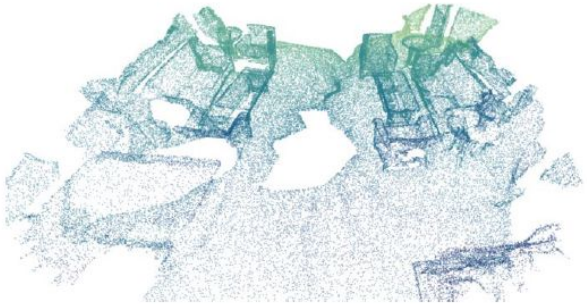
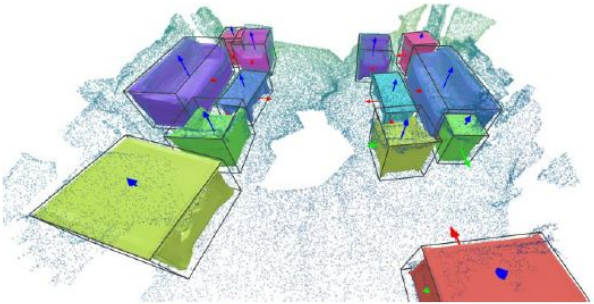
Weakly-supervised 3D Semantic Scene Reconstruction

Alexander Baumann, Sophia Wagner

Agenda

1. Introduction
2. Related work
3. Approach
4. Results
5. Outlook
6. References





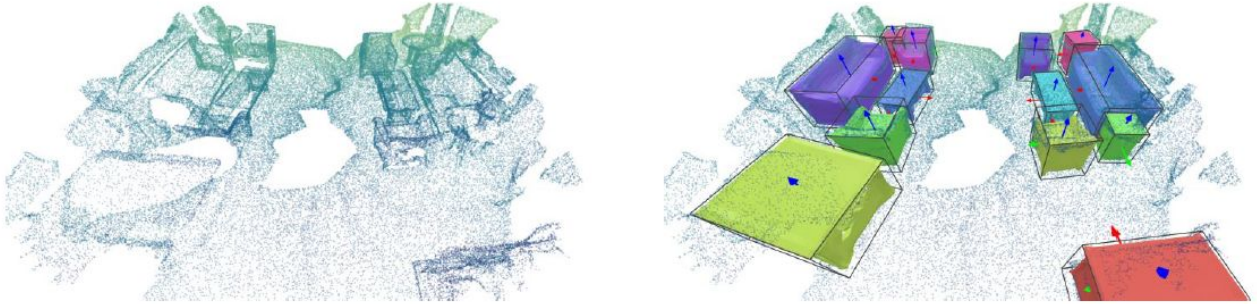
1. Introduction

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

Challenge: full 3D geometry of objects often unknown due to occlusions



2. Related work

RfD-Net: Point Scene Understanding by Semantic Instance Reconstruction [1]

- Reconstruction of 3D objects directly from raw point clouds
  ➔ efficient learning
- Decoupling the problem into global object detection and local shape prediction
- Drawback: requiring full supervision

2. Related work

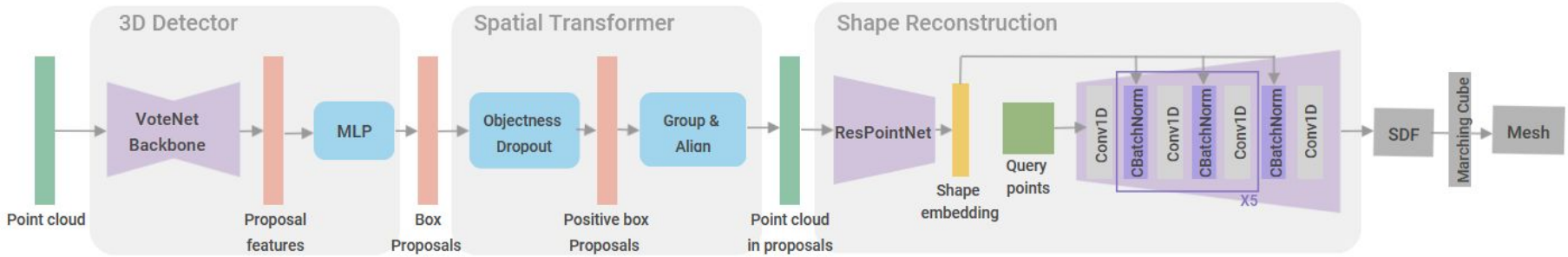
DOPS: Learning to Detect 3D Objects and Predict their 3D Shapes [2]

- Weakly-supervised approach for detecting and estimating 3D shapes
- Training a Shape Prior on a synthetic dataset
  ➜ no ground-truth shape information required
- Drawback: working with voxelized scenes

3. Approach

1. 3D Detection

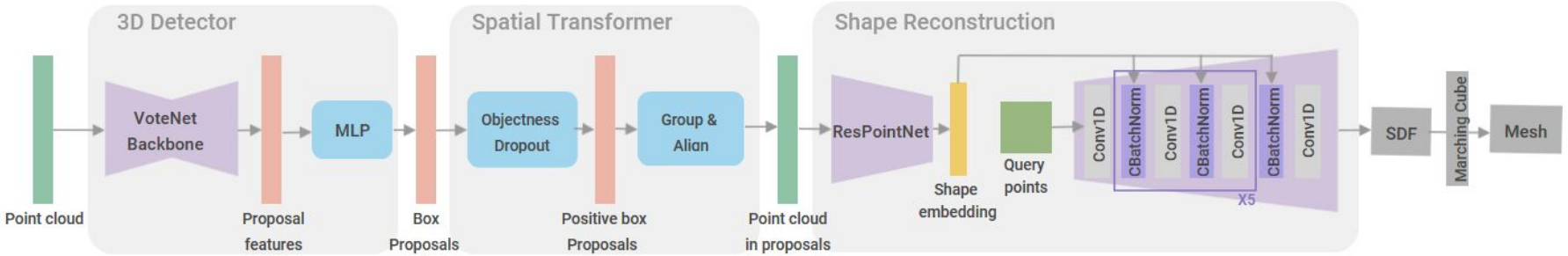
- VoteNet [3] & MLP: for each box predict center, scale, heading angle, semantic label and objectness score (indicates if the predicted box is close to any ground-truth object)
- supervised using ScanNet Data [4]



3. Approach

2. Spatial Transformation

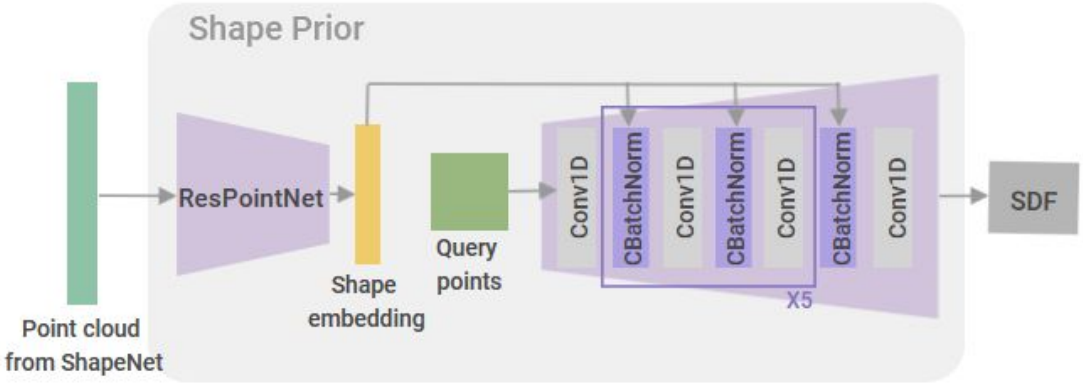
- Objectness Dropout: only keep the box proposals with the highest objectness score
- Group: sample points from point cloud within the box proposals and cluster them
- Align: normalize each point cluster



3. Approach

3. Shape Reconstruction

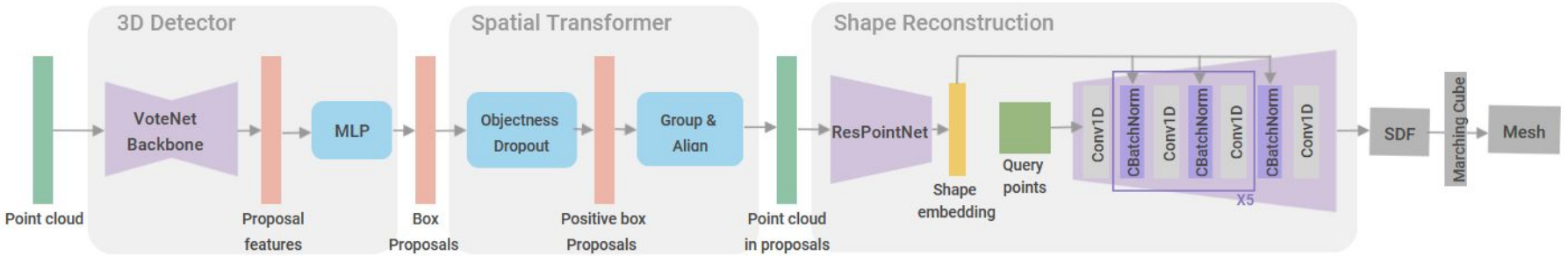
- Train a Shape Prior on the synthetic ShapeNet [5] dataset
- ResPointNet [6] encoder: extract a shape embedding from a sampled point cloud
- Decoder: predict Signed Distance Function values from query points conditioning on the shape embedding



3. Approach

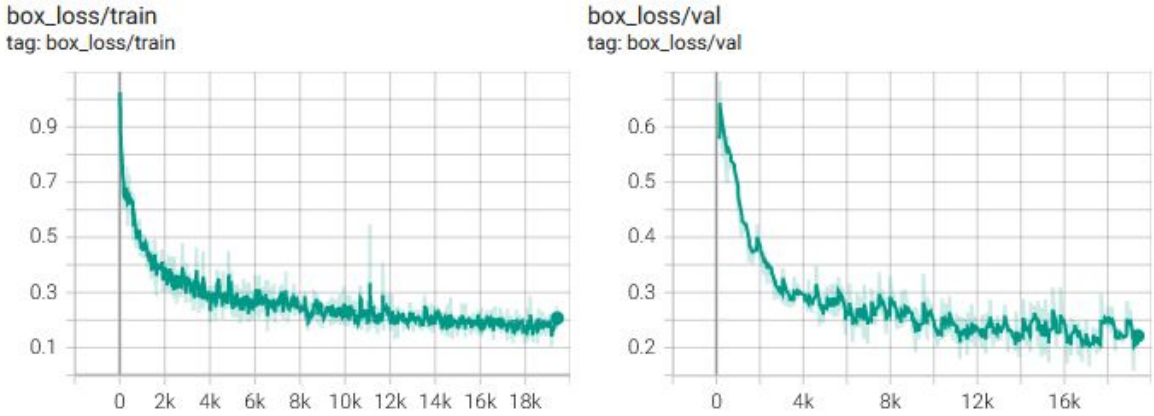
3. Shape Reconstruction

- Testing: use the pre-trained Shape Prior to predict the Signed Distance Function from the input point cloud and to reconstruct the mesh
  (here: ScanNet dataset)



4. Results

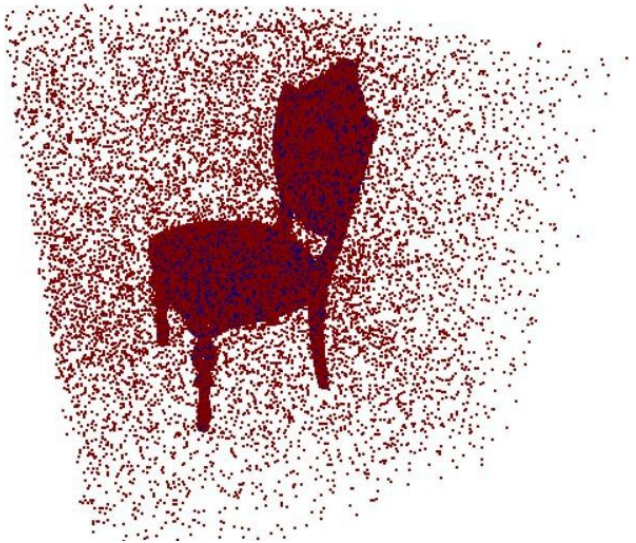
Training of 3D Detection on ScanNet data



4. Results

ShapeNet Data preparation

- sampling point clouds from 3D object meshes
- sampling query points near the object surface and calculate Signed Distance Values



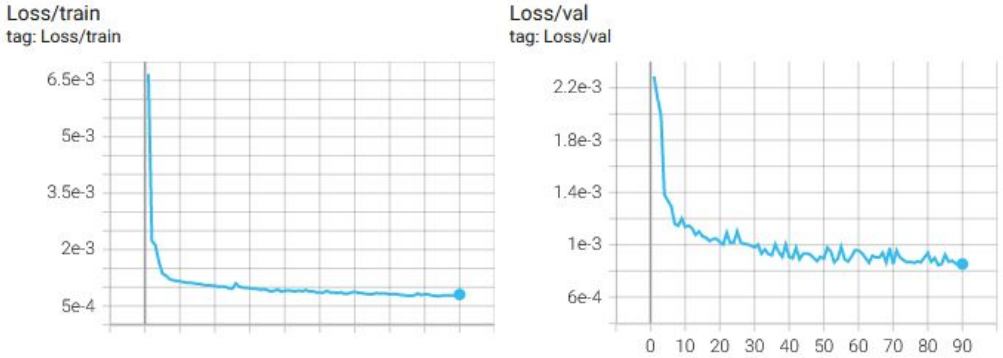
4. Results

- Implementation of the Shape Prior using the following loss:

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

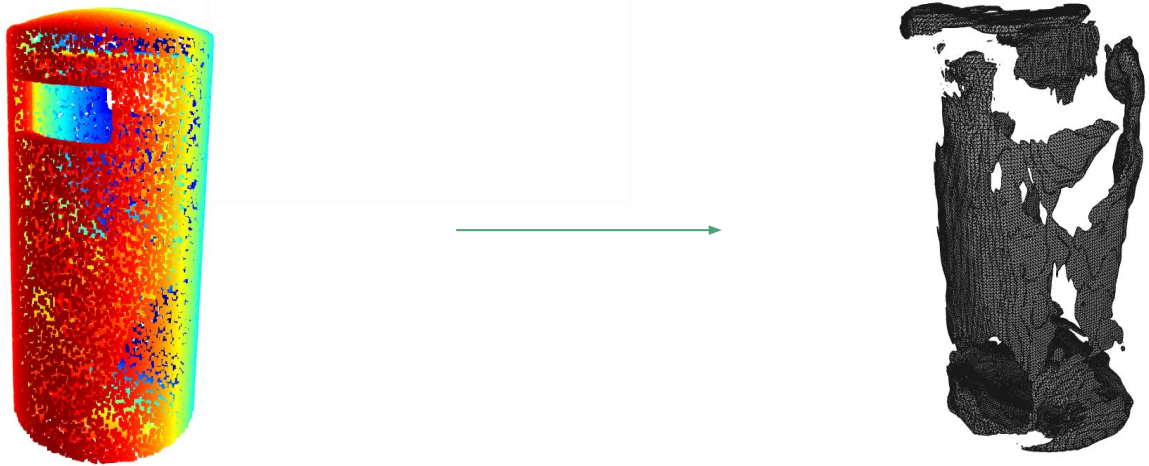
query point — shape embedding — ground-truth SDF value

- Training of the Shape Prior for 90 epochs



4. Results

Creating 3D instance meshes from the Signed Distance Function



5. Outlook

- Improve Shape Prior training (using different losses, tune hyperparameters)
- Create 3D instance meshes from Signed Distance Function
- Integrate trained Shape Prior into full pipeline
- Compare our results to state-of-the-art results

6. References

[1]: 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.

[2]: Najibi, Mahyar, et al. "Dops: Learning to detect 3d objects and predict their 3d shapes." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

[3]: Qi, Charles R., et al. "Deep hough voting for 3d object detection in point clouds." *proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

[4]: Dai, Angela, et al. "Scannet: Richly-annotated 3d reconstructions of indoor scenes." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[5]: Chang, Angel X., et al. "Shapenet: An information-rich 3d model repository." *arXiv preprint arXiv:1512.03012* (2015).

[6]: Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.