



Multi-animal tracking in transition: Comparative insights into established and emerging methods

Anne Marthe Sophie Ngo Bibinbe ^a, Patrick Gagnon ^b, Jamie Ahloy-Dallaire ^{a, },
Eric R. Paquet ^{a, ,*}

^a Animal Science Department, Laval University, Quebec City, QC, Canada

^b Centre de développement du porc du Québec (CDPQ), Lévis, QC, Canada

ARTICLE INFO

Dataset link: <https://github.com/ngobibinbe/Tracking-Benchmark>

Keywords:

Benchmark
Multi-animal tracking
Multi-object tracking
Pigs
Livestock

ABSTRACT

Precision livestock farming requires advanced monitoring tools to meet the increasing management needs of the industry. Computer vision systems capable of long-term multi-animal tracking (MAT) are essential for continuous behavioral monitoring in livestock production. MAT, a specialized subset of multi-object tracking (MOT), shares many challenges with MOT, but also faces domain-specific issues, including frequent animal occlusion, highly similar appearances among animals, erratic motion patterns, and a wide range of behavior types.

While some existing MAT tools are user-friendly and widely adopted, they often underperform compared to state-of-the-art MOT methods, which can result in inaccurate downstream tasks such as behavior analysis, health state estimation, and related applications. In this study, we benchmarked both MAT and MOT approaches for long-term tracking of pigs. We compared tools such as DeepLabCut and idTracker with MOT-based methods including ByteTrack, DeepSORT, Cross-Input Consistency, and newer approaches like Track-Anything and PromptTrack.

All methods were evaluated on a 10-minute pig tracking dataset. Our results demonstrate that, overall, MOT approaches outperform traditional MAT tools, even for long-term tracking scenarios. These findings highlight the potential of recent MOT techniques to enhance the accuracy and reliability of automated livestock tracking.

1. Introduction

Animal production represents 40% of the global agri-food sector's income. To improve the quality of animal production, the precision livestock sector, estimated at 3.2 billion dollars in 2022 and projected to reach 7 billion by 2030, is being developed to ensure effective livestock management [1]. To do so, it will require to be able to monitor productivity, health parameters, and animal welfare in real time [31]. One innovative approach involves using computer vision to track animals via video cameras, enabling further analyses such as behavior analysis and abnormal event detection. With recent developments in deep learning, computer vision-based tracking is promising because it is a non-invasive approach to closely monitor animal behavior over time. As a consequence, multi-animal tracking (MAT) has garnered increasing interest in the livestock sector. Multi-animal tracking is a subclass of multi-object tracking (MOT) focused specifically on animals, which

involves detecting objects in each frame of a video and assigning them unique identities throughout the video.

Like MOT, MAT faces challenges such as frequent occlusions, similar animal appearances, and interactions between objects [17]. However, MAT also presents domain-specific challenges that are often more pronounced, including random and erratic movements, similar appearances, a wide range of behavioral patterns, and complex motion dynamics which might induce occlusion [39]. These factors require careful consideration when developing and applying MAT approaches.

While MOT has been extensively explored by the community, MAT remains underexplored in comparison, and there is a lack of comprehensive recent benchmarks evaluating MAT approaches in livestock particularly on long-term scenarios (e.g. several minutes). On the other hand, the application of MAT in livestock has focused mainly on animal tracking tools, as seen in existing benchmarks related to animal tracking, which compare unsupervised approaches using methods such as thresholding, ellipse fitting, and filter by size for detection [25,37].

* Corresponding author.

E-mail address: eric.paquet@fsaa.ulaval.ca (E.R. Paquet).

<https://doi.org/10.1016/j.atech.2025.101543>

Received 25 August 2025; Received in revised form 14 October 2025; Accepted 16 October 2025

These approaches are often emphasized because they do not require annotations, even for the detection step, or because they offer graphical interfaces that assist non-AI specialists in performing tracking. However, these methods are less effective than methods that perform supervised detections such as the method proposed in [16], especially in scenarios that involve many animals.

With recent advances in MOT, several approaches have emerged that are well-suited to address these livestock-specific MAT challenges [16,26,30,27]. Furthermore, recent developments in foundational models have opened the door to deep learning based detection and segmentation with no need of training data, thanks to models such as SAM [13], MDETR [12], and OWLv2 [19]. These models offer promising new directions for annotation-free tracking in livestock contexts.

Briefly, MOT is composed of two main underlying tasks: detection and identification of objects from one frame to another [17].

The two main tasks could further be classified as unsupervised or supervised depending on how they perform object detection and identification. The existing methods can be grouped into the following three main categories.

- Unsupervised detection and unsupervised identification
- Supervised detection and unsupervised identification
- Supervised detection and supervised identification

In each category, detection and identification can be performed jointly or separately. The literature includes methods from all of these categories. However, approaches that rely on supervised identification are often impractical in the livestock context, as each new animal would require reannotation at the different stages of the life of the animal. This makes such methods difficult to apply effectively in real-world livestock settings. Nevertheless, each group of MOT approaches includes methods capable of addressing the specific challenges of MAT in livestock.

Contribution. We conducted a benchmark study to compare state-of-the-art MOT and MAT approaches in livestock context across the main categories previously described, excluding methods based on supervised identification, which are less applicable in livestock contexts.

We included two widely used MAT methods: DeepLabCut [16] and idTracker [27].

We also evaluated MOT approaches with supervised detection: ByteTrack [40], DeepSORT [36], and Cross-Input Consistency [3], as well as fully unsupervised methods such as Track-Anything [38] and PromptTrack [22].

The benchmark was carried out on a 10-minute video featuring 15 active growing-finishing pigs in a pen [23]. The video was annotated at approximately one-second intervals, with five keypoints labeled per animal (ears, nose, neck, and tail), along with individual identities.

The goal of this benchmark is to assess the performance of leading MAT and MOT tools for long-term pig tracking. Our primary contribution is a comprehensive analysis of the strengths and limitations of each method, providing practical guidance for researchers and practitioners seeking to select and apply the most appropriate tools for livestock tracking and behavior analysis.

Our results indicate that MOT approaches (with supervised detection) outperform current MAT tools. Furthermore, among the fully unsupervised methods, Track-Anything and PromptTrack showed better performance than idTracker and produced results comparable to ByteTrack (approach with supervised detection), primarily due to the superior quality of their detection modules.

Paper Structure. The remainder of this paper is structured as follows: Section 2 reviews related work on MAT, MOT, and existing benchmarks. Section 3 presents the experimental protocol. Section 4 reports the results, which are analyzed to highlight the strengths and limitations of each approach. Finally, Section 5 concludes the paper and outlines directions for future research.

2. Background

2.1. Animal tracking benchmark

Animal tracking benchmarks have already been proposed in the literature [39,25,37,21], but unfortunately they have not fully addressed the diversity of the selected approaches (e.g., unsupervised vs. supervised detection, MAT vs. MOT). Panadeiro et al. [25] provided a detailed comparison of MAT tools, but lacked an experimental analysis of these tools. On the other hand, Wurtz et al. [37] experimentally compared four MAT tools ToxTrac [30], BioTracker [20], idTracker [27], and Ctrax [6] on three short-term one-minute videos of pigs in the livestock sector. Su et al. [21] experimentally compared DeepSORT [36], StrongSORT [10], and their own method for black cattle tracking. Zhang et al. [39] offered a broader benchmark for MOT approaches in the context of wild animal videos, encompassing a wide range of species. The dataset selection in [39] covers diverse animal species, with videos averaging 426 frames (14.2s x 30 FPS), and the longest video containing 2,268 frames (75.6s x 30 FPS). While [25] and [37] focused only on unsupervised MAT tools, [21] and [39] explored supervised MOT approaches. In our work, we evaluate both MAT and MOT approaches on a 10-minute, 25 FPS video (total of 15,000 frames with 782 annotated) and also incorporate fully unsupervised MOT methods, which have emerged with recent advances in zero-shot object detection and segmentation [23].

2.2. MOT approaches

There have been various propositions to tackle MOT in the literature, with most of the improvements being made in tracking by detection methodology. In this approach, object detection is first performed to detect the objects, and then objects are associated across frames through matching matrices. The matrices are built based on the probability that two detected objects in consecutive frames correspond to the same entity (i.e., have the same identity). The first prominent approach in this category is SORT [5], which uses a Kalman filter [35] to predict the positions of objects in the next frame. To create a correspondence matrix between consecutive frames, distances between the predicted positions and detected objects in the next frame are used as matching weights. Then it uses the Hungarian algorithm [15] to match the detected objects on the next frame and the current frame. The idea behind the use of the Hungarian algorithm is to provide the set of associations (assignments) with the lowest distance cost. Based on this strategy, many approaches have been developed. DeepSORT [36] improves SORT by adding appearance features as a matching criterion between objects in consecutive frames. In the two aforementioned approaches, a threshold is fixed on detection before starting the association, so detections with low confidence are not considered. However, when an object is occluded, its detection confidence score becomes low, increasing the chances of losing track of that object. To overcome this limitation, ByteTrack [40] proposed keeping low-confidence objects and matching them after high-confidence detected objects to recover missed objects. Many other approaches have improved DeepSORT, such as StrongSORT [10], BoT-SORT [2], BoostTrack [32], and many others, which enhance its feature extraction for appearance matching with advanced models and improve its data association methods.

Other methods have proposed joint detection and tracking approaches in a supervised manner. In this category, TransMOT [9] proposed an approach based on graph transformers that treats tracking as a prediction problem. FairMOT [41], on the other hand, formulates tracking as a multitask learning problem with two prediction heads: one for detection and another for extracting appearance features (the re-identification (Re-ID) head). With both heads sharing the same backbone, each head contributes to improving the other during the learning stage.

To work around the lack of annotations in MOT, some have proposed self-supervised strategies for tracking individuals. In this category, some

prominent works include [34] and [3]. In [34] objects are tracked by training a model to predict where the pixel colors of one frame have moved in the next frame, based on the input provided to the model. This process allows the model to detect where pixels of an object have moved and locate its new position in the next frame. In [3], on the other hand, they train two Recurrent Neural Networks (RNNs) to predict object identities. The two RNNs are given different inputs, where some frames are hidden from one or the other, and they are trained to provide the same identities for objects in shared frames.

More recently, with improvements in foundational models capable of performing segmentation and detection on images, fully deep learning-based unsupervised tracking approaches have emerged. For example, Track-Anything [38] leverages Segment Anything Model (SAM) [14] and XMem [8] for tracking. Similarly, PromptTrack [22] uses OWLv2 [19] for textual prompt-based detection and ByteTrack [40] for tracking. [7] have also proposed a tracking methodology using an image-level model for segmentation, combined with a model for temporal propagation, merging the two to provide coherent tracking. With this architecture, as well as prompt-based foundational detectors, foundational models for segmentation could also be utilized. In this category, we also find Type-to-track [24], which uses text descriptions of objects and foundational models to propose generalized tracking.

2.3. MAT approaches

A lot of researchers track animals to understand their behavior and interactions inside a group. To achieve this, many MAT tools have been developed to facilitate the life of researchers. These tools are mostly user-friendly and come with many features, such as behavior analysis, labeling facilities, GUI, etc. Some prominent tools in this category include idTracker, ToxTrac, DeepLabCut, SLEAP, and others [26,16,30,27]. [16] proposed a MAT methodology using the popular tool DeepLabCut. The first step in this method is pose estimation, where keypoints of the animals and the connectivity score between these points (PAF) are provided by a pre-trained model. These points are then grouped into individual animals using their affinity scores, and tracklets are created using a box or ellipse tracker. As input, DeepLabCut requires the number of animals in the scene. With this input and the tracklets, DeepLabCut is able to stitch these tracklets together to maintain the correct number of animals in the scene. This is achieved by reducing the stitching problem to an s-t cut problem, where the nodes are the tracklets and the weights between the nodes are derived from the distance in both time and space between the end of an earlier tracklet and the beginning of a later tracklet. Similarly, SLEAP [26] proposed a MAT tool that works similarly to DeepLabCut in that it performs keypoint detection before reconstructing individual animals using PAFs, and then tracks the animals using optical flow. On the other hand, idTracker [27] proposed a completely unsupervised methodology, even for detection. First, each frame is normalized to its mean to remove the background. Then, blocks with normalized intensity higher or lower than a certain threshold are assumed to represent animals. It detects animals by subtracting the background using an average model and segments them using thresholding and a filter based on the given size of the animals. The animals are then tracked using the overlap of blocks across frames, with a deep neural network learning the features of animals in case of occlusion. ToxTrac [30], is a fully unsupervised tracking methodology similar to idTracker and uses adaptive thresholding to detect animals before tracking them based on distance, speed, direction, and regions of interest.

3. Methods

3.1. Protocol

For our experiments, we used the dataset from [23], which consists of a 10-minute video featuring 15 pigs actively interacting in a pen, annotated every second on average using Visual Object Tagging Tool (VoTT).

Table 1

MAT and MOT approaches benchmarked in this work. (sup: supervised, uns: unsupervised, boxes: using bounding boxes, points: using keypoints).

Method	Detection	Identification	Point/boxes
idTracker	uns	uns	uns
PromptTrack	uns	uns	uns
Track-Anything	uns	uns	uns
DeepLabCut	sup	uns	points
ByteTrack	sup	uns	boxes
Cross-Input Consistency	sup	uns	boxes
DeepSORT	sup	uns	boxes

Table 2

Keypoints detection performance for the 3 tested DeepLabCut configurations (2 keypoints, 3 keypoints and 5 keypoints) by considering $IOU > 0$. Here the minimal enclosing bounding box is considered to evaluate the performance.

Methods	F1 score	Recall	Accuracy
2 keypoints (neck + tail)	0.92	0.92	0.92
3 keypoints (2 ears + tail)	0.85	0.85	0.85
5 keypoints (nose, 2 ears, neck, and tail)	0.98	0.98	0.98

There are two types of annotations in this dataset: keypoints and bounding boxes for each animal. Regarding point annotations, each pig has its two ears, neck, tail, and nose annotated when visually available. From these keypoints, a bounding box was generated as the minimum bounding rectangle. In total, 782 frames were annotated, yielding 11,730 annotations of pigs, including bounding boxes, five keypoints, and their identity [23].

3.2. Selected tracking approaches

For our analysis, we selected the methods listed in Table 1. The purpose is to assess one SOTA (State of the art) approach in each MOT category to compare them consistently. For MAT, idTracker (version 5.2.12) and ToxTrac are SOTA approaches widely used by researchers since they are user-friendly with graphical user interfaces to assist the user. Unfortunately, we were unable to run ToxTrac on our 10-minute video due to its length. As a result, we were unable to evaluate its performance on our dataset. We also included PromptTrack (version 1.0.2) and Track-Anything (last accessed on May 2024) to our benchmark, as both are unsupervised tracking approaches leveraging OWLv2 and SAM, thus requiring no prior supervised training. The purpose of including them in the analysis is to determine if they offer a better alternative to traditional unsupervised MAT approaches, which are commonly used.

DeepLabCut, another widely used tool by researchers, has the particularity of detecting keypoints before reconstructing and tracking the animal. Although DeepLabCut requires annotated keypoint data for detection, we decided to compare it to other SOTA MOT approaches that also require bounding box annotations for detection to evaluate their performance in scenarios with many animals and erratic movement, such as in pig livestock. Since DeepLabCut performance depends on the number of keypoints, we assessed its performance with different numbers of keypoints: 2 keypoints (neck and tail), 3 keypoints (2 ears and tail) or 5 keypoints (nose, 2 ears, neck, and tail). DeepLabCut (version 2.3.9) was trained to detect keypoints and PAFs over 130 K epochs splits of 90% for training and 10% for test since it is the default configuration. The detection performance is reported in Table 2. DeepLabCut also requires the definition of a skeleton, for the 2 keypoint we used: (neck and tail), for 3 keypoints we used: (tail, left ear), (tail, right ear), (left ear, right ear), and for 5 keypoints we used: (nose, left ear), (nose, right ear)(nose, neck), (left ear, right ear), (neck, tail) (neck, right ear), (left ear, neck).

For the supervised tracking approaches, we selected ByteTrack (last accessed on May 2024), DeepSORT (last accessed on May 2024), and

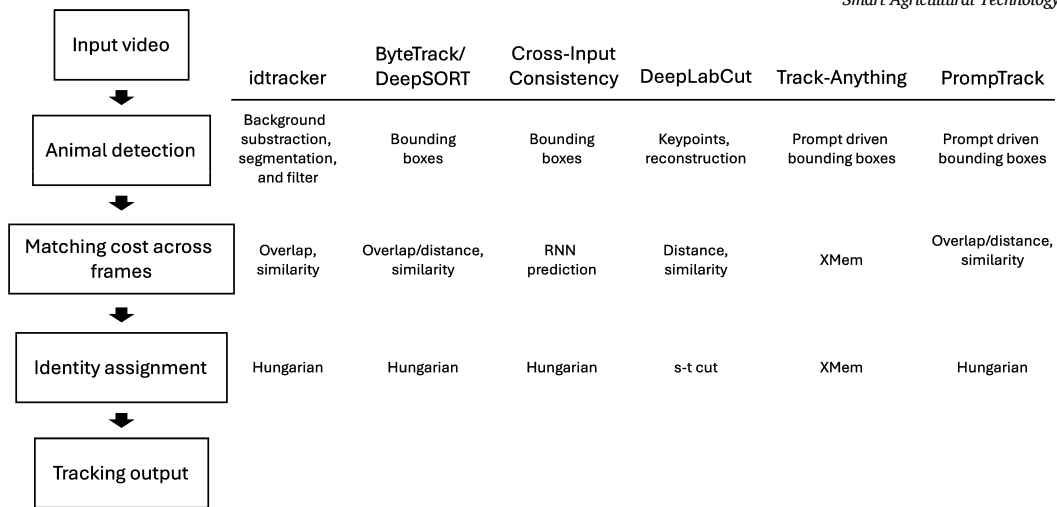


Fig. 1. Diagram presenting the different steps necessary to perform tracking and the specificity of the selected tracking algorithms.

Cross-Input Consistency (last accessed on May 2024) to cover a variety of tracking approaches. All three use a pretrained detection model that was trained using bounding boxes generated with the enclosing rectangle of the 5 keypoints of each animal. We trained a YOLOX [11] model from 11,700 annotations of pigs (80% training, 20% validation) using 500 epochs and got an mAP of 96.5% for the training set and 95% for the validation set. On the full video the F1 score on detection ($IOU > 0$) is 98%. Due to tracker post-processing, some detections may be filtered out if no identity is assigned. Consequently, we report in Table 5 the detection performances obtained after tracking. The differences between the selected approaches are presented in Fig. 1.

3.3. Evaluation metrics

To compare the performance of the different MOT and MAT approaches first we used IDF1 [4], and MOTA [29] which are evaluation metrics generally used for tracking. IDF1 is an identity-aware F1 score that evaluates how consistently a tracker maintains the correct identities over time. It differs from a standard F1 score by focusing on identity preservation, not just detection accuracy. MOTA is a global metric that penalizes missed detections, false positives, and identity switches to assess overall tracking performance. However, generally, when tracking is applied to animals, the objective is to perform analysis related to each animal, so it is important to also evaluate each model on its ability to give the exact identity of each animal. For this reason, the identification accuracy, recall and F1 score of each approach were also evaluated [33,28,18].

4. Results and discussion

4.1. Comparison of supervised detection based MAT and MOT approaches

In this section, we compare the performance of MAT and MOT approaches that are supervised for detections. Among the selected methods that have a supervised detection, we benchmarked DeepLabCut, DeepSORT, ByteTrack, and Cross-Input Consistency. We noticed that the performance of DeepLabCut depends on the number of keypoints used so we tested 3 configurations: DeepLabCut with 2 keypoints (neck and tail), 3 keypoints (2 ears and tail), and 5 keypoints (2 ears, tail, neck, and nose). The F1 score, recall, accuracy, MOTA, IDF1 of those approaches tested on our 10-minute video are reported in Table 3. We also estimated those performance per frame over the entire video to see how the performance evolves over time. Our findings reveal a significant challenge for all tested methods in maintaining performance over long-term tracking scenarios, as evidenced by a decline in F1 scores over time (Fig. 2).

Table 3

Average tracking performance of the different benchmarked approaches over the entire 10-minute video. The color in the different cells goes from white (lower performance) to green (higher performance). The best approaches for a given metric are presented in bold.

Method	IDF1	MOTA	F1 score	Recall	Accuracy
ByteTrack	0.79	0.73	0.59	0.58	0.60
DeepSORT	0.74	0.72	0.50	0.50	0.51
Cross-Input Consistency	0.31	0.66	0.14	0.09	0.29
DeepLabCut 2 keypoints	0.32	0.56	0.13	0.13	0.13
DeepLabCut 3 keypoints	0.26	0.53	0.14	0.14	0.14
DeepLabCut 5 keypoints	0.52	0.84	0.20	0.20	0.20
idTracker	0.16	0.32	0.10	0.09	0.10
PromptTrack	0.66	0.48	0.76	0.74	0.79
Track-Anything	0.70	0.45	0.79	0.76	0.82

Table 4

Percentage of pigs with at least one matching keypoint with a distance greater than 5, 10, or 100 pixels over the entire DeepLabCut tracking in the 10-min video.

Methods	5 px	10 px	100 px
2 keypoints	42.35%	11.17%	0.33%
3 keypoints	50.44%	4.63%	0.41%
5 keypoints	63.07%	7.58%	0.82%

The results presented in Table 3 show that DeepLabCut performs best when using 5 keypoints, both in overall tracking metrics and in its performance over time (Table 2). Despite this, and except for MOTA, DeepLabCut's results with 5 keypoints remain lower than those of most supervised MOT approaches. This is notable considering that DeepLabCut requires lower-level annotations (keypoints instead of bounding boxes), which are more time-consuming to obtain before the tool can be used.

When analyzing identity switches over time, DeepLabCut shows a consistently higher number of switches compared to other approaches. Upon investigating this behavior, we found that DeepLabCut occasionally reconstructs skeletons by mistakenly incorporating keypoints from neighboring animals (Fig. 3, Table 4). For instance, when a keypoint is missing from a given animal, DeepLabCut may substitute it with a keypoint from a nearby pig. Given that the average pig in our videos measures about 100 pixels in length, we identified that approximately 0.82% of pigs had at least one keypoint located around 100 pixels away, likely originating from a different pig. From this, we estimated that on average, a pig has one misassigned keypoint every 9 frames.

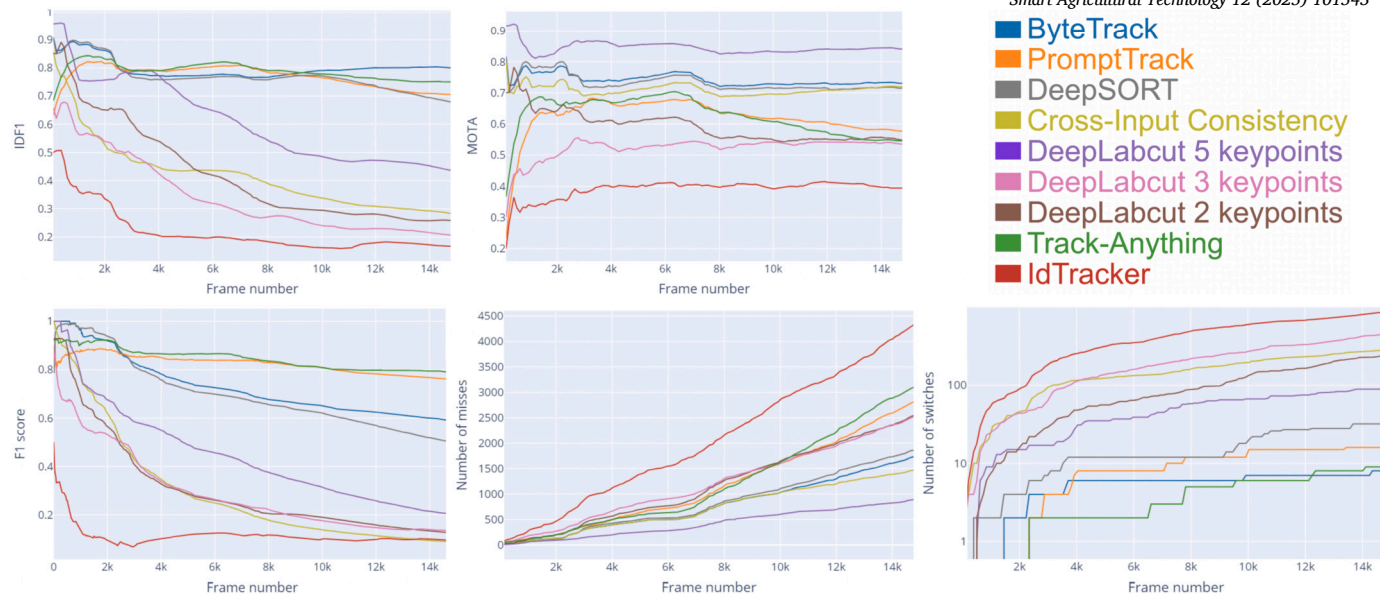


Fig. 2. Tracking performance (IDF1, MOTA, F1 score, number of misses, and number of identity switches) of the different benchmarked approaches over time in function of the number of processed frames on our long-term 10 minutes video at 25 FPS.

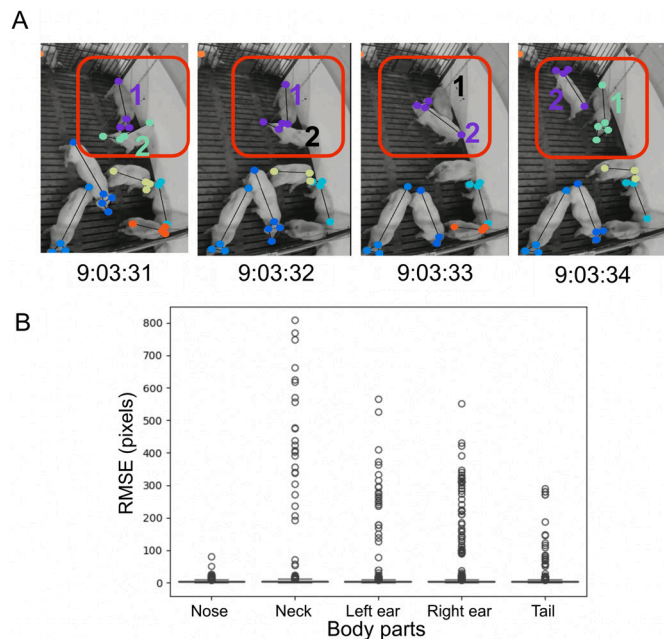


Fig. 3. Illustration of DeepLabCut identity switch errors and corresponding keypoint RMSE distributions. A) Reconstruction of the skeleton of the purple pig by DeepLabCut using keypoints from another pig, which led to an identity switch. B) RMSE (or distance between the ground truth point and the detected point by DeepLabCut) in function of the different body parts. A pig is roughly 100 px in our video.

While its susceptibility to keypoint misassignments results in many identity switches, DeepLabCut’s use of multiple keypoints provides robustness against missed detections of entire animals, leading to its relatively high MOTA (Fig. 2).

Looking at the supervised methods in Fig. 2, Cross-Input Consistency exhibited the lowest performance overall, primarily due to a high number of identity switches. Among the approaches relying on supervised detection, ByteTrack appears to be the most promising. It maintained relatively stable IDF1 and MOTA scores over time and achieves average performance in terms of F1 score, number of misses, and identity

Table 5

Detection performance with IoU > 0. Note that all methods rely on the same detector, and the values correspond to detection metrics after the tracking phase of each method.

Methods	F1 score	Recall	Accuracy
ByteTrack	0.97	0.96	0.99
DeepSORT	0.97	0.95	0.98
Cross-Input Consistency	0.98	0.98	0.98

switches. As an enhanced version of DeepSORT, ByteTrack shows similar performance, with a slight overall advantage.

Across all supervised tracking methods, we observed difficulties in maintaining consistent identities over extended periods, as indicated by declining F1 scores over time (Fig. 2). This degradation is primarily driven by an increasing number of identity switches and missed detections. Unfortunately, once an identity switch occurs, these methods are unable to recover or correct it.

4.2. Comparison of unsupervised detection based MAT and MOT approaches

We benchmarked idTracker, Track-Anything, and PromptTrack for the unsupervised detection category of trackers. None of these approaches required pre-training a model to detect animals in the videos. Their tracking performance is summarized in Table 3 and Fig. 2.

As shown in Table 3, Track-Anything and PromptTrack achieve performance comparable to ByteTrack, which yields the best results among supervised methods. In particular, Track-Anything performs well not only due to its detection capabilities but also because it uses SAM to segment each animal, which it then leverages through XMem to ensure accurate tracking.

As illustrated in Fig. 2, both Track-Anything and PromptTrack outperformed idTracker, which had the lowest overall performance among all evaluated approaches.

We observed that the main limitation of most unsupervised approaches lies in the detection stage, which negatively impacts overall tracking performance. Detection performance was evaluated based on each method’s ability to identify ground truth boxes (recall) and its precision in each detection (precision). These results are presented in Table 6.



Fig. 4. Visual comparison of detection performance between idTracker (left), PromptTrack (center), and Track-Anything (right). (left) idTracker snapshot showing several undetected animals (marked with red points). (center and right) Track-Anything and PromptTrack snapshots showing all animals detected.

Table 6

Detection performance with $IOU > 0$.

Methods	F1 score	Recall	Accuracy
idTracker	0.86	0.77	0.97
PromptTrack	0.96	0.94	0.99
Track-Anything	0.95	0.92	0.99

As shown in Table 6, the detection performance of unsupervised methods is generally lower than that of supervised approaches (mAP of 95% on the validation set), particularly in terms of recall. However, Track-Anything and PromptTrack outperform idTracker in both recall and accuracy. These superior detection capabilities improved their tracking performance, as illustrated in Fig. 4.

Currently, Track-Anything and PromptTrack do not offer standalone graphical user interfaces, which may pose a challenge for users outside the computer science field. However, both tools are relatively easy to install and use, and they benefit from active GitHub communities. In Track-Anything, the user is required to manually click on each animal in the first frame of the video. In contrast, PromptTrack only requires a text prompt, which can be especially useful for scaling to multiple videos without the need for manual annotation in each one.

4.3. Comparison with other benchmarks

Our results can be compared with other benchmarking efforts. Su et al. [21] evaluated deep learning-based trackers on black cattle, reporting that while modern MOT methods achieved high short-term accuracy, identity consistency degraded under occlusions and dense grouping. This is consistent with our findings: supervised MOT approaches such as ByteTrack performed strongly on detection but accumulated identity switches over our longer 10-minute pig video. Panadeiro et al. [25] reviewed 28 open-source animal-tracking tools and emphasized that most legacy MAT software (e.g. idTracker, ToxTrac) suffered from limited robustness and struggled in crowded or long-duration scenarios. Our benchmarks confirm this limitation, with idTracker yielding the lowest recall and F1 scores among all tested methods. Wurtz et al. [37] specifically tested open-source MAT programs on pigs and concluded that they were not sufficiently reliable for commercial farm conditions, with tracking accuracies rarely exceeding the mid-80% range. Our evaluation is in line with these results for idTracker, but also shows that newer prompt- and segmentation-based methods (PromptTrack and Track-Anything) improve both detection ($F1 \geq 0.95$) and tracking precision, approaching the performance of supervised MOT. Finally, the AnimalTrack benchmark [39] highlighted that applying standard MOT pipelines to wildlife sequences led to substantial performance degradation compared to human datasets, particularly in long-term identity preservation. Our results support this observation: supervised MOT methods still work better than traditional MAT tools, but they lose track of identities over time. We also show that newer foundation-based unsupervised methods can reduce this problem and reach accuracy close to ByteTrack without any pre-training.

5. Conclusion

In this study, we benchmarked several SOTA and legacy MAT and MOT approaches using the same long-term 10-minute video of pigs in a livestock setting. Our results show that MOT approaches consistently outperform MAT methods in terms of tracking performance. They also reveal a significant challenge for all tested methods in maintaining performance over long-term tracking scenarios, as evidenced by a decline in F1 scores over time.

Among the supervised detection-based methods, object detection was generally robust, achieving high accuracy. However, tracking performance declined over time due to frequent identity switches. Of these methods, ByteTrack provided the most balanced and reliable results. DeepLabCut, which reconstructs animal identities from detected keypoints rather than bounding boxes, struggled with detection consistency and was more prone to errors during tracking.

For unsupervised methods, detection posed a greater challenge. However, Track-Anything and PromptTrack successfully addressed this limitation by integrating SAM and OWLv2, respectively. Track-Anything, in particular, benefited from segmentation-level detection, resulting in superior tracking precision. Despite being unsupervised, both methods achieved performance comparable to ByteTrack. Additionally, they require no pre-trained detection models: Track-Anything allows users to manually select seed animals in the first frame via a simple GUI, while PromptTrack uses natural language prompts for initialization, an approach that scales well across multiple videos.

One key advantage of traditional MAT tools like idTracker is their user-friendly, standalone graphical interface, which makes them more accessible to users without a technical background. While Track-Anything and PromptTrack currently lack some of the tracking utilities found in other MAT tools, they nonetheless offer a fast, relatively accurate, and intuitive tracking solution. An important direction for future development would be the creation of standalone GUI-based versions for these new tools, enhancing their usability and adoption in non-technical research settings, and ultimately contributing to more accurate and reliable automated livestock tracking that supports improved animal welfare and productivity.

CRediT authorship contribution statement

Anne Marthe Sophie Ngo Bibinbe: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Patrick Gagnon:** Writing – review & editing, Investigation, Funding acquisition. **Jamie Ahloy-Dallaire:** Writing – review & editing, Visualization, Validation, Supervision, Investigation, Formal analysis. **Eric R. Paquet:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used chatGPT in order to improve language and readability. After using this service, the

authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Eric Paquet reports financial support was provided by MAPAQ Innov'action (IA120640). Patrick Gagnon reports financial support was provided by MAPAQ Innov'action (IA120595). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors want to acknowledge the financial support from MAPAQ Innov'action (IA120640 and IA120595), FRQNT for the PhD scholarship to A.M.S.N.B and all fruitful discussions with members of the Paquet lab.

Data availability

Data will be made available on request. A GitHub repository is available for the scripts used in this work: <https://github.com/ngobibinbe/Tracking-Benchmark>.

References

- [1] Precision livestock farming market size, share, opportunities & forecast, 2024.
- [2] Nir Aharon, Roy Orfaig, Ben-Zion Bobrovsky, Bot-sort: robust associations multi-pedestrian tracking, arXiv preprint arXiv:2206.14651, 2022.
- [3] Favven Bastani, Songtao He, Sam Madden, Self-supervised multi-object tracking with cross-input consistency, in: Proceedings of the 35th International Conference on Neural Information Processing Systems, NIPS '21, Red Hook, NY, USA, Curran Associates Inc., 2024.
- [4] Keni Bernardin, Rainer Stiefelhagen, Evaluating multiple object tracking performance: the clear mot metrics, EURASIP J. Image Video Process. 2008 (1) (5.2008) 1–10.
- [5] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, Ben Upcroft, Simple online and realtime tracking, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 3464–3468.
- [6] Kristin Branson, Alice A. Robie, John Bender, Pietro Perona, Michael H. Dickinson, High-throughput ethomics in large groups of drosophila, Nat. Methods 6 (6) (2009) 451–457.
- [7] Ho Kei Cheng, Seoung Wug Oh, Brian Price, Alexander Schwing, Joon-Young Lee, Tracking anything with decoupled video segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 1316–1326.
- [8] Ho Kei Cheng, Alexander G. Schwing, Xmem: long-term video object segmentation with an Atkinson-Shiffrin memory model, in: European Conference on Computer Vision, Springer, 2022, pp. 640–658.
- [9] Peng Chu, Jiang Wang, Quanzeng You, Haibin Ling, Zicheng Liu, Transmot: spatial-temporal graph transformer for multiple object tracking, in: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023, pp. 4859–4869.
- [10] Yunhao Du, Zhicheng Zhao, Yang Song, Yanyun Zhao, Fei Su, Tao Gong, Hongying Meng, Strongsort: make deepsort great again, Trans. Multimed. 25 (January 2023) 8725–8737.
- [11] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, Jian Sun, Yolox: Exceeding yolo series in 2021, arXiv, 5:12, 7.2021.
- [12] Aishwarya Kamath, Mannat Singh, Yann LeCun, Gabriel Synnaeve, Ishan Misra, Nicolas Carion, Mdetr-modulated detection for end-to-end multi-modal understanding, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1780–1790.
- [13] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, et al., Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 4015–4026.
- [14] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, et al., Segment anything, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 4015–4026.
- [15] H.W. Kuhn, The Hungarian method for the assignment problem, Nav. Res. Logist. Q. 2 (3.1955) 83–97.
- [16] Jessy Lauer, Mu Zhou, Shaokai Ye, William Menegas, Steffen Schneider, Tanmay Nath, Mohammed Mostafizur Rahman, Valentina Di Santo, Daniel Soberanes, Guoping Feng, et al., Multi-animal pose estimation, identification and tracking with deeplabcut, Nat. Methods 19 (4) (2022) 496–504.
- [17] Wenhao Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, Tae-Kyun Kim, Multiple object tracking: a literature review, Artif. Intell. 293 (April 2021) 103448.
- [18] S. Divya Meena, Loganathan Agilandeeswari, Smart animal detection and counting framework for monitoring livestock in an autonomous unmanned ground vehicle using restricted supervised learning and image fusion, Neural Process. Lett. 53 (2) (2021) 1253–1285.
- [19] Matthias Minderer, Alexey Gritsenko, Neil Houlsby, Scaling open-vocabulary object detection, Adv. Neural Inf. Process. Syst. 36 (2023) 72983–73007.
- [20] Hauke Jürgen Mönck, Andreas Jörg, Tobias von Falkenhausen, Julian Tanke, Benjamin Wild, David Dormagen, Jonas Piotrowski, Claudia Winklmayr, David Bierbach, Tim Landgraf, Biotracker: an open-source computer vision framework for visual animal tracking, arXiv preprint arXiv:1803.07985, 2018.
- [21] Su Myat Noe, Thi Thi Zin, Pyke Tin, Ikuo Kobayashi, Comparing state-of-the-art deep learning algorithms for the automated detection and tracking of black cattle, Sensors 23 (1) (2023).
- [22] Anne Marthe Sophie Ngo Bibinbe, Jamie Ahloy-Dallaire, Eric Paquet, Prompttrack: Demo library, 2024. (Accessed 30 March 2025).
- [23] Anne Marthe Sophie Ngo Bibinbe, Patrick Gagnon, Jamie Dallaire-Ahloy, Eric R. Paquet, An HMM-based framework for identity-aware long-term MOT from sparse and uncertain identification: use case on long-term tracking in livestock, arXiv:2509.09962, 2025.
- [24] Pha Nguyen, Kha Gia Quach, Kris Kitani, Khoa Luu, Type-to-track: retrieve any object via prompt-based tracking, Adv. Neural Inf. Process. Syst. 36 (2024).
- [25] Veronica Panadeiro, Alvaro Rodriguez, Jason Henry, Donald Wlodkowic, Magnus Andersson, A review of 28 free animal-tracking software applications: current features and limitations, Nat. Cell Biol. 50 (9) (2021) 246–254.
- [26] Talmo D. Pereira, Nathan Tabris, Arie Matsliah, David M. Turner, Jingfan Li, Shruthi Ravindranath, Eli Papadoyannis, Erik Normand, David S. Deutsch, Z. Jonathan Wang, et al., Sleep: a deep learning system for multi-animal pose tracking, Nat. Methods 19 (4) (2022) 486–495.
- [27] Alfonso Pérez-Escudero, Javier Vicente-Page, Robert C. Hinz, Sara Arganda, Gonzalo G. de Polavieja, idtracker: tracking individuals in a group by automatic identification of unmarked animals, Nat. Methods 11 (7) (2014) 743–748.
- [28] Yongliang Qiao, Daobilige Su, He Kong, Salah Sukkarieh, Sabrina Lomax, Cameron Clark, Bilstm-based individual cattle identification for automated precision livestock farming, in: 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), 2020, pp. 967–972.
- [29] Ergys Ristani, Francesco Solera, Roger S. Zou, Rita Cucchiara, Carlo Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: ECCV Workshops, 2016.
- [30] Alexander Rodriguez, Hongzhen Zhang, Jonatan Klaminder, Tomas Brodin, Per L. Andersson, Mats Andersson, Toxtrac: a fast and robust software for tracking organisms, Methods Ecol. Evol. 9 (3) (2018) 460–464.
- [31] Juliette Schillings, Richard Bennett, David Christian Rose, Exploring the potential of precision livestock farming technologies to help address farm animal welfare, Front. Anim. Sci. (5.2021) 13.
- [32] Vukasin D. Stanojevic, Branimir T. Todorovic, Boosttrack: boosting the similarity measure and detection confidence for improved multiple object tracking, Mach. Vis. Appl. 35 (3) (2024) 1–15.
- [33] Eric T. Psota, Ty Schmidt, Benny Mote, Lance C. Pérez, Long-term tracking of group-housed livestock using keypoint detection and map estimation for individual animal identification, Sensors 20 (13) (2020).
- [34] Carl Vondrick, Abhinav Shrivastava, Alireza Fathi, Sergio Guadarrama, Kevin Murphy, Tracking emerges by colorizing videos, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 391–408.
- [35] Greg Welch, Gary Bishop, An introduction to the Kalman filter, Technical report, USA, 1995.
- [36] Nicolai Wojke, Alex Bewley, Dietrich Paulus, Simple online and realtime tracking with a deep association metric, in: Proceedings - International Conference on Image Processing, ICIP, 2017-September, 2.2018, pp. 3645–3649.
- [37] Kaitlin Wurtz, Tomas Norton, Janice Siegford, Juan Steibel, Assessment of open-source programs for automated tracking of individual pigs within a group, in: Practical Precision Livestock Farming, Wageningen Academic, 2022, pp. 213–230.
- [38] Jinyu Yang, Mingqi Gao, Zhe Li, Shang Gao, Fangxing Wang, Feng Zheng, Track anything: segment anything meets videos, arXiv preprint arXiv:2304.11968, 2023.
- [39] Libo Zhang, Junyuan Gao, Zhen Xiao, Heng Fan, Animaltrack: a benchmark for multi-animal tracking in the wild, Int. J. Comput. Vis. 131 (2) (2023) 496–513.
- [40] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, Xinggang Wang, Bytetrack: multi-object tracking by associating every detection box, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13682, 10.2021, pp. 1–21.
- [41] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, Wenyu Liu, Fairmot: on the fairness of detection and re-identification in multiple object tracking, Int. J. Comput. Vis. 129 (2021) 3069–3087.