Received: 29 August 2023 | Revised: 19 December 2023 | Accepted: 16 January 2024 | Published online: 17 January 2024

RESEARCH ARTICLE

Artificial Intelligence and Applications
2024, Vol. 2(2) 115–125
DOI: 10.47852/bonviewAIA42021603

BON VIEW PUBLISHING

Mask YOLOv7-Based Drone Vision System for Automated Cattle Detection and Counting

Rotimi-Williams Bello[1,*] and Mojisola Abosede Oladipo[2]

[1]*Department of Mathematics and Computer Science, University of Africa, Nigeria*

[2]*Department of Agricultural Economics and Extension, University of Africa, Nigeria*

Abstract: Conventional method of counting animals is one of the most challenging tasks in livestock management; moreover, counting of animals in drone-acquired imagery, though promising, is more challenging in intelligent livestock management. In this paper, we apply state-of-the-art object detection model, Mask YOLOv7, for detection and counting of cattle in different scenarios such as in controlled (feedlot) environment and uncontrolled (open-range) environment. Mask mechanism was embedded into the backbone of the YOLOv7 algorithm (Mask YOLOv7) for instance segmentation of individual cattle object. We evaluate the performance of the model proposed in this study using Intersection over Union threshold of 0.5, average precision (AP), and mean average precision. The results of the experiment conducted in this study show that the proposed model achieves an accuracy of 93% in counting cattle in controlled environment and 95% in uncontrolled environment. These results affirm the potential of the model, Mask YOLOv7, to perform competitively with any other existing object detection and instance segmentation models in terms of accuracy and AP especially when the speed of object detection matters. Moreover, the research has potential applications in livestock inventory, which helps in tracking, monitoring, and reporting vital information about individual cattle.

Keywords: cattle, deep learning, livestock management, object detection, Mask YOLOv7

1. Introduction

Many large-scale agriculture producing countries generate greater percentage of their agricultural revenue from animal husbandry. According to Sishodia et al. (2020), meat consumption is on the increase in demand by populace, and to meet their demand, there is a need to address the challenges confronting livestock production and their management. Lack of expertise to manage the livestock, problems with remote monitoring of the livestock, high costs of managing the livestock, and government policies are some of the harsh challenges confronting the operation and maintenance of large-scale livestock production and management systems. Therefore, it is necessary to address the aforementioned challenges in animal husbandry with appropriate methods. Animals' behavior is a reflection of their state and conditions, and recent advancements in smart agricultural technology have enabled automatic monitoring of animals' behaviors for their health, which include their body weights and eating habits, etc. for the improvement and maximization of meat production (Kumar & Ilango, 2018).

1.1. Above-ground animal detection and monitoring using drone vision systems

The use of sensors has in no small measure assisted in remote monitoring of animals by providing useful information in real-time for uninterrupted monitoring of animals in the open-range; by this, great changes have been brought to the perception of farmers toward the possibility of remote monitoring and management of animals (Kumar & Ilango, 2018). Monitoring of movement and behavior of animals by the farmers are also made possible by wearing sensor-based devices on the animals, which further help in monitoring their physiological and morphological conditions to prevent their unhealthy growth and death rate for overall production gain (Auclair-Ronzaud et al., 2020; Du & Zhou, 2018; González et al., 2018; Halachmi et al., 2019; Kumar & Ilango, 2018; O'Leary et al., 2020; Sharma & Koundal, 2018; Wang et al., 2021). Camera-trapped imagery has been employed for automatic identification and counting of species of animals (Willi et al., 2019); moreover, camera traps and thermal infrared imagery have been employed as methods by Sharma and Koundal (2018), Beaver et al. (2020), and Tabak et al. (2020) for capturing animal activities at different locations and diagnosing various degrees of disease militating against them.

Nigeria is a country with millions of cattle species alone excluding other animal species, mostly with nomadic animal rearing methods. These methods make it difficult to monitor activities of individual animals during grazing. Real-time monitoring can be a good method of controlling the intrusion of animals to restricted areas and keeping the animals safe during grazing. Unmanned aerial vehicle (UAV) imagery could serve as an alternative method to land-based survey of animals. The emergence and the application of drone vision system in monitoring of animals show promising prospect, which if combined with deep learning models will turn around animal farming and management.

Drone vision systems are on a par with other technologies for tasks such as viewing from above-ground, acquisition of high-resolution image outputs without delay irrespective of the terrain

*Corresponding author: Rotimi-Williams Bello, Department of Mathematics and Computer Science, University of Africa, Nigeria. Email: sirbrw@yahoo.com

and weather conditions. Although reliability and accuracy are very important factors to be considered when counting animals in drone vision system-acquired imagery, they still present some difficulties in the management of intelligent husbandry (Alanezi et al., 2022). For drone vision systems to be regarded as a reliable and efficient monitoring method for livestock activities, the embedded algorithms for processing the acquired images must match with the corresponding functions (Tsouros et al., 2019).

1.2. Machine learning-based drone vision systems for animal detection and monitoring

The recent advancement in machine learning has greatly increased the application of drone vision systems in animal detection and counting (Eikelboom et al., 2019). UAV applications cut across different tasks such as estimation of livestock population (Chabot et al., 2018; Eikelboom et al., 2019; Han et al., 2019; O'Leary et al., 2020; Ulhaq et al., 2021). Image segmentation is one of the most employed techniques for automatic detection and counting of animals in images; it works on either the instance or the semantic of objects (animals) in the images using their pixels with a specific threshold (Chabot et al., 2018; Dujon et al., 2021). In addition, image segmentation performs better when there is a clear difference between the image foreground objects (animals) and the image background.

To perform a multi-stage counting in a UAV embedded with imagery sensors for images with complex features, a hybrid of template matching and spectral characteristics approach was improved by Sadgrove et al. (2021). To obtain accurate results from the above-ground animal survey, thermal imagery and UAV were integrated for wildlife detection, segmentation, classification, and tracking by González et al. (2018), and this was made possible by using a pixel with a specific threshold and binary mask that matches a template in different instances. The prospect in using computer vision for livestock detection from UAV imagery was revealed in Sadgrove et al. (2021), this is in addition to different machine learning models that have been employed solely for detection and counting of animals from UAV imagery. Supervised pixel-based image classification method (Chabot et al., 2018) and unsupervised pixel-based image classification method (Han et al., 2019) have been used for animal identification and counting, and animal population overestimation, respectively, with proper preprocessing such as labeling and augmentation of the images for supervised data training method using either manual or automatic labeling method such as LabelMe (Russell et al., 2008) for effectual results.

Convolutional neural network (CNN) and deep CNN-based models such as R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2017), and Mask R-CNN (He et al., 2017) have received awesome acceptance in the computer vision community for exhibiting great object detection, segmentation, and classification in complex images with speed, accuracy, and precision combined. Among the researchers who have utilized CNN-based models for tasks involving the counting and monitoring of animals are Eikelboom et al. (2019) and Xu et al. (2020), who proposed a hybrid of CNN and UAV systems for tracking and counting animals that were detected in UAV-based video recordings. It is necessary to consider some factors when conducting animal detection tasks such as variability in illumination, occlusion, and similarity between foreground objects (target objects) and their backgrounds.

Although R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, and SSD (Liu et al., 2016) have been applied as solutions to animal detection, identification, and monitoring tasks with great results (Bello et al., 2021a; Bello et al., 2021b; Bello et al., 2021c), there is a need to address detection speed of the models when applying to animal monitoring, which is what YOLOv7 (Wang et al., 2022) represents. Mask YOLOv7, a deep learning model popular for its speed and good accuracy among the community of precision agriculture, is a neural network architecture developed purposely for object detection and image segmentation in real time. Among the numerous works that utilized YOLO algorithms for agricultural tasks are Hatton-Jones et al. (2021), Hu et al. (2023), Madasamy et al. (2021), and Yang et al., (2023).

Hence, the primary goal of the work carried out in this paper is to utilize the state-of-the-art algorithm, Mask YOLOv7, a speed and accurate method, embedded in drone vision systems, for automatic detection and counting of cattle in images.

2. Related Work

Andrew et al. (2016) proposed an automated visual identification system for individual Holstein Friesian cow from dorsal RGB-D-based imagery. By using support vector machine (SVM) and radial basis function kernels, which were based on ASIFT descriptor structure, predictions were generated. The system was able to perform segmentation of the animal regions by fitting a depth model; this was followed by extracting ASIFT descriptors over the area that was detected. The essence of using SVM is to learn a species-wide predictor of descriptor individuality utilized for the selection and usage of features to recover the identity of the cow. A method based on image entropy was proposed by Gu et al. (2017) to recognize and identify the behavior of cow object on motion against a complex background. For automated capturing of behavior and characteristic features displayed by the cow, they employed minimum bounding box and contour mapping. By demonstration, Andrew et al. (2017) posited the appropriateness of computer vision pipelines that make use of the architectures of deep neural network to perform the automated detection and identification of individual Holstein Friesian cow in a farm setup using dorsal coat patterns. With the available datasets, they have demonstrated the possibility of performing robust detection and localization of Holstein Friesian cow with 99.3% accuracy.

Cheema and Anand (2017) proposed object detection based on Faster R-CNN framework for efficient detection of animals in images. They trained a linear SVM classifier for the recognition of individual animals using the features extracted from AlexNet of the animal's flank. The techniques of deep learning were employed in Zin and Tin (2018) for exploration and examination of the image processing technologies utilization in analyzing and identifying individual cattle. The main features considered for the identification are the black and white body patterns of the cow. The body of the cow which was placed on the Rotary Milking Parlor was detected by using inter-frame differencing and horizontal histogram-based method. The predefined distance value was used for the extraction of body region of the cow, which served as input data to train the deep CNN.

A method of artificial intelligence-based CNNs was employed in Rivas et al. (2018) for the analyses of images captured by a camera-aided drone for the identification of individual objects in the images. The approach they used is such that the trained CNNs can detect not only cow but any other object by using the same algorithmic process of CNNs training. A computer vision system was proposed by Zhao et al. (2019) that could identify individual dairy cows. This was made possible by making use of videos that show the side view of cow in motion. Detection and location of the cow object and its body area as the individual identity information were made possible by the system. To determine the identity of unfamiliar images, a

template database was created for matching and comparing the images. Their experiment results reveal the possibility of calculating accurately the feature points in the body pattern of cows by the use of SIFT method. When FAST, SIFT, and FLANN are used for the detection, extraction, and matching points, 96.72% accuracy of one-step identification was achieved.

Liu et al. (2020) in their proposed practical system employed multiple methods to detect recorded structural information about cattle in a video. To come up with the cow structural model, key features were employed for the representation of positions of the specific body parts of the cow and its overall spatial location, such as the connections between the head, the trunk, and the legs. For the extraction of the key features from the raw images and selection of individual features for conversion into a structural model, two CNNs were applied to the detection system. In order to enable the system work with different quality of videos collected from a public farm during normal operation, a post-processing model was developed. A non-contact method based on deep parts features fusion for identifying cow was proposed by Hu et al. (2020). For the extraction of the cow object in the side view image and the cow's head, trunk, and leg parts, they applied YOLO method and a part segmentation algorithm using frame differencing and segmentation span analysis.

Three independent fine-tuned AlexNet models were used in extracting the deep features of the cow's head, trunk, and leg parts. While a weighted summation strategy was employed for the features fusion, a trained SVM classifier was used for the cow image classification. An automated method based on Mask R-CNN capable of counting cattle in a quadcopter vision system was proposed by Xu et al. (2020). The application of the Mask R-CNN framework was demonstrated for instance segmentation of the detected cow images in the counting experiment in open-range and feedlots environment. Performance evaluation method was used in verifying the optimal Intersection over Union (IOU) threshold (0.5) and the detection performance of the algorithm for full appearance.

Similar work to Xu et al. (2020) was carried out by Shao et al. (2020) to aid in managing open-range cattle; they proposed a system based on CNNs for detecting and counting cow using UAV-captured images. They improved the system performance for detection by utilizing the UAV images, thereby enabling the approximate size prediction of the object when the assumption can be made of the height of UAV from the ground to be approximately constant. They resized to an optimal resolution the input image for training and testing the CNN, which is determined by the object's size and the down-sampling rate of the network. To prevent repetitive image counting, they applied a 3D model reconstructed by using the UAV images for clustering detection results.

3. Materials and Methods

The primary materials used in performing the work in this paper include dataset of cow images, drone system, open-range, and feedlots. For the methods, they include overview of the proposed model architecture and our framework, image acquisition, dataset preparation and preprocessing, the algorithm for detection and counting cattle, Mask YOLOv7 implementation details, and performance evaluation metrics.

3.1. Overview of the proposed model architecture and our framework

This section presents the general idea behind the proposed model architecture and our framework pipeline for processing the drone-captured images solely for cattle detection and counting using Mask YOLOv7 algorithm. Figure 1 (Wang et al., 2022) shows the extended efficient layer aggregation networks (E-ELAN) of the YOLOv7 algorithm, which primarily concentrate on a model's number of parameters and computational density. The VovNet (CNN seeks to make DenseNet more efficient by combining all features only once in the last feature map) model and the CSPVoVNet model analyze the influence of the input/output channel ratio and the element-wise operation on the network inference speed. YOLOv7 extended ELAN and called it E-ELAN. The major advantage of ELAN was that by controlling the gradient path, a deeper network can learn and converge more effectively. The gradient transmission path of the original architecture is not changed by the E-ELAN; however, the cardinality of the added features is increased by it using group convolution, and the features of different groups are combined in a shuttle and merged cardinality manner. The essence of carrying out the operation in this manner is to ensure the enhancement of the features learned by different feature maps and the improvement of the use of parameters and computations.

While the architecture in the computational block is majorly changed by E-ELAN, the entire transition layer architecture is not changed. It employs expansion technique in addition to shuffle and merge techniques to enhance the network learning ability without collapsing the original gradient path. The approach in this scenario

Figure 1
Extended-efficient layer aggregation networks (E-ELAN)

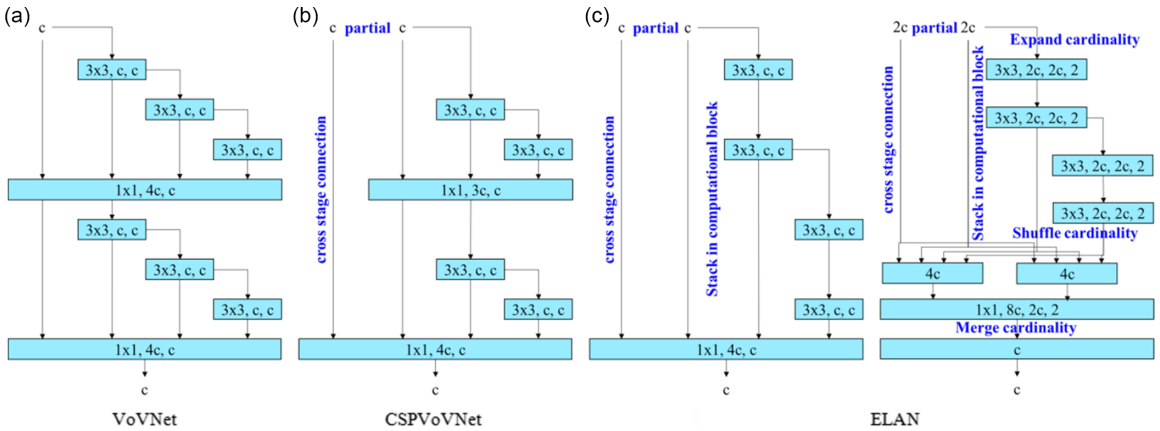
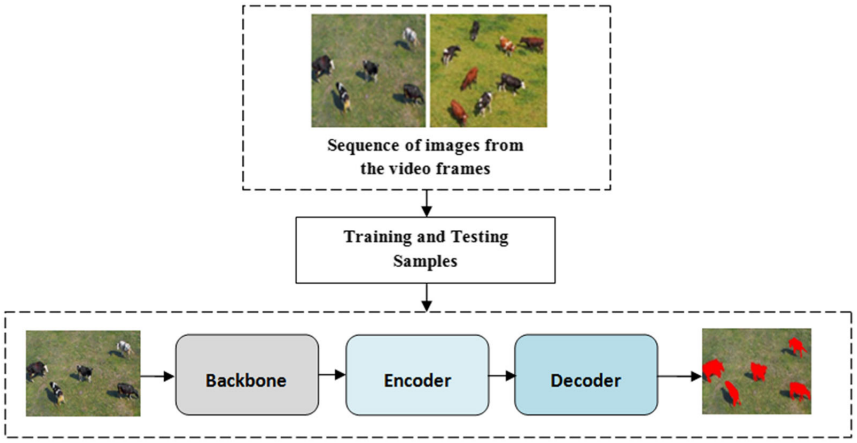


Figure 2
Flow diagram of the algorithm for cattle detection and counting



is to employ group convolution for the expansion of the channel and number of computational blocks, which employs the same group parameter and channel multiplier to all the computational blocks of a computational layer. Subsequently, the feature map computed by each computational block is shuffled and after that concatenated together. Therefore, the number of channels in each group of the feature maps will be equal to the number of channels in the original architecture. After all these, these groups of feature maps are merged. The capability of E-ELAN to learn more diverse features necessitated applying it in this paper.

Mask YOLOv7 has all-purpose detection pipeline, which comprises three different parts, namely (1) backbone, (2) encoder, and (3) decoder. The structure of the Mask YOLOv7 model primarily comprises three parts, which are (1) input, (2) backbone feature extraction network, and (3) the part for strengthen feature extraction network and predictions. As shown in Figure 2, the input cattle features are detected and extracted by the convolution layers from the image acquired by the drone to form a feature map, and then, the YOLO detection module detects the feature map sent to it. The output results in the feature map are then framed by the detection module and the selection decision is made by the detection module whether to frame coordinates, label the confidence, and categorize information in accordance with the program settings.

Manual annotation was carried out on the ground truth for all the cattle training datasets before training the network with optimized parameters; all these processes were followed by testing the model on the testing dataset for cattle detection and counting.

3.2. Image acquisition and datasets preprocessing

Inaccessibility and lack of suitable open datasets are two major reasons for ineffectiveness recorded by machine learning in cattle detection and counting tasks; moreover, few available public datasets such as FriesianCattle dataset (Andrew et al., 2017) have several flaws such as distorted images, blurred images, similarity between images, limited number of cattle per image, and many more disadvantages. In order to leverage the shortage of open datasets, we employed drone-based data collection method to collect image datasets from the two dataset collection sources, that is, from the feedlot and the open-range environments. The input datasets employed for the detection and counting experiment conducted in this research were collected from (1) the cattle ranch containing a group of Nigerian beef cattle and other complicated background objects and (2) the surrounding housed beef cattle.

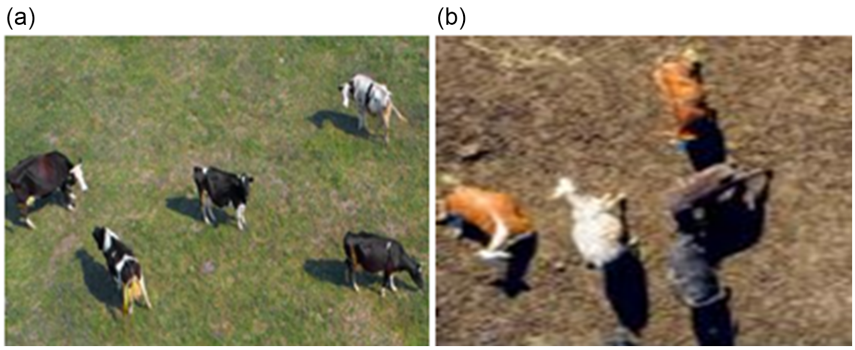
For the application of this proposed model in different scenarios and backgrounds, two cattle ranches and one housed-cattle farm were chosen. As shown in Figure 3, the employed drone DJI Phantom 4 Rtk has the following technical specifications: ISO range of 12800, image sensor of 1" CMOS, maximum image size of 4096 × 2160 pixels, video processor of H.265 4 k at 60 fps, WiFi of 2.4 and 5.8 GHz, flight time of up to 30 min, speed of up to 72 km/h, remote control maximum distance of 5 km, and battery of 6000 mAh. The integrated camera has the following technical specifications: ISO range of 100–6400, glasses of 18–55 mm, image sensor of CMOS, maximum image size of 6016 × 4000 pixels, effective pixels of 24 megapixels, and video processor of full HD 1,920 × 1,080/30 fps.

Figure 3
Drone DJI phantom 4 Rtk

Both the videos and photos used in this study were captured by the camera; however, we preferred video recordings to photos because of many factors including the good qualities that video recordings possess, whereby the captured cattle datasets (for both the feedlot datasets and the open-range datasets) were collected in different scenarios and saved in MOV format as an MPEG 4 video container file, which, by cropping, were later converted to original images in JPEG format. As standard practice, the original images were reduced to the size (512 × 512 pixels for the feedlot cattle datasets and 1280 × 1280 pixels for the open-range cattle datasets) suitable for features extraction by CNN; not only is this reduction method guides against over-fitting during network training, it increases the speed of the model also.

The datasets in the feedlot comprise 800 training images and 200 testing images, and in the open-range, the datasets comprise

Figure 4
Sample of cattle dataset depicting cattle. (a) In the open-range and (b) in the feedlot



800 training images and 200 testing images, making it ratio 4:1 for both training and testing datasets. LabelMe, the web-based image annotation tool was employed in labeling the ground truth of the datasets, which include the cattle heads and their whole body. Figure 4 shows sample of cattle datasets in open-range and feedlot. These labeled data were then stored in a format that conforms to Mask YOLOv7 framework for image annotation.

3.3. The algorithm for detection and counting cattle

YOLOv7, an extended version of the family of YOLO models (YOLO (Redmon et al., 2016), YOLOv2 (Redmon & Farhadi, 2017), YOLOv3 (Redmon & Farhadi, 2018), YOLOv4 (Bochkovskiy et al., 2020), YOLOv5 (Jocher et al., 2022), YOLOv6 (Li et al., 2022)), has high detection speed and accuracy. Mask YOLOv7, just like other models in that family, is a single-stage object detector. Frames of images in the form of features are extracted through a backbone in a YOLO model; the extracted features are mapped in the neck and forwarded to the network head. Just as in YOLO model, both the object's locations and classes are predicted by Mask YOLOv7 with the help of bounding boxes generated for them. To arrive at a final prediction, Mask YOLOv7 carries out a post-processing through non-maximum suppression.

Mask YOLOv7 sets the standard in object detection by possessing a network architecture that predicts bounding boxes accurately more than any known algorithms at similar inference speeds. To achieve this feat, a number of changes were made to the network and training routines of Mask YOLOv7. Four notable improvements of YOLOv7 on the existing YOLOs are (1) extended efficient layer aggregation, (2) model scaling techniques, (3) re-parameterization planning, and (4) auxiliary head coarse-to-fine. The final head trains efficiently more than the auxiliary head because of presence of fewer networks between the auxiliary head and the prediction. Therefore, different levels of supervision were conducted for this head in Mask YOLOv7 resulting in accepting a coarse-to-fine definition where at different granularities, supervision is passed back from the lead head.

3.3.1. Loss function

YOLOv7 loss function comprises three different parts, namely (1) bounding box loss function, (2) objectness loss function, and (3) class loss function. The primary function of bounding box loss function is to measure the prediction box error for the error of coordinate positioning. While the prediction box confidence error is reflected by the objectness loss function, the class loss function gives a reflection of the error committed by the prediction box error for the target category. Mask mechanism was embedded into the backbone of YOLOv7 for instance segmentation, and the mask loss function of the Mask YOLOv7 is defined as the average binary cross-entropy loss, which carries out a sigmoid function on each pixel in the target category.

3.4. Mask YOLOv7 implementation details

The implementation of Mask YOLOv7 was set up on a Python environment inferred with a pre-trained model. The collected data were prepared for training, and the Mask YOLOv7 model was trained using the prepared data before testing and evaluating the model. Mask YOLOv7 model, being a model that was developed not only for object detection but also for image segmentation, was implemented in this work for cattle detection, cattle instance segmentation, and cattle counting. By applying this technique, we locate cattle objects in the images with great precision. Before conducting the training on the model, GPU accessibility is a pre-requirement; this is to avoid training with the CPU, which is time-consuming and inefficient, especially for resource-demanding instance segmentation task. Mask YOLOv7 and its dependencies were installed by cloning the repository and changing the git branch from main to u7, where instance segmentation can be found.

Unlike other YOLO series such as YOLOv5, where all tasks are stored in one codebase, Mask YOLOv7 stores each task on a separate branch. Instead of using test inference, we use Mask YOLOv7 instance segmentation model pre-trained on the COCO dataset to test whether the installation of the environment was successful. We used polygon annotations for the labeling tasks in addition to bounding box around the cattle objects; this is to ensure the model learns the precise shape of each cattle object for both detection and instance segmentation. We applied preprocessing and augmentation after labeling the data to supplement the dataset and stabilize the model from facing object prediction difficulty. The parameter values that we pass matter; therefore, most notably, attention was paid to epochs, batch size, and image size; this is because they are very crucial to performance of model training more than any other parameters.

Epochs are the number of times it will take the model to make a cycle through the data in the course of training. The batch size is the number of samples per gradient update, and the image size is the input image dimensions, which determines the number of pixels the model has to process for each image. Model performance can be improved by increasing the epochs, batch size, and image size parameters; however, this improvement may require more training time and computational resources. To measure Mask YOLOv7

generalization performance as a deep learning model, we run the model on a test dataset; this was carried out to ensure the effectiveness of the model in predicting outcomes for new and unseen data. Test images are usually selected by randomly chosen a sample of the collected data and excluded the sample from the training process.

The Mask YOLOv7 implementation has been executed by employing Google Colab and GPU for the model training. To complete the training of the model, we based the parameters on the total number of images in our datasets, which are 1000 images per feedlot dataset and open-range dataset. Therefore, with a batch size of 50, it takes 20 gradient updates to complete 1 epoch. Furthermore, we trained the network using stochastic gradient descent with 0.001 weight decay, 0.9 momentum, 0.01 initial learning rate, and 0.5 confidence thresholds. After the training was completed, the generated weight was used for the evaluation and inference. Other specifications used are 64-bit of Windows 10 Operating System with 16 GB RAM.

To evaluate the performance of the proposed method in this paper, precision, average precision (AP), and recall are employed as the performance evaluation metrics. Precision refers to the proportion of true positive prediction in all the positive prediction Equation (1); recall refers to the proportion of true positive prediction in all of the positives Equation (2). The precision–recall curve measures the performance of the model based on how large the area enclosed by the curve at different IOU thresholds. AP is expressed in Equation (3). IOU, which stands for Intersection Over Union, is defined as the area of intersection of predicted bounding box and the ground-truth bounding box over the area of their union as expressed in Equation (4).

$$P = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{1}$$

$$R = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{2}$$

$$AP = \sum_{n=1}^{N} [R(n) - R(n-1)].\max P(n) \tag{3}$$

where N is the calculated number of PR points.

$$IOU = \frac{A \cap B}{A \cup B} \times 100 \tag{4}$$

4. Results and Discussion

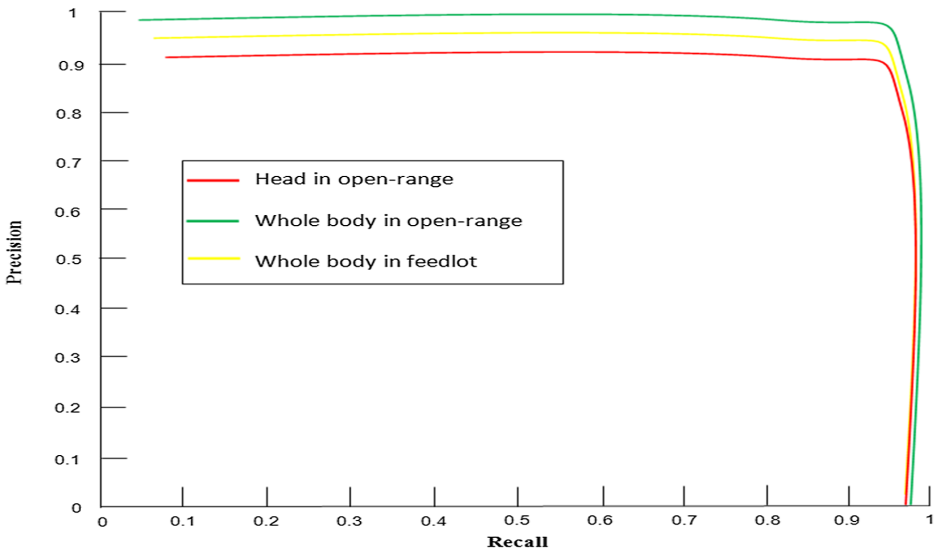
As mentioned earlier in the previous section, the performance evaluation of the proposed method for cattle detection and counting in both feedlot and open-range was performed to compare it with other state-of-the-art detection algorithms. The comparison experiments were carried out on the head and whole body of the cattle.

4.1. Evaluation results on detection and counting

The IOU threshold employed in this work ranges from 0.1 to 0.95 for the APs for bounding box prediction. The precision–recall curve results for whole body detection of the cattle in the feedlot, and head and whole body detection of the cattle in the open-range are shown in Figure 5. Cattle on the open-range are very easy to be detected by the detection algorithm from their head to every other part of their body more than cattle in the feedlot. The cattle in the feedlot are restricted in motion with tendency to be found in different positions such as lying down in their stock densities; this, most often, makes it difficult for the detection algorithm to detect their head, although the detection algorithm sometimes find it difficult to detect cattle's heads in the open-range especially when the cattle on grazing are eating with their head bending down.

Head detection matters in any cattle counting tasks; therefore, due to the aforementioned issues, the performance of the proposed method on detecting cattle's head in the feedlot was very low compared to the full-appearance detection of the cattle. As mentioned earlier, IOU, which stands for Intersection Over Union, is defined as the area of intersection of predicted bounding box and the ground-truth bounding box over the area of their union as expressed in Equation (4); and the threshold, whose dependent variable changes whenever its values which are between 0 and 1

Figure 5
Curves showing precision–recall metrics for the three detection cases



reach optimal, is significant to the performance of object detection tasks.

Choosing either a too large or too small threshold will lead to predicting a bounding box that overlaps. For single label detection, precision was chosen over any other metrics as standard evaluation measure for the evaluation of variable thresholds where three different detection cases were considered. At threshold of 0.5, which is known as the equilibrium point, having the same values in the three cases by precisions and recalls means that all the predictions that are positive are the true positives. This paper conducts experiment on detection and instance segmentation of cattle object for their counting in an image; this is the major reason why precision is preferred to any other metrics for the instance segmentation task which is all about boundary extraction of each cow in the image.

As presented in Table 1, the APs are computed for (1) bounding box prediction which is used for the detection results and (2) mask prediction which is used for the cattle object counting by instance segmentation of the three detection and counting cases over different values of IoU threshold at the equilibrium points. As presented in Tables 1 and 2, the detection of cattle instances and their counting accuracy in the three detection and counting cases show great effectiveness of the proposed method. Table 1 shows that the proposed Mask YOLOv7 method achieved detection accuracy of 90% AP for bounding box in whole body detection in the feedlot, 83% AP in head detection in the open-range, and 95% AP for whole body detection in the open-range.

Table 1
AP scores for bounding box and mask detection for three detection cases

Detection case	AP% (bounding box)	AP% (mask)
Head in open-range	83	80
Whole body in open-range	95	91
Whole body in feedlot	90	88

Table 2
Counting results for three detection cases

Detection case	Counting accuracy (CA)%	CA error (%)
Head in open-range	91	9
Whole body in open-range	95	5
Whole body in feedlot	93	7

Furthermore, Table 1 also shows that the proposed Mask YOLOv7 method achieved detection accuracy of 88% AP for mask in whole body detection in the feedlot, 80% AP in head detection in the open-range, and 91% AP for whole body detection in the open-range. Table 2 shows that the proposed Mask YOLOv7 method achieved 93% counting accuracy in whole body detection result in the feedlot with 7% counting error, 91% counting accuracy in head detection result in the open-range with 9% counting error, and 95% counting accuracy in whole body detection result in the open-range with 5% counting error.

4.2. Comparisons of Mask YOLOv7 with other mainstream object detection models

When evaluated and compared with other state-of-the-art models such as YOLOv3 (regression-based technique), SSD (regression-based technique), and Faster R-CNN (region proposals-based technique) using the same datasets, Mask YOLOv7 (regression-based technique) shows high speed and accuracy as presented in Table 3 where YOLOv3 achieved detection accuracy of 89% AP for bounding box in whole body detection in the feedlot, 81% AP in head detection in the open-range, and 93% AP for whole body detection in the open-range; SSD achieved detection accuracy of 87% AP for bounding box in whole body detection in the feedlot, 79% AP in head detection in the open-range, and 90% AP for whole body detection in the open-range; Faster R-CNN achieved 89% AP for bounding box in whole body detection in the feedlot, 82% AP in head detection in the open-range, and 92% AP for whole body detection in the open-range.

For the counting results, YOLOv3 achieved 91% counting accuracy in whole body detection result in the feedlot, 89% counting accuracy in head detection result in the open-range, and 93% counting accuracy in whole body detection result in the open-range; SSD achieved 89% counting accuracy in whole body detection result in the feedlot, 88% counting accuracy in head detection result in the open-range, and 90% counting accuracy in whole body detection result in the open-range; Faster R-CNN achieved 91% counting accuracy in whole body detection result in the feedlot, 90% counting accuracy in head detection result in the open-range, and 92% counting accuracy in whole body detection result in the open-range. Mask YOLOv7, as employed in the work reported in this paper, has achieved the most accurate detection (AP) and counting accuracy among the compared existing object detection algorithms in the three detection and counting cases. Going by these results, Mask YOLOv7 represents effectiveness in real-world applications regardless the scenes and circumstances under which the images that formed the datasets were collected such as images with complex background, overlapping, occlusion, similarity in cattle coat color, and variation in illumination. Figure 6 shows the comparisons of the prediction performance of

Table 3
Performance comparisons of counting results with competing models

	AP%			Counting accuracy (CA)%		
Method	Head detection (open-range)	Whole body detection (open-range)	Whole body detection (feedlot)	Head detection (open-range)	Whole body detection (open-range)	Whole body detection (feedlot)
YOLOv3	81	93	89	89	93	91
SSD	79	90	87	88	90	89
Faster R-CNN	82	92	89	90	92	91
Mask YOLOv7	**83**	**95**	**90**	**91**	**95**	**93**

Figure 6
Comparisons of the prediction performance of the four object detection algorithms on open-range test images

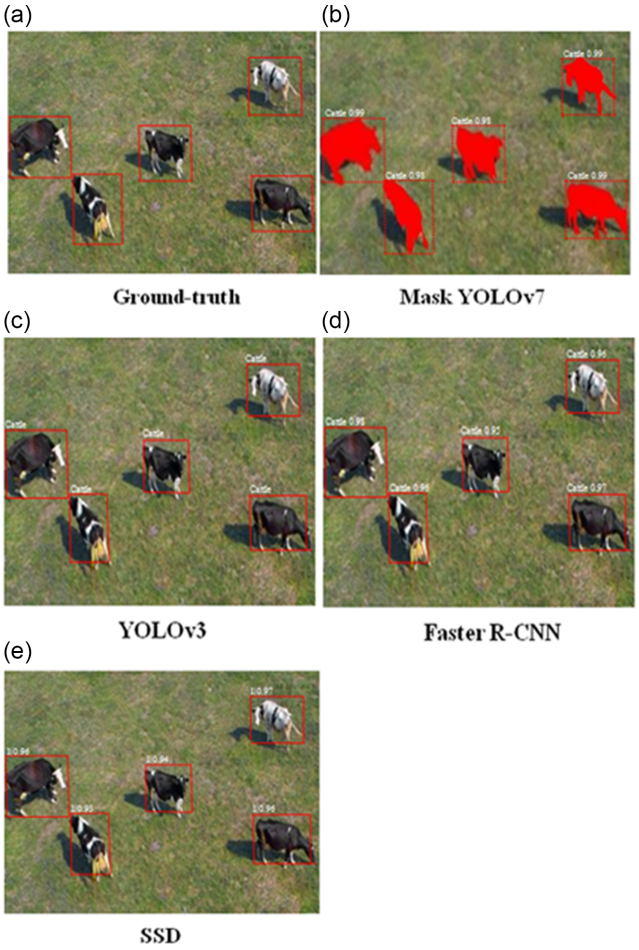
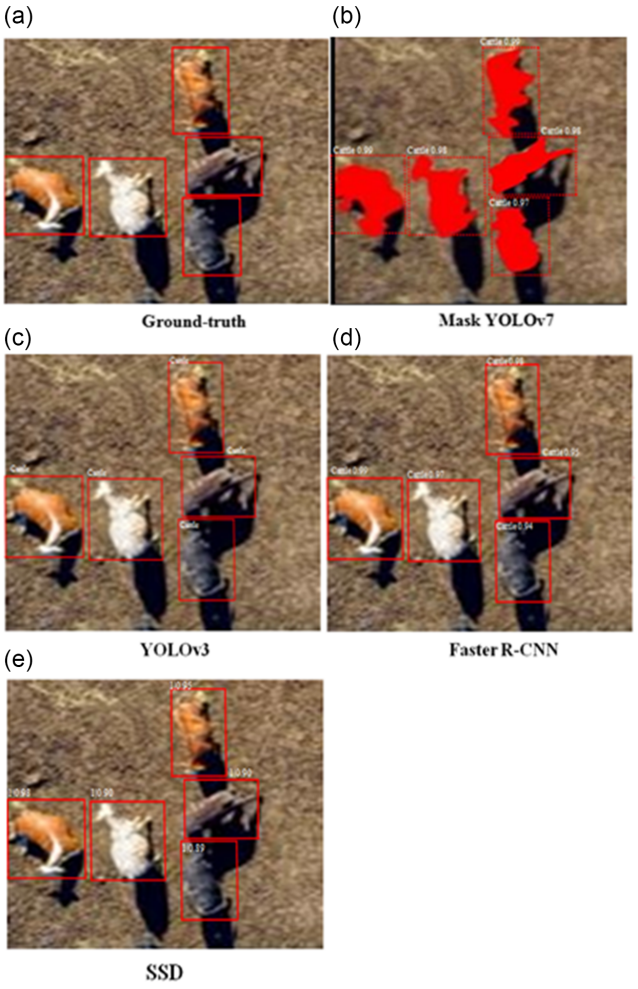


Figure 7
Comparisons of the prediction performance of the four object detection algorithms on feedlot test images



the four object detection algorithms considered in this paper on open-range test images, and Figure 7 shows the comparisons of the prediction performance of the four object detection algorithms considered in this paper on feedlot test images.

5. Discussion

Mask YOLOv7, a state-of-the-art object detection algorithm, is proposed in this paper as a method for achieving cattle detection and counting in drone vision system imagery. The major contribution of this work lies in the high speed and accuracy of Mask YOLOv7 algorithm when applied to the cattle detection and counting tasks. Mask YOLOv7 classifier was designed for binary classification of object in the image (1 for cattle and 0 for no cattle) with the confidence score and mask in place. Mask YOLOv7 uses the regression-based technique to carry out the instance segmentation of the detected cattle object in the image, thereby making it achieve high speed and accuracy compared to other aforementioned state-of-the-art models. Instance segmentation is a popular method used in object detection; it was applied in this paper to aid the counting of cattle unlike the existing works in which the both of bounding box and mask formulation are poorly addressed (Rivas et al., 2018).

The real-time monitoring of livestock for feeding, mating, resting, and other behaviors as telltale for health-related conditions requires a reliable detection technique such as keypoint detection in an image (Mayo et al., 2019; Wang et al., 2022), which is addressed by Mask YOLOv7 in this paper as instance segmentation method for real-time monitoring of farm animals (Piette et al., 2020). The performance of Mask YOLOv7 in detecting cattle in the image requires that the input cattle features are detected and extracted by the convolution layers from the image acquired by the drone to form a feature map, and then, the Mask YOLOv7 detection module detects the feature map sent to it. Different precisions and recalls metrics at different thresholds were measured quantitatively in order to give accurate assessment of the Mask YOLOv7 performance; the evaluation revealed threshold of 0.5 as a better value with AP greater than 90%.

However, threshold of 0.5 is adjustable to fit the application scenario as there is no one-fit-all threshold in object detection. In Bello et al. (2021a), threshold of 0.5 was used to achieve best results in the cattle instance segmentation task. Cattle instance segmentation helps in head detection of cattle although not as accurate as it does for whole body detection. When compared, the detection of cattle head for counting achieved 91% and 95% counting accuracy in whole body detection result in the open-range. Many factors were responsible for the unequal performance and difficulty in detecting cattle head among which is pose variation caused by either movement or grazing behavior of the cattle. As presented earlier, the

comparisons of the proposed Mask YOLOv7 method with other state-of-the-art algorithms on the same datasets for the three detection cases justify the performance of Mask YOLOv7 over others. Mask YOLOv7 was applied to the cattle instance segmentation, thereby adding to the performance of the YOLOv7 in the detection and counting of the cattle. However, there was difficulty by the proposed model to detect heads of cattle in the open range. This further confirms the struggle that YOLOv7, like many other detection algorithms, goes through to detect small objects.

The techniques employed in the monitoring of cattle using drones and the challenges involved are considered in Alanezi et al. (2022) where a strong case was presented for the application of drone systems for the detection and counting of cattle over extensive properties with much interest from animal husbandry. Conclusively, drone applications in animal farming keep expanding geometrically especially in the feedlot operations for monitoring livestock production and activities (Bello et al., 2021b; Ghazali et al., 2022) so much so that its applications are spreading with no barrier for industrial benefits.

6. Conclusion

Drone system application in animal farming is a technology that made possible the detection and counting of animals such as cattle for their inventory and welfare monitoring. In this paper, Mask YOLOv7 model was embedded in the drone system for the cattle object detection and instance segmentation. Annotated imagery acquired with the aid of drone system was employed for the performance evaluation of the proposed method, Mask YOLOv7. The proposed Mask YOLOv7 method achieved 93% counting accuracy in whole body detection result in the feedlot, 91% counting accuracy in head detection result in the open-range, and 95% counting accuracy in whole body detection result in the open-range. The evaluation revealed threshold of 0.5 as a better value with AP greater than 90% at different precisions and recalls metrics. We have as our future work automated cattle inventory system based on drone integrated with enhanced Mask YOLOv7.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

Alanezi, M. A., Shahriar, M. S., Hasan, M. B., Ahmed, S., Sha'Aban, Y. A., & Bouchekara, H. R. (2022). Livestock management with unmanned aerial vehicles: A review. *IEEE Access*, *10*, 45001–45028. http://doi.org/10.1109/ACCESS.2022.3168295

Andrew, W., Greatwood, C., & Burghardt, T. (2017). Visual localisation and individual identification of Holstein Friesian cattle via deep learning. In *IEEE International Conference on Computer Vision*, 2850–2859. https://doi.org/10.1109/ICCVW.2017.336

Andrew, W., Hannuna, S., Campbell, N., & Burghardt, T. (2016). Automatic individual holstein friesian cattle identification via selective local coat pattern matching in RGB-D imagery. In *IEEE International Conference on Image Processing*, 484–488. http://doi.org/10.1109/ICIP.2016.7532404

Auclair-Ronzaud, J., Benoist, S., Dubois, C., Frejaville, M., Jousset, T., Jaffrézic, F., . . . , & Chavatte-Palmer, P. (2020). No-contact microchip monitoring of body temperature in yearling horses. *Journal of Equine Veterinary Science*, *86*, 102892.https://doi.org/10.1016/j.jevs.2019.102892

Beaver, J. T., Baldwin, R. W., Messinger, M., Newbolt, C. H., Ditchkoff, S. S., & Silman, M. R. (2020). Evaluating the use of drones equipped with thermal sensors as an effective method for estimating wildlife. *Wildlife Society Bulletin*, *44*(2), 434–443. https://doi.org/10.1002/wsb.1090

Bello, R. W., Mohamed, A. S. A., & Talib, A. Z. (2021a). Contour extraction of individual cattle from an image using enhanced Mask R-CNN instance segmentation method. *IEEE Access*, *9*, 56984–57000. https://doi.org/10.1109/ACCESS.2021.3072636

Bello, R. W., Mohamed, A. S. A., & Talib, A. Z. (2021b). Enhanced Mask R-CNN for herd segmentation. *International Journal of Agricultural and Biological Engineering*, *14*(4), 238–244. https://doi.org/10.25165/j.ijabe.20211404.6398

Bello, R. W., Mohamed, A. S. A., Talib, A. Z., Olubummo, D. A., & Enuma, O. C. (2021c). Enhanced deep learning framework for cow image segmentation. *IAENG International Journal of Computer Science*, *48*(4), 1–10.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv Preprint:2004.10934*.

Chabot, D., Dillon, C., & Francis, C. M. (2018). An approach for using off-the-shelf object-based image analysis software to detect and count birds in large volumes of aerial imagery. *Avian Conservation & Ecology*, *13*(1), 15. http://doi.org/10.5751/ACE-01205-130115

Cheema, G. S., & Anand, S. (2017). Automatic detection and recognition of individuals in patterned species. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 27–38. https://doi.org/10.1007/978-3-319-71273-4_3

Du, X., & Zhou, J. (2018). Application of biosensors to detection of epidemic diseases in animals. *Research in Veterinary Science*, *118*, 444–448. https://doi.org/10.1016/j.rvsc.2018.04.011

Dujon, A. M., Ierodiaconou, D., Geeson, J. J., Arnould, J. P., Allan, B. M., Katselidis, K. A., & Schofield, G. (2021). Machine learning to detect marine animals in UAV imagery: Effect of morphology, spacing, behaviour and habitat. *Remote Sensing in Ecology and Conservation*, *7*(3), 341–354. https://doi.org/10.1002/rse2.205

Eikelboom, J. A., Wind, J., van de Ven, E., Kenana, L. M., Schroder, B., de Knegt, H. J., . . . , & Prins, H. H. (2019). Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, *10*(11), 1875–1887. https://doi.org/10.1111/2041-210X.13277

Ghazali, M. H. M., Azmin, A., & Rahiman, W. (2022). Drone implementation in precision agriculture–A survey. *International Journal of Emerging Technology and Advanced Engineering*, *12*(4), 67–77. http://doi.org/10.46338/ijetae0422_10

Girshick, R. (2015). Fast R-CNN. In *IEEE International Conference on Computer Vision*, 1440–1448. https://doi.org/10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic

segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/CVPR.2014.81

González, L. A., Kyriazakis, I., & Tedeschi, L. O. (2018). Precision nutrition of ruminants: Approaches, challenges and potential gains. *Animal*, *12*, s246–s261. https://doi.org/10.1017/S1751731118002288

Gu, J., Wang, Z., Gao, R., & Wu, H. (2017). Cow behaviour recognition based on image analysis and activities. *International Journal of Agricultural and Biological Engineering*, *10*(3), 165–174.

Halachmi, I., Guarino, M., Bewley, J., & Pastell, M. (2019). Smart animal agriculture: Application of real-time sensors to improve animal well-being and production. *Annual Review of Animal Biosciences*, *7*, 403–425. https://doi.org/10.1146/annurev-animal-020518-114851

Han, L., Tao, P., & Martin, R. R. (2019). Livestock detection in aerial images using a fully convolutional network. *Computational Visual Media*, *5*, 221–228. https://doi.org/10.1007/s41095-019-0132-5

Hatton-Jones, K. M., Christie, C., Griffith, T. A., Smith, A. G., Naghipour, S., Robertson, K., . . . , & du Toit, E. F. (2021). A YOLO based software for automated detection and analysis of rodent behaviour in the open field arena. *Computers in Biology and Medicine*, *134*, 104474. https://doi.org/10.1016/j.compbiomed.2021.104474

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969.

Hu, H., Dai, B., Shen, W., Wei, X., Sun, J., Li, R., & Zhang, Y. (2020). Cow identification based on fusion of deep parts features. *Biosystems Engineering*, *192*, 245–256. https://doi.org/10.1016/j.biosystemseng.2020.02.001

Hu, T., Yan, R., Jiang, C., Chand, N. V., Bai, T., Guo, L., & Qi, J. (2023). Grazing sheep behaviour recognition based on improved YOLOV5. *Sensors*, *23*(10), 4752. https://doi.org/10.3390/s23104752

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., . . . , & Jain, M. (2022). Ultralytics/yolov5: v7. 0-YOLOv5 SOTA realtime instance segmentation. *Zenodo*. http://doi.org/10.5281/zenodo.7347926

Kumar, S. A., & Ilango, P. (2018). The impact of wireless sensor network in the field of precision agriculture: A review. *Wireless Personal Communications*, *98*, 685–698. https://doi.org/10.1007/s11277-017-4890-z

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., . . . , & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv Preprint:2209.02976v1*.

Liu, H., Reibman, A. R., & Boerman, J. P. (2020). A cow structural model for video analytics of cow health. *arXiv Preprint:2003.05903*.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In *European Conference on Computer Vision*, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Madasamy, K., Shanmuganathan, V., Kandasamy, V., Lee, M. Y., & Thangadurai, M. (2021). OSDDY: Embedded system-based object surveillance detection system with small drone using deep YOLO. *EURASIP Journal on Image and Video Processing*, *2021*(1), 19. https://doi.org/10.1186/s13640-021-00559-1

Mayo, L. M., Silvia, W. J., Ray, D. L., Jones, B. W., Stone, A. E., Tsai, I. C., . . . , & Heersche Jr, G. (2019). Automated estrous detection using multiple commercial precision dairy monitoring technologies in synchronized dairy cows. *Journal of Dairy Science*, *102*(3), 2645–2656. https://doi.org/10.3168/jds.2018-14738

O'Leary, N. W., Byrne, D. T., O'Connor, A. H., & Shalloo, L. (2020). Invited review: Cattle lameness detection with accelerometers. *Journal of Dairy Science*, *103*(5), 3895–3911. https://doi.org/10.3168/jds.2019-17123

Piette, D., Norton, T., Exadaktylos, V., & Berckmans, D. (2020). Individualised automated lameness detection in dairy cows and the impact of historical window length on algorithm performance. *Animal*, *14*(2), 409–417. https://doi.org/10.1017/S1751731119001642

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 779–788.

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7263–7271.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv Preprint:1804.02767*.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Rivas, A., Chamoso, P., González-Briones, A., & Corchado, J. M. (2018). Detection of cows using drones and convolutional neural networks. *Sensors*, *18*(7), 2048. https://doi.org/10.3390/s18072048

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*, 157–173. https://doi.org/10.1007/s11263-007-0090-8

Sadgrove, E. J., Falzon, G., Miron, D., & Lamb, D. W. (2021). The segmented colour feature extreme learning machine: Applications in agricultural robotics. *Agronomy*, *11*(11), 2290. https://doi.org/10.3390/agronomy11112290

Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., & Naemura, T. (2020). Cows detection and counting in UAV images based on convolutional neural networks. *International Journal of Remote Sensing*, *41*(1), 31–52. https://doi.org/10.1080/01431161.2019.1624858

Sharma, B., & Koundal, D. (2018). Cattle health monitoring system using wireless sensor network: A survey from innovation perspective. *IET Wireless Sensor Systems*, *8*(4), 143–151. https://doi.org/10.1049/iet-wss.2017.0060

Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote Sensing*, *12*(19), 3136. https://doi.org/10.3390/rs12193136

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Newton, E. J., Boughton, R. K., Ivan, J. S., . . . , & Miller, R. S. (2020). Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. *Ecology and Evolution*, *10*(19), 10374–10383. https://doi.org/10.1002/ece3.6692

Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A review on UAV-based applications for precision agriculture. *Information*, *10*(11), 349. https://doi.org/10.3390/info10110349

Ulhaq, A., Adams, P., Cox, T. E., Khan, A., Low, T., & Paul, M. (2021). Automated detection of animals in low-resolution airborne thermal imagery. *Remote Sensing*, *13(*16), 3276. https://doi.org/10.3390/rs13163276

Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv Preprint:2207.02696*.

Wang, F. K., Shih, J. Y., Juan, P. H., Su, Y. C., & Wang, Y. C. (2021). Non-invasive cattle body temperature measurement using infrared thermography and auxiliary sensors. *Sensors*, *21*(7), 2425. https://doi.org/10.3390/s21072425

Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., …, & Fortson, L. (2019). Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, *10*(1), 80–91.https://doi.org/10.1111/2041-210X.13099

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., …, & Schneider, D. (2020). Automated cows counting using Mask R-CNN in quadcopter vision system. *Computers and Electronics in Agriculture*, *171*, 105300. https://doi.org/10.1016/j.compag.2020.105300

Yang, X., Bist, R., Subedi, S., Wu, Z., Liu, T., & Chai, L. (2023). An automatic classifier for monitoring applied behaviors of cage-free laying hens with deep learning. *Engineering Applications of Artificial Intelligence*, *123*, 106377. https://doi.org/10.1016/j.engappai.2023.106377

Zhao, K., Jin, X., Ji, J., Wang, J., Ma, H., & Zhu, X. (2019). Individual identification of Holstein dairy cows based on detecting and matching feature points in body images. *Biosystems Engineering*, *181*, 128–139. https://doi.org/10.1016/j.biosystemseng.2019.03.004

Zin, T., & Tin, P. (2018). Image technology based cow identification system using deep learning. In *Proceedings of the International MultiConference of Engineers and Computer Scientists, 1*, 320–323.

How to Cite: Bello, R. W., & Oladipo, M. A. (2024). Mask YOLOv7-Based Drone Vision System for Automated Cattle Detection and Counting. *Artificial Intelligence and Applications*, *2*(2), 115–125. https://doi.org/10.47852/bonviewAIA42021603