Comparison of Conventional and Deep Learning Based Methods for Pulmonary Nodule Segmentation in CT Images

Joana Rocha[1,2(✉)], António Cunha[2,3], and Ana Maria Mendonça[1,2]

[1] Faculdade de Engenharia, Universidade do Porto,
FEUP Campus, Dr. Roberto Frias 4200-465, Porto, Portugal
joana866rocha@gmail.com
[2] INESC TEC – INESC Technology and Science,
FEUP Campus, Dr. Roberto Frias 4200-465, Porto, Portugal
[3] Universidade de Trás-os-Montes e Alto Douro,
Quinta de Prados, 5001-801 Vila Real, Portugal

Abstract. Lung cancer is among the deadliest diseases in the world. The detection and characterization of pulmonary nodules are crucial for an accurate diagnosis, which is of vital importance to increase the patients' survival rates. The segmentation process contributes to the mentioned characterization, but faces several challenges, due to the diversity in nodular shape, size, and texture, as well as the presence of adjacent structures. This paper proposes two methods for pulmonary nodule segmentation in Computed Tomography (CT) scans. First, a conventional approach which applies the Sliding Band Filter (SBF) to estimate the center of the nodule, and consequently the filter's support points, matching the initial border coordinates. This preliminary segmentation is then refined to include mainly the nodular area, and no other regions (e.g. vessels and pleural wall). The second approach is based on Deep Learning, using the U-Net to achieve the same goal. This work compares both performances, and consequently identifies which one is the most promising tool to promote early lung cancer screening and improve nodule characterization. Both methodologies used 2653 nodules from the LIDC database: the SBF based one achieved a Dice score of 0.663, while the U-Net achieved 0.830, yielding more similar results to the ground truth reference annotated by specialists, and thus being a more reliable approach.

Keywords: Computer-aided diagnosis · Conventional ·
Deep Learning · Lung · Nodule · Segmentation · Sliding Band Filter ·
U-Net

1 Introduction

Pulmonary nodules can be associated with several diseases, but a recurrent diagnosis is lung cancer, which is the main cause of cancer death in men and the second cause in women worldwide [10]. For this reason, providing an early detection and diagnosis to the patient is crucial, considering that any delay in cancer

detection might result in lack of treatment efficacy. The advances of technology and imaging techniques such as computed tomography (CT) have improved nodule identification and monitoring. In a Computer-Aided Diagnosis (CAD) system, segmentation is the process of 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. Data imbalance also poses a challenge, as in a CT scan less than 5% of the voxels belong to these lesions [11].

In biomedical image analysis, early methods (generally described as conventional) consisted of following a sequence of image processing steps (e.g. edge/line detectors, region growing) and mathematical models [4]. Among other conventional techniques, lesion detection and segmentation often imply the use of filters; e.g. the Sliding Band Filter can be used to develop an automated method for optic disc segmentation [2] and cell segmentation [6]. Such filter also proved to perform better than other local convergence index filters in pulmonary nodule detection [5].

Afterwards, the idea of extracting features and feeding them to a statistical classifier made supervised techniques become a trend. More recently, the trend is to use Deep Learning to develop models that are able to interpret what features better represent the data, but these require a large amount of annotated data, and have large computational cost. Medical imaging commonly relies on Convolutional Neural Networks (CNNs) [3]. For example, Fully Convolutional Networks are often applied for medical imaging segmentation [8], including encoder-decoder structures such as the SegNet, which are frequently used in semantic segmentation tasks [1]. The U-Net is a particular example of an encoder-decoder network for biomedical segmentation [7].

This work aims to precisely segment pulmonary nodules using a conventional approach, based on the Sliding Band Filter, and a Deep Learning based approach, more specifically the U-Net, and therefore evaluate and compare the performance of both methodologies.

2   Local Convergence Filters and the Sliding Band Filter

Local Convergence Filters (LCFs) estimate the convergence degree, C, 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. LCFs aim to maximize the convergence index at each image point, which is calculated 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 that connects (k, l) to P. The overall convergence is obtained by averaging the individual convergences at all M points in R, as written in Eq. 1, taken from [2].

$$C(x, y) = \frac{1}{M} \sum_{(k,l) \in R} \cos \theta_i(k, l) \tag{1}$$

LCFs perform better than other filters because they are not influenced by gradient magnitude, nor by the contrast with the surrounding structures. Being a member of the LCFs, the SBF also outputs a measure which estimates the degree of convergence of the gradient vectors. However, the position of the support region, which is a band of fixed width, is adapted according to the direction and the gradient orientation. The SBF studies the convergence along that band, ignoring the gradient's behaviour at the center of the object, which is considered irrelevant for shape estimation. Such feature makes this filter more versatile when it comes to detecting different shapes, even when they are not perfectly round, because the support region can be molded.
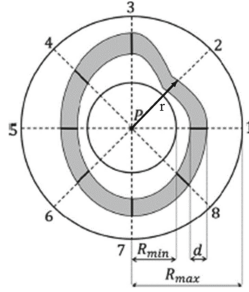

Fig. 1. Schematics of the Sliding Band filter with 8 support region lines (dashed lines, $N = 8$). Retrieved from [2].

The SBF searches each one of the N radial directions leading out of P for the position of the band of fixed width d that maximizes the convergence degree, as represented in Fig. 1. The search is done within a radial length that varies from a minimum ($Rmin$) to a maximum ($Rmax$) values, and so the filter's response is given by Eq. 2, where $\theta_{i,m}$ represents the angle of the gradient vector at the point m pixels away from P in direction i [2]. The coordinates of the band's support points $(X(\theta_i), Y(\theta_i))$ are obtained using Eq. 3, assuming that the center of the object is (x_c, y_c), and r_{shape} corresponds to the radius in direction i [9].

$$SBF(x,y) = \frac{1}{N} \sum_{i=0}^{N-1} \max_{Rmin \leq r \leq Rmax} \left[\frac{1}{d} \sum_{m=r-\frac{d}{2}}^{r+\frac{d}{2}} cos\theta_{i,m} \right] \quad (2)$$

$$\begin{cases} X(\theta_i) = x_c + r_{shape}(i) \times \cos(\theta_i) \\ Y(\theta_i) = y_c + r_{shape}(i) \times \sin(\theta_i) \\ r_{shape}(i) = \underset{Rmin \leq r \leq Rmax}{argmax} \left[\frac{1}{r} \sum_{m=r-\frac{d}{2}}^{r+\frac{d}{2}} cos\theta_{i,m} \right] \end{cases} \quad (3)$$

3   U-Net

The U-Net is a Fully Convolutional Network specially developed for biomedical image segmentation, thus requiring a smaller amount of parameters and computing time [7]. It resorts to data augmentation to simulate realistic elastic deformations, the most common variations in tissue, and so this network is able to work with less training data and still achieves a great performance.

The network, which is an example of an encoder-decoder architecture for semantic segmentation, includes a contracting path and an expansive path. The contracting path is also known as encoder, and can be seen as a common convolutional network with several convolutional layers, each followed by a ReLU and a max pooling layer. This path downsamples the input image into feature representations with different levels of abstraction. In other words, information about the features is gathered, while spatial information is shortened.

On the other hand, the expansive path takes advantage of a series of upconvolutions and concatenations, and thus the feature and spatial knowledge from the contracting path is associated with higher resolution layers. This causes the expansive path to be symmetrical to the contracting one, resulting in a U-shaped network. The precise segmentation is achieved with upsampling (resolution improvement) combined with high resolution features to localize the object.

4   Methodology

The following algorithms were applied on nodule candidates 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 method will be explained for a single plane (Fig. 2a for the conventional approach and Fig. 3a for the Deep Learning based one).

4.1   Conventional Approach

The SBF is first applied to get a better estimation of the nodule's center coordinates. Considering that most nodules have an overall uniform intensity, the nodules' images were processed by truncating any intensities much higher and lower than the nodule's. To do so, the nodule's average intensity was determined by calculating the mean of a matrix centered in the image. These steps result in a truncated mask, where there is already a very primitive segmentation (Fig. 2b) involving a low computational cost, which now needs substantial refinement.

The original nodule image, as well as the truncated nodule mask, are fed to the SBF and the filter's response in each pixel around the center of the image is calculated. The estimated nodule's center corresponds to the pixel which maximizes the response of the filter. With those coordinates, the SBF then evaluates the corresponding set of support points, returning the N border coordinates
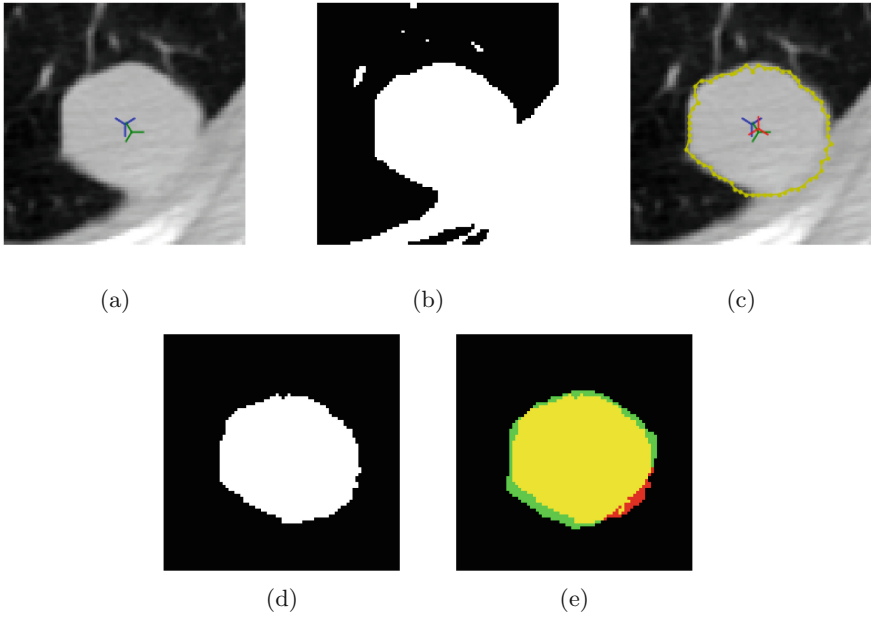
Fig. 2. Exemplification of the conventional methodology steps, where the blue mark is the center of the image, the green mark is the ground truth center of the nodule, and the red mark is the estimated center of the nodule. (Color figure online)

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. An outlier attenuation/removal step was implemented, minding the distance between consecutive border coordinates, and afterwards a binary mask with the initial SBF segmentation is created.

To further refine the segmentation and specifically select the nodular area, only the intersection of the SBF segmentation mask and the truncated nodule mask is considered, thus eliminating unwanted regions. 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. This can be done by eliminating from the mask all regions that do not encompass the center of the image, as the nodule is always centered. After this step, the final segmentation mask is achieved, as exemplified in Fig. 2d.

4.2   Deep Learning Based Approach

The Deep Learning algorithm presented in this work was implemented using *Keras*, with a *TensorFlow* backend. The 2D images are imported and split into training, validation, and test sets. A condition was added to ensure that a nodule belongs exclusively to one of the sets, as it would not make sense to train and test the model using the same nodules. This way, the set of 2653 nodules was first split into training and test sets (80%–20%, respectively), and then 20% of the training set was used as validation set. 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, as well as grayscale variations.

The architecture of the U-Net was kept as proposed in [7], where the contracting path includes four sets of 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. The number of feature channels is doubled with each downsampling phase, which is repeated four times as mentioned above. Two 3×3 convolutions with ReLU, batch normalization, and dropout, are present at the end of this path, creating a bottleneck before the following path. The expansive path is comprised of four blocks which repeatedly have a deconvolution layer with stride equal to 2 (reducing the number of feature channels by half), a concatenation with the corresponding cropped feature map from the contracting path, and two 3×3 convolutions, each followed by ReLU with batch normalization. In order to achieve the segmentation result (pixel-wise classification task with two classes), the endmost layer includes a 1×1 convolution, using softmax activation. The pixel-wise probabilities face a 50% threshold to decide whether a pixel is nodular or not.

The network was trained with the Adam optimizer, to achieve faster a stable convergence. 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 model is guided by two evaluation metrics: accuracy and Jaccard Index.

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 is the default provided in the original paper for the Adam optimizer, and will be reduced by a factor of 0.1 (new learning rate = learning rate $\times$ 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), allowing a maximum of 100 epochs. 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. Figure 3b is the segmentation achieved by the U-Net for the nodule in Fig. 3a.
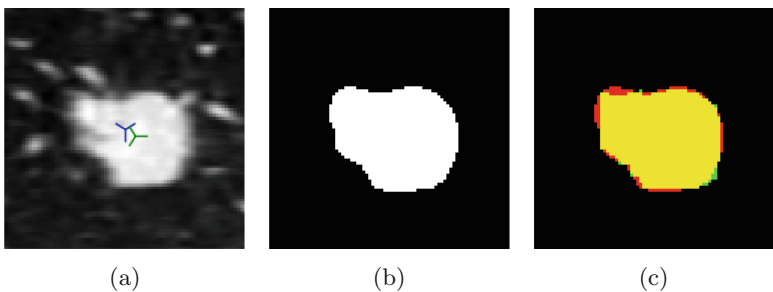
(a)   (b)   (c)

Fig. 3. Exemplification of the U-Net results, where the blue mark is the center of the image, the green mark is the ground truth center of the nodule. (Color figure online)

5   Results

The methods were evaluated on 2653 nodules, using as ground truth the segmentation masks from the LIDC database, which is publicly available and consists of lung cancer screening thoracic CT scans from 1010 patients. The results for the conventional approach, exhibited in Table 1, were obtained with the SBF parameter values N = 64, d = 7, Rmin = 1 and Rmax = 25, which were established empirically to maximize the algorithm's performance. This method achieved a Dice score of 0.663, while having Precision and Recall values of 0.710 and 0.732, respectively. On the other hand, the U-Net test set predictions achieved a loss value of 0.172, accuracy of 0.992, and Dice score of 0.830, while having Precision and Recall values of 0.792 and 0.898. The U-Net model exhibited fast convergence (after 14 epochs), and did not overfit to the training data, considering the validation loss is similar to the training loss (Fig. 4). The SBF needed approximately 5 hours to run, while the U-Net had a reasonable training time of roughly 8 hours on a NVidia GeForce GTX 1080 GPU (8 GB).

Table 1. Evaluation metrics achieved by the conventional and Deep Learning based methods.

	Dice coefficient	Precision	Recall
SBF	0.663	0.710	0.732
U-Net	0.830	0.792	0.898

To illustrate the performance of the methods presented in this work, a comparison plot was created (Figs. 2e and 3c), in which the green pixels belong exclusively to the ground truth mask, the red pixels belong exclusively to the achieved result using the proposed method, and the yellow pixels belong to both - meaning that the yellow pixels mark the correct predictions made by the algorithm. Figure 5 compares the performance of each algorithm, using these plots.
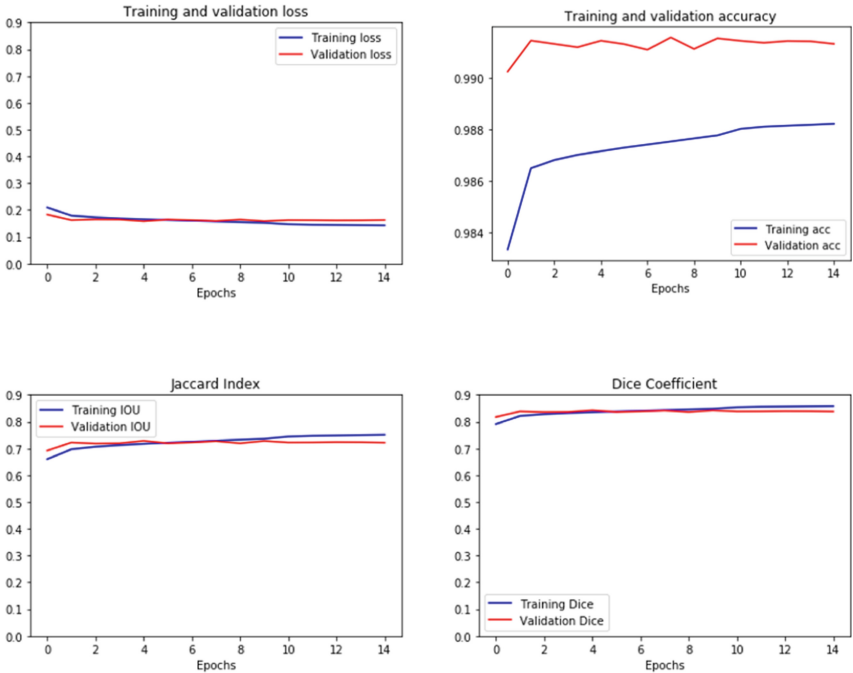
Fig. 4. Training and validation results of the U-Net model.

Considering that the SBF was evaluated on a set of 2653 nodules, and the U-Net was evaluated on 531 nodules (20% of the database), the following analysis is done minding the performance of both methodologies for the same set of 531 nodules, randomly selected in the U-Net algorithm. The proposed conventional approach exhibits a highly satisfactory performance when dealing with well-circumscribed solid nodules. The nodules that have a pleural tail generally have the thin structure ignored by the algorithm, which does not include it in the segmentation mask, while the specialists consider the tail as part of the nodule. In spite of vascularized nodules entailing an inherent difficulty when it comes to distinguishing the nodule from the attached vessels, the SBF based approach is frequently able to separate them and create a mask which does not include the non-nodular structures. The algorithm's performance is also satisfying when dealing with nodules whose intensities vary within their border (e.g. calcific and cavitary nodules), as it is able to ignore the cavities/calcific regions during the segmentation process. The main flaws of the algorithm appear when dealing with juxtapleural nodules, since it often does not know where the nodule ends and the pleura begins, thus only being able to estimate the boundary to some extent. Overall, the less satisfying results are mainly due to the unexpected shape of the nodule, or because the nodule does not have a clear margin (e.g. non-solid nodules/ground glass opacities).
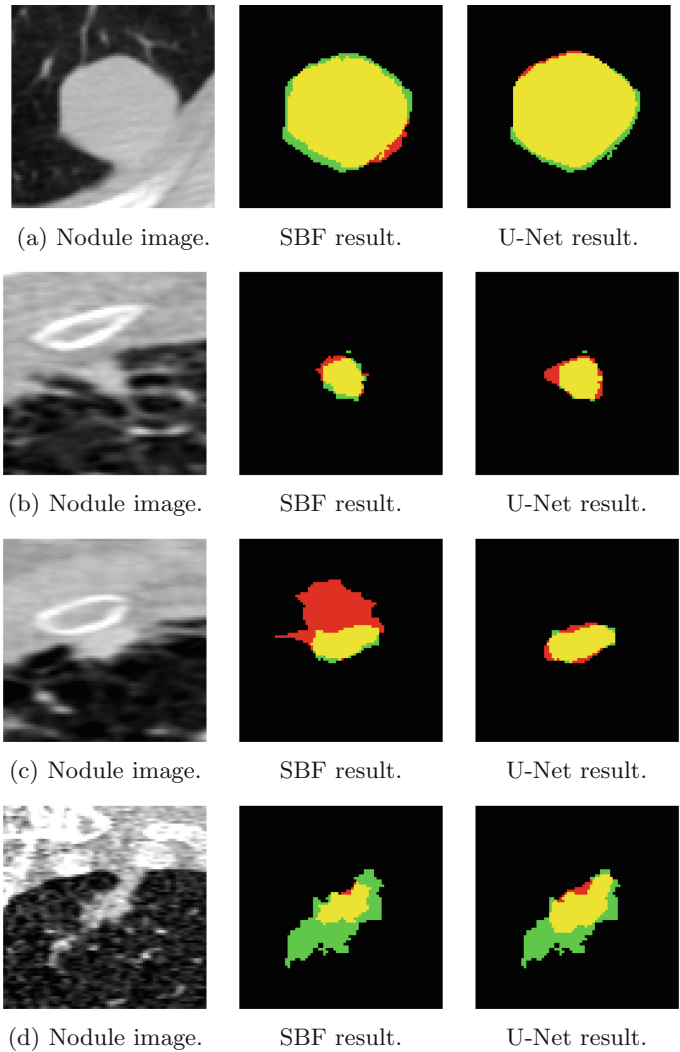
(a) Nodule image.   SBF result.   U-Net result.

(b) Nodule image.   SBF result.   U-Net result.

(c) Nodule image.   SBF result.   U-Net result.

(d) Nodule image.   SBF result.   U-Net result.

Fig. 5. Comparison between methodologies: 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.

Similarly to the SBF based approach, the U-Net is able to clearly segment well-circumscribed solid nodules, as well as cavitary, calcific, and vascularized nodules. This network also agrees with the previous approach when it comes to nodules with pleural tail, in the sense that it does not encompass the tail in the segmentation mask, unlike suggested by the specialists. However, the U-Net certainly outperforms the conventional approach, as it is able to deal nearly

perfectly with juxtapleural nodules, achieving very accurate segmentation results in such cases. It also achieves better results when presented with non-solid nodules/ground glass opacities, in spite of their inherent segmentation complexity, and so the results for this type of nodules may also be slightly different from the ground truth. Even though the algorithm demonstrates a certain degree of dexterity when dealing with irregular shaped nodules in comparison to the SBF, it still produces less accurate results when presented with these situations. However, the U-Net segmentation results are in general very similar to the ones given by the specialists, hence its great performance in comparison to the SBF based approach. In some cases, the U-Net is even able to achieve a more uniform detailed segmentation in comparison to the specialists', which may be justified by its pixel-wise perspicacity.

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 capability of the segmentation process. 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 approach, 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.

6   Conclusion

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 with the LIDC database. On the other hand, the Deep Learning based approach achieves a Dice score of 0.830, using the U-Net. Both performances are impaired by irregular shaped/non-solid nodules, and in the case of the convencional approach by juxtapleural nodules as well, and so future work includes the refinement of the methods to deal with these particular challenges. However, they are promising segmentation tools for well-circumscribed, solid, cavitary, calcific, and vascularized nodules. This way, the comparison of the conventional and Deep Learning based approaches explored the advantages and disadvantages of each technique, establishing the U-Net as the most efficient method in this case - particularly efficient for obvious lesions, and able to overcome to a certain extent the high variability of nodular structures. Consequently, the satisfactory segmentation results achieved by the U-Net 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 to establish a reliable diagnosis based on the analysis of such characteristics.

References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deepconvolutional encoder-decoder architecture for imagesegmentation, November 2015. arXiv:1511.00561 [cs]. http://arxiv.org/abs/1511.00561
2. Dashtbozorg, B., Mendonça, A.M., Campilho, A.: Optic disc segmentation using the sliding band filter. Comput. Biol. Med. **56**, 1–12 (2015). https://doi.org/10.1016/j.compbiomed.2014.10.009. https://linkinghub.elsevier.com/retrieve/pii/S0010482514002832
3. Jiang, F., et al.: Medical image semantic segmentation based on deep learning. Neural Comput. Appl. **29**(5), 1257–1265 (2018). https://doi.org/10.1007/s00521-017-3158-6. https://link.springer.com/10.1007/s00521-017-3158-6
4. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005. https://linkinghub.elsevier.com/retrieve/pii/S1361841517301135
5. Pereira, C.S., Mendonça, A.M., Campilho, A.: Evaluation of contrast enhancement filters for lung nodule detection. In: Kamel, M., Campilho, A. (eds.) ICIAR 2007. LNCS, vol. 4633, pp. 878–888. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74260-9_78
6. Quelhas, P., Marcuzzo, M., Mendonca, A.M., 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. https://ieeexplore.ieee.org/document/5477157/
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
8. Roth, H.R., et al.: Deep learning and its application to medical image segmentation, March 2018. arXiv:1803.08691 [cs]. https://doi.org/10.11409/mit.36.63. http://arxiv.org/abs/1803.08691
9. Shakibapour, E., Cunha, A., Aresta, G., Mendonça, A.M., Campilho, A.: An unsupervised metaheuristic search approach for segmentation and volume measurement of pulmonary nodules in lung CT scans. Exp. Syst. Appl. **119**, 415–428 (2019)
10. Torre, L.A., Siegel, R.L., Jemal, A.: Lung cancer statistics. In: Ahmad, A., Gadgeel, S. (eds.) Lung Cancer and Personalized Medicine. AEMB, vol. 893, pp. 1–19. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-24223-1_1
11. Wang, S., et al.: Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. Med. Image Anal. **40**, 172–183 (2017). https://doi.org/10.1016/j.media.2017.06.014. https://linkinghub.elsevier.com/retrieve/pii/S1361841517301019