杭州電子科技大学
HANGZHOU DIANZI UNIVERSITY

ICIP 2024 Workshop 11

City Traffic Aware Multi-Target Tracking Prediction With Multi-Camera

Presenter： **Jianhui Zhang**

Kanglei Peng, Tuo Dong, Wanqin Zhang, Jianhui Zhang*

Hangzhou Dianzi University, Hangzhou China

计算机学院
COMPUTER & SOFTWARE SCHOOL

ICIP 2024
2024.10.30， Abu Dhabi

CONTENTS

01 Background
02 Introduction
03 Related Work
04 Method
05 Experiment
06 Conclusion

BACKGROUND

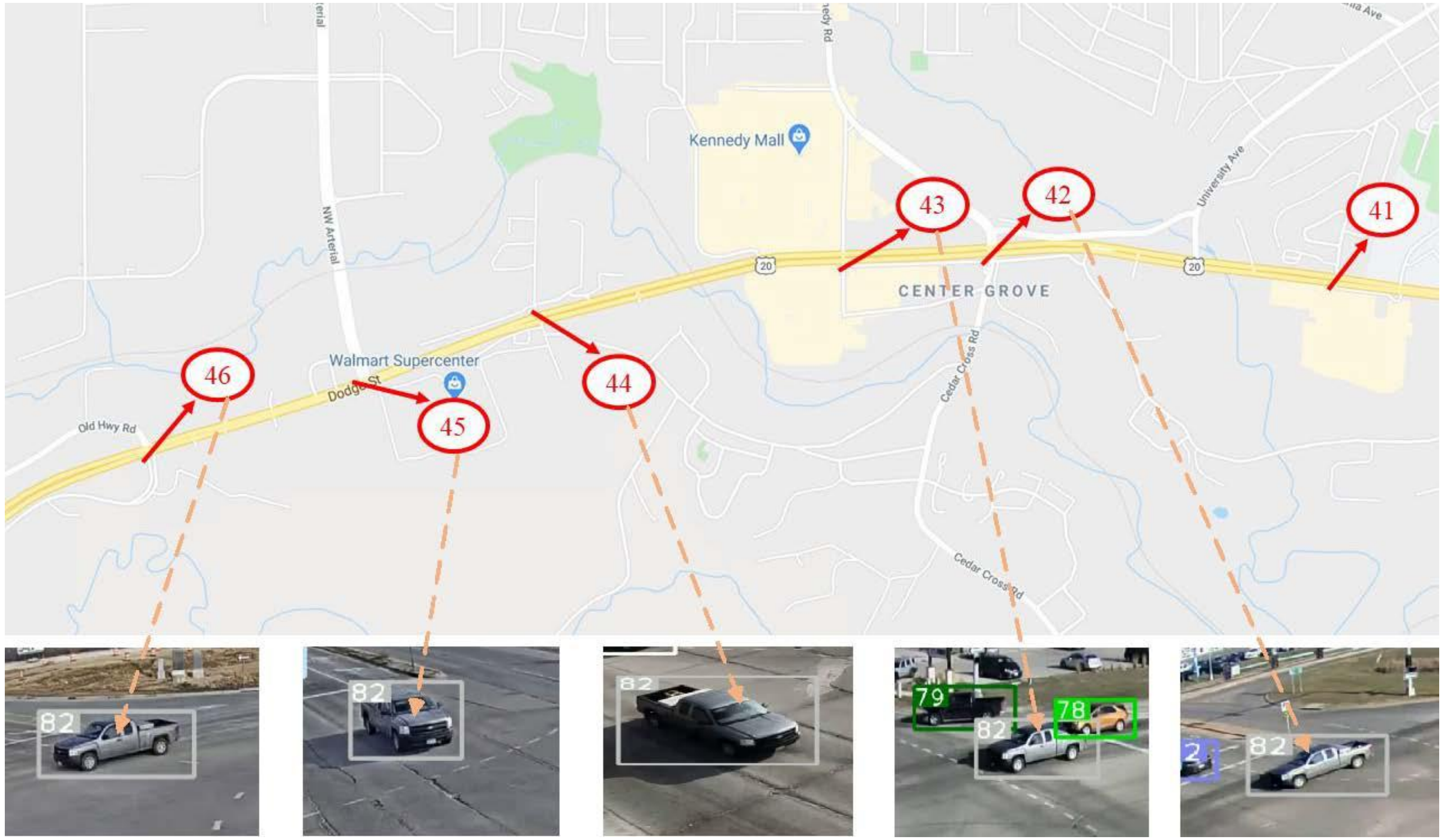
Several cameras on each pole

Intersection has several camera poles

1. The current MCMT research mainly focuses on data association and trajectory detection, which mostly use pre selected small-scale datasets for experiments and cannot effectively cope with the huge amount of data in urban areas.

2. Most existing researches ignore environmental factors such as traffic flow, as well as the spatiotemporal correlations between cameras, especially at the urban scale

INTRODUCTION



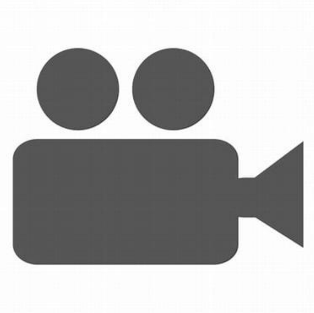
Demand and challenging:

- Large area and long time object track
- Tracking one object from large amount of neighboring ones
- Each location has huge amount of recorded data (raw data).

Need an object tracking model (OTM) based on graph convolutional networks (GCN) to capture spatiotemporal correlations between cameras, and a raw data selection algorithm (RDM) to reduce the amount of raw data that needs to be processed by predicting the velocity and position of the target object.

INTRODUCTION

Advantages
- Reduce the amount of raw data
- Improve the accuracy of target tracking



Multi-Camera Video

OTM+RDM →

Multi-Object Tracking

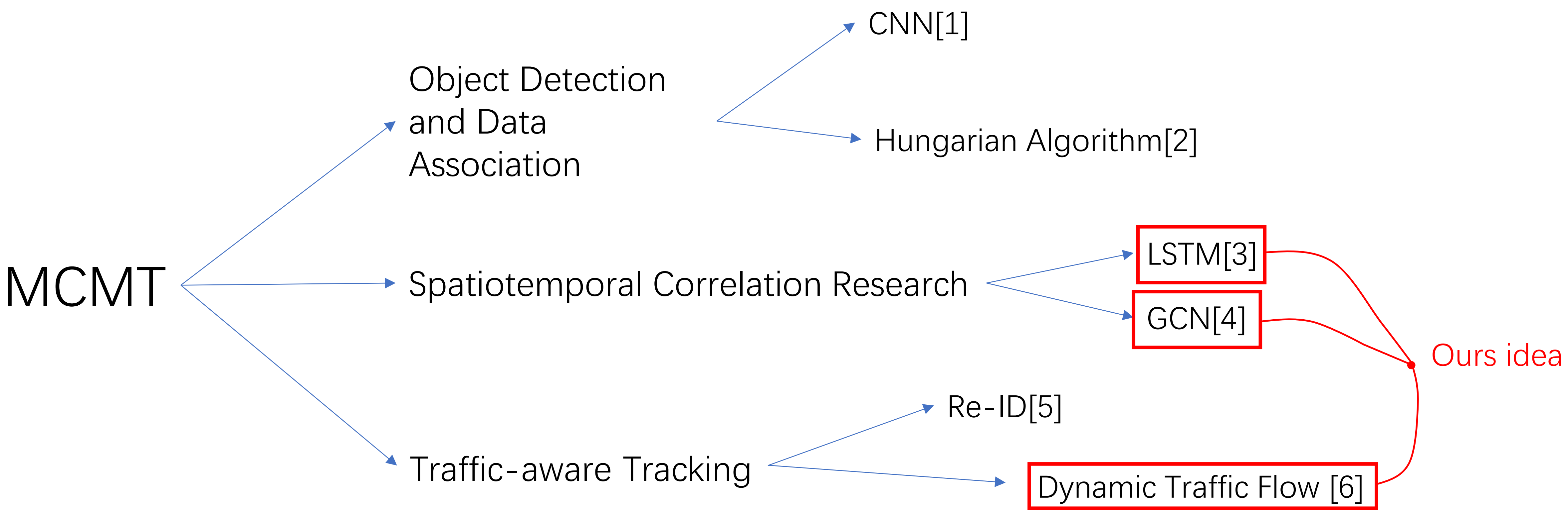
Apply →

Intelligent Transportation

Autonomous Driving

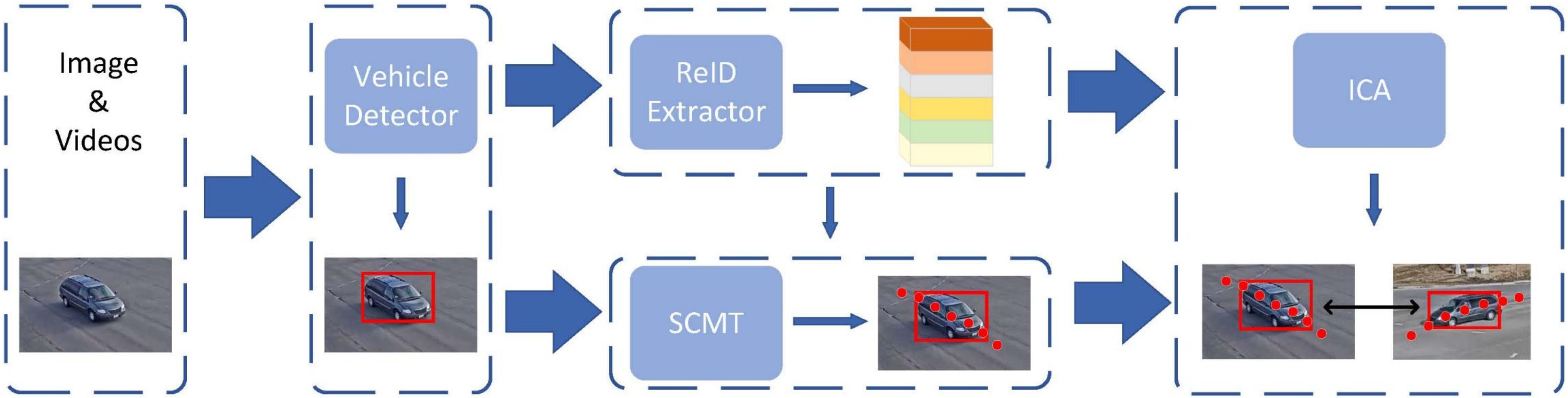
Smart City

RELATED WORK



[1] X. Zhang, Y. Feng, P. Angeloudis, and Y. Demiris, “Monocular visual traffic surveillance: A review,” IEEE Transactions on Intelligent Trans portation Systems, vol. Early Access, pp. 1–18, 2022.
[2] Y. He, X. Wei, X. Hong, W. Shi, and Y. Gong, “Multi-target multi camera tracking by tracklet-to-target assignment,” IEEE Transactions on Image Processing, vol. 29, pp. 5191–5205, 2020.
[3] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social lstm: Human trajectory prediction in crowded spaces,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 961–971.
[4] Lee J, Jeong M, Ko B C. Graph convolution neural network-based data association for online multi-object tracking[J]. IEEE Access, 2021, 9: 114535-114546.
[5] M. Ye, J. Shen, G. Lin, T. Xiang, and S. Hoi, “Deep learning for person re-identification: A survey and outlook,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PP, no. 99, pp. 1–1, 2021.
[6] D. M. Jim´ enez-Bravo, ´ Alvaro Lozano Murciego, A. Sales Mendes, H. S´anchez San Bl´as, and J. Bajo, “Multi-object tracking in traffic environments: A systematic literature review,” Elsevier Neurocomputing, vol. 494, pp. 43–55, 2022.

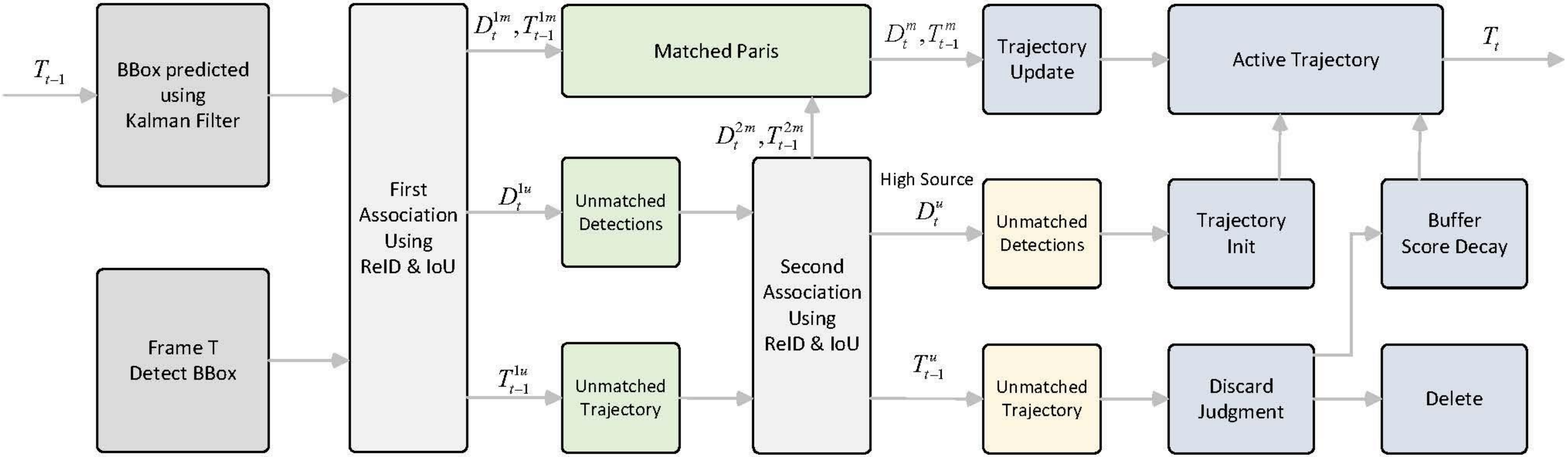
Method-basics



MCMT: A system designed to tackle the MCMT task typically consists of four sub-modules: **Object detection**, **Re-Identification (ReID)**, **Single-Camera Multi-Target tracking (SCMT)**, and **Inter-Camera Association (ICA)**.

Object detection outputs vehicle coordinates and categories in frames and extracts vehicle features by ReID. Then, based on the vehicle location and learned features, the **SCMT** module generates candidate trajectories for every single camera. At last, the **ICA** module matches these candidate trajectories across different cameras to associate targets with global identities.

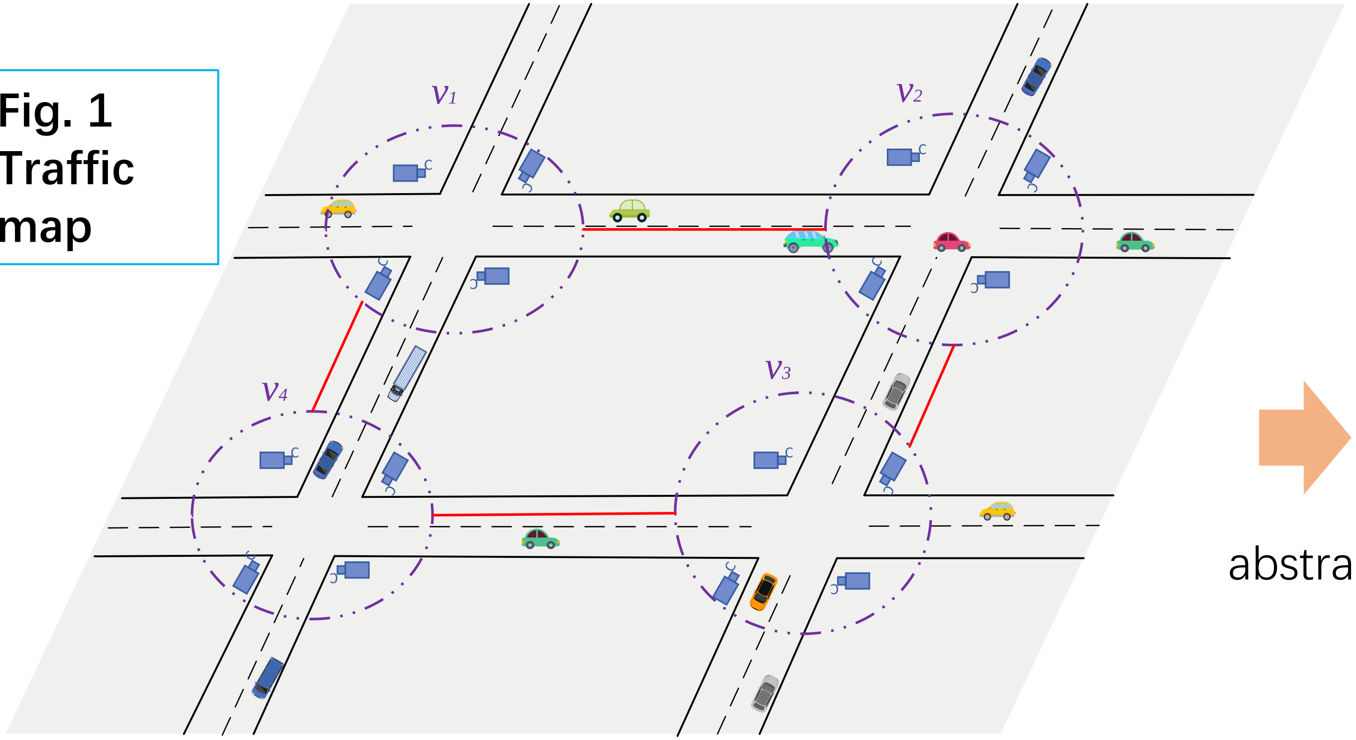
Method-basics



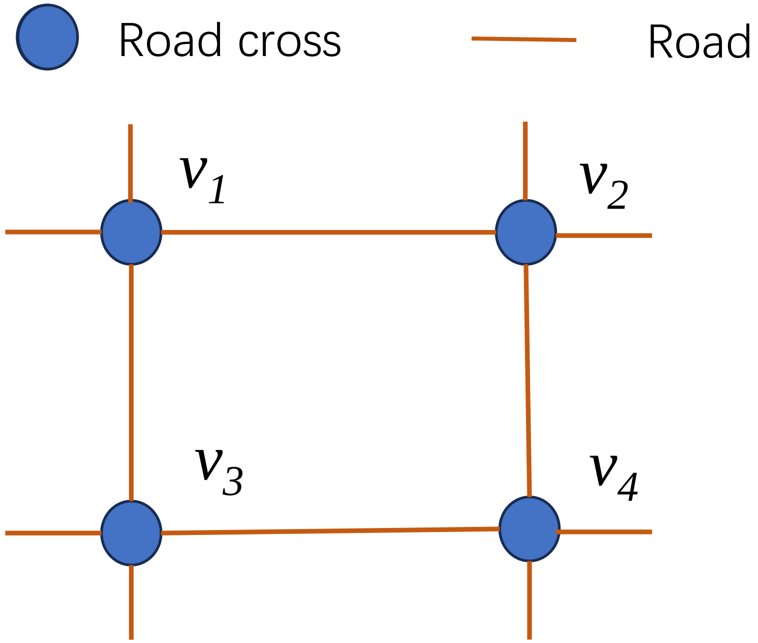
Data Association: Data correlation is the core part of MCMT, Data association requires associating targets between different frames under the same camera, and associating the same target across different cameras to form a complete trajectory. It usually uses similarity metrics such as IoU, distance measurement, and appearance features, and matches them through algorithms such as greedy and Hungarian algorithms.

METHOD

Fig. 1 Traffic map



abstracting



Plane graph

Idea: explore the information behind the traffic map with the graph

Node Connectivity Graph(NCG) → How many input and output, for example traffic lane?

Neighborhood Graph(NNG) → How affects the current node?

Node Type correlation Graph(NTG) → How similar to the others the current node is?

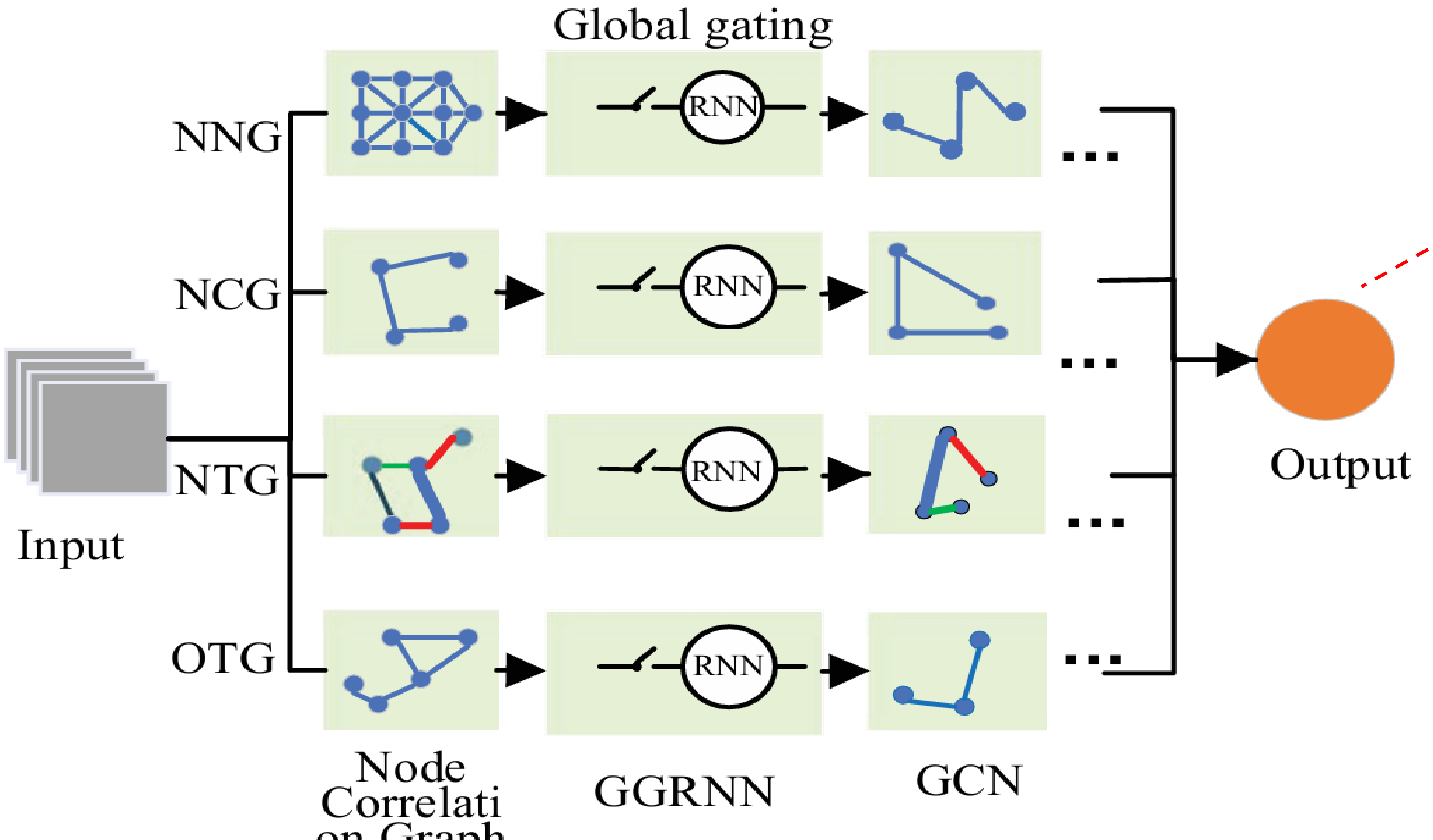
Object Tracking correlation Graph(OTG) → How similar to the others the current object is sensed?

METHOD

Multi-graph convolution model including the graphs:

Neighborhood Graph(NNG), Node Connectivity Graph(NCG), Node Type correlation Graph(NTG)

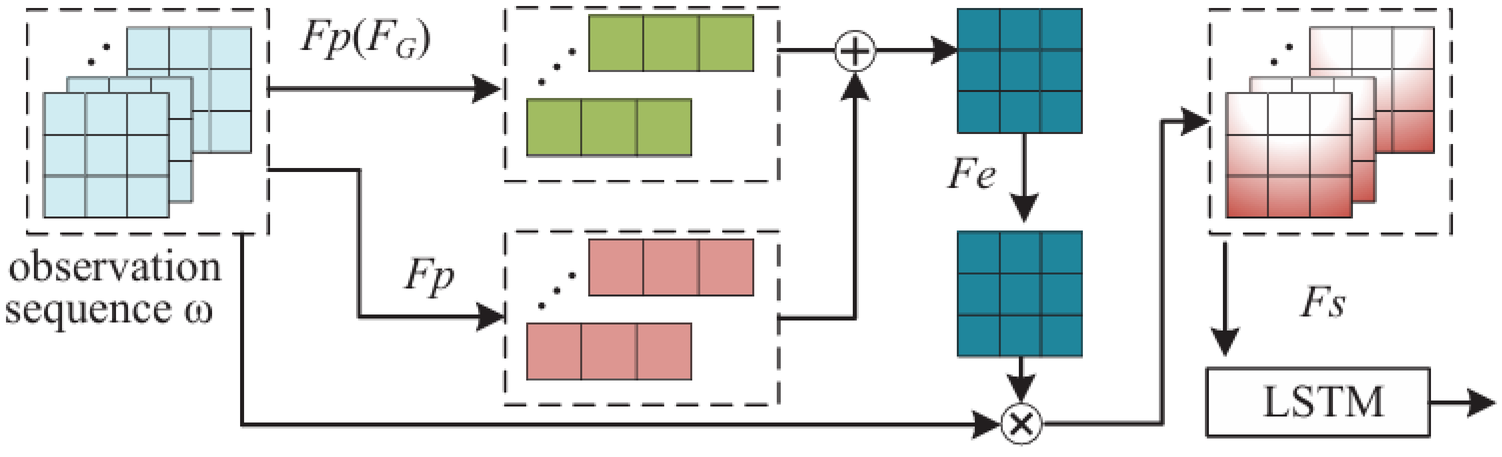
Object Tracking correlation Graph(OTG)



Output is the feature information from the traffic map.

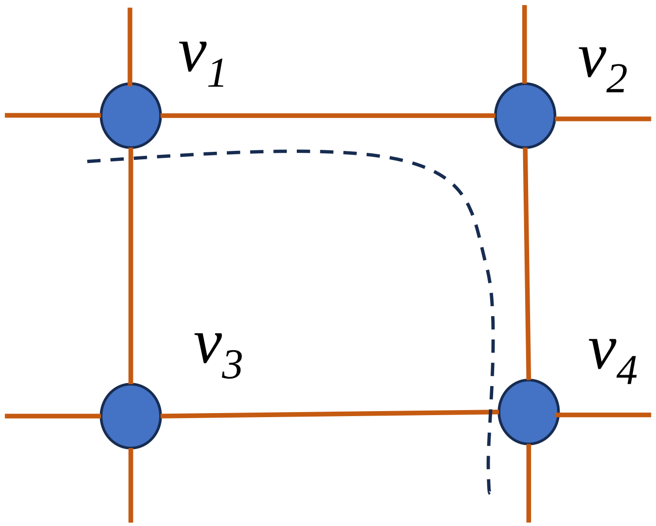
METHOD

Each single node have full time object recording and how to reduce it?



Associating one node with others.

Track object with its feature
based on the node association forward and backward.



Time correlation model is devised to gather observational data for each node depicted in Fig.1 across different time intervals.

Data set

CityFlowV2 is a dataset comprising 3.58 hours (215.03 minutes) of videos collected from 46 cameras spanning 16 inter sections. The distance between the two furthest simultaneous cameras is 4 km. The dataset covers a diverse set of location types, including intersections, stretches of roadways, and highways.

EXPERIMENT

We use the IDF1 metric to evaluate our proposed method

$$IDF1 = \frac{2 \times IDTP}{2 \times IDTP + IDFP + IDFN}$$

IDF1 measures the ratio of correctly identified detections over the average number of ground-truth and computed detections.

Our proposed method achieved an IDF1 score of 0.8275 on the CityFlow dataset

Team and paper	IDF1
FourBeauties [1]	0.8166
SKKU [2]	0.8129
UCAS [3]	0.8095
HCMIU [4]	0.7255
ours	**0.8275**

[1] T. T. Nguyen, H. H. Nguyen, M. Sartipi, and M. Fisichella, "Multi-vehicle multi-camera tracking with graph-based tracklet features," IEEE Transactions on Multimedia, 2023.
[2] D. N.-N. Tran, L. H. Pham, H.-J. Jeon, H.-H. Nguyen, H.-M. Jeon, T. H.-P. Tran, and J. W. Jeon, "A robust traffic-aware city-scale multi-camera vehicle tracking of vehicles," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 3150–3159
[3] C. Liu, Y. Zhang, H. Luo, J. Tang, W. Chen, X. Xu, F. Wang, H. Li, and Y.-D. Shen, "City-scale multi camera vehicle tracking guided by crossroad zones," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4129–4137.
[4] N. M. Chung, H. D.-A. Le, V. A. Nguyen, Q. Q.-V. Nguyen, T. D.-M. Nguyen, T.-T. Thai, and S. V.-U. Ha, "Multi-camera multi-vehicle tracking with domain gen eralization and contextual constraints," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 3327–3337

CONCLUSION

- Study MCMT in the city scale with CMN, where the cameras at different locations are usually non-overlapped, so much raw data with redundant video are created during objects monitoring, which brings quite huge computation challenge.
- Follow the way to select the subset of the raw data exactly containing the objects so as to reduce the computation challenge while guaranteeing a certain tracking accuracy.

Thanks!

Q&A

Presenter: JianHui Zhang, Pof., Leader of IoT Group

College of Computer Science & Tech.
Hangzhou Dianzi Univ., Hangzhou China