Pneumonia Detection Using Chest X Rays

Rahul Sharma
Student- AI/ML, IIIT D
rahultheogre@gmail.com

Shubhangi Gupta
Student–AIML, IIITD

Sanchit Bhardwaj
Student–AIML, IIITD

Hema Bhavani
Student- AI/ML, IIIT D

Aarti Dangi
Student–AIML, IIITD

Jishnu Chatterjee
Student–AIML, IIITD

Pratiti Basu
Student- AI/ML, IIIT D

Abstract — Purpose: This work investigates deep learning models that can automate the detection of pneumonia in chest x-ray images. We propose a modified U-Net Architecture. We validated our solution on a famous dataset call RSNA Pneumonia Detection Challenge publicly available on Kaggle.

Keywords — Pneumonia Detection, Medical Imaging, Transfer Learning, CNN Architecture, U-Net.

1. Introduction

Pneumonia is an inflammatory condition of the lung that primarily affects the small air sacs called alveoli, and it is caused by bacterial or viral infection. Symptoms include cough with phlegm, chest pain, difficulty breathing, and fever, and the severity of symptoms can vary. Pneumonia diagnosis is a time-consuming process that involves highly skilled professionals to analyze a chest radiograph or chest X-ray (CXR) and confirm the diagnosis with clinical history, vital signs, and laboratory tests. It helps doctors to work out the extent and placement of the infection in the lungs. Respiratory illness manifests as a neighborhood of inflated opacity on X-Ray. However, pneumonia diagnosis is complicated because increased opacity on CXRs could represent several other lung conditions, such as pulmonary edema, bleeding, volume loss, and lung cancer.

In recent years, deep-learning methods based on convolutional neural networks (CNNs) have exhibited increasing potential and efficiency in image recognition tasks, such as robotics, self-driving cars, and medical applications. For application to CXR, deep-learning models can achieve detection performance close to that of radiologists.

2. Literature review

Roth et al. demonstrated the power of deep convolutional neural network (CNN) to detect the lymph node in clinical diagnostic task and got drastic results even in the presence of low contrast surrounding structures obtained from computer tomography. In another study, Shin et al. addressed the problems of thoraco-abdominal lymph detection and interstitial lung disease classification using deep CNN. They developed different CNN architectures and got promising results with 85 percent sensitivity at three false positives per patient.

Ronneburger et al. developed a CNN approach with the use of data augmentation. They suggested that even trained on small samples of image data obtained from transmitted light microscopy; the developed model could capture high accuracy. Jamaludin et al. applied CNN architecture to analyze the data obtained from spinal lumber magnetic resonance imaging (MRI). They developed an efficient CNN model to generate radiological grading of spinal lumber MRIs. All these studies have performed well on radiological data, except that the size of the data was restricted to a few hundred samples of patients. Therefore, a detailed study is required to use the power of deep learning over a thousand samples of patients to achieve the accurate and reliable predictions.

Kallianos et al. presented a state of art review stating the importance of artificial intelligence in chest X-ray image classification and analysis. Wang et al. addressed this issue and prepared a new database ChestX-ray8 with 108,948 front view X-ray images of 32,717 unique patients. Each of the X-ray images could have multiple labels. They used deep convolutional neural networks to validate the results on this data and obtained promising results. They mentioned that chestX-ray8 database can be extended by including more disease classes and would be useful for other research studies. Rajpurkar et al. developed a 121 layer deep convolutional layer network chestX-ray14 dataset. This dataset is publicly available, with over 0.1 million front view X-ray images with 14 disease labels. They mentioned their algorithm can predict all 14 disease categories with high efficiency.

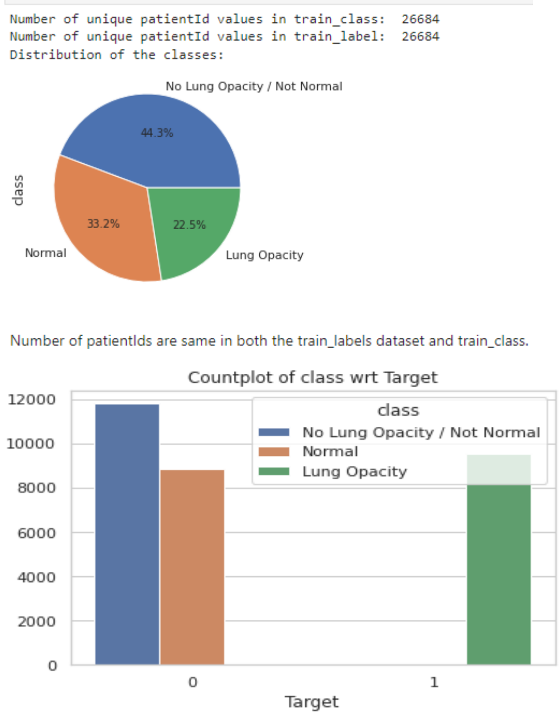
Rajpurkar et al. developed CheXNet10 by using the ChestX-ray14 dataset11, which contains 112,120 frontal-view chest x-ray images individually labeled with 14 different pathologies, including pneumonia. Tey et al. estimated and compared the performance of the model and four radiologists, revealing that the model exceeded the average performance of the radiologists on the pneumonia detection task. Hwang et al. applied a deep-learning method for chest radiograph diagnosis in the emergency department, with their results showing that the diagnostic performance of radiology residents using CXR readings improved after radiographs were reinterpreted through the deep-learning algorithm output.

Irvin et al. stated that a large labeled dataset is the key to success in prediction and classification tasks. They presented a huge dataset that comprises 224,316 chest radiographic images of 65,240 patients. They named this dataset as CheXpert. Then they used convolutional neural networks to assign labels to them based on the probability assigned by the model. Model used frontal and lateral radiographs to output the probabilities of each observation. Further, they released the dataset as a benchmark dataset. Besides the availability of a large dataset, it is highly desirable that every object in the image should be detected carefully and segmentation of each instance should be done precisely. Therefore, a different approach is required to handle both instance segmentation and object detection.

Such powerful methods are faster region-based CNN (F-RCNN) and FCN (Fully Convolutional Network). F-RCNN can be extended with an additional branch for segmentation mask prediction on each region of interest, along with existing branches for the classification task. This extended network is called Mask R-CNN, and it is better than F-RCNN in terms of efficiency and accuracy. Kaiming He et al. presented Mask R-CNN approach for object instance segmentation. They compared their results with the best models from COCO 2016. Luc et al. extended their approach by introducing an instance level segmentation by predicting convolutional features.

3. Materials and Methods

The labeled dataset of the chest X-Ray images and patients' metadata was publicly provided for a Kaggle challenge by the US National Institutes of Health Clinical Center. The database comprises frontal-view X-ray images from 26684 unique patients. Each image is labeled with one of three different classes from the associated radiological reports: 31.61% lung opacity, 39.11% -no lung opacity, 29.28% normal. In the target class, there are 31.61% of pneumonia class, 68.38% of non-pneumonia images. Bounding boxes for patients having pneumonia are defined in the train labels file. (Check Figure for representation) There are 9555 positive patients in this file. Each X-ray image has a metadata associated with it. It gives information about the patient, the view position etc. 3543 duplicate entries suggest presence of different X-ray views for the same patient.



```
Number of unique patientId values in train_class:  26684
Number of unique patientId values in train_label:  26684
Distribution of the classes:
```

Number of patientIds are same in both the train_labels dataset and train_class.

Initial images received from the dataset are in the DICOM format. They contain a combination of header metadata as well as underlying raw image arrays for pixel data. Most of the standard headers containing patient identifiable information have been removed. Understanding the data structure, imaging file format and label types, the primary objective is to detect the bounding boxes comprising binary classification, i.e. presence (or) absence of Pneumonia.

3.1 BASE CNN MODEL

The base model architecture had 4 convolutional layers along with MaxPooling (2 by 2) and dropout layers. The first two convolutional layers had 32 and 64 filters with size (3,3). The activation function used for each convolutional layer was Relu. The final output layer had 1 neuron for binary classification with activation function Sigmoid.

3.2 TRANSFER LEARNING WITH Inception V3

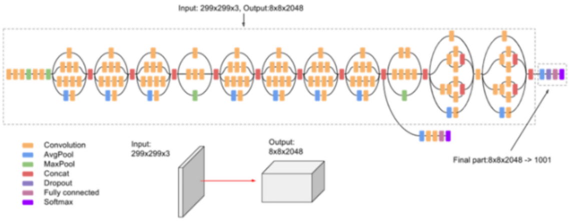


Figure 1- InceptionV3 Architecture

3.3 U- NET

U-Net is one of the most famous image segmentation architectures. It was proposed in 2015 by Olaf Ronneberger, Philipp Fischer, Thomas Brox (University of Freiburg, Germany). An end-to-end segmentation technique- U-Net takes a raw image in and outputs a segmentation map of the image. The U-Net architecture is a U-shaped (Figure 3), symmetric convolutional network with a down-sampling contraction path and an up-sampling expansion path. The resulting segmented output image is much smaller than the raw input image. U-net only has convolutional layers.

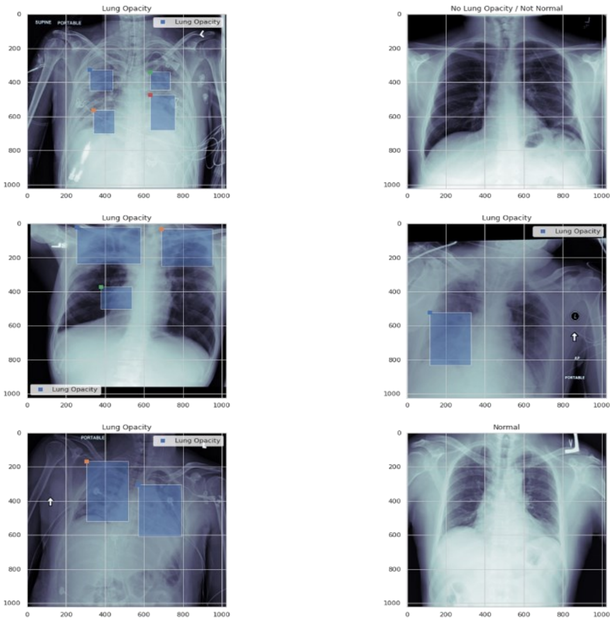


Figure 2: Sample Images with Bounding Boxes of all categories

And the input image is fed into the network, the data is propagated through the network, resulting in a segmented map as output. Contraction/down sampling path (Encoder Path). The encoder path captures the context of the image. It is just a stack of convolution and max pooling layers. The encoding path has 4 blocks. Each block comprises 1) Two 3 x 3 convolution layers + ReLU activation function (with batch normalization). 2) And. One 2 x 2 max pooling layer.

Decoder enables precise localization using transposed convolutions (An upsampling technique). The expansion path has 4 blocks. Each block comprises: 1) Deconvolution layer with stride 2. 2) Concatenation with the corresponding cropped feature map from the contracting path. i.e. At every step of the decoder, we use skip connections. These skip connections are important. Every output of every transposed convolution layer is concatenated with feature maps from the Encoder at the same level. 3) And. Two 3 x 3 convolution layers + ReLU activation function (with batch normalization).

Let's look at the operations/steps a bit more closely.

1. 3 x 3 convolution layer: It is a standard 3x3 convolution followed by a non-linear activation function (ReLU). It uses only the valid part of the convolution (no padding), which is why a one-pixel border of the is lost. This allows processing large images in individual tiles. This layer helps to extract features from the images, i.e., meaningful information from the data.

2. Max-pooling Operation: It reduces the size of the feature map. It propagates maximum activation from each 2x2 window to the next feature map. The resulting map has factor 2 lower spatial resolution. Max-pooling is used to reduce the features' dimensions to extract the heavily weighted features, which become easy to compute and require less computation power for parameter learning.

3. After each max-pooling operation, feature channels are scaled by a factor of 2.

4. The above steps 2 and 3 together result in the sequence of convolutions and the max-pooling operation results in a spatial contraction of the image. This contraction portion helps us understand the "what" of the image. The contraction path of the networks makes it learn the useful content/information in the image.

5. The expansion path has a series of steps of up-convolutions and concatenation with high-resolution features from the contraction path. The expansion path creates a high-resolution segmentation map.

6. 2x2 up-convolution: The up-convolution uses a learned kernel to map each vector to a 2x2 output window, followed by a non-linear activation function. The resulting maps have a factor 2 higher resolution. The up-sampling allows the propagation of more information to the high-resolution layers.

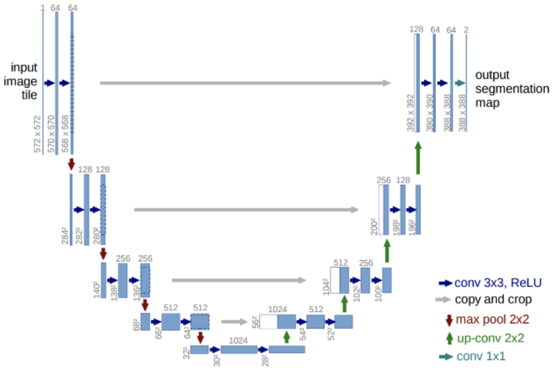


Fig. 1. U-net architecture (example for 32x32 pixels in the lowest resolution). Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

Figure 3: Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention — MI

7. The series of convolutional operations are followed with concatenations via skip connections. The skip connections are essential. They help to figure out the "where" of the various objects in the image. These give valuable information about the spatial and graphical information of the image.

8. Finally, 1 × 1 convolution operation yields the final segmented image. The output segmentation map has two channels, one for foreground and another for background classes. Because of unpadded convolutions, the output is smaller than the input image.

In this way, the series of contraction and expansion paths of the U-Net architecture allows leveraging the useful information of feature mapping and pooling while retaining the spatial and graphical resolution, yielding a more robust segmentation. This architecture can tackle challenges like low training data availability, touching and overlapping objects, partially invisible borders between different objects, fuzzy borders, low contrast edges, and objects with strong shape variations. Given the computational power we have, and the theoretical resources at hand, in this project, we will focus on this architecture as our own.

4. Results

Result of Basic CNN- Training accuracy is around 60 percent whereas validation accuracy is round 58 percent. We have avoided over-fitting, but it seems to be clear that a normal CNN will not help us. Test accuracy is also 61 percent is. At least our model is consistent.

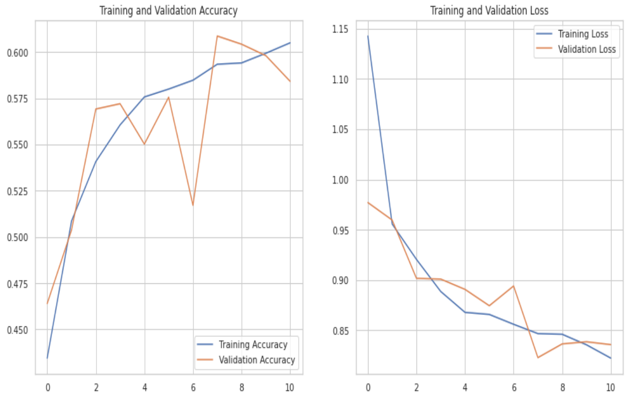


Figure 4: Base CNN model result

TRANSFER LEARNING

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task. It is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems. In transfer learning, we first train a base network on a base dataset and task, and then we repurpose the learned features, or transfer them, to a second target network to be trained on a target dataset and task. This process will tend to work if the features are general, meaning suitable to both base and target tasks, instead of specific to the base task.

Result with Inception V3-

Test loss: 0.52. Test accuracy: 0.59. The scores are very low and the model is not useful.

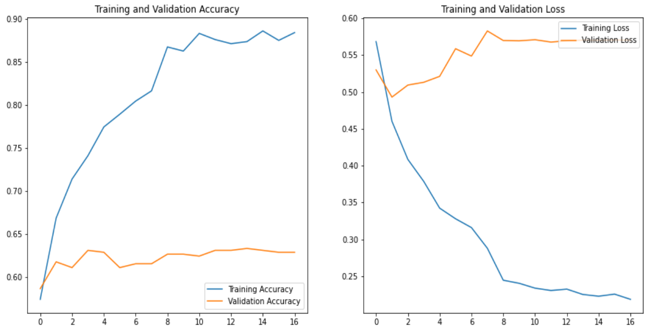


Figure 5- Inceptionv3 results

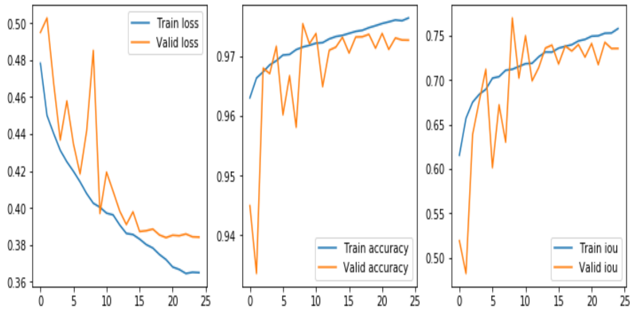


Figure 6- Modified U - Net result

5. CONCLUSIONS

Thus, we took up a challenging task which is still being researched upon by radiologist and deep learning engineers worldwide and by going a little deeper into the literature and various models available online; we improved upon existing knowledge. We are keen to go deep into image segmentation and understand how Mask R-CNN improves upon U-Net and Faster R-CNN.

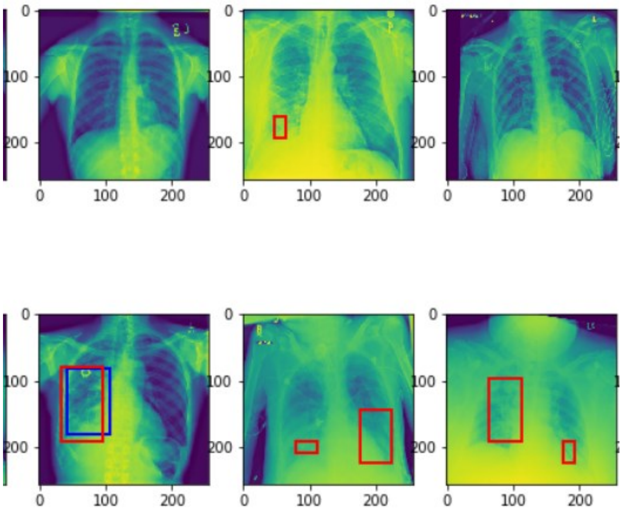


Figure 7- Plotting predictions in one batch- U-Net Model

6. REFERENCES

1. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. *RadioGraphics* 2017; 37:2113–2131 [Crossref] [Medline] [Google Scholar]

2. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015; 115:211–252 [Crossref] [Google Scholar]

3. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, eds. *Advances in neural information processing systems 25 (NIPS 2012)*. San Diego, CA: Neural Information Processing Systems Foundation, 2012 [Google Scholar]

4. Prevedello LM, Halabi SS, Shih G, et al. Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. *Radiol Artific Intell* 2019; 1:e180031 [Crossref] [Google Scholar]

6. Shih G, Wu CC, Halabi SS, et al. Augmenting the National Institutes of Health chest radiograph dataset with expert annotations of possible pneumonia. *Radiol Artif Intell* 2019; 1:e180041 [Crossref] [Google Scholar]

7. RSNA Pneumonia Detection Challenge: overview. Kaggle website. www.kaggle.com/c/rsna-pneumonia-detection-challenge. Accessed July 17, 2019 [Google Scholar]

9. Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Piscataway, NJ: IEEE, 2017:764–773 [Google Scholar]

10. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the impact of

residual connections on learning. arXiv website. arxiv.org/abs/1602.07261. Published February 23, 2016. Accessed May 18, 2018 [Google Scholar]

11. Chollet F. Xception: deep learning with depthwise separable convolutions. arXiv website. arxiv.org/abs/1610.02357. Published October 7, 2016. Accessed May 18, 2018 [Google Scholar]

12. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. arXiv website. arxiv.org/abs/1608.06993. Published August 24, 2016. Accessed May 18, 2018 [Google Scholar]

13. ImageNet website. ImageNet. www.image-net.org/. Accessed February 20, 2019 [Google Scholar]

14. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: *2017 IEEE International Conference on Computer Vision.* Los Alamitos, CA: IEEE, 2017:2999–3007 [Google Scholar]

15. Dai J, Li Y, He K, Sun J. R-FCN: object detection via region-based fully convolutional networks. arXiv website. arxiv.org/abs/1605.06409v2. Published May 20, 2016. Accessed February 22, 2019 [Google Scholar]

16. Hu H, Gu J, Zhang Z, Dai J, Wei Y. Relation networks for object detection. arXiv website. arxiv.org/abs/1711.11575v2. Published November 30, 2017. Accessed February 22, 2019 [Google Scholar]

17. Code for 1st place solution in Kaggle RSNA Pneumonia Detection Challenge. GitHub website. github.com/i-pan/kaggle-rsna18. Accessed July 17, 2019 [Google Scholar]

18. Ng A. Convolutional neural networks. Coursera website. www.coursera.org/learn/convolutional-neural-networks. Accessed December 27, 2018 [Google Scholar]

19. Howard J. Cutting edge deep learning for coders: part 2. Onwards fast ai website. course18.fast.ai/part2.html. Accessed January 26, 2019 [Google Scholar]

20. Gaiser H. Keras implementation of RetinaNet object detection: keras-retinanet. github.com/fizyr/keras-retinanet. GitHub website. Accessed January 19, 2019 [Google Scholar]

21. 3rd Place solution for RSNA Pneumonia Detection Challenge. GitHub website. github.com/pm-cheng/rsna-pneumonia. Accessed March 2019 [Google Scholar]

22. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition.* Los Alamitos, CA: IEEE, 2009:248–255 [Google Scholar]

23. Pan I, Agarwal S, Merck D. Generalizable inter-institutional classification of abnormal chest radiographs using efficient convolutional neural networks. *J Digit Imaging* 2019 Mar 5 [Epub ahead of print] [Google Scholar]