Sparse 3D Convolutional Neural Networks for Large-Scale Shape Retrieval

Alexandr Notchenko, Ermek Kapushev, Evgeny Burnaev
{avnotchenko,kapushev,burnaevevgeny}@gmail.com

Skoltech
Skolkovo Institute of Science and Technology
&
RUSSIAN ACADEMY OF SCIENCES
INSTITUTE FOR INFORMATION TRANSMISSION PROBLEMS
(Kharkevich Institute)

3D Deep Learning Workshop at NIPS 2016

3D Shape representations

- Meshes
- Point clouds
- Implicit surfaces / potentials
- Voxels
- Set of 2D projections

3D Shape representations

- Meshes
- Point clouds
- Implicit surfaces / potentials
- Voxels
- Set of 2D projections

Irregular size, not clear how to use in NN

Regular size, good to go in CNN

3D Shape representations

- Meshes
- Point clouds
- Implicit surfaces / potentials
- Voxels
- Set of 2D projections

Irregular size, not clear how to use in NN

Regular size, good to go in CNN

Not really 3D, 2D CNNs are powerful enough already

Sparsity of voxel representation

Mean sparsity for all classes of ModelNet40 train dataset at voxel resolution 40 equal to 5.5%.

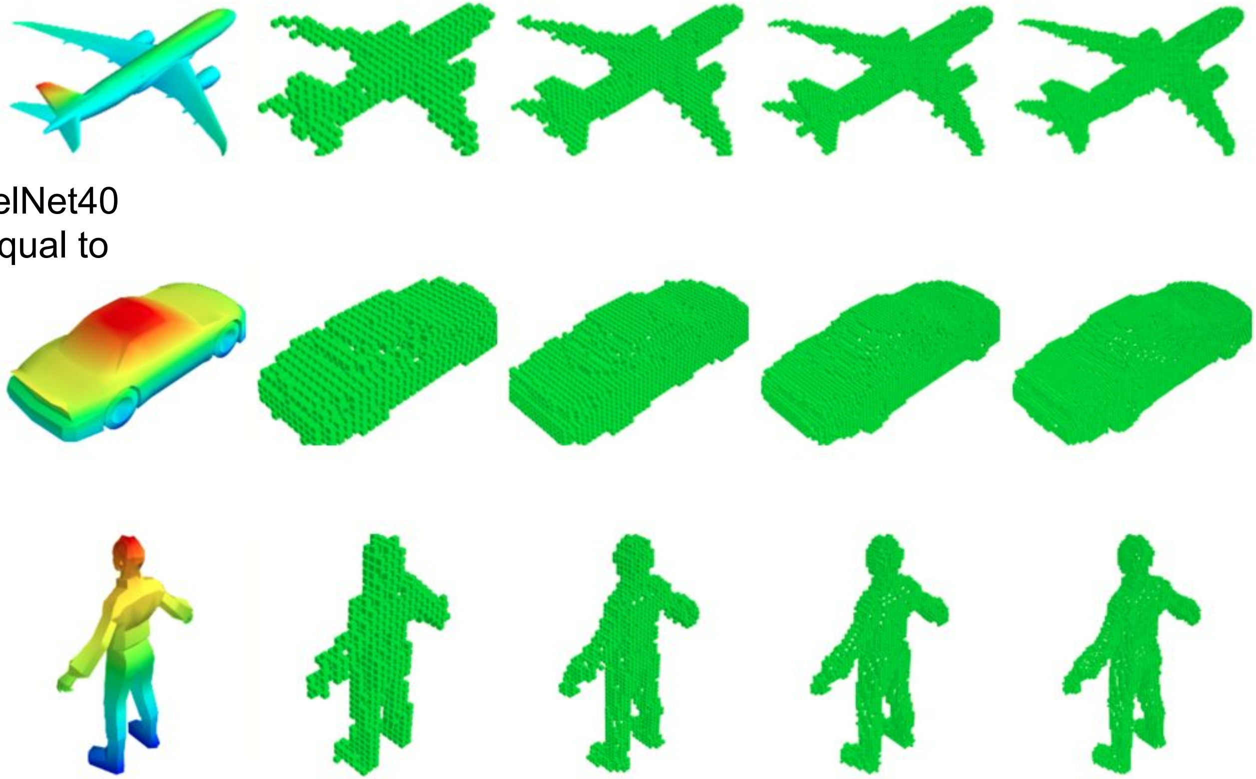
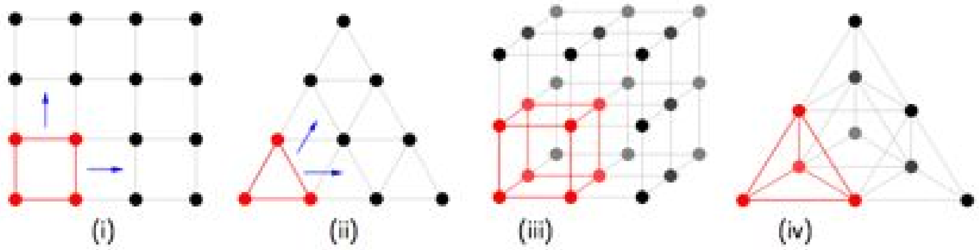


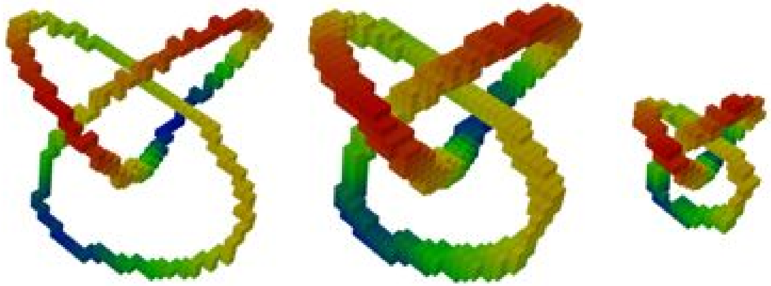
Figure: Examples of some objects voxelizations at different resolutions **30**, **50**, **70**, **100** (from left to right), left-most objects are depicted using original meshes

SparseConvNet

Dr. Benjamin Graham
formerly: Associate Professor at Warwick University
now at Facebook AI Research, Paris Lab



Convolutional filter shapes for different lattices: (i) A 4×4 square grid with a 2×2 convolutional filter. (ii) A triangular grid with size 4, and a triangular filter with size 2. (iii) A $3 \times 3 \times 3$ cubic grid, and a $2 \times 2 \times 2$ filter. (iv) A tetrahedral grid with size 3, and a filter of size 2.



To show how sparsity operates in 3D, consider a trefoil knot has been drawn in the cubic lattice (left). Applying a $2 \times 2 \times 2$ convolution, the number of active/non-zero sites increases (middle). Applying a $2 \times 2 \times 2$ pooling operation reduces the scale, which tends to decrease the number of active sites (right).

http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/graham/bmvc.pdf

PySparseConvNet

- Python wrapper for SparseConvNet, with extended functionality.
- Fixed several Memory issues that prevented large scale learning.
- Made possible to use different loss functions.
- Made layer activations accessible to debugging.
- Interactivity for exploration of models — a way to perform operations step by step, to explore properties of models.

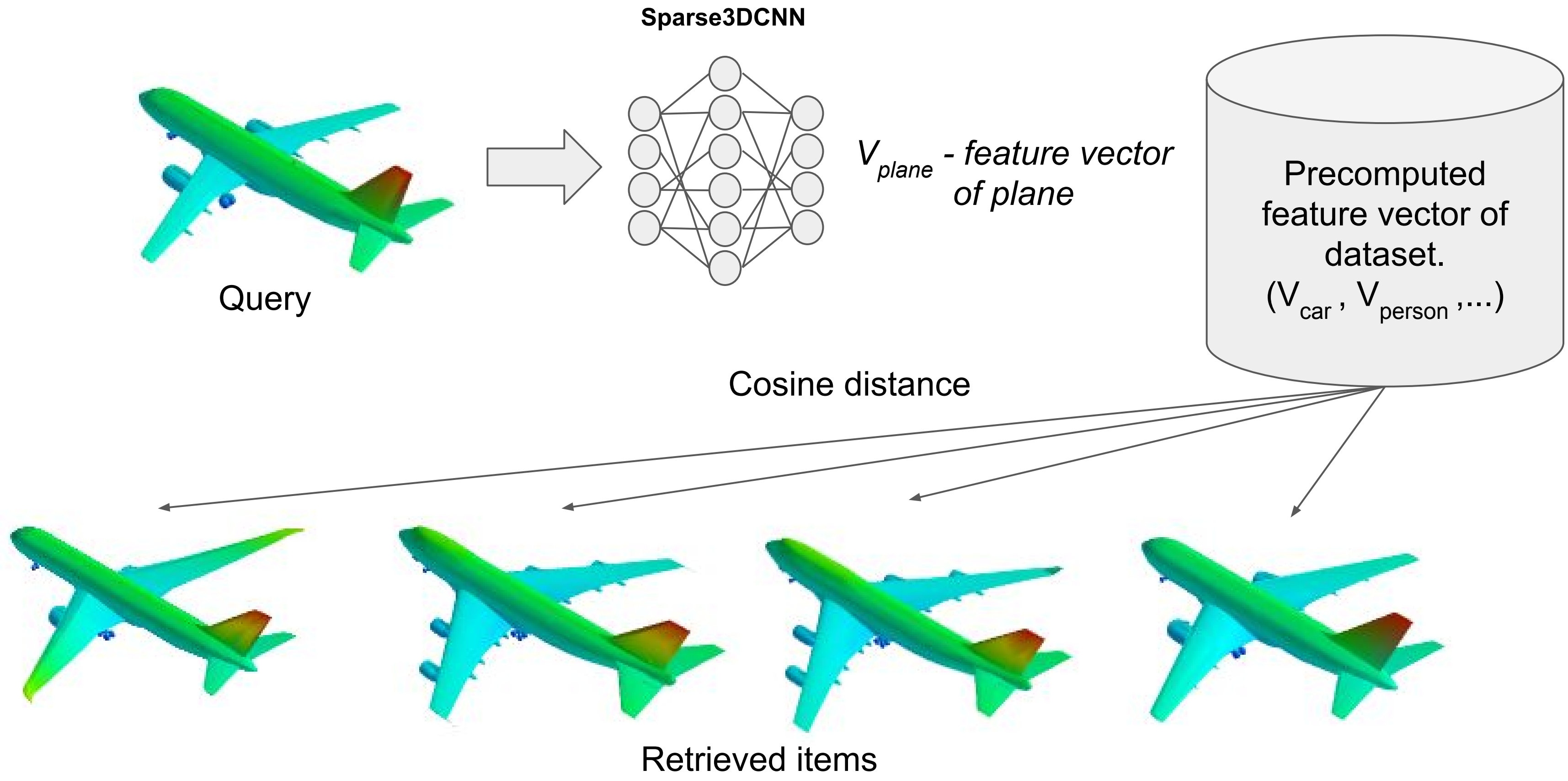
Shape Retrieval

Problem statement

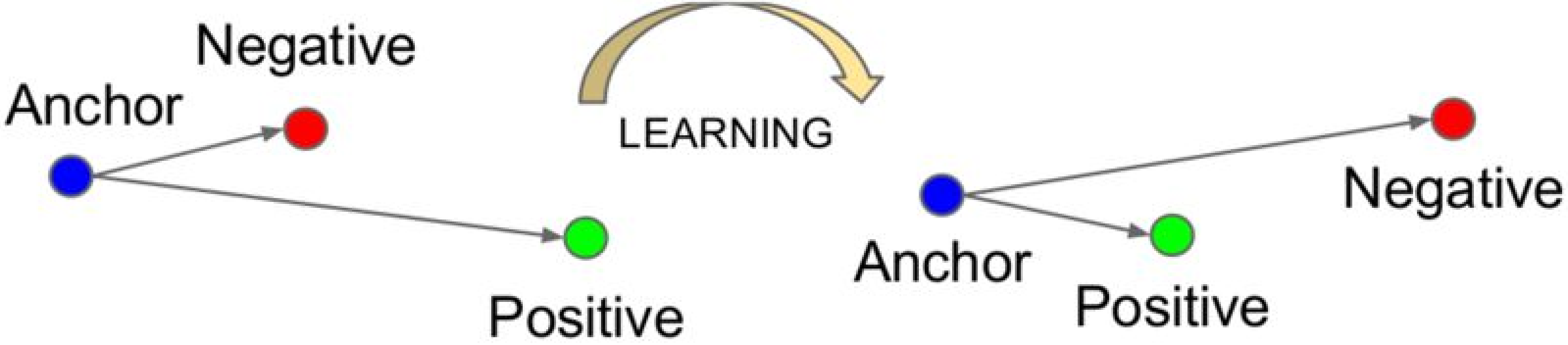
Given a query object find several the most “similar” to the query objects from the given database.

The objects are considered to be similar if they belong to the same category of objects and have similar shapes.

Shape Retrieval



Triplet loss



The representation can be efficiently learned by minimizing triplet loss.

Triplet is a set (*a, p, n*), where

- *a* is an anchor object
- *p* is a positive object - an object that is similar to anchor object
- *n* is a negative object - an object that is not similar to anchor object

$$\lambda(\delta_+, \delta_-) = \max(\mu + \delta_+ - \delta_-)$$

where μ is a margin parameter, δ_+ and δ_- are distances between *p* and *a* and *n* and *a*

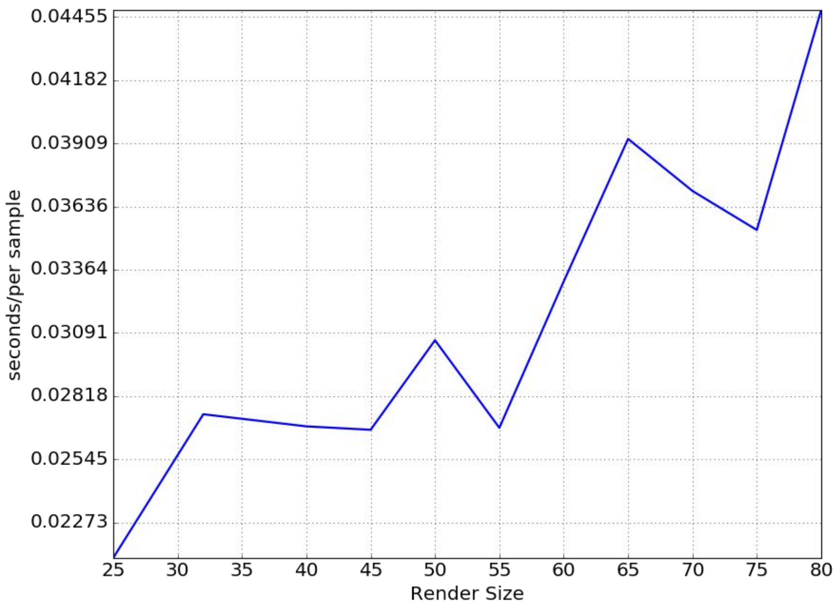
Our approach

- Use very **large** resolutions, and **sparse** representations.
- Used **triplet learning** for 3D shapes.
- Used **Large Scale Shape Datasets** ModelNet.

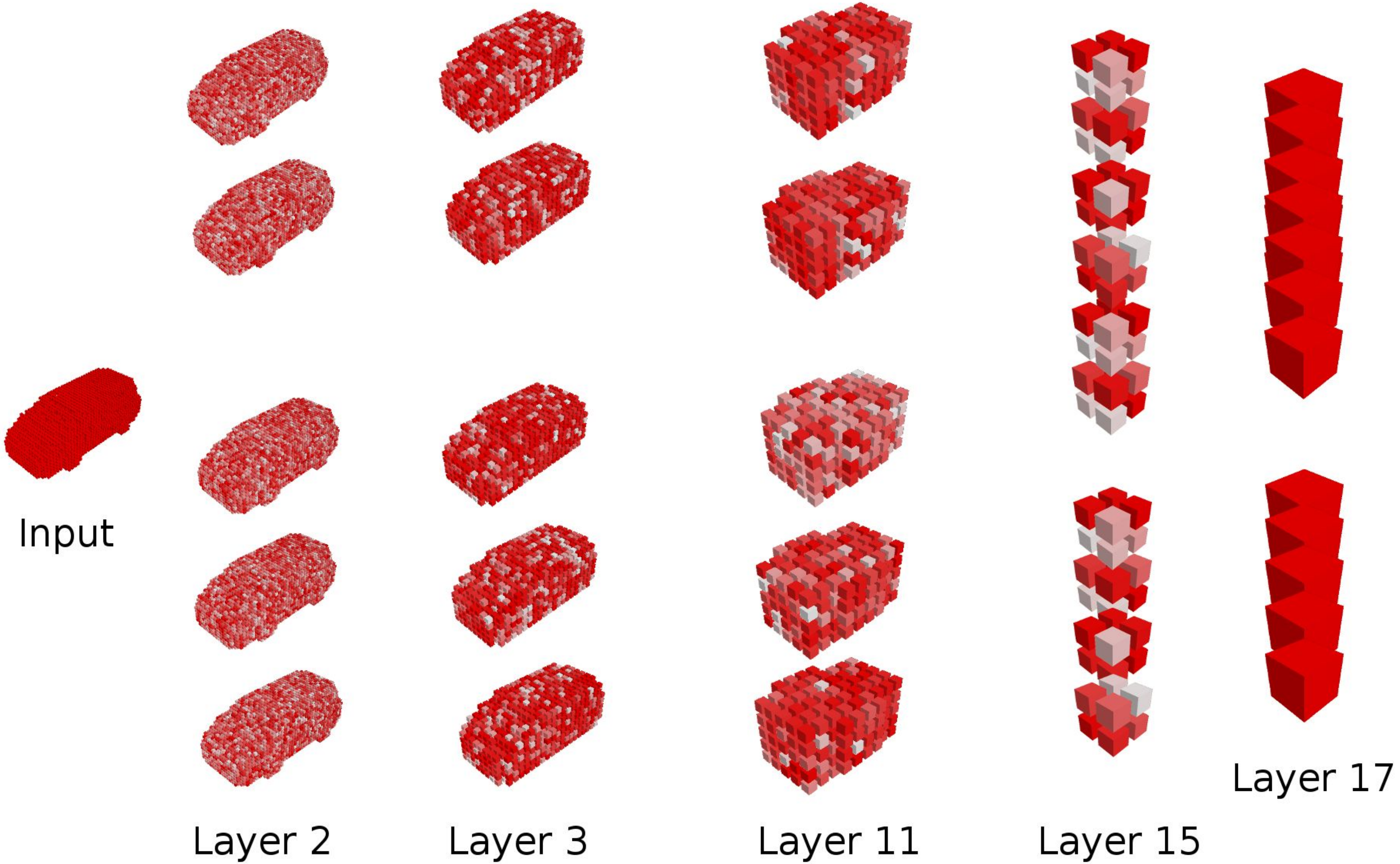
Network description

layer #	layer type	size	stride	channels	spatial size	mean sparsity (%)[1]
0	Data input	-	-	1	126	0.18
1	Sparse Convolutional Layer	2	1	8	125	-
2	Leaky ReLU ($\alpha = 0.33$)	-	-	32	125	0.35
3	Sparse MaxPooling Layer	3	2	32	62	0.69
4	Sparse Convolutional Layer	2	1	256	61	-
5	Leaky ReLU ($\alpha = 0.33$)	-	-	64	61	1.07
6	Sparse MaxPooling Layer	3	2	64	30	1.93
7	Sparse Convolutional Layer	2	1	512	29	-
8	Leaky ReLU ($\alpha = 0.33$)	-	-	96	29	3.26
9	Sparse MaxPooling Layer	3	2	96	14	7.32
10	Sparse Convolutional Layer	2	1	768	13	-
11	Leaky ReLU ($\alpha = 0.33$)	-	-	128	13	15.14
12	Sparse MaxPooling Layer	3	2	128	6	46.30
13	Sparse Convolutional Layer	2	1	1024	5	-
14	Leaky ReLU ($\alpha = 0.33$)	-	-	160	5	97.54
15	Sparse MaxPooling Layer	3	2	160	2	100.00
16	Sparse Convolutional Layer	2	1	1280	1	-
17	Leaky ReLU ($\alpha = 0.33$)	-	-	192	1	100.00

[1]Last column "sparsity" is computed for render size = **40**

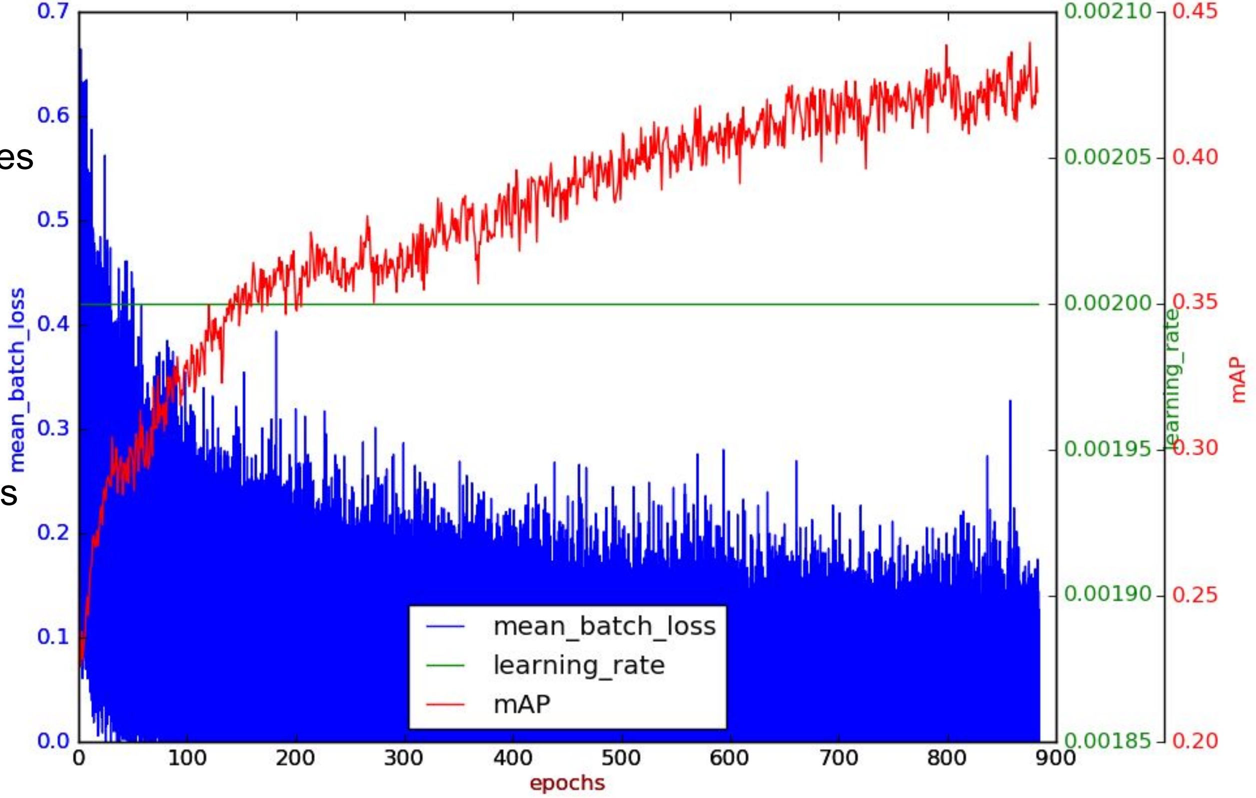


Forward Pass Activations



Training Dynamics

Can finish learning when all samples outside of margin.

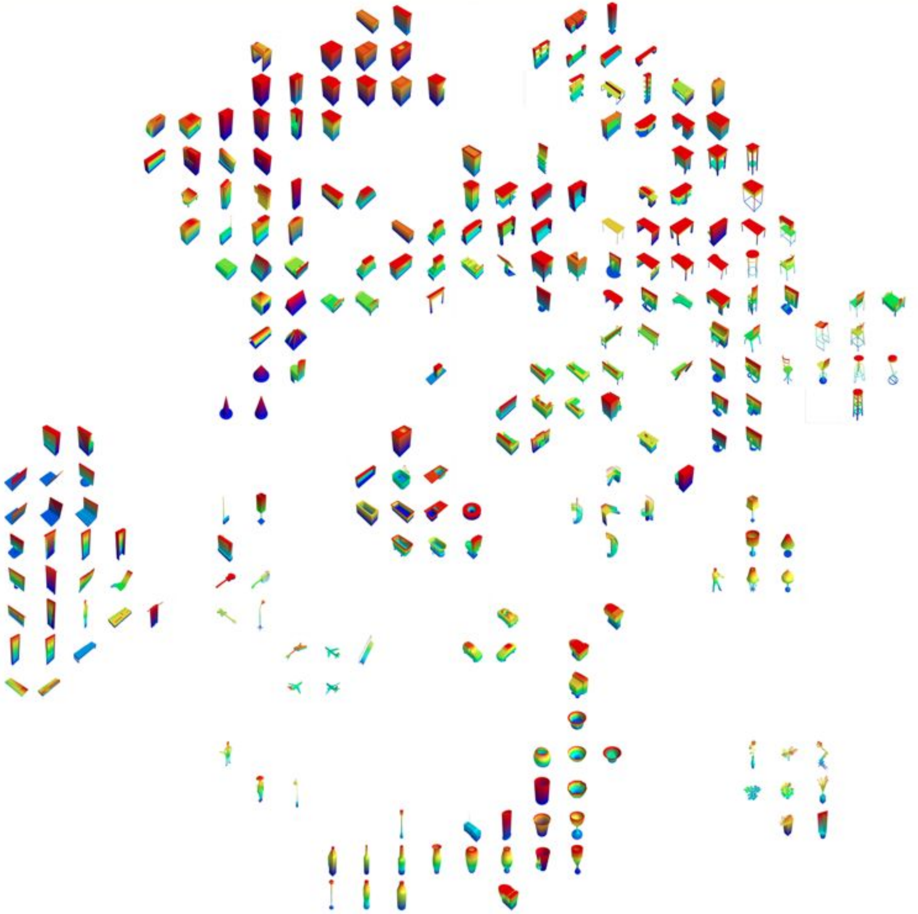


Can finish learning when all samples outside of margin.

Optimisation algorithm:
Nesterov Accelerated Gradient with momentum = 0.99

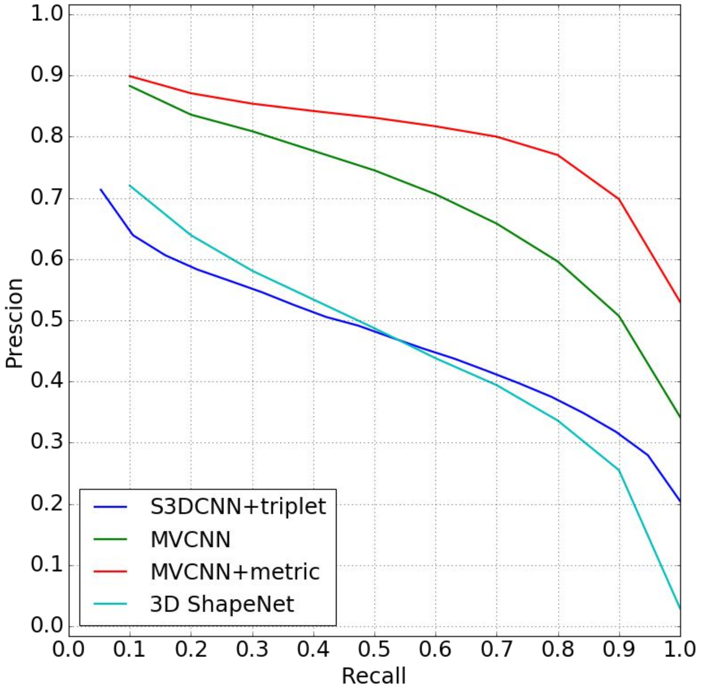
Constant Learning Rate = 0.002

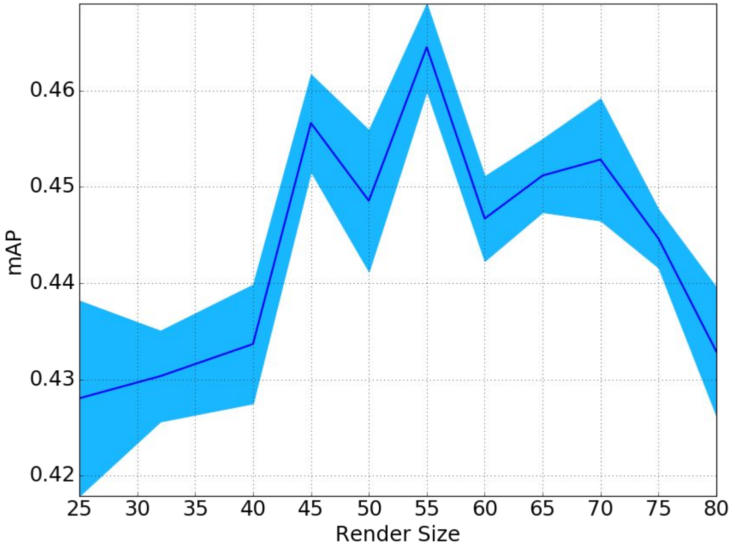
Obligatory t-SNE



Experimental results

method	Classification	Retrieval AUC	Retrieval mAP
3DShapeNet	77.32%	**49.94%**	49.23%
MVCNN	90.10%	---	**80.20%**
3DSCNN	**90.3%**	47.30%	45.16%
S3DCNN + triplet	---	48.81%	46.71%





State-of-the-art

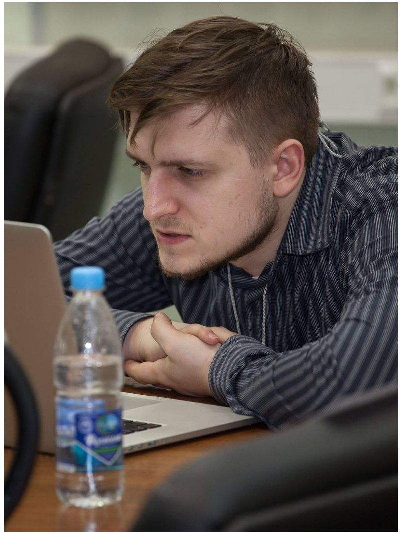
Algorithm	ModelNet40 Classification	ModelNet40 Retrieval (mAP)
Geometry Image [13]	83.9%	51.3%
Set-convolution [11]	90%	
3D-GAN [10]	83.3%	
VRN Ensemble [9]	95.54%	
FusionNet [7]	90.8%	
Pairwise [6]	90.7%	
MVCNN [3]	90.1%	79.5%
GIFT [5]	83.10%	81.94%
VoxNet [2]	83%	
DeepPano [4]	77.63%	76.81%
3DShapeNets [1]	77%	49.2%

[1] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao. 3D ShapeNets: A Deep Representation for Volumetric Shapes. CVPR2015.
[2] D. Maturana and S. Scherer. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition. IROS2015.
[3] H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller. Multi-view Convolutional Neural Networks for 3D Shape Recognition. ICCV2015.
[4] B Shi, S Bai, Z Zhou, X Bai. DeepPano: Deep Panoramic Representation for 3-D Shape Recognition. Signal Processing Letters 2015.
[5] Song Bai, Xiang Bai, Zhichao Zhou, Zhaoxiang Zhang, Longin Jan Latecki. GIFT: A Real-time and Scalable 3D Shape Search Engine. CVPR 2016.
[6] Edward Johns, Stefan Leutenegger and Andrew J. Davison. Pairwise Decomposition of Image Sequences for Active Multi-View Recognition CVPR 2016.
[7] Vishakh Hegde, Reza Zadeh 3D Object Classification Using Multiple Data Representations.
[8] Nima Sedaghat, Mohammadreza Zolfaghari, Thomas Brox Orientation-boosted Voxel Nets for 3D Object Recognition.
[9] Andrew Brock, Theodore Lim, J.M. Ritchie, Nick Weston Generative and Discriminative Voxel Modeling with Convolutional Neural Networks.
[10] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, Joshua B. Tenenbaum. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. NIPS 2016
[11] Siamak Ravanbakhsh, Jeff Schneider, Barnabas Poczos. Deep Learning with sets and point clouds
[12] A. Garcia-Garcia, F. Gomez-Donoso†, J. Garcia-Rodriguez, S. Orts-Escolano, M. Cazorla, J. Azorin-Lopez PointNet: A 3D Convolutional Neural Network for Real-Time Object Class Recognition
[13] Ayan Sinha, Jing Bai, Karthik Ramani Deep Learning 3D Shape Surfaces Using Geometry Images ECCV 2016

Conclussions

- For Modelnet in voxel form - resolution beyond 30^3 doesn’t improves much
- More voxels - change scale of features, probably needs more layers
- Quality of representation depends on RS non smoothly but is maxed around render size of 55

Thank you.



Alexandr Notchenko, Ermek Kapushev, Evgeny Burnaev