Learning Optical Flow with Limited Data

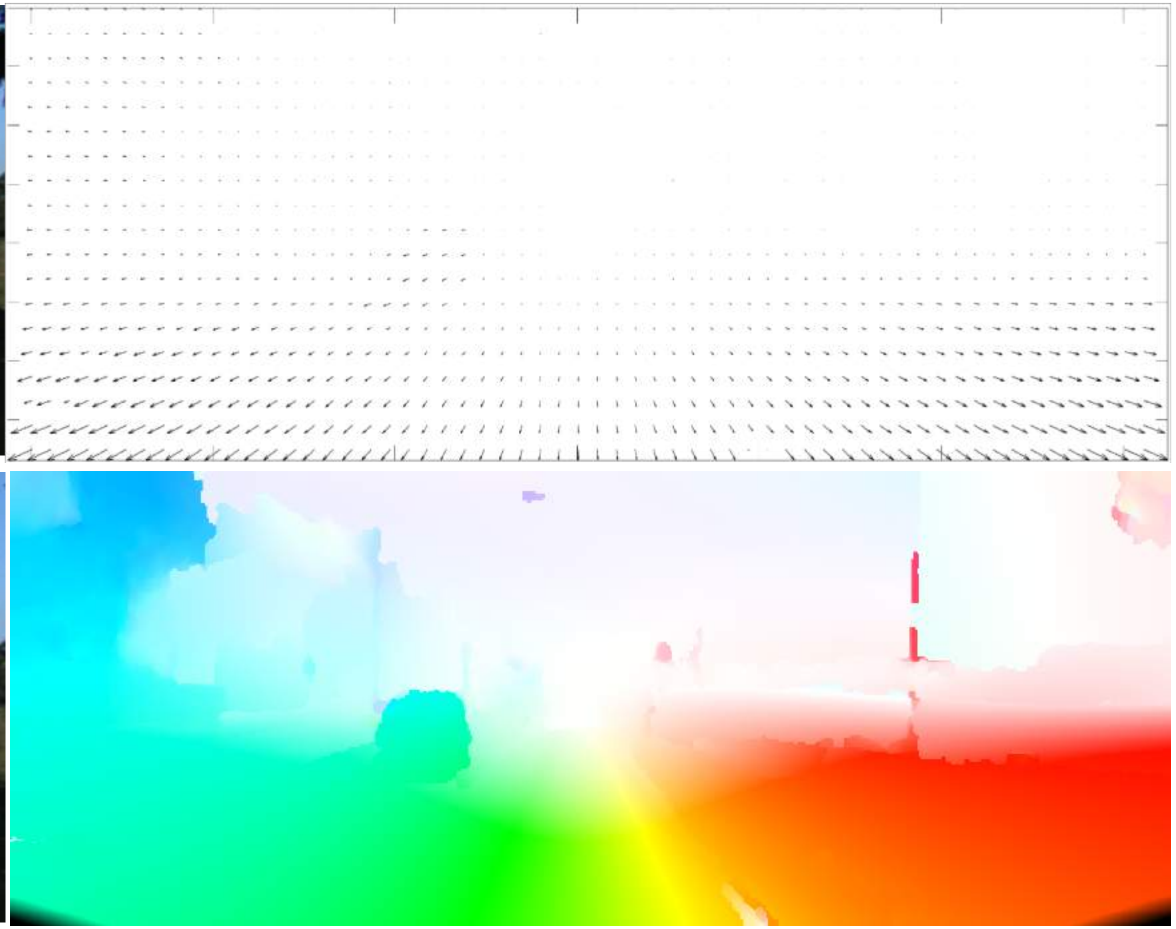
Jia Xu (许佳)
Tencent AI Lab

2019-03-14

Introduction

Input

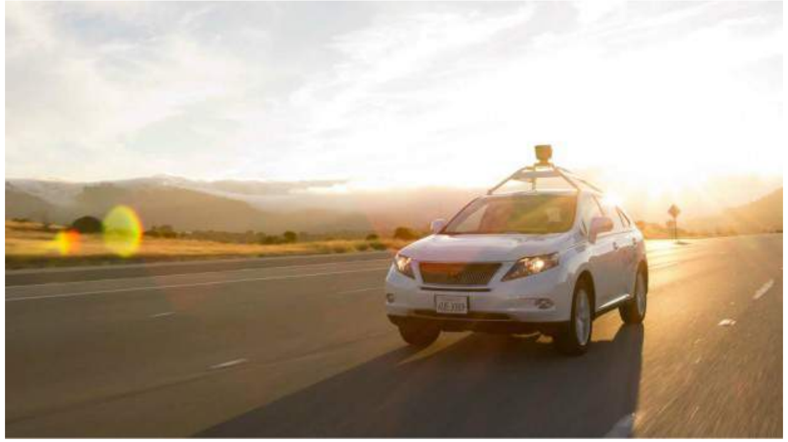
Output



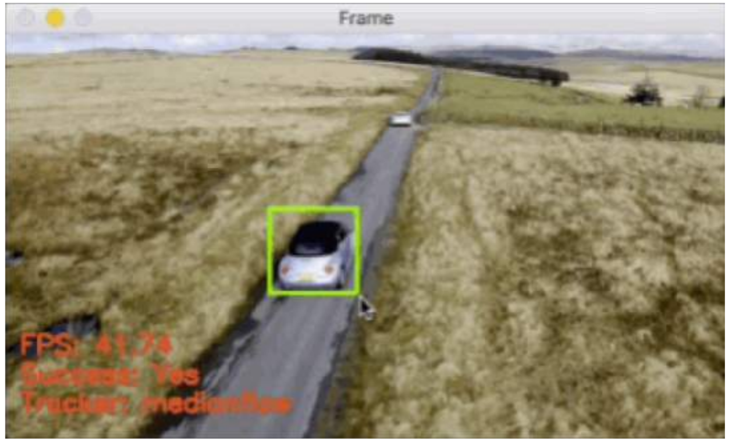
Dense correspondence for each pixel between two frames

Why Optical Flow?

- Optical flow has a wide range of applications.



Autonomous Driving

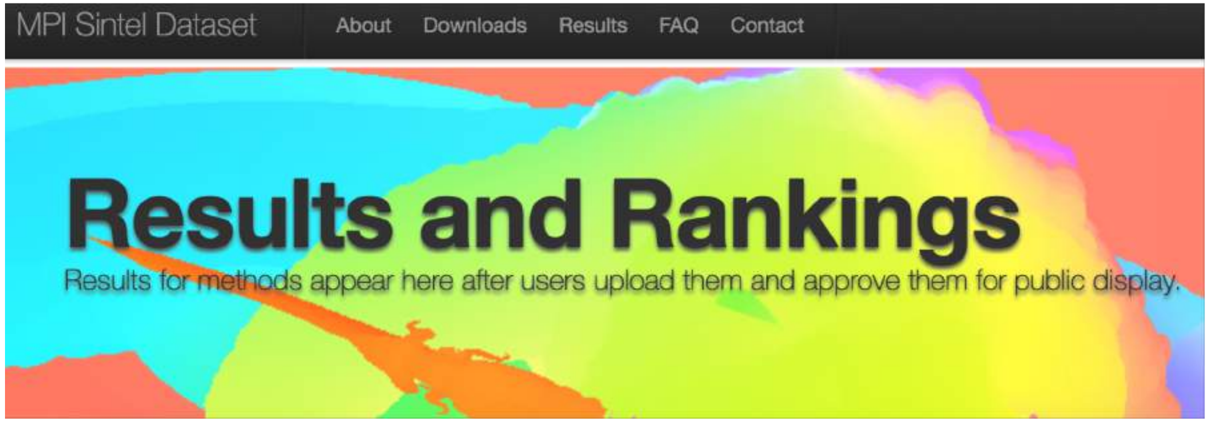


Object Tracking

3D Shape Reconstruction

Video Action Recognition

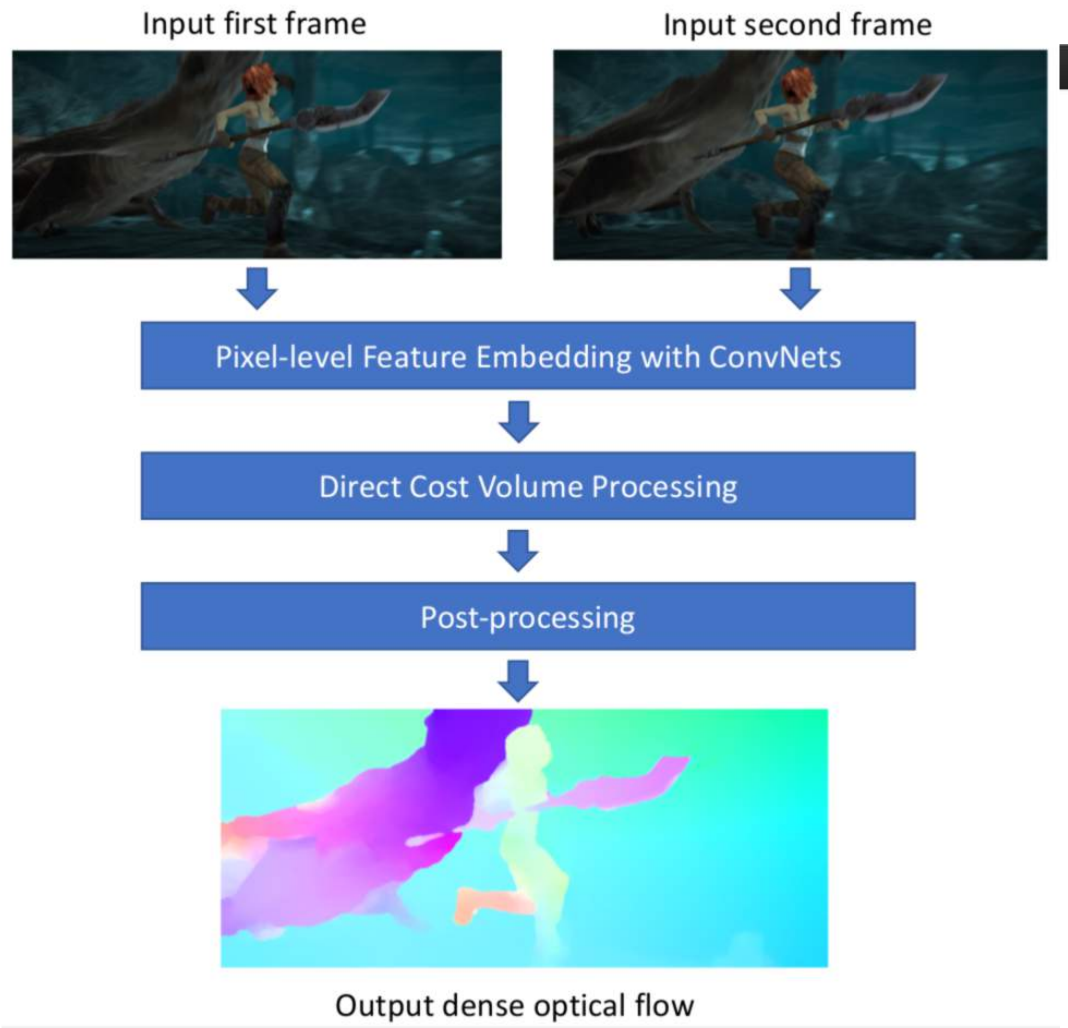
History of Optical Flow Estimation

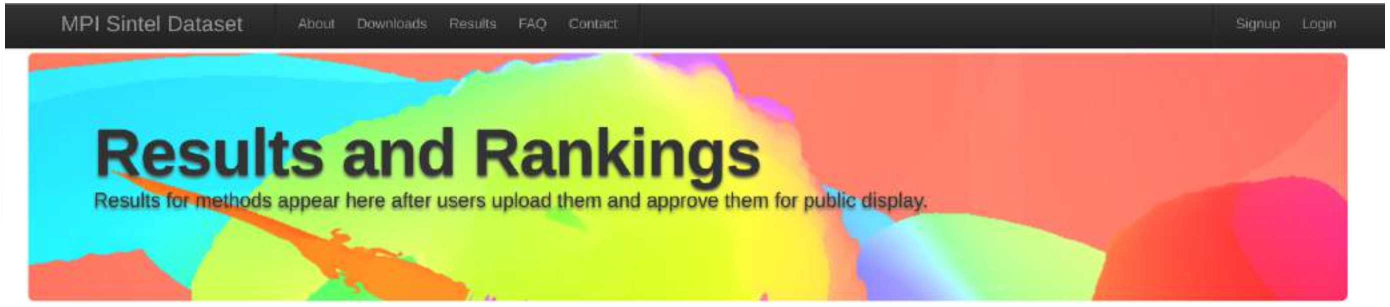


Horn and Schunck
Brox et al.
Sun et al.
EpicFlow
FlowNet
FullFlow
DCFlow
PWC-Net
DDFlow
SelFlow

1981
2004
2010
2015
2016
2017
2018
2019

DC Flow





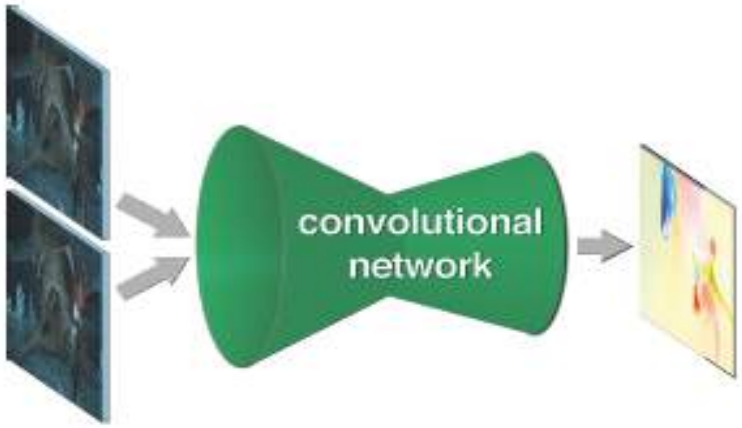
Final Clean

	EPE all	EPE matched	EPE unmatched	d0-10	d10-60	d60-140	s0-10	s10-40	s40+	
GroundTruth [1]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Visualize Results
DCFlow [2]	5.119	2.283	28.228	4.665	2.108	1.440	1.052	3.434	29.351	Visualize Results
FlowFieldsCNN [3]	5.363	2.303	30.313	4.718	2.020	1.399	1.032	3.065	32.422	Visualize Results
MR-Flow [4]	5.376	2.818	26.235	5.109	2.395	1.755	0.908	3.443	32.221	Visualize Results
S2F-IF [5]	5.417	2.549	28.795	4.745	2.198	1.712	1.157	3.468	31.262	Visualize Results
InterpoNet_ff [6]	5.535	2.372	31.296	4.720	2.018	1.532	1.064	3.496	32.633	Visualize Results
RegionalFF [7]	5.562	2.595	29.741	4.921	2.393	1.639	1.122	3.477	32.625	Visualize Results
PGM-C [8]	5.591	2.672	29.389	4.975	2.340	1.791	1.057	3.421	33.339	Visualize Results
RicFlow [9]	5.620	2.765	28.907	5.146	2.366	1.679	1.088	3.364	33.573	Visualize Results
InterpoNet_cpm [10]	5.627	2.594	30.344	4.975	2.213	1.640	1.042	3.575	33.321	Visualize Results
CPM_AUG [11]	5.645	2.737	29.362	4.707	2.150	1.918	1.087	3.306	33.925	Visualize Results
ProbFlowFields [12]	5.707	2.545	31.471	4.730	2.153	1.666	1.126	3.688	33.303	Visualize Results

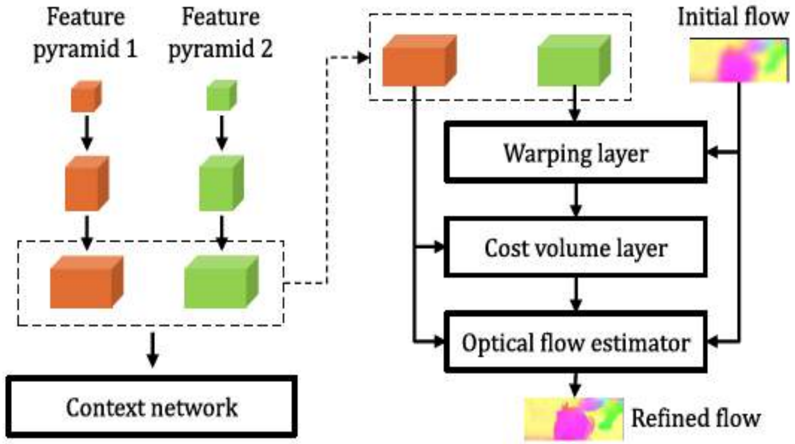
Xu, Ranftl, Koltun. Accurate Optical Flow via Direct Cost Volume Processing. CVPR 2017

CNNs for Optical Flow

- ❑ Advantage: high performance while running at real time.
- ❑ Disadvantage: need a large amount of labeled data ➔ **difficult** to obtain.



FlowNet



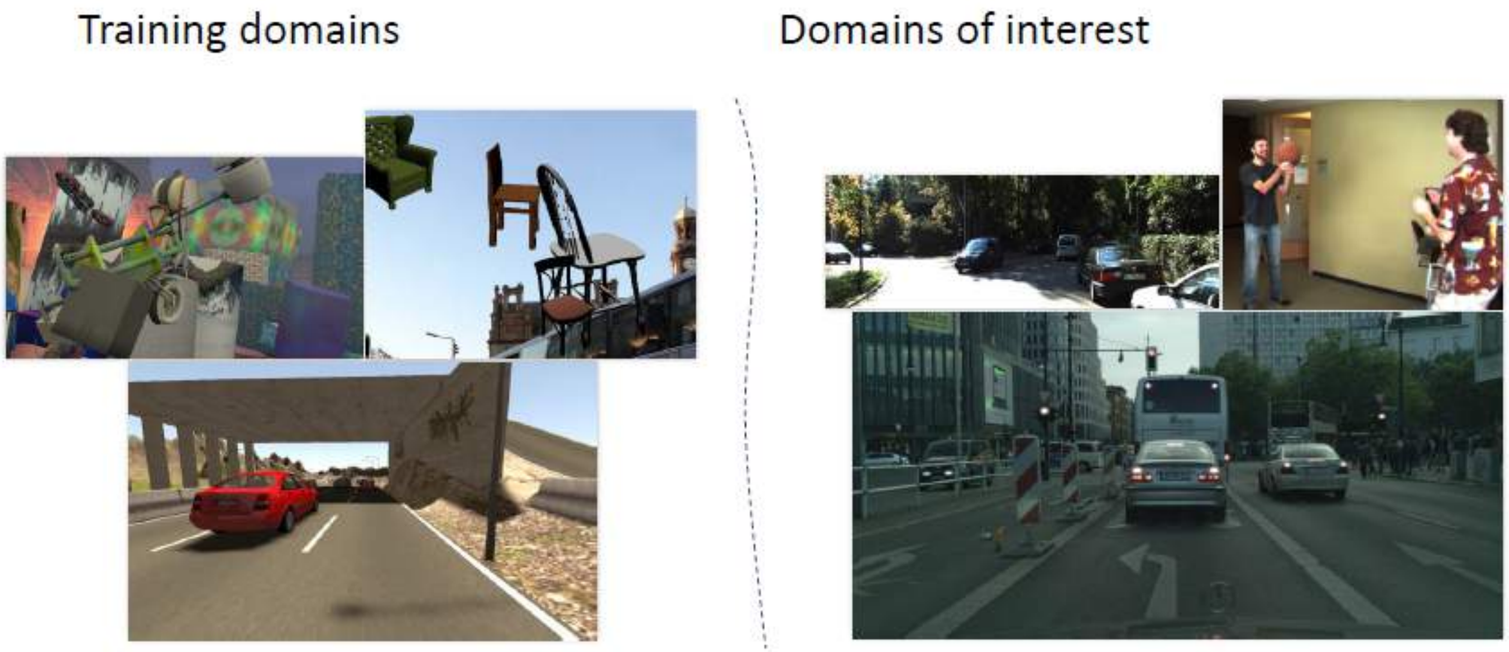
PWC-Net

Fischer et al. 2015, "FlowNet: Learning Optical Flow with Convolutional Networks"

Sun et al. 2018, "PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume"

CNNs for Optical Flow

- ❑ Advantage: high performance while running at real time.
- ❑ Disadvantage: need a large amount of labeled data → **difficult** to obtain.
  - ■ Pre-training on synthetic dataset: domain gap.
  - ■ Unsupervised learning: performance gap, cannot predict flow of occluded pixels.



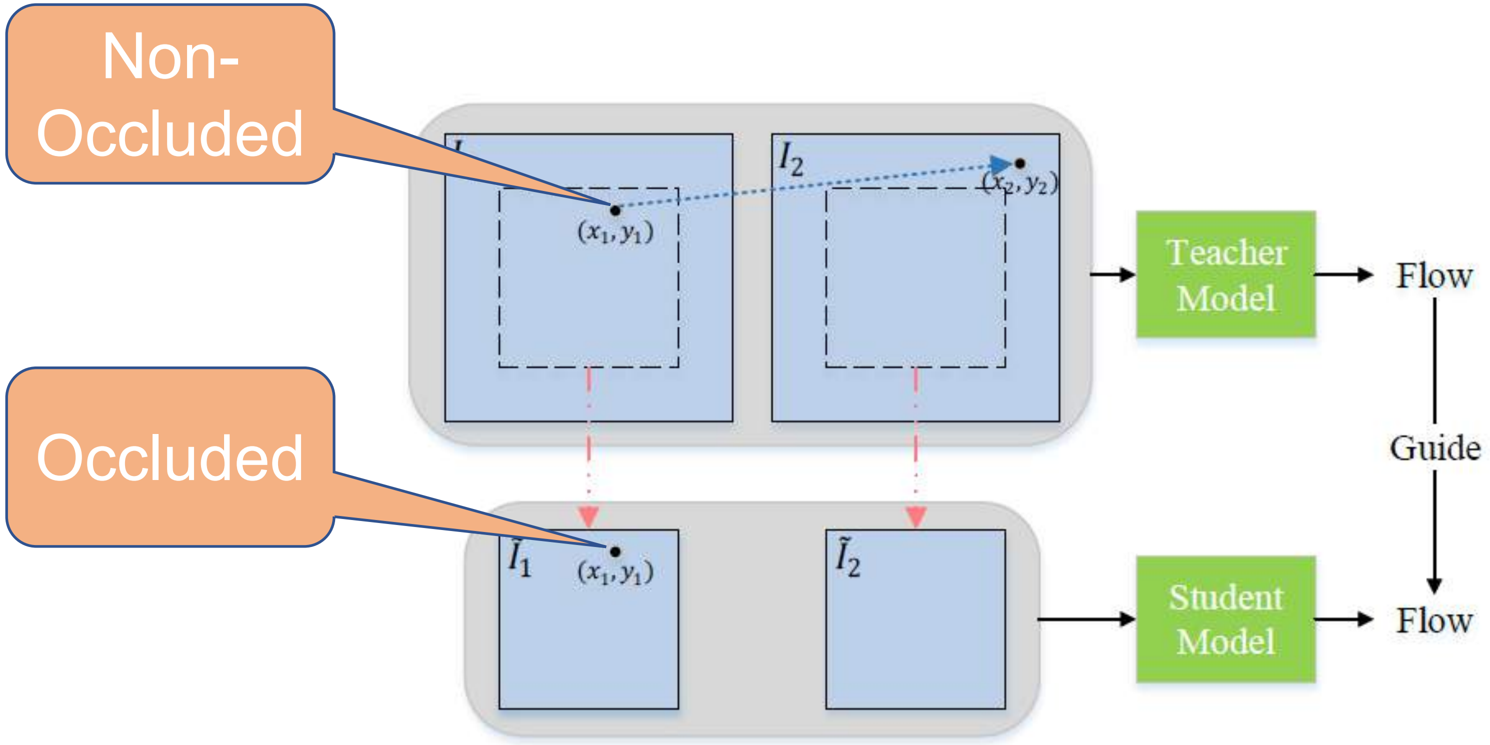
Meister et al. 2018, “UnFlow: Unsupervised Learning of Optical Flow with a Bidirectional Census Loss"

Unsupervised Learning for Optical Flow

How to learn optical flow of **occluded** pixels in a totally unsupervised way?

Key Observation

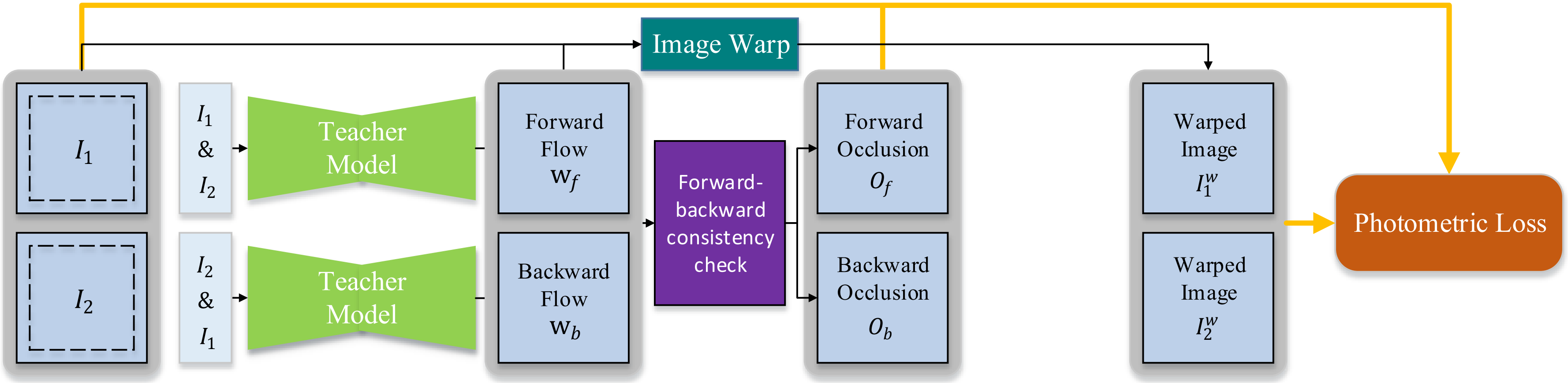
- Unsupervised Learning: **detect** occlusion and **exclude** occluded pixels.
  - The optical flow of **non-occluded** pixels can be **accurately** estimated.
  - How do we fully utilize those reliable non-occluded predictions?
  - Data Distillation!



Liu, King, Lyu, **Xu.** DDFlow: Learning Optical Flow with Unlabeled Data Distillation. AAAI 2019

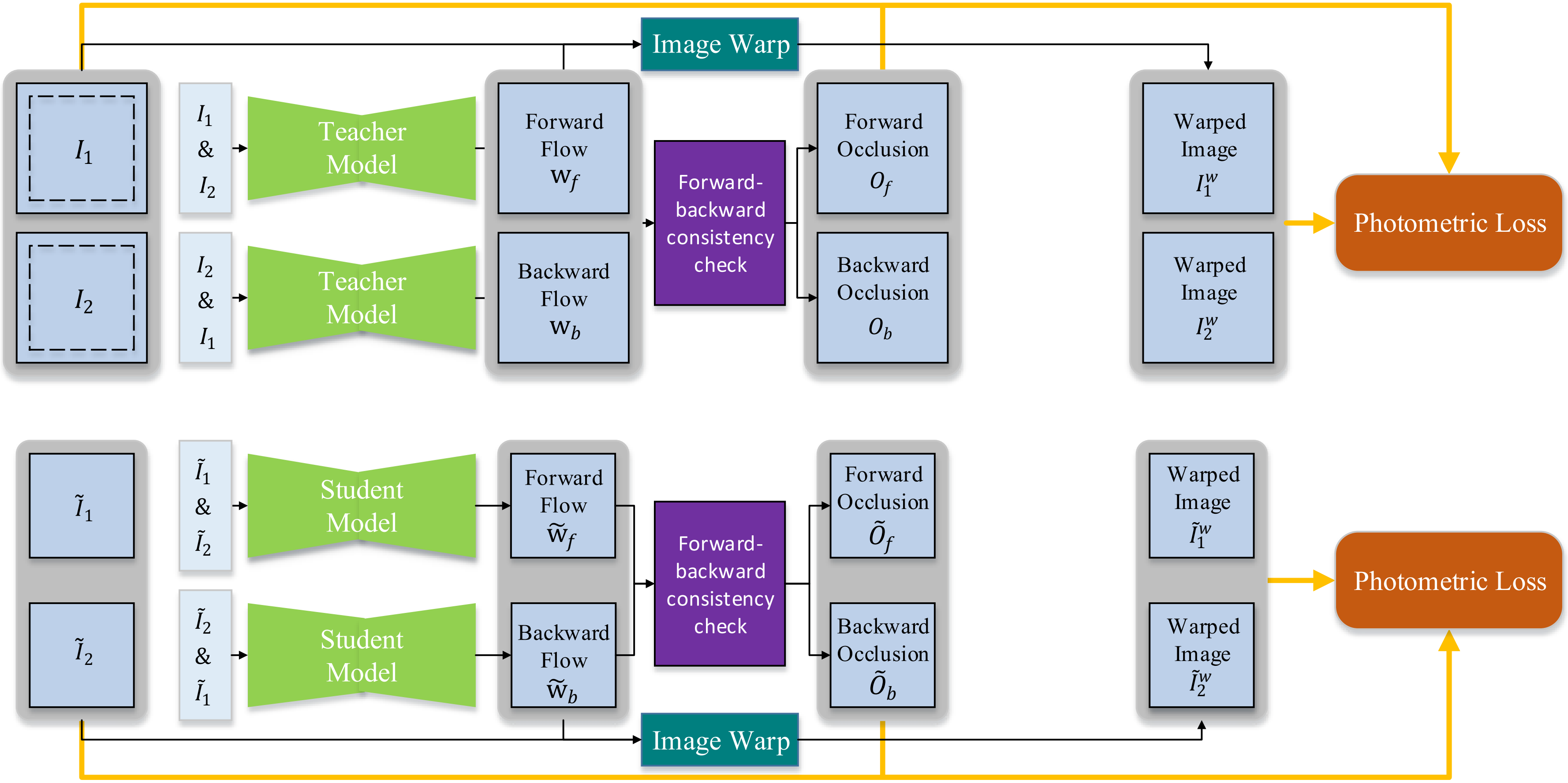
Framework

- Teacher model is trained with photometric loss L_p for non-occluded pixels.



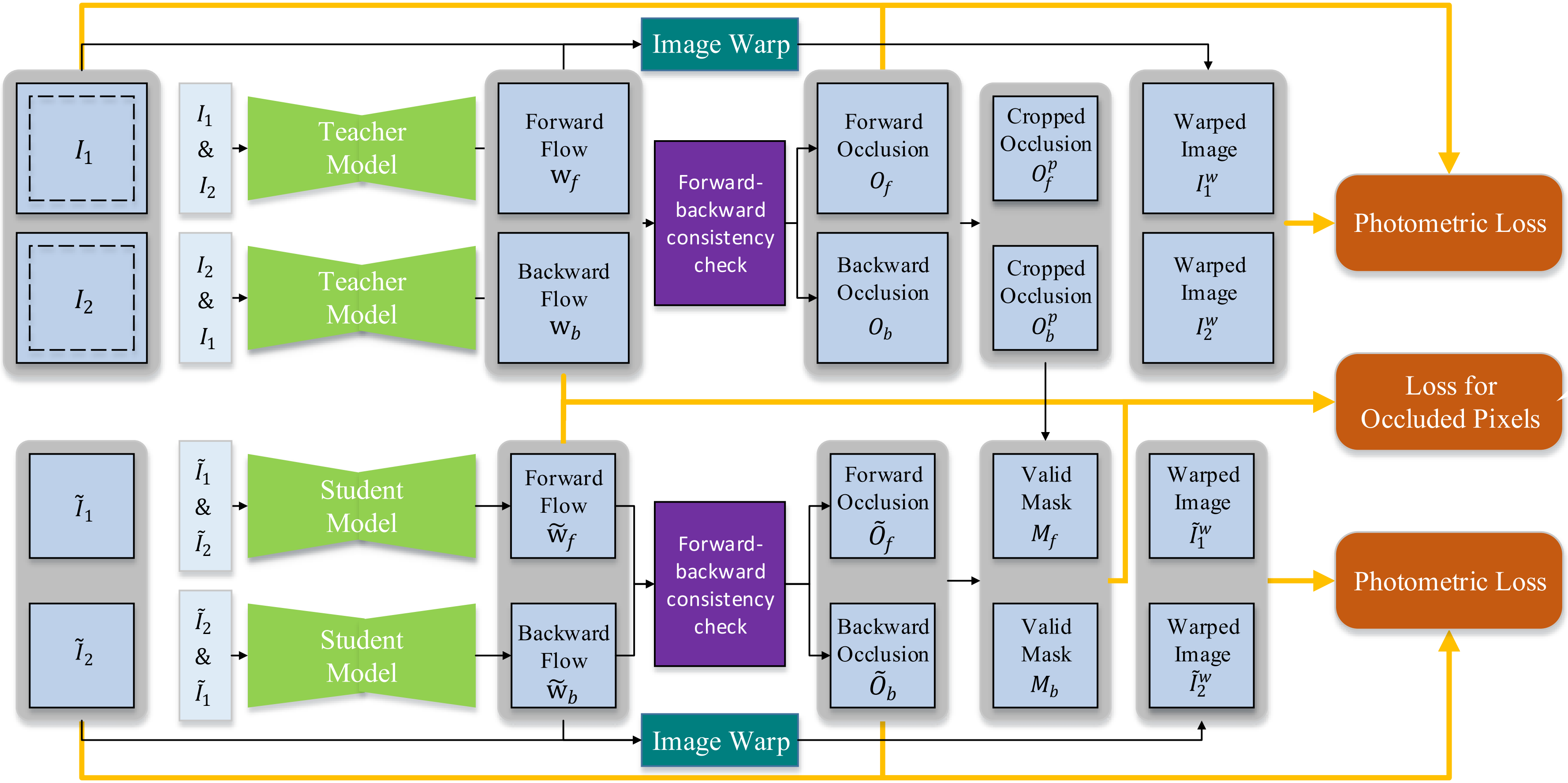
Framework

- ❑ Student model has the same network structure as teacher model.



Framework

- Student model is trained with both L_p for non-occluded pixels and L_o for occluded pixels. Only student model is needed during testing.



L_o only functions on pixels that are non-occluded in original images but occluded in cropped patches.

Loss Functions

- Occlusion estimation: based on the forward-backward consistency prior

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

- Photometric loss L_p

$$\begin{aligned} L_p = & \sum \psi(I_1 - I_2^w) \odot (1 - O_f) / \sum (1 - O_f) \\ & + \sum \psi(I_2 - I_1^w) \odot (1 - O_b) / \sum (1 - O_b) \end{aligned}$$

- Loss for occluded pixels L_o

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

$$\begin{aligned} L_o = & \sum \psi(\mathrm{w}_f^p - \widetilde{\mathrm{w}}_f) \odot M_f / \sum M_f \\ & + \sum \psi(\mathrm{w}_b^p - \widetilde{\mathrm{w}}_b) \odot M_b / \sum M_b \end{aligned}$$

- Teacher model: $L = L_p$
- Student Model: $L = L_p + L_o$

No hyperparameter !

Evaluation Metrics

- Optical Flow
  - EPE: average endpoint error between the predicted flow and the ground truth flow over all pixels.
  - Fl: percentage of erroneous pixels. A pixel is considered to be correctly estimated if flow end-point error is < 3 pixels or <5%.

- Occlusion estimation
  - F-score: the harmonic average of the precision and recall.

Quantitative Comparisons

- DDFlow outperforms all existing unsupervised flow learning methods on all datasets.

	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%**

Quantitative Comparisons

- Our pre-trained model on Flying Chairs even outperforms the finetuned state-of-the-art unsupervised models on Sintel dataset.

	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%**

Quantitative Comparisons

- 28.6 % relative improvement on KITTI 2012, 37.7% relative improvement on KITTI 2015.

	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%**

Quantitative Comparisons

- 28.6 % relative improvement on KITTI 2012, 37.7% relative improvement on KITTI 2015.
- On KITTI 2012, DDFlow outperforms Flownet 2.0 for ranking metric Fl-noc.

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

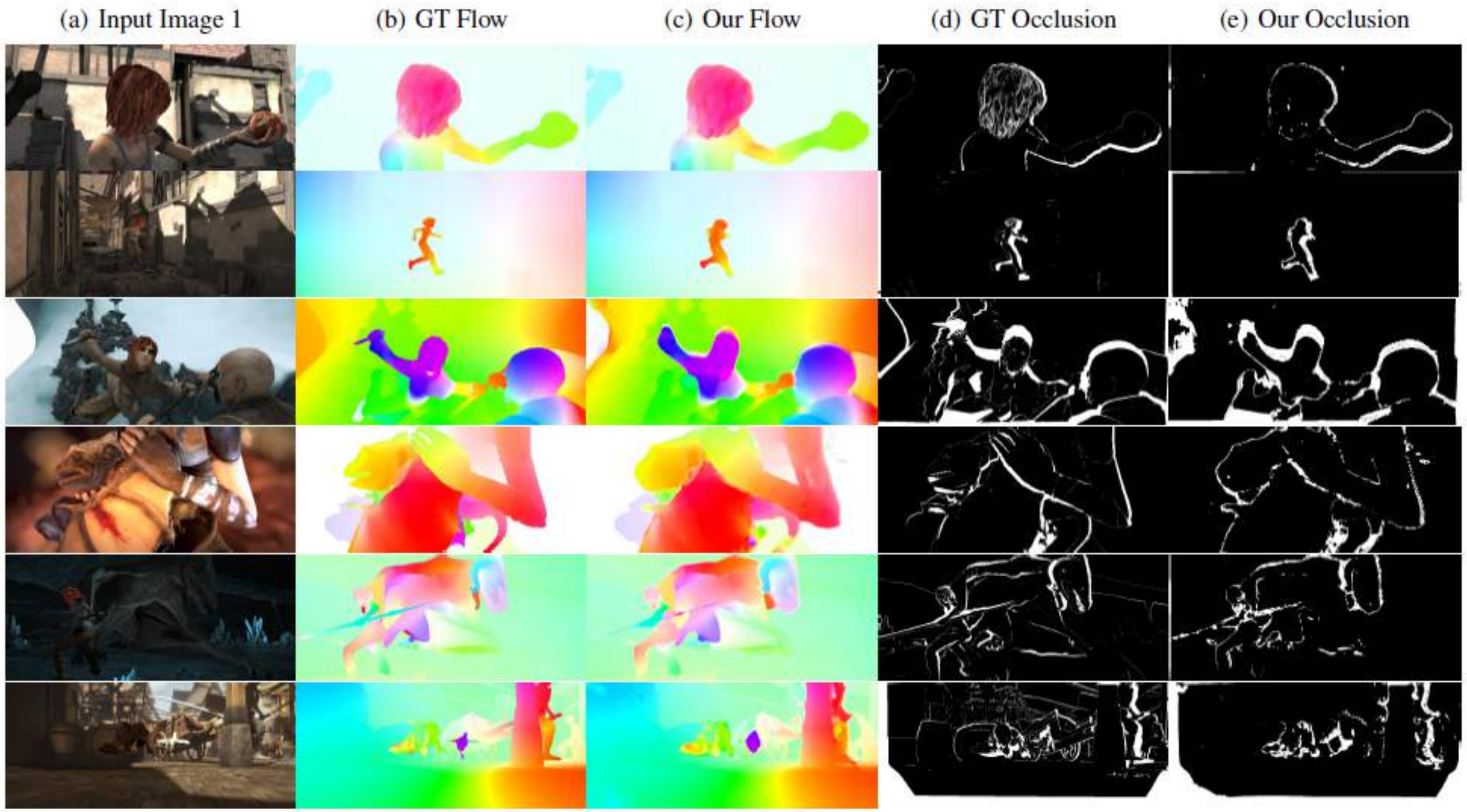
Quantitative Comparisons

- DDFlow achieves the best occlusion estimation performance on Sintel Clean and Sintel Final datasets.
- On KITTI dataset, the ground truth occlusion masks only contain pixels moving out of the image boundary. Under such setting, our method can achieve comparable performance.

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	$\boxed{\textbf{(0.59)}}$	$\boxed{\textbf{(0.52)}}$	0.94*	0.86*

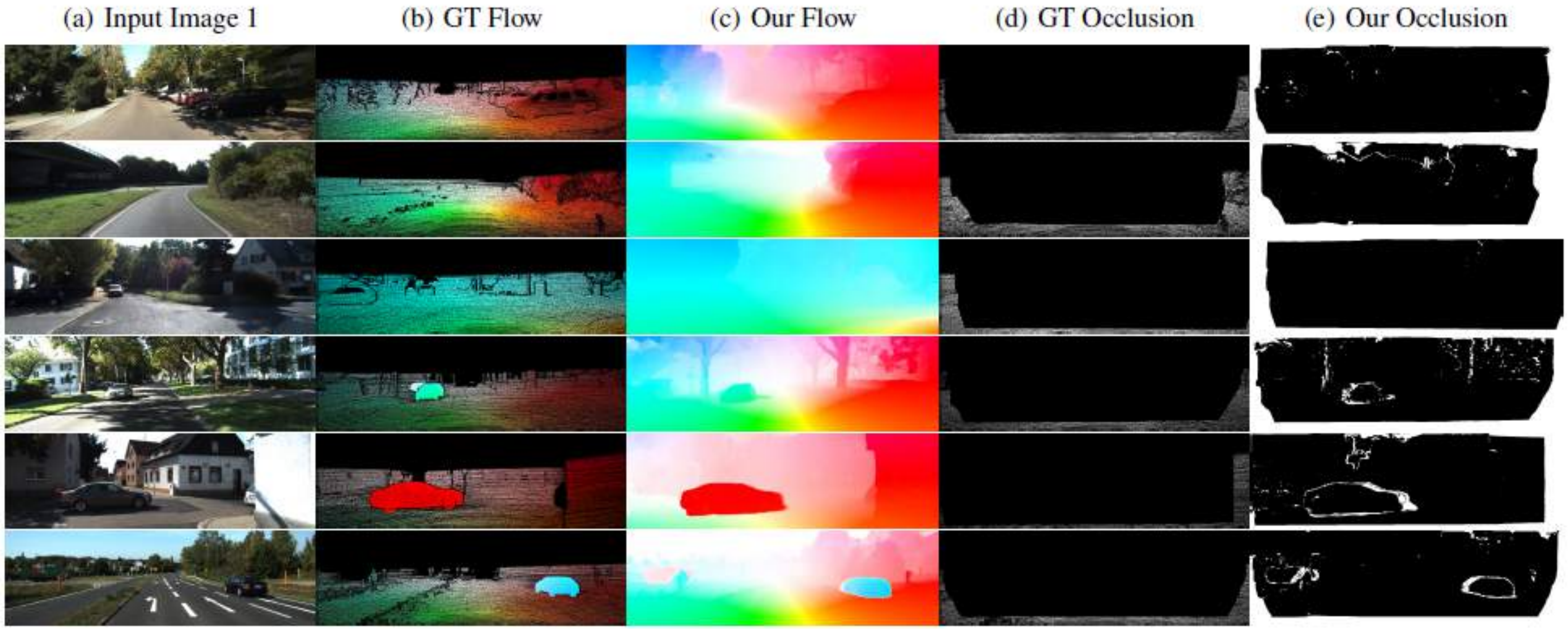
Qualitative Comparisons

- Sample results on Sintel datasets. The first three rows are from Sintel Clean, while the last three are from Sintel Final.



Qualitative Comparisons

- Example results on KITTI datasets. The first three rows are from KITTI 2012, and the last three are from KITTI 2015.
- Note that on KITTI datasets, the occlusion masks are sparse and only contain pixels moving out of the image boundary.



Quantitative: Ablation Study

- Comparing row 1, 2 and row 3, 4: occlusion handling can improve flow estimation performance on all datasets.

Occlusion Handling	Census Transform	Data Distillation	Chairs ALL	Sintel Clean ALL	Sintel Clean NOC	Sintel Clean OCC	Sintel Final ALL	Sintel Final NOC	Sintel Final OCC	KITTI 2012 ALL	KITTI 2012 NOC	KITTI 2012 OCC	KITTI 2015 ALL	KITTI 2015 NOC	KITTI 2015 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**

Quantitative: Ablation Study

- ❑ Comparing row 1, 2 and row 3, 4: occlusion handling can improve flow estimation performance on all datasets.
- ❑ Comparing row 1, 3 and row 2, 4: census transform constantly improve performance.

Occlusion Handling	Census Transform	Data Distillation	Chairs ALL	Sintel Clean ALL	Sintel Clean NOC	Sintel Clean OCC	Sintel Final ALL	Sintel Final NOC	Sintel Final OCC	KITTI 2012 ALL	KITTI 2012 NOC	KITTI 2012 OCC	KITTI 2015 ALL	KITTI 2015 NOC	KITTI 2015 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**

Quantitative: Ablation Study

- ❑ Comparing row 1, 2 and row 3, 4: occlusion handling can improve flow estimation performance on all datasets.
- ❑ Comparing row 1, 3 and row 2, 4: census transform constantly improve performance.
- ❑ Comparing row 4, 5: data distillation can greatly improve the performance, especially for **occluded** pixels, with EPE-OCC decreases 18.5% on Sintel Clean, 16.1% on Sintel Final, 58.2% on KITTI 2012 and 42.1% on KITTI 2015.

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**

Video Flow Estimation on Sintel Dataset

- The top part is the input frame and the bottom part is the corresponding optical flow estimated by DDFlow.

DDFlow code

- Code and models available on https://github.com/ppliuboy/DDFlow.

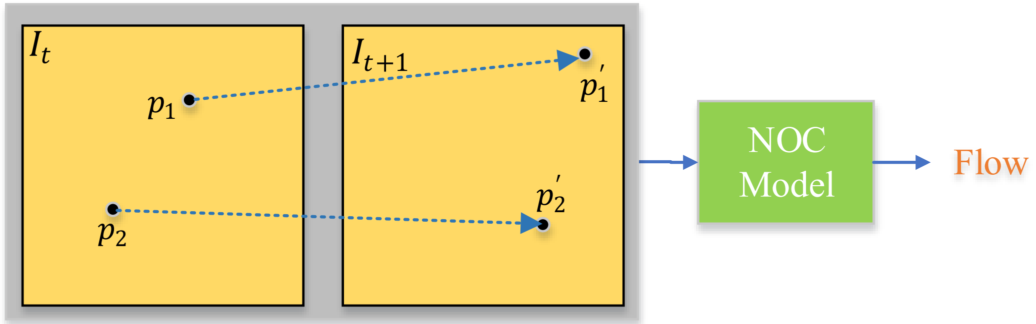
What is Next?

Motivation

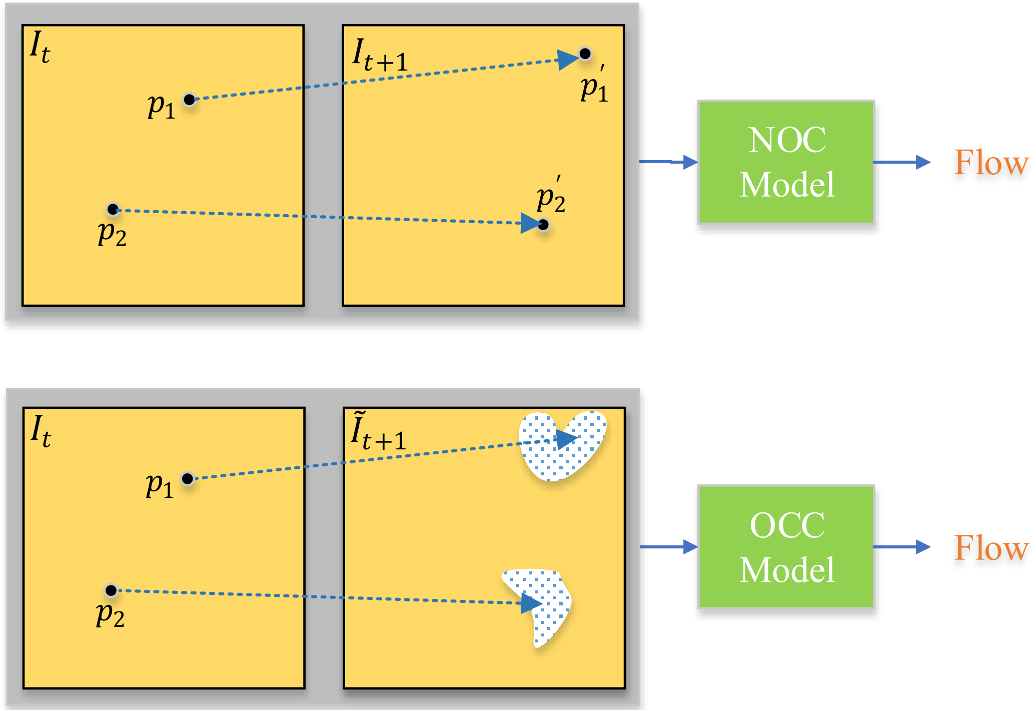
- Can we completely get rid of synthetic data?
- Can we win Sintel back?

Liu, King, Lyu, **Xu.** SelFlow: Self-Supervised Learning of Optical Flow. CVPR 2019

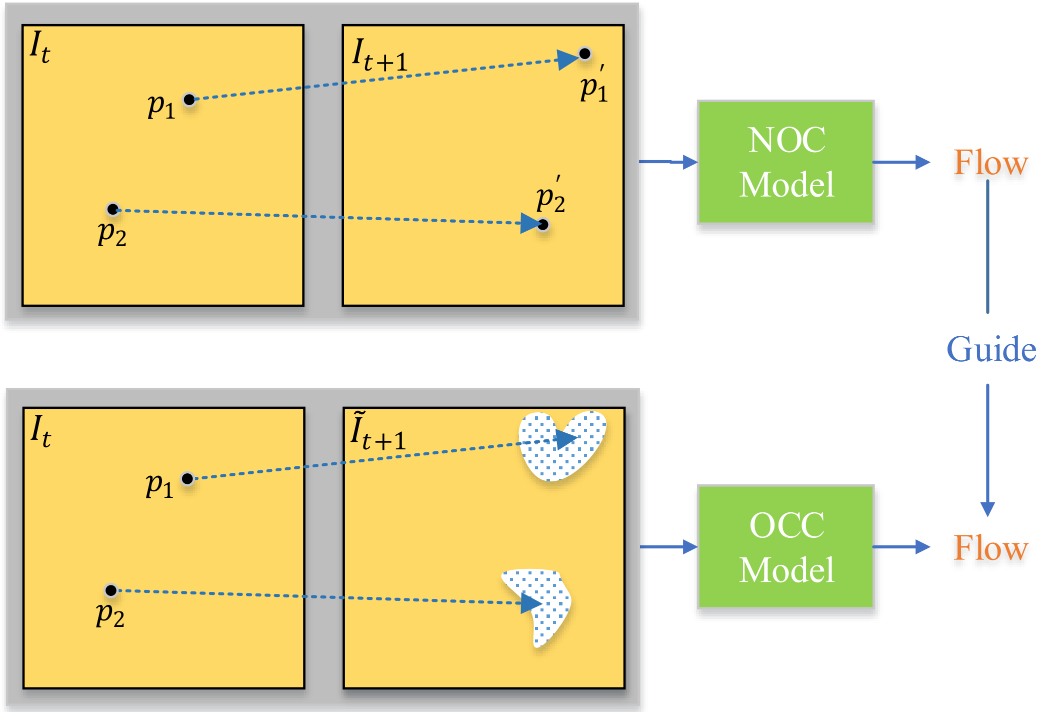
Initially, p_1 and p_2 are non-occluded from I_t to I_{t+1}, p_1' and p_2' are their corresponding pixels. NOC-Model can accurately estimate the flow of p_1 and p_2 using photometric loss.



We inject random noise to I_{t+1} and let noise cover p_1 and p_2, then p_1 and p_2 become occluded from I_t to $\tilde{I}_{t+1}$. OCC-Model cannot accurately estimate flow of p_1 and p_2 using photometric loss.



We distill reliable flow estimations of p_1 and p_2 from NOC-Model to guide the flow learning for OCC-Model. The guidance is only employed to pixels that are occluded from I_t to $\tilde{I}_{t+1}$ but non-occluded from I_t to I_{t+1}, such as p_1 and p_2.



Quantitative Results

Our unsupervised results outperform all existing unsupervised results on all datasets by a large margin.

	Method	Sintel Clean		Sintel Final		KITTI 2012			KITTI 2015	
		train	test	train	test	train	test	test(Fl)	train	test(Fl)
Unsupervised	BackToBasic+ft [19]	–	–	–	–	11.3	9.9	–	–	–
	DSTFlow+ft [30]	(6.16)	10.41	(6.81)	11.27	10.43	12.4	–	16.79	39%
	UnFlow-CSS [25]	–	–	(7.91)	10.22	3.29	–	–	8.10	23.30%
	OccAwareFlow+ft [39]	(4.03)	7.95	(5.95)	9.15	3.55	4.2	–	8.88	31.2%
	MultiFrameOccFlow-Hard+ft [17]	(6.05)	–	(7.09)	–	–	–	–	6.65	–
	MultiFrameOccFlow-Soft+ft [17]	(3.89)	7.23	(5.52)	8.81	–	–	–	6.59	22.94%
	Ours	**(2.82)**	**6.56**	**(3.87)**	**6.57**	**1.69**	**2.2**	**7.68%**	**4.84**	**14.19%**
Supervised	FlowNetS+ft [9]	(3.66)	6.96	(4.44)	7.76	7.52	9.1	44.49%	–	–
	FlowNetC+ft [9]	(3.78)	6.85	(5.28)	8.51	8.79	–	–	–	–
	SpyNet+ft [28]	(3.17)	6.64	(4.32)	8.36	8.25	10.1	20.97%	–	35.07%
	FlowFieldsCNN+ft [2]	–	3.78	–	5.36	–	3.0	13.01%	–	18.68 %
	DCFlow+ft [42]	–	3.54	–	5.12	–	–	–	–	14.83%
	FlowNet2+ft [14]	(1.45)	4.16	(2.01)	5.74	(1.28)	1.8	–	(2.3)	11.48%
	UnFlow-CSS+ft [25]	–	–	–	–	(1.14)	1.7	8.42%	(1.86)	11.11%
	LiteFlowNet+ft-CVPR [13]	(1.64)	4.86	(2.23)	6.09	(1.26)	1.7	–	(2.16)	10.24%
	LiteFlowNet+ft-axXiv [13]	**(1.35)**	4.54	(1.78)	5.38	(1.05)	1.6	7.27%	(1.62)	9.38%
	PWC-Net+ft-CVPR [36]	(2.02)	4.39	(2.08)	5.04	(1.45)	1.7	8.10%	(2.16)	9.60%
	PWC-Net+ft-axXiv [35]	(1.71)	3.45	(2.34)	4.60	(1.08)	**1.5**	6.82%	(1.45)	7.90%
	ProFlow+ft [23]	(1.78)	**2.82**	–	5.02	(1.89)	2.1	7.88%	(5.22)	15.04%
	ContinualFlow+ft [27]	–	3.34	–	4.52	–	–	–	–	10.03%
	MFF+ft [29]	–	3.42	–	4.57	–	1.7	7.87%	–	**7.17%**
	Ours+ft	(1.68)	3.74	**(1.77)**	**4.26**	**(0.76)**	**1.5**	**6.19%**	**(1.18)**	8.42%

Our unsupervised results even outperform several famous fully-supervised methods.

	Method	Sintel Clean		Sintel Final		KITTI 2012			KITTI 2015	
		train	test	train	test	train	test	test(Fl)	train	test(Fl)
Unsupervised	BackToBasic+ft [19]	–	–	–	–	11.3	9.9	–	–	–
	DSTFlow+ft [30]	(6.16)	10.41	(6.81)	11.27	10.43	12.4	–	16.79	39%
	UnFlow-CSS [25]	–	–	(7.91)	10.22	3.29	–	–	8.10	23.30%
	OccAwareFlow+ft [39]	(4.03)	7.95	(5.95)	9.15	3.55	4.2	–	8.88	31.2%
	MultiFrameOccFlow-Hard+ft [17]	(6.05)	–	(7.09)	–	–	–	–	6.65	–
	MultiFrameOccFlow-Soft+ft [17]	(3.89)	7.23	(5.52)	8.81	–	–	–	6.59	22.94%
	Ours	**(2.82)**	**6.56**	**(3.87)**	**6.57**	**1.69**	**2.2**	**7.68%**	**4.84**	**14.19%**
Supervised	FlowNetS+ft [9]	(3.66)	6.96	(4.44)	7.76	7.52	9.1	44.49%	–	–
	FlowNetC+ft [9]	(3.78)	6.85	(5.28)	8.51	8.79	–	–	–	–
	SpyNet+ft [28]	(3.17)	6.64	(4.32)	8.36	8.25	10.1	20.97%	–	35.07%
	FlowFieldsCNN+ft [2]	–	3.78	–	5.36	–	3.0	13.01%	–	18.68 %
	DCFlow+ft [42]	–	3.54	–	5.12	–	–	–	–	14.83%
	FlowNet2+ft [14]	(1.45)	4.16	(2.01)	5.74	(1.28)	1.8	–	(2.3)	11.48%
	UnFlow-CSS+ft [25]	–	–	–	–	(1.14)	1.7	8.42%	(1.86)	11.11%
	LiteFlowNet+ft-CVPR [13]	(1.64)	4.86	(2.23)	6.09	(1.26)	1.7	–	(2.16)	10.24%
	LiteFlowNet+ft-axXiv [13]	**(1.35)**	4.54	(1.78)	5.38	(1.05)	1.6	7.27%	(1.62)	9.38%
	PWC-Net+ft-CVPR [36]	(2.02)	4.39	(2.08)	5.04	(1.45)	1.7	8.10%	(2.16)	9.60%
	PWC-Net+ft-axXiv [35]	(1.71)	3.45	(2.34)	4.60	(1.08)	**1.5**	6.82%	(1.45)	7.90%
	ProFlow+ft [23]	(1.78)	**2.82**	–	5.02	(1.89)	2.1	7.88%	(5.22)	15.04%
	ContinualFlow+ft [27]	–	3.34	–	4.52	–	–	–	–	10.03%
	MFF+ft [29]	–	3.42	–	4.57	–	1.7	7.87%	–	**7.17%**
	Ours+ft	(1.68)	3.74	**(1.77)**	**4.26**	**(0.76)**	**1.5**	**6.19%**	**(1.18)**	8.42%

Our fine-tuned models achieve state-of-the-art results without using any external labeled data.

	Method	Sintel Clean		Sintel Final		KITTI 2012			KITTI 2015	
		train	test	train	test	train	test	test(Fl)	train	test(Fl)
Unsupervised	BackToBasic+ft [19]	–	–	–	–	11.3	9.9	–	–	–
	DSTFlow+ft [30]	(6.16)	10.41	(6.81)	11.27	10.43	12.4	–	16.79	39%
	UnFlow-CSS [25]	–	–	(7.91)	10.22	3.29	–	–	8.10	23.30%
	OccAwareFlow+ft [39]	(4.03)	7.95	(5.95)	9.15	3.55	4.2	–	8.88	31.2%
	MultiFrameOccFlow-Hard+ft [17]	(6.05)	–	(7.09)	–	–	–	–	6.65	–
	MultiFrameOccFlow-Soft+ft [17]	(3.89)	7.23	(5.52)	8.81	–	–	–	6.59	22.94%
	Ours	**(2.82)**	**6.56**	**(3.87)**	**6.57**	**1.69**	**2.2**	**7.68%**	**4.84**	**14.19%**
Supervised	FlowNetS+ft [9]	(3.66)	6.96	(4.44)	7.76	7.52	9.1	44.49%	–	–
	FlowNetC+ft [9]	(3.78)	6.85	(5.28)	8.51	8.79	–	–	–	–
	SpyNet+ft [28]	(3.17)	6.64	(4.32)	8.36	8.25	10.1	20.97%	–	35.07%
	FlowFieldsCNN+ft [2]	–	3.78	–	5.36	–	3.0	13.01%	–	18.68 %
	DCFlow+ft [42]	–	3.54	–	5.12	–	–	–	–	14.83%
	FlowNet2+ft [14]	(1.45)	4.16	(2.01)	5.74	(1.28)	1.8	–	(2.3)	11.48%
	UnFlow-CSS+ft [25]	–	–	–	–	(1.14)	1.7	8.42%	(1.86)	11.11%
	LiteFlowNet+ft-CVPR [13]	(1.64)	4.86	(2.23)	6.09	(1.26)	1.7	–	(2.16)	10.24%
	LiteFlowNet+ft-axXiv [13]	**(1.35)**	4.54	(1.78)	5.38	(1.05)	1.6	7.27%	(1.62)	9.38%
	PWC-Net+ft-CVPR [36]	(2.02)	4.39	(2.08)	5.04	(1.45)	1.7	8.10%	(2.16)	9.60%
	PWC-Net+ft-axXiv [35]	(1.71)	3.45	(2.34)	4.60	(1.08)	**1.5**	6.82%	(1.45)	7.90%
	ProFlow+ft [23]	(1.78)	**2.82**	–	5.02	(1.89)	2.1	7.88%	(5.22)	15.04%
	ContinualFlow+ft [27]	–	3.34	–	4.52	–	–	–	–	10.03%
	MFF+ft [29]	–	3.42	–	4.57	–	1.7	7.87%	–	**7.17%**
	Ours+ft	(1.68)	3.74	**(1.77)**	**4.26**	**(0.76)**	**1.5**	**6.19%**	**(1.18)**	8.42%

MPI Sintel Dataset　About　Downloads　Results　FAQ　Contact　Signup　Login

Results and Rankings

Results for methods appear here after users upload them and approve them for public display.

Final　Clean

	EPE all	EPE matched	EPE unmatched	d0-10	d10-60	d60-140	s0-10	s10-40	s40+	
GroundTruth [1]	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Visualize Results
SelFlow [2]	4.262	2.040	22.369	4.083	1.715	1.287	0.582	2.343	27.154	Visualize Results
ContinualFlow_ROB [3]	4.528	2.723	19.248	5.050	2.573	1.713	0.872	3.114	26.063	Visualize Results
MFF [4]	4.566	2.216	23.732	4.664	2.017	1.222	0.893	2.902	26.810	Visualize Results
PWC-Net+ [5]	4.596	2.254	23.696	4.781	2.045	1.234	0.945	2.978	26.620	Visualize Results
CompactFlow_mix [6]	4.691	2.307	24.142	4.327	1.933	1.458	0.863	2.552	28.873	Visualize Results
CompactFlow_ROB [7]	4.720	2.334	24.185	4.305	1.959	1.529	0.815	2.668	29.065	Visualize Results

Qualitative Results

Effect of Self-supervision

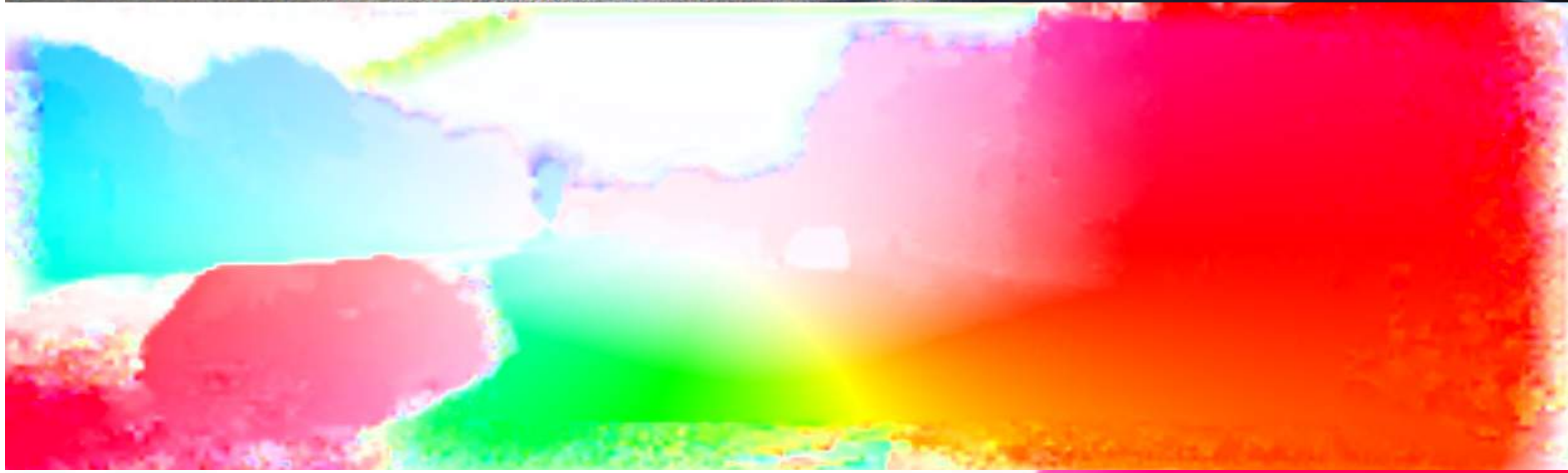
Reference Image

Flow Estimation
without Self-supervision

Flow Estimation
with Self-supervision

Reference Image

Flow Estimation
without Self-supervision



Flow Estimation
with Self-supervision

Reference Image

Flow Estimation without Self-supervision

Flow Estimation with Self-supervision

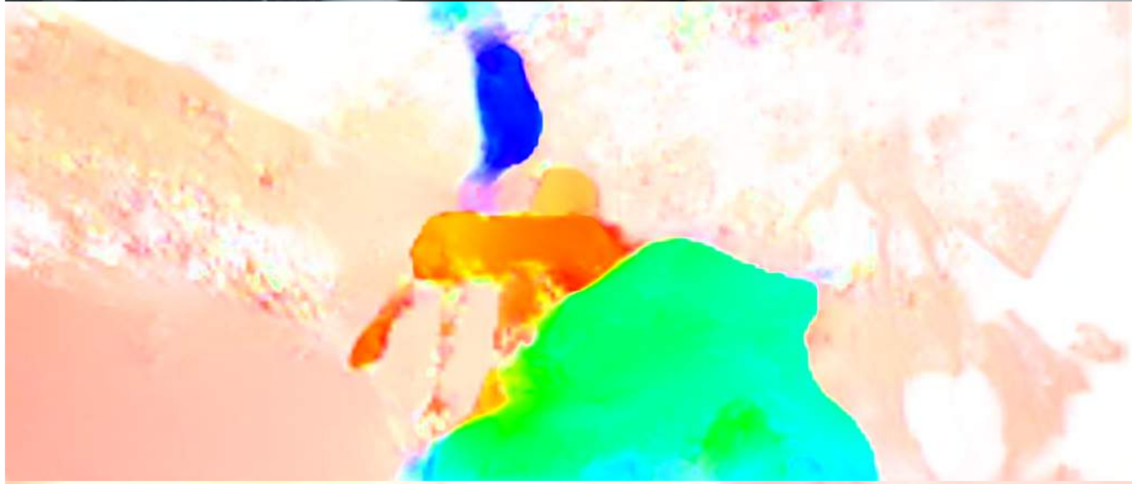
Reference Image

Flow Estimation
without Self-supervision

Flow Estimation
with Self-supervision

Reference Image

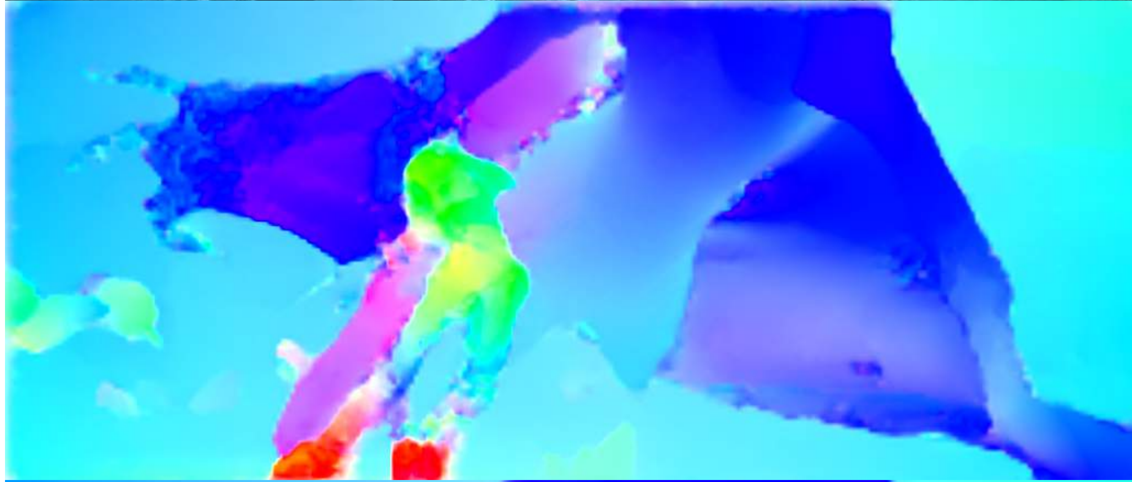
Flow Estimation
without Self-supervision



Flow Estimation
with Self-supervision

Reference Image

Flow Estimation
without Self-supervision



Flow Estimation
with Self-supervision

Reference Image

Flow Estimation
without Self-supervision

Flow Estimation
with Self-supervision

Reference Image

Flow Estimation
without Self-supervision

Flow Estimation
with Self-supervision

Compared with PWC-Net, our fine-tuned model estimates optical flow with more accurate details.

Reference Image

Flow Estimation using PWC-Net

Flow Estimation using Our Fine-tuned Model

Reference Image

Flow Estimation using PWC-Net

Flow Estimation using Our Fine-tuned Model

Reference Image

Flow Estimation
using PWC-Net

Flow Estimation
using Our Fine-tuned
Model

Reference Image

Flow Estimation
using PWC-Net

Flow Estimation
using Our Fine-tuned
Model

To demonstrate the generalization ability of our model, we further show our flow estimation on real-word videos (from the DAVIS dataset).

Reference Image

Flow from Our Unsupervised Model

Flow from Our Fine-tuned Model

Reference Image

Flow from Our Unsupervised Model

Flow from Our Fine-tuned Model

Q & A

Hiring in Vision and Graphics ;)
http://pages.cs.wisc.edu/~jiaxu/