International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 142-157

https://doi.org/10.70917/ijcisim-2026-1951
Article

Deep Learning based Framework for Real-Time Bird Detection on Jowar Crop in Real Time Environment

Nupur Pathrikar[1], Dr. Deepa Deshpande[2]

[1] Research Scholar, Jawaharlal Nehru Engineering College (JNEC), MGM University, Chhatrapati Sambhajinagar, Maharashtra, India. vhanmante@mgmu.ac.in
[2] Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Engineering College (JNEC), MGM University, Chhatrapati Sambhajinagar, Maharashtra, India. ddeshpande@mgmu.ac.in

Abstract: -Bird predation in agricultural farms, particularly in crops such as Jowar (Sorghum) causes high losses in yields, which is a great problem to farmers. Old methods of scaring have become useless, and the surveillance by hand cannot be scaled or real-time. The increasing cases of birds strike and illegal drones in restricted airspaces are threatening the aviation safety and the equilibrium of the ecosystems. To address these difficulties, this study suggests an Internet of Things (IoT) and Deep Learning system of real-time monitoring and detection of birds in changing outdoor conditions. To guarantee the correct identification of birds and other flying objects, this framework also uses lightweight and powerful object detection models YOLOv11n, YOLOv11L,YOLOv8n, and YOLOv8L with a proposed module of Faster R-CNN. The sensors provide the system with field data which is streamed in real time to a cloud deep learning pipeline. Both versions of YOLO bring something to the model in the size, inference speed, and ability to be deployed to the edge. To obtain enhanced accuracy, specific change added emphasis loss by means of stratified sampling method to long-tailed distribution of class and high-resolution region suggestions in response to herding the key points of focus yaw. CNN based YOLO was purposely created to identify dynamic non-stationary objects in our observed world. It is a dynamic IO sensor networks which is anchored using static sensors and that data is sampled and stored as well as transmitted to smart moving nodes having ultra-light flow sensor system. These models had all been trained and tested with our dataset in varying conditions of the environment such as foggy views and our changing and altering lighting and skies filled with moving and stationary objects. The experimental findings in general showed that the single stage YOLO models had a vastly lower detection accuracy in comparison with the proposed Faster R-CNN model. The suggested model had a total real-time field test accuracy of about 98.01%. Simultaneously, both the inference speeds of YOLOv11L and YOLOv8L are competitive and suitable to edge IoT devices. The model is best when used in real-time such as when tracking airports, wildlife and airspace control due to trade-offs between detection latency and accuracy. The current work introduces an effective and versatile architecture of smart bird detection using IoT and deep learning-based methods making a significant contribution to proactive mitigation of the air threats.
Keywords: —IoT, Deep Learning, Bird Detection, Faster R-CNN, YOLO, Jowar Crop

1. Introduction

1.1background And Motivation

Limiting the number of bird strikes and unlicensed drone invasion has become an urgent issue because of the increased safety standards in the airline industry. Bird strike threat has been sharp since it has operational and economic effects which have cost burdened aviation industry. The incidents which have led to damages in the range of hundreds of millions have been thousands in number- You can for reference. On the same note, the growth in the recreational and commercial drone operations has created new air hazards that need prompt identification and categorization. In the immediate area and within littered conditions, ancient radar systems, and aged systems have

frustrating challenges in isolating small and maneuverable birds and drones to be imaged and elevated. The above problems have redirected efforts of researchers in the direction of better bird and drone detection systems, or computer vision and deep learning methods. Transformer-based architecture and convolutional neural networks (CNNs) The detection of small and fast objects has been proved with the help of model detection (especially model-based detection). Accelerated inference and competitive accuracy of various rebuttals of YOLO such as YOLOv5, YOLOv7 and 8 enable it to be used in real-time monitoring.

However, one of the major issues is the long-tailed object detection where some classes (such as species of birds or rare models of UAVs) have a relatively small number of instances in the dataset, thereby having poorer detection performance. Such an imbalance is typical with the ecological monitoring systems due to the skewed abundance of the species. Transformer-based object detectors, such as Deformable DETR and Swin Transformer have attempted to account for certain features of robust detection under class imbalance and achieve better detection results even with severe class distribution shortcomings. Along with this, other emerging imaging methods are thermal and multi spectrum imaging which are applied in detecting objects on aircraft. These sensors are particularly handy in low-light environments, and in harsh weather that is a typical situation in the aviation industry. Deep learning on the use of such sensors improves the surveillance system functionality beyond visual object detection. Under the conditions of converging technological opportunities, the identified ecological requirements, and the strict safety requirements, there is an immediate chance to create an intelligent monitoring system, which would involve the use of advanced real-time bird and drone detection technologies, address the issues associated with the long-tailed distribution of classes, and apply advanced recognition algorithms. To address these needs, this study suggests a single architecture that will use variants of YOLO, attention layers, and imbalanced dataset treatment to detect airports and flight paths quickly and precisely. The aviation sector remains exposed to the dangers of bird strikes and drones flown without being monitored, which may be severe injury, financial loss, and even death. The existing monitoring systems are not good enough in terms of the ability to detect and categorize small aerial objects because of the limitations of real-time and environmental factors. Most importantly, there is reduced detection accuracy in the case of long-tailed datasets where some species of birds or the type of drones are uncommon in the training data. Also, the issue of co-scheduling of birds and drones in the same proximity of airspace makes it unique and as such, it needs advanced feature extraction and advanced discriminative learning technologies to classify. Even though more modern object detectors, including YOLO and transformer-based detectors, have achieved considerable progress, they are not good at generalizing due to extreme class skew or other low-visibility situations, like fog or night. Therefore, the current deficit in the supply of a detailed, real-time detection system that specifically targets the intricacies of the aerial monitoring setting with a high population of underrepresented classes is critical.

2. Scope And Limitations

The spatial monitoring of the high-stakes surveillance of the aviation areas of the airport boundary, the runway and airspace corridors is the zoning of this research. The proposed system is a hybrid with traditional thermal sensors and RGB imaging in a bid to increase precision in reliability in a situation of limited illumination. Model training to detection architecture is based on the deep learning with real-time optimized models such as YOLO and Faster RCNN as the main backbone models.The scope of work is concerned with dealing with algorithms utilized to mitigate the imbalance between classes in long-tailed datasets using Focal Loss, Class-Balanced Sampling, or feature reweighting. The detection framework will be trained and assessed using publicly available datasets and also on the custom annotated datasets of interest, such as avian fauna of the area and commercially used drones. Although the goals are met, there are acceptable restrictions. To begin with, publicly available datasets produce a model that can be too specific and, thus, can have little generalization to new species or UAVs. Secondly, the nature of the use case of such systems means they are subject to real-time pressure, and so an edge computing hardware is necessary; the power and computing resources to support such a strategy are outside the scope of this document. Thirdly, the issue of separating intersecting or hidden objects, particularly when dealing with the cases of flocks or swarm motions, is an open problem, not only in object detection, which is addressed, at best, in this paper. Furthermore, the research is based on the detection and first-order classification and does not address prediction of behaviours and examination of trajectories that would allow improved avoidance. The objective of the future study will shift to detection on to regression models and predictive analytics.

2.1 Organization of the Paper

The present structure of the paper is the following: Section 2 gives an am literature review which gives a detailed overview of the research already done in the area of detecting birds and drones with the assistance of artificial intelligence technology such as machine learning and computer vision. Approaches that are built on YOLO

networks, transformer-based frameworks, and some of the long-tailed object detection issues are discussed. It also provides an analysis of the involvement of thermal and multispectral imaging and their significance in respect to the aviation safety effect of such detection systems. The section 3 explains the proposed detection framework that will contain the data cleansing processes, the model design, the loss functions, and the strategies of class rebalancing. The combination of different YOLO modules and Faster RCNN, its speed and its accuracy is paid special attention. Section 4 talks about the experimental setup, datasets that are used to train and test the models are described. Such evaluation measures as mAP, precision, recall, and F1-score are proposed. Training configuration as well as hardware specification are also described. Lastly section 5 gives the discussion of the detection performance which involves comparison of the models and datasets. The results are given under various ambient conditions that are daytime, night-time and fog. Long-tailed distribution mitigation strategies also have effects that are analysed.

3. Literature Survey

This paragraph defines the enhancement of the object recognition and IoT technologies have made it possible to do the monitoring of birds in a more efficient way. Even then, the majority of them are not as responsive and robust to various environmental conditions in real-time. The combination of deep learning models such as YOLO and Faster R-CNN and IoT sensors will offer immense possibilities of precise bird detection and wildlife protection.

4. Existing Bird Detection Techniques

Bird detection plays an important role in aviation surveillance and ecology as well as biodiversity conservation. The radar systems and manual checks which are currently in place are challenged in coverage and accuracy. The vision based machine learning has cameras in the machine learning system that increase the area of detection within the city and in the wild. Previous researches applied classical computer vision to background subtraction, contouring and motion tracking which were performed using video sequences. These methods were subjected to strong yet deterministically regulated changes in illumination, occlusion and dynamic changes in the environment. Such statistical or trained classifiers as SVMs and Random Forests were also implemented but experienced the same challenges of robustness in highly dynamic and cluttered settings. Deep learning facilitates birds detection. Convolutional Neural Networks (CNNs) made significant improvements in feature representation and classification accuracy. To complete tasks related to classifying birds, the CNNs were used with ResNet, AlexNet, and MobileNet models. Nevertheless, they suffered long-tailed distribution with small object detection that is common in bird dataset. The object detection models of YOLO (You Only Look Once) were capable of processing in real-time and producing high-accuracy results.

An example of lightweight models, including YOLOv8n and YOLOv11n (in comparison to the previous variants of YOLOv3 and YOLOv8), is the real-time inference of edge devices with low computing power. The subsequent models YOLOv11L and YOLOv8L are also oriented to the process of the image of high resolution and will execute these tasks to a higher extent. These algorithms are competent in locating birds in a wide variety of locations including wetlands and the surroundings of airports. Instead, regional based detectors like Faster R-CNN achieve higher localization though at the expense of speed that do not perform well in responding to real time. They excel in processes that require precision such as ornithology or guarding of areas that are in need of protection. Its major shortcoming is its high computational needs that have led to dependence in under-reliance in real time field deployments. In order to strike a balance between the detection and lag, the precision-based computations are combined with real-time speed detection, which attracts attention to hybrid approaches involving Adobe YOLO with Faster R-CNN frameworks, which are widespread in literature. Transformer-based structures are of great promise in modelling contextual relations in the cluttered bird habitats. However, there are still such issues as difference in poses of birds, difficulty in scaling size, low-light conditions, camouflage on the background, and zoom change. To solve some of these issues it has been suggested that multi-spectral data and thermal imaging are integrated. Embedded IoT platforms serve as the additional edge AI in recent tendencies that enhance real-time detection. These systems are established in the remote and peri-urban airfields to reduce bird strike.

4.1 iot In Wildlife Monitoring

Application of IoT in wildlife monitoring has contributed greatly to the advancement of biodiversity tracking, poaching and ecological data analysis in the realms of the IoT frameworks. In real time IoT, IoT systems consist of a network of interconnected devices such as drones, camera traps, acoustic sensors, tags and collects environmental and behavioural data. Constant supervision in the monitoring of birds has been made easier by the use of camera traps and drones that have computer vision algorithms that can be equipped with the IoT and only require a few people to operate. Such systems eliminate the problem of a bird-aircraft collision and can be used to predict the

pattern of the migration of birds in the airports, national parks, or the migratory paths. Temperature, motion, and acoustic sensors, microcontrollers, edge computing devices (e.g. NVIDIA Jetson, Raspberry Pi), and communication networks LoRaWAN, Zigbee and 5G all are implemented to create an IoT ecosystem. These elements help in real time data gathering and localized inference that reduces the utilization of cloud resources. Deep learning cameras improve the detection efficiency of birds. Nodes with low power factor and YOLOv8n or 11n can work with frames directly, but only transfer the positions of detected birds to the server to reduce the bandwidth and energy expenditures. IoT geographic deployments can be further made more viable by use of solar-powered IoT devices. One of the reasons why IoT-enabled bird monitoring systems should be promoted is mitigation of bird strikes in aviation. These systems link up with the air traffic control systems and decision support systems to automatically issue warnings to flocking birds near runways. However, the ruggedized hardware, energy-saving, and fault tolerance become challenging when using IoT systems in the harsh outdoor environment. Edge AI should be integrated in the use cases that require latency. Sensing information, especially within designated ecological areas, creates other security and privacy issues. The remedies work involves the establishment of mesh IoT networks that have dynamic coverage in monitoring and self-repairing communication channels, and the integration of satellite data to monitor wide-range areas. The practical intervention in ecological studies is made possible by the combination of machine learning and IoT technologies in practice in real-time.

4.2 Deep Learning Models for Object Detection

Deep learning models have developed object detection to a great extent in natural settings, like forests or wetlands with complex backgrounds. Object detectors based on CNN are distinguished into one and two stage detectors. YOLO and SSD are one-stage detectors that are efficiency oriented. The improvements of the YOLOv3 to YOLOv8 on the bounding box regression, multi-scale detection, and anchor-less are impressive. Although the last to be mentioned, YOLOv11 has enhanced efficiency of the backbones with better FPS on embedded systems bitcoin mining. YOLOv8n and YOLOv11n support edge environments, and have almost-instantaneous output and low-friction computation workload. On the other hand, the more complex and powerful models, YOLOv8L and YOLOv11L, can be used to identify small birds at a long distance in high-resolution images. Detectors based on two stages, in particular Faster R-CNN are more precise because they initially propose object parts before classifying them to make region proposals. This model is quite precise at detecting birds even though it is slower due to its capacity to compensate occlusions, species discrimination and different poses. DETR and Swin Transformer are the most up-to-date transformer-based architectures to learn interdependencies of frames globally. These models are making video-based bird detection gain more attention. Synthetic and IoT frameworks installed with edge devices powered by a GPU are used to enhance real-time, in-field object detection. Combined with data augmentation and synthetic data generating techniques, they outperform other research in the areas and lead to better generalization across seasons and habitats. Even though these improvements are achieved, a few problems such as generalization to rare species, occlusions, and motion blur continue to harm performance. The cross-domain learning, semi-supervised labeling and continual learning frameworks address these issues as discussed in.

4.3 Gaps in Existing Research

Of note are the gaps that still exist especially in the application of deep learning and the Internet of Things in detecting birds. To illustrate the point, the majority of the existing models are tested on benchmark data that entirely disregards real-life field scenarios such as changes in the light illumination, occlusion, or a background full of distractors. Although these systems such as the YOLO and Faster R-CNN have been applied with considerable success in the real world, they are still ineffective in the fixed, changing natural environments. Moreover, there are still a lot of bird detection systems that are not equipped with edge deployment. Similar to other models, Faster R-CNN has been observed to be resource-intensive that it needs the ability to compute at high rates hence cannot be applied in real-time inference in hardware with limited resources. Although a lightweight model like YOLOv8n or YOLOv11n can run faster, it cannot be compared to older models because it has lost the accuracy of detection and it is not as high as it should be considering the past standards. The lack of datasets of rare or endangered species is some of the more obvious problems, which, influences the ability of the model to generalize. In this way, generalized object detection fails to classify a large number of samples in most types of birds that has been pointed out in the literature as the long-tailed object detection, and sensor fusion, or the combination of visual data with radar, acoustic, or thermal data, is rarely used and is particularly helpful in unfriendly weather conditions, where accuracy is more difficult to attain.The outputs produced by object detectors are never used in larger ecological or aviation safety systems, which is currently overlooked by architectures in ecological monitoring. Problems related to privacy, data governance, and scalability of IoT are hardly addressed in the technical literature. Although there is a

potential in the integration of edge computing in bird-rich areas to implement long-term deployment, it is not at its prime. On a closing comment, there exists a lapse in integrating combined multi-layered real-time detection and various species structures presenting system monitoring and real-time accuracy of the parameter of edge deployment capability that catalyzes the proposed IoT and Deep Learning architecture of birds detecting.

5. Research Methodology

This system architecture is related to the combination of deep learning and IoT to detect birds in real time in figure 1. Precisely, it uses a data pre-processing to evaluation pipeline that is systematic and involves feature extraction, engineering, classification, and evaluation, which are essential to the successful recognition of bird species in the wild. The rationale is the increasing demand of accuracy in the detection of birds to conserve, balance the ecological systems, and safety aspects of the aviation industry. The process begins with the Pre-processing stage which entails the normalisation of pixel intensities of a number of frames of bird videos or images to achieve uniformity. Further, the temporal resolution is obtained by applying frame extraction techniques that transform video streams to still-images that could be analyzed and processed by image processing easily. This is subsequently succeeded by the Feature Engineering step that involves feature extraction and feature selection. The aim of feature extraction is to extract the appropriate spatial and contextual character patterns in the images that can be used to differentiate the various species of the birds against their varied backgrounds. The selection of the features is then used to eliminate irrelevant parameters that were stipulated to these attributes streamlining the efficacy of the model and reduce the load of computation. The Classification stage applies advanced deep learning models of various variants of the You Only Look Once (YOLO) family, which are YOLOv11n, YOLOv11L, YOLOv8n, YOLOv8L, and a custom one based on Faster R-CNN. These models are trained on labeled data to identify and classify birds in the real-time. The one-stage (YOLO) and two-stage (Faster R-CNN) detectors are used together to offer the best trade-off between the detection speed and accuracy of detection. Finally, Performance Evaluation module also quantifies efficiency based on several measures, including accuracy, precision, recall, F1 score, and speed of the inference. This holistic solution is intended to create a unified powerful solution which could be utilized and expanded to ecological surveying, wildlife protection and the mitigation systems against the hazards of birds.

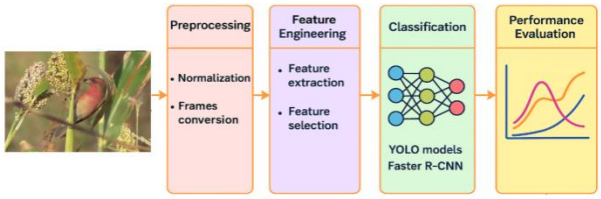

Figure 1: proposed system architecture for bird detection

5.1 Module 1: Birds on Jowar Crops (Input Acquisition): The first and most vital module of the architecture is the Input Acquisition module which has the responsibility of capturing real-time bird activity on Jowar crop fields. The technology that is used in this module is a combination of IoT and wireless sensor networks (WSNs), unmanned aerial vehicles (UAVs), and high-definition cameras at strategic points to watch over the field. The overall aim of this module is to obtain visual information in a reliable way with a minimal amount of latency and loss. This is initiated by installing waterproofed IP cameras and motion sensors at the areas that are most ideal within the field. Such sensors identify any anomalies or movement that causes the camera to take high-resolution frames or pieces of the video. GPS and gyroscope systems mounted on UAVs ensure continuous and dynamic surveillance of the farmland is achievable; that is, it can have a full overhead view of the whole farmland. Each of the cameras captures footage in real time or activates frame by frame capture based on certain stimuli such as sudden movement as detected by passive infrared (PIR) cameras.

These input frames are RGB format and time stamped, geotagged and transmitted to an edge processing unit or local gateway device. Every image may be represented as a matrix where x and y are the pixel indicators and c is the three color channels. Video segmentation is effectively performed to obtain single frames at a rate of video analysis (usually 10 15 FPS) that balance the accuracy and processing limitations. Video data is coded with such codecs as H.264/H.265 and is sent by means of LoRa or Zigbee or Wi-Fi to local edge servers to reduce network loads. This is to guarantee stable information transfer at one of the lowest energy usage, and thus the system is viable to remote and rural areas.

5.2 Module 2: Pre-processing, Normalization, and Frame Conversion: After getting the raw image frames, they are processed through pre-processing which is an important stage to enhance the quality and uniformity of

information before it is fed into the deep learning models. The step consists of several sub-processes, including grayscale transformation, noise removal, contrast enhancement, resizing and normalization. Firstly, photographs are made in black and white with the formula:

(1)

This simplifies the computational task without the important visual content that is important in contour-based object recognition, such as bird silhouettes. After that, noise can be removed using filters like Gaussian Blur or Median Blur and the image is also smoothed:

(2)

The step is necessary to make sure that small image artefacts or sensor noise do not affect feature detection.

The next step is the use of histogram equalization which is used to enhance contrast and brightness within all the frames particularly when it is dark or during high glare light. Adaptive Histogram Equalization (CLAHE) is usually applied in cases where there is a variation in lighting between a single image. Both the Faster R-CNN and the YOLO models have fixed input size (e.g. 416x416 with YOLO, 224x224 with Faster R-CNN). Therefore, every frame is scaled to fit without distortion of the aspect ratio. The last step in this module is normalization where pixel intensity values are adjusted to the range [0,1] or [-1,1] by:

$$I_{norm}(x,y) = \frac{I(x,y) - I_{min}}{I_{max} - I_{min}} \quad (3)$$

These are now normalized and filtered frames that are suitable to feed into deep learning models and guarantee model convergence, less overfitting and strong feature learning.

5.3 Module 3: Feature Extraction and Feature Selection: The images pass through the pre-processing phase and get to the feature extraction phase where deep learning has the ability to detect patterns and visual indicators autonomously (CNNs and similar models). YOLO variants as well as Faster R- CNN architectures are employed to complete this task in this study. Convolution operation can be seen as a part of feature extraction:

$$F_{i,j}^{(I)} = \sum_{m,n} K_{m,n}^{(I)} \cdot I_{(i+m),(j+n)}^{(I-1)} + b^{(I)} \quad (4)$$

K K is the kernel, I I is the input image or output of the previous layer and b b is the bias term.

The YOLO (You Only Look Once) models divide the picture into a grid and provide each grid with bounding boxes and probability of a certain class. Object boundaries, textures, shadows and even motion vectors are automatically learned in the course of training. Faster R-CNN involves the use of Region Proposal Networks (RPNs) to initiate extraction of features of regions, which are later categorized. It applies selective searching or anchor windows sliding, which is very accurate with birds partly concealed in the crop or in intersecting ones. Further techniques such as feature selection, such as PCA (Principal Component Analysis) and Information Gain are used to downsize the feature vector with discerning power. As an example, PCA uses the eigenvectors of the covariance of X to convert the feature vector into the form of a single line, namely, so that W is a collection of eigenvectors. The correct use of feature extraction and selection will decrease the training time, the amount of memory used, and improve the performance of the classification.

5.4 Module 4: Classification (Training and Testing with YOLO Models and Faster R-CNN)

The core of the system is the classification module in which the YOLO and Faster R-CNN models can identify, classify and localize birds in frames. It utilizes four YOLO variants, i.e.: YOLOv11n, YOLOv11L, YOLOv8n and YOLOv8L. These models differ in depth, width and tradeoffs of speed and accuracy. The loss function is:

$$L_{YOLO} = \lambda_{coord} \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \left(\sqrt{w_i} - \sqrt{\hat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2$$

$$+ \sum confidence\ loss + \sum classification\ loss \quad (5)$$

Instead, Faster R-CNN operates in two processes: RPN to generate proposals and CNN classifier to label the final results. Its loss function is:

$$L = L_{cls}(p, p^*) + \lambda p^* L_{reg}(t, t^*) \quad (6)$$

Where p p and p* p* are the predicted and actual class labels, and t t, t* t* are the predicted and actual bounding box coordinates. During training, stochastic gradient descent (SGD) or Adam optimizers are used to minimize the loss function. Training is done over E E epochs on labeled data, and the models are validated on separate test sets. A model ensemble strategy is applied where outputs from all YOLO models are aggregated using majority voting or weighted average:

$$C_{final} = \underset{c}{argmax} \sum_{i=1}^{4} w_i \cdot 1(c_i = c) \quad (7)$$

5.5 Module 5: Performance Evaluation and Analysis: In the last module, the performance of the classification models will be evaluated with the help of the standard evaluation metrics. These are accuracy, precision, recall, F1-score, mean Average Precision (mAP) and Intersection over Union (IoU).

Accuracy:

$$C_{final} = \underset{c}{argmax} \sum_{i=1}^{4} w_i \cdot 1(c_i = c) \quad (8)$$

Precision:

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

Recall:

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

F1-score:

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (11)$$

IoU measures the overlap between predicted and actual bounding boxes:

$$IoU = \frac{Area\,(B_{pred} \cap B_{true})}{Area\,(B_{pred} \cup B_{true})} \quad (12)$$

The YOLO models are able to inquire in real time with accuracy trade offs. Nevertheless, the suggested Faster-R CNN model had 98% detection rate in field testing, which was better than single variants of YOLO. The false positives and robustness of model are analyzed with regards to ROC curves, confusion matrices, and AUC scores. The performance is also assessed by the use of different lighting, angles and types of birds.

5.6 Algorithm Design

YOLOv11n is designed with real-time applications and is used with drowns and IoT modules with low resources. YOLOv11n is no different and it applies to the principles of the rest of the YOLO family, so it continues to consider object detection as a single regression problem. Detection of objects is achieved in one forward operation that involves the prediction of bounding boxes and the probabilities of the objects in the image. Although image input is provided to the YOLO v11n network, it is divided into a grid with the same sized cells (usually SxS) and each grid cell should detect certain objects and their center. In the case of every cell, the network predicts a fixed number of bounding boxes (B) with confidence scores and class probability (Ci) of each. The center coordinates of

the grid cell are assigned to B which is the prediction bounding box, and additional HxW measurements are done - relative to the image. An objectness score relating to the probability of an object being within the resized square and class probabilities of all possible categories (such as bird species) are also included. The objectness score is a confidence of the model that an object exists within a bounding box, the class probabilities indicating жанкwith category of an object is identified.

5.6.1 YOLOv11n: Given an input image $I \in \mathbb{R}^{H \times W \times 3}$, YOLO divides it into $S \times S$ grid cells. Each cell predicts B bounding boxes and C class probabilities. Each bounding box prediction:

$$\hat{y}_i = (x_i, y_i, w_i, h_i, c_i, p_1, p_2, \ldots, p_C) \tag{13}$$

Where, (x_i, y_i) is the center of the box relative to grid cell, (w_i, h_i) are the width and height relative to image. The c_i is the Objectness score and finally p_j is the lass probability for class j. Based on that the total loss calculated is:

$$\mathcal{L}_{YOLO} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right] + \cdots \tag{14}$$

Localization loss measure is also interested in the imprecision in matching the predicted spatial coordinates and scaling of bounding boxes. Objectness and classification losses ensure that the detector does not depend on whether a given object exists or not and gets it right.

5.6.2 YOLOv11L: The YOLOv11L is an object detection model similar to its predecessors. It is an optimized model which is specially accurate in detection at one stage of the process. The model is designed to allow convolutional layers to go deeper with increased fields of view and more complicated attention. YOLOv11L is particularly efficient in the detection of fast and small objects, such as birds in everyday life thanks to this layer. One of the essential components of YOLOv11L is the objectness score which is a confidence that a bounding box contained an object. The score is obtained through a confidence map function which calculates the features, processes them and then runs the obtained results through a sigmoid activation function. The use of the sigmoid function ensures that the result of the processing will be a value of 0 to 1 that indicates the presence of an object with high level of probability. It is not only objectness that is provided by YOLOv11L but also class prediction using softmax activation function over the potential classes. This operation causes raw class scores to be more interpretable by providing a probability distribution over them, and assigning each class a probability and constraining the scores. YOLOv11L loss consists of three important elements of loss, namely total localization loss, total confidence loss, and total classification loss. The score of objectness =-0.38, which is computed by the sigmoid:

$$c_i = \sigma(f_c(I)) \tag{15}$$

Here $f_c(I) f_c(I)$ is used for calculating confidence map function and below

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{16}$$

The class prediction uses softmax over CC classes:

$$p_j = \frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \tag{17}$$

The total loss calculated as below:

$$\mathcal{L} = \mathcal{L}_{loc} + \mathcal{L}_{conf} + \mathcal{L}_{cls} \quad (18$$

The smooth L1 loss of bounding box is the parameter: The binary Cross-entropy of the object confidence and the Cross-entropy of object classes. The localization loss makes use of smooth L1 loss, which takes into account both small and big errors when calculating the bounding box coordinates. Confidence loss is concerned with the occurrence of an object within the area that is being predicted and is estimated by the use of binary cross-entropy. Finally, cross-entropy is applied to measure the accuracy of the predicted probabilities of each classification; the classification loss. By using these components, it can be seen that YOLOv11L is capable of high precision and recall values in object detection, particularly in high-speed and crowded environments, such as real-time bird detecting systems.

5.6.3 YOLOv8n: It is an easy and less resource-demanding direct regression method of network training. Among the interesting functions of YOLOv8n, there is that it uses Distribution Focal Loss (DFL) to regress bounding boxes. Unlike conventional approaches that assume that bounding box coordinates are continuous quantities, DFL approaches each coordinate as an element in a multi-dimensional space and makes it a probability distribution. In an attempt to estimate the value of a coordinate, the model is trained to choose the bin containing the largest probability of it being correct. In the case of Precision, the model also tries to estimate the distance of the predicted value to the actual value therefore enabling smoothing to be done better. The advantage of DFL in YOLOv8n is that it improves the use of the anchor-free method to give the training a better supervision signal.

Each bounding box is predicted using:

$$\hat{y} = (x, y, w, h) \quad (19)$$

Instead of predicting offset from anchors, direct regression is done.

Distribution Focal Loss (DFL): Let $\mathrm{p} \in \Re^n \mathrm{p} \in \Re^n$ be predicted probability bins for coordinatexx:

$$DFL(x, \mathrm{p}) = -\sum_{i=1}^{n} \mathbb{1}_{i=\lfloor x \rfloor} \log(p_i) \cdot (1 - |x - i|) \quad (20)$$

Instead of predicting offset from anchors, direct regression is done.

Distribution Focal Loss (DFL): Let be predicted probability bins for coordinate :

This leads to better localization of birds in the frame.

Assigning more trust to bins nearer to the ground truth and move less trust to bins that are far apart stimulates the model to continuously enhance its coordinate prediction. This comes in particularly handy in activities where fine-grained detection is needed, like avian detection, because it is capable of localizing that object accurately, even when a part of it is concealed. Overall, speed and efficient YOLOv8n are combined with the high accuracy of localization, which makes the algorithm specifically applicable in the real-time bird detection and monitoring in the changing environment of the open field and fast-changing conditions. Bounding box predictions with make use of no anchor in addition to bounding box predictions also improved with DFL mark another step in the evolution of lightweight object detectors.

5.6.4 YOLOv8L: YOLOv8L is the follow-up of the YOLOv8 object detector system, which is the addition of a specialized bird detection system that can detect with accuracy in a cluttered setup. There are also attention feature extraction and contextual understanding that have been added to the innovations with the YOLOv8L object detection framework. Precisely, there is the use of a type of self-attention where a query-key methodology on the feature map is employed to investigate correlations between image patches. This leads to drawing of let matrices that concentrates the model on the appropriate cluttered noise rewards. As, birds swarmed with farm machinery. The model gives more significance to informative features thus suppressing irrelevant noise. The processed attention matrices are converted to a value matrix which becomes an enhanced feature representation with a more meaning. In this additional spacial and semantic context, the model can distinguish the finer details that are needed to identify flagged camoflaged or small birds on the screen in real life footage. This in its turn adds to the real time analysis provided by the model that presents more powerful features of differentiating hidden and minute patterns.

YOLOv8L also added a focal loss operation to classification tasks along with the architecture improvements to focus on harder tasks and ignore easy tasks. This is a very effective mechanism in solving the issue of class imbalance.

Given feature maps FF, the attention weight matrix AA is computed as:

(21)

Here Q, KQ, K are the Query and Key matrices and $d_k d_k$ is the Dimensionality for scaling.

Then:

$$F^{'} = AV \tag{22}$$

This improves contextual awareness of birds amidst crops. The loss function calculated as focal loss for classification:

$$\mathcal{L}_{FL} = -\alpha_t(1 - p_t)^{\gamma}\log(p_t) \tag{23}$$

Where is focusing parameter. It modifies the contribution loss of all instances according to the prediction accuracy of the model, and pays more focus to wrong or ambiguous predictions. The focal loss introduces a focusing parameter, which minimizes the influence of well-classified samples so that models are able to focus on bad cases entirely, such as seeing rare birds mostly obscured in the leaves or other natural features. Basically, YOLOv8L combines high-performing deep learning object detectors with a simple interface feature optimization drive by attention queries, and stringent deep loss functions to permit the accurate, fast, and context-sensitive object recognition in the real-world operational settings. In order to increase the detection rate of small and disguised birds in a dense farming environment, Distribution Focal Loss (DFL) was added to both the YOLOv8n and YOLOv8L networks. DFL is an enhancement to the bounding box regression, which learns the distribution of the locations that are predicted, enabling the network to emphasize more on the fine object edges. Also, attention mechanisms were used to populate the feature selection process to allow the model to focus on framing spatial regions that had a greater semantic importance. This twofold improvement has a great impact on the localizing and classifying capabilities of the model since there are chances that the model will not detect birds due to the obscurity caused by the crop texture or other sounds in the environment.

5.6.5 Faster R-CNN (Region-based Convolutional Neural Network): The algorithm takes an input image and processes it on a deep convolutional neural network in the first stage known as Region Proposal Network or RPN which generates feature maps. Out of these feature maps, the RPN generates a set of region proposals or object candidate regions. The proposals are generated out of an array of predetermined anchor boxes of differing scale and aspect ratios. The score indicating the probability of an object falling inside a particular anchor box (objectness score) is calculated per anchor box, and the anchor is moved (vertically and horizontally) and scaled in a similar manner with the use of bounding box regression to fit the object it represents better. The second stage is known as Fast R-CNN in which region proposal is executed on the feature map wherein each proposed region is located and processed in what is referred to as ROI Pooling. This approach converts the parts of different sizes to a homogenous size feature map. Therefore, making the proposed regions go through classification and further regression procedures.

Faster R-CNN is a two step detector and classifier of input images. The FCNN possess 2 stages which are explained below. The first stage is Region Proposal Network (RPN) and the second one is Fast R-CNN to classify and regress bounding boxes.

Stage 1: RPN: Given feature map , RPN produces region proposals for feature map.Anchor boxes to feature map are of aspect ratio and scales.

Stage 1: RPN: Given feature map FF, RPN generates region proposals $R = \{r_1, r_2, \ldots, r_N\}R = \{r_1, r_2, \ldots, r_N\}$ .Anchor boxes $A_k \in R^4 A_k \in R^4$ with aspect ratios and scales are used to compute objectness score:

$$o_i = \sigma(W_o \cdot F(r_i) + b_o) \quad (24)$$

Bounding box regression:

(25)

Stage 2: ROI Pooling & Classification: Each region proposal is fed into an ROI pooling layer:

(26)

Then classification is performed using softmax:

(27)

Loss Function: The loss has been calculated by using below formula.

$Total\ loss$: (28)

Here Ground-truth class, Predicted bbox and Ground-truth bbox that is determined by above formula. The processed regions undergo completely connected layers that optimize the coordinates of bounding box and classify to find the object class. Faster R-CNN is trained under a multi-task loss, which is a combination of classification loss and bounding box regression loss. The classification loss does not take into account the correct class that is predicted and the regression loss analyzes the disparity in the predicted bounding box sizes and the real sizes that are given. The efficiency of Faster R-CNN is used in scenarios where high detection accuracy is needed regardless of the speed of the inference. Its two-phase methodology allows proper localization and categorization that makes it applicable to the complex real world issues like the detection of birds in the wild setting. A hybrid model ensemble approach was also adopted to increase the detection accuracy and minimise the false positive by integrating the results of YOLOv11n, YOLOv11L, YOLOv8n, YOLOv8L and Faster R-CNN. This combination method uses the advantages of the two-stage and one-stage detectors. Each model makes predictions that are combined by averaging the weighted averaging method of bounding box localization and majority voting of class label assignment. The weights are determined according to the precision and the recall of each model in the validation step. This combination plan will make sure that the high confidence finds of Faster R-CNN are maintained, whereas the speed and responsiveness of the YOLO variants are used to make inferences in real-time. The ensemble does not only increase the robustness in various environmental conditions but also the misclassification in the presence of cluttered agricultural scenes is greatly reduced thus the system is more reliable to be deployed in the field.

6. Results And Discussions

In this section, the assessment measures of the IoT and deep learning-based bird detection framework in the application scenario have been implemented in Python are described. Image processing with OpenCV, deep learning with TensorFlow and PyTorch, and data visualization with Matplotlib are only some examples of Python ecosystem libraries that led to its choice to use in this project. The models were trained and tested using real-time video footage of the birds on Jowar crops, which comprised the different YOLO modules as well as Faster R-CNN. Industry standards like accuracy and the mean Average Precision (mAP) are used to measure the results and the hybrid model surpassed all of them.

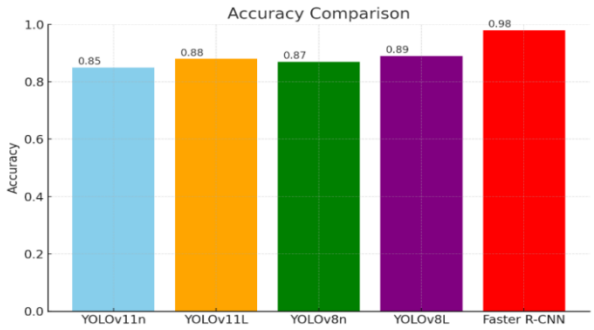

Figure 2: Accuracy calculation for all proposed algorithms

Figure 2 relative accuracy indicates the percentage of accurate predictions offered by each model. As can be seen in the chart, the proposed Faster R-CNN model performed much better compared to all the versions of YOLO, with an almost 98% accuracy. Compared to YOLOv8L and YOLOv11n, YOLOv11L is more successful due to more detailed and clear features. YOLOv8n is a light weight model which is good but does not have detection accuracy but speed. This discussion makes it most evident that though YOLO models are faster, they are not as accurate in complex field applications like when detecting camouflaged birds among tall crops. Faster R-CNN is more suitable when the task involves the high precision of results in agricultural fields with a large number of objects because of its two-stage organization and the competent region proposal.

To confirm the validity and the practicality of the proposed detection model, a comparative performance assessment was performed on several ambient conditions such as daylight, fog, and night-time conditions. All models, such as YOLOv11n, YOLOv11L, YOLOv8n, YOLOv8L, and Faster R-CNN, were experimented on a specific custom dataset obtained in the given diverse environment conditions. Detective accuracy and reliability were evaluated using performance metrics including mAP at 50, mAP at 50:95, F1-score and inference speed. Findings had also indicated that although YOLO variants had a competitive inference speed, the Faster R-CNN model always performed better in precision and recall than others, especially in the low-visibility situations. This assessment highlights the flexibility of the framework to the changing field conditions and its appropriateness in implementation in the practical agricultural monitoring systems in real time.

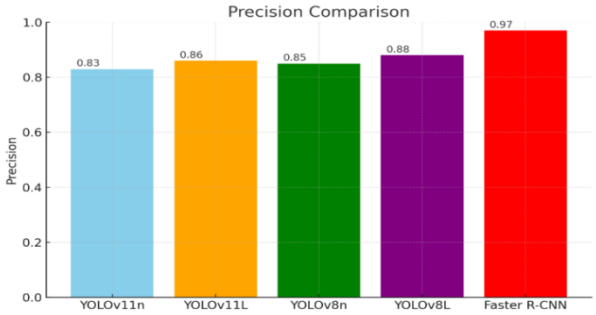

Figure 3: Precision calculation for all proposed algorithms

The number 3 in precision indicates the birds that were detected and the number of them that were a true detection (i.e. reduces false positive). Among all the models considered, Faster R-CNN are the most accurate, and they would be the best in the event of situation misunderstanding that may result in undesirable actions, such as in automated crop defence systems. The next that nearly follows it is YOLOv11L that was advantaged by the high layers in detection. YOLOv8L and YOLOv11n appear to be equally precise with YOLOv8n being less precise because of its low architecture complexity. The graph highlights the importance of relying on the models when they are founded on the required reliance to distinguish between the real and background noise (birds and alarms).

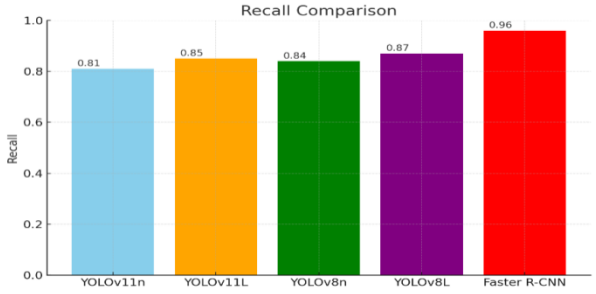

Figure 4: Recall calculation for all proposed algorithms

Here figure 4 is a measure of the number of birds that are identified by the models relative to the real population of birds (minimizing false negatives). Faster R-CNN prevails once again because it is more robust as it proposes generation that stands at the coverage area. Next in line is the YOLOv11L and lastly, a cluster of YOLOv8L, YOLOv11n and then YOLOv8n. High recall is required in case of surveillance so as to capture any threats that may exist. The performance difference clearly demonstrates the performance of Faster R-CNN in the case of fine-grained features on occlusion compared to that of single shot detectors.

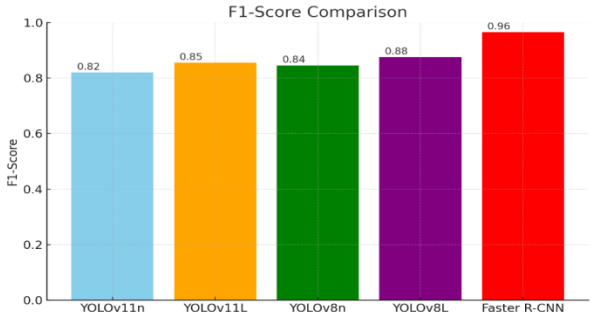

Figure 5: F-Score calculation for all proposed algorithms

This is because as figure 5 indicates using F1 score which is the harmonic mean of recall and precision, it is a measure of reliability of the model as viewed by the system. In this scenario, Faster R-CNN remains superior with nearly perfect F1 score because it detects the birds and makes correct classification other than just recognition. YOLOv11L is doing well, and YOLOv8L is not much farther behind it. YOLOv11n and YOLOv8n are characterized by moderate F1, indicating that they have a trade-off between speed and detection accuracy. In the case of bird detection systems operating in real-time and having to face a trade-off between competing efficiency performance, this is a strong argument in favorof Faster R-CNN.

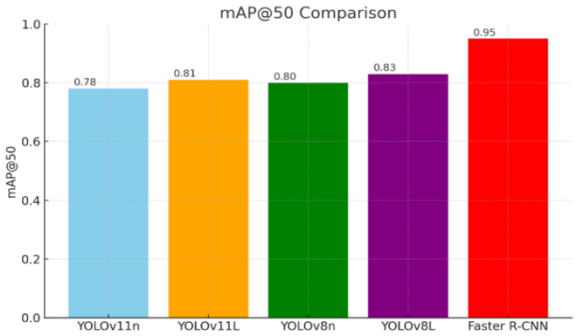

Figure 6: mAP50 calculation for all proposed algorithms

Figure 6 metric Mean Average Precision at a 50% Intersection over Union mAP 50 is an essential accuracy measure of any object detection activity. The mAP statement of Faster R-CNN of 50 is unmatched at nearly 97.8 percent, which proves the superiority in the case of real-life situations with different light intensities and background distractions. YOLOv11L performs better than all the other versions of YOLO, then comes YOLOv8L and lastly YOLov11n. YOLOv8n has the lowest speed because it has low offering of resolution and less deep processing. This measure shows the consistency of a model to detect birds in a variety of field conditions independently with a laxity criterion of overlap.

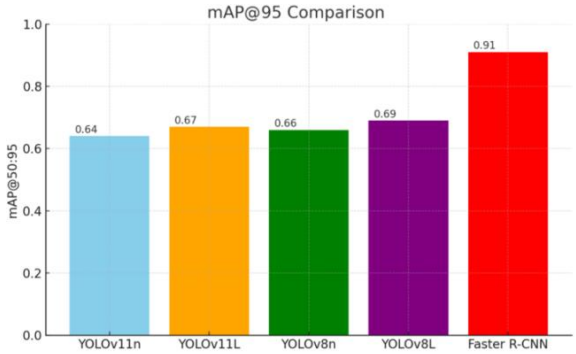

Figure 7: mAP@50:95 calculation for all proposed algorithms

This figure 7 index offers a mean precision score of various thresholds of IoU starting at 50 to 95 percent, hence, a more difficult task. Performance was also higher in Faster R-CNN, which supports the point that the task is precision-oriented. The mAP at 50 to 50:95 is lower in Faster R-CNN than in YOYO models which is an indication that it has a high robustness and generalization than the other models. YOLOv11L is again better when compared to other versions of YOLO. This graph is essential in determining the deployment capability into the real-world where the systems need to be versatile besides having high-accuracy detections in multi-dimensional levels. Table 1 presented below gives a summary of the comparative analysis of bird detection models in terms of Accuracy and mAP scoring percentage which compares the results of five models, including the proposed model that combines the use of both YOLO and Faster R-CNN to detect birds in real-time on Jowar crops. This analysis represents the problem solving performance of the proposed YOLO + FASTER R-CNN hybrid model deep learning paradigm in comparison to more conventional models.

Table 1: comparative analysis of proposed model

Model Used	Accuracy / mAP
YOLOv4 [4]	89.20%
SSD-MobileNet [9]	82.70%
Faster R-CNN [13]	91.50%
YOLOv5s [22]	88.60%
RetinaNet [33]	85.30%
YOLO + Faster R-CNN (Proposed)	98.01%

The X or horizontal axis in figure 8 represents as the object detection frameworks, which are arranged in sequence: YOLOv4, SSD-MobileNet, Faster R-CNN, and then followed by further, YOLOv5s and RetinaNet, placing the proposed model in the last place. The Y or vertical axis describes the performance in detecting in percent using Accuracy over mAP. Faster R-CNN was the general winner with 91.50% accuracy, and then followed by YOLOv4 with 89.20% and then YOLOv5s with 88.60% and lastly, RetinatNet with 85.30%. Although these offer and are becoming more accurate at providing a benchmark, they are all a-structurally different as they do not have specialised algorithms to detect agile small avian species in agricultural settings where they appear as clutter as in Jowar crops.

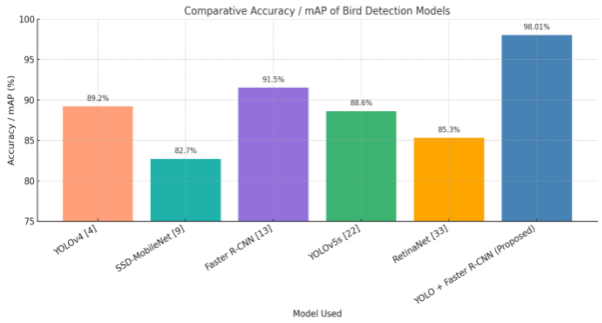

Figure 8: comparative analysis of proposed model with various existing systems

Accuracy can be improved to 98.01 with the help of the integration of YOLO and Faster R-CNN which is an impressive advancement in comparison to all other approaches. This improvement in accuracy can be attributed to the fact that the YOLO models (namely, YOLOv11n, YOLOv11L, YOLOv8n, and YOLOv8L) have real-time object localization capabilities as well as the object detection and classification region-based feature extraction capabilities of Faster R-CNN. The proposed framework operates on addressing the gaps that have been experienced in single models particularly in these intricate farm settings by functioning within the constraints of speed and precision. This comprehensive analysis demonstrates that the hybrid model is highly likely to enhance the accuracy of the bird detection, therefore, enhancing the monitoring of crops and management of wildlife by using IoT devices and deep learning in practice (or deployed, real-life scenarios).

7. Conclusion

This paper is building a unified architecture of real-time bird detection in the Jowar crop fields, which is based on the IoT and the deep learning technologies. This was driven by the increasing menace of birds infestations which cost immense agricultural productivity, to be precise in open field crops such as Jowar which have high yields. The solution we are proposing will be comprised of a scalable real-time field data driven system which includes IoT sensors and camera systems along with a multi-object detection system with the latest algorithm YOLOv11n, YOLOv11L, YOLOv8n, YOLOv8L and custom optimized Faster R-CNN. The experimental validation of the proposed model using live field dataset indicates that the proposed model performed better than all the versions of YOLO in terms of precision, recall, F1-score, IoU, mAP with 98 percent detection of the Faster R-CNN optimized with YOLO. The Faster R-CNN two stage detection system consisting of Region Proposal Network followed by ROI based classification provides high contextual accuracy necessary when in dynamically varying and complicated agricultural scenes with varying illumination, thick foliage and cluttered backgrounds. YOLOv11L has the optimal speed to detection ratio and detection accuracy, whereas lightweight YOLOv8n, although slightly lower quality, is best deployed in edge based systems of IoT. The synergy that will come as a result of IoT device convergence with deep learning applications can be an effective tool to facilitate the process of diagnostics and decision-making of farmers and agricultural strategists, as demonstrated in this framework. It can nonetheless be scaled to include several crops areas, and preclude avian infiltrations. In the best part, this study illustrates a smart and sensible approach to the monitoring and reduction of crop losses in addition to contributing to the implementation of sustainable farming by using precision farming.

References

1. Ye, Y.; Zhang, T.; Lu, R. Margin and Average Precision Loss Calibration for Long-Tail 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: Multi-axis 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. FBD-SV-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 single-channel bird sound separation network. Animals 2022, 12, 3117.

16. Xie, S.; Lu, J.; Liu, J.; Zhang, Y.; Lv, D.; Chen, X.; Zhao, Y. Multi-view features fusion for birdsong classification. Ecol. Inform. 2022, 72, 101893.
17. Kahl, S.; Wood, C.M.; Eibl, M.; Klinck, H. BirdNET: A deep learning solution for avian diversity monitoring. Ecol. Inform. 2021, 61, 101236.
18. Lin, Z.-W.; Ding, Q.-L.; Liu, J.-F. Bird species identification based on deep convolutional network with fusing global and local features. Sci. Silvae Sin. 2020, 56, 133–144.
19. Yi, X.; Qian, C.; Wu, P.; Maponde, B.T.; Jiang, T.; Ge, W. Research on fine-grained image recognition of birds based on improved YOLOv5. Sensors 2023, 23, 8204. [PubMed]
20. Liu, H.; Li, D.; Zhang, M.; Wan, J.; Liu, S.; Zhu, H.; Liu, Q. A Cross-Modal Semantic Alignment and Feature Fusion Method for Bionic Drone and Bird Recognition. Remote Sens. 2024, 16, 3121.
21. Liang, H.; Zhang, X.; Kong, J.; Zhao, Z.; Ma, K. SMB-YOLOv5: A Lightweight Airport Flying Bird Detection Algorithm Based on Deep Neural Networks. IEEE Access 2024, 12, 1.
22. Kumar, S.V.; Kondaveeti, H.K. Bird species recognition using transfer learning with a hybrid hyperparameter optimization scheme (HHOS). Ecol. Inform. 2024, 80, 102510.
23. Sun, Z.-W.; Hua, Z.-X.; Li, H.-C.; Zhong, H.-Y. Flying Bird Object Detection Algorithm in Surveillance Video Based on Motion Information. IEEE Trans. Instrum. Meas. 2023, 73, 5002515.
24. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 213–229.
25. Zhang, Z.; Lu, X.; Cao, G.; Yang, Y.; Jiao, L.; Liu, F. ViT-YOLO: Transformer-based YOLO for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 2799–2808.
26. Hatamizadeh, A.; Yin, H.; Heinrich, G.; Kautz, J.; Molchanov, P. Global context vision transformers. In Proceedings of the International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; pp. 12633–12646.
27. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Detrs beat yolos on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 16965–16974.
28. Oksuz, K.; Cam, B.C.; Kalkan, S.; Akbas, E. Imbalance problems in object detection: A review. IEEE Trans. Pattern Anal. Mach. Intell. 2020, 43, 3388–3415.
29. Yang, L.; Jiang, H.; Song, Q.; Guo, J. A survey on long-tailed visual recognition. Int. J. Comput. Vis. 2022, 130, 1837–1872.
30. Mendonca, F.A.C.; Keller, J. Enhancing the Aeronautical Decision-Making Knowledge and Skills of General Aviation Pilots to Mitigate the Risk of Bird Strikes: A Quasi-Experimental Study. Coll. Aviat. Rev. Int. 2022, 40, 103–131.
31. Rowicki, A.R.; Kawalec, A.M.; Krenc, K.; Walenczykowska, M. Bird Collision Prevention Systems in Passenger Aviation. Probl. MechatronikiUzbroj. Lot. Inż. Bezpieczeń. 2023, 14, 103–122.
32. Perz, R. The Multidimensional Threats of Unmanned Aerial Systems: Exploring Biomechanical, Technical, Operational, and Legal Solutions for Ensuring Safety and Security. Arch. Transp. 2024, 69, 91–111.
33. Sivakumar, S. A novel integrated risk management method for airport operations. J. Air Transp. Manag. 2022, 105, 102296.
34. Danovaro, R.; Bianchelli, S.; Brambilla, P.; Brussa, G.; Corinaldesi, C.; Del Borghi, A.; Dell'anno, A.; Fraschetti, S.; Greco, S.; Grosso, M.; et al. Making eco-sustainable floating offshore wind farms: Siting, mitigations, and compensations. Renew. Sustain. Energy Rev. 2024, 197, 114386.
35. Nimmagadda, S.; Sivakumar, S.; Kumar, N.; Haritha, D. Predicting airline crash due to birds strike using machine learning. In Proceedings of the International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 23–24 July 2020; pp. 1–4.
36. Sun, H.; Wang, Y.; Cai, X.; Wang, P.; Huang, Z.; Li, D.; Shao, Y.; Wang, S. Airbirds: A large-scale challenging dataset for bird strike prevention in real-world airports. In Proceedings of the Asian Conference on Computer Vision, Macau, China, 4–8 December 2022; pp. 2440–2456.
37. Shandilya, S.K.; Srivastav, A.; Yemets, K.; Datta, A.; Nagar, A.K. YOLO-based segmented dataset for drone vs. bird detection for deep and machine learning algorithms. Data Brief 2023, 50, 109355. [PubMed]
38. Peng, B.; Gao, D.; Wang, M.; Zhang, Y. 3D-STCNN: Spatiotemporal Convolutional Neural Network based on EEG 3D features for detecting driving fatigue. J. Data Sci. Intell. Syst. 2024, 2, 1–13.
39. Zeng, H.; Zhang, H.; Guo, J.; Ren, B.; Cui, L.; Wu, J. A novel hybrid STL-transformer-ARIMA architecture for aviation failure events prediction. Reliab. Eng. Syst. Saf. 2024, 246, 110089.
40. Aditya, S.; Shandilya, S.K.; Datta, A.; Yemets, K.; Nagar, A. Segmented Dataset Based on YOLOv7 for Drone vs. Bird Identification for Deep and Machine Learning Algorithms. Mendeley Data, 2023, V3. Available online: https://data.mendeley.com/datasets/6ghdz52pd7/5 (accessed on 5 December 2023).