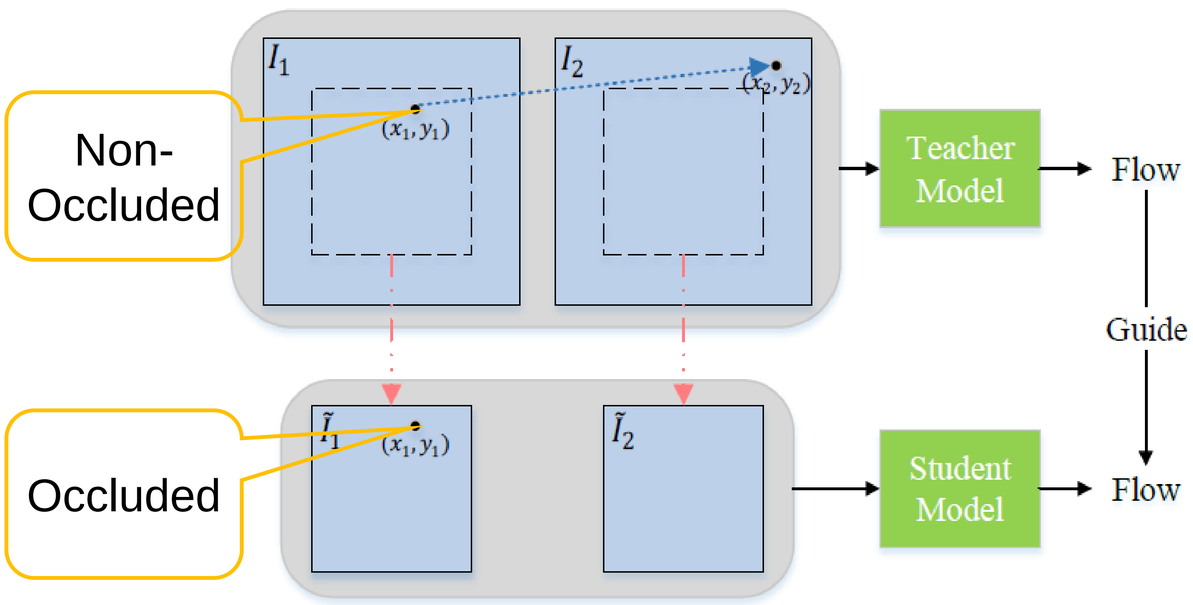
DDFlow: Learning Optical Flow with Unlabeled Data Distillation

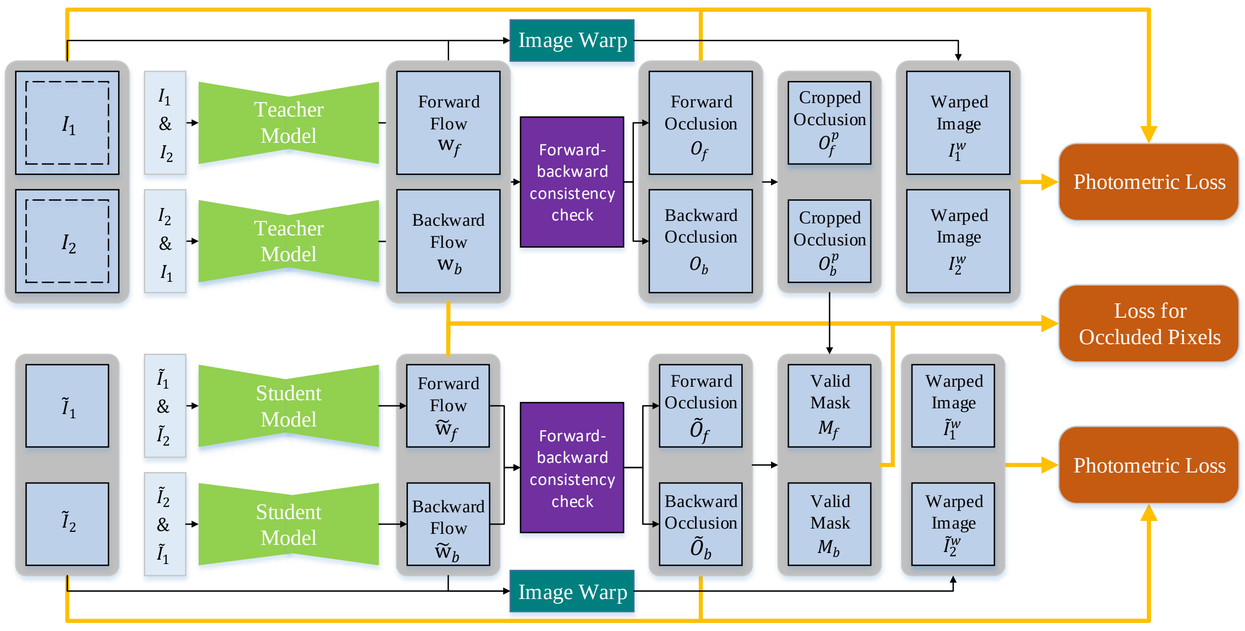
Pengpeng Liu[1], Irwin King[1], Michael Lyu[1], Jia Xu[2]
[1] The Chinese University of Hong Kong [2] Tencent AI Lab

Introduction

- Optical Flow: motion of pixels between two images
- Challenges
  - Traditional methods: time costing, complex
  - CNNs: need a large amount of labeled data, difficult to obtain
  - Pre-train on synthetic dataset ➔ domain gap
- Unsupervised Learning
  - Photometric loss: measure the difference between reference image and warped target image
  - Detect occlusion and exclude occluded pixels
  - Produce reliable optical flow for **non-occluded** pixels, but lack the ability to learn the flow of **occluded** pixels
- How to fully utilize those reliable non-occluded predictions?



Our Approach



- L_o only functions on pixels that are non-occluded in original images but occluded in cropped patches
- Occlusion estimation: forward-backward consistency check

$$\begin{cases} |\mathrm{w}_f + \hat{\mathrm{w}}_f|^2 < \alpha_1(|\mathrm{w}_f|^2 + |\hat{\mathrm{w}}_f|^2) + \alpha_2, \\ \mathbf{p} + \mathrm{w}_f(\mathbf{p}) \in \Omega, \end{cases}$$

- Photometric loss

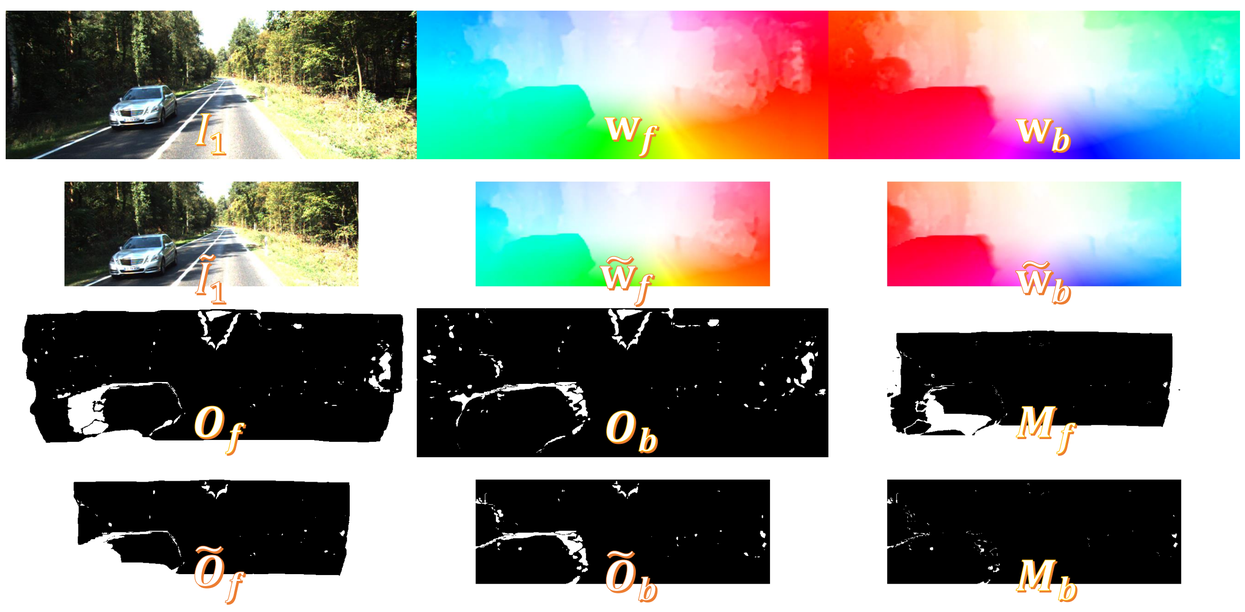
$$L_p = \sum \psi(I_1 - I_2^w) \odot (1 - O_f) / \sum (1 - O_f) + \sum \psi(I_2 - I_1^w) \odot (1 - O_b) / \sum (1 - O_b)$$

- Loss for occluded pixels

$$M_f = \mathrm{clip}(\widetilde{O}_f - O_f^p, 0, 1)$$

$$L_o = \sum \psi(\mathrm{w}_f^p - \tilde{\mathrm{w}}_f) \odot M_f / \sum M_f + \sum \psi(\mathrm{w}_b^p - \tilde{\mathrm{w}}_b) \odot M_b / \sum M_b$$

- Example intermediate results

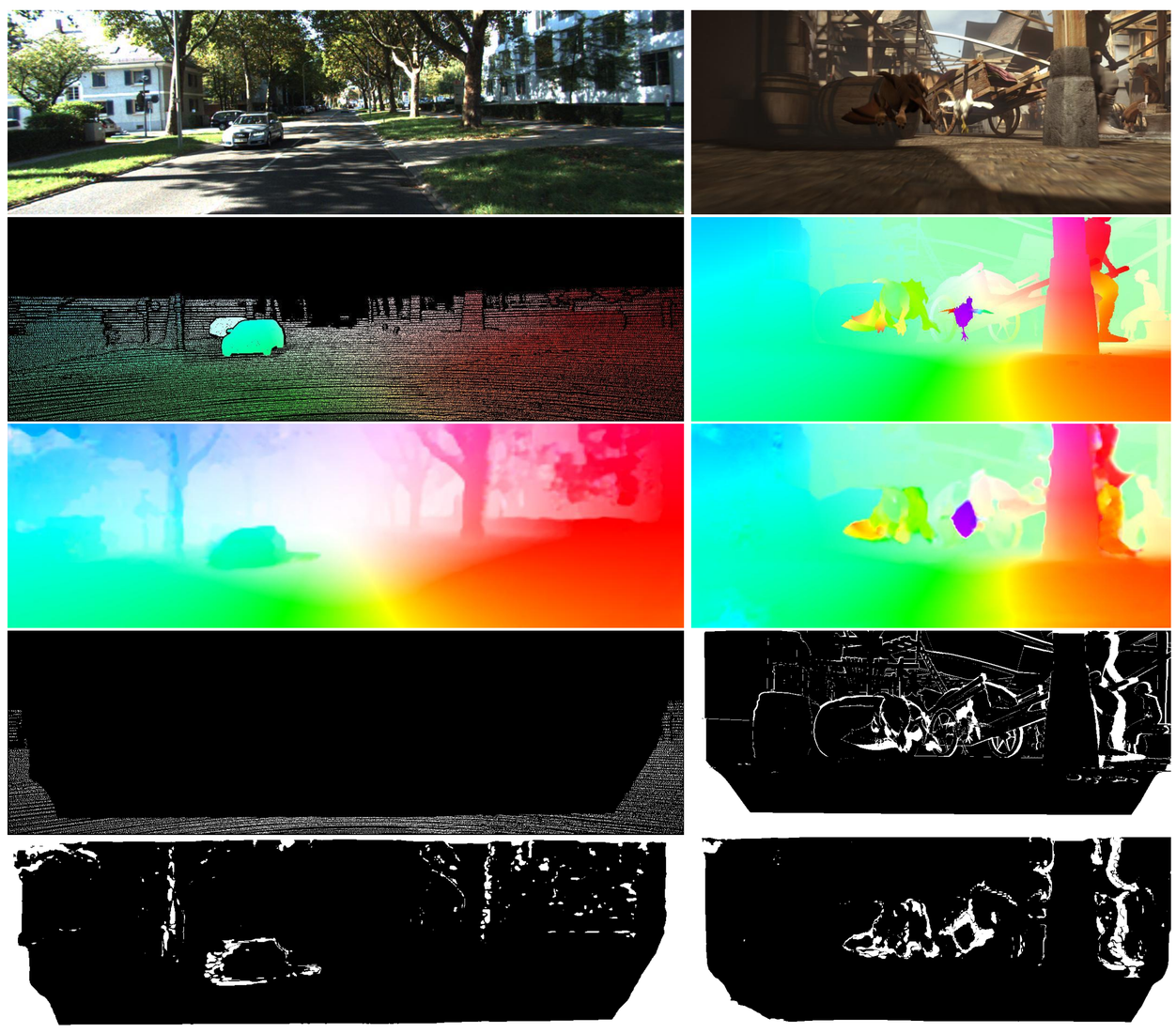


Experiment

	Method	Chairs	Sintel Clean		Sintel Final		KITTI 2012			KITTI 2015	
		test	train	test	train	test	train	test	Fl-noc	train	Fl-all
Supervise	FlowNetS (Dosovitskiy et al. 2015)	2.71	4.50	7.42	5.45	8.43	8.26	–	–	–	–
	FlowNetS+ft (Dosovitskiy et al. 2015)	–	(3.66)	6.96	(4.44)	7.76	7.52	9.1	–	–	–
	SpyNet (Ranjan and Black 2017)	2.63	4.12	6.69	5.57	8.43	9.12	–	–	–	–
	SpyNet+ft (Ranjan and Black 2017)	–	(3.17)	6.64	(4.32)	8.36	8.25	10.1	12.31%	–	35.07%
	FlowNet2 (Ilg et al. 2017)	–	2.02	3.96	3.14	6.02	4.09	–	–	10.06	–
	FlowNet2+ft (Ilg et al. 2017)	–	(1.45)	4.16	(2.01)	5.74	(1.28)	1.8	4.82%	(2.3)	11.48%
	PWC-Net (Sun et al. 2018)	**2.00**	3.33	–	4.59	–	4.57	–	–	13.20	–
	PWC-Net+ft (Sun et al. 2018)	–	**(1.70)**	**3.86**	**(2.21)**	**5.13**	**(1.45)**	**1.7**	**4.22%**	**(2.16)**	**9.60%**
Unsupervise	BackToBasic+ft (Jason, Harley, and Derpanis 2016)	5.3	–	–	–	–	11.3	9.9	–	–	–
	DSTFlow+ft (Ren et al. 2017)	5.11	(6.16)	10.41	(6.81)	11.27	10.43	12.4	–	16.79	39%
	UnFlow-CSS+ft (Meister, Hur, and Roth 2018)	–	–	–	(7.91)	10.22	3.29	–	–	8.10	23.30%
	OccAwareFlow (Wang et al. 2018)	3.30	5.23	8.02	6.34	9.08	12.95	–	–	21.30	–
	OccAwareFlow+ft-Sintel (Wang et al. 2018)	3.76	(4.03)	7.95	(5.95)	9.15	12.9	–	–	22.6	–
	OccAwareFlow-KITTI (Wang et al. 2018)	–	7.41	–	7.92	–	3.55	4.2	–	8.88	31.2%
	MultiFrameOccFlow-Hard+ft (Janai et al. 2018)	–	(6.05)	–	(7.09)	–	–	–	–	6.65	–
	MultiFrameOccFlow-Soft+ft (Janai et al. 2018)	–	(3.89)	7.23	(5.52)	8.81	–	–	–	6.59	22.94%
	DDFlow	**2.97**	3.83	–	4.85	–	8.27	–	–	17.26	–
	DDFlow+ft-Sintel	3.46	**(2.92)**	**6.18**	**(3.98)**	**7.40**	5.14	–	–	12.69	–
	DDFlow+ft-KITTI	6.35	6.20	–	7.08	–	**2.35**	**3.0**	**4.57%**	**5.72**	**14.29%**

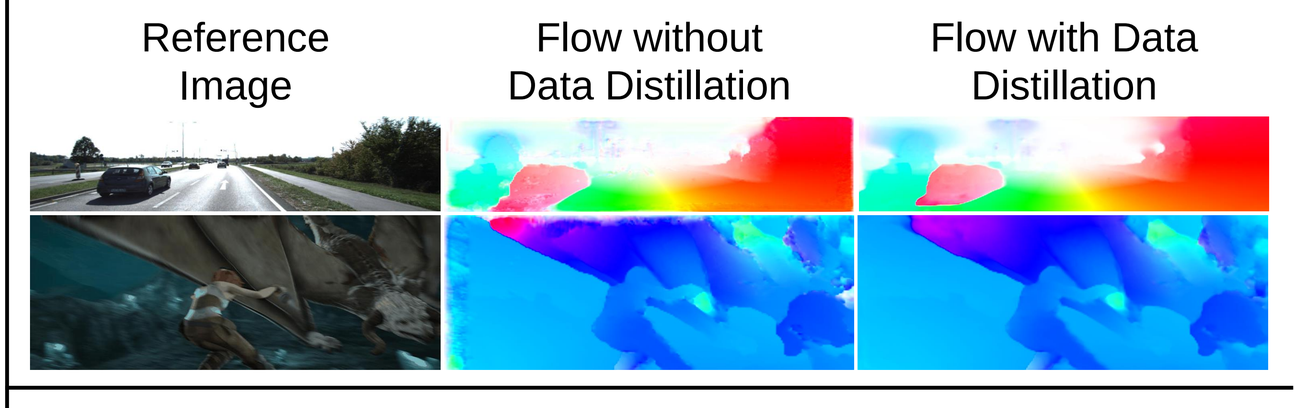
F-score for Occlusion Estimation

Method	Sintel Clean	Sintel Final	KITTI 2012	KITTI 2015
MODOF	–	0.48	–	–
OccAwareFlow-ft	(0.54)	(0.48)	**0.95***	0.88*
MultiFrameOccFlow-Soft+ft	(0.49)	(0.44)	–	**0.91***
Ours	**(0.59)**	**(0.52)**	0.94*	0.86*



Ablation Study

Occlusion Handling	Census Transform	Data Distillation	Chairs	Sintel Clean			Sintel Final			KITTI 2012			KITTI 2015		
			ALL	ALL	NOC	OCC	ALL	NOC	OCC	ALL	NOC	OCC	ALL	NOC	OCC
✗	✗	✗	4.06	(5.05)	(2.45)	(38.09)	(7.54)	(4.81)	(42.46)	10.76	3.35	59.86	16.85	6.45	82.64
✓	✗	✗	3.95	(4.45)	(2.16)	(33.48)	(6.56)	(4.12)	(37.83)	6.67	1.94	38.01	12.42	5.67	60.59
✗	✓	✗	3.75	(3.90)	(1.60)	(33.31)	(5.23)	(2.80)	(36.35)	8.66	1.47	56.24	14.04	4.06	77.16
✓	✓	✗	3.24	(3.37)	(1.34)	(29.36)	(4.47)	(2.32)	(31.86)	4.50	1.10	27.04	8.01	3.02	42.66
✓	✓	✓	**2.97**	**(2.92)**	**(1.27)**	**(23.92)**	**(3.98)**	**(2.21)**	**(26.74)**	**2.35**	**1.02**	**11.31**	**5.72**	**2.73**	**24.68**



Summary

- Propose a data distillation approach to learn optical flow from unlabeled data and can predict the optical flow of occluded pixels
- Achieve the highest accuracy among all prior unsupervised methods on all challenging optical flow benchmarks including Flying Chairs, KITTI 2012, KITTI 2015 and Sintel
- Code and models available on https://github.com/ppliuboy/DDFlow