

Audio-Visual Floorplan Reconstruction

Senthil Purushwalkam¹, Sebastian Vicenc Amengual Gari², Vamsi Krishna Ithapu², Carl Schissler², Philip Robinson², Abhinav Gupta¹³, Kristen Grauman³⁴

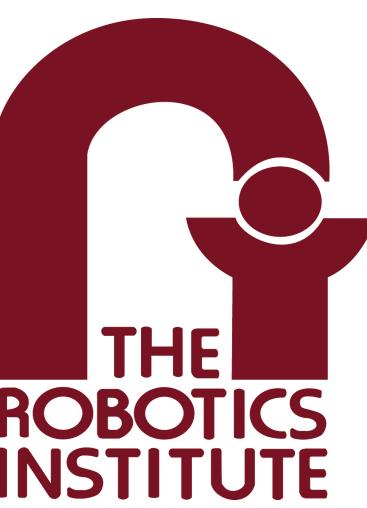
¹Carnegie Mellon University

²Facebook Reality Labs

³Facebook AI Research

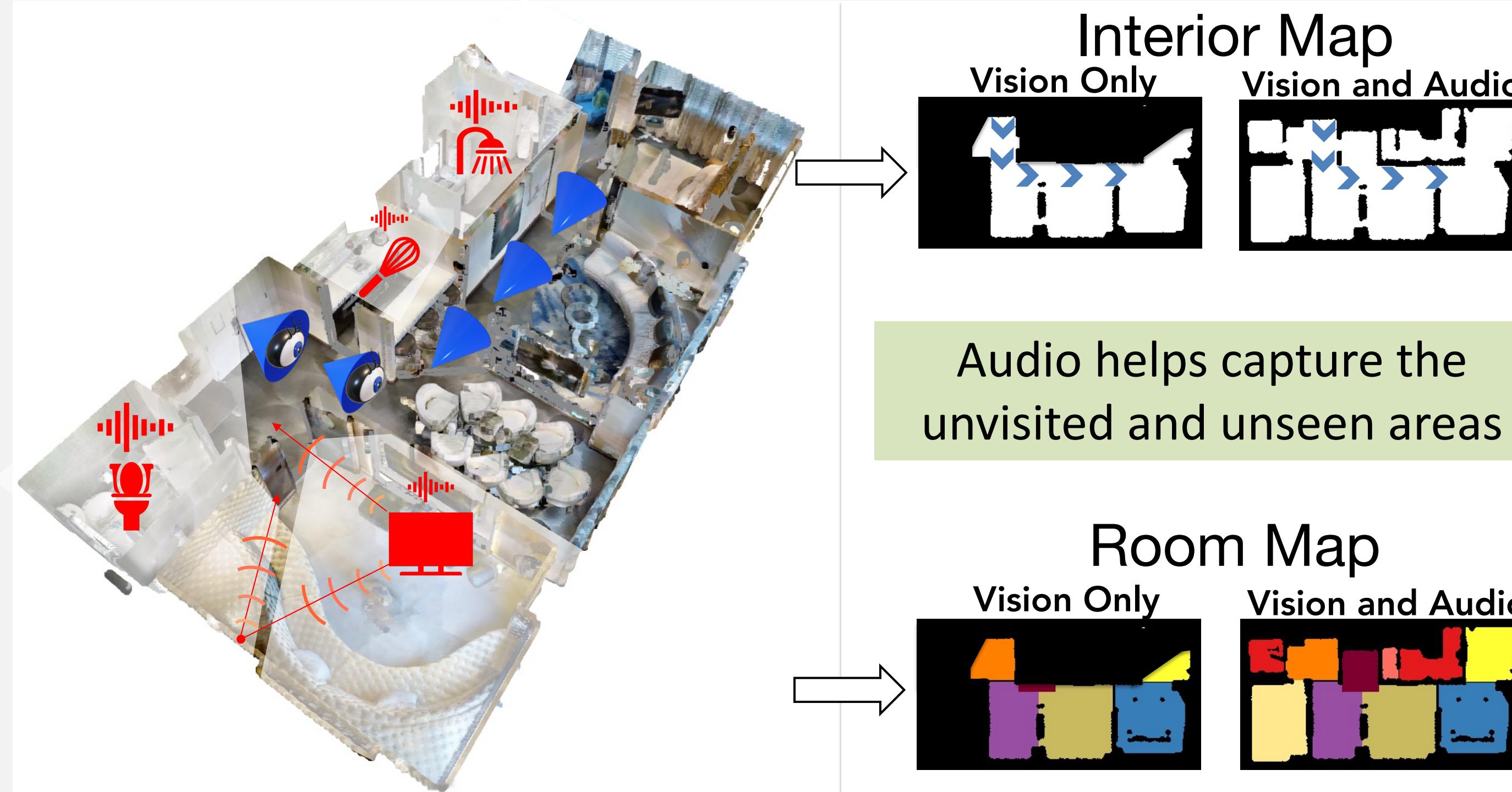
⁴University of Texas at Austin

Code: <http://www.cs.cmu.edu/~spurushw/publication/avmap/>

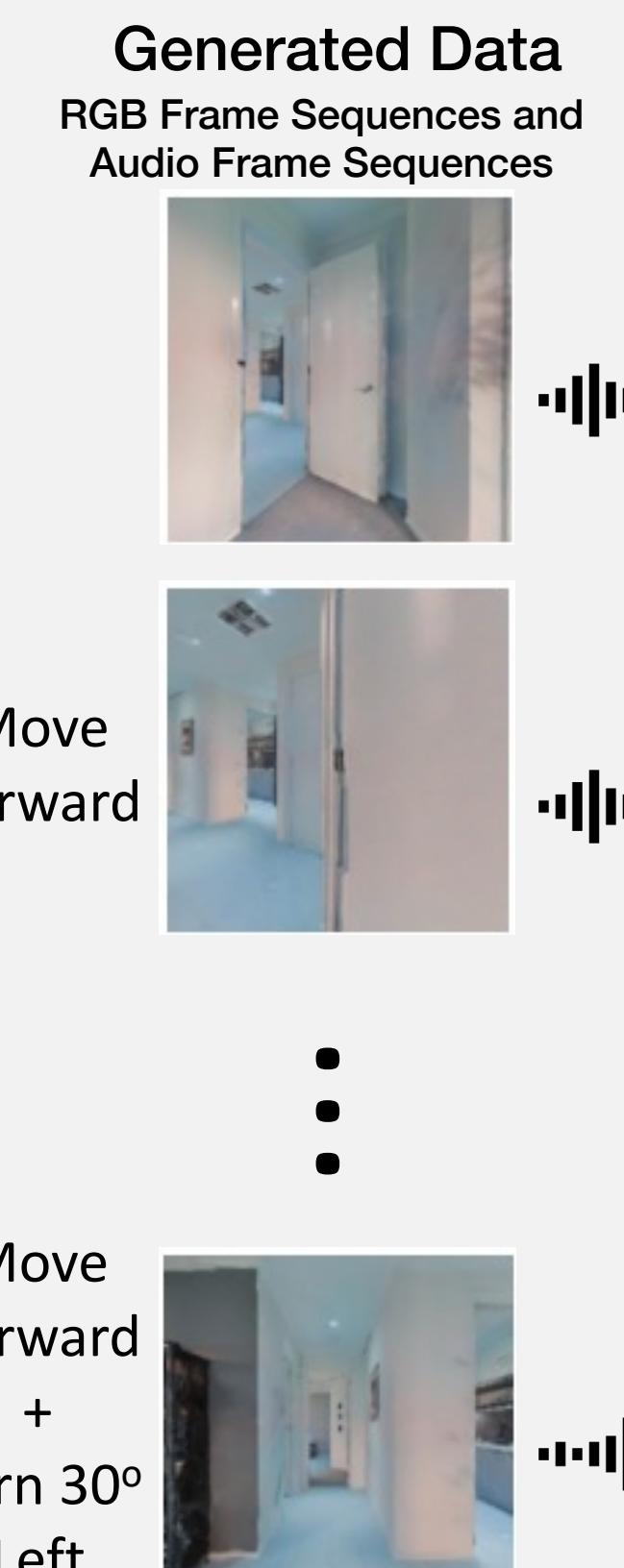
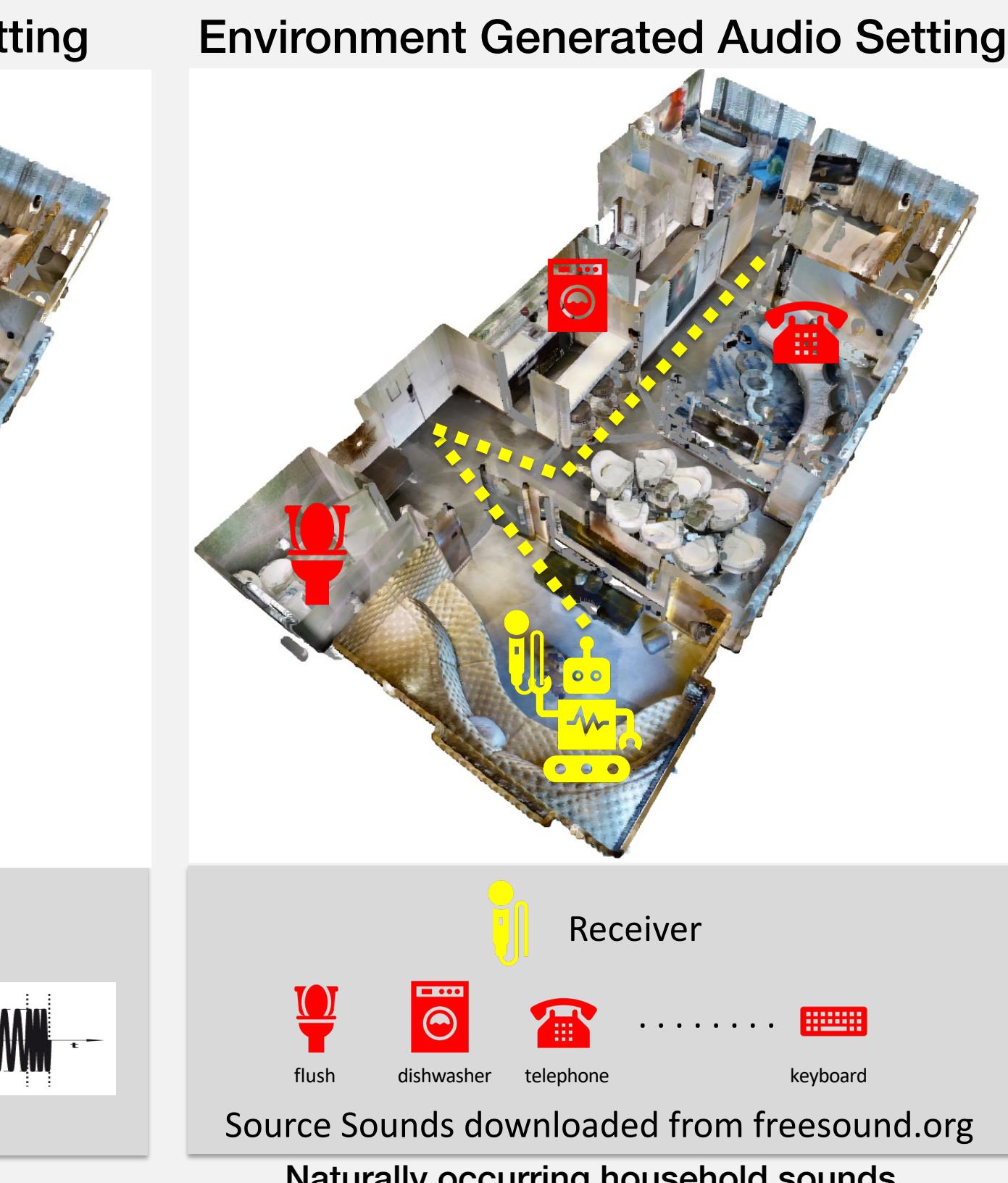
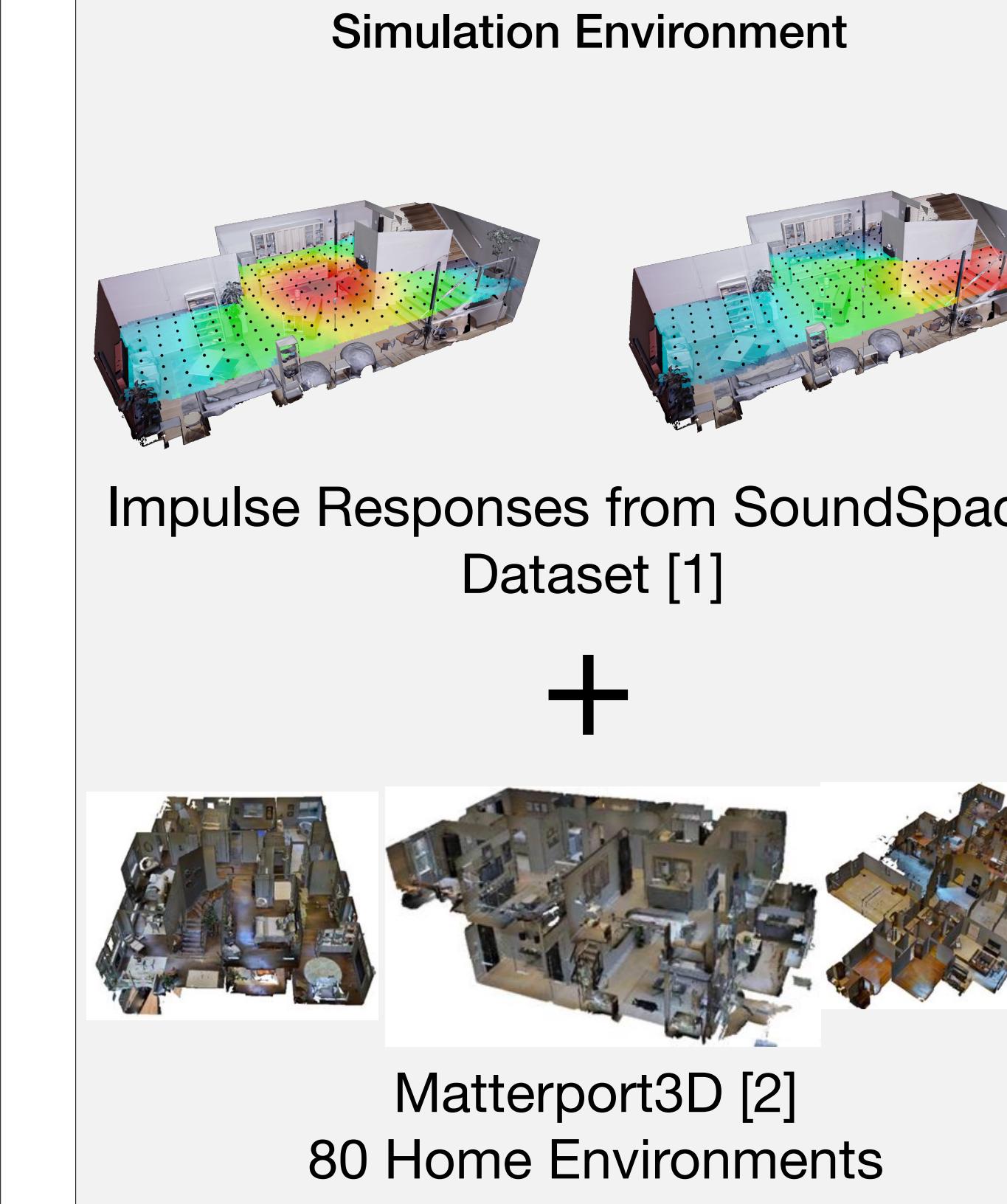


Overview

Goal: given a short video, infer a full floorplan



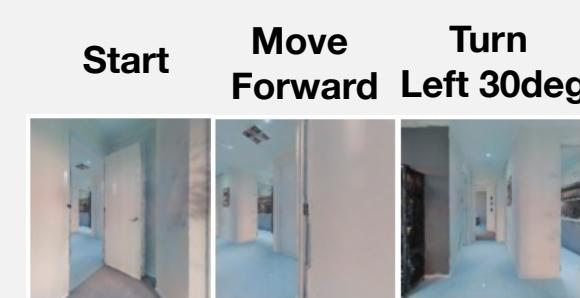
Generating Realistic Audio-Visual Data



AV-Map Model

Input: RGB+Audio Sequence

RGB Feature Extraction



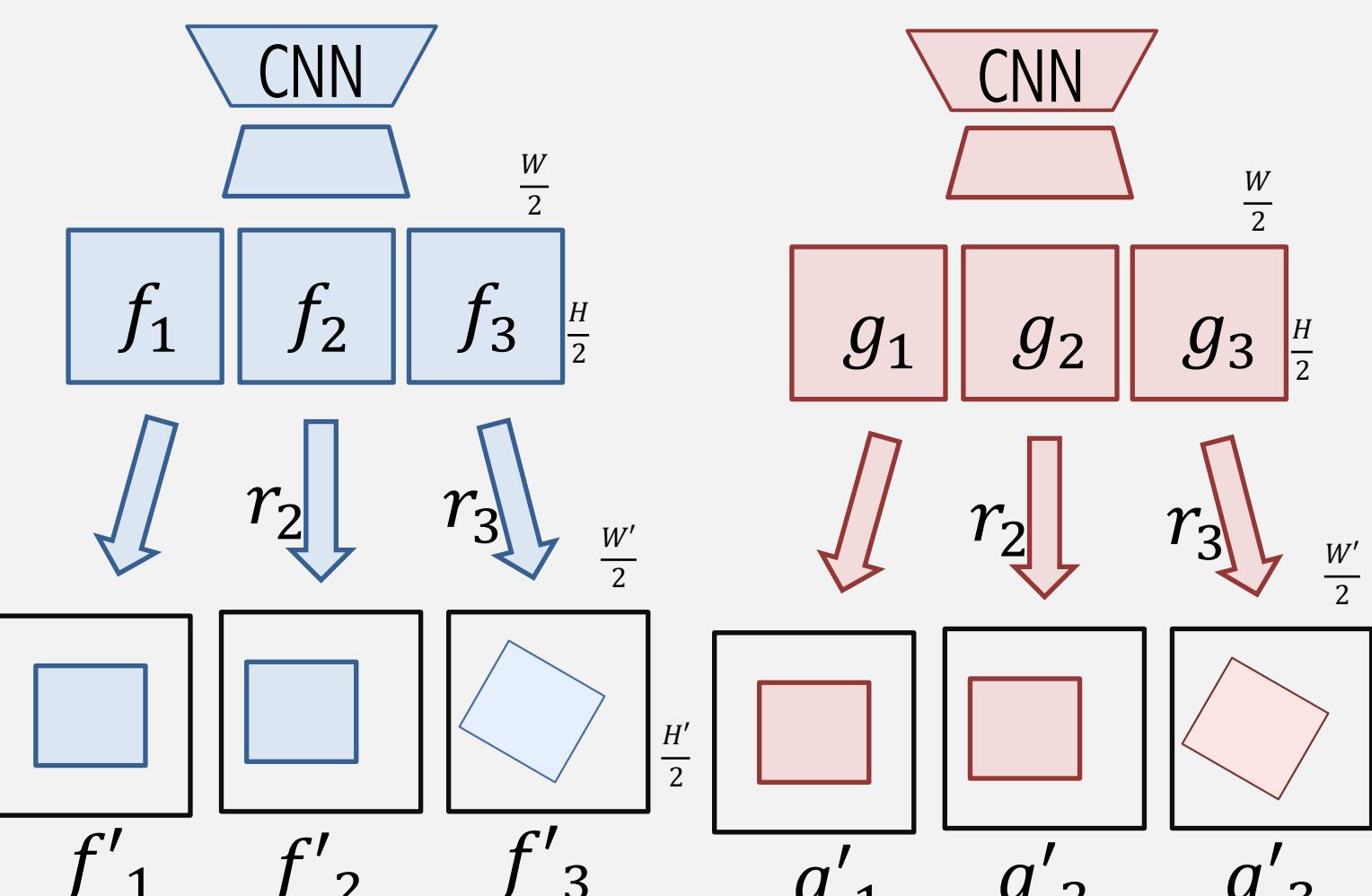
x_1, x_2, x_3

Audio Feature Extraction

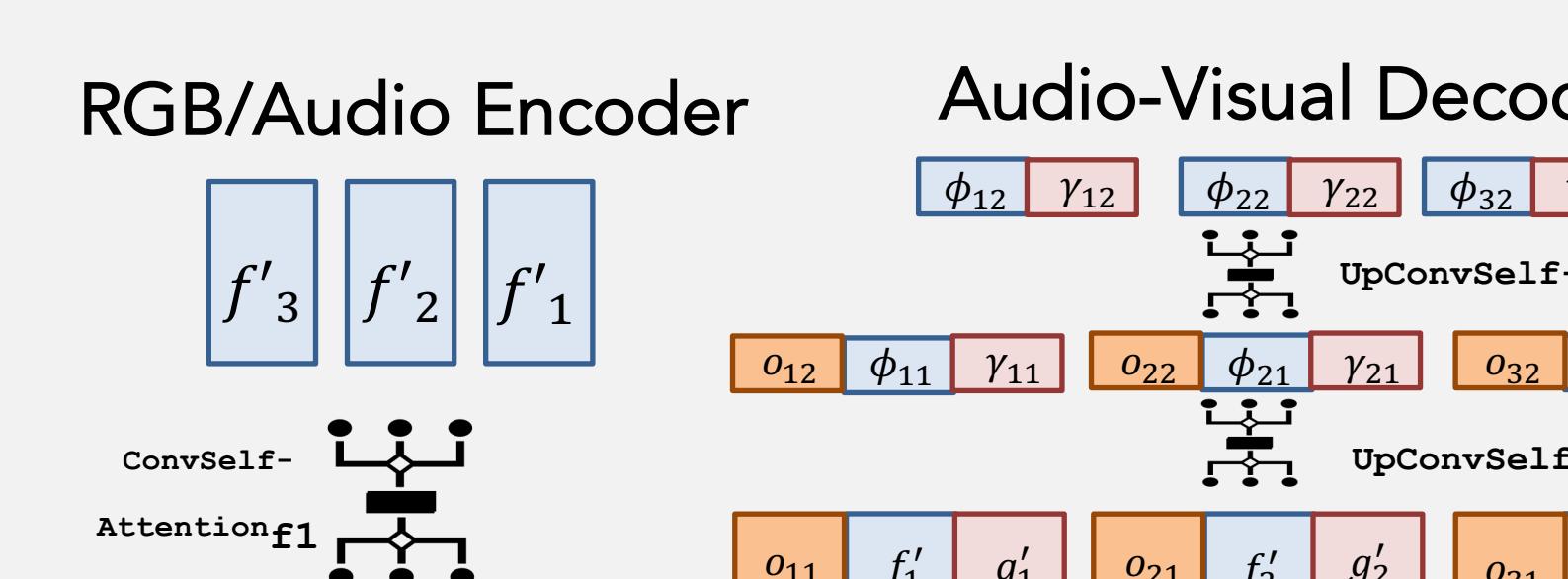


r_1, r_2, r_3

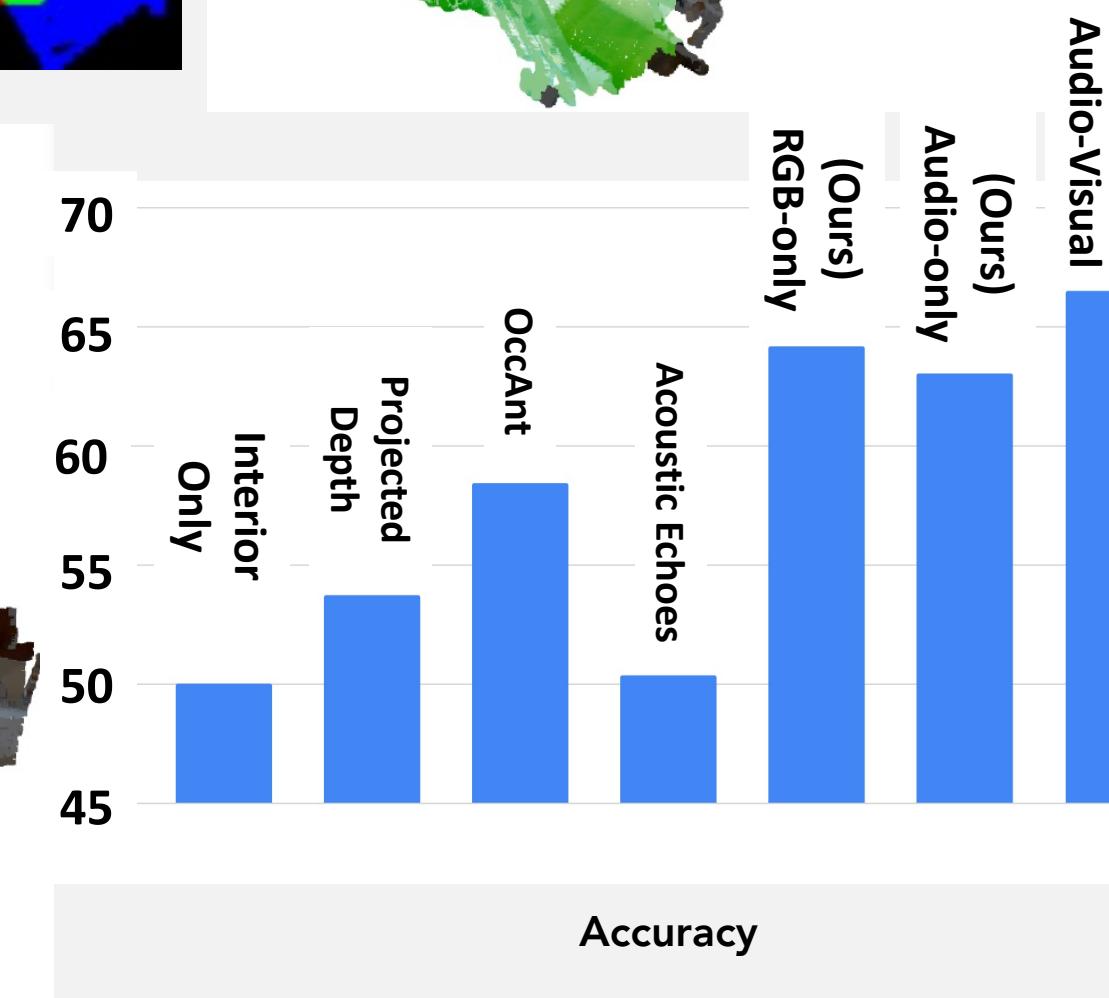
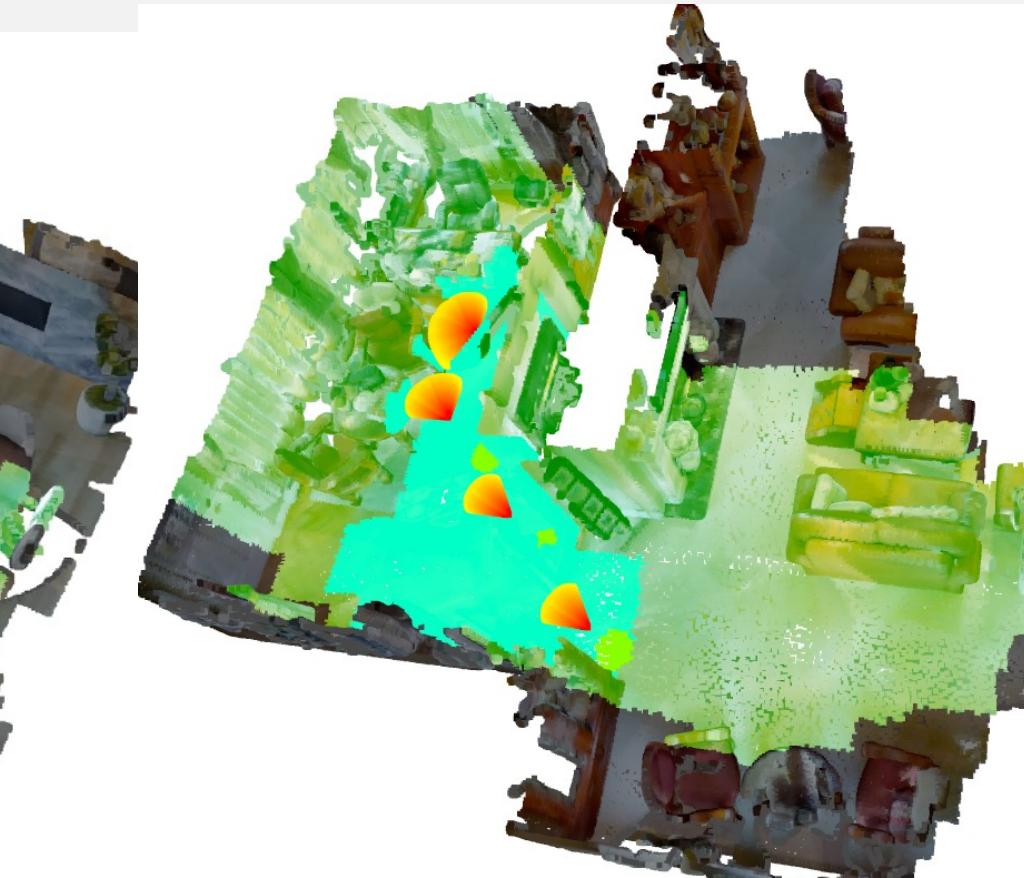
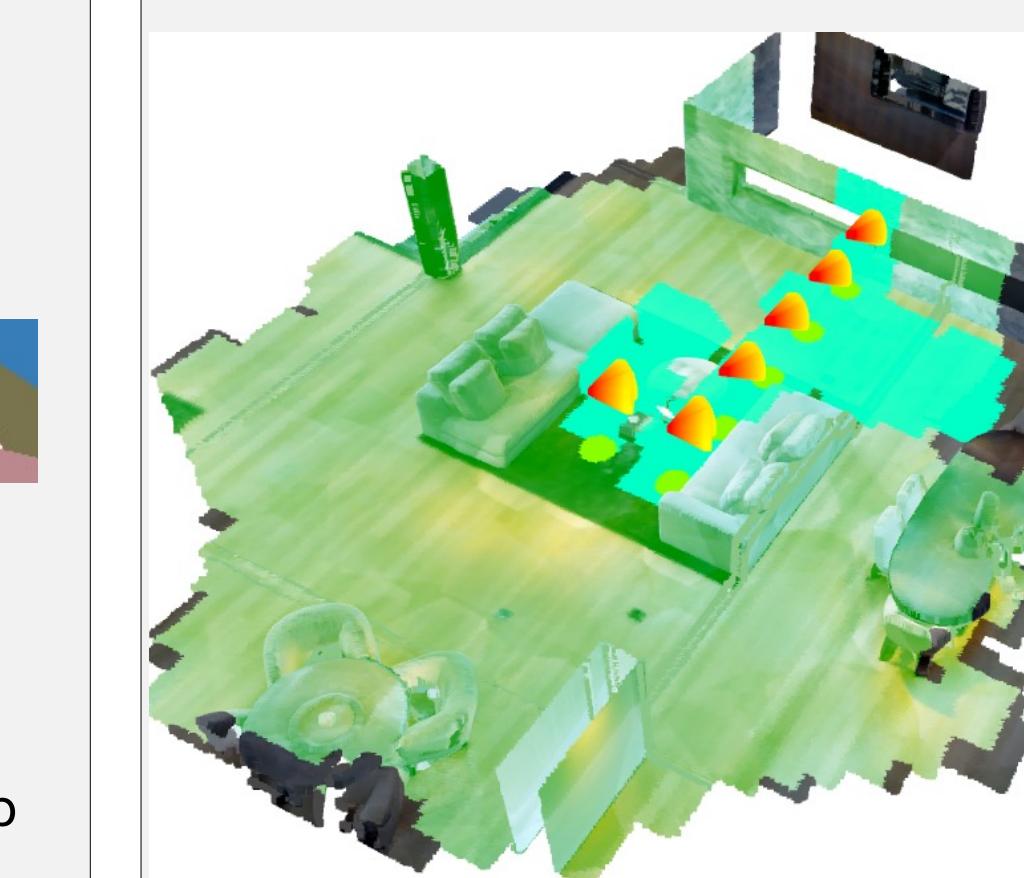
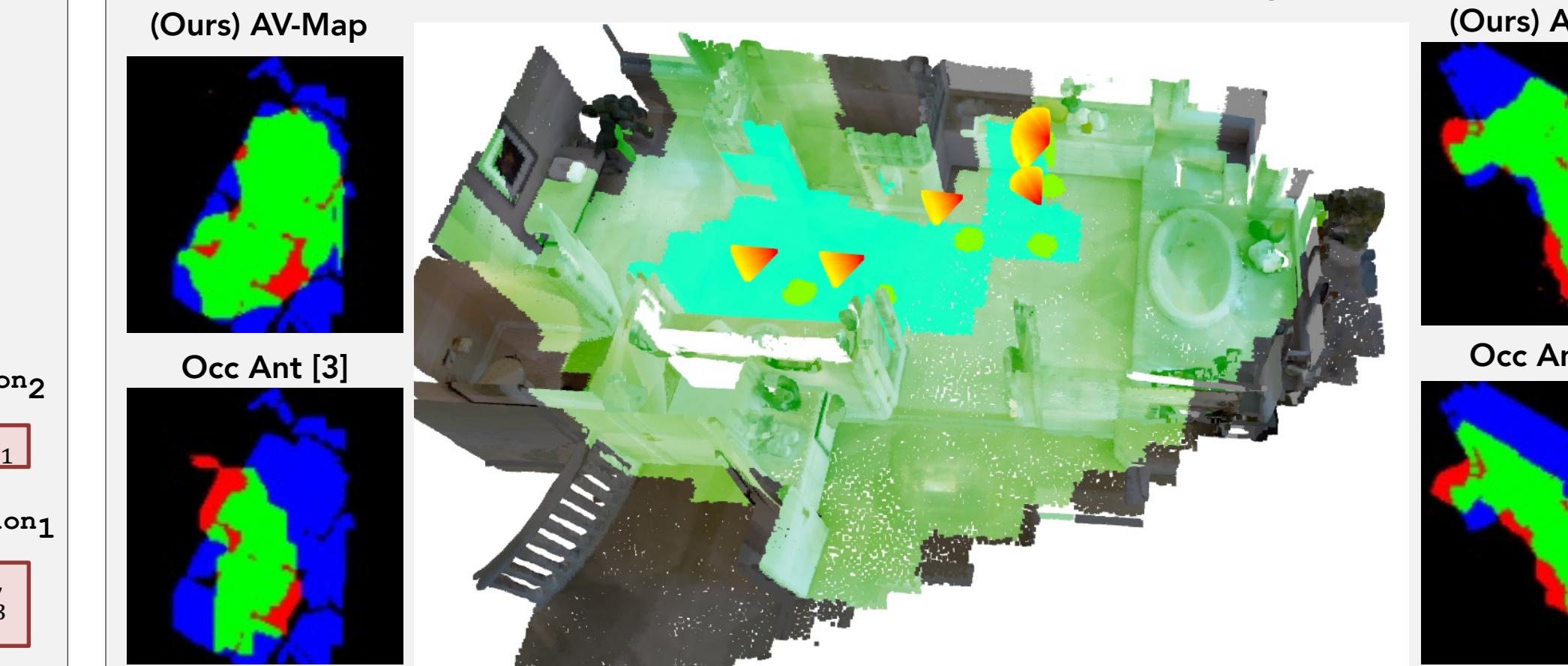
Feature Alignment



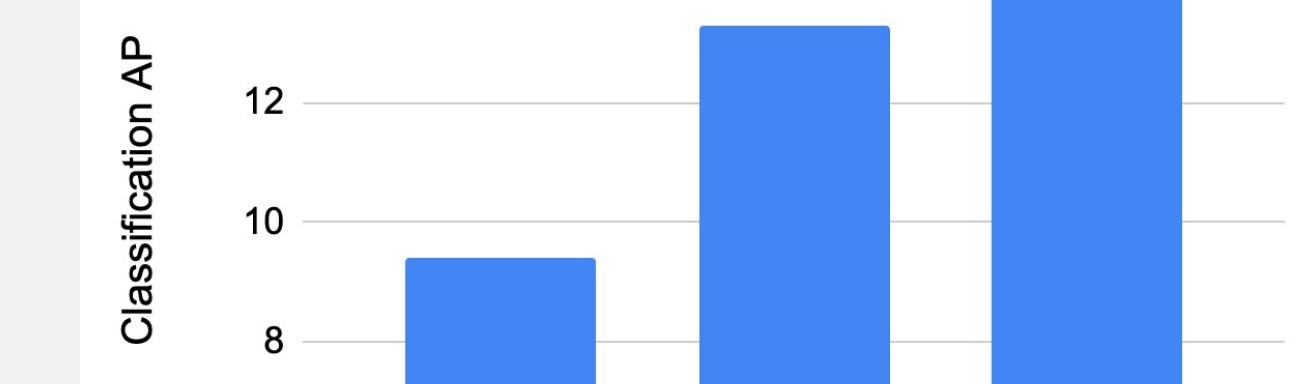
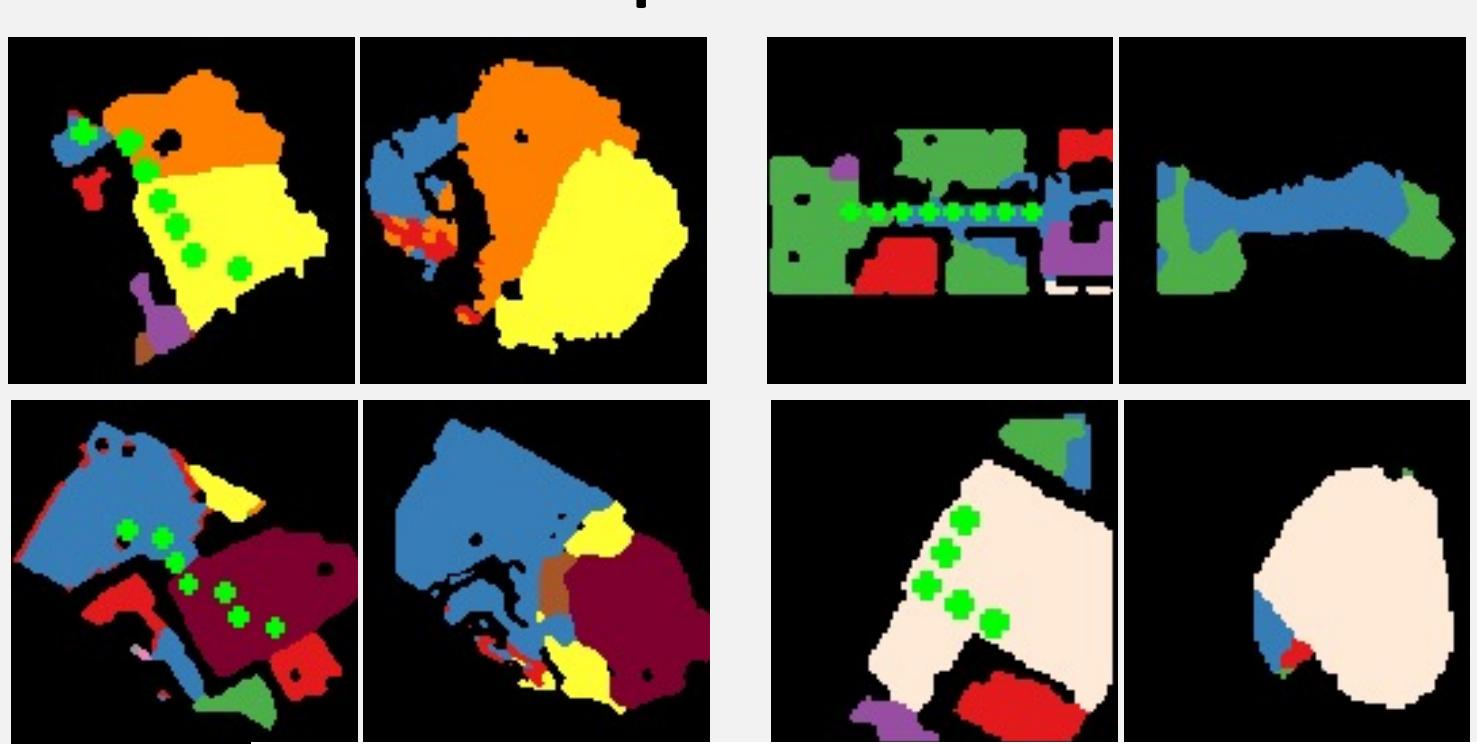
Output: 2D Binary Interior Map, N-d Room Map



Interior Map Predictions



Room Map Predictions



References:

- Chen, Changhan, et al. "Soundspaces: Audio-visual navigation in 3d environments." Computer Vision–ECCV 2020
- Chang, Angel, et al. "Matterport3d: Learning from rgb-d data in indoor environments." arXiv preprint arXiv:1709.06158 (2017).
- Ramakrishnan, Santhosh K., Ziad Al-Halah, and Kristen Grauman. "Occupancy anticipation for efficient exploration and navigation." European Conference on Computer Vision. Springer, Cham, 2020.