# Localization of Radiographic Evidence for Pneumonia

**Ekaterina Kastrama**
kastrama@stanford.edu

**Robert Pinkerton**
rpinkerton@stanford.edu

**Karina Samuel-Gama**
ksamuelg@stanford.edu

## Abstract

*Deep Neural Networks were applied to Chest Radiographs to accurately classify patients with pneumonia and draw bounding boxes over regions within the radiograph that had strong evidence of pneumonia. Pneumonia, an infection of the lungs and a leading cause of death worldwide, is characterized by fluid in the lungs and can be identified by regions of opacity in radiographs. Pneumonia is difficult to diagnosis, because it requires expert interpretation of chest radiographs and laboratory results to differentiate pneumonia related opacity from other pulmonary cavity diseases. Two different one-stage detection models were evaluated - RetinaNet and YOLOv3. Additional steps were taken to determine if state of the art performance could be recovered by detection algorithm that ran in real-time or near real time. The final models used were RetinaNet, YOLO, and YOLO + CheXNet and were evaluated on mean Average Precision of bounding box predictions compared to labels generated by expert radiologists. RetinaNet demonstrated higher performance than YOLO even in cases were YOLO was paired with robust classifiers. YOLO ensembles performs marginally better than YOLO as a single model. In addition, some steps of ChexNet improvements were taken such as different weights initialization.*

## 1. Introduction

Pneumonia is an infection of the lungs that causes inflammation and fluid build up in the alveoli. The build-up causes symptoms such as chest pain, fever, and severe cough. In some cases where a patient has a weakened immune system or the condition is not caught early enough, the disease may result in death [1]. Pneumonia is one of the leading causes of death in the U.S. and worldwide [2]. Currently, diagnosis is made by highly trained clinical experts interpreting chest radiographs (CXR) and laboratory exams [3]. Areas of increased opacity are usually clear indicators for the potential presence of pneumonia, because, the fluid, characteristic of pneumonia, preferentially attenuates the x-ray beam and therefore appears more opaque than the surrounding area [4]. Diagnosis of pneumonia via CXR is complicated because a number of other pulmonary conditions present as opaque regions in the CXR, such as fluid overload (pulmonary edema), bleeding, volume loss (atelectasis or collapse), lung cancer, or post-radiation or surgical changes [4]. Our proposed model uses convolutional neural networks to processes CXRs and output bounding boxes localized to opaque regions indicating evidence for solely pneumonia. A robust method for accurately identifying cases of radiological evidence for pneumonia would speed diagnosis time and aid healthcare providers with providing a higher quality of care for patients and hopefully reduce the number of deaths caused by pneumonia worldwide.

## 2. Related Work

### 2.1 Medical Image Classification

Recent advancement in image recognition have shown that Deep Neural networks can be successfully used for medical images classification and localization [9]. CheXnet a 121 layer Densenet architecture model achieved an F1 score of 0.435 when classifying 14 different pulmonary conditions including pneumonia, which was markedly higher than the 0.387 average of radiologists [5]. Despite this performance the model has received criticism on implementation and source data [6]. Despite high performance of deep learning models on image classification (medical images or otherwise), interpretation of chest radiographs is still done manually [6].

### 2.2 One Stage Detection

Most predictive models implemented in clinical settings prioritize simplicity of implementation and output [7]. One stage detection and localization models such as YOLO, and SSD, have been shown to have lower accuracies but reduced model complexity [10][11]. RetinaNet, uses focal loss to improve accuracy of one stage detectors to a level comparable to that of a two stage detectors [8]. The models proposed in this work focuses on the uses of one-stage

detection frameworks for the classification and localization of radiographic evidence for pneumonia.

## 3. Dataset and Features

The CXRs and the accompanying bounding box labels are sourced from the Radiological Society of North America (RSNA) via the RSNA Pneumonia Detection Kaggle competition [12]. The dataset consists of ~37,000 unique patient IDs labeled as 31% with opacity, 41% no lung opacity (normal), and 29% other (not normal, no opacity). For images labeled as pneumonia positive, bounding boxes of the abnormalities have also been included. An image can have anywhere between 0 and 4 associated bounding boxes. Patient's sex and age also included in the data but was not used for the model. Compared to previously used in research datasets such as NIH the RSNA dataset claims to be more precisely labeled [4].

The CXRs are stored in dicom format at 1024x1024 solution. The images have been converted to jpeg and scaled down (various sizes) for further analysis. Very basic data augmentation, flip transform of random images, has been used to increase size of training set. Additionally in some models, randomly concatenated pairs of positive and negative images have been used as training input. The 30,200 images in the training data was split into an 80/10/10 train/dev/test split. There was an additional set of 3000 images designated as the test set by the Kaggle sponsors for which labels were not provided. This set was used only for comparison against other leaderboard results.

*Negative, Normal* *Positive, Pneumonia* *Negative, Not Normal*

## 4. Methods

### *4. 1 RetinaNet*

RetinaNet is one stage object detection algorithm that consistently outperforms other models in terms of accuracy [8]. This model was selected as a representation for state of the art one shot object detection. Compared to other one-stage object detection algorithms, RetinaNet uses a novel loss function that recovers accuracy seen in two stage detectors. The loss function, focal loss, applies changes to the standard cross entropy loss to weight losses from harder classes over easier classes. This idea is captured in the term $(1 - p_t)^{\gamma}$ in which $p_t < 0.5$ indicates a difficult class and $\gamma > 0$ controls reduces weight for easier classes [8].

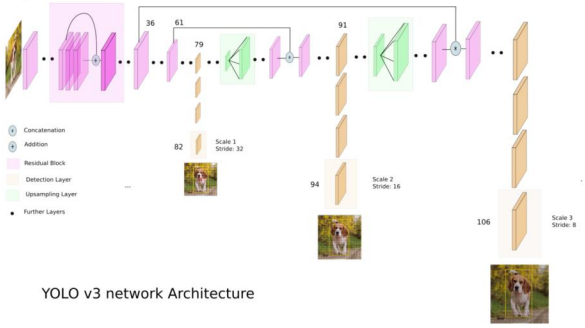
$$FL(p_t) = -(1-p_t)^{\gamma} * log(p_t)$$

The model used in this research was adapted from the keras-RetinaNet repository [13]. In this implementation, negative images were ignored in training, changes were made to calculate loss on all images. A learning rate of 0.001 and batch size of 64 were chosen to match the hyperparameters used in YOLOv3. Despite better performance shown by selecting ResNet101 for the RetinaNet backbone [8], ResNet51 pre-trained on ImageNet was selected for decreased training time. Backbones other than ResNet were not explored. The models were trained for 6 hours on two p100s.

### *4. 2 YOLOv3*

YOLO is a model known for fast, robust predictions of objects in real time. The original implementation struggled with small and unknown objects [10]. Subsequent iterations improved limitations of the first. Compared to YOLOv2, YOLOv3 architecture was improved by adding new blocks, such as residual blocks, skip connections and upsampling. These additional layers allow for a much deeper network. YOLOv3 is built on Darknet 53, compared to Darknet 19 of previous iterations.

For the task of detection, 53 more layers have been stacked onto the backbone, giving a total of 106 layers.




YOLO v3 network Architecture

YOLO's innovation lies within its loss function, which is a 5 term cross entropy loss that captures error in the width, height, coordinates, class probability, and confidence. This function captures the idea of a regression approach to predicting bounding boxes. In YOLOv3 class and confidence are also predicted in a regressive manner [10][14].

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(x_i - \hat{x}_i\right)^2 + \left(y_i - \hat{y}_i\right)^2$$
$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2$$
$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i\right)^2$$
$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i\right)^2$$
$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c)\right)^2 \quad (3)$$

YOLO Loss Function

YOLOv3 is an appropriate choice for localization because it is a one stage detection network that allows for a simpler, less computationally expensive pipeline. The model was initialized with weights from darknet53.conv.74 that had been trained on ImageNet. A learning rate of 0.001 and batch size of 64 with a subdivision of 8 were selected on recommendation of YOLO creators and hw constraints. No changes were made to bounding box configurations. The model was trained on the positive samples from our training data set on two P100's for around 10 hours. The first 1000 iterations were trained on a single P100 for stability.

### *4.3 DenseNet 121 (CheXNet)*

To improve YOLOv3 classification capabilities [14], the model was paired with CheXNet, a classification network. Given the similarity of the task, classifying chest x-rays, CheXNet is an appropriate first choice for classification modules. The model implementation [15] was adapted to train on three classes - Normal, Opacity, and Not normal/opacity compared to the 14 conditions initially used in the paper [5]. CheXnet uses a simple log loss function in which the loss is equal to the negative log of the predicted probability of an image being of a certain class, such that the error decreases the closer the probability is to 1 or 0 depending on the class label. This loss is summed across all three classes [5].

$$L(X, y) = \sum_{c=1}^{3} [-y_c \log p(Y_c = 1|X) - (1 - y_c) \log p(Y_c = 0|X)]$$

Four DenseNet models were trained, three were using pre-trained weights from the NIH dataset and the one was using pre-initialized weights from ImageNet: as weights initialization , with and without layers freeze and drop out to overcome overfitting. DenseNet was run using scaled down images at 224x224 resolution, with horizontal flipping, batch size of 32 and a decreasing learning rate ("reduced on Plateau") in a range (0.001, 1e-9). However, the low learning rate was causing the network to overfit on training data, so future iterations were limited to the smallest learning rate of 1e-7. The network were trained for approximately 30 epochs. The result of prediction was ensembled with Yolo predictions to improve the classification.

## 5. Results

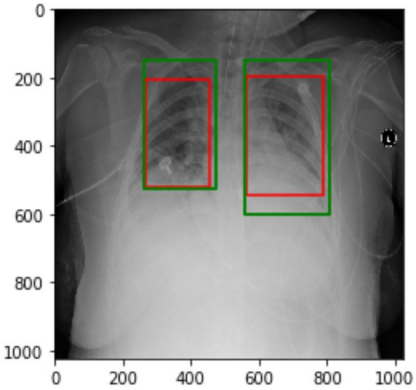



*Image of patient with pneumonia. Green indicates ground truth label and red indicates YOLO bounding box predictions.*

Both YOLO and RetinaNet were able to locate regions of pneumonia related opacity in CXRs. Above we can see example patient with opaque regions related to pneumonia bounded.

Given the dual nature of the task, classification and localization, we evaluated our models using F1 and AUROC for classification and mean average precision (mAP) for localization. mAP is defined as the average precision of the bounding boxes at different intersection over union (IoU) thresholds. The metric sweeps over a range of IoU thresholds, at each point calculating an average precision value. The threshold values used range from 0.4 to 0.75 with a step size of 0.05: (0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75). It has gained popularity as benchmark metric and is used in several object detection challenges (including the RSNA kaggle competition) as well as in both the RetinaNet and YOLOv3 papers.

For each model, the results for the best implementation on the test set are listed below.

| Model | F1 Score | AUROC | mAP |
|---|---|---|---|
| CheXNet with NIH pretrain | 0.40 | 0.9058 | N/A |
| RetinaNet | N/A | N/A | 0.157 |
| YOLOv3 | 0.81 | N/A | 0.141 |
| CheXNet + YOLOv3 | 0.8385 | N/A | 0.144 |

As expected, RetinaNet outperformed YOLOv3 in terms of bounding box precision. Pairing YOLO with Chexnet to improve classification did not achieve performance comparable to RetinaNet. The F1 score of ensembling two models slightly increased.

$$\frac{1}{|thresholds|} \sum_{t} \frac{TP(t)}{TP(t) + FP(t) + FN(t)}$$

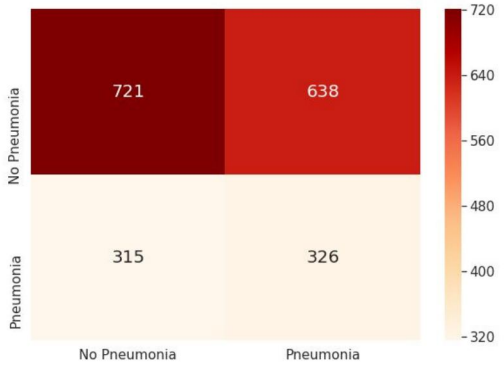
mAP formula

## 6. Discussion

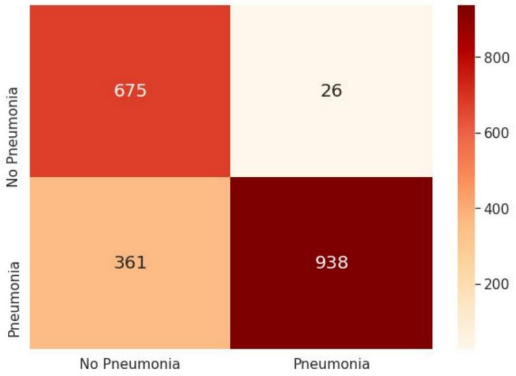
Models trained on only positive images containing bounding boxes showed good localization and classification when tested with both positive and negative images. Ensembled CheXNet + YOLO showed increased precision even when trained on the entire dataset compared to the single YOLO model trained on only the positive labels. Despite the marginal improvement of the ensemble model, we were not able to recover the performance of the state of the art baseline - RetinaNet. To understand the difference in performance we investigated the two components, classification and localization, separately.

DenseNet classification algorithm showed better performance for classifying 'Normal' vs Localization algorithms. For the class "No opacity/ Not normal' the performance was the worst. Binary (Pneumonia/Not Pneumonia) F1 score for trained DenseNet on RSNA data was close to original CheXNet result (0.43 CheXnet and 0.40 DenseNet with NIH activation).

ChexNet Confusion Matrix on Test Set:



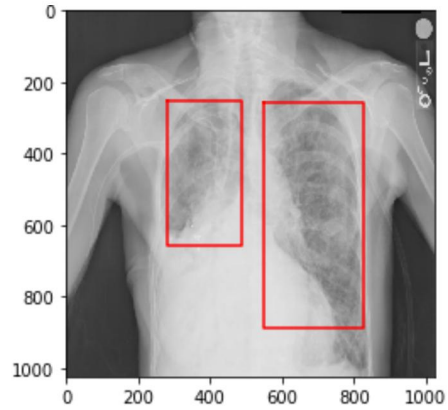

YOLO Confusion Matrix on Test Set:




*Confusion matrix of YOLO and Chexnet showed that YOLO is better in True Positive classification and and DenseNet slightly better in True negative prediction.*

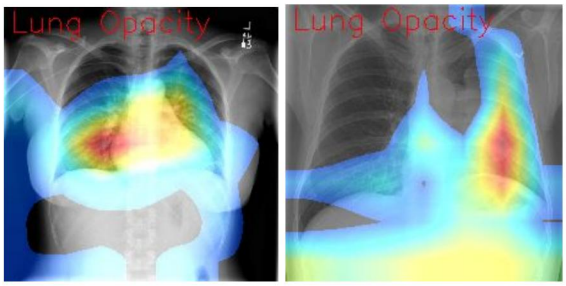
## Error analysis

In some cases YOLO was unable to distinguish between no pneumonia/not normal and pneumonia patients. In the cases were YOLO failed the baseline illumination was higher compared to other negatives, which indicates poor radiograph. Other cases of failure showed similar noisiness at baseline. To avoid these false positives, additional training on these cases would be necessary either via feeding in twomers (positive/negative pairs for training) or data augmentation that includes lumosity shift/change.




*Sample patient with no pneumonia with incorrectly drawn bounding boxes.*

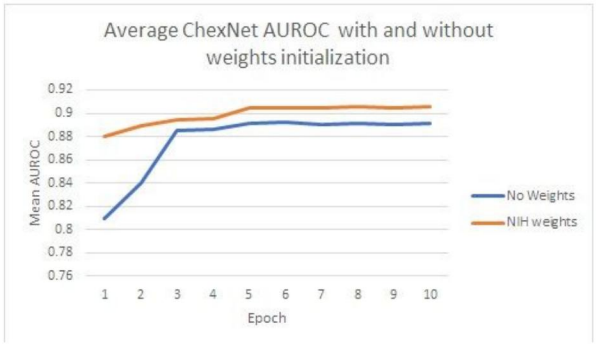
DenseNet CAM analyses showed that network usually focusing on the right areas, but in some cases it looked at extraneous regions. These cases can be explained by different image quality. Adding Deformable Layers to identify offset might improve this [16].




*Generated CAM of CheXNet model to verify classification weighted important regions, evidence of misclassification of other sources of opacity*

### Weights initialization experimentation

Experimentation with previously trained NIH weights with DenseNet 121 led to increase of AUROC metric but overfitting on the training set (train and validation loss difference: train loss: 0.0673 - val loss: 0.5557 after 4 epochs). Validation loss never passed below 0.37 for all four models. Changing architecture and adding Dropout layer improved the difference but overfitting was still observed. Freezing layers led to underperforming model. Adding more Dropout layers may potentially improve the performance.




## 7. Conclusion

Retinanet and Yolo were able to predict and localize radiographic evidence for pneumonia. Pairing YOLO with more robust classifier marginally improved YOLO. Both YOLO and Densenet were better at predicting negatives and had difficulty distinguishing from true positives and not normal.no pneumonia class. Retinanet outperformed all other models used for pneumonia classification, most likely due to weighted emphasis on harder classes.

## 8. Future Steps

There is value in a cheap algorithm like YOLO that could be run in near real time that works on low resolution data. This model would be especially useful for areas with limited access to computational and medical resources. Future work would involve boosting the performance of YOLO so it is more comparable to RetineNet, while still maintaining relatively simple architecture. Though we were not able to show that pairing YOLO with CheXNet boosted performance, further work could be done to find a cheaper classifier that improves performance compared to YOLO trained with only positive labels. Another option would be to use the RetinaNet framework with a YOLO backbone in attempt to capture the improvement gained by using Focal Loss with a simpler, faster model than ResNet.

Other methods that could improve accuracy include image segmentation using bounding box distribution could be performed that prelocalized the region of interest. Future training could hone in on the errors produced by misclassification of no pneumonia/no normal class of images, such as creative concatenations of train images, up sampling this class, etc. Similarly, class activation maps for all models could be investigated to determine potential regions prone to misclassification. Additionally, many of our models overfit, experimentation with dropout regularization on the classification layers may improve overall performance.

## 9. Contributions

Each team member spearheaded a different approach - Ekaterina worked on the simple CNN, CheXNet modification for RSNA data (preparing data, labels, applying different weights and architecture change, script for prediction and adding F1 metric to the implementation) and CAM visualization; Robert worked on the development environment, exploratory data analysis, the YOLOv3 model with the mAP score; Karina worked on the YOLOv3 and RetinaNet models and the exploratory data analysis. All members contributed equally to the report.

## 10. Code

**Software:** Described in Requirements.txt on github
**Hardware:** We used various GPU's on a Google Cloud VM Instance
**Github:** https://github.com/pinkertr/cs230project/
**TeamDrive**:
https://drive.google.com/drive/u/1/folders/0AJnJSPxbRhX8Uk9PVA
**Libraries**


- Pandas: McKinney, Wes. "Data structures for statistical computing in python." *Proceedings of the 9th Python in Science Conference*. Vol. 445. 2010.
- Sci-Kit Learn: Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of machine learning research* 12.Oct (2011): 2825-2830.

- Keras: Chollet, François. "Keras: The python deep learning library." *Astrophysics Source Code Library* (2018).
- Numpy: Walt, Stéfan van der, S. Chris Colbert, and Gael Varoquaux. "The NumPy array: a structure for efficient numerical computation." *Computing in Science & Engineering* 13.2 (2011): 22-30.
- OpenCV: Bradski, Gary, and Adrian Kaehler. "OpenCV." *Dr. Dobb's journal of software tools* 3 (2000).
- Pydicom: Mason, D. "SU-E-T-33: Pydicom: an open source DICOM library." *Medical Physics* 38.6Part10 (2011): 3493-3493.

## References


1. LaCroix, Andrea Z., et al. "Prospective study of pneumonia hospitalizations and mortality of US older people: the role of chronic conditions, health behaviors, and nutritional status." *Public health reports* 104.4 (1989): 350.
2. Duthey, Béatrice. "Priority medicines for europe and the world:"a public health approach to innovation"." *WHO Background paper*6 (2013).
3. "Diagnosing and Treating Pneumonia." *Lung.Org*, American Lung Association, www.lung.org/lung-health-and-diseases/lung-disease-lookup/pneumonia/diagnosing-and-treating.html.
4. Franquet T. Imaging of community-acquired pneumonia. J Thorac Imaging 2018 (epub ahead of print). PMID 30036297
5. Rajpurkar, Pranav, et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." *arXiv preprint arXiv:1711.05225* (2017).
6. Oakden-Rayner, Luke. "CheXNet: an in-depth review." *URL: https://lukeoakdenrayner. wordpress. com/2018/01/24/chexnetan-in-depth-review* (2018).
7. Shah, Nigam. "Predictions." Data Driven Medicine. Biomedin 215, Stanford, Gates B3.
8. Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *IEEE transactions on pattern analysis and machine intelligence*(2018).
9. Li, Qing, et al. "Medical image classification with convolutional neural network." *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*. IEEE, 2014.
10. Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).
11. Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016.
12. 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.
13. Keras-RetinaNet . Keras implementation of RetinaNet object detection. GitHub. https://github.com/fizyr/keras-retinanet
14. Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
15. CheXNet-Keras. Keras tool to build CheXNet like models. GitHub. https://github.com/brucechou1983/CheXNet-Keras
16. Dai, Jifeng, et al. "Deformable convolutional networks." *CoRR, abs/1703.06211* 1.2 (2017): 3.