Graph Networks for Multiple Object Tracking

Jiahe Li, Xu Gao, Tingting Jiang (ttjiang@pku.edu.cn)

NELVT, Department of Computer Science, Peking University, China

WACV 2020
March 1-5, Aspen, CO

北京大学数字媒体研究所
INSTITUTE OF DIGITAL MEDIA, PEKING UNIVERSITY

Background

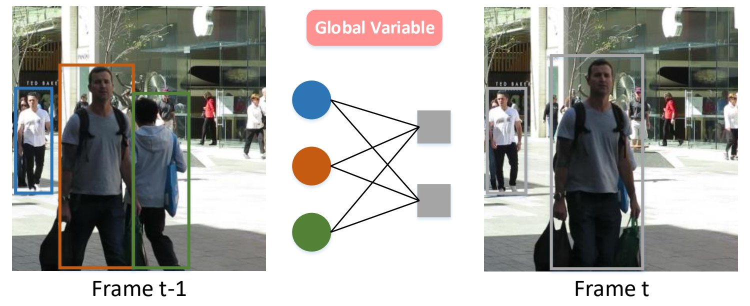
Most current graph models are **static**:

- Nodes and edges are fixed.
- The global relationship among objects is not modeled.

Motivation

Make use of the graph network [1] to enable the update of nodes and edges.

Pipeline



We construct a graph:

- Nodes: the objects and the detections.
- Edges: the associations between objects and detections.
- Global variable: the global relationship among objects.

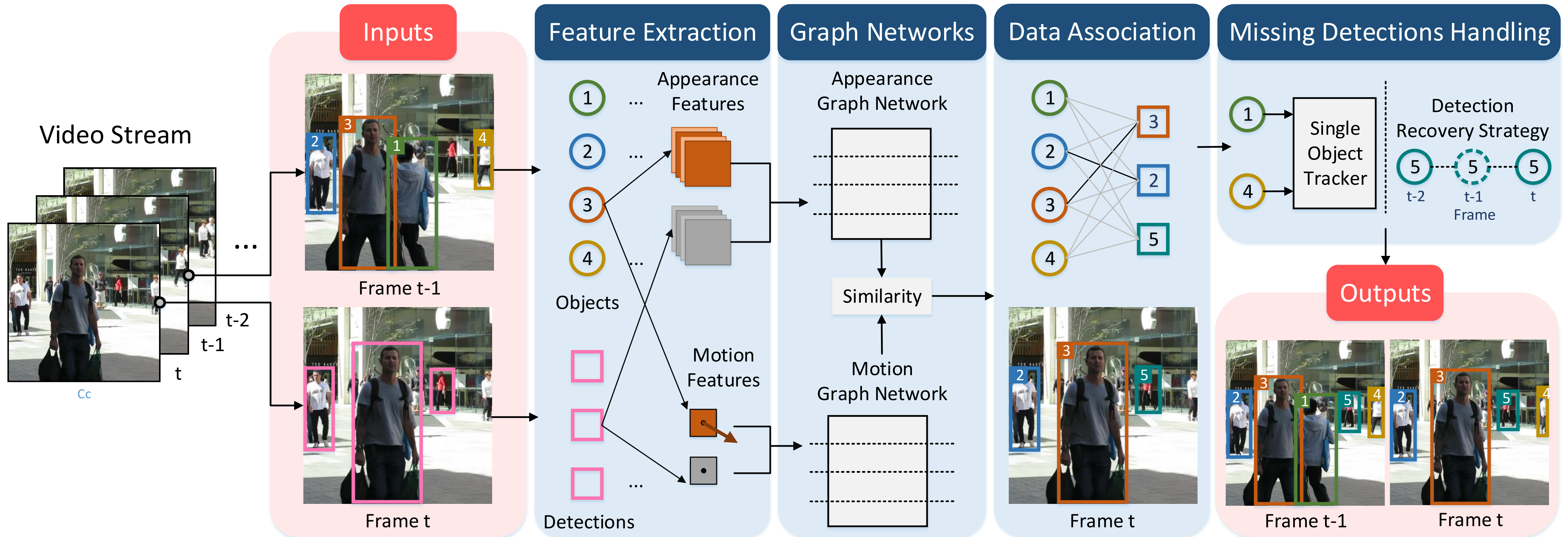


Figure 1. Pipeline of our MOT model. There are four procedures: feature extraction, graph networks, data association and missing detection handling.

4-step Graph Network

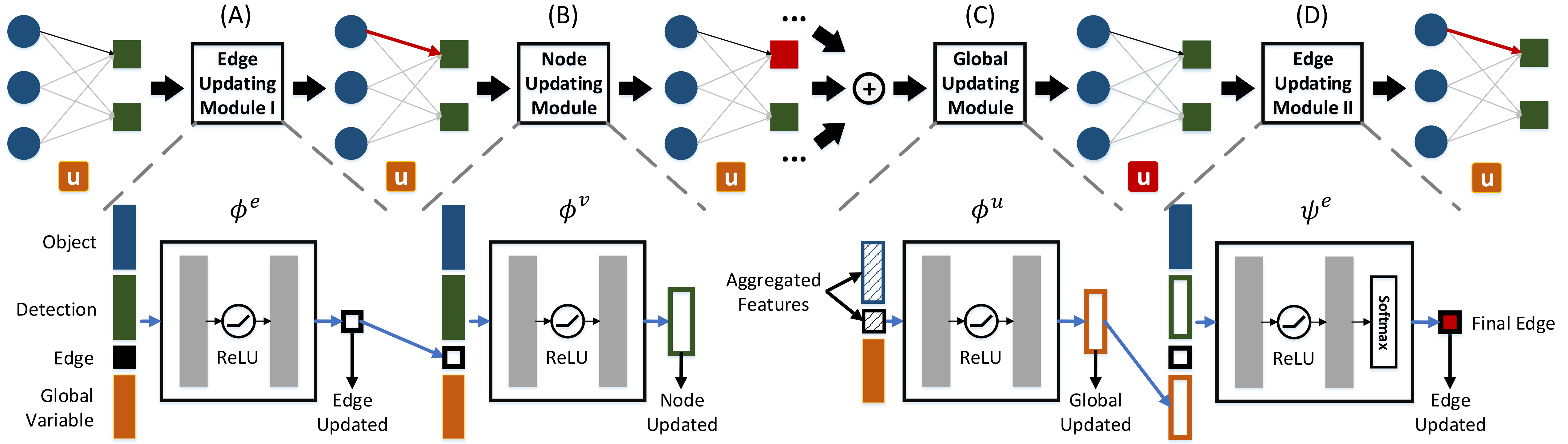
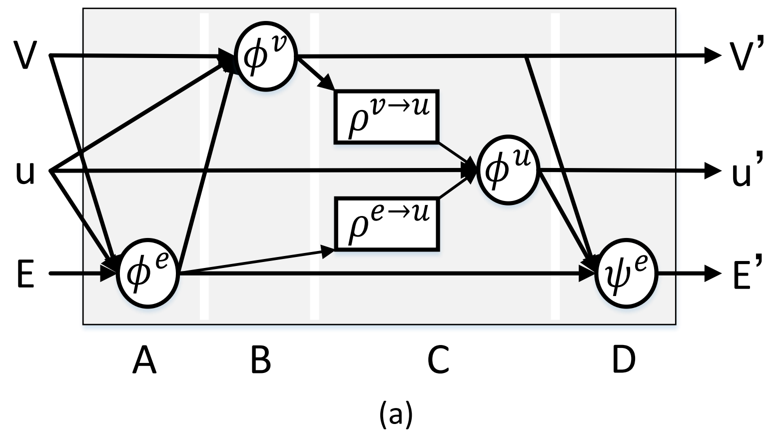


Figure 2. **Upper part**: The structure of the 4-step graph network. **Lower part**: The corresponding networks for the four modules.



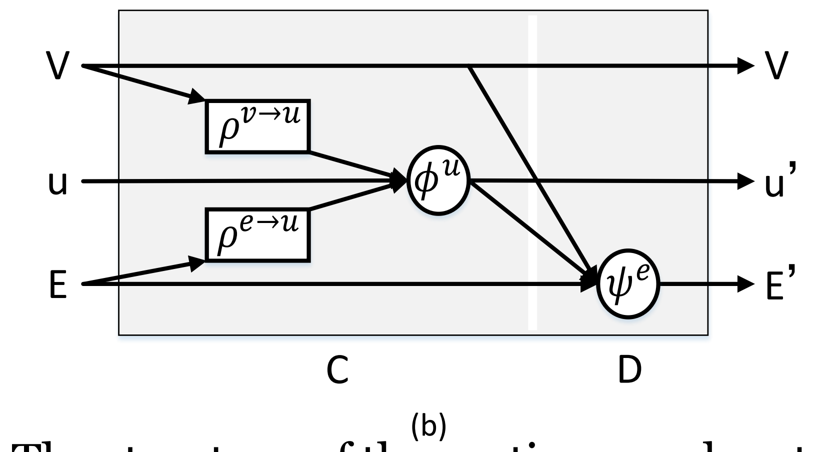


Figure 3. **(a)** The structure of the appearance graph network. **(b)** The structure of the motion graph network.

Experiments

Dataset	Detection	Methods	MOTA	IDF1	MT	ML	FP	FN	IDS	FM
MOT16	Public	LINF [2], *ECCV 2016*	41.0	45.7	11.6%	51.3%	7896	99224	430	963
		MHT_bLSTM [3]*, *ECCV 2018*	42.1	47.8	14.9%	44.4%	11637	93172	753	1156
		NOMT [4], *ICCV 2015*	46.4	**53.3**	**18.3%**	41.4%	9753	87565	**359**	**504**
		Ours without SOT	47.4	42.6	14.5%	34.4%	**7795**	86178	1931	3389
		Ours	**47.7**	43.2	16.1%	**34.3%**	9518	**83875**	1907	3376
	Private	Ours without SOT	**58.4**	**54.8**	**27.3%**	**23.2%**	**5731**	**68630**	**1454**	**1730**
MOT17	Public	MHT_bLSTM [3]*, *ECCV 2018*	47.5	**51.9**	18.2%	41.7%	25981	268042	**2069**	**3124**
		Ours without SOT	50.1	46.3	18.6%	33.3%	**25210**	250761	5470	8113
		Ours	**50.2**	47.0	**19.3%**	**32.7%**	29316	**246200**	5273	7850

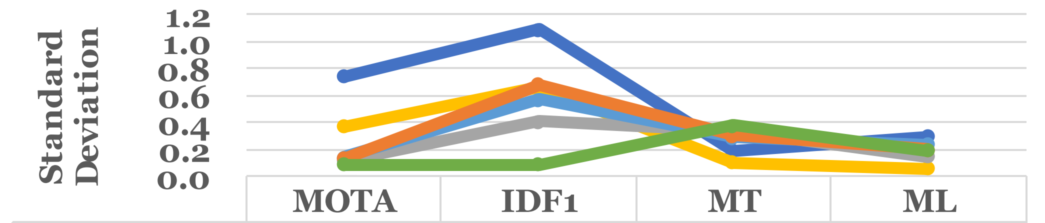
Table 1. Experiments on MOT16 and MOT17 test set. The best result in each metric is highlighted in bold, and the second best result is underlined. * indicates the use of additional training data.

Methods	MOTA	IDF1	MT	ML	FP	FN	IDS	FM
A*	**52.7**	**56.3**	**31.5**	33.0	**1455**	28882	**1161**	913
A*/g	52.6	55.8	31.2	**32.9**	1545	**28819**	1174	**885**
M	**53.9**	**61.4**	**31.9**	**32.2**	**1390**	**28570**	**690**	**772**
M/g	52.6	60.0	31.6	32.8	1392	28621	1521	802
A*+M	**54.5**	**63.7**	**33.2**	32.3	**1525**	**28210**	**511**	**683**
A*/g+M/g	54.3	62.3	32.9	**32.0**	1622	28247	517	692

Table 2. Performance of models with/without the global variable. **A***, **M** and **A*+M** denote the appearance graph network, the motion graph network and the merged graph network respectively. **A*/g** denotes **A*** without the global variable. **M/g** denotes **M** without the global variable. **A*/g+M/g** denotes **A*+M** without the global variable. The best result is highlighted in bold.

Methods	MOTA	IDF1	MT	ML	FP	FN	IDS	FM
$L_C + \lambda L_N$	**52.7**	**56.3**	31.5	**33.0**	**1455**	28882	**1161**	**913**
L_C	52.5	56.0	**32.0**	33.9	1539	**28811**	1253	939

Table 3. Performance of **A*** trained with/ without L_N. L_C and L_N denotes the cross-entropy loss and the node cost loss respectively.



	MOTA	IDF1	MT	ML
M	0.7	1.1	0.2	0.3
M/g	0.4	0.7	0.1	0.1
A*	0.1	0.4	0.3	0.2
A*/g	0.1	0.6	0.3	0.2
A*+M	0.1	0.7	0.3	0.2
A*/g+M/g	0.1	0.1	0.4	0.2

Figure 4. Standard deviation and mean of MOTA, IDF1, MT and ML of our methods over five initializations.

Reference

[1] Battaglia et al. Relational inductive biases, deep learning, and graph networks. arXiv, 2018.

[2] Fagot-Bouquet et al. Improving multi-frame data association with sparse representations for robust near-online multi-object tracking. ECCV, 2016.

[3] Kim et al. Multi-object tracking with neural gating using bilinear LSTM. ECCV, 2018.

[4] W. Choi. Near-online multi-target tracking with aggregated local flow descriptor. ICCV, 2015.

The introduction video is available at http://jiaheli.wacv.cc/