Automatic Detection and Segmentation of Pulmonary Lesions on CT scans using Deep Convolutional Neural Networks

João Francisco Lourenço Borges de Sá Carvalho
jbsacarvalho@ist.utl.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2018

Abstract

Early detection of lung cancer has been shown to significantly improve patient survival. Apart from lesion detection, tumour segmentation is critical for developing *radiomic* signatures. In this work, we propose a novel hybrid approach for lung lesion detection and segmentation on CT scans, where the segmentation task is assisted by prior detection of regions containing lesions. For the detection task, we introduce a 2.5D *residual deep convolutional neural network (CNN)* working in a sliding-window fashion, whereas segmentation is tackled by a modified *residual U-Net* with a weighted-dice plus cross-entropy loss. Experimental results on the *LIDC-IDRI* dataset and on the lung tumour task dataset, within the *Medical Segmentation Decathlon*, show competitive detection performance of the proposed approach (0.902 recall) and superior segmentation capabilities (0.709 dice score). These results confirm the high potential of simpler models, with lower hardware requirements, thus of more general applicability.
Keywords: Radiomics, lung cancer, deep learning, convolutional neural networks, residual connections

1. Introduction

1.1. Lung Cancer and Radiomics

Lung cancer is the most prevalent cancer in men and the third most common cancer in women, being the worldwide leading cause of cancer death (∼2 million cases per year with a 60% mortality rate). It is well known that early detection and subsequent follow up play a key role in improving treatment outcomes, including survival rates [27]. Screening trials have been implemented using low-dose thoracic *computer tomography* (CT) scans, and have now become the standard practice for early nodule detection. Specifically, the *National Lung Screening Trial* has shown that patients who are screened with thoracic CT have a 15-20% increase in survival rate when compared to patients who underwent standard chest X-ray imaging [30]. In parallel, a variety of software has been developed and approved by the *Food and Drug Administration* (FDA) aiming at improving workflow efficiency and performance in early detection of lung lesions on thoracic CT scans, namely by focusing on detecting and characterizing nodules [5].

The large datasets arising from those efforts together with the advent of high-performance computing have enabled the development of the field of *radiomics*. It is now possible to extract high-dimensional features from medical images, the so-called *radiomic signatures*, which, combined with machine learning methods, are used to predict cancer-related outcomes [7]. The segmentation of *volumes of interest* (VOI) around lesions, and possibly sub-regions (*i.e.*, habitats) within the tumour, is critical in the radiomics pipeline [20], as radiomic features are extracted from the VOI. Therefore, the development of semi- or fully-automated methods is of great value in clinical application, to avoid the intrinsic inter-reader variability of human segmentations [21, 22]. Moreover, manual detection and segmentation of lung nodules/tumours require a high amount of tedious, expensive, and time-consuming work by human experts (radiologists) [7, 23]. Automatic tools can thus be extremely useful in daily clinical work, where the tasks of detection of pulmonary lesions and segmentation of lung tumours both require an undesirably high level of manual processing [22]. These tools promise to play a pivotal role in increasing the quality of early cancer detection and diagnosis.

1.2. Deep Learning for Medical Image Analysis

Generic object detection and image segmentation are old problems, with an extremely vast literature, and to which many different image processing tools have been applied. A recent trend in these tasks is to resort to machine learning (ML), which allows

designing algorithms that "learn" to solve the problem from examples provided by experts; a particular class of ML techniques, known as *deep learning* (DL) [11, 17, 20], has been found particularly effective in solving demanding image analysis tasks, such as image classification, object detection, and image segmentation. These tools are expected to substitute or complement human-intensive procedures, allowing for the fast analysis of large collections of images and reducing human-caused variability [23]. Technological advances, namely high-end graphical processing units (GPUs) and central processing units (CPUs), and new learning algorithms, together with the increasing availability of large amounts of data, has led DL to unprecedented success and many are the studies that use such technology for several tasks medical imaging [26].

Deep convolutional neural networks (DCNNs) are a particular type of DL architecture that has shown promising results in several medical imaging tasks [11, 26]. Yet, these models face several major challenges within this field [11]. Training DCNNs requires large amounts of labelled data, which is inherently difficult to obtain in the medical domain, due to the limited availability in both expert annotation and data of lesions throughout their several stages. This is a crucial aspect, since the use of insufficient amounts of training data is known to increase the risk of *overfitting* [7], *i.e.*, of obtaining a model that performs well on the training data, but generalizes poorly when applied to unseen cases. Furthermore, the sheer amount of variation between acquired images [20], due both to the inherent disparity in the acquisition process, as well as the multiple possible medical imaging machines used for the acquisition [7, 20], lead to a high probability of overfiting to specific features that are not shared by all images in the dataset [7, 11]. To overcome this problem, several strategies have been proposed: data augmentation [11, 24], where more data is artificially created by modifying existing data; *dropout*, where each hidden neuron has a probability of being set to zero, therefore not contributing for neither forward or backpropagation in each iteration through the training set [29]; other regularization techniques, such as ℓ_1 and/or ℓ_2 regularization of the weights of the DCNN.

In the particular case of medical imaging, it is common to deal with 3D images (volumes), which is usually the case in CT scans. However, several approaches have only relied on models which use information from single slices of the original volume to perform their tasks, therefore ignoring the intrinsic 3D nature of the data [20]. Recently, there has been increased interest in using 2.5D models, which are based on multiple slices of the image data [11], as well as in using pure 3D architectures, where the convolution operations of the DCNN are performed on the 3D data [6]. Although there are clear advantages in using 3D DCNNs, there are also severe limitations, due to very high GPU memory requirements and higher potential for overfitting [15]. Finally, DCNNs with *skip connections* (also known as *residual connections*) have been extensively used for image classification and object detection [34, 13], and have also played a significant role in improving the segmentation performance in several medical imaging applications [1, 9].

1.3. Lung Nodule Detection and Segmentation

Several DCNNs have been applied to lung nodule detection with very good results [8, 32, 2]. However, most of the architectures used had been originally developed for object detection on colour images, a task characterized by an extremely large number of classes [34]. Therefore, those architectures are very hard to train (*e.g.*, they include several hyper-parameters that need to be fine-tuned), requiring very long training times involving huge amounts of computation [34, 14]. To adapt those architectures to the different requirements of medical imaging tasks, several works have proposed to add deconvolution layers to extend the resolution of the feature map [8, 32], as well as using ensemble methods within a boosting approach [25] for the classification sub-network, which may also lead to increased overfitting to the training set [32].

The use of DL for lesion segmentation has also recently flourished [23], specifically using models based on the U-Net [24], which is a DCNN with two components: an encoder path, which produces a high-density low-resolution feature map; a decoder path, which translates the feature map into the final segmentation. Image segmentation is an extremely challenging task when applied to tumours, not only due to the high morphological and phenotypic heterogeneity of lesions, but also to the small amount of labelled data available (which is even more critical in the presence of that high heterogeneity). Recent work has used 3D architectures with input size reduction by using cascaded models [15]. Other recent approaches have also tried to perform a voxel-by-voxel classification using DCNNs; however, this has an extremely high computational cost [31].

1.4. Proposed Approach

To move beyond the limitations of prior DCNN-based models, this paper proposes a novel approach, where the task of segmentation is assisted by first detecting regions that contain lesions. For that lesion detection task, we develop a new 2.5D residual DCNN applied to the CT scan, working in a sliding-window fashion, therefore reducing the problem to a patch classification task. Furthermore, we develop a modified 2D residual U-Net for the tumour seg-

mentation task, for the training of which we adopt a weighted-dice plus cross-entropy loss function tailored to the task at hand. Moreover, we present a full pipeline for image pre-processing, as well as data augmentation and example re-sampling to improve the quality of the training data.

2. Methods

2.1. Model Architecture

In this work, we propose the LungSD-Net, a novel hybrid architecture that first performs the detection of lesions followed by the segmentation of the detected lesions (see Figure 2). The detection is performed by first reducing the problem to a patch classification task. This is possible by first transforming the input CT scan, $X \in \mathbb{R}^n$, where n is number of voxels in the image, to a set of patches $P = \{p_1, ..., p_n\}$, such that each $p_i \in \mathbb{R}^{N \times N \times Z}$ is a three dimensional patch with $N \times N \times Z$ voxels. The detection model, a 2.5D residual DCNN, is then responsible for accurately mapping from each individual p_i to a binary classification, c_i, depending on having or not a lesion. The second stage of the proposed hybrid architecture performs the segmentation of the positively classified patches, using a 2D residual U-Net.

Table 1: Architecture details of the detection network, divided into *convolutional* (Conv), *max pooling* (M. Pool.), and *fully connected* (FC) layers.

Layer	**Conv layers**		
	kernel size	nb of kernels	stride
Conv 1	3×3	64	1
Conv 2	3×3	64	1
Conv 3	3×3	64	1
M. Pool. 1	2×2	-	2
Conv 4	3×3	128	1
Conv 5	3×3	128	1
Conv 6	3×3	128	1
M. Pool. 2	2×2	-	2
Conv 7	3×3	256	1
Conv 8	3×3	256	1
Conv 9	3×3	256	1
M. Pool. 3	2×2	-	2
	FC layers		
	nb of units	act. function	
FC 1	512	leaky-ReLu	
FC 2	512	softmax	

2.1.1. Detection Network

Detection follows a sliding-window approach with a step of 15 voxels, where three 64x64 patches are extracted from the CT image and fed to the network, which classifies the middle slice as as having or not a lesion. Figure 2 depicts the detection network, which is comprised of two parts: three convolutional layer blocks that extract features from the data; two fully connected layers that perform classification. Convolutions and pooling are all applied with appropriate padding, with convolutions having 3x3 kernels with stride 1, and max-pooling using 2x2 kernels and stride 2.

One of the main goals of this architecture is to avoid overfitting, therefore each convolutional layer block has a residual connection between the first and the last layer [9], followed by dropout and max-pooling. We applied drop-out with a retaining rate of 0.9 for the convolutional layers and 0.5 for the fully connected layers, due to the increase in accuracy reported in computer vision tasks with this configuration [29]. The output layer has a binary softmax activation function, and every other layer is equipped with a *leaky-ReLu* to avoid the vanishing gradient problem[33, 19]. In order to balance accuracy and model size, only two fully connected layers were used. The details of the architecture can be seen in Table 1.

2.1.2. Segmentation Network

Our adaptation of the U-Net aims at optimising its architecture for the task of segmenting the detected lesions. As the input size for this network follows the input size of the detection network, our U-net was reduced to three layers in both the encoding and decoding paths (Figure 2), instead of four as in the original work. Leaky-ReLU [33, 19] was used as activation function for every layer, except the last one, where softmax was used. Each block of convolutional layers was also equiped with a residual connection between the first and the second layer, followed by max-pooling. Kernel sizes and strides, as well as padding, follow the original U-Net arquitecture [24].

2.2. Datasets

The two datasets used in this work are next described.

The *Lung Image Database Consortium - Image Database Resource Initiative (LDC-IDRI)*, used for the detection task, is a public and free dataset that contains 1018 CT scans, both standard and low-dose CT scans, all annotated by four experienced radiologits in a two-phase blind annotation process [3]. In order to increase the quality of the dataset, all scans with slice thickness greater than 3mm, inconsistent slice spacing, or missing slices, were excluded, yielding 888 slides. Taking into account recent metrics for follow-up of detected pulmonary lesions, only nodules larger than 6mm were included [4]. The final distribution of the nodule size is shown in Figure 1, where the diameter of the lesions ranges from 6mm to 32mm with high deviation towards smaller nodules. This dataset was split

into three parts: 80% of the patients constitute the training set and the remaining were equally split into validation and test sets. From each patient, a variable number of patches was extracted for training and validation, maintaining a balanced distribution of classes, and totalling 1908 patches from 704 scans for the training set, and 344 patches from 88 scans for the validation set. Patches for the test set were sampled following a sliding-window approach from a total of 88 scans, with each CT scan producing 1156 patches per slice with a highly unbalanced class distribution.

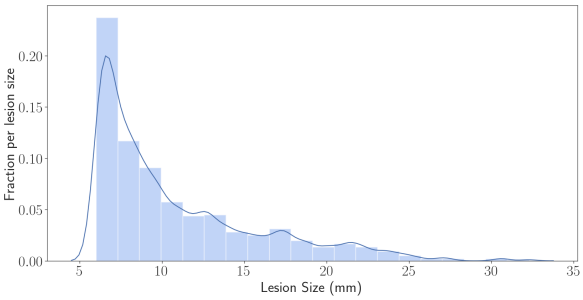


Figure 1: Final size distribution of the lesions in the training set of the detection network.

The dataset used to train and evaluate the segmentation model (*Lung task from the Decathlon competition*), contains 64 labelled CT scans with singular, non-small lesions, annotated by an expert thoracic radiologist on a representative cross-section [28]. As the segmentation of pulmonary lesions is only relevant in the context of the radiomics pipeline for larger ones, tumours smaller than 25mm were excluded. The final dataset totalled 32 scans, and a variable number of 64×64 patches was extracted from the VOI around the lesions. Each VOI had the same number of slices with and without lesion, the latter being equally distributed above and bellow it. Finally, 5-fold cross-validation (CV) was used, with 21 whole scans used to train the model, and the rest split by assigning 6 examples to the validation set and 7 to the test set.

2.3. Pre-processing

In order to prepare the images for both the detection and segmentation models, several steps were adopted, following the methodology presented next. These are partially outlined in Figure 3.

2.3.1. Lung Extraction

One of the main issues in automatic detection of pulmonary nodules is the false positives outside of the lung region. To avoid this issue, the lungs were extracted from all images by using the masks provided in the LIDC-IDRI dataset. To allow for the detection of *juxta-pleural nodules* – nodules attached to the side-wall of the lung – a small dilation was applied to all the masks. Due to the hybrid approach to the detection and segmentation tasks, only the CT scans used for lesion detection will have the lungs removed. This effort also serves the purpose of reducing the amount of information that needs to be encoded by the model, as all the regions outside of the lung will be easily detected as not containing any lesion. Since the accuracy of the segmentation is not a factor for the model performance (as long as within certain quality criteria that guarantee that all the lung region is included in the mask) the extraction of the lungs could have been made by non-supervised methods.

2.3.2. Resampling

In general, DCNNs are not prepared to handle the voxel spacing heterogeneity of CT images, which arises from different scanners and different acquisition protocols. To overcome this issue, re-sampling was applied to all patients (from both datasets) using second-order B-spline interpolation to the median of the dataset voxel size. The resampling is reported to improve the system's performance by close to 20% in other 2.5D CNN-based systems [2].

2.3.3. Normalization

CT scans have absolute intensity scale, arising from the estimation of the attenuation coefficients. This yields a scale that can go from -1000 to 1000, which may introduce instability when training a CNN. Therefore, the final input images for both networks were volume-level normalised, by using histogram-based normalisation and clipping to the [1, 99] intensity value percentiles.

2.4. Training Procedure

Both models were trained in its entirety. The 2.5D residual DCNN was trained with a binary cross-entropy loss, using the Adam optimizer [16] with batch size 300 and learning rate 3×10^{-4}, whereas the modified U-Net was trained, also using Adam, with batch size 12 and learning rate 10^{-5}. Training was stopped when the loss reached a *plateau*.

The U-Net usually struggles to segment objects occupying small fractions of the total number of voxels/pixels of the input image, as is usually the case of lung nodules or tumours. To overcome this hurdle, which would lead to failure to convergence to a non-zero solution, the network was trained with a combination of a *weighted dice* (w-dice) and *cross-entropy* (CE) losses:

$$L_{\text{total}} = L_{\text{w-dice}} + L_{\text{CE}}. \tag{1}$$

The w-dice loss is a weighted version of the dice loss [15], computed as

$$L_{\text{w-dice}} = -\sum_{k\in\{0,1\}} w_k \frac{\sum_{i\in B} o_i^k g_i^k + \epsilon}{\sum_{i\in B} o_i^k + \sum_{i\in B} g_i^k + \epsilon}, \tag{2}$$

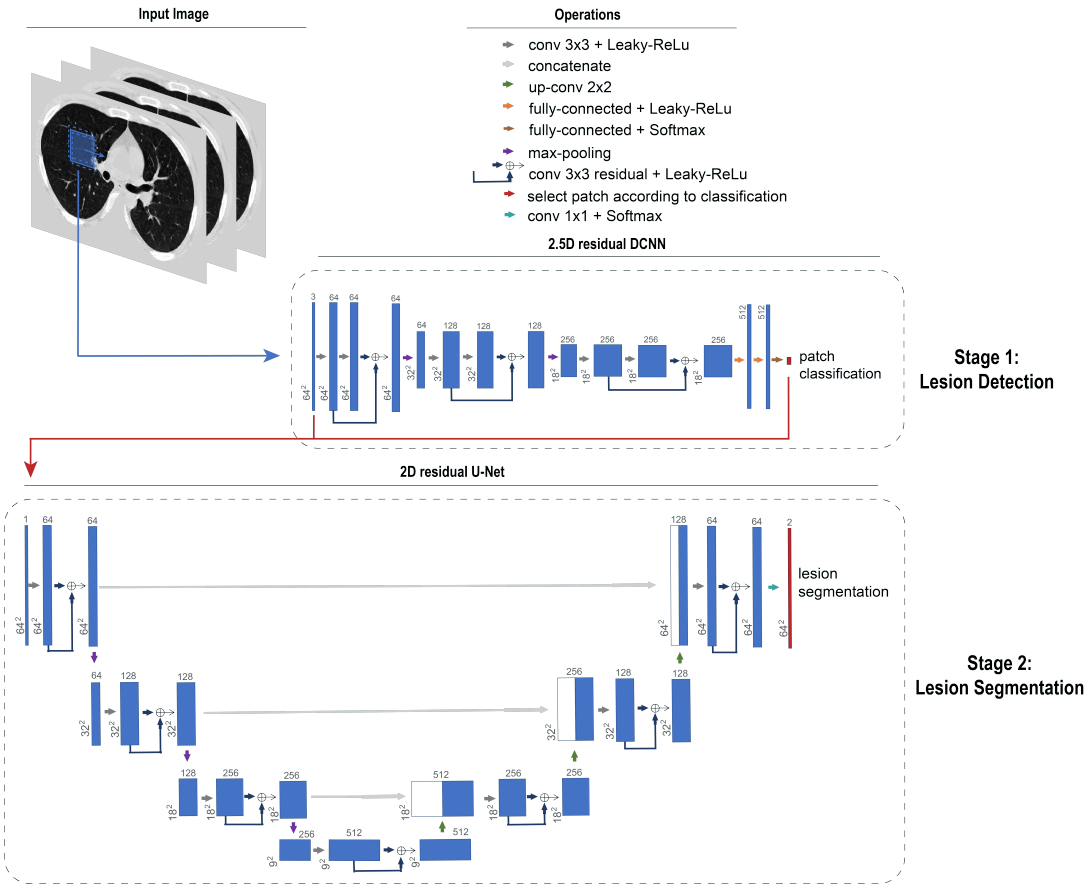


Figure 2: Overview of the LungSD-Net, which integrates the detection and segmentation architectures with a sliding-window approach. Three 64x64 patches are extracted with a shift of 15 voxels, and used as input for the detection model, the 2.5D residual DCNN that classifies the middle patch as having or not a lesion. The patches positively classified are used as input for the second step of the model, the 2D residual U-Net, that yields a final segmentation.

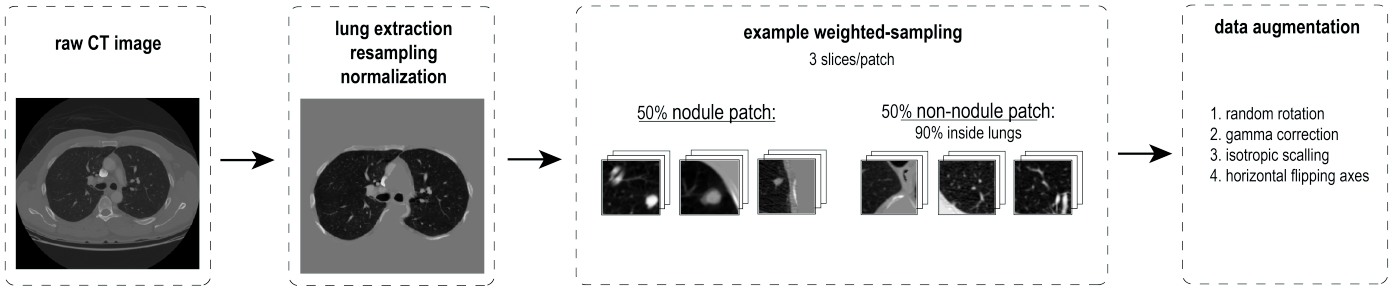


Figure 3: Depiction of the pre-processing and data curation steps used to prepare the training set for the detection model. The raw thoracic CT image was first pre-processed by having the the lungs extracted, followed by resolution resampling to the median of the training-set voxel size and histogram-based normalization. Sets of 3 patches were extracted, maintaining a balanced class distribution and focusing on highly informative patches. The training was optimized by using several data augmentation techniques.

where B is the set of voxels in the training batch, o_i^k is the k-th softmax output for the i-th voxel, and g_i^k is the corresponding one-hot encoding of the ground-truth segmentation; the weights w_0 and w_1 control the contribution of each class to the loss, with $w_0 < w_1$, thus down-weighting the background contribution.

All experiments were run on a machine with an Intel Xeon 1620 CPU and two NVidia Geforce GTX 1080 (totalling 22 GB on-board memory). The networks were implemented using Niftynet, a TensorFlow-based platform which provides a modular DL pipeline with components dedicated to data loading, data augmentation, network architectures, loss functions and evaluation metrics tailored for medical imaging tasks [10]. The best iteration of the model was chosen as that with the best detection accuracy, or best dice score for the segmentation task, on the validation dataset, in order to minimise overfitting.

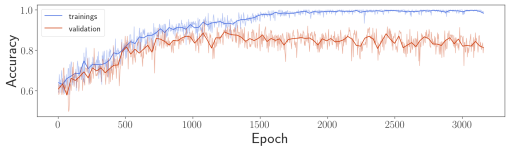
2.4.1. Data Augmentaiton and Weighted-Sampling
Due to the limited amount of training data in both tasks, besides the architecture changes aimed at avoiding overfitting, extensive data curation of training examples was performed. These are outlined in Figure 3. First, weighted sampling of significant examples was applied, with a balanced sampling of examples for training the 2.5D residual DCNN: equal numbers of patches with and without nodules were extracted from each image. Due to the extraction of the lungs from the image, learning the false examples from outside the lungs is a trivial task for the network, so 90% of the negative examples were chosen from within the lungs. A similar strategy was applied to train the 2D residual U-Net with examples selected through weighted sampling where the sampling likelihood of each voxel, and window around it, is proportional to its class frequency. Secondly, several data augmentation schemes were applied, with a focus oin techniques that retained relevant features for the tasks at hand. Random isotropic scaling, and horizontal mirroring were applied to images for both tasks. Random rotations and vertical mirroring was only used for segmentation, as orientation may encode relevant features for detection.

2.4.2. Regularization
Extensive regularization was implemented through penalization terms added to the loss function. For the detection task, the network was first trained for 500 iterations with ℓ_1 regularization, in an attempt to select the weights that contribute the most to the improvement in performance, and to leave the others as zero. This was followed by ℓ_2 regularization until convergence. Both weights were set to 0.1. The modified U-Net was trained with ℓ_2 regularization, with weight set to 0.001.

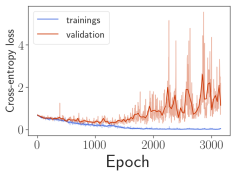
3. Results

3.1. Detection

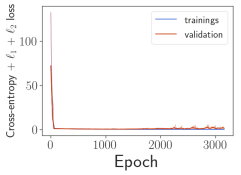
The evaluation of the model during training is showcased in Figure 4. From Figure 4(a), the final model was chosen by maximizing the accuracy in the validation-set. In Figure 4(b), it is possible to note that this model has a minimal cross-entropy loss in the validation-set when the accuracy is maximized, which occurs right before the model starts diverging due to overfitting.



(a) *Accuracy measured during training*



(b) *Cross-entropy loss during training*



(c) *Cross-entropy and weight penalization loss during training*

Figure 4: Metrics assessed during training for the final detection model.

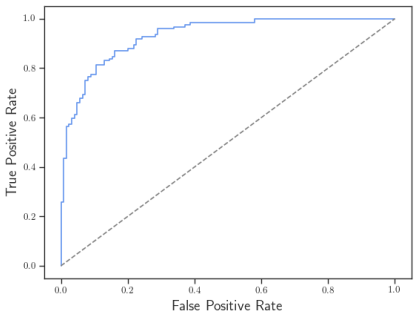


Figure 5: ROC of the lesion detection task (AUC = 0.87)

The ROC curve computed on the test set is shown in Figure 5. From this ROC curve, an AUC of 0.87 was computed. Taking into account the many false examples in the test set, it can be claimed that AUC = 0.87 is a meaningful value for a sliding-window approach. Nonetheless, a true negative rate (TNR) of 0.83 leads to 196.52 false positives per slice, which

is a number comparable to [32], but higher than the 15.0 candidates/slice, reported in [2]. Therefore, it is possible to suggest that the LungsSD-Net would greatly benefit from a false positive reduction model, such as those presented in [32, 2].

In order to assess the impact of several hyperparameters of the 2.5D residual DCNN, as well as the contribution of the data pre-processing, curation, and augmentation, an ablation study was performed, where six different models were trained:

1. *detection model 1*: 2D residual DCNN with no weight penalization, and no drop-out;
2. *detection model 2*: 2D residual DCNN with both ℓ_1 and ℓ_2 regularization, but no drop-out;
3. *detection model 3*: 2D residual DCNN trained with only raw images and no data augmentation;
4. *detection model 4*: 2D residual DCNN trained with pre-processed images, but no data augmentation;
5. *detection model 5*: 2D residual DCNN;
6. *detection model 6*: 2.5D residual DCNN.

The comparison of the models is provided in Table 2, tith the AUC and the TNR evaluated patch-wise, whereas the TPR was computed slice-wise, as the fraction of true posives and the total number of lesions in the test-set. First, it is possible to note that all the performance metrics assessed - AUC, TPR and TNR - increase from the *detection model 1* to the *detection model 2*, and then to the *detection model 5*. This leads to the conclusion that both regularization strategies lead to improved model performance. It is noticeble that the added weight penalization terms largely contribute to the increase in performance, with a relevant benefit in the decrease of false positives. From the comparison of *detection model 3*, *detection model 4*, and *detection model 5* it is possible to conclude that both data augmentation and image pre-processing have an impact in the final models' performance. Specifically, the removal of all pre-processing largely contributes to the decrease in performance of the model, which may be due to the large values of the images' voxels, which are a known cause of instability when training neural networks, as well as the multiple steps to curate the examples for the training of the model. Finally, the increased performance in the final detection model (*detection model 6*), when compared with the one with only one slice as input (*detection model 5*), leads to the conclusion that additional anatomical context leads to more accurate lesion detection. It is also relevant to note that a decrease in performance in the 2D residual CNN is more noticeable in terms of TNR. One possible explanation is that the added context allows the model to better detect common false positives, such as vascular structures, which are frequent lesion's confounders in axial plane of CT scans [2]. This is supported by the evidence that anatomical information, specially in the transverse direction, might encode relevant features for the detection of these structures,

Table 2: Summary of the all the models performance in the test-set for the ablation study. Evaluation metrics are AUC from the ROC curve, true positive rate (TPR), and true negative rate (TNR).

Model	**Evaluation Metrics**		
	AUC	TNR	TPR
detection model 1	0.66	0.645	0.710
detection model 2	0.78	0.734	0.846
detection model 3	0.56	0.548	0.597
detection model 4	0.74	0.718	0.806
detection model 5	0.80	0.766	0.871
detection model 6	**0.87**	**0.830**	**0.901**

In Table 3, we we compare our results with the state of the art for lung lesion detection.

In comparison to much more complex architectures and training methodologies, the proposed, much simpler, 2.5D residual DCNN achieves recall values comparable to the state-of-the art. As demonstrated in the ablation study, regularization and data augmentation largely contributed to the good results herein reported. Also, reducing the problem to the inside of the lungs allowed to benefit from resampling of the training examples, which may have contributed to the ability of the network to learn more relevant features from within the lungs. This may have led to the comparable results achieved as none of the non 3D models resorted to lung extraction, and only one of the 3D models used it as pre-processing step. Moreover, it is important to mention that the 2.5D *Faster RCNN* was also helped by extensive data augmentation as well as positive oversampling. Finally, it is important to note that currently there is a disparity in the size of lesions that are considered for analysis, but that this work followed the parameters set by the *Fleischner Society*[4].

In Figure 6, the comparison between the number of false negatives per lesion size, and the size distribution of examples in training set is presented. Even though the number of nodules drastically decreases in the dataset, the number of positive patch examples decreases with a slower rate, as it is possible to extract a greater number of examples per lesion. As expected, due to a smaller amount of larger examples, most of the undetected lesions co-

Table 3: State-of-the-art 2D detection models comparison.

Model	Recall
3D dual path network R-CNN[35]	**0.958**
3D Deep CNN w/ lung extraction[12]	0.947
2.5D Faster RCNN w/ deconv layer[8]	0.946
YOLOv2 w/ InceptionV3 backbone[2]	0.890
triple 2D Faster RCNN w/ deconv layer and RPN ensemble[32]	0.864
2.5D residual DCNN (proposed)	0.902

incide with larger lesions. It is also possible to hypothesise that this trend is counterbalanced for larger nodules, as its sheer size might make them easier to detect, is presented in Figure 6. Nonetheless, a more comprehensive analysis of this detection model needs to be performed using a larger test set.

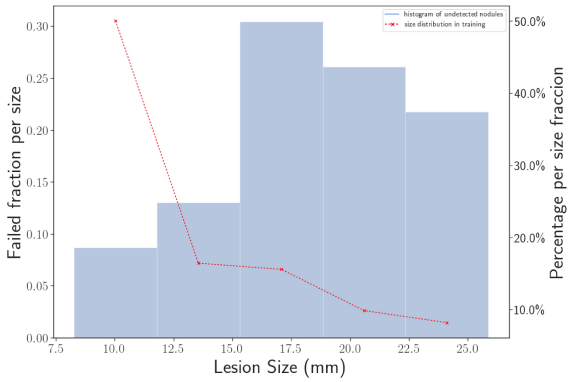


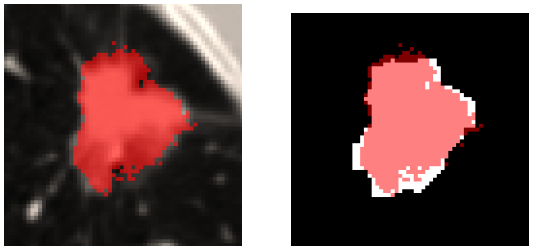
Figure 6: Comparison between false negatives count in the test set and percentage of examples, both with respect to the lesion size. An inverse pattern trend is noticeable with an increase of false negative counts as the percentage of examples decrease and the size of the lesions increases.

3.2. Segmentation task

The modified U-Net achieved 0.709 ± 0.124 dice score (assessed by 5-fold CV). This score outperforms the best model so far (Table 4), without using a 3D approach nor a cascaded network. The proposed architecture proves to be viable for the segmentation task at hand, without an extensive training setup, even being trained with a small number of examples, and still fully retaining the ability to generalize. Figure 7 shows two examples of segmented lesions, depicting one of the highest-scoring segmentations (Figure 7(a) and (b)), and then one of lowest-scoring segmentations (Figure 7(c) and (d)). These results illustrate the robustness of the model for the segmentation of lesions with challenging textures.

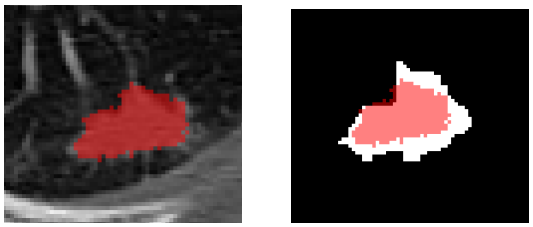
Table 4: Comparison of the dice score with state-of-the-art model in Decathlon-Lung task for lesion segmentation.

Model	Mean dice score
3D Cascade U-Net[15]	0.692
2D residual U-Net (proposed)	**0.709**



(a) Output mask overlaying the original scan

(b) Output mask overlaying the ground-truth segmentation



(c) Output mask overlaying the original scan

(d) Output mask overlaying the ground-truth segmentation

Figure 7: Example of segmentation with the *modified residual U-Net*. Top row: one of the segmentations with the best dice-score; bottom row: one of the lowest scoring segmentations.

In order to evaluate the impact of residual connection on the task of lung lesion segmentation, a vanilla U-Net was also trained, following the same image preparation pipeline of the residual U-Net. Even though the two model shared a similar pattern of loss convergence, the final assessment of the vanilla U-Net achieved 0.692 ± 0.115 dice score. Even though there is a marginal decrease in deviation, the residual U-Net achieved a better mean

performance.

A 2.5D approach was also tested, with training results comparable to the proposed model, but presenting worse testing performance. Arguably, for the specific case of the detected lesions, the added slices in a 2.5D approach contribute to overfitting the model to the training set. An exploratory assessment of a recent 3D segmentation model that allowed the improvement of multi-class segmentation of brain lesions, the *HighResNet* [18], was also performed. Nonetheless, early tests performed for lung cancer in CT scan indicated a failure of the model to fully segment the lesions.

This comparison validates the *LungSD-Net* as a viable approach for lesion segmentation. However, in order to achieve clinically relevant results, it is still mandatory to implement several improvements to its first stage, as these are directly connected to the segmentation task. Nonetheless, the proposed architecture proves to be viable for the segmentation task at hand, without an extensive training setup, even if trained with a small dataset.

4. Conclusions

This paper introduced a novel region-based hybrid model for detection and segmentation of lung lesions on 3D CT scans, which is comprised of a new 2.5D residual *deep convolutional neural network* (DCNN) and a modified residual U-Net. The results obtained support the hypothesis that simpler DCNN architectures are able to achieve relevant results in medical imaging, through a careful design of the training scheme, powerful data augmentation, and state-of-the-art architectural modifications. These simpler, but highly accurate, models have lower hardware requirements, thus are potentially of more widespread applicability. The results herein presented provide evidence for the advantages of reducing the problem of segmentation to the *volume of interest* (VOI) around each detected lesion. This is a first step towards the creation of a fully automated segmentation tool based on deep learning that first relies on the detection of lesions in a sliding-window fashion.

Naturally, there is still much room for improving lesion detection accuracy, namely by using a fully 3D architecture, which would benefit from the lower requirements of using a sliding-window approach. A semi-supervised method for the segmentation task would also be expected to yield a more clinically valid mask of the VOI. The integration of this work in a radiomics pipeline and its validation using external data-sets is the logic next step.

Acknowledgements

I wish to thank both my supervisors, Mário Figueiredo and Nickolas Papanikolaou, for all the insight and support provided during this project. This work was also enabled by the Champalimaud Foundation, which both granted access to computational power that made the model training possible, as well as funded my participation in the IEEE BIBE conference. I would like to thank José Moreira for the sustained assistance in this work.

References

[1] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018.

[2] G. Aresta, T. Araújo, C. Jacobs, B. van Ginneken, A. Cunha, I. Ramos, and A. Campilho. Towards an automatic lung cancer screening system in low dose computed tomography. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 310–318. Springer, 2018.

[3] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. 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.

[4] A. A. Bankier, H. MacMahon, J. M. Goo, G. D. Rubin, C. M. Schaefer-Prokop, and D. P. Naidich. Recommendations for measuring pulmonary nodules at ct: a statement from the fleischner society. *Radiology*, 285(2):584–600, 2017.

[5] L. Bogoni, J. P. Ko, J. Alpert, V. Anand, J. Fantauzzi, C. H. Florin, C. W. Koo, D. Mason, W. Rom, M. Shiau, et al. Impact of a computer-aided detection (CAD) system integrated into a picture archiving and communication system (PACS) on reader sensitivity and efficiency for the detection of lung nodules in thoracic CT exams. *Journal of digital imaging*, 25(6):771–781, 2012.

[6] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.

[7] T. A. Data, R. J. Gillies, P. E. Kinahan, and H. Hricak. Radiomics: Images Are More than Pictures, They Are Data. 278(2), 2016.

[8] J. Ding, A. Li, Z. Hu, and L. Wang. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 559–567. Springer, 2017.

[9] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*, pages 179–187. Springer, 2016.

[10] E. Gibson, W. Li, C. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, et al. Niftynet: a deep-learning platform for medical imaging. *Computer methods and programs in biomedicine*, 158:113–122, 2018.

[11] H. Greenspan, B. Van Ginneken, and R. M. Summers. Deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, 2016.

[12] R. Gruetzemacher, A. Gupta, and D. Paradice. 3D deep learning for detecting pulmonary nodules in CT scans. *Journal of the American Medical Informatics Association*, 25(10):1301–1310, 2018.

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[14] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7310–7311, 2017.

[15] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

[16] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[17] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[18] W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren. On the compactness, efficiency, and representation of 3D convolutional networks: brain parcellation as a pretext task. In *International Conference on Information Processing in Medical Imaging*, pages 348–360. Springer, 2017.

[19] A. L. Maas, A. Y. Hannun, and A. Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3, 2013.

[20] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *Journal of Magnetic Resonance Imaging*, 49(4):939–954, 2019.

[21] A. Nair, E. C. Bartlett, S. L. Walsh, A. U. Wells, N. Navani, G. Hardavella, S. Bhalla, L. Calandriello, A. Devaraj, J. M. Goo, et al. Variable radiological lung nodule evaluation leads to divergent management recommendations. *European Respiratory Journal*, 52(6):1801359, 2018.

[22] N. Papanikolaou and J. Santinha. An introduction to radiomics: Capturing tumour biology in space and time. *Hellenic Journal of Radiology*, 3(1), 2018.

[23] M. I. Razzak, S. Naz, and A. Zaib. Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps*, pages 323–350. Springer, 2018.

[24] O. Ronneberger, P. Fischer, and T. 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.

[25] R. E. Schapire and Y. Freund. Boosting: Foundations and algorithms. *Kybernetes*, 2013.

[26] D. Shen, G. Wu, and H.-I. Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

[27] R. L. Siegel, K. D. Miller, and A. Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, 2019.

[28] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.

[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[30] N. L. S. T. R. Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011.

[31] S. Wang, M. Zhou, Z. Liu, Z. Liu, D. Gu, Y. Zang, D. Dong, O. Gevaert, and J. Tian. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Medical image analysis*, 40:172–183, 2017.

[32] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognition*, 85:109–119, 2019.

[33] B. Xu, N. Wang, T. Chen, and M. Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.

[34] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 2019.

[35] W. Zhu, C. Liu, W. Fan, and X. Xie. Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 673–681. IEEE, 2018.