Technische Universität München

Point Cloud Generation from a Single 2D Image

Can Gümeli, Tymur Mirlas

can.gumeli@gmail.com, tymur.mirlas@tum.de

Introduction

- **Motivation:** generating point clouds from 2D images will enable future research to utilize 3D geometry in all image analysis challenges, ranging from scene retrieval to object detection.
- **Goal:** introduce an end-to-end deep learning architecture to generate point clouds from 2D images.
- **Mathematical formulation of the problem:**

$$S_{pred} = DNN_{\theta}(I_{rgb}) = \{(x_i, y_i, z_i)\}_{i=1}^{N}$$

- **Dataset:** synthetic benchmarks such as Shapenet [1] or ModelNet [2].

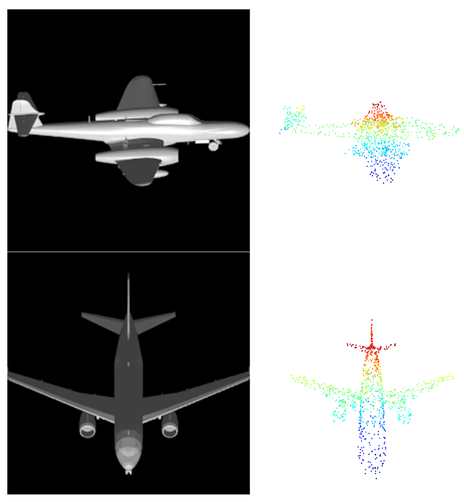

Figure 1. Examples of inputs and corresponding point clouds. Note, that we‘re interested in generating point clouds, which are aligned with the input. The color of points in a point cloud corresponds to their depth.

Method

- **Problem:** Exponentially increasing dimensionality of an output produced by a CNN with skip connections.
- **Solution:** Multi-branch CNN with the Spatial Transformer Network [3] as an attention mechanism.
- **Loss function – Chamfer distance:**

$$J_{\theta}(\hat{S}, S) = \sum_{x \in \hat{S}} \min_{y \in S} \|x - y\| + \sum_{x \in S} \min_{y \in \hat{S}} \|x - y\|$$

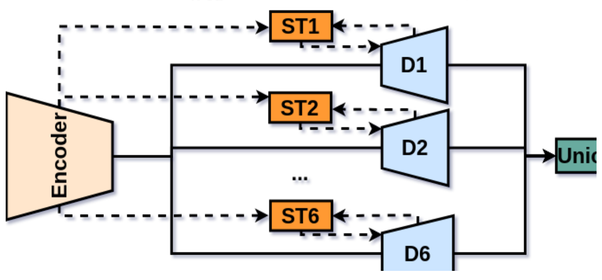

Figure 2. Multi-branch CNN for point cloud generation.

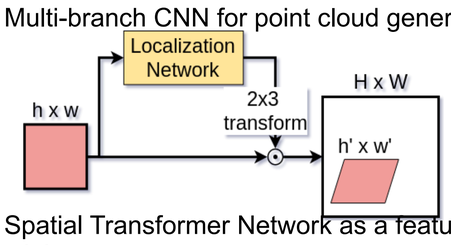

Figure 3. Spatial Transformer Network as a feature pooling module.

Results

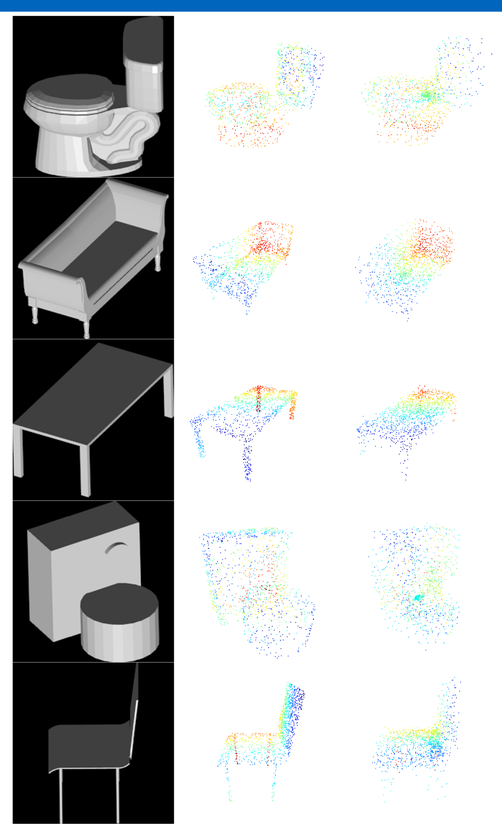

Figure 4. Predictions from the validation samples. **Input – GT – Predictions** from left to right respectively.

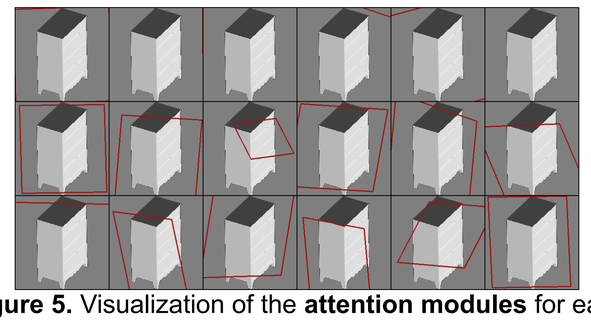

Figure 5. Visualization of the **attention modules** for each of 6 branches (each column corresponds to one branch).

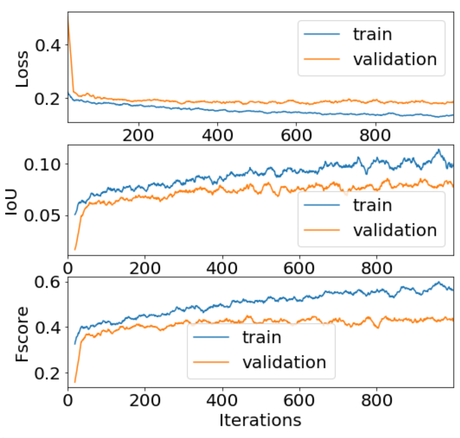

Figure 6. Learning curves.

	Loss		*Fscore*		*IoU*	
	Ours	**PSG**	**Ours**	**PSG**	**Ours**	**PSG**
bathtub	**0.1721**	0.1740	**0.4617**	0.4465	**0.0864**	0.0834
bed	**0.1499**	0.1543	**0.5328**	0.5092	**0.0876**	0.0846
chair	**0.1733**	0.1748	**0.4657**	0.4436	**0.0970**	0.0900
desk	**0.2140**	0.2182	**0.3703**	0.3445	**0.0578**	0.0552
dresser	**0.2100**	0.2142	**0.2998**	0.2839	**0.0403**	0.0360
monitor	0.1543	**0.1534**	**0.5502**	0.5376	**0.1082**	0.1010
night stand	**0.2446**	0.2447	**0.2561**	0.2395	0.0383	**0.0384**
sofa	**0.1435**	0.1466	**0.5259**	0.5139	**0.1029**	0.1002
table	**0.2255**	0.2374	**0.3513**	0.3232	0.0848	**0.0884**
toilet	**0.1613**	0.1646	**0.4606**	0.4300	**0.0815**	0.0743

Figure 7. Class-wise comparison of ours model vs. PointSetGenerationNetwork [4] trained on **ModelNet10**.

Final Layer. Ablation Study

The ablation study was conducted to choose the final layer. There were 3 options:
- Grid Deformation [5]
- Linear Transformation
- Tanh

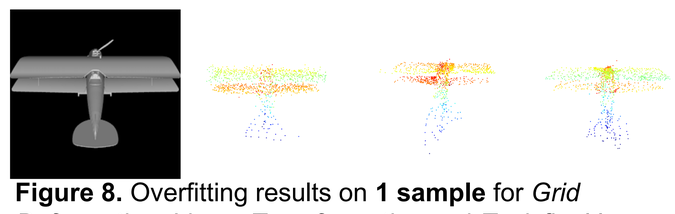

Figure 8. Overfitting results on **1 sample** for *Grid Deformation, Linear Transformation* and *Tanh* final layers.

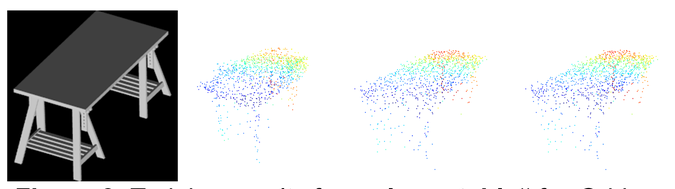

Figure 9. Training results for a **class „table“** for *Grid Deformation, Linear Transformation* and *Tanh* final layers.

Conclusions

- During the ablation study, we have checked the hypothesis, that final layer of the network affects its performance. As a conclusion, we can say, that final layer indeed matters only in case of overfitting to one sample, and doesn‘t matter in case of large-scale training of deep networks.
- As the baseline, we have the Point Set Generation Network, which utilizes convolutional layers to regress point clouds.
- We have reached the state of the art performance in comparison with the baseline. Moreover, our model has more capacity, is more interpretable, and scales better for architectural enhancements.

References

[1] Angel X. Chang, Thomas A. Funkhouser, Leonidas J.Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, SilvioSavarese, Manolis Savva, Shuran Song, Hao Su,Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. CoRR, abs/1512.03012, 2015.

[2] 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.

[3] Max Jaderberg, Karen Simonyan, Andrew Zisserman,and Koray Kavukcuoglu. Spatial transformer networks. CoRR, abs/1506.02025, 2015.

[4] 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.

[5] Chen-Hsuan Lin, Chen Kong, and Simon Lucey. Learning efficient point cloud generation for dense 3d object reconstruction. CoRR, abs/1706.07036, 2017.

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Computer Vision & Artificial Intelligence