Joint Learning of Geometry and Motion with 3D Holistic Understanding

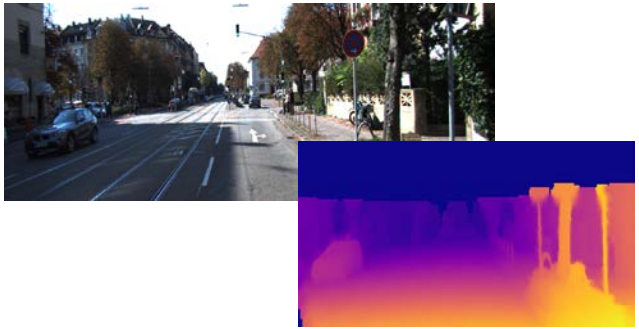
Chenxu Luo*
Zhenheng Yang*, Peng Wang*, Yang Wang,
Wei Xu, Ram Nevatia, and Alan Yuille
https://arxiv.org/abs/1810.06125

Tasks

Training

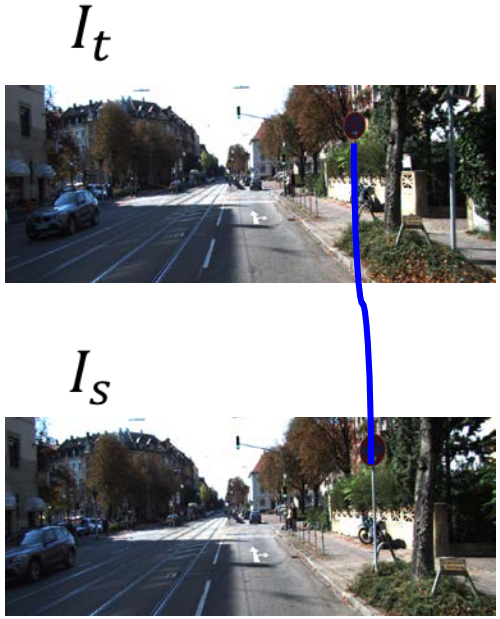
Testing

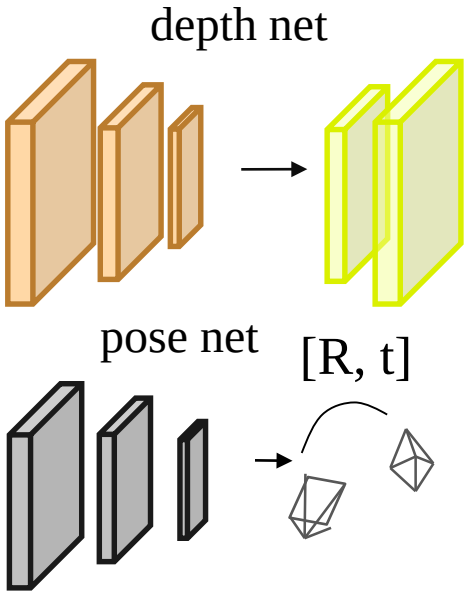
Depth

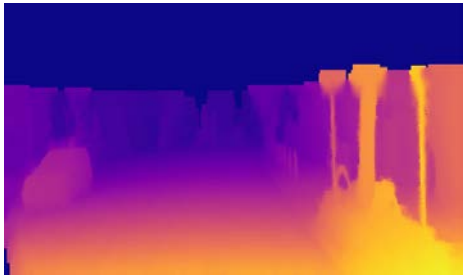


Flow & Motion

SfMLearner (Zhou et at CVPR 2017)







$p_s = KT_{t\to s}D(p_t)K^{-1}p_t$

Recon Loss: $| I(p_s) - I(p_t) |$

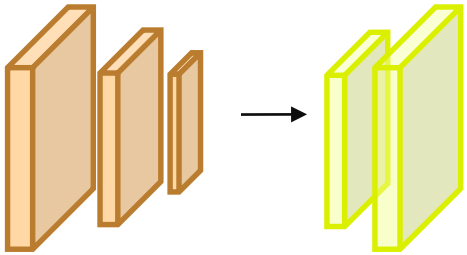
Smooth Loss: $| \nabla d_{i,j} |$

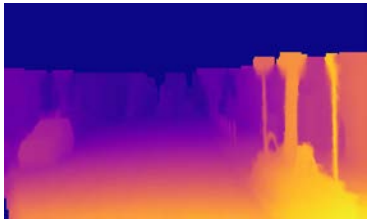
Depth Flow (Rigid Flow)

I_t

I_s

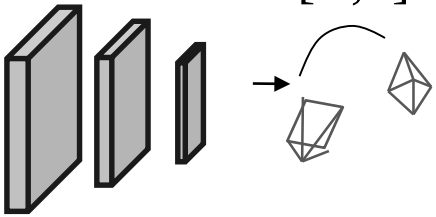
depth net

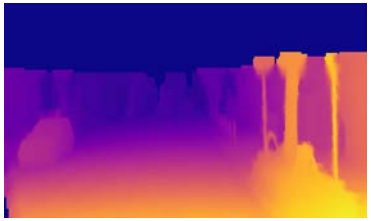




pose net

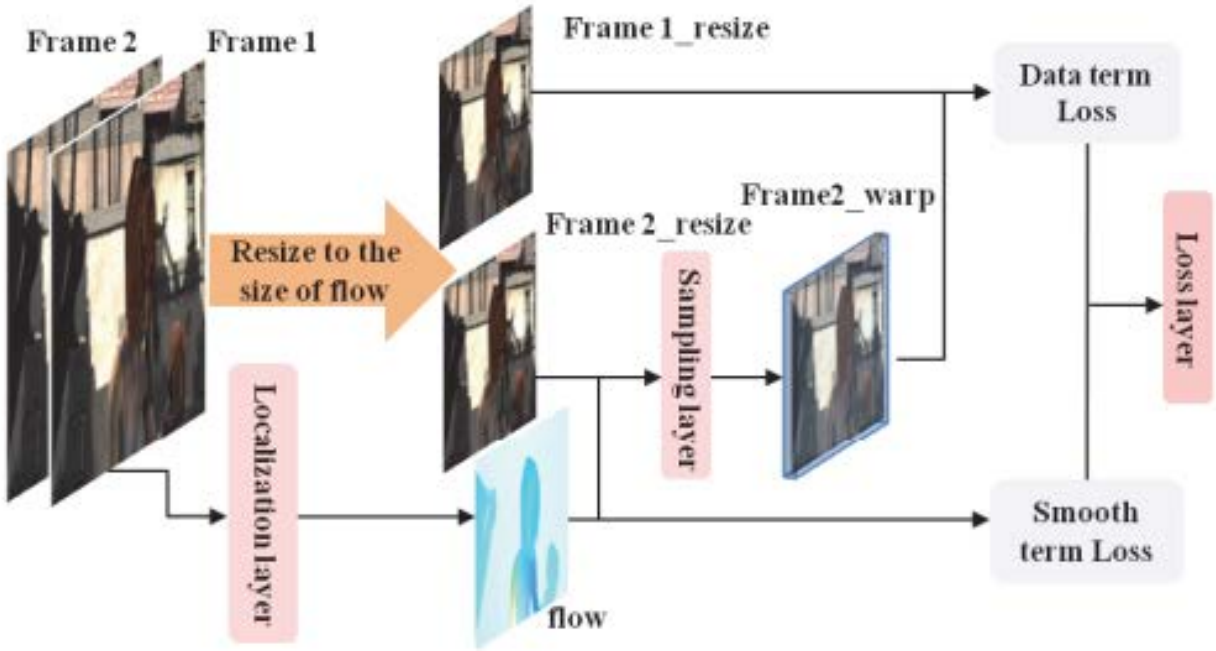
[R, t]



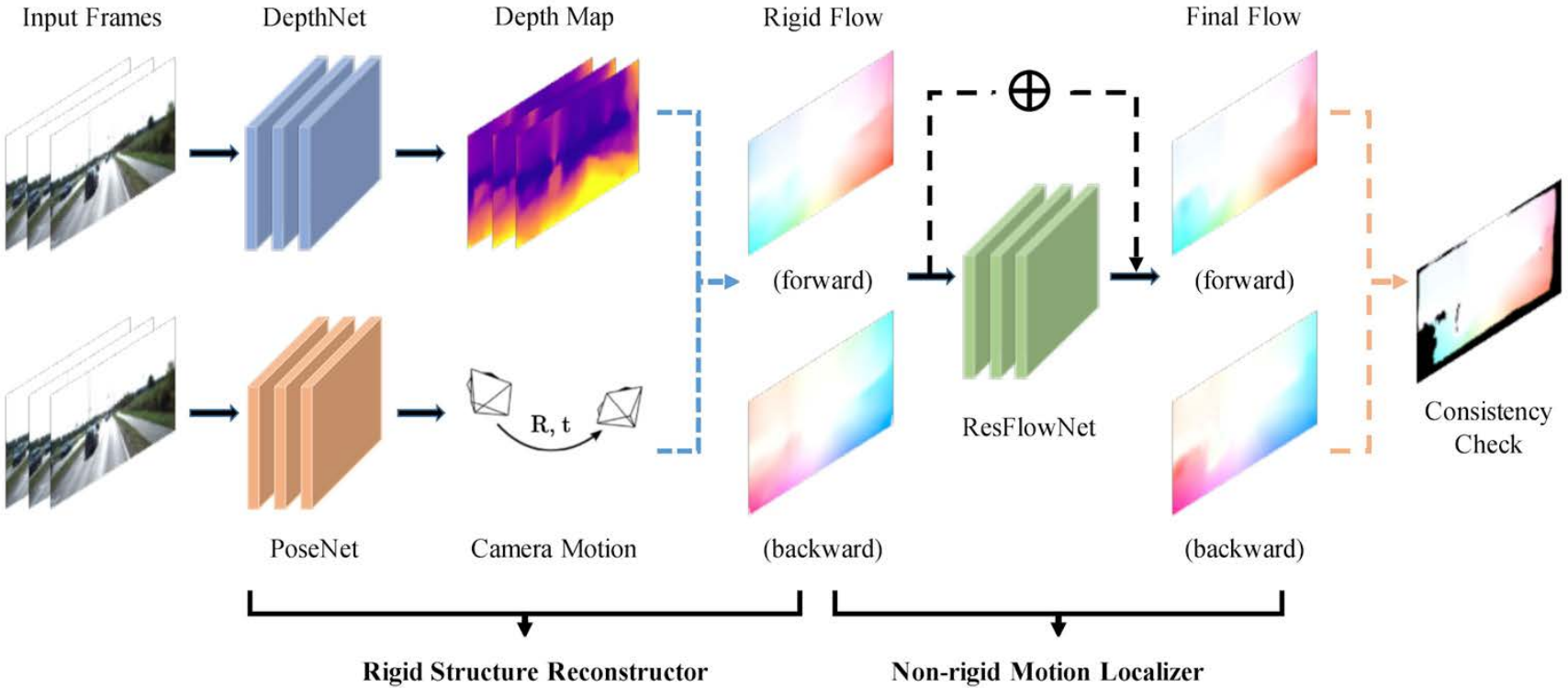


Depth flow

DSTFlow (Zhe Ren AAAI 2017)



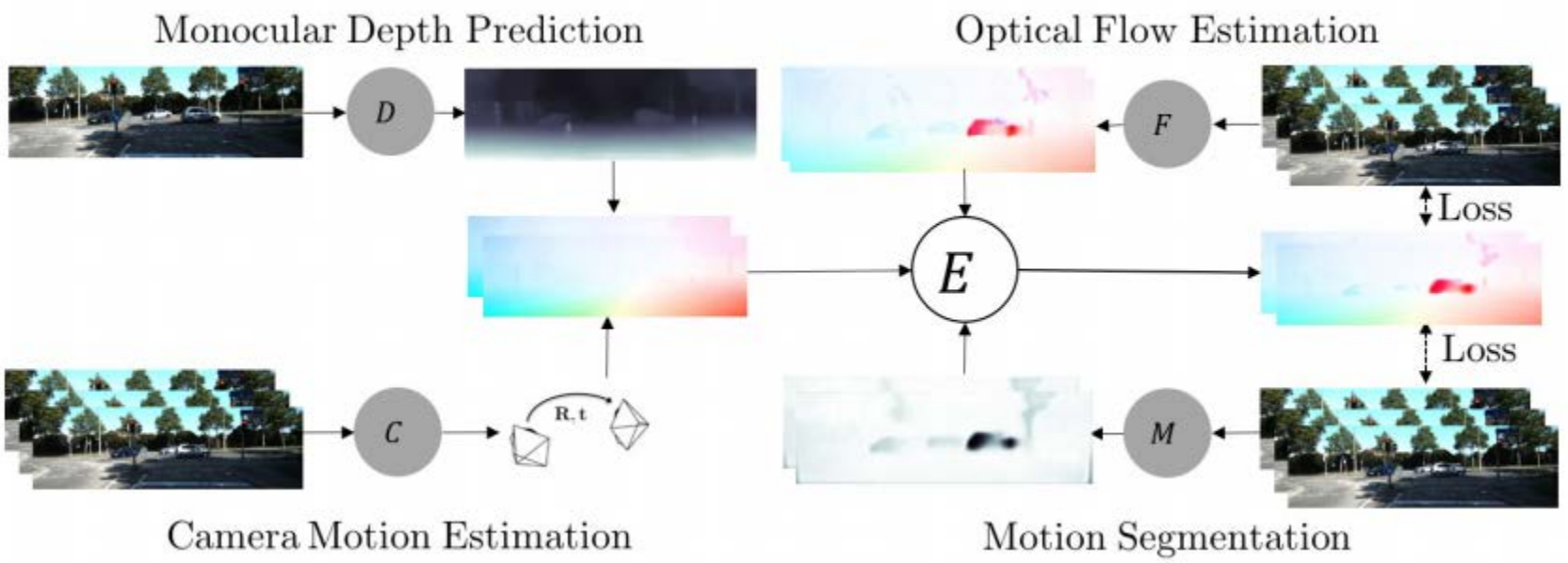
Previous Works --GeoNet (CVPR 2018)



GeoNet -- KITTI 2015 TRAIN

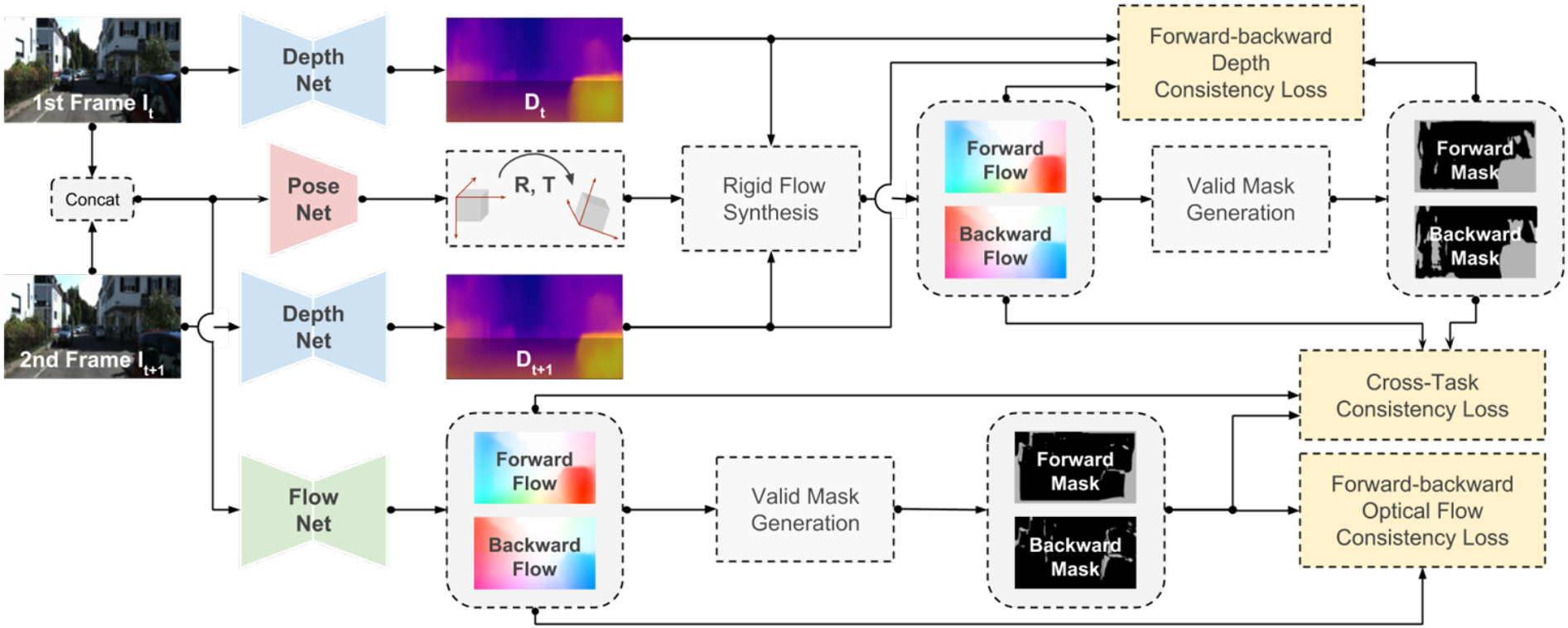
Method	Dataset	Noc	All
EpicFlow [38]	-	4.45	9.57
FlowNetS [8]	C+S	8.12	14.19
FlowNet2 [18]	C+T	4.93	10.06
DSTFlow [37]	K	6.96	16.79
Our DirFlowNetS (no GC)	K	6.80	12.86
Our DirFlowNetS	K	**6.77**	12.21
Our Naive GeoNet	K	8.57	17.18
Our GeoNet	K	8.05	**10.81**

Previous Works -- Competitive Collaboration (arxiv)



Method	Average EPE		
	SP	MP	Total
FlowNet2[11]	-	-	10.06
SPyNet[24]	-	-	20.26
UnFlow-C[22]	-	-	8.80
UnFlow-CSS[22]	-	-	8.10
Geonet [32]	-	-	10.81
Ours (basic, R)	7.51	32.75	13.54
Ours (basic, F)	15.32	6.20	14.68
Ours (CC)	**6.35**	**6.16**	**7.76**

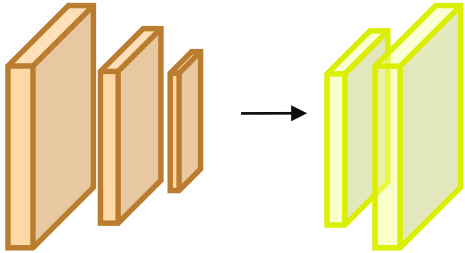
Previous Works --- DF-Net (ECCV 2018)

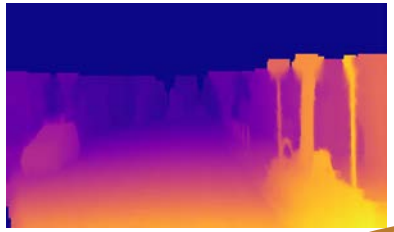


Our Approach

I_t

depth net



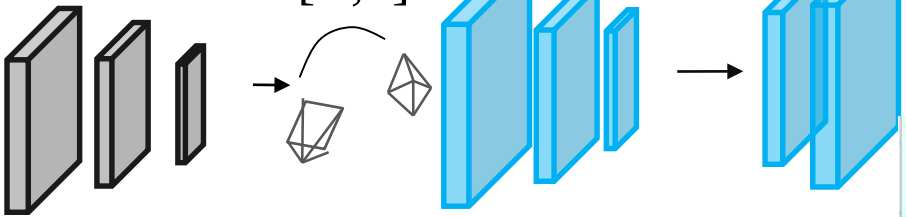


Depth flow

Non-occluded & rigid

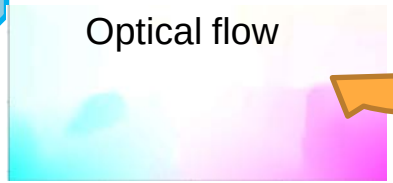
pose net

[R, t]



flow net

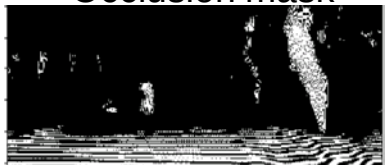
Optical flow



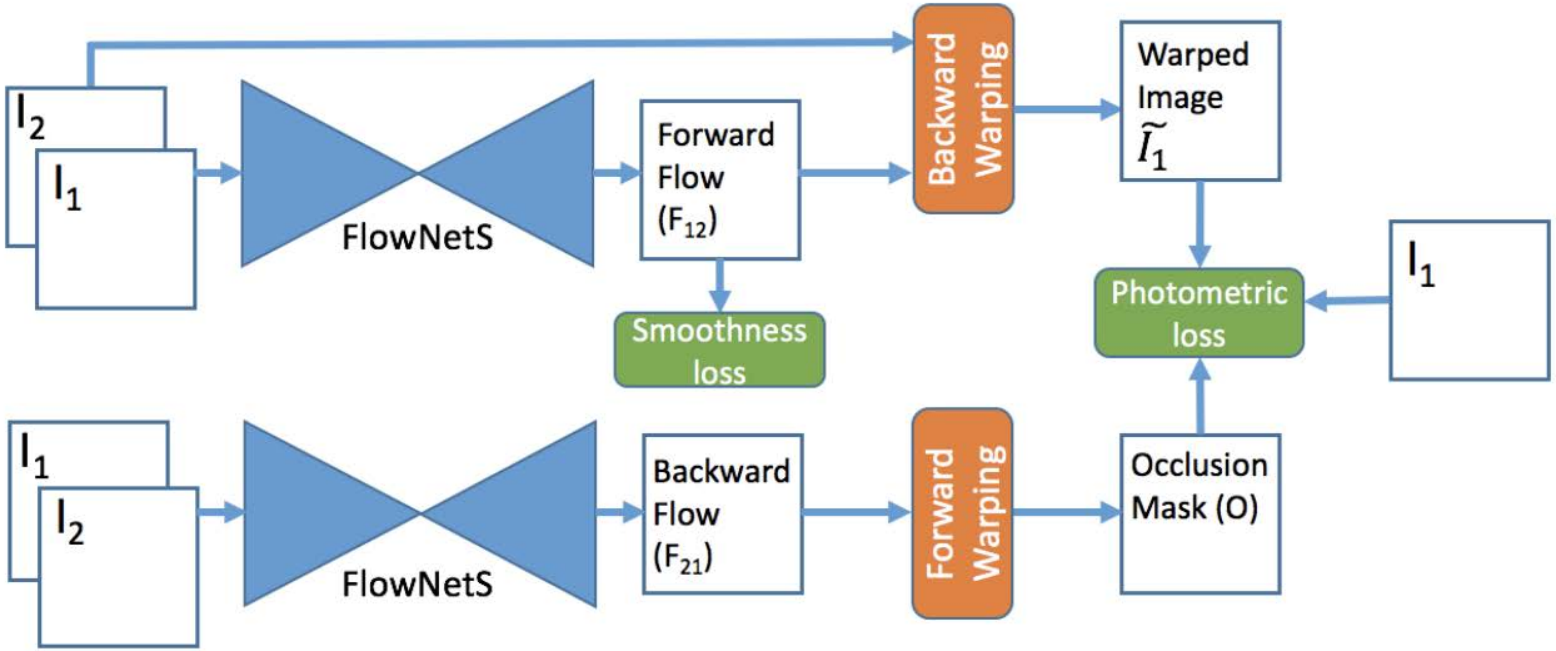
occluded & rigid

I_s

Occlusion mask



Occlusion Map Estimation
Unflow (AAAI 2018)
Occlusion Aware (CVPR 2018)



Training schedule

- Initialize Depth/Pose net and FlowNet by training each of them individually
- Iterative training:

  Repeat:

  1. train Depth/Pose net using optical flow as supervision on non-occluded and static regions.
  2. train FlowNet using depth flow as supervision on occluded and static regions.

Flow-guided Depth and Motion Estimation

On non-occluded and static regions, optical flow performs better and can be used for supervising depth flow: $L = \lambda || (depth\_flow - optical\_flow) * occ\_mask * rigid\_mask ||_1$

Method	Abs	Sq	RMSE	RMSE log	$\delta<1.25$	$\delta<1.25^2$	$\delta<1.25^3$
+flow consist(all)	0.1478	1.0339	5.5460	0.2224	80.21	93.77	97.55
+flow consist(noc + rigid)	0.1437	1.0425	5.3579	0.2179	81.30	94.09	97.58
iterative	0.1412	1.0296	5.3499	0.2161	81.60	94.12	97.58

Depth-guided Flow Estimation

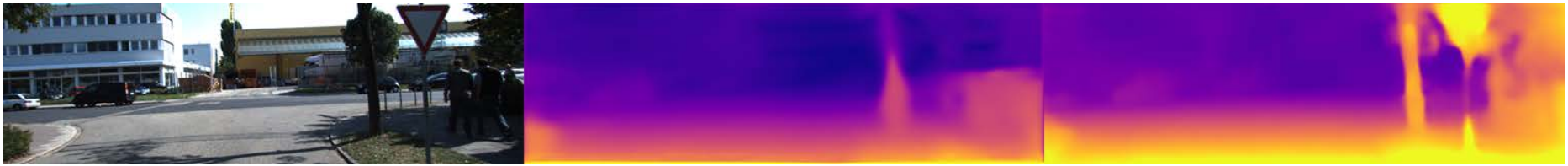
On occluded regions, Depth flow often performance better than optical flow.

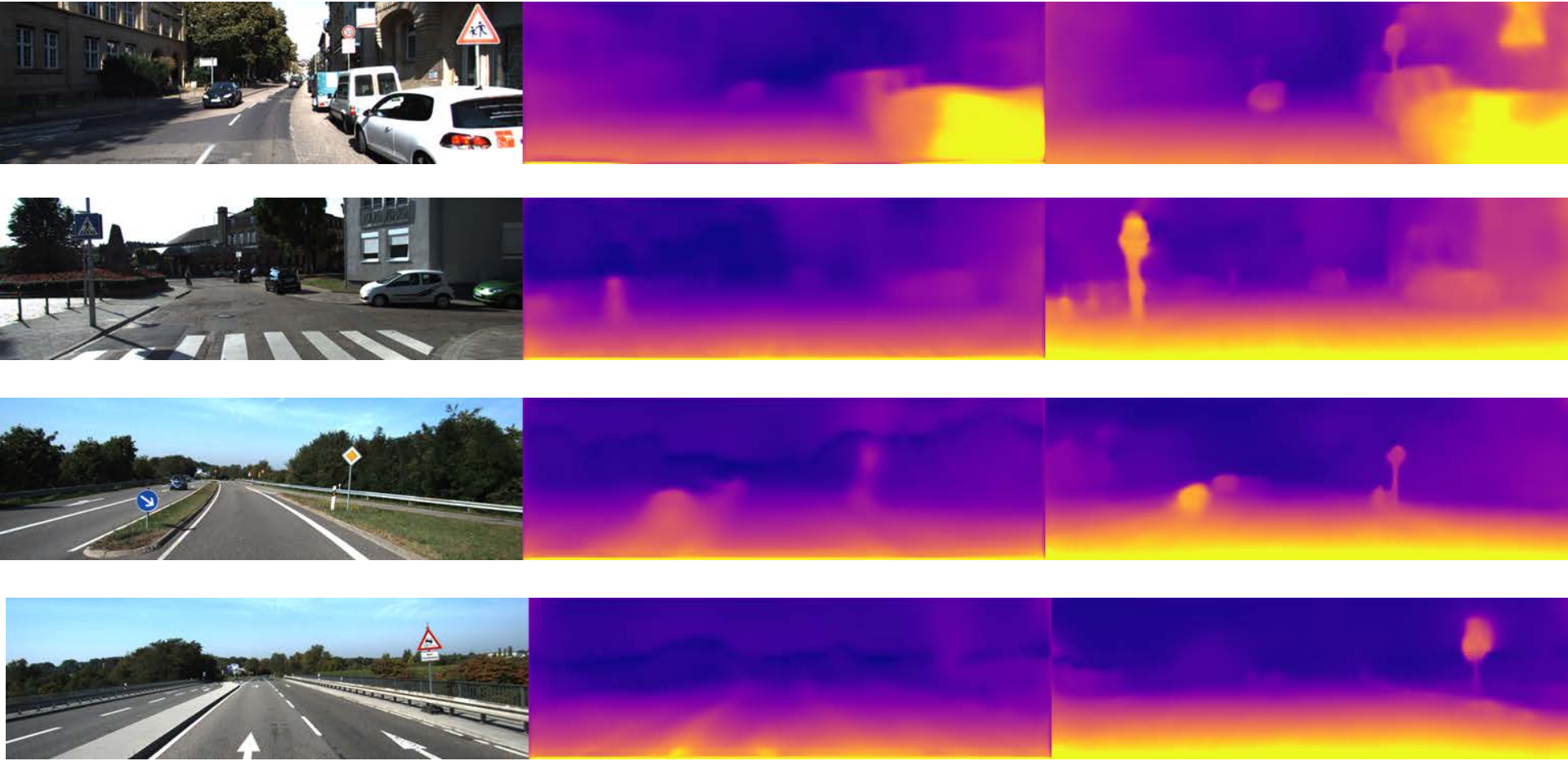
(eigen split) $L = ƛ|| (depth\_flow - optical\_flow) * (1 - occ\_mask) * rigid\_mask ||_1$

	all	noc	occ
baseline	7.07	3.66	23.07
Train together	7.40	4.00	22.96
Train flow (all)	6.46	4.33	16.20
Train flow (noc)	7.17	3.97	21.79
Train flow (occ)	5.84	3.84	15.72

Flow Results on KITTI 2015 test set

Method		F1-bg	F1-fg	F1-all
unsupervised	Ours	**20.61%**	26.32%	**21.56%**
	Multi-frame (ECCV18)	22.67 %	**24.27 %**	22.94 %
	Occlusion Aware(CVPR 18)	-	-	31.2%
	DSTFlow (AAAI17)	-	-	39%
Supervised	Unflow-ft (AAAI 18) DFNet-ft (ECCV 18)	10.15 %	15.93 %	11.11 % 22.82%
	PWC-Net (CVPR 18)	7.87 %	8.03 %	7.90 %
	LiteFlowNet (CVPR18)	9.66 %	**7.99 %**	9.38 %





Moving object mask

GroundTruth Results

Applications

Deep Virtual Stereo Odometry: Leveraging Deep Depth Prediction for Monocular Direct Sparse Odometry (ECCV 2018)