Identification and Classification of Diseases Based on Object Detection and Majority Voting of Bounding Boxes

Satanat Kitsiranuwat [1], Thitipong Kawichai [2], and Paisit Khanarsa [3,*]

[1] Department of Computer Science, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand; Email: satanat@tu.ac.th (S.K.)
[2] Department of Mathematics and Computer Science, Academic Division, Chulachomklao Royal Military Academy, Nakhon Nayok 26001, Thailand; Email: thitipong.kaw@crma.ac.th (T.K.)
[3] Institute of Field Robotics, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
*Correspondence: paisit.kha@kmutt.ac.th (P.K.)

***Abstract*—The intelligence identification and classification of diseases could assist in the early diagnosis and treatment of diseases. With deep learning models, object detection has been applied for real-time detection of plant and human diseases, including apple leaf diseases and skin diseases. However, multiple bounding boxes of lesions can be identified in the same image without advantageously summarizing the information of bounding boxes for classifying a disease image. In this work, a new approach that integrates object detection with Majority Voting of Bounding boxes (MVB) is proposed for the classification of 1,500 images of three common apple leaf diseases and three well-known skin diseases. With MVB, Fast-Region-based Convolutional Neural Network (Fast-RCNN), Faster-RCNN, and YOLOv5 are selected to implement the proposed models. Consequently, Fast-RCNN and Faster-RCNN with MVB outperform the modern deep learning models (EfficientNetB7, RestNet152V2, VGG19, and MobileNetV2) with an accuracy of more than 99% in the classification of apple leaf diseases. Furthermore, Faster-RCNN and YOLOv5 with MVB possess the highest accuracy of 87% in the classification of skin diseases, when compared to the common and recently developed classification models. This research indicates that the proposed approach provides an effective way to identify lesions and classify disease images based on MVB, which can efficiently take advantage of information of bounding boxes for image classification.**

***Keywords*—deep learning, image classification, majority voting, object detection**

I. Introduction

Image classification is an important research topic in pattern recognition and computer vision. It has a wide range of applications, including an identification of plant and human diseases from images containing lesions. Apple is one of the most popular fruits in the world which is globally grown. One of the major causes that can reduce yields and quality of apples is apple leaf diseases [1]. Early detection and identification of apple leaf diseases are critical things to control the spread of infections [2]. To support efficient and accurate identification of apple leaf diseases, Deep Learning (DL) methods have been proposed for identifying diseases from images of apple leaves [3]. Liu *et al*. [2] developed a new Convolutional Neural Network (CNN) model based on AlexNet to diagnose four common apple leaf diseases [2]. Yu *et al*. [4] proposed Region-of-Interest-aware Deep Convolutional Neural Network (ROI-aware DCNN) to divide an input image into three areas (i.e., background, leaf area, and lesion area) and then classify apple leaf diseases principally based on identified lesion areas [4]. Thapa *et al*. [5] introduced a classification model based on ResNet50 for identifying foliar diseases of apples [5].

With almost one-third of the world's population affected by skin illness, skin diseases are ranked as the fourth most prevalent human disease [6]. Early detection of skin diseases and immediately receiving treatment to limit the growth and spread of abnormal skin cells can increase the probability of successful treatment of skin diseases [7]. To support automatic identification of skin diseases from skin images, many computational methods have been developed based on various machine learning and DL models. Alamdari *et al*. [7] presented several image segmentation techniques with two machine learning models (i.e., support vector machine and fuzzy-c-means models) for detecting and classifying types of acne lesions from skin images taken by mobile phones [8]. Bajwa *et al.* [9] deployed several deep neural network models for diagnosis of skin diseases [9]. Srinivasu *et al.* proposed a DL method based on MobileNet V2 and Long Short-Term Memory (LSTM) to classify images of seven skin diseases [10].

To perform real-time detection of apple leaf diseases and skin diseases, various object detection models have been introduced to detect and classify lesions identified in images. Jiang *et al*. [11] proposed the single-shot multibox detector with inception module and rainbow

Manuscript received May 29, 2023; revised July 18, 2023; accepted July 31, 2023; published December 1, 2023.

concatenation (INAR-SSD) model to detect five common apple leaf diseases [11]. Matthew *et al*. utilized You Only Look Once version 3 (YOLOv3) to detect three common diseases and healthy areas in apple leaves [12]. Nie *et al*. [13] proposed automatic detection of melanoma using variants of YOLO [13]. Dwivedi *et al.* [14] introduced a method for automated detection of skin diseases using the fast Region-based Convolutional Neural Network (R-CNN) model [14].

By deploying object detection models, multiple bounding boxes of lesions can be identified in the same apple leaf or skin image and could be served as clear evidence of diagnosis. However, there is still not a method that advantageously utilizes information of bounding boxes for diagnosis or classification of apple leaf and skin disease images. In addition, inaccurate bounding boxes may be created due to several challenging factors in detection of apple leaf diseases and skin diseases. In case of apple leaf diseases, detection failures may be caused by tiny sizes of disease spots, similar characteristics of some different diseases, and inappropriate environmental factors (e.g., background, blur, and lighting) [11]. Erroneous identification of skin diseases can occur by different lesion types of the same disease, confusing characteristics among different diseases, varying skin colors, and distinct skin types [15]. Direct interpretation of false detected bounding boxes may lead to inaccurate classification of apple leaf diseases and skin diseases. To alleviate the effect of false detection, an ensemble method of bounding boxes to classify apple leaf and skin diseases is required.

In this study, a new approach that integrates object detection and a classification technique based on majority voting is proposed for identification and classification of apple leaf diseases and skin diseases from images of apple leaves and skin lesions. To classify apple leaf diseases and skin diseases, we introduce the Majority Voting of Bounding boxes (MVB) algorithm to summarize the information of multiple bounding boxes identified in the images. Several common DL classification models are used as standard models and compared with the proposed method to evaluate the efficiency of the proposed method. Additionally, we evaluate the performance of the proposed method by comparing with some existing methods recently developed for the classification of apple leaf diseases and skin diseases. The main contributions of this study are summarized as follows:

- A new approach which integrates object detection with a majority voting technique is proposed for identification and classification of apple leaf and skin diseases from images containing disease lesions. The outputs of the proposed method are a disease class label of an image and bounding boxes of lesions, which could be served as clear evidence of an automated diagnosis of apple leaf diseases and skin diseases.
- By the proposed majority voting algorithm, the effect of erroneous detection of diseases can be mitigated in disease classification. Additionally, the average strategy, another common ensemble technique, is compared with the majority voting classification technique to demonstrate the superiority of MVB.
- The superiority of the proposed method is validated with two distinct domains of images, which are apple leaf diseases and human skin diseases, to suggest the practicality of the proposed approach in other domains of disease images.

II. Materials and Methods

The overview of this study is summarized into five steps, as shown in Fig. 1. First, images of two case studies, including apple leaf diseases and skin diseases, are collected from public databases. Second, image preparation is performed to create data in a suitable format for further using in model creation. Third, the prepared image data are split into three groups which are testing sets, training sets, and validating sets for developing image classification models. Fourth, the image classification models based on some common DL algorithms and the proposed approach are executed with the training and validating sets. Finally, the performance results of two approaches, including the DL-based classification models and the proposed models, are evaluated and compared using the testing sets.

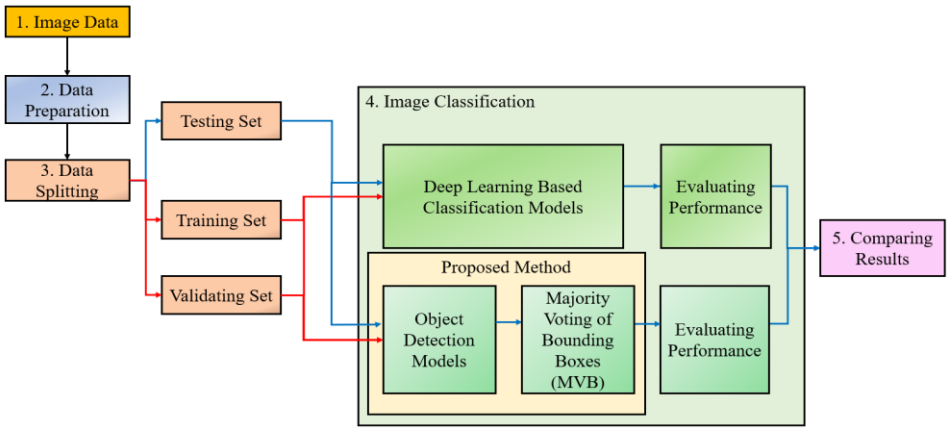


Figure 1. The overview of this study.

A. Datasets

In this study, the datasets of apple leaf diseases and skin diseases are used to evaluate the classification performance of the proposed method. The images of apple leaf diseases were obtained from the PlantVillage dataset [16]. This dataset contains apple leaf images of three most common diseases including apple scab, black rot, and cedar apple rust. Each image is labeled with only one apple leaf disease from those three apple leaf diseases. In case of skin diseases, the image data of skin diseases were extracted from DermNet [17], a database freely providing the collections of skin lesion images. According to the statistical information from the institute of dermatology, Thailand [18], we focused on only the images of top three most common skin diseases, which are acne, dermatitis, and psoriasis. Similarly, each skin image from DermNet represents only one skin disease.

B. Data Preparation

The images in both PlantVillage and DermNet datasets are randomly selected to obtain 250 representing images per class label. In total, there are 750 images of apple leaf diseases and 750 images of skin diseases. Then, all selected images are resized to (640, 640) to prepare the images to be all identical sizes. The batch size of the images is set to 32. After that, the graphical annotation tool, named LabelImg [19], is used to label object bounding boxes in the images of both apple leaf diseases and skin diseases. All annotations are exported for further using with object detection models.

The prepared image data is divided into three groups which are the training sets, validating sets, and testing sets. The training sets are used to create DL-based classification models and object detection models while the validating sets are applied to prevent overfitting of weighted parameters. The testing sets are generated for evaluating the performance of DL-based classification models and the proposed method. In both datasets, images of each class are split into 170, 30, and 50 images for serving as a training set, validating set, and testing set, respectively. In total, there are 510 training images, 90 validating images, and 150 testing images.

C. DL-Based Classification Models

During several last years, DL-based algorithms have been widely utilized for image classification, including the classification of apple leaf and skin disease images. In this study, we used DL-based classification models as the standard models to compare with the proposed method. Some common DL classification algorithms, including EfficientNetB7 [20], ResNet152V2 [21], VGG19 [22], and MobileNetV2 [23], from Keras applications are selected based on accuracy [24].

EfficientNetB7 is an architecture of pre-trained models based on the uniform scaling method of depth, width, and resolution for effective accuracy [20]. ResNet152V2 is an identity mapping in deep residual building blocks to analyze the propagation formulations of connection mechanisms [21]. Visual Geometry Group 19, also known as VGG19, is a deep convolutional neural network for large-scale image recognition [22]. It consists of 19 layers (16 convolution layers, 3 fully connected layers, five max-pooling layers, and a SoftMax layer). MobileNetV2 is a mobile architecture based on inverted residual blocks with linear bottleneck features to reduce parameter counts and support image sizes greater than (32, 32) [23]. MobileNetV2 contains fully convolution layers with 32 filters and 19 residual bottleneck layers. For the pre-trained deep convolutional neural network models, a convolutional layer is trained to generate customized classifiers, to regularize with a dropout layer, and to activate a densely connected classifier with a SoftMax layer.

D. Majority Voting of Bounding Boxes (MVB)

In the proposed approach, an object detection model is first applied to automatically identify bounding boxes of disease lesions in each image. Then, all detected bounding boxes are used to predict a disease class of an image using the proposed algorithm named Majority Voting of Bounding boxes (MVB). Three modern object detection models selected to implement our method are Fast-RCNN [25], Faster-RCNN [26], and YOLOv5 [27]. Fast-RCNN and Faster-RCNN are region-based deep convolutional neural networks which classify object proposal [25, 26]. Faster-RCNN is the modified version of the Fast-RCNN model, which avoids the computational cost of selective search and utilizes the Region Proposal Net (RPN) to generate candidate regions proposal. YOLO is a real-time object detection algorithm which identifies a detection region and categorize objects in an image [27]. The original version of YOLO cannot detect small targets due to lack of detailed grid division [27]. The fifth version of YOLO (YOLOv5) was proposed to solve this problem by transmitting a batch through the data loader to perform data enhancements.

By applying an object detection model, bounding boxes of disease lesions found in an image are produced. Each bounding box has its disease class label and confidence score, which is the probability that a bounding box precisely contains a disease lesion. A high confidence score of a bounding box implies high importance of that region for disease classification. Sometimes, multiple bounding boxes of disease lesions may be detected in an image by an object detection algorithm. Therefore, we propose the Majority Voting of Bounding boxes (MVB) algorithm to classify an entire image based on the class labels (c_i) and confidence scores (s_i) of n bounding boxes detected in an image, where $i = 1, 2, \ldots, n$.

According to Algorithm, the lists of class labels (C) and confidence scores (S) of all bounding boxes detected in an image are inputs for classifying the image. c_i and s_i represent the class label and confidence score of the i^{th}-bounding box from n bounding boxes, respectively. Initially, the class label that appears most often in C, also known as mode, is identified. If there is one mode, then that most frequent class label is given as a predicted class of the image. In cases that there is no mode or more than one mode, MVB will find the class label which possesses the highest confidence score and use that class label as a

predicted class of the image. According to algorithm, the time complexity of MVB directly depends on the time complexity of the mode function and the argmax function. Therefore, MVB has a time complexity of $O(n)$, where n is the number of bounding boxes.

Algorithm: Majority Voting of Bounding Boxes (MVB)

Input: List of class labels of all bounding boxes $C = [c_1, c_2, \ldots, c_n]$ and list of confidence scores of all bounding boxes $S = [s_1, s_2, \ldots, s_n]$

Output: A predicted class of an image (*predicted_class*)

1. Find the most frequent class label(s) in C, denoted as *mode*.
2. **if** *mode* has only one value **then**
3. *predicted_class* = *mode*
4. **else**
5. $k = \underset{i}{\text{argmax}}\ s_i$
6. *predicted_class* = c_k
7. **end if**
8. **return** *predicted_class*

E. Performance Evaluation

For evaluating the performance of both DL-based classification models and the proposed method, we computed precision, recall, accuracy, and F1-score (F1) following to Eqs. (1)–(3), where *TP*, *FP*, *FN*, *TN* are the number of true positives, false positives, false negatives, and true negatives, respectively.

$$\text{Precision} = \frac{TP}{TP + FP}, \ \text{Recall} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{2}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

For object detection, an intersection over union (IoU) is a typical way to assess whether one object proposal is correct or not. An IoU can be calculated by dividing an overlapping area between a predicted bounding box (B_p) and a ground truth bounding box (B_{gt}) by a union area between those two bounding boxes, as in Eq. (4).

$$\text{IoU} = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p \cup B_{gt})} \tag{4}$$

Based on IoU values, *TP*, *FP*, and *FN* can be calculated. *TP* represents the number of detections that have IoU values greater than or equal to a threshold, whereas *FP* is the number of detections that have IoU values less than a threshold. *FN* indicates the number of no detection. A threshold usually sets to 50%, 75%, or 95% [28]. A precision-recall curve can be created by varying IoU threshold values, generally from 0.5 to 0.95. For each class, an area of a precision-recall curve, or an Average Precision (AP), can be estimated to reflect the performance of an object detection model. Additionally, a mean Average Precision (mAP) can be calculated by averaging AP scores over all object classes.

III. Results

In this section, the image classification of apple leaf diseases (apple scab, black rot, and cedar apple rust) and skin diseases (acne, dermatitis, and psoriasis) are investigated. The classification performance of the DL-based classification models, including EfficientNetB7, ResNet152V2, VGG19, and MobileNetV2, are demonstrated. Moreover, the performance scores such as Average Precision (AP) and mean Average Precision (mAP) are calculated to primarily show the efficiency of the selected object detectors, including Fast-RCNN, Faster-RCNN, and YOLOv5. In the proposed approach, we applied the MVB technique to classify images based on all detected objects produced by the object detection algorithms. The precision, recall, F1, and accuracy scores are evaluated to indicate the classification performance of both the DL-based classification models and the proposed models. In addition, the best classification performance of the DL-based classification models, the proposed models and the previous work are investigated to compare the efficiency of three approaches in classifying images of apple leaf and skin diseases. More details of the results can be found at https://github.com/PaisitPor/JAIT-Object-Detection-and-Majority-Voting-of-Bounding-Boxes.

A. Performance of DL-Based Classification Models

The selected DL-based classification models are EfficientNetB7, ResNet152V2, VGG19, and MobileNetV2. The Stochastic Gradient Descent (SGD) optimizer was used to train all DL-based classification models with the learning rates ranging from 0.0001 to 0.01. The performance of the DL-based classification models, including the precision, recall, F1, and accuracy scores, are evaluated in the classification of apple leaf disease images (Table I) and skin disease images (Table II).

TABLE I. Performance of the DL-based Classification Models in the Classification of Apple leaf Diseases

Performance	DL-based classification models			
	Efficient NetB7	**ResNetV2**	**VGG19**	**Mobile NetV2**
Precision	0.99	0.99	0.90	0.83
Recall	0.99	0.99	0.89	0.75
F1	0.99	0.99	0.88	0.75
Accuracy	0.99	0.99	0.89	0.75

In Table I, EfficientNetB7 and ResNetV2 show the highest precision, recall, F1, and accuracy scores. According to accuracy, the EfficientNetB7 and RestNetV2 models 99% correctly classify apple disease images. For the other DL-based classification models, VGG19 can separate apple scab, black rot, and cedar apple rust images with the performance scores up to 90% while MobileNetV2 can separate those images with the performance scores only up to 83%.

TABLE II. PERFORMANCE OF THE DL-BASED CLASSIFICATION MODELS IN THE CLASSIFICATION OF SKIN DISEASES

Performance	DL-based classification models			
	Efficient NetB7	ResNetV2	VGG19	Mobile NetV2
Precision	0.85	0.85	0.78	0.80
Recall	0.83	0.85	0.76	0.80
F1	0.83	0.84	0.76	0.80
Accuracy	0.83	0.85	0.76	0.80

In Table II, ResNetV2 has the highest precision, recall, F1, and accuracy scores with values more than 0.83. According to accuracy, ResNetV2 85% correctly classifies skin disease images. A model whose performance values are very close to those of ResNetV2 is EfficientNetB7. Its performance values are in the range of 0.83 to 0.85. The performance scores of MobileNetV2 and VGG19 in the classification of skin disease images are only up to 0.80 and 0.78, respectively. In summary, RestNetV2 and EfficientNetB7 outperform the other DL-based classification models when classifying the images of both apple leaf and skin diseases.

B. Performance of Object Detectors

The Average Precision (AP) and mean Average Precision (mAP) are calculated to primarily observe the performance of the selected object detectors, including Fast-RCNN, Faster-RCNN, and YOLOv5. AP is used to compare the performance of object detectors for each image label, whereas mAP indicates an average value of AP scores over all disease classes. Additionally, AP (IoU = 0.50:0.95) and mAP (IoU = 0.50:0.95) are AP and mAP scores averaged over different IoU thresholds ranging from 0.50 to 0.95 with a step size of 0.05.

The performance results of the object detection models in identifying apple leaf diseases and skin diseases are shown in Table III and Table IV, respectively. Table IV shows that the mAP scores of Fast-RCNN, Faster-RCNN, and YOLOv5 are 0.2172, 0.2170, and 0.2010, respectively. This means that Fast-RCNN, Faster-RCNN and YOLOv5 can detect approximately up to 22% of all lesions in an apple leaf image. When considering each apple disease class, we found that Faster-RCNN, YOLOv5, and Fast-RCNN show the best performance in identifying apple scab, black rot, and cedar apple rust, respectively.

TABLE III. PERFORMANCE OF THE OBJECT DETECTION MODELS IN THE DETECTION OF APPLE LEAF DISEASES

Models	AP (IoU = 0.50:0.95)			mAP (IoU = 0.50:0.95)
	Apple scab	Black rot	Cedar apple rust	
Fast-RCNN	0.1005	0.3393	0.2119	0.2172
Faster-RCNN	0.1175	0.3367	0.1957	0.2170
YOLOv5	0.0863	0.3760	0.1410	0.2010

In Table IV, the mAP scores of Fast-RCNN, Faster-RCNN, and YOLOv5 are 0.1721, 0.1799, and 0.1840, respectively. This means that Fast-RCNN, Faster-RCNN and YOLOv5 can detect approximately 17%–18% of all skin disease lesions in a skin image. When comparing within the same disease class, Faster-RCNN shows the best performance in detecting acne, whereas YOLOv5 outperforms the others in identifying dermatitis and psoriasis diseases.

TABLE IV. PERFORMANCE OF THE OBJECT DETECTION MODELS IN THE DETECTION OF SKIN DISEASES

Models	AP (IoU = 0.50:0.95)			mAP (IoU = 0.50:0.95)
	Acne	Dermatitis	Psoriasis	
Fast-RCNN	0.1502	0.2535	0.1126	0.1721
Faster-RCNN	0.1744	0.2809	0.0843	0.1799
YOLOv5	0.1330	0.3050	0.1140	0.1840

C. Performance of Object Detection Models with Majority Voting of Bounding Boxes

In the proposed approach, the object detection models including Fast-RCNN, Faster-RCNN, and YOLOv5 were employed to initially predict regions or lesions of apple leaf diseases and skin diseases with bounding boxes and their confidence scores. However, those kinds of algorithms are not intended for assigning a disease class for an entire image. To classify apple leaf and skin diseases, the proposed algorithm, named Majority Voting of Bounding boxes (MVB), is applied to the image results obtained from the object detection models. MVB predicts a disease class of an image based on the most commonly predicted class and the confidence scores of bounding boxes.

The performance results of the proposed method in classifying apple leaf diseases and skin diseases are presented in Table V and Table VI, respectively. According to Table V, both Fast-RCNN and Faster-RCNN with MVB reach the best performance in the classification of apple leaf diseases. Nevertheless, all selected object detection models with MVB can correctly classify apple leaf diseases with more than 99% of all performance values.

TABLE V. PERFORMANCE OF THE PROPOSED MODELS IN THE CLASSIFICATION OF APPLE LEAF DISEASES

Performance	Object detection models with MVB		
	Fast-RCNN	Faster-RCNN	YOLOv5
Precision	1.00	1.00	0.99
Recall	1.00	1.00	0.99
F1	1.00	1.00	0.99
Accuracy	1.00	1.00	0.99

TABLE VI. PERFORMANCE OF THE PROPOSED MODELS IN THE CLASSIFICATION OF SKIN DISEASES

performance	Object detection models with MVB		
	Fast-RCNN	Faster-RCNN	YOLOv5
Precision	0.84	0.87	0.87
Recall	0.83	0.87	0.87
F1	0.83	0.87	0.87
Accuracy	0.83	0.87	0.87

In Table VI, Faster-RCNN and YOLOv5 with MVB outperform Fast-RCNN with all performance values of

0.87. This means that both Faster-RCNN and YOLOv5 correctly classify 87% of all skin disease images on average in the testing sets. Meanwhile, Fast-RCNN can correctly classify 83% of all skin disease images in the testing sets.

D. Comparison of the Classification Performance

In this section, the classification performance of the DL-based classification models, the proposed models, and the models proposed in some previous works are compared in classifying apple leaf diseases and skin diseases. In case of apple leaf disease classification, the Liu's model [2] and the Thapa's model [5] are selected to compare with our method. The Liu's model was proposed in 2017 and is based on AlexNet precursor and cascade inception [2]. The Thapa's model was proposed in 2020 and is based on ResNet50, the best architecture for classifying foliar disease of apples on Pathology Challenge 2020 [5]. Moreover, the Bajwa's model [9], proposed in 2020, is represented as the model of the previous works in the classification of skin diseases. This model is the ensemble model consisting of ResNet-152, DenseNet-161, SE-ResNeXt-101, and NASNet. In addition, the prediction images from the DL-based classification models (i.e., EfficientNetB7, ResNet152V2, VGG19, and MobileNetV2) are further analyzed and illustrated using Grad-CAM. Also, the prediction images from the proposed models with Fast-RCNN, Faster-RCNN, and YOLOv5 are shown.

1) Apple leaf diseases

In classifying apple leaf diseases, the performance scores of the best DL-based classification models (EfficientNetB7 and ResNetV2), the best proposed models (Fast-RCNN and Faster-RCNN with MVB) and the models of the previous works (the Thapa's and Liu's [5] model) are selected to compared and presented in Table VII. The results show that the best scores of all metrics, including precision, recall, F1, and accuracy, of the proposed models are greater than those of the DL-based classification models and the previous models. The proposed models, both Fast-RCNN and Faster-RCNN with MVB, can all correctly predict apple leaf diseases with an accuracy of 100%, which improves the classification performance of the DL-based models and the models of the previous works.

TABLE VII. COMPARISON OF THE BEST PERFORMANCE SCORES IN THE CLASSIFICATION OF APPLE LEAF DISEASES

Performance	Deep Learning		Proposed Models		Previous Works	
	EfficientNetB7	ResNetV2	Fast-RCNN	Faster-RCNN	Liu's Model [2]	Thapa's Model [5]
Precision	0.99	0.99	1.00	1.00	0.89	0.97
Recall	0.99	0.99	1.00	1.00	0.89	0.97
F1	0.99	0.99	1.00	1.00	0.89	0.97
Accuracy	0.99	0.99	1.00	1.00	0.89	0.97

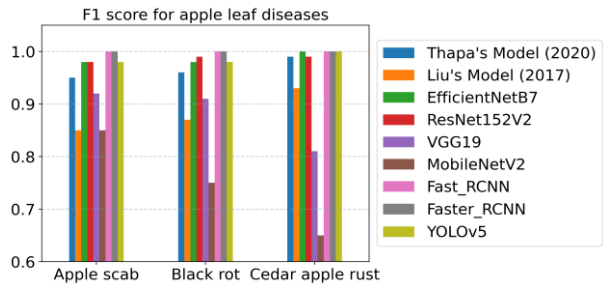


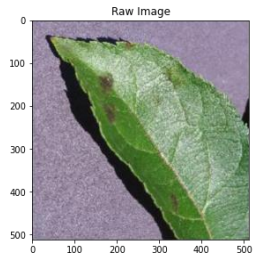
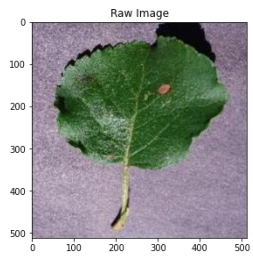
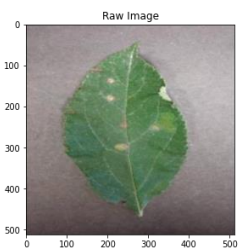
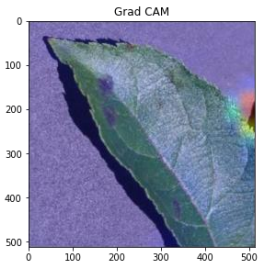
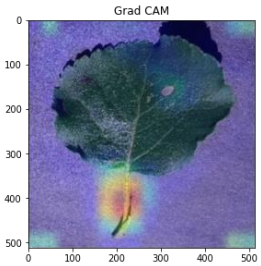
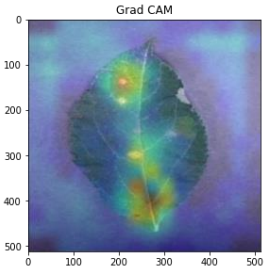
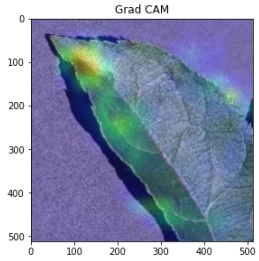
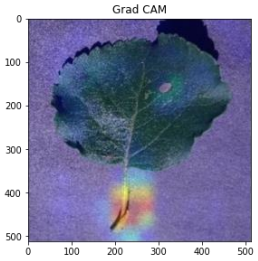
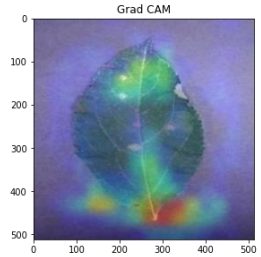
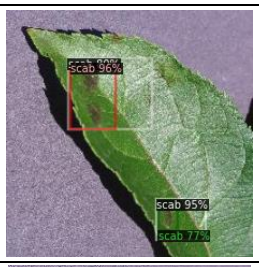
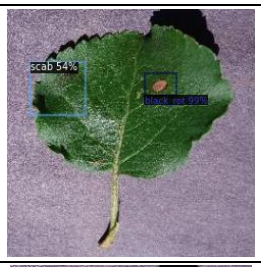
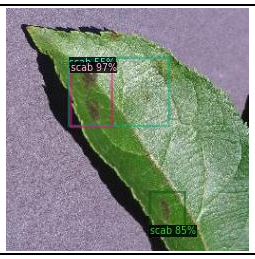
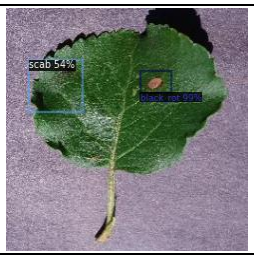
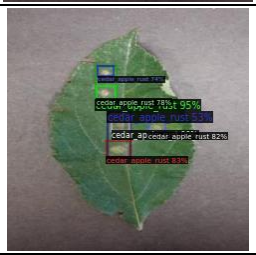
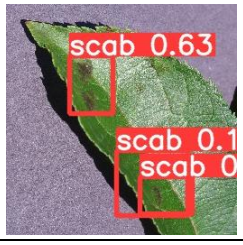
Figure 2. F1 scores of the DL-based classification models and the object detection models with MVB in the classification of apple leaf diseases.

Fig. 2 shows F1 scores of the previous works (the Thapa's model and the Liu's model), the DL-based classification models (i.e., Efficient-NetB7, ResNet-152V2, VGG19, and MobileNetV2), and the proposed models, which combines MVB with the object detection models (i.e., Fast-RCNN, Faster-RCNN, and YOLOv5) in classifying each apple leaf disease. The models that perform remarkably better than the others are EfficientNetB7, ResNet152V2, Fast-RCNN, Faster-RCNN, and YOLOv5 with the F1 scores of more than 98%. Moreover, the performance scores such as precision and recall for detecting apple leaf diseases are illustrated at https://github.com/PaisitPor/JAIT-Object-Detection-and-Majority-Voting-of-Bounding-Boxes. The comparison results showed that precision, recall, and F1 scores of the proposed models (object detection models with MVB) still outperform the others.

In Section III(A–C), the results show that EfficientNetB7, ResNet152V2, and all object detectors (i.e., Fast-RCNN, Faster-RCNN, and YOLOv5) with MVB outperform the others in predicting apple leaf diseases and skin diseases. Therefore, we created the prediction images of those algorithms as illustrated in Table VIII. The first row of the table represents the raw images of diseases apple scab, black rot, and cedar apple rust, respectively.

For the DL-based classification models, the prediction images show the heat-map visualization of Grad-CAM, which can be used to identify important areas for classifying apple leaf diseases. Dark tones are the highlights of highly important regions for image classification. For Fast-RCNN, Faster-RCNN, and YOLOv5, the prediction images present bounding boxes of lesions identified in the images. In Table VIII, most DL-based classification models may incorrectly spot the regions of apple disease lesions. However, they can still accurately classify scab, black rot, and cedar apple rust. Furthermore, most object detection models with the MVB algorithm can detect apple disease lesions precisely.

TABLE VIII. PREDICTION IMAGES FOR APPLE SCAB, BLACK ROT, AND CEDAR APPLE RUST

Actual class	Apple scab	Black rot	Cedar apple rust
Raw image			
EfficientNetB7			
ResNet152V2			
Fast-RCNN			
Faster-RCNN			
YOLOv5			

2) Skin diseases

In classifying skin diseases, the performance scores of the best DL-based classification model (ResNetV2), the best proposed models (Faster-RCNN and YOLOv5), and the Bajwa's model are compared and presented in Table IX. It is noticeable that the proposed models, both Faster-RCNN and YOLOv5 with MVB, outperform the others with an accuracy of 87%. Meanwhile, ResNetv2 (the best DL-based classification model) and the Bajwa's model can achieve an accuracy of up to 85% in the classification of skin diseases.

TABLE IX. COMPARISON OF THE BEST CLASSIFICATION PERFORMANCE SCORES FOR SKIN DISEASES

Performance	Deep Learning	Proposed Models		Previous Work
	ResNet V2	Faster-RCNN	YOLOv5	Bajwa's Model [9]
Precision	0.85	0.87	0.87	0.85
Recall	0.85	0.87	0.87	0.83
F1	0.84	0.87	0.87	0.84
Accuracy	0.85	0.87	0.87	0.84

Fig. 3 shows F1 scores of the previous work (i.e., Bajwa's model in 2020), the DL-based classification models (i.e., Efficient-NetB7, ResNet152V2, VGG19, and MobileNetV2), and the object detection models (i.e., Fast-RCNN, Faster-RCNN, and YOLOv5) with MVB in the classification of acne, dermatitis, and psoriasis. The results show that the YOLOv5, Faster-RCNN, and Fast-RCNN models with the MVB technique outperform the others with F1 scores up to 92%, 85%, and 88% in classifying acne, dermatitis, and psoriasis, respectively. Noticeably, the proposed models achieve superior F1 scores when compared to those of the DL-based classification models. Moreover, the other performance scores such as precision and recall for detecting skin diseases are illustrated in https://github.com/PaisitPor/JAIT-Object-Detection-and-Majority-Voting-of-Bounding-Boxes. The comparison results show that precision, recall, and F1 scores of the proposed models dominate the others.

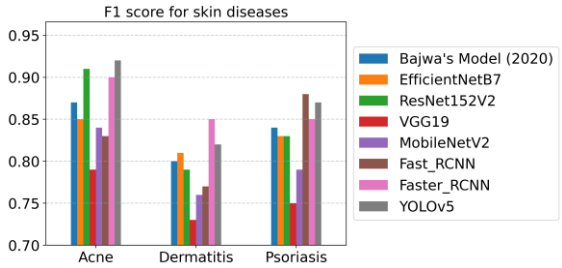
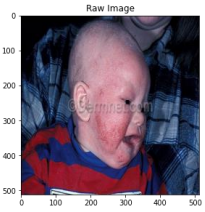
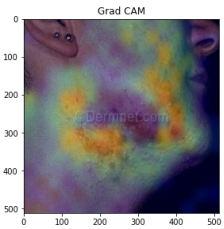
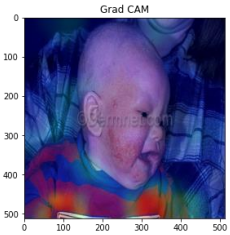
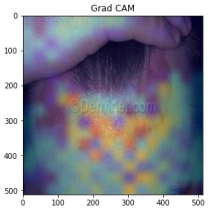
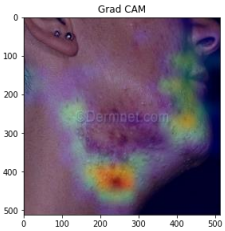
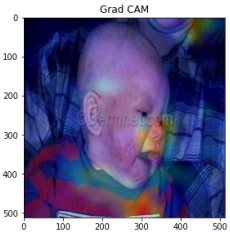
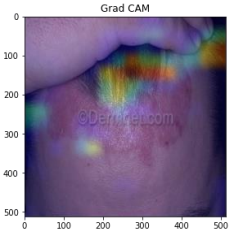
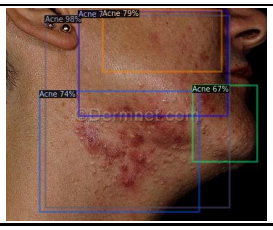
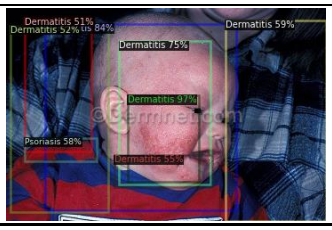
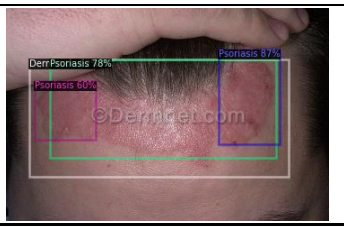
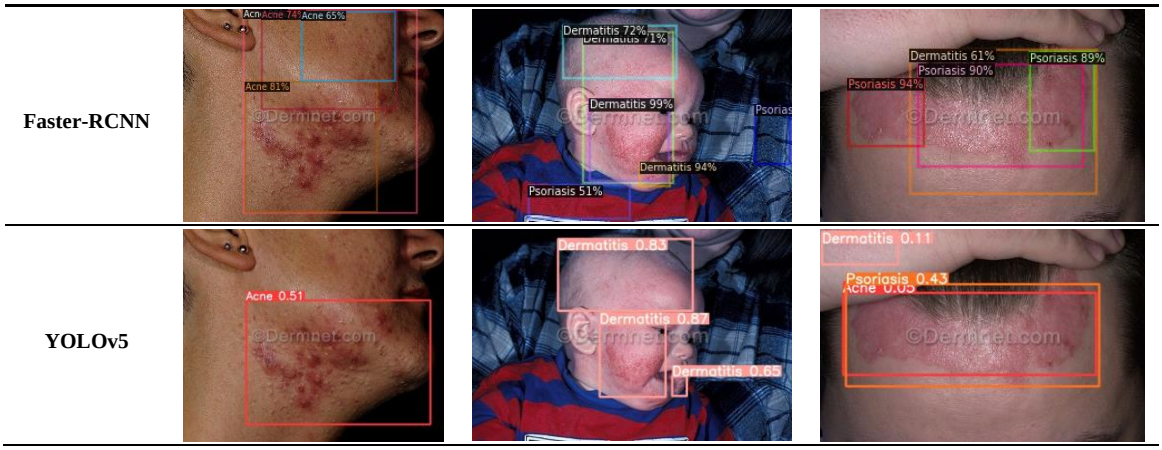


Figure 3. F1 scores of the DL-based classification models and the object detection models with MVB in the classification of skin diseases.

In addition, the prediction images of the DL-based classification models (i.e., EfficientNetB7 and ResNet152V2) and the proposed models, which are based on the object detection models (i.e., Fast-RCNN, Faster-RCNN, and YOLOv5), are illustrated in Table X. The first row of the table represents the raw images of acne, dermatitis, and psoriasis, respectively.

TABLE X. PREDICTION IMAGES FOR ACNE, DERMATITIS, AND PSORIASIS

Actual class	Acne	Dermatitis	Psoriasis
Raw image			
EfficientNetB7			
ResNet152V2			
Fast-RCNN			



The results show that although most DL-based classification models may incorrectly spot the areas of the skin lesions in the images, the DL-based classification models still correctly classify acne and dermatitis. However, the DL models still confuse in classifying psoriasis. Furthermore, most models of object detection algorithms with MVB can detect skin lesions precisely and correctly classify skin diseases of bounding boxes detected in the images.

IV. Discussion

This study introduces a new approach for disease image classification based on majority voting of bounding boxes from object detection. Two study cases, including apple leaf diseases and skin diseases, are employed to validate the performance of the proposed method. Due to big differences in lesion characteristics and their surrounding areas of apple leaf diseases and skin diseases, the performance values of all models in the image classification of apple leaf diseases and skin diseases are moderately different. In case of apple leaf diseases, the surrounding areas of disease lesions are almost the same. With this reason, both DL-based classification models and the proposed models achieves the high-performance values in the classification of apple leaf diseases. In contrast, skin lesions can appear on skin of different parts of body, such as on the chest, hands, feet, or face. With different characteristics and surrounding areas of skin lesions, the performance scores of almost models in the classification of skin diseases are worse than those in the classification of apple leaf diseases, as shown in Section III(A) and III(C).

According to the results in Section III(B), the mAP values of the object detection models are only 20% and 17% in the detection of apple leaf disease lesions and skin lesions, respectively. Those low values of mAP might be because there is a small number of training sets, only 170 images per class for both datasets. Zoph *et al.* studied data augmentation strategies, such as translating, flipping, resizing, brightness, and cropping, for object detection, and they revealed that data augmentation can improve the performance of object detection models [29]. Despite low precision of lesion detection in both apple leaf images and skin images, the majority voting strategy or MVB can alleviate the effects of erroneous detection in the disease classification by ignoring the noise bounding boxes, which present low class frequencies and low confidence scores. This results in the superior performance of the purposed models when compared to the common DL methods and the models of the previous works in both the classification of apple leaf diseases and skin diseases. Without data augmentation, this can be implied that the proposed approach can advantageously learn from the small number of training sets.

According to the visual explanation results using Grad-CAM, it is revealed that the DL-based classification models (i.e., EfficientNetB7, ResNet152V2, VGG19, and MobileNetV2) sometimes inaccurately identify lesion regions in the classification of both apple leaf diseases and skin diseases. Under this situation, the DL-based classification models may produce inaccurate predictions of apple leaf diseases and skin diseases. On the contrary, the object detection models (i.e., Fast-RCNN, Faster-RCNN, and YOLOv5) correctly detect the lesion areas of both apple leaf diseases and skin diseases despite some inaccurate disease predictions of the detected bounding boxes. However, the MVB technique can effectively mitigate the effects of false bounding boxes and improve the performance in the classification of an entire image. In addition, the capability of identifying disease lesions are a major profit of the proposed method when compared to the DL models, automatically extracting features that are difficult to understand. The proposed method classifies apple leaf diseases and skin diseases by imitating the diagnosis of professional physicians or agriculturists, who initially search for suspect regions of disease lesions and then specify diseases from those detected lesions. In practical applications of the proposed approach, the bounding boxes of suspect lesion regions could be reported as supporting evidence of the disease predictions for additional verifications by humans.

Despite many advantages of the proposed approach, there are some limitations of its use. Firstly, the models proposed in this work are developed under the presumption that each image contains only lesions of one disease. This means that the proposed models are still not applicable for identification and classification of multiple diseases in the same image. Secondly, varying values of some parameters, including an intersection over union

(IoU) threshold and a level of confidence threshold in object detection models, have direct impacts on the proposed method. Since the IoU threshold is used to identify and remove duplicate proposals from the set of final proposals, too low values of the IoU threshold could result in too few received final proposals, which could affect MVB depending on mode of disease classes of bounding boxes. The threshold of confidence scores is for including proposals possessing the confidence scores greater than the threshold value into the set of final predictions. Too high values of the confidence threshold could lead to few or no proposals finally obtained and then affect the classification of the proposed method mainly based on mode of classes of bounding boxes. However, by using mode, MVB can well manipulate some false bounding boxes received from object detection models in the image classification (e.g., when an object detection model produces numerous correct bounding boxes with moderate confidence scores and a few of incorrect bounding boxes with a higher confidence score). Moreover, by comparing majority voting with another common ensemble strategy (i.e., an averaging technique), the results demonstrate that the majority voting technique is more suitable for the classification of apple leaf diseases and skin diseases based on bounding boxes received from object detectors (https://github.com/PaisitPor/JAIT-Object-Detection-and-Majority-Voting-of-Bounding-Boxes). Thirdly, because the performance of object detectors could be affected by imbalanced data [30, 31], the proposed method which is mainly based on object detection may not perform well when faced with highly imbalanced data. However, data augmentation techniques, such as translating, flipping, and resizing, can be used to improve the performance of object detection models [29].

V. Conclusion

A new approach based on object detection and a classification technique named majority Voting of Bounding Boxes (MVB) is proposed to identify areas of lesions and classify images of three common apple leaf diseases and three well-known skin diseases. Although the mAP scores of Fast-RCNN, Faster-RCNN, and YOLOv5 are up to 20% in the lesion identification of both apple leaf diseases and skin diseases, MVB can mitigate the effect of false detection of lesions and efficiently classify disease images. In the classification of apple leaf diseases, Fast-RCNN and Faster-RCNN with MVB outperform the common DL models and the models of the previous works with accuracies of greater than 99%. In the classification of skin diseases, Faster-RCNN and YOLOv5 with MVB reach the highest accuracy of 87%, when compared to the common deep learning models and the models of the previous works. According to the superior results of the classification of those two different domains of disease images, it can be suggested that the proposed approach can be potentially extended for the classification of other diseases besides apple leaf diseases and skin diseases. In addition to a disease class label of an image, the proposed models also provide bounding boxes of lesions identified in the image, which could be served as clear evidence of an intelligence diagnosis for further validation by humans.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, S.K., P.K., and T.K.; methodology, S.K. and P.K.; validation, S.K., P.K., and T.K.; formal analysis, S.K., P.K., and T.K.; writing—original draft preparation, S.K.; writing—review and editing, S.K., P.K., and T.K.; all authors had approved the final version.

References

[1] P. Bansal, R. Kumar, and S. Kumar, "Disease detection in apple leaves using deep convolutional neural network," *Agriculture*, vol. 11, no. 7, 617, 2021.

[2] B. Liu, Y. Zhang, D. He, and Y. Li, "Identification of apple leaf diseases based on deep convolutional neural networks," *Symmetry*, vol. 10, no. 1, 11, 2018.

[3] M. Brahimi, M. Arsenovic, S. Laraba, S. Sladojevic, K. Boukhalfa, and A. Moussaoui, "Deep learning for plant diseases: Detection and saliency map visualisation," *Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent*, pp. 93–117, 2018.

[4] H. J. Yu, C. H. Son, and D. Lee, "Apple leaf disease identification through region-of-interest-aware deep convolutional neural network," *Journal of Imaging Science and Technology*, vol. 64, 2020.

[5] R. Thapa, K. Zhang, N. Snavely, S. Belongie, and A. Khan, "The plant pathology challenge 2020 data set to classify foliar disease of apples," *Appl Plant Sci*, vol. 8, no. 9, e11390, Sep. 2020.

[6] T. A. D. Goswami, K. Vipul, and B. P. H. Kumar, "Skin disease classification from image—A survey," presented at the 6th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 12, 2020.

[7] C. Flohr and R. Hay, "Putting the burden of skin diseases on the global map," *British Journal of Dermatology*, vol. 184, no. 2, pp. 189–190, 2021.

[8] N. Alamdari, K. Tavakolian, M. Alhashim, and R. F. Rezai, "Detection and classification of acne lesions in acne patients: A mobile application," in *Proc. 2016 IEEE International Conference on Electro Information Technology (EIT)*, pp. 0739–0743, 2016.

[9] M. N. Bajwa *et al.*, "Computer-aided diagnosis of skin diseases using deep neural networks," *Applied Sciences*, vol. 10, no. 7, 2488, 2020.

[10] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with mobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, 2852, 2021.

[11] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019.

[12] M. P. Mathew and T. Y. Mahesh, "Determining the region of apple leaf affected by disease using YOLO V3," in *Proc. 2021 International Conference on Communication, Control, and Information Sciences (ICCISc)*, vol. 1, pp. 1–4, 2021.

[13] Y. Nie, P. Sommella, M. O'Nils, C. Liguori, and J. Lundgren, "Automatic detection of melanoma with Yolo deep convolutional neural networks," in *Proc. 2019 E-Health and Bioengineering Conference (EHB)*, pp. 1–4, 2019.

[14] P. Dwivedi, A. A. Khan, A. Gawade, and S. Deolekar, "A deep learning based approach for automated skin disease detection using Fast R-CNN," in *Proc. 2021 Sixth International Conference on Image Information Processing (ICIIP)*, 2021, vol. 6, pp. 116–120.

[15] T. Goswami, V. K. Dabhi, and H. B. Prajapati, "Skin disease classification from image—A survey," in *Proc. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 599–605, 2020.
[16] D. P. H. A. M. Salath, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," arXiv preprint, arXiv:1511.08060v2v, 2015.
[17] DermNet. DermNet: Dermatology pictures-skin diseases pictures. Nonprofits organizations. [Online]. Available: https://dermnetnz.org/image-library
[18] I. O. D. Thailand. Institue of Dermatology Thailand Website. Government of Thailand. [Online]. Available: https://www.iod.go.th/
[19] Tzutalin. LabelImg. [Online]. Available: https://github.com/tzutalin/labelImg
[20] M. T. A. Q. V. Le. "EfficientNet: Rethinking model scaling for convolutional neural networks," arXiv preprint, arXiv:1905.11946, 2019.
[21] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," arXiv preprint, arXiv:1603.05027, 2016.
[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint, arXiv:1409.1556, 2014.
[23] M. A. H. Sandler *et al*., "Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation," *CoRR*, 1801, 2018.
[24] A. Gulli and S. Pal, *Deep Learning with Keras*, Packt Publishing Ltd, vol. 8, 2017.
[25] R. B. Girshick, "Fast R-CNN," arXiv preprint, arXiv:1504.08083, 2015.
[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," arXiv preprint, arXiv:1506.01497, 2015.
[27] W. Wu *et al*., "Application of local fully convolutional neural network combined with YOLO v5 algorithm in small target detection of remote sensing image," *PloS One*, vol. 6, no. 10, 2021.
[28] S. L. N. R. Padilla and E. A. B. D. Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. International Conference on Systems, Signals, and Image Processing (IWSSIP)*, 2020, pp. 237–242.
[29] B. Zoph, E. D. Cubuk, G. Ghiasi, T. Y. Lin, J. Shlens, and Q. V. Le, "Learning data augmentation strategies for object detection," in *Proc. Computer Vision – ECCV 2020*, 2020, pp. 566–583.
[30] K. Oksuz, B. C. Cam, S. Kalkan, and E. Akbas, "Imbalance problems in object detection: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3388–3415, 2021.
[31] J. Chen, Q. Wu, D. Liu, and T. Xu, "Foreground-background imbalance problem in deep object detectors: A review," in *Proc. 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2020, pp. 285–290.