Vitis Vinera L. Leaf Detection using Faster R-CNN

Moechammad Sarosa[1], *Puteri Nurul* Ma'rifah[1*], *Mila* Kusumawardani[1], and *Dimas Firmanda Al* Riza[2]

[1]Department of Electrical Engineering, State Polytechnic of Malang, 65141 Malang, Indonesia
[2]Faculty of Agricultural Technology, University of Brawijaya, 65141 Malang, Indonesia

Abstract. Grapes are a type of vine that belongs to the Vitaceae family and has many health benefits. There are dozens of grape varieties that are widespread in Indonesia. Grape varieties can be differentiated based on their various leaf shapes. At first glance, it might look the same. However, if you look at the shape and character of each leaf, grapes have different types and leaf variants. In recent years, various plant leaf classification methods based on deep learning have been proposed. This research uses a deep learning method with the Faster R-CNN ResNet-50 algorithm and uses pre-trained COCO weights to classify grape varieties through leaf images. For this purpose, a dataset of grape leaf images from five varieties was taken independently. Based on the tests that have been carried out, it shows that the improved network can effectively increase the efficiency of network operation. After testing four times ranging from 3,000 steps to 8,000 steps, the accuracy of recognizing leaf variations reached the highest level of 90.11% at 8,000 test steps with a loss of 0.134721. The results of this research show that the algorithm can classify types of grapes based on their leaves.

1 Introduction

Grapes are a type of vine that belongs to the Vitaceae family and has many health benefits [1]. There are dozens of grape varieties that are widespread in Indonesia. Cultivation, potential quality, and selling price of grapes depend on the variety [2]. Therefore, the interest of grape growers in the accurate identification of grape varieties is increasing, whether to determine the correct way of propagating grape cuttings according to the variety or to estimate the supply price of grapes. Grape varieties can be differentiated based on their various leaf shapes [3]. At first glance, it might look the same. However, if you look at the shape and character of each leaf, grapes have different types and leaf variants.

In recent years, various methods of classifying plant leaves have been proposed [4 – 9]. Of some methods to classify plant leaves, most of the research applies deep learning methods, which can help classify objects effectively, but none of the studies have discussed the classification of grapes through the leaves. Deep learning methods have been widely used for research, such as image detection and classification [10 – 15]. Convolutional neural networks (CNN) are one of the most common and widely used models of deep learning and

* Corresponding author: pnurulmarifah@gmail.com

have good performance because most CNN architectures have superior capabilities in learning the features of an object [16]. The CNN method continues to develop until the newest method proposed by Shaoqing Ren et al., namely Faster R-CNN [17].

This research uses a deep learning method with the Faster R-CNN ResNet-50 algorithm and uses pre-trained COCO weights to classify grape varieties through leaf images. The classification of grape leaves used in this research consists of five types of grape leaves: academic, jupiter, local, taldun, and transfiguration. It is hoped that this research can help grape farmers classify grape plant types so that the management of these plants can be optimal.

2 Research methods

This section discusses the stages used in this research. Figure 1 depicts the stages of the research carried out. The tools used in this research are smart phones and laptops, where smartphones are used for collecting datasets while laptops are used to process datasets as well as analyze the results of the research. The material used in this research is a google collaboratory that is used to train and test datasets that have been processed.

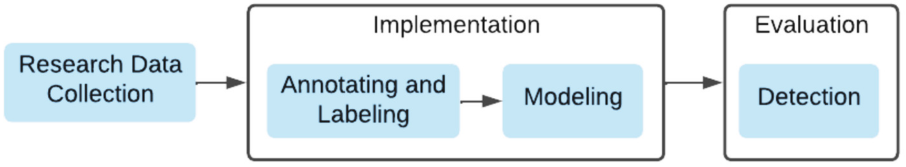

Fig. 1. Research methods

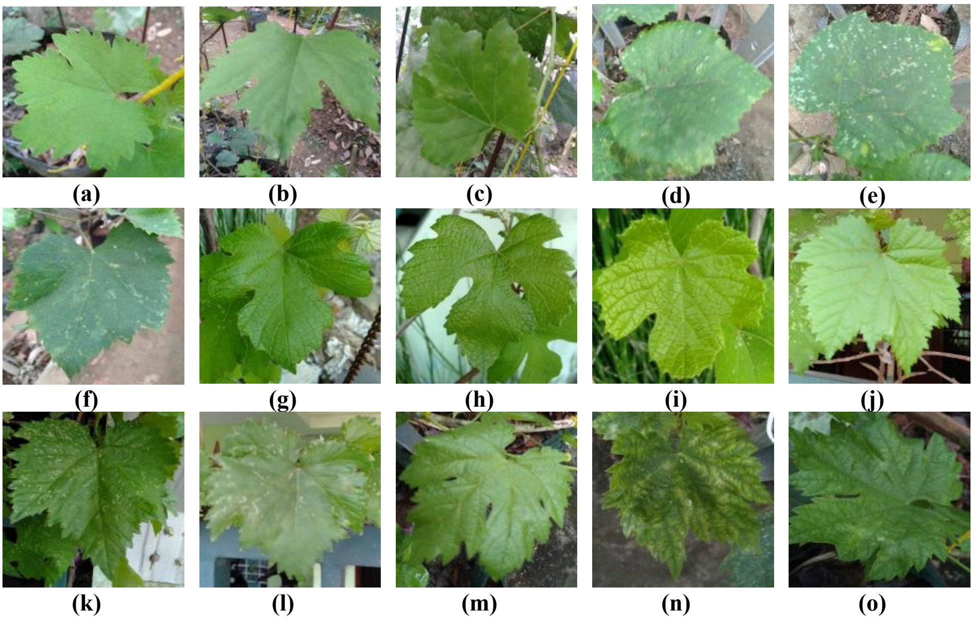

Fig. 2. Datasets. Academic : a, b, c; Jupiter: d, e, f; Local: g, h, i; Taldun: j, k, l; Transfiguration: m, n, o

2.1 Research data collection

Research data collection is a process of collecting images of vine leaves that will be used for data sets in this research. The number of images collected is 500, of which 100 are of each type of leaf variety of vineyards. The amount of pictures collected represents the accuracy of the researchers in order to produce good results at the time of classification. There are five varieties of grapes used in this study, namely academic grapes, Jupiter, local, taldun, and transfiguration. The image used is a picture of the leaves of the vineyard in the garden of one of our researchers. So the process of taking pictures is actually done outdoors with taking pictures from 11.00 - 13.00 WIB (UTC +7). When taking pictures, no light and saturation adjustments are performed.

After collecting data on grape plant leaf images, the image resolution is then adjusted to the provisions of the pre-trained weight that we use, namely the COCO pre-trained method. COCO's pre-trained weight is 640 x 640. Examples of images used for the dataset are shown in Figure 2, which represent each class. Class here refers to the type of grape variety used.

2.2 Annotation and labeling

Annotation and labeling is the process of marking the grape leaves in an image. This process begins by making a bounding box on the leaf object of the grape plant and then labeling it 'academic', 'jupiter', 'local', 'taldun', and 'transfiguration'. The result of image annotation and labeling is a file with the extension '.xml' in PASCAL VOC format. The annotation and labeling process functions to determine the ground truth, which is used to calculate the loss regression of bounding box detection locations on objects detected during the training process. The annotation and labeling processes are carried out using LabelImage software. Figure 3 shows an example of annotation and labeling.

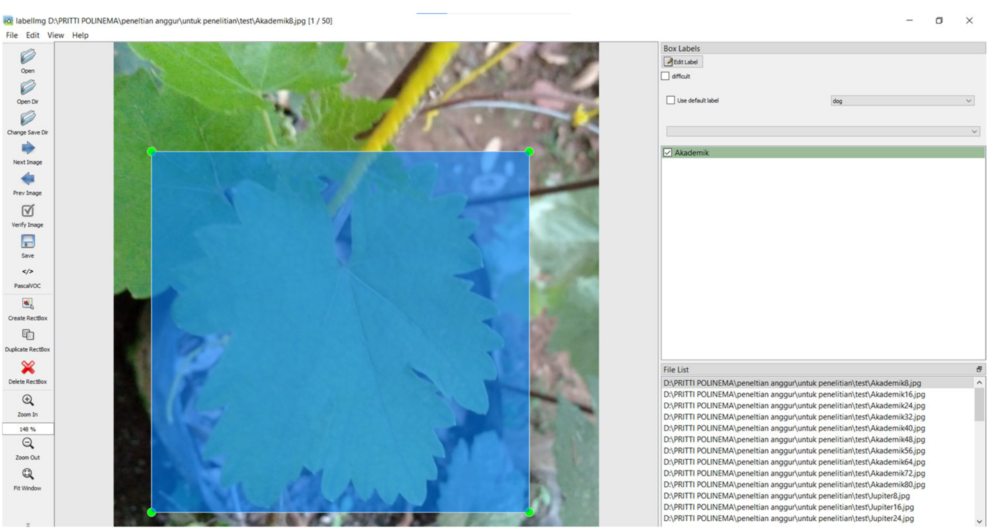

Fig. 3. Labeling Process with LabelImg software

The results of image annotation and labeling cannot be directly used for the training process. Therefore, the next stage is converting the resulting annotation and labeling files with the extension format '.xml' into '.csv' format so that they can be used in the training process and during the modeling process. Table 1 shows the results of converting the file into '.csv' format.

Table 1. XML to CSV File Conversion Results

File name	Wide	Tall	Class	Xmin	Ymin	Xmax	Ymax
Academic8.jpg	640	640	academic	50	144	548	620
Jupiter16.jpg	640	640	jupiter	82	1	616	590
Local24.jpg	640	640	local	1	6	595	522
Taldun32.jpg	640	640	taldun	17	1	567	534
Transfiguration72.jpg	640	640	transfiguration	118	80	640	558

2.3 Modeling

The modeling process was carried out using the Faster R-CNN algorithm with the ResNet-50 architecture. The modeling process was carried out using Google Collaboratory. The Faster R-CNN ResNet-50 architecture model can be seen in Figure 4.

Faster R-CNN is a CNN method that has developed from region-based CNN (R-CNN) and Fast R-CNN. The novelty of the previous method is replacing selective search with a region proposal network (RPN) [18 – 22]. The Faster R-CNN architectural model begins by inputting an image, which is then carried out by extracting grape plant leaf features using a backbone CNN.

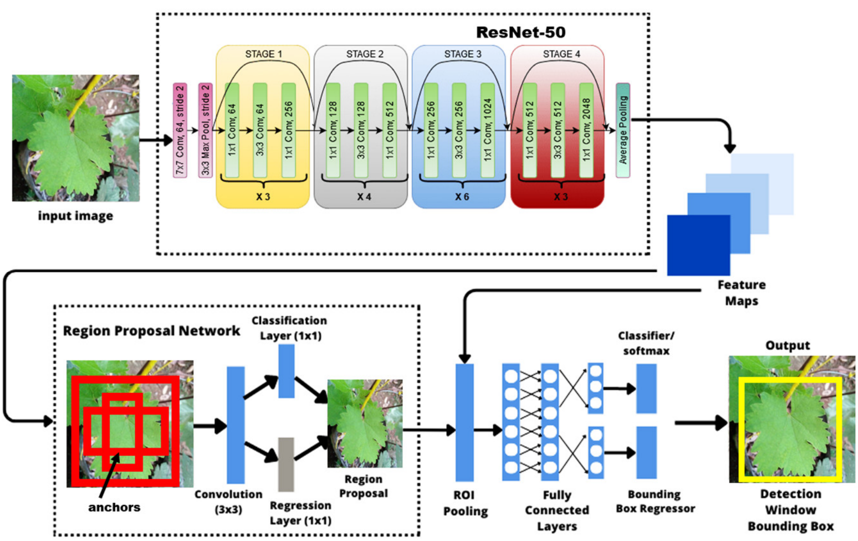

Fig. 4. Illustration of the Faster R-CNN ResNet-50 architecture model

A convolutional neural network (CNN) is an artificial neural network that uses a collection of convolutional layers for feature extraction [23 , 24] during the training phase to produce a layer of concern for the final result [25] by knowing the characteristics of objects in the images being trained. It has been proven to be used by groups for large image classification data, which is often called ImageNet, COCO, Computer Vision, and Natural Language Processing (NLP).

The backbone CNN used in this research is ResNet-50. Residual Network (ResNet) [26] is a deep-CNN architecture that has a total of 50 convolution layers and was the first deep-

CNN architecture to utilize residual learning in 2015. ResNet-50 has succeeded in improving accuracy in computer vision benchmarking challenges [27].

The next section is the region proposal network (RPN), which requires future map output from the backbone network. The RPN process is carried out by placing a set of "Anchors" on the input image for each location on the output feature map. These anchors show possible objects in various sizes and ratios. For PASCAL VOC, the anchors used have 3 scale box sizes (128^2, 256^2, 512^2) and 3 aspect ratios (1:1, 1:2, and 2:1), so there are 9 possible anchors placed on the image input to the output feature map [28 , 29]. The output in this process is used to provide the probability of whether each point on the backbone feature map contains objects in the nine anchors at that point or not [29].

The third part is an ROI polling layer that uses a maximum polling operation to collect features from the feature map and change their dimensions so that they have a fixed size. In the fully connected layer, the output of the ROI polling layer is organized into a one-dimensional feature vector as the input of the fully connected layer. In the final part, after passing it through the fully connected layers, the features are fed into the classification and regression branches that predict the correct match of the object. Thus producing images with objects marked by bounding boxes and possible classification results [29].

2.4 Measurement

The grape plant leaf classification system was tested using a measurement method that obtained true positive (TP), false positive (FP), and false negative (FN) parameters. True positive (TP) means that the bounding box successfully detected grape plant leaves correctly; false positive (FP) indicates that the bounding box can detect objects but does not identify grape plant leaves; and false negative (FN) indicates that the bounding box does not contain objects in the given figure. These parameters are used to calculate precision, recall, and the F1 score (Eq. 1-3). Precision is the level of accuracy of the resulting detection, while recall is the level of detection success. The F1 score results from a balance between precision and recall. The following is the formula for calculating precision, recall, and F1 score [30].

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

3 Results and discussion

3.1 Test result

We assessed the faster R-CNN algorithm with the ResNet-50 network architecture with a data set consisting of 500 images of grape plant leaves. The F1 score for object recognition was 90.11% at step 8000; this is the highest value obtained during 4 trials with different numbers of training steps. Meanwhile, the lowest F1 score value was produced at step 3000, namely 82.35%. Table 2 shows the increase in accuracy value for each step increase in grape plant leaf object recognition, although the increase obtained is not much or very significant.

Table 2. Test Result

Step	Precision (%)	Recall (%)	F1 Score (%)	TP	FP	FN
3000	87.50	77.78	82.35	35	5	10
4000	85.37	79.55	82.35	35	6	9
5000	87.80	80.00	83.72	36	5	9
8000	95.35	85.42	90.11	41	2	7

3.2 Loss during modelling

In the process of modeling grape leaf varieties, the loss function is an important result. Each step of the modeling process produces its own weaknesses. So we can combine these three weaknesses to see the difference in performance of the modeling process, as shown in Figure 5. We can see that the highest loss is produced at step 5000, with a loss of 0.247958. Step 8000 has the lowest loss, namely 0.247958.

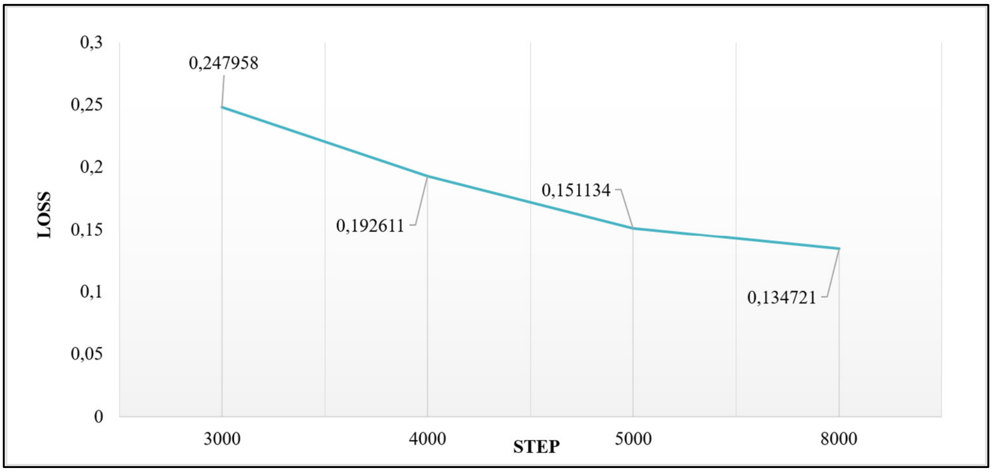

Fig. 5. Loss of Faster R-CNN ResNet-50 in the modeling process

3.3 Analysis of the results

A sample of the accuracy of this research is shown in Figure 6. From the test image shown in Figure 6, there is a detection accuracy that has reached 100%, which means the system has succeeded in learning the model and finally detecting and classifying it accurately and appropriately. The process thus producing like Figure 6 has been explained in Figure 4.

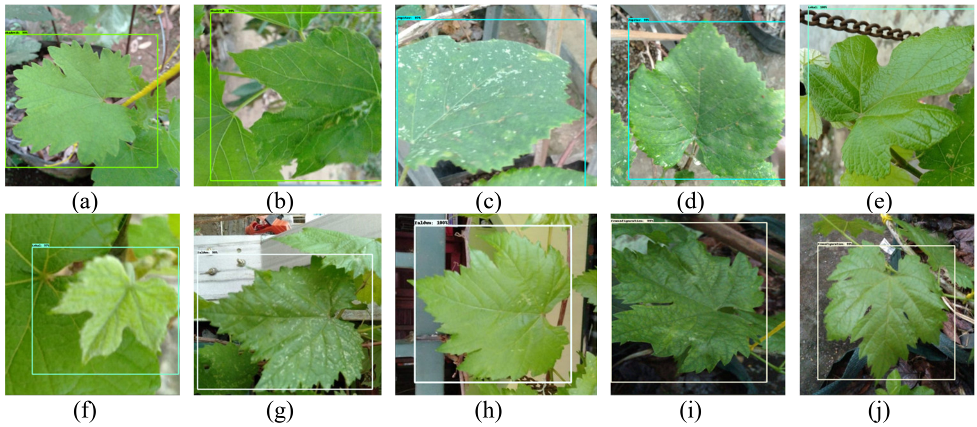

Fig. 6. Detection Results. Academic: a, b; Jupiter: c, d; Local: e, f; Taldun: g, h; Transfiguration: i, j.

4 Conclusion

This paper analyzes the grape leaf detection method using the Faster R-CNN algorithm with the ResNet-50 network architecture, which is used as a pre-training network to extract Faster R-CNN features. The experimental results show that the network used can effectively improve the efficiency of network operations. The highest recognition accuracy of grape leaves reached 90.11% in 8,000 test steps with a loss of 0.134721. The results of the evaluation of the grape leaf detection system using Faster R-CNN show that the smaller the total loss in the training process, the better the system is at learning object features.

The researchers would like to express gratitude towards the Directorate General of Research Enhancement and Development, Ministry of Education, Culture, Research, and Technology, for the grant that was granted under contract number: 12882/PL2.1/HK/2023.

References

1. T. Nagata, A. Hosaka-Sasaki, and S. Kikuchi, The Evolutionary Diversification of Genes that Encode Transcription Factor Proteins in Plants. Elsevier Inc. (2016). http://dx.doi.org/10.1016/B978-0-12-800854-6/00005-1.
2. G. Hasanaliyeva et al., Effects of production region, production systems and grape type/variety on nutritional quality parameters of table grapes; results from a UK retail survey, Foods, **9**, 12 (2020). doi: 10.3390/foods9121874.
3. Y. Khan et al., Antioxidant potential in the leaves of grape varieties (Vitis vinifera L.) grown in different soil compositions, Arab. J. Chem., **14**, 11, 103412 (2021), https://doi.org/10.1016/j.arabjc.2021.103412.
4. C. P. Lee, K. M. Lim, Y. X. Song, and A. Alqahtani, Plant-CNN-ViT: Plant Classification with Ensemble of Convolutional Neural Networks and Vision Transformer, Plants, **12**, 14, 1–21 (2023). doi: 10.3390/plants12142642.
5. M. Faisal, J. S. Leu, and J. T. Darmawan, Model Selection of Hybrid Feature Fusion for Coffee Leaf Disease Classification, IEEE Access, **11**, June 62281–62291 (2023), doi: 10.1109/ACCESS.2023.3286935.
6. B. Dwi Mardiana, W. Budi Utomo, U. Nur Oktaviana, G. Wasis Wicaksono, and A. Eko Minarno, Herbal Leaves Classification Based on Leaf Image Using CNN Architecture Model VGG16, J. RESTI (Rekayasa Sist. dan Teknol. Informasi), **7**, 1, 20–26 (2023). doi: 10.29207/resti.v7i1.4550.

7. E. Elfatimi, R. Eryigit, and L. Elfatimi, Beans Leaf Diseases Classification Using MobileNet Models, IEEE Access, **10**, 9471–9482 (2022). doi: 10.1109/ACCESS.2022.3142817.
8. K. Yang, W. Zhong, and F. Li, Leaf segmentation and classification with a complicated background using deep learning, Agronomy, **10**, 11 (2020). doi: 10.3390/agronomy10111721.
9. Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, Plant leaf disease classification using EfficientNet deep learning model, Ecol. Inform., **61**, 101182 (2021). https://doi.org/10.1016/j.ecoinf.2020.101182.
10. S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace, and U. Bagci, Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches, IEEE Trans. Med. Imaging, **38**, 8, 1777–1787 (2019). doi: 10.1109/TMI.2019.2894349.
11. S. Roy et al., Deep Learning for Classification and Localization of COVID-19 Markers in Point-of-Care Lung Ultrasound, IEEE Trans. Med. Imaging, **39**, 8, 2676–2687 (2020). doi: 10.1109/TMI.2020.2994459.
12. M. Sarosa, N. Muna, and E. Rohadi, Detection of natural disaster victims using You Only Look Once (YOLO), IOP Conf. Ser. Mater. Sci. Eng., **1098**, 3, 032076 (2021), doi: 10.1088/1757-899x/1098/3/032076.
13. A. Alem and S. Kumar, Deep Learning Models Performance Evaluations for Remote Sensed Image Classification, IEEE Access, **10**, October (2022). doi: 10.1109/ACCESS.2022.3215264.
14. M. Sarosa, N. Muna, E. Rohadi, M. Junus, A. Setiawan, and I. Muhiqqin, Circle Hough Transform (CHT) method applied to coin identifier and recognition based on computer vision and android developer, J. Phys. Conf. Ser., **1402**, 7 (2019). doi: 10.1088/1742-6596/1402/7/077036.
15. M. Sarosa, N. Muna, and E. Rohadi, Performance of faster r-cnn to detect plastic waste, Int. J. Adv. Trends Comput. Sci. Eng., **9**, 5, 7756–7762 (2020). doi: 10.30534/ijatcse/2020/120952020.
16. H. Xu, X. Lv, X. Wang, Z. Ren, N. Bodla, and R. Chellappa, Deep Regionlets: Blended Representation and Deep Learning for Generic Object Detection, IEEE Trans. Pattern Anal. Mach. Intell., **43**, 6, 1914–1927 (2021). doi: 10.1109/TPAMI.2019.2957780.
17. J. Chen, J. Mao, C. Thiel, and Y. Wang, iWaste: Video-Based Medical Waste Detection and Classification, 2020 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 5794–5797 (2020). doi: 10.1109/EMBC44109.2020.9175645.
18. B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, and K. Ouni, Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3, 2019 1st Int. Conf. Unmanned Veh. Syst. UVS 2019, 1, 1–6 (2019). doi: 10.1109/UVS.2019.8658300.
19. K. bing Chen, Y. Xuan, A. jun Lin, and S. hua Guo, Esophageal cancer detection based on classification of gastrointestinal CT images using improved Faster RCNN, Comput. Methods Programs Biomed., **207**, 106172 (2021). https://doi.org/10.1016/j.cmpb.2021.106172.
20. H. Yuan, Y. Shao, Z. Liu, and H. Wang, An Improved Faster R-CNN for Pulmonary Embolism Detection from CTPA Images, IEEE Access, **9**, 105382–105392 (2021), doi: 10.1109/ACCESS.2021.3099479.
21. Z. Li et al., A high-precision detection method of hydroponic lettuce seedlings status based on improved Faster RCNN, Comput. Electron. Agric., **182**, February (2021), doi: 10.1016/j.compag.2021.106054.
22. V. Sharma and R. N. Mir, Saliency guided faster-RCNN (SGFr-RCNN) model for object detection and recognition, J. King Saud Univ. - Comput. Inf. Sci., **34**, 5, 1687–

1699 (2022). https://doi.org/10.1016/j.jksuci.2019.09.012.
23. Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects, IEEE Trans. neural networks Learn. Syst., **33**, 12, 6999–7019 (2022). doi: doi: 10.1109/TNNLS.2021.3084827.
24. L. Fan, T. Zhang, and W. Du, Optical-flow-based framework to boost video object detection performance with object enhancement, Expert Syst. Appl., **170**, December 114544 (2021). https://doi.org/10.1016/j.eswa.2020.114544.
25. A. Ajit, K. Acharya, and A. Samanta, A Review of Convolutional Neural Networks, Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020, 1–5 (2020). doi: 10.1109/ic-ETITE47903.2020.049.
26. A. S. B. Reddy and D. S. Juliet, Transfer learning with RESNET-50 for malaria cell-image classification, Proc. 2019 IEEE Int. Conf. Commun. Signal Process. ICCSP 2019, 945–949 (2019). doi: 10.1109/ICCSP.2019.8697909.
27. M. Sankupellay and D. Konovalov, Bird call recognition using deep convolutional neural network, ResNet-50, Aust. Acoust. Soc. Annu. Conf. AAS 2018, October 200–207 (2019). doi: 10.13140/RG.2.2.31865.31847.
28. Y. P. Chen, Y. Li, and G. Wang, An enhanced region proposal network for object detection using deep learning method, PLoS One, **13**, 9, 1–26 (2018). doi: 10.1371/journal.pone.0203897.
29. W. Gu et al., High Accuracy Thyroid Tumor Image Recognition Based on Hybrid Multiple Models Optimization, IEEE Access, **8**, 128426–128439 (2020). doi: 10.1109/ACCESS.2020.3008401.
30. N. A. Prasetyo, Pranowo, and A. J. Santoso, Automatic detection and calculation of palm oil fresh fruit bunches using faster R-CNN, Int. J. Appl. Sci. Eng., **17**, 2, 121–134 (2020). doi: 10.6703/IJASE.202005.