Radiological Physics and Technology (2024) 17:124–134
https://doi.org/10.1007/s12194-023-00756-9

RESEARCH ARTICLE

Hierarchical approach for pulmonary-nodule identification from CT images using YOLO model and a 3D neural network classifier

Yashar Ahmadyar[1] · Alireza Kamali-Asl[1] · Hossein Arabi[2] · Rezvan Samimi[1] · Habib Zaidi[2,3,4,5]

Received: 5 May 2023 / Revised: 16 October 2023 / Accepted: 17 October 2023 / Published online: 18 November 2023

Abstract

This study aimed to assist doctors in detecting early-stage lung cancer. To achieve this, a hierarchical system that can detect nodules in the lungs using computed tomography (CT) images was developed. In the initial phase, a preexisting model (YOLOv5s) was used to detect lung nodules. A 0.3 confidence threshold was established for identifying nodules in this phase to enhance the model's sensitivity. The primary objective of the hierarchical model was to locate and categorize all lung nodules while minimizing the false-negative rate. Following the analysis of the results from the first phase, a novel 3D convolutional neural network (CNN) classifier was developed to examine and categorize the potential nodules detected by the YOLOv5s model. The objective was to create a detection framework characterized by an extremely low false positive rate and high accuracy. The Lung Nodule Analysis 2016 (LUNA 16) dataset was used to evaluate the effectiveness of this framework. This dataset comprises 888 CT scans that include the positions of 1186 nodules and 400,000 non-nodular regions in the lungs. The YOLOv5s technique yielded numerous incorrect detections owing to its low confidence level. Nevertheless, the addition of a 3D classification system significantly enhanced the precision of nodule identification. By integrating the outcomes of the YOLOv5s approach using a 30% confidence limit and the 3D CNN classification model, the overall system achieved 98.4% nodule detection accuracy and an area under the curve of 98.9%. Despite producing some false negatives and false positives, the suggested method for identifying lung nodules from CT scans is promising as a valuable aid in decision-making for nodule detection.

Keywords Lung cancer · Pulmonary nodules · Object detection · Classification · CT scan

1 Introduction

Lung cancer significantly impacts global mortality rates. The Lung Cancer Research Foundation predicts that approximately 236,740 new lung cancer cases will occur in the United States in 2022, leading to 130,180 lung cancer-related deaths [1]. The early detection of lung cancer heavily relies on detecting pulmonary nodules. These nodules are abnormal growths within the lung, and while nodules smaller than 5 mm are frequently noncancerous, they can indicate the early stages of cancer [2]. Therefore, identifying these nodules is essential for the early diagnosis of lung cancer.

Lung cancer screening can be conducted using various methods, such as bronchoscopy, a procedure that enables physicians to explore the inside of the airways and obtain cell samples. Biopsy samples are analyzed in a laboratory to detect abnormal cells. Another method is computed tomography (CT) scan-guided biopsy, wherein, for nodules on the outer portion of the lung, CT images are used to guide a thin needle through the skin and into the lungs. This procedure aims to obtain tissue samples from a nodule and examine them for abnormalities. Positron emission tomography (PET) is also used to detect cancerous cells in organs [3].

✉ Alireza Kamali-Asl
a_kamali@sbu.ac.ir

1 Department of Medical Radiation Engineering, Shahid Beheshti University, Tehran, Iran

2 Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, CH-1211 Geneva, Switzerland

3 Geneva University Neurocenter, Geneva University, 1205 Geneva, Switzerland

4 Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

5 Department of Nuclear Medicine, University of Southern Denmark, 500 Odense, Denmark

The United States Preventive Services Task Force suggests that individuals with a heightened likelihood of developing lung cancer should undergo annual low-dose CT scans. Low-dose chest computed tomography (LDCT) can be used to identify lung cancer in its early stages, potentially enhancing survival rates. Consequently, the widespread use of LDCT has prompted a renewed emphasis on lung cancer screening [4].

The automated identification of lung nodules frequently relies on artificial intelligence techniques such as deep convolutional neural networks (DCNNs) [5]. DCNNs offer great potential in nodule detection for lung cancer, providing faster, more cost-effective, and more accurate results. This method streamlines the analysis of CT scan images, reducing detection time and improving diagnostic precision. Recently, three-dimensional (3D) convolutional neural networks (CNNs) have emerged as popular methods for detecting nodules in CT scans. These networks utilize a segmentation framework centered on UNet networks to locate and outline nodules in the lung [6–8]. Subsequently, a classifier is applied to increase the accuracy of discerning false-positive and false-negative cases [9–12]. A more rapid and accurate method, Mask-R-CNN, has been developed, which includes a two-stage object detector that combines a region proposal network (RPN) with a region-based CNN (R-CNN) and a semantic segmentation model (MASK) [13–15]. The first stage of this method involves using the selective search technique to create a bounding box around the target object, followed by implementing a CNN layer to classify the detected objects.

Nguyen et al. [16] employed an adaptive anchor box fast R-CNN model to detect lung nodules in CT images. The fast R-CNN model was trained on nodules of various dimensions in the training dataset and utilized adaptive anchor boxes of different sizes. In contrast to the fixed anchor box typically used in R-CNNs, this method adjusts the anchor box size to enhance the detection accuracy for nodules of diverse dimensions. To minimize false positives in the fast R-CNN output, the authors suggested a post-processing residual CNN architecture (ResNet) [17] for the detected nodules.

In another study, a hierarchical method consisting of an R-CNN was proposed for nodule detection, and a 3D ResNet was employed to reduce false-positive outputs [18]. Because of the slow processing of R-CNN models, these approaches are inappropriate for real-time applications [19]. Furthermore, Agnes et al. [20] employed a two-stage model consisting of a UNet-based network (Atrous Unet +) for node detection and a pyramid-dilated convolutional long short-term memory (LSTM) network to reduce false positives.

The YOLO [21] method, also known as "you only look once," is a significant object recognition algorithm. In a single forward pass, it can identify objects and classify them based on their labels in real time. YOLO has been used in medical domains, such as detecting and classifying breast masses in mammography images, which was one of its earliest applications [22, 23]. It has also been employed in skin cancer detection and melanoma identification [24]. In this context, George et al. [25] introduced an object detection approach for identifying lung nodules in CT scans by integrating the DetectNet and GoogleNet architectures. DetectNet is based on the YOLO architecture, which explores an entire image to detect suspicious lesions and classify them into nodule or non-nodule cases. In another approach presented by Huang et al. [26], lung-nodule detection was performed using a 3D OSAF-YOLOv3 model. This model is an integration of 3D YOLOv3 and a one-shot aggregation module (OSA), receptive field block (RFB), and feature fusion scheme (FFS). Despite the promising performance of the YOLO-based nodule detection model, it may not always be accurate, and a significant number of nodules may remain undetected, or non-nodule structures may be mistakenly identified as nodules. This limitation can be addressed by introducing a compartment to the model output to decrease false-positive outputs.

YOLO offers several advantages over methods like Faster R-CNN and UNet in terms of speed and efficiency [27–29]. YOLO achieves faster detection by performing object recognition in a single pass, eliminating the need for a time-consuming region proposal step. This allows YOLO to process images more quickly than the multi-stage approach of Faster R-CNN and the pixel-level segmentation of UNet. However, YOLO's speed comes with a trade-off between model lightweightness and accuracy. While YOLO may sacrifice some precision compared to more complex models, its real-time capabilities make it well-suited for applications where speed is crucial. To improve the accuracy of YOLO, integrating a 3D classifier, such as a 3D CNN, brings significant advantages. The 3D classifier modifies YOLO by enhancing the classification of detected objects, reducing false positives, and improving the overall precision of lung nodule detection. This combination of YOLO and a 3D classifier offers a promising approach for achieving both speed and accuracy in the automated identification of lung nodules in CT scans.

We aimed to utilize lightweight models to achieve highly accurate nodule detection in LDCT images with a low false-positive rate. Hence, a hierarchical approach for detecting and classifying lung nodules was developed. Initially, an object detection mechanism was employed to identify potentially concerning nodules. Subsequently, these nodules were analyzed using a 3D convolutional classifier to determine their status. The study used a modified pre-trained YOLOv5s model, known for its simplicity and efficiency, adapted for nodule identification in CT scans to detect all suspicious nodules. The aim of this step was to detect suspicious nodules. To minimize the false-positive rates in the nodules detected by the YOLOv5s model, we developed

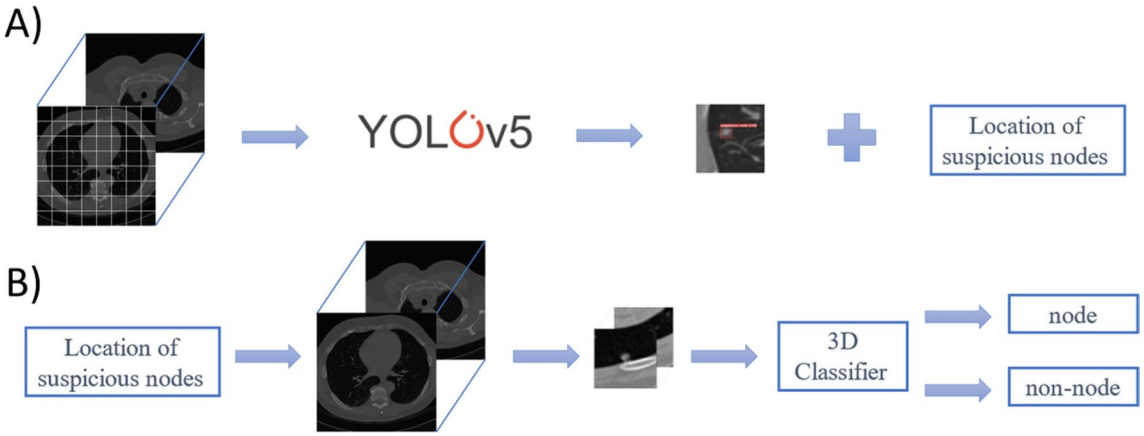

Fig. 1 Proposed framework (HND) for detecting cancerous nodules in CT images. The entire 3D CT image is fed into the YOLOv5s model to determine the location of the entire suspicious nodules (**A**). Given the location of the suspicious nodules, 3D patches of the image are fed into the proposed 3D classifier to classify them as nodules or non-nodules (**B**)

a streamlined 3D CNN classifier (without using residual blocks) and incorporated it into the framework to classify suspicious nodules accurately.

2 Material and methods

2.1 Overview

The proposed model, referred to as hierarchical nodule detection (HND), involves two steps. In the first step, the entire CT image is analyzed using the YOLOv5s algorithm, which is commonly used for object detection. The YOLOv5s algorithm determines the location and probability (confidence score) of the nodules. Given the locations of all the suspicious nodules detected by the YOLOv5s network (using a low-confidence score), a 3D bounding box (containing the nodule and background tissue) can be defined around each nodule to be fed to the next module. In the next phase, a 3D CNN classifier processes these 3D patches to determine whether the candidates are actual nodules. In summary, the first stage with the YOLOv5s algorithm detects all the nodules as a course classification, and the second module performs fine classification by focusing on a single nodule at a time (Fig. 1). The 3D CNN classifier aims to minimize the false-positive rates. The source code for YOLOv5 models is available at https://github.com/ultralytics/yolov5.

2.2 Dataset and preprocessing

2.2.1 Internal dataset

We employed 888 CT images and labeled nodules from the LUNA 16 dataset to assess the efficacy of the proposed deep learning-based framework. The dataset represents a subset of the LIDC-IDRI dataset [30] (Lung Image Database Consortium and Image Database Resource Initiative), in which each subject includes a low-dose CT image of 512×512 pixels, where the number of slices can vary between 100 and 500. A team of four expert radiologists examined the CT images and classified the nodules and non-nodules. At least three of the four radiologists were required to have labeled positives to be considered a nodule (with a diameter of at least 3 mm). In total, 1186 nodules were contained in this dataset. Furthermore, the dataset encompasses a total of 400,000 areas labeled as non-nodular. These non-nodules refer to specific regions within the CT scans that, upon expert evaluation, were not demarcated as nodules by experienced radiologists. These particular regions encapsulate typical anatomical features, benign formations, or zones devoid of clinical pertinence in the context of nodule detection. The data are available at https://luna16.grand-challenge.org/Data.

After thoroughly reviewing all the annotations for the nodules, we decided to exclude a total of 61 annotations. This exclusion was based on their incorrectness, impreciseness, or low quality compared to other images. As a result, we selected 1125 nodules belonging to 597 different patients. All the CT images were converted to Hounsfield Units (HU) and resampled into an isotropic voxel size of 1 mm. The CT images were normalized to adjust their intensity range to between 0 and 1 using a global normalization factor (maximum intensity in the entire dataset).

71% of the data was randomly chosen for the training/validation process (20% was designated for validation). The remaining 29% of the data was kept aside as unseen data to assess the model's ultimate performance, thereby preventing any bias or data leakage. Due to the random selection of training, testing, and validation data, the distribution

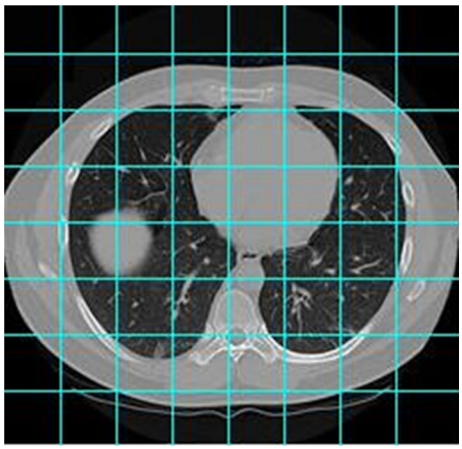

Fig. 2 Transaxial slice of a CT image, divided into 64×64 sub-images for the training of the YOLO network (The sub-images containing lung tissue were employed to train the YOLO model)

characteristics of nodules, such as their size and shape, are likewise random within the respective sets.

2.2.2 External dataset

Additional evaluations were conducted on a different independent dataset to evaluate the generalizability of our model. The dataset was collected at Khatam's PET/CT Center. The subjects comprised 47 patients (37 male, 10 female; mean age 65.04 years, range 41 to 85) with 60 lung nodules. All patients underwent [^{18}F]-FDG PET/CT scanning on a Biograph mCT equipped with 128 CT slice capability (Siemens Healthcare). We acquired a low-dose CT scan using the Siemens CARE Dose package, with automatic exposure control (AEC), according to the lowest possible patient dose, maintained at 120 kV for all exams. An effective tube current of 80 mA, pitch of 0.8, and reconstructed slice thickness of 2 mm were used. Even though lung nodules were confirmed with PET data, we only evaluated the low-dose CT images.

2.3 Network training

2.3.1 YOLOv5s network

The YOLOv5s network was retrained to identify nodules within an entire input CT image. YOLO v5s is characterized by a relatively smaller model size and fewer parameters compared to other YOLO v5 variants. The smaller model size of YOLOv5s allows for faster inference. The reduced number of parameters also contributes to its efficiency and makes it more suitable for deployment on resource-constrained devices or scenarios where real-time performance is crucial. YOLOv5s training was performed in a 2D mode to explore all the slices. Regarding the small size of the nodules compared with the 2D slices of the CT images, the CT slices were divided into 64×64 sub-images (Fig. 2). Sub-images containing nodule tissues were used to train the YOLOv5s model.

The COCO (Common Objects in Context) dataset, designed for activities such as object identification, partitioning, and description [31], was employed to pre-train the YOLO model, which was then fine-tuned for nodule identification. YOLOv5 leverages a CNN framework [32] composed of an entry phase, backbone, neck, and head (as shown in Fig. 3). Initial preprocessing occurs in the entry phase, whereas the backbone, composed of cross-stage partial systems (CSPs) [33] and spatial pyramid pooling (SPP), extracts characteristics from the input information. The

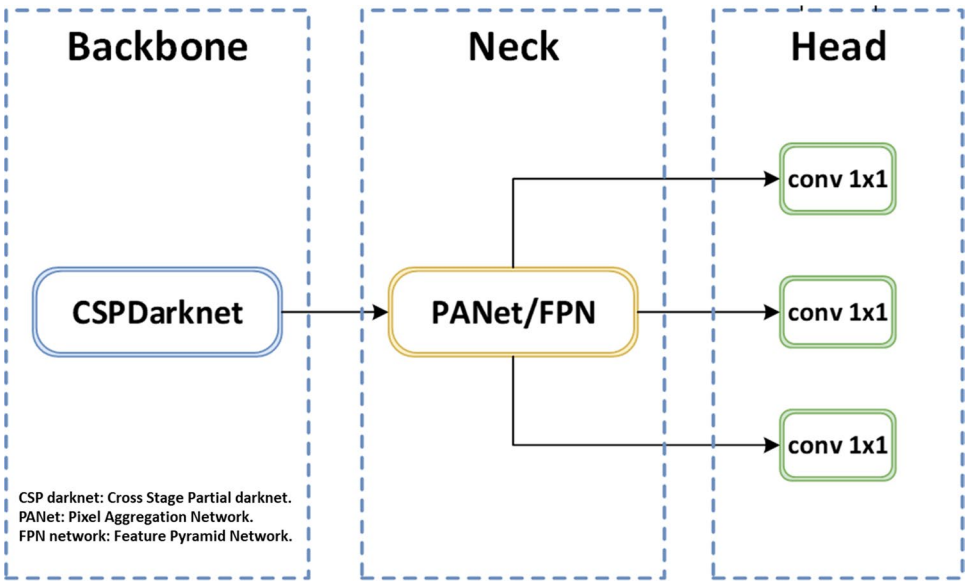

Fig. 3 Architecture of the YOLO v5 network. (CSP darknet: Cross Stage Partial darknet, PANet: Pixel Aggregation Network, FPN network: Feature Pyramid Network)

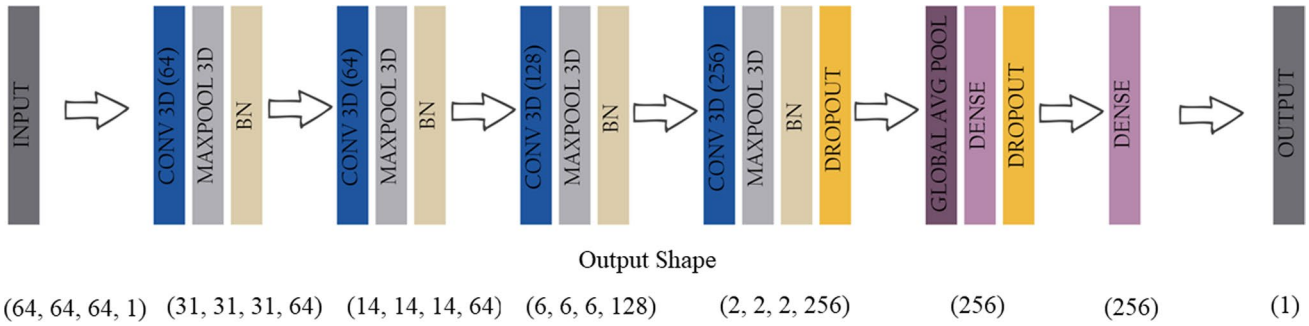

Fig. 4 Architecture of the 3D classifier for nodule classification

neck comprises a feature pyramid network (FPN) and pixel aggregation network (PAN) that transmit extensive semantic data from the upper to the lower feature map. The PAN [34] transfers feature maps from lower- to higher-level localization. When these two structures are merged, they provide the benefits of multi-scale feature representation and contextual information integration. Finally, the resulting layers in the head identify objects of varying dimensions based on feature maps. The YOLO algorithm was retrained to recognize all the nodules in the CT scans, irrespective of whether they were nodules or non-nodules.

To retrain the YOLOv5s model, we employed 804 nodules from 397 individuals across 300 iterations using stochastic gradient descent (SGD), an initial learning rate of 0.01, and a batch size of 16. The YOLOv5s model generated a bounding rectangle encircling the identified object, confidence score, and center of the nodule. The confidence score was computed as $C = Probability(object) \times Intersection\ over\ Union\ (IoU)$, where IoU denotes the intersection over the union between the predicted rectangle and actual value, signifying the likelihood of object recognition. The confidence threshold of the YOLOv5s model was reduced from 0.5 to 0.3 to ensure the detection of all nodules in the input CT scan.

2.3.2 Proposed 3D CNN network

To minimize the number of false positives, we introduced a new deep-learning structure consisting of four units. Each unit incorporated a 3D CNN, 3D max pooling, and batch normalization. Every unit featured a convolutional layer, with the filter count increasing from 64 in the initial layer to 256 in the final layer. As the filter quantity increased, more low-level characteristics could be extracted. Additionally, 3D max pooling and batch normalization layers were implemented following each convolutional layer. Using max pooling layers facilitated image analysis at four distinct complexity levels. A pair of dropout layers with a rate of 0.3 was included prior to the output to avoid overfitting. The ReLU function served as the activation function for the internal layers, and a sigmoid function was employed in conjunction with a dense layer for binary classification (Fig. 4).

A $64 \times 64 \times 64$ voxel patch was established around each nodule for 3D CNN training to account for the small dimensions of the lung nodules. The 3D CNN classified these patches as nodules or non-nodules. The 3D CNN model was trained with a learning rate of 0.0001 and a decay factor of 0.96 using the Adam optimizer and binary cross-entropy as the loss function. The training process consisted of 100 epochs with a batch size of 64, and 804 nodules were used for training and validation following the same approach as for YOLOv5s. The outputs of the YOLOv5s model were used to evaluate the network.

In this study, we developed the 3D CNN model using the Python programming language version 3.7, specifically leveraging the Keras and Tensorflow libraries. Additionally, the YOLO model has been developed using the PyTorch framework. Our experiments were conducted on a system equipped with an NVIDIA Tesla K80 GPU with a maximum capacity of 149W and memory of 11441MiB. The GPU was running NVIDIA-SMI version 495.44 with CUDA version 11.2.

3 Evaluation strategy

To evaluate the performance of the model, we conducted experiments with different detection threshold values. Our findings indicate that setting the threshold at 0.3 yielded the most favorable results in terms of identifying nodules within the YOLO model. To ensure the detection of all suspicious nodules, we examined a confidence level of 0.3. After the YOLOv5s model identified suspicious nodules in full-sized CT images, consisting of 64 patches with dimensions of 64×64, we extracted image 3D patches (with size of $64 \times 64 \times 64$) around the suspicious nodules. These 3D patches were then utilized as input for the 3D classifier. This

process was performed on 200 CT images (200 patients) containing 321 nodules as the test (unseen) dataset for the model evaluation. The results of the YOLOv5s model with the default threshold (0.5) and the adjusted threshold of 0.3 will be reported in the results section.

Ultimately, we assessed the final model using the external dataset under identical conditions and evaluation metrics as those employed for the LUNA 16 dataset.

To evaluate the nodule classification performance of the model, we used the accuracy, precision, recall, and F1 scores as follows:

$$Accuracy\ Score = (TP + TN) / (TP + FN + TN + FP) \quad (1)$$

$$Precision\ Score = TP / (FP + TP) \quad (2)$$

$$\text{Recall (sensitivity) Score} = \text{TP} / (\text{FN} + \text{TP}) \quad (3)$$

$$\text{F1 score} = 2 \times \text{Precision Score} \times \text{Recall Score} / (\text{Precision Score} + \text{Recall Score}) \quad (4)$$

$$(True\ Positive = TP,\ False\ Positive = FP,\ True\ Negative = TN,\ False\ Negative = FN) \quad (5)$$

In addition, receiver operating characteristics (ROC) were plotted using the true-positive rate (TPR) versus the false-positive rate (FPR) for varying thresholds, and the area under the curve (AUC) was determined.

$$TPR = TP / (TP + FN) \quad (6)$$

$$FPR = FP / (FP + TN) \quad (7)$$

4 Results

The YOLOv5s model was evaluated using 200 CT images containing 321 nodules. The YOLOv5s model identified 187 of 321 real nodules as suspicious using a confidence score of 0.5 (107 were false positives). Employing a confidence score of 0.3, the model detected 459 potential objects among 321 actual nodules, with 138 false positives, leading to an FPR of 28%. This comparatively high rate may be attributed to the lower confidence score used by the YOLOv5s model, primarily intended to identify larger objects in contrast to minuscule nodules. All nodules (321 real nodules) were detected in the CT images of 200 patients using a confidence level of 0.3. Table 1 presents the results of the assessment. Figure 5 depicts a representative true positive and false positive identified using the YOLOv5s model. In Fig. 6, cases initially undetected at a confidence level of 0.5 but later detected at a lower confidence level of 0.3 are shown. Furthermore, Fig. 7 shows the false negative cases classified by the HND model. To address the high FPR, we inputted the output of the YOLOv5s model into the 3D CNN classifier, improving the nodule detection precision from 69 to 100%. This improvement was owing to the ability of the 3D CNN classifier to differentiate between nodules and non-nodules. Nevertheless, we also encountered 7 instances where false negatives occurred. The ROC plot for the model is presented in Fig. 8. In addition, Fig. 9 and Table 2 present the outcomes of the 3D CNN model on YOLOv5s outputs with a 0.3 confidence level (HND model).

Table 1 Outcome of the YOLO model before being processed by the 3D classifier

YOLO confidence Level	Number of real nodules	True Positive	False Positive	Precision (%)	Detected nodule per real nodule (%)
0.5	321	187	107	58	58
0.3	321	321	138	69	100

To evaluate the model's generalizability, we conducted tests on 47 patients from the external dataset, encompassing a total of 60 nodules. The detection part of the model detected 98 suspicious nodules, out of which 58 were true positives, while it missed 2 nodules. The classification model correctly identified 35 false positives but incorrectly classified 5 nodules as false, which means we ultimately reached to 53 true positives and 35 true negatives. Figure 9 provides further clarification about the model's performance on the external datasets' samples. Overall, when applied to the external dataset, the model achieved an accuracy rate of 88.0% and a sensitivity of 88.3%.

Specifically, for YOLOv5s, the breakdown of the average detection processing time per image is as follows: 0.5ms for preprocessing, 7.0ms for inference, and 1.2ms for non-maximum suppression (NMS). This results in a cumulative average of 8.7ms per image. Furthermore, the decision-making duration for each 3D patch, driven by the 3D classifier, is approximately 0.32ms.

5 Discussion

This study aimed to detect entire nodules with a low false-negative rate using lightweight models. False negatives, where nodules are present but undetected, can occur due to factors such as small nodule size, subtle appearances, and overlapping structures. False positives, on the other hand, arise when non-nodule structures are mistakenly identified

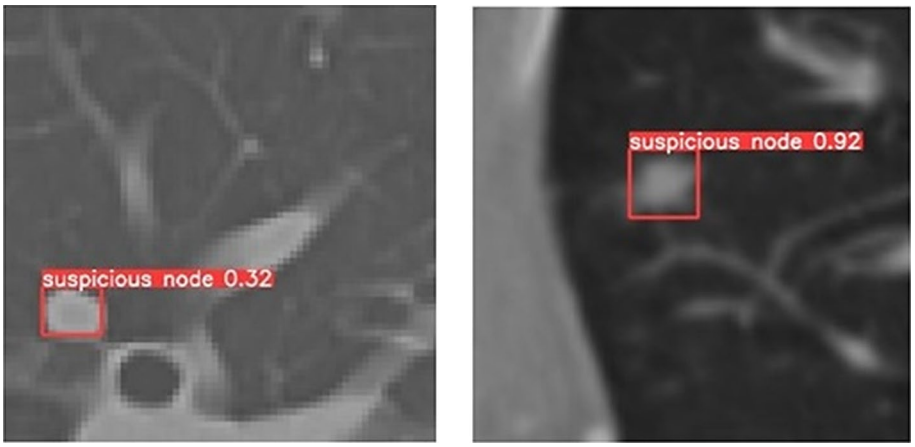

Fig. 5 Representative false-positive (left) and true-positive (right) nodules detected by the YOLO v5s model with a confidence level of 0.3

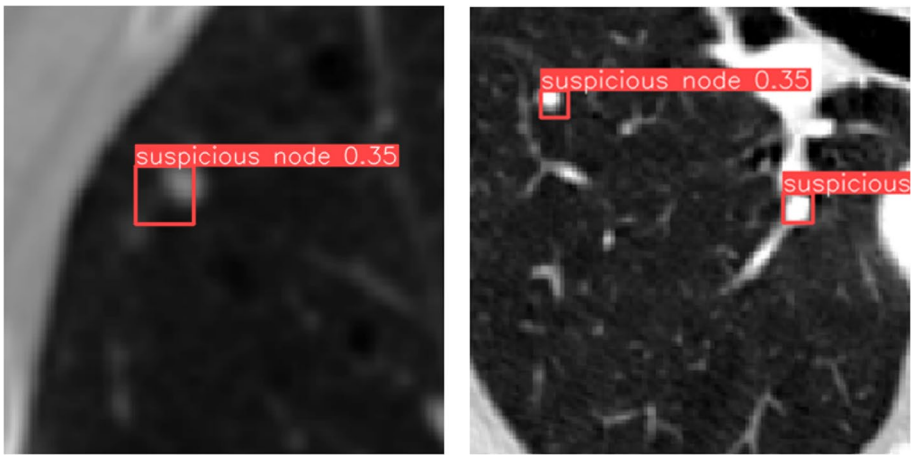

Fig. 6 Representative cases that were not detected with a confidence score of 0.5 but were detected with a confidence score of 0.3 by the YOLO v5s model. (The right detected suspicious nodule in the right image is a false positive)

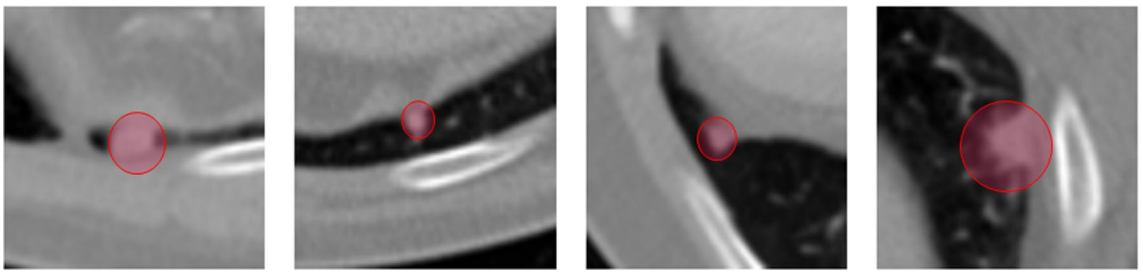

Fig. 7 Representative images of the false-negative cases using the HND (YOLO 0.3 + 3D classifier) model

as nodules. These can result from imaging artifacts, anatomical structures resembling nodules, and model limitations. Therefore, the YOLOv5s model was retrained to achieve a very high sensitivity for nodule detection (using a low confidence level) with high FPRs. Understanding that missing a nodule and foregoing subsequent screenings is riskier than mischaracterizing healthy tissue as a nodule, we moderated our detection model's specificity. This was designed to optimize the detection of suspicious lesions, albeit with a concurrent rise in false positives. The incorporation of the 3D CNN model was aimed at reducing the occurrence of false positives by utilizing the outcomes of the YOLOv5s model. The presence of the 3D CNN resulted in a significant improvement in object detection accuracy, demonstrating its potential as a promising tool for lung nodule classification. In the initial stage of nodule detection throughout the entire image, the model successfully identified nodules but also labeled numerous non-nodules as nodules. However, during the subsequent stage of classifying the outputs from the first stage, the

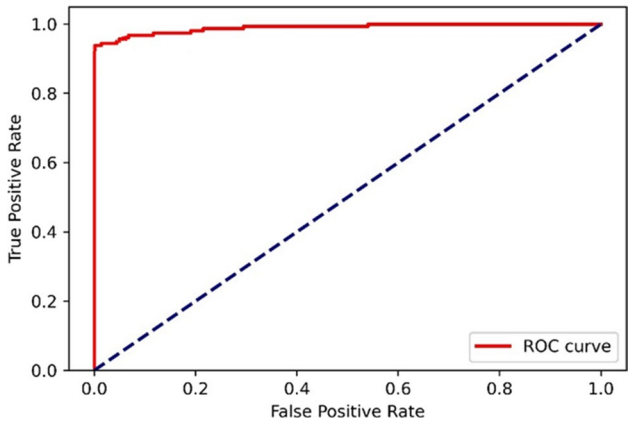

Fig. 8 ROC plot for the HND model

model classified 314 objects as nodules and 145 as non-nodules. This indicates that in the second stage, the model missed 7 nodules, initially identifying them as nodules but later classified them as non-nodules by the classifier. In essence, there is a trade-off between the model missing 7 nodules or incorrectly classifying 138 non-nodules as nodules. These false negative cases shared common characteristics, specifically their location along the lung walls and their small size. The HND model (YOLO 0.3 + 3D classifier) could accurately detect nodules in the lung with an accuracy and precision of 98.4% and 100%, respectively.

To prevent missing any nodules, we deliberately lowered the sensitivity of the model. This allowed us to detect a broader range of potential nodules, including subtle or smaller ones that could have been overlooked. Consequently, the increased sensitivity also led to a higher false positive rate, as the model became more prone to detecting false signals or noise as nodules. To address this trade-off, we employed the 3D classifier as a post-processor to refine the nodule detection and eliminate incorrect responses. Striking a balance between sensitivity and false positive rate is essential, considering the specific requirements of the application and the potential consequences of false positive detections. When a confidence level of 0.5 was used for the YOLOv5s model, 37% of the nodules were missed in the first stage. Thus, the confidence level was reduced to 0.3 to enhance the sensitivity of the model in detecting all suspicious nodules (although the model resulted in high FPRs). By reducing the confidence level of the YOLOv5s model to 0.3, all the nodules were detected (100%).

One of the advantages of adjusting model parameters, such as the confidence level, is that it allows for fine-tuning and optimizing the model's behavior without significantly altering its architecture. This flexibility can save computational resources and training time, as adding layers or external parameters may require retraining the entire model from scratch. By simply modifying the confidence threshold, we can achieve substantial improvements in detection performance with minimal overhead.

When using an external dataset to assess the generalizability of the model, our accuracy decreased from 98.4% to 88.0%, and our sensitivity dropped from 97.8% to 88.3%. Nevertheless, it is important to note that the low-dose CT of this dataset was utilized solely for PET attenuation correction and localization, making them unsuitable for clinical diagnosis. Despite these challenges, our model still demonstrated a promising performance on this type of data, which

Table 2 Results of the 3D classifier on the outputs of YOLO with a confidence level of 0.3

Accuracy (%)	Recall (%)	Precision (%)	AUC (%)
98.4	97.8	100	98.9

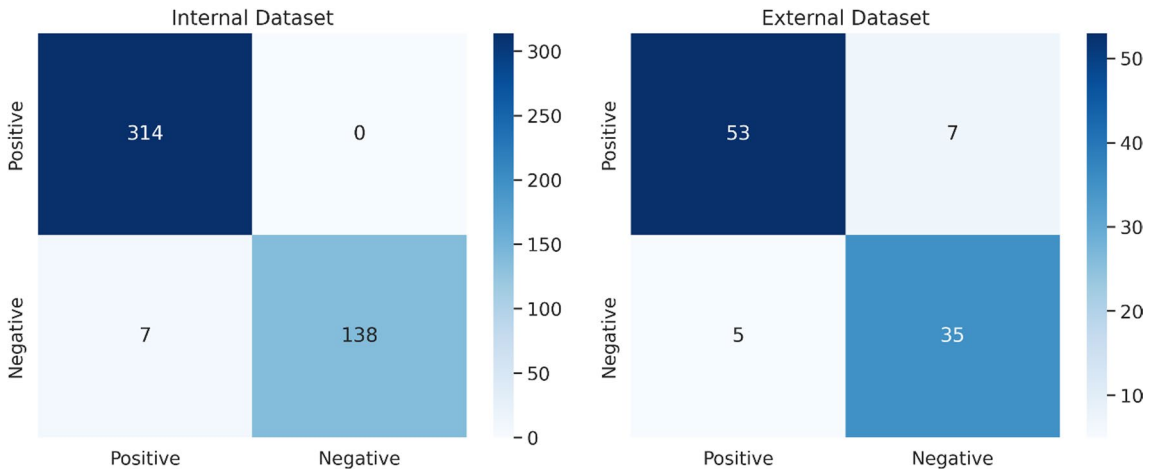

Fig. 9 Confusion matrix of the HND model evaluated with internal dataset (left) and external dataset (right)

Table 3 Performance comparison with studies on LUNA 16

Study	Methodology	Data augmentation	Cross validation	Sensitivity (%)	Accuracy (%)
Nguyen et al. [16]	Fast R-CNN and 2D network	Yes	Tenfold	93.8	95.7
Agnes et al. [20]	UNet + Pyramid Dilated Conv. LSTM	Yes	Fivefold	93.0	96
Fan et al. [18]	R-CNN and 3D ResNet	No	No	93.6	–
George et al. [25]	YOLOv5	Yes	No	–	93
Huang et al. [26]	3D-YOLOv3	Yes	Fivefold	96.2	–
Our study	HND model	No	No	97.8	98.4

is inherently more challenging to detect compared to diagnostic CT images. Two nodules were missed in the detection process. Both undetected nodules were tiny and adhered to the lung's wall. This indicates that our challenge in misclassifying these types of nodules as non-nodules by the model may worsen when dealing with low-dose images.

We conducted a comparative analysis between our study and other studies that were trained and tested using the LUNA 16 dataset. To decrease false positives, Nguyen et al. [16] used a fast R-CNN and a 2D network. As a result, an accuracy of 95.7% was achieved with a sensitivity of 93.8% in a tenfold cross-validation. Distinctly, our methodology employs a 3D network, enabling an in-depth analysis of the data surrounding lung nodes. This captures a broader context, positioning the proposed technique as a marked improvement over the constraints of 2D methodologies. Our 3D architecture effectively mitigates variations across slices, translating to fewer false positives and an enhanced ability to discern nuanced patterns. This is achieved through the 3D model's ability to interpret inter-slice connections, a capability lacking in 2D models that treat each slice in isolation. Furthermore, this approach harmonizes with the study's core focus on analyzing nodules and their adjacent area within a 3D framework, a task that benefits from the inherent ease of 3D analysis. Additionally, utilizing an external dataset decreased their sensitivity to 89.3%. This sensitivity reduction closely mirrors the decrease seen in our model's sensitivity upon employing the external dataset. However, the nature of our dataset, involving low-dose correction CT scans and the utilization of a more challenging 3D mode as opposed to a 2D mode, potentially enhances the reliability of our results. Moreover, in a study conducted by Agnes et al. [20], a UNet-based model combined with a pyramid-dilated convolutional LSTM resulted in a sensitivity of 93%. Their proposed CAD system comprises several modules, with the initial module utilizing Atrous UNet + for semantic segmentation to identify candidate nodules from CT axial slices. The subsequent module incorporates the Pyramid Dilated Convolutional LSTM (PD-CLSTM) classifier, trained on the LUNA16 dataset, to discern genuine nodules from false ones. They proposed a comprehensive but complex model, which may not always be efficient. In addition to employing a data augmentation technique and reserving 20% of the dataset for testing, the study obtained results that are likely similar to our own findings. In another study [18], Fan et al. used an R-CNN model and 3D ResNet (consisting of 50 deep layers) to detect nodules and achieved a sensitivity of 93.6%, whereas the HND model in this study exhibited a sensitivity (recall) of 97.8%. Consequently, a complex classifier is not required for nodule classification when the nodules are cropped. The 3D CNN classifier yielded superior results with fewer layers (17 vs. 50), which could explain the better convergence and lower likelihood of overfitting. George et al. [25] developed an end-to-end process by analyzing all input images for lesion detection and classification. They achieved a precision of 89%, which was better than that of the YOLOv5s model in this study but inferior to the overall precision of the HND model. This improvement in accuracy suggests that the 3D CNN classifier performed well in minimizing the occurrence of incorrect positive results, also known as false positives. Huang et al. [26] applied a single-stage 3D-YOLOv3 model and achieved a sensitivity of 96.2% on a fivefold validation strategy using data augmentation, which was inferior to that observed in this study. On the other hand, our approach was to maintain the study as natural as possible without augmenting the data. Table 3 presents a detailed comparison among the studies.

The study's limitations include the reliance on a specific dataset for evaluation, which raises concerns about the generalizability of the developed hierarchical system to different datasets and clinical settings. Validating the system using larger and more diverse datasets, conducting external validation, and performing prospective clinical studies are recommended to overcome these limitations. Furthermore, in our study, we solely relied on YOLOv5s as our chosen architecture. However, we also recognized the value of modifying the YOLOv5s architecture, taking inspiration from studies like Huang's [26] for a single-stage detection/classification, which incorporated attention layers into the model. Finally, optimizing the computational requirements and resource utilization would enhance the practical implementation of the system.

This study demonstrated that hierarchical networks would provide an efficient pipeline for nodule detection in the lungs. Additionally, the 3D CNN classifier can be used in conjunction with other algorithms, such as segmentation frameworks, to ensure the validity of the results and improve the overall precision of the model.

6 Conclusion

This study developed a hierarchical method consisting of two phases for detecting and categorizing lung nodules in CT scans. In the first phase, the YOLOv5s model is applied to the entire CT image, achieving a high sensitivity for identifying nearly all suspicious nodules. In the second phase, a 3D classifier identifies false-positive cases (i.e., non-nodules), significantly improving the overall precision of the model. The model achieved an accuracy of 98.4% and an AUC of 98.9%.

Funding We would like to state that no external funding or financial support was received for this research. The study was conducted independently without any monetary contributions from any organization or entity.

Declarations

Conflict of interest The authors declare no financial or non-financial conflicts of interest.

Ethical approval All procedures performed were in accordance with the ethical standards of the internal institution's ethical committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

1. Wood DE, Kazerooni EA, Aberle D, Berman A, Brown LM, Eapen GA, et al. NCCN Guidelines® Insights: Lung Cancer Screening, Version 1.2022: Featured Updates to the NCCN Guidelines. J Nat Compreh Cancer Net. 2022. https://doi.org/10.6004/jnccn.2022.0036.
2. Cruickshank A, Stieler G, Ameer F. Evaluation of the solitary pulmonary nodule. Intern Med J. 2019. https://doi.org/10.1111/imj.14219.
3. Philip B, Jain A, Wojtowicz M, Khan I, Voller C, Patel RS, et al. Current investigative modalities for detecting and staging lung cancers: a comprehensive summary. Indian journal of thoracic and cardiovascular surgery. 2023. https://doi.org/10.1007/s12055-022-01430-2.
4. MacRedmond R, McVey G, Lee M, Costello R, Kenny D, Foley C, et al. Screening for lung cancer using low dose CT scanning: results of 2 year follow up. Thorax. 2006. https://doi.org/10.1136/thx.2004.037580.
5. Pehrson LM, Nielsen MB, Ammitzbøl LC. Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review. Diagnostics. 2019. https://doi.org/10.3390/diagnostics9010029.
6. Xiao Z, Liu B, Geng L, Zhang F, Liu Y. Segmentation of lung nodules using improved 3D-UNet neural network. Symmetry. 2020. https://doi.org/10.3390/sym12111787.
7. Nguyen T, Hua B-S, Le N. 3d-ucaps: 3d capsules unet for volumetric image segmentation. Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24: Springer. 2021; https://doi.org/10.1007/978-3-030-87193-2_52.
8. Gu Y, Lai Y, Xie P, Wei J, Lu Y. Multi-scale prediction network for lung segmentation. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019): IEEE; 2019; https://doi.org/10.1109/ISBI.2019.8759207.
9. Chon A, Balachandar N, Lu P (2017). Deep convolutional neural networks for lung cancer detection. Standford University https://doi.org/10.14445/22312803/IJCTT-V67I11P104.
10. Tran GS, Nghiem TP, Nguyen VT, Luong CM, Burie J-C. Improving accuracy of lung nodule classification using deep learning with focal loss. Journal of healthcare engineering. 2019. https://doi.org/10.1155/2019/5156416.
11. Alakwaa W, Nassef M, Badr A (2017). Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). Int J Adv Comp Sci Appl https://doi.org/10.14569/IJACSA.2017.080853.
12. Zhang G, Yang Z, Gong L, Jiang S, Wang L. Classification of benign and malignant lung nodules from CT images based on hybrid features. Phys Med Biol. 2019. https://doi.org/10.1088/1361-6560/ab2544.
13. Kopelowitz E, Engelhard G (2019). Lung nodules detection and segmentation using 3D mask-RCNN. arXiv preprint arXiv:190707676. https://doi.org/10.48550/arXiv.1907.07676.
14. Cai L, Long T, Dai Y, Huang Y. Mask R-CNN-based detection and segmentation for pulmonary nodule 3D visualization diagnosis. IEEE Access. 2020. https://doi.org/10.1109/ACCESS.2020.2976432.
15. Pereira FR, De Andrade JMC, Escuissato DL, De Oliveira LF. Classifier ensemble based on computed tomography attenuation patterns for computer-aided detection system. IEEE Access. 2021. https://doi.org/10.1109/ACCESS.2021.3109860.
16. Nguyen CC, Tran GS, Burie J-C, Nghiem TP. Pulmonary nodule detection based on faster R-CNN with adaptive anchor box. IEEE Access. 2021. https://doi.org/10.1109/ACCESS.2021.3128942.
17. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recognit. 2016. https://doi.org/10.1109/CVPR.2016.90.
18. Fan R, Kamata S-i, Chen Y. Pulmonary nodule detection using improved faster R-CNN and 3D Resnet. Thirteenth International Conference on Digital Image Processing (ICDIP 2021): SPIE. 2021; https://doi.org/10.1117/12.2599884.
19. Kim J-a, Sung J-Y, Park S-h. Comparison of Faster-RCNN, YOLO, and SSD for real-time vehicle type recognition. 2020 IEEE international conference on consumer electronics-Asia (ICCE-Asia): IEEE. 2020; https://doi.org/10.3390/s20174938.
20. Agnes SA, Anitha J, Solomon AA. Two-stage lung nodule detection framework using enhanced UNet and convolutional LSTM networks in CT images. Comput Biol Med. 2022. https://doi.org/10.1016/j.compbiomed.2022.106059.
21. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. Proc IEEE Conf Comput Vis Pattern Recognit. 2016. https://doi.org/10.1109/CVPR.2016.91.
22. Al-Masni MA, Al-Antari MA, Park J-M, Gi G, Kim T-Y, Rivera P, et al. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD

system. Comput Methods Programs Biomed. 2018. https://doi.org/10.1016/j.cmpb.2018.01.017.
23. Baccouche A, Garcia-Zapirain B, Olea CC, Elmaghraby AS (2021). Breast Lesions Detection and Classification via YOLO-Based Fusion Models. Comp Mater Cont https://doi.org/10.32604/cmc.2021.018461.
24. Nie Y, Sommella P, O'Nils M, Liguori C, Lundgren J. Automatic detection of melanoma with yolo deep convolutional neural networks. 2019 E-Health and Bioengineering Conference (EHB): IEEE. 2019; https://doi.org/10.1109/EHB47216.2019.8970033.
25. George J, Skaria S, Varun V (2018). Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans. Medical Imaging 2018: Computer-Aided Diagnosis: SPIE. https://doi.org/10.1117/12.2293699.
26. Huang Y-S, Chou P-R, Chen H-M, Chang Y-C, Chang R-F. One-stage pulmonary nodule detection using 3-D DCNN with feature fusion and attention mechanism in CT image. Comput Methods Programs Biomed. 2022. https://doi.org/10.1016/j.cmpb.2022.106786.
27. Srivastava S, Divekar AV, Anilkumar C, Naik I, Kulkarni V, Pattabiraman V. Comparative analysis of deep learning image detection algorithms. Journal of Big data. 2021. https://doi.org/10.1186/s40537-021-00434-w.
28. Ahmed KR. Smart pothole detection using deep learning based on dilated convolution. Sensors. 2021. https://doi.org/10.3390/s21248406.
29. Mahendrakar T, Ekblad A, Fischer N, White R, Wilde M, Kish B, et al. Performance study of yolov5 and faster r-cnn for autonomous navigation around non-cooperative targets. 2022 IEEE Aerospace Conference (AERO): IEEE. 2022; https://doi.org/10.1109/AERO53065.2022.9843537.
30. Armato S. rd, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. Med Phys https://doi.org/10.1118/1.3528204.
31. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13: Springer. 2014; https://doi.org/10.1007/978-3-319-10602-1_48.
32. Li Z, Tian X, Liu X, Liu Y, Shi X. A two-stage industrial defect detection framework based on improved-yolov5 and optimized-inception-resnetv2 models. Appl Sci. 2022. https://doi.org/10.3390/app12020834.
33. Kim D, Park S, Kang D, Paik J. Improved center and scale prediction-based pedestrian detection using convolutional block. 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin): IEEE. 2019; https://doi.org/10.1109/ICCE-Berlin47944.2019.8966154.
34. Wang W, Xie E, Song X, Zang Y, Wang W, Lu T, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network. Proceedings of the IEEE/CVF international conference on computer vision. 2019; https://doi.org/10.48550/arXiv.1908.05900.