Volumetric and Multi-View CNNs for Object Classification on 3D data

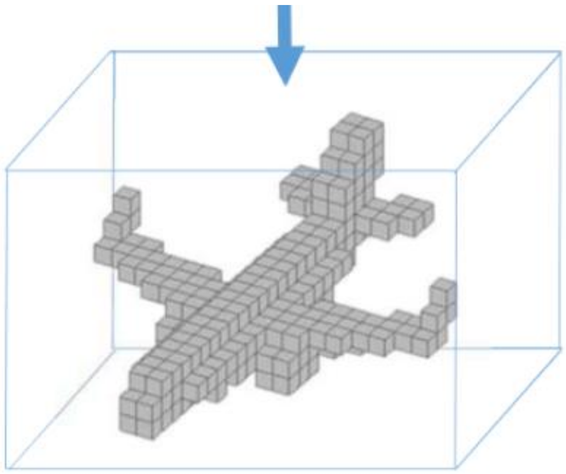
Charles R. Qi; Hao Su; Matthias Nießner; Angela Dai; Mengyuan Yan; Leonidas J. Guibas

Stanford University

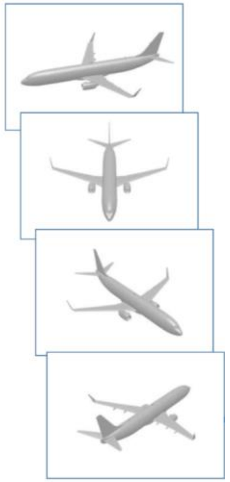
Introduction

- Using CNNs (Convolutional Neural Networks) to learn 3D features
- Object classification

Volumetric CNNs

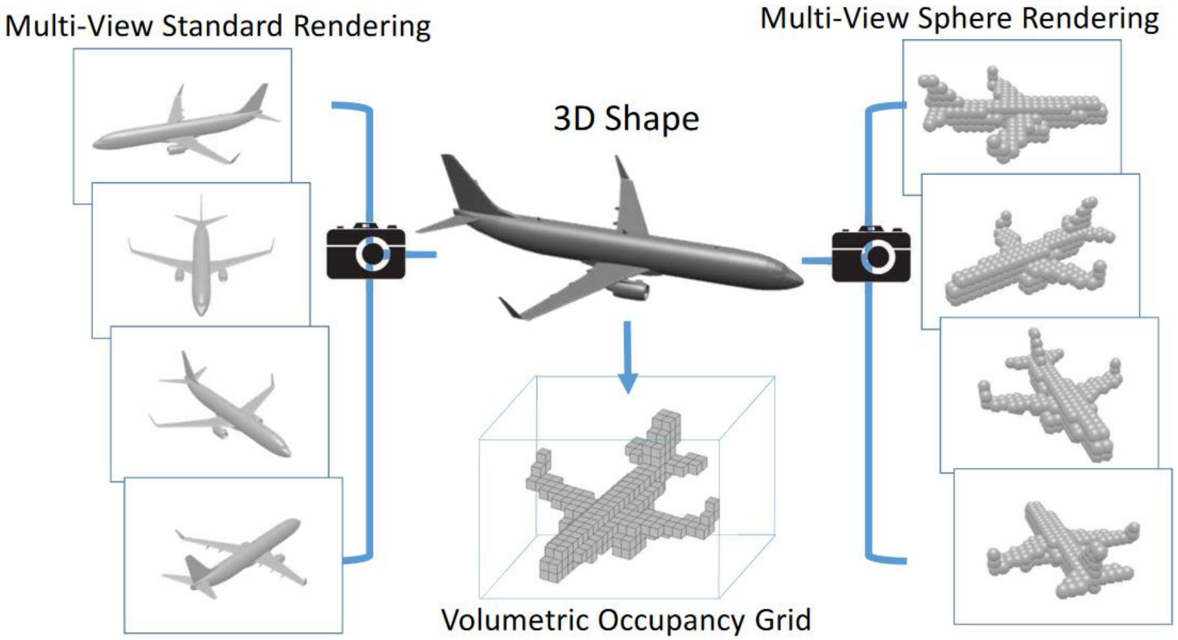


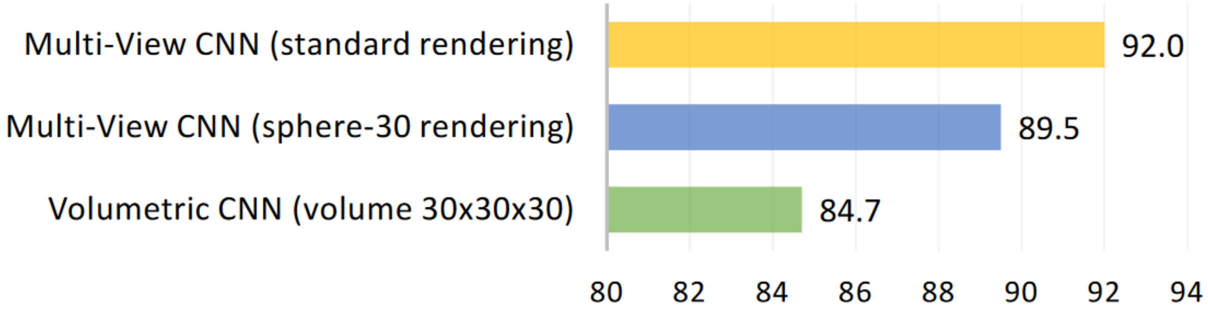
Multi-view CNNs



State-of-the-art

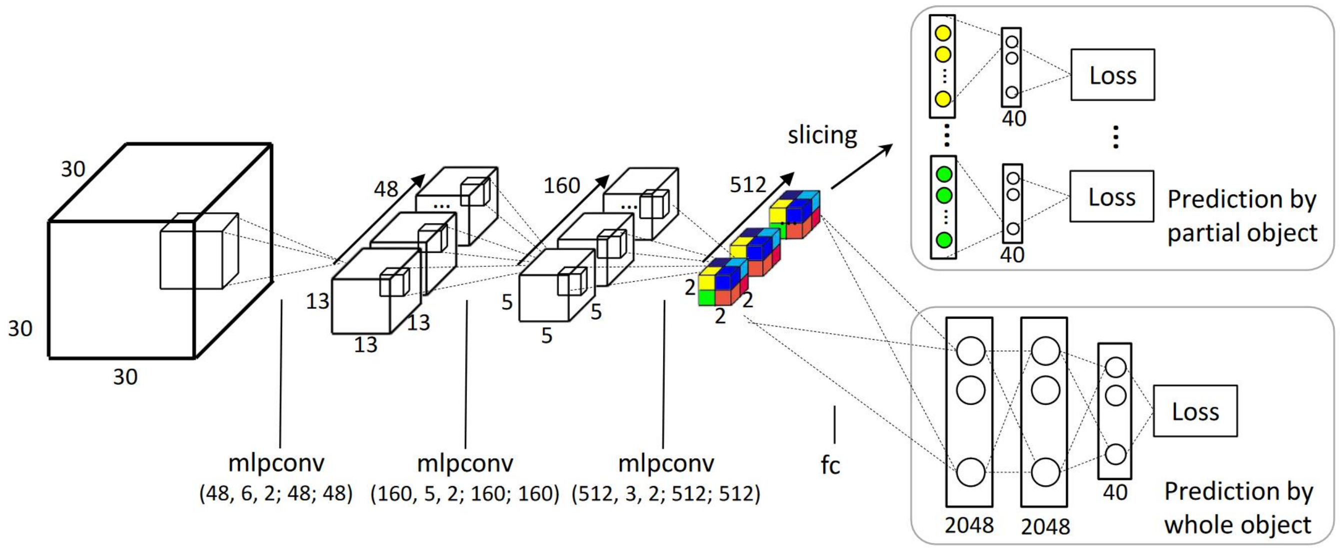
- Volume: **3D tensor** of binary/real values
  - 30x30x30
- Multi-view: **Collection of renderings** (2D) from multiple viewpoints
  - 227x227
- Similar details: Render sphere for each voxel





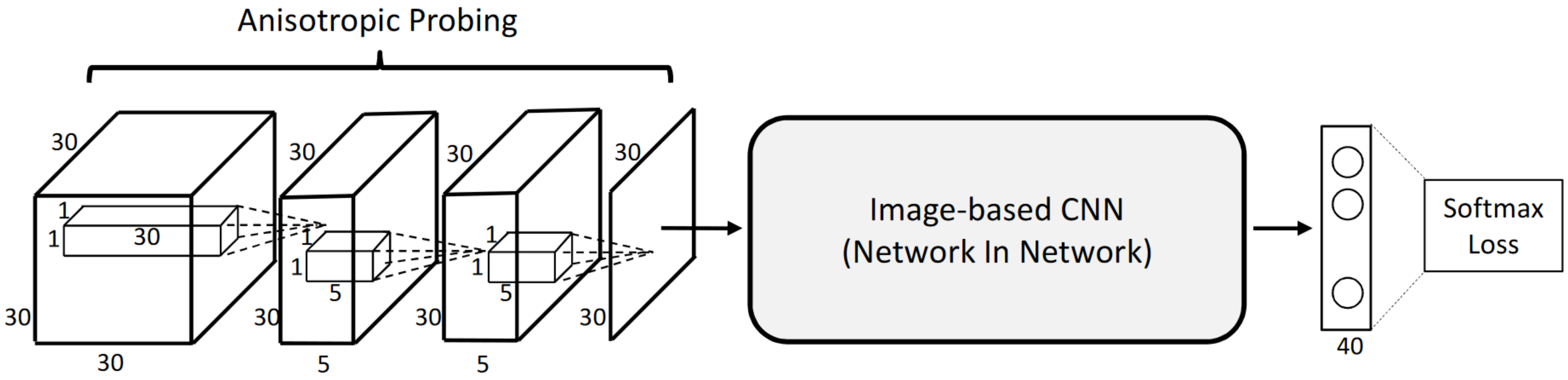
Volumetric CNNs

- Network architecture
  1. Auxiliary training tasks – closely related to main task
     - Classify using local subvolumes – difficult to overfit!



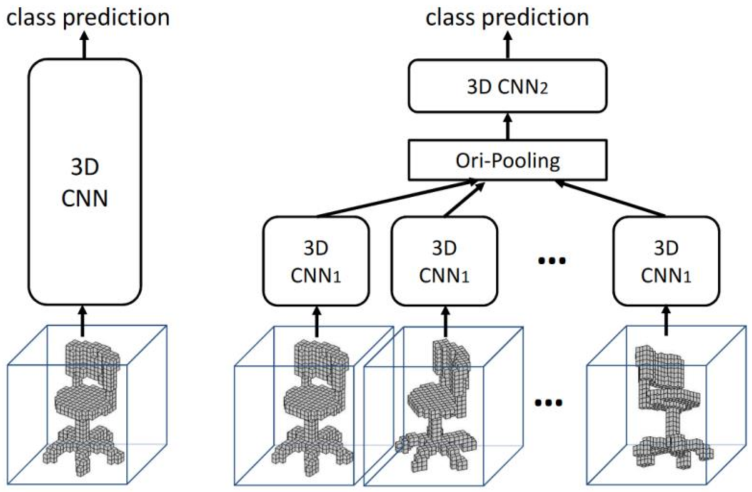
Volumetric CNNs

- Network architecture
  2. Anisotropic Probing
     - 3D-to-2D-projection using elongated kernels (“X-ray”)



Volumetric CNNs

- Data augmentation
  - Input varying object orientations into one network
- Multi-orientation pooling
  - Decompose network with various orientations of 3D input



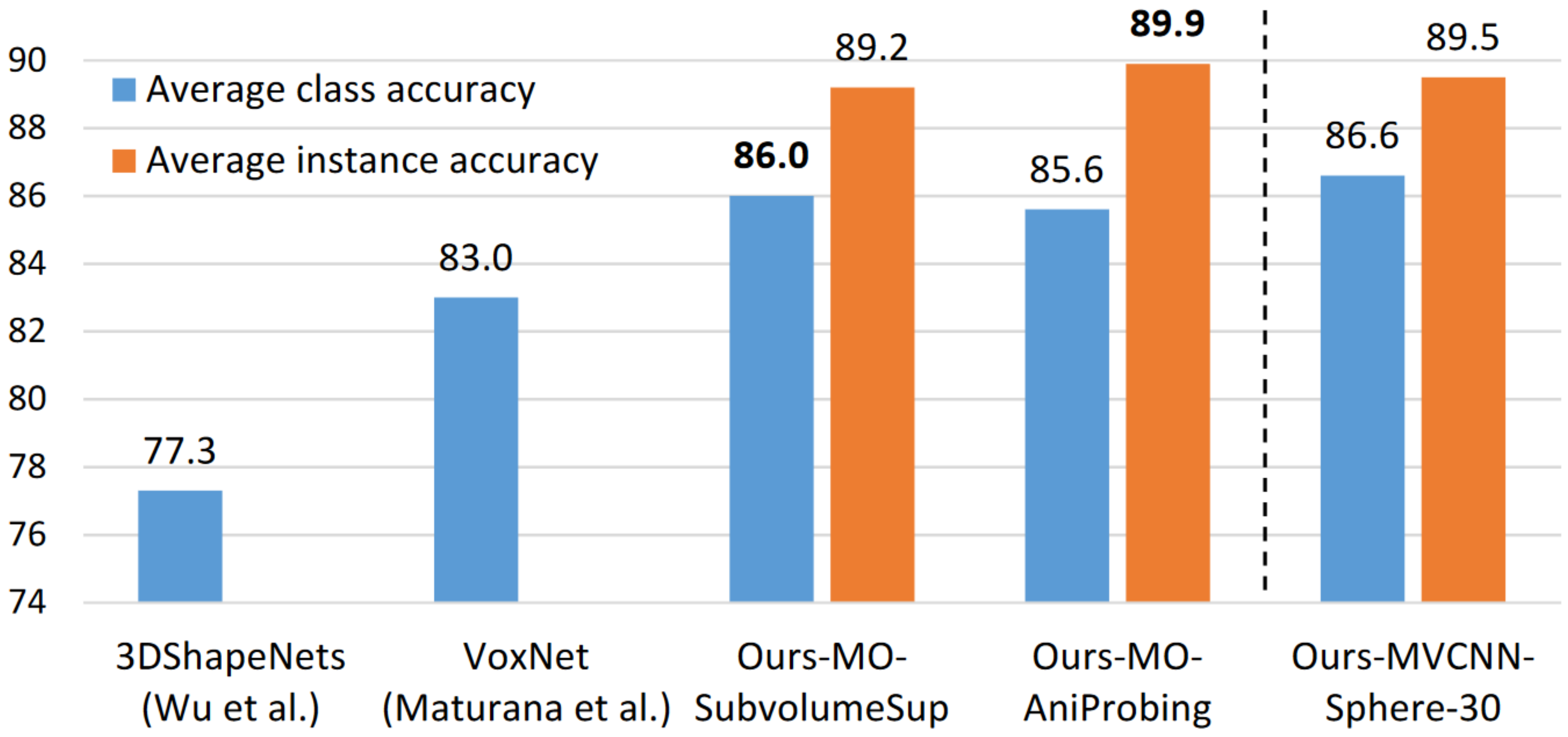
Multi-view CNNs

- Data augmentation – **Multi-resolution** sphere rendering
  - Spheres are view-invariant
  - Smooth out noise

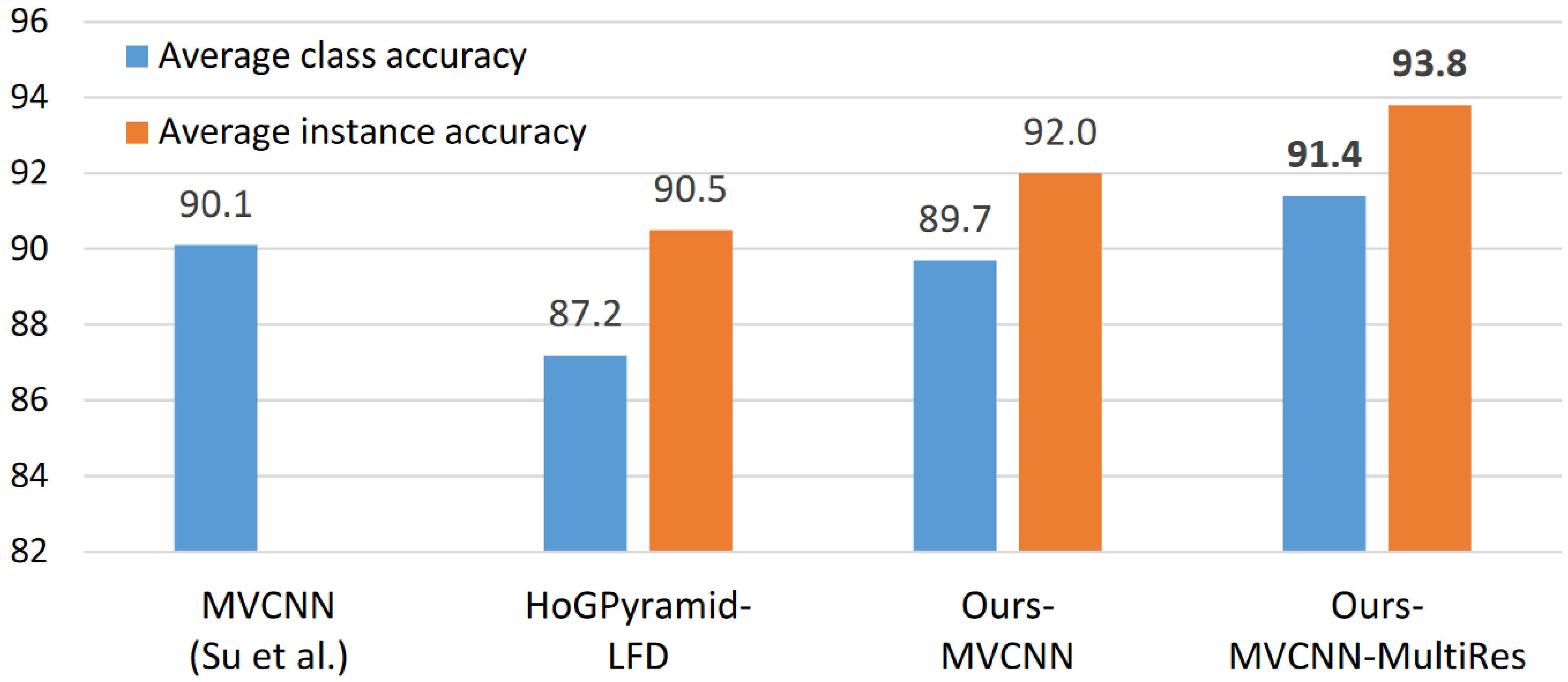
Results – Volumetric CNNs

Class accuracy = average of all (class hits / class preds)
Instance accuracy = total hits / total preds

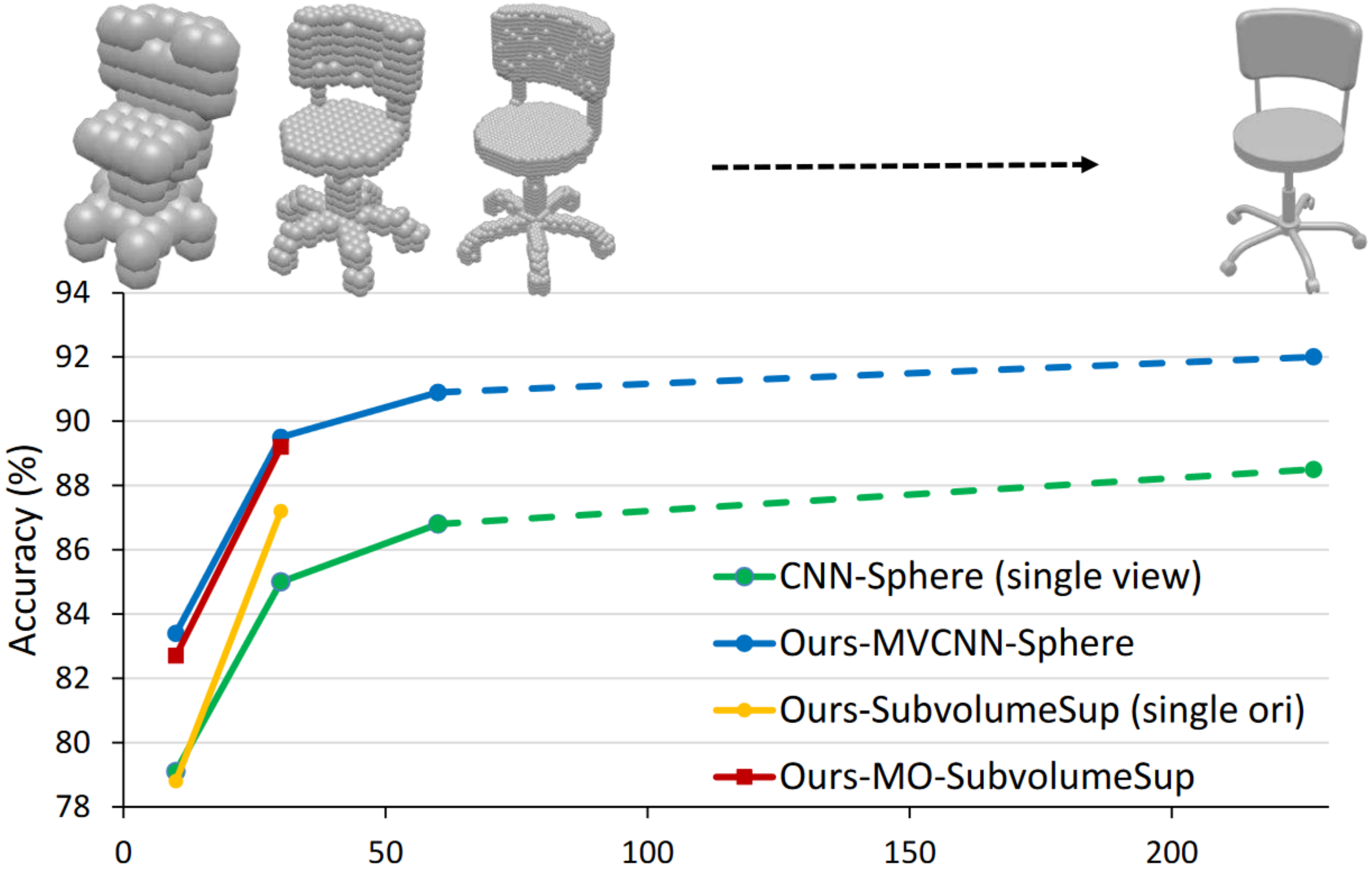
- ModelNet40



Results – Multi-view CNNs



Results



94
92
90
88
86
84
82
80
78
Accuracy (%)
0
50
100
150
200
CNN-Sphere (single view)
Ours-MVCNN-Sphere
Ours-SubvolumeSup (single ori)
Ours-MO-SubvolumeSup

Results

Data Augmentation	Single-Ori	Multi-Ori	Δ
Azimuth rotation (AZ)	84.7	86.1	1.4
AZ + translation	84.8	86.1	1.3
AZ + elevation rotation	83.0	**87.8**	4.8

Results – scanned models

Method	Classification	Retrieval MAP
E2E-[30]	69.6	-
Su-MVCNN [29]	72.4	35.8
Ours-MO-SubvolumeSup	73.3	39.3
Ours-MO-AniProbing	70.8	40.2
Ours-MVCNN-MultiRes	**74.5**	**51.4**