Geometric Deep Learning for Monocular Object Orientation Estimation

Rene Vidal, Johns Hopkins University

Estimating the 3D orientation of an object given a single 2D image of the object, is an important component of traditional computer vision problems like scene understanding and 3D reconstruction as well as modern vision challenges like autonomous driving, augmented reality and robot manipulation. This talk will present Geometric Deep Learning models for the orientation estimation task, which incorporate the geometry of the orientation space, the SO(3) manifold, into the deep learning pipeline by carefully choosing and designing representations, loss functions and network architectures well suited for this application. We first consider the problem of estimating the orientation of an object in an image assuming we are given the object category and a bounding box containing the object in the image. We show that modeling the orientation space correctly by designing Riemannian CNNs i.e. regression and classification CNNs that use axis-angle or quaternion representations of rotation matrices and geodesic loss functions, leads to good performance on a challenging benchmarking dataset. We also propose a family of Bin & Delta models that combine pose classification CNNs (bin model) to get a coarse estimate of the object orientation and pose regression CNNs (delta model) that refine the coarse orientation estimate. Such models achieve state-of-the-art performance in the benchmark dataset. Additionally, we have extended these models to the scenarios of unknown categorization and unknown localization by designing novel Integrated Networks to solve these multi-task problems.