

BatVision: Learning to See 3D Spatial Layout with Two Ears

Jesper Haahr Christensen

Technical University of Denmark

Sascha Hornauer

Stella Yu

University of California, Berkeley / ICSI



Existing Sensors have some Drawbacks

- Vision is valuable sensor but sometimes fails
- Ultrasound, Radar or LIDAR sensors are often costly, complex and provide limited information



Image: Dark shadows by [Christopher Kyba](#) is licensed under a [Creative Commons Attribution 4.0 International License](#).
<https://creativecommons.org/licenses/by/4.0/legalcode>

Approach Inspired by Nature

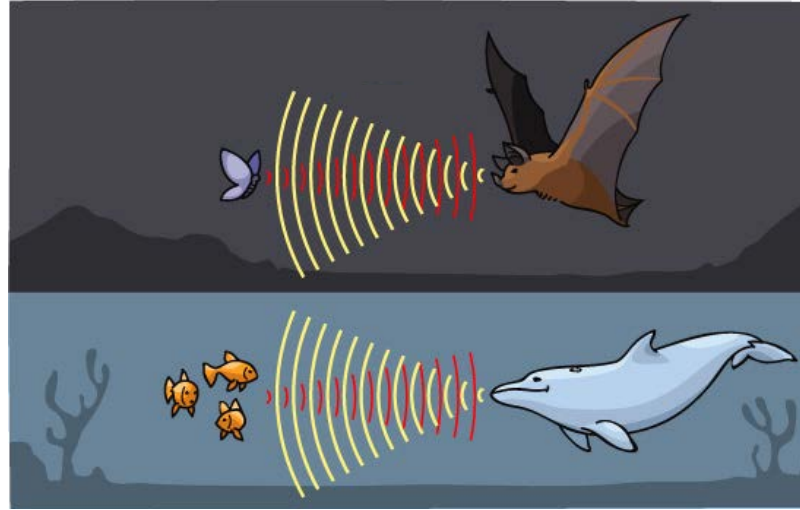
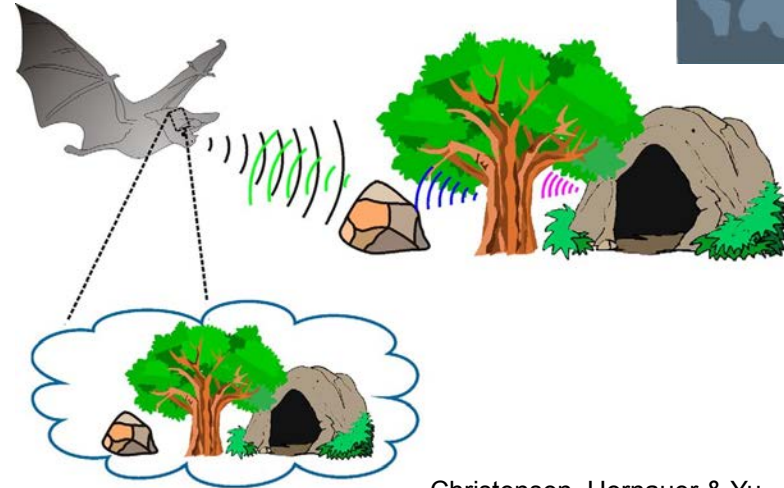


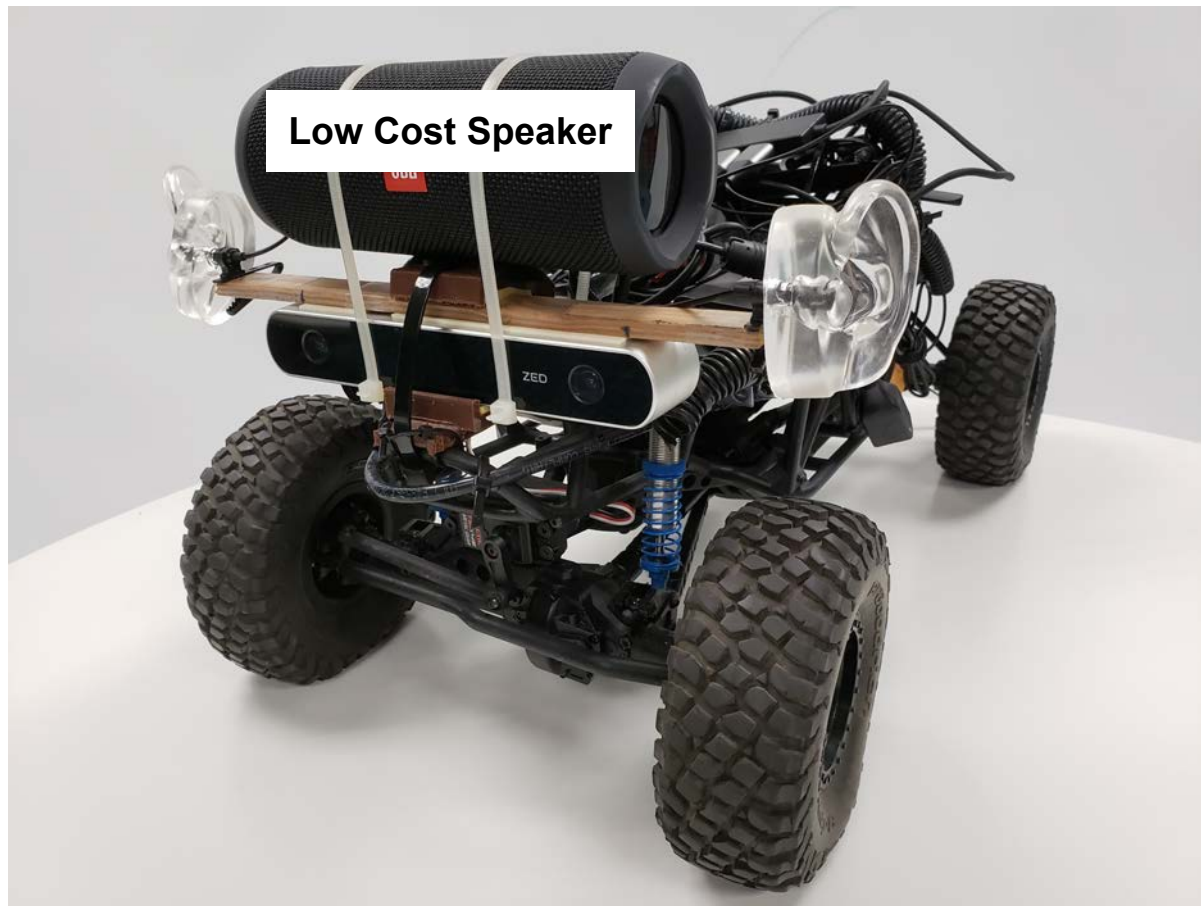
Image credit: <https://askabiologist.asu.edu/echolocation>



Batvision System



Batvision System



Batvision System

Microphones in Artificial Ears

Microphones in Artificial Ears

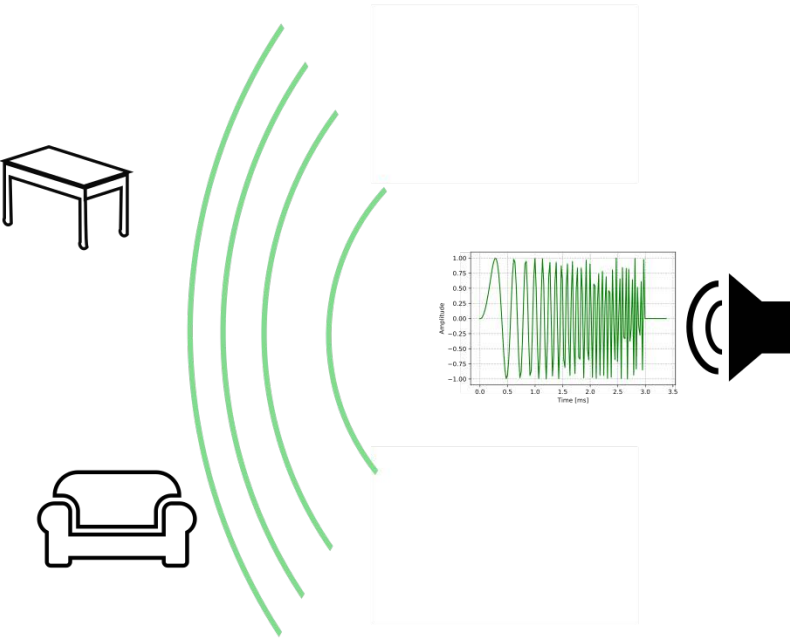


Batvision System



Stereo Camera

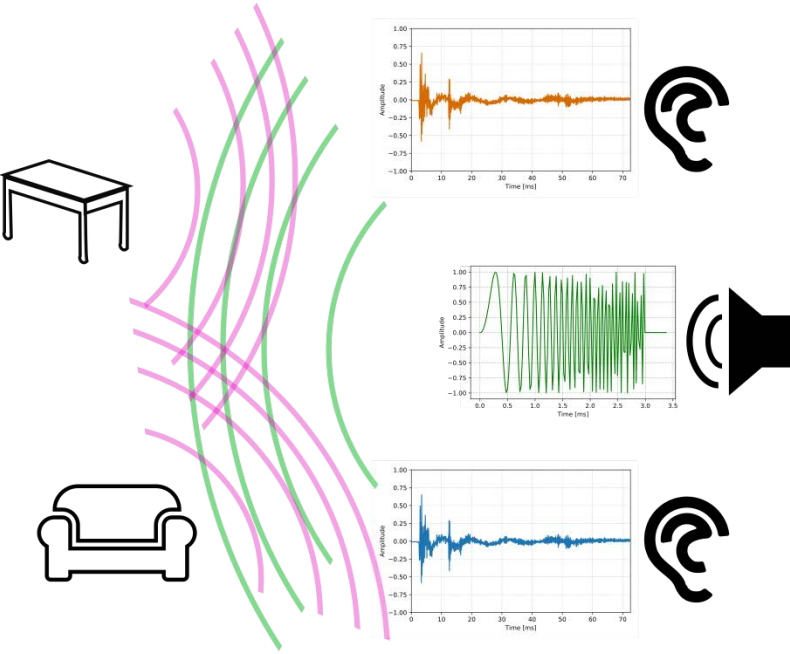
Batvision System



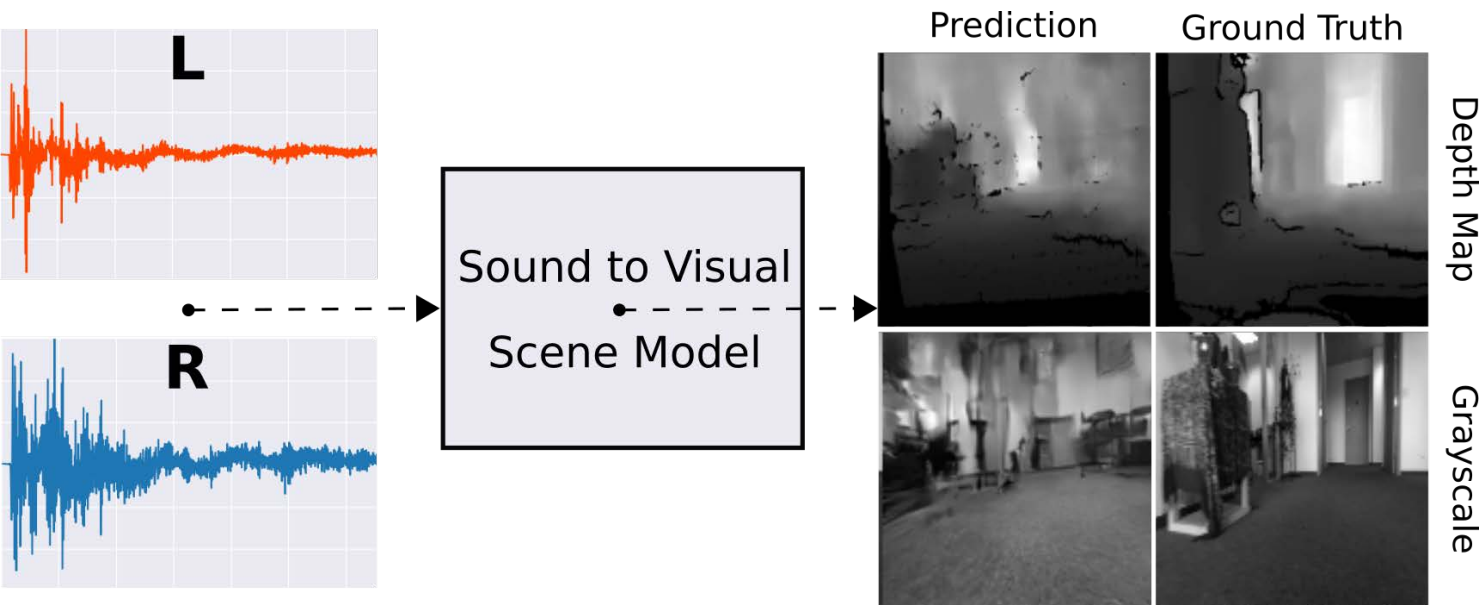
Chirp:

- 3 Milliseconds
- From 20hz to 20kHz

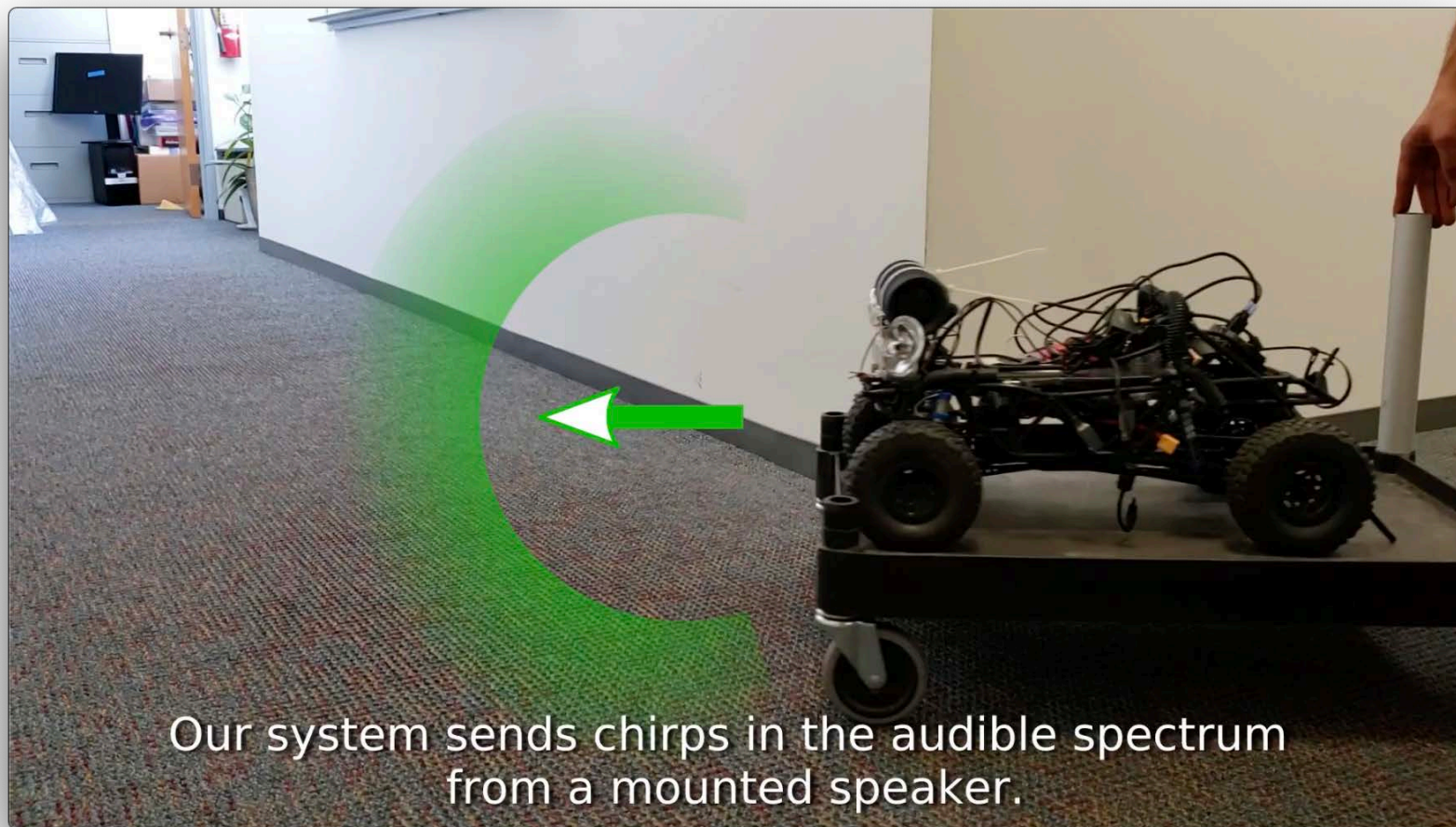
Batvision System



Neural Network Prediction of Visual Layout from Sound

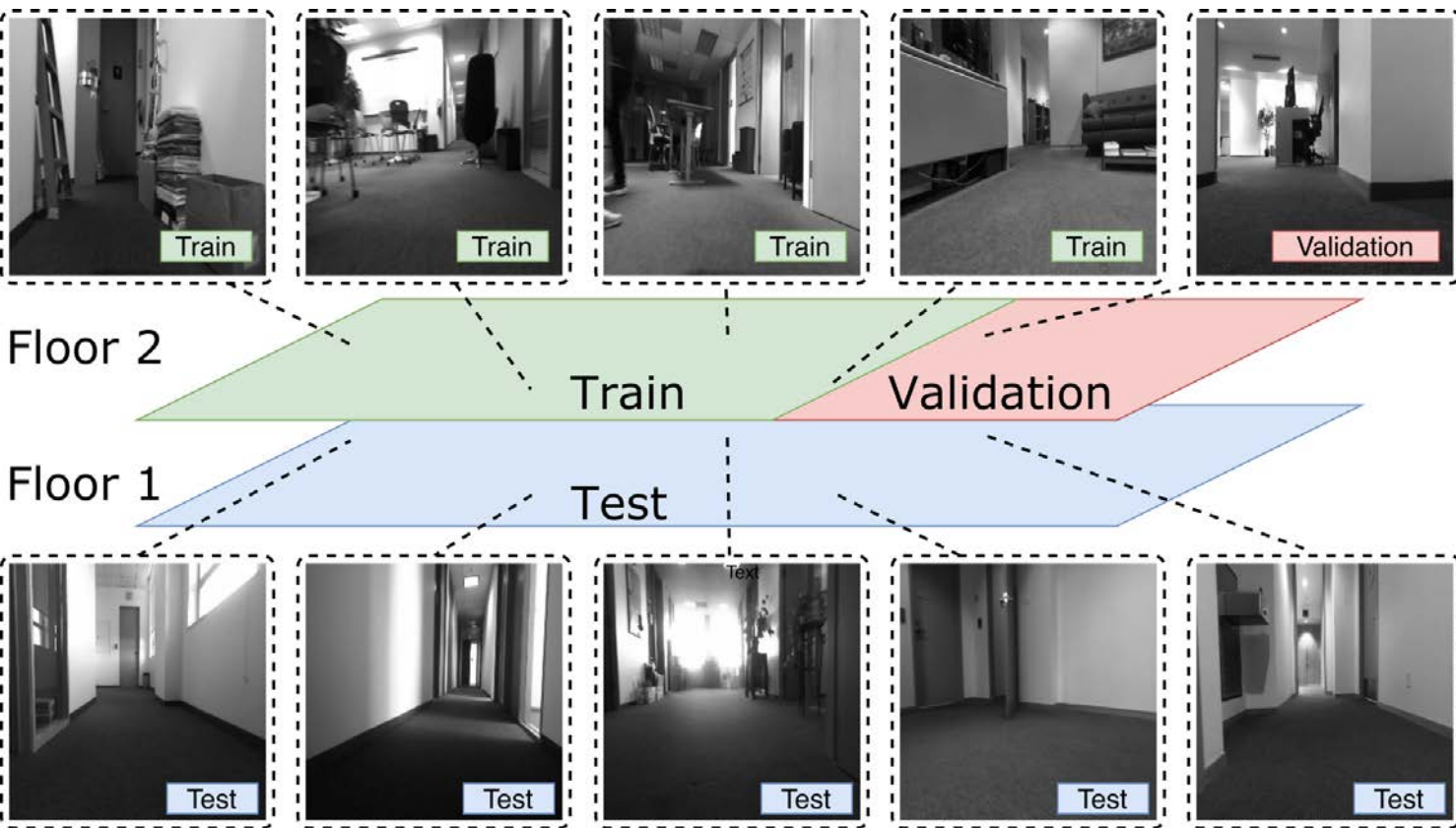


Dataset: Binaural echo to depth



Our system sends chirps in the audible spectrum from a mounted speaker.

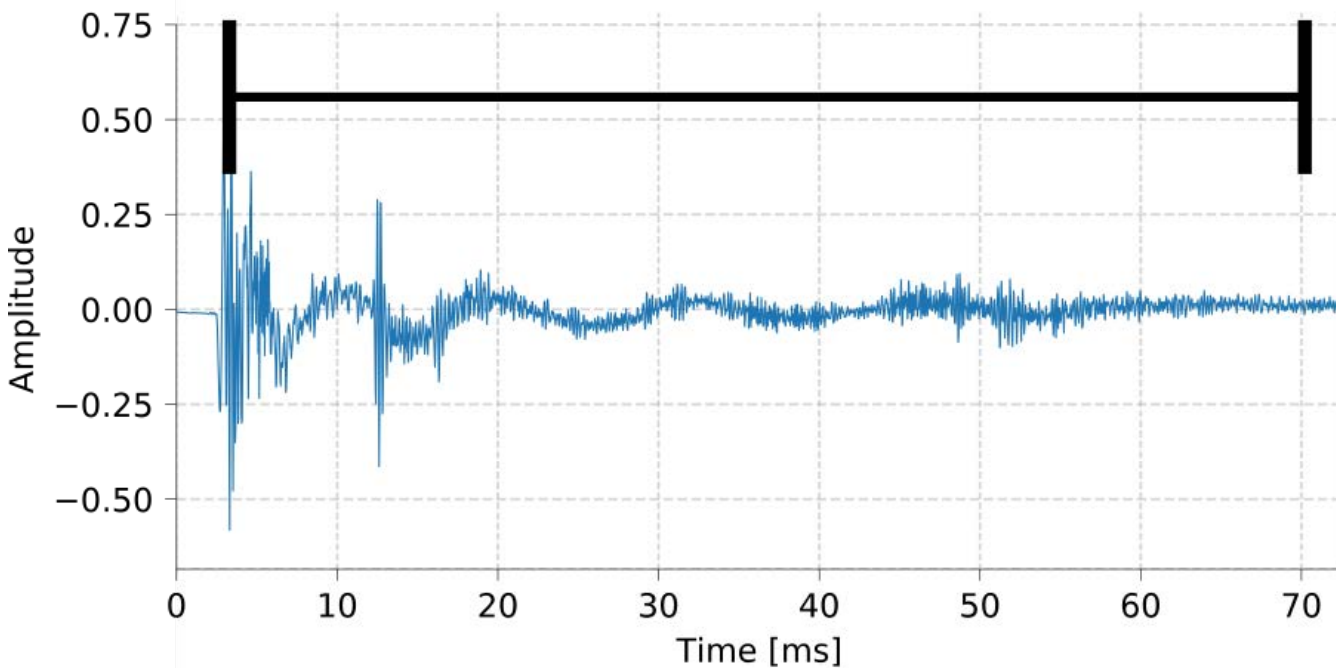
Data Collection in Office Building



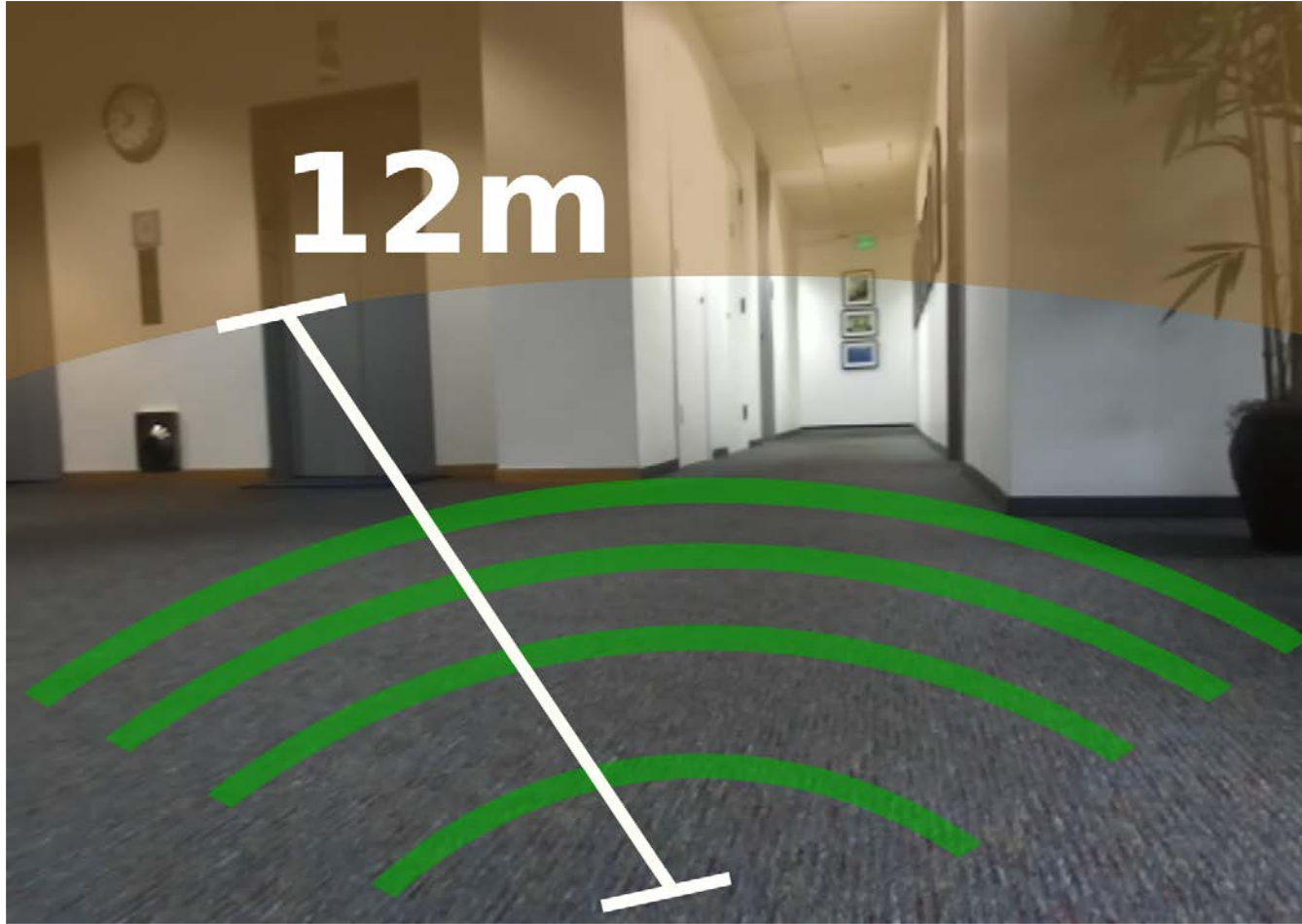
Echos cut off at fixed distance



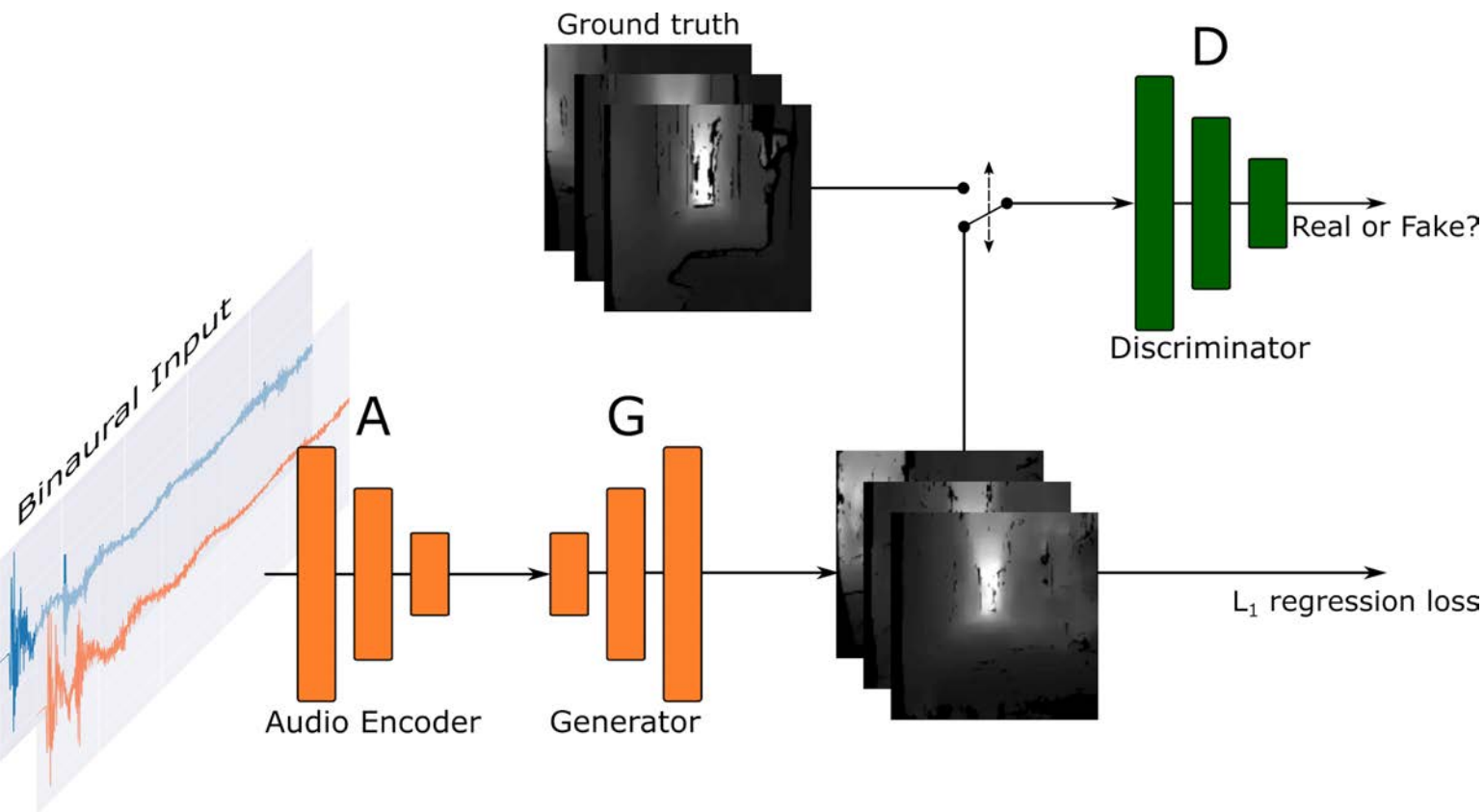
72.5 ms = 12 m



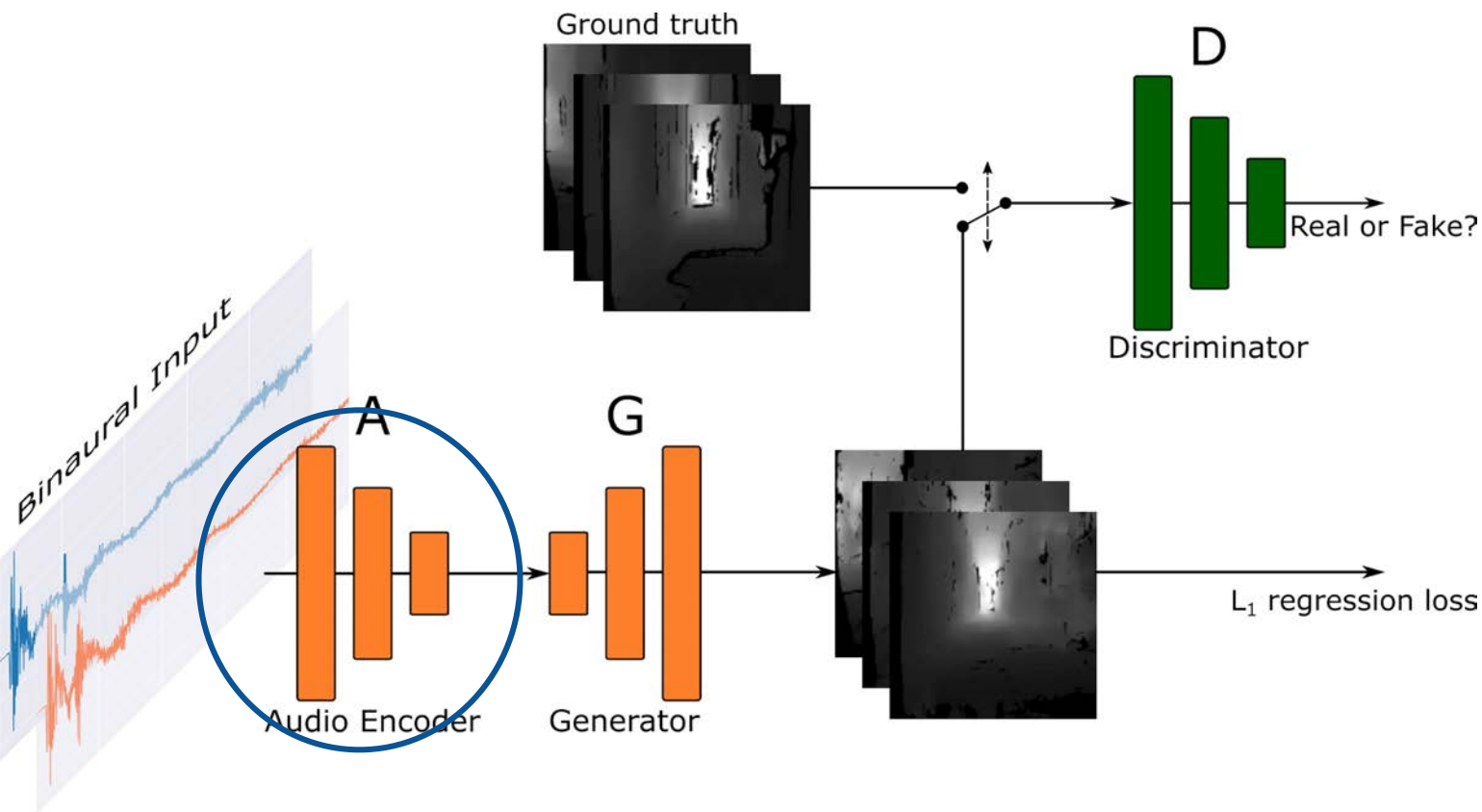
Structure Beyond can not be Directly Observed



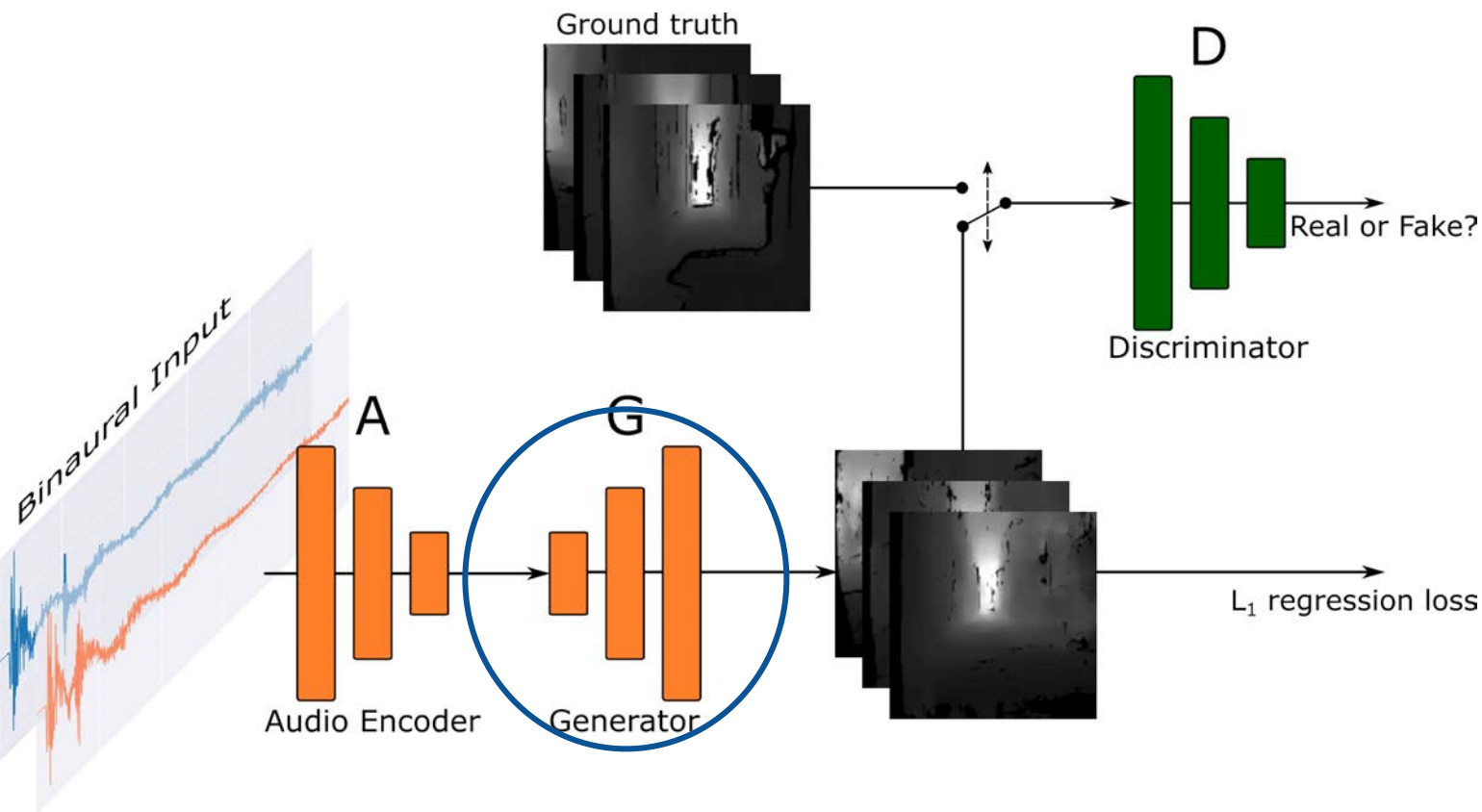
Architecture Overview



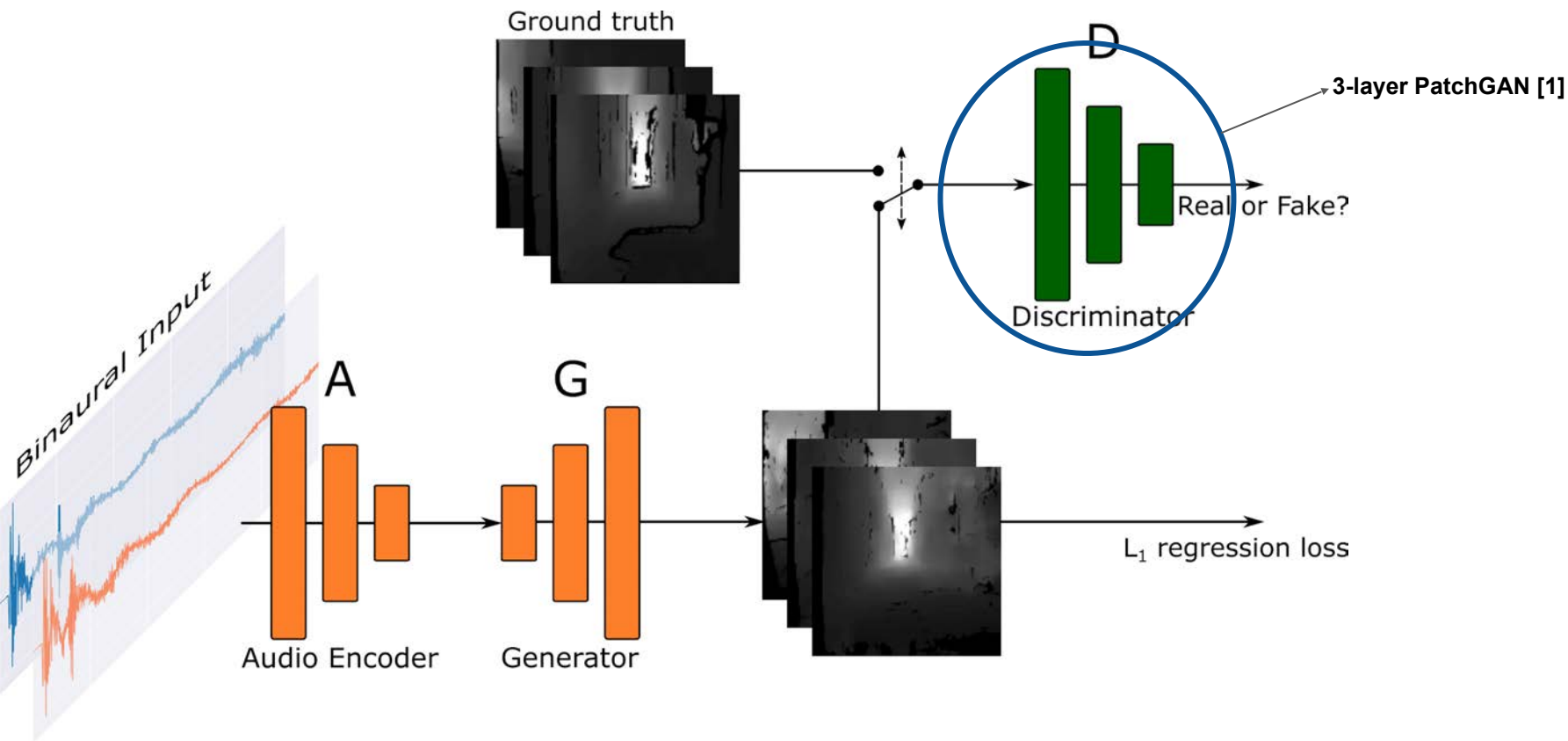
Architecture Overview



Architecture Overview

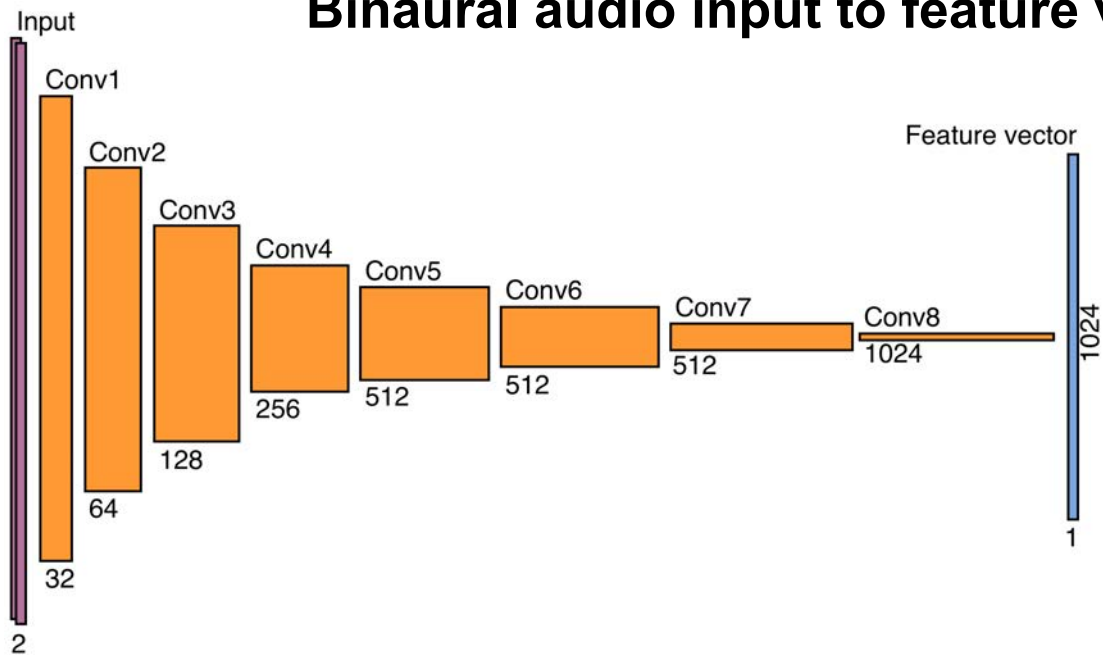


Architecture Overview



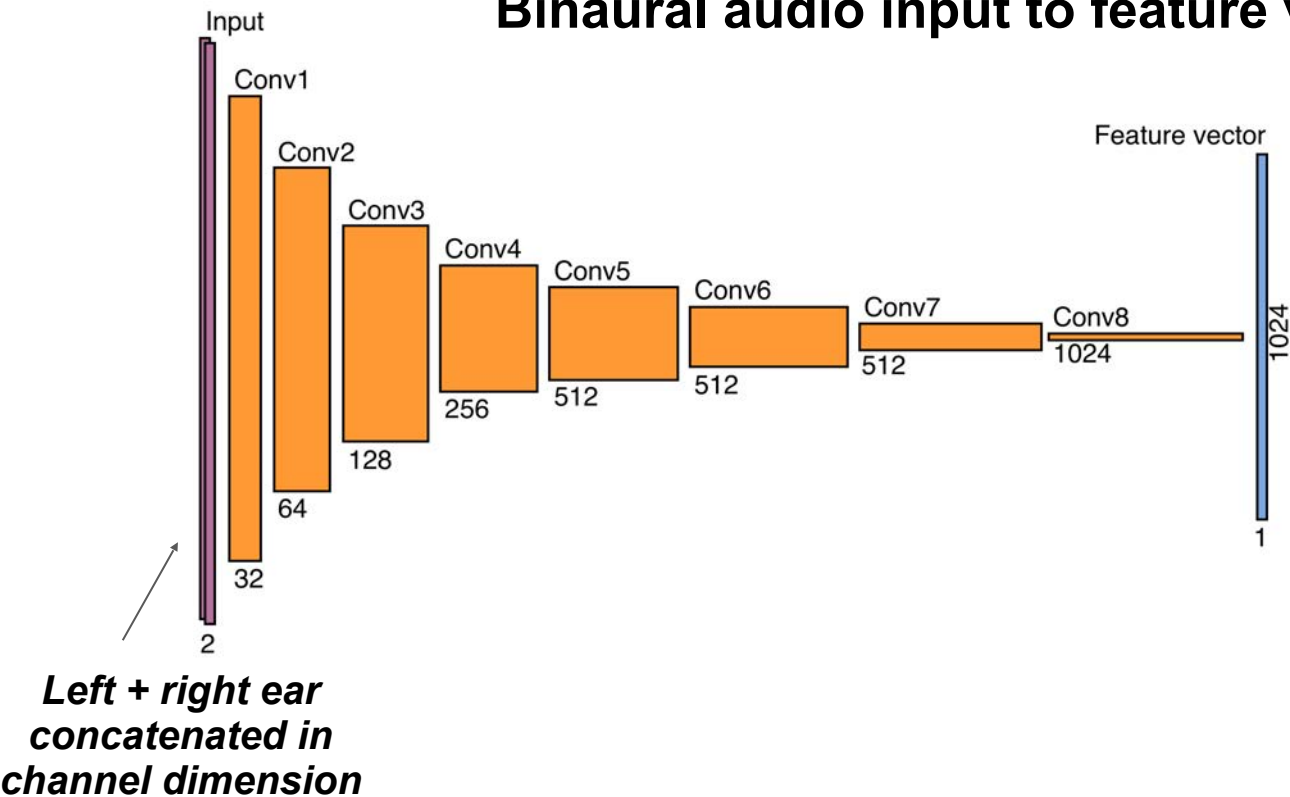
Architecture Overview

Binaural audio input to feature vector



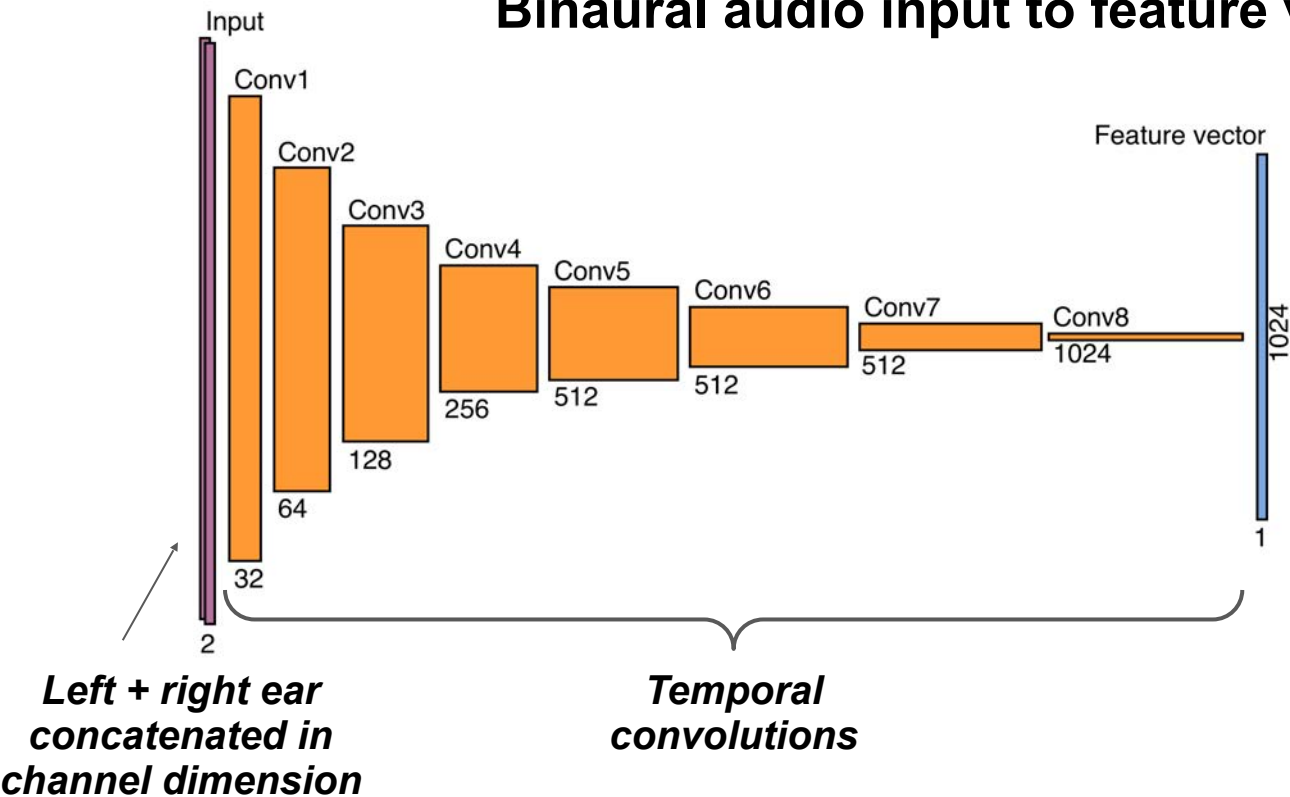
Architecture Overview

Binaural audio input to feature vector



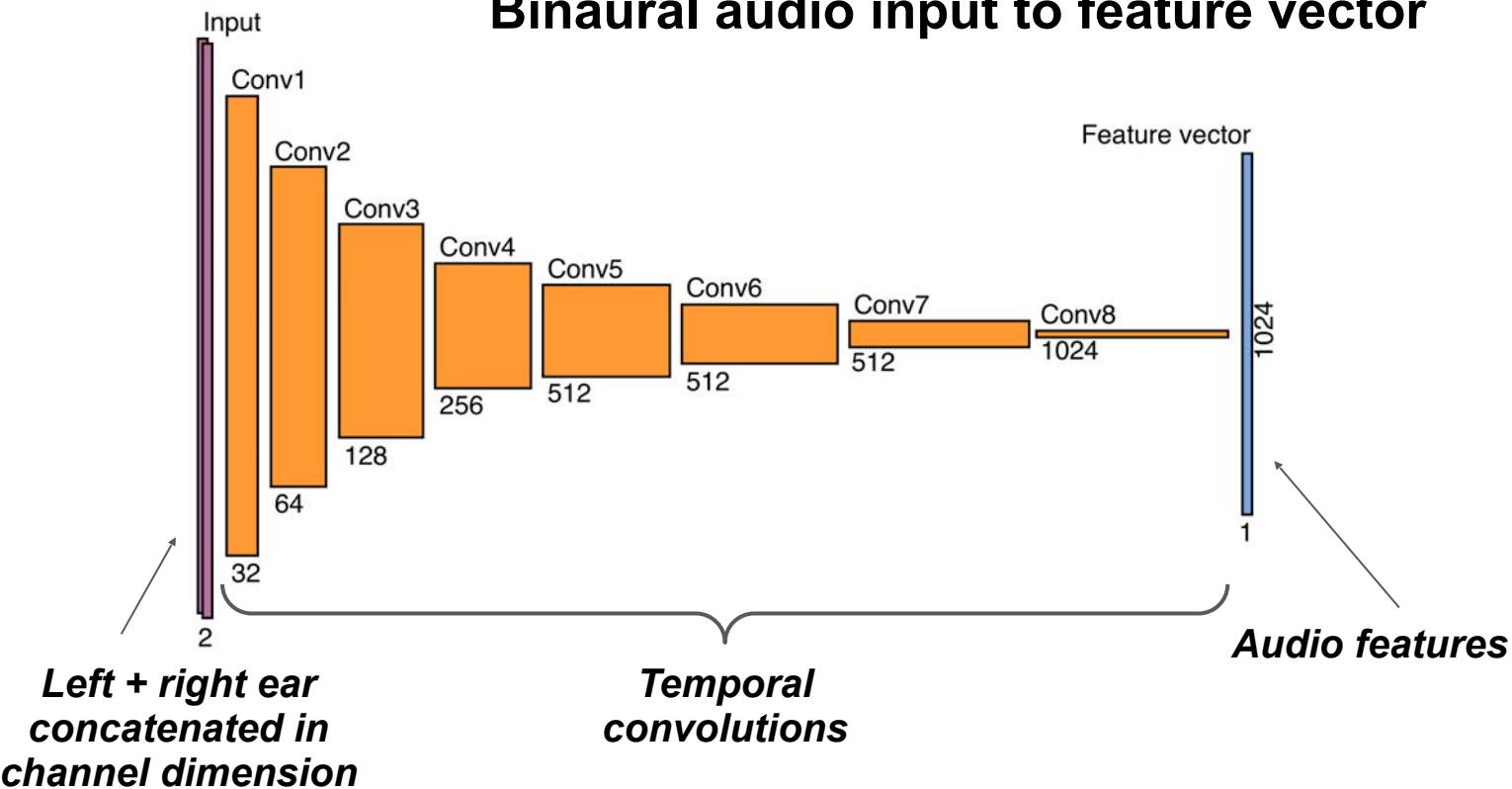
Architecture Overview

Binaural audio input to feature vector



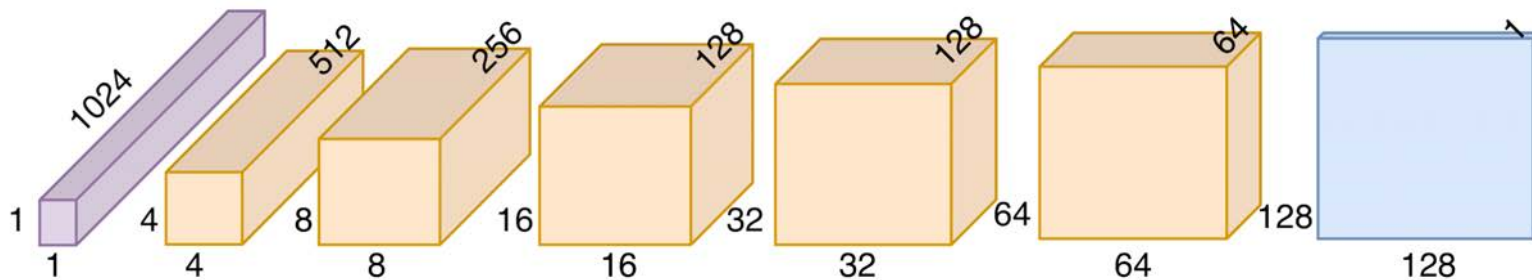
Architecture Overview

Binaural audio input to feature vector



Architecture Overview

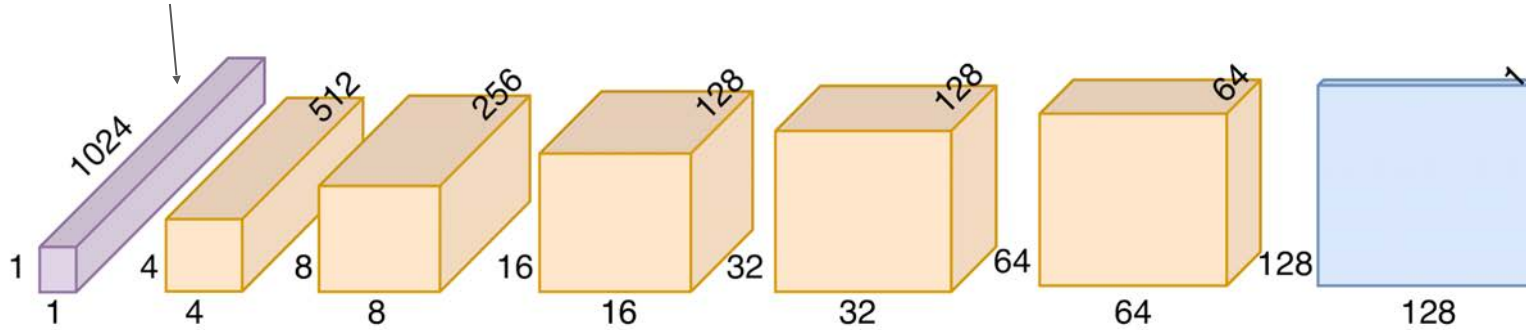
Audio features to depth prediction



Architecture Overview

Audio features to depth prediction

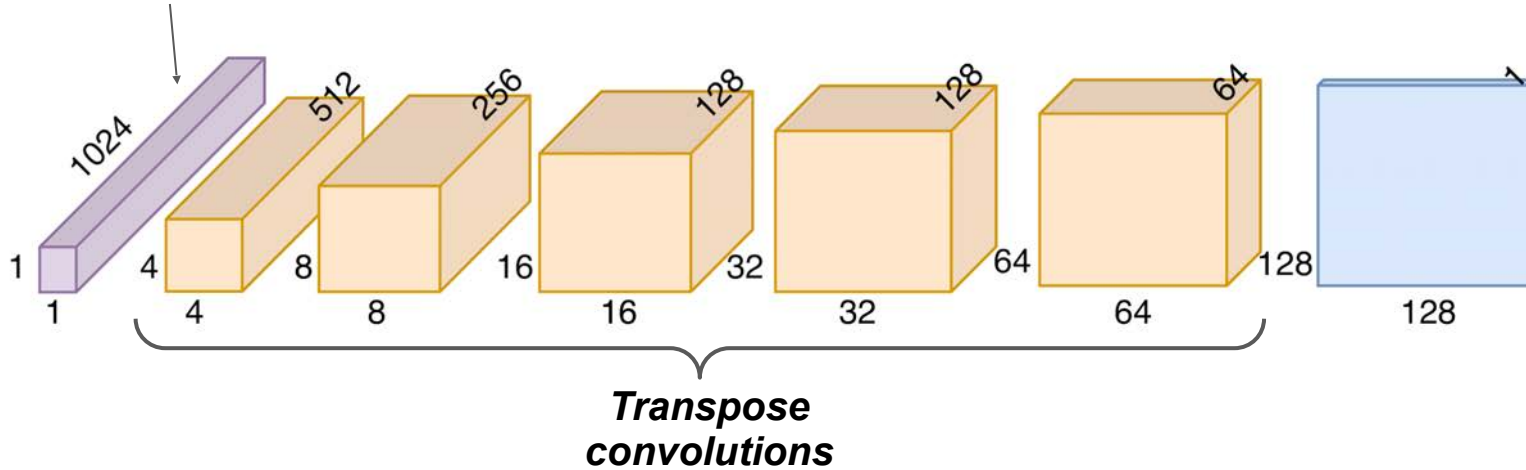
*Tiled and 1x1
augmented audio
features*



Architecture Overview

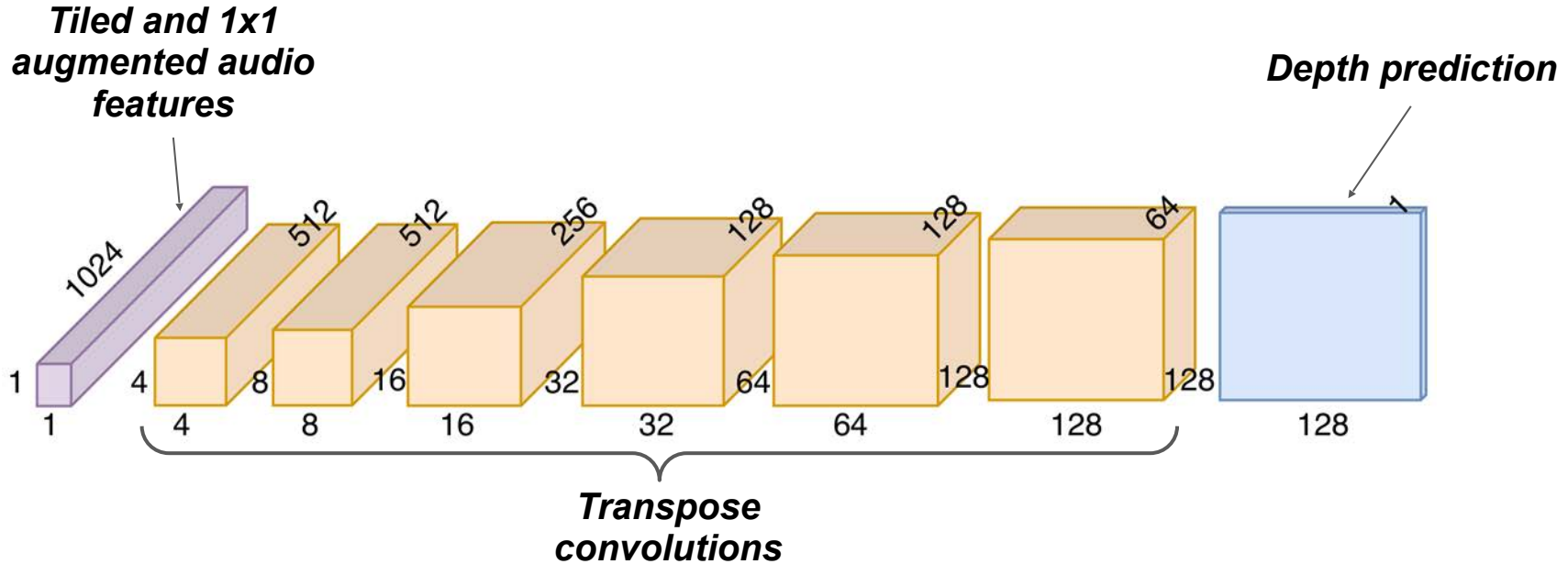
Audio features to depth prediction

*Tiled and 1x1
augmented audio
features*

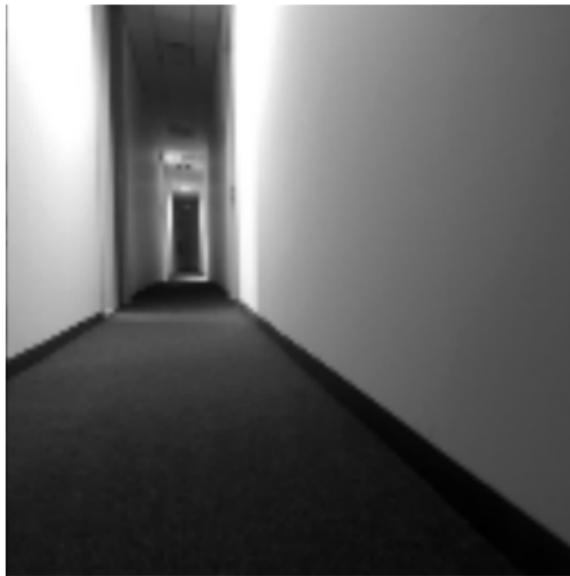


Architecture Overview

Audio features to depth prediction

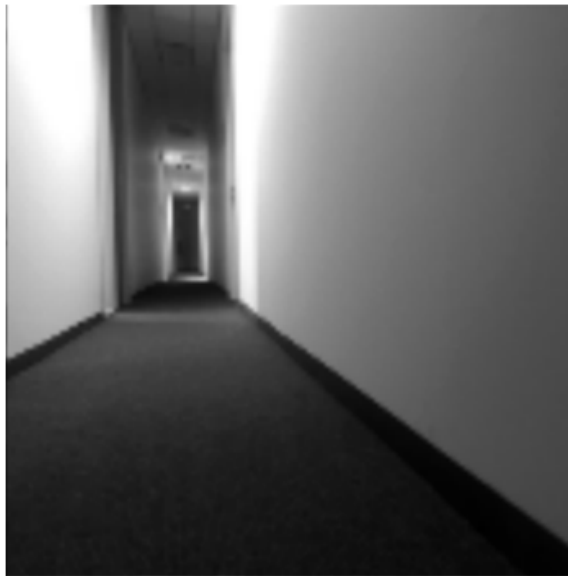


Results: Depth map prediction



Grayscale image from camera

Results: Depth map prediction

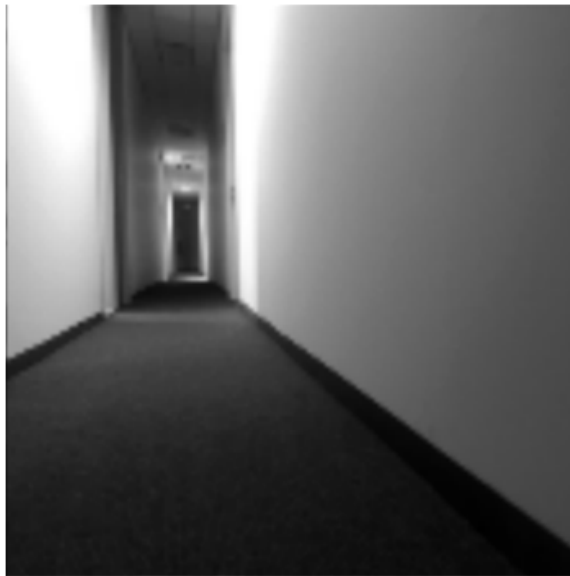


Grayscale image from camera



Depth map from stereo camera

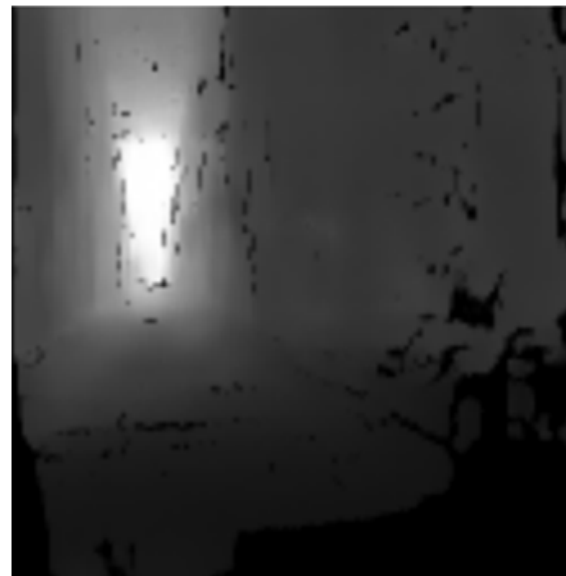
Results: Depth map prediction



Grayscale image from camera



Depth map from stereo camera



**Predicted depth map from
a single binaural echo**

Results: Depth map prediction

Ground Truth
Depth



Predicted
Depth

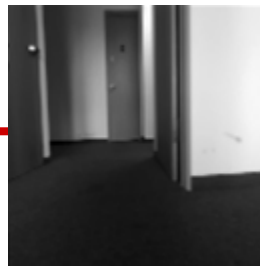
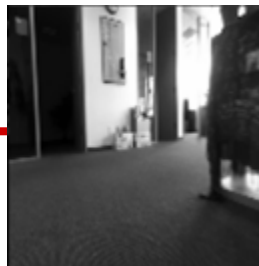
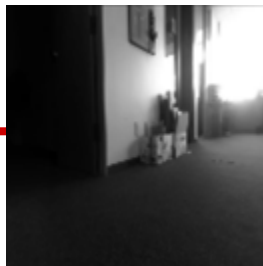
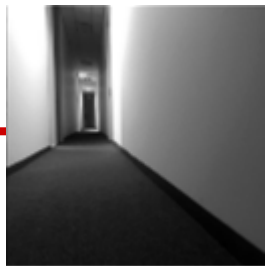


Ground Truth
Grayscale



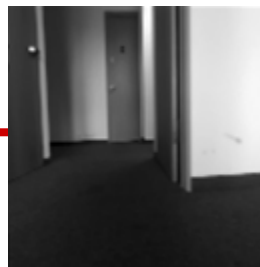
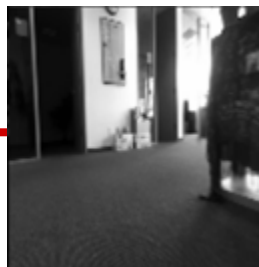
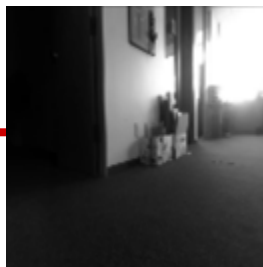
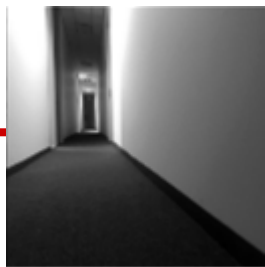
The depth reconstruction is robust against vision artefacts from failed depth calculations of the stereo to depth algorithm.

Results: Error prone to vision artifacts

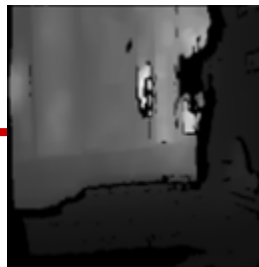


Scene

Results: Error prone to vision artifacts

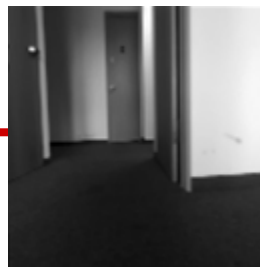
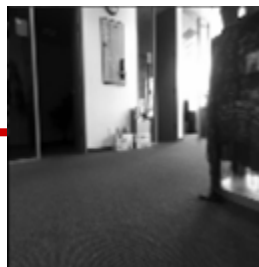
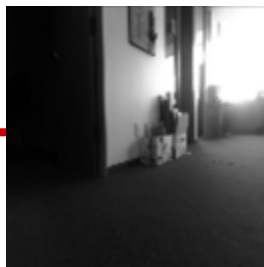
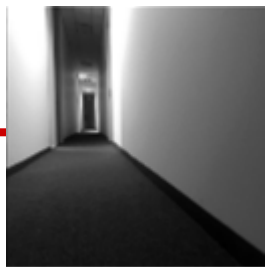


Scene

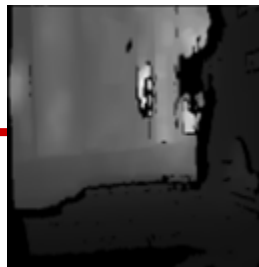


Stereo Camera

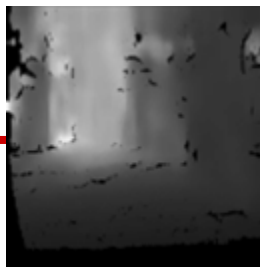
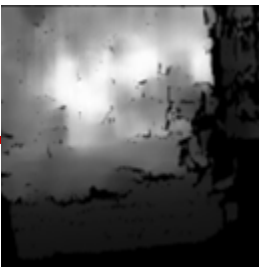
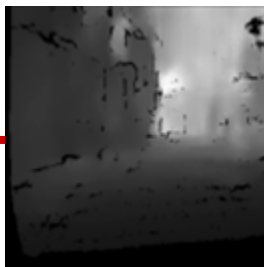
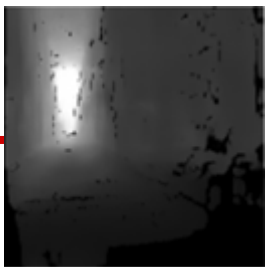
Results: Error prone to vision artifacts



Scene



Stereo Camera



Prediction

Results: Grayscale layout

Plausible layout of free space / obstacles

*No depth used
for training!*

Results: Grayscale layout

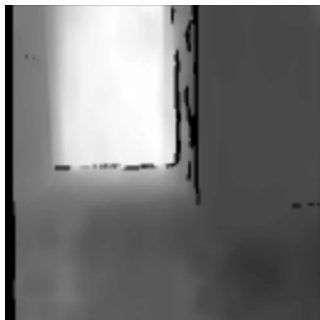
Plausible layout of free space / obstacles

*No depth used
for training!*

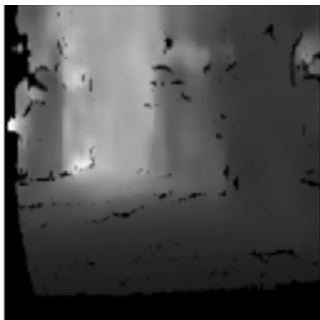
Conclusion

- First sound-to-**vision** model
- Depth prediction
- Grayscale “layout” prediction
- Failure cases at close range and dense scenes

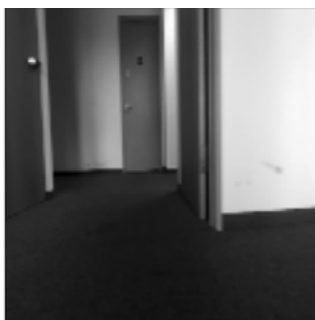
Stereo depth



BatVision depth



Grayscale image



BatVision layout

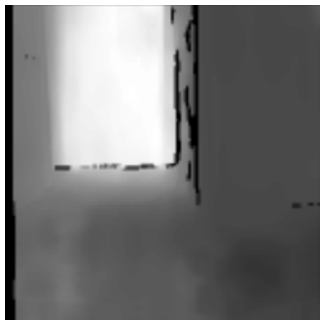


Paper: <https://arxiv.org/abs/1912.07011>

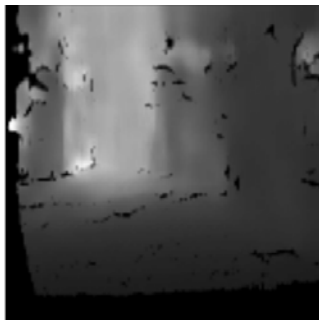
Improving BatVision

Generalized Cross-Correlation Features + architecture modifications

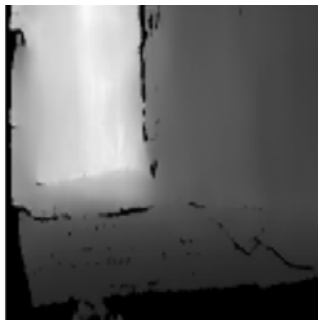
Stereo depth



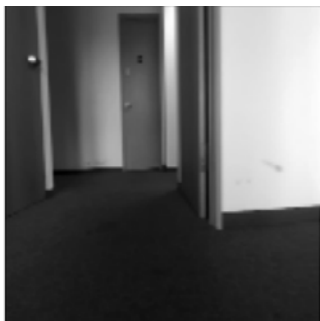
BatVision depth



New



Grayscale image



BatVision layout



New

