IMAGE & SIGNAL PROCESSING

Conventional Filtering Versus U-Net Based Models for Pulmonary Nodule Segmentation in CT Images

Joana Rocha[1,2] (ID) · **António Cunha[2,3]** · **Ana Maria Mendonça[1,2]**

Abstract

Lung cancer is considered one of the deadliest diseases in the world. An early and accurate diagnosis aims to promote the detection and characterization of pulmonary nodules, which is of vital importance to increase the patients' survival rates. The mentioned characterization is done through a segmentation process, facing several challenges due to the diversity in nodular shape, size, and texture, as well as the presence of adjacent structures. This paper tackles pulmonary nodule segmentation in computed tomography scans proposing three distinct methodologies. First, a conventional approach which applies the Sliding Band Filter (SBF) to estimate the filter's support points, matching the border coordinates. The remaining approaches are Deep Learning based, using the U-Net and a novel network called SegU-Net to achieve the same goal. Their performance is compared, as this work aims to identify the most promising tool to improve nodule characterization. All methodologies used 2653 nodules from the LIDC database, achieving a Dice score of 0.663, 0.830, and 0.823 for the SBF, U-Net and SegU-Net respectively. This way, the U-Net based models yield more identical results to the ground truth reference annotated by specialists, thus being a more reliable approach for the proposed exercise. The novel network revealed similar scores to the U-Net, while at the same time reducing computational cost and improving memory efficiency. Consequently, such study may contribute to the possible implementation of this model in a decision support system, assisting the physicians in establishing a reliable diagnosis of lung pathologies based on this segmentation task.

Keywords Computer-aided diagnosis · Conventional · Deep learning · Image analysis · Lung · Nodule · Segmentation

Introduction

Pulmonary nodules can be associated with several pathologies, among them lung cancer, which is the main cause of cancer death in men and the second cause in women worldwide [14]. For this reason, providing an early detection and diagnosis to the patient is crucial, considering that any delay

✉ Joana Rocha
joana866rocha@gmail.com

1 Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias s/n, 4200 - 465, Porto, Portugal

2 INESC Technology and Science, Rua Dr. Roberto Frias s/n, 4200 - 465, Porto, Portugal

3 Universidade de Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801, Vila Real, Portugal

in cancer detection might result in lack of treatment efficacy. The advances of technology and imaging techniques such as Computed Tomography (CT) have contributed to nodule identification and monitoring; more specifically, the segmentation process within a Computer-Aided Diagnosis (CAD) system has facilitated their location and characterization, thus differentiating the nodule from other structures. However, this task is quite complex considering the heterogeneity of the size, texture, position, and shape of the nodules, and the fact that their intensity can vary within the borders. When it comes to biomedical image segmentation, early methods (generally described as conventional) followed a set of principles and logical rules to deduce new information [4]. Among other conventional techniques, lesion segmentation often implies the use of filters, such as the Sliding Band Filter, which has previously been used e.g. to develop an automated method for optic disc [2] and cell segmentation [10]. Machine Learning approaches have then appeared and lately have been preferred over the conventional ones. They became a trend by extracting features and feeding them to a statistical classifier. Support

Vector Machines, Decision Trees and K-Nearest Neighbors are common techniques implemented within this trend for lesion segmentation [9]. More recently, Deep Learning established its dominant role in CAD systems and segmentation tasks [5], by automatically extracting knowledge from a large quantity of data. Convolutional Neural Networks and, more specifically, Fully Convolutional (FC) Networks are a frequent example that is usually applied for medical imaging segmentation [12], encompassing e.g. encoder-decoder structures for semantic segmentation tasks such as the SegNet [1], or the U-Net, a particular example of an encoder-decoder network initially developed for biomedical segmentation [11]. More recently, hybrid networks have appeared, where the SegNet and the U-Net can be combined to achieve a more memory-efficient model, which is also able to capture fine details. That is the case of the hybrid networks proposed for knee bone tumor segmentation [3], and brain tissue segmentation [7]. Both these algorithms proved to perform better than the SegNet and the U-Net individually.

This work aims to precisely segment pulmonary nodules using not only a conventional approach based on the Sliding Band Filter, but also two Deep Learning based approaches focused on FC encoder-decoder networks, as is the case of the U-Net and the SegU-Net, a novel hybrid network introduced in the current paper. Therefore, this paper presents the evaluation and comparison of the selected methodologies.

Conventional approach

The conventional approach selected for this paper is based on a Local Converge Filter (LCF) since these tend to work well for noisy low-contrasted images, which is often the case of biomedical images. LCFs estimate and maximize the convergence degree of the gradient vectors within a support region R, toward a central pixel of interest $P(x, y)$, assuming that the studied object has a convex shape and limited size range. The overall convergence degree is obtained by averaging the individual convergences at all points in R - each of them minding the orientation angle $\theta_i(k, l)$ of the gradient vector at point (k, l) with respect to the line with radial direction i [2]. Being a member of the LCFs, the SBF is not influenced by gradient magnitude, nor by the contrast with the surrounding structures. Instead, the band of fixed width d which comprises the support region is adapted in each one of the N radial directions leading out of P, minding the gradient orientation in order to maximize the convergence degree. Such feature makes this filter more versatile when it comes to detecting different shapes, since

the support region can be molded. More specifically, the support region is adapted in each direction i along a radial length that varies from a minimum ($Rmin$) to a maximum ($Rmax$) values, originating the filter's response minding the angle of the gradient vector at the point m pixels away from P. The coordinates of the band's support points $(X(\theta_i), Y(\theta_i))$ are obtained, assuming that the center of the object is known [13].

Deep learning based approaches

The state of the art has been greatly improved when it comes to object detection and segmentation, and overall region recognition thanks to Deep Learning [8]. This work involves semantic segmentation in the context of a binary classification problem, where a pixel is either nodular or non-nodular. To complete the task, two FC encoder-decoder style architectures are implemented: the U-Net, and a novel hybrid between the SegNet and the U-Net, designated as SegU-Net. Both structures include an encoder and decoder components, ending with a classification layer whose output is a pixel-wise segmentation mask. The encoder receives the input image and downsamples it into features maps with different levels of abstraction (low-to-high level features), while the decoder receives those feature maps and upsamples them until the output has the same resolution as the input, ensuring a precise pixel location of the object.

The U-Net is an improved FC Network specially developed for biomedical image segmentation, thus requiring a smaller amount of parameters and computing time [11]. The architecture of the U-Net was defined as follows: an encoder block comprises two 3×3 unpadded convolutions, each followed by a ReLU with batch normalization, a 2×2 max pooling layer with stride equal to 2, and a dropout layer; while a decoder block includes a upconvolution layer with stride equal to 2, a concatenation with the corresponding cropped feature map from the encoder block, and two 3×3 convolutions, each followed by ReLU with batch normalization. This way, the U-Net has four encoder blocks, a bottleneck, and four decoder blocks, ending with an end-most layer that includes a 1×1 convolution with sigmoid activation. The number of feature channels is doubled with each downsampling phase of the encoder, and then doubled again in each upsampling of the decoder, creating a symmetric architecture. The skip connections at the same depth level in the U-Net are an extremely important and effective way to transfer the low-level features from the encoder to the decoder, where these are concatenated with high-level features, generating precise spacial information.
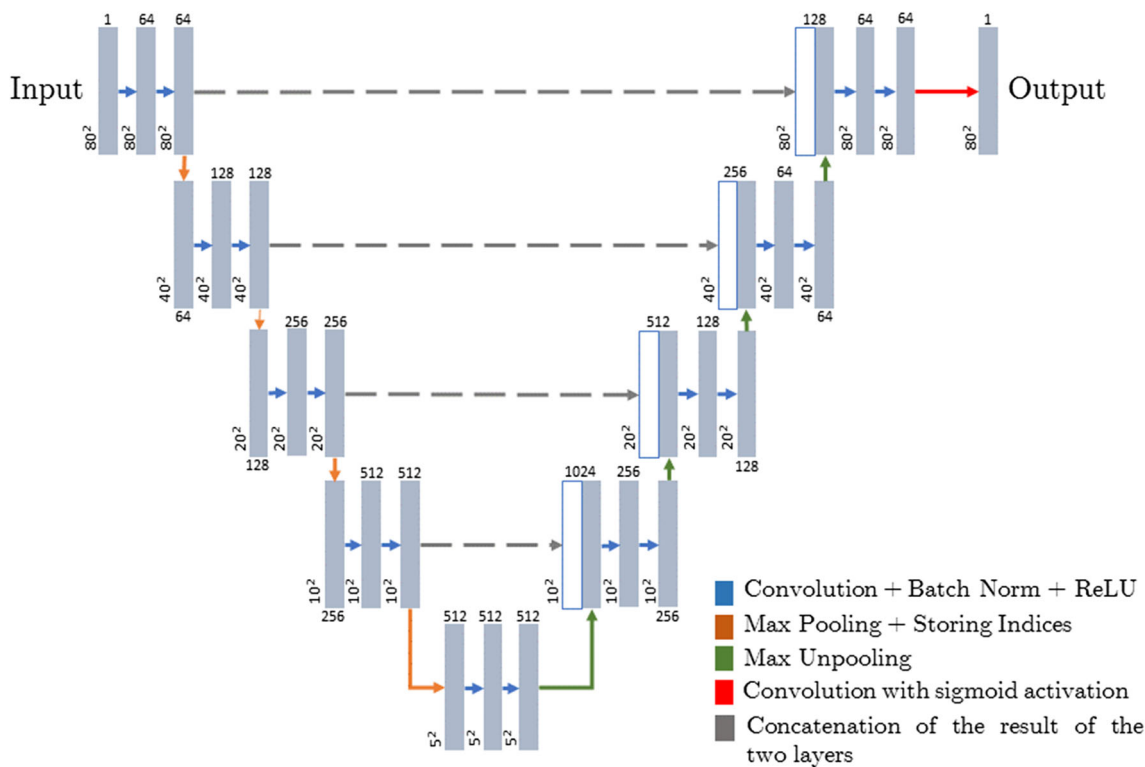
Fig. 1 SegU-Net's model schematics

The SegU-Net was developed and applied to the same dataset, by replacing the U-Net's upsampling method by the SegNet's. In other words, the previous description of the U-Net still applies, in the sense that the SegU-Net has four encoder blocks, a bottleneck, and four decoder blocks. However, the encoder block performs downsampling using max pooling and storing the max pooling indices, while the decoder block uses the SegNet's 2×2 max unpooling, instead of the U-Net's upconvolution operation. Figure 1 outlines the SegU-Net's architecture. The skip connections of the U-Net are kept in this new hybrid network, to ensure the concatenation between the encoder and decoder features. This way, the SegU-Net takes advantage of distinct characteristics from the SegNet and the U-Net, combining both into a new architecture dedicated to restoring pixel position information and ideally achieve finer edge details, while at the same time reducing computational cost and increasing memory efficiency.

Methodologies for nodule segmentation

The ground truth for this exercise consisted of segmentation masks from the LIDC database, which is publicly available and consists of lung cancer screening thoracic CT scans from 1010 patients. The following algorithms were applied on 2653 nodule candidates from that database, whose images are the output of a detection scheme. For each nodule, the 3D volume around its center was split into three anatomical planes (sagittal, axial, and coronal), resulting in three 80×80 pixel images per nodule. For clarity and brevity reasons, the conventional method will be explained for a single plane.

Sliding band filter

The conventional methodology can be split into three main steps. The SBF is first applied to get a better estimation

Fig. 2 Exemplification of the conventional methodology steps, where the blue mark is the center of the image, the green mark is the ground truth center of the nodule, and the red mark is the estimated center of the nodule
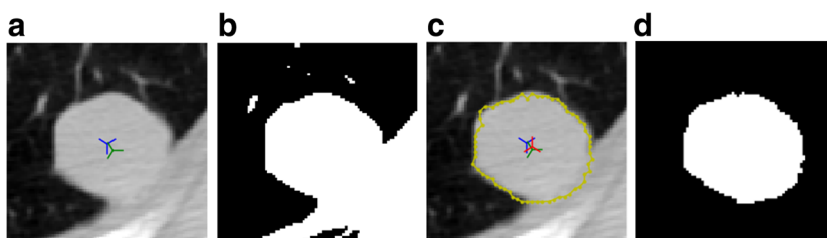
of the nodule's center coordinates, as shown in Fig. 2a. Considering that most nodules have an overall uniform intensity, a truncated binary mask is generated (Fig. 2b), containing exclusively the pixels with similar intensity of to the nodule's. The SBF receives the original nodule image and the truncated mask, and calculates its response for each pixel, defining as the estimated nodule's center the pixel which maximizes that response. With those coordinates, the SBF then evaluates the corresponding set of support points, returning the N border coordinates marked in Fig. 2c with yellow. To ensure the SBF is as precise as possible, a condition was added to force the cosine of the gradient vector's orientation angle to be null when the pixel which is being evaluated in a certain direction is null in the truncated mask. Ideally, this keeps the SBF from including in the segmentation non-nodular regions within the $Rmin$ and $Rmax$ limits. The calculations for the SBF were obtained with the parameter values $N = 64$, $d = 7$, $Rmin = 1$, and $Rmax = 25$, which were established empirically to maximize the algorithm's performance.

To refine the initial segmentation, only the intersection of the SBF segmentation mask and the truncated nodule mask is considered. Any cavities within the intersected binary masks are filled. By labeling all the different regions present in the intersected masks, which are identified by their connected components, it is possible to eliminate any region that has no connection to the nodule and specifically select the nodular area. After this step, the final segmentation mask is achieved, as exemplified in Fig. 2d.

U-Net and SegU-Net

Both Deep Learning algorithms presented in this work are implemented using *Keras*, with a *TensorFlow* backend. The 2D images are imported and split into training, validation, and test sets, as described in Table 1. Real-time data augmentation is applied to the training set, replicating tissue deformations through affine transformations (e.g. 0.2 degrees of shear and random rotation within a 90 degree range), and generating more training data with horizontal and vertical flips.

It was necessary to take into consideration the class imbalance within a sample (generally, there are more non-nodular pixels in an image than nodular ones), and so a Dice based loss function was selected. The training stage of the

Table 1 Composition of the train, validation and test sets

	Train set	Validation set	Test set	Total
#Nodules	1591 (60%)	531 (20%)	531 (20%)	2653

model is guided by two evaluation metrics: accuracy and Jaccard Index.

The networks were trained with the Adam optimizer to achieve a faster stable convergence, using the default hyperparameter values [6]. While training the model, callbacks were included. First, early stopping ensures the training ends when the validation loss stops improving. At the same time, the learning rate also reduces on plateau, meaning that it is reduced when the validation loss cannot reach a lower value. More specifically, the initial learning rate value was the default for the Adam optimizer, and will be reduced by a factor of 0.1 (new learning rate = learning rate × 0.1), having a minimum accepted value of 0.00001. The model is fit on batches with real-time data augmentation (using a batch size of 64 samples for the U-Net and 32 samples for the SegU-Net), allowing a maximum of 100 epochs and minding a dropout regularization of 50%. After analyzing the validation loss for every epoch, the training weights which maximize the evaluation metrics and minimize the loss are stored, to get the predictions of the test set. The pixel-wise probabilities that resulted from the sigmoid activation faced a 50% threshold to decide whether a pixel is nodular or not.

Results and discussion

Several adversities are encountered when evaluating medical segmentation, and for this reason it is very important to establish an adequate standard evaluation system, and consequently select which evaluation metrics to use as a mean of comparison of the algorithms. This comparison is also done using plots, present in Figs. 3 and 4, which directly assess how close the segmentation is to the ground truth by marking the True Positives (yellow pixels), False Positives (red pixels), and False Negatives (green pixels). The scores for the selected approaches are exhibited in Table 2.

The SBF needed approximately 5 hours to run, while the U-Net and SegU-Net had a reasonable training time of roughly 8 and 7,5 hours respectively on an NVidia GeForce GTX 1080 GPU (8 GB). This indicates that the SBF requires less computational power, followed by the SegU-Net. Each Deep Learning model, i.e. its trained parameters, was stored in a *.h5* format file - the U-Net's file required 121.5 MB and the SegU-Net's required 52.5 MB. Taking these values into consideration, one can confirm that the SegU-Net translates to a smaller number of parameters and consequently higher memory efficiency.

Both the U-Net and SegU-Net models exhibited fast convergence and did not overfit to the training data, considering the validation loss is similar to the training loss.
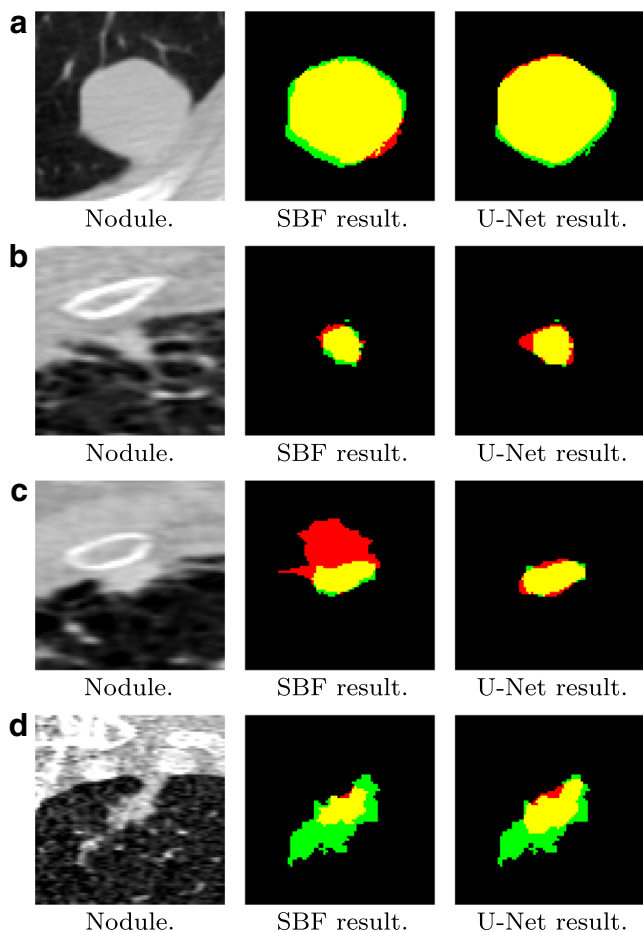
Fig. 3 Comparison between the SBF and the U-Net: in nodule **a** both methodologies are successful, in nodule **b** the SBF outperforms the U-Net, in nodule **c** the U-Net outperforms the SBF, and finally in nodule **d** none of the methodologies have a satisfying result

Fig. 4 Comparison between the SBF and the SegU-Net: in nodule **a** both methodologies are successful, in nodule **b** the SBF outperforms the SegU-Net, in nodule **c** the SegU-Net outperforms the SBF, and finally in nodule **d** none of the methodologies have a satisfying result

Fig. 5 displays the convergence curves for the loss and the accuracy of the models.

The conventional approach exhibits a highly satisfactory performance when dealing with well-circumscribed solid nodules, with overall defined sharp margins. In these cases, both smaller and larger nodules tend to be segmented in accordance with the specialists. The algorithm also deals very well with nodules whose intensities vary within their border (i.e. cavitary and calcific nodules), as it is able to ignore the cavities and calcific regions during the segmentation process. Vascularized nodules have the potential to pose a challenge, considering the inherent difficulty in distinguishing the nodule from the attached vessels. However, the SBF based approach is frequently able to separate them and create a mask which does not include the vessels. Such feat is possible thanks to the truncated
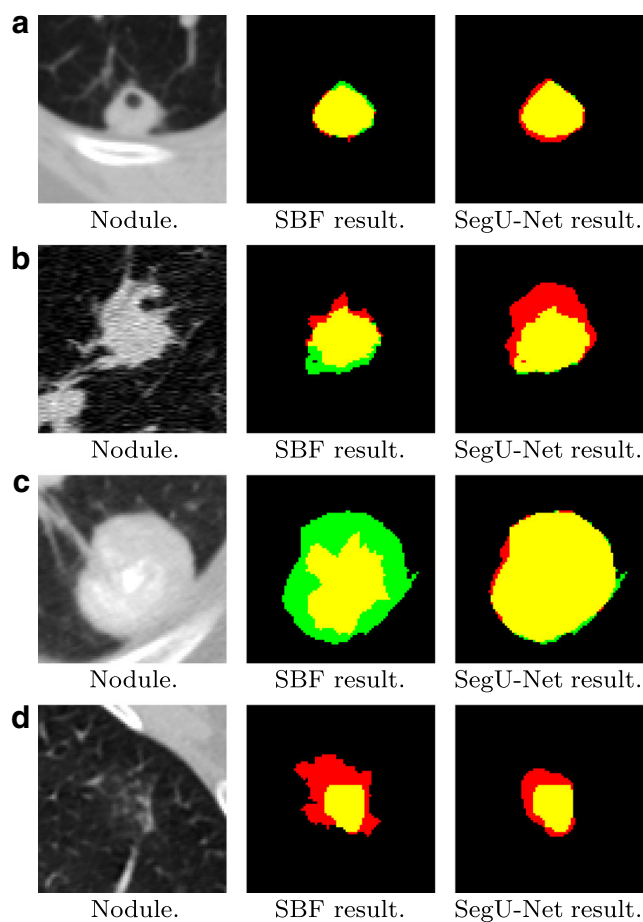
mask, which guides the SBF and consequently is able to remove vascular structures from the segmentation.

The main flaws of the SBF algorithm appear when dealing with juxtapleural nodules: since these lesions are attached to the pleural wall and do not exhibit a sharp margin, the algorithm often does not know where the nodule ends and the pleura begins. In some cases, it is able to estimate to some extent where the nodule ends, while in

Table 2 Evaluation metrics achieved by the conventional and Deep Learning based methods. The number of nodules used to evaluate each method is presented in the last column

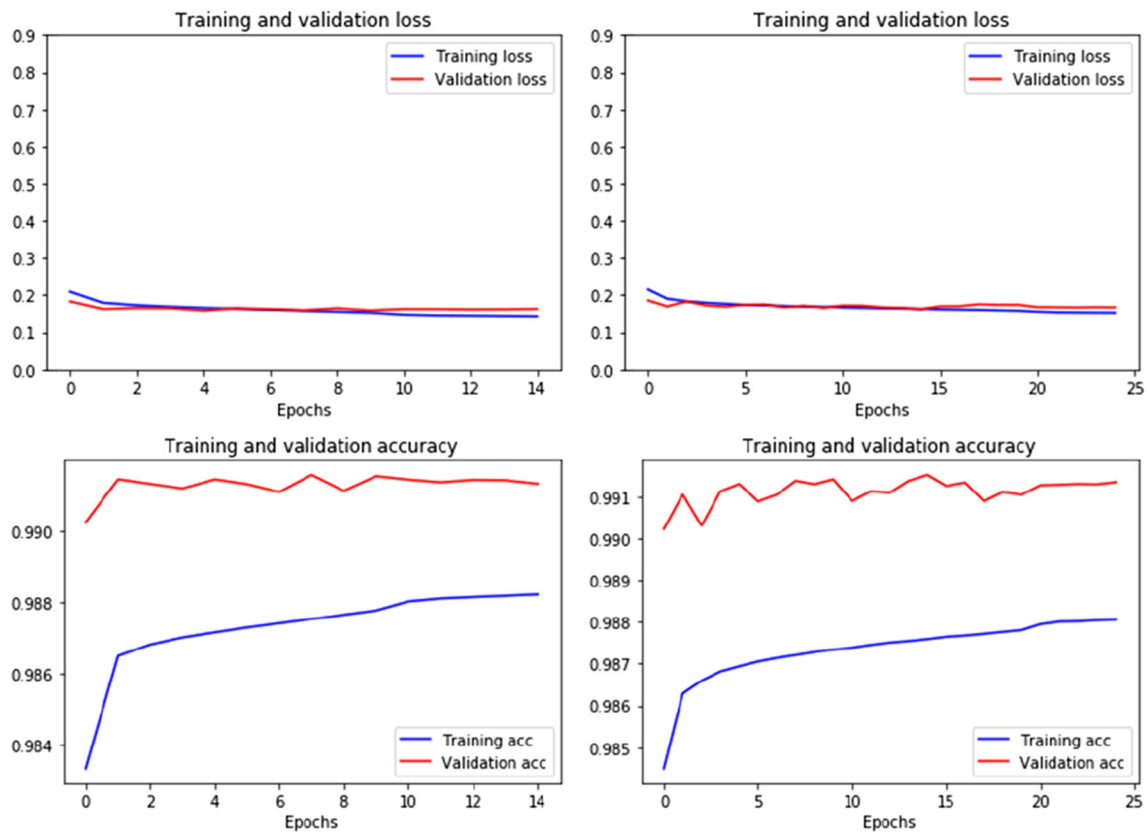|	Dice coefficient	Precision	Recall	#Nodules
SBF	0.663	0.710	0.732	2653
U-Net	0.830	0.792	0.898	531
SegU-Net	0.823	0.787	0.858	531

Fig. 5 Loss and accuracy convergence of the U-Net (left) and SegU-Net (right)

other cases part of the pleural wall is included in the segmentation. The less satisfying results are also due to the unexpected irregular shape of the nodule, or because the nodule does not have a clear margin (non-solid).

Similarly to the SBF based approach, the U-Net is able to clearly segment well-circumscribed solid nodules, independently of their size. It also functions correctly with cavitary and calcific nodules, as well as vascularized nodules. In this last case, the U-Net is able to exclude the vessels from the segmentation. However, unlike the SBF, the U-Net demonstrates a great skill when segmenting juxtapleural nodules, being able to tell almost perfectly where the nodule ends and the pleura begins. The U-Net experiences some degree of difficulty when segmenting irregular shapes and non-solid lesions, but even in such cases this approach is still able to clearly outperform the first. In general, the U-Net segmentation results are very similar to the ones given by the specialists, hence its great performance. In some measure, the U-Net even achieves a more uniform detailed segmentation in comparison to the specialists', which may be justified by its pixel-wise perspicacity. The SegU-Net displayed similar scores to the U-Net, being able to precisely segment solid nodules with defined margins, with or without vasculature and other

structures near them. The same can be stated when working with cavitary and calcific nodules. The algorithm's slightly inferior performance in comparison to the U-Net is mostly due to an increase of difficulty when segmenting irregular shapes, or non-solid lesions. Implementing a U-Net based model has clear advantages - among them, the need for fewer training images and the skip connections which allow the merging of the encoder's coarse contextual information with the more precise spatial information achieved in the decoder, ultimately leading to higher accuracy values. One may argue that in this particular exercise the U-Net's upconvolutions proved to be slightly more efficient due to their learnable parameters, but the SegU-Net's difference in terms of performance is not significant. In fact, applying the SegU-Net's max unpooling also reveals an improved boundary delineation, while at the same time decreasing the amount of parameters needed for end-to-end training (values of the max pooling indices are kept and the rest of the feature map values are zeroed). Additional efforts can be done in future work to improve the algorithms' performance, namely develop a more adequate pre-processing for the input images, in order to promote the accuracy of the segmentation process. More specifically, in the conventional approach, the border coordinates may

be refined for the juxtapleural nodules by establishing a more efficient post-processing stage. In the Deep Learning approaches, the most straightforward way to enhance its performance in non-solid or irregular shaped lesions would be to add more of these examples to the training set, promoting a more advanced and perceptive learning.

Conclusions

The segmentation of pulmonary nodules contributes to their characterization, which makes it a key to assess the patient's health state. This way, a segmentation step implemented within a CAD system can help the physician to establish a more accurate diagnosis. However, the automation of such task is hampered by the diversity of nodule shape, size, position, lighting, texture, etc. The proposed conventional approach deals with these challenges by implementing the Sliding Band Filter to find the coordinates of the borders, and achieves a Dice score of 0.663 when tested using the LIDC database. On the other hand, the Deep Learning based approaches go even further and yield a Dice score of 0.830 and 0.823 for the U-Net and the SegU-Net respectively. All the approaches perform as expected for well-circumscribed obvious nodules, with sharp margins and/or solid texture, and tend to fail when segmenting non-solid or irregular shaped lesions. The conventional approach exhibited the lowest performance scores, which may be justified by its difficulty segmenting juxtapleural nodules, unlike the other algorithms. Based on previous work referenced in the state-of-the-art revision, it was expected for the SegU-Net's hybrid network to outperform the U-Net. The SegU-Net was not able to outperform it, but the difference between their performance is not significant, meaning that they have the same ability to segment pulmonary nodules. Taking into consideration that the main distinction between these two models are their upsampling method, one may highlight that the SegU-Net's max unpooling can be quickly integrated into any encoder-decoder structure in order to achieve fine detail in a segmentation task. By doing so, the computational cost of the operation is reduced and the memory efficiency of the training process is increased, in opposition to the higher number of learnable parameters required in an upconvolution. For these reasons, it can be considered beneficial to choose the SegU-Net over the U-Net.

As mentioned above, the presented methods can be promising segmentation tools for well-circumscribed, solid, cavitary, calcific, and vascularized nodules, and so future work includes the refinement of these methods to deal with their particular challenges. When it comes to the most successful techniques, it would be wise to focus on non-solid or irregular shaped nodules in order to improved the models' ability to precisely segment them, for example by collecting more of these examples for the training set and consequently adjust their parameters to succeed in these types of lesions. To conclude, the comparison of the conventional and Deep Learning based approaches explored the advantages and disadvantages of each technique, establishing the U-Net based models as the most efficient methods in this case - particularly efficient for obvious lesions, and able to overcome to a certain extent the high variability of nodular structures. Comparing the similar performance of both Deep Learning approaches, it is possible to state that the SegU-Net is a promising novel network, since it translates to a lower computational cost and higher memory efficiency. Consequently, the satisfactory segmentation results achieved by the U-Net based models in this work lead to further insights on nodule characterization, contributing to the development of a decision support system, which may be able to assist the physicians in establishing a reliable diagnosis based on the analysis of such characteristics.

Compliance with Ethical Standards

Conflict of interests Joana Rocha declares that she has no conflict of interest. António Cunha declares that he has no conflict of interest. Ana Maria Mendonça declares that she has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Badrinarayanan, V., Kendall, A., and Cipolla, R., SegNet: A deep convolutional encoder-decoder, architecture for image segmentation. arXiv:1511.00561 [cs], 2015.
2. Dashtbozorg, B., Mendonça, A. M., and Campilho, A., Optic disc segmentation using the sliding band filter, Vol. 56. https://linkinghub.elsevier.com/retrieve/pii/S0010482514002832, 2015.
3. Do Nhu, T., Joo, S. D., Yang, H. J., Taek Jung, S., and Kim, S., Knee Bone Tumor Segmentation from radiographs using Seg-Unet with Dice Loss, 2019.
4. Dubois, D., Hájek, P., and Prade, H., Knowledge-driven versus data-driven logics. Journal of logic, Language and Information, pp. 65–89. https://link.springer.com/article/10.1023/A:1008370109997, 2000.
5. van Ginneken, B., Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiol. Phys. Technol.* 10(1):23–32, 2017.
6. Kingma, D. P., and Ba, J., Adam: A method for stochastic optimization. arXiv:1412.6980, 2014.

7. Kumar, P., Nagar, P., Arora, C., and Gupta, A., U-SegNet,: Fully Convolutional Neural Network based Automated Brain tissue segmentation Tool. arXiv:1806.04429 [cs], 2018.

8. LeCun, Y., Bengio, Y., and Hinton, G., Deep learning. *Nature* 521(7553):436–444, 2015. https://doi.org/10.1038/nature14539. http://www.nature.com/articles/nature14539.

9. Markel, D., Caldwell, C., Alasti, H., Soliman, H., Ung, Y., Lee, J., and Sun, A., Automatic segmentation of lung carcinoma using 3d texture features in 18-fdg pet/ct. International journal of molecular imaging 2013, 2013.

10. Quelhas, P., Marcuzzo, M., Mendonca, A. M., and Campilho, A., Cell nuclei and cytoplasm joint segmentation using the sliding band filter. *IEEE Trans. Med. Imaging* 29(8):1463–1473, 2010. https://doi.org/10.1109/TMI.2010.2048253. http://ieeexplore.ieee.org/document/5477157/.

11. Ronneberger, O., Fischer, P., and Brox, T., U-Net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*. pp. 234–241. Springer, 2015.

12. Roth, H. R., Shen, C., Oda, H., Oda, M., Hayashi, Y., Misawa, K., and Mori, K., Deep learning and its application to medical image segmentation. arXiv:1803.08691 [cs], 2018.

13. Shakibapour, E., Cunha, A., Aresta, G., Mendonça, A. M., and Campilho, A., An unsupervised metaheuristic search approach for segmentation and volume measurement of pulmonary nodules in lung ct scans. *Expert Syst. Appl.* 119:415–428, 2019.

14. Torre, L. A., Siegel, R. L., and Jemal, A., Lung cancer statistics. In: Ahmad, A., and Gadgeel, S. (Eds.) *Lung Cancer and Personalized Medicine*, Vol. 893, pp. 1–19. Cham: Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-24223-1_1.