Chest X-ray Thoracic Disease Diagnosis with Convolutional Neural Networks

Team 19 @ 2020 Fall BD4H:
Xiaojie Du, Wenqin You, Jing Zhao, Jingyao Zhu

Github Repo

Motivation

- Clinical diagnosis of thoracic diseases using X-ray images is expensive and requires expertise.

- Build end-to-end deep learning models to facilitate physicians for the clinical diagnosis.

Data Pipeline

Raw Data

API

kaggle

Preprocess

amazon web services™

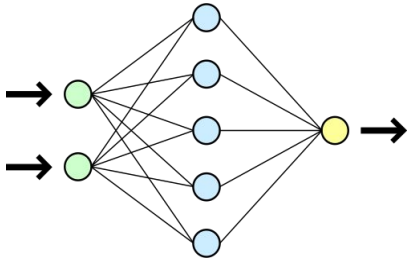
PySpark

python

Framework

Keras

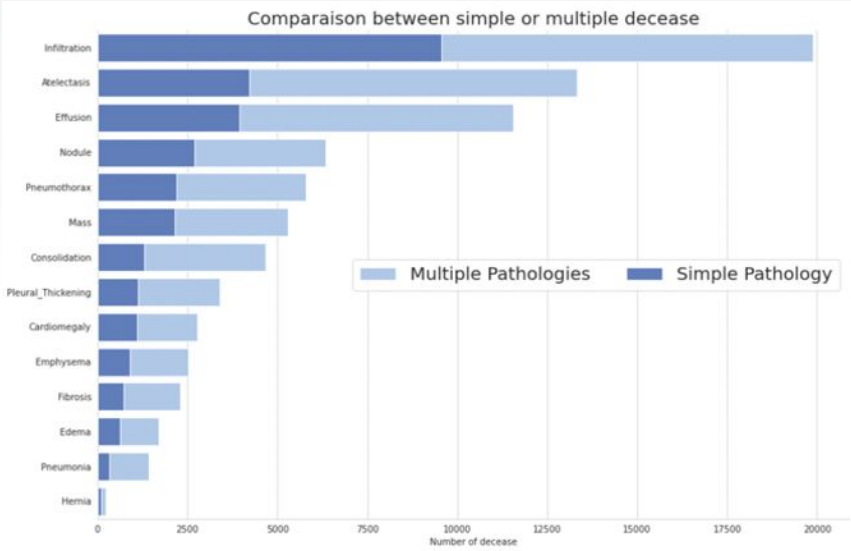
Models



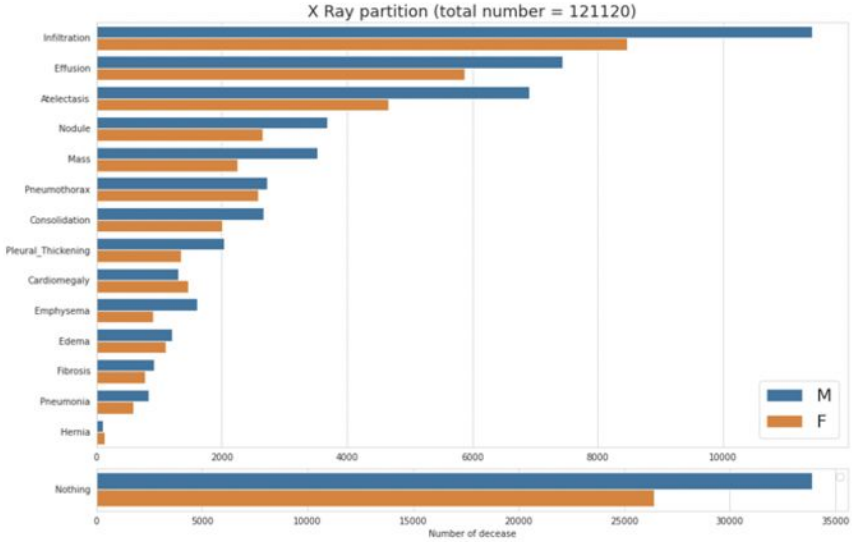
Raw Data

- NIH Chest X-ray Dataset
  - 112,120 Chest X-ray images
  - Bounding box coordinates: BBoxlist2017.csv
  - Class labels and patient data: Dataentry2017.csv

Data Preprocessing

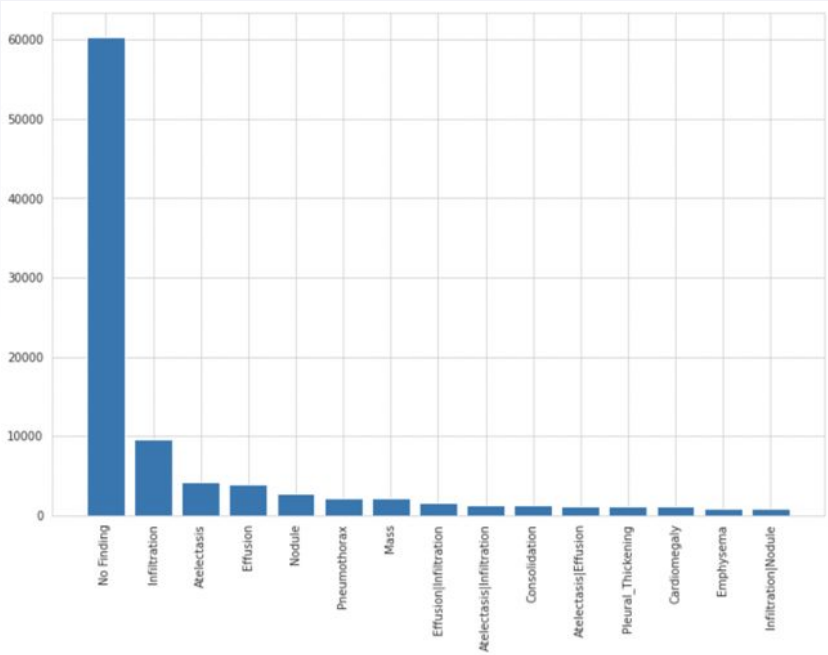


(a)simple or multiple labels per image

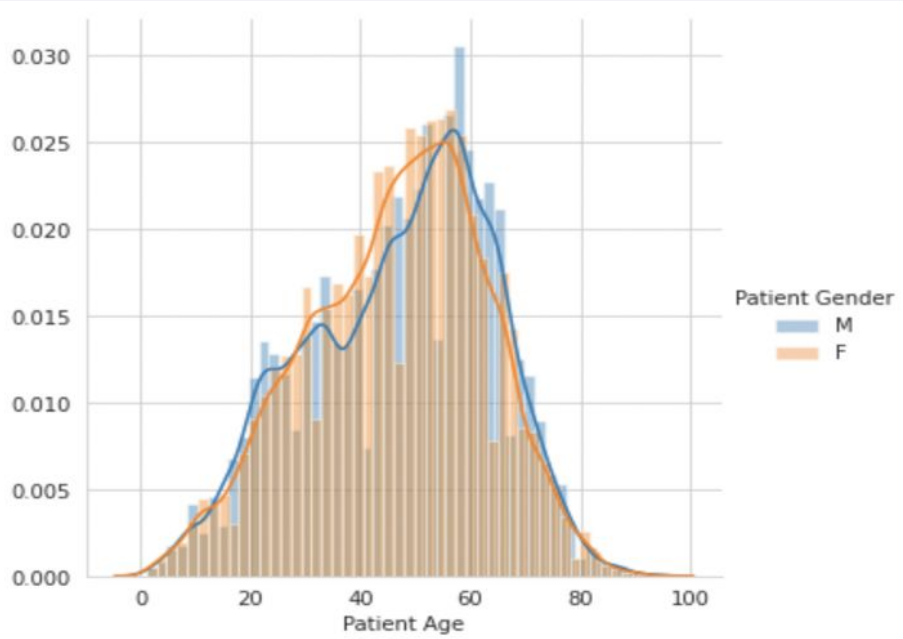


(b) Disease Distribution by Gender

Data Preprocessing

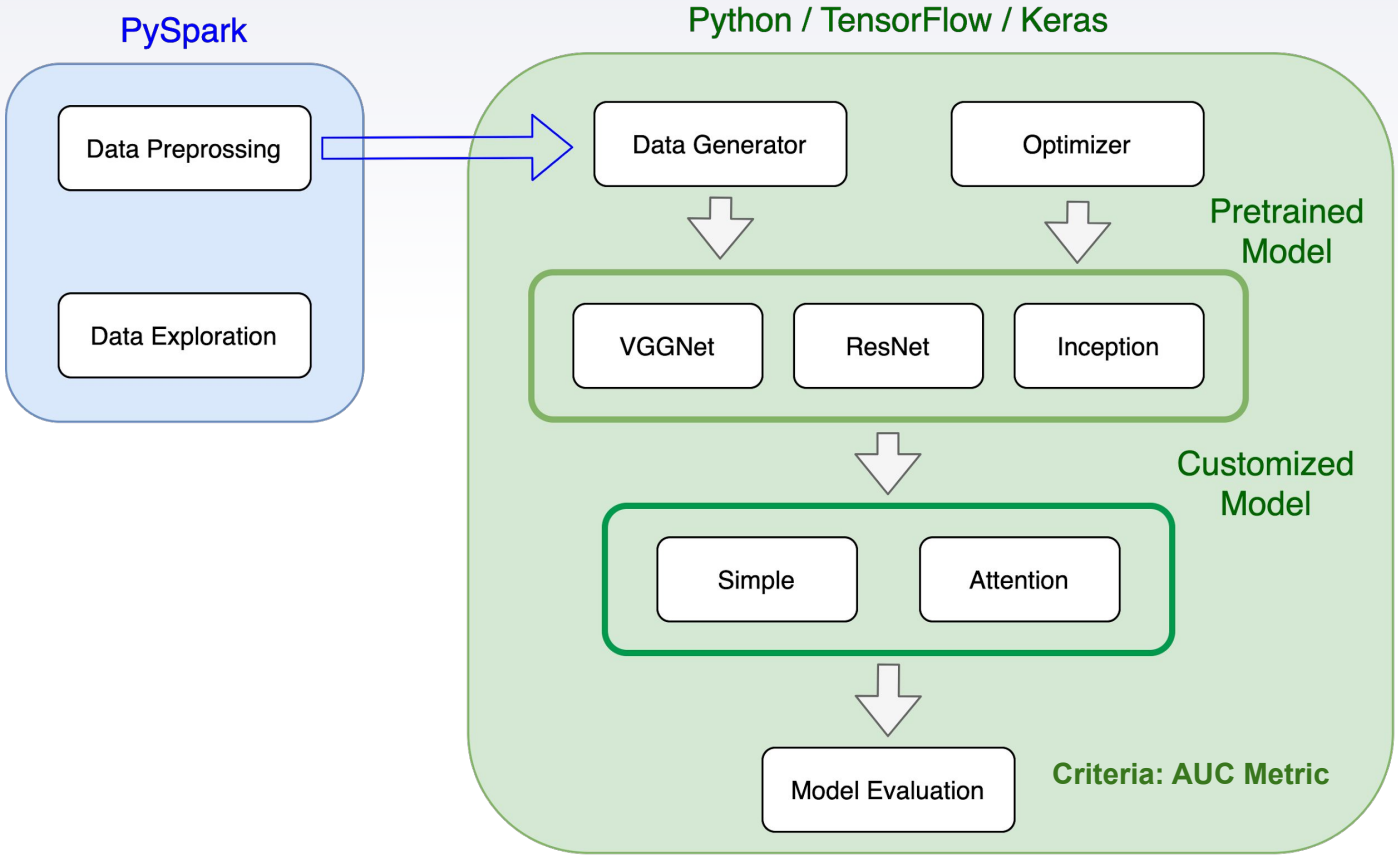


(a) Labels Distribution

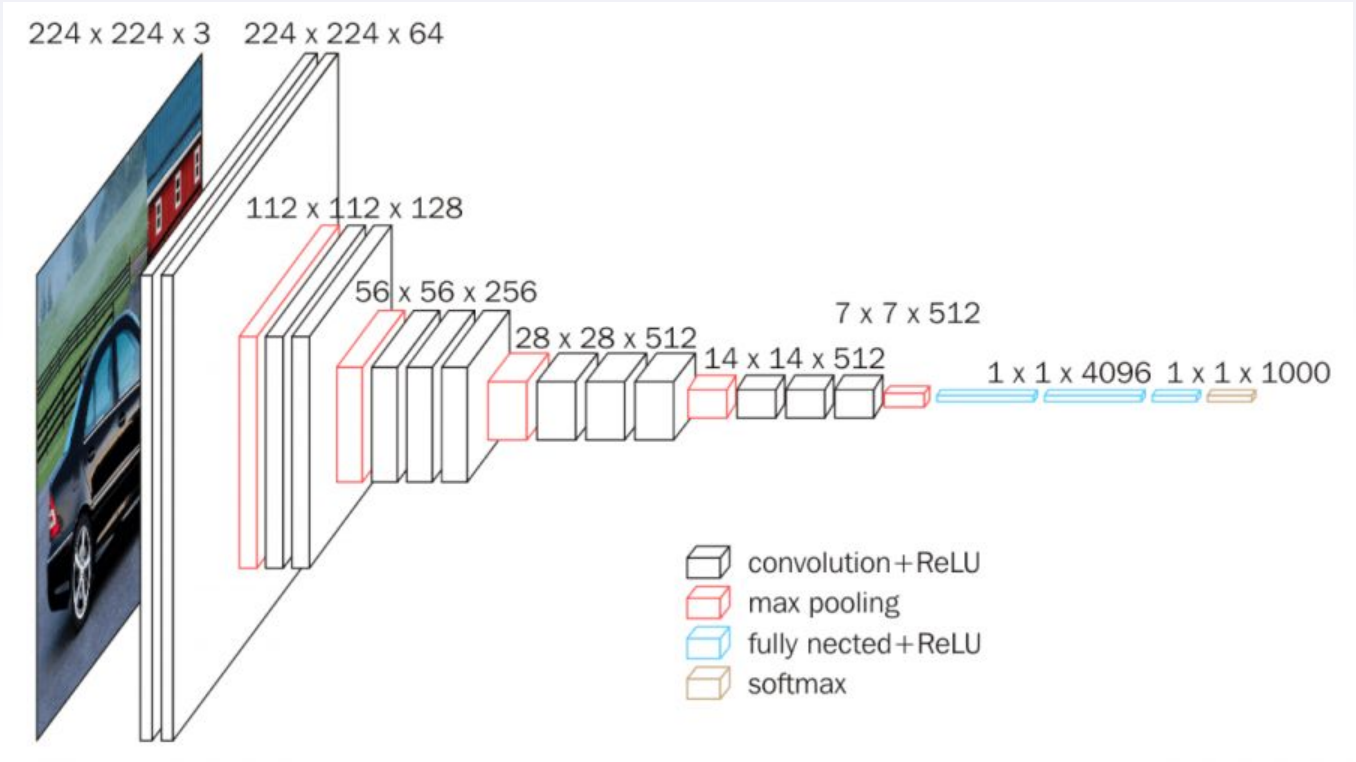


(b) Age Distribution by Gender

Project Pipeline

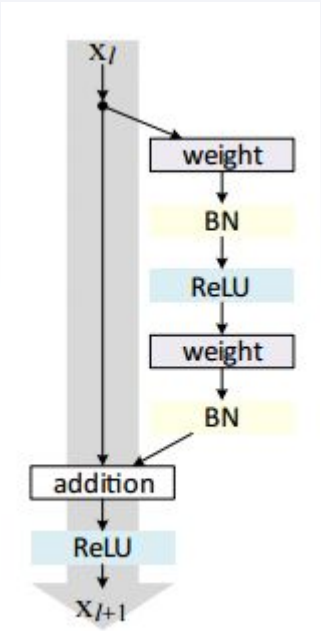


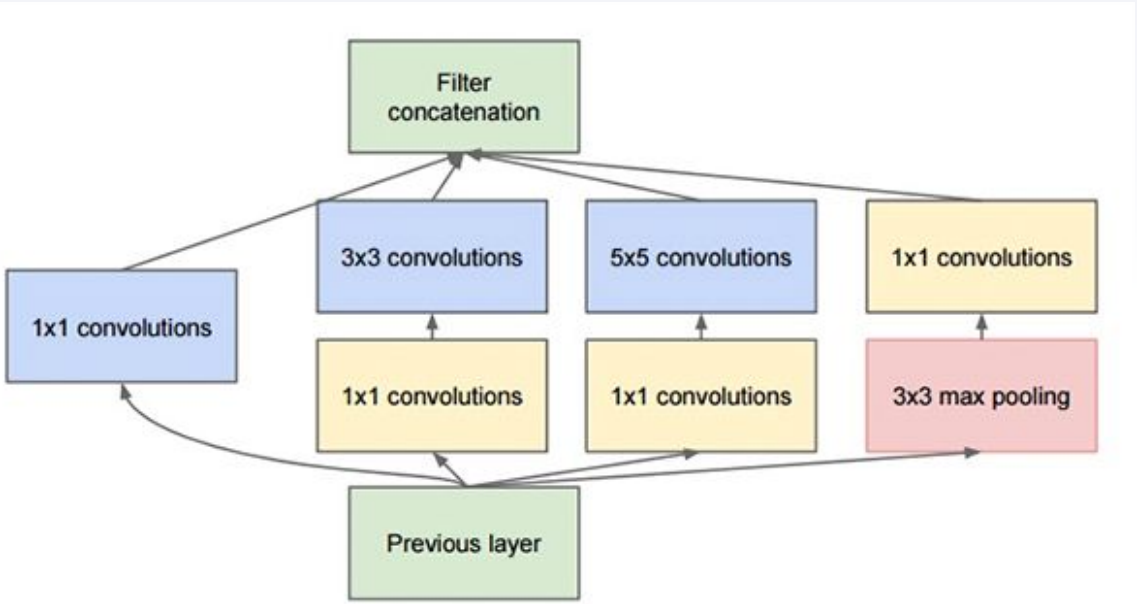
VGGNet Model



https://www.cs.toronto.edu/~frossard/post/vgg16/

ResNet and Inception Model

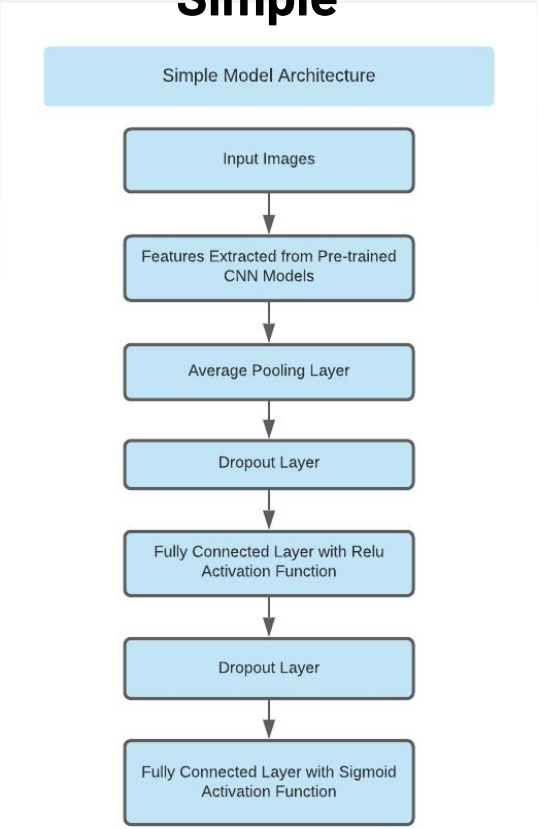




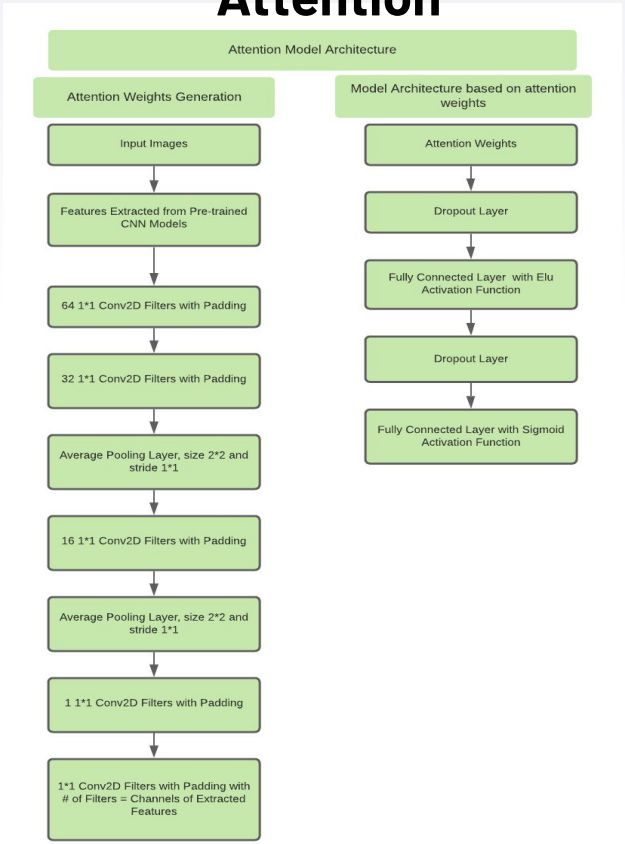
https://www.analyticsvidhya.com/blog/2018/10/understanding-inception-network-from-scratch/

Customized Model Architecture

Simple

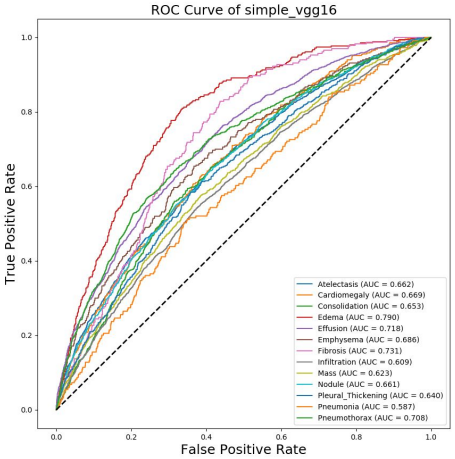


Attention



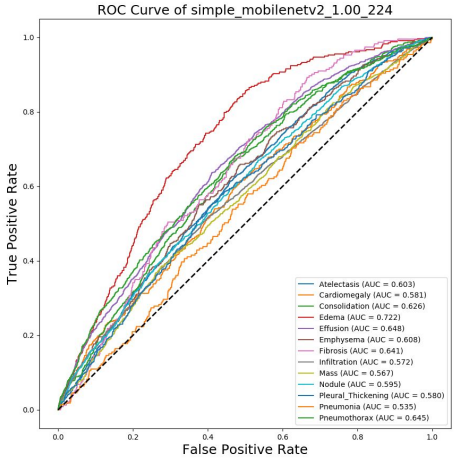
Pretrained Models

VGG16



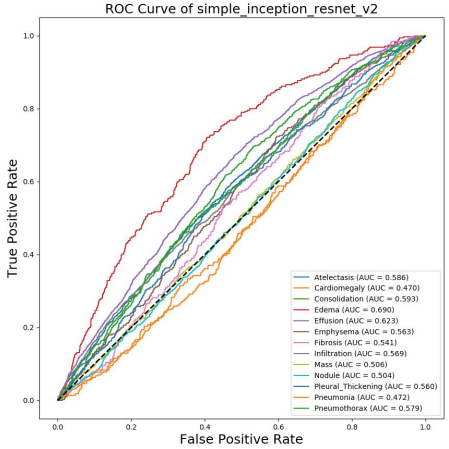
Average AUC = 0.672

MobileNet_V2



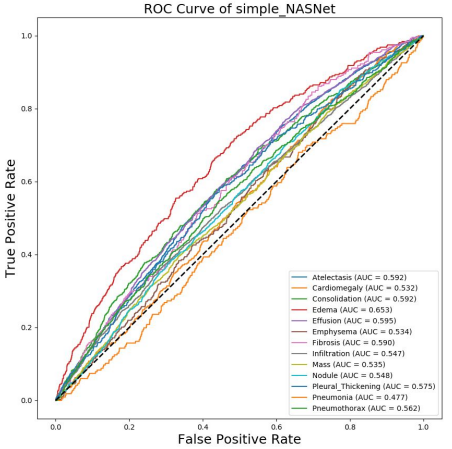
Average AUC = 0.609

Inception_resnet_v2



Average AUC = 0.558

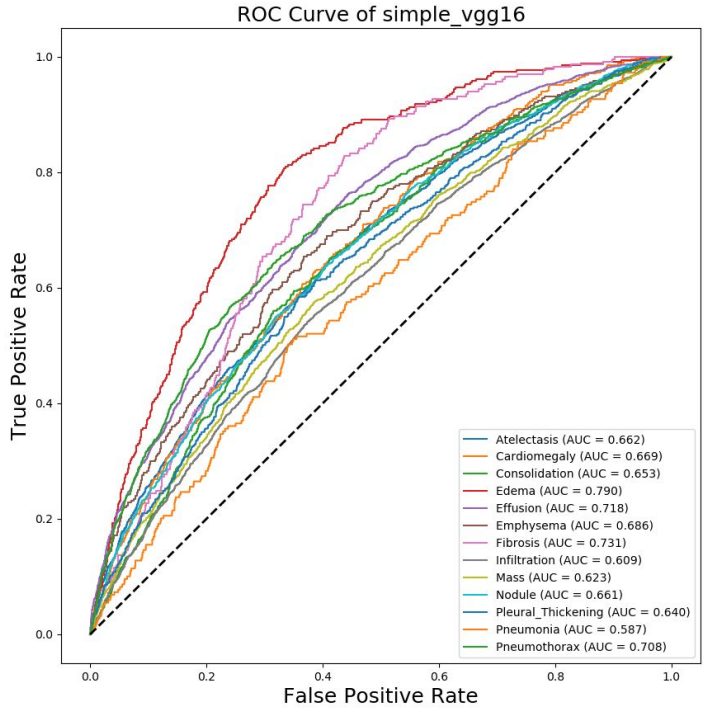
NASNet



Average AUC = 0.564

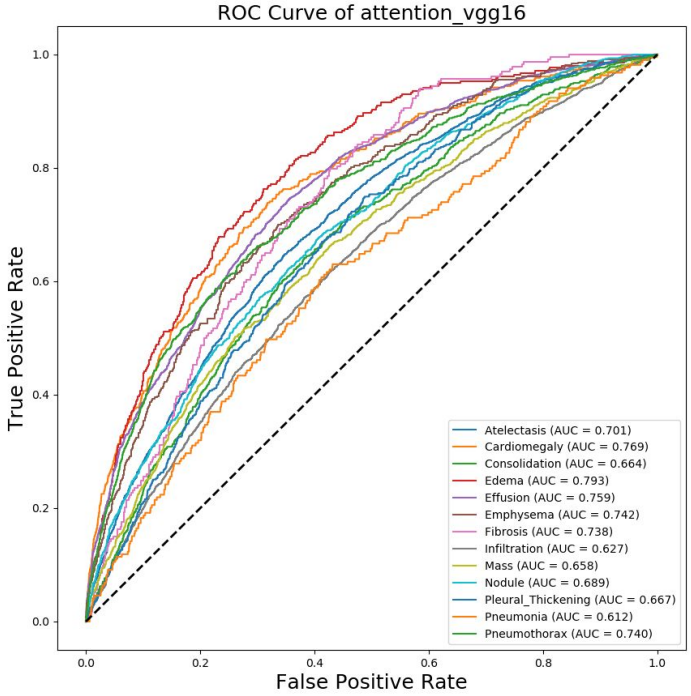
Simple vs Attention

Simple



Average AUC = 0.672

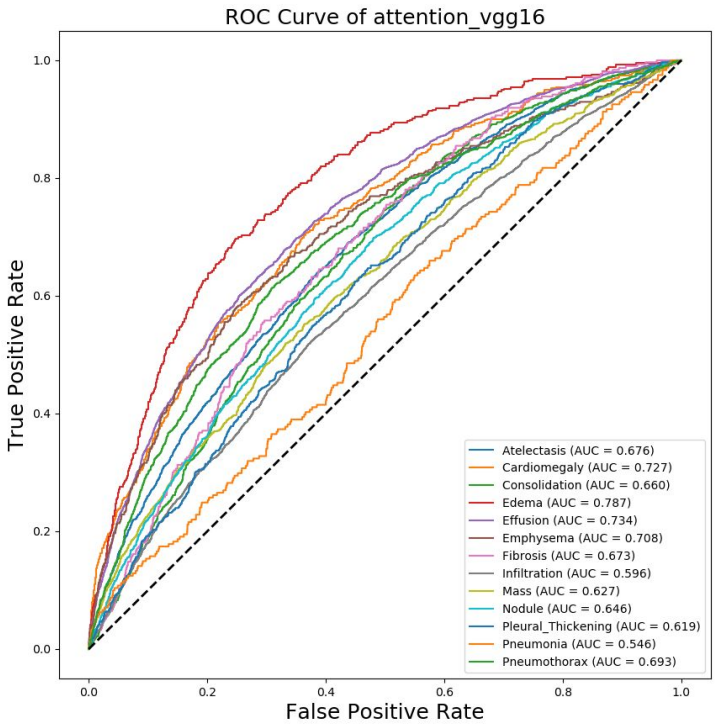
Attention



Average AUC = 0.705

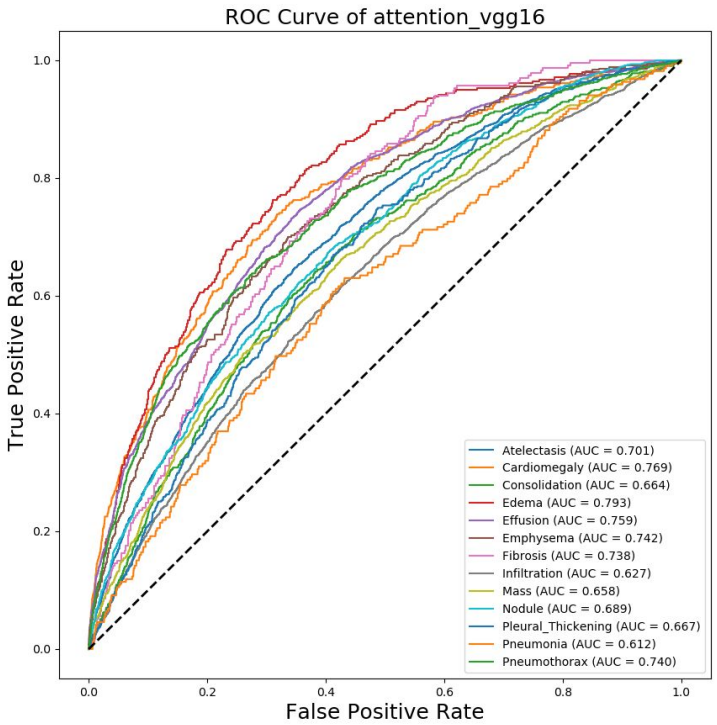
Optimizer: Adam vs Accumulated Adam

ADAM



Average AUC = 0.669

Accumulated Adam



Average AUC = 0.705

Conclusion

- Model Performance:
  - VGGNet has the best performance over all of the other pre-trained CNN models.
  - VGGNet16 plus Attention architecture achieves the best performance in terms of the AUC.

- Achievement:
  - Our end-to-end CNN models achieve satisfactory performance for classifying thoracic diseases using chest X-ray dataset. The AUC is comparable to the literature result.

Thanks for your attention!