A Weakly Supervised Adaptive DenseNet for Classifying Thoracic Diseases and Identifying Abnormalities

Bo Zhou
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA

Yuemeng Li
Department of Bioengineering
University of Pennsylvania
Philadelphia, PA, USA

Jiangcong Wang
Department of Bioengineering
University of Pennsylvania
Philadelphia, PA, USA

Abstract

We present a weakly supervised deep learning model for classifying thoracic diseases and identifying abnormalities in chest radiography. In this work, instead of learning from medical imaging data with region-level annotations, our model was merely trained on imaging data with image-level labels to classify diseases, and is able to identify abnormal image regions simultaneously. Our model consists of a customized pooling structure and an adaptive DenseNet front-end, which can effectively recognize possible disease features for classification and localization tasks. Our method has been validated on the publicly available ChestX-ray14 dataset. Experimental results have demonstrated that our classification and localization prediction performance achieved significant improvement over the previous models on the ChestX-ray14 dataset. In summary, our network can produce accurate disease classification and localization, which can potentially support clinical decisions.

1. Introduction

Large scale annotated visual datasets have boosted performance of deep learning methods on many challenging computer vision problems [1, 2, 3]. Tasks like object detection, classification, tracking, and segmentation have been successfully tackled by techniques built on top of these large-scale dataset with annotations [4, 5, 6]. There are increasing numbers of applications utilizing deep learning methods in medical imaging analysis over the last decade [7]. In clinical procedures, visual evidence such as segmentation or spatial localization of abnormal regions that supports the diagnosis results, is an vital part of clinical diagnosis. This provides a comprehensive interpretation of diagnosis results and potentially decreases the false positive rate.

In this work, we focus on the automatic disease diagnosis and localization in chest radiography released by [8] named ChestX-ray14, which is one of the largest public chest radiography dataset with image-level disease labels and contains a small subset of region-level disease localizations (bounding boxes). Our goal is to develop a deep learning scheme capable of both classifying the disease and localizing the associated lesion sites. Divergent from standard strongly supervised object detection, our model does not require ground truth localization annotations during training. Firstly, we adopted and modified a pre-trained classification CNN as our front-end feature extractor [9]. The pre-trained front-end encodes the information from a large perceptual field. Then, after passing through a simple bridging structure, the extracted feature from the front-end are fed into a customized two-stage pooling network structure, which produces both classification and associated localization simultaneously.

Both quantitative and qualitative visual evaluations show that our proposed model obtains significant improvement over the previous published state-of-the-art results on disease classification and localization. Visual evaluations indicate a strong alignment and correspondence between the clinical annotations and the predicted disease candidate regions shown in figure 1.

Contributions: The contributions of this work are two-folds. First, we developed a weakly supervised end-to-end learning structure that learns from chest radiography images containing multiple common thoracic diseases by explicitly searching over possible disease features and locations in the image. Second, we performed extensive experimental analysis of our model on the large-scale ChestX-ray14 datasets. With the optimal parameter settings found from our analysis, we observed that our model (i) can predict accurate thoracic disease classifications (ii) output better approximate disease locations than previous methods with only image-level labels available.
Figure 1. Examples of the disease localizations generated from our network. The localization output from our network (red bounding box), that trained only with image-level annotation, match with the ground truth localization (green bounding box).

2. Related Work

**CAD for chest radiography:** The chest radiography is the most ordered and common radiological examination for chest diseases. It is a low-cost, low radiation and fast imaging exam. Recent studies also demonstrated the chest radiography’s application on detection and evaluation of coronary artery diseases, which are usually evaluated using Computed Tomography (CT) with expensive cost and high radiation dose [10, 11, 12]. CAD techniques have been widely applied in chest radiography for task such as automatic diagnosis and patient image retrieval. Previously, Bar et al. adapted the Decaf-Net for 8 thoracic diseases classification on a relatively small chest x-ray dataset [13]. Lajhani et al. proposed the DCNN for tuberculosis classification task. Their model ensembled both Alex-Net and Google-Net and achieved impressive results [14]. Anavi et al. has worked on image retrieval in medicine, specifically for chest radiography given the pathology [15]. All the aforementioned research showed promising results. However, they only conducted their experiments on relatively small dataset, ranging from 10 to 500 images.

**ChestX-ray14:** Recently, National Institute of Health (NIH) released one of the largest public chest radiography dataset, consisting of 108,948 posterior-anterior view images from 32,797 patients with eight major chest diseases [8]. A small subset of this dataset is provided with hand labeled bounding boxes for evaluation. Lately, NIH further expanded this dataset to 112,120 frontal-view images with 6 additional thoracic diseases, named ChestX-ray14. Several deep learning methods have been addressed the application of CNN on this dataset for thoracic disease classification and localization [16, 17, 18]. Wang et al. [8] applied a pre-trained Res-Net as the backbone to generate heatmap as localization, and subsequently used a global maxpooling to obtain classification. However, most of the localizations generated by their network mismatched with the ground-truth bounding boxes due to only limited extracted feature is used. Yao et al. [16] used DenseNet [9] to extract features. To harness the correlation between some of the diseases, they used a LSTM module to repeatedly decode the feature vector from a DenseNet [9] front end and produced one disease prediction at each step. They achieved improved results compared to the baseline in [8]. One of the most recent work from Rajpurkar et al. achieved a good multi-label classification results by fine-tuning a pre-trained DenseNet-121 [9, 18]. Their classification results outperformed the previous methods [8, 16, 17]. However, there is no localization component and weakly supervision in their work. Another most recent work from Li et al. [17] used a pre-trained Res-Net to extract features and divided them into patches. They passed the extracted patches through a fully-convolutional classification CNN to obtain a disease probability map which is supervised by both image-level annotations and limited pixel-level annotations. In the mean time, a classification score was acquired by multiplying all probability values together. They achieved significantly better localization results as compared to the baseline [8] for certain diseases. But this approach requires certain amount of bounding box training data for improving their model performance.

**Weakly supervised learning:** The ChestX-ray14 dataset [8] contains mostly classification labels but few bounding boxes annotations. Therefore, we seek a model that is capable of localizing the diseased regions given only image wise labels. This falls into category of weakly supervised learning (WSL), which often refers as the task of capturing object location through a customized deep learning model that trained with only image-level labels. Oquab et al. [19] proposed a WSL scheme. They used a pre-trained CNN to generate class probability map across spatial locations and applied max-pooling across spatial locations to get a single binary score vector. With their strategy, they were able to get both decent classification accuracy and localization mAP on the benchmark dataset [20, 3]. Recently, Durand et al. proposed a more sophisticated WSL approach.
3. Method

In this work, we aim to use chest radiographs as input with only image-wise labels to train a model that generates classification for thoracic disease along with the localization heatmaps. The architecture of our proposed network can be summarized into three parts: 1) using an Adaptive DenseNet to generate feature maps and 2) using a bridging layer to convert the feature maps into class-specific sub-maps 3) using a WSL pooling structure to pool the sub-maps into class-specific heatmaps and single probability score for each disease.

3.1. Adaptive DenseNet

The Adaptive DenseNet inherits the basic structure from DenseNet [9]. The dense block of DenseNet consists of direct connections from preceding layers to all subsequent layers which pass image information in deep model’s training. We maintained the DenseNet structures before the fully connected classification layers as our basic structure. However, the output feature map size after the final dense block is greatly decreased (7 × 7), losing too much spatial resolution as compared to the input image (224 × 224), which is not suitable for our localization task. To overcome this, firstly, we took out the average pooling operation at the third transition layer to achieve a fine resolution. Secondly, in order to maintain the same receptive field of each convolution kernel, we dilated all kernels in the fourth dense block. The Adaptive DenseNet outputs a feature map with spatial size of 14 × 14. Details of our Adaptive Densenet is shown in Figure 2. In this work, we chose the DenseNet-169 as our basic structure. Our adaptation can be applied to other DenseNet structures as well, such as DenseNet-121 and Densenet-201. With the Adaptive DenseNet, we were able to acquire a fine resolution feature map.

[21]. They presented the idea of class pooling based on the assumption that an object inherently can be decomposed into different sub-maps to capture better semantic segmentation in WSL. Their experimental results tested on PASCAL 2007, PASCAL 2012, MS COCO, 15 Scene, MIT67 [22, 20, 3, 23, 24] indicated the advantage of such design and achieve significant improvement over [19]’s model.
### 3.2. Bridging layer

To connect the output of the Adaptive DenseNet and the pooling structure described in section 3.3, a bridging layer is proposed and used here. Specifically, the output of the Adaptive DenseNet is transformed by a $1 \times 1$ convolutional layer, which learns the mapping from the Adaptive DenseNet’s output feature maps to class specific sub-maps. Through the bridging layer, the channel size of the feature map is transferred from $1664$ (channel size of Adaptive DenseNet’s output) to $M \times C$, as shown in Figure 3. This bridging layer connects the Adaptive DenseNet and the WSL pooling structure together to form the overall network.

### 3.3. WSL pooling

To obtain accurate disease classification along with disease localization heatmap, we integrated a customized weakly supervised learning strategy into our network. Similar to [21], we implemented two stages of global pooling layers to summarize all information for each class in feature map outputted from the Adaptive DenseNet, which include: 1) a class-wise pooling within the same class, and 2) a spatial pooling along different classes. The class-wise pooling summarizes submaps into final localization heatmap, and the spatial-wise pooling summarizes heatmap into classification. Details of the pooling designs are demonstrated in Figure 4.

**class-wise pooling:** The class-wise pooling combines the $M$ submaps from each of the $C$ classes from the bridging layer. In specific, our class-wise pooling combines the $M$ maps for all disease classes independently through Equation 1.

$$D_{i,j}^c = \frac{1}{M} \sum_{m=1}^{M} D_{i,j}^{c,m}$$  \hspace{1cm} (1)$$

where $M$ is the number of sub-maps with the assumption that each sub-map contains different features of the corresponding disease. Unlike traditional WSL with only one map considering only one extracted feature, our multiple maps strategy is useful for the disease classification/localization tasks by utilizing multiple extracted features. $H$, $W$, $C$ are feature map’s height, width, and the number of class, respectively. The input feature $D$ with size of $H \times W \times MC$ is divided into $C$ branches with each size of $H \times W \times M$. Then, the sub-maps information are composed into one final feature map for each class through an average pooling. This results in a transformed class heatmap of size $H \times W \times C$.

**spatial-wise pooling:** Given a class heatmap with size of $H \times W \times C$, we used a spatial-wise pooling to generate our final classification output. Our spatial-wise pooling extracts a classification vector ($1 \times 1 \times C$) from the class
we looked at regions with the highest probability of simply choosing the max gives the network higher performance. We introduced a weighting factor to control the relative importance between these two terms. During training phase, we have:

\[ d_{final}^c = d_{r+}, \]

where \( d_{r+} \in D^c_+ \) is a single element uniformly random sampled from \( D^c_+ \). During testing phase, we have:

\[ d_{final}^c = \frac{1}{k} \sum_{d_+ \in D^c_+} d_+ + \alpha \left( \frac{1}{k} \sum_{d_- \in D^c_-} d_- \right) \]

where \( d_+/d_- \) are elements in \( D^c_+/D^c_- \) and \( \alpha \) is the weighting factor added on bottom scores.

### 3.4. Network’s training

For network initialization, we transferred the weights from the pre-trained models on ImageNet for our Adaptive DenseNet. Then, we randomly initialized parameters for bridging and WSL pooling layers. We used learning rate of 0.002 with weight decay of 0.1 for every 10 epochs during training. Our model were implemented with Pytorch after each convolution and additional drop out of rate of 0.1 for each DenseBlock inside Adaptive DenseNet to prevent overfitting. Since there is a large data imbalance between each class, we trained our model with weighted binary cross entropy loss:

\[ L(X, y) = -w_+ \cdot y \log p(Y = 1 | X) - w_- \cdot (1 - y) \log p(Y = 0 | X) \]

where \( w_+ \) represents percent of positive sample and \( w_- \) represents percent of negative sample among all dataset.

### 3.5. Disease classification and localization

The heatmap obtained from our class-wise pooling is further used as a reference map to generate disease bounding boxes. A thresholding followed by connected component analysis are applied for each disease class to generate bounding boxes. Empirically, we used a threshold value of 0.8 for “Cardiomegaly” and threshold values of 0.9 for the rest of classes to obtain the best results. The disease classification score is acquired from the spatial-wise pooling on the heatmap. In short, our network structure is designed for both disease classification and localization tasks, yet only use chest radiographs along with image-level annotations for our training.

### 4. Experiments and Results

#### 4.1. Data for experiments

We performed our experiment using the ChestX-ray14 dataset. The diseases in the corresponding chest radiograph are diagnosed by radiologists and the labeled ground truth are obtained through text mining on the patients’ diagnostic reports. We used exactly same published data split as in [8, 16]. The dataset is divided into training (70%), validation (10%) and testing (20%). The original resolution of image is 1024 × 1024 and was downscaled to 256 × 256. In order to fit dataset into ImageNet pre-trained models, we normalized the image by mean and standard deviation of the images from ImageNet. We used random crop of size 224 × 224 from the downscaled image as the network input for training and we used center crop of same size for testing. For localization validation, the ChestX-ray14 dataset contains 880 images with 983 disease bounding boxes annotated by board-certified radiologists. We calculated the intersection over union (IoU) between our prediction and the ground truth bounding boxes.

#### 4.2. Disease classification

We calculated the Area under Receiver Operating Characteristic curve (AUROC) for each class to evaluate the classification performance of our model. Our results are selected based on the best performance of classifier for each class throughout our training iterations. Our classification performance is compared with other state-of-the-art results [8, 16, 17, 18]. The experiment is tested with the same dataset splitting as demonstrated in [8, 16, 18].

In all 14 diseases, our model achieved better classification AUROC than [8, 16] as shown in Table 1. In comparison to [17] with supervision from both image-level labels and bounding box annotations, our model achieved significant better classification AUROC in 13 out of 14 diseases.
Table 1. AUROC is used to evaluate model classification performance for 14 diseases. The best results are shown in bold text. In order to obtain a comprehensive comparison, performance of all previous networks to our knowledge are included. All results are obtained from their latest update.

| Pathology          | Wang [8] | Yao [16] | Li [17] | Rajpurkar [18] | Ours     |
|--------------------|----------|----------|---------|----------------|----------|
| Atelectasis        | 0.7003   | 0.772    | 0.80    | 0.8094         | 0.8121   |
| Cardiomegaly       | 0.8100   | 0.904    | 0.87    | **0.9248**     | 0.9066   |
| Effusion           | 0.7585   | 0.859    | 0.87    | 0.8638         | **0.8786**|
| Infiltration       | 0.6614   | 0.695    | 0.70    | **0.7345**     | 0.7065   |
| Mass               | 0.6933   | 0.792    | 0.83    | **0.8676**     | 0.8534   |
| Nodule             | 0.6687   | 0.717    | 0.75    | 0.7802         | **0.7852**|
| Pneumonia          | 0.6580   | 0.713    | 0.67    | 0.7680         | **0.7810**|
| Pneumothorax       | 0.7993   | 0.841    | 0.87    | 0.8887         | **0.8911**|
| Consolidation      | 0.7032   | 0.788    | 0.80    | 0.7901         | **0.8115**|
| Edema              | 0.8052   | 0.882    | 0.88    | 0.8878         | **0.8953**|
| Emphysema          | 0.8330   | 0.829    | 0.91    | 0.9371         | **0.9373**|
| Fibrosis           | 0.7859   | 0.767    | 0.78    | 0.8047         | **0.8187**|
| Pleural Thickening | 0.6835   | 0.765    | 0.79    | **0.8062**     | 0.7792   |
| Hernia             | 0.8717   | 0.914    | 0.77    | 0.9164         | **0.9487**|

In particular, our model obtained more than 3% higher AU-ROC in "Cardiomegaly", "Nodule", "Pneumonia", "Fibrosis" and "Hernia". Comparing our results with [18]'s work which only focus on classification, our model outperformed theirs in 10 out of 14 diseases.

In Table 1, our network performed better on diseases classification with large lesions than one with smaller lesions. "Cardiomegaly" and "Emphysema" are the representative classes for disease with large lesion and our model achieved classification score over 90% on both. Although our network performed less accurate on identifying smaller lesions such as mass and nodule, our model is improved from [8, 16, 17, 18] on these classes as well. Unlike the traditional maxpooling which only output the maximal response from a network output, our class-wise pooling can pool multiple response from the submaps which contain different feature responses of the target lesion. In this case, the classification of the large lesions that potentially contains more appearance feature than small lesions can be better improved with our customized pooling methods.

In addition, we performed hyper-parameter searching for best classification performance (details shown in Supplemental Materials). In class-wise pooling, we tested number of sub-maps (M) from number 2 to 18 and found 14 sub-maps yielded best result. We then kept this setting for the rest of experiment. In spatial pooling, we tested hyper-parameters on two different strategies separately. During training, we tested our model using top $k_+$ scores, $k_+$ ranging from 1 to 20. We found that $k_+$ at 10 yielded best network. During testing, we tested $k_+$ from 1 to 20 and $k_-$ from 1 to 25. We found that both $k_+$ and $k_-$ at 15 yielded best classification result. We tested $\alpha$ ranging from 0.25 to 1 and $\alpha$ equals at 1 gave the best importance balance between $k_+$ and $k_-$.  

4.3. Disease localization

We generated our predicted bounding boxes by applying a naive thresholding ($T = 0.1 \sim 0.7$) on the normalized
Table 2. Comparison of disease localization using IoU, where threshold of IoU are set to 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 for evaluation, respectively. Best results between Wang et al.’s and ours are shown in bold. Last row shows average IoU for all diseases. * means results obtained from same dataset with expanded annotation from certified radiologist.

| T(IoU) | Model | Atelectasis | Cardiomegaly | Effusion | Infiltration | Mass | Nodule | Pneumonia | Pneumothorax |
|-------|-------|-------------|--------------|----------|-------------|------|--------|-----------|-------------|
| 0.1   | Wang  | 0.6888      | 0.9383      | 0.6601   | 0.7073      | 0.4000 | 0.1392 | 0.6333    | 0.3775      |
|       | Ours  | 0.4113      | 0.9659      | 0.5950   | 0.8134      | 0.5292 | 0.0758 | 0.7421    | 0.3267      |
|       | Ours* | 0.4283      | 0.9657      | 0.6926   | 0.8211      | 0.5539 | 0.2156 | 0.7752    | 0.3263      |
| 0.2   | Wang  | 0.4722      | 0.6849      | 0.4509   | 0.4796      | 0.2588 | 0.0506 | 0.35      | 0.2346      |
|       | Ours  | 0.2388      | 0.9453      | 0.3341   | 0.6745      | 0.3058 | 0.0125 | 0.5752    | 0.2043      |
|       | Ours* | 0.2557      | 0.9453      | 0.5036   | 0.6833      | 0.3292 | 0.1522 | 0.6331    | 0.2052      |
| 0.3   | Wang  | 0.2444      | 0.4589      | 0.3006   | 0.2764      | 0.1529 | 0.0379 | 0.1666    | 0.1326      |
|       | Ours  | 0.1458      | 0.9316      | 0.1448   | 0.5223      | 0.1531 | 0.0031 | 0.4420    | 0.1432      |
|       | Ours* | 0.1613      | 0.9246      | 0.3334   | 0.5368      | 0.1768 | 0.1393 | 0.5010    | 0.1432      |
| 0.4   | Wang  | 0.0944      | 0.2808      | 0.2026   | 0.1219      | 0.0705 | 0.0126 | 0.075     | 0.0714      |
|       | Ours  | 0.0887      | 0.8772      | 0.0590   | 0.3575      | 0.1174 | 0.0025 | 0.2582    | 0.1124      |
|       | Ours* | 0.1062      | 0.8769      | 0.2678   | 0.3911      | 0.1430 | 0.1388 | 0.3252    | 0.1123      |
| 0.5   | Wang  | 0.0500      | 0.1780      | 0.1111   | 0.0650      | 0.0117 | 0.0126 | 0.0333    | 0.0306      |
|       | Ours  | 0.0448      | 0.7806      | 0.0133   | 0.2523      | 0.0592 | 0.00022 | 0.1419    | 0.0514      |
|       | Ours* | 0.0612      | 0.7812      | 0.2286   | 0.2843      | 0.0827 | 0.1394 | 0.2086    | 0.0522      |
| 0.6   | Wang  | 0.0222      | 0.0753      | 0.0457   | 0.0243      | 0.0000 | 0.0126 | 0.0166    | 0.0306      |
|       | Ours  | 0.0116      | 0.5211      | 0.0000   | 0.1552      | 0.0238 | 0.0000 | 0.0422    | 0.0413      |
|       | Ours* | 0.0280      | 0.5208      | 0.2159   | 0.1873      | 0.0475 | 0.1393 | 0.1081    | 0.0409      |
| 0.7   | Wang  | 0.0055      | 0.0273      | 0.0196   | 0.0000      | 0.0000 | 0.0000 | 0.0083    | 0.0204      |
|       | Ours  | 0.0063      | 0.2342      | 0.0000   | 0.0742      | 0.0123 | 0.0000 | 0.0165    | 0.0108      |
|       | Ours* | 0.0223      | 0.2332      | 0.2161   | 0.1062      | 0.0358 | 0.1395 | 0.0836    | 0.0104      |
| IoU   | Wang  | 0.2254      | 0.3776      | 0.2558   | 0.2392      | 0.1277 | 0.0379 | 0.1833    | 0.1282      |
|       | Ours  | 0.1353      | 0.7508      | 0.1637   | 0.4071      | 0.1715 | 0.0134 | 0.3169    | 0.1272      |
|       | Ours* | 0.1519      | 0.7497      | 0.3511   | 0.4300      | 0.1956 | 0.1520 | 0.3764    | 0.1272      |

5. Discussion

In clinical procedures, visual evidence such as segmentation or spatial localization of disease lesions, in support of disease classification results, is a vital part of clinical diagnosis. It provides a comprehensive insight into the dis-
ease and potentially decreases the false positive of diagnosis. In this work, we proposed a weakly supervised adaptive DenseNet architecture. It only trains on image-level annotation, yet is able to provide both disease classification and corresponding visual evidence, making it potentially valuable in clinical setting.

In our experiment, we specifically looked at 14 different thoracic diseases. Our network demonstrated its ability to precisely identify disease patterns which generated accurate disease classification and corresponding heatmaps for disease localization. In our classification experiments, our network outperformed the current state-of-the-art method on 10 out of 14 diseases. Specifically, our network has shown significant classification improvements on diseases with large lesion such as "Pneumonia" and "Emphysema". Our model also gave robust classification results on diseases with small lesions such as "Nodule". In our localization experiments, our network achieved significant better localization performance on 5 out of 8 diseases with mean T(IoU)=0.7, as compared to the NIH baseline [8]. Compared to previous methods, the possible reasons for our network to achieve higher classification and localization accuracy is two-fold. First of all, the use of a two-stage pooling (class-wise & spatial-wise) allows the network to capture complex intra-class variations and highlight different useful lesion features on multiple sub-maps for one disease class. Secondly, the removal of average pooling helps to improve spatial resolution and maintain more image details for the lesion features capturing in the WSL pooling.

We also evaluated the effect of localization label ambiguity. We collaborated with a certified radiologist to expand the ground truth bounding box annotations on 80 cases. We found that our localization has significantly better overlap with the bounding box in the expanded annotation than the one provided in ChestX-ray14 dataset on certain cases. Examples of the ambiguity cases are illustrated in our Supplementary Materials with detailed explanations.

For future work, it is worth noting that the annotations in the ChestX-ray14 dataset is generated by text mining instead of manual annotation, which guarantees 90% correctness of labeling. In this case, the labels can potentially be further improved. The model should be able to achieve enhanced performance if such training data is available. Furthermore, a dataset with more disease categories included will be favorable as it is closer to real life clinic settings. Future works include expanding the current ChestX-ray14 dataset with more manual label and more disease categories such as coronary artery disease [10, 11] for evaluation. Applications of our model on different medical imaging dataset that targeting multiple diseases should also be evaluated in the future.

6. Conclusion

In summary, we presented a weakly supervised adaptive DenseNet to classify and localize 14 thoracic diseases by using only image-level annotation during training. Extensive experiments demonstrated the effectiveness of our network which achieved both the best classification and localization results among the previous published state-of-the-art approaches on the ChestX-ray14 dataset.
References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

[2] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3):211–252, 2015.

[3] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In European conference on computer vision, pages 740–755. Springer, 2014.

[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In Computer Vision (ICCV), 2017 IEEE International Conference on, pages 2980–2988. IEEE, 2017.

[5] Ross Girshick. Fast r-cnn. arXiv preprint arXiv:1504.08083, 2015.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[7] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. Medical image analysis, 42:60–88, 2017.

[8] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammad Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3462–3471. IEEE, 2017.

[9] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, volume 1, page 3, 2017.

[10] Bo Zhou, Di Wen, Katelyn Nye, Robert C Gilkeson, Brendan Eck, David Jordan, and David L Wilson. Detection and quantification of coronary calcium from dual energy chest x-rays: Phantom feasibility study. Medical physics, 44(10):5106–5119, 2017.

[11] Bo Zhou, Yi Jiang, Di Wen, Robert C Gilkeson, Jun Hou, and David L Wilson. Visualization of coronary artery calcium in dual energy chest radiography using automatic rib suppression. In Medical Imaging 2018: Image Processing, volume 10574, page 105740E. International Society for Optics and Photonics, 2018.

[12] Di Wen, Katelyn Nye, Bo Zhou, Robert C Gilkeson, Amit Gupta, Shiraz Ranin, Spencer Couturier, and David L Wilson. Enhanced coronary calcium visualization and detection from dual energy chest x-rays with sliding organ registration. Computerized Medical Imaging and Graphics, 2018.

[13] Yaniv Bar, Idit Diamant, Lior Wolf, Sivan Lieberman, Eli Konen, and Hayit Greenspan. Chest pathology detection using deep learning with non-medical training. In Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on, pages 294–297. IEEE, 2015.

[14] Paras Lakhani and Baskaran Sundaram. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology, 284(2):574–582, 2017.

[15] Yaron Anavi, Ilya Kogan, Elad Gelva, and Hayit Greenspan. A comparative study for chest radiograph image retrieval using binary texture and deep learning classification. In Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, pages 2940–2943. IEEE, 2015.

[16] Li Yao, Eric Poblenz, Dmitry Dagunts, Ben Covington, Devon Bernard, and Kevin Lyman. Learning to diagnose from scratch by exploiting dependencies among labels. arXiv preprint arXiv:1710.10501, 2017.

[17] Zhe Li, Chong Wang, Mei Han, Yuan Xue, Wei Wei, Li-Jia Li, and Fei-Fei Li. Thoracic disease identification and localization with limited supervision. arXiv preprint arXiv:1711.06373, 2017.

[18] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level lung nodule detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225, 2017.

[19] Maxime Oquab, Léon Bottou, Ivan Laptev, and Josef Sivic. Is object localization for free?-weakly-supervised learning with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 685–694, 2015.

[20] M Everingham, L Van Gool, CKI Williams, J Winn, and A Zisserman. The pascal visual object classes challenge 2012 (voc2012) results (2012). In URL http://www.pascal-network.org/challenges/VOC/voc2011/workshop/index.html, 2011.

[21] Thibaut Durand, Taylor Mordan, Nicolas Thome, and Matthieu Cord. Wildcat: Weakly supervised learning of deep cnns for image classification, pointwise localization and segmentation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), 2017.

[22] Mark Everingham, L Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge 2007 (voc2007) results (2007). In URL http://www.pascal-network.org/challenges/VOC/voc2007/index.html, 2007.

[23] Ariadna Quattoni and Antonio Torralba. Recognizing indoor scenes. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 413–420. IEEE, 2009.

[24] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Computer vision and
[25] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[26] Vincent Tinto. Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, 45(1):89–125, 1975.