Multi-Object Tracking Algorithm for Rodents Based On Re-Identification Enhancement

Kaifang Cheng, Hongchen Zhu

***Abstract*—In modern biological research, rodents play a crucial role as key experimental subjects in numerous fields. Accurately tracking the movements and behaviors of rodents is of great significance for obtaining valuable research data. Traditional tracking methods, such as marker - based techniques, often interfere with natural behaviors and fail in complex scenarios. This paper proposes a novel multi - object rodent tracking algorithm enhanced by re - identification. This algorithm combines motion features and re - identification features to reduce identity switches and false detections, enabling precise tracking of rodents. Experimental results on the AnimalTrack dataset show a significant improvement in the MOTA and IDF1 metrics, especially in complex scenarios. The proposed algorithm provides a robust solution for rodent behavior analysis and has potential applications in neuroscience, ethology, and drug research.**

***Index Terms*—Behavior analysis,Deep learning,Multi - object tracking,Re - identification,Rodents**

I. Introduction

Rodents, owing to their physiological structure and genetic characteristics that bear some resemblance to humans, coupled with their short reproductive cycles and ease of breeding and handling, have emerged as ideal models for a plethora of disease research, drug development, and neuroscience exploration. For instance, in neuroscience, the observation and analysis of rodent behavior in specific environments can provide profound insights into cognitive functions, learning and memory mechanisms, and the pathogenesis of neurological diseases. Accurate tracking of animal movement trajectories, behavioral patterns, and social interactions is fundamental to obtaining valuable research data.

Traditional animal tracking methods primarily include marker-based tracking and some simple visual tracking techniques. Marker-based methods, such as attaching special tags, collars, or dyes to animals, can distinguish individuals to a certain extent. However, these markers may interfere with the natural behavior of animals, affecting the authenticity of experimental results. Moreover, in long-term tracking experiments, markers may fall off or become damaged, leading to tracking failures. Simple visual tracking techniques, such as those based on color, texture, or shape features, exhibit significant limitations when faced with animals of similar appearance, frequent posture changes, and occlusions. For example, when multiple rodents gather together, their small size and similar fur color and shape make it easy for traditional visual feature-based tracking algorithms to confuse individual identities, failing to accurately track specific animals continuously.

With the continuous advancement of computer vision technology, techniques such as object detection, feature extraction, and image recognition have become increasingly sophisticated. In the field of animal tracking, computer vision technology can automatically extract animal feature information from video images without the need for physical markers, thereby reducing interference with the animals. Re-identification algorithms, leveraging the feature representation capabilities of deep learning, can learn unique and distinctive features of individual animals. Even in complex experimental environments, they can effectively identify the same animal across different times and scenes, providing robust technical support for high-precision and robust rodent tracking. This has gradually become a popular direction in animal tracking research with broad application prospects.

Based on this, we propose a re-identification-enhanced rodent tracking algorithm aimed at achieving precise localization of rodents by integrating their motion features. The algorithm first employs deep learning techniques to extract rodent features, then constructs stable multi-frame connections to continuously track multiple rodent targets. The entire system is divided into three core modules: the Rodent Tracking Module (RTM), the Re-Identification Module (RIM), and the Data Association Module (DAM).

II. Related Work

A. Re-Identification Algorithms

With the rapid advancement of computer vision technology, re-identification (Re-ID) has emerged as a critical research direction in the visual domain, particularly in cross-camera and multi-environment recognition of targets such as pedestrians, vehicles, and animals. Traditional Re-ID methods predominantly relied on handcrafted features. However, the rise of deep learning — marked by the introduction of representation learning, metric learning, local feature extraction, video sequence analysis, and generative adversarial networks (GANs)—has significantly improved the accuracy, robustness, and applicability of existing algorithms.

Representation learning, as a core concept in deep learning, focuses on automatically extracting effective feature

Manuscript received February 15, 2025

Kaifang Cheng, School of Computer Science and Technology, TianGong University, TianJin, Chin

Hongchen Zhu, School of Computer Science and Technology, TianGong University, TianJin, China

representations from data [1]. This approach drastically reduces dependence on manual feature engineering, enabling models to autonomously learn and capture subtle details of targets. In pedestrian Re-ID, for instance, representation learning allows models to accurately identify appearance features such as clothing and facial characteristics, thereby enhancing cross-camera recognition capabilities [2]. Similarly, this technique applies to vehicle and animal Re-ID. Through convolutional neural networks (CNNs), details like vehicle license plates, headlights, and animal body shapes or fur textures can be effectively captured, further improving recognition accuracy.

Metric learning provides theoretical support for addressing cross-camera and cross-view challenges in Re-ID tasks [3]. By learning an optimal metric space, it ensures that similar targets are closer in the feature space, while dissimilar ones are farther apart. This approach is widely adopted in pedestrian, vehicle, and animal Re-ID. For example, Triplet Loss [4] and Contrastive Loss [5] optimize inter-sample distance relationships, enabling models to better distinguish between targets and significantly improving cross-view and cross-camera recognition accuracy. Models incorporating metric learning efficiently handle complex scene variations while maintaining stability and precision.

Local feature extraction has become indispensable for further enhancing recognition accuracy [6]. By segmenting target images and extracting localized features, models can focus on key regions while mitigating background interference. In pedestrian Re-ID, Part-based Convolutional Neural Networks (PCB) [7] divide images into multiple regions for independent feature extraction, substantially boosting performance. Analogously, for vehicle and animal Re-ID, local feature extraction helps identify details such as vehicle logos, license plates, or animal fur patterns, strengthening cross-view and cross-camera recognition [8][9].

The integration of video sequences introduces temporal dynamics to Re-ID tasks. Compared to static images, video sequences capture temporal variations and behavioral patterns, allowing models to analyze both appearance and motion features. In pedestrian Re-ID, video-based methods leverage motion trajectories and behavioral cues to improve cross-camera accuracy. When combined with temporal models like Long Short-Term Memory (LSTM) networks, video sequence analysis enhances performance in complex scenarios [10]. Similarly, temporal information in vehicle and animal Re-ID helps models adapt to dynamic environments by understanding target movements [11].

Beyond traditional deep learning methods, Generative Adversarial Networks (GANs) [12] have demonstrated immense potential in Re-ID. By generating realistic synthetic data, GANs expand training datasets and improve model generalization. In pedestrian, vehicle, and animal Re-ID, GANs synthesize images under varying lighting conditions and viewpoints, enabling models to handle real-world environmental changes. For instance, GANs enhance pedestrian Re-ID accuracy across diverse poses, lighting, and backgrounds [13]. Likewise, synthetic data generation for vehicles and animals improves system robustness against environmental variations [14].

These advancements collectively drive the evolution of Re-ID systems, enabling reliable identification across complex, real-world scenarios.

B. Multi-Object Tracking

Multi-Object Tracking (MOT) is a critical task in computer vision, aiming to simultaneously track multiple targets in a video while assigning each a unique identifier. With the rapid development of deep learning and convolutional neural network (CNN) technologies, MOT plays a vital role in addressing target detection and tracking challenges in complex scenarios. In recent years, the adoption of **Tracking-by-Detection (TBD)** methods, joint optimization of detection and tracking, and attention mechanisms has led to significant progress in tracking pedestrians, vehicles, and animals. These advancements demonstrate strong potential in handling dynamic and cluttered environments [15].

Tracking-by-Detection (TBD), one of the mainstream approaches in MOT, decouples target detection and tracking. It first detects targets in video frames and then matches and tracks them based on detection results. This framework offers flexibility by allowing independent optimization of detection and tracking modules. For pedestrian tracking, DeepSORT [16] is a classic TBD-based method. It integrates pedestrian detection with deep feature extraction, leveraging deep neural networks to extract appearance features and the Hungarian algorithm for target matching. This enables robust tracking accuracy in challenging scenarios involving occlusions or target intersections. Similarly, TBD excels in vehicle tracking. Combining YOLOv4 with DeepSORT, systems efficiently handle multi-vehicle tracking in traffic scenes [17], maintaining reliable performance even in congested, intersecting, or occluded environments. In animal tracking, TBD is widely applied to wildlife monitoring [18], where deep learning models detect and track multiple species, achieving stable individual tracking in complex natural settings.

Despite its broad adoption, TBD faces challenges such as target loss and ID switching in dynamic scenes. To address these, joint detection-and-tracking optimization has emerged. By training end-to-end deep learning models to jointly optimize detection and tracking tasks, this approach mitigates information loss caused by separately trained models, enhancing overall performance. For example, TrackNet [19] is a representative joint optimization model that directly learns spatial and temporal features from video streams. It employs a unified loss function to refine detection and tracking results, significantly improving accuracy and robustness. In vehicle tracking, DeepMOT [20] integrates YOLOv3 detections with LSTM-based temporal modeling, reducing trajectory matching errors and boosting accuracy in multi-vehicle scenarios. For animal tracking, joint optimization methods simultaneously process multi-animal feature extraction and temporal dependencies, improving long-term tracking stability, especially in complex natural environments.

Attention mechanisms have also gained traction in MOT, particularly for addressing occlusions, cluttered backgrounds, and inter-target interference. By focusing models on critical target information, attention enhances tracking stability and accuracy. In pedestrian tracking, integrating attention

modules into DeepSORT enables automatic adjustment of focus based on key frame regions, effectively resolving occlusion and intersection issues in crowded scenes. This

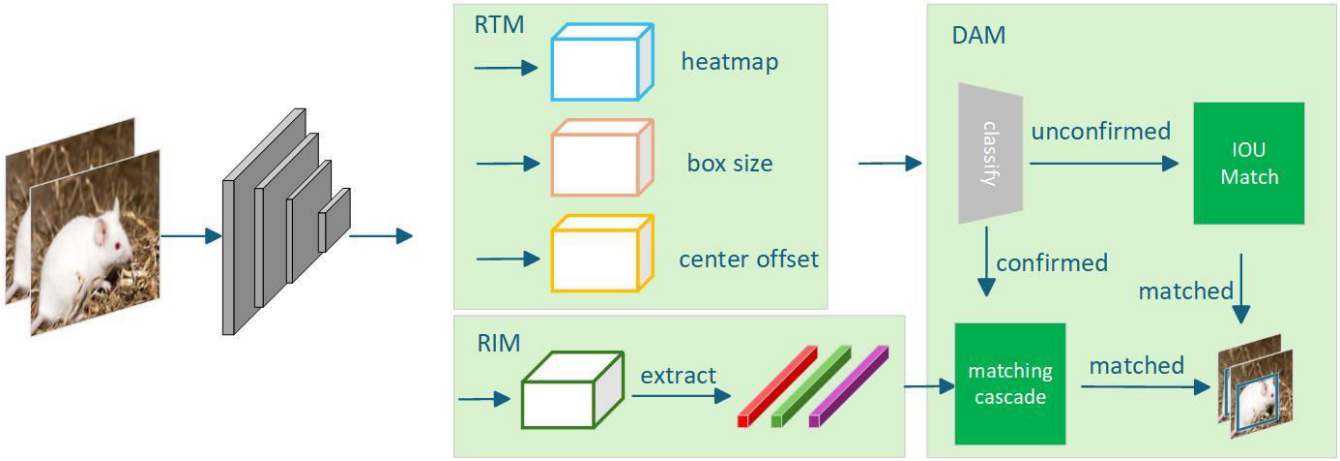

Fig 1 Overall Framework Diagram of the Re - identification - Enhanced Multi - object Tracking Model for Rodents

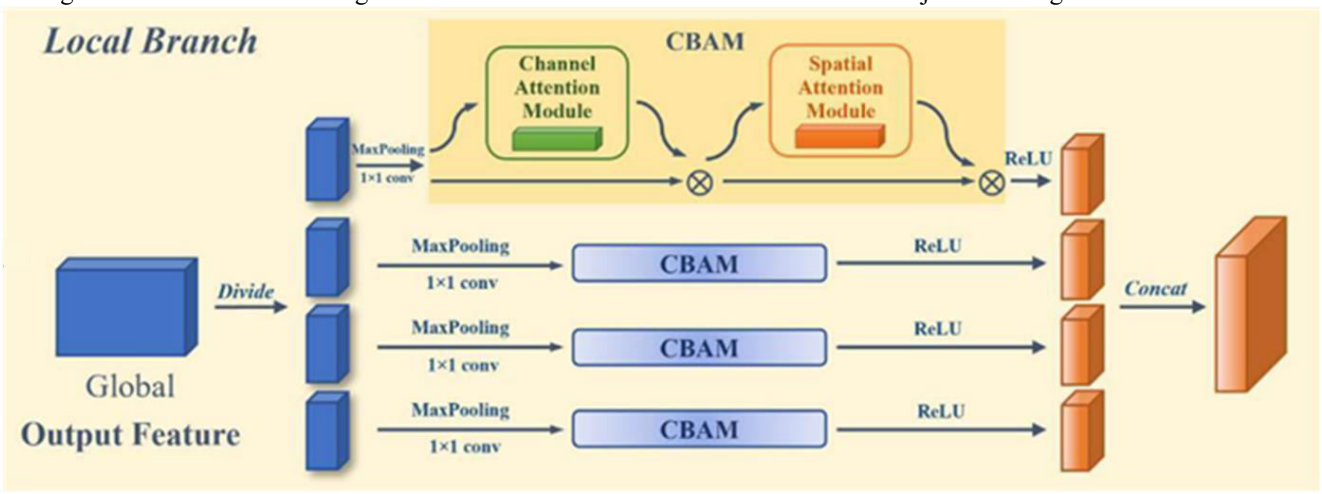

Fig 2 Local Attention Enhancement Module Based on Convolutional Block Attention Module

allows models to prioritize features like heads or facial characteristics, improving precision. For vehicle tracking,Attention-based Recurrent Neural Networks (ARNN) [21] leverage attention to identify and emphasize vehicle-specific details, minimizing mutual interference in complex traffic scenarios and ensuring long-term tracking consistency. In animal tracking, attention mechanisms excel by highlighting distinctive features such as heads, tails, or body shapes, reducing confusion caused by high inter-individual similarity and background noise. These innovations collectively advance MOT systems, enabling robust and accurate tracking across diverse applications, from urban surveillance to ecological studies.

I. MATH

The overall framework of the proposed re-identification enhanced multi-object tracking model for rodents is illustrated in Fig 1. It employs an anchor-free object detection method (YOLOX) to estimate the target center and location on high-resolution feature maps. Additionally, a parallel branch is incorporated to estimate pixel-level Re-ID features, which are used to predict the target's ID.

A. Rodent Detection Module (RTM)

Our detection branch is built upon YOLOX, though other anchor-free methods can also be utilized. We briefly describe the methodology to ensure the self-contained nature of this work. Specifically, three parallel heads are attached to the RTM (Realtime Target Model) to estimate the heatmap, object center offsets, and bounding box sizes, respectively. Each head is implemented by applying a 3×3 convolutional layer (256 channels) to the RTM's output features, followed by a 1×1 convolutional layer to generate the final predictions.

Heatmap head estimates the locations of object centers using a heatmap-based representation, the de facto standard for ground-truth localization tasks. The heatmap has dimensions of 1×H×W. If a position in the heatmap aligns with the collapsed center of a ground-truth object, the response at that position is expected to be 1. The response decays exponentially with the distance between the heatmap position and the target center.

For each ground-truth (GT) box in the image, we compute the object center. Its position on the feature map is obtained by dividing by the stride ss. The heatmap response at position (x,y) is calculated as:

$$M_{xy} = \sum_{i=1}^{N} exp^{\frac{(x-\tilde{c}_x^i)^2+(y-\tilde{c}_y^i)^2}{2\sigma_c^2}} \quad (1)$$

where NN denotes the number of objects in the image, and σ c is the standard deviation. The loss function is defined as pixel-wise logistic regression with focal loss:

$$L_{heat} = -\frac{1}{N}\sum_{xy}\begin{cases}(1-\widehat{M}_{xy})^{\alpha}\log(\widehat{M}_{xy}) & M_{xy}=1\\ (1-M_{xy})^{\beta}(\widehat{M}_{xy})^{\alpha}\log(1-\widehat{M}_{xy}) & otherwise\end{cases} \quad (2)$$

where M^ is the predicted heatmap, and α, β are predefined hyperparameters for the focal loss.

The detection box offset head aims to locate objects more precisely. Since the stride of the final feature map is four, it will introduce a quantization error of up to four pixels. This branch estimates the continuous offset of each pixel relative to the center of the object to mitigate the impact of downsampling. The detection box size head is responsible for estimating the height and width of the target box at each position. Denote the outputs of the size head and the offset head as S and O respectively.For each ground truth (GT) box in the image, we calculate its size. Denote the estimated size and offset at the corresponding position s as o and respectively. Then, we compute the loss for the two heads:

$$L_{box} = \sum_{i=1}^{N}\left\|o^i-\hat{o}^i\right\|_1 + \lambda_s\left\|s^i-\hat{s}^i\right\|_1 \quad (3)$$

among them, λ is the weighting parameter and is set to 0.1 as the original input.

B. *Re-Identification Module (RIM)*

Based on the feature tensor obtained from the YOLOX Backbone, a dual - branch structure is constructed to achieve the requirements of rodent position localization in images and fine - grained re - identification. We connect the two branches in series. The feature tensor obtained from the output of the global branch is sent to the local branch for further feature extraction and fusion. Finally, the feature tensors of the global branch output and the local branch output, $F_{Global}\in R^{1024\times1\times1}$and $F_{Local}\in R^{1024\times1\times1}$, are obtained, and they are concatenated and sent to the classifier layer to generate the re - identification feature vector.

Specifically, we define $C\in R^{C\times H\times W}$ to represent the features output by each layer of the YOLOX Backbone, where H×W corresponds to the spatial dimensions of the feature map, and C represents the number of channels. We design a multi - scale feature fusion connection strategy. We use the features ,$C_3\in R^{256\times80\times80}$, $C_4\in R^{512\times40\times40}$, $C_5\in R^{1024\times20\times20}$ extracted from layers 3 to 5 in the Backbone as the input features. Based on the H×W dimensions of the Neck layer, C_4 and C_5 are successively up - sampled, subjected to convolutional feature extraction, fusion, concatenation, and ReLU activation to complete the feature connection and fusion at each stage. Through continuous mapping from low - level features to high - level features, we obtain , $P_3\in R^{256\times80\times80}$, $P_4\in R^{512\times40\times40}$, $P_5\in R^{1024\times20\times20}$, . The features obtained from each layer are subjected to average pooling to obtain the complete global multi - scale feature $F_{Global}\in R^{1024\times1\times1}$, which integrates key information at different levels and can better capture small but unique features such as the whisker texture, eye details, and subtle color differences of the fur of mice.

We adopt a Local Attention Enhancement Module (LAEM) based on the Convolutional Block Attention Module (CBAM) to enhance the feature extraction performance of multiple local blocks in the local branch. CBAM is an attention mechanism module used to enhance the performance of convolutional neural networks. It improves the model's perception ability by introducing a hybrid attention of channel attention and spatial attention. Channel attention helps to enhance the feature representation of different channels, while spatial attention helps to extract key information at different positions in space.

The local branch of this network differs from the global branch in terms of fusing features of different scales. The local branch further enhances and extracts features in different ranges through horizontal occlusion, which helps to extract and optimize local details such as the fur texture and stripes of rodents, improving the accuracy of re - identification.

We divide the global feature obtained from the global branch into 4 blocks from left to right, and each block is a local feature block $\in R^{1024\times20\times5}$. In the local branch, we perform adaptive Max - pooling and 1×1 convolutional operations on each local block feature to obtain a reduced local feature of 256. Then, we feed each simplified feature block into the CBAM. Channel and spatial attention feature enhancement is applied to the local block features to enhance the dual - domain feature representation and enhance important features such as local stripes after block segmentation, obtaining $\{L1, L2, L3, L4\}\in R^{256\times1\times1}$. Finally, the 4 feature blocks obtained through attention enhancement are activated using the activation function ReLU, and then concatenated to obtain the final feature output $F_{Local}\in R^{1024\times1\times1}$ of the local branch.

C. *Data Association Module (DAM).*

We follow MOTDT [22] and employ a hierarchical online data association method. First, we initialize multiple trajectories based on the detected bounding boxes in the first frame. Subsequently, in the following frames, we use a two - stage matching strategy to link the detected bounding boxes to the existing trajectories. In the first stage, we utilize the Kalman filter and re - identification features to obtain the initial tracking results. Specifically, the Kalman filter is used to predict the trajectory positions in the next frame, and we calculate the Mahalanobis distance D_m between the predicted bounding box after Deep - SORT and the detected bounding box. We fuse the Mahalanobis distance with the cosine distance calculated based on the re - identification features: $D_r = D_r+(1-\lambda)D_m$, where λ is a weighting parameter, which is set to 0.98 in our experiments. According to JDE [95], if the Mahalanobis distance is greater than the threshold, we set it to infinity to avoid getting trajectories with large motions. We use the Hungarian algorithm with a matching threshold of 0.4 to complete the first - stage matching. In the second stage, for the unmatched detections and trajectories, we attempt to match them according to the overlap between their bounding boxes. Specifically, we set a matching threshold of 0.5. We update the appearance features of the trajectories at each time step to handle appearance changes. Finally, we initialize the unmatched detections as new trajectories and retain the unmatched trajectories for 30 frames in case they reappear in the future. sequence is accurately annotated, providing the bounding boxes, movement trajectories, and behavior labels of individual animals, which is suitable for tasks such as object detection, multi - object tracking, and behavior

analysis.

Experiments were conducted using an Intel(R) Xeon(R) E5 - 2603V4 CPU processor and an NVIDIA TITAN Xp GPU running Ubuntu 22.04.2. YOLOX was trained with an input size of 640×640, a batch size of 16. The Stochastic Gradient Descent (SGD) optimizer was employed, with a momentum of 0.937, a learning rate of 0.01, and a weight decay of 0.0005. To prevent overfitting, an early - stopping mechanism was used. The number of epochs was set to 300, and the patience was set to 30.

Table 1 Comparison Results between YOLOXs and the Proposed Tracking Framework

Methods	HOTA	MOTA	IDF1	MT	ML	IDs
YOLOXs	56.8	60.8	58.5	324	329	1976
Our Proposed	63.3	65.3	67.1	367	317	1309

Table 2 Results of the Performance Comparison Experiments

Model	Params(M)	GFLOPs	HOTA	MOTA	IDF1
YOLOv5s	7015519	15.8	56.8	55.1	63.9
YOLOv7-tiny	6010302	13	53.3	52.7	62.5
YOLOv8s	11126358	28.4	58.3	58.3	66
Our Proposed	8940000	8.1	63.3	65.3	67.1

Table 3 Ablation Experiment Results of the Proposed Model

Methods	HOTA	MOTA	IDF1	MT	ML	IDs
YOLOXs	58.8	63.8	58.5	333	527	2613
RTM	53.3	59.3	47.4	327	534	2215
RTM+DAM	54.5	64.5	59.6	351	428	1973
RTM+RIM+DAM	63.3	65.3	67.1	367	317	1309

Table 4 Ablation Experiment Results of the Multi - object Tracking Model for Rodents

Methods	HOTA	MOTA	IDF1	MT	ML	IDs
ResNet-34	52.5	55.8	58.5	333	327	1613
ResNet-50	58.8	57.3	59.4	322	334	1654
DLA-34	53.3	61.4	63.9	339	329	1473
HarDNet-85	54.5	58.8	66.5	337	327	1542
Our Proposed	63.3	65.3	67.1	367	317	1309

We adopted a single - image training method for training on the image - level object detection dataset. Instead of using two simulated consecutive frames as input, we only used a single image as input. We assigned a unique identifier to each bounding box, treating each object instance in the dataset as a separate class. Different transformations were applied to the entire image, including HSV enhancement, rotation, scaling, translation, and shearing. The single - image training method has significant empirical value. The training of re - identification features further enhanced the associative ability of the tracker. Secondly, it can be fine - tuned on other datasets to further improve the final performance.

A. Evaluation Metrics

The performance of the proposed algorithm was evaluated using standard MOT metrics, including MOTA (Multiple Object Tracking Accuracy), IDF1 (Identity F1 - score), and HOTA (Higher - Order Tracking Accuracy). These metrics measure the accuracy of the tracking algorithm in terms of detection, identity maintenance, and overall tracking performance.

B. Experimental Results

The experimental results of evaluating the multi - object tracking algorithm in this chapter based on video standards are shown in Table 4 - 1. Compared with the object tracking method using the general - purpose object detector YOLOX, the MOTA of the proposed re - identification - enhanced multi - object tracking algorithm for rodents increased to 65.3%, and the IDF1 increased from 58.5% to 67.1%. Meanwhile, as can be seen from Table 1, RIM improved the tracking accuracy of rodents. It can be seen that the re-identification features can effectively supplement the global features and solve the problem of a high identity - switching rate in complex backgrounds.

To comprehensively evaluate the performance of the proposed tracking framework, this chapter selects several currently leading detection frameworks in the YOLO series for detailed comparative experiments. The results are shown in Table 2. In the experiments, the image input size of all models is set to 640×640 pixels. This is considered for the lightweight characteristics of the models, while also ensuring sufficient input information. Given the small scale of the animal tracking dataset used in this study, directly applying pre - trained weights may lead to model overfitting. Therefore, none of the models involved in the comparison are trained with pre - trained weights to ensure the fairness of the experimental results. In addition, an early - stopping mechanism is adopted during the experiment to prevent the model from prematurely stopping learning on the validation set, so as to select the best performance of each model. Through this strategy, it can be ensured that the model can not only learn sufficiently during the training process, but also avoid overfitting.

It is easy to find that, despite significant differences in the number of parameters and computational complexity, the tracking and detection model proposed in this chapter demonstrates superior performance in two key evaluation metrics, MOTA and IDF1.

C. Ablation Experiments

An ablation study was conducted to evaluate the performance of each proposed module. The RTM, RIM, and DAM were detected respectively. The results are shown in Table 3. The three modules in the method of this chapter can

all improve the model performance. In particular, the increase in the MOTA value indicates a significant improvement in the overall tracking performance of the model.

Meanwhile, we compared several backbones, namely ResNet, DLA, and HarDNet. For a fair comparison, the remaining factors of these methods were controlled to be the same. In particular, the stride of the final feature map of all methods was 4. We performed three up - sampling operations on ResNet to obtain the feature map of the fourth step. We divided these backbone networks into two categories: one without multi - layer fusion and the other with multi - layer fusion. The results are shown in Table 4.

From the comparison results, we found that blindly using a larger network did not significantly improve the overall tracking results measured by MOTA. In particular, the quality of the re - identification function hardly benefited from a larger network. For example, the IDF1 only increased from 58.5% to 59.4%. In addition, the number of ID switches even increased from 613 to 654.

We evaluated different methods for balancing the losses of different tasks, including Uncertainty, GradNorm, and MGDA - UB. We also evaluated the baseline of fixed weights obtained through grid search. We implemented two versions for the Uncertainty - based method. The first is “Uncertainty - task”, which learns two parameters for the detection loss and the re - identification loss respectively. The second is “Uncertainty - branch”, which learns four parameters for the heatmap loss, box size loss, offset loss, and re - identification loss respectively.

II. Conclusion

This paper proposes a novel multi - object rodent tracking algorithm enhanced by re - identification. This algorithm combines motion features and re - identification features to achieve accurate and continuous tracking of rodents in complex environments. Experimental results on the AnimalTrack dataset show a significant improvement in tracking accuracy and identity retention rate. The proposed algorithm provides a robust solution for rodent behavior analysis and has potential applications in neuroscience, ethology, and drug research.

Acknowledgment

This work is supported by the Humanities and Social Sciences Youth Foundation of Ministry of Education of China (No. 22YJC870018), the Open Project of Tianjin Key Laboratory of Autonomous Intelligence Technology and Systems (No. TJKL-AITS-20241006, No. TJKL-AITS-20241004), the Science and Technology Development Fund of Tianjin Education Commission for Higher Education (No. 2020KJ112, No. KYQD1817) and Science and Technology Project of Putian City (No. 2021R4001-10), and Haihe Lab. of Information Technology Application Innovation (No. 22HHXCJC00002).

References

[1] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning Deep Features for Discriminative Localization," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2921-2929, doi: 10.1109/CVPR.2016.319.

[2] D. Yi, Z. Lei, S. Liao and S. Z. Li, "Deep Metric Learning for Person Re-identification," 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 2014, pp. 34-39, doi: 10.1109/ICPR.2014.16.

[3] J. Ramod, P. Shrivastav, R. Shetty, V. Nimbalkar and L. Ragha, "Signature Authentication Verification using Siamese Network," 2023 6th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2023, pp. 558-562, doi: 10.1109/ICAST59062.2023.10454931.

[4] Hoffer, E., Ailon, N. (2015). Deep Metric Learning Using Triplet Network. In: Feragen, A., Pelillo, M., Loog, M. (eds) Similarity-Based Pattern Recognition. SIMBAD 2015. Lecture Notes in Computer Science(), vol 9370. Springer, Cham. https://doi.org/10.1007/978-3-319-24261-3_7

[5] R. Hadsell, S. Chopra and Y. LeCun, "Dimensionality Reduction by Learning an Invariant Mapping," 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 2006, pp. 1735-1742, doi: 10.1109/CVPR.2006.100.

[6] Bruggisser, Sebastian, et al. "Baryon asymmetry from a composite Higgs boson." Physical review letters 121.13 (2018): 131801.

[7] Sun, Yifan, et al. "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)." Proceedings of the European conference on computer vision (ECCV). 2018.

[8] Khan, Sultan Daud, and Habib Ullah. "A survey of advances in vision-based vehicle re-identification." Computer Vision and Image Understanding 182 (2019): 50-63.

[9] Xu, Zeyu, et al. "A review of deep learning techniques for detecting animals in aerial and satellite images." International Journal of Applied Earth Observation and Geoinformation 128 (2024): 103732.

[10] C. Eom, G. Lee, J. Lee and B. Ham, "Video-based Person Re-identification with Spatial and Temporal Memory Networks," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 12016-12025, doi: 10.1109/ICCV48922.2021.01182.

[11] H. -G. Kim, Y. Na, H. -W. Joe, Y. -H. Moon and Y. -J. Cho, "Vehicle Re-identification with Spatio-temporal Information," 2023 14th International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, Republic of, 2023, pp. 1825-1827, doi: 10.1109/ICTC58733.2023.10392420.

[12] M. Krichen, "Generative Adversarial Networks," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-7, doi: 10.1109/ICCCNT56998.2023.10306417.

[13] Zheng Z, Zheng L, Yang Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]//Proceedings of the IEEE international conference on computer vision. 2017: 3754-3762.

[14] Z. Zhou et al., "GAN-Siamese Network for Cross-Domain Vehicle Re-Identification in Intelligent Transport Systems," in IEEE Transactions on Network Science and Engineering, vol. 10, no. 5, pp. 2779-2790, 1 Sept.-Oct. 2023, doi: 10.1109/TNSE.2022.3199919.

[15] Ciaparrone G, Sánchez F L, Tabik S, et al. Deep learning in video multi-object tracking: A survey[J]. Neurocomputing, 2020, 381: 61-88.

[16] N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017, pp. 3645-3649, doi: 10.1109/ICIP.2017.8296962.

[17] M. A. Bin Zuraimi and F. H. Kamaru Zaman, "Vehicle Detection and Tracking using YOLO and DeepSORT," 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2021, pp. 23-29, doi: 10.1109/ISCAIE51753.2021.9431784.

[18] Liu Y, Li W, Liu X, et al. Deep learning in multiple animal tracking: A survey[J]. Computers and Electronics in Agriculture, 2024, 224: 109161.

[19] Huang Y C, Liao I N, Chen C H, et al. Tracknet: A deep learning network for tracking high-speed and tiny objects in sports applications[C]//2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2019: 1-8.

[20] Xu Y, Ban Y, Alameda-Pineda X, et al. Deepmot: A differentiable framework for training multiple object trackers[J]. arXiv preprint arXiv:1906.06618, 2019, 10(11).

[21] Qin Y, Song D, Chen H, et al. A dual-stage attention-based recurrent neural network for time series prediction[J]. arXiv preprint arXiv:1704.02971, 2017.

[22] L. Chen, H. Ai, Z. Zhuang and C. Shang, "Real-Time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification," 2018 IEEE International Conference on

Multimedia and Expo (ICME), San Diego, CA, USA, 2018, pp. 1-6,
doi: 10.1109/ICME.2018.8486597.