Audio-Visual Floorplan Reconstruction

Senthil Purushwalkam[1]

Sebastià V. Amengual Garí[3]

Vamsi Krishna Ithapu[3]

Carl Schissler[3]

Abhinav Gupta[12]

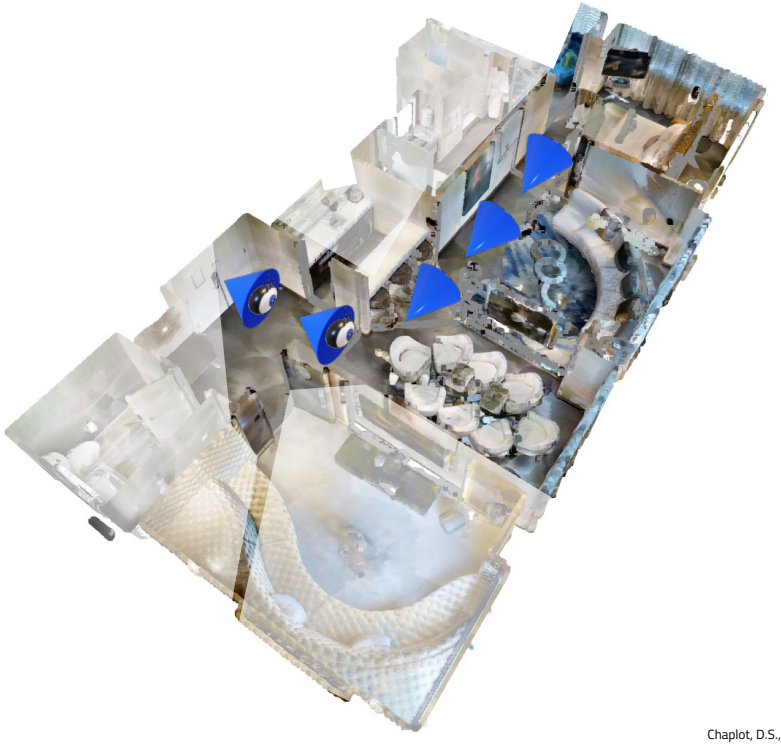
Kristen Grauman[24]

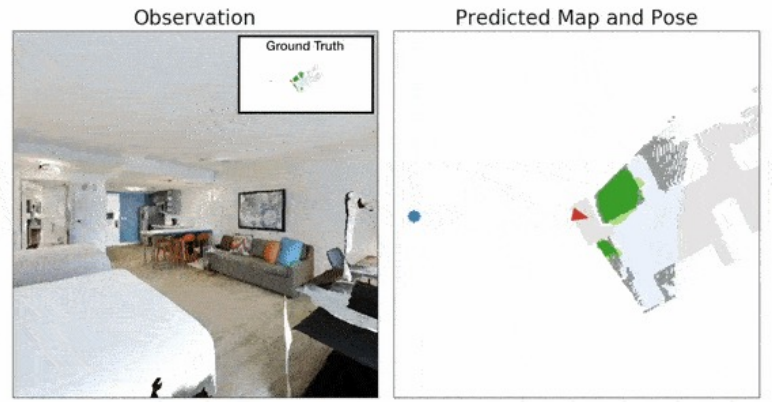
[1]Carnegie Mellon University [2]Facebook AI Research [3]Facebook Reality Labs [4]University of Texas at Austin

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

Walk through an Active Household

Walk through an Active Household

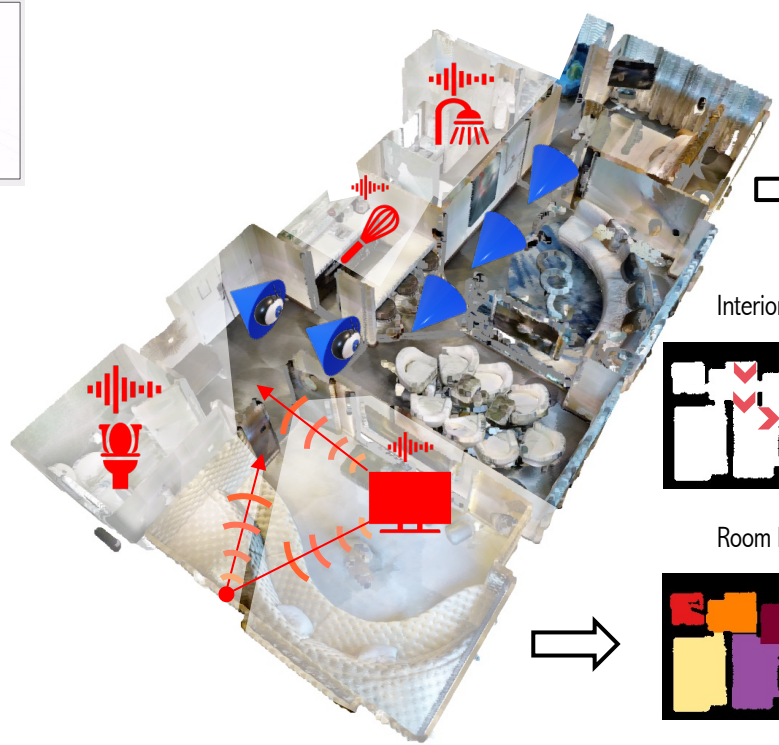




Chaplot, D.S., Gandhi, D., Gupta, S., Gupta, A. and Salakhutdinov, R., 2020. Learning To Explore Using Active Neural SLAM. In International Conference on Learr Representations (ICLR).

Walk through an Active Household

Geometric Structure

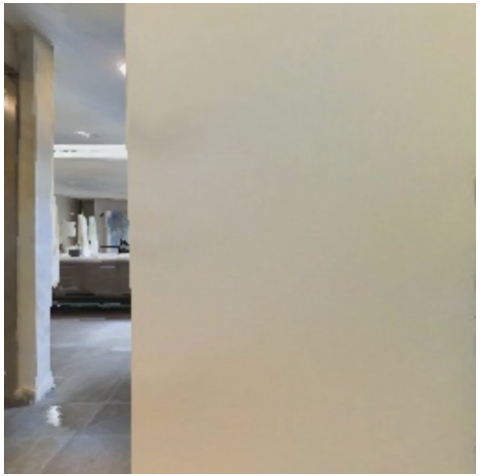


Interior Map

Room Map

Problem Formulation

Video with Audio



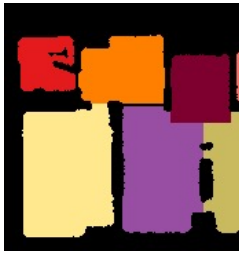
Sequence of RGB Frames

Sequence of Audio Clips

Binary Map

Free space,
Small Objects

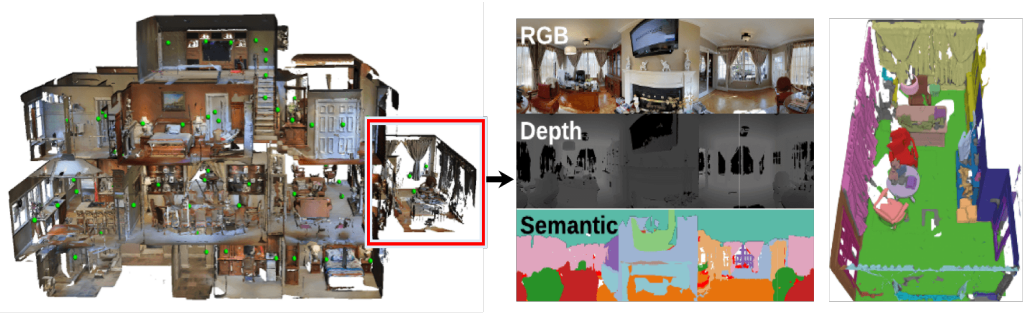
N-channel Map



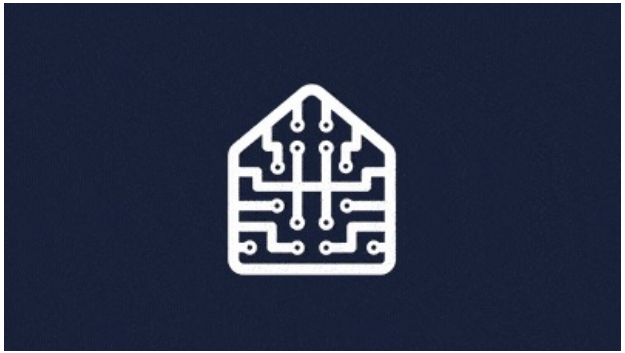
Room Labels

Data Generation - Visual

Matterport3D Dataset: 85 Home Environments



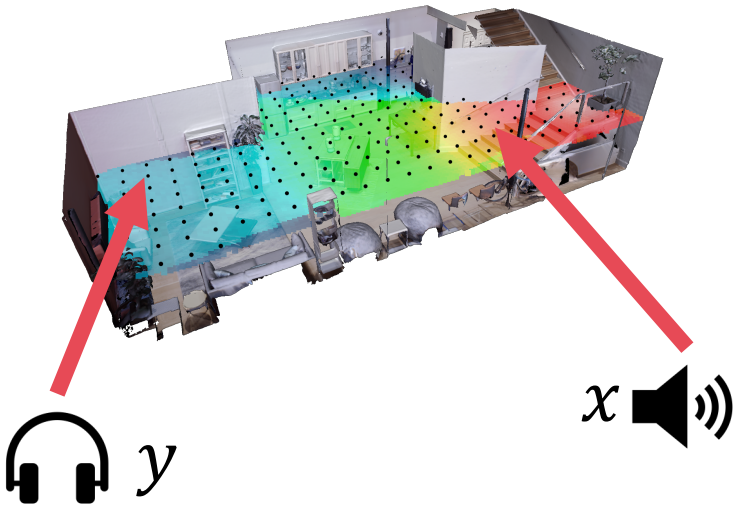
Habitat-Sim to generate RGB-sequences



Chang, Angel, et al. "Matterport3d: Learning from rgb-d data in indoor environments." arXiv preprint arXiv:1709.06158 (2017).
Savva, Manolis, et al. "Habitat: A platform for embodied ai research." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

Data Generation - Audio

SoundSpaces Dataset



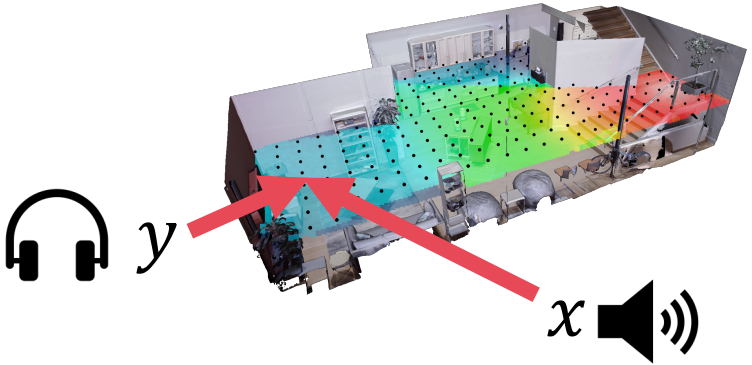
- Impulse Responses (IR) for a dense grid of source-receiver positions
- Ambisonics and Binaural IRs

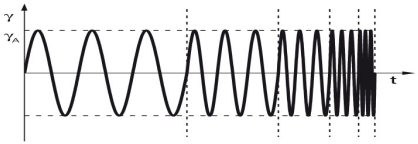
$$y = x \circledast IR$$

Chen, Changan, et al. "Soundspaces: Audio-visual navigation in 3d environments." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2020.

Data Generation - Audio

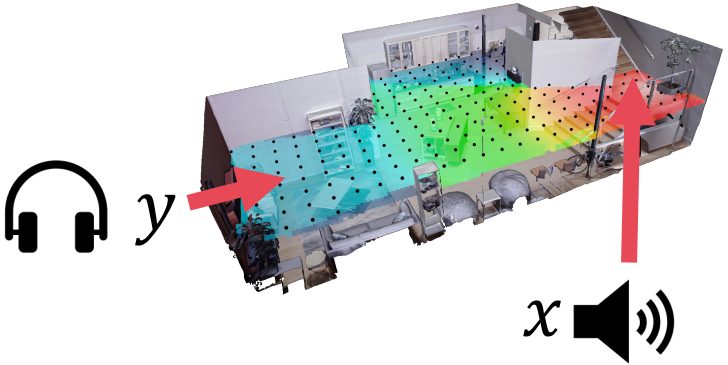
Device Generated Audio





Freq Sweep

Environment Generated Audio



56/32/32 Train/Val/Test Sounds

Telephone ring, flush, keyboard, …

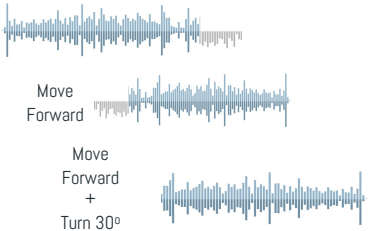
AV Map Model

Move Forward

Move Forward + Turn 30° Left

Features + Top-Down Alignment

Self-Attn + UNet Enc.



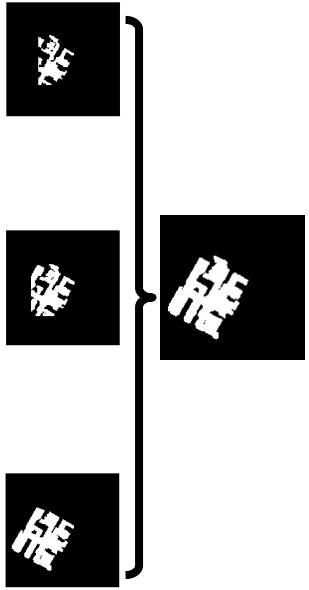
Move Forward

Move Forward + Turn 30° Left

Features + Top-Down Alignment

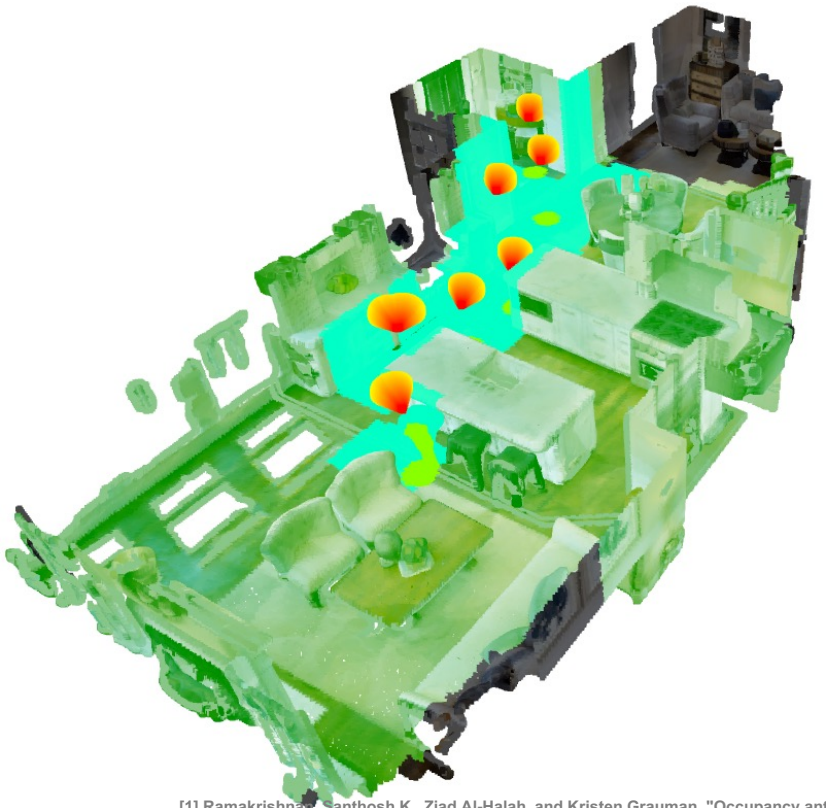
Self-Attn + UNet Enc.

Self-Attn + UNet Decoder

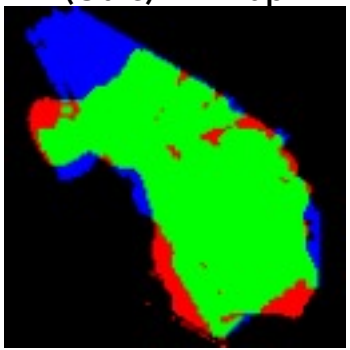


Pool Predictions

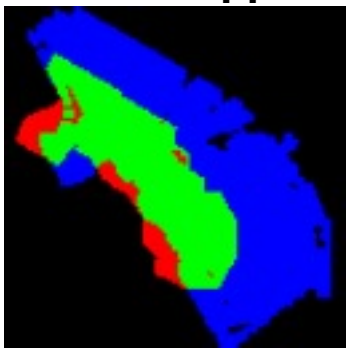
Interior Map Predictions

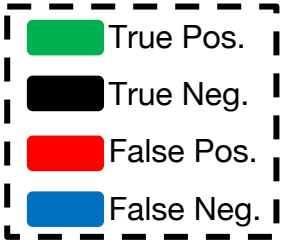


(Ours) AV-Map



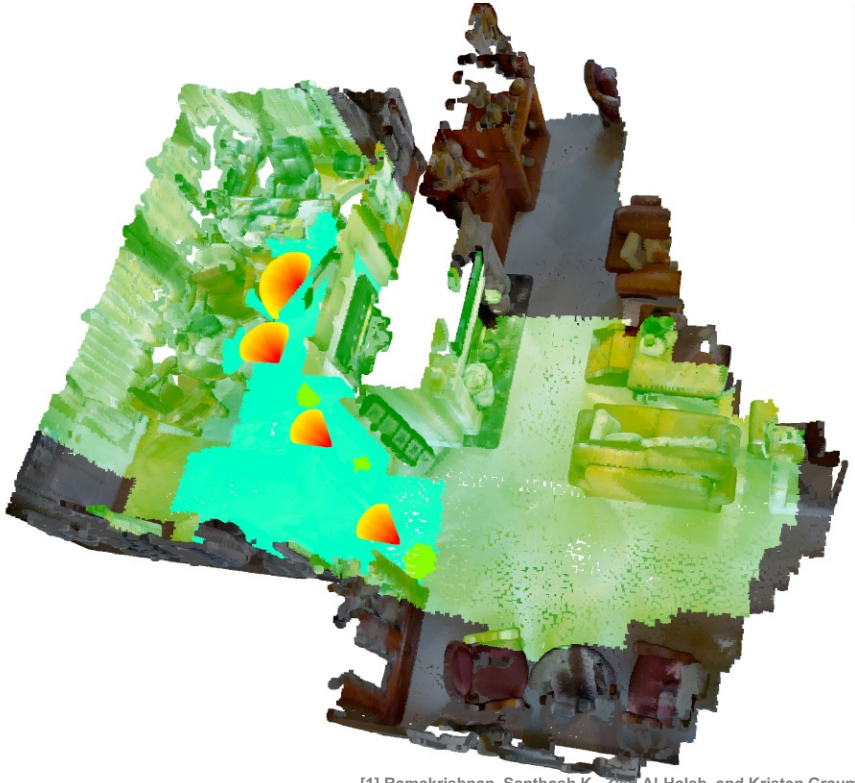
Occ Ant [1]



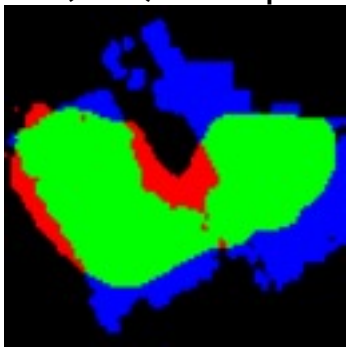


[1] Ramakrishnan, Santhosh K., Ziad Al-Halah, and Kristen Grauman. "Occupancy anticipation for efficient exploration and navigation." European Conference on Computer Vision. Springer, Cham, 2020.

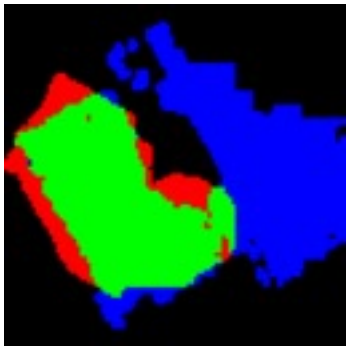
Interior Map Predictions

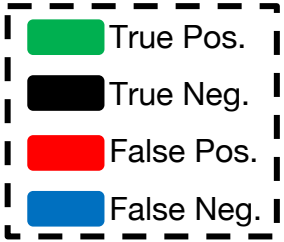


(Ours) AV-Map



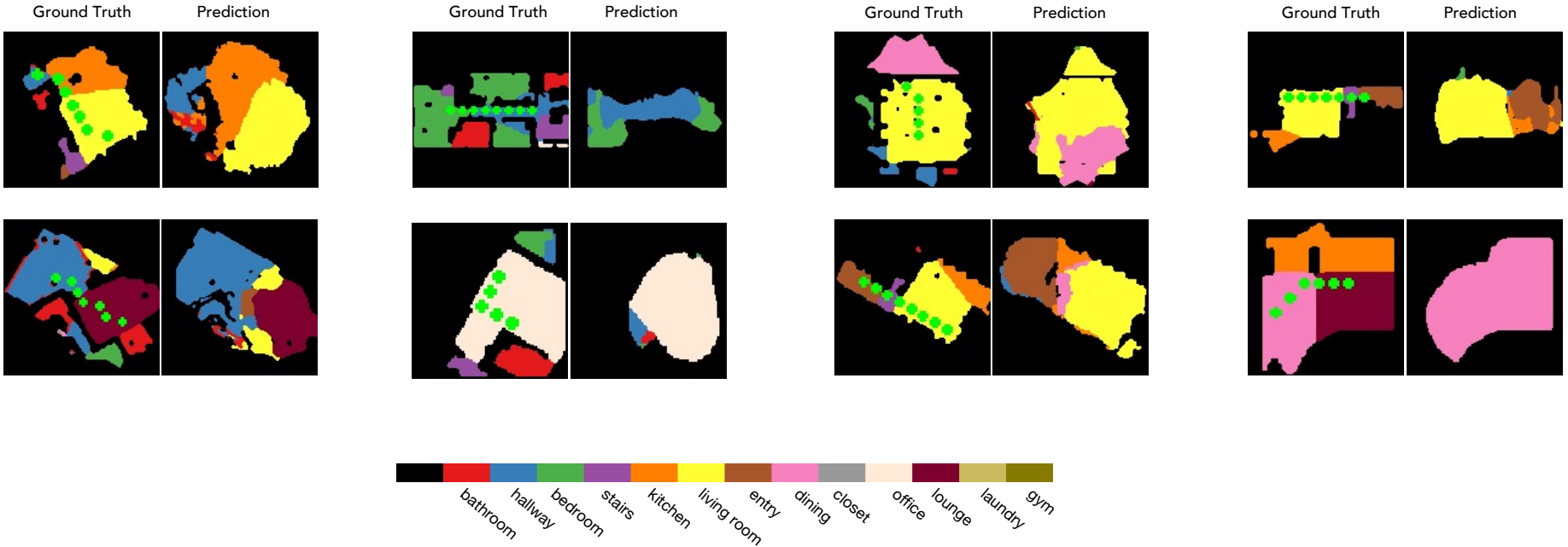
Occ Ant [1]



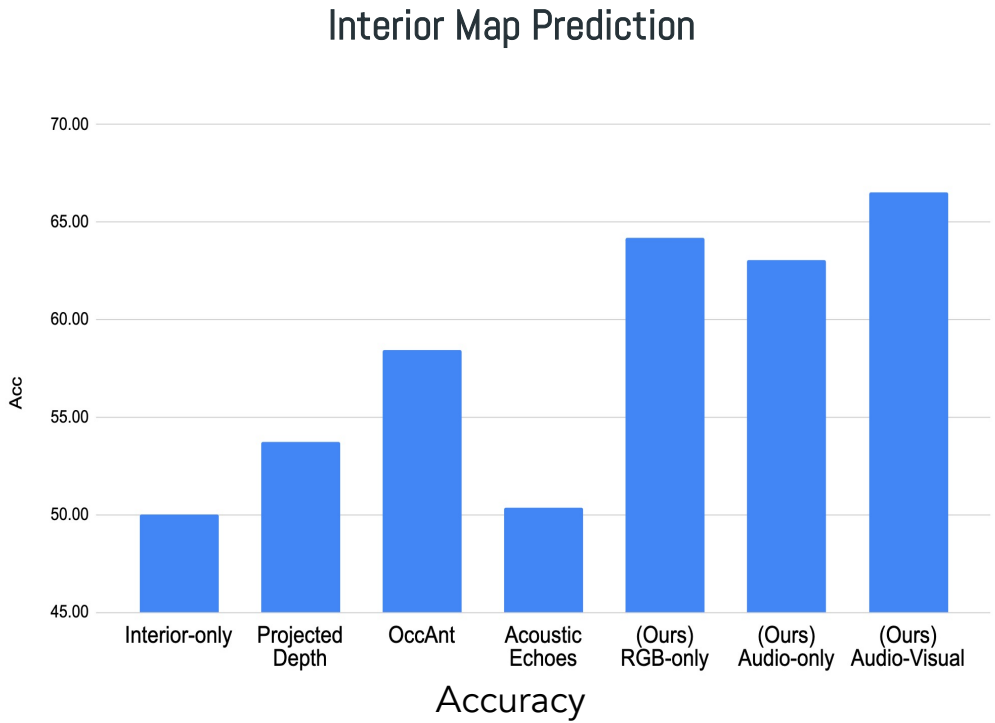


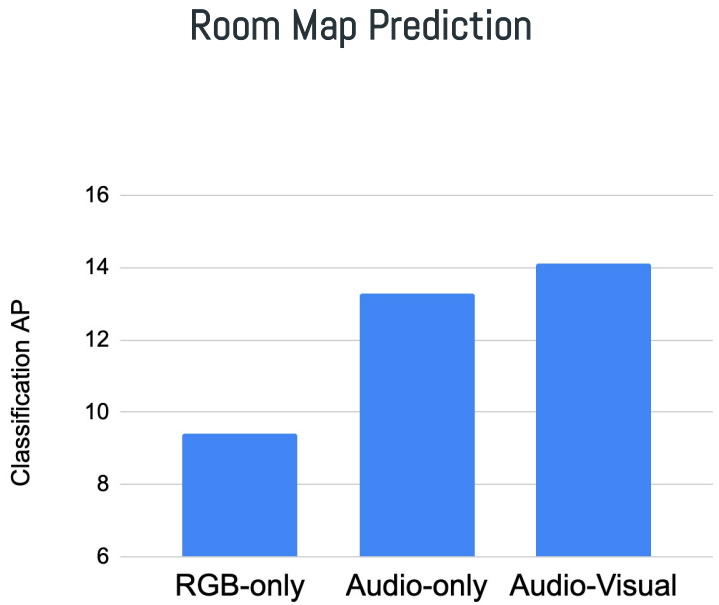
[1] Ramakrishnan, Santhosh K., Ziad Al-Halah, and Kristen Grauman. "Occupancy anticipation for efficient exploration and navigation." European Conference on Computer Vision. Springer, Cham, 2020.

Room Map Predictions



Quantitative Results (See paper for details)





Thank you for Listening!

Checkout the website for video demos with audio:
http://www.cs.cmu.edu/~spurushw/publication/avmap/