Multi-task Learning in Autonomous Driving Scenarios via Adaptive Feature Refinement Networks

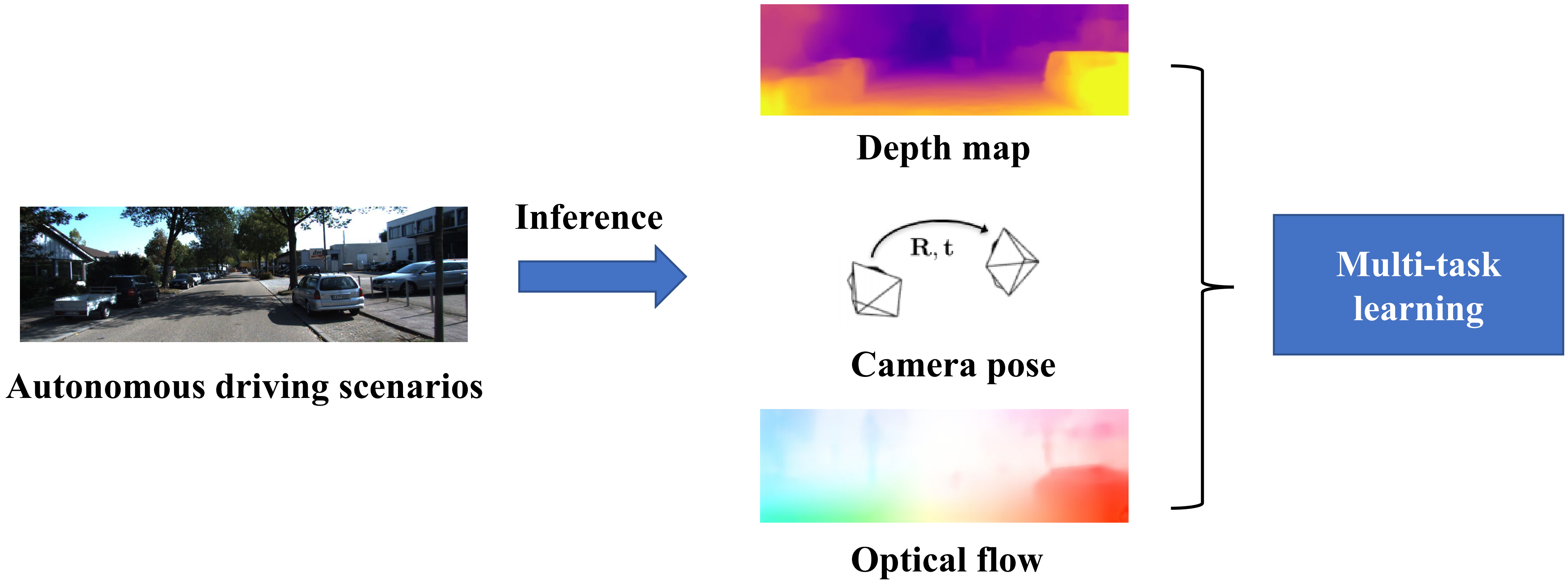
Mingliang Zhai[1], Xuezhi Xiang[1], Ning Lv[1] and Abdulmotaleb EI Saddik[2]

1. Harbin Engineering University

2. University of Ottawa

Contents

Multi-task learning in autonomous driving scenarios

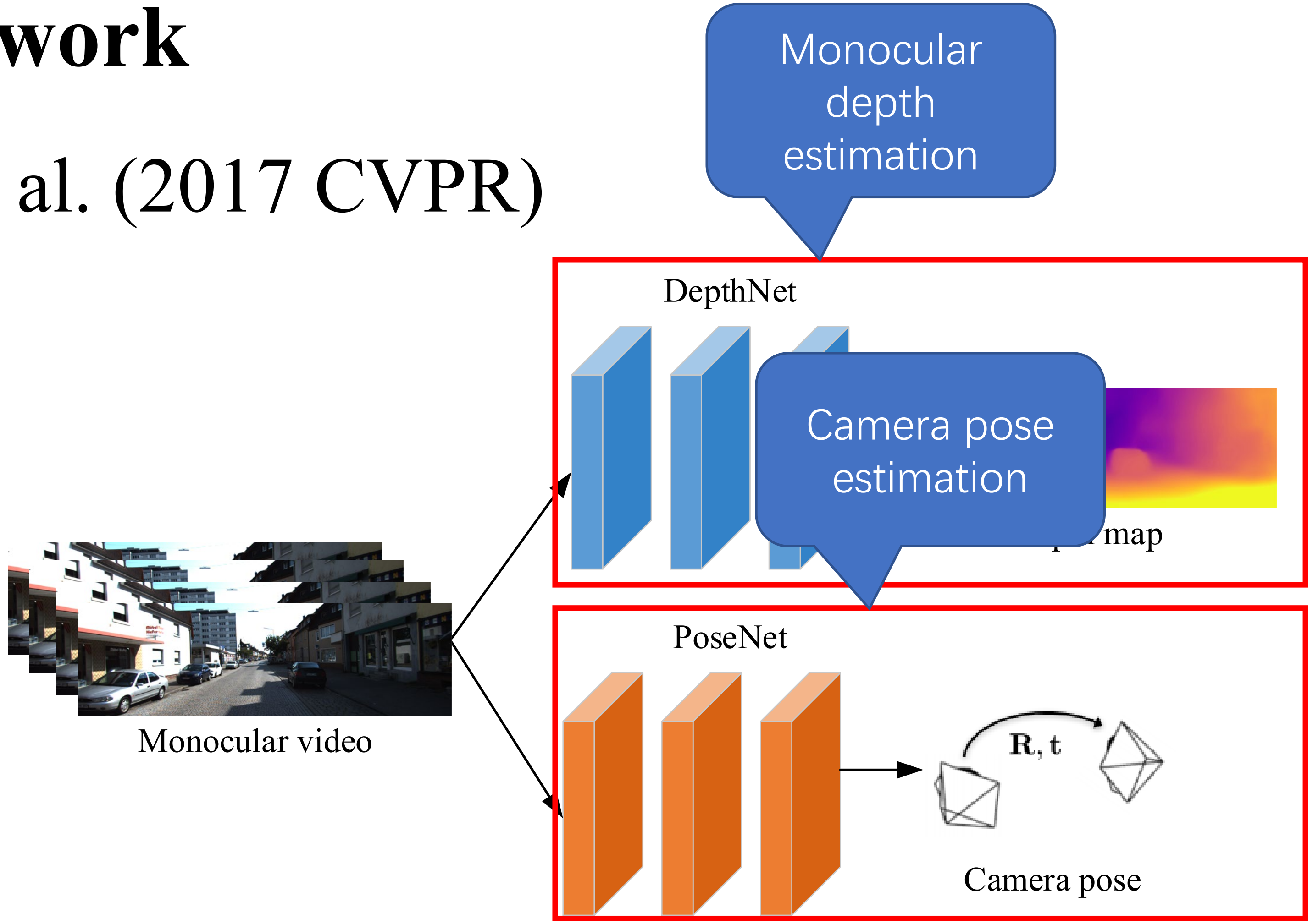


Motivation

- Although existing approaches consider to exploit 3D scene geometry information and can infer flow, depth and camera pose in a unified network, these approaches ignore capturing global channel and spatial dependencies during feature learning and lack the ability to exploit rich contextual information.

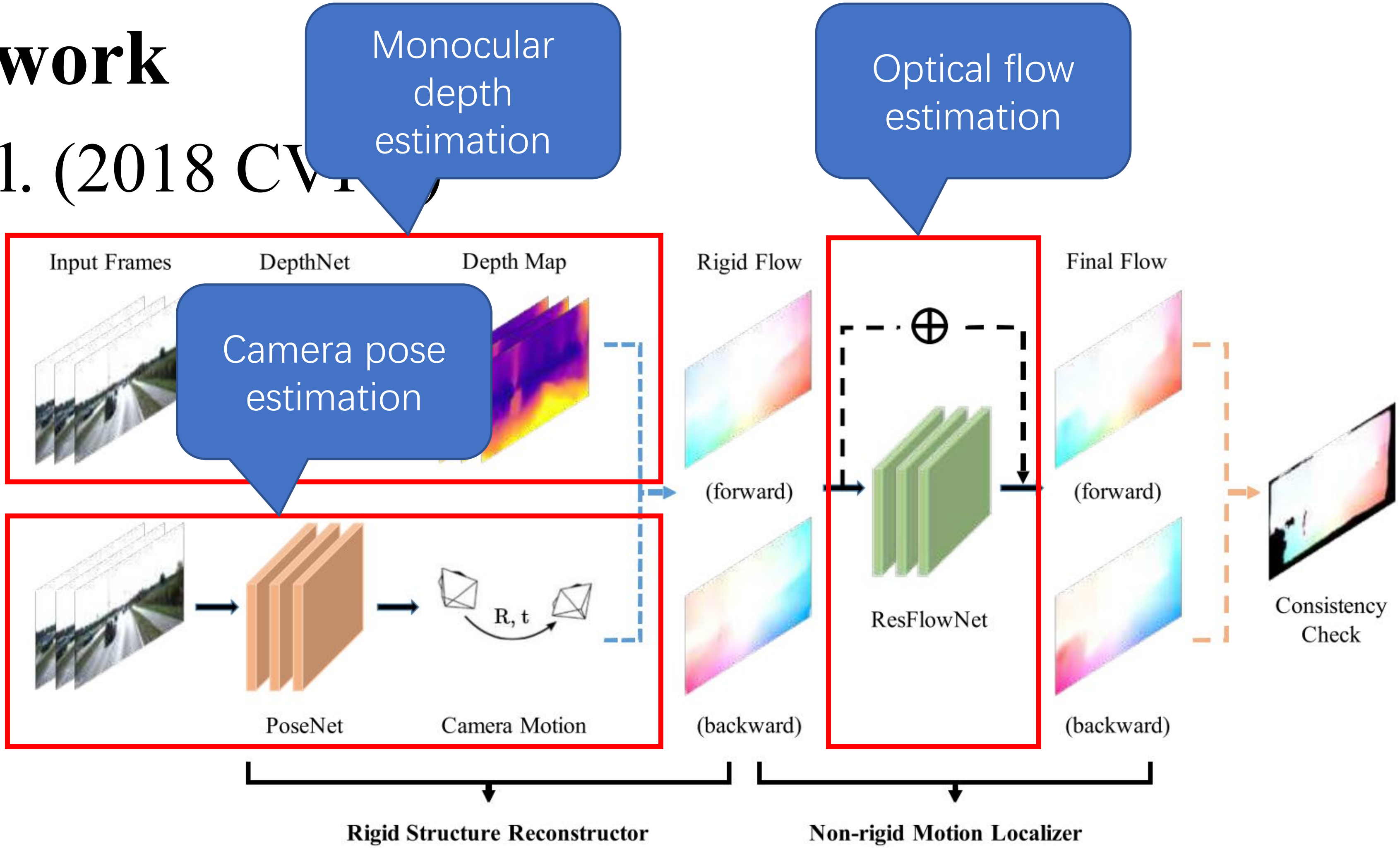
Recent work

- Zhou et al. (2017 CVPR)



Recent work

- Yin et al. (2018 CVPR)



Recent literature

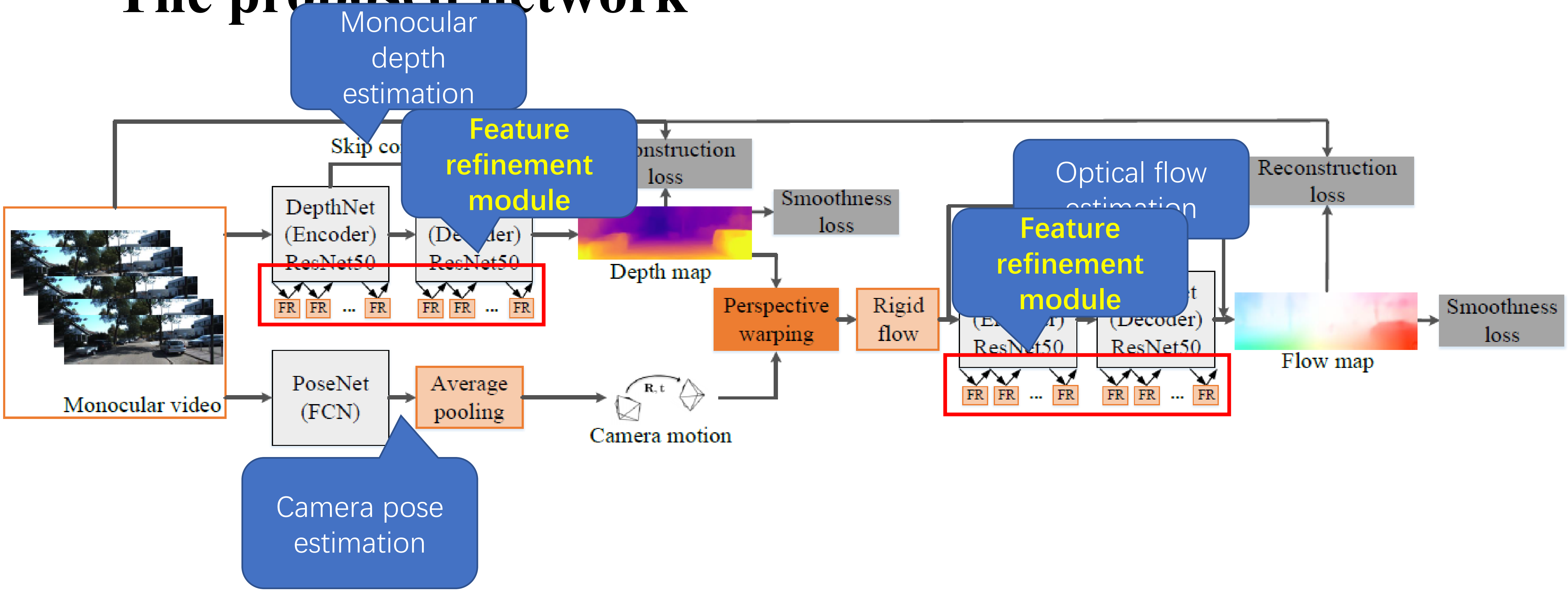
- T. Zhou, M. Brown, N. Snavely, and D. G. Lowe, "Unsupervised learning of depth and ego-motion from video," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017, pp. 6612–6619.
- Z. Yang, P.Wang,W. Xu, L. Zhao, and R. Nevatia, "Unsupervised learning of geometry from videos with edgea ware depth-normal consistency," in AAAI Conference on Artificial Intelligence, 2018.
- Z. Yang, P. Wang, Y. Wang, W. Xu, and R. Nevatia, "Lego: Learning edge with geometry all at once by watching videos," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 225–234.
- R. Mahjourian, M. Wicke, and A. Angelova, "Unsupervised learning of depth and ego-motion from monocular video using 3d geometric constraints," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 5667–5675.
- Z. Yin and J. Shi, "Geonet: Unsupervised learning of dense depth, optical flow and camera pose," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 1983–1992.

Many works are proposed to joint learning of depth, optical flow and camera pose. However, these approaches ignore capturing global channel and spatial dependencies during feature learning and lack the ability to exploit rich contextual information.

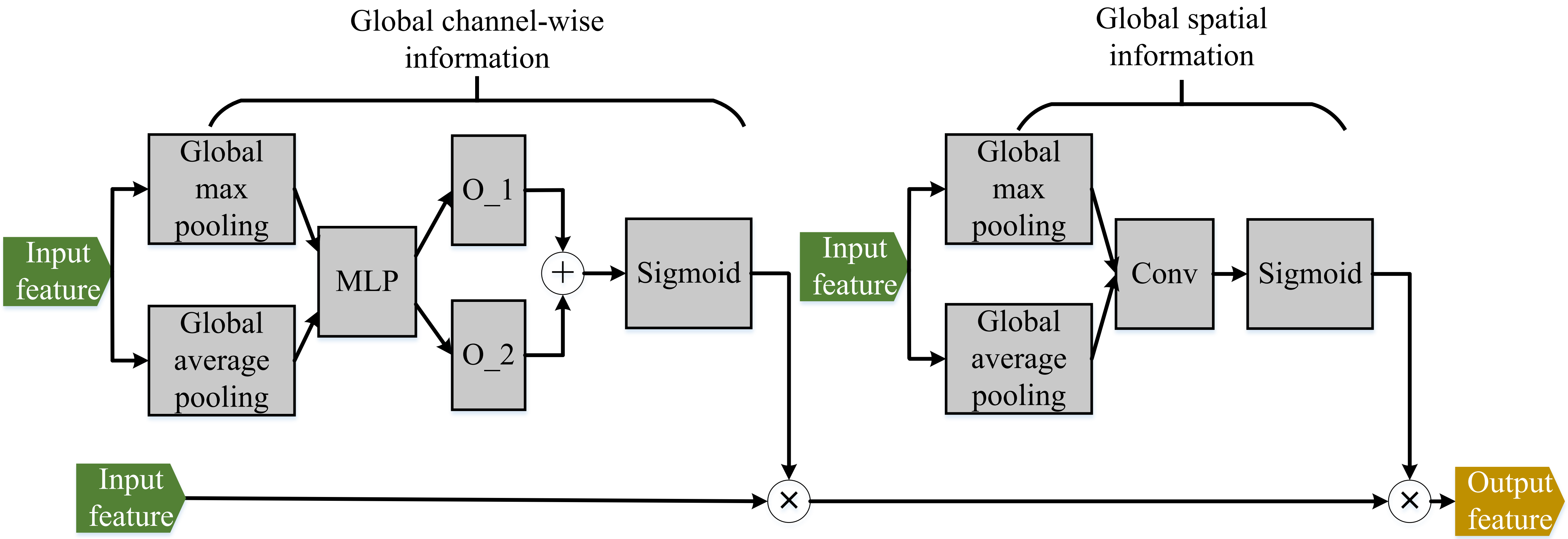
Our Main Contributions

- We combine an adaptive **feature refinement module** and a unified framework for joint learning of optical flow, depth and camera pose estimation in an unsupervised setting.
- The feature refinement **is conducted on both optical flow an depth tasks** for boosting the quality of flow and depth map.
- We observe that our proposed network can achieve comparable results on KITTI dataset.

The proposed network



Feature refinement module



Dataset

KITTI dataset

- Unlabeled monocular image sequence for training. (Autonomous Driving Scenarios)

Quantitative Results (Depth)

Table 1. Performance comparison on KITTI eigen split dataset

Method	Training data	Abs Rel	Sq Rel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
		Lower the better				Higer the better		
Eigen *et al.* [5]	Single image	0.203	1.548	6.307	0.282	0.702	0.890	0.958
Zhan *et al.* [19]	Stereo pair	**0.144**	1.391	5.869	**0.241**	**0.803**	**0.928**	**0.969**
Godard *et al.* [7]	Stereo pair	0.148	1.344	5.927	0.247	**0.803**	0.922	0.964
Gavg *et al.* [6]	Stereo pair	0.152	**1.226**	**5.849**	0.246	0.784	0.921	0.967
Zhou *et al.* [8]	Monocular video	0.208	1.768	6.856	0.283	0.678	0.885	0.957
Yang *et al.* [9]	Monocular video	0.156	1.360	6.641	0.248	0.750	0.914	0.969
Mahjourian *et al.* [11]	Monocular video	0.163	1.240	6.220	0.250	0.762	0.916	0.968
Yang *et al.* [10]	Monocular video	0.162	1.352	6.276	0.252	-	-	-
Yin *et al.* [12]	Monocular video	0.155	1.296	5.857	0.233	0.793	0.931	0.973
Ours	Monocular video	**0.152**	**1.103**	**5.608**	**0.230**	**0.796**	**0.935**	**0.974**

Quantitative Results (Optical flow and camera pose)

Table 2. Performance comparison on KITTI2015 flow training dataset

Method	Supervised	KITTI2015
		Train (AEE)
FlowNetS [1]	Yes	14.19
FlowNetC [1]	Yes	11.49
FlowNet2.0 [2]	Yes	**10.06**
PWC-Net [3]	Yes	10.35
Yin *et al.* [12]	No	10.81
Ren *et al.* [4]	No	16.79
Ours	No	**10.19**

Table 3. Absolute Trajectory Error (ATE) on the KITTI odometry dataset.

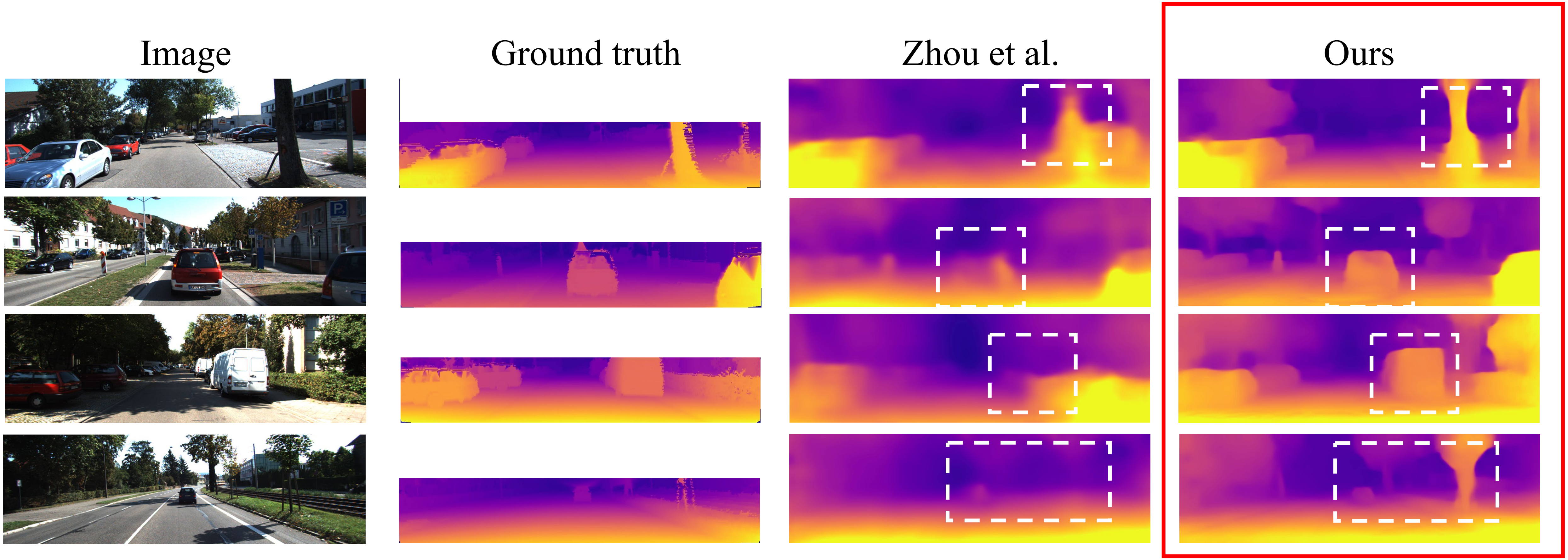
Method	Seq.09	Seq.10
Mean Odom.	0.032±0.026	0.028±0.023
Zhou *et al.* [8]	0.021±0.017	0.020±0.015
Mahjourian *et al.* [11]	0.013±0.010	0.012±0.011
Ours	**0.012±0.013**	**0.012±0.007**

Ablation study

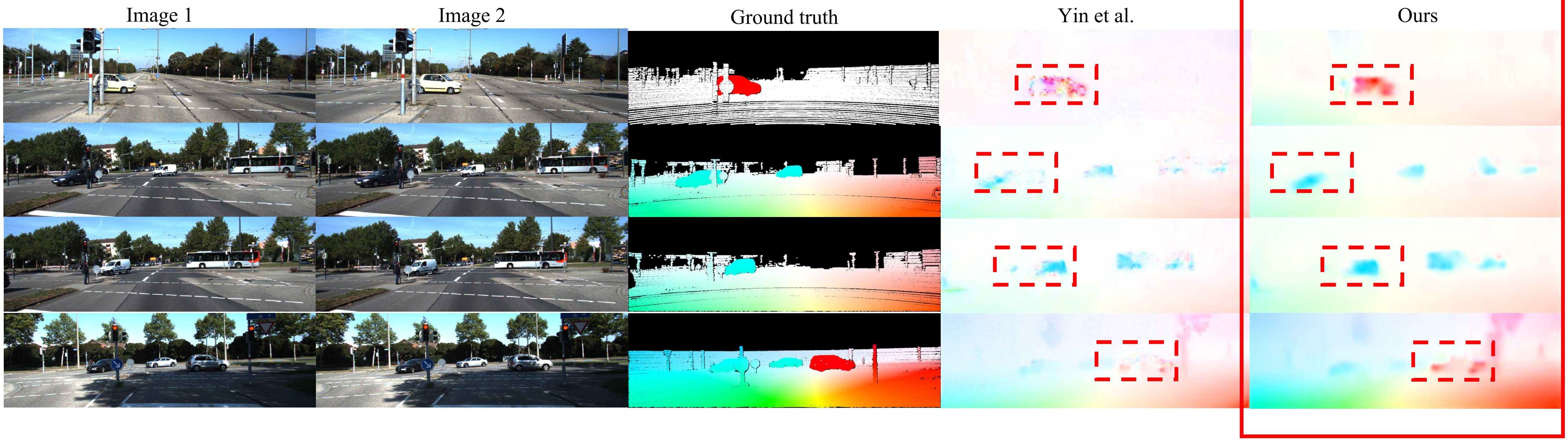
Table 4. Ablation study

Method	Depth			Flow
	Abs Rel	Sq Rel	RMSE	AEE
Ours (w/o FR)	0.155	1.296	5.857	10.81
Ours (full)	**0.152**	**1.103**	**5.608**	**10.19**

Visual Samples (Depth map)



Visual Samples (Flow map)



Conclusion

- In this paper, we introduce an adaptive feature refinement into multi-task learning based framework for depth, optical flow and camera pose estimation.
- The entire network is accomplished in two parts. The first part is design to estimate depth and camera pose, and further calculates rigid flow. The second part is design to estimate the incremental flow. Moreover, the feature refinement module is embedded into depth and flow sub-networks, which can draw global dependencies along channel and spatial aspects.
- To verify the effectiveness of our method, we conduct comprehensive experiments on KITTI dataset. The experimental results show that our model can achieve comparable results on depth, flow and camera pose tasks.

Q&A

If you have any questions, please contact us for more details.
Thank you!

E-mail: zmlshiwo@outlook.com, xiangxuezhi@hrbeu.edu.cn