Generation of 3D data by deep neural networks has been attracting increasing attention in the research community. The majority of extant works resort to regular representations such as volumetric grids or collections of images; however, these representations obscure the natural invariance of 3D shapes under geometric transformations, and also suffer from a number of other issues. In this paper we address the problem of 3D reconstruction from a single image, generating a straight-forward form of output – point cloud coordinates. Along with this problem arises a unique and interesting issue, that the groundtruth shape for an input image may be ambiguous. Driven by this unorthodox output form and the inherent ambiguity in groundtruth, we design architecture, loss function and learning paradigm that are novel and effective. Our ﬁnal solution is a conditional shape sampler, capable of predicting multiple plausible 3D point clouds from an input image. In experiments not only can our system outperform state-ofthe-art methods on single image based 3d reconstruction benchmarks; but it also shows strong performance for 3D shape completion and promising ability in making multiple plausible predictions.

We present Im2Pano3D, a convolutional neural network that generates a dense prediction of 3D structure and a probability distribution of semantic labels for a full 360◦ panoramic view of an indoor scene when given only a partial observation (≤ 50%) in the form of an RGB-D image. To make this possible, Im2Pano3D leverages strong contextual priors learned from large-scale synthetic and realworldindoorscenes. Toeasethepredictionof3Dstructure, we propose to parameterize 3D surfaces with their plane equations and train the model to predict these parameters directly. Toprovidemeaningfultrainingsupervision,weuse multiple loss functions that consider both pixel level accuracy and global context consistency. Experiments demonstrate that Im2Pano3D is able to predict the semantics and 3D structure of the unobserved scene with more than 56% pixel accuracy and less than 0.52m average distance error, which is signiﬁcantly better than alternative approaches.

里面

It is important and helpful to get a full 3D view in auto-drive scene. There have been many researches about 3D reconstruction these years in the field of computer vision and deep learning. In this work, we focus on the use of deep learning in reconstruction and especially emphasize the auto-drive scene, thus there’s needs to mimicking radar data, for what we use the KITTI Velodyne radar data as input and expect a point cloud output. The main task of this work is preidcting the corresponding point cloud representation with the given image of a autodrive scene. Experimental results showed the effectiveness of the approach.