ISSN: 2414-1895 DOI: 10.6919/ICJE.202505_11(5).0045

Multimodal Bird Monitoring System Integrating Micro-Motion Detection and Gimbal Control YOLOv8 Optimization and Laser Deterrence System

Yaping Zhang[1], Jiayu Pan[2], Jiaxuan Zhang[3], Zixuan Li[4]

[1] College of Science, North China University of Science and Technology, Tangshan 063210, Hebei, China

[2] College of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, Hebei, China

[3] College of Neuropsychiatry, North China University of Science and Technology, Tangshan 063210, Hebei, China

[4] College of Economics and Management, North China University of Science and Technology, Tangshan 063210, Hebei, China

Abstract

To mitigate aviation safety risks, agricultural damage, and power outages caused by bird activities, this study introduces a bird recognition and dynamic deterrence system that combines You Only Look Once version 8 (YOLOv8) object detection, advanced micro-motion detection algorithms, and intelligent gimbal control. Traditional bird deterrent methods are often inefficient and lack adaptability, while current detection systems struggle with accuracy and real-time tracking of small, fast-moving targets. By integrating the Lucas-Kanade optical flow technique with the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm, this research develops a multiscale, adaptive micro-motion detection model that effectively captures subtle bird movements while minimizing noise interference. This model is further enhanced by YOLOv8, utilizing an upgraded Cross-Stage Partial Network to 2-Stage FPN (C2f) architecture, the TaskAlignedAssigner for more accurate sample allocation, and dynamic Mosaic data augmentation for improved detection precision and speed. In addition, a gimbal control algorithm based on spatial registration and adaptive zoom adjustment ensures precise target localization and laser deterrence, employing a two-degree-of-freedom coordinate transformation and a 30x optical zoom field-of-view fitting function.

Keywords

YOLOv8 Object Detection; Micro-Motion Detection; DBSCAN; Gimbal Camera Control Algorithm.

1. Introduction

1.1 Background

The accelerating process of urbanization and the continuous evolution of the ecological environment are causing a clear contradiction between bird activities and human production and life. At airports, bird strikes on airplanes are frequent, posing a serious threat to aviation safety and causing significant economic losses. In farmland, birds peck at crops, resulting in a large number of food production losses. In the field of electric power facilities, bird activities often lead to short-circuit lines and other faults, affecting the stability of power supply.

Conventional avian repellent methodologies, including sound repulsion and the establishment of physical barriers, are often characterized by inefficiency, limited adaptability, and substantial expense, thereby falling short of contemporary practical requirements. In recent years, the rapid development of computer vision and artificial intelligence technology has opened up a brand new technical way for the intelligent identification and repulsion of birds. Nevertheless, existing bird detection systems continue to grapple with significant technical challenges, including suboptimal detection accuracy and inadequate real-time performance when confronted with small and rapidly moving bird targets. Consequently, the development of an efficient and intelligent bird recognition and dynamic repellent system is of both significant academic research value and urgent practical application significance.

1.2 Status of Research

Presently, the automatic detection and identification of birds has attracted considerable interest both domestically and internationally. Driven by the rapid advancement of information technology, deep learning methodologies, and the development of Pan/Tilt/Zoom (PTZ) camera algorithms, the integration of these technologies in the field of bird identification is becoming increasingly evident, offering significant benefits. For example, Redmon Joseph et al. proposed the You Only Look Once (YOLO) algorithm, which had a significant impact on the You Only Look Once version 8 (YOLOv8) image recognition technique in this project[1]. James Bergstra discusses hyperparametric optimization methods that provide a theoretical basis for the optimization of micromotion logic models[2]. Karen Simonyan et al. present a study of deep convolutional neural networks that informs the design of convolutional networks in YOLOv8[3]. Ren Shaoqing et al. proposed Fast Region-based Convolutional Neural Networks (Faster R-CNN) as a region proposing network, which provides inspiration for the improvement of real-time detection ability of YOLOv8[4]. Alex Krizhevsky presents research on Deep Convolutional Neural Networks (AlexNet) to advance image classification and detection[5]. Liu Wei proposed a Single Shot MultiBox Detector (SSD) target detection method, which provides a reference for the multi-task detection method of YOLOv8[6]. Li Min et al. explored the application of target detection using YOLOv3 and SSD algorithms in different environments to enhance the understanding of bird detection in complex scenes[7]. Andreas Geiger et al. provided research on the application of deep learning to autonomous driving, inspiring dynamic object detection[8]. Bruce D. Lucas et al. proposed the Lucas-Kanade optical flow method, which provides the theoretical basis for the micromotion detection algorithm in this project[9]. Toby Sharp proposed a moving target tracking technique to support target identification and localization in laser bird repellent systems[10]. Alexey Bochkovskiy provides an in-depth analysis of the advantages of YOLOv4, and YOLOv8 as an improved version of YOLOv8 for real-time detection and accuracy improvement[11]. Seyed Mojtaba Marvasti-Zadeh explores the application of deep learning models for visual tracking and motion detection, shedding light on the dynamic detection of birds[12]. Applications and Challenges of Deep Learning Techniques in Wildlife Monitoring Presented by Devis Tuia et al. Provides Insights into Bird Identification and Repellent Systems[13].

The objective of this study is to develop an intelligent bird identification and repelling system that utilizes deep learning technology to achieve optimal efficiency. The integration of YOLOv8 image recognition technology and a laser repelling device enables real-time monitoring and precise identification of avian species. The system's capacity for real-time location of birds is enabled by the integration of a camera and a gimbal, facilitating precise repulsion through laser beams. This approach is designed to mitigate the potential hazards posed by birds to agricultural crops and flight safety. The paper further proposes an enhanced micromotion detection algorithm Density-Based Spatial Clustering of Applications with Noise (DBSCAN) that employs the optical flow method in conjunction with the optical flow method. This algorithm is optimized for image recognition and serves to mitigate the leakage of the YOLOv8 algorithm, thereby enhancing the system's recognition accuracy and robustness in dynamic environments. The system incorporates a gimbal and camera

Volume 11 Issue 5, 2025
DOI: 10.6919/ICJE.202505_11(5).0045

algorithm to regulate the velocity and inclination of camera rotation, facilitating the swift and precise location of birds and the emission of lasers. This system exhibits a high degree of intelligence and adaptability to diverse environments and avian species, thereby ensuring a wide range of applications.

1.3 Innovative Points and Development Ideas

In this paper, we propose a micromotion detection model based on the improved optical flow method with DBSCAN clustering algorithm, which is combined with YOLOv8 to achieve accurate recognition of birds and design an intelligent gimbal control algorithm for dynamically optimizing the monitoring viewing angle and focal length.

(1) While traditional optical flow methods struggle with detecting subtle movements, our enhanced approach integrates adaptive thresholding and multiscale analysis to significantly improve micro-motion capture capabilities for small targets.

(2) The detection framework employs DBSCAN clustering for post-processing optical flow results, effectively eliminating noise while maintaining detection robustness. This refined micro-motion detection mechanism synergizes with the YOLOv8 architecture through multi-level feature fusion, enabling real-time avian species identification with enhanced classification accuracy.

(3) A novel pan-tilt-zoom control system was developed using spatial coordinate registration and adaptive zoom algorithms. By implementing precise coordinate transformations and intelligent magnification strategies, the system ensures high-quality image acquisition across diverse environmental conditions, establishing a reliable foundation for subsequent recognition and deterrence operations.

Organization of the paper:

Section 2 outlines the experimental framework, covering YOLOv8 architecture optimization, micro-motion detection principles, and laser targeting mechanisms. Section 3 presents comparative analyses between baseline YOLOv8 and our micro-motion enhanced variant, along with performance evaluations of the original versus algorithm-enhanced PTZ camera systems in practical scenarios.

2. Introduction to Experimental Methods and Modeling Principles

2.1 Experimental Workflow Diagram

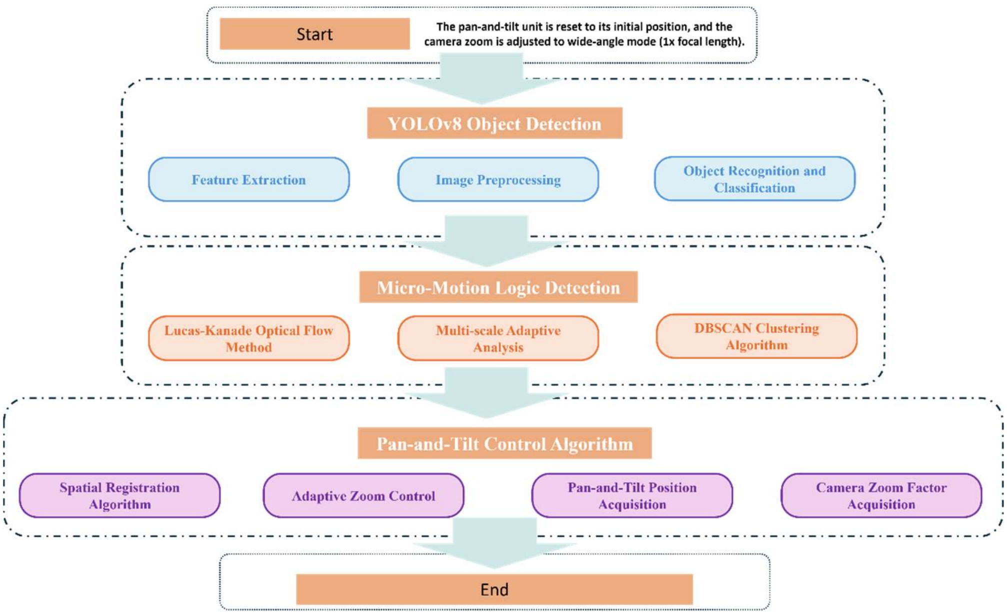

Figure 1. Experimental Workflow Diagram

2.2 YOLOv8 Object Detection Algorithm

The YOLO series has gained widespread popularity in the computer vision community for its speed, efficiency, and accuracy, making it a go-to choice for real-time object detection tasks. As the latest

Volume 11 Issue 5, 2025
DOI: 10.6919/ICJE.202505_11(5).0045

iteration, YOLOv8 builds on this foundation with notable performance improvements while preserving the simplicity and usability the series is known for.

YOLOv8 introduces several key architectural and training optimizations to boost detection accuracy. Drawing from the Efficient Layer Aggregation Networks (ELAN) design in YOLOv7, it replaces the Cross-Stage Partial Network (C3) module from YOLOv5 with the more gradient-rich Cross-Stage Partial Network to 2-Stage FPN (C2f) module, which improves feature extraction and enhances the backbone's representation capabilities. For loss calculation, it adopts the TaskAlignedAssigner for positive sample assignment and incorporates Distribution Focal Loss v5 to address class imbalance. These refinements lead to more stable training and faster convergence.

Data augmentation also sees improvements-YOLOv8 modifies the randomness and stitching method in Mosaic augmentation, increasing the density of objects within a single image. This significantly enhances performance on small and densely packed targets, making the model more robust across varied datasets. Training efficiency is further optimized through improved distributed strategies, including Gate Recurrent Unit (GRU) support, gradient synchronization, mixed-precision training, and mechanisms like early stopping and intelligent model checkpointing to reduce overfitting.

When applied to laser-based bird deterrence, YOLOv8's high-precision, multi-scale detection enables accurate identification of birds across different species and sizes. Combined with multi-object tracking, it allows for continuous monitoring and trajectory prediction, making it possible to dynamically locate and follow targets. This capability greatly improves control precision, enhancing both the intelligence and effectiveness of bird deterrence systems.

2.3 Micro-Motion Logic Detection Algorithm

Motion logic detection analyzes subtle movement patterns in video sequences to identify, track, and interpret targets' micro-scale displacements. By scrutinizing temporal-spatial variations in movement signatures, this technique amplifies and deciphers minute behavioral changes to infer operational logic and intent.

Combining computer vision, signal processing, and machine learning, the workflow involves three critical phases:

(1) Hyperparameter Optimization

Tuning parameters like learning rates and regularization coefficients enhances model convergence speed and generalization capacity during training.

(2) Architectural Refinement

Adjusting network depth, layer configurations, and channel dimensions balances computational efficiency with feature representation capabilities.

(3) Model Compression

Techniques including pruning, quantization, and knowledge distillation reduce parameter volume and computational overhead while maintaining inference speed.

This study examines three motion detection approaches: background subtraction, optical flow, and frame differencing. Our enhanced solution integrates optical flow with DBSCAN clustering for avian motion analysis.

(1) Background Subtraction

Establishes reference models to isolate foreground objects through frame-background comparison. While effective for complete target extraction, it requires frequent background updates due to sensitivity to environmental dynamics.

(2) Optical Flow

Calculates pixel-level motion vectors to capture micro-movements, excelling in non-rigid target tracking with detailed velocity/direction data. The proposed system upgrades Lucas-Kanade optical flow through adaptive parameterization.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202505_11(5).0045

(3) Frame Differencing

Detects moving edges via inter-frame pixel variation analysis. Though computationally lightweight, it fails to reconstruct full target silhouettes.

Building on these insights, our hybrid architecture combines optimized optical flow with clustering-based noise suppression, enabling precise avian motion detection across diverse scenarios.

2.4 Laser Pan-and-Tilt Control Algorithm

On the hardware control side, this study adopts a direct coupling approach between the motor and the gimbal shaft. Paired with a custom-tuned control algorithm, the system uses input from the vision module to achieve precise gimbal camera control. When a flying bird is detected, the control unit receives real-time positional and velocity data and adjusts the gimbal accordingly. The laser is then guided to aim just near the bird, creating an effective deterrence without physical contact.

The core function of the gimbal control algorithm is to manage both horizontal and vertical angles. Using the incoming parameters, it calculates the target position and rotates the gimbal to align. Once the position is reached, the system allows a brief stabilization phase, during which fine adjustments are made to smooth out any residual motion.

To enhance stability during this process, a micro-motion detection mechanism is introduced. Inspired by the principles of Proportional-Integral-Derivative (PID) control, the algorithm ensures that the gimbal and camera maintain a steady focus on the target area. During operation, micro-motion detection continuously compares the current and previous positions of the gimbal and camera. If a significant discrepancy is detected, it signals a possible disturbance—triggering actions like re-centering, zoom adjustment, or temporarily halting movement until stabilization is restored.

Once micro-motion analysis confirms a stable state, a "False" signal is sent to indicate the detection phase is complete. During all gimbal movements and adjustments, the system also communicates with the recognition module, using zoom control to keep the bird target clearly in view. This integration of gimbal control with micro-motion detection ensures consistent stability, enabling accurate target identification and reliable tracking throughout the process.

2.4.1 Spatial Registration Algorithm

The object target frame obtained by the target detection algorithm is the relative target frame established under the camera field of view. The establishment of camera facing coordinates and field of view range is an important guide for the conversion of the target frame obtained from target detection to gimbal movement and zoom.

For the registration of multiple targets, the system will calculate based on the images in different camera states and zoom in on the targets one by one in angular order to enhance the confidence of the targets through clear images. Since the camera moves with the two-degree-of-freedom gimbal and the camera field of view is determined by the camera magnification, we constructed a field of view model for the smart gimbal camera:

$$FOV = 2\arctan\left(\frac{S}{2M \cdot D}\right) \tag{1}$$

where FOV denotes the camera field of view, S is the width of the sensor, M is the camera magnification, and D is the distance from the camera to the object.

Camera orientation is expressed as an angle vector, in order to guide the movement of the gimbal, the system constructs a field of view determined by the angle range of the actual coordinate system, the camera orientation vector is the center vector of the field of view, and the change of the focal length maps out the change of the camera's field of view. The camera position is defined to be at the coordinate origin, and the composed image can be mapped on a spherical surface centered at the origin.

 Volume 11 Issue 5, 2025
DOI: 10.6919/ICJE.202505_11(5).0045

The mapping formula for camera pixels to spatial pixels in the camera coordinate system is as follows, with x representing different focal lengths:

$$x = f \cdot \tan(\theta) \tag{2}$$

If there is a detection frame in the target detection image, the system can register it as an angular range in space based on the coordinates of the camera coordinate system. Through this conversion, the orientation range of the stationary object relative to the coordinate origin can be obtained, thus providing guidance for the movement of the gimbal. Taking a 30x zoom camera as an example, the system fits a continuous field of view range function through the field of view ranges corresponding to different magnifications toward space, which provides data support for head doubling.

2.4.2 Pan-and-Tilt Position Acquisition Algorithm

After spatial registration during object detection, the system performs optical zoom on the target to capture a clearer image. This helps verify whether small, distant objects are actual targets or simply noise, and it also improves the clarity of local features for better analysis. Because the system processes short video streams for recognition, it can generate relatively stable bounding boxes. However, this also increases the complexity of handling those boxes.

To manage this, the system applies Non-Maximum Suppression (NMS) in world coordinates to filter out overlapping or redundant detections. It converts multiple bounding boxes into spatial target boxes, then computes a weighted average of their coordinates to determine the central direction of the target. This calculated direction is then used to guide the gimbal's rotation, ensuring it stays locked onto the most probable target location.

2.4.3 Camera Zoom Factor Acquisition Algorithm

Using the two-degree-of-freedom field-of-view constraints derived from the spatial NMS algorithm, the system applies a continuous field-of-view fitting function to calculate the appropriate focal length for the camera based on the size of the target bounding box. This calculated value is then sent to the optical zoom module for execution.

The computation and execution of gimbal movement and camera zoom are carried out simultaneously. After spatial registration, the system progressively zooms in on each target, performing high-resolution inspection to eliminate false positives. Once confirmed, the bird deterrent mechanism is activated.

2.4.4 Adaptive Zoom Algorithm Based on Anchor Box Size

The object target frame obtained by the target detection algorithm is the relative target frame established under the camera field of view. The establishment of camera facing coordinates and field of view is of great significance in guiding the conversion of the target frame obtained from target detection to the world coordinate system, as well as the movement of the gimbal and the zoom of the camera. For the registration of multiple targets, the system will calculate according to the images under different camera states and zoom in on the targets one by one in angular order to enhance the confidence of the targets through clear images, thus improving the detection accuracy of bird targets.

The gimbal control algorithm consists of two main components: a spatial registration method that establishes a world coordinate system, and a zooming algorithm that adapts to the size of the target's bounding box. The video stream is captured by a camera mounted at the end of a two-degree-of-freedom gimbal. By applying rotation and translation to the camera's local coordinate frame, the system computes a six-degree-of-freedom representation of its pose relative to the gimbal's base.

This representation can be used to derive the camera's orientation, expressed as an angular vector （θ_h, θ_v）. In this context, θ_h corresponds to horizontal rotation, and θ_v to vertical rotation. The vector is defined in absolute coordinates-each point in 3D space has a unique orientation vector

Volume 11 Issue 5, 2025
DOI: 10.6919/ICJE.202505_11(5).0045

depending on its direction from the camera. The first joint of the gimbal controls horizontal movement, while the second joint adjusts the vertical angle.

The direction vector points to the center of the field of view. As the focal length changes, so does the angle of view, effectively altering what the camera sees. By setting the camera's position at the origin of the coordinate system, the field of view can be imagined as a projection onto a spherical surface centered at that point.

For a specific anchor box, the corresponding field of view can be described using the following formula:

$$FOV_h = 2\arctan(\frac{W}{2f}), FOV_V = 2\arctan(\frac{H}{2f}) \tag{3}$$

The gimbal control logic is divided into two key parts. One handles spatial registration using a world coordinate system. The other focuses on zoom control, adjusting focal length based on the size of the anchor box.

In this setup, FOV_h and FOV_V represent the horizontal and vertical field of view angles for the i-th anchor box. Meanwhile, $\theta_{h_{\min}}$、$\theta_{h_{\max}}$、$\theta_{v_{\min}}$、$\theta_{v_{\max}}$ contains the angular boundaries of that anchor box in the world coordinate system. Specifically, it includes the minimum and maximum horizontal angles, as well as the highest and lowest vertical angles associated with the target's orientation.

$$\begin{aligned}
\theta_{h_{\min}} &= \arctan(\frac{x_{\min} - W/2}{f}) \\
\theta_{h_{\max}} &= \arctan(\frac{x_{\max} - W/2}{f}) \\
\theta_{v_{\min}} &= \arctan(\frac{y_{\min} - H/2}{f}) \\
\theta_{v_{\max}} &= \arctan(\frac{y_{\max} - H/2}{f})
\end{aligned} \tag{4}$$

Here, $x_{\min}, x_{\max}, y_{\min}, y_{\max}$ represents the pixel coordinate boundaries of the i-th anchor box. It includes the minimum and maximum values along the x-axis, and the minimum and maximum values along the y-axis. W and H refer to the width and height of the image in the pixel coordinate system.

The center direction of this anchor box-used to calculate its orientation-can be derived using the following formula:

$$\begin{aligned}
\theta_h &= \frac{\theta_{h_{\min}} + \theta_{h_{\max}}}{2} \\
\theta_v &= \frac{\theta_{v_{\min}} + \theta_{v_{\max}}}{2}
\end{aligned} \tag{5}$$

In this context, θ_h and θ_v represent the horizontal and vertical angles of the i-th anchor box. θ_i refers to the directional coordinates of that anchor box in the world coordinate system.

Since the camera moves with the two-degree-of-freedom gimbal, its field of view depends on the zoom level. Both the horizontal and vertical viewing angles (FOV) are influenced by the focal length f and the physical dimensions of the camera sensor.

The horizontal FOV angle θ_h and the vertical FOV angle θ_v can be calculated using the following equations:

$$\begin{aligned} \theta_h &= 2\arctan(\frac{W}{2f}) \\ \theta_v &= 2\arctan(\frac{H}{2f}) \end{aligned} \tag{6}$$

Here, W and H represent the width and height of the camera sensor, while f is the focal length. As the focal length increases, the horizontal and vertical field-of-view angles, θ_h and θ_v, become narrower. Conversely, reducing the focal length results in a wider field of view.

If the initial focal length is f_0 and the current zoom level is m, then the updated focal length f can be expressed as:

$$f = f_0 \cdot m \tag{7}$$

The effect of the zoom magnification m on the field of view range can be expressed as:

$$\begin{aligned} \theta_h &= 2\arctan(\frac{W}{2f_0 \cdot m}) \\ \theta_v &= 2\arctan(\frac{H}{2f_0 \cdot m}) \end{aligned} \tag{8}$$

As the zoom level m increases, the focal length becomes larger, and the field of view angles decrease. Conversely, when the zoom level decreases, the focal length becomes smaller, and the field of view angles increase.

The system is equipped with a 30x optical zoom camera. By gradually adjusting from 1x to 30x, the system determines the camera’s field of view at different zoom levels, including both horizontal and vertical ranges. Using this data, the system fits continuous functions for the horizontal and vertical field-of-view ranges. These functions help determine the appropriate zoom level for different anchor boxes.

For each anchor box, the system calculates the required vertical and horizontal field-of-view ranges using the fitting function. This allows the system to identify the optimal zoom level for each box.

The system then converts the initially identified anchor boxes into orientation coordinates in the world coordinate system, sorting them by horizontal angle from low to high. Using the spatial registration algorithm and the adaptive zoom algorithm, it calculates the required gimbal movement angles and zoom levels for each anchor box.

Based on the horizontal angles of the anchor boxes, the system performs a second zoomed-in recognition to confirm whether there are birds at each anchor box's location. If birds are detected during the second recognition, the system will activate the laser to deter them. If no birds are found, the system moves to the next anchor box and repeats the process. This continues until all anchor boxes have been re-checked.

After all the anchor boxes are processed, the gimbal returns to its initial position, performing a random turn to complete the loop of the detection task.

3. Conclusion

In this study, a high-precision and strong robust bird intelligent identification and dynamic repellent system is constructed by integrating YOLOv8, micro-motion logic detection and gimbal control algorithms. The innovativeness is reflected in three aspects: 1) proposing a micromotion detection model based on the optical flow method and DBSCAN, which improves the detection accuracy of tiny moving targets through multi-scale analysis and noise suppression techniques; 2) optimizing the YOLOv8 network architecture and training strategy to improve the model accuracy and robustness while maintaining real-time; 3) designing a spatial registration and adaptive zoom gimbal control algorithm to achieve high accuracy and strong robustness through 30x optical zoom's continuous field of view fitting to achieve sub-angle level target localization accuracy. The experiments verify the excellent performance of the system in dynamic environments, and future research will explore the multi-sensor fusion and reinforcement learning strategies to further improve the adaptability of the system in extreme weather and night scenes.

Acknowledgments

This research was supported by a grant from the Innovative Entrepreneurship Training Program for College Students (Project No. X2024101), for which I would like to express my sincere thanks.

I have received help and support from many people during the research and writing of this project/dissertation, and I would like to express my sincere gratitude. I would like to thank all the members of the laboratory who provided me with a dynamic and creative research environment. They gave me a lot of valuable advice and help in experimental design and data analysis. I would like to thank the laboratory of North China University of Science and Technology for providing technical support and internship opportunities for this project, which enabled me to combine theoretical knowledge with practical applications and enhanced my understanding and application of professional knowledge.

References

[1] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[2] Bergstra J, Bengio Y. Random search for hyper-parameter optimization [J]. The journal of machine learning research, 2012, 13(1): 281-305.

[3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arxiv preprint arxiv:1409.1556, 2014.

[4] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[5] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

[6] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[7] Li M, Zhang Z, Lei L, et al. Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD[J]. Sensors, 2020, 20(17): 4938.

[8] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 3354-3361.

[9] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]//IJCAI'81: 7th international joint conference on Artificial intelligence. 1981, 2: 674-679.

[10] Sharp T, Keskin C, Robertson D, et al. Accurate, robust, and flexible real-time hand tracking[C]//Proceedings of the 33rd annual ACM conference on human factors in computing systems. 2015: 3633-3642.

[11] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arxiv preprint arxiv:2004.10934, 2020.

[12] Marvasti-Zadeh S M, Cheng L, Ghanei-Yakhdan H, et al. Deep learning for visual tracking: A comprehensive survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(5): 3943-3968.

[13] Tuia D, Kellenberger B, Beery S, et al. Perspectives in machine learning for wildlife conservation[J]. Nature communications, 2022, 13(1): 792.