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

Charles R. Qi, Li Yi, Hao Su and Leonidas J. Guibas
Stanford University

Project Website

Fundamental Question: *How to learn deep geometric representations from point sets in a metric space?*

Applications: robot perception, augmented reality, industrial design, molecular biology etc.

Motivation & Background

Image: a regular **array** of pixels

Point Cloud: an unordered **set** of points

In most previous works, point clouds are converted to other data representations before they are fed into deep networks:

Conversion	Deep Net	
Voxelization	3D CNN	quantization, high computation cost
Projection/Rendering	2D CNN	loss of 3D geometry
Feature extraction	Fully Connected	not end-to-end optimized for tasks

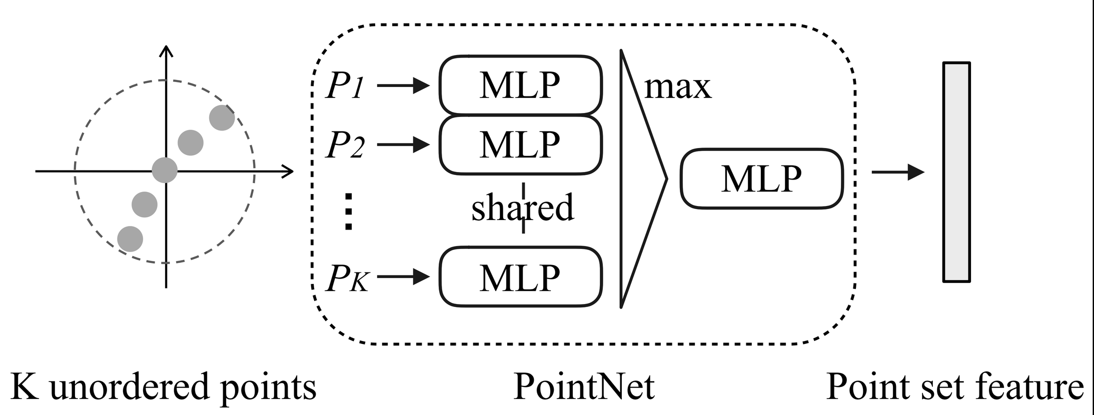
This work: Deep hierarchical feature learning on raw point clouds. Our proposed architecture, which is built on top of PointNet, is called **PointNet++**

PointNet Review

While the previous work PointNet by Qi et al. also consumes raw point clouds, **it's limited in capturing interactions among points** — it only learns either global or single-point features, thus lags behind in generalizability to large-scale scenes.

Basic PointNet Architecture:

$$f(x_1, x_2, ..., x_n) = \gamma\left(\underset{i=1,...,n}{\text{MAX}} \{h(x_i)\}\right)$$

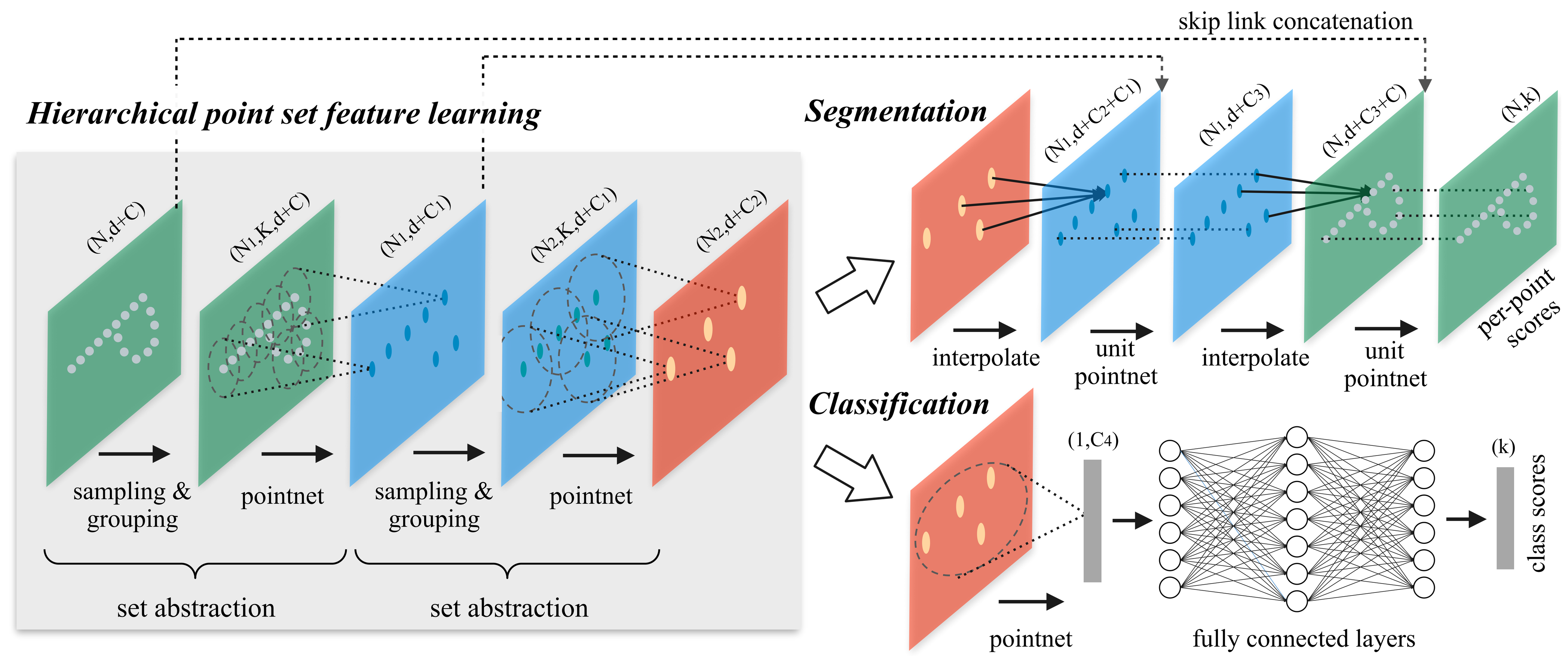


Referecne: C. R. Qi, H. Su, K. Mo, and L. J. Guibas. *PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation.* CVPR 2017.

PointNet++ Network Architecture

1. Hierarchical Point Set Feature Learning

We first partition the set of points into overlapping local regions by the distance metric of the underlying space (region centers chosen by furthest point sampling — FPS). Similar to CNNs, we extract local features capturing fine geometric structures from small neighborhoods; such local features are further grouped into larger units and processed to produce higher level features.



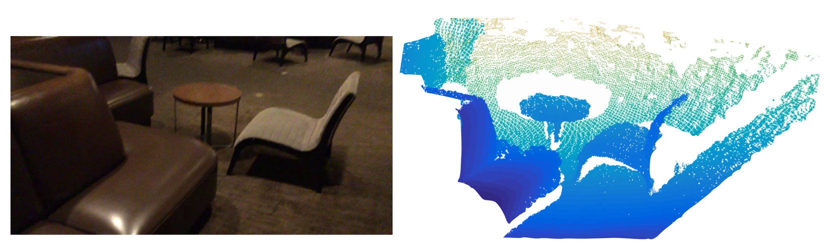
2. Feature Propagation for Set Segmentation

We adopt a hierarchical propagation strategy with 3D interpolation based on metric space distance and skip links.

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

3. Robust Features for Non-Uniform Sampling Density

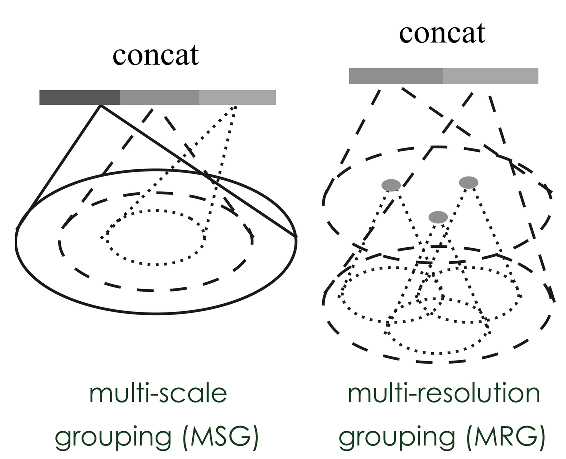
It is common that a point set comes with non-uniform density in different regions. Such non-uniformity introduces a significant challenge for point set feature learning. Features learned in dense data may not generalize to sparsely sampled regions and vise versa.



Visualization of a scan captured from a Structure Sensor (left: RGB; right: point cloud).

In PointNet++, each abstraction level extracts multiple scales of local patterns and combine them intelligently according to local point densities. Two types of robust layers are shown on the right:

We train the network to learn an optimized strategy to combine multi-scale features by randomly dropping out input points with a randomized probability for each instance.



Application Results

MNIST Digit 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 results. Positive pixels are converted to 2D point cloud (x,y) to feed to PointNet[++].

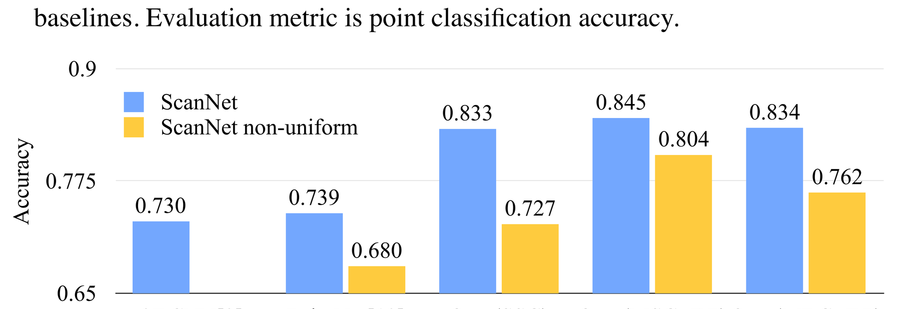
3D Shape Classification

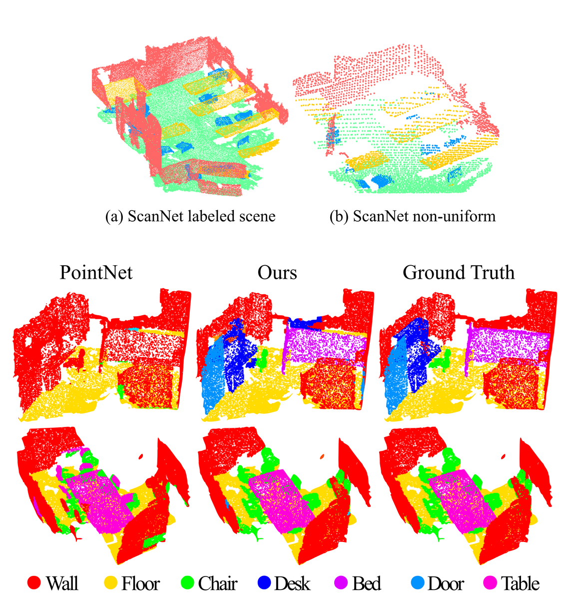
Method	Input	Accuracy (%)
Subvolume [21]	vox	89.2
MVCNN [26]	img	90.1
PointNet (vanilla) [20]	pc	87.2
PointNet [20]	pc	89.2
Ours	pc	90.7
Ours (with normal)	pc	**91.9**

Table 2: 3D shape classification results on ModelNet40.

3D Scene Semantic Segmentation

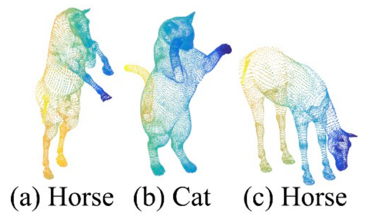
With hierarchical features and robust layers (MSG and MRG) for non-uniform density, our PointNet++ models significantly outperform 3D CNN and PointNet baselines. Evaluation metric is point classification accuracy.





Point Set Classification in Non-Euclidean Metric Space

Although (a) and (b) are similar in pose but they are different in class. For non-rigid shape, we extends our models to geodesic distance space with intrinsic features.



	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: Non-rigid shape classification results on SHREC15.

Analysis Experiments

Robust Layers for Non-Uniform Density

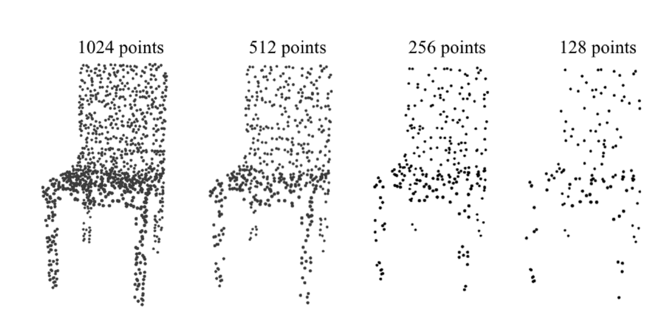
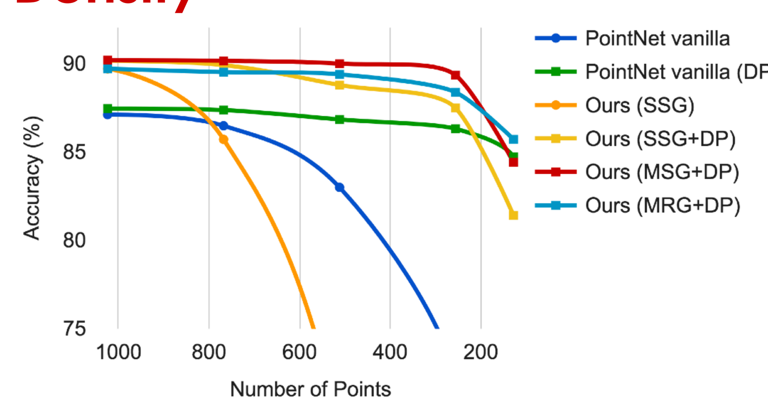
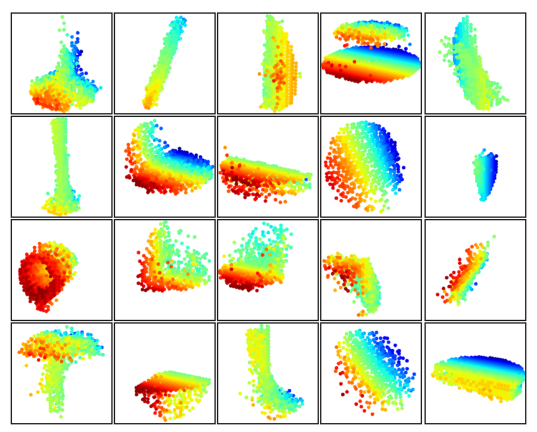


Illustration of point clouds with varying sampling densities.



Learned Feature Visualization



Twenty representative point set patterns learned by the first-level neurons in PointNet++.

More Experiments

Feature difference std.	Accuracy std.
0.021	0.0017

Table 4: Effects of randomness in FPS (3D shape classification).

kNN (k=16)	kNN (k=64)	radius (r=0.1)	radius (r=0.2)
89.3	90.3	89.1	90.7

Table 5: Effects of neighborhood choices. Evaluation metric is classification accuracy for 3D shape classification.

	PointNet (vanilla) 1	PointNet 1	Ours (SSG)	Ours (MSG)	Ours (MRG)
Model size (MB)	9.4	40	8.7	12	24
Forward time (ms)	11.6	25.3	82.4	163.2	87.0

Table 6: Runtime and size of PointNet and PointNet++ models.

Conclusion and Future Works

In this work, we propose PointNet++, a powerful neural network architecture for processing point sets sampled in a metric space. PointNet++ learns hierarchical point cloud features and is able to adapt to non-uniform sampling densities in local regions. These contributions enable us to achieve state-of-the-art performance on challenging benchmarks of 3D point clouds.

In the future, it's worthwhile thinking how to accelerate inference speed of our proposed network as well as to find applications in higher dimensional metric spaces where CNN based method would be computationally unfeasible while our method can scale well.