SemARFlow: Injecting Semantics into Unsupervised Optical Flow Estimation for Autonomous Driving

Shuai Yuan, *Shuzhi Yu*, *Hannah Kim*, and *Carlo Tomasi*
Department of Computer Science, Duke University, USA

ICCV23 PARIS

Introduction

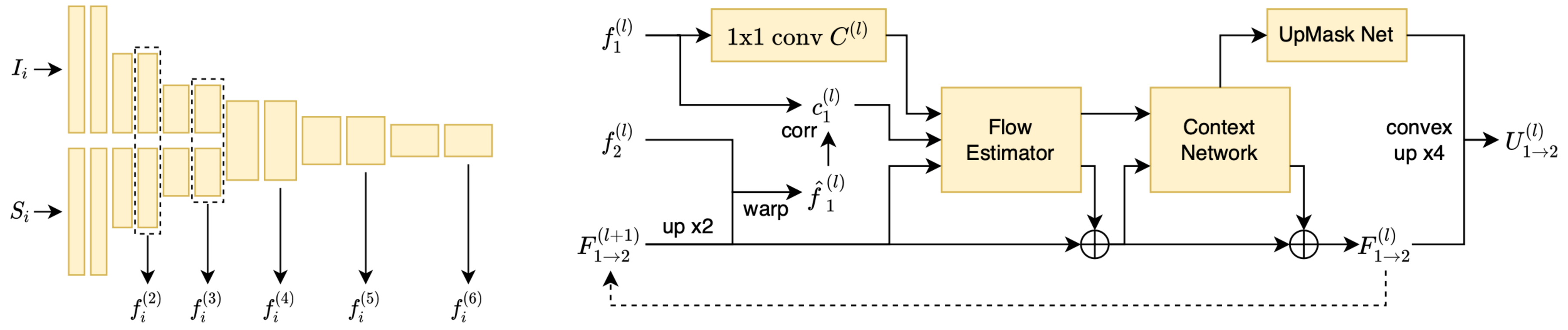
- **Unsupervised optical flow estimation : poorly constrained** (occlusions, motion boundaries, poor texture, illumination change, etc.)
- **Can we inject additional information to help?** Example: semantics and domain knowledge
- **Autonomous driving**: labeling flow is hard, but labeling semantics is feasible (and indeed available)
- **SemARFlow**: add **Sem**antic Segmentation inputs (estimated by trained semantic models) to an unsupervised optical flow network (**ARFlow**[1])
- Improved performance; sharper edges; better results on vehicles; better generalization ability

Network Architecture

Adapted from ARFlow[1]

- +enc: adding semantic encoder
- +up: adding learned upsampler
- +no_sm: turning off smoothness loss

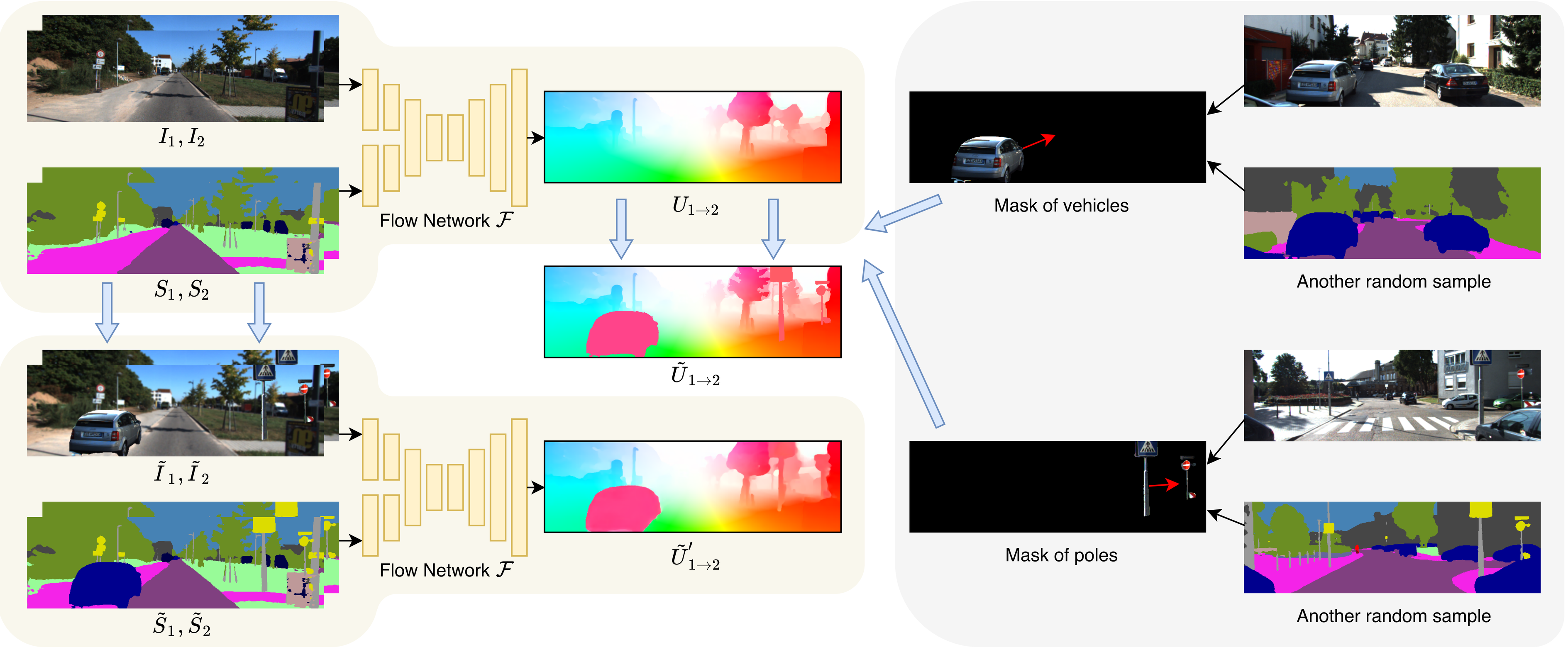
Code: https://github.com/duke-vision/semantic-unsup-flow-release



(a) shared pyramidal encoder ($i \in \{1, 2\}$)

(b) one iteration (at the lth-level) of the iterative decoder ($l \in \{6, 5, 4, 3, 2\}$)

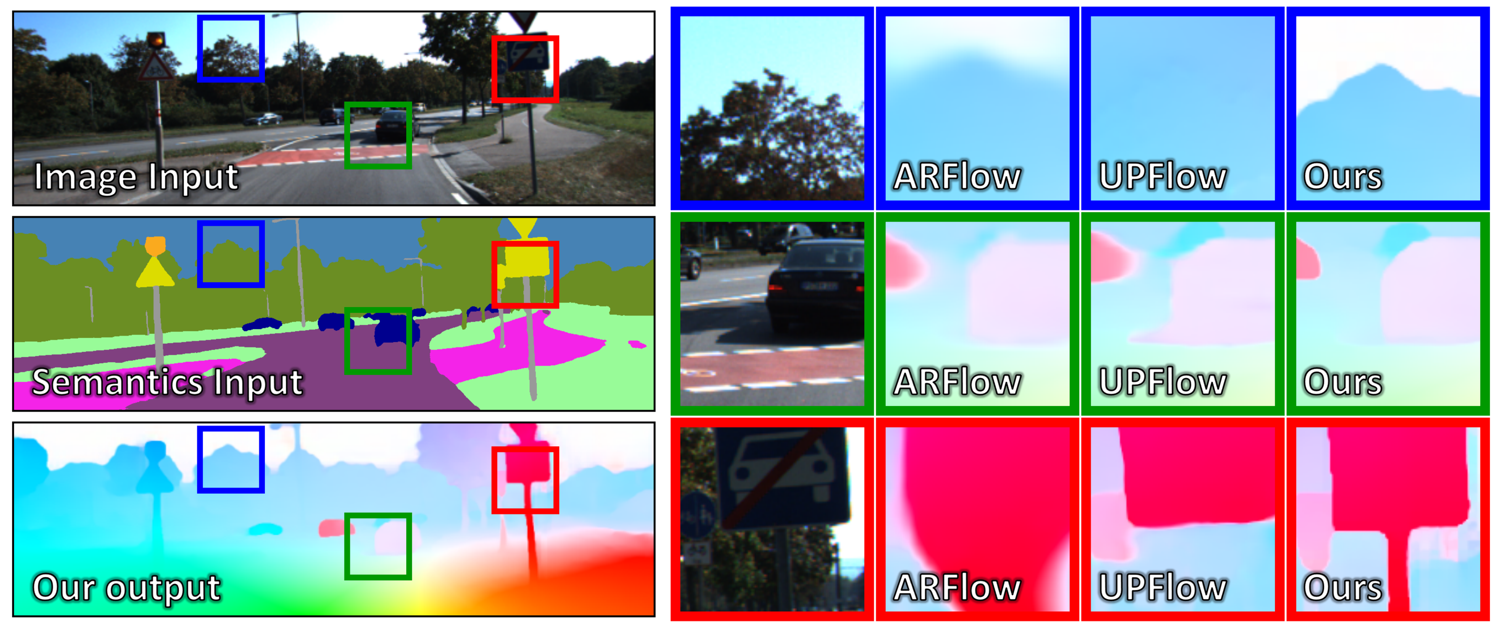
Semantic Augmentation as Self-Supervision



1. Estimate flow in the first forward pass
2. Extract vehicles and poles from other random samples
3. Copy and paste vehicles/poles with augmented motions to generate augmented samples and pseudo-labels
4. Estimate flow for the augmented samples in the second forward pass, self-supervised by pseudo-labels

Demo

Our method outputs **sharper edges around objects**



Evaluation

We **greatly outperform state-of-the-arts** on KITTI [2]

	Method	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
supervised	PWC-Net+ [66]	-	(1.50)	3.36	1.4	7.72	4.91	7.69	7.88
	IRR-PWC [22]	-	(1.63)	3.21	1.6	7.65	4.86	7.68	7.52
	RAFT [68]	-	(0.63)	-	-	5.10	3.07	4.74	6.87
	Separable Flow [85]	-	(0.69)	-	-	4.53	2.78	4.25	5.92
unsupervised	SelFlow [39]	1.69	4.84	4.31	2.2	14.19	9.65	12.68	21.74
	SimFlow [24]	-	5.19	-	-	13.38	8.21	12.60	17.27
	ARFlow [37]	1.44	2.85	-	1.8	11.80	-	-	-
	UFlow [27]	1.68	2.71	4.26	1.9	11.13	8.41	9.78	17.87
	UPFlow [42]	**1.27**	2.45	-	**1.4**	9.38	-	-	-
	Ours (baseline)	1.39	2.61	4.30	1.7	9.89	6.98	8.82	15.21
	Ours (+enc)†	1.29	2.42	3.97	1.5	8.99	3.97	8.19	13.01
	Ours (+enc +aug)†	1.28	**2.18**	**3.90**	1.5	**8.38**	**3.90**	**7.48**	**12.91**

Generalization: train on Cityscapes[3], test on KITTI[2]

Method	Fl-all	EPE-all	EPE-noc	EPE-occ
ARFlow (our impl.)	13.21	4.08	2.88	9.40
Ours (baseline)	12.27	3.81	2.43	9.91
Ours (+enc)†	11.28	3.33	2.12	8.75
Ours (+enc +aug)†	**10.32**	**2.64**	**1.56**	**7.12**

Other Info

Full Paper

Code

Video

Acknowledgments:
This research is based upon work supported by the National Science Foundation under Grant No. 1909821.

[1] Liu, L., *et al.*: Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6489–6498 (2020)
[2] Menze, M. and Geiger, A.: Object scene flow for autonomous vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
[3] Cordts, M., *et al.*: The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3213–3223 (2016)