

A SR-NET 3D-TO-2D ARCHITECTURE FOR PARASEPTAL EMPHYSEMA SEGMENTATION

D. Bermejo-Peláez¹, Y. Okajima², G. R. Washko², M. J. Ledesma-Carbayo^{†1}, R. San José Estépar^{†2}

¹Biomedical Image Technologies, Universidad Politécnica de Madrid & CIBER-BBN, Madrid, Spain

²Applied Chest Imaging Laboratory, Brigham and Womens Hospital, Boston, MA, USA

ABSTRACT

Paraseptal emphysema (PSE) is a relatively unexplored emphysema subtype that is usually asymptomatic, but recently associated with interstitial lung abnormalities which are related with clinical outcomes, including mortality. Previous local-based methods for emphysema subtype quantification do not properly characterize PSE. This is in part for their inability to properly capture the global aspect of the disease, as some the PSE lesions can involved large regions along the chest wall. It is our assumption, that path-based approaches are not well-suited to identify this subtype and segmentation is a better paradigm. In this work we propose and introduce the Slice-Recovery network (*SR-Net*) that leverages 3D contextual information for 2D segmentation of PSE lesions in CT images. For that purpose, a novel convolutional network architecture is presented, which follows an encoding-decoding path that processes a 3D volume to generate a 2D segmentation map. The dataset used for training and testing the method comprised 664 images, coming from 111 CT scans. The results demonstrate the benefit of the proposed approach which incorporate 3D context information to the network and the ability of the proposed method to identify and segment PSE lesions with different sizes even in the presence of other emphysema subtypes in an advanced stage.

Index Terms— Parasetal emphysema, Deep learning, Convolutional neural networks, Segmentation

1. MOTIVATION

Pulmonary emphysema is characterized by abnormal permanent enlargement of airspaces distal to the terminal bronchiole, accompanied by the destruction of their walls, and without obvious fibrosis [1]. Emphysema is divided into three major subtypes: centrilobular, panlobular and paraseptal emphysema.

[†] RSJE and MJLC (ORCID: 0000000168463923) contributed equally to this work. This study was supported by project RTI2018-098682-B-I00. DBP (ORCID: 0000000201813957) was supported by a FPU grant by the Spain's Ministry of Education. RSJ was supported by NHLBI grant 5R01HL116931 and 5R21HL140422. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

Paraseptal emphysema (PSE), also known as distal acinar emphysema, is characterized by subpleural and peribronchovascular regions of low attenuation separated by intact interlobular septa [2]. PSE is a relatively unexplored emphysema subtype with a very distinct clinical manifestation that is often underrecognized clinically. However, recent studies demonstrated a significant association between paraseptal emphysema (PSE) and interstitial lung abnormalities (ILA) [3], which the latter have been associated with clinical outcomes, including mortality [4].

Densitometric analysis in computed tomography (CT) is commonly used and widely accepted to identify and quantify the extent of pulmonary emphysema. This method discriminates emphysematous and non-emphysematous tissue by thresholding the intensity levels in the CT image within the lung region.

A densitometric approach disregards the information present in the morphology of the emphysema subtypes. Thus, several methods have been previously proposed to carry out emphysema subtype classification exploiting local texture and intensity patterns [5, 6]. The justification for using a local-based approach is usually founded on the pseudo-independence between adjacent regions of interest (ROI) that correspond to the secondary pulmonary lobule (smallest subunit of the lung). Although this approach is valid for detecting centrilobular and panlobular subtypes, it has been observed that is not able to resolve the identification of PSE properly. Given its particular structure, PSE is not only defined as a local pattern but it is a context-based lesion with a notion of location, always appearing adjacent to a pleural surface, as well as lobar septal walls, and which may have a considerable extension. This characteristics make the PSE identification problem consistent with a segmentation approach.

In this work, we present the first technique to particularly address paraseptal emphysema identification by proposing the Slice-Recovery architecture (*SR-Net*) that exploits 3D information to perform segmentation of PSE lesions on 2D slices of CT images.

2. METHODS

3D manual annotation carries a high burden from experts, and when they make 2D annotations they usually use 3D information to place value on the spatial context around the region to

segment. In this work, we propose the new Slice-Recovery network (*SR-Net*) for automatic segmentation of paraseptal emphysema (PSE) on CT images. The proposed CNN architecture, fully described in Section 2.1.2, leverages 3D information for 2D segmentation of PSE in CT axial images. We hypothesize that incorporating 3D context information can improve 2D segmentation results in medical imaging by capturing volumetric information around the lesion.

2.1. SR-Net

2.1.1. Volume extraction

The input of the proposed network consists on a 3D volume of size $384 \times 384 \times 8$ voxels, where the central axial slice corresponds to the target 2D slice to be segmented. This input volume is further pre-processed by segmenting and masking only the lung region.

2.1.2. CNN architecture

The proposed SR-Net is illustrated in Figure 1 and its innovative dense efficient architecture decodes 2D segmentation maps from a 3D encoded volume through a Slice-Recovery layer. The encoding pathway of the proposed CNN is composed by encoder blocks. Each encoder block consists of a convolutional dense block with an iterative concatenation of previous feature maps. Dense blocks introduced in [7] have been previously used to deal with the problem of semantic segmentation [8]. These dense blocks are subsequently processed by an ENet-style block which combines max pooling and strided convolutions to avoid representation bottlenecks [9]. We have previously shown that a purely 3D version of the proposed architecture outperforms state-of-the-art segmentation CNNs such as U-Net [10] when segmenting 3D pulmonary arteries in CTA images [11].

On the other hand, in the decoding pathway, transposed convolutional layers (deconvolution) and skip-connections are subsequently processed by a dense block. Skip-connections of the SR-Net are designed to pass higher resolution feature maps to the decoding path. These connections are based on a Slice-Recovery (SR) layer. We designed the SR layer to select the higher resolution feature maps that correspond to the original target 2D image to be segmented (highlighted in red in Figure 1). Thus, the decoding path can generate a 2D segmentation map based only on the higher resolution features of the slice under study rather than the whole input 3D volume. Deconvolution operation in the decoding path is performed to generate 2D segmentation maps from contracted features extracted by the encoding path from the 3D volume by fixing the third dimension of the 3D convolutional kernels to be 1. The decoding path ends with a $1 \times 1 \times 1$ convolution with a sigmoid activation function to generate the final probability map.

The network were trained on data augmented with random non-rigid distortions. In this work we used Adam Stochastic Optimization (with an initial learning rate $1e-4$ and plateau learning rate decay with a factor of 0.2 when the validation loss is not improved after 5 epochs) to minimize a class-weighted Dice-based loss function to consider class imbalance.

The network was implemented using Keras with TensorFlow, and built in a PC with GPU NVIDIA TITAN X Pascal 12GB, CPU Intel Core i7 3.6 GHz and 32GB of RAM.

2.2. Comparative methods

To prove the potential of incorporating 3D context to the network for 2D segmentation problems, we compared the proposed 3D-to-2D SR-Net to a purely 2D version of the latter (2D-Net). Additionally, we compared the proposed method to the UNet [10], a strong baseline method in medical image segmentation problems, and to the converted 3D-to-2D version of the UNet following the proposed SR-Net approach (SR-UNet). We therefore prove the possibility and potential of converting previously designed architectures using the proposed SR-Net paradigm. Additionally, we will compare the proposed method to a local-based approach [6].

3. EXPERIMENTS AND RESULTS

3.1. Dataset and evaluation approach

An experienced radiologist manually segmented PSE lesions on 664 2D axial images coming from 111 CT scans from the COPDGene study. COPDGene centers obtained approval from their Institutional Review Boards and all subjects provided written informed consent. Of all of these subjects, 89 were randomly selected for training and validation, while the remaining 22 subjects were used for testing. Both training (55%) and validation (20%) sets comprised 498 axial slices, while the independent test set was composed by 166 images (25%). The evaluation of the proposed method as a well as the comparison to other methods were carried out on the test set. Training and validation sets were used to hyperparameter tuning of the proposed architecture.

Segmentation accuracy was evaluated using Dice Similarity Coefficient (DSC), Jaccard Similarity Coefficient (JSC), and Mean Accuracy (MA), defined as:

$$\begin{aligned} \text{DSC} &= \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \\ \text{JSC} &= \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \\ \text{MA} &= \frac{\text{TNR} + \text{TPR}}{2} \end{aligned} \quad (1)$$

where TP, FP, TN and FN denote the number of true positives, false positives, true negatives and false negatives respectively. TNR and TPR denote true negative and true positive rates.

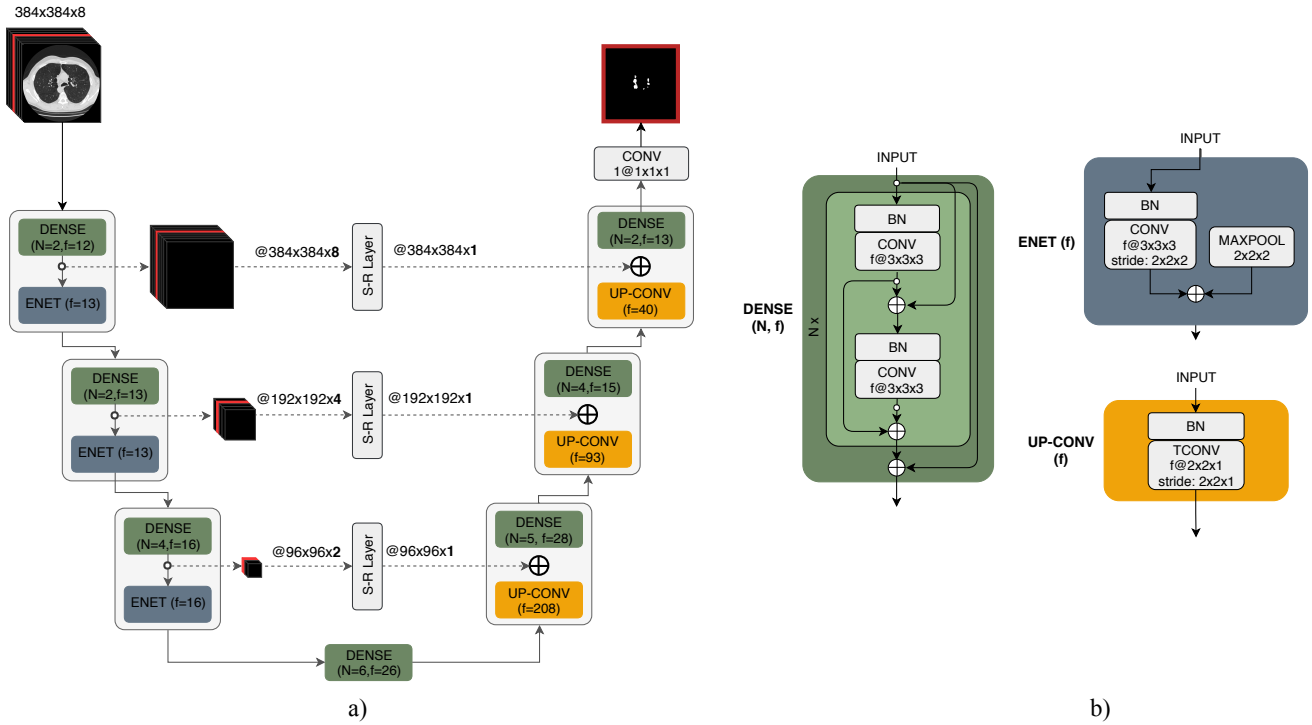


Fig. 1. a) SR-Net architecture which generates a 2D segmentation map given a 3D input volume. b) Example of a dense block of 2 layers ($N=2$) (green), ENET-style block (blue) and Up-convolutional block (yellow).

The lung area affected with PSE can vary substantially from one patient to another, ranging from 0.23 cm^2 up to 134 cm^2 . Overlap ratio measures have been shown unsuitable for comparing segmentation accuracy on objects that differ in size. To compensate for this drawback, all reported metrics correspond to a weighted average of the previous metrics according to the area of the lesions, thus giving a higher weight to those lesions with a greater extension. These weighted metrics are consistent with the overall measure used to report disease extension that would be percentage of volume affected by PSE.

3.2. Results

The evaluation metrics are computed between the manually segmented images and the thresholded probability maps generated by the CNN. The threshold was chosen so that it maximizes the performance on the validation set.

An overview of the results are shown in Table 1, providing a comparison of the proposed SR-Net with a purely 2D version of the architecture as well as with UNet and a converted version of UNet using the proposed SR-Net approach. All the networks were trained with the same training samples, and tested with the same test set. Table 1 proves the potential of incorporating volumetric 3D context information to the network for a 2D segmentation problem. We can observe how both converted 3D-to-2D versions of the baseline architectures outperform their 2D versions by 22% and 5.1%

Table 1. Overview of the results obtained with the proposed method and comparative methods (weighted mean \pm std). Results are computed on the independent test set. DSC: Dice Similarity Coefficient, JSC: Jaccard Similarity Coefficient, MA: Mean Accuracy.

	DSC	JSC	MA
2D-UNet	0.492 (\pm 0.047)	0.355 (\pm 0.044)	0.713 (\pm 0.020)
SR-UNet	0.712 (\pm 0.013)	0.563 (\pm 0.016)	0.844 (\pm 0.007)
2D-Net	0.715 (\pm 0.019)	0.572 (\pm 0.021)	0.864 (\pm 0.006)
SR-Net	0.764 (\pm 0.014)	0.630 (\pm 0.017)	0.908 (\pm 0.004)
Local-based [6]	0.448 (\pm 0.014)	0.295 (\pm 0.008)	0.730 (\pm 0.003)

respectively on terms of mean DSC.

Figure 2 shows the segmentation results of 5 cases selected from the test set with different PSE extension and in the presence of other emphysema subtypes, even with an advanced stage.

Results shown in Table 1 and Figure 2 also prove how a local-based approach fails at properly characterize PSE.

4. DISCUSSION AND CONCLUSIONS

In this work, we present a dense efficient 3D-to-2D CNN architecture (referred as SR-Net) for automatic paraseptal emphysema (PSE) segmentation in axial 2D images leveraging and exploiting contextual 3D information. Additionally, we

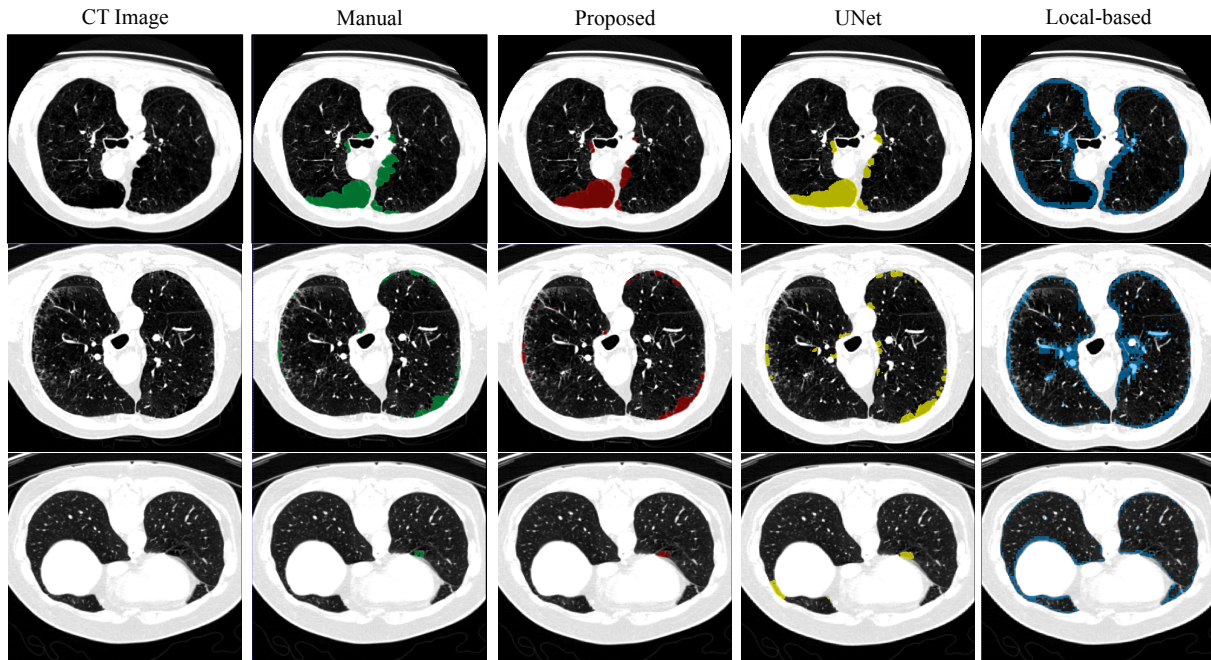


Fig. 2. Comparison of the segmentation results for the proposed method, UNet and a local-based approach [6] for 3 cases from the test set with different emphysema extension.

have shown the benefit of this approach when applied to previous state-of-the-art CNN architectures.

Manual annotation of regions in 3D is a non-trivial task which demands an intensive and time-consuming labour from experts. On the other hand, when experts annotate 2D images belonging to a 3D volume, they usually use 3D information to reach optimal 2D annotations placing value on contextual information around the region to segment. Considering these premises, we proposed an approach based on 2D annotations but incorporating spatial 3D information.

It has been shown that previous local-based methods to emphysema subtype identification, achieve remarkable accuracy in characterizing an identifying centrilobular and panlobular emphysema subtypes, but are not able to properly characterize PSE. The proposed method tackles PSE identification considering the specific nature of PSE, consistent with a context-based segmentation approach.

The results, which were obtained on an independent test set, demonstrated that the proposed SR-Net outperforms state-of-the-art methods when segmenting PSE lesions on CT images.

5. REFERENCES

- [1] G. L. Snider, et al., “The definition of emphysema - report of a national-heart-lung-and-blood-institute, division of lung-diseases workshop,” *Am Rev Respir Dis*, vol. 132, no. 1, pp. 182–185, 1985.
- [2] David M. Hansell, et al., “Fleischner society: Glossary of terms for thoracic imaging,” *Radiology*, vol. 246, no. 3, pp. 697–722, 2008.
- [3] Tetsuro Araki, et al., “Paraseptal emphysema: Prevalence and distribution on ct and association with interstitial lung abnormalities,” *Eur J Radiol*, vol. 84, no. 7, pp. 1413–1418, 2015.
- [4] George R. Washko, et al., “Identification of early interstitial lung disease in smokers from the copdgene study,” *Acad Radiol*, vol. 17, no. 1, pp. 48–53, 2010.
- [5] L. Sorensen, et al., “Quantitative analysis of pulmonary emphysema using local binary patterns,” *IEEE Trans Med Imaging*, vol. 29, no. 2, pp. 559–569, 2010.
- [6] David Bermejo-Peláez, et al., “Emphysema classification using a multi-view convolutional network,” *IEEE ISBI*, pp. 519–522, 2018.
- [7] G. Huang, et al., “Densely connected convolutional networks,” in *CVPR 2017*, 2017, pp. 2261–2269, ID: 1.
- [8] Simon Jégou, et al., “The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation,” *CVPRW 2017*, pp. 1175–1183, 2017.
- [9] Adam Paszke, et al., “Enet: A deep neural network architecture for real-time semantic segmentation,” *CoRR*, vol. abs/1606.02147, 2016.
- [10] Olaf Ronneberger, et al., “U-net: Convolutional networks for biomedical image segmentation,” *MICCAI 2015*, vol. 9351, pp. 234–241, 2015.
- [11] Karen López-Linares Román, et al., “3D pulmonary artery segmentation from CTA scans using deep learning with realistic data augmentation,” in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. 2018, pp. 225–237, Springer International Publishing.