INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

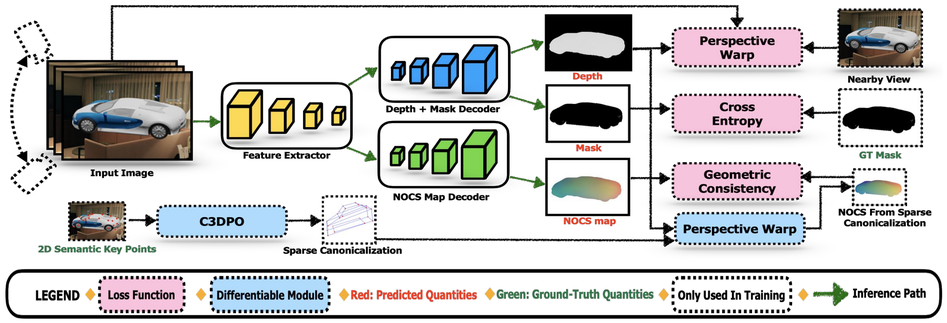
R&D SHOWCASE 2021

DRACO: Weakly Supervised Dense Reconstruction And Canonicalization of Objects

Robotics Research Center

ABSTRACT

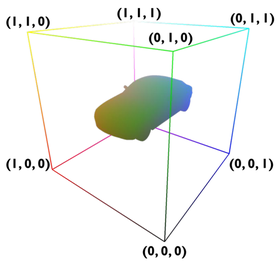
We present DRACO, a method for **D**ense **R**econstruction **A**nd **C**anonicalization of **O**bject shape from one or more RGB images. Canonical shape reconstruction— estimating 3D object shape in a coordinate space canonicalized for scale, rotation, and translation parameters—is an emerging paradigm that holds promise for a multitude of robotic applications. Prior approaches either rely on painstakingly gathered dense 3D supervision, or produce only sparse canonical representations, limiting real-world applicability. DRACO performs dense canonicalization using only weak supervision in the form of camera poses and semantic key points at train time. During inference, DRACO predicts dense object-centric depth maps in a canonical coordinate-space, solely using one or more RGB images of an object. Extensive experiments on canonical shape reconstruction and pose estimation show that DRACO is competitive or superior to fully-supervised methods.



(a) DRACO Pipeline

NOCS Canonical Space

We adopt the recently proposed "Normalised Object Coordinate Space"---or NOCS---as our canonical representation. NOCS is a 3D space contained within a unit cube. Each location in the unit cube is mapped to its corresponding color in the RGB intensity space. "NOCS maps" are 2D projections of a 3D reconstruction where the RGB intensity of each pixel directly corresponds to its 3D location within the unit cube. In short, NOCS enables us to represent all objects from a specific category with respect to a common reference frame.

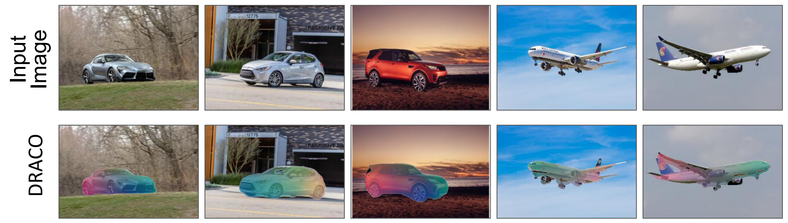


(b) NOCS

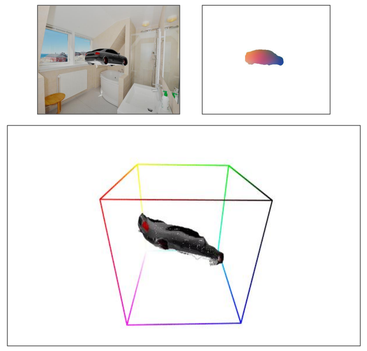
Please Visit https://bit.ly/37lUhlx To Learn More About DRACO

MOTIVATION & OBJECTIVE

Given one or more images of an object, DRACO recovers a dense canonical reconstruction of the visible surfaces. What's special about the reconstructions is that they are "canonicalized" for 6D pose and size, meaning that all instances from a given category are reconstructed in the same 3D coordinate system, regardless of how they are imaged. This can greatly benefit robotics tasks like pose estimation, manipulation, object-based SLAM, and more.



(c) DRACO on Real Images

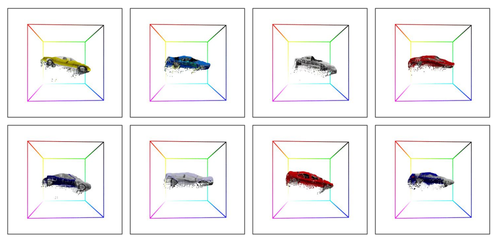


(d) DRACO on DRACO20K

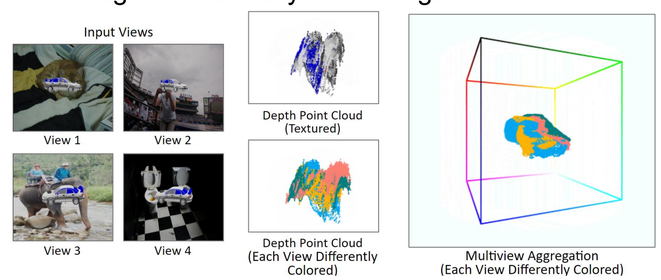
Prior approaches either require hard to obtain dense 3D supervision or only produce sparse canonical reconstructions unsupervised approaches. We address both these limitations by performing dense canonicalization using weak supervision in the form of camera motion and 2D semantic keypoint information.

DRACO - METHOD

DRACO takes as input a single image and predicts an object-centric depth map, a foreground mask, and a "NOCS map". We do not require any dense 3D supervision, and only assume camera motion and 2D key points to be specified at train time. At inference time, DRACO only needs one or more RGB images to perform dense canonicalization. In our method, the encoded features of the image are passed through a depth and mask decoder and a NOCS decoder. We obtain an independent NOCS supervisory signal by leveraging sparse canonicalization and predicted depth. DRACO generalizes remarkably well to real-world images without any fine tuning.



(e) Additional Results on DRACO20K



(f) Multiview Aggregation

Authors: Rahul Sajnani, AadilMehdi Sanchawala, K. M. Jatavallabhula, Srinath Sridhar, K. Madhava Krishna Research Center Name:Robotics Research Center