Title

VTag: a semi-supervised pipeline for tracking pig activity with a single top-view camera

Permalink

https://escholarship.org/uc/item/8mh5m5np

Journal

Journal of Animal Science, 100(6)

ISSN

0021-8812

Authors

Chen, Chun-Peng J
Morota, Gota
Lee, Kiho
et al.

Publication Date

2022-06-01

DOI

10.1093/jas/skac147

Peer reviewed

OXFORD

AMERICAN SOCIETY of **ANIMAL SCIENCE**

VTag: a semi-supervised pipeline for tracking pig activity with a single top-view camera

Chun-Peng J. Chen,[†,] Ⓘ **Gota Morota,**[‡,||,] Ⓘ **Kiho Lee,**[$,] Ⓘ **Zhiwu Zhang,**[¶,] Ⓘ **and Hao Cheng**[†,1,] Ⓘ

[†]Department of Animal Science, University of California, Davis, CA 95616, USA
[‡]Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[||]Center for Advanced Innovation in Agriculture, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[$]Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA
[¶]Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA

[1]Corresponding author: qtlcheng@ucdavis.edu

Abstract

Precision livestock farming has become an important research focus with the rising demand of meat production in the swine industry. Currently, the farming practice is widely conducted by the technology of computer vision (CV), which automates monitoring pig activity solely based on video recordings. Automation is fulfilled by deriving imagery features that can guide CV systems to recognize animals' body contours, positions, and behavioral categories. Nevertheless, the performance of the CV systems is sensitive to the quality of imagery features. When the CV system is deployed in a variable environment, its performance may decrease as the features are not generalized enough under different illumination conditions. Moreover, most CV systems are established by supervised learning, in which intensive effort in labeling ground truths for the training process is required. Hence, a semi-supervised pipeline, VTag, is developed in this study. The pipeline focuses on long-term tracking of pig activity without requesting any pre-labeled video but a few human supervisions to build a CV system. The pipeline can be rapidly deployed as only one top-view RGB camera is needed for the tracking task. Additionally, the pipeline was released as a software tool with a friendly graphical interface available to general users. Among the presented datasets, the average tracking error was 17.99 cm. Besides, with the prediction results, the pig moving distance per unit time can be estimated for activity studies. Finally, as the motion is monitored, a heat map showing spatial hot spots visited by the pigs can be useful guidance for farming management. The presented pipeline saves massive laborious work in preparing training dataset. The rapid deployment of the tracking system paves the way for pig behavior monitoring.

Lay Summary

Collecting detailed measurements of animals through cameras has become an important focus with the rising demand for meat production in the swine industry. Currently, researchers use computational approaches to train models to recognize pig morphological features and monitor pig behaviors automatically. Though little human effort is needed after model training, current solutions require a large amount of pre-selected images for the training process, and the expensive preparation work is difficult for many farms to implement such practice. Hence, a pipeline, VTag, is presented to address these challenges in our study. With few supervisions, VTag can automatically track positions of multiple pigs from one single top-view RGB camera. No pre-labeled images are required to establish a robust pig tracking system. Additionally, the pipeline was released as a software tool with a friendly graphical user interface, that is easy to learn for general users. Among the presented datasets, the average tracking error is 17.99 cm, which is shorter than one-third of the pig body length in the study. The estimated pig activity from VTag can serve as useful farming guidance. The presented strategy saves massive laborious work in preparing training datasets and setting up monitoring environments. The rapid deployment of the tracking system paves the way for pig behavior monitoring.

Key words: computer vision; pig activity; object tracking; RGB camera
Abbreviations: CV, computer vision; GPU, graphics processing unit; POI, pixels of interest; RGB, red, green, and blue; VTag, virtual tag

Introduction

Precision livestock farming, which collects detailed measurements of animals through sensors or cameras, has become an important research focus with the rising demand of animal production (Morota et al., 2018; Clark et al., 2020). Monitoring animal activity can facilitate the management of animal production, and it is conventionally conducted by frequently visiting the farms or manually reviewing recorded videos (Larsen et al., 2019). However, these approaches can be subjective and laborious. Alternatively, the technology of computer vision (CV), which is inspired by human vision that can intuitively focus on objects of interest and exclude noisy signals, can automate monitoring animal activity solely based on information obtained from video recordings. Automation is fulfilled by deriving imagery features from a series of computational tasks, such as video segmentation and edge detection, guiding the CV models to recognize animals' body contours, positions, and behavioral categories.

With the CV technology, many studies have shown promising results in practicing precision livestock farming. For example, it used to be costly to manage cattle in large-scale pasture lands. Coupling with unmanned aerial vehicles, CV is

possible to automate cattle counting in real-time with labor costs substantially reduced (Xu et al., 2020). In smaller-scale indoor farms, CV systems were also used to detect body cleanness (Li et al., 2019a), entirety (Fang et al., 2020), structure (Liu et al., 2020), and behaviors (Okura et al., 2019; Fuentes et al., 2020; Ren et al., 2021) for animal production. This technology is particularly beneficial to the swine industry, as pigs are usually group-housed in indoors settings. By deploying one top-view RGB colors camera, producers can track pig activity by capturing their positions and identities in a high-throughput manner (Yang et al., 2018; Li et al., 2019b; Zhang et al., 2019; Huang et al., 2020). To assess complicated traits that are labor intensive to be measured, the deployment of multiple cameras or RGB-D depth sensing cameras can provide extra dimensions of information to enhance the CV system. Many successful applications have also demonstrated automation in assessing body weight (Yu et al., 2021), feeding behaviors (Leonard et al., 2019), and more precise measurement of real-time pig positions (Tu et al., 2020) in recent years.

Nevertheless, challenges still exist in the current CV systems and make CV difficult to be widely implemented in most farming environments. First, the performance of CV systems is sensitive to the imagery features, which are derived based on the observed imagery patterns under certain environmental conditions (e.g., lighting). When the CV system is deployed in a new environment, its performance may decrease as the features are not generalized enough to be associated with pig morphological patterns under different illumination conditions (Chen et al., 2021). Second, most CV systems are established by supervised learning, in which intensive effort in labeling ground truths for the training process is required. The insufficient size or quality of training datasets may be a harmful factor for the model robustness. Lastly, many successful CV systems are built by deep learning models, in which tens of millions of unknown parameters are estimated (He et al., 2015; Simonyan and Zisserman, 2015; Bochkovskiy et al., 2020). Such nature makes CV systems challenging to implement, limiting a wider deployment of high-throughput monitoring in animal industry.

Due to the difficulties described in the current CV approaches, tracking pig activity is a challenging task without considerable labor efforts. The objective of this paper is to develop a semi-supervised pipeline, Virtual Tag (VTag), to automate long-term tracking of group-housed pigs. In this pipeline, successful tracking algorithms (hereafter trackers) are implemented. They include Sparse Optical Flow proposed by Lucas and Kanade (LK) (Lucas and Kanade, 1981), multiple instance learning (MIL) (Babenko et al., 2009), and channel and spatial reliability (CSRT) (Lukežič et al., 2018) that learn representatives from the object of interest and to find the similar image region in the next input video frame. These algorithms are lightweight and require no specific computing resources such as graphics processing units (GPU). The implemented tracker substantially reduce efforts in labeling pig positions for every single frame. To start tracking, users can either assign initial positions, or VTag can predict the positions based on their motion, which is anticipated to be effective features under different monitoring environments. We validated VTag by four three-hundred-frame videos collected from our farming trials, and the benchmark test is performed to compare the performance and detected frames per second (FPS) of the implemented

trackers and other state-of-the-art models, such as YOLOv5 (Jocher et al., 2022) and Mask R-CNN (He et al., 2017). In addition, VTag is released as a friendly software tool in both a graphical user interface (GUI) and a Python library, allowing users to freely utilize the labeled data for their following research. Therefore, neither hard-coded features selected by human experts nor large training datasets labeled from a massive manual work are required in our pipeline. The complete algorithm and source code are available at https://github.com/vt-ads/vtag.

Materials and Methods

Data acquisition
All animal experiments were approved and carried out in accordance with the Virginia Tech Institutional Animal Care and Use Committee (IACUC) under protocol #19-182. The demonstrated video recordings were obtained from (Yu et al., 2021), which reported the image-based live body weight prediction of non-restrained grower pigs. The pigs entered the trial at 5 wk post-weaning. The imaging system was built with a laptop-controlled camera (Intel RealSense D425) that captured RGB and depth videos with resolution of 848 × 480 pixels. The camera was installed at a height of 2.25 m perpendicularly to the floor in each 5 × 7 ft pen, where pigs can freely move and walk during the entire recording. In each day, each monitored pen was recorded in a three-hundred-frame video at a rate of 6 frames per second. Raw videos were saved in Robot Operating System bag video format, and the decoder Intel RealSense Viewer was applied to obtain sequential image files as the input data. In this study, only RGB-converted grayscale images were used, and depth and color information were excluded from the pipeline. Each video clip had 300 timeframes. There were four video clips being evaluated for the performance of the presented pipeline: 3 clips contain 1, 2, and 3 pigs (denoted as 1-pig, 2-pig, 3-pig datasets in the following paragraphs), respectively. The last clip also contains 2 pigs, but more motion was observed (denoted as 2-pig (high motion)).

Implemented trackers
In VTag, 3 trackers, LK, CSRT, and MIL, are implemented by Python OpenCV (Bradski, 2000). To simplify the pipeline, users only need to provide 2 parameters to the trackers. First, positions coded in (x, y) coordinates are required to be set to start the tracking task. Each pig needs to be assigned to at least one starting position. The position can be provided automatically by VTag, which uses detected motion to propose starting positions for users. The detailed process is described in the following section. The second parameter is the size of the bounding box that covers the tracking area. The bounding box is centered on the starting position assigned from the first parameter, and it is desired to cover diverse textures (e.g., contour edges) for the best tracking performance.

Motion detection
Motion detection is useful to automatically propose potential pig positions for trackers. VTag will detect the pixels of interest (POI) over time frames, which are expected to cover entire animal bodies and help locate positions of pigs. The motion is quantified by the variation of pixel values in a time range of multiple neighboring time frames (Figure 1a). Assuming p_{xyt} is a pixel value at coordinate x, y of the time frame t, and p_{xy}.
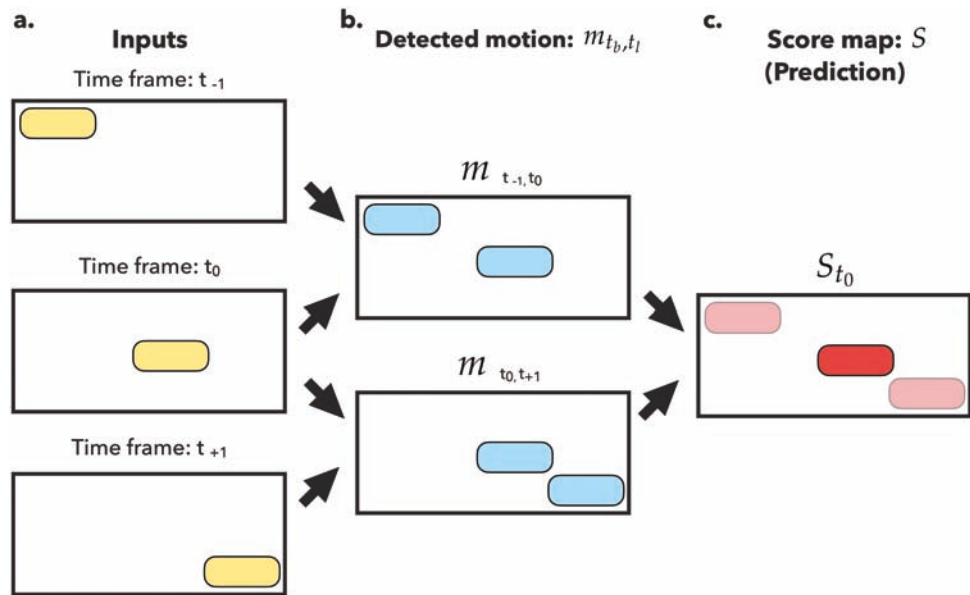
Figure 1. The computational approach to detect motions. (a) In this example, the input video containing three frames from $t-1$ to $t+1$. Yellow blocks represent the ground truths of animal positions. (b) Motions detected in the time range of (t_b, t_l). Blue blocks show the area with high motions. (c) The pixels with larger gradient of red means higher chance the areas are occupied by the studied objects.

is an averaged value in the range (t_b, t_l) at coordinate x, y. The detected motion $m_{t_b, t_l}(x, y)$ during the time range (t_b, t_l) is:

$$m_{t_b, t_l}(x, y) = \frac{\sum_{t \in (t_b, t_l)} (p_{xyt} - p_{xy\cdot})}{t_l - t_b}$$

where t_b, t_l are the beginning and the last time frames in the scanned time range, respectively. High pixel variation represents high motion values m for a pixel at coordinate x, y. If the studied pigs show up or leave out of the scope (pixel at (x, y)) in the middle of the time range (t_b, t_l), the pixel values therefore change from background colors to animal skin textures or from animal skin textures to background colors, resulting in high pixel variation (Figure 1b).

Pig positions

By computing pixel motion for 2 flanking, equal-length time ranges, the positions of all monitored pigs can be located. For example, to infer the livestock positions at time frame t_0, pixel motion m_{t_{-k}, t_0} and m_{t_0, t_k} are calculated, where t_{-k} and t_k are k frames before and after the observed frame t_0, respectively. Then, summing 2 motion values can obtain a score map S_{t_0} where high-value pixels are more likely to be occupied by the studied objects at the moment (Figure 1c). In this study, pixel scores $s_{t_0}(x, y)$ in the score map S_{t_0} are defined with 2 frames before and after the observed frame t_0 as

$$s_{t_0}(x, y) = \sum_{k=1}^{2} \left(m_{t_{-k}, t_0}(x, y) + m_{t_0, t_k}(x, y) \right)$$

The scores s were hypothesized to follow a skewed distribution in which most pixels are observed with low scores. Hence, the 99th quantile of scores was set as the threshold to binarize pixels into 2 categories: POI is assigned a value of 1 if its score is greater than the threshold, otherwise value 0 is assigned to represent a background pixel. Finally, for each

time frame t, we have a binary map B_t indicating the positions of the studied objects (i.e., pixel areas with non-zero values).

Refining motion detection and proposals of tracking points

As the pig positions were inferred from observed motion, motion caused by irrelevant sources, such as human activities or pigs from other pens, should be avoided. Although the camera scope is limited to the studied pigs, some unrelated movements will be detected by our score functions and become noisy signals. For example, the noisy movements can be vibrations caused by the occasional collision between the pigs and pens or human activities. Such noisy signals are usually small in pixel areas and can be removed by "blurring" the binary map B_t with a Gaussian kernel ω (Appendix) in convolutional operations. Each refined pixel $B'_t(x, y)$ is:

$$B'_t(x, y) = f(\omega * B_t(x, y))$$
$$= f \left(\sum_{dx=-m}^{m} \sum_{dy=-n}^{n} \omega(dx, dy) B_t(x + dx, y + dy) \right)$$

$$f(x) = \begin{cases} 1, & \text{if } x \geq 0.5 \\ 0, & \text{otherwise} \end{cases}$$

where B'_t is the refined binary map in which POIs are represented by values of ones, and background colors are noted as zeros at the time frame t. The lower and upper boundaries of the kernel ω are denoted by $-m, m$ of x-dimension and $-n, n$ of y-dimension, respectively. To save memory usage and avoid including noisy signals, we do not keep the positions of every POI but track their contours as representatives. We can obtain the finalized map B''_t containing the contours of objects with edge detection kernel γ (Appendix a) in the convolutional operation:
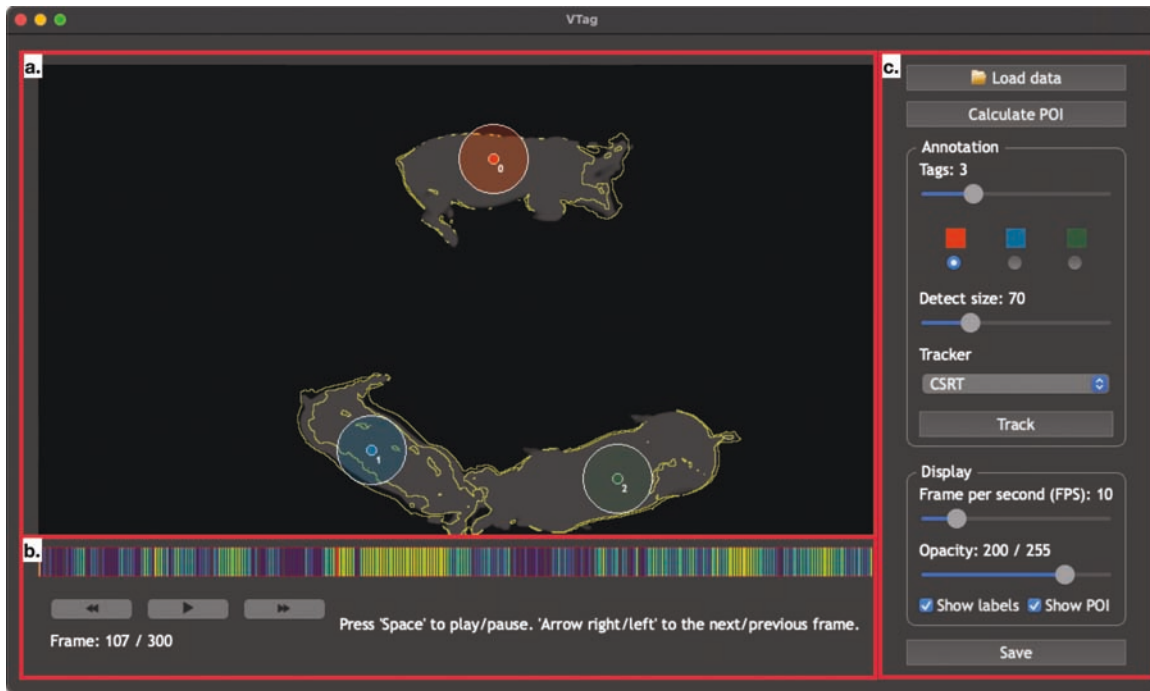
Figure 2. VTag graphical user interface. (a) Preview panel to display the tracking results. (b) Playback controller to traverse video frames and inspect tracking errors. (c) Configuration panel to fine-tune the tracking algorithm and import/export video data.

$$B''_t = \gamma * B_t$$

Instead of keeping the entire area of POIs with high motion, the derived map B''_t only records the contours of POIs. The contours of POIs in B''_t are further clustered into object identities. A similarity matrix of each POI coordinate was calculated as a clustering constraint. With the defined constraint, agglomerative hierarchical clustering from the Python library, scikit-learn (Pedregosa et al., 2018), was performed to cluster POIs into each object identity. The centroids of each cluster are proposed as initial tracking points for the implemented trackers.

Benchmark test

The implemented trackers were evaluated for their precision and computing time. To evaluate the precision, we manually labeled the central positions of each pig body as the ground truths. The precision error was determined by the Euclidean distance between the ground truth and the centroid of the predicted bounding box. To make the results comparable with other studies, the error was standardized by being divided by the diagonal distance of the video frame ranging from 0 to 1. In addition, as the tracking process may be unsuccessful when the similarity of two consecutive frames is low, human supervision is needed to provide new tracking positions to resume tracking. Hence, we also evaluated the number of supervision is needed to complete tracking the 300 frames in each dataset. To evaluate the computing time, the elapsed time to track one single frame is measured for 100 iterations. The time is presented in FPS by inverting the observed elapsed time. In addition to the implemented trackers, the object detection models, YOLOv5 and Mask R-CNN, pre-trained by the COCO dataset (Lin et al., 2014) are also included in the evaluation of computation time. It can help explore the possibility of adapting these pre-trained deep learning models in the pig tracking

tasks. The evaluation was run on a personal laptop, MacBook Pro (14-in., 2021) with Apple M1 Max chip, 10 CPU cores, and 32 GB RAM. The GPU resources were not utilized during the evaluation.

Results

Software interface

The VTag pipeline is released as a Python software and can be accessed by a GUI or an interactive Python session. There are 3 components that users can interact with in the GUI: the video previewer, the playback controller, and the configuration. The preview (Figure 2a) shows the video overlayed by the tracking results, which are presented by a centroid and its tracking window area. Different tracking points are colored differently to show pig entities. The video can be played, paused, and traversed to any video frame by interacting with the playback controller (Figure 2b). Each frame in the progress bar is colored in a gradient scale from yellow to blue, showing the tracking errors estimated from the implemented tracker. In the configuration panel (Figure 2c), parameters needed for the tracking task are tunable. In the panel, users can load a directory containing the video to start the tracking tasks, adjust the number of tracked objects and tracking size, and optimize the quality for displaying the tracking results. If users need to work with their own analysis in an interactive programming session, users can load VTag in Python as a library. The library has commands available that correspond to all the actions in the VTag GUI. In sum, VTag provides a friendly platform to annotate video data and generate informative farming guidance for pig activity.

Benchmark testing

The precision evaluation is presented in Figure 3, the standardized errors over frames were plotted in boxplots. Every 0.1 of
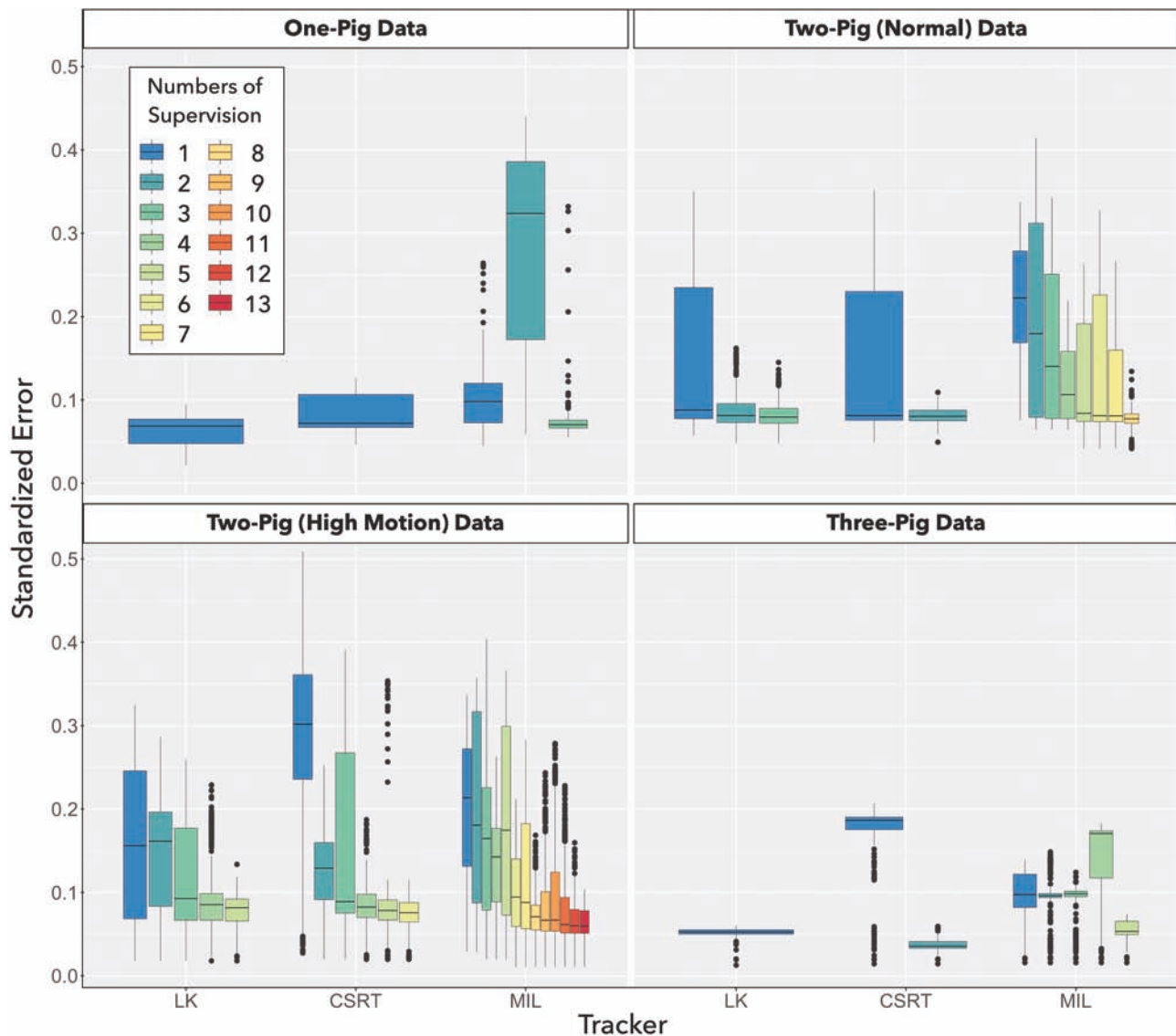
Figure 3. Evaluation of tracking precision. The standardized errors are plotted in box plots, which are colored in corresponding to the number of supervisions. Trackers, LK (sparse optical flow by Lucas and Kanade), CSRT (channel and spatial reliability), and MIL (multi-instance learning) are listed on the *x*-axis.
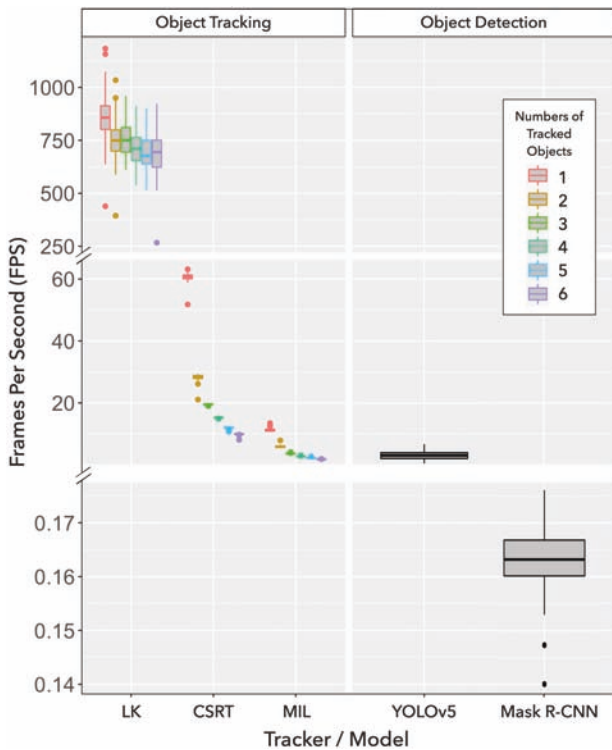
the standardized error is 26.29 cm in the presented datasets. The colors represent different supervisions. For example, the results shown in red are evaluated after 8 times of human supervision. With adequate human supervision, all trackers can precisely track pig activity with errors less than 22.82 cm in all the 4 datasets. In particular, the tracker LK can complete the tasks without any resuming supervision with the median errors of 18.03 and 13.81 cm for the datasets of one-pig and three-pig, respectively. The tracker CSRT performed similarly well with only one additional supervision with the median error of 16.3 cm in the studied datasets except the dataset of two-pig (high motion). Among the studied trackers, MIL has similar precision but requires more human supervision than others in all the dataset. It is noted that the number of tracked objects is not a major limiting factor when it comes to tracking precision. In this study, more supervision is needed when the objects are found to move rapidly and create blurry image features. When the pigs move rapidly, the input video with low FPS had latencies to display object positions timely. In the

2-pig dataset (high motion) although with similar precision, 7, 5, and 13 supervisions were needed to complete tracking the 300 frames for the 3 trackers, respectively.

The computing time is presented by FPS, which indicates how many frames the tracker can process per second (Figure 4). As results, LK tracked averaged FPS of 900 and showed outperformance in computing speed to other trackers by more than 100 folds. CSRT is the second fast tracker with a performance ranging from 9.9 FPS to 60.81 FPS in the tasks of tracking different number of pigs. MIL is found to be the slowest tracker, with as low FPS as 1.8 FPS when it tracked six pigs. It is also found that for the trackers CSRT and MIL, the numbers of tracked objects affect the tracking speed nonlinearly. Additionally, the pre-trained object detection models are evaluated in this study as well. Without enabling GPU resources, both models predict the studied videos slower than the presented trackers. Only 4 FPS and 0.17 FPS are processed by YOLOv5 and Mask RCNN, respectively.

Social interaction

The distance between studied subjects implies 2 types of general social interactions: separated or engaged. When the subjects engage closely, the distance values are low during the period of time frames. Otherwise, subjects are separated apart without much interaction. A line chart of the distance against the 300 timeframes was visualized to monitor such patterns, showing 4 peaks and 4 valley values from the 2-pig data (Figure 5). To examine whether the distance is an effective indicator for the interactions, video frames with peak and valley values were displayed. Consequently, in the frames with peak values, interaction was observed among pigs, and they were observed staying in 2 different corners of the pen at the examined time frame. On the other hand, in the frame with valley values, social interactions were observed for all inspected frames. Pigs were in the status of in-taking feeds alongside or chasing each other. From the examined 300 frames, the estimated distance between pigs is an accurate indicator to filter time frames where social interactions may occur.

By knowing the tracks of each pig, pixel movements per time frame were studied to monitor the activities individually. In the presented data, 2 studied pigs were denoted as "Pig_1" and "Pig_2". The median movement of Pig_1 and Pig_2 is 21.1 and 21.98 pixels per frame, which show no significant difference (P-value = 0.953) in overall activity (Figure 6a and b). However, individual-specific temporal pattern can be discovered by dissecting the activity at certain time frames. For example, during the first 50 frames, Pig_2 was much more active, the difference between Pig_1 and Pig_2 was especially revealed in those peak movements. Moreover, after the 50th frame, Pig_2 continuously had greater changes of accumulated movements over Pig_1. The superiority was 1739.7 pixels at the 50th frame, and it was later expanded to 3612.9 pixels at the 250th frame (Figure 6c). Finally, we inspected the synchronicity between pigs by comparing their movements per frame (Figure 6d). A moderate correlation (r = 0.605, P-value < 0.01) was observed in the studied data, which implied that the activity of each individual was not independent and were partially determined by its neighboring pig.

In addition to monitoring the temporal activity, spatial patterns of pig movement can be informative for herd management. Heat maps (Figure 7) generated from pixel-wise variation across all time frames provided insightful guide on what areas were visited most (coded in yellow). In the one-pig data (Figure 7a), middle-top and bottom-left regions have found to be the hot spots, which were the places to engage with neighboring pigs and the feeding area, respectively. Whereas in the 2-pig data (Figure 7b), there was no clear spatial trend of the subject activity. Most corners of the pen were visited by both pigs except the central area.
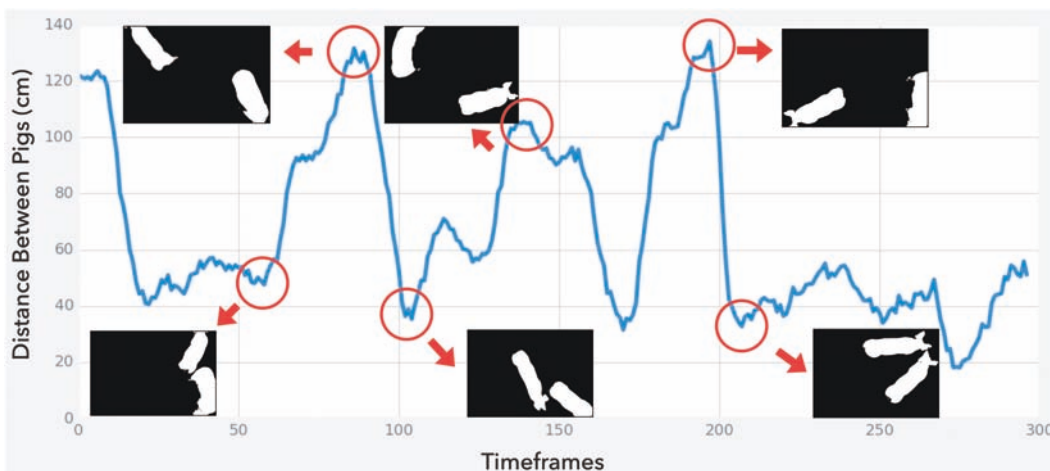

Figure 4. Evaluation of computing time by frame per second (FPS). FPS of each tracking task is plotted in box plots, which are colored corresponding to the number of tracked objects. Trackers, LK (sparse optical flow by Lucas and Kanade), CSRT (channel and spatial reliability), and MIL (multi-instance learning) are listed on the left column. The object detection models, YOLOv5 and Mask R-CNN, are plotted on the right column.

Figure 5. The predicted distances between pigs over all time frames in the two-pig data. The distance in centimeter is shown on the *y*-axis, and the *x*-axis represents the 300 timeframes. Six snapshots of the selected timeframes show the 2 extreme scenarios when pigs are closely engaged or separated apart.
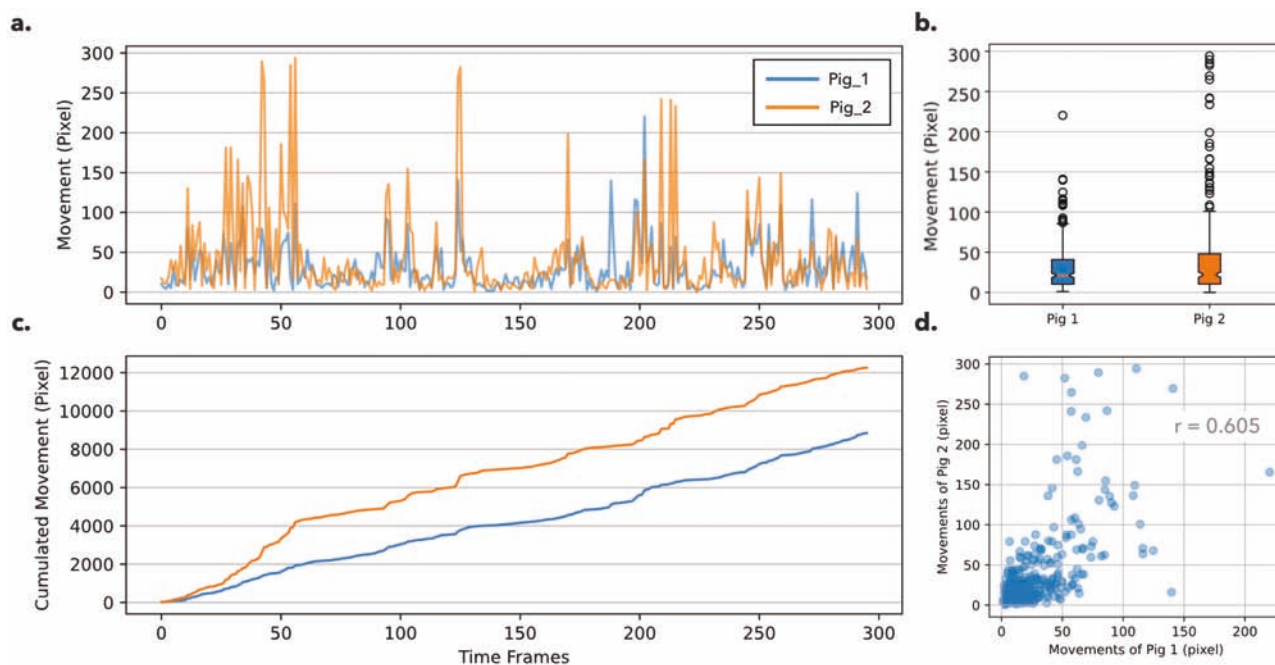
Figure 6. The movement of pigs in pixels in the two-pig dataset. Two pigs are colored by blue and orange. (a) A line chart showing the pixel movement (*y*-axis) per time frame (*x*-axis). (b) A boxplot comparing the movement between two pigs. (c) A line chart showing the accumulated pixel movement (*y*-axis) per time frame (*x*-axis). (d) A correlation plot between the movement of Pig 1 (*x*-axis) against Pig 2 (*y*-axis), and each dot represents the movements of the observed time frame.

Discussion

Continuously tracking pig activity from videos is an important initial step to monitor farming conditions in swine industry. Including animal diseases, welfare, and pen-scale social interactions, such complex monitoring tasks require detailed observation of pig activity. Many existing works have automated the tasks through the aid of CV technology but required massive human effort in preparing data sets to build an effective system. In contrast, this paper presented a semi-supervised pipeline, VTag, which does not require laborious work in setting up the training system. Solely relying on a top-view and grayscale video, VTag provides an efficient approach to continuously track the positions of group-housed pigs with an average error of 17.99 cm in the presented datasets. The results can serve as preliminary farming guidance to infer complex traits that used to require intensive labor resources.

For example, by continuously tracking pig positions with VTag, individual-level activity per unit time and walking speed can be estimated. This is important information for the trait assessment of pig lameness, which can be potential indicators of fractures, lesions, and development disease (Heinonen et al., 2013; Benjamin and Yik, 2019), and diminishes welfare in pigs. Hence, effectively evaluating lameness allows farmers to control economic losses from losing pigs with poor body conditions (Anil et al., 2009). Another important monitoring task that can be improved with VTag is tail biting in pigs. Because tail biting is linked to stressful farming conditions (D'Eath et al., 2014) and lower body weights (Marques et al., 2012), detecting the negative events at an early growing stage can be beneficial to both animal welfare and production. As the real-time pig positions are obtained automatically, the relative distance between individuals in the pen can be estimated. Behavioral researchers can use this information to filter a specific time range from an hour-long video: When the relative distance is low, it is more likely to observe tail-biting events.

In addition to evaluating pig behaviors through RGB videos, automation can be further facilitated by other sources. For example, videos containing depth information are useful to estimate pig body weights. Body weight is a critical trait associated with growing rate, feed efficiency, and meat biomass. Conventionally, pigs are weighed on the electronic scale in the pen, but it can be either inaccurate when more than one pig is standing on the scale or expensive if the scale is integrated into the feeding system. A past study has presented a video-based pipeline that can successfully estimate pig body weight with an RGB-depth camera by segmenting pig contours (Yu et al., 2021). By combining the existing work and VTag, which can continuously track pig positions, the fully automatic system of pig weighing is feasible for farms with limited resources.

The implemented trackers have successfully shown their great performance of tracking objects in their published papers (Babenko et al., 2009; Lukežič et al., 2018). However, the trackers failed to track pigs without any human supervision in our presented results. The potential reasons may be explained by the difference in the monitoring context. In our study, the tracked objects share similar morphological features. Even for the feature-rich areas, such as pig heads and tails where unique spatial patterns are observed, they were hard for trackers to distinguish the difference between different pig individuals. The trackers easily lost tracking the correct individual when two pigs frequently interact with each other in a short period. Another reason is the video quality. In the papers where the trackers were published, the demonstrated videos recorded at least 20 FPS. Whereas the studied datasets only have 6 FPS, which is a common setting in practical farming to reduce
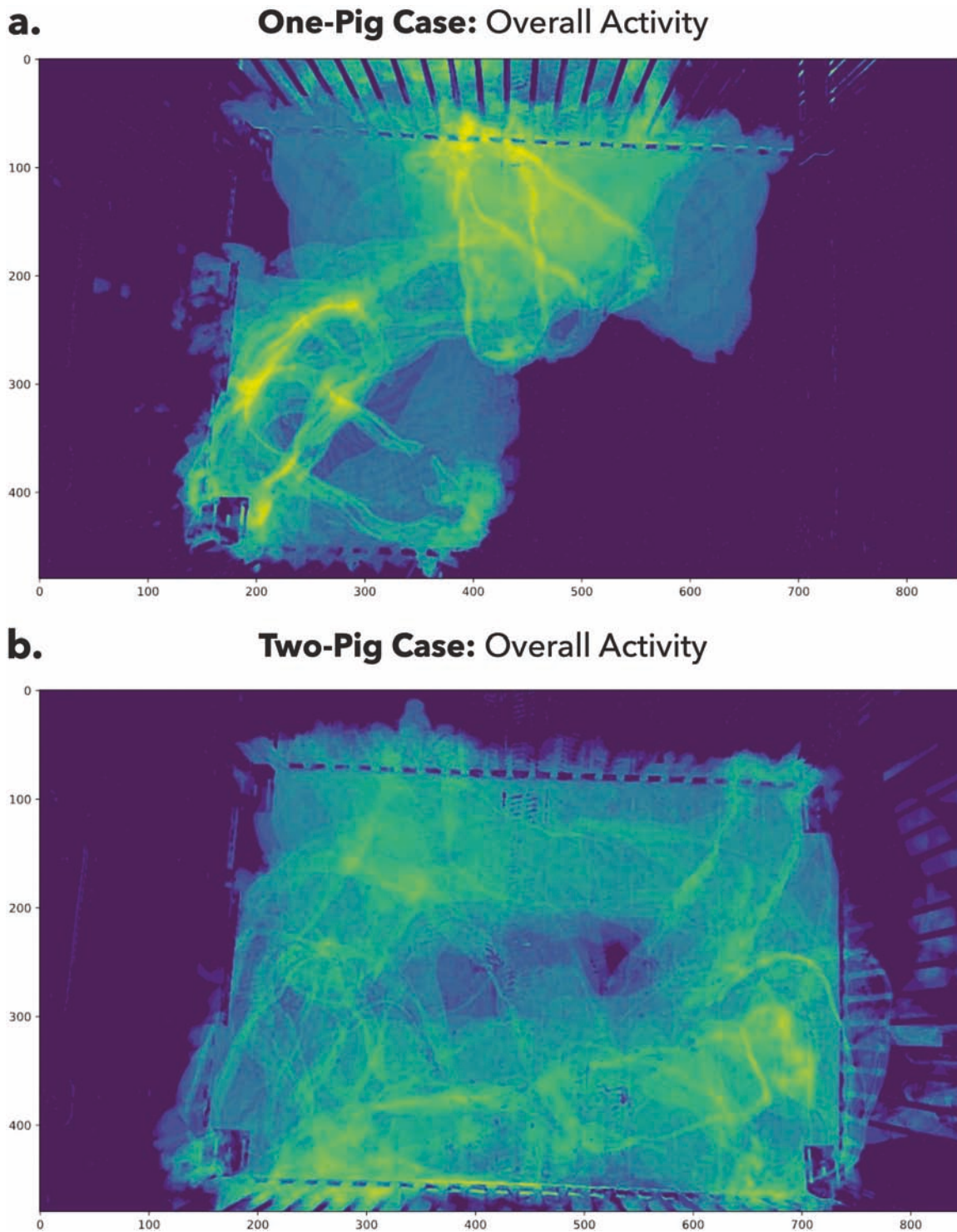
Figure 7. Heatmaps showing the spatial activity of pigs. The yellow areas are visited the most by the pigs, and the purple area indicates the least visited region in the view scope. (a) One-pig dataset; (b) two-pig dataset.

power dissipation and save data storage. Such low FPS videos reduce the similarity between consecutive frames and increase the chance of mismatching tracked features over time (Porikli and Tuzel, 2005; Li et al., 2007). Additionally, when the tracked object moves rapidly, the track features are more likely to become blurry in a low-FPS video. These limitations make the tracking task in livestock farming more difficult than the regular tracking task, where videos have 30 FPS, and the video frames are assumed to be similar in adjacent frames and the tracked features are unique compared to other objects.

We also included pre-trained models, YOLOv5 and Mask R-CNN, in our benchmark study. The low-FPS results indicate that it is difficult to fulfill real-time long-term monitoring in livestock farming without accessing GPU resources. Although we did not show their precision in the current work,

the detection results are not comparable with the presented trackers. It is because the models were trained by COCO datasets, in which top-view pig images are not included. In some video frames, pigs are either not detected, or 2 adjacent pigs are identified as the same object. Besides, without further modification of the models, it cannot force tracking the certain number of objects. These limitations make the evaluation difficult when we want to compare the precision of tracking the same number of pigs. In conclusion, the results suggest that the object detection models are not as suitable as object tracking algorithms in the pig monitoring tasks.

Further improvement can be made in the current version of VTag. For example, VTag is found to mis-identified pig identities when individuals frequently contact each other in a short time as described earlier. Although the wrong labels can be corrected manually, it still requires time and effort from humans' supervision. One way to reduce such error is to utilize a strategy called template learning, which was discussed in the literature (Lan et al., 2017; Wang et al., 2019). The general idea of this strategy is to first select the video frames in which pigs are not in close range of their neighbors. Then, the pig morphology observed in the selected frames is learned as "templates". Finally, the model can use the templates to update the predictions in the frames where pig positions are mis-identified due to the closed distance between pigs.

In addition to improving the algorithm, adding information by wearable devices is also helpful to increase the monitoring precision. The devices, including motion sensors, magnetometers, gyroscopes, and GPS receivers, have been widely used to monitor behavioral patterns in large farming environments (i.e., pastures and barns) (Chapa et al., 2021; Li et al., 2021; Nikodem, 2021; Perisho and Hajnal, 2021; dos Reis et al., 2021). Specifically, with the use of tracking collar wore by pastured livestock, grazing behaviors were successfully identified for cattle (Brennan et al., 2021) and sheep (dos Reis et al., 2021). The spatial resolution for outdoor studies was further improved to centimeter-level by coupling sensor collars with signal receivers deployed around the farm (Li et al., 2021). In group-housed scenarios, Smartbow (Weibern, Austria), a commercialized ear-tag sensor system, also demonstrated promising results in monitoring complex interactions on reproduction traits in swine cohorts (Perisho and Hajnal, 2021) and feeding behaviors of cows (Chapa et al., 2021). In conclusion, by coupling the VTag algorithm and the described improvement, the automation of the assessment system is expected to monitor more complex farm settings.

Conclusion

The presented semi-supervised pipeline, VTag, can track pigs in the video clips with a minimal human supervision to achieve decent precision. Among the tested trackers, a simple algorithm as spare optical flow can achieve ideal balance between precision and computing speed for the pig tracking tasks. The observed median error in the studied videos is no larger than 17.99 cm in predicting positions of each individual. Such performance provides individual-level guidance for farming management in knowing animal activity, visiting hot spots, and social interaction. In conclusion, this study reports a rapid, precise deployment of high-throughput assessment for continuous pig tracking. With the efficient monitoring system, animal health and derived products can be greatly improved.

Conflict of Interest Statement

The authors declare no real or perceived conflicts of interest.

Appendix

$$\text{Gaussian kernel}\,\omega \;=\; \begin{bmatrix} 1 & 4 & 1 \\ 4 & 9 & 4 \\ 1 & 4 & 1 \end{bmatrix}$$

$$\text{Edge detection kernel}\,\gamma \;=\; \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

References

Anil, S. S., L. Anil, and J. Deen. 2009. Effect of lameness on sow longevity. *J. Am. Vet. Med. Assoc.* 235:734–738. doi:10.2460/javma.235.6.734.

Babenko, B., M.-H. Yang, and S. Belongie. 2009. Visual tracking with online Multiple instance learning. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA; p. 983–990. doi:10.1109/cvpr.2009.5206737.

Benjamin, M., and S. Yik. 2019. Precision livestock farming in swine welfare: a review for swine practitioners. *Animals* 9:133–153. doi:10.3390/ani9040133.

Bochkovskiy, A., C. -Y. Wang, and H. M. Liao. 2020. YOLOv4: optimal speed and accuracy of object detection. doi:10.48550/arXiv.2004.10934.

Bradski, G. 2000. *The OpenCV Library*. Dr. Dobb's Journal of Software Tools.

Brennan, J., P. Johnson, and K. Olson. 2021. Classifying season long livestock grazing behavior with the use of a low-cost GPS and accelerometer. *Comput. Electron. Agric.* 181:1–7. doi:10.1016/j.compag.2020.105957.

Chapa, J. M., L. Lidauer, A. Steininger, M. Öhlschuster, et al. 2021. Use of a real-time location system to detect cows in distinct functional areas within a barn. *JDS Commun.* 2:217–222. doi:10.3168/jdsc.2020-0050.

Chen, C., W. Zhu, and T. Norton. 2021. Behaviour recognition of pigs and cattle: journey from computer vision to deep learning. *Comput. Electron. Agric.* 187:1–23. doi:10.1016/j.compag.2021.106255.

Clark, M. A., N. G. G. Domingo, K. Colgan, S. K. Thakrar, et al. 2020. Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets. *Science* 370:705–708. doi:10.1126/science.aba7357.

D'Eath, R. B., G. Arnott, S. P. Turner, T. Jensen, et al. 2014. Injurious tail biting in pigs: how can it be controlled in existing systems without tail docking? *Animal* 8:1479–1497. doi:10.1017/S1751731114001359.

Fang, C., J. Huang, K. Cuan, X. Zhuang, and T. Zhang. 2020. Comparative study on poultry target tracking algorithms based on a deep regression network. *Biosyst. Eng.* 190:176–183. doi:10.1016/j.biosystemseng.2019.12.002.

Fuentes, A., S. Yoon, J. Park, and D. S. Park. 2020. Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. *Comput. Electron. Agric.* 177:1–11. doi:10.1016/j.compag.2020.105627.

He, K., G. Gkioxari, P. Dollár, and R. Girshick. 2017. Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy; p. 2980–2988. doi:10.1109/ICCV.2017.322.

He, K., X. Zhang, S. Ren, and J. Sun. 2015. Deep residual learning for image recognition. doi:10.48550/arXiv.1512.03385.

Heinonen, M., O. Peltoniemi, and A. Valros. 2013. Impact of lameness and claw lesions in sows on welfare, health and production. *Livestock Science* 156:2–9. doi:10.1016/j.livsci.2013.06.002.

Huang, W., W. Zhu, C. Ma, and Y. Guo. 2020. Weber texture local descriptor for identification of group-housed pigs. *Sensors*. 20:1–18. doi:10.3390/s20164649.

Jocher, G., A. Chaurasia, A. Stoken, J. Borovec, et al. 2022. ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference. doi:10.5281/zenodo.3908559. Last accessed date: May 17, 2022.

Lan, X., P. C. Yuen, and R. Chellappa. 2017. Robust MIL-based feature template learning for object tracking. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 31, San Francisco, California, USA. Available from: https://ojs.aaai.org/index.php/AAAI/article/view/11220.

Larsen, M. L. V., H. M. Andersen, and L. J. Pedersen. 2019. Changes in activity and object manipulation before tail damage in finisher pigs as an early detector of tail biting. *Animal* 13:1037–1044. doi:10.1017/S1751731118002689.

Leonard, S. M., H. Xin, T. M. Brown-Brandl, and B. C. Ramirez. 2019. Development and application of an image acquisition system for characterizing sow behaviors in farrowing stalls. *Comput. Electron. Agric.* 163:1–19. doi:10.1016/j.compag.2019.104866.

Li, B., L. Liu, M. Shen, Y. Sun, and M. Lu. 2019a. Group-housed pig detection in video surveillance of overhead views using multi-feature template matching. *Biosyst. Eng.* 181:28–39. doi:10.1016/j.biosystemseng.2019.02.018.

Li, W., J. D. Bakker, Y. Li, S. Zheng, and F. Y. Li. 2021. Applying a high-precision tracking system to distinguish the spatiotemporal patterns of animal movement in grassland ecology. *Biol. Conserv.* 255:1–6. doi:10.1016/j.biocon.2021.109016.

Li, X., C. Cai, R. Zhang, L. Ju, and J. He. 2019b. Deep cascaded convolutional models for cattle pose estimation. *Comput. Electron. Agric.* 164:1–11. doi:10.1016/j.compag.2019.104885.

Li, Y., H. Ai, T. Yamashita, S. Lao, and M. Kawade. 2007. Tracking in low frame rate video: a cascade particle filter with discriminative observers of different lifespans. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA; p. 1–8.

Lin, T. -Y., M. Maire, S. Belongie, L. Bourdev, et al. 2014. Microsoft COCO: common objects in context. doi:10.48550/arXiv.1405.0312.

Liu, H., A. R. Reibman, and J. P. Boerman. 2020. Video analytic system for detecting cow structure. *Comput. Electron. Agric.* 178:1–13. doi:10.1016/j.compag.2020.105761.

Lucas, B. D., and T. Kanade. 1981. Lucas and Kanade - An Iterative Image Registration Technique with an application to stereo vision. In: Proceedings of Imaging Understanding Workshop, Pittsburgh, Pennsylvania, USA; p. 121–130.

Lukežič, A., T. Vojíř, L. Zajc, J. Matas, and M. Kristan. 2018. Discriminative correlation filter tracker with channel and spatial reliability. *Int. J. Comput. Vis.* 126:671–688. doi:10.1007/s11263-017-1061-3.

Marques, B. M. F. P. P., M. L. Bernardi, C. F. Coelho, M. Almeida, et al. 2012. Influence of tail biting on weight gain, lesions and condemnations at slaughter of finishing pigs. *Pesq. Vet. Bras.* 32:967–974. doi:10.1590/S0100-736X2012001000003.

Morota, G., R. V. Ventura, F. F. Silva, M. Koyama, and S. C. Fernando. 2018. Big data analytics and precision animal agriculture symposium: machine learning and data mining advance predictive big data analysis in precision animal agriculture. *J. Anim. Sci.* 96:1540–1550. doi:10.1093/jas/sky014.

Nikodem, M. (2021). Bluetooth Low Energy Livestock Positioning for Smart Farming Applications. In: M. Paszynski, D. Kranzlmüller, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot, editors. *Computational Science – ICCS 2021*. Springer International Publishing, Cham; p. 55–67.

Okura, F., S. Ikuma, Y. Makihara, D. Muramatsu, et al. 2019. RGB-D video-based individual identification of dairy cows using gait and texture analyses. *Comput. Electron. Agric.* 165:1–9. doi:10.1016/j.compag.2019.104944.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, et al. 2018. Scikit-learn: machine learning in Python. *arXiv*. doi:10.48550/arXiv.1201.0490.

Perisho, S., and A. Hajnal. 2021. Characterization of swine behavior and production using measurements collected via indoor positioning system. *Appl. Anim. Behav. Sci.* 234:1–7. doi:10.1016/j.applanim.2020.105164.

Porikli, F., and O. Tuzel. 2005. Object tracking in low-frame-rate video. In: A. Said and J. G. Apostolopoulos, editors. *Electronic Imaging 2005*. San Jose, CA; p. 72. doi:10.1117/12.587907

dos Reis, B. R., Z. Easton, R. R. White, and D. Fuka. 2021. A LoRa sensor network for monitoring pastured livestock location and activity1. *Translat. Anim. Sci.* 5:1–9. doi:10.1093/tas/txab010

Ren, K., G. Bernes, M. Hetta, and J. Karlsson. 2021. Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning. *J. Syst. Archit.* 116:1–7. doi:10.1016/j.sysarc.2021.102139.

Simonyan, K., and A. Zisserman. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv*:1409.1556 [cs]. Available from: http://arxiv.org/abs/1409.1556

Tu, S., H. Liu, J. Li, J. Huang, et al. (2020). Instance segmentation based on mask scoring R-CNN for group-housed pigs. In: *2020 International Conference on Computer Engineering and Application (ICCEA);* p. 458–462.

Wang, N., Y. Song, C. Ma, W. Zhou, et al. (2019). Unsupervised deep tracking. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). p. 1308–1317.

Xu, B., W. Wang, G. Falzon, P. Kwan, et al. 2020. Automated cattle counting using Mask R-CNN in quadcopter vision system. *Comput. Electron. Agric.* 171:1–12. doi:10.1016/j.compag.2020.105300.

Yang, A., H. Huang, C. Zheng, X. Zhu, et al. 2018. High-accuracy image segmentation for lactating sows using a fully convolutional network. *Biosyst. Eng.* 176:36–47. doi:10.1016/j.biosystemseng.2018.10.005.

Yu, H., K. Lee, and G. Morota. 2021. Forecasting dynamic body weight of nonrestrained pigs from images using an RGB-D sensor camera. *Translational Animal Science* 5:1–9. doi:10.1093/tas/txab006.

Zhang, L., H. Gray, X. Ye, L. Collins, and N. Allinson. 2019. Automatic individual pig detection and tracking in pig farms. *Sensors* 19:1188–1207. doi:10.3390/s19051188.