ProjectX: Classification, localization and tracking progression of Thoracic Diseases

Dhananjay Singh, Vibhu Jawa, Vishal Rao

***Abstract*—We developed a system for finding and localizing thoracic diseases in X-rays using latest computer vision techniques. Then we attempted to highlight the progression of a disease over the multiple visits of a patient.**

I. Introduction

A large number of people in the world are affected every year by Thoracic diseases. In the United States alone, it is estimated that more than 100,000 die of diseases such as COPD[1] and Pneumonia[2] each year, with 16 million Americans diagnosed with COPD. With COPD being the third leading cause of death in the United States (2014)[3] and millions more undiagnosed [3] it is a clear that we need techniques to speed up diagnoses. Using computer vision techniques we attempt to do that by analyzing Xrays. We train a residual neural network to be able to classify and localize thoracic diseases and then use structure similarity techniques to detect structural changes in X-Rays over multiple visits of a patient.

A. Related Work

Advancement in deep learning, computer vision and availability of large data sets has enabled machine learning to surpass the performance of human professionals in a variety of vision related tasks Skin cancer classification (Esteva et al., 2017)[4], Mitosis detection in breast cancer (Cireş et al., 2013) [5] and hemorrhage detction (van Grinsven, 2016) [6]

There has been a lot of recent work done in computer vision driven automated diagnosis from chest radiographs, High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks (Rajkomar et al. 2017)[7], pediatric frontal chest radiograph screening with fine-tuned convolutional neural networks (Gerrand, J et al. 2017)[8] using the publicly available OpenI dataset (Demner-Fushman et al., 2015)[9].

Wang et al. (2017) released ChestX-ray-14 in 2017, which was an order of magnitude larger than previous datasets of its kind and also bench marked different convolutional neural network architectures on pre-trained on ImageNet. Yao et al. (2017) exploited statistical dependencies between labels in order make more accurate predictions, outperforming Wang et al. (2017) on 13 of 14 classes . Very recently Andrew ng et. al (Nov 25,2017) published their work on Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.

We focus our work towards classification as well as disease localization using heat maps from the convulsion layers and using the heat maps to track the progression in the diseases.

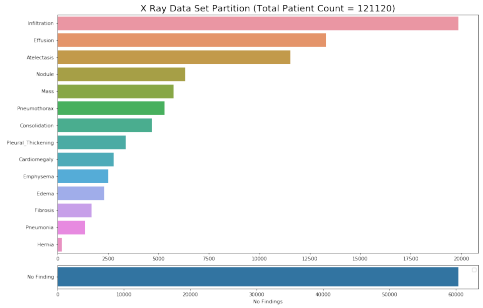

Fig. 1. Distribution of Multiple pathologes in the data

The magnitude of this dataset with the availability of multiple patient follow ups has made tracking disease progression through time possible and is a completely novel attempt.

B. Dataset and Analysis

We used the ChestX-ray14 dataset [10] containing 112,120 frontal-view X-ray images of 30,805 unique patients. Wang et al. (2017) annotated these images with up to 14 different thoracic pathology labels using automatic extraction methods on radiology reports.

We analyzed the dataset for age,gender and disease distribution to check if there exists a statistical relationship which we could use for prediction. [11]

The figure 1 shows the distribution of multiple pathology in the data.

The figure 2 shows the distribution of male and female patients for each pathology.

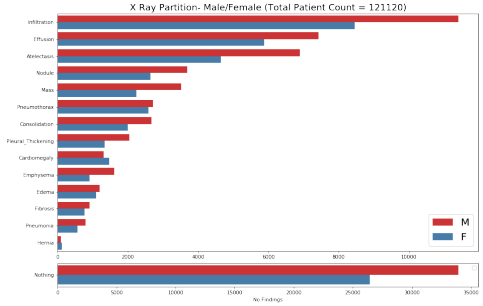

Fig. 2. Distribution of pathologes with Gender Partitions

We can infer from figure 1 and figure 2 that the ratio of male and female patents for all the pathology's is quite balanced so

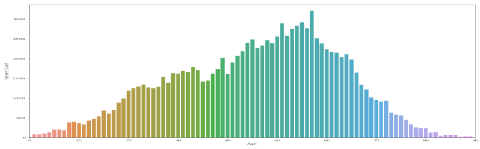

Fig. 3. Overall Distribution of Patholgoes with age

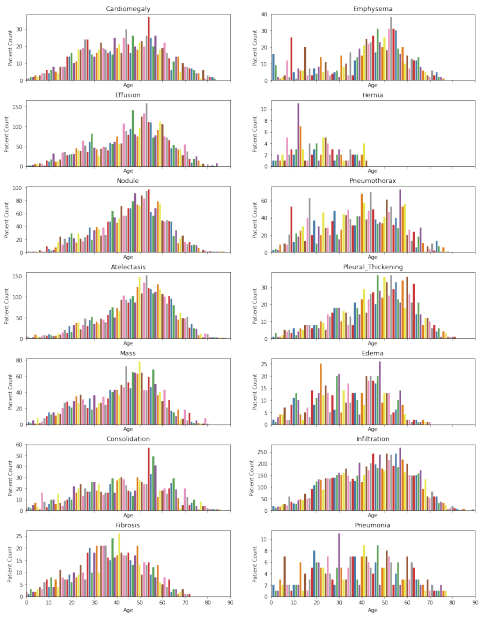

Fig. 4. Disease Distribution with Age

there is no apparent relationship between these diseases and gender.

The figure 4 and 3 shows the relationship between age and the different diseases.

We plot the the distribution of data in figure 5 for patients along with follow ups as we are interested in the progression tracking of diseases.

We plot the distribution in figure 6 between patients with multiple pathology's vs those with single pathology. This gives

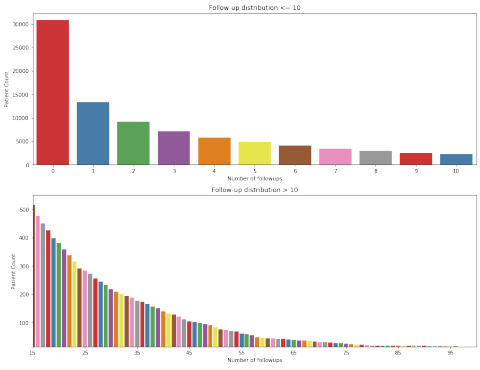

Fig. 5. Distribution or Follow Up Data

us insights for out multi-class classification.

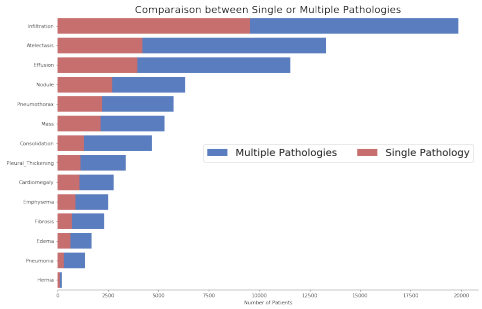

Fig. 6. Multiple Pathologies vs Single Pathology

II. PROJECT-X

A. Preprocessing

We randomly split the dataset into training (21505 patients, 78028 images), validation (3175 patients, 11629 images), and test (6125 patients, 22464 images). [11]

There is no patient overlap between the sets to ensure the unbiased results especially for progression tracking.

Before inputting the images into our network, we down-scaled them to $224 * 224$ and normalized based on the mean and standard deviation of images in the ImageNet training set because we are applying transfer learning on images trained on ImageNet. We also augmented the training data with random horizontal flipping.

B. Problem Formulation

The task at hand is that of multi-label multi classification, since a patient may have more than one disease. We have used the multi label soft margin loss based on max-entropy.

$$loss(x, y) = -\sum_i y[i] * \log \frac{1}{1 + e^{-x[i]}} + (1 - y[i]) * \log \frac{e^{-x[i]}}{1 + e^{-x[i]}} \quad (1)$$

C. Architecture

1) Disease Multi Class Classification: We have developed this solution using Pytorch and it's pre trained version of Resnet50. While deeper nets are likely to give better performance, the training time and resource requirements deterred us from pursuing them. The last fully connected layer is modified to output 15 logits, for each of our categories. The objective function is Multi label soft margin loss.

2) Disease Localization: The Resnet network can also be used to provide information of the affected areas of the lungs. To achieve this, the X-Ray image to be analyzed is passed through the Resnet network, and information of the output of the final convolutional layer and the predicted disease probabilities are extracted. Using the fully connected layer weights associated with the most probable disease (in cases where actual disease is not known) or the weights associated

with the known disease (in cases where the actual disease is known), a weighted average of the output of the final convolutional layer of Resnet is derived. The resultant 8x8 grayscale image is resized to match the size of the display image. Contours are generated around areas in the image with intensities greater than a threshold value, and the corner points of the box bounding the contours are identified. The grayscale image is converted into a colored heat-map to better represent the variations in image intensities, and then blended with the original image. Bounding boxes are drawn to highlight the most probable affected area using corner-points identified in the previous step. Figure 7 shows an overview of the disease localization process.

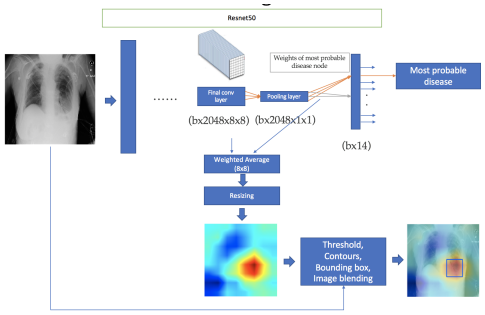

Fig. 7. Overview of Disease Localization Process

3) Disease Progression Analysis: We developed two techniques to identify the changes that occur over a period of time. The first technique involves identifying structural differences and the second involves identifying differences in heat maps.

Structural differences: In this approach, we differentiate between the X-Ray images based on the observed structural differences. At a broad level this involves three steps:

- Pre-processing: The X-Ray images could differ because of differences in orientation and image brightness. This would impact the precision of the differences identified. To address this, the images are first normalized to have similar brightness. To address the issues related to orientation differences, we use affine transformation to align the second image to the first one. The affine matrix required for this transformation is computed using the ECC algorithm.
- Analysis: Once the two images are aligned and normalized, the structural similarity around each point is computed using the structural similarity index (SSIM). A binary image, which highlights points with large differences, is generated by identifying points having the similarity score below a configured threshold value.
- Post-processing Annotation: The resultant binary image would have many minor difference areas which may have occurred due to medically inconsequential factors like orientation differences. Morphological opening technique is used to remove such areas. Contours are generated around areas of difference and the two images are displayed with bounding boxes around the contours.

Figure 8 gives an overview of the process.

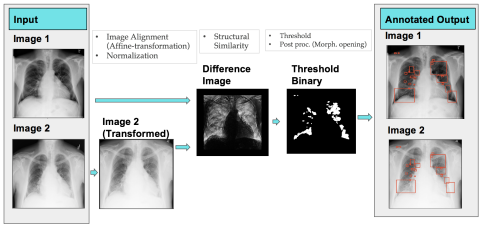

Fig. 8. Identification of Structural Differences in X-Rays

Heat-map based differences: We first generate the heat-maps for the two images to be compared, and also extract the gray-scale images which were originally used for generating the heat-maps. The points in the gray-scale images which are brighter than a given threshold are identified, and binary images representing the most affected areas are generated. The two binary images are aligned by transforming the second X-ray image using the affine matrix generated for aligning image 2 to image 1. By comparing the two aligned images, we identify points which are present in one image but absent in the other. The original X-ray images and their corresponding heat-maps are blended, and bounding boxes are drawn around the points identified in the previous step. A black box is drawn around the affected area present only in the particular image, and a red box is drawn around the affected area in the other image which was not an affected area in the first image.

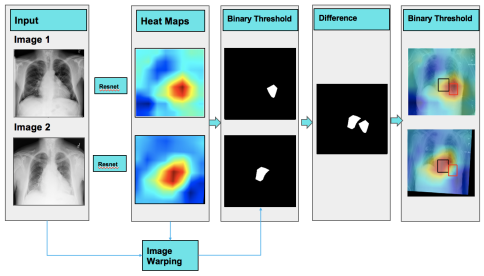

Fig. 9. Identification of Differences in X-Rays Using Heat-Maps

D. Experiments

1) Disease Classification :

- We tried image augmentation at training by adding rotation and random horizontal flipping.We did not see improvements in accuracy through rotation and thus only used random flipping in our results.
- We tried out both multi label margin loss and multi label soft margin loss but archived better results with the latter.

2) Disease Progression Analysis:

- Difference threshold: While generating the binary image containing information of the structurally different areas, different techniques were tested. The Otsu threshold technique gave decent results in images which actually contained many differences. However, the false positive

TABLE I. AUC COMPARISON OF CLASSIFICATION RESULTS

Pathology	Wang et al. (2017)	Andrew NG et al. (2017)	Our Results
Atelectasis	0.716	0.8094	0.73
Cardiomegaly	0.807	0.9248	0.82
Effusion	0.784	0.8638	0.72
Infiltration	0.609	0.7345	0.55
Mass	0.706	0.8676	0.606
Nodule	0.671	0.7802	0.611
Pneumonia	0.633	0.768	0.621
Pneumothorax	0.806	0.8887	0.77
Consolidation	0.708	0.7901	0.68
Edema	0.835	0.8878	0.75
Emphysema	0.815	0.9371	0.77
Fibrosis	0.769	0.8047	0.69
Pleural Thickening	0.708	0.8062	0.64
Hernia	0.767	0.9164	0.66

rates were also high. Thus, the standard threshold technique in which the value is configured as a parameter was used.

- Similarity technique for structural differences: Mean squared error (MSE) was initiallly used for finding similarity between the region around pixels. SSIM, however, outperformed MSE.
- Opening kernel size: Different values of opening kernel size were tested and it was observed that a value of 10 gave optimum results.

III. LIMITATIONS

- The training of a networks can be really slow, especially when hyperparameters are being tuned.
- The localization accuracy drops in cases where the neural network incorrectly predicts the disease.

IV. RESULTS

A. Disease Classification

The results of disease classification are show cased in table I.

We made improvements on some classes as compared to the original paper but are still behind when compared to the state of the art. We believe that we can improve the accuracies by more training but are currently being bottle neck by computation limitations.

B. Disease Localization

The disease localization analysis was performed on X-rays of patients suffering from different diseases, and the degree of overlap of the predicted and actual bounding boxes were observed. Figure 10 shows sample results of disease localization prediction on X-rays of patients suffering from Cardiomegaly and Atelectasis. The blue box indicates the predicted affected area, and the red box indicates the area manually identified by radiologists.

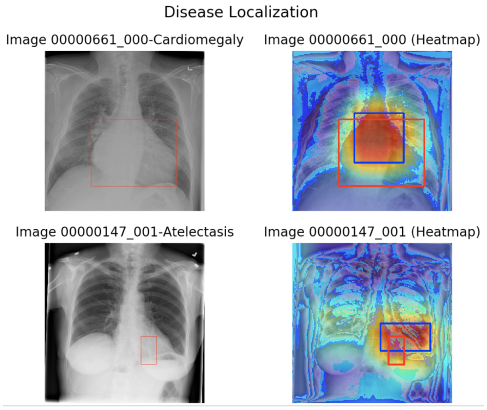

Fig. 10. Identification of Differences in X-Rays Using Heat-Maps

It was observed that the heat-map localization accuracy varied with the classification accuracy of the neural network. The localization accuracy was higher for diseases which had a higher classification accuracy. The overlap of the actual and predicted bounding boxes was highest for X-rays of patients with Cardiomegaly, Atelectasis and Effusion, and it was the lowest for X-rays of patients with nodules.

C. Disease Progression Analysis

To provide a quick snapshot of probable diseases and their progression, we created a dashboard which compared X-Rays of patients over multiple visits and highlighted the key differences.

Figure 11 shows a dashboard which compares the X-Rays of a patient who initially suffered from Effusion and who later also developed Atelectasis. The probabilities of diseases predicted by the neural network for these images are shown at the bottom. The two 'Probable Affected Area' images indicate the regions affected by Effusion and Atelectasis respectively. The two 'Heat-Map Difference' images compare these heat-maps. In these images, the black box indicates the region which was affected only during the first visit (when the patient suffered from Effusion), and the red box indicates the region which was affected only during the second visit (when the patient suffered from Atelectasis too). These images provide a clear view of how the diseases have progressed over time. The 'Structural difference' images highlight the differences based on the variations in structures.

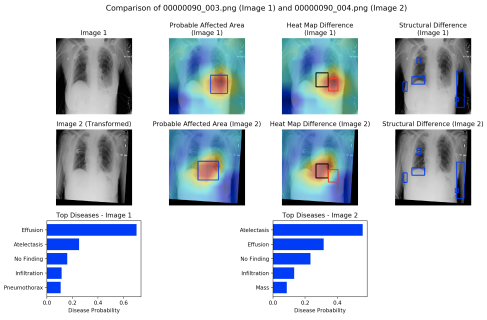

Fig. 11. X-Ray Comparison Dashboard

On comparing the heat-map based difference images and structural difference images of several patients, it was observed that the differences identified by the heat-map technique were generally more medically relevant. Additionally, they were also more tolerant to variations due to subject-orientation, brightness and presence of equipment like heart pace makers.

V. Conclusion

We achieved good classification accuracy but more importantly show cased that the heat map obtained from the neural network can act as an important aid for disease localization and tracking progression of diseases which is difficult to infer when we only compare x ray images.

References

[1] CDC, "COPD Statistics." https://www.cdc.gov/copd/data.html. Accessed: 2017-12-16.

[2] CDC, "Pneumonia Feature." https://www.cdc.gov/features/pneumonia/index.html. Accessed: 2017-12-16.

[3] CDC, "COPD Homepage." https://www.cdc.gov/copd/index.html. Accessed: 2017-12-16.

[4] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[5] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 411–418, Springer, 2013.

[6] M. J. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, "Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1273–1284, 2016.

[7] A. Rajkomar, S. Lingam, A. Taylor, M. Blum, and J. Mongan, "High-throughput classification of radiographs using deep convolutional neural networks," *Journal of Digital Imaging*, vol. 30, no. 1, pp. 95–101, 2017. cited By 6.

[8] J. Gerrand, Q. Williams, D. Lunga, A. Pantanowitz, S. Madhi, and N. Mahomed, "Paediatric frontal chest radiograph screening with fine-tuned convolutional neural networks," *Communications in Computer and Information Science*, vol. 723, pp. 850–861, 2017. cited By 0.

[9] D. Demner-Fushman, M. Kohli, M. Rosenman, S. Shooshan, L. Rodriguez, S. Antani, G. Thoma, and C. McDonald, "Preparing a collection of radiology examinations for distribution and retrieval," *Journal of the American Medical Informatics Association : JAMIA*, vol. 23, pp. 304–310, 3 2016.

[10] 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," *arXiv preprint arXiv:1705.02315*, 2017.

[11] S. Bernadac, "Chest x ray data analysis," 2017.