


Chest X-ray Disease Diagnosis with Deep Learning Architecture

Team 37

CSE-6250



Motivation

- The chest X-ray is one of the most commonly accessible radiological examinations for screening and diagnosis of many lung diseases. However, current diagnostics using X-ray images in medical practices still highly depend on the expert of radiologists
- Deep Learning is a technique that shows high potential for robust image classification and object recognition
- Our project aims at using deep convolutional neural network to detect and localize diseases in chest X-ray images

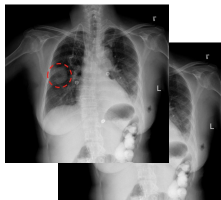
Objective

Dataset: ChestX-ray14 which contains 112,120 frontal-view X-ray images of 30,805 unique patients and each image was annotated with 14 different thoracic pathology labels

- Lung disease diagnosis: Multiple-label classification
- Lung disease localization

Pipeline

Data acquisition



Preprocessing → Training ↔ Evaluation



Disease Classification

Data preprocessing

	Nodule	Effusion	Mass	Consolidation	Edema	Pneumonia	Fibrosis	Atelectasis	Pleural_Thick	Infiltration	Emphysema	Cardiomegaly	Pneumothorax	Hernia	SUM
Original count	4708	8659	4034	2852	1378	876	1251	8280	2242	13782	1423	1707	2637	141	53970
Original ratio	8.7%	16.0%	7.5%	5.3%	2.6%	1.6%	2.3%	15.3%	4.2%	25.5%	2.6%	3.2%	4.9%	0.3%	100.0%
Reduced training set	2000	2433	2000	2118	1378	876	1251	2000	2037	2520	1423	1707	2105	141	23989
Reduced set ratio	8.3%	10.1%	8.3%	8.8%	5.7%	3.7%	5.2%	8.3%	8.5%	10.5%	5.9%	7.1%	8.8%	0.6%	100.0%

Training set: 16,000 images

Validation set: 4,000 images

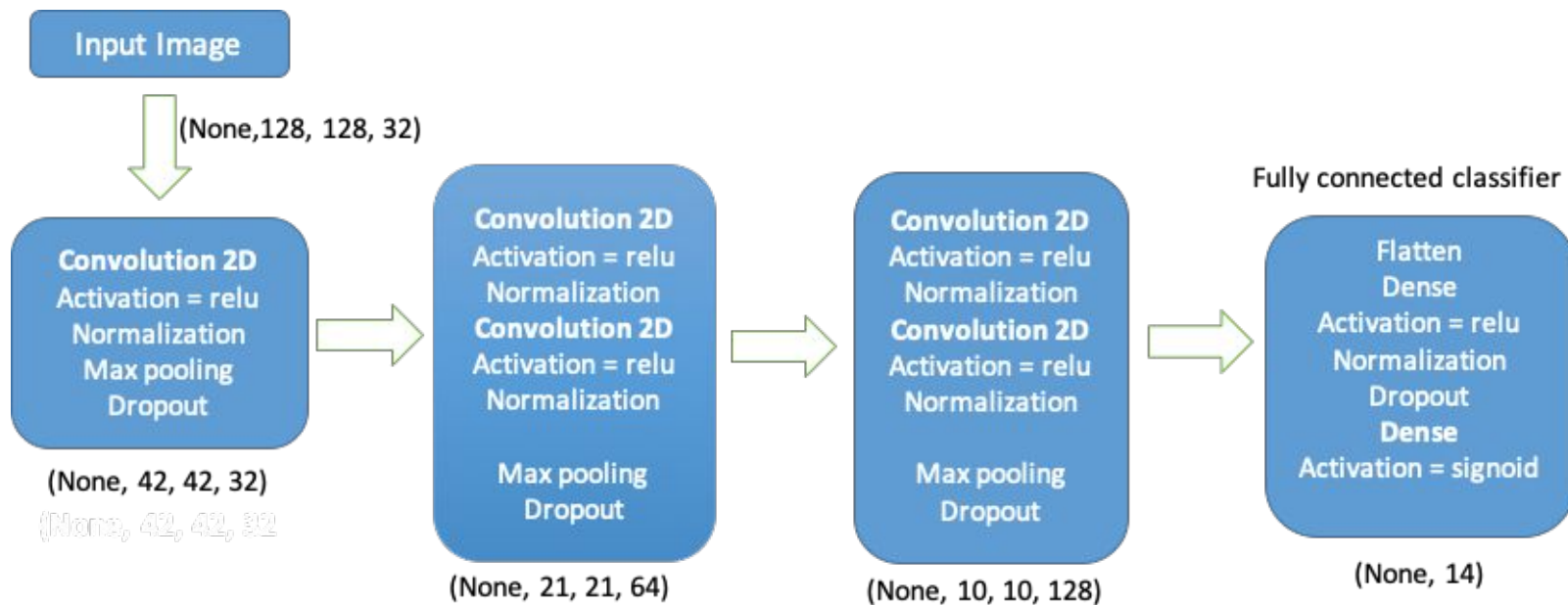
Test set: 20,000 images

Downscaled the images to 128×128

Scaled the raw pixel intensities to the range [0, 1]

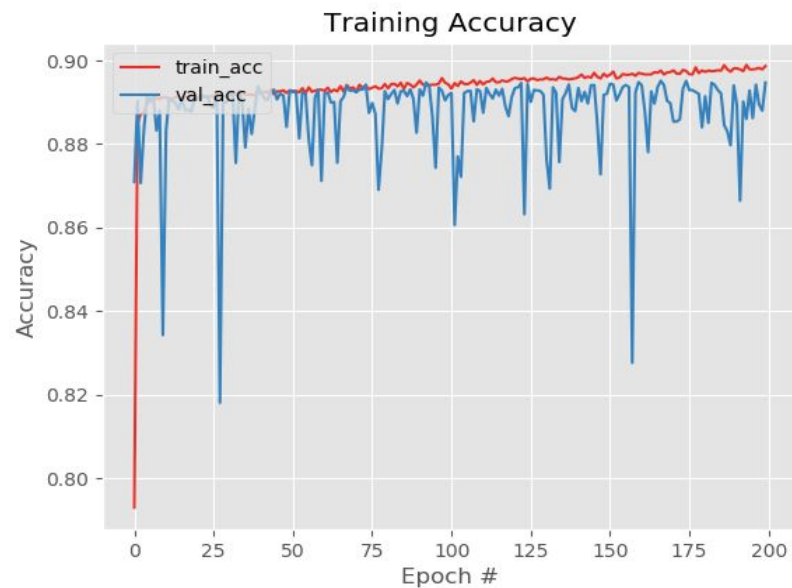
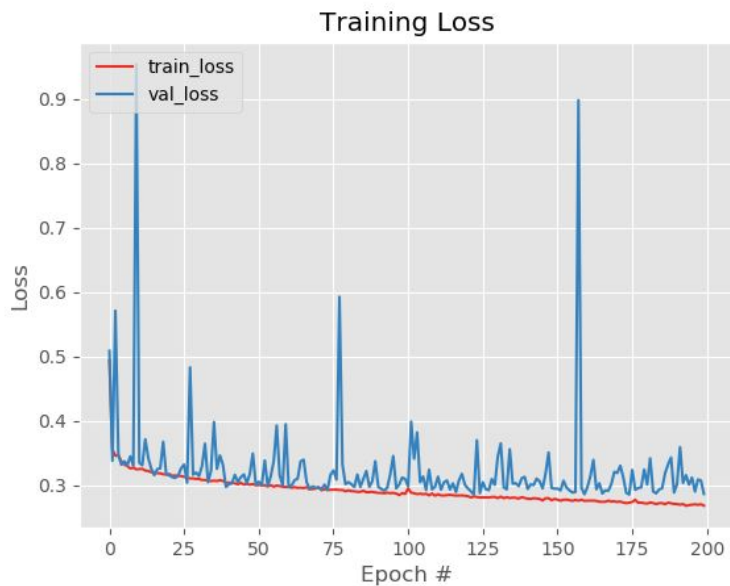
Disease Classification

Structure of CNN



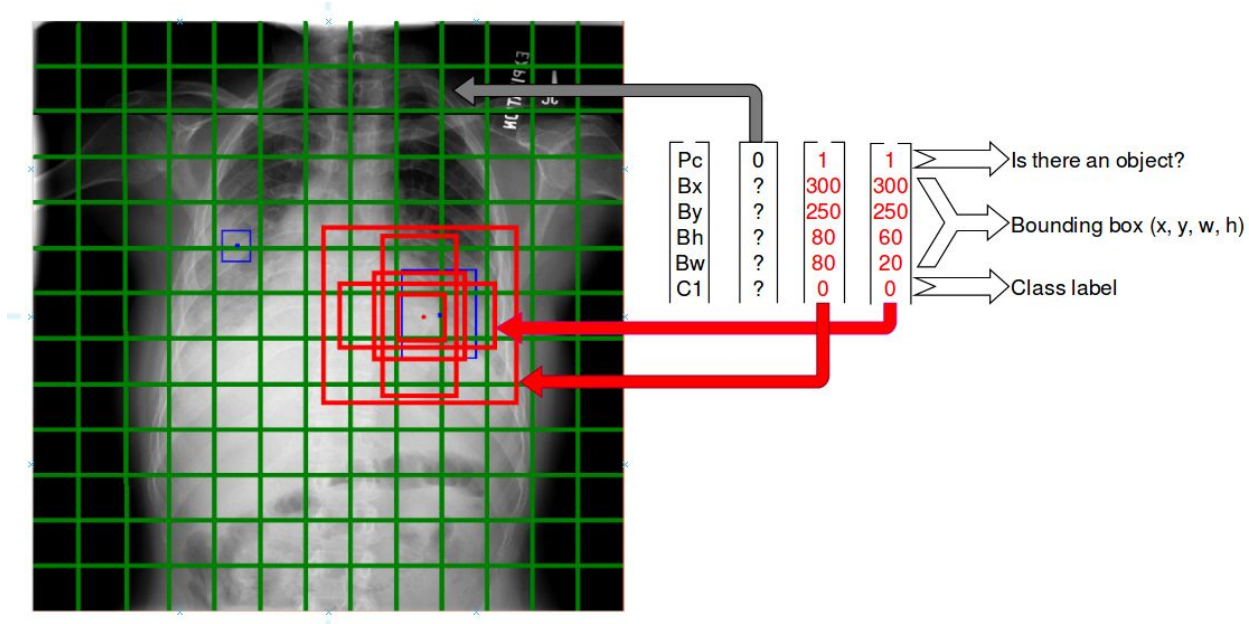
Disease Classification

Experimental evaluation



Lung disease detection and localization

- YOLO (You Only Look Once) is a state-of-the-art object detection architecture.



Lung disease detection and localization

Data preprocessing

	Nodule	Effusion	Mass	Consolidation	Edema	Pneumonia	Fibrosis	Atelectasis	Pleural_Thick	Infiltration	Emphysema	Cardiomegaly	Pneumothorax	Hernia	SUM
Bounding box count	79	153	85	0	0	120	0	180	0	123	0	146	98	0	984
Bounding box percentage	8.0%	15.5%	8.6%	0.0%	0.0%	12.2%	0.0%	18.3%	0.0%	12.5%	0.0%	14.8%	10.0%	0.0%	100.0%

ChestX-ray14 comes with very limited (about 1000) ground truth bounding boxes

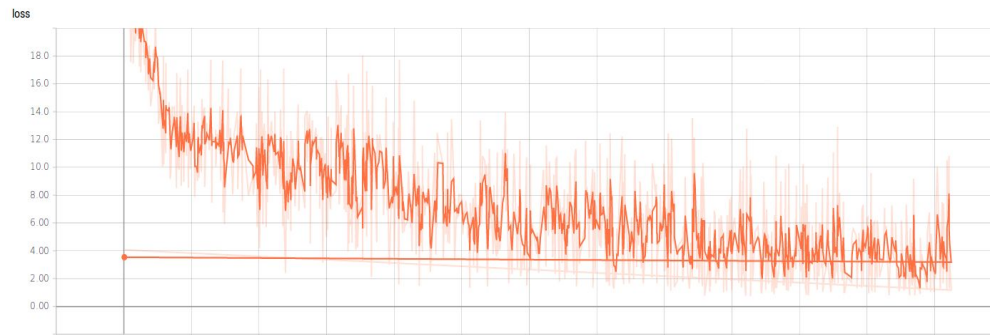
Treat the 14 diseases as a single type of lung disease in order to provide enough training samples for detection and localization algorithm

Intentionally created more variants randomly from available training data through image augmentation techniques

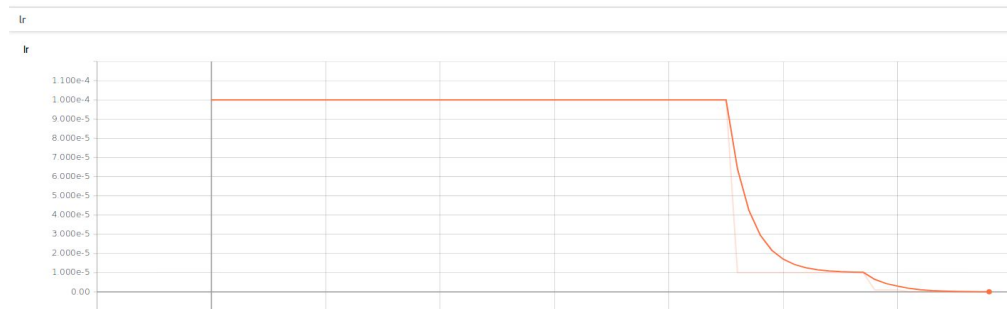
Lung disease detection and localization

Experimental evaluation

Training loss



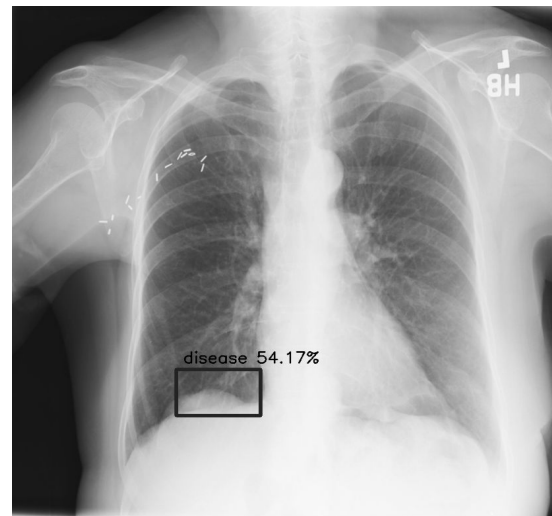
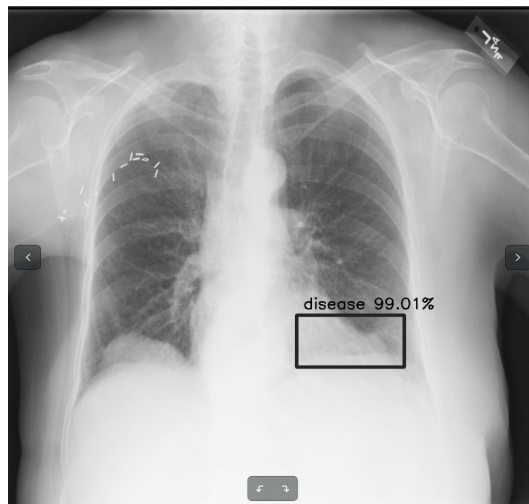
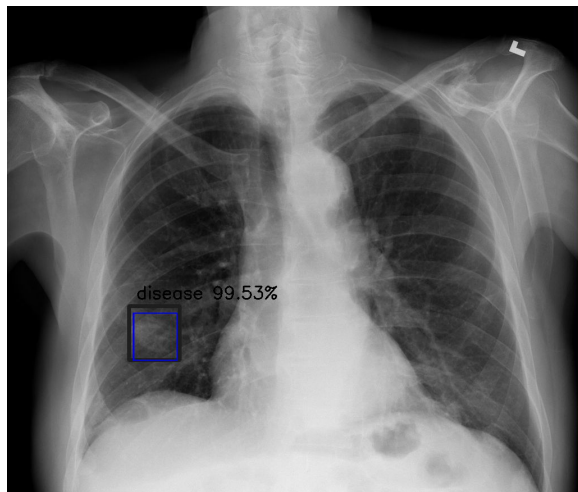
Learning rate decay



Lung disease detection and localization

Experimental evaluation

mAP (mean accuracy precision): 0.29



Future Work

- Use more complex models with appropriate regularization to avoid overfitting
- Create more labeled bounding boxes to improve the detection results