



Article

Advanced Deep Learning Methods for Multiple Behavior Classification of Cage-Free Laying Hens

Sachin Subedi , Ramesh Bahadur Bist , Xiao Yang , Guoming Li and Lilong Chai *

Department of Poultry Science, College of Agricultural & Environmental Sciences, University of Georgia, Athens, GA 30602, USA

* Correspondence: lchai@uga.edu

Abstract: The welfare of hens in cage-free systems is closely linked to their behaviors, such as feeding, drinking, pecking, perching, bathing, preening, and foraging. To monitor these behaviors, we developed and evaluated deep learning models based on YOLO (You Only Look Once), an advanced object detection technology known for its high accuracy, speed, and compact size. Three YOLO-based models—YOLOv5s_BH, YOLOv5x_BH, and YOLOv7_BH—were created to track and classify the behaviors of laying hens in cage-free environments. A dataset comprising 1500 training images, 500 validation images, and 50 test images was used to train and validate the models. The models successfully detected poultry behaviors in test images with bounding boxes and objectness scores ranging from 0 to 1. Among the models, YOLOv5s_BH demonstrated superior performance, achieving a precision of 78.1%, surpassing YOLOv5x_BH and YOLOv7_BH by 1.9% and 2.2%, respectively. It also achieved a recall of 71.7%, outperforming YOLOv5x_BH and YOLOv7_BH by 1.9% and 2.8%, respectively. Additionally, YOLOv5s_BH recorded a mean average precision (mAP) of 74.6%, exceeding YOLOv5x_BH by 2.6% and YOLOv7_BH by 9%. While all models demonstrated high detection precision, their performance was influenced by factors such as stocking density, varying light conditions, and obstructions from equipment like drinking lines, perches, and feeders. This study highlights the potential for the automated monitoring of poultry behaviors in cage-free systems, offering valuable insights for producers.



Academic Editor: In-Bok Lee

Received: 30 November 2024

Revised: 11 January 2025

Accepted: 21 January 2025

Published: 23 January 2025

Citation: Subedi, S.; Bist, R.B.; Yang, X.; Li, G.; Chai, L. Advanced Deep Learning Methods for Multiple Behavior Classification of Cage-Free Laying Hens. *AgriEngineering* **2025**, *7*, 24. <https://doi.org/10.3390/agriengineering7020024>

Copyright: © 2025 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/>).

Keywords: poultry production; cage-free housing; animal welfare; precision management

1. Introduction

The traditional cage-based farming of laying hens suppresses animal natural behavior expression; as a result, poultry welfare policies and corresponding public concerns are increasing [1]. The way of raising birds is changing towards cage-free so that the birds can express normal behaviors like preening, foraging, perching, and dust bathing [2–5]. There is a pressing need to enhance agricultural methods for identifying anomalies in chicken behavior, health, and welfare while minimizing reliance on manual labor. To achieve this, the development and implementation of automated systems are essential. Advanced technologies like computers, sensors, cloud computing, machine learning (ML), and artificial intelligence (AI) are revolutionizing various industries by driving substantial improvements and efficiencies [6]. For instance, these technologies enhance the profitability and productivity of commercial poultry farming by reducing the reliance on human labor for routine bird monitoring [7]. There is an increasing trend in computer vision and deep learning techniques in solving various problems in hematology, cell biology or botany,

agriculture, and livestock. Computer vision combines mathematics and computer science to enable fully automated, non-invasive, real-time monitoring and process control in poultry production by analyzing raw images, extracting patterns, classifying them, and making predictions [8,9].

Six behaviors of laying hens, namely standing, sitting, sleeping, grooming, feeding, and drinking, were classified using deep learning [10], classifying the latter two behaviors with an accuracy of 89% [11]. The image classification methodology, developed to identify broiler breeders' behaviors using machine vision techniques, demonstrated a success rate of over 70% in recognizing behaviors such as wing spreading, bristling, drinking, scratching, resting, stretching, preening, and mounting [4]. In addition, flock behaviors (crowding near feeders) were recognized using a CNN architecture with an accuracy of 99.17% [12]. A machine vision method proved to be a practical tool for tracking abnormal behaviors in poultry, such as pecking [13]. The mask region-based convolutional neural network was developed for detecting preening behaviors in birds [14]. An improved deep learning model based on the YOLOv5 architecture was developed to monitor the spatial and floor patterns of cage-free hens, including the real-time tracking of bird numbers in zones like perching, feeding, drinking, and nesting, achieving an accuracy of 87% to 94% across all chicken ages and zones [15].

The automatic detection of poultry behaviors helps identify deviations from normal behavior patterns that will notify producers in real time [16]. The early detection of unexpected changes in activity and behaviors can benefit poultry's well-being and welfare. The objectives of this research were to develop a machine vision method to (1) identify poultry behaviors in cage-free housing systems, (2) evaluate its performance in research cage-free facilities, and (3) explore strategies to enhance the model's detection accuracy. In this study, we classify different poultry behaviors (Feeding, Drinking, Perching, Preening, Perching, Dust Bathing, and Foraging). We compare three deep learning models (YOLOv5s, YOLOv5x, and YOLOv7) from multiple behavior classification. YOLOv5s is fast and lightweight, making it suitable for real-time use, while YOLOv5x provides higher accuracy for more detailed tasks. YOLOv7, a newer version compared to YOLOv5, brings advanced features that improve both efficiency and accuracy, although even later versions now exist. These models were chosen to evaluate the balance between speed, accuracy, and resource use, which are important for behavior classification in real-time or limited-resource settings [17–19].

2. Materials and Methods

2.1. Experimental Station

The study was carried out in a cage-free layer facility located at the University of Georgia's poultry research farm in Athens, Georgia. The GPS coordinates of the research farm are approximately 33.90879373981556, −83.37708374133594. Approval for animal use and management was granted by the University's Institutional Animal Care and Use Committee (IACUC; AUP# A2020 08-014-A2). We housed 800 W36 Hy-Line hens across four separate cage-free rooms, assigning 200 birds to each room from the first day. Each room measured 7.3 m in length, 6.1 m in width, and stood 3 m high. Before introducing the hens, the floors were evenly covered with pine shavings at a depth of 5 cm, and the birds were provided with unlimited access to commercial feed. The management practices adhered to the guidelines for Hy-Line W-36 commercial layers. An automated environmental control system maintained the rearing conditions, regulating the air temperature between 21 and 23 °C and ensuring a light intensity of 30 lux during a photoperiod of 19 h light and 5 h dark. Additionally, the growth of the hens and the environmental param-

eters were monitored daily in accordance with the Standard Operating Procedures of the UGA Poultry Research Center.

2.2. Dataset Collection and Preparation

We installed six PRO-1080MSB night-vision network cameras from Swann Communications USA Inc. (Figure 1; Santa Fe Springs, CA, USA) strategically above the drinking stations, feeders, perches, and on walls approximately three meters high to capture both overhead and lateral video footage. The continuous monitoring of various hen behaviors and activities was performed, with all video data being recorded on DVR-4580 digital video recorders from the same manufacturer. The recordings were saved in .avi format at a resolution of 1920×1080 pixels and a frame rate of 15 frames per second, then converted to .jpg image files using Free Video to JPG Converter software (version 5.0).



Figure 1. Imaging system and data collection.

2.3. Definition of Laying Hens' Behaviors and Labeling

To develop a behavior detection and classification system (Figure 2), we manually annotated seven distinct activities of laying hens—namely Feeding, Drinking, Perching, Preening, Dust Bathing, Pecking, and Foraging—by drawing bounding boxes as specified in the ethogram outlined in Table 1. This annotation process resulted in a comprehensive dataset of 2050 images, which were divided into 1500 images for training, 500 for validation, and 50 for testing, encompassing hens aged between 22 and 28 weeks. Utilizing the open-source platform Makesense.AI, we meticulously created bounding boxes around each behavior of interest within the images. The dataset was then organized into separate directories for training and validation, each containing subdirectories labeled “Images” and “Labels”. Ultimately, the annotations were saved in .txt format files, preparing the dataset for effective model development.

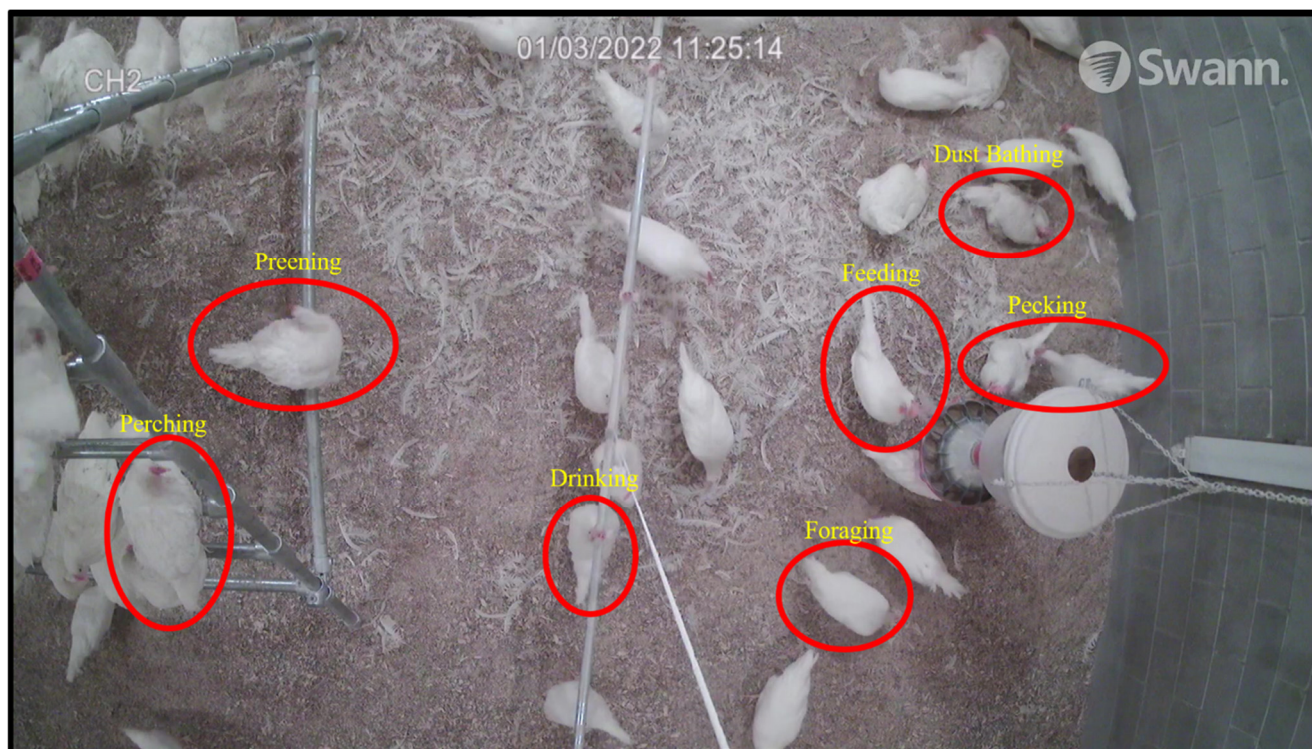


Figure 2. Different behaviors identified in cage-free laying hens.

Table 1. Different behaviors of laying hen.

Behavior	Description	References
Feeding	Birds approach feeders for feeding	[20,21]
Drinking	Birds approach drinkers to drink	[21]
Perching	Birds roost on the elevated structure	[22]
Preening	Birds bend and twist their bodies to access their uropygial glands, using their beaks to clean and groom their feathers	[15]
Dust Bathing	Birds crouch down to bath in the litter and use their wings to throw dust	[5]
Pecking	Birds peck at the feathers of another bird	[23]
Foraging	Birds scratch or peck the ground	[24]

2.4. You Only Look Once (YOLO) Model for Poultry Image Analysis

Introduced in 2015 by researchers Joseph Redmon and Ali Farad, YOLO (You Only Look Once) is a pioneering single-stage object detection framework. Unlike the R-CNN family and its faster variants, YOLO leverages single stage detectors (SSDs) to achieve superior speed and precision by eliminating the need for a region proposal network [25]. Over the years, YOLO has evolved into various iterations, including TinyYOLO, YOLOv2 through YOLOv8, YOLOx, and scaled YOLOv7, each incorporating different backend enhancements. Among these, YOLOv3 remains the most prevalent in industry applications, while the Ultralytics implementations of YOLOv5 and YOLOv7 are also widely adopted [26]. Pretrained YOLO models, especially those trained on the extensive COCO dataset—which encompasses object detection, segmentation, keypoint detection, and captioning tasks—are readily available and straightforward to deploy. When applied to datasets focused on poultry behaviors, these models are annotated with bounding boxes and image segmentation

labels for specific activities such as Feeding, Drinking, Perching, Preening, Pecking, Dust Bathing, and Foraging. In our enhanced model, images are collected from diverse locations and scales, prioritizing regions with high detection confidence to improve tracking and identification accuracy. A significant advantage of YOLO over traditional classifier-based systems is its capability to analyze the entire image in a single pass, facilitating efficient and accurate predictions [27]. The process begins by dividing the input image into an $S \times S$ grid. For each grid cell, K anchor boxes are generated, each predicting six values: the coordinates (x, y) and dimensions (w, h) of the bounding box, a confidence score, and the probability (Pr) of each behavior category. Here, Pr indicates the likelihood that a bounding box contains the target behavior, while the confidence score reflects both the presence of an object and the accuracy of the bounding box prediction. The confidence score is calculated as the product of Pr and the intersection over union (IOU) between the predicted box and the ground truth [28]. If a ground truth object resides within a grid cell, the corresponding probability is set to 1; otherwise, it is 0. To refine the predictions and eliminate redundant bounding boxes, the model employs non-maximum suppression. This ensures that only the most accurate and relevant detections are retained. Consequently, each grid cell can produce multiple bounding boxes, each associated with a specific behavior probability, enhancing the model's ability to accurately identify and classify various poultry behaviors within diverse and complex environments.

2.5. Architecture of YOLOv5_BH Model (YOLOv5s_BH and YOLOv5x_BH)

We tailored the YOLOv5_BH and YOLOv5x_BH models by modifying the YOLOv5s and YOLOv5x architectures, which are structured into three key components (Figure 3): backbone, neck, and detection head [29]. The input stage of YOLOv5s is optimized with mosaic data augmentation, adaptive anchor boxes, and dynamic image resizing, enhancing feature extraction and boosting dataset performance. The backbone employs CSPDarkNet53 to extract features at four different scales, corresponding to behaviors such as Feeding, Drinking, Perching, Preening, Pecking, Dust Bathing, and Foraging [26,30]. This backbone integrates the Focus, C3, and SPP modules: the Focus module reduces feature dimensions through image slicing, the C3 module enhances computational efficiency and lowers parameter complexity with a bottleneck design, and the SPP module captures multi-scale information by pooling and merging feature maps of varying sizes [29]. These refined features are then passed to the Neck, which combines up-sampling and down-sampling techniques to create a feature pyramid, thereby improving object detection accuracy by refining the features extracted by the backbone [31,32]. The Neck utilizes Feature pyramid networks (FPNs) and path aggregation networks (PANs) to effectively merge features, transferring high-level semantic information to lower layers and enhancing localization by propagating detailed features upwards [30,32]. This enriched feature set is forwarded to the detection head, which employs a 1×1 convolutional layer to classify and predict outputs, producing three distinct results per batch for final target detection [29]. The detection head generates outputs including class probabilities, object confidence scores, and bounding box coordinates. Additionally, the architecture incorporates Convolution-BatchNorm-Leaky ReLU (CBL) modules for convolution, normalization, and activation, and utilizes Cross-Stage Partial (CSP) networks within both the backbone and neck to enhance inference speed and maintain accuracy by reducing the overall model size [33]. The Spatial Pyramid Pooling (SPP) module further enhances feature integration by applying max pooling with various kernel sizes and concatenating the resulting feature maps, thereby unifying multiple scales into a single, cohesive representation [34]. This comprehensive design ensures efficient feature extraction, integration, and accurate object detection, optimizing the model's performance for diverse applications.

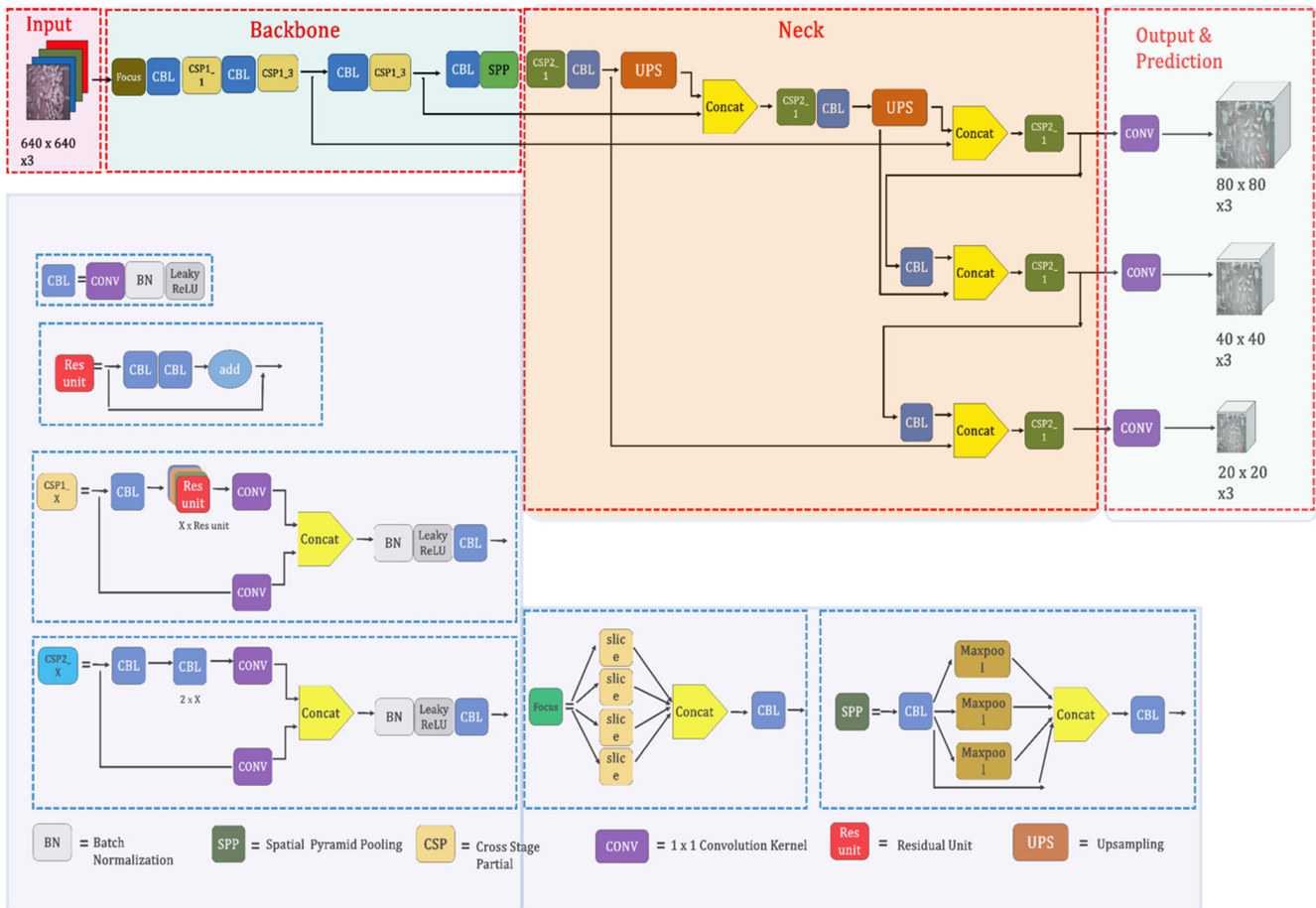


Figure 3. YOLOv5_BH architecture was used for this study.

2.6. Architecture of YOLOv7_BH Model

YOLOv7, launched in July 2022 by the original YOLOv4 team, enhances single-stage object detection by building on YOLOv4, scaled YOLOv4, and YOLO-R architectures and integrating YOLOv5’s mosaic data augmentation, effectively balancing parameter count and computational efficiency and achieving high speed and accuracy in detecting small objects [19,26]. YOLOv7_BH incorporates several critical enhancements that elevate its performance, including the efficient E-ELAN module, advanced re-parametrization techniques, refined label assignment strategies, and an innovative auxiliary head training approach. Architecturally, it features an expanded ELAN (E-ELAN) module within its backbone that employs expansion, shuffling, and cardinality merging to boost learning capacity without disrupting gradient flows, while group convolutions and compound model scaling optimize channel capacity and feature diversity for superior overall performance [35]. The input stage of YOLOv7_BH involves preprocessing steps such as mixup and mosaic augmentation, followed by resizing images to 640 pixels before feeding them into the backbone [35]. The backbone integrates BConv layers, E-ELAN modules, and MP layers. Each BConv layer includes convolution operations, batch normalization (BN), and activation functions [20]. These components work together to extract rich, multi-scale features from the input images [36]. After the backbone, YOLOv7_BH’s Neck utilizes Spatial Pyramid Pooling and Cross-Stage Partial Channel (SPPCSPC) networks to further enhance and integrate features [36]. This involves applying multiple pooling operations with varying kernel sizes to capture multi-scale information, which are then merged using the PANet structure through up-sampling and down-sampling, resulting in feature maps of different sizes that are subsequently processed by the Head. The Head of YOLOv7_BH incorporates

reparameterized RepConv layers for enhanced recognition and classification, along with an auxiliary head that assists in loss calculation during training [35]. This auxiliary head introduces multi-way branching, which improves training efficiency and model performance without increasing computational costs. Additionally, YOLOv7 utilizes bag-of-freebies (BoF) strategies—techniques that enhance model performance without additional training expenses—to further optimize its detection capabilities [19]. Overall, YOLOv7’s comprehensive design, featuring advanced modules and innovative training strategies, ensures robust and accurate object detection across diverse applications, including poultry behavior detection (Figure 4). Its ability to efficiently process and integrate multi-scale features, combined with optimized computational strategies, makes YOLOv7 a leading choice for real-time object detection tasks.

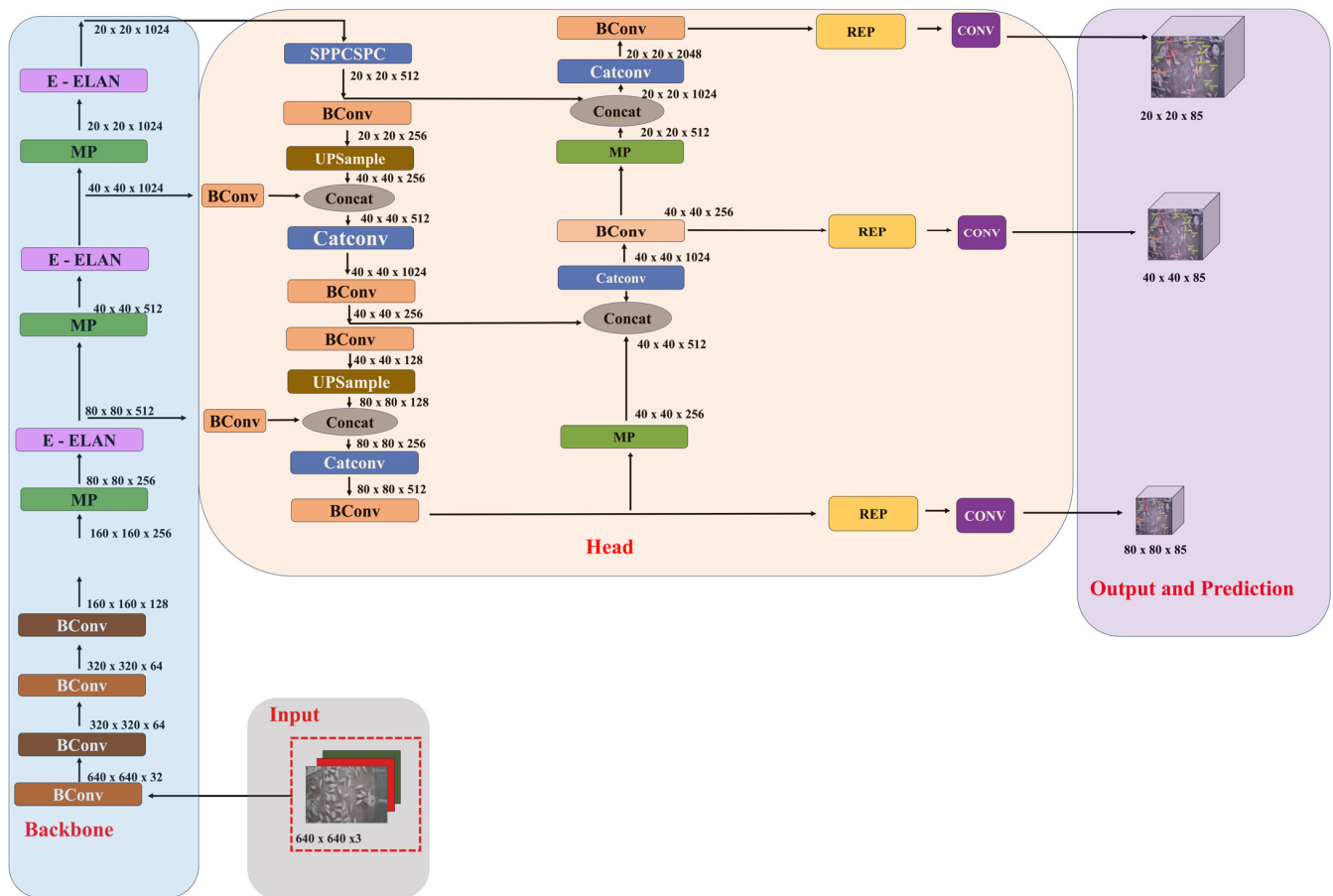


Figure 4. YOLOv7_BH architecture was used for monitoring chicken behaviors in this study.

2.7. Detection Training and Validation

The input to YOLOv5_BH and YOLOv7_BH models includes images resized to 640×640 pixels [YOLO], normalized to the range $[0, 1]$ [37] and annotated with YOLO-compatible bounding boxes stored in .txt files. We trained the behavior dataset utilizing pre-trained COCO weights, continuously monitoring loss metrics throughout the training process. Each of the three models was configured with a tailored “yaml” file that included image data and adjusted labels for both training and validation phases [38]. The training process used a batch size of 16, an initial learning rate of 0.001 and a total of 1000 epochs. The training environment was established on Oracle Linux 7, equipped with a Tesla V100-SXM2 GPU. Development was carried out using Python 3.8 and PyTorch 2.0 for implementing deep learning procedures. The system operated with NVIDIA driver version 510.108.03 and CUDA version 11.6.

2.8. Evaluation Metrics

To assess the performance of the model, we utilized precision, recall, and mean average precision (mAP) as key evaluation indicators.

Precision measures the accuracy of the positive predictions made by the model, defined as the ratio of correctly identified positive instances (behaviors: Feeding, Drinking, Perching, Preening, Dust Bathing, Pecking, and Foraging) to the total number of positive predictions.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

Recall quantifies the model's ability to identify all relevant positive instances. It is calculated by dividing the number of correctly detected positive instances by the total number of actual positive instances in the dataset—behaviors (Feeding, Drinking, Perching, Preening, Dust Bathing, Pecking and Foraging).

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

Mean average precision (mAP) provides a comprehensive measure of the model's accuracy across different classes. It is computed by summing the precision values at each recall level and multiplying by the incremental change in recall.

$$mAP = \sum_{i=1}^N P(i) * \Delta R(i) \quad (3)$$

$P(i)$ is the precision, and $\Delta R(i)$ is the change in recall from the i th detection. For all the above metrics, a value closer to 100% reflects the better performance of the detectors.

Higher values for precision, recall, and mAP, approaching 100%, indicate the superior performance of the detection model.

3. Results and Discussion

3.1. Model Performance in Computing System Use

Figure 5 illustrates the attributes of YOLOv5s, YOLOv5x, and YOLOv7 models, including their usage of GPU memory, model size, and the duration of their training time. Regarding the duration of the training process, YOLOv5s_BH required nearly 5 h, whereas YOLOv5x_BH and YOLOv7_BH had a longer training time of 12.5 h and 14 h, respectively. YOLOv5x_BH is the biggest model, measuring 173.1 mb, while YOLOv7_BH comes in second with a size of 74.8 mb, and YOLOv5s_BH is the smallest with a size of 14.4 mb. YOLOv5s_BH stands out as the most efficient model regarding both time and GPU usage, but it has the smallest size among the three models. YOLOv5x_BH, on the other hand, is the largest model with the highest GPU usage, while YOLOv7_BH requires less GPU usage compared to YOLOv5x_BH. This information is crucial as it can affect the selection of computer systems and the speed of data analysis.

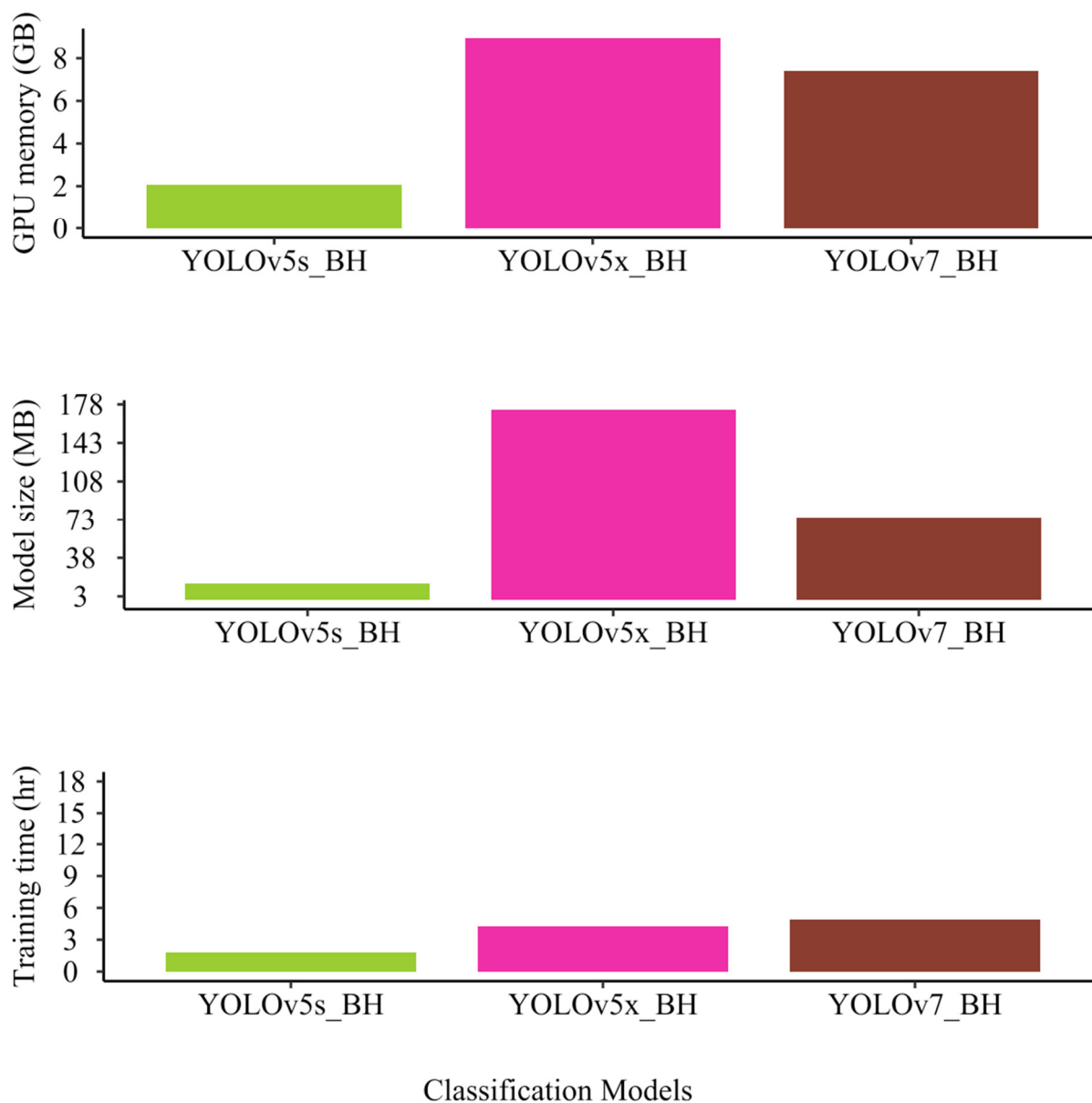


Figure 5. YOLOv5s_BH, YOLOv5x_BH, and YOLO7_BH models in memory size, model size, and training time.

3.2. Performance of Behavior Detection Models in Data Training and Validation

(a) Total training loss

All three models' full training loss function decreased while running in 1000 epochs (Figure 6). The total loss in the YOLO model is the sum of three loss functions (box/localization loss, classification loss, and objectness loss) used to optimize the model's performance. In addition, the total loss is used to update the model weights during back-propagation to minimize the loss and improve the model's accuracy on the training data.

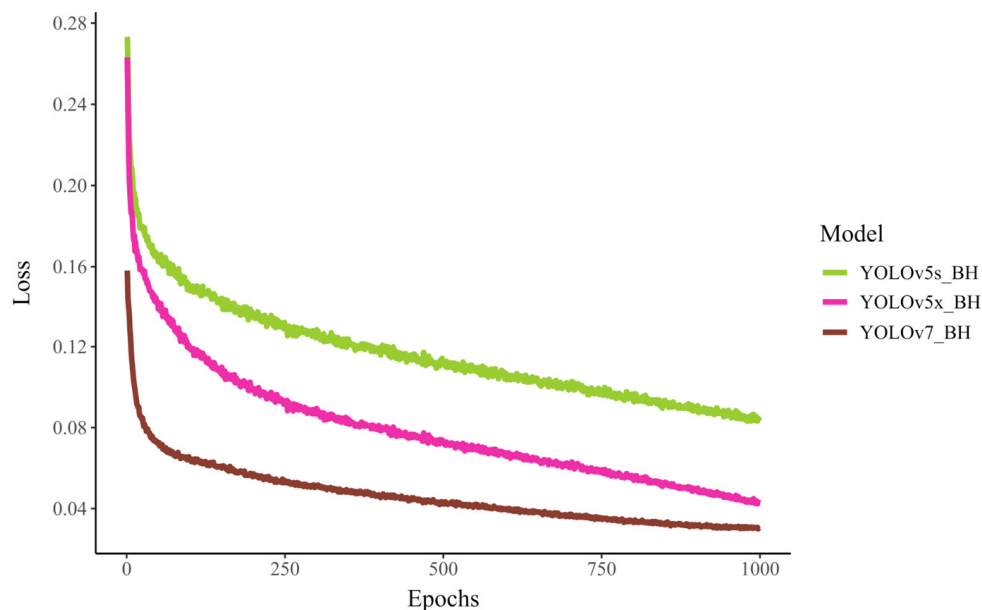


Figure 6. Total training losses from all three models while training behavior dataset.

(b) Precision

Precision is the ratio of true positive detections to the total number of positive detections. It measures the accuracy of the detection algorithm. Figure 7 shows the precision of the model on a poultry behavior classification over 1000 epochs of training.

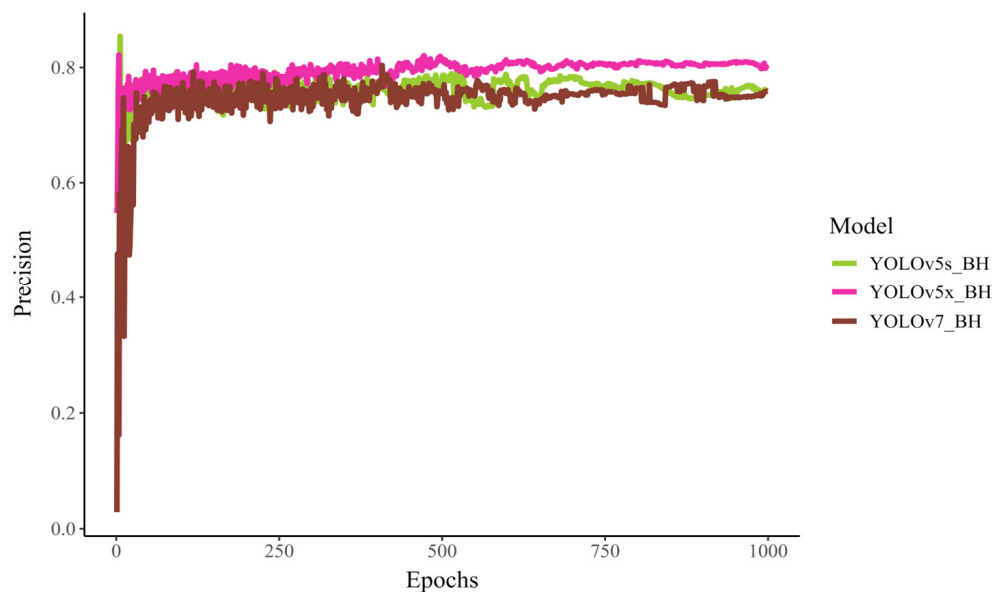


Figure 7. Precision from all three models while training behavior dataset.

(c) Recall

Recall is the ratio of true positive detections to the total number of actual positive instances. Figure 8 shows the recall of the model on a poultry behavior classification over 1000 epochs of training.

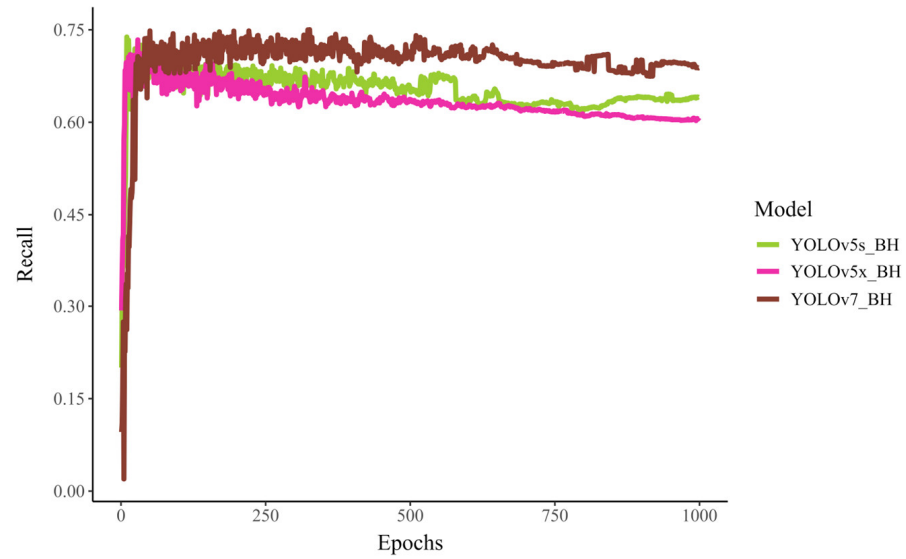


Figure 8. Recall from all three models while training behavior dataset.

(d) Mean average precision

mAP (mean average precision) is a measure that combines both precision and recall. It is the average precision calculated at various levels of recall. Figure 9 shows the mAP of the model on a poultry behavior classification over 1000 epochs of training.

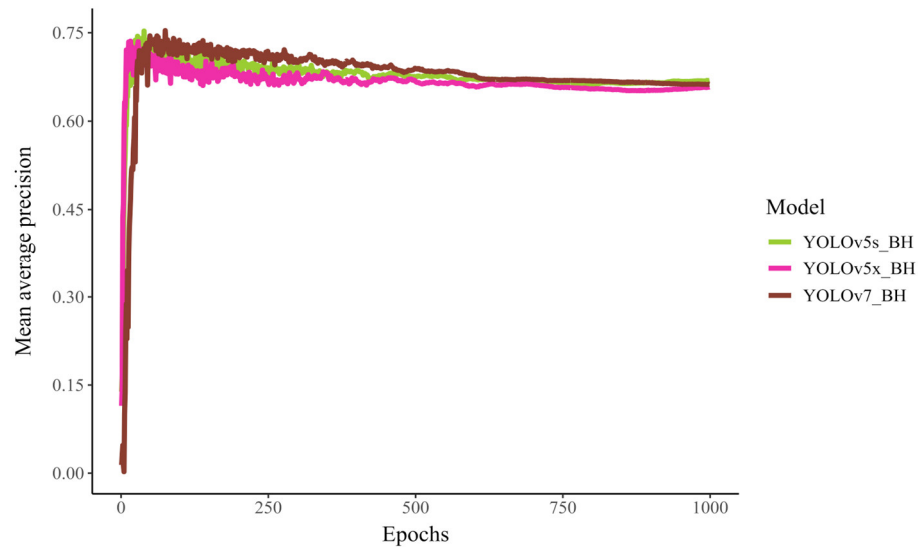


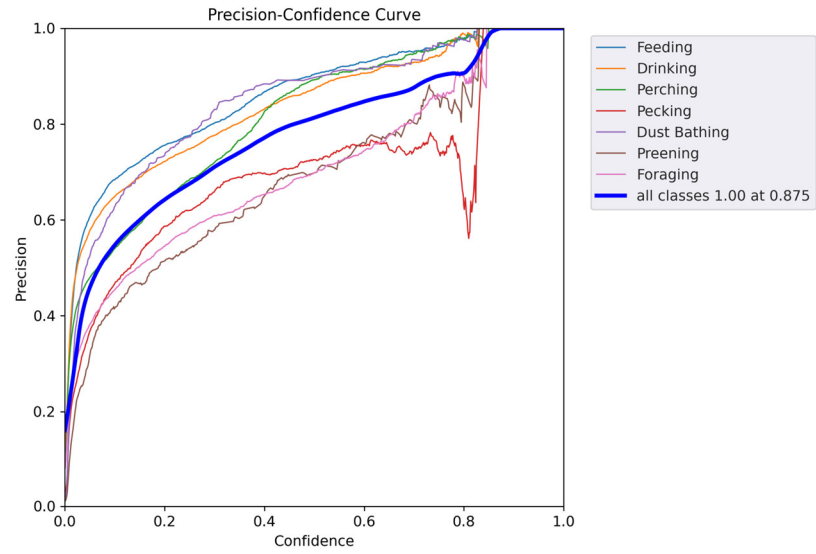
Figure 9. mAP from all three models while training behavior dataset.

3.3. Models' Performance Curves After Training

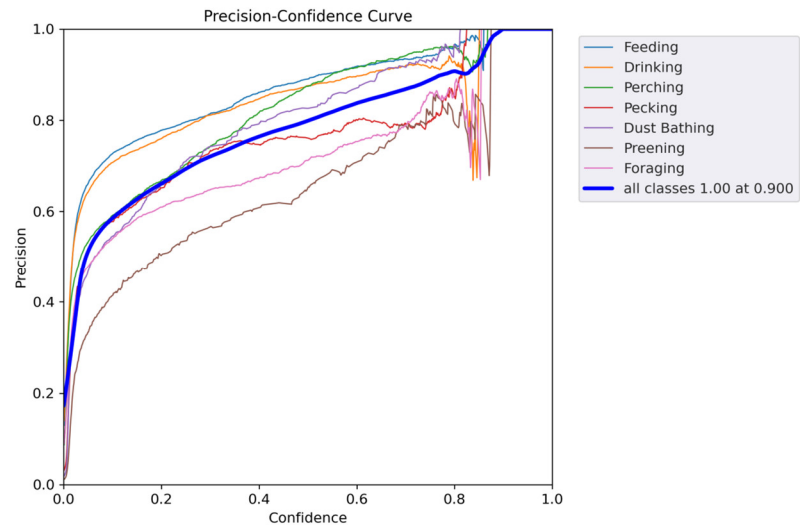
The precision, recall, and PR curve (precision × recall) were tested concerning confidence scores.

(a) Precision Curves

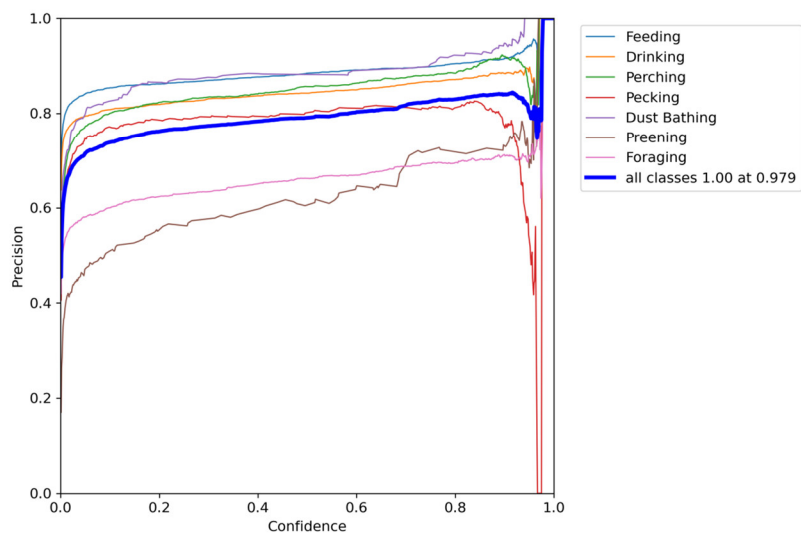
The precision plots demonstrate that the YOLOv5s_BH model initially exhibits lower behavior detection precision, which steadily improves as the confidence score increased. Specifically, as shown in Figure 10a, precision achieves 100% at a confidence score of 0.875. Similarly, the YOLOv5x_BH model in Figure 10b reaches 100% precision when the confidence score is 0.9. Additionally, the YOLOv7_BH model depicted in Figure 10c gradually increases its precision to 100% at a confidence score of 0.979.



(a) Precision Curve of YOLOv5s_BH



(b) Precision Curve of YOLOv5x_BH

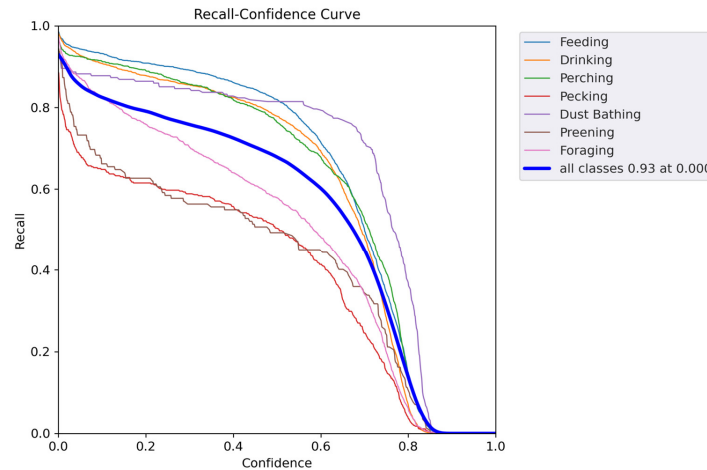


(c) Precision Curve of YOLOv7_BH

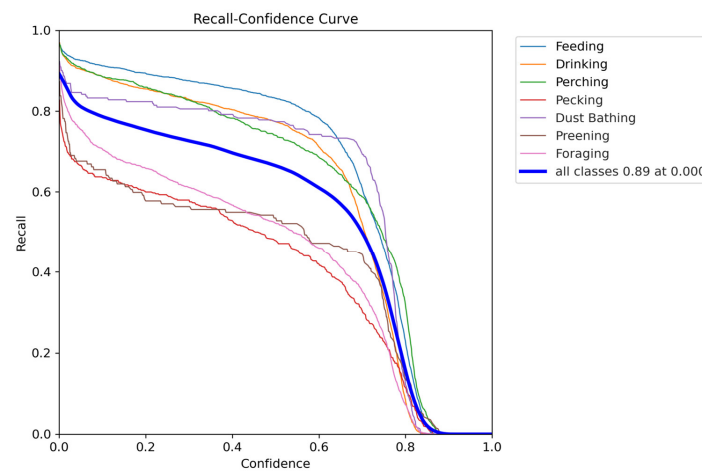
Figure 10. The precision curve of the YOLOv5s_BH, YOLO-v5x_BH, and YOLOv7_BH in behavior detection.

(b) Recall Curves

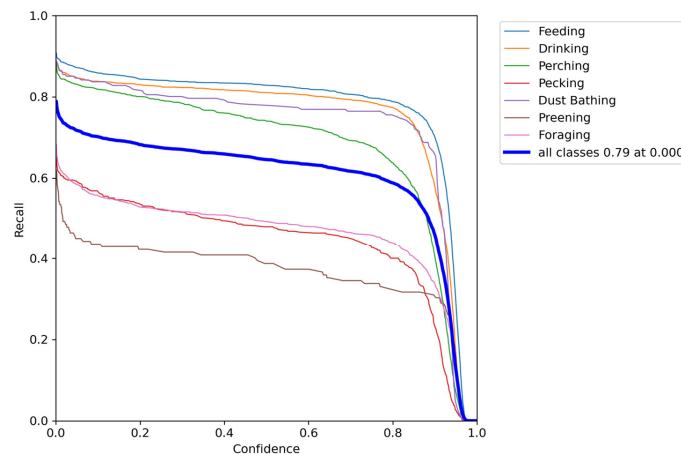
In Figure 11a, the YOLOv5s_BH model initially achieves a recall of 93% for the behavior classes, which then declines as the confidence score approaches 1. Similarly, Figure 11b illustrates that the YOLOv5x_BH model starts with a recall of 89%, mirroring the YOLOv5s_BH model, and gradually sees a decrease in recall as the confidence score increases to 1. In Figure 11c, the YOLOv7_BH model begins with a recall of 79%, which progressively diminishes as the confidence score rose to 1.



(a) Recall Curve of YOLOv5s_BH



(b) Recall Curve of YOLOv5x_BH

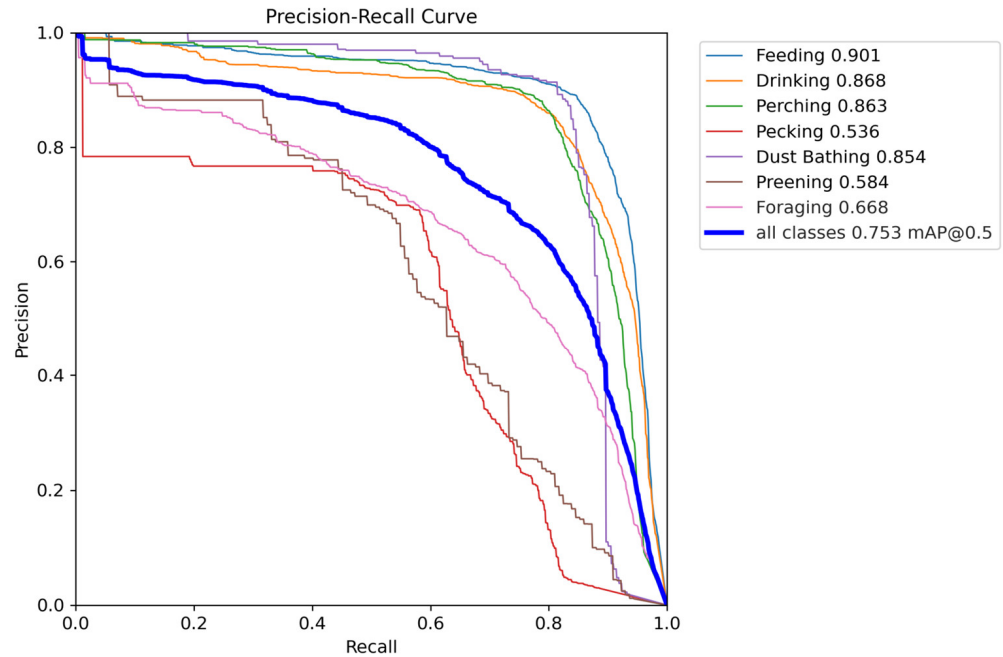


(c) Recall Curve of YOLOv7_BH

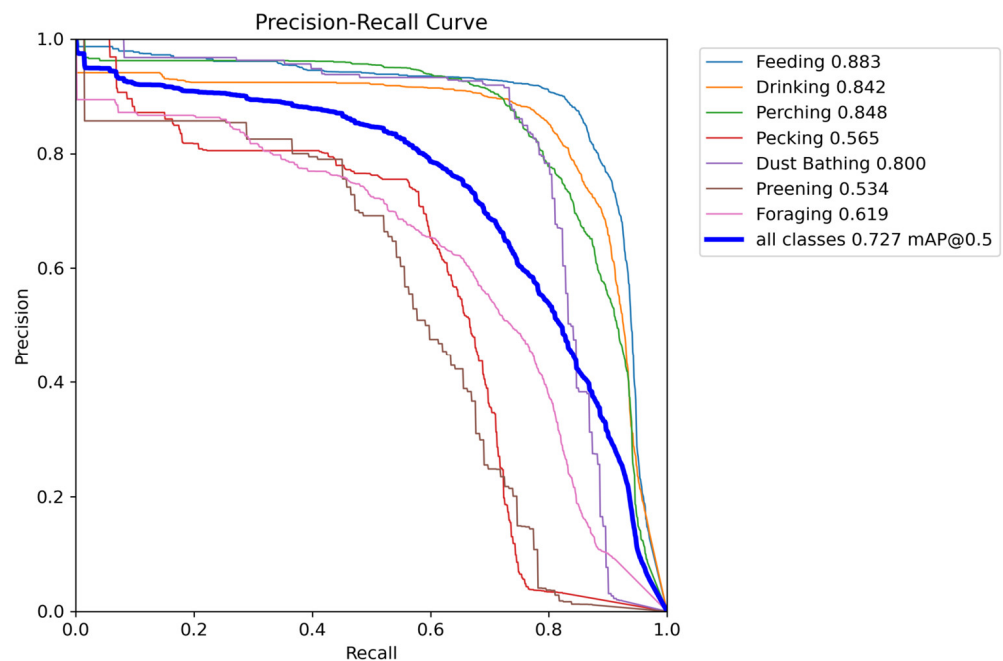
Figure 11. The recall curve of the YOLOv5s_BH, YOLOv5x_BH, and YOLOv7_BH in behavior detection.

(c) Precision–Recall Curves

The precision–recall curves were analyzed to assess the behavior detection performance of all three models as confidence scores varied for each class. This evaluation method helps determine the models’ ability to maintain high precision while increasing recall. As shown in Figure 12a–c, the combined precision \times recall values and mean average precision (mAP) for all behavior classes are 75.3% for YOLOv5s_BH, 72.7% for YOLOv5x_BH, and 66.3% for YOLOv7_BH.

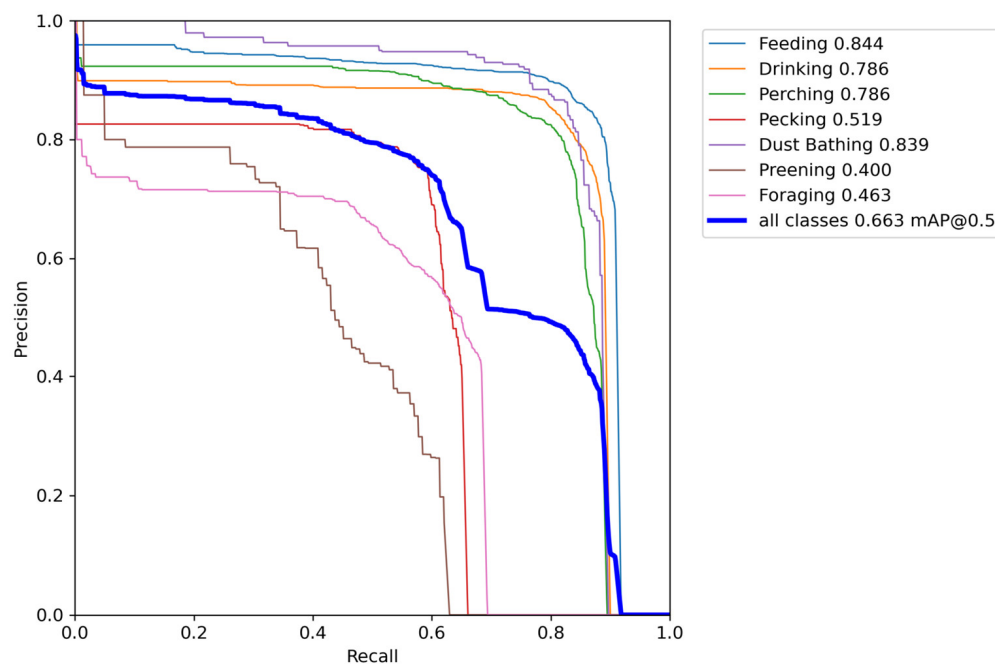


(a) Precision-Recall Curve of YOLOv5s_BH



(b) Precision-Recall Curve of YOLOv5x_BH

Figure 12. Cont.



(c) Precision-Recall Curve of YOLOv7_BH

Figure 12. The precision x recall curve of the YOLOv5s_BH, YOLOv5x_BH, and YOLOv7_BH in behavior detection.

3.4. Testing Results of New Models in Detecting Laying Behaviors

After training, the optimized model was employed to identify laying hen behaviors in new, unlabeled images. Examples of automatic behavior detection are depicted in Figure 13a–d. Table 2 provides a comparative overview of the YOLOv5s_BH, YOLOv5x_BH, and YOLOv7_BH models: the YOLOv5s_BH achieves a precision of 78.1%, a recall of 71.7%, and a mean average precision (mAP) of 75.3%; the YOLOv5x_BH attains a precision of 76.2%, a recall of 69.8%, and an mAP of 72.7%; and the YOLOv7_BH reaches a precision of 75.9%, a recall of 68.9%, and an mAP of 66.3%. Overall, the YOLOv5s_BH model outperformed both the YOLOv5x_BH and YOLOv7_BH models in terms of precision, recall, and mAP.

Table 2. Results of comparing YOLOv5s_BH, YOLOv5x_BH and YOLOv7_BH models for behavior detection.

Model	YOLOv5s_BH			YOLOv5x_BH			YOLOv7_BH		
Parameters	7.1 million			86.1 million			37 million		
Layers	157			322			-		
Class Behaviors	Precision	Recall	mAP@50	Precision	Recall	mAP@50	Precision	Recall	mAP@50
	0.781	0.717	0.753	0.762	0.698	0.727	0.759	0.689	0.663

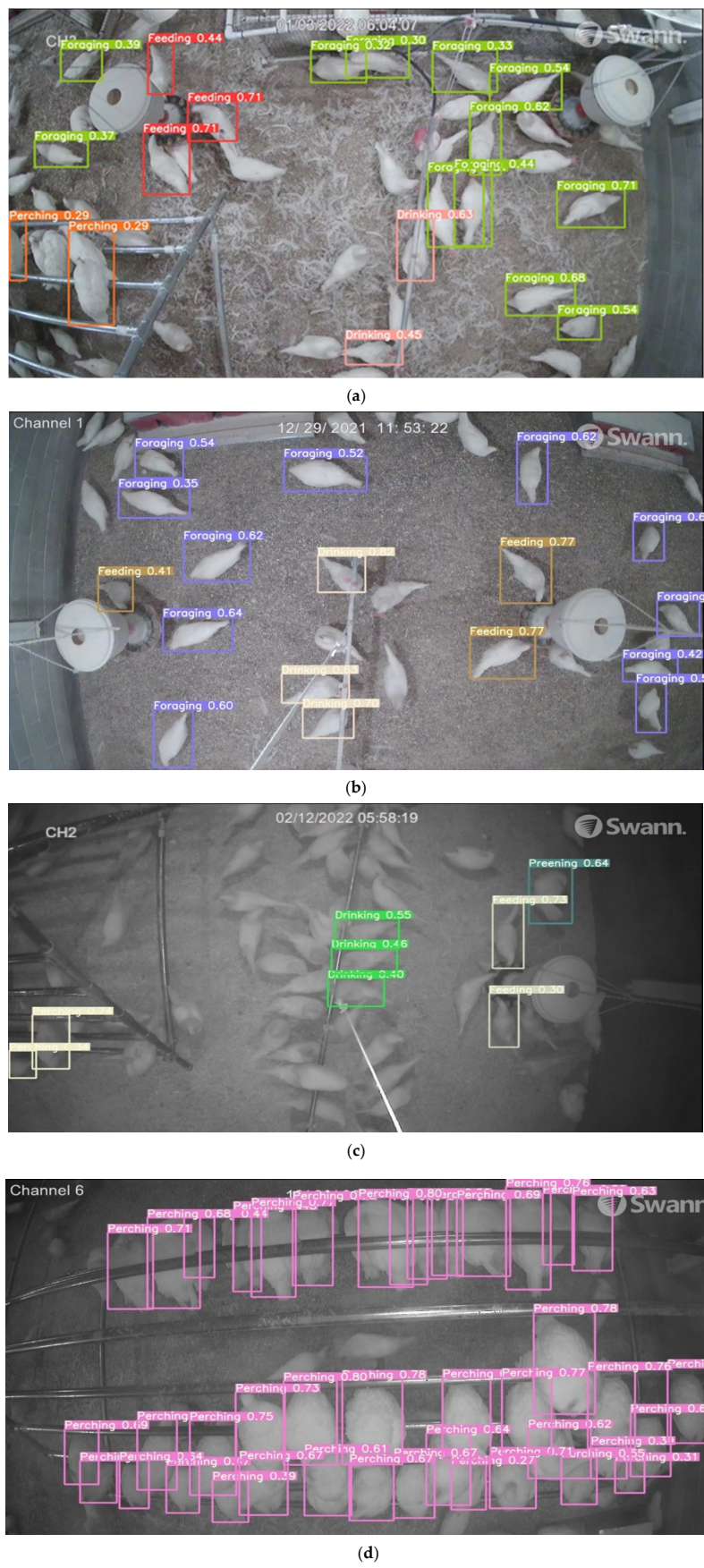


Figure 13. Performance of all three models in test dataset. (a) Feeding, Foraging, Perching and Drinking behaviors were detected. (b) Feeding and Foraging behaviors were detected. (c) Detection (Perching, Feeding and Drinking) of behaviors in an image affected by dust. (d) Perching behaviors detected by our models.

3.5. Evaluation Metrics of All Three Models in Detecting Laying Behaviors

Precision: Precision measures how accurate our behavior detector is when it predicts the presence of a behavior. It is calculated as the ratio of true positives (correctly predicted behaviors) to the sum of true positives and false positives (incorrectly predicted behaviors). From Figure 14, we can see that the highest precision is achieved for Dust Bathing with YOLOv5s_BH (0.887), followed by Feeding with YOLOv5s_BH (0.887) and Dust Bathing with YOLO7_BH (0.867). On the other hand, the lowest precision is achieved for Preening with YOLOv7_BH (0.542).

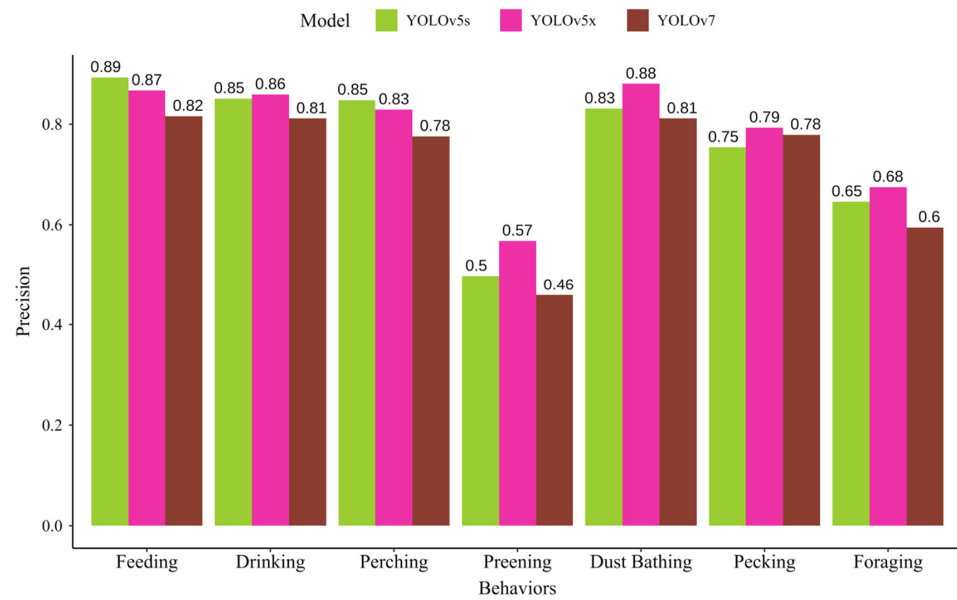


Figure 14. Precision of all models in Behavior detection.

Recall: Recall measures how well the behavior detector can detect all the behaviors in an image. It is calculated as the ratio of true positives to the sum of true positives and false negatives (missed behaviors). From Figure 15, we can see that the highest recall is achieved for all Feeding behaviors with YOLOv5x_BH (0.858), YOLOv5s_BH (0.854) and YOLOv7_BH (0.849). On the other hand, the lowest recall is achieved for Preening with YOLOv7_BH (0.43).

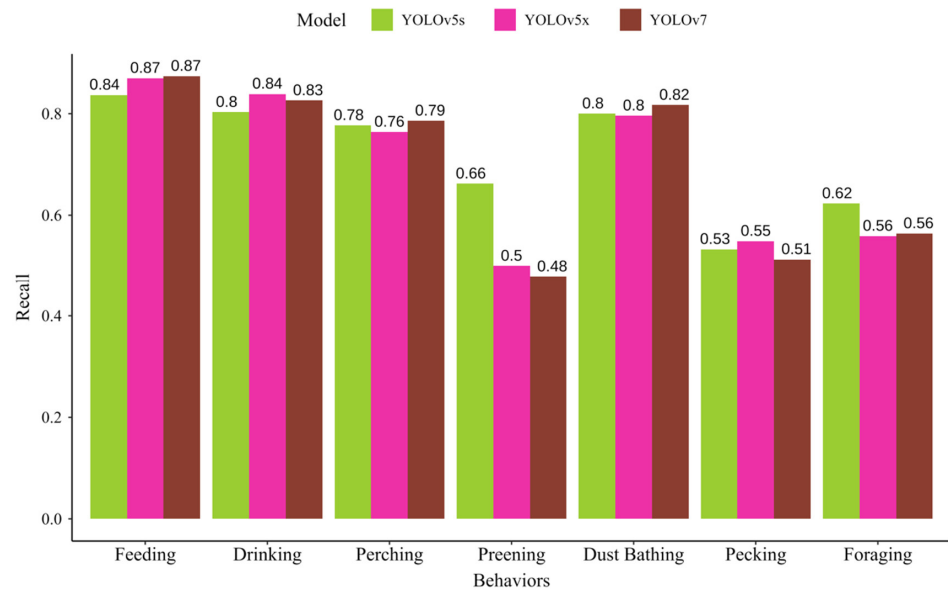


Figure 15. Recall of all models in Behavior detection.

Mean average precision: Object detection models often use the mean average precision (mAP) metric to balance precision and recall. The mAP is the average of the precision–recall curve over all the behavior categories in the dataset. From Figure 16, we can see that the highest mAP is achieved for Feeding with YOLOv5s_BH (0.901), followed again by Feeding with YOLOv5x_BH (0.883) and Drinking with YOLOv5s_BH (0.868). On the other hand, the lowest mAP is achieved for Preening with YOLOv7_BH (0.4).

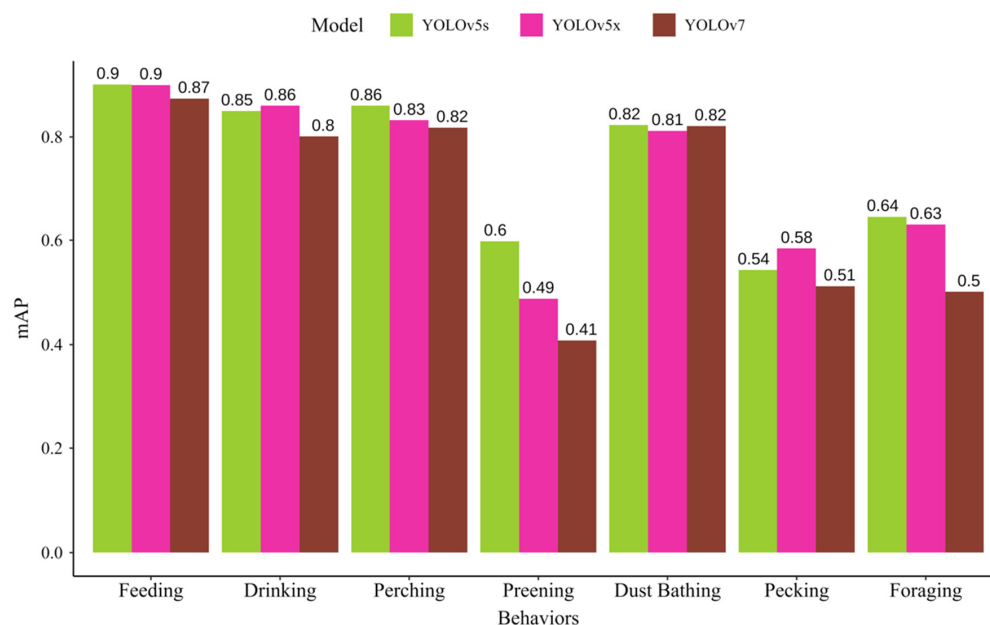


Figure 16. mAP of all models in Behavior detection.

Feeding and Dust Bathing behavior monitoring showed the highest precision, recall, and mAP across all models. This can be attributed to their distinct postures and movements, which made them easier for the models to detect and classify. Additionally, these behaviors had a relatively higher number of labeled samples in the training dataset, contributing to better model performance. On the other hand, Preening and Foraging behaviors exhibited the lowest performance metrics, likely due to the subtlety of movements in Preening and the increased complexity of detecting Foraging behaviors in crowded environments. The overlapping of birds and environmental factors such as dust may have further contributed to the challenges in detecting these behaviors accurately.

Using deep learning and machine vision techniques in classifying and analyzing poultry behaviors has yielded substantial advancements in our understanding of avian welfare and husbandry practices. Six behaviors of laying hens—standing, sitting, sleeping, grooming, feeding, and drinking—were also classified [10]. Our research also uses a single-stage detector (YOLO) to classify the different behaviors in the laying hens. There have been works on classifying feeding, drinking, and resting behavioral behavior using deep learning with the addition of different zones and interference, and we attempted to classify other behaviors expressed by the birds [21]. The volume of the training dataset also plays an important role in obtaining high evaluation metrics. The evaluation metrics for some behaviors are not high. The reason could be the small training data for that behavior. Our attempt to detect multiple classes of laying behaviors was challenging due to the complexity of the data with precision, recall, and mAP values of 78.1%, 71.7%, and 75.3%, which are higher than the 70% success rate of broiler’s behavioral detection [4]. Behaviors were classified using different zones like perching, feeding, drinking, and nesting [21] are different than using specific images of behaviors as the posture of birds showcasing the behaviors plays an important role in detection. The detection of individual poultry

behaviors is different than using a collective class of different behaviors of birds in a cage-free setting. The task was quite challenging due to the environment with high dust affecting the data quality in images and the overcrowding of birds, making the task harder for behavior detection.

In summary, the YOLOv5s_BH model seems to perform better than YOLOv5x_BH and YOLOv7_BH models for most behaviors. Feeding and Dust Bathing are the behaviors with the highest overall performance in terms of precision, recall, and mAP. Preening and Foraging are the behaviors with the lowest overall performance in terms of precision, recall, and mAP. The efficiency of the model is impacted by the density of the stock, as overcrowding can increase the likelihood of errors in detecting objects. This is because in environments with higher stocking densities, there is a greater probability of flock crowding, which can result in increased detection errors. In cage-free environments, whether in experimental or commercial settings, the overlapping of the head and body is frequent.

4. Conclusions

Precision poultry farming technology is necessary to detect laying hen behaviors automatically. An advanced object detection technology (i.e., the YOLO model) was used as the model structure. In this study, three new deep learning models, the “YOLOv5s_BH”, “YOLOv5x_BH”, and “YOLOv7_BH” networks, were developed to detect and classify behaviors in four research study cage-free facilities at the University of Georgia. Our model automatically detects and classifies behaviors in cage-free facilities. Furthermore, the smallest model, YOLOv5s_BH, performed well in terms of precision, recall and mAP compared to YOLOv5x_BH and YOLOv7_BH. The YOLOv5s_BH model had a 1.9%, 1.9%, and 2.6% higher performance in precision, recall, and mAP than the YOLOv5x_BH model, respectively. Also, the YOLOv5s_BH model had a 2.2%, 2.8%, and 9% higher performance in precision, recall, and mAP than the YOLOv7_BH model, respectively. The better performance of YOLOv5s_BH is likely due to its smaller architecture, which helps prevent overfitting on the smaller datasets often used in behavioral research. Its lightweight design allows for faster training and inference, making it more effective for detecting laying hen behaviors with high precision. In contrast, the larger YOLOv5x_BH and YOLOv7_BH models, built for more complex tasks, may add unnecessary complexity and computational demands, which can slightly lower their performance in this context. This study provides a reference for cage-free producers that poultry behaviors could be monitored automatically. Future studies are guaranteed to test the system in commercial houses.

Author Contributions: Conceptualization, L.C.; Methodology, S.S. and L.C.; Validation, S.S.; Formal analysis, S.S.; Investigation, S.S., R.B.B., X.Y. and L.C.; Resources, L.C.; Writing—original draft, S.S., R.B.B., X.Y., G.L. and L.C.; Supervision, L.C.; Project administration, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This study was sponsored by the USDA-NIFA AFRI program (2023-68008-39853); Georgia Research Alliance; Egg Industry Center; Oracle for Research Grant, Oracle America (Award Number: CPQ-2060433); and UGA Institute for Integrative Precision Agriculture.

Data Availability Statement: The datasets generated, used and/or analyzed during the current study will be available from the corresponding author on reasonable request.

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

References

1. Hewson, C. What is animal welfare? Common definitions and their practical consequences. *Can. Vet. J.* **2003**, *44*, 496. Available online: <https://www.semanticscholar.org/paper/What-is-animal-welfare-Common-definitions-and-their-Hewson/3b61285ac62162b63487d397f6e45927d92ef95e> (accessed on 28 November 2024). [PubMed]

2. Vestergaard, K.S.; Skadhauge, E.; Lawson, L. The Stress of Not Being Able to Perform Dustbathing in Laying Hens. *Physiol. Behav.* **1997**, *62*, 413–419. [[CrossRef](#)] [[PubMed](#)]
3. Xu, D.; Shu, G.; Liu, Y.; Qin, P.; Zheng, Y.; Tian, Y.; Zhao, X.; Du, X. Farm Environmental Enrichments Improve the Welfare of Layer Chicks and Pullets: A Comprehensive Review. *Animals* **2022**, *12*, 2610. [[CrossRef](#)]
4. Pereira, D.F.; Miyamoto, B.C.; Maia, G.D.; Sales, G.T.; Magalhães, M.M.; Gates, R.S. Machine vision to identify broiler breeder behavior. *Comput. Electron. Agric.* **2013**, *99*, 194–199. [[CrossRef](#)]
5. van Liere, D.W.; Bokma, S. Short-term feather maintenance as a function of dust-bathing in laying hens. *Appl. Anim. Behav. Sci.* **1987**, *18*, 197–204. [[CrossRef](#)]
6. Neethirajan, S. The role of sensors, big data and machine learning in modern animal farming. *Sens. Bio-Sens. Res.* **2020**, *29*, 100367. [[CrossRef](#)]
7. Okinda, C.; Nyalala, I.; Korohou, T.; Okinda, C.; Wang, J.; Achieng, T.; Wamalwa, P.; Mang, T.; Shen, M. A review on computer vision systems in monitoring of poultry: A welfare perspective. *Artif. Intell. Agric.* **2020**, *4*, 184–208. [[CrossRef](#)]
8. Loddo, A.; Di Ruberto, C. On the Efficacy of Handcrafted and Deep Features for Seed Image Classification. *J. Imaging* **2021**, *7*, 171. [[CrossRef](#)] [[PubMed](#)]
9. Black, K.M.; Law, H.; Aldoukhi, A.; Deng, J.; Ghani, K.R. Deep learning computer vision algorithm for detecting kidney stone composition. *BJU Int.* **2020**, *125*, 920–924. [[CrossRef](#)]
10. Leroy, T.; Vranken, E.; Van Brecht, A.; Struelens, E.; Sonck, B.; Berckmans, D. A computer vision method for on-line behavioral quantification of individually caged poultry. *Trans. ASABE* **2006**, *49*, 795–802. [[CrossRef](#)]
11. Li, G.; Zhao, Y.; Purswell, J.L.; Du, Q.; Chesser, G.D.; Lowe, J.W. Analysis of feeding and drinking behaviors of group-reared broilers via image processing. *Comput. Electron. Agric.* **2020**, *175*, 105596. [[CrossRef](#)]
12. Pu, H.; Lian, J.; Fan, M. Automatic Recognition of Flock Behavior of Chickens with Convolutional Neural Network and Kinect Sensor. *Int. J. Pattern Recognit. Artif. Intell.* **2018**, *32*, 1850023. [[CrossRef](#)]
13. Subedi, S.; Bist, R.; Yang, X.; Chai, L. Tracking pecking behaviors and damages of cage-free laying hens with machine vision technologies. *Comput. Electron. Agric.* **2023**, *204*, 107545. [[CrossRef](#)]
14. Ren, Y.; Huang, Y.; Wang, Y.; Zhang, S.; Qu, H.; Ma, J.; Wang, L.; Li, L. A High-Performance Day-Age Classification and Detection Model for Chick Based on Attention Encoder and Convolutional Neural Network. *Animals* **2022**, *12*, 2425. [[CrossRef](#)] [[PubMed](#)]
15. Yang, X.; Bist, R.; Subedi, S.; Chai, L. A deep learning method for monitoring spatial distribution of cage-free hens. *Artif. Intell. Agric.* **2023**, *8*, 20–29. [[CrossRef](#)]
16. Massari, J.M.; de Moura, D.J.; Nääs, I.d.A.; Pereira, D.F.; Branco, T. Computer-Vision-Based Indexes for Analyzing Broiler Response to Rearing Environment: A Proof of Concept. *Animals* **2022**, *12*, 846. [[CrossRef](#)]
17. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
18. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Xie, T.; Fang, J.; Michael, K.; Montes, D.; Nadar, J.; et al. ultralytics/yolov5: v6.1—TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference. *Zenodo* **2022**. [[CrossRef](#)]
19. Wang, C.Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696. [[CrossRef](#)]
20. Li, G.; Xu, Y.; Zhao, Y.; Du, Q.; Huang, Y. Evaluating Convolutional Neural Networks for Cage-Free Floor Egg Detection. *Sensors* **2020**, *20*, 332. [[CrossRef](#)]
21. Automatic Monitoring of Chicken Movement and Drinking Time Using Convolutional Neural Networks. Available online: <https://doi.org/10.13031/trans.13607> (accessed on 10 January 2025).
22. Wood-Gush, D.G.M.; Duncan, I.J.H. Some behavioural observations on domestic fowl in the wild. *Appl. Anim. Ethol.* **1976**, *2*, 255–260. [[CrossRef](#)]
23. Mens, A.; van Krimpen, M.; Kwakkel, R. Nutritional approaches to reduce or prevent feather pecking in laying hens: Any potential to intervene during rearing? *Worlds Poult. Sci. J.* **2020**, *76*, 591–610. [[CrossRef](#)]
24. Ferreira, V.H.B.; Barbarat, M.; Lormant, F.; Germain, K.; Brachet, M.; Løvlie, H.; Calandreau, L.; Guesdon, V. Social motivation and the use of distal, but not local, featural cues are related to ranging behavior in free-range chickens (*Gallus gallus domesticus*). *Anim. Cogn.* **2020**, *23*, 769–780. [[CrossRef](#)] [[PubMed](#)]
25. ATulbure, A.-A.; Dulf, E.-H. A review on modern defect detection models using DCNNs—Deep convolutional neural networks. *J. Adv. Res.* **2022**, *35*, 33–48. [[CrossRef](#)]
26. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [[CrossRef](#)]
27. Guo, Z.; Wang, C.; Yang, G.; Huang, Z.; Li, G. MSFT-YOLO: Improved YOLOv5 Based on Transformer for Detecting Defects of Steel Surface. *Sensors* **2022**, *22*, 3467. [[CrossRef](#)]

28. Luo, Y.; Xia, J.; Lu, H.; Luo, H.; Lv, E.; Zeng, Z.; Li, B.; Meng, F.; Yang, A. Automatic Recognition and Quantification Feeding Behaviors of Nursery Pigs Using Improved YOLOV5 and Feeding Functional Area Proposals. *Animals* **2024**, *14*, 569. [[CrossRef](#)] [[PubMed](#)]
29. Lv, M.; Su, W.-H. YOLOV5-CBAM-C3TR: An optimized model based on transformer module and attention mechanism for apple leaf disease detection. *Front. Plant Sci.* **2024**, *14*, 1323301. [[CrossRef](#)]
30. Shen, H.; Dong, Z.; Yan, Y.; Fan, R.; Jiang, Y.; Chen, Z.; Chen, D. Building roof extraction from ASTIL echo images applying OSA-YOLOv5s. *Appl. Opt.* **2022**, *61*, 2923–2928. [[CrossRef](#)]
31. Liu, W.; Quijano, K.; Crawford, M.M. YOLOv5-Tassel: Detecting Tassels in RGB UAV Imagery with Improved YOLOv5 Based on Transfer Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3206399. [[CrossRef](#)]
32. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768. [[CrossRef](#)]
33. Zhang, D.-Y.; Luo, H.-S.; Wang, D.-Y.; Zhou, X.-G.; Li, W.-F.; Gu, C.-Y.; Zhang, G.; He, F.-M. Assessment of the levels of damage caused by Fusarium head blight in wheat using an improved YoloV5 method. *Comput. Electron. Agric.* **2022**, *198*, 107086. [[CrossRef](#)]
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *arXiv* **2015**, arXiv:1406.4729. [[CrossRef](#)]
35. Wang, K.; Hu, X.; Zheng, H.; Lan, M.; Liu, C.; Liu, Y.; Zhong, L.; Li, H.; Tan, S. Weed detection and recognition in complex wheat fields based on an improved YOLOv7. *Front. Plant Sci.* **2024**, *15*, 1372237. [[CrossRef](#)] [[PubMed](#)]
36. Yang, Z.; Ni, C.; Li, L.; Luo, W.; Qin, Y. Three-Stage Pavement Crack Localization and Segmentation Algorithm Based on Digital Image Processing and Deep Learning Techniques. *Sensors* **2022**, *22*, 8459. [[CrossRef](#)] [[PubMed](#)]
37. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [[CrossRef](#)]
38. Hossain, A.; Islam, M.T.; Almutairi, A.F. A deep learning model to classify and detect brain abnormalities in portable microwave based imaging system. *Sci. Rep.* **2022**, *12*, 6319. [[CrossRef](#)]

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.