Review for Unsupervised Learning of Depth and Ego Motion from Video

Can Koz

March 7, 2021

1 Summary

Authors proposed end to end, unsupervised learning method for depth and camera motion estimation. They evaluated their algorithm in KITTI dataset and the method is able to achieve comparable performance in depth and pose estimation with respect to supervised methods.

2 Strengths

The method is unsupervised so it can be trained using sequences of images with no labeling. Authors modeled the entire view synthesis pipeline as the inference procedure of a CNN so that training on large dataset forces the network to learn intermediate tasks of depth and camera motion estimation. Achieving comparable performance with supervised methods is crucial in terms of labeling cost. They simultaneously trained an explainability prediction network, indicating the networks belief on view synthesis.

3 Weaknesses

Performance results are very few and the explanations are not sufficient enough. They used 2 datasets for evaluation and 2 datasets are not enough to evaluate generalization ability of their method. They do not share any measurements for efficiency and computational load. The ablation study is only applied on explainability mask, so it requires more work on ablation study.

4 Evaluation

They used KITTI and Make3D dataset for evaluating cross-dataset generalization ability. They evaluated single view depth estimation and pose estimation. Also they try to evaluate explainability of the method with explainability masks. They performed ablation study on explainability masks. It is hard to make a fair comparison with supervised methods but their results are comparable to that of the supervised baselines.

5 Final Comments and Future Work

The authors try to develop an end to end, unsupervised approach for depth and camera motion estimation. Their approach is valuable since they get rid of the labeling cost. The method can be improved by using additional properties of images such as geometric constraints.