GMFlow: Learning Optical Flow via Global Matching

https://github.com/haofeixu/gmflow

Haofei Xu[1] Jing Zhang[2] Jianfei Cai[1] Hamid Rezatofighi[1] Dacheng Tao[3,2]

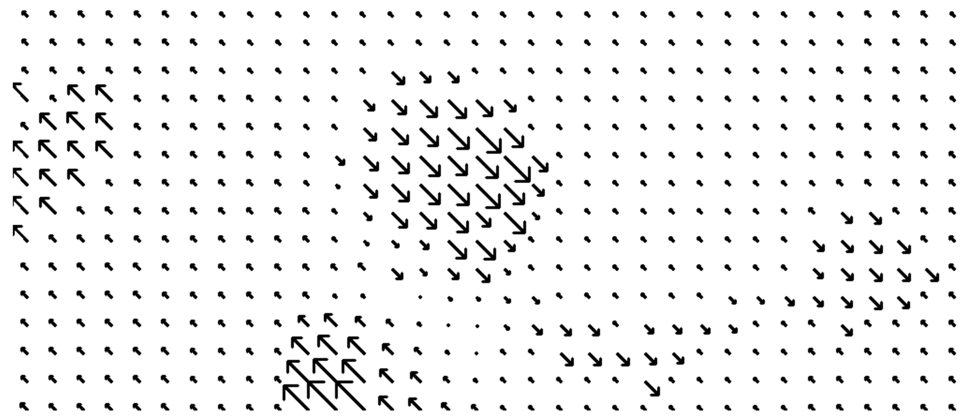
[1]Monash University, Australia [2]The University of Sydney, Australia [3]JD Explore Academy, China

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Problem: Optical Flow

- Apparent motion between two video frames

frame1



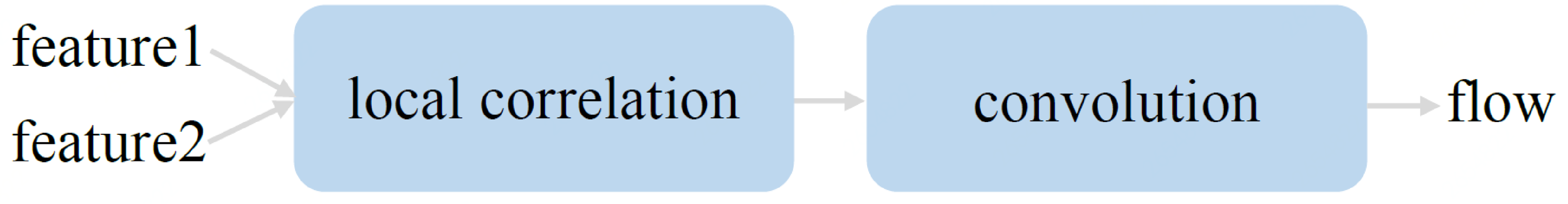
optical flow

frame2

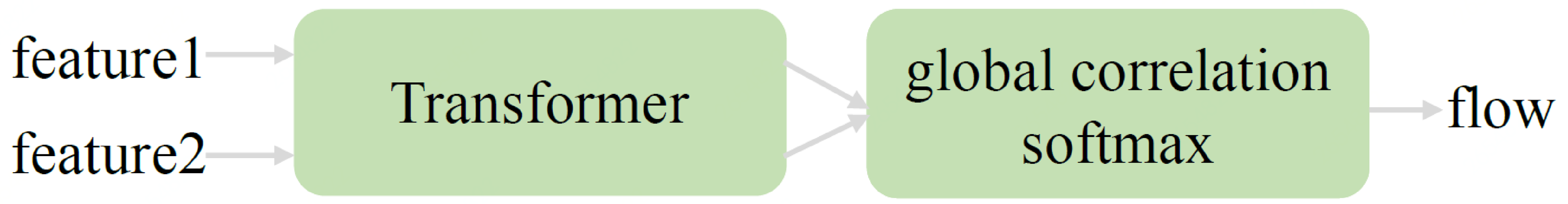
optical flow (visualization)

Local Regression *vs.* Global Matching

- Directly compare all pair-wise feature similarities (i.e., global matching), with a Transformer to obtain stronger features



(a) previous flow estimation approach

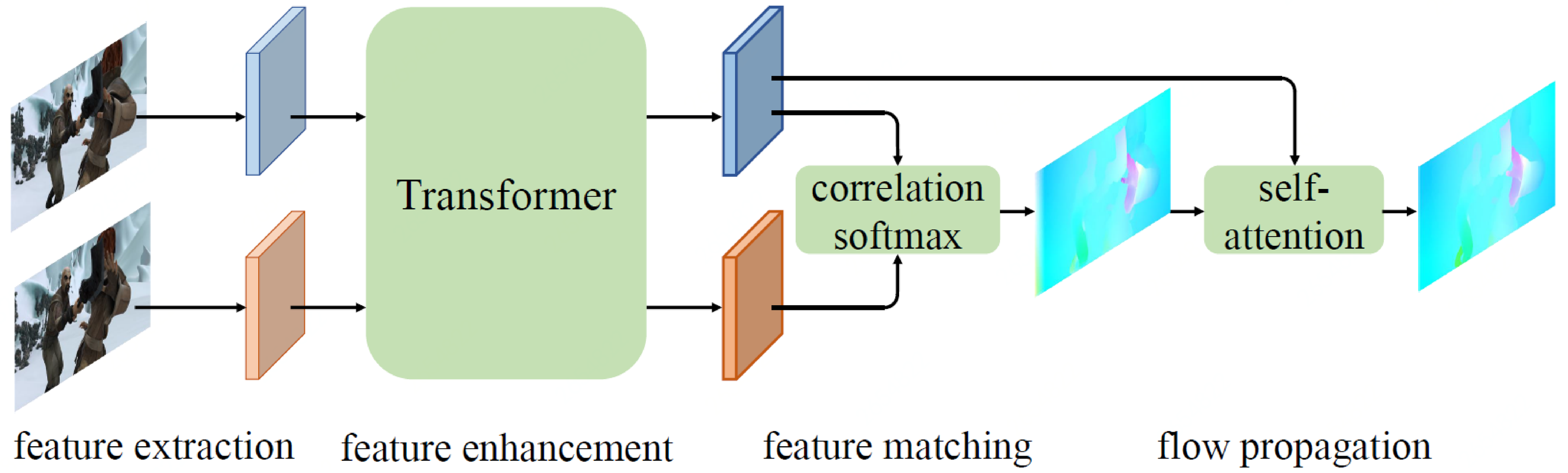


(b) GMFlow

Formulation: Global Matching

- Inputs: I_1 I_2
- Feature extraction: $F_1, F_2 \in \mathbb{R}^{H \times W \times D}$
- Global correlation: $C = \frac{F_1 F_2^T}{\sqrt{D}} \in \mathbb{R}^{H \times W \times H \times W}$
- Softmax normalization: $M = \mathrm{softmax}(C) \in \mathbb{R}^{H \times W \times H \times W}$
- Correspondence: $\hat{G} = MG \in \mathbb{R}^{H \times W \times 2}$ $G \in \mathbb{R}^{H \times W \times 2}$
- Optical flow: $V = \hat{G} - G \in \mathbb{R}^{H \times W \times 2}$

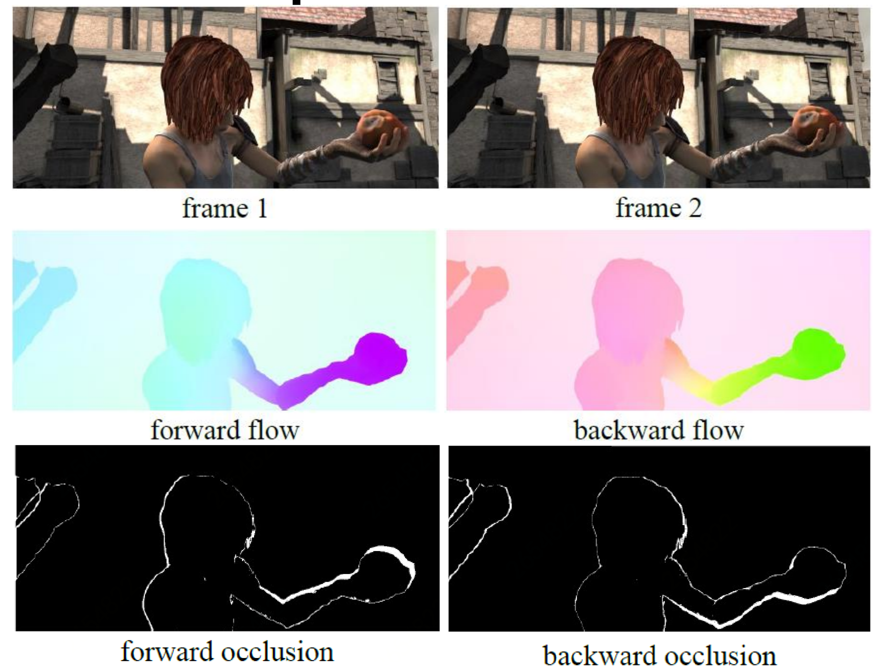
Framework



Benefits

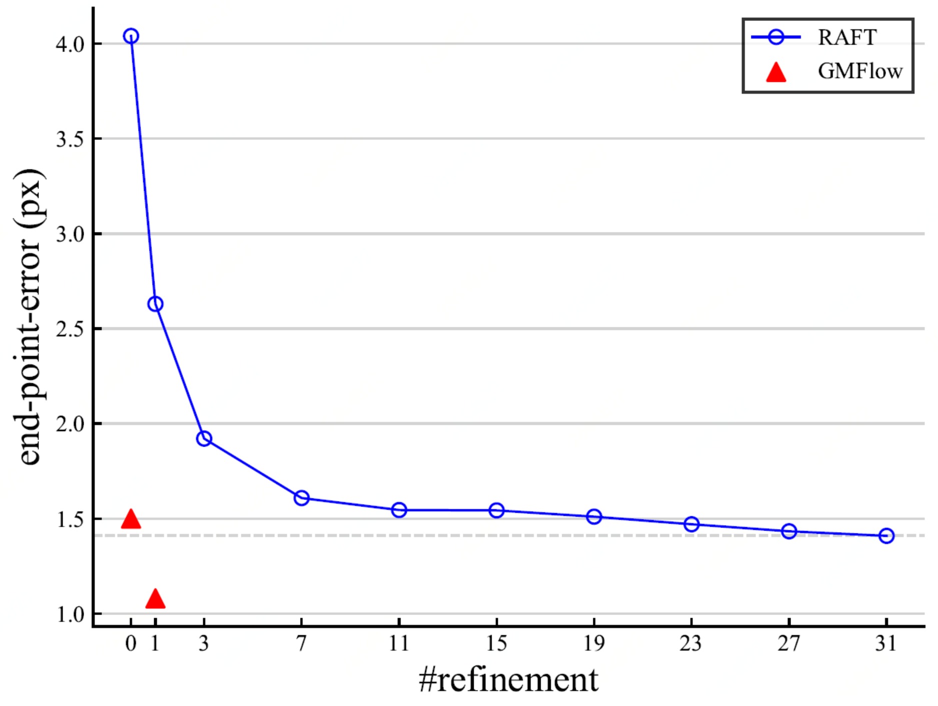
- Much better than local correlation + conv
- Simplify bidirectional optical flow prediction

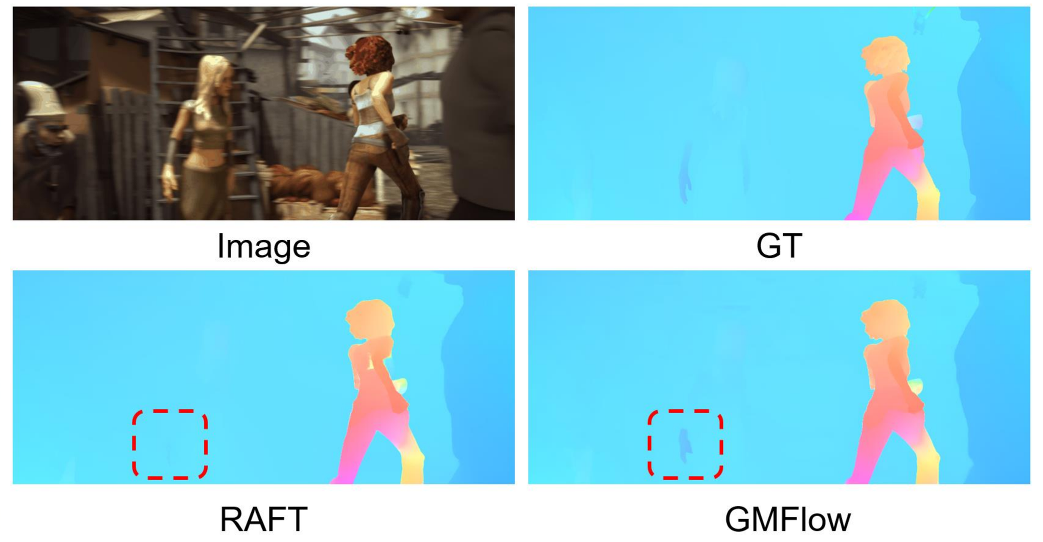
Method	#blocks	Things (val, clean)			
		EPE	s_{0-10}	s_{10-40}	s_{40+}
cost volume + conv	0	18.83	3.42	6.49	49.65
	4	10.99	1.70	3.41	29.78
	8	9.59	1.44	2.96	26.04
	12	9.04	1.37	2.84	24.46
	18	8.67	1.33	2.74	23.43
Transformer + softmax	0	22.93	8.57	11.13	52.07
	1	11.45	2.98	4.68	28.35
	2	8.59	1.80	3.28	21.99
	4	7.19	1.40	2.62	18.66
	6	**6.67**	**1.26**	**2.40**	**17.37**



Comparison with RAFT

- With only one refinement, GMFlow outperforms 31-refinements RAFT, while running faster





More Visual Results

code & models