End-to-End Supervised Lung Lobe Segmentation

Filipe T. Ferreira*, Patrick Sousa*, Adrian Galdran*, Marta R.Sousa† and Aurélio Campilho*‡
*INESC TEC, Porto, Portugal
†Centro Hospitalar de Entre o Douro e Vouga, E.P.E., Santa Maria da Feira, Portugal
‡Faculdade de Engenharia da Universidade do Porto - FEUP, Porto, Portugal

***Abstract*—The segmentation and characterization of the lung lobes are important tasks for Computer Aided Diagnosis (CAD) systems related to pulmonary disease. The detection of the fissures that divide the lung lobes is non-trivial when using classical methods that rely on anatomical information like the localization of the airways and vessels. This work presents a fully automatic and supervised approach to the problem of the segmentation of the five pulmonary lobes from a chest Computer Tomography (CT) scan using a Fully Regularized V-Net (FRV-Net), a 3D Fully Convolutional Neural Network trained end-to-end. Our network was trained and tested in a custom dataset that we make publicly available. It can correctly separate the lobes even in cases when the fissure is not well delineated, achieving 0.93 in per-lobe Dice Coefficient and 0.85 in the inter-lobar Dice Coefficient in the test set. Both quantitative and qualitative results show that the proposed method can learn to produce correct lobe segmentations even when trained on a reduced dataset.**

***Index Terms*—Lung Segmentation, Lobe Segmentation, 3D Segmentation, Deep Learning**

I. Introduction

Segmentation of the lung anatomical structures is an important task of Computer Assisted Diagnosis (CAD) systems based on Chest Computer Tomography (CT) scans. Information about localization, volume or shape of these structures is necessary to complete other diagnostic tasks, provide a precise quantification of the extent and heterogeneity of pulmonary diseases and for treatment planning.

The lungs are composed of five lobes (two in the left lung and three in the right lung) separated by the lobar fissures, represented in Fig. 1. Lobe segmentation can be a trivial task when fissures are clearly delineated in the CT scan. However, this is often not the case due to fissure incompleteness, the presence of other structures and lung parenchymal abnormalities surrounding them.

Since fissures appear at the boundary between two adjacent lobes, most methods for lobe segmentation have, as the initial step, a fissure detection procedure. Searching for this type of structure can result in many false positives due to the presence of structures with strong resemblance inside the lungs. As such, many methods rely on anatomical information, such as airway or vessel segmentation, bronchial tree or even pre-existing atlases [5].

Following this kind of approach, Bargman et al. [2] applied a probabilistic method based on the model of the fissures. Using a two-class Gaussian Mixture Model with prior anatomical information, a non-parametric surface fitting was performed

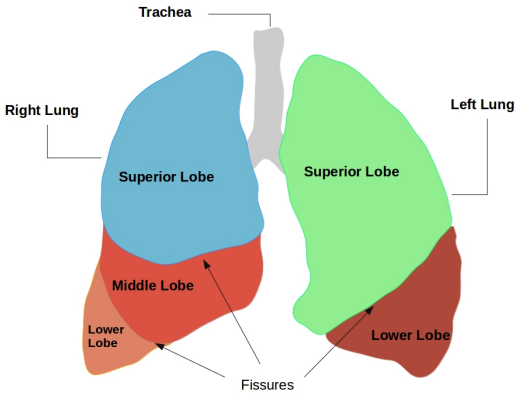


Fig. 1: Representation of the lungs with the respective lobes and fissures.

to obtain a final segmentation. The algorithm achieved a relatively high performance in terms of DICE and F1 scores, but the dependence on a correct modelization of fissures adds a substantial complexity to the method.

Over the past years, supervised deep learning methods and, more precisely Convolutional Neural Networks (CNNs), have become the methodology of choice in the medical image domain, being applied to a wide range of CAD-related tasks, e.g detection, classification and segmentation of structures and organs [15]. For both 2D and 3D scenarios, CNNs present state-of-the-art results in challenging segmentation tasks. Ronneberger et al.[17], proposed a method for segmentation of microscopy images using a fully-convolutional neural network (F-CNN) called U-Net, which has become the most popular CNN architecture for segmentation of 2D medical images. For 3D medical images, 2D CNN based methods are usually applied slice-per-slice to the volumetric data. For example, Zhou et al. [24] used a 2D F-CNN trained on the slices of an abdominal CT scan to perform 3D segmentation of the pancreas. 3D CNN based methods have been recently proposed to exploit the 3D spatial information and integrate context for a better volume segmentation. For instance, Yu et al. [4] proposed a Volumetric ConvNet with mixed residual connections for prostate segmentation from 3D MR images. Some 3D architectures have also been developed based on the

978-1-5090-6014-6/18/$31.00 ©2018 IEEE

U-Net. An example that is particularly relevant for the method proposed in this paper is V-Net, first proposed by Milletari et al. [16] to segment images of the pancreas. Qi Dou et al. [6], developed a 3D Deeply Supervised Network (3D DSN) for segmentation of the liver from 3D CT scans and heart and large vessels from 3D MR images, achieving competitive state-of-the-art results with a substantially improved speed.

For the segmentation of lungs in CT Scans, Harrison et al. [10], developed a 2D method using a 2D Progressive Holistically-Nested Network (P-HNN) slice-per-slice. For the segmentation of the lobes of the lungs, based on the previous approach, George et al. [7] employed the same P-HNN algorithm to identify potential lobar boundaries. After the identification, a Random Walker algorithm, seeded and weighted by the P-HNN output, generates the final segmentation.

The lack of annotated data is a common burden in medical images. To overcome this difficulty and also to avoid overfitting the training data, regularization techniques can be used. Regularization techniques have been described as modifications made to learning algorithms in order to reduce their generalization error but not their training error [9]. Multi-Task Learning [3], Deep Supervision [14], Batch Normalization [12] and Dropout [21] are some examples of these techniques.

The main contribution presented in this work consists of a novel approach for the segmentation of lungs and their lobes from CT scans. In contrast with previous approaches, which typically process 2D slices sequentially, our technique can directly receive and process three-dimensional data. Our proposed method is a 3D Fully Convolutional Neural Network, based on the V-Net architecture, with the addition of carefully selected advanced regularization techniques. The resulting model, named Fully Regularized V-Net (FRV-NET), is shown to be effective for producing highly accurate three-dimensional segmentations of the lobes, without relying on heavy pre-processing and post-processing schemes, and without needing large quantities of data to be trained. The performance of the method is thoroughly analyzed on two different datasets, evaluating the resulting segmentations by means of a new inter-lobar overlap-based measuring metric. We also provide a rigorous ablation study, where we individually disable each regularization technique one at a time, to analyze their influence on the overall performance of the model.

II. Method

In this section we present the architecture of our system. The regularizing techniques applied are described as well as the loss function of the model and its implementation. The code and data to replicate the experiments are publicly available. [1]

A. Model Architecture

The architecture explored in this study is based on the V-Net [16], a 3D extension of the U-net [17], which is widely used in biomedical image because of its capability to solve segmentation problems relying on small sets of training data.

[1]https://github.com/filipetrocadoferreira/end2endlobesegmentation

Using a similar architecture as V-Net, our model (Fig. 2) receives as input images with size $128 \times 128 \times 64$ and has as main output a voxel-wise prediction for the six target classes, the five lobes plus the background. The architecture is constituted by an encoding path followed by a decoding path. The encoding part follows a typical CNN architecture where convolutional layers iteratively decrease the feature resolution while the number of channels is increased in the same order. To achieve per-voxel prediction in the same resolution as the input image, it is necessary to sequentially upsample the feature maps. This takes place in the decoding path. Skip (or residual) connections are then employed between encoding and decoding paths, concatenating features of the same resolution. This way small details in the image, lost during downsampling, are recovered.

We employ only $3 \times 3 \times 3$ convolutions in the whole model instead of bigger kernels due to its efficiency. It is possible to achieve bigger receptive fields using these filters sequentially with a reduction of the computing time. The activation function is the Parametric Rectified Linear Unit (PReLU) [11]. We choose this function to alleviate the problem of vanishing gradient [8] during the training of the network. PReLUs allow a learned parametric gradient even when its input is negative in opposition to traditional Rectified Linear Unit (ReLU). To reduce the feature maps in the downsampling path, in opposition to usual Pooling layers, we perform strided $2 \times 2 \times 2$ convolutions, which are known to lead to a greater accuracy [20]. In the upsample path we use 3D upsample layers with dimension $2 \times 2 \times 2$ to increase the size of the feature space.

The last layer is a $1 \times 1 \times 1$ convolution followed by a soft-max activation function that produces the probability of each voxel to be classified as one each of the six classes of interest.

B. Regularization techniques

Deep neural networks can easily overfit in small training sets, resulting in poor generalization. In order to avoid the problem of overfitting we extend the original V-Net architecture with additional regularizing techniques described in the following paragraphs.

1) Batch Normalization: Batch Normalization [12] deals with the change of the feature space distribution along the model during the training, also known as the *internal covariate shift*. It addresses the problem, by normalizing layer's input and keeping its mean close to 0 and standard deviation approximately equal to 1.

This step can act as a regularizer, but also speeds up training, allows higher learning rates and reduces the dependence on weights initialization.

2) Dropout: One of the simplest yet most effective regularization method for deep neural networks is Dropout [21]. It consists of the random disabling of neurons during training with probability p. Ignoring temporarily some activations forces the other neurons to learn a more robust representation of the input data while training and leads to a reduction of the sensitivity of specific neurons. At test time, dropout is disabled

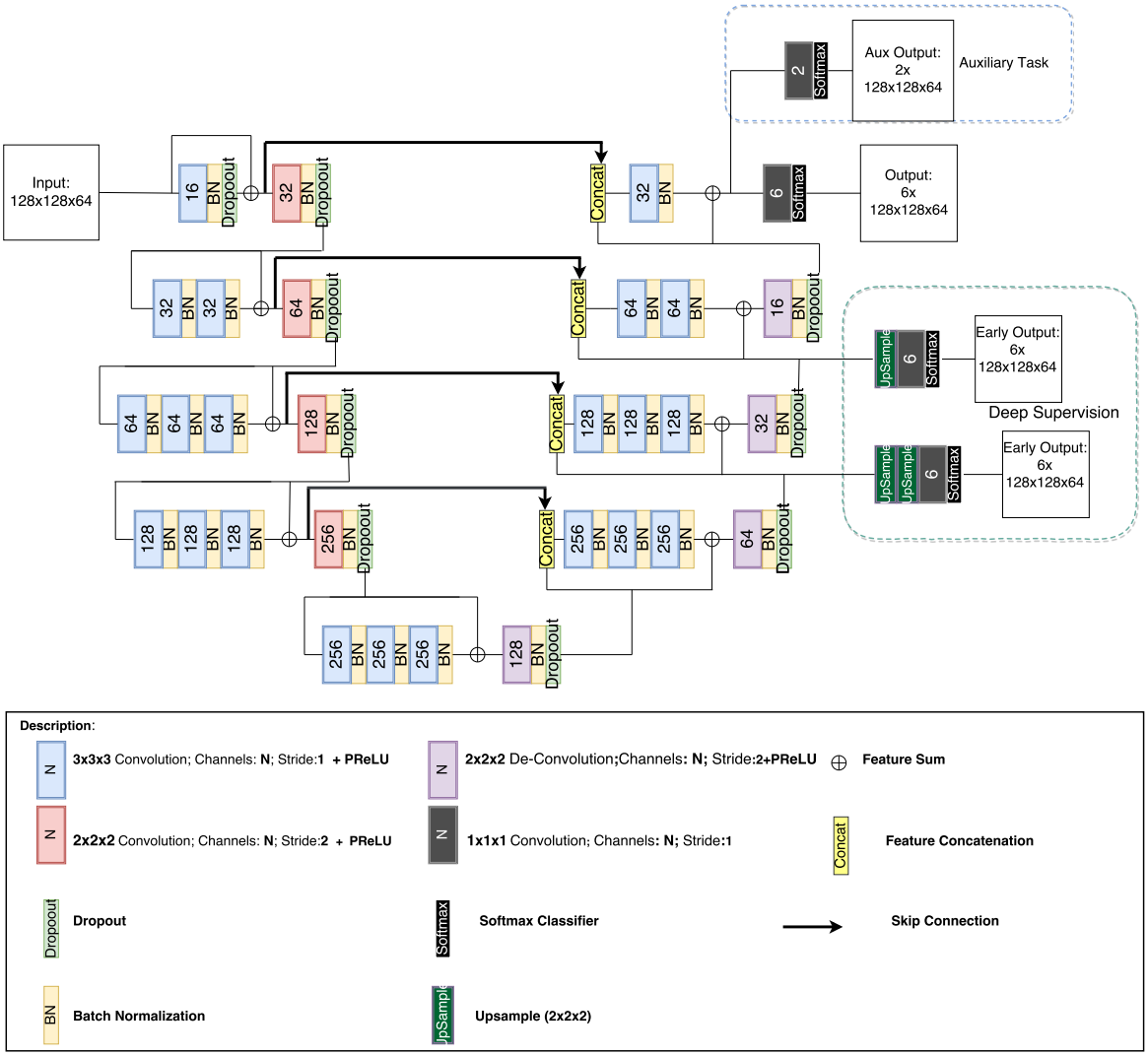


Fig. 2: Schematic of the proposed model based on the V-Net[16]. Like the V-Net the model is formed by an encoding path followed by a decoding path receiving input images with size 128 x 128 x 64. It contains regularization techniques such as Batch Normalization, Deep Supervision, Multi-task learning and Dropout represented in the figure. Please note that all the calculation are performed in 3 dimensions; 2D icons are just used for schematics. Best seen in an electronic version

and the weights are scaled by a factor of p to compensate for the increased number of active neurons [21].

We apply Dropout after the activation layer in every down-sampling and upsampling phases, both in the encoder-decoder networks.

3) Deep Supervision: Normally, the supervision of the network is performed in the output of the model with the labels of the dataset. In deep networks and networks with small training datasets, due to the loss in the representation capability in the first layers of the model, the norm of the gradients can fast decrease to zero during training. This low norm of the gradient affects the back-propagation turning the training phase slower and leading to the vanishing gradients problem, which compromises convergence. Deep supervision pays attention to the hidden layers of the network adding cost functions on those layers. It is considered as a regularization technique suitable to overcome the vanishing gradient phenomenon [14].

In our model, we adopt Deep supervision on the latest two scales of the upsampling path, where the ouput is upsampled by nearest neighbor interpolation to the ouput size of the main model. At test time the part of the model used in deep supervision is simply ignored.

4) Multi-Task Learning: Multi-Task Learning is a method consisting of using the same network core to solve multiple tasks simultaneously. Training a model in different but related tasks has shown to lead to better performance than training a model for each task separately [3]. Moreover, Multi-Task Learning allows better generalization on the main task using the shared representation of auxiliary tasks, acting this way as regularization in the training procedure [18]. As aforementioned, our main goal is to predict the correct probability of

each voxel of belonging to each one of the lobes. However, this task is harder in volume regions surrounded by elements of other classes, in our case, near to other lobes and background. In lobar segmentation these areas usually correspond to the fissures and lung walls. Therefore we introduce an auxiliary loss function, which will be described in the next section in order to penalize wrong segmentations of lobe borders. This supplementary objective is used to focus the attention [18] of the model in the most difficult scenarios, such as lungs not well formed or with pathologies, improving its representation power and the ability to separate the lobes.

C. Loss

Inspired on the evaluation method of the LObe and Lung Analysis 2011 (LOLA11) challenge [23] we apply the Dice coefficient to each Lobe (per-lobe) to train the network. In the challenge, the loss employed was the mean per-lobe Dice coefficient. However, due to difference of lobe sizes and consequently of class frequency we exploit a weighted average per-lobe Dice for the loss function of the main-task, also proposed in [22].

Let $\mathcal{L}(P, G)$ be the loss function based on the weighted average of the per-lobe Dice coefficient. G represents the ground-truth segmentation distributed in N voxels r_{n_c}. The class of each voxel is represented through one-hot encoded, where each voxel is represented by a binary vector with size of the number of the classes C.

We define P as the output of the $softmax$ layer composed by a vector with the probability of each voxel p_{n_c} to belong to each one of the C target classes. Accordingly, we can represent the loss as:

$$\mathcal{L}(P, G) = -2 \cdot \frac{\sum_c w_c \sum_n p_{n_c} r_{n_c} + \delta}{\sum_c w_c \sum_n (p_{n_c} + r_{n_c}) + \delta} \tag{1}$$

being w_c the inverse frequency of each target class in the training batch, $c \in \{1, \ldots, C\}$, and $w_c = \sum_n \frac{1}{r_{n_c} + \delta}$ with δ being a small number to avoid the zero-division.

For the auxiliary task mentioned in subsection II-B4 we use the Dice coefficient in the loss function $\mathcal{L}_{aux}$ between the prediction P_{aux} and the ground-truth of the lobe borders G_{aux}:

$$\mathcal{L}_{aux}(P_{aux}, G_{aux}) = -2 \cdot \frac{\sum_n p_{n_a} r_{n_a} + \delta}{\sum_n (p_{n_a} + r_{n_a}) + \delta} \tag{2}$$

In order to perform deep supervision, the loss function utilized in the E early predictions of the network P_e is the same as in the main task, see equation (1) with $e \in \{1, 2, \ldots, E\}$. So we write the total loss of the deep supervision path as:

$$\mathcal{L}_E(P_E, G) = \frac{\sum_E -2 \cdot \frac{\sum_c w_c \sum_n p_{e_{n_c}} r_{n_c} + \delta}{\sum_c w_c \sum_n (p_{e_{n_c}} + r_{n_c}) + \delta}}{E} \tag{3}$$

It must be noted that the loss of each one of the E deeply supervised paths will just optimize the part of the network upstream. Here we refer to the sum of each component due to ease of notation.

Finally, we can state the loss function employed during the training of the entire network as:

$$\mathcal{L}_{total} = \lambda_1 \cdot \mathcal{L}(P, G) + \lambda_2 \cdot \mathcal{L}_{aux}(P_{aux}, G_{aux}) + \lambda_3 \cdot \mathcal{L}_E(P_E, G) \tag{4}$$

where the $\lambda_{1,2,3}$ are the weights of each component in the final loss.

D. Implementation

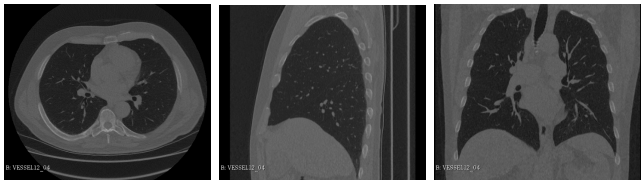
For the implementation and training of the model described above, the following steps were performed:

1) Data Preparation: To fit the input size of the architecture, we would need to resize the $512 \times 512 \times 256$ scan dimension to $128 \times 128 \times 64$. However, this drastic reduction may result in loosing important details on lobe borders, that are thin surface volumes. In order to avoid this, we opted by resizing the scans to $256 \times 256 \times 128$ and, from this volume randomly sample in $128 \times 128 \times 64$ patches. With this approach we are able to decrease memory requirements while effectively keeping the resolution needed to extract good representations of the lung lobes. The values of the scans were clipped between $[-1000; 400]$ Hounsfield units since, within this range, all the relevant information is preserved. Additionally, the images are normalized to zero mean and unit variance. To increase efficiently the training data, the original dataset was augmented by means of linear transformations. Small random rotations around the Z-axis, voxel translations in the X and Y axis and zoomed in/out operations around the Z-axis were applied. Note that in our implementation, no mirroring was applied for data augmentation to keep the relative position of the lobes static.

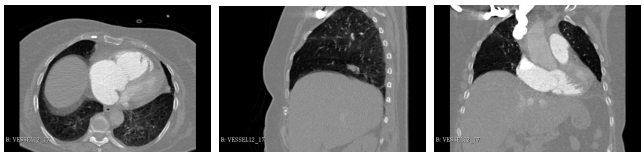
2) Post-Processing: The only post-processing operation is applied at inference phase, to recover the original shape of the scan after the downsampling and division in patches. During inference, division of patches uses a predefined stride of 25% of the patch dimension. The reconstruction of the entire scan from a set of patches holding voxel-wise predictions is using the mean of the overlapped patches. Finally, the prediction is upsampled by a nearest neighbor interpolation to match the shape of the original input scan.

3) Training: The model is trained with standard backpropagation and mini-batch gradient descent for 8000 epochs. The employed optimizer is Adam [13]. On each epoch, the model sees 200 random volumetric patches. On each batch, single patches are due to memory constraints. The initial learning rate is set to 10^{-4} up to the first 4000 epochs, and it is reduced to 10^{-5} for the remaining epochs. Dropout is applied with a probability of 50% and the weights for the losses in the Equation 4 are $\lambda_{1..3} = [0.375, 0.375, 0.25]$.

Each training experiment took approximately 50 hours in the workstation described in the next paragraph. The inference time was less than 45 seconds per scan taking in consideration all the pre and post-processing.



(a) Healthy lung



(b) Pathological lung

Fig. 3: Axial, Sagittal and Coronal CT scan views of an Healthy and a Pathological Lung from the VESSEL12 dataset.

4) System Settings: The proposed method was implemented in Python 2.7 using Keras (v.:2.0.4) framework with Tensorflow (v.:1.1.0) backend. The workstation has a CPU: Intel® Core™ i7-6700K, 16 Gb of RAM and a GPU: NVIDIA 1080.

III. Experiments and Results

A. Data

We evaluated our method with data from two different public sources. From VESSEL12 Challenge [19] all the 20 scans were used, 14 for training and the remaining for testing. This dataset contains 3D CT scans with a maximum slice spacing of $1mm$ and size of $512 \times 512 \times 256$ and it includes both healthy and pathological lungs (Fig. 3).

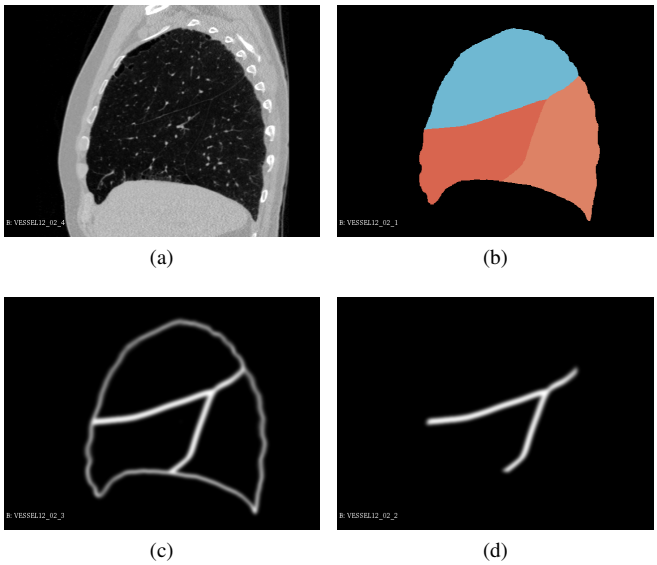
The other source of data is the LIDC-IDRI [1] database, from where we extracted randomly 5 scans for testing. This dataset was built for nodule detection and for this reason it is relatively easy in the task of lobe segmentation since the lungs are structurally healthy.

Due to the lack of labeled lobe segmentations, for each one of the 25 scans, ground-truth data was obtained by a radiologist, who manually delineated the lung lobes using the Chest Imaging Platform [2] in 3D Slicer environment[3]. The annotation included a pre-segmentation of the lungs by 3D Slicer and then the user was asked to mark fissure points on each one of the three fissures. Those points were then interpolated generating the labeled fissure surfaces, which split the lungs in the corresponding five lobes (**Fig. 4 (a) and (b)**).

The lung lobe borders and fissure maps used in the auxiliary task of the network were automatically obtained from manual lobe segmentations. The lobes were extracted after subtracting a morphological erosion with a unitary square kernel to the original lobe map. Finally, the border maps were calculated applying a Gaussian filter on the resulting binary maps to smooth the results (**Fig. 4 (c)**).

[2]https://chestimagingplatform.org/about

[3]https://www.slicer.org/



(a) (b) (c) (d)

Fig. 4: Examples of the data employed in training: **(a)** original CT scan; **(b)** ground-truth for lobe segmentation; **(c)** lobe borders used as auxiliary task during training; **(d)**: fissure maps adopted in evaluation

B. Experiments

In order to evaluate prediction of the model, our evaluation metrics were based on Dice coefficient overlap metric (equation 5), with P being the result of the segmentation and GT the ground truth. We calculated the average of the per-lobe Dice Coefficient (*pl-DC*), similarly to the LOLA11 challenge [23]. However, since lobes are much larger than its borders, the borders will not have a great impact in the dice coefficient, and so dice coefficient will not be sufficient to assess the capacity to separate the lobes. So we proposed inter-lobar Dice Coefficient (*il-DC*) that is based on the overlap of the fissure maps of the predicted and ground-truth lobe as shown in **Fig. 4 (d)**.

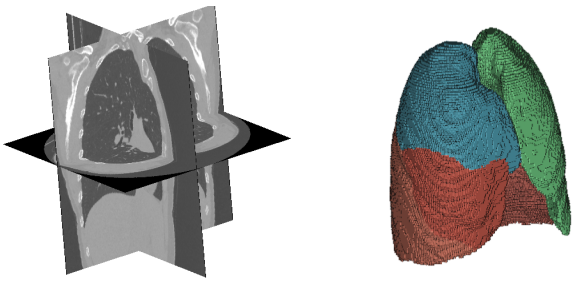
$$\mathcal{DC}(P, GT) = 2 \cdot \frac{P \cdot GT}{P + GT} \tag{5}$$

In order to assess the influence of each one of the proposed regularization methods, we removed each regularization technique alternately and trained the model again with the same hyper-parameters. We also compared the network output without any of the proposed techniques to assess the performance of the joint regularization scheme.

C. Results

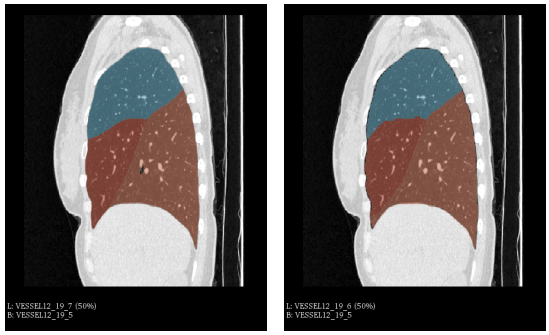
In this section we provide both qualitative visual results of the predictions produced by the proposed model and quantitative assessment of its performance for the task of lobe and fissure segmentation.

1) Qualitative Results: Our system receives a CT scan input and yields a 3D segmentation of the lungs and its lobes. In the

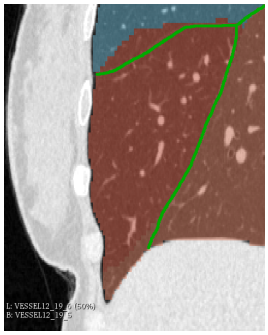


(a) (b)

Fig. 5: Example of a segmentation with the scan **(a)** as input and the resultant 3D volume **(b)**



(a) (b)



(c)

Fig. 6: Comparison of the Ground-truth with the Prediction of the proposed method: (**a**) Ground-truth; (**b**) Prediction; (**c**) Representation of part of the predicted Segmentation with the highlight of the true fissure map in green.

3D segmentation, an average pl-DC of 93.6% and an il-DC Coefficient of 76.2% were achieved.

Examples of segmentations are shown in Fig. 6 and 7. In Fig. 6, the predicted (a) and the ground-truth (b) lobe segmentation of the lung are presented. Fig. 6(c) shows a zoomed-in representation of Fig. 6(b) with the overlay of the ground-truth fissure map in green. In Fig. 7, it is presented the prediction of lobe segmentation with the ground-truth of the fissure highlighted in black.

Fig. 8 presents a comparison between the predictions produced by our method and the three variations that produced

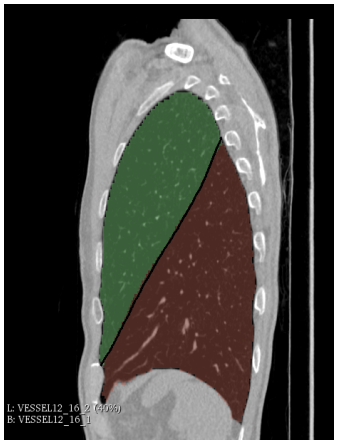


Fig. 7: Prediction of the lobe segmentation, highlight of the true fissure map being the black line

the best results as seen in Table I. The results are presented in a 2D sagittal view for ease of visualization.

It is possible to verify that even without the fissure well delineated, the network is able to correctly separate the lobes. However the model has some difficulty to segment the lung in the presence of large vessels or airways leaving some fragments classified as background. In the case of the model trained without deep supervision, these fragments are bigger. Deep supervision seems to improve contextual information although the quantitative results are just marginally better.

The model trained without multi-task learning seemed to have more difficulty to deal with the lung and lobe borders, leaving some misclassified lobe parts outside the lung region. This was probably caused by the sampling procedure, which provided a weaker contextual representation of the lobes.

As previously stated, results yield without Dropout are very similar, or in the case of the last slice, even better than the proposed method.

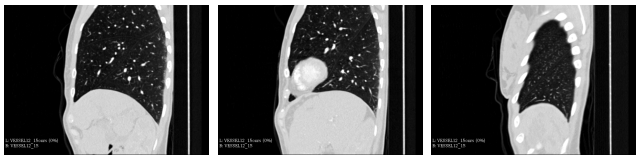
In Fig. 6 and 7, it is possible to observe that the proposed method learned roughly to separate the lobes. This can be verified by producing the fissure maps for both ground-truth and predicted segmentations. In the presented case, the difference is minimal even with the fissures in the scan being difficult to distinguish visually. We can deduce that the model learns fissure segmentation but also infers it from the context learned from the remaining lung structure.

2) Quantitative Performance Evaluation: Table I presents the performance of the proposed method and the variations of the method to show the influence of each one of the regularization techniques applied to the model. As aforementioned we used two different evaluation metrics, the per-lobe Dice Coefficient (pl-DC) and the inter-lobar Dice Coefficient (il-DC). For the first metric, we also show the Dice Coefficient for each one of the lobes and then, finally, the corresponding average and standard deviation.

It can be clearly appreciated that the regularization techniques were fundamental in these experiments. The FRV-Net achieved the best results. Furthermore, the ablation studies

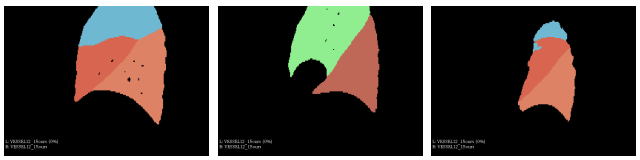
TABLE I: Performance evaluation on the test set for the proposed method and variations. **pl-DC**: per-lobe Dice Coefficient; **il-DC**: inter-lobar Dice Coefficient; **UR**: Upper Right Lobe; **MR**: Medium Right Lobe; **LR**: Lower Right Lobe; **UL**: Upper Left Lobe; **LL**: Lower Left Lobe. **$** refers for the methods that were not able to converge and were trained reducing the network size.

Methods	**pl-DC**							**il-DC**	
	UR	MR	LR	UL	LL	**Avg**	std	**Avg**	std
FRV-Net	0.93	0.87	0.95	0.95	0.94	**0.93**	0.07	**0.85**	0.05
V-Net (w/o any regularization technique) $	0.82	0.68	0.89	0.87	0.79	0.81	0.09	0.62	0.13
FRV-Net w/o Dropout	0.94	0.88	0.94	0.96	0.95	**0.93**	0.06	**0.85**	0.06
FRV-Net w/o Batch Normalization $	0.84	0.75	0.89	0.86	0.87	0.84	0.12	0.66	0.11
FRV-Net w/o Deep Supervision	0.91	0.88	0.93	0.93	0.94	0.92	0.05	0.84	0.07
FRV-Net w/o Multi-Task Learning	0.92	0.86	0.91	0.77	0.91	0.88	0.08	0.80	0.06

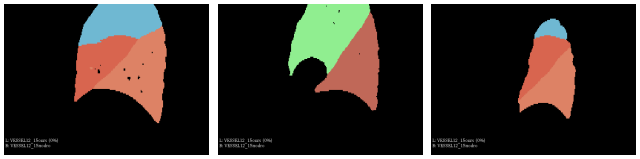


(a) Original scan

(b) Ground-Truth

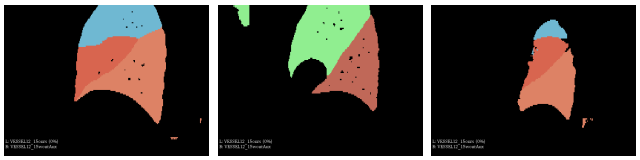


(c) FRV-Net



(d) FRV-Net w/o Dropout

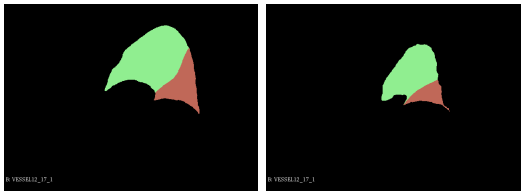
(e) FRV-Net w/o Deep Supervision



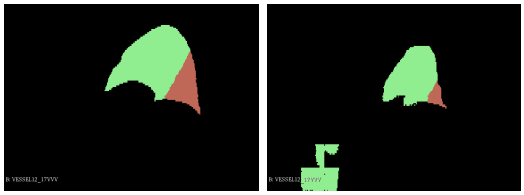
(f) FRV-Net w/o Multi-Task Learning

Fig. 8: Qualitative comparison in a scan of the test set. **(a)** Sagittal slices of the same scan; **(b)** Ground-truth; **(c)** FRV-Net output; **(d)** FRV-Net without dropout; **(e)** FRV-Net without deep supervision; **(f)** FRV-Net without multi-task learning.

(a) Original Scan



(b) ground-truth



(c) Prediction

Fig. 9: Example of some failures that usually happen with the proposed method.

allow to understand how regularization techniques influence the results. For instance, Dropout does not seem to improve the accuracy, since the FRV-Nets with and without Dropout have the same results. Without the Batch Normalization (V-Net and FRV-Net w/o Batch Normalization) this network was not able to converge, we decided to train an analogous model, but decreasing its complexity by simply halving the number of channels on each convolution. The Batch Normalization became fundamental for training this deep network.

Multi-task learning is important for an efficient training since it gives focus for the most difficult scenarios allowing more representability to the network. Deep Supervision, in our case, represents a marginal increase on the accuracy from 0.92 to 0.93 in the average pl-DC and 0.84 to 0.85 in the average il-DC.

3) Failure Analysis: The proposed method is trained with few examples and mostly because of that it still does not

deal correctly with some situations. For instance, in absence or large difference in the size of the lobar structure among the training examples, the method usually predicts a wrong separation of the lobes. An example of this behavior can be seen in Fig. 9. Another mistake that we observed in our model is the residual segmentation of lobar regions completely outside the lung region. However, this drawback could be easily addressed with some simple post-processing operations, such as simply select the larger volumes of each lobe.

IV. Conclusion

In this paper, we have presented FRV-Net, the first supervised method for lobar segmentation in CT scans of the lung that can process 3D information in an end-to-end fashion. FRV-Net does not depend on heavy pre and post-processing strategies, and it can bypass the lack of training data by making use of carefully selected regularization techniques, including Batch Normalization, Multi-Task Learning, Deep Supervision, and Dropout. Our experiments show that, among all these techniques, Batch Normalization is essential for an effective training. In addition, Multi-Task Learning (with the additional task of locating inter-lobar fissures) helped focusing the attention of the network in the most difficult cases, whereas Deep Supervision produced a more stable representation of the lobar structures. Finally, Dropout seemed to add only marginal improvements, with a less critical impact in the overall performance of the model.

Numerical performance was assessed by means of a new overlap metric, inter-lobar Dice coefficient, allowing a more suitable verification of the capability of the model to correctly separate the lobes. An interesting open goal for future research will be to extend the method to segment simultaneously lung lobes and other anatomical structures like airways and vessels.

Acknowledgement

This work is supported by Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[1] Samuel G Armato, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38(2):915–931, 2011.

[2] F. J. S. Bragman, J. R. McClelland, J. Jacob, J. R. Hurst, and D. J. Hawkes. Pulmonary lobe segmentation with probabilistic segmentation of the fissures and a groupwise fissure prior. *IEEE Transactions on Medical Imaging*, 36(8):1650–1663, Aug 2017.

[3] Rich Caruana. Multitask learning. In *Learning to learn*, pages 95–133. Springer, 1998.

[4] Hao Chen, Qi Dou, Lequan Yu, Jing Qin, and Pheng-Ann Heng. Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. *NeuroImage*, 2017.

[5] Tom Doel, David J Gavaghan, and Vicente Grau. Review of automatic pulmonary lobe segmentation methods from ct. *Computerized Medical Imaging and Graphics*, 40:13–29, 2015.

[6] Qi Dou, Lequan Yu, Hao Chen, Yueming Jin, Xin Yang, Jing Qin, and Pheng-Ann Heng. 3D deeply supervised network for automated segmentation of volumetric medical images. *Medical Image Analysis*, 41(Supplement C):40 – 54, 2017.

[7] Kevin George, Adam P. Harrison, Dakai Jin, Ziyue Xu, and Daniel J. Mollura. Pathological pulmonary lobe segmentation from ct images using progressive holistically nested neural networks and random walker. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pages 195–203, Cham, 2017. Springer International Publishing.

[8] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.

[9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[10] 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. Medical Image Computing and Computer-Assisted Intervention, pages 621–629, Cham, 2017. Springer International Publishing.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, pages 1026–1034, Washington, DC, USA, 2015. IEEE Computer Society.

[12] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

[13] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[14] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Artificial Intelligence and Statistics*, pages 562–570, 2015.

[15] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42(Supplement C):60 – 88, 2017.

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

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[18] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.

[19] Rina D Rudyanto, Sjoerd Kerkstra, Eva M Van Rikxoort, Catalin Fetita, Pierre-Yves Brillet, Christophe Lefevre, Wenzhe Xue, Xiangjun Zhu, Jianming Liang, İlkay Öksüz, et al. Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the vessel12 study. *Medical Image Analysis*, 18(7):1217–1232, 2014.

[20] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin A. Riedmiller. Striving for simplicity: The all convolutional net. *CoRR*, abs/1412.6806, 2014.

[21] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[22] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sébastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *arXiv preprint arXiv:1707.03237*, 2017.

[23] E van Rikxoort, B van Ginneken, and S Kerkstra. Lobe and lung analysis 2011 (lola11), 2011.

[24] Yuyin Zhou, Lingxi Xie, Wei Shen, Yan Wang, Elliot K. Fishman, and Alan L. Yuille. A fixed-point model for pancreas segmentation in abdominal ct scans. In *MICCAI*, 2017.