Classification of Radiographic Evidence for Pneumonia

Ekaterina Kastrama, Robert Pinkerton, Karina Samuel-Gama

{kastrama, rpinkerton, ksamuelg}@stanford.edu

Stanford

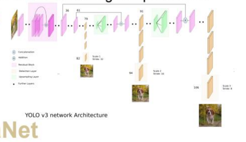
Introduction

Pneumonia is one of the leading causes of death in the U.S. Timely, accurate diagnosis is a critical factor in determining patient outcomes. Currently, pneumonia diagnosis is made by highly trained clinical experts interpreting chest radiographs (CXR) and laboratory exams. Areas of increased opacity are usually clear indicators for the presence of pneumonia, because the fluid, characteristic of pneumonia, *preferentially attenuates the x-ray beam and therefore appears more opaque than the surrounding area.* **Our proposed model takes in chest radiographs as input and outputs bounding boxes localized to opaque regions indicating presence of pneumonia.** An autonomous method for accurately identifying cases of radiological evidence for pneumonia would speed diagnosis time and hopefully reduce the number of deaths caused by pneumonia world

Model Architecture

The main compared were YOLOv3 and RetinaNet. RetinaNet uses ResNet51 for classification backbone. YOLOv3 uses Darknet53 as backbone. Both models are one-stage detectors where one optimizes speed and the other is lauded for its high accuracy focal loss function. Much of our investigations centered around recovering similar accuracy using YOLOv3 as the main component for bounding box prediction.

YOLO



YOLO v3 network Architecture

Yolov3 Hyperparameters

- batch size - 64
- lr - 0.0001
- batch norm
- pretrained - ImageNet

RetinaNet

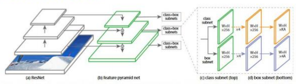


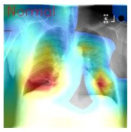
Figure 3. The one-stage **RetinaNet** network architecture uses a Feature Pyramid Network (FPN) [19] backbone on top of a feedforward ResNet architecture [15] (a) to generate a rich, multi-scale convolutional feature pyramid (b). To this backbone RetinaNet attaches two subnetworks, one for classifying anchor boxes (c) and one for regressing from anchor boxes to ground-truth object boxes (d). The network design is intentionally simple, which enables this work to focus on a novel focal loss function that eliminates the accuracy gap between our one-stage detector and state-of-the-art two-stage detectors like Faster R-CNN with FPN [19] while running at faster speeds.

RetinaNet Hyperparameters

- batch size - 64
- lr - 0.0001
- batch norm
- ResNet 51 backbone pretrained on ImageNet

Workflow Optimization

- Trained models on only positive images to reduce train time
- Scaled image resolution to reduce train time
- Concatenated positive/negative images to improve performance



Generated cmap of ChexNet model to verify classification weighted important regions, evidence of misclassification of other sources of opacity

Discussion

- The NIH weight initialization improved performance of the ChexNet F1 score and convergence speed
- The category No Opacity/Not normal produced the most misclassification errors
- Chaining YOLO and ChexNet showed decreased performance most likely due to reduced training space and messy training space (mix of positive and negatives)
- Chaining YOLO and ChexNet increased train time considerably
- We were unable to effectively compare class activation maps of the various models to each other and will be left for future work.
- RetinaNet Vs ChexNet + Yolo ensemble: The state of the art retinanet model returned a much higher AP IOU score

Future Work

Although Chexnet + YOLO boosted accuracy, DenseNet121 is expensive to train, future work could be

- YOLO with cheaper classifiers that maintains accuracy
- All model class activation map comparison
- Low fold ensembling
- Crop image based on box distribution
- Focus on No Lung Opacity / Not Normal prediction - predicting 3rd class, concatenating train images, etc.
- Dropout regularization to sync classification and bounding box prediction

There is value in a cheap algorithm like YOLO that could be run in near real time working with low resolution data for areas with limited access to computational resources.

Dataset & Features

The chest radiographs and the corresponding bounding boxes are provided by the Radiological Society of North America (RSNA) via the Pneumonia Detection Kaggle competition. 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). The images are stored in dicom format at 1024x1024 solution. The images have been converted to .jpg and scaled down (various sizes) for further analysis. We have split the data into an 80/10/10 train/dev/test split.

Negative, Normal — Positive, Lung opacity — Negative, not normal

Comparison of Loss

Chexnet Classification Loss

$$L(X,y) = \sum_{c=1}^{14}[-y_c \log p(Y_c = 1|X) - (1-y_c)\log p(Y_c = 0|X)]$$

Improved classification performance comes more from the dense layer architecture rather than loss which is just log loss on the 14 class predictions.

YOLO Loss vs. Focal Loss

$$\begin{aligned} loss_t = \sum_{i=0}^{W}\sum_{j=0}^{H}\sum_{k=0}^{A} \; & 1_{Max\,IOU < Thresh}\,\lambda_{noobj} * (-b_{ijk}^{o})^2 \\ & + 1_{t<12800}\,\lambda_{prior} * \sum_{r\,\epsilon\,(x,y,w,h)} (prior_k^r - b_{ijk}^r)^2 \\ & + 1_k^{truth}(\,\lambda_{coord} * \sum_{r\,\epsilon\,(x,y,w,h)} (truth^r - b_{ijk}^r)^2 \\ & + \lambda_{obj} * (IOU_{truth}^k - b_{ijk}^o)^2 \\ & + \lambda_{class} * (\sum_{c=1}^{C} (truth^c - b_{ijk}^c)^2)\,) \end{aligned}$$

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

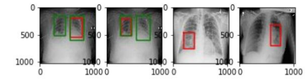
Both YOLO and retinanet are one stage detectors in which classification and object localization occur in the same model. The difference in performance is largely due to calculation of loss.

In YOLOv3 the loss is calculated as the cross entropy loss of the class probabilities and the confidence as well ass the error in the overlap of the predicted bounding boxes. In contrast focal loss uses parameters to weight negative and positive images differently so that a model isn't incentivized to maximize reward from classification of 'easy' negatives.

Results

AP: mean average precision at different intersection over union (IoU) thresholds.

Model	AUROC dev	AP* train
ChexNet on RSNA without weights	0.8916	na
ChexNet on RSNA data with NIH weights	0.9058	na
RetinaNet (state of the art, Kaggle winner)	na	0.157
ChexNet + YOLOv3	na	0.09
YOLOv3	na	0.141



References

1. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning https://arxiv.org/pdf/1711.05225.pdf
2. Foca lLoss for Dense Object Detection https://arxiv.org/pdf/1708.02002.pdf
3. YOLOv3: An Incremental Improvement. https://pjreddie.com/media/files/papers/YOLOv3.pdf
4. Keras-retinanet https://github.com/fizyr/keras-retinaneto
5. Darknet https://pjreddie.com/darknet/yolo/

Youtube Video Presentation: https://www.youtube.com/watch?v=F_Mil9y8KQE&feature=youtu.be