



Cattle Segmentation and Contour Detection Based on Solo for Precision Livestock Husbandry

*¹BELLO, RW; ¹IKEREMO, ES; ¹OTOBO, FN; ²OLUBUMMO, DA; ³ENUMA, OC

¹Department of Mathematics and Computer Science, University of Africa, 561101 Sagbama, Bayelsa, Nigeria

²Department of Computer and Information Systems, Robert Morris University, Moon-Township, Pennsylvania, United States of America

³Department of Planning, Research and Statistics, Ministry of Health, Bayelsa State, Nigeria

*Corresponding Author Email: sirbrw@yahoo.com

Co-Authors Email: samplefiye@gmail.com; otobofirstman@gmail.com; adanielolu@yahoo.co.uk; charlieyogo@yahoo.com

ABSTRACT: Segmenting objects such as herd of cattle in natural and cluttered images is among the herculean dense prediction tasks of computer vision application to agriculture. To achieve the segmentation goal, we based the segmentation on the model of single objects by locations (SOLO) which is capable of exploiting the contextual cues and segmenting individual cattle by their locations and sizes. For its simple approach to instance segmentation with the use of instance categories, SOLO outperforms Mask R-CNN which uses detect-then-segment approach to predict a mask for each instance of cattle. The model is trained using synchronized stochastic gradient descent (SGD) over GPU to achieve a mAP of 0.94 making it 0.02 higher than the result recorded by the Mask R-CNN model. By using the focal loss, the proposed approach achieved 32.23 ADE on cattle contour detection making its performance better than the Mask R-CNN's performance.

DOI: <https://dx.doi.org/10.4314/jasem.v26i10.15>

Open Access Policy: All articles published by **JASEM** are open access articles under **PKP** powered by **AJOL**. The articles are made immediately available worldwide after publication. No special permission is required to reuse all or part of the article published by **JASEM**, including plates, figures and tables.

Copyright Policy: © 2022 by the Authors. This article is an open access article distributed under the terms and conditions of the **Creative Commons Attribution 4.0 International (CC-BY- 4.0)** license. Any part of the article may be reused without permission provided that the original article is clearly cited.

Cite this paper as: BELLO, RW; IKEREMO, ES; OTOBO, FN; OLUBUMMO, DA; ENUMA, OC (2022). Cattle Segmentation and Contour Detection Based on Solo for Precision Livestock Husbandry. *J. Appl. Sci. Environ. Manage.* 26 (10) 1713-1720

Keywords: Agriculture; Cattle; Instance segmentation; Livestock husbandry

Before any performance accuracy can be recorded in object detection task, there are many important factors to consider, among which is image segmentation method used. Image segmentation process involves regions-by-regions partitioning of an image based on some measures whereby disjointed but informative regions are achieved. This process typically facilitates the detection of objects' locations and boundaries in images. However, image segmentation is a challenging and herculean task of computer vision that is negatively influenced by many aspects such as occlusion conditions, poor illumination, low contrast, noise, and irregular edges and boundaries of the object (Guo and Ashour, 2019; Bello *et al.*, 2021a; Bello *et al.*, 2021b). Moreover, the existing database of cattle images are inadequate for the task involve in learning to detect categories of cattle species in the object detection problem in their thousands in cluttered images as investigated in this paper. Hence the objective of this paper is to overcome the

aforementioned problems by using the model of single objects by locations (SOLO) (Wang *et al.*, 2020) which is capable of exploiting the contextual cues and segmenting individual cattle by their locations and sizes. The top-down approach to instance segmentation used in Mask R-CNN (He *et al.*, 2017) is a detect-then-segment approach, and popular for its ability to detect bounding boxes before segmenting the masks of instances present in each bounding box, however, being a step-wise approach, Mask R-CNN has the shortcoming of completely relying on the accuracy of bounding box detection. On the other hand, the affinity relation is learned in bottom-up approach and each pixel is assigned an embedding vector by repelling the pixels that are not of the same instances and attract those pixels that are of the same instances, making grouping post-processing a necessity for the separation of instances. In contrast to what is obtained from the aforementioned instance segmentation approaches, in SOLO, the masks of each instance is directly

*Corresponding Author Email: sirbrw@yahoo.com

segmented by using the annotations of a complete instance mask rather than the mask of instance present in each bounding box or the learning of affinity relations, thereby quantizing locations of the center and the sizes of the object for enablement of objects segmentation by locations. This is one of the reasons we have preferred SOLO to other mainstream instance segmentation models such as Mask R-CNN for the cattle instance segmentation problem. The framework of SOLO enables the optimization of the neural network in a fashion that is end-to-end in such a way that notable limitations associated with other instance segmenters are completely removed using exclusive mask annotations for performing pixel-level instance segmentation without the need for detecting local box and grouping of pixel. The approach used in SOLO for instance segmentation enables it to achieve on par results with the mainstream methods on COCO dataset (Lin et al., 2014) which comprises different objects of different sizes and classes including cattle, our experiment data. Another generality of SOLO is in the task of detecting the contour of each instance by considering edge contours of the instance as a one-hot binary mask out of the restrictions of modification. Proposed in this paper is a method for achieving on a par cattle segmentation and contour detection accuracy with mainstream segmentation model in terms of performance in complex background surroundings. The proposed approach consists of the following steps: (1) reformulating the cattle instance segmentation as: (a) prediction of category and (b) mask generation tasks for each cattle instance, (2) dividing the input cattle image into a uniform grids, i.e., $G \times G$ grid cells in such a way that a grid cell can predict the category of the semantic and masks of the cattle instances

provided the center of the cattle object falls into the grid cell, and (3) conducting cattle segmentation and contour detection experiments. The work in this paper is a step towards cost-effective measures (Song et al., 2018; Bello et al. 2021c; Bello et al. 2021d; Bello *et al.* 2021e) to sustain a precision livestock husbandry.

MATERIALS AND METHODS

This section presents the following: a sketchy diagram of acquisition platform of cattle images, information on dataset employed for the experiment and the method for image enhancement, overview of the proposed approach, and instance segmentation by SOLO.

Data acquisition: 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 input data was 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.

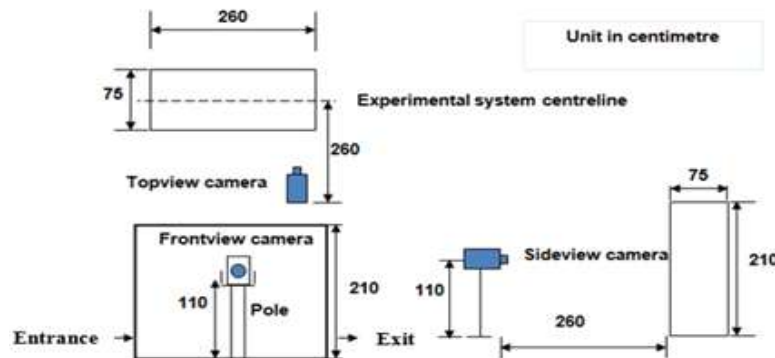


Fig. 1. A sketchy diagram of acquisition platform of cattle images

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

Overview of the proposed approach: The SOLO-based approach is proposed for realization of cattle instance segmentation in a group of Nigerian beef cattle, and other complicated background objects. The proposed approach takes the central approach of SOLO framework by reformulating the cattle instance segmentation as two concurrent “category-aware prediction” problems. In actual fact, the input cattle image was divided into uniform grids, i.e., $G \times G$ by the SOLO model in such a way that a grid cell can predict the category of the semantic and masks of the cattle instances provided the center of the cattle object falls into the grid cell and finally, the cattle instances segmented and their contour detected.

Semantic category and instance mask generation: For each grid, C -dimensional output (where C is the number of classes) is predicted by the proposed SOLO to show the probabilities of the semantic class which are conditioned on cells of the grid. By dividing the input cattle image into $G \times G$ grids, the output space becomes $G \times G \times C$. This approach is based on the premise that each $G \times G$ grid cell must contain one individual cattle instance, hence only containing one semantic category. During inference, the C -dimensional output shows the probability of the class for each object (cattle) instance. Each corresponding cattle instance mask is generated by each positive cell of grid in parallel with the prediction of semantic category. The fully convolutional networks (FCNs) (Long et al., 2015) are adopted as a direct approach to predict the cattle instance mask, though, their operations, to some extent, are spatially invariant making them more suitable for image classification for robust result. Therefore, SOLO, being spatially invariant, is considered as a solution for the segmentation task since the segmentation masks must be conditioned on the cells of the grid and separated by different feature channels to achieve the desired result. For simple solution to the cattle instance segmentation task using

SOLO, normalized pixel coordinates are directly fed to the networks at the initial stage of the network, and this is inspired by ‘CoordConv’ operator (Liu et al., 2018) for its simplicity and easy implementation. The spatial functionality is added to the FCN model by allowing the convolution access to its own input coordinates. Finally, cattle instance segmentation is formed based on the knowledge of naturally associating semantic category prediction and the corresponding instance mask by their reference cell of the grid. The results obtained from this segmentation could be beneficial to precision livestock husbandry for obtaining parameter information about each cattle such as size, height, width, and length.

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 acquired datasets need enhancement: (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

Cattle instance segmentation by SOLO: Instance segmentation task involves distinguishing between

individual cattle and their background in an image by segmentation procedure. It is on this premise that we

conducted our image segmentation experiment on individual cattle image in order to segment the image from the background. The proposed SOLO-based cattle segmentation was conducted by: (1) predicting the semantic category and mask generation tasks for each cattle instance, (2) dividing the input cattle image into a uniform grids, i.e., $G \times G$ grid cells in such a way

that a grid cell can predict the category of the semantic and masks of the cattle instances provided the center of the cattle object falls into the grid cell, and (3) performing cattle segmentation and contour detection experiments. Framework of SOLO-based cattle instance segmentation is illustrated in Fig. 3.

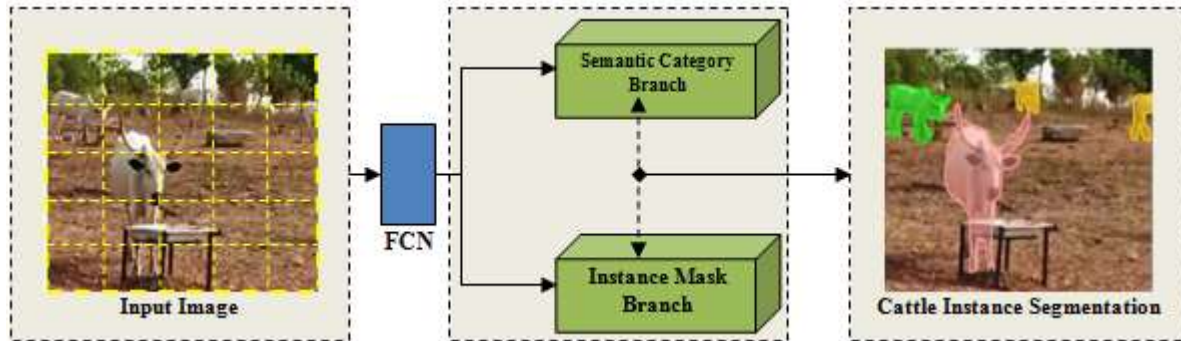


Fig. 3. Framework of SOLO-based cattle instance segmentation

SOLO network architecture: The construction of SOLO is such that its network is attached to the backbone of convolutional neural network. Feature pyramid network (FPN) (Lin et al., 2017) is used for generating feature maps pyramid of different sizes with channels of a fixed number for each level. These maps serve as input for semantic category and instance mask as the two prediction heads, with the head weights shared across different levels, excluding only the last conv from being shared as shown in Fig. 4. SOLO architecture is instantiated with multiple architectures for its generalization and effectualness.

At different pyramids, there may be variance in the number of grid, and for each cattle image, the differences consist of: (a) the backbone architecture employed for extracting convolutional features, (b) the network head for computing the results of the cattle instance segmentation, and (c) using training loss function for optimizing the model. The head architecture handles most of the experiments (cattle instance segmentation) with utilization of different variants for more generalization.

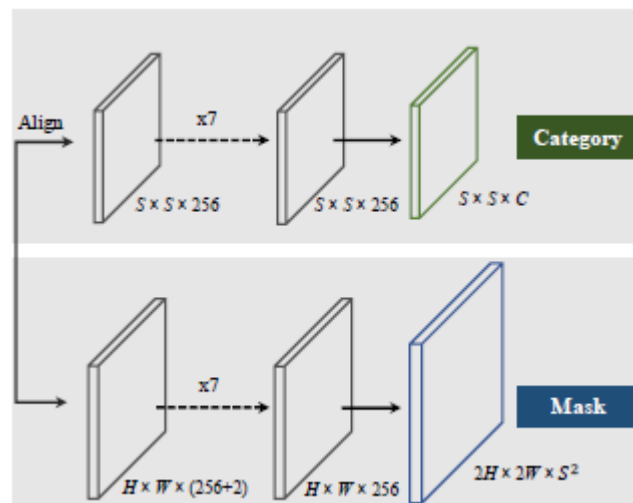


Fig. 4. Architecture of SOLO head for instance segmentation (Wang et al., 2020). Each of the attached two sibling sub-networks at feature level of FPN is responsible for predicting instance category (top) and segmenting instance mask (bottom). This architecture is applied to cattle instance segmentation as illustrated in Fig. 3

Cattle instance segmentation and SOLO behavior: The generated network outputs for the cattle instance

segmentation is shown by $G = 5$ grids as illustrated in Fig. 3. In the figure, the right column is the cattle

instance segmentation generated result from the networks, the left column is the input cattle image to the networks with $G = 5$ grids, and the middle columns are the branches for predicting category and activating cattle instance masks. At each grid, only one instance is allowed for activation, however, such instance prediction may be allowed by more than two mask channels in adjacent positions. At different positions, instances are explicitly segmented so that instance (individual cattle) segmentation problem can be converted by SOLO into a classification task that is position-cognizant. Non-maximum suppression (NMS) is used during inference to suppress the masks that are redundant.

SOLO learning via label assignment and loss function: Concerning the branch that is responsible for predicting the instance category, there is a need for probability to be given by the network to the object (cattle) category for each of $G \times G$ grid. Explicitly, grid (i, j) which its responsibility is to indicate the generated results of mask prediction by the corresponding mask channel in the activation map is regarded as a positive sample provided the mask of ground truth has it in its center region, or else, it is regarded as a negative sample. Because there is effectualness in sampling the center region as contain in literatures on object (cattle) detection (Tian et al., 2019; Kong et al., 2020), a similar technique is utilized for classifying mask category. The constant scale factors $\epsilon: (c_x, c_y, \epsilon w, \epsilon h)$ controls the center region, if the mass center (c_x, c_y) , with height h , and width w of the ground truth mask are given. ϵ is set to 0.2 and there are 3 positive samples on average for each mask of the ground truth. There is a binary segmentation mask for each positive sample in addition to the label for instance category; and the binary mask of the corresponding target will be annotated for each positive sample. For each image (cattle image), there is S^2 output masks resulting from S^2 grids. "The loss function for the training is defined as

$$L = L_{cate} + \lambda L_{mask} \quad (1)$$

Where L_{cate} is the conventional focal loss (Lin et al., 2017) for semantic category classification. L_{mask} is the loss for mask prediction, and it is expressed as

$$L_{mask} = \frac{1}{N_{pos}} \sum_k 1_{(p_{ij}^* > 0)} d_{mask}(\mathbf{m}_k, \mathbf{m}_k^*) \quad (2)$$

Here, indices $i = [K / S]$, $j = k \bmod S$, if we index the grid cells (instance category labels) from left to right and top to down. N_{pos} denotes the number of positive samples, \mathbf{p}^* and \mathbf{m}^* represent category and mask

target respectively. 1 is the indicator function, being 1 if $\mathbf{p}_{ij}^* > 0$ and 0 otherwise.

We have compared different implemetations of d_{mask} (...): binary cross entropy (BCE), focal loss and dice loss (Milletari et al., 2016). Finally, we employ dice loss for its effectiveness and stability in training. λ in Equation (1) is set to 3. The dice loss is defined as

$$L_{Dice} = 1 - D(\mathbf{p}, \mathbf{q}) \quad (3)$$

Where D is the dice coefficient which is defined as

$$D(\mathbf{p}, \mathbf{q}) = \frac{2 \sum_{x,y} (p_{x,y} \cdot q_{x,y})}{\sum_{x,y} p_{x,y}^2 + \sum_{x,y} q_{x,y}^2} \quad (4)$$

Here, $\mathbf{p}_{x,y}$ and $\mathbf{q}_{x,y}$ refer to the value of pixel located at (x, y) in predicted soft mask \mathbf{p} and ground truth mask \mathbf{q} ".

Cattle instance contour detection: Cattle instance contour detection helps in obtaining vital information regarding body parameters of individual cattle (Gomes et al., 2016). The proposed model is extended to cattle instance contour detection using the segmented cattle images. This is achieved by firstly converting the masks of ground truth in MS COCO dataset and in our cattle dataset (manually labeled using LabelMe tool (Russell et al., 2008) to get the ground truth) into instance contours by using findContours function of OpenCV, and the mask branch is optimized using the binary contours in parallel with the semantic category branch. While the focal loss is used for optimizing the contour detection, the instance segmentation baseline and other settings are the same.

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 (5)) and its mean (Equation (6)) are calculated as follows

$$AP = \sum_{n=1}^N [R(n) - R(n-1)] \cdot \max P(n) \quad (5)$$

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.

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (6)$$

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

Experimental set up: This section presents the experimental setup for performing and achieving the objective of the work. Also, the segmentation approach and results generated are compared with the mainstream methods.

Table 2. The software and hardware specifications

Software	Type/Version
Operating system	64-bit Windows 10
IDE	Jupyter
Python library	OpenCV
	MATLABR2021b
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

The cattle segmentation model is trained using synchronized stochastic gradient descent (SGD) over GPU with a total of 1050 images sampled to equal pixels; 850 of the images were used as the training dataset and the remaining 200 images formed the testing dataset with equal number per mini batch. For the ground truth, LabelMe tool is employed for manual labeling of the cattle parts. Table 2 illustrates the detail information about the software and hardware specifications in this experiment. The proposed

SOLO-based segmentation models are trained for reasonable number of epochs with 0.001 learning rate, 0.0001 weight decay and 0.9 momentum on the raw image datasets. To speed up the training process, the models are initialized from COCO pre-trained weights.

Comparison of the proposed method and mainstream method: Mask R-CNN is the only mainstream segmentation method compared with the proposed SOLO-based cattle segmentation method, reason being that their architectural pipelines are fairly similar as illustrated in Fig. 5. 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 SOLO. Notable limitation of Mask R-CNN is its complete reliance on the accuracy of bounding box detection. SOLO, in order to address this limitation directly segmented the masks of each instance by using the annotations of a complete instance mask rather than the mask of instance present in each bounding box or the learning of affinity relations, thereby quantizing locations of the center and the sizes of the object for enablement of objects segmentation by locations.

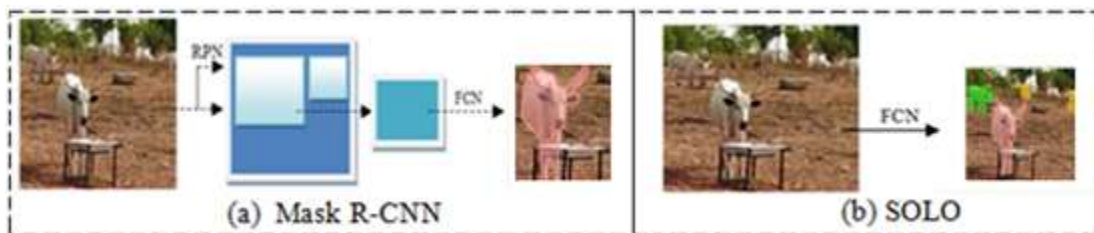


Fig. 5. Comparison of the pipelines of (a) Mask R-CNN and (b) SOLO

RESULTS AND DISCUSSION

The experimental results of this work are in two forms; the SOLO-based cattle instance segmentation results and the SOLO-based cattle instance contour detection result. Both are presented and discussed in this section.

SOLO-based cattle instance segmentation results: The qualitative results of the SOLO-based cattle instance segmentation experiment performed on the cattle

image datasets are as earlier presented in Fig. 3 and compared with Mask R-CNN in Fig. 5. Although there was little or no overlapping case among the cattle in the image which would have justified the capability of the proposed SOLO-based 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. Presented in Table 3 are the comparison results of segmentation accuracies and time taken for the process to be completed.

Table 3. Cattle segmentation accuracies and time taken to complete processing

Methods	mAP	Time
Mask R-CNN	0.92	0.76
SOLO	0.94	0.69

SOLO-based cattle contour detection results: Segmentation results of SOLO-based approach played huge role in detecting the contour lining of individual cattle. This was manifested in the qualitative results generated from the SOLO-based cattle contour

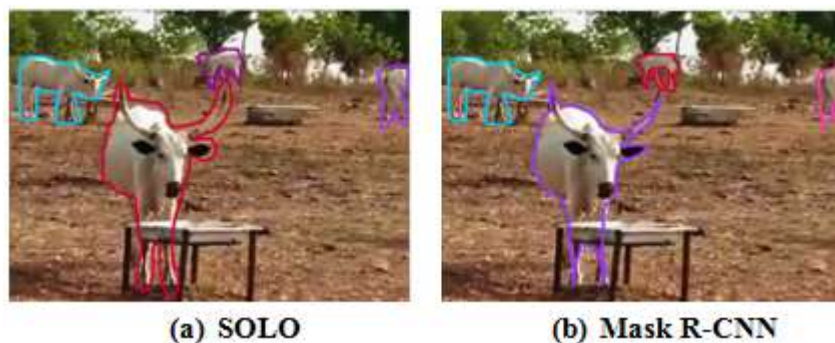


Fig. 6. Comparison of the contour detection performance of (a) SOLO and (b) Mask R-CNN

Table 4. Cattle contour detection and time taken to complete processing

Methods	ADE	Time
Mask R-CNN	34.37	0.76
SOLO	32.23	0.69

Conclusion: Cattle segmentation and contour detection based on SOLO for precision livestock husbandry has been proposed in this paper. The proposed model is important for the evaluation of cattle welfare in precision animal management. The cattle segmentation model is trained using synchronized stochastic gradient descent (SGD) over GPU to achieve a mAP of 0.94 making it 0.02 higher than the result recorded by the Mask R-CNN model. By using the focal loss, the proposed approach achieved 32.23 ADE on cattle contour detection making its performance better than the Mask R-CNN's performance. Our future research is on poultry instance segmentation.

REFERENCES

Bello, RW; Mohamed, ASA; Talib, AZ. (2021a). Contour extraction of individual cattle from an image using enhanced Mask R-CNN instance segmentation method. *IEEE Access*, 9, 56984-57000

detection experiment performed on the cattle image datasets. Fig. 6 illustrates example of cattle contour detection obtained by the proposed model and Mask R-CNN. The overcoming of the little overlapping situation in the image datasets was by the arrangement of firstly converting the masks of ground truth in MS COCO dataset and in our cattle dataset into instance contours and by optimizing the mask branch in parallel with the semantic category branch. The instance segmentation baseline and other settings are the same while the focal loss is used for optimizing the contour detection. Presented in Table 4 are the comparison results of contour detection and time taken for the process to be completed.

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

Bello, RW; Mohamed, ASA; Talib, AZ. (2021c). Cow Image Segmentation Using Mask R-CNN Integrated With Grabcut. In: *Proceedings of International Conference on Emerging Technologies and Intelligent Systems, Lecture Notes in Networks and Systems*, 322, 23-32. Springer, Cham.

Bello, RW; Mohamed, ASA; Talib, AZ; Olubummo, DA; Enuma, OC. (2021d). Enhanced Deep Learning Framework for Segmenting Cow Images. *IAENG Inter. J. Computer Sci.* 48(4), 1182-1191

Bello, RW; Mohamed, ASA; Talib, AZ; Olubummo, DA; Enuma, OC. (2021e). Computer Vision-based Techniques for Cow Object Recognition. In *IOP Conference Series: Earth and Environmental Science*, 7th International Conference on Agricultural and Biological Sciences, 858(012008), 1-9.

Gomes, RA; Monteiro, GR; Assis, GJF; Busato, KC; Ladeira, M. M; Chizzotti, ML. (2016). Estimating

- body weight and body composition of beef cattle trough digital image analysis. *Journal of Animal Science*, 94(12), 5414-5422
- Guo, Y; Ashour, AS. (Eds.). (2019). *Neutrosophic Set in Medical Image Analysis*. Academic Press
- He, K; Gkioxari, G; Dollár, P; Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969)
- Kong, T; Sun, F; Liu, H; Jiang, Y; Li, L; Shi, J. (2020). Foveabox: Beyond anchor-based object detection. *IEEE Transactions on Image Processing*, 29, 7389-7398
- Lin, TY; Dollár, P; Girshick, R; He, K; Hariharan, B; Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125)
- Lin, TY; Goyal, P; Girshick, R; He, K; Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988)
- Lin, TY; Maire, M; Belongie, S; Hays, J; Perona, P; Ramanan, D; Zitnick, CL. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Springer, Cham
- Liu, R; Lehman, J; Molino, P; Petroski Such, F; Frank, E; Sergeev, A; Yosinski, J. (2018). An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in neural information processing systems*, 31
- Liu, Z; Wang, D; Liu, Y; Liu, X. (2016). Adaptive adjustment algorithm for non-uniform illumination images based on 2D gamma function. *Transactions of Beijing Institute of Technology*, 36(2), 191-196
- Long, J; Shelhamer, E; Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *arXiv preprint*, 1-10. arXiv:1411.4038v2
- Milletari, F; Navab, N; Ahmadi, SA. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). IEEE
- Russell, BC; Torralba, A; Murphy, KP; Freeman, WT. (2008). LabelMe: a database and web-based tool for image annotation. *International J. Comp. Vision*, 77(1), 157-173
- Song, X; Bokkers, EAM; van der Tol, PPJ; Koerkamp, PG; Van Mourik, S. (2018). Automated body weight prediction of dairy cows using 3-dimensional vision. *J. Dairy Sci.* 101(5), 4448-4459
- Tian, Z; Shen, C; Chen, H; He, T. (2019). Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9627-9636)
- Wang, X; Kong, T; Shen, C; Jiang, Y; Li, L. (2020). Solo: Segmenting objects by locations. In *European Conference on Computer Vision* (pp. 649-665). Springer, Cham