One Stage vs Two Stage Detectors: Which one is better for lung nodule detection in CT Images ?

Luís Fernandes[1,2]
luis.e.pereira@inesctec.pt
Hélder P. Oliveira[1,3]
helder.f.oliveira@inesctec.pt

[1] INESC TEC - Institute for Systems and Computer Engineering, Technology and Science 4200-465 Porto, Portugal
[2] FEUP - Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal
[3] FCUP - Faculty of Science, University of Porto, 4169-007 Porto, Portugal

Abstract

Cancer results from the accumulation of mutations in the genetic material of the cells. Furthermore, due to its evasive nature and sometimes rapid development, it is one of the deadliest diseases worldwide. Among the different types of cancer, lung cancer is one of the deadliest, being responsible for millions of deads worldwide. However, this alarming numbers can be reduced if the lung cancer is detected in its early stages. Therefore, there is a great interest in the scientific community to develop new early detection tools for lung cancer. In this paper, we access two main approaches for the detection of lung cancer in CT images: One Stage Detectors and Two Stage Detectors. Our results show that One Stage Detectors are able to achieve better performance than the Two Stage Detectors.

1 Introduction

Lung cancer is the leading cause of cancer death, resulting in 2 million diagnoses and 1.8 million deaths at a world wild level. Furthermore, projections indicate an increase of lung cancer in the future due to also the increase of tobacco smoking habits in the general population [1]. However, if the cancer is detected in its early stages, the patient has a 5-year survival rate of 95%, while if it is detected in its late stages, the same patient will have a 5-year survival rate of only 10% [2]. During the treatment of such patients, one important step is the accurate segmentation of the lung nodule for the subsequent clinical planing and also the radiotherapy treatment [3]. To do such segmentation, the healthcare professional must manually highlight the lung cancer that was detected in the CT scan. Considering the facts that several patients are admitted daily in the hospital with suspicion of lung cancer, it is possible to see how the segmentation of lung nodules can become a tiresome task.

Deep Learning (DL) models have been used in several health applications to assist in different types of tasks, such as lung nodule detection in CT images. The current frameworks that are employed for this type of task can be categorized in two groups: One Stage Detectors (OSD) and Two Stage Detectors (TSD). TSD are denominated as such because they often have an intermediate step where they predict a rough estimation of the objects localizations and their respective classification. An example of an TSD framework is the Faster-RCNN model [4]. On the other hand, there are models that predict directly the bounding box and classification of each object on the input image, without the need of an intermediate step. These DL models are generally denominated as OSD. A popular example of an OSD is the You Only Look Once (YOLO) model proposed in [5]. Both of these two types of framework present their advantages and disadvantages. For instances, while OSD are a more recent and faster than TSD, the latter is still used in some applications because although they tend to be slower, they also tend to have better performance in some applications [6].

In the current state of the art, it is possible to find several works that employ one of the two aforementioned approaches, achieving interesting results. Nonetheless, the performance of the models is still underwhelming and more research is required into developing a DL model that can be used in a clinical setting. To this end, it is important to understand which of the two aforementioned frameworks has the potential to achieve the best results in the lung nodule detection task. Although different studies already explored both of these approaches, they employ different experimental settings, such has data enhancement, which makes the comparison between studies a challenging task. In this work, we do a comparative study between the TSD and the OSD so that we can understand which framework as the potential to achieve the best results. By knowing which framework achieve the best results, it is possible to develop new DL models with the potential to achieve better performance than the current state of the art.

2 Related Work

Lung nodule detection is a challenging task due to the fact that the nodules are significantly smaller than the CT image. For instances, while the convectional CT slice as a dimension of 512×512 pixels, the lung nodule can be contained in a bounding box of 64×64 pixels. To detect the small nodules in such large area, the current papers in the state-of-the-art attempt to implement OSD and TSD models to execute the object detection task. For example, in [7] the authors adopted the Faster-RCNN model for the detection and classification of lung nodules. Furthermore, the authors attempt to develop a new method so that the Faster-RCNN model learns the anchor box sizes from the ground truth dimensions of the lung nodules by employing adaptive anchor boxes. By using the method established by the authors, there is no need to manually design the anchor sizes or to use simpler clustering methods. The dataset used to train and test such framework was the LUNA16 dataset. Data augmentation such has horizontal flip and random rotation were used to balance out the classes in the dataset. After training, the DL model was evaluated, reaching an Area Under Curve (AUC) of 95.7%. In [8] the authors modified the YOLO-v5 model in an attempt to improve its performance in the lung nodule detection task. The authors proposed three modifications of the YOLO-v5 model architecture: replacing max-pooling with the stochastic-pooling method, multiscale feature fusion with the use of a bidirectional feature fusion pyramid and finally, the authors also improved the training of the YOLO-v5 model by adopting the EIoU loss function. To train and validate such framework, the authors used the LUNA16 dataset. After validation, the proposed framework was able to achieve and mean Average Precision (mAP) of 95.9% which is higher than the Faster-RCNN baseline, which only achieved an mAP of 91.9%. In [9] the authors proposed the I3DR-Net model that implements 3D feature extraction at different scales. The proposed model is composed by three main parts: an inception 3D backbone for feature extraction, a Feature Pyramid Network (FPN) for feature extraction at different scales and then a detection head for the prediction of the final bounding boxes and their respective classification. The I3DR-Net model was trained and evaluated on the publicly available LIDC-IDRI dataset and also the Moscow Private Datasets. The proposed framework was able to achieve an AUC of 81.84% and 69.00% in the LIDC-IDRI and Moscow Private dataset, respectively.

As it is possible to see from the presented works from the state of the art, the development of new frameworks that are based on OSD and TSD is still an ongoing process. However, due to the difference in experimental settings between the presented works, it is hard to make a far comparison between the presented frameworks. Due to this limitation, and to assist future works in choosing between the two aforementioned frameworks as foundation for their DL models, in this paper we study the generally used frameworks in a control experimental setting, to understand which has the best performance.

3 Materials and Methods

The object detection models that were used in this study where: Faster-RCNN, YOLO-v3, YOLO-v5 and YOLO-v8. These models were chosen to compare the performance between OSD and TSD, and also to access if the latest versions of the YOLO model have better performance. The LIDC-IDRI dataset was used to train the aforementioned models [10]. The used dataset has 1018 clinical cases and in total has 1186 annotated nodules. Each clinical case has its CT scan, the labels of each nodule in cancer and non-cancer, the segmentation masks and also the clinical data.

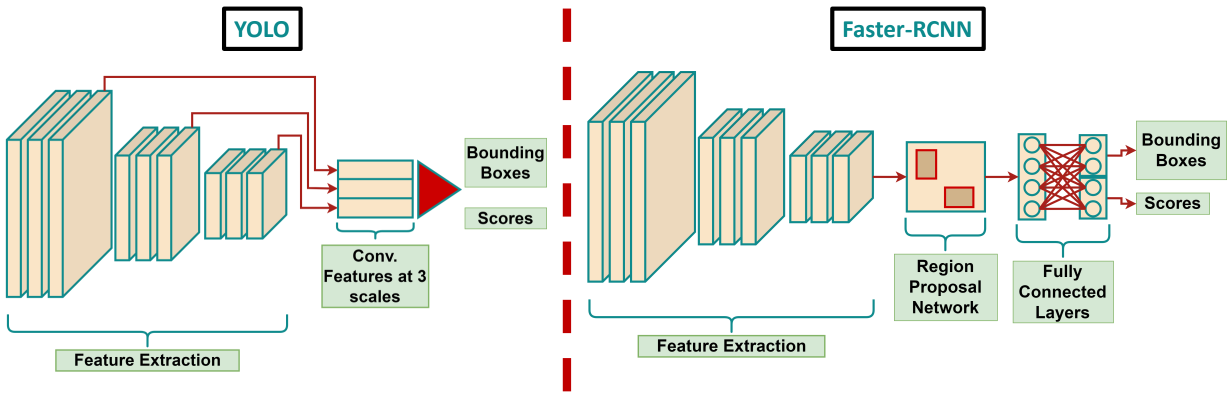

Figure 1: Illustration of the YOLO and Faster-RCNN DL models.

Table 1: Experimented values for each hyperparameter in the DL models.

	Hyperparameter		
Model	Batch Size	Learning Rate	Optimizer
YOLO models	16	0.00001	SGD
Faster-RCNN	8	0.001	SGD

For each model, this dataset was split in training set and also test set. The performance metric that was used to evaluate the performance of the studied DL models was the mAP. An illustration of the models used in this work can be seen in Figure 1. Also, the hyperparameters used in each of the studied DL models can be seen in Table 1.

4 Results and Discussion

After training and testing the aforementioned DL models, it is possible to see their performance metrics in Table 2.

Table 2: Performance of the DL models in the test dataset.

Model	**mAP-0.5 (%)**
YOLOv8	62.05
YOLOv5	58.61
YOLOv3	59.71
Faster-RCNN	37.24

When comparing the mAP of the different models, it is possible to see that the YOLO models tend to have significantly higher performance, than the Faster-RCNN model. However, it is important to notice that the YOLO models are more recent than the Faster-RCNN model and also, there as been more effort in studying different strategies to increase the performance of the YOLO models. Furthermore, this results show that contrary to popular belief, OSD can have a better performance than TSD [6].

5 Conclusions

Although several efforts have already being made in the fight against cancer, there are still a number of applications where DL models can assist healthcare professionals in their tasks. One example of such tasks in the detection of lung nodules. This part of the clinical procedure is not only revenant for the diagnosis of the patient but also for the planing of the possible radiotherapy. Nonetheless, due to the several numbers of patients with suspicion of lung cancer, the detection of each lung nodule on the CT scan of each patient can become a tiresome task. In an attempt to assist the radiologist in such tasks, several frameworks that are based on OSD or in TSD have already proposed in the literature. However, these frameworks tend to have underwhelming performance and therefore further research is necessary to develop capable object detection models. In this study, we attempt to access which type of the two aforementioned frameworks for object detection has the potential to be the foundation for the development of DL models for the lung nodule detection task. Our results show that on contrary to popular belief, OSD tend to have better performance than TSD in the detection of lung nodules. This means that during the development of new frameworks, OSD should be chosen over TSD.

6 Acknowledgements

This work is financed by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within the project LUCCA, with reference 2022.03488.PTDC and by the Portuguese funding agency FCT—Fundação para a Ciência e a Tecnologia—within the PhD grant “2022.13862.BD”.

References

[1] Krishna Chaitanya Thandra, Adam Barsouk, Kalyan Saginala, John Sukumar Aluru, and Alexander Barsouk. Epidemiology of lung cancer. *Contemporary Oncology/Współczesna Onkologia*, 25(1):45–52, 2021.

[2] Ankur R Parikh. Lung cancer genomics. *Acta Medica Academica*, 48(1), 2019.

[3] Sergey P Primakov, Abdalla Ibrahim, Janita E van Timmeren, Guangyao Wu, Simon A Keek, Manon Beuque, Renée WY Granzier, Elizaveta Lavrova, Madeleine Scrivener, Sebastian Sanduleanu, et al. Automated detection and segmentation of non-small cell lung cancer computed tomography images. *Nature communications*, 13(1):3423, 2022.

[4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[5] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[6] Manuel Carranza-García, Jesús Torres-Mateo, Pedro Lara-Benítez, and Jorge García-Gutiérrez. On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing*, 13(1):89, 2020.

[7] Chi Cuong Nguyen, Giang Son Tran, Jean-Christophe Burie, Thi Phuong Nghiem, et al. Pulmonary nodule detection based on faster r-cnn with adaptive anchor box. *Ieee Access*, 9:154740–154751, 2021.

[8] Kehong Liu. Stbi-yolo: A real-time object detection method for lung nodule recognition. *IEEE Access*, 10:75385–75394, 2022.

[9] Ivan William Harsono, Suryadiputra Liawatimena, and Tjeng Wawan Cenggoro. Lung nodule detection and classification from thorax ct-scan using retinanet with transfer learning. *Journal of King Saud University-Computer and Information Sciences*, 34(3):567–577, 2022.

[10] 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.