Livestock classification and counting in quadcopter aerial images using Mask R-CNN

Beibei Xu, Wensheng Wang, Greg Falzon, Paul Kwan, Leifeng Guo, Zhiguo Sun & Chunlei Li

Published online: 21 Apr 2020.

Submit your article to this journal

Article views: 4112

View related articles

View Crossmark data

Citing articles: 21 View citing articles

Taylor & Francis
Taylor & Francis Group

Livestock classification and counting in quadcopter aerial images using Mask R-CNN

Beibei Xu [ID][a], Wensheng Wang[a,b], Greg Falzon[c,d], Paul Kwan[c,e], Leifeng Guo[a,b], Zhiguo Sun[a] and Chunlei Li[a]

[a]Agricultural Information Institute, Chinese Academy of Agriculture Science, Beijing, China; [b]Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing, China; [c]School of Science & Technology, University of New England, Armidale, Australia; [d]Precision Agriculture Research Group, University of New England, Armidale, Australia; [e]School of Information Technology & Engineering, Melbourne Institute of Technology, Melbourne, Australia

ABSTRACT

Quadcopters equipped with machine learning vision systems are bound to become an essential technique for precision agriculture applications in pastures in the near future. This paper presents a low-cost approach for livestock counting jointly with classification and semantic segmentation which provide the potential of biometrics and welfare monitoring in animals in real time. The method used in the paper adopts the state-of-the-art deep-learning technique known as Mask R-CNN for feature extraction and training in the images captured by quadcopters. Key parameters such as IoU (Intersection over Union) threshold, the quantity of the training data and the effect the proposed system performs on various densities have been evaluated to optimize the model. A real pasture surveillance dataset is used to evaluate the proposed method and experimental results show that our proposed system can accurately classify the livestock with an accuracy of 96% and estimate the number of cattle and sheep to within 92% of the visual ground truth, presenting competitive advantages of the approach feasible for monitoring the livestock.

1. Introduction

In order to meet the growing population demand for meat and improve meat quality, livestock monitoring including behaviours and health has become a hot research topic among livestock management. The application of information technologies known as the Internet of things, remote sensing and computer vision are desires for managing the pasture automatically and intelligently (Ray 2017; O'Grady and O'Hare 2017). Smart pasture plays an increasingly important role in the construction of precision agriculture especially in agriculture-developed countries such as Australia and New Zealand which are scarcely populated countries with vast rangelands but are developing intensive and large-scale farms (O'Grady and O'Hare 2017). Traditionally, livestock farming is normally done in a wildlife environment (Qazi et al. 2018) and the challenges including animal's

CONTACT Wensheng Wang ✉ wangwensheng@caas.cn 🖅 Agricultural Information Institute, Chinese Academy of Agriculture Science, No.12 South Street, Zhongguancun, Haidian District, Beijing, China

death or loss due to hunting or unintentional factors such as drown in the river or landslides, and dangerous infectious diseases carried from invasive species, pose great threats to livestock management. With accurate knowledge of species and quantity of livestock, the farmers or the farm managers can efficiently monitor the animals to avoid animals' loss or invasion by other species causing vandalization of crops such as hares and wild boars (Priyadharshini et al. 2018). In addition, the farmers or the farm managers need to consciously control the number of livestock to adapt to the carrying capacity of the grazed pastures, or the pastures will be overgrazed which leads to soil degradation and environmental damage (Evju et al. 2006; Oesterheld, Sala, and McNaughton 1992).

The general approach for farmers to get information on numerous livestock is visual observation which is useful but very costly and time-consuming. For some farms with advanced technologies, wearable sensors including GPS collars (Bailey et al. 2018), ear tags (Kumar and Singh 2016) and Radio Frequency Identification (RFID) (Voulodimos et al. 2010; Ismail and Ariff 2018) are becoming important options for farm management which also emphasize on individual wellbeing. Constrained by high costs on materials and limited transmission scope for network especially in large geographic ranges and inaccessible habitats, biometrics is only demonstrated on some farmlands at a range of distances. Other existing monitoring techniques including surveillance cameras (Gula et al. 2010; Gogoi 2015), thermal cameras (Ward et al. 2016) and camera traps (Verma and Gupta 2018; Norouzzadeh et al. 2018) typically require considerable investment in time and resources. In addition, these facilities are expensive to maintain and are also limited in the flexibility of ranges and disturbance from surroundings (Gonzalez et al. 2016). Recent advances in machine vision in agriculture and automation enable us to obtain huge data about the visual aspects for animals over larger space and time domain. The availability of economical quadcopters offers a potential solution to address the above challenges by diminishing cost owing to longer endurance and high repeatability and by performing flight paths autonomously at almost everywhere (Sa et al. 2016; Gonzalez et al. 2016). As such, quadcopters have been widely used in animals and wildlife monitoring in recent years such as goat groups (Qazi et al. 2018), yak (Su et al. 2018), sea turtles (Bevan et al. 2015), birds, large herbivores and mammal (Linchant et al. 2015). However, extracting useful knowledge from these quadcopter-based images remains a time-consuming and costly manual task (Norouzzadeh et al. 2018). Therefore, in addition to the restrictions of quadcopter regulations, the development of vision system with automatic image processing for some special tasks is also very important. There is a dire need for visual system to process images captured by quadcopters to detect and recognize the species accurately and automatically which is a fundamental but crucial step for animals' populations as well as behaviours and health monitoring. With subtle changes in illumination, the colour similarity between animals and background, overlapping among animals and obstacles like rocks and branches, it is really a challenge to do species identification and counting in real-world setting such as pastures. Despite these challenges, the rapid developments of object detection (He et al. 2017; Ren et al. 2015; Liu et al. 2016; Redmon et al. 2016) of machine learning in computer vision provide promising techniques for animal detection and classification.

Studies on animal detection and counting in the scene using convolutional neural network detector were implemented (Ardo et al. 2017; Ardö et al. 2017; Chamoso et al.

2014; Guzhva et al. 2018; Shao et al. 2020). However, these researches above just focus on the single species detection and the experiment is in dairy barn or feedlot. Yu et al. (2013) improved sparse coding spatial pyramid matching (ScSPM) to recognize 18 animal species which achieved an average classification accuracy of 82% (Yu et al. 2013). A deep convolutional neural network was proposed in (Chen et al. 2014) based species recognition algorithm for wild animal classification, but the results were unsatisfactory. Cao et al. (2015) combined convolutional neural network (CNN) with hand-designed images features to classify marine animals, yielding better classification results than existing approaches (Cao et al. 2015). Kumar, Manohar, and Chethan (2015) proposed graph-cut based technique with K-nearest neighbours classifier for the classification of animals (Kumar, Manohar, and Chethan 2015). Gomez Villa, Salazar, and Vargas (2017) and Nguyen et al. (2017) both also adopted the deep neural networks using Snapshot Serengeti (SSe) dataset and Wildlife Spotter project dataset, respectively, but the former did improvements in data processing to identify animal species (Gomez Villa, Salazar, and Vargas 2017; Nguyen et al. 2017). The former results showed it outperformed other previous approaches for the most common 26 species and the latter achieved 90.4% accuracy for identifying the three most common species. Norouzzadeh et al. (2018) trained the deep convolutional neural networks using 3.2 million images of Serengeti wildlife and then automatically classified 48 species with 95% accuracy (Norouzzadeh et al. 2018). More recently, Tabak et al. (2019) used convolutional neural networks with the ResNet-18 architecture to automatically classify wildlife species (Tabak et al. 2019). The model achieved 98% accuracy within the sample dataset and out-of-sample also correctly identified at least 82% of images containing an animal.

Despite these advances in animal classification, the scenarios examined above were in the wild and only single animal exists in per image. Pastures, the most common cases of livestock breeding, are different from the wild. The animals in the pastures are likely to be tightly packed herds. Visual clutter (vegetation and other natural elements) and strong lighting contrast and shadows (from farm infrastructure) should also be considered. Therefore, in this paper, our aim is to take the first step towards livestock monitoring vision system applying in actual scenario. This paper will present a state-of-art object detection framework, Mask R-CNN algorithm, on livestock dataset for effective classification and counting which can work well in quadcopter system. Mask R-CNN performs not only on object detection and classification but also on instance segmentation (associating specific image pixels to the detected object). The benefits provided by instance segmentation allow for diverse future applications including estimation of animal pose and direction of travel to monitor abnormal behaviours.

In the previous researches, Mask R-CNN was used to explore road damage classification (Singh and Shekhar 2018), classification of magnetic resonance images of the knee (Couteaux et al. 2019) and identification of whale species as well as length measurement of whale individuals (Gray et al. 2019). Mask R-CNN has been demonstrated the availability of pixel-level instance segmentation applying in counting cattle with stronger robustness (Danish 2018). We also evaluated the algorithm in different cases using full-appearance detection and head detection and achieved good performance (Xu et al. 2019). However, there are many other factors which will influence the accuracy of classification and counting including the number of species, the quantity of training data and the density

of animals in the pasture. As a result, this paper will examine the effect of different densities and various number of training set on classification and counting of species to optimize the proposed model (Smith 1989).

2. Materials and methods

2.1. Data collection and processing

A publicly available dataset from real pasture scenario consisting of at least two livestock species which meets the requirements in the paper has not been produced to date due to the expensive and time-consuming nature of image annotation. The publicly available datasets like FriesianCattle dataset in the paper (Andrew 2017) are from the very specific but controversial scenarios, which was one of the less challenging computer vision scenarios for object detection and localization consisting of cattle with distinctive black and white coat patterns contrast with lush green pastures. The dataset used in the paper was collected from a private farmland in Armidale in Australia. The observation videos of 10 flight campaigns for livestock were recorded by the MAVIC PRO drone from April to October. The drone is equipped with an integrated Pan-Tilt- Zoom (PTZ) camera shown in Figure 1. The camera has a 2.3^{-1}-inch CMOS image sensor that can rotate flexibly in the lateral and vertical. The video data captured by this stabilized camera are a frame resolution 4096×2160 pixels and 30 Frames Per second (FPS). Figure 2 is showing the frame examples of cattle and sheep in the pasture.

Considering avoiding fractions when downscaling and upscaling in the convolutional neural network, the input image size used in this method must be divisible by 2 at least 6 times. In addition, the size of the input images is limited by computational capabilities and memory of computer system, so the frame images for dataset extracted from videos in our cases are clipped automatically using MATLAB to the size of 512×512 pixels. The dataset consists of 1000 images with 3737 livestock in total, of which 80% are used for training
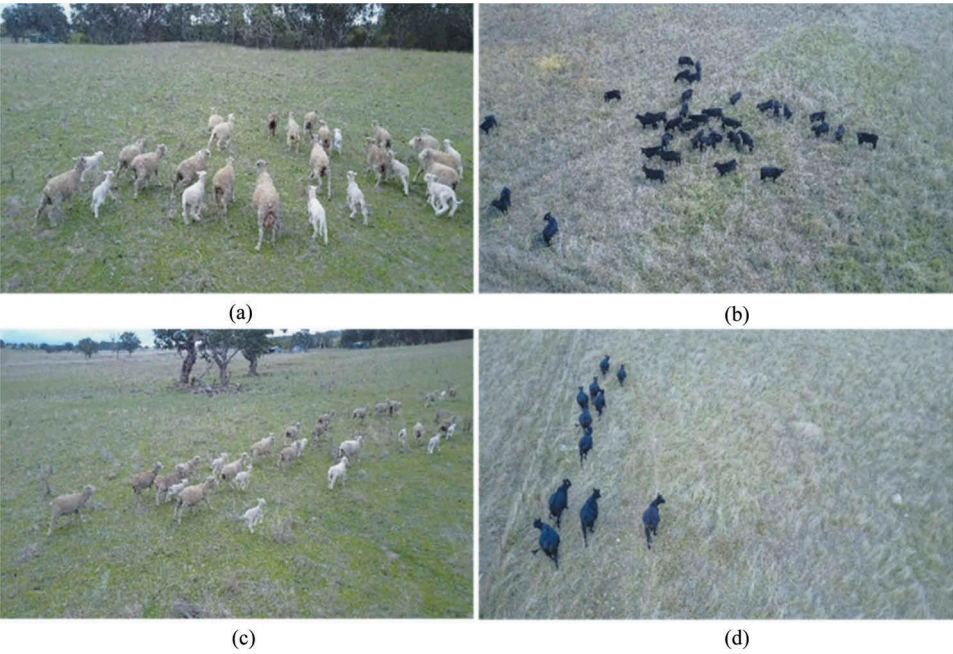

Figure 1. MAVIC PRO drone.

Figure 2. (a) and (c) are example frames of sheep and (b) and (d) are example frames of cattle.

and the rest for testing (validation dataset is not used in the research). Details of training and testing dataset are given in Table 1. The ground truth annotation for training dataset is done by the image annotation tool, LabelMe (Russell et al. 2008). Each cow and sheep are clicked along the outside edge with points until connected into a closed loop in the images and the label name also needs to be marked (see Figure 3). Then, the ground truth data are stored in a table format aligned with that required by the Mask-RCNN framework for data annotation.

2.2. Proposed system overview

In this section, the proposed livestock classification and counting system based on machine learning is illustrated in detail. Figure 4 is presenting the working block diagram for the proposed system. For the manifestation of detection and classification of livestock (sheep and cattle) in the vast pasture, the drone is used for continuous acquisition of livestock data from the captured videos in the proposed system. Further, this system extracts the video frames and performs segmentation on them. The images

Table 1. (a) Details of training images. (b) Details of testing images.

	Number of images	Number of objects
Cattle	400	1392
Sheep	400	1533
Cattle	100	600
Sheep	100	592

(a)   (b)

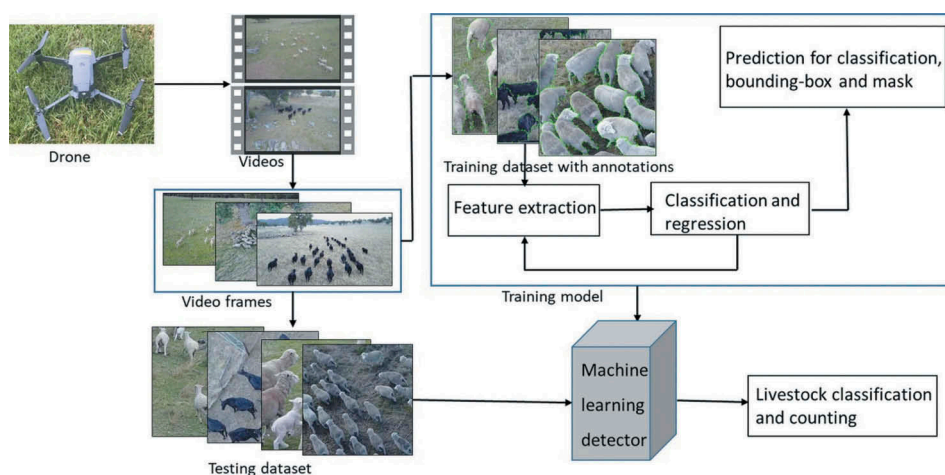**Figure 3.** (a) and (b) are examples of annotations for cattle and sheep using LabelMe.

Figure 4. Flowchart of the livestock classification and counting based on machine learning.

of 512 × 512 pixels cropped and resized from previous segmentation are used for training and testing, respectively. For training the machine learning model, livestock features are extracted from training dataset based on the annotations and then are classified and regressed simultaneously in a multitasking way. The output is the classification for livestock with pixel-level segmentation with localization through many repeated trainings of parameters optimization for the model. The machine learning detector for classification and counting is formed after optimizing some key parameters for the training model. When running the detector with testing dataset, we could get the instance segmentation, localization, classification and confidence scores for livestock and classification for sheep and cattle.

2.3. Machine learning network

The machine learning detector in the proposed system employs Mask R-CNN algorithm for livestock classification and counting. Mask R-CNN is an object detection framework with instance segmentation. It is an extension of Faster R-CNN, which is added a mask prediction branch composed of a Fully Convolutional Network for segmenting each Region of Interest (RoI). The RoI Pooling in Faster R-CNN is also replaced with RoIAlign for Mask R-CNN using the bilinear interpolation to remove the harsh quantization of RoI Pooling, properly aligning the extracted features with the input to improve the accuracy of predicting pixel-level masks (He et al. 2017). Figure 5 illustrates the architecture of Mask R-CNN network for livestock classification and counting.

Specifically, the network of Mask R-CNN applied in this research is composed of three functional modules: (*i*) the convolutional backbone architecture used for feature extraction over an entire image, and (*ii*) the network head for bounding-box recognition (classification and regression) and (*iii*) mask prediction that is applied separately to each RoI (He et al. 2017).

(*i*) Feature Extraction: The procedure starts with inputting a three-channel image (RGB image) into a pre-trained convolutional network. But due to computational constraints, the image is constrained to 512×512 dimensions. For feature extractor, the ResNet-101 performs better than others in terms of accuracy and speed including VGG16, AlexNet, and Inception and so on (He et al. 2016, 2017; Huang et al. 2017; LeCun, Bengio, and Hinton 2015). Compared with VGG16, AlexNet, Inception (He et al. 2017, 2015; Huang et al. 2016), ResNet-101 achieves competitive performance in scale-invariant feature extraction (He et al. 2017). So, the pretrained ResNet-101 using COCO dataset (Lin et al. 2014) is employed as a backbone network to extract features for Mask R-CNN. The 101 layers of ResNet are mainly constructed by six parts: conv1, conv2_x, conv3_x, conv4_x, and conv5_x which consist of three-layer blocks (conv1 excluded).

(*ii*) Region Proposal Network (RPN): On the feature maps from the last shared convolutional layer, sliding window method is conducted to obtain the feature vector at each location of the feature map. RPN, a newly high-sufficient proposal generation network in the Faster R-CNN, is used to propose Region of Interests (RoIs) based on the feature vector. It replaces the selective search method in the previous RCNN and Fast RCNN. For every minimum unit point on the feature map of last convolutional layer in conv4_x, fifteen anchors can be generated in different size (128×128, 64×64, 32×32, 16×16, 8×8) of different ratios (0.5, 1, 2) and thereby multiple ROIs are produced. All the ROIs are fed into RPN to perform binary classification (foreground or background) and preliminary Bounding-box regression to filter out some candidate ROIs using Non-maximum
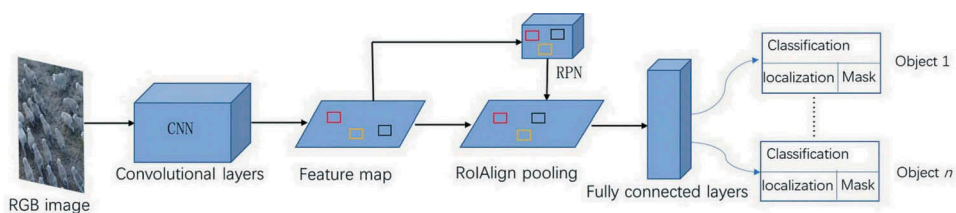

Figure 5. Architecture of the Mask R-CNN network for livestock classification and counting.

suppression. The RoiAlign then operates on the above proposed regions of different sizes, which are mapped precisely to generate fixed-size feature maps equal to the size of the convolutional network input. The ROI of 16 × 16 on the feature map with a stride substitutes for 32 × 32 with a stride of 2 in the conv5_x according to (He et al. 2015).

(*iii*) The head architecture for bounding-box recognition (classification and regression) and mask prediction: The RoIAlign layer selects the features corresponding to each RoI on the feature map, and sends them to the fully connected layer for classification prediction, mask prediction and bounding-box prediction which typically use a per-pixel Softmax loss (He et al. 2015)and average binary cross-entropy loss. The average pooling layer following the conv5_x is used for 2048-dimensional features to classify and regress the bounding-boxes through the fully connected layer with softmax. Additionally, the end-to-end Fully Convolutional Network carries out convolution and deconvolution successively to achieve accurate mask segmentation.

3. Results

3.1. Implementation details

Experiments are performed on the dataset that includes training dataset of two annotated categories and testing dataset on which the evaluation is done. Transfer learning is used in this study for training which means that the parameters of the learned and trained model on a similar task are transferred and fine-tuned to a new model for the required target task (Karri, Chakraborty, and Chatterjee 2017; Zhu et al. 2011; Pan and Yang 2010). Specifically, the network used in the proposed algorithm was initialized by the pre-trained Resnet101 with bounding box annotations. To avoid destroying the parameters learned of convolutional layers during pre-training, only the network head is trained independently for the first stage in training using the training dataset of 800 images while all the backbone layers remain fixed. Then, we conduct our training dataset to fine-tune the global layers to form the Mask R-CNN-based livestock counting detector which could take advantage of the generalized features from large-scale data. The network is trained using the Stochastic Gradient Descent (SGD) algorithm with a weight decay of 0.001 and a momentum of 0.9 and the initial learning rate is 0.01. All the training experiments have a batch size of 10 images and the epochs is 1000. The proposed system has been implemented using the deep learning framework and visual component libraries including TensorFlow, Keras and Matplotlib with an Intel core i7 2.4 GHz Processor running on CPU in a 64-bit version of Windows 10 laptop with 16GB of RAM. It takes about 3 min for per epoch on the machine. The livestock in the pastures are separate and rarely mixed together, so only cattle or sheep in each image of the dataset whose details are previously shown in Table 1. The following results of this proposed system are tested over 200 randomly chosen images, half containing cattle and half sheep. The resultant loss curve of training is shown in Figure 6 which demonstrates that the proposed system achieves the error rate at 0.2–0.3.

3.2. Evaluation protocol

The performance evaluation for the proposed system is based on IoU which is defined as a measure of similarity between the bounding boxes for predicted
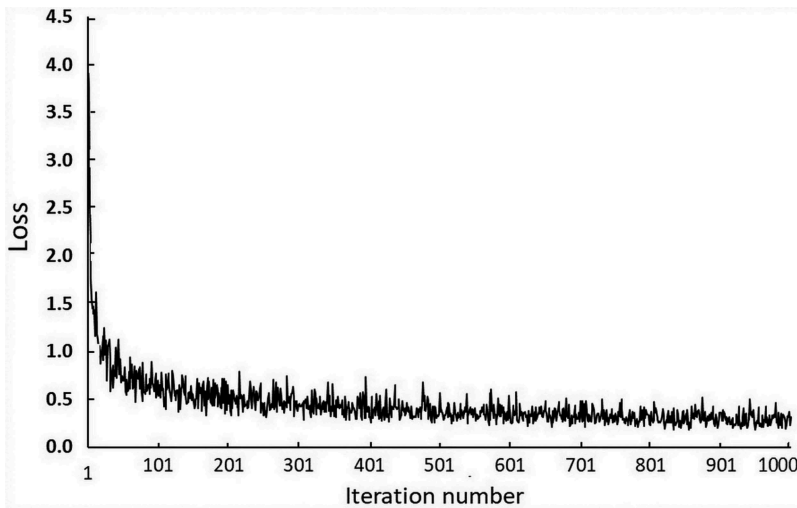
Figure 6. Training loss curve of the proposed system.

objects and ground truth. True positive, TP, is considered if the value of IoU outperforms a certain threshold ranging from 0 to 1, and false positive, FP otherwise. For evaluating the accuracy of the livestock classification and counting, we utilize the precision, recall, F1 score, confusion matrix and mean average precision (mAP) as the evaluation metrics in this paper. The precision reflects the proportion of true predicted positive in all the predicted positive while the recall reflects the proportion of true predicted positive in all the true positives. Recall is computed as the fraction of ground truth objects covered above an IoU threshold. F1 score is a statistical measure which is defined as the harmonic average between precision and recall, where it achieves the best performance at the best IoU threshold. Confusion matrix is a table with two dimensions of predicted class and true classes to do more detailed analysis than be limited to the true classification accuracy. The mAP is the average precision of all classes in the livestock and the AP is the value of area enclosed by the precision-recall curve at different IoU threshold that is introduced in the Pascal VOC Challenge. To evaluate the counting and classification of results in the work, we compare the detected results using the proposed approach with the ground truth. The classification accuracy and counting accuracy follow Equation (1) and (2):

$$A(\%) = (1 - \frac{|d - g|}{g}) \times 100\% \tag{1}$$

$$B(\%) = \frac{c}{g} \times A \times 100\% \tag{2}$$

Here, *A* is the counting accuracy and *B* is the conditional classification accuracy. *A* can be calculated via the detected counting (*d*) and ground truth (*g*) and *B* calculates the probability of correct classification (*c*) in the condition of correct detection.

### 3.3. Selection of iou thresholds for the proposed system

Most of the output proposals generated from RPN are duplicates and thus need to be filtered out. The quality of bounding box proposals for cattle or sheep in images is typically evaluated by IoU score which measures the union area of box proposal and ground-truth box. Only high scoring proposals are selected to the RoIAlign layer. The threshold of IoU from 0.4 to 0.7 is used to indicate successful detection in many cases (for example, 0.5 in PASCAL VOC challenge) (Kuo, Hariharan, and Malik 2015; Ghodrati et al. 2015). The performance of the network for detecting cattle or sheep will be poor if the threshold is not set appropriately which means overlapping bounding-box predictions for higher threshold and missing objects for lower threshold. However, the threshold may vary for different dataset, in this section, we evaluate the performance over varying IoU thresholds to decide the optimal value for the proposed system.

The precision, recall and F1 scores are utilized to compare the performance of different IoU thresholds shown in Figure 7. The green solid lines are the overall precision rate and recall rate for livestock in Figure 7(a). The higher the threshold is, the lower the recall that is considered to be true object rate, and the higher the precision that is considered to be object detection accuracy. At about IoU = 0.4, the precision and the recall achieve the same value considered as a balance point. Both precision and recall can get best performance because the predicted objects are the true objects. As can be seen in Figure 7(b), the F1 score for the overall livestock achieves the highest value at about IoU = 0.4 with a precision rate of 0.955 and a recall rate of 0.952, which is consistent with the previous results. Compared the precisions, recalls an F1 scores of cattle and sheep, respectively, under different IoU thresholds in Figure 7, similar trends are with the overall livestock and the performance of the threshold at around 0.4 is significantly better than others. Furthermore, the sheep detection outperforms the cattle.

3.4. Impact of number of images on livestock classification and counting

Typically, large amounts of data are essential for training the deep learning model in order to avoid overfitting. Since there are many parameters in deep neural networks, so if there is not enough data for training them, they tend to remember the entire training set, which will result in good training, but bad performance on testing set. Therefore, we evaluate the performance of different number of data sets (200,400,600,800 and 1000, respectively) using the training loss curve and precision-recall metric.

Figures 8 and 9 explore the trade-off between the performances measured with training loss, precision and recall rate on the testing set and the quantity of training images. For fair of comparison, the number of epochs for training are set to be 1000 for all the cases and the overlap threshold is set to be 0.4 we tested in last section. Results in Figure 8 present a consistent trend across the curves of different number of training images which indicates that the training loss drops rapidly with the increase of the epochs
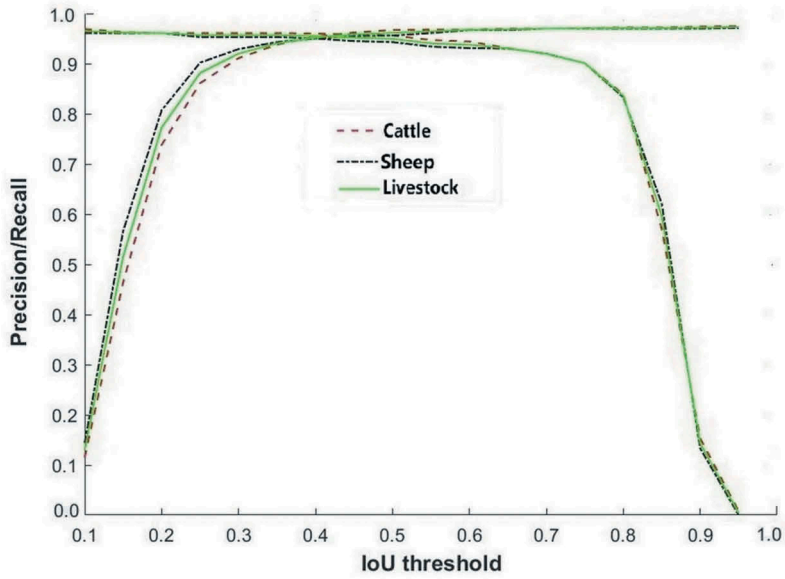
(a)
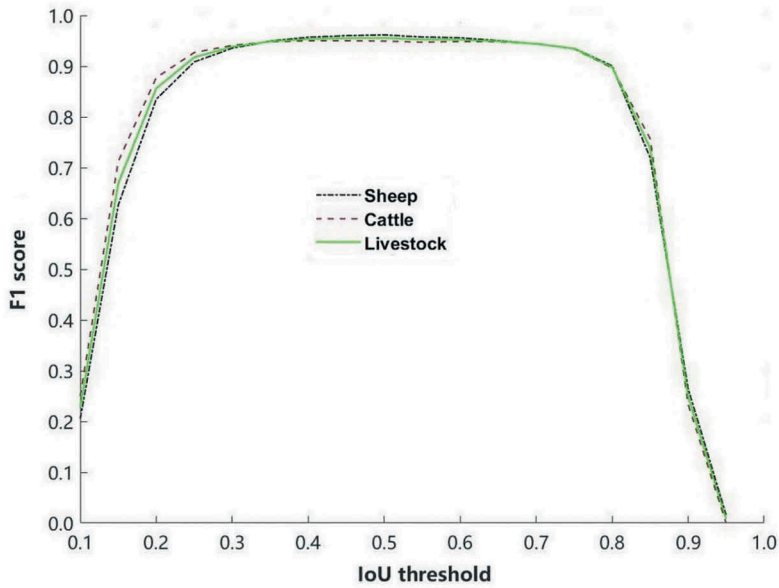
(b)

Figure 7. (a) is precision and recall as a function of IoU threshold and (b) is F1 scores as a function of IoU threshold.

and then decrease at a slow rate after about 200 epochs, and finally tend to be stable at around 900 epochs. However, the number of 1000 and 200 images seem to outperform than others considering the values of loss and the stability of network convergence during training.
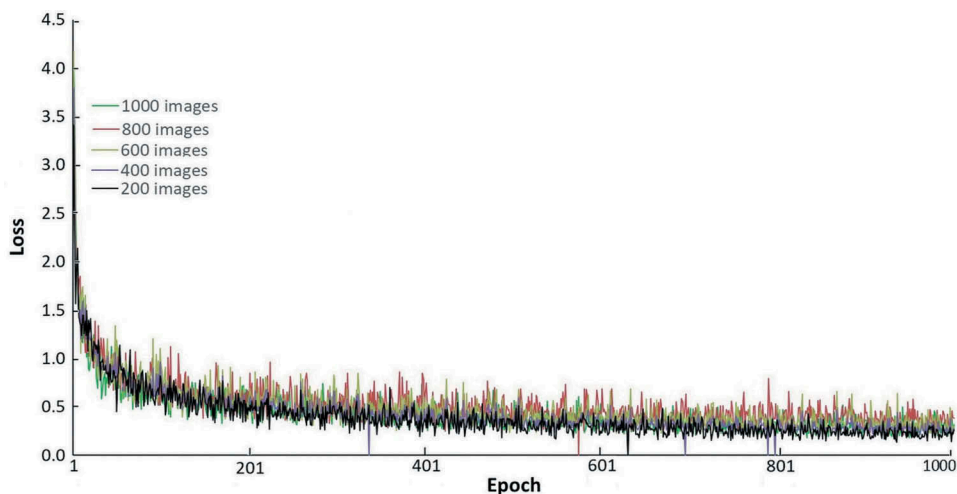
Figure 8. Training loss curves of different number images for the proposed system.
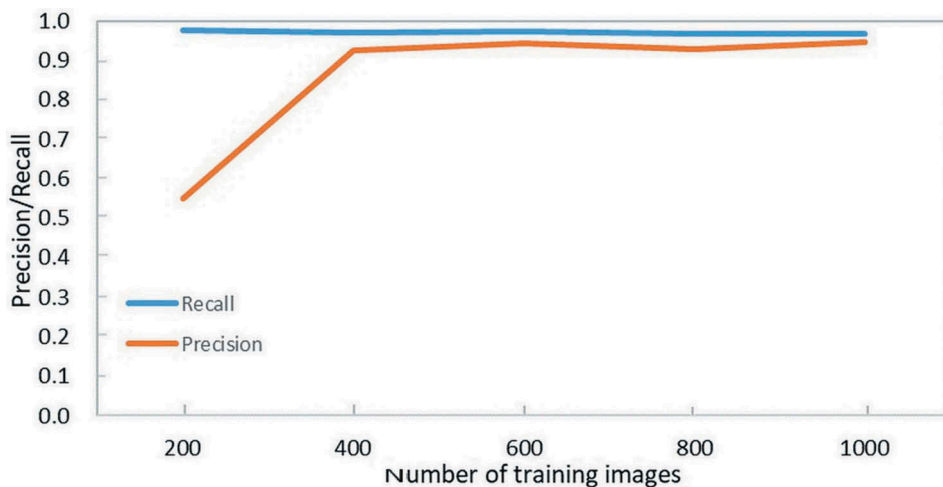

Figure 9. Recall and precision versus different number of training images at IoU threshold of 0.5.

Figure 9 reflects the overall quality of the results for a given number of images. When inspecting Figure 9 from fewer images to more images, one notices that the proposed system outputs low precision on testing set with only 200 training images compared with other more training images. Furthermore, results show that the proposed system achieves both high and stable recall as well as precision across almost the whole range of the number of training images (except number 200). Therefore, we have a reasonably consistent quality & quantity trade-off considering both the training loss and precision and recall of testing set.

In terms of the computation cost and detection accuracy, when 1000 images are adopted for training model with 1000 epochs, we obtain a good quantity & quality trade-off whose recall and precision are higher than 90%, which shows that the proposed system is effective for livestock classification of high quality. On the contrary, although

equally low in training loss with the former, the training model of 200 images doesn't produce good performance. The reason is that the limited data couldn't train the complicated networks of many parameters and feature dimensions very well, so the training model is overfitting without considering the generalization ability.

3.5. Results of livestock classification and counting

The performance evaluation of livestock classification and counting follows the same training and testing settings described in Section 2.2 and Section 3.1, and the IoU threshold is set as 0.4. Even though the precision and the recall as a function of IoU thresholds, respectively, are previously presented in Section 3.3, we need to combine these two metrics to evaluate the performance of livestock classification. Figure 10 shows the precision-recall curves over varying IoU thresholds. It can be observed that the performance of the detection of sheep is better than cattle which is consistent with the conclusion in Section 3.3. The higher the precision and the recall are, the better the performance of livestock classification is, therefore, inflection points at curves are chosen known as balance points where they could achieve the optimal values. The best precision rate is 0.955, 0.960 and 0.950 for livestock, sheep and cattle at IoU threshold = 0.4 while the best recall is 0.952, 0.950 and 0.954.
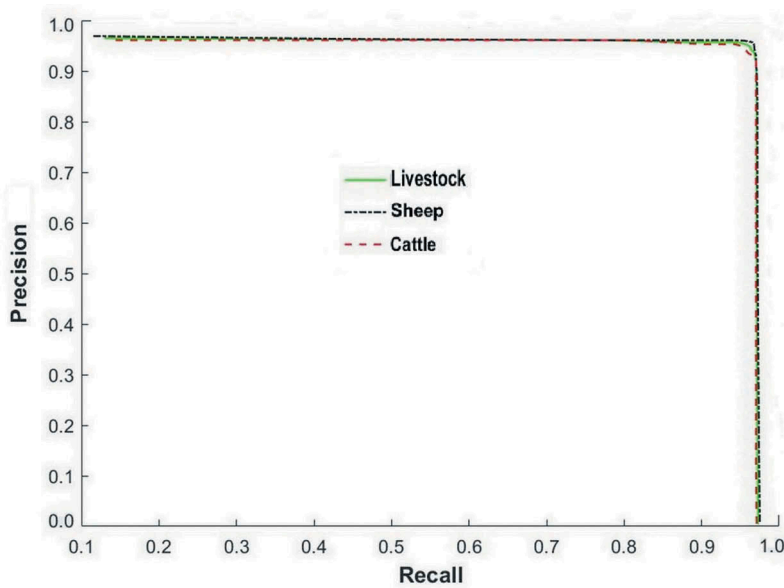

Figure 10. Precision-recall curves of livestock classification.

Table 2. Confusion matrix of the livestock classification results.

prediction	True Cattle	Sheep	Non-livestock
Cattle	573	6	53
Sheep	0	569	39

Table 3. Livestock classification and counting results.

	No. in reference data set	No. livestock detected	No. correct identification	No. missed	Counting accuracy (%)	Classification accuracy (%)
Cattle	600	632	573	18	94.7	90.4
Sheep	592	608	569	23	97.3	93.5
Total	1192	1240	1142	41	96.0	92.0

In order to visually evaluate the performance of the classifier, the confusion matrix of the testing results is presented in Table 2. Notice that the classification error of cattle mistaken for sheep and non-livestock mistaken for cattle are bit higher than the sheep mistaken for cattle or non-livestock mistaken for sheep which indicates the accuracy of sheep detection is higher than the cattle on testing dataset. The detailed results of livestock classification and counting is shown in Table 3. As observed, the proposed approach yields an accuracy of 96.0% for livestock counting and 92.0% for classification. However, the results considering the counting and classification accuracy for sheep (97.3% and 93.5%) are both precisely than the cattle (94.7% and 90.4%).

We further demonstrate the classification results for each image in testing dataset in Figure 11(a). The image index from 1 to 20 is with more livestock, than the index ranging (21 ~ 40) and (60 ~ 80) which is also more than the rest. The classification accuracy is mainly between 50% and 100% and the accuracy changes a lot with different image index especially for cattle. In the meanwhile, we also compute the counting accuracy of 200 images whose values are also mainly between 50% and 100% shown in Figure 11(b). It can be inferred from above analyses that either the classification or counting, the sheep performs better than cattle because of the complexity of the environment for cattle and various postures of cattle which is difficult to distinguish from other obstructions, such as trees and rocks in the pastures.

We note that the discrepancy of accuracy in Figure 11 may suffer from the density of livestock in images. Specifically, the higher accuracy occurs when the livestock is dense in contrast with images consisting of sparse objects. In Table 4, we show how the proposed system performs on different range of livestock in images. For both cattle and sheep, we regard the number of ground truth below 4 as low density, larger than 10 (included) as high and middle otherwise. The performance results of above every case are obtained from 20 images of each testing dataset for cattle and sheep, respectively. The comparative results show that the classification and counting for sheep and cattle keep the similar trend with the density changing. With the density increasing, although the missed counting increase, the livestock classification and counting accuracy also improve gradually. The reason can be attributed to the small percentage of the misidentified accounting for the total number of livestock. Due to limited testing data, there is no consistent conclusion on the classification and counting that which livestock is better over varying density. However, the performance of high density of cattle and sheep illustrates the advantage of the proposed approach over the occlusion between livestock which can effectively detect the class and location of cattle and sheep. The instances of livestock classification with different density is presented throughout Figure 12.
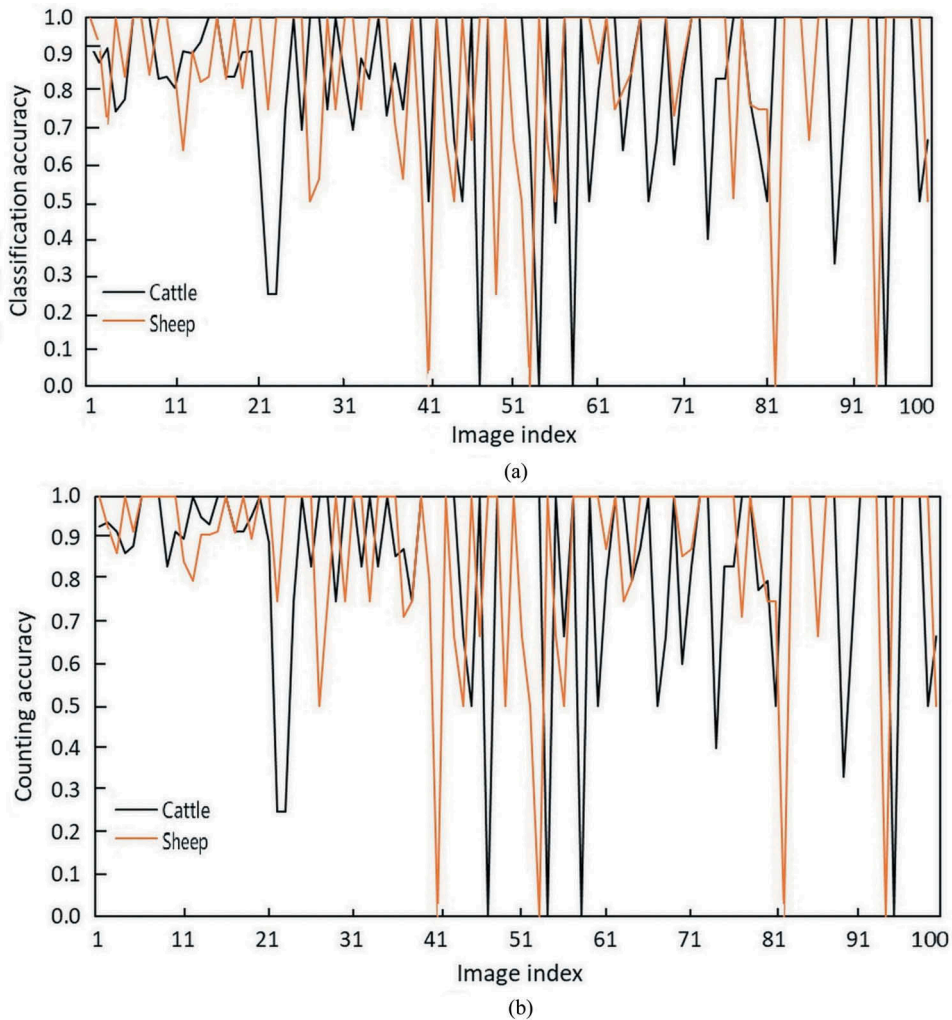
Figure 11. (a) and (b) are the classification accuracy and counting accuracy of 200 images, respectively.

Table 4. Performance comparisons of different density.

Range of in reference data set		No. in reference data set	No. detected	No. correct		No. missed		Counting accuracy (%)	Classification accuracy (%)
				Cattle	Sheep	Cattle	Sheep		
Low (<4)	Cattle	38	43	35	–	2	–	86.8	79.9
	Sheep	44	55	–	43	–	0	75	73.3
Middle (<10)	Cattle	113	120	106	–	4	–	93.8	88.0
	Sheep	107	111	–	104	–	3	96.3	93.6
High (≥10)	Cattle	292	289	278	–	11	–	99.0	94.3
	Sheep	286	282	–	273	–	15	98.6	94.1

4. Discussion

This paper has presented a novel vision-based statistical and recognition approach that takes advantage of machine learning to automate the identification process of livestock
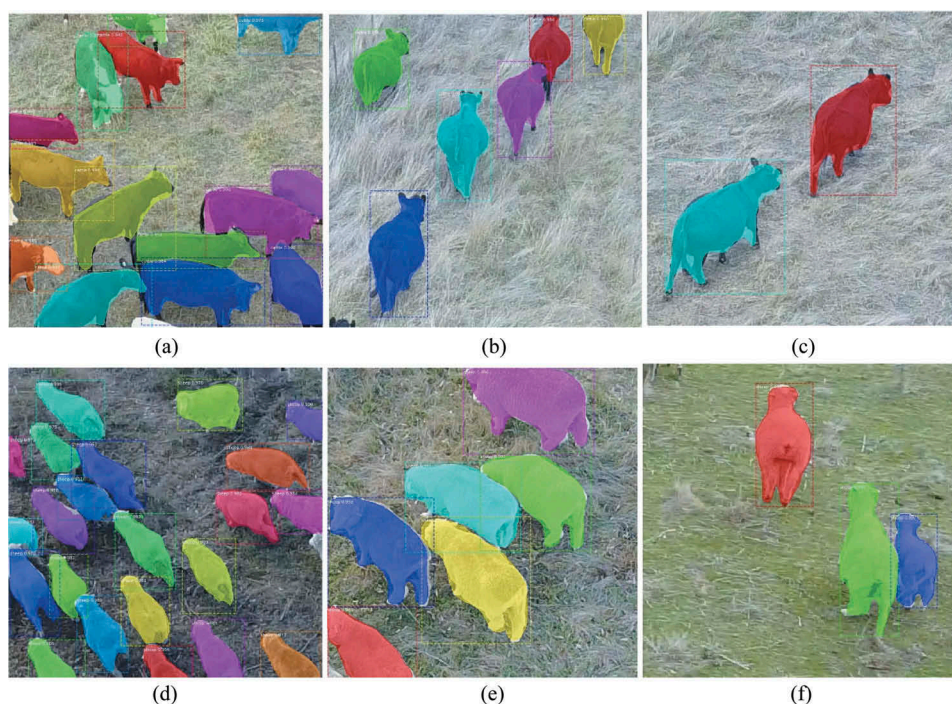
Figure 12. (a), (b) and (c) are sample detection results on the cattle dataset. (d), (e) and (f) are sample detection results on the sheep dataset.

for the quadcopter vision system in the farmland. The key novelty of the study is the application of the Mask R-CNN algorithm and the demonstration of its effectiveness for this important livestock monitoring task. The essence of the livestock classification in this paper is object-based segmentation and classification with confidence and mask, that is, the result is whether it is cattle or sheep. Previous studies have demonstrated the demand for quadcopters and real-time remote sensing capability of quadcopters to rapidly record livestock (Rahnemoonfar, Foster, and Starek 2017), but the existing research in single animal species counting suffer from the deviation of bounding-box and the challenge for mask detection (Alberto et al. 2017). A major advantage of the Mask R-CNN approach is the ability to perform both detection and classification as well as instance segmentation of livestock within the imagery, this allows the development of further algorithms to perform tasks such as welfare monitoring from the imagery. Specifically, Mask R-CNN can also be used for key point detection (He et al. 2017), which can be used for real-time detection of behaviours of the animals to provide early warning for diseases like oestruses (Dolecheck et al. 2015; Tian et al. 2013). Livestock instance segmentation in the paper is the first step towards real-time animal monitoring in farming environments that have different applications, such as early lameness detection (Viazzi et al. 2012) and other animal welfare improvements.

Owing to the relatively low-flying altitudes and high-resolution imaging, the quadcopter presents the potential capabilities of quick monitoring of large areas and detailed aerial imagery of livestock (Windrim et al. 2019) to manage the livestock, providing the benefits of low cost, time efficiency and convenient operation. The aim is to build an

accurate, fast and reliable livestock classification system, which plays a vital part in an autonomous robotic system for livestock management (Van Hertem et al. 2018); it is a key element for automated livestock monitoring such as the individual behaviour activities, housing welfare and grazing estimation of grassland (Nasirahmadi, Edwards, and Sturm 2017; Nir et al. 2018). In this study, Mask R-CNN deep learning network is adopted which we modify and fine-tune on our own training data to detect and classify cattle and sheep. The multi-class classification approach achieves an accuracy above 92% without the need of any pre-processing steps such as data augmentation. It is worth repeating that while validation data is not used in this study due to limited labelled data. The validation data of machine learning is used to help supervise the model for overfitting and adjust para-meters. As a replacement for the validation set, this study uses TensorBoard provided by TensorFlow for visualizing the training process and repeatedly readjusts the parameters manually for optimizing the model. The study reports evaluation metrics of test data based on the optimal model for the task of livestock detection and classification.

In previous experiments for the application of cattle instance segmentation, cell nucleus segmentation and pose estimation using Mask R-CNN (Danish 2018; Ter-Sarkisov et al. 2018; Johnson 2018; Chen et al. 2018), we find that the IoU threshold changes a lot under a variety of conditions ranging from 0.5 to 0.7. Therefore, we compute the precision, recall and F1 score for each IoU from 0.1 to 0.9, among which the IoU of 0.4 leads to a good detector for livestock classification, to optimize the parameters for the model. Although the optimal threshold is not same with that in our previous research (0.5), it confirms the conclusion that this IoU threshold of object detection framework should be properly adjusted depending on the circumstances and applications.

The number of training images is also an important parameter which directly affects the quality of training model and, hence, indirectly decides on the final object detection and classification performance. The extensive evaluation enables researchers to make more informed decisions when considering on the quantity of training data. The number of 1000 are found to provide the best compromise regarding the training loss versus testing quality in the proposed approach. This is mainly because the dataset is collected in the relatively simple environment without too much interference, and the feature difference between a herd of cattle or sheep flock in the same region is not obvious. However, it is unlikely that others could achieve the same results in a more heterogeneous environment, especially since the sheep and cattle stand out so much more than animals that are more similar in colour or texture to their environment. In the future work, we will enlarge the dataset under complex scenes to further demonstrate the generalization of this approach.

We also investigate how the proposed system performs on images of varying density. The proposed approach is experimentally shown that it tends to overestimate the classification in cases of images with more than 10 cattle or sheep per image but under-estimate below 10. This estimation error could possibly be a consequence of the insuffi-cient number of training images with such large crowds in the dataset. Moreover, the detection of sheep outperforms cattle due to diverse postures which indicates that there is still huge space for adopting heterogeneous visual feature integration to further boosting the object detection performance by designing more powerful feature representations.

5. Conclusions

This paper addresses the challenge of automated livestock detection and classification in quadcopter imagery by describing a system that combines a quadcopter and artificial intelligence image processing for aerial survey. This technique would be especially useful in vast and inaccessible rugged terrain. The proposed system could process the images captured in-place by quadcopters rather than manual observation which is challenging and time-consuming or the sensors with limited transmission distance. It provides a practical and applicable solution for detection and instance segmentation of livestock within the imagery, which in fact represents a comparable trend with other approaches for livestock monitoring. The proposed machine-learning-based quadcopter vision system has proven to be effective at livestock recognition and counting in a convenient and timely fashion that can perform up to with an accuracy of 96% for livestock classification and 92% for livestock counting. The goal of the work is to build the more appropriate quadcopter vision system for our task. Therefore, the experimental results and comparisons over different IoU thresholds as well as the impact on number of training dataset demonstrate how the proposed system is able to effectively recognize two different categories of livestock.

The proposed system is expected to make a significant contribution to the agriculture research area especially for the balance between carrying capacity and the stocking rate of the grassland (Pádua et al. 2017). Future works will involve the integration of the proposed algorithm with a quadcopter and improve the current accuracy. The NVIDIA GPU (Graphics Processing Unit) device is also required to improve the processing time if we deploy the system into the quadcopter system. A promising quadcopter vision system could be extended to protect the crops against animal intrusion (Priyadharshini et al. 2018) and for animal search or rescue remote sensing in agricultural environments. Furthermore, the multispectral cameras will be deployed on quadcopters for land cover and vegetation mapping applications to help assess the grazing capacity of pastures (Ahmed et al. 2017; Marcial-Pablo et al. 2019; Yang and Everitt 2010).

Disclosure statement

The authors declare no conflicts of interest.

Ethics statement

The experimental procedures related with animals used in the study conform to the provision of the Declaration of Helsinki. Animals involved in this study did not interfere with the research equipment, all the interactions with animals and their environment were approved by farmers.

Funding

ORCID

Beibei Xu http://orcid.org/0000-0001-5804-2906

References

Ahmed, O. S., A. Shemrock, D. Chabot, C. Dillon, G. Williams, R. Wasson, and S. E. Franklin. 2017. "Hierarchical Land Cover and Vegetation Classification Using Multispectral Data Acquired from an Unmanned Aerial Vehicle." *International Journal of Remote Sensing* 38 (8–10): 2037–2052. doi:10.1080/01431161.2017.1294781.

Alberto, R., C. Pablo, G.-B. Alfonso, and C. Juan. 2018. "Detection Of Cattle Using Drones and Convolutional Neural Networks." *Sensors* 18 (7): 2048.

Andrew, W. 2017. *Visual Localisation and Individual Identification of Holstein Friesian Cattle via Deep Learning*. Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, October 2850-2859..

Ardö, H., O. Guzhva, M. Nilsson, and A. H. Herlin. 2017. "Convolutional Neural Network-based Cow Interaction Watchdog." *IET Computer Vision* 12 (2): 171–177. doi:10.1049/iet-cvi.2017.0077.

Ardo, H., O. Guzhva, M. Nilsson, and A. Herlin. 2017. "A CNN-based Cow Interaction Watchdog." *IET Computer Vision* 2017: 1–7.

Bailey, D. W., M. G. Trotter, C. W. Knight, and M. G. Thomas. 2018. "Use of GPS Tracking Collars and Accelerometers for Rangeland Livestock Production Research." *Translational Animal Science* 2 (1): 81–88. doi:10.1093/tas/txx006.

Bevan, E., T. Wibbels, B. M. Z. Najera, M. A. C. Martinez, L. A. S. Martinez, F. I. Martinez, J. M. Cuevas, T. Anderson, A. Bonka, and M. H. Hernandez. 2015. "Unmanned Aerial Vehicles (UAVs) for Monitoring Sea Turtles in Near-shore Waters." *Marine Turtle Newsletter* 145: 19–22.

Cao, Z., J. C. Principe, B. Ouyang, F. Dalgleish, and A. Vuorenkoski. 2015. "Marine Animal Classification Using Combined CNN and Hand-designed Image Features." Paper presented at the OCEANS 2015 - MTS/IEEE, Washington, October 19–22.

Chamoso, P., W. Raveane, V. Parra, and G. Angélica 2014. "UAVs Applied to the Counting and Monitoring of Animals."

Chen, G., T. X. Han, Z. He, R. Kays, and T. Forrester. 2014. "Deep Convolutional Neural Network Based Species Recognition for Wild Animal Monitoring." Paper presented at the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, October 27–30.

Chen, Y., Z. Wang, Y. Peng, Z. Zhang, Y. Gang, and J. Sun. 2018. "Cascaded Pyramid Network for Multi-person Pose Estimation." Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City,USA.

Couteaux, V., S. Si-Mohamed, O. Nempont, T. Lefevre, A. Popoff, G. Pizaine, N. Villain, I. Bloch, A. Cotten, and L. Boussel. 2019. "Automatic Knee Meniscus Tear Detection and Orientation Classification with Mask-RCNN." *Diagnostic and Interventional Imaging* 100 (4): 235–242. doi:10.1016/j.diii.2019.03.002.

Danish, M. 2018. "Beef Cattle Instance Segmentation Using Mask R-Convolutional Neural Network."

Dolecheck, K., W. Silvia, G. Heersche, Y. Chang, D. Ray, A. Stone, B. Wadsworth, and J. Bewley. 2015. "Behavioral and Physiological Changes around Estrus Events Identified Using Multiple Automated Monitoring Technologies." *Journal of Dairy Science* 98: 8723–8731. doi:10.3168/jds.2015-9645.

Evju, M., A. Mysterud, G. Austrheim, and R. H. Økland. 2006. "Selecting Herb Species and Traits as Indicators of Sheep Grazing Pressure in a Norwegian Alpine Habitat." *Ecoscience* 13 (4): 459–468. doi:10.2980/1195-6860(2006)13[459:SHSATA]2.0.CO;2.

Ghodrati, A., A. Diba, M. Pedersoli, T. Tuytelaars, and L. Van Gool 2015. "DeepProposal: Hunting Objects by Cascading Deep Convolutional Layers."

Gogoi, M. 2015. "Protection of Crops from Animals Using Intelligent Surveillance System." *Journal of Applied and Fundamental Sciences* 1 (2): 200.

Gonzalez, L. F., G. A. Montes, E. Puig, S. Johnson, K. Mengersen, and K. J. Gaston. 2016. "Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation." *sensors* 16 (1): 97. doi:10.3390/s16010097.

Gray, P. C., K. C. Bierlich, S. A. Mantell, A. S. Friedlaender, J. A. Goldbogen, and D. W. Johnston. 2019. "Drones and Convolutional Neural Networks Facilitate Automated and Accurate Cetacean Species Identification and Photogrammetry." *Methods in Ecology and Evolution* 10 (9): 1490–1500. doi:10.1111/2041-210X.13246.

Gula, R., J. Theuerkauf, S. Rouys, and A. Legault. 2010. "An Audio/video Surveillance System for Wildlife." *European Journal of Wildlife Research* 56 (5): 803–807. doi:10.1007/s10344-010-0392-y.

Guzhva, O., H. Ardö, M. Nilsson, A. Herlin, and L. Tufvesson. 2018. "Now You See Me: Convolutional Neural Network Based Tracker for Dairy Cows." *Frontiers in Robotics and AI* 5: 107. doi:10.3389/frobt.2018.00107.

He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017. "Mask R-CNN." Paper presented at the Computer Vision (ICCV), 2017 IEEE International Conference on, Venice, Italy.

He, K., X. Zhang, S. Ren, and J. Sun. 2015. "Deep Residual Learning for Image Recognition."

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." Paper presented at the Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition, Las Vegas, USA.

Huang, J., V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, and S. Guadarrama. 2017. "Speed/accuracy Trade-offs for Modern Convolutional Object Detectors." Paper presented at the IEEE CVPR, Hawaii, USA.

Huang, J., V. Rathod, S. Chen, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, S. Yang, and S. Guadarrama. 2016. "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors."

Ismail, I., and M. H. Ariff. 2018. "RFID Application Development for a Livestock Monitoring System." In *Bioresources Technology in Sustainable Agriculture*, 81–94. Apple Academic Press: Palm Bay, FL, USA.

Johnson, J. W. 2018. "Adapting Mask-rcnn for Automatic Nucleus Segmentation." *arXiv Preprint* arXiv:1805.00500.

Karri, S. P. K., D. Chakraborty, and J. Chatterjee. 2017. "Transfer Learning Based Classification of Optical Coherence Tomography Images with Diabetic Macular Edema and Dry Age-related Macular Degeneration." *Biomedical Optics Express* 8 (2): 579–592. doi:10.1364/BOE.8.000579.

Kumar, S., and S. K. Singh. 2016. ""Monitoring of Pet Animal in Smart Cities Using Animal Biometrics." *Future Generation Computer Systems 83:553-563*.

Kumar, Y. H., S. N. Manohar, and H. K. Chethan. 2015. "Animal Classification System: A Block Based Approach." *Procedia Computer Science* 45: 336–343. doi:10.1016/j.procs.2015.03.156.

Kuo, W., B. Hariharan, and J. Malik. 2015. "DeepBox: Learning Objectness with Convolutional Networks." Paper presented at the IEEE International Conference on Computer Vision, Santiago, Chile.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–444. doi:10.1038/nature14539.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. Lawrence Zitnick. 2014. "Microsoft Coco: Common Objects in Context." Paper presented at the European Conference on Computer Vision, Zurich, Switzerland.

Linchant, J., J. Lisein, J. Semeki, P. Lejeune, and C. Vermeulen. 2015. "Are Unmanned Aircraft Systems (UAS S) the Future of Wildlife Monitoring? A Review of Accomplishments and Challenges." *Mammal Review* 45 (4): 239–252. doi:10.1111/mam.12046.

Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, F. Cheng-Yang, and A. C. Berg. 2016. "Ssd: Single Shot Multibox Detector." Paper presented at the European Conference on Computer Vision, Amsterdam, Netherlands.

Marcial-Pablo, M., A. de Jesús, S. Gonzalez-Sanchez, I. Jimenez-Jimenez, R. E. Ontiveros-Capurata, and W. Ojeda-Bustamante. 2019. "Estimation of Vegetation Fraction Using RGB and Multispectral Images from UAV." *International Journal of Remote Sensing* 40 (2): 420–438. doi:10.1080/01431161.2018.1528017.

Nasirahmadi, A., S. A. Edwards, and B. Sturm. 2017. "Implementation of Machine Vision for Detecting Behaviour of Cattle and Pigs." *Livestock Science* 202: 25–38. doi:10.1016/j.livsci.2017.05.014.

Nguyen, H., S. J. Maclagan, T. D. Nguyen, T. Nguyen, P. Flemons, K. Andrews, E. G. Ritchie, and D. Phung. 2017. "Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring." Paper presented at the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, October 19–21.

Nir, O., Y. Parmet, D. Werner, G. Adin, and I. Halachmi. 2018. "3D Computer-vision System for Automatically Estimating Heifer Height and Body Mass." *Biosystems Engineering* 173: 4–10. doi:10.1016/j.biosystemseng.2017.11.014.

Norouzzadeh, M. S., A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune. 2018. "Automatically Identifying, Counting, and Describing Wild Animals in Camera-trap Images with Deep Learning." *Proceedings of the National Academy of Sciences*: 201719367. doi:10.1073/pnas.1719367115

O'Grady, M. J., and G. M. P. O'Hare. 2017. "Modelling the Smart Farm." In *Information Processing in Agriculture. Inf. Process. Agric* 4: 179-187.

Oesterheld, M., O. E. Sala, and S. J. McNaughton. 1992. "Effect of Animal Husbandry on Herbivore-carrying Capacity at a Regional Scale." *Nature* 356 (6366): 234–236. doi:10.1038/356234a0.

Pádua, L., J. Vanko, J. Hruška, J. J. Telmo Adão, E. P. Sousa, and R. Morais. 2017. "UAS, sensors, and Data Processing in Agroforestry: A Review Towards Practical Applications." *International Journal of Remote Sensing* 38 (8–10): 2349–2391. doi:10.1080/01431161.2017.1297548.

Pan, S. J., and Q. Yang. 2010. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering* 22 (10): 1345–1359. doi:10.1109/TKDE.2009.191.

Priyadharshini, G., S. Rajya, S. A. Bhat, M. Vigneswar, and R. Ranjith. 2018. "Convolution Neural Network Based Farm Protection System."

Qazi, K. A., Z. Mehmood, T. Nawaz, and H. A. Habib. 2018. "Goat Flock Surveillance: A Video Analytics Framework Using Quadcopter." *Pakistan Journal of Science* 70 (1): 63–70.

Rahnemoonfar, M., J. Foster, and M. J. Starek. 2017. "Automatic Detection and Counting of Cattle in UAV Imagery Based on Machine Vision Technology (Conference Presentation)." *SPIE Commercial + Scientific Sensing and Imaging: SPIE*, Anaheim, USA, Vol. 10217.

Ray, P. P. 2017. "Internet of Things for Smart Agriculture: Technologies, Practices and Future Direction." *Journal of Ambient Intelligence and Smart Environments* 9 (4): 395–420. doi:10.3233/AIS-170440.

Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. 2016. "You Only Look Once: Unified, Real-time Object Detection." Paper presented at the Proceedings of the IEEE Conference on Computer Vision And Pattern Recognition, Las Vegas, USA.

Ren, S., H. Kaiming, R. Girshick, and J. Sun. 2015. "Faster R-cnn: Towards Real-time Object Detection with Region Proposal Networks." Paper presented at the Advances in Neural Information Processing Systems, Montreal, Canada.

Russell, B. C., A. Torralba, K. P. Murphy, and W. T. Freeman. 2008. "LabelMe: A Database and Web-Based Tool for Image Annotation." *International Journal of Computer Vision* 77 (1–3): 157–173. doi:10.1007/s11263-007-0090-8.

Sa, I., G. Zongyuan, F. Dayoub, B. Upcroft, T. Perez, and M. Chris. 2016. "Deepfruits: A Fruit Detection System Using Deep Neural Networks." *Sensors* 16 (8): 1222. doi:10.3390/s16081222.

Shao, W., R. Kawakami, R. Yoshihashi, S. You, H. Kawase, and T. Naemura. 2020. "Cattle Detection and Counting in UAV Images Based on Convolutional Neural Networks." *International Journal of Remote Sensing* 41 (1): 31–52. doi:10.1080/01431161.2019.1624858.

Singh, J., and S. Shekhar. 2018. "Road Damage Detection And Classification In Smartphone Captured Images Using Mask R-CNN." *Eprint arXiv:1811.04535*: arXiv:1811.04535.

Smith, M. P. L. 1989. "Improving Multispecies Rocky Reef Fish Censuses by Counting Different Groups of Species Using Different Procedures." *Environmental Biology of Fishes* 26 (1): 29–37. doi:10.1007/BF00002473.

Su, X., S. Dong, S. Liu, A. P. Cracknell, Y. Zhang, X. Wang, and G. Liu. 2018. "Using an Unmanned Aerial Vehicle (UAV) to Study Wild Yak in the Highest Desert in the World." *International Journal of Remote Sensing* 39 (15–16): 5490–5503. doi:10.1080/01431161.2018.1441570.

Tabak, M. A., M. S. Norouzzadeh, D. W. Wolfson, S. J. Sweeney, K. C. Vercauteren, N. P. Snow, J. M. Halseth, et al. 2019. "Machine Learning to Classify Animal Species in Camera Trap Images: Applications in Ecology." *Methods in Ecology and Evolution* 10 (4): 585–590. doi:10.1111/2041-210X.13120.

Ter-Sarkisov, A., R. Ross, J. Kelleher, B. Earley, and M. Keane. 2018. "Beef Cattle Instance Segmentation Using Fully Convolutional Neural Network." *arXiv Preprint arXiv*: 1807.01972.

Tian, F., R. Wang, M. Liu, Z. Wang, and F. Li. 2013. "Oestrus Detection and Prediction in Dairy Cows Based on Neural Networks." *Nongye Jixie Xuebao/transactions of the Chinese Society of Agricultural Machinery* 44 (4): 277–281.

Van Hertem, T., A. S. Tello, S. Viazzi, M. Steensels, C. Bahr, C. E. B. Romanini, K. Lokhorst, E. Maltz, I. Halachmi, and D. Berckmans. 2018. "Implementation of an Automatic 3D Vision Monitor for Dairy Cow Locomotion in a Commercial Farm." *Biosystems Engineering* 173: 166–175. doi:10.1016/j.biosystemseng.2017.08.011.

Verma, G. K., and P. Gupta. 2018. "Wild Animal Detection Using Deep Convolutional Neural Network." Paper presented at the Proceedings of 2nd International Conference on Computer Vision & Image Processing Roorkee, Uttarakhand, India.

Viazzi, S. C., A. Bahr, S. Tello, T. Van Hertem, C. E. B. Romanini, A. Pluk, I. Halachmi, C. Lokhorst, and D. Berckmans. 2012. "Analysis of Individual Classification of Lameness Using Automatic Back Posture Measurement in Dairy Cattle." *Journal of Dairy Science* 96. doi:10.3168/jds.2012-5806.

Villa, A., A. S. Salazar, and F. Vargas. 2017. "Towards Automatic Wild Animal Monitoring: Identification of Animal Species in Camera-trap Images Using Very Deep Convolutional Neural Networks." *Ecological Informatics* 41: 24–32. doi:10.1016/j.ecoinf.2017.07.004.

Voulodimos, A. S., C. Z. Patrikakis, A. B. Sideridis, V. A. Ntafis, and E. M. Xylouri. 2010. "A Complete Farm Management System Based on Animal Identification Using RFID Technology." *Computers and Electronics in Agriculture* 70 (2): 380–388. doi:10.1016/j.compag.2009.07.009.

Ward, S., J. Hensler, B. Alsalam, and L. F. Gonzalez. 2016. "Autonomous UAVs Wildlife Detection Using Thermal Imaging, Predictive Navigation and Computer Vision." Paper presented at the Proceedings of the 2016 IEEE Aerospace Conference, Big Sky, Montana, USA.

Windrim, L., M. Bryson, M. McLean, J. Randle, and C. Stone. 2019. "Automated Mapping of Woody Debris over Harvested Forest Plantations Using UAVs, High-Resolution Imagery, and Machine Learning." *Remote Sensing* 11: 733. doi:10.3390/rs11060733.

Xu, B., W. Wang, G. Falzon, P. Kwan, L. Guo, G. Chen, A. Tait, and D. Schneider. 2019. "Counting Cattle: Livestock Management Technology Using a Quadcopter Machine Vision System." *Review of under review*.

Yang, C., and J. H. Everitt. 2010. "Comparison of Hyperspectral Imagery with Aerial Photography and Multispectral Imagery for Mapping Broom Snakeweed." *International Journal of Remote Sensing* 31 (20): 5423–5438. doi:10.1080/01431160903369626.

Yu, X., J. Wang, P. A. Roland Kays, T. W. Jansen, and T. Huang. 2013. "Automated Identification of Animal Species in Camera Trap Images." *EURASIP Journal on Image and Video Processing* 2013 (1): 52. doi:10.1186/1687-5281-2013-52.

Zhu, Y., Y. Chen, L. Zhongqi, S. J. Pan, G.-R. Xue, Y. Yong, and Q. Yang. 2011. "Heterogeneous Transfer Learning for Image Classification." Paper presented at the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, USA.