Research on the Algorithm of YOLO Visual Image Classification System for Pests and Diseases

Li Yuanqi
Faculty of Innovation Engineering, Macau University of Science and Technology, Computer Science Major, Macau, 999078, China MUST
1220029051@student.must.edu.mo

Abstract

During the internship, it was found that artificial intelligence algorithms have problems such as low accuracy and low efficiency in the visual graphic classification of agricultural pests and diseases. These issues not only affect the healthy development of the industry but also make it difficult to meet the development needs of large-scale and precise agriculture. To solve these practical problems, this research topic was proposed in combination with the professional knowledge learned ——the Algorithm of YOLO Visual Image Classification System for Pests and Diseases. This article conducts a comprehensive review of the YOLO series algorithms using the literature review method. Firstly, conduct a comprehensive review of the current research status of the YOLO series of algorithms and the visual image classification of pests and diseases, and the research objectives were clarified(Conduct an in-depth analysis of the technical features and performance, Summarize the existing problems and challenges, Summarize research trends and development directions, Build a systematic review framework).Then, the basic principles, network structures and improved versions (such as YOLOv3, YOLOv4, etc.) of the YOLO series algorithms are discussed. The advantages and disadvantages of different versions of the algorithms and their performance in image classification tasks are compared. The shortcomings and areas for improvement of the existing research are summarized.

Keywords: algorithm; Computer Vision; Object Detection; Pest and Disease Identification; Smart Agriculture YOLO (key words)

1. Introduction (Significance of the subject)

The output of agricultural products is an important foundation for the steady development of a country. The issue of agricultural product safety is related to people's lives and health as well as social stability and public security and is highly valued by the state. Crop diseases and pests are one of the key factors that restrict agricultural production efficiency and threaten food security. Traditional pest and disease identification relies on manual experience, which has problems such as low efficiency, strong subjectivity and limited coverage, making it difficult to meet the development needs of large-scale and precise agriculture. The automatic identification technology of pests and diseases based on computer vision, especially the visual image classification system combined with the YOLO (You Only Look Once) algorithm, has become an important breakthrough to solve this problem.

1.1 theoretical significance

1.1.1 Enrich the theory of object detection algorithms in agricultural scenarios

Agricultural pest and disease images have unique attributes such as a high proportion of small targets, subtle category differences, and complex background interferences. The feature expression ability of the existing YOLO algorithm in this scenario is limited. This study can enhance the adaptability of the algorithm to complex agricultural scenarios by improving the feature fusion module of the algorithm (such as introducing an attention mechanism to increase the weights of key features) and optimizing the loss function (such as designing a dynamic weight strategy for category imbalance problems), providing a theoretical supplement for the application of object detection algorithms in unstructured scenarios.

1.1.2 Promote the integration of interdisciplinary theories

The research should integrate the morphology intelligence antics and occurrence patterns of pests and diseases in plant protection with the deep learning theory in computer vision to construct a mapping relationship of "agricultural biological features - image features - algorithm models", providing a new research perspective for the theoretical system of the

intersection of agriculture and artificial intelligence, and promoting the integration and innovation of cross-disciplinary knowledge.

1.2 practical significance

1.2.1 Improve the efficiency and accuracy of agricultural pest and disease control

The optimized YOLO classification system can be installed on devices such as drones and field cameras to achieve real-time identification and classification of pests and diseases, solving the problem of lag in manual identification. Farmers can apply pesticides in a targeted manner based on the identification results, reducing the number of pesticides used, lowering production costs while minimizing ecological pollution, and promoting the development of green agriculture.

1.2.2 Support the construction of a smart agriculture system

As the core perception link of smart agriculture, this system can be linked with the Internet of Things and big data platforms to build an integrated pest and disease management system of "monitoring - early warning - decision-making".

2. The Technical Evolution and Core Principle of the YOLO Series Algorithms

2.1 The Development Trajectory of the YOLO Series Algorithms

The YOLO algorithm was initially proposed by Redmon et al. in 2016 and has undergone multiple generations of iterations, achieving continuous breakthroughs in detection accuracy and speed. Its development can be divided into three stages:

Initial Exploration Stage (YOLOv1-YOLOv2):

YOLOv1 first introduced the concept of "single-stage detection," dividing an image into an S×S grid where each grid predicts B bounding boxes and class probabilities. It achieved a detection speed of 45 FPS but suffered from lower accuracy in small object detection. YOLOv2 improved mAP to 76.8% and increased speed to 67 FPS by incorporating batch normalization, an anchor box mechanism, and a multi-scale training strategy. It also supported real-time detection for images of different resolutions, laying the foundation for dynamic monitoring in agricultural scenarios.

Performance Enhancement Stage (YOLOv3-YOLOv5):

YOLOv3 adopted multi-scale feature fusion and a residual network structure to enhance small object feature extraction, while introducing a Softmax classifier to optimize class prediction. Building on YOLOv3, YOLOv4 integrated the CSPDarknet53 backbone network, an SPP module, and a SAM attention mechanism to further improve detection robustness in complex scenarios. YOLOv5 achieved lightweight design and enhanced generalization through adaptive anchor boxes and Mosaic data augmentation. Its smallest version, YOLOv5s, reached a detection speed of 140 FPS, making it suitable for deployment on edge devices.

Refinement and Optimization Stage (YOLOv6-YOLOv8):

YOLOv6 was designed for industrial applications, balancing accuracy and speed using a RepVGG backbone network and a decoupled head structure. YOLOv7 introduced an ELAN network structure and Mosaic augmentation strategy, achieving a mAP of 56.8% on the COCO dataset. As the latest representative framework in the YOLO ecosystem,

YOLOv8—developed independently by Ultralytics—integrates the strengths of previous YOLO iterations while standing as a distinct advancement rather than an official continuation of the series originally initiated by Redmon/Jocher. It supports multiple tasks including object detection, segmentation, and classification, and also offers lightweight model variants, thereby providing more flexible options for the multi-dimensional recognition of agricultural pests and diseases.

2.2 Core principle of the YOLO algorithm

The core logic of the YOLO algorithm is "end-to-end" object detection. Its workflow primarily consists of three stages:

Image Preprocessing: The input image is resized to a fixed dimension (e.g., 416×416) and normalized to mitigate the impact of pixel value variations on model training. Some improved versions employ data augmentation techniques such as Mosaic and CutMix to enhance dataset diversity and improve the model’s generalization capability.

Feature Extraction and Prediction：The backbone network (e.g., Darknet, CSPDarknet) extracts image features, which are then fused through a neck network (e.g., FPN, PANet) for multi-scale feature integration. Finally, the head network outputs the target bounding box coordinates, confidence scores, and class probabilities. Taking YOLOv5 as an example, it uses

CSPDarknet53 to extract deep features and integrates multi-level features via an FPN-PAN structure, enabling comprehensive detection of large, medium, and small objects.

Postprocessing: The Non-Maximum Suppression (NMS) algorithm is applied to filter redundant bounding boxes, retaining only the predictions with the highest confidence scores. The final output includes the location and category information of the detected objects.

3. Current research status at home and abroad

3.1 the latest domestic research achievements and technological advancements

Considerable progress has also been made in the research of algorithms for visual image classification systems of pests and diseases based on YOLO in China. Many universities and research institutions have actively carried out relevant research and achieved outstanding results in both algorithm optimization and practical application.

The team led by Pan Wei from Hefei University has developed a detection method based on YOLOv7 and a general pruning process of the YOLO model to address the issues of complex algorithms and large computational load in agricultural pest and disease detection. This enables automated monitoring of agricultural product pests and diseases, allowing for early detection of problems and the implementation of control measures, thereby reducing crop losses. The experimental results show that after pruning, the computational load and parameter count of the model have decreased by 62% and 76% respectively, while the average accuracy and reasoning speed have increased by 3.9% and 40.4% respectively. The research has significant practical reference value for enhancing the productivity of new products such as intelligent detection of agricultural pest and disease problems.[1] The team led by Tian Bin from Suzhou Vocational College of Agriculture enhanced the feature extraction ability of the backbone network by introducing the attention mechanism SENet(Squeeze - and - Excitation Networks) on the basis of the PP-YOLO network, thereby improving the recognition accuracy of small target tomato leaf spot disease. Experiments have shown that the detection performance of the method proposed in this paper on the tomato leaf spot disease dataset in Plant Village is superior to that of PP-YOLO, with an average accuracy of 86.44%, and it has certain application value.[2] Professor Ji-Yuan Ding's team from Kyungnam National University proposed the DM-YOLO model, which is an enhanced version of the YOLOv8 framework and aims to improve the detection accuracy of cucumber diseases. Traditional detection models have difficulty identifying small-scale and overlapping symptoms, especially when key features are obscured by changes in lighting, occlusion and background noise. The proposed DM-YOLO model combines three innovative modules and collectively improves the detection performance. Compared with the baseline YOLOv8, DM-YOLO improved the average precision (mAP50) by 1.2%p on the custom dataset and by 3.2%p on the PlantDoc dataset. These results highlight the enhanced ability of the model to detect small-scale and overlapping disease symptoms, demonstrating its effectiveness and robustness in various agricultural monitoring environments. Compared with the original algorithm, the improved model has made significant progress and has strong competitiveness compared with other advanced object detection models.[3] A team of professors from Central South University has introduced their enhanced version, pest-yolo. The improved pest-YOLO integrates two key advancements: an effective channel attention (ECA) mechanism for enhanced feature extraction and a transformer encoder for capturing global features. The original extrusion incentive attention mechanism was replaced by the ECA mechanism, effectively improving the model's ability to extract basic features from pest images. In addition, the transformer encoder is introduced into the convolutional neural network (CNN) architecture to enhance its ability to capture global context information. To further enhance the expression ability of small target features such as agricultural pests, a feature fusion method based on cross-stage feature fusion (CSFF) is proposed. This method significantly improves the representation of small targets in the feature fusion stage. Through experiments on the pes24 dataset, an impressive average accuracy of 73.4% was achieved, surpassing the most advanced methods. These results demonstrate the effectiveness of the improved Pest-YOLO model in accurately detecting pests in real-time scenarios.[4]The team from Guizhou University addressed the issues of poor real-time performance, low detection accuracy, and high hardware power consumption in the identification of Clivia pests and diseases in greenhouse and garden environments, and proposed a Clivia pest and disease detection method suitable for mobile operation: YOLO V4-TLite.Firstly, based on YOLO V4-Tiny, a sampling convolution layer was added to the input end of the model, and low-cost partial convolutions were used to replace the traditional convolutions in the backbone network. Secondly, an inverted residual network structure was adopted to form a lightweight backbone network. Then, a channel fusion sampling layer mechanism was employed. Finally, the improved model was migrated and deployed on the ROCK 5B mobile terminal, and experiments were conducted on typical pests and diseases of Clivia.The experimental results show that when the YOLO V4-TLite model is used for Clivia pest and disease detection on mobile terminals with low resources and low power consumption, it has better competitiveness

compared with the existing mainstream YOLO series models.[6]The team from South China Agricultural University, based on YOLO v4, adopted GhostNet — a lighter and faster lightweight network — as the backbone network for feature extraction. They also incorporated the core design of GhostNet by introducing the more cost-effective Ghost Module (a type of convolution) to replace the traditional convolutions in the neck structure, resulting in the lightweight YOLO v4-G model. On this basis, they further improved YOLO v4-G by using a new feature fusion method and the CBAM (Convolutional Block Attention Module) attention mechanism. This improvement enhanced detection accuracy without compromising detection speed or the model's lightweight nature, leading to the proposal of the YOLO v4-GCF litchi pest and disease detection model. The proposed YOLO v4-GCF litchi pest and disease detection model can effectively suppress interference from complex backgrounds, accurately and quickly detect litchi pest and disease targets in images, and thus provide a reference for research on real-time detection of crop pests and diseases in complex, unstructured backgrounds under natural environments.[7]The team from China Agricultural University addressed two key challenges in cotton leaf disease detection under natural environments: the high difficulty of detection, and the struggle of manually designed feature extractors to capture feature representations similar to those of cotton leaf pests and diseases. To tackle these issues, they proposed an improved attention mechanism-based YOLO v7 algorithm (CBAM-YOLO v7). Built on the foundation of the original YOLO v7 model, this improved model incorporates the CBAM (Convolutional Block Attention Module) attention mechanism between the Backbone and the Head. Additionally, a 4× down sampling operation is implemented in the Head part. Subsequently, the CBAM-YOLO v7 model is applied to the identification of cotton leaf pests and diseases, and comparative experiments are conducted against the conventional YOLO v5 and YOLO v7 models.[8]

3.2 the latest research achievements and technological progress from abroad

Internationally, few research teams and institutions have actively engaged in the study of visual image classification of pests and diseases based on the YOLO algorithm and have achieved a series of remarkable results. Professor Vilar-Andreu's team has focused on enhancing pest control in precision agriculture: a generalized deep learning method based on yolov8 insect detection, with the goal of real-time detection of the presence of any type of insect in any type of crop. Conduct a comprehensive performance evaluation using well-known datasets. Then the results of the training, validation and testing phases were discussed. It was obtained that the mAP50 value of the m model was 0.967 and the MAP50-95 value of the l model was 0.632.[5]

3.3 limitations of Existing Research

Based on current research landscape domestically and internationally, existing vision-based pest and disease image classification systems utilizing YOLO still face three prominent issues:

Inadequate Suppression of Complex Background Interference: Background elements such as weeds, soil, and crop branches in farmland environments are prone to be confused with pest and disease features. Although existing algorithms have incorporated attention mechanisms, most are optimized for single background factors and struggle to handle multi-factor composite interference.

Difficulty in Discriminating Similar Categories：

Some pests and diseases (e.g., early and late blight in tomatoes, aphids and whiteflies) exhibit similar morphological characteristics. Most current loss functions rely on cross-entropy loss, which has limited capacity to capture inter-class differences, resulting in ambiguous classification.

Imbalance Between Lightweight Design and Accuracy: While some lightweight approaches have been explored, they often sacrifice accuracy for speed, making it difficult to simultaneously meet the practical agricultural requirements of high accuracy, real-time performance, and low computational demand. This is particularly evident in scenarios requiring the detection of multiple pests and diseases at once, where performance fluctuations are significant.

4. The key problems to be solved

How to reduce the interference of complex backgrounds on the classification of pests and diseases?

How to enhance the algorithm's ability to distinguish pests and diseases with high similarity between classes and solve the problem of ambiguous classification.

Under the premise of ensuring classification accuracy, how can the lightweight and real-time performance of the algorithm be achieved to meet the requirements for system response speed in actual agricultural scenarios?

5. Research objective of the subject

Comprehensively review the research status of the YOLO series of algorithms and visual image classification of pests and diseases, clarify the application bottlenecks and development potential of current technologies in the identification of agricultural pests and diseases, and provide theoretical basis and direction guidance for subsequent algorithm optimization.

In-depth analysis of technical features and performance: Summarize the technical implementation methods of the YOLO algorithm in the classification of visual images of pests and diseases in existing research, including dataset construction and preprocessing methods, model training strategies, feature extraction methods, etc. Meanwhile, the system summarizes the classification performance indicators of the YOLO algorithm in different studies (such as accuracy rate, recall rate, F1 value, detection speed, etc.), compares and analyzes the advantages and disadvantages of different versions of the YOLO algorithm in the task of pest and disease classification, as well as the performance differences in complex environments (such as light changes, occlusion, background interference, etc.).

Summarize the existing problems and challenges Based on the analysis of the current application status and performance, the common problems and challenges faced by the current visual image classification system for pests and diseases based on the YOLO algorithm in practical applications are extracted. For instance, the detection accuracy of small target pests and diseases is insufficient, the samples in the dataset are unbalanced, the generalization ability of the model among different crops or pest and disease types is weak, and the robustness in complex field environments needs to be improved, etc.

Summarize research trends and development directions Combining the development demands of intelligent agriculture and the advancements in computer vision technology, this paper summarizes the research trends in the field of visual image classification of pests and diseases based on the YOLO algorithm. Such as the integrated application with other technologies (such as attention mechanisms, transfer learning, lightweight network design, etc.), the exploration of pest and disease classification methods based on multimodal data fusion, and the strategy of balancing real-time performance and accuracy for actual production scenarios, etc., provide direction guidance for subsequent related research.

Construct a systematic review framework: Through the above research, a comprehensive and systematic review framework on the application of the YOLO algorithm in the field of visual image classification of pests and diseases is formed, providing valuable references for researchers to quickly understand the research progress, key technologies and future directions in this field.

6. Research Contents

The basic principles, network structures and improved versions (such as YOLOv3, YOLOv4, YOLOv5, etc.) of the YOLO series algorithms, as well as the comparison of the advantages and disadvantages of different versions of the algorithms and their performance in image classification tasks. Investigate the relevant research achievements of applying the YOLO algorithm in the field of pest and disease image classification at home and abroad and summarize the deficiencies and areas for improvement of the existing research.

Firstly, conduct a systematic review and construct a domain knowledge framework. Starting from the fundamental level of systematic learning, this paper comprehensively combs through the core knowledge of the YOLO algorithm and the field of pest and disease visual image classification. On the one hand, it sorts out the development context of the YOLO series algorithms, including technical iterations from the initial version to subsequent versions such as YOLOv3, YOLOv4, YOLOv5, and YOLOv7. It clarifies the core improvements of each version in terms of network structure, loss function, feature extraction method, etc., and establishes an overall understanding of the YOLO algorithm system. On the other hand, it combs through the basic information of agricultural pest and disease visual image classification, covering the types of common pests and diseases, image features (color, shape, texture, etc.), the definition and boundary of classification tasks, as well as the key technical indicators in this field (such as accuracy, recall, detection speed, etc.), laying a knowledge foundation for subsequent application analysis and comparison. At the same time, it combs through important literature, industry standards and typical datasets (such as IP102, Plant Village, etc.) in this field over the past decade, forming a systematic data system.

Secondly, analyze the application and explore the characteristics of technology implementation. Based on the collated data, this paper conducts an in-depth analysis of the practical application of the YOLO algorithm in pest and disease visual image classification from the application level of systematic learning. It is classified by crop types (e.g., food crops, cash crops, fruits and vegetables, etc.) and pest/disease categories (diseases, pests) and analyzes the application cases of the

YOLO algorithm in different scenarios. Specifically, it includes the following aspects: （1）Analyze the basis for selecting YOLO algorithm versions in each application case, and explore the adaptation logic between algorithm versions and application scenarios by combining the characteristics of crops, pests, and diseases. （2）Analyze the construction and application details of datasets, such as data acquisition equipment (cameras, UAVs, etc.), environments (field, greenhouse, etc.), annotation methods, and the impact of dataset scale on algorithm classification performance.

（3）Analyze the technical workflow in practical applications, including image preprocessing (enhancement, denoising, etc.), model training parameter settings, and deployment methods (cloud, edge devices, etc.). Summarize the technical commonalities and individual characteristics in different application scenarios and explore the implementation rules of the YOLO algorithm in actual agricultural scenarios.

Thirdly, conduct comparative studies and reveal performance and differences. Through systematic comparative analysis, this paper presents the performance differences of YOLO algorithms in pest and disease visual image classification by combining quantitative and qualitative methods. From the perspective of algorithm versions, it compares the performance metrics of different YOLO versions in the same or similar pest and disease classification tasks, including accuracy, mAP, FPS, etc., and clarifies the advantages and disadvantages of each version in terms of precision and speed. For instance, it compares the performance of YOLOv5 in detecting small target pests/diseases with the real-time performance of YOLOv7.From the perspective of application scenarios, it compares the adaptability of the same YOLO version in classifying pests and diseases of different crops, analyzes its performance fluctuations under conditions such as complex backgrounds (e.g., weed interference), lighting changes, and occlusions, and summarizes the degree of influence of scenario factors on algorithm performance. From the perspective of technical solutions, it compares the differences in technical routes based on YOLO algorithms in different studies, such as the impact of data preprocessing methods and training strategies on the final classification results. For example, it compares the difference in generalization ability of the algorithm when data augmentation is adopted versus when it is not, providing a reference for technical choices in subsequent research.

REFERENCES

[1] Pan Wei; Zhu Rong; Shi Runfa ; Hu Guohua ;Lian Shun ;Mei Jian ; He Fenglian；Journal of Xiangnan University, 2024-04, Vol.45

[2] TIAN Bin;GU Bin;FEI Chen;WANG Xiaotuo;SUN Yan;NI Chengong;XIAO Junchun(Suzhou Polytechnic Institute of Agriculture, Suzhou 215200, China; Suzhou Wanlian Magnetic Induction Communication Technology Co., Ltd,Suzhou 215200, China)Using Agricultural Machinery & Maintenance, 2025 (3), p.74-77

[3] Ji-Yuan Ding, Wang-Su Jeon and Sang-Yong Rhee（2024）DM-YOLOv8: Cucumber Disease and Insect Detection using Detailed Multi-Intensity Features，Department of IT Convergence Engineering, Department of Computer Engineering，University of Kyungnam, South Korea, Changwon City

[4] Zhe Tang, Jiajia Lu, Zhengyun Chen, Fang Qi, Lingyan Zhang (2023) Improved Pest-YOLO: Real-time pest detection based on efficient channel attention mechanism and transformer encoder，Vol.78

[5] Vilar-Andreu, M (Vilar-Andreu, Mario) [1] ; García, L (Garcia, Laura) [1] ; Garcia-Sanchez, AJ (Garcia-Sanchez, Antonio-Javier) [1] ; Asorey-Cacheda, R (Asorey-Cacheda, Rafael) [1] ; Garcia-Haro, J (Garcia-Haro, Joan)；IEEE ACCESS

[6] SONG Zhiwen; LI Wei; TAN Wei; QIN Tao; LIU Jie; YANG Jing(2025)Detection method for Clivia Miniata pests and diseases on mobile terminal based on YOLO V4-TLite；Transactions of the Chinese Society of Agricultural Engineering

[7] WANG Weixing；LIU Zeqian；GAO Peng；LIAO Fei；LI Qiang ；XIE Jiaxing（2023）Detection of Litchi Diseases and Insect Pests Based on Improved YOLO v4 Model；

[8] ZHANG Nannan；ZHANG Xiao ；BAI Tiecheng ； SHANG Peng ；WANG Wenhan；LI（2023）Identification Method of Cotton Leaf Pests and Diseases in Natural Environment Based on CBAM － YOLO v7