Article

Not peer-reviewed version

Highly Accurate and Reliable Tracker for UAV-Based Herd Monitoring

Wei Luo , Guoqing Zhang , Quanqin Shao , Xiaoliang Li , Zhiguo Wang * , Xia Zhu , Zihui Zhao , Longfang Duan , Ke Liu , Dongliang Wang , Xiongyi Zhang , Yongxiang Zhao , Jiandong Liu , Zhongde Yu

Posted Date: 23 June 2023

doi: 10.20944/preprints202306.1669.v1

Keywords: Livestock monitoring; Open source UAV; Depth sorting; Kalman filter; Optical flow; Visual servo

Article

Highly Accurate and Reliable Tracker for UAV-Based Herd Monitoring

Wei Luo [1,2,3,4], Guoqing Zhang [1], Quanqin Shao [2,5], Xiaoliang Li [1], Zhiguo Wang [6,7,*], Xia Zhu [1,3,4], Zihui Zhao [1,3,4], Longfang Duan [1,3,4], Ke Liu [1,3,4], Dongliang Wang [2], Xiongyi Zhang [2], Yongxiang Zhao [1], Jiandong Liu [1] and Zhongde Yu [1]

[1] North China Institute of Aerospace Engineering, Langfang 065000, China
[2] Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[3] Aerospace Remote Sensing Information Processing and Application Collaborative Innovation Center of Hebei Province, Langfang 065000, China
[4] National Joint Engineering Research Center of Space Remote Sensing Information Application Technology, Langfang 065000, China
[5] University of Chinese Academy of Sciences, Beijing 101407, China
[6] GuangDong polytechnic of Science and Technology, School of Internet of Things Engineering, Guangzhou 510640, China
[7] St.Paul University Philippines,School of Information Technology and Engineering,Tuguegarao 3500, Philippines
* Correspondence: mralex119@163.com

Abstract: It is a challenging and meaningful task to carry out drone-based livestock monitoring in high-altitude and cold regions. The purpose of AI is to execute automated tasks and to solve practical problems in actual applications by combining the software technology with the hardware carrier to create integrated advanced devices. Only in this way, the maximum value of AI could be realized. In this paper, a real-time tracking system with dynamic target tracking ability is proposed. It is developed based on the tracking-by-detection architecture using YOLOv7 and DeepSORT algorithms for target detection and tracking, respectively. To address the existing problems of the DeepSORT algorithm, the following two optimizations are made: (1) Optical flow is used to compensate the Kalman filter for improvement of the prediction accuracy; (2) A low-confidence trajectory filtering method is adopted to reduce the influence of unreliable detection on target tracking. In addition, an visual servo controller for the UAV is designed to enable the automated tracking task. Finally, the system is tested using the Tibetan yaks living in the Tibetan Plateau as the tracking targets, and the results reveal the real-time multiple tracking ability and the ideal visual servo effect of the proposed system.

Keywords: livestock monitoring; open source UAV; depth sorting; kalman filter; optical flow; visual servo

1. Introduction

The fourth agricultural revolution, also known as Agriculture 4.0, aims at improving productivity, efficiency, quality, and resilience of agricultural systems, as well as reducing environmental impacts, resource use, and costs [7]. It is a new technology revolution in agriculture supported by policy-makers around the world. It refers to the utilization of advanced technologies including artificial intelligence (AI), biotechnology, big data, internet of things (IoT), and robotics to achieve sustainable agriculture development [27]. These technologies have been used to improve various aspects of agriculture such as precision farming, smart irrigation, and crop monitoring [15,48].

 doi:10.20944/preprints202306.1669.v1

Traditional animal husbandry usually uses manual operation to identify and track moving herds to obtain real-time information of the herd status, which is a time-consuming, laborious and inefficient method, especially in high-altitude pastoral areas with poor natural conditions. With the emergence of agriculture 4.0 and the rapid development of the UAV technology, autonomous UAV equipped with embedded computer can conduct real-time tracking and individual differentiation of herds without human intervention [42–44] to achieve the substantive progress of intelligent grazing. However, it is a challenging task to change the tracking targets from individual animals to groups with the help of AI algorithms, and to realize offline real-time monitoring using autonomous drones.

Currently, drones have become a popular and powerful tool for monitoring wildlife and studying animal behavior, as they can rapidly cover large and inaccessible areas, reduce human risk and disturbance, as well as provide high-resolution imageries and videos of wildlife [11,43,45,71]. In addition, UAVs have been used for a variety of animal research, such as mapping habitats, estimating species abundance and distribution, monitoring individual and group behaviour, measuring physiological parameters, assisting in anti-poaching efforts and studying anti-predator responses [14,55]. Drones can also help dairy farmers manage their livestock, including counting cows, detecting health issues, monitoring grazing patterns, tracking lost cows, and improving security[33,36].

In recent years, deep learning has also been applied to various domains of animal science, such as wildlife ecology, conservation biology, animal behaviour, animal welfare and animal breeding [28,49,51]. One of the main advantages of deep learning is its ability to automatically extract features and patterns from large and complex datasets, including images, videos, sounds, texts, etc. This can reduce the needs for human intervention, manual annotation and domain-specific knowledge, and simultaneously improve the efficiency and accuracy of data analysis. For example, deep learning can automatically identify, describe and count wildlife in the camera-trap images, which are widely used for monitoring wild animal populations and habitats [49]. Besides, deep learning can also automatically detect and track animal movements and postures in videos, which are valuable information for scientific study of animal behaviour and welfare [1,19,51,60–62].

Object detection based on deep learning architectures can be categorized into fast detection, shortcut connection and region-based networks. These networks are effective from perspectives of processing speed, accuracy and so on. Therefore, they are widely used for animal farming. Particularly, SSD and YOLO V3 have the advantage of processing speed, and R-CNN is advantageous with respect to the processing speed as well as the accuracy, so that they are the mostly applied networks currently. In relatively simple cases such as optimal lighting and clear view, the combination of different shallow networks (e.g., VGG + CNN) might achieve satisfactory performance [4]. However, in complex scenarios such as real commercial environments, to enhance the model capacity for sufficient environmental variations, it is necessary to combine multiple networks, for example, UNet + Inception V4 [23], and VGGNet + SSD [70]. Besides, even for the same model, a parallel combination to create the two-streamed connection could also improve the detection performance [36,72].

Multi-object tracking (MOT) typically refers to the detection, the recognition, and the tracking of multiple objects (e.g., pedestrians, cars, and animals) in videos without prior knowledge of the target number. Different targets have different IDs to achieve subsequent trajectory prediction, accurate search, and other tasks. In recent years, various MOT methods have been proposed and widely applied, such as monitoring [30], traffic monitoring [8], autonomous driving [18], and animal monitoring, aiming at object collision avoidance [39] or target tracking [26]. However, due to the crowded environments and the occluded objects, the results of MOT could be influenced by the difficult problem configuration, which leads to performance limitations in such scenarios. In addition, due to the extensive application of MOT methods, the importance of the MOT is still a challenging subject for the relevant research [13,63,67].

In recent years, with the rapid development of deep learning technology, the target detection performance has been significantly improved. With the emergence of the deep learning-based object detectors, tracking through detection has already become the most-focused method in MOT research

[67]. This method utilizes the knowledge of object location to establish an model that can associate with objects over time. In recent studies, the algorithm of Kalman Filter (KF) has been used as the motion model to improve the object correlation over time [9,12,24,64]. In 2016, SORT was proposed [9], which applies KF to estimate the object states and associates KF prediction with new object detection using the Hungarian algorithm [34]. One year later, an optimized Deep SORT was proposed by Wojke et al. [64], which includes a new cascading association procedure using the object appearance characteristics based on CNN. In this data association algorithm, the similarity of the object appearance characteristics and the Mahalanobis distances between the object states are combined, and the SORT data association is used in the later stage of mismatched states. Despite using CNN, high frame rates on target tracking benchmarks were achieved with the Deep SORT approach. Chen et al. proposed an algorithm similar to the Deep SORT, namely MOTDT [12], which employs a scoring function completely based on CNN to optimally select candidates. The Euclidean distance in the extracted object appearance characteristics is also adopted to optimize the association steps. Recently, He et al., [24] proposed a GMT-CT algorithm, which combines deep feature learning and graph partitioning. The graph is constructed using extracted object appearance characteristics for association steps to more accurately model the correlation between the measurements and the trajectories.

With the rapid development of autonomous UAVs, the abilities of UAVs for low-speed flying, hovering, laterally flying, and maneuvering in confined spaces make visual servo control a promising platform for performing tasks such as inspection, surveillance, and monitoring. In recent years, various studies have been conducted on the visual servo control of the UAVs, including quadrotors [22,54], airships [5] and drones [65]. Strategies for navigation and control of UVAs using only vision with feedback loops for monitoring known objects were proposed previously [10]. Stability control methods for quadrotor helicopters using vision as the primary sensor were also reported [2]. In this work, the helicopter attitude is estimated and used for vehicle control. Some studies on vision-based autonomous flight have been reported previously [3,52,57]. Among the different visual servo control models based on images, adaptive control [69], PID control [56], sliding model control [46] and neural network control [21] have been mainly used to enable one-camera UAVs to explore environments and avoid obstacles. Especially, the PID controller has very wide application in this field due to its high robustness.

DeepSORT As a multi-target tracking method with a competitive advantage, we propose a method with appearance features based on the Kalman filter and the Hungarian algorithm. The motion model is built using a Kalman filter and the correlation between objects is optimally solved by the Hungarian algorithm. The proposed algorithm can achieve excellent performance at real-time speed. However, when the motion of the object is complex and the detection of the object in the current frame is lost, the bounding box predicted by the Kalman filter cannot match the input. Based on the above motivation, we compensate the Kalman filter with optical flow to overcome the problem just discussed.

We may note that, in Deep SORT, detection confidence thresholds are used to filter all detections with low confidence. This is based on the assumption that tests with confidence below the threshold are likely to be a false positive, and tests with confidence above the threshold should be a true positive. However, state-of-the-art assays have not fully followed this hypothesis. Therefore, we used a low-confidence trajectory filtering extension in Deep SORT that which average detection confidence within the first few frames after initialization. Trajectories with low average confidence were filtered out to reduce false positive trajectories. Average confidence prevents missing true correct tracks with little low confidence detection caused by occlusion or noise environments. At the same time, false positive traces with relatively high confidence are more likely to be discarded.

Many vision-based algorithms ignore the height and rolling motion of helicopters. Since the camera is fixed to the drone, the rapid movement of the drone causes a dramatic change in the view of the camera, resulting in a tracking failure or a complete disappearance of objects from the field of view. Therefore, in this paper, a visual servo controller is designed to control the UAV to automatically complete the tracking task.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

The main contributions of this paper are follows:

- Combine the Kalman filter and the optical flow to predict the motion state of the object to improve the prediction accuracy.
- A low confidence tracking filtering extension was added to the Deep SORT tracking algorithm to reduce false positive tracks.
- Use the visual servo controller to assist the UAV to automatically complete the tracking task, and has no negative impact on other controllers.

The contents of this paper are arranged as follows. Section 2 describes the area and objects of the study, and introduces the overall framework of this system, including detector, tracker and servo control server. In Sections 3 and 4, experimental results are presented and discussed, respectively. Section 5 summarizes the conclusions.

2. Materials and Methods

2.1. Area and objects of study

The area selected for this study is in Maduo County, under the jurisdiction of Golog Tibetan Autonomous Prefecture, in the southern part of Qinghai Province (Figure 1a). Maduo County locates at the source of the Yellow River and belongs to a typical plateau area, with an average annual temperature of -4.0 °C. Due to its unique geographical location and ecological environment, the local flora and fauna resources are very abundant, and animal husbandry is particularly developed. It is highly reasonable to choose this area to conduct research on AI-based precise grazing technology.

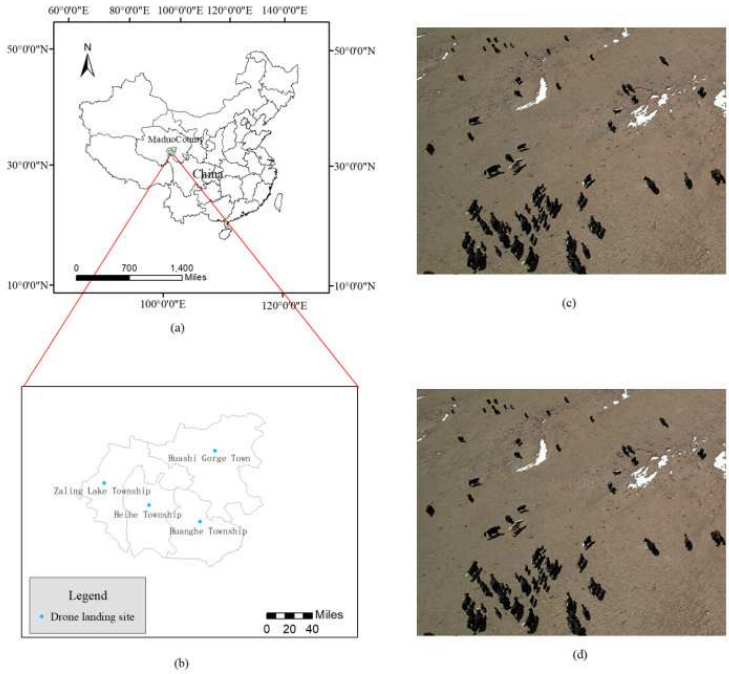

Figure 1. The area selected for the present study: (a) The location of Maduo County; (b) the distribution of UAV sampling points in Maduo County; (c, d) aerial images of the studied area.

In April 2023, the authors of this paper and research colleagues went to Maduo County for aerial photography, flying a total of 20 sorties at height of 100 m for sampling. The sampling points are shown Figure 1b. Finally, the domestic Tibetan yaks were selected as the research objects (Figure 1c and 1d), which have a color characteristic of mainly black and gray, and rarely white. The yaks move very slowly and steadily, and their stride frequency is usually between 120-140 steps per minute.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

2.2. System overview

To acquire data in the selected area, a P600 type intelligent UAV (Chengdu Bobei Technology Co., Ltd., China) was used (Figure 2). In addition, Q10F 10x single light pod equipped with a USB interface was incorporated with the P600 UAV, and a specific robot operating system (ROS2) driver was developed for P600. This equipment is able to capture real-time images through the pod within the airborne computer. It could also follow the targets and adjust the position to always keep a constant distance from moving targets. During the target tracking process, both UAV and pod can achieve fully autonomous control via ROS2.

Figure 2. Data acquisition equipment used for this study.

Based on the function, the system can be divided into three components, including the controller, the detector, and the tracker. In this study, two walking Tibetan yaks were selected as the tracking objectives. When each of the camera frames processes, several confirmed tracking paths are sent to the control system, which calculates the required speed in 4 different control variables in accordance with the embedded algorithms and the real-time location of UAV. Afterwards, the speed is sent to the autopilot to control UAV for tracking the targets. As a basic component of the control system, ROS2 plays a crucial role in information exchange between UAV and the tracking program. Moreover, the algorithms for speed calculation in 4 control variables differ from each other, so that they are described separately in Section 2.5.

2.3. Detector

Since this study aims at tracking and identifying target objects in scenarios with high dynamic density and low training data, YOLOv7 was chosen as the baseline model for balancing the limited computational power and the airborne computer speed. The YOLOv7 model was developed in 2022 by Wang and Bochkovskiy et al., integrating strategies including E-ELAN (Extended Efficient Layer Aggregation Network) [19], cascade-based model scaling [16] and model reparameterization [58] to appropriately balance the detection efficiency and accuracy. It can be seen in Figure 3 that the YOLOv 7 network comprises 4 different modules: Input module, backbone network, head network, and prediction network. Detailed description of the YOLOv7 model can be found at http://dx.doi.org/10.3390/jmse11030677.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

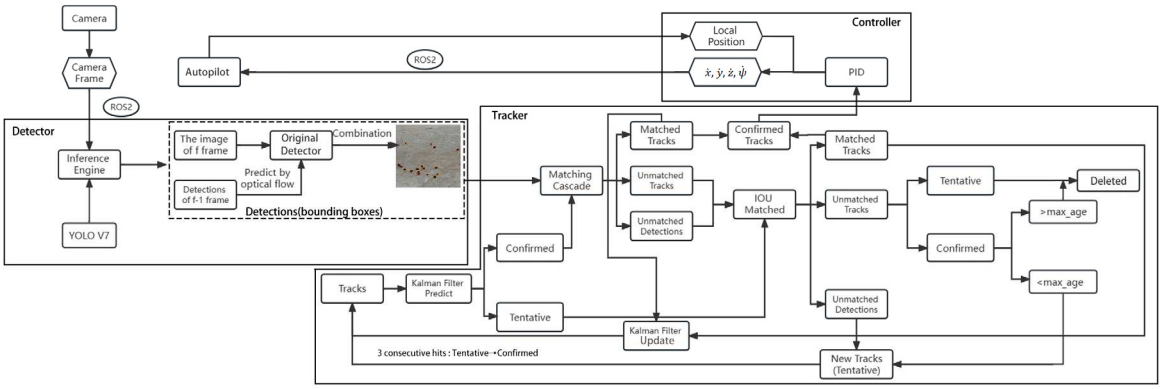

Figure 3. The overall technical framework proposed in this study.

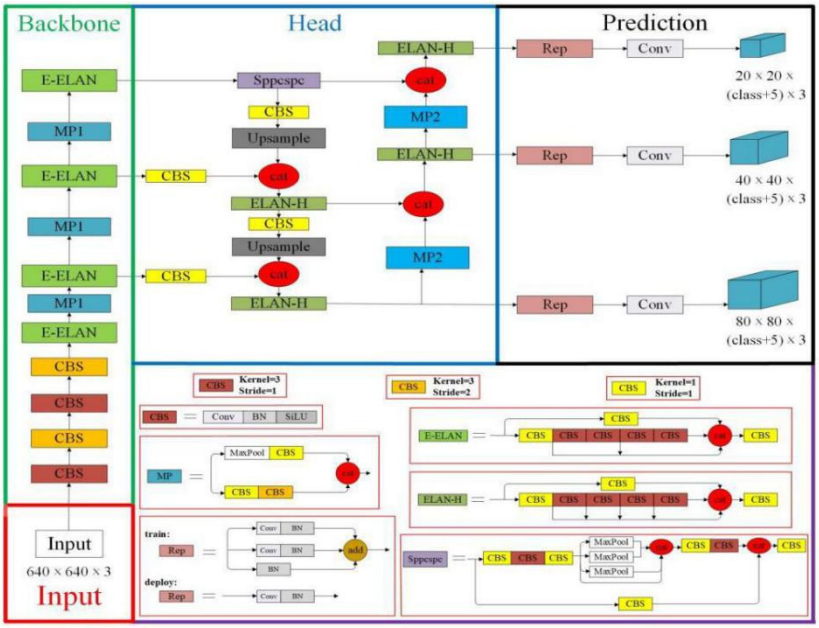

Figure 4. Network structure of the YOLOv7 model.

In this work, 20 aerial yak video data were taken from the studied area, among which 10 video sequence data were used as target detection data set and the other 10 video sequence data were applied as target tracking data set. The two sets of data were utilized as two benchmarks for yak detection and tracking, and the YOLOv7 model and Deepsort algorithm were employed to detect and track yaks.

To improve the model's ability to detect yaks, the target detection dataset was further divided into 3 sets, including training, validation and test sets with a ratio of 7:2:1. The YOLOv7 model is adopted to train the dataset by adjusting the model parameters to achieve high stability of the model. The yak hair color is generally pure black or black and white. In order to obtain more yak hair texture features, 2400 yak hair images from 10 video sequences in the target detection dataset were intercepted and the dataset was divided with the ratio of 7:2:1, which was trained again using YOLOv7.

2.4. Tracker

The Deep SORT algorithm was used as the baseline algorithm for the tracker and two improvements in the algorithm were made. Firstly, optical flow for motion estimation [43] was introduced into the scheme to improve the motion prediction accuracy of KF. Secondly, an extended

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

version of the original tracking method, named as low confidence track filtering method, was used to improve the ability of the tracker for handling unreliable detection results, which might occur in the real-world target detection due to the complex environment. By this means, the quantity of the false positive paths could be significantly reduced, avoiding the unreliable detection.

In order to apply DeepSORT to track and monitor yaks, a big amount of yak datasets are required to extract the appearance characteristics of trained yaks. Since there are only 10 video sequence data in the target tracking dataset, the quantity is insufficient, so the target tracking dataset was regenerated by setting truncation rate and occlusion rate parameters, clipping, rotating and synthesizing the video frame images. To reduce the impact of the noise, the yak data with a truncation rate above 0.5 or a blocking rate higher than 0.5 were removed. Finally, around 100 video data with the same interval were selected as a batch to intercept video frames, which were then adjusted to generate a total of 6000 yak JPEG images with the same size (500,500). Subsequently, the Labelimage software was used to annotate these images and store them in the XML format as target tracking dataset, which is considered as the benchmark for yak tracking.

2.4.1. Object tracking method

The Deep SORT algorithm adopted in this work uses KF to estimate the existing track in the current frame. The states applied in KF are defined as $(x, y, \gamma, h, \dot{x}, \dot{y}, \dot{\gamma}, \dot{h})$, in which (x, y, γ, h) represents the bounding box position, and $(\dot{x}, \dot{y}, \dot{\gamma}, \dot{h})$ represents the single coordinate velocity. KF involved in Deep SORT is the standard version using a constant velocity and a linear observation. When each new frame appears, the position of each existing track will be estimated based on the previous one, and the track estimation only needs spatial information.

In order to achieve the appearance information of the detection results and tracks, appearance descriptors were used for extracting features from the detection images and tracking the images from the previous frames. As a CNN model trained on a large-scale recognition dataset, the appearance descriptor is capable of extracting features in the feature space based on that the features from same identity are similar to each other.

By estimating the position and appearance information of existing tracks, in each future frame new detection results could be associated with the existing tracks. New detection results need to have confidence levels above the detection confidence threshold t_d to become candidates for data association. All the detections do not meet this criterion will be filtered out. A cost matrix is used in Deep SORT for representing spatial and visual similarity between the new detections and the existing tracks, which contains two distance parameters. The first one is the Mahalanobis distance represented by formula (1) for spatial information:

$$d^{(1)}(i, j) = (d_j - y_i)^T s_i^{-1} (d_j - y_i) \tag{1}$$

where y_i represents the i-th orbit, s_i^{-1} represents the covariance of d and y, (y_i, s_i) represents the projection of the i-th orbit in the space of measurement, and d_j represents the j-th new detection. It is the distance between the estimated position of the i-th orbit and the j-th new detection. The second distance represents the appearance information as shown below by formula (2):

$$d^{(2)}(i, j) = \min\left\{1 - r_j^T r_k^{(i)} \mid r_k^{(i)} \in R_i\right\} \tag{2}$$

where r represents an appearance descriptor, R_i represents the appearance of the last one hundred objects associated to the i-th track. Besides, each of the distance is accompanied by gate matrix $b_{i,j}^{(1)}$ and $b_{i,j}^{(2)}$, if the distance is less than a predefined threshold, it is equal to 1, otherwise it is equal to 0. The comprehensive cost matrix is presented in formula (3):

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j) \tag{3}$$

The gate function $b_{i,j} = \prod_{m=1}^{2} b_{i,j}^{(m)}$ is used to set the threshold, it is equal to 1 only when both the space and the appearance gate functions are 1, otherwise, it is equal to 0, indicating whether (i, j)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

effectively matches both space and appearance. The cost matrix is used for each of the new frame to associate the new detection with the tracks of the existing gate matrix.

In case of a successful association of the new detection with the existing track, the new detection is included into the track, and track shows a non-association age of zero. In case the new detection cannot be associated with the existing track in the F-frame, it is initialized as a tentative track. The original algorithm of Deep SORT verifies whether the tentative track is associated to the new detection in the frame $(f+1),(f+2),\ldots(f+t_{tentative})$. In case of a successful association, an update of the track to a confirmed one will be conducted. Otherwise, the temporary track will be immediately deleted. For existing tracks without successful association with the new detection in each frame, their non-association ages increase by 1. In case that the non-association ages exceed the threshold, the corresponding tracks will also be removed.

2.4.2. Combination of KF and optical flow

As a classic tracking algorithm, the Lucas-Kanad (LK) optical flow [40] algorithm has been widely applied due to its competitive real-time speed and strong robustness. To address the problems derived from KF, optical flow is also used to estimate objects in this study, and several assumptions are made, including constant brightness between the adjacent frames, slow movement of the targets, and similar motion pixels of the same images. There is no doubt that the loss of the object detection will challenge the updating of KF and lead to the interruption of trajectory. Therefore, the boundary frames of objects are predicted by using the light flow. In addition to the bounding frame of the F-frame generated with original detector in the data set, optical flow is also adopted to predict the position of the object based upon information in the previous frame. It could provide more historical clues to the information of the previous frame.

It can be observed that the former produces a more accurate trace input, nevertheless, the primitive detection in complex environments cannot be ignored. To compensate for the adverse effect on performance, combination of them as input for current frame tracking is required, which could provide more reliable state of motion for KF. At the same time, a constant velocity of the object in the frame is assumed, and KF is used to construct a model of linear motion defined in 8-dimensional space:

$$S=(x,y,\gamma,h,\dot{x},\dot{y},\dot{\gamma},\dot{h}) \tag{4}$$

where (x, y) represent bounding box center coordinates, γ represents the aspect ratio, h means high, and $(\dot{x},\dot{y},\dot{\gamma},\dot{h})$ represents the speed of objects in the frame.

2.4.3. Filtering of low confidence tracks

False positive tracks derived from unreliable detection results seriously affect the performance of the tracker. At present, the most advanced detection tracking technology still faces a large number of false positive tracks and other problems. To better solve this problem, a filter for low confidence tracks was included into our tracker. In this tracker, not only a confidence threshold t_d is used to filter out detections with confidence below this threshold, but also average confidence values are calculated for new detections in the frame $(f+1),(f+2),\ldots(f+t_{tentative})$ related to tentative tracks. Only when these average values are greater than the predefined threshold t_{ave_d}, update of the corresponding tentative tracks to the confirmed tracks could be performed. Otherwise, these tentative tracks will be deleted. By this means, the detection results are filtered by two threshold stages of t_d and t_{ave_d} rather than simply by t_d alone. Therefore, the threshold t_d with a preset lower value can avoid losing detection, and extraction helps for suppressing false positive tracks produced with low t_d. The algorithm used in this study to filter low confidence tracks is detailed in Appendix A.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

2.5. Visual servo control

In this study, a servo control system using helicopters and cameras [38] is applied for MOT. The system consists of 4 control variables, including lateral control, longitudinal control, vertical control, as well as yaw rate control.

The lateral control aims at keeping the camera frame center aligning with the horizontal middle of tracked objects by using a PID controller that takes the sum of the horizontal distances of each object as the proportional input, the sum of the differences between the current and previous centers as the derivative input, and the cumulative error as the integral input.

The longitudinal control adjusts the forward and backward speed of the helicopter based on the heights of bounding boxes of the objects, which indicate the distance of objects to the camera. This control unit uses a PID controller that takes the sum of differences between current and minimum heights and between current and maximum heights as the proportional input for calculation of forward and backward speeds, respectively. Besides, it takes the sum of height change rates of each object as the derivative input.

The vertical control loosely regulates the height of the drone based on a predefined range. In comparison with the response to the lateral speeds, the response of the autopilot to low vertical speeds to achieve accurate height adjustment is relatively slower. Therefore, it is often that after the autopilot receives such a vertical speed command, the height of drone does not change.

The yaw rate control rotates the helicopter around its vertical axis to keep it perpendicular to the line connecting the two objects outermost of the camera frame, which estimate the yaw angle by using a ratio between horizontal distance and image width, and a ratio between height difference and standard height for each class of objects. Afterwards, this angle is divided by the processing time and multiplied by a coefficient to achieve the yaw rate.

3. Results

The intelligent unmanned field platform embedded with Jetson AGX Xavier launched by NVIDIA [44] was used for onboard image processing in the experiments. This modular supercomputer has a 512 CUDA-core NVIDIA Volta GPU with a 8-core ARMv8.2 CPU and strong power of AI computation. It shows a 10 times higher power consumption ratio and a 20 times higher performance compared to the previous Jetson TX2 platform (256 CUDA-core NVIDIA Pascal GUP with a CPU of quad-core ARM).

3.1. Metrics for tracking

To objectively compare the performance of different trackers, the experimental results from this study were evaluated based on the metrics defined in the CLEAR MOT metrics and :

PR-MOTA: Under different confidence thresholds, the values of precision and recall are obtained separately, and then the corresponding PR-MOTA can be obtained based on the different precision and recall. MOTA is the multi-target tracking accuracy , a key score for evaluating the tracking performance. It is composed of 3 calculation errors, including false positive (FP), lost target (FN), and identity switch (IDs). It measures the performance of the tracker in detecting targets and maintaining trajectories, independent of the accuracy of the target location estimation.

- PR-MOTP: It is derived from the values of precision and recall under different confidence thresholds. MOTP is the multi-target tracking accuracy , which is a measure of the tracker's ability to estimate the target position.
- PR-MT: It is originated from the values of precision and recall for different confidence thresholds. MT is the number of primary tracking traces that are successfully tracked during at least 80% of the target’s lifetime.
- PR-ML: It is derived from the values of precision and recall under different confidence thresholds. ML is the quantity of the mostly lost tracks that are not successfully tracked during minimum 20% of the target's lifetime.
- PR-FP: It is the total quantity of FPs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

- PR-FN: It means total quantity of FNs (target not met).
- PR-FM: With different confidence thresholds, the PR-FM is derived from the values of precision and recall. FM is the times of interruption for a track due to missing detection.
- PR-IDSw: It is found under different confidence thresholds based on the values of precision and recall. IDSw, also known as IDs, is the times of the IDs switch for the same target due to misjudgment of the tracking algorithm. The ideal IDs in the tracking algorithm should be 0. It is the total number of identity switches.

3.2. Evaluation of benchmarks

When the motion of the object is complex and the detection of the object in the current frame is lost, the bounding box predicted by KF cannot match the input. Therefore, the optical flow was introduced into the scheme to improve the motion prediction accuracy of KF. As shown in Figure 5, the yellow-colored bounding boxes are the original detection results and the red ones are the results of the optical flow.

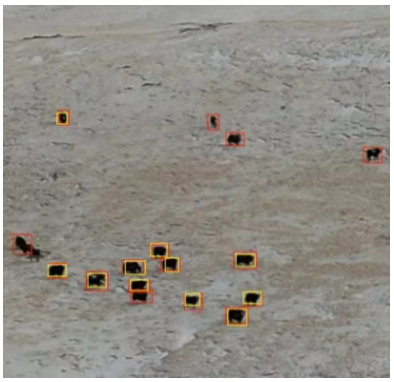

Figure 5. Comparison of the detection results (yellow bounding boxes: Original detection results; red bounding boxes: Detection results from optical flow).

The resulting mean detection confidence threshold was chosen experimentally. 10 sequences were selected from the target-tracking dataset that were filmed in a relatively more complex environment. It was found that t_{ave_d} = 0.0~1.0 for these 10 sequences. The tracker in this study used the YOLOV7 detection method as detection input. The 10 sequences were tested, and the final tracking results were evaluated. Figure 6 shows a comparison of these results.

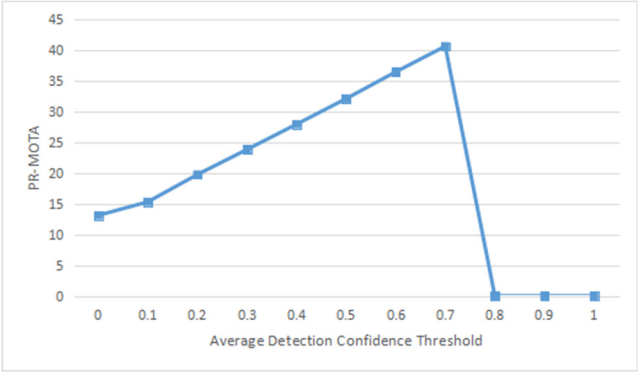

Figure 6. Comparison of the MOTA values of tracking results for 10 training sequences using YOLOV7 detection method under different average detection confidence thresholds.

The tracking results obtained with the proposed tracker in the train sequence "Tibetanyak2023042607" of the target tracking dataset are shown in Figure 7. This is the result of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

tracking using the YOLOv7 detection, where the t_d and t_{ave_d} thresholds were set to 0.0 and 0.7, respectively.

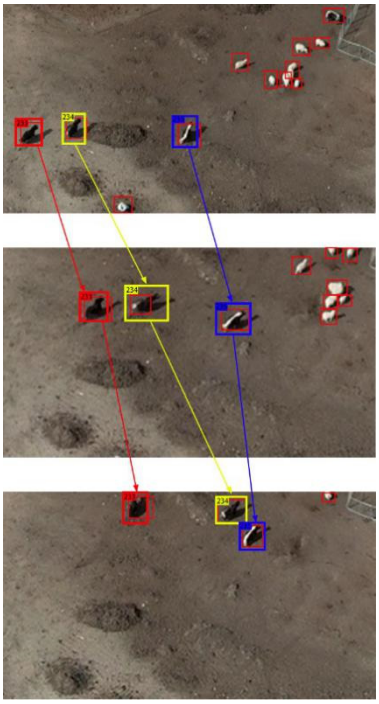

Figure 7. The tracking results on the self-made training sequence "Tibetanyak2023042607" using YOLOv7 detections (td = 0.0 and taved = 0.7).

The proposed method was tested on an overall test dataset of the target tracks containing ten sequences. As described above, the YOLOv7 tracking results were used as the test input. This detector was chosen for this study because it is the baseline detection method used by most of other trackers, and it exhibits a general good performance. Table 1 shows a comparison of the proposed tracker and state-of-the-art ones with respect to the tracking performance. All these trackers were divided into batch class and online class. In the batch trackers, both the previous and the future information are used for generating tracks in the current frame, while in the online trackers, only the previous information is applied for generating tracks

Table 1. Comparison of the results from the proposed method and other methods (tests conducted on the self-made training date set).

Tracker	Detector	Method	PR-MOTA	PR-MOTP	PR-MT	PR-ML	PR-FM	PR-FP	PR-FN	PR-IDs
IOU[17]	R-CNN[53]	Batch	18.3%	41.9%	14.3%	20.6%	523	2313.5	19845.1	513
IOU[17]	Comp ACT[68]	Batch	18.4%	41.3%	14.7%	20.1%	379	2459.2	17125.6	245
IOU[17]	EB[41]	Batch	23.5%	33.2%	17.5%	16..7	248	1456.6	17054.4	233
IOU[17]	YOLOv7	Batch	33.8%	40.2%	34.6%	19.4%	88	1731.5	17945.5	70
Deep SORT	EB[41]	Online	20.6%	45.3%	18.1%	17.2%	201	3501.9	16874.5	180
Ours	EB[41]	Online	22.9%	45.3%	17.8%	17.3%	205	2009.7	17012.4	166

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

Deep SORT	YOLOv7	Online	30.4%	39.1%	34.3%	18.5%	159	6456.6	16456.7	245
Ours	YOLOv7	Online	33.6%	39.2%	32.9%	19.7%	126	2013.1	17913.2	198

3.3. Validation in actual scenarios

In this study, visual servo controller was used to control parameters from four aspects, i. e., lateral, longitudinal, vertical, and yaw rate controls, to assist P600 intelligent UAV flight, and to track and identify multiple yaks. To simultaneously test the comprehensive performances of the visual servo controller in all directions, an experiment with relatively complex object trajectories was designed. In this experiment, a pure black yak and a yak with black and white color were chosen as target objects to verify the tracking ability of the UAV with the visual servo controller. The two yaks walked along concentric arcs of different radii. No overlapped trajectories of yaks and drones were observed. Figure 8 shows the relevant trajectories in real scenes.

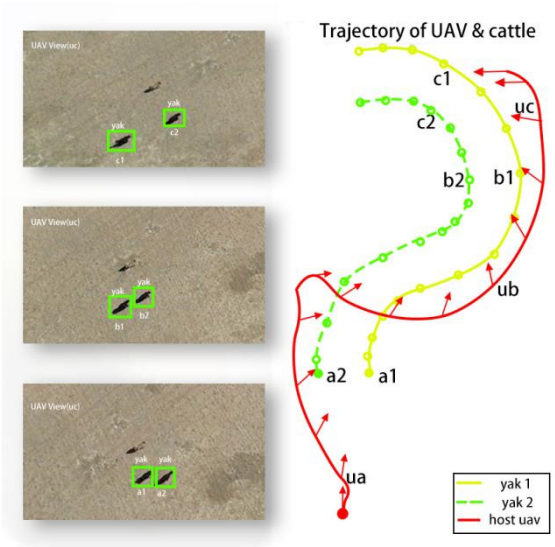

Figure 8. The trajectories of the UAV (red line) and the yaks (yellow and green lines). The left 3 images show the drone views at the corresponding positions of ua, ub and uc.

Aiming at further verifying the yaw rate performance of the visual servo controller, the starting points of the trajectories were marked with solid points of the corresponding colors, just similar to the trajectories in Figure 8. Fifteen arrows were presented on the UAV P600 trajectory to point out the current YUAV axis directions. In addition, on each locus of the two objects, fifteen hollow circles were presented, which correspond to the arrows in the experiment. The fifteen time points were also marked and exhibited in Figure 9, which shows the angle between the two object connection lines plus Π/2 and Xworld, and the angle curve between YUAV and Xworld. YUAV is the trajectory of the drone and Xworld is the trajectory of the two objects.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 June 2023 doi:10.20944/preprints202306.1669.v1

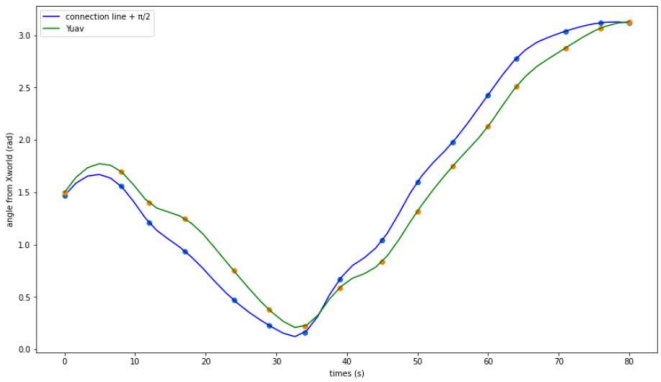

Figure 9. Angle between two object connectors plus Π/2 and Xworld and the angle curve between YUAV and Xworld.

4. Discussion

As shown in Figure 6, MOTA is better in the presence of t_{ave_d} (not equal to 0) than in the absence of t_{ave_d} (equal to 0). The tracking accuracy of the YOLOV7 method keeps improving until t_{ave_d} reaches 0.7. Therefore, t_{ave_d} = 0.7 was selected for the experiment of the tracker on the YOLOV7 detection results. It is worth noting that for this detection method, the tracker does not work when t_{ave_d} is greater than 0.7, because all trajectories are filtered out under that condition.

It can be seen from Figure 7 that there are many red boxes without the identification label. They are the false positive detection filtered out by the filtering algorithm for low confidence tracking. When t_d = 0.0, many red bounding boxes appear, however, they do not affect the final result of tracking.

Table 1 shows that 33.6% of PR-MOTA on YOLOV7 detection was achieved with the proposed tracker. It's the highest value for all the online trackers and comparable to the highest PR-MOTA for the batch IOU tracker. Note that the original algorithm of Deep SORT trained on appearance descriptors of our dataset was already able to achieve high PR-MOTA of 30.4% on YOLOV7 detection. Based on the improved algorithm, the proposed tracker could further enhance PR-MOTA of the original tracker by about 3.2%. Furthermore, a 4443.5 PR-FP decrease on the improved algorithm compared to Deep SORT was observed, revealing that PR-FP could be significantly reduced using the algorithm optimized in this paper. Meanwhile, the ID of our tracker significantly decreased compared to the ID of Deep SORT on the YOLOV7 detection.Results from experiments indicate that the improved algorithm could reduce false positive targets, and, it shows better accuracy performance than the other relevant algorithms.

As shown in Figure 9, the maximum distance between the UAV trajectory line and the trajectory lines of the two yaks occurred at 21.086 s was -0.342 rad. This was because that the positions of the two target objects (Xworld,Yworld) changed rapidly, and the yaw rate visual servo controller was unable to respond quickly to the abrupt change. During the later half period of the experiment, the controller always kept a distance around 0.26 rad, and at the final stage, it adjusted YUAV nearly perpendicular to the connection line. These results can demonstrate the performance of the visual servo controller to some extent.

5. Conclusion

A real-time target tracking system that can always keep the tracking targets within the view of a UAV camera is presented in this paper. The system uses the YOLOv7 algorithm for target detection and the DeepSORT algorithm optimized in this work for target tracking. In addition, a visual servo controller for the UAV is designed to complete the automated tracking task. Based on the experimental results, the following conclusions are summarized:

Using the optical flow compensation Kalman filter for motion prediction can solve the problem that the bounding box of the target predicted by the Kalman filter cannot match the input when the detection of the target in the current frame is complex, and thereby achieve better accuracy results.

Low confidence trajectory filtering methods can significantly reduce false positive trajectories generated by DeepSORT to avoid unreliable detection.

The DeepSORT algorithm modified in this work combines optical flow with a Kalman filter for motion prediction and adds a low-confidence trajectory filtering method to achieve a 3.2% of improvement in PR-MOTA, 4443.5 of reduction in PR-FP and the ID significantly decreased compared to original Deep SORT.

Using the visual servo controller can assist the UAV in multi-target tracking and identification of the yak group. In addition, it has no negative effect on the other controllers.

CRediT authorship contribution statement: Wei Luo: Conceptualization, Zhiguo Wang: Methodology, Xiaoliang Li: Supervision Yongxiang Zhao: Writing- Reviewing and Editing, Zihui Zhao: Software, Xia Zhu: Datacuration, Quanqin Shao: finances, Dongliang Wang: Software, Guoging Zhang: WritingOriginal draft preparation, Longfang Duan: Visualization, Ke Liu: Investigation, XiongyiZhang: Validation, Jiandong Liu: field investigation , Zhongde Yu: format editing..

Data availability: Data will be made available on request.

Acknowledgements: This research was funded by the National Natural Science Foundation of China (No.: 42071289).

Declaration of competing interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Low Confidence Track Filtering Algorithm

Algorithm Low Confidence Track Filtering

Input：Tentative tracks T_t; Tentative threshold $t_{tentative}$;

Average detection confidence threshold t_{ave_d}; Associated detection confidence P_t.

Output：Confirmed tracks T_c; Deleted tracks T_d.

1：for sequential frames do

2： for $t \in T_t$ do

3： if t is new in T_t then

4： hits = 0

5： total_prob = 0

6： hits = hits + 1

7： total_prob = total_prob + P_t

8： if $hits \geq t_{tentative}$ then

9： if $\frac{total\_prob}{hits} < t_{ave_d}$ then

10： $T_d = T_d \bigcup t$ and $T_t = T_t \setminus t$

11： else

12： $T_c = T_c \bigcup t$ and $T_t = T_t \setminus t$

References

1. Andrew William,Gao Jing,Mullan Siobhan,Campbell Neill,Dowsey Andrew W.,Burghardt Tilo. Visual identification of individual Holstein-Friesian cattle via deep metric learning[J]. Computers and Electronics in Agriculture,2021,185(1)

2. Altuˇg, E., Ostrowski, J.P., Mahony, R.: Control of a quadrotor helicopter using visual feedback. In: Proceedings of the 2002 IEEE International Conference on Robotics & Automation. Washington, DC (2002)
3. Altuˇg, E., Ostrowski, J.P., Taylor, J.: Control of a quadrotor helicopter using dual camera visual feedback. Int. J. Robot. Res. 24(5), 329–341 (2005)
4. Ardö, H.; Guzhva, O.; Nilsson, M. A CNN-based cow interaction watchdog. In Proceedings of the 23rd International Conference Pattern Recognition, Cancun, Mexico, 4–8 December 2016; pp. 1–4.
5. Azinheira, J.R., Rives, P., Carvalho, J.R.H., Silveira, G.F., de Paiva, E.D., Bueno, S.S.: Visual servo control for the hovering of an outdoor robotic airship. ICRA 3, 2787–2792 (2002).
6. Baraniuk, R., Donoho, D., & Gavish, M. (2020). The science of deep learning. Proceedings of the National Academy of Sciences, 117(48), 30029–30032. https://doi.org/10.1073/pnas.2020596117
7. Barrett H., & Rose D.C. (2020). Perceptions of the fourth agricultural revolution: What's in, what's out, and what consequences are anticipated? Sociologia Ruralis, 60(3), 631–652. https://onlinelibrary.wiley.com/doi/full/10.1111/soru.12324
8. Behrendt, K.; Novak, L.; Botros, R. A Deep Learning Approach to Traffic Lights: Detection, Tracking, and Classification. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017.
9. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple online and realtime tracking. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016.
10. Bourquardez, O., Chaumette, F.: Visual Servoing of an Airplane for Auto-landing. IROS, San Diego (2007).
11. Chabot, D., & Bird, D. M. (2015). Wildlife research and management methods in the 21st century: Where do unmanned aircraft fit in? Journal of Unmanned Vehicle Systems, 3(4), 137–155. https://doi.org/10.1139/juvs-2015-0021
12. Chen, L.; Ai, H.; Zhuang, Z.; Shang, C. Real-Time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), San Diego, CA, USA, 23–27 July 2018.
13. Ciaparrone, G.; Sánchez, F.L.; Tabik, S.; Troiano, L.; Tagliaferri, R.; Herrera, F. Deep learning in video multi-object tracking: A survey. Neurocomputing 2020, 381, 61–88. [CrossRef]
14. Corcoran, E., Winsen, M., Sudholz, A., & Hamilton, G. (2021). Automated detection of wildlife using drones: Synthesis, opportunities and constraints. Methods in Ecology and Evolution, 12(4), 674–687. https://doi.org/10.1111/2041-210X.13581
15. David Christian Rose, Rebecca Wheeler, Michael Winter, Matt Lobley, Charlotte-Anne Chivers, Agriculture 4.0: Making it work for people, production, and the planet, Land Use Policy, Volume 100, 2021, 104933, https://doi.org/10.1016/j.landusepol.2020.104933.
16. Dollár, Piotr.; Mannat, Singh.; Ross, Girshick. Fast and accurate model scaling. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. pp. 924-932.
17. E. Bochinski, V. Eiselein, and T. Sikora. High-speed tracking-by-detection without using image information. In International Workshop on Traffic and Street Surveillance for Safety and Security at IEEE AVSS 2017, Lecce, Italy, Aug. 2017. 1, 5
18. Ess, A.; Schindler, K.; Leibe, B.; Gool, L.V. Object Detection and Tracking for Autonomous Navigation in Dynamic Environments. Int. J. Robot. Res. 2010, 29, 1707–1725. [CrossRef]
19. Gao, J., Burghardt, T., Andrew, W., Dowsey, A. W., & Campbell, N. W. (2021). Towards Self-Supervision for Video Identification of Individual Holstein-Friesian Cattle: The Cows2021 Dataset. Paper presented at Conference on Computer Vision and Pattern Recognition Workshop on Computer Vision for Animal Behavior Tracking and Modeling (CV4Animals). https://arxiv.org/abs/2105.01938
20. Gao, P.; Lu, J.; Li, H.; Mottaghi, R.; Kembhavi, A. Container: Context aggregation network. arXiv preprint arXiv:2106.01401, 2021.
21. Guo, Z., Pan, Y., Sun, T., Zhang, Y., and Xiao, X., "Adaptive neural network control of serial variable stiffness actuators," Complexity, Vol. 2017, 2017.
22. Hamel, T., Mahony, R.: Visual servoing of an under actuated dynamic rigid-body system: an image-based approach. IEEE Trans. Robot. Autom. 18(2), 187–198 (2002).
23. Han, L.; Tao, P.; Martin, R.R. Livestock detection in aerial images using a fully convolutional network. Comput. Vis. Media 2019, 5, 221–228.

24. He, J.; Huang, Z.; Wang, N.; Zhang, Z. Learnable Graph Matching: Incorporating Graph Partitioning with Deep Feature Learning for Multiple Object Tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021.
25. H. Pirsiavash, D. Ramanan, and C. C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In CVPR 2011, pages 1201–1208, June 2011. 5
26. Islam, M.; Hong, J.; Sattar, J. Person-following by autonomous robots: A categorical overview. Int. J. Robot. Res. 2019, 38, 1581–1618. [CrossRef]
27. Jellason, N.P.; Robinson, E.J.Z.; Ogbaga, C.C. Agriculture 4.0: Is Sub-Saharan Africa Ready? Appl. Sci. 2021, 11, 5750. https://doi.org/10.3390/app11125750
28. Jiménez López, J., & Mulero-Pázmány, M. (2019). Drones for conservation in protected areas: Present and future. Drones, 3(1), 10. https://doi.org/10.3390/drones3010010
29. Jin, J.; Li, X.; Li, X.; Guan, S. Online Multi-object Tracking with Siamese Network and Optical Flow. In Proceedings of the IEEE 5th International Conference on Image, Vision and Computing (ICIVC), Beijing, China, 10–12 July 2020.
30. Kamal, R.; Chemmanam, A.J.; Jose, B.; Mathews, S.; Varghese, E. Construction Safety Surveillance Using Machine Learning. In Proceedings of the International Symposium on Networks, Computers and Communications (ISNCC), Montreal, QC, Canada, 20–22 October 2020.
31. Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay Ç., Backus G., Bee MA., Bohn K., Cao Y., Carter G., Cäsar C., Coen M., DeRuiter SL., Doyle L., Edelman S., Ferrer-i-Cancho R., Freeberg TM., Garland EC., Gustison ML., Harley HE., Huetz C., Hughes M., Bruno JHJr.. Ilany A., Jin DZ., Johnson M., Ju C., Karnowski J., Lohr B., Manser MB., McCowan B., Mercado E III.. Narins PM.. Piel A.. Rice M.. Salmi R.. Sasahara K.. Sayigh L.. Shiu Y.. Taylor C.. Vallejo EE.. Waller S.& Zamora-Gutierrez V.(2020). Acoustic sequences in non-human animals: a tutorial review and prospectus.Biological Reviews95(1)13–52. https://doi.org/10.1111/brv.12556
32. K. He, G. Gkioxari, P. Doll´ar, and R. B. Girshick.Mask R-CNN. CoRR, abs/1703.06870, 2017. 4, 5
33. Krul, S.; Pantos, C.; Frangulea, M.; Valente, J. Visual SLAM for Indoor Livestock and Farming Using a Small Drone with a Monocular Camera: A Feasibility Study. Drones 2021, 5, 41. https://doi.org/10.3390/drones5020041
34. Kuhn, H.W. The Hungarian method for the assignment problem. Nav. Res. Logist. Q. 1955, 2, 83–97. [CrossRef]
35. Lee, S.; Kim, E. Multiple Object Tracking via Feature Pyramid Siamese Networks. IEEE Access 2019, 7, 8181–8194. [CrossRef]
36. Li, G.; Huang, Y.; Chen, Z.; Chesser, G.D., Jr.; Purswell, J.L.; Linhoss, J.; Zhao, Y. Practices and Applications of Convolutional Neural Network-Based Computer Vision Systems in Animal Farming: A Review. Sensors 2021, 21, 1492. https://doi.org/10.3390/s21041492
37. Lin, T.; Dollár, P.; Girshick, R.B.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
38. Liu, J. and Yuqiang Yao. “Real-time Multiple Objects Following Using a UAV.” AIAA SCITECH 2023 Forum (2023): n. pag.
39. Lo, S.; Yamane, K.; Sugiyama, K. Perception of Pedestrian Avoidance Strategies of a Self-Balancing Mobile Robot. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 4–8 November 2019.
40. Lucas, B.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI), Vancouver, BC, Canada, 24–28 August 1981.
41. L. Wang, Y. Lu, H. Wang, Y. Zheng, H. Ye, and X. Xue. Evolving boxes for fast vehicle detection. In IEEE International Conference on Multimedia and Expo (ICME), pages 1135–1140, 2017. 4, 5
42. Luo, W.; Li, X.; Zhang, G.; Shao, Q.; Zhao, Y.; Li, D.; Zhao, Y.; Li, X.; Zhao, Z.; Liu, Y.; et al. High-Accuracy and Low-Latency Tracker for UAVs Monitoring Tibetan Antelopes. Remote Sens. 2023, 15, 417. https://doi.org/10.3390/rs15020417
43. Luo, W.; Zhao, Y.; Shao, Q.; Li, X.; Wang, D.; Zhang, T.; Liu, F.; Duan, L.; He, Y.; Wang, Y.; Zhang, G.; Wang, X.; Yu, Z. Procapra Przewalskii Tracking Autonomous Unmanned Aerial Vehicle Based on Improved Long and Short-Term Memory Kalman Filters. Sensors 2023, 23, 3948. https://doi.org/10.3390/s23083948

44. Luo, W.; Zhang, Z.; Fu, P.; Wei, G.; Wang, D.; Li, X.; Shao, Q.; He, Y.; Wang, H.; Zhao, Z.; et al. Intelligent Grazing UAV Based on Airborne Depth Reasoning. Remote Sens. 2022, 14, 4188. https://doi.org/10.3390/rs14174188
45. Mackenzie Weygandt Mathis, Alexander Mathis, Deep learning tools for the measurement of animal behavior in neuroscience, Current Opinion in Neurobiology, Volume 60, 2020, Pages 1-11. https://doi.org/10.1016/j.conb.2019.10.008.
46. Ma, Z., and Sun, G., "Dual terminal sliding mode control design for rigid robotic manipulator," Journal ofthe Franklin Institute, Vol. 355, No. 18, 2018, pp. 9127–9149.
47. Milan, A.; Leal-Taixé, L.; Reid, I.D.; Roth, S.; Schindler, K. MOT16: A benchmark for multi-object tracking. arXiv 2016, arXiv:1603.00831.
48. Mohd Javaid, Abid Haleem, Ravi Pratap Singh, Rajiv Suman, Enhancing smart farming through the applications of Agriculture 4.0 technologies, International Journal of Intelligent Networks, Volume 3, 2022, 150-164. https://doi.org/10.1016/j.ijin.2022.09.004.
49. Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences, 115(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115
50. NVIDIA, "NVIDIA TensorRT," , 2021. URL https://developer.nvidia.com/tensorrt.
51. Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., & Shaevitz, J. W. (2020). Fast animal pose estimation using deep neural networks. Nature Methods, 17(1), 59–62. https://doi.org/10.1038/s41592-019-0667-1
52. Proctor, A.A., Johnson, E.N., Apker, T.B.: Visiononly control and guidance for aircraft. J. Field Robot. 23(10), 863–890 (2006).
53. R. B. Girshick, J. Donahue, T. Darrell, and J. Malik.Rich feature hierarchies for accurate object detection and semantic segmentation. CoRR, abs/1311.2524,2013. 5
54. Romero, H., Benosman, R., Lozano, R.: Stabilizational Location of a Four Rotor Helicopter Applying Vision, pp. 3930–3936. ACC, Minneapolis (2006).
55. Schad, L., & Fischer, J. (2022). Opportunities and risks in the use of drones for studying animal behaviour. Methods in Ecology and Evolution, 13(1), 3–16. https://doi.org/10.1111/2041-210X.13922
56. Subramanian, R. G., Elumalai, V. K., Karuppusamy, S., and Canchi, V. K., "Uniform ultimate bounded robust model reference adaptive PID control scheme for visual servoing," Journal ofthe Franklin Institute, Vol. 354, No. 4, 2017, pp. 1741–1758.
57. Sukhatme, G.S., Mejias, L., Saripalli, S., Campoy, P.: Visual servoing of an autonomous helicopter in urban areas using feature tracking. J. Field Robot. 23(3/4), 185–199 (2006).
58. Vasu, P. K. A.; Gabriel, J.; Zhu, J.; Tuzel, O.; Ranjan, A. An improved one millisecond mobile backbone. arXiv preprint arXiv:2206.04040, 2022.
59. Wang, X., "TensorRTx," , 2021. URL https://github.com/wang-xinyu/tensorrtx.
60. 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 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2019, pp. 237-243, doi: 10.1109/IROS40897.2019.8968555.
61. W. Andrew, C. Greatwood and T. Burghardt, "Deep Learning for Exploration and Recovery of Uncharted and Dynamic Targets from UAV-like Vision," 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 2018, pp. 1124-1131, doi: 10.1109/IROS.2018.8593751.
62. W. Andrew, C. Greatwood and T. Burghardt, "Visual Localisation and Individual Identification of Holstein Friesian Cattle via Deep Learning," 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 2017, pp. 2850-2859, doi: 10.1109/ICCVW.2017.336.
63. Wang, Q.; Zhang, L.; Bertinetto, L.; Hu, W.; Torr, P.H. Fast Online Object Tracking and Segmentation: A Unifying Approach. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
64. Wojke, N.; Bewley, A.; Paulus, D. Simple Online and Realtime Tracking with a Deep Association Metric. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017.

65. Wu, A.D., Johnson, E.N., Proctor, A.A.: Vision-aided inertial navigation for flight control. In: AIAA Guidance, Navigation and Control Conf. and Exhibit, San Francisco, USA (2005).
66. Xiaohui Li, Li Xing, Use of Unmanned Aerial Vehicles for Livestock Monitoring based on Streaming K-Means Clustering. 6th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture AGRICONTROL 2019, Volume 52, Issue 30, 2019, 324-329. https://doi.org/10.1016/j.ifacol.2019.12.560.
67. Xu, Y.; Osep, A.; Ban, Y.; Horaud, R.; Leal-Taixe, L.; Alameda-Pineda, X. How To Train Your Deep Multi-Object Tracker. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020.
68. Z. Cai, M. J. Saberian, and N. Vasconcelos. Learning complexity-aware cascades for deep pedestrian detection. CoRR, abs/1507.05348, 2015. 5
69. Zhang, D., and Wei, B., "A review on model reference adaptive control of robotic manipulators," Annual Reviews in Control, Vol. 43, 2017, pp. 188–198.
70. Zhang, Y.; Cai, J.; Xiao, D.; Li, Z.; Xiong, B. Real-time sow behavior detection based on deep learning. Comput. Electron. Agric. 2019, 163, 104884.
71. Zhou, M.; Elmore, J.A.; Samiappan, S.; Evans, K.O.; Pfeiffer, M.B.; Blackwell, B.F.; Iglay, R.B. Improving Animal Monitoring Using Small Unmanned Aircraft Systems (sUAS) and Deep Learning Networks. Sensors 2021, 21, 5697. https://doi.org/10.3390/s21175697
72. Zhu, X.; Chen, C.; Zheng, B.; Yang, X.; Gan, H.; Zheng, C.; Yang, A.; Mao, L.; Xue, Y. Automatic recognition of lactating sow postures by refifined two-stream RGB-D faster R-CNN. Biosyst. Eng. 2020, 189, 116–132.