Weakly Supervised Pneumonia Localization

Shih-Cheng (Mars) Huang, Emanuel Cortes, Medi Monam
{mschuang, ecortes, mmonam} @stanford.edu

Motivation and Objective

Motivation:

- Globally, Pneumonia is responsible for over 15% of all deaths of children under the age of 5. Unfortunately, this is usually due to a lack of professionals who can expertly identify the presence of pneumonia in an X-ray image. Furthermore, by knowing the location and area of the infection, doctors can get a better idea of the cause and severity of Pneumonia. However, Chest X-Ray images generated from hospitals are usually only labeled with Pneumonia diagnosis, without specifying the location. The lack of fully labeled data commonly observed in a hospital setting motivated us to explore a weakly supervised training method to do automated diagnosis. Our approach is comparable to the performance of a supervised method as demonstrated in our paper.

Objectives:

- Create a weakly supervised model (CAM + CNN) , which extracts features learned from the classifier and does not require location training labels, to predict Pneumonia location with performance comparable to supervised models.

Data and Features

Data:

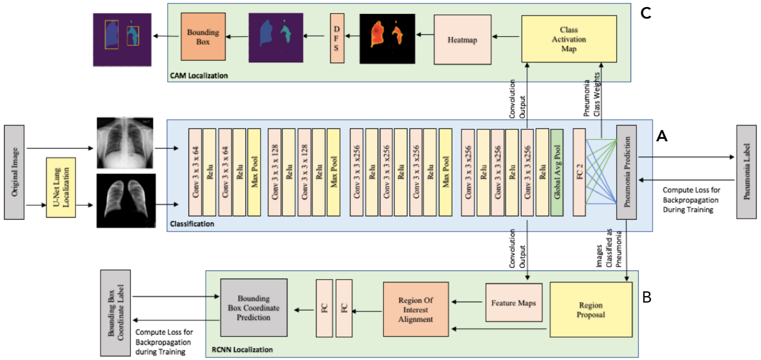
- We acquired our dataset from Kaggle's RSNA Pneumonia Detection Competition. The dataset consists of 17,489 x-ray images (8964 with pneumonia, 8525 no pneumonia). Each Pneumonia image is labeled with the X, Y, weight and height as the ground truth of their bounding boxes.
- The data is almost perfectly balanced 51.25 / 48.74) so we split our dataset to a 70:20:10 train, valid and test split without sampling.

Figure 1: lung segmented from original and lung location provided data

Feature Engineering:

- To expedite the training of our neural network models, we compressed the original 1024*1024 pixels to 128*128.
- All images were normalized by their mean and standard deviation.
- A U-Net neural network was used to predict the confidence of each pixel belonging to the lung. Then we segmented the lung by multiplying the original image matrix with the localization matrix.
- Both the original and the segmented images were fed into our model as inputs to provide our model with a hypothesis of lung location.

Models



Figures 2 a) classification model b) R-CNN localization baseline c) CAM localization model

Classification

- The first part of our study is to build a CNN model that can accurately classify whether a given image is labeled as Pneumonia or not. The images that were predicted as Pneumonia positive are then fed into our localization model.
- SVM, Random Forest and Logistic Regression were used as a baseline to evaluate our classification model.
- Our best model contains 10 convolutional layers, each with zero padding to keep the size of the original image and ReLU as the activation function. The Class Activation Mapper requires our model to only have one fully connected layer and a Global Average Max Pool layer before that. The model were need trained with an Adam optimizer with 0.0001 learning rate on 20 epochs.

Localization:

- The second part of our project is to build a weakly supervised model that can predict Pneumonia location without training labels, on the positively predicted images from the classifier.
- To benchmark the performance our of localization model, we created a supervised RCNN model that was trained with ground truth location labels (figure 2B).
- Our weakly supervised consists of the following components (figure 2C):

$$M_c(x, y) = \sum_k w_k^c f_k(x, y)$$

Formula 1 : Class Activation Mapper

1. A Class Activation Map (CAM) that takes in the output of the final convolutional neural network model and the fully connect layer weights for the Pneumonia class neuron and sum up the weighted outputs (Formula 1) .
2. The output from CAM were then scaled into a 3-channel RGB heatmap.
3. To find individual clusters of predictions on the heatmap, we applied a Depth First Search on a random non-zero pixel on the heatmap, and repeat until all non-zero pixels are clustered.
4. Lastly, we drew a bounding box around each clusters by finding the minimum and maximum X,Y coordinates of the clusters, and only kept boxes that are within 2 standard deviations of all predictions.

Results

Classification	LR	SVM	RF	Our Model
Train	75.86%	74.17%	86.39%	93.07%
Test	73.02%	58.18%	83.00%	C

Localization	R-CNN	CNN +CAM
Train	0.1859	N/A
Test	0.1266	0.1508

Table 1A, Classification accuracies, 1B Localization IoU

Classification

- Since we have a balanced dataset, we used accuracy as a metric to evaluate the performance of our classifier. Our CNN significantly outperformed traditional classifiers without overfitting (Table 1)

Localization:

$$IoU(A, B) = \frac{A \cap B}{A \cup B}$$

Formula 2

- To evaluate the localization task, we used the IoU metric (Formula 2) by calculating the intersection over union of the prediction and ground truth bounding boxes.
- An example prediction heatmap and bounding boxes is shown in figure 4, where the blue boxes are the prediction drawn over the heatmap cluster, while the red boxes are the ground truth.
- Our mode is able to predict Pneumonia location even better than the supervised approach, with an increase of 0.0242 IoU increase (Table 1B)

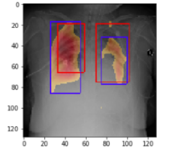


Figure 3: Bounding Box Prediction

Discussion and Future Work

- Based on our result, we have shown that our weakly supervised method is able to localize Pneumonia just as well as a supervised method.
- We predict that our model can perform even better if we have the computing power to train our model on the full images, as a lot of information are lost during compression. We also expect improvements by including more training data or transferring learned models from similar works, such was ChestXNet.
- If improved to human level performance, our weakly supervised model can not only automate pneumonia location annotation and classification tasks, but can also be used to automate other medical image datasets.

References

- He, Kaiming, et al. "Mask r-cnn." Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017.
- Zhou, Bolei, et al. "Learning deep features for discriminative localization." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).