Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/48/20241260

Application and evaluation of deep learning based image recognition techniques in agriculture

Ningning Wu

Glasgow College, University of Electronic Science and Technology of China, Chengdu, 611731, China

2021190501016@std.uestc.edu.cn

Abstract. Agriculture, as an important industry in society, is facing problems such as an aging population and rural labor exodus, which leads to rising labor costs and uncertainty in agricultural production. Deep learning techniques are considered as a key tool to solve this problem. In this paper, three popular deep learning algorithms, namely, Region-based Convolutional Neural Network, You Only Look Once, and Single Shot MultiBox Detector, are introduced and their working principles are described in detail, while the advantages and disadvantages of these algorithms are briefly analyzed. Additionally, this paper specifically analyzes the application of these three algorithms in three agricultural scenarios, such as timber species recognition, fruit picking, and pest identification. The results show that although the three algorithms are slightly different in terms of accuracy and detection speed, they all demonstrate the potential for a wide range of applications in the agricultural field. Therefore, deep learning technology is of great significance in solving the problem of rural labor shortage, especially when combined with advanced equipment, which is expected to significantly improve the efficiency of identification, monitoring, and harvesting in agriculture and promote the development of automated agriculture.

Keywords: Deep Learning Algorithms, Automated Agriculture, Seed Recognition, Pest Recognition, Fruit Harvesting.

1. Introduction

Agriculture has always been the backbone of human society; however, with the rapid growth of urbanization and population aging, the supply of labor in rural areas is gradually decreasing and labor costs are rising. This trend may bring great uncertainty to agricultural production and adversely affect the sustainability of rural communities. In the context of this challenge, deep learning techniques have emerged as a key tool to address the rural labor shortage [1,2].

The aim of this paper is to investigate three currently popular deep learning algorithms, namely Region-based Convolutional Neural Network, You Only Look Once, and Single Shot MultiBox Detector, to address the challenges in the field of agriculture such as timber species identification, fruit picking, and pest identification [3, 4]. Region-based Convolutional Neural Network is a classical two-stage target detection algorithm, which is implemented through convolutional Neural Networks and Support Vector Machines to extract features from images and to classify and regress candidate regions [5, 6]. In contrast, You Only Look Once adopts a single-stage target detection method, which can

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

classify and localize objects with one forward propagation, and is therefore also known as a "region-free" method [7, 8]. Single Shot MultiBox Detector is also a single-stage target detection algorithm, which simultaneously detects and localizes multiple objects in an image by applying anchor frames to the multi-scale feature map [8, 9]. Each of these three algorithms has unique strengths and weaknesses, and each has its own strengths in agricultural applications.

In summary, deep learning is important in solving the problem of rural labor shortage. Using deep learning technology, combined with advanced equipment, it is possible to automate agricultural tasks, including harvesting and monitoring, which improves the efficiency of agricultural production and contributes to the modernization and sustainable development of agriculture.

2. Analysis of the theoretical foundations of the algorithm

2.1. R-CNN Algorithms

Region-based Convolutional Neural Network is a classic two-stage model for target detection, representing the forefront of deep learning applications in the field of object detection. The fundamental architecture of Region-based Convolutional Neural Network relies on various algorithms, including Convolutional Neural Networks, Linear Regression, and Support Vector Machine, to accomplish target detection. This is achieved by conducting Convolutional Neural Networks-based feature extraction and subsequent classification and regression for each candidate region [3].

The working principle of the Region-based Convolutional Neural Network algorithm can be summarized as follows: One of the core innovations in Region-based Convolutional Neural Network is the introduction of the Region Proposal Network for selective searching within the input image. The Region Proposal Network employs a sliding window technique and anchor boxes to generate candidate regions that potentially contain target objects. Typically, Region-based Convolutional Neural Network also applies a pre-trained Convolutional Neural Network model, such as AlexNet or VGGNet, to extract features from each of the candidate regions. The extracted features from the candidate regions are then fed into a Support Vector Machine classifier to create a classification model. Additionally, Region-based Convolutional Neural Network combines the bounding boxes of each region as examples to train a linear regression model. This regression model helps in predicting the ground-truth bounding boxes for training purposes. To enhance the accuracy of detection results, Region-based Convolutional Neural Network typically employs Non-Maximal Suppression. This technique eliminates overlapping bounding boxes, ensuring the selection of the final detection outcomes [9]. The schematic diagram of Region-based Convolutional Neural Network is shown in Figure 1.

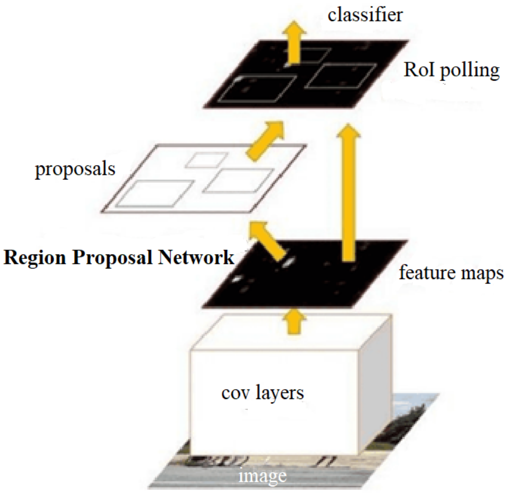

Figure 1. Schematic diagram of R-CNN [9].

In summary, Region-based Convolutional Neural Network is a powerful target detection approach that achieves precise detection and localization of objects in images. It accomplishes this by integrating various techniques, including region proposal networks, Convolutional Neural Network-based feature extraction, Support Vector Machine-based classification, and linear regression [9].

2.2. YOLO Algorithms

You Only Look Once is a state-of-the-art target detection algorithm known for its unique single-stage detection approach. Compared to traditional two-stage detection methods (e.g., Region-based Convolutional Neural Network), You Only Look Once is unique in that it can simultaneously localize objects and classify categories in a single forward pass, and is therefore also known as a "region-free" method [10].

You Only Look Once works as follows: First, it splits the input image into fixed-size grid cells that will be used to detect objects. Then, within each grid cell, You Only Look Once simultaneously predicts multiple bounding boxes (typically 5), each consisting of a set of coordinate values (center coordinates x, y, width, and height) and a score indicating the confidence level. Additionally, the algorithm needs to predict for each bounding box the classes of objects it may contain and their probability scores. Typically, You Only Look Once supports multi-category detection with an associated score for each category. To generate the final detection results, You Only Look Once also applies the Non-Maximal Suppression technique. This technique is used to reduce overlapping bounding boxes and ensure that each object is detected only once. Finally, You Only Look Once combines the confidence of each bounding box with the probability of the object category it belongs to, so that the final detection result of the object can be produced [10]. The schematic diagram of You Only Look Once is shown in Figure 2.

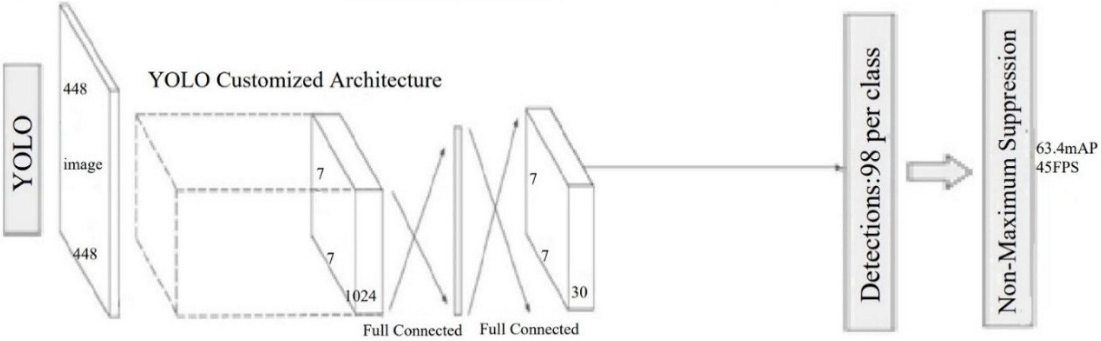

Figure 2. Schematic diagram of YOLO [10].

This single-stage object detection approach empowers You Only Look Once with an admirable equilibrium between swiftness and precision, consequently rendering it highly effective in practical real-world applications.

2.3. SSD Algorithm

Single Shot MultiBox Detector is also a deep learning algorithm for target detection. Similar to the You Only Look Once algorithm, Single Shot MultiBox Detector also belongs to the single-stage target detector, which is capable of detecting and localizing multiple objects in an image simultaneously by applying anchor frames, performing classification and regression tasks on a multi-scale feature map [9]. Its working principle is as follows: First, the Single Shot MultiBox Detector processes the input image using a convolutional neural network and generates multi-scale feature maps in different convolutional layers. These feature maps contain information at different resolutions, enabling Single Shot MultiBox Detector to detect objects of different sizes. Second, on the cell of each feature map, the Single Shot MultiBox Detector defines a set of anchor frames. These anchor frames have different widths and heights to cover different sizes and aspect ratios of various objects. Each anchor box is associated with one or more predefined object categories. For each anchor frame, Single Shot MultiBox Detector performs two main tasks: categorization and regression. The categorization task aims to determine whether each anchor frame contains an object and to which category the object belongs. The regression task, on the other hand, is used to accurately localize the bounding boxes of the objects, i.e., to precisely adjust the anchor boxes to better fit the objects. To train the model, Single Shot MultiBox Detector also uses a combination of loss functions, including a classification loss and a regression loss, to ensure correct classification and accurate localization of objects. Finally, the Single Shot MultiBox Detector algorithm also uses Non-Maximal Suppression to filter the detection results [9]. The schematic diagram of Single Shot MultiBox Detector is shown in Figure 3.

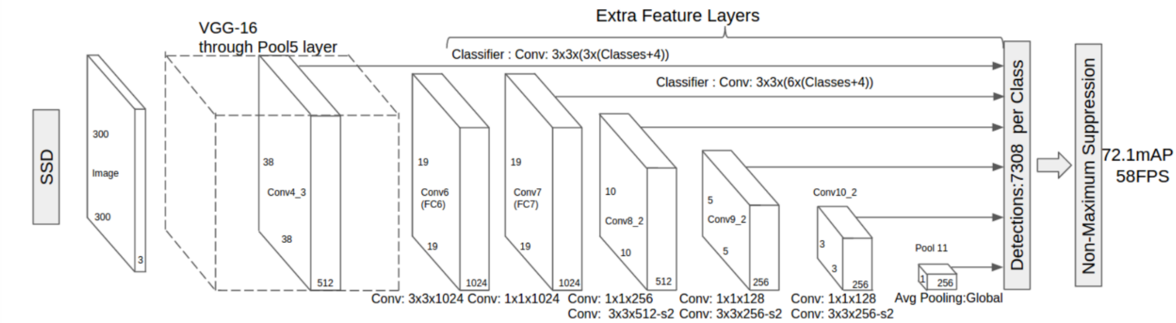

Figure 3. Schematic diagram of SSD [9].

This comprehensive approach allows Single Shot MultiBox Detector to achieve efficient object detection and localization in a single stage.

2.4. Comparison of three algorithms

All the above three algorithms represent the main direction of the current deep learning image recognition technology. It is worth noting that all three algorithms employ Non-Maximal Suppression techniques to eliminate redundant bounding boxes so that the accuracy of the detection results can be improved.

In terms of prediction speed, the Region-based Convolutional Neural Network algorithm employs a two-stage approach involving three different models (Convolutional Neural Network, Support Vector Machine, and regression model) for target detection [9]. These multi-stage processes result in longer image prediction times. In contrast, the Single Shot MultiBox Detector and You Only Look Once algorithms use a one-stage approach, where the main idea is to densely sample the image at various locations and then go through a single Convolutional Neural Network feature extraction, classification, and regression stage. This one-stage design allows them to have faster detection speed [6, 10].

However, in terms of accuracy, the Region-based Convolutional Neural Network algorithm achieves extremely high detection accuracy through its two-stage approach. In contrast, although the Single Shot MultiBox Detector and You Only Look Once algorithms have faster detection speeds, their accuracy is relatively low [10]. In particular, the You Only Look Once algorithm may suffer from the disadvantages of difficulty in detecting small targets and inaccurate position localization [10]. Single Shot MultiBox Detector, however, partially overcomes these drawbacks by employing a priori frames with different scales and aspect ratios, which improves the detection accuracy for multi-scale targets [3].

3. Scenario analysis of applications in agriculture

3.1. Wood Identification Technology

Wood recognition technology is widely used in several fields, including wood species identification, wood furniture inspection, and wood structure ancient building protection [4]. Traditional wood recognition methods are mainly divided into two categories: one is recognition technology based on macroscopic features, which is convenient for sampling but less accurate; the other is recognition technology based on microscopic anatomical features, which is more accurate but more time-consuming. However, traditional image segmentation techniques are less efficient as they face difficult and complex computational processes in extracting wood features. Contrarily, image recognition techniques are based on deep learning, which can autonomously learn the features of wood, thus achieving more accurate and efficient wood detection.

Tao Wang et al. addressed the problem of species recognition of broadleaf timber [4]. First, using CT technology to obtain multiple micro images of broadleaf timber. Subsequently, features in these images were labeled according to category using LabelIm software. The dimensions and other relevant information of each microscopic image were also recorded, and then these datasets were converted into a training set. Then, four different types of broadleaf timber conduit images were classified and recognized using two deep learning frameworks, You Only Look Once v3 and Single Shot MultiBox Detector. By calculating four identification correctness metrics such as precision, recall, average precision, and average precision mean, as well as the detection time of each framework, Tao Wang et al. derived the detection precision and speed of the two frameworks. the average precision mean of You Only Look Once v3 was 0.9157 and the average detection time was 1.185 s, while the average precision mean of Single Shot MultiBox Detector was 0.9117 and the average detection time was 0.608 s [4].

By analyzing the above data, it can be concluded that both algorithmic frameworks are able to achieve efficient and accurate wood species recognition. You Only Look Once v3 algorithmic framework has a relatively higher recognition accuracy due to its use of residual networks, inverse convolution, and multi-feature layer ideas. The Single Shot MultiBox Detector algorithmic framework, however, is able to perform object classification and regression of prediction frames concurrently, thus its time spent is significantly shorter. Taken together, Wang Tao et al. concluded that the Single Shot MultiBox Detector algorithm can process the samples more quickly under the premise of guaranteeing the correct recognition rate, thus realizing a more efficient automatic recognition of the four broadleaf timber species.

3.2. Fruit Visual Inspection Technology

Fruit harvesting in most areas of China is based on manual picking, but problems such as population aging and rural labor outflow have led to rising rural labor costs. The emergence of intelligent picking robots provides an effective solution to this problem, and fruit visual detection technology has become a crucial part of fruit-picking robots. Deep learning-based target detection algorithms can efficiently detect various types of fruits in natural environments or complex orchard contexts [1, 3, 6].

For fruits like apples and citrus, they have relatively homogeneous features, and the colors and sizes of fruits at maturity are basically the same. Therefore, the algorithms for fruit visual detection techniques focus on reducing the leakage problem caused by occlusion or strong exposure. Tongyang Huang et al. proposed a citrus fruit recognition method based on the You Only Look Once v5 improved model [1]. The model introduces a Convolutional Block Attention Module to enhance the feature extraction capability of the network, which effectively improves the omission detection problem of occluded and small targets [1]. In addition, they used the α-IoU loss function instead of the GIoU loss function as the bounding box regression loss function, thus improving the accuracy of bounding box localization. The experimental results show that the average precision (mAP) of the improved model can reach 0.916 in the natural environment, and the detection time of a single citrus fruit image is only 16.7 ms. The improved You Only Look Once v5 algorithm can actually realize the efficient recognition of citrus fruits in the natural environment [1].

However, for fruits like blueberries, the long span of the ripening period leads to differences in ripeness, size, and color of the whole cluster. Xu Zhu proposed a way to solve this problem with an improved Faster Region-based Convolutional Neural Network algorithm [6]. They built a recognition model for processing blueberry images with different maturity levels, extracted the target features of blueberry fruits, and calibrated them. The Faster Region-based Convolutional Neural Network algorithm used an alternative training approach for four categories of images, namely, unripe fruits, semi-ripe fruits, ripe fruits, and backgrounds, and calculated their recognition accuracies. The results of the study show that the Faster Region-based Convolutional Neural Network algorithm recognizes all three types of blueberries with an accuracy of 93% or more, with an average recognition accuracy of 94.05%. This provides technical support for the automated picking of berry fruits [6].

3.3. Plant Pest and Disease Identification Techniques

Plant pests and diseases have a significant impact on crop yield and quality, so the use of pesticides has become a common method of suppressing pests and diseases. However, excessive use of pesticides can lead to serious economic losses and environmental problems [2, 5, 8]. For a long time, the identification of pests has mainly relied on agricultural professionals and experienced farmers to observe the morphological characteristics of pests. This manual identification method suffers from high subjectivity, low efficiency, and lagging response. Therefore, there is an urgent need for an objective and efficient pest detection method for more rational planning of pesticide use [8].

As Yong He put forward the idea that there are imperfections in the current publicly available pest databases, he proposed to create an oilseed rape pest imaging database containing 12 typical oilseed rape pests and to select the Single Shot MultiBox Detector w/Inception model, which is characterized by multiscale and high-precision features, for performing efficient pest identification [5]. During the design process, he also raised the use of data augmentation (DA) to improve the generalization performance of the model to address the lack of publicly available oilseed rape pest datasets. Furthermore, he added a dropout layer to cope with the overfitting problem of neural networks, thus reducing the difficulty of model training. Experimental results show that the improved Single Shot MultiBox Detector w/Inception model achieves excellent performance with an average accuracy of 0.7714 and an average detection time of 0.052 s, which is significantly faster than the other architectures [5]. Yong He believes that this model, with its improved adaptability to different environments, responsiveness, and accuracy, can be used for drones and the Internet of Things in pest monitoring tasks.

Qingru Chen [8] also mentioned that when optimizing the best path for UAVs to spray pesticides, the key is to identify pests in orchards, and the You Only Look Once v3 neural network model plays an important role in this regard. He conducted an experiment using an embedded computer, the Jetson TX2, to compare the performance of two different models, You Only Look Once v3 and Tiny-You Only Look Once v3, in terms of speed and accuracy in recognizing pests such as T. papillosa. In addition, experiments were conducted to augment the T. papillosa database with data to improve the learning ability of the models. The experimental results show that You Only Look Once v3 has a mAP of 0.93, while Tiny-You Only Look Once v3 has a mAP of 0.89, which is a similar performance. However, the recognition speed of the Tiny-You Only Look Once v3 model is more than three times faster than that of the You Only Look Once v3 model, and thus it is more suitable for the real-time pest recognition module of UAVs, which helps to better plan the pesticide spraying path of agricultural UAVs [8].

4. Conclusion

Each of the three algorithms for deep learning has its own strengths and weaknesses. The Region-based Convolutional Neural Network algorithm utilizes a two-stage approach and achieves superior detection accuracy, however, its detection speed is relatively low. By contrast, the You Only Look Once and Single Shot MultiBox Detector algorithms are faster but perform lower in terms of accuracy. In particular, Single Shot MultiBox Detector overcomes the problem of detecting small targets and inaccurate location localization by employing a priori frames with different scales and aspect ratios. All three deep learning algorithms show good potential for application in agriculture. With the support of these deep learning techniques and devices, the agricultural field can realize the prediction of pests, automatic fruit picking, and efficient wood recognition, thus promoting the development of automated agriculture.

Despite the great potential of deep learning in agricultural automation, specific challenges remain. Deep learning models require large amounts of labeled data for training; however, in agriculture, acquiring and labeling large-scale agricultural data is often expensive and time-consuming. Additionally, agricultural data often exhibits category imbalance, such as uneven distribution of positive and negative samples in pest detection, which may lead to models that are biased toward predicting the majority of categories. In the future, the development of transfer learning and data augmentation techniques is expected to effectively address these challenges and provide better support for deep learning, thus further improving agricultural productivity, monitoring and management, and contributing to sustainable agricultural development.

References

[1] Huang, T., Huang, H., Li, Z., et al. Citrus Fruit Recognition Method Based on Improved YOLOv5 Model. Journal of Huazhong Agricultural University, 2022, 41(4): 170-177.

[2] Liu J, Wang X. Plant diseases and pests detection based on deep learning: a review. Plant Methods, 2021, 17: 1-18.

[3] Li, W., Wang, D., Ning, Z., et al. A Review of Fruit Object Detection Algorithms in Computer Vision. Computer and Modernization, 2022, (06): 87.

[4] Wang, T., Yang, X., Gao, Y., et al. Material Identification of Four Broadleaf Tree Species Based on Deep Learning. Computer Era, 2022.

[5] He Y, Zeng H, Fan Y, et al. Application of deep learning in integrated pest management: A real-time system for detection and diagnosis of oilseed rape pests. Mobile Information Systems, 2019.

[6] Zhu, X., Ma, H., Ji, J., et al. Detection and Recognition Analysis of Blueberry Canopy Fruit Based on Faster R-CNN. Journal of Southern Agriculture, 2020, 51(6): 1493-1501.

[7] Li, S., Wang, P., Lv, Z., et al. A Novel Agricultural System Based on Jetson Nano and Deep Learning. Southern Agricultural Machinery, 2022, 53(20): 13-15.

[8] Chen C J, Huang Y Y, Li Y S, et al. Identification of fruit tree pests with deep learning on embedded drone to achieve accurate pesticide spraying. IEEE Access, 2021, 9: 21986-21997.

[9] Zhang, W. Research on Object Detection Algorithms and Applications Based on Deep Learning. Doctoral dissertation, Jiangnan University, 2023.

[10] Li, M., Wu, C., Bao, Y., et al. On the Application Principles of the YOLO Algorithm in Machine Vision. Education Modernization, 2018, 5(41): 174-176.