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

Supplementary Material

1. MC-MOT Dynamic Graph Processing Flow

This section briefly summarizes our proposed DyGLIP pipeline for building dynamic graphs and assigning ID for tracklets globally.

Algorithm 1 DyGLIP process for Graph Building & Assigning global tracklets' ID

Init $t \leftarrow 0$ /* Time */, $V \leftarrow \emptyset$
while $t < t_{\max}$ **do**
  Obtain the set of tracklets $\mathcal{L}^{(t)}$ (from the underlying MOT tracker).
  for $\mathbf{l}_j \in \mathcal{L}^{(t)}$ **do**
    $\mathcal{V}^{(t)} \leftarrow \mathcal{V}^{(t-1)} \cup \mathbf{l}_j$ /* Add new nodes to graph */
    /* Use the vector $\mathbf{l}_i$ as node features. */
  end for
  for $v_i \in \mathcal{V}^{(t)}$ **do**
    Obtain node embedding $\mathbf{e}'_{v_i}$ /* Section 3.3.2 */
  end for
  for $v_i \in \mathcal{N}^{(t)}$ **do**
    Obtain link prediction to the remaining nodes /* Section 3.4 */
    Assign ID based on predicted link.
  end for
  $t \leftarrow t + 1$
end while

Fig. 1 and Fig. 2 illustrate the structure of SAL and TAL in our proposed DyGLIP, respectively.

1.1. Additional Results

In this section, we include the additional results for ablation studies and visualization of the final MC-MOT.

1.1.1 Ablation Studies

We conduct additional ablation studies to evaluate the effects on the configuration of the attention modules in DyGLIP, including the number of attention layers. Table **??** shows the performance of DyGLIP in terms of ID accuracy, i.e., ID F1, IDP, IDR, and IDS, using various configurations of the attention modules. We change the number of layers for structural attention and the temporal attention layers independently. We use a fixed number of layers, i.e., 2, for SAL and TAL while changing the other, respectively.

Method	ID F1 (%) ↑	IDP (%) ↑	IDR (%) ↑	IDS ↓
No Attention	39	50.1	36.5	135
SAL 1-layer	48.0	57.2	48.0	**32**
SAL 2-layer	**56.2**	**59.5**	**56.2**	44
SAL 3-layer	55.4	58.8	55.4	34
TAL 1-layer	55.9	56.2	55.9	**53**
TAL 2-layer	**56.2**	**59.5**	**56.2**	**44**
TAL 3-layer	51.5	52.5	51.5	45

Table 1. Ablation study on different configuration for structural and temporal attention modules.

1.1.2 MC-MOT Visualization Results

Illustration results of our DyGLIP on various test datasets, i.e., CAMPUS, PETS, and MCT We visualize tracking results at a certain frame taken from PETS, Cam-

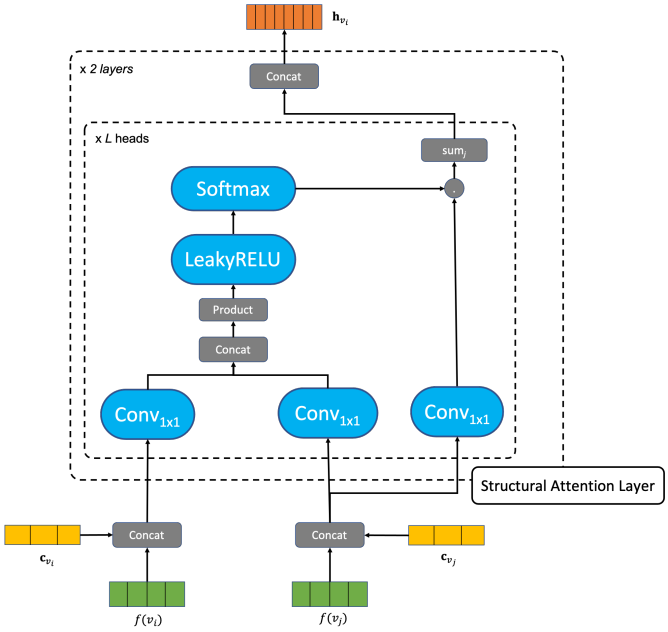

Figure 1. Structure of SAL in the attention modules

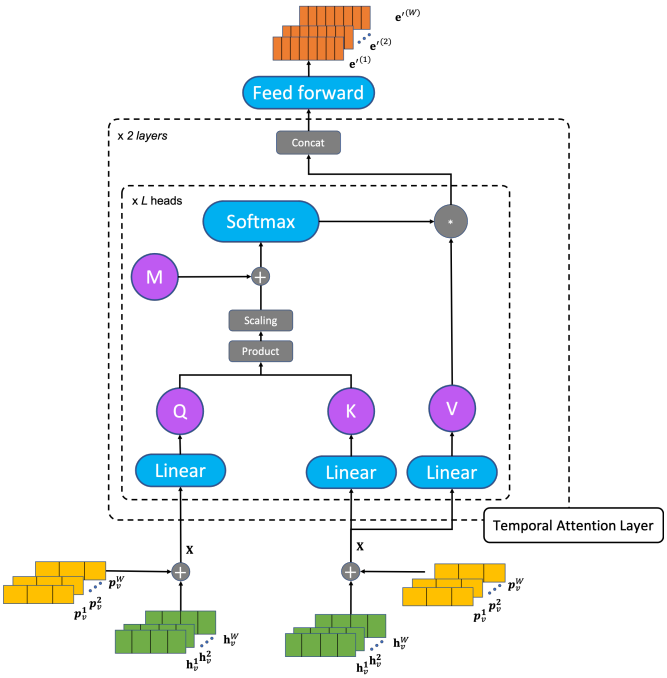

Figure 2. Structure of TAL in the attention modules

pus, CityFlow and MCT datasets in Figures 3, 4, 5 and 6, respectively.

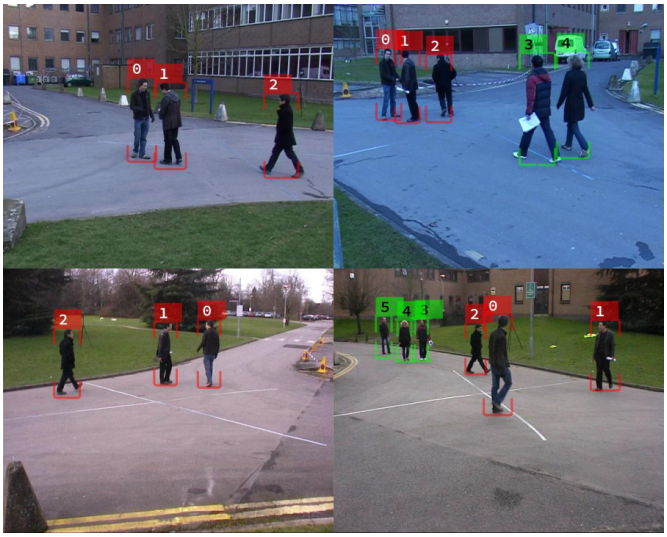

Figure 3. Results on PETS dataset. (**Best viewed in color**)

Illustration results on recovering from single-camera MOT error Fig. 7 shows some more cases that our DyGLIP can recover from local tracklet errors.

Figure 4. Results on CAMPUS dataset. (**Best viewed in color**)

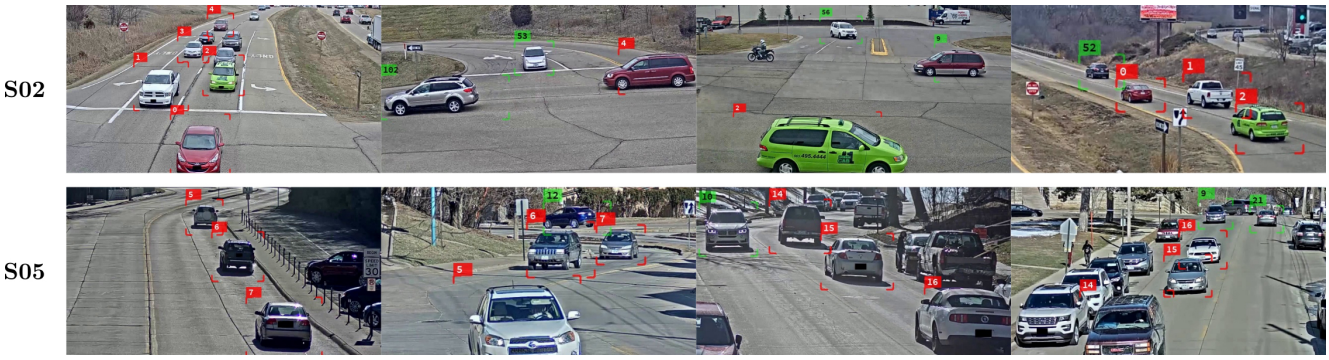

Figure 5. Results on CityFlow dataset. (**Best viewed in color**)

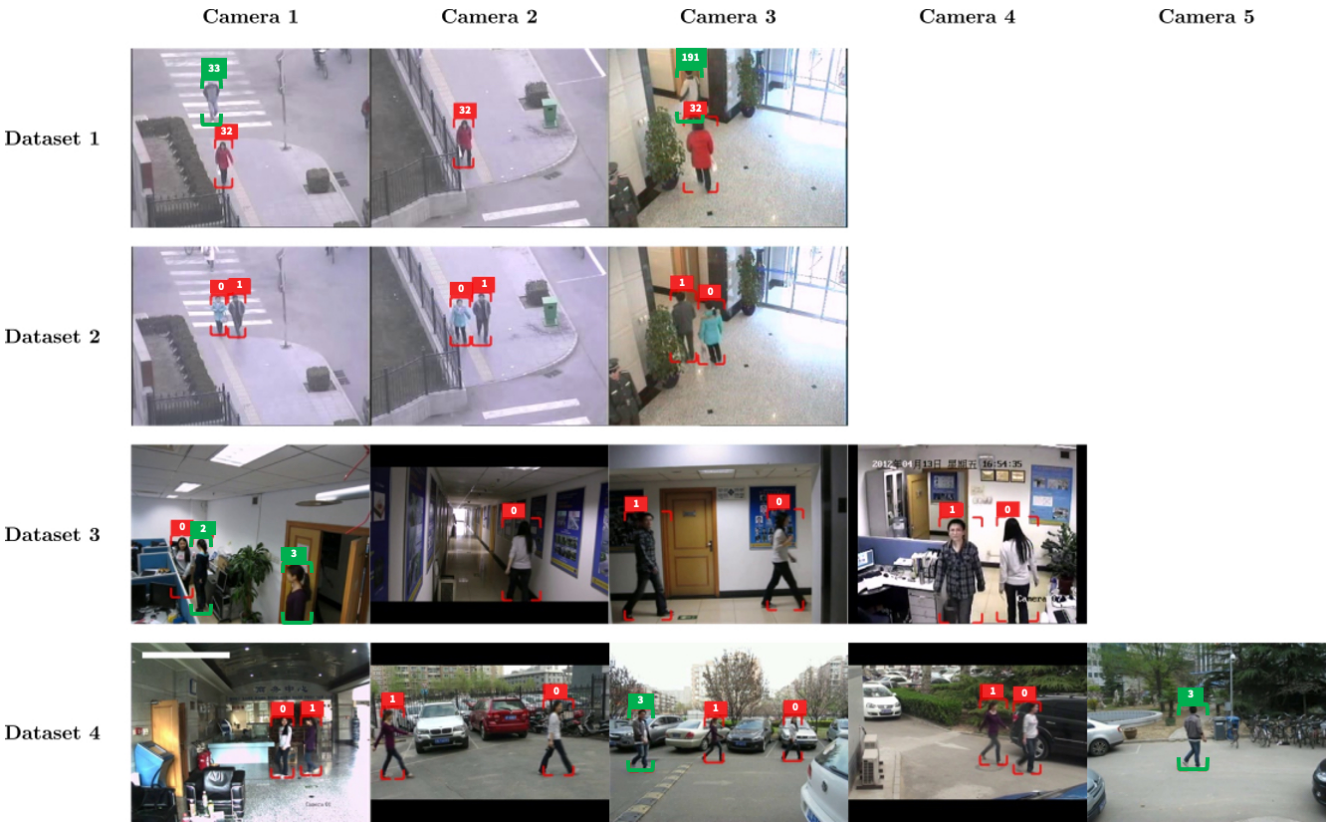

Figure 6. Results on MCT dataset. (**Best viewed in color**)

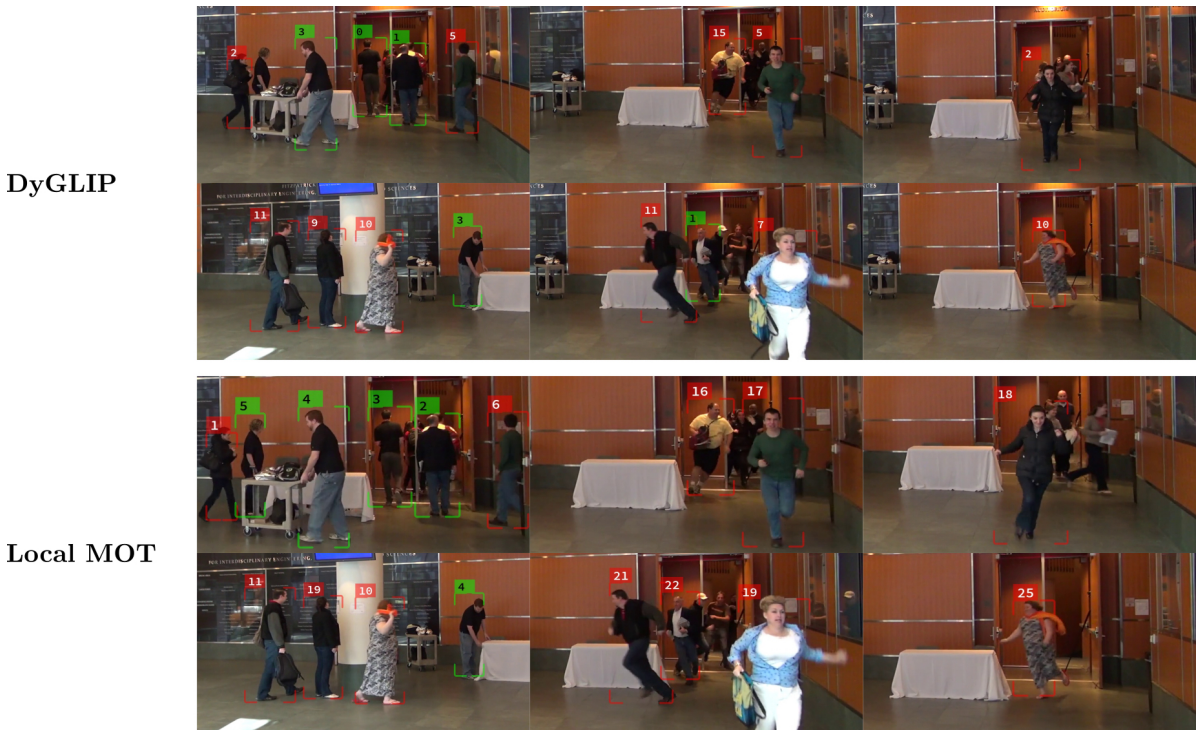

Figure 7. Our proposed method (up) corrects a negative matched case caused by a short-memorized MOT system (down). Determined ID is highlighted by the red bounding box.