Research Progress of Real-time Counting and Localization of Rice Planthoppers Based on Fully Convolutional Networks

Li Yong-Yue [1], Huang Xiu-Qin [2], Tian Xue-Jun [3], Li Lu [4], He Zhang-Zhang [5*]
[1-5] Innovation Team for Crop Growth and Stress Adaptation, Research Team for Crop Micronutrient Metabolism and Stress Resistance, College of Food and Biology, Jingchu University of Technology, Jingmen 448000, Hubei, China

* Corresponding Author: **He Zhang-Zhang**

Article Info

ISSN (online): 3049-0588
Volume: 03
Issue: 06
Received: 20-08-2026
Accepted: 22-09-2026
Published: 24-10-2026
Page No: 05-09

Abstract
As small-sized agricultural pests with wide-ranging hazards, rice planthoppers are difficult to monitor accurately by traditional manual inspection, which suffers from low efficiency, large measurement error and obvious time-lag and fails to satisfy the requirements of precise pest prevention and control. This paper reviews real-time counting and localization technologies for rice planthoppers based on the Fully Convolutional Network (FCN). Its core mechanism extracts multi-scale features of rice planthoppers through a lightweight backbone network. Combined with the channel-spatial hybrid attention mechanism and Feature Pyramid Network structure, the recognition performance for tiny targets is enhanced. Pixel-level localization and blob prediction of rice planthoppers are realized by adopting object-oriented loss-function, so as to complete accurate counting. Breaking through the detection bottleneck caused by the tiny size, frequent overlapping and complex background of rice planthoppers, this technology achieves real-time response under field conditions. It provides an efficient tool for dynamic monitoring of rice-planthopper populations, supports precise pesticide application and green pest-control strategies, reduces the risk of pesticide residues, improves agricultural productivity and sustainability, and supplies technical support for the development of intelligent plant-protection and precision agriculture.

DOI: https://doi.org/10.54660/GARJ.2026.3.6.05-09

Keywords: Fully Convolutional Network, Rice Planthopper, Loss Function, Feature Pyramid Network

1. Research Background

Agricultural production is always severely threatened by various plant diseases and insect pests. Characterized by fast reproduction, strong adaptability and wide damage range, insect pests have become one of the key factors restricting high yield and high quality of crops. Some pests gnaw stems, leaves, flowers and fruits of crops and directly destroy the normal growth of plants[1]. For example, cotton bollworm larvae feed on cotton buds, flowers and bolls, as well as fruits and stalks of corn, tomato and other crops, causing abscission of flowers and bolls and fruit rot, and damaging plant reproductive organs and nutrient-transport structures[2]. Control measures include deep ploughing to eliminate overwintering pupae, application of sex pheromone attractants for adult trapping, spraying chlorantraniliprole and methoxyfenozide at the larval-hatching stage, and conservation of natural enemies such as lady-beetles. Some insects act as transmission vectors of plant viruses and trigger large-scale disease outbreaks, which may lead to partial or total yield loss in farmland. Aphids, for instance, can transmit multiple plant viruses including cucumber mosaic virus and potato virus Y[3]. The prevention of virus-carrying aphids can be carried out by repelling aphids with silver-grey mulch, yellow-sticky-trap trapping, insect-proof net isolation, releasing lady-beetles or aphid parasitoids, early application of imidacloprid at seedling stage, crop rotation and planting of virus-resistant varieties to interrupt

virus transmission chains. Certain pests suck plant sap and result in weak crop growth and quality degradation. Rice planthoppers pierce crop epidermis with piercing-sucking mouthparts and ingest sap, which causes curled and whitened leaves and scarred fruits. Integrated management strategies including field-cleaning deep ploughing, blue-sticky-board trapping, physical isolation with insect-proof nets, release of predatory natural enemies, and spraying of biological agents such as ethyl-multi-fungin on the leaf abaxial surface at dawn and dusk can be adopted to control rice-planthopper infestation. In recent years, Artificial Intelligence (AI) technology has developed rapidly and been applied to a wide range of industries including industry, commerce and medical care[4]. Compared with conventional technologies, computer vision and deep-learning solutions are more convenient and efficient, and can greatly save time and labor costs. Artificial-intelligence technologies have been widely deployed to replace or even surpass manual labor and significantly improve work efficiency. In agriculture, AI-related applications are gaining popularity, covering intelligent agricultural equipment, crop-growth monitoring, and field pest-and-disease surveillance. Based on previous research achievements, scholars have proposed prediction and counting methods for rice planthoppers using the Fully Convolutional Network (FCN), so as to avoid economic losses caused by large-scale pest outbreaks once the population reaches the threshold and deliver feasible protection strategies for agricultural production[5].

2. Research Status

Monitoring technologies for large-bodied insects are relatively mature and achieve high detection accuracy. For instance, the YOLO computer-vision architecture has become a core prediction tool for large pests such as locusts and fall armyworms. Based on a one-stage detection mechanism, the algorithm divides monitored images into dense grids, where each grid independently predicts target bounding boxes and object categories. Combined with multi-scale feature fusion and non-maximum-suppression optimization, YOLO can effectively cope with interferences in complex field environments including vegetation occlusion and illumination fluctuation. It outperforms traditional two-stage algorithms significantly in both recognition accuracy and detection speed for large-size pests[6]. On this basis, by integrating time-series monitoring data and environmental factors (e.g. temperature-humidity, vegetation coverage), pest population diffusion models can be further constructed to realize early warning of outbreak risks, which supplies key technical support for regional joint-prevention and-control. The integration of artificial intelligence and unmanned-aerial-vehicle (UAV) remote-sensing provides an efficient technical route for pest prediction. Equipped with multi-modal sensors for visible-light, multispectral, hyperspectral and thermal-infrared imaging, UAVs can rapidly acquire morphological, spectral and thermal features of crop canopies and capture subtle physiological changes induced by pest stress. A single UAV can monitor more than 3 000 mu of farmland per day, building an integrated space-sky-ground multi-scale perception network. AI algorithms serve as the core of data interpretation. Convolutional Neural Networks (CNN) can precisely extract features such as pest speckle textures and spectral anomalies, while transfer-learning technology enables high-accuracy recognition with small-sample datasets[7]. Time-series models such as Long-Short-Term Memory (LSTM) networks can fuse meteorological, soil and historical pest data, mine the rules of pest occurrence and diffusion, and achieve accurate prediction of outbreak time and scope. This technology has been applied on a large scale in major grain-producing regions, shifting plant-protection management from passive control to active early warning and laying a solid foundation for food security[8].

Multiple monitoring approaches are also available for tiny insect pests. Infrared sensors detect pests based on the emission, reception and reflection of infrared light. By identifying infrared-beam occlusion or differences in self-emitted infrared radiation, these sensors can detect pest presence, count individuals, track movement trajectories and preliminarily classify pest species. Three core mechanisms are adopted: occlusion detection, infrared-radiation detection and near-infrared scattering/reflection recognition. Such sensors have been widely deployed in intelligent insect traps and pest-monitoring warning lamps. Radar-based monitoring is another feasible technique. Dual-polarization weather-radar data combined with fuzzy-logic algorithms can build recognition models for echoes from precipitation, insects and birds. These algorithms have been applied to quantitative precipitation estimation and the analysis of rice-planthopper migration[9]. A review of current insect-recognition and monitoring research shows that a multi-technical monitoring system consisting of infrared sensing, machine vision and biological sensors has been well-developed for large-bodied insects. Benefiting from prominent target features and easily-captured signals, high-precision counting, localization and species-identification can be realized, with high technological maturity and wide application coverage. Nevertheless, existing monitoring techniques generally show obvious shortcomings for tiny-sized pests such as rice planthoppers. Technical solutions are relatively scarce. Restricted by tiny body size, inconspicuous morphological features and susceptibility to environmental interference, these technologies fail to meet the practical demands of precision pest control in farmland, forming a critical bottleneck for the construction of green pest-management systems targeting small-size pests[10]. To tackle this problem, this paper proposes a real-time counting and localization method for rice planthoppers based on Fully Convolutional Networks (FCN). Built upon conventional deep-learning object-detection frameworks, this approach innovatively introduces multi-modal image-fusion technology and a modified loss function. By fusing image features with different resolutions and spectral dimensions, the contrast between rice-planthopper targets and background environments is effectively enhanced, solving the problem of insufficient feature extraction for tiny-sized objects[11]. Meanwhile, optimized loss-function design balances counting accuracy and localization precision, substantially improving the robustness of rice-planthopper detection under complex field conditions. This technical scheme not only achieves accurate quantitative counting of rice-planthopper populations and precise localization of individual pests, but also enables data-driven inversion of population dynamic trends. It provides scientific data support and decision-making basis for early warning, precision pesticide application and optimization of control strategies against rice-planthopper infestations in agricultural production, and

bears great significance for promoting the intelligent and standardized development of monitoring technologies for small-sized insect pests[12].

3. Introduction to Artificial-Intelligence-Related Technologies

Artificial intelligence boasts a wide range of application fields and a rich, diversified technical system. It achieves remarkably higher accuracy in object detection and recognition tasks than manual operation, with prominent advantages in cost control. This section elaborates on its core relevant technologies.

3.1. I3D Network

As a three-dimensional spatiotemporal convolutional neural network extended from 2D convolutional networks, I3D is applied to video-based action recognition and spatiotemporal-feature extraction. Its core idea is to "expand" the convolution kernel of 2D-CNN into 3D, so as to capture both spatial features (image content) and temporal features (motion variation) from videos and satisfy the demand for spatiotemporal modeling of video sequences. Its main feature lies in extending 2D convolution (H×W×C) to 3D convolution (T×H×W×C, where T denotes the temporal dimension) to realize joint spatiotemporal-feature extraction. It is often fed with two-stream inputs including RGB stream (for appearance information) and optical-flow stream (for motion information) to boost action-recognition accuracy[12]. Pre-trained 2D-model weights can be adopted to initialize the 3D network, reducing training difficulty and improving generalization performance. In the medical field, for instance, this technology can recognize abnormal human behaviors for early prevention[13].

3.2. R(2+1)D Network

R(2+1)D is an efficient convolutional neural network for video action recognition. Its core design decomposes 3D convolution into 2D spatial convolution and 1D temporal convolution to balance model performance and computational efficiency. Adopting a "spatiotemporal-separation" strategy, it splits a single 3D-convolution operation into two steps: 2D convolution first extracts spatial features from individual frames; then 1D convolution captures temporal dynamics across successive frames[14]. This scheme cuts computational overhead, accelerates feature-learning efficiency, avoids gradient vanishing and delivers more stable training. Its basic building block is the R(2+1)D residual block, which contains two 2D spatial-convolution layers and one 1D temporal-convolution layer. Residual connections are introduced to propagate gradients, and blocks can be stacked to build deep networks of different depths (e.g. R(2+1)D-18/34/50). Taking a (3×3×3) convolution kernel as an example, it is decomposed into a 3×3×1 2D spatial convolution and a 1×1×3 1D temporal convolution, which greatly reduces parameters and computational cost[15]. Shallow layers extract low-and-mid-level motion features through spatiotemporal convolution, while deep layers enhance semantic information on the basis of spatial features to support classification, detection and other downstream tasks. Compared with 3D-CNN, R(2+1)D requires far fewer parameters and computations while maintaining high accuracy. It achieves competitive performance on datasets such as Kinetics-400 and UCF101, outperforms 2D-CNN and approaches the performance of optimal 3D-CNN models. Hybrid convolution (shallow 3D + deep 2D) is also supported to adapt to diverse task scenarios, such as video-based human-action recognition (abnormal-behavior detection, sports-motion analysis), intelligent surveillance and autonomous-driving video analysis[16].

3.3. SlowFast Network

SlowFast is a two-branch asynchronous convolutional neural network designed for hierarchical spatiotemporal modeling in video understanding. It processes "slow-changing semantic information" and "fast-changing motion information" via two parallel branches, and fuses cross-branch features to strike an excellent balance between prediction accuracy and computational efficiency. It has been widely adopted in video action classification, detection and temporal reasoning[17]. The slow branch captures spatial semantics and appearance information. It runs at a low sampling frame rate with a large temporal stride. Built on deep 3D-ResNet with a large-dimension-D channel and strong spatial convolution capability, it gradually downsamples along the temporal dimension and undertakes core semantic learning, consuming approximately 80 % of the total computation resources. By contrast, the fast branch captures fine-grained temporal motion and dynamic information (such as motion trajectories and rapid shape changes). It samples frames at a high frame-rate with a short temporal stride[18]. Structurally, it adopts a shallow 3D-ResNet with β-D channels and no extra temporal down-sampling, so that high temporal resolution can be maintained. Its lightweight design guarantees computational efficiency, consuming around 20 % of the total computation cost. SlowFast has become a classic basic model for various video-understanding tasks, including action recognition, behavior detection and video segmentation[19].

3.4. TimeSformer

TimeSformer is the first pure-Transformer-based classic model for spatiotemporal feature modeling in video-understanding tasks. Its core innovation extends self-attention into both spatial and temporal dimensions. By adopting decomposed spatiotemporal attention, it greatly cuts computational costs while maintaining competitive performance, achieving excellent results on video action-recognition and other tasks. Video patch embedding splits each input video frame into non-overlapping image patches of fixed size. For instance, a 224×224 frame can be divided into 16×16-sized patches, yielding 14×14=196 patches in total. Each patch is linearly embedded, followed by position encoding and temporal encoding to form input vectors of spatiotemporal sequences[20]. An extra classification token is appended for final prediction. The decomposed spatiotemporal-attention mechanism computes attention among different patches within one single frame to capture spatial structures and local features such as object location and posture. Meanwhile, attention is calculated between patches at identical spatial positions across different frames, so as to extract temporal dynamics including action continuity and changing trends. This mechanism splits the joint spatiotemporal computation into two separate stages,

which reduces computational complexity and substantially improves efficiency. Multiple encoder layers are stacked. Each layer contains decomposed spatiotemporal attention, layer normalization, and an MLP block. GELU is commonly selected as the activation function, and regularization strategies such as Dropout are applied to avoid over-fitting. Finally, features of the classification token are fed into a fully-connected layer to generate video classification outputs. Compared with 3D convolutional networks, TimeSformer achieves approximately 3-fold faster training speed and up to 10-fold higher inference speed. It can process much longer video clips (over one minute). Since self-attention has no restriction of local receptive fields, long-range spatiotemporal dependencies can be captured, enabling more precise understanding of complex multi-step action sequences. The model can be scaled-up by increasing network depth, head number and embedding dimension, so as to adapt to high-resolution or long-duration video inputs. With strong generalization capacity, it has been widely deployed in video action-recognition tasks such as human-activity analysis, sports-motion evaluation and abnormal-behavior detection[21].

3.5. TSM Network

Temporal Shift Module (TSM) is a lightweight spatiotemporal-feature modeling network for video understanding. Built upon mature 2D-Convolutional Neural Networks (CNN), it introduces a parameter-free or light-parameter temporal-shift module. Without computation-intensive 3D convolutions, TSM efficiently fuses temporal features between successive video frames. While preserving the powerful spatial-feature extraction capability of 2D-CNN, it endows the model with temporal-modeling ability at low computational cost. It is a well-recognized video-understanding framework that balances inference speed, modeling accuracy and parameter consumption. TSM shifts partial feature channels of video frames forwards or backwards along the temporal dimension (±1 frame shift) to realize information fusion between adjacent frames, requiring no extra convolution operations or additional parameters and producing nearly zero inference latency[22]. Two operation modes are available: - Offline mode (bidirectional shift): 1/8 of channels are shifted to the previous frame and another 1/8 to the subsequent frame, suitable for high-throughput offline tasks; - Online mode (unidirectional shift): only 1/8 of channels from the previous frame are shifted to the current frame, which fits low-latency real-time scenarios. Only data rearrangement is performed, with no extra FLOPs or parameters. Taking 2D-CNN backbones such as ResNet as the main framework, TSM modules are embedded inside residual branches of residual blocks. After temporal-information fusion via TSM on input video frame sequences, global average pooling and fully-connected layers produce final classification outputs. TSM runs at the speed of pure 2D-CNN and achieves accuracy close to that of 3D-CNN. Owing to no extra parameters and computations, it is well-suited for edge-side equipment and real-time applications such as intelligent monitoring and mobile-terminal video classification, and has been widely used in human-behavior recognition and motion-tracking tasks[23].

4. OutLook

At present, machine vision technology based on deep learning has become mainstream. As a core technology in current artificial intelligence, deep learning has two major advantages. First, it boasts powerful learning capacity: it can automatically learn and extract features from data, and its performance continuously improves as the size of datasets increases. Second, it features an end-to-end architecture, which eliminates the need for complicated preprocessing or feature engineering and delivers stronger fitting capability for complex and high-dimensional data.

After years of development, intelligent detection technologies for rice planthoppers have gradually evolved from traditional image processing and machine learning approaches to deep-learning-based artificial intelligence technologies, achieving remarkable improvements in dataset scale, model architecture and detection accuracy.

References

1. Abbas M，Ramzan M，Hussain N，*et al*. Role of light traps in attracting, killing and biodiversity studies of insect pests in Thal [J]. *Pakistan Journal of Agricultural Research*，2019，32（4）: 684-690.
2. Abbas M，Ramzan M，Hussain N，et al. Role of light traps in attracting, killing and biodiversity studies of insect pests in Thal [J]. Pakistan Journal of Agricultural Research，2019，32（4）: 684-690.
3. Cai ZW，Vasconcelos N. Cascade R-CNN: delving into high quality object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)，2018.
4. Chithambarathanu M，Jeyakumar MK. Survey on crop pest detection using deep learning and machine learning approaches [J]. Multimedia Tools and Applications，2023，82（27）: 42277-42310.
5. Dorj U，Lee M，Yun S. An yield estimation in citrus orchards via fruit detection and counting using image processing [J]. Computers and Electronics in Agriculture，2017，140: 103-112.
6. He K，Zhang X，Ren S，et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)，2016.
7. Huang W，Huo Y，Yang S，et al. Detection of Laodelphax striatellus (small brown planthopper) based on improved yolov5 [J]. Computers and Electronics in Agriculture，2023，206: 107657.
8. Ibrahim MF，Bejo SK，Hanafi M，et al. Deep CNN-based planthopper classification using a high-density image dataset [J]. Agriculture，2023，13（6）.
9. Karam C，Awad M，Abou Jawdah Y，et al. GAN-based semi-automated augmentation online tool for agricultural pest detection: a case study on whiteflies [J]. Frontiers in Plant Science，2022，13.
10. LeCun Y，Bengio Y，Hinton G. Deep learning [J]. Nature，2015，521（7553）: 436-444.
11. Li R，Wang R，Xie C，et al. A multi-branch convolutional neural network with density map for aphid

counting [J]. Biosystems Engineering, 2022, 213: 148-161.
12. Liu L, Xie C, Wang R, et al. Deep learning based automatic multiclass wild pest monitoring approach using hybrid global and local activated features [J]. IEEE Transactions On Industrial Informatics, 2021, 17（11）: 7589-7598.
13. Lu Y, Chen D, Olaniyi E, et al. Generative adversarial networks (GANs) for image augmentation in agriculture: a systematic review [J]. Computers and Electronics in Agriculture, 2022, 200: 107208.
14. Nazri A, Mazlan N, Muharam F. Penyek: automated brown planthopper detection from imperfect sticky pad images using deep convolutional neural network [J]. Plos One, 2018, 13（12）: e208501.
15. Park Y, Han M, Kim H, et al. Density estimation of rice planthoppers using digital image processing algorithm [J]. Korean Journal of Applied Entomology, 2003: 57-63.
16. She J, Zhan W, Hong S, et al. A method for automatic real-time detection and counting of fruit fly pests in orchards by trap bottles via convolutional neural network with attention mechanism added [J]. Ecological Informatics, 2022, 70: 101690.
17. Wang F, Wang R, Xie C, et al. Convolutional neural network based automatic pest monitoring system using hand-held mobile image analysis towards non-site-specific wild environment [J]. Computers and Electronics in Agriculture, 2021, 187: 106268.
18. Wang Q, Zhang S, Dong S, et al. Pest24: a large-scale very small object data set of agricultural pests for multi-target detection [J]. Computers and Electronics in Agriculture, 2020, 175: 105585.
19. Wang R, Jiao L, Xie C, et al. S-RPN: sampling-balanced region proposal network for small crop pest detection [J]. Computers and Electronics in Agriculture, 2021, 187: 106290.
20. Wei M, Zhan W. Yolo_mrc: a fast and lightweight model for real-time detection and individual counting of tephritidae pests [J]. Ecological Informatics, 2024, 79: 102445.
21. Zhan W, Sun C, Wang M, et al. An improved yolov5 real-time detection method for small objects captured by uav [J]. Soft Computing, 2022, 26（1）: 361-373.
22. Zhang N, Yang G, Pan Y, et al. A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades [J]. Remote Sensing, 2020, 12（19）.
23. Zhang Z, Zhan W, He Z, et al. Application of spatio-temporal context and convolution neural network (CNN) in grooming behavior of Bactrocera minax (Diptera: Trypetidae) detection and statistics [J]. Insects, 2020, 11（9）: 565.

How to Cite This Article
Li YY, Huang XQ, Tian XJ, Li L, He ZZ. Research progress of real-time counting and localization of rice planthoppers based on fully convolutional networks. Glob Agron Res J. 2026;3(6):5-09. doi:10.54660/GARJ.2026.3.6.05-09.

Creative Commons (CC) License
This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.