universität innsbruck

Point Clouds and Deep Learning

Martin Berger

Warm up

Warm up

Can you tell me a joke about 3D point clouds and AI?

Warm up

Can you tell me a joke about 3D point clouds and AI?

Why did the AI refuse to analyze the 3D point cloud?

Warm up

Can you tell me a joke about 3D point clouds and AI?

Why did the AI refuse to analyze the 3D point cloud?
Because it heard the data points were just "plane" boring!

Warm up

Warm up

That’s not really funny. Give me another one.

Warm up

That’s not really funny. Give me another one.

Why did the AI become a master at processing 3D point clouds?

Warm up

That’s not really funny. Give me another one.

Why did the AI become a master at processing 3D point clouds?
Because it knew how to "point" out all the right features!

Overview

- Why point clouds?
- PointNet
- FoldingNet
- Cluster-Based Autoencoders for Volumetric Point Clouds

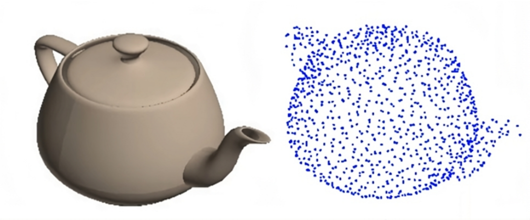
Point Clouds

What are 3D point clouds?

img source: https://shorturl.at/oHOZ8

Why though?

- Simplest representation: only points, no connections
- Collection of (x, y, z) coordinates



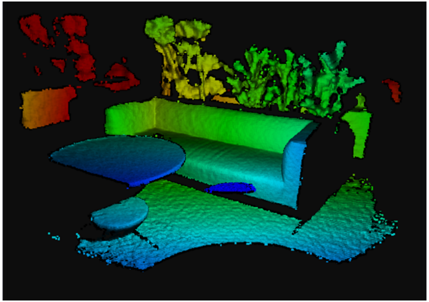
Why though?

- Laser (LIDAR, StreetView)

source: https://shorturl.at/hjnzK

Why though?

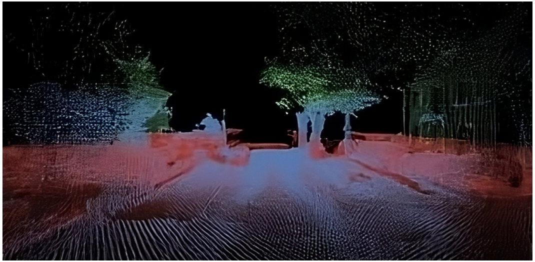
- Laser (LIDAR, StreetView)
- Infrared (Kinect)



source: https://shorturl.at/owEX9

Why though?

- Laser (LIDAR, StreetView)
- Infrared (Kinect)
- Stereo Cameras



source: https://shorturl.at/nvFP5

Challenges from a Data Science POV

- A point cloud is a set of sparse 3D points
- Irregular format is hard to handle

Challenges from a Data Science POV

- A point cloud is a set of sparse 3D points
- Irregular format is hard to handle

First attempts:

- Transform data to regular 3D voxel grids (Volumetric CNNs)

Challenges from a Data Science POV

- A point cloud is a set of sparse 3D points
- Irregular format is hard to handle

First attempts:

- Transform data to regular 3D voxel grids (Volumetric CNNs)
- Turn point clouds into a collection of images (Multiview CNNs)

PointNet

https://arxiv.org/abs/1612.00593

Observation

Observation

- Points are unordered

Observation

- Points are unordered
  $\rightarrow$ **invariance under permutations**

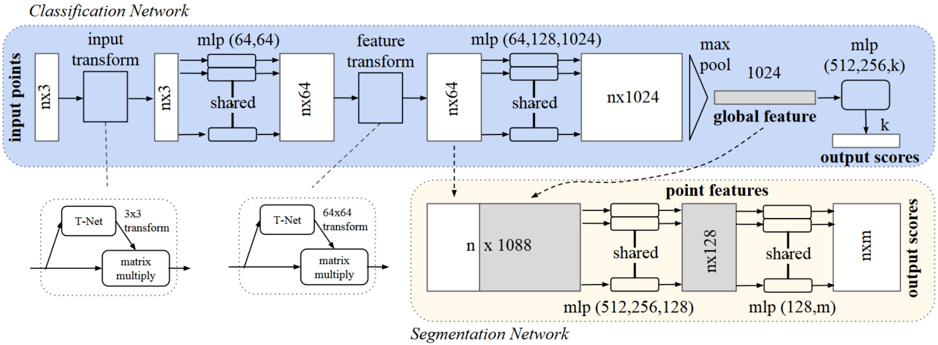
Observation

- Points are unordered
  $\rightarrow$ **invariance under permutations**
- Point clouds don't change after rotations/translations

Observation

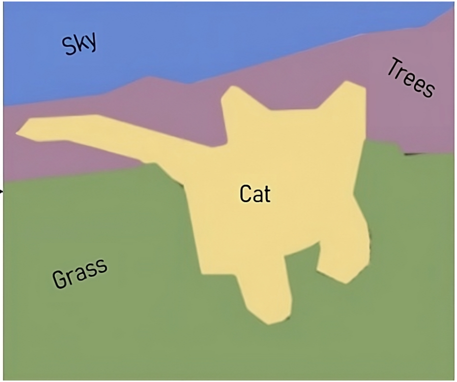
- Points are unordered
  $\rightarrow$ **invariance under permutations**
- Point clouds don't change after rotations/translations
  $\rightarrow$ **invariance under transformations**

PointNet



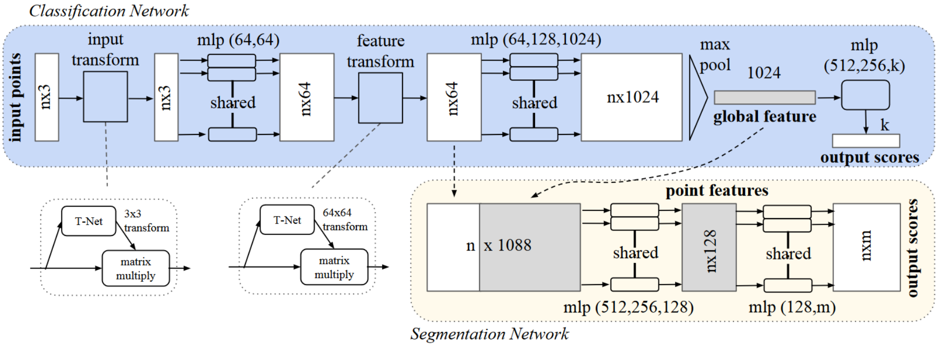
Classification Network
input points
nx3
input transform
nx3
mlp (64,64)
shared
nx64
feature transform
nx64
mlp (64,128,1024)
shared
nx1024
max pool
1024
global feature
mlp (512,256,k)
k
output scores
T-Net
3x3 transform
matrix multiply
T-Net
64x64 transform
matrix multiply
point features
n x 1088
shared
nx128
shared
nxm
output scores
mlp (512,256,128)
mlp (128,m)
Segmentation Network

Semantic Segmentation



source: https://shorturl.at/bvBQY

PointNet

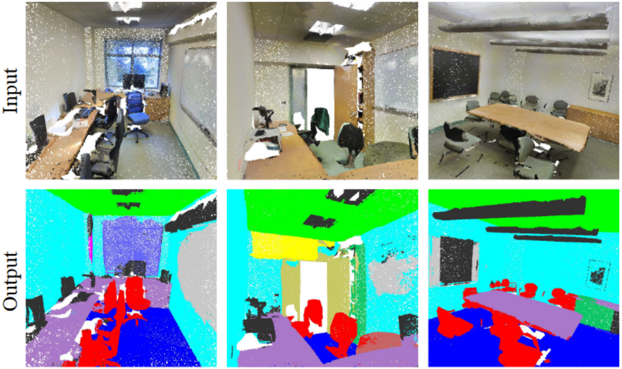

Classification Network
input points
nx3
input transform
nx3
mlp (64,64)
shared
nx64
feature transform
nx64
mlp (64,128,1024)
shared
nx1024
max pool
1024
global feature
mlp (512,256,k)
k
output scores
T-Net
3x3 transform
matrix multiply
T-Net
64x64 transform
matrix multiply
point features
n x 1088
shared
nx128
shared
nxm
output scores
mlp (512,256,128)
mlp (128,m)
Segmentation Network

PointNet - Results Classification

	input	#views	accuracy avg. class	accuracy overall
SPH [11]	mesh	-	68.2	-
3DShapeNets [28]	volume	1	77.3	84.7
VoxNet [17]	volume	12	83.0	85.9
Subvolume [18]	volume	20	86.0	**89.2**
LFD [28]	image	10	75.5	-
MVCNN [23]	image	80	**90.1**	-
Ours baseline	point	-	72.6	77.4
Ours PointNet	point	1	86.2	**89.2**

PointNet - Results Segmentation

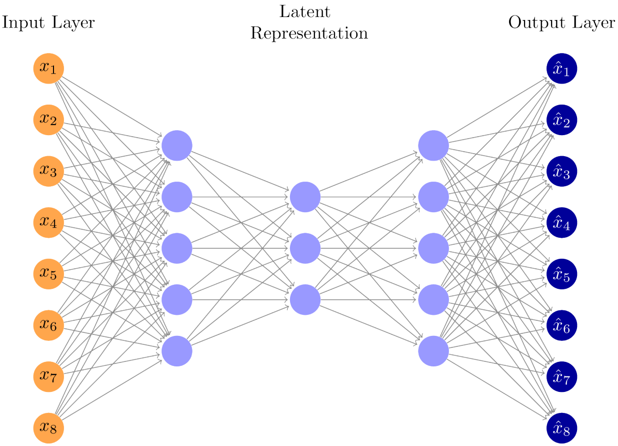
PointNet - Results Semantic Segmentation



FoldingNet

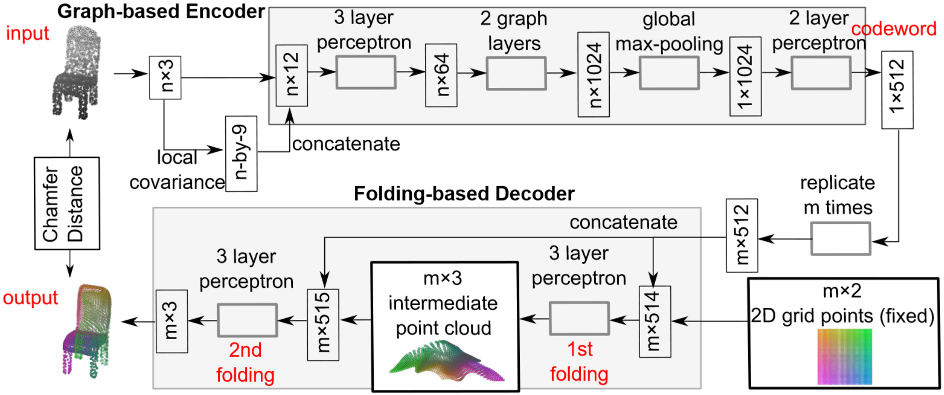
https://arxiv.org/abs/1712.07262

FoldingNet - Autoencoder for Point Clouds



source: https://shorturl.at/bhDP8

FoldingNet - Autoencoder for Point Clouds



FoldingNet - Local Covariance

FoldingNet - Local Covariance

- Let P be a point in the given point cloud.

FoldingNet - Local Covariance

- Let P be a point in the given point cloud.
- Let $Q_1, \ldots, Q_K$ denote the K nearest neighbors of P.

FoldingNet - Local Covariance

- Let P be a point in the given point cloud.
- Let $Q_1, \ldots, Q_K$ denote the K nearest neighbors of P.
- Compute the centroid C of the K points and set $\tilde{Q}_i = Q_i - C$

FoldingNet - Local Covariance

- Let P be a point in the given point cloud.
- Let $Q_1, \ldots, Q_K$ denote the K nearest neighbors of P.
- Compute the centroid C of the K points and set $\tilde{Q}_i = Q_i - C$
- Then

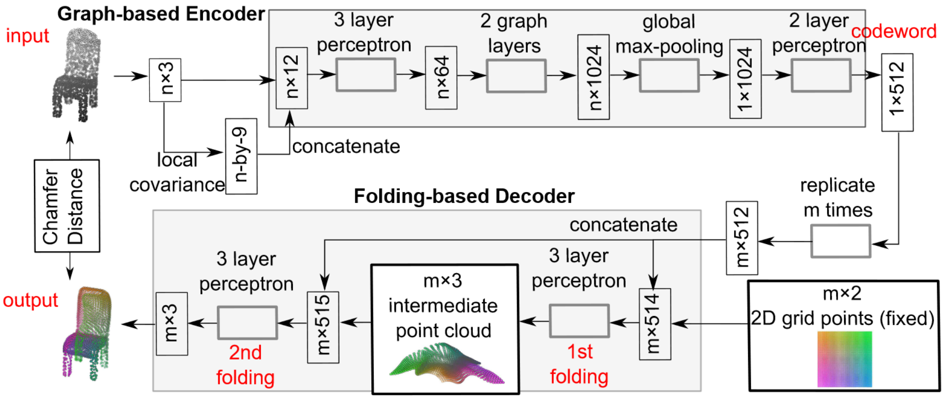
$$A = \begin{bmatrix} \tilde{Q}_1 \\ \vdots \\ \tilde{Q}_K \end{bmatrix} \in \mathbb{R}^{K \times 3}$$

and the covariance matrix is given by

$$A^T A = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix} \in \mathbb{R}^{3 \times 3}$$

.

FoldingNet - Autoencoder for Point Clouds

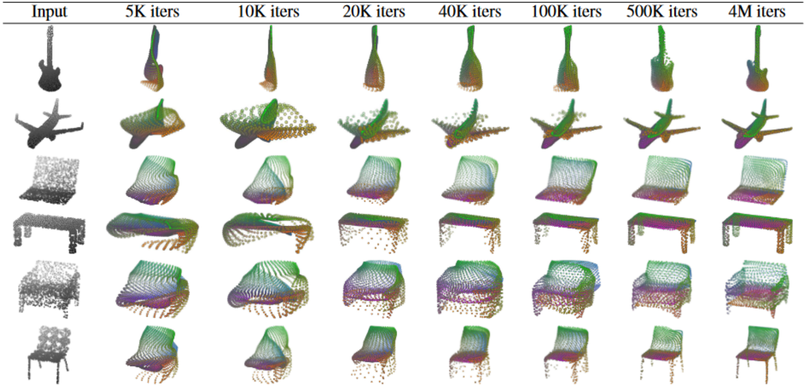


FoldingNet - Loss

Given two point clouds $\mathcal{X}, \mathcal{Y}$ their Chamfer distance is

$$d_{CH}(\mathcal{X}, \mathcal{Y}) = \max \left\{ \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} \|x - y\|_2, \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \|x - y\|_2 \right\}.$$

FoldingNet - Results



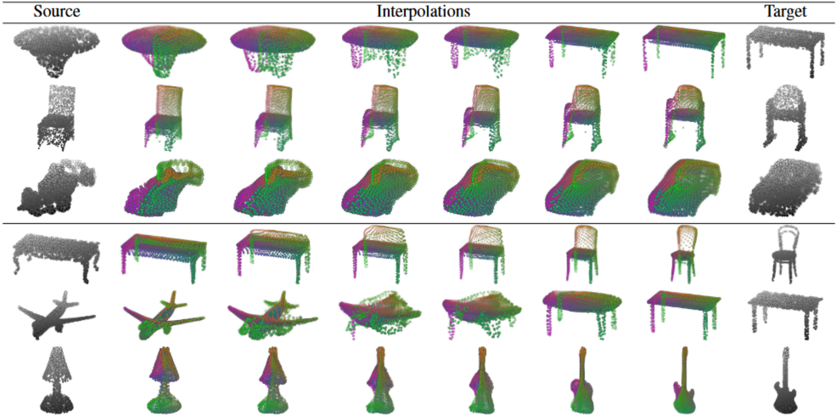
Blending between Datapoints

Given two point clouds P_1, P_2, encode them by codewords C_1, C_2 within the feature space then decode

$$\lambda C_1 + (1 - \lambda) C_2$$

for some $\lambda \in [0, 1]$ with the decoder.

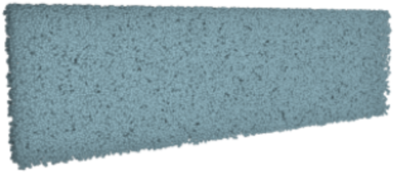
FoldingNet - Results

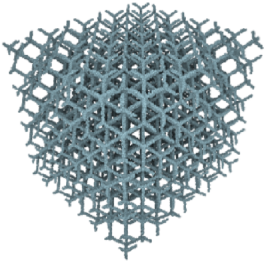


Cluster-Based Autoencoders for Volumetric Point Clouds

https://arxiv.org/abs/2211.01009

Density based Style Transfer

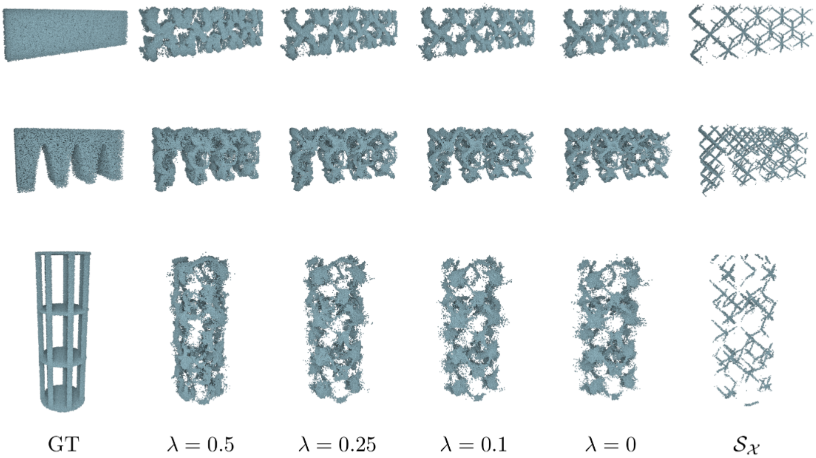




Density based Style Transfer

- Given a point cloud $\mathcal{X}$ and a style point cloud $\mathcal{S}$.
- Assume $\mathcal{X}$ was sampled according to the density $f_{\mathcal{X}}$.
- Sample a point cloud $\mathcal{S}_{\mathcal{X}}$ from $\mathcal{S}$ according to $f_{\mathcal{X}}$.
- Use the autoencoder to blend between $\mathcal{X}$ and $\mathcal{S}_{\mathcal{X}}$.

Density based Style Transfer



Cluster Based AC

- FoldingNet only works for small point clouds
- In practice resolution is much higher

Cluster Based AC

Idea:

Cluster Based AC

Idea:

- Split point clouds into smaller pieces

Cluster Based AC

Idea:

- Split point clouds into smaller pieces
- Feed the smaller point clouds into the autoencoder

Cluster Based AC

Idea:

- Split point clouds into smaller pieces
- Feed the smaller point clouds into the autoencoder
- Reassemble

Cluster Based AC

- Given two point clouds $\mathcal{X}$ and $\mathcal{Y}$

Cluster Based AC

- Given two point clouds $\mathcal{X}$ and $\mathcal{Y}$
- Cluster $\mathcal{X}$ into k clusters $\mathcal{X}_1, \ldots, \mathcal{X}_k$

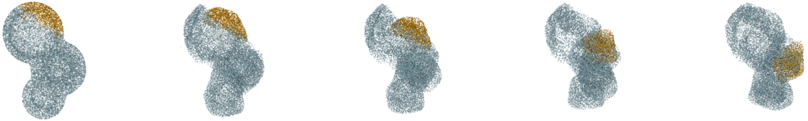
Cluster Based AC

- Given two point clouds $\mathcal{X}$ and $\mathcal{Y}$
- Cluster $\mathcal{X}$ into k clusters $\mathcal{X}_1, \ldots, \mathcal{X}_k$
- Cluster $\mathcal{Y}$ into k clusters $\mathcal{Y}_1, \ldots, \mathcal{Y}_k$

Cluster Based AC

- Given two point clouds $\mathcal{X}$ and $\mathcal{Y}$
- Cluster $\mathcal{X}$ into k clusters $\mathcal{X}_1, \ldots, \mathcal{X}_k$
- Cluster $\mathcal{Y}$ into k clusters $\mathcal{Y}_1, \ldots, \mathcal{Y}_k$
- Blend between $\mathcal{X}_i$ and $\mathcal{Y}_i$ for $i = 1, \ldots, k$

Cluster Based AC

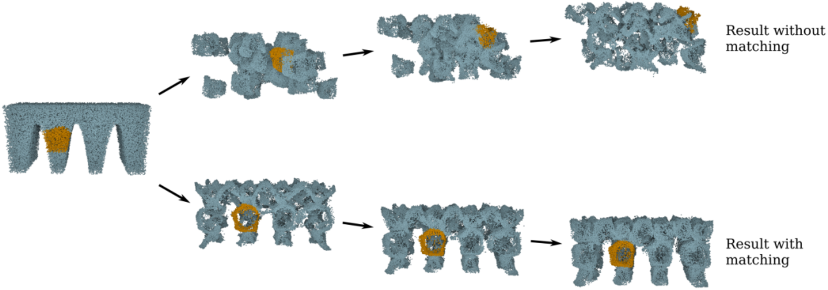


GT $\lambda = 0.1$ $\lambda = 0.25$ $\lambda = 0.5$ $\lambda = 0.75$

Cluster Based AC

- Given two point clouds $\mathcal{X}$ and $\mathcal{Y}$.
- Cluster $\mathcal{X}$ into k pieces and let $C_1, \ldots, C_k$ denote the centroids of these clusters.
- Cluster $\mathcal{Y}$ into k clusters with centers $C_1, \ldots, C_k$.
- Blend between $\mathcal{X}_i$ and $\mathcal{Y}_i$ for $i = 1, \ldots, k$

Cluster Based AC

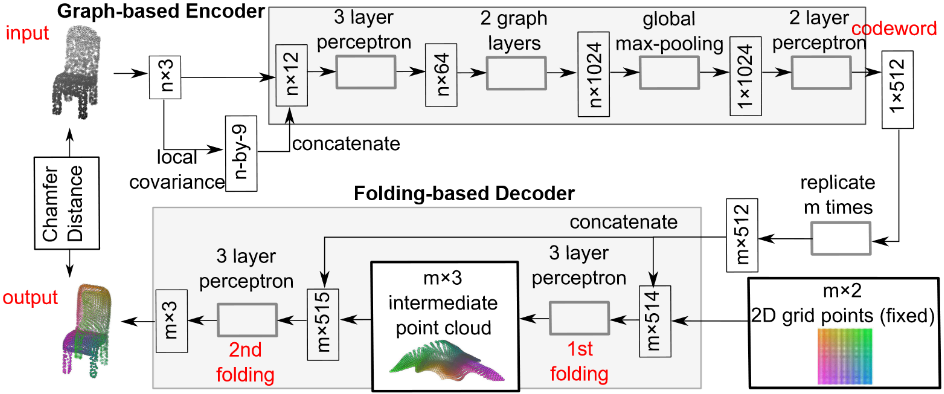


Thank you!

References I

- Stephan Antholzer, Martin Berger, and Tobias Hell.
  Cluster-based autoencoders for volumetric point clouds, 2022.
- Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas.
  Pointnet: Deep learning on point sets for 3d classification and segmentation.
  CoRR, abs/1612.00593, 2016.
- Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian.
  Foldingnet: Point cloud auto-encoder via deep grid deformation, 2018.

Appendix - Graph Layer



Appendix - Graph Layer

- Let X denote the input of the Graph layer.
- Create a matrix A via

$$A_{ij} = \mathrm{ReLU}(\max_{k \in \mathcal{N}(i)} x_{kj})$$

  where $\mathcal{N}(i)$ denotes the K nearest neighbors of ith row of X.
- Then the output is given via

$$A \cdot M$$

  where M is the feature mapping matrix.