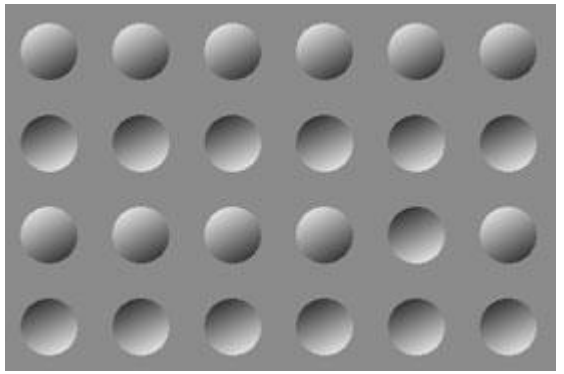
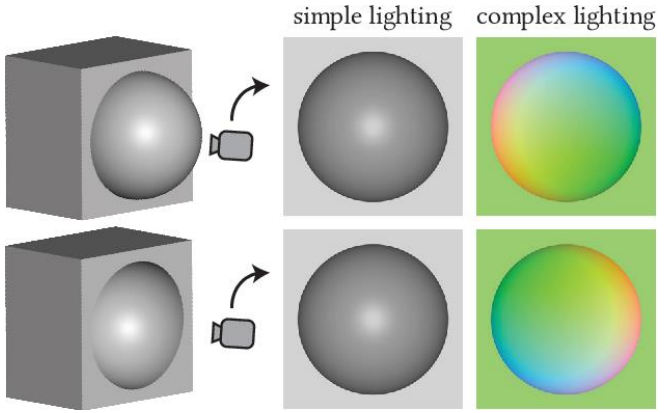
3D Shape Representations

Background

- Deep relationship between human visual perception & 3D representations
  - Marr's "Stages of Vision"
  - Explain visual input in a hierarchy of representations
  - 2D features -> 2.5D sketch -> 3D representation

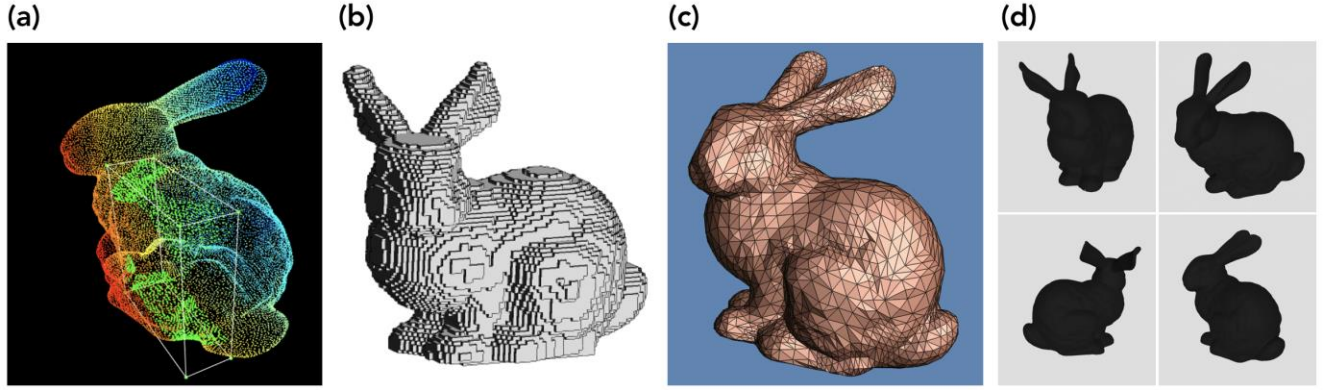


- Ambiguity in visual input
  - Requires strong priors
  - Neural, VAE, Adversarial



- Representation level
  - How to represent the 3D shape?

3D representations



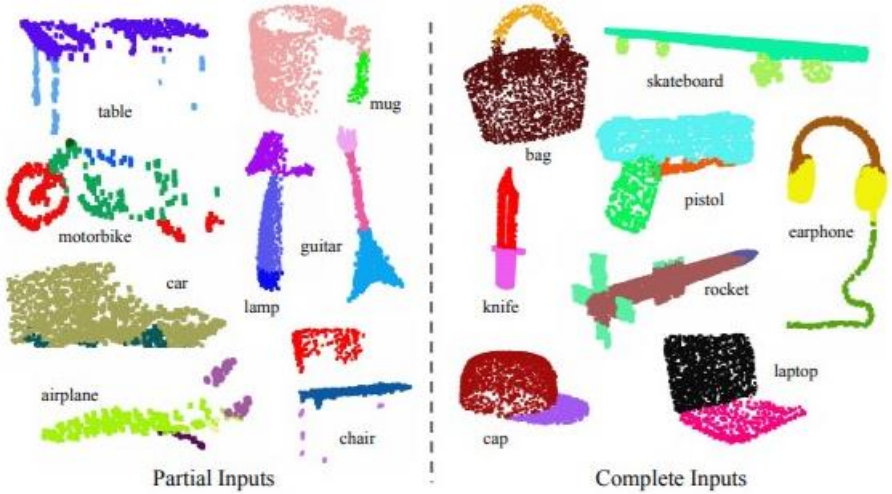
- Point Cloud
  - Very sparse
  - Geometrically ambiguous (no surface orientation)
  - Requires “marching cubes”
- Voxel
  - Not sparse (memory heavy)
  - Huge amount of existing machinery (Convolutions)
- Mesh
  - Very sparse
  - Surface orientation preserved
  - Limited amount of existing machinery (Topologically adaptive graph convs)
- Implicit surface representations
  - New popular method
  - Infinite detail (See also NeRF & SRN for scenes)
  - Requires “marching cubes”

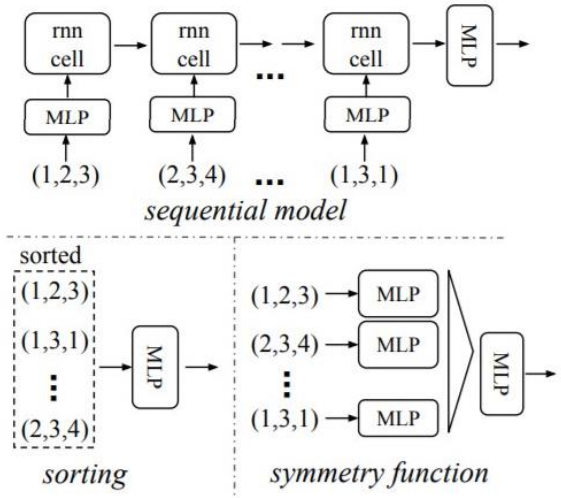
Prior work on learning 3D priors

- Voxel based
  - 3D GAN (NeurIPS 2016 -- Wu, Jiajun and Zhang, Chengkai and Xue, Tianfan and Freeman, William T and Tenenbaum, Joshua B)
  - 3D Descriptor Nets (CVPR 2018 -- Xie, Jianwen and Zheng, Zilong and Gao, Ruiqi and Wang, Wenguan and Zhu, Song-Chun and Nian Wu, Ying)
- Mesh based
  - No work so far on Mesh GANs or Mesh VAEs
- Point Cloud based
  - Latent 3D Points (ICML 2018 -- Achlioptas, Panos and Diamanti, Olga and Mitliagkas, Ioannis and Guibas, Leonidas J)
- Implicit based
  - IM-Net (CVPR 2019 -- Chen, Zhiqin and Zhang, Hao)

PointNet

- Points are scattered
  - Sort the points
  - Data augmentation & train an RNN
  - Symmetric pooling function (their approach)
    - Max/Avg Pool

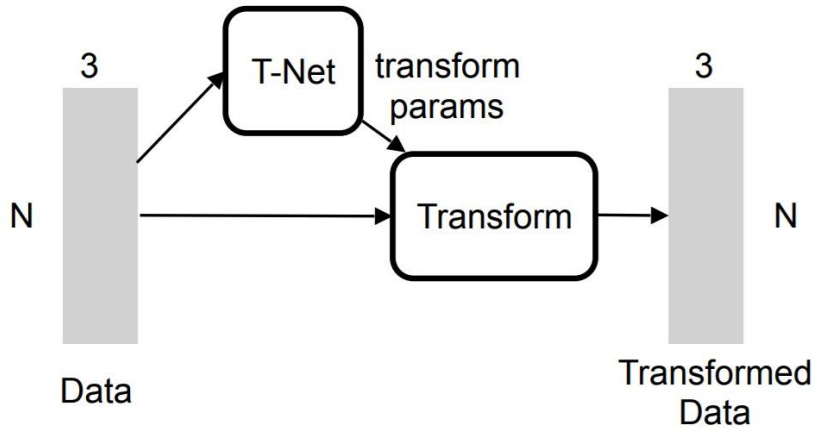


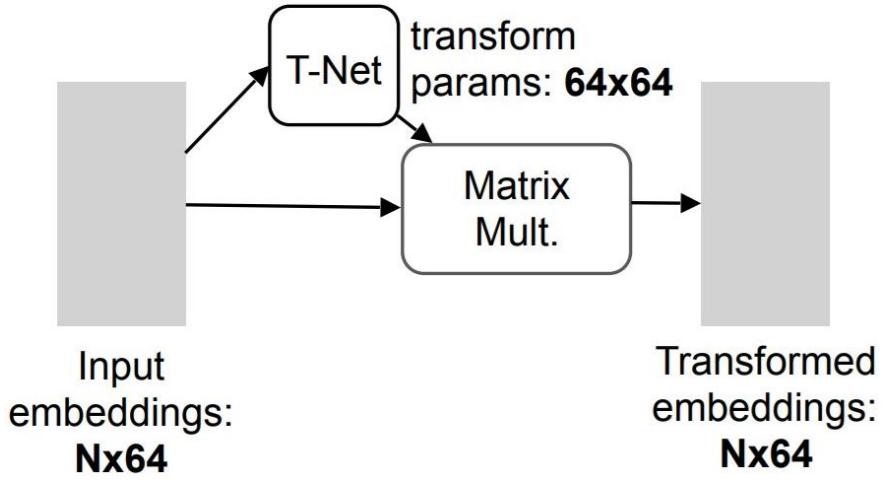


	accuracy
MLP (unsorted input)	24.2
MLP (sorted input)	45.0
LSTM	78.5
Attention sum	83.0
Average pooling	83.8
Max pooling	**87.1**

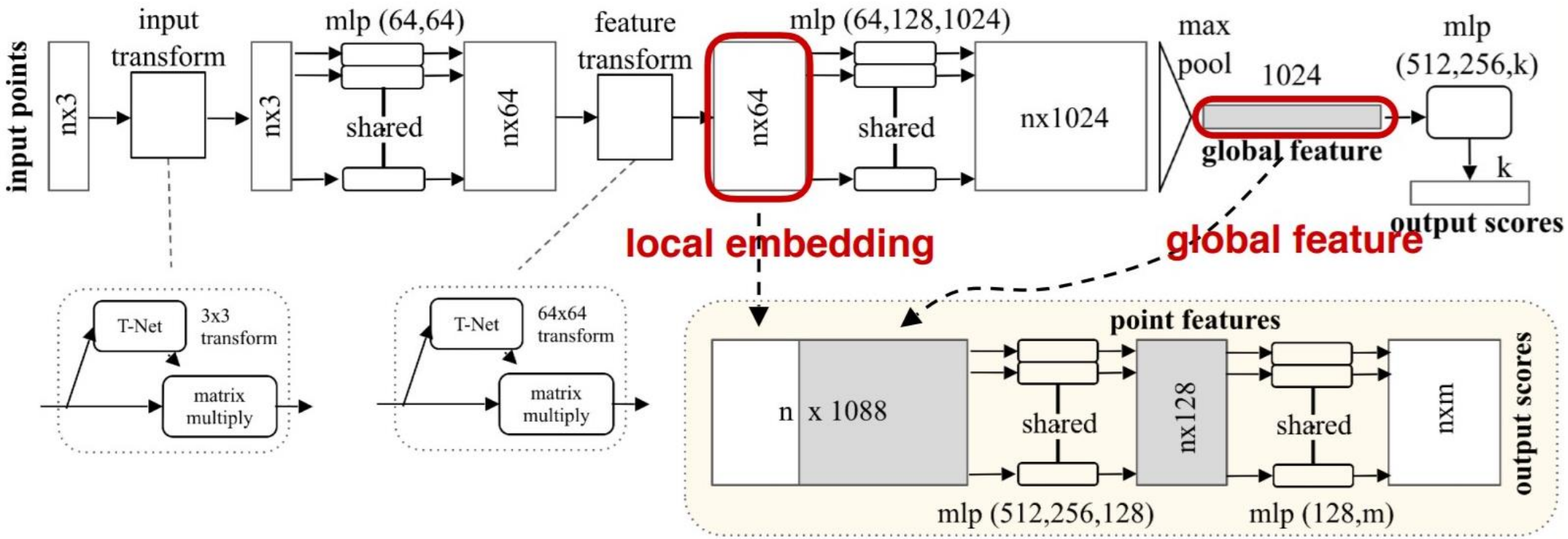
Rotations & translations

- Ideally, geometric transformations should not change the classification of the shape
  - Learn a data-driven spatial transformation to “correct” the input
  - Spatial Transformer Networks (2015)
  - Driven by global feature



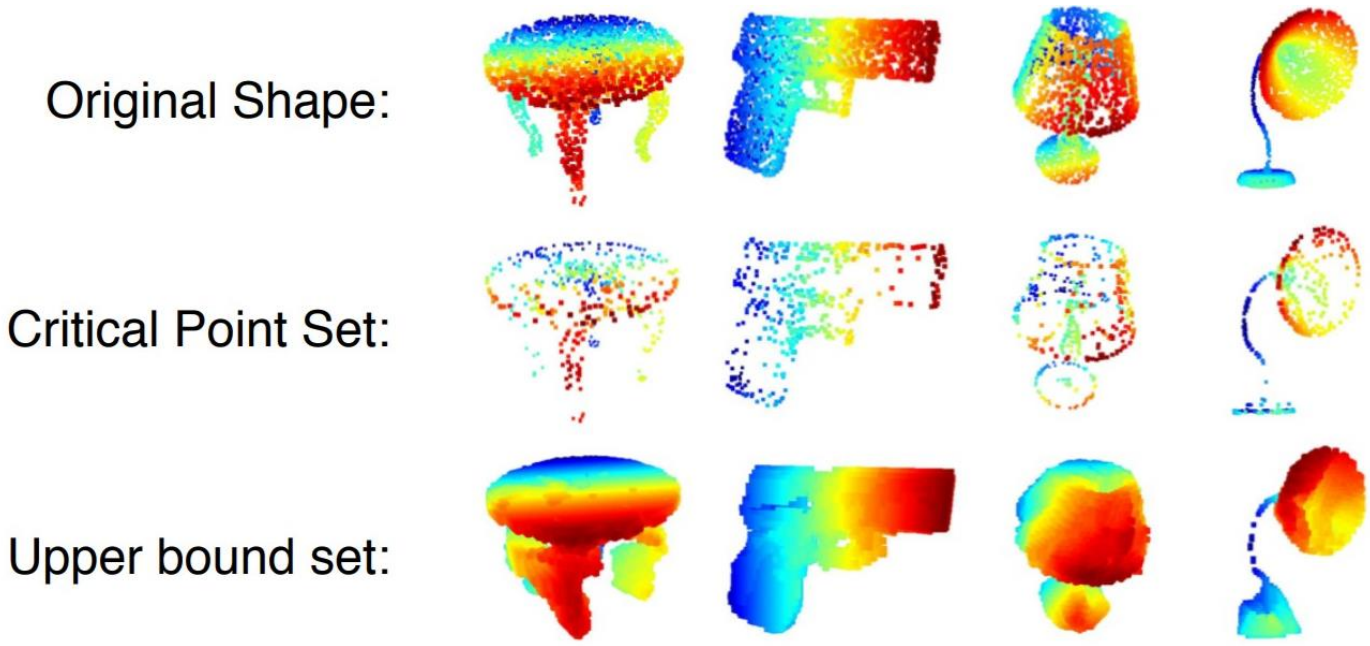


Full Network



Internal Sparsity

- The MAX bottleneck leads to forced sparsity
- Visualization of points that contribute to final MAX vector



Bottleneck tradeoff

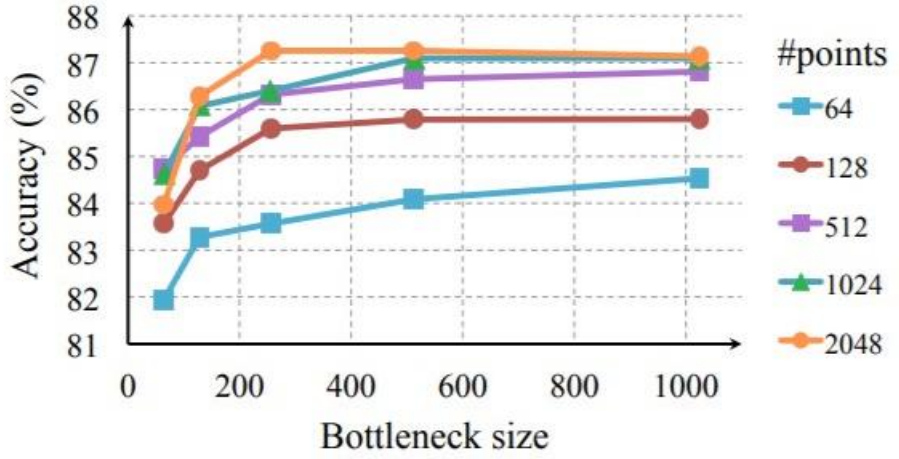
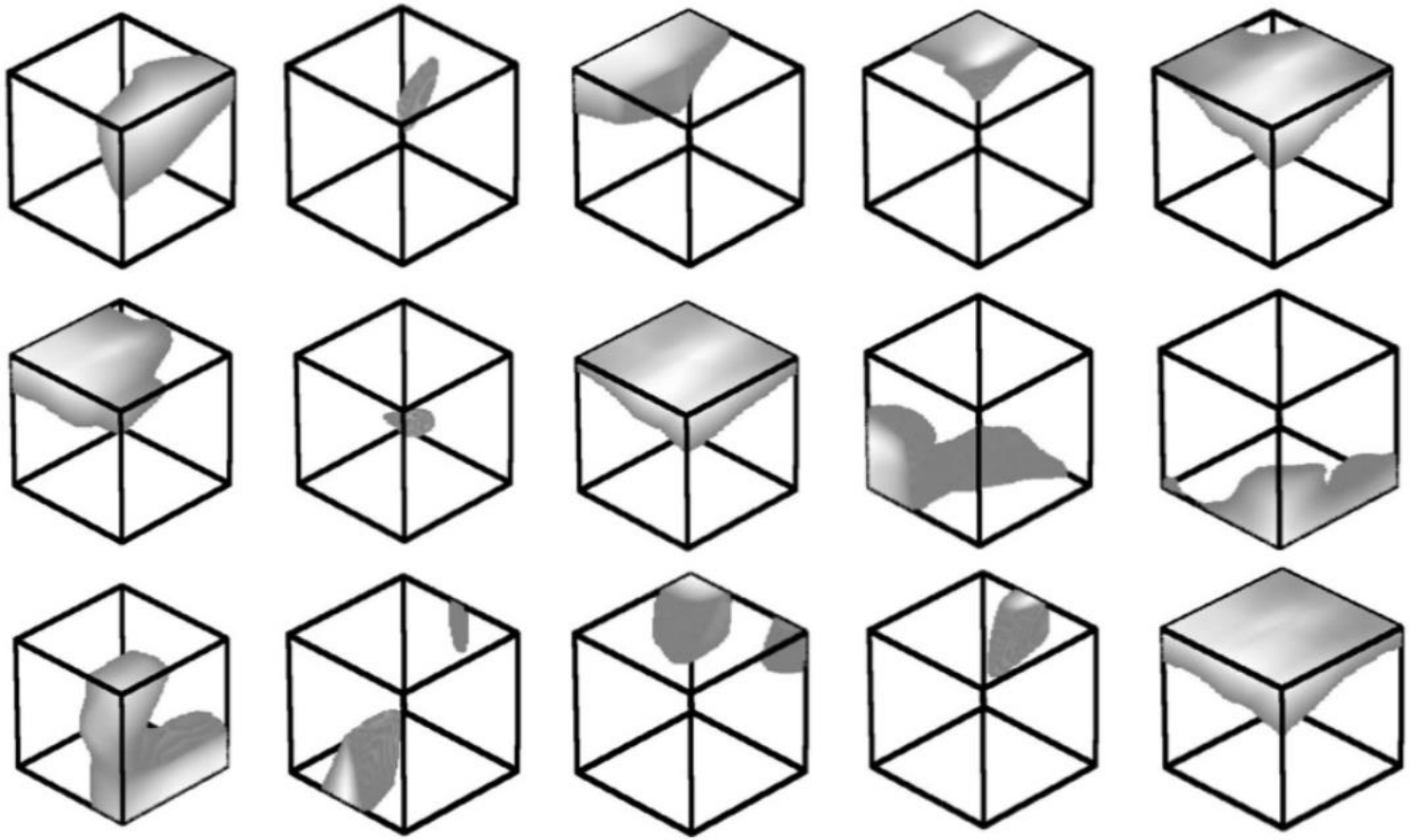


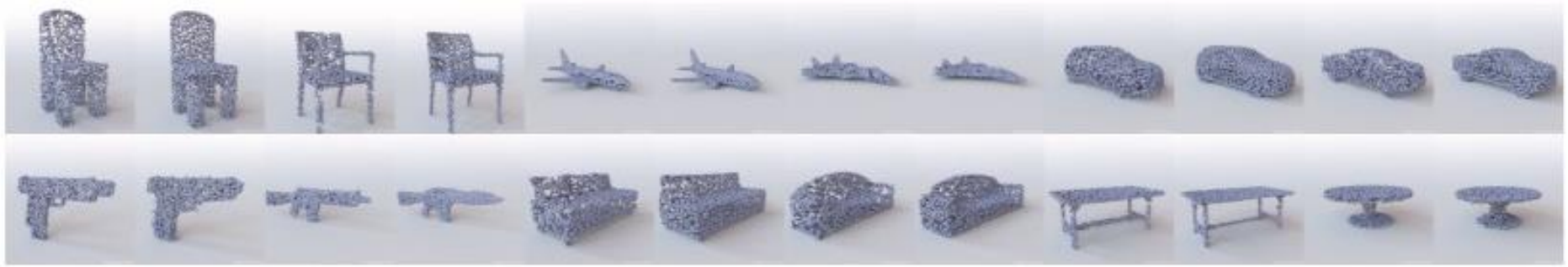
Figure 15. **Effects of bottleneck size and number of input points.** The metric is overall classification accuracy on ModelNet40 test set.

Learned functions from local/global fusion



GAN for point clouds

- Learning Representations and Generative Models for 3D Point Clouds
  - Not to be confused with “Point Cloud GAN”



Generator in latent versus point cloud space

- Adversarial Auto Encoder (latent GAN)
  - First train an Encoder-Decoder Pair
    - Encoder is PointNet based (Max Pool)
    - Decoder generates (2048 x 3 vector, reshaped to [2048, 3])
  - Then train a generator to generate samples from latent space
  - Discriminator also in latent space
- Raw GAN
  - Takes in z (random Normal)
  - Discriminator based on Encoder-Decoder

Training objectives

- Chamfer Distance (CD)
  - Does not require same number of points

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2$$

- Earth Movers Distance (EMD)
  - Optimal transport based
  - Much slower to calculate but better
  - Requires same number of points

$$d_{EMD}(S_1, S_2) = \min_{\phi: S_1 \to S_2} \sum_{x \in S_1} \|x - \phi(x)\|_2$$

where $\phi : S_1 \to S_2$ is a bijection.

Model	Type	JSD	MMD-CD	MMD-EMD	COV-EMD	COV-CD
A	MEM	0.017	0.0018	0.063	78.6	79.4
B	RAW	0.176	0.0020	0.123	19.0	52.3
C	CD	0.048	0.0020	0.079	32.2	59.4
D	EMD	0.030	0.0023	0.069	57.1	59.3
E	EMD	0.022	0.0019	0.066	66.9	67.6
F	GMM	**0.020**	**0.0018**	**0.065**	**67.4**	**68.9**

Table 3. Evaluating 5 generators on the *test* split of the chair dataset on epochs/models selected via minimal JSD on the validation-split. We report: A: sampling-based memorization baseline, B: r-GAN, C: l-GAN (AE-CD), D: l-GAN (AE-EMD) , E: l-WGAN (AE-EMD), F: GMM (AE-EMD).