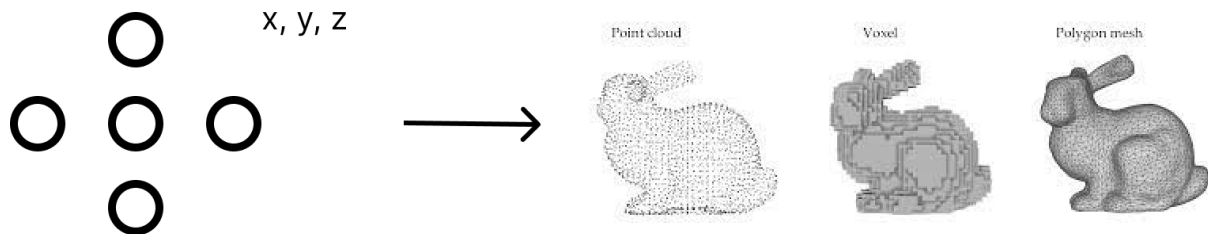


Giải thích về PointCloud và PointNet

Đăng Anh Quân – 20203909

Introduction:

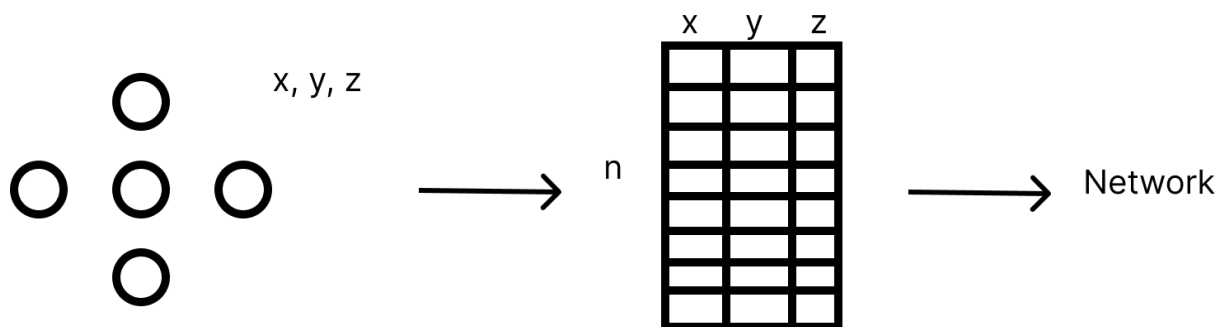
- Pointnet là phương pháp đầu tiên sử dụng tạo ra mạng neural xử lý trực tiếp điểm 3D mà không qua cách biểu diễn trung gian (điều mà không hiệu quả).



Point Clouds

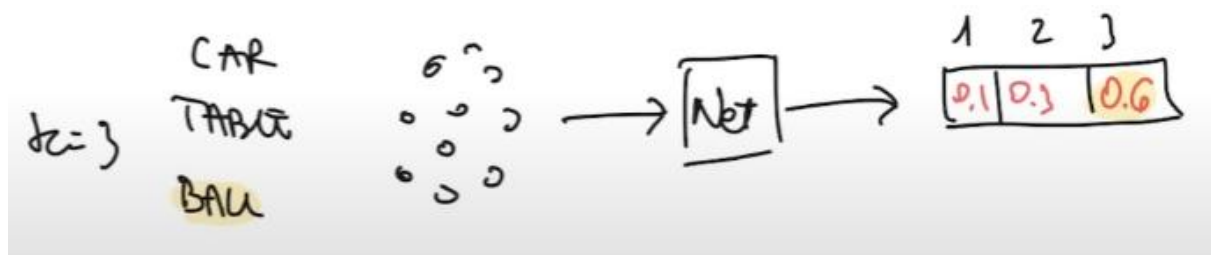
3D meshes, voxels, ...

- Đầu vào của network sẽ là n điểm Point Clouds thực tế là một ma trận 2 chiều gồm $n \times 3$ tương ứng n điểm và tọa độ x, y, z của chúng.

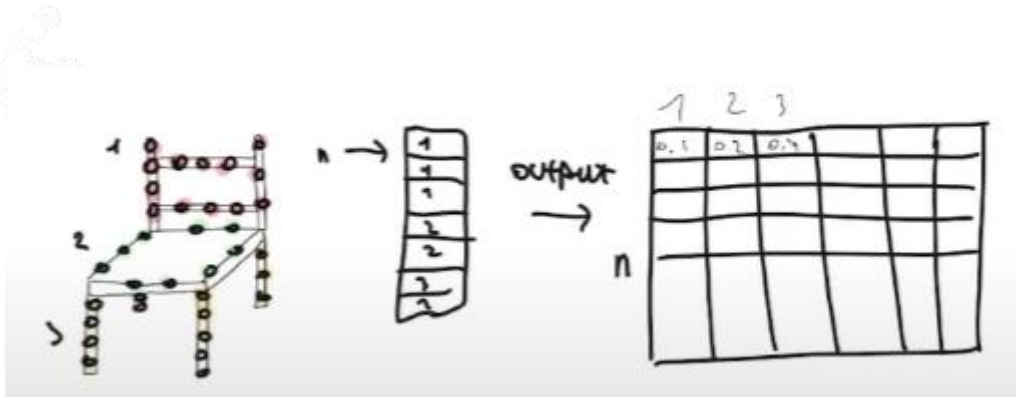


Problem Statement:

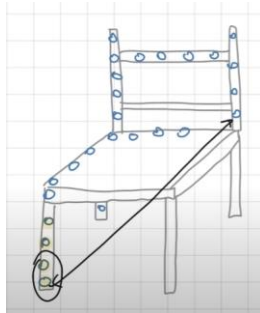
- Đối với tập điểm Point Clouds đầu vào, chúng ta muốn tìm được 2 networks gồm:
 - o Classification network: lấy n điểm Point Clouds đầu vào và đầu ra là số điểm K tương ứng với K lớp phân loại.



- o Semantic segmentation:



- 3 đặc điểm chính của Point Clouds:
 - Không có thứ tự: Không giống các cách biểu diễn khác, các điểm Point Cloud có tọa độ riêng nên không quan trọng về thứ tự khi lưu trữ cũng như làm đầu vào cho network. Do đó, khi kể cả lưu trữ các điểm Point Clouds với thứ tự khác nhau thì đầu ra sẽ cho cùng 1 kết quả.
 - Các điểm tương tác lẫn nhau: Mỗi quan hệ giữa hai điểm ở gần nhau (khoảng cách gần) khác với mỗi quan hệ giữa hai điểm ở khoảng cách xa.



- Bất biến trông qua các biến đổi (invariance under transformations): Dù vật thể có bị xoay, lật, ... hay các phép biến đổi khác thì kết quả sau khi cho vào network sẽ vẫn như nhau.

Overview:

- Chúng ta có 3 networks khác nhau:
 - Classification network: sẽ phân loại cho chúng ta biết đây là vật thể gì
 - Segmentation network: sẽ nói cho ta biết điểm Point Clouds thuộc bộ phận nào.
 - T-Net: có nhiệm vụ tìm ma trận phù hợp để nhân với ma trận đầu vào để điều chỉnh đầu vào (realign) đưa vào không gian nhỏ nhất.

Break down:

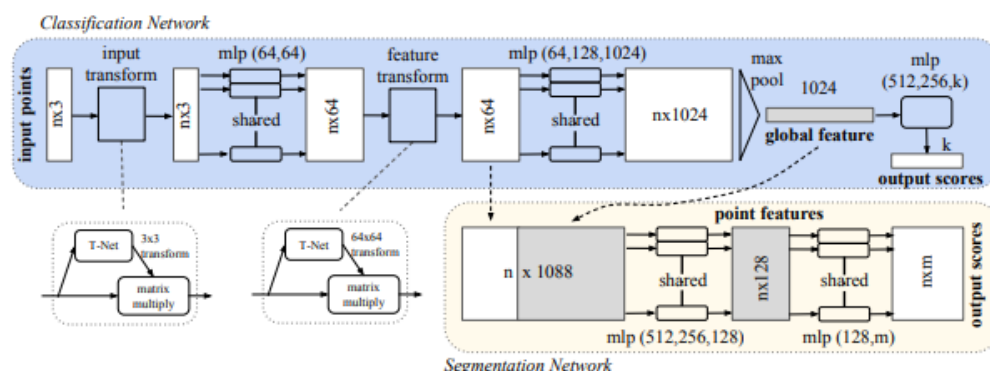


Figure 2. **PointNet Architecture.** The classification network takes n points as input, applies input and feature transformations, and then aggregates point features by max pooling. The output is classification scores for k classes. The segmentation network is an extension to the classification net. It concatenates global and local features and outputs per point scores. "mlp" stands for multi-layer perceptron, numbers in bracket are layer sizes. Batchnorm is used for all layers with ReLU. Dropout layers are used for the last mlp in classification net.

- Đầu vào, ta có n điểm Point Clouds tương ứng ma trận $n \times 3$ với n điểm và 3 tọa độ x, y, z của chúng.
- Input transform (biến đổi đầu vào) là một ma trận 3×3 dùng để xoay hay di chuyển các điểm về tọa độ phù hợp để không gian vào là bé nhất, các đặc trưng của vật thể sẽ không bị mất đi.
- Multi layer perceptron 1: mở rộng số chiều của feature vector từ 3 thông qua filter sẽ trở thành 64 chiều. Kết quả ta có feature map với số chiều là 64 đối với mỗi điểm đầu vào. Từ đó, có ma trận feature mới là $n \times 64$
- Feature transform (biến đổi ma trận đặc trưng) là một ma trận 64×64 dùng để xoay hay di chuyển các điểm về tọa độ phù hợp để không gian vào là bé nhất, các đặc trưng của vật thể sẽ không bị mất đi.
- Multi layer perceptron 2: mở rộng số chiều của feature vector từ 64 thông qua filter sẽ trở thành 1024 chiều. Kết quả ta có feature map với số chiều là 1024 đối với mỗi điểm đầu vào. Từ đó, có ma trận feature mới là $n \times 1024$
- Max pool operation: lấy giá trị max của từng channel đặc trưng của tất cả các điểm thành 1 số duy nhất do đó output sẽ là 1 vector 1024 chiều chính là global feature
- Multi layer perceptron 3: sử dụng fully connected layer với đầu vào là 1024 điểm đặc trưng và đầu ra là một ma trận với k chiều (k là số lớp) thể hiện output score của mô hình.
- Segmentation network:
 - o Concatenate: kết nối feature matrix 64×64 với global feature 1024 thành ma trận $n \times 1088$
 - o Multi layer perceptron 4: đổi từ ma trận 1088 sang ma trận với số chiều là $n \times 128$
 - o Multi layer perceptron 5: đổi từ ma trận 128 sang ma trận với số chiều là $n \times m$ (m là số bộ phận được xác định trước) – output score

