

# Weakly supervised classification and localization of common thorax diseases

Shivam Mittal, Dr. Abhinav Dhall

IIT Ropar



# Main motivation

1. X-ray - one of the most common test with lots of information.
2. Automatically detecting disease can be used as a assistive technology by radiologists to increase speed and prevent errors. It can increase accessibility of health-care in remote areas.
3. Localization by weak supervision can help us take leverage of unlabelled/partially labelled data.

# Problem statement

1. We use the NIH Chest X-ray Dataset which comprises of over 1 lakh frontal-view X-ray images with labels for each image specifying the presence/absence of 14 common thorax diseases, and no bounding box information for training.

# Problem statement

1. Given the chest x-ray dataset, the problem is formulated as a weakly supervised multi-label classification and localization problem. The task is given an input x-ray image, we have to detect the diseases (multi-label) present and also localize (draw a bounding box around) around the diseases. It is a weakly supervised task because in the training set, only class labels are present for each image, the bounding box information is not present.

# Core idea

- Train the CNN on the multi-label classification task using a network which preserves spatial information (by using global average pooling).

- Then use the features and weights of the network to obtain class activation mapping (heatmaps).

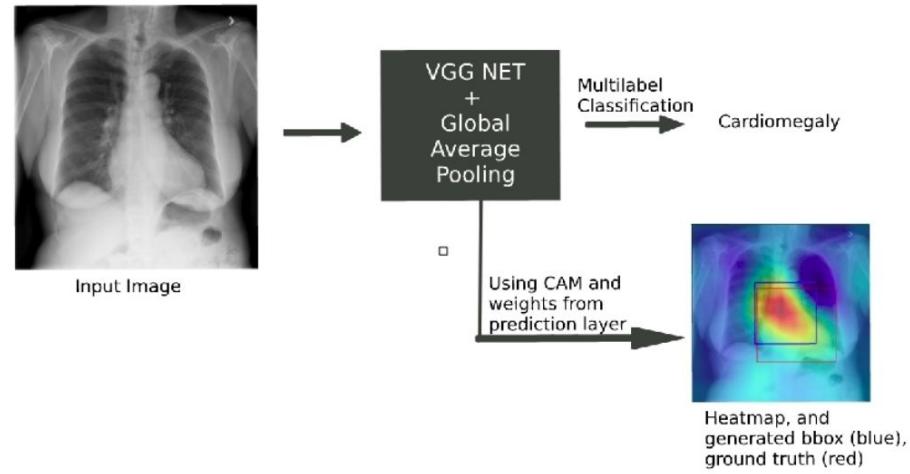


Figure 1: Given an input x-ray image (containing pathology cardiomegaly), our model produces the multilabel classification of the pathology (more than one can be present) present in the xray, and the bounding box localizing the pathology based on the generated heat map.

# Related Work

- Similar work done by Wang et al, the authors who assembled the dataset.
- Class activation mapping concept taken from the paper by Zhou et al - Learning Deep Features for Discriminative Localization



# Method

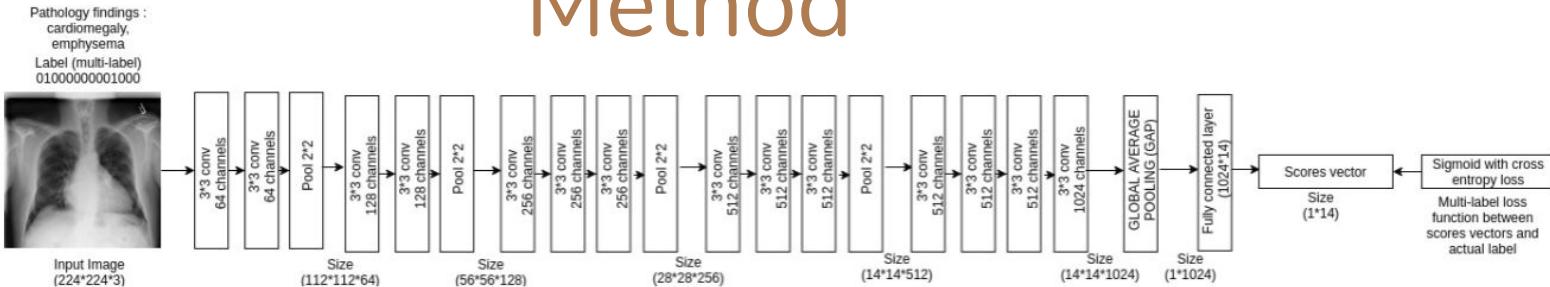
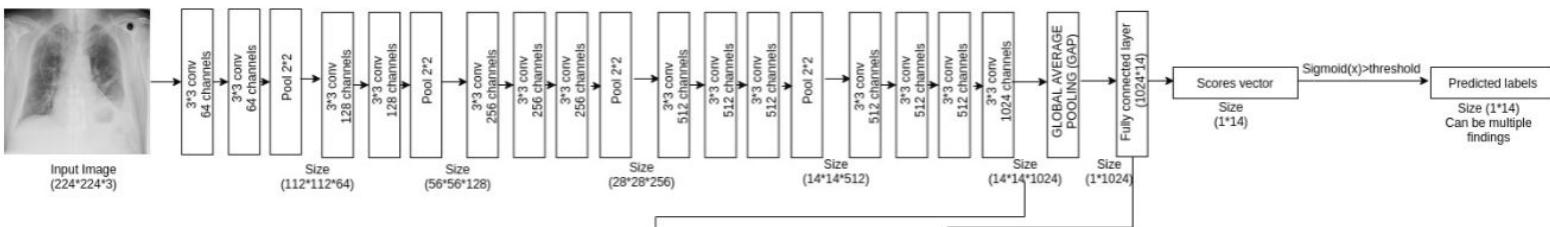


Figure 2: Our model, and the training process



Unified model for weakly supervised multi-label classification and pathology localization.

Figure 3: Our model, and the inference process

# Method - CNN framework

Conv model - Initial layers from pre-trained VGG network pre-trained on Imagenet data.

Removed the fully connected layers present in VGG-net to preserve spatial information in the feature maps.

Added another convolutional layer at the end to increase the depth of the feature maps.

After that we add a global average pooling layer, a prediction layer (fully connected layer) and a loss layer at the end.

# Method - Multi label setup

For representing the labels of the images, we define a 14 dimensional vector  $y = [y_1, y_2, \dots, y_{14}]$  where each  $y_i \in \{0,1\}$  ( $1 \leq i \leq 14$ ).

Each  $y_i$  represents the presence/absence of one disease. The presence of no disease or "Normal" is represented by an all zero vector.

# Method - Loss function

We use the sigmoid cross entropy loss, i.e. apply sigmoid function on the scores obtained from the network to scale them between 0 and 1, and then apply cross entropy loss between the obtained probabilities and the true label.

Loss = sum ( $z * -\log(\text{sigmoid}(x)) + (1-z) * -\log(1-\text{sigmoid}(x))$ ) where x=logits (scores from network) and z = labels.

# Method - Localization task

A class activation mapping or the heatmap for a particular class basically identifies those discriminative spatial locations which are used by the CNN to identify the that class.

The initial parts of the network consists of convolutional and max-pool layers to give convolutional feature map which is fed to a global average pooling layer. Global average pooling averages across one 2-D feature map to generate one value.

So a  $S * S * D$  where D is the number of channels in the feature map gets reduced to  $1 * D$ .

# Method - Localization task

This  $1 \times D$  vector is used as a feature for a fully connected layer which produce the final classification scores.

The classification scores are produced by the weighted sum (through the weights of the fully connected layer) of the  $1 \times D$  vector.

Similarly using the same weights, the weighted sum of the feature maps of the last convolutional layer are used to generate the class activation maps (CAM).

# Method - Bounding box generation

Heatmaps produced from the framework indicate the spatial locations which are likely to contain the pathology.

We normalize the scores of the heatmap between 0 and 1, and apply a simple threshold to consider the values above 0.65 (determined empirically) only.

Then, we generate a bounding box which covers all those locations.

# Method

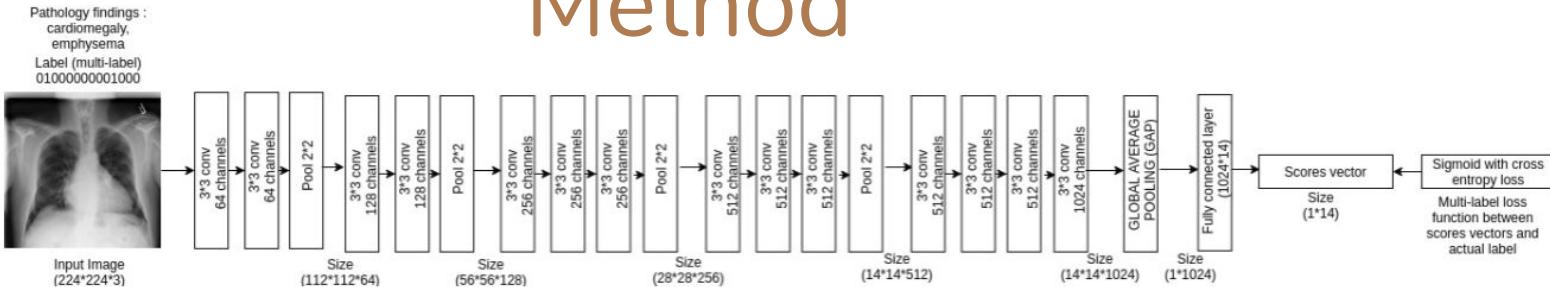
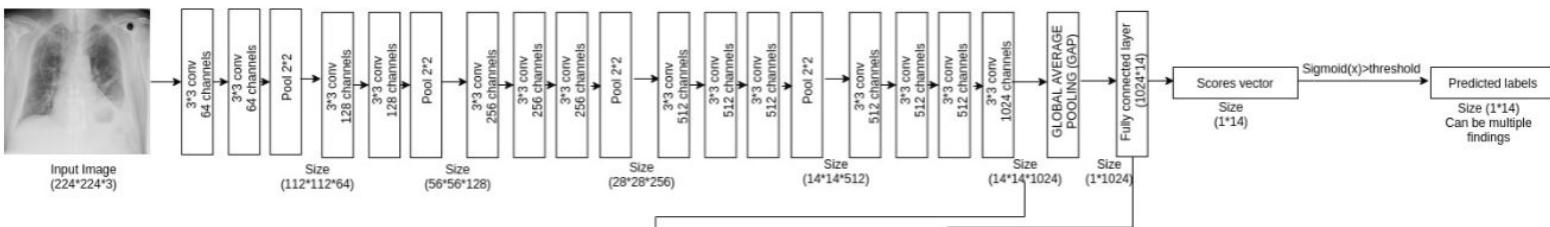
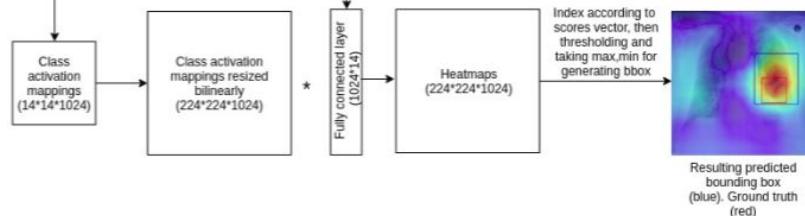


Figure 2: Our model, and the training process



Unified model for weakly supervised multi-label classification and pathology localization.

Figure 3: Our model, and the inference process



# EXPERIMENT - Evaluation protocol

## Classification task

- For the classification problem, different thresholds used - [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1].
- For each threshold value, and for each class, we calculate the true positive rate (tpr) and the false positive rate (fpr). :  $tpr = tp / (tp + fn)$ ,  $fpr = fp / (tn + fp)$
- Then for each class, the area under the ROC curve (AUC) is used as a performance measure.

# EXPERIMENT

## Results

### Classification task

Table 1: Multi-label classification performance (measured by classwise AUC) of our model, baseline given by the original authors (Wang et al) on VGG-net, and the state of the art Chexnet (trained on 121 layers densenet by [18])

| Pathologies        | Our model | Baseline | Chexnet |
|--------------------|-----------|----------|---------|
| Atelectasis        | 0.6893    | 0.6281   | 0.8094  |
| Cardiomegaly       | 0.7197    | 0.7084   | 0.9248  |
| Effusion           | 0.7646    | 0.6502   | 0.8638  |
| Infiltration       | 0.6563    | 0.5896   | 0.7345  |
| Mass               | 0.6744    | 0.5103   | 0.8676  |
| Nodule             | 0.6459    | 0.6556   | 0.7802  |
| Pneumonia          | 0.5194    | 0.5100   | 0.7680  |
| Pneumothorax       | 0.71974   | 0.7516   | 0.8887  |
| Consolidation      | 0.6150    | NA       | 0.7901  |
| Edema              | 0.6898    | NA       | 0.8878  |
| Emphysema          | 0.7128    | NA       | 0.9371  |
| Fibrosis           | 0.5533    | NA       | 0.8047  |
| Pleural_Thickening | 0.6100    | NA       | 0.8062  |
| Hernia             | 0.5116    | NA       | 0.9164  |

# EXPERIMENT - Evaluation protocol

## Localization task

- 1000 images with bounding box labels to evaluate performance.
- Since, the labels in the localization task has only one pathology present corresponding to one x-ray, we take that pathology which has the highest probability.
- Generate the bounding box corresponding to that heatmap.
- Intersection over union calculated with the ground truth box.
- A correct localization is defined if  $\text{IOU} > T(\text{IoU})$  where  $T(\text{IoU}) \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ .
- Accuracy and Average false positive (AFP) calculated for each class and each  $T(\text{IoU})$  value.  $\text{acc} = (\text{tp} + \text{tn}) / (\text{tp} + \text{fp} + \text{fn} + \text{tn})$ ,  $\text{afp} = \text{fp} / (\text{fp} + \text{tn})$

# EXPERIMENT - Results

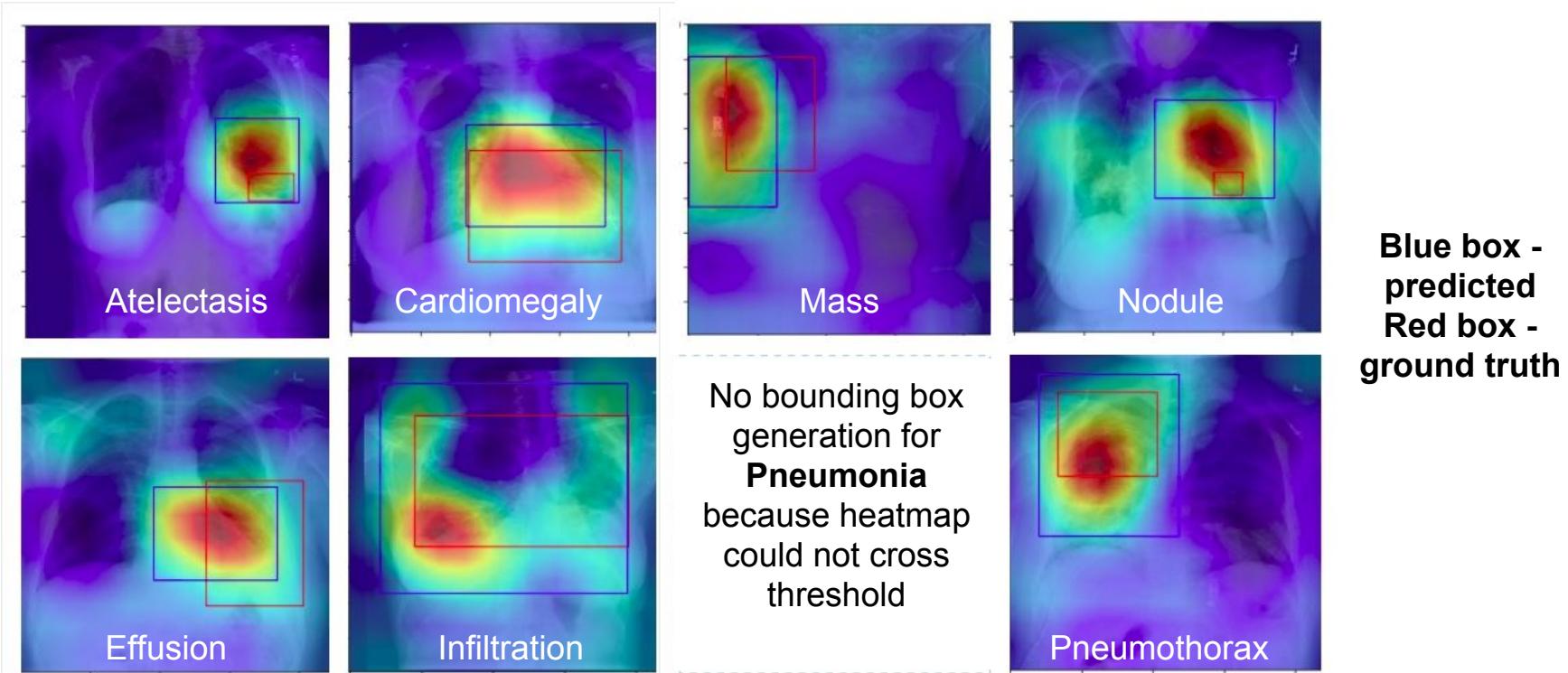
## Localization task

Table 2: Pathology localization accuracy and average false positive (AFP) for each T(IoU) value as mentioned in the evaluation protocol. The diseases are Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural\_Thickening and Hernia

|               | Atel.  | Card.  | Effu.  | Infil. | Mass   | Nod.   | Pneu.  | Pneumo. | Conso. | Edema  | Emph.  | Fibr.  | Pleu.  | Hernia |
|---------------|--------|--------|--------|--------|--------|--------|--------|---------|--------|--------|--------|--------|--------|--------|
| T(IoU) = 0.1  |        |        |        |        |        |        |        |         |        |        |        |        |        |        |
| Acc.          | 0.7488 | 0.9011 | 0.7204 | 0.6670 | 0.8784 | 0.8795 | 0.8659 | 0.8772  | 0.9988 | 0.9954 | 0.9852 | 0.9988 | 0.9931 | 1.     |
| AFP.          | 0.1481 | 0.0228 | 0.2446 | 0.2912 | 0.0591 | 0.0663 | 0.     | 0.0403  | 0.0011 | 0.0045 | 0.0147 | 0.0011 | 0.0068 | 0.     |
| T(IoU) = 0.25 |        |        |        |        |        |        |        |         |        |        |        |        |        |        |
| Acc.          | 0.7375 | 0.9    | 0.7011 | 0.6534 | 0.875  | 0.8795 | 0.8659 | 0.8693  | 0.9988 | 0.9954 | 0.9852 | 0.9988 | 0.9931 | 1.     |
| AFP.          | 0.1592 | 0.0241 | 0.2604 | 0.3017 | 0.0626 | 0.0663 | 0.     | 0.0486  | 0.0011 | 0.0045 | 0.0147 | 0.0011 | 0.0068 | 0.     |
| T(IoU) = 0.5  |        |        |        |        |        |        |        |         |        |        |        |        |        |        |
| Acc.          | 0.7318 | 0.8545 | 0.6852 | 0.6477 | 0.8681 | 0.8795 | 0.8659 | 0.8659  | 0.9988 | 0.9954 | 0.9852 | 0.9988 | 0.9931 | 1.     |
| AFP.          | 0.1647 | 0.0737 | 0.2730 | 0.3059 | 0.069  | 0.0663 | 0.     | 0.0522  | 0.0011 | 0.0045 | 0.0147 | 0.0011 | 0.0068 | 0.     |
| T(IoU) = 0.75 |        |        |        |        |        |        |        |         |        |        |        |        |        |        |
| Acc.          | 0.7318 | 0.8272 | 0.6818 | 0.6454 | 0.8670 | 0.8795 | 0.8659 | 0.8659  | 0.9988 | 0.9954 | 0.9852 | 0.9988 | 0.9931 | 1.     |
| AFP.          | 0.1647 | 0.1012 | 0.2756 | 0.3076 | 0.0706 | 0.0663 | 0.     | 0.0522  | 0.0011 | 0.0045 | 0.0147 | 0.0011 | 0.0068 | 0.     |
| T(IoU) = 0.9  |        |        |        |        |        |        |        |         |        |        |        |        |        |        |
| Acc.          | 0.7318 | 0.8272 | 0.6806 | 0.6443 | 0.8670 | 0.8795 | 0.8659 | 0.8659  | 0.9988 | 0.9954 | 0.9852 | 0.9988 | 0.9931 | 1.     |
| AFP.          | 0.1647 | 0.1012 | 0.2765 | 0.3085 | 0.0706 | 0.0663 | 0.     | 0.0522  | 0.0011 | 0.0045 | 0.0147 | 0.0011 | 0.0068 | 0.     |

# EXPERIMENT - Qualitative results

## Localization task



# Conclusion and future work

- Developed a unified model for multilabel classification and weakly supervised localization.
  - Using CAM, decent performance in localization can be obtained.
- 
- Future work includes experimenting with the loss function (probably with triplet loss to better learn the discriminative features between different pathologies).
  - Using data augmentation to increase the positive examples.

