Autonomous Driving

1 Recap: Introduction to Deep Learning

2 3D Data Representation

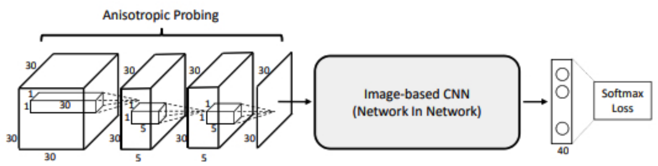
- Explicit Representations: Voxel grids, Point Clouds, Polygonal Meshes
- Implicit Representations: Signed Distance Functions (SDFs), Occupancy Networks

2.1 Volumetric Grids

Volumetric Data Structures

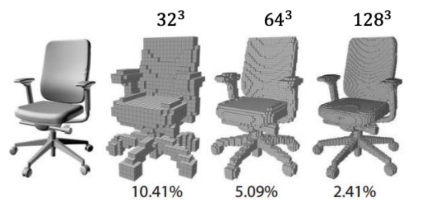
- Occupancy grids
- Ternary grids (2-D triangular coordinate system, e.g., Barycentric coordinate)
- (Signed) Distance fields

Extension of AlexNet/2D CNNs to 3D CNNs by using 3D convolutional kernels: Object Classification with 3DCNNs, 2016



Summary:

+ Simple data structure
+ Encode free space and distance fields
+ Naturally extend 2D CNNs and other concepts
+ Faster training due to an additional dimension per sample (but often lack 3D data)
– High memory consumption (cubic growth), a lot is used for empty space the higher the resolution
– Require a lot of processing time (use sliding window or fully-convolutions)



Surface occupancy gets smaller with higher resolutions

2.2 Volumetric Hierarchies

Octrees

- Hierarchical data structure that recursively subdivides 3D space into eight octants (like a quadtree in 2D).
- 3D partitioning: higher resolution at surfaces, lower resolution in empty space for applying 3DCNN kernels
- Can be used for discriminative tasks (like object classification, segmentation) or generative tasks (like 3D reconstruction).

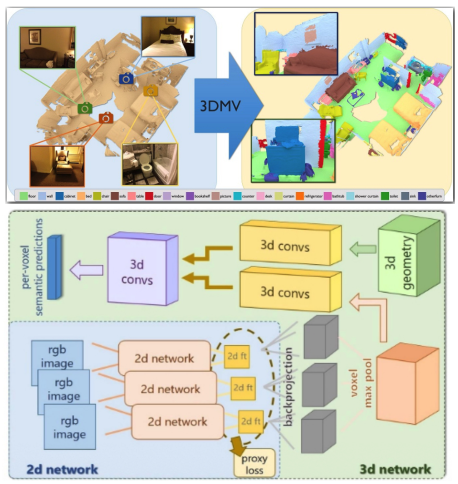
Summary:

+ Great for reducing memory and runtime
+ Increases strongly performance of 3D CNNs
– Easier for discriminative tasks when structure is known, more complex for generative tasks (split voxel that are partially occupied)

2.3 Hybrid: Volumetric + Multi-View

- Combine volumetric representation with multi-view images (colors) to improve segmentation
- Separate 2D (image) and 3D (voxel) feature extraction + backprojection of 2D features into 3D space

3DMV: 3D Voxel + Multi View Semantic Segmentation, 2016



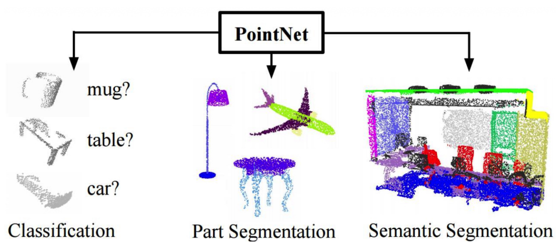
Summary:

+ Nice way to combine color and geometry
+ Good performance
– End-to-End training helps less than expected
– Could be faster

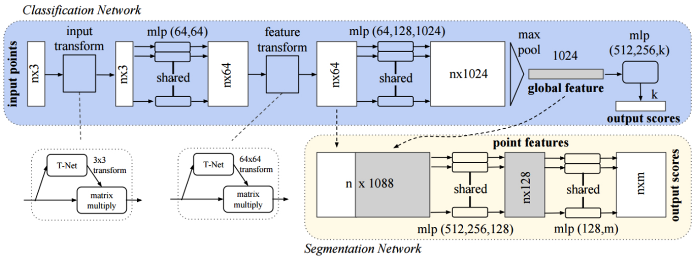
2.4 Point Clouds

- Unordered set of 3D points (often from LiDAR or depth sensors)
- Each point can have additional features (e.g., color, intensity, normals)
- Direct processing of point clouds using specialized neural networks (e.g., PointNet, PointNet++)

PointNet: 3D Classification and Segmentation for Point Clouds, 2017



- take n points as input
- apply feature transformations and aggregate features by max pooling
- classification net: output scores for k classes
- segmentation net: concatenate global and local features for per-point classification



PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space, 2017

- learn hierarchical representation of point clouds
- apply multiple PointNets at different locations and scales
- multiscale grouping (MSG) and multi-resolution grouping (MRG) for local features

Point Convolutions

- Transform points to continuous representation using radial basis functions (RBFs) or kernel density estimation (KDE)
- Apply convolutional operations on the continuous representation

Point Transformer

- Use self-attention mechanisms to capture relationships between points
- Learn point features by attending to neighboring points

2.5 Sparse Convolutional Networks

- Efficiently process sparse 3D data (e.g., point clouds, voxels with many empty spaces)
- Use sparse convolutional operations that only compute on non-empty voxels
- To improve efficiency, use sparse hashes (e.g. QSParseConv, MinkowskiEngine) instead of masking conv output
- Latency on GPU for hash lookup is hidden by parallelism

Summary:

+ Efficient representation, only store surface points
+ Fast training and testing
+ Cover large space in one shot
– Can not represent free space
– Perform worse than volumetric methods in some tasks (a lot on-going research)

2.6 Polygonal Meshes

- Represent 3D surfaces using vertices, edges, and faces (typically triangles or quadrilaterals)
- Widely used in computer graphics and 3D modeling
- Process via graph neural networks
  - Message Passing
  - Graph Convolutions
  - Transformers
- Process via specialized mesh convolutional networks
  - MeshCNN
  - Geodesic CNNs
  - Spectral CNNs

2.7 Signed Distance Functions (SDFs)

2.8 Occupancy Networks

3 3D Datasets

3.1 3D Shapes

ShapeNet

- Main Dataset, synthetic 3D models
- 51.3k shapes, 55 classes
- mostly chairs, mediocre textures

Objaverse-XL

- 10mio 3D shapes
- heterogeneous distributed

3.2 3D Scenes

3D-FRONT

- furnished rooms with layouts and semantics
- 18k rooms, 7k furniture objects