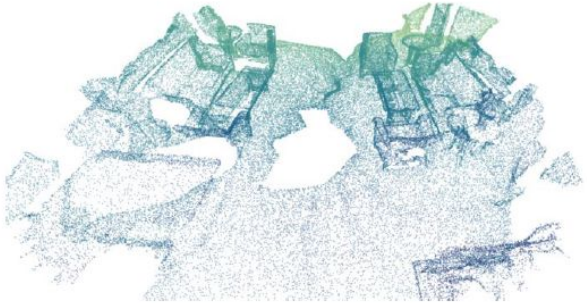
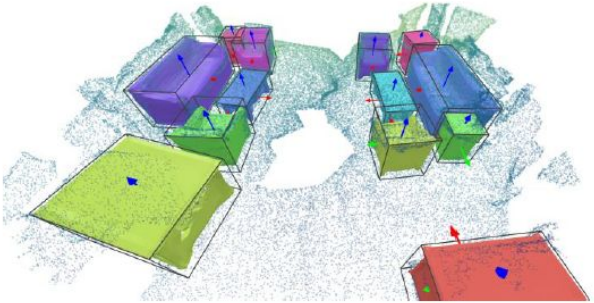
Weakly-supervised 3D Semantic Scene Reconstruction

Alexander Baumann, Sophia Wagner

Agenda

1. Recap
2. Results
3. Integration of Shape prior into full pipeline
   a. pretrained Shape Prior
   b. Shape retrieval
4. Outlook
5. References



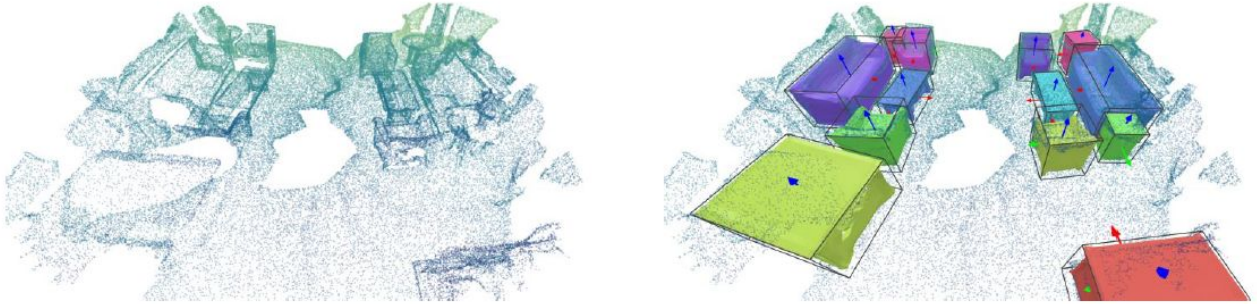


1. Recap

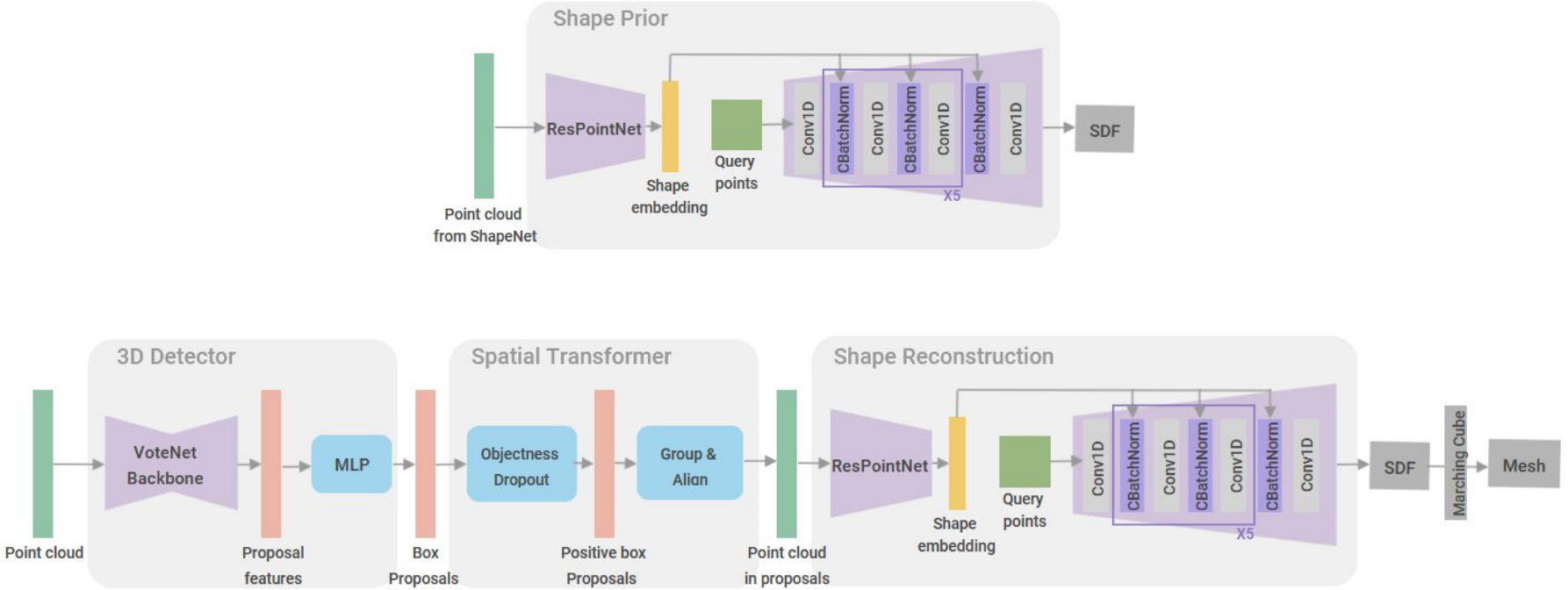
Weakly-supervised semantic scene reconstruction

- Input: Point cloud scan of a 3D indoor scene
- Output: Semantic instance meshes of 3D objects

Extract the features directly from the point cloud without voxelizing the scene



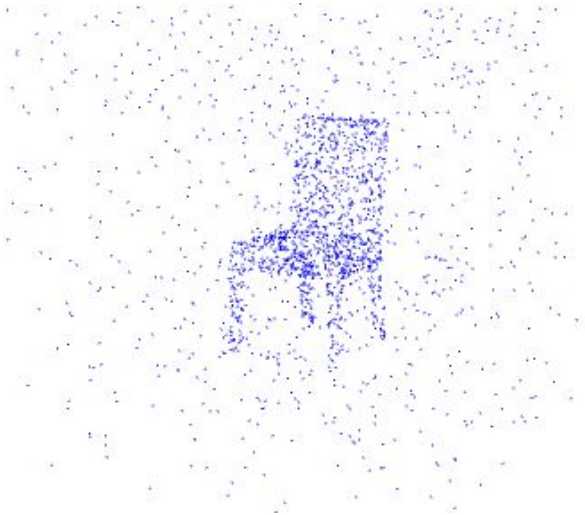
1. Recap


Shape Prior
Point cloud
from ShapeNet
ResPointNet
Shape
embedding
Query
points
Conv1D
CBatchNorm
Conv1D
CBatchNorm
Conv1D
CBatchNorm
Conv1D
X5
SDF
3D Detector
Point cloud
VoteNet
Backbone
Proposal
features
MLP
Box
Proposals
Spatial Transformer
Objectness
Dropout
Positive box
Proposals
Group &
Align
Point cloud
in proposals
Shape Reconstruction
ResPointNet
Shape
embedding
Query
points
Conv1D
CBatchNorm
Conv1D
CBatchNorm
Conv1D
CBatchNorm
Conv1D
X5
SDF
Marching Cube
Mesh

2. Results - Shape prior

ShapeNet Data preparation

- sampling point clouds from 3D object meshes using farthest point sampling
- sampling query points near the object surface and uniformly in a hypercube and calculate their Signed Distance Values



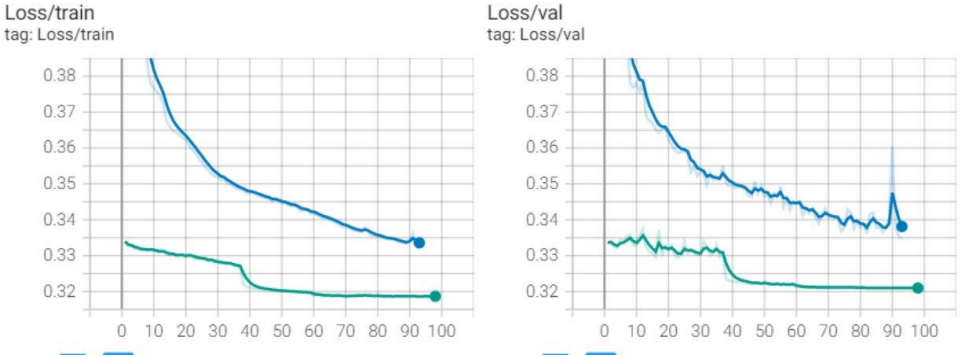
2. Results - Shape prior

- Implementation of the Shape Prior using the following loss:

$$\mathcal{L}(Q, S, E|f) = \frac{1}{N}\sum_{i=1}^{N} \|f(q_i|e_i) - sign(s_i)\|^2$$

query point — shape embedding — ground-truth SDF value

- Training of the Shape Prior

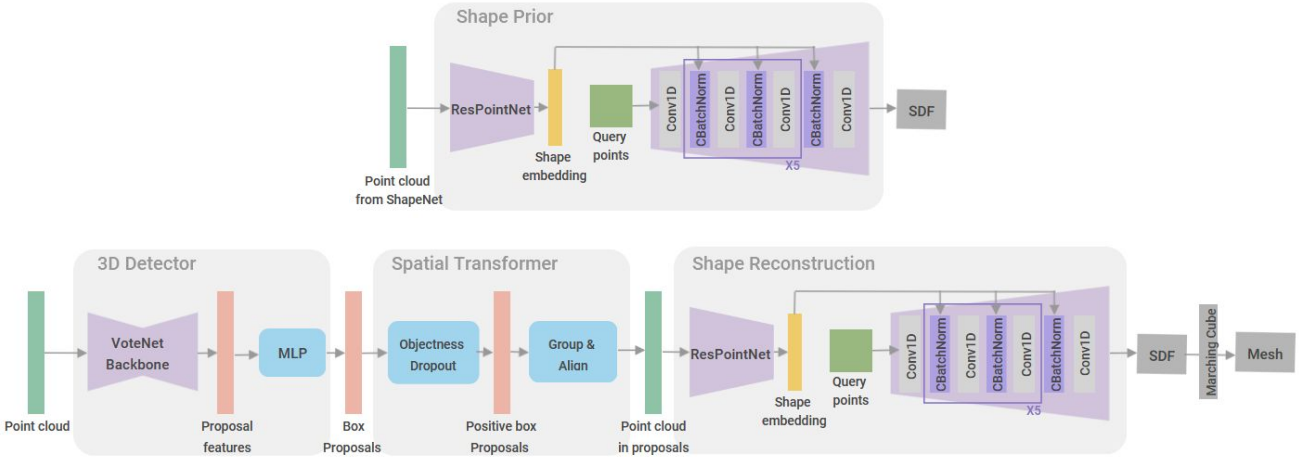


2. Results - Shape prior

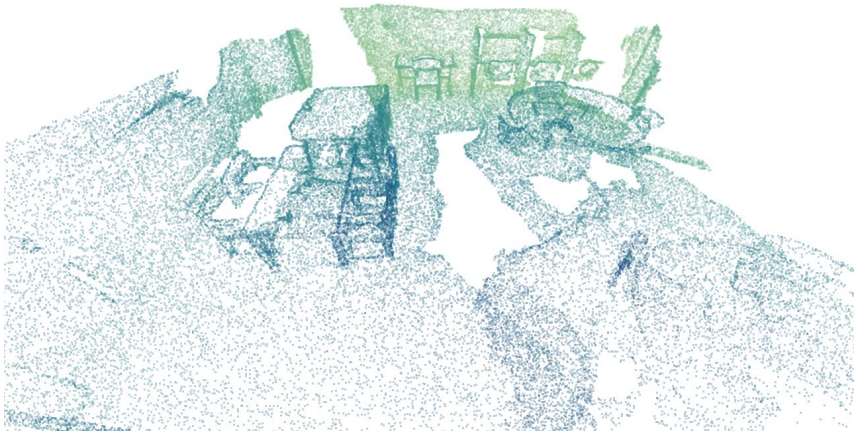
Creating 3D instance meshes from the Signed Distance Function

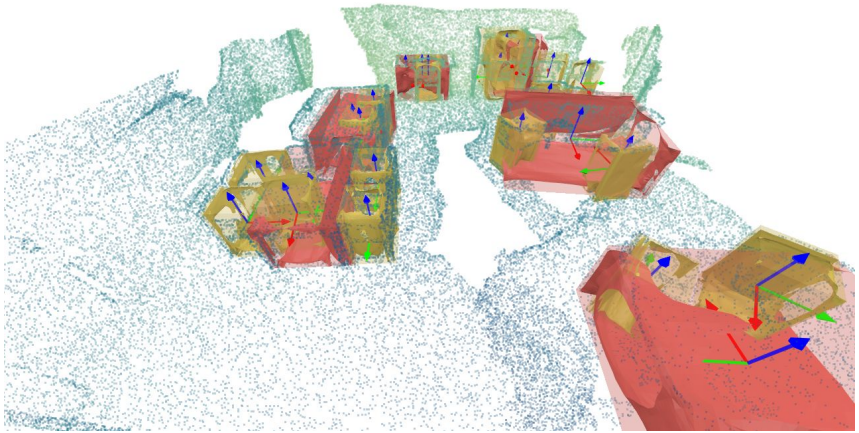
3. Integration of Shape prior into full pipeline - pretrained Shape prior

- Use the pretrained encoder & decoder of the shape prior for predicting the SDF
- Input point cloud looks significantly different to the input point cloud from the ShapeNet data



3. Integration of Shape prior into full pipeline - pretrained Shape prior





3. Integration of Shape prior into full pipeline - pretrained Shape prior

Category	table	chair	bookshelf	sofa	trash bin	cabinet	display	bathtub	Avg
IoU (RevealNet [1])	17.43	29.61	18.05	28.73	17.90	20.78	18.68	12.64	20.48
IoU (RfD-Net [2]) (best)	19.22	37.05	35.14	46.73	36.92	49.25	39.28	32.55	37.02
IoU (Ours)	12.72	15.31	14.14	19.01	26.23	29.81	12.52	17.46	18.40

3. Integration of Shape prior into full pipeline - Shape retrieval

- Freeze only the pretrained decoder for predicting the SDF from a shape embedding
- Extract the shape embedding via ResPointNet (trainable)
- Retrieval of shape embeddings between the detected objects in Scannet scene and the corresponding ShapeNet category

3. Integration of Shape prior into full pipeline - Shape retrieval

Using a triplet loss from [3]:

$$L_{ret} = \max\left(\|z_i - \tilde{z}_i^{+}\|_2^2 - \|z_i - \tilde{z}_i^{-}\|_2^2 + 0.5,\ 0\right)$$

z_i : extracted shape embedding of object with predicted semantic class x
$\tilde{z}_i^{+}$: mean of shape embeddings of class x from ShapeNet data
$\tilde{z}_i^{-}$: mean of shape embeddings of class $y \neq x$ from ShapeNet data

4. Outlook

- Train shape retrieval method
- Improve Shape Prior
- Compare our results to state-of-the-art results

5. References

[1]: Hou, Ji, Angela Dai, and Matthias Nießner. "Revealnet: Seeing behind objects in rgb-d scans." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

[2]: Nie, Yinyu, et al. "Rfd-net: Point scene understanding by semantic instance reconstruction." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

[3]: Gümeli, Can, Angela Dai, and Matthias Nießner. "ROCA: Robust CAD Model Retrieval and Alignment from a Single Image." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.