IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

Smart Agriculture: Deep Learning Approach for Bird Detection on Crops in the Field

Trupti Rajput[1] and Trupti Sachin Bhase[2]
Department of Computer Engineering[1,2]
Siddhant College of Engineering Pune, Maharashtra, India

Abstract: *Bird invasions in agricultural regions result in significant crop damage, thereby diminishing food output and impacting farmers' income. Traditional deterrence techniques, such as scarecrows and auditory devices, sometimes prove ineffective owing to their restricted flexibility and absence of automation. Progress in deep learning and computer vision has facilitated the creation of sophisticated bird detection systems that can monitor and identify avian behaviors in real time. This article examines contemporary deep learning methodologies employed for avian identification and their significance in current smart agricultural practices. It emphasizes the accessible datasets, prevalent detection algorithms (including YOLO and Faster R-CNN), assessment measures, and IoT-based hardware integration for practical applications. Finally, the report examines existing problems and proposes future research avenues for developing dependable and cost-effective automated avian monitoring systems.*

Keywords: Visual Impairment, Raspberry Pi, CNN Algorithm, Computer Vision, Object Recognition

I. INTRODUCTION

Agriculture consistently faces obstacles from pests, phytopathogens, and avian disruptions. Birds, specifically, can inflict harm on crops during several growth stages, particularly in rice paddies and orchards. Traditional deterrence methods require significant manpower and are insufficient for large agricultural areas. Conversely, contemporary smart agriculture employs deep learning, IoT sensors, and edge computing to provide automated and efficient monitoring.

Deep learning-based avian identification systems enable ongoing surveillance using image or video feeds, precisely recognizing bird species and triggering suitable deterrent measures. This technique enhances crop protection while promoting precision and sustainable agriculture.

II. LITERATURE SURVEY

Birds Detection Techniques

Bird detection methodologies can be broadly categorized into several types, including CNN-based detection, region-based detection, single-shot detectors, and attention-based methods. Recent advancements in deep learning (DL) have significantly improved automated bird identification and classification, enabling accurate detection of both common and rare species under diverse environmental conditions.

CNN-Based Detection:

Convolutional Neural Networks (CNNs) automatically extract hierarchical features from images, such as edges, textures, and shapes, which are crucial for distinguishing bird species. CNN architectures like VGGNet, ResNet, and DenseNet are widely used for bird classification tasks. While highly effective in feature extraction, CNNs alone may struggle with detecting birds in complex or cluttered backgrounds.

Region-Based Detection:

Region-based Convolutional Neural Networks (R-CNN), Fast R-CNN, and Faster R-CNN focus on object detection by first proposing candidate regions (bounding boxes) and then classifying them. Faster R-CNN integrates region proposal networks (RPNs) into a CNN for faster and more accurate detection. This approach is effective for identifying multiple birds in a single image but can be computationally intensive.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

Single-Shot Detectors (SSD and YOLO):

Single-shot detectors, such as SSD and YOLO, predict bounding boxes and class probabilities in a single pass, making them suitable for real-time bird detection. YOLOv5 and YOLOv8 are especially efficient in detecting birds at various scales and in dynamic environments. These models provide a balance between speed and accuracy, enabling applications in drones and surveillance systems.

II. LITERATURE SURVEY

Ye, Y., Zhang, T., & Lu, R. (2024) [1] a Margin and Average Precision Loss Calibration method to address the long-tail item detection issue, characterized by certain categories having markedly fewer samples than others. The suggested method enhances model training by dynamically modifying class margins and improving precision-oriented loss functions, leading to equitable detection across all categories. Experimental findings demonstrate significant enhancements in the recognition of underrepresented classes, with performance exceeding current benchmarks. The approach exhibits adaptability in areas like remote sensing and autonomous systems, where data imbalance often arises. This research improves fairness and reliability in contemporary object detection systems.

Gao, X., Zhao, D.,et.al (2024) [2] The authors introduce YOLO-Parallel, an innovative deep learning model designed to enhance long-tail remote sensing object recognition. The framework implements a Positive Gradient Modeling approach to mitigate class imbalance and enhance feature learning for rare object categories. By parallelizing gradient updates, the model attains enhanced training stability and accelerated convergence. Experiments utilizing satellite imaging datasets demonstrate that YOLO-Parallel attains more accuracy than traditional YOLO variations. The system is especially adept in detecting small or unusual items in aerial imagery, hence enhancing the use of deep detectors in environmental monitoring and remote sensing applications.

Jocher, G., Chaurasia, A.,et.al (2023) [3] This reference offers extensive technical material for Ultralytics YOLOv8, a significant advancement in the YOLO series of realtime object detectors. YOLOv8 integrates advancements including decoupled detection heads, dynamic input scaling, and improved loss optimization, facilitating superior accuracy in classification, segmentation, and posture estimation applications. The model is optimized for deployment on both edge devices and cloud infrastructures, ensuring adaptability in real applications. Its open-source framework and modular architecture provide the seamless adaption to bespoke datasets. YOLOv8 represents a significant advancement in the evolution of adaptable and cohesive vision frameworks.

Pan, X., et al. (2022) [4] Pan and colleagues offer a hybrid vision system that amalgamates self-attention with convolutional processes to capture both global context and local spatial intricacies. This integration utilizes the strengths of convolutional networks in spatial representation and the ability of transformers to describe long-range relationships. Comprehensive evaluations on established benchmarks demonstrate constant performance improvements in both classification and detection tasks. The study highlights the significance of integrating CNN and transformer frameworks for thorough visual comprehension and has impacted later vision transformer architectures. The study successfully integrates the effectiveness of CNNs with the interpretability of transformers.

Tu, Z., et al. (2020) [5] This study introduces MaxViT (Multi-Axis Vision Transformer), a hybrid deep learning model that integrates convolutional layers with transformer blocks. MaxViT utilizes local and global attention processes to effectively collect multi-scale picture characteristics. The design attains cutting-edge performance in classification, segmentation, and detection, all while preserving computational economy.

MaxViT has greater versatility than pure CNN or transformer models and has proven beneficial in several domains, including autonomous navigation and environmental analysis. This effort established the groundwork for a new age of hybrid vision architectures that harmonize precision and efficiency.

Rajagopal, A., & Nirmala, V. (2021) [6] This study presents the Convolutional Gated MLP (gMLP) architecture, which combines convolutional processes with gated multilayer perceptrons to improve visual identification tasks. The hybrid architecture enhances spatial awareness and feature generalization beyond the capabilities of conventional CNNs or independent MLPs. The integration of convolutional layers maintains local spatial dependencies, but the gating technique enhances non-linear feature representation. Experimental findings on benchmark picture datasets provide

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, November 2025 Impact Factor: 7.67

high accuracy and processing efficiency. The suggested approach provides enhanced scalability and interpretability for complex visual input, laying the groundwork for efficient and high performing detection systems.

Li, M., Cheung, Y.-M.,et.al (2022) [7] Li and colleagues provide the Gaussian Clouded Logit Adjustment, a method designed to address class imbalance in long-tailed visual recognition tasks. The approach enhances classifier logits using Gaussian distribution modeling, facilitating improved calibration between majority (head) and minority (tail) classes during training. Empirical assessments on difficult vision datasets demonstrate substantial enhancements in accuracy and stability compared to previous long-tail methodologies. The approach interfaces effortlessly with deep object detection frameworks like YOLO and Faster R-CNN, enhancing equitable and consistent recognition of uncommon item categories.

Fujii, S., Akita, K., et.al(2021) [8] Fujii and collaborators introduce a bird detection framework designed to guarantee safe autonomous drone operation by averting bird-drone accidents. The research presents a customized dataset consisting of annotated avian photos obtained by drone-mounted cameras. The device employs computer vision models to detect birds in real time, enabling dynamic modifications to flight paths for improved aviation safety. The researchers underscore the importance of early identification and the development of lightweight algorithms appropriate for on-board processing. Experimental findings validate the model's dependability across diverse illumination and motion conditions, facilitating the progress of safer and more sophisticated aerial systems.

Kondo, Y., et al. (2023) [9] This document presents findings from the MVA2023 Small Object Detection Challenge, focused on the identification of birds in aerial photography.

The authors provide a novel benchmark dataset and assess various detection algorithms designed to recognize tiny, rapidly moving avian objects. Research demonstrates that traditional YOLO-based detectors exhibit suboptimal performance with small object sizes, highlighting the necessity for tailored feature extraction layers. The challenge outcomes delineate the highest-performing models and their parameter tuning methodologies. The dataset and insights encourage more study in wildlife monitoring and aviation safety, establishing a significant benchmark for assessing tiny object detection systems.

Sun, Z.-W., et al. (2024)[10] Sun and collaborators introduce FBD-SV-2024, an extensive dataset for Flying Bird Detection, designed for surveillance and monitoring purposes. The collection comprises many bird species recorded under different lighting, weather, and environmental situations, enabling study in bird identification and classification for ecological and aviation applications. Several advanced detectors, such as YOLOv8 and transformer-based devices, were evaluated, demonstrating notable enhancements in performance. The research highlights the significance of dataset diversity in enhancing model generalization and resilience. FBD-SV-2024 serves as a crucial resource for creating robust detection algorithms tailored for dynamic real-world contexts.

Table 1: Comparison table for the literature review

Title	Methodology	Algorithm Used	Limitations
Contributors, Y. You Only Look Once Version 5 (YOLOv5), 2021. [11]	Utilized a real-time object detection framework optimized for speed and accuracy in various environments.	YOLOv5	Struggles with detecting very small or overlapping objects; limited performance in low-light or cluttered agricultural conditions.
Li, C. et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications, arXiv, 2022.[12]	Introduced an optimized single-stage detection model tailored for industrial and high-speed visual tasks.	YOLOv6	Although efficient, it requires high computational resources and fine-tuning for outdoor agricultural applications.
Wang, C.-Y., Yeh, I.-H., Liao, H.-Y.M. YOLOv9: Learning What You Want to	Proposed a new gradient programming mechanism to enhance feature learning and	YOLOv9	Complexity in training and requires large, well-annotated datasets to achieve optimal accuracy.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

Learn Using Programmable Gradient Information, arXiv, 2024. [13]	detection precision.		
Wang, W. et al. Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction Without Convolutions, IEEE/CVF ICCV, 2021.[14]	Introduced a transformer-based architecture for dense visual prediction, eliminating convolutional layers.	Pyramid Vision Transformer (PVT)	Transformer models require large-scale data and computational power; slower inference compared to CNN-based models.
Zhang, C. et al. An Efficient Time-Domain End-to-End Single-Channel Bird Sound Separation Network, Animals, 2022.[15]	Developed an end-to-end deep learning network for isolating bird sounds in real-time audio streams.	Deep Sound Separation Network	Focuses on audio detection only; lacks integration with visual-based bird detection for field applications.

Gap Analysis

Most current studies focus on general-purpose object detection rather than specialized agricultural use cases, thereby reducing the efficiency of systems designed explicitly for identifying birds in crop protection and precision farming contexts [1, 2, 3, 9].

Many detection models struggle with the issue of long-tail data imbalance, as they are not effectively tuned to handle uneven sample distributions across diverse bird species or varying flight patterns [1, 2, 7].

A significant limitation of existing detection techniques is their inability to operate in real time, hindering their suitability for continuous monitoring over large agricultural or ecological landscapes [3, 8, 12].

Available bird detection datasets often lack sufficient variability in illumination, climatic conditions, and crop types, resulting in limited model adaptability to complex and changing outdoor environments [9, 10].

Although hybrid deep learning models that integrate CNN and Transformer architectures have emerged, their potential remains largely unexplored for detecting small, fast-moving objects such as birds in open-field agricultural scenarios [4, 5, 9].

Temporal and behavioral dynamics of bird activity are seldom incorporated, as most algorithms depend solely on static images instead of leveraging motion-based or sequence-aware analysis [2, 3, 13].

Traditional YOLO-based detection systems encounter challenges in identifying small or swiftly moving avian targets, which leads to diminished precision and recall in aerial image interpretation [3, 9, 12].

Research efforts are largely centered around visual data alone, with limited incorporation of multimodal information such as audio signals, radar sensing, or environmental parameters that could significantly improve detection reliability under real-world conditions [8, 10, 15].

The lack of standardized datasets and evaluation metrics for bird detection hampers consistency, reproducibility, and fair benchmarking across different research studies [9, 10, 12].

Advanced deep learning architectures such as YOLOv8 and MaxViT require substantial computational power, posing challenges for their integration into low-energy, real-time IoT or edge-based agricultural surveillance systems [3, 5, 12].

III. CHALLENGES IN BIRDS DETECTION

Birds often appear as small targets in wide-field or aerial imagery, making it difficult for detection models to accurately identify and localize them, especially at greater distances or under low-resolution conditions. Birds exhibit fast, erratic flight patterns, leading to motion blur and inconsistent object trajectories that hinder accurate frame-by-frame detection and tracking.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

Table II: Key Challenges in Birds Detection

Sr. No.	Challenges	References
1	Small object size and limited visibility in large agricultural or aerial imagery reduce detection accuracy.	[3], [9], [10]
2	Rapid and unpredictable flight motion causes motion blur and hinders continuous detection and tracking.	[2], [8], [9]
3	Environmental variability, including lighting, weather, and crop background, affects model robustness.	[8], [9], [10]
4	Data imbalance across bird species and flight patterns leads to biased learning and poor generalization.	[1], [2], [7]
5	Occlusion by vegetation and background similarity causes misclassification and false detections.	[4], [5], [9]
6	Limited availability of large, diverse, and annotated datasets restricts model scalability and adaptability.	[9], [10], [15]
7	High computational cost and latency make real-time deployment difficult on IoT and edge devices.	[3], [5], [12]
8	Lack of standardized benchmark datasets prevents fair comparison across detection models.	[9], [10], [12]
9	Integration complexity between deep learning models, IoT sensors, and automation systems hinders field deployment.	[8], [10], [12]
10	Energy inefficiency and high operational costs limit continuous surveillance in large-scale farming environments.	[3], [5], [12]

IV. APPLICATIONS OF MACHINE OR DEEP LEARNING TECHNIQUES ON DIFFERENT DATASETS

Machine learning and deep learning models have been widely employed in avian detection and monitoring systems, especially within agricultural and ecological sectors. These models are trained on varied picture and video datasets that encompass avian species, flying behaviors, and ecological fluctuations. Utilizing extensive datasets, algorithms may acquire intricate geographical and temporal characteristics that facilitate precise bird categorization, motion tracking, and behavioral analysis.

Table III: Overview of Machine OR DEEP Learning Applications on Different Datasets

Authors / Year	Dataset Used	Machine Learning / Deep Learning Technique
Ye, Y.; Zhang, T.; Lu, R. (2024) [1]	Long-Tail Object Detection Dataset	Margin and Average Precision Loss Calibration for long-tail detection
Gao, X.; Zhao, D.; Yuan, Z. (2024) [2]	Remote Sensing Object Dataset	YOLO-Parallel with Positive Gradient Modeling for imbalanced data
Jocher, G.; Chaurasia, A.; Qiu, J. (2023) [3]	Ultralytics YOLOv8 Dataset	YOLOv8 framework for real-time object detection
Pan, X.; Ge, C.; Lu, R.; et al. (2022) [4]	Vision Dataset (CVPR)	Integration of Self-Attention and Convolution for feature enhancement
Tu, Z.; Talebi, H.; Zhang, H.; et al. (2020) [6]	Image Classification Dataset	MaxViT: Multi-Axis Vision Transformer for dense prediction
Rajagopal, A.; Nirmala, V. (2021) [7]	Big Data Image Dataset	Convolutional Gated MLP combining Convolutions and gMLP
Li, M.; Cheung, Y.-M.; Lu, Y. (2022) [8]	Long-Tailed Visual Recognition Dataset	Gaussian Clouded Logit Adjustment for improved classification
Fujii, S.; Akita, K.; Ukita, N.	Bird Detection Dataset for	Distant Bird Detection for safe UAV flight

(2021) [8]	Drone Safety	
Kondo, Y.; Ukita, N.; Yamaguchi, T.; et al. (2023) [9]	MVA 2023 Small Object Detection Dataset	Bird spotting challenge dataset and detection results
Sun, Z.-W.; Hua, Z.-X.; Li, H.-C.; et al. (2024)[10]	FBDSV-2024 Dataset	Flying Bird Detection Dataset in Surveillance Videos
Contributors, Y. (2021) [11]	YOLOv5 Dataset	YOLOv5: Single-stage real-time object detection
Li, C.; Li, L.; Jiang, H.; et al. (2022) [12]	Industrial Object Detection Dataset	YOLOv6: Object detection framework for industrial and agricultural applications
Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. (2024) [13]	Object Detection Dataset	YOLOv9: Programmable Gradient-based Object Detection Framework
Wang, W.; Xie, E.; Li, X.; et al. (2021) [14]	Vision Transformer Dataset	Pyramid Vision Transformer (PVT) for dense prediction without convolutions
Zhang, C.; Chen, Y.; Hao, Z.; Gao, X. (2022) [15]	Bird Sound Dataset	Time-domain end-to-end single-channel bird sound separation network

V. TRENDS IN Birds DETECTION

State-of-the-art models such as YOLOv5, YOLOv8, and YOLOv9 have become the preferred choice for real-time bird detection, offering exceptional speed, precision, and the ability to identify small, rapidly moving objects.

Meanwhile, Vision Transformers (ViT), Pyramid Vision Transformers (PVT), and MaxViT have emerged as powerful architectures for capturing global contextual cues and enhancing detection performance in visually complex agricultural environments.

To handle class imbalance where certain bird species appear infrequently, methods like Margin Loss Calibration and Gaussian Clouded Logit Adjustment are employed.

Modern bird detection frameworks are increasingly leveraging IoT-enabled smart cameras and edge computing technologies to facilitate real-time observation and automated deterrent actions in the field.

Moreover, integrating visual data with acoustic inputs from bird sound separation networks strengthens detection accuracy in noisy, outdoor environments.

Lightweight architectures such as YOLOv6 and Convolutional Gated MLPs are being developed to ensure efficient operation on low-power and embedded agricultural hardware.

Recent advancements also focus on combining detection with motion tracking and behavioral analysis to gain deeper insights into bird activity and movement patterns.

Additionally, GAN-based data augmentation and synthetic dataset generation techniques are used to enhance model generalization, particularly for rare or unseen bird species.

Ultimately, modern intelligent bird monitoring systems integrate AI-driven detection, decision-making, and automated deterrent modules, creating autonomous, adaptive frameworks for sustainable crop protection.

VI. FUTURE RESEARCH DIRECTIONS IN BIRDS DETECTION

Design of unified, multi-species detection frameworks capable of performing consistently across varied and challenging environmental settings.

Optimization of algorithms to enhance detection accuracy for small, distant, and partially obscured birds within complex natural scenes.

Integration of multimodal learning approaches that combine visual and acoustic information to improve recognition precision and robustness.

Expansion of synthetic and augmented datasets to effectively mitigate data scarcity and class imbalance issues.

Development of lightweight edge-AI architectures tailored for real-time inference and deployment in field-based agricultural environments.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, November 2025
Impact Factor: 7.67

Implementation of adaptive loss calibration techniques to address long-tail distributions and underrepresented bird species in datasets.

Utilization of drone- and UAV-assisted monitoring systems for large-scale and remote bird activity surveillance.

Adoption of continual and transfer learning strategies to enable models to adapt dynamically to changing environmental and seasonal conditions.

Incorporation of environmental and climatic parameters into detection pipelines to enhance system reliability and contextual awareness.

Development of explainable and transparent AI frameworks to increase model interpretability, trust, and practical usability among end-users.

VII. CONCLUSION

In conclusion, the Deep Learning-Based Framework for Real-Time Bird Detection on Crops effectively illustrates the application of artificial intelligence and computer vision in addressing practical agricultural issues. The YOLO deep learning model enables the system to precisely detect and monitor birds in real time, assisting farmers in mitigating crop damage and decreasing manual monitoring requirements. The study underscores the efficacy of automation and deep learning in enhancing agricultural output, providing a dependable, scalable, and effective answer for contemporary smart farming methodologies.

REFERENCES

[1]. Ye, Y.; Zhang, T.; Lu, R. Margin and Average Precision Loss Calibration for Long- ail Object Detection. In Proceedings of the 2024 9th International Conference on Computer and Communication Systems (ICCCS), Xi'an, China, 19–22 April 2024; pp. 26–32.

[2]. Gao, X.; Zhao, D.; Yuan, Z. YOLO-Parallel: Positive Gradient Modeling for Long-Tail Remote Sensing Object Detection. In Proceedings of the IEEE Geoscience Remote Sensing Letters, Athens, Greece, 7–12 July 2024.

[3]. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8. 2023. Available online: https://docs.ultralytics.com/zh/models/yolov8/ (accessed on 8 November 2022).

[4]. Pan, X.; Ge, C.; Lu, R.; Song, S.; Chen, G.; Huang, Z.; Huang, G. On the integration of self-attention and convolution. In Proceedings of the IEEE/CVF Conference On Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 815–825.

[5]. Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; Li, Y. Maxvit: Multiaxis vision transformer. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 459–479.

[6]. Rajagopal, A.; Nirmala, V. Convolutional Gated MLP: Combining Convolutions and gMLP. In Proceedings of the International Conference on Big Data, Machine Learning, and Applications, Kenitra, Morocco, 5–6 June 2021; pp. 721–735.

[7]. Li, M.; Cheung, Y.-m.; Lu, Y. Long-tailed visual recognition via gaussian clouded logit adjustment. In Proceedings of the IEEE/CVF Conference On Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 6929–6938.

[8]. Fujii, S.; Akita, K.; Ukita, N. Distant bird detection for safe drone flight and its dataset. In Proceedings of the 2021 17th International Conference on Machine Vision and Applications (MVA), Singapore, 20–22 February 2021; pp. 1–5.

[9]. Kondo, Y.; Ukita, N.; Yamaguchi, T.; Hou, H.-Y.; Shen, M.-Y.; Hsu, C.-C.; Huang, E.-M.; Huang, Y.-C.; Xia, Y.-C.; Wang, C.-Y. Mva2023 small object detection challenge for spotting birds: Dataset, methods, and results. In Proceedings of the 2023 18th International Conference on Machine Vision and Applications (MVA), Singapore, 10–12 March 2023; pp. 1–11.

[10]. Sun, Z.-W.; Hua, Z.-X.; Li, H.-C.; Qi, Z.-P.; Li, X.; Li, Y.; Zhang, J.-C. FBDSV- 2024: Flying Bird Object Detection Dataset in Surveillance Video. arXiv 2024, arXiv:00317.

[11]. Contributors, Y. You Only Look Once Version 5. 2021. Available online: https://github.com/ultralytics/yolov5/ (accessed on 7 October 2022).

[12]. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W. YOLOv6: A single-stage object detection framework for industrial applications. arXiv 2022, arXiv:02976.

[13]. Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. Yolov9: Learning what you want to learn using programmable gradient information. arXiv 2024, arXiv:13616.

[14]. Wang, W.; Xie, E.; Li, X.; Fan, D.-P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 568–578.

[15]. Zhang, C.; Chen, Y.; Hao, Z.; Gao, X. An efficient time-domain end-to-end singlechannel bird sound separation network. Animals 2022, 12, 3117.