PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation [4]

David Stutz

June 1-2, 2017

Motivation

Which is the best 3D representation for deep learning?

- Voxel grids,
- triangular meshes,
- point clouds,
- projections ...

How to efficiently train deep models on 3D data?

- Efficient convolutions (e.g. [2]),
- efficient data structures (e.g. OctNets [12, 11, 16, 17]) ...

Related Work

"Traditional" 3D CNNs
([9, 14, 1, 6, 13, 7, 10, 18, 5] ...)

- Meshes or point clouds are voxelized;
- efficient convolution possible [2].

Related Work

"Traditional" 3D CNNs
([9, 14, 1, 6, 13, 7, 10, 18, 5] ...)

- Meshes or point clouds are voxelized;
- efficient convolution possible [2].

Efficient Data Structure
([12, 11, 16, 17])

- Octree-based,
- learning structure difficult.

Related Work

“Traditional” 3D CNNs
([9, 14, 1, 6, 13, 7, 10, 18, 5] ...)

- Meshes or point clouds are voxelized;
- efficient convolution possible [2].

Efficient Data Structure
([12, 11, 16, 17])

- Octree-based,
- learning structure difficult.

“Manifold” CNNs
([8] – more?)

- Extends Euclidean CNNs to Riemannian manifolds.

Related Work

“Traditional” 3D CNNs
([9, 14, 1, 6, 13, 7, 10, 18, 5] ...)

- Meshes or point clouds are voxelized;
- efficient convolution possible [2].

Efficient Data Structure
([12, 11, 16, 17])

- Octree-based,
- learning structure difficult.

“Manifold” CNNs
([8] – more?)

- Extends Euclidean CNNs to Riemannian manifolds.

Point Clouds
([3, 4])

- Sample points from meshes;
- operate directly on unordered point sets.

Point Set Properties

Consider point set $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^3$:

- unordered: in contrast to voxel grids no inherent order – invariance to $n!$ permutations required;
- distance metric: interactions between points through a distance on $\mathbb{R}^3$;
- invariance: invariance to 3D transformations required.

Unordered Point Sets

Goal: invariance to $n!$ permutations.

Idea:

$$f(\{x_1, \ldots, x_n\}) \approx g(h(x_1), \ldots, h(x_n))$$

symmetric function

feature transform

In practice:

- $h(x)$ modeled as multi-layer perceptron;
- $g = \max$.

Unordered Point Sets (cont'd)

Theorem (informal): any continuous function can be approximated as

$$f(\{x_1, \ldots, x_n\}) \approx \max(h(x_1), \ldots, h(x_n)).$$

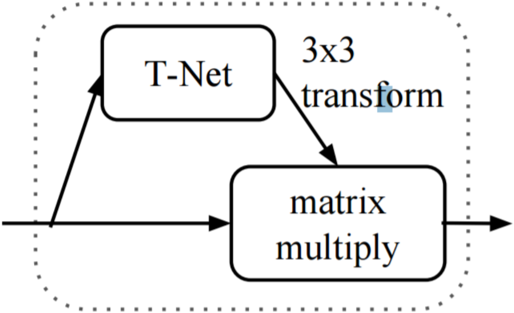
Further result: "Intuitively, our network learns to summarize shape by a sparse set of key points".

- It exists a critical set of points.

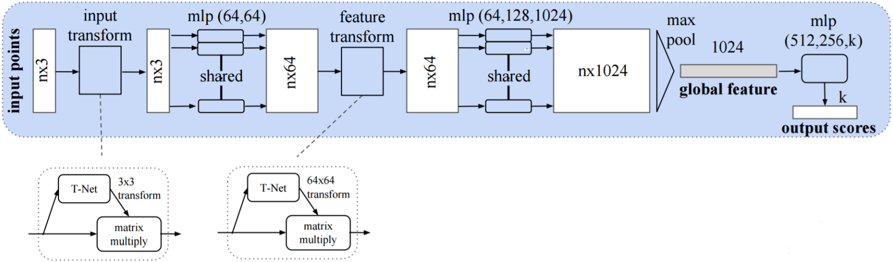
Invariances

Goal: learn invariances regarding rotations, translations, noise ...

Idea: joint alignment network ...



Shape Classification



Shape Classification (cont'd)

Experimental setup:

- ModelNet40 [18];
- **1024** sampled points;
- normalized to unit sphere;
- randomly rotated;
- with Gaussian noise ($\sigma = 0.02$).

Shape Classification (cont'd)

		overall accuracy	average class accuracy
VoxNet [9]	volumes	83	85.9
Subvolume [10]	volumes	86	89.2
MVCNN [15]	images	–	90.1
PointNet	points	89.2	86.2

Table: Accuracy on ModelNet40.

numbers hard to trace in literature

Shape Classification (cont'd)

green one should be 89.2

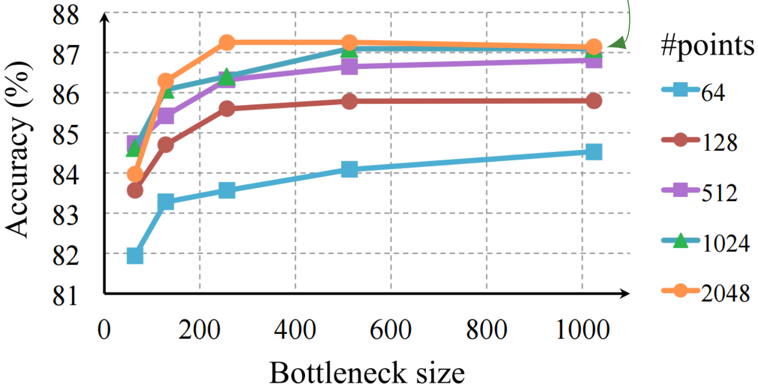


Figure: Overall accuracy on ModelNet40.

Critical Sets

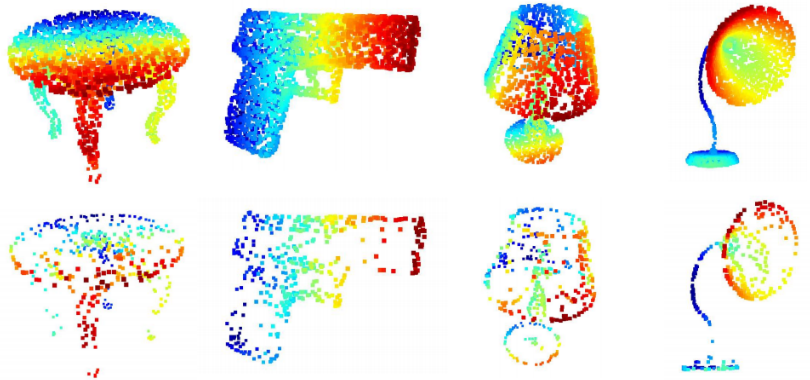


Figure: Critical point sets, i.e. points that contribute to the max-pooling features.

My 2 Cents ...

Revisiting properties of point sets:

- unordered;
- interactions between points;
- invariances;

My 2 Cents ...

Revisiting properties of point sets:

- ~~unordered~~ symmetric function ✓;
- interactions between points;
- invariances;

My 2 Cents ...

Revisiting properties of point sets:

- ~~unordered~~ symmetric function ✓;
- interactions between points?;
- invariances;

My 2 Cents ...

Revisiting properties of point sets:

- ~~unordered~~ symmetric function ✓;
- interactions between points?;
- ~~invariances~~ alignment network ✓;

My 2 Cents ... (cont'd)

For meshes:

- additional local structure (faces) available;
- invariance to points sampled from the same mesh;
- no notion of "surface" without point normals.

Conclusion

PointNet = (shared) multilayer perceptrons on individual points + max pooling over points.

Interesting alternative to OctNets to apply deep learning on sparse 3D data.

- Still some open questions to investigate ...

References I

- Hao Chen, Qi Dou, Lequan Yu, and Pheng-Ann Heng.
  Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation.
  CoRR, abs/1608.05895, 2016.

- Martin Engelcke, Dushyant Rao, Dominic Zeng Wang, Chi Hay Tong, and Ingmar Posner.
  Vote3deep: Fast object detection in 3d point clouds using efficient convolutional neural networks.

References II

CoRR, abs/1609.06666, 2016.

Haoqiang Fan, Hao Su, and Leonidas J. Guibas.
A point set generation network for 3d object reconstruction from a single image.
CoRR, abs/1612.00603, 2016.

Alberto Garcia-Garcia, Francisco Gomez-Donoso, José García Rodríguez, Sergio Orts-Escolano, Miguel Cazorla, and Jorge Azorín López.

References III

Pointnet: A 3d convolutional neural network for real-time object class recognition.
In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pages 1578–1584, 2016.

Rohit Girdhar, David F. Fouhey, Mikel Rodriguez, and Abhinav Gupta.
Learning a predictable and generative vector representation for objects.
CoRR, abs/1603.08637, 2016.

References IV

Vishakh Hegde and Reza Zadeh.
Fusionnet: 3d object classification using multiple data representations.
CoRR, abs/1607.05695, 2016.

Yangyan Li, Sören Pirk, Hao Su, Charles Ruizhongtai Qi, and Leonidas J. Guibas.
FPNN: field probing neural networks for 3d data.
CoRR, abs/1605.06240, 2016.

References V

- Jonathan Masci, Davide Boscaini, Michael M. Bronstein, and Pierre Vandergheynst.
  Shapenet: Convolutional neural networks on non-euclidean manifolds.
  CoRR, abs/1501.06297, 2015.
- Daniel Maturana and Sebastian Scherer.
  Voxnet: A 3d convolutional neural network for real-time object recognition.
  In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2015,*

References VI

Hamburg, Germany, September 28 - October 2, 2015, pages 922–928, 2015.

Charles Ruizhongtai Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J. Guibas.
Volumetric and multi-view cnns for object classification on 3d data.
CoRR, abs/1604.03265, 2016.

References VII

- Gernot Riegler, Ali Osman Ulusoy, Horst Bischof, and Andreas Geiger.
  Octnetfusion: Learning depth fusion from data.
  CoRR, abs/1704.01047, 2017.
- Gernot Riegler, Ali Osman Ulusoy, and Andreas Geiger.
  Octnet: Learning deep 3d representations at high resolutions.
  CoRR, abs/1611.05009, 2016.

References VIII

Nima Sedaghat, Mohammadreza Zolfaghari, and Thomas Brox.
Orientation-boosted voxel nets for 3d object recognition.
CoRR, abs/1604.03351, 2016.

Abhishek Sharma, Oliver Grau, and Mario Fritz.
Vconv-dae: Deep volumetric shape learning without object labels.
CoRR, abs/1604.03755, 2016.

References IX

- Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik G. Learned-Miller.
  Multi-view convolutional neural networks for 3d shape recognition.
  CoRR, abs/1505.00880, 2015.

- Maxim Tatarchenko, Alexey Dosovitskiy, and Thomas Brox.
  Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs.
  CoRR, abs/1703.09438, 2017.

References X

- Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Sun Chun-Yu, and Xin Tong.
  O-cnn: Octree-based convolutional neural networks for 3d shape analysis.
  ACM Transactions on Graphics (SIGGRAPH), 36(4), 2017.

- Zhirong Wu, Shuran Song, Aditya Khosla, Xiaoou Tang, and Jianxiong Xiao.
  3d shapenets for 2.5d object recognition and next-best-view prediction.
  CoRR, abs/1406.5670, 2014.