#DLUPC

Day 4 Lecture 1

3D Analysis

Javier Ruiz Hidalgo
j.ruiz@upc.edu

Associate Professor
Universitat Politecnica de Catalunya
Technical University of Catalonia

[course site]

Outline

- Motivation
- Point Clouds
- 3D datasets
- Deep Learning considerations
- Techniques
    - 2.5D
    - Voxelization
    - Projection
    - Direct
- Conclusions

Acknowledgments

Belen Luque López
CV Master student

Alba Pujol Miró
PhD Student

Motivation

- New / cheaper / smaller sensors to acquire 3D structure of the scene
  - Microsoft Kinect, Structure Sensor , Primesense Carmine
  - New datasets
- Virtual and Augmented reality applications

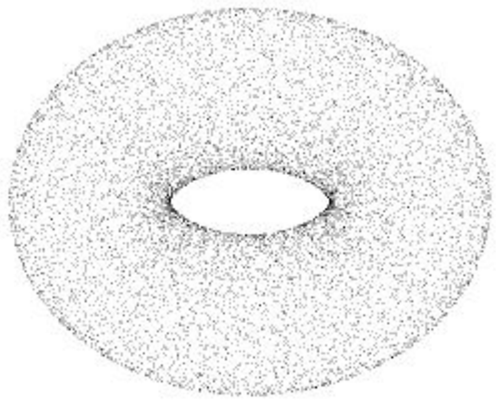
Primesense Carmine

Oculus rift

Kinect

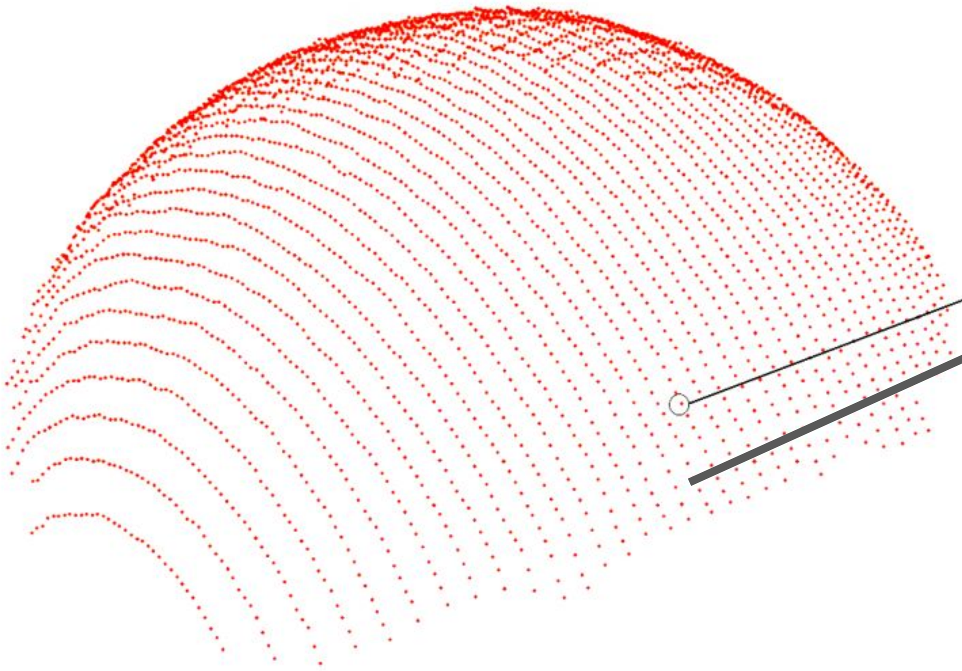
Point clouds (1)

- Common representation for 3D data
- Collection of data points defined by a given coordinates system
- Represents **surface** of objects
- Usually in Cartesian Coordinate System
  - X,Y,Z coordinates for each point of the cloud



3D Point cloud of a Torus

Point clouds (2)



[Example of a point cloud]

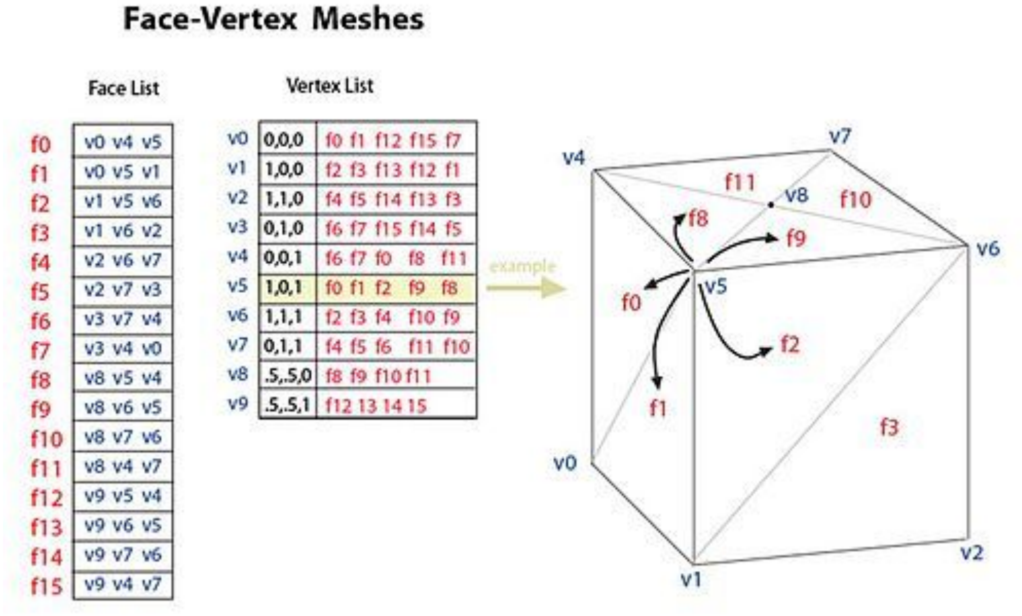
X	Y	Z
26716	-5456	12120
26716	-5424	12120
26716	-5392	12120
26704	-5364	12120
26696	-5336	12120
26692	-5308	12120
26684	-5276	12120
26676	-5248	12120
26668	-5216	12120
26656	-5184	12120
26652	-5156	[illegible]
[illegible]	[illegible]	15840
26644	-5096	15840
26636	-5064	15840
26628	-5036	15840
26616	-5004	15840
26608	-4976	15840
26600	-4944	15840

Extra data can be added for each point:

- Color (RGB)
- Orientation
- Curvature

Point clouds vs. Meshes

- A more complete 3D representation may include faces defined between points / vertices



Face-Vertex Meshes

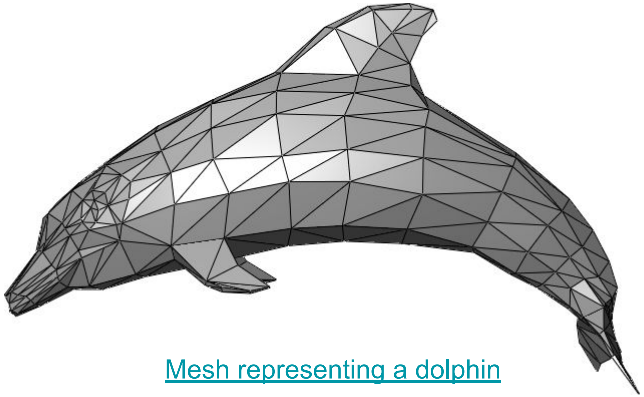
Face List

f0	v0 v4 v5
f1	v0 v5 v1
f2	v1 v5 v6
f3	v1 v6 v2
f4	v2 v6 v7
f5	v2 v7 v3
f6	v3 v7 v4
f7	v3 v4 v0
f8	v8 v5 v4
f9	v8 v6 v5
f10	v8 v7 v6
f11	v8 v4 v7
f12	v9 v5 v4
f13	v9 v6 v5
f14	v9 v7 v6
f15	v9 v4 v7

Vertex List

v0	0,0,0	f0 f1 f12 f15 f7
v1	1,0,0	f2 f3 f13 f12 f1
v2	1,1,0	f4 f5 f14 f13 f3
v3	0,1,0	f6 f7 f15 f14 f5
v4	0,0,1	f6 f7 f0 f8 f11
v5	1,0,1	f0 f1 f2 f9 f8
v6	1,1,1	f2 f3 f4 f10 f9
v7	0,1,1	f4 f5 f6 f11 f10
v8	.5,.5,0	f8 f9 f10 f11
v9	.5,.5,1	f12 13 14 15

Mesh representation as a list of face-Vertex



Mesh representing a dolphin

3D datasets: Classification

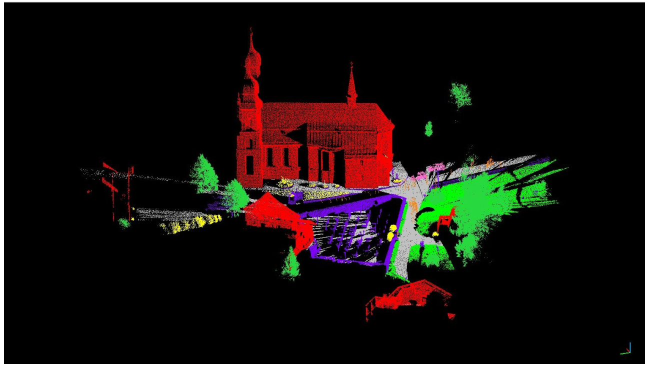
- Large Dataset of Object Scans
  - PrimeSense Carmine sensor
  - 10k scans
  - 43 objects

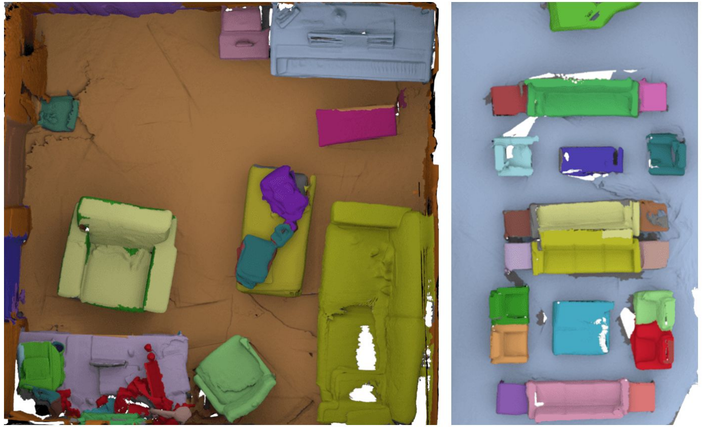
3D datasets: Pose estimation

- [T-less: An RGB-D Dataset for 6D Pose Estimation of Texture-less Objects]
    - Primesense Carmine, Kinect v2 and Canon IXUS 950 sensors
    - 38k (training) + 10k (test) scans
    - 30 objects + groundtruth pose

3D datasets: Segmentation

- Sematic3d
  - Velodyne LIDAR sensor
  - 30 scenes, 1 billion labelled points
  - 8 classes

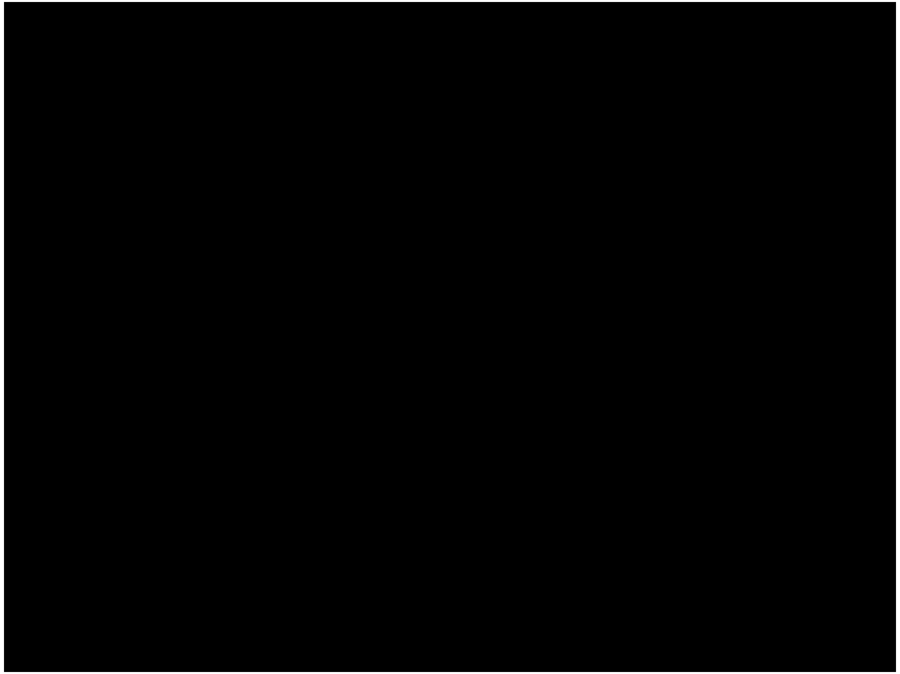




- ScanNet
  - Structure sensor
  - 1.5k scenes, 2.5M views
  - 20 classes

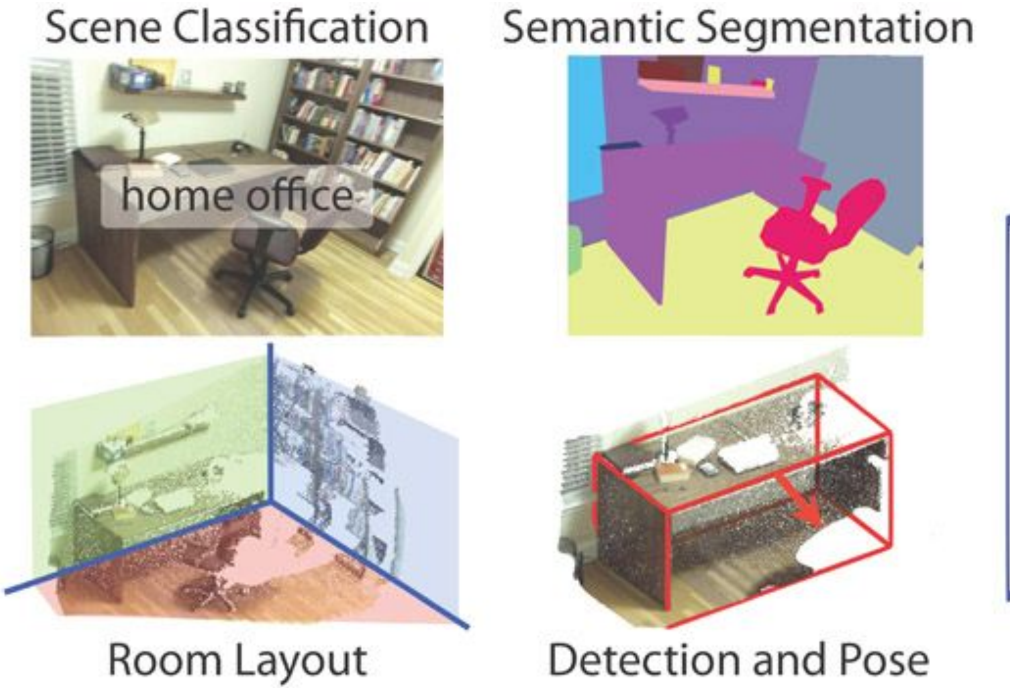
3D datasets: Autonomous driving

- Cityscapes: semantic understanding of urban street scenes
    - Stereo cameras
    - 5 cities, 20k images
    - 20 classes (instances)



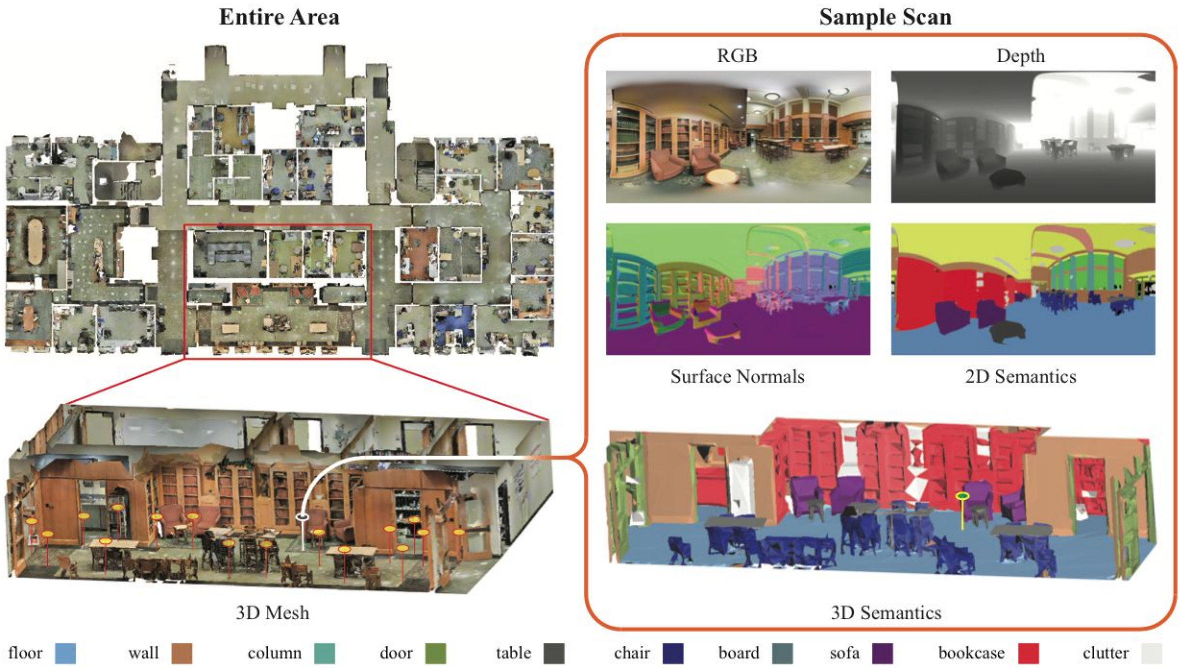
3D datasets: Scene understanding (1)

- [SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite]
  - Kinect sensor
  - 10k scans



3D datasets: Scene understanding (2)

- [Stanford 2D-3D-Semantics dataset]
  - Kinect sensor
  - 70k scans, 6 areas with over 6000 m^2
  - 13 classes



Deep learning from 3D point clouds? (1)

There are several challenges when using 3D point clouds in a deep learning framework:

1. Undefined neighborhood

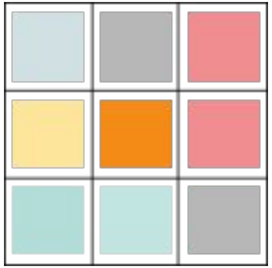
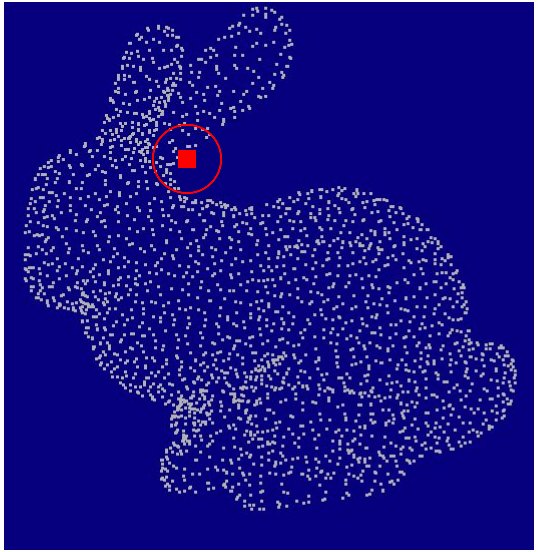


Image neighbours are easily defined by their connectivity

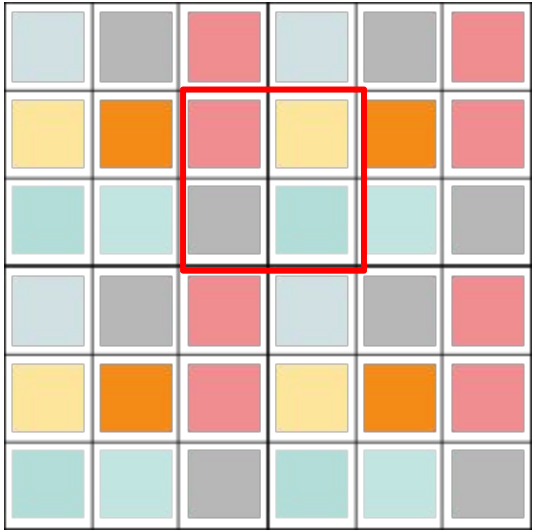


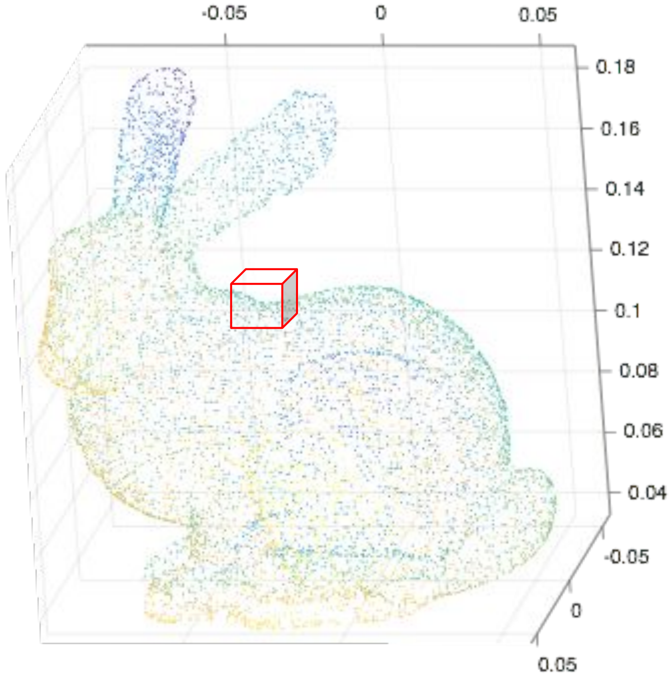
Point cloud neighbours need to be explicitly defined (Euclidean distance)

Deep learning from 3D point clouds? (2)

There are several challenges when using 3D point clouds in a deep learning framework:

2. No lattice (convolution layers?)

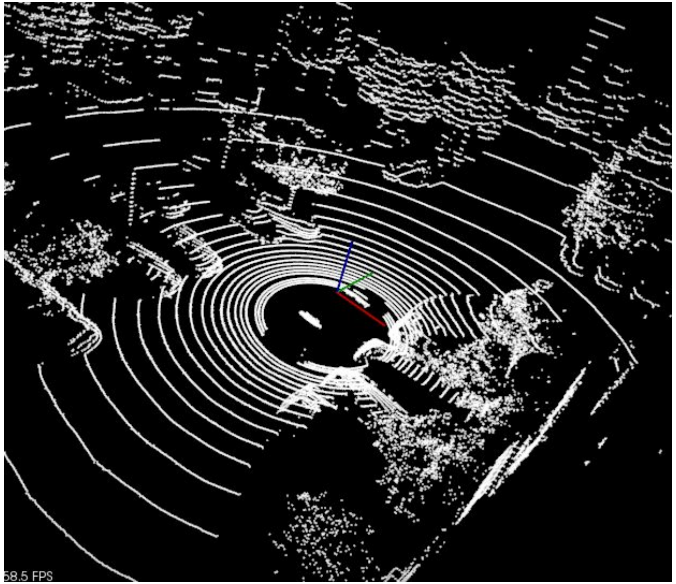




Deep learning from 3D point clouds? (3)

There are several challenges when using 3D point clouds in a deep learning framework:

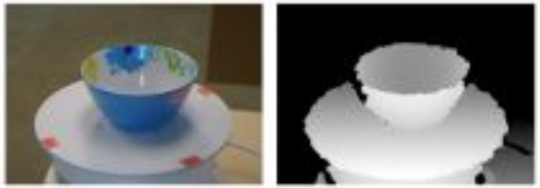
3. Different density



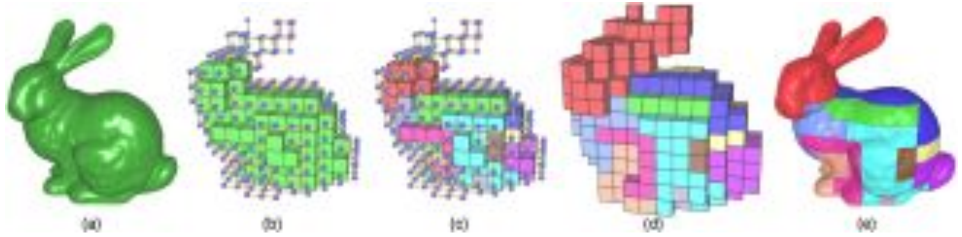
Velodyne LIDAR data

Techniques

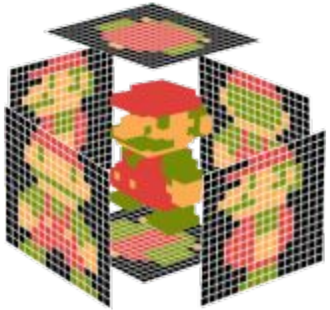
RGB-D / 2.5D Data



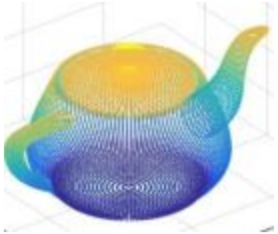
Voxelization



Multiview Projection



Point clouds



RGB-D / 2.5D data (1)

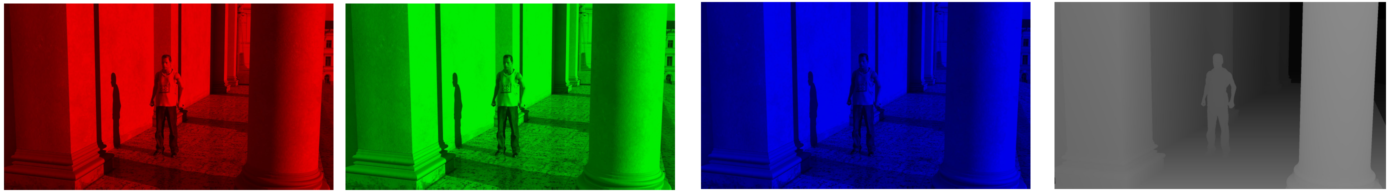
Use depth as RGB + Depth (RGB-D) images

- Very common and used in the deep learning literature
    - RAW data from Kinect / Structure / Primesense sensors
- Multiple applications (classification, gesture recognition, semantic segmentation, etc.)

RGB DEPTH POINT CLOUD

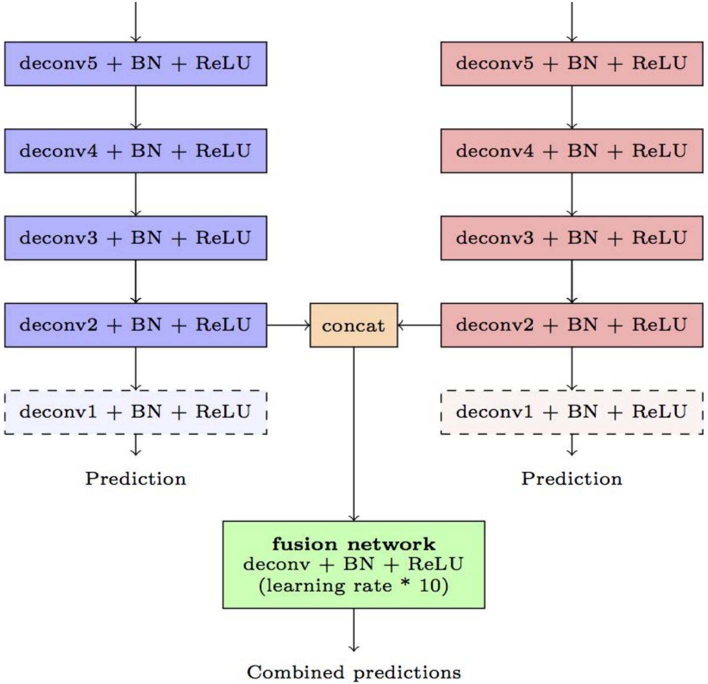
RGB-D / 2.5D data (2)

- Straight-forward solution ➔ include depth as a new channel (RGBD input)



RGB-D / 2.5D data (3)

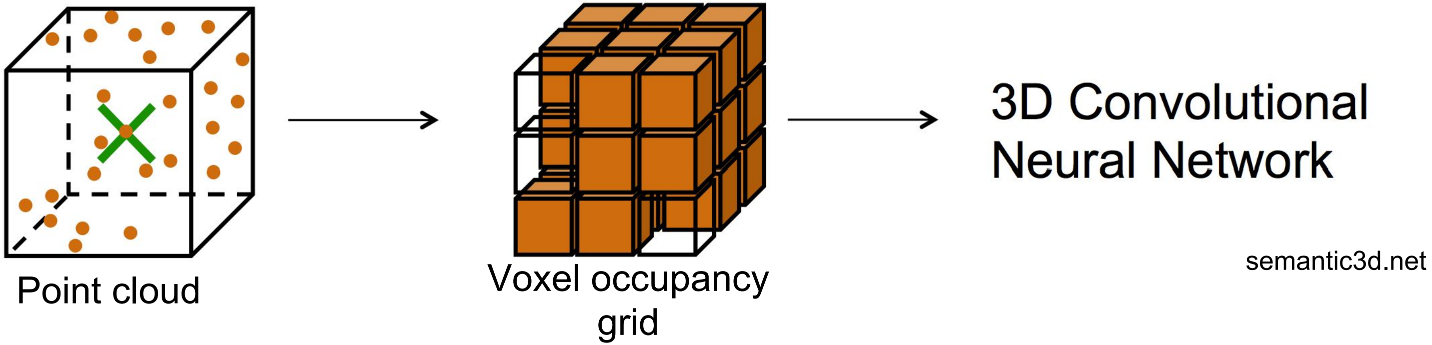
- However, better results are obtained when depth is incorporated as a two-stream network



N. Audebert, et al. [Fusion of heterogeneous data in convolutional networks for urban semantic labeling](), JURSE 2017

Voxelization

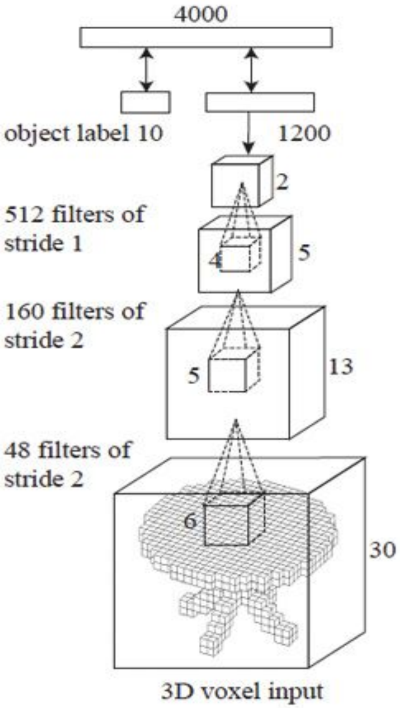
- Discretize 3D space with occupancy grid



- Difficult to define a voxel size for all applications (density of point clouds)
- Use 3D convolutional layers

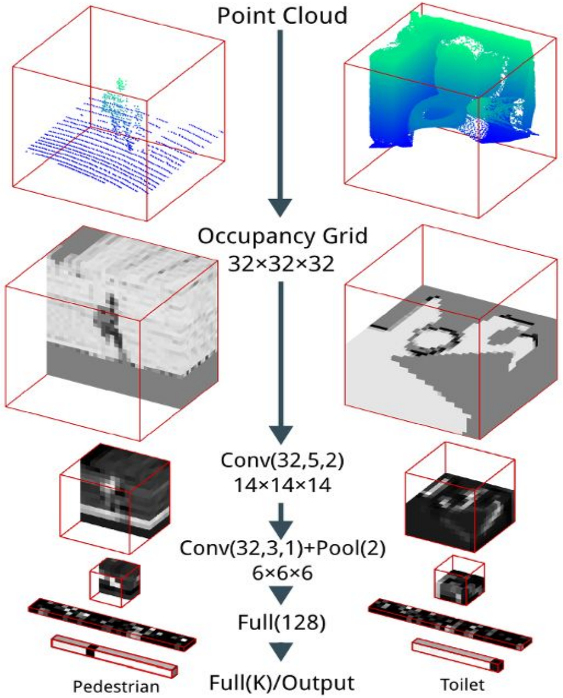
Voxelization: Architecture examples

3D ShapeNets



Z. Wu, et al. [3D ShapeNets: A Deep Representation for Volumetric Shapes](), CVPR 2015

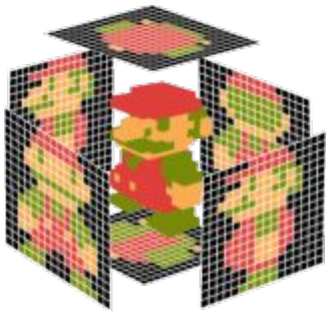
VoxNet



D. Maturana, et al. [VoxNet: A 3D Convolutional Neural Network for real-time object recognition](), IROS 2015

Projection

- Instead of using the point cloud directly or voxelize it, project back the point cloud into 1 (single) or several (multi-view) images
- Use the projected images as input tensors for the network



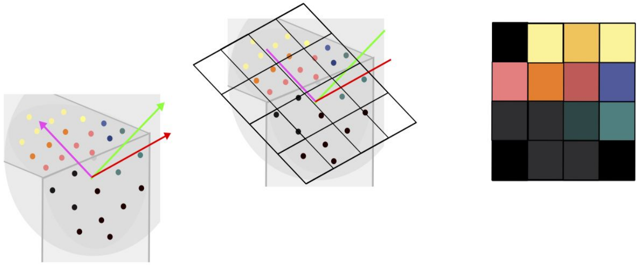
Projection: Example (1)

Point cloud 1

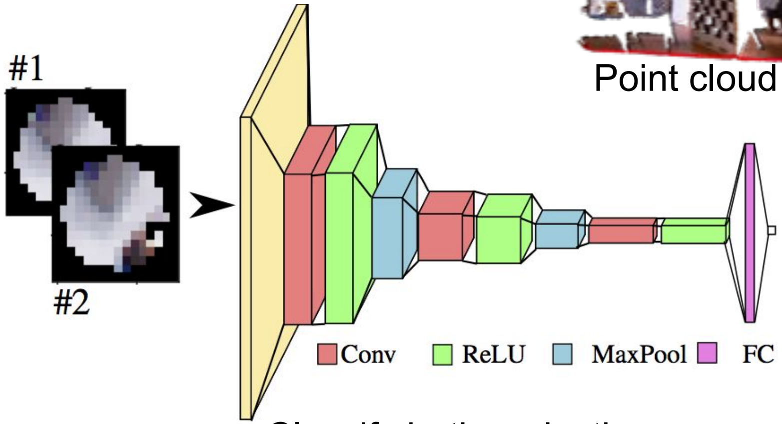
Point cloud 2

Correspondence matching

- Find correspondent 3D points between two point clouds



Project neighbouring points into principal plane for the two candidates

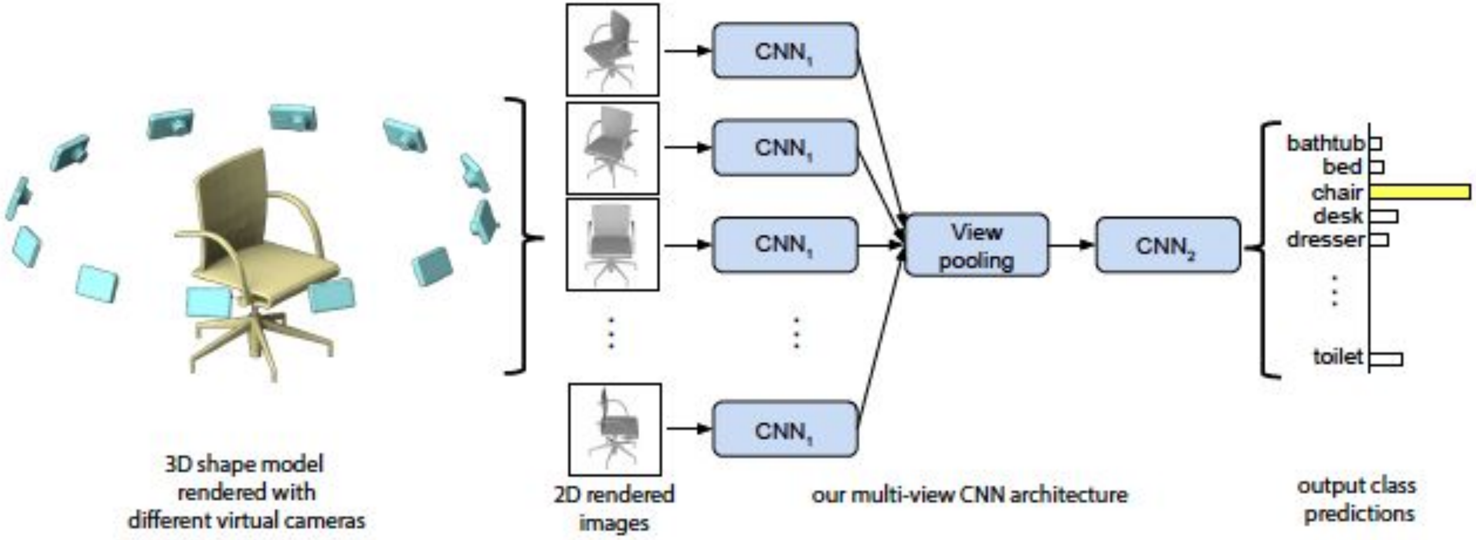


Classify both projections as correspondences or not (binary classification)

A Pujol, J.R. Casas, J. Ruiz-Hidalgo, (Work in progress)

Projection: Example (2)

Multiview projections

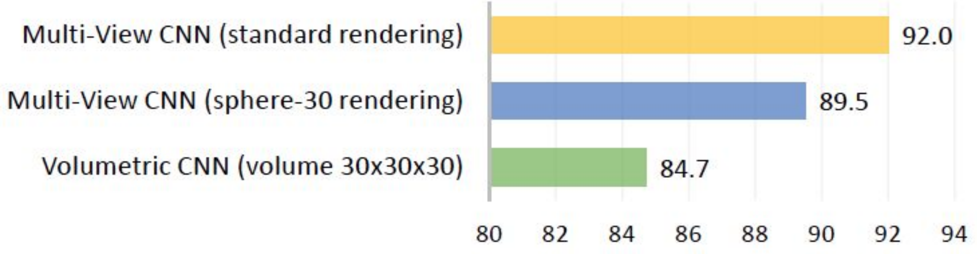


H. Su, et al. **Multi-view Convolutional Neural Networks for 3D Shape Recognition**, ICCV, 2015

Voxelization vs. Projection (1)

Experiments indicate that projection based multiview CNNs perform better than voxelized volumetric representations

- Accuracy in object classification

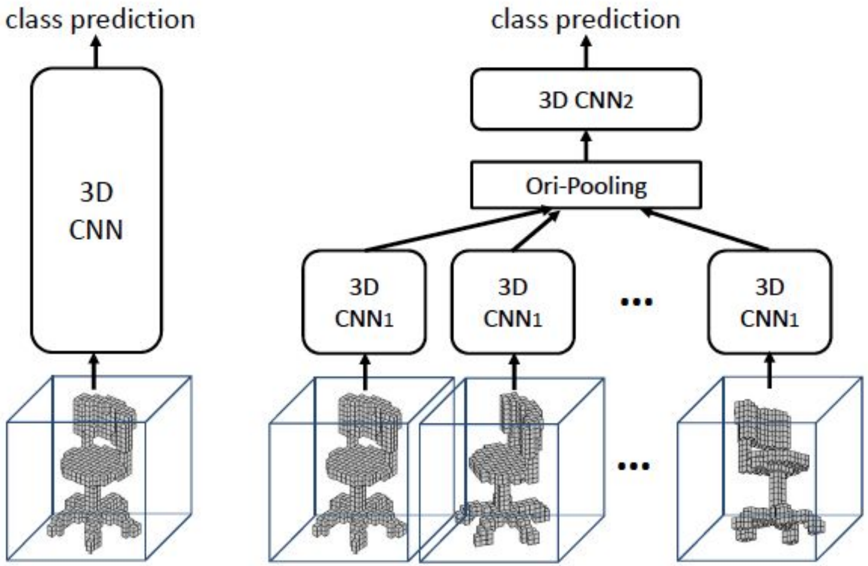


C. R. Qi, et al. [Volumetric and Multi-View CNNs for Object Classification on 3D Data](), CVPR 2016

Voxelization vs. Projection (2)

Experiments indicate that projection based multiview CNNs perform better than voxelized volumetric representations

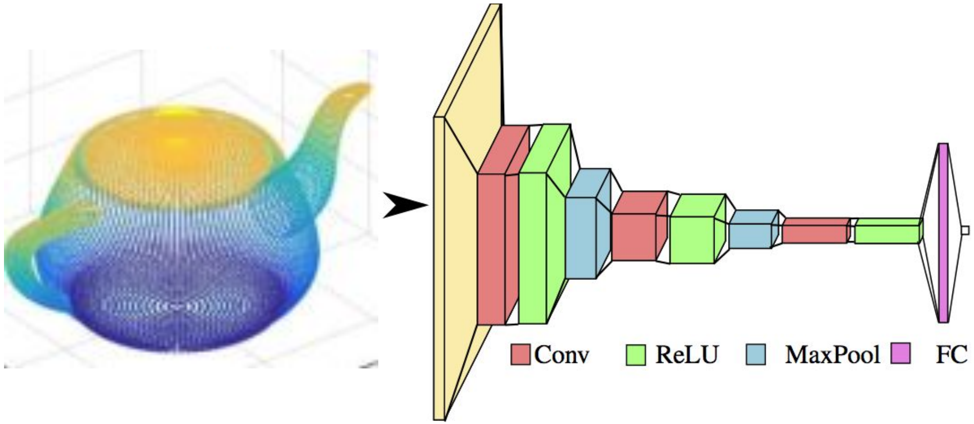
- New architectures are closing the gap



C. R. Qi, et al. [Volumetric and Multi-View CNNs for Object Classification on 3D Data](), CVPR 2016

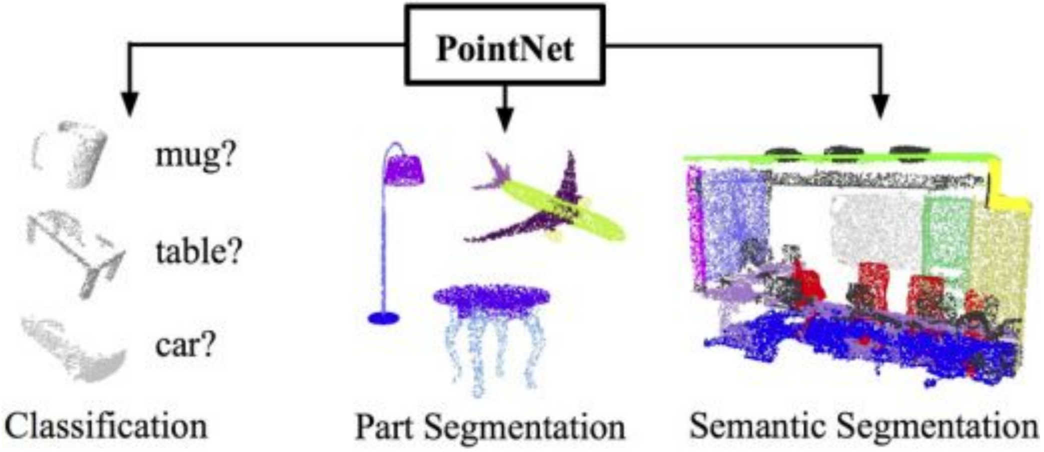
Point clouds

Is it possible to use point clouds directly in a learning framework?



PointNet (1)

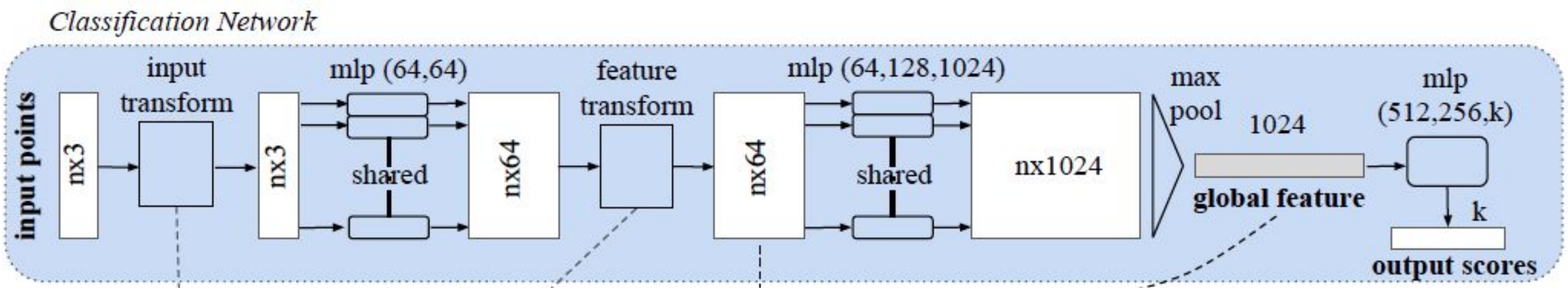
Novel deep learning architecture that directly consumes point clouds



C. R. Qi, et al. [PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,]() CVPR 2017

PointNet (2)

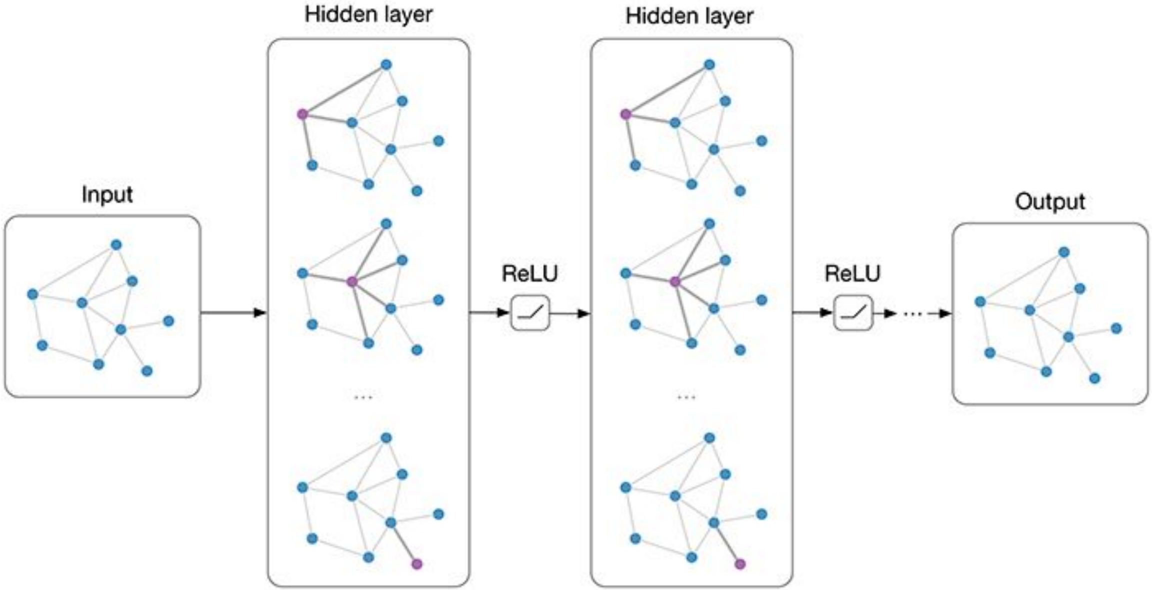
Input is an unordered list of XYZ coordinates (point cloud)



C. R. Qi, et al. [PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation](), CVPR 2017

Graph Neural Networks - GNNs (1)

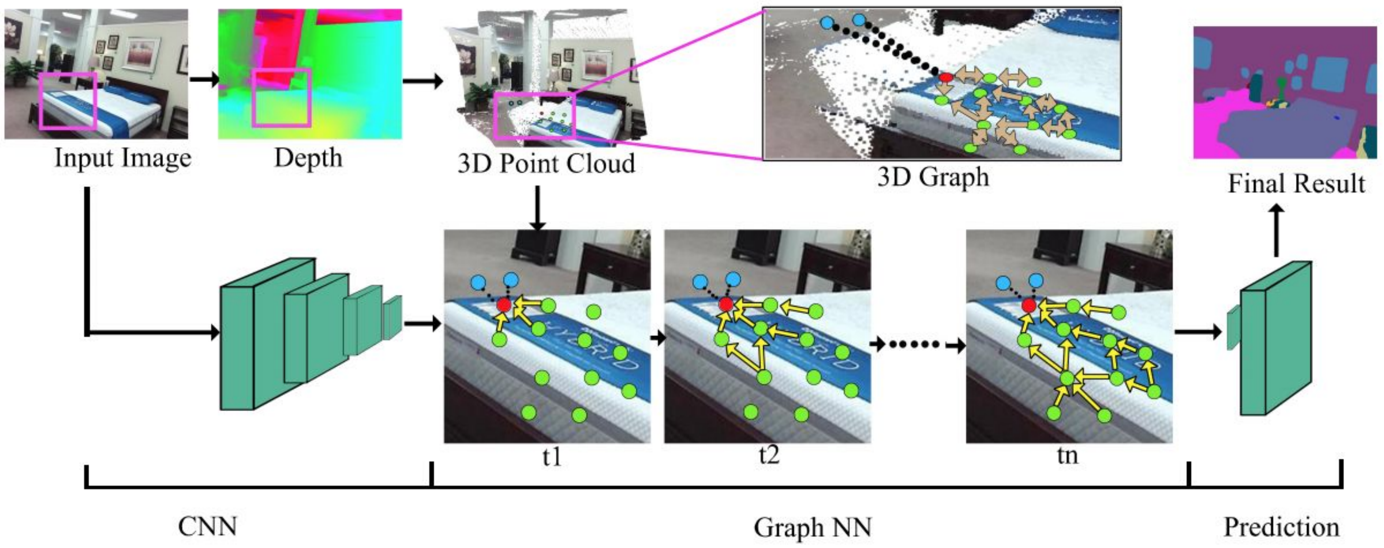
Generalization of CNNs to work on arbitrarily structured graphs



T. N. Kipf, M, Welling, SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS, ICLR 2017

Graph Neural Networks - GNNs (2)

Build graphs from 3D point clouds



X, Qi, et.al., [3D Graph Neural Networks for RGBD Semantic Segmentation](), ICCV 2017

Conclusions

- Point clouds rich representation of 3D data
- Working with point clouds on deep learning frameworks is a challenge due to the unorganized nature of point clouds
- Techniques
  - RGB-D / 2.5D data
  - Voxelization
  - Multi-view projections
  - Point clouds as input gaining momentum

Questions?

PointNet (3)

Classification results

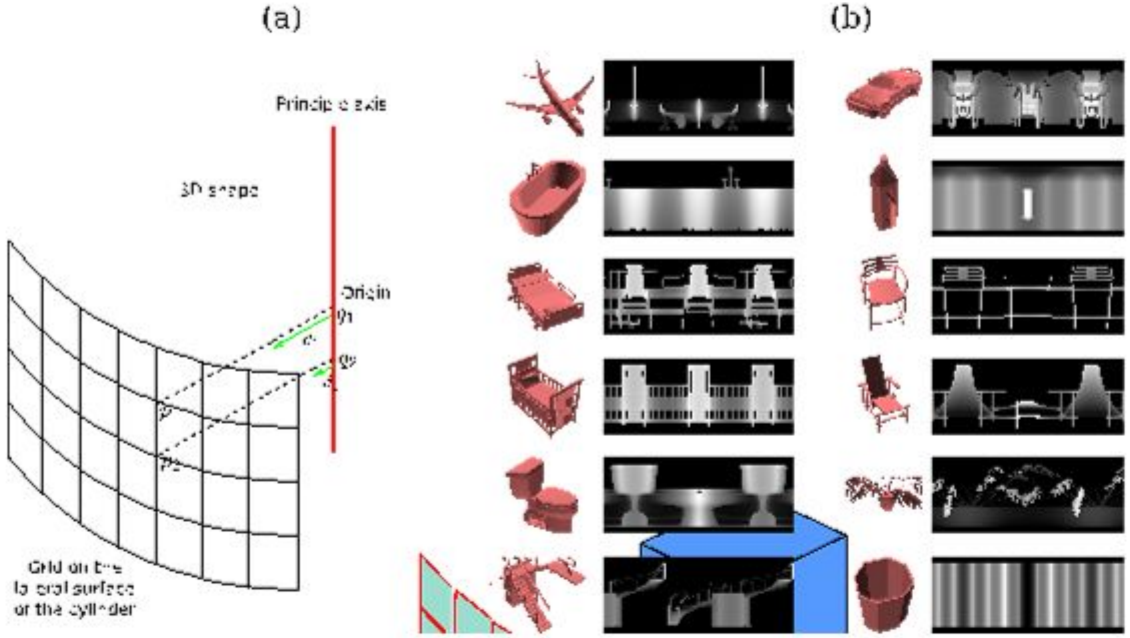
	input	#views	accuracy avg. class	accuracy overall
SPH [11]	mesh	-	68.2	-
3DShapeNets [28]	volume	1	77.3	84.7
VoxNet [17]	volume	12	83.0	85.9
Subvolume [18]	volume	20	86.0	**89.2**
LFD [28]	image	10	75.5	-
MVCNN [23]	image	80	**90.1**	-
Ours baseline	point	-	72.6	77.4
Ours PointNet	point	1	86.2	**89.2**

Table 1. **Classification results on ModelNet40.** Our net achieves state-of-the-art among deep nets on 3D input.

C. R. Qi, et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, CVPR 2017

Projection: Example (3)

Panoramic projections: DeepPano



D. Shi, et al. [DeepPano: Deep Panoramic Representation for 3-D Shape Recognition](), IEEE Signal Processing Letters, 2015