DyGLIP: A Dynamic Graph Model with Link Prediction for Accurate Multi-Camera Multiple Object Tracking

Kha Gia Quach, Pha Nguyen, Huu Le, Thanh-Dat Truong, Chi Nhan Duong, Minh-Triet Tran, Khoa Luu

CVPR VIRTUAL JUNE 19-25

1. Introduction

Most of previous matching approaches rely on an underlying re-ID model and perform classical matching algorithms, i.e. clustering, nearest neighbors, hungarian algorithm.

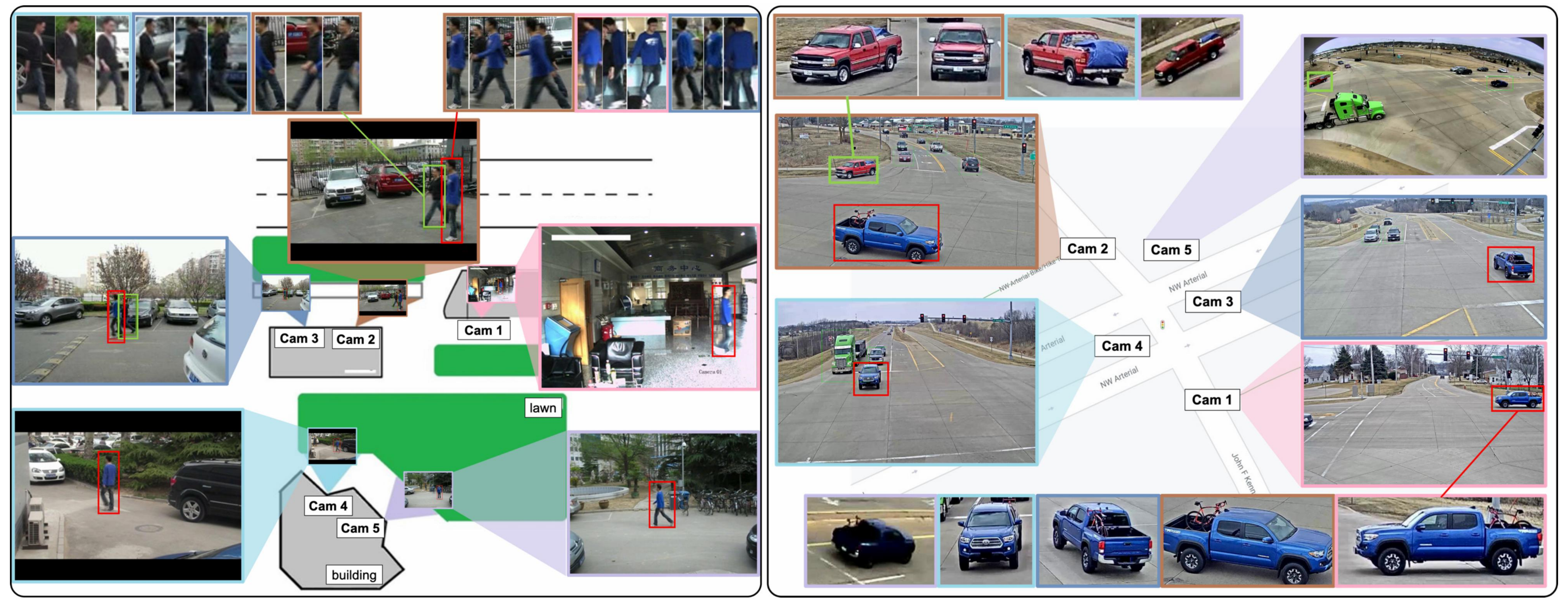

Fig 1: Objects' transition from an indoor to an outdoor environment can be totally different. Left: MCT dataset [1], Right: CityFlow dataset[2].

2. Contribution

- We introduce a novel MC-MOT framework which uses the link prediction in conjunction with a dynamic graph formulation.
- The proposed dynamic graph allows dynamically accumulating temporal and spatial information by incorporating with *attention mechanism*.

3. Dynamic Graph Formulation

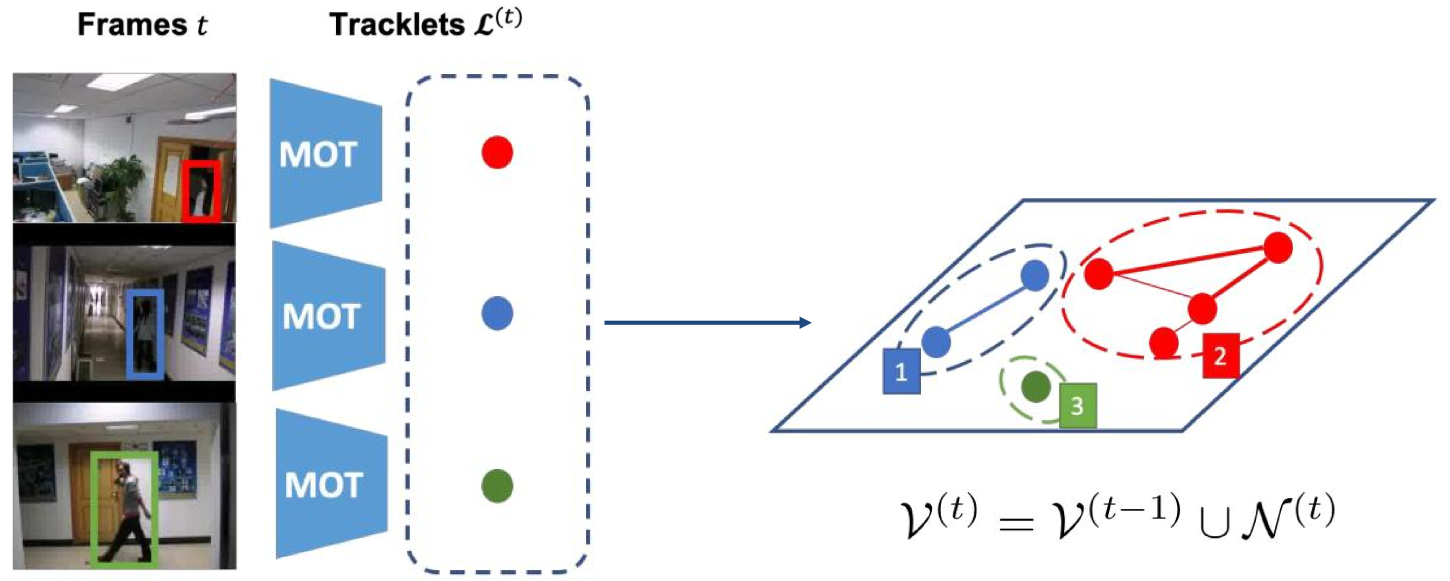

Fig 2: We maintain a dynamic graph during our tracking process.

4. Dynamic Graph with Self-Attention Module

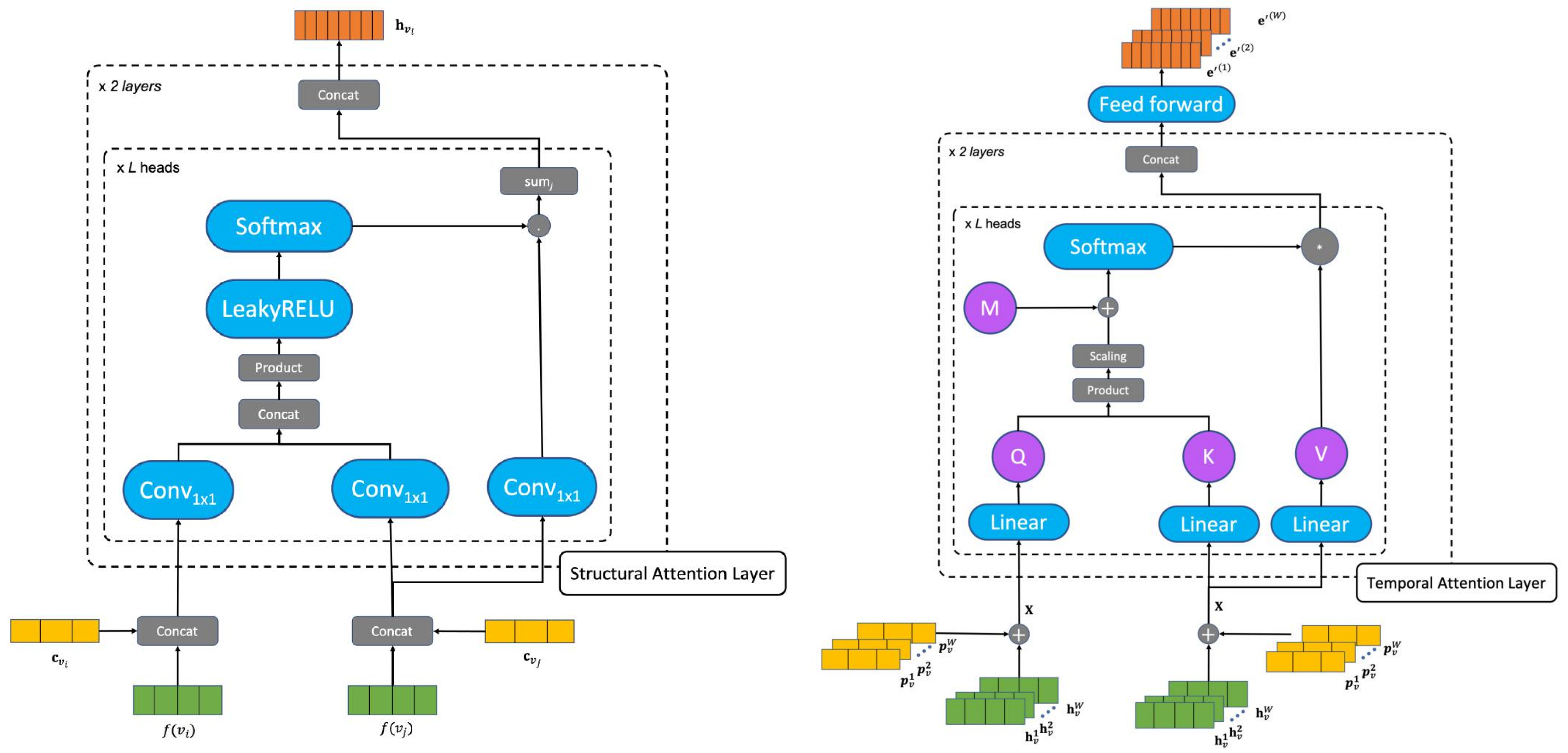

Fig 3: Structure of Structural Attention and Temporal Attention Layers.

Our attention layers take into account not only the provided embedding features but also the camera information and timestamp position encoding.

5. Link prediction and model learning

We compute dot product of two transformed vector to measure their similarity. The loss function is a combination between binary cross-entropy and classification:

$$\mathcal{L}(v_i) = \sum_t^T \Bigg(\sum_{v_j \in \mathcal{N}_b^{(t)}(v_i)} -\log\left(\sigma\left(< e'_{v_i}, e'_{v_j} >\right)\right) - w_g \sum_{v_k \in \mathcal{N}_g^{(t)}(v_i)} \log\left(1 - \sigma\left(< e'_{v_i}, e'_{v_k} >\right)\right) + \sum_{v_j \in \mathcal{N}_a^{(t)}(v_i)} \mathcal{L}_c(v_i, v_j) \Bigg)$$

6. Quantitative results

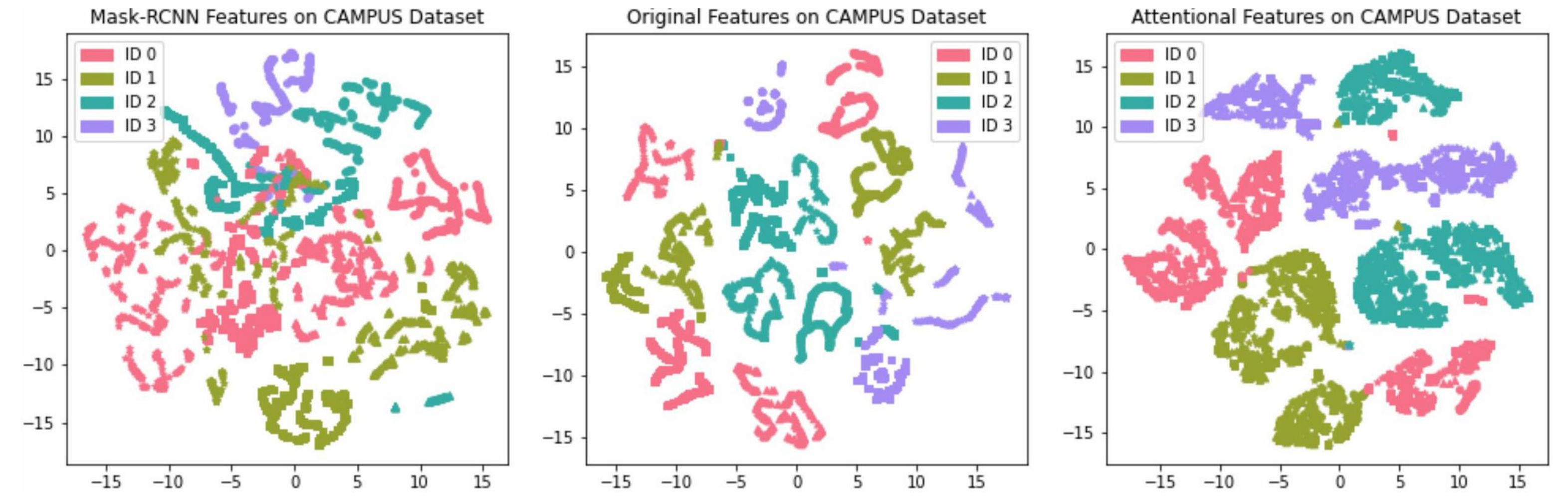

Fig 4: Node embedding in t-SNE space.

Features produced by DyGLIP form better clusters than the original features.

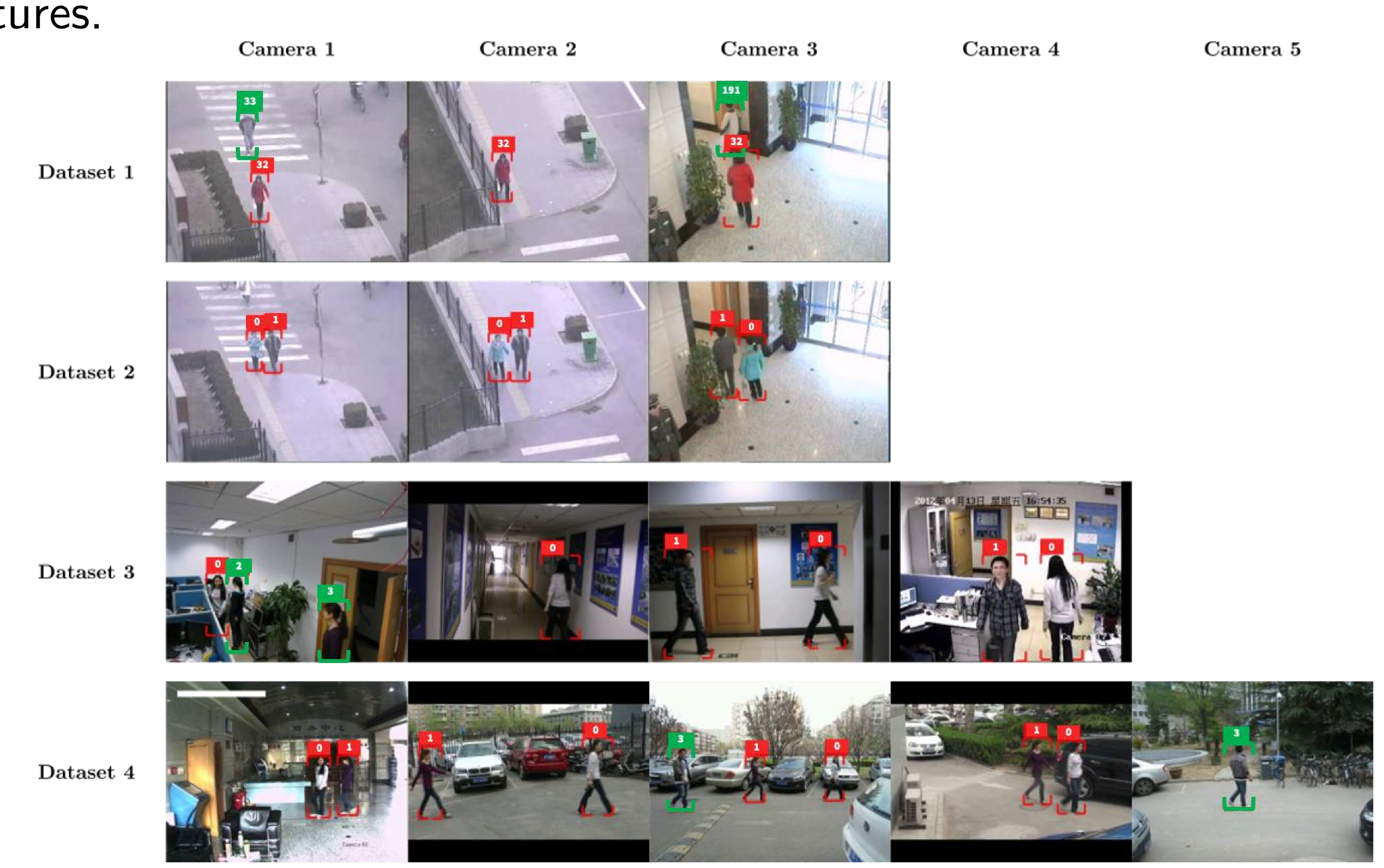

Fig 5: Results on MCT [1] dataset.

7. References

[1] Weihua Chen, Lijun Cao, Xiaotang Chen, and Kaiqi Huang. An equalized global graph model-based approach for multi-camera object tracking. IEEE Transactions on Circuits and Systems for Video Technology.

[2] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Ratnesh Kumar, David Anastasiu, and Jenq-Neng Hwang. Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. CVPR.