IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 May 2022, accepted 6 July 2022, date of publication 28 July 2022, date of current version 4 August 2022.
Digital Object Identifier 10.1109/ACCESS.2022.3194507

TOPICAL REVIEW

A Systematic Literature Review on the Use of Deep Learning in Precision Livestock Detection and Localization Using Unmanned Aerial Vehicles

D. B. MAMEHGOL YOUSEFI[1], A. S. MOHD RAFIE[1], S. A. R. AL-HADDAD[2], (Senior Member, IEEE), AND SYARIL AZRAD[1]
[1]Department of Aerospace Engineering, Universiti Putra Malaysia, Seri Kembangan, Selangor 43400, Malaysia
[2]Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia, Seri Kembangan, Selangor 43400, Malaysia

Corresponding authors: D. B. Mamehgol Yousefi (mamehgol.yousefi@gmail.com) and Syaril Azrad (syaril@upm.edu.my)

This work was supported by the Ministry of Education, Malaysia, through the Fundamental Research Grant Scheme (FRGS) under Grant FRGS/1/2020/TK0/UPM/02/33.

ABSTRACT With the ever-increasing importance of dairy and meat production, precision livestock farming (PLF) using advanced information technologies is emerging to improve farming production systems. The latest automation, connectivity, and artificial intelligence developments open new horizons to monitor livestock in the pasture, controlled environments, and open environments. Due to the significance of livestock detection and tracking, this systematic review extracts and summarizes the existing deep learning (DL) techniques in PLF using unmanned aerial vehicles (UAV). In the context of livestock recognition studies, UAVs are receiving growing attention due to their flexible data acquisition and operation in different conditions. This review examines the implemented DL architectures and scrutinizes the broadly exploited evaluation metrics, attributes, and databases. The classification of most UAV livestock monitoring systems using DL techniques is in three categories: detection, classification, and localization. Correspondingly, this paper discusses the future benefits and drawbacks of these DL-based PLF approaches using UAV imagery. Additionally, this paper describes alternative methods used to mitigate issues in PLF. The aim of this work is to provide insights into the most relevant studies on the development of UAV-based PLF systems focused on deep neural network-based techniques.

INDEX TERMS Livestock monitoring, deep learning, detection, localization, UAVs.

I. INTRODUCTION

The livestock sector is one of the fastest-growing sectors supporting the food safety of 1.3 billion people and contributing 40% of the global value of agricultural output. Protecting livestock is a practical risk mitigation approach for small communities and a rapidly growing demand for livestock products [1]. To output reliable, healthy, cost-effective, welfare, and environmentally safe dairy products in an increasingly complex international agricultural market, timely production-related data must be accessible to livestock producers. Various information communication technologies (ICT) have helped the continuous expansion of precision livestock farming (PLF), through which the livestock sector has implemented welfare breeding to meet the quality of livestock products [2]. Long-term livestock monitoring can produce meaningful information for researchers and engineers developing PLF technologies. The general information acquisition on several farm animals is the visual data that is effective but expensive and tedious. Other existing systems tag every animal using either Radio-Frequency Identification (RFID) [3], [4], ear tags [5], and global positioning system (GPS)

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Chen.

collars [6]. These tags identify and pinpoint the location of livestock, thereby avoiding extortion. However, these technologies' implementation and setup are constrained by high power consumption, physical size, cost, and local communication networks, specifically in vast geographic ranges and remote environments [7]. Other implemented monitoring methods like camera traps [8], thermal cameras [9], and surveillance cameras typically require significant time and investment. Also, these devices are expensive to implement and have flexibility issues of ranges and disturbance from surroundings. The image sequences obtained from the above methods comprise highly cluttered images that result in low detection rates.

In recent years, object detection and localization architectures based on deep learning (DL) techniques have found extensive applications to resolve the issues with the methods mentioned above. For example, using a camera trap database, Verma *et al.* [10] proposed an animal detection model based on a self-learned Deep Convolutional Neural Network (DCNN). The work done by [11] found that Convolutional Neural Networks (CNNs) are used widely for livestock identification. However, the current few commercialized PLF techniques depend on integrating other advanced information and sensory technologies.

The accessibility of UAVs offers a feasible solution to solve these challenges by reducing cost owing to extended endurance and practical flight planning autonomously almost everywhere [12]. The advancement of new technologies like UAVs and artificial intelligence techniques have alluded to promising results to support livestock farming. Livestock precision farming exploits these techniques for different purposes, such as the animals' identification and counting. The efficient monitoring of livestock welfare demonstrated by industrious farming techniques has shown the effective implementation of UAVs for constant real-time data (e.g., image, video) acquisition. However, visual species detection is challenging, and counting in real-world scenarios is difficult. This difficulty is due to imaging the animal concerning the environment. The differences in illumination, the livestock's natural camouflage, overlapping vegetation, other animals, and terrain obstacles can hinder machine vision systems. Despite these limitations, the recent advancement of DL-based object detection [13], [14] encourages animal detection and classification solutions.

Furthermore, recent studies extended to extract the features of acoustic sounds using audio-visual learning techniques. In this technique, the audio and visual modalities are introduced to overcome the limits of perception tasks in each modality. The audio and visual modalities represented by electrical voltage and RGB colour space are designed to be perceived by humans. However, the major challenge is to learn the mapping of the audio and vision and find constraints. Therefore, unsupervised learning methods provide a well-perceived solution while finding correct implicit supervision [137]. A perceived audio-visual learning system processes information captured by UAVs to recognize and identify livestock precisely and automatically essential for livestock population, behaviour, and health monitoring. Recently, [15] employed a UAV with onboard radar motion sensors to identify COVID-19 patients from their breathing patterns. This exploratory work has the potential for life health sensing applications by sensing variations in breathing rate and depth in COVID-19 patients.

Different machine learning (ML) methods have historically played a central role in agricultural and farming remote sensing research. Garcia *et al.* [16] conducted a review study assessing the current research on ML application in PLF, aiming only at animal health and grazing issues. A significant challenge with commonly applied ML methods (e.g., k nearest neighbour (kNN), random forest (RF), support vector machine (SVM)) is feature engineering (i.e., feature extraction, labelling, and determining the relevant features). In another systematic literature review by [17] for livestock body weight measurement, linear regression was the most applied algorithm, and DL application is minimal. However, UAVs have been extensively deployed in agricultural remote sensing over the last decade, leading to an increasing amount of UAV-based data.

In contrast with the ML models, DL techniques have often found applications with larger datasets. In particular, the UAV-based data (e.g., images) usually provide richer spatiotemporal-spectral information; they have more complicated and diverse patterns, thus imposing higher requirements on the processing ways of remotely sensed data [18]. The DL's strong ability in feature representation, multi-layer learning, and great superiorities in multilevel and multi-scale feature extraction contribute to high performance in image processing and classification problems [19]. DL-based livestock detection and localization received much attention in the PLF field due to their ability to extract high-quality features from raw data. As a result, long-term livestock monitoring can produce meaningful information for both researchers and engineers developing PLF technologies. Most publications centre on monitoring the animals using classical ML classification techniques, but more work is needed to explore the DL technologies in this domain.

The above discussion on the several published informative and original articles on livestock detection and localization reveals that a comprehensive review article is needed to address the lack of existing methods. In this sense, the most important contributions of this research contributions are as follows:

- Describe the fundamental principles of DL-based prediction algorithms for object recognition.
- Discuss the current state of deep neural networks (DNNs) in PLF and give a broad description of the extensively applied evaluation metrics and UAV-based databases.
- Provide an extensive critical review of the DNN algorithms in UAV-based livestock monitoring; examination of used features in PLF; comparable assessment of the relevant documents.

- Give an exhaustive analysis of DL algorithms in UAV-based livestock farming, classified into three categories of (1) detection, (2) classification, and (3) localization.
- Expand on the prospective research scopes on DL-based livestock detection and localization from UAV imagery. The motivation here is to discover and present expected difficulties and opportunities.
- To capture the various features and DL techniques in the PLF to support and advance the current adaptation of DL-based livestock detection, classification, and localization. algorithms in the livestock production industry.

This review is a precursor to the design of an intelligent system to detect, classify, and localize a target. Hence, the following sections scrutinize those methods employed to achieve that objective.

II. METHODOLOGY

In this context, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [20] is used to determine the current status of PLF relevant to computer vision. With a specific focus on current DL methods used in UAV-based PLF and identify the potential gaps for the subsequent researchers. The PRISMA comprises five phases:

1. Study guides and scope of the study
2. Identification of the need for vision (research questions)
3. Article selection process
4. Performing the review
5. Deficiencies and new perspectives in the knowledge of the problem

The first phase describes the research objectives and the need for a comprehensive review in this domain. The next step specifies the questions that lead the research and the research methods; the article selection method (i.e., the criteria for documents identification, screening, exclusion, and inclusion). Further, performing a comprehensive review of the included documents to answer the research questions. Lastly, to present the state of the knowledge, deficiencies, and future opportunities.

A. RESEARCH OBJECTIVES

There have been several published review articles on the computer vision-based PLF. These surveys on PLF systems mainly focused on a few aspects of animal behaviour monitoring (grazing or health) using sensor data. A review by [21] investigated existing knowledge about livestock identification and counting-based computer vision techniques using data captured by UAVs. Table 1 lists the recent related review papers; it shows that a comprehensive review article is required to address the lack of current studies. This paper distinguishes itself from other works as it focuses on various livestock detection and localization approaches based on DL algorithms using different data types captured by UAVs.

TABLE 1. Evaluation of the existing review articles on DL-based livestock farming.

Reference	Objectives	Limitations of the review
[17]	Computer vision-based livestock body weight estimation techniques, the features of models, underlying approaches, performance evaluation metrics, challenges, and the recommendation for future research.	This study is limited to the livestock body weight estimation and avoids UAV and DL methods
[16]	Listed the potential applications of ML algorithms in precision livestock farming, particularly for the analysis of grazing and animal health, and the main form of data acquisition in PLF	This review lacks the critical evaluation of DL algorithms for different PLF applications, their ability, and the potential of UAVs for data acquisition.
[21]	It presents the computational vision technologies and tools used for livestock identification and counting using data captured by UAVs	This review lacks the critical evaluation, existing research gaps, and future search opportunities
[22]	This review presents the livestock farming digitization research focusing on biometric sensors, big data, and blockchain technologies in PLF	However, it lacks the ML algorithms, automation of livestock activities, data acquisition
[23]	This review summarized the precision cattle farming technologies focusing on identification, body condition score evaluation, and live weight estimation	This review lacks the critical examination of DL algorithms for a wide range of livestock identification tasks, especially those performed using UAVs
[24]	It presents a detailed analysis of the existing hardware-and software-based livestock counting tasks, the UAV-based systems compared to traditional methods of monitoring, and object detection algorithms from images and videos	This review lacks the critical evaluation of the DL algorithm's performance using data captured by UAVs for livestock detection and counting, analysis of commonly used features, and data

B. RESEARCH QUESTIONS

The research is performed by including the outlined concepts related to the outlook of this review. The article uses the selection process of preferred reporting items for PRISMA methods demonstrated in Figure 1.

The most well-established search engines, namely Google Scholar, IEEE explorer, and ScienceDirect, were exploited to extract the matched articles. We filtered our search using the following strings: Keywords= (livestock detection, Deep Learning, precision farming, classification, localization, and UAV). Ultimately, we searched these keywords together or in combination with other keywords comprising ''audio-visual'', ''features and databases,'' ''livestock detection'', ''photogrammetry or sensors'', ''animals or wildlife'', and ''remote sensing''. The restriction of the search results is those that qualified the following criteria in the reviewing process.

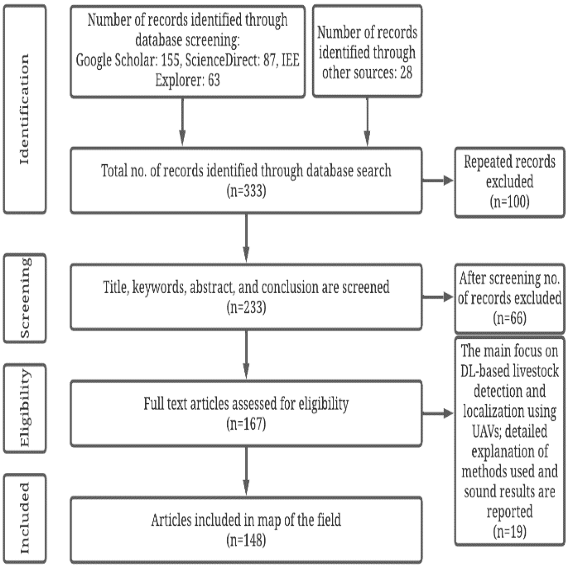

FIGURE 1. **Flow diagram of article selection setup.**

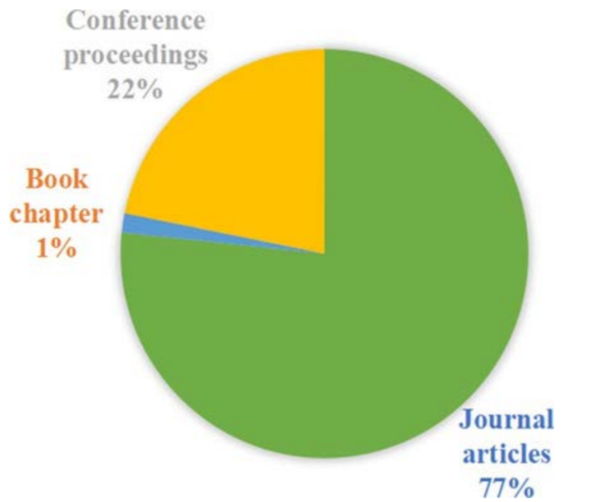

FIGURE 2. **Number of journal articles, conference proceedings, and book chapters.**

- Documents published in peer-reviewed journals, peer-reviewed conferences, review articles, book chapters, and research articles from computer science and engineering organizations
- Documents published in English
- Documents published between 2011 and 2021 (both years inclusive)

The first round of review involved inclusion criteria based on the title, abstract, conclusion, and keywords of a given document. Then further the examination of papers that satisfied the scope of this study, i.e., articles that investigated DL-based livestock detection and localization solutions applied to the precision farming environment and livestock industry. Finally, the documents were separated and reviewed concerning their applications in UAV-based livestock detection, classification, and localization in DL techniques. After restricting search papers, it resulted in a total number of screened 148 documents. Consequently, the following information led to dataset-related information, extensively examined feature sets, and evaluation metrics, and each source drives prediction algorithms. Considering the total number of included articles, those classified as journal articles comprise 77%, conference proceedings were 22%, and book chapters only 1%, as outlined in Figure 2.

III. FUNDAMENTAL OF DEEP LEARNING-BASED OBJECT RECOGNITION

Deep learning object recognition approaches comprise several phases, namely data collection, learning networks, and output. Well-established object detection techniques based on manually extracted features and shallow trainable architectures performance deteriorates by building deep ensembles with complex contexts from scene classifiers and object detectors. For instance, feature-detection algorithms, including speeded-up robust features (SURF) [25] or scale-invariant feature transform (SIFT) [26], are applied on PLF. However, the biggest drawback of computational complexity associated with costly computations needed for SIFT feature calculation and matching limits their application. The development of more powerful DNN object detection methods capable of learning semantic and deep features resolved the drawbacks of conventional methods. Figure 3 shows the architecture of the DNN object recognition structure.

Recent results in CNNs have shown them outperform computer vision methods (e.g., classical classifiers) in object detection [28]. CNN-based feature extraction solutions are developing extensively, especially in PLF detection, classification, and localization issues, since these networks can input the original images directly with lower pre-processing complexity. The CNN-based architecture contains at least one layer of convolution to process input images. This layer comprises several feature maps designed as a plane with equal-weighted neurons on each plane [29]. Next, the pooling layer minimizes the number of training variables while keeping valuable knowledge regarding the input through a down-sampling operation.

Consequently, this leads to a final prediction output, such as classification generated for a fully connected layer [30]. The fully configured CNN integrates the series of base layers, multiple filters in each layer, and other tunning parameters selected by the network architecture [31]. Essentially, CNNs combine base layers and networks, each of which demonstrates specific application scenes and characteristics.

Livestock DL-based detection, classification, and localization have been recognized over the years to establish recognition and localization techniques with the ever-increasing diverse sets of records. For this purpose, two categories of DL architectures consisting of one-stage, and two-stage detectors, have been assessed. Examples of the one-stage

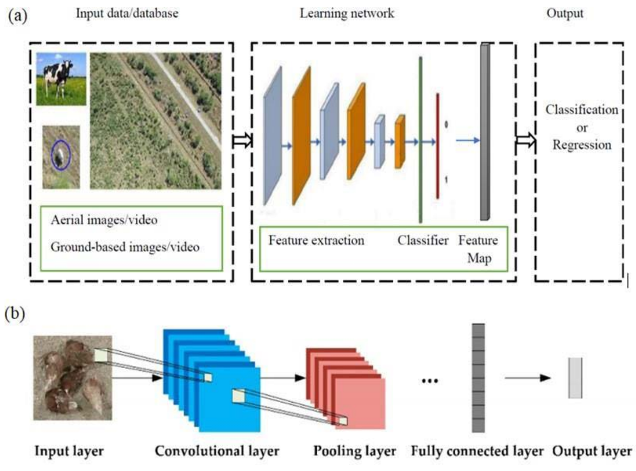

FIGURE 3. The general framework of (a) deep neural network and (b) convolutional neural network-based object recognition [27].

detector include "You Only Look Once" (YOLO) [32] and a single-shot multi-box detector (SSD) [13] that predicts the bounding boxes and their probabilities. In comparison, the two-stage detectors such as CNN [33], CRNN [34], CNN, and region-based convolutional neural networks (R-CNN) use regional proposal networks. The former produce class-agnostic regions from input images before bounding boxes classification and regression. While the latter is fast with reduced elements for object detection, classification, and localization tasks without a region proposal classification, they tend to achieve lower accuracy when dealing with small objects. Moreover, they are computational time and cost-intensive for precise image annotation.

Despite the differences between the two object detectors (one or two-stage) categories, both deal with a class imbalance issue that can reduce training accuracy [35]. In addition, the choice of architecture could depend on multiple features (e.g., training methodology, speed, and inference time), and thus there is no determined winner. Breed recognition tasks with imbalanced datasets used the YOLOv3 [36], RetinaNet [37], and Faster R-CNN [14]. Redmon *et al.* [38] introduced the YOLO to perform classification and localization tasks simultaneously in its network by estimating the probability distribution over image grids, the bounding boxes, and the confidence scores of every grid.

DL-based modelling for livestock detection includes four main techniques [39]:

1. Semantic segmentation focuses on labelling every pixel belonging to a particular class
2. Instance segmentation performs object detection and localization by creating segmentation maps for every detected instance.
3. Detection that defines the object of the interest bounding box
4. Heat mapping or probability distribution using CNNs that display the position of the herds in an image

CNNs' ability to recognize patterns is essential for livestock detection and localization applications, where livestock exhibit immense visual diversity due to their colour, species, posture changes, and other external conditions like blurriness effect due to motion and sensor changes in various illumination backgrounds. Hence, most object detection approaches create bounding boxes for every object and then classify them. For instance, the R-CNN combines the two functions of candidate object localization and object proposals classification through region proposals networks (RPNs) [40]. In Faster R-CNN, RPNs and Fast R-CNN extract high-level features from the images. It combines the bounding boxes and classification to achieve multi-task learning for classification and regression problems [41]. These methods aim to characterize the object of interest at the pixel-wise structure. Semantic segmentation techniques developed from object detection distinguish the class of objects at the pixel level by drawing regions around the boundary of an object. However, this technique cannot distinguish objects in the same class as every pixel is labelled uniquely [42]. The CNN variant, fully convolutional networks (FCN), transforms image pixels into pixel categories extensively used in semantic segmentation models to hold the position data discarded in the pooling layer. This method can control pixel labelling in dense images used in livestock recognition.

On the other hand, blurred hidden features cause the network to ignore the specifics of an image and fail to distinguish the link between the local and the whole. Most remote sensing applications have benefited from pixel-wise semantic segmentation and object detection algorithms like DNN networks. However, these approaches require large, labelled datasets to draw multiple objects bounding boxes in an image, unlike scene-wise classification requiring only the class labels' annotation. Henceforth, the Faster R-CNN model with the RPN shares convolutional features with the detection network was proposed [43]. The authors of [44] used A CNN-based algorithm to process an immense amount of image data, comprising objects' size, location, and posture. In dense environments, most studies utilized CNN as the framework [45]. Overall, image recognition and computer vision applications commonly use CNN approaches [46]. Figure 4 presents the CNN-based object detection pipeline for livestock monitoring tasks.

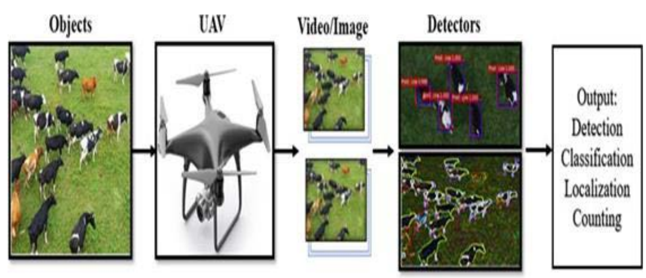

FIGURE 4. Livestock recognition pipeline.

UAVs provide recorded data in video or image formats for object detection in PLF; this happens by applying state-of-the-art object detectors like YOLO and R-CNN. Those architectures can detect individual or multiple objects of various sizes and scales overlapping UAV-based images or video records.

A. PERFORMANCE EVALUATION METRICS

As each research uses various datasets, data mining models, and metrics, performance comparisons between existing techniques become difficult without clear evaluation metrics. Evaluation metrics measure the model's performance by differentiating the outcomes of different learning models and, in practice, help researchers select the optimal solution. Several metrics compare and evaluate the DNN models' prediction performance. First, construct a confusion matrix table for each algorithm to examine the distribution of correct and incorrect prediction rates and evaluate a given classifier's performance to identify tuples of various classes [47]. In [48], a confusion matrix evaluates the CNN visual classification accuracy trained by the augmented UAV-based data for cattle counting. Next, they measured the mean pixel accuracy (MPA) and average distance error (ADE) to examine the performance of cattle segmentation and contour extraction results. MPA is a standard measurement tool for evaluating image segmentation acquired from precisely segmented pixels. Finally, average distance error is measured to evaluate the derived contour line [49]. In [50], the authors selected the mean average precision (mAP) to compare three detection algorithms (i.e., YOLOv3, Faster R-CNN, RetinaNet) performance applying the area under the curve (AUC) for generated precision-recall curves over every cross-validation fold. It measures the average precision (AP) value for recall as the region covered by the precision-recall curve at the various intersection over union (IoU) thresholds (from 0 to 1). If the IoU value exceeds a certain threshold ranging from 0 to 1 is considered true positive (TP) or else false positive (FP) [51]. Another study by [30] performed IoU to compare predicted bounding box regions and ground truth denoted as ''bbox_p'' and ''bbox_{gt}'', respectively shown in Table 2.

The appropriate threshold selection is crucial for the performance evaluation of cattle or sheep detection networks. Such systems will be weak if a threshold is incorrectly selected, indicating overlapping bounding boxes and missing objects. Nonetheless, determining the optimal value requires evaluating the performance over different IoU thresholds [52]. Specific confidence score threshold metrics, such as Non-Maximum Suppression (NMS) and IoU, are commonly used to evaluate DL-based object detection and localization techniques [11]. In object detection methods where an intense set of duplicate predictions appear, NMS removes these duplicate FPs. To evaluate a cattle detection framework trained by YOLOv2, [53] calculated the IoU between bounding boxes and ground truth, and the selected threshold values were 0.5 and 0.2, respectively. Favourably, precision and recall are applied using the values of TP, FP, and False Negatives (FN). Weighted mean classification accuracy is another evaluation measure used to evaluate the CNN

TABLE 2. Threshold metrics for object detection algorithms.

Metrics	Formula	Evaluation focus
Intersection over Union (IoU)	$\frac{bbox_{gt} \cap bbox_p}{bbox_{gt} \cup bbox_p}$	Measures the overlap between two boundaries, the overlap between the ground truth and predicted box regions
Mean Pixel Accuracy (MPA)	$\frac{1}{k}\sum_{i=0}^{R}\frac{p_{ii}}{\sum_{j=0}^{K}P_{ij}}$	Commonly used to evaluate the image segmentation models and it's measured from correctly segmented pixels
Average Distance Error (ADE)	$\frac{A_{union} - A_{overlap}}{T_{contour}}$	To evaluate the obtained contour line
Precision	$\frac{N_{TP}}{N_{TP} + N_{FP}}$	Measures the percentage of correct prediction
Mean Average Precision (mAP)	$\frac{1}{n}\sum_{r\in\{0,0,\ldots,1\}} P_{interp^{(r)}}$	Is widely applied to measure the accuracy of object detectors, such as Faster R-CNN A popular metric used to measure the accuracy of object detectors like Faster R-CNN etc.
F-measure (FM)	$2 \times \frac{Recall \times Precision}{Recall + Precision}$	F1Score or F-measure is calculated as a harmonic mean of precision and recall
Recall	$\frac{N_{TP}}{N_{TP} + N_{FN}}$	It measures the percentage of true positives over all possible positives
Accuracy	$\frac{(TP + TN)}{(TP + TN + FP + FN)}$	Calculates the ratio between the sum of correct predictions (TP and TN) to the total number of values in the matrix
Mean Square Error (MSE)	$L(y,\hat{y}) = \frac{1}{n}\sum_{i=0}^{m}(y,\hat{y})$	Used to quantitatively compare the performance of the resultant segmented image of different methods estimates to actual values

performance on aerial UAV datasets [54]. The precision, recall, and F-measure commonly derives from TP, FP, and FN using a confusion matrix. However, recall is more important in the counting context, indicating the number of detected objects (e.g., sheep, cattle). Precision measures the relevancy of the correctly classified instances, while recall indicates the model's percentage of all correctly classified instances.

The precision-recall curve [53] evaluates the accuracy of cattle detection at various input resolutions. Moreover, its area under the curve (AUC) shows the effect of input resolution resize when the aim is the resolution optimization. In [55], they evaluate the performance of the CNN-based detection models using the precision-recall curve showing the disjoint of the actual centroids and the estimated centroids of the sheep. Similarly, [39], [56] assessed fifteen different CNN architectures using four global performance metrics, accuracy, precision, recall, and F1Score. For livestock classification and counting approaches, [52] evaluated the performance of the Mask R-CNN model using the precision, recall, F-measure, confusion matrix, and mAP and another cattle detection framework using IoU, precision, recall, and F-measure [53]. The F-measure combines the trade-off of recall and precision, indicating the prediction performance fits a classifier with imbalanced classes [57], [58]. "Overall accuracy" is a classic metric evaluated based on correct predictions over total predictions from the confusion matrix. Aside from classification problems, the study by [59] uses Mean Square Error (MSE) as a loss function to alleviate image segmentation problems. The study [60] proposes evaluation methods to assess animal detection models by comparing the number of detections to the actual number of objects with fewer restrictions on the positional accuracy of predictions. Table 2 summarizes PLF studies' most used evaluation metrics for object detector evaluations.

IV. ARCHITECTURE OF UAV SYSTEMS

Modern PLF requires continuous monitoring to ensure optimal functioning. UAVs or drones with mainly audio-visual sensors provide sufficient data within a short period. According to [61], UAVs are typically equipped with cameras, LiDAR, multi-spectral, and obstacle avoidance sensors to provide a live feed for livestock monitoring. A UAV hardware classifies into two groups: fixed-wing and multi-rotor aerial vehicles. The fixed-wing UAVs require flight planning and control using GPS digital map navigation. Fixed-wing UAVs have a mounted gateway to collect data from the sensors and a greater load-bearing capacity, speed, and longer flight time, and can carry many sensors.

On the other hand, Multirotor UAVs have slower flight speeds and smaller payloads, and their maximum flight time is approximately half that of fixed-wing ones. However, they are most suited for smaller spaces that are not easily accessible and are considerably cheaper than fixed-wing ones [62]. In general, flight time and range, initial cost (e.g., imaging sensors, hardware, software, and tools), national laws, connectivity, weather dependency, and the need for specialized knowledge and skills are specific blockages for UAV-based farming development [18], [63]. As a result, multi-rotor UAVs are the most deployed types in livestock monitoring, following the quadcopter and fixed wings.

A multi-rotor UAV with six engines and six arms bearing a single board computer with a 700 MHz processor and 512 MB memory capable of running a Linux distribution was used for monitoring and counting animals [48].

The authors of [56] deployed a DJI Phantom 4 Pro equipped with a 20-MPixel to detect the Canchim breed that is visually like the Nelore breed. They observed significant contrast variations between animals and background from images captured at different times of the year. For instance, excessive brightness and motion blur caused low contrast images. Furthermore, observing different degrees of animal occlusion. Nevertheless, CNN models trained by UAV-based images under far from ideal conditions could reliably distinguish the animals and show UAVs' possible application for cattle monitoring. In terms of flight height, the minimum elevation flown was 10 m, and the maximum elevation tested was 120 m above the ground.

Several potential challenges in intelligent aerial robotics include computational demand and system capabilities, online learning, expedited learning time, uncertainties in data distribution and representation, stability-plasticity dilemma, the curse of dimensionality, limited payloads, and challenging flight environments [64]. In addition, operators require an advanced level of skill to fly drones in compliance with federal aviation regulations for UAV operations.

A. UAV-BASED DATA AND FEATURES FOR PLF

Exploiting deep learning for feature extraction and knowledge discovery from existing databases helps to improve decision support systems [66]. However, various factors influence the quality of livestock prediction algorithms in which feature lists are the essential components. Moreover, livestock behaviour and location across different temporal and spatial scales can reveal the factors driving resource selection, growth, reproduction and survival, response to disease, and coping mechanism with environmental conditions [67]. Therefore, real-time livestock behaviour monitoring can facilitate timely and accurate decision-making. However, introducing only a few successful commercialization of precision farming technologies recently. Therefore, to establish a successful PLF, the most significant factors, such as incorporating technology components, efficient interpretation of captured data from remote sensing tools, and relevant yet straightforward decision-making systems able to communicate continuously, are prerequisites in this industry. Concretely, UAVs, unmanned ground vehicles (UGVs), ML, DL, Big data, image processing, WSN, and cloud computing have brought positive and sustainable changes [68].

Remote sensing technologies, particularly UAVs, can cover large-scale livestock monitoring tasks as timely data collection and analysis become increasingly valuable with farms expanding and individual animal observation is not feasible. Traditionally, real-time image processing and analysis employ many data acquisition tools (e.g., GPS, thermal, accelerometer). However, UAVs are contact-free and present no risks of disease transfer, infection, or stress on animals while recording measurements. Also, UAVs are cost-effective when using a single camera or microphone to observe many animals, eliminating the need for sensor recovery on livestock.

Several types of research used acceleration data for animals' behaviour classification using ML algorithms. [69] applied AdaBoost ensemble learning algorithm to identify dairy cows' seven behaviour patterns, including feeding, standing, lying, standing up, lying down, regular walking, and active walking. Similarly, [70] implemented the Adaboost algorithm to classify dairy calves' behaviours using merged collar-mounted sensor signals and video camera records. In addition, they developed an adjusted count quantification model identifying many behaviours: locomotor play, ruminating, self-grooming, non-nutritive suckling, nutritive suckling, active lying, and non-active lying. Despite very low behaviour prevalence in real-world conditions, these multi-class classification and quantification learning algorithms demonstrated high accuracy. In [71], the accelerometer and gyroscope sensors extract eleven features: mean, standard deviation, kurtosis, minimum and maximum value, interquartile range, signal area, total signal area, zero crossings, dominant frequency, and dominant frequency and spatial entropy to test ML classification algorithms. The results demonstrated that the RF outperformed other classifiers (kNN and SVM) with the gyroscope-based features significantly contributing to the classification of sheep eating behaviours.

Similarly, [72] confirmed five features, including mean, standard deviation, root mean square, median, and range, for the highest performance classification of cows' seven behaviours, including lying, lying down, feeding, standing, regular walking, and active walking, and standing up. Ter-Sarkisov *et al.* [73] obtained nine features, including overall mean, maximum, and three quartiles vector features from pixel intensities, size of bounding boxes, and coordinates of the bounding box's centroid the distance feature used for cow tracking application using video data. While the currently deployed remote sensing tools generate a massive amount of data, the transmission of this big data is expensive, time, and energy-consuming. For instance, they utilize extensive data logger and accelerometer tools to examine livestock behaviour. These data recording tools provide animals' head inclination and acceleration features to distinguish dairy cattle lying time, the number of lying bouts, lying patterns, cow's head posture identification while grazing and differentiating non-grazing ones [74]. In addition, the investigation of stock density on livestock behaviour, productivity, and comfort of dairy cows used data logging tools [75]. Although these tools have provided valuable information to detect livestock behaviour patterns, animals experience stress when removing these tags from their bodies for data extraction [76].

Therefore, it is necessary to discover ways to acquire data from animals and transform that data remotely. Combining DL and UAV technologies with minimal interference offers visual and audio-based recognition abilities in livestock farming. The main phases of UAV-based object detection methods include three stages information region selection, feature extraction, and classification [43]. As an example, UAV-based video data fitted into DL (e.g., CNN) and image processing methods facilitated livestock monitoring of sheep [55], [77], and calves [78]. However, video sequencing is challenging for several reasons, each of which causes a specific issue. These challenges are 1) the animals changing position and different postures; 2) the view of animals can suffer from a great degree of partial occlusion; 3) changing illumination and lighting conditions can pose issues with learning algorithms to erroneously learn patches or shadows as animals' features; and 4) in many cases the noisy and dark background cannot be differentiated from the segmented objects even by the most advanced algorithms [73], [79], [80]. Different CNN structures, such as the Mask R-CNN, have enhanced the object detection and segmentation (e.g., cattle) performance by extracting more features and thus, overcoming the challenges mentioned earlier. It is also not affected by cattle coat colour, such as brown, white, dark, and poses [49]. Likewise, [80] presented a novel instance segmentation model that extends the FCN to label objects independently without predicting regions of interest. In addition, [81] used the residual learning theory to build an FCN able to harness multilevel contextual feature representations learned from various residual blocks from aerial images and video. To this end, DL methods extract predictive features automatically and are an active research area for animal behaviour studies. However, the user cannot interpret the extracted features from DL, unlike handcrafted features used in classical ML techniques. Thus, the choice of the algorithm varies by aiming for an optimum prediction accuracy or feature extraction.

Moreover, developing algorithms that capture relevant information at the lowest possible level, enabling the transfer of relevant features and knowledge rather than data, resolves the challenge of excessive data accumulation and transmission. Also, these innovative algorithms combining the livestock data with other public data can facilitate the PLF management standards [82]. Another possible way to enable new livestock decision support models is to expand available data sources. By bridging the geospatial data, social media feeds, and other data sources using interchangeable standards to develop new livestock decisions and support models. Table 3 lists the existing datasets utilized in UAV-based livestock recognition.

V. REVIEW RESULTS

Three parts discuss the results regarding DL-based PLF using UAVs: 1) Deep learning-based livestock monitoring using UAVs, 2) livestock detection and classification 3) livestock localization.

A. DEEP LEARNING-BASED LIVESTOCK MONITORING USING UAV

UAV and DL vision systems will soon provide autonomous solutions for precision farming applications in pastures. They are being exploited as a cost-effective and fast choice to collect data from specific regions. Most proposed livestock monitoring mechanisms rely on mounting tags, sensors, and nodes. However, it is troublesome to attach devices to many animals in terms of time and cost in the real-world environment. Several studies utilized satellite and aerial imagery and field data for independent livestock survey and detection; however, significant limitations are low image resolutions, high cost, and low sampling frequency. Object and pixel-based models have been successfully applied for livestock detection, counting, and positioning over small and relatively homogeneous areas with few images [12], [85], [86]. Livestock detection is the only primary level of more complicated tasks, including animal counting and anomaly recognition, addressing multiple technical problems, such as the size of the target object changing variously in image recognition tasks. Another presumed challenge is ML algorithms coping with object brightness variation beyond the training set. Alternatively, the most common and basic pixel-based supervised,

TABLE 3. Precision livestock farming databases in deep learning.

Database	Application	Resolution	Data format	No. of observations	No. of objects	Platform	References
FriesianCattle2015	Automatic individual Holstein Friesian cattle identification	Image 512×424 Video 1920×1080	RGB Video	764 RGB-D images	92 individual cattle	UAV	[83]
FriesianCattle2017	Holstein-Friesian cattle detection	Image 512×424 Video 1920×1080	RGB Video	940 RGB images	89 individual Holstein Friesian	UAV	[30]
AerialCattle2017	Visual Localization and Individual Identification of Holstein Cattle	3840×2160	RGB Video	46,439 image frames	23 individual cattle	UAV	[30]
Cattle UAV-based aerial images dataset	UAV-based cattle detection	224×224	-	14,489 images labelled as cattle 1,448 non-cattle	-	UAV	[56]
A real pasture surveillance dataset	Livestock classification and counting	Image 512×512 Video 4096×2160	-	1,000 images	37 37	UAV	[52]
OpenCows2020	Detection, localization, identification	Image 512×424 1920×1080	RGB image	3,707 images	6,917 cattle annotations	UAV and ground-based imaging	[50]
Aerial-livestock-datasets	Counting grazing animals on grassland	224×384	Image	2,716 images for training 348 images for testing	-	UAV	[59]
UAV-based wildebeest count data	Animal counting	4288×2848	Video	2,018	-	UAV	[84]
Aerial UAV datasets	Cattle counting and grazing management	400×3000	Image	-	212 individual targets	UAV	[53]
UAV-based dataset	Extensive mammal census in African savanna wildlife	300×4000	Image	654	-	UAV	[60]

unsupervised, and thresholding methods based on mapping assign every pixel to a class because large animals can cover single to multiple pixels in very high resolution (VHR) space-borne imagery.

The authors of [87] identified DL-based PLF as one of the most common topics between 2016 to 2019. While, in review by [88], few works considered issues related to livestock farming using DL, with only three papers published after 2015. Recently, several state-of-the-art object detection approaches based on CNNs such as DenseNet, ResNet, and NASNet have achieved reliable outcomes for wild animal and livestock recognition. However, achieving high automation degrees and accuracy in livestock recognition, trained explicitly on UAVs' data, requires improvements [89]. Moreover, the additional post-processing efforts negate UAVs' time savings and convenience compared to traditional monitoring techniques. Thus, to deliver a reliable UAV-based livestock detection and counting over data collection to analyse powerful data analytics capabilities is necessary to use the data collected by UAVs. Livestock monitoring and detection studies in UAV imagery are emerging more recently. [90] examined the possible integration of generic object recognition solutions for onboard animal detection and counting. They concluded that the object detection deformable part-based model (DPM) is well-suited to automate the animal detection and counting onboard UAVs.

TABLE 4. Application of DL algorithms in PLF using UAV.

Application	Method	Backbone network	Type of livestock	Type of data	Environment	Platform	Altitude	References
Segmentation and body contour extraction	Mask R-CNN	FPN based ResNet101	Cattle	Image	Feedlot, Australia	Camera	-	[49]
Individual cattle unrecognizable boundary and body shape recognition and segmentation	Enhanced Mask R-CNN	ResNet101+RPN+Feature Pyramid Network (FPN)	Cattle	Image	Ranch, Nigeria	Camera	-	[97]
Detection and counting	FCN, CNN	Network-I, Network II, AlexNet, GoogLeNet, VGG16, VGG19, ResNet50, and U-Net	Sheep	Video	Farm, New Zealand	UAV- a DJI Phantom 3 Pro	80-120 m	[98]
Detection	CNN	VGG-16/VGG-19/ResNet-50 v2/ResNet-101 v2/ ResNet-152 v2/ MobileNet/MobileNet v2/ MobileNet 121/DenseNet 169/ DenseNet 201/Xception/ Inception v3/Inception ResNet v2/NASNet Mobile/NASNet Large	Cattle	Image	Farm, Brazil	UAV- a DJI Phantom 4 Pro	30 m	[56]
Detection	CNN	64x64-18C7-MP4-96C5-MP2-4800L-2	Cattle	Video	Farm, Brazil	Multi-rotor UAV	-	[99]
Visual localization and identification	R-CNN	VGG CNN M1024	Holstein Friesian cattle	Image Video	Farm, UK	Camera and DJI Inspire MKL	25 m	[30]
Detection	YOLOv2	AlexNet, GoogleLeNet, Inceptionv3	Holstein Friesian cattle	Video	Farm, UK	DJI Matrice 100 quadrotor UAV	10 m	[100]
Visual identification	CNN	ResNet50	Holstein Friesian cattle	Video	Farm, UK	UAV	-	[101]
Detection	YOLOv2 Faster R-CNN	AlexNet	Wildlife	Video	Wildlife, Namibia	UAV	-	[102]
Detection, identification, and identity estimation	CNN	DarkNet	Cattle	Image	Barn, Italy	Camera	3.5 m	[103]
Counting and monitoring	CNN	y64x64-18C7-MP4-96C5-MP2 -4800L-2	Cow	Image	Spain	Multirotor UAV	-	[48]
Detection and counting	R-CNN	-	Sheep	Image	Farm, New Zealand	UAV	80 m	[55]
Classification and counting	Mask R-CNN	-	Cattle and sheep	Image	Farm, Australia	Quadcopter, MAVIC PRO	-	[52]
Detection	CNN	U-net, Google	Livestock	Video	Grassland, China	Quadcopter UAV	-	[59]
Identification and localization	Faster R-CNN	Inception-v4	Goat		Farm, India	Mobile Camera	-	[104]
Detection	CNN	Inception-ResNet-v2	Cattle		Farm, Brazil	DJI Phantom 4 Pro	30 m	[78]
Visual instance segmentation	FCN	FCN-16s and FCN-32s	Beef cattle		Farm, Irland	Camera	-	[80]

Extensive UAV-based wildlife management and conservation are advancing due to its versatile real-time data acquisition performance and timely decision making. According to the literature, cattle monitoring studies comprise detection and counting [39], [51], the distance between calf and cow [91], and feeding behaviour [92] using DL methods to extract information from images obtained by UAVs. A comprehensive examination of intelligent cattle monitoring approaches showed that the non-invasive, consistent, and real-time cattle detection, body condition score evaluation, and live weight estimation significantly contribute to future studies [23]. In [57], the developed system based on CNN's achieved a body condition score results whose application might extend beyond dairy production. Consequently, technologies such as 3D model reconstruction and DL should be combined to develop non-contact, high precision, automated systems for intelligent cattle farming. For the cattle detection and counting task, [93] took advantage of computer vision techniques (i.e., CNN and structure from motion (SfM)); to construct the candidate bounding boxes and the 3D surface construction of the pasture used aerial records.

Further, the per-frame detection outputs joined with the 3D surface; hence, the multiple detected cows were removed and reported a correctly counted number of cows. CNNs play a significant role in remote sensing tasks, considering the recent DL-based object detection model developments. Furthermore, they have been the base building blocks for nearly all computer vision applications [94].

Several researchers have demonstrated the possible integration of UAVs and DL algorithms for livestock monitoring; however, screening the entire farm in a single flight and detecting too small objects are the main drawbacks [95]. This issue can increase error rates related to the time interval between flights, sudden weather changes, dynamic livestock behaviour, and the angle of flight incidence will vary. Several studies achieved promising results in monitoring animals by performing flights at specific time intervals [96] with recorded images at the nadir position (orthogonally). Still, the complete area coverage was limited, which would improve with multiple UAVs. The commonly selected vertical angle provides the same ground sample distance (GSD) on all images that facilitate the object of interest detection. Alternatively, [78] investigated a feasible way to monitor the entire farm using an oblique UAV image and tested it using deep CNN. Experimental results showed that a tilted angle could increase the area covered by each image under certain conditions by addressing precisely view obstructions and determining the border.

Also, constant weather conditions challenge farmers to monitor livestock active spread and congregation. The only literature exception to mitigate this issue was image acquisition at an angle with all pints assigned the same GSD [51], [52]. The GSD on these images varies in oblique images, occlusion become severe, and other object detections and measurements become challenging. However, the area screening capacity increases and significantly reduces the quantity of UAVs required. Table 4 briefly describes DL algorithms for livestock recognition tasks using UAV records.

B. LIVESTOCK DETECTION AND CLASSIFICATION

Deep learning-based object detection and semantic segmentation techniques received attention in livestock monitoring in pasture and open space environments. State-of-the-art CNN is a new approach that showed remarkable results for object detection, classification, and localization tasks in computer vision [79]. Furthermore, the in-depth features learned by CNNs provide necessary semantic and spatial knowledge [30] to further evaluate cattle welfare in precision livestock management. An individual cattle segmentation (for classification and localization) and contour extraction technique based on the Mask R-CNN framework [49]. The experimental results showed Mask RCNN outperforming other instance segmentation techniques (DeepMask and SharpMask) with a high MPA of 0.92 and ADE of 33.56 for cattle segmentation and contour extraction. However, it is essential to focus on enhanced segmentation methods for the overlapping cattle regions and their different body parts. To support this, [97] proposed an enhanced Mask R-CNN instance segmentation model to detect indistinguishable body patterns and boundaries of cattle. The transformed instance segmentation and contour extraction results, depicted in Figure 5.

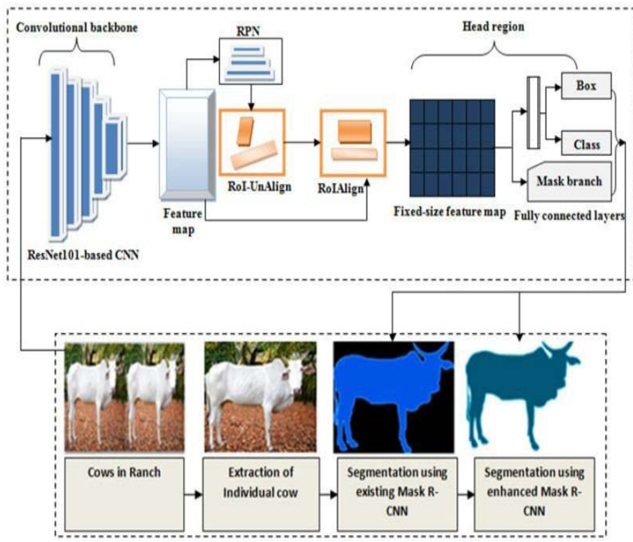

FIGURE 5. Enhanced Mask R-CNN for cattle image instance segmentation [97].

The enhanced Mask R-CNN framework effectively extracted smaller and overlapped features with optimal filter size, unlike Mask R-CNN, and enhanced segmentation with sub-network integration. In addition, the implemented approach obtained an MPA of 0.93 and achieved precise simultaneous localization and mapping. However, due to the variation in cattle contour, its manual pixel-wise annotation is costly and time-consuming in CNN-based approaches. Accordingly, [105] presented a cattle segmentation DNN Bonnet method based on different data augmentation approaches. This strategy has improved the segmentation accuracy of cattle from the complicated background with 99.50% mean accuracy and 97.31% mean IoU. For beef cattle segmentation, [80] developed a new instance segmentation algorithm based on FCNs, eliminating the region of interest (RoI) prediction from video data. The presented instance segmentation MaskSplitter framework with FCN8s as the last layer outperformed state-of-the-art methods such as Mask R-CNN.

The Authors of [98] explored Different FCN and CNN networks for sheep detection and counting using UAV-based videos recorded from 80 m and 120 m altitude. The results showed U-Net-Ms model outperformed the other networks (AlexNet, GoogleLeNet VGG16, and ResNet50). UAV frames captured from a height of 120 m (maximum legal altitude in many countries) covered a larger area of paddock; however, various other objects (e.g., tree branches, wooden logs, farm barriers) obscured sheep recognition accuracy. In-depth, altitude determines livestock contour's distinction; thus, the lower the measurement, the more distinguishable the objects are from UAV records using CNN networks. Furthermore, the fully connected network applies centroids of the objects instead of bounding boxes in CNNs, which fits composite, small, and overlapping objects detection. That framework used a small Synthetic Aperture Radar (SAR) image using a multi-view network for Automatic Target Recognition (ATR), which exhibited an excellent recognition performance with the limited number of records required for network training. It also performed remarkably well in various conditions, such as depression angles and configuration changes [106]. Even though UAVs provide distinctive features, their livestock detection and localization applications are far from fully operational; they solve various issues.

Several studies experimented with different DNNs and UAV imagery to determine the highest achievable detection accuracy in the breed-wise livestock recognition task. The proposed approaches achieved the ideal GSD and the most accurate CNN architecture. Furthermore, most models obtained considerable robustness against restricting conditions, such as illumination, extreme brightness, blurriness, and partly visible animal body. For example, in a study by [56], 900 models (CNN architectures $\times$ three spatial resolution $\times$ two datasets $\times$ 10- fold cross-validation) were extensively analyzed and tested with 15 CNN architectures to achieve the most precise CNN model and the optimal GSD using UAV images. Overall, CNNs networks showed accuracies higher than 95%, with sizeable exceptional accuracy close to 100% for robust detection of two similar Canchim and Nelore breeds. In addition, these approaches could overcome restricting factors in UAV-based images under undesirable conditions, revealing the reliable application of UAVs for livestock monitoring. Notably, the captured records at relatively high altitudes produced the highest accuracies with a GSD of 2 cm/pixel for animal detection. However, model drift is a challenging factor using pre-trained models.

Further, to examine the main challenges of small object detection, [107] applied Sig-NMS-based R-CNN with transfer learning using remote sensing images and automatically labelled them. Also, [108] introduced a self-reinforced network for small object detection named remote sensing R-CNN (R2-CNN), comprised of a network backbone Tiny-Net, intermediate global attention block, end classifier and detector. [109] showed feature pyramid network (FPN) to generate integrated feature maps that significantly improve small object detection accuracy. Hopefully, these approaches will inspire future real-time livestock remote sensing systems developments. Most developed object detection and counting techniques based on computer vision approaches use non-orthogonal photos captured horizontally with optical cameras. Rivas *et al.* [99] designed a cattle detection system using video data captured by the auxiliary camera equipped on a multirotor UAV. The sliding window technique was employed to analyse small, bordering, and intersecting image tails located over the frame. Subsequently, obtaining the output results using a trained CNN network. Few objects in the frame were reduced with overlapping boundaries, achieving the highest accuracy of 98.78%. They also concluded that vertically taken aerial images by UAVs could display tiny objects of interest with only a top view presenting a blob shape that lacks valuable features. Furthermore, the area of interest is usually overlapped or like other objects in the background. One animal can pass through the UAV path multiple times, creating complex problems to solve. While corresponding ground images contain numerous distinctive features (e.g., head, body, legs) important for pattern recognition [110].

Alternatively, [30] introduced a video processing pipeline using long-term recurrent convolutional networks (LRCNs) for Holstein Friesian cattle visual detection and localization using video captured by a UAV. The LRCNs were first introduced by [111] as the combination of CNNs and long short-term memory (LSTM) using UAV-based imagery datasets. Convolutional visual features were extracted using the Inception V3 CNN and fed into an LSTM layer for visual identification of an individual cattle. Consequently, the video-based LRCNs successfully distinguished the unique dorsal patterns and structures exposed by particular species with an accuracy of 99.3%. For path planning, UAV and wireless sensor network (WSN) based on the Markov decision process (MDP) model were applied [112]. Even though deployed sensors in WSN can provide animal monitoring information in an outdoor environment, long-distance wireless communication in large remote wildlife regions is expensive and impractical. For example, long information transmission delays can lead to livestock loss. Currently, some systems use UAV communication using sockets [48]; but it is limited to fixed-wing UAVs. This system encourages the multi-rotor UAVs' automatic control and information acquisition regarding the status of the arrays it carries like GPS, and it can send video footage. Ultimately, timely and efficient animal data collection and transmission are crucial for livestock monitoring and localization tasks.

Livestock tracking in challenging environments comes with several significant issues, such as crowded background and high similarity within a group of in-motion objects, which significantly reduces the robustness of existing methods. A design for an Unmanned Aerial Vehicle onboard system used a DL inference process for visual detection of seventeen heifer Holstein Friesian cattle in an open pasture environment. The proposal used three layers of deep CNNs architectures comprising a species detection using YOLOv2, a dual-stream deep network to create an exploratory agency, and an InceptionV3- based biometric LRCN for single object detection. The experimental results proved the concept with error-free aerial identification performance (100% accuracy) in freely moving herds in an open environment that supports tag-less technology towards autonomous farming technology [100]. However, in successfully performing independent aerial identification of animals, computational resource limitations alongside payload, weight, and more should be considered for future farming automation and productivity.

Similarly, [50] proposed a non-intrusive automated cattle detection, localization, and identification technique from aerial images. Initially, three object detectors such as YOLOv3, Faster R-CNN, and RetinaNet, were applied for cattle RoI extraction. A ResNet embedded the output for image clustering based on coat patterns. Lastly, carrying out the classification step using kNN to create cattle identities. The state-of-the-art object detectors such as YOLOv3, Faster R-CNN, and RetinaNet networks have yielded decent results in primary cattle identification pipelines. Even though the deep object detectors were highly precise, the combination of basic localizing methods within video frames is likely to remove any possibly occurring errors. For significant animal detection from UAV images, [102] implemented CNNs. The proposed architecture on a pre-trained AlexNet adopted two approaches to identify the presence and size of the animals by assigning the local likelihood scores.

Consequently, the proposed architecture outperformed the Fast R-CNN with a far better average image processing time per frame. Studies compared the two models, Faster R-CNN and YOLOv2, to recognize and detect the animal species in images. The trained models achieved 83% and 76.7% accuracy for Faster R-CNN and YOLOv2. Although Faster R-CNN was more accurate than YOLOv2, YOLOv2 was faster than Faster R-CNN. Overall, both have demonstrated high detection performance, but they performed unsatisfactorily in the presence of small targets in images obtained by UAVs [113]. The latest study [103] applied YOLOv3 neural network using high-resolution video frames to identify a single or a group of cows, their movement, actions, and positions in real-time. The model was trained and validated with four cows' video records from the barn. The designed framework achieved satisfying IoU scores between 0.75 to 0.81 overcoming the global object detection threshold (0.7 to 1.0). The outcome showed that the cow coat pattern is most fitted for detection. Moreover, single cow recognition is possible based on its outward appearance, and the piebald spotting figures of a cow's coat depict a discernible feature for the computer vision networks.

C. COUNTING APPLICATION

For counting livestock and animals in nature, the algorithms require extracting every animal's unique feature to consider every individual only once. For wildlife and animal monitoring, such as goat groups [114], recently quadcopter UAVs have been extensively employed. However, aerial cattle detection and counting is a deceptively challenging task with livestock and low background contrast, their profile movement with body changes, large groups of animals, and the presence of small livestock (i.e., calves). Automatic livestock identification and counting could enhance animal monitoring and management and population estimation. Generally, livestock manual counting is labour-intensive during drenching, shearing, or loading and tends to cause slight but substantial errors. Recently, farming industries are moving towards automated methods to efficiently manage and monitor their herd. Accordingly, researchers have applied several statistical and biological techniques to cope with challenges, such as species-specific characteristics, diversity of background, and spatial clustering of animals. Most of the relevant research in this domain has shown promising outcomes using ML classification algorithm [90], template matching algorithm [115], deformable part-based model (DPM), and power spectral-based methods. However, these methods for detecting and counting livestock on UAVs require handcrafted training data. Also, these methods apply to images where the livestock are low in number and require high-resolution records.

Furthermore, various animal species' remote monitoring and counting under different climate conditions is advancing with the evolution of DL algorithms and UAVs. For instance, [48] successfully performed a livestock recognition and counting system via CNNs and video data captured by UAVs. Recent counting approaches revolve around data from scattered regions that favour detection networks. However, these techniques fail to detect sparsely located dense objects. In [110], the authors developed a novel DisCountNet model based on two-stage networks of DiscNet for coarse detection and CountNet conducted on dense areas of the sparse matrix to produce a density map using the UAV dataset. The system consisted of detection, and heat-map networks presented a simple yet effective solution by processing a large and high-resolution image. However, even though the network improved various object counting and localization, its outcome was highly occluded. Another issue of dealing with hundreds of small animals per image still lacks information and reliable processing methods. In this regard, [55] explored two sheep detection and counting approaches based on R-CNNs and an expert system using blob analysis from UAV video. While the proposed expert system technique indicated great potential for the intended application, the CNN technique required more work to detect practical objects, especially when dealing with small objects in the background. A variety of cattle and sheep recognition and counting system based on Mask R-CNN from UAV images have been proposed [51], [52]. The results demonstrated its classification and counting effectiveness with 96% and 92% accuracies, respectively. Figure 6 presents a successful livestock classification and counting framework based on Mask R-CNN and UAV. Even though the sheep and cattle classification and counting followed the identical pattern with the changing density, the classification and counting certainty of livestock over changing thickness is not investigated.

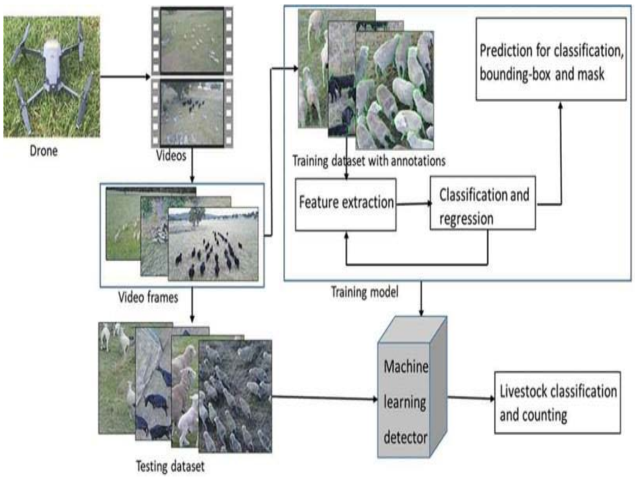

FIGURE 6. Enhanced mask R-CNN for cattle image instance segmentation [97].

Deep learning detectors such as R-CNN, Mask R-CNN, Enhanced Mask R-CNN, and Faster R-CNN based on instance segmentation have effectively employed background discrimination and identifying fewer uniform livestock like cattle with multi-colour. In addition, UAVs equipped with these approaches require large datasets. Contrary, most livestock monitoring techniques have been conducted on limited observations, yielding numerous false alarms when tested with the large-scale and actual study scenes for extensive animal monitoring. To this end, it presents a method to train CNN for animal censuses on a UAV-based image dataset, including many images with sparse animals [60]. As a result, the CNN yielded a considerably higher prediction performance and generated reliable results expanding over fewer image tiles. Furthermore, these research results demonstrate CNN's effectiveness in producing a substantially higher precision at high recall values than state-of-the-art detectors and achieving more confined predictions across fewer image tiles.

Furthermore, CNN reduced the number of image tiles significantly compared to the baseline, at a recall level of 90%. Thus, these evaluation methods and the recommended model-agnostic are straightforward and can complement any DL object detector for animal monitoring with accessible results during an actual UAV movement, counting images without wildlife records. [59] compares the livestock detection algorithm based on U-net and Google Inception-v4 networks against YOLOv3 and Faster R-CNN. The proposed model based on U-net and Google Inception-v4 achieved better detection results than YOLOv3 and Faster R-CNN. The lower detection performance of YOLOv3 and Faster R-CNN was attributed to their ineffective light-coloured instance segmentation. Other research gaps, such as high-speed algorithms, fence lines detection, robust parameter detection, and the integration of tracking technologies to resolve minor errors in object detection, should be tackled. An overview of the latest achievements with CNN techniques for livestock counting and density estimation is more proficient at handling large density crowds with object scales and background changes. Also, these methods drastically enhance the estimation error when incorporated with scale and contextual information (i.e., time, location, temperature) [48].

Another UAV-based cattle counting by [39] applied a pre-trained CNNs network known as Nas-Net Large to depict the area of interest. The process included colour transformation and object segmentation from the background, mathematical morphology to distinguish clusters and false objects elimination, and image matching to match image overlapping. The proposed structure achieved over 90% accuracy under different conditions and backgrounds with highly present young calves. UAVs and autonomous approaches would be especially applicable in broad and remote rugged terrain. It represents a comparable solution for livestock monitoring to other methods.

Further possibilities of covering large areas at once should be tested to eliminate the long flights and image overlap limitations [8]. Wang *et al.* [89] proposed a livestock population estimation model with the animal crowd features extracted from UAV imagery and satellite data. According to experimental results, both the UAV and satellite imagery achieved reliable results in yak counting, but satellite imagery failed to detect sheep due to low-resolution observation. Additionally, the UAV operation did not cause changes in livestock distribution and reactions at 700m altitudes. UAV shows to be a promising, affordable, and reliable platform for PLF applications. However, monitoring animals over extended distances and vast study regions requires caution.

Additionally, remote imagery using UAVs for small animals is highly constrained by the imagery resolution. The small object detection from UAV imagery is advancing, taking advantage of DL algorithms to automatically detect livestock, which needs constant counting between and within years. However, these techniques fail when fed with too small sizes of the object of interest, overlapping background, and low resolution of the images. Even though these issues challenge researchers, having a uniform scene reduces the diversity and complexity of geospatial object presence. To sum up, animal counting is possible using DL-based object detection techniques embedded in UAVs for their practical application to deliver diverse aerial imagery.

D. LIVESTOCK LOCALIZATION

Farm animals' body location and motion identification greatly determine their well-being and efficient monitoring [116]. Specifically, few livestock monitoring systems have been commercialized for livestock behaviour and position inspection from motions; however, these are relatively expensive. For instance, [117] proposed the Gea CowView system able to detect the presence of cows in the 0.5-3 m range with over 70% sensitivity using ultra-wideband (UWB) highly spatial and directional data. [118] employed GPS tags and Bluetooth Low Energy (BLE) technology for herds localization and identification. For outdoor localization of cows, [119] evaluated the matching method and DNN using the received signal strength indicator (RSSI) from BLE tags. Deep learning methods facilitated livestock localization with higher precision and speed than the matching approach. The matching method uses pattern matching between RSSI datasets during localization [120]. However, this approach suffers from the large fingerprint datasets that lead to long retrieval time and processing costs over a vast pasture region. Moreover, authors in [121] and [122] proposed low-cost BLE tags for cow's location tracking with an achieved precision of 3.3 m and an accuracy of 3.27 ∓ 2.11 m in a barn (10×40 m2). However, the amount of noise due to tag positioning, orientation sensitivity, and pasture environment structure decline the developed systems learning ability.

Visual information recorded by UAVs provides rich chromatic details of the scene. DNNs are novel in the field of Aerial imagery research; the training of object localization uses UAV recorded data based on deep network architectures. A study by [104] applied a trained Faster R-CNN model using fine-tuned transfer learning for automatic goat breed identification and localization using images. Localization testing used coordinate information of the animal in each training image and obtained the ground truth data using a bounding box predictor algorithm. Creating bounding boxes around the goats in the training phase performs the image annotations task. The evaluated fine-tuned Faster R-CNN with Inception-ResNet-v2 as the feature extraction backbone showed positive results as animals breed detection system. This approach could automatically dismiss low-resolution images limiting the identification of animal breed detection. CNN's achieved good performance for ground-to-aerial localization studies [123]. Recent object detection and tracking promoted motion features next to the task of localization. The authors in [124] showed the reliable application of motion features for animal detection and tracking wildlife using UAV videos. Although high-speed cameras can capture information to perceive and detect fast-moving objects, the major drawback is the possibility of them being out of the field of view and do not assume animals are constantly in motion.

Similarly, fast-moving animals within the field of view can be too small and barely visible to be detected by computer vision algorithms [125]. Technological advancements in data collection with UAVs provide a wide range of data resources to analyze animal behaviour in unprecedented detail. Livestock classification and localization from audio-visual data are becoming an important research area to tackle the dynamic behaviour of animals on farms. Given that animals have audio and visual signatures, both audio and visual modalities can be used to study animals' behaviour. Specifically, audio provides temporal segments, and visual data provide visual-spatial learning. Technically, feature extraction from this data type is filled with low contrast, poor lighting conditions, and frequent occlusion issues when the aim is to monitor animal behaviour directly. Audio-visual analysis of field data using labelling methods allows real-time monitoring to manage individual cows [126]. Alternatively, a cow tracking algorithm was developed using video data based on DL algorithms, heuristic techniques, and an ensemble learning method [73]. Specifically, a novel object segmentation algorithm named conditional random fields as recurrent neural networks (CRF as RNN) was designed for localization. So far, this approach has demonstrated very high performance on video data compared to other methods, including FCN.

Specifically, one-stage detectors, such as YOLO and SSD, were used to localize and segment bird localization [13], [127]. A CNN can effectively perform deep-level feature extraction and is essential for species categorization and environmental sound classification. Another cattle monitoring system was built based on CNN and MFCCs for real-time cattle audio detection and corresponding behaviour matching. Two CNN networks were applied to evaluate cattle conditions and behaviour classification by classifying cattle audio and eliminating background noise from existing datasets. The audio-based cattle vocal detection monitored their actions through a four-behavioural classification model with an accuracy of 81.96% following SIFT analysis used for audio filtering [128]. Likewise, [129] investigated animal audio classification based on CNN and MFCCs to classify ten different animal types. The Nesterov-accelerated Adaptive Moment Estimation (NAME) reports the best accuracy of 75%. However, the low number of training samples harmed the model recognition performance as it is well-known that DL requires big data for better generalization. Thus, recommends increasing the number of samples or unsupervised data augmentation methods to increase data volume where data acquisition is restricted. Alternatively, use data augmentation by randomly cropping and patching images to increase the size of livestock data. This strategy has improved the segmentation accuracy of the DNNs to resolve the limited labelled training datasets in farming applications [105]. To classify animal species, a multi-view CNN framework with a wireless acoustic sensor network (WASN) reported a high accuracy and outperformed classical classification algorithms (e.g., SVM, kNN) when the environmental noise dominates the audio signal [130]. Audio detection as a system to identify livestock is a field with significant potential. Recently a Siamese neural network (SNN) generated dissimilarity descriptors to determine and enlarge the distance function within multiple classes for animal sound classification [131], [132]; this leads to the idea of future improvements used in conjunction with a CNN.

Furthermore, the authors of [133] developed a novel Audio-visual Fusion Block (AVFB) and Segment-Wise Attention Block (SWAB) to address the sound source positioning issue. The proposed model was automated to learn the fusion between audio and video, containing high variability in data without imposing any constraints, thereby assuring generalizability and robustness. The developed model enabled visual scene perception using audio localization for robotics. Recent approaches used network localization methods to locate sound sources using DNNs measuring the correlations between visual and audio inputs. However, these models are suited for relative positioning tasks and can handle only sequential data [134]. Deep learning algorithms depend significantly on data representation, and real-work data such as videos, images, and audio does not retain specific algorithmically defined features [135]. Alternatively, introducing Audio-visual Correspondence (AVC) could discover single embeddings that illustrate the sound across an image—led to a new network architecture design for cross-modal retrieval and sound source localization using the unsupervised AVC the objective function. Audio-visual object localization (AVOL-Net) exhibited impressive object localization capabilities [136] using the unlabelled video's overlapping visual and audio systems; these cross-modal auditory localization approaches can tackle the labour-intensive and objects' bounding box annotations. Although the heterogeneous nature of the discrepancy in audio-visual learning (i.e., a large gap between audio tracks and visual modality) poses another research challenge. Hence, attention mechanisms and memory banks may improve the performance of mapping audio attributes and objects in an image or a video in a PLF application [137]. Most established livestock health assessment methods are performed manually by farmers, effective for only a limited number of livestock, and prone to human error. Few studies classified sheep eating behaviour with relatively high accuracies using sensors [76], [138]. However, these studies used highly unbalanced datasets [139] or had very few algorithmic classification data points [140], [141]. Therefore, livestock timely behavioural classification systems combined with UAVs can provide unusual behaviour and health features where different behaviour changes can imply health-related issues.

VI. PROSPECTIVE LIMITATIONS AND FUTURE SOLUTIONS

UAVs or drones in this application have abundant contextual and spatial data for animals scattered around the pasture [18]. Nonetheless, finding possible solutions to some of the reported and addressed significant limitations. Primary, UAV-based images are associated with a big scene-understanding issue as ground-based images. In [142], semantic segmentation architectures like U-net and Inception were ineffective for light-colour aerial image segmentation. However, the two neural networks, Faster R-CNN and Yolov3 obtained better prediction results than the semantic segmentation models. Overall, most techniques for livestock monitoring and detection in literature use two approaches: 1) CNNs to create a probability heat map for livestock localization [48], [99], [110], and 2) creating bounding boxes around the object [53], [100]. However, the former architecture performed poorly on identifying small objects, generalizing visual feature ratios, and not well-established datasets. The latter struggled with precise image annotation, which is time-consuming and labour-intensive. Even though DL models dominated visual recognition, their performance dropped with either or both semantic and covariate shifts, and they tend to overfit. To this end, the authors of [65] proposed to use Semantically Coherent Out-Of-Distribution (SC-OOD) with an Unsupervised Dual Grouping (UDG) to distinguish between distribution and OOD samples. Using that method is particularly important in modelling UAV-based data collected at different times and under different conditions. Specific livestock detection and localization challenges in vast rangelands comprise 1) livestock blocked by trees and bushes, 2) contrasting illumination, 3) weak features from small-scale livestock images, and 4) herding animals' distinction. The findings showed numerous potential uses of UAVs and DL technologies to support livestock PLF activities to detect, localize, and classify. The future of livestock farming can hold swarms of UAVs [143] and hybrid aerial-ground actors that could support data collection and monitoring tasks [21], [144]. These will mitigate increased processing delays due to computational processing-intensive tasks like visual identification and tracking of animals. However, UAV-based lower resolution images can pose challenges for computer vision analysis.

Overall, these issues represent lesser information for learning algorithms. Recent livestock counting and recognition techniques such as Fast-CNN and YOLO have indicated high detection accuracy but were unsatisfactory when trained with small targets using UAV images [113]. However, the multi-scale feature-based object detector SSD has demonstrated better detection performance than other well-known models. Notably, from the literature, most of these works have focused on using classification techniques that those applied deep object detection methods require modifications to detect small objects. The convolutional neural network-based livestock recognition frameworks embody several desirable characteristics that can handle recordings of different lengths and segment sizes, producing and adjusting secondary outputs. The network does not require additional pre-processing steps and can reach advanced audio and visual fusion systems. Faster R-CNN and YOLO are proven architectures to automate cattle detection and identification in pasture environments where livestock move free.

Deep neural network approaches trained on limited data can be fed with diverse training sets and avoid over-fitting using data augmentation and transfer learning techniques. The results in [54] showed that fine-tuned CNNs trained on augmented instances outperform fine-tuned CNNs trained on a set of the original dataset. However, these data augmentation methods demand high computational resources and filters. Generative Adversarial Nets (GANs) [145] performed unsupervised data generation. However, their deployment is complex and leads to a long computation time. The DL approaches employed above have strengths and limitations that need to be acquainted with mixed methods to bring together the differences and non-overlapping weaknesses. Extensively performed DNNs in PLF tasks require lots of training data. A novel less than one shot (LO-shot) learning method that learns from a smaller number of the training set in domains where little training data is available would be very beneficial [146]. Furthermore, livestock classification and localization using the attention-based audio-visual localization techniques described in [125], [147] demonstrate their feasible application. The multi-modal audio-visual transformers will improve upon the two-stream audio-visual neural networks.

These models utilize nearly the same data and deploy similar externally pre-trained embeddings that illustrate the power of the transformer architecture for the task at hand. Subsequently, many expect to see more approaches adopting LSTM or other RNN algorithms exploiting the time dimension to improve prediction performance. Also, the effect of label density (i.e., weekly labels) and corruption (e.g., noise in labels) are other limiting parameters for large-scale AED that needs to develop algorithms able to address these challenges [148]. While these solutions for livestock classification and localization using different UAV-based data sources (e.g., image, video, sound) can be explored in detail to minimize the data sample requirement. Additionally, the scalability of the above-discussed techniques on a large population should be investigated by acquiring more significant instances from an individual target discovering unique features exhibited by animals. In doing so, more generalized models will be available to transfer on unfamiliar farms and herds.

Other technical issues related to UAVs that may influence the data quality and formats are hardware maintenance and inadequate energy sources in fields with harsh weather conditions and remote areas. Hence, reliable hardware and energy-efficient solutions are desirable for the end devices [18].

VII. CONCLUSION

Early detection and prevention play a crucial role in the modern livestock farming domain. Ultimately, farmers require an appropriate, timely guideline to monitor and predict livestock position and behaviour in enormous pasture and rangeland environments. Deep learning techniques provide a clear understanding of the process by analysing a diverse set of data and elucidating the achieved information. In remote sensing, livestock detection, classification, and localization techniques based on deep computational networks have received increased attention over the last few years by advancing monitoring techniques, such as UAVs providing diverse sets of data. A recent family of object detection and semantic segmentation techniques (e.g., CNN) demonstrated higher precision for livestock monitoring in pasture and open space environments using UAV-based data. Faster R-CNN extensively tested for livestock detection has shown the most accurate results despite exploring the wide range of region-based object detectors. Livestock detection models based on YOLO demonstrated high performance in locating and classifying the object of interest within the image by drawing the bounding boxes and the class probabilities. However, they struggle with small objects within the images. To this end, Mask R-CNN extended the livestock classification and localization tasks to other instance segmentation solutions.

Further, the unified and self-reinforced R-CNN and attention-based mechanisms will hopefully inspire the real-time monitoring tasks using appearance and motion information from UAV-based audio-visual data. The present review also shows the potential application of audio-based perception for searching and livestock management for a person with vision disability (e.g., fog, fire). Several introduced combinations of DL (e.g., CNN) and feature extraction methods (e.g., MFCCs) apply to unique acoustic feature extraction from limited data available to evaluate livestock conditions and behaviour. However, most studies rely on visual perception, and a limited number of studies investigated audio-visual perception using UAVs. Despite all investigations, training robust DNNs, need large datasets that do not exist in many cases.

REFERENCES

[1] P. K. Thornton, "Livestock production: Recent trends, future prospects," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 365, no. 1554, pp. 2853–2867, Sep. 2010.
[2] A. Nasirahmadi, S. A. Edwards, and B. Sturm, "Implementation of machine vision for detecting behaviour of cattle and pigs," *Livestock Sci.*, vol. 202, pp. 25–38, Aug. 2017.
[3] M. Ariff and I. Ismail, "RFID application development for a livestock monitoring system," in *Bioresources Technology in Sustainable Agriculture: Biological and Biochemical Research*. New York, NY, USA: Apple Academic Press, 2018, p. 81.
[4] T. M. Brown-Brandl, F. Adrion, J. Maselyne, A. Kapun, E. F. Hessel, W. Saeys, A. Van Nuffel, and E. Gallmann, "A review of passive radio frequency identification systems for animal monitoring in livestock facilities," *Appl. Eng. Agricult.*, vol. 35, no. 4, pp. 579–591, 2019.
[5] B. Wolfger, B. W. Jones, K. Orsel, and J. M. Bewley, "Technical note: Evaluation of an ear-attached real-time location monitoring system," *J. Dairy Sci.*, vol. 100, no. 3, pp. 2219–2224, Mar. 2017.
[6] D. W. Bailey, M. G. Trotter, C. W. Knight, and M. G. Thomas, "Use of GPS tracking collars and accelerometers for rangeland livestock production research," *Transl. Animal Sci.*, vol. 2, no. 1, pp. 81–88, 2018.
[7] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, "An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3758–3773, Oct. 2018.
[8] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 25, Jun. 2018.
[9] S. Ward, J. Hensler, B. Alsalam, and L. F. Gonzalez, "Autonomous UAVs wildlife detection using thermal imaging, predictive navigation and computer vision," in *Proc. IEEE Aerosp. Conf.*, Mar. 2016, pp. 1–8.
[10] G. K. Verma and P. Gupta, "Wild animal detection using deep convolutional neural network," in *Proc. 2nd Int. Conf. Comput. Vis. Image Process.* Singapore: Springer, 2018, pp. 327–338.
[11] V. Sharma and R. N. Mir, "A comprehensive and systematic look up into deep learning based object detection techniques: A review," *Comput. Sci. Rev.*, vol. 38, Nov. 2020, Art. no. 100301.
[12] L. F. Gonzalez, G. A. Montes, E. Puig, S. Johnson, K. Mengersen, and K. J. Gaston, "Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation," *Sensors*, vol. 16, no. 1, p. 97, 2016.
[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.
[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*.

[15] S. M. M. Islam, C. Grado, V. Lubecke, and L. C. Lubecke, "UAV radar sensing of respiratory variations for COVID-type disorders," in *Proc. IEEE Asia–Pacific Microw. Conf. (APMC)*, Dec. 2020, pp. 737–739.

[16] R. García, J. Aguilar, M. Toro, A. Pinto, and P. Rodríguez, "A systematic literature review on the use of machine learning in precision livestock farming," *Comput. Electron. Agricult.*, vol. 179, Dec. 2020, Art. no. 105826.

[17] R. Dohmen, C. Catal, and Q. Liu, "Computer vision-based weight estimation of livestock: A systematic literature review," *New Zealand J. Agricult. Res.*, vol. 65, pp. 1–21, May 2021.

[18] N. Islam, M. M. Rashid, F. Pasandideh, B. Ray, S. Moore, and R. Kadel, "A review of applications and communication technologies for Internet of Things (IoT) and unmanned aerial vehicle (UAV) based sustainable smart farming," *Sustainability*, vol. 13, no. 4, p. 1821, Feb. 2021.

[19] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang, J. Gao, and L. Zhang, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, May 2020, Art. no. 111716.

[20] P. V. Torres-Carrion, C. S. Gonzalez-Gonzalez, S. Aciar, and G. Rodriguez-Morales, "Methodology for systematic literature review applied to engineering and education," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2018, pp. 1364–1373.

[21] F. de Lima Weber, M. I. Cagnin, D. M. B. Paiva, V. A. de Moraes Weber, S. R. de Medeiros, R. da Costa Gomes, and H. Pistori, "Use of computational vision and UAVs in livestock: A literature review," in *Proc. 14th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2019, pp. 1–7.

[22] S. Neethirajan and B. Kemp, "Digital livestock farming," *Sens. Bio-Sensing Res.*, vol. 32, Jun. 2021, Art. no. 100408.

[23] Y. Qiao, H. Kong, C. Clark, S. Lomax, D. Su, S. Eiffert, and S. Sukkarieh, "Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation," *Comput. Electron. Agricult.*, vol. 185, Jun. 2021, Art. no. 106143.

[24] F. Sarwar *et al.*, "Survey of livestock counting and tracking methods," in *Proc. Nusantara Sci. Technol.*, Indonesia, Nov. 2020, pp. 150–157.

[25] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[26] A. I. Awad, A. E. Hassanien, and H. M. Zawbaa, "A cattle identification approach using live captured muzzle print images," in *Proc. Int. Conf. Secur. Inf. Commun. Netw.* Berlin, Germany: Springer, 2013, pp. 143–152.

[27] G. Li, Y. Huang, Z. Chen, G. D. Chesser, J. L. Purswell, J. Linhoss, and Y. Zhao, "Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review," *Sensors*, vol. 21, no. 4, p. 1492, Feb. 2021.

[28] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.

[29] C. Yao, Y. Zhang, Y. Zhang, and H. Liu, "Application of convolutional neural network in the classification of high resolution agricultural remote sensing images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 2, pp. 989–992, Sep. 2017.

[30] W. Andrew, C. Greatwood, and T. Burghardt, "Visual localisation and individual identification of Holstein Friesian cattle via deep learning," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 2850–2859.

[31] J. Hendriks and P. Dumond, "Exploring the relationship between preprocessing and hyperparameter tuning for vibration-based machine fault diagnosis using CNNs," *Vibration*, vol. 4, no. 2, pp. 284–309, Apr. 2021.

[32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Piscataway, NJ, USA, Jun. 2016, pp. 779–788.

[33] T. Grill and J. Schluter, "Two convolutional neural networks for bird detection in audio signals," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 1764–1768.

[34] N. Singh, "Classification of animal sound using convolutional neural network," M.S. dissertation, AAROW Technol. Univ. Dublin, 2020.

[35] J. Chen, Q. Wu, D. Liu, and T. Xu, "Foreground-background imbalance problem in deep object detectors: A review," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, China, 2020, pp. 285–290.

[36] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[37] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[38] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[39] J. G. A. Barbedo, L. V. Koenigkan, P. M. Santos, and A. R. B. Ribeiro, "Counting cattle in UAV images—Dealing with clustered animals and animal/background contrast changes," *Sensors*, vol. 20, no. 7, p. 2126, Apr. 2020.

[40] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.

[41] R. Girshick, "Fast R-CNN object detection with caffe," in *Proc. 13th Eur. Conf. Comput. Vis.*, Oct. 2014, pp. 6233–6236.

[42] X. Wu, D. Sahoo, and S. C. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, Jul. 2020.

[43] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[44] H. Hong, D. L. Yamins, N. J. Majaj, and J. J. DiCarlo, "Explicit information for category-orthogonal object properties increases along the ventral stream," *Nature Neurosci.*, vol. 19, no. 4, p. 613, 2016.

[45] Q. Zhang, Y. Liu, C. Gong, Y. Chen, and H. Yu, "Applications of deep learning for dense scenes analysis in agriculture: A review," *Sensors*, vol. 20, no. 5, p. 1520, Mar. 2020.

[46] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019.

[47] S. V. Stehman, "Selecting and interpreting measures of thematic classification accuracy," *Remote Sens. Environ.*, vol. 62, no. 1, pp. 77–89, 1997.

[48] P. Chamoso, W. Raveane, V. Parra, and A. Gonzalez, "UAVs applied to the counting and monitoring of animals," in *Ambient Intelligence-Software and Applications*. Cham, Switzerland: Springer, 2014, pp. 71–80.

[49] Y. Qiao, M. Truman, and S. Sukkarieh, "Cattle segmentation and contour extraction based on mask R-CNN for precision livestock farming," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104958.

[50] W. Andrew, J. Gao, S. Mullan, N. Campbell, A. W. Dowsey, and T. Burghardt, "Visual identification of individual Holstein-Friesian cattle via deep metric learning," *Comput. Electron. Agricult.*, vol. 185, Jun. 2021, Art. no. 106133.

[51] B. Xu, W. Wang, G. Falzon, P. Kwan, L. Guo, G. Chen, A. Tait, and D. Schneider, "Automated cattle counting using mask R-CNN in quadcopter vision system," *Comput. Electron. Agricult.*, vol. 171, Apr. 2020, Art. no. 105300.

[52] B. Xu, W. Wang, G. Falzon, P. Kwan, L. Guo, Z. Sun, and C. Li, "Livestock classification and counting in quadcopter aerial images using mask R-CNN," *Int. J. Remote Sens.*, vol. 41, no. 21, pp. 8121–8142, Nov. 2020.

[53] W. Shao, R. Kawakami, R. Yoshihashi, S. You, H. Kawase, and T. Naemura, "Cattle detection and counting in UAV images based on convolutional neural networks," *Int. J. Remote Sens.*, vol. 41, no. 1, pp. 31–52, Jan. 2020.

[54] E. Okafor, L. Schomaker, and M. A. Wiering, "An analysis of rotation matrix and colour constancy data augmentation in classifying images of animals," *J. Inf. Telecommun.*, vol. 2, no. 4, pp. 465–491, Oct. 2018.

[55] F. Sarwar, A. Griffin, P. Periasamy, K. Portas, and J. Law, "Detecting and counting sheep with a convolutional neural network," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2018, pp. 1–6.

[56] J. G. A. Barbedo, L. V. Koenigkan, T. T. Santos, and P. M. Santos, "A study on the detection of cattle in UAV images using deep learning," *Sensors*, vol. 19, no. 24, p. 5436, Dec. 2019.

[57] J. R. Alvarez, M. Arroqui, P. Mangudo, J. Toloza, D. Jatip, J. M. Rodríguez, A. Teyseyre, C. Sanz, A. Zunino, C. Machado, and C. Mateos, "Body condition estimation on cows from depth images using convolutional neural networks," *Comput. Electron. Agric.*, vol. 155, pp. 12–22, Dec. 2018.

[58] R. Liu, M. Kuffer, and C. Persello, "The temporal dynamics of slums employing a CNN-based change detection approach," *Remote Sens.*, vol. 11, no. 23, p. 2844, Nov. 2019.

[59] L. Han, P. Tao, and R. R. Martin, "Livestock detection in aerial images using a fully convolutional network," *Comput. Vis. Media*, vol. 5, no. 2, pp. 221–228, Jun. 2019.

[60] B. Kellenberger, D. Marcos, and D. Tuia, "Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning," *Remote Sens. Environ.*, vol. 216, pp. 139–153, Oct. 2018.

[61] Y. Alghamdi, A. Munir, and H. M. La, "Architecture, classification, and applications of contemporary unmanned aerial vehicles," *IEEE Consum. Electron. Mag.*, vol. 10, no. 6, pp. 9–20, Nov. 2021.

[62] S. Petkovic, D. Petkovic, and A. Petkovic, "IoT devices vs. drones for data collection in agriculture," in *DAAAM International Scientific Book*, vol. 16, Vienna, Austria, 2017, pp. 63–80.

[63] H. Pathak, G. Kumar, S. Mohapatra, B. Gaikwad, and J. Rane, "Use of drones in agriculture: Potentials, problems and policy needs," ICAR-Nat. Inst. Abiotic Stress Manage., India, Tech. Rep. 300, 2020.

[64] F. Santoso, M. A. Garratt, and S. G. Anavatti, "State-of-the-art intelligent flight control systems in unmanned aerial vehicles," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 2, pp. 613–627, Apr. 2018.

[65] J. Yang, H. Wang, L. Feng, X. Yan, H. Zheng, W. Zhang, and Z. Liu, "Semantically coherent out-of-distribution detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8301–8309.

[66] F. N. Fote, A. Roukh, S. Mahmoudi, S. A. Mahmoudi, and O. Debauche, "Toward a big data knowledge-base management system for precision livestock farming," *Proc. Comput. Sci.*, vol. 177, pp. 136–142, Jan. 2020.

[67] D. M. Anderson, R. E. Estell, and A. F. Cibils, "Spatiotemporal cattle data—A plea for protocol standardization," *Positioning*, vol. 4, no. 1, pp. 115–136, 2013.

[68] V. Moysiadis, P. Sarigiannidis, V. Vitsas, and A. Khelifi, "Smart farming in Europe," *Comput. Sci. Rev.*, vol. 39, Feb. 2021, Art. no. 100345.

[69] J. Wang, Z. He, G. Zheng, S. Gao, and K. Zhao, "Development and validation of an ensemble classifier for real-time recognition of cow behavior patterns from accelerometer data and location data," *PLoS ONE*, vol. 13, no. 9, Sep. 2018, Art. no. e0203546.

[70] C. Carslake, J. A. Vázquez-Diosdado, and J. Kaler, "Machine learning algorithms to classify and quantify multiple behaviours in dairy calves using a sensor: Moving beyond classification in precision livestock," *Sensors*, vol. 21, no. 1, p. 88, Dec. 2020.

[71] E. Walton, C. Casey, J. Mitsch, J. A. Vázquez-Diosdado, J. Yan, T. Dottorini, K. A. Ellis, A. Winterlich, and J. Kaler, "Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour," *Roy. Soc. Open Sci.*, vol. 5, no. 2, Feb. 2018, Art. no. 171442.

[72] P. C. P. Khanh, D.-T. Tran, T. Van Duong, N. H. Thinh, and D.-N. Tran, "The new design of cows' behavior classifier based on acceleration data and proposed feature set," *Math. Biosci. Eng.*, vol. 17, no. 4, pp. 2760–2780, 2020.

[73] A. Ter-Sarkisov, R. Ross, and J. Kelleher, "Bootstrapping labelled dataset construction for cow tracking and behavior analysis," in *Proc. 14th Conf. Comput. Robot Vis. (CRV)*, May 2017, pp. 277–284.

[74] P. P. Nielsen, "Automatic registration of grazing behaviour in dairy cows using 3D activity loggers," *Appl. Animal Behav. Sci.*, vol. 148, nos. 3–4, pp. 179–184, Oct. 2013.

[75] F. X. Wang, D. F. Shao, S. L. Li, Y. J. Wang, A. Azarfar, and Z. J. Cao, "Effects of stocking density on behavior, productivity, and comfort indices of lactating dairy cows," *J. Dairy Sci.*, vol. 99, no. 5, pp. 3709–3717, May 2016.

[76] D. S. Rodriguez-Baena, F. A. Gomez-Vela, M. García-Torres, F. Divina, C. D. Barranco, N. Daz-Diaz, M. Jimenez, and G. Montalvo, "Identifying livestock behavior patterns based on accelerometer dataset," *J. Comput. Sci.*, vol. 41, Mar. 2020, Art. no. 101076.

[77] N. Al-Thani, A. Albuainain, F. Alnaimi, and N. Zorba, "Drones for sheep livestock monitoring," in *Proc. IEEE 20th Medit. Electrotech. Conf. (MELECON)*, Jun. 2020, pp. 672–676.

[78] J. G. A. Barbedo, L. V. Koenigkan, and P. M. Santos, "Cattle detection using oblique UAV images," *Drones*, vol. 4, no. 4, p. 75, Dec. 2020.

[79] Y. Qiao, C. Cappelle, Y. Ruichek, and T. Yang, "ConvNet and LSH-based visual localization using localized sequence matching," *Sensors*, vol. 19, no. 11, p. 2439, May 2019.

[80] A. Ter-Sarkisov, R. Ross, J. Kelleher, B. Earley, and M. Keane, "Beef cattle instance segmentation using fully convolutional neural network," 2018, *arXiv:1807.01972*.

[81] L. Mou and X. X. Zhu, "Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6699–6711, Nov. 2018.

[82] C. Bahlo, P. Dahlhaus, H. Thompson, and M. Trotter, "The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review," *Comput. Electron. Agricult.*, vol. 156, pp. 459–466, Jan. 2019.

[83] W. Andrew, S. Hannuna, N. Campbell, and T. Burghardt, "Automatic individual Holstein Friesian cattle identification via selective local coat pattern matching in RGB-D imagery," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 484–488.

[84] C. J. Torney, A. P. Dobson, F. Borner, D. J. Lloyd-Jones, D. Moyer, H. T. Maliti, M. Mwita, H. Fredrick, M. Borner, and J. G. C. Hopcraft, "Assessing rotation-invariant feature classification for automated wildebeest population counts," *PLoS ONE*, vol. 11, no. 5, May 2016, Art. no. e0156342.

[85] M. A. Olivares-Mendez, C. Fu, P. Ludivig, T. F. Bissyande, S. Kannan, M. Zurad, A. Annaiyan, H. Voos, and P. Campoy, "Towards an autonomous vision-based unmanned aerial system against wildlife poachers," *Sensors*, vol. 15, no. 12, p. 31 362-31 391, 2015.

[86] N. Rey, M. Volpi, S. Joost, and D. Tuia, "Detecting animals in African Savanna with UAVs and the crowds," *Remote Sens. Environ.*, vol. 200, pp. 341–351, Oct. 2017.

[87] Z. Ünal, "Smart farming becomes even smarter with deep learning—A bibliographical analysis," *IEEE Access*, vol. 8, pp. 105587–105609, 2020.

[88] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Aug. 2018.

[89] D. Wang, Q. Song, X. Liao, H. Ye, Q. Shao, J. Fan, N. Cong, X. Xin, H. Yue, and H. Zhang, "Integrating satellite and unmanned aircraft system (UAS) imagery to model livestock population dynamics in the longbao wetland national nature reserve, China," *Sci. Total Environ.*, vol. 746, Dec. 2020, Art. no. 140327.

[90] J. C. van Gemert, C. R. Verschoor, P. Mettes, K. Epema, L. P. Koh, and S. Wich, "Nature conservation drones for automatic localization and counting of animals," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 255–270.

[91] J. T. Mufford, D. J. Hill, N. J. Flood, and J. S. Church, "Use of unmanned aerial vehicles (UAVs) and photogrammetric image analysis to quantify spatial proximity in beef cattle," *J. Unmanned Vehicle Syst.*, vol. 7, no. 3, pp. 194–206, Sep. 2019.

[92] S. Nyamuryekung'e, A. F. Cibils, R. E. Estell, and A. L. Gonzalez, "Use of an unmanned aerial vehicle-mounted video camera to assess feeding behavior of Raramuri Criollo cows," *Rangeland Ecol. Manage.*, vol. 69, no. 5, pp. 386–389, Sep. 2016.

[93] J. Tao, L. Shao, Z. Guan, W. Ho, and S. Talluri, "Incorporating risk aversion and fairness considerations into procurement and distribution decisions in a supply chain," *Int. J. Prod. Res.*, vol. 58, no. 7, pp. 1950–1967, Apr. 2020.

[94] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[95] J. G. A. Barbedo and L. V. Koenigkan, "Perspectives on the use of unmanned aerial systems to monitor cattle," *Outlook Agricult.*, vol. 47, no. 3, pp. 214–222, Sep. 2018.

[96] Y. Sun, S. Yi, F. Hou, D. Luo, J. Hu, and Z. Zhou, "Quantifying the dynamics of livestock distribution by unmanned aerial vehicles (UAVs): A case study of yak grazing at the household scale," *Rangeland Ecol. Manage.*, vol. 73, no. 5, pp. 642–648, Sep. 2020.

[97] R.-W. Bello, A. S. A. Mohamed, and A. Z. Talib, "Contour extraction of individual cattle from an image using enhanced mask R-CNN instance segmentation method," *IEEE Access*, vol. 9, pp. 56984–57000, 2021.

[98] F. Sarwar, A. Griffin, S. U. Rehman, and T. Pasang, "Detecting sheep in UAV images," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106219.

[99] A. Rivas, P. Chamoso, A. González-Briones, and J. M. Corchado, "Detection of cattle using drones and convolutional neural networks," *Sensors*, vol. 18, no. 7, p. 2048, 2018.

[100] W. Andrew, C. Greatwood, and T. Burghardt, "Aerial animal biometrics: Individual Friesian cattle recovery and visual identification via an autonomous UAV with onboard deep inference," 2019, *arXiv:1907.05310*.

[101] W. Andrew, J. Gao, S. Mullan, N. Campbell, A. W. Dowsey, and T. Burghardt, "Visual identification of individual Holstein-Friesian cattle via deep metric learning," *Comput. Electron. Agricult.*, vol. 185, 2021, Art. no. 106133

[102] B. Kellenberger, M. Volpi, and D. Tuia, "Fast animal detection in UAV images using convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 866–869.

[103] P. Tassinari, M. Bovo, S. Benni, S. Franzoni, M. Poggi, L. M. E. Mammi, S. Mattoccia, L. Di Stefano, F. Bonora, A. Barbaresi, E. Santolini, and D. Torreggiani, "A computer vision approach based on deep learning for the detection of dairy cows in free stall barn," *Comput. Electron. Agricult.*, vol. 182, Mar. 2021, Art. no. 106030.

[104] P. Ghosh, S. Mustafi, and S. N. Mandal, "Image-based goat breed identification and localization using deep learning," *Int. J. Comput. Vis. Image Process.*, vol. 10, no. 4, pp. 74–96, Oct. 2020.

[105] Y. Qiao, D. Su, H. Kong, S. Sukkarieh, S. Lomax, and C. Clark, "Data augmentation for deep learning based cattle segmentation in precision livestock farming," in *Proc. IEEE 16th Int. Conf. Automat. Sci. Eng. (CASE)*, Aug. 2020, pp. 979–984.

[106] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T.-S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2196–2210, Apr. 2017.

[107] R. Dong, D. Xu, J. Zhao, L. Jiao, and J. An, "Sig-NMS-based faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8534–8545, Nov. 2019.

[108] J. Pang, C. Li, J. Shi, Z. Xu, and H. Feng, "R2-CNN: Fast tiny ob-ject detection in large-scale remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5512–5524, 2019.

[109] Y. Liu, P. Sun, N. Wergeles, and Y. Shang, "A survey and performance evaluation of deep learning methods for small object detection," *Expert Syst. Appl.*, vol. 172, Jun. 2021, Art. no. 114602.

[110] M. Rahnemoonfar, D. Dobbs, M. Yari, and M. J. Starek, "DisCountNet: Discriminating and counting network for real-time counting and localization of sparse objects in high-resolution UAV imagery," *Remote Sens.*, vol. 11, no. 9, p. 1128, May 2019.

[111] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, T. Darrell, and K. Saenko, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2625–2634.

[112] J. Xu, G. Solmaz, R. Rahmatizadeh, D. Turgut, and L. Boloni, "Internet of Things applications: Animal monitoring with unmanned aerial vehicle," 2016, *arXiv:1610.05287*.

[113] M. Razaak, H. Kerdegari, E. Davies, R. Abozariba, M. Broadbent, K. Mason, V. Argyriou, and P. Remagnino, "An integrated precision farming application based on 5G, UAV and deep learning technologies," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Cham, Switzerland: Springer, 2019, pp. 109–119.

[114] K. Qazi, Z. Mehmood, T. Nawaz, and H. Habib, "Goat flock surveillance: A video analytics framework using a quadcopter," *Pakistan J. Sci.*, vol. 70, no. 1, pp. 63–70, 2018.

[115] M. Parikh, M. Patel, and D. Bhatt, "Animal detection using template matching algorithm," *Int. J. Res. Mod. Eng. Emerg. Technol.*, vol. 1, no. 3, pp. 26–32, 2013.

[116] I. Halachmi, M. Guarino, J. Bewley, and M. Pastell, "Smart animal agriculture: Application of real-time sensors to improve animal well-being and production," *Annu. Rev. Animal Biosci.*, vol. 7, pp. 403–425, Feb. 2019.

[117] B. Meunier, P. Pradel, K. H. Sloth, C. Cirié, E. Delval, M. M. Mialon, and I. Veissier, "Image analysis to refine measurements of dairy cow behaviour from a real-time location system," *Biosystems Eng.*, vol. 173, pp. 32–44, Sep. 2018.

[118] F. Maroto-Molina, J. Navarro-García, K. Príncipe-Aguirre, I. Gómez-Maqueda, J. E. Guerrero-Ginel, A. Garrido-Varo, and D. C. Pérez-Marín, "A low-cost IoT-based system to monitor the location of a whole herd," *Sensors*, vol. 19, no. 10, p. 2298, May 2019.

[119] T. Jikyo, T. Yamanishi, T. Kamada, R. Nishide, C. Ohta, K. Oyama, and T. Ohkawa, "A study on outdoor localization method based on deep learning using model-based received power estimation data of low power wireless tag," *IEICE Commun. Exp.*, vol. 8, no. 12, pp. 524–529, 2019.

[120] Q. Li, W. Li, W. Sun, J. Li, and Z. Liu, "Fingerprint and assistant nodes based Wi-Fi localization in complex indoor environment," *IEEE Access*, vol. 4, pp. 2993–3004, 2016.

[121] J. Trogh, D. Plets, L. Martens, and W. Joseph, "Bluetooth low energy-based location tracking for livestock monitoring," in *Proc. 8th Eur. Conf. Precis. Livestock Farming (EC-PLF)*, 2017, pp. 469–475.

[122] V. Bloch and M. Pastell, "Monitoring of cow location in a barn by an open-source, low-cost, low-energy Bluetooth tag system," *Sensors*, vol. 20, no. 14, p. 3841, Jul. 2020.

[123] T.-Y. Lin, Y. Cui, S. Belongie, and J. Hays, "Learning deep representations for ground-to-aerial geolocalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5007–5015.

[124] Y. Fang, S. Du, R. Abdoola, K. Djouani, and C. Richards, "Motion-based animal detection in aerial videos," *Proc. Comput. Sci.*, vol. 92, pp. 13–17, Jan. 2016.

[125] J. Martinez-Carranza and C. Rascon, "A review on auditory perception for unmanned aerial vehicles," *Sensors*, vol. 20, no. 24, p. 7276, Dec. 2020.

[126] T. Norton and D. Berckmans, "Developing precision livestock farming tools for precision dairy farming," *Animal Frontiers*, vol. 7, no. 1, pp. 18–23, Jan. 2017.

[127] L. Fanioudakis and I. Potamitis, "Deep networks tag the location of bird vocalisations on audio spectrograms," 2017, *arXiv:1711.04347*.

[128] D.-H. Jung, N. Y. Kim, S. H. Moon, C. Jhin, H.-J. Kim, J.-S. Yang, H. S. Kim, T. S. Lee, J. Y. Lee, and S. H. Park, "Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering," *Animals*, vol. 11, no. 2, p. 357, Feb. 2021.

[129] E. Sasmaz and F. B. Tek, "Animal sound classification using a convolutional neural network," in *Proc. 3rd Int. Conf. Comput. Sci. Eng. (UBMK)*, Sep. 2018, pp. 625–629.

[130] W. Xu, X. Zhang, L. Yao, W. Xue, and B. Wei, "A multi-view CNN-based acoustic classification system for automatic animal species identification," *Ad Hoc Netw.*, vol. 102, May 2020, Art. no. 102115.

[131] Y. R. Pandeya, D. Kim, and J. Lee, "Domestic cat sound classification using learned features from deep neural nets," *Appl. Sci.*, vol. 8, no. 10, p. 1949, Oct. 2018.

[132] Z. Zhao, S.-H. Zhang, Z.-Y. Xu, K. Bellisario, N.-H. Dai, H. Omrani, and B. C. Pijanowski, "Automated bird acoustic event detection and robust species classification," *Ecol. Informat.*, vol. 39, pp. 99–108, May 2017.

[133] J. Ramaswamy and S. Das, "See the sound, hear the pixels," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2970–2979.

[134] C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised visual representation learning by context prediction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1422–1430.

[135] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[136] R. Arandjelovic and A. Zisserman, "Objects that sound," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 435–451.

[137] H. Zhu, M.-D. Luo, R. Wang, A.-H. Zheng, and R. He, "Deep audio-visual learning: A survey," *Int. J. Autom. Comput.*, vol. 18, no. 3, pp. 1–26, 2021.

[138] V. Giovanetti, M. Decandia, G. Molle, M. Acciaro, M. Mameli, A. Cabiddu, R. Cossu, M. G. Serra, C. Manca, S. P. G. Rassu, and C. Dimauro, "Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer," *Livestock Sci.*, vol. 196, pp. 42–48, Feb. 2017.

[139] F. A. P. Alvarenga, I. Borges, L. Palkovič, J. Rodina, V. H. Oddy, and R. C. Dobos, "Using a three-axis accelerometer to identify and classify sheep behaviour at pasture," *Appl. Animal Behav. Sci.*, vol. 181, pp. 91–99, Aug. 2016.

[140] J. Marais, S. P. Le Roux, R. Wolhuter, and T. Niesler, "Automatic classification of sheep behaviour using 3-axis accelerometer data," in *Proc. 25th Annu. Symp. Pattern Recognit. Assoc. South Afr. (PRASA)*, 2014, pp. 97–102.

[141] M. Radeski and V. Ilieski, "Gait and posture discrimination in sheep using a tri-axial accelerometer," *Animal*, vol. 11, no. 7, pp. 1249–1257, 2017.

[142] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2014.

[143] A. Mukherjee, S. Misra, A. Sukrutha, and N. S. Raghuwanshi, "Distributed aerial processing for IoT-based edge UAV swarms in smart farming," *Comput. Netw.*, vol. 167, Feb. 2020, Art. no. 107038.

[144] A. Ronzhin, G. Rigoll, and R. Meshcheryakov, *Interactive Collaborative Robotics*. Berlin, Germany: Springer, 2016.

[145] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*.

[146] I. Sucholutsky and M. Schonlau, "'Less than one'-shot learning: Learning n classes from M<N samples," 2020, *arXiv:2009.08449*.

[147] Y.-B. Lin and Y.-C. F. Wang, "Audiovisual transformer with instance attention for audio-visual event localization," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 1–17.

[148] A. Shah, A. Kumar, A. G. Hauptmann, and B. Raj, "A closer look at weak label learning for audio events," 2018, *arXiv:1804.09288*.

D. B. MAMEHGOL YOUSEFI received the Diploma degree in experimental sciences in Golestan, Iran, in 2009, the Bachelor of Agricultural Engineering degree from Islamic Azad University, Gorgan, in 2013, and the Master of Innovation and Design Engineering degree from Universiti Putra Malaysia, in 2018. Upon completion of her master's degree, in 2017, she continued her research with the Department of Aerospace Engineering, Universiti Putra Malaysia, where she is a Research Assistant. Her research interest includes multi-disciplinary design optimization techniques using modeling approaches.

A. S. MOHD RAFIE received the Diploma degree, the B.Eng. degree in mechanical engineering and aeronautic, the M.Eng. degree in aeronautic engineering, and the Ph.D. degree in aerospace engineering, under the tutor scheme, from Universiti Teknologi Malaysia, in 1992, 1998, 2002, and 2007, respectively. He was with Universiti Putra Malaysia as a Research Assistant, from 1998 to 2001. After his M.Eng. degree, he joined Universiti Putra Malaysia as a Tutor, where he is a Lecturer with the Aerospace Engineering Department. His research is dedicated to experimental aerodynamics and aeroelasticity. His current research interests include developing experimental aerodynamics rotating cylinder for wind harvesting, drag reduction mechanism, and vertical lifting mechanism.

S. A. R. AL-HADDAD (Senior Member, IEEE) received the Ph.D. degree in electrical, electronic, and systems engineering from the National University of Malaysia, specialized in sound processing, bio signal data processing, artificial intelligence, and big data. He taught students for undergraduate and graduate and managed to get international and national grants. He is a Professor with the Department of Computer and Communications Systems Engineering, Universiti Putra Malaysia. Further than that, he has few patents and copyrights and actively join professional society such as the Malaysian IEEE Section and the Systems Man and Cybernetics Society as the Malaysia Chapter Past Chair.

SYARIL AZRAD received the Bachelor of Engineering and the Master of Engineering degrees in mechanical engineering from the Tokyo University of Science, in 2000 and 2002, respectively, and the Ph.D. degree in mechanical engineering from Chiba University, in 2012. He is currently a Senior Lecturer with the Department of Aerospace Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM). He is engaged in various international and national projects and collaborations funded by the Ministry of Higher Education Malaysia, international universities, and companies. His research interests include unmanned aerial vehicle control systems, high altitude balloons and platforms, cube-sat systems, vision-based control, and systems engineering. He is the regional facilitator for HEPTA-Sat training for Malaysia and a member of the University Space Engineering Consortium, UNISEC-Global, Japan, which is envisioned to enable students from all countries to participate in space-related projects, by 2030. He is also a Committee Member for Southeast Asia Network in Aerospace Engineering (SANAE), which organizes yearly workshops and research collaboration between aerospace-related education and research institutions. He is the contact point for Belt and Road Aerospace Innovation Alliance (BRAIA) for UPM.