# Deep Neural Networks for Lung Cancer Tumor Region Segmentation

Runze Zhang, Zhong Guan, Shun-Cheung Lai, Jun Xiao, and Kin-Man Lam
Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

*Abstract*— **In this paper, we investigated the use of deep neural networks to perform segmentation and prediction of lung cancer tumor regions, via screening Computed Tomography (CT) scans. In order to achieve satisfactory results, we have adopted U-Net, a deep neural network, for the segmentation of tumor regions in the CT scans. U-Net was designed for medical image segmentation. With the segmentation results generated by U-Net, some false-positive segments will appear. Therefore, a process similar to radiomic analysis was carried out to rectify this issue. After evaluating the performance of various networks, ResNet-18 can achieve the best performance, in terms of reducing the false positives. Using the extracted deep features for classification, each segmented region generated by U-Net is classified as to whether it contains a real tumor or not. With the use of U-Net only, the four evaluation metrics, i.e., dice coefficient, mean surface distance, Hausdorff distance, and patient-wise accuracy, are 0.5471, 12.505, 29.336 and 90%, respectively. After using the radiomic analysis, the four metrics become 0.592, 9.786, 20.061, and 90%, respectively.**

***Keywords—deep learning, segmentation, U-Net, radiomic analysis, lung tumor***

## I. Introduction

Medical diagnosis, based on images, involves a huge amount of data. The interpretation of such a large amount of data has always relied on the experience of the radiologist, and thus has made it a very time-consuming process. This outlines the fact that automated analysis and diagnosis of medical images is crucial. In this paper, efficient methods have been studied for the segmentation and prediction of lung cancer tumor regions via screening computed tomography (CT) scans.

To achieve an accurate segmentation and prediction performance, efficient segmentation and prediction methods for tumor detection are indispensable. Our proposed method uses a deep neural network for segmentation, followed by a radiomic analysis [1] for removing false positives from the segmentation results to further improve the accuracy. In order to achieve better performance, another deep neural network is employed. In Section II, we will discuss the methods we considered, and those finally used in our tumor-segmentation method. The experiment set-up and results will be presented in Section III, and a conclusion will be given in Section IV.

## II. Methods for Lung Tumor Region Detection and Segmentation

It is a challenging task to detect and segment lung tumor regions in CT scans. We employed various learning algorithms, which were trained based on the ground truth images of a data set. In this section, we will describe the techniques used in our method, including the deep neural networks and radiomic analysis.

### *A. Deep Neural Network for Tumor Region Segmentation*

To detect and segment a tumor region, various types of deep networks can be adopted. Two approaches can be used for this purpose. The first one is to use object-detection methods, including the Faster RCNN [2], SSD [3], and Yolo [4], to locate the tumor regions first, followed by segmentation. The second approach is to perform detection and segmentation from end to end, with the potential deep models for this purpose including Mask RCNN [5], DeepLab [6], and U-Net [7]. We have tested the performance of all the models. In conclusion, we have found that U-Net achieves the best performance for this task.

The U-Net model [7], which was proposed by Ronneberger et al., has shown outstanding performance on segmentation task of medical images. As this model was designed for medical image segmentation, it could also achieve a better performance in tumor region segmentation. With the original U-Net structure, we have made some minor modifications, so that the network can be adapted for lung tumor segmentation. The structure used in the challenge is shown in Fig. 1. Each blue block represents two convolutional layers, with 3×3 filters, separately followed by a rectified linear unit (ReLU) layer. The white block denotes the copied feature maps. Each downward arrow represents a 2×2 max pooling layer, with a stride of 2. In this step, the number of feature channels is doubled. The upward arrow denotes a 2×2 upsampling convolution, while the number of feature maps will be halved. Finally, a 1×1 convolution is used, followed by a sigmoid layer, to normalize the results. With such a structure, the output segmentation map will maintain a similar size to the input slice. Furthermore, U-Net can extract features on different levels and can achieve good performance with a limited amount of training data.

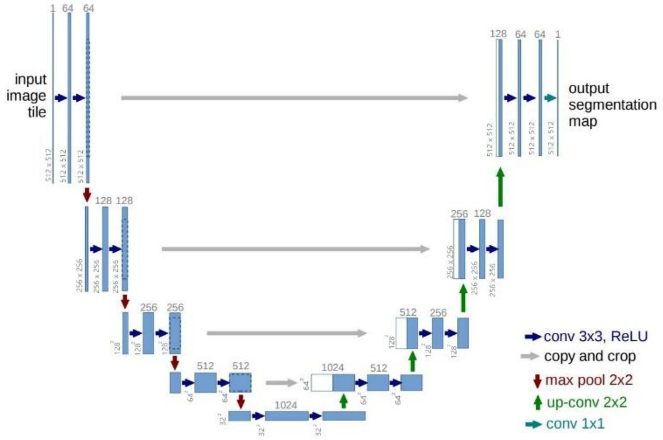

Fig. 1. Structure of the U-Net model (modified from [7]).

**Loss Function:** As with most medical scans, the tumor of interest usually occupies only a very small region in an image. If the cross-entropy loss, as in [7], is used for learning, the final segmentation map will tend to be the background. In our network, we use the dice loss [8], which is based on dice coefficient, as our training loss function. The reason for using the dice loss is because it focuses on foreground regions during training. The dice coefficient loss function $D$ is given as follows:

$$D = \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i}, \quad (1)$$

where $N$ is the number of pixels in a CT slide, and $p_i$ and $g_i$ are the predicted binary segmentation result and the binary ground truth for pixel $i$ in the CT slide. With the use of Pytorch, backward propagation can be implemented easily.

**Training:** In the training stage, we firstly extracted slices from the dicom files provided in the data set. There are approximately 100 slices for each patient. We filtered all the slices and only kept those that contained a tumor for the training purpose. The information about all the tumor positions was saved in terms of the *x* and *y* coordinates. We drew the masks for all the images, according to the tumor coordinates, and each mask and the corresponding image formed a training pair. In order to accelerate the training and maximize GPU memory usage, we trained the network with 100 epochs and a batch size of 10. The validation data set was subsequently used to confirm our results by calculating the dice coefficient so as to avoid overfitting. To improve the final result, we adopted the Adam optimizer [9], with a learning rate of 0.01.

**Image Augmentation:** To train the network, we had the lung slices from 260 different patients. Due to the limited number of samples in the training data set, data augmentation was adopted to enlarge the size of the training data set. The size of the images is 512×512. We randomly cropped the boundary of the images with a random boundary size and then resized them to 512×512. The images were also rotated and flipped. The total number of samples for training, after augmentation, is 70,000.

## B. *Radiomic Analysis*

Radiomics [1], in the field of medical study, refers to the use of a large number of quantitative features from medical images for the comprehensive quantification of tumor phenotypes. In [1], 440 features, which were extracted from the CT data of 1,019 patients with lung and head-and-neck cancer, were used to quantify tumor image intensity, shape, and texture. As there are a number of false positives in the segmentation results obtained by the U-Net deep learning model, a radiomic analysis is important for removing the false-positive results and preserving the true-positive results as many as possible. In our study, we used different visual or deep features to represent tumor regions and then employed a classifier to determine whether a segmented region was a tumor or not.

It is very time consuming if many features are to be extracted from the images. In our study, we first used some of the features in [1], as well as the features used for facial-image analysis [10]. All of these are handcrafted features, whose performance for tumor detection and classification were found not optimal. In order to achieve better accuracy, deep features were applied for radiomic analysis. Therefore, we used a deep neural network to determine if each segmented region from U-Net was a true tumor region or not. Since different network models have different characteristics and levels of performance in tumor classification, we have evaluated the AlexNet [11] and ResNet-18 [12] neural network models for our task. We found that the pre-trained ResNet-18 can achieve the best performance for removing the false-positive results and, at the same time, keeping the true-positive results.

**ResNet-18 Classification:** The classifier is the main component in a statistical pattern-recognition system. The regions, where the deep features will be extracted and used for classification, depend on the segmentation results from U-Net. Fig. 2 shows two segmentation results generated by U-Net. We consider a 170×170 square window centered at each of the segmented regions, where deep features are extracted and then classified. The size of the window is 1.15 times the largest size of the tumor in the data set, which is large enough to cover all of the tumors. With the original ResNet-18 structure, the number of the output of its last fully-connected layer is changed from 1000 to 2 to classify whether a region has a tumor or not.

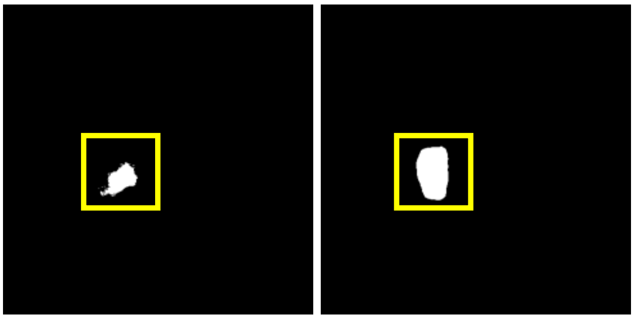

Fig. 2. The region surrounding a potential segmented tumor is considered for feature extraction and classification, for removing false positives as well as preserving true positives. The window size is 170×170.

## III. Experiment Setup and Results

The data sets used in our experiments come from the 2018 IEEE SPS Video and Image Processing (VIP) Cup competition [13]. The training data set contains the data from 260 patients, with their labels provided. The validation and test data sets contain 40 subjects each, but only the validation data set is provided with labels. In our experiments, we evaluated the performance of our proposed approach, with the U-Net only, then with radiomic analysis based on AlexNet and ResNet-18. Six evaluation metrics were measured – the dice coefficient, mean surface distance, 95% Hausdorff distance, slice-wise missing rate, false-alarm rate, and CT-scan-based accuracy. Their definitions are given in the Appendix.

### *A. Results based on U-Net*

We evaluated our model with the validation data set by calculating the dice coefficient, mean surface distance and 95% Hausdorff distance, as well as the slice-wise missing rate and false-alarm rate. Since it is impossible to calculate the first three metrics on those slices without a tumor, we only focused on the slices with a tumor. The results, in terms of dice coefficient, mean surface distance, 95% Hausdorff distance, slice-wise missing rate, and false-alarm rate, based on U-Net only, were 0.547, 12.505, 29.336, 25.4%, and 33.9%, respectively. The slice-wise missing rate is defined as the percentage of missed detection of slices with a tumor. Although the slice-wise missing rate is 25.4%, we can make the decision whether a patient has a tumor or not based on all the slices in a CT scan. For the CT scan of a patient, if there are a certain number of consecutive slices containing a predicted mask and the Intersection-over-Union (IoU) between the masks of two adjacent slices reaches a certain degree, then the patient is considered to have a tumor. In this case, by considering 3 consecutive slices, the accuracy is equal to 90.0%, or the missed detection is reduced from 25.4% to 10%. We call this accuracy CT-scan-based accuracy.

According to the above results, we can find that the false-alarm rate is very high. Deep feature extraction and classification are necessary to further improve the overall performance in the next step.

### *B. Radiomic Analysis based on AlexNet and ResNet-18*

In our study, we used pre-trained AlexNet and ResNet-18 to perform deep feature extraction and classification to remove false positives. To fine-tune the pre-trained AlexNet and ResNet-18, we kept all the positive samples and randomly selected negative samples. The ratio of positive samples and negative samples was 1:3. Adam optimizer [9] with a learning rate of 0.0001 was adopted to improve the result.

By applying the trained AlexNet and ResNet-18 to the segmentation results from U-Net, the dice coefficient, mean surface distance, 95% Hausdorff distance, slice-wise missing rate, false-alarm rate, and CT-scan-based accuracy are 0.611, 8.137, 18.143, 30.3%, 11.1%, and 85.0%, respectively, for AlexNet, and 0.592, 8.835, 21.176, 26.4%, 15.6% and 85.0%, respectively, for ResNet-18, respectively. With the deep feature extraction and classification performed by pretrained network, we can see that the false-alarm rate is reduced significantly, while the missing rate is almost the same.

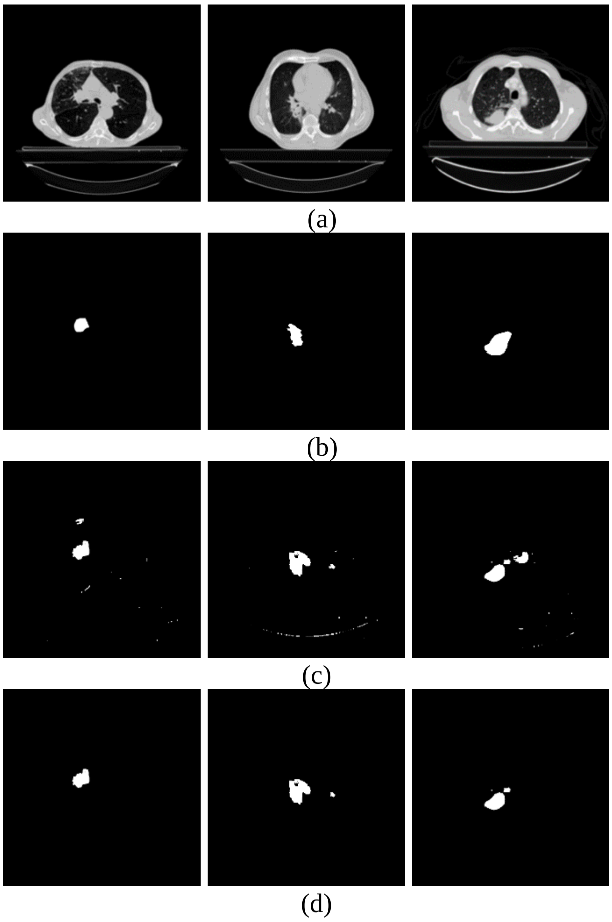

Fig. 3. Segmentation results: (a) original images, (b) ground truth, (c) results generated by U-Net, and (d) results after radiomic analysis.

Although the missing rate and false-alarm rate can both reflect the accuracy of the algorithm, a low missing rate is more important than a low false alarm rate when evaluating a tumor detection algorithm. This is because missing a true tumor would cause severe harm to a patient, compared to detecting a false tumor. As ResNet-18 can achieve a lower missing rate and a higher accuracy, when compared to AlexNet, ResNet-18 is adopted in our algorithm.

Fig. 3 shows some of the segmentation results on the validation data set. The first row shows the images of some CT scans. The second row illustrates the ground truth, in the form of a binary mask, of the tumor regions in the images. The third row shows the segmentation results, generated by our U-Net. When compared to the ground truth, some false positives can be observed. The fourth row shows the results after a radiomic analysis based on ResNet-18. Some false positives were removed.

### *C. Fine-tuning U-Net with Radiomic Analysis*

The result from radiomic analysis shows that the classifier can significantly reduce the false-alarm rate by correctly removing some false-positive masks. To improve the accuracy, we fine-tune U-Net again to increase the true-positive rate, while the false-alarm rate remains unchanged. We mainly focus on adjusting the threshold, used in U-Net, for deciding whether an output pixel is a tumor pixel or not. As shown in Fig. 4, by reducing the value of the threshold, we can reduce the slice-wise missing rate. However, there are more false positives. With ResNet-18, most of the new false-positive masks can be removed, so the false-alarm rate can be maintained. After evaluating the results, we found that by using ResNet-18 and setting the U-Net threshold to 0.0001, our algorithm can achieve the best overall performance. The overall results on the validation data set, in terms of dice coefficient, mean surface distance, 95% Hausdorff distance, slice-wise missing rate, false-alarm rate, and CT-scan-based accuracy, are 0.563, 10.896, 26.052, 23.3%, 24.6%, and 90.0%, respectively. The performance of our proposed method using U-Net only, U-Net and AlexNet, and U-Net and ResNet-18, with the threshold set at 0.5 and 0.0001 are tabulated in Tables 1 and 2, respectively.

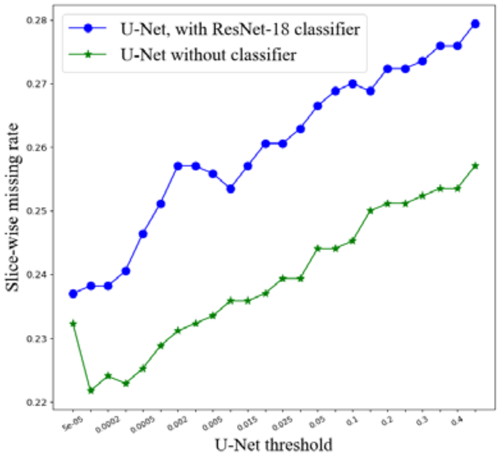

Fig. 4. Slice-wise missing rate with different U-Net thresholds.

TABLE. 1 PERFORMANCE OF OUR PROPOSED ALGORITHM AT DIFFERENT STAGES WITH THRESHOLD OF U-NET = 0.5

| | U-Net only | U-Net and AlexNet | U-Net and ResNet-18 |
|---|---|---|---|
| Dice coefficient | 0.547 | 0.611 | 0.592 |
| Mean surface distance | 12.505 | 8.137 | 8.835 |
| 95% Hausdorff distance | 29.336 | 18.143 | 21.176 |
| Slice-wise missing rate | 25.4% | 30.3% | 26.4% |
| False-alarm rate | 33.9% | 11.1% | 15.6% |
| CT-scan-based Accuracy | 90.0% | 85% | 85% |

TABLE. 2 PERFORMANCE OF OUR PROPOSED ALGORITHM AT DIFFERENT STAGES WITH THRESHOLD OF U-NET = 0.0001

| | U-Net only | U-Net and AlexNet | U-Net and ResNet-18 |
|---|---|---|---|
| Dice coefficient | 0.475 | 0.588 | 0.563 |
| Mean surface distance | 27.014 | 9.336 | 10.896 |
| 95% Hausdorff distance | 75.978 | 21.243 | 26.052 |
| Slice-wise missing rate | 22.2% | 28.7% | 23.3% |
| False-alarm rate | 75.2% | 17.6% | 24.6% |
| CT-scan-based Accuracy | 95.0% | 82.5% | 90% |

## IV. CONCLUSION

We have implemented a method, based on a deep neural network and a radiomic analysis, for segmenting the tumor regions in lung CT scans. We have evaluated different deep models, including U-Net, DeepLab, Mask RCNN, etc., and found that U-Net can achieve the best performance. For the radiomic analysis, different types of features for shape and texture representation, features used for recognition, and deep features, have all been considered. Experimental results show that the deep feature can achieve the best performance. After evaluated several deep neural networks for classification, we found that ResNet-18 can achieve the best classification performance. With the use of the best methods evaluated, our method can achieve the accuracy of 90%.

## REFERENCES

[1] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Cavalho, J. Bussink, et al. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach,” Nature Communications, 5 (1), 2014.

[2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, June 2017.

[3] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg, “SSD: Single Shot MultiBox Detector,” in the Proceedings of IEEE International Conference on Computer Vision 2016.

[4] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in the proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2016.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick, “Mask R-CNN,” in the Proceedings of IEEE International Conference on Computer Vision 2016, pp. 2961 – 2969.

[6] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834-848, 2018.

[7] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, Springer, pp. 234–241. 2015.

[8] O. Milletari, N. Navab, and S. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” arXiv preprint arXiv:1606.04797, 2016.

[9] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in the Proceedings of ICLR, 2015.

[10] Cigdem Turan and Kin-Man Lam, “Histogram-based Local Descriptors for Facial Expression Recognition (FER): A comprehensive Study,” Journal of Visual Communication and Image Representation, vol. 55, Pages 331-341, August 2018.

[11] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” In Advances in Neural Information Processing Systems, 25, pp. 1106–1114, 2012.

[12] He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[13] 2018 IEEE Video and Image Processing Cup (VIP-Cup): https://users.encs.concordia.ca/~i-sip/2018VIP-Cup/index.html.

## Appendix - Evaluation Metrics

To evaluate the performance, we used 6 evaluation metrics: Dice Score, Mean Surface Distance, 95% Hausdorff Distance, Slice-wise Missing Rate, False-Alarm Rate, and CT-scan-based Accuracy. The overall score for each of these metrics is calculated in our experiments.

### *A. Dice Coefficient*

Dice coefficient calculates two times the intersection area of the ground truth and the predicted tumor regions. Dice coefficient will only consider those slices with a tumor in the ground truth. For those slices without predicted tumor regions, the dice coefficient for that single image will be counted as 0. The equation for those slices with a tumor in both the ground truth and the predicted image utilized the following equation:

$$D = \frac{2|X| \cap |Y|}{|X| + |Y|}$$

### *B. Mean Surface Distance and 95% Hausdorff Distance:*

Because of the definition of these two metrics have its limitation, we only consider those slices having tumor in both the ground truth and the predicted mask.

The mean surface distance is calculated through the following two equations:

$$\vec{d}_{H,avg}(X,Y) = \frac{1}{|X|} \sum_{x \in X} \min_{y \in Y} d(x,y)$$

$$d_{H,avg}(X,Y) = \frac{\vec{d}_{H,avg}(X,Y) + \vec{d}_{H,avg}(Y,X)}{2}$$

The 95% Hausdorff Distance is calculated through the following two equations:

$$\vec{d}_{H,r}(X,Y) \; = K_r(\min_{y \in Y} d(x,y)) \; \forall x \in X$$

$$d_{H,r}(X,Y) \; = \frac{\vec{d}_{H,r}(X,Y) + \vec{d}_{H,r}(Y,X)}{2}$$

### *C. Slice-wise Missing Rate:*

A slice is counted as a miss if it has a tumor but cannot be detected, or if the value of Intersection-over-Union (IoU) between the detected mask and ground truth is lower than 0.1. We have utilized the following equation to calculate the missing rate:

$$\mathrm{MR} \; = \frac{\text{number of false negative masks in a CT scan}}{\text{number of masks with a tumor in the CT scan}}.$$

### *D. False-Alarm Rate (FAR) or False-Positive Rate:*

The false-alarm rate is a measure of the number of slices that have no tumor but are detected to have a tumor. It is calculated as follows:

$$\mathrm{FAR} = \frac{\text{no. of false positive masks in a CT scan}}{\text{no. of ground} - \text{truth masks without tumor in a CT scan}}$$

### *E. CT-scan-based Accuracy*

For each patient, if 3 consecutive slices are detected as true positives, and the IoU value between any two of the adjacent slices is larger than 0.5, then the patient or the CT scan is declared to have a tumor. The accuracy rate is determined using the following equation:

$$\text{Accuracy} = \frac{\text{number of patients with tumor detected}}{\text{number of patients with a tumor}}$$