LEARNING TO SEGMENT THE LUNG VOLUME FROM CT SCANS BASED ON SEMI-AUTOMATIC GROUND-TRUTH

Patrick Sousa[⋆] *Adrian Galdran*[⋆] *Pedro Costa*[⋆] *Aurélio Campilho*[⋆,†]

[⋆] INESC TEC Porto, Portugal
[†]Faculdade de Engenharia da Universidade do Porto, Portugal

ABSTRACT

Lung volume segmentation is a key step in the design of Computer-Aided Diagnosis systems for automated lung pathology analysis. However, isolating the lung from CT volumes can be a challenging process due to considerable deformations and the potential presence of pathologies. Convolutional Neural Networks (CNN) are effective tools for modeling the spatial relationship between lung voxels. Unfortunately, they typically require large quantities of annotated data, and manually delineating the lung from volumetric CT scans can be a cumbersome process. We propose to train a 3D CNN to solve this task based on semi-automatically generated annotations. For this, we introduce an extension of the well-known V-Net architecture that can handle higher-dimensional input data. Even if the training set labels are noisy and contain errors, our experiments show that it is possible to learn to accurately segment the lung relying on them. Numerical comparisons on an external test set containing lung segmentations provided by a medical expert demonstrate that the proposed model generalizes well to new data, reaching an average 98.7% Dice coefficient. The proposed approach results in a superior performance with respect to the standard V-Net model, particularly on the lung boundary.

***Index Terms*—** Lung Volume Segmentation, CT scans

1. INTRODUCTION

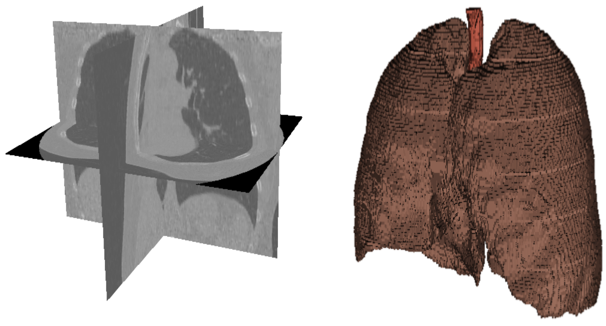
In Computer-aided diagnosis of pulmonary diseases, lung volume segmentation is a key preliminary pre-processing stage intended to isolate the lung from the background. Accurate lung segmentation allows to avoid processing irrelevant information and enables false positive removal, thereby preventing potentially incorrect diagnosis.

Automated methods for lung segmentation have been developed along the years, especially on Computer Tomography (CT) images. Most of them are threshold [1] or region-based [2], relying on intensity levels, contrast and neighborhood homogeneity. More sophisticated methods are based on prior anatomical knowledge, like atlas-based methods, which rely on the registration of the target image to a template containing labels of the thoracic region [3]. Neighboring anatomy-guided methods use spatial information about the surrounding organs to delineate lung regions, simplifying the segmentation task when abnormalities or artifacts are present. Hybrid approaches combining fast traditional threshold-based techniques with more sophisticated multi-atlas methods have also been proposed [4]. In this case, a segmentation obtained using conventional approaches is examined for errors, and corrected by means of more time-consuming atlas-based methods.

Deep Learning techniques has also been proposed for segmenting organs from CT scans. Dou *et al.* [5] proposed a 3D deeply supervised model based on a Fully-Convolutional Neural Network (F-CNN) to automatically segment the liver on CT images. For segmenting multiple organs in CT scans, Roth *et al.* [6] adapted an existent architecture called 3D U-Net [7]. For lung segmentation in CT scans, Harrison *et al.* [8] introduced a deep architecture termed Holistically-Nested Network. This model was particularly accurate at finely delineating lung borders. A progressive multi-path scheme was also implemented in order to deal with output ambiguity and coarsening resolution, resulting in an extended method called Progressive Holistically-Nested Network.

Deep neural networks are known to require large quantities of annotated data. Unfortunately, for the problem of lung segmentation, few public data sources exists. However, semi-automatic segmentations of the lung in CT scans can be easily generated. In the LUng Nodule Analysis 2016 (LUNA16) challenge [9], such ground-truth was provided based on CT scans from the Lung Image Database Consortium and Image Database Resource Initiative. Lung segmentations were generated by a semi-automatic method [4], resulting in reasonably accurate annotations, see Figure 1. However, these annotations were not perfect nor validated by a medical doctor, and hence not usable for clinical evaluation purposes. Noisy ground-truth and pseudo-labels have recently proven useful for training deep learning-based segmentation models on brain MRI images achieving good performance when evaluated with clinically corrected ground-truth [10].

In this paper we propose two main contributions. First, we show that semi-automatically generated (and thus imperfect) volumetric lung segmentations can be employed for training a deep neural network, resulting in great performance when evaluated in expert-labeled data. Second, we introduce



(a) (b)

Fig. 1: (a) A volumetric lung CT scan from the LUNA16 dataset [9] (b) Automatically generated lung segmentation.

a methodological modification to a popular 3D deep architecture in order to handle input of high spatial resolution without losing the ability to capture fine details at lung borders.

2. METHODOLOGY

2.1. Model Architecture

The U-Net architecture [11] is one of the most popular techniques for medical image segmentation. Its architecture is an extension of F-CNNs consisting of a downsampling followed by an upsampling path. In the downsampling path the input traverses several layers of convolutional blocks that reduce the spatial resolution of the output volumes. Along the upsampling path, skip connections add back information from higher scales of the image present in the encoder part of the model. Several three-dimensional versions of the U-Net, *e.g.* 3-D U-Net [7] or V-Net [12]. In both cases, the spatial resolution of the input images remained limited ($245 \times 244 \times 56$ for the former, $128 \times 128 \times 64$ for the latter) due to memory constraints. A simple solution is resizing the spatial dimensions of the data to fit the input of the architecture, but this strategy can lead to a critical loss of relevant information. An alternative approach consists of dividing the data into volumetric input patches. The output of the model is then a spatial reconstruction of several outputs, corresponding to the initial division of the input volume. With this procedure, small details are not lost, but certain contextual information is missed at the boundaries of the divided sub-volumes.

We introduce a novel strategy to deal with large data dimensionality while avoiding information loss. For this, we modify the original V-Net architecture, introducing a max-pooling layer early in the model in order to reduce its dimensionality from $512\times512\times256$ to the conventional $128\times128\times64$ of V-Net. To mitigate information loss, we introduce a skip connection between the input of our architecture and the last convolutional layer. We also introduce another relevant modification to the original V-Net design: we reduce the number of filters in our model to $2/3$ of the ones used in the original architecture. We experimentally verified that the resulting model is less prone to overfitting, and still produces highly accurate predictions. Finally, ReLU non-linearities were replaced by PReLu activation functions that with their adaptive learning coefficient for negative values have shown better results [13], and batch normalization was also incorporated. An overall diagram of the proposed architectural design is represented in Fig. 2.

2.1.1. Loss Function

The loss function minimized during training the proposed model is a modified version on the Dice coefficient, which can measure overlap volume between three-dimensional objects:

$$xD(P,G) = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}, \tag{1}$$

where N is the number of voxels on each image, $p_i \in P$ is a voxel p_i within the predicted segmentation P, and $g_i \in G$ is the binary ground-truth.

2.2. Semi-Automatically Generated Ground-Truth and Training

The data used for training the above model consisted of 888 CT scans, accompanied by volumetric lung segmentations for each scan. Such segmentations were generated automatically, and may contain certain amount of errors. Accordingly, these annotations should not be used as a reference in any segmentation study. It is important to stress that these segmentations are not used for testing our algorithm, but only to train the model. The main hypothesis we aim to verify is that a deep CNN can be trained on such imperfect noisy ground-truth and still learn useful representations. The model is thus trained to generate lung segmentations that are to be validated in test time with a separate dataset of manually delineated lung volumes, see Section 3.1.

For training, the available data was randomly divided into 700 scans and 188 for validation. The scans had a fixed spatial resolution of 512 and a slice thickness ranging between 0.6mm to 2.5mm. Depth resolution was mapped to a common value of 256 voxels. Since in CT scans most relevant information lies in the Houndsfield units range of $[-1000; 400]$, information outside this range was omitted. Standard data augmentation techniques (spatial shifting, zooming along the depth axis) were applied to increase the training data.

The model was trained with standard backpropagation for 12 epochs using the Adam Optimizer [14] and a learning rate of $1e^{-3}$. The loss defined in eq. (1) was monitored in the validation set, and training was early-stopped when no improvement was observed for a pre-determined number of epochs.

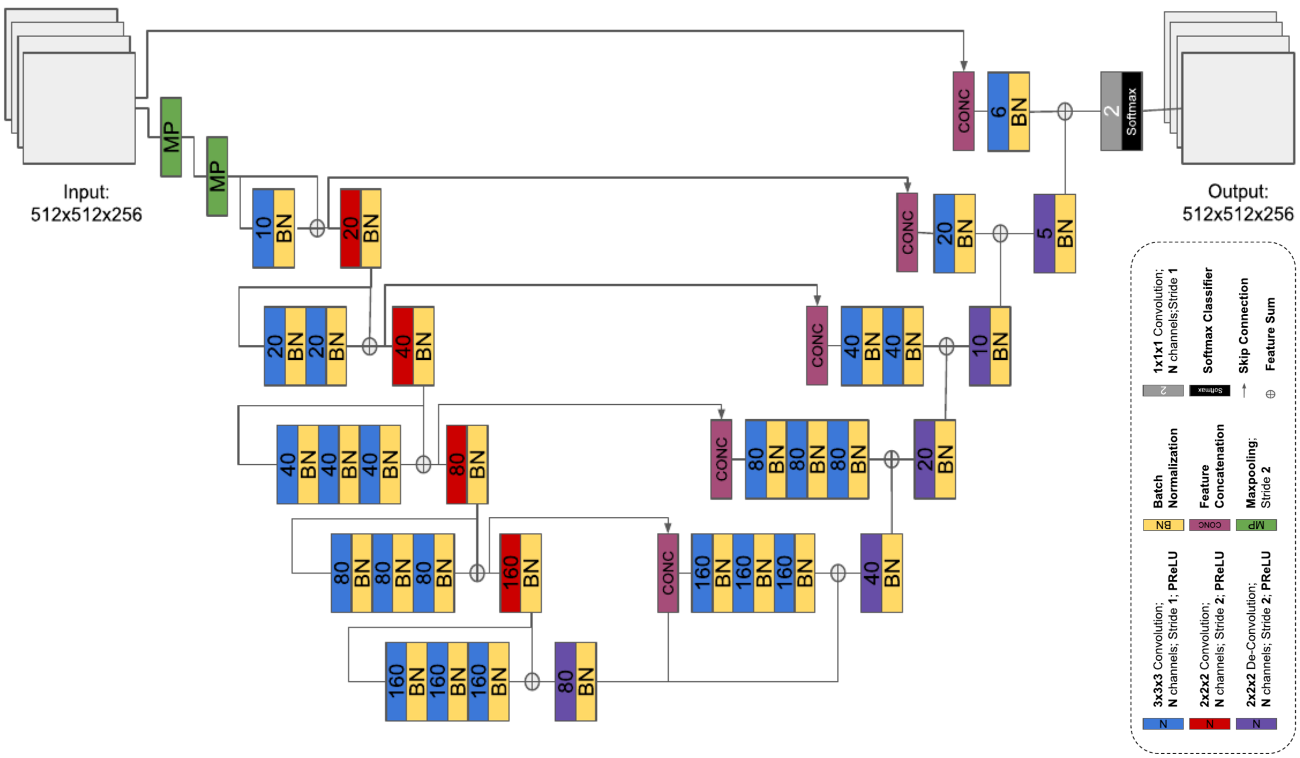


Fig. 2: The proposed model, based on the V-Net, with the addition of an extra initial layer and skip connections allowing to deal with $512 \times 512 \times 256$ size input CT scans.

3. EXPERIMENTAL SETTING

3.1. Test Set

The proposed method was evaluated with data from the VESSEL12 Challenge consisting 20 healthy and pathological chest CT scans of size 512×512, with a variable depth resolution of a maximum spacing of 1mm. The lung volume ground-truth data was acquired and validated by an expert radiologist. It contains labels for different lung lobes intended to train lobe segmentation algorithms [15] [16]. Since we were only interested in the overall lung region, we merge the annotations from all lobes into a single label.

3.2. Experimental Evaluation

For performance evaluation, two different metrics were considered. First the DICE coefficient was computed to assess the degree of overlap between expert ground-truth and predictions of our model, as compared with the standard V-Net. However, comparing overlap between relatively large objects with the DICE score may be slightly misleading, since errors present in borders have a low impact. As such, for a more fair experimental evaluation, the predicted segmentations were also assessed by the Average Symmetric Surface Distance (ASD). ASD is a surface distance metric that measures the average distance of all the points of the surface of the 3D segmentation with their closest points in the surface associated to the ground-truth:

$$ASD = \frac{\sum_{x \in B_{seg}} d(x, B_{gt}) + \sum_{x \in B_{gt}} d(y, B_{seg})}{|B_{seg}| + |B_{gt}|}.$$

4. RESULTS AND DISCUSSION

Figure 3 b) shows an example of a prediction generated by our method, which in this case produces a fine lung volume segmentation. For comparison purposes, we trained a standard V-Net on the same dataset. The model received as input the downsampled CT scan to $128 \times 128 \times 64$. In Figure 3 c) we present a prediction generated by the V-Net of the same CT scan shown. A more detailed visual comparison is provided in Figure 3 d), e) and f). As can be observed, the volume in general is well predicted by both models, although the approach introduced in this paper achieves a more finely delineated boundary, as opposed to the stair-casing effects along the lung borders produced by V-Net. Nevertheless, both results are relatively satisfactory, validating our hypothesis that a deep 3D CNN can be effectively trained on the kind of noisy ground-truth used in this work.

Dice coefficients and ASM scores on the entire available test set are reported in Table 1. Both methods achieve satisfactory Dice scores, which demonstrates that both models

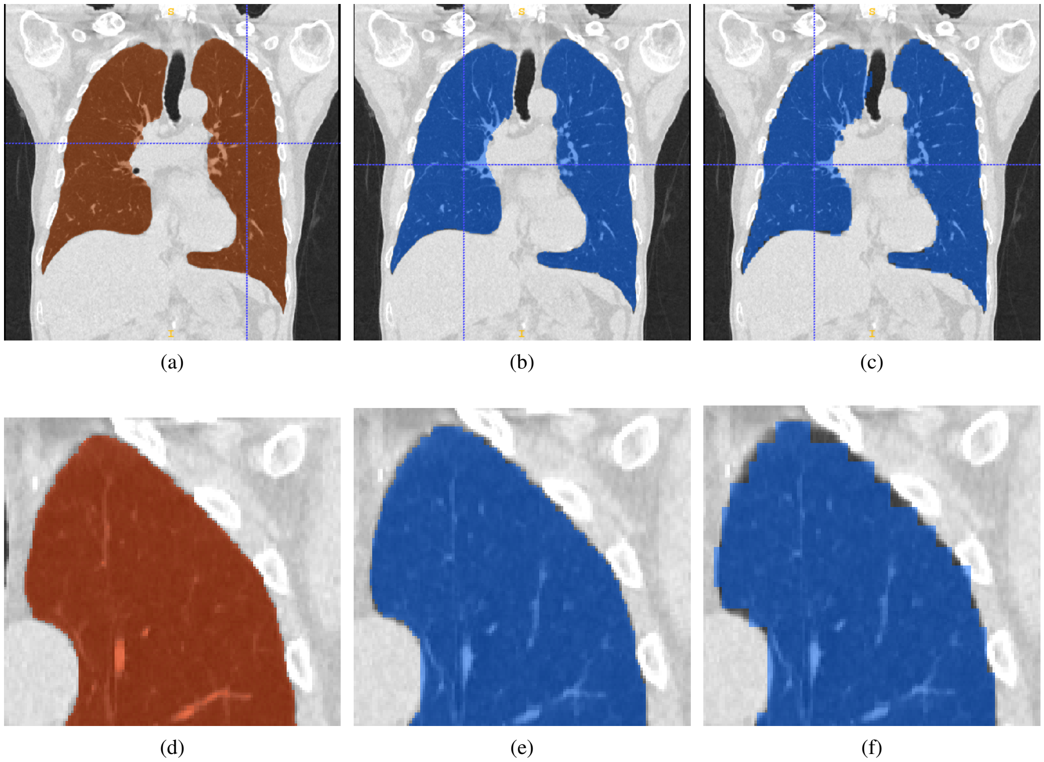


Fig. 3: Coronal view view of a CT scan from our test set. a), d): expert-labeled ground-truth. b), e): Segmentation produced by the proposed model. c), f) Result generated by a standard V-Net model.

Table 1: Comparison of the Dice Coefficient and Average Symmetric Surface Distance (ASD) of the results from the proposed model and V-Net.

	Dice Coefficient (%)	ASD (mm)
V-Net	97.2	2.627
Ours	**98.7**	**0.576**

learned to properly segment the lung volume, even when trained on imperfect ground-truth. Furthermore, the proposed model achieves a slightly better Dice coefficient than the standard V-Net. This is confirmed by the ASD values obtained by each model. The ASD achieved by the proposed extension to V-Net seems to be better capable of handling lung surface voxels, resulting in better segmented boundaries.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have demonstrated that modern 3D segmentation methods based on Deep CNNs can be effectively trained on imperfect automatically generated ground-truth for the task of lung volume segmentation from CT scans. In addition, we introduce an extension of the well-known V-Net architecture that can handle better surface voxels inside the lung. The proposed model can accepts scans of a $512 \times 512 \times 256$ resolution, thereby avoiding any initial information loss, and properly dealing with memory constraints. The proposed model produces highly accurate lung volume segmentations when validated in an external test set containing ground-truth provided by a medical expert, achieving a Dice Coefficient of 98.7% and an Average Surface Distance of 0.576mm. which are superior to results produced by a standard V-Net.

In future work, modification to the loss function driving the optimization process will be explored, in order to dedicate more attention to boundary errors. Another interesting research direction is the potential extension of the segmentation method to other pulmonary regions for which manual ground-truth is hard to acquire, based on automatically generated segmentations that may be used for training such models.

Acknowledgments

Work funded by North Portugal Regional Operational Programme (NORTE 2020), PORTUGAL 2020 Partnership Agreement, and European Regional Development Fund (ERDF), within the project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016".

6. REFERENCES

[1] Samuel G. Armato III and William F. Sensakovic, "Automated lung segmentation for thoracic ct: Impact on computer-aided diagnosis," *Academic Radiology*, vol. 11, no. 9, pp. 1011 – 1021, 2004.

[2] Jiahui Wang, Feng Li, and Qiang Li, "Automated segmentation of lungs with severe interstitial lung disease in ct," vol. 36, pp. 4592–9, 10 2009.

[3] Baojun Li, Gary E. Christensen, Eric A. Hoffman, Geoffrey McLennan, and Joseph M. Reinhardt, "Establishing a normative atlas of the human lung: Intersubject warping and registration of volumetric ct images," *Academic Radiology*, vol. 10, no. 3, pp. 255 – 265, 2003.

[4] Eva van Rikxoort, Bartjan de Hoop, Max A Viergever, Mathias Prokop, and Bram Ginneken, "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection," vol. 36, pp. 2934–47, 08 2009.

[5] Qi Dou, Hao Chen, Yueming Jin, Lequan Yu, Jing Qin, and Pheng-Ann Heng, "3d deeply supervised network for automatic liver segmentation from ct volumes," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Cham, 2016, pp. 149–157, Springer International Publishing.

[6] Holger R. Roth, Hirohisa Oda, Yuichiro Hayashi, Masahiro Oda, Natsuki Shimizu, Michitaka Fujiwara, Kazunari Misawa, and Kensaku Mori, "Hierarchical 3d fully convolutional networks for multi-organ segmentation," *CoRR*, vol. abs/1704.06382, 2017.

[7] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *MICCAI*, 2016.

[8] Adam P. Harrison, Ziyue Xu, Kevin George, Le Lu, Ronald M. Summers, and Daniel J. Mollura, "Progressive and Multi-path Holistically Nested Neural Networks for Pathological Lung Segmentation from CT Images," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, Cham, 2017, pp. 621–629, Springer International Publishing.

[9] Colin Jacobs, Arnaud Arindra, Adiyoso Setio, Alberto Traverso, and Bram van Ginneken, "Lung nodule analysis (luna16)," 2016, https://luna16.grand-challenge.org/home/.

[10] Mohsen Ghafoorian, Jonas Teuwen, Rashindra Manniesing, Frank-Erik de Leeuw, Bram van Ginneken, Nico Karssemeijer, and Bram Platel, "Student beats the teacher: Deep neural networks for lateral ventricles segmentation in brain MR," *Proc.SPIE*, vol. 10574, pp. 10574 – 10574 – 6, 2018.

[11] "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015, pp. 234–241, Springer International Publishing.

[12] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, 2016.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 1026–1034.

[14] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[15] Rina Dewi Rudyanto, Rina Dewi Rudyanto, Sjoerd Kerkstra, and Bram van Ginneken, "Vessel segmentation in the lung 2012 (vessel12)," 2012, https://vessel12.grand-challenge.org/home/.

[16] F. T. Ferreira, P. Sousa, A. Galdran, M. R. Sousa and A. Campilho, "End-to-end supervised lung lobe segmentation," *International Joint Conference on Neural Networks(IJCNN), Rio de Janeiro(Brazil)*, Jul/2018.