A Point Set Generation Network for 3D Object Reconstruction from a Single Image [4]

David Stutz

June 1-2, 2017

Motivation

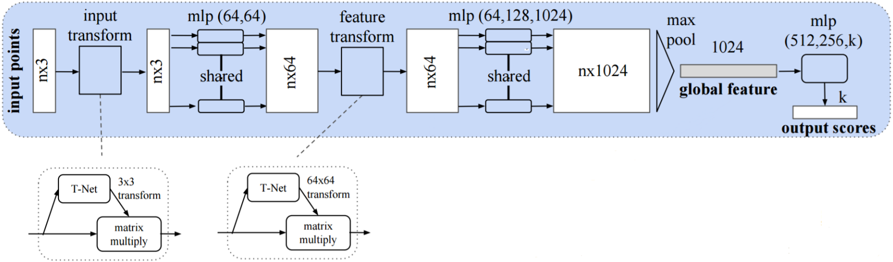
Addressed problem: 3D reconstruction from a single image.

PointNet [5]: deep learning on point sets.

Underlying motivation:

- How to learn how to generate shapes, i.e. point sets?

Recap: PointNet



Problems

Problems when predicting point sets:

- How to appropriately compare two unordered point sets?
- How to model uncertainty?

Comparing Point Sets

Chamfer Distance (CD) for point sets
$X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_m\}$:

$$d_{\mathrm{CD}}(X, Y) = \sum_{x \in X} \min_{y \in Y} \|x - y\|_2^2 + \sum_{y \in Y} \min_{x \in X} \|x - y\|_2^2.$$

Easily implemented (and parallelizable).

Comparing Point Sets (cont'd)

Earth Mover Distance (EMD) for point sets X, Y with $|X| = |Y|$:

$$d_{\text{EMD}} = \min_{\phi : X \mapsto Y} \sum_{x \in X} \|x - \phi(x)\|_2$$

with ϕ being a bijection.

Exact computation not feasible; $(1 + \epsilon)$ approximation [1] used.

Model Uncertainty

Possibilities:

- (Conditional) generative adversarial networks [6, 9];
- (Conditional) variational auto-encoders [8, 7, 10].

Model Uncertainty (cont'd)

Inject randomness as additional input (i.e. noise vector): $\boldsymbol{G}(\boldsymbol{I}, \epsilon)$.

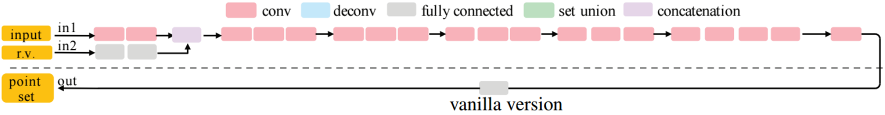
Minimize

$$\sum_k \min_{\epsilon_j \in \mathcal{N}(0,1)} d(G(I_k, \epsilon_j), S_k)$$

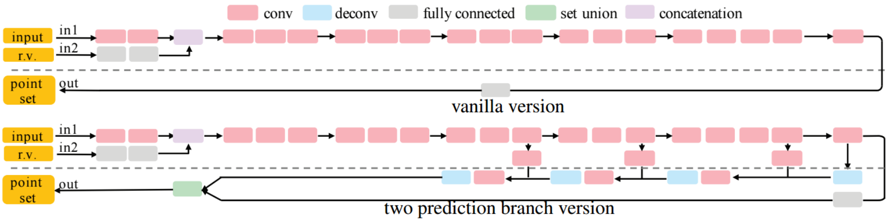
for $\boldsymbol{j} \in \{\mathbf{0}, \ldots, \boldsymbol{n}\}$, $\boldsymbol{I_k}$ being the input image and $\boldsymbol{S_k}$ the ground truth point set.

So-called "Min-of-N" loss.

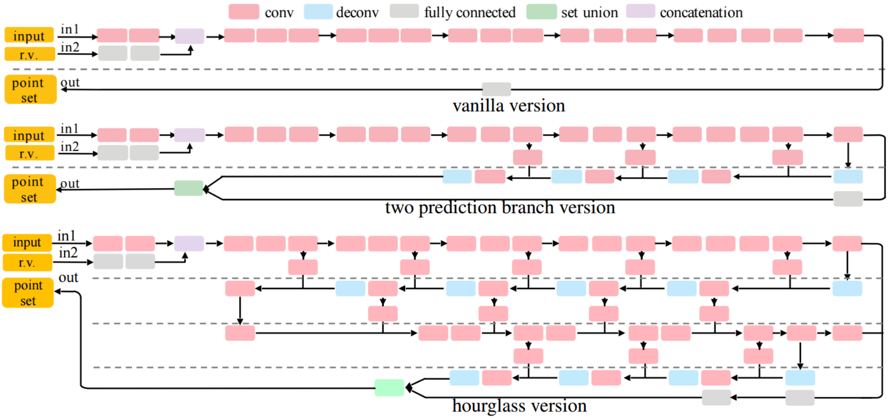
Network Architecture



Network Architecture (cont'd)



Network Architecture (cont'd)



3D Reconstruction (RGB)

Experimental Setup:

- rendered models (RGBD) from ShapeNet [2];
- unclear how many points are predicted – for two-branch: 256 + 768 points;
- unclear how the ground truth point sets are chosen;
- evaluation using Intersection-over-Union (IoU) on voxels.

3D Reconstruction (RGB) (cont'd)

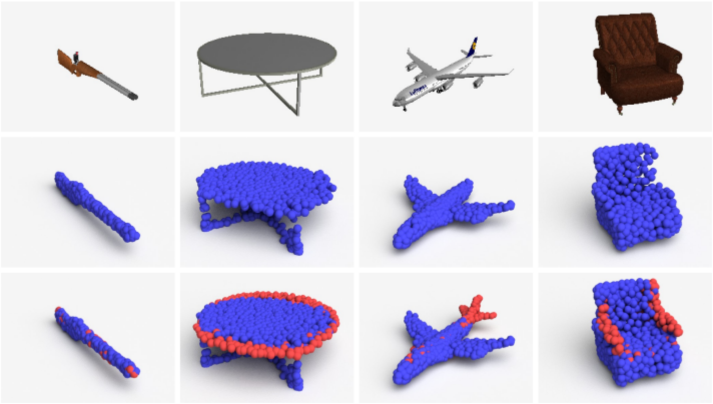
	Mean IoU
3D-R2N2 [3] 1 views	0.560
3D-R2N2 [3] 3 views	0.617
3D-R2N2 [3] 5 views	0.631
PointNet	0.64

Table: Intersection-over-Union results on ShapeNet [2] compared to 3D-R2N2 [3].

3D Reconstruction (RGB) (cont'd)

3D Shape Completion (RGBD)

Deconvolution vs. Fully Connected



My 2 Cents ...

Some experiments missing:

- Quantitative comparison of (C)VAE or (C)GAN with MoN loss;
- comparison of vanilla, two branches and hourglass with respect to IoU.

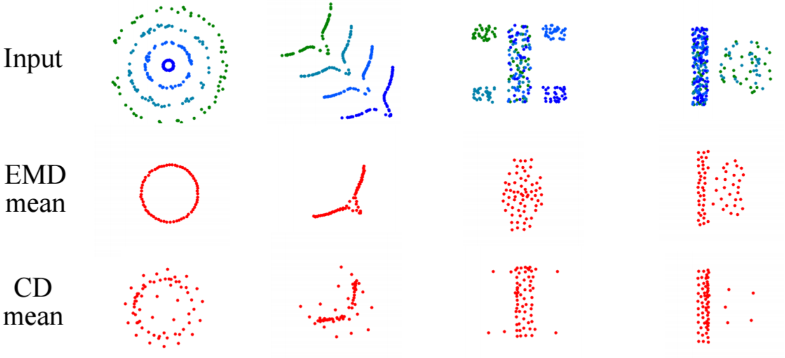
Conclusion

3D reconstruction by using a CNN to predict a set of points – trained on synthetic data.

Interesting in combination with PointNet, allows to directly operate on and predict unordered sets of points.

- The only two works on deep learning on 3D point sets ...

Distances



VAE Comparison

Figure: Multiple predictions using the MoN loss.

VAE Comparison (cont'd)

Figure: Multiple predictions using a (C)VAE formulation.

Real Examples

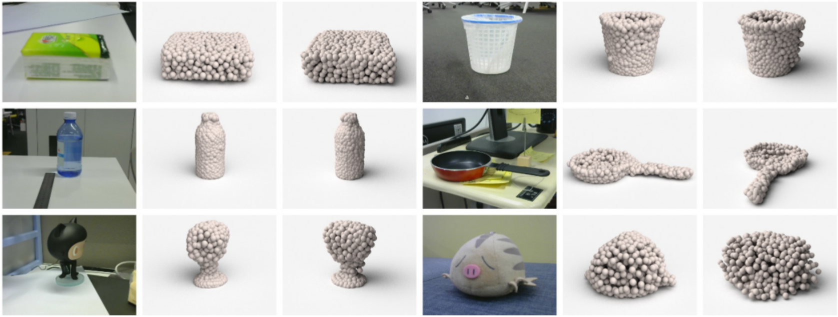


Figure: Multiple predictions on real examples; objects were masked out.

References I

D. P. Bertsekas.
A distributed asynchronous relaxation algorithm for the assignment problem.
In *1985 24th IEEE Conference on Decision and Control*, pages 1703–1704, 1985.

Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu.

References II

Shapenet: An information-rich 3d model repository.
CoRR, abs/1512.03012, 2015.

Christopher Bongsoo Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese.
3d-r2n2: A unified approach for single and multi-view 3d object reconstruction.
CoRR, abs/1604.00449, 2016.

References III

- 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.
  Pointnet: A 3d convolutional neural network for real-time object class recognition.

References IV

In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pages 1578–1584, 2016.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial networks. *CoRR*, abs/1406.2661, 2014.

References V

Diederik P. Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3581–3589, 2014.

References VI

- Diederik P. Kingma and Max Welling.
  Auto-encoding variational bayes.
  CoRR, abs/1312.6114, 2013.
- Mehdi Mirza and Simon Osindero.
  Conditional generative adversarial nets.
  CoRR, abs/1411.1784, 2014.
- Kihyuk Sohn, Honglak Lee, and Xinchen Yan.
  Learning structured output representation using deep conditional generative models.
  In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural*

References VII

Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, pages 3483–3491, 2015.