

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

Date: 2022.07.13

Name: ChangMin An

Github: [Link](#)

I. Introduction

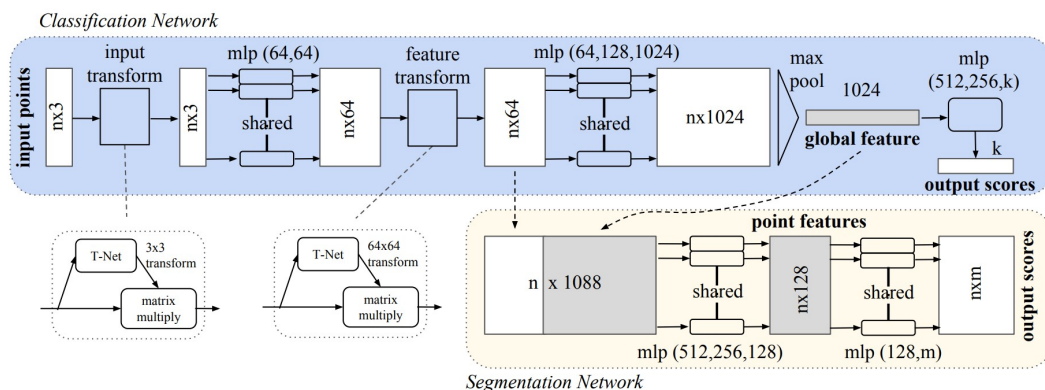
센서 스캔 결과로 얻어지는 3차원 point cloud를 인식하기 위해, 많은 연구자들은 point cloud를 Voxel 형식으로 변환한다. 복셀은 요약하기 좋으나, 빈 공간이 많이 발생하여 비효율적인 측면이 있다. PointNet은 입력 시 포인트 순열의 불변성을 이용한다. PointNet을 통해, Classification, Segmentation, 의미 해석 등에 필요한 아키텍처를 제공한다.

II. Properties of Point sets

PCD(point cloud data)는 $\{P_i \mid i = 1, \dots, n\}$, $P = (x, y, z)$ 로 정의한다. PointNet은 k 개의 output을 가지도록 class를 정의하며, 이 모델을 통해 $n \times m$ scores output을 가질 수 있다. 이 때, n : PCD m : semantic subcategory이다. PointNet의 input은 Euclidean space에서 포인트들의 subset이다.

- Unordered: 점들 사이에 순서가 없다.
- Interaction among points: 점들간 상호영향을 받는다.
- Invariance under transformations: 변환에 불변이어야 한다.

III. PointNet Architecture



Architecture에는 3가지 주요 모듈이 있다. **max pooling layer**는 모든 점들로부터 정보를 모으기 위한 symmetric 함수이며, **지역 및 전역 정보 조합 구조**, 입력 점군과 점 특징들을 정렬하는 2개의 **joint alignment network**로 구성된다.

1. Symmetry Function for Unordered Input

정렬은 2차원에서 좋은 solution이지만 높은 차원(3D)의 자료 정렬은 존재하지 않는다. 예를 들어, 고차원 공간의 점들을 1차원 실수 선으로 projection한 후 정렬할 수 있으나, 이에 대한 역변환으로 원 데이터를 복구할 수 없다. 이를 해결하기 위한 PointNet의 아이디어는 변환된 요소에 대한 **symmetric function**(max-pooling)을 적용한 점군을 정의하는 것이다. 이는 순서에 상관없이 결과가 일정하게 나오기 위함이다.

$$f(\{x_1, \dots, x_n\}) = g(h(x_1), \dots, h(x_n))$$

이 때 g 가 max pooling을 해주는 symmetric function이다.

2. A Local and Global Information Combination Structure

벡터 f_1, \dots, f_k 형태의 출력은 입력 집합에 대한 global information이다. SVM이나 MLP를 이용해 형상의 전역 특징을 학습하는 것은 쉽지만, local 및 global information을 구분해 얻는 것이 필요하다. **전역 특징이 계산된 후, 각 점들의 특징과 전역 특징을 연결하여 포인트 특징을 얻는다.**

3. Joint Alignment Network

점군의 labeling은 형상 변환(translation, rotation)에 대해 불변이어야 한다. 이를 위해 mini-network(T-net)을 정의하고 적용한다. 이 때 T-net에서 transformation된 matrix를 추정하여 사용한다.