UnSAMFlow: Unsupervised Optical Flow Guided by Segment Anything Model

Shuai Yuan, Lei Luo, Zhuo Hui, Can Pu, Xiaoyu Xiang, Rakesh Ranjan, Denis Demandolx (*Meta Reality Labs*)

CVPR SEATTLE, WA JUNE 17-21, 2024

Introduction

Unsupervised Optical Flow Estimation

- Important due to lack of labels
- Assuming appearance constancy and flow smoothness
- **Current challenges**:
  - Occlusion: **Objects** cover each other
  - Motion boundary: **Objects** move differently

→ **Optical flow is *low-level*, but we still need *object-level* info!**

What kind of object-level info?

- Previous work: Semantic Segmentation

  ✗ Separate instances of the same class ✗ Novel objects
- Ours: Segment Anything Model (SAM)

  ✓ Separate objects of different levels ✓ Open-world objects

→ **Use SAM masks to guide unsupervised optical flow!**

Method Overview

Two settings: We use SAM …

- Setting 1: only during training
- Setting 2: both training and inference (Inference speed ↓, Accuracy ↑)

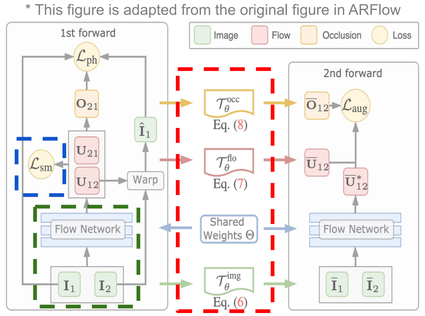
Baseline: ARFlow[1]

Three proposed adaptations:
For Setting 1:

- **Semantic augmentation**
- **Homography smoothness**

For Setting 2:

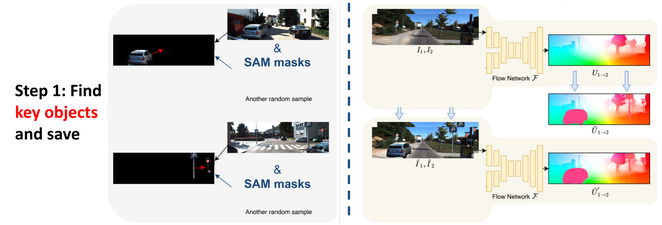
- **Mask feature module**



Method Details

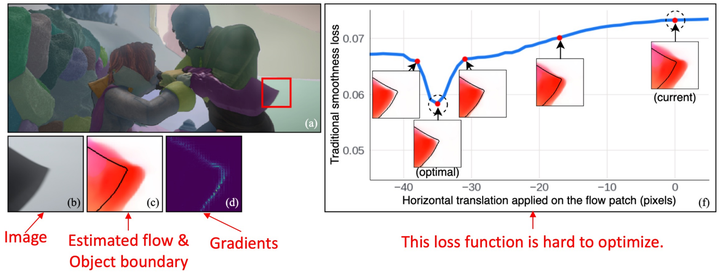
Semantic augmentation ("+aug")

Inspired by SemARFlow[2]; self-supervision by augmented samples



Homography smoothness loss ("+hg")

Previous edge-aware smoothness losses are too local!

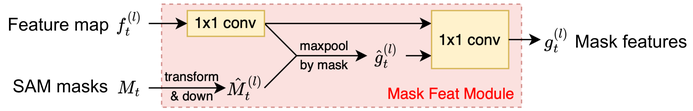


We evaluate "smoothness" based on piece-wise **homography**:

1. Pick SAM objects with occlusions
2. Estimate homography for each selected object
3. Refine and self-supervise

Mask feature module ("+mf")

Simply added to the decoder; aggregate features on each SAM object

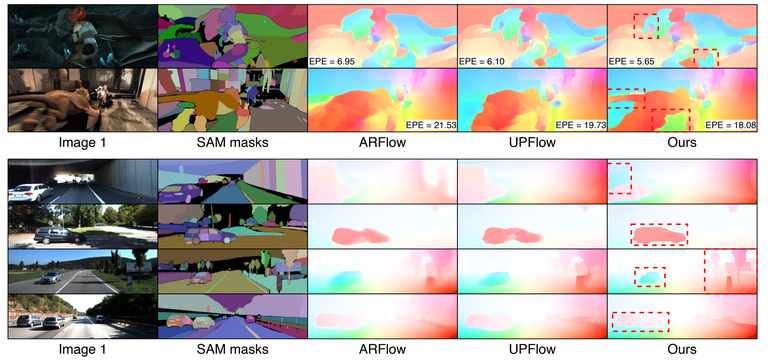


Experiments

Benchmark tests

	Method	Train Clean all	Train Final all	Test Clean all	Test Clean noc	Test Clean occ	Test Final all	Test Final noc	Test Final occ	Train 2012 EPE	Train 2015 EPE	Test 2012 Fl-noc	Test 2012 EPE	Test 2015 Fl-all	Test 2015 Fl-noc	Test 2015 Fl-bg	Test 2015 Fl-fg	Param.
Supervised	PWC-Net+ [55]	(1.71)	(2.34)	3.45	1.41	20.12	4.60	2.25	23.70	-	(1.50)	3.36	1.4	7.72	4.91	7.69	7.88	8.8M
	IRR-PWC [22]	(1.92)	(2.51)	3.84	1.47	23.22	4.58	2.15	24.36	-	(1.63)	3.21	1.6	7.65	4.86	7.68	7.52	6.4M
	RAFT [57]	(0.77)	(1.27)	1.61	0.62	9.65	2.86	1.41	14.68	-	(0.63)	-	-	5.10	3.07	4.74	6.87	5.3M
	FlowFormer [20]	(0.48)	(0.74)	1.16	0.42	7.16	2.09	**0.96**	11.30	-	(0.53)	-	-	4.68	**2.69**	**4.37**	**6.18**	18.2M
	SAMFlow [73]*†	-	-	**1.00**	**0.38**	**5.97**	**2.08**	1.04	**10.60**	-	-	-	-	**4.49**	-	-	-	-
Unsupervised	UnFlow-CSS [42]	-	7.91	9.38	5.37	42.11	10.22	6.06	44.11	3.29	8.10	-	-	23.27	-	-	-	116.6M
	DDFlow [34]	(2.92)	(3.98)	6.18	2.27	38.05	7.40	3.41	39.94	2.35	5.72	4.57	3.0	14.29	9.55	13.08	20.40	4.3M
	SelFlow [35]	(2.88)	(3.87)	6.56	2.67	38.30	6.57	3.12	34.72	1.69	4.84	4.31	2.2	14.19	9.65	12.68	21.74	4.8M
	SimFlow [23]	(2.86)	(3.57)	5.93	2.16	36.66	6.92	3.02	38.70	-	5.19	-	-	13.38	8.21	12.60	17.27	-
	ARFlow [33]	(2.79)	(3.73)	4.78	1.91	28.26	5.89	2.73	31.60	1.44	2.85	5.02	1.8	11.80	8.91	10.30	19.32	2.2M
	UFlow [26]	(2.50)	(3.39)	5.21	2.04	31.06	6.50	3.08	34.40	1.68	(2.71)	4.26	1.9	11.13	8.41	9.78	17.87	-
	UPFlow [39]	(2.33)	(2.67)	4.68	1.71	28.95	5.32	**2.42**	28.93	1.27	2.45	-	**1.4**	9.38	-	-	-	3.5M
	Ours (baseline)	(2.67)	(3.63)	4.29	1.64	25.96	5.81	2.76	30.60	1.32	2.44	4.05	1.6	9.60	6.77	8.74	**13.89**	2.5M
	Ours (+aug)*	(2.35)	(3.33)	4.00	**1.58**	23.76	5.33	2.53	28.17	1.33	2.26	4.15	1.6	9.05	6.46	7.96	14.55	2.5M
	Ours (+aug +hg)*	(2.25)	(3.10)	4.00	1.76	22.36	5.22	2.62	**26.40**	1.27	2.11	3.89	1.5	8.18	6.04	6.67	15.72	2.5M
	Ours (+aug +hg +mf)*†	(2.21)	(3.07)	**3.93**	1.67	**22.34**	**5.20**	2.56	26.75	**1.26**	**2.01**	**3.79**	**1.4**	**7.83**	**5.67**	**6.40**	14.98	2.6M
										Sintel [3]				KITTI [4]				

Qualitative examples



References

[1] Liu, L., *et al.*: Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In *CVPR*, pages 6489–6498, 2020.
[2] Yuan, S., *et al.*: Semarflow: Injecting semantics into unsupervised optical flow estimation for autonomous driving. In *ICCV*, pages 9566–9577, 2023.
[3] Butler, D., *et al.*: A naturalistic open source movie for optical flow evaluation. In *ECCV*, pages 611–625, 2012.
[4] Menze, M., *et al.*: Object scene flow for autonomous vehicles. In *CVPR*, pages 3061–3070, 2015.

Paper

Code