Detecting Chest Diseases from Chest X-ray Using Deep Learning

Partho Ghosh, Prasun Datta, Raisa Bentay Hossain, Ratul Kundu, Taseen Afrid, Shahed Ahmed, Talha Ibn Mahmud
dept. of Electrical and Electronic Engineering
Bangladesh University of engineering and Technology

***Abstract*—In the sector of screening thoracic diseases till now chest X-rays are the most affordable in radio-logical examination. The pathological information usually depends on the lung and heart regions. However, model training mainly relies on image-level class labels in a weakly supervised manner, which is quite challenging for computer-aided chest X-ray screening. Regarding the continuity, some methods have been come across recently to determine local regions containing pathological information, usually vital for thoracic diseases. We propose a novel deep framework for the multi-label classification of thoracic diseases in chest X-ray images. To exploit disease-specific cues effectively, we locate lung and heart regions containing pathological information by a well-trained pixel-wise segmentation model to generate masks. Compared to existing methods fusing global and local features, we adopt feature weighting to avoid weakening visual cues unique to lung and heart regions. For training such systems, existing deep learning-based algorithms frequently require substantial supervision, such as annotated bounding boxes, which are difficult to gather on a wide scale. We offer PCAM pooling, a unique global pooling procedure for lesion localization that requires just image-level supervision. Our method with pixel-wise segmentation can help overcome the deviation of locating local regions. Evaluated by the benchmark split on the publicly available chest X-ray14 data-set, the comprehensive experiments show that a DenseNet-121 model trained with PCAM pooling beats state-of-the-art baselines. When compared to the localization heat-map obtained by CAM, the probability maps generated by PCAM pooling exhibit distinct and crisp boundaries around lesion. Our proposed network aims to effectively exploit pathological regions containing the main cues for chest X-ray screening.**

***Index Terms*—Chest X-rays; Thoracic Diseases Classification; Pixel-wise Segmentation; Lung and Heart Regions; PCAM pooling; Multi-scale Attention**

I. Introduction

In recent years, deep convolutional neural networks (DCNNs) have significantly improved computer-aided thoracic disease identification based on chest X-ray [1], [2]. The bulk of these approaches are based on a multi-task binary classification problem in which a CNN is trained to predict the probability of various thoracic illnesses. The use of heatmaps or segmentation masks to visualize lesions on chest X-ray is commonly advocated in clinical practice to provide interpretable supports for categorization findings. Beyond image-level labeling, proper lesion localization typically involves the use of bounding boxes to train CNNs under strict supervision.

On the other hand, accurately labeling lesion locations is difficult, time-consuming, and impossible to execute on a large scale. For example, the CheXpert dataset, which is one of the biggest publicly accessible chest X-ray datasets, comprises over a hundred thousand pictures with image-level labels, with just a few thousand images additionally annotated with bounding boxes [1] for benchmarking. As a result, based on image-level labeling, poorly guided lesion localization on chest X-ray remains a difficult but critical topic for computer-aided thoracic illness diagnosis. The recent work of Class Activation Map (CAM) shows that CNNs trained on nature photos with only image-level supervision have great localization capabilities [1]. CAM and its variants have also been used to locate lesions on chest X-ray images.

In this report, we present a unique and easy expansion to the CAM-based framework for image-level lesion localization on chest X-ray. We propose Probabilistic-CAM (PCAM) pooling, a new global pooling technique that explicitly uses CAM for localization during training in a probabilistic manner. PCAM pooling adds no new training parameters and is simple to deploy. On the CheXpert dataset [1], we test the effectiveness of PCAM pooling for lesion localization with image-level supervision. In both the classification and localization tasks, a DenseNet-121model trained with PCAM pooling considerably outperforms the CheXpert baseline. In comparison to the normal class activation map, the probability maps created by PCAM pooling feature distinct and strong borders surrounding lesion sites, according to qualitative visual assessment.

II. Dataset

In this experiment, CheXpert - a large dataset of chest X-rays was used which features uncertainty labels and radiologist-labeled reference standard evaluation sets. It is a large public dataset for chest radiograph interpretation, consisting of 224,316 chest radiographs of 65,240 patients. It has been retrospectively collected the chest radiographic examinations from Stanford Hospital, performed between October 2002 and July 2017 in both inpatient and outpatient centers, along with their associated radiology reports which we use to test lesion localization with image-level supervision. The official train valid set was randomly divided into 75 percent training and 25 percent validation [3]. We assess the classification task on the 14 diseases on the official test set and the localization job on the 8 diseases with bounding boxes on the official test set. We train models that take as input a single-view chest radio-graph and output the probability of each of the 14 observations. When more than one view is available, the

models output the maximum probability of the observations across the views.

III. First Baseline Model : Probabilistic-CAM Pooling

A new global pooling operation that explicitly leverages CAM for localization during training in a probabilistic manner, which is called "Probabilistic-CAM (PCAM) Pooling." The network is named after the concept of probabilistic-CAM (PCAM). A fully convolutional backbone network analyses the input chest X-ray picture and outputs a feature map, which is then used to train the network [4]. Each feature embedding within the feature map is then subjected to a fully connected (fc) layer, which is implemented as a 1 1 convolutional layer, in order to generate a class activation score that monotonically measures the disease likelihood of each embedding for each label of thoracic disease, such as "Pneumonia." To distinguish ourselves from traditional methods that employ only the class activation score for localization, we bound it with the sigmoid function and interpreted the result as the disease probability of each embedding, as opposed to the normal technique. Once this is done, the resulting probability map is normalized to the attention weights of each embedding, which is done in accordance with the multiple-instance learning (MIL) framework, which was used to pool the initial feature map through weighted average pooling.

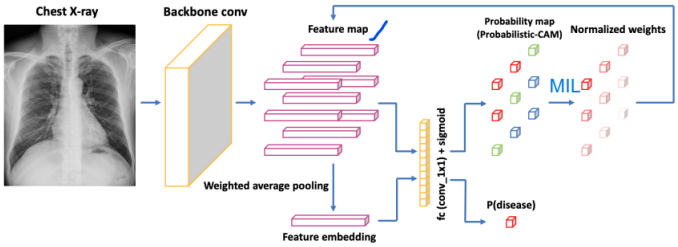

Fig. 1. Global Probabilistic-CAM(PCAM) Pooling Network

After passing through the same fc layer as previously described, the pooled embedding is used to create illness probabilities at the image level for training. The probability map is used directly for lesion localization during inference time, and a simple probability thresholding procedure is applied in order to determine the disease regions and bounding boxes.

IV. Second Baseline Model

A supervised multi-label classification framework based on deep convolutional neural networks (CNNs) is presented by Hieu H. Pham et. al. for predicting the presence of 14 common thoracic diseases and observations [5]. They tackle this problem by training state-of-the-art CNNs that exploit hierarchical dependencies among abnormality labels. They also propose to use the label smoothing technique for a better handling of uncertain samples.

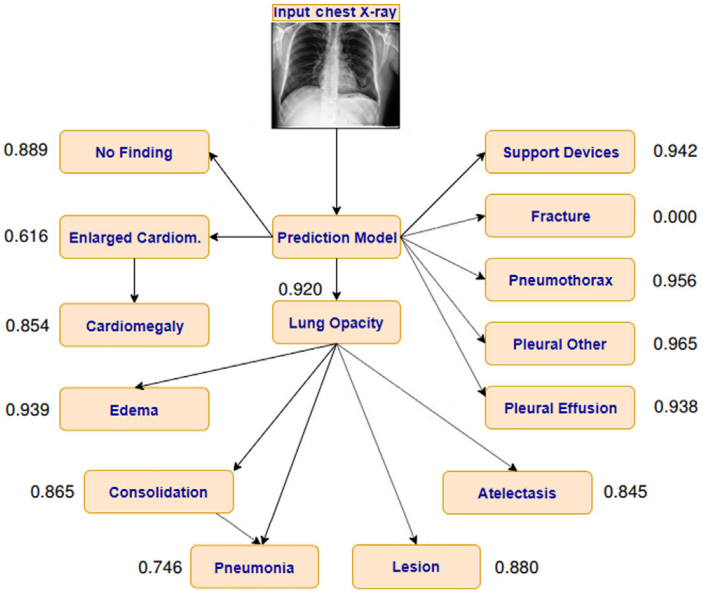

Fig. 2. Illustration of our classification task, which aims to build a deep learning system for predicting probability of presence of 14 different pathologies or observations from the CXRs.

A. Conditional training to learn dependencies among labels

Labels are commonly grouped into trees or directed acyclic graphs in medical imaging (DAG). Domain specialists, e.g. radiologists in the case of CXR data, build these hierarchies. CXR diagnoses or observations are frequently label dependent. So use it during model training and prediction. However, most present CXR classification methods treat each label independently, ignoring label structure. Flat classification methods are a subset of this group. When applied to hierarchical data, a flat learning model fails to simulate illness dependence. For example, in Fig. 2, cardiomegaly suggests enlarged cardiomediastinum. Also, some lower-level labels, particularly leaf nodes, have extremely few positive examples, making the flat learning model readily biased towards the negative class. [5]

The hierarchies are constructed in a way that the root nodes correspond to the most general classes (like Opacity) and the leaf nodes correspond to the most specific ones (like Pneumonia). One common approach to exploit such a hierarchy is to (1) train a classifier on conditional data, ignoring all samples with negative parent-level labels, and then (2) add these samples back to finetune the network on the whole dataset.

B. Leveraging uncertainty in CXRs with label smoothing regularization

A new advance in machine learning called label smoothing regularization (LSR) is proposed for a better handling of uncertainty samples. The method has been efficiently used to boost the performances of multi-class classification models via smoothing out the impulse-like label vector of each sample. We adapt this idea of LSR to the binary classification of a CXR into positive/negative for each of the 14 categories [9].

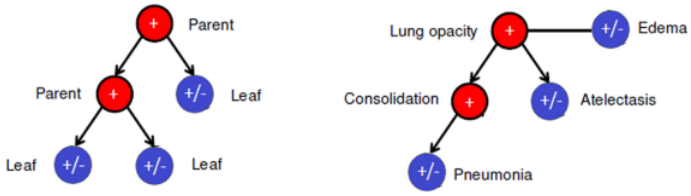

Fig. 3. Illustration of the key idea behind the conditional training (left). In this stage, a CNN is trained on a training set where all parent labels (red nodes) are positive, to classify leaf labels (blue nodes), which could be either positive or negative. For example, we train a CNN to classify Edema, Atelectasis, and Pneumonia on training examples where both Lung Opacity and Consolidation are positive.

Our main goal is to exploit the large amount of uncertain CXRs and, at the same time, to prevent the model from overconfident prediction of the training examples that might contain mislabeled data. Specifically, the U-ones approach is softened by mapping each uncertainty label (1) to a random number close to 1.

V. Experimental Method

Our proposed deep framework covers several parts: a feature extractor, attention module, a pixel-wise segmentation model, and a feature weighting module. The feature extractor is a backbone that creates a feature map. The multi-scale attention module helps the feature extractor to focus on salient regions and detect subtle texture abnormality. We pass the feature map through attention module and concatenate the output with that feature map. After passing through a sigmoid it gives us a probability map. Simultaneously, the well-trained pixel segmentation model identifies areas of the lung and heart, following binarized as a global mask in which pixels are 1 for lung and heart region and 0 for other regions. Then we conduct an element-wise summation operation on the probability attention map and the global mask to generate a local attention map. By weighing the lung and heart region features, the local attention map only contains visual cues unique to the lung and heart region containing pathological information and discards features of non-lung and heart regions by zeroing operation [10]. Following the local attention map, an average pooling layer and a fully connected layer are introduced to train disease-specific probability by binary cross-entropy loss.

We used Probabilistic Class Activation Map (PCAM) pooling, a global pooling operation for lesion localization with only image-level supervision. PCAM pooling explicitly leverages the excellent localization ability of CAM during training in a probabilistic fashion.

A. A feature extractor

The feature extractor consists of a backbone. Each chest X-ray image X is resized into 3 × 256 × 256 and firstly inputted into the backbone. We use the pre-trained 121-layer DenseNet as the backbone. We take out the last convolutional feature map from backbone and input it to the SAM attention module [5].

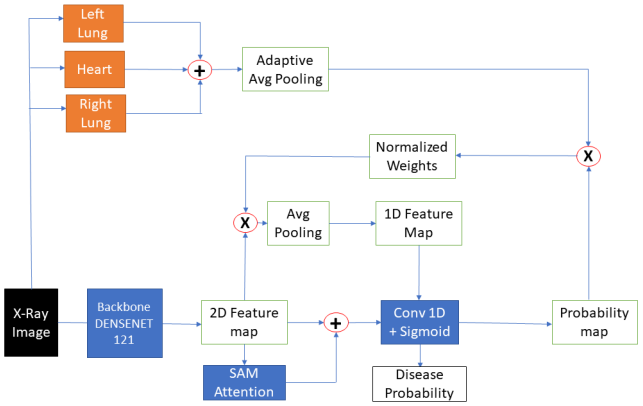

Fig. 4. Our proposed Model Architecture.

B. Attention Modules

We have used spatial attention module (SAM) here. The attention module takes the feature map and generates a spatial attention map by utilizing the inter-spatial relationship of features. Different from the channel attention, the spatial attention focuses on where an informative part is, which is complementary to the channel attention. To compute the spatial attention, we first apply average-pooling and max-pooling operations along the channel axis and concatenate them to generate an efficient feature descriptor. The output of the attention module is then concatenated with the feature map. After a sigmoid activation function, we find the probability map.

C. A pixel-wise segmentation model

A well-trained pixel segmentation model locates the lung and heart regions to binarize a mask in which pixels are 1 for lung and heart regions and 0 for other regions. We apply U-Net [6] to train a segmentation model for the left lung, right lung, and heart on the JSRT dataset by using dice loss. The dice loss is formulated as:

$$dice = \frac{2|Mgt \cap Mprob|}{|Mgt| + |Mprob|} \quad (1)$$

where **Mgt denotes the ground truth mask, and Mprob is the predicted mask.** The dice loss is minimized for optimization and the model with the smallest loss was saved. The image pre-processing of U-Net follows the same pipeline of the feature extractor to enable automatic region segmentation for the chest X-ray14 dataset. We first input the chest X-ray image X into the well-trained segmentation model to generate three pixel-wise masks for the left lung, right lung, and heart. Then we merge the three pixel-wise masks into a pixel-wise mask Mg in which pixels are either 1 for the lung and heart regions or 0 for other regions by pixel-wise summation. The pixel-wise mask Mg further is resized into a size equal to the width and height of the probability attention map by adaptive average pooling.

D. Feature weighting

The probability attention map is taken out from the backbone and SAM attention module of the image classifier. Further, we generate a local attention map from the probability attention map and the pixel-wise mask by element-wise multiplication. We introduce the logical AND operator on the global attention map and the pixelwise mask. The local attention map contains the zero pixels of non-lung and heart regions and the non-zero pixels of the lung and heart regions. Hence, only the pixel values of the lung and heart region containing pathological information in the local attention map are embedded into the average pooling layer for label prediction by a channel-wise average operation, and the pixel values of other regions in the attention map are zeroed [5]. **The feature weighting for the global attention map Fg and the pixel-wise mask Mg** is defined as:

$$F1 = Fg \oplus Mg \tag{2}$$

With the help of the SAM attention module, the probability attention map effectively learns the salient information from the chest X-ray image, containing the discriminative information in the lung and heart. The pathological regions are typically located in the lung and heart, hence, we introduce the binary masks on the probability attention map to generate the local attention map [8]. The generated local attention map suppresses the information of other regions and remains the information of the lung and heart regions. By logical AND operation, we locate features of the lung and heart regions containing pathological information.

E. PCAM Pooling

The main idea of PCAM pooling is to explicitly leverage the localization ability of CAM through the global pooling operation during training.A fully convolutional network trained for multi-task binary classification, the class activation map of a particular thoracic disease is given by

$$s(i,j) = w(X(i,j) + b)(i,j) \in H, W \tag{3}$$

X(i,j) is the feature embedding of length C at the position (i, j) of a feature map X with shape (C, H, W) from the last convolutional layer. w, b are the weights and bias of the last fc layer for binary classification. In other words, s(i,j) is the logit before sigmoid function under the binary classification setting. s(i,j) monotonically measures the disease likelihood of X(i,j) , and is used to generate the localization heatmap after the model is trained in the standard CAM framework [7]. Here we have used PCAM pooling instead of average pooling layer. The 2D feature map from the backbone and the weighted local attention map undergo through a logical AND operation. The resultant is then sent to the PCAM pooling layer. After going through a fc (fully connected) layer, the diseases probability is found.

VI. Experiments and Results

In the PCAM pooling paper, x-ray images of size 512*512 are used whereas in our implementation images of size 256*256 are used. Eventually, an AUC score of 0.896 was reported in the paper and in our experimental part 0.892 was achieved. Our target was to beat the baseline score of 0.892 and for that we did a lot of experiments. Firstly, we looked

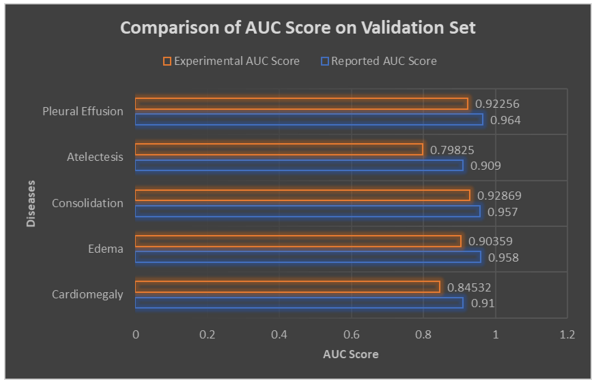

Fig. 5. Comparison of classwise AUC Score between reported and experimental part.From implementing this baseline, we have achieved a mean AUC of 0.88 on the validation set

at the different backbone architecture such as Densenet201, Densenet169, InceptionV3 and their performance on the dataset which scaled to image size of 256*256.

Secondly, we focused on several attention modules, including Class Activation Map (CAM), Feature Pyramid Attention (FPA), and Special Attention Map (SAM) (SAM). The performance of the CAM attention map is not up to par; however, the SAM attention map produces an AUC score of 0.893 by increasing the baseline value.

Next, we implemented the SAM Attention module and Inception V3 model since they functioned well in our prior phase trials. Let's explore whether we can attain our goal score using both the SAM attention module and the Inception V3 backbone architecture. Unfortunately, our previous high score of 0.893 was not improved.

So, with the larger picture size of 320*320, we installed this attention module and backbone model to see whether it can achieve the desired outcome. However, it did not improve.

We transformed the positional weights for the next phase of our experiments because we noticed in previous experiments that class 1 and 4 did not perform well in terms of accuracy score, so we changed the positional weights for class 1 and 4 to see if we could achieve our target, but the best possible result we got was 0.893, which was also our previous best.

Conclusion

In this work, we propose a novel deep framework for the multi-label classification of thoracic diseases in chest X-ray images. The proposed network aims to effectively exploit pathological regions containing the main cues for chest X-ray screening. We present a feature extractor equipped with a multi-scale attention module to effectively learn pathological

information from chest X-ray images. At the same time, we apply the pixel-level segmentation to identify the lung and heart regions containing pathological information to overcome location deviation. Then, we adopt the feature weighting strategy to filter out the non-lung and heart regions. Based on our deep framework, the class-probability layer mainly rely on the information of the lung and heart regions. Evaluated on the benchmark split of the cheXpert dataset, we establish a new state-of-the-art baseline. Our proposed network has been used in clinic screening to assist the radiologists. Chest X-ray accounts for a significant proportion of radiological examinations. It is valuable to explore more methods for improving performance.

ACKNOWLEDGEMENT

The authors would like to thank Shahed Ahmed and Talha Ibn Mahmud for providing this opportunity to work on this interesting topic.

REFERENCES

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2097–2106,2017.

[2] Y. Tang, X. Wang, A. P. Harrison, L. Lu, J. Xiao, and R. M. Summers. Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs. In International Workshop on Machine Learning in Medical Imaging, pages 249–258. Springer, 2018.

[3] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 590–597 (2019)

[4] Ye, Wenwu, et al. "Weakly supervised lesion localization with probabilistic-cam pooling." arXiv preprint arXiv:2005.14480 (2020).

[5] Pham, H. H., Le, T. T., Tran, D. Q., Ngo, D. T., Nguyen, H. Q. (2021). Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. Neurocomputing, 437, 186-194.

[6] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015, p. 234–241.

[7] Fang, J., Xu, Y., Zhao, Y., Yan, Y., Liu, J., Liu, J. (2021). Weighing features of lung and heart regions for thoracic disease classification. BMC Medical Imaging, 21(1), 1-12.

[8] Yan, C., Yao, J., Li, R., Xu, Z., Huang, J.: Weakly supervised deep learning for thoracic disease classi
cation and localization on chest x-rays. In: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, pp. 103110 (2018)

[9] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., et al.: Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: Proceedings of the AAAI Conference on Arti
cial Intelligence, vol. 33, pp. 590597 (2019)

[10] Tang, Y., Wang, X., Harrison, A.P., Lu, L., Xiao, J., Summers, R.M.: Attention-guided curriculum learning for weakly supervised classi
cation and localization of thoracic diseases on chest radiographs. In: International Workshop on Machine Learning in Medical Imaging, pp. 249258 (2018). Springer