PRINT ISSN 1119-8362
Electronic ISSN 2659-1502

Full-text Available Online at
https://www.ajol.info/index.php/jasem
http://www.bioline.org.br/ja

J. Appl. Sci. Environ. Manage.
Vol. 27 (5) 953-958 May 2023

YOLOv7 Applied to Livestock Image Detection and Segmentation Tasks in Cattle Grazing Behavior, Monitor and Intrusions

[1]MORADEYO, OM; [1]OLANIYAN, AS; [1]OJOAWO, AO; [1]OLAWALE, JA; *[2]BELLO, RW

[1]Department of Computer Science, Adeseun Ogundoyin Polytechnic, Eruwa, Oyo State, Nigeria
*[*2]Department of Mathematics and Computer Science, University of Africa, 561101 Sagbama, Bayelsa State, Nigeria*

**Corresponding Author Email Address: sirbrw@yahoo.com*
Other Authors Email: tomilola.moradeyo@gmail.com;308mobeni@gmail.com; akinwale.rototech@gmail.com;bimcrown2001@gmail.com

ABSTRACT: You only look once (YOLO) is a state-of-the-art, real-time object detection system. YOLO version 7 (YOLOv7) model is a variant of YOLO. The objective of this paper is to apply YOLOv7 to livestock image detection and segmentation tasks in cattle grazing behavior, monitor and intrusions. Data obtained revealed that YOLOv7 performs better in terms of speed and accuracy with a mAP of 0.95 than the baseline techniques.

DOI: https://dx.doi.org/10.4314/jasem.v27i5.10

Cite this paper as: MORADEYO, O. M; OLANIYAN, A.S; OJOAWO, A.O; OLAWALE, J.A; BELLO, R.W. (2023). YOLOv7 Applied to Livestock Image Detection and Segmentation Tasks in Cattle Grazing Behavior, Monitor and Intrusions. *J. Appl. Sci. Environ. Manage.* 27 (5) 953-958

Dates: Received: 17 February 2023; Revised: 08 April 2023; Accepted: 16 April 2023
Published: 31 May 2023

Keywords: You only look once; Cattle; Image segmentation; Livestock; YOLOv7

The agricultural research community, within the past decades has witnessed a great advancement in deep learning, which was made possible by machine learning and computer vision. Although the applications of deep learning in many different fields have been encouraging, such cannot be said in the agricultural domain because of some challenges. Attempts have been made by different researchers over the years to mitigate these challenges by devising breakthrough techniques (Bello and Abubakar, 2019; Bello and Moradeyo, 2019; Bello *et al.*, 2021; Bello *et al.*, 2021); this is in addition to the new research that is perpetually leading to the growth of more original and ground-breaking techniques. One such technique is You Only Look Once (YOLO), a deep learning model popular for its speed and accuracy in the community of agricultural imaging and object detection. YOLO is a neural network architecture developed purposely for object detection and image segmentation in real time with a good accuracy. YOLO training and implementation are based on Darknet, which is an open source architectural framework of neural network coded in C and CUDA. The installation and implementation are easy and fast, in addition to the CPU and GPU computation that it supports. In 2016, 2017, and 2018, YOLOv1 (unified, real-time object detection) (Redmon *et al.*, 2016), YOLOv2 (YOLO9000: better, faster, stronger) (Redmon and Farhadi, 2017), and YOLOv3 (an incremental improvement) (Redmon and Farhadi, 2018) were proposed and released, respectively. YOLOv3 was implemented on PyTorch. However, YOLOv4 (optimal speed and accuracy of object detection) (Bochkovskiy *et al.*, 2020) and YOLOv5 (Jocher, 2022) were released, respectively within a shortest time interval. The difference between YOLOv2 and YOLOv3 is that YOLOv2 has the ability to process images at 40-90 FPS (Frame per Second) while YOLOv3 allows easy tradeoff between speed and accuracy by changing the size of the model with

**Corresponding Author Email Address: sirbrw@yahoo.com*

no restriction. YOLOv4 was based on state-of-art BoF (bag of freebies) and several BoS (bag of specials). The BoF major role is to improve the detector accuracy without increasing the inference time, but the training cost. While the BoS increases the inference cost by a small amount, it substantially improve the object detection accuracy. YOLOv4 was implemented on Darknet and has obtained 43.5% AP (Average Precision) on the COCO dataset with 65 FPS on the Tesla V100. There is a difference between YOLOv5 and all other prior releases; it was implemented on PyTorch rather than Darknet. YOLOv5 has a CSP backbone and PA-NET neck as YOLOv4. Mosaic data augmentation and auto learning bounding box anchors are the major improvements of YOLOv5. In 2022, YOLOv6 (a single-stage object detection framework for industrial applications) was proposed and released by Li *et al*., (2022) According to the authors; YOLOv6 accomplishes the best trade-off in terms of accuracy and speed. In terms of speed and accuracy, YOLOv7 (Wang *et al.*, 2022) version E6 performed better than transformer-based detectors like SWINL Cascade-Mask R-CNNR-CNN. Moreover, YOLOv7 outperformed PP-YOLO, YOLOX, YOLOR, Scaled-YOLOv4, YOLOv5, DETR, Deformable DETR, DINO-5scale-R50, and Vit-Adapter-B. Among the numerous works that utilized YOLO algorithms for agricultural tasks are (Hatton-Jones *et al.*, 2021; Schütz *et al.*, 2021; Jintasuttisak *et al.*, 2022; Siriani *et al.*, 2022). Hence, the objective of this paper is to apply YOLOv7 to livestock image detection and segmentation tasks in cattle grazing behavior, monitor and intrusions.

MATERIALS AND METHODS

A sketchy diagram of acquisition platform of cattle images is presented in this section, additionally; the following items are also presented: dataset acquisition and enhancement, model architecture and training, and performance evaluation.

Data acquisition and enhancement: The input data employed for the segmentation experiment was acquired using cattle capturing system in a cattle ranch containing a group of Nigerian beef cattle, and other complicated background objects as depicted in Fig. 1. A high-resolution video camera that can capture and retain the video quality of every successive frame of cattle images was employed. To obtain a better image, the video camera was positioned on a very high pole object away from the centerline of the experimental system for the capturing of the cattle images. The image processing system was located in a site where the light and shadow diffusion could be reduced so that clear images with little or no noise could be obtained. The system was installed near a location in which the cattle pass or graze frequently each day. 80% of the 1050 acquired images were used as training data and 20% as testing data. The number of cattle, their breeds, and heights are among the information made available in Table 1.

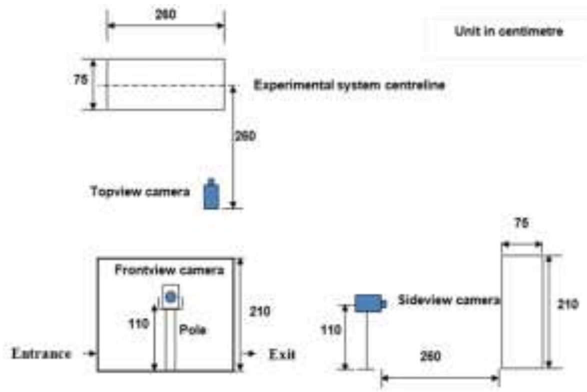

Fig. 1. A sketchy diagram of acquisition platform of cattle images (Bello *et al*., 2020)

Table 1. Information on the acquired input data

Information in the cattle ranch	Description
Number of cattle	50
Breeds of cattle	Keteku and Muturu
Body length of cattle (cm)	93.5-107
Body height of cattle (cm)	88-95.4

Enhancement of the acquired images: The cattle datasets were enhanced due to the circumstances that surrounded their acquisition and also, to make it easy for the image segmenter during segmentation. The following aspects are some of the reasons why we enhanced the acquired datasets: (1) frequent change of cattle posture; (2) similarity in body patterns among cattle of the same species and complexity in differentiating the segmented cattle from their background; (3) variance in lighting from time to time especially during image capturing. Because image is a product of the illumination and reflection images, this often result to huge disparity in brightness among image frames. So, by extracting components of the illumination, adjusting the adaptive illumination, and reconstructing the RGB images, using an image adaptive correction algorithm (Liu *et al*., 2016) that is based on 2-Dimensional gamma function to get rid of influence of shadow and illumination that is not in uniformity, the aforementioned issues were addressed for the overall improvement of the image quality. Fig. 2 displays the qualitative result of the cattle image enhancement. The experiment in this work was performed on the sampled and enhanced images.

Fig. 2. Comparison of raw and enhanced cattle images; the enhancement follows the image enhancement steps

Model Architecture and Training: Extended efficient layer aggregation networks primarily focus 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 analyses 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. Fig. 3 shows the extended efficient layer aggregation networks (E-ELAN). 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.

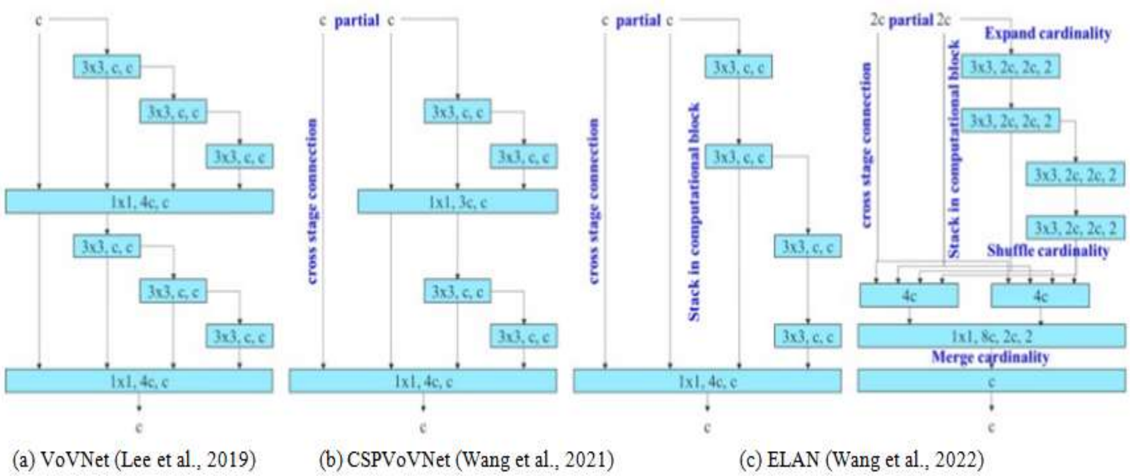

Fig. 3. Extended efficient layer aggregation networks (E-ELAN) (Wang *et al*., 2022)

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 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 study. We used Google Colab and GPU for the model training, and trained the model for 55 epochs with 16 batch size and 0.5 confidence thresholds. After completing the training, we used the generated weight for the evaluation and inference.

Performance Evaluation: The performance evaluation of the proposed approach is by using the average precision (AP) and mean average precision (mAP) being common tools for measuring and evaluating object detection and image segmentation tasks. The average precision (Equation (1)) and its mean (Equation (2)) are calculated as follows:

$$\mathrm{AP} = \sum_{n=1}^{N}[R(n) - R(n-1)] \cdot \max P(n) \qquad (1)$$

Where N is the calculated number of precision-recall (PR) points produced. P(n) and R(n) are the precision and recall with the lowest n-th recall, respectively.

$$\mathrm{mAP} = \frac{1}{n}\sum_{n=1}^{n} AP_i \qquad (2)$$

Where AP_i is the AP of class i, and n is the number of classes.

RESULTS OF EXPERIMENT

The experimental results of this work are presented and discussed in this section. Microsoft COCO dataset was employed to carry out the experiments and validation of the YOLOv7 model used in this study. YOLOv7, being a model that was trained from scratch disallowed the experiments conducted in this study from using any pre-trained models. Moreover, this study relies on the Train 2017 set and val 2017 set used in training and verifying the model during the development process with choosing hyper-parameters. Although there was little or no overlapping case among the cattle in the image which would have justified the capability of YOLOv7 as cattle instance segmentation approach to handle such situation, it still performs well on multi-cattle segmentation by detecting and segmenting each individual cattle from their complex background irrespective of the posture and body patterns or color of the cattle.

Ablation Experiments to Analyze YOLOv7: A number of ablations was conducted to analyze YOLOv7. Although it is not automatic scenario for all frameworks to benefit from deeper or advanced networks, YOLOv7 model built on the frameworks of the YOLO variants benefits from their deeper networks and advanced their designs. In view of the fact that YOLOv7 employs multiple pyramids for jointly prediction of object detection results, auxiliary head can be directly connected to the middle layer pyramid for training (Wang *et al*., 2022). This form of training arrangement can recover the lost information when predicting pyramid in the next level. For the above-mentioned reasons, partial auxiliary head was designed in the E-ELAN architecture. The visual results of the ablation experiments are shown in Fig. 4.

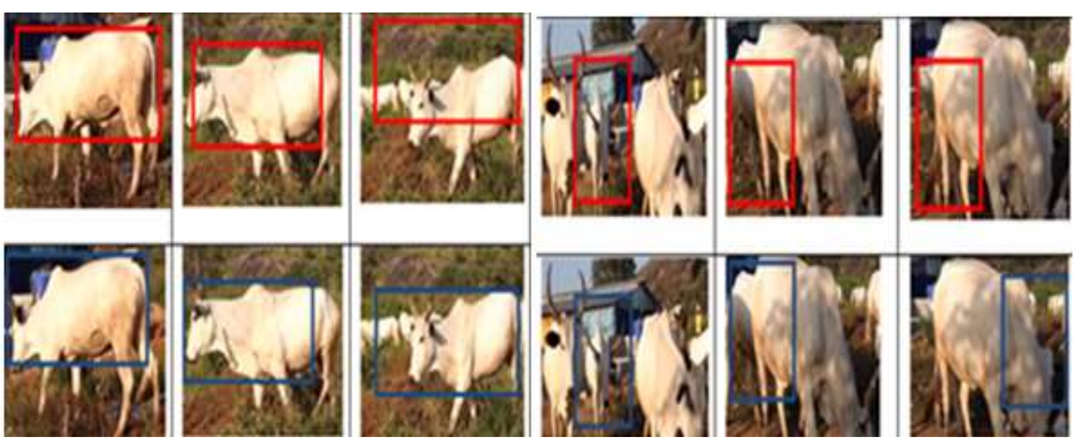

Fig. 4. Visual results of ablation experiments of YOLOv7 conducted on test images.

Mask R-CNN is the only baseline segmentation model compared with the YOLOv7 in this study for the cattle segmentation task, this is because Mask R-CNN is a state-of-the-art object detector different from the previous versions of YOLO that have similar characteristics as YOLOv7. Mask R-CNN uses detect-then-segment approach for its instance segmentation and is widely accepted for its ability to detect bounding boxes before segmenting the masks of instances present in each bounding box. For the experiment of Mask R-CNN in term of segmentation approach, Mask R-CNN is built on the model of Faster R-CNN to include RPN with alignment (for region proposals) and mask (for instance effect). The code employed for the Mask R-CNN experiment was acquired from the open source code. The learning rate of 0.003 was recorded for Mask R-CNN when trained on the datasets used for YOLOv7. Notable limitation of Mask R-CNN is its complete reliance on the accuracy of bounding box detection. YOLOv7 was applied in this study to address this limitation in agricultural applications. Presented in Table 2 are the comparison results of segmentation accuracies of Mask R-CNN and YOLOv7 and time taken for the process to be completed. Fig. 5 shows the visual result of masked YOLOv7 on cattle image segmentation.

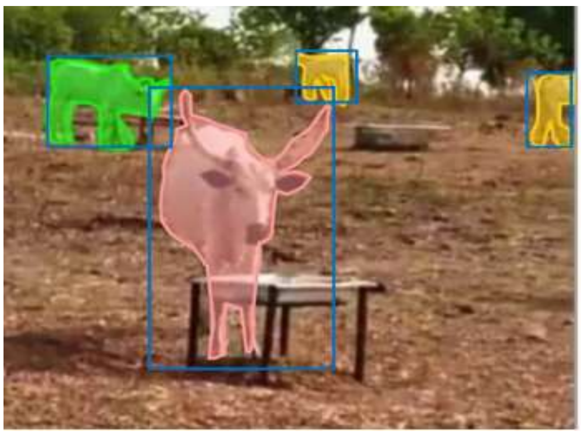

Fig. 5. Masked YOLOv7 applied to cattle image segmentation

Table 2. Cattle segmentation accuracies and time taken to complete processing

Methods	mAP	Time
Mask R-CNN	0.92	0.76
YOLOv7	0.95	0.68

Conclusion: Yolov7 is the new state-of-the-art object detection model applied to different agricultural domains. In this study, we applied it to detection and segmentation of cattle instances in an image. The model is important for the evaluation of cattle welfare in livestock management. We used Google Colab and GPU for the model training, and trained the model for 55 epochs with 16 batch size and 0.5 confidence thresholds. After completing the training, we used the generated weight for the evaluation and inference.

REFERENCES

Bello, RW; Abubakar, S. (2019). Development of a software package for cattle identification in Nigeria. *J. Appl. Sci. Environ. Manage.* 23(10), 1825-1828.

Bello, RW; Moradeyo, OM. (2019). Monitoring cattle grazing behavior and intrusion using global positioning system and virtual fencing. *Asian J. Math. Sci.* 3(4), 1-11.

Bello, RW; Mohamed, ASA; Talib, AZ. (2021). Enhanced Mask R-CNN for herd segmentation. *Inter. J. Agric. Biol. Engineer.* 14(4), 238-244.

Bello, RW; Mohamed, ASA; Talib, AZ; Olubummo, DA; Enuma, OC. (2021). Enhanced deep learning framework for cow image segmentation. IAENG *Inter. J. Comp. Sci.* 48(4), 1-10.

Bello, RW; Talib, AZ; Mohamed, ASA; Olubummo, DA; Otobo, FN. (2020). Image-based individual cow recognition using body patterns. *Image*. 11(3), 92-98.

Bochkovskiy, A; Wang, CY; Liao, HYM. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv*:2004.10934, 1-17.

Hatton-Jones, KM; Christie, C; Griffith, TA; Smith, AG; Naghipour, S; Robertson, K; du Toit, EF (2021). A YOLO based software for automated detection and analysis of rodent behaviour in the open field arena. Comp. Biol. Med. 134(104474), 1-9.

Jintasuttisak, T; Leonce, A; Sher Shah, M; Khafaga, T; Simkins, G; Edirisinghe, E. (2022). Deep learning based animal detection and tracking in drone video footage. In Proceedings of the 8th International Conference on Computing and Artificial Intelligence (pp. 425-431).

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.

Lee, Y; Hwang, JW; Lee, S; Bae, Y; Park, J. (2019). An energy and GPU-computation efficient backbone network for real-time object detection. *arXiv*:1904.09730v1, 1-11.

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 arXiv*:2209.02976v1, 1-17.

Liu, Z; Wang, D; Liu, Y; Liu, X (2016). Adaptive adjustment algorithm for non-uniform illumination images based on 2D gamma function. *Transact. Beijing Institute of Technol*. 36(2), 191-196.

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 (pp. 779-788).

Redmon, J; Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).

Redmon, J; Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 1-6.

Schütz, AK; Schöler, V; Krause, ET; Fischer, M; Müller, T., Freuling, CM; Lentz, HH (2021). Application of YOLOv4 for detection and Motion monitoring of red Foxes. *Animals*. 11(6), 1-18.

Siriani, ALR; Kodaira, V; Mehdizadeh, SA; de Alencar Nääs, I; de Moura, DJ; Pereira, DF (2022). Detection and tracking of chickens in low-light images using YOLO network and Kalman filter. *Neural Comp. Applicad*. 34(24), 21987-21997.

Wang, CY; Bochkovskiy, A; Liao, HYM (2021). Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/cvf conference on computer vision and pattern recognition (pp. 13029-13038).

Wang, CY; Bochkovskiy, A; Liao, HY M (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696, 1-15.