PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation

CVPR 2017

Summary

This paper presents a novel network architecture called PointNet which is designed to extract local and global features from a raw point cloud. The architecture takes into account the permutation and transformation invariance of point sets and can be used as a feature extractor for a variety of downstream tasks such as 3D object classification, object part segmentation as well as object retrieval and object correspondence problems. PointNet further displays impressive robustness properties which stem from an internal representation based on the skeleton of the observed object.

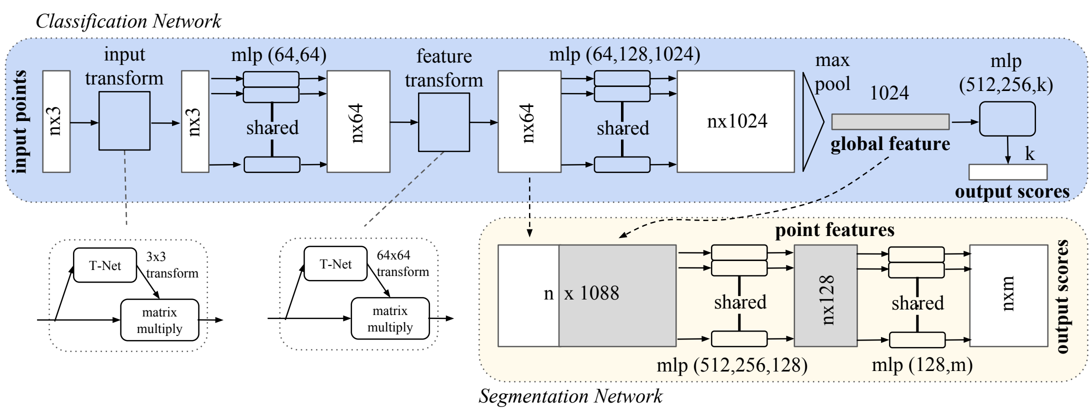

Figure 1: Illustration of the PointNet architecture. The model relies on input and feature transform modules, global and local feature aggregation and symmetric max-pooling operations.

Main Contributions

- **Novel Network Architecture** The proposed architecture processes point clouds based on the raw set representation. Consequently, the unordered nature of the point set, the relation between points in a metric space as well as the invariance under certain transformations have to be accounted for by the network. PointNet consists of three main modules: *Joint Alignment Networks* for both input points and point features, an aggregation structure for local and global features and the max-pooling function for symmetric information aggregation.

- **Theoretical Analysis** The authors provide a theoretical treatment of the robustness properties of the model w.r.t. input corruption and noise. The main insight is that PointNet essentially learns to represent a shape by a set of key points $\mathcal{C}_s$ that corresponds to the skeleton of the observed object.

- **Evaluation on Classification and Segmentation Tasks** PointNet is designed as a generic feature extractor for point sets. Consequently, it can be readily applied to a number of different downstream tasks. The authors evaluate PointNet on object classification, part segmentation as well as scene segmentation and further present possible extensions to different tasks.

Implementation Details

- **Permutation Invariance** Three different options are evaluated that ensure permutation invariance of the computations: 1) input sorting; 2) input as sequence; 3) symmetric information aggregation. Option 3) performs best, using a combination of an MLP and max-pooling to process the input points.

- **Joint Alignment Network** Semantic labelling of points should be invariant to rigid-body transformations. A mini-network (T-net) is used to predict an affine transformation matrix which is applied to the input points. The concept is also applied to feature space alignment. A special regularization component is introduced to keep the learned transformation matrix close to orthogonal.

- **Information Aggregation** For accurate segmentation, a global feature is concatenated with per-point features. Based on the concatenation, new context-aware local features are computed for each point.

Evaluation

- **Model Robustness** The robustness properties of PointNet are shown by evaluating performance on reduced and corrupted input sets. Furthermore, the critical point set $\mathcal{C}_s$ and the upper-bound shape $\mathcal{N}_s$ are introduced which define all sets for which the network computes the same global feature and thus act as robustness measures.

- **3D Object Part Segmentation** The model is evaluated on ShapeNet which contains 16,881 shapes from 16 categories and 50 annotated parts in total. The problem is formulated as per-point classification using the segmentation network displayed in Figure 1. Performance is evaluated using mIoU for each category. PointNet is compared to two traditional methods based on shape correspondences and a 3D-CNN baseline. PointNet outperforms the other methods by a large margin on most object categories.

- **Task Extension** As PointNet computes a global feature representation for each input set, it can be applied to a variety of other tasks. The authors apply PointNet to retrieval tasks based on NN-search in feature space as well as shape correspondence problems by comparing feature space activations of the critical point sets $\mathcal{C}_s$.

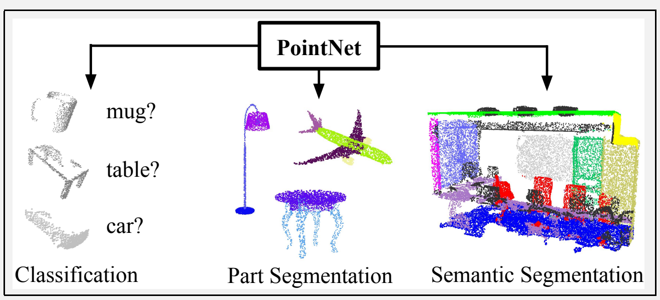

Figure 2: Illustration of applicability of PointNet to different tasks. PointNet acts as a generic feature extractor for local and global features which can be used for a variety of tasks.

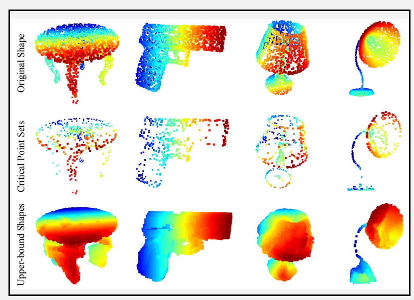

Figure 3: Illustration of critical point set $\mathcal{C}_s$ and upper-bound shape $\mathcal{N}_s$. $\mathcal{C}_s$ justifies statement that PointNet extracts a representation based on the skeleton of the observed object.

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/1612.00593.pdf`