Apple Furthers Point Cloud Feature Recognition

In a paper released on arXiv, Apple engineers Yin Zhou and Oncel Tuzel have detailed an end-to-end deep neural network solution for accurate feature recognition in point clouds.

One way we currently perform feature recognition is to project the point cloud into a perspective view and apply image-based feature recognition techniques. Another approach is to rasterise the point clouds into a 3D voxel grid, and encode the features by hand (think of a voxel as a pixel in three dimensions instead of two). The first approach loses us some valuable depth information, and both create an extra step. According to the researchers, this “informational bottleneck” keeps us from “effectively exploiting 3D shape information and the required invariances for the detection task.”

Apple’s solution is the VoxelNet which will scale up a deep neural network and process sparse point clouds and recognise features without requiring manual input. The trick is simultaneously learning how to discriminate features in a point cloud and predicting how to bound that feature in a 3D box. This allows the network to divide the point cloud into voxels, encode them using VFE layers, and then generate a volumetric representation. The deep learning network learns how to break features down into voxels, and then relates the right ones together to create 3D representations of each feature.

According to the researchers, the results are promising. VoxelNet “outperforms state-of-the-art liDAR based 3D detection methods by a large margin. On more challenging tasks, such as 3D detection of pedestrians and cyclists, VoxelNet also demonstrates encouraging results showing that it provides a better 3D representation.”

If the network can recognise cyclists in a sparse point cloud captured by a moving car, couldn’t it recognise features in a point cloud gathered by a cell phone, a handheld scanner, or a terrestrial scanner? There would certainly be a lot of uses for that.

Click here to read the paper.