

## Article

# IoT-Based Object-Detection System to Safeguard Endangered Animals and Bolster Agricultural Farm Security

Mohaimenul Azam Khan Raiaan , Nur Mohammad Fahad , Shovan Chowdhury , Debopom Sutradhar , Saadman Sakib Mihad  and Md. Motaharul Islam \*

Department of Computer Science and Engineering, United International University, Madani Avenue, Dhaka 1212, Bangladesh; mraiaan19122@bscse.uiu.ac.bd (M.A.K.R.); nfahad191040@bscse.uiu.ac.bd (N.M.F.); schowdhury191177@bscse.uiu.ac.bd (S.C.); dsutradhar201046@bscse.uiu.ac.bd (D.S.); smihad201063@bscse.uiu.ac.bd (S.S.M.)

\* Correspondence: motaharul@cse.uiu.ac.bd; Tel.: +880-171-2644837

**Abstract:** Significant threats to ecological equilibrium and sustainable agriculture are posed by the extinction of animal species and the subsequent effects on farms. Farmers face difficult decisions, such as installing electric fences to protect their farms, although these measures can harm animals essential for maintaining ecological equilibrium. To tackle these essential issues, our research introduces an innovative solution in the form of an object-detection system. In this research, we designed and implemented a system that leverages the ESP32-CAM platform in conjunction with the YOLOv8 object-detection model. Our proposed system aims to identify endangered species and harmful animals within farming environments, providing real-time alerts to farmers and endangered wildlife by integrating a cloud-based alert system. To train the YOLOv8 model effectively, we meticulously compiled diverse image datasets featuring these animals in agricultural settings, subsequently annotating them. After that, we tuned the hyperparameter of the YOLOv8 model to enhance the performance of the model. The results from our optimized YOLOv8 model are auspicious. It achieves a remarkable mean average precision (mAP) of 92.44% and an impressive sensitivity rate of 96.65% on an unseen test dataset, firmly establishing its efficacy. After achieving an optimal result, we employed the model in our IoT system and when the system detects the presence of these animals, it immediately activates an audible buzzer. Additionally, a cloud-based system was utilized to notify neighboring farmers effectively and alert animals to potential danger. This research's significance lies in its potential to drive the conservation of endangered species while simultaneously mitigating the agricultural damage inflicted by these animals.

**Keywords:** object-detection model; IoT; endangered animals; agriculture farms; cloud; YOLOv8



**Citation:** Raiaan, M.A.K.; Fahad, N.M.; Chowdhury, S.; Sutradhar, D.; Mihad, S.S.; Islam, M.M. IoT-Based Object-Detection System to Safeguard Endangered Animals and Bolster Agricultural Farm Security. *Future Internet* **2023**, *15*, 372. <https://doi.org/10.3390/fi15120372>

Academic Editor: Gianluigi Ferrari

Received: 21 August 2023

Revised: 18 September 2023

Accepted: 24 October 2023

Published: 21 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The rapid expansion of industries into forested areas has had a profound and long-lasting impact on the dynamics between animals and surrounding communities [1,2]. Animals are being increasingly pushed closer to human settlements, leading to potential conflicts and posing significant challenges for farmers [3]. These conflicts result in crop destruction, encroachment on farmland and a loss of productivity. Necessitating increased financial resources for recovery after such damages occur [4]. Traditionally, farmers have relied on electric fences to prevent animals from entering their fields, aiming to mitigate the effects of these challenges [5]. However, it is crucial to recognize that while the primary objective is safeguarding crops and maintaining productivity, ensuring humans' and animals' health and safety remains the utmost priority in this complex coexistence [6]. In our research, we resolve the challenges posed by industrial expansion into forested areas by implementing an IoT-based detection model linked to cloud computing infrastructure and specialized hardware solutions. This integrated approach enables real-time monitoring

of animal movements, efficient data analysis and the automatic activation of deterrent devices, ensuring the protection of crops and the safety of humans and wildlife in areas where agriculture and animal habitats are prone to conflict.

Despite its importance, the detection and identification of animals have yet to be thoroughly investigated promptly [7]. Implementing an advanced surveillance system capable of automatically monitoring the area and detecting animal presence is becoming crucial to address this issue effectively [8]. The development of various technical alternatives for intrusion detection has frequently surpassed the limitations associated with traditional methods. One such system is intrusion detection, which focuses on identifying moving targets in the surrounding area [9].

Moreover, the IoT and other sensing technologies have revolutionized environmental monitoring and natural safeguarding efforts. These low-cost, pervasive technologies allow for the collection and analysis of data in real time, making them valuable instruments for protecting natural ecosystems and biodiversity [10]. IoT systems and technologies such as Wireless Sensor Networks (WSNs) and Unmanned Aerial Vehicles (UAVs) are now widely used to monitor and protect the environment, including wildlife conservation, forest fire detection and climate change monitoring. These technologies provide real-time data, facilitating prompt response and proactive environmental protection measures [10]. Computer vision has gained immense popularity for its classification capabilities and robust detection capabilities [11]. In recent years, several studies [3,12–16] involving Internet of Things (IoT) [17] and deep learning (DL) [18] have been conducted to detect animals that can damage farmers' crops. In their research, they utilized machine learning (ML) and deep learning (DL) techniques; however, most of the previous studies focused solely on detecting moving objects without considering the safety of animals. To address this limitation and prioritize animal welfare, our research aims to develop an animal-detection system that protects crops and ensures endangered animals' well-being.

This research focuses on developing an animal detector that activates an alarm system when animals approach fields, prompting them to retreat and eliminating the need for harmful measures like killing. The project workflow starts with preparing a dataset, where we collect images or recordings of endangered animals near agricultural areas and label them accordingly. Next, we deployed a model to detect animals accurately and save the best-performing model weights for future hardware implementation. An IoT system is then created to detect and notify farmers and animals, promoting ecological balance. Finally, via a performance analysis, the system's effectiveness in protecting crops and ensuring animal safety is evaluated. By adopting this approach, a comprehensive solution safeguards agricultural productivity and fosters a harmonious coexistence between farming communities and wildlife.

The significant contributions of the research are as follows:

- A comprehensive dataset is compiled, comprising high-resolution images of endangered animals near agricultural areas. The images of those species hold significance in maintaining an ecological balance within biodiversity while also posing a threat to agricultural lands and they are primarily collected.
- The dataset has been meticulously annotated with bounding boxes and labels, enabling the creation of a robust object-detection model.
- An optimized version of the object-detection model, YOLOv8, has been introduced. It demonstrates remarkable reliability in recognizing and identifying animals, even under adverse environmental conditions. Hyperparameters such as batch size, learning rate and epoch number have been fine-tuned to enhance the stability of the model.
- The animal-detection model's integration into IoT systems allows for real-time surveillance of agricultural areas, providing swift identification of animal intrusions.
- To mitigate potential conflicts, the system activates an alarm when animals approach the fields, prompting them to retreat without resorting to harmful measures. This ensures the safety of both humans and animals.

- The cloud-based systems and communication protocol swiftly relay messages to farmers about animal intrusions, enabling prompt response and effective crop protection.

This research is organized as follows: Section 2 presents a comprehensive review of the existing research concerning the detection of endangered animals through the utilization of different hybrid methodologies and a proper gap analysis is also performed for a better understanding of the limitations of the state-of-the-art works. Section 3 contains a detailed discussion of the research methodology, including introducing an optimized YOLOv8 object-detection model and integrating our IoT system. Section 4 summarizes the result section of the research employing different performance metrics to assess our technique's robustness by integrating a runtime cost analysis and comparison with state-of-the-art models. Section 5 broadly discusses the proposed system's novelty and Section 6 illustrates the limitations and future scope of the study. Finally, the research is concluded in Section 7.

## 2. Related Works

Animal encroachment is causing extensive damage to agricultural lands and crops worldwide. Various technologies have already been implemented to mitigate the intrusion of animals. The integration of artificial intelligence with IoT setups has played a significant role in tackling this problem. In this section, we discuss existing research and technologies that have been employed for the detection and prevention of endangered and harmful animals.

In the research by K. Balakrishna et al. [12], an IoT-based application with ML integration is presented. The researchers created an image dataset comprising five different animals to conduct their experiment. Their proposed R-CNN model outperformed other methods, achieving an impressive mean average precision of 85.22%.

In another study, Simla et al. [13] utilized IoT devices and deep learning (DL) with a machine-to-machine (M2M) communication protocol to develop an intruder-detection system. They introduced a modified convolutional neural network (CNN) architecture to recognize animals. The IoT application captured the animals' motion and sent a message to the farm owners using the M2M protocol.

Radhakrishnan et al. [14] employed image-processing techniques and ML algorithms to detect animal intrusion in agricultural fields. They extracted Gabor features from the images and utilized support vector machines (SVMs) for classification. Their proposed approaches achieved a remarkable accuracy of 99.48%.

Additionally, Mamat et al. [3] developed an animal-intrusion-detection system using the YOLOv8 model. They conducted experiments using a dataset of animal images and employed the CSP network as the backbone of the model's architecture. The proposed approach demonstrated its effectiveness by achieving an accuracy of 95%.

Meena et al. [15] introduced a hybrid algorithm to detect unwanted wildlife intrusion efficiently. They combined the YOLOv8 with the layer from CNN architecture to develop this algorithm. The proposed system was evaluated on the collected dataset and obtained an accuracy of 92.5%.

A study conducted by Priya et al. [19] developed a deep ensembled model to detect intruders on agricultural land. The primary focus was to enhance the security and sustainability of agriculture. They integrated different networks and layers to build the proposed object-detection model. The model outperforms other architectures by achieving precision and recall values of 97% and 96%, respectively.

In another study, Bapat et al. [20] developed an application using wireless sensor networks to divert animal intrusions. They integrated multiple hardware equipment items with sensors to effectively prevent animal intrusion. The application was tested in the lab and its robustness was enhanced to improve the reliability of the farm.

Moreover, Varun et al. [16] designed a framework to detect animal intrusion from the field and also classify intruders. They deployed a hybrid CNN model in the system and experimented with this model on three datasets, namely ATRW, ADID and Google V6+. They achieved the best accuracies of 92%, 99.6% and 95.6% for the corresponding dataset.

Ravoor et al. [21] integrated computer vision and deep learning techniques to design an animal-intrusion-detection system. They utilized a Pi camera to capture the movement of the images and deployed MobileNetV2 architecture for the animals' identification. The research uses three animal species and yields accuracies of 80%, 89.47% and 92.56%, respectively.

Finally, Ding et al. [22] utilized deep learning techniques to identify wild animal species. They used the fusion weight of the bidirectional feature pyramid network and channel attention to propose an improvised version of YOLOv8. The model effectively outperforms the benchmark network by obtaining an accuracy of 95.5% for the test data.

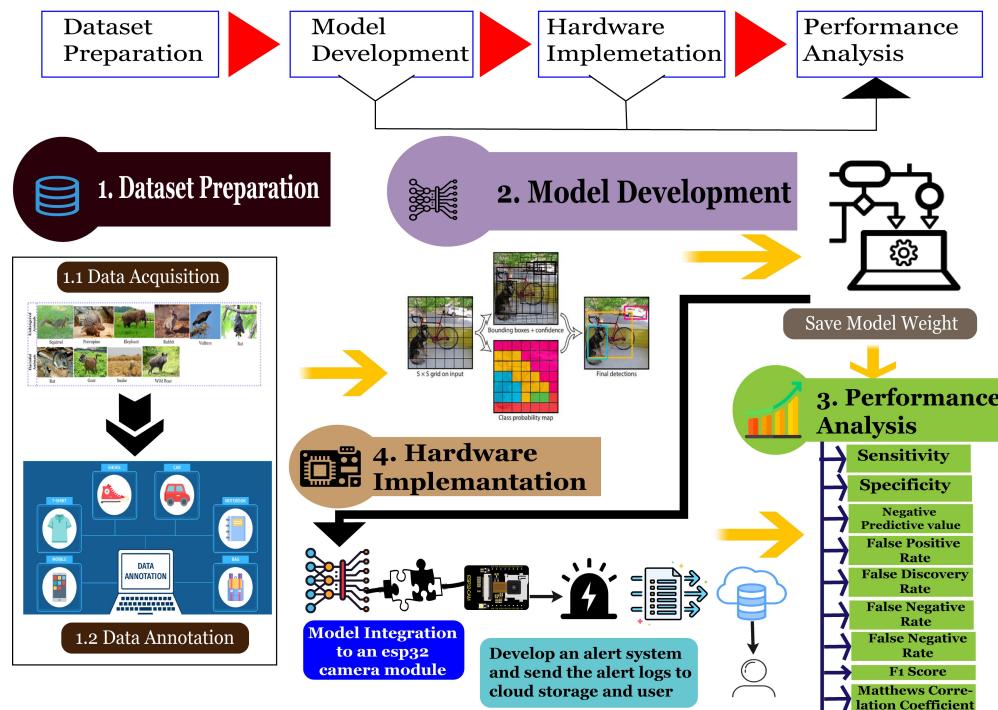
The limitations of the existing research are depicted in Table 1 and our research aims to address these limitations by focusing on filling those gaps. The objective is to develop a system that promotes ecological balance between endangered animals and humans.

**Table 1.** Comprehensive Analysis of Literature Gaps.

Reference	Dataset Employed	Identified Gaps
[12]	Animal Images Dataset	The study lacks the identification and consideration of rare species within the employed dataset. Additionally, the sustainability of the deployed models and their integration with Internet of Things (IoT) devices are not addressed.
[13]	Private Dataset	The research fails to provide individual class classification for animals in the utilized dataset. Furthermore, the value and significance of protecting rare animal species are not adequately emphasized.
[14]	Field Image Dataset	The study lacks the capability to detect animal intrusions and does not incorporate an alarm system for safeguarding crops and animals. Furthermore, the potential impact of extinct animals on ecosystems is not discussed.
[3]	Bird Images Dataset	The research focuses solely on bird detection across various ecosystems, neglecting discussions on extinct animal species and lacking integration with hardware components.
[15]	General Image Dataset	The study's scope is limited to animal classification, without offering a comprehensive framework for crop protection. Additionally, the importance of extinct animal preservation is not given due emphasis.
[23]	Image Data Representation	While predicting crop pathology, the research overlooks the protection mechanisms for crops and lacks discourse on the preservation of extinct animal species.
[20]	Animal Images Compilation	The study focuses solely on animal intrusion detection, lacking a real-time monitoring system and neglecting the significant aspects of rare species within ecosystems.
[16]	Customized Image Dataset	The research is centered on animal classification, without implementing real-time alerts for agricultural environments. Furthermore, the crucial subjects of biodiversity and protection of rare species are not adequately covered.
[21]	Wild-Animal Camera-Trap Data	The study does not adequately address real-time monitoring requirements and lacks provisions for alerting against animal intrusions. Additionally, the absence of hardware setup is notable.
[24]	Whale Images Dataset	The research confines its focus to detecting and classifying whales, disregarding other extinct animal species. Moreover, no real-time alarming system is discussed in the context of the study.

### 3. Methodology

Figure 1 presents this research's workflow diagram, organized into four distinct sections.



**Figure 1.** Workflow of the research methodology.

Figure 1 presents an illustrative depiction of the workflow diagram of this study. The first stage of the process entails obtaining images of endangered animals and animals that cause damage to crops. These images are then manually annotated to train the object-detection model. In the second stage, an object-detection model is introduced and the parameters are fine-tuned to obtain the optimal configuration of the model and improve its performance. The resulting model's weight is then stored for implementation on hardware. The hardware phase comprises implementing an alarm system that utilizes a cloud-based approach, effectively notifying farmers about the presence of the animals and the potential risks to their crops. Finally, this study assesses the performance of the system by employing metrics such as sensitivity, recall, positive rate, etc. These metrics determine the efficacy of this comprehensive solution.

#### 3.1. Dataset Description

This research has a great novelty in simultaneously detecting endangered animals and harmful animals for crops. But such datasets are not publicly available. Consequently, a new dataset has been introduced in this experiment. According to the perspective of our region, the endangered species of our biodiversity were consciously figured out. Along with this, we also listed the common animals that attack the crop fields occasionally. We comprehensively collected 2362 high-resolution images of eight species of animals in various habitats. The dataset includes at least 400 images of each species. The species are squirrel (*Sciurus vulgaris*), indian crested porcupine (*Hystrix indica*), wild elephant (*Elephas maximus*), hispid hare (*Caprolagus hispidus*), rat (*Rattus*), bat (*Chiroptera*), vulture and goat (*Capra aegagrus hircus*). These animals are all threatened by various factors, including habitat loss, poaching and climate change. The ecological balance of our planet depends on the survival of all species, including these threatened animals. Table 2 includes all the information of our dataset.

**Table 2.** Dataset description.

Risk Criteria	Class Name	# of Images	Source	Format
Endangered	Indian Crested Porcupine	500		
	Wild Elephant	500		
	Squirrel	462		
	Wild Boar	400	Google, Freepik, Unsplash, Pixabay, Pexels	JPG, JPEG
	Vulture	500		
Deleterious	Bat	500		
	Rat	500		
	Goat	500		

### 3.2. Model: YOLOv8

Deep learning has significantly advanced image classification, detection and segmentation [11]. The CNN-based deep learning model is exceptionally adept at accurately detecting and classifying images. This research primarily focuses on recognizing extinct animals and wildlife animals surrounding an agriculture field. To address this, an object-detection model, YOLOv8 was deployed using the deep learning architecture. YOLOv8 presents the most recent advancement in YOLO models, demonstrating state-of-the-art capabilities in object detection [25]. It is comprised of mainly three components, the backbone, neck and output layer. There is a prevailing assumption that the YOLO-v8 model will prioritize the deployment of limited edge devices with an emphasis on achieving high-inference speed [26]. Figure 2 illustrates the model of this research.

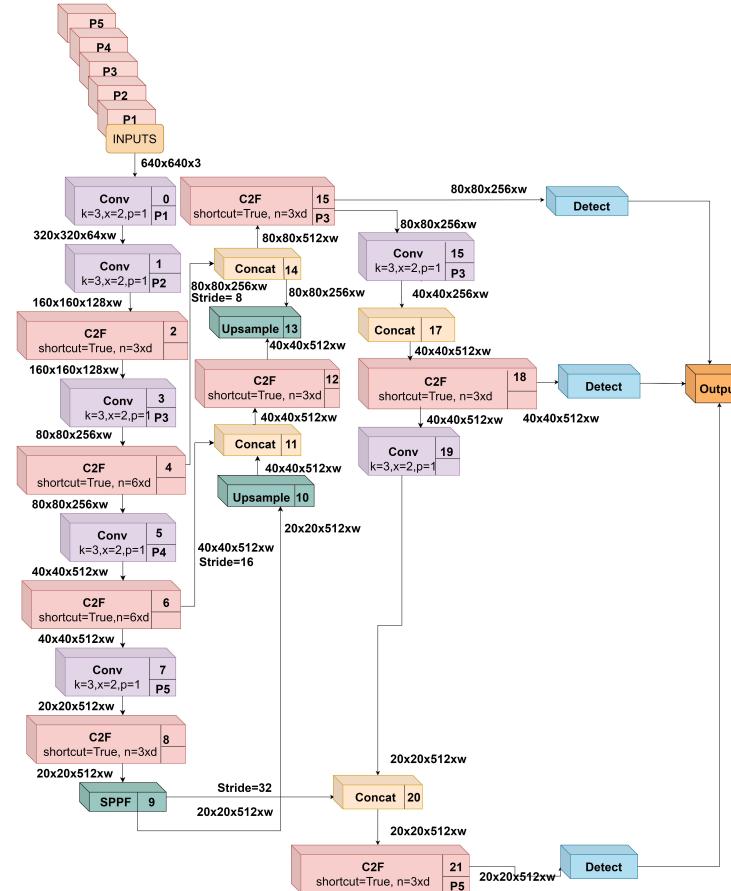
**Figure 2.** The architecture of the YOLOv8 model.

Figure 2 presents the architectural description of this research. The proposed model has an input layer taking the input image dimensions  $640 \times 640 \times 3$ . It consists of 9 convolutional layers for feature extraction and among these convolutional layers, there are 8 CSP layers incorporated to enhance the feature-extraction process and simplify information flow. YOLOv8 employs a comparable underlying architecture to YOLOv5, though with certain modifications made to the CSPLayer, which has been renamed the C2f module. The C2f module, referred to as the cross-stage partial bottleneck with two convolutions, was designed to improve the accuracy of detection by incorporating contextual information with high-level features. The CSPDarknet53 is used as a backbone which is further utilized to extract features from the images [27]. It creates a complementary low-resolution image and moves from pixel to pixel for feature extraction. A Cross-Stage Partial (CSP) module creates a features map from the extracting features and minimizes the duplication of information to enhance the model scalability. It also optimizes the process for a large-scale network. It uses a spatial pyramid pooling (SPP) [28] network to distinguish the contextual features and increase the receptive field in the final segment.

After completing the feature-extraction process, the model includes a series of sampling and concatenation processes, which enhance its capability to identify animals. Next, the neck structure performs the feature-enhancement tasks. It employs an attention-based pyramid network to improve the feature-extraction ability. For this reason, the same object with different categories and shapes can easily be recognized [29]. Consequently, the overall detection accuracy is increased substantially. Finally, the model's output layer makes predictions, which consist of bounding boxes that outline the detected animals.

The YOLOv8 algorithm incorporates the CIoU [30] and DFL [31] loss functions to calculate the bounding box loss, while binary cross-entropy is employed for classification loss. The aforementioned losses have demonstrated enhanced performance in the task of object detection, particularly in scenarios involving tiny objects.

An additional contribution of the proposed work involves the development of a precisely refined and optimized detection model. Various hyperparameter settings were utilized in our study, including adjustments to the number of layers, batch size and learning rate, in order to optimize the results. The epoch numbers of 80, 100 and 150 were modified and the outcome of the experiment was evaluated based on the completion of 100 epochs. Furthermore, it has been shown that the optimal learning rate for the model is established at 0.001, while the ideal batch size is identified as 16. The activation function used in the model is leaky relu and sigmoid. The leaky relu is used for corresponding hidden layers and softmax is used in the final output layer to detect the accurate label of the object. This setup provides the system with the capability to achieve outstanding reliability in detecting and classifying animals, even under challenging conditions.

Algorithm 1 explains the process of optimized YOLOv8 detection model.

---

**Algorithm 1:** Optimized YOLOv8 Object Detection

---

**Data:** Input image

**Result:** Detected objects with bounding boxes and class labels

Load pre-trained weights;

Preprocess input image;

Pass image through backbone network;

Extract features at multiple scales;

**for each scale do**

Apply detection head to predict boxes, class probabilities and objectness scores;  
Filter out low-confidence detections;

Apply non-maximum suppression to remove duplicate detections;

---

### 3.3. IoT Systems

By leveraging the power of the model, we seamlessly integrated the ESP32-CAM and buzzer to craft an innovative IoT system. The ESP32-CAM acted as the central hub, efficiently capturing and transmitting real-time video and image data to our network. This allows us to monitor and control devices remotely, enhancing the system's versatility. The buzzer played a vital role in providing audio feedback and alerts, enabling timely notifications for critical events.

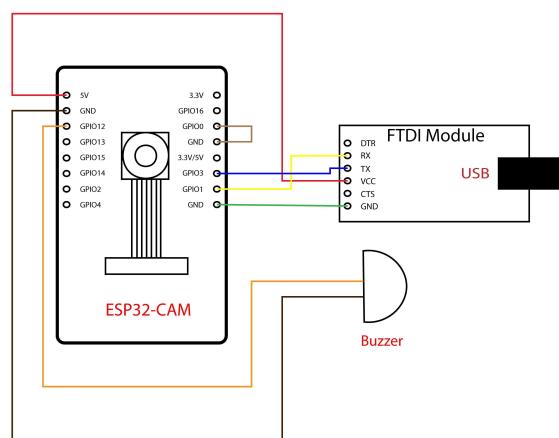
#### (1) Hardware Components

**ESP32-CAM:** The ESP32-CAM is a highly versatile and compact development board [32], bringing together an ESP32 microcontroller module, an OV2640 Camera Module with 2MP resolution and a small 802.11b/g/n Wi-Fi BT SoC module. This powerful combination allows for a wide range of applications, including the exciting possibility of face recognition [33] and object detection [34]. With its minimal  $40 \times 27$  mm footprint, the ESP32-CAM can function independently, making it an excellent choice for projects requiring a small and self-contained system. The convenience of its DIP package and features like GPIO pins and a microSD card slot further enhance its utility. Additionally, programming the ESP32-CAM is made easy using the Arduino IDE with the ESP32 core installed.

**Buzzer:** A buzzer is a fundamental electronic sound-producing device that finds applications in a wide array of scenarios, such as alarms, timers, notifications and electronic games. By being driven by a DC voltage, this small PCB mountable 5V active device generates a simple, continuous and often monotonous sound. It typically consists of two positive and negative pins, allowing for straightforward integration into various circuits and systems.

**Future Technology Devices International Limited (FTDI) module:** ESP32-CAM does not have a program chip. To program this board, an FTDI [35] to TTL module was used. This device handles all serial data communication to the board.

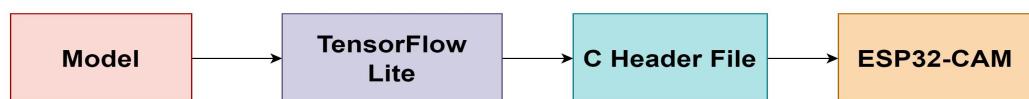
Figure 3 depicts the diagram of our proposed IoT system.



**Figure 3.** Hardware diagram of ESP32-CAM and buzzer.

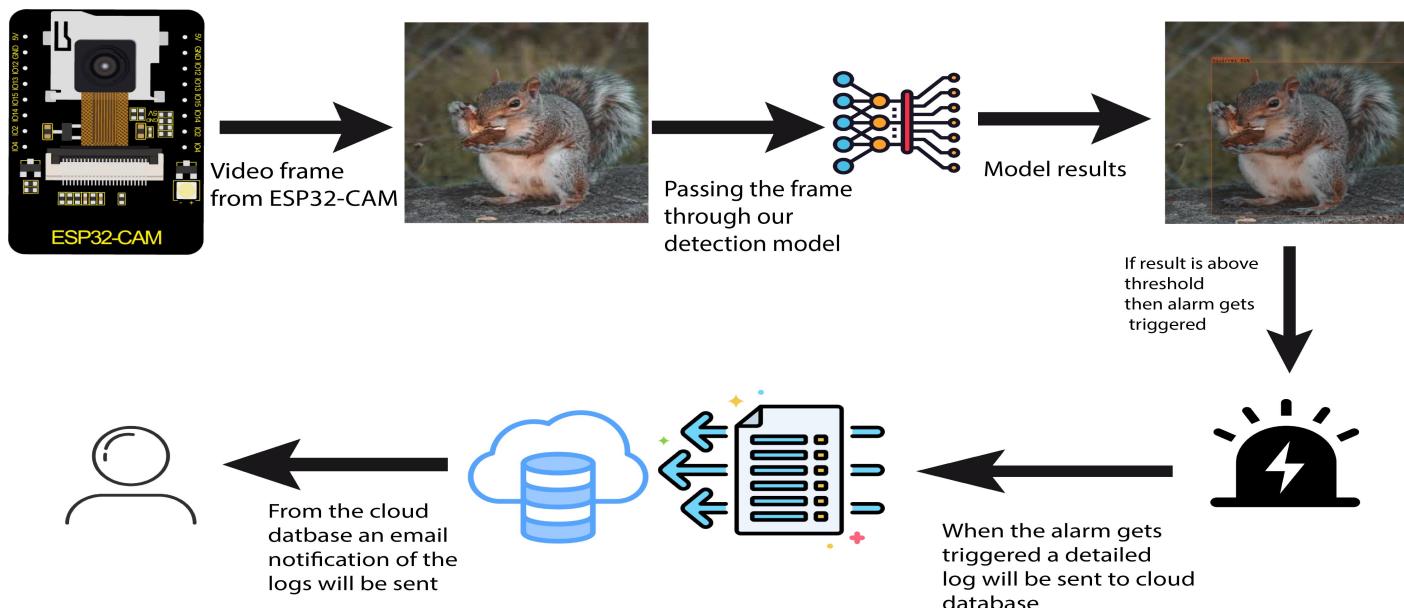
#### (2) Hardware implementation

For hardware implementation, the ESP32-CAM module was used. Since the board does not have a program chip, an FTDI module was connected to the ESP32-CAM in order to program it. Figure 3 shows the pin diagram of our IoT system. A modified YOLOv8 model was trained with 2362 images. Since the ESP32-CAM module has limited memory, the model needed to be as light as possible. Therefore, the model was converted into TensorFlow (TF) Lite format and that TFLite model was converted to a C array header file in order to implement it in ESP32-CAM. Figure 4 depicts the proposed IoT system's implementation overview.



**Figure 4.** Implementation of IoT device.

Figure 5 displays the workflow of the IoT device.



**Figure 5.** Workflow of IoT device.

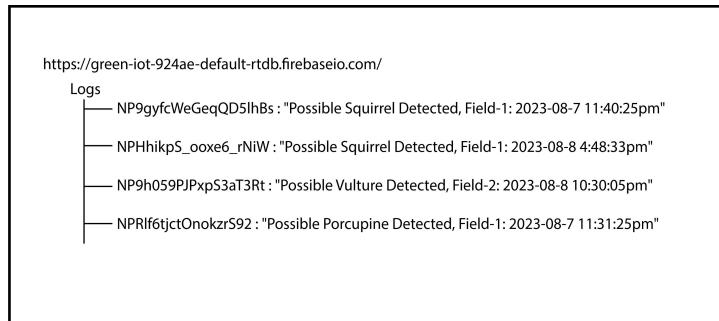
The incoming frames from the ESP32-CAM are resized to  $800 \times 600$  pixels with JPEG pixel formatting to ensure compatibility with the trained model. After preprocessing, the frames are routed through a C array, which facilitates the detection of endangered animals within the frame. If the detection confidence is greater than 80%, the buzzer trigger is activated, activating the alarm system to notify stakeholders of the presence of an endangered animal on the farm. This procedure is repeated for every frame captured by the ESP32-CAM, with the trained model used to determine the presence of endangered animals in the scene. Recognizing that a single device might not adequately cover a vast agricultural field, the possibility of integrating multiple IoT devices is raised. Multiple autonomously operating devices would be dispersed across the field in this scenario. Any device detecting an endangered animal within its field of view would independently activate an alarm, notifying the appropriate parties of the potential threat as soon as possible.

### (3) Cloud

**Firebase Cloud:** Firebase is a real-time cloud-hosted NoSQL database platform that Google develops [36,37]. It lets users store and sync data in real-time across multiple end devices. Firebase also provides secure storage for content such as images and videos. The data are encrypted at rest and in transit. Firebase Cloud is designed to simplify the development process and eliminate the need for extensive backend infrastructure management [38].

**Cloud implementation:** For storing the logs of alarms, we used the firebase real-time database. Firebase is a free-to-use database platform developed by Google. Since the ESP32-CAM module has Wi-Fi 802.11b/g/n/e/i, it will be connected to a nearby Wi-Fi access point. When the buzzer is triggered, the ESP32-CAM module will send a detailed message to the Firebase database. This message contains the data of the animal that was detected, the timestamp of the sighting and the location of where the animal was detected. When a log is updated in the database, an email will be sent to the user to notify them about the alert. The email will also contain the entire log message in detail.

Figure 6 displays the example logs of the Firebase database.



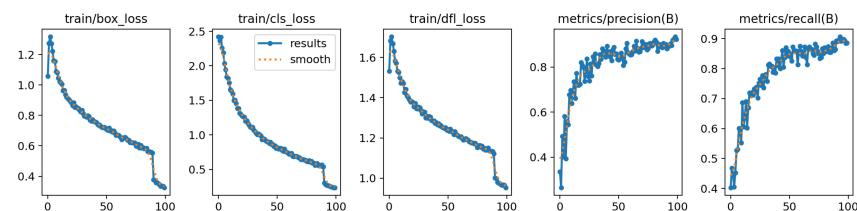
**Figure 6.** Example of database logs.

#### 4. Result Analysis of Detection Model

In this section, the analysis of the results of the detection model is presented.

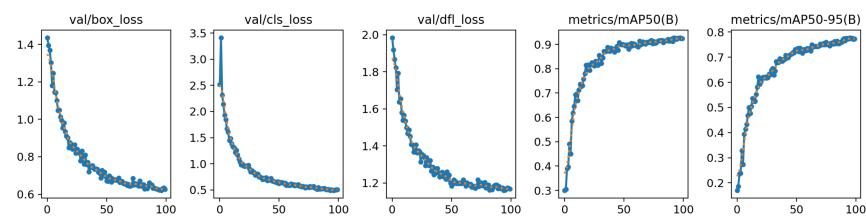
##### 4.1. Training Phase

Figure 7 illustrates various loss curves and precision and recall metrics to evaluate the effectiveness of the proposed object-detection models. The three-loss curve, which consists of ‘box loss’, ‘class loss’ and distributive focal loss, also known as ‘dfl loss’, refers to the model’s loss in terms of bounding box, class and the data imbalance problem. The model is run for 100 epochs. At the beginning of the epoch, the model had the highest loss value. The maximum box loss was greater than 1.2 at the outset of the epoch and it began to decrease as it progressed. After 100 epochs, the model possessed a minimum loss value of 0.1. After 100 epochs, class loss and dfl loss tended to decrease to 0.1 from their initial maximums of 2.5 and 1.8, respectively. The mean average precision (mAP) method evaluates the efficacy of object-identification models, whereas the recall curve is used to classify object detection as a true positive [39]. During the early epochs, the mAP precision and recall curves may have a minimum value of 0. As the training process advances and the model’s parameters endure improvements, the model’s capacity to accurately identify objects is enhanced, resulting in an increase in mean average precision (mAP). The initial precision and recall curve values are 0 and 0.4, respectively, and increase to 1 after 100 epochs.



**Figure 7.** Box loss, class Loss, DFL loss, map during training phase during model training.

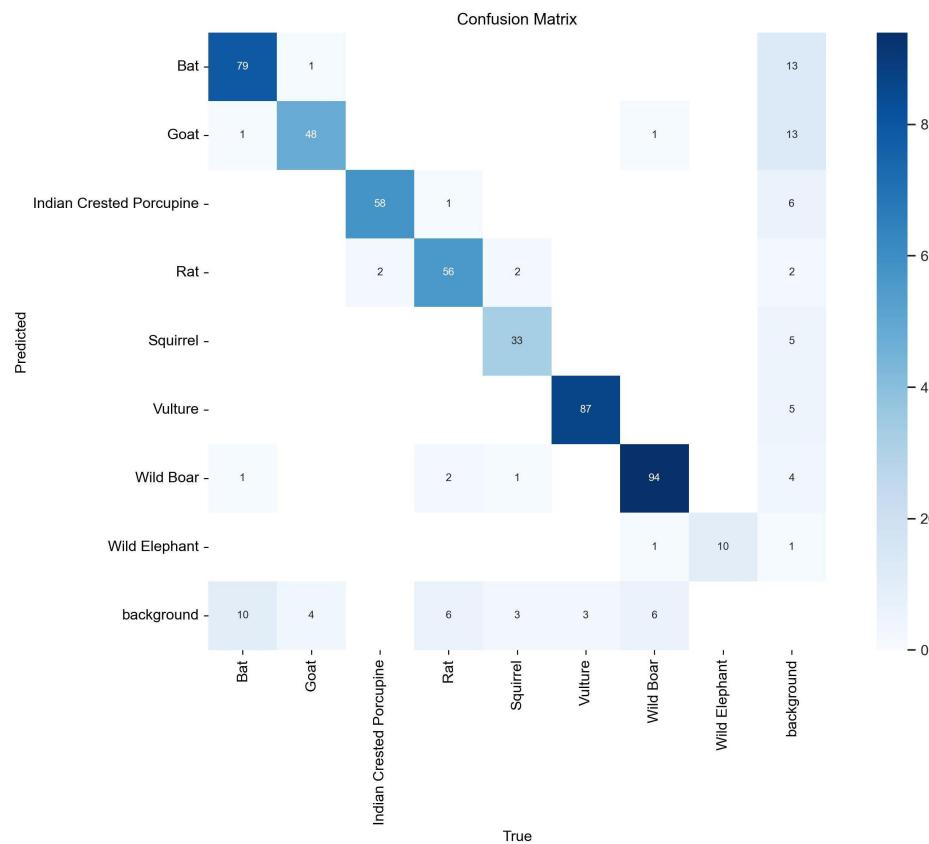
The precision–recall and loss function curves are fully depicted in Figure 8’s visual representation of the model’s validation phases. The curves display similar trends in the validation phase as in the training phase. In order to assess how well the model generalizes to new data, it is essential to consider the validation loss. The loss curve revealed a recurrent pattern where the model first showed the highest loss and increased performance by gradually lowering the loss throughout 100 epochs. The maximum box loss value observed was 1.5 at the beginning of the period, which then decreased to 0.6 upon completion of all epochs. At the beginning, the maximum class loss and dfl loss values were 3.5 and 2, respectively. After all 100 epochs were complete, these values gradually decreased to 0.5 and below 1.2. The validation loss shows a decreasing trend over time, supporting the effectiveness of our approach. The precision and recall curve exhibits an initial value of 0, which progressively increases to 1 after 100 epochs during the validation phase. This observation indicates that the object-detection model has shown substantial progress in validating the unseen data.



**Figure 8.** Box loss, Class Loss, DFL loss, map during validation phase during model training.

#### 4.2. Testing Phase

The confusion matrix is a commonly employed tool for evaluating the performance of a model [40]. The evaluation assesses the performance of individual classes and presents various performance metrics of the model, including precision, recall, specificity, F1 score and others. This analysis examines the distribution of accurate and inaccurate predictions obtained from the model, aiding in the assessment of the model's reliability. Figure 9 depicts the confusion matrix generated for our suggested model. The matrix is additionally used to estimate several performance metrics, as indicated in Tables 3 and 4.



**Figure 9.** Confusion matrix of the detection model.

The F1 confidence curve provides insights about the model's predictions and helps to balance between the precision and recall [41]. Figure 10 depicts the F1 score of the model in relation to different confidence threshold values, as well as the confidence scores affiliated with its predictions. In our study, the F1 score is tuned at 89% to balance both precisions and recall. Moreover, this value was determined at a confidence threshold of 0.524, which was the optimal threshold. This observation suggests that the model has satisfactory performance in terms of precision and recall across all classes at the aforementioned value. The precision–recall curve is typically a curve that starts at a high value of precision and gradually decreases as recall increases [42]. It provides a trade-off between precision and recall for different classification thresholds. According to Figure 10, the model attained a

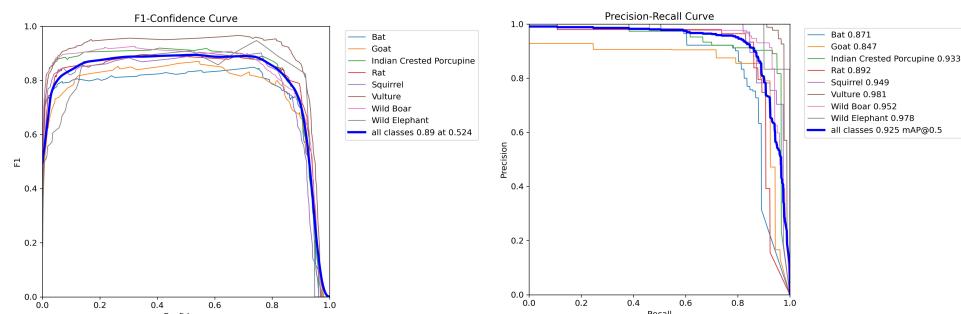
precision value of 98.1% for the vulture class, which is the highest value obtained. The total precision value across all classes is reported to be 92.5%, with a threshold set at 0.5.

**Table 3.** Performance metrics analysis for individual class.

Measure	Bat	Goat	Porcupine	Rat	Squirrel	Vulture	Wild Boar	Elephant
Sensitivity	98.75	96	98.31	93.33	100	100	95.92	90.91
Specificity	99.5	99.7	99.41	99.28	99.26	99.74	99.74	100
Precision	97.53	97.96	96.67	94.92	91.67	98.86	98.95	100
Negative Predictive Value	99.75	99.39	99.7	99.05	100	100	98.96	99.77
False Positive Rate	0.5	0.3	0.59	0.72	0.74	0.26	0.26	0
False Discovery Rate	2.47	2.04	3.33	5.08	8.33	1.14	1.05	0
False Negative Rate	1.25	4	1.69	6.67	0	0	4.08	9.09
F1 Score	98.14	96.97	97.48	94.12	95.65	99.43	97.41	95.24
Matthews Correlation Coefficient	97.76	96.52	97.04	93.29	95.39	99.3	96.77	95.24

**Table 4.** Overall Performance Analysis of the Proposed Model.

Measure	Value (%)
Sensitivity	96.65
Specificity	99.57
Precision	97.07
Negative Predictive Value	99.57
False Positive Rate	0.42
False Discovery Rate	2.93
False Negative Rate	3.34
F1 Score	96.8
Matthews Correlation Coefficient	96.41

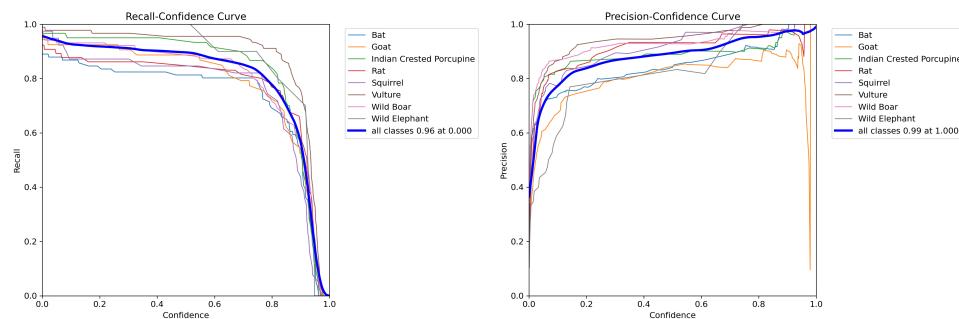


**Figure 10.** F1 confidence and precision–recall curve.

Precision confidence evaluates the accuracy of accurate positive predictions, highlighting the model's reliability [43]. On the other hand, recall confidence refers to the level of confidence associated with the model's capacity to identify all objects of a particular class [44]. The trade-off between them is critical because increasing precision reduces the number of alarms generated by the system but increases the risk of missing some detections. In contrast, increasing recall may increase the number of alarms, including some false ones, to minimize missed detections. Object-detection models assign each detected object a confidence score [45]. These scores are used as thresholds for determining the validity of detections. By conducting a comparison analysis of precision and recall metrics across various confidence score thresholds, it is feasible to enhance the model's decision-making process and effectively monitor the effectiveness of the object-detection model over time.

Figure 11 depicts a precision–confidence and a recall–confidence graph. The recall–confidence graph illustrates the correlation between recall and confidence thresholds within the context of a detection task. In addition, it provides valuable insights into the intrinsic trade-off between recall and precision during the adjustment of the confidence threshold.

Figure 11 shows that the model obtains a recall value of 96% at a very low threshold value. This implies that the model successfully detects positive cases across all classes. In a similar manner, the model demonstrates its robustness by achieving a maximum precision value of 99% when the confidence threshold is set to the highest value of 1.



**Figure 11.** Recall–confidence and precision–confidence curve.

The performance assessment of a classification model on several animal species is summarized in Table 3. The measurements provide valuable insights into the model’s ability to distinguish between various species effectively. The model has a notable level of sensitivity, as evidenced by the high sensitivity for bats (98.75%) and porcupines (98.31%). These results indicate the model’s proficiency in accurately identifying positive cases. The model demonstrates the ability to accurately detect negative instances, as indicated by the specificity metric. Specifically, the wild boar class exhibits a specificity value of 95.92%, while the elephant class exhibits a specificity value of 90.91%. The precision of the model remains consistently high for all categories, particularly for goat (97.96%) and squirrel (98.86%), indicating its ability to generate accurate positive predictions. This table also demonstrates the model’s inclination towards developing adverse predictions, as indicated by the high unfavorable predictive values observed for most species. The research exhibits a commendable achievement in achieving low rates of false positives, revealing just a limited number of cases where predictions were incorrectly positive. The results of this study illustrate the model’s high level of precision and reliability when it comes to detecting as well as classifying various animal species.

Table 4 contains the key performance indicators for a detection model within a research endeavor. These metrics comprehensively evaluate the model’s ability to distinguish between diverse groups. The model’s sensitivity of 96.65% demonstrates how well it can identify pleasurable events. With a specificity rating of 99.57%, the model can identify unfavorable circumstances. The model’s ability to make accurate optimistic predictions is illustrated by the model’s accuracy rate of 97.07%. The model’s high negative predictive value of 99.57% demonstrates how accurately it predicts adverse outcomes. False positive occurrences are exceedingly uncommon, as indicated by their low frequency (0.42%). With a mistaken discovery rate of 2.93%, false positive predictions are distinctive. The false negative rate of 3.34% represents instances in which true positives were neglected. Due to its low value, the F1 score of 9.68% should be interpreted cautiously; further investigation may be required. The 96.41% Matthews Correlation Coefficient demonstrates the overall correlation between the model’s predictions and the actual outcomes. These metrics support the comprehensive evaluation of the model’s performance in the research environment by providing an in-depth analysis of the model’s benefits and drawbacks.

The model achieves a remarkable average precision of 94% for the “Indian Crested Porcupine” class on both the validation and test sets. This high score demonstrates the model’s ability to predict this class’s instances accurately. The model’s performance remains strong across different classes, with average precision scores ranging from 86% to 94%. Particularly noteworthy is the fact that both the “Indian Crested Porcupine” and “Wild Boar” classes achieve the highest precision score of 94%, showcasing the model’s proficiency in identifying these classes correctly. The model exhibits outstanding capabilities in this

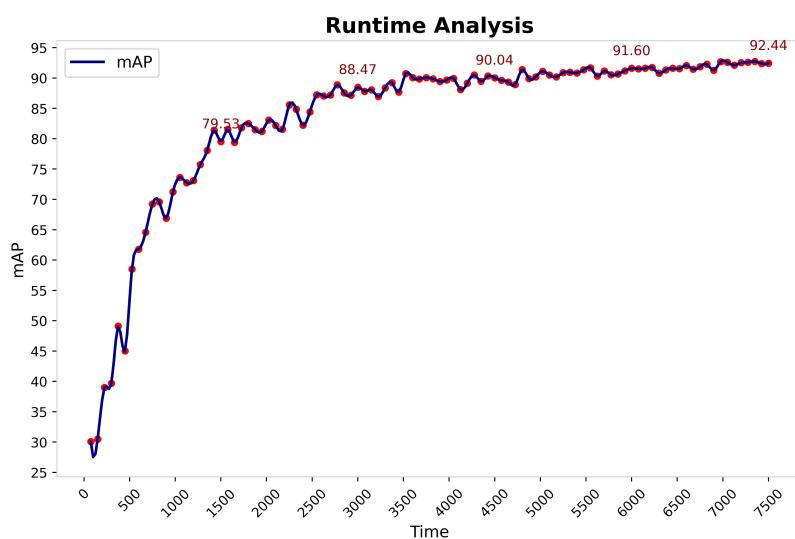
detection task by analyzing the overall performance. With an average precision of 90% on the validation set and 89% on the test set for all classes combined, the model demonstrates high accuracy in classifying instances accurately. Furthermore, the comparable performance of the model on the test set to that on the validation set indicates its ability to generalize well to unseen data. This aspect highlights the model's robustness and demonstrates that it has been well trained.

Therefore, it is evident that our model has proven to be robust, efficient and valuable for detecting endangered animals and giving extra protection to farms.

#### 4.3. Runtime Cost Analysis

In this section, we analyze the runtime of our proposed optimized YOLOv8 model.

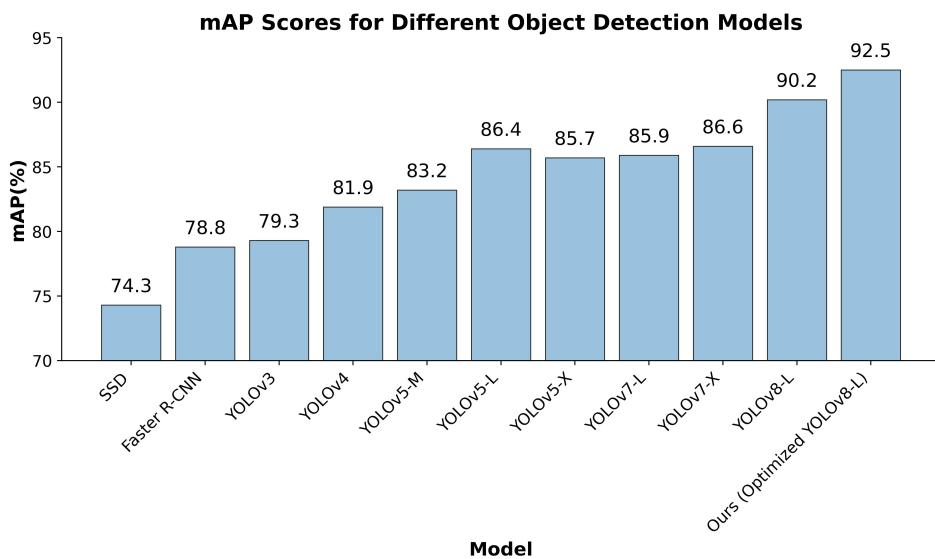
Figure 12 depicts the runtime analysis of the experiments, as mentioned earlier. Investigating runtime complexity is vital to this study as it refers to the model's efficacy in processing data and making predictions within a suitable timetable. In this research, we trained our object-detection model for 100 epochs, where each epoch needed 75 s to be completed. This research utilized an AMD Ryzen 5 5600X 6-core Central Processing Unit (CPU) and 16 GB of RAM for all the experiments. It is paired with a Graphical Processing Unit (GPU) named ZOTAC GAMING GeForce RTX 3060 Twin Edge OC GDDR6 with 12 GB video RAM (VRAM). It was previously stated that, for a single epoch, the device required 75 s; for 100 epochs, the total time stands at 7500 s, just exceeding 2 h. Since we conducted the experiment on a good number of images, this shortest training duration shows the efficiency of the proposed model. Moreover, optimizing the model facilitates convergence while maintaining the model's mAP, which exhibited consistent improvement with each epoch. In addition to the hardware setup, the optimal configuration of this model played a significant role in enhancing the efficiency of our research. Also, it highlights the possibility of timely and precise object detection for practical applications.



**Figure 12.** Runtime analysis of the optimized YOLOv8 model.

#### 4.4. Comparison with State-of-the-Art Models

In this section, we conducted a performance comparison of our proposed model with several state-of-the-art object-detection models, including Single Shot Detectors (SSDs), faster regions with convolutional neural networks (R-CNN), YOLOv3, YOLOv4, YOLOv5-M, YOLOv5-L, YOLOv5-X, YOLOv7-L, YOLOv7-X and our optimized version of YOLOv8-L. We evaluated their mAP performance, with higher values indicating more precise object detection. The objective was to identify each model's strengths and limitations and highlight our proposed model's exceptional performance. Figure 13 illustrates the comparison of our models with other object-detection models.



**Figure 13.** Comparison with state-of-the-art models.

SSD emerged with the lowest mAP score of 74.3% among the models we evaluated, indicating its relatively inferior performance in endangered-animal-detection tasks. The next fastest R-CNN had a modest mAP of 78.8%, a marginal improvement. The mAP scores of YOLOv3 and YOLOv4 were 79.3% and 81.9%, respectively, reflecting their moderate effectiveness in this context. The YOLOv5 family of models (YOLOv5-M, YOLOv5-L and YOLOv5-X) achieved mAP scores ranging from 83.2% to 85.7%, demonstrating their capacity to improve object-detection precision. This trend was reinforced by YOLOv7-L and YOLOv7-X, which had mAP values of 85.9% and 86.6%, respectively. With a mAP score of 90.2%, YOLOv8-L performed better in our evaluation than its predecessors and the majority of other models. Nonetheless, the “Optimized YOLOv8-L” model attained the most outstanding mAP score (92.44%), representing the pinnacle of our investigation. This significant margin of improvement over other models demonstrates our optimization efforts’ efficacy, highlighting our proposed model’s robustness in detecting objects within images.

Therefore, our performance comparison revealed that our optimized YOLOv8-L model emerged as the top performer, outperforming all baseline models with an exceptional mAP score of 92.44%, indicating the robustness of our model.

## 5. Discussion

This research aims to design an object-detection model integrated with an alarm system to reduce human–wildlife conflict in agricultural environments. The primary objective of our study is to implement conservation strategies for endangered animal species while concurrently mitigating the negative impact of wildlife invasions on crop fields. The significance of this study lies in the concern observed among farmers, who are increasingly resorting to killing endangered animals to prevent their intrusion. But, to maintain the ecological balance, the role of these animals is beyond description. Consequently, this research emphasizes formulating a proficient solution to yield advantages for farmers and endangered wildlife. Existing research has discussed the challenges associated with human–wildlife conflict, reinforcing the necessity for safe approaches. However, none of them emphasized the significance of conserving endangered animals.

Moreover, previous methodologies have frequently proven inadequate because of constraints in detecting and generating real-time alarms. This research overcomes all these limitations substantially. We compiled a dataset of eight species of animals by acquiring images of the animals on the verge of extinction and commonly harming crops. To effectively train the model, we meticulously annotated all the images. The YOLOv8 model

was employed for object detection in our study. The model's parameters were meticulously tuned to maximize its performance. The outcome of this optimized model demonstrates its efficacy and stability by obtaining an impressive mean average precision (mAP) of 92.44%. Various performance metrics such as precision–confidence, recall–confidence and F1–confidence figures are illustrated to assess the model's ability to detect objects properly. The aforementioned findings provide evidence of the model's robustness and applicability in real-world implementations. The optimal weight of the model was stored to implement the hardware component. The hardware component involves a variety of devices such as cameras, buzzers and sensors and makes a proper alarm system. The images are subsequently subjected to processing by our object-detection model in order to ascertain the presence of any potential threats. In the case in which a positive detection occurs, an alarm is activated to notify farmers or necessary authorities, facilitating prompt response and intervention. The deployment of a cloud-based approach for storing alarm logs provides several benefits, including improved data accessibility and collaboration.

Additionally, our system enables real-time updates and offers cost-efficiency. Integrating hardware components with the object-detection system facilitates a comprehensive and practical strategy for addressing the issue of human–wildlife conflict while concurrently reducing the adverse impact on endangered animals and crop fields. In summary, this study allows real-time monitoring investigation and aims to safeguard endangered animals and protect crops.

## 6. Limitation and Future Scope

Acknowledging and addressing certain limitations within this research is crucial in developing an object-detection model for recognizing endangered and hazardous animals in crop fields. There are some limitations identified in this research. The first one is a constraint with sample size. The study focuses extensively on eight species of animals that are primarily on the verge of extinction and cause agricultural damage. Still, many species could make the model more effective at detecting different animals. In addition, the experiments did not consider the influence of environmental factors, such as temperature, lighting, terrain, etc. Furthermore, the experiments were conducted exclusively with RGB images.

To address these limitations, we proposed future works that will increase the prediction accuracy of the proposed system. In particular, providing detailed explanations and exploring possible paths for further development and research might yield significant insights. The inclusion of a broader range of animal species in the dataset will enhance the model's efficacy across various geographical locations. Moreover, the application of multi-domain learning will be utilized to improve the capacity for generalization. Incorporating weather and seasonal data, such as rainfall and temperature, will provide valuable insights into the growth stages of crops and the frequency of animal intervention within a specific time frame. The use of hyperspectral, multispectral or satellite images will increase the sustainability of this system, as it will enable the detection of animals using UAVs or other compatible devices. This research will focus on integrating video input to conduct temporal analysis to utilize motion and temporal patterns observed in video footage to enhance detection capabilities and minimize false positive rates. Additionally, the implementation of edge computing enables efficient execution on embedded devices such as Raspberry Pi deployed throughout the farm, hence facilitating the generation of alarms with little delay.

## 7. Conclusions

An innovative and unique approach to tackle the threats posed by the extinction of animal species and its impact on agricultural farms has been introduced in this research. The research successfully achieves real-time detection and classification of endangered and harmful animals in farming using a sophisticated object-detection system that combines the ESP32-CAM and the YOLOv8 model. The system's impressive performance is evident through its efficiency and accuracy, with a mean average precision (mAP) of 92.44% and a sensitivity rate of 96.65% on the unseen test dataset. The significance of this research

lies in its dual purpose: promoting the conservation of endangered species and mitigating the agricultural damage caused by these animals. By providing farmers with real-time warnings when animals approach their fields, the proposed system enables a humane approach to protect crops, eliminating the need for harmful and lethal measures against the animals. This research approach fosters a harmonious coexistence between farming communities and wildlife, ensuring the safety and well-being of humans and endangered animals. The research's comprehensive workflow, from data collection and model training to implementing the IoT system, showcases a holistic solution that preserves agricultural productivity and ecological balance. This research represents a crucial stride towards sustainable agriculture and wildlife conservation, addressing the challenges posed by the encroachment of human settlements into natural habitats. The novel contributions of this work lie in seamlessly integrating deep learning object-detection models and IoT systems to develop an animal-detection system that prioritizes animal welfare while safeguarding crops. The resulting system benefits farmers by protecting their livelihoods and contributes to the preservation of endangered species, thereby promoting a balanced ecosystem.

**Author Contributions:** Conceptualization, M.A.K.R., N.M.F., D.S. and M.M.I.; methodology, M.A.K.R., N.M.F., S.C., D.S. and S.S.M.; validation, M.A.K.R., N.M.F. and M.M.I.; formal analysis, M.A.K.R., N.M.F., S.C., D.S. and S.S.M.; investigation, M.A.K.R., N.M.F., S.C., D.S. and S.S.M.; writing—original draft preparation, M.A.K.R., N.M.F., S.C., D.S. and S.S.M.; writing—review and editing, M.M.I.; supervision, M.A.K.R. and M.M.I.; project administration, M.M.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is funded by the Institute for Advanced Research Publication Grant of the United International University, Ref. No.: IAR-2023-Pub-034.

**Data Availability Statement:** Dataset access is only granted for legitimate research purposes.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mi, X.; Feng, G.; Hu, Y.; Zhang, J.; Chen, L.; Corlett, R.T.; Hughes, A.C.; Pimm, S.; Schmid, B.; Shi, S.; et al. The global significance of biodiversity science in China: An overview. *Natl. Sci. Rev.* **2021**, *8*, nwab032. [[CrossRef](#)] [[PubMed](#)]
2. Liu, X.; Guo, W.; Feng, Q.; Wang, P. Spatial correlation, driving factors and dynamic spatial spillover of electricity consumption in China: A perspective on industry heterogeneity. *Energy* **2022**, *257*, 124756. [[CrossRef](#)]
3. Mamat, N.; Othman, M.F.; Yakub, F. Animal Intrusion Detection in Farming Area using YOLOv5 Approach. In Proceedings of the 2022 22nd International Conference on Control, Automation and Systems (ICCAS), Jeju, Republic of Korea, 27 November–1 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.
4. Manzoor, Z.; Ehsan, M.; Khan, M.B.; Manzoor, A.; Akhter, M.M.; Sohail, M.T.; Hussain, A.; Shafi, A.; Abu-Alam, T.; Abioui, M. Floods and flood management and its socio-economic impact on Pakistan: A review of the empirical literature. *Front. Environ. Sci.* **2022**, *10*, 2480. [[CrossRef](#)]
5. Vogel, S.M.; Songhurst, A.C.; McCulloch, G.; Stronza, A. Understanding farmers' reasons behind mitigation decisions is key in supporting their coexistence with wildlife. *People Nat.* **2022**, *4*, 1305–1318. [[CrossRef](#)]
6. Jeevitha, S.; Kumar, S.V. A study on sensor based animal intrusion alert system using image processing techniques. In Proceedings of the 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 12–14 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 20–23.
7. Fennell, M.; Beirne, C.; Burton, A.C. Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology. *Glob. Ecol. Conserv.* **2022**, *35*, e02104. [[CrossRef](#)]
8. Fuentes, A.; Han, S.; Nasir, M.F.; Park, J.; Yoon, S.; Park, D.S. Multiview Monitoring of Individual Cattle Behavior Based on Action Recognition in Closed Barns Using Deep Learning. *Animals* **2023**, *13*, 2020. [[CrossRef](#)]
9. Xue, W.; Jiang, T.; Shi, J. Animal intrusion detection based on convolutional neural network. In Proceedings of the 2017 17th International Symposium on Communications and Information Technologies (ISCIT), Cairns, QLD, Australia, 25–27 September 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–5.
10. Fascista, A.; Coluccia, A.; Ravazzi, C. A Unified Bayesian Framework for Joint Estimation and Anomaly Detection in Environmental Sensor Networks. *IEEE Access* **2022**, *11*, 227–248. [[CrossRef](#)]
11. Raiaan, M.A.K.; Fatema, K.; Khan, I.U.; Azam, S.; ur Rashid, M.R.; Mukta, M.S.H.; Jonkman, M.; De Boer, F. A Lightweight Robust Deep Learning Model Gained High Accuracy in Classifying a Wide Range of Diabetic Retinopathy Images. *IEEE Access* **2023**, *11*, 42361–42388. [[CrossRef](#)]

12. Balakrishna, K.; Mohammed, F.; Ullas, C.; Hema, C.; Sonakshi, S. Application of IOT and machine learning in crop protection against animal intrusion. *Glob. Transitions Proc.* **2021**, *2*, 169–174. [CrossRef]
13. Simla, A.J.; Chakravarthi, R.; Leo, L.M. Agricultural intrusion detection (AID) based on the internet of things and deep learning with the enhanced lightweight M2M protocol. *Soft Comput.* **2023**, *1*–12. [CrossRef]
14. Radhakrishnan, S.; Ramanathan, R. A support vector machine with Gabor features for animal intrusion detection in agriculture fields. *Procedia Comput. Sci.* **2018**, *143*, 493–501. [CrossRef]
15. Meena, D.; Jahnnavi, C.N.V.; Manasa, P.L.; Sheela, J. Efficient Wildlife Intrusion Detection System using Hybrid Algorithm. In Proceedings of the 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 21–23 September 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 536–542.
16. Sajithra Varun, S.; Nagarajan, G. DeepAID: A design of smart animal intrusion detection and classification using deep hybrid neural networks. *Soft Comput.* **2023**, *1*–12. Available online: [https://www.researchgate.net/publication/370948001\\_DeepAID\\_a\\_design\\_of\\_smart\\_animal\\_intrusion\\_detection\\_and\\_classification\\_using\\_deep\\_hybrid\\_neural\\_networks](https://www.researchgate.net/publication/370948001_DeepAID_a_design_of_smart_animal_intrusion_detection_and_classification_using_deep_hybrid_neural_networks) (accessed on 1 September 2023). [CrossRef]
17. Azam, Z.; Islam, M.M.; Huda, M.N. Comparative Analysis of Intrusion Detection Systems and Machine Learning Based Model Analysis through Decision Tree. *IEEE Access* **2023**, *11*, 80348–80391. [CrossRef]
18. Islam, M.M.; Bhuiyan, Z.A. An Integrated Scalable Framework for Cloud and IoT Based Green Healthcare System. *IEEE Access* **2023**, *11*, 22266–22282. [CrossRef]
19. Singh, P.; Krishnamurthi, R. Object detection using deep ensemble model for enhancing security towards sustainable agriculture. *Int. J. Inf. Technol.* **2023**, *1*–14. Available online: [https://www.researchgate.net/publication/371843834\\_Object\\_detection\\_using\\_deep\\_ensemble\\_model\\_for\\_enhancing\\_security\\_towards\\_sustainable\\_agriculture](https://www.researchgate.net/publication/371843834_Object_detection_using_deep_ensemble_model_for_enhancing_security_towards_sustainable_agriculture) (accessed on 1 September 2023). [CrossRef]
20. Bapat, V.; Kale, P.; Shinde, V.; Deshpande, N.; Shaligram, A. WSN application for crop protection to divert animal intrusions in the agricultural land. *Comput. Electron. Agric.* **2017**, *133*, 88–96. [CrossRef]
21. Ravoor, P.C.; Sudarshan, T.; Rangarajan, K. Digital Borders: Design of an Animal Intrusion Detection System Based on Deep Learning. In Proceedings of the International Conference on Computer Vision and Image Processing, Prayagraj, India, 4–6 December 2020; Springer: Singapore, 2020; pp. 186–200.
22. Ma, D.; Yang, J. Yolo-animal: An efficient wildlife detection network based on improved yolov5. In Proceedings of the 2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), Xi'an, China, 28–30 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 464–468.
23. Hassija, V.; Batra, S.; Chamola, V.; Anand, T.; Goyal, P.; Goyal, N.; Guizani, M. A blockchain and deep neural networks-based secure framework for enhanced crop protection. *Ad Hoc Netw.* **2021**, *119*, 102537. [CrossRef]
24. Shiu, Y.; Palmer, K.; Roch, M.A.; Fleishman, E.; Liu, X.; Nosal, E.M.; Helble, T.; Cholewiak, D.; Gillespie, D.; Klinck, H. Deep neural networks for automated detection of marine mammal species. *Sci. Rep.* **2020**, *10*, 607. [CrossRef]
25. Aboah, A.; Wang, B.; Bagci, U.; Adu-Gyamfi, Y. Real-time multi-class helmet violation detection using few-shot data sampling technique and Yolov8. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 5349–5357.
26. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* **2023**, *11*, 677. [CrossRef]
27. Terven, J.; Cordova-Esparza, D. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. *arXiv* **2023**, arXiv:2304.00501.
28. Tai, S.K.; Dewi, C.; Chen, R.C.; Liu, Y.T.; Jiang, X.; Yu, H. Deep learning for traffic sign recognition based on spatial pyramid pooling with scale analysis. *Appl. Sci.* **2020**, *10*, 6997. [CrossRef]
29. Sharma, N.; Baral, S.; Paing, M.P.; Chawuthai, R. Parking Time Violation Tracking Using YOLOv8 and Tracking Algorithms. *Sensors* **2023**, *23*, 5843. [CrossRef] [PubMed]
30. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
31. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
32. Dietz, H.; Abney, D.; Eberhart, P.; Santini, N.; Davis, W.; Wilson, E.; McKenzie, M. ESP32-Cam as a programmable camera research platform. *Imaging* **2022**, *232*. Available online: <http://aggregate.org/DIT/ei2022esp.pdf> (accessed on 1 September 2023). [CrossRef]
33. Kumar, S.; Sharma, K.; Raj, G.; Datta, D.; Ghosh, A. Arduino and ESP32-CAM-based automatic touchless attendance system. In Proceedings of the 3rd International Conference on Communication, Devices and Computing: ICCDC 2021, Haldia, India, 16–18 August 2021; Springer: Berlin/Heidelberg, Germany, 2022; pp. 135–144.
34. Mehendale, N. Object Detection using ESP 32 CAM. 2022. Available online: <https://www.studocu.com/vn/document/truong-dai-hoc-bach-khoa-ha-noi/vat-ly-i/ssrn-id4152378-wqequew/65206956> (accessed on 1 September 2023).
35. Bagchi, T.; Mahapatra, A.; Yadav, D.; Mishra, D.; Pandey, A.; Chandrasekhar, P.; Kumar, A. Intelligent security system based on face recognition and IoT. *Mater. Today Proc.* **2022**, *62*, 2133–2137. [CrossRef]

36. Tran, T.D.; Huynh, K.T.; Nguyen, P.Q.; Ly, T.N. AttendanceKit: A set of Role-Based Mobile Applications for Automatic Attendance Checking with UHF RFID Using Realtime Firebase and Face Recognition. In Proceedings of the International Conference on Future Data and Security Engineering, Ho Chi Minh City, Vietnam, 23–25 November 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 432–446.
37. Tripathy, S.; Das, S.S. Automated home using firebase and Google assistant. *J. Inf. Optim. Sci.* **2022**, *43*, 1021–1028. [[CrossRef](#)]
38. Calderoni, L.; Maio, D.; Tullini, L. Benchmarking cloud providers on serverless iot back-end infrastructures. *IEEE Internet Things J.* **2022**, *9*, 15255–15269. [[CrossRef](#)]
39. Wu, X.; Sahoo, D.; Hoi, S.C. Recent advances in deep learning for object detection. *Neurocomputing* **2020**, *396*, 39–64. [[CrossRef](#)]
40. Theissler, A.; Thomas, M.; Burch, M.; Gerschner, F. ConfusionVis: Comparative evaluation and selection of multi-class classifiers based on confusion matrices. *Knowl.-Based Syst.* **2022**, *247*, 108651. [[CrossRef](#)]
41. Zeng, X.; Hu, Y.; Shu, L.; Li, J.; Duan, H.; Shu, Q.; Li, H. Explainable machine-learning predictions for complications after pediatric congenital heart surgery. *Sci. Rep.* **2021**, *11*, 17244. [[CrossRef](#)]
42. Yang, J.; Chen, J.; Li, J.; Dai, S.; He, Y. An Improved Median Filter Based on YOLOv5 Applied to Electrochemiluminescence Image Denoising. *Electronics* **2023**, *12*, 1544. [[CrossRef](#)]
43. Padilla, R.; Netto, S.L.; Da Silva, E.A. A survey on performance metrics for object-detection algorithms. In Proceedings of the 2020 international conference on systems, signals and image processing (IWSSIP), Niteroi, Brazil, 1–3 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 237–242.
44. Boyd, K.; Eng, K.H.; Page, C.D. Area under the precision–recall curve: Point estimates and confidence intervals. In Proceedings of the Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, 23–27 September 2013; Proceedings, Part III 13; Springer: Berlin/Heidelberg, Germany, 2013; pp. 451–466.
45. Erhan, D.; Szegedy, C.; Toshev, A.; Anguelov, D. Scalable object detection using deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2147–2154.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.