Weakly supervised learning of rigid 3D scene flow

Zan Gojcic, Or Litany, Andreas Wieser, Leonidas J. Guibas, Tolga Birdal

Project page

3D Scene Flow Estimation

Given two point clouds $\mathbf{X} \in \mathbb{R}^{N \times 3}$ and $\mathbf{Y} \in \mathbb{R}^{M \times 3}$ acquired by the same observer in two time steps, the goal is to estimate the displacement vectors $\mathbf{V} = f_\theta(\mathbf{X}, \mathbf{Y}) \quad s.t. \quad \mathbf{X} + \mathbf{V} \approx \mathbf{Y}$.

Prior Work

- **Fully supervised methods:** trained on sythetic data such as FT3D (large domain gap)
- **Unsupervised:** trained on the target domain, but fail to reach competitive performance

FlyingThings3D[1]

stereoKITTI[2]

References

[1] Mayer, Nikolaus, et al. "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation." In CVPR, 2016.
[2] Menze, Moritz, et al. "Object scene flow." ISPRS Journal of Photogrammetry and Remote Sensing , 2018.
[3] Liu, Xingyu, et al. "Flownet3d: Learning scene flow in 3d point clouds." In CVPR, 2019.
[4] Gu, Xiuye, et al. "Hplflownet: Hierarchical permutohedral lattice flownet for scene flow estimation on large-scale point clouds." In CVPR, 2019.
[5] Wu, Wenxuan, et al. "Pointpwc-net: A coarse-to-fine network for supervised and self-supervised scene flow estimation on 3d point clouds." In ECCV, 2020.
[6] Tishchenko, Ivan, et al. "Self-Supervised Learning of Non-Rigid Residual Flow and Ego-Motion." In 3DV, 2020.
[7] Puy, Gilles, Alexandre Boulch, and Renaud Marlet. "FLOT: Scene Flow on Point Clouds guided by Optimal Transport." In ECCV, 2020.
[8] Geiger, Andreas, et al. "Are we ready for autonomous driving? the kitti vision benchmark suite." In CVPR, 2012.
[9] Sun, Pei, et al. "Scalability in perception for autonomous driving: Waymo open dataset." In CVPR, 2020.

Our Formulation

Scenes can be decomposed into agents that move as rigid bodies

This object level representation:

- prevents physically non viable flow estimates
- provides a hollistic understanding of the scene
- enables test-time optimization
- allows reducing the supervision requirements

Network Architecture

- **Shared backbone network with task-specific heads**
- **Can be trained in a weakly supervised manner directly on the target domain**

Loss function: $\mathcal{L} = \mathcal{L}_{\mathrm{BG}} + \mathcal{L}_{\mathrm{ego}} + \mathcal{L}_{\mathrm{FG}}$

Results

LidarKITTI[8]:

Input point clouds   3D endpoint error   Aligned point clouds

Generalization to Waymo Open[9]:

Input point clouds   Aligned point clouds   Inferred object masks

Comparison to FLOT[7]:

Input point clouds   FLOT   Ours