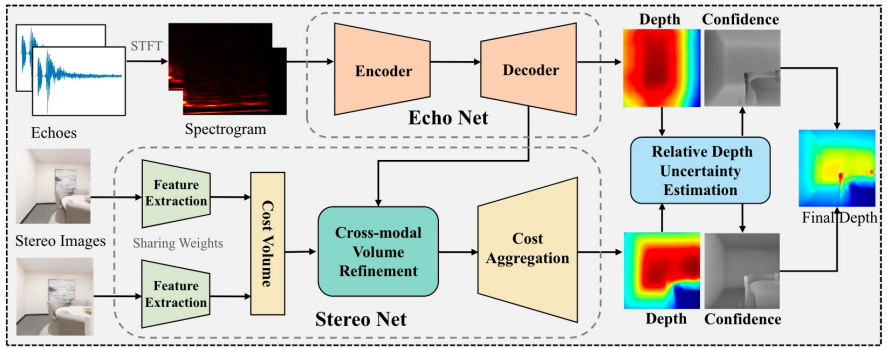
The Audio-Visual BatVision Dataset for Research on Sight and Sound

Amandine Brunetto[1*], Sascha Hornauer[1*], Stella X. Yu[2], Fabien Moutarde[1]

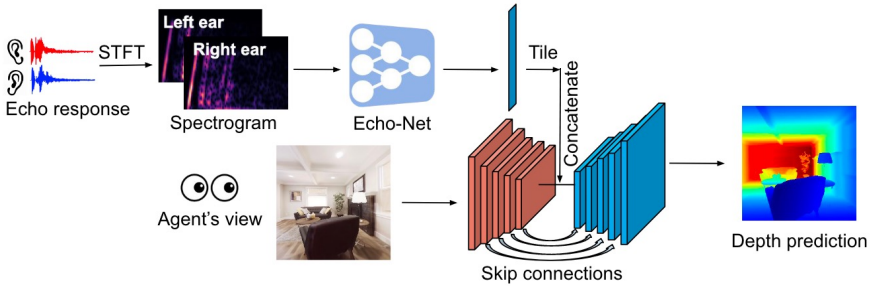
[1]Center for Robotics, MINES Paris, Université PSL, Paris, France
[2]University of Michigan, Ann Arbor, United States of America
*Equal Contribution

Sound & Vision

Improved depth prediction

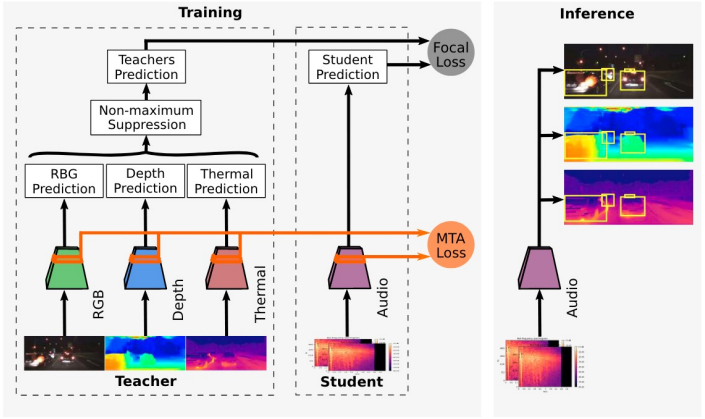


Zhang, Chenghao et al., "Stereo Depth Estimation with Echoes". ECCV 2022.



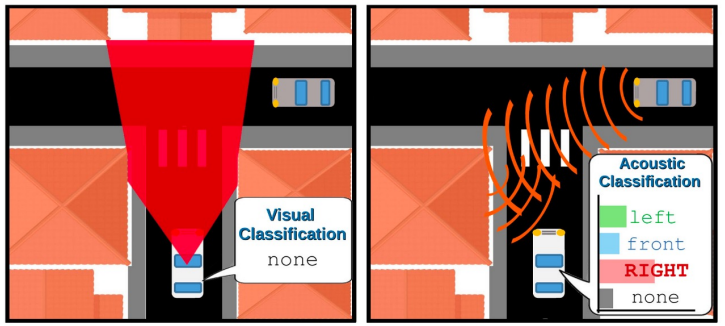
Gao, Ruohan et al., "VisualEchoes: Spatial Image Representation Learning through Echolocation". ECCV 2020.

Object tracking



Valverde, Francisco et al., "There is more than meets the eye: Self-supervised multi-object detection and tracking with sound by distilling multimodal knowledge". CVPR 2021.

None-line-of-sight detection



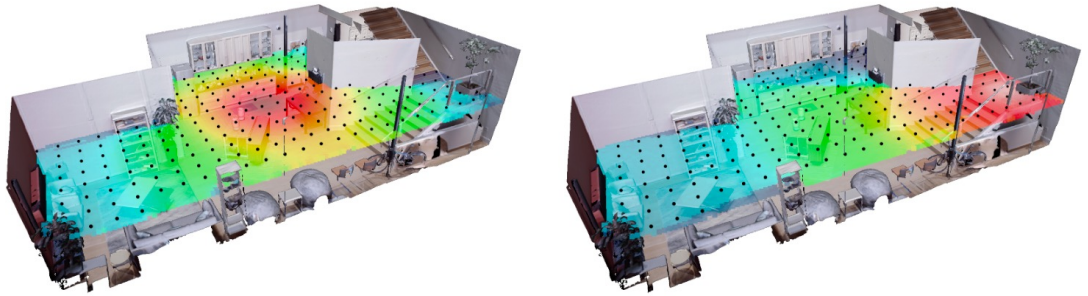
(a) line-of-sight sensing (b) directional acoustic sensing

(c) sound localization with a vehicle-mounted microphone array detects the wall reflection of an approaching vehicle behind a corner before it appears

Schulz, Yannick et al., "Hearing what you cannot see: Acoustic Vehicle Detection Around Corners". *IEEE Robotics and Automation Letters* 6.2 (2021).

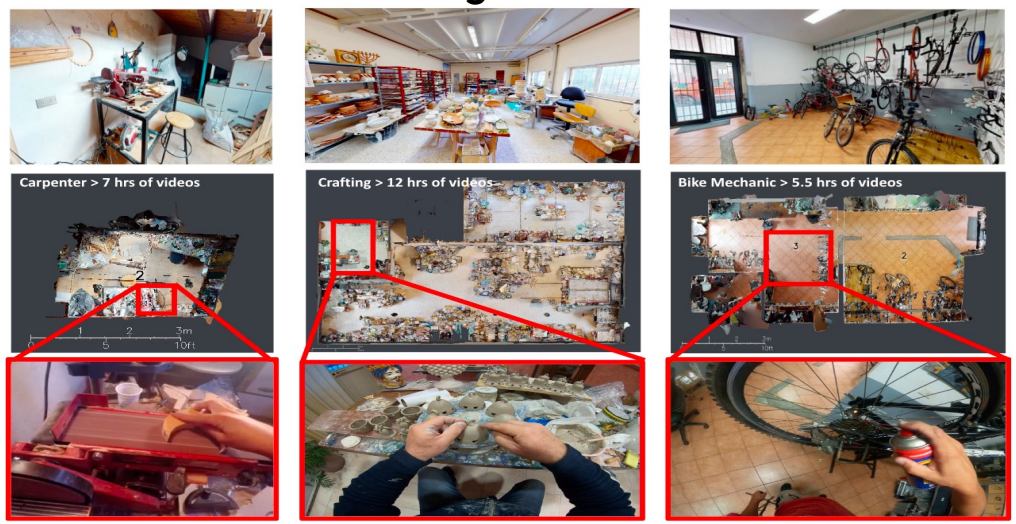
Audio-Visual Datasets

A Widely Used Simulation Dataset: SoundSpaces[1]



Real-World Dataset with Audio

Ego4D[2]



The Greatest Hits[3]

[1] Chen, Changan, et al. "Soundspaces: Audio-visual navigation in 3d environments." ECCV 2020
[2] Grauman, Kristen, et al. "Ego4d: Around the world in 3,000 hours of egocentric video." CVPR 2022
[3] Owens, Andrew, et al. "Visually indicated sounds." CVPR 2016

Robot Echolocation

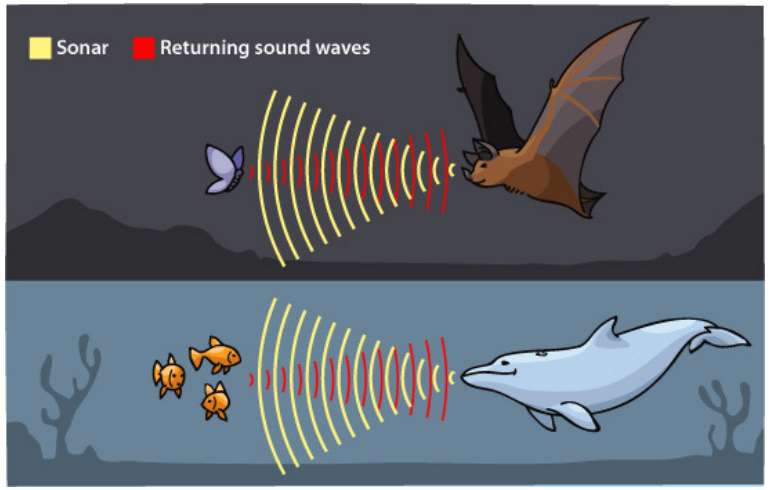
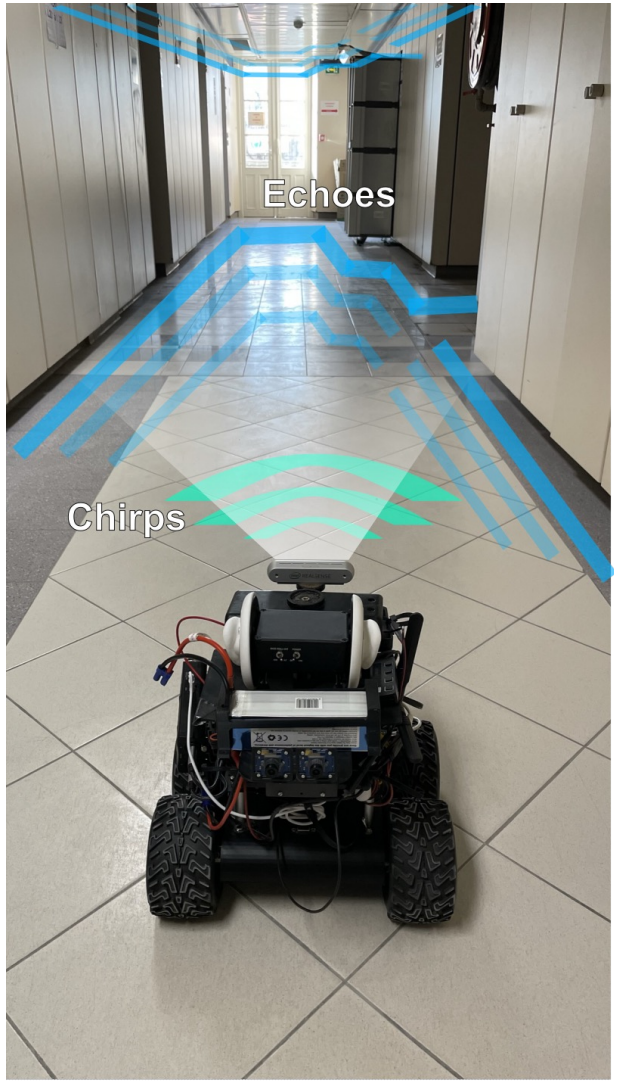
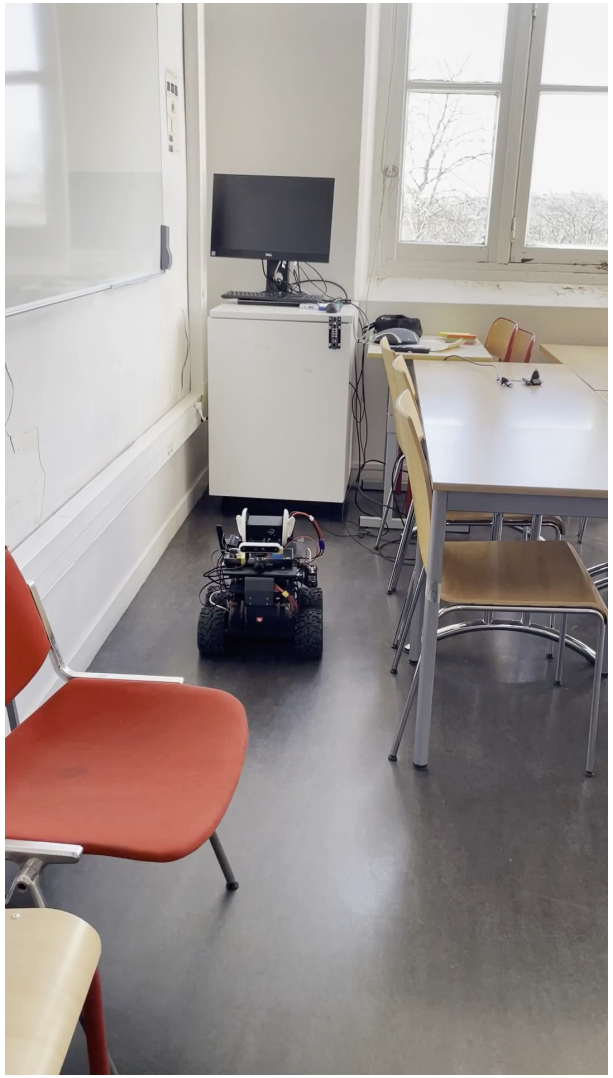
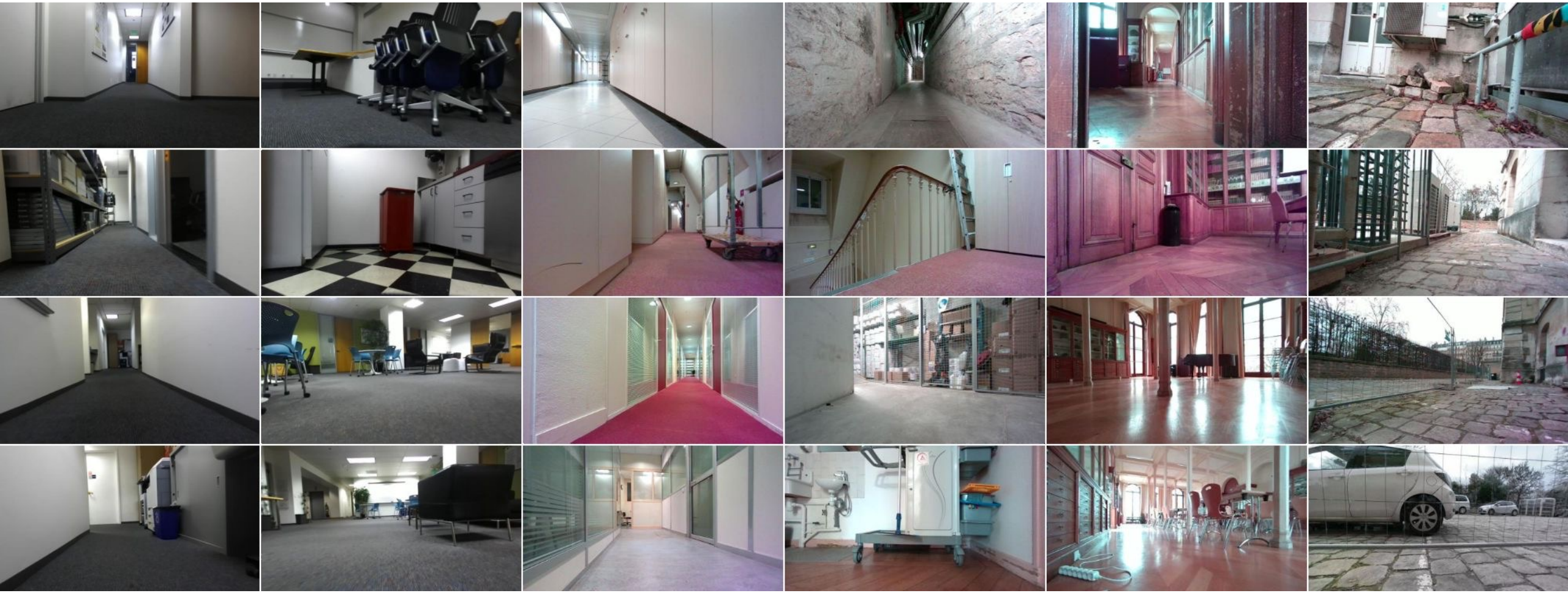


Image from https://askabiologist.asu.edu/echolocation





Dataset Overview



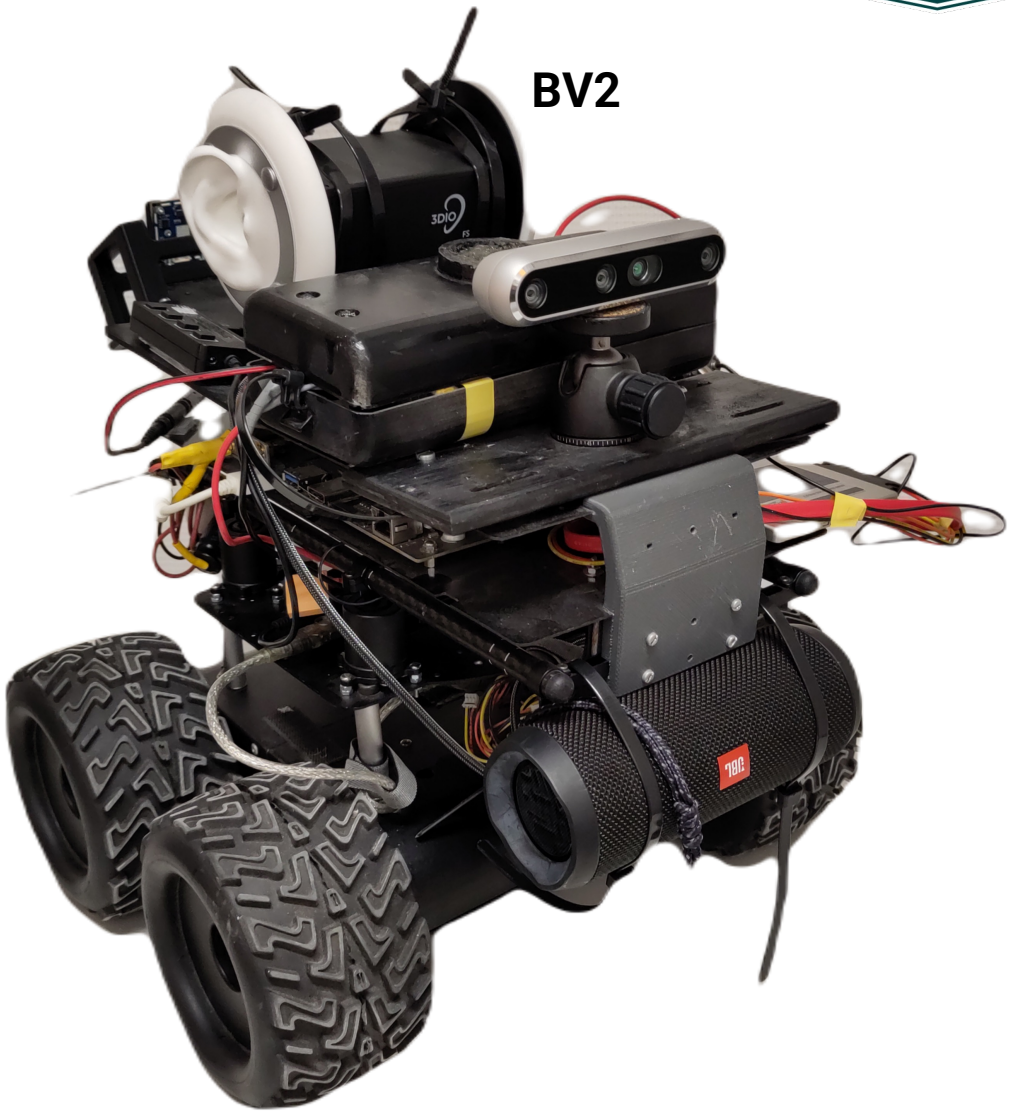
UC Berkeley (BV1): 52,220 instances

Mines Paris (BV2): 3,120 instances

Robots

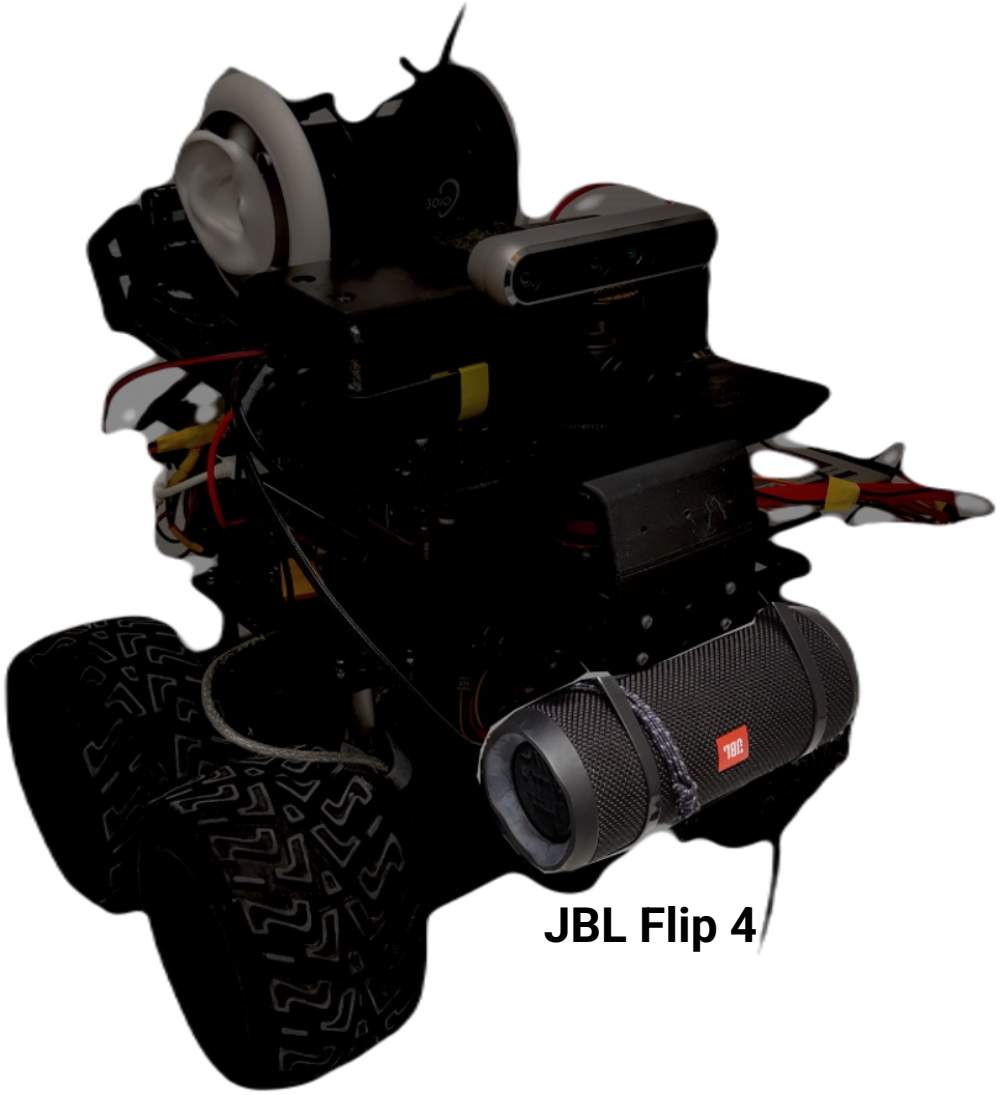
BV1

BV2



Speakers

JBL Flip 4

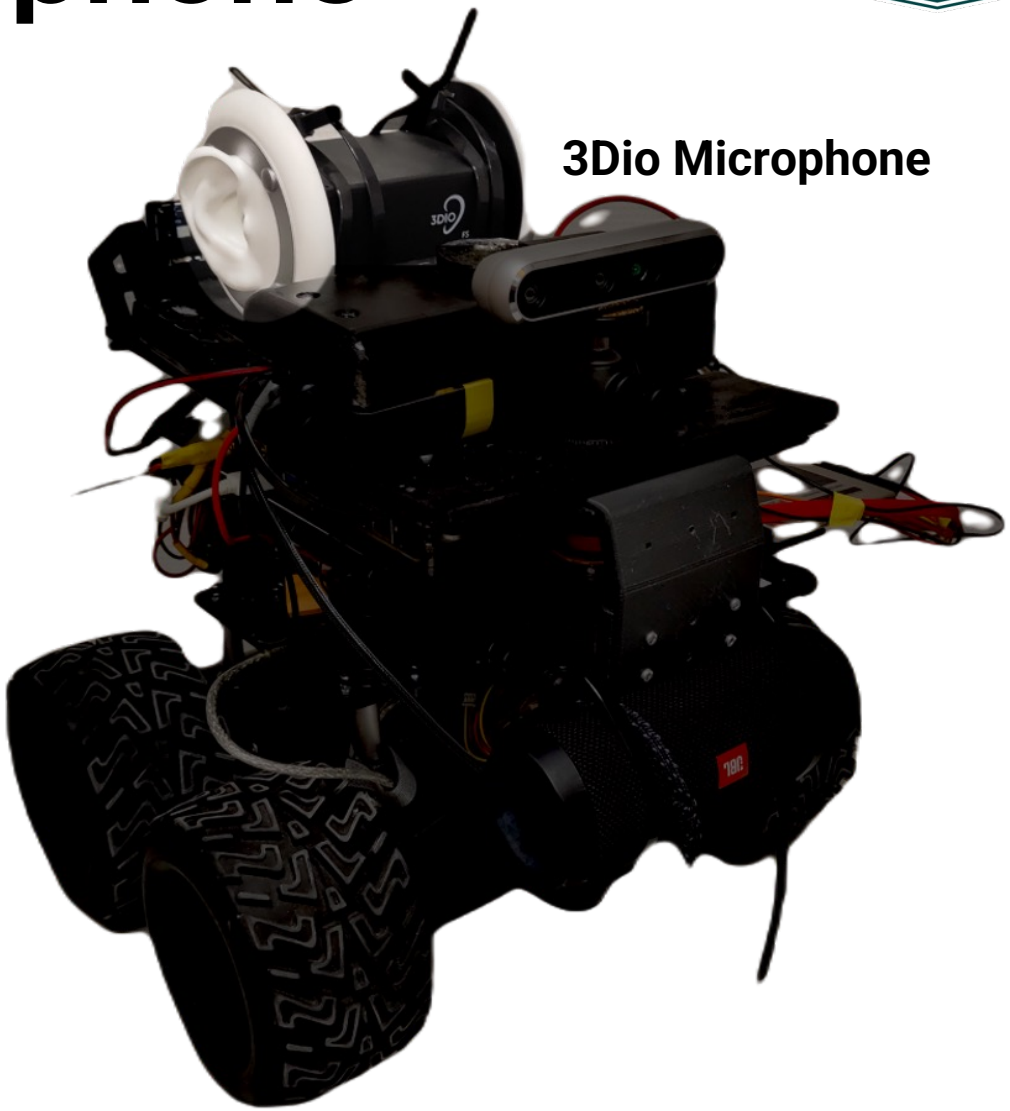


JBL Flip 4

Binaural Microphone

Microphones in silicon ears

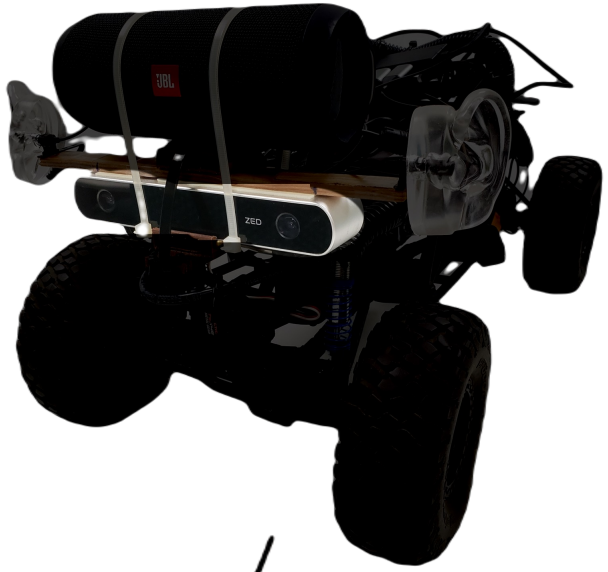
3Dio Microphone



Vision

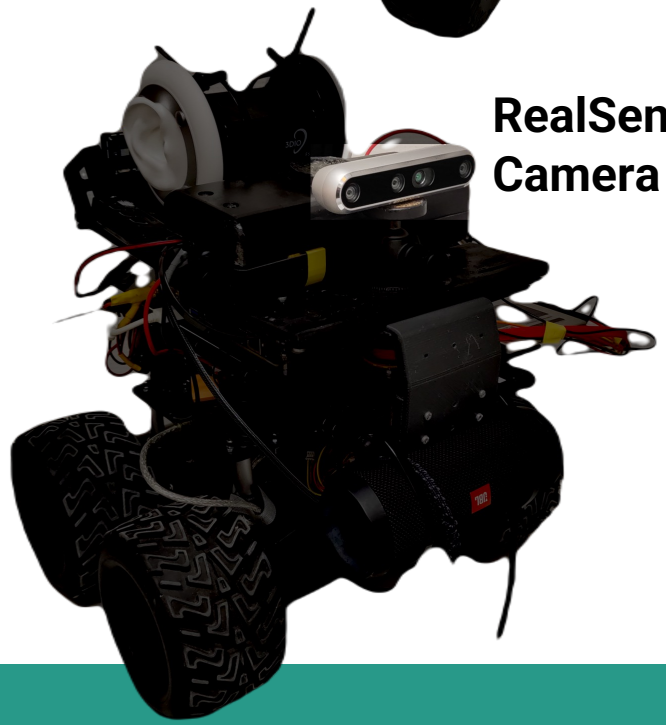
ZED Stereo Camera

BV1

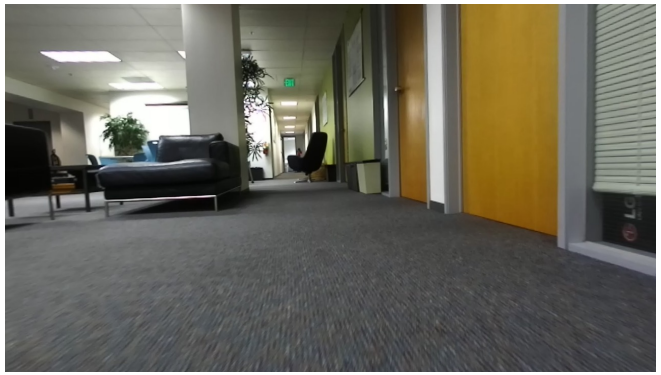


RealSense RGB-D Camera

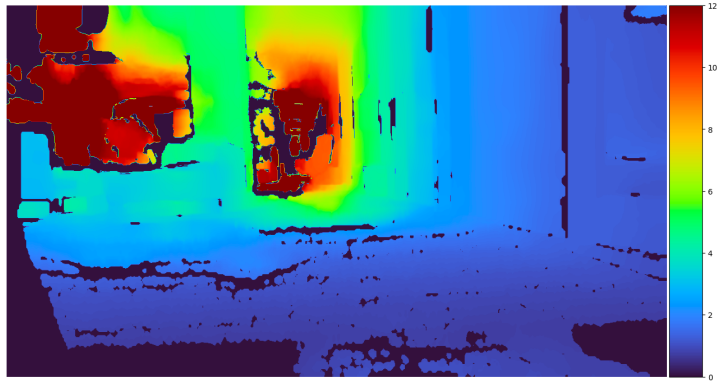
BV2



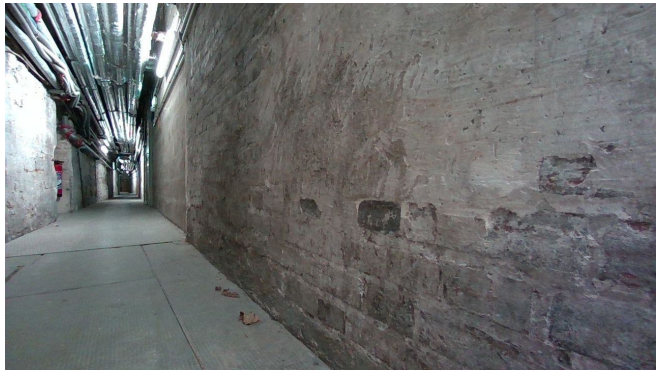
RGB



Depth

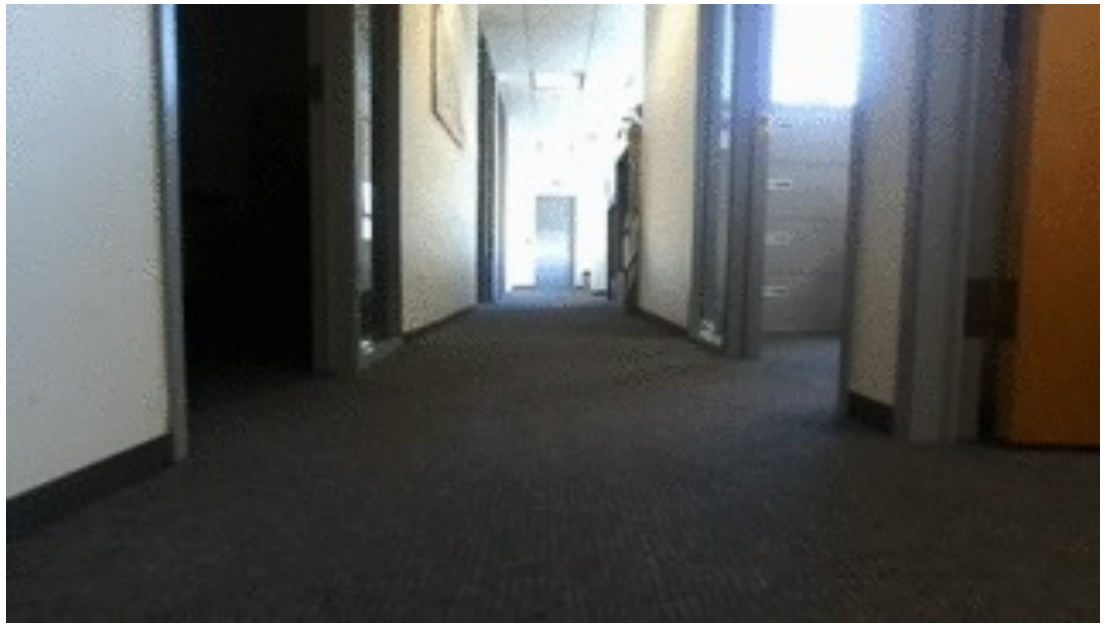


RGB



Depth

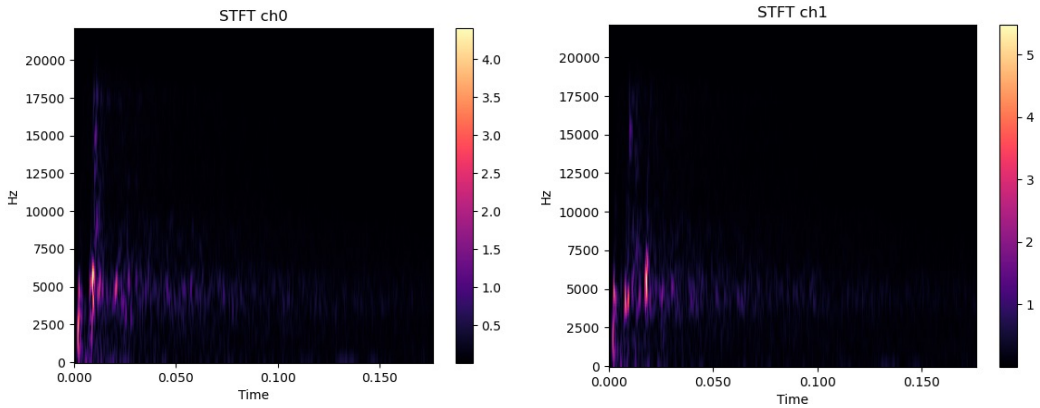
BV1

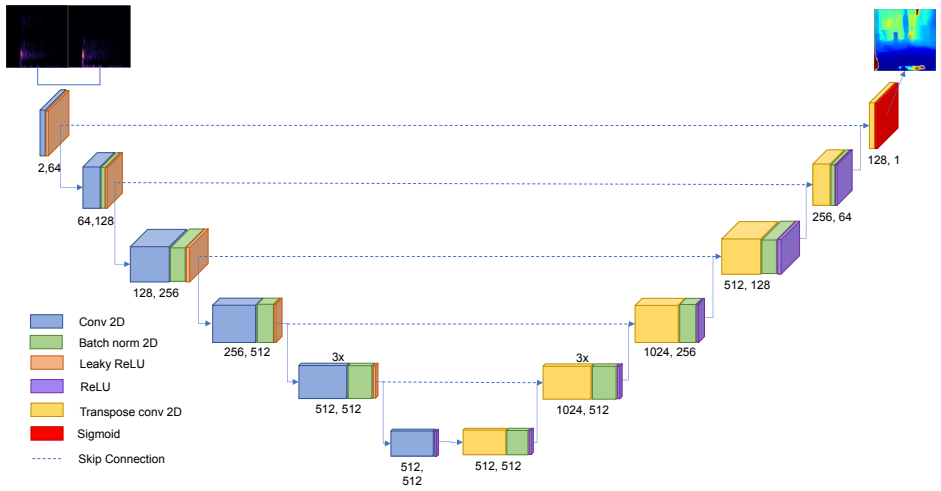


BV2

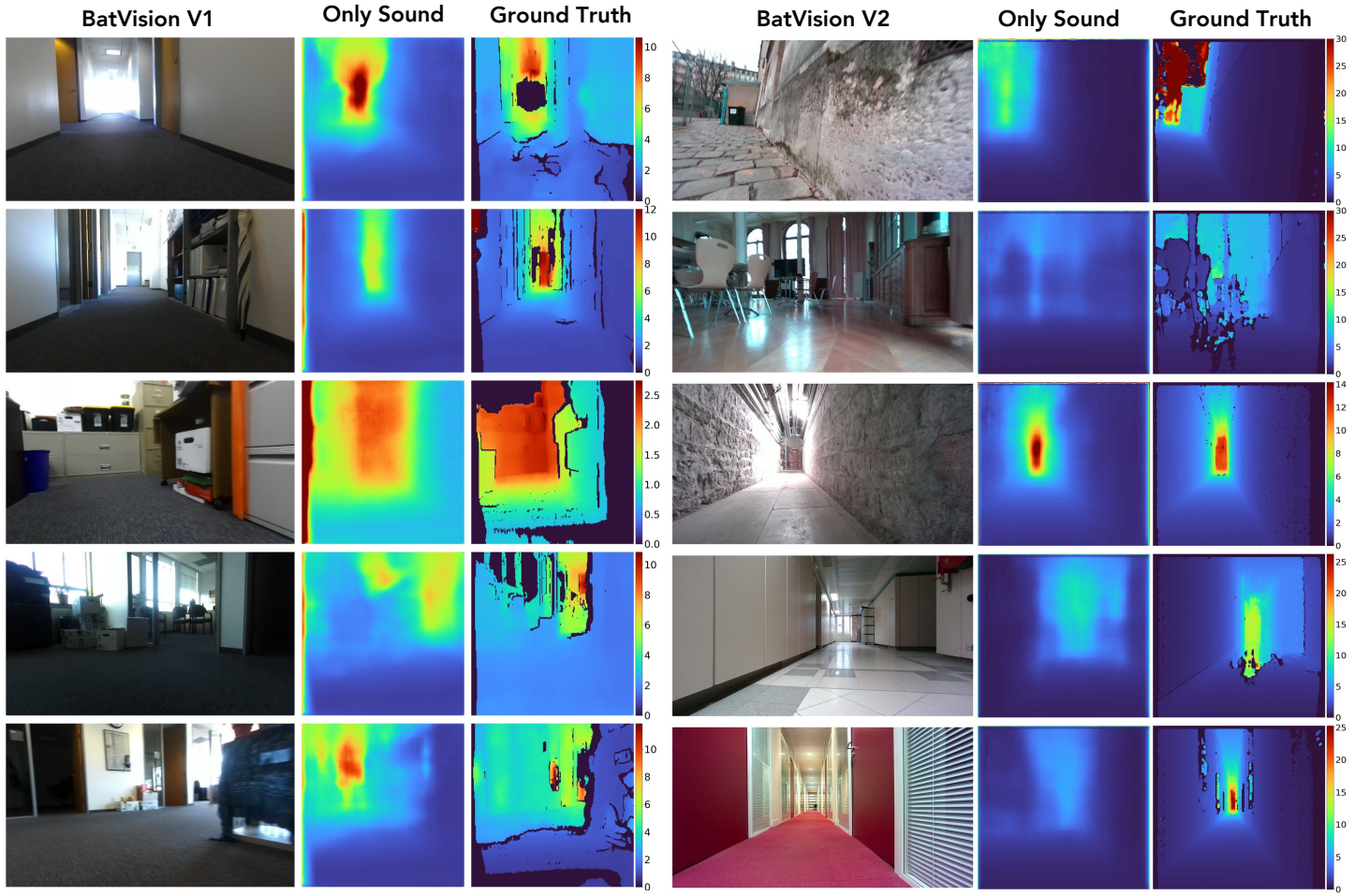
Depth Prediction from Audio-Only: UNet Baseline

Audio Spectrograms





UNet Architecture

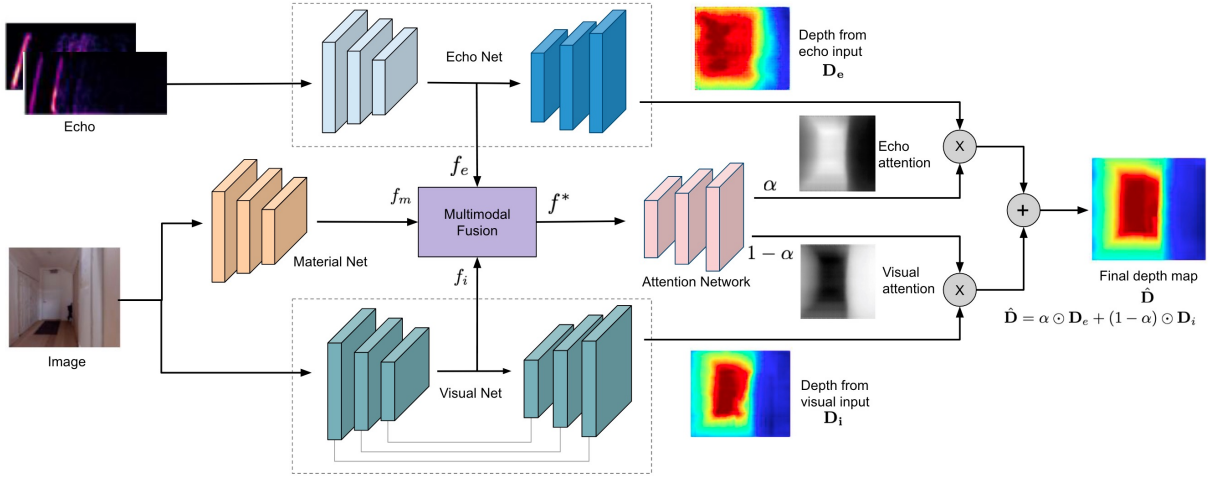


Results

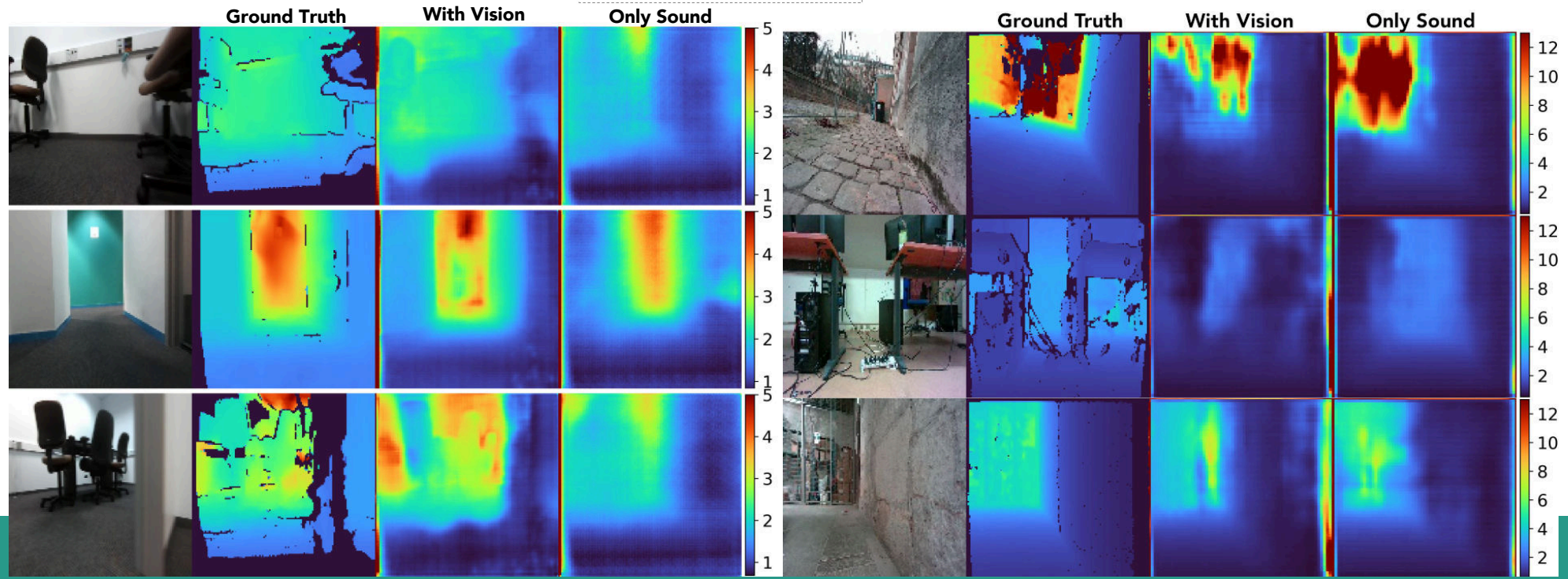
Depth Prediction State-of-the-Art[1]

[1] Parida, et al. "Beyond Image to Depth: Improving Depth Prediction using Echoes."

Original paper: Simulated Data



Real Data



Dataset is available at the project page: https://amandinebtto.github.io/Batvision-Dataset/
Code: https://github.com/AmandineBtto/Batvision-Dataset

Contact: amandine.brunetto@minesparis.psl.eu
Sascha.hornauer@minesparis.psl.eu