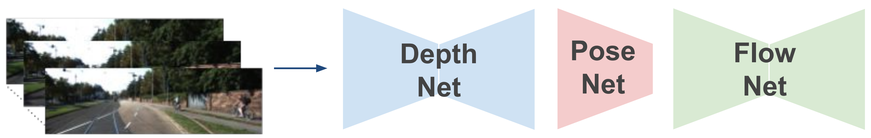
DF-Net: Unsupervised Joint Learning of Depth and Flow using Cross-Task Consistency

Yuliang Zou, Virginia Tech
Zelun Luo, Stanford University
Jia-Bin Huang, Virginia Tech

Code available at: http://yuliang.vision/DF-Net

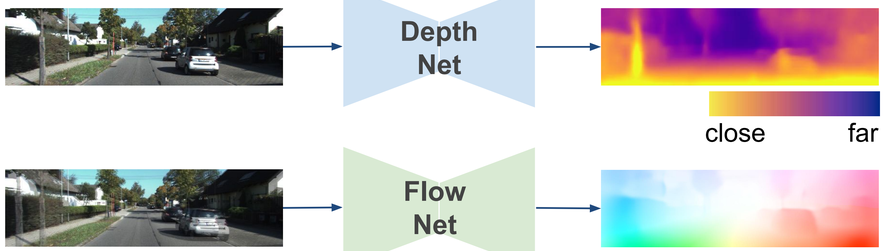
Introduction

Train

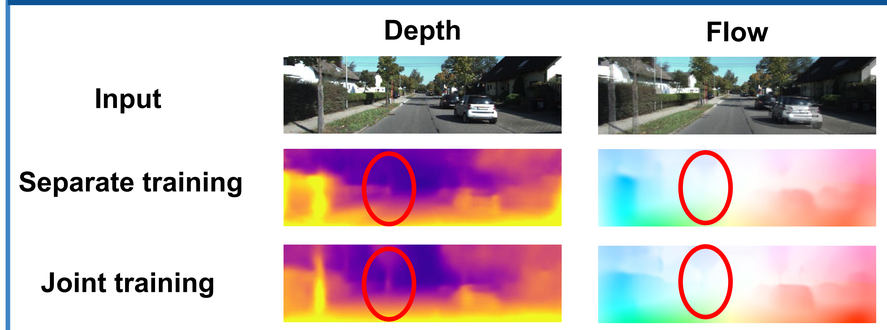


Unlabeled video clips

Test

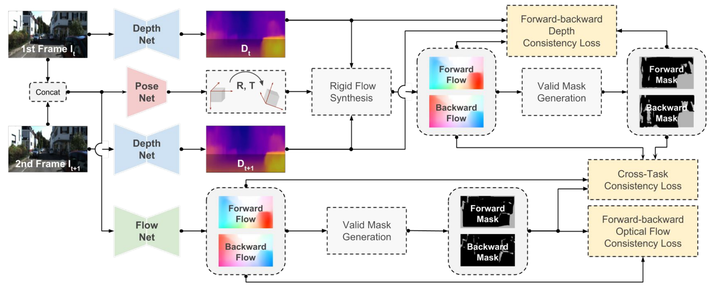


Motivation



Unsupervised Joint Learning

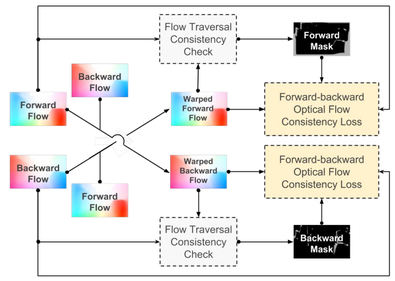
Overview



Overall Objective

$$L = L_{\text{photometric}} + \lambda_s L_{\text{smooth}} + \lambda_f L_{\text{forward-backward}} + \lambda_c L_{\text{cross}}$$

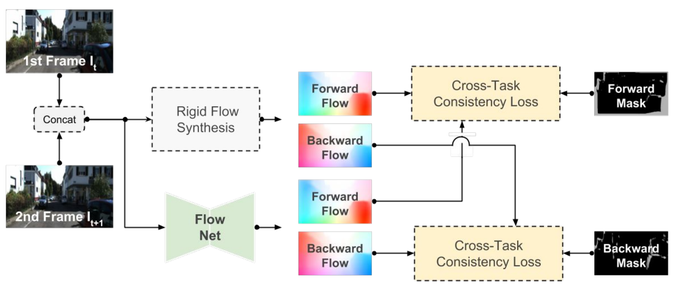
Forward-backward consistency



- A pixel moves along forward-backward flow should stay at the same position
- Large disagreement indicates invalid regions (e.g., occlusions)
- Enforce the forward-backward consistency for valid regions

$$L_{\text{forward-backward, flow}} = \sum_{p \in V_{\text{flow}}} ||F_{t \to t+1}(p) + F_{t+1 \to t}(p + F_{t \to t+1}(p))||_1$$

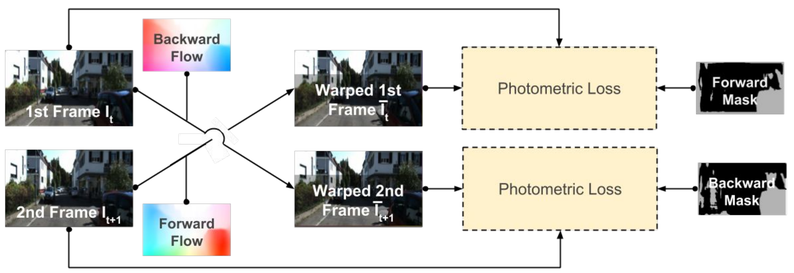
Cross-task consistency



- The **synthesized rigid flow** should be consistent with the **estimated optical flow** in valid regions

$$L_{\text{cross}} = \sum_{p \in V_{\text{depth}} \cap V_{\text{flow}}} ||F_{\text{rigid}}(p) - F_{\text{flow}}(p)||_1$$

Brightness constancy

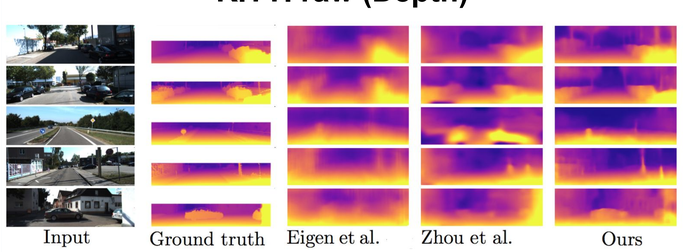


- A warped frame should be similar to the target frame

$$L_{\text{photometric}} = \sum_{p \in V} \rho\big(I_t(p), \bar{I}_t(p)\big)$$

Experimental Results

KITTI raw (Depth)



Method	Dataset	Error metric ↓ Abs Rel	Sq Rel	RMSE	log RMSE	Accuracy metric ↑ $\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Zhou et al. 2017	K (M)	0.208	1.768	6.856	0.283	0.678	0.885	0.957
Yang et al. 2017	K (M)	0.182	1.481	6.501	0.267	0.725	0.906	0.963
Mahjourian et al. 2018	K (M)	0.163	1.240	6.220	0.250	0.762	0.916	0.968
Yang et al. 2018	K (M)	0.162	1.352	6.276	0.252	-	-	-
Yin et al. 2018	K (M)	0.155	1.296	5.857	0.233	0.793	0.931	**0.973**
Godard et al. 2018	K (M)	0.154	1.218	5.699	0.231	0.798	0.932	**0.973**
Ours	K (M)	**0.150**	**1.124**	**5.507**	**0.223**	**0.806**	**0.933**	**0.973**
Zhou et al. 2017	CS+K (M)	0.198	1.836	6.565	0.275	0.718	0.901	0.960
Yang et al. 2017	CS+K (M)	0.165	1.360	6.641	0.248	0.750	0.914	0.969
Mahjourian et al. 2018	CS+K (M)	0.159	1.231	5.912	0.243	0.784	0.923	0.970
Yang et al. 2018	CS+K (M)	0.159	1.345	6.254	0.247	-	-	-
Yin et al. 2018	CS+K (M)	0.153	1.328	5.737	0.232	0.802	0.934	0.972
Ours	CS+K (M)	**0.146**	**1.182**	**5.215**	**0.213**	**0.818**	**0.943**	**0.978**

KITTI 2012 / 2015 (Flow)

Method	Dataset	KITTI 2012 Train EPE	KITTI 2012 Test EPE	KITTI 2015 Train EPE	KITTI 2015 Train F1	KITTI 2015 Test F1
FlowNetC	C (S)	9.35	-	12.52	47.93%	-
SemiFlowGAN	C (S) / K (U)	7.16	-	16.02	38.77%	-
FlowNet2	C (S) + T (S)	**4.09**	-	**10.06**	**30.37%**	-
UnsupFlownet	C (U) + K (U)	11.3	9.9	-	-	-
DSTFlow	C (U)	16.98	-	24.30	52.00%	-
DSTFlow	K (U)	10.43	12.4	16.79	36.00%	39.00%
UnFlowC	SYN (U) + K (U)	3.78	4.5	**8.80**	28.94%	29.46%
Ours	SYN (U) + K (U)	**3.54**	**4.4**	8.98	**26.01%**	**25.70%**
UnFlowC-ft-kitti	SYN (U) + K (U) + K (S)	(2.13)	3.0	(3.67)	(17.78%)	24.20%
Ours-ft-kitti	SYN (U) + K (U) + K (S)	(1.75)	3.0	(2.85)	(13.47%)	**22.82%**