Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

# Small Object Detection Method for Farmed Animals in UAV Images Based on Improved YOLOv7-Tiny

**Zewei Kang[1,2], Jocelyn F. Villaverde[1*], Ji Zhang[2]**

[1] School of Electrical Electronics and Computer Engineering, Mapúa University, Manila, Philippines
[2] School of Electrical and Electronic Engineering, Guangdong Technology College, Zhaoqing, China
E-mail: jfvillaverde@mapua.edu.ph

**Received:** 6 June 2025; **Revised:** 21 June 2025; **Accepted:** 25 June 2025

**Abstract:** Unmanned Aerial Vehicle (UAV) technology plays a vital role in the animal husbandry industry. High-resolution images can monitor the spatial distribution and behavior patterns of animal populations in real time, thereby significantly improving the efficiency of animal breeding management. In view of the technical difficulties commonly found in animal husbandry, such as small target size, frequent occlusion, and unbalanced category distribution, this study proposes an enhanced animal small target detection algorithm based on the YOLOv7-tiny framework, named SPF-YOLOv7-tiny. The algorithm contains three key innovations: first, the Segment Anything Model (SAM) image segmentation technology is integrated to optimize the Mosaic data enhancement strategy, significantly improving the diversity of training samples; second, the dedicated small target detection head module enhances the feature extraction capability of tiny targets. Third, the Focal_DIoU loss function is used to replace the original SIoU, which effectively alleviates the impact of category imbalance on detection accuracy. In order to verify the performance of the algorithm, a special farmed animal image dataset was constructed and a comparative experiment was conducted. Experimental data show that the improved SPF-YOLOv7-tiny algorithm achieved a recognition accuracy of 92.6% on the self-built data set, and the detection speed reached 103 FPS, which is 2.5% higher than the baseline YOLOv7-tiny model mean Average Precision (mAP). Compared with Faster Region-based Convolutional Neural Network (R-CNN), YOLOX, YOLOv5s, NanoDet and Single Shot multiBox Detector (SSD), the SPF-YOLOv7-tiny algorithm has excellent detection performance. Although the detection speed is slightly lower than NanoDet, it can also meet the needs of real-time detection, providing strong technical support for real-time detection applications in actual farming scenarios.

***Keywords*:** target detection, YOLOv7-tiny, SAM, deep learning

**MSC:** 68T07

## Abbreviation

| | |
|---|---|
| UAV | Unmanned Aerial Vehicle |
| SAM | Segment Anything Model |
| Faster R-CNN | Region-based Convolutional Neural Network |
| SSD | Single Shot multiBox Detector |


DOI: https://doi.org/10.37256/cm.6520257392

| | |
|---|---|
| YOLO | You Only Look Once |
| RS | Remote Sensing |
| FPS | Frames Per Second |
| DCGAN | Deep Convolutional Generative Adversarial Networks |
| mAP | mean Average Precision |
| ELAN | Extended Latent Attention Network |
| MP | Max Pooling |
| CIoU | Complete Intersection over Union |
| DIoU | Distance Intersection over Union |
| TP | True Positive |
| FN | False Negative |
| PR | Precision-Recall |
| AP | Average Precision |

## 1. Introduction

Remote Sensing (RS) imagery, acquired through satellite, aerial, or UAV platforms, provides essential earth surface observation data. With the widespread adoption of UAV technology, remote sensing applications have progressively extended to agricultural and animal husbandry scenarios. The integration of UAV systems with advanced object detection models offers a promising approach for intelligent management in farming operations, potentially enhancing production efficiency and scalability, which are crucial for realizing smart farming systems. In smart farm management, it is relatively easy to use drones to obtain image information of farmed animals and crops, and use this information to monitor and analyze the growth of farmed animals and crops. However, due to the way drones collect images, the targets in the images have low resolution, small size, easy occlusion and Complex visual backdrops, which makes it harder to assess farming scenes. Especially when using existing target detection algorithms, detection failures and recognition inaccuracies frequently occur with small or partially obscured targets.

Recent advancements in artificial intelligence have significantly enhanced the performance of object detection techniques, particularly those utilizing deep learning frameworks. Among these advances, YOLOv4 [1] achieves an excellent balance between detection accuracy and computational efficiency, achieving top results on the widely recognized Microsoft Common Objects in Context (MS COCO) dataset, outperforming many contemporary methods in terms of both accuracy and real-time processing capabilities. The YOLOv7 [2] model proposed in 2022 has achieved better detection results in more complex scene detection tasks after incorporating more complex data enhancement technology and training strategies. These classic algorithm models have great advantages in many specific data sets, but in the detection scenes of animal husbandry, there are generally small target sizes, occlusions, complex backgrounds, etc., which greatly reduce the effectiveness of these detection algorithms. In view of the characteristics of animal husbandry scene images, researchers have proposed a series of improvement measures for classic algorithm models. Xiao et al. developed an enhanced model of improved YOLO, which is specifically used to identify duck flock behavior under different lighting conditions [3]. Zheng et al. introduced a two-stage methodology combining object detection and classification networks for determining gender ratios in domestic ducks [4]. Li et al. implemented a point-supervised approach utilizing a fully convolutional network for chicken counting, achieving an accuracy of 93.84% with a processing speed of 9.27 Frames Per Second (FPS) [5]. The improved algorithm combined with high-performance computing resources has a significant effect on the single-category farmed animal target detection scenario, but as the farm scale expands and the number of target categories increases, the algorithm performance will be greatly reduced. The extensive use of high-performance computing resources will also increase the management cost of smart farms.

Contemporary research in UAV image processing demonstrates that deep neural network architectures have become the established framework for target detection tasks, but there are still some difficulties and challenges in real-time detection scenarios. Contemporary benchmark algorithms typically exhibit three fundamental limitations: substantial training data requirements, high parameter complexity, and compromised inference performance when deployed on edge

computing platforms. The target size in the farmed animal images collected by the drone is relatively small, and the target activities are frequent and easy to be occluded. These factors will have adverse effect on the detection effect of the target detection algorithm. To address these issues in livestock target detection, this study proposes an enhanced SPF-YOLOv7-tiny model, which systematically improves upon the lightweight YOLOv7-tiny architecture. The specific enhancements encompass three key innovations: Firstly, to tackle the challenges of species diversity and class imbalance in livestock environments, we introduce an improved Mosaic data augmentation method integrated with SAM [6] image segmentation, thereby enhancing dataset diversity and representativeness. Secondly, to overcome the difficulties posed by small target sizes and frequent occlusions, we integrate a dedicated small object detection head in the YOLOv7-tiny architecture, adopt an extended receptive field through higher resolution feature maps; finally, we replace the traditional SIoU with Focal_DIoU to enhance the recognition of difficult samples and solve the problem of imbalanced class distribution. The proposed SPF-YOLOv7-tiny was tested on the farmed animal image dataset built by this paper. The data showed that the optimization algorithm achieved good detection performance with lower model complexity, providing new ideas for the development of real-time monitoring systems in animal husbandry.

## 2. Related work

### 2.1 *Small object detection in UAV remote sensing images*

With the advancement of drone technology, agricultural production methods are undergoing a revolutionary change. Drone technology can provide high-resolution images of a large number of farms. These image data are crucial for research in the fields of agricultural monitoring, disaster assessment, and environmental protection. However, the images collected by drones usually have the characteristics of small target size, complex background, and complex background, which leads to poor results of some traditional detection methods and makes it difficult to deploy and apply them in real scenes. Therefore, contemporary research has established various experimental protocols to deal with these difficulties. They mainly include feature extraction and deep learning.

The researchers used simple rectangular features combined with integral image technology to quickly calculate eigenvalues, and combined the AdaBoost algorithm for feature selection and combination to build a strong classifier, owing to its exceptional computational efficiency and instantaneous processing capabilities, this methodology has become prevalent in preliminary facial recognition applications [7]. The Histogram of Oriented Gradients (HOG) descriptor operates by partitioning the input image into local cell regions, computing the distribution of gradient directions within each region, and aggregating these distributions into a comprehensive feature representation, which has shown remarkable results in human detection applications [8]. Such algorithms are limited by the quality of handcrafted features, and the calculation process is complicated, making it difficult to apply to large-scale data sets and complex scenes.

The emergence of deep learning has precipitated a paradigm shift in object detection methodologies, with CNN-based approaches progressively supplanting conventional computer vision techniques due to their superior feature learning capabilities. The R-CNN framework [9] introduced a novel two-stage detection paradigm by generating region proposals and extracting convolutional features, which are then classified using a Support Vector Machine (SVM) classifier. This groundbreaking architecture not only significantly improved detection accuracy, but also laid the conceptual foundation for much subsequent research in this area. The YOLO [1] family of algorithms is a completely different single-stage approach that formulates detection as a unified regression problem and achieves unprecedented inference speed through its end-to-end trainable architecture. While both methods perform well on standard benchmark datasets, their performance in small object detection scenarios and real-time applications with strict latency requirements highlights important directions for future algorithmic improvements.

### 2.2 *Latest progress in image enhancement technology*

Image enhancement technology is a crucial data preprocessing method in computer vision tasks. The training dataset is augmented by applying operations such as scaling and adding noise to the original images to improve the model's generalization ability. Classical image enhancement techniques primarily include geometric transformations,

color transformations, noise addition, mixing enhancements, and synthetic data generation. Study [10] proposed a reinforcement learning-based automatic data augmentation strategy search method, which can automatically search for optimal augmentation strategies from data. However, this method incurs high computational costs during the search process, requiring substantial computational resources. To enhance the practicality of augmentation strategies, research [11] introduced the RandAugment method, which simplifies the strategy search process of AutoAugment by reducing the search space. Additionally, research [12] proposed the CutMix method, which generates new training samples by randomly cropping and pasting regions of one image onto another. However, this approach may introduce irrelevant features into the generated images, potentially impairing the model's discriminative ability. In the application of deep learning features, research [13] introduced an interpolation-based data generation method, which enhances the diversity of data distribution to improve model generalization and reduce the risk of overfitting. Although these methods can increase data diversity and enhance the accuracy of target detection and model generalization to some extent, they typically require additional computational resources and time. Research [14] proposed an image generation optimization method based on Deep Convolutional Generative Adversarial Networks (DCGAN), combined with improvements to YOLOv4, significantly enhancing the success rate and detection speed of garbage detection. These studies demonstrate that combining advanced feature extraction and generation technologies can further improve the effectiveness of image enhancement and its application value in practical tasks.

## 2.3 *YOLO series algorithms*

The earliest YOLO model exhibited significant advantages in detection speed but had limitations in small object detection and bounding box localization accuracy [15]. With the rapid evolution of deep learning architectures and methodologies, the detection performance of the YOLO series of algorithms has been continuously improved by introducing new modules and optimization strategies, and has gradually been applied to various fields. The YOLOv2 architecture adopts several key innovations, including batch normalization layers for more stable training dynamics, support for higher resolution image inputs to preserve fine-grained visual information, and a refined loss function optimized specifically for small object detection. These improvements work together to improve the accuracy and robustness of the detection framework [16]. The YOLOv3 model further optimized the network architecture and introduced a multi-scale prediction mechanism to achieve real-time target detection at different scales [17]. Medina et al. used YOLOv5 to develop a fish disease monitoring system, which achieved efficient monitoring of goldfish images and real-time camera module images [18]. In 2022, the YOLOv7 model achieved high-precision and low-latency target detection in complex scenarios through deeper network architecture optimization, including structural improvement and strategy optimization. At the same time, the YOLOv7-tiny model with a smaller parameter scale was proposed. This model can still achieve good detection performance on devices with limited computing resources, which promotes the application of target detection algorithms in practical scenarios [3].

In specific target detection tasks, researchers have proposed a series of optimized versions by improving the modules in the YOLO series models. Yu et al. introduced an improved model for face mask recognition based on YOLOv4, incorporating an enhanced CSPDarkNet53 backbone network and Path Aggregation Network (PANet) structure, achieving a mean Average Precision (mAP) of 98.3% [19]. Xu et al. implemented a lightweight mask detection by integrating ShuffleNetV2 and Coordinate Attention into YOLOv5, achieving a high accuracy and speed [20]. For agricultural scenarios, researchers proposed the YOLOv7-CS model, a lightweight target detection model for bayberries, effectively addressing the challenge of high-density target detection in complex backgrounds [21]. The researchers used YOLOv7 and multiple data-enhanced camellia fruit detection methods to demonstrate excellent detection performance in complex field scenarios [22].

YOLO series algorithms have progressively overcome the limitations of early models in detection accuracy and small object detection through continuous optimization of network architectures and detection mechanisms. These improvements have not only significantly enhanced the overall performance of the models but also provided valuable technical references and practical experience for subsequent research.

### 2.4 *Research status of small object detection in different application fields*

Detection algorithms such as YOLO, Faster R-CNN [9], and SSD [23] have been extensively adopted for small object detection. Research [24] introduced a new feature fusion layer into YOLOv5 to optimize the shortcomings of algorithm in small object detection. This feature fusion layer, with a smaller receptive field, effectively captures minute details in the feature maps. Research [25] developed an attention-based feature fusion framework by selectively aggregating local salient features and global contextual features and optimizing cross-layer information integration. When deployed for traffic scene analysis, the proposed architecture achieves enhanced small object detection accuracy (measured in mAP) without compromising real-time processing requirements. YOLO-sea [26] is a specific model applied to maritime search and rescue missions. By optimizing the detection head and introducing a parameter-free attention mechanism, it achieves real-time detection of small-sized people and ships at sea.

Small object detection technology has made great progress in agriculture, remote sensing and rescue, and various improved models based on YOLO and RCNN series algorithms have been proposed. However, in many specific tasks, the detection accuracy of small object and low-latency real-time detection still cannot meet production needs.

## 3. Methodology

YOLOv7 uses a multi-branch stacking structure and introduces a novel downsampling structure that can perform upsampling and downsampling operations simultaneously, thereby achieving parallel compression and feature extraction. A specialized structure with a larger residual branch is designed to assist in model optimization and image feature extraction. At the same time, an adaptive multi-positive sample matching strategy is adopted to increase the number of positive samples by assigning multiple Prior Boxes to each Ground Truth Box for prediction. During training, the Prior Box that best matches each Ground Truth Box is determined based on the adjusted Prior Box prediction results. These innovative improvements give YOLOv7 better target detection capabilities. YOLOv7-tiny is optimized for edge computing deployment and implements a compressed network architecture that accelerates inference speed while maintaining compact memory requirements. This lightweight architecture is particularly effective in resource-constrained environments.

Figure 1 illustrates the architectural composition of the YOLOv7-tiny framework. The schematic reveals a tripartite structural organization comprising: (1) a feature extraction backbone that processes input imagery to generate discriminative feature representations, (2) an intermediate feature fusion neck that hierarchically combines and refines multi-scale features, (3) a detection head that performs the final localization and classification tasks. In light of the characteristics of UAV remote sensing images, we propose a series of improvements based on the lightweight YOLOv7-tiny network, aiming to enhance the model's detection accuracy for small and occluded targets while retaining its lightweight characteristics to ensure efficient operation on edge devices.

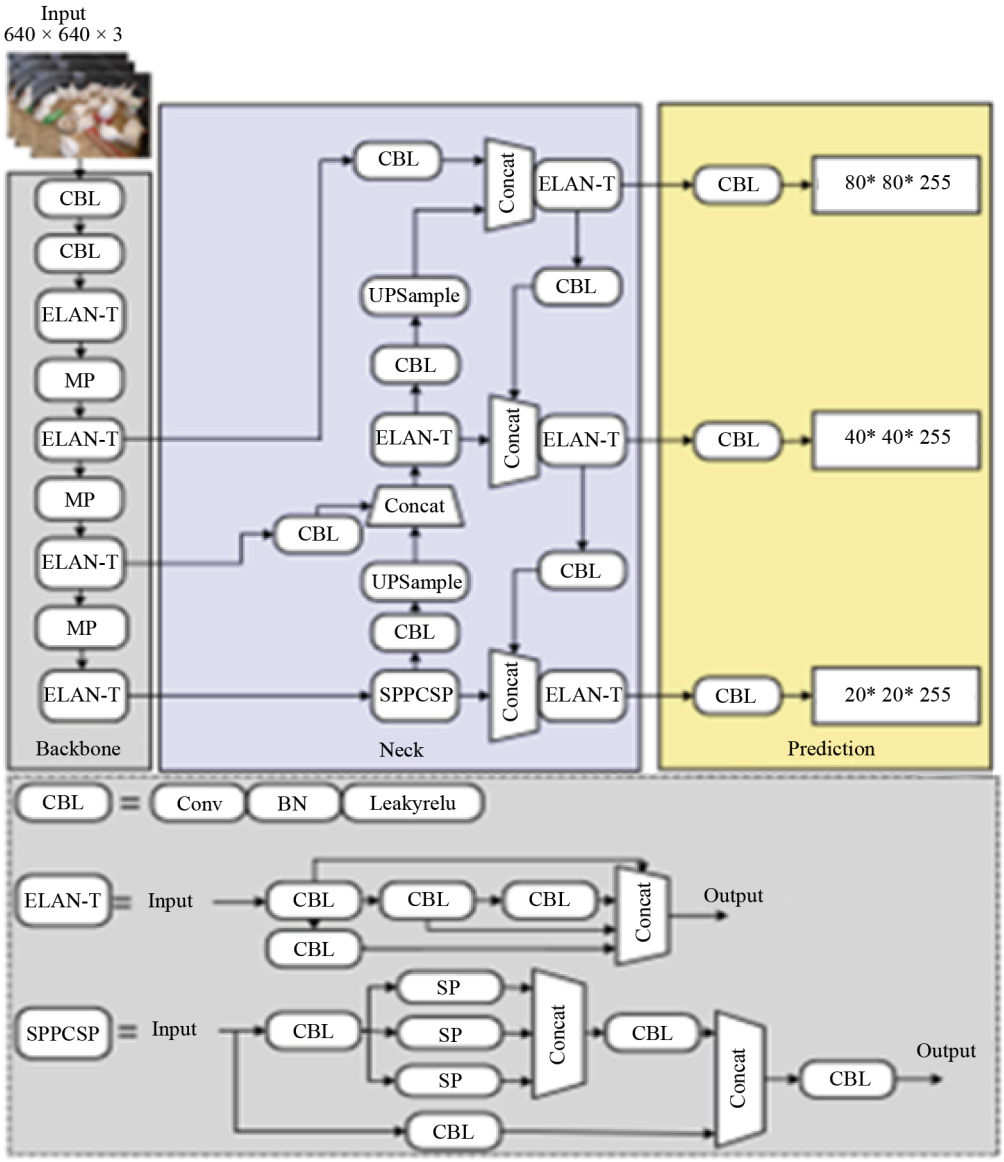


**Figure 1.** YOLOv7-tiny model structure diagram

## 3.1 *Mosaic image enhancement method based on SAM improvement*

The Mosaic data augmentation technique, employed by the YOLOv7 series models, is a widely used data augmentation method. Its core principle involves generating new training samples by randomly combining multiple images. During the model training process, four samples are randomly selected, and these images are randomly scaled, cropped, and arranged, and finally spliced together to form new training samples. This random combination method greatly expands the diversity of training data. By randomly scaling the images, the proportion of small objects in the training samples is increased. However, in actual farming scenarios, some target categories, due to their small size, are difficult for conventional training methods to fully learn their features, leading to class imbalance issues that affect the accuracy of the detection model. Traditional Mosaic augmentation methods use random cropping strategies, which may result in small or densely blurred targets being excessively cropped, causing training samples to contain only background information and reducing the model's learning effectiveness. Furthermore, due to inconsistent scales of the original

images, the stitched images often produce a large number of white border regions, and this irrelevant feature information can interfere with the model's training process, thereby affecting the model's convergence speed.

Based on these problems, this paper proposes a mosaic enhancement method based on SAM image segmentation. First, four images are randomly selected from the training data set, and after scaling and geometric transformation, they are spliced to obtain a four-grid spliced image. Then, a spliced image is randomly selected, and the silent SAM segmentation mode is applied to segment all the target masks in the image. Finally, 20% of the target mask is randomly selected and pasted onto another spliced image to obtain a new four-grid spliced image. An example of the enhancement process is shown in Figure 2.


**Figure 2.** Schematic diagram of improved Mosaic enhancement method

The improved Mosaic enhancement method is used to obtain new stitched images with more targets and a higher proportion of occluded targets. This effectively enriches the training dataset, increases data diversity, and alleviates the problem of class imbalance. This enables the model to learn smoother and more robust decision boundaries during training, thereby improving the generalization ability of the model. In addition, by integrating targets into training images, the probability of occlusion between targets can be increased, allowing the model to learn more diverse feature representations during training and reducing the possibility of model overfitting.

The Mosaic enhancement method based on SAM image segmentation proposed in this section significantly improves the diversity of the dataset and the target masking ratio through four-frame spliced images and randomized target mask porting. The method mitigates the category imbalance problem, while effectively suppressing the risk of overfitting by increasing the inter-target masking probability. Although the data enhancement strategy optimizes the training samples, the underlying structure of YOLOv7-tiny still suffers from insufficient feature extraction for small objects, and structural improvements are needed to enhance the accuracy of small object detection.

## 3.2 *Adding small object detection head*

The original YOLOv7-tiny target detection network contains three detection heads of different scales. When the input size is 640 × 640, the detection head can output feature maps of 80 × 80, 40 × 40 and 20 × 20, which are used to detect targets with pixel sizes greater than or equal to 8 × 8, 16 × 16 and 32 × 32, respectively. The feature fusion network mainly realizes multi-scale feature fusion through feature pyramid structure and path aggregation network. When detecting small object, due to the lack of lower-level feature fusion, it is difficult to fully utilize the detailed information of small targets in the image, which makes the recall rate and accuracy of small object detection relatively low, resulting in the easy loss of small targets during the detection process. The lowest detection layer of YOLOv7-tiny is the P3 layer. The feature map

of this layer only retains 1/8 of the original image information after downsampling. The information extraction of small targets is not sufficient, resulting in an increased probability of missed detection and false detection. Therefore, this study introduces an additional P2 detection layer in the YOLOv7-tiny network to make the features of small targets in the high-resolution layer clearer, so as to improve the accuracy of small object detection. First, under the same input conditions, the feature map size output by the detection layer is 160 × 160. The higher resolution can retain more original image information, especially the details of small object. Secondly, by fusion of lower-level features, finer-grained features are combined with higher-level features, so that the details of small object can be better captured; finally, the P2 layer undergoes fewer convolution and pooling operations, and the feature map retains more spatial information. By adding a detection head to the P2 layer, the feature expression ability of small object can be improved, so that the model can better distinguish the differences between small object.

Based on the original model, this paper adds a detection head with 4× downsampling capability, which can identify targets with a resolution as small as 4 × 4. After adding a new detection head, the feature fusion structure of the head part also needs to be further optimized. As shown in Figure 3, it is the feature fusion structure after adding the P2 detection head. Among them, C2, C3, C4 and C5 correspond to the feature maps extracted by the backbone network after 4×, 8×, 16× and 32× downsampling, respectively, while P2 is the detection layer added in this paper, and P3, P4 and P5 are the original detection layers of the model. The feature fusion structure after adding the P2 detection layer first performs an upsampling operation on the feature map C5 input from the backbone network, and then fuses it with the C4 and C3 feature maps in turn to generate the feature maps of F4 and F3. Then, the feature map of F3 continues to be upsampled and fused with the C2 feature map, and finally the detection result is output through the P2 layer. Finally, starting from P2, the feature map is first downsampled and then fused with the F3 feature map, and then the P3 detection layer outputs the result. Continue to downsample and fuse with the F4 and F5 feature maps, and finally output the result through the P4 and P5 layers.

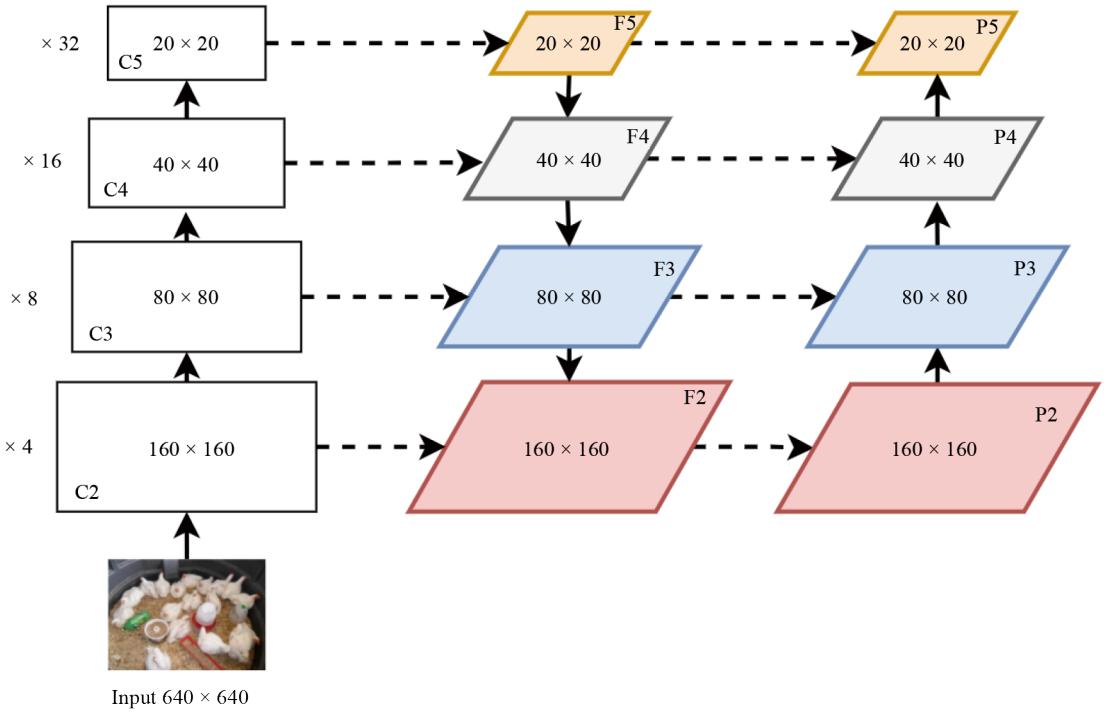


**Figure 3.** Adding a small target detection layer structure diagram

In this section, to address the insufficient performance of small target detection in YOLOv7-tiny networks, it is proposed to optimize the multiscale feature fusion structure by adding P2 detection layer, and introducing C2 feature

maps at lower layers, which achieves higher resolution feature retention and cross-layer fusion. At the same time, the geometric modeling of improved target detection puts higher requirements on the loss function. Although the original YOLOv7 adopts the CIoU loss function, which can comprehensively consider the target shape and spatial location, its complicated computation process is prone to reduce the training efficiency and gradient instability.

## 3.3 *Improved loss function*

The loss function of YOLOv7 adopts a multi-task learning framework, combining three basic components: bounding box regression loss for accurate localization, object prediction loss for foreground-background distinction, and categorical cross entropy loss for semantic classification. Among these, the localization loss is calculated using the Complete Intersection over Union (CIoU) as the regression loss function. It comprehensively considers factors such as target shape, spatial position, and orientation, enabling it to more accurately capture the geometric characteristics of targets, thereby contributing to improved accuracy of the target detection model. However, due to its involvement of more computational steps and multiple data inputs, the CIoU loss function increases the computational complexity during the detection process, prolongs training time, and is prone to causing gradient explosion issues. These problems may lead to increased localization deviations, thus affecting the detection effect.

Focal Loss introduces a regulation factor to make the model pay more attention to samples that are difficult to classify, effectively solving the problem of category imbalance. The Distance Intersection over Union (DIoU) loss function accelerates model convergence and improves positioning accuracy by considering the distance between the center of the predicted box and the center of the true box. The calculation process of Focal Loss and DIoU is shown in Equations (1) and (2). This study uses the FOCAL_DIoU loss function to replace the CIoU. The FOCAL_DIoU combines the advantages of Focal Loss and DIoU, and shows obvious advantages in dealing with category imbalance problems and improving positioning accuracy. The calculation process of FOCAL_DIoU is shown in Equation (3). By introducing FOCAL_DIoU, the stability and detection accuracy of the model are improved.

$$L_{\mathrm{DIOU}} = 1 - \mathrm{IoU} + \frac{\rho^2\left(b,\, b^{gt}\right)}{c^2} \tag{1}$$

$$FL(p_t) = -(1-p_t)^{\gamma}\log(p_t) \tag{2}$$

$$L_{\text{Focal-DIOU}} = -(1-\mathrm{IoU})^{\gamma}\log(\mathrm{IoU})L_{\mathrm{DIOU}} \tag{3}$$

Where: $t$ is the category balance factor; $\gamma$ is the focus factor; $p_t$ is the predicted probability of the target detection task; $\rho$ is the Euclidean distance between the center points of the two boxes; $c$ is the diagonal length of the minimum closure area containing the two boxes; IoU is the intersection over union ratio.

In this section, we propose to replace YOLOv7-tiny's CIoU with the FOCAL_DIoU composite loss function, which combines the difficult sample focusing mechanism of Focal Loss and the advantages of DIoU's centroid distance metric: on the one hand, it mitigates the problem of category imbalance through the conditioning factor, and on the other hand, it utilizes the distance constraints to accelerate the convergence of the model and enhance the localization accuracy. Experiments show that FOCAL_DIoU significantly improves model stability and detection accuracy while maintaining the ability of multi-scale feature fusion.

## 3.4 *Farmed animal object detection dataset*

The use of farmed animal target detection algorithms can timely detect abnormal situations in the breeding process, realize smart breeding and improve breeding efficiency. However, most of the current research focuses on manually processed high-definition images. These data can enable the model to obtain a high detection accuracy, but ignore the

impact of target size on detection performance in real scenes, resulting in the model being difficult to meet actual needs in actual breeding scenes. The actual breeding environment is often complex and noisy, which poses a huge challenge to the detection model. At present, there is little research on farmed animal target detection, and the detection model is only effective in single-category target detection. Compared with large-sized targets, small targets have sparse features, and there are fewer feature points or areas available for extraction, resulting in information loss during feature extraction. These are important challenges for small object detection.

**Figure 4.** Dataset image examples

This study constructed a dataset by mixing real farmed animal images with artificially processed high-definition images, including five common farmed animals, namely chicken, duck, cattle, rock pigeon, and sheep. In order to allow the model to learn more livestock target features, additional images of the same species were obtained from Google Images. Ultimately, these two data sources were combined to create a livestock dataset comprising 3,156 images, including 572 chickens, 448 ducks, 541 cows, 927 rock pigeons, and 668 sheep. As shown in Figure 4, this is an example of the image data set used in this article. The image categories shown in the example are duck, sheep, rock pigeon, cow, and chicken. In this study, small object refer to targets with a size less than 32 × 32 pixels, and large targets refer to targets with a size greater than 96 × 96 pixels, and the remaining sizes as medium targets. The distribution of target sizes across categories in the self-constructed livestock dataset is illustrated in Figure 5. As shown, blue indicates the number of small objects, orange indicates the number of medium-sized objects, and gray indicates the number of large-sized objects. Each category contains different numbers of large, medium and small objects. In all categories, the number of small objects, medium-sized objects and large-sized objects accounts for 37.5%, 57.5% and 5% of the total number of objects respectively. There is also a clear imbalance in the number of objects between different categories, which may significantly affect the performance of the object detection model.

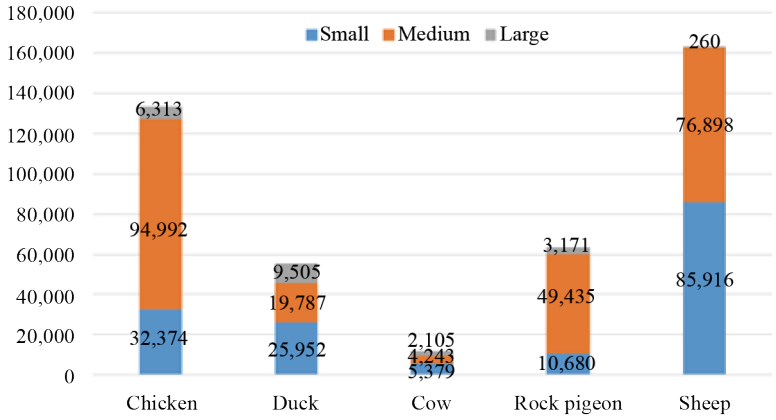


**Figure 5.** Distribution of the number of objects of different sizes in the dataset

# 4. Experiment and analysis

## 4.1 *Experimental data set processing*

In this study, a variety of data augmentation techniques were applied to the self-constructed dataset, including random cropping, random scaling, and random rotation, to enhance data diversity and robustness. From the augmented images, a random sampling process was employed to reduce the dataset size while ensuring a balanced ratio of positive to negative samples, thereby improving training efficiency and strengthening the model's generalization capability. The dataset was randomly partitioned into training, validation, and testing subsets using an 8 : 1 : 1 ratio to ensure balanced data distribution across all phases of model development.

## 4.2 *Experimental setup and parameter configuration*

The experimental environment for the experiments conducted in this study is presented in Table 1. This implementation uses the Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.01, a weight decay of 0.005, and 150 epochs. The batch size is set to 32, and the adaptive image scaling size is configured to 640 × 640. Three epochs are preheated. The mean Average Precision (mAP) and the number of model parameters (params) are used as evaluation metrics. mAP is calculated by the detection accuracy of each category, and the detection accuracy of each category is calculated by Precision and Recall. The calculation process is as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{5}$$

In the formula, True Positive ($TP$) means that the target is correctly classified as a positive example; False Positive ($FP$) means that the target is incorrectly classified as a positive example but is actually a negative example; False Negative ($FN$) means that the target is incorrectly classified as a negative example but is actually a positive example. Precision is defined as the proportion of samples correctly classified as positive examples among all samples classified as positive examples; recall is defined as the proportion of samples correctly classified as positive examples among all samples that are actually positive examples. Construct a Precision-Recall ($PR$) curve. The Average Precision ($AP$) is determined by the area under the $PR$ curve, and its calculation process is shown in formula (6). The dataset used in this paper contains

five categories, each category corresponds to a different $AP$ value, and the average precision ($mAP$) is calculated by taking the average $AP$ value of all categories, and the calculation process is shown in formula (7).

$$AP = \int_0^1 P(R)dR \tag{6}$$

$$mAP = \frac{\sum_{i=1}^{5} AP_i}{5} \tag{7}$$

The number of model parameters and Frames Per Second (FPS) play a crucial role in the feasibility of real-time detection. The number of parameters and FPS will directly affect the complexity of the model, the required storage space, and the speed of real-time detection.

**Table 1.** Software and hardware environment

| Environment configuration | Parameter |
|---|---|
| CPU | Xeon(R) Platinum 8362C |
| GPU | RTX 3090 (24 GB) |
| Operating system | ubuntu |
| RAM capacity | 45 G |
| CUDA | 11.8 |
| Framework | PyTorch 2.1.2 |

## 4.3 *Comparative experimental results and analysis*

### 4.3.1 *P2 detection head and improved mosaic enhancement*

The YOLOv7 series models employ the Mosaic data augmentation technique, which enhances the model's generalization capability by randomly combining multiple images to create new training samples. However, the traditional Mosaic method is prone to losing small targets during random cropping and often generates substantial white border regions, leading to the inclusion of irrelevant features in training samples and consequently impairing model performance. Therefore, this paper proposes a SAM-based mosaic enhancement method. This approach, after generating new training samples, utilizes SAM segmentation technology to randomly extract target regions from the composite images and paste them onto another randomly selected training image. This process enriches dataset diversity, mitigates class imbalance issues, and enhances the simulation of occlusion scenarios, ultimately improving the model's generalization and noise resistance capabilities.

Furthermore, to tackle the problem of small target loss in UAV aerial imagery, an additional small object detection layer is added to the YOLOv7-tiny network, called P2. When processing input images of size 640 × 640, the P2 layer outputs feature maps of size 160 × 160, enabling the detection of small objects as small as 4 × 4 pixels while expanding the receptive field. Adding this detection layer slightly increases memory usage, but it enables the model to extract more features of small objects, thereby improving performance in occlusion and small object detection tasks. The experimental results of adding the P2 detection head and the improved mosaic enhancement method are shown in Table 2, compared with YOLOv7-tiny, map is improved by 1.4% and 1.5% respectively, demonstrating that our proposed enhancement strategies significantly improve detection accuracy and robustness.

**Table 2.** Experimental results of adding small object detection head and improving mosaic enhancement

| Method | mAP (%) | Parameter |
|---|---|---|
| YOLOV7-Tiny | 90.1 | 6.02 M |
| +P2 | 91.5 | 6.12 M |
| +Improved mosaic | 91.6 | 6.02 M |

### 4.3.2 *Comparative analysis of optimized loss functions*

The YOLOv7-Tiny model conventionally employs CioU as its loss function. While CIoU effectively enhances localization accuracy in object detection tasks, it suffers from high computational complexity and is prone to gradient explosion issues, particularly with small bounding boxes. To address these limitations, we propose the adoption of Focal_DIoU as an alternative loss function, which resolves class imbalance issues, improves localization accuracy, and accelerates model convergence. Focal_DIoU is designed to synergistically combine sample reweighting from Focal Loss with geometric alignment from DIoU. Through an adaptive weighting mechanism that dynamically adjusts hard example emphasis and imposes centroid displacement penalties, this composite loss function simultaneously mitigates class imbalance and enhances bounding box regression accuracy.

We demonstrate the effectiveness of Focal_DIoU by training on the YOLOv7-Tiny model using various common loss functions and comparing the mAP of each set of experiments. The experimental results demonstrate that Focal_DIoU achieves superior detection accuracy in livestock detection tasks compared to other loss functions, as detailed in Table 3. It can be seen that when using Focal_DIoU as the loss function of YOLOv7-tiny, the mAP of the model is improved by 1.2% compared with using Ciou, and is higher than the mAP when using other loss functions. At the same time, it can be found that the accuracy of Ciou, Siou and Giou has been improved to a certain extent after combining FOCAL, which proves the effectiveness of FOCAL loss function in class imbalanced samples.

**Table 3.** Experimental results of improved loss function

| Method | Method | mAP (%) |
|---|---|---|
| YOLOv7-tiny | Ciou (default) | 90.1 |
| | Eiou | 89.5 |
| | Diou | 91.0 |
| | Giou | 89.9 |
| | Focal_Ciou | 90.5 |
| | Focal_Siou | 90.3 |
| | Focal_Giou | 90.1 |
| | Focal_Diou | 91.3 |

### 4.3.3 *Ablation experiment*

In order to test the effectiveness of the improvements in each part of the SPF-YOLOv7-tiny architecture we proposed, we used YOLOv7-tiny as the baseline model and conducted ablation experiments on the farmed animal dataset for each improvement. The designed experimental groups and results are shown in Table 4. Experiment 1 represents the original YOLOv7-tiny model, which is the baseline model of this paper, and its mAP is 90.1%, the inference speed is 127 FPS. Experiment 2 improves the mosaic image enhancement on the basis of Experiment 1, and the mAP is improved by 1.5%. Experiment 3 adds the P2 detection head on the basis of Experiment 1, and the mAP is improved by 1.4%. Experiment 4

replaces the loss function with FOCAL_Diou on the basis of Experiment 1, and the mAP is improved by 1.2%. Experiment 5 adds the P2 detection head on the basis of Experiment 2, and the mAP is improved by 0.2%. Experiment 6 replaces the loss function with FOCAL_Diou on the basis of Experiment 2, and the mAP is improved by 0.6%. Experiment 7 is the improved algorithm proposed in this paper. Based on Experiment 1, the mosaic image enhancement is improved, the P2 detection head is added, and the loss function is replaced with FOCAL_Diou. The mAP is improved by 2.5% compared with Experiment 1, and the inference speed of the model is 103 FPS. It can be seen that on the basis of the baseline model, the mAP of the model is improved to varying degrees after using the improved mosaic enhancement method, adding the P2 detection head, and using the FOCAL_Diou loss function, indicating that each improvement measure adopted in this paper has brought positive incentives to the accuracy of farmed animal target detection, further proving the effectiveness of the various methods proposed in this paper. Although the inference speed of the model is reduced from 107 FPS to 103 FPS, it still meets the demand of real-time detection.

**Table 4.** Ablation experiment results

| NO. | YOLOv7-tiny | Improved mosaic | P2 | FOCAL_Diou | mAP (%) | Parameter | FPS |
|---|---|---|---|---|---|---|---|
| 1 | √ | | | | 90.1 | 6.02 M | 127 |
| 2 | √ | √ | | | 91.6 | 6.12 M | 121 |
| 3 | √ | | √ | | 91.5 | 6.02 M | 110 |
| 4 | √ | | | √ | 91.3 | 6.02 M | 125 |
| 5 | √ | √ | √ | | 91.8 | 6.12 M | 105 |
| 6 | √ | √ | | √ | 92.2 | 6.02 M | 115 |
| 7 | √ | √ | √ | √ | 92.6 | 6.12 M | 103 |

### 4.3.4 *Performance comparison and analysis with other models*

In order to further verify the superiority of the algorithm proposed in this paper, we compared SPF-YOLOv7-tiny with YOLOv5, YOLOX, Faster R-CNN, NanoDet [28], and SSD under the same conditions. The results are shown in Table 5. It can be seen that the mAP of the SPF-YOLOv7-tiny proposed in this paper on the farmed animal detection dataset is higher than these comparison models, which are 16.8%, 3%, 2.7%, and 3.4% higher than the mAP of the comparison models, and 2.5% higher than the baseline model, reflecting the effectiveness of the SPF-YOLOv7-tiny model proposed in this paper. At the same time, in terms of detection speed, the FPS parameter of the SPF-YOLOv7-tiny model is also significantly higher than YOLOv5, YOLOX, Faster R-CNN, and SSD, reaching 103 FPS. Although it is lower than the baseline model and the dedicated drone detection model NanoDet, it can meet the needs of real-time detection and basically achieve a balance between detection accuracy and speed.

**Table 5.** Comparison of detection results with other models

| Methods | Parameters | mAP (%) | FPS |
|---|---|---|---|
| SSD | 26.28 M | 75.8 | 34 |
| YOLOv5s | 7.2 M | 89.6 | 92 |
| YOLOX [27] | 25.3 M | 89.9 | 95 |
| NanoDet [28] | 0.95 M | 81.2 | 198 |
| Faster RCNN | 138 M | 89.2 | 14 |
| YOLOv7-tiny | 6.02 M | 90.1 | 127 |
| SPF-YOLOv7-tiny | 6.12 M | 92.6 | 103 |


**Figure 6.** Comparison of detection results of SPF-YOLOV7-tiny and YOLOV7-tiny

In order to intuitively demonstrate the advantages of the improved YOLOv7-tiny model in the recognition of small objects of farmed animals, this paper compares the performance differences between the improved algorithm and the original YOLOv7-tiny algorithm in target detection. Different categories of images were selected for testing, and the detection results of the two algorithms were compared. The detection results are shown in Figure 6. The first column in the figure is the original image, the second column is the detection result of the baseline model YOLOv7-tiny, and the third column is the test result of the improved model SPF-YOLOv7-tiny in this paper. It can be clearly seen from the comparison

results in the figure that in the recognition of small objects in farmed animal images under complex backgrounds, the improved algorithm shows a better detection effect than the original YOLOv7-tiny model, and significantly reduces the common missed detection and false detection problems of the original YOLOv7-tiny model. For example, in the first row of images, the original YOLOv7-tiny model identifies the debris as chickens, and in the fifth row of images, the original YOLOv7-tiny model misjudges the bucket as a cow target, and the cow farther away in the upper left corner is not detected. The improved algorithm does not have such a situation. In addition, the average accuracy of the improved algorithm in target detection is generally higher than that of the original YOLOv7-tiny model. In summary, the improved algorithm SPF-YOLOv7-tiny proposed in this paper increases the receptive field and extracts richer context information. It shows strong anti-interference ability in the face of complex background information and effectively improves the missed detection and false detection of small objects.

## 5. Discussion

In this study, we aimed at the problem of low detection accuracy of small objects and occluded environments in farming scenes. Based on YOLOv7-tiny, we proposed the SPF-YOLOV7-tiny model and verified it, providing a reference for the application of drone technology in complex agricultural farming scenes.

In terms of data enhancement, although the traditional Mosaic method increases the diversity of training samples by randomly combining multiple images, its random cropping process easily loses small objects and introduces a large number of useless white boundary areas. To address this problem, we try to introduce SAM segmentation technology to extract the target area from the combined image and paste it back into other images. This method can effectively avoid the loss of small objects and increase the simulation of occlusion conditions. The generated new images can further improve the generalization ability of the model. Although the CutMix method also enhances data diversity through image splicing and mixing. However, while retaining the advantages of Mosaic, our method also improves the limitations of Mosaic and improves the detection accuracy of the model.

To address the problem of small objects being easily missed in drone aerial images, we added a P2 detection layer to the YOLOv7-tiny network to improve the accuracy of small objects detection by outputting larger feature maps. Compared with feature fusion methods such as FPN, the design of the P2 layer has lightweight features. Experiments show that the introduction of the P2 layer enables the model to achieve good results in small-size object detection.

Farmed animal datasets often have class imbalance problems. This paper attempts to replace the traditional CIoU with Focal_DIoU. Although CIoU also has very good performance in target detection tasks, its computational complexity is high and it is easy to cause gradient explosion problems when the prediction box is small. Focal_DIoU not only improves the class imbalance problem, but also further improves the positioning accuracy. This improvement makes the detection model more robust in complex farming scenarios.

These improved strategies show significant advantages in farmed animal detection tasks, but also have some limitations. Although the improved Mosaic enhancement method improves sample quality, it increases computational cost, which may affect training efficiency when processing large-scale data sets. The introduction of the P2 layer improves the small objects detection capability, but also increases the memory usage of the model, which may become a bottleneck on resource-constrained edge devices. Although Focal_DIoU optimizes positioning accuracy, its generalization ability on different data sets still needs further verification.

In summary, the method proposed in this paper provides a new idea for small-size target detection in UAV aerial images. These improvements not only provide technical support for real-time monitoring and intelligent management of farmed animals, but also provide references for other similar tasks (such as traffic monitoring, agricultural monitoring, etc.). However, how to further reduce computational cost and memory usage while maintaining performance is still a key issue that needs to be solved in the future.

## 6. Conclusions

This paper proposes a small objects detection algorithm for farmed animals based on image enhancement and improved YOLOv7-tiny, named SPF-YOLOv7-tiny. In order to improve the model's detection ability for small objects and occluded targets of farmed animals, a detection head dedicated to small objects is added to the YOLOv7-tiny model, and the Mosaic data enhancement method of the original model is improved. In addition, in order to solve the problem of class imbalance and reduce positioning deviation, the Focal_DIoU loss function is used to replace the SIoU in the original model. Experimental results show that the detection accuracy of the SPF-YOLOv7-tiny model in the farmed animal target detection task is significantly better than that of mainstream algorithms such as YOLOv5, Faster R-CNN and SSD. At the same time, the SPF-YOLOv7-tiny model only increases the number of parameters by 0.1 M compared to the baseline model, retains the lightweight feature of the baseline model, and has certain advantages when deployed on edge devices. When deploying the SPF-YOLOv7-tiny model in real farms or embedded systems, the problems of light variations, complex backgrounds, and occlusions that exist in farm environments may affect the robustness of model detection. Embedded devices have limited computational resources and memory, and despite the lightweight nature of the model, further quantization or pruning is required to reduce the computational overhead to ensure real-time detection requirements.

## Conflict of interest

The authors declare no competing financial interest.

## References


[1] Bochkovskiy A, Wang CY, Liao HYM. Yolov4: Optimal speed and accuracy of object detection. *arXiv:2004.10934*. 2020. Available from: https://arxiv.org/abs/2004.10934.

[2] Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada: IEEE; 2023. p.7464-7475. Available from: https://doi.org/10.1109/CVPR52729.2023.00721.

[3] Xiao D, Wang H, Liu Y, Li W, Li H. DHSW-YOLO: A duck flock daily behavior recognition model adaptable to bright and dark conditions. *Computers and Electronics in Agriculture*. 2024; 225: 109281. Available from: https://doi.org/10.1016/j.compag.2024.109281.

[4] Zheng X, Li F, Lin B, Xie D, Liu Y, Jiang K, et al. A two-stage method to detect the sex ratio of hemp ducks based on object detection and classification networks. *Animals*. 2022; 12(9): 1177. Available from: https://doi.org/10.3390/ani12091177.

[5] Cao L, Xiao Z, Liao X, Yao Y, Wu K, Mu J, et al. Automated chicken counting in surveillance camera environments based on the point supervision algorithm: LC-DenseFCN. *Agriculture*. 2021; 11(6): 493. Available from: https://doi.org/10.3390/agriculture11060493.

[6] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE; 2023. p.4015-4026. Available from: https://doi.org/10.1109/ICCV51070.2023.00371.

[7] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Kauai, HI, USA: IEEE; 2001. p.8-14. Available from: https://doi.org/10.1109/CVPR.2001.990517.

[8] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, CA, USA: IEEE; 2005. p.886-893. Available from: https://doi.org/10.1109/CVPR.2005.177.

[9] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2016; 39(6): 1137-1149. Available from: https://doi.org/10.1109/TPAMI.2016.2577031.

[10] Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV. Autoaugment: Learning augmentation policies from data. *arXiv:1805.09501*. 2018. Available from: https://doi.org/10.48550/arXiv.1805.09501.

[11] Cubuk ED, Zoph B, Shlens J, Le QV. Randaugment: Practical automated data augmentation with a reduced search space. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Seattle, WA, USA: IEEE; 2020. p.702-703. Available from: https://doi.org/10.1109/CVPRW50498.2020.00359.

[12] Yun S, Han D, Oh SJ, Chun S, Choe J, Yoo Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE; 2019. p.6023-6032. Available from: https://doi.org/10.1109/ICCV.2019.00612.

[13] Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. Mixup: Beyond empirical risk minimization. *arXiv:1710.09412*. 2017. Available from: https://doi.org/10.48550/arXiv.1710.09412.

[14] Fan J, Cui L, Fei S. Waste detection system based on data augmentation and YOLO_EC. *Sensors*. 2023; 23(7): 3646. Available from: https://doi.org/10.3390/s23073646.

[15] Jiang P, Ergu D, Liu F, Cai Y, Ma B. A review of Yolo algorithm developments. *Procedia Computer Science*. 2022; 199: 1066-1073. Available from: https://doi.org/10.1016/j.procs.2022.01.135.

[16] Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE; 2017. p.7263-7271. Available from: https://doi.org/10.1109/CVPR.2017.690.

[17] Zhang X, Dong X, Wei Q, Zhou K. Real-time object detection algorithm based on improved YOLOv3. *Journal of Electronic Imaging*. 2019; 28(5): 053022. Available from: https://doi.org/10.1117/1.JEI.28.5.053022.

[18] Medina JK, Tribiana PJP, Villaverde JF. Disease classification of Oranda goldfish using YOLO object detection algorithm. In: *2023 15th International Conference on Computer and Automation Engineering (ICCAE)*. Sydney, Australia: IEEE; 2023. p.249-254. Available from: https://doi.org/10.1109/ICCAE56788.2023.10111494.

[19] Yu J, Zhang W. Face mask wearing detection algorithm based on improved YOLO-v4. *Sensors*. 2021; 21(9): 3263. Available from: https://doi.org/10.3390/s21093263.

[20] Xu S, Guo Z, Liu Y, Fan J, Liu X. An improved lightweight yolov5 model based on attention mechanism for face mask detection. In: *International Conference on Artificial Neural Networks*. Cham: Springer Nature Switzerland; 2022. p.531-543. Available from: https://doi.org/10.1007/978-3-031-15934-3_44.

[21] Li S, Tao T, Zhang Y, Li M, Qu H. YOLO v7-CS: A YOLO v7-based model for lightweight bayberry target detection count. *Agronomy*. 2023; 13(12): 2952. Available from: https://doi.org/10.3390/agronomy13122952.

[22] Wu D, Jiang S, Zhao E, Liu Y, Zhu H, Wang W, et al. Detection of *Camellia oleifera* fruit in complex scenes by using YOLOv7 and data augmentation. *Applied Sciences*. 2022; 12(22): 11318. Available from: https://doi.org/10.3390/app122211318.

[23] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. Ssd: Single shot multibox detector. In: *Computer Vision–ECCV 2016*. Springer International Publishing; 2016. p.21-37. Available from: https://doi.org/10.1007/978-3-319-46448-0_2.

[24] Singh I, Munjal G. Improved Yolov5 for small target detection in aerial images. *SSRN*. 2022. Available from: http://dx.doi.org/10.2139/ssrn.4049533.

[25] Lian J, Yin Y, Li L, Wang Z, Zhou Y. Small object detection in traffic scenes based on attention feature fusion. *Sensors*. 2021; 21(9): 3031. Available from: https://doi.org/10.3390/s21093031.

[26] Zhao H, Zhang H, Zhao Y. Yolov7-sea: Object detection of maritime uav images based on improved Yolov7. In: *2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*. Waikoloa, HI, USA: IEEE; 2023. p.233-238. Available from: https://doi.org/10.1109/WACVW58289.2023.00029.

[27] Ge Z, Liu S, Wang F, Li Z, Sun J. Yolox: Exceeding yolo series in 2021. *arXiv:2107.08430*. 2021. Available from: https://doi.org/10.48550/arXiv.2107.08430.

[28] Miao D, Wang Y, Yang L. Foreign object detection method of conveyor belt based on improved Nanodet. *IEEE Access*. 2023; 11: 23046-23052. Available from: https://doi.org/10.1109/ACCESS.2023.32536244.