Review for Self-Supervised Monocular Scene Flow Estimation

Can Koz

March 21, 2021

1 Summary

Monocular scene flow estimation is obtaining 3D structure and 3D motion from two consecutive images. The difference between optical flow and scene flow is scene flow analyzes 3D motion fields in real world while optical flow analyzes 2D motion field in an image. In this paper, authors propose a novel monocular scene flow technique based on optical flow approach that achieves competitive accuracy and real time performance. They tested their method on KITTI scene flow dataset and Eigen Split. Instead of multi task complex approaches, using a joint decoder for depth and scene flow, authors achieve state of the art performance in monocular scene flow.

2 Strengths

- Scene flow estimation is an ill posed problem due depth-scale ambiguity. Authors use an inverse problem technique by estimating the scene flow by utilizing a optical flow pipeline where optical flow is 2D projection of 3D points.
- Authors use joint decoder for depth and flow estimation to simplify the problem. Other methods use multi task CNNs that require complex training schedule and joint optimization.
- Authors introduce a self-supervised loss function and data augmentation scheme.

3 Weaknesses

- While it is mentioned in the abstract, there is no result section for real-time performance. Similarly there is no result section for algorithm complexity or parameter counts.
- Ablation study results are provided for single decoder vs seperate decoders. These experiments can be extended. Authors proposed a loss function but the effect of this function is not evaluated in ablation study.

4 Evaluation

For evaluation KITTI raw dataset, KITTI Split, KITTI Scene Flow Training and Eigen Split are used. They experimented data augmentation, different loss functions and single decoder vs seperate decoders. Their results show superior accuracy for monocular depth estimaton and comparable accuracy on flow estimation.

5 Final Comments and Future Work

In this paper, authors proposed a simpler approach to an ill posed problem, scene flow estimation and achieved comparable results. As they mentioned in the conclusion their method provides a solid foundation for CNN based monocular scene flow estimation. Instead of multi stage algorithms, end to end methods for scene flow estimation can be improved by enabling additional features specific to scene flow problem.