PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space

Vo Tuan Kiet

May 22, 2024

Point Cloud

- Simplest, only points, no connectivity
- (x,y,z) coordinates, with some features
- Easy to get directly from existing devices

Previous Work

Point cloud is converted to other representations before it's fed to a deep neural network

Conversion	Deep Net
Voxelization	3D CNN
Projection/Rendering	2D CNN
Feature extraction	Fully Connected

Question

Can we archive effective feature learning directly on point clouds ?

Challenges

- Unordered set as input : Model need to be invariant to N! permutations.

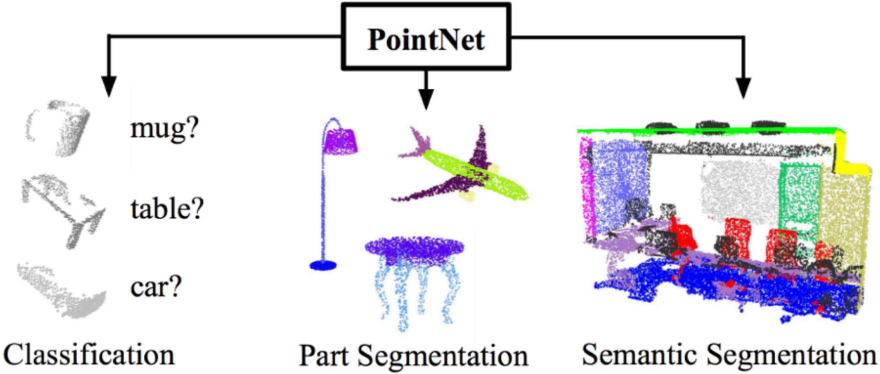
- Interaction among points : Model needs to be able to capture local structures from nearby points, and the combinatorial interactions among local structures.
- Invariance under geometric transformation : Point cloud rotations should not alter classification results.

Proposed Structure

- Unordered set as input $\rightarrow$ Max pooling
- Interaction among points $\rightarrow$ Global-local feature concatenation
- Invariance under geometric transformation $\rightarrow$ T-net

PointNet

End to end learning for scatterd, unordered point data. Unified framework for various tasks



Permutation Invariance : Symetric Function

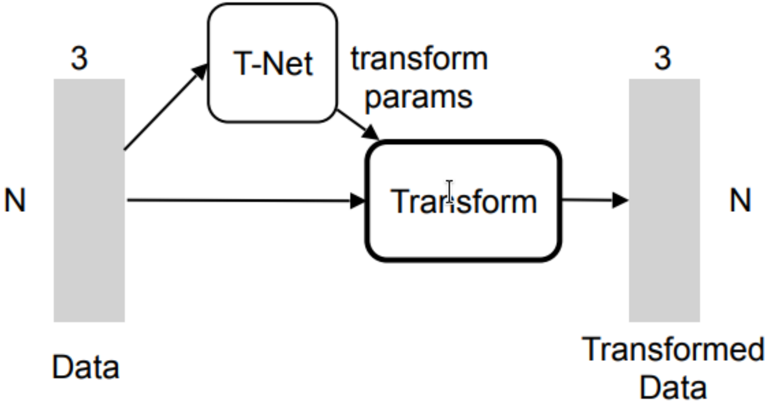
- Fundamental idea : approximate a general functions defined on a point set by appling a symetric function on transformed elements in the set :

$$f(\{x_1, \ldots, x_n\}) = \gamma\left(\max_{i=1,\ldots,n} h(x_i)\right)$$

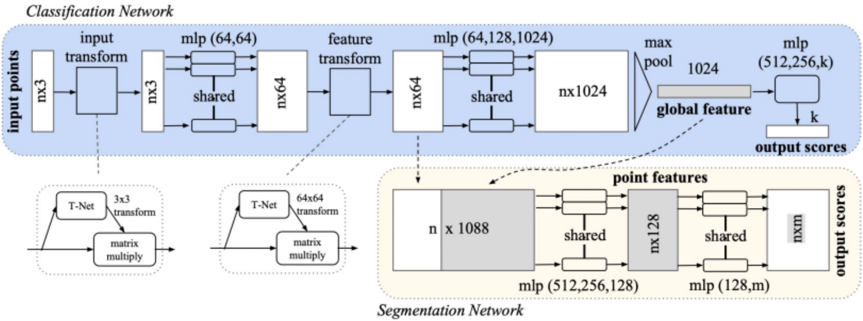
- Symetric function :
$f(x_1, x_2, \ldots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}), \quad x_i \in \mathbb{R}^d$

Input Alignment by Transformer Network

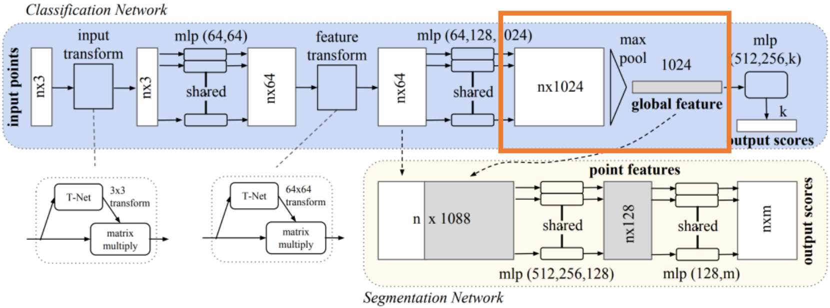
Idea : Data dependent transformation for automatic alignment

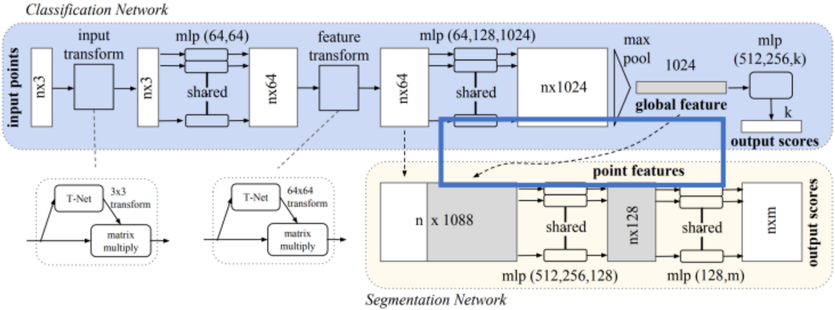


PointNet Architecture

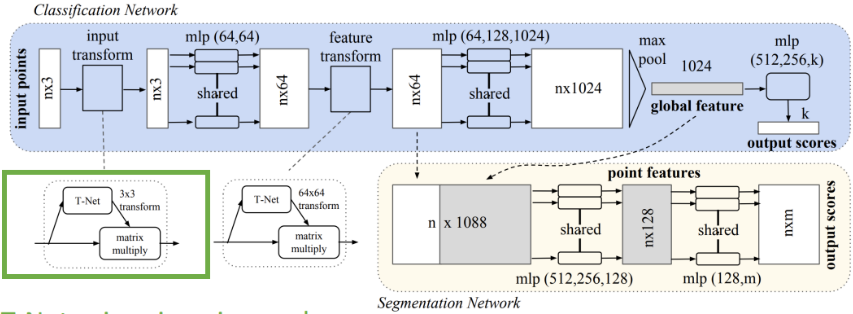


Max pooling gives order invariance





Concatenating global-to-local features combines local + global info



T-Nets gives invariances by transforming to canonical pose

Limitations of PointNet

- Does not take into account density variability
- Does not capture both local and global features

Desired Properties

- Hierarchical feature learning
- Robustness to density changes
- Point feature propagation for set segmentation

Architecture choices

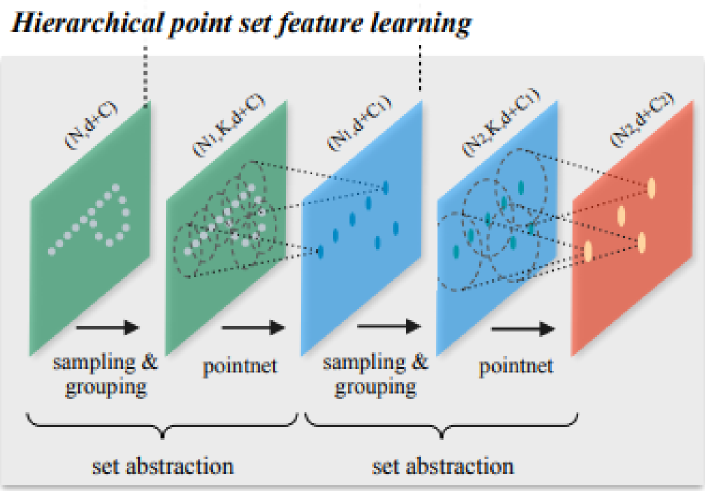
- Hierarchical feature learning → Set abstraction layer
- Robustness to density changes → Density-adaptive grouping
- Point feature propagation for set segmentation → Distance-based interpolation

Hierarchical Feature Learning

Applying PointNet recursively on a nested partitioning of the point set
Set abstraction layer:

1. Sampling layer : Select centroid with iterative farthest point sampling (FPS)
2. Grouping layer : Select group of point for each neighborhood centroid (KNN, ball queries)
3. PointNet layer : Apply small PointNet on each group

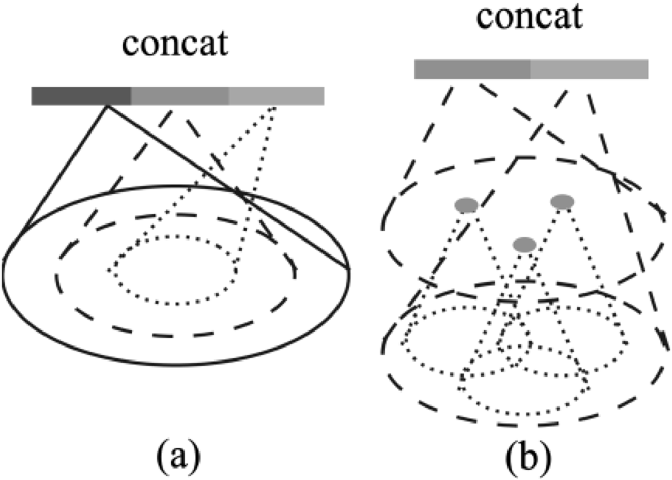
Hierarchical Feature Learning



Robustness to density changes

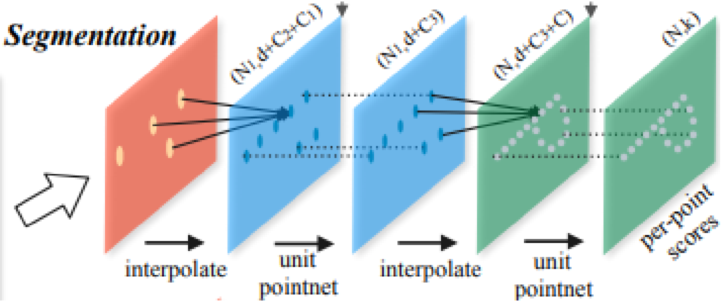
Adaptively group by density

- In high density area, group tightly
- In low density area, group widely



Propagation For Set Segmentation

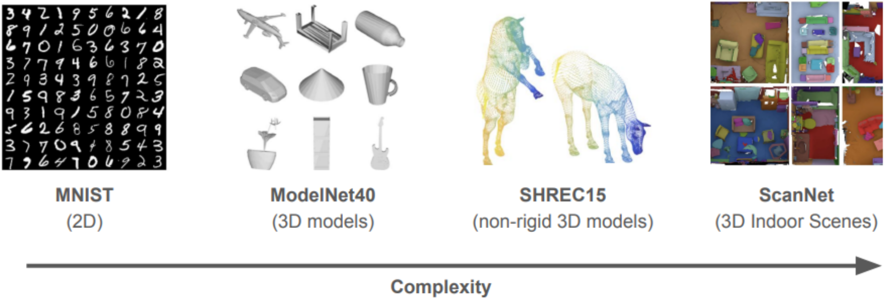
Propagate features from subsampled points to the original points.



Inverse distance weight interpolation

$$f^{(j)}(x) = \frac{\sum_{i=1}^{k} w_i(x) f_i^{(j)}}{\sum_{i=1}^{k} w_i(x)} \quad \text{where} \quad w_i(x) = \frac{1}{d(x, x_i)^p}, \quad j = 1, \ldots, C$$

PointNet++ was evaluated on four datasets in various domains:



Result : MNIST classification

Method	Error rate (%)
Multi-layer perceptron [24]	1.60
LeNet5 [11]	0.80
Network in Network [13]	**0.47**
PointNet (vanilla) [20]	1.30
PointNet [20]	0.78
Ours	0.51

Table 1: MNIST digit classification.

Result : ModelNet40 classification

Method	Input	Accuracy (%)
Subvolume [21]	vox	89.2
MVCNN [26]	img	90.1
PointNet (vanilla) [20]	pc	87.2
PointNet [20]	pc	89.2
Ours	pc	90.7
Ours (with normal)	pc	**91.9**

Table 2: ModelNet40 shape classification.

Result : Robustness to density changes

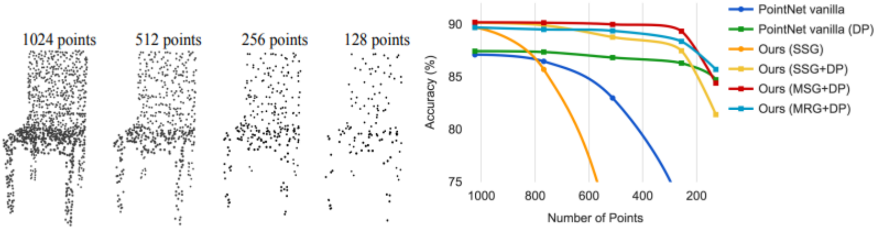


Figure: Left: Point cloud with random point dropout. Right: Curve showing advantage of our density adaptive strategy in dealing with non-uniform density. DP means random input dropout during training; otherwise training is on uniformly dense points

Result : ScanNet40 labeling

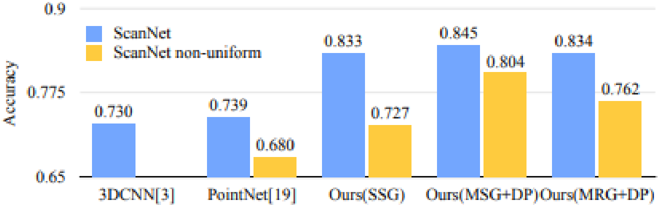
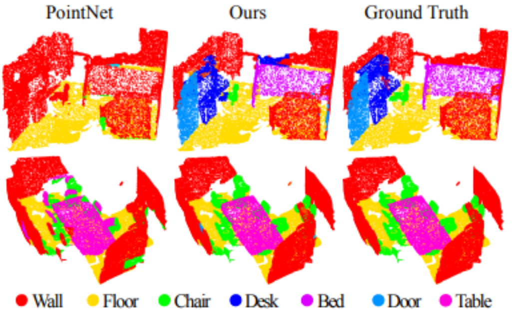


Figure 5: Scannet labeling accuracy.



Result : SHREC15 Non-rigid shape classification

	Metric space	Input feature	Accuracy (%)
DeepGM [14]	-	Intrinsic features	93.03
Ours	Euclidean	XYZ	60.18
	Euclidean	Intrinsic features	94.49
	Non-Euclidean	Intrinsic features	**96.09**

Table 3: SHREC15 Non-rigid shape classification.