International Journal of Computer Vision (2023) 131:796–812
https://doi.org/10.1007/s11263-022-01726-1

Project to Adapt: Domain Adaptation for Depth Completion from Noisy and Sparse Sensor Data

Adrian Lopez-Rodriguez[1] · Benjamin Busam[2,3] · Krystian Mikolajczyk[1]

Received: 30 April 2021 / Accepted: 21 November 2022 / Published online: 16 December 2022

Abstract

Depth completion aims to predict a dense depth map from a sparse depth input. The acquisition of dense ground-truth annotations for depth completion settings can be difficult and, at the same time, a significant domain gap between real LiDAR measurements and synthetic data has prevented from successful training of models in virtual settings. We propose a domain adaptation approach for sparse-to-dense depth completion that is trained from synthetic data, without annotations in the real domain or additional sensors. Our approach simulates the real sensor noise in an RGB + LiDAR set-up, and consists of three modules: simulating the real LiDAR input in the synthetic domain via projections, filtering the real noisy LiDAR for supervision and adapting the synthetic RGB image using a CycleGAN approach. We extensively evaluate these modules in the KITTI depth completion benchmark.

Keywords Depth completion · Domain adaptation · LiDAR · Sensor fusion

1 Introduction

Motivation Active sensors such as LiDAR determine the distance of objects within a specified range via a sparse sampling of the environment whose density decreases quadratically with the distance. RGB cameras densely capture their field of view, however, monocular depth estimation from RGB is an ill-posed problem that can be solved only up to a geometric scale. The combination of RGB and depth modalities form a rich source for mutual improvements where each sensor can benefit from the advantage of the other.

Many pipelines have been proposed for a fusion of these two inputs [46,47,54,66,74,76]. Ground-truth annotations for this task, however, require elaborate techniques, manual adjustments and are subject to hardware noise or costly and time-consuming labeling. The most prominent publicly available data for this task [65] creates a ground-truth by aligning consecutive raw LiDAR scans that are cleaned from measurement errors, occlusions, and motion artifacts in a post-processing step involving classical stereo reconstruction. Even after the use of this additional data and tedious processing, the signal is not noise-free as discussed in [65]. To avoid the need for such annotations, some methods perform self-supervision [47,69–71,76], where a photometric loss is employed with stereo or video data. The dependence on additional data such as stereo or temporal sequences brings other problems such as line-of-sight issues and motion artifacts from incoherently moving objects. Modern 3D engines can render highly realistic virtual environments [13,20,58] with perfect ground-truth. However, a significant domain gap between real and virtual scenes prevents from successful training on synthetic data only.

Contributions and Outline In contrast to the self-supervised methods [47,69–71,76], we propose to use a domain adaptation approach to address the depth completion problem without real data ground-truth as shown in Fig. 1. We train our method from the synthetic data generated with the driving simulator CARLA [13] and evaluate it on the real KITTI depth completion benchmark [65]. The real LiDAR data is noisy with the main source of noise being the see-through

Communicated by Cheng-Lin Liu.

✉ Adrian Lopez-Rodriguez
al4415@imperial.ac.uk

Benjamin Busam
b.busam@tum.de

Krystian Mikolajczyk
k.mikolajczyk@imperial.ac.uk

1 Imperial College London, London, UK

2 Technische Universität München, Munich, Germany

3 Huawei Noah’s Ark, London, UK