Lung Semantic Segmentation using Convolutional Neural Networks

N. Sarada, K. Thirupathi Rao

Abstract: In the fields of computer vision, semantic segmentation is widely used to build a broad range of applications, specifically, medical image processing is where it majorly impacts. It helps in the prediction of abnormalities and the description of accurate analysis on CT-Scans and X-Rays. With the rise of deep learning, neural networks are performing and achieving accurate results with respect to the metrics in the fields of image classification and segmentation. In this work, we use the Convolutional Neural Network on Chest-X-Rays for segmenting the lungs. The architecture being used for image segmentation is U-Net which applies on the train and test data of NIH dataset. The outputs of the model include preprocessed images, accuracies on the test and train dataset, loss and prediction with intersection over union values (how dynamically the images are segmented) of Chest X Rays over two thousand examples.

Index Terms: Convolutional Neural Networks, Deep Learning, Image Segmentation, Lung Abnormalities, Medical Image Processing, Tensorflow.

I. INTRODUCTION

Chest abnormalities and lung cancer are the most commonly observed problems in the current era due to the increase in the pollution. Lung cancer is one of the small non-cell types of cancer that is being diagnosed in the recent years. Few abnormalities like tuberculosis, pneumonia are also becoming strenousto identify with the naked eye by radiologists. Additionally, due to the increase in the treatment costs and lack of required medical equipment, it is difficult to locate where the diagnosis needs to be done. Also, high cost and very less availability of these, results in loss of crucial information needed for the diagnosis. There have been several advancements in computer vision to identify these problems in the recent years. However, it is difficult to determine and tackle these problems without the involvement of radiologists as it will be a challenging task for a machine to predict accurate results. There are several Artificial Intelligence and Machine Learning techniques that have been applied to solving the same but which couldn't achieve a high accuracy. Instead deep learning can produce better diagnosis and results than a radiologist. Few architectures like MobileNet, ResNet and DeepMind Net can also be used for image segmentation.

One of the drawbacks of using neural network is that it consumes ample of time for the training process. If the data is not consistent or processed, it results in false predictions. Therefore, Neural Networks need a lot of computational

 N. Sarada, Research Scholar, Department of Compute Science and Engineering, K L E F, Guntur, India.
  Dr. K. Thirupathi Rao, Professor & BOS Chairman, Associate Dean-Academics, Department of Compute Science and Engineering, K L E F, Guntur, India.

power which CPU's cannot afford; on the flip side, these are powered by GPU's to quicken the training process. Approximately, CPU consumes 15 mins to run a simple Convolutional Neural Network, whereas a GPU can do the same in 60 seconds. GPU's, therefore, make great disruptive progress in computing, especially in the fields of deep learning.

The primary goal of this research is to demonstrate the efficient way of two-dimensional Medical Image Segmentation and to identify the abnormalities, which is being done on Chest X-Rays (CXR) in this work.

II. PROBLEM AND RELATED WORK

Previously, image segmentation techniques were processed in the field of computer vision by applying several filters and kernels. Image segmentation techniques now are carried on neural network architectures like AlexNet [1], VGGNet, GoogleNet [2] which return the probabilities of edges that segment the entire image based on the pixel densities. This is how convolutional neural networks such as the ones described above segment objects and patterns present in the images. f they have a huge number of max-pooling layers, it results in the loss of spatial information [3].

Inspired from computational neuroscience, few image segmentations use techniques that resemble the visual cortex for handling multiscale resolutions of the image [5]. It uses a particular type of filter known as Gabor filter which is usually used in the inception models. This inception model is further developed into a 22-layer Convolutional Neural Network architecture that can be seen in the google net model. Moreover, few probabilistic models are used for coupling the ConvNets with a "Conditional Random Field (CRF)" that gives better results than resorting to conventional segmentation techniques. The advantage of using this conditional probability is that it relates the pixels with the other neighbouring pixels processing standalone segmenting process [6].

Baris, Kayalıbay et al. explains the image segmentation process on medical images in theory in "CNN-based Segmentation of Medical Imaging Data." In this, he uses a four-layer basic convolutional architecture with a single softmax activation function to deal with medical images. This is done by converting a medical model of pixels into tensors. Tensor is an n-dimensional matrix that stores the data. The architecture uses a U-NET convolutional model. It has two parts, namely, image segmentation, and deconvolution. In the first part, the input tensors are sent into convolutional layers where the image gets segmented with an increase

in the stride value and decrease in the resolution of the image. Deconvolution uses n-th index techniques [7].

III. CONVOLUTIONAL NEURAL NETWORKS FOR MEDICAL IMAGE PROCESSING

Convolutional neural networks perform exceptionally well for image classification and image segmentation giving a maximum accuracy. In this work, we discuss briefly about how convolutional neural networks perform on medical images for identifying abnormalities, pathologies, white and grey matter inside brain tissues etc. These are a particular type of neural networks that uses a three-layered architecture. The primary layer in this architecture is known as the convolutional layer, the one which is initially defined, later a convolutional filter/kernel is used which multiplies itself with the given image to produce feature maps. Next to convolution kernel is a pooling layer which defines the most effective elements to be extracted once the convolution layer is filtered. There are two types of pooling techniques, one is max pooling and the other is average pooling. In max pooling, the feature with the highest value is elevated. While in average pooling, the average of the significant features is taken and selected among the elements. Followed by which is a fully connected layer. It is used after a series of convolution and pooling layers for high-level reasoning in the neural network.
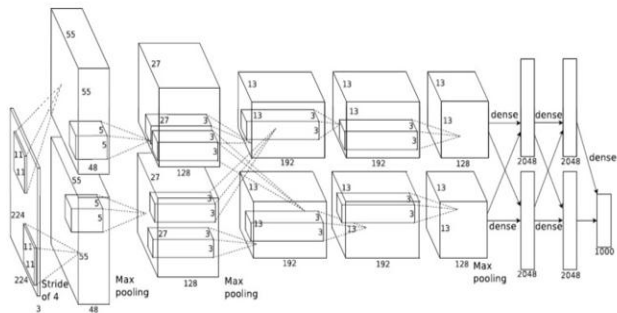

Fig 1. Convolution Architecture Proposed and Used for Image Classification by Alex Krizhevsky et al. [3]

IV. IMAGE SEGMENTATION ON CRX IMAGES

A. Dataset

The below mentioned segmentation convolutional neural networks are applied on NIH dataset, an open dataset where research can be carried out. This dataset contains over one lakh images of the X-Rays and CT Scans of the chest. All the photos are not classified based on abnormalities or any cancer identification. The total number of unique patients can be recognised in the data are around 30,000. Few radiology reports which are diagnosed are available, but almost many other models on the same dataset gave about 90% of accuracies using neural networks with weakly supervised learning [10]. CheXNet [8] was a model developed in the same dataset which almost identified several abnormalities and pathologies in the lungs. The architecture used on this dataset is a 121-layer convolutional network. In this work, we use the same data for lung segmentation. Below are the few sample images where the image segmentation using our model is carried out.
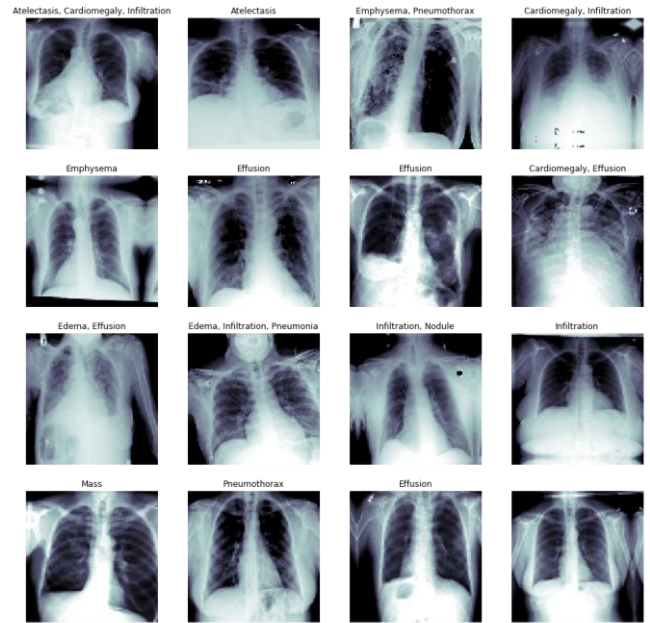

Fig 2. NIH dataset Peek Over, Images and their abnormalities described on the labels that are provided.

V. U-NET ARCHITECTURE FOR IMAGE SEGMENTATION

U-Net is a particular convolutional architecture that is applied to medical images to segment nerve channels or organs present. In this work, we separate the left and right lung from the given CT-Scan or the X-Ray using U-Nets. The architecture of U-Net is described below.

The U-Net architecture is comprised of two parts one on it's left, and one on it's right namely, contracting portion and expansive path respectively. Two 3x3 convolutions are repeated followed by ReLU activation function with two max pooling and two downsampling kernels with stride s = 2. In the right side which is expansive path the sampling was done to the feature map are served by 2x2 convolutions which reduce the number of feature channels to the half these are added with the cropped feature map from the left part i.e., contracting path wherein again two convolution operation of kernel size 3x3 are done followed by ReLU activation layers. In the last segment, a 1x1 sized convolution operation is done to harmonise every feature vector of 64-component. As we see the figure which is proposed as the official architecture of U-Net [13], it consists of 23 convolution operations. We also require to determine a proper input size so that max operations are implemented with even x and y values.
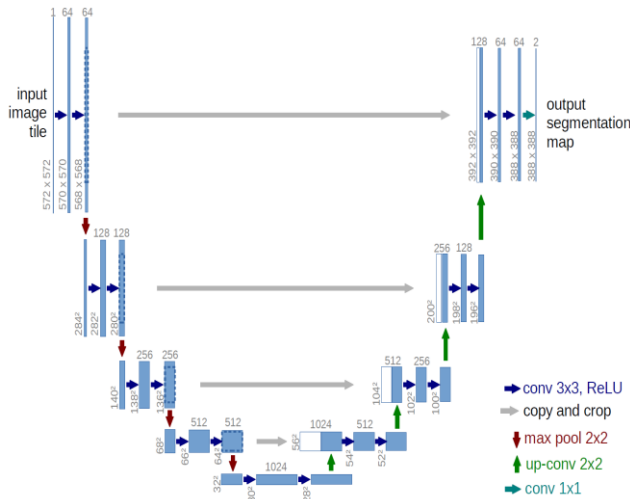
Fig 3. U-Net: Convolutional Networks for Biomedical Image Segmentation proposed by Olaf Ronneberger et al. [9]

VI. PROPOSED U-NET MODEL FOR SEGMENTATION

The proposed U-Net model for segmentation is a 22-layer convolutional neural network with 11 convolutional kernels, and 11 upsampling layers, wherein each layer starts with upsampling followed by convolutional layer and RELU (Linear rectified unit) activation function. The last layer is followed by Sigmoid activation function returning the segmented lungs from the medical image that is passed into the model. The total number of parameters that the model plays with is more than six thousand.

There are five crucial features that need to be identified while executing the model. Initially, the original image is converted into tensor, a multidimensional array representation, and the dimensions are passed into the network as input. The work is carried out on the preprocessed original images of variable size 256*256 and 126*126. A ground truth binary mask is generated over the lung parts using computer vision techniques by finding the mean and median of the pixel array of the input image. Below are the tabulated values of the mean and median of the sample chest x-ray images. Using these mean and median values, the model thereafter generates a new binary mask which is acquired from the training dataset. Ground Truth Mask overlay on the original Image is applied where most of the segmentation process is done. These are adjusted with the trained weights based on the accuracy that is generated by the model. Finally, the model creates a mask overlay on the original image where deconvolutional layers help in scaling up the image resolution.

Table I. Proposed U-Net Based Convolution Architecture

Layer	Layer Input Dimensions
input_1 (InputLayer)	(None, 256, 256, 3)
conv2d_1 (Conv2D)	(None, 256, 256, 64)
conv2d_2 (Conv2D)	(None, 256, 256, 64)
max_pooling2d_1 (MaxPooling2D)	(None, 128, 128, 64)
conv2d_3 (Conv2D)	(None, 128, 128, 64)
conv2d_4 (Conv2D)	(None, 128, 128, 64)
max_pooling2d_2 (MaxPooling2D)	(None, 64, 64, 64)
conv2d_5 (Conv2D)	(None, 64, 64, 64)
conv2d_6 (Conv2D)	(None, 64, 64, 64)
max_pooling2d_3 (MaxPooling2D)	(None, 32, 32, 64)
conv2d_7 (Conv2D)	(None, 32, 32, 64)
conv2d_8 (Conv2D)	(None, 32, 32, 64)
max_pooling2d_4 (MaxPooling2D)	(None, 16, 16, 128)
conv2d_9 (Conv2D)	(None, 16, 16, 128)
up_sampling2d_1 (UpSampling2D)	(None, 32, 32, 256)
conv2d_10 (Conv2D)	(None, 32, 32, 256)
conv2d_11 (Conv2D)	(None, 32, 32, 256)
concatenate_1 (Concatenate)	(None, 32, 32, 256)
conv2d_12 (Conv2D)	(None, 32, 32, 256)
up_sampling2d_2 (UpSampling2D)	(None, 64, 64, 256)
conv2d_13 (Conv2D)	(None, 64, 64, 256)
conv2d_14 (Conv2D)	(None, 64, 64, 256)
concatenate_2 (Concatenate)	(None, 64, 64, 256)
conv2d_15 (Conv2D)	(None, 64, 64, 256)
up_sampling2d_3 (UpSampling2D)	(None, 128, 128, 256)
conv2d_16 (Conv2D)	(None, 128, 128, 256)
conv2d_17 (Conv2D)	(None, 128, 128, 256)
concatenate_3 (Concatenate)	(None, 128, 128, 256)
conv2d_18 (Conv2D)	(None, 128, 128, 256)

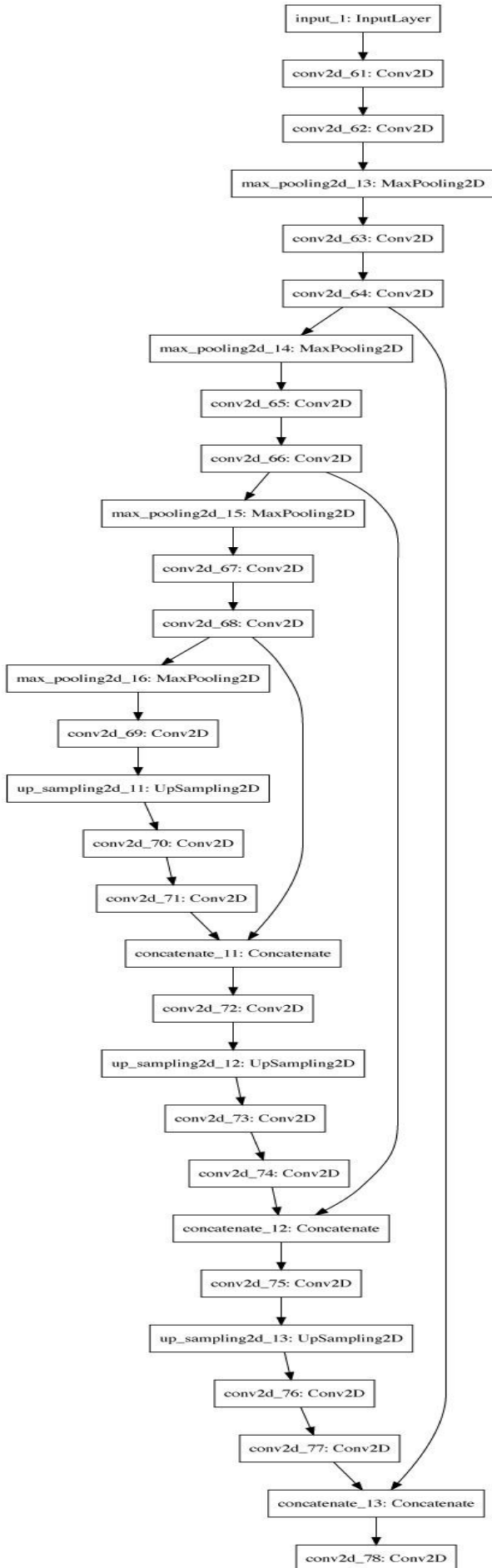# Lung Semantic Segmentation using Convolutional Neural Networks

Fig 4. Model Flow with Sampling, Pooling and Deconvolution Architectures

VII. TRAINING THE MODEL AND ANALYSIS

The U-Net proposed U-Net architecture have 5,469,473 parameters in all the layers with 23 layers in total including fully connected layers. The kernel size for all the convolutional layers is almost 3. The training model takes the input image dimensions as 256*256, else if the image is given in other formats or shapes it automatically reshapes the image into 256*256 size or 128*128 size based on the input parameters. The batch size considered for the training process is 16, which means the number of training examples that are processed at a time. The model uses cross entropy as the loss function which is mathematically represented by equation,

$$H_{y`}(y): = - \sum_i (y`_i \log(y_i) + (1 - y`_i) \log(1-y_i)) \qquad (1)$$

Let f_k be the k_{th} feature map and let w_c, k be the weight in the final classification layer for feature map k leading to pathology c. We obtain a map M_c of the most salient features used in classifying the image as having pathology c by taking the weighted sum of the feature maps using their associated weights. Formally,

$$M_c := \sum_K (w_{c,k} f_k)$$

For this U-Net architecture developed based on the training time and accuracy the ideal batch size was found to be 16. The standard 11 convolutional layer architecture was used to train on the NIH dataset, for preprocessing, feature extraction as well as image segmentation. The model took almost forty-five minutes of training time on a single CPU (8 cores of Intel i7), the same architecture and the model was trained on Single GPU which almost reduced over seventy per cent of training time. The received results from GPU performed 3.0 times faster when compared to CPU. The output images that are acquired from the trained model are also of the size 256*256 and are presented below. The mean and mode of the images are also calculated which gives the accuracy of the segmented images. The loss that has been iterated over epochs over test and validation data of the proposed U-Net model.
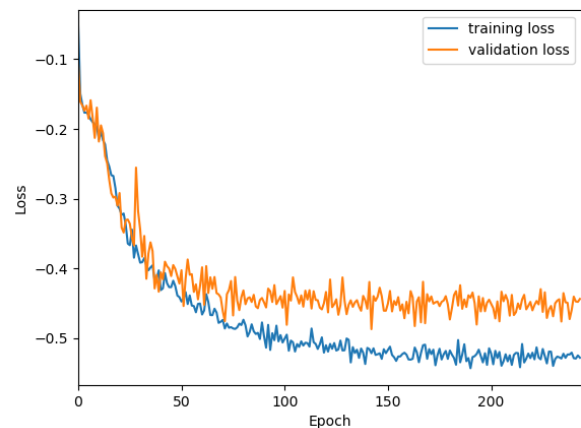

Fig 5. Graph showing the training loss and the validation loss of the developed U-Net Model.

VIII. RESULTS

The segmented results are given below with the filter and IoU values which represent the measure of how accurate the predictions are. The model was trained with seventy percent of NIH dataset and evaluated with the remaining thirty percent of the dataset. We have evaluated using IoU values and ROC curves. Below are the results and the segmented lung images evaluated from the NIH dataset.
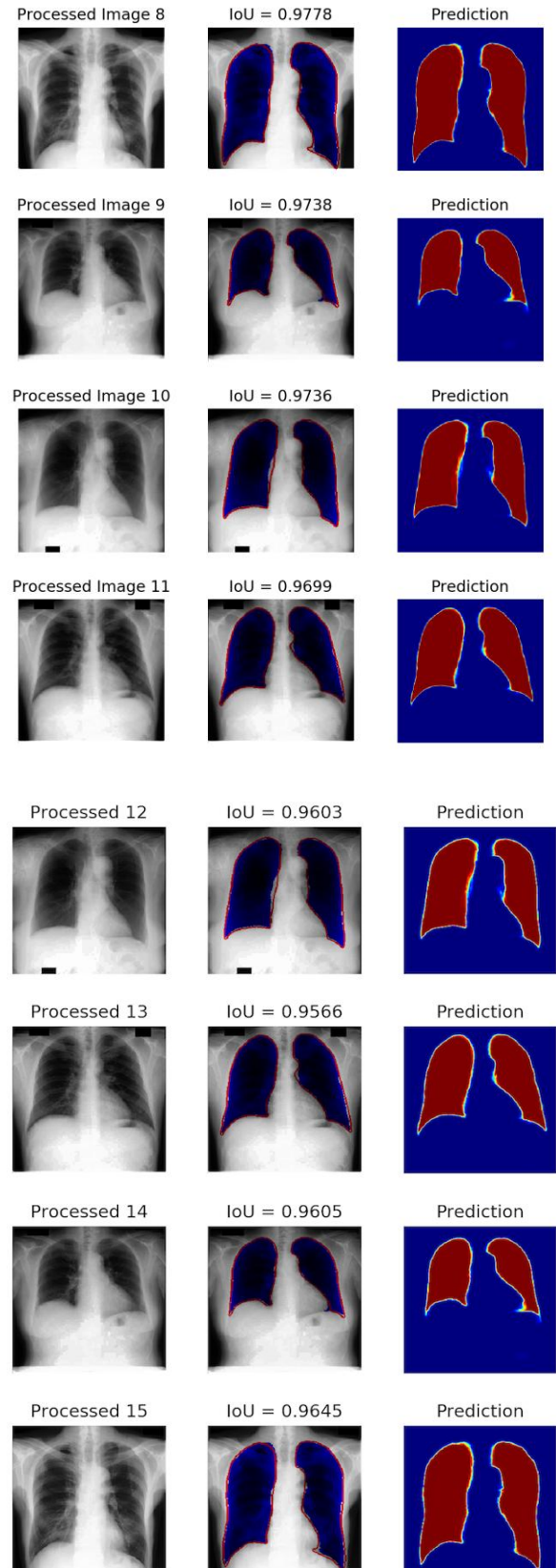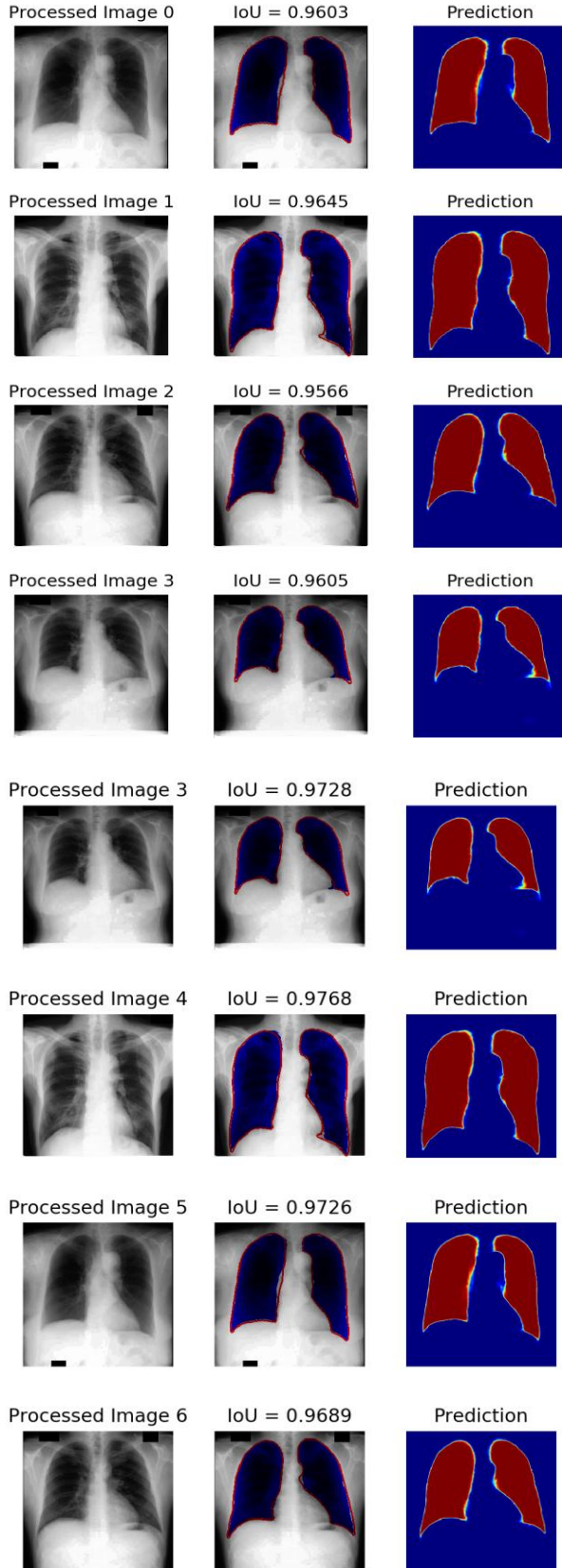
Table II. Metrics form Mask Overlays of Segmented Images

Image	Mean	Dice Values
Processed Image 0	0.956614709	0.9778257118
Processed Image 1	0.960528436	0.9798660476
Processed Image 2	0.960258242	0.9797257168
Processed Image 3	0.964508957	0.9819334999
Processed Image 4	0.956614703	0.9778257112
Processed Image 5	0.960528436	0.9798660475
Processed Image 6	0.960258243	0.9797257168
Processed Image 7	0.964508957	0.9819334655
Processed Image 8	0.940258242	0.9266147034
Processed Image 9	0.967389534	0.9233284360
Processed Image 10	0.934661563	0.9345682430
Processed Image 11	0.934561560	0.9232373829
Processed Image 12	0.934634616	0.9642983567
Processed Image 13	0.944661564	0.9835309378
Processed Image 14	0.923615600	0.9633328920
Processed Image 15	0.908302923	0.9302929101

IX. CONCLUSION

In this work, we have implemented a new convolutional architecture used for the segmentation of chest X rays. The modified U-Net architecture performs better than the traditional techniques, in here the training accuracy is greater than 98\% as observed in the above results. The training metrics are calculated using a statistical attribute called IoU (intersection over union) and ROC curves. We have tabulated these IOU and statistical characteristics for comparative analysis. The network is trained over 40,000 images of NIH dataset and evaluated over 15,000 copies. Future work includes detecting pathologies in the segmented lung X-Rays using CheXNet architecture.

REFERENCES

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
2. Going deeper with convolutions Christian Szegedy Google Inc. Wei Liu University of North Carolina, Chapel Hill Yangqing Jia Google Inc. Pierre Sermanet Google Inc. Scott Reed University of Michigan Dragomir Anguelov Google Inc. Dumitru Erhan Google Inc. Vincent Vanhoucke Google Inc. Andrew Rabinovich Google Inc. arXiv:1409.4842v1 [cs.CV] 17 Sep 2014.
3. Alex Krizhevsky, Ilya Sutskever, and Geoff Hinton. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25, pages 1106–1114, 2012.
4. Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. CoRR, abs/1312.4659, 2013.
5. Thomas Serre, Lior Wolf, Stanley M. Bileschi, Maximilian Riesenhuber, and Tomaso Poggio. Robust object recognition with cortex-like mechanisms. IEEE Trans. Pattern Anal. Mach. Intell., 29(3):411–426, 2007.
6. P. Krahenbuhl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge ¨ potentials. In NIPS, 2011.
7. CNN-based Segmentation of Medical Imaging Data Barıs¸ Kayalıbay, Grady Jensen† Patrick van der Smagt arXiv:1701.03056v1 [cs.CV] 11 Jan 2017.
8. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng arXiv:1711.05225 [cs.CV]
9. U-Net: Convolutional Networks for Biomedical Image Segmentation Olaf Ronneberger, Philipp Fischer, Thomas Brox.
10. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation Ozg¨un C¸ i¸cek ¨ 1,2 , Ahmed Abdulkadir.

AUTHORS PROFILE

N. Sarada is currently pursuing Ph.D at Koneru Lakshmaiah Education Foundation, Guntur Dist., A.P., India. She completed her M.Tech from Acharya Nagarjuna University, Guntur, A.P. Her Research interests include Neural Networks, Computer Vision & Image Processing, and Cloud Computing. She had published five Research Papers in Scopus Indexed Journals on Cloud Computing & Deep learning algorithms.

Dr. K. Thirupathi Rao, M.Tech., Ph.D., is currently working as a Professor and BOS Chairman Department of Computer Science and Engineering, Associate Dean Academics ,at Koneru Lakshmaiah Education Foundation, Guntur Dist., A.P., India. His research interests include Cloud Computing, Operating Systems, Neural Networks and Algorithm Analysis. He got published more than 30 technical papers in International journals and Conferences. At present he is the Program Committee Member (PC) of various International journals. He received his Doctoral degree from Acharya Nagarjuna University, Guntur, A.P. for the topic related to Scheduling in the area of Cloud Computing. He is the Chief Technical Advisory Board Member, Chief Editor, Editor and Technical Reviewer of many International Journals. He has guided and is guiding a number of research scholars.