PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud

CVPR 2019

Summary

In this paper PointRCNN is presented as a novel 2-stage 3D object detection architecture. Based on a foreground segmentation of the input points, accurate 3D proposals are generated. Bounding box refinement is then performed using canonical transformations and aggregated local and global features. Furthermore, the authors present a bin-based regression loss for highly accurate box center prediction.

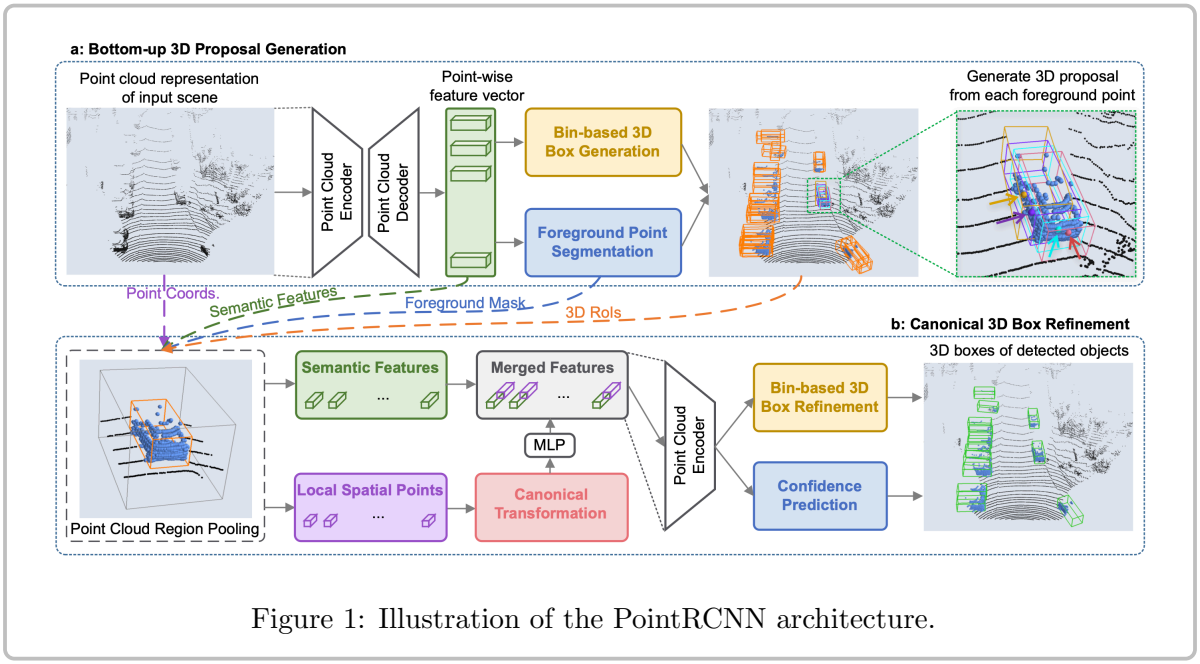

Figure 1: Illustration of the PointRCNN architecture.

Main Contributions

- **Bottom-up Proposal Generation** The authors introduce a novel proposal generation algorithm that computes accurate 3D proposals in a bottom-up manner by segmenting the point cloud into foreground and background points while simultaneously generating proposal bounding boxes. This approach is based on the observation that natural 3D scenes feature separated objects whose segmentation masks can easily be obtained by considering all points within the respective bounding boxes.
- **Bin-based Box Regression Loss** The quality of regression losses is considered highly dependent on the magnitude of the regressed values. In order to minimize the required regression values, a bin-based regression loss is proposed. The area around each point is turned into a grid based on a specified search range $\mathcal{S}$ and a bin size δ. The regression of the bounding box center is then treated as a classification problem and ordinary regression is only used for the offset within the predicted bin, keeping the regression ranges to a minimum.
- **SOTA on KITTI 3D Object Detection Benchmark** The authors report state-of-the-art performance on the KITTI 3D Object Detection Benchmark. Performance increases can be observed especially for the hardest category, showing an 8% increase in AP at 0.7 IoU compared to previous arts.

Implementation Details

- **Foreground Segmentation** One part of the proposal generation is the binary segmentation of the input point cloud. It is argued that the extraction of relevant features for point-wise segmentation benefits the quality of the generated bounding box proposals. The segmentation head is appended to a backbone feature extraction network, in this case PointNet++. The segmentation masks are provided by the bounding box annotations and the training is supervised using focal loss to account for severe class imbalance.

- **Bin-based Box Regression** The concept of bin-based box regression is displayed in Figure 2. By treating the center regression of the bounding box as a combination of classification and regression problem, the quality of the estimated box position can be improved compared to a pure regression loss. The idea of binning is also applied for the orientation of the bounding box, while the size of the bounding box is regressed directly w.r.t the average size of the object in the training set.

- **Canonical Box Refinement** For the refinement of the box proposal, points are pooled from each slightly enlarged proposal and transformed to the canonical coordinates of the respective box to allow for more accurate local features. The transformed local features are combined with the global semantic features from the proposal generation stage. These features are concatenated and augmented by the distance between point and LiDAR to account for the sparsity at large distances. The loss functions used for the bounding box refinement are similar to the bin-based regression loss used for the proposal generation.

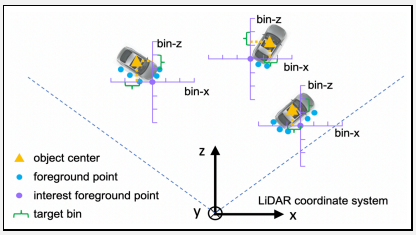

Figure 2: Illustration of the bin-based regression loss.

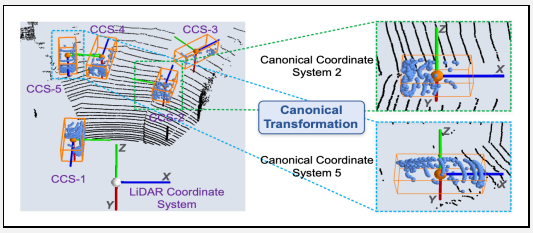

Figure 3: Illustration of the canonical box refinement.

Evaluation

- **3D Object Detection** PointRCNN achieves state-of-the-art on KITTI 3D Object Detection on the car class across all three categories. It outperforms several previous arts that leverage a combination of geometric and RGB information. On the pedestrian class, PointRCNN outperforms other methods that are purely based on point clouds. However, for small objects, the sparsity of the observation leads to less accurate detections compared to RGB-based methods.

- **Proposal Generation** The authors evaluate the quality of the proposals by computing the recall for a given number of proposals. For only 50 proposals, a recall of 96% is obtained at IoU threshold 0.5. At IoU threshold 0.7, a recall rate of 82.3% is obtained using 300 proposals.

References

This summary is solely based on my understanding of the original paper. All images used here are taken from the original paper as well. The paper can be found under the following link:
`https://arxiv.org/pdf/1812.04244.pdf`