



Automatic Segmentation of Pulmonary Lobes Using a Progressive Dense V-Net

Abdullah-Al-Zubair Imran^{1,2}, Ali Hatamizadeh^{1,2}, Shilpa P. Ananth², Xiaowei Ding^{1,2}(✉), Demetri Terzopoulos^{1,2}, and Nima Tajbakhsh²

¹ University of California, Los Angeles, CA 90095, USA

² VoxelCloud Inc., Los Angeles, CA 90024, USA

xding@voxelcloud.io

Abstract. Reliable and automatic segmentation of lung lobes is important for diagnosis, assessment, and quantification of pulmonary diseases. The existing techniques are prohibitively slow, undesirably rely on prior (airway/vessel) segmentation, and/or require user interactions for optimal results. This work presents a reliable, fast, and fully automated lung lobe segmentation based on a progressive dense V-network (PDV-Net). The proposed method can segment lung lobes in one forward pass of the network, with an average runtime of 2 s using 1 Nvidia Titan XP GPU, eliminating the need for any prior atlases, lung segmentation or any subsequent user intervention. We evaluated our model using 84 chest CT scans from the LIDC and 154 pathological cases from the LTRC datasets. Our model achieved a Dice score of 0.939 ± 0.02 for the LIDC test set and 0.950 ± 0.01 for the LTRC test set, significantly outperforming a 2D U-net model and a 3D dense V-net. We further evaluated our model against 55 cases from the LOLA11 challenge, obtaining an average Dice score of 0.935—a performance level competitive to the best performing team with an average score of 0.938. Our extensive robustness analyses also demonstrate that our model can reliably segment both healthy and pathological lung lobes in CT scans from different vendors, and that our model is robust against configurations of CT scan reconstruction.

Keywords: Lung lobe segmentation · CT · Progressive dense V-Net Fissure · 3D CNN

1 Introduction

Human lungs are divided into five lobes. The right lung has three lobes, namely, right upper lobe (RUL), right middle lobe (RML), and right lower lobe (RLL), which are separated by a minor and a major fissure, whereas the left lung has

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-00889-5_32) contains supplementary material, which is available to authorized users.

two lobes, namely, left upper lobe (LUL) and left lower lobe (LLL), separated by a major fissure. Figure 1 shows the five lobes separated by major and minor fissures in a coronal CT slice. Each of the five lobes is functionally independent as they have separate bronchial and vascular systems.

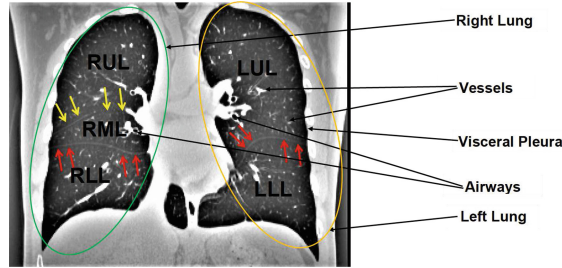


Fig. 1. A coronal lung CT slice with visible fissures. Major fissures are denoted by red arrows and yellow arrows denote the minor fissure.

Automatic lobe segmentation is important for both clinical and technical purposes. In clinical practice, doctors very often base their assessment of a disease severity and the corresponding treatment plan on the affected lung lobe. As such, upon encountering a disease or lesion in the lung, radiologists may navigate through the nearby slices to identify the affected lobe, especially when the fissure lines are not clearly visible in the target slice. An automatic lobe segmentation model can therefore shorten the CT reading session by continually informing the radiologists about their location in the lung anatomy. From the technical perspective, accurate lung lobe segmentation can improve several subsequent clinical tasks, including nodule malignancy prediction (cancers mostly occur in the left or right upper lobes), automatic lobe-aware report generation for each nodule, and assessment and quantification of pulmonary diseases, by narrowing down the search space to the lung lobes most-likely to be affected. However, identifying fissures poses a challenge for both human and machine perception. *First*, fissures are most often incomplete, not extending to the lobar boundaries. Several studies in the literature have confirmed the incompleteness of fissures as a very common phenomenon [1]. *Second*, the visual characteristics of lobar boundaries can change in presence of pathologies. Such morphological changes could also be related to the varying thicknesses, locations, and shapes of the fissures. *Third*, there also exist other fissures in the lungs that can be misinterpreted as the major or minor fissures that separate the lobes (e.g., accessory fissures andazygos fissures).

To address the need for accurate and robust lobe segmentation, we have proposed a fully automatic and reliable deep learning solution via progressive dense V-net (PDV-net). The PDV-net model takes entire CT volume and through three dense feature blocks, generates the segmentation progressively improving at each pathway. Our model generates accurate segmentation of the lung lobes

in about 2 s in only a single forward pass of the network, eliminating the need for any user interactions or any prior segmentation of lungs, vessels, or airways, which are common assumptions in the design of existing models.

2 Related Work

Various automatic and semi-automatic approaches have been proposed for lung lobe segmentation. Despite the methodological differences, the existing approaches are similar in that they require either prior segmentation of airways and vessels (e.g., Bragman et al. [2]) or demand previously defined atlases (e.g., van Rikxoort et al. [11] and Ross et al. [13]). Therefore, they suffer from slow execution time, cumbersome process of generating the atlas, and relatively lower performance for pathological cases. A significant shift from this common trend is the work of George et al. [4] wherein a 2D fully convolutional neural network followed by a 3D random walker algorithm is used to segment lobes. However, their method still relied on the random walker algorithm whose optimal parameters could change from one dataset to another. It is most desirable to have an end-to-end solution that does not rely on any subsequent heuristic method.

In the presented work, we mitigate the aforementioned limitations, namely reliance on prior masks, slow runtime, and lack of robustness by an end-to-end, single-pass, deep-learning-based framework that does not rely on any prior airway/vessel segmentation, anatomical knowledge, or atlases.

3 Method

We combine the ideas from dense V-network [5] and progressive holistically nested networks [7] to obtain a new architecture: progressive dense V-network (PDV-net), an end-to-end solution for organ segmentation in 3D volumetric data. Our proposed architecture is illustrated in Fig. 2. As seen, the input to the network is first down-sampled and concatenated with a strided $5 \times 5 \times 5$ convolution of the input with 24 kernels. The concatenation result is then passed to 3 dense feature blocks, each consisting of 5, 10, and 10 densely-wired convolution layers respectively. The growth rates of dense blocks are set to 4, 8, and 16 respectively. All the convolutional layers in a dense block have a kernel size of $3 \times 3 \times 3$ and are followed by batch normalization and parametric rectified linear units (PReLU).

Consecutively, the outputs of the dense feature blocks are utilized in low and high resolution passes via convolutional down-sampling and skip connections. This enables the generation of feature maps at three different resolutions. The outputs of the skip connections of the second and third dense feature blocks are further up-sampled in order to be consistent with the size of the output in the first skip connection. The feature maps from skip1 are passed to a convolutional layer followed by a softmax, which outputs the probability maps. In the second pathway, the feature maps from skip1 and skip2 are merged and the output probability maps are produced by a convolutional layer followed by softmax.

Similarly, we get the final segmentation result from the merged feature maps resulted from the skip2 and skip3 connections. Unlike dense V-net, PDV-net generates the final output by progressively improving the outputs at previous pathways. To train the suggested architecture, we choose to use a dice-based loss function [10] at each stage of the progressive architecture.

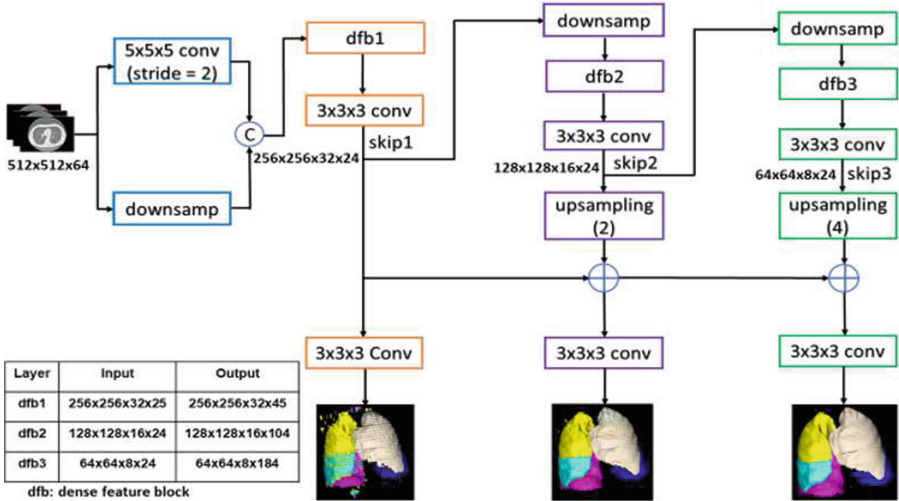


Fig. 2. PDV-net model for the segmentation of lung lobes. Segmentation outputs at different pathways are progressively improved for the final result.

4 Experiments

Datasets: We used 3 public datasets to evaluate our models. **First**, we selected a subset of chest CTs (354 cases) from the publicly available LIDC dataset for annotation. To ensure variability in the data, CT scans were selected such that both challenging and visible fissures are well-represented in the dataset. The ground truth masks were generated in a semi-automatic fashion by multiple observers using 3D Slicer. To mitigate bias in the ground truth, the generated masks were later refined and validated by an expert radiologist. The dataset was split into 270 training and 84 test cases. 10% of the training set was utilized as the validation set. **Second**, we selected 154 CTs from LTRC database. The LTRC dataset includes lobe masks for pathological cases that have clear evidence of COPD or ILD diseases, including emphysema and fibrosis. The LTRC cases allow us to measure the robustness of our model against pathologies in the lungs. **Third**, we used 55 cases of the Lobe and Lung Analysis (LOLA11) challenge [9] and submitted the results to the challenge organizers for evaluation.

Baselines for Comparison: We used a U-Net architecture [12] and a dense V-Net for comparison. The former is used in the most recent published article [4] for lung lobe segmentation and the latter is a strong baseline for comparison, which we use for the first time for lung lobe segmentation.

Implementation Details: For the proposed model and dense V-Net, the training volumes were first normalized, followed by rescaling to $512 \times 512 \times 64$, using 1 NVIDIA Titan XP GPU. Due to the large memory footprint of the model, the gradient check-pointing method [3] was used for memory-efficient back-propagation. In addition, batch-wise spatial dropout [5] is incorporated for regularization purposes. The training was performed on a Intel(R) Xeon(R) CPU E5-2697 v4@2.30 GHz machine. We used the Adam optimizer [8] with a learning rate of 0.01 and a weight decay of 10^{-7} .

For the 2D U-net implementation, we trained the network with axial slices from all the training volumes, each sized 512×512 and normalized to have values between 0 and 1. To avoid over-fitting to the background class, we used only the axial slices, wherein at least one lung lobe is present. We further used the Adam optimizer with a learning rate of 5×10^{-5} and batches of 10 images.

Table 1. Performance comparison of the proposed 3D progressive dense V-net with the 2D U-net and 3D dense V-net models in segmenting 84 LIDC and 154 LTRC cases. Mean Dice score and standard deviation for each lobe have been reported.

| Dataset | Model | RUL | RML | RLL | LUL | LLL | Overall |
|-----------|------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| LIDC(84) | 2D U-Net | 0.908 ± 0.049 | 0.844 ± 0.076 | 0.940 ± 0.054 | 0.959 ± 0.042 | 0.949 ± 0.056 | 0.920 ± 0.043 |
| | 3D DV-Net | 0.929 ± 0.036 | 0.873 ± 0.058 | 0.951 ± 0.018 | 0.958 ± 0.020 | 0.949 ± 0.041 | 0.932 ± 0.023 |
| | 3D PDV-Net | 0.937 ± 0.031 | 0.882 ± 0.057 | 0.956 ± 0.017 | 0.966 ± 0.014 | 0.966 ± 0.037 | 0.939 ± 0.020 |
| LTRC(154) | 2D U-Net | 0.914 ± 0.039 | 0.866 ± 0.054 | 0.952 ± 0.023 | 0.961 ± 0.023 | 0.954 ± 0.021 | 0.929 ± 0.025 |
| | 3D DV-Net | 0.949 ± 0.013 | 0.901 ± 0.021 | 0.959 ± 0.009 | 0.961 ± 0.007 | 0.958 ± 0.012 | 0.946 ± 0.008 |
| | 3D PDV-Net | 0.952 ± 0.011 | 0.908 ± 0.020 | 0.961 ± 0.008 | 0.966 ± 0.006 | 0.960 ± 0.010 | 0.950 ± 0.007 |

LIDC Results: Table 1 shows the calculated overall and lobe-wise Dice scores for each of the models. The proposed progressive dense V-net model, with an overall score of 0.939 ± 0.020 , significantly outperformed the 2D model, with an overall score of 0.9201 ± 0.0431 . As is evident in Table 1, the 3D progressive dense V-net yields consistently larger Dice score for each of the lung lobes against both dense V-net and U-net. Moreover, the lower standard deviation for each lobe indicates that the progressive model is more robust. We have also shown a qualitative comparison between the 3 models in Fig. 3 where the lung fissures are better captured by our progressive dense V-net model than by 2D U-net and dense V-net.

We further used Bland-Altman plots to measure the agreement between our progressive dense V-net and ground truth segmentations of the 84 LIDC cases (Fig. 4). A good agreement was observed between our segmentation model and ground truth in every plot (Lung and LLL being the two best agreements).

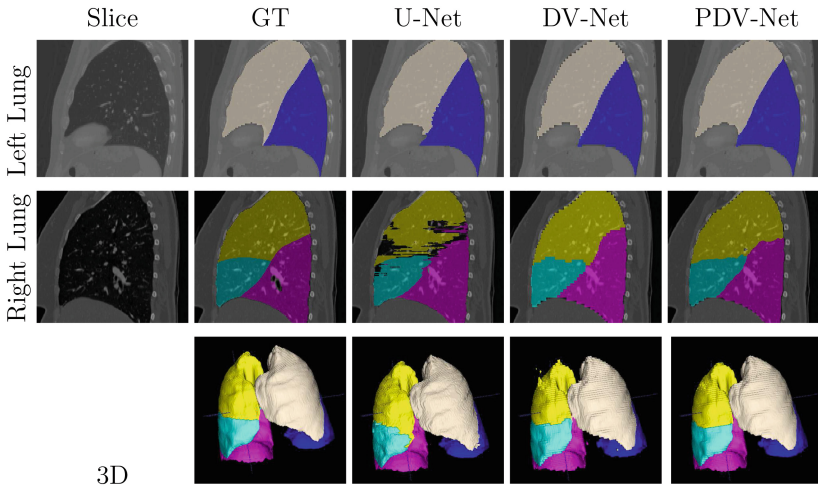


Fig. 3. Qualitative comparison between the proposed 3D progressive dense V-Net (PDV-Net), dense V-Net (DV-Net), and U-net. Note how noisy patches are removed from the final segmentation generated by PDV-Net. Color coding: almond: LUL, blue: LLL, yellow: RUL, cyan: RML, pink: RLL. (Color figure online)

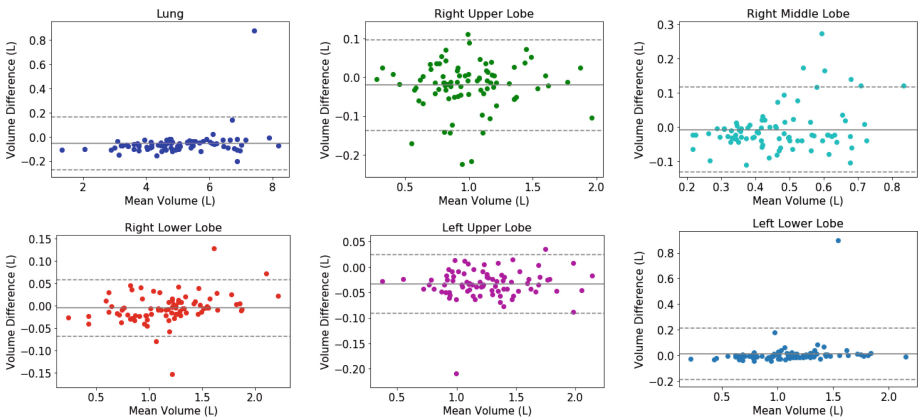


Fig. 4. Bland-Altman plots show the agreement between 3D progressive dense V-net and ground truth.

Pearson correlation showed that all six volume sets in ground truth are strongly correlated with the corresponding six volume sets in the PDV-net segmentation, with $p < 0.001$.

LTRC Results: Table 1 shows that the 3D progressive dense V-net achieves an average Dice score of 0.950 ± 0.007 , significantly improving the dense V-net (0.946 ± 0.008). Once again, the progressive dense V-net model outperformed the

2D U-net model with an average Dice score of 0.929 ± 0.025 . Individual lobes were segmented better in the proposed 3D progressive dense V-net model than in the 3D dense V-net and the 2D U-net models (Table 1). Note that the LTRC dataset includes many pathological cases where the fissure lines are either invisible, distorted, or absent in presence of pathologies such as emphysema, fibrosis, etc. As a result, lobe segmentation becomes more challenging. Nevertheless, our model performed well in segmenting lobes in pathological cases from the LTRC dataset. Moreover, our model outperformed the model of George et al. [4] in segmenting the LTRC cases both in Dice score (0.941 ± 0.255) and inference speed (4–8 minutes per case).

LOLA11 Results: The segmentation results on the LOLA11 cases submitted online were evaluated as overlap (Jaccard) scores. To be consistent with our previous analyses, we converted the Jaccard scores to Dice scores. The results are shown in Table 2. Our method achieved an overall Dice score of 0.934, which is very competitive with the state-of-the-art [2] with a Dice score of 0.938, while outperforming the methods of Giuliani et al. [6] and van Rikxoort et al. [11].

Table 2. Performance evaluation of 3D PDV-Net models on 55 LOLA cases: showing lobe-wise mean Dice scores, standard deviations, median scores, first quartiles, and third quartiles

| Lobe | Mean \pm SD | Q_1 | Median | Q_3 |
|---------|---------------------|--------|--------|--------|
| RUL | 0.9518 ± 0.1750 | 0.9371 | 0.9688 | 0.9881 |
| RML | 0.8621 ± 0.4149 | 0.8107 | 0.9284 | 0.9663 |
| RLL | 0.9581 ± 0.1993 | 0.9621 | 0.9829 | 0.9881 |
| LUL | 0.9551 ± 0.2160 | 0.9644 | 0.9834 | 0.9924 |
| LLL | 0.9342 ± 0.3733 | 0.9546 | 0.9805 | 0.9902 |
| Overall | 0.9345 | | | |
| [6] | 0.9282 | | | |
| [2] | 0.9384 | | | |
| [11] | 0.9195 | | | |

*Jaccard score to Dice score conversion:
 $\text{Dice} = 2 \times \text{Jaccard} / (1 + \text{Jaccard})$

Robustness Analysis: We further investigated the robustness of our model by grouping the 84 LIDC cases in three ways. For the first grouping, the Dice scores were put in three different Z-spacing buckets: Z-spacing ≤ 1 , $1 < \text{Z-spacing} < 2$, and Z-spacing ≥ 2 . In the second grouping, the Dice scores were put in four manufacturer buckets: GE, Philips, Siemens, and Toshiba. In the third grouping, the Dice scores were grouped according to the reconstruction kernel into 3 buckets: soft, lung, and bone. The one-way ANOVA analysis confirmed that there were no significant differences between the average Dice scores of the buckets within each

grouping, suggesting that our model is robust against the choice of reconstruction kernel, size of reconstruction interval, and different CT scan vendors. Moreover, nodule volume in each of the 84 cases does not affect the lobe segmentation performance. There is no correlation between nodule volume and lobe segmentation accuracy, found from Pearson correlation.

We also studied how the segmentation correlation is affected by lung pathologies. For this purpose, we analyzed the correlation between Dice scores and emphysema index (proportion of the lung affected by emphysema) in LTRC cases. According to the Pearson correlation, it was found that lobe segmentation accuracy is not correlated with emphysema index, indicating the robustness of our proposed model in segmenting lobes from pathological cases.

Speed Analysis: The proposed 3D progressive dense V-net model takes approximately 2 s to segment lung lobes from one CT scan using 1 Nvidia Titan XP GPU, which is six times faster than the 2D U-net model. As per our knowledge from the lung lobe segmentation models available in literature, this is by far the fastest model. Note that no prior published research has yet considered a 3D convolutional model for lung lobe segmentation.

5 Conclusions

Automatic and reliable lung lobe segmentation is a challenging task in the presence of chest pathologies and in the absence of visible, complete fissures. In this paper, we introduced a new 3D segmentation approach, namely, progressive dense V-networks for the automatic, fast, and reliable segmentation of lung lobes from chest CT scans, without any prior segmentation. We evaluated our method using 3 test datasets: 84 cases from LIDC, 154 cases from LTRC, and 55 cases from LOLA11. Our results demonstrated that the suggested model outperforms, or at worst performs comparably to, the state-of-the-art while running at an average speed of 2 s per case. Our analyses further demonstrated the robustness of the suggested method against varying configurations of CT reconstruction, choice of CT vendor, and presence of lung pathologies.

References

1. Aziz, A., Ashizawa, K., Nagaoki, K., Hayashi, K.: High resolution CT anatomy of the pulmonary fissures. *J. Thoracic. Imag.* **19**(3), 186–191 (2004)
2. Bragman, F.J., McClelland, J.R., Jacob, J., Hurst, J.R., Hawkes, D.J.: Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior. *IEEE Tran. Med. Imag.* **36**(8), 1650–1663 (2017)
3. Bulatov, Y.: Saving memory using gradient-checkpointing (2018). <https://github.com/openai/gradient-checkpointing>
4. George, K., Harrison, A.P., Jin, D., Xu, Z., Mollura, D.J.: Pathological pulmonary lobe segmentation from CT images using progressive holistically nested neural networks and random walker. In: Cardoso, M.J. (ed.) *DLMIA/ML-CDS -2017*. LNCS, vol. 10553, pp. 195–203. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_23

5. Gibson, E., et al.: Automatic multi-organ segmentation on abdominal CT with dense v-networks. *IEEE Tran. Med. Imag.* (2018)
6. Giuliani, N., Payer, C., Pienn, M., Olschewski, H., Urschler, M.: Pulmonary lobe segmentation in ct image using alpha-expansion. In: *Proceedings of VISIGRAPP*, pp. 387–394 (2018)
7. Harrison, A.P., Xu, Z., George, K., Lu, L., Summers, R.M., Mollura, D.J.: Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images. In: Descoteaux, M., Maier-Hein, L., Franz, A., Janin, P., Collins, D.L., Duchesne, S. (eds.) *MICCAI 2017. LNCS*, vol. 10435, pp. 621–629. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66179-7_71
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
9. LOLA11: Lobe and lung analysis (2011). <http://lola11.com>
10. Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: *Proceedings of 3DV*, pp. 565–571. IEEE (2016)
11. van Rikxoort, E., Prokop, M., de Hoop, B., Viergever, M., Pluim, J., van Ginneken, B.: Automatic segmentation of pulmonary lobes robust against incomplete fissures. *IEEE Tran. Med. Imag.* **29**(6), 1286–1296 (2010)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
13. Ross, J., et al.: Automatic lobe segmentation using particles, thin plate splines, and a maximum a posteriori estimation. *Proc. MICCAI* **6363**, 163–171 (2010)