Systematic Literature Review on the Application of Machine Vision to Outdoor Livestock Monitoring

A systematic literature review on the state of the current research in using machine vision and drones to noninvasively monitor animals in the wild.

ELI-HENRY DYKHNE, STACEY D. SCOTT

School of Computer Science, University of Guelph, Guelph, Ontario, Canada

Abstract:

Advances in machine vision technologies have permeated many aspects of modern society, enabling automated security authentication and intrusion detection, automotive safety features, and everyday conveniences like rapid computer login. Artificial intelligence, especially machine learning, has been recently combined with image processing techniques to improve automated classification and identification of many types of artifacts and even people and animals in certain contexts. This paper presents a systematic literature review of current machine vision approaches used to monitor livestock and wildlife in outdoor habitats. This review was conducted to better understand the feasibility and potential opportunities for using machine vision approaches to support precision livestock farming applications on farms using outdoor habitats. The review found that automated monitoring of livestock and wildlife outdoors using machine vision shows promise but is still in early stages. Somewhat reliable techniques exist for basic animal monitoring capabilities, such as counting animals, but more advanced monitoring capabilities, such as specific animal identification and detection of health and welfare issues are open areas that continue to pose many technical and contextual issues. The review also found that the field lacks exploration of data fusion techniques (e.g., combining camera sources during analysis) that may address some existing technical challenges.

Keywords: outdoor monitoring, livestock, cattle, wildlife, machine vision, CNN, OBIA, thresholding, drone, review

1 INTRODUCTION

Competitive markets and increasing labor costs are forcing farmers to manage larger herds with fewer staff. Accordingly, a worker's attention must now be split over more animals leading to missing cattle and unnoticed health issues.

The ability to monitor large herds autonomously has the potential to bring significant cost savings for farmers with large herds due to reduction in staff costs. The appeal of more frequent monitoring also exists with autonomous methods having the potential to preform multiple checks on massive herds per day rather than one every week. Additionally, the monitoring of livestock with machine vision allows the technology to scale more easily as one does would not need to purchase additional monitoring equipment for each individual animal, in contrast to monitoring via RFID tags and other technologies placed directly on each animal.

1.1 Research Questions

To address the problem above, we conducted a systematic literature review to better understand the current state-of-the-art in outdoor livestock monitoring. The following questions were used to guide and scope this literature review:

i. What challenges have been identified in monitoring livestock outdoors with machine vision?
ii. What techniques have been used to monitor livestock or wildlife outdoors with machine vision?

Special consideration was given during the review to examples of multisource imaging techniques and the potential for data fusion, as early investigations into the problem and general knowledge from the machine vision field show promise in these areas.

Dykhne, E.-H., Scott, S.D. (2020). Systematic Literature Review on the Application of Machine Vision to Outdoor Livestock Monitoring. Report CSL2020-01, School of Computer Science, University of Guelph, Guelph, Canada.

1.2 Purpose of the Review

This review examines and summarizes the work that has been done regarding the use of machine vision to monitor animals in an outdoor, unconstrained environment. That is, monitoring is done while animals move freely in large open-range outdoor areas (e.g., pasture, rangeland, open wilderness). This scope contrasts many existing machine vision techniques used to monitor the health and welfare of livestock in indoor, constrained settings, such as in pens or chutes, that situate livestock in ideal lighting or viewing positions to favour machine vision algorithms (Rainwater-Lovett et al., 2009; Viazzi et al., 2014; Zhao et al., 2018). The review also identifies common problems, current best solutions, and potential research directions to move this research area forward.

1.3 Background

Prior reviews have covered a diverse group of topics. Hollings et al.'s (2018) review, focused on detecting and counting wildlife from remotely sensed imagery, including satellite, uninhabited aerial vehicles (UAVs), and aerial imagery from aircraft. Their paper reviewed and categorized existing methods, results and compared findings from the literature. Relevant findings are discussed in detail in this review.

Brack et al.'s (2018) review focused on the types of errors that current animal counting and animal population estimation algorithms used with UAV-based imagery run into as well as methods to correct them. These techniques are discussed in detail in this review.

Finally, Valletta et al.'s (2017) review focused on a host of machine learning techniques machine learning (ML) techniques currently being used in animal behavior studies. Many of these techniques do not relate to image analysis, but instead involve analysis of other types of data. However, they do mention the use of ML techniques for classification of animal images using random forests and histograms of oriented gradients (HOG's), the second of which is discussed below.

2 KEY CONCEPTS AND TERMS

To help readers understand the technical concepts discussed in the review findings, the following key concepts are defined. Many of these concepts relate to the artificial intelligence (AI) and machine learning (ML) techniques that were uncovered in the review. But others are more general machine vision concepts that are relevant for any image analysis technique applied.

The following four potential outcomes are important to understand for any machine vision approach related to detection or classification of an object (or an animal) in each image.

True Positive (TP): A test result that correctly indicates that the condition being tested for is present (i.e. an animal is detected / classified when an animal is actually in the image).

False Positive (FP): A test result that incorrectly indicates that the condition being tested for is present when it is not (i.e. an animal is detected / classified when there is no animal (or no animal of that type) in the image).

True Negative (TN): A test result that correctly indicates that the condition being tested for is not present (i.e. no animal is detected / classified when no animal (or no animal of that type) is actually in the image).

False Negative (FN): A test result that incorrectly indicates that the condition being tested for is not present when it is (i.e. no animal is detected / classified when an animal is actually in the image).

How well a machine vision algorithm performs relative to the above detection / classification outcomes is often presented and assessed using the following tools and metrics.

Confusion Matrix: A confusion matrix is a table that shows the predicted results (i.e. output) of a detection / classification algorithm compared to what the actual classes are, aggregated by class type.

Precision: Precision is defined by the equation $\frac{TP}{TP+FP}$. As such, having good precision relies on having a low number of false positives.
Recall/Sensitivity: Recall is defined by the equation $\frac{TP}{TP+FN}$. As such, having good recall relies on having a low number of false negatives.
F1-Score: The harmonic mean of precision and sensitivity. It is defined as $2 \times \frac{Percision \times Recall}{Percision + Recall}$.

Two commonly used tools and techniques in machine vision are:
Thresholding: Thresholding is the act of setting cut-offs for some value and accepting everything in the regions you are interested in while rejecting the rest. This technique helps to focus an image analysis on areas that are more likely to contain the desired content. For example, when analyzing thermal imagery, it is common to set a threshold to a certain temperature value and ignore all areas of the image that are above or below that temperature value. In animal contexts, this can help separate “warm” bodies from “cooler” surroundings.
OBIA: Object Based Image Analysis (OBIA) is a technique that works on the idea that it is not enough to simply evaluate what a pixel is by its colour. Instead, it preforms segmentation on an image to divide it into many small portions before classifying those portions based a host of potential features such as spectral signature, shape or size before taking extra user configurable steps such as combining adjacent sectors of the same type, or cross referencing with other types of data such as lidar, infrared, multispectral or even vector based data (Blaschke, 2010). For example, when analyzing a combination of infrared and RGB imagery, one may choose to preform segmentation on the RGB imagery and use both the color and the infrared heat values in the same location to classify each segment as animal or background.

The review found a strong trend of incorporating artificial intelligence, in particular, machine learning, techniques to help improve the detection and classification of animals in outdoor contexts. Core concepts to help understand these findings are provided below.

Machine learning: Computer algorithms that improve automatically through experience, without being explicitly programmed to do a specific task. In the context of livestock detection, it refers to a class of algorithms that optimize themselves based on correct examples of counting to learn what an image needs to look like for us to consider there to be an animal there.

Artificial Neural Network (ANN / NN): An artificial neural network is a machine learning model based on the biological neurons within the human brain. The model includes at least three layers of connected nodes, input, hidden, and output layers. More complex models will have more than one hidden layers (i.e. layers of connected nodes). By determining the difference between the correct output and the output of the model (network), one can automatically optimize the weights (nodes in the hidden layers) within the network to approach the correct answer over a group of training examples (Jain et al., 1996)**.**

Convolutional Neural Network (CNN): CNN’s extract learned features from images, often iteratively and feed them to a fully connected network for localization and classification of objects (LeCun et al., 1999).

Recurrent Neural Network (RNN): Recurrent Neural Networks are used when analyzing sequences such as frames of a video. They work by using part of the output of the network as part of the input for the analysis of the next frame, thus letting the network choose what to remember from old frames (Sherstinsky, 2020).

The following concept is used in many fields, including machine vision, to help account for limitations in relying on any one single data source.

Data Fusion: Data fusion refers to the combination of multiple data sources, often of different types (e.g., lidar, RGB, infrared) in order to produce more accurate and useful information than one source on its own.

Optical flow: Optical flow illustrates the pattern of apparent motion between the observer and objects in visual scene (Fang et al., 2016).
The next subsections provide instructions on how to insert figures, tables, and equations in your document.

3 METHODOLOGY

As this is a systematic literature review, search terms were generated and iterated on to produce as many relevant articles as possible. The specific search terms used in both the Web of Science (WoS) and EBSCO Host databases are provided below:
"(Comput OR Machine OR Algorithm OR Neural OR Processing OR artificial) AND (Aerial OR Outside OR Outdoors OR Remote Sens*) AND (Sensor* OR Detect* OR Monitor OR Camera* OR Drone OR UAV OR UAS) AND (Vision OR Imag*) AND (Livestock OR Bovine OR Pig OR Cow OR Herd OR Animal OR Cattle OR wildlife) AND (farm* or agri* or ecolog* or wildlife) NOT (Crop* OR Plant* OR Bird OR Rodent OR "human health" OR Microb* OR Marine OR Insect)"*

The generated articles were filtered to include papers published in 2005 or later. Afterwards the papers were filtered to manually remove papers with the following exclusion criteria:

- Extended abstracts.
- Papers lacking an English translation.

Papers were also filtered according to the following inclusion criteria:

- Use of machine vision technologies, or those that directly support machine vision.
- Topic must involve livestock or wildlife in an outdoor setting.

The initial search found 136 papers in WoS and 188 papers in EBSCO. After duplicate records were removed, 296 papers remained. These papers were then filtered using these criteria based on their abstracts (70 papers remained), and then again based on their full contents (31 papers remained). After this initial screening, each paper was read. While reviewing these full papers, the "snowball" method was also used to source additional key papers: Papers cited within the reviewed papers which both appeared relevant and passed the aforementioned criteria were added to a list of related papers and also included in the review. This added an additional 37 papers to the selected dataset. The paper selection process is summarized in Figure 1.

This research was completed in the span of a summer work term and thus some older snowball papers may have been left out due to the time constraint.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

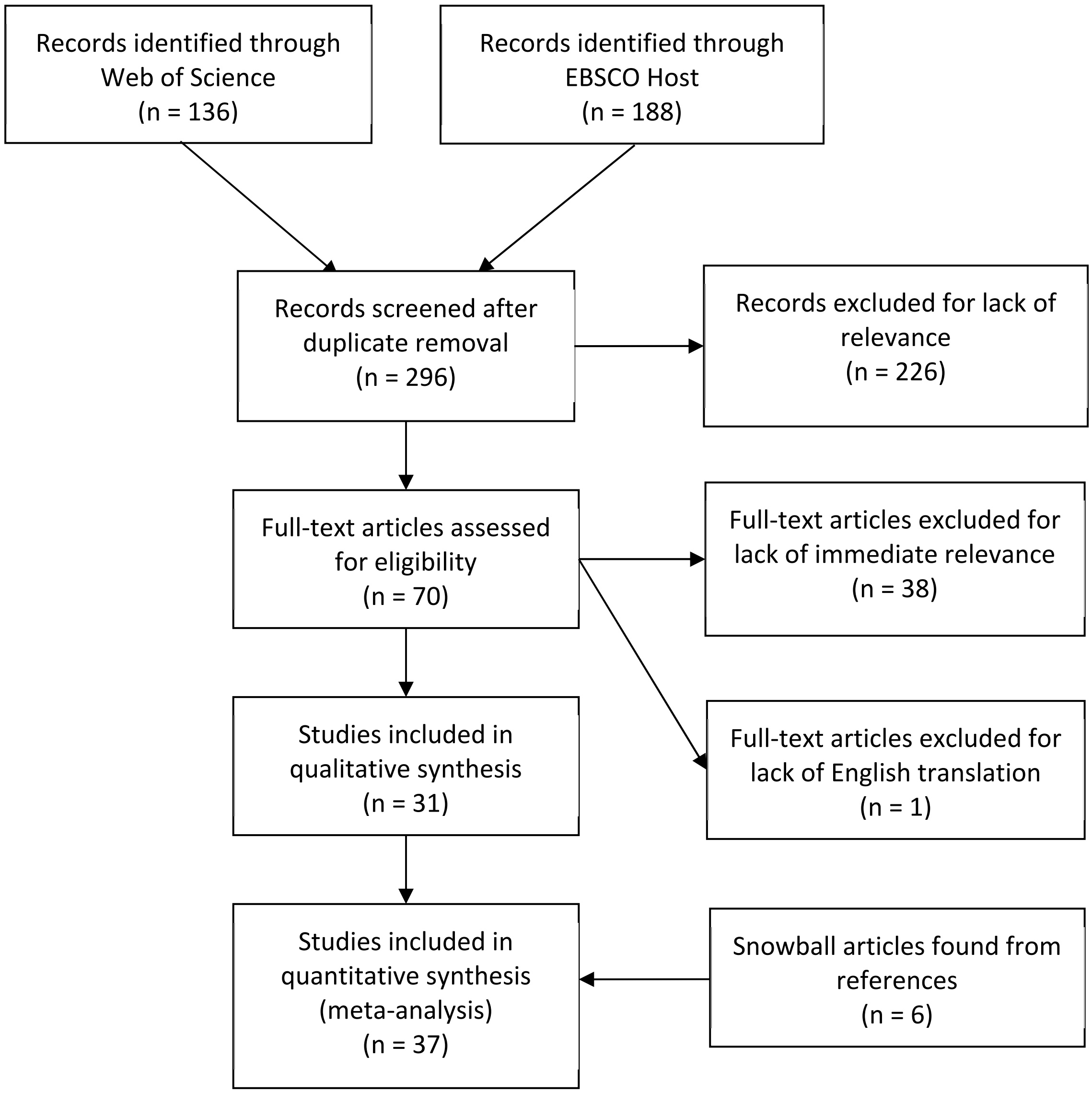


Figure 1. PRISMA chart describing the selection and filtering process that results in the reviewed articles.

4 RESULTS

The animal monitoring approaches found in the review involved the use of different types of imagery sourced from cameras that capture different parts of the light spectrum. Different spectra can be used in different contexts. For instance, infrared wavelengths can be used for thermal detection to monitor animals at night when it is too dark to use the visible spectrum. Different spectra can be combined to provide different information about the same scene as well. The types of imagery used in the reviewed papers are first discussed to provide a understanding of the type of data that the uncovered machine vision techniques are using. Next, the key tools and techniques for monitoring animals outdoors found in the review are presented. These findings are organized by the main vision techniques found in the review. Then, common themes shared across multiple papers, such as data augmentation and data fusion, are overviewed. papers that introduce novel concepts are also briefly discussed. Note that a paper being mentioned in one section does not preclude it from appearing in another as some papers fall into multiple categories when it comes to the type of techniques used.

4.1 Types of Imagery

4.1.1 Visible spectrum (RGB)

Images in the visible spectrum usually comes courtesy of RGB cameras, which have the advantage of being some of the highest resolution per dollar options when compared to infrared, multispectral or time of flight cameras. Taking advantage of all the information in RGB images has been done in the literature through methods that evaluate the image in chunks rather than each pixel in isolation such as OBIA (Barbedo & Koenigkan, 2018; Chrétien et al., 2016) or CNN's (Andrew et al., 2017; Barbedo et al., 2019, 2020; Barbedo & Koenigkan, 2018; Bonneau et al., 2020; Bowley et al., 2017, 2019; Corcoran et al., 2019; Eikelboom et al., 2019; Kellenberger, Marcos, & Tuia, 2018; Kellenberger, Marcos, Courty, et al., 2018; Kellenberger et al., 2017; Sadgrove et al., 2018; Shao et al., 2020; Xu, Wang, Falzon, Kwan, Guo, Sun, et al., 2020). Thresholding has also been attempted on RGB imagery in conjunction with further pre and postprocessing with mixed results (Barbedo et al., 2020; Vayssade et al., 2019).

Techniques using RGB imagery have achieved a wide variety of results depending on the specific detection technique used. Shao et al. (2020) managed to achieve F1 scores of 95.2% when testing on images from the same dataset as the training images and 71.3% when testing on images from a set with a different background than the training set. A different paper working with a wide variety of backgrounds (Kellenberger, Marcos, & Tuia, 2018) and using RGB imagery achieved an F1 score of 60% when recall was set to 70%. Most other papers had results in the same range as these papers.

4.1.2 Infrared

Infrared imaging for counting animals, though less popular lately due to the rise of CNN's, has been used to great effect due to it being able to effectively separate warm-blooded animals from the background via thresholding. Papers using exclusively infrared imaging focused on ways to filter out false positives caused by objects heated by the sun as well as segmentation of herds of animals into individuals (Christiansen et al., 2014; Corcoran et al., 2019; Lhoest et al., 2015; Longmore et al., 2017; Ward et al., 2016).

Longmore et al. (2017) achieved a detection accuracy of approximately 70% with infrared images on cattle and the authors claim that over multiple frames of video, all cows were eventually detected. A separate paper relying on infrared data to count hippos achieved count errors of under 15% percent on four testing images containing approximately 100 hippos each (Lhoest et al., 2015).

A common thread among papers using infrared imagery is the recommendation to take videos early in the morning and late at night as these two times of day would provide the highest thermal contrast between animals and the background.

4.1.3 Multispectral

Multispectral imaging has largely been used to detect vegetation (Barbedo & Koenigkan, 2018; Gil et al., 2014). Much like infrared imagery, the capture devices necessary suffer from high cost and low resolution. Some of the reviewed literature has claimed high spectral fidelity has potential to detect and distinguish animals as different species have different spectral signatures (Terletzky et al., 2012). Little work has been done using multispectral imagery excepting studies that have explored its fusion with panchromatic (visible spectrum) imagery in order to overcome the resolution limitations (Witharana et al., 2016).

4.2 Tools and Techniques for Identification

A Note on Study Results

Directly comparing results and statistics between techniques and studies is difficult due to varied methods of measuring accuracy and categorizing true positives. Moreover, most papers apply their techniques to different datasets, so direct comparison is not feasible.

4.2.1 Thresholding

A common approach in many reviewed papers was to use some form of thresholding to help filter out irrelevant parts of an image. For example, several papers applied thresholding to infrared data where all pixels above a chosen intensity (heat) are kept while the rest are discarded giving algorithms an easy way to filter out parts of the image unlikely to contain animals (because animals are typically warm blooded). Thresholding is a fundamental tool for image processing that is often paired with the others described in this review and is not limited to any particular type of data (Barbedo et al., 2020; Barbedo & Koenigkan, 2018; Fang et al., 2016; Gonzalez et al., 2016; Hollings et al., 2018; Longmore et al., 2017; Vayssade et al., 2019; Ward et al., 2016).

Challenges that come with thresholding involve understanding the best threshold settings since a slight shift in hardware, environment, or target object can render thresholds inaccurate. This is clearly presented in Vayssade et al. (2019), who used different thresholds to deal with different coloured goats and different backgrounds, shown in Figure 2. Other researchers have accounted for lighting and ambient heat variance using dynamically calculated thresholds (Barbedo et al., 2020).

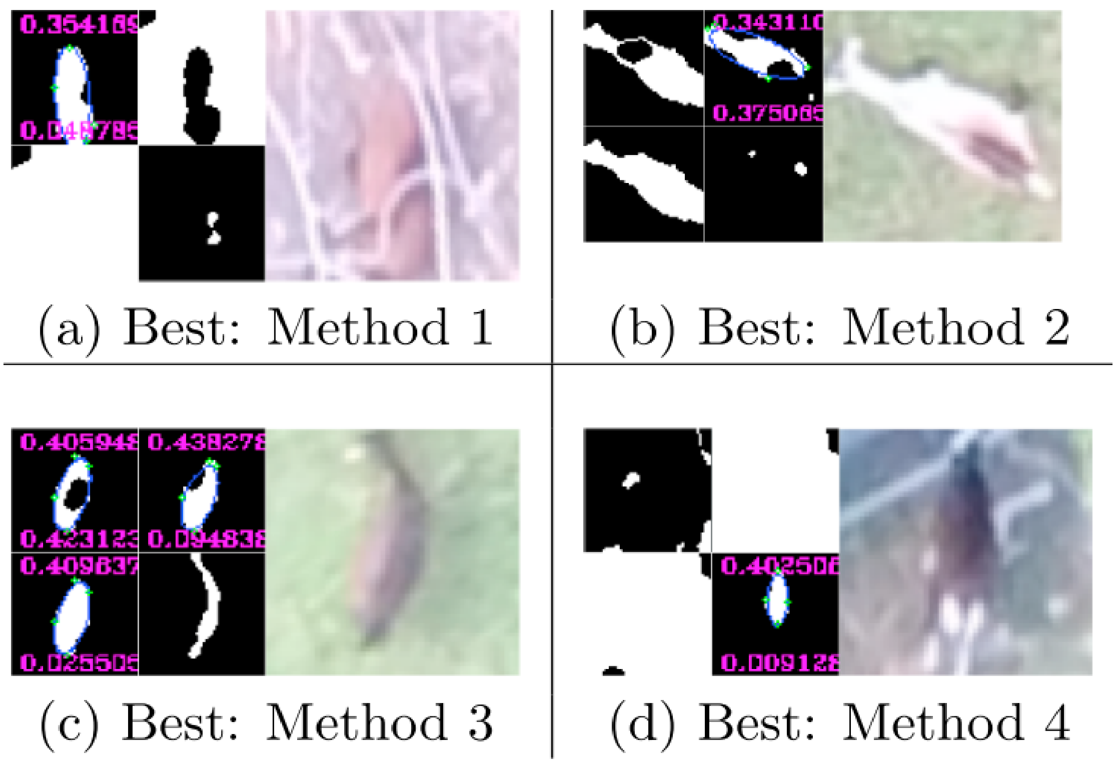


Figure 2. The results of 4 different thresholding methods on 4 different animal images. (Vayssade et al., 2019)

4.2.2 Object Based Image Analysis (OBIA)

OBIA is an expert system (Blaschke, 2010). This means that it is programmed to make decisions using expert knowledge and hand coded set of rules. It is not limited to any one type of data and has shown to be effective in the fusion of data sources (infrared and RGB). When used to count white tailed deer from UAV footage, it outperformed both supervised and unsupervised pixel based approaches by an order of magnitude in terms of accuracy by reducing false positives (Chrétien et al., 2016). Both supervised and unsupervised methods experience overestimations by a factor of 20 to 51 on the various test images. Using OBIA with RGB and infrared data, they achieved perfect detection under the best conditions with no false negatives or false positives, but it should be noted that both the number of test images and the number of animals within each image was under 10. The average detection rate was 50%, with all missed deer being attributed to tree cover, meaning that visible identification was perfect. This result contrasts with work done earlier by Yang et al. (2012) which also compared pixel and OBIA methods for animal detection of migrating mammals in African savannahs. It failed to find a difference in performance of the two algorithms with the results showing a false positive rate of 11%-15% and a false negative rate of 12% for pixel based methods while OBIA achieved similar results with false positive rates of 7%-13% and false negative rates of 13%-16%.

The main drawbacks of OBIA appear to be the need to craft expert rulesets for each type of animal manually, and the long amount of time it takes to process a single image (Hollings et al., 2018).

4.2.3 Convolutional Neural Network (CNN) Based Techniques

When applied to the counting and detection of animals, CNNs have proven effective in their ability to operate on RGB imagery (Andrew et al., 2017; Barbedo et al., 2019, 2020; Bowley et al., 2017, 2019; Chamoso et al., 2014; Eikelboom et al., 2019; Kellenberger et al., 2017; Kellenberger, Marcos, & Tuia, 2018; Rivas et al., 2018; Shao et al., 2020).

Almost all studies in the review using CNN's with the exception of Corcoran et al. (2019) have used pure RGB data, likely to take advantage of networks that have been pretrained on large datasets. These networks have not been trained on infrared data or any combination of RGB and infrared. Pretrained CNNs are those that are trained on existing, usually very large, image datasets, and thus, are already trained to identify many everyday items, and image features. They are not specifically trained for livestock or wildlife applications.

CNN's are often employed in different ways depending on the goals of the study and the precise type of CNN being used.

CNN's have been applied with great success in animal detection. When detecting and classifying multiple species (elephant, zebra, giraffe) with ResNet, (another open source CNN architecture) as described in Eikelboom et al. (2019), they were able to achieve results as presented in the precision/recall graph in Figure 3 with both precision and recall scores of over 70%.

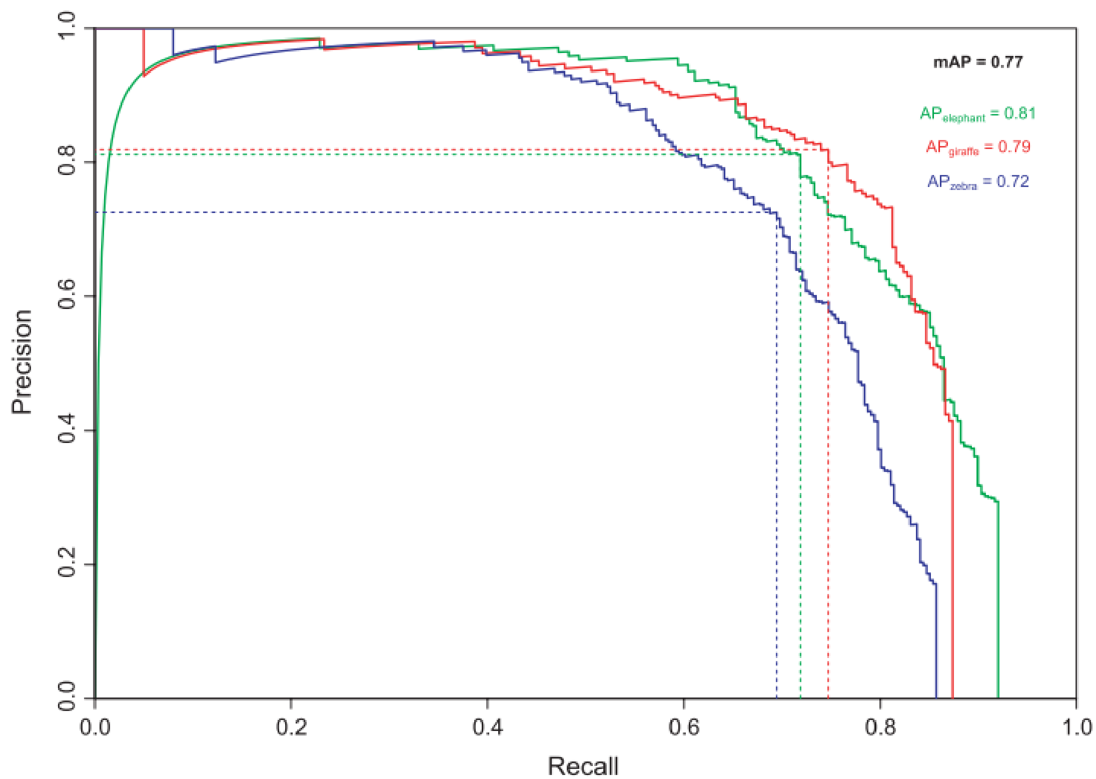


Figure 3. Average precision (AP) Precision/recall curves of the animal detection algorithm for elephant (AP ≈ 0.81), giraffe (AP ≈ 0.79), and zebra (AP ≈ 0.72) on the test set (mean AP ≈ 0.77). The precision/recall combinations with the highest F1-scores (0.76 for elephant, 0.78 for giraffe and 0.71 for zebra) are marked with dashed lines (Eikelboom et al., 2019)

A grid based technique was used by Barbedo et al (2020) where a CNN was trained to identify if there was a part of a cow in a small image. This network was run over a large image in a grid, blacking out squares without cattle parts and leaving the squares with cattle as seen in Figure 4. Once only squares with cattle were identified, colour space manipulations used to enhance contrast and thresholds were applied to generate binary masks. These masks were then fused together, and the resulting images had the now standing out cattle counted via feature matching. When predicting the number of cattle in a cluster, they achieved both precision and recall scores of over 80% for group sizes of 1-8 and 97.4% for both the precision and recall of all cluster sizes.

It should be noted that the environment did not contain other animals and was largely uncluttered and well lit. This provides good contrast and may have made it easier to detect cattle. In cases with sheds and feeders, the CNN would largely ignore them, but parts of the structure would remain when there were animals nearby. This would inevitably cause overcounts through false detections.

Figure 4. The results of the first step of image parsing via CNN to limit the search area (Barbedo et al., 2020)

Other techniques use the activations (how sure the network is that an animal feature is in each part of the image) as input to a bounding box predictor such as in Figure 5. Here, both the activations of the recognition network and the original features work to generate bounding boxes across the image. Often the activations are generated via a sliding window rather than a grid. Using multiple window sizes can be prove useful for detecting animals that vary in size, such as fully grown cows and their calves (Rivas et al., 2018).

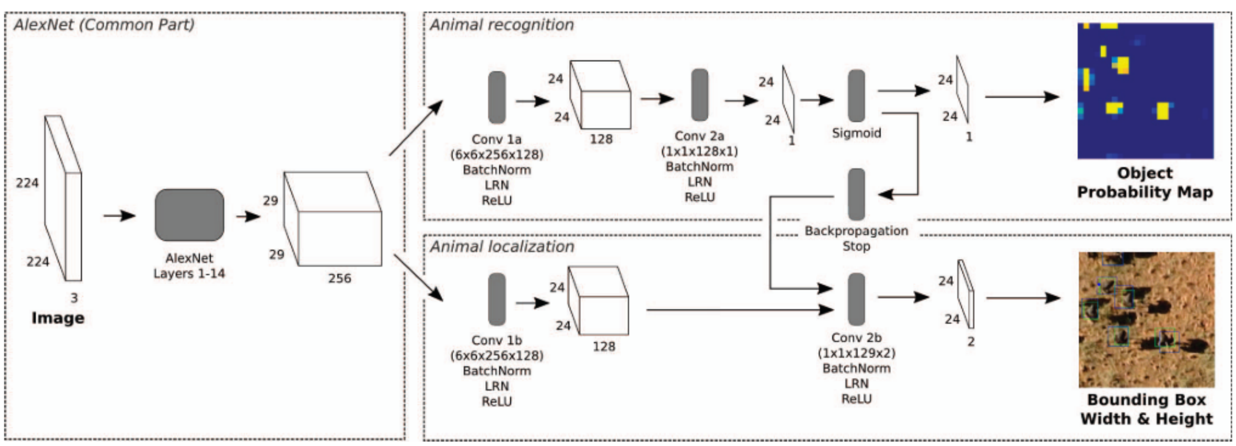


Figure 5. Architecture of the CNN based animal detection model (Kellenberger et al., 2017)

R-CNN's seek to combine the abilities of CNN's to process images with the abilities of RNN's to process sequence information. With this, video can be analyzed which is often more informative than a single frame. An example of how each frame (f) is fed into the network can be seen in Figure 6. Andrew et al. (2017) used this technique to locate cattle and distinguish individual cattle based on their black and white markings with the increased temporal information. For the purpose of detection and localization, the R-CNN was able to achieve well above 95% precision and recall with mistakes cropping up around animals close together. For individual identification, they were able to achieve 98.13% accuracy. It should be noted that the cattle were of the Holstein Friesian variety and their spotting is what made it possible to distinguish them from one another.

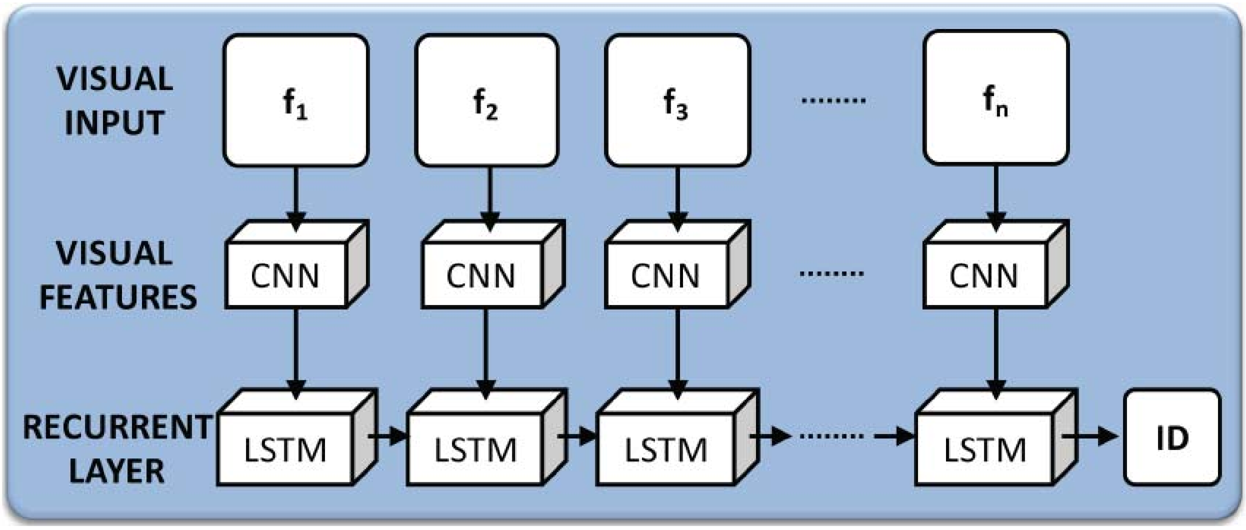


Figure 6. Architecture of recurrent cattle detection network (Andrew et al., 2017)

R-CNN's come in a variety of network architectures, each with their own strengths and weaknesses. The architectures mentioned in this section are open source.

Mask R-CNN (He et al., 2020) is a variant that has shown promising results in the detection and segmentation of animals including cattle and sheep (Xu, Wang, Falzon, Kwan, Guo, Chen, et al., 2020; Xu, Wang, Falzon, Kwan, Guo, Sun, et al., 2020). Unlike other algorithms, Mask R-CNN generates tight masks over individual selection in addition to bounding boxes. The selection of cattle, as well as just the selection of their heads were demonstrated on small groups with a variety of angles and backgrounds including pasture and feedlot. Mask R-CNN showed counting accuracies consistently above 90% for the tasks of appearance detection and head detection. It performed similarly to comparable algorithms such as Faster R-CNN (Ren et al., 2017) and YOLO v3.

Finally, Andrew et al. (2019) used 3 different types of CNN architectures onboard their UAV: A YOLOv2 (Redmon & Farhadi, 2016) based species detector, a dual-stream network to determine the next grid location to explore, and an InceptionV3 (Szegedy et al., 2016) based R-CNN for individual animal identification. The study achieved detection accuracies of 92.4% and individual identification accuracies of 93.6% (although it should be noted that much like their previous study, the cattle were of the Holstein Friesian variety with distinct spots. Additionally, there were only 17 unique cows). For the optimal navigation to locate the maximum number of cattle within a designated area, they found that their algorithm made the optimal direction decision 72.45% of the time.

4.2.4 Invariants

Animals do not always appear in the middle of images, or in the same posture and orientation. Therefore, features that are posture, rotation and translation invariant can be useful for classification of areas highlighted by other means (infrared thresholding, CNN, etc.). Invariants such as thermal contour signatures (Christiansen et al., 2014) and Histograms of Oriented Gradients (HOGs) (Torney et al., 2016; Valletta et al., 2017) have been used with moderate success to classify potential animals. The algorithm implemented by Torney et al. (2016) was able to distinguish between wildebeest and zebra and achieved a precision of 75.15% and recall of 85.83% from greyscale imagery. It was noted that, for the purpose of population estimation, the algorithm achieved a better score than either expert achieved on their first try due to the fact that while its mean absolute error was higher than that of experts, it was not biased to one side (overcounting or undercounting bias), leading to a more accurate final count.

4.2.5 Motion Detection

Motion has been used to detect wildlife in a number of papers (Fang et al., 2016; Y Oishi & Matsunaga, 2010; Yu Oishi & Matsunaga, 2014). As shown in Figure 7, Fang et al. (2016) computed optical flow across a sequence of frames in order to estimate the velocity of each pixel before correcting for drone motion. They then applied thresholds to single out moving pixels and morphological operations to remove noise.

This technique was successfully used to identify zebra and antelope with false positive rates of 2.03% and 12.50% respectively and false negative rates of 17.97% and 11.39%. The high false positive rate for antelope was said to be caused by poor optical flow estimation. The authors do not elaborate further but this may have been caused by their camouflage texture when overlaid on certain types of backgrounds.

The authors note that an advantage of relying on motion is that it takes little account of colour, texture, and the influence of illumination (Fang et al., 2016). As noted by Haalck et al. (2020) which used a separate motion detection/tracking technique, motion analysis is not dependent on colour or wavelength and can thus be applied to any type of video including infrared, multispectral, night vision and more. Additionally, because the technique does not rely on machine learning, it does not require training.

A clear drawback of this approach is the potential to miss still animals. For this reason, work should be considered by combining texture and colour space segmentation for stationary animals. Fang et al. (2016) also mention the effect of parallax makes their method of thresholding less effective for pixels further from the camera.

Other relevant work on the subject of animal motion detection was done by Oishi & Matsunaga (2014) where they extracted features such as trees in order to help match images more easily.

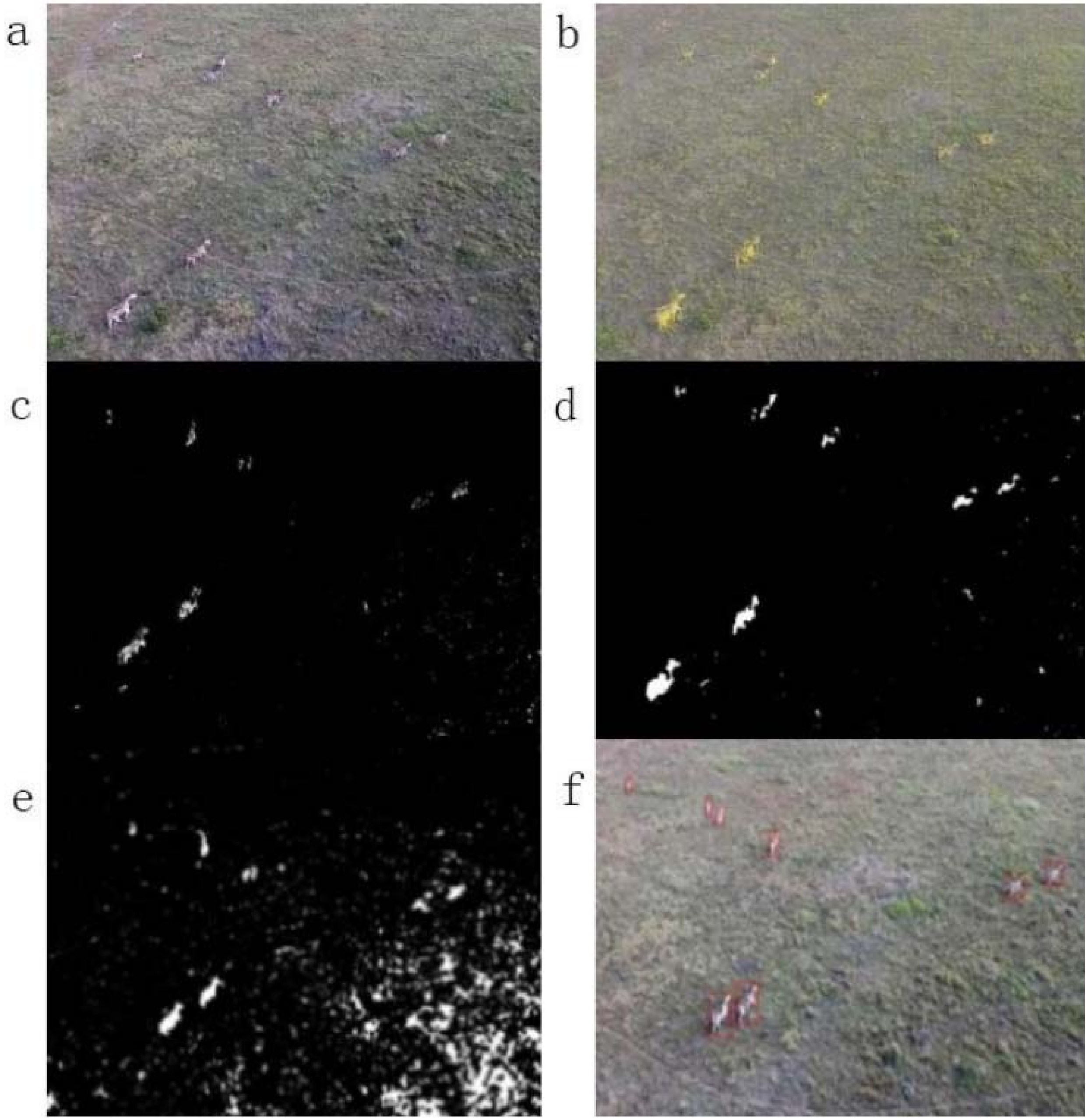


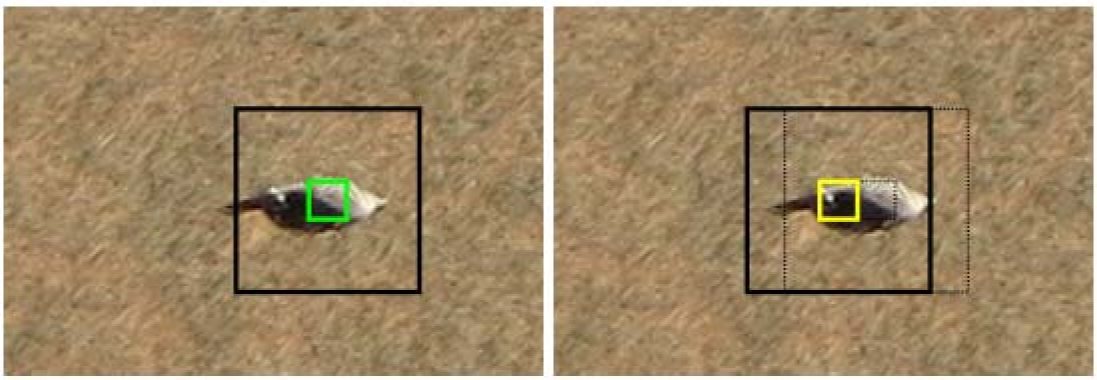
Figure 7. (a) Original video frame (b) Optical flow vectors (c) Binary mask without morphologic operation (d) Binary mask after morphologic operation (e) Segmentation with great errors (f) Detection result. (Fang et al., 2016)

4.3 Miscellaneous Techniques

4.3.1 Border Classes

In the work by Kellenberger, Marcos, & Tuia (2018), it was noted that training a model on patches of images patches instead of single examples had the effect of “spillage” along the spatial boundaries between locations of different classes as shown in Figure 8. Their CNN classified locations into animal and background by considering the location’s neighboring pixels via the model’s receptive field. This may fail when the receptive field is part background and part animal as it should be labeled background to avoid multiple central detections. Training with these sections labeled as background risks causing the CNN to learn that patches influenced by animals through the receptive field belong to the background class. Ignoring the issue in turn causes problems with animal clusters as all sections are predicted to be of the animal class and they cannot be properly separated anymore.

Figure 8. A CNN takes into account multiple neighbors of a specific location (“receptive field”) to determine the likelihood of the center location being an animal (a). To avoid confusion of animals not in the center, but still in the receptive field (b), they introduce a border



(a) Receptive field on an animal (b) Receptive field at the border

class around a true location (Kellenberger, Marcos, & Tuia. 2018).

Their solution involved creating a third border class which allowed those regions to be categorized separately and only the centers of animals to be recorded as true animal detections. This technique showed a significant improvement over an identical model trained without it with significantly higher precision across all recall rates as shown in Figure 9.

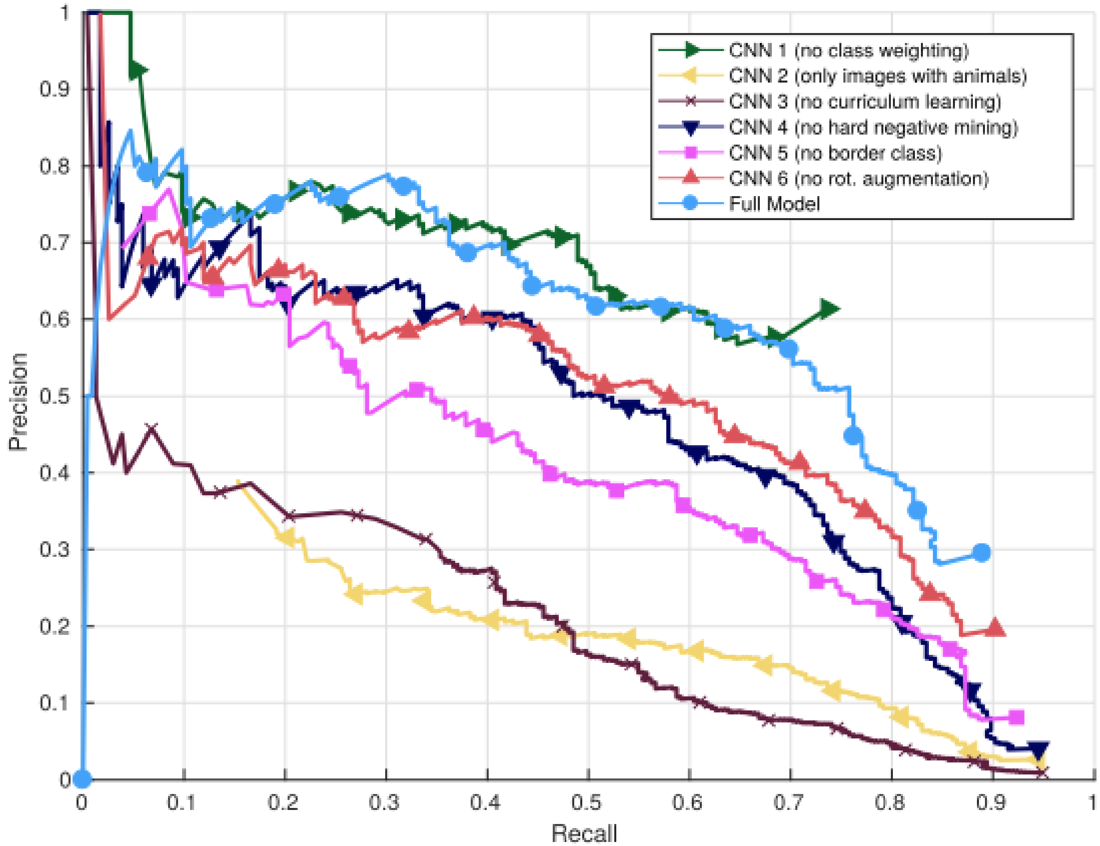


Figure 9. The results of ablation studies to show the difference in performance when individual parts of the training process were changed or removed. Important are the light blue and pink lines which follow the full model and model with no border class respectively. (Kellenberger, Marcos, & Tuia. 2018).

4.4 Data Techniques for Improving Detection and Classification

4.4.1 Examples of Data Fusion

One type of application for fusion was the upsampling of multispectral imagery through fusion with the much higher resolution panspectral imagery taken from a similarly positioned camera. The study was conducted with imagery of Antarctic wildlife gathered from satellite footage.

Various algorithms for data fusion were qualitatively and quantitatively evaluated by a group of experts and by an ensemble of benchmarks respectively for the purpose of counting wildlife (Witharana et al., 2016). The experts evaluated the algorithms visually on how easy they made it to accurately count the animals in the resultant images which were displayed in false colour (false color is used to display images in color which were recorded partially or entirely outside the visible spectrum), while the benchmarks gave scores for spectral and spatial accuracies. The results of the benchmarks are assumed to indicate the best fusion algorithms for automatic counting and classification systems.

Of the seven algorithms evaluated, the High-Pass fusion (Chavez et al., 1991) and University of New Brunswick fusion (Zhang, 2002) algorithms were found to be best for manual wildlife detection while Gram-Schmidt fusion (Laben & Brower, 2000) and University of New Brunswick fusion algorithms were found to be best for automated classification (Witharana et al., 2016).

It should be noted that the authors of the studies themselves state that some of the 14 benchmarks used may be redundant or correlated, and of the 14 benchmarks only 3 of them gave scores for the spatial accuracy while the rest focused on spectral accuracy. Furthermore, no actual wildlife counting tests were done using any type of automated system to verify the validity of the authors interpretations of the benchmarks.

Another example of data fusion was its use in the counting of white tailed deer by taking advantage of both RGB and infrared imagery (Chrétien et al., 2016). The authors experimented with supervised and unsupervised pixel classification, as well as OBIA. Each method was tested with pure RGB, pure infrared and combined data. The pixel classification methods were also tested using the results of Principal Component Analysis (PCA) on the combined data. The results of passing both data sources and using PCA fusion were mixed when it came to the pixel based algorithms, with no general improvement over pure infrared (though significant improvement over pure RGB). When it came to OBIA however, using both data sources resulted in a complete reduction of false positives leading to a perfect detection on the test set.

Important to consider in this study is the small testing set and the small number of animals to be counted in each image.

4.4.2 Data Augmentation

For algorithms that relied on training a detector, (such as CNN's) the training data was the most important part. Machine learning systems often require an enormous amount of training data and acquisition of data for the counting of cattle or various wildlife is time consuming and expensive. To increase the amount of data available for training, studies relied on various methods of data augmentation including rotation, mirror, translation, and the addition of gaussian noise.

Some publications excluded non 90-degree rotations. Kellenberger et al. (2018) stated, "We limit rotation angles to multiples of 90 degrees to avoid spatial shifts of the (reduced size) ground truth due to nearest-neighbor interpolation". Andrew et al. (2017) noted that not all algorithm implementations currently support the parametrization of object rotation leading newly generated orthogonal bounding boxes to include more background pixels. Other papers however have allowed more rotations without adverse effects (Shao et al., 2020).

4.4.3 Citizen Science

Crowdsourcing of labeling has been used to process large amounts of images for wildlife monitoring. Bowley et al. (2019) explored the use of images labeled by untrained citizen scientists in the training of a CNN detector for white phase snow geese in aerial imagery. Blue phase snow geese where considered background. The results of the CNN were then run through a blob detector to amass a final count. The paper compared CNN's trained on expert labels to those trained on citizen labels, as well as those trained on citizen labels that had been matched with those of other citizen scientists to remove false positives.

After images that had mosaicking artifacts were removed, the expert trained CNN resulted in an overcount of 88%. Matched and unmatched citizen science results showed overcounts of 150% and 250% respectively showing that while expert data is still best, matched citizen scientist labels provide a significant improvement over unmatched data and have the potential to increase the size of machine learning datasets for this area of research.

5 DISCUSSION

5.1 Monitoring Trends

Research on the topic of counting wildlife with machine vision in recent years has been highly focused on machine learning, more specifically, on CNN based techniques, mostly with RGB imagery. Efforts made at finding techniques to deal with clusters of animals have achieved moderate success (Barbedo et al., 2020; Xu, Wang, Falzon, Kwan, Guo, Chen, et al., 2020; Xu, Wang, Falzon, Kwan, Guo, Sun, et al., 2020).

5.2 Challenges

The challenges associated with monitoring livestock in an uncontrolled environment are highly related to the data collection aspect that goes into training the models. Papers that focused on large animals rarely had datasets larger than a few hundred animals (Hollings et al., 2018) with at most a few thousand annotations in a few hundred images. To put this in perspective, Luo et al. (2017) who explored both vehicle recognition and face recognition used datasets of nearly 90,000 images and 5000 images respectively for the two tasks.

The small datasets and the large amount of time it takes to generate them creates a barrier for this type of research and casts doubt on how robust some of the generated models are. It should be noted that most datasets collected by the studies only focused on one geographical area and one animal subspecies (Andrew et al., 2017; Bowley et al., 2019; Chrétien et al., 2016). It has been shown by Kellenberger et al. (2018) that the simple act of changing to a new camera, altitude or year could have significant adverse effects on the accuracy of an animal counting model. Their solution (optimal transport) produces a small improvement but still leaves precision and recall very low.

Animal size variation has been shown to be a significant problem with calves being missed by detection algorithms, especially when in close proximity to larger animals (Barbedo et al., 2020). Algorithms tested with grounded, static cameras also had issues detecting young goats (kids), with them making up over half of all false negatives in the paper by Bonneau et al. (2020).

Double counting, or missing animals due to their movement between scans is an issue occasionally brought up. This problem becomes more pressing with low altitude scans due to the increased opportunities for animals to move into already scanned sections from unscanned sections, or vice versa. In the ideal scenario the entire field would be captured in one shot from high above, but this is infeasible given the size of large pastures. On the subject of altitude, much of the imagery is taken below 200m. Papers that experimented with changing the altitude often found detection accuracy would sometimes drop significantly above 60m and often much lower (Barbedo et al., 2019) as shown in Figure 10. It is unclear if a higher resolution camera is the solution to this problem; however, altitude may be limited by transportation regulations in some regions.

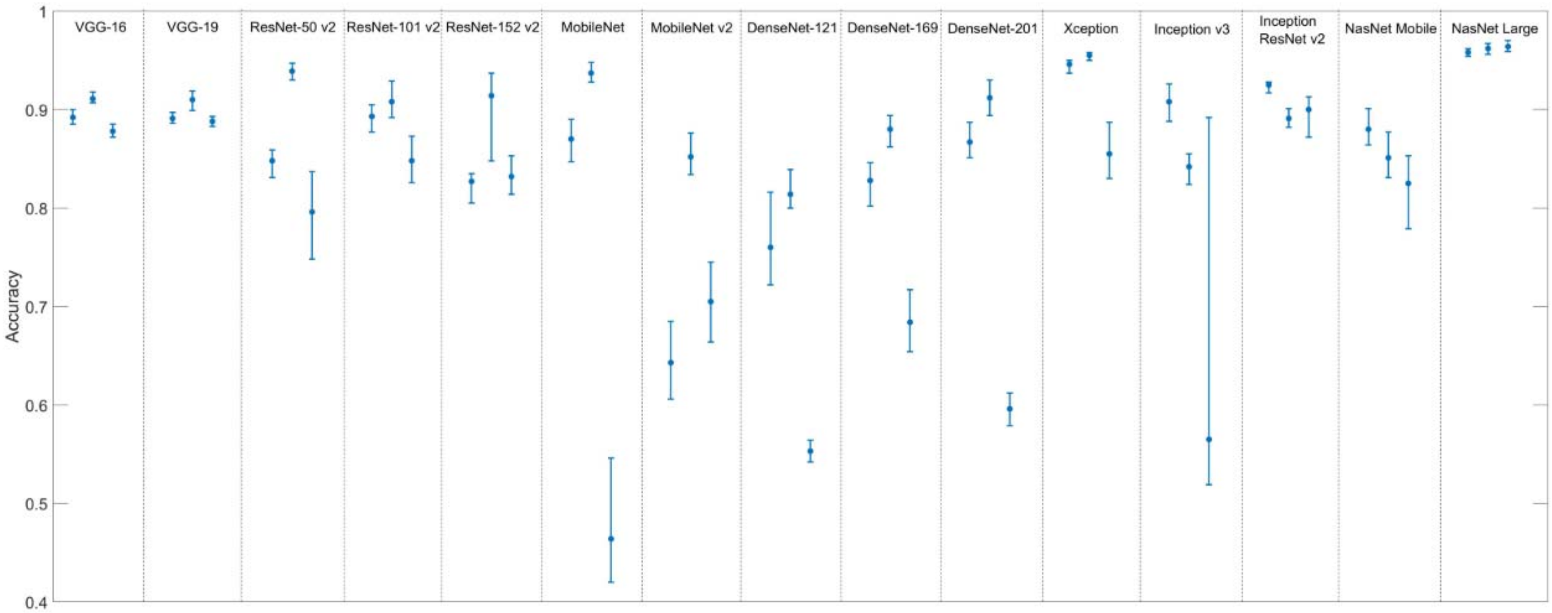


Figure 10. The range of accuracies obtained for each CNN architecture with each of the three bars representing input at scaled resolution to simulate different altitudes. Left to right: 30m, 60m, 120m (Barbedo et al., 2019)

5.3 Directions for Future Research

The fusion of various sources of data appears to be underexplored in this space and may be a possible direction for improving algorithm performance. In addition to improving current algorithms, incorporating more data sources (such as infrared and hyperspectral imagery) may allow algorithms to make new inferences about the health of a particular animal (Barbedo & Koenigkan, 2018). Moreover, the integration of multiple camera sources may be an effective way to deal with availability errors in blind spots.

Health in general has been an underexplored topic in the reviewed papers. The ability to use remote imagery to assess the general health and potential injuries of an animal or an entire herd has not been explored. Research into grazing patterns and herd dynamics could allow UAV imagery to be analyzed to determine if a herd is healthy and stable and relay the information to farmers as an early warning.

6 CONCLUSION

Studies on the automatic monitoring of wildlife via machine vision outdoors have produced a variety of effective techniques. Of these, CNNs seem to be the most effective method when it comes to counting animals. A variety of image sources have been successful including RGB and infrared, but most CNN papers limit their use to RGB for the moment. Current challenges include the cost of data acquisition and labeling, animal size variation, and animal proximity leading to missed detections. The field currently has a lack of exploration on the topic of data fusion and animal health beyond counting. Combinations of camera sources have yet to be explored. These are the key areas to explore for improvements and new research directions.

7 REFERENCES

Andrew, W., Greatwood, C., & Burghardt, T. (2019). Aerial Animal Biometrics: Individual Friesian Cattle Recovery and Visual Identification via an Autonomous UAV with Onboard Deep Inference. *2019 IEEE/RSJ International*

Conference on Intelligent Robots and Systems (IROS). https://doi.org/10.1109/iros40897.2019.8968555

Andrew, W., Greatwood, C., & Burghardt, T. (2017). Visual Localisation and Individual Identification of Holstein Friesian Cattle via Deep Learning. *2017 IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION WORKSHOPS (ICCVW 2017)*, 2850–2859. https://doi.org/10.1109/ICCVW.2017.336

Barbedo, J. G. A., & Koenigkan, L. V. (2018). Perspectives on the use of unmanned aerial systems to monitor cattle. *OUTLOOK ON AGRICULTURE*, *47*(3), 214–222. https://doi.org/10.1177/0030727018781876

Barbedo, J. G. A., Koenigkan, L. V., Santos, P. M., & Ribeiro, A. R. B. (2020). Counting Cattle in UAV Images—Dealing with Clustered Animals and Animal/Background Contrast Changes. *Sensors (14248220)*, *20*(7), 2126. http://10.0.13.62/s20072126

Barbedo, J. G. A., Koenigkan, L. V., Santos, T. T., Santos, P. M. P. M., Arnal Barbedo, J. G., Koenigkan, L. V., Santos, T. T., & Santos, P. M. P. M. (2019). A study on the detection of cattle in UAV images using deep learning. *Sensors (Switzerland)*, *19*(24), 1–14. https://doi.org/10.3390/s19245436

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Bonneau, M., Vayssade, J.-A., Troupe, W., & Arquet, R. (2020). Outdoor animal tracking combining neural network and time-lapse cameras. *Computers & Electronics in Agriculture*, *168*, N.PAG-N.PAG. http://10.0.3.248/j.compag.2019.105150

Bowley, C., Mattingly, M., Barnas, A., Ellis-Felege, S., & Desell, T. (2017). Toward using citizen scientists to drive automated ecological object detection in aerial imagery. *Proceedings - 13th IEEE International Conference on EScience, EScience 2017*, 99–108. https://doi.org/10.1109/eScience.2017.22

Bowley, C., Mattingly, M., Barnas, A., Ellis-Felege, S., & Desell, T. (2019). An analysis of altitude, citizen science and a convolutional neural network feedback loop on object detection in Unmanned Aerial Systems. *JOURNAL OF COMPUTATIONAL SCIENCE*, *34*, 102–116. https://doi.org/10.1016/j.jocs.2019.04.010

Brack, I. V, Kindel, A., & Oliveira, L. F. B. (2018). Detection errors in wildlife abundance estimates from Unmanned Aerial Systems (UAS) surveys: Synthesis, solutions, and challenges. *METHODS IN ECOLOGY AND EVOLUTION*, *9*(8), 1864–1873. https://doi.org/10.1111/2041-210X.13026

Chamoso, P., Raveane, W., Parra, V., & González, A. (2014). UAVs Applied to the Counting and Monitoring of Animals. In *Ambient Intelligence - Software and Applications* (pp. 71–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-07596-9_8

Chavez, P. S., Sides, S. C., & Anderson, J. A. (1991). Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. In *Photogrammetric Engineering & Remote Sensing*.

Chrétien, L.-P. P., Théau, J. J., Ménard, P., Chretien, L.-P., Theau, J., & Menard, P. (2016). Visible and thermal infrared remote sensing for the detection of white-tailed deer using an unmanned aerial system. *Wildlife Society Bulletin*, *40*(1), 181–191. https://doi.org/10.1002/wsb.629

Christiansen, P., Steen, K. A., Jorgensen, R. N., Karstoft, H., Detection, A., & Cameras, T. (2014). Automated Detection and Recognition of Wildlife Using Thermal Cameras. *SENSORS*, *14*(8), 13778–13793. https://doi.org/10.3390/s140813778

Corcoran, E., Denman, S., Hanger, J., Wilson, B., & Hamilton, G. (2019). Automated detection of koalas using low-level aerial surveillance and machine learning. *SCIENTIFIC REPORTS*, *9*. https://doi.org/10.1038/s41598-019-39917-5

Eikelboom, J. A. J. J., Wind, J., van de Ven, E., Kenana, L. M., Schroder, B., de Knegt, H. J., van Langevelde, F., & Prins, H. H. T. T. (2019). Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods in Ecology and Evolution*, *10*(11), 1875–1887. https://doi.org/10.1111/2041-210X.13277

Fang, Y., Du, S., Abdoola, R., Djouani, K., & Richards, C. (2016). Motion Based Animal Detection in Aerial Videos. In S. Patnaik (Ed.), *2ND INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING, COMMUNICATION & CONVERGENCE, ICCC 2016* (Vol. 92, pp. 13–17). https://doi.org/10.1016/j.procs.2016.07.316

Gil, A., Yu, Q., Abadi, M., & Calado, H. (2014). Using aster multispectral imagery for mapping woody invasive species in pico da vara natural reserve (Azores Islands, Portugal). *Revista Árvore*, *38*(3), 391–401. https://doi.org/10.1590/s0100-67622014000300001

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

Haalck, L., Mangan, M., Webb, B., & Risse, B. (2020). Towards image-based animal tracking in natural environments using a freely moving camera. *Journal of Neuroscience Methods*, *330*, N.PAG-N.PAG. http://10.0.3.248/j.jneumeth.2019.108455

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 386–397. https://doi.org/10.1109/tpami.2018.2844175

Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T., & Robinson, A. (2018). How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. *METHODS IN ECOLOGY AND EVOLUTION*, *9*(4), 881–892. https://doi.org/10.1111/2041-210X.12973

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: a tutorial. *Computer*, *29*(3), 31–44. https://doi.org/10.1109/2.485891

Kellenberger, B., Marcos, D., Courty, N., & Tuia, D. (2018). Detecting animals in repeated uav image acquisitions by matching cnn activations with optimal transport. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2018-July*, 3643–3646. https://doi.org/10.1109/IGARSS.2018.8519012

Kellenberger, B., Marcos, D., & Tuia, D. (2018). Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *REMOTE SENSING OF ENVIRONMENT*, *216*, 139–153. https://doi.org/10.1016/j.rse.2018.06.028

Kellenberger, B., Volpi, M., & Tuia, D. (2017). Fast animal detection in UAV images using convolutional neural networks. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2017-July*, 866–869. https://doi.org/10.1109/IGARSS.2017.8127090

Laben, C. ., & Brower, B. . (2000). Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. *United States Patent 6*.

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). Object recognition with gradient-based learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/3-540-46805-6_19

Lhoest, S., Linchant, J., Quevauvillers, S., Vermeulen, C., & Lejeune, P. (2015). HOW MANY HIPPOS (HOMHIP): ALGORITHM FOR AUTOMATIC COUNTS OF ANIMALS WITH INFRA-RED THERMAL IMAGERY FROM UAV. In G. Mallet, C and Paparoditis, N and Dowman, I and Elberink, SO and Raimond, AM and Sithole, G and Rabatel, G and Rottensteiner, F and Briottet, X and Christophe, S and Coltekin, A and Patane (Ed.), *ISPRS GEOSPATIAL WEEK 2015* (Vols. 40–3, Issue W3, pp. 355–362). https://doi.org/10.5194/isprsarchives-XL-3-W3-355-2015

Longmore, S. N., Collins, R. P., Pfeifer, S., Fox, S. E., Mulero-Pazmany, M., Bezombes, F., Goodwin, A., Ovelar, M. D. J., Knapen, J. H., & Wich, S. A. (2017). Adapting astronomical source detection software to help detect animals in thermal images obtained by unmanned aerial systems. *INTERNATIONAL JOURNAL OF REMOTE SENSING*, *38*(8–10), 2623–2638. https://doi.org/10.1080/01431161.2017.1280639

Luo, X., Shen, R., Hu, J., Deng, J., Hu, L., & Guan, Q. (2017). A Deep Convolution Neural Network Model for Vehicle Recognition and Face Recognition. *Procedia Computer Science*, *107*, 715–720. https://doi.org/https://doi.org/10.1016/j.procs.2017.03.153

Oishi, Y, & Matsunaga, T. (2010). SURVEY OF WILD ANIMALS IN THE SNOW IN AIRBORNE REMOTE SENSING IMAGES USING COMPUTER AIDED DETECTION OF THE MOVING WILD ANIMALS ALGORITHM. In S. Kajiwara, K and Muramatsu, K and Soyama, N and Endo, T and Ono, A and Akatsuka (Ed.), *NETWORKING THE WORLD WITH REMOTE SENSING* (Vol. 38, Issue 8, pp. 498–500).

Oishi, Yu, & Matsunaga, T. (2014). Support system for surveying moving wild animals in the snow using aerial remote-sensing images. *International Journal of Remote Sensing*, *35*(4), 1374–1394. https://doi.org/10.1080/01431161.2013.876516

Rainwater-Lovett, K., Pacheco, J. M., Packer, C., & Rodriguez, L. L. (2009). Detection of foot-and-mouth disease virus infected cattle using infrared thermography. The Veterinary Journal, 180(3), 317–324. https://doi.org/10.1016/J.TVJL.2008.01.003

Redmon, J., & Farhadi, A. (2016). YOLO9000: Better, Faster, Stronger. *CoRR*, *abs/1612.0*. http://arxiv.org/abs/1612.08242

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2016.2577031

Rivas, A., Chamoso, P., González-Briones, A., & Corchado, J. (2018). Detection of Cattle Using Drones and Convolutional Neural Networks. *Sensors*, *18*(7), 2048. https://doi.org/10.3390/s18072048

Sadgrove, E. J., Falzon, G., Miron, D., & Lamb, D. W. (2018). Real-time object detection in agricultural/remote environments using the multiple-expert colour feature extreme learning machine (MEC-ELM). *COMPUTERS IN INDUSTRY*, *98*, 183–191. https://doi.org/10.1016/j.compind.2018.03.014

Shao, W., Kawakami, R., Yoshihashi, R., You, S., Kawase, H., & Naemura, T. (2020). Cattle detection and counting in UAV images based on convolutional neural networks. *International Journal of Remote Sensing*, *41*(1), 31–52. http://10.0.4.56/01431161.2019.1624858

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306. https://doi.org/10.1016/j.physd.2019.132306

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2016.308

Terletzky, P., Ramsey, R. D., & Neale, C. M. U. (2012). Spectral Characteristics of Domestic and Wild Mammals. *GIScience & Remote Sensing*, *49*(4), 597–608. https://doi.org/10.2747/1548-1603.49.4.597

Torney, C. J., Dobson, A. P., Borner, F., Lloyd-Jones, D. J., Moyer, D., Maliti, H. T., Mwita, M., Fredrick, H., Borner, M., & Hopcraft, J. G. C. (2016). Assessing Rotation-Invariant Feature Classification for Automated Wildebeest Population Counts. *PLOS ONE*, *11*(5). https://doi.org/10.1371/journal.pone.0156342

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. https://doi.org/10.1016/j.anbehav.2016.12.005

Vayssade, J.-A., Arquet, R., & Bonneau, M. (2019). Automatic activity tracking of goats using drone camera. *COMPUTERS AND ELECTRONICS IN AGRICULTURE*, *162*, 767–772. https://doi.org/10.1016/j.compag.2019.05.021

Viazzi, S., Bahr, C., Van Hertem, T., Schlageter-Tello, A., Romanini, C. E. B., Halachmi, I., Lokhorst, C., & Berckmans, D. (2014). Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. Computers and Electronics in Agriculture, 100, 139–147. https://doi.org/10.1016/J.COMPAG.2013.11.005

Ward, S., Hensler, J., Alsalam, B., & Gonzalez, L. F. (2016). Autonomous UAVs Wildlife Detection Using Thermal Imaging, Predictive Navigation and Computer Vision. *2016 IEEE AEROSPACE CONFERENCE*.

Witharana, C., LaRue, M. A., & Lynch, H. J. (2016). Benchmarking of data fusion algorithms in support of earth observation based Antarctic wildlife monitoring. *ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING*, *113*, 124–143. https://doi.org/10.1016/j.isprsjprs.2015.12.009

Wu, C. (2013). Towards linear-time incremental structure from motion. *Proceedings - 2013 International Conference on 3D Vision, 3DV 2013*. https://doi.org/10.1109/3DV.2013.25

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., & Schneider, D. (2020). Automated cattle counting using Mask R-CNN in quadcopter vision system. *Computers and Electronics in Agriculture*, *171*, 105300. https://doi.org/10.1016/j.compag.2020.105300

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Sun, Z., & Li, C. (2020). Livestock classification and counting in quadcopter aerial images using Mask R-CNN. *INTERNATIONAL JOURNAL OF REMOTE SENSING*. https://doi.org/10.1080/01431161.2020.1734245

Yang, Z., Wang, T., Skidmore, A. K., & Weir, M. J. (2012). *Evaluating high resolution GeoEye-1 satellite imagery for mapping wildlife in open savannahs*.

Zhang, Y. (2002). A new automatic approach for effectively fusing Landsat 7 as well as IKONOS images. *International Geoscience and Remote Sensing Symposium (IGARSS)*. https://doi.org/10.1109/igarss.2002.1026567

Zhao, K., Bewley, J. M., He, D., & Jin, X. (2018). Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique. Computers and Electronics in Agriculture, 148, 226–236. https://doi.org/10.1016/J.COMPAG.2018.03.014