



Fast and automatic segmentation of pulmonary lobes from chest CT using a progressive dense V-network

Abdullah-Al-Zubaer Imran^{a,b}, Ali Hatamizadeh^{a,b}, Shilpa P. Ananth^b, Xiaowei Ding^{a,b}, Nima Tajbakhsh^b and Demetri Terzopoulos^{a,b}

^aComputer Science Department, University of California, Los Angeles, CA, USA; ^bVoxelCloud, Inc., Los Angeles, CA, USA

ABSTRACT

Automatic, reliable lobe segmentation is crucial to the diagnosis, assessment, and quantification of pulmonary diseases. Existing pulmonary lobe segmentation techniques are prohibitively slow, undesirably rely on prior (airway/vessel) segmentation, and/or require user interactions for optimal results. We introduce a reliable, fast, and fully automated lung lobe segmentation method 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 seconds using a single Nvidia Titan XP GPU. An extensive robustness analysis of our method demonstrates reliable lobe segmentation of both healthy and pathological lungs in CT images acquired by scanners from different vendors, across various CT scan protocols and acquisition parameters.

ARTICLE HISTORY

Received 16 November 2018
Accepted 22 September 2019

KEYWORDS

Lung lobe segmentation; CT; progressive dense V-Net; fissure; 3D CNN

1. Introduction

Human lungs are divided into five lobes. The inner membrane of the lung (visceral pleura) folds towards the centre of the lung and creates double layer fissures that define the five lobes. The lobar boundaries are made of two major (oblique) fissures and a minor (horizontal) fissure. As shown in [Figure 1](#), the left lung has two lobes separated by a major fissure – the upper (superior) lobe and the lower (inferior) lobe. Along with upper and lower lobes, the right lung has a middle lobe; a major fissure separates the upper lobe from the middle lobe and a minor fissure separates the lower lobe from the middle lobe. Each of the five lobes is functionally independent, with its own bronchial and vascular systems.

Automatic segmentation of the lung lobes is important for both clinical and technical purposes. From the clinical perspective, automatic lung lobe segmentation can help radiologists review chest CT scans more efficiently. This is because radiologists often report their pulmonary findings by indicating the affected lung lobe, whose identification requires them to navigate through the nearby slices and search for fissure lines, which are often visually indistinct. Automatic lung lobe segmentation can eliminate the need for such a tedious and time-consuming process. From the technical perspective, accurate lung lobe segmentation can assist 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 (see [Figure 2\(b\)](#)), and assessment and quantification of chronic obstructive pulmonary diseases (COPD) and interstitial lung diseases (ILD), 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. This is shown in [Figure 2\(a\)](#) where the horizontal fissure is incomplete,

unlike the oblique fissures. Several studies in the literature have confirmed the incompleteness of fissures as a very common phenomenon. After reviewing 100 fixed and inflated lung specimens, Raasch et al. (1982) found incomplete right major fissures in 70% of the cases, left major in 46% of the cases, and 94% across the minor fissures. Moreover, the studies of Gulsun et al. (2006) and Aziz et al. (2004) also showed more than 50% incompleteness in pulmonary fissures. Second, the visual characteristics of lobar boundaries change in the presence of pathologies. For example, in [Figure 1](#), the major fissure in the left lung is invisible. The changes could also be related to their thicknesses, locations, and shapes. Third, there also exist other fissures in the lungs that can be misinterpreted as the major and minor fissures that separate the lobes. Examples include accessory fissures (see the sagittal slice in [Figure 2\(c\)](#)) and azygos fissures (see the axial slice in [Figure 2\(d\)](#)).

To address the need for accurate and robust lobe segmentation, we have pursued a fully automatic and reliable deep learning solution based on a Progressive Dense V-Network (PDV-Net) (Imran et al. 2018). Our PDV-Net model inputs an entire CT volume and generates accurate segmentation of the lung lobes in about 2 seconds in only a single forward pass of the network, eliminating the need for any user interaction or any prior segmentation of the lungs, vessels, or airways, which are common assumptions in the design of existing models. Extensive robustness analyses demonstrate that our proposed method performs reliably for CT scans acquired using various imaging protocols from both healthy and pathological patients.

2. Related work

There have been several efforts to segment lung lobes using semi-automatic and automatic techniques. We categorise these approaches into two groups: *reliant approaches*, which rely on

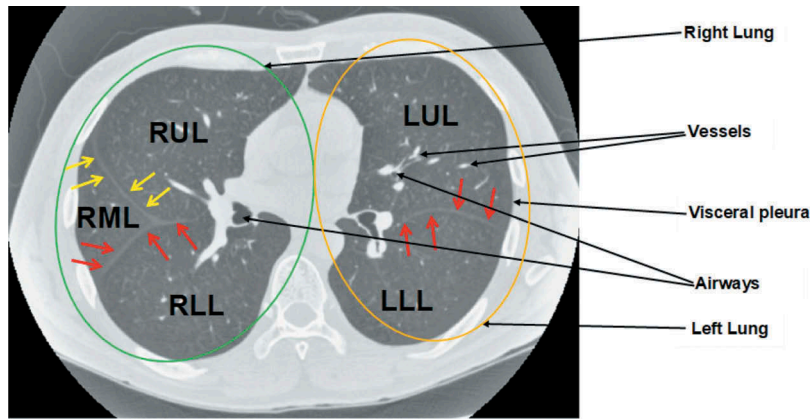
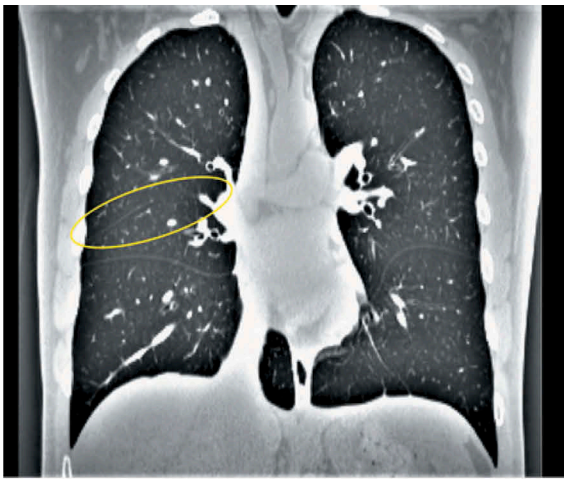
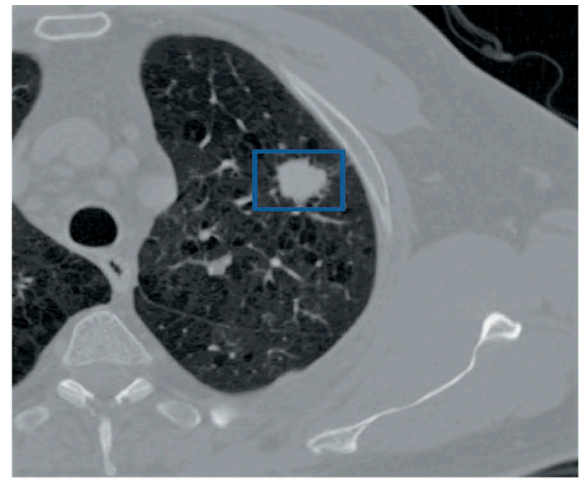


Figure 1. An axial lung CT slice with visible fissures. Two lobes in the left upper lobe (LUL) and left lower lobe (LLL) are defined by a major fissure (indicated by red arrows); Three lobes namely right upper lobe (RUL), right middle lobe (RML), and right lower lobe (RLL) are defined by a major fissure (indicated by red arrows) and a minor fissure (indicated by yellow arrows).



(a)



(b)



(c)



(d)

Figure 2. (a) A coronal slice where the major fissures are complete and visible, but the minor fissure (circled) is incomplete. (b) Nodule shown in the bounding box. (An example nodule report: 5mm nodule found in the left upper lobe.) (c) Accessory fissure (arrows) in a left lung sagittal slice, which looks similar in shape to a minor fissure. (d) Azygos fissure (arrow) in an axial slice creates an extra lobe (azygos lobe) in the right lung.

a prior segmentation or anatomical information, and *non-reliant approaches*, which do not rely on such prior segmentations.

2.1. Reliant approaches

2.1.1. Prior-based segmentation

Reliant approaches require as input a segmentation mask of lungs or lobes (different modalities), airways and vessels, or fissure initialisation. A good example of the latter is the work by Doel et al. (2012), in which lobe segmentation is performed based on an initialisation via fissure detection. In another example of fissure initialisation, Iwano et al. (2013) proposed semi-automatic and automatic lobe segmentation methods based on region-growing. The semi-automatic approach requires major and minor fissure initialisation, whereas for the automatic approach, recognition of lobar bronchi and localisation of fissures are performed prior to the final lobar segmentation. On average, the semi-automatic approach takes approximately 80 seconds and the automatic approach takes approximately 44 seconds per case.

A number of works depend on prior segmentation of airways, vessels, or fissures. The work by Bragman et al. (2017) is a good representative, wherein the suggested method relies on the prior segmentation of airways and vessels. Specifically, a population model of fissure priors was constructed and combined with patient-specific anatomical information for non-parametric surface fitting. Despite the promising results, the model lacks robustness and its reliance on prior knowledge limited the study. In recent work, Giuliani et al. (2018) proposed an approach to segment lobes from an approximate segmentation based on the airway tree. The final lobe segmentation was generated by combining the approximate segmentation with all the lung structures (airways, vessels, lungs, and fissures) segmentation using a multilevel graph cut algorithm. This segmentation method is highly reliant on the quality of the prior airway and vessel segmentations, as well as anatomical knowledge. Lassen and van Rikxoort (2013) proposed a watershed-based lobe segmentation method by combining anatomical information from lungs, fissures, vessels, and bronchi. Despite reporting improved segmentation in the presence of incomplete fissures, the failure of individual prior segmentations limited the performance of the overall segmentation. Based on this work, Lassen-Schmidt et al. (2017) proposed an interactive lobe segmentation method to interactively correct lobe segmentation error through user inputs. However, this improvement was obtained at the price of prolonged segmentation sessions. Lim et al. (2016) performed quantification of emphysema in 66 patients with moderate to severe emphysema who had undergone CT for lung volume reduction planning. They used lobar segmentation from four different prototypes for inter-software variability in lobe-wise emphysema quantifications. Although the lobe segmentation performance is not reported, it is dependent on prior airway and vessel segmentation.

Other works also rely on prior lung or lobe segmentation masks. For example, Bauer et al. (2018) segmented the lung lobes in the expiration phase based on a prior lobe segmentation mask obtained from a CT image acquired in the inspiration phase. An automated lung and lobe segmentation pipeline was

proposed by Blaffert et al. (2010), in which a lung model mesh based on watershed segmentation is adapted to lobar segmentation. Final lobe regions are obtained by adjusting based on overlaid lungs in a post-processing step. However, the authors do not report a quantitative evaluation of lobar segmentation. The model takes 20 seconds to perform lobar segmentation in each CT scan.

2.1.2. Atlas-based segmentation

Another variation of reliant segmentation is registration using mutual information with a previously segmented atlas. The performance of final lobe segmentation is greatly dependent on the performance of the segmentation algorithm used in creating a reference atlas. Among atlas-based approaches for lobe segmentation, Ross et al. (2010) employed the thin-plate spline and a maximum a posteriori estimation method using a manually-defined atlas as a reference. Fissure points were selected based on the atlas and the final lobe segmentation was generated after a post-processing step. Although this method did not rely on any prior airway and vessel segmentation, the execution time was long. Moreover, the creation of the atlas is very cumbersome and prone to poor results in pathological lung cases. By contrast, Pu et al. (2009) performed lobe segmentation by fitting an implicit function to fissures without reliance on prior airway or vessel segmentation. Although they achieved good accuracy for healthy lungs, the performance of their method degraded in the case of lungs with abnormal orientations. Unlike the other atlas-based segmentations, van Rikxoort et al. (2010) made use of multiple atlases for lobe segmentation. Their method showed promise albeit at the expense of slow execution.

2.2. Non-reliant segmentation

Recently, a few convolutional neural-network-based lobe segmentation techniques have been proposed (George et al. 2017; Ferreira et al. 2018; Wang et al. 2018). The segmentation method of George et al. (2017) employs a 2D fully convolutional network followed by a 3D random walker algorithm. This approach does not rely on a prior segmentation of airways or vessels nor on any pre-computed atlases; however, it cannot generate lobe segmentation in a single pass, nor in an end-to-end manner. Furthermore, the 3D random walker algorithm relies on a number of heuristics for the initialisation of seeds and weights. Ferreira et al. (2018) proposed a lobe segmentation model based on a fully regularised V-Net model with deep supervision and carefully chosen regularisation. Although the performance looks impressive, the model was trained with few examples, so it lacks generalisability and may not be effective for varying CT scan cases. A 3D Dense Net-based lobe segmentation method was proposed by Wang et al. (2018). Although they reported good accuracy for pathological lungs, their lobe segmentation method relies on prior lung segmentation and assumes the presence of five lobes, which might not always be the case (e.g. LOLA11 2011)).

Our work (Imran et al. 2018) mitigates the aforementioned limitations – namely, reliance on prior masks, slow runtime, and lack of robustness – through an end-to-end learning network. Without relying on any prior airway/vessel segmentation or anatomical knowledge or atlases, our method performs lobe

segmentation in a single pass of the network. Owing to the full utilisation of the 3D context in our model, the resulting lobe segmentation is smooth and nearly noise-free, which eliminates the need for any subsequent post-processing to fill holes or remove noisy patches from outside the lung area. Our method shows promise for the potential clinical use in quantification of pulmonary diseases and automatic generation of radiological reports.

3. Materials and methods

3.1. Progressive dense V-Net

Combining ideas from dense V-networks (Gibson et al. 2018) and progressive holistically-nested networks (Harrison et al. 2017), we propose a new architecture – the Progressive Dense V-Network (PDV-Net), an end-to-end solution for organ segmentation in 3D volumetric data.

As shown in Figure 3, 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 the 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 normalisation and parametric rectified linear units (PReLU). The outputs of the dense feature blocks are consecutively utilised 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 obtain

the final segmentation from the merged feature maps resulting from the skip2 and skip3 connections. Unlike the dense V-Net, the PDV-Net generates the final output by progressively improving the outputs from the previous pathways.

The PDV-Net is trained using a subset S of a volumetric medical image dataset \mathcal{D} . The training set S contains 3D CT scan images and their corresponding ground truth labels. So, $S = (\mathcal{X}_n, \mathcal{Y}_n)$, for $n = 1, \dots, N$, where the input volumes $\mathcal{X}_n^{(m)} = x_i^{(n)}, i = 1, \dots, |\mathcal{X}|_n$, and the corresponding ground truth labelled volumes $\mathcal{Y}_n^{(m)} = y_i^{(n)}, i = 1, \dots, |\mathcal{Y}|_n, y_i^{(n)} \in \{0 \dots L\}$. Here, $|S|$ is the total number of training examples passed to the network and L is the number of labels provided in the ground truth data through per-voxel labelling (l). To train the PDV-Net, we use a Dice loss function (Milletari et al. 2016) at each level of the progressive network, which directly maximises the similarity between the predicted values and the ground truth over all voxels. This loss properly handles the class imbalance problem prevalent in lung lobe segmentation: lung lobes have different sizes and background regions can be large. We employ a multi-class Dice for the segmentation task:

$$d = \sum_{l=1}^L \frac{\sum_{j=1}^Z p_j^l g_j^l}{\sum_{j=1}^Z (p_j^l)^2 + \sum_{j=1}^Z (g_j^l)^2}, \quad (1)$$

where Z is the total number of voxels, L is the number of classes, p_j^l denotes the predicted probabilities for each class, and g_j^l denotes the corresponding ground truth for each class.

3.2. Data

We used 3 public datasets to evaluate our model:

- (1) We selected a subset of chest CT volumes (354 cases) from the LIDC dataset (Armato et al. 2011) for annotation. To ensure variation in the data, the CT scans were selected such that both challenging and visible fissures are well-represented in the dataset. The lobe segmentation ground truth masks were generated in a semi-automatic fashion by

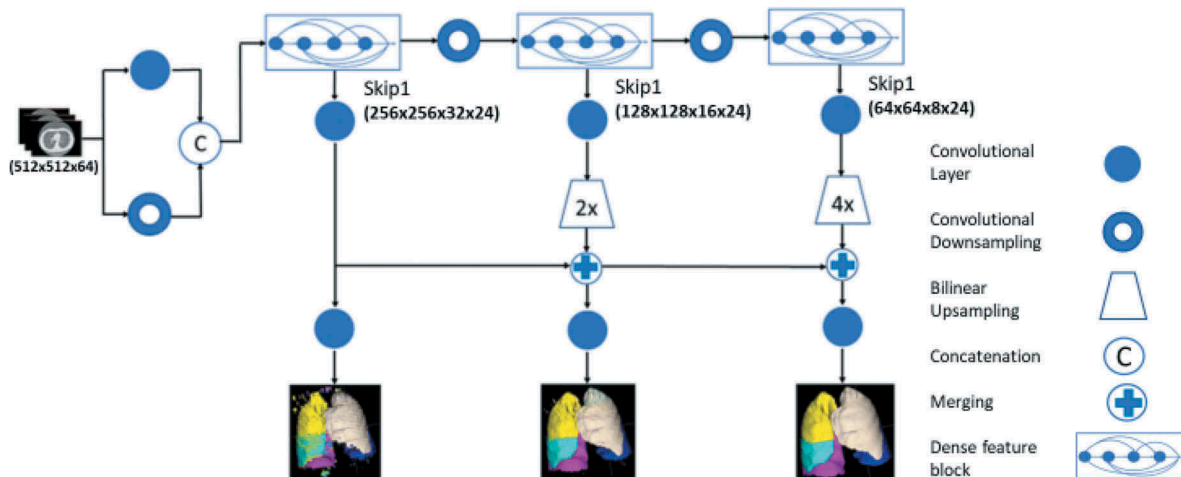


Figure 3. PDV-Net model for the segmentation of lung lobes. Segmentation outputs in different pathways are progressively improved to yield the final result.

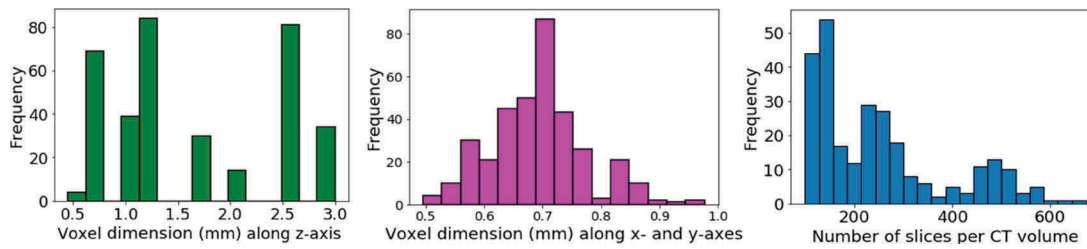


Figure 4. Histograms (from left) of the number of slices per volume; voxel dimensions along the x and y axes; and voxel dimensions along the z axis of lung CT scans in the entire LIDC dataset.

multiple human annotators using the chest imaging platform feature of 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 utilised as the validation set to select values for the hyper-parameters. The CT scans used in the experiment have a variable number of slices with each CT volume containing 100 to 672 slices of size 512×512 pixels. Figure 4 shows the histograms of the number of slices per volume, and of the voxel dimensions which vary between 0.49–0.98 mm, 0.49–0.98 mm, and 0.45–3.00 mm along the x , y , and z axes, respectively. Therefore, the selected CT scans used for pulmonary lobe segmentation not only exhibit varying shapes of fissures and lobes, but also show a variable number of slices and voxel sizes.

- (2) We selected 154 CTs from the LTRC database (Karwoski et al. 2008). 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.
- (3) We used 55 cases of the Lobe and Lung Analysis (LOLA11) challenge (LOLA11 2011) and submitted our results to the challenge organizers for evaluation.

4. Experiments

4.1. Baselines for comparison

For our baseline comparison, we used a U-Net architecture (Ronneberger et al. 2015) and a dense V-Net. The former is used in the most recent published article for lung lobe segmentation (George et al. 2017) and the latter is a strong baseline for comparison, which we are the first to employ for lung lobe segmentation.

4.2. Implementation details

For our PDV-Net and the dense V-Net, the training volumes were first normalised, followed by rescaling to $512 \times 512 \times 64$, using one NVIDIA Titan XP GPU. Due to the large memory footprint of the model, the gradient check-pointing method (Bulatov 2018) was used for memory-efficient back-propagation. Additionally, batch-wise spatial dropout (Gibson et al. 2018) is incorporated for regularisation purposes. The

training was performed on a 64-bit Intel Xeon E5-2697 v4 2.30 GHz CPU system with 256 GB of RAM. We used the Adam optimiser (Kingma and Ba 2014) with a learning rate of 0.01 and a weight decay of 10^{-7} .

For the 2D U-Net model, the implemented architecture is symmetric and consists of four contracting and expanding layers, starting with 16 features in the first layer and doubling the number of features in each step. Each contracting layer consists of two 3×3 convolutions and a ReLU activation followed by a 2×2 max-pooling layer. The expansion path consists of an up-convolution with feature concatenation from the respective contracting layer, and two 3×3 convolutions. In addition, all the ReLU layers are preceded by a batch-normalisation layer. To improve the training process, we also used a generalised Dice score as the loss function, such that the contribution of each class in the image to the gradients is balanced. We trained the network with axial slices from all the training volumes, each sized 512×512 pixels and normalised 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 used the Adam optimiser with a learning rate of 5×10^{-5} and batches of 10 images.

5. Results and discussion

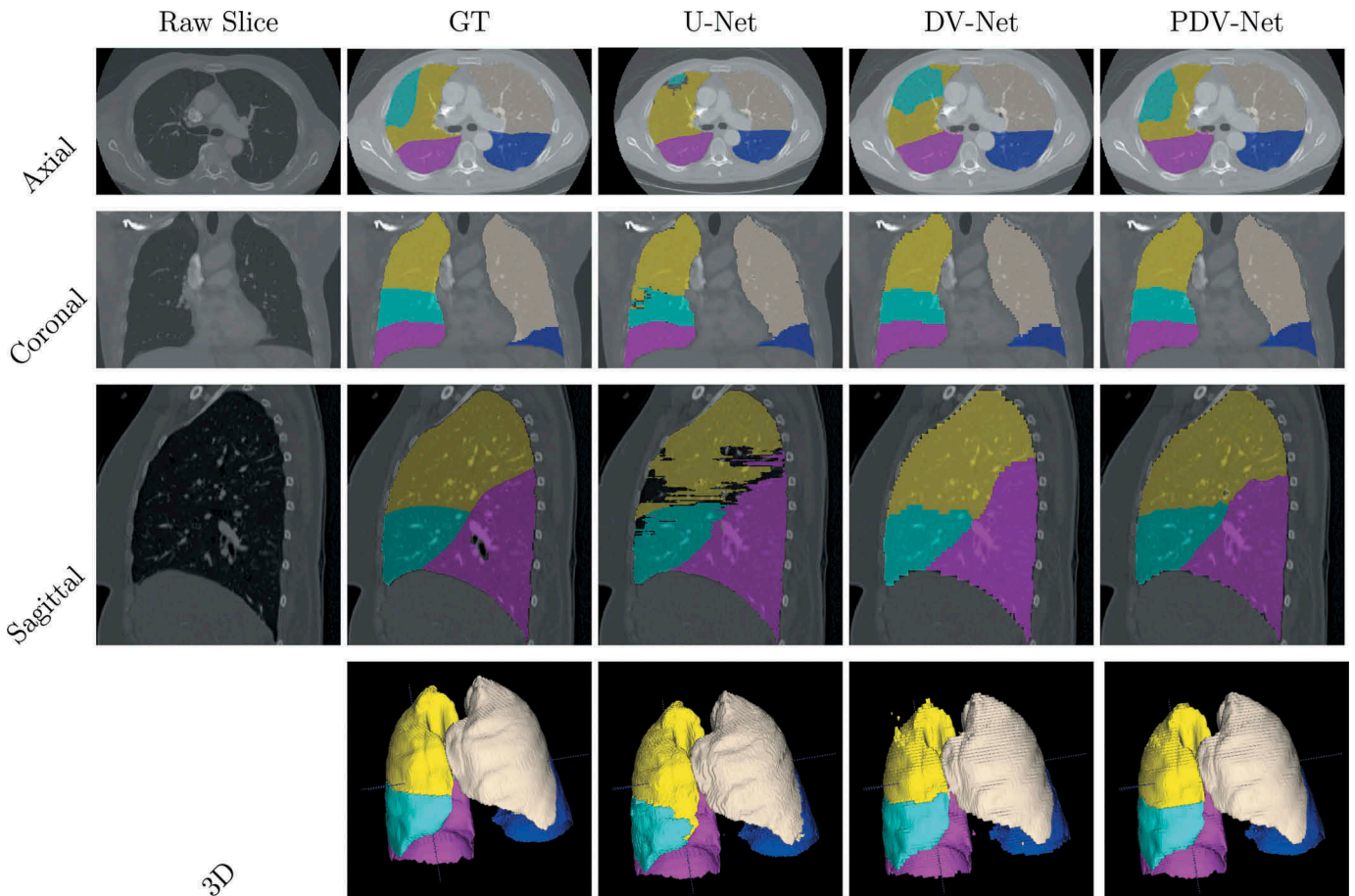
5.1. LIDC results

Table 1 shows the calculated overall and lobe-wise Dice scores and standard deviations for each of the models. Our PDV-Net model, with an overall score of 0.939 ± 0.020 , significantly outperformed the 2D model and yielded consistently larger Dice scores for each of the lung lobes against both the DV-Net and U-Net. Moreover, the lower standard deviation for each lobe indicates that our progressive model is more robust. Figure 5 provides a qualitative comparison between the three models, showing that our PDV-Net model captures lung fissures better than the 2D U-Net and DV-Net. The superiority of our PDV-Net model is evident both in slice (axial, coronal, sagittal) and 3D views.

We further used Bland-Altman plots to measure the agreement between our PDV-Net and ground truth segmentations of the 84 LIDC cases (Figure 6). Good agreement was observed between our segmentation model and ground truth in every plot (Lung and LLL being the two best agreements). 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 $p < 0.001$.

Table 1. Performance comparison of our 3D progressive dense V-Net against 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 of the five lobes are 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

**Figure 5.** Qualitative comparison of PDV-Net's superior performance, both in slice and volume level, against DV-Net and U-Net. Note how noisy patches and rough boundaries are removed from the final segmentation generated by the PDV-Net. Colour coding: almond: LUL, blue: LLL, yellow: RUL, cyan: RML, pink: RLL.

5.2. 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, our 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 by our 3D progressive dense V-Net model than by 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 the 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. (2017) in segmenting the LTRC cases both in Dice score (0.941 ± 0.255) and inference speed (4–8 minutes per case).

5.3. LOLA11 results

Our segmentation results for the LOLA11 cases were evaluated by the organisers of LOLA11. To be consistent with our previous analyses, the Jaccard scores computed by the organisers were converted to Dice scores. The results are shown in Table 2. Our method achieved an overall Dice score of 0.934, which is very competitive to the state-of-the-art reliant method (Bragman et al. 2017) with a Dice score of 0.938, while outperforming the methods of Giuliani et al. (2018) and van Rikxoort et al. (2010).

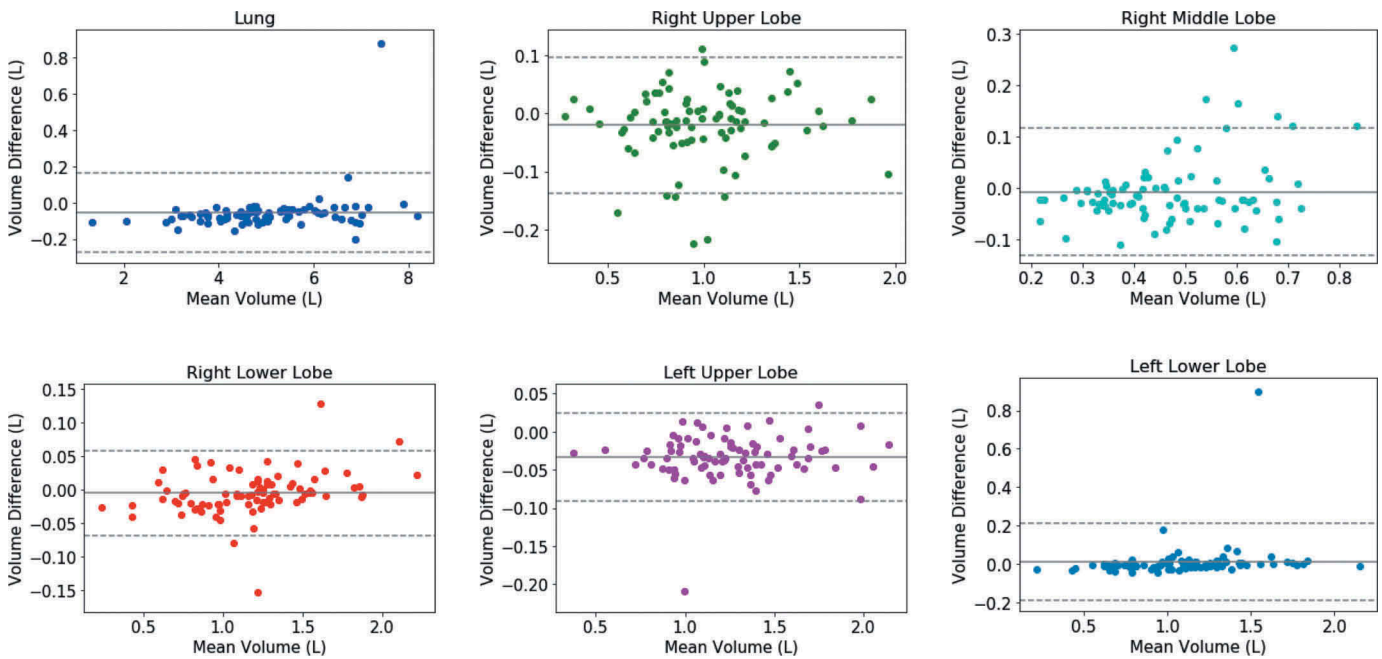


Figure 6. Bland-Altman plots show the agreement between our 3D PDV-Net and ground truth.

Table 2. Performance evaluation of our 3D PDV-Net model 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 \pm 0.1750	0.9371	0.9688	0.9881
RML	0.8621 \pm 0.4149	0.8107	0.9284	0.9663
RLL	0.9581 \pm 0.1993	0.9621	0.9829	0.9881
LUL	0.9551 \pm 0.2160	0.9644	0.9834	0.9924
LLL	0.9342 \pm 0.3733	0.9546	0.9805	0.9902
Overall	0.9345			
(Giuliani et al. 2018)	0.9282			
(Bragman et al. 2017)	0.9384			
(van Rikxoort et al. 2010)	0.9195			

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

Figure 7 shows the segmentation results for the LOLA11 cases. For the left lung in Case 8, the LUL and LLL Dice scores were 0.9940 and 0.9926, respectively. For the right lung in Case 6, the scores are as follows: RUL: 0.9580, RML: 0.9480, and RLL: 0.9869. Again, for the left lung of Case 21, the segmentation Dice scores were relatively low. For the left lung in Case 21, the LUL score was 0.8170 and the LLL score was 0.3035. For the right lung in Case 55, although the right lower lobe was segmented with a high Dice score of 0.9818, because of the invisibility of the horizontal fissure, the RUL and RML had low segmentation Dice scores of 0.6827 and 0.7499, respectively.

5.4. 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: $Z\text{-spacing} \leq 1$, $1 < Z\text{-spacing} < 2$, and $Z\text{-spacing} \geq 2$. 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. A one-way ANOVA analysis confirmed that there were no significant differences ($p\text{-value} < 0.05$) 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 scanner 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, as indicated by the Pearson correlation ($p\text{-value} < 0.05$).

We also studied how the segmentation correlation is affected by lung pathologies, by analysing the correlation between Dice scores and the emphysema index; i.e. the proportion of the lungs affected by emphysema (in the range 0–1). For the LTRC cases, we associated lobe-wise emphysema indices by calculating the proportion of emphysema voxels (voxels marked as emphysema in the LTRC ground truth) in each of the lobes, as well as overall emphysema indices for both lungs. Figure 8 shows plots of the per-lobe and overall emphysema indexes versus segmentation performance. The small Pearson correlation ($p\text{-value} < 0.05$) reveals that the lobe segmentation accuracy is uncorrelated with the emphysema index, confirming the robustness of our model in segmenting lobes in pathological cases.

5.5. Speed analysis

Our 3D PDV-Net model takes approximately 2 seconds to segment lung lobes from one CT scan using a single Nvidia Titan XP GPU, which is six times faster than the 2D U-Net model. To our knowledge from the lung lobe segmentation models reported in literature, ours is by far the fastest model. Note that no prior published research considered a 3D convolutional model for lung lobe segmentation.¹

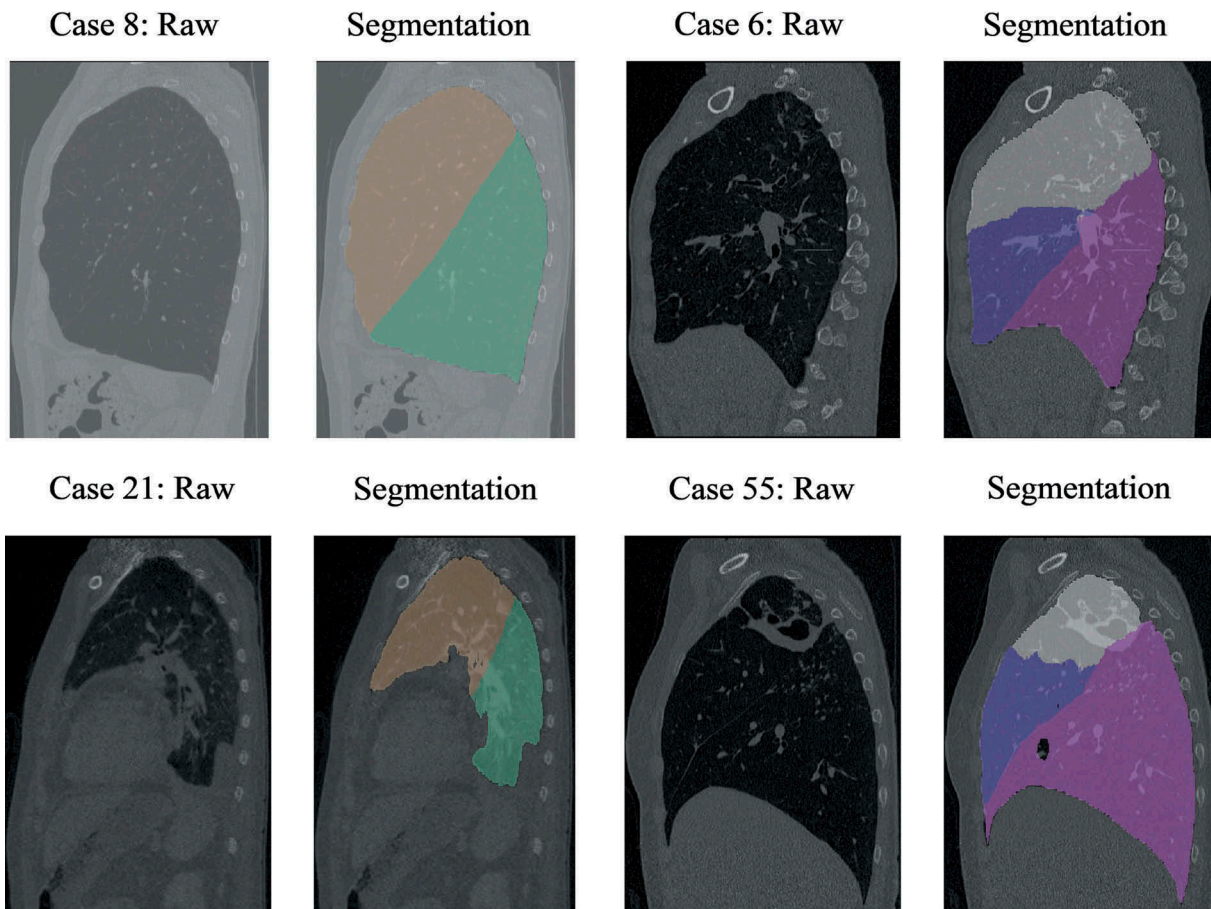


Figure 7. Sagittal plane visualisation of LOLA11 segmentation by our 3D PDV-Net: good cases (upper row) and failure cases (bottom row).

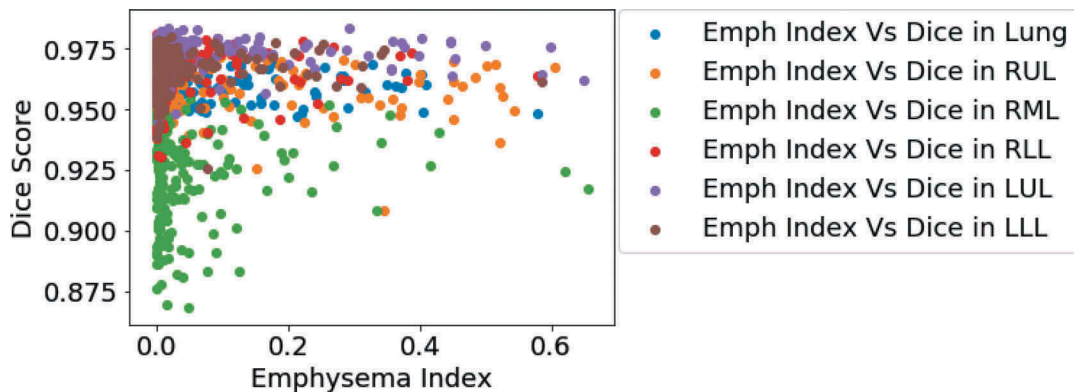


Figure 8. Plots of lobe-wise and overall segmentation accuracy (Dice scores) of our model versus the emphysema indices of the LTRC test cases reveal insignificant correlation.

6. Conclusions

Reliable and automatic lung lobe segmentation is a challenging task, especially in the presence of pathologies and incomplete fissures. We introduced a new 3D CNN-based segmentation technique, namely, Progressive Dense V-Networks (PDV-Nets), and applied it to the automatic, fast, and reliable segmentation of lung lobes from chest CT scans. We evaluated our method using three test datasets – 84 cases from LIDC, 154 cases from LTRC, and 55 cases from LOLA11. Our results demonstrated that our model outperforms, or at worst performs comparably to, the

state-of-the-art while running at an average speed of 2 seconds per case, without requiring any prior segmentation. Furthermore, we demonstrated the robustness of our method against varying configurations of CT reconstruction, choice of CT imaging device vendor, and the presence of lung pathologies.

Note

1. At the time of the original conference paper (Imran et al. 2018) submission, no prior work had used 3D CNN-based models for lung lobe segmentation.

Acknowledgments

The authors are thankful to Gerard Nguyen, MD, of the Washington University School of Medicine, for his extensive support and assistance with the annotation of lung lobes in the LIDC cases. A. Imran and A. Hatamizadeh were supported in part by an unrestricted gift to UCLA from VoxelCloud, Inc.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by VoxelCloud, Inc.

Notes on contributors

Abdullah-Al-Zubaer Imran is a PhD candidate in the Computer Science Department at the University of California, Los Angeles. He received his MS degree in Computer Science from Delaware State University in 2016. His research is focused on machine learning, generative modeling, computer vision, and medical image analysis.

Ali Hatamizadeh is a PhD candidate in the Computer Science Department at the University of California, Los Angeles. He received MS degrees in Mechanical Engineering in 2017 and Computer Science in 2019, both from UCLA. He is the recipient of the 2018 UCLA Henry Samueli School of Engineering and Applied Science Edward K. Rice Outstanding Masters Student Award. His research interests include computer vision, computational medical imaging, and artificial intelligence in medicine.

Shilpa P. Ananth is a Machine Learning Engineer at VoxelCloud, Inc. She received her Master's degree in Biomedical Engineering from Carnegie Mellon University in 2017. Her research interests include deep learning and computer vision for medical imaging. At VoxelCloud, she is currently working on the comprehensive image analysis of Fundus Images.

Xiaowei Ding is Co-Founder and Chief Executive Officer at VoxelCloud, Inc. He is also a Research Assistant Professor of Computer Science at the University of California, Los Angeles. He received his PhD degree in Computer Science from UCLA in 2015. He has published papers on cardiopulmonary image analysis and deep learning applied to various medical imaging modalities.

Nima Tajbakhsh is the Lead Scientist at VoxelCloud, Inc. He received his PhD degree in Biomedical Informatics at Arizona State University in 2015. His research interests lie in the development of innovative deep learning solutions for medical image analysis from various imaging modalities, particularly with minimal annotations. His work has led to more than 40 peer-reviewed publications and 10 US patents.

Demetri Terzopoulos is a Chancellor's Professor of Computer Science at the University of California, Los Angeles, where he holds the rank of Distinguished Professor and directs the UCLA Computer Graphics & Vision Laboratory. He is also Co-Founder and Chief Science Officer of VoxelCloud, Inc. He is a fellow of the ACM, IEEE, Royal Society of Canada, and Royal Society of London. He received his PhD degree in Artificial Intelligence from MIT in 1984. He has (co-)authored more than 400 published research papers and several volumes, primarily in computer graphics, computer vision, medical imaging, computer-aided design, and artificial intelligence/life.

References

Armato SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, et al. 2011 May. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys*. 38(2):915–931. 5.

Aziz A, Ashizawa K, Nagaoki K, Hayashi K. 2004 Feb. High resolution CT anatomy of the pulmonary fissures. *J Thoracic Imag*. 19 (3):186–191. 2.

Bauer C, Eberlein M, Beichel R. 2018 Mar. Pulmonary lobe separation in expiration chest CT scans based on subject-specific priors derived from inspiration scans. *J Med Imag*. 5(1):014003. 3.

Blaffert T, Wiemker R, Barschdorf H, Kabus S, Klinder T, Lorenz C, Schadewaldt N, Dharaiya E. 2010 Mar. A completely automated processing pipeline for lung and lung lobe segmentation and its application to the LIDC-IDRI data base. In: Dawant BM, Haynor DR, editors. *Medical imaging 2010: image processing*. Vol. 7623. International Society for Optics and Photonics; p. 762347.

Bragman FJ, McClelland JR, Jacob J, Hurst JR, Hawkes DJ. 2017. Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior. *IEEE Tran Med Imag*. 36(8):1650–1663.

Bulatov Y. 2018 Jun. Saving memory using gradient-checkpointing [accessed 2018 Mar 1]. <https://github.com/openai/gradient-checkpointing>.

Doel T, Matin T, Gleeson F, Gavaghan D, Grau V. 2012 Mar. Pulmonary lobe segmentation from CT images using fissureness, airways, vessels, and multilevel b-splines. *Proc ISBI*. 9:1491–1494.

Ferreira FT, Sousa P, Galdran A, Sousa MR, Campilho A. 2018. End-to-end supervised lung lobe segmentation. 2018 International Joint Conference on Neural Networks (IJCNN); Rio de Janeiro, Brazil. IEEE. p. 1–8.

George K, Harrison A, Jin D, Xu Z, Mollura D. 2017. Pathological pulmonary lobe segmentation from CT images using progressive holistically nested neural networks and random walker. *DLMIA/ML-CDS 2017, LNCS 10553*. Cham: Springer; p. 195–203.

Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, Davidson B, Pereira SP, Clarkson MJ, Barratt DC. 2018. Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE Tran Med Imag* 37(8):18221834.

Giuliani N, Payer C, Pienn M, Olschewski H, Urschler M. 2018. Pulmonary lobe segmentation in CT image using alpha-expansion. *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) – Volume 4: VISAPP*, p. 387–394.

Gulsun M, Ariyurek O, Comert R, Karabulut N. 2006. Variability of the pulmonary oblique fissures presented by high-resolution computed tomography. *Surg Radiologic Anat*. 28(3):293–299.

Harrison AP, Xu Z, George K, Lu L, Summers RM, Mollura DJ. 2017. Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images. *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Sep 10. p. 621–629.

Imran A, Hatamizadeh A, Ananth SP, Ding X, Terzopoulos D, Tajbakhsh N. 2018. Automatic segmentation of pulmonary lobes using a progressive dense V-network. In: Stoyanov D, et al. editors. *Proceedings of the fourth MICCAI international workshop on deep learning in medical image analysis (DLMIA 18)*, Granada, Spain, September 2018. In *deep learning in medical image analysis and multimodal learning for clinical decision support (DLMIA 2018, ML-CDS 2018)*, Lecture Notes in Computer Science. Vol. 11045. Berlin: Springer-verlag, Berlin; p. 282–290.

Iwano S, Kitano M, Matsuo K, Kawakami K, Koike W, Kishimoto M, Inoue T, Li Y, Naganawa S. 2013. Pulmonary lobar volumetry using novel volumetric computer-aided diagnosis and computed tomography. *Interact Cardiovasc Thorac Surg*. 17(1):59–65.

Karwoski RA, Bartholmai B, Zavaletta VA, Holmes D, Robb RA. 2008. Processing of CT images for analysis of diffuse lung disease in the lung tissue research consortium. In: Hu XP, Clough AV, editors. *Medical imaging 2008: physiology, function, and structure from medical images*. Vol. 6916. International Society for Optics and Photonics; p. 691614.

Kingma DP, Ba J. 2014. Adam: a method for stochastic optimization. *arXiv:1412.6980*.

Lassen B, van Rikxoort E. 2013. Automatic segmentation of the pulmonary lobes from chest CT scans based on fissures, vessels, and bronchi. *IEEE Tran Med Imag*. 32(2):210–222.

Lassen-Schmidt B, Kuhnigk J, Konrad O, Van Ginneken B, Van Rikxoort E. 2017. Fast interactive segmentation of the pulmonary lobes from thoracic computed tomography data. *Phys Med Biol*. 62(16):6649.

- Lim HJ, Weinheimer O, Wielpütz MO, Dinkel J, Hielscher T, Gompelmann D, Kauczor HU, Heussel CP. 2016. Fully automated pulmonary lobar segmentation: influence of different prototype software programs onto quantitative evaluation of chronic obstructive lung disease. *PLoS One*. 11(3):e0151498.
- LOLA11: Lobe and lung analysis. 2011. [accessed 2018 Mar 1]. <http://lola11.com>.
- Milletari F, Navab N, Ahmadi S. 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. *Proceedings of 3DV*; Stanford, CA. IEEE. p. 565–571.
- Pu J, Zheng B, Leader J, Fuhrman C, Knollmann F, Klym A, Gur D. 2009. Pulmonary lobe segmentation in CT examinations using implicit surface fitting. *IEEE Tran Med Imag*. 28(12):1986–1996.
- Raasch B, Carsky E, Lane E, OCallaghan J, Heitzman E. 1982. Radiographic anatomy of the interlobar fissures: a study of 100 specimens. *Am J Roentgenol*. 138(6):1043–1049.
- Ronneberger O, Fischer P, Brox T. 2015. U-net: convolutional networks for biomedical image segmentation. *Proceedings of MICCAI*; Munich, Germany. Springer. p. 234–241.
- Ross J, Estpar RSJ, Kindlmann G, Daz A, Westin C, Silverman E, Washko G. 2010. Automatic lobe segmentation using particles, thin plate splines, and a maximum a posteriori estimation. *Proc MICCAI*. 6363:163–171.
- van Rikxoort E, Prokop M, de Hoop B, Viergever M, Pluim J, van Ginneken B. 2010. Automatic segmentation of pulmonary lobes robust against incomplete fissures. *IEEE Tran Med Imag*. 29(6):1286–1296.
- Wang X, Teng P, Lo P, Banola A, Kim G, Abtin F, Goldin J, Brown M. 2018. High throughput lung and lobar segmentation by 2D and 3D CNN on chest CT with diffuse lung disease. In: Stoyanov D, et al., editors. *Image analysis for moving organ, breast, and thoracic images. RAMBO 2018, BIA 2018, TIA 2018*. *Lecture Notes in Computer Science*. Vol. 11040. Cham: Springer; p. 202–214.