FusionNet: 3D Object Classification Using Multiple Data Representations

Vishakh Hegde
Reza Zadeh

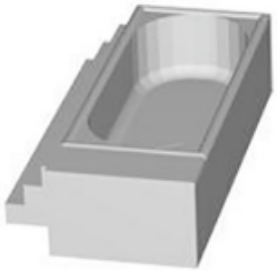
STANFORD
UNIVERSITY

Matroid

Paper: http://matroid.com/papers/fusionnet.pdf Twitter: @Reza_Zadeh

Object recognition

Given 3D model, figure out what it is



» bathtub

Princeton ModelNet

662 object classes, 127,915 CAD models

ModelNet40: 40 class subset

http://modelnet.cs.princeton.edu

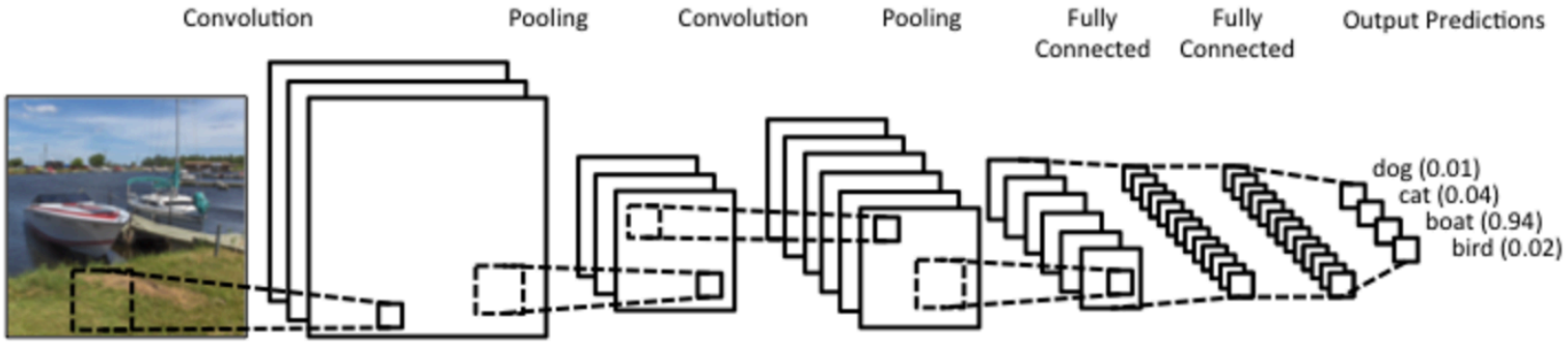
Princeton ModelNet

Problem of input representation

Try using image recognition on projections, but that only goes so far.

From Image Recognition to Object Recognition

Convolutional Network

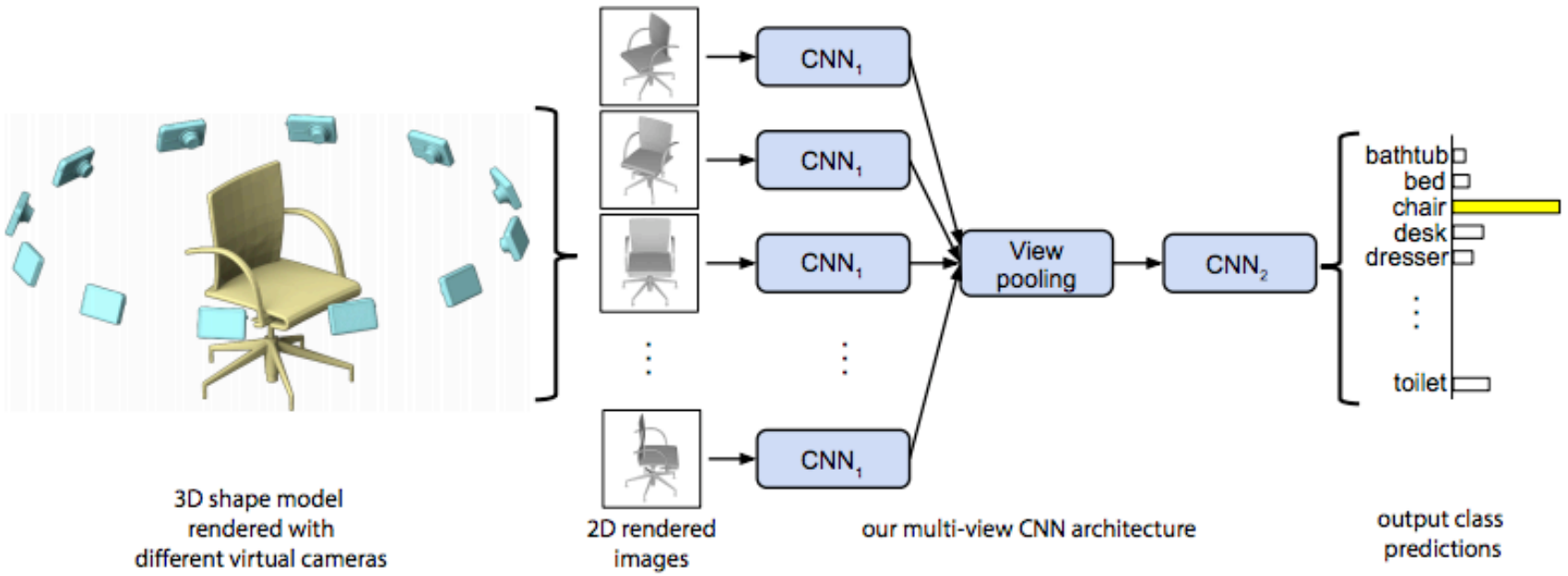


Slide a two-dimensional patch over *pixels*.

How to adapt to three dimensions?

Figure: Google image search for “convolutional neural network”

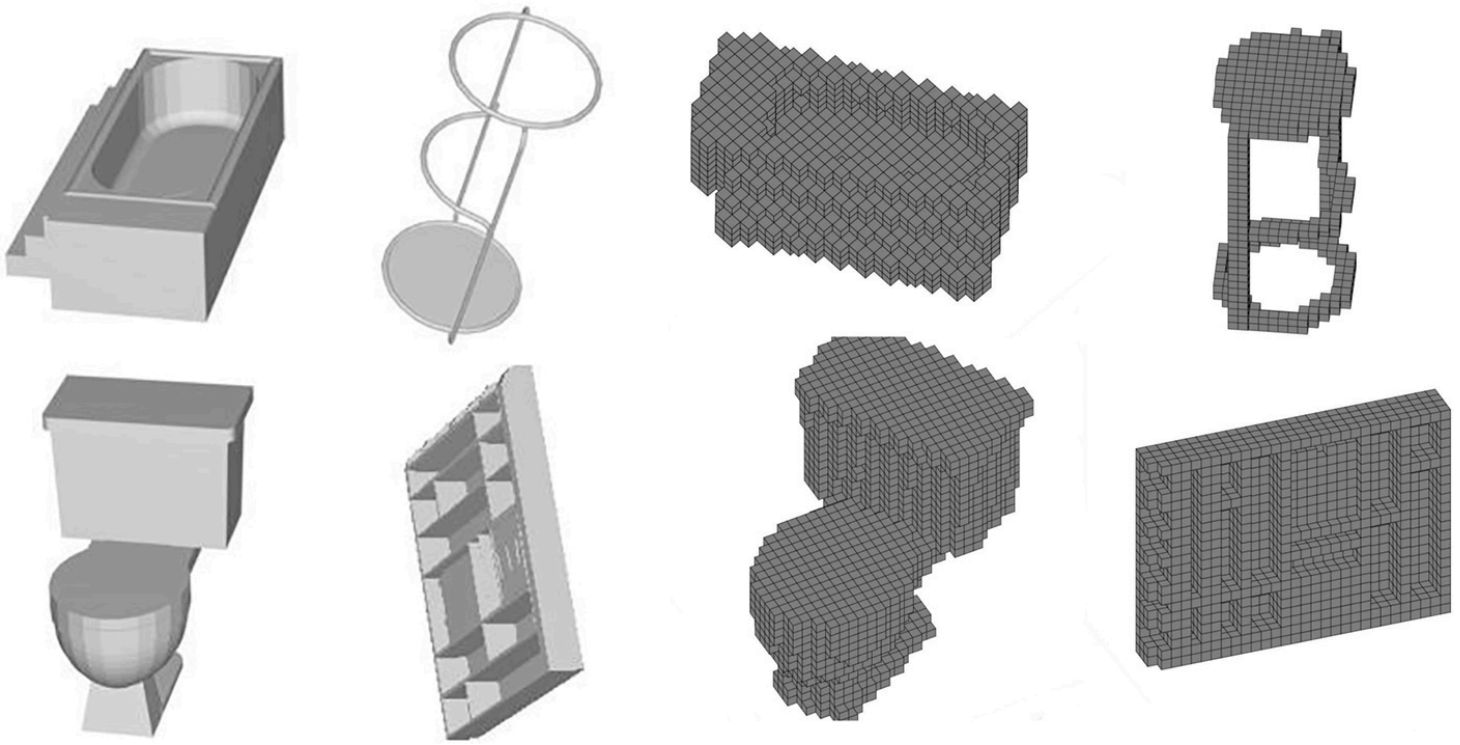
Multi-View CNN



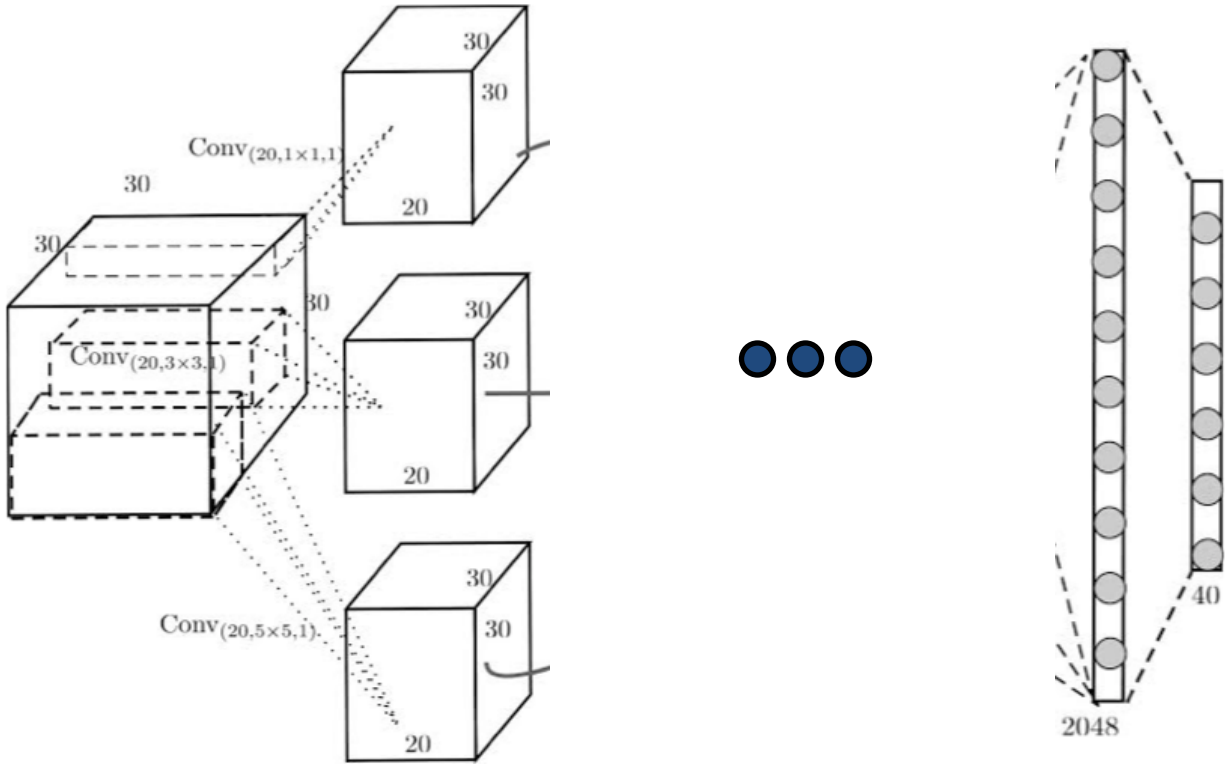
Rotate camera around object

Figure: H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller. Multi-view Convolutional Neural Networks for 3D Shape Recognition. ICCV2015

Representations



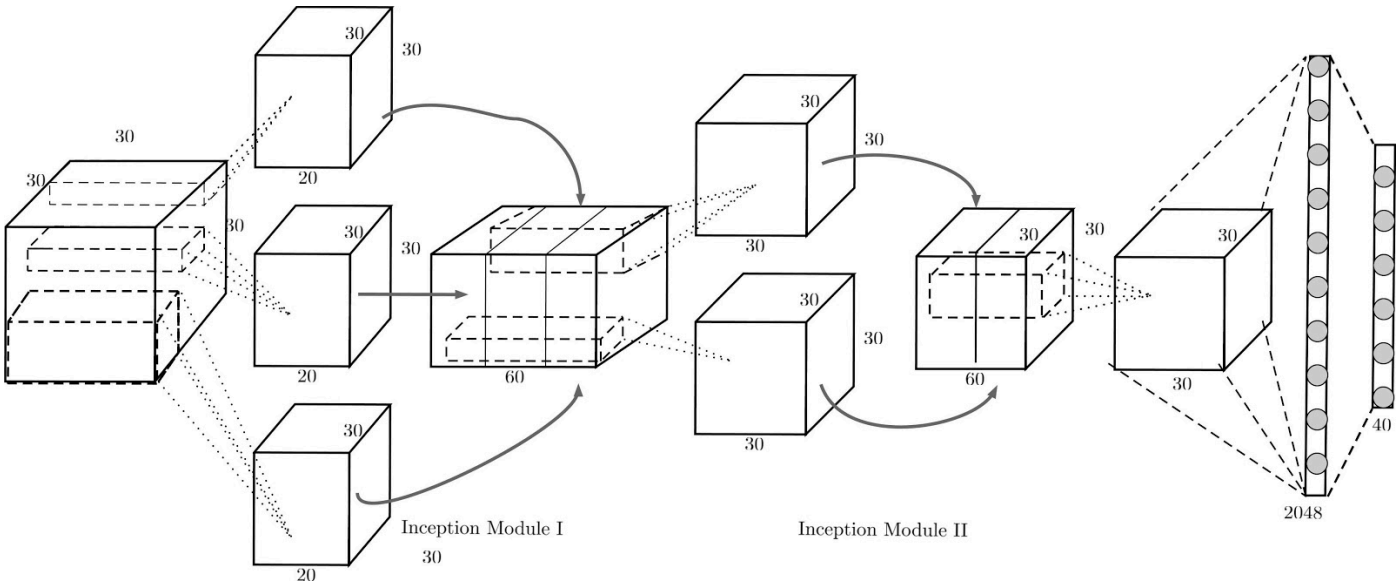
Volumetric (V-CNN)



Simple idea: slide a three-dimensional volume over *voxels*.

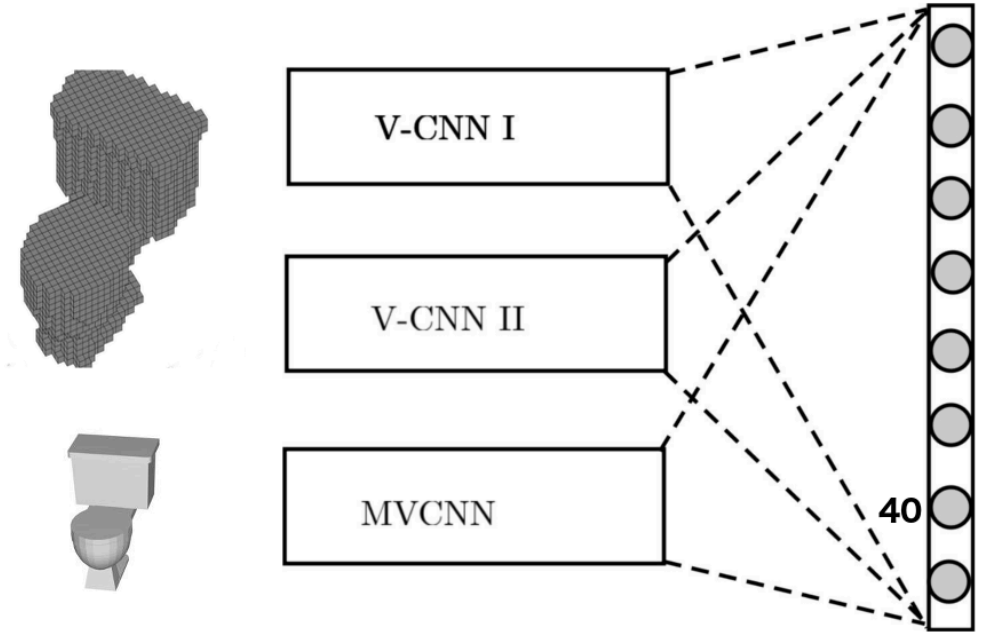
Volumetric CNNs

Use two different Volumetric CNNs (VCNN-I and VCNN-II). Example of one:



FusionNet

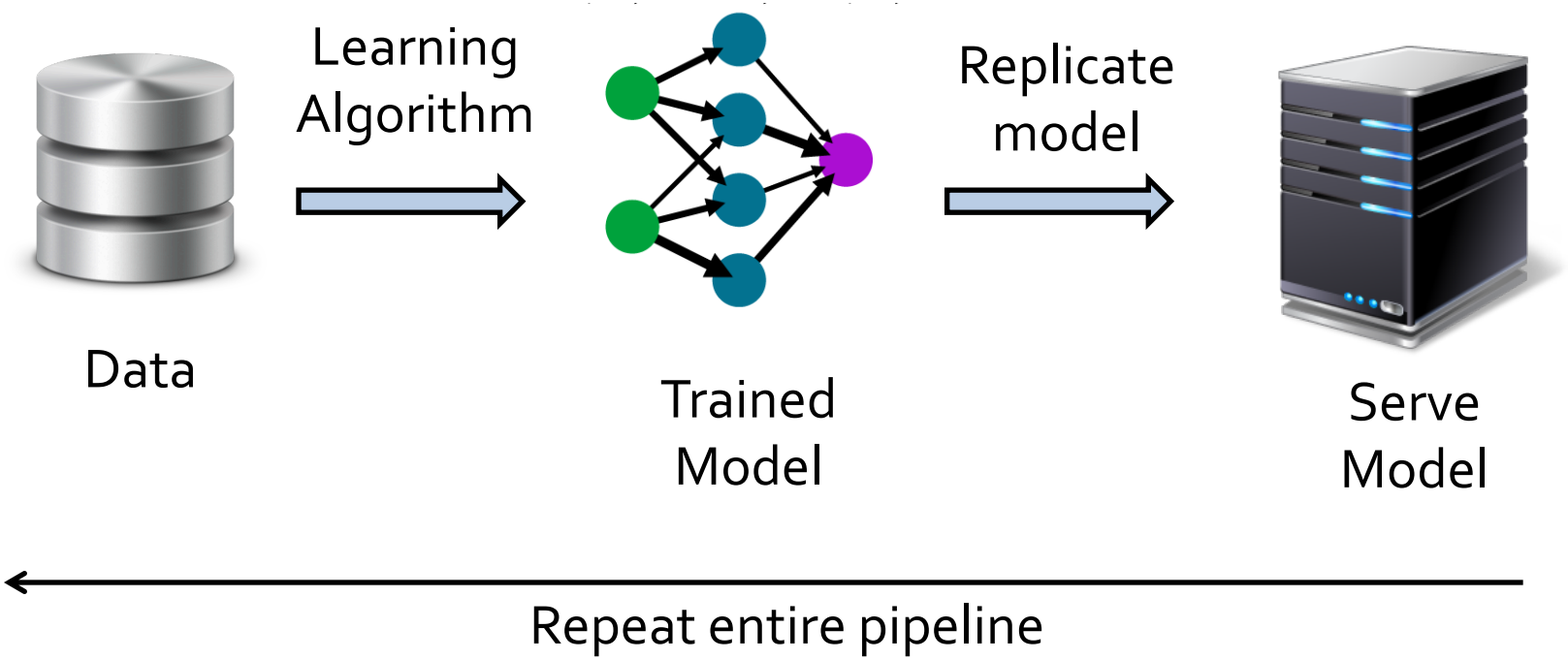
Fusion of two volumetric representation CNNs and one pixel representation CNN



Hyper-parameters tuned on a cluster

http://matroid.com/papers/fusionnet.pdf

Machine Learning Pipeline



Deeper Dive into Networks

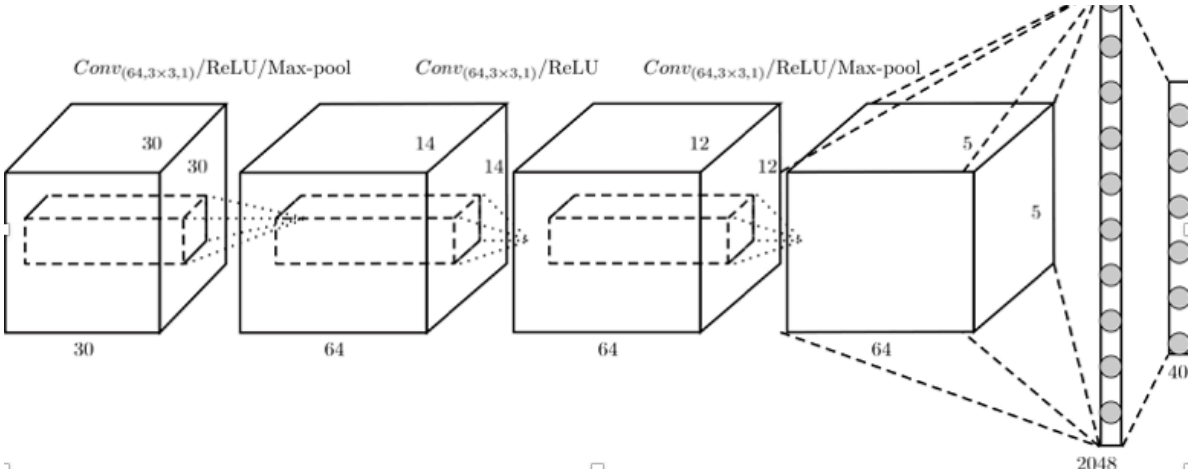
Multi-View CNN

View positions: Corners of icosahedron (20 faces)

Base network: AlexNet (# parameters ~ 60M)

Pre-training on ImageNet, fine-tune last three layers.

VCNN-I

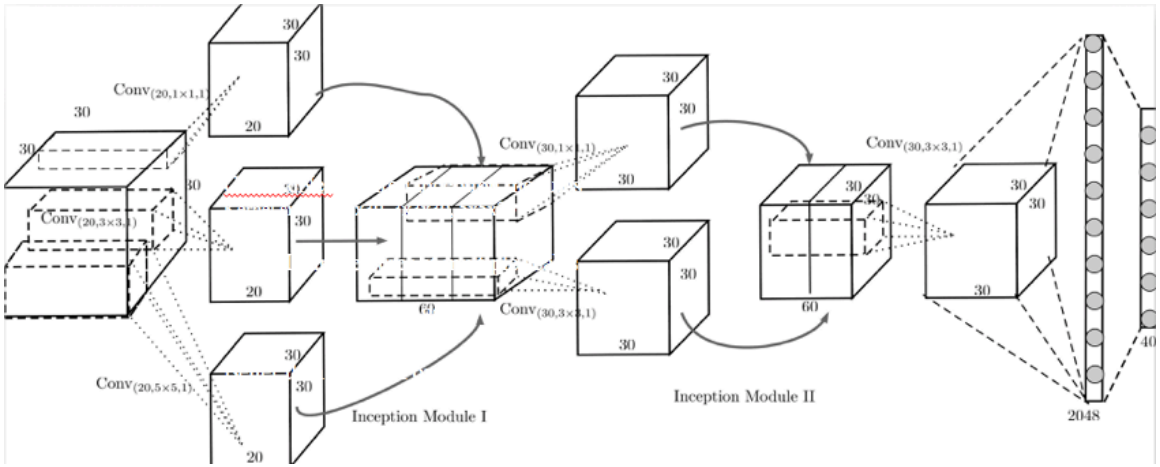


Long kernels learn features spanning the size of the 3D model

Data Augmentation: Gaussian noise added to vertex coordinates in CAD model

Better than VCNN II on: Table, Plant, Bench

VCNN-II



GoogLeNet inspired inception modules
Kernel sizes: 1x1x30, 3x3x30, 5x5x30

Hope: Learn features at multiple scales

Better than VCNN I: Radio, Wardrobe, Xbox

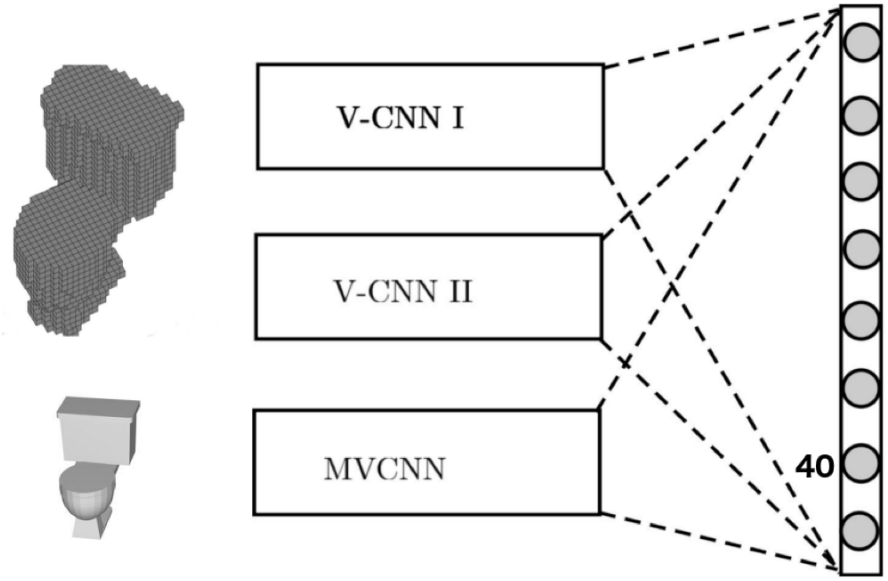
Results

Results

Network	No. Views	Acc. (ModelNet10)	Acc. (ModelNet40)
Volumetric CNN (V-CNN) 1	60	91.48	82.41
V-CNN I*	60	–	80.63
V-CNN II	60	90.22	82.11
V-CNN II + V-CNN II	60	90.32	83.31
V-CNN I + V-CNN II	60	91.95	83.78
AlexNet (random) MV-CNN	20	–	85.82
AlexNet (FT) MV-CNN	20	92.69	86.92
AlexNet (FT) MV-CNN + V-CNN I	20, 60	93.04	88.50
FusionNet	**20, 60**	**93.11**	**90.80**

Table 3: Classification accuracy for individual models. FT = Fine Tuning. * = Data augmentation with Gaussian noise. random = Random initialization of all network weights. FusionNet achieves the best performance.

FusionNet



Algorithm	ModelNet40 Classification (Accuracy)	ModelNet40 Retrieval (mAP)	ModelNet10 Classification (Accuracy)	ModelNet10 Retrieval (mAP)
FusionNet [7]	90.8%		93.11%	
Pairwise [6]	90.7%		92.8%	
MVCNN [3]	90.1%	79.5%		
GIFT [5]	83.10%	81.94%	92.35%	91.12%
VoxNet [2]	83%		92%	
DeepPano [4]	77.63%	76.81%	85.45%	84.18%
3DShapeNets [1]	77%	49.2%	83.5%	68.3%

At the time of submission (July 17th 2016)

ModelNet now

Algorithm	ModelNet40 Classification (Accuracy)	ModelNet40 Retrieval (mAP)	ModelNet10 Classification (Accuracy)	ModelNet10 Retrieval (mAP)
Geometry Image [13]	83.9%	51.3%	88.4%	74.9%
Set-convolution [11]	90%			
PointNet [12]			77.6%	
3D-GAN [10]	83.3%		91.0%	
VRN Ensemble [9]	95.54%		97.14%	
ORION [8]			93.8%	
FusionNet [7]	90.8%		93.11%	
Pairwise [6]	90.7%		92.8%	
MVCNN [3]	90.1%	79.5%		
GIFT [5]	83.10%	81.94%	92.35%	91.12%
VoxNet [2]	83%		92%	
DeepPano [4]	77.63%	76.81%	85.45%	84.18%
3DShapeNets [1]	77%	49.2%	83.5%	68.3%

Recent (December 5th 2016)

Conclusions

3D convolutions on different kernel sizes help

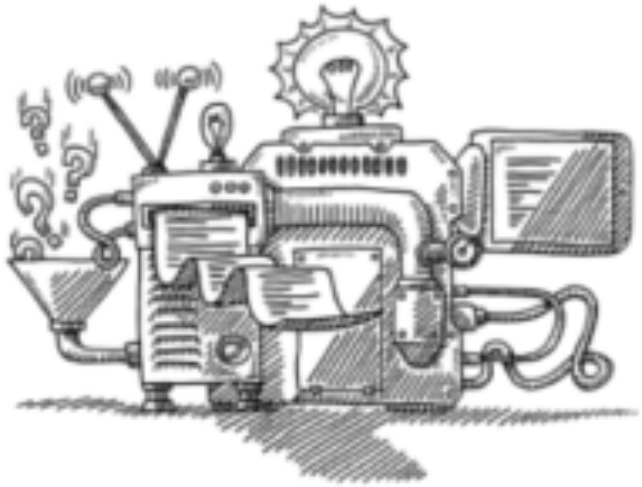
Combination MVCNN + VCNN helps

Hyper-parameter tuning helps

DEEM workshop

Held in conjunction with SIGMOD/PODS

May 14th, 2017 – Submissions open!



DEEM

1st Workshop on Data Management for End-to-End Machine Learning

Matroid

Thank you!

FusionNet paper

http://matroid.com/papers/fusionnet.pdf