Ensemble Lung Segmentation System Using Deep Neural Networks

Redha Ali, Russell C. Hardie
Department of Electrical and Computer Engineering
University of Dayton
300 College Park, Dayton, Ohio 45469
{almahdir1, rhardie1}@udayton.edu

Hussin K. Ragb
Department of Engineering
School of Electrical and Computer Engineering
Christian Brothers University
Memphis, Tennessee
hragb@cbu.edu

2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) | 978-1-7281-8243-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/AIPR50011.2020.9425311

***Abstract*—Lung segmentation is a significant step in developing computer-aided diagnosis (CAD) using Chest Radiographs (CRs). CRs are used for diagnosis of the 2019 novel coronavirus disease (COVID-19), lung cancer, tuberculosis, and pneumonia. Hence, developing a Computer-Aided Detection (CAD) system would provide a second opinion to help radiologists in the reading process, increase objectivity, and reduce the workload. In this paper, we present the implementation of our ensemble deep learning model for lung segmentation. This model is based on the original DeepLabV3+, which is the extended model of DeepLabV3. Our model utilizes various architectures as a backbone of DeepLabV3+, such as ResNet18, ResNet50, Mobilenetv2, Xception, and inceptionresnetv2. We improved the encoder module of DeepLabV3+ by adjusting the receptive field of the Spatial Pyramid Pooling (ASPP). We also studied our algorithm's performance on a publicly available dataset provided by Shenzhen Hospital, that contains 566 CRs with manually segmented lungs (ground truth). The experimental result demonstrate the effectiveness of the proposed model on the dataset, achieving an Intersection-Over-Union (IoU, Jaccard Index) score of 0.97 on the test set.**

***Index Terms*—Chest Radiographs, Lung Segmentation, Convolutional Neural Networks, DeepLabV3+, Ensemble model, Computer Aided Diagnosis.**

I. Introduction

Automated lung segmentation of pulmonary Chest Radiography (CR) images is crucial to most CAD systems. CR is a widely used imaging modality to evaluate numerous lung diseases that include but are not limited to COVID-19, pneumonia, lung cancer, and tuberculosis. Segmentation of the lungs plays a crucial role in these diagnoses [1]. CR's availability and low cost have made CR the most widely used medical imaging modality for detecting lung disease. Detecting the lung regions in CR images provides radiologists insights in terms of shape, size, and other related geometrical properties [2]. Shape, size, and total lung volume vary for each patient, which could provide signs of serious diseases [3]. This makes it challenging for many researchers to develop CAD systems to help radiologists evaluate CR images.

Lung segmentation has attracted great interest in the field of Machine Learning. Various methods have been proposed in the literature [1]–[12] to develop computational methods to help radiologists read chest images. In [1], [5] lung segmentation is based on optimal thresholding to select a threshold value using the unique characteristics of the lungs. Another lung segmentation is based on wavelength transformation and optimal thresholding [4]. Several segmentation methods are presented in [3], [5]–[7], based on anatomical atlases-based, active shape models (ASM) and a pixel-based classification method. Recently, favorable results were obtained in the field of medical imaging segmentation by deep learning, which related to a CXR image analysis [2], [8]–[11]. The research work presented in [2] serves as benchmarks for our model.

Recently, deep learning-based semantic segmentation is providing state-of-the-art performance in multiple applications, including medical imaging applications [8], [12], [13]. DeepLabV3+ has been a favorite segmentation model in the field of medical imaging [14]. DeepLabV3+ has beendeveloped by Google's DeepLab as discussed in L.C. Chen et al.'s paper 'Encoder decoder with atrous separable convolution for semantic image segmentation.' [15]. DeepLab deep learning-based semantic segmentation series has come along for three versions from DeepLabV1 [16], DeepLabV2 [17], and DeepLabV3 [18]. In this research, we propose an Ensemble DeepLabV3+ based architecture for automated lung segmentation for CRs publicly available datasets, thus setting a benchmark for future study efforts. We evaluate our ensemble approach by performing hold-out validation. We show that our ensemble approaches perform well using this modality compared to its benchmark algorithms. We study the performance of different backbone architectures to improve segmentation accuracy further.

The remainder of this paper is organized as follows. Section II provides a brief description of the dataset that are employed in this research. Section III presents the ensemble DeepLabV3+ architecture proposed in this research. Section IV presents the experimental results obtained using the proposed methods. Finally, conclusions are offered in Section

V.

II. Materials

In this study, we utilized the Shenzhen Hospital (SH) dataset of CR images. The SH dataset has been captured from Shenzhen No. 3 People's Hospital in Shenzhen, China [19]. The CRs images have been collected from patients as part of the daily routine using Philips digital radiography. The SH dataset was provided in PNG image format with a resolution of 3000×3000 pixels. SH holds 566 CR samples with manual lung segmentation masks. These ground truth masks for SH dataset were prepared manually by domain experts. We utilize this dataset to study the robustness of our proposed model. The dataset is publicly available in https://lhncbc.nlm.nih.gov/publication/pub9931 .

Fig. 1: Example of the original image with manual ground truth mask for a random case from the SH dataset

We divided the dataset into a group of 466 CRs and 100 CRs for training and testing, respectively. Each image is resized to fit the input of Xception and InceptionResnetV2 with an image input size of 299×299 pixels and the input of ResNet18, ResNet50, and Mobilenetv2 with an image input size of 224×224 pixels. Figure 1 shows an example of a random case from the SH dataset with a ground truth lung mask.

III. DeepLabV3+ Architecture

In this section, we briefly describe the DeepLabV3+ as shown in Figure 2 which is used as our base encoder and decoder module and to build the ensemble model. We also present a modified encoder module of the DeepLabV3+ model, which further improves the performance with competitive computation.

A. Encoder decoder architecture

The segmentation tasks require both encoding and decoding units. Figure 2 shows the encoder and decoder blocks. One of the most recent encoder-decoder type architectures is a deep learning-based semantic segmentation. The encoder unit features extraction and downsamples blocks that progressively decrease the feature dimensions and capture deeper semantic information. The decoder unit reconstructs or retrieves the spatial information and upsamples the feature map size to equal the original image from the downsampled feature map.

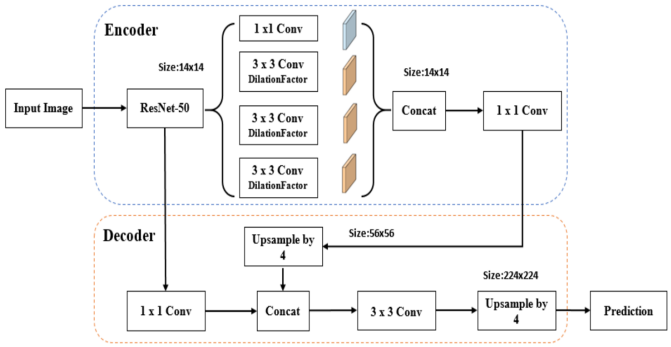

Fig. 2: DeepLabV3+ architecture.

In this study, We have used several different CNNs architecture as backbone including ResNet-18, ResNet-50, Mobilenetv2, Xception and InceptionResnetV2. The encoder block involves the input feature maps that are first downsampled by a 1×1 convolution with an output that feeds a subsequent 3×3 convolution layer. Before adding the result to the input feature map, another 1×1 convolution is executed to meet the input feature maps' depth as shown in Figure 2. The inverted residual block introduced in [20] has been applied. It first utilizes a 1×1 convolution to increase the depth of feature maps and a 3×3 depthwise convolution. The inverted residual block technique is more computationally efficient and requires fewer learnable parameters than the conventional residual block.

Moreover, the atrous convolution, also known as dilated convolution, has been used in the encoder block. The atrous convolution functions on an input feature map (x) as follows:

$$y[i] = \sum_{k==1}^{k} x[i + r.k]\, w[k], \tag{1}$$

where i is the pixel location in the output feature map y. The convolution filter w is applied over the input feature map x where the atrous rate r corresponds to the stride with which we sample the input signal. When $r = 1$, it is the standard

convolution. When $r > 1$, it is the atrous convolution, which is the stride to sample the input sample during convolution. Also, the input signal is sampled alternatively. Atrous convolution enables the model to expand the field of view of filters to incorporate a more extensive context. Atrous convolution blocks consist of one 1×1 convolution layer and three 3×3 convolution layers with different dilation rates = $(6, 12, 18)$, as we can see in Figure 2. However, we have adjusted those rates in our ensemble architecture to $(4, 8, 16)$ in the second encoder, as shown in Figure 3.

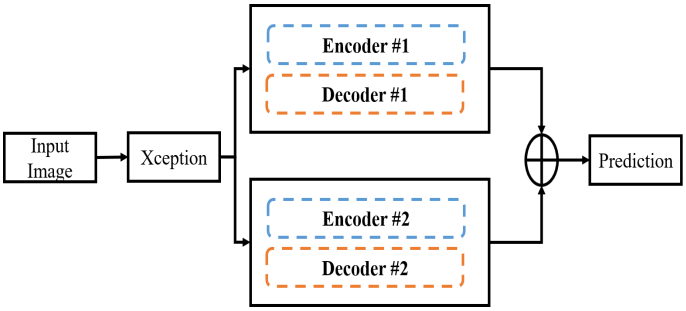

Fig. 3: Illustration of our ensemble DeepLabV3+ architecture for lung segmentation.

Our ensemble model consists of the Xception and InceptionResnetV2 networks, for Versions 1 and 2, respectively. The feature extraction step, followed by the atrous spatial pyramid pooling (ASPP). The final feature map in Encoder 1 has a different dilation rate than Encoder 2. After that, we have to reconstruct the encoder's output to the original sized segmentation map from a small feature map. Instead of immediately upsampling the feature map, it is accomplished in two steps, as demonstrated in Figure 3 . We first upsample the ASPP output feature map by factor 4 using a transposed convolution layer with bilinear upsampling and then concatenated with the corresponding feature map from the encoder stage. Finally, a 3×3 convolution is conducted before upsampling that feature map by 4 to generate the final segmentation output. We then created the final probability matrix by taking the average of the two output segmentation masks that have been generated. The probability matrix is later thresholded to produce the final segmentation masks, as shown in Figure 3 .

IV. EXPERIMENTAL RESULTS

This section presents the results obtained using the ensemble DeepLapv3+ approach for the SH dataset. Furthermore, we present the performance measure in terms of the confusion matrix obtained using the pixel-based classification approach. Also, we present the segmentation results, including Global Accuracy, Mean Accuracy, Mean Intersection over Union (IOU), based on the obtained segmentation and ground truth.

First, we evaluate our model's performance on a test set that consists of 100 CRs from the SH dataset. Figure 4 shows the normalized confusion matrix achieved using the pixel-based classification method on test cases. The overall accuracy of 96.99% and 99.51% for lung and not lung, respectively, in terms of classification of pixels.

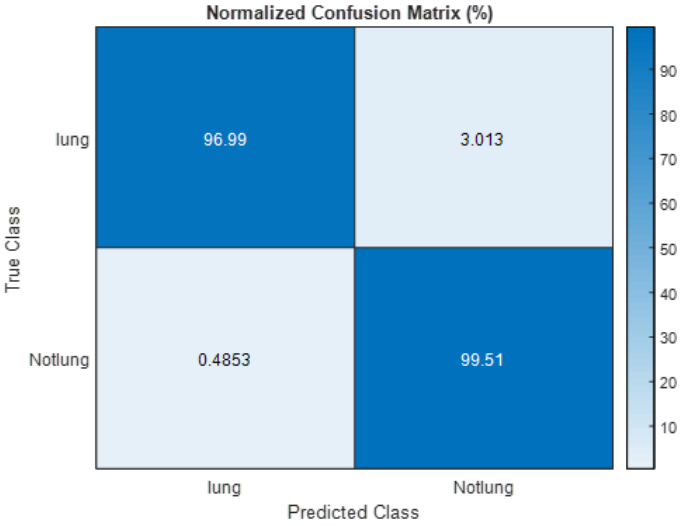

Fig. 4: Normalized Confusion Matrix obtained for SH Dataset.

Second, we offer a visual result of the proposed methods applied to the SH testing set. Figure 5 shows the output of the ensemble model against the ground truth. The results indicate that the proposed method's segmentation performance is better than the state-of-the-art methods reported in [2].

In this study, ResNet18, ResNet50, Mobilenetv2, Xception, and InceptionResnetV2 have been used as backbones for the DeepLabV3+ model. Only Xception and InceptionResnetV2 have been used as the backbone for the ensemble DeepLabV3+ model named Ensemble Model V1 and Ensemble Model V2, respectively. We trained these techniques on CRs images from the SH dataset. Each approach has a Global Accuracy, Mean Accuracy, Mean Intersection over Union (IOU), and BF Score, demonstrating their performance on several metrics. Table I shows that Ensemble Model V2 achieved the highest efficiency in all metrics. ResNet18 yielded the lowest performance values with accuracy and an IOU of 97.48% and 0.950, respectively.

Moreover, we offer a computation time comparison in Table II. The computation time was computed for the entire testing set of 100 test cases on a desktop with an i7 processor at 2.8 GHz with 32 GB RAM and NVIDIA TITAN RTX. We compare the performance and computation time for lung segmentation using the proposed algorithm and other algorithms. In terms of computation time, ResNet-18 outperforms other approaches.

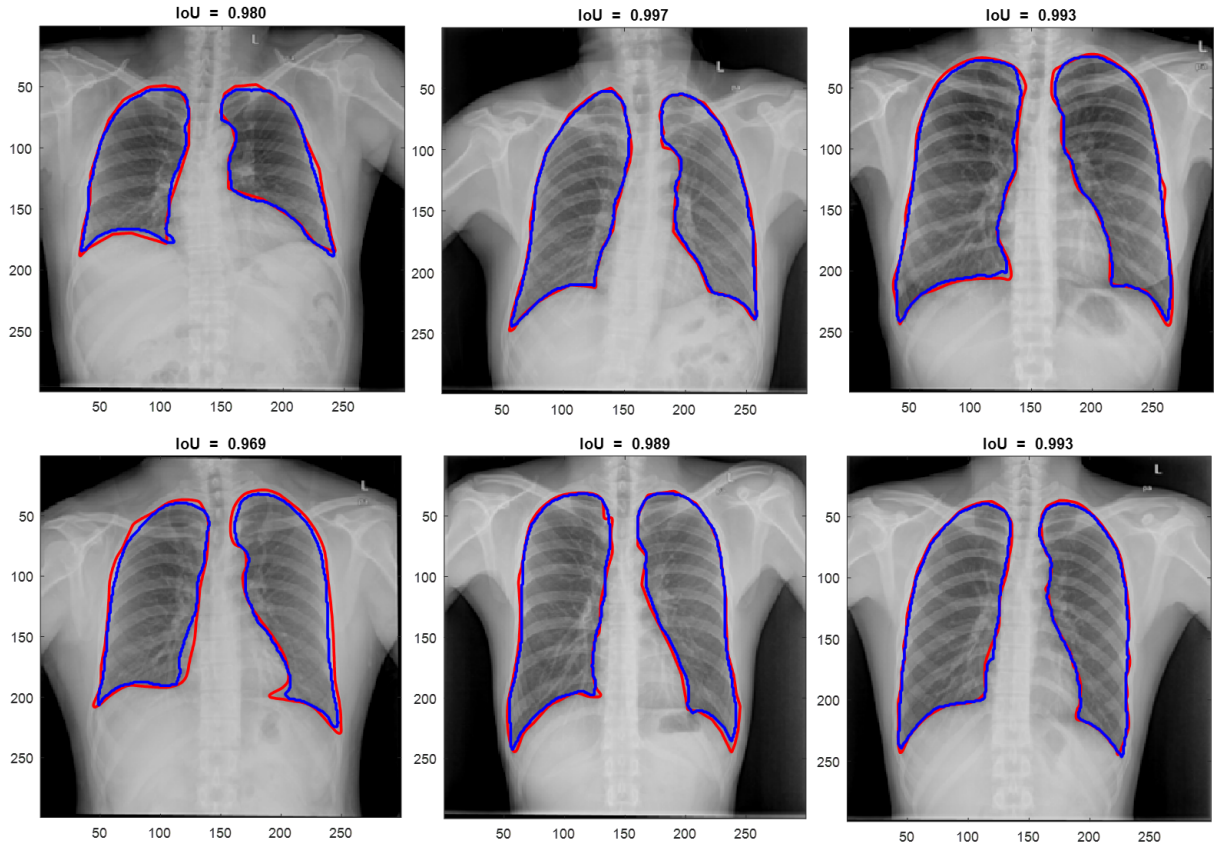

Fig. 5: Visual comparison of lung segmentation of our proposed method with IOU scores. The ground truth (red) and proposed methods (blue).

TABLE I: DeepLabV3+ performance comparison with different backbones.

Model	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BF Score
ResNet-18	98.14%	97.48%	0.95	0.963	0.905
ResNet-50	98.12%	97.73%	0. 950	0.963	0.905
Mobilenetv2	98.18%	97.37%	0. 951	0. 964	0.902
Xception	98.45%	97.41%	0. 958	0. 969	0. 919
InceptionResnetV2	98.73%	98.11%	0. 965	0. 975	0. 937
Ensemble Model V1	98.89%	**98.25%**	**0.97**	**0. 978**	0.946
Ensemble Model V2	**98.91%**	**98.25%**	**0.97**	**0. 978**	**0.948**

TABLE II: Performance comparison of computation time for SH dataset.

Model	Computation time (Entire test set) (Seconds)	
	GPU	CPU
ResNet-18	1.74	22.17
ResNet-50	1.89	23.56
Mobilenetv2	1.85	30.92
Xception	2.43	31.16
InceptionResnetV2	3.07	58.85
Ensemble Model V1	2.37	36.34
Ensemble Model V1	3.47	70.09

V. CONCLUSIONS

In this research, we introduced the ensemble method that produces computationally efficient and accurate lung segmentation for CRs. Our model integrates two different CNN models that have different architectures. The proposed method provided good performance under different testing conditions proving its robustness and efficacy. We utilized two encoders and decoders with different atrous convolution dilation rates, which provide a wider field of view at the same computational cost. Results indicate that the ensemble model produces significantly higher accuracy than the use of a single encoder and decoder. The proposed method could be used in computer aided detection systems for various lung disease detection applications. Automated lung segmentation would be crucial for medical imaging specialists and would improve their workflow.

REFERENCES

[1] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images," *IEEE transactions on medical imaging*, vol. 20, no. 6, pp. 490–498, 2001.

[2] B. N. Narayanan and R. C. Hardie, "A computationally efficient u-net architecture for lung segmentation in chest radiographs," in *2019 IEEE National Aerospace and Electronics Conference (NAECON)*. IEEE, 2019, pp. 279–284.

[3] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE transactions on medical imaging*, vol. 33, no. 2, pp. 577–590, 2013.

[4] P. Korfiatis, S. Skiadopoulos, P. Sakellaropoulos, C. Kalogeropoulou, and L. Costaridou, "Combining 2d wavelet edge highlighting and 3d thresholding for lung segmentation in thin-slice ct," *The British journal of radiology*, vol. 80, no. 960, pp. 996–1004, 2007.

[5] R. C. Hardie, S. K. Rogers, T. Wilson, and A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs," *Medical Image Analysis*, vol. 12, no. 3, pp. 240–258, 2008.

[6] B. N. Narayanan, R. C. Hardie, T. M. Kebede, and M. J. Sprague, "Optimized feature selection-based clustering approach for computer-aided detection of lung nodules in different modalities," *Pattern Analysis and Applications*, vol. 22, no. 2, pp. 559–571, 2019.

[7] S. G. Armato III, M. L. Giger, and H. MacMahon, "Automated lung segmentation in digitized posteroanterior chest radiographs," *Academic radiology*, vol. 5, no. 4, pp. 245–255, 1998.

[8] R. Ali, R. C. Hardie, B. Narayanan Narayanan, and S. De Silva, "Deep learning ensemble methods for skin lesion analysis towards melanoma detection," in *2019 IEEE National Aerospace and Electronics Conference (NAECON)*, 2019, pp. 311–316.

[9] J. C. Souza, J. O. Bandeira Diniz, J. L. Ferreira, G. L. França da Silva, A. Corrêa Silva, and A. C. de Paiva, "An automatic method for lung segmentation and reconstruction in chest x-ray using deep neural networks," *Computer Methods and Programs in Biomedicine*, vol. 177, pp. 285 – 296, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260719303517

[10] B. N. Narayanan, M. S. De Silva, R. C. Hardie, N. K. Kueterman, and R. Ali, "Understanding deep neural network predictions for medical imaging applications," *arXiv preprint arXiv:1912.09621*, 2019.

[11] P. Gang, W. Zhen, W. Zeng, Y. Gordienko, Y. Kochura, O. Alienin, O. Rokovyi, and S. Stirenko, "Dimensionality reduction in deep learning for chest x-ray analysis of lung cancer," in *2018 tenth international conference on advanced computational intelligence (ICACI)*. IEEE, 2018, pp. 878–883.

[12] H. K. Ragb, I. T. Dover, and R. Ali, "Fused deep convolutional neural network for precision diagnosis of covid-19 using chest x-ray images," *arXiv preprint arXiv:2009.08831*, 2020.

[13] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[14] E. Yi and Y. Liu, "A pilot exploratory study of the potentials of deep learning methods in cancer image segmentation and classification," *bioRxiv*, p. 820563, 2019.

[15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[18] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[19] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative imaging in medicine and surgery*, vol. 4, no. 6, p. 475, 2014.

[20] A. Howard, A. Zhmoginov, L.-C. Chen, M. Sandler, and M. Zhu, "Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation," 2018.