Signal, Image and Video Processing (2023) 17:2473–2480
https://doi.org/10.1007/s11760-022-02464-0

ORIGINAL PAPER

3D multi-scale vision transformer for lung nodule detection in chest CT images

Hassan Mkindu[1] · Longwen Wu[1] · Yaqin Zhao[1]

Received: 8 June 2022 / Revised: 2 September 2022 / Accepted: 21 December 2022 / Published online: 17 January 2023

Abstract
Lung cancer becomes the most prominent cause of cancer-related death in society. Normally, radiologists use computed tomography (CT) to diagnose lung nodules in lung cancer patients. A single CT scan for a patient produces hundreds of images that are manually analyzed by radiologists which is a big burden and sometimes leads to inaccuracy. Recently, many computer-aided diagnosis (CAD) systems integrated with deep learning architectures have been proposed to assist radiologists. This study proposes the CAD scheme based on a 3D multi-scale vision transformer (3D-MSViT) to enhance multi-scale feature extraction and improves lung nodule prediction efficiency from 3D CT images. The 3D-MSViT architecture adopted a local–global transformer block structure whereby the local transformer stage individually processes each scale patch and forwards it to the global transformer level for merging multi-scale features. The transformer blocks fully relied on the attention mechanism without the inclusion of the convolutional neural network to reduce the network parameters. The proposed CAD scheme was validated on 888 CT images of the Lung Nodule Analysis 2016 (LUNA16) public dataset. Free-response receiver operating characteristics analysis was adopted to evaluate the proposed method. The 3D-MSViT algorithm obtained the highest sensitivity of 97.81% and competition performance metrics of 0.911. Therefore, the 3D-MSViT scheme obtained comparable results with low network complexity related to the counterpart deep learning approaches in prior studies.

Keywords Computer-aided diagnosis · Computed tomography · Vision transformer · Lung nodule · 3D-MSViT

1 Introduction

According to the American cancer society, lung cancer is more prevalent among cancer-associated deaths. In 2022, cancer statistics show that the overall cancer cases and deaths were approximately 1.9 million cases and 609,360 deaths, respectively, whereby lung cancer contributed to closely 350 deaths per day [1]. Usually, late detection of cancer nodules leads to complex treatment of the patients and results in death. There is only a 10% to 16% chance of living five years more for late-diagnosed patients. Nevertheless, early diagnosed patients can increase their survival rate to 70% [2]. In clinical diagnosis, CT images are most preferred for lung cancer patients diagnosis owing to their fast acquisition and high resolution [3, 4]. A single CT scan produces multiple images and is manually diagnosed by radiologists which is a time-consuming task that results in inaccuracy [5]. Hereafter, computer-aided detection (CAD) systems have been introduced to assist radiologists and improve diagnosis efficiency [6].

In the first stage, CAD systems screen all positive nodule candidates in a given number of CT images with high sensitivity which is resulting in the inclusion of many false-positive nodules [7, 8]. To reduce false-positive nodules, a second stage for classifying the true positive nodules from non-nodules is conducted [9]. In the prior studies, conventional CAD systems based on hand-engineered features have been suggested for nodule candidate detection [10, 11]. The conventional approaches resulted in low performance due to their detection inefficiency from nodules with various sizes, textures, and shapes. Lu et al. [12] and Murphy et al. [13] applied machine learning algorithms that relied on manual nodule feature extraction for candidate nodule screening. The

✉ Longwen Wu
wulongwen@hit.edu.cn

Hassan Mkindu
hmkindu@hit.edu.cn

Yaqin Zhao
yaqinzhao@hit.edu.cn

1 School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China