Investigating Multi-Camera Tracking and Re-Identification Pipelines Based on YOLOv8

Boshen Ruan, Beifei Wei and Stefan Twieg
Institut für Medizin und Technik e.V., Bernburger Str. 57, 06366 Koethen, Germany
Boshen.Ruan@student.hs-anhalt.de, Beifei.Wei@student.hs-anhalt.de, Stefan.Twieg@hs-anhalt.de

Keywords: Deep Learning, YOLO, ByteTrack, BoT-SORT, Object Detection, Multi-Camera Tracking, Re-Identification, Machine Learning, Animal.

Abstract: Multi-camera identification and tracking of individual animals is a challenging task in complex monitoring environments due to frequent occlusions, viewpoint changes, and visual similarity between individuals. This challenge is particularly relevant for future animal monitoring scenarios, such as goose tracking in agricultural settings. However, collecting large-scale, well-annotated multi-camera animal datasets remains difficult. In this work, we investigate a conceptual multi-camera tracking and re-identification (ReID) framework based on modern detection, tracking, and appearance-based re-identification methods. The proposed pipeline integrates YOLOv8 for object detection with ByteTrack and BoT-SORT for single-camera multi-object tracking, generating stable tracklets as the basis for cross-camera identity association. To evaluate the feasibility of the re-identification component, experiments are conducted on publicly available human benchmark datasets, which are widely used for appearance-based ReID research. The results demonstrate that the proposed modular pipeline can reliably perform detection, tracking, and identity embedding extraction in multi-camera scenarios. Furthermore, we discuss the transferability of the framework to animal monitoring applications and analyze the limitations arising from domain gaps between human and animal data. The presented study serves as a proof-of-concept and provides a practical foundation for future deployment in real-world goose monitoring scenarios.

1 INTRODUCTION

Continuous monitoring of individuals in complex environments is an important task in applications such as smart farming and automated surveillance [1]. Multi-camera systems are commonly used to cover large areas and reduce blind spots that cannot be handled by single-camera setups [2]. However, maintaining consistent identities over time and across different camera views remains challenging due to occlusions, viewpoint changes, and visually similar objects [3].

In animal monitoring scenarios, including future goose husbandry applications [4], these challenges are further intensified. Animals often exhibit similar appearances and frequent interactions, making reliable long-term identity tracking difficult. Manual monitoring is inefficient, while single-camera systems easily lose identities when objects leave the field of view.

To address these issues, this paper investigates a conceptual multi-camera tracking and re-identification framework based on modern deep learning methods [5]. The proposed pipeline combines YOLOv8 for object detection with ByteTrack [6], [7] and BoT-SORT [8] for single-camera multi-object tracking, producing stable tracklets for further identity association [9].

Due to the limited availability of annotated multi-camera animal datasets, the re-identification component is evaluated using well-known human benchmark datasets, including Market1501 and DukeMTMC [10], [11]. This allows a controlled validation of the framework before its application to animal monitoring. In this study, the term conceptual framework refers to a modular pipeline architecture that integrates object detection, single-camera tracking, tracklet generation, and multi-camera identity association. The primary objective of this work is to establish a practical and modular framework that can be adapted to real-world goose tracking scenarios in future studies.

Therefore, the main contribution of this work is not the development of a fundamentally new tracking algorithm, but the engineering integration and experimental analysis of a modular multi-camera tracking pipeline. By evaluating YOLOv8-based detection, ByteTrack- and BoT-SORT-based tracking, and geometry- and appearance-based cross-camera identity association under a unified workflow, this study provides practical insights into the strengths and limitations of current multi-camera tracking and re-identification methods in dense scenes.

2 THEORETICAL FOUNDATIONS

2.1 Object Detection in Agriculture

Deep learning-based object detection has been widely applied in agricultural and livestock monitoring scenarios to automate the observation of animals and farming environments. In recent years, YOLO-based models have gained significant attention due to their real-time performance and competitive detection accuracy [1]. Several studies have demonstrated the effectiveness of YOLO variants for detecting livestock such as cattle, pigs, and poultry in complex farm environments, making them suitable candidates for practical monitoring systems [12]. While two-stage detectors (e.g., Faster R-CNN, Mask R-CNN), anchor-free detectors (e.g., FCOS, CenterNet), and transformer-based models (e.g., DETR, Deformable DETR) are also applicable to agricultural object detection tasks, YOLOv8 is selected in this study. Although YOLOv8 may experience performance degradation in scenarios with heavy object overlap or when distinguishing visually similar animals, it provides superior real-time performance, seamless integration with tracking algorithms such as ByteTrack and BoT-SORT, and a stable baseline supported by a large and active community.

2.2 Animal Re-Identification

Animal re-identification aims at recognizing individual animals across time and different viewpoints [13]. Existing approaches can be broadly categorized into biometric-based methods, which rely on natural appearance features, and marker-based methods, which utilize artificial identifiers such as tags or collars. While biometric approaches have shown promising results in controlled settings, their performance often degrades in real-world livestock scenarios due to strong appearance similarity between individuals. Marker-based identification, when available, can provide more reliable identity cues but requires additional infrastructure and data acquisition [13]. Alternative approaches for individual recognition, including trajectory-based features, gait-based features, spatio-temporal behavior signatures, and multi-modal sensing, have demonstrated promising potential. However, these methods are not considered in the present study due to the absence of specialized annotated datasets for animal gait analysis and the high system complexity associated with multi-modal hardware deployment. For visual feature extraction, established software libraries, namely FastReID and the OpenAI implementation of CLIP, are utilized.

2.3 Multi-Object Tracking in Livestock Monitoring

Multi-object tracking (MOT) plays a crucial role in maintaining consistent identities over time [14]. Recent tracking-by-detection methods, such as ByteTrack [6] and BoT-SORT [3], have achieved strong performance in crowded scenes by effectively associating detection results into stable trajectories. Although these methods were primarily developed and evaluated on human-centered datasets, their robustness and modular design make them attractive candidates for exploratory studies in animal monitoring and livestock tracking scenarios [15]. ByteTrack and BoT-SORT are selected instead of alternatives such as DeepSORT or StrongSORT. The latter relies heavily on appearance features for data association, which can be unreliable in livestock monitoring scenarios due to the high visual similarity between individuals. In contrast, ByteTrack and BoT-SORT emphasize motion consistency and high-confidence detection results, enabling more stable and continuous trajectories.

Table 1: Comparison of evaluated multi-camera identity association strategies.

Strategy	Features	Limitation
Geometry-based	Ground-plane coords	Calibration, density
Appearance-based	Visual embeddings	Viewpoint, similarity
Semantic-based	CLIP embeddings	Noise in dense scenes

Table 1 summarizes the key characteristics of the evaluated multi-camera identity association strategies under a unified detection and tracking setup, highlighting the trade-offs between geometry-based, appearance-based, and semantic-based approaches in dense livestock monitoring scenarios.

2.4 Proposed Conceptual Pipeline for Multi-Camera Identity Tracking

The proposed multi-camera tracking and identity association pipeline follows a modular structure, as illustrated in Figure 1. The process starts with synchronized multi-camera video input, in which each camera stream is processed independently. Single-camera object detection is first performed using YOLOv8, followed by single-camera tracking with ByteTrack or BoT-SORT to link detections over time and generate continuous trajectories within each view. These trajectories are then temporally aggregated into tracklets, which provide more stable units for subsequent cross-camera analysis.

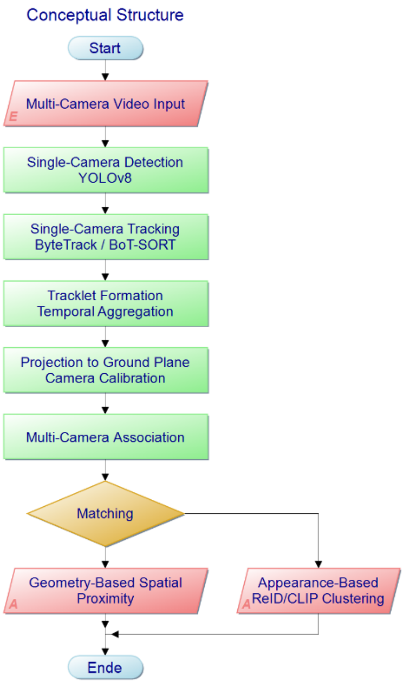

Figure 1: Conceptual pipeline for multi-camera object tracking and identity association.

The resulting tracklets are projected onto a shared ground-plane coordinate system using camera calibration to enable spatial alignment across different camera views. Based on these aligned representations, multi-camera identity association is performed by matching tracklets across cameras, either through geometry-based spatial proximity or through appearance-based clustering using ReID or CLIP embeddings.

3 PRACTICAL IMPLEMENTATION AND RESULTS

3.1 Dataset and Experimental Setup

To validate the complete multi-camera tracking and identification pipeline, the Wildtrack dataset is employed as a proof-of-concept benchmark. Wildtrack provides synchronized multi-camera videos together with precise ground-truth annotations and calibrated camera parameters, making it suitable for evaluating multi-camera tracking workflows.

In our experiments, videos from Cameras 1, 3, and 5 are selected to represent different viewpoints and spatial configurations. The dataset is primarily used to verify the feasibility of the proposed pipeline rather than to achieve state-of-the-art performance.

3.2 Detection and Single-Camera Tracking

In the first stage, the original annotations provided by the Wildtrack dataset are used to train a YOLOv8-based object detector [7]. The training process converges steadily, and the detector achieves high detection performance on the validation set. As shown in Figure 2, the mAP@0.5 increases rapidly during training and stabilizes at a high value, indicating reliable localization accuracy for pedestrian detection in crowded scenes.

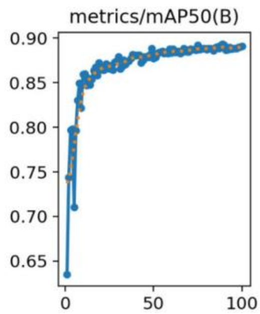

Figure 2: Evolution of mAP50 on the validation set during training.

In addition to quantitative metrics, qualitative detection results further confirm the robustness of the trained detector. Figure 3 present representative detection outputs on Wildtrack video frames, where pedestrians are accurately localized despite strong perspective variations and high object density. These results demonstrate that the trained YOLOv8 model provides sufficiently reliable detections as input for subsequent tracking and identity association stages.

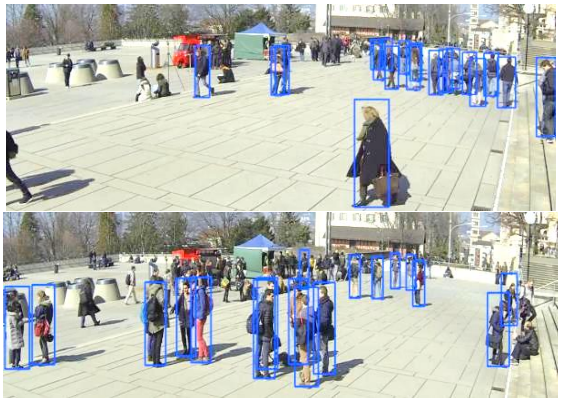

Figure 3: Qualitative detection results of the trained YOLOv8 detector on the Wildtrack dataset.

Subsequently, the trained detector is applied to reprocess the Wildtrack videos from Camera 1, Camera 3, and Camera 5. Detection outputs are integrated with a tracking-by-detection framework to generate videos with persistent track IDs and corresponding label files [6]. Visual inspection of the tracking results shows that identities remain consistent over consecutive frames within a single camera view. As illustrated in Figure 4, tracked individuals preserve stable track IDs even under moderate viewpoint changes and partial occlusions.

Figure 4: Single-camera tracking results on the Wildtrack dataset.

Overall, both quantitative detection metrics and qualitative visualization results confirm that the proposed detection and single-camera tracking pipeline operates reliably. This stable performance provides a solid foundation for the subsequent investigation of cross-camera identity association and re-identification.

3.3 Geometry-Based Multi-Camera Association

To explore multi-camera identity association without appearance-based features, the generated label files are first converted from text format to CSV files. Using the intrinsic and extrinsic camera parameters provided by Wildtrack, detections are projected onto a common ground-plane coordinate system.

Based on this representation, identities across different cameras are associated by comparing the spatial proximity of individuals within a manually defined temporal window. Additional constraints, including mutual nearest-neighbor matching and multi-frame consistency, are introduced to reduce incorrect associations.

Although the number of overlapping frame pairs is abundant (over 540,710 pairs with ⩾10 frames overlap, see Table 2), no tracklets were actually merged during the process. The final number of global IDs equals 29,478, almost identical to the input tracklets (Table 2). This indicates that geometric information alone is insufficient to resolve identities in dense and occluded scenarios, as present in the Wildtrack dataset.

Table 2: Results of geometry-based multi-camera association.

Metric	Value
Number of valid tracklets	27,512
Number of candidates tracklets for multi-camera association	27,512
Number of merge operations (union)	0
Total number of tracklet pairs across cameras	242,326,032
Tracklet pairs with ≥1 frame overlap	789,754
Tracklet pairs with ≥3 frames overlap	763,743
Tracklet pairs with ≥10 frames overlap	540,710
Final number of global IDs	29,478

3.4 Re-Identification and Embedding-Based Association

To further improve cross-camera identity association, appearance-based re-identification (ReID) is integrated into the pipeline. ReID models are trained using Market1501 and DukeMTMC to learn discriminative feature embeddings. Using the YOLOv8 detection results, video frames are cropped to generate person image patches. For each track, embeddings are extracted using a trained FastReID model and aggregated via track-level mean pooling. Subsequently, pairwise cosine distances between track embeddings are computed and the DBSCAN clustering algorithm is applied to assign global identities across cameras.

The intrinsic capability of the ReID embeddings is first validated using standard ReID evaluation metrics. As shown in Figure 5, the model trained/evaluated on Market1501 achieves consistently high retrieval performance (approximately Rank-1 ≈ 95%, Rank-5 ≈ 99%, Rank-10 ≈ 100%), indicating strong top-k identification accuracy on that benchmark. In contrast, the DukeMTMC setting shows noticeably lower performance (approximately Rank-1 ≈ 40%, Rank-5 ≈ 55%, Rank-10 ≈ 62%), suggesting that appearance-based discrimination becomes significantly harder under that dataset’s camera conditions.

Beyond rank-based metrics, mAP and mINP provide additional insight into overall ranking quality and hard-case robustness. In Figure 5, Market1501 maintains a high mAP (≈ 90%) and a relatively strong mINP (≈ 65%), which implies that correct matches tend to appear early in the ranked list even for more challenging identities. Conversely, DukeMTMC exhibits much lower mAP (≈ 25%) and near-zero mINP (≈ 5%), indicating frequent failure cases where true matches are ranked far down the list. This gap suggests substantial sensitivity of appearance embeddings to viewpoint, illumination, and background changes.

In practice, only 6 merge operations occurred (Table 3), and the final number of global IDs was 29,472, almost equal to the input tracklets. The average cosine similarity of 0.755 indicates high overlap between embeddings of different individuals, resulting in poor clustering performance. This demonstrates that ReID embeddings alone are insufficient for robust identity matching in dense, occluded scenarios.

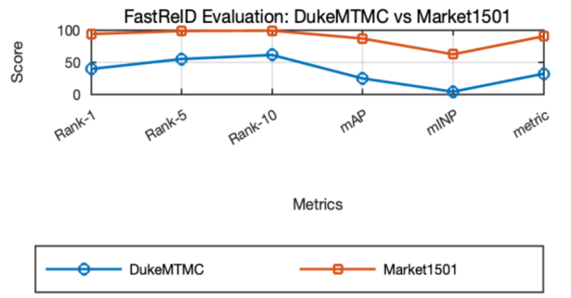

Figure 5: Comparison of ReID models trained on DukeMTMC and Market1501.

Table 3: Results of ReID- and embedding-based association.

Metric	Value
Original number of tracklets	29,478
Number of valid tracklets (sufficient frames + embedding)	27,512
Tracklets with ground coordinates	27,512
Number of candidate tracklets for multi-camera association	27,512
Number of merge operations (union)	6
Cosine similarity range	0.629 – 0.904
Average cosine similarity	0.755
Final number of global IDs	29,472

3.5 Appearance-Based Identity Matching using CLIP

CLIP-based association [16] uses pre-trained visual embeddings for each tracked individual and assigns global IDs via DBSCAN clustering [17]. Track-level embeddings are computed by averaging frame-wise embeddings, and pairwise cosine distances are used for clustering. The final number of global IDs is 245, with 3,739 tracklets marked as noise (Table 4). Although CLIP provides a stronger visual representation than ReID, clustering remains unstable in dense and occluded scenarios [3], [18]. This suggests that relying solely on appearance features is insufficient. Compared to geometric and ReID methods, CLIP shows some improvement but still requires additional strategies such as trajectory-based association, motion consistency, or gait features for robust multi-camera tracking [19], [20].

Table 4: Results of CLIP-based visual identity matching.

Metric	Value
Total number of tracklets	26,283
Final number of global IDs	245
Tracklets marked as noise (-1)	3,739
DBSCAN parameter eps	0.035
DBSCAN parameter min_samples	6

3.6 Comparative Analysis and Observations

The results obtained from both FastReID-based and CLIP-based appearance matching approaches reveal similar limitations in dense multi-camera scenarios. In both cases, the global identity assignment is highly sensitive to clustering parameters, resulting either in identity fragmentation-where the same individual is assigned multiple global identities-or identity merging, where multiple individuals are incorrectly grouped [3].

Despite the stronger representation capability of CLIP, no consistent improvement over traditional ReID embeddings is observed [16]. These findings indicate that appearance-based features alone are insufficient to reliably resolve cross-camera identity association under heavy occlusion, strong viewpoint variation, and crowded conditions, as exemplified by the Wildtrack dataset [18].

These findings highlight the need to investigate alternative multi-camera collaboration strategies, such as trajectory-based association, motion consistency, or gait-related features [19], which may offer better robustness than appearance-based methods in animal monitoring contexts.

It is important to note that Table 3, which reports tracklet-level statistics for ReID, and Table 4, which summarizes clustering outcomes for CLIP, are based on fundamentally different evaluation protocols. Specifically, FastReID is assessed using retrieval-based metrics, including Rank-1/5/10 accuracy and mAP, whereas CLIP is evaluated using clustering statistics. Consequently, these tables should be interpreted as diagnostic indicators rather than as results intended for direct quantitative comparison. In the absence of unified multi-object tracking metrics (e.g., IDF1 or MOTA), the analysis is therefore limited to a qualitative assessment of the feasibility of the respective methods.

4 DISCUSSION AND FUTURE WORK

4.1 Environment-Related Challenges in Real-World Animal Monitoring

The experimental results clearly reveal both the strengths and limitations of the proposed multi-camera tracking and identity association framework. While object detection and single-camera tracking achieve stable performance, cross-camera identity association remains a major challenge, especially in dense scenes [9]. Experiments on the Wildtrack dataset demonstrate that under conditions of high target density and frequent occlusions, both geometry-based methods and appearance-based re-identification struggle to maintain consistent identities.

A key challenge arises from occlusion handling. When multiple animals overlap or move in proximity, identity switches frequently occur, and simple spatial or appearance-based cues are insufficient to preserve identity continuity [3]. Addressing such cases requires more robust occlusion-aware strategies, including long-term trajectory reasoning, temporal consistency constraints, and identity recovery mechanisms after occlusion events.

A further limitation is the domain gap between human-centered benchmark datasets and animal monitoring scenarios. The Wildtrack dataset contains pedestrians with upright body structures, relatively stable shapes, and more distinguishable clothing-based appearance cues. In contrast, geese and other livestock animals often have highly similar visual appearances, non-rigid body poses, smaller visible body regions, and frequent physical interactions. In addition, animal monitoring environments are affected by dirt, changing illumination, partial occlusions, and camera viewpoints close to the ground. As a result, ReID models trained or validated on human datasets cannot be directly transferred to animal tracking without further adaptation.

In addition, real-world animal monitoring environments introduce challenges that are not fully represented in existing human-centered datasets. Factors such as environmental contamination, uneven ground surfaces, and unfavorable lighting conditions inside barns or livestock facilities can significantly degrade detection and tracking performance. These environmental factors further amplify identity ambiguities and reduce the reliability of appearance-based representations. Overcoming these challenges will require dedicated animal-specific data collection and the incorporation of illumination-robust feature representations and adaptive preprocessing techniques.

4.2 Beyond Appearance-Based Identity Cues

Given the limitations observed in appearance-based re-identification, it is necessary to explore alternative identity cues beyond visual appearance. In particular, trajectory-based and gait-based information represent promising directions for multi-camera identity association in animal monitoring scenarios.

Compared to appearance features, motion-related cues rely less on texture, color, and fine-grained visual details, which are often unstable in real-world animal monitoring environments. For animals such as geese, appearance-based ReID is especially challenging due to the high visual similarity between individuals and the limited availability of distinctive visual traits. Environmental factors such as dirt, mud attachment, and lighting variations further reduce the reliability of appearance-based representations.

In contrast, trajectory-based features, including movement direction, speed patterns, and temporal-spatial consistency, can provide more stable identity cues across multiple camera views [3], [21]. Such features capture behavioral regularities over time rather than instantaneous visual similarity. Similarly, gait-based features focus on periodic motion patterns instead of static appearance, making them potentially more robust to occlusion and viewpoint changes.

Although gait recognition has been extensively studied in human identification, its application in animal monitoring remains limited and largely unexplored. For geese, relatively consistent walking rhythms and motion dynamics may form implicit identity signatures when observed over extended time periods. Therefore, integrating gait- and trajectory-based cues with existing detection and tracking pipelines represents an important direction for future research.

Overall, these findings suggest that relying solely on appearance-based ReID is insufficient for reliable identity association in animal monitoring. Future work should focus on combining motion-based cues with detection and tracking frameworks, and, where applicable, fusing weak appearance information to achieve more robust multi-camera identity tracking in livestock environments.

For future adaptation to goose monitoring, several concrete steps are required. First, animal-specific multi-camera datasets should be collected and manually annotated under different lighting conditions, camera viewpoints, and density levels. Second, the YOLOv8 detector should be fine-tuned on goose images to improve the detection of small, occluded, and overexposed animals. Third, cross-camera association should combine multiple cues, including calibrated ground-plane positions, temporal movement consistency, trajectory direction, gait-related motion patterns, and, where available, external identity markers such as neck collars. Fourth, appearance-based ReID should not be used as the only identity cue, but rather as a weak complementary feature together with motion-based and scene-geometric information. These steps would allow the proposed pipeline to move from a proof-of-concept on human benchmark data toward a practical multi-camera animal monitoring system.

5 CONCLUSIONS

This paper presents a conceptual multi-camera tracking and identification framework and evaluates its feasibility using the Wildtrack dataset. By integrating YOLOv8-based detection with modern tracking methods and cross-camera association strategies, the complete processing pipeline is successfully implemented and validated in practice.

Experimental results show that object detection and single-camera tracking achieve stable performance in crowded scenes. However, the study also demonstrates that both geometry-based association and appearance-based re-identification face significant limitations in dense multi-camera environments. These findings highlight the challenges of maintaining consistent identities under occlusions and complex scene conditions.

Overall, this work provides valuable insights into the strengths and weaknesses of current multi-camera collaboration approaches. Rather than delivering a finalized identification system, the presented framework serves as a practical baseline and reference for future research. The conclusions drawn from this study will guide further development toward more robust multi-camera animal monitoring systems, including potential applications in goose tracking and smart farming.

ACKNOWLEDGMENTS

The research project underlying this contribution (NEXON) was funded by the European Regional Development Fund (EFRE). The responsibility for the content of this publication lies with the author.

REFERENCES

[1] D. Wang, W. Cao, F. Zhang, Z. Li, S. Xu, and X. Wu, “A Review of Deep Learning in Multiscale Agricultural Sensing,” Remote Sens., vol. 14, no. 3, p. 559, Jan. 2022, [Online]. Available: https://doi.org/10.3390/rs14030559.

[2] T. I. Amosa et al., “Multi-camera multi-object tracking: A review of current trends and future advances,” Neurocomputing, vol. 552, p. 126558, Oct. 2023, [Online]. Available: https://doi.org/10.1016/j.neucom.2023.126558.

[3] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, “Multiple object tracking: A literature review,” Artif. Intell., vol. 293, p. 103448, Apr. 2021, [Online]. Available: https://doi.org/10.1016/j.artint.2020.103448.

[4] R. A. Rajagukguk et al., “Deep learning for visual animal monitoring (detection, tracking, pose estimation, and behavior classification): A comprehensive review,” Smart Agric. Technol., vol. 12, p. 101539, Dec. 2025, [Online]. Available: https://doi.org/10.1016/j.atech.2025.101539.

[5] M. A. Ingrisch, R. M. Schilling, I. Chmielewski, and S. Twieg, "Determining the Optimal T-Value for the Temperature Scaling Calibration Method Using the Open-Vocabulary Detection Model YOLO-World," Appl. Sci., vol. 15, no. 22, p. 12062, Jan. 2025, [Online]. Available: https://doi.org/10.3390/app152212062.

[6] Y. Zhang et al., "ByteTrack: Multi-object Tracking by Associating Every Detection Box," in Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXII, Berlin, Heidelberg: Springer-Verlag, Oct. 2022, pp. 1-21, [Online]. Available: https://doi.org/10.1007/978-3-031-20047-2_1.

[7] L. You, Y. Chen, C. Xiao, C. Sun, and R. Li, "Multi-Object Vehicle Detection and Tracking Algorithm Based on Improved YOLOv8 and ByteTrack," Electronics, vol. 13, no. 15, p. 3033, Jan. 2024, [Online]. Available: https://doi.org/10.3390/electronics13153033.

[8] J. Zhao and J. Chen, "YOLOv8 Detection and Improved BOT-SORT Tracking Algorithm for Iron Ladles," in Proceedings of the 2024 7th International Conference on Image and Graphics Processing, in ICIGP '24, New York, NY, USA: Association for Computing Machinery, May 2024, pp. 409-415, [Online]. Available: https://doi.org/10.1145/3647649.3647713.

[9] E. Ristani and C. Tomasi, "Features for Multi-target Multi-camera Tracking and Re-identification," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 6036-6046, [Online]. Available: https://doi.org/10.1109/CVPR.2018.00632.

[10] Z. Tian, S. Chen, D.-H. Wang, and J. Lu, "A Survey of Person Re-identification Based on Deep Learning," in Proceedings of the 2021 10th International Conference on Computing and Pattern Recognition, in ICCPR '21, New York, NY, USA: Association for Computing Machinery, Feb. 2022, pp. 36-42, [Online]. Available: https://doi.org/10.1145/3497623.3497630.

[11] Z. Hao, H. Ge, and J. Huang, "Research on an unsupervised person re-identification based on image quality enhancement method," Eng. Appl. Artif. Intell., vol. 123, p. 106392, Aug. 2023, [Online]. Available: https://doi.org/10.1016/j.engappai.2023.106392.

[12] Z. Zheng, J. Li, and L. Qin, "YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows," Comput. Electron. Agric., vol. 209, p. 107857, Jun. 2023, [Online]. Available: https://doi.org/10.1016/j.compag.2023.107857.

[13] H. Meng et al., "Livestock Biometrics Identification Using Computer Vision Approaches: A Review," Agriculture, vol. 15, no. 1, p. 102, Jan. 2025, [Online]. Available: https://doi.org/10.3390/agriculture15010102.

[14] M. Galinski, I. Jatz, and L. Soltes, "Head-On Collision Avoidance Using V2X Communication," in 2023 International Symposium ELMAR, Sep. 2023, pp. 37-40, [Online]. Available: https://doi.org/10.1109/ELMAR59410.2023.10253932.

[15] J. Yang, K. Wang, R. Li, Z. Qin, and P. Perner, "A novel fast combine-and-conquer object detector based on only one-level feature map," Comput. Vis. Image Underst., vol. 224, p. 103561, Nov. 2022, [Online]. Available: https://doi.org/10.1016/j.cviu.2022.103561.

[16] S. Li, L. Sun, and Q. Li, "CLIP-ReID: Exploiting Vision-Language Model for Image Re-identification without Concrete Text Labels," Proc. AAAI Conf. Artif. Intell., vol. 37, no. 1, pp. 1405-1413, Jun. 2023, [Online]. Available: https://doi.org/10.1609/aaai.v37i1.25225.

[17] J. Kim and J. Cho, "DBSCAN-Based Tracklet Association Annealer for Advanced Multi-Object Tracking," Sensors, vol. 21, no. 17, p. 5715, Jan. 2021, [Online]. Available: https://doi.org/10.3390/s21175715.

[18] T. Chavdarova et al., "WILDTRACK: A Multi-camera HD Dataset for Dense Unscripted Pedestrian Detection," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: IEEE, Jun. 2018, pp. 5030-5039, [Online]. Available: https://doi.org/10.1109/CVPR.2018.00528.

[19] A. Sepas-Moghaddam and A. Etemad, "Deep Gait Recognition: A Survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 264-284, Jan. 2023, [Online]. Available: https://doi.org/10.1109/TPAMI.2022.3151865.

[20] Y. Li, L. Wu, Y. Chen, X. Wang, G. Yin, and Z. Wang, "Motion estimation and multi-stage association for tracking-by-detection," Complex Intell. Syst., vol. 10, no. 2, pp. 2445-2458, Apr. 2024, [Online]. Available: https://doi.org/10.1007/s40747-023-01273-3.

[21] R. Menghani, S. Rajanayagam, and S. Twieg, "Optimized Indoor and Outdoor SLAM for a Quadruped-Based Robot Walking Partner to Safely Promote Elderly Mobility and Support Nursing Staff," Jun. 2025, [Online]. Available: https://doi.org/10.1109/ISSC67739.2025.11291245.