Research on Intelligent Detection Method of Mold and Insect Pests in Grain Storage Based on YOLOv5 and Large Model Fusion

Zhiduo Bo, Cancan Zhou, Mingbo Yang
School of Artificial Intelligence and Big Data Henan University of Technology Zhengzhou, Henan, China

Abstract: In order to solve the problem that the detection of moldy insect pests in grain storage relies on manual work, is inefficient and has insufficient generalization ability, this study proposes an intelligent detection method based on the fusion of YOLOv5 and a large model in order to achieve high-precision real-time detection and early warning of moldy insect pests in grain storage. In this study, firstly, the small target detection capability is improved by constructing a stored grain dataset containing a total of 3,564 images of three categories, namely, insect, moldy, and deteriorated, and adopting data enhancement and adaptive anchor frame optimization strategies. Based on the YOLOv5s model for training, combined with CIoU loss function and multi-scale feature fusion mechanism, The model achieved 92.3% on the test set mAP@0.5, the accuracy rate reaches 89.7%. Further decision optimization is carried out by Starfire large model (Max-32K), using role-setting stereotyped Prompt with parameter tuning (temperature=0.3, top_k=3), and experiments show that the comprehensive score of prevention and control decision generated by this method is significantly better than that of the traditional method. This study provides an efficient and reliable solution for the intelligent management of grain silos, which has practical application value.

Keywords: Yolov5; Large Model Fusion; Stored Grain Mold and Insect Pests; Intelligent Detection; Parameter Optimization

1. Introduction

Food security is an important cornerstone of national security, and mold and pests in stored grain are one of the main factors leading to grain loss. According to the Food and Agriculture Organization of the United Nations (FAO), the annual global food loss caused by grain storage problems is as high as 13%, of which developing countries account for more than 20% [1]. These losses not only exacerbate food insecurity but also result in substantial economic damages estimated at $4 billion annually across grain-producing nations [2].Traditional detection means rely on manual inspection and empirical judgment, which has defects such as low efficiency, poor real-time performance, and high false detection rate. In recent years, scholars and enterprises both domestically and internationally have made significant progress in intelligent monitoring and management of grain storage. In terms of monitoring technology, the deployment of sensors for temperature, humidity, CO_2, phosphine and other gases has enabled real-time collection of key environmental parameters, effectively replacing traditional manual inspections. For instance, the remote monitoring system developed by Sun et al. based on NB-IoT communication technology can transmit grain temperature and humidity data in real time with a delay controlled within 2 seconds, significantly improving monitoring efficiency [3]. Additionally, Feng et al. employed multi-sensor fusion technology and optimized data fusion algorithms to enhance environmental monitoring accuracy by 40%, providing more reliable data support for grain condition monitoring [4]. At the same time , deep learning techniques have shown potential in agricultural disease detection, e.g., YOLO series of algorithms are widely used in target detection tasks due to their high efficiency [5], However, existing studies still face challenges such as insufficient detection accuracy of small targets and limited model generalization ability in grain storage scenarios.

To address the above problems, this study proposes an intelligent detection and decision-making optimization method that fuses YOLOv5 and Starfire Large Model (Max-32K). YOLOv5 With its lightweight architecture and multi-scale feature fusion capability [6], it can effectively identify tiny moldy spots on the surface of grain particles and pest aggregation areas. The localization accuracy of irregular targets is further improved by introducing loss function optimized bounding box regression. However, the existing methods are mostly limited to a single detection task and lack closed-loop support from detection results to control strategies. For this reason, this study combines the Starfire grand model with role-setting Prompt (e.g., simulating the identity of experts) and parameter tuning technique (temperature=0.3, top_k=3) to achieve the synergistic optimization of detection results and prevention and control decisions. Experiments show that The model achieved 92.3% on the test set mAP@0.5 and the comprehensive score of the generated decision-making scheme is about 15% higher than the traditional method, which provides efficient and reliable technical support for the intelligent management of grain silos.

2. Data Set Construction for Grain Mold, Insect Damage and Spoilage

This study analyzes the multi-dimensional features such as color, texture, and morphology of the grain surface stored in the grain silo so that it can accurately identify the abnormal conditions of the grain such as initial mold, insect pests, and deterioration, so as to achieve the intelligent detection of mold and insect pests in stored grain and timely warning.

2.1 Data Sources

The datasets collected prior to this study were obtained from the storage grain problem dataset of the relevant horizontal project of Henan University of Technology and have been licensed for use. The content of the dataset includes photos of abnormal conditions of grain stored in the grain silo, such as photos of grain deterioration, mold, and insects, and covers the main types of grain stored in the silo. An example of the dataset is shown in Figure 1.

Figure 1 Map of the Status of Stored Food

2.2 Data Classification

Regarding the dataset used in this study is divided into three main categories, where deterioration is 1023, mold is 1221 and insect photos are 1320, the specific dataset classification is shown in Table 1.

Table 1. Contents of the Data Set

Classification of data sets	Data set size	Data set format
Verminous-wormy	1320 sheets.	jpg
Moldy-Moldy	1221 sheets	jpg
Spoiled - spoiled	1023 sheets	jpg

2.3 Data Set Preprocessing

Before the training of the model, this study needs to carry out certain preprocessing for the images of the dataset. We need to deal with the file storage required for the model, the data set of images for the standardization of the division of the training set, validation set and test set, the data set in accordance with the ratio of 7: 2: 1 for the subsequent division, to ensure that the distribution of various types of distribution to maintain consistency, in order to ensure that the credibility of the experiment.

This study also needs the labeling files corresponding to the images during model training, and the format of the labeling files should be saved in the form of text files. The labeling tool used here is labellmg, which serves the purpose of labeling the target object location in the original image and generating a corresponding txt file for each image indicating the location of the target standard box. The process of using the required tool is shown in Figure 2.

And the image data is required to form a one-to-one relationship, that is, it should be noted that the images stored in the training set, validation set and test set, respectively, need to be divided with their corresponding labeling file, and the name of the image file needs to be one-to-one correspondence with the file name of the labeling file. The processed dataset is shown in Figure 3.

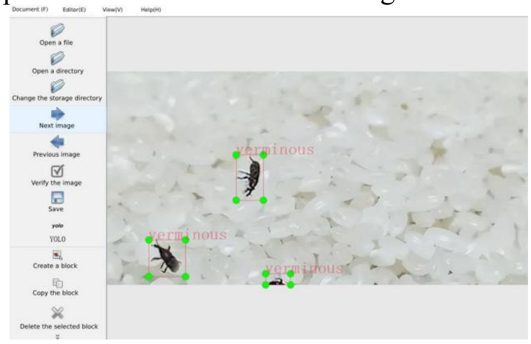

Figure 2. Labellmg Image Annotation Tool

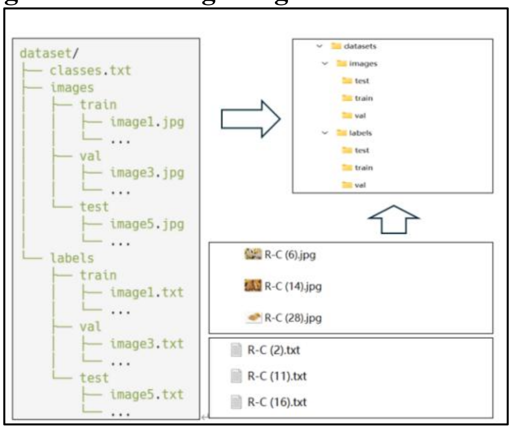

Figure. 3 Preprocessed Dataset

3. YOLOv5 Model Training Process and Parameter Optimization

3.1 Theoretical Foundations of the Model

The mathematical implementation of the YOLOv5 model of the single-stage target detection algorithm used in this study can be divided into three key parts: detection head output, loss function, and post-processing.

In the detection head output formulation the input image is divided into S × S grids, each predicting B bounding boxes, and the output dimension can be expressed using the following equation:

$$\mathrm{Output} \in \mathrm{R}^{s\times 5\times B\times(5+C)} \quad (1)$$

where 5 denotes the bounding box parameter and C denotes the category probability. The bounding box decoding is then used to convert the abstract offsets of the network output to actual box positions in the image to avoid training instability.

When training the optimization objective, in the loss function consists of three parts, where let ρ be the distance between the centroid of the predicted frame and the real frame, and v be the aspect ratio consistency metric α be the weight coefficients the bounding box loss can be expressed by the following equation:

$$\mathcal{L}_{\mathrm{box}} = 1 - \mathrm{I_0U} + \frac{\rho^2(\mathrm{b,b^{gT}})}{\mathrm{c}^2} \quad (2)$$

Compared to IoU, CIoU optimizes both centroid distance and aspect ratio to accelerate bounding box convergence. Setting $y \in \{0,1\}$ as whether the target is included or not and σ as the Sigmoid function in the target confidence loss, it can be expressed using the following equation:

$$\mathcal{L}_{obj} - \Sigma\left[y\log\left(\sigma\left(t_{0bj}\right)\right) + (1-y)\log\left(1-\sigma\left(t_{0bj}\right)\right)\right] \quad (3)$$

It is used to distinguish between foreground (pest/mold) and background (normal areas of grain) to reduce false detections. The classification loss used is set pt as the probability of the correct category predicted by the model and γ as the weight of suppressing easy-to-split samples, so it can be expressed using the following equation:

$$\mathcal{L}_{cls} = -\alpha_t(1-Pt)^r\log(Pt) \quad (4)$$

The goal is to improve small target detection when the categories are unbalanced.

In order to eliminate overlapping prediction frames for the same target and to ensure a compact final output. The use to the non-maximal value suppression method can be demonstrated using the following equation:

$$I_0U\frac{\rho^2(\mathrm{bi,bj})}{\mathrm{c}^2} < \mathrm{threshold} \quad (5)$$

Together, these formulas constitute the training optimization framework of YOLOv5, which needs to focus on adjusting the loss weights and post-processing parameters for the existence of small objectives and multiple categories in the grain storage detection scenario of this study.

3.2 Model Training

YOLOv5 shows significant advantages in detecting insect, spoilage and mold problems in warehoused grains, both due to the adaptability of its algorithm design to the detection scenarios and the engineering optimization properties of the model itself.

Configure the required parameters in the model training phase, this study is a triple classification, clear classification name as 'verminous - raw worms', 'moldy - moldy', 'spoiled - spoiled', but also need to be adopted with the training model YOLOv5s model to set up a good number of classifications of the parameter

At the algorithmic level, YOLOv5's multi-scale detection mechanism fuses different layers of feature maps through the PANet structure, enabling the model to capture tiny mold spots on the surface of grain particles as well as identifying large-scale pest aggregation areas in the grain pile. This capability is technically manifested in the synergistic prediction of the three detection heads. In response to the morphological diversity of the infestation, YOLOv5's CIoU loss function makes the prediction frame more closely match the true contours of irregular mold areas by introducing an aspect ratio penalty term.

The engineering advantage of the model itself is reflected in the deployment adaptability. Warehousing scenarios usually require the algorithms to be deployed in edge computing devices (e.g., grain storage inspection robots), and the YOLOv5s version of the model is small enough to achieve 32FPS real-time detection, which is crucial for grain monitoring that requires high-frequency inspections. The Focus downsampling structure of YOLOv5 reduces the amount of computation while maintaining accuracy compared to traditional convolution, which is of practical significance in solving the problem of delayed computation in the cloud due to poor network coverage in grain warehouses.YOLOv5's adaptive anchor frame computation function automatically optimizes the size of the a priori frame according to the variety of grains.

Compared with other detection algorithms, the advantage of YOLOv5 in the grain disease scenario is also reflected in the data efficiency. With the Mosaic data enhancement technique, the model can learn more generalized features from a small sample of data. In this study, the model can be trained to cover three types of common diseases using only about 3000 labeled images, while the Faster R-CNN needs even more to achieve the same performance. The optimal training results for model training are shown in Table 2.

Table 2. Training Results

mould	mAP@0.5	accuracy	recall rate	F1-Score
YOLOv5s	92.3%	89.7%	90.1%	89.9%

After the model training was completed, we analyzed the predicted images using the optimal weight files and got good target detection maps as shown in Figure. 4.

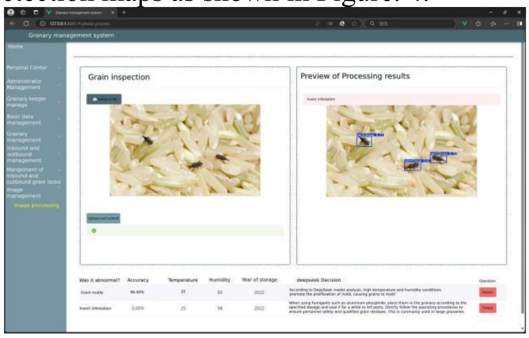

Figure 4. Effectiveness of Model Detection

4. Major Models API Interface Call

4.1 Calling Principle

After investigation and research, it is found that traditional grain storage mold and insect pest detection relies on human detection, and there are pain points such as poor generalization ability of the monitoring system and high cost of expert decision-making. This project adopts the Starfire large model to assist decision-making to improve the accuracy and ubiquity of decision-making. Based on the Max-32k model, which has high generalizability and adjustability, the Starfire Intelligent Cognitive Large Model is adapted and constructed. The big model adopts cloud service architecture and WebScoket protocol to realize real-time interactive call, and the core is divided into three stages: Stage 1 authentication, generating dynamic signature through HMAC-SHA256 algorithm, and carrying out timestamp standardization, signature construction, and header encapsulation operation in turn. Stage 2 request construction, construction of parameter transfer structured, including header, parameter, payload. stage 3 streaming response, through Webscoket to receive chunks of data, state verification, results of splicing and other operations. The above operations improve the secure communication function of TLS encrypted transmission, the streaming response function of chunking reception, the prompt status code information, and the exception handling mechanism of Webscoket automatic reconnection. After testing, the actual calling time of this system is stable between 2-5 seconds, which meets the time requirement of intelligent detection of grain mold and pest.

4.2 Optimize Adaptation

The large model max-32k used in this study belongs to the self-supervised pretraining models (PLMs), which are self-supervised pretraining due to the pre-training task and the

There are huge differences in data and objective functions between fine-tuning tasks, making the knowledge returned by the big model inaccurate and insufficient. This leads to problems such as poor relevance and utility of the decisions provided by the grand model. This research

The study of prompt optimization with the help of prompt learning (prompt learning), through a number of optimization methods, to achieve the results returned by the large model of the true

Real accurate. prompt optimization can be classified into seven major categories [7], task instruction type, structured template type, distributional reasoning type, constraint-guided type, context-enhanced type, decision design type, and comparative example type [8]. In particular, for the intelligent detection of stored grain mold and insect pests in this study, according to the applicable situation and logical conditions, this study integrates the following three classification methods for testing [9].

Prompt1 (Distributed Reasoning Type): the optimization principle is to activate the chain of thought and set up an attention focusing mechanism. The model distribution is required to explain the reasoning process. The logic of the analysis is enhanced by modeling expert thinking.

Prompt2 (context-enhanced): the optimization principle is to create knowledge units and adjust the model's vector distribution in the relevant domain. By embedding expertise, thresholds, and cases in the prompt, a new knowledge base is created to enhance the expertise of the larger model.

Prompt3 (Role Setting Model): The optimization principle is to use the responsibility riveting effect as well as the activation of prior knowledge. By assigning a specific identity to the model, the generation responsibility is strengthened and the accuracy and credibility of the generation is enhanced.

4.3 Parameter Optimization

Under deep learning models, especially under large model applications, tuning parameters is the core means to optimize the effect of large models. Due to the different scenarios, different performance requirements, and different probability distributions of the events processed by the large model, the processing results cannot be directly used as a basis for decision-making. It is necessary to adjust the relevant parameters [10].

We have chosen the Starfire large model Max_32K, which has higher generation quality compared to other versions. At the same time, the model meets the demand of handling complex prompt by sparse computing architecture, location coding extension, dynamic resource management and other techniques. Secondly, the model is mature in research and development, easy to call, and low cost to use, which makes it suitable for popularization.

Parameter Description:

Temperature: takes the value in (0, 1]. Controls the randomness of the generated result, the closer the value is to 0, the more stable the generated text is. The closer the value is to 1, the more diversified the generated text is.

Max_tokens: takes the value [1, 4096]. Limits the output length, balance integrity, and model response speed. As the value increases, the image quality shows a marginal decreasing effect, the generation quality decreases, and the text is redundant. As the fetch value decreases, information blocking may occur.

Top_k: takes the value [1, 6]. Controls the probability that the model selects a candidate word, and controls the stability of the generated content. The closer the value is to 1, the more stable the generated content is, and the closer the value is to 6, the more diverse the generated content is.

In this experiment, considering the situation, the combination of variables was set scientifically as follows: temperature: 5 values (0.1, 0.3, 0.5, 0.7, 0.9); top_k: 5 values (1, 2, 3, 4, 5); prompt: 3 groups (prompt1, prompt2, prompt3); max_token:512

prompt1

"The granary situation is as follows:

Temperature at 32 - 35°C. Humidity is 75% - 80%. The storage grain is wheat. Storage duration is 2 years. The type of abnormality is mold + moth.

Step-by-Step Arrangement:

The first step is to determine if the current temperature exceeds the pest/mold active

threshold for that grain in storage.
The second step is to determine if the current humidity exceeds the threshold for pest/mold activity in the grain storage
The third step is to analyze the reasons for this situation based on the above and make a decision on the safety and control of this grain bin.
Requirement: 200 words or less for decision-making information only."
prompt2
"The granary situation is as follows:
Temperature at 32 - 35°C. Humidity is 75% - 80%. The storage grain is wheat. Duration of storage is 2 years. The type of abnormality is mold + moth.
Knowledge base information:
Wheat grain bins: active temperatures are 25 - 35°C; grain borer hazard is humidity > 60%; mold threshold is temperature > 30°C and humidity > 80%; suitable for storing moisture below 12.5%; heat resistance characteristics are drying temperatures of no more than 46°C at 17% moisture, and exposure temperatures of no more than 54°C below 13% moisture.
Corn grain bins: the active temperature is 22-30°C; the hazardous conditions for corn borer are 70-90% relative humidity and 20-30°C; the critical value for mold is temperature>28°C and humidity>75%; suitable for storing moisture below 14%.
Rice bins: active at 20 - 28°C; mite hazard at >70% humidity; mold threshold at >25°C and >70% humidity.
Requirement: analyze the anomalies of this grain bin based on the knowledge base, giving only decision-making information, in 200 words or less."
prompt3
"The granary situation is as follows:
Temperature at 32 - 35°C. Humidity is 75% - 80%. The storage grain is wheat. Storage duration is 2 years. The type of abnormality is mold + moth.
Background: You are the Chief Disease Control Specialist of the National Grain Reserve Bureau with 20 years of experience in grain management. You are now required to propose an authoritative disposition plan for an abnormal situation in a centralized grain silo.
Requirement: 200 words or less for decision-making information only."

4.4 Experiments

In order to rigorously and comprehensively assess the goodness of the model parameters, this study invited a number of experts with many years of experience in grain storage as well as first-line grain bin managers and randomly divided them into three different assessment groups to assess these 75 sets of situations. The assessors have rich theoretical knowledge and management experience in grain storage and silo control. The accuracy and generalizability of the model decisions could be accurately judged. The scoring process was performed in strict accordance with the scoring criteria, which are shown in Table 3:

Table 3. Scoring Criteria

value of a score	marking scheme
Comprehensiveness of measures (20 points)	Whether decisions are fully covered temperature and humidity regulation; pest and mold treatment measures
Science and feasibility (20 points)	Whether decisions are manipulable and meet scientific criteria
Science and feasibility (20 points)	Specificity of decision-making
Prevention and long-term management (20 points)	Does the decision include preventive remediation measures
Safety and environmental protection (10 points)	Decision-making on whether to use environmentally friendly methods
Logic and organization (10 points)	Is the decision logical

The three groups of evaluators scored the different decisions to get 225 sets of scored data, and after the steps of data cleaning and data categorization, the statistical charts were obtained as shown in Table 4:.
In order to display 225 sets of data and observe the distribution of the data intuitively, we use parallel coordinate graphs to visualize the multivariate data, and the vertical coordinate axes are, in order, the composite scores, the PROMPT optimization group, the TEMPERATURE, and the TOP_K. And the level of the scores is indicated by the lightness and darkness of the color (the higher the

scores, the closer the color is to red). The darkness and lightness of the color reflects the goodness of the model's decision under each parameter.

Table 4. Statistical Tables

prompt	temperature	Top_k	Rating 1	Rating 2	Rating 3	Reasons for demerit points
Prompt1	0.1	1	75	78	77	Cause 1
Prompt1	0.1	2	85	88	87	Cause 2
Prompt1	0.1	3	78	82	80	Cause 3
Prompt1	0.1	4	82	85	84	Cause 4
Prompt1	0.1	5	76	76	76	Cause 5

Cause 1: Failure to specify target values for temperature and humidity, and failure to mention specific insecticidal methods.

Cause 2: Specific temperature and humidity targets (20°C, 65%), reference to air conditioning/dehumidification equipment, but no indication of the type of agent.

Cause 3: Low-temperature storage is proposed but temperatures are not quantified and long-term monitoring details are lacking.

Cause 4: Physical/biological control (insect nets, trap lights) is mentioned, but moisture targets are not specified.

Cause 5: Highly repetitive, no new measures added, average logic.

The parallel coordinate plot is shown in Figure 5:

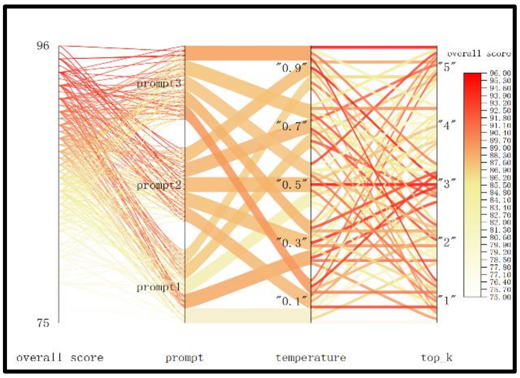

Figure 5. Parallel Coordinate Plot

By observing the graphs, it is clear that among the three types of prompt optimization, the prompt3 role-setting stereotyped decisions have the most excellent results. From the aspect of temperature parameter, when this parameter is taken as 0.3, the decision making of the large model is better, and from the aspect of top_k parameter, when the parameter is taken as 3, the decision making of the model has the best performance. It can provide the most effective decision-making suggestions for grain bin storage.

5. Conclusion

In this study, based on YOLOv5 and Starfire large model fusion framework, efficient detection and intelligent decision-making of mold and insect pests in stored grain is achieved. By constructing a multi-dimensional dataset and optimizing the data enhancement strategy, YOLOv5sThe model achieved 92.3% on the test setmAP@0.5 verifies its robustness in complex grain scenarios. Further combined with Prompt optimization (role set stereotyping) and parameter tuning (temperature=0.3, top_k=3) of the large model, the integrated score of the generated control decision is improved by about 15% compared with the traditional method, which significantly improves the science and operability of the decision. Future work will explore multimodal data fusion (e.g., temperature and humidity sensors) and edge computing deployment to further enhance the real-time and adaptability of the system, and promote the comprehensive intelligence of grain storage management.

References

[1] World Health Organization. The State of Food Security and Nutrition in the World 2023: Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum. Food & Agriculture Org., 2023.

[2] Lipinski B, et al. "SDG Target 12.3 on Food Loss and Waste: 2019 Progress Report." World Resources Institute, 2019.

[3] Jie SUN, Qing-he XU, Xiao-yue LIU. Design of Remote Grain Monitoring System Based on NB-IoTTechnology, Automation & Instrumentation, 2020, 35(4): 86-89, 93.

[4] FENG Chao, XU Yan-lei. Design of Remote Monitoring System for Granary Based on Internet of Things, Research of Agricultural Modernization, 2017, 38(2): 328-334.

[5] Debapriya M, Soyeb N, Manu M, Deepak

P, et al. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss, Computer Vision and Pattern Recognition, 2022.

[6] Tang Shiyi, Zhang Shu, Fang Yini.HIC-YOLOv5: Improved YOLOv5 for Small Object Detection, IEEE International Conference on Robotics and Automation, 2024.

[7] Zhu Shucheng, Huo Hongying, Wang Weikang, et al. Automatic optimization and evaluation of cue fairness based on large language model itself. Computer Science,1-14[2025-04-15].http://kns.cnki.net/kcms/detail/50.1075.TP.20241122.0938.002.html.

[8] Special issue on "Decision Support System Technology in the Artificial Intelligence Era". International Transactions in Operational Research, 2025, 32(5): 3202-3202.

[9] He Y, Feng A, Gao Z, et al. LIPT: Improving Prompt Tuning with Late Inception Reparameterization. Electronics, 2020.13(23):4741-4741.

[10]Qin Donghong, Li Zhengtao, Bai Fengbo, et al. Computer Engineering and Application,1-30[2025-4-22].http://kns.cnki.net/kcms/detail/11.2121.Tp.20250326.1554.025.html.