Enhanced Mask R-CNN for herd segmentation

Rotimi-Williams Bello[1,2], Ahmad Sufril Azlan Mohamed[1*], Abdullah Zawawi Talib[1]

(1. *School of Computer Sciences, Universiti Sains Malaysia, 11800, Pulau Pinang, Malaysia*;
2. *Department of Mathematics/Computer Sciences, University of Africa, Toru-Orua, Bayelsa State, Nigeria*)

Abstract: Livestock image segmentation is an important task in the field of vision and image processing. Since utilizing the concentration of forage in the grazing area with shielding the surrounding farm plants and crops is necessary for making effective cattle ranch arrangements, there is a need for a segmentation method that can handle multiple objects segmentation. Moreover, the indistinct boundaries and irregular shapes of cattle bodies discourage the application of the existing Mask Region-based Convolutional Neural Network (Mask R-CNN) which was primarily modeled for the segmentation of natural images. To address this, an enhanced Mask R-CNN model was proposed for multiple objects instance segmentation to support indistinct boundaries and irregular shapes of cattle bodies for precision livestock farming. The contributions of this method are in multiple folds: 1) optimal filter size smaller than a residual network for extracting smaller and composite features; 2) region proposals for utilizing multiscale semantic features; 3) Mask R-CNN's fully connected layer integrated with sub-network for an enhanced segmentation. The experiment conducted on pre-processed datasets produced a mean average precision (mAP) of 0.93, which was higher than the results from the existing state-of-the-art models.

Keywords: livestock farming, image segmentation, mask R-CNN, herd, enhancement
DOI: 10.25165/j.ijabe.20211404.6398

Citation: Bello R W, Mohamed A S A, Talib A Z. Enhanced Mask R-CNN for herd segmentation. Int J Agric & Biol Eng, 2021; 14(4): 238–244.

1 Introduction

Nigeria is a country with a national herd comprising 18.4 million cattle[1]. These herds are mostly managed by semi-sedentary and transhumant pastoralists. Nigeria practices three systems of cattle production, namely the pastoral system, the agro-pastoral system, and the commercial system. With the perpetual dependence of human beings on cattle and cattle by-products, there is a great need to continue providing a grazing environment for the cattle with supplementary feeds. This is mostly the method adopted by the agro-pastoral system having the grazing environment on demarcated rangelands. Proper monitoring of individual cows in such an arrangement can assist in the early detection of any abnormality and thereby preventing bad occurrences[2]. In recent years, different researchers have applied many and different state-of-the-art methods in monitoring the activities of cattle, namely radio-frequency identification method, biometrics identification method, sensor identification method, and computer vision identification method[3].

Among the methods mentioned above, computer vision occupies the frontline as a technology that deals with how computers can achieve advanced understanding from digital images or videos[4], in which image segmentation is one of the prerequisites. Computer vision also tries to find an easy way to comprehend and automate the tasks performed by the human visual system for the benefit of agricultural practice[5]. This practice involves getting access to information about individual cattle's health status and behavior, thereby making a substantial contribution to the management decision-making of livestock farming[6,7]. The contour extraction of an individual animal from the background and the image analysis that follows enables the monitoring of health and productivity-related variables such as body structures, body measurement, body condition score, live weight regression, and disease detection of the animals by the farmers throughout the life cycle of the animals[8-16]. The accurate extraction of the different features of the animal from the image greatly depends on the image segmentation efficiency.

But, the quality of the image segmentation can become a challenge due to both internal and external factors such as poor illumination and heterogeneous background[17]. To address the segmentation challenges iterated in this study, different contributions have been made in different studies using convolutional neural networks (CNN) based approaches with powerful abilities to learn the spatial-rich and semantic-information features[18-21]. In this study, an enhanced Mask Region-based-Convolutional Neural Network (Mask R-CNN) is proposed for herd segmentation so that an accurate segmentation can be accomplished in an environment full of complex backgrounds. The proposed model is in multiple folds, and the folds are presented in section 2 under model development.

Recently, He et al.[22] proposed an instance segmentation framework called Mask R-CNN for object detection. Also Li et al.[23] proposed an instance segmentation technique capable of learning implicit structure before any further improvement. Moreover, Mask SSD[24], MaskSplitter[17], DeepMask[25], and SharpMask[26] directly produced segmentation proposals of the object from the image pixels before classifying them. Bounding boxes are generated by most of the object detection methods for each target object that is detected with proper classification[27]. In the selective search method, the R-CNN method generates region

Received date: 2021-01-02 **Accepted date:** 2021-05-09
Biographies: Rotimi-Williams Bello, PhD candidate, research interest: computer vision and image processing with focus on object recognition from images and videos, Email: sirbrw@yahoo.com; **Abdullah Zawawi Talib**, PhD, Professor, research interest: computer graphics and visualization, geometric computing and scientific computing, Email: azht@usm.my.
***Corresponding author: Ahmad Sufril Azlan Mohamed**, PhD, Senior Lecturer, research interest: image processing, video tracking, facial recognition and medical imaging. School of Computer Sciences, Universiti Sains Malaysia, 11800, Pulau Pinang, Malaysia. Tel: +60-4-6536351, Email: sufril@usm.my.

proposals before object proposals classification by employing a deep CNN[28,29]. Nevertheless, it is not cost-effective extracting features of the proposal regions using R-CNN. The generation of region proposals is carried out in Faster R-CNN by a region proposal network (RPN) by taking as input an image and produces a set of several object proposals rectangular in shape that are used on feature maps by a sliding window to detect cow object. RPN is one of the two branches found in Faster R-CNN. The second branch of Faster R-CNN is the branch that is responsible for features extraction, bounding box prediction, and classification[27].

A fully convolutional network is a variant of CNN and a popular semantic segmentation model that can transform image pixels-to-pixel categories[30]. Ter-Sarkisov et al.[17] built a fully convolutional network to achieve a beef cattle segmentation model. Chen et al.[31] and Zhang et al.[24] have proposed some instance segmentation methods. Mask R-CNN, which this study enhances is an instance segmentation that identifies the instance of each object in an image using a mask representation with the simultaneous prediction of the object class and bounding box regression[32]. Instance segmentation usually involves three steps, namely 1) region proposals using RPN; 2) object class prediction; 3) object mask generation. Put together, the aforementioned achievements show the practicality of the approaches that are CNN-based for cattle segmentation even in complex environments.

This work is a contributory step in precision livestock farming towards the realization of real-time evaluation of cattle wellbeing.

2 Materials and methods

2.1 Datasets and pre-processing

The datasets used in the experiment of this study were in two categories, namely 1) the datasets acquired by authors (own dataset); 2) the cow dataset acquired from the Microsoft common objects in context (MS coco) datasets. The own dataset which was acquired from 10 cows (Keteku and Muturu breeds) using a camera contains 1000 images of the cows from which 800 images were used as a training dataset and the remaining 200 images were used for testing the model. In order to expand the own dataset, a data augmentation method was applied to the dataset. For the annotation, LabelMe[33] was used in labeling the dataset. The labeled images represented the ground-truth bounding box against which the predicted object bounding box regression was evaluated. On the other hand, the MS coco dataset[34], which contains 1986 cow training images, and 85 validating and testing images, were used together with the own dataset for the model training, fine-tuning, and testing. The pre-trained coco weights of Mask R-CNN were used in the training of the proposed model in the form of transfer learning as shown in Figure 1.

2.2 Model development

The proposed model was an enhanced Mask R-CNN comprising 1) optimal filter size smaller than a residual network for extracting smaller and composite features; 2) region proposals for utilizing multiscale semantic features; 3) Mask R-CNN's fully connected layer integrated with sub-network for enhanced segmentation. The proposed model was presented in three sections, namely 1) an abridged model of the residual network; 2) the enhanced Mask R-CNN structure; (3) the loss function, as shown in Figure 1. Figure 2a shows the existing model and Figure 2b shows the enhanced model. Table 1 shows the hyper-parameters for the model, and the different models in this study were obtained by changing the hyper-parameters.

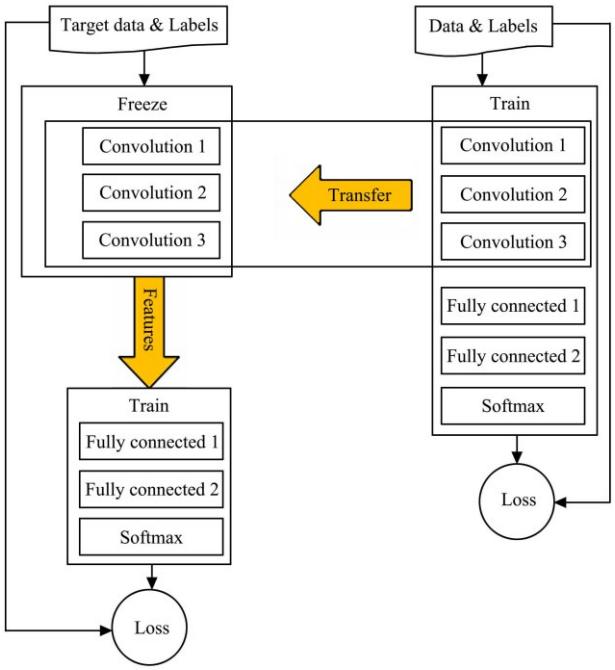
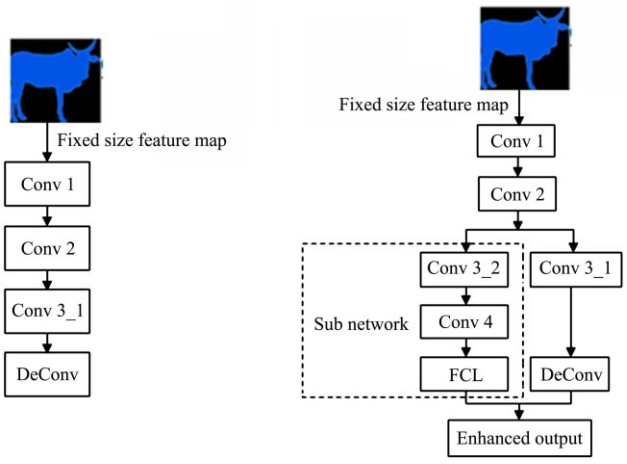


Figure 1 Transfer learning



a. Existing model b. Enhanced model

Figure 2 Comparison of the existing model and the enhanced model

Table 1 Model hyper-parameters

Specifications	Amount in number
Learning rate	0.001
Weight decay	0.0001
Momentum of learning	0.90
The dimension of the image (minimum)	512
The dimension of the image (maximum)	512
Detection confidence (minimum)	0.50
Number of batches	5
Size of batch	200
Epochs	5
Iterations per epoch	5
Steps per epoch	1 000
Validation steps	5
Mask shape	28×28
Number of anchor classes (cow and background)	2

2.2.1 An abridge model of residual network

In the backpropagation process, the skipping of activation layers in the residual network (ResNet) is the major reason for most

of the issues the network confronts. The absence of an equation that can describe the instability in the ResNet parameters is also an issue, thereby leading to the inaccuracy of the gradient equation.

In addition, the training process does not clarify the layer that has more training advantage over another. Therefore, by using the algorithm of the backpropagation, an abridged model of a residual network was proposed that is capable of solving the issues. By using the new gradient equation, new rules made of different parameters for ResNet were obtained whereby optimal filter size smaller than ResNet was provided for extracting smaller and composite features. By that means, there was a decrease in the number of parameters needed for the training, thereby leading to an increase in the computation efficiency. Figures 3-5 show the architecture of the residual network, the block of the residual network, and the enhanced architecture of the residual network respectively.

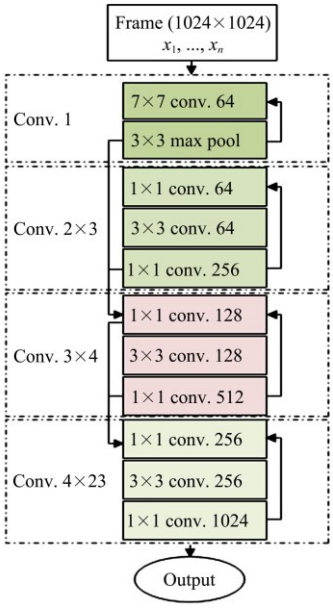


Figure 3 Architecture of a residual network

By using a deep network comprising a set of blocks that could solve the gradient vanishing issues[22], there was a great improvement in the performance of ResNet. Figure 4 shows the residual network block. Equation (1) presents the building formula of the two-layer block.

$$H(x)=F(x,\{W_i\})+x \quad (1)$$

where, x is the building block input; $H(x)$ is the building block output vectors, and $F(x,\{W_i\})$ is the learned residual mapping in the training process.

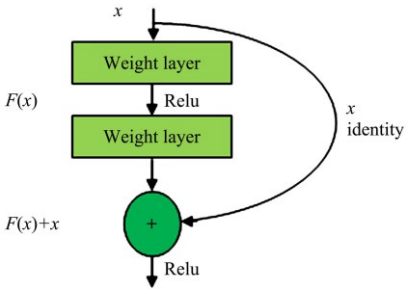


Figure 4 Block of a residual network

Based on Figure 5, training for convolutional layers with the best block was carried out as enumerated below: 1) 1 repetition for the 1st convolution block; 2) 4 repetitions for the 2nd convolution block; 3) 4 repetitions for the 3rd convolution block; 4) 14 repetitions for the 4th convolution block.

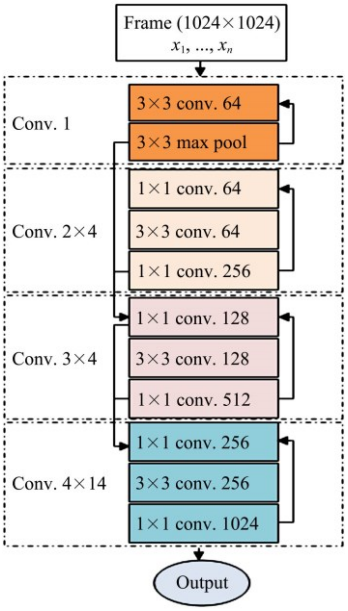


Figure 5 Enhanced architecture of a residual network

2.2.2 Enhanced Mask R-CNN structure

The enhanced Mask R-CNN structure comprises three separate branches, namely 1) the network backbone branch which is used for extracting features and generating region of interest (ROI) alignment. The branch comprises ResNet101+RPN+Feature Pyramid Network (FPN). In addition, the branch generates multi-scale feature maps before mapping each of its points to the input image so that matching ROI can be acquired; 2) the ROI alignment (ROIalign) branch which is used for pooling generated ROIs from the network backbone to feature maps (fixed in size) so that any form of quantization error can be overcome; 3) the mask generating branch. The fully connected layer (FCL) serves as a gateway through which all the feature maps (fixed in size) from the ROIalign pass through before generating the object mask, the bounding box regression, and the object classification. Figure 6 illustrates the operation of the above three modules on the cows.

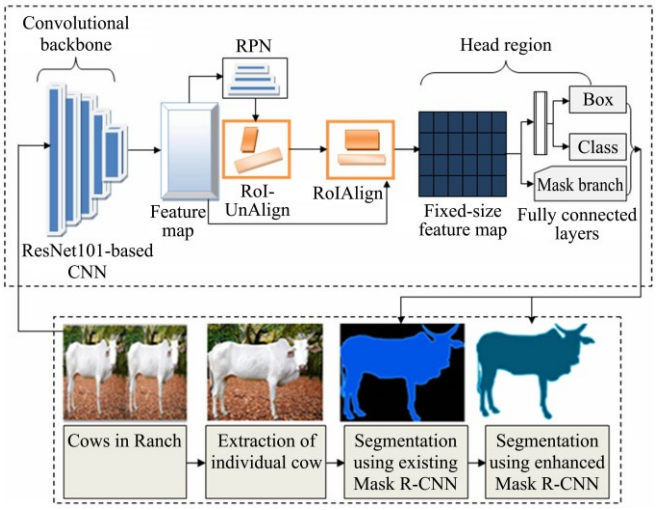


Figure 6 Framework of the proposed enhanced Mask R-CNN

2.3 Loss function

The generated masks are soft masks with their representation in float numbers making them hold more details than binary masks. While the ground-truth masks are scaled down to 28×28 pixels during training to enable the computation of loss, the predicted masks on the other hand are scaled up to the size of the bounding box's ROI during inferencing. This process leads to the

generation of final masks for each of the objects. In training the network, the loss function defines the difference between the predicted value and the ground-truth value. Furthermore, the position of the loss function in model training for cow instance segmentation is very essential as there is a need to calculate and reduce the error associated with the neural networks during the process of optimization. The value produced by the loss function as illustrated in Figure 1 is referred to as a loss, and the function can be referred to as a loss function, an error function, or a cost function when the function is being minimized. It is through the loss function that all aspects of the model are distilled down into a single number for improvement, thereby leading to a better model. The function to perform this task must be able to capture the problem and its attributes, and reliably represent the goal of the work. In order to achieve satisfactory results in the proposed framework, a combination of loss functions was applied in the training of bounding box regression, object class prediction, and mask branch segmentation.

Equation (2) represents the loss function that was used in accomplishing this task.

$$L=L_{ce}+L_{be}+L_{me} \tag{2}$$

where, L represents loss function; L_{ce} represents classification error; L_{be} represents bounding box regression error, and L_{me} represents mask error.

The following equation is used in measuring the segmentation accuracy,

$$\text{IOU}=\frac{A\cap B}{A\cup B} \tag{3}$$

where, IOU defines the extent of overlap between the predicted and ground-truth bounding boxes; A and B are the bounding boxes of the predicted objects and their ground truth respectively.

The IOU values from 0.50 to 0.95 with mAP@X notation are considered for this work, where X is the value of the threshold employed to compute the metric. Only after all the matches for the image are established can the precision-recall be computed. Precision is the total number of correct instances that the model produces, and it is computed as follows:

$$P=\frac{\text{True positive}}{\text{True positive}+\text{False positive}} \tag{4}$$

A recall measures the total positive instances that the model can produce, and it is computed as follows:

$$R=\frac{\text{True positive}}{\text{True positive}+\text{False negative}} \tag{5}$$

where, P is the precision; R is the recall; True Positive is an outcome where the model correctly predicts the positive instance; False Positive is an outcome where the model incorrectly predicts the positive instance; False Negative is an outcome where the model incorrectly predicts the negative instance. Average precision (AP) is calculated by taking the area under the precision-recall curve and by segmenting the recalls evenly to different parts. AP is calculated as follows:

$$\text{AP}=\sum_{n=1}^{N}[R(n)-R(n-1)]\cdot\max P(n) \tag{6}$$

where, N is the calculated number of PR points produced; a PR point is a point with a pair of x and y values in the PR space where x is recall and y is precision. $P(n)$ and $R(n)$ are the precision and recall with the lowest n-th recall, respectively.

The mean average precision (mAP) is calculated as follows:

$$\text{mAP}=\frac{1}{n}\sum_{n=1}^{n}AP_i \tag{7}$$

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

3 Implementation

A graphic processing unit (GPU), Tensorflow[35], Keras, and Opencv-python are some of the main required hardware and software packages installed on the system on which the proposed framework was implemented. Keras is a popular python deep learning application programming interface (API) that has the low-level flexibility for implementing arbitrary research ideas while voluntarily presenting high-level expediency features to speed up experimentation processes. TensorFlow on the other hand is an end-to-end open-source python deep learning application that serves as a platform for machine learning. TensorFlow possesses an all-inclusive and flexible network of tools, and libraries that help research push the state-of-the-art in machine learning, and developers effortlessly build and deploy machine learning-powered applications.

The effectiveness of Tensorflow in code optimization and in handling high-performing computation makes it suitable for the detection and segmentation task. The information about the hardware and software employed in performing this study is presented in Table 2.

Table 2 Requirements for the segmentation

Software	Type/Version
Operating system	64-bit Windows 10
IDE	Visual studio 2019
Python library	Keras
MATLAB	R2019b
Hardware	**Type/Version**
CPU	Intel Core i5 processor@2.4GHz
RAM	16 Gigabytes
Graphics card	GeForce GTX 1080 Ti
Hard-disk	2 Terabytes
Camera module	Vision Datum LEO 640H-200gc High-Speed 200fps Sharp RJ33 CCD Gigabit Ethernet 3d
Monitor	10.1 inch IPS HD Portable LCD Gaming Monitor PC display VGA HDMI interface for PS3/PS4/XBOx360/CCTV/Camera

4 Results and discussion

As shown in Figure 6, the proposed model is an extension of the existing Mask R-CNN with modification of the ResNet network and the segmentation mask branch as presented in Section 2. The results of the segmentation tasks as shown in Table 3 were obtained from the experiment performed on the acquired datasets with two different versions of the enhanced Mask R-CNN model (model 1 and model 2) built by changing the hyper-parameters. The hyper-parameters of model 1 are listed in Table 1, whereas the learning rate, weight decay, and momentum of learning of the second model were 0.01, 0.001, and 0.95 respectively. The precision, recall, average precision (AP), IOU, and mean average precision (mAP) were the main evaluation metrics used in evaluating the results of the experiment. Figure 7 shows the graph interpretation of Table 3 where the generated results of IOU, precision, recall, AP, and mAP were according to Equations (3)-(7), respectively. The three models were tested on both the own dataset and the MS coco cow dataset. The own dataset comprised the two classes of cow objects (muturu and keteku), and MS coco cow dataset comprised the standard cow dataset. As shown in Table 3 and Figure 7, at threshold value of 0.50, both model 1 and

model 2 produced an mAP of 0.93 which was higher than the 0.92 mAP produced by the existing model.

Moreover, the experiment of the enhanced Mask R-CNN on the MS coco cow dataset produced an mAP of 0.90, this implies that the own dataset performed better than the MS coco cow dataset.

Table 3 Results of the enhanced model and the existing model for own dataset and MS coco cow dataset at different values of IOU

Model	Cow object	Metric	Average Precision (AP) at different values of IOU									
		IOU	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
Enhanced Mask R-CNN (Model 1)	Muturu cow	AP	0.90	0.89	0.87	0.78	0.64	0.56	0.51	0.39	0	0
	Keteku cow	AP	0.96	0.71	0.81	0.80	0.67	0.54	0.50	0	0	0
		mAP	**0.93**	**0.80**	**0.84**	**0.79**	**0.66**	**0.55**	**0.51**	**0.20**	**0**	**0**
Enhanced Mask R-CNN (Model 2)	Muturu cow	AP	0.90	0.88	0.86	0.78	0.64	0.56	0.51	0	0	0
	Keteku cow	AP	0.96	0.82	0.80	0.71	0.67	0.54	0.50	0.37	0	0
		mAP	**0.93**	**0.85**	**0.83**	**0.75**	**0.66**	**0.55**	**0.51**	**0.19**	**0**	**0**
Existing Mask R-CNN (Model 3)	Keteku cow	AP	0.91	0.90	0.89	0.78	0.64	0.56	0.51	0	0	0
	Muturu cow	AP	0.92	0.90	0.86	0.84	0.64	0.56	0.52	0	0	0
		mAP	**0.92**	**0.90**	**0.88**	**0.81**	**0.64**	**0.56**	**0.52**	**0**	**0**	**0**
Enhanced Mask R-CNN	MS coco cow_1	AP	0.93	0.94	0.91	0.82	0.90	0.57	0.53	0	0	0
	MS coco cow_2	AP	0.87	0.92	0.90	0.83	0.89	0.59	0.51	0	0	0
	MS coco cow_3	AP	0.88	0.93	0.88	0.85	0.88	0.58	0.53	0	0	0
	MS coco cow_4	AP	0.93	0.90	0.89	0.84	0.92	0.49	0.52	0	0	0
	MS coco cow_5	AP	0.89	0.91	0.92	0.86	0.87	0.57	0.47	0	0	0
		mAP	**0.90**	**0.92**	**0.90**	**0.84**	**0.89**	**0.56**	**0.51**	**0**	**0**	**0**

Note: Model 1 and Model 2 are the enhanced models with different hyper-parameters. Model 3 is the existing Mask R-CNN model, and the hyper-parameters for it are found in reference[22], the same as below.

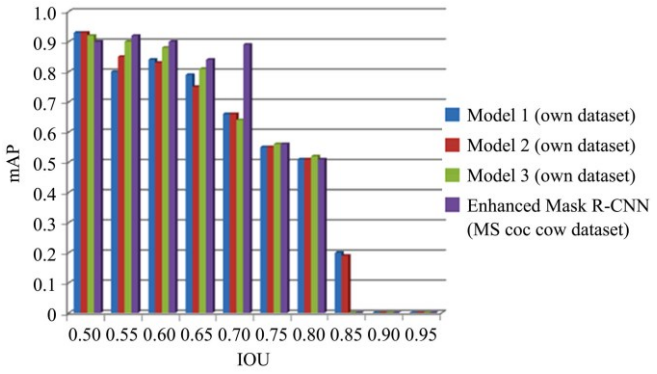


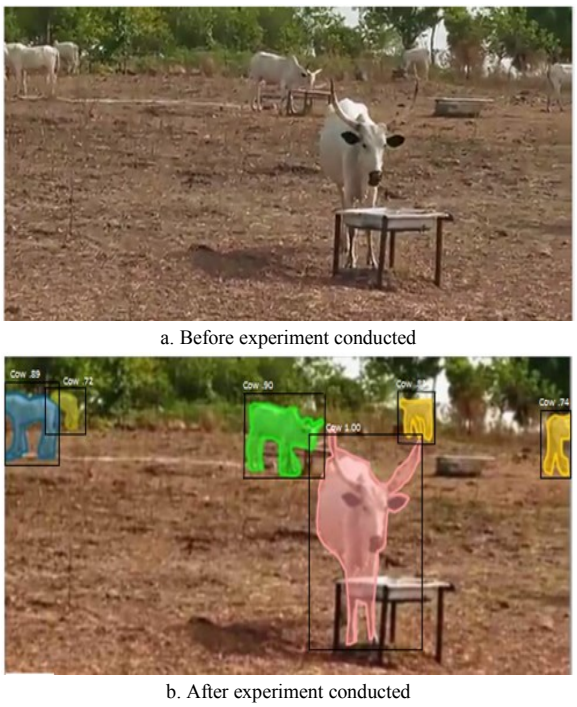
Figure 7 Graphical result of the enhanced model and the existing model for own dataset and MS coco cow dataset at different values of IOU

From the results obtained, the proposed model achieved an mAP of 0.93 compared to that of Mask R-CNN and other existing models (Table 4). Going by these results, the enhanced Mask R-CNN shows great segmentation accuracy over other state-of-the-art segmentation methods.

Table 4 Comparison of the enhanced model with the state-of-the-art models

Segmentation model	Backbone network	mAP
Mask R-CNN[22]	ResNet101	0.92
MaskSplitter[17]	VGG16	0.71
FCIS[36]	ResNet101-C5-dilated	0.56
Faster R-CNN[27]	ResNet101-FPN	0.90
YOLO v2[37]	DarkNet19	0.91
Mask Single Shot Detector (Mask SSD)[24]	ResNet101-FPN-B6	0.82
Multi-task Network Cascades (MNC)[38]	ResNet101-C4	0.42
DeepMask[25]	VGGNet[39]	0.53
SharpMask[26]	VGGNet[39]	0.82
Enhanced Mask R-CNN (Proposed)	ResNet101	**0.93**

Figure 8b shows the results of another experiment conducted on the own dataset involving six cows in a ranch (Figure 8a) using the same framework of Figure 6 but with more emphasis on utilizing the multiscale semantic features. By applying the multiscale semantic features, the enhanced Mask R-CNN was able to achieve multiple objects segmentation as shown in Figure 8b, each cow with its generated bounding box, mask, and confidence score. The computation efficiency between the existing Mask R-CNN and the enhanced Mask R-CNN on both the raw data and the enhanced data was measured in terms of their speed, and the results are presented in Table 5. Raw data are unprocessed data that are unfit for training a model because they are noisy, unreliable, and missing in values. So, in order to not produce misleading results, such data need to be enhanced to fit for training a model. With the results presented in Table 5, it can be easily concluded that the enhanced dataset performed better than the raw dataset in training the two models for herd segmentation.



a. Before experiment conducted

b. After experiment conducted

Figure 8 Cattles in the ranch before experiment conducted and after experiment conducted

Table 5 Computation efficiency of the existing model and the enhanced model for instance segmentation

Model	Data type	Time/s
Mask R-CNN	Raw	0.73
	Enhanced	0.72
Enhanced Mask R-CNN	Raw	0.72
	Enhanced	**0.70**

5 Conclusions

An enhanced Mask R-CNN was presented in this study that should support precision livestock farming for the segmentation of multiple cow objects in a typical agricultural environment. The framework of the proposed model was an extension of the existing Mask R-CNN model with three main enhancements, namely 1) optimal filter size smaller than a residual network for extracting smaller and composite features; 2) region proposals for utilizing multiscale semantic features; 3) Mask R-CNN's fully connected layer integrated with sub-network for an enhanced segmentation. The results showed a mAP of 0.93 was achieved by the proposed model with improved computation efficiency. The model demonstrated accurate simultaneous localization and mapping. Future work involves developing a separate-mask prediction model for segmenting overlapping regions and differentiating cattle body parts explicitly.

Acknowledgements

The authors received funding from the Division of Research and Innovation (RCMO), Universiti Sains Malaysia for the publication of this work.

[References]

[1] FAO. The future of livestock in Nigeria: Opportunities and challenges in the face of uncertainty. Food and Agriculture Organization of the United Nations Rome, 2019; 46p.

[2] Bello R, Talıb A, Mohamed A. Deep learning-based Architectures for recognition of cow using cow nose image pattern. Gazi University Journal of Science, 2020; 33(3): 831–844.

[3] Bello R W, Olubummo D A, Seiyaboh Z, Enuma O C, Talib A Z, Mohamed A S A. Cattle identification: the history of nose prints approach in brief. In IOP Conference Series: Earth and Environmental Science, IOP Publishing, 2020; 594(1): 1–9.

[4] Shao W, Kawakami R, Yoshihashi R, You S, Kawase H, Naemura T. Cattle detection and counting in UAV images based on convolutional neural networks. International Journal of Remote Sensing, 2020; 41(1): 31–52.

[5] Mao Y, He D, Song H. Automatic detection of ruminant cows' mouth area during rumination based on machine vision and video analysis technology. Int J Agric & Biol Eng, 2019; 12(1): 186–191.

[6] He D, Liu D, Zhao K. Review of perceiving animal information and behavior in precision livestock farming. Transaction of the CSAM, 2016; 47: 231–244. (in Chinese)

[7] Bos J M, Bovenkerk B, Feindt P H, Van Dam Y K. The quantified animal: Precision livestock farming and the ethical implications of objectification. Food Ethics, 2018; 2: 77–92.

[8] Xudong Z, Xi K, Ningning F, Gang L. Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. Computers and Electronics in Agriculture, 2020; 178: 1–11.

[9] Gomes R A, Monteiro G R, Assis G J F, Busato K C, Ladeira M M, Chizzotti M L. Estimating body weight and body composition of beef cattle trough digital image analysis. Journal of Animal Science, 2016; 94(12): 5414–5422.

[10] Bello R W, Abubakar S. Development of a software package for cattle identification in Nigeria. J. Appl. Sci. Environ. Manag., 2019; 23(10): 1825–1828.

[11] Hansen M F, Smith M L, Smith L N, Jabbar K A, Forbes D. Automated monitoring of dairy cow body condition, mobility and weight using a single 3d video capture device. Comput. Ind, 2018; 98: 14–22.

[12] Zhou C, Lin K, Xu D, Liu J, Zhang S, Sun C, et al. Method for segmentation of overlapping fish images in aquaculture. Int J Agric & Biol Eng, 2019; 12(6): 135–142.

[13] Xiao D, Feng A, Liu J. Detection and tracking of pigs in natural environments based on video analysis. Int J Agric & Biol Eng, 2019; 12(4): 116–126.

[14] Tebug S F, Missohou A, Sourokou S S, Juga J, Poole E J, Tapio M, et al. Using body measurements to estimate live weight of dairy cattle in low-input systems in Senegal. J. Appl. Anim. Res, 2018; 46: 87–93.

[15] Chen F E, Liang X M, Chen L H, Liu B Y, Lan Y B. Novel method for real-time detection and tracking of pig body and its different parts. Int J Agric & Biol Eng, 2020; 13(6): 144-149.

[16] Liu H, Reibman A R, Boerman J P. A cow structural model for video analytics of cow health. 2020; arXiv pre-print. arXiv:2003.05903, 2020; 1–13.

[17] Ter-Sarkisov A, Ross R, Kelleher J, Earley B, Keane M. Beef cattle instance segmentation using fully convolutional neural network. 2018; arXiv pre-print. arXiv:1807.01972, 2018; 1–11.

[18] Lyu S, Noguchi N, Ospina R, Kishima Y. Development of phenotyping system using low altitude UAV imagery and deep learning. Int J Agric & Biol Eng, 2021; 14(1): 207–215.

[19] Bello R W, Talib A Z, Mohamed A S A, Olubummo D A, Otobo F N. Image-based individual cow recognition using body patterns. International Journal of Advanced Computer Science and Applications, 2020; 11(3): 92–98.

[20] Salau J, Krieter J. Instance segmentation with Mask R-CNN applied to loose-housed dairy cows in a multi-camera setting. Animals, 2020; 10(12): 1–19.

[21] Bello R W, Talib A Z H, Mohamed A S A B. Deep belief network approach for recognition of cow using cow nose image pattern. Walailak Journal of Science and Technology, 2021; 18(5): 1–14.

[22] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, Venice-Italy, 2017; pp.2961–2969. doi: 10.1109/TPAMI.2018.2844175.

[23] Li K, Hariharan B, Malik J. Iterative instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016; pp.3659–3667. doi: 10.1109/CVPR.2016.398.

[24] Zhang H, Tian Y, Wang K, Zhang W, Wang F Y. Mask SSD: An effective single-stage approach to object instance segmentation. IEEE Transactions on Image Processing, 2019; 29(1): 2078–2093.

[25] Pinheiro P O, Collobert R, Dollár P. Learning to segment object candidates. 2015; arXiv pre-print. arXiv:1506.06204.

[26] Pinheiro P O, Lin T Y, Collobert R, Dollár P. Learning to refine object segments. 2016; arXiv pre-print. arXiv:1603.08695v2, 2016; 1–18.

[27] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 2016; 39(6): 1137–1149

[28] Girshick R, Donahue J, Darrell T, Malik J. Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell, 2015; 38: 142–158.

[29] Girshick R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago-Chile, 2015; 1440–1448.

[30] Bello R W, Mohamed A S A, Talib A Z. Contour extraction of individual cattle from an image using enhanced Mask R-CNN instance segmentation method. IEEE Access, 2021; 9: 56984–57000.

[31] Chen L C, Hermans A, Papandreou G, Schroff F, Wang P, Adam H. Masklab: Instance segmentation by refining object detection with semantic and direction features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018; pp.4013–4022. doi: 10.1109/CVPR.2018.00422.

[32] Zhao K, Kang J, Jung J, Sohn G. Building extraction from satellite images using mask R-CNN with building boundary regularization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, USA, 2018; pp.247–251. doi: 10.1109/CVPR.2018.00422.

[33] Russell B C, Torralba A, Murphy K P, Freeman W T. Labelme: A database and web-based tool for image annotation. Int. J. Comput. Vision, 2008; 77: 157–173.

[34] Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer, 2014; pp.740–755. doi: 10.1007/978-3-319-10602-1_48.

[35] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. In 12th Symposium on Operating Systems Design and Implementation, 2016; pp.265–283.

[36] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39(4): 640–651.

[37] Redmon J, Farhadi A. YOLO9000: better, faster, stronger. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017; pp.6517– 6525. doi: 10.1109/CVPR.2017.690.

[38] Dai J, He K, Sun J. Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016; pp.3150–315. doi: 10.1109/CVPR.2016.343.

[39] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2015; arXiv pre-print. arXiv:1409.1556v6.