DS Journal of Artificial Intelligence and Robotics
Volume 1 Issue 2, 15-25, Oct - Dec 2023
ISSN: 2583-9926/ https://doi.org/10.59232/AIR-V1I2P102

Original Article

Mask the Target: Mask R-CNN-Based Approach for Open-Range Individual Cattle Segmentation

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

[1]*Department of Mathematics and Computer Science, University of Africa, Toru-Orua, Sagbama, Bayelsa, Nigeria.*
[2]*Department of Agricultural Economics and Extension, University of Africa, Toru-Orua, Sagbama, Bayelsa, Nigeria.*

[1]sirbrw@yahoo.com

Received: 10 July 2023; Revised: 31 August 2023; Accepted: 15 September 2023; Published: 12 October 2023;

Abstract - Targeting individual cattle in an open-range has been a computer vision problem with great agricultural implications for animal health monitoring, behavior monitoring, and economic benefit. In this paper, the individual cattle targets in the open-range are segmented using Mask R-CNN's detect and segment approach. We evaluate the performance of the model proposed in this study using the Intersection Over Union (IOU) threshold of 0.5, Average Precision (AP) and mean Average Precision (mAP). The results of the experiment conducted with the Acquired dataset in this study show that the proposed model achieves an accuracy of 95%, thereby affirming the potential of the Mask R-CNN model to perform competitively with any other existing object detection and instance segmentation models for cattle target segmentation with high accuracy and average precision.

Keywords - Cattle, Mask R-CNN, Open-range, Segmentation, Target.

1. Introduction

According to [1], meat consumption is on the increase demand by the populace, and to meet their demand, there is a need to address the challenges confronting livestock production and their management. Lack of expertise in managing the livestock, problems with remote livestock monitoring, 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. 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., the improvement and maximization of meat production [2]. Therefore, it is necessary to address the aforementioned challenges in animal husbandry with appropriate methods.

To address the aforementioned challenges, [3] proposed an automated visual identification system for individual Holstein Friesian cows from dorsal RGB-D-based imagery. Predictions were generated using Support Vector Machine (SVM) and Radial Basis Function (RBF) kernels based on the ASIFT descriptor structure. 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 selecting and using features to recover the cow's identity. A method based on image entropy was proposed by [4] to recognize and identify the behavior of cow objects in 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, [5] posited the appropriateness of computer vision pipelines that make use of the architectures of deep neural networks to perform the automated detection and identification of individual Holstein Friesian cows 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 cows with 99.3% accuracy.

[6] Proposed object detection based on the Faster R-CNN framework for efficient detection of animals in images. They trained a linear SVM classifier to recognize individual animals using the features extracted from AlexNet of the animal's flank. The techniques of deep learning were employed in [7] 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 cow's black and white body patterns. The cow's body placed in the Rotary Milking Parlor was detected using inter-frame differencing and horizontal histogram-based methods. The predefined distance value was used to extract the cow's body region, which served as input data to train the deep convolutional neural network.

A method of artificial intelligence-based Convolutional Neural Networks (CNNs) was employed in [8] 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 using the same algorithmic process of CNN training. A computer vision system was proposed by [9] that could identify individual dairy cows. This was made possible by using videos showing the cow's side view in motion. Detection and location of the cow object and its body area as the individual identity information was 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 accurately calculating the feature points in the body pattern of cows using the 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.

[10] 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 to represent the 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. Two convolutional neural networks were applied to the detection system to extract the key features from the raw images and select individual features for conversion into a structural model. In order to enable the system to work with different quality 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 cows was proposed by [11]. For the extraction of the cow object in the side view image and the cow's head, trunk and leg parts, they applied the YOLO method [12] 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 Support Vector Machine (SVM) classifier was used for the cow image classification. An automated method based on Mask R-CNN capable of counting cattle in a quad copter vision system was proposed by [13]. 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. The performance evaluation method was used in verifying the optimal IOU threshold (0.5) and the detection performance of the algorithm for full appearance.

Similar work to [13] was carried out by [14] to aid in managing open-range cattle; they proposed a system based on Convolutional Neural Networks (CNNs) for detecting and counting cows using Unmanned Aerial Vehicle (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 the 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.

2. Materials and Methods

Datasets of cow images in an open-range and the camera for capturing them made up the primary materials in this paper. For the methods they include an overview of the proposed model architecture and our framework, image acquisition, dataset preparation and preprocessing, the implementation details of Mask R-CNN [15], and performance evaluation metrics.

2.1. 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 captured images solely for cattle target detection and segmentation using the Mask R-CNN algorithm. Figure 1 shows the Mask R-CNN architecture for performing the cattle target segmentation task. Mask R-CNN has an all-purpose detection and segmentation pipeline which comprises the following structural parts: 1) input, 2) residual network-based convolutional backbone for feature extraction, 3) the Region Proposal Network (RPN) for proposing region of interest, and 4) the head region part which comprises feature map of fixed size and fully connected layers for object's bounding box, class, and mask. As shown in Figure 1, the features of the input cattle target are detected and extracted by the convolution layers from the image acquired to form a feature map. Then, the RPN detects the feature map sent to it and aligns all the unaligned regions of interest. The output results from the aligned regions of interest are fixed into the same size feature map for fully connected layers to process into bounding box, class and mask using the mask branch, which is a unique addition to the Faster R-CNN algorithm [16].

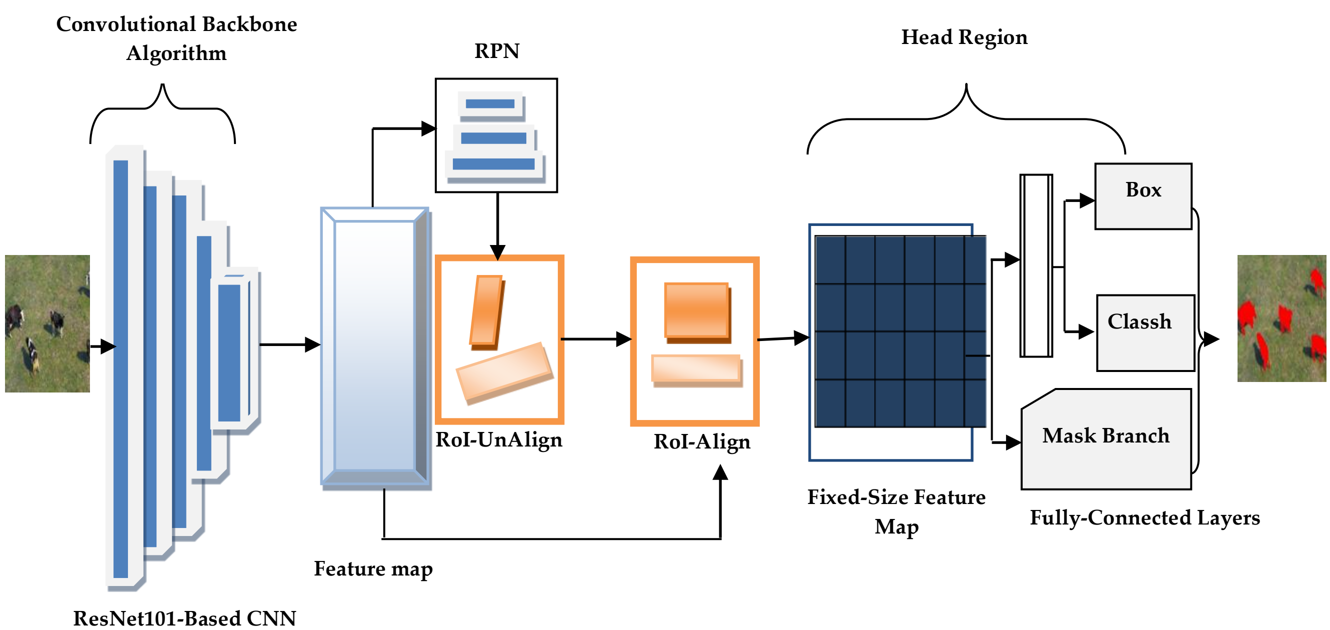


Fig. 1 Mask R-CNN architecture for performing cattle target segmentation task

2.2. Image Acquisition and Datasets Preprocessing

Inaccessibility and lack of suitable open datasets are two major reasons for the ineffectiveness recorded by machine learning in cattle detection and segmentation tasks; moreover, few available public datasets, such as the Friesian Cattle dataset [5], have several flaws, such as distorted images, blurred images, the similarity between images, a limited number of cattle per image and many more disadvantages. In order to leverage the shortage of open datasets, we employed a camera-trapped data collection method to collect image datasets from the open-range environment. The input datasets employed for the detection and segmentation experiment conducted in this research were collected from the cattle ranch containing a group of cattle and other complicated background objects. For the application of this proposed model in different scenarios and backgrounds, the cattle ranch was chosen.

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 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 a size 1280 × 1280 pixels suitable for feature extraction by CNN; not only does this reduction method guide against over-fitting during network training, it also increases the model's speed. The datasets comprise 800 training images and 200 testing images, making it a ratio of 4:1 for both training and testing datasets. LabelMe [17], the web-based image annotation tool, was employed to label the ground truth of the datasets, including the cattle heads and their whole body. Figure 2 shows a sample of cattle datasets in the open-range. These labeled data were then stored in a format that conforms to the Mask R-CNN framework for image annotation.

Fig. 2 Sample of cattle images from the open-range dataset

2.3. Mask R-CNN Implementation Details

The implementation of Mask R-CNN was set up on a Python environment inferred with a pre-trained model. The collected data were prepared for training, and the Mask R-CNN model was trained using the prepared data before testing and evaluating the model. Mask R-CNN model, being a model that was developed not only for object detection but also for image segmentation, was implemented in this work for cattle target detection and segmentation in the open-range. By applying this technique, we locate individual 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 tasks. Mask R-CNN and its dependencies were installed by cloning the repository. We use the Mask R-CNN instance segmentation model pre-trained on the COCO dataset to test whether the environment installation was successful. We used polygon annotations for the labeling tasks in addition to bounding boxes 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 we pass matter; therefore, most notably, attention was paid to epochs, batch size and image size. This is because they are very crucial to the 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 R-CNN generalization performance as a deep learning model, we ran 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 choosing a sample of the collected data and excluding the sample from the training process. The Mask R-CNN implementation has been executed by employing Google Colab and GPU for the model training. To complete the model's training, we based the parameters on the total number of images in our datasets, 1000 images per 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 (SGD) 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 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 predictions (Equation 1); recall refers to the proportion of true positive prediction in all the positives (Equation 2). The precision-recall curve measures the model's performance based on how large the area is enclosed by the curve at different IOU thresholds. "Average precision is expressed in Equation (3). IOU, which stands for Intersection Over Union, is defined as the area of intersection of the predicted bounding box and the ground-truth bounding box over the area of their union as expressed in Equation (4).

$$P= (\text{True Positive})/(\text{True Positive}+\text{False Positive}) \quad (1)$$

$$R= (\text{True Positive})/(\text{True Positive}+\text{False Negative}) \quad (2)$$

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

Where N is the calculated number of PR points.

$$IOU= (A\cap B)/(A\cup B)\times 100 \quad (4)$$

For satisfactory results, the loss function was used during the training of the following: a) bounding box regression, b) object class prediction, and c) mask branch segmentation. Equation (5) is the loss function representation:

$$L=L_{ce}+L_{be}+L_{me} \quad (5)$$

Where L is the loss function, L_{ce} is the classification error, L_{be} is the bounding box regression error, and L_{me} is the mask error.

3. Results and Discussion

As mentioned earlier in the previous section, the performance evaluation of the proposed method for cattle target detection and segmentation in the 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.

3.1. Evaluation Results on Detection and Segmentation

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 head and whole-body detection of the cattle in the open-range are shown in Figure 3. Although the detection algorithm sometimes finds it difficult to detect cattle's heads in the open-range especially when the cattle on grazing are eating with their head bending down, the cattle are very easy to be detected by the segmentation algorithm from their head to every other part of their body irrespective of their background.

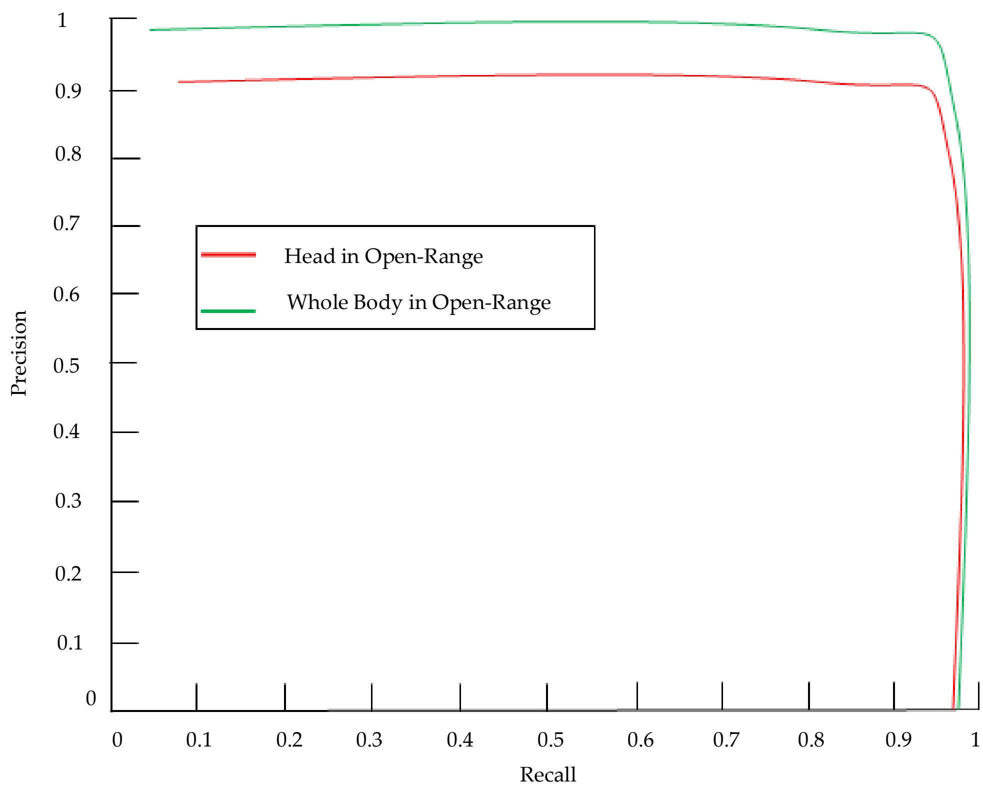


Fig. 3 Curves showing precision-recall metrics for the detection cases

As mentioned earlier, IOU, which stands for Intersection Over Union, is defined as the area of intersection of the 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 reach optimal is significant to the performance of object detection tasks.

Choosing a too-large or too-small threshold will predict a bounding box that overlaps. Precision was chosen over any other metrics for single-label detection as a standard evaluation measure for evaluating variable thresholds where different detection cases were considered. At a threshold of 0.5, known as the equilibrium point, having the same values in the considered cases by precisions and recalls means that all the positive predictions are the true positives. This paper experiments on the detection and instance segmentation of individual cattle targets 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 the 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 target by instance segmentation of the detection cases over different values of IOU threshold at the equilibrium points. As presented in Table 1, the detection of cattle instances and their segmentation accuracy in the detection and segmentation cases show great effectiveness of the proposed method. Table 1 shows that the proposed Mask R-CNN method achieved detection accuracy of 83% AP in head detection and 95% AP for whole body detection.

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

Detection Case	AP% (Bounding Box)	AP% (Mask)
Head	83	80
Whole Body	95	91

Furthermore, Table 1 also shows that the proposed Mask R-CNN method achieved detection accuracy of 80% and 91% AP for whole-body detection.

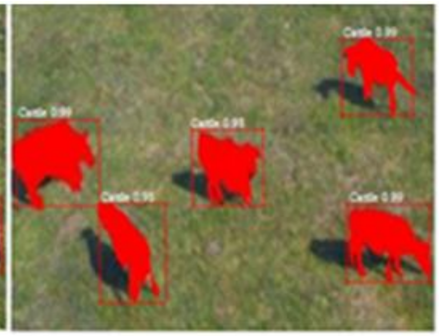
3.2. Comparisons of Mask R-CNN with other Mainstream Models

When evaluated and compared with other state-of-the-art models such as YOLOv3 (regression-based technique) [18], SSD (regression-based technique) [19], and Faster R-CNN (region-proposals-based technique) using the same datasets, Mask R-CNN shows high accuracy as presented in Table 2, where YOLOv3 achieved detection accuracy of 81% AP in head detection and 93% AP for whole body detection; SSD achieved detection accuracy of 79% AP in head detection and 90% AP for whole body detection; Faster R-CNN achieved 82% AP in head detection and 92% AP for whole body detection.

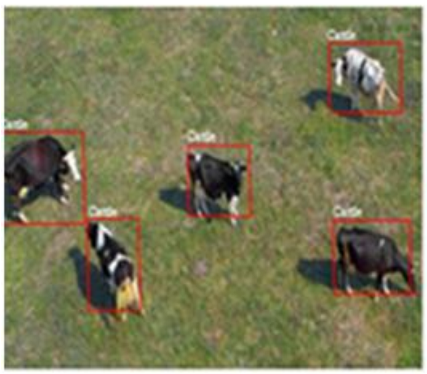
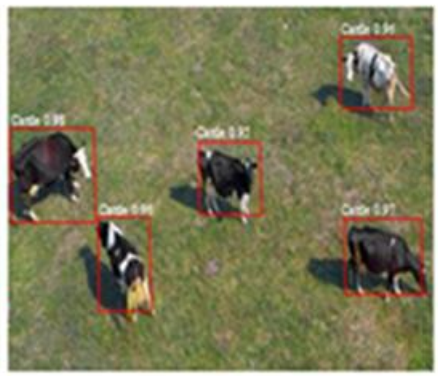
Mask R-CNN, as employed in the work reported in this paper, has achieved the most accurate detection (AP) and segmentation accuracy among the compared existing object detection and segmentation algorithms in the detection and segmentation cases. By these results, Mask R-CNN represents effectiveness in real-world applications regardless of the scenes and circumstances under which the images that formed the datasets were collected, such as images with a complex background, overlapping, occlusion, similarity in cattle coat color, variation in illumination, and so on. Figure 4 compares the prediction performance of the four object detection algorithms considered in this paper on open-range test images.

Table 2. Performance comparisons of segmentation results with competing models

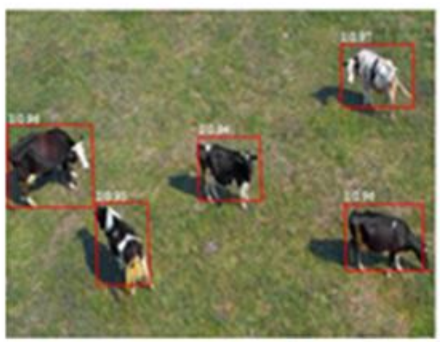
Method	AP (%)		Classification Accuracy (%)	
	Head Detection	Whole Body Detection	Head Detection	Whole Body Detection
YOLOv3	81	93	89	93
SSD	79	90	88	90
Faster R-CNN	82	92	90	92
Mask R-CNN	83	95	91	95

(a) Ground-Truth **(b) Mask R-CNN**

(c) YOLOv3 **(d) Faster R-CNN**



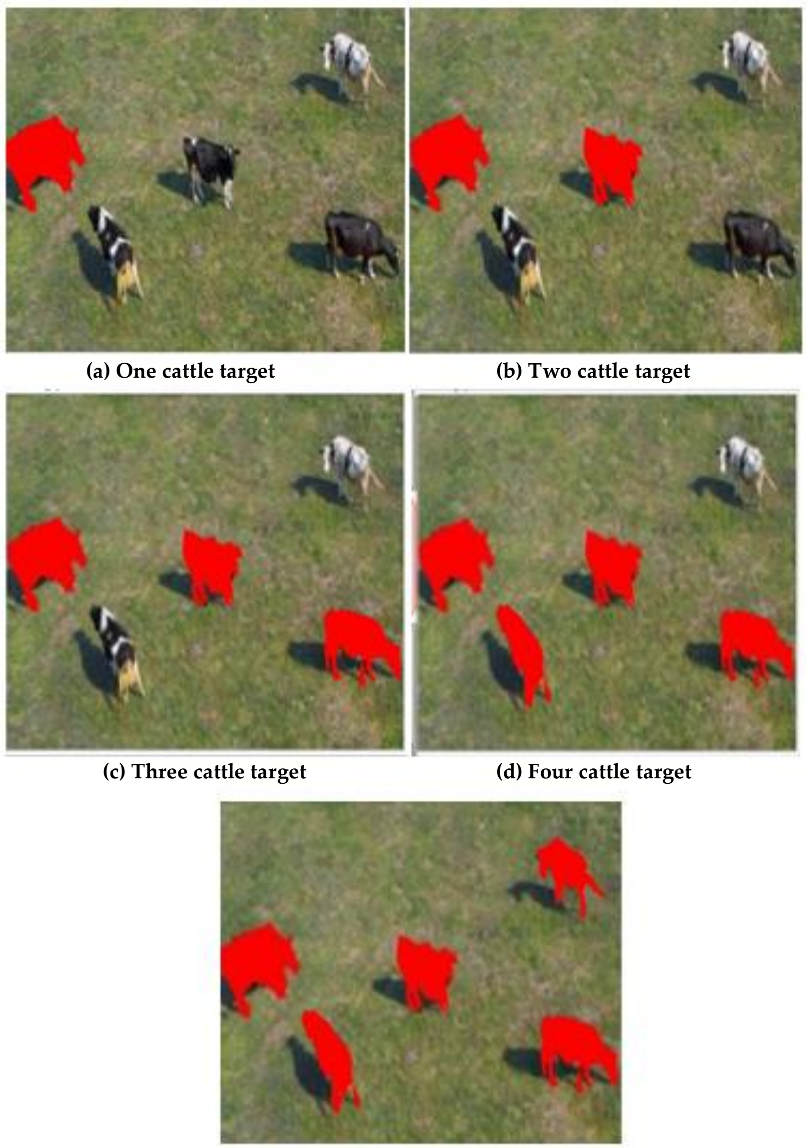
(e) SSD

Fig. 4 Comparisons of the prediction performance of the four object detection algorithms on test images

4. Discussion

Mask R-CNN, a state-of-the-art object detection and instance segmentation algorithm, is proposed in this paper as a method for achieving cattle target segmentation in open-range captured images. The major contribution of this work lies in the high accuracy of the Mask R-CNN algorithm when applied to cattle target detection and segmentation tasks. Mask R-CNN classifier was designed for binary classification of objects in the image (1 for cattle and 0 for no cattle) with the confidence score and mask in place. Mask R-CNN uses the region proposals-based technique to carry out the instance segmentation of the detected cattle object in the image, thereby making it achieve high 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 segmentation of individual cattle targets, unlike the existing works in which mask formulation for individual targets is not well addressed [13, 20].

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 [21-23], which is addressed by Mask R-CNN in this paper as instance segmentation method for real-time monitoring of farm animals [24-25]. The performance of Mask R-CNN 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 camera to form a feature map. Then, the Mask R-CNN's Region Proposal Network (RPN) detects the feature map sent to it and proposes a Region of Interest (ROI) on the image, followed by the fully connected layers comprising box, class, and mask branch for generating the masked output. Different precisions and recall metrics at different thresholds were measured quantitatively to give an accurate assessment of the Mask R-CNN performance; the evaluation revealed a threshold of 0.5 as a better value with AP greater than 90%.



(a) One cattle target **(b) Two cattle target**

(c) Three cattle target **(d) Four cattle target**

(e) Five cattle target

Fig. 5 Targeting individual cattle in an open-range using Mask R-CNN

However, the threshold of 0.5 is adjustable to fit the application scenario as there is no one-fit-all threshold in object detection. In [20], the threshold of 0.5 was used to achieve the 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. Many factors were responsible for the unequal performance and difficulty in detecting cattle head, among which is pose variation caused by the cattle's movement or grazing behaviour. As presented earlier, the comparisons of the proposed Mask R-CNN method with other state-of-the-art algorithms on the same datasets for the detection cases justifies the performance of Mask R-CNN over others. Mask R-CNN was applied to the cattle target segmentation, thereby adding to the algorithm's potential for possible tracking and counting of animals in the near future. More photos of individual cattle target segmentation are shown in Figure 5.

5. Conclusion

Mask R-CNN approach in animal farming is a technology that made possible the detection, segmentation and classification of animals such as cattle for their inventory, health and behavioral monitoring. In this paper, the Mask R-CNN model was applied for cattle target detection and segmentation. Annotated imagery acquired with the aid of a camera was employed for the performance evaluation of the proposed Mask R-CNN method. The proposed method achieved detection accuracy of 83% AP in head detection and 95% AP for whole body detection. The evaluation revealed a threshold of 0.5 as a better value with AP greater than 90% at different precisions and recall metrics. We have as our future work to enhance the proposed algorithm for possible applications in tracking and counting animals in the open-range.

Authors' Contribution

"Conceptualization, R.W. Bello; Methodology, R.W. Bello; Software, R.W. Bello; Validation, R.W. Bello and M.A. Oladipo; Formal Analysis, R.W. Bello; Investigation, R.W. Bello and M.A. Oladipo; Resources, R.W. Bello and M.A. Oladipo; Data Curation, M.A. Oladipo; Writing – Original Draft Preparation, R.W. Bello; Writing – Review & Editing, R.W. Bello; Visualization, R.W. Bello; Supervision, R.W. Bello and M.A. Oladipo; Project Administration, R.W. Bello and M.A. Oladipo; Funding Acquisition, R.W. Bello and M.A. Oladipo".

References

[1] Rajendra P. Sishodia, Ram L. Ray, and Sudhir K. Singh, "Applications of Remote Sensing in Precision Agriculture: A Review," *Remote Sensing*, vol. 12, no. 19, pp. 1-31, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Subramania Ananda Kumar, and Paramasivam Ilango, "The Impact of Wireless Sensor Network in the Field of Precision Agriculture: A Review," *Wireless Personal Communications*, vol. 98, pp. 685-698, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[3] William Andrew et al., "Automatic Individual Holstein Friesian Cattle Identification via Selective Local Coat Pattern Matching in RGB-D Imagery," *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 484-488, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[4] Gu Jingqiu et al., "Cow Behaviour Recognition Based on Image Analysis and Activities," *International Journal of Agricultural and Biological Engineering*, vol. 10, no. 3, pp. 165-174, 2017. [Google Scholar] [Publisher Link]

[5] William Andrew, Colin Greatwood, and Tilo Burghardt, "Visual Localisation and Individual Identification of Holstein Friesian Cattle via Deep Learning," *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 2850-2859, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[6] Gullal Singh Cheema, and Saket Anand, "Automatic Detection and Recognition of Individuals in Patterned Species," *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 27-38, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] Thi Thi Zin et al., "Image Technology Based Cow Identification System Using Deep Learning," *Proceedings of the International MultiConference of Engineers and Computer Scientists*, pp. 1-4, 2018. [Google Scholar] [Publisher Link]

[8] Alberto Rivas et al., "Detection of Cows Using Drones and Convolutional Neural Networks," *Sensors*, vol. 18, no. 7, pp. 1-15, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[9] Kaixuan Zhao et al., "Individual Identification of Holstein Dairy Cows Based on Detecting and Matching Feature Points in Body Images," *Biosystems Engineering*, vol. 181, pp. 128-139, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] He Liu, Amy R. Reibman, and Jacquelyn P. Boerman, "A Cow Structural Model for Video Analytics of Cow Health," *arXiv*, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Hengqi Hu et al., "Cow Identification Based on Fusion of Deep Parts Features," *Biosystems Engineering*, vol. 192, pp. 245-256, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Joseph Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[13] Beibei Xu et al., "Automated Cows Counting Using Mask R-CNN in Quadcopter Vision System," *Computers and Electronics in Agriculture*, vol. 171, pp. 1-12, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Wen Shao et al., "Cattle Detection and Counting in UAV Images Based on Convolutional Neural Networks," *International Journal of Remote Sensing*, vol. 41, no. 1, pp. 31-52, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Kaiming He et al., "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980-2988, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[16] Shaoqing Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[17] Bryan C. Russell et al., "LabelMe: A Database and Web-Based Tool for Image Annotation," *International Journal of Computer Vision*, vol. 77, pp. 157-173, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[18] Joseph Redmon, and Ali Farhadi, "YOLOv3: An Incremental Improvement," *arXiv*, pp. 1-6, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] Wei Liu et al., "SSD: Single Shot Multibox Detector," *European Conference on Computer Vision*, pp. 21-37, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[20] Tao Feng et al., "Cattle Target Segmentation Method in Multi-Scenes Using Improved DeepLabV3+ Method," *Animals*, vol. 13, no. 15, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Yongliang Qiao, Matthew Truman, and Salah Sukkarieh, "Cattle Segmentation and Contour Extraction Based on Mask R-CNN for Precision Livestock Farming," *Computers and Electronics in Agriculture*, vol. 165, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] Tiara Lestari Subaran, Transmissia Semiawan, and Nurjannah Syakrani, "Mask R-CNN and GrabCut Algorithm for an Image-Based Calorie Estimation System," *Journal of Information Systems Engineering and Business Intelligence*, vol. 8, no. 1, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Jennifer Salau, and Joachim Krieter, "Instance Segmentation with Mask R-CNN Applied to Loose-Housed Dairy Cows in a Multi-Camera Setting," *Animals*, vol. 10, no. 12, pp. 1-19, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Venkata Sai Praveen Gunda et al., "A Hybrid Deep Learning Based Robust Framework for Cattle Identification," *2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, Bhubaneswar, India, pp. 1-5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] D. Piette et al., "Individualised Automated Lameness Detection in Dairy Cows and the Impact of Historical Window Length on Algorithm Performance," *Animal*, vol. 14, no. 2, pp. 409-417, 2020. [CrossRef] [Google Scholar] [Publisher Link]