RRc-UNet 3D for lung tumor segmentation from CT scans of Non-Small Cell Lung Cancer patients

Van-Linh Le[1,2,3,4]
van-linh.le@u-bordeaux.fr

Olivier Saut[1,2]
olivier.saut@inria.fr

[1] MONC team - INRIA Bordeaux Sud-Ouest, Talence-33400, France
[2] Bordeaux Mathematics Institute UMR 5251 (IMB), University of Bordeaux, CNRS and Bordeaux INP, Talence, France
[3] Bordeaux Institute of Oncology, BRIC U1312, INSERM, University of Bordeaux, Institute Bergonié, 33000 Bordeaux, France
[4] Department of Data and Digital Health, Institute Bergonié, Bordeaux, France

Abstract

Lung cancer is a grave disease that accounts for more than one million deaths, and Non-Small Cell Lung Cancer (NSCLC) accounts for 85% *of all lung cancers. Rapid detection of lung cancer could reduce the mortality rate and increase the patient's survival rate, in which tumor segmentation plays a significant role in the diagnosis and treatment of lung cancer. Nevertheless, manual segmentation by radiologists can be time-consuming and labor-intensive. In recent years, deep learning methods have achieved good results in medical image segmentation. In this paper, RRc-UNet 3D, a variant of the Unet model, was proposed to perform tumor segmentation in Computed Tomography (CT) images of NSCLC patients. This network was trained end-to-end from a small set of CT scans of NSCLC patients, then the trained model was validated on another set of CT scans of NSCLC patients. The experimental results showed that our model can provide a highly accurate segmentation of tumors in the 3D volume of CT images.*

1. Introduction

Lung cancer is the second most common cancer worldwide with more than $230,000$ new cases annually. Early detection of lung cancer is critical to saving patients. A lung tumor refers to a small and visible lesion in the lung. Tumor segmentation plays a vital role in the used diagnosis and aggressive treatment of lung cancer. It can help the diagnosis of radiologists and help train new radiologists. In most modalities to acquire the image of the cancer patient, Computed Tomography (CT) is an effective medical screening that can be used for the diagnosis and detection of lung cancer. As CT procedures represent a large volume of CT scans with millions of voxels, manual diagnosis and realization of lung diseases are challenging and time-consuming even for experienced radiologists. Thus, automatic computer-aided diagnosis (CAD) for lung CT is a powerful solution to help radiologists. As usual, CAD involves several steps, of which accurate segmentation of tumors is the first step toward the success of the entire CAD system. Although many studies have been conducted, this problem remains a challenge for experts experienced for decades. Since manual segmentation is time-consuming and labor-intensive, an automatic method for nodule segmentation is highly desirable.

Over the past decade, deep learning methods have been at the forefront of computer vision, such as image classification, segmentation, and object detection with several proposed Deep Convolutional Neural Network (DCNN)[1, 2, 3, 4, 5, 6, 7]. The methods most often focus on image classification tasks on very large-scale datasets like ImageNet [4], where the outputs are single labels or probability values for each input. Usually, the model consists of some convolutional layers with activation functions followed by down-sampling layers, which reduce the dimensions of the feature maps. As an input traverse through the layers of the network, the number of feature maps increases, but the dimensions of feature maps decrease. Eventually, an activation function (e.g., Sigmoid or Softmax) is applied at the end of the model to compute the probability of the target classes.

Besides classification, segmentation is another success story of deep learning methods in computer vision. A

limited number of classification architectures can be transformed and used for the segmentation tasks. For example, a Fully Convolutional Network (FCN)[6] replaces the last fully connected layer in the neural network with a convolutional layer, which improves the efficiency of segmentation. However, that transformation is not enough, thus the model should be particular with specific operations to provide a better feature map for the segmentation task. In the last years, UNet architecture [7] was a good model that succeeded in segmenting the medical image. Principally, the UNet model [7] consists of two paths: a contracting and an expansive path. The contracting is like a classic CNN to learn from low-level features to the high-level of an input representation. On another side, the expansive consists of an up-sampling of the feature map followed by up-convolution and convolution layers to rebuild the feature at different scaled inputs. A concatenation is made between the convolution layer in the contracting and the up-convolution of the expansive path to get better precise locations. At the final layer of the expansive, a convolution layer is used to map the features to the desired number of classes. This model opened a new benchmark in the segmentation domain by introducing an encoder-decoder (contracting-expansive) architecture. This model bested all previous state-of-the-art segmentation techniques based on CNN and pixel-wise classification.

Generally, the model for segmentation tasks like the UNet model requires both convolutional encoding and decoding paths. The encoding encodes the input image into a large number of feature maps with small dimensions. The decoding path performs up-convolution operations to produce segmentation maps with the same dimension as the original input image. Therefore, the architecture model for the segmentation task has double the number of parameters compared to the classification model. Thus, it is interesting to design the model for segmentation tasks that can ensure performance with a small number of parameters.

In this work, we introduce another 3D architecture that inspired from UNet [7], Residual Neural Network [3], Recurrent CNN [5] and R2U-net [8], named RRc-UNet 3D, for segmenting lung tumor from 3D volume of CT image of a NSCLC patients. Our model has the same structure as Unet [7], but we have modified the core inside each level of the encoder and decoder paths. Our model achieved an accurate segmentation of tumor in CT scans. The highlighted points of our method are summarized as follows:

- The proposed model inputs lung CT volume (3D) and outputs the segmentation of the tumors inside the lung without any post-processing operation.

- The proposed model can provide the results in a short time, e.g., the model provides the predictions of lung tumors for 41 testing patients in less than 3 minutes.

- This prediction can be used to assess the development of the tumor region, as well as diagnosis and treatment of cancer.

This paper is organized as follows: The following section discusses the related works. Section 3 presents the methodology, which details the RRc-UNet 3D architecture and the loss function used to train the model. Section 4 details the experimental process: Firstly, we describe the datasets for the experiment. Secondly, we present the evaluation scores in this study and the model's implementation. Finally, we show the experimental results. Section 5 points out some concluding remarks from our works.

2. Related Work

Initial approaches for lung/tumor segmentation consider the characteristics of the image or based on the shape knowledge by using the traditional image processing technique such as thresholding based on the grey-level, adaptive thresholding technique, region growing, image registration [9], or the others [10]. However, these methods are computationally expensive, and it is difficult to generalize the learning features.

Among diverse modalities for acquiring medical images, the CT scan is a usual choice of the clinician to work with cancer patients, for example, for lung or liver cancer [11]. In a lung CT scan, it is easy to distinguish the lung parenchyma because this region is large and it takes up most of the space on an image. However, the tumor region is a difficult question when the number of tumor voxels is very tiny compared to the total lung. Therefore, tumor segmentation could be a challenge in many studies about lung diseases [12, 13, 14].

In recent years, Convolutional Neural Networks (CNNs) have been applied to various medical image segmentation tasks and it has achieved numerous success. On lung disease, CNNs models can be used to analyze lung CT scans [15]. Their applications vary from detecting the lung pathology segmentation [15], classifying the lung region [15, 16] to segmenting the lung volume [17]. In the approaches to provide whole volume of lung segmentation, most of applications have chosen to consider independently each slice of CT volume: Ravindra Patil et al. [18] and Brahim A. S. et al. [17] proposed different modifications on UNet model to provide the segmentation of slices in CT volume. Lei Geng et al. [19] presented a combination of VGG-16 [20] and dilated convolution to provide the lung segmentation. Swati P. Pawar et al. [21] have proposed a conditional generative adversarial network to encode the features from the slices of CT image. Then, the encoder and decoder were used to extract the multi-scales features and to give the lung segmentation, respectively. In whole 3D approach, Negahdar et al. [22] have proposed a volumetric segmentation network based on V-net [23] for 3D

volumetric medical segmentation. Md Zahangir Alom et al. [8] proposed RU-Net and R2U-Net models to provide the segmentation of medical images on several benchmarks.

In the lung tumor segmentation approach, Hongtao Xie et al. [24] have proposed a framework with 2D CNN to assist the CT reading process consisting of two steps: (1) nodule candidature detection based on a Faster R-CNN model integrated a deconvolution layer to enlarge the feature map. They trained the model on the slices of the CT scan, then merged the candidates to obtain the results; (2) false positive reduction based on training a classifier to reduce the false positive produced by the first step. Fuli Zhang et al. [25] have modified ResNet and applied it to segment the gross tumor in the CT scans of NSCLC. Evi Kopelowitz et al. [26] employed a MaskRCNN to handle 3D images to detect and segment lung nodules from CT scans. Uday Kamal et al. [27] proposed a Recurrent 3D-DenseUNet based on a recurrent multiple Convolutional Long Short-Term Memory for lung tumor segmentation.

3. Methodology

In this section, we present the architecture of the proposed model, the RRc-UNet 3D model. Then, we describe the loss functions used to train the model.

3.1. RRUnet 3D architecture

Figure 1 shows the structure of the RRc-UNet 3D. Depending on the requirements of the applications, the depth and layers at each model's level are different. After testing different values for the depth of the model, we have decided to use five levels for the proposed model. In addition, we made some modifications to the structure of the model to adapt to our problem: (1) the layers have been changed to adapt with 3D input features; (2) each level of contracting and the expansive path contains a Residual Recurrent block (RRc); (3) the layers in each RRc block have been re-organized to adapt with our question; (4) the number of channels of the input is doubled before reducing the space to the next level.

Figure 2 illustrates the structure of a RRc block consisting of two Recurrent Blocks (RBs). The input traverses a 3D Convolutional layer. Then, the convoluted features will be fed to the RBs. Finally, it is concatenated to the output of RBs for providing the outcome of the RRc block. In our design, a RB is a repetition of t times (e.g., t = 2) of a group including three consecutive layers: Convolutional 3D layer, Group Normalization (GN) [28], and PreLU activation [2] function. Firstly, the GN is used instead of Batch normalization (BN) [29] because BN increases the error rapidly caused by inaccurate batch statistics estimation when using a small batch. On the opposite, GN divides the channels into groups and computes the mean and variance for normalization within each group. GN's computation is independent

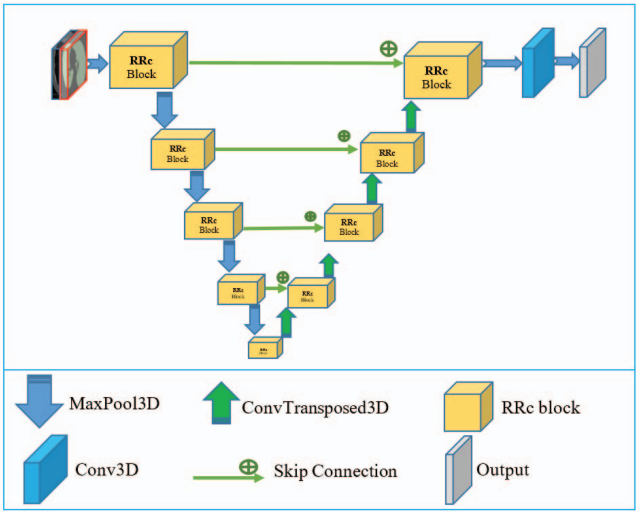

Figure 1: The architecture of the RRc-UNet 3D

of batch sizes, and its accuracy is stable in a wide range of batch sizes. Second, the PreLU activation function [2] is used instead of the ReLU activation function [30] to make the leakage coefficient a parameter learned along with the other network parameters. Really, RB takes full advantage of the spatial continuity of the slices in the volume.

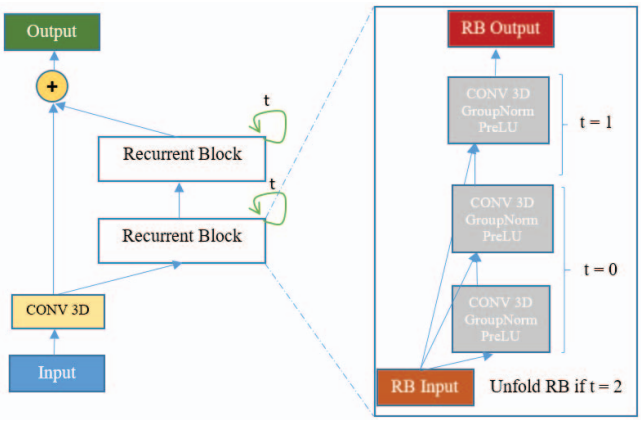

Figure 2: The structure of a Residual Recurrent block

As presented in Figure 1, RRc-UNet 3D receives a two-channel input: CT scan and lung parenchyma map. In fact, the lung map was obtained from the CT scan by using another model in our work [31]. Each level of the contracting path includes one RRc block followed by a Dropout layer. The filter begins with 32 at the first level. Then, the filter is doubled at the end of the RRc block before sending to the next level in the encoder path. At the decoder path, every level consists of an up-sampling layer that halves the

number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting, followed by a RRc block to provide the information context of the feature map and a Dropout layer. Finally, a 1×1 Convolution layer maps the feature to the number of classes. In this work, we trained the model to classify the voxels into three categories: background, lung parenchyma, and tumor. However, the main goal is to provide the segmentation of tumors. The output of the lung parenchyma map is to improve the segmentation of lung parenchyma from the previous step in our application.

Table 1 details the input/output images at each level of the RRc-UNet 3D model. The input is a two-channel image consisting of a CT scan and lung parenchyma segmentation. In the encoding path, we decrease the dimensions of the images and increase the number of channels after each level; in the decoder path, we reduce the number of channels but increase the dimensions of the images.

There are several advantages of the proposed model compared to U-Net based on the comparison of network parameters. The proposed model with two folds (t=2) in Recurrent Block has the same parameters as the Unet model but shows better performance on the segmentation tasks.

3.2. Loss function

The main objective of RRc-UNet 3D is to provide the segmentation of tumors, while it predicts the pixels into three categories: background, lung parenchyma, and tumor region. Thus, the tumor region is a tiny part compared to the background region and lung parenchyma; therefore, it is difficult to extract the features of tumor regions, which reduces the confidence of the model. So, the tumor regions are considered difficult samples. The background and lung parenchyma are large regions, it is easier to extract these regions than the tumor ones. These samples are seen as easy samples. To increase the model's confidence, we have to remark on this point while training the model. Therefore, the Focal loss (FL) [32], expressed in Equation 1, can be used to solve this concern. It down-weights the contribution of easy samples and enables the model to focus more on learning difficult samples. In addition, Focal loss works well for highly imbalanced class scenarios.

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma} log(p_t) \tag{1}$$

Here, $\gamma > 0$ and α generally range from $[0, 1]$. When $\gamma = 1$, Focal loss become Cross-entropy loss [33]. p_t is the model's estimated probability, α is used to adjust the distribution of the easy sample, and $(1-p_t)^{\gamma}$ is a dynamic scaling factor to adjust the distribution of hard samples. In our implementation, we have set $\gamma = 2$ and $\alpha = 1$, but they can be treated as hyper-parameters.

As usual, Dice loss (DL) [23], expressed in Equation 2, is usually used for the segmentation problems in deep learning.

$$DL = \frac{1}{N}(1 - DSC) \tag{2}$$

Where, N is the number of samples, DSC is dice coefficient expressed by Equation 5.

According to Dice loss, it considers only the segmented objective regions to be sure that the model focuses more on extracting the tumor regions than the others. Following that, the other regions (healthy regions) are ignored. However, it is unreasonable to ignore healthy regions, which also contain important information, e.g., the features of lung parenchyma. If the model can compare the features from objective regions and healthy regions, the performance of the model can be enhanced. To solve this problem, the Generalized Dice loss (GDL) [34] is a good candidate. It is the multi-class extension of Dice loss where the weight of each class is inversely proportional to the square of label frequencies (Eq. 3).

$$GDL = 1 - \frac{2|W(2\hat{Y}-1) \cap W(2Y-1)|}{|W(2\hat{Y}-1)| + |W(2Y-1)|} \tag{3}$$

Where, W, Y, and $\hat{Y}$ are the weights generated according to the segmentation labels, ground truth, and predicted segmentation, respectively.

Thus, our application meets both of the two mentioned problems. Therefore, we have combined the two losses for training the model (Eq. 4).

$$Loss = Focal\ Loss + G.Dice\ Loss \tag{4}$$

4. Experiments and Results

4.1. Dataset and Pre-processing

RRc-UNet 3D was trained and validated on 3D CT scans of NSCLC patients. Images were collected from 2 sources: the public and local datasets. The public dataset consists of 494 CT scans from 3 public datasets: NSCLC-Radiomics-Interobserver1 dataset [35], NSCLC-Radiomics-Genomics dataset [36] and LungDecath dataset [37]. The local dataset consists of 41 CT scans of local NSCLC patients. Each image has a variable size from $(512 \times 512 \times 60)$ to $(512 \times 512 \times 600)$ pixels. All images were reformatted from standard DICOM to Neuroimaging Informatics Technology Initiative (NIfTI) format created by the National Institutes of Health [38]. The NIfTI file held the 3D image matrix and diverse metadata. Figure 3 shows a 3D image for volumetric measurements of lung parenchyma across three axes.

Data splitting and pre-processing:

The images from 2 sources were divided into two sets: train/validation and testing sets. The train/validation set

Level	Input ($c \times w \times h \times d$)	Output ($c \times w \times h \times d$)	Level	Input ($c \times w \times h \times d$)	Output ($c \times w \times h \times d$)
Encoder path			Decoder path		
Input	-	$2 \times 256 \times 256 \times 48$	Output	$32 \times 256 \times 256 \times 48$	$1 \times 256 \times 256 \times 48$
Level 0	$2 \times 256 \times 256 \times 48$	$32 \times 256 \times 256 \times 48$	Level 0	$96 \times 256 \times 256 \times 48$	$32 \times 256 \times 256 \times 48$
Level 1	$32 \times 128 \times 128 \times 24$	$64 \times 128 \times 128 \times 24$	Level 1	$192 \times 256 \times 256 \times 48$	$64 \times 128 \times 128 \times 24$
Level 2	$64 \times 64 \times 64 \times 12$	$128 \times 64 \times 64 \times 12$	Level 2	$384 \times 64 \times 64 \times 12$	$128 \times 64 \times 64 \times 12$
Level 3	$128 \times 32 \times 32 \times 6$	$256 \times 32 \times 32 \times 6$	Level 3	$768 \times 32 \times 32 \times 6$	$256 \times 32 \times 32 \times 6$
Level 4	$256 \times 16 \times 16 \times 3$	$512 \times 16 \times 16 \times 3$	Level 4	$512 \times 16 \times 16 \times 3$	$512 \times 16 \times 16 \times 3$

Table 1: The dimensions of input/output features at each level of proposed model.

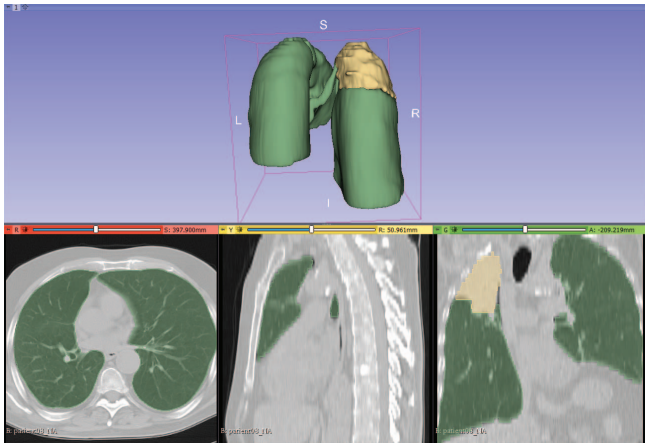

Figure 3: 3D volume of lung with XYZ axes. Green/ yellow is the segmentation of lung parenchyma/ tumor, respectively.

consists of 494 images from the public datasets. The testing set consists of the remaining 41 images of the local dataset. Table 2 details the number of patients in each dataset and the ground truth in each dataset. During the training process, we divided the 494 images into two sets corresponding to the training and validation process with a ratio of $0.8 : 0.2$.

Dataset	Nb. patients	Lung mask	Tumor mask
R.Interob.1	21	✓	✓
R.Genomics	410	×	✓
LungDecath	63	×	✓
Local	41	✓	✓

Table 2: The number of patients (CT scans) in each dataset and the ground truth masks.

According to the guideline of CT imaging [11], a CT scan consists of pixel spacing, axial slice thickness, and view in the z-axis with various scans. Therefore, the input images are uniformly pre-processed to minimize the variability within the database. In this work, the input images go through the three following steps for preparing the images:

1. The image intensity of each slice was first truncated in the range of $[-1200, 600]$.
2. Z-normalization was performed on each CT scan.
3. The CT scans were cropped to focus on the lung region and converted to $256 \times 256 \times 48$ pixels.

As mentioned, our model receives an input of two channels which consist of a CT scan and segmentation of lung parenchyma. However, the two datasets: R.Genomics and LungDecath do not provide the ground truth for the lung parenchyma. Thus, we have designed another variant of Unet with helping of the Coordinated Convolutional layer [39] to extract the lung parenchyma in another task [31]. This model has been trained and validated on other datasets. The results showed that it can be used as an extractor to provide the segmentation of lung parenchyma. The outputted segmentation was good enough to be used as ground truth (input) for the task of segmentation of tumors.

Data augmentation

Due to the limitation of the number of samples in the training set, some augmentation operations have randomly been applied to the image during the training process, for example, flipping, deformation, and affine transformation. Figure 4 shows an augmented example in our dataset.

4.2. Implementation details

The model was implemented using the PyTorch library [40]. Models were trained in 1000 epochs using an Adam optimization [41] with a weight decay of 10^{-4}. The learning rate was set to 3×10^{-4}, batch size was equal to 1, and the loss was backward after obtaining forward values of all patients. An early stopping strategy was applied by monitoring the validation loss to prevent over-fitting.

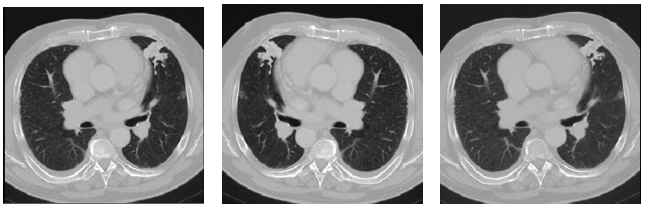

Figure 4: A slice in a 3D studied image with its augmentation. From left to right: original slice, flipped slice, and deformed slice.

4.3. Evaluation metrics

The evaluation of 3D lung parenchyma segmentation is based on comparing ground truth and outputted segmentation from the model. For quantitative analysis of the predicted segmentation, several performance metrics are considered, including the Dice score coefficient (DSC), Jaccard similarity (IOU), and F1 score [42].

DSC is expressed as in Eq. 5, the IOU score is represented using Eq. 6 according to [43] and [44], respectively. Here, GT and SP refer to the ground truth and predicted segmentation, respectively.

$$DSC = \frac{2 * (|GT \cap SP|)}{|GT| + |SP|} \tag{5}$$

$$IOU = \frac{|GT \cap SP|}{|GT \cup SP|} \tag{6}$$

Furthermore, F1 is calculated by using the Eq. 7.

$$F1 = 2 * \frac{(precision * recall)}{(precision + recall)} = \frac{2TP}{2TP + FP + FN} \tag{7}$$

where True Positive (TP), False Positive (FP) and False Negative (FN) are the variables.

4.4. Results and Analysis

As mentioned, we trained the RRc-UNet 3D model to classify the voxels into three categories: background, lung parenchyma, and tumor; but the main objective was to segment the tumor regions. To evaluate the model's performance, we computed the scores on three labels and only on tumor label. Table 3 and 4 show the evaluation scores on three categories and tumor only, respectively.

Score	Dice coeff.	IOU	F1
Value	0.863	0.9971	0.9982

Table 3: The evaluation scores of RRc-UNet 3D on three labels.

Score	Dice coeff.	IOU	F1
Value	0.8777	0.7274	-

Table 4: The evaluation scores of RRc-UNet 3D on tumor only.

These scores showed that the RRc-UNet 3D model highly perform on the validation set to provide the prediction for three categories. It consistently provides high evaluation scores for the three classes. It is evident because the lung parenchyma segmentation has supported it to pre-indicate the tumor location. Concerning the tumor region, the model has provided a significant Dice coefficient, even higher than the three categories. On other metrics, we obtained a good score as well, but they are still tiny compared to the performance of the three classes.

The trained model has been used to predict the segmentation for each image in the testing set (local dataset). Then, the predictions were compared with ground truth by calculating the Dice coefficient. It is worth noting that we take into account only the segmentation of the tumor in this step. Finally, the average Dice coefficient has been computed for all testing images. We have obtained an average Dice coefficient of 0.7682 for all 41 CT scans in the testing set. Figures 5 and 6 illustrate the predictions of two samples in the testing set with high and low dice coefficients, respectively. In the first case, the RRc-UNet 3D model provides perfect segmentation of the entire tumor inside the lung. In the second one, the model detects only a limited tumor region, which is big and has spread to the outside of the lung.

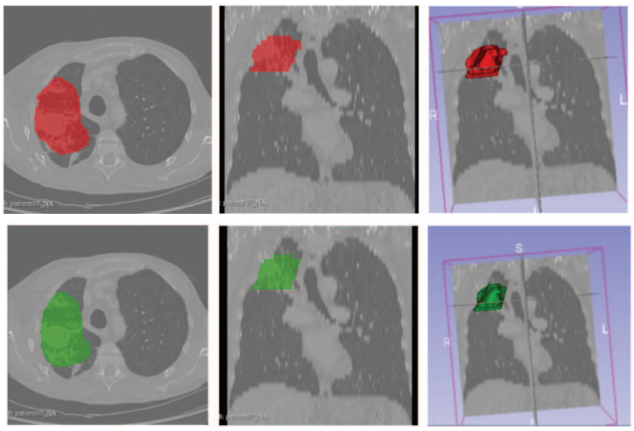

Figure 5: A good prediction in the testing set(top = prediction, bottom = ground truth). From left to right: X-axis, Z-axis, and 3D view.

To better understand the difference between the results on the two datasets (validation and test), we look at the histogram of the distribution of Dice scores of the patients in

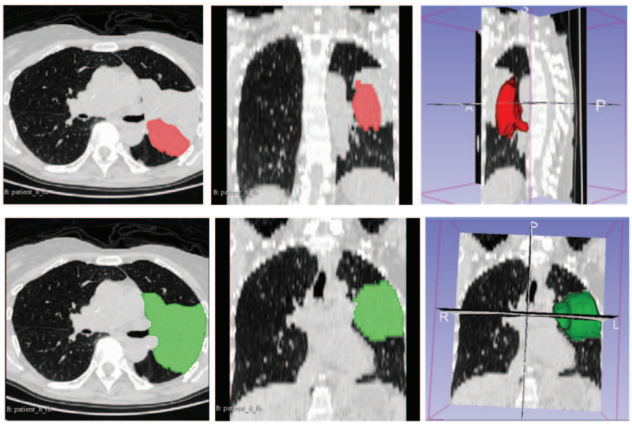

Figure 6: A bad prediction in the test set (top = prediction, bottom = ground truth). From left to right: X-axis, Z-axis, and 3D view.

the testing set. Figure 7 illustrates the number of images in a different range of Dice scores. We have 24 well predictions with dice scores greater than 0.8; 6 good predictions with scores from 0.75 to 0.8; 5 cases have scored in the range from 0.5 to 0.75; and 5 cases have dice scores less than 0.5. In summary, we obtain 30 out of 41 CT scans with Dice scores greater than 0.75; 11 CT scans have scores smaller than 0.75.

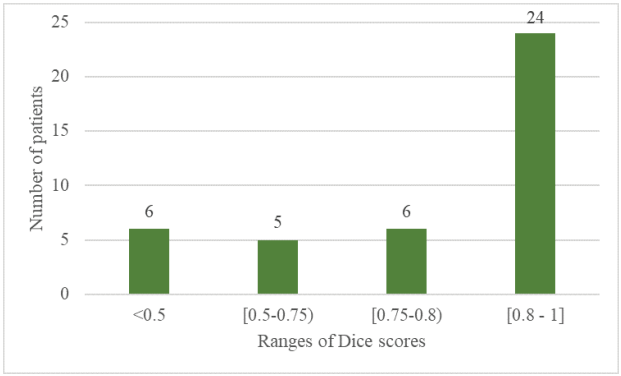

Figure 7: The number of predictions (patients) in each range of Dice score

Thus, our model has worked fine to segment the tumor inside the lung. It provides a perfect segmentation for the small tumor or the tumor located inside the lung. The model meets the difficulties of the large tumor staying at the border of the lung. Figure 8 illustrates a slice of the easy and hard cases in the testing set. Both two tumors are big, the difference is the location of the tumors: the tumor of the hard case mostly stays on the wall of the lung, while the other begins to contact the wall of the lung.

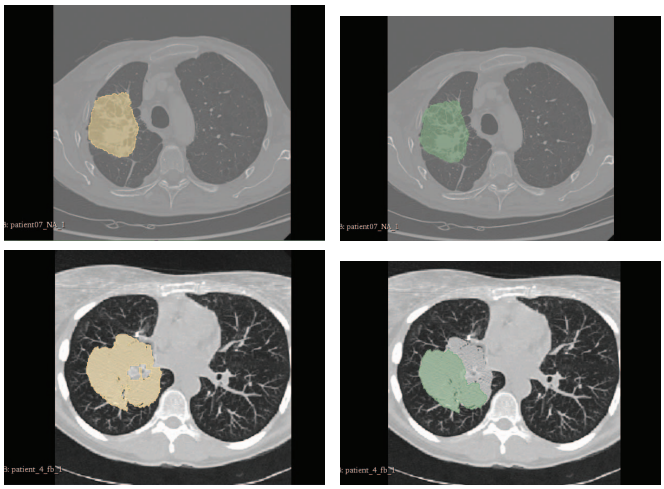

Figure 8: A slice of a 3D studied scan. Top/Bottom: The easy/hard cases in the testing set. Left/Right: The ground truth/ prediction of tumor segmentation over the CT scans.

4.5. Comparison with other block concepts

To audit the efficiency of the RRc block, we replaced the RRc block with other concepts. In this work, we investigated the efficiency of the CGPD block, the Residual block, and the Recurrent block. For each concept, we placed two blocks at each level of the model instead of a RRc block. The modified models have been trained in the same scenarios as we have done with the RRc block.

CGPD block: consists of a Convolution layer, a Group normalization (GN) layer [28], a PreLU activation function [2], and a Dropout layer [45] (Figure 9). The number of input channels is doubled at the second block before applying a max pooling layer to reduce the spatial size and send it to the next level.

Residual block: is constructed from the convolutional layer, group normalization, and PreLU activation as well (Fig. 10). In the Unet model, we placed two Residual blocks followed by two dropout layers.

Recurrent block: includes three consecutive layers: Convolutional layer, Group Normalization, and PreLU activation. These layers can repeat in several times (e.g., t = 2) to provide the output.

Table 5 presents the evaluation scores of the model by using different concepts of blocks. We see that models with different concepts of blocks provide a good score in 3 categories. It is obvious because we have support from the lung parenchyma segmentation to pre-indicate the tumor location. However, we see the difference in the scores that concern the tumor region. The best Dice coefficient is obtained

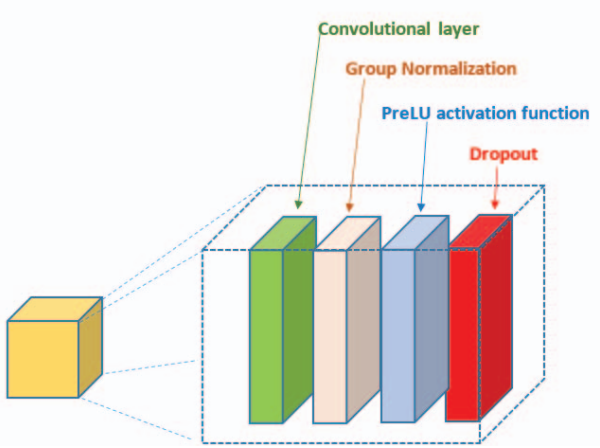

Figure 9: Layers in a CGPD block

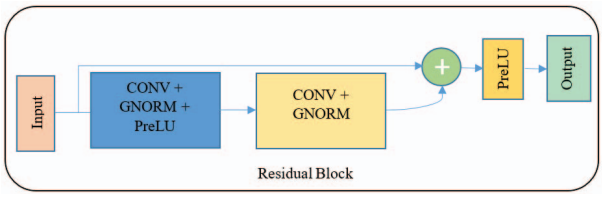

Figure 10: Layers in a Residual block

with the RRc block. The same score on CGPD is also good, but it is very worst on the Residual and Recurrent blocks.

Concepts	IOU 3 labels	F1 3 labels	Dice coeff. tumor	IOU tumor
CGPD block	0.9955	0.9974	0.7837	-
Residual block	0.9955	0.9973	0.1737	0.5843
Recurrent block	0.9957	0.9973	0.1992	0.5825
RRc block	**0.9971**	**0.9982**	**0.8777**	**0.7274**

Table 5: The comparison of the evaluation scores among different concepts of block for Unet model.

It is not clear to compare our results with other results because we are on different approaches [25, 26, 27]. For example, we consider the whole CT scan with the assistance of lung parenchyma segmentation; while Uday Kamal et al. [27] focus only on the local regions (several slices) around the tumors. However, RRc-UNet 3D model can segment tumor regions with very satisfactory performance and have the potential to locate and analyze tumor lesions.

5. Conclusion

In this work, we presented an RRc-UNet 3D model to provide tumor segmentation from CT scans of NSCLC patients. We have obtained an accurate segmentation with a Dice coefficient of 0.8777 for the validation set. The trained model has been used to predict tumor segmentation from the CT scans of a local dataset. The average Dice score on the testing set showed that our model worked fine to provide tumor segmentation (average Dice coefficient on the testing set of 0.7682). We saw also a bias between the Dice scores on the validation and testing set. The problem was found as the difficulties of the big tumors, which stay at the border of the lung. The advantage of the method is the model can work with a whole 3D volume of the CT scan, and it can be applied to a wide area of different medical image segmentation tasks. Another objective is to generalize the approach for use in another work, for example, to automatically prelocate the tumor position. Future work could be focused on considering the two sets of small and big tumors.

6. Acknowledgments

This work was supported by the Fondation MSDAvenir and Fondation Inria for the Pimiento project. The experiments presented in this paper were carried out using the PlaFRIM platform[1].

References

[1] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015. 1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 1, 3, 7

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 2

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017. 1

[5] Ming Liang and Xiaolin Hu. Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3367–3375, 2015. 1, 2

[6] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1, 2

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 1, 2

[1] https://www.plafrim.fr

[8] Md Zahangir Alom, Chris Yakopcic, Mahmudul Hasan, Tarek M Taha, and Vijayan K Asari. Recurrent residual u-net for medical image segmentation. *Journal of Medical Imaging*, 6(1):014006–014006, 2019. 2, 3

[9] Kenneth R Castleman. *Digital image processing*. Prentice Hall Press, 1996. 2

[10] Selin Uzelaltinbulat and Buse Ugur. Lung tumor segmentation algorithm. *Procedia computer science*, 120:140–147, 2017. 2

[11] Dominique Delbeke, R Edward Coleman, Milton J Guiberteau, Manuel L Brown, Henry D Royal, Barry A Siegel, David W Townsend, Lincoln L Berland, J Anthony Parker, Karl Hubner, et al. Procedure guideline for tumor imaging with 18f-fdg pet/ct 1.0. *Journal of nuclear Medicine*, 47(5):885–895, 2006. 2, 5

[12] David S Ettinger, Wallace Akerley, Gerold Bepler, Matthew G Blum, Andrew Chang, Richard T Cheney, Lucian R Chirieac, Thomas A D'Amico, Todd L Demmy, Apar Kishor P Ganti, et al. Non–small cell lung cancer. *Journal of the national comprehensive cancer network*, 8(7):740–801, 2010. 2

[13] Jan P Van Meerbeeck, Dean A Fennell, and Dirk KM De Ruysscher. Small-cell lung cancer. *The Lancet*, 378(9804):1741–1755, 2011. 2

[14] Elise Noel-Savina and Renaud Descourt. Focus on treatment of lung carcinoid tumor. *OncoTargets and therapy*, pages 1533–1537, 2013. 2

[15] A Asuntha and Andy Srinivasan. Deep learning for lung cancer detection and classification. *Multimedia Tools and Applications*, 79:7731–7762, 2020. 2

[16] Aleksandr Zotin, Yousif Hamad, Konstantin Simonov, and Mikhail Kurako. Lung boundary detection for chest x-ray images classification based on glcm and probabilistic neural networks. *Procedia Computer Science*, 159:1439–1448, 2019. 2

[17] Brahim Ait Skourt, Abdelhamid El Hassani, and Aicha Majda. Lung ct image segmentation using deep neural networks. *Procedia Computer Science*, 127:109–113, 2018. 2

[18] Ravindra Patil, Leonard Wee, and Andre Dekker. Auto segmentation of lung in non-small cell lung cancer using deep convolution neural network. In *Advances in Computing and Data Sciences: 4th International Conference, ICACDS 2020, Valletta, Malta, April 24–25, 2020, Revised Selected Papers 4*, pages 340–351. Springer, 2020. 2

[19] Lei Geng, Siqi Zhang, Jun Tong, and Zhitao Xiao. Lung segmentation method with dilated convolution based on vgg-16 network. *Computer Assisted Surgery*, 24(sup2):27–33, 2019. 2

[20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 2

[21] Swati P Pawar and Sanjay N Talbar. Lungseg-net: Lung field segmentation using generative adversarial network. *Biomedical Signal Processing and Control*, 64:102296, 2021. 2

[22] Mohammadreza Negahdar, David Beymer, and Tanveer Syeda-Mahmood. Automated volumetric lung segmentation of thoracic ct images using fully convolutional neural network. In *Medical Imaging 2018: Computer-Aided Diagnosis*, volume 10575, pages 356–361. SPIE, 2018. 2

[23] Fausto Milletari et al. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016. 2, 4

[24] Hongtao Xie, Dongbao Yang, Nannan Sun, Zhineng Chen, and Yongdong Zhang. Automated pulmonary nodule detection in ct images using deep convolutional neural networks. *Pattern Recognition*, 85:109–119, 2019. 3

[25] Fuli Zhang, Qiusheng Wang, and Haipeng Li. Automatic segmentation of the gross target volume in non-small cell lung cancer using a modified version of resnet. *Technology in Cancer Research & Treatment*, 19:1533033820947484, 2020. 3, 8

[26] Evi Kopelowitz and Guy Engelhard. Lung nodules detection and segmentation using 3d mask-rcnn. *arXiv preprint arXiv:1907.07676*, 2019. 3, 8

[27] Uday Kamal et al. Lung cancer tumor region segmentation using recurrent 3d-denseunet. In *Thoracic Image Analysis: Second International Workshop, TIA 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2*, pages 36–47. Springer, 2020. 3, 8

[28] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. 3, 7

[29] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015. 3

[30] Andreas Maier, Christopher Syben, Tobias Lasser, and Christian Riess. A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik*, 29(2):86–101, 2019. 3

[31] L. Van Linh and S. Olivier. Coordinate-Unet 3D for segmentation of lung parenchyma. *Computer Science Research Notes*, 3301:36–42, 2023. 3, 5

[32] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 4

[33] Jason Brownlee. *Probability for machine learning: Discover how to harness uncertainty with Python*. Machine Learning Mastery, 2019. 4

[34] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, DLMIA 2017*, pages 240–248. Springer, 2017. 4

[35] Petros Kalendralis et al. Fair-compliant clinical, radiomics and dicom metadata of rider, interobserver, lung1 and head-neck1 tcia collections. *Medical Physics*, 47(11):5931–5940, 2020. 4

[36] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*, 5(1):4006, 2014. 4

[37] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022. 4

[38] Michele Larobina and Loredana Murino. Medical image file formats. *Journal of digital imaging*, 27:200–206, 2014. 4

[39] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in neural information processing systems*, 31, 2018. 5

[40] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017. 5

[41] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[42] Pierre Baldi, Søren Brunak, Yves Chauvin, Claus AF Andersen, and Henrik Nielsen. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):412–424, 2000. 6

[43] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945. 6

[44] Paul Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912. 6

[45] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 7