Learnable Graph Matching: Incorporating Graph Partitioning with Deep Feature Learning for Multiple Object Tracking

Jiawei He[1,3] Zehao Huang[2] Naiyan Wang[2] Zhaoxiang Zhang[1,3,4]
[1]CASIA [2]TuSimple [3]UCAS [4]Centre for Artificial Intelligence and Robotics, HKISI_CAS

CVPR VIRTUAL JUNE 19-25

Abstract

Some problems in recent Tracking-by-Detection paradigm: 1) ignoring intra-frame context information; 2) Solely relying on fitting data by end-to-end training or inconsistent between training and inference.
So, in this paper, we

- Focus on the object association across frames in Multiple Object Tracking problem, and propose a novel learnable graph matching method.
- Utilize the edge in the graph to model the relationship between tracklets and intra-frame detections.
- Propose a differentiable quadratic programming layer based on the continuous relaxation of the problem, with the help of implicit function theorem and KKT conditions to derive the gradient.
- Achieve state-of-the-art performance on MOTChallenge benchmark.

(a) Hungarian Algorithm

(b) Graph Matching

Problem Formulation

Graph Matching problem between the tracklet and the detection graph. The original graph matching problem (QAP):

$$\mathbf{\Pi}^* = \arg\min_{\mathbf{\Pi}} \frac{1}{2}||\mathbf{A_1\Pi} - \mathbf{\Pi A_2}||_F^2 - \text{tr}(\mathbf{B}^\top \mathbf{\Pi}), \quad (1)$$

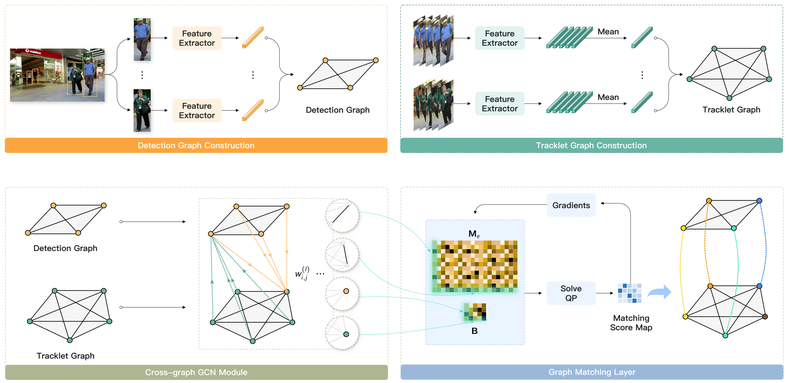
$\mathbf{\Pi}$: permutation matrix between two graphs, $\mathbf{A}$: weighted adjacency matrix, $\mathbf{B}$: vertex affinity matrix.
We expand the Eq. 1 from adjacency matrix $\mathbf{A}$ to adjacency tensor $\mathbf{H}$, i.e., we consider the edge feature instead of edge weight:

$$\mathbf{\Pi}^* = \arg\min_{\mathbf{\Pi}} \sum_{c=1}^{d} \frac{1}{2}||\mathbf{H}_1^c\mathbf{\Pi} - \mathbf{\Pi}\mathbf{H}_2^c||_F^2 - \text{tr}(\mathbf{B}^\top \mathbf{\Pi}). \quad (2)$$

Simplifying & relaxing to QP:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \mathbf{x}^\top((n-1)^2\mathbf{I} - \mathbf{M})\mathbf{x} - \mathbf{b}^\top\mathbf{x}. \quad (3)$$

Method



- **Feature encoding in tracklet and detection graphs:**
  - The vertex feature is appearance feature, in the detection graph.
  - In the tracklet graph, the vertex feature is the mean of appearance feature across frame.
- **Feature enhancement by cross-graph GCN:**
  - Only inter-graph GCN.
  - The weight $w_{i,j} = \cos(\mathbf{h}_i\mathbf{h}_j) + \text{IoU}(\mathbf{g}_i, \mathbf{g}_j)$ when video is taken by static camera.
  - The weight $w_{i,j} = \cos(\mathbf{h}_i, \mathbf{h}_j)$ when camera moves.
  - We normalize the features before aggregation as [DeeperGCN, 2020].
- **Matching by a differentiable graph matching layer:**
  - Solving QP (Eq. 3) and calculating the gradients with the help of the implicit function theorem and KKT conditions (refer to Appendix A).
  - The score map is normalized by Softmax with temperature.
- **Training & Inference:**
  - Training with weighted BCE Loss:

$$\mathcal{L} = \frac{-1}{n_d n_t} \sum_{i=1}^{n_d} \sum_{j=1}^{n_t} k y_{i,j} \log(\hat{y}_{i,j}) + (1 - y_{i,j}) \log(1 - \hat{y}_{i,j})$$

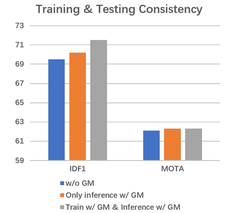
  - Greedy rounding to obtain final matching results during inference.
  - Determining matching or not by thresholds, including feature similarity gate, Kalman Filter gate and IoU gate.

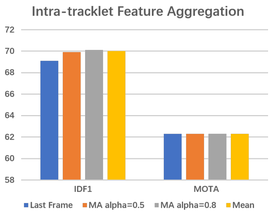
Experiments & Results

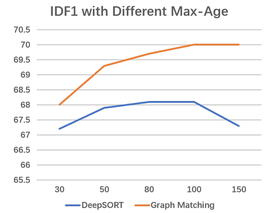
Ablation study on MOT17 val set:

GM	App. Enc.	GCN	Geo	Inter.	IDF1 ↑	MOTA ↑	MT ↑	ML ↓	FP ↓	FN ↓	ID Sw. ↓
					68.1	62.1	556	371	1923	124480	1135
✓					70.0	62.3	555	374	1735	124292	1128
✓			✓		70.2	62.2	555	374	1744	124301	1140
✓	✓				70.4	62.3	554	375	1741	124298	1058
✓	✓	✓			70.6	62.2	556	374	1748	124305	1399
✓	✓	✓	✓		71.5	62.3	555	375	1741	124298	1017
				✓	68.9	62.9	678	361	11440	112853	723
✓				✓	71.6	64.0	669	365	7095	113392	659
✓			✓	✓	71.7	64.0	666	364	6816	113778	724
✓	✓			✓	72.0	64.2	671	368	7701	112370	627
✓	✓	✓		✓	72.1	63.3	676	364	10888	111869	716
✓	✓	✓	✓	✓	73.0	63.8	672	361	9579	111683	570

Discussions:







Comparisons:

Methods	Refined Det	IDF1 ↑	HOTA↑	MOTA ↑	MT↑	ML↓	FP ↓	FN ↓	IDS ↓	AssA↑	DetA↑	LocA↑
MOT17												
GNMOT (O*)	-	47.0	-	50.2	19.3	32.7	29316	246200	5273	-	-	-
FAMNet (O)	-	48.7	-	52.0	19.1	33.4	14138	253616	3072	-	-	-
JBNOT (O*)	-	50.8	41.3	52.6	19.7	35.8	31572	232659	3050	39.8	43.3	80.2
Tracktor++ (O)	Tracktor	52.3	42.1	53.5	19.5	36.6	12201	248047	2072	41.7	42.9	80.9
Tracktor++v2 (O)	Tracktor	55.1	44.8	56.3	21.1	35.3	8866	235449	1987	45.1	44.9	81.8
GNNMatch (O)	Tracktor	56.1	45.4	57.0	23.3	34.6	12283	228242	1957	45.2	45.9	81.5
GSM_Tracktor (O)	Tracktor	57.8	45.7	56.4	22.2	34.5	14379	230174	**1485**	47.0	44.9	80.9
CTTrackPub (O)	CenterTrack	59.6	48.2	61.5	**26.4**	**31.9**	14076	200672	2583	47.8	49.0	81.7
GMTracker(Ours) (O)	Tracktor	63.8	49.1	56.2	21.0	35.5	**8719**	236541	1778	53.9	44.9	81.8
GMT_CT(Ours) (O)	CenterTrack	**66.9**	**52.0**	**61.5**	26.3	32.1	14059	**200655**	2415	**55.1**	**49.4**	**81.8**
TPM	-	52.6	41.5	54.2	22.8	37.5	13739	242730	1824	40.9	42.5	80.0
eTC17	-	58.1	44.9	51.9	23.1	35.5	36164	232783	2288	47.0	43.3	79.4
MPNTrack	Tracktor	61.7	49.0	58.8	28.8	33.5	17413	213594	1185	51.1	47.3	81.5
Lif_TsimInt	Tracktor	65.2	50.7	58.2	28.6	33.6	16850	217944	**1022**	54.9	47.1	81.5
LifT	Tracktor	65.6	51.3	60.5	27.0	33.6	14966	206619	1189	54.7	48.3	81.3
GMT_simInt (Ours)	Tracktor	65.9	51.1	59.0	29.0	33.6	20395	209553	1105	55.1	47.6	81.2
GMT_VIVE (Ours)	Tracktor	65.9	51.2	60.2	26.5	33.2	**13142**	209812	1675	55.1	47.8	81.3
GMTCT_simInt (Ours)	CenterTrack	**68.7**	**54.0**	**65.0**	**29.4**	**31.6**	18213	**177058**	2200	**56.4**	**52.0**	**81.5**
MOT16												
Tracktor++v2 (O)	Tracktor	54.9	44.6	56.2	20.7	35.8	2394	76844	617	44.6	44.8	82.0
GNNMatch (O)	Tracktor	55.9	44.6	56.9	22.3	35.3	3235	74784	564	43.7	45.8	81.7
GSM_Tracktor (O)	Tracktor	58.2	45.9	57.0	22.0	34.5	4332	73573	**475**	46.7	45.4	81.1
GMTracker(Ours) (O)	Tracktor	63.9	48.9	55.9	20.3	36.6	**2371**	77545	531	53.7	44.6	**82.1**
GMT_CT (Ours) (O)	CenterTrack	**68.6**	**53.1**	**62.6**	**26.7**	**31.0**	5104	**62377**	787	**56.3**	**50.4**	81.8
TPM	-	47.9	36.7	51.3	18.7	40.8	**2701**	85504	569	34.6	39.3	79.1
eTC	-	56.1	42.0	49.2	17.3	40.3	8400	83702	606	44.5	39.9	78.8
MPNTrack	Tracktor	61.7	48.9	58.6	27.3	34.0	4949	70252	354	51.1	47.1	81.7
Lif_TsimInt	Tracktor	64.1	49.6	57.5	25.4	34.7	4249	72868	**335**	53.3	46.5	**81.9**
LifT	Tracktor	64.7	50.8	61.3	27.0	34.0	4844	65401	389	53.1	48.9	81.4
GMT_simInt (Ours)	Tracktor	66.2	51.2	59.1	27.5	34.4	6021	68226	341	55.1	47.7	81.5
GMT_VIVE (Ours)	Tracktor	66.6	51.6	61.1	26.7	33.3	3891	66550	503	55.3	48.5	81.5
GMTCT_simInt (Ours)	CenterTrack	**70.6**	**55.2**	**66.2**	**29.6**	**30.4**	6355	**54560**	701	**57.8**	**53.1**	81.5