OinkTrack: An Ultra-Long-Term Dataset for Multi-Object Tracking and Re-Identification of Group-Housed Pigs

Feng-Kai Huang
National Taiwan University
Taipei, Taiwan
leonelhuang@cmlab.csie.ntu.edu.tw

Hong-Wei Xu
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
hw413504009.ee13@nycu.edu.tw

Chu-Chuan Lee
Chung Yuan Christian University
Taoyuan, Taiwan
g11378032@cycu.edu.tw

Hong-Yi Tu
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
hongyi.ee11@nycu.edu.tw

Hong-Han Shuai*
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
hhshuai@nycu.edu.tw

Wen-Huang Cheng
National Taiwan University
Taipei, Taiwan
wenhuang@csie.ntu.edu.tw

Abstract

Long-term multi-animal tracking in densely group-housed agricultural settings is critical for automated behavior monitoring and early anomaly detection in precision livestock farming. However, it poses significant challenges due to persistent occlusions from feeders and water dispensers, high inter-individual appearance similarity, and drastic visual changes across day and night cycles. Existing multi-object tracking datasets rarely capture the combined difficulty of these real-world conditions. To address this, we introduce *OinkTrack*, a large-scale benchmark for continuous multi-pig tracking in commercial farm environments. The dataset comprises over five hours of annotated video across sixteen sequences, covering day, night, night-to-day, and day-to-night transitions. Each sequence ranges from one minute to one hour, featuring an average of thirty-six pigs per frame. In total, *OinkTrack* provides 573,700 bounding boxes linked to 574 consistent pig identities. It enables detailed behavior analysis under varying lighting and crowding conditions. We describe the data collection and annotation process, present statistical insights into tracking difficulty, and benchmark 11 state-of-the-art tracking methods. *OinkTrack* provides a robust foundation for developing long-term tracking models and supports downstream applications such as individual activity profiling and early detection of abnormal behavior in real-world, high-density animal populations. The complete dataset and supplementary materials are publicly accessible at https://leohuang0511.github.io/oinktrack-page.

CCS Concepts

• **Computing methodologies** → **Tracking**.

*Corresponding author, hhshuai@nycu.edu.tw

MM '25, Dublin, Ireland

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3758189

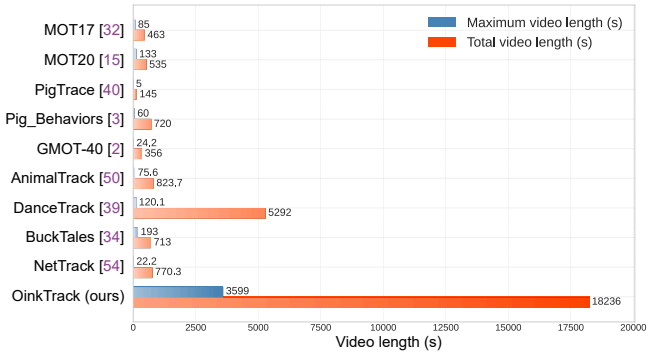

Figure 1: Comparison of maximum and total video lengths between *OinkTrack* and other representative MOT datasets [2, 3, 15, 32, 34, 39, 40, 50, 54].

Keywords

Multi-object tracking; Long-term tracking; Animal tracking; Benchmark dataset; Precision agriculture

ACM Reference Format:
Feng-Kai Huang, Hong-Wei Xu, Chu-Chuan Lee, Hong-Yi Tu, Hong-Han Shuai, and Wen-Huang Cheng. 2025. OinkTrack: An Ultra-Long-Term Dataset for Multi-Object Tracking and Re-Identification of Group-Housed Pigs. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3746027.3758189

1 Introduction

Automated animal monitoring has become a key component of modern precision agriculture, offering crucial benefits in animal welfare, health diagnostics, and operational efficiency [7, 30, 35]. Among the enabling technologies, Multi-Object Tracking (MOT) plays a central role in enabling fine-grained, identity-preserving behavior analysis. However, achieving accurate long-term tracking in real-world agricultural settings remains a formidable challenge. In group-housed environments such as pig farms, tracking systems must contend with severe inter-individual appearance similarity [50], frequent occlusions caused by dense interactions, and infrastructure such as

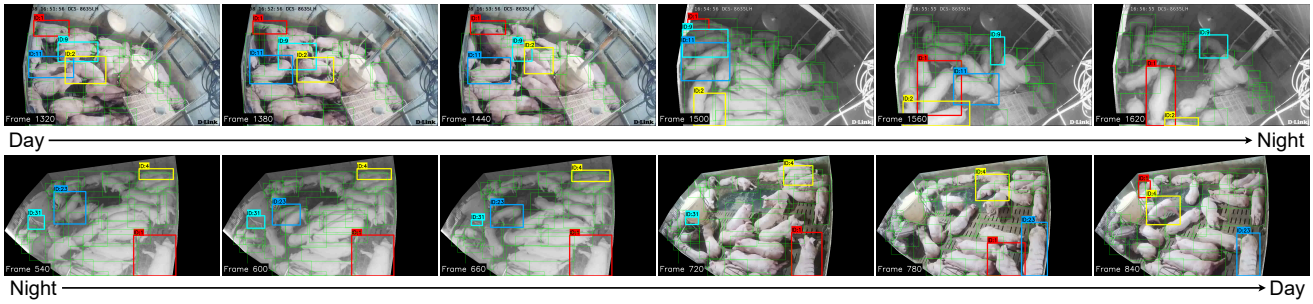

Figure 2: Annotated *OinkTrack* sequences showcasing day-night transitions. Four pigs are highlighted for ID consistency, with all other individuals also tracked. These scenes exemplify the core challenges of *OinkTrack*: dense pig groups, complex interaction, diverse motion, and frequent occlusion/reappearance events.

feeders and water dispensers [40], and complex social dynamics and motion patterns [54]. These challenges are further intensified by the need for continuous tracking across full diurnal cycles, during which dramatic changes in illumination significantly degrade the reliability of appearance-based models [28].

Despite recent advances in MOT algorithms [18, 19, 48, 51, 55], most existing benchmarks are short in duration and focused on human-centric scenarios [12, 15, 32, 39]. Datasets involving animals are emerging [34, 40, 50, 54], but they typically consist of short clips, limited numbers of individuals, or restricted environmental diversity. While long-duration datasets such as LaSOT [17] have been proposed, they target single-object tracking and do not capture the multi-agent interactions or identity-switch challenges present in densely housed animal settings. Although collecting farm video footage is relatively easy, large-scale annotation under persistent identity supervision remains a major bottleneck, particularly in dynamic, high-density environments like commercial pig farms.

To fill this gap, we introduce *OinkTrack*, a new large-scale benchmark designed to advance research in Long-Term Multi-Animal Tracking (LTMAT) under real-world farming conditions. *OinkTrack* contains 16 continuous video sequences totaling over 5 hours of annotated footage. These sequences range from one minute to one hour, with an average length of over 1,100 seconds and up to 35.88 pigs per frame. The dataset spans diverse lighting conditions, including daytime, nighttime, and full day-to-night transitions. In total, it includes 573,700 bounding boxes and 574 identity-consistent trajectories, each annotated with high-quality 2D bounding boxes and persistent IDs. As illustrated in Fig. 1, *OinkTrack* significantly outperforms existing MOT datasets in both maximum and total video length, establishing a new benchmark standard for long-duration tracking in complex, high-density environments.

We describe the data collection and annotation process, provide statistical analysis of the dataset's unique properties and benchmark 11 state-of-the-art MOT models. Our experiments reveal that *OinkTrack* presents significant challenges in identity preservation under visual ambiguity, handling long-term occlusions, and maintaining robust tracking through drastic illumination changes. As such, *OinkTrack* sets a new standard for evaluating long-term tracking performance and provides a critical resource for both multimedia and agricultural research communities.

The main contribution of this work is three-fold. 1) We introduce *OinkTrack*, the first benchmark tailored for long-term multi-animal tracking in commercial pig farming environments. 2) The dataset enables fine-grained behavior analysis with over 500K annotations across identity-consistent trajectories. 3) We benchmark state-of-the-art tracking algorithms, revealing key challenges in identity preservation, occlusion handling, and robustness to visual changes.

2 Related Work

2.1 Multi-animal Tracking

Recent years have seen substantial advances in Multi-Object Tracking (MOT), including tracking-by-detection [1, 20, 29, 51, 52], joint detection and tracking [4, 41, 44, 56], and propagation-based methods [31, 37, 49, 53]. These methods are primarily evaluated on short-term, human-centric datasets such as MOT16 [32], MOT20 [15], DanceTrack [39], and SportsMOT [12]. Although some general-purpose MOT datasets [2, 14] include animal categories, they lack the scale needed for detailed behavior analysis. In response, several animal-specific datasets have emerged, targeting natural environments [34, 50, 54] or within specific contexts [3, 24, 38, 40]. However, most of these datasets contain short sequences (typically seconds to a few minutes), which limits their applicability to long-term behavior and health monitoring in livestock environments.

2.2 Long-term Tracking

Long-term tracking focuses on maintaining object identities over extended periods despite occlusions, re-entries, and drastic appearance changes. While prior work has proposed dedicated methods [13, 25, 28, 45, 47] and single-object benchmarks [17, 23, 33, 42], these efforts largely overlook multi-object and densely populated scenarios. Although some MOT models support longer sequences [8, 10, 21, 26, 34], datasets that capture the compounded challenges of tracking group-housed livestock over long durations remain scarce [3, 38, 40]. To address this, we propose *OinkTrack*, a benchmark designed for long-term multi-animal tracking in commercial pig farms. With sequences up to one hour and consistent identities across full diurnal cycles, it introduces significant challenges in identity preservation, occlusion handling, and visual robustness—advancing research in both MOT and agricultural AI.

Table 1: Comparison of *OinkTrack* with popular MOT benchmarks. *OinkTrack* stands out for its long video durations, dense annotations, full day–night transitions, and inclusion of synchronized audio—features rarely combined in existing MOT datasets. "-" represents values that are unavailable.

Benchmark	Videos	Classes	Avg. len. (min)	Max. len. (min)	Total len. (min)	Avg. tracks	Total tracks	Frame rate	Anno. FPS	Total boxes	Total frames	Day-night transition	Audio
KITTI [22]	50	5	0.17	-	8.30	52.00	2600	30	10	80K	14K	✗	✗
MOT17 [32]	14	1	0.55	1.42	7.72	95.07	1331	30	30	300K	11K	✗	✗
MOT20 [15]	8	1	1.11	2.22	8.92	479.12	3833	25	25	2.1M	13K	✗	✗
TAO [14]	2907	833	0.61	-	1782.96	5.90	17287	30	1	333K	2.6M	✗	✗
PigTrace [40]	29	1	0.08	0.08	2.42	28.31	821	3, 6	3, 6	1.5K	0.5K	✗	✗
Pig_Behaviors [3]	12	1	1.00	1.00	12.00	8.00	96	10	10	1.2K	7.2K	✗	✗
GMOT-40 [2]	40	10	0.15	0.40	5.93	50.65	2026	24-30	24-30	256K	9.6K	✗	✗
AnimalTrack [50]	58	10	0.24	1.26	13.73	33.00	1927	30	30	429K	24K	✗	✗
DanceTrack [39]	100	1	0.88	2.00	88.2	9.00	990	20	20	877K	105K	✗	✗
SportsMOT [12]	240	3	0.42	-	100.25	14.17	3401	25	25	1.6M	150K	✗	✗
BuckTales [34]	12	2	0.99	3.22	11.88	62.17	746	30	30	1.2M	21K	✗	✗
NetTrack [54]	106	22	0.12	0.37	12.84	6.30	668	25	25	85K	19K	✗	✗
OinkTrack (ours)	16	1	19.00	60.00	303.93	35.88	574	15	1	573K	273K	✓	✓

3 OinkTrack

To advance research in long-term multi-animal tracking (LTMAT), we present *OinkTrack*, a novel dataset featuring real-world video sequences with dense identity annotations for group-housed pigs. Designed to capture the complex dynamics of commercial livestock environments, *OinkTrack* enables benchmarking of tracking algorithms under extreme real-world conditions, including persistent occlusions, high inter-object similarity, and drastic illumination changes. This section details our data collection process, annotation methodology, and dataset characteristics.

3.1 Data Collection

The video data in *OinkTrack* were collected from a systematic commercial pig farm, where each pen measures 3.10m × 4.85m and houses between 32 and 40 pigs (mean ≈ 36). Each pen includes a central feeder and side-mounted water dispensers, which frequently cause significant occlusions. Recordings were made from two pens using RGB cameras mounted at elevated corners above the pens, angled downward at approximately -45° to provide a comprehensive top-down view while minimizing occlusions. The videos were recorded at a resolution of 1280×720 pixels and a frame rate of 15 FPS. The cameras were also equipped with a removable IR-cut filter (ICR) to support low-light capture during nighttime, and audio was recorded in parallel to support future multimodal behavior analysis and acoustic event detection.

To ensure robust evaluation across real-world variability, data collection followed four core principles: (1) inclusion of ultra-long-term sequences, (2) dense and consistent trajectory annotations, (3) coverage of diverse motion and interaction patterns, and (4) capture of full diurnal cycles. All recordings were conducted under standard farming conditions, without introducing external stimuli or stress to the animals. Specifically, continuous video was recorded over a 60-day period (November to December 2024). To ensure tracking remains grounded in observable behavior rather than external disruptions, we retained over 48 hours of footage with clear, unobstructed views—excluding only segments with prolonged human interference. From this filtered pool, 16 long-duration clips were selected based on activity level, behavioral diversity, and visual clarity. These clips range from 1 minute to 1 hour in length and capture complex group behaviors such as resting, playing, feeding, and overlapping. The selected clips include 57 minutes of daytime footage, 97 minutes of nighttime footage, and 150 minutes of transition periods (day-to-night and night-to-day), ensuring comprehensive environmental coverage for robust model evaluation.

3.2 Annotation

Each visible pig in the *OinkTrack* dataset is annotated with a 2D bounding box and a unique, persistent identity, following the ID consistency protocol adopted in DanceTrack [39]. Annotations are performed using the Supervisely[1] platform, with its tools for bounding box tracking and sequence navigation. For a focused and consistent annotation scope, masks are applied to exclude regions outside the primary pen in view (*e.g.*, adjacent enclosures partially captured by the camera). Examples of annotated frames are shown in Fig. 2.

A team of trained annotators, guided by a domain expert (*i.e.*, a PhD researcher), followed strict guidelines to maintain annotation quality and identity consistency across long video sequences. One of the challenges lies in preserving identity across occlusions and re-entries. Annotators reviewed each clip both forward and backward, using contextual cues such as position, velocity, orientation, and neighboring interactions to re-identify individuals. A new ID is assigned only when consensus cannot be reached after multi-pass reviews. Each sequence underwent a two-stage quality control process: a full review by the initial annotator team followed by a secondary verification and correction by two senior annotators.

Nocturnal clips present additional difficulty due to low lighting. To address this, annotators adjusted video brightness, contrast, and gamma levels to improve visibility during the labeling process.

[1]Annotation tool available at https://supervisely.com.

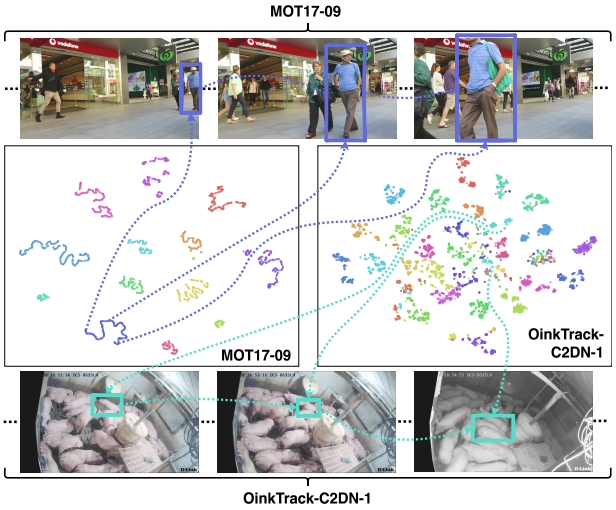

Figure 3: Visualization of re-ID features from sampled videos in MOT17 and *OinkTrack* using t-SNE [43]. Different colors are used to express different individual pigs. For better visualization, we only select 200 frames in each video sequence.

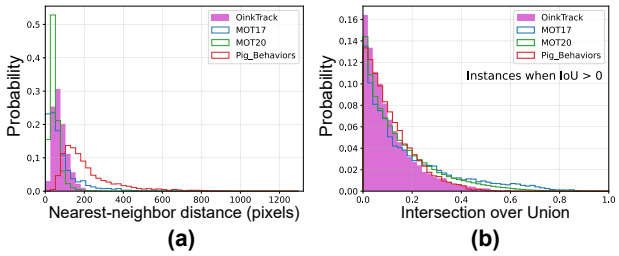

Figure 4: Spatial properties of *OinkTrack* compared to MOT17 [32], MOT20 [15], and Pig_Behaviors [3]: (a) nearest-neighbor distance, and (b) inter-object IoU.

Annotations were performed at 1 frame per second (FPS). While sparser than in some short-term datasets, this sampling rate effectively captures salient motion and identity continuity over long periods, yielding high-quality, persistent trajectories.

3.3 Dataset Analysis

Dataset Statistics. Tab. 1 compares *OinkTrack* with representative MOT datasets across domains such as pedestrian, animal, and general object tracking. *OinkTrack* stands out with an average video length of 19 minutes and a maximum of 60 minutes, with over 273K frames and 573K annotated bounding boxes across 574 identity-consistent trajectories. Unlike most existing datasets, *OinkTrack* includes synchronized audio and diverse lighting conditions, with 5 daytime, 7 nighttime, and 4 transition sequences covering more than 300 minutes in total. Its dense group-housing setup, averaging 35.88 pigs per sequence, presents challenges in occlusion, identity preservation, and visual similarity. These characteristics make *OinkTrack* a rare and comprehensive benchmark for long-term, multimodal, multi-animal tracking in real-world farm environments.

Visual Ambiguity and Re-identification Challenge. To illustrate the difficulty of distinguishing visually similar individuals in *OinkTrack*, we compare Re-ID feature embeddings from our dataset with those from the MOT17 pedestrian benchmark [32]. Separate Re-ID models are trained on each dataset, and the resulting features are visualized using t-SNE [43], as shown in Fig. 3. In MOT17, pedestrian identities form relatively distinct clusters, indicating effective visual separation. In contrast, pigs in *OinkTrack* produce highly overlapping feature embeddings, revealing significant visual ambiguity across individuals. This makes re-identification substantially more difficult than in human-centric tracking scenarios. The challenge is further intensified by appearance changes across varying lighting conditions, especially during day-to-night transitions.

Spatial Properties. We analyze *OinkTrack*'s crowding level and compare it with MOT17 [32], MOT20 [15], and Pig_Behaviors [3] using two metrics. First, we measure the distribution of minimum distance from each object to its nearest neighbor, as depicted in Fig. 4(a). The distances in *OinkTrack* are concentrated in the 0-200 pixel range, generally shorter than in MOT17 and Pig_Behaviors, indicating closer proximity among pigs. While MOT20 exhibits even shorter raw pixel distances, this characteristic is largely attributable to its typically distant viewpoints and smaller object scales. Second, we analyze the distribution of Intersection over Union (IoU) values between all distinct pairs of objects within each frame (Fig. 4(b)). In instances of overlap, *OinkTrack*'s IoU distribution is similar to that of Pig_Behaviors. Compared to MOT17 and MOT20, where objects are smaller and scenes are more crowded (often leading to larger IoUs for overlapping objects), *OinkTrack*'s IoUs for overlapping pigs are somewhat smaller. Despite this difference in magnitude, *OinkTrack*'s overall IoU distribution for overlapping pairs remains broadly comparable. These analyses collectively demonstrate *OinkTrack*'s higher spatial congestion and more frequent inter-individual interactions, thereby increasing the likelihood of ID switches and track fragmentation for MOT algorithms.

Temporal Properties. The ultra-long sequences in *OinkTrack* mean that pigs frequently disappear and later reappear due to prolonged occlusion; we term this phenomenon a "gap". We analyze the cumulative length of gaps per track in *OinkTrack* and compare it with MOT17, MOT20, and Pig_Behaviors in Fig. 5(a). Notably, the proportion of tracks with a gap in *OinkTrack* is higher than in the other datasets, while the sharp increase for Pig_Behaviors is attributable to frequent omissions in that dataset. Additionally, the longest gap in *OinkTrack* exceeds 1,400 frames, underscoring the challenge *OinkTrack* poses to long-term ID maintenance.

Furthermore, we investigate pig movement patterns. An analysis of total accumulated movement versus the number of frames tracked (Fig. 5(b)) reveals significant individual variation in activity. A strong correlation exists between tracking duration and distance traveled. Interestingly, total movement is generally greater in daytime than in nighttime scenes, consistent with the diurnal nature of pigs. To delve deeper, cumulative movement curves for individuals in representative daytime and nighttime sequences are shown in Fig. 5(c), respectively. These plots highlight substantial intra-scene variability in activity levels, with some pigs being significantly more active than others. The most active individuals during the day cover roughly twice the distance of their most active nighttime counterparts, and overall daytime movement is higher.

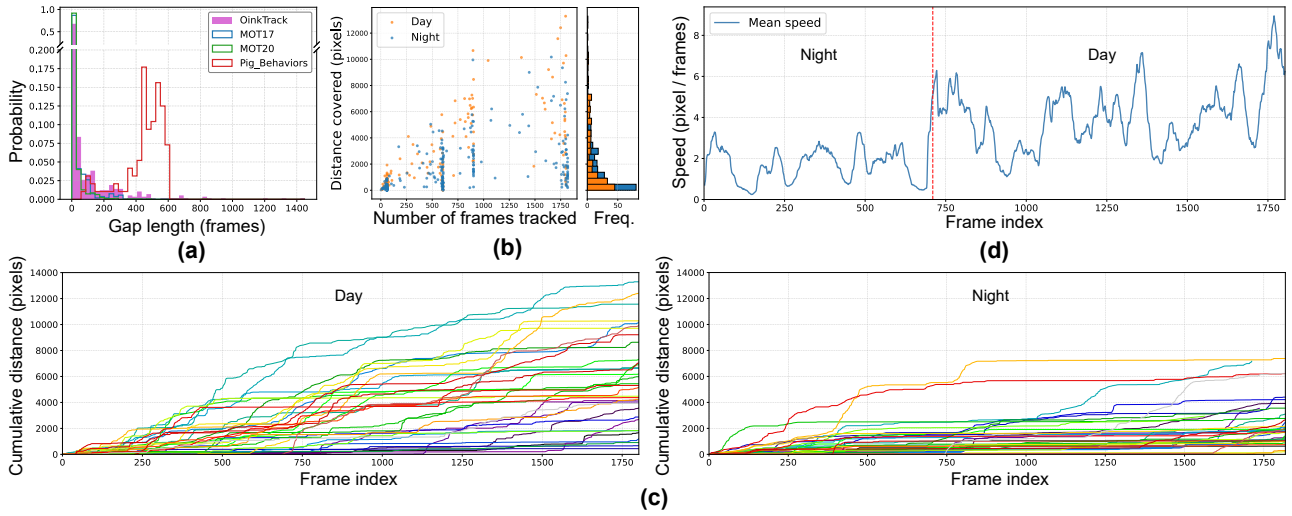

Figure 5: Temporal properties of the *OinkTrack* dataset: (a) gap durations, (b) distance covered by individuals, (c) curves of cumulative path lengths in a long sequence, and (d) mean group speed over time during a cross-day-and-night sequence.

Finally, we analyze the change in mean velocity for all individuals in a cross-day-and-night sequence (Fig. 5(d)). A significant increase in overall pig movement speed is observed as the scene transitions to daytime. Such pronounced differences in day-night activity, coupled with variations in visual appearance due to lighting changes, substantially elevate the difficulty of continuous tracking. In summary, our analysis of *OinkTrack*'s temporal characteristics reveals the challenges posed by its extreme video lengths, including the gap phenomenon, diverse individual activity levels, and the complexities introduced by day-night transitions.

3.4 Dataset Split and Availability

OinkTrack is made publicly available to encourage research and standardized benchmarking in LTMAT. We split the dataset into training, validation, and test sets in a 7:2:7 ratio, corresponding to total video lengths of 117 minutes, 45 minutes, and 142 minutes, respectively. This partition was stratified by key attributes, including illumination conditions (day, night, and transition) and pen identity (Pen 1 vs. Pen 2), to ensure a balanced distribution of characteristics across all three sets. The complete dataset, including frame images, detailed annotations in DanceTrack format, and relevant metadata, can be accessed and downloaded from our project website[2]. The data are released under the Creative Commons CC BY-NC-SA 4.0 license, permitting free use for academic and research purposes.

4 Experiments

4.1 Evaluation Metrics

For a comprehensive evaluation of tracking performance on *OinkTrack*, we adopt a suite of well-established metrics. Our primary metric is the Higher Order Tracking Accuracy (HOTA) [27], which effectively balances and assesses different aspects of tracking by decomposing into Detection Accuracy (DetA) and Association Accuracy (AssA). To specifically evaluate identity preservation, we utilize the IDF1 score [36]. Additionally, we report the widely-used Multi-Object Tracking Accuracy (MOTA) and the number of Identity Switches (IDsw), both components of the CLEAR MOT metrics [5], to ensure broader comparability with existing literature.

[2] *OinkTrack* project website: https://leohuang0511.github.io/oinktrack-page

4.2 Evaluated Trackers

To establish robust baselines and stimulate future research on *OinkTrack*, we comprehensively evaluate 11 state-of-the-art MOT algorithms. Our selection represents two categories: 1) tracking-by-detection approaches, which include SORT [6], DeepSORT [46], MOTDT [11], ByteTrack [51], OC-SORT [9], StrongSORT and its enhanced variant StrongSORT++ [16], and Hybrid-SORT [48]; and 2) recent transformer-based methods, such as MOTR [49], MeMOTR [21], and MOTIP [20]. This diverse selection facilitates a thorough assessment of how existing trackers handle the unique challenges of long-term tracking, crowded environments, and high appearance similarity that *OinkTrack* presents.

4.3 Benchmark Results

Overall Performance. The experimental results on *OinkTrack*, summarized in Tab. 2, reveal the profound challenges our dataset poses to state-of-the-art MOT algorithms. Transformer-based methods consistently demonstrate superior performance. Notably, MOTIP significantly outperforms all other evaluated trackers, as it achieves the highest scores across HOTA (43.7), MOTA (74.8), DetA (63.3), AssA (30.4), and IDF1 (47.8). This substantial performance gain underscores the inherent advantage of such models in modeling long-range spatio-temporal dependencies via attention mechanisms. In particular, the query-based method maintains a more persistent "memory" of each individual, which facilitates robust tracking throughout the challenging long sequences in our dataset. However, while MOTIP excels in accuracy, its high number of ID switches (7854) suggests its aggressive track management can compromise identity stability. In a compelling counterpoint, MeMOTR not only

Table 2: MOT algorithms benchmarking on *OinkTrack*.

Method	HOTA↑	MOTA↑	DetA↑	AssA↑	IDF1↑	IDsw↓
SORT [6]	30.4	53.9	46.4	20.3	31.7	2752
DeepSORT [46]	29.4	53.0	46.5	19.1	29.8	3289
MOTDT [11]	25.7	51.5	46.1	14.8	25.6	8381
ByteTrack [51]	30.1	55.0	46.9	19.8	32.4	2326
MOTR [49]	34.5	64.3	56.0	21.4	33.5	2436
OC-SORT [9]	29.8	52.3	44.9	20.0	31.7	2224
StrongSORT [16]	28.2	52.0	45.3	18.1	29.0	3946
StrongSORT++ [16]	28.1	47.2	44.5	18.4	28.5	3305
MeMOTR [21]	42.0	73.6	61.6	28.9	44.6	**1835**
Hybrid-SORT [48]	30.8	53.3	47.4	20.3	31.6	2933
MOTIP [20]	**43.7**	**74.8**	**63.3**	**30.4**	**47.8**	7854

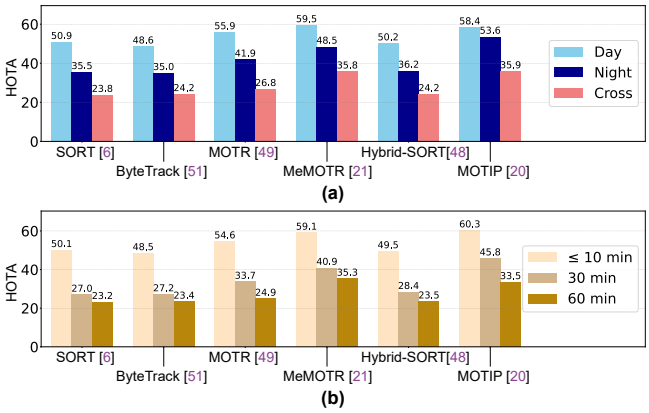

Figure 6: Scene analysis of different (a) illumination conditions and (b) video lengths.

rivals MOTIP in accuracy (42.0 HOTA) but also achieves the best identity persistence with 1835 ID switches.

Among the tracking-by-detection methods, those with sophisticated association logic exhibit specific strengths. For instance, ByteTrack's strategy of leveraging low-confidence detections yields the highest MOTA (55.0) within this category, while OC-SORT's observation-centric motion model delivers the second-best identity stability with 2224 ID switches. In summary, our benchmarks validate that *OinkTrack* effectively differentiates tracker capabilities, confirming the dominance of transformer-based approaches for LTMAT and revealing that high accuracy and state-of-the-art identity stability are not mutually exclusive. This pinpoints a clear direction for future work: optimizing the architectural trade-offs within these powerful end-to-end frameworks.

Scene Analysis. Our scene-specific analysis delves into tracker performance (HOTA) under key challenging conditions within *OinkTrack*: varying illumination and extended sequence durations. First, we evaluate performance on daytime, nighttime, and day-night-transition sequences from the test set in Fig. 6(a), revealing critical insights into the robustness of current MOT algorithms under these varied illuminations. Performance consistently degrades as conditions deviate from daytime scenarios. All methods exhibit a noticeable decline in HOTA scores during nighttime, with a more substantial drop in the particularly challenging cross-day-and-night sequences. This underscores the significant impact of illumination and associated appearance variability on tracking fidelity.

Second, we categorize the test set by sequence length into subsets of ≤10 minutes, 30 minutes, and 60 minutes, showing the profound challenge of maintaining tracking fidelity over extended periods. As depicted in Fig. 6(b), all methods exhibit a significant degradation in HOTA scores as video length increases, demonstrating the effects of error accumulation and re-identification difficulty in extreme long-term tracking. Notably, MeMOTR outperforms other methods in the 60-minute sequence, while it still represents a 40.3% decrease compared to its score on sequences ≤10 minutes. These findings highlight that extreme long-term tracking remains a formidable open problem, and *OinkTrack* effectively reveals the effectiveness of these approaches under such demanding temporal scales.

5 Potential Application

Continuous individual monitoring over extended periods, as facilitated by *OinkTrack*, allows for the derivation of critical health and welfare indicators from activity levels, feeding patterns, or social engagement; these indicators potentially signal early illness or distress. Research leveraging *OinkTrack* extends beyond algorithmic advancements to encompass the integration of these technologies into comprehensive farm management solutions. This integration includes the development of user interfaces that translate tracking data into actionable insights for farmers and veterinarians. Furthermore, coupling these systems with cloud-based analytics enables large-scale, continuous monitoring, which in turn offers unprecedented opportunities to enhance livestock health and elevate global animal welfare standards. The dataset's multimodal nature, which includes audio recordings, further invites exploration into richer, multi-sensory animal behavior understanding.

6 Conclusion and Future Work

OinkTrack is the first benchmark for Long-Term Multi-Animal Tracking (LTMAT) in challenging agricultural environments, offering hour-long, identity-consistent annotated trajectories of group-housed pigs from commercial farms. We introduce *OinkTrack*'s characteristics: extreme appearance similarity, dense crowding, diverse pig motion, and extended-duration tracking. Furthermore, we benchmarked 11 state-of-the-art MOT trackers, assessing their capabilities under these demanding conditions. *OinkTrack* is a valuable new resource to drive MOT research, advancing algorithms towards robust, long-term tracking in complex scenarios with direct applications in automated farming, animal health, and welfare.

Limitations and Future Work. While *OinkTrack* provides robust instance-level LTMAT data, future enrichment is possible. One limitation is the lack of fine-grained behavioral labels (*e.g.*, feeding, sleeping, agonistic interactions), which would advance ethological understanding and fuel automated health monitoring systems. Future work could also explore richer group-level annotations, capturing collective dynamics and social interactions via natural language to foster multi-agent behavior analysis and a holistic welfare view. Moreover, *OinkTrack*'s inherent challenges (extreme similarity, prolonged occlusions, day-night visual shifts over long durations) demand novel MOT algorithms for robust re-identification, long-range temporal reasoning, and adaptive appearance modeling.

Acknowledgments

We gratefully acknowledge Da Fong Hog Producers for their invaluable assistance and for providing access to their facilities for the data collection of the *OinkTrack* dataset. This research was supported in part by the National Science and Technology Council (NSTC) of Taiwan under grant number NSTC-114-2640-B-005-001.

References

[1] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. 2022. BoT-SORT: Robust associations multi-pedestrian tracking. *arXiv:2206.14651* (2022).

[2] Hexin Bai, Wensheng Cheng, Peng Chu, Juehuan Liu, Kai Zhang, and Haibin Ling. 2021. Gmot-40: A benchmark for generic multiple object tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*. 6719–6728.

[3] Luca Bergamini, Stefano Pini, Alessandro Simoni, Roberto Vezzani, Simone Calderara, Rick BD Eath, and Robert B Fisher. 2021. Extracting accurate long-term behavior changes from a large pig dataset. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. 524–533.

[4] Philipp Bergmann, Tim Meinhardt, and Laura Leal-Taixe. 2019. Tracking without bells and whistles. In *International Conference on Computer Vision*. 941–951.

[5] Keni Bernardin and Rainer Stiefelhagen. 2008. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing* 2008 (2008), 1–10.

[6] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. 2016. Simple online and realtime tracking. In *IEEE International Conference on Image Processing*. 3464–3468.

[7] Sofia Broomé, Marcelo Feighelstein, Anna Zamansky, Gabriel Carreira Lencioni, Pia Haubro Andersen, Francisca Pessanha, Marwa Mahmoud, Hedvig Kjellström, and Albert Ali Salah. 2023. Going deeper than tracking: A survey of computer-vision based recognition of animal pain and emotions. *International Journal of Computer Vision* 131, 2 (2023), 572–590.

[8] Jiarui Cai, Mingze Xu, Wei Li, Yuanjun Xiong, Wei Xia, Zhuowen Tu, and Stefano Soatto. 2022. Memot: Multi-object tracking with memory. In *IEEE Conference on Computer Vision and Pattern Recognition*. 8090–8100.

[9] Jinkun Cao, Jiangmiao Pang, Xinshuo Weng, Rawal Khirodkar, and Kris Kitani. 2023. Observation-centric sort: Rethinking sort for robust multi-object tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*. 9686–9696.

[10] Orcun Cetintas, Guillem Brasó, and Laura Leal-Taixé. 2023. Unifying short and long-term tracking with graph hierarchies. In *IEEE Conference on Computer Vision and Pattern Recognition*. 22877–22887.

[11] Long Chen, Haizhou Ai, Zijie Zhuang, and Chong Shang. 2018. Real-time multiple people tracking with deeply learned candidate selection and person re-identification. In *International Conference on Multimedia and Expo*. IEEE, 1–6.

[12] Yutao Cui, Chenkai Zeng, Xiaoyu Zhao, Yichun Yang, Gangshan Wu, and Limin Wang. 2023. Sportsmot: A large multi-object tracking dataset in multiple sports scenes. In *International Conference on Computer Vision*. 9921–9931.

[13] Kenan Dai, Yunhua Zhang, Dong Wang, Jianhua Li, Huchuan Lu, and Xiaoyun Yang. 2020. High-performance long-term tracking with meta-updater. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 6298–6307.

[14] Achal Dave, Tarasha Khurana, Pavel Tokmakov, Cordelia Schmid, and Deva Ramanan. 2020. Tao: A large-scale benchmark for tracking any object. In *European Conference on Computer Vision*. Springer, 436–454.

[15] Patrick Dendorfer, Hamid Rezatofighi, Anton Milan, Javen Shi, Daniel Cremers, Ian Reid, Stefan Roth, Konrad Schindler, and Laura Leal-Taixé. 2020. Mot20: A benchmark for multi object tracking in crowded scenes. *arXiv:2003.09003* (2020).

[16] Yunhao Du, Zhicheng Zhao, Yang Song, Yanyun Zhao, Fei Su, Tao Gong, and Hongying Meng. 2023. Strongsort: Make deepsort great again. *IEEE Transactions on Multimedia* 25 (2023), 8725–8737.

[17] Heng Fan, Liting Lin, Fan Yang, Peng Chu, Ge Deng, Sijia Yu, Hexin Bai, Yong Xu, Chunyuan Liao, and Haibin Ling. 2019. Lasot: A high-quality benchmark for large-scale single object tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*. 5374–5383.

[18] Teng Fu, Xiaocong Wang, Haiyang Yu, Ke Niu, Bin Li, and Xiangyang Xue. 2023. Denoising-mot: Towards multiple object tracking with severe occlusions. In *ACM Multimedia*. 2734–2743.

[19] Jie Gao, Bineng Zhong, and Yan Chen. 2023. Unambiguous object tracking by exploiting target cues. In *ACM Multimedia*. 1997–2005.

[20] Ruopeng Gao, Ji Qi, and Limin Wang. 2025. Multiple object tracking as id prediction. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[21] Ruopeng Gao and Limin Wang. 2023. MeMOTR: Long-term memory-augmented transformer for multi-object tracking. In *International Conference on Computer Vision*. 9901–9910.

[22] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition*. 3354–3361.

[23] Lianghua Huang, Xin Zhao, and Kaiqi Huang. 2019. Got-10k: A large high-diversity benchmark for generic object tracking in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 5 (2019), 1562–1577.

[24] Zhiyu Jin, Hanyang Yu, Chen Haul, Linxiang Wang, Zuobin Zhu, Qiu Shen, and Xun Cao. 2023. WormTrack: Dataset and Benchmark for Multi-Object Tracking in Worm Crowds. In *ACM Multimedia*. 5756–5763.

[25] Xiaohai Li, Bineng Zhong, Qihua Liang, Guorong Li, Zhiyi Mo, and Shuxiang Song. 2025. MambaLCT: Boosting Tracking via Long-term Context State Space Model. In *Association for the Advancement of Artificial Intelligence*, Vol. 39. 4986–4994.

[26] Ningxin Liang, Guile Wu, Wenxiong Kang, Zhiyong Wang, and David Dagan Feng. 2018. Real-time long-term tracking with prediction-detection-correction. *IEEE Transactions on Multimedia* 20, 9 (2018), 2289–2302.

[27] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. 2021. Hota: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision* 129 (2021), 548–578.

[28] Chao Ma, Xiaokang Yang, Chongyang Zhang, and Ming-Hsuan Yang. 2015. Long-term correlation tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*. 5388–5396.

[29] Gerard Maggiolino, Adnan Ahmad, Jinkun Cao, and Kris Kitani. 2023. Deep OC-Sort: Multi-Pedestrian Tracking by Adaptive Re-Identification. In *IEEE International Conference on Image Processing*. 3025–3029.

[30] Md Sultan Mahmud, Azlan Zahid, Anup Kumar Das, Muhammad Muzammil, and Muhammad Usman Khan. 2021. A systematic literature review on deep learning applications for precision cattle farming. *Computers and Electronics in Agriculture* 187 (2021), 106313.

[31] Tim Meinhardt, Alexander Kirillov, Laura Leal-Taixe, and Christoph Feichtenhofer. 2022. Trackformer: Multi-object tracking with transformers. In *IEEE Conference on Computer Vision and Pattern Recognition*. 8844–8854.

[32] Anton Milan, Laura Leal-Taixé, Ian Reid, Stefan Roth, and Konrad Schindler. 2016. MOT16: A benchmark for multi-object tracking. *arXiv:1603.00831* (2016).

[33] Abhinav Moudgil and Vineet Gandhi. 2019. Long-term visual object tracking benchmark. In *Asian Conference on Computer Vision*. 629–645.

[34] Hemal Naik, Junran Yang, Dipin Das, Margaret Crofoot, Akanksha Rathore, and Vivek Hari Sridhar. 2024. BuckTales: A multi-UAV dataset for multi-object tracking and re-identification of wild antelopes. *Advances in Neural Information Processing Systems* 37 (2024), 81992–82009.

[35] Suresh Neethirajan. 2023. SOLARIA-SensOr-driven resiLient and adaptive monitoRIng of farm Animals. *Agriculture* 13, 2 (2023), 436.

[36] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. 2016. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision*. Springer, 17–35.

[37] Mattia Segu, Luigi Piccinelli, Siyuan Li, Yung-Hsu Yang, Bernt Schiele, and Luc Van Gool. 2025. Samba: Synchronized Set-of-Sequences Modeling for Multiple Object Tracking. In *International Conference on Learning Representations*.

[38] Aniket Shirke, Aziz Saifuddin, Achleshwar Luthra, Jiangong Li, Tawni Williams, Xiaodan Hu, Aneesh Kotnana, Okan Kocabalkanli, Narendra Ahuja, Angela Green-Miller, et al. 2021. Tracking grow-finish pigs across large pens using multiple cameras. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*.

[39] Peize Sun, Jinkun Cao, Yi Jiang, Zehuan Yuan, Song Bai, Kris Kitani, and Ping Luo. 2022. Dancetrack: Multi-object tracking in uniform appearance and diverse motion. In *IEEE Conference on Computer Vision and Pattern Recognition*. 20993–21002.

[40] Bhavesh Tangirala, Ishan Bhandari, Daniel Laszlo, Deepak K Gupta, Rajat M Thomas, and Devanshu Arya. 2021. Livestock Monitoring with Transformer. In *British Machine Vision Conference*.

[41] Pavel Tokmakov, Jie Li, Wolfram Burgard, and Adrien Gaidon. 2021. Learning to track with object permanence. In *International Conference on Computer Vision*. 10860–10869.

[42] Jack Valmadre, Luca Bertinetto, Joao F Henriques, Ran Tao, Andrea Vedaldi, Arnold WM Smeulders, Philip HS Torr, and Efstratios Gavves. 2018. Long-term tracking in the wild: A benchmark. In *European Conference on Computer Vision*. 670–685.

[43] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008).

[44] Zhongdao Wang, Liang Zheng, Yixuan Liu, Yali Li, and Shengjin Wang. 2020. Towards real-time multi-object tracking. In *European Conference on Computer Vision*. 107–122.

[45] Xing Wei, Yifan Bai, Yongchao Zheng, Dahu Shi, and Yihong Gong. 2023. Autoregressive visual tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*. 9697–9706.

[46] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. 2017. Simple online and realtime tracking with a deep association metric. In *IEEE International Conference on Image Processing*. 3645–3649.

[47] Jinxia Xie, Bineng Zhong, Zhiyi Mo, Shengping Zhang, Liangtao Shi, Shuxiang Song, and Rongrong Ji. 2024. Autoregressive queries for adaptive tracking with spatio-temporal transformers. In *IEEE Conference on Computer Vision and Pattern Recognition*. 19300–19309.

[48] Mingzhan Yang, Guangxin Han, Bin Yan, Wenhua Zhang, Jinqing Qi, Huchuan Lu, and Dong Wang. 2024. Hybrid-sort: Weak cues matter for online multi-object tracking. In *Association for the Advancement of Artificial Intelligence*, Vol. 38. 6504–6512.

[49] Fangao Zeng, Bin Dong, Yuang Zhang, Tiancai Wang, Xiangyu Zhang, and Yichen Wei. 2022. Motr: End-to-end multiple-object tracking with transformer. In *European Conference on Computer Vision*. 659–675.

[50] Libo Zhang, Junyuan Gao, Zhen Xiao, and Heng Fan. 2023. Animaltrack: A benchmark for multi-animal tracking in the wild. *International Journal of Computer Vision* 131, 2 (2023), 496–513.

[51] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. 2022. Bytetrack: Multi-object tracking by associating every detection box. In *European Conference on Computer Vision*. 1–21.

[52] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. 2021. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision* (2021), 3069–3087.

[53] Yuang Zhang, Tiancai Wang, and Xiangyu Zhang. 2023. Motrv2: Bootstrapping end-to-end multi-object tracking by pretrained object detectors. In *IEEE Conference on Computer Vision and Pattern Recognition*. 22056–22065.

[54] Guangze Zheng, Shijie Lin, Haobo Zuo, Changhong Fu, and Jia Pan. 2024. Nettrack: Tracking highly dynamic objects with a net. In *IEEE Conference on Computer Vision and Pattern Recognition*. 19145–19155.

[55] Tao Zhou, Wenhan Luo, Zhiguo Shi, Jiming Chen, and Qi Ye. 2022. Apptracker: Improving tracking multiple objects in low-frame-rate videos. In *ACM Multimedia*. 6664–6674.

[56] Tianyu Zhu, Markus Hiller, Mahsa Ehsanpour, Rongkai Ma, Tom Drummond, Ian Reid, and Hamid Rezatofighi. 2022. Looking beyond two frames: End-to-end multi-object tracking using spatial and temporal transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 11 (2022), 12783–12797.