CO-MOT: Enhancing End-to-End Multi-Object Tracking with Cooperative Label Assignment and Shadow Concept

ICLR 2025

Feng Yan, Weixin Luo, Yujie Zhong, Yiyang Gan, Lin Ma

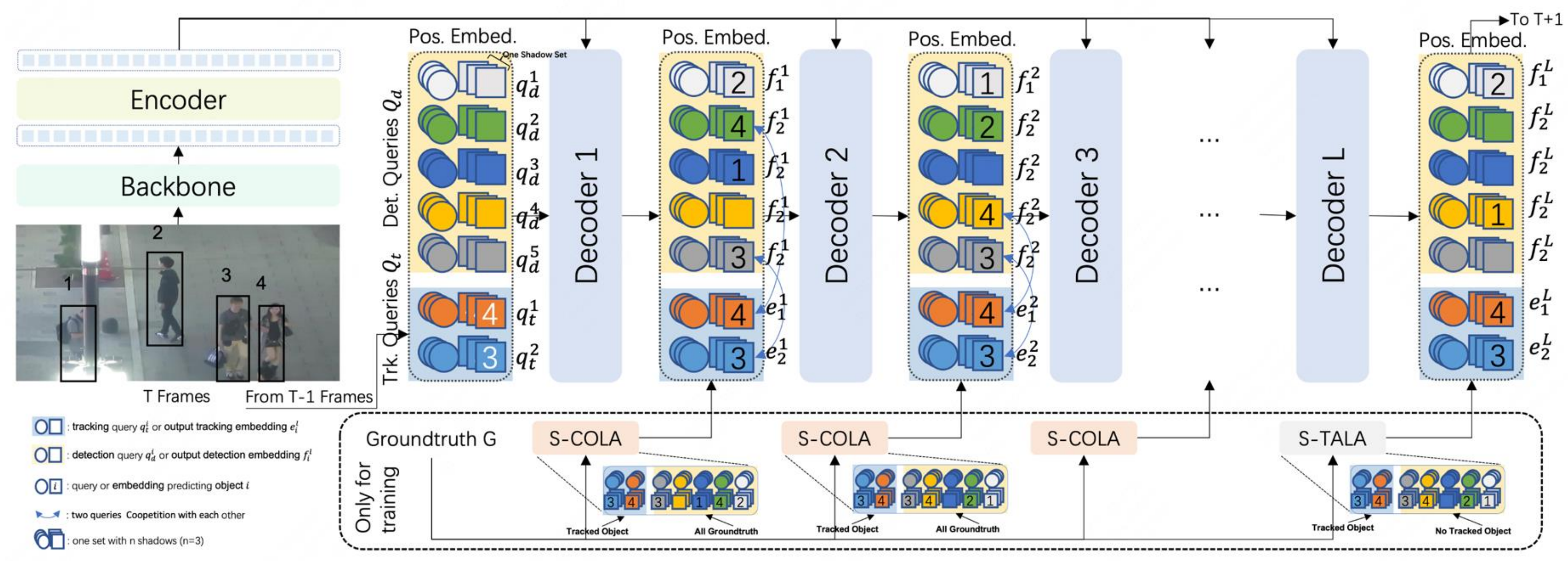
March, 2025

Project: https://github.com/BingfengYan/CO-MOT

Introduction

- **Background**:
  - Traditional MOT tackles tasks separately, achieving optimal solutions for each but lacking global optimization.
  - End-to-end MOT models eliminate pre- and post-processing steps but have not yet surpassed traditional methods.
- **Problem**:
  - Label assignment strategy results in scarce positive samples for detection queries, affecting detection capability.

Method Overview



Components: CNN backbone, deformable encoder, deformable decoder.

Key Components: Role of **COLA and Shadow Set** in the framework.

Detailed Methodology

- **COLA Strategy**:
  - Allows tracked objects to be reassigned to detection queries in intermediate decoders, fostering collaboration and enhancing the representation through self-attention between similar identities.
- **Shadow Queries**:
  - Each query is augmented with shadows that act as counterparts, enabling robust handling of crowded scenes and easing one-to-set optimization.

ICLR 2025

Experimental Results

Table 2: Comparison to existing methods on the DanceTrack test set. "*" and "+" respectively represent the use of DAB-Deformable backbone and joint training with CrowdHuman. For static images in CrowdHuman dataset, we apply random shifts as in CenterTrack to generate video vlips with pseudo tracks.

	Source	HOTA	DetA	AssA	MOTA	IDF1
	Non-End-to-end					
CenterTrack (Zhou et al., 2020)	ECCV'20	41.8	78.1	22.6	86.8	35.7
TransTrack (Sun et al., 2020)	arXiv'20	45.5	75.9	27.5	88.4	45.2
FairMOT (Zhang et al., 2021)	IJCV'21	39.7	66.7	23.8	82.2	40.8
QDTrack (Fischer et al., 2022)	CVPR'21	54.2	80.1	36.8	87.7	50.4
TraDeS (Wu et al., 2021)	CVPR'21	43.3	74.5	25.4	86.2	41.2
ByteTrack (Zhang et al., 2022b)	ECCV'22	47.7	71.0	32.1	89.6	53.9
GTR (Zhou et al., 2022)	CVPR'22	48.0	72.5	31.9	84.7	50.3
MT-IoT^{+} (Yan et al., 2022)	arXiv'22	66.7	84.1	53.0	94.0	70.6
OC-SORT (Cao et al., 2023)	CVPR'23	55.1	80.3	38.3	92.0	54.6
C-BIoU (Yang et al., 2023)	WACV'23	60.6	81.3	45.4	91.6	61.6
MOTRv2^{+} (Zhang et al., 2023)	CVPR'23	69.9	83.0	59.0	91.9	71.7
FineTrack (Ren et al., 2023)	CVPR'23	52.7	72.4	38.5	89.9	59.8
GHOST (Seidenschwarz et al., 2023)	CVPR'23	56.7	81.1	39.8	91.3	57.7
Walker (Segu et al., 2024)	ECCV'24	52.4	36.1	76.5	89.7	55.7
GeneralTrack (Qin et al., 2024)	CVPR'24	59.2	82.0	42.8	91.8	59.7
MotionTrack (Xiao et al., 2024b)	arXiv'24	58.2	81.4	41.7	91.3	58.6
ConfTrack (Jung et al., 2024)	WACV'24	56.1	-	-	89.6	56.2
MambaTrack (Xiao et al., 2024a)	arXiv'24	56.8	80.1	39.8	90.1	57.8
Hybrid-SORT (Yang et al., 2024)	AAAI'24	62.2	-	-	91.6	63.0
UCMCTrack (Yi et al., 2024)	AAAI'24	63.6	-	51.3	88.8	65.0
DiffusionTrack (Luo et al., 2024)	AAAI'24	52.4	82.2	33.5	89.3	47.5
	End-to-end					
MOTR (Zeng et al., 2022)	ECCV'22	54.2	73.5	40.2	79.7	51.5
DNMOT (Fu et al., 2023)	arXiv'23	53.5	-	-	89.1	49.7
MeMOTR (Gao & Wang, 2023)	ICCV'23	63.4	77.0	52.3	85.4	65.5
MeMOTR* (Gao & Wang, 2023)	ICCV'23	68.5	80.5	58.4	89.9	71.2
MOTRv3^{+} (Yu et al., 2023)	arXiv'23	68.3	-	-	91.7	70.1
SUSHI (Cetintas et al., 2023)	CVPR'23	63.3	80.1	50.1	88.7	63.4
MambaTrack+ (Huang et al., 2024)	arXiv'24	56.1	80.8	39.0	90.3	54.9
OuTR (Liu et al., 2024)	arXiv'24	54.5	-	-	88.3	55.7
DiffMOT (Lv et al., 2024)	CVPR'24	62.3	**82.5**	47.2	**92.8**	63.0
ByteSSM (Hu et al., 2024)	arXiv'24	57.7	81.5	41.0	92.2	57.5
CO-MOT	-	65.3	80.1	53.5	89.3	66.5
CO-MOT^{+}	-	**69.4**	82.1	**58.9**	91.2	**71.9**

(a) MOT17 Test Dataset

	HOTA	AssA	MOTA	IDF1
Non-End-to-end				
CenterTrack	52.2	51.0	67.8	64.7
TransTrack	54.1	47.9	74.5	63.9
FairMOT	59.3	58.0	73.7	72.3
QDTrack	63.5	64.5	77.5	78.7
ByteTrack	63.1	62.0	80.3	77.3
OC-SORT	63.2	63.2	78.0	77.5
DiffusionTrack	60.8	58.8	77.9	73.8
MOTRv2	62.0	60.6	78.6	75.0
End-to-end				
TrackFormer	-	-	65.0	63.9
MOTR	57.8	55.7	73.4	68.6
MeMOT	56.9	55.2	72.5	69.0
MeMOTR	58.8	58.4	72.8	71.5
DNMOT	58.0	-	**75.6**	68.1
CO-MOT	**60.1**	**60.6**	72.6	**72.7**

(b) BDD100K Validation Set

	TETA	LocA	AssocA	ClsA
Non-End-to-end				
DeepSORT	48.0	46.4	46.7	51.0
QDTrack	47.8	45.8	48.5	49.2
TETer	50.8	47.2	52.9	52.4
MOTRv2	54.9	49.5	51.9	63.1
End-to-end				
MOTR	50.7	35.8	51.0	-
CO-MOT	**52.8**	**38.7**	**56.2**	**63.6**

(c) MOT20 Test Dataset

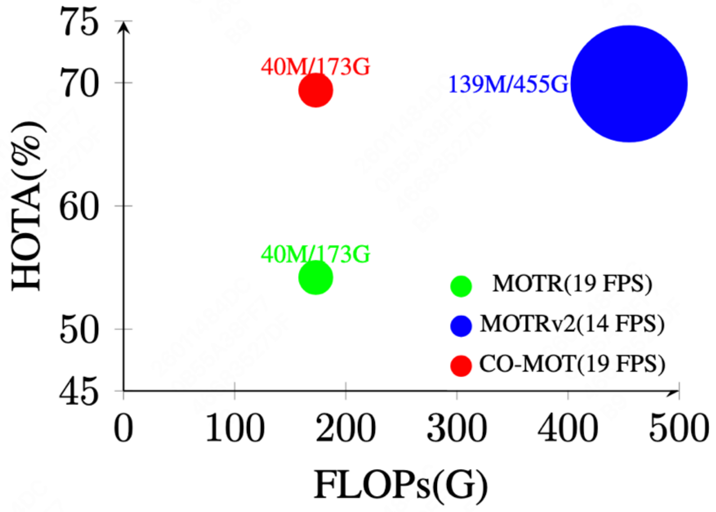
	HOTA	AssA	MOTA	IDF1
End-to-end				
MeMOT	54.1	55.0	63.7	66.1
TrackFormer	54.7	-	68.6	65.7
CO-MOT	57.5	65.7	60.1	70.5

- **Datasets**:
  - Evaluated on DanceTrack, MOT17, BDD100K and MOT20.
- **Performance Comparison**:
  - Use tables to compare CO-MOT with other methods
  - Highlight key metrics: HOTA, DetA, AssA, etc.

Efficiency Comparison

ICLR 2025

- **Resource Utilization**:
  - With only 173G FLOPs and 40M parameters, CO-MOT achieves 69.4% HOTA, comparable to MOTRv2's HOTA but without the extra computational overhead of a separate detector.
- **Inference Speed**:
  - CO-MOT demonstrates a 1.4× faster inference speed compared to MOTRv2, highlighting its deployment efficiency.



Challenges and Summary

- **Challenges**:
  - Dataset size impacts the model's performance, as seen in MOT17's smaller data volume leading to less robust results.
  - Handling of small objects remains a challenge, affecting detection and tracking accuracy.
- **Summary**:
  - CO-MOT significantly improves the performance of end-to-end Transformer models in multi-object tracking.
  - Acts as a plug-in solution to advance end-to-end MOT research.

Question & Answer
Thank You!