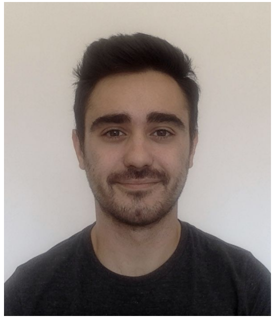
http://bit.ly/dlcv2018

#DLUPC

Day 4 Lecture 2

3D Reconstruction



Eduard Ramon
eduard.ramon.maldonado@upc.edu

PhD Student
Universitat Politècnica de Catalunya
Technical University of Catalonia

crisalix VIRTUAL AESTHETICS

telecom BCN

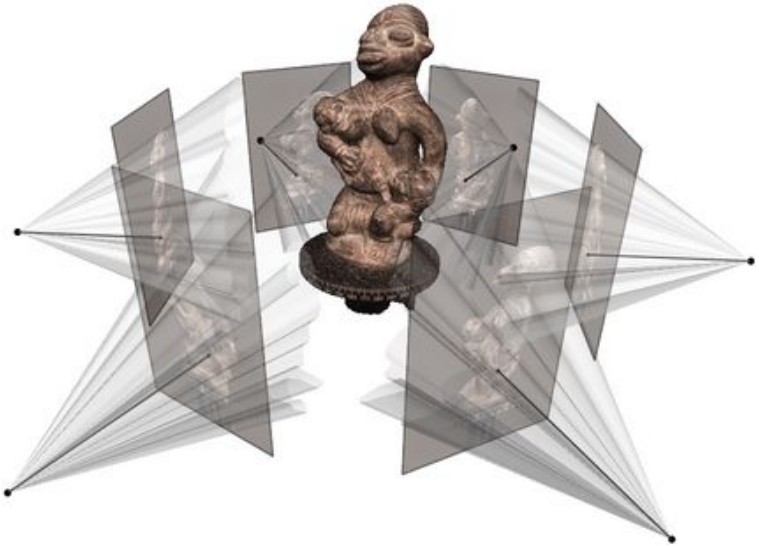
UPC

Outline

- Introduction
- Motivation
- Classical methods
    - Geometric and Stereo
    - Limitations
- Deep learning approaches
    - Volumetric grids: 3D-R2N2
    - Point clouds: PointOutNet
    - Meshes: Pixel2Mesh

Introduction

3D Reconstruction is the process of obtaining the geometric properties of a scene by processing and combining visual cues from a set of views.

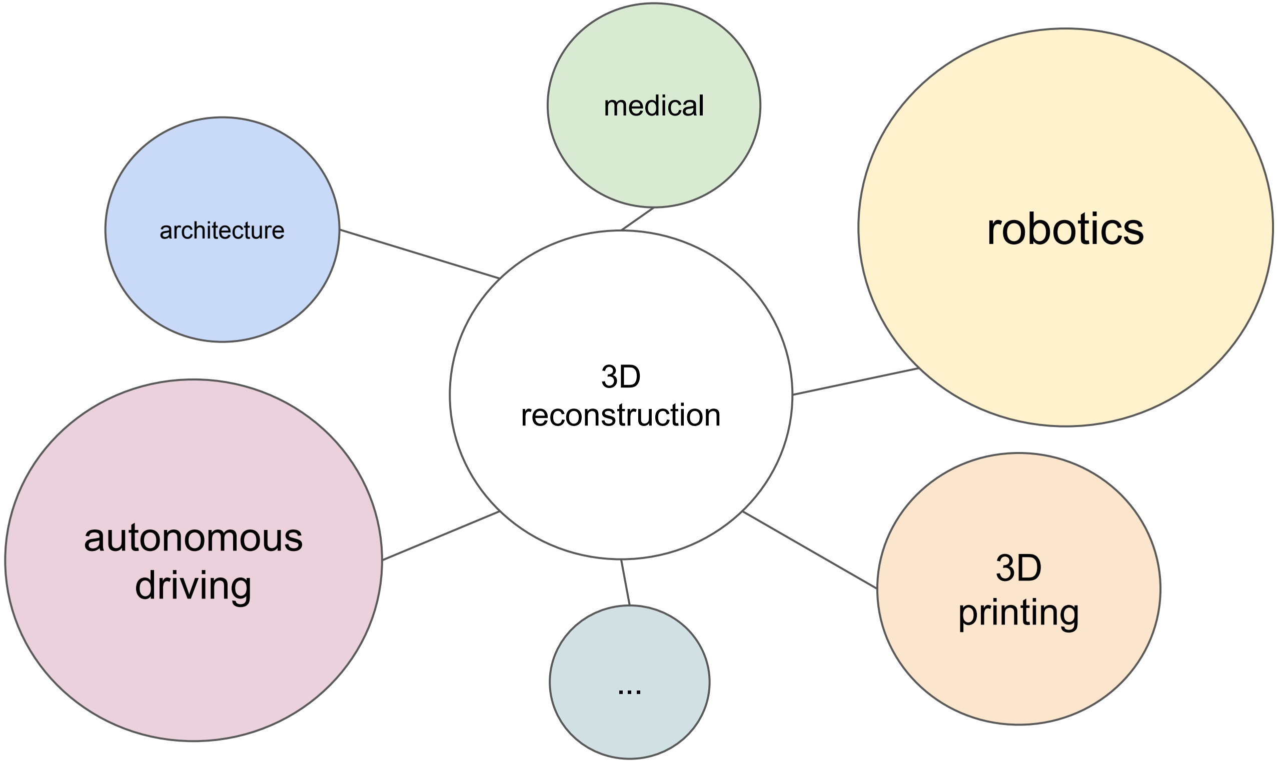


- Surface
- Camera poses
- Albedo (percentage of reflected radiation or base color)
- Illumination
- ...

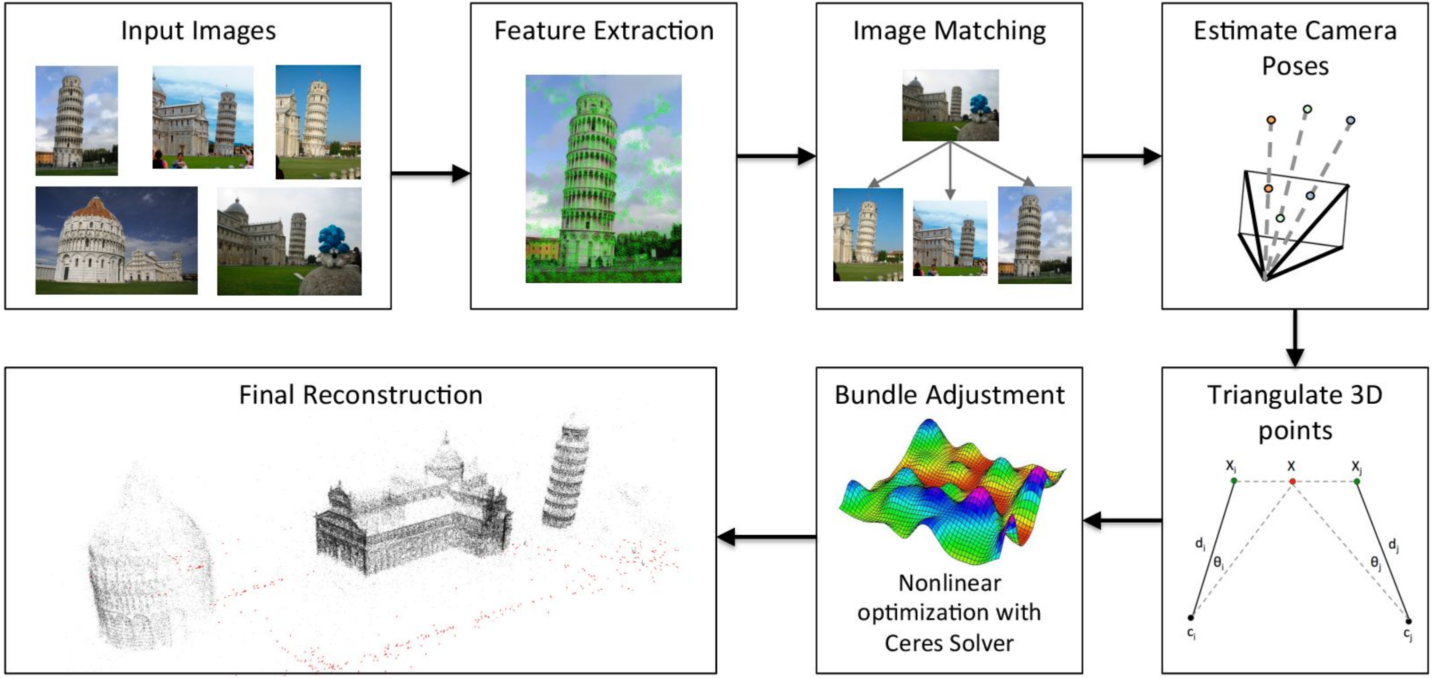
[Multi-view Stereo: A tutorial. - Carlos Hernández, Google Inc.]

Motivation

There's a bunch of applications which may benefit from 3D reconstruction algorithms.

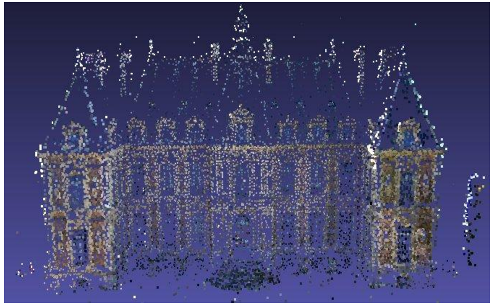


Classical methods: Multiview Geometry



http://www.theia-sfm.org/sfm.html

Classical methods: Multiview Stereo

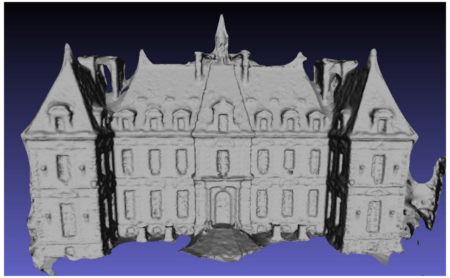


Output from Multiview Geometry

Point cloud densification

Meshing

Texturing



Output from Multiview Stereo

https://github.com/cdcseacave/openMVS

Classical methods: Limitations

When do they fail?

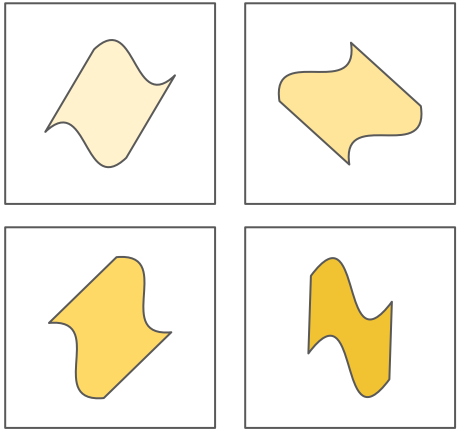
- Not enough images with enough overlap.
- Featureless or reflecting surfaces.
- Pure rotations.
- Repeated structures.
- Thin structures.
- Non-Lambertian surfaces.

Deep Learning:

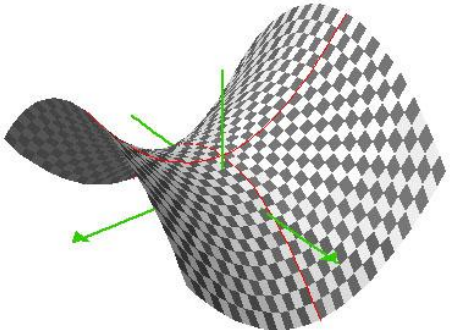
- More descriptive, robust and problem specific image representations.
- Encoding of prior knowledge.

Deep learning approaches

Goal: Estimate an irregular surface and its properties from a set of RGB images.

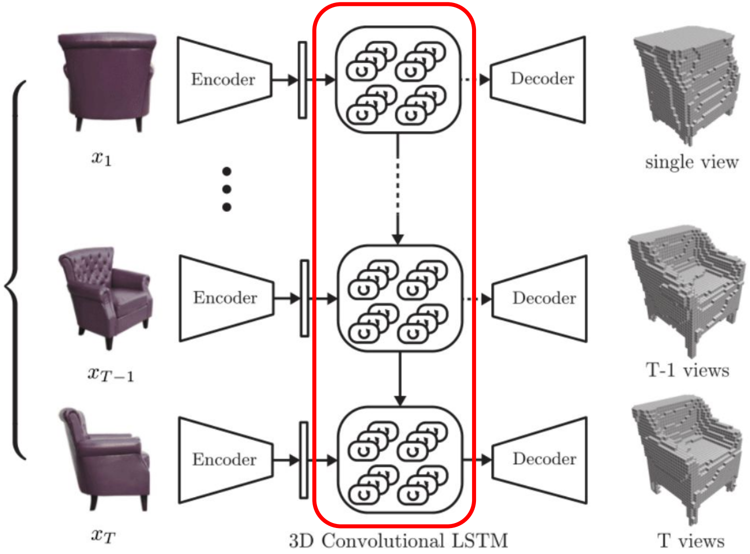


$$S = f(I_1, I_2, \ldots, I_N)$$



Irregular surfaces do not lie on an **Euclidean Space** and, in principle, we cannot use the standard convolution to model them.

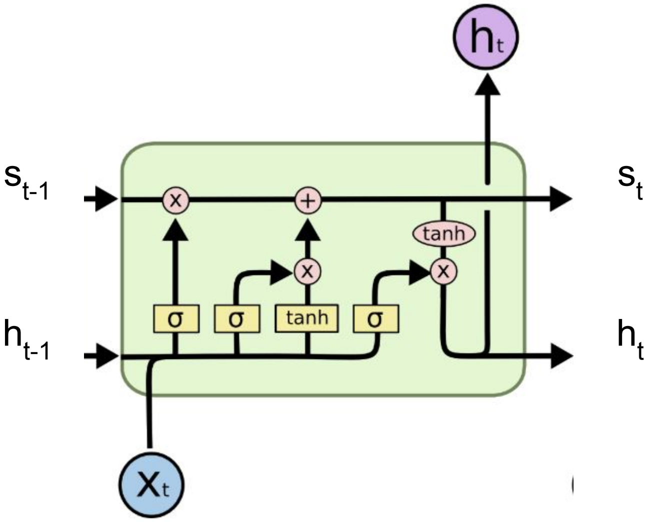
Volumetric grids: 3D-R2N2



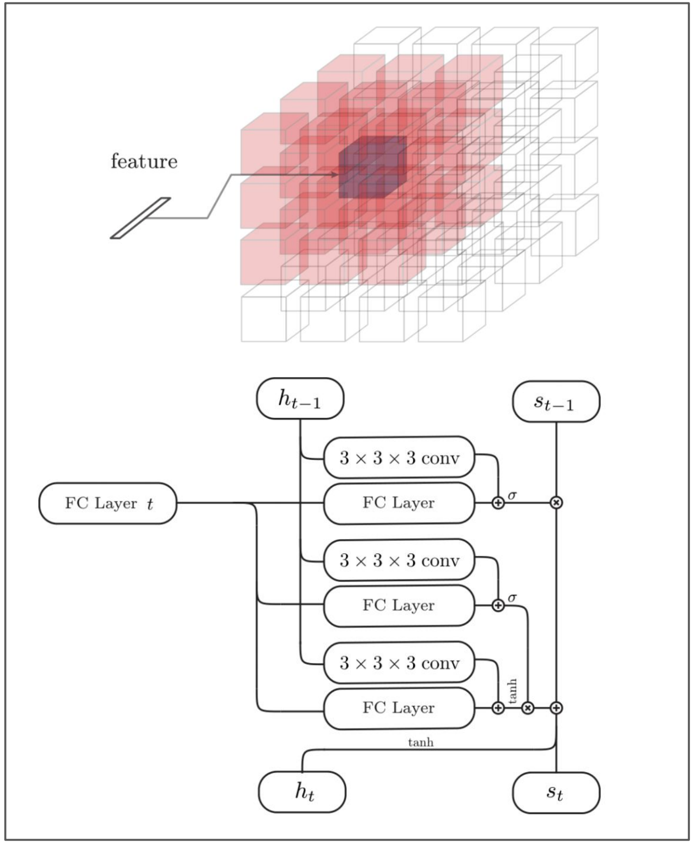
$$L(\mathcal{X}, y) = \sum_{i,j,k} y_{(i,j,k)} \log(p_{(i,j,k)}) + (1 - y_{(i,j,k)}) \log(1 - p_{(i,j,k)})$$

3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction [ECCV 2016]

Volumetric grids: 3D-R2N2



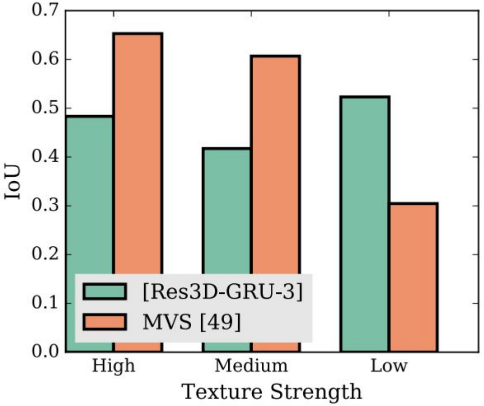
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$
$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$
$$s_t = f_t \odot s_{t-1} + i_t \odot \tanh(W_s x_t + U_s h_{t-1} + b_s)$$
$$h_t = o_t \odot \tanh(s_t)$$



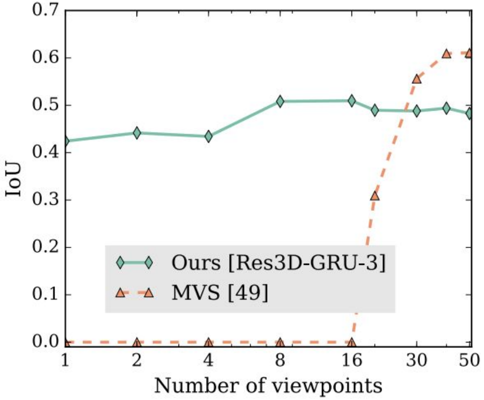
[3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction [ECCV 2016]]

Volumetric grids: 3D-R2N2

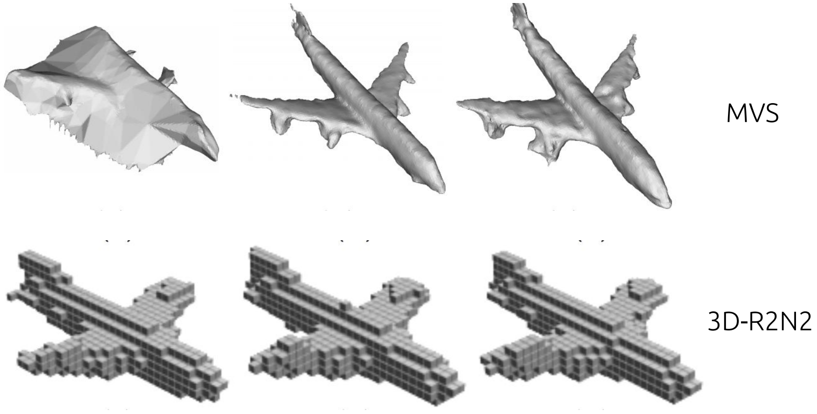
bit.ly/DLCV2018
#DLUPC



Reconstruction quality with respect the texture strength, averaging results from 20, 30 and 40 views.



Reconstruction quality with respect the number of views, averaging on all texture strengths.



20 views 30 views 40 views

[3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction [ECCV 2016]]

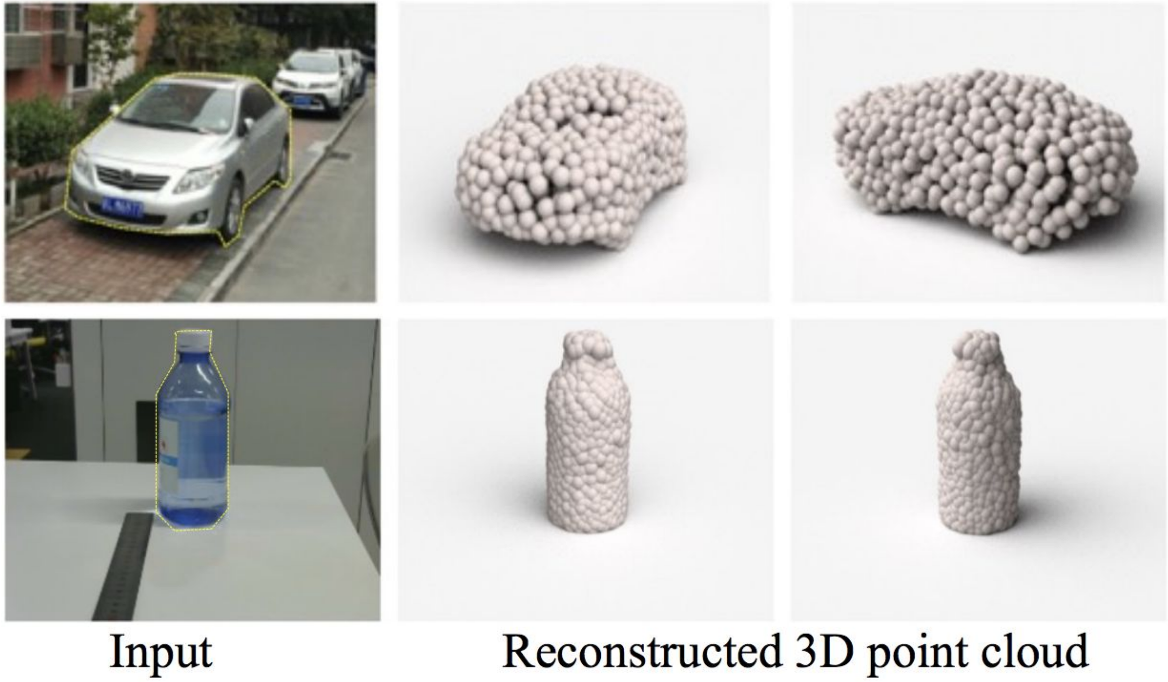
Volumetric grids: 3D-R2N2

Limitations

- Inefficient use of the representation space.
- Limited resolution 32^3 due to memory constraints (cubic growth).
- Modeling views as sequences instead of sets.
- Requires 3D supervision.
- Requires post-processing.

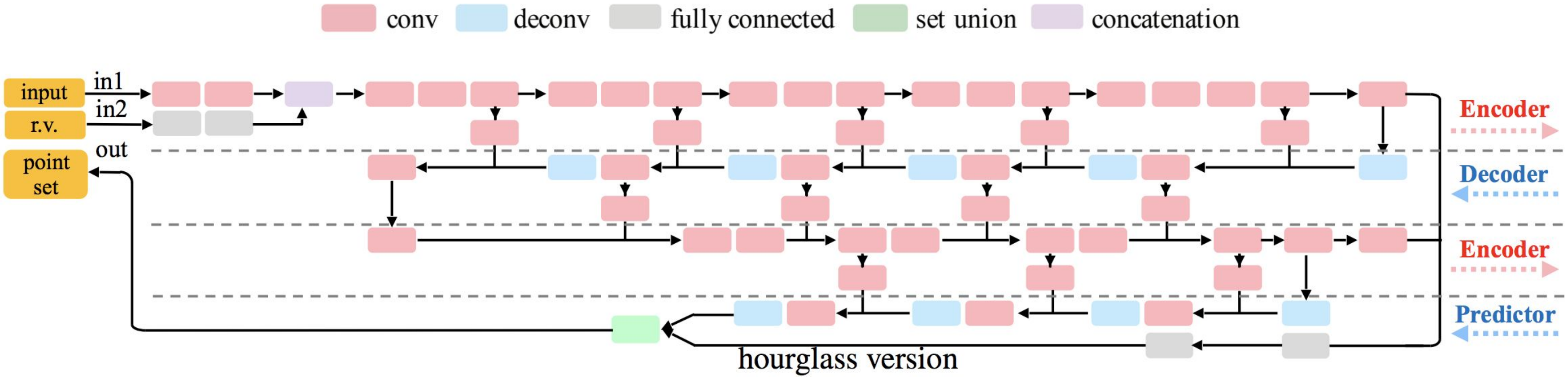
3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction [ECCV 2016]

Point clouds: PointOutNet

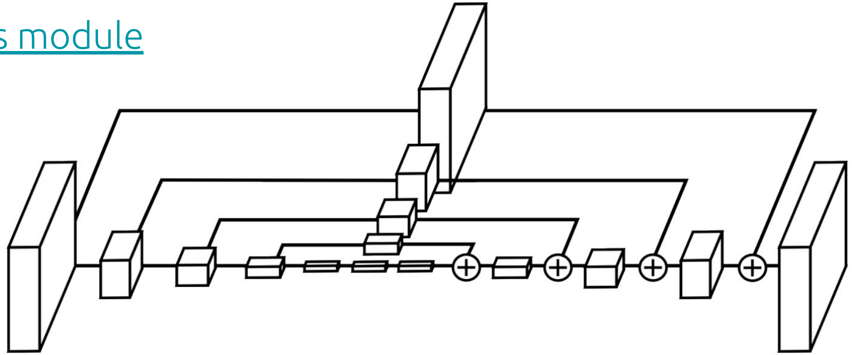


[A Point Set Generation Network for 3D Object Reconstruction from a Single Image [CVPR 2017]]

Point clouds: PointOutNet



[Hourglass module](#)



[A Point Set Generation Network for 3D Object Reconstruction from a Single Image [CVPR 2017]](#)

Point clouds: PointOutNet

Data terms (not both at the same time):

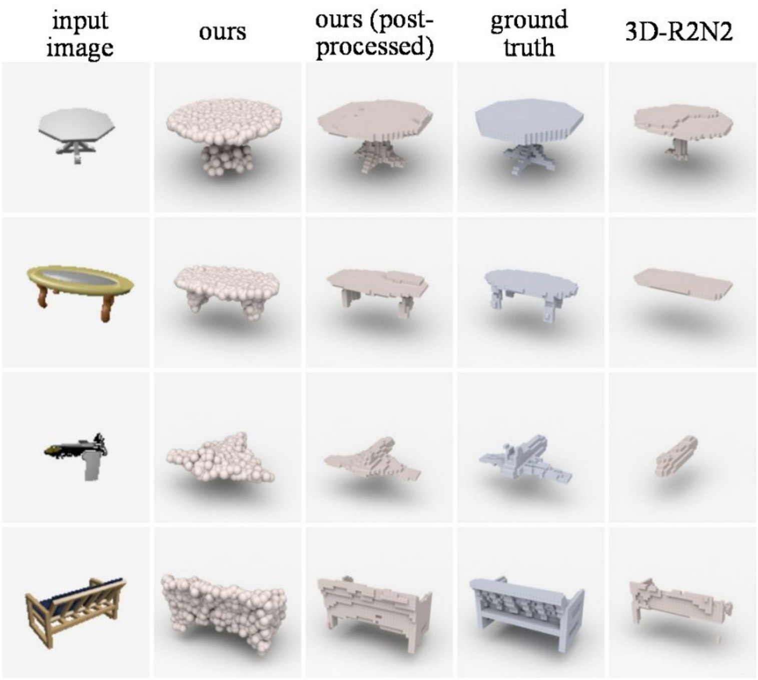
- Chamfer distance $$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2$$
- Earth Mover's distance $$d_{EMD}(S_1, S_2) = \min_{\phi: S_1 \to S_2} \sum_{x \in S_1} \|x - \phi(x)\|_2$$

Statistical term (not both at the same time):

- Minimum-of-N (MoN) $$\underset{\Theta}{\textbf{minimize}} \sum_k \min_{\substack{r_j \sim \mathbb{N}(\mathbf{0}, \mathbf{I}) \\ 1 \le j \le n}} \{d(\mathbb{G}(I_k, r_j; \Theta), S_k^{gt})\}$$
- Conditional VAE

A Point Set Generation Network for 3D Object Reconstruction from a Single Image [CVPR 2017]

Point clouds: PointOutNet



category	Ours	3D-R2N2		
	1 view	1 view	3 views	5 views
plane	**0.601**	0.513	0.549	0.561
bench	**0.550**	0.421	0.502	0.527
cabinet	0.771	0.716	0.763	**0.772**
car	0.831	0.798	0.829	**0.836**
chair	0.544	0.466	0.533	**0.550**
monitor	0.552	0.468	0.545	**0.565**
lamp	**0.462**	0.381	0.415	0.421
speaker	**0.737**	0.662	0.708	0.717
firearm	**0.604**	0.544	0.593	0.600
couch	**0.708**	0.628	0.690	0.706
table	**0.606**	0.513	0.564	0.580
cellphone	0.749	0.661	0.732	**0.754**
watercraft	**0.611**	0.513	0.596	0.610
mean	**0.640**	0.560	0.617	0.631

Intersection over Union (IoU) on ShapeNet

[A Point Set Generation Network for 3D Object Reconstruction from a Single Image [CVPR 2017]]

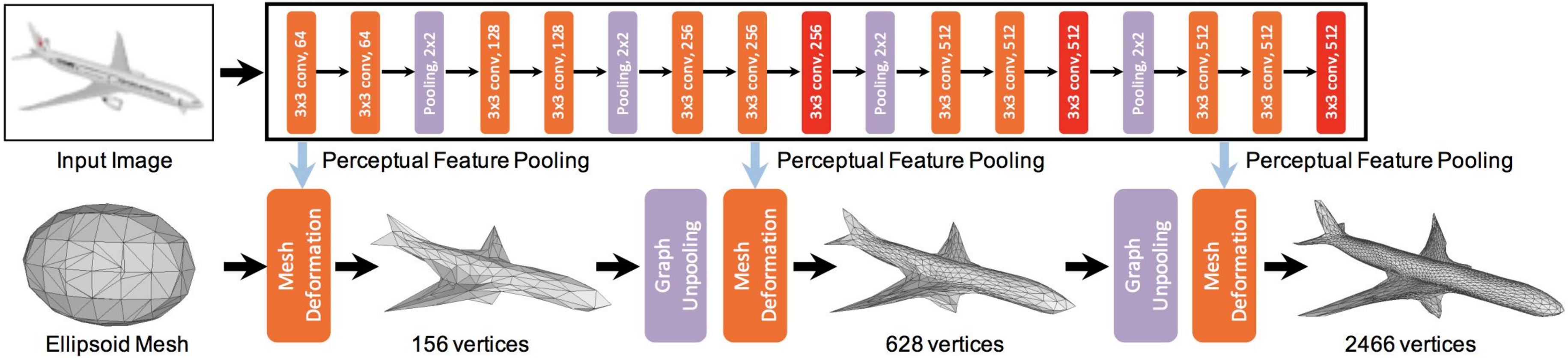
Point clouds: PointOutNet

Limitations

- Lack of details.
- Single view.
- Poor performance on unseen categories.
- Struggling with compositionality.
- Requires post-processing.

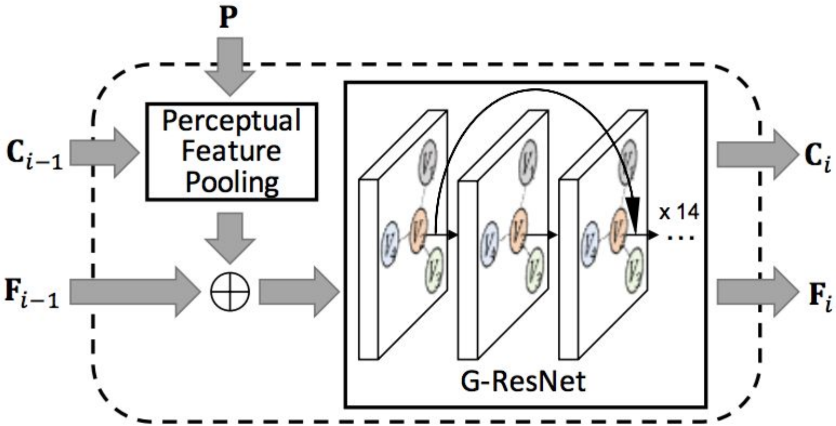
A Point Set Generation Network for 3D Object Reconstruction from a Single Image [CVPR 2017]

Meshes: Pixel2Mesh



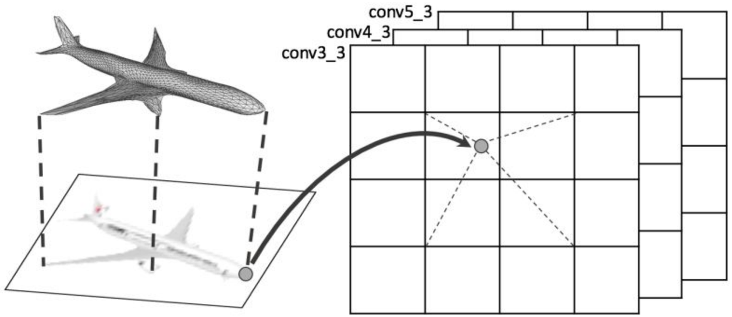
[Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images [arXiv 2018]]

Meshes: Pixel2Mesh



$$f_p^{l+1} = w_0 f_p^l + \sum_{q \in \mathcal{N}(p)} w_1 f_q^l$$

Perceptual Feature Pooling



Uses vertex position c_{i-1} and known camera pose to project to feature space and estimate features using bilinear interpolation.

[Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images [arXiv 2018]]

Meshes: Pixel2Mesh

Data terms:

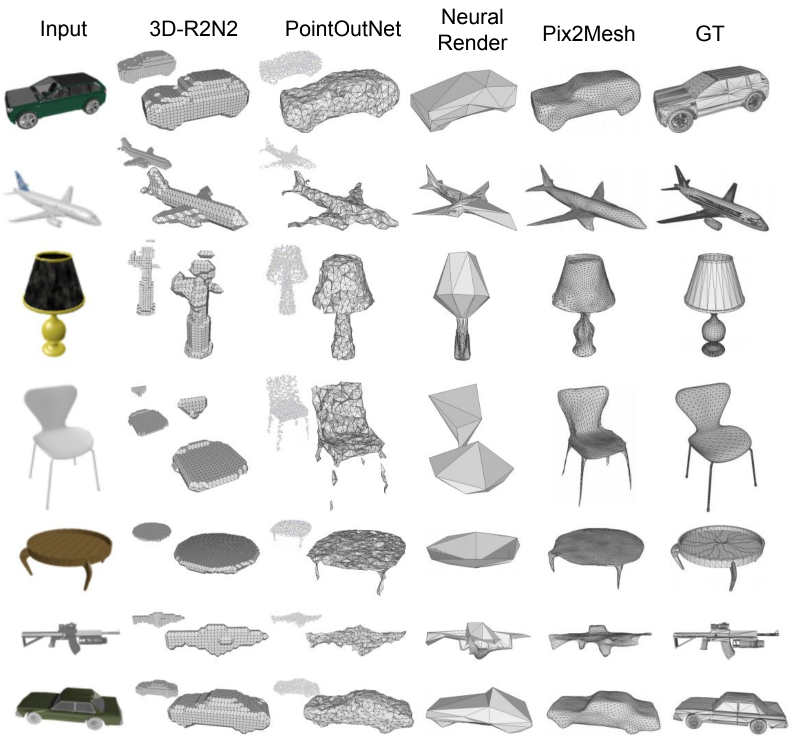
- Chamfer loss $$l_c = \sum_p \min_q \|p - q\|_2^2 + \sum_p \min_q \|p - q\|_2^2$$
- Normal loss $$l_n = \sum_p \sum_{\substack{q=\arg\min_p(\|p-q\|_2^2) \\ k \in \mathcal{N}(p)}} \|\langle p - k, \mathbf{n}_q \rangle\|_2^2$$

Regularizations

- Laplacian loss $$l_{lap} = \sum_p \|\delta'_p - \delta_p\|_2^2 \qquad \delta_p = p - \sum_{k \in \mathcal{N}(p)} \frac{1}{\|\mathcal{N}(p)\|} k$$
- Edge length loss $$l_{loc} = \sum_p \sum_{k \in \mathcal{N}(p)} \|p - k\|_2^2$$

[Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images [arXiv 2018]]

Meshes: Pixel2Mesh



Threshold	τ				2τ			
Category	3D-R2N2	PSG	N3MR	Ours	3D-R2N2	PSG	N3MR	Ours
plane	41.46	68.20	62.10	**71.12**	63.23	81.22	77.15	**81.38**
bench	34.09	49.29	35.84	**57.57**	48.89	69.17	49.58	**71.86**
cabinet	49.88	39.93	21.04	**60.39**	64.83	67.03	35.16	**77.19**
car	37.80	50.70	36.66	**67.86**	54.84	77.79	53.93	**84.15**
chair	40.22	41.60	30.25	**54.38**	55.20	63.70	44.59	**70.42**
monitor	34.38	40.53	28.77	**51.39**	48.23	63.64	42.76	**67.01**
lamp	32.35	41.40	27.97	**48.15**	44.37	58.84	39.41	**61.50**
speaker	45.30	32.61	19.46	**48.84**	57.86	56.79	32.20	**65.61**
firearm	28.34	69.96	52.22	**73.20**	46.87	82.65	63.28	**83.47**
couch	40.01	36.59	25.04	**51.90**	53.42	62.95	39.90	**69.83**
table	43.79	53.44	28.40	**66.30**	59.49	73.10	41.73	**79.20**
cellphone	42.31	55.95	27.96	**70.24**	60.88	79.63	41.83	**82.86**
watercraft	37.10	51.28	43.71	**55.12**	52.19	**70.63**	58.85	69.99
mean	39.01	48.58	33.80	**59.72**	54.62	69.78	47.72	**74.19**

F-score (%) on ShapeNet

[Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images [arXiv 2018]]

Meshes: Pixel2Mesh

Limitations

- Single view.
- Graph convolutional not based on geometric operator.
- Generates only meshes with genus 0.

[Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images [arXiv 2018]]

Outline

- Introduction
- Motivation
- Classical methods
  - Geometric and Stereo
  - Limitations
- Deep learning approaches
  - Volumetric grids: 3D-R2N2
  - Point clouds: PointOutNet
  - Meshes: Pixel2Mesh

THANK YOU !

Q & A