Region Proposal Network for Lung Nodule Detection and Segmentation

Mohammad Hesam Hesamian,[1] Wenjing Jia, Xiangjian He, Paul Kennedy

Abstract. Lung nodule detection and segmentation play a critical role in detecting and determining the stage of lung cancer. This paper proposes a two-stage segmentation method which is capable of improving the accuracy of detecting and segmentation of lung nodules from 2D CT images. The first stage of our approach proposes multiple regions, potentially containing the tumour, and the second stage performs the pixel-level segmentation from the resultant regions. Moreover, we propose an adaptive weighting loss to effectively address the issue of class imbalance in lung CT image segmentation. We evaluate our proposed solution on a widely adopted benchmark dataset of LIDC. We have achieved a promising result of 92.78% for average DCS that puts our method among the top lung nodule segmentation methods.

key words: Nodule segmentation, Deep learning, Region proposal network

1 Introduction

Lung cancer, as one of the deadliest cancers, is responsible for major cancer deaths worldwide [16]. Early detection and accurate classification of the lung nodules plays a significant role in increasing the survival rate of the patients. Manual process of detection and segmentation of lung nodule is a challenging task which requires lots of time, proficiency and yet various types of errors may occur. With the increasing growth in the availability of medical images such as computed tomography (CT) scans, automatic nodule detection and segmentation has become a reliable tool to help the radiologist in their tough task of lung image analysis.

Recently, convolutional neural networks (CNNs) have shown the capability to effectively extract image features for successful pattern detection and segmentation across a variety of situations from scene to medical images [18, 3, 4]. Similarly, deep learning approaches have been used for various tasks of medical image analysis, including organ detection, lesion classification and tumour segmentation [13, 6, 7]. Among all those applications, lung nodule segmentation is known to be a challenging task due to the heterogeneous appearance of the lung tumour and also the great similarity between tumour and non-tumour substances in the lung area.

Another severe challenge in medical image segmentation is to deal with the class imbalance [20]. In a fully annotated CT image of the lung, the area occupied by tumour is much smaller than the rest of the lung. This is due to the sparse distribution of pulmonary nodules in the lung. This issue gets more severe when we are performing a semantic segmentation task in which each pixel is considered as one sample. In such a case, the number of samples (pixels) corresponding to the tumour is significantly lower than the rest of the lung area. The class imbalance issue affects the fully convolutional networks [23] more than others. Moreover, the ratio of tumour pixels to background pixels significantly varies from sample to sample. To address this problem, we propose an adaptive weighted loss to alleviate the class imbalance issue at the sample level.

The main contributions of this paper can be summarized in three folds. First, we propose a two-stage network to reduce the dependency on dense sliding window searching for accurate segmentation of the lung nodules. Second, we propose an adaptive weighted loss function to address the class imbalance issue to improve segmentation accuracy. Third, we perform a far distant transfer learning strategy in which the weights are transferred from a general object detection model.

2 Related works

Two major categories of studies have explored lung image segmentation. The first one is the whole lung segmentation, and the second is the lung tumour segmentation. Whole lung segmentation aims to distinguish the border of the entire lung from the rest of the elements appearing in a cardiac CT image [2, 5]. It helps to determine the size and shape of the lung. It is also often used as the first stage for the lung tumour segmentation with the purpose of reducing the false positive cases caused by non-lung areas of CT image.

Recently, many CAD systems based on deep learning are proposed for automatic lung cancer detection. For example, ZNET [24] employed U-Net fully convolutional network architecture for candidate selection on axial slices. For the subsequent false positive reduction, three orthogonal slices of each candidate were fed to the same wide residual network. Wang et al. in [22] proposed a model that can capture a set of nodule-sensitive features from CT images. The 2D branch of the model learns multi-scale 2D features from 2D patches. In the 3D branch, a novel central pooling layer helped the model to select the features around the target voxels effectively. In another study, a region CNN (R-CNN) is proposed for lung nodule segmentation from 3D lung patches [23]. In this model, Deep Active Self-paced Learning (DASL) was introduced to reduce the dependency of the network to fully annotated data. It utilized unannotated samples by taking to account both the knowledge know before training and the knowledge made during the training. Jiang et al. [12] proposed a residually connected multiple resolution network, which was able to combine the features in various resolution inputs simultaneously. Images with different resolution were passed through two separate residual networks, and the extracted features were refined and concatenated. This technique helped them to improve the localization

[1] School of Electrical and Data Engineering, University of Technology Sydney. Email: mh.hesamian@gmail.com

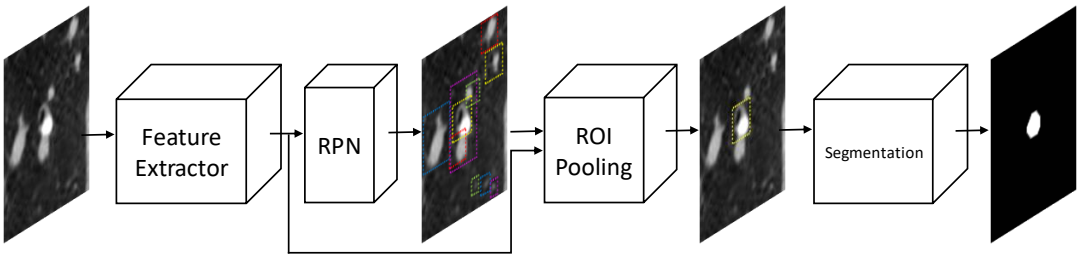

Figure 1. The proposed network model

and negative effect of multiple pooling operation.

Generally, 3D models have a huge demand for memory and high processing cost. Due to these limitations, the algorithms that can be implemented on 3D image analysis are restricted. Moreover, CT scans usually have different slice thicknesses which are not recommended to be treated uniformly in a 3D model [24]. Moreover, 2D models require fewer resources for training and also not affected by the slice thickness.

In two-stage detections methods, region proposal networks (RPN) attracted lots of attention. RPN was first introduced to general object detection tasks by [19]. In the proposed structure, there is a classifier which determines the probability of having a target object at the anchor. Then the regression regresses the coordinates of the proposal. The candidature multiple anchors and sharing the features make the RPN efficient in time and detection. Employing the bounding box regression enables the RPN to produce more precise proposals. RPN has many successful variation and application[17, 15] yet it has not been explored adequately in lung nodule segmentation task.

According to all shortcomings of the current methods mentioned above, and the enormous potential of the RPN networks, we were motivated to design and develop a model to combine the benefits of two-stage detection and segmentation models with higher segmentation accuracy.

3 The Proposed Method

There are several image modalities, which can be used for lung abnormality detection such as PET, SPECT, X-Ray, MRI and CT. PET and SPECT are mainly utilized for metabolism characterization. Therefore, they can unveil the functional abnormalities in an organ. CT, X-Ray and MRI are structural image modalities which are able to hand out anatomical information about the organ. The simplicity and lower price of CT scan have changed it to primary image modality for cancer detection and screening [25]. For this study, we used the benchmark dataset of LIDC-IDRI [1], consisting of 1024 chest CT scans. Each case is associated with an XML file, denoting the boundary pixels of the tumour, marked by four radiologists. For this study, the nodules with size $>$ 3mm are selected. These scans are broken down to slices and converted to JPG format in the original size.

3.1 Network Structure

There are some two-stage nodule detection systems, where in the first stage the nodule candidates are detected, and in the second stage, the false positive is reduced [24]. But in our approach, inspired by the development of general object detection [8], we propose a two-stage method capable of nodule detection and segmentation. The model segments the elements of CT scan into two classes of tumour and background. The general building block of the network is presented in Fig. 1. The first section of the network is the feature extractor in which general feature maps are created. The ResNet101 [9] is used as the backbone of the system. The residual connections of the ResNet allow using deeper structures without facing the gradient vanishing problem.

The extracted feature maps are then passed to the RPN. By this method, the convolution layers are reuses. Thus, it saves lots of computations. This module produces several candidate regions, containing the potential malignant tumours. Since our model has only two classes of tumour and background, we limit the number of proposed regions to 300. Having a limited number of regions will help to reduce the possibility of false positive as well as speed up the training process. Each of the candidate bounding boxes is then given to the ROI pooling. At this stage, the network evaluates the proposed regions according to intersection over union (IoU) value. The IoU is set to 0.5 to reduce the false positive. It means a fixed number of the ROIs (50) with the IoU more than 0.5 will be passed to the segmentation module for semantic segmentation task. In case of no ROI satisfying the condition, the input sample is considered as a negative sample. The last block of the network will produce the segmentation mask for the proposed ROI.

3.2 Adaptive Loss Coefficients

One of the main reasons of accuracy drop in nodule segmentation is the class imbalance of the training samples, due to which, during the training model will not learn equally from all classes. Therefore, one of the main challenges in medical image analysis is how to effectively modify the model to overcome the class imbalance issue and maximize the learning capability of the model.

In the segmentation section of the proposed model, samples will be segmented in the two classes of background and tumour. Thus, we applied a binary cross-entropy function as the segmentation loss. This is shown in Eq. 1.

$$L(S) = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log \hat{y_n} + (1-y_n)\log(1-\hat{y_n})\right] \quad (1)$$

In this equation, y and $\hat{y}$ represent the ground truth and the predicted value of each pixel, and N is the total number of the pixels in the given sample of S.

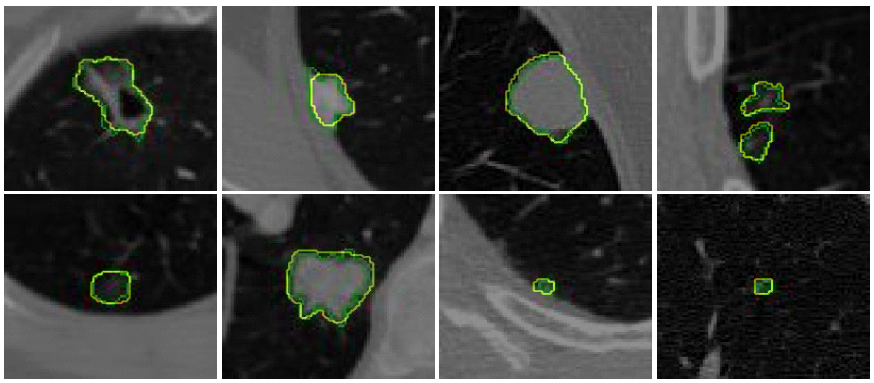

Figure 2. Visualization of the network output. The green color is the ground truth marked by the radiologists. The yellow is the prediction, generated by the proposed model. Best viewed in color.

Then the dynamic loss coefficients of γ_1 and γ_2 are calculated for each of the given ROIs representing the proportion of tumour and background pixels. These coefficients compensate the imbalance between the numbers of pixels in the two classes.

Since in our model, each pixel is considered as one sample, the loss will be calculated at the pixel level. Therefore, the final loss function of Eq. 2 is formed by applying the dynamic loss coefficient to the Eq. 1. The loss is calculated for each pixel of p_i, where p_i denotes the i^{th} pixel of a given flattened ROI.

$$Loss(p_i) = \begin{cases} \gamma_1 \times \Big[- y(p_i) \log \hat{y}(p_i) \Big], & \text{if } p_i \in tumour \\ \gamma_2 \times \Big[- y(p_i) \log(1 - \hat{y}(p_i)) \Big], & \text{if } p_i \in bg \end{cases} \tag{2}$$

The proposed weighted loss updates its coefficients for each ROI proposed by ROI selection module. This process helps to address the class imbalance more accurately because the proportion of tumour to background varies significantly from sample to sample. The common weighting loss such as what proposed in [11] applies a fixed set of the coefficient for all the samples. These coefficients are calculated by counting and averaging the tumour pixel and background pixels over the entire dataset. Application of a fixed set of weight to all the samples does not seem to be an optimal solution while the adaptive weighted loss allows the network to address the class imbalance more effectively within individual samples.

4 Experiments and Discussion

In this section, we experimentally evaluated our solution on the widely used benchmark dataset LIDC [1]. Our model is implemented with the Keras library, and all the experiments are conducted on Linux (REH7.0) with an Nvidia Quadro P5000 GPU of 16 GB memory.

4.1 Data preparation

We evaluated our solution on the publicly available dataset of LIDC [1], which contains 1024 lung scans each annotated by at least four radiologists. All data are selected and used for training and testing of the network with the 5-fold cross validation approach.

Similar to most of the other medical datasets, our data was also imbalanced toward the tumour class. Training a model on such data will lead to learning more features from one class and ignoring the others. As another technique to combat this issue, we have selected a patch-wise training strategy. For this purpose, we have extracted the patches around the tumour and use them as training samples. According to nodule size distribution statistics, 85% of nodules can be covered by a patch of 30×30 (voxels), and 99% of nodules can be covered by a 40×40 (voxels) patch [24]. Hence, we have extracted the patches of 76×76 pixel to ensure almost all of the tumours are covered. At the next stage, we heavily augment our training data by using filipping, rotating, zooming and shrinking the input samples. Data augmentation is reported that data augmentation can also help to improve the performance and robustness of CNNs [10]. This data augmentation serves two purposes of avoiding the overfitting and extending the training dataset size.

4.2 Training

We have employed a two-step full network adaptation strategy to transfer the weights from a far distant source. In this process, the weights are initialized from a pre-trained model trained on COCO [14] dataset for the general object detection task. Transferring the weights from such model and using them for a different task of medical object segmentation, is called 'far transfer learning'. Generally, transfer learning is proven to alleviate the overfitting issue on the small training dataset and improve the convergence speed of the training. Theoretically, transfer learning has better performance when the task of source and target models are more similar [10]. Thus, some believe that far transfer learning may not produce good results [21] but, our achieved results are demonstrating that far transfer learning combined with a careful fine-tuning strategy can deliver competitive results.

As a part of our weight transferring, the weights for the feature extractor layers are initialized from the pre-trained model, and the last layers are initialized randomly. During the first stage of training, all the network weights except for the feature extractor weights are fixed. At this stage, only the feature extractor is trained on the input data for a couple of epochs. The rest of the network weights are injected at the second stage of the training, and all the network layers are trained together afterwards.

Table 1. The detection rate of nodules of various sizes, obtained with our proposed approach.

Size	Measurements (%)			
	IoU=0.3	IoU=0.4	IoU=0.5	IoU= 0.6
Tumour size <10mm	97.54 ± 0.56	95.81 ± 0.67	92.32 ± 1.07	84.76 ± 1.55
$10 <$ Tumour size < 30 mm	98.13 ± 0.72	97.56 ± 0.64	95.57 ± 1.06	92.04 ± 0.95
Tumour size > 30mm	95.91 ± 4.56	94.25 ± 4.21	93.26 ± 4.46	90.16 ± 7.95
Average	97.21 ± 1.95	95.87 ± 1.84	93.73 ± 2.19	88.99 ± 3.78

5-fold cross-validation is employed to validate the results of the testing. The data set is divided into five equal subsets, and each time four of them are used for training, and the one remaining is used for testing. The LIDC dataset has not a separate set of train and test set. Hence, we divided the data into two main sets of train and test for the purpose of 5-fold cross-validation. Moreover, a portion of training data (10%) has already been used for validation at the end of each epoch. By this method, we will ensure that all the samples are used at least once for training and testing.

4.3 Experiment results

Fig. 2 shows the qualitative results of tumours detected and segmented for various tumour types and sizes.

To highlight the performance of our proposed model, we quantitatively evaluate its performance for two tasks, *i.e.,* detection and segmentation, respectively.

4.3.1 Results of tumour detection

Our proposed method performs the detection task through a segmentation approach. This detection method differentiate our model from the pure detection models. For the detection part, we measure the detection accuracy of the tumours under various IoU values of 0.3, 0.4, 0.5 and 0.6. Table 1 shows the results of tumour detection for three tumour size categories. The results are presented with various IoU values to analyse the detection performance clearly. For instance, in the case of $IoU = 0.5$, if the IoU of prediction mask with the ground truth is more than 0.5, the tumour is considered as correctly detected. As expected, the detection accuracy is dramatically increased when the tumour size increased.

The results in Table 1 highlights the significant detection accuracy of the proposed model in detecting the small size tumours.

4.3.2 Results of tumour segmentation

To evaluate the segmentation performance, we measure the Dice score of segmentation (DCS), sensitivity and positive predictive value (PPV) as defined as in Eqs. 3, 4 and 5, respectively, as:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

and

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

TP and FP refer to true positive and false positive while TN and FN denote the true negative and false negative.

The results presented in the Table 2 show that the proposed method is able to perform the challenging task of lung nodule segmentation with higher accuracy than the state-of-the-art methods. The dice score of our segmentation model is almost 10% higher than the listed methods. Similarly, the precision of our method is much higher than all the listed methods. In case of the recall, we got slightly (less than 1 percent) lower than CF-CNN, but CF-CNN produced wider standard deviation. It shows that our model was more stable throughout all the input cases.

4.3.3 Discussion

The main reason for this improvement is the structure of the network. In the proposed structure, at the first stage, some ROIs are extracted as the potential tumour. Then, the most tumour-like ROIs are selected through the ROI pooling module. Then the second stage performs the segmentation on the selected ROIs. This two-step strategy eliminates the necessity to scan the entire input image via a sliding window that leads to creating the prediction at every position. By performing these two steps, many of the nodule-like patterns which may confuse the model are eliminated. Therefore, the accuracy of segmentation is improved by only focusing on the more relevant tumour features. Moreover, the training and testing speed will increase as there would be no full input search via the sliding window.

The second reason for the improved accuracy is the application of novel adaptive loss in training of the segmentation module. This adaptive loss helps to address the class imbalance issue of the medical images within the individual sample. By application of a dynamic pair of class weight coefficients, the model can derive more balanced features from each sample.

The transfer learning strategy used for training of the model helped the model to be more stable and prevent it from overfitting. Moreover, one of the reasons for achieving a smaller standard deviation value compared with other studies could be the application of transfer learning.

5 Conclusion

We have proposed a two-stage framework for accurate segmentation of the lung nodules. The first stage of the network provides some potential nodule areas. The second stage accurately segments the selected ROIs. To effectively address the class imbalance issue of the small organ segmentation, we proposed an adaptive weighted loss function, where the weight coefficients are calculated per sample. This approach leads to extract more accurate features and therefore, more precise segmentation of the input. The model was tested on the

Table 2. Quantitative evaluation of the achieved results and comparison. Models marked with * use 3D processing.

Methods	Measurements		
	Dice (%)	Recall (%)	Precision (%)
Nodule R-CNN* [23]	64 ± 0.44	-	-
Hesamian et al. [11]	81.24 ± 1.24	-	79.75 ± 4.08
CF-CNN* [22]	80.47 ± 10.76	92.75 ± 12.83	75.84 ± 13.14
Jiang et al. [12]	68 ± 0.23	85 ± 0.13	67 ± 0.22
Proposed method	92.78 ± 0.1	92.31 ± 0.27	93.17 ± 0.18

publicly available dataset of LIDC and was able to deliver an average detection accuracy of 93.73% (IoU = 0.5) and average dice score of 92.78% for the segmentation part. At last, the achieved results demonstrate that a far distant transfer learning with careful weight initialization can perform competitive results.

REFERENCES

[1] Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric 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).

[2] Geng Chen, Dehui Xiang, Bin Zhang, Haihong Tian, Xiaoling Yang, Fei Shi, Weifang Zhu, Bei Tian, and Xinjian Chen, 'Automatic pathological lung segmentation in low-dose ct image using eigenspace sparse shape composition', *IEEE transactions on medical imaging*, (2019).

[3] 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 transactions on pattern analysis and machine intelligence*, **40**(4), 834–848, (2018).

[4] Noel CF Codella, David Gutman, M Emre Celebi, Brian Helba, Michael A Marchetti, Stephen W Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, et al., 'Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging ISBI, hosted by the international skin imaging collaboration ISIC', in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, pp. 168–172. IEEE, (2018).

[5] Yu Gordienko, Peng Gang, Jiang Hui, Wei Zeng, Yu Kochura, O Alienin, O Rokovyi, and S Stirenko, 'Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-Ray analysis of lung cancer', in *International Conference on Theory and Applications of Fuzzy Systems and Soft Computing*, pp. 638–647. Springer, (2018).

[6] Nazia Hameed, Antesar M Shabut, Miltu K Ghosh, and M Alamgir Hossain, 'Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques', *Expert Systems with Applications*, **141**, 112961, (2020).

[7] Sardar Hamidian, Berkman Sahiner, Nicholas Petrick, and Aria Pezeshk, '3D convolutional neural network for automatic detection of lung nodules in chest CT', in *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, p. 1013409. International Society for Optics and Photonics, (2017).

[8] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, 'Mask r-cnn', in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, (2017).

[9] 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*, pp. 770–778, (2016).

[10] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy, 'Deep learning techniques for medical image segmentation: Achievements and challenges', *Journal of digital imaging*, 1–15, (2019).

[11] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul J Kennedy, 'Atrous convolution for binary semantic segmentation of lung nodule', in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1015–1019. IEEE, (2019).

[12] Jue Jiang, Yu-Chi Hu, Chia-Ju Liu, Darragh Halpenny, Matthew D Hellmann, Joseph O Deasy, Gig Mageras, and Harini Veeraraghavan, 'Multiple resolution residually connected feature streams for automatic lung tumor segmentation from ct images', *IEEE transactions on medical imaging*, **38**(1), 134–144, (2018).

[13] Yang Lei, Tonghe Wang, Sibo Tian, Xue Dong, Ashesh B Jani, David Schuster, Walter J Curran, Pretesh Patel, Tian Liu, and Xiaofeng Yang, 'Male pelvic multi-organ segmentation aided by cbct-based synthetic mri', *Physics in Medicine & Biology*, **65**(3), 035013, (2020).

[14] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, 'Microsoft coco: Common objects in context', in *European conference on computer vision*, pp. 740–755. Springer, (2014).

[15] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg, 'Ssd: Single shot multibox detector', in *European conference on computer vision*, pp. 21–37. Springer, (2016).

[16] Kimberly D Miller, Ann Goding Sauer, Ana P Ortiz, Stacey A Fedewa, Paulo S Pinheiro, Guillermo Tortolero-Luna, Dinorah Martinez-Tyson, Ahmedin Jemal, and Rebecca L Siegel, 'Cancer statistics for hispanics/latinos, 2018', *CA: a cancer journal for clinicians*, **68**(6), 425–445, (2018).

[17] Joseph Redmon and Ali Farhadi, 'Yolo9000: better, faster, stronger', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271, (2017).

[18] Joseph Redmon and Ali Farhadi, 'Yolov3: An incremental improvement', *arXiv preprint arXiv:1804.02767*, (2018).

[19] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, 'Faster r-cnn: Towards real-time object detection with region proposal networks', in *Advances in neural information processing systems*, pp. 91–99, (2015).

[20] Skylar Stolte and Ruogu Fang, 'A survey on medical image analysis in diabetic retinopathy', *Medical Image Analysis*, 101742, (2020).

[21] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri, 'Deep end2end voxel2voxel prediction', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 17–24, (2016).

[22] Shuo Wang, Mu Zhou, Zaiyi Liu, Zhenyu Liu, Dongsheng Gu, Yali Zang, Di Dong, Olivier Gevaert, and Jie Tian, 'Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation', *Medical image analysis*, **40**, 172–183, (2017).

[23] Wenzhe Wang, Yifei Lu, Bian Wu, Tingting Chen, Danny Z Chen, and Jian Wu, 'Deep active self-paced learning for accurate pulmonary nodule segmentation', in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 723–731. Springer, (2018).

[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).

[25] Junjie Zhang, Yong Xia, Hengfei Cui, and Yanning Zhang, 'Pulmonary nodule detection in medical images: a survey', *Biomedical Signal Processing and Control*, **43**, 138–147, (2018).