Spatio-Temporally Consistent Depth Estimation for Dynamic Scenes using 3D Scene Flows

Yu Cai, Tianjiao Jing, Chang Liu, Zhengxuan Lian, Shi-sheng Huang, Hua Huang

Beijing Normal University

Motivation

Dynamic depth estimation continues to be crucial but challenging mainly due to the violation of multi-view consistency raised by dynamic areas. Recent approaches have made impressive progress by implicitly fusing the intra-relation features, but is still limited for heterogeneous dynamic scenes. In this paper, we propose a new intra-relation feature fusion, but significantly improve the fusion quality using an explicit regularization from 3D scene flow cues.

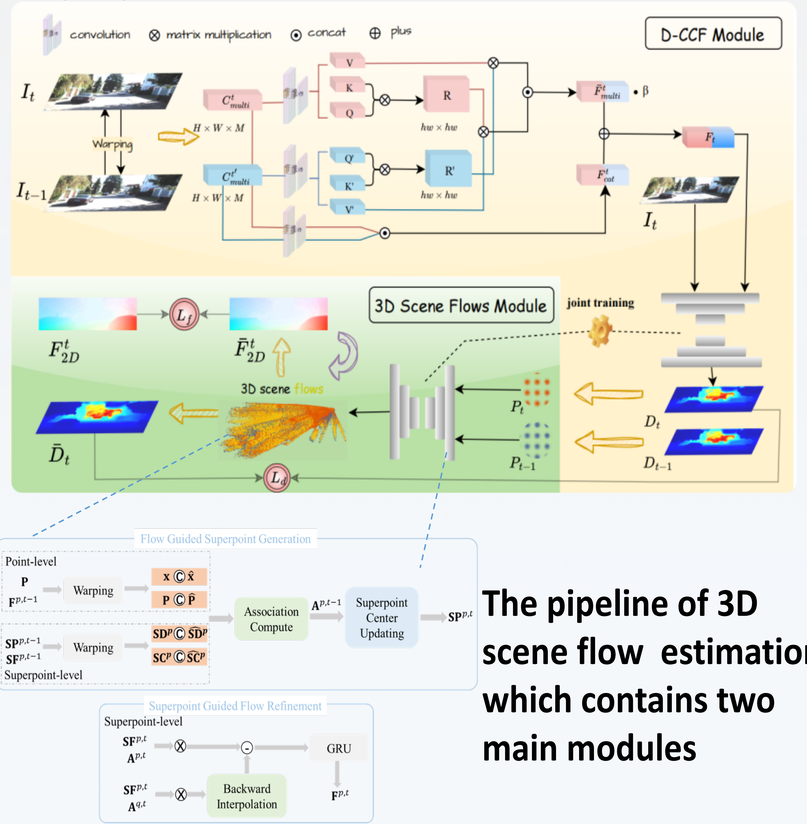
Contribution

- **Introducing the D-CCF Module:** We put forward a Double Cross Cue Fusion (D-CCF) module for depth prediction, which improves the accuracy of depth estimation by effectively fusing the features of successive frames
- **Building 3D Scene Flow Estimation:** We establish an efficient 3D scene flow estimation as explicit 3D spatio-temporal corresponding priors to regularize the depth prediction. This step is crucial as it introduces an effective mechanism to correct the non-coherent depth prediction caused by dynamic areas.
- **Joint Unsupervised Learning:** We perform joint unsupervised learning of both the depth prediction and 3D scene flow estimation. This innovative approach leverages the 3D scene flows to regulate the D-CCF depth prediction, resulting in more accurate dynamic depth prediction towards spatio-temporal consistency.

Method

The framework of the proposed approach, which mainly contains two key components: (1) a **D-CCF module** to predict depth map using multiple frame feature fusion, and (2) a **3D scene flow estimation module**, which is used to effectively regularize the depth prediction learning.

In order to jointing learn both the depth prediction and 3D scene flow estimation simultaneously, we formulate a loss function L by combining the **depth warping loss Ld** and **flow warping loss Lf** introduced by the 3D scene flow performed on the depth prediction.



The pipeline of 3D scene flow estimation, which contains two main modules

Result

Test results of KITTI and DDAD datasets

TABLE I
COMPARISON OF DIFFERENT METHODS FOR DEPTH ESTIMATION (KITTI)

Method	Error Metric (lower is better)				Accuracy Metric (higher is better)		
	AbsRel	SqRel	RMSE	$RMSE_{log}$	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Manydepth	0.071	0.343	3.184	0.108	0.945	0.991	0.998
DynamicDepth	0.068	0.296	3.067	0.106	0.945	0.991	0.998
MonoRec	0.050	0.290	2.266	0.082	0.972	0.991	0.996
MaGNet	0.057	0.215	2.597	0.088	0.967	0.996	0.999
CCF Module	0.046	0.155	2.112	0.076	0.973	0.996	0.999
AFNet	0.044	0.132	**1.712**	0.069	0.980	0.997	0.999
Ours	**0.036**	**0.098**	1.721	**0.057**	**0.985**	**0.998**	**1.000**

TABLE II
COMPARISON OF DIFFERENT METHODS FOR DEPTH ESTIMATION (DDAD)

Method	Error Metric (lower is better)				Accuracy Metric (higher is better)		
	AbsRel	SqRel	RMSE	$RMSE_{log}$	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Monodepth2	0.381	8.387	21.277	0.371	0.587	-	-
Manydepth	0.146	3.258	14.098	-	0.836	-	-
MonoRec	0.158	3.102	**7.553**	0.227	0.854	0.931	0.961
CCF Module	0.158	2.416	9.855	0.299	0.747	0.894	0.947
MaGNet	0.208	2.641	10.739	0.382	0.620	0.878	0.942
Ours	**0.099**	**1.449**	8.311	**0.153**	**0.911**	**0.965**	**0.980**

Visualization on KITTI

Visualization on DDAD