Deep 3D Representation Learning for Visual Computing



Hao Su

UC San Diego

July 6, 2017



Overview of 3D deep learning

3D deep learning algorithms

Conclusion

Outline

Overview of 3D deep learning Background 3D deep learning tasks 3D deep learning algorithms

Conclusion

The world around us is comprised of 3D geometry





Robotics





Robotics





Augmented Reality



Robotics





Augmented Reality



Autonomous driving



Robotics



Autonomous driving



Augmented Reality



Medical Image Processing





Historically, most 3D visual computing techniques focus on single models, lacking robustness



Autonomous driving



Medical Image Processing

Lacking 3D data has been the major bottleneck

Status as of 2010:



Stanford bunny



Utah teapot

1800 models in 90 categories



Princeton shape benchmark [Shilane et al. 04]

Recent rise of Internet 3D models



Nowadays millions of 3D models in online repositories

Recent rise of Internet 3D models

Growing market of crowd-sourcing for 3D modeling



Nowadays millions of 3D models in online repositories

Recent rise of Internet 3D models

Growing market of crowd-sourcing for 3D modeling



Nowadays millions of 3D models in online repositories

Learning for 3D data



Build 3D knowledge base

- -



Build 3D knowledge base

Design deep learning methods



The surge of 3D deep learning

- Arguably started from **2015** along with of big 3D datasets (ShapeNet & ModelNet)
- Very active due to huge industry interests!



- Robotics
- Autonomous driving
- Virtual/augmented reality
- Smart manufacturing

• • • •

3D geometry analysis

3D-assisted image analysis

3D synthesis

3D geometry analysis







Classification

Parsing (object/scene)

Correspondence

3D synthesis









Monocular 3D reconstruction

Shape completion

Shape modeling

3D-assisted image analysis







Results

Cross-view image retrieval

Intrinsic decomposition

All about Data and Network

3D geometry analysis

3D-assisted image analysis

3D synthesis





All about **Data** and **Network**

3D geometry analysis



3D synthesis



Overview of 3D deep learning

3D deep learning algorithms

3D Representation issue

Deep learning on different 3D representations

Conclusion

Images: Unique representation with regular data structure



1	44	33	12	20	23	35	14
51	16	40	32	46	48	28	17
29	60	3	63	49	55	36	7
52	22	26	41	38	10	61	53
2	24	19	11	34	43	5	8
2 57	24 9	19 37	11 42	34 25	43 21	5 27	8 18
2 57 30	24 9 56	19 37 50	11 42 64	34 25 4	43 21 59	5 27 6	8 18 13

3D has many representations:

3D has many representations:



multi-view RGB(D) images
volumetric
polygonal mesh
point cloud
primitive-based CAD models

Novel view image synthesis

3D has many representations:





3D has many representations:



3D has many representations:

3D has many representations:





3D has many representations:

Rasterized form (regular grids)

Geometric form (irregular) multi-view RGB(D) images

volumetric

polygonal mesh

point cloud

primitive-based CAD models

3D deep learning algorithms (by representations)



[Su et al. 2015] [Kalogerakis et al. 2016]

. . .



Volumetric

[Maturana et al. 2015] [Wu et al. 2015] (GAN) [Qi et al. 2016] [Liu et al. 2016] [Wang et al. 2017] (O-Net) [Tatarchenko et al. 2017] (OGN)

. . .

3D deep learning algorithms (by representations)



Multi-view

[Su et al. 2015] [Kalogerakis et al. 2016]

. . .



Volumetric

[Maturana et al. 2015] [Wu et al. 2015] (GAN) [Qi et al. 2016] [Liu et al. 2016] **[Wang et al. 2017] (O-Net) [Tatarchenko et al. 2017] (OGN)**

. . .

[Qi et al. 2017] (PointNet) [Fan et al. 2017] (PointSetGen) [Defferard et al. 2016] [Henaff et al. 2015] **[Yi et al. 2017] (SyncSpecCNN)** ... [Tulsiani et al. 2017] [Li et al. 2017] (GRASS)

Point cloud

Mesh (Graph CNN)

Part assembly

Cartesian product space of "task" and "representation"

3D geometry analysis



3D synthesis


Can we directly apply CNN on 3D data?



Can we directly apply CNN on 3D data?



14433122023351451164032464828172960363495536752222641381061532241911344358579374225212718305650644596135847453139156254								
51 16 40 32 46 48 28 17 29 60 3 63 49 55 36 7 52 22 26 41 38 10 61 53 2 24 19 11 34 43 5 8 57 9 37 42 25 21 27 18 30 56 50 64 4 59 6 13 58 47 45 31 39 15 62 54	1	44	33	12	20	23	35	14
29 60 3 63 49 55 36 7 52 22 26 41 38 10 61 53 2 24 19 11 34 43 5 8 57 9 37 42 25 21 27 18 30 56 50 64 4 59 6 13 58 47 45 31 39 15 62 54	51	16	40	32	46	48	28	17
52 22 26 41 38 10 61 53 2 24 19 11 34 43 5 8 57 9 37 42 25 21 27 18 30 56 50 64 4 59 6 13 58 47 45 31 39 15 62 54	29	60	3	63	49	55	36	7
2 24 19 11 34 43 5 8 57 9 37 42 25 21 27 18 30 56 50 64 4 59 6 13 58 47 45 31 39 15 62 54	52	22	26	41	38	10	61	53
579374225212718305650644596135847453139156254	2	24	19	11	34	43	5	8
30 56 50 64 4 59 6 13 58 47 45 31 39 15 62 54	57	9	37	42	25	21	27	18
58 47 45 31 39 15 62 54	30	56	50	64	4	59	6	13
	58	47	45	31	39	15	62	54

Convolution needs an underlying structure $(f * g)[n] = \sum_{m=-M}^{M} f[n - m]g[m]$

3D has many representations:

Rasterized form (regular grids)

- Can directly apply CNN
- But has other challenges

multi-view RGB(D) images volumetric

3D has many representations:

Rasterized form (regular grids)

Geometric form (irregular)

Cannot directly apply CNN

multi-view RGB(D) images volumetric

polygonal mesh point cloud primitive-based CAD models

Deep learning on Multiview representation

Multi-view representation as 3D input

Leverage the huge CNN literature in image analysis

Multi-view representation as 3D input

Classification



Hang Su, Subhransu Maji, Evangelos Kalogerakis, Erik Learned-Miller, "Multi-view Convolutional Neural Networks for 3D Shape Recognition", Proceedings of ICCV 2015

Multi-view representation as 3D output

Novel-view RGB(D) image synthesis (direct prediction)



Maxim Tatarchenko, Alexey Dosovitskiy, Thomas Brox,

"Multi-view 3D Models from Single Images with a Convolutional Network", ECCV2016

Multi-view representation as 3D output

Novel-view RGB(D) image synthesis (flow prediction)



Tinghui Zhou, Shubham Tulsiani, Weilun Sun, Jitendra Malik, Alexei A. Efros **"View Synthesis by Appearance Flow**" ECCV2016

• Each view only contains partial information

- Each view only contains partial information
- Not trivial to predict across viewpoints



- Each view only contains partial information
- Not trivial to predict across viewpoints
- Cannot see through the surface





- Regular structures in 3D cannot be well captured
 - e.g., symmetry, straightness, roundish

• Each view only contains partial information

- Not trivial to agg
 A true 3D representation is
 Cannot see throit
 More natural for 3D learning
- Regular structures in 3D cannot be well captured
 - e.g., symmetry, straightness, roundish

Deep learning on volumetric representation

3D CNN on volumetric data

3D convolution uses 4D kernels



[Credit: Su et al.]

Computational complexity issue

3D convolution uses 4D kernels



High space/time complexity $O(N^3)$

[Credit: Su et al.]

Computational complexity issue





3DShapeNets, 2015

Input resolution: 30x30x30

224x224=27000

AlexNet, **2012**

Input resolution: 224x224

224x224=50176

Computational complexity issue



Polygon Mesh

Occupancy Grid 30x30x30

Information loss in voxelization

The sparsity characteristic of 3D data

Yangyan Li, Sören Pirk, Hao Su, Charles R. Qi, Leonidas J. Guibas FPNN: Field Probing Neural Networks for 3D Data



Store only the occupied grids

Octree: recursively partition the space Each internal node has exactly eight children



Skip the computation of empty cells



Gernot Riegler, Ali Osman Ulusoy, Andreas Geiger "OctNet: Learning Deep 3D Representations at High Resolutions" CVPR2017

Pengshuai Wwang, Yang Liu, Yuxiao Guo, Chunyu Sun, Xin Tong "O-CNN: Octree-based Convolutional Neural Network for Understanding 3D Shapes" *SIGGRAPH2017*

Volumetric representation as input

Define convolution and pooling along the octree



Challenge: how to implement efficiently — build a hash table to index the neighborhood Restrict the convolution stride to be 2

Volumetric representation as output



Christopher B. Choy, Danfei Xu*, JunYoung Gwak*, Kevin Chen, Silvio Savarese, **3D-R^2N^2: A unified approach for single and multi-view 3D object reconstruction** *ECCV2016*

Towards higher spatial resolution



Maxim Tatarchenko, Alexey Dosovitskiy, Thomas Brox "Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs" arxiv (March, 2017)

Progressive voxel refinement



Key challenges for volumetric representation

- Computational complexity (seems to have been resolved)
- Regular structures in 3D cannot be well captured in reconstruction
 - e.g., symmetry, straightness, roundish

Typical artifacts of volumetric reconstruction



Missing thin structures due to improper shape space structure

hard for the network to rotate / deform / interpolate

How to design neural networks for geometric forms?

3D has many representations:

Rasterized form (regular grids)

Geometric form (irregular)

Cannot directly apply CNN

multi-view RGB(D) images volumetric polygonal mesh

point cloud

primitive-based CAD models

Deep learning on polygonal mesh

!! math heavy, you can take a break if you do not like math that much. Be normal soon.

Two different strategies for deep learning on graphs

Directly conduct convolution on graphs

Conduct convolution on 2D parameterization of 3D surfaces

Two different strategies for deep learning on meshes

Directly conduct convolution on graphs

Spatial construction (Geodesic CNN) Spectral construction (Spectral CNN)

Conduct convolution on 2D parameterization of 3D surfaces

Meshes can be represented as graphs



Geometry aware convolution can be important



image credit: D. Boscaini, et al.

convolutional considering underlying geometry

image credit: D. Boscaini, et al.

convolutional along spatial coordinates

How to define convolution kernel on graphs?

- Desired properties:
 - locally supported (w.r.t graph metric)
 - allowing weight sharing across different coordinates



from Shuman et al. 2013

How to allow multi-scale analysis?



from Michaël Defferrard et al. 2016
How to allow multi-scale analysis?



from Michaël Defferrard et al. 2016

Spatial construction: Geodesic CNN

- Constructing convolution kernels:
 - Local system of geodesic polar coordinate
 - Extract a small patch at each point x



Issues of Geodesic CNN

- The local charting method relies on a fast marching-like procedure requiring a triangular mesh.
- The radius of the geodesic patches must be sufficiently small to acquire a topological disk.
- No effective pooling, purely relying on convolutions to increase receptive field.

Spectral construction: Spectral CNN

Fourier analysis

Convert convolution to multiplication in spectral domain

Convolution Theorem in non-Euclidean domain

Generalized convolution of $f,g \in L^2(X)$ can be defined by analogy



Bases on meshes: eigenfunction of Laplacian-Bertrami operator



Synchronization of functional space across meshes

Functional map



Li Yi, Hao Su, Xingwen Guo, Leonidas Guibas "SyncSpecCNN: Synchronized Spectral CNN for 3D Shape Segmentation" CVPR2017 (spotlight)

Two different strategies for deep learning on meshes

Directly conduct convolution on graphs

Conduct convolution on 2D parameterization of 3D surfaces

Surface parameterization

Map curved 3D surfaces to 2D Euclidean plane





Ayan Sinha, Jing Bai, Karthik Ramani "Deep Learning 3D Shape Surfaces Using Geometry Images" ECCV2016 Maron et al.

"Convolutional Neural Networks on Surfaces via Seamless Toric Covers" SIGGRAPH2017

Deep learning on surface parameterization

Use CNN to predict the parameterization, then convert to 3D mesh



Ayan Sinha, Asim Unmesh, Qixing Huang, Karthik Ramani "SurfNet: Generating 3D shape surfaces using deep residual networks" CVPR2017

Key challenges for mesh representation

• Good progress seems to have been made for meshes as input

- Mesh as output is very challenging:
 - Need consistent surface parameterization
 - Not clear how to generate shapes with topology variation

Deep learning on point cloud

PointNet: Directly process point cloud data



Hao Su, Charles Qi, Kaichun Mo, Leonidas Guibas **PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation** *CVPR 2017 (oral)*

PointNet: Directly process point cloud data

. . .



Properties of a desired neural network on point clouds

Point cloud: N orderless points, each represented by a D dim coordinate



2D array representation

Properties of a desired neural network on point clouds

Point cloud: N orderless points, each represented by a D dim coordinate



2D array representation

Permutation invariance

Transformation invariance

Properties of a desired neural network on point clouds

Point cloud: N orderless points, each represented by a D dim coordinate



2D array representation

Permutation invariance

Permutation invariance: Symmetric function

$$f(x_1, x_2, \dots, x_n) \equiv f(x_{\pi_1}, x_{\pi_2}, \dots, x_{\pi_n}), x_i \in \mathbb{R}^D$$

Examples:

. . .

$$f(x_1, x_2, \dots, x_n) = \max\{x_1, x_2, \dots, x_n\}$$
$$f(x_1, x_2, \dots, x_n) = x_1 + x_2 + \dots + x_n$$







Q: What symmetric functions can be constructed by PointNet?



PointNet (vanilla)

Theorem:

A Hausdorff continuous symmetric function $f: 2^{\mathcal{X}} \to \mathbb{R}$ can be arbitrarily approximated by PointNet.



Robustness to data corruption



Robustness to data corruption



Segmentation from partial scans



Non-uniform Sampling Density

Density variation is a common issue of 3D point cloud

- perspective effect, radial density variation, motion etc.



PointNet++: Robust learning under varying sampling density



Charles R. Qi, Li Yi, Hao Su, Leonidas J. Guibas **PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space** *arxiv*

Point cloud as output



Input Reconstructed 3D point cloud

Hao Su, Haoqiang Fan, Leonidas Guibas "A Point Set Generation Network for 3D Object Reconstruction from a Single Image" CVPR2017 (oral)

Volumetric upconvolution?



Another representation possibility: Point clouds

Transformation friendly for networks

? Usable as network output?

No prior works in deep learning community!





CVPR '17, Point Set Generation



Recent work on 3D prediction by point clouds



Input Reconstructed 3D point cloud
The first work to generate a set in deep learning [CVPR'2017(oral)]

Comparison to direct 3D volumetric upconvolution



Input

Volumetric upconv (ECCV 2016, 3D-R2N2)

(post-processed to volumetric)

Groundtruth

CVPR '17, Point Set Generation

Pipeline



Loss function: Earth Mover's Distance (EMD)

• Given two sets of points, measure their discrepancy:



Differentiable

Admit fast computation

Quantitative evaluation



CVPR '17, Point Set Generation
Quantitative evaluation



CVPR '17, Point Set Generation

Real-world results



CVPR '17, Point Set Generation

Generalization to unseen categories

input 90° observed view input observed view 90° Out of training categories

CVPR '17, Point Set Generation

Key challenges for point cloud representation

- Point cloud as output is still very challenging:
 - The global structure is reasonable but details are missing

• Combined with volumetric representation seems to give better results. Need more study on optimal combination strategy. Deep learning on primitives

Primitive-based assembly



Shubham Tulsiani, Hao Su, Leonidas Guibas, Alexei A. Efros, Jitendra Malik Learning Shape Abstractions by Assembling Volumetric Primitives *CVPR 2017*

Approach



We predict primitive parameters: size, rotation, translation of M cuboids.

Variable number of parts? We predict "primitive existence probability"

GRASS



Jun Li, Kai Xu, Siddhartha Chaudhuri, Ersin Yumer, Hao Zhang, Leonidas Guibas "GRASS: Generative Recursive Autoencoders for Shape Structures" SIGGRAPH 2017 How to introduce other primitives types?

Towards image based modeling, how to add more operations to edit those primitives?

• e.g., Deform? Extrude? Loop cut?

How to use it for design purposes? For example, with certain structural and functional constraints.

Ultimately, we expect to automate the modeling process from images, as artists do.



Overview of 3D deep learning

3D deep learning algorithms

Conclusion

The surge of 3D deep learning

- A field with very short history arguably started from **2015**
- But very active due to huge industry interests!



- Robotics
- Autonomous driving
- Virtual/augmented reality
- Smart manufacturing

Based upon a new course at Stanford

Course (Machine Learning on 3D data) website:

http://graphics.stanford.edu/courses/cs468-17-spring/schedule.html

Tutorial on 3D deep learning at CVPR, see you at Hawaii!

http://3ddl.stanford.edu/

Workshop on Learning to see 3D data at ICCV'17, Venice, Italy

Opening for PhD/Postdoc/Visiting Scholar positions

Deep learning for computer vision, computer graphics, and robotics

More information on my personal homepage

Thank you!