Journal of Computer Technology and Software

ISSN:2998-2383

Vol. 3, No. 4, 2024

Enhanced Lung Lesion Detection with Improved Attention Mechanisms

Carter Evans

University of California, Berkeley

carter.evans9076@berkeley.edu

Abstract: In recent years, artificial intelligence (AI) technology has been extensively utilized to aid radiologists in the diagnosis and analysis of medical images. AI proves particularly effective in assisting with the localization of lesions. Nevertheless, the current mainstream target detection models face practical implementation challenges within medical systems due to their reliance on large backbone networks and high input resolutions, which result in reduced accuracy and significant computational resource demands. This paper introduces the IAB-RefineDet, a lung lesion detection network characterized by rapid detection speeds and high accuracy. By enhancing the channel and spatial attention mechanisms and integrating an improved attention module into RefineDet, the accuracy of lesion detection is markedly increased without a substantial rise in parameter numbers. We conduct comprehensive experiments on VinDr-CXR, the largest publicly available chest radiograph detection dataset, alongside comparative studies with existing mainstream target detection models. The experimental outcomes demonstrate that IAB-RefineDet achieves a mean Average Precision (mAP) of 16.24%, significantly outperforming leading deep learning models in lesion detection performance.

Keywords:Lesion Detection; Deep Learning; Attention Mechanism; Chest X-ray.

1.Introduction

With the acceleration of global industrialization, human lung health has emerged as a critical challenge. The World Health Organization reports that five of the top ten global causes of death are lung-related Consequently, early screening for lung diseases is essential to reduce mortality rates. Given the significant disparity in the ratio of radiologists to patients, radiologists must review numerous CXR (Chest X-ray) images daily within a limited time frame, leading to potential misdiagnoses influenced by the radiologist's experience and subjective factors. To alleviate radiologists' workload and enhance diagnostic accuracy, computer-aided diagnosis using artificial intelligence techniques has gained popularity. The rapid advancements in artificial intelligence have spurred a trend of leveraging deep learning to assist in medical diagnoses. Specifically, in lung disease screening, target detection models significantly enhance radiologists' efficiency by identifying potential lung lesion areas. Medical studies indicate that computer-aided diagnosis systems are highly effective in disease screening when the false-positive rate is minimized, and sensitivity exceeds 81% .

Deep learning-based target detection in medical imaging often follows general target detection models such as YOLO and Faster RCNN. One-stage target detection methods offer faster detection speeds, while two-stage methods provide higher detection accuracy. Although one-stage algorithms are faster and have smaller models, making them more suitable for hospital auxiliary detection systems, they lag behind two-stage algorithms in detection accuracy. Conversely, two-stage models are challenging to deploy in hospital systems due to their larger size and higher computational resource demands.

Zhang et al. introduced a one-stage target detection network, RefineDet, which incorporates the ARM (Anchor Refinement Module) and ODM (Object Detection Module) for initial filtering and further screening of anchor frames. The TCB (Transfer Connection Block) module fuses features between ARM and ODM, enabling the one-stage network to achieve two-stage accuracy while maintaining fast detection speeds, thus facilitating model deployment. This paper uses RefineDet as a benchmark, enhancing its architecture to address insufficient feature extraction and low lesion detection accuracy. The aim is to develop a high-accuracy, fast detection algorithm that more effectively aids in lung disease diagnosis.

We validate the performance of IAB-RefineDet through experiments on VinDr-CXR, the largest publicly available chest X-ray detection dataset. Comparative experiments with mainstream deep learning models focus on mAP and parameter count. The results, presented in Table 1, indicate that IAB-RefineDet achieves a 16.24% mAP with a moderate parameter count. This represents a 6.94% performance improvement over the benchmark network, significantly surpassing mainstream deep learning models in detection accuracy.

2. Related Work

The utilization of artificial intelligence (AI) in medical diagnostics has significantly advanced with the incorporation of deep learning and data preprocessing techniques. Xu et al. [1] demonstrate the potential of deep learning in improving medical diagnostic accuracy, emphasizing the importance of effective data preprocessing to enhance model performance. This foundational work

provides critical insights into the preprocessing strategies that can be applied to medical imaging datasets to improve diagnostic outcomes. In the domain of medical image segmentation, Zhu et al. [2] introduce the Attention-Unet, a deep learning model that leverages attention mechanisms for faster and more accurate segmentation. Their approach underscores the value of integrating attention mechanisms to refine model focus on relevant image features, which is particularly pertinent to our work on enhancing lesion detection accuracy in chest X-rays. Zhang et al. [3] present RefineDet, a one-stage target detection network that achieves high accuracy and fast detection speeds by incorporating modules such as the ARM (Anchor Refinement Module) and ODM (Object Detection Module). This network serves as a benchmark in our research, and our enhancements build upon its architecture to address specific challenges in lesion detection accuracy and feature extraction.Yao et al. [4] explore advancements in 3D semantic scene completion using monocular vision, contributing to the broader field of computer vision and providing insights into model optimization techniques that can be adapted for medical imaging applications. Their work in normalizing device coordinates space offers valuable perspectives on enhancing spatial understanding in image analysis, which parallels the spatial attention mechanisms employed in our study.Yang et al. [5] focus on the predictive accuracy of neural networks in various applications, including the prediction of laser-induced shock wave velocities. Although not directly related to medical imaging, the methodologies discussed for improving neural network predictions are relevant to optimizing deep learning models for lesion detection.Song and Liu [6] compare norm-based feature selection methods on biological omics data, providing a comparative analysis of feature selection techniques that can be beneficial for refining the input features in medical imaging models. Their findings support the selection of optimal features to enhance model performance, which is crucial for the accuracy of lesion detection networks. Liu et al. [7] analyze the impact of external factors on predictive models, specifically studying the effects of the COVID-19 epidemic on New York taxi data. Their use of machine learning algorithms to adapt to changing data conditions offers insights into model robustness and adaptability, principles that are applicable to developing resilient medical imaging models.

Collectively, these works contribute to the development of more accurate, efficient, and robust deep learning models for medical imaging applications. By integrating advanced attention mechanisms and optimizing feature extraction processes, our proposed IAB-RefineDet network aims to achieve superior lesion detection performance, addressing the critical need for reliable and efficient diagnostic tools in medical practice.

2.Method

2.1.The Overall Structure of proposed IAB-RefineDet

The network structure of IAB-RefineDet, depicted in Fig. 1, integrates several key modules to enhance detection accuracy and efficiency. The ARM (Anchor Refinement Module) conducts preliminary screening by eliminating non-target frames and adjusting the anchor positions to provide optimal initial anchor frames for subsequent target regression. The ODM (Object Detection Module) then refines the anchor frame positions and predicts their class information. Positioned between the ARM and ODM modules, the TCB (Transfer Connection Block) module enhances contextual information integration by transferring features from the ARM module's various output layers to the corresponding ODM module, thereby boosting the ODM module's detection capabilities.

In chest X-ray lesion detection, RefineDet faces the challenge of underutilized features. To address this, we designed the IAB (Improved Attention-Based) module, which combines attention mechanisms and is integrated before the TCB module. This design ensures that features extracted by the backbone network are first processed by the IAB module. By incorporating the IAB module prior to the TCB, the model's lesion detection performance is markedly improved.

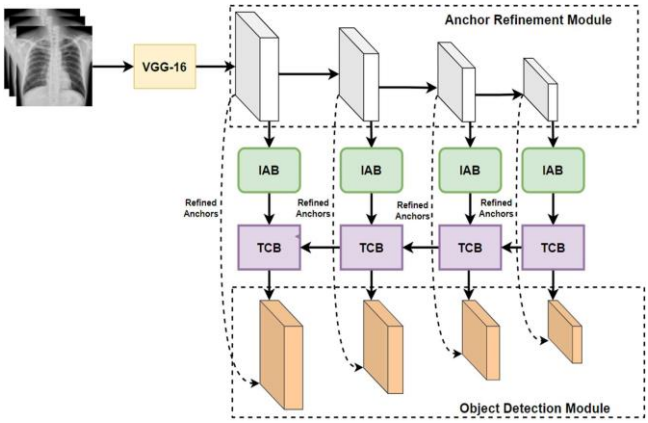

Figure 1 The overall structure of IAB-RefineDet

2.2.The Convolutional Block Attention Module

The attention mechanism in deep learning is designed to autonomously learn and selectively focus on crucial features, enabling the network to allocate more computational resources to significant feature information. The Convolutional Block Attention Module (CBAM) introduced by Woo et al., is a straightforward yet effective attention module that extracts features in both the channel and spatial dimensions.

The CBAM structure serially connects the channel attention module and the spatial attention module between its input and output. Both modules utilize global maximum pooling and global average pooling to extract rich global and local semantic information. For feature maps generated by the

convolutional network, CBAM first calculates the channel attention templates based on the channel dimensions. These templates are then used to recalibrate the channel weights of the original feature images by multiplying them with the original feature images. The resulting feature maps, which now include extracted channel attention, are subsequently fed into the spatial attention module and the channel attention module for further processing.

By incorporating CBAM into the IAB-RefineDet network, the model enhances its ability to focus on significant features, thereby improving lesion detection accuracy and efficiency in chest X-ray images.

2.3.The Improved Attention Block

Inspired by CBAM, we designed the IAB module, as illustrated in Fig. 2. The IAB comprises a channel attention mechanism and a spatial attention mechanism. The feature map undergoes initial processing by the channel attention mechanism, which adjusts the channel dimension weights of the original feature image by multiplying the resulting feature map with the original one. This adjusted feature map is then processed by the spatial attention mechanism, with the resulting feature map subsequently multiplied with the output of the channel attention mechanism. This approach allows the network to adaptively learn the importance of different pixel positions within the same channel, ultimately isolating the most salient features. These filtered salient features are then fused with the original feature map through a shortcut connection, ensuring that the refined features enhance the model's overall performance without losing crucial information from the initial feature map.

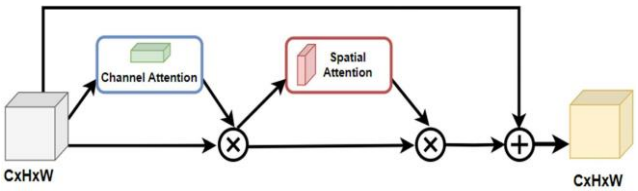

Figure 2 The improved attention block includes both spatial and channel attention components

2.4.Channel Attention Block

To adjust the weights of different channels in the feature map more effectively, our channel attention mechanism employs both global maximum pooling and global average pooling to extract semantic information. Maximum pooling focuses on the elements with maximum values in each pooled region, thereby highlighting the salient features of the target and preserving texture, structure, and contour information. Conversely, average pooling computes the average value of all elements in each pooled region, retaining more background information.

Using both pooling methods simultaneously helps eliminate redundant information while ensuring richer high-level feature extraction. This process results in two feature vectors of size HxWx1. These vectors are then processed through a 1x1 convolutional layer, which captures more complex channel attention features. Finally, the feature vectors from both paths are fused and processed by a sigmoid function to generate the final channel attention map.

This combined approach enhances the ability of the network to adjust channel weights dynamically, improving the extraction of significant features while maintaining comprehensive information from the original image.

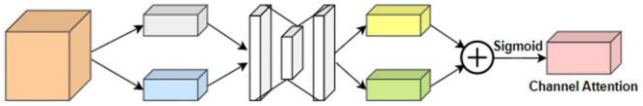

Figure 3 The structure of channel attention block

2.5.Spitial Attention Block

The spatial attention block follows the channel attention block, focusing on extracting key information from various locations within the same feature map and generating a spatial attention map based on the spatial relationships of the features. The input to the spatial attention module is the recalibrated feature map obtained from the channel attention module.

First, semantic information is extracted from the HxWxC feature map along the spatial dimension using global average pooling and global maximum pooling, resulting in an HxWx2 feature map. This pooled feature map is then processed through three convolutional layers of different sizes to extract multiscale information. The information obtained from each convolutional branch is then fused and processed by a sigmoid function to generate the final spatial attention map.

This process allows the network to adaptively focus on important spatial locations within the feature map, enhancing the model's ability to detect and highlight significant features across different scales and spatial contexts.

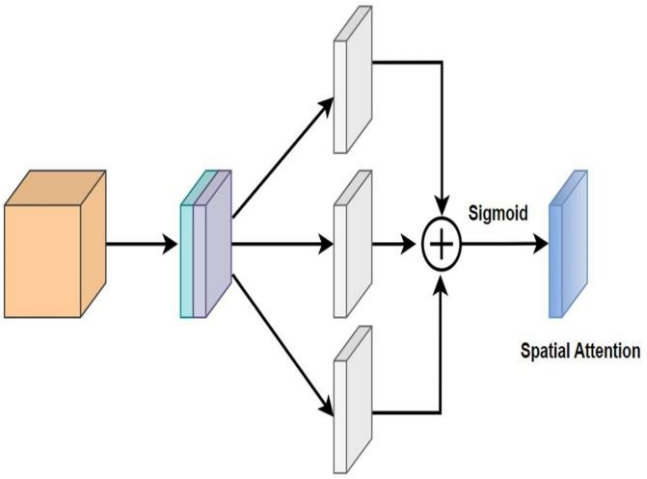

Figure 4 The structure of spatial attention block

3.Comparison with Mainstream Target Detection Algorithms

To demonstrate the superiority of our model, we conducted comparative experiments with mainstream target detection models, evaluating their performance using the metrics of mean Average Precision (mAP) and the number of parameters. The experimental results, presented in Table 1, highlight these comparisons.

In terms of detection accuracy, the benchmark model RefineDet achieves a mAP of 9.29%. Other models, such as SSD and RetinaNet , also have mAP values below 11%, with Yolov3 reaching a slightly higher detection accuracy of 10.85%. The two-stage detection network Faster R-CNN shows a modest increase in detection accuracy due to its architecture but still falls short of our model, IAB-RefineDet. Additionally, Faster R-CNN has a larger number of parameters compared to our model.

The experimental results reveal that IAB-RefineDet achieves a mAP of 16.24%, significantly outperforming the mainstream deep learning models while maintaining a moderate number of parameters. This indicates that IAB-RefineDet can sustain high detection speeds alongside superior detection accuracy, making it a highly effective solution for lung lesion detection.

Table 1 Performance comparison with other models

Methods	Backbone	mAP(%)	Params(M)
RetinaNet	ResNet-50	9.33	38.67
SSD	VGG-16	6.20	36.59
Faster R-CNN	VGG-16	11.70	42.33
	ResNet-50	12.4	45.58
Yolov3	DarkNet-53	10.85	40.72
RefineDet	VGG-16	9.29	36.75
IAB-RefineDet	VGG-16	16.24	40.22

4.Conclusion

In this paper, we introduce IAB-RefineDet, a highly efficient and accurate lung lesion detection network. By incorporating an attention processing method that includes a channel attention module and a spatial attention module, we significantly enhance the network's lesion detection performance. Our IAB-RefineDet achieves a mean Average Precision (mAP) that is 6.96% higher than the benchmark network, and its mAP scores surpass those of the current mainstream deep learning models. This improvement addresses the clinical need for more accurate lung disease diagnoses and effectively mitigates the existing challenges of low detection accuracy and high resource consumption in lung lesion detection.

References

[1] R. Xu, Y. Zi, L. Dai, H. Yu, and M. Zhu, "Advancing Medical Diagnostics with Deep Learning and Data Preprocessing," International Journal of Innovative Research in Computer Science & Technology, vol. 12, no. 3, pp. 143-147, 2024.

[2] Z. Zhu, Y. Yan, R. Xu, Y. Zi, and J. Wang, "Attention-Unet: A Deep Learning Approach for Fast and Accurate Segmentation in Medical Imaging," Journal of Computer Science and Software Applications, vol. 2, no. 4, pp. 24-31, 2022.

[3] H. Yang, Y. Zi, H. Qin, H. Zheng, and Y. Hu, "Advancing Emotional Analysis with Large Language Models," Journal of Computer Science and Software Applications, vol. 4, no. 3, pp. 8-15, 2024.

[4] J. Yao et al., "Ndc-scene: Boost Monocular 3D Semantic Scene Completion in Normalized Device Coordinates Space," in 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9421-9431, 2023.

[5] H. Yang, "Improving Prediction Accuracy of Laser-Induced Shock Wave Velocity Prediction Using Neural Networks," Scientific Reports, vol. 14, no. 1, p. 13576, 2024.

[6] J. Song and Z. Liu, "Comparison of Norm-Based Feature Selection Methods on Biological Omics Data," in Proceedings of the 5th International Conference on Advances in Image Processing, pp. 109-112, 2021.

[7] Z. Liu, X. Xia, H. Zhang, and Z. Xie, "Analyze the impact of the epidemic on New York taxis by

machine learning algorithms and recommendations for optimal prediction algorithms," in Proceedings of the 2021 3rd International Conference on Robotics Systems and Automation Engineering, pp. 46-52, 2021.