Mask3D 🎭
Mask Transformer for 3D Instance Segmentation

published at ICRA 2023

spotlight presentation at the T4V Workshop, CVPR 2022 and Urban3D, ECCV 2022

Jonas Schult[1], Francis Engelmann[2,3], Alexander Hermans[1], Or Litany[4], Siyu Tang[3], Bastian Leibe[1]

[1]RWTH Aachen University, [2]ETH AI Center, [3]ETH Zurich, [4]NVIDIA

Paper   Code   Video   ScanNet Visualizations

ScanNet200 Visualizations   S3DIS Visualizations

STPLS3D Visualizations   Try it out on your own scans!

Input 3D Scene   Instance Heatmaps   3D Semantic Instances

Mask3D predicts accurate 3D semantic instances achieving state-of-the-art on ScanNet, ScanNet200, S3DIS and STPLS3D.

News 🗞️

- **Mar 2023**: Video presenting Mask3D released.
- **Jan 2023**: Mask3D is accepted at ICRA'23.

- **Oct 2022**: Mask3D ranks 2nd on the STPLS3D Challenge (Urban3D Workshop) at ECCV'22. (Talk)
- **Oct 2022**: Mask3D preprint released on arXiv.
- **Sep 2022**: Code released.
- **Jun 2022**: A precursor of Mask3D is accepted at the T4V Workshop, CVPR'22. (Paper, Poster, Talk)

Abstract

Modern 3D semantic instance segmentation approaches predominantly rely on specialized voting mechanisms followed by carefully designed geometric clustering techniques. Building on the successes of recent Transformer-based methods for object detection and image segmentation, we propose a Transformer-based approach for 3D semantic instance segmentation. We show that we can leverage generic Transformer building blocks to directly predict instance masks from 3D point clouds. In our model called Mask3D each object instance is represented as an *instance query*. Using Transformer decoders, the instance queries are learned by iteratively attending to point cloud features at multiple scales. Combined with point features, the instance queries directly yield all instance masks in parallel. Mask3D has several advantages over current state-of-the-art approaches, since it neither relies on (1) voting schemes which require hand-selected geometric properties (such as centers) nor (2) geometric grouping mechanisms requiring manually-tuned hyper-parameters (e.g. radii) and (3) enables a loss that directly optimizes instance masks. Mask3D sets a new state-of-the-art on ScanNet test (+6.2 mAP), S3DIS 6-fold (+10.1 mAP), STPLS3D (+10.9 mAP) and the recent ScanNet200 test (+12.4 mAP).

Video 🎬

Mask3D: Mask Transformer for 3D Instance Segmentati...

Visual Comparison to SoftGroup

We compare Mask3D with SoftGroup, the currently best performing voting-based 3D instance segmentation approach. We highlight two error cases for SoftGroup and show Mask3D for comparison.
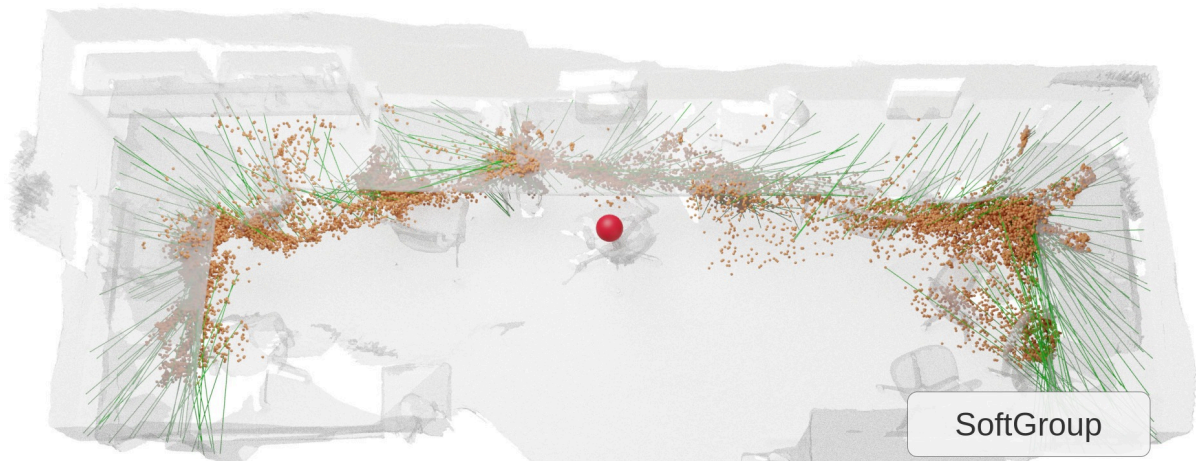
Example 1: Large Non-Convex Shapes

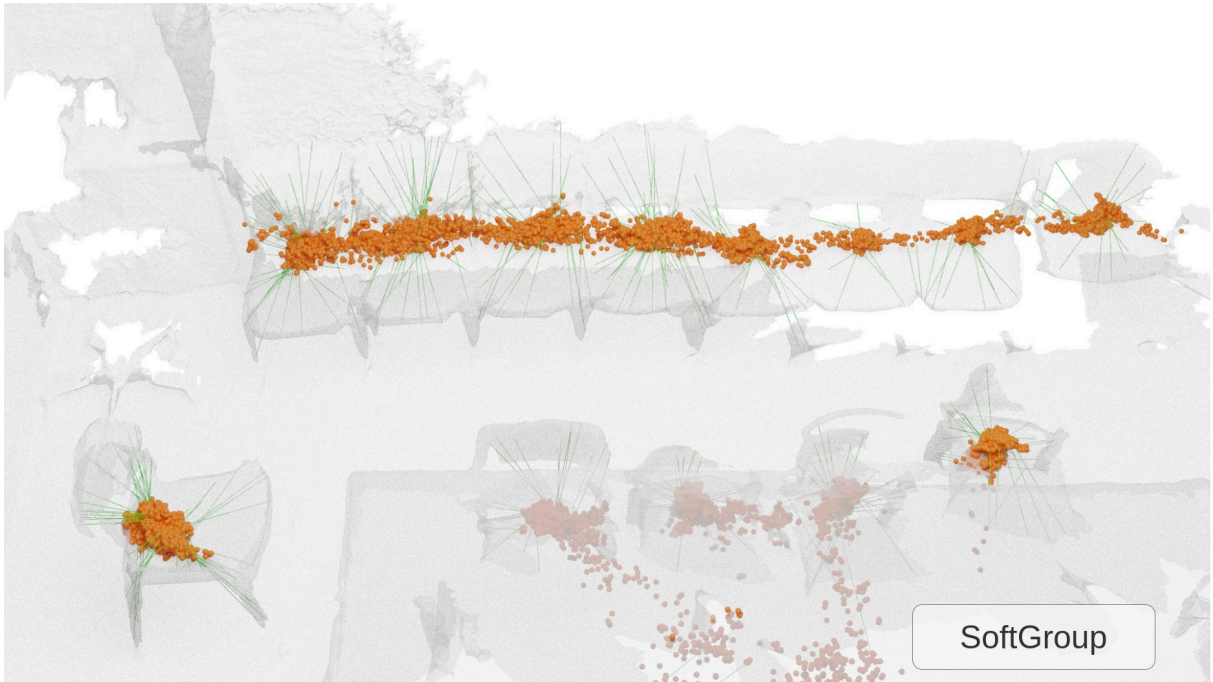SoftGroup

Example 2: Nearby Instances

A Closer Look: Paradigm Comparison

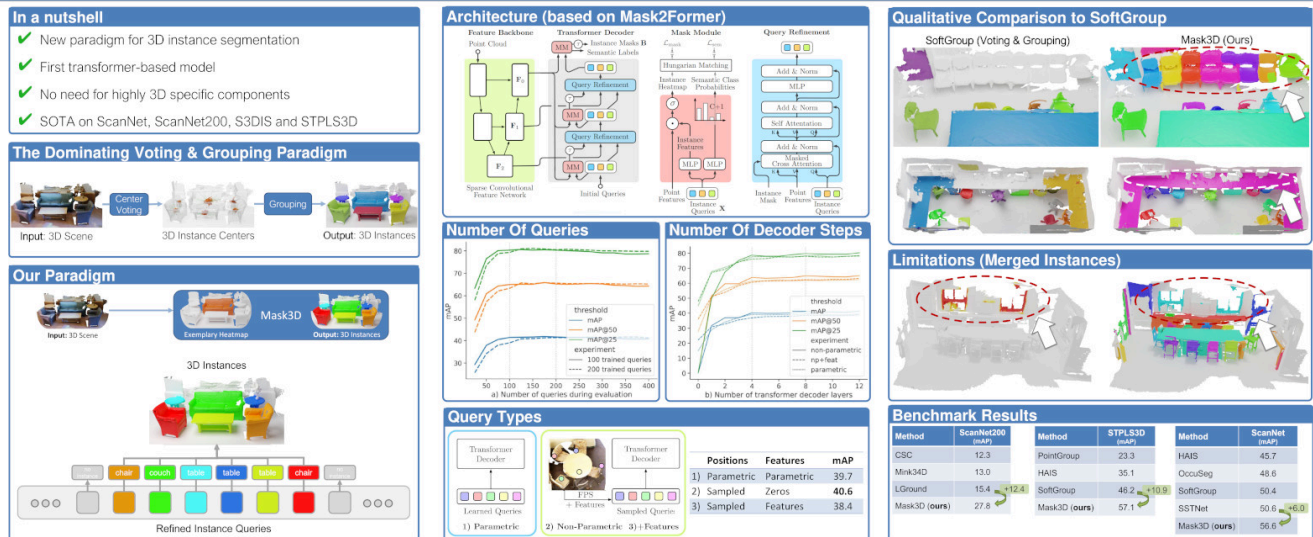## Example 1: Large Non-Convex Shapes

Example 2: Nearby Instances

Conference Paper

This work has been accepted at ICRA 2023.

Conference Poster

Mask3D: Mask Transformer for 3D Instance Segmentation

Jonas Schult[1] Francis Engelmann[2,3] Alexander Hermans[1] Or Litany[4] Siyu Tang[3] Bastian Leibe[1]

[1]RWTH Aachen University   [2]ETH AI Center   [3]ETH Zürich   [4]NVIDIA

https://JonasSchult.github.io/Mask3D/

BibTeX 🙏

```
@article{Schult23,
    title     = {{Mask3D: Mask Transformer for 3D Semantic Instance Segmentation}},
    author    = {Schult, Jonas and Engelmann, Francis and Hermans, Alexander and
    booktitle = {{International Conference on Robotics and Automation (ICRA)}},
    year      = {2023}
}
```