Semi-Supervised Disparity Estimation with Deep Feature Reconstruction
Julia Guerrero-Viu* Sergio Izquierdo* Philipp Schröppel Thomas Brox

guerrero, izquierd, schroepp, brox@cs.uni-freiburg.de

Problem Statement

- Disparity estimation models trained on synthetic data have limited generalization. Obtaining labels for supervised fine-tuning on the target domain is expensive.

- Self-supervision based on view reconstruction allows label-free training on the target domain, but performs worse, especially on challenging areas due to limitations of the common photometric consistency.

- Self-supervision based on deep feature reconstruction may help to overcome photoconsistency limitations. However, this requires further analysis.

Contributions

- We propose a semi-supervised pipeline for disparity estimation with supervised training on labeled synthetic data and self-supervised training on unlabeled real data.
  → Improves cross-domain generalization.

- We perform a thorough analysis of deep feature reconstruction.
  → Shows the potential of deep feature reconstruction and analyses problems that limit its effectiveness.

Approach Overview

- We use a DispNet architecture [5] that predicts occlusion masks $M$ and disparity maps $D$ at multiple scales.

- We use a supervised loss for disparity and occlusion predictions on synthetic data and a self-supervised reconstruction-based loss for disparity predictions on real data.

- For self-supervised training, we experiment with either photometric or deep feature reconstruction as supervisory signal.

- We consider feature maps $F_l$, $F_r$ from the Dispnet encoder (first three conv layers) for the deep feature reconstruction.

We perform a self-supervision based on view reconstruction. We use a DispNet architecture trained on synthetic data have disparity estimation models trained on synthetic data have limited generalization. Obtaining labels for supervised fine-tuning on the target domain is expensive.

Self-supervision based on view reconstruction allows label-free training on the target domain, but performs worse, especially on challenging areas due to limitations of the common photometric consistency. Self-supervision based on deep feature reconstruction may help to overcome photoconsistency limitations. However, this requires further analysis.

Experiments

- Training data: FlyingThings3D (synthetic, labeled), KITTI RAW (real, unlabeled)
- Test data: FlyingThings3D-Test, KITTI2015, ETH3D, Middlebury

Deep Feature Reconstruction (DFR)

- Most works apply self-supervision based on view reconstruction: warp the right image according to the predicted disparities and measure reconstruction via photometric consistency.

- Instead, we apply DFR: warp and compute loss on feature maps. → loss is based on consistency of the warped right image feature map and the respective original left image feature map.

Analysis

We identity the following problems of DFR:

1. Higher sensitivity to occlusions.
2. Large dependence on the distance metric and resampling strategy.
3. Tainted information around disparity discontinuities due to convolutional aggregation.
5. High gradient locality that complicates optimization.

On the left figure we illustrate (3) and (4). Despite DFR’s clearly better response in texture-less areas (road), it fails near disparity discontinuities (sky-pole). Due to its higher entropy curve, DFR is less precise on object boundaries (van).

- Our results show improved generalization across domains, outperforming previous works in this setting.
- Our network is drastically faster than most SOTA models.

References


Acknowledgements

We acknowledge partial funding by ‘la Caixa’ Foundation (LCF/BQ/EU19/3110058), by the German Research Foundation (BR 3815/10-1) and by the German Federal Ministry for Economic Affairs and Energy within the project “KI Delta Learning”.