Mediastinal Lesion Detection from CT Scans with NoduleNet

Shuang Song[1], Rui Xu[1], Yong Luo[1], Bo Du[1], Zhijian Yang[1], Jiancheng Yang[2], Kaiming Kuang[3], and Bingbing Ni[2]

[1] Wuhan University, Wuhan, China
{songshuang327, rui.xu, luoyong, dubo, zjyang.math}@whu.edu.cn
[2] Shanghai Jiao Tong University, Shanghai, China
{jekyll4168, nibingbing}@sjtu.edu.cn
[3] Dianei Technology, Shanghai, China
{kaiming.kuang}@dianei-ai.com

Abstract. Diagnosis of mediastinal lesions is critical for facilitating early screening and timely treatment. However, few prior studies investigate deep learning methods on this labor-intensive task. In this paper, we first build a mediastinal lesion dataset that contains a train-dev dataset of 880 computed tomography (CT)scans and a test dataset of 220 CT scans. Then we build a benchmark by applying NoduleNet, a well-known detection technique based on 3D convolutional neural network (CNN), to automatically detect mediastinal lesions from CT scans. Our benchmark achieves a sensitivity of 82.57% with average false positives per scan ≤ 6. To the best of our knowledge, this is the first study that applies 3D deep learning techniques to automatically detect mediastinal lesions in a large-scale cohort.

Keywords: Mediastinal lesion · computed tomography · deep learning · 3D convolutional neural network.

1 Introduction

The mediastinum contains many vital anatomic structures. It is a site of non-neoplastic and neoplastic lesions, many of which present as mediastinal masses [1]. Chest computed tomography (CT) remains the study of choice for evaluating mediastinal pathologic conditions due to its advantage in providing a clear distinction between different tissues [2]. However, detection of mediastinal lesions is a challenging task for the following reasons: 1) mediastinal lesions vary greatly between cases in their size, shape, and location (as shown in Fig. 1); 2) the contrast between the lesions in the mediastinum and surrounding background is relatively low in CT scans. Hence it is difficult to discriminate one tissue from others; and 3) a large number of CT slices must be evaluated sequentially, which is tedious and labor-intensive. Therefore, it is desirable to design automatic mediastinal lesion detection techniques.

In recent years, deep learning has become a dominant methodology in medical image analysis and has been applied in various medical imaging tasks, such

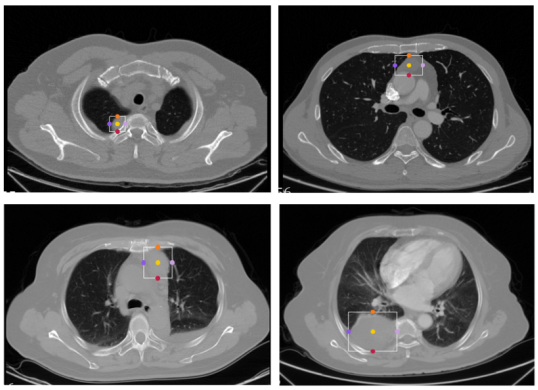

Fig. 1. The mediastinal lesions in CT scans from the view of axial dimension.

as lesion and organ classification, detection, and segmentation [3]. However, few prior studies explore automatic detection of mediastinal lesions. Deeplesion [4] established a large-scale lesion image dataset that contains a variety of lesions with annotations on key CT slices, which also includes lesions in the mediastinum. Some universal lesion detectors [4–6] were then developed based on 2D CNN to detect all types of lesions in Deeplesion. However, these studies explore detecting lesions from CT images of different body parts, rather than focusing on detecting mediastinal lesions. Besides, lesion detection from 2D images actually simplifies this problem, lacking the capability of accurate localization and count of all mediastinal lesions.

To remedy these drawbacks, we first propose a dataset that only contains mediastinal lesions. Experienced radiologists are required to draw bounding box wrapping around the lesion from the axial, coronal and sagittal direction as close as possible in each CT scan. The dataset is randomly split into a training set of 770 CT scans, a validation set of 110 CT scans, and a test set contains 220 CT scans. Then we explore a fully automatic approach based on NoduleNet [7] to detect mediastinal lesions from CT scans. NoduleNet is a 3D deep convolutional neural net (DCNN), which is originally designed for pulmonary nodule detection and nodule segmentation. Since NoduleNet has been shown to achieve excellent performance in detecting nodules in the lung, we propose to employ it for mediastinal lesion detection. The key contributions of this paper are:

- To the best of our knowledge, this is the first dataset that only focuses on mediastinal lesion detection.
- We build a benchmark for mediastinal lesion detection from CT scans based on the 3D deep learning technique. The performance is average detection sensitivity of 82.57% and average false positives per scan ≤ 6.

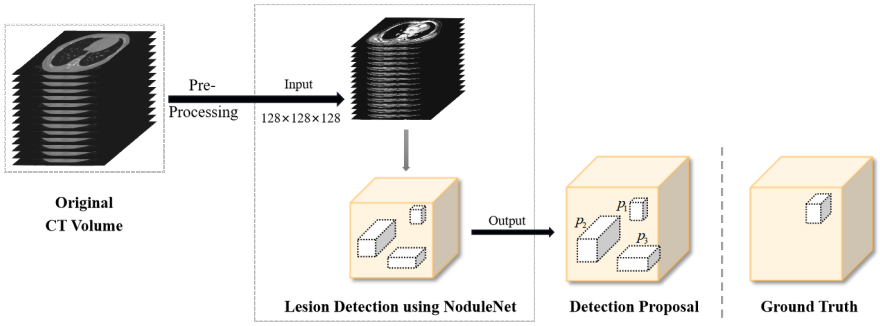

Fig. 2. The pipeline for detecting mediastinal lesions from CT scans.

2 Datasets and Methods

2.1 Data Collection and Annotation

This study is approved by the institutional review board of Shanghai Pulmonary Hospital Affiliated to Tongji University, which waives the requirement for patients written informed consent referring to the CIOMS guideline.

A custom dataset termed *MELA* (mediastinal lesion analysis) is collected from the Shanghai Pulmonary Hospital, where 1100 CT scans from patients with one or more mediastinal lesions are included. Experienced radiologists annotated each mediastinal lesion in each CT scan by drawing a bounding box wrapping around the lesion from the axial, coronal and sagittal direction as close as possible. Each mediastinal lesion corresponds to an annotation consisting of the coordinates and lengths of the ground truth bounding box in three-dimension. The CT scans are in ".nii" format, and the annotations are in ".csv" format. The *MELA* dataset is further split into a subset of 770 CT scans for training, a subset of 110 CT scans for validation, and a test set of 220 CT scans for evaluation.

2.2 Model Overview

As shown in Fig. 2, our method includes the following stages to detect mediastinal lesions from CT scans automatically: 1) **pre-processing** the input CT volumes; 2) **lesion detection** using NoduleNet.

Pre-Processing. (1) Windowing. To highlight the mediastinum structures in CT scans, we adjust the intensity of input using mediastinum window (window level: 40, window width: 350), and normalize it to [0, 255]; (2) Resampling. All of the CT scans are resampled to the spacing of $1 \times 1 \times 1$. In this way, the voxels of different sizes in CT scans are normalized to the same size; (3) Cropping. Extra backgrounds are cropped according to the values of the CT images after step (1), to be specific, the outermost coordinates with a non-zero gray value are used as the cropping boundary.

Lesion Detection. NoduleNet [7] is applied to perform the task of detection, with an input size of $128 \times 128 \times 128$. The two modules of NoduleNet, i.e.,

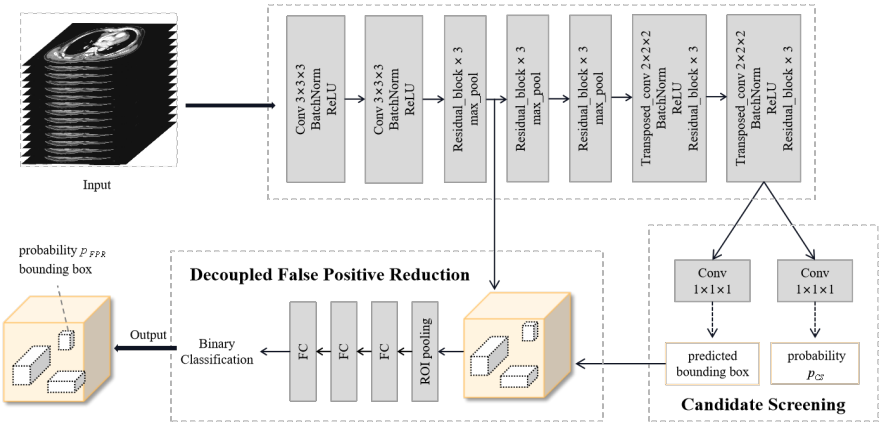

Fig. 3. NoduleNet Network Architecture.

Candidate Screening and Decoupled False Positive Reduction, are adopted for detection and segmentation is not involved. After lesion detection, the network outputs bounding boxes as detection results together with confidence.

2.3 NoduleNet Network Architecture

NoduleNet [7] is a unified framework, which integrates nodule detection, false positive reduction, and nodule segmentation. The framework is proposed to solve the problems of pulmonary nodule detection and segmentation in a multi-task manner. Since it performs well on nodule detection, we utilize NoduleNet to build our benchmark on the *MELA* dataset for mediastinal lesion detection. The architecture is illustrated in Fig. 3.

Candidate screening (CS) generates predicted probabilities and bounding box regression terms associated with anchors of different sizes, by using two parallel $1\times1\times1$ convolutional layers on the feature map. Decoupled false positive reduction (DFPR) uses a 3D region of interest (ROI) pooling layer to pool features from the early feature map, which ensures a smaller receptive field and can learn more feature representation. The same multi-task loss function of the two modules is defined as:

$$L = L(p_i, t_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*), \qquad (1)$$

where i is the index of i-th anchor in one patch. N_{cls} and N_{reg} are the numbers of anchors considered for computing classification (lesion or not lesion) loss L_{cls} and regression loss L_{reg}, respectively. λ is a parameter used to balance the two losses. p_i is the predicted probability of i-th anchor being a lesion, p_i^* is 1 if the anchor is positive and is 0 if the anchor is negative. t_i is a vector denoting

the predicted bounding box coordinates for lesion position, and t_i^* is the ground-truth bounding box. What's more, cross-entropy loss is used for L_{cls} and smooth L_1 loss for L_{reg} here.

3 Experiments

3.1 Experiment Setting

Model Training. Stochastic gradient descent (SGD) is used for training the model, with an initial learning rate of 0.01, momentum 0.9, and l_2 penalty 0.0001, for 400 epochs. The learning rate is scheduled to decrease to 0.001 after 180 epochs and to 0.0001 after another 100 epochs. Data augmentation including horizontal flip, rescaling and rotation is applied during training stage.
Detection Metric. Due to the various sizes, shapes, and positions of mediastinal lesions, the Intersection over Union (IoU) between the ground truth and prediction results may vary significantly. Similar to the P3D IoU Evaluation Metric [8], a detection proposal counts for a "hit" if its predicted 3D bounding box has an IoU > 0.3 with a certain ground truth lesion annotation.
Model Evaluation. Free-Response ROC (FROC) analysis [9] is adopted for evaluating the performance of lesion detection. FROC is a curve of sensitivity vs. false positives per scan (FP/s), by varying the acceptance probability score. Given an acceptance probability threshold, we are able to compute the true positives (TP), false positives (FP), and false negatives (FN). That is,

$$sensitivity = TP/(TP + FN), \tag{2}$$

$$FP/s = FP/Num, \tag{3}$$

where *Num* denotes the size of evaluation set.

3.2 Results

We train and tune the model on the *MELA* train-dev set, where 770 CT scans are randomly selected for training, and 110 CT scans are used for validation. The trained model is evaluated on the *MELA* test set, and the performance are reported in Table. 1. From the results, we observe that our benchmark model can achieve a sensitivity over 90% when $FP/s \geq 1$. The average sensitivity is 82.57%, and average false positives per scan is $5.96 \leq 6$.

Table 1. Mediastinal lesion detection sensitivities (unit: %) w.r.t. different false positive rate per CT scan (FP/s).

	0.125	0.25	0.5	1.0	2.0	4.0	6.0	Avg.
NoduleNet	53.47	72.27	80.54	90.04	92.04	94.57	95.02	82.57

4 Conclusion

In this paper, we build a novel dataset for mediastinal lesion analysis, and explore a 3D deep learning benchmark based on NoduleNet to automatically detect mediastinal lesions from CT scans. As the first effort focusing on automatic mediastinal lesion detection, we hope this work could facilitate the algorithm development and attract more studies for this vital problem.

References

1. S Karki and S Chalise. Analysis of mediastinal lesions: A study of 27 cases. *Journal of pathology of Nepal*, 1(2):114–117, 2011.
2. Pamela M Tecce, Elliot K Fishman, and Janet E Kuhlman. Ct evaluation of the anterior mediastinum: spectrum of disease. *Radiographics*, 14(5):973–990, 1994.
3. Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
4. Ke Yan, Xiaosong Wang, Le Lu, and Ronald M Summers. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of medical imaging*, 5(3):036501, 2018.
5. You-Bao Tang, Ke Yan, Yu-Xing Tang, Jiamin Liu, Jin Xiao, and Ronald M Summers. Uldor: a universal lesion detector for ct scans with pseudo masks and hard negative example mining. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 833–836. IEEE, 2019.
6. Qingbin Shao, Lijun Gong, Kai Ma, Hualuo Liu, and Yefeng Zheng. Attentive ct lesion detection using deep pyramid inference with multi-scale booster. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 301–309. Springer, 2019.
7. Hao Tang, Chupeng Zhang, and Xiaohui Xie. Nodulenet: Decoupled false positive reduction for pulmonary nodule detection and segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 266–274. Springer, 2019.
8. Jinzheng Cai, Adam P Harrison, Youjing Zheng, Ke Yan, Yuankai Huo, Jing Xiao, Lin Yang, and Le Lu. Lesion-harvester: Iteratively mining unlabeled lesions and hard-negative examples at scale. *IEEE Transactions on Medical Imaging*, 40(1):59–70, 2020.
9. H. L. Kundel, K. S. Berbaum, and D. D. Dorfman. Receiver operating characteristic analysis in medical imaging. *Journal of the ICRU*, 8(1):p.1–1, 2008.