What Matters for 3D Scene Flow Network

Guangming Wang[1], Yunzhe Hu[1], Zhe Liu[1], Yiyang Zhou[2], Masayoshi Tomizuka[2], Wei Zhan[2], and Hesheng Wang[1]
[1]Shanghai Jiao Tong University [2]University of California, Berkeley

ECCV TEL AVIV 2022

Task

- To predict 3D scene flow (point-wise displacement vectors) from two consecutive frames of point clouds.

Preliminary

- Su et al. [1] proposed Pyramid, Warping, and Cost volume (PWC) structure in the optical flow network to refine the optical flow through a coarse-to-fine approach. Wu et al. [2] introduced the PWC structure to 3D scene flow estimation.
- Wang et al. [3] developed the attentive cost volume between two frames of point clouds to estimate 3D scene flow.

Motivation

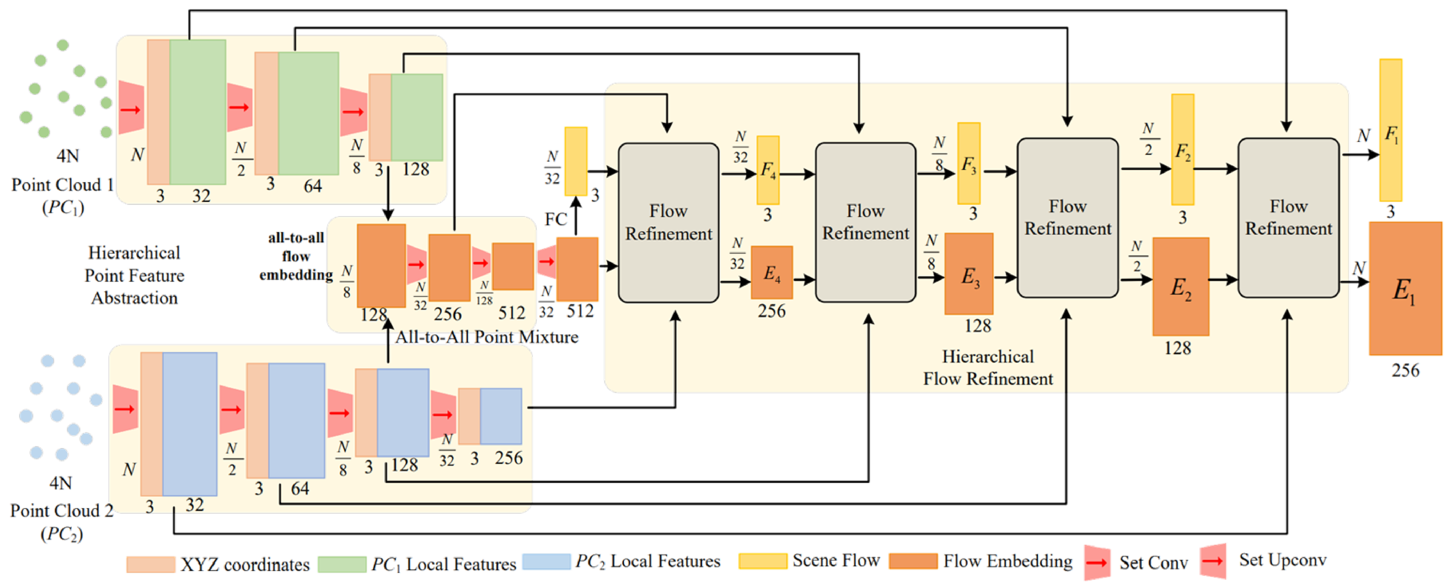
- Previous flow-embedding based methods tend to miss correct yet distant matching point due to limited search range of KNN.
- The estimated correspondence may not conform to the bidirectional consistency.
- The performance gain of designed components and techniques from recent related works are unclear for a PWC-based scene flow network.

Contribution

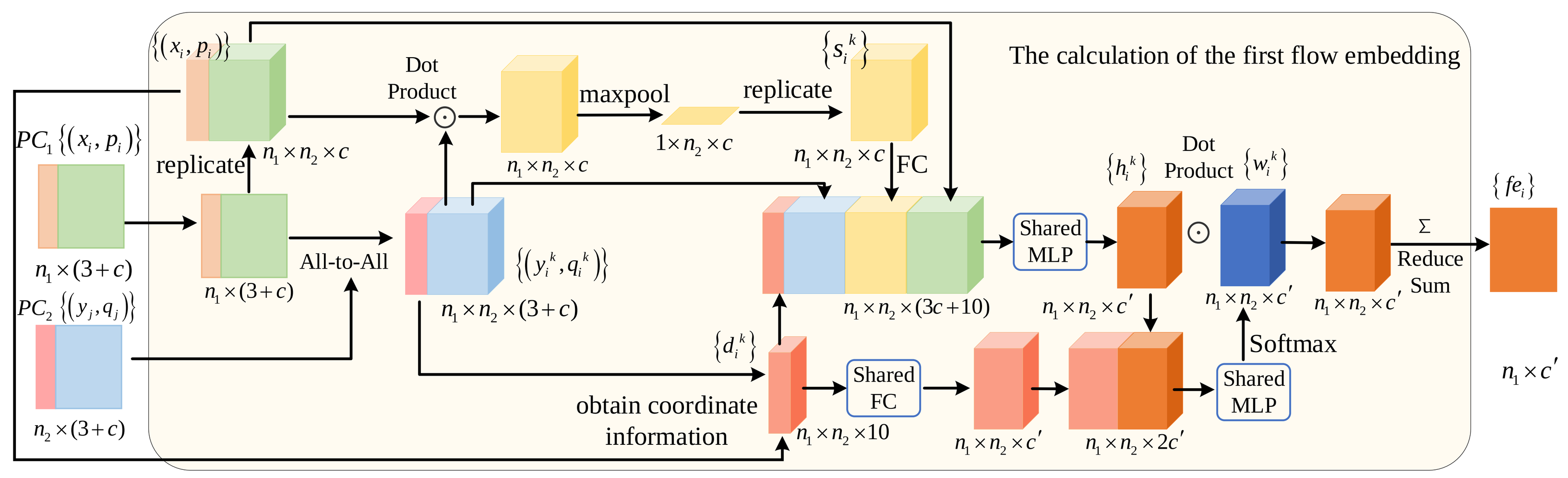
- A novel all-to-all flow embedding module with backward reliability validation is proposed for reliable correlation between point clouds. The all-to-all mechanism is adopted to capture reliable match candidates from the distance, and backward information is integrated in the inference process to validate the matching consistency.
- Different designs and techniques of 3D scene flow network are widely compared and analyzed. Point Similarity Calculation, Designs of Scene Flow Predictor, Input Elements of Scene Flow Predictor, and Flow Refinement Level Design are individually discussed and evaluated to showcase what matters in 3D scene flow network.
- Experiments demonstrate that our model achieves state-of-the-art performance, reducing EPE3D metric by at least 38.2% on FlyingThings3D dataset and 24.7% on KITTI Scene Flow dataset.

3DFlow-Net

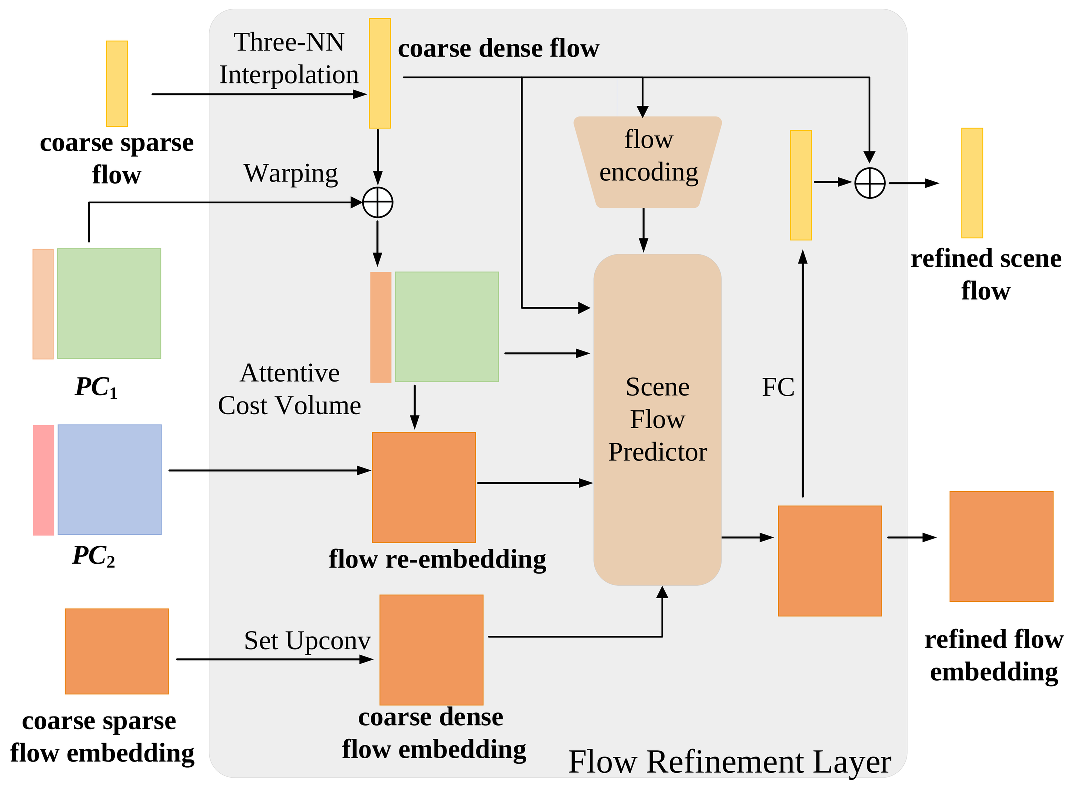
- 3D Scene Flow Network



- All-to-all flow embedding with backward validation



- Flow refinement module



Results

- Results on datasets without occlusion

Evaluation Dataset	Method	Training Data	Input	Sup.	EPE3D	Acc3D Strict	Acc3D Relax	Outliers3D	EPE2D	Acc2D
FlyingThings 3D dataset [23]	FlowNet3 [10]	Quarter	RGB stereo	Full	0.4570	0.4179	0.6168	0.6050	5.1348	0.8125
	ICP [1]	No	Points	Full	0.4062	0.1614	0.3038	0.8796	23.2280	0.2913
	FlowNet3D [19]	Quarter	Points	Full	0.1136	0.4125	0.7706	0.6016	5.9740	0.5692
	SPLATFlowNet [31]	Quarter	Points	Full	0.1205	0.4197	0.7180	0.6187	6.9759	0.5512
	HPLFlowNet [4]	Quarter	Points	Full	0.0804	0.6144	0.8555	0.4287	4.6723	0.6764
	HPLFlowNet [4]	Complete	Points	Full	0.0696	—	—	—	—	—
	PointPWC-Net [50]	Complete	Points	Full	0.0588	0.7379	0.9276	0.3424	3.2390	0.7994
	HALFlow [44]	Quarter	Points	Full	0.0511	0.7808	0.9437	0.3093	2.8739	0.8056
	HALFlow [44]	Complete	Points	Full	0.0492	0.7850	0.9468	0.3083	2.7555	0.8111
	FLOT [28]	Complete	Points	Full	0.0520	0.7320	0.9270	0.3570	—	—
	HCRF-Flow [17]	Quarter	Points	Full	0.0488	0.8337	0.9507	0.2614	2.5652	0.8704
	PV-RAFT [49]	Complete	Points	Full	0.0461	0.8169	0.9574	0.2924	—	—
	FlowStep3D [16]	Complete	Points	Full	0.0455	0.8162	0.9614	0.2165	—	—
	Ours	Quarter	Points	Full	0.0317	0.9109	0.9757	0.1673	1.7436	0.9108
	Ours	Complete	Points	Full	**0.0281**	**0.9290**	**0.9817**	**0.1458**	**1.5229**	**0.9279**
KITTI dataset [25]	FlowNet3 [10]	Quarter	RGB stereo	Full	0.9111	0.2039	0.3587	0.7463	5.1023	0.7803
	ICP [1]	No	Points	Full	0.5181	0.0669	0.1667	0.8712	27.6752	0.1056
	FlowNet3D [19]	Quarter	Points	Full	0.1767	0.3738	0.6677	0.5271	7.2141	0.5093
	SPLATFlowNet [31]	Quarter	Points	Full	0.1988	0.2174	0.5391	0.6575	8.2306	0.4189
	HPLFlowNet [4]	Quarter	Points	Full	0.1169	0.4783	0.7776	0.4103	4.8055	0.5938
	HPLFlowNet [4]	Complete	Points	Full	0.1113	—	—	—	—	—
	PointPWC-Net [50]	Complete	Points	Full	0.0694	0.7281	0.8884	0.2648	3.0062	0.7673
	HALFlow [44]	Quarter	Points	Full	0.0692	0.7532	0.8943	0.2529	2.8660	0.7811
	HALFlow [44]	Complete	Points	Full	0.0622	0.7649	0.9026	0.2492	2.5140	0.8128
	FLOT [28]	Complete	Points	Full	0.0560	0.7550	0.9080	0.2420	—	—
	HCRF-Flow [17]	Quarter	Points	Full	0.0531	0.8631	0.9444	0.1797	2.0700	0.8656
	PV-RAFT [49]	Complete	Points	Full	0.0560	0.8226	0.9372	0.2163	—	—
	FlowStep3D [16]	Complete	Points	Full	0.0546	0.8051	0.9254	**0.1492**	—	—
	Ours	Quarter	Points	Full	0.0332	0.8931	0.9528	0.1690	1.2186	0.9373
	Ours	Complete	Points	Full	**0.0309**	**0.9047**	**0.9580**	0.1612	**1.1285**	**0.9451**

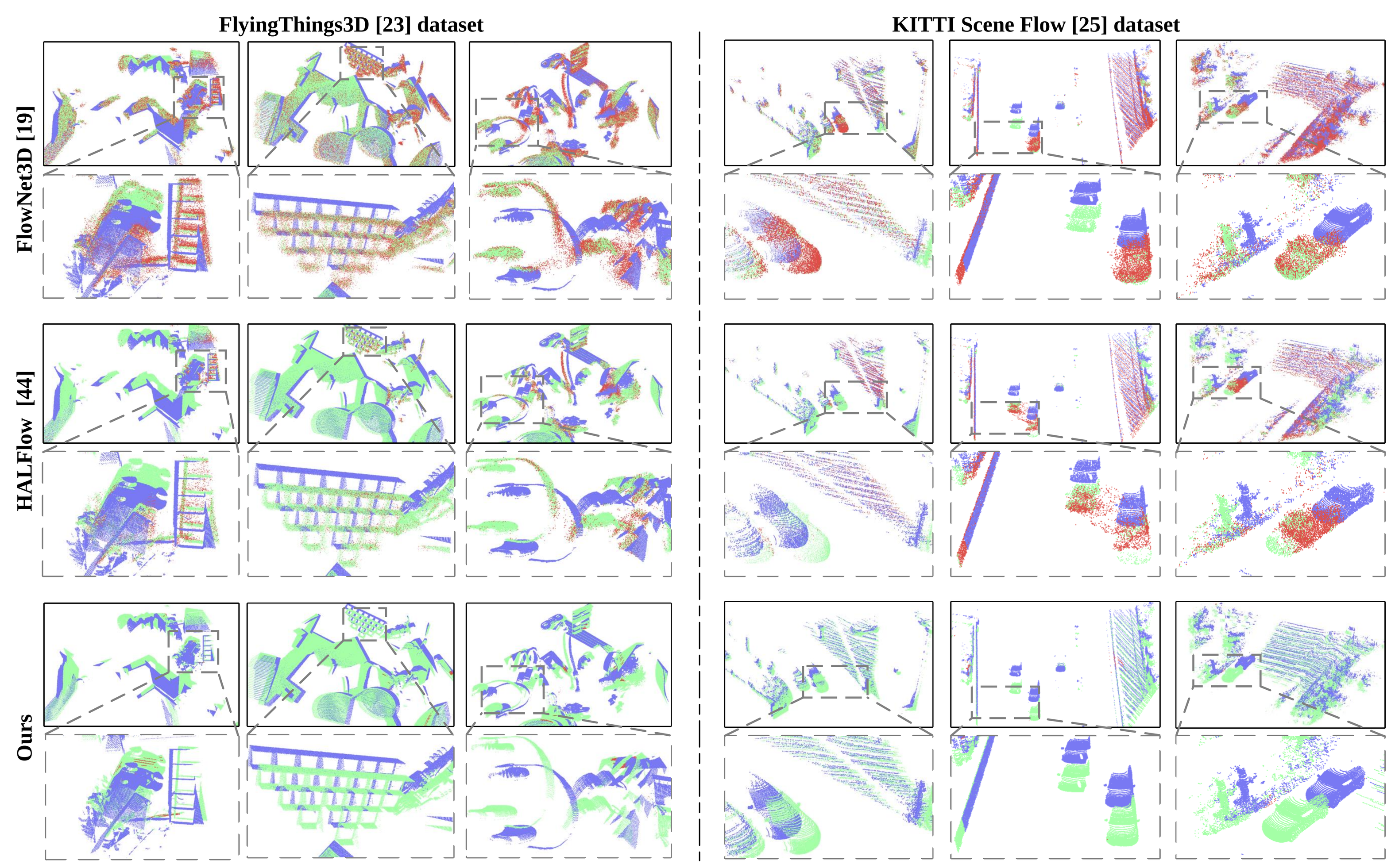
- Results on datasets with occlusion

Evaluation Dataset	Method	Input	Sup.	EPE3D	Acc3D Strict	Acc3D Relax	Outliers
FlyingThings 3D dataset [23]	FlowNet3D [19]	Points	Full	0.169	0.254	0.579	0.789
	FLOT [28]	Points	Full	0.156	0.343	0.643	0.700
	FESTA [47]	Points	Full	0.111	0.431	0.744	—
	Ours	Points	Full	**0.063**	**0.791**	**0.909**	**0.279**
KITTI dataset [25]	Self-Point-Flow [18]	Points	Self	0.105	0.417	0.725	0.501
	FlowNet3D [19]	Points	Full	0.173	0.276	0.609	0.649
	FLOT [28]	Points	Full	0.110	0.419	0.721	0.486
	FESTA [47]	Points	Full	0.097	0.449	0.833	—
	Ours	Points	Full	**0.073**	**0.819**	**0.890**	**0.261**

- Ablation study on FT3D dataset

	Method	EPE3D	Acc3D Strict	Acc3D Relax	Outliers	EPE2D	Acc2D
(a)	Ours w/o backward validation	0.0332	0.9044	0.9743	0.1766	1.8221	0.9065
	Ours w/o backward validation and all-to-all mechanism	0.0349	0.9001	0.9725	0.1798	1.9819	0.9032
	Ours (full, with backward validation and all-to-all mechanism)	**0.0317**	**0.9109**	**0.9757**	**0.1673**	**1.7436**	**0.9108**
(b)	Ours (with product similarity)	0.0356	0.8872	0.9692	0.1953	1.9872	0.8870
	Ours (with cosine product similarity)	0.0370	0.8755	0.9670	0.2142	2.0637	0.8746
	Ours (with normalized product similarity)	0.0339	0.8971	0.9724	0.1845	1.8790	0.8965
	Ours (full, with concatenated similarity)	**0.0317**	**0.9109**	**0.9757**	**0.1673**	**1.7436**	**0.9108**
(c)	Ours (replace Scene Flow Predictor with GRU)	0.0350	0.8892	0.9668	0.1827	1.9274	0.8896
	Ours (full, with Scene Flow Predictor)	**0.0317**	**0.9109**	**0.9757**	**0.1673**	**1.7436**	**0.9108**
(d)	Ours w/o features of PC_1 in Scene Flow Predictor	0.0333	0.9047	0.9743	0.1740	1.8428	0.9073
	Ours w/o up-sampled flow embedding in Scene Flow Predictor	0.0380	0.8732	0.9642	0.2099	2.0953	0.8785
	Ours w/o coarse flow in Scene Flow Predictor	0.0323	0.9076	0.9750	0.1717	1.7760	0.9083
	Ours w/o flow feature in Scene Flow Predictor	0.0327	0.9061	0.9748	0.1740	1.8063	0.9074
	Ours (full, with complete five inputs in Scene Flow Predictor)	**0.0317**	**0.9109**	**0.9757**	**0.1673**	**1.7436**	**0.9108**
(e)	Ours (with interpolation estimating 2048 points' flow)	0.0359	0.8844	0.9691	0.2004	1.9511	0.8911
	Ours (with interpolation estimating 8192 points' flow)	0.0332	0.9043	0.9739	0.1740	1.8039	0.9076
	Ours (full, with flow refinement estimating 2048 points' flow)	**0.0317**	**0.9109**	**0.9757**	**0.1673**	**1.7436**	**0.9108**

Visualization



[1] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pages 8934–8943, 2018.
[2] Wenxuan Wu, Zhi Yuan Wang, Zhuwen Li, Wei Liu, and Li Fuxin. In European conference on computer vision (ECCV), pages 88-107, 2020.
[3] Guangming Wang, Xinrui Wu, Zhe Liu, and Hesheng Wang. Hierarchical attention learning of scene flow in 3d point clouds. In IEEE Transactions on Image Processing, 2021, doi: 10.1109/TIP.2021.3079796.