8th IEEE International Conference and Workshops on Recent Advances and Innovations in Engineering- ICRAIE 2023
(IEEE Record #59459)

Fruit Detection and Recognition Using Faster R-CNN with FPN30 Pre-trained Network

Muhammad Izzat Roslan
College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA, Shah Alam
40450, Selangor, Malaysia
Muhdizzatroslan94@gmail.com

Zaidah Ibrahim
College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA, Shah Alam
40450, Selangor, Malaysia
zaida782@uitm.edu.my

*Nur Aina Khadijah Adnan
College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA, Shah Alam
40450, Selangor, Malaysia
ainakhadijah@uitm.edu.my

Norizan Mat Diah
College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA, Shah Alam
40450, Selangor, Malaysia
norizan289@uitm.edu.my.

Nur Azima Alya Narawi
College of Computing, Informatics, and Mathematics
Universiti Teknologi MARA, Shah Alam
40450, Selangor, Malaysia
azimaalya@uitm.edu.my

Yunifa Miftachul Arif
Informatics Engineering, Faculty of Science and Technology Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia
yunif4@ti.uin-malang.ac.id

***Abstract*— Accurate and reliable fruit detection and recognition in orchards is critical for enabling higher-level agriculture tasks such as fruit picking. However, detecting and recognizing fruits with occlusion by neighboring fruits is extremely difficult. Faster R-CNN (Faster Region-based Convolutional Neural Network) is a well-known deep learning technology for object detection and recognition. Thus, this study investigates the application of Faster R-CNN for apple detection and recognition. Two different datasets have been constructed under variable illumination conditions and occlusion; an inter-class dataset that consists of images of apples and oranges, and an intra-class dataset that consists of images of two types of apples, namely fuji and royal gala apples. Results indicate that Faster R-CNN can detect and recognize apples from oranges, and the fuji apple in the orchards, with high accuracy. This suggests that Faster R-CNN can be used practically in the real orchard context.**

Keywords—artificial intelligence, fruit detection, Faster R-CNN, feature pyramid network, FPN, pre-trained network

I. Introduction

Image recognition, also known as computer vision, is an artificial intelligence (AI) domain focused on enabling machines to interpret and understand visual data. This technology empowers computers to recognize patterns, objects, and actions within images or videos, akin to how humans categorize and comprehend visual stimuli [1], [2].

The applications of image recognition are diverse and far-reaching. From self-driving cars identifying obstacles and pedestrians [3] to food classification applications [4], its potential impact is profound. With the rise of smartphones and social media, people frequently share various types of images that they encounter, making it possible to create large datasets for training artificial intelligence models. And one of the objects is fruit. By leveraging AI techniques and deep learning algorithms, fruit recognition systems can automatically analyze images and accurately identify the type of fruit. Convolutional neural networks (CNNs) are widely used for image recognition tasks due to their ability to learn by extracting distinctive patterns and features from images [1].

An automatic fruit recognition system using computer vision is a challenging task due to the similarities in appearance between fruits. Research in inter-class fruit recognition, classifying different types of fruits using deep learning, is important for precision agriculture. The decisions from trained humans to inspect the quality of fruit by seeing can be inconsistent, and biased [5][6]. This research involves the recognition of inter-class fruit images. There is not much research conducted on intra-class fruit detection and recognition.

After the enormous achievements of CNN [2] in classification, region-based CNN (R-CNN) [7] [8] achieves remarkable results in object detection. Object detection frameworks combine both recognition and localization into one system to detect and draw boxes around objects in images. Girshick et al. propose the R-CNN, which consists of two parts. A series of regions are generated by selective search in the first section [9]. All regions are then warped to a fixed size and sent to a convolutional network for feature extraction. Then, all extracted features are fed into the support vector machine (SVM) for recognition, and a regression layer is used to refine the position bounding boxes simultaneously.

Another popular method for image detection and recognition is a Faster R-CNN (regional convolutional neural network) with an FPN (feature pyramid network) pre-trained model [10]. The Faster R-CNN architecture is a two-stage object detection system that utilizes a region proposal network (RPN) to generate object proposals and a separate network for recognizing the objects within those proposals. The FPN architecture is a feature pyramid network that extracts features at multiple scales, allowing for better detection of objects of various sizes. By using a pre-trained model, the model can leverage knowledge from a large dataset and achieve high accuracy on new image recognition tasks.

2023 IEEE 8th International Conference on Recent Advances and Innovations in Engineering (ICRAIE) | 979-8-3503-1551-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICRAIE59459.2023.10468169

979-8-3503-1551-6/23/$31.00 ©2023 IEEE

The Faster R-CNN [11], the latest generation of region-based deep learning object detection model, has achieved good detection results in many public datasets [12], [13]. With VOC2007 and 2012 training sets for training, the mAP (mean Average Precision) of the VOC2007 test set reached 73.2% and the mAP of the VOC2012 test set reached 70.4%, allowing the object detection speed to reach 5 frames per second. Fig. 1 depicts the network architecture of R-CNN, which consists mostly of two primary modules; the VGG16 network's convolutional network module, which extracts regional characteristics, makes up the first module, and the Fast R-CNN detector makes up the second [14]. Before sending the feature maps of the final layer to the Region Proposal Networks (RPN), it first extracts features from the images using the VGG16 network [10] that has been trained on ImageNet [12]. RPN, the most significant module of Faster R-CNN, employs a sliding window to apply 3 * 3 convolution kernels to feature maps to create nine proposal regions with three sizes (128 x 2, 256 x 2, 512 x 2) and three aspect ratios (11:1, 1:2, and 21:1). The proposal regions then sent to the ROI Pooling layer, which extracts the proposal features from the regions; finally, input the pooled layer's proposal region features to the fully connected layer, which contains the classification and regression subnets. The classification sub-network is used to categorize the target, and the regression sub-network is used to change the position of the target frame to achieve a more accurate target category, area score, and coordinate position of the area in the indicated region of the image [16].

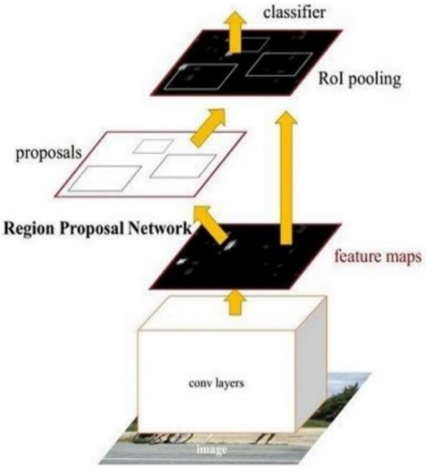

Fig. 1. Architecture of R-CNN [7]

Faster R-CNN is a type of convolutional neural network (CNN) architecture for object detection. It consists of three main components: a deep CNN feature extractor, a region proposal network (RPN), and a classifier. The deep CNN feature extractor, typically a ResNet or VGG network, is used to extract features from an input image [16][17]. These features are then fed into the RPN, which generates a set of region proposals (i.e., potential object locations) in the image. The classifier then takes these region proposals, classifies each one as an object or not, and assigns a class label to each object. One of the key features of Faster R-CNN is that it uses a region proposal network (RPN) to generate region proposals rather than using a separate module, as in the original R-CNN architecture. This allows for a more efficient and faster detection process. Overall, Faster R-CNN is a two-stage object detection algorithm that uses a CNN to extract features from an input image and a separate RPN to generate region proposals. The classifier then uses these region proposals to classify and locate objects in the image. The architecture of the Faster R-CNN model is shown in Fig.2.

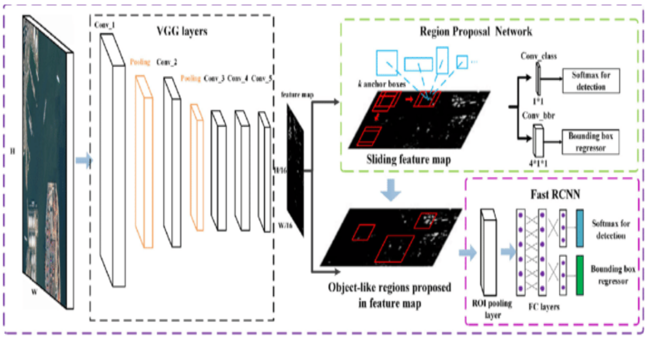

Fig. 2. Architecture of Faster R-CNN Model [14]

II. Methodology

The task aims to experiment with Deep Learning Algorithms for image recognition and classification using inter-class and intra-class datasets on an Apple. The resource for the dataset was taken from the website [18]. The dataset was used to implement image recognition using the Faster R-CNN Deep Learning Model and evaluate its performance in image recognition. In addition, the performance results of the algorithms are then analyzed to understand their application. The research method consists of constructing datasets, annotating the datasets, constructing models, and evaluating the models.

A. *Constructing Datasets*

In this study, a dataset consisting of two types of images; intra-class and inter-class, was constructed [19]. Intra-class images, depict variations of the same apple type, in this study, royal gala and fuji apples; while inter-class images, depict different types of fruits, in this study, apples and oranges. The dataset was created to investigate the ability of the image detection and recognition algorithm (Faster R-CNN) to distinguish between two different types of fruits, namely apples and oranges; and variations within the same apple, namely royal gala and fuji.

i. Datasets of Inter-Class for Apples and Oranges

When constructing a dataset for deep learning and data analysis, it is important to consider the diversity and distribution of the samples within the dataset. One common example is the distinction between different classes within the dataset, such as apples and oranges as in Fig. 3 and Fig. 4. The Inter-Class dataset is a prime example of this concept in action. This dataset consists of a collection of samples that have been labeled and classified as either apples or oranges. The training dataset consists of 125 images of apples and 125 images of oranges while the testing dataset consists of 12 mages of apples and 12 images of oranges. One of the key advantages of the Inter-Class dataset is that it provides a wide range of samples for apples and oranges. This allows for a more comprehensive understanding of the characteristics and features of each class, which can be used to improve the performance of deep learning models. Additionally, the Inter-Class dataset provides a good opportunity to evaluate the performance of different deep-learning algorithms on different datasets.

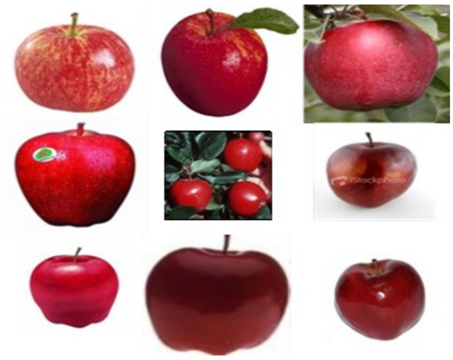

Fig. 3. Sample image of apples in Inter-Class Datasets

Fig. 4. Sample images of oranges in Inter-Class Datasets

ii. *Datasets of Intra-Class for Royal Gala Apples and Fuji Apples*

A dataset containing multiple examples of the same class or category is known as an Intra-Class dataset. In this case, a dataset of Gala and Fuji apples will be constructed as in Fig. 5 and Fig. 6. Many images of Gala and Fuji apples will be gathered, which can be done by taking photographs in a controlled environment such as a studio or laboratory or using existing images from the internet. The images must be high quality and accurately represent the apples in question. Once sufficient images are gathered, they will be labeled as either Gala or Fuji. This can be done manually by reviewing each image and assigning a label or through automated image classification algorithms. The labels must be accurate and consistent to ensure the reliability and usefulness of the dataset. The training dataset consists of 194 images of fuji apples and 96 images of gala apples while the testing dataset consists of 10 images of gala apples and 21 images of fuji apples.

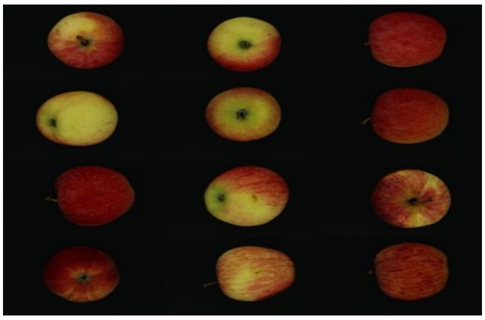

Fig. 5. Sample images of a royal gala apple in Intra-Class Datasets

Fig. 6. Sample images of fuji Apples in intra-Class Datasets

B. *Annotating the Datasets*

Annotating datasets is an important step in training deep-learning models [20]. It involves adding labels or tags to the data, which provide information about the content of the data and allow the model to learn how to classify or identify the data. One tool that can be used to annotate datasets is Roboflow. Roboflow is an annotation platform designed to simplify and streamline the annotation process. It allows users to upload their datasets, create annotation projects, and assign tasks to annotators. The platform also provides various tools and features to aid in the annotation process, including annotation templates, quality control tools, and data export capabilities. One of the key advantages of using Roboflow for annotation is its ability to automate many of the tedious and time-consuming tasks associated with annotation. For example, the platform can automatically resize images and convert them to the appropriate format, which saves time and reduces the risk of errors. Additionally, it can automatically assign tasks to annotators and provide real-time feedback on the progress of the annotation project as in Fig. 7 and Fig. 8.

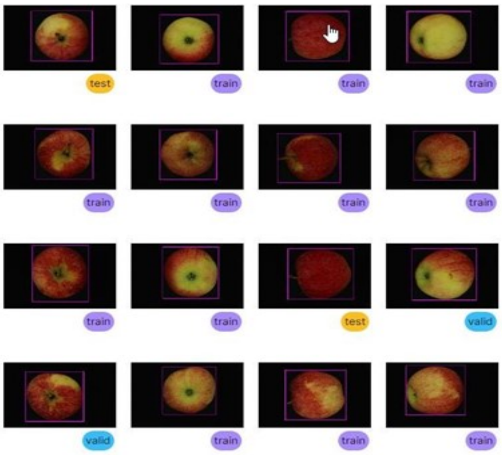

Fig. 7. Annotating the Datasets using Roboflow for Royal Gala Apple

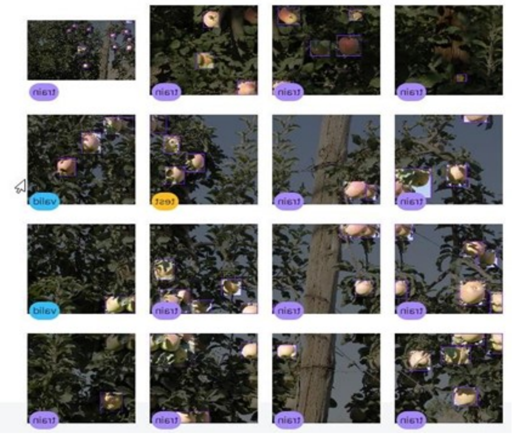

Fig. 8. Annotating the Datasets using Roboflow for Fuji Apple

C. *Construct Model and Evaluate Model*

The effectiveness of the use of a constructed Faster R-CNN model with an FPN30 pre-trained network for training datasets was demonstrated in this study. A popular object detection method, the Faster R-CNN algorithm, was utilized in a constructed form with an FPN30 pre-trained network on a dataset of images to evaluate its performance in training datasets. High values of precision, recall, and F1-score were achieved, indicating that the objects in the images were effectively detected and classified.

Additionally, it was observed that the training process converged in a relatively short number of iterations, indicating that the provided dataset was effectively learned from. These results suggest that the use of a constructed Faster R- CNN model with an FPN30 pre-trained network is an effective approach for training datasets in object detection and recognition tasks.

III. Experiment Results and Discussion

i. Training Datasets using Constructed Faster R-CNN Model with FPN30 Pre-trained Network (Inter-Class Datasets).

The evaluation criteria applies is as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} \quad APi$$

$$F - measure = \frac{2\ x\ recall\ x\ precision}{precision + recall}$$

where
TP is for the correctly recognized true output;
FP is for the incorrectly recognized true output;
FN is for the incorrectly recognized false output.

Fig. 9 illustrates the results of training a Faster R-CNN model on Inter-Class Datasets. The graph shows the class accuracy of the model. Upon examination of the graph, it is clear that the results of training indicate good performance. The class accuracy, which measures the percentage of correctly classified objects in the test set, is high. This indicates that the model is able to accurately identify a large number of objects within the images it is presented. A high class accuracy is a crucial aspect of object detection, as it ensures that the model is able to accurately identify objects of interest within an image.

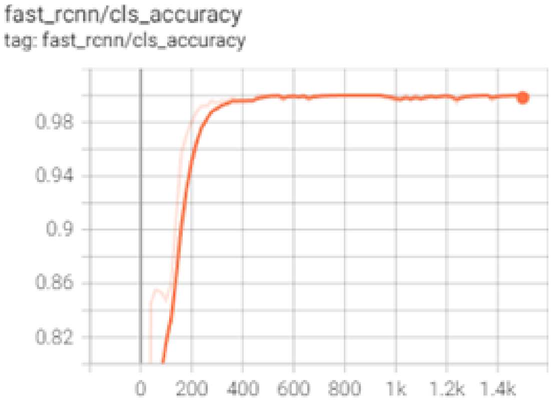

Fig. 9. Shows the class accuracy for Inter-class classification.

ii. Training Datasets using Constructed Faster R-CNN Model with FPN30 Pre-trained Network (Intra-Class Datasets)

Similarly, based on Fig. 12 shows the results of training the Faster R-CNN model with FPN30 pre-trained network using intra-class datasets, as shown in the graph in the paper's subsection "Training Datasets using Constructed Faster R-CNN Model with FPN30 Pre-trained Network (Intra-Class Datasets)", indicate good performance. The graph demonstrates high class accuracy, low false negative rate, and strong foreground class accuracy, suggesting that the model is able to accurately identify a large number of objects within images and able to identify objects in the foreground of the image. This makes the model well-suited for object detection tasks, object tracking, and scene understanding using intra-class datasets.

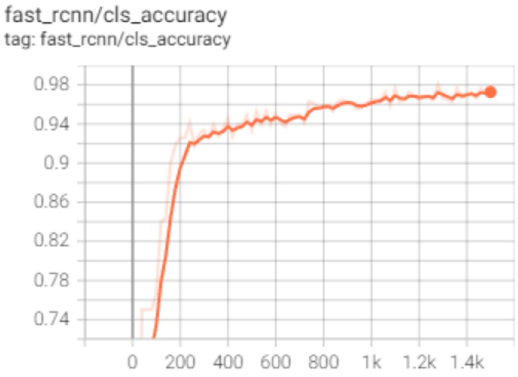

Fig. 10. Shows the class accuracy for Intra-class classification.

B. *Faster R-CNN Performance Metrics (Inter-Class)*

Based on Table I, the evaluation results for the bounding box (inter-class) show strong performance across all metrics, including average precision (AP), AP at 50, 75, small, medium, and large object scales (AP50, AP75, APs, APm, APl). The AP score indicates a high level of accuracy in detecting objects within the class categories, and the APs, APm, and APl scores demonstrate the model's ability to accurately detect objects of various sizes. Additionally, the AP50 and AP75 scores indicate that the model is able to accurately detect objects at different intersections over union (IoU) thresholds. Overall, these results demonstrate that the bounding box (inter-class) model is a robust and reliable solution for object detection tasks.

Based on Table II, the results of the per-category detect and identify apples within the image. However, the "Orange" category has a lower score of 82.885, indicating that the model may not perform as well in detecting and identifying oranges. Overall, these results demonstrate that the model has a strong performance in detecting certain object classes, but may have difficulty with others.

TABLE I. EVALUATION RESULT FOR BBOX (INTER-CLASS)

AP	AP50	AP75	APs	APm	APl
91.443	91.443	91.443	N/A	N/A	91.443

TABLE II. PER-CATEGORY BBOX AP (INTER-CLASS)

Category	AP
Apple	100.00
Orange	82.885

C. Faster R-CNN Performance metrics (Intra-Class)

Based on Table III, shows the evaluation results for the bounding box (intra-class) showing an average precision (AP) score of 57.486, indicating that the model has moderate performance in detecting objects within the same class. The AP50 score of 82.757 and the AP75 score of 59.861 suggest that the model is able to accurately detect objects at different intersections over union (IoU) thresholds. The APS score is not available, which suggests that the model is not able to accurately detect small objects. This suggests that the model may have difficulty detecting objects of various sizes within the same class. Overall, these results indicate that while the model has moderate performance in detecting objects within the same class, there is room for improvement in detecting objects of different sizes.

Based on Table IV, it shows the results of the per-category bounding box average precision (AP) for intra-class evaluation shows a high level of accuracy for the "Fuji Apple" category, with a score of 100.00. This suggests that the model is able to accurately detect and identify Fuji Apples within the image. However, the "Royal Gala Apple" category has a lower score of 82.885, indicating that the model may not perform as well in detecting and identifying Royal Gala Apples. This suggests that the model is able to detect and identify certain sub-categories of apples well but may have difficulty with others. Overall, these results demonstrate that the model has a strong performance in detecting certain object sub-categories, but may have difficulty with others.

TABLE III. EVALUATION RESULT FOR BBOX (INTRA-CLASS)

AP	AP50	AP75	APs	APm	APl
57.486	82.757	59.861	N/A	27.370	60.791

TABLE IV. PER-CATEGORY BBOX AP (INTRA-CLASS)

Category	AP
Fuji Apple	100.00
Royal Gala Apple	82.885

D. Image Recognition Test Using Faster R-CNN (Inter-Class)

The Faster R-CNN model can accurately detect and locate various classes of apples within an image, as demonstrated by the bonding boxes that were successfully drawn around the objects in Fig. 13.

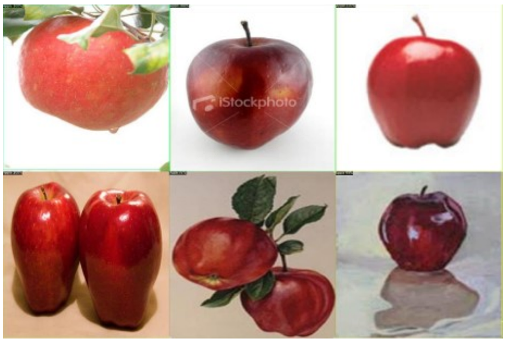

Fig. 11. Shows Detection Test (Inter-Class)

E. Image Recognition Test Using Faster R-CNN (Intra-Class)

Similarly, the Faster R-CNN model is able to accurately detect and locate various classes for fuji apple and royal gala apple within an image, as demonstrated by the bonding boxes that were successfully drawn around the objects in Fig 12.

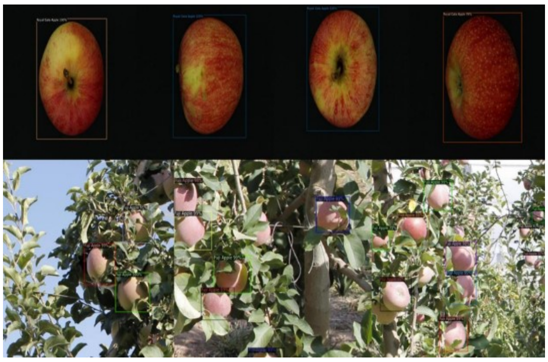

Fig. 12. Shows Detection Test (Intra-Class)

IV. CONCLUSION

In precision agriculture, it is important to apply deep learning solutions to detect and recognize the correct and good quality of fruits. This study investigates the use of deep learning method, namely, Faster R-CNN, in agriculture sectors. The objective of the experiments was successfully achieved through the construction of two datasets containing apples and oranges for inter-class recognition and Royal Gala apples and Fuji apples for intra-class recognition. The Faster R-CNN model with FPN 30 pre-trained network was also successfully constructed and its performance was evaluated in image recognition tasks. The results of the experiments indicate that the Faster R-CNN model was able to accurately detect and recognize different classes of apples and oranges within an image, demonstrating its effectiveness in object recognition tasks. Overall, this study highlights the potential of using the Faster R-CNN model with a pre-trained network for image recognition tasks in the real orchard environment.

ACKNOWLEDGMENT

The authors would like to thank the School of Computing Sciences, College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, for sponsoring this research.

REFERENCES

[1] S. Shakya, "Analysis of Artificial Intelligence based Image Classification Techniques," Journal of Innovative Image Processing, vol. 2, no. 1, pp. 44–54, Mar. 2020, doi: 10.36548/JIIP.2020.1.005.

[2] Mandal M, "CNN for Deep Learning | Convolutional Neural Networks," 2021. https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/ (accessed Aug. 23, 2023).

[3] Gupta, A., Anpalagan, A., Guan, L., & Khwaja, A. S. (2021). Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. https://doi.org/10.1016/j.array.2021.100057.

[4] Zhang, Y., Deng, L., Zhu, H., Wang, W., Ren, Z., Zhou, Q., Lu, S., Sun, S., Zhu, Z., Gorriz, J. M., & Wang, S. (2023). Deep learning in food category recognition. https://doi.org/10.1016/j.inffus.2023.101859

[5] Nguyen, H. H. Cuong; Luong, A. T.; Tribh, T. H.; Ho, P. H.; Meesad, P. and Nguyen, T. T., (2021). Intelligent Fruit Recognition System Using Deep Learning. IC2IT 2021, LNNS 251, pp. 13–22, 2021. https://doi.org/10.1007/978-3-030-79757-7_2

[6] Albarrak, K.; Gulzar, Y.; Hamid, Y.; Mehmood, A. and Soomro, A. B., 2022]. A Deep Learning-Based Model for Date Fruit Classification. Sustainability 2022, 14, 6339.https://doi.org/10.3390/su14106339

[7] R. Girshick, J. Donahue, T. Darrell, J. Malik, U. C. Berkeley, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, p. 5000, Sep. 2014, doi: 10.1109/CVPR.2014.81.

[8] Gandhi R, "R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms | by Rohith Gandhi | Towards Data Science," 2018. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e (accessed Aug. 23, 2023).

[9] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective Search for Object Recognition," Int J Comput Vis, vol. 104, no. 2, pp. 154–171, Sep. 2013, doi: 10.1007/S11263-013-0620-5.

[10] Y. Youssef and M. Elshenawy, "Automatic vehicle counting and tracking in aerial video feeds using cascade region-based convolutional neural networks and feature pyramid networks," Transp Res Rec, vol. 2675, no. 8, pp. 304–317, Mar. 2021, doi: 10.1177/0361198121997833/ASSET/IMAGES/LARGE/10.1177_0361198121997833-FIG7.JPEG.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans Pattern Anal Mach Intell, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[12] S. Pha, H. Zheng, Y. Sun, L. Hao, F. Jiang, and W. Li, "Analysis of Object Detection Performance Based on Faster R-CNN, J Phys Conf Ser, vol. 1827, p. 12085, 2021, doi: 10.1088/1742-6596/1827/1/012085.

[13] L. Ma, F. Zhang, and L. Xu, "Fruit detection using faster R-CNN based on deep network," Smart Innovation, Systems, and Technologies, vol. 128, pp. 193–199, 2019, doi: 10.1007/978-3-030-04585-2_23.

[14] Y. Zhang, Y. Chen, C. Huang, and M. Gao. "Object Detection Network Based on Feature Fusion and Attention Mechanism". Future Internet. 2019;11(1):9. https://doi.org/10.3390/fi11101000

[15] S. Ma, Y. Song, N. Cheng, Y. Hao, Z. Chen, and X. Fu, "IOP Conference Series: Earth and Environmental Science Structured Light Detection Algorithm based on Deep Learning" 2019, doi: 10.1088/1755-1315/252/4/042050.

[16] Z. Deng, H. Sun, S. Zhou, J. Zhao, L. Lei, and H. Zou, (2018). Multi-scale object detection in remote sensing imagery with convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing. 145. 10.1016/j.isprsjprs.2018.04.003.

[17] P. Simon and V. Uma, "Deep Learning based Feature Extraction for Texture Classification," Procedia Comput Sci, vol. 171, pp. 1680–1687, Jan. 2020, doi: 10.1016/J.PROCS.2020.04.180.

[18] B. Ashwath, "Apple2orange Dataset", 2020, https://www.kaggle.com/datasets/balraj98/apple2orange-dataset (accessed Mar. 23, 2023).

[19] A. Venkataramanan, M. Laviale, C. Figus, P. Usseglio-Polatera, and C. Pradalier, "Tackling Inter-Class Similarity and Intra-Class Variance for Microscopic Image-based Classification," Sep. 2021, Accessed: Nov. 08, 2023. [Online]. Available: http://arxiv.org/abs/2109.11891

[20] T. Hylander and M. Alvenkrona, "Semi-Automatic Image Annotation Tool," 2023, Accessed: Nov. 08, 2023. [Online]. Available: http://liu.diva-portal.org/smash/get/diva2:1791647/FULLTEXT01.pdf