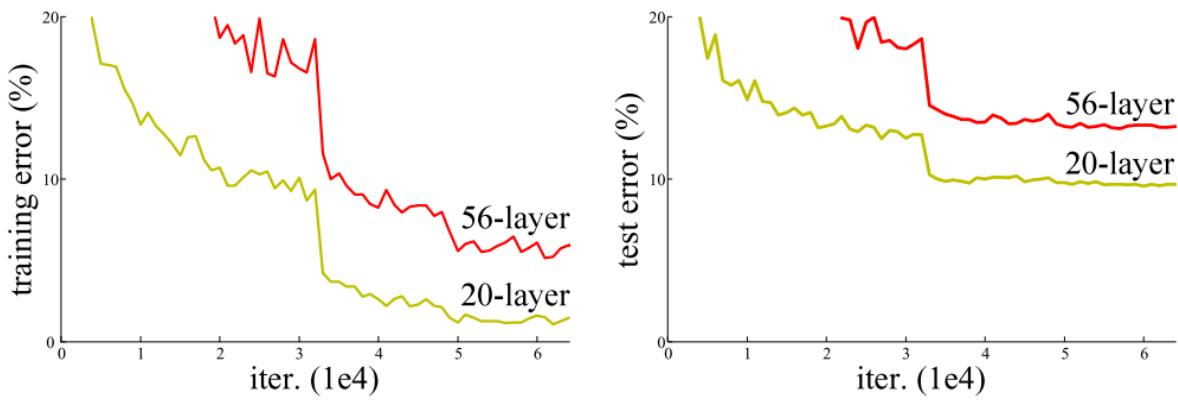


Figure 5: the training error (left) and test error (right) on CIFAR-10 using 20-layer and 56-layer traditional network. The performance degraded as the network get deeper.

[15] proposed a revolutionized idea: instead of learning direct mapping from x to $H(x)$ in convolution layer, let learn the difference, the “residual” between them. So let say $F(x) = H(x) - x$, the network is now trying to learn $x + F(x)$ instead of direct mapping $H(x)$

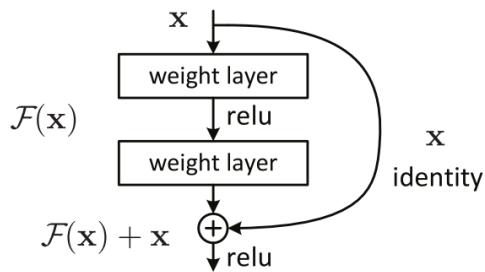


Figure 6: the “residual” connection.

This “residual” idea has made training deeper network more easily, researchers have tried network with 100, 200 or even 1000 layer deep. [11] put this idea a litter bit further by concatenating the output of previous layer instead of summation.

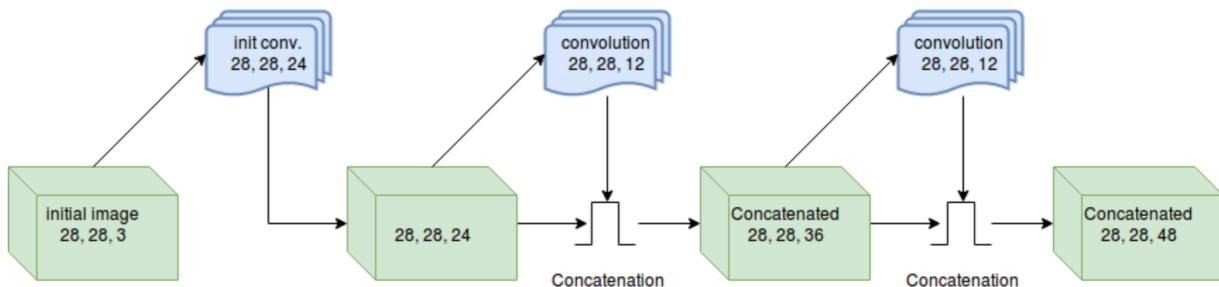


Figure 7: block of convolution layer with result concatenated



Vietnam National Universities – HCMC International University Biomedical Engineering Department

This concatenated setting improves the flow of gradients when backpropagation. This is because of concatenate operator, every transition layer has direct access to gradient from the last layer, leading to a much deeper network.

3.4 Pooling layer

The pooling layer plays an important role of choosing what information to pass down to next layer. Beside conventional max and average pooling, Log-Sum-Exp (LSE) pooling is also be used.

Equation 2

$$x_p = x_{\max} + \frac{1}{r} \log \left[\frac{1}{S} \sum_{(i,j) \in S} \exp[r(x_{ij} - x_{\max})] \right]$$

Where:

s: pooling kernel size

S = s*s: total number of pixel inside kernel

r: Hyper-parameter, by controlling r, the pool ranges from maximum in S ($r \rightarrow \infty$) to average ($r \rightarrow 0$). LSE pooling is an adjustable option between max and average pooling

3.5 Training setting

In order to train a good CNN, there are many hyper-parameters to control. After doing many experiments, the following hyper-parameters are chosen for the final model:



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

- DenseNet121 with pretrained weight from ImageNet
- [16] Adam optimizer with default hyper-parameter ($\beta_1 = 0.9, \beta_2 = 0.999$).
- Initial learning rate = 0.0001 with decay by 10 scheduler
- Epochs = 100, batches = 500, batch size = 32

3.6 Heatmap Localization

To generate the localization heatmap, we use the method of Class Activation Map (CAM) which is proposed by [17]. To generate CAM image, a image is first feed into the network and a feature map of last convolution layer is extracted. This feature map is then multiply with the weight of the desired class from the last fully connected layer and added altogether to produce an activation map. The activation map is then blended with the original image to produce the final localized heatmap.

The class activation map is given by the following formula

Equation 3

$$M_c = \sum_k \omega_k^c f_k$$

Where:

M_c activation map for disease c

Number of feature map k, k= 1024 for our DenseNet121

ω_k^c weight of last fully connected layer that match the k-th feature to the c-th class

f_k feature map at k

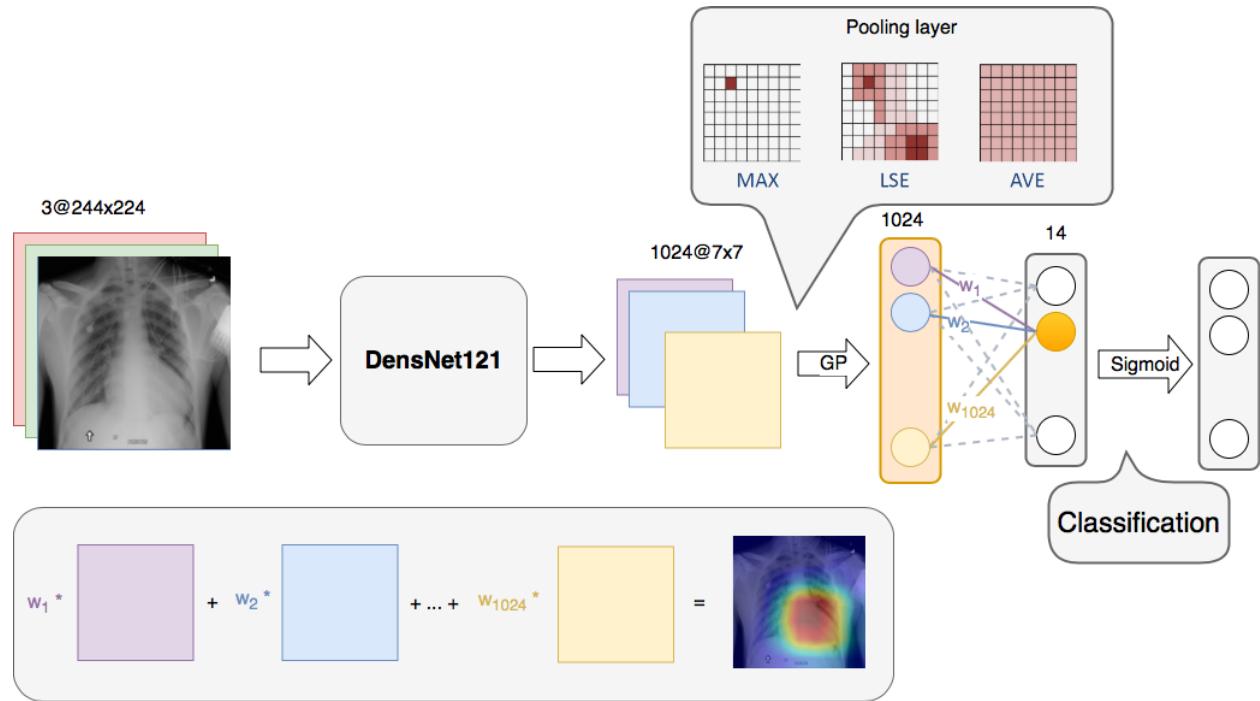


Figure 8: The overall flowchart of CNN model for disease classification and localization.

Chapter 4: Result

4.1 Experimental Result

4.1.1 Different Learning Epochs

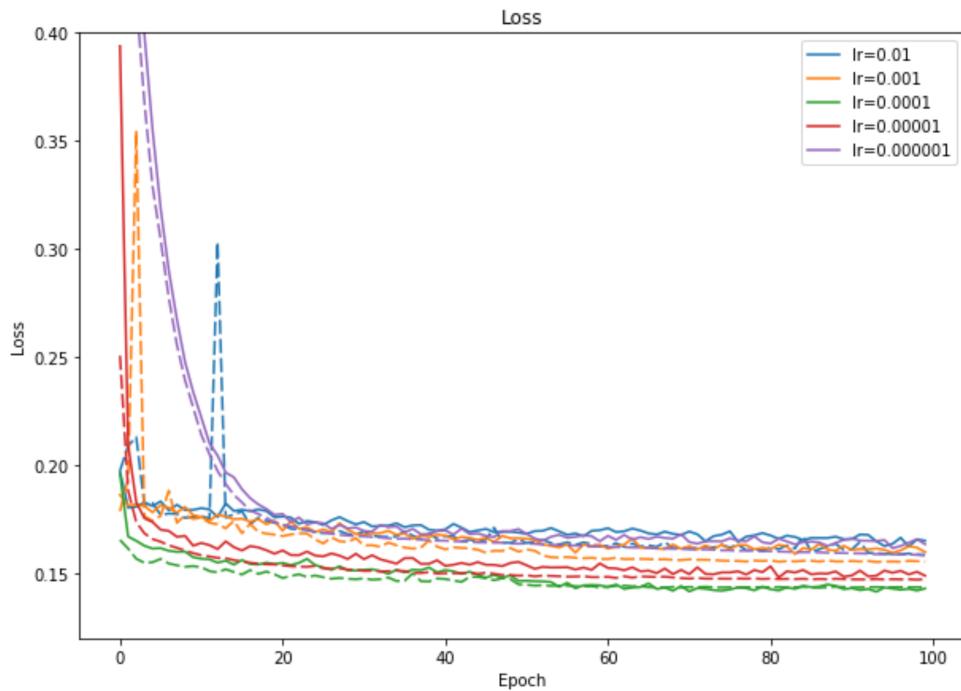


Figure 9: The Loss of different learning rate over 100 epochs. The solid line represents training phase and dash line represents validation phase.

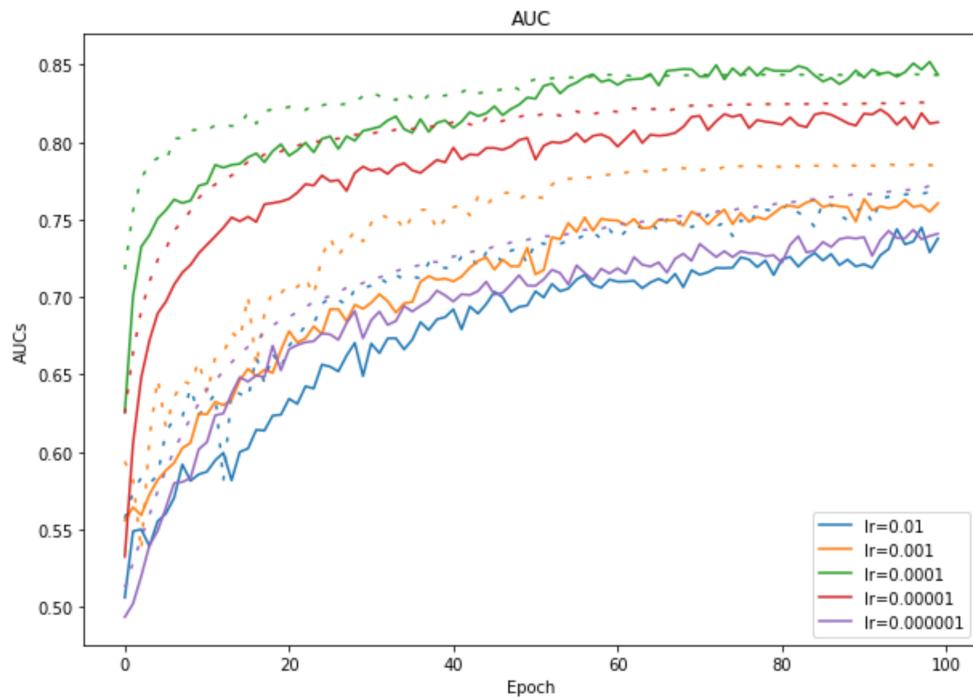


Figure 10: The AUC of different learning rate over 100 epochs. The solid line represents training phase and dash line represents validation phase.

Discussion: If the learning rate is too high (i.e. $lr=0.01$) the loss will be bouncing back and forth and cannot converge. While if the loss is too small (i.e. $lr = 0.0000001$) the loss will converge slowly. In this experiment, I have found that appropriate learning rate (i.e. $lr = 0.0001$) make the training process better

4.1.2 Different Batch Size

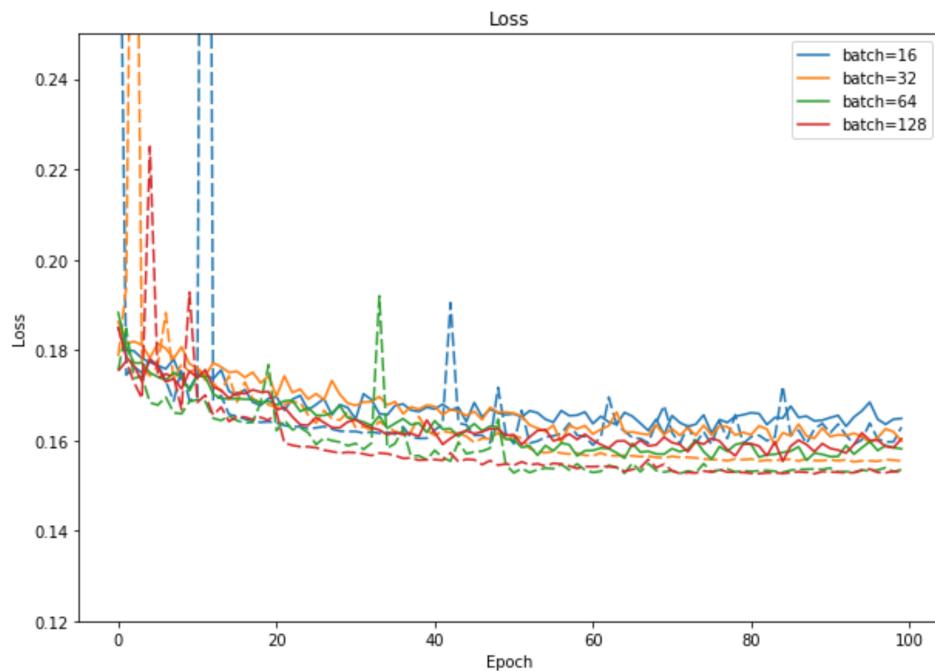


Figure 11: The Loss of different batch size over 100 epochs. The solid line represents training phase and dash line represents validation phase.

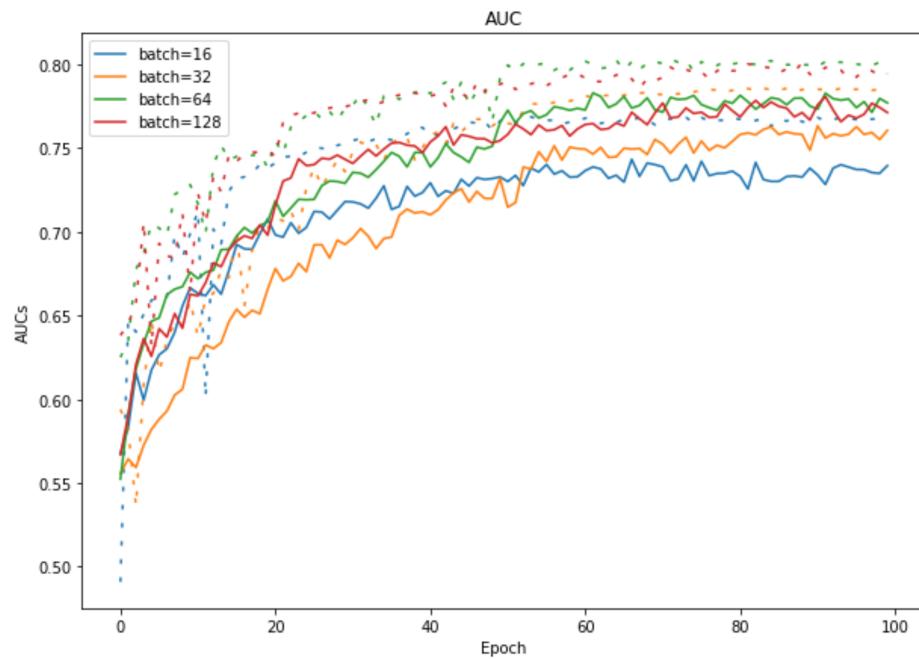


Figure 12: The AUC of different batch size over 100 epochs. The solid line represents training phase and dash line represents validation phase.

Discussion: Smaller batch size add more noise to the as they are not represented for the entire dataset. After doing several experiment on how many images for each batch, I choose 64 images for the best performance gain.

4.1.3 Longer training



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

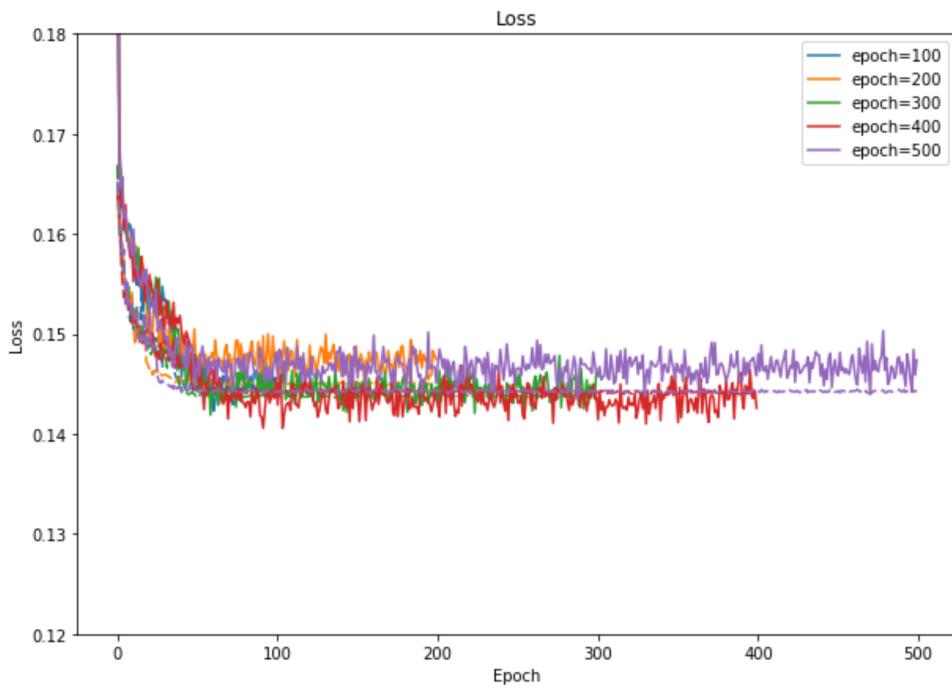


Figure 13: The progress of Loss over different epoch size. The solid line represents training phase and dash line represents validation phase.



Vietnam National Universities – HCMC International University Biomedical Engineering Department

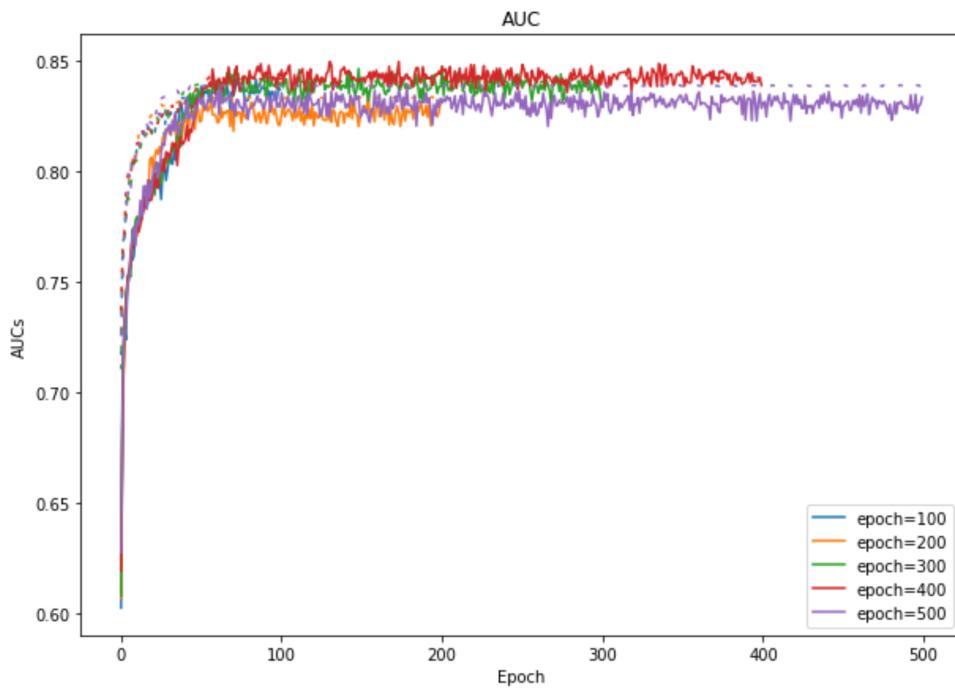


Figure 14: The progress of AUC over different epoch size. The solid line represents training phase and dash line represents validation phase.

Discussion: Can longer training improve the performance of the final model? In this experiment, I train all the model in the same setting with 100 longer epochs each time. The final conclusion is that after around 100 epochs, the performance have no significant improve.

4.2 Compare to current SOTA



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

Table 1: Evaluation of different model on ChestX-ray 14 dataset, the performance is reported using AUC score.

Model	Wang et al.	Yao et al.	Rajpurkar et al	Peking	Microsoft	dattran.xray
Atelectasis	0.716	0.772	0.8094	0.8311	0.828543	0.8238291392
Cardiomegaly	0.807	0.904	0.9248	0.922	0.891449	0.9147935856
Effusion	0.784	0.859	0.8638	0.8891	0.817697	0.8847050351
Infiltration	0.609	0.695	0.7345	0.7146	0.907302	0.710563024
Mass	0.706	0.792	0.8676	0.8627	0.895815	0.8604432331
Nodule	0.6701	0.717	0.7802	0.7883	0.907841	0.792681043
Pneumonia	0.633	0.713	0.768	0.782	0.817601	0.7737050722
Pneumothorax	0.806	0.841	0.8887	0.8844	0.881838	0.8839498983
Consolidation	0.708	0.788	0.7901	0.8148	0.721818	0.8148896447
Edema	0.835	0.882	0.8878	0.8992	0.868002	0.8973931359
Emphysema	0.815	0.829	0.9371	0.9343	0.787202	0.9254988388
Fibrosis	0.769	0.767	0.8047	0.8385	0.826822	0.841977836
Pleural Thickening	0.708	0.765	0.8062	0.7914	0.793416	0.78972166
Hernia	0.767	0.914	0.9164	0.9206	0.889089	0.9361743983
Mean	0.74	0.8	0.841	0.848	0.845	0.846



Vietnam National Universities – HCMC International University Biomedical Engineering Department

Discussion: [9] create the dataset and establish the baseline using ResNet-50 in which we and other teams competed on. [12] using LSTM model with DenseNet backbone to learn the dependencies between different labels from scratch and improve the performance by 0.6-fold. [10] uses pretrain model from ImageNet and improve the AUC by 0.4-fold. Base on that result, team from Peking reimplement with improvement in training setting establish the SOTA, while team from Microsoft clean up the dataset. Base on three previous implementation, my model has successfully achieve very competitive result.

4.3 CAM Heat map

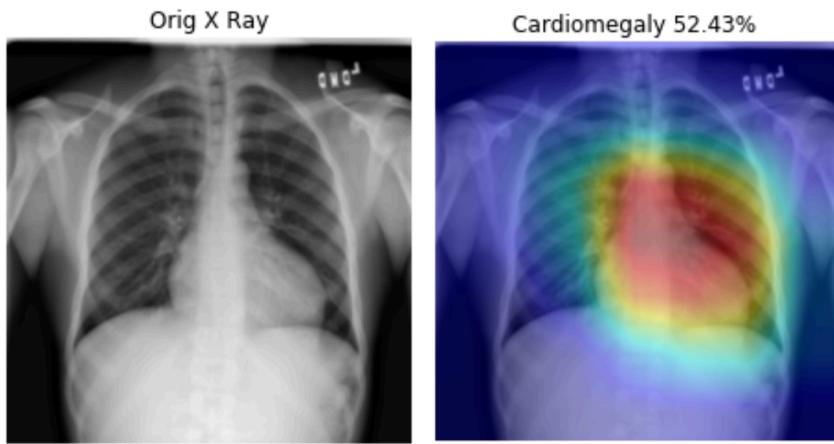


Figure 15: Patient with cardiomegaly (enlarged heart).

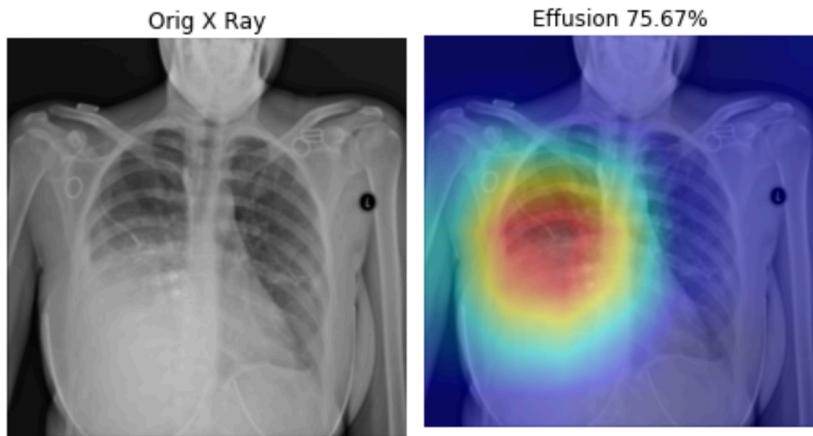


Figure 16: Patient with right pleural effusion (fluid in pleural space).



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

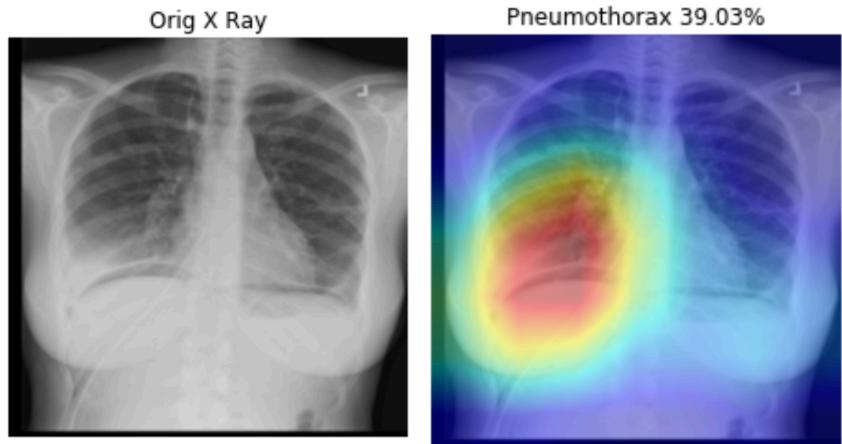


Figure 17: Patient with right side Pneumothorax (collapsed lung).



Chapter 5: Conclusion

Chest X-ray is one of the most common medical imaging data available. It also plays an important role in diagnosis many different lung diseases. The development of a CAD (Computer Aided Diagnosis) which can help the radiologist to make better decision faster is very important. Our thesis aim to create a prediction model that will be the brain of such system. Here in the pre-thesis project, we have successfully reproduce the current State-of-the-Art on the ChestX-ray 14 dataset. However, this is still far from the final destination and many more works need to be done. These are some ideas that can be done to make the model work better and can be applicable in the wild:

- Model the dependencies between labels using LSTM
- Instead of using the whole chest X-ray image, only a chest part is cropped using lung segmentation for downstream process
- Testing the model on real world data



Reference

- [1] L. L. G. Oliveira, S. A. e. Silva, L. H. V. Ribeiro, R. M. de Oliveira, C. J. Coelho, and A. L. S. Ana Lúcia, “Computer-aided diagnosis in chest radiography for detection of childhood pneumonia,” *Int. J. Med. Inform.*, vol. 77, no. 8, pp. 555–564, 2008.
- [2] N. R. S. Parveen and M. M. Sathik, “Detection of Pneumonia in chest X-ray images.,” *J. Xray. Sci. Technol.*, vol. 19, no. 4, pp. 423–8, 2011.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” 2012.
- [4] V. Gulshan *et al.*, “Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs,” *JAMA*, vol. 316, no. 22, p. 2402, Dec. 2016.
- [5] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
- [6] H.-C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, “Learning to Read Chest X-Rays: Recurrent Neural Cascade Model for Automated Image Annotation,” Mar. 2016.
- [7] D. Demner-Fushman *et al.*, “Preparing a collection of radiology examinations for distribution and retrieval,” *J. Am. Med. Informatics Assoc.*, vol. 23, no. 2, pp. 304–310, Mar. 2016.
- [8] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, “Abnormality Detection and



Vietnam National Universities – HCMC

International University

Biomedical Engineering Department

Localization in Chest X-Rays using Deep Convolutional Neural Networks,” 2017.

- [9] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,” May 2017.
- [10] P. Rajpurkar *et al.*, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” Nov. 2017.
- [11] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” Aug. 2016.
- [12] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, “Learning to diagnose from scratch by exploiting dependencies among labels,” Oct. 2017.
- [13] C. Tataru, D. Yi, A. Shenoyas, and A. Ma, “Deep Learning for abnormality detection in Chest X-Ray images,” 2017.
- [14] Y. LeCun, B. Leon, B. Yoshua, and H. Patrick, “Gradient-Based Learning Applied to Document Recognition,” 1998.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015.
- [16] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Dec. 2014.
- [17] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning Deep Features for Discriminative Localization.”