EDIC Seminar 2022

Amir Zamir, VILAB.

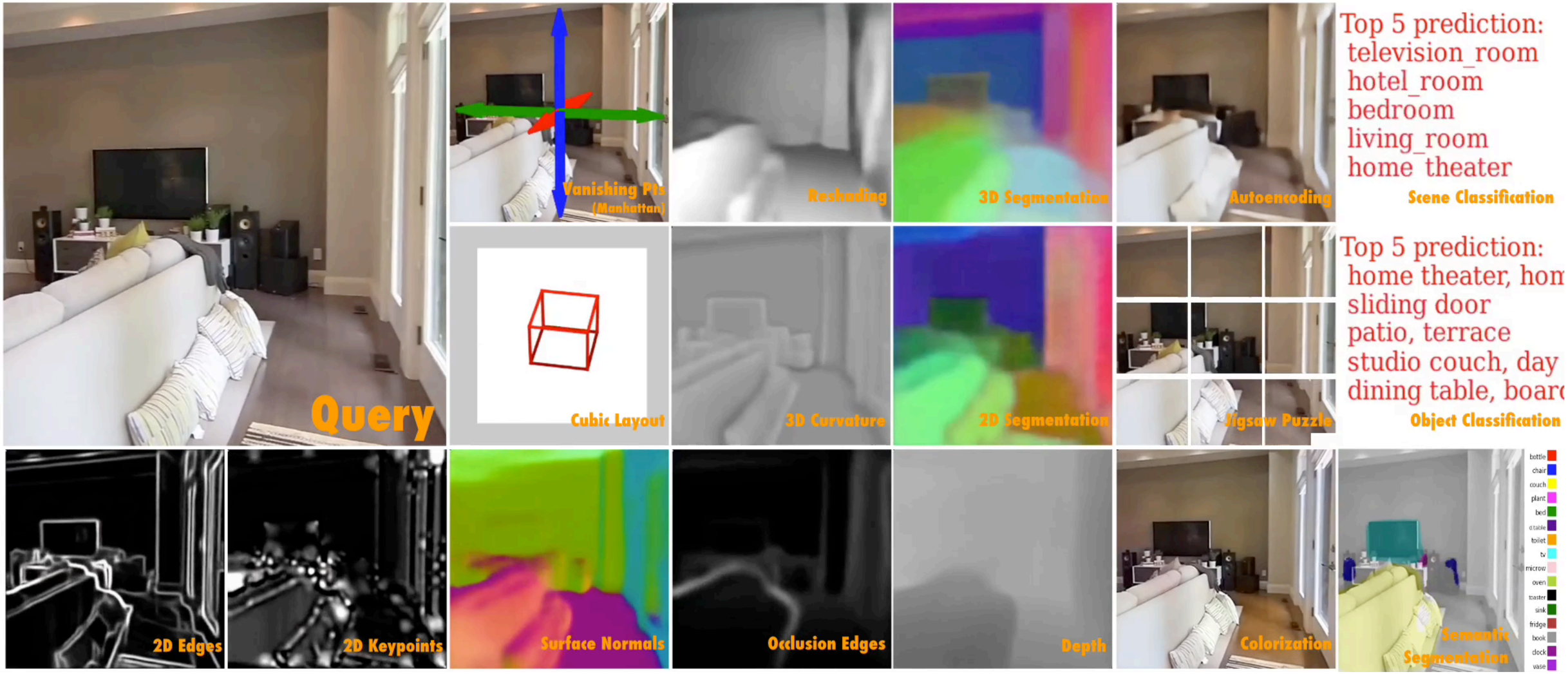
06.09.2022

École
polytechnique
fédérale
de Lausanne

Research Areas

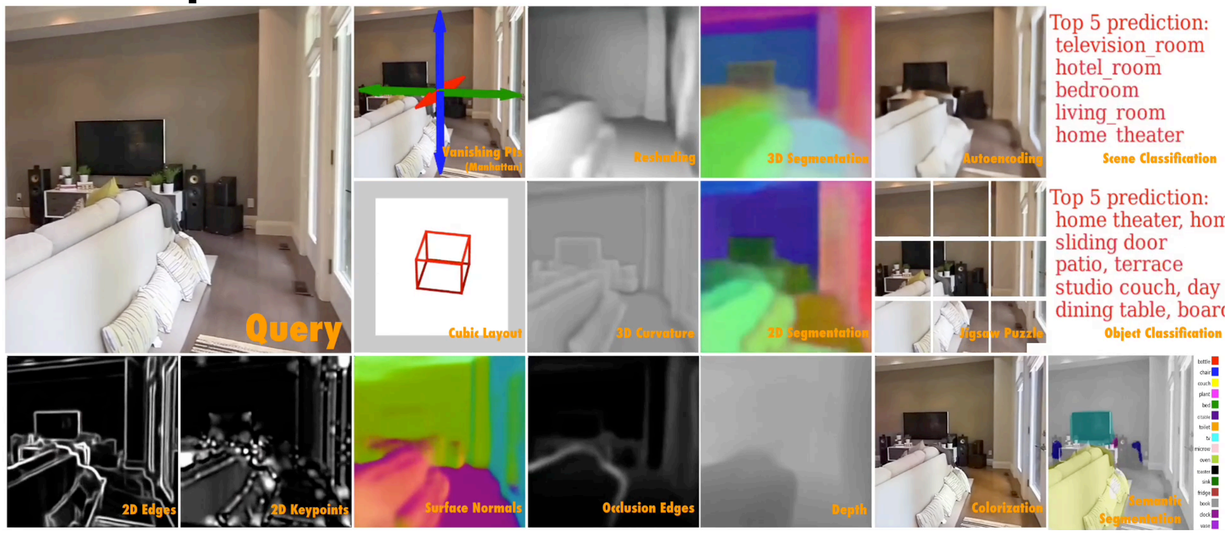
Research Areas

- Computer Vision:

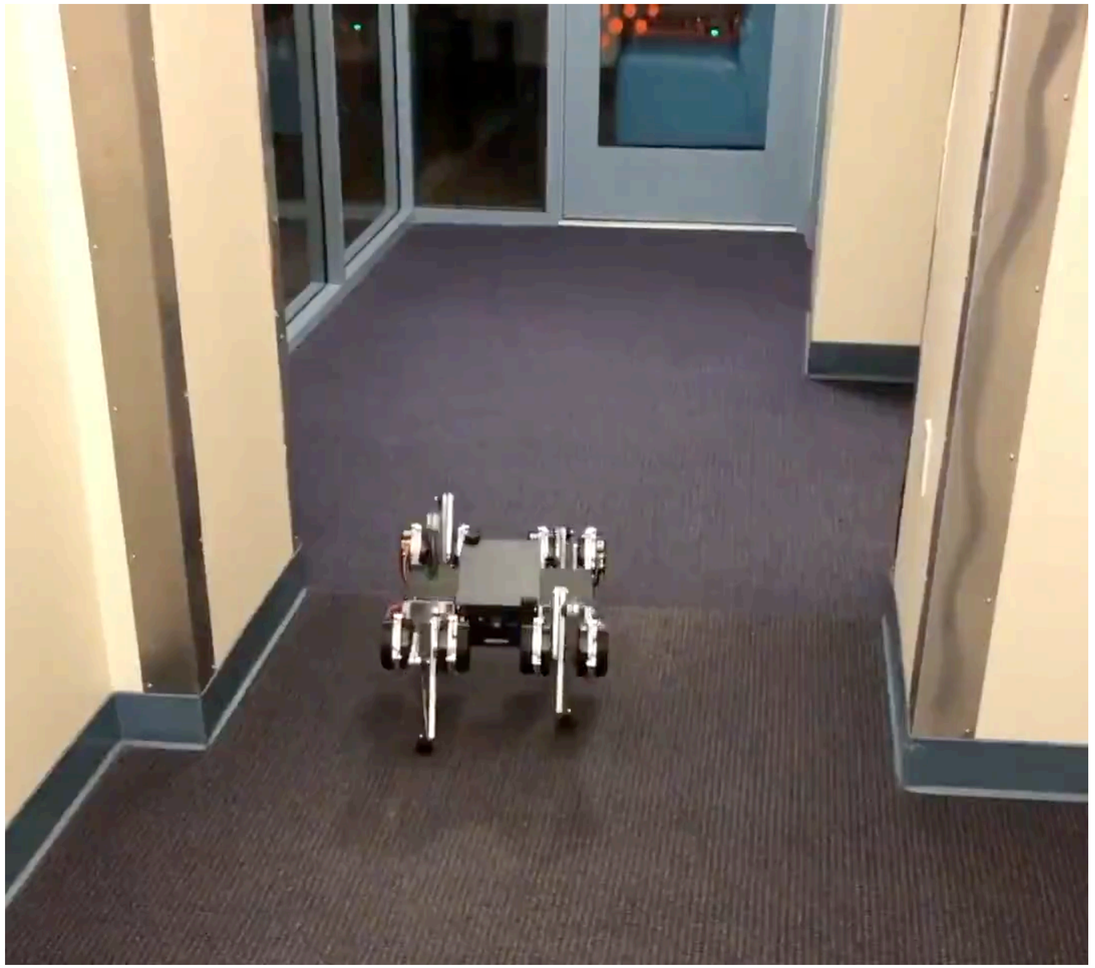


Research Areas

- Computer Vision:

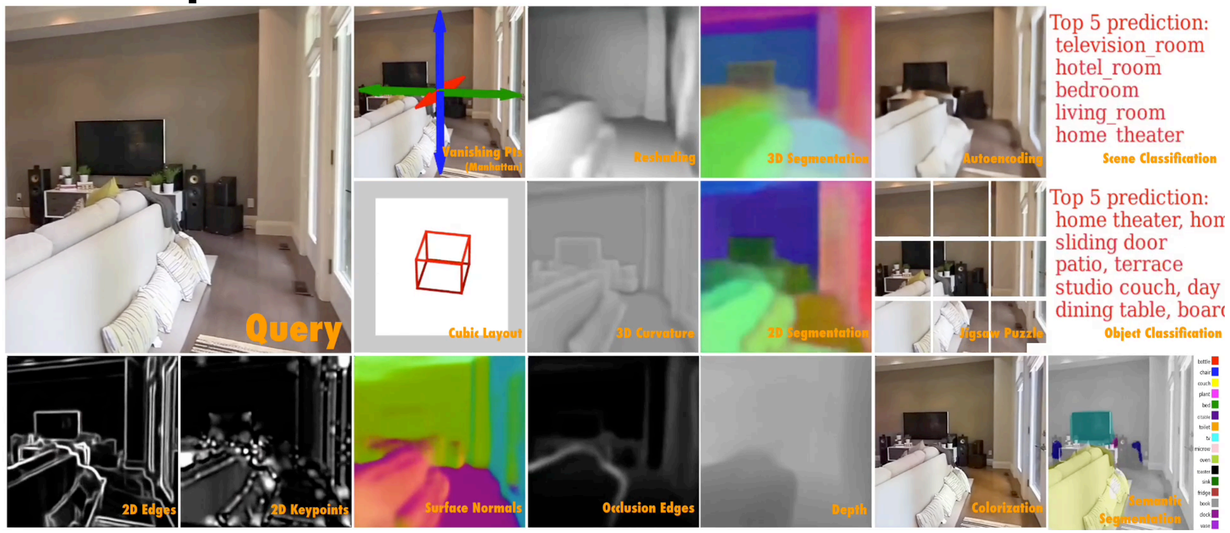


- Perception-for-Robotics:

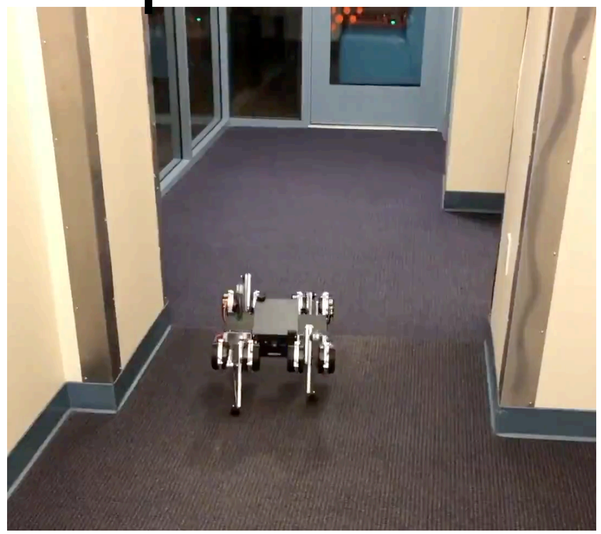


Research Areas

- Computer Vision:



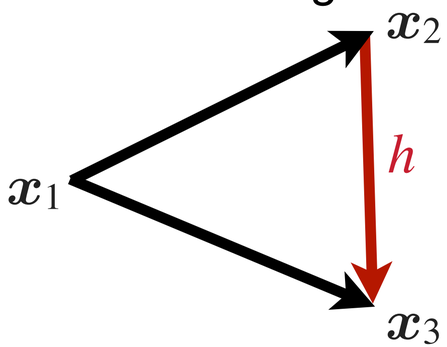
- Perception-for-Robotics:



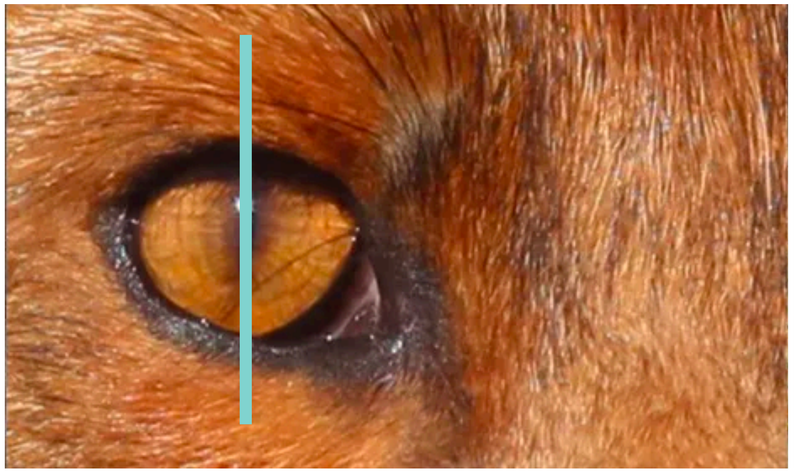
- Machine Learning:

Baseline Learning

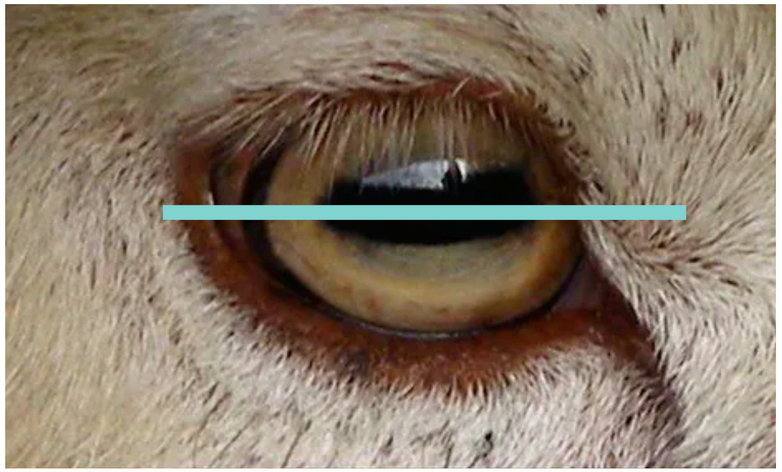
Cross-Task Consistent Learning



Vision ⟷ Action



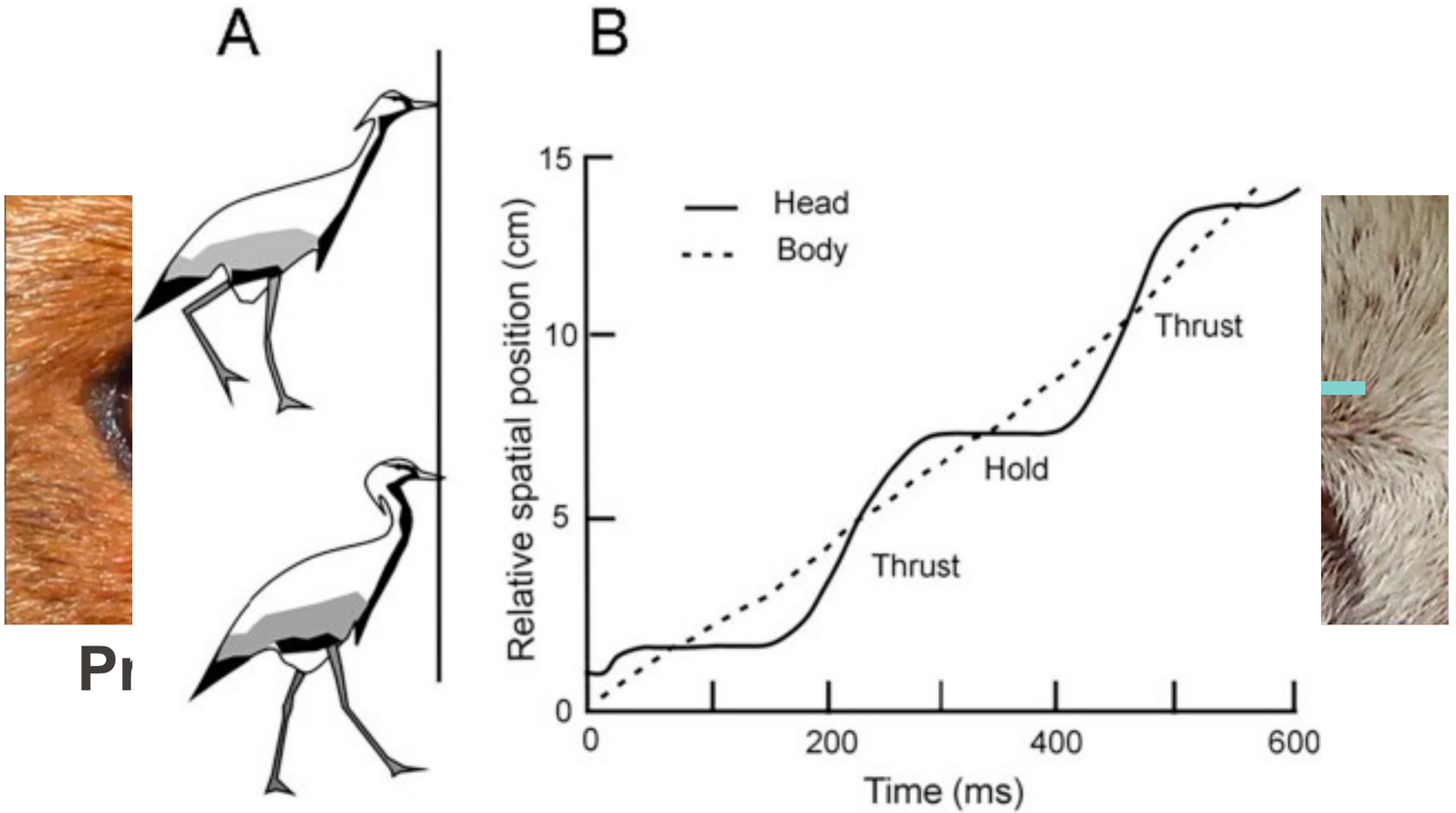
Predator



Prey

M. Bank et al., 2015.
M. Land. 2002

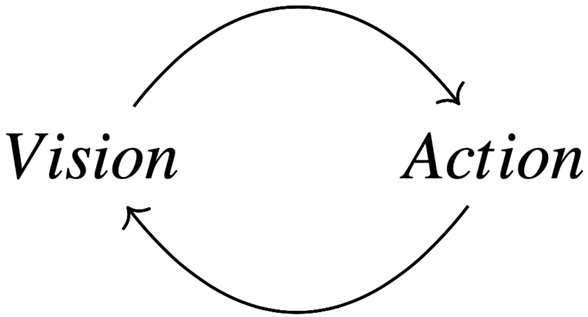
Vision ⟷ Action

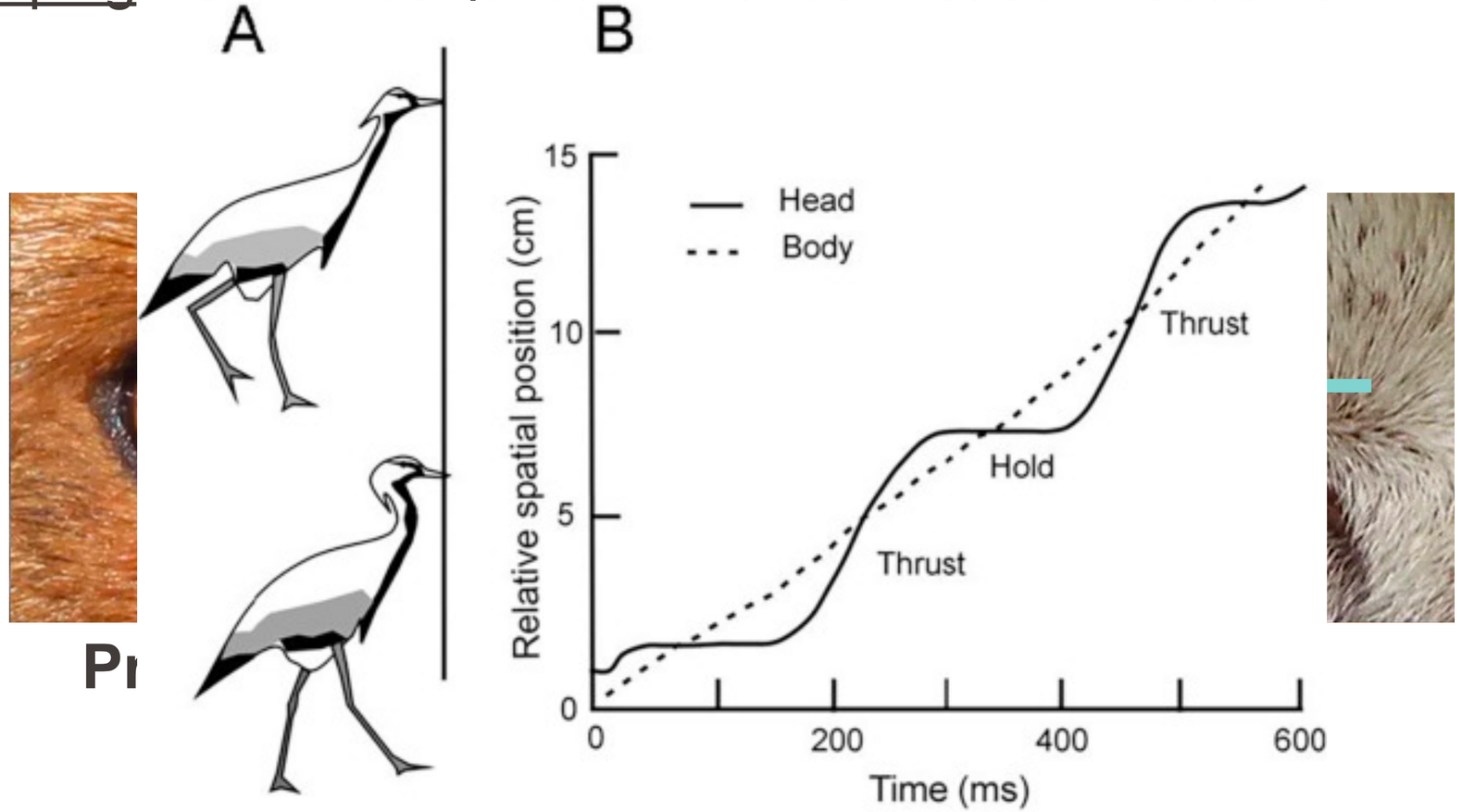


M. Bank et al., 2015.
M. Land. 2002

Vision ⟷ Action

- **Conventional Computer Vision**: primarily developed in isolation from action.
- Understanding and developing vision in conjunction with downstream “action”.



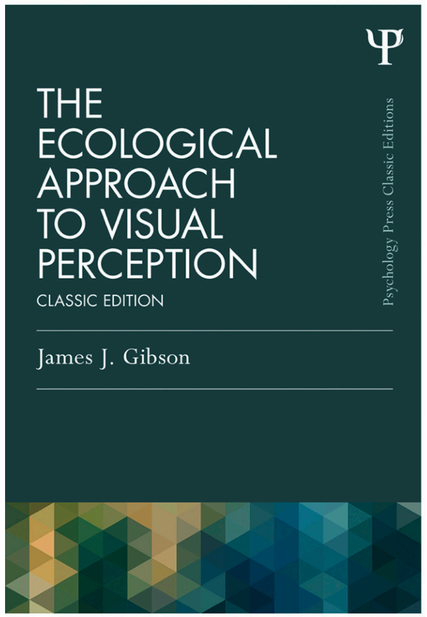


M. Bank et al., 2015.
M. Land. 2002

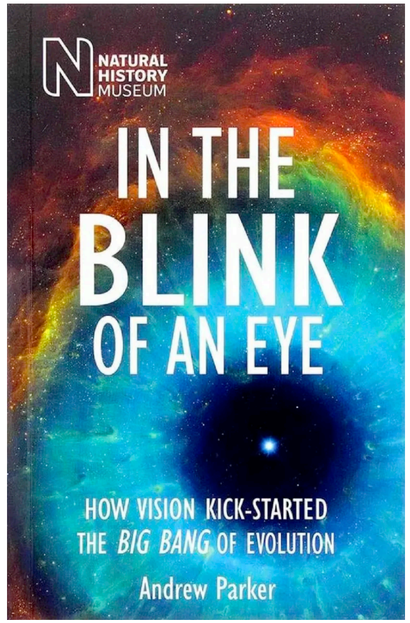
Vision ⟷ Action

- **Conventional Computer Vision**: primarily developed in isolation from action.
- Understanding and developing vision in conjunction with downstream “action”.

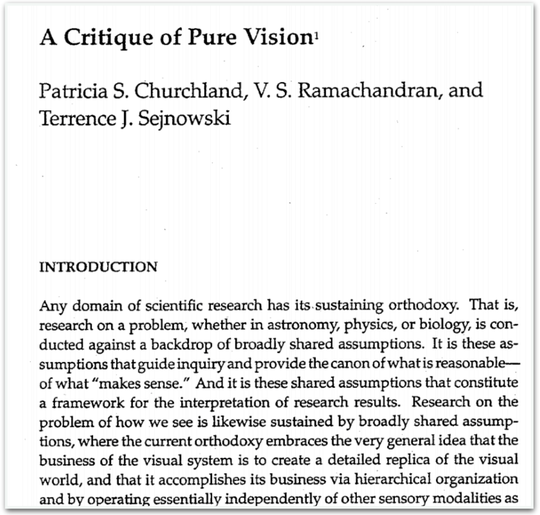
The “ecological vision” argument



Light switch theory, A. Parker



A Critique of Pure Vision, Churchland et al.:
Vision is meant to support “the four Fs: feeding, fleeing, fight, and reproducing”.



A Critique of Pure Vision[1]

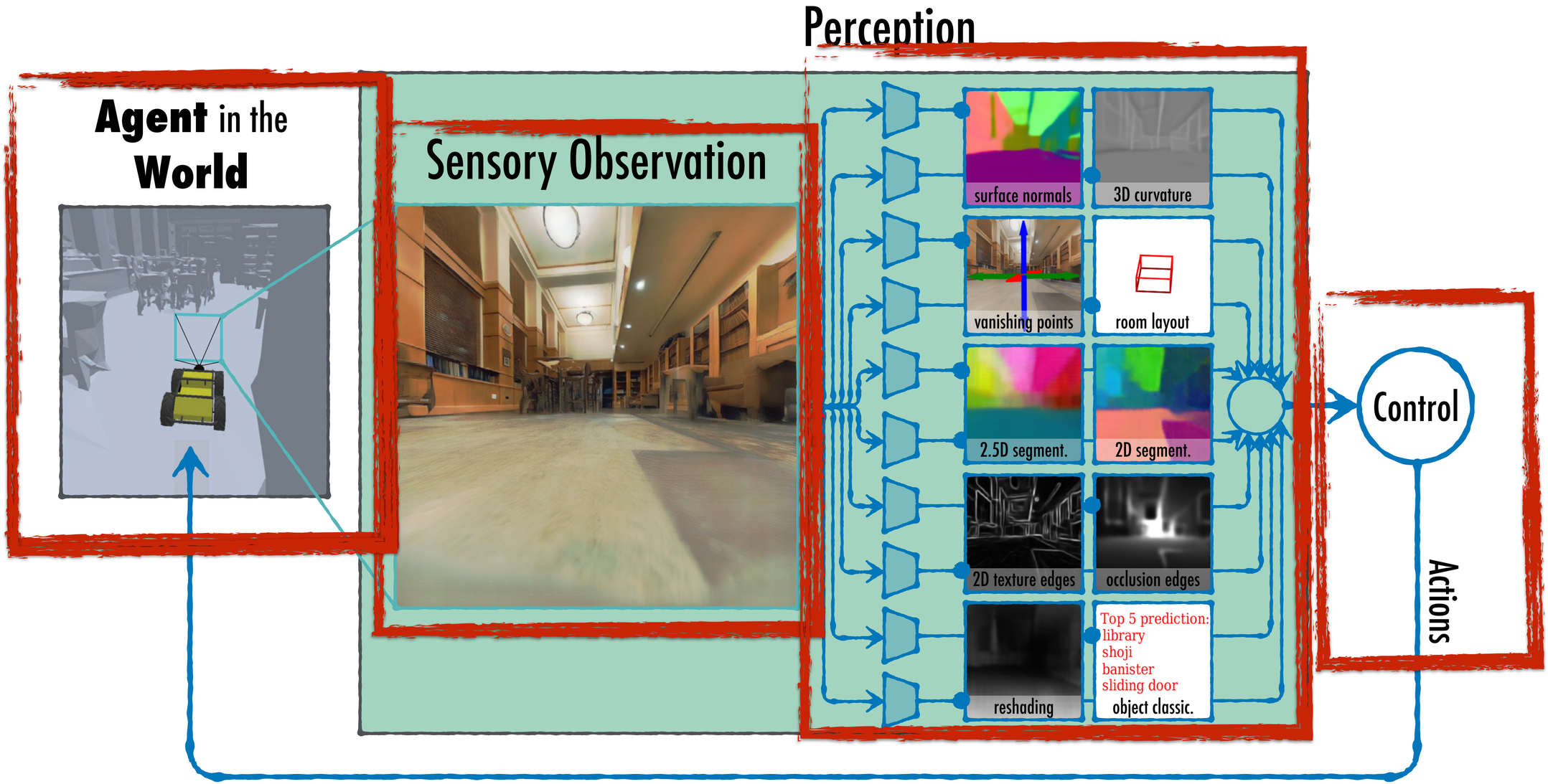
Patricia S. Churchland, V. S. Ramachandran, and Terrence J. Sejnowski

INTRODUCTION

Any domain of scientific research has its sustaining orthodoxy. That is, research on a problem, whether in astronomy, physics, or biology, is conducted against a backdrop of broadly shared assumptions. It is these assumptions that guide inquiry and provide the canon of what is reasonable—of what “makes sense.” And it is these shared assumptions that constitute a framework for the interpretation of research results. Research on the problem of how we see is likewise sustained by broadly shared assumptions, where the current orthodoxy embraces the very general idea that the business of the visual system is to create a detailed replica of the visual world, and that it accomplishes its business via hierarchical organization and by operating essentially independently of other sensory modalities as

Animal Eye, M. Land

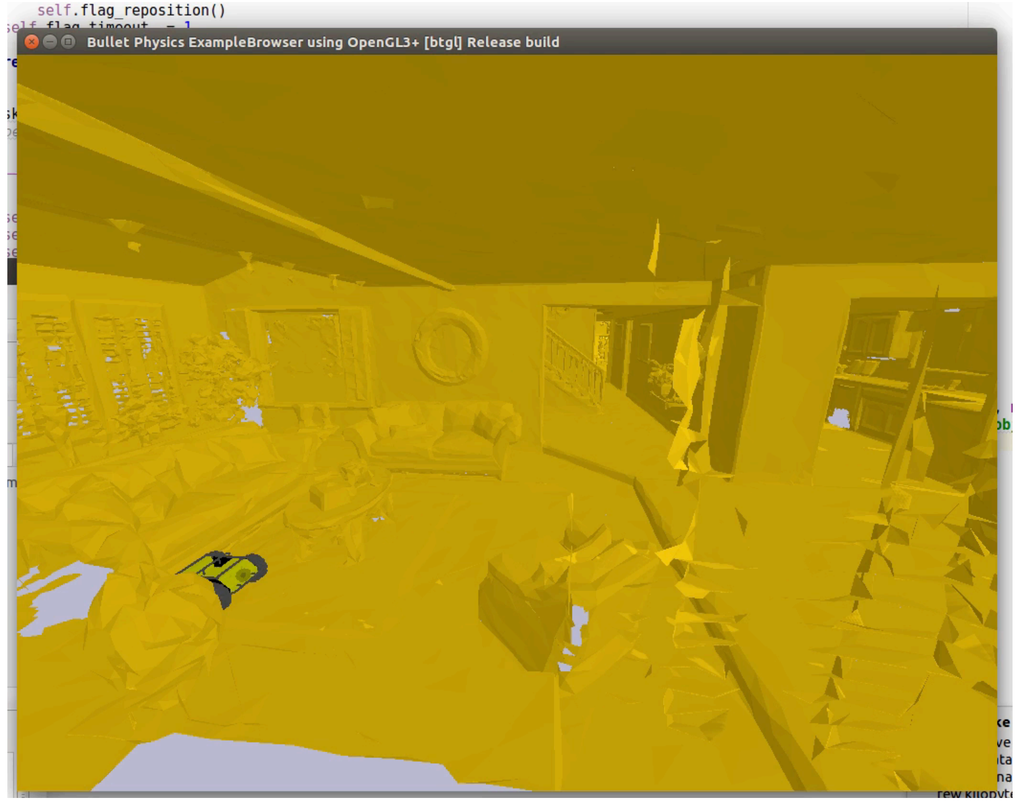
Vision In-the-loop



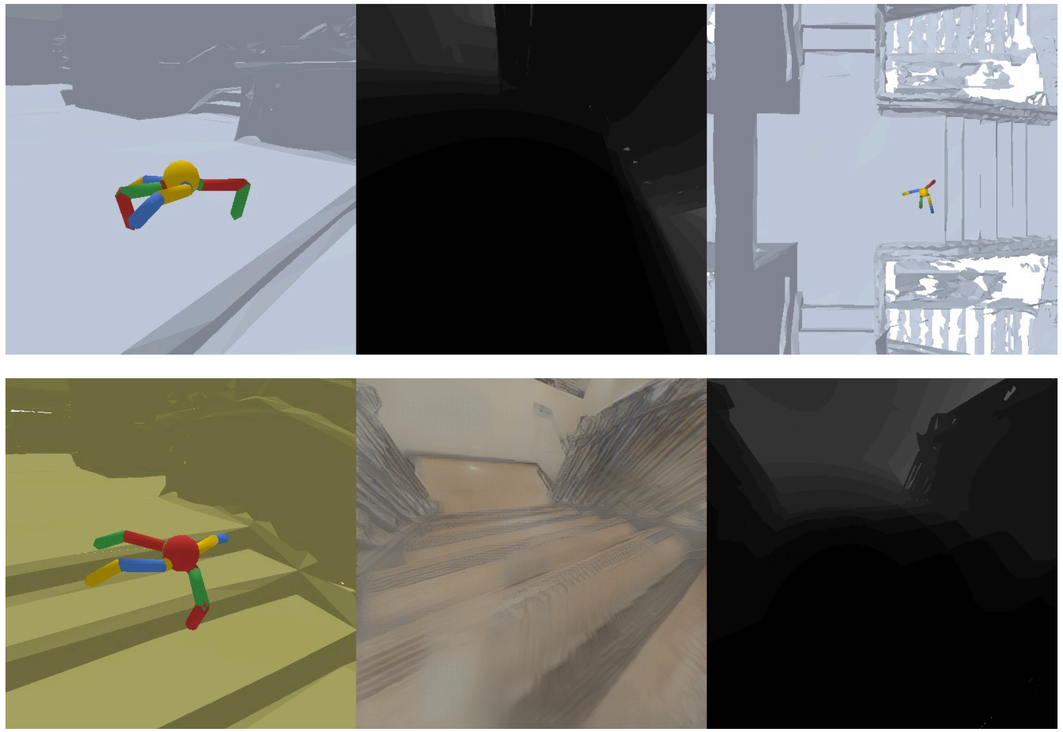
Perceptual Robotic Simulators

- Gibson Environment, Omnidata. 2018, 2021.

Perceptual Robotics



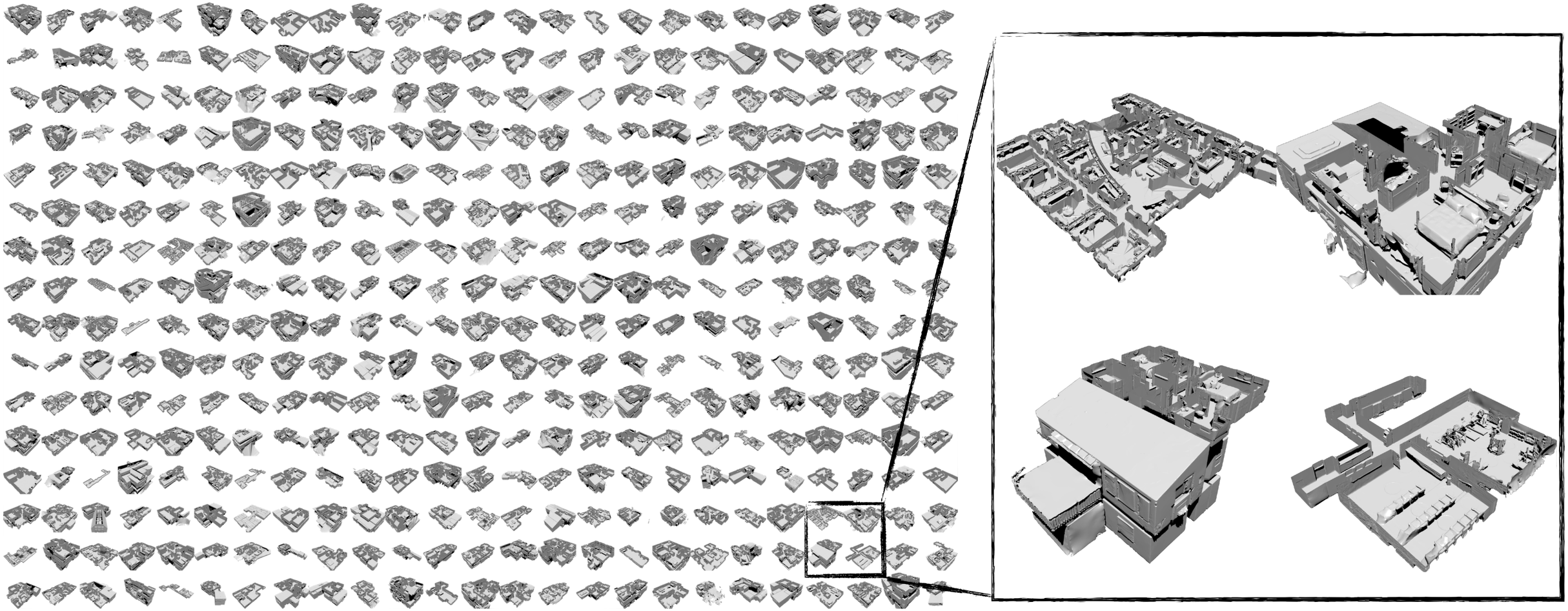
Local planning ("go to the target")



Stair climb

- Mid-Level Vision for Robotics. 2018, 2019, 2020.

3D Vision



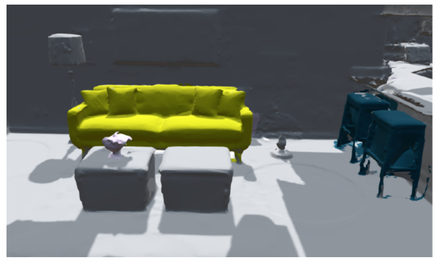
- Cross-Task Consistency, Cross-Domain Ensembles, Omnidata, 2D-3D-S, 3D Common Corruptions, etc. 2018, 2020, 2021, 2022

OmniData, ICCV2021.

3D textured mesh

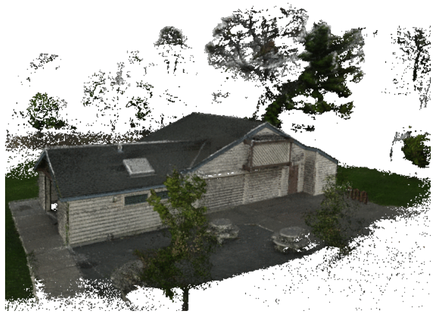
(E.g. Yao et al 2020)

3D mesh + Aligned RGB



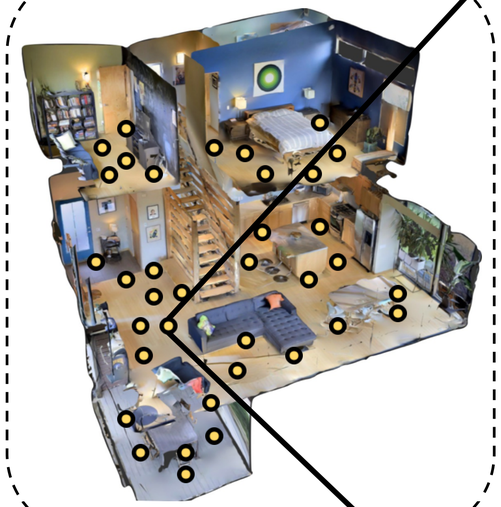
(E.g. Armeni et al 2017)

3D pointcloud + Aligned RGB

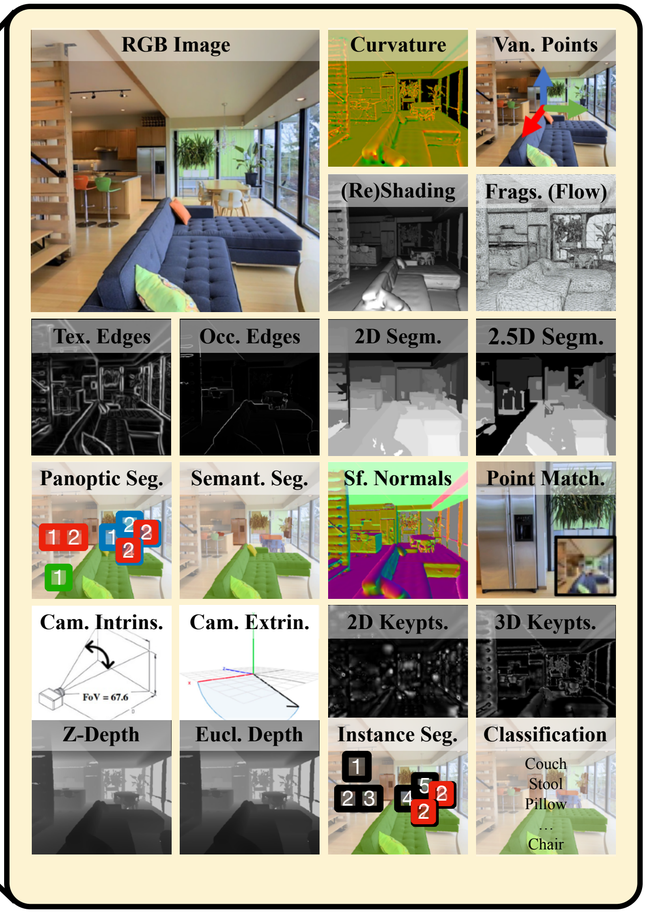


(E.g. Knapitsch et al 2017)

Dense Viewpoint Sampling

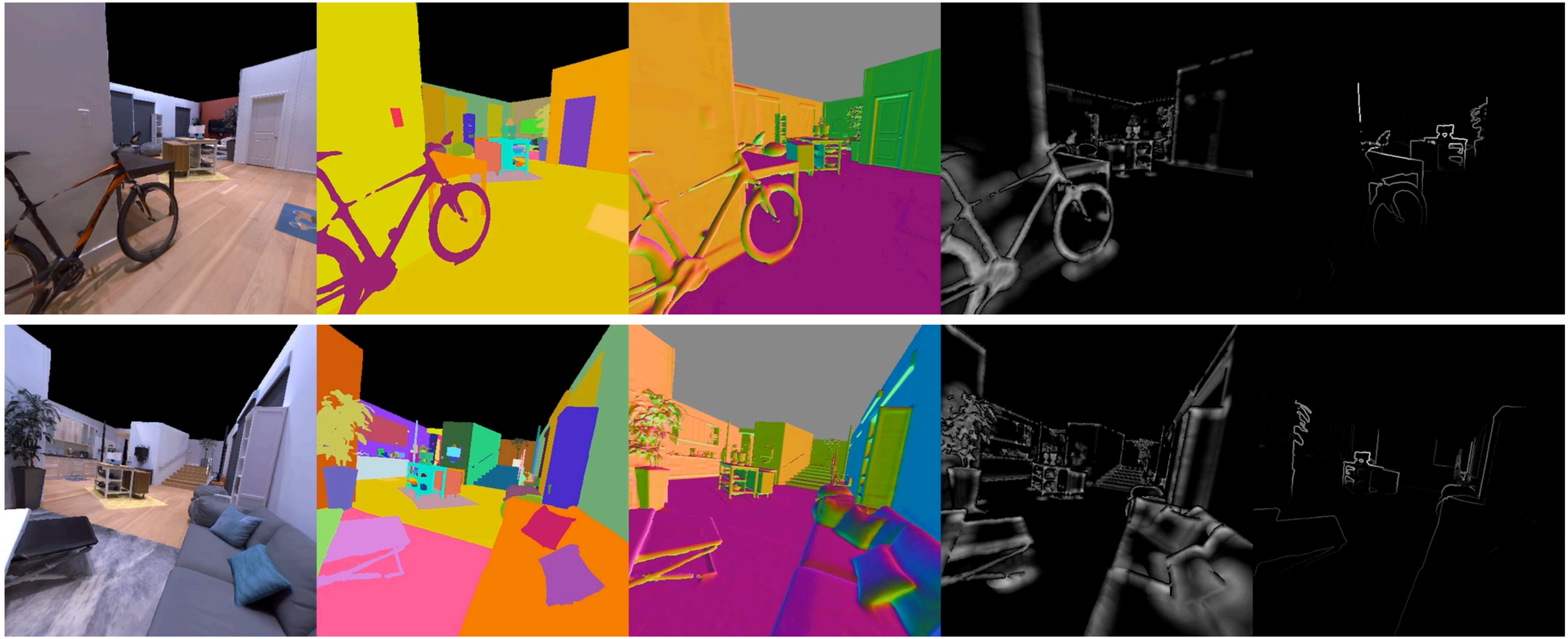


Mid-Level Cues



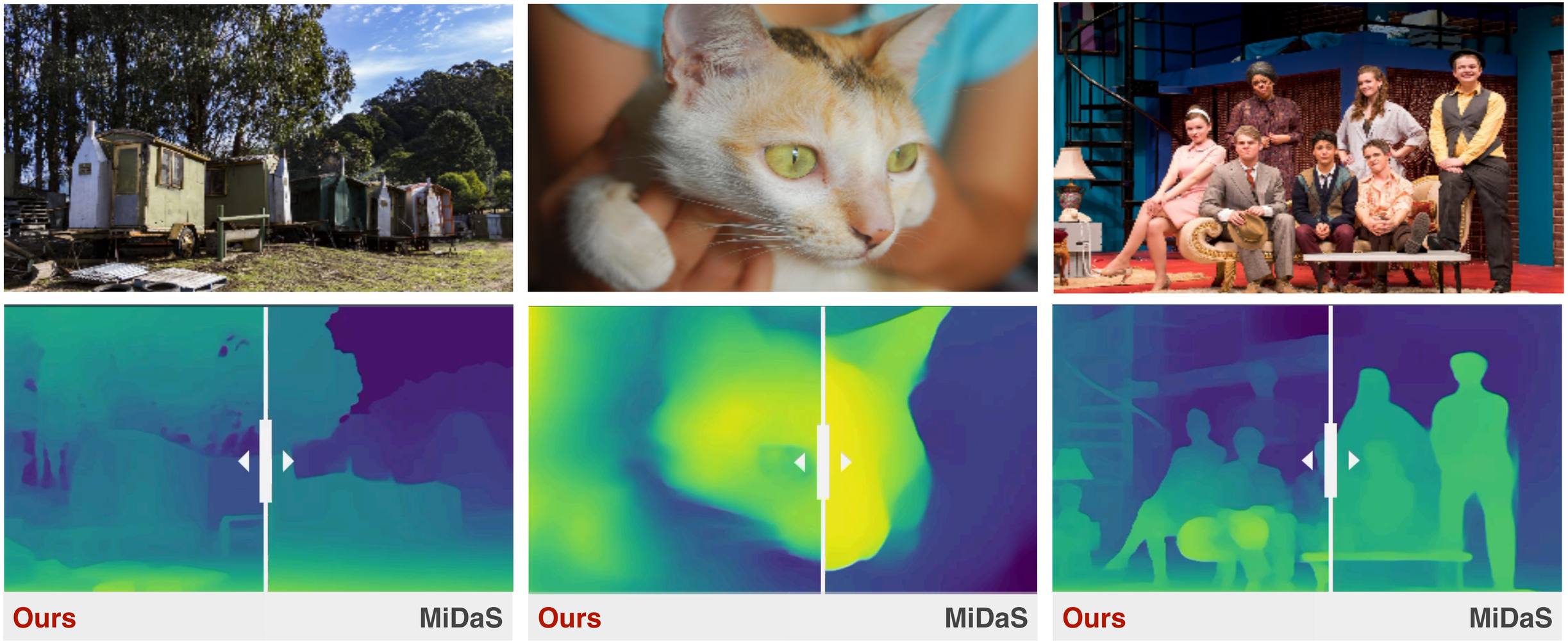
OmniData, ICCV2021.

RGB | Semantics | Surface Normals | 3D Keypoints | Occlusion Edges



OmniData, ICCV2021.

Task : Depth Estimation

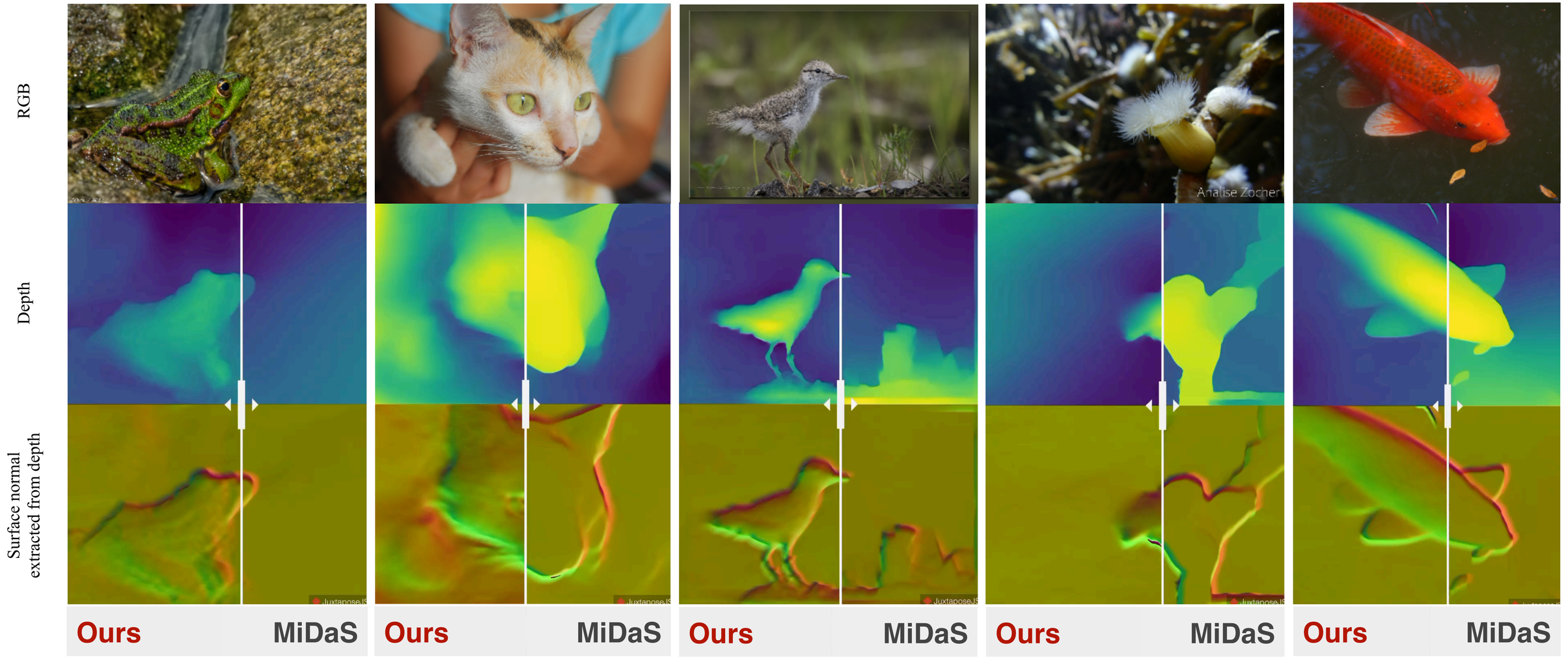


[Ranftl et al. TPAMI 2020]

[Ranftl et al. 2021]

OmniData, ICCV2021.

Task : Depth Estimation



[Ranftl et al. TPAMI 2020]

[Ranftl et al. 2021]

Input Image

Upload your own image or click on one of the sample queries below. Click on the cube to use a random query image from previous uploads.

refresh to upload new image

Verification expired. Check the checkbox again.

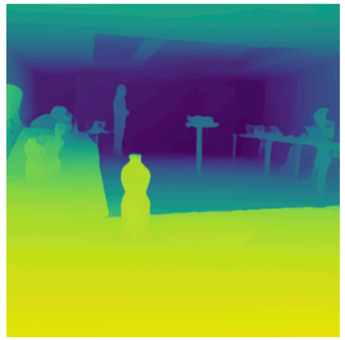
I'm not a robot

reCAPTCHA

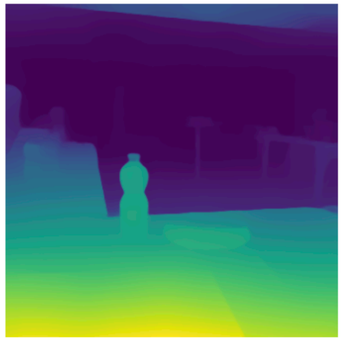
Privacy - Terms

Depth Estimation

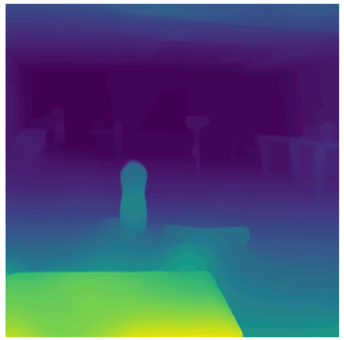
Omnidata Depth



MiDaS Depth

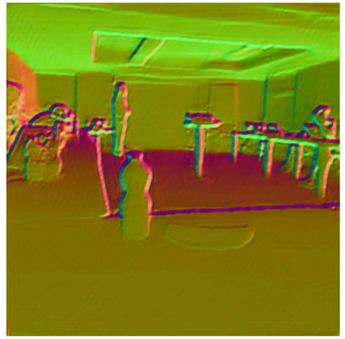


Taskonomy Depth (X-TC)

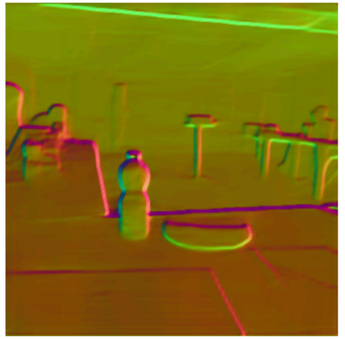


Surface Normals Extracted from Predicted Depth

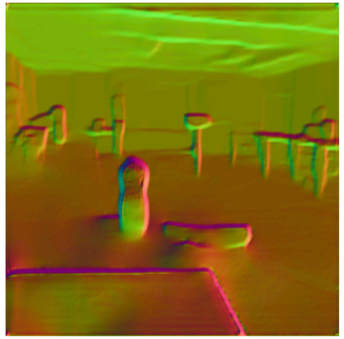
Depth → Normal (Omnidata)



Depth → Normal (MiDaS)

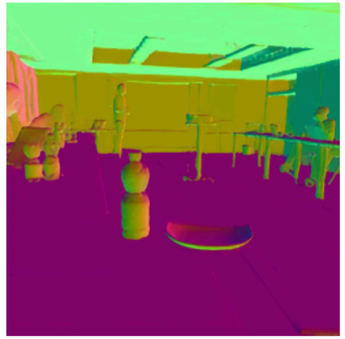


Depth → Normal (X-TC)

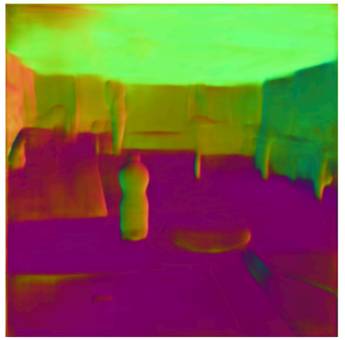


Surface Normal Estimation

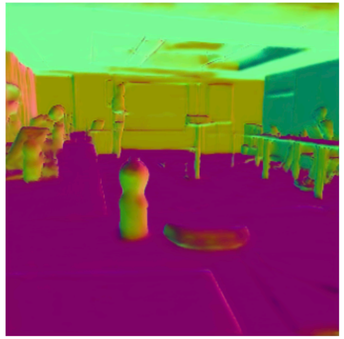
Omnidata Normal



Oasis Normal



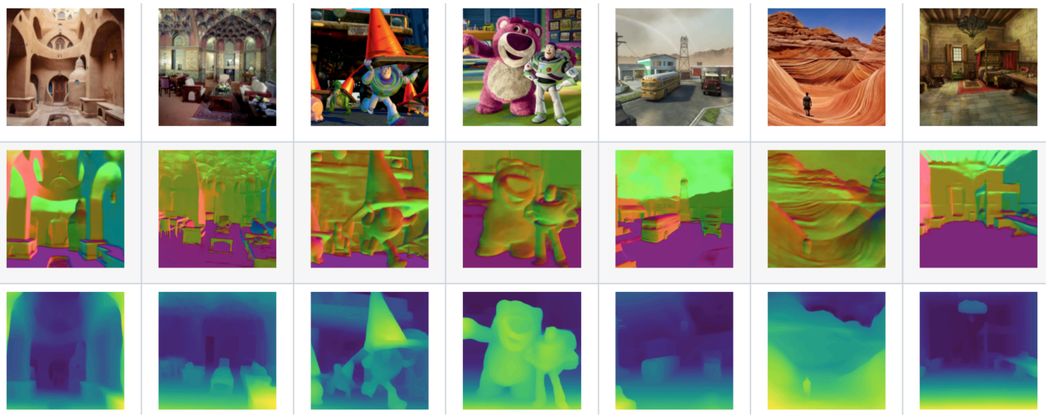
Taskonomy Normal (X-TC)



- Omnidata, ICCV2021.
- 3D Common Corruptions CVPR

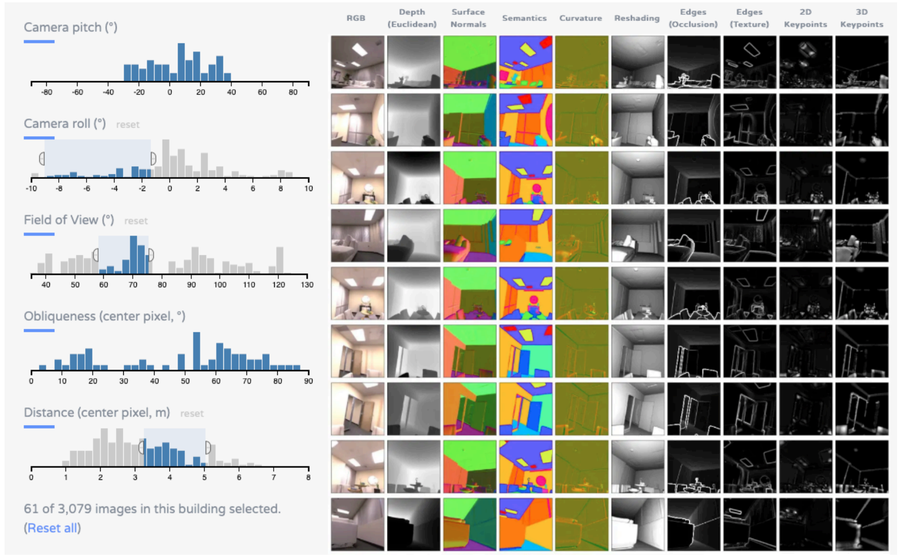
Live Demo

https://omnidata.vision/demo/



Dataset Explorer

https://omnidata.vision/designer/

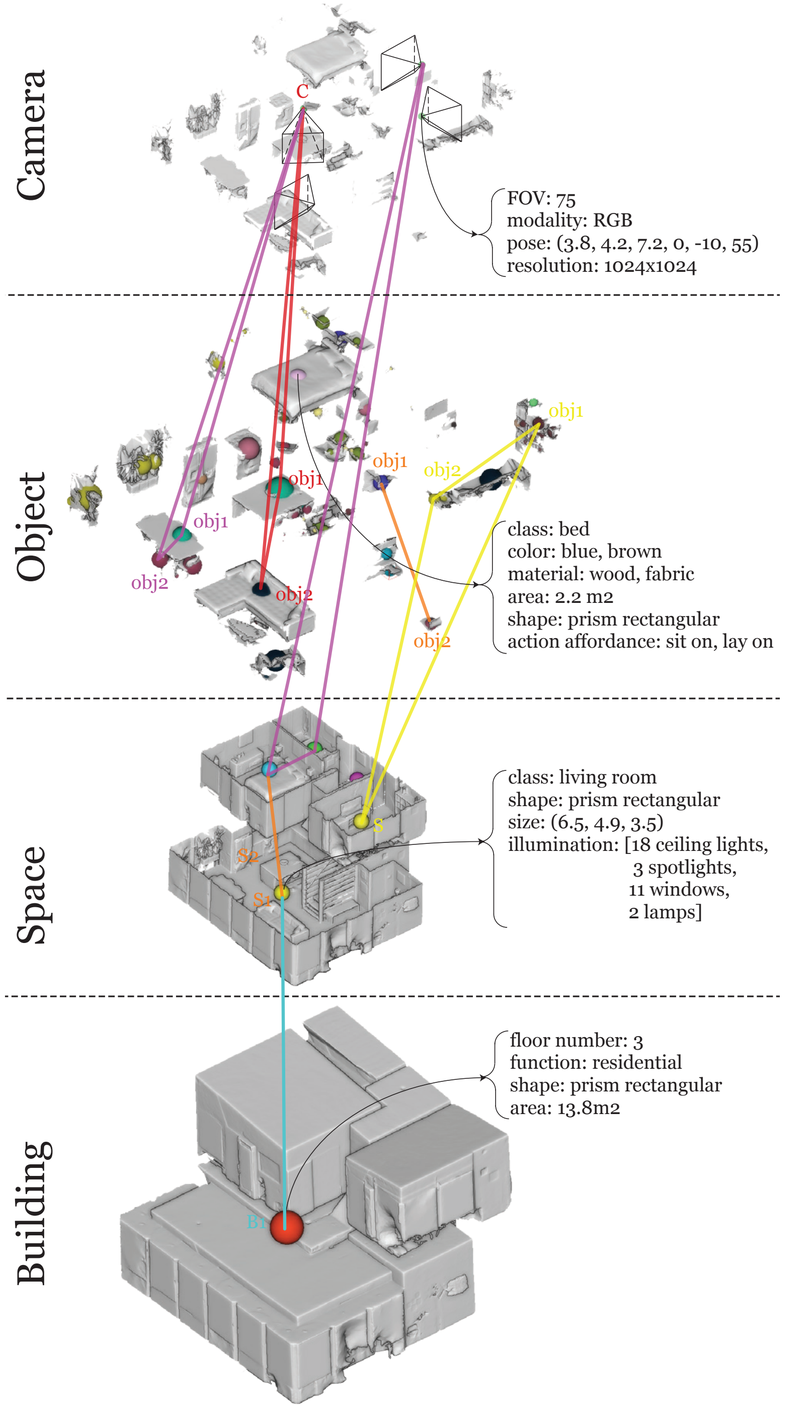


Omnidata: A Scalable Pipeline for Making Multi-Task Mid-level Vision Datasets from 3D Scans

Ainaz Eftekhar* Alexander Sax* Roman Bachmann Jitendra Malik Amir Zamir

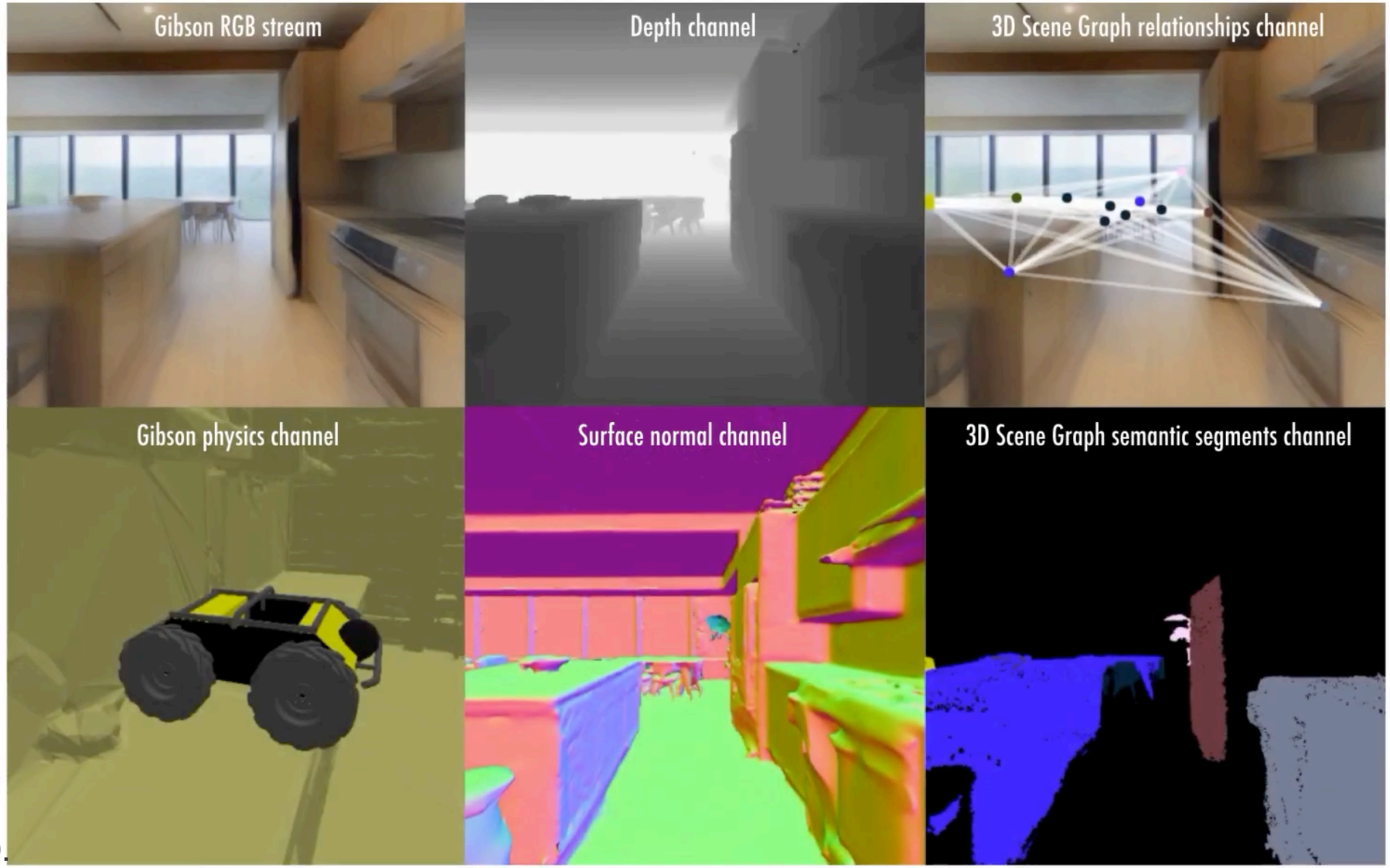
https://omnidata.vision

Graphical Scene Representations



- 3D Scene Graphs. 2020.

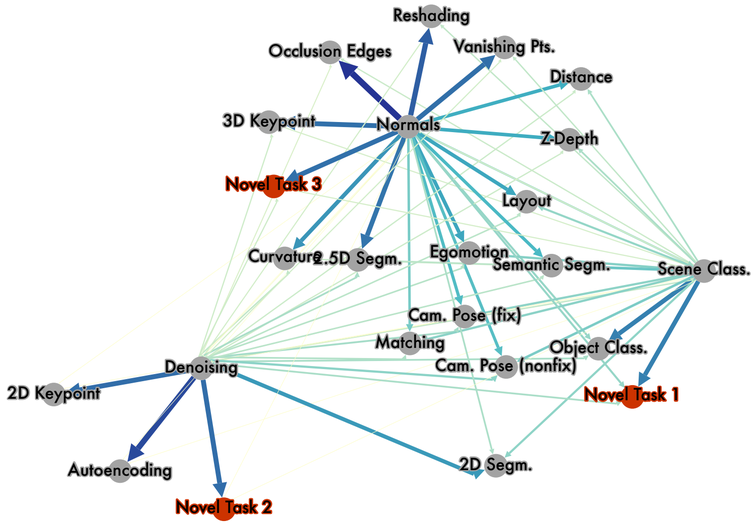
Graphical Scene Representations



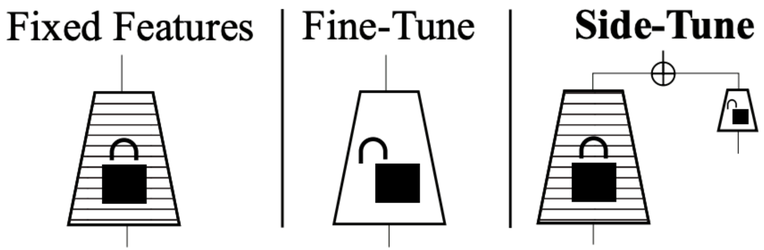
3D Scene Graphs. 2020.

Multi-Task Learning

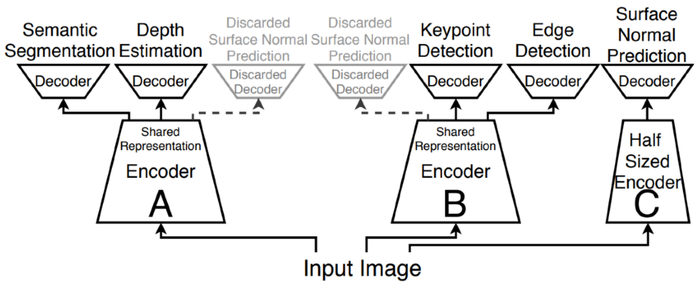
- Transfer Learning:
  - "Taskonomy: Disentangling Task Transfer Learning", CVPR18 Best Paper.



- Incremental Learning:
  - "Side-Tuning: Network Adaptation via Additive Side Networks", ECCV20.



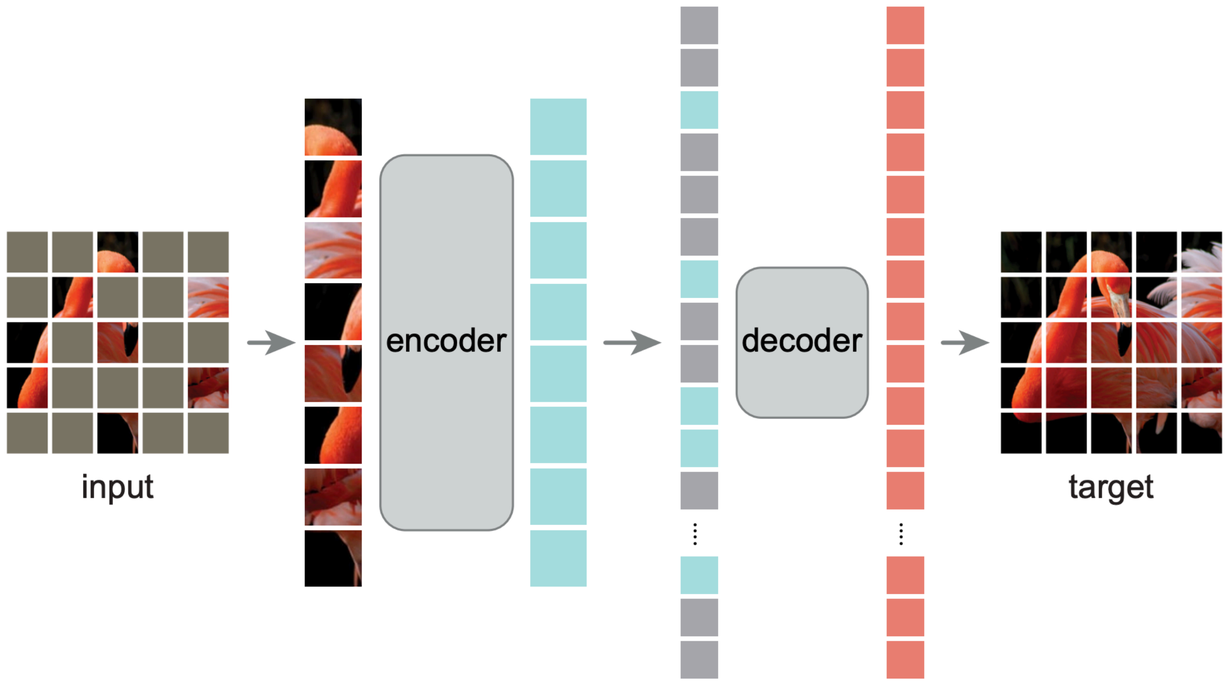
- Task Grouping:
  - "Which Tasks Should Be Learned Together in Multi-task Learning?", ICML20.

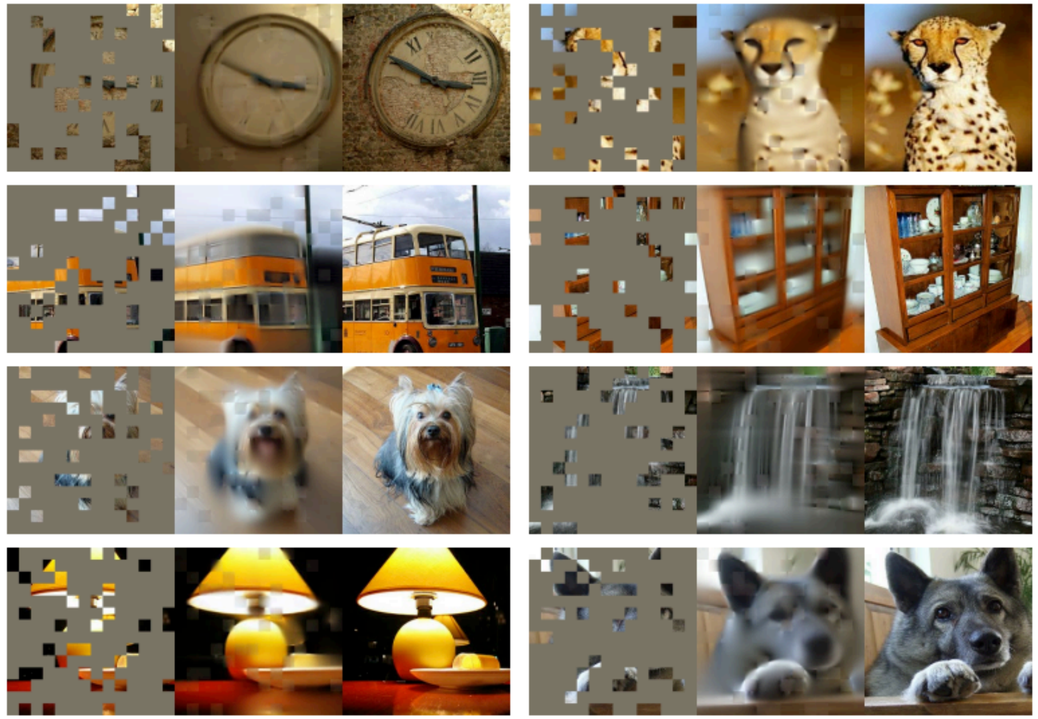


22

MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.

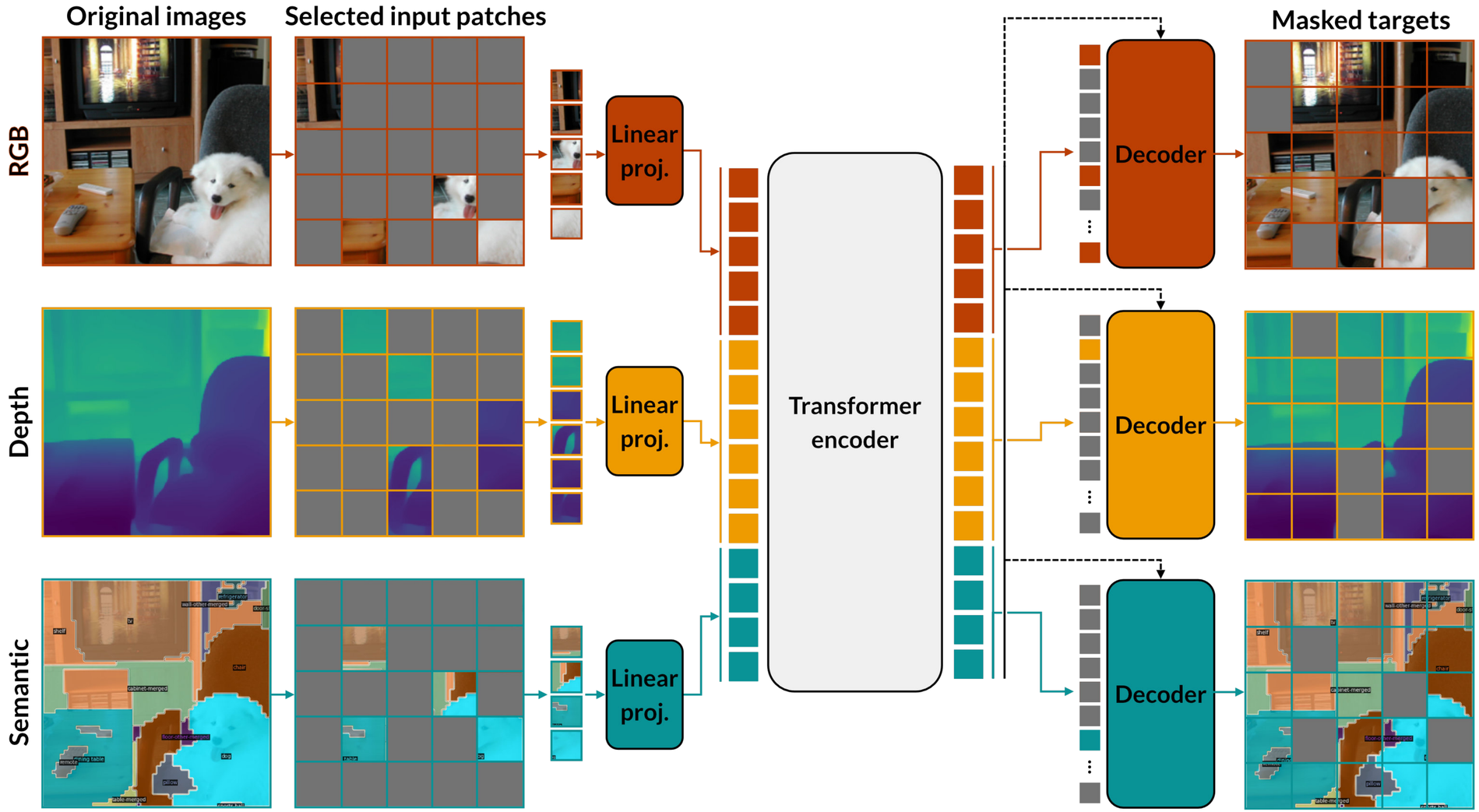
MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.





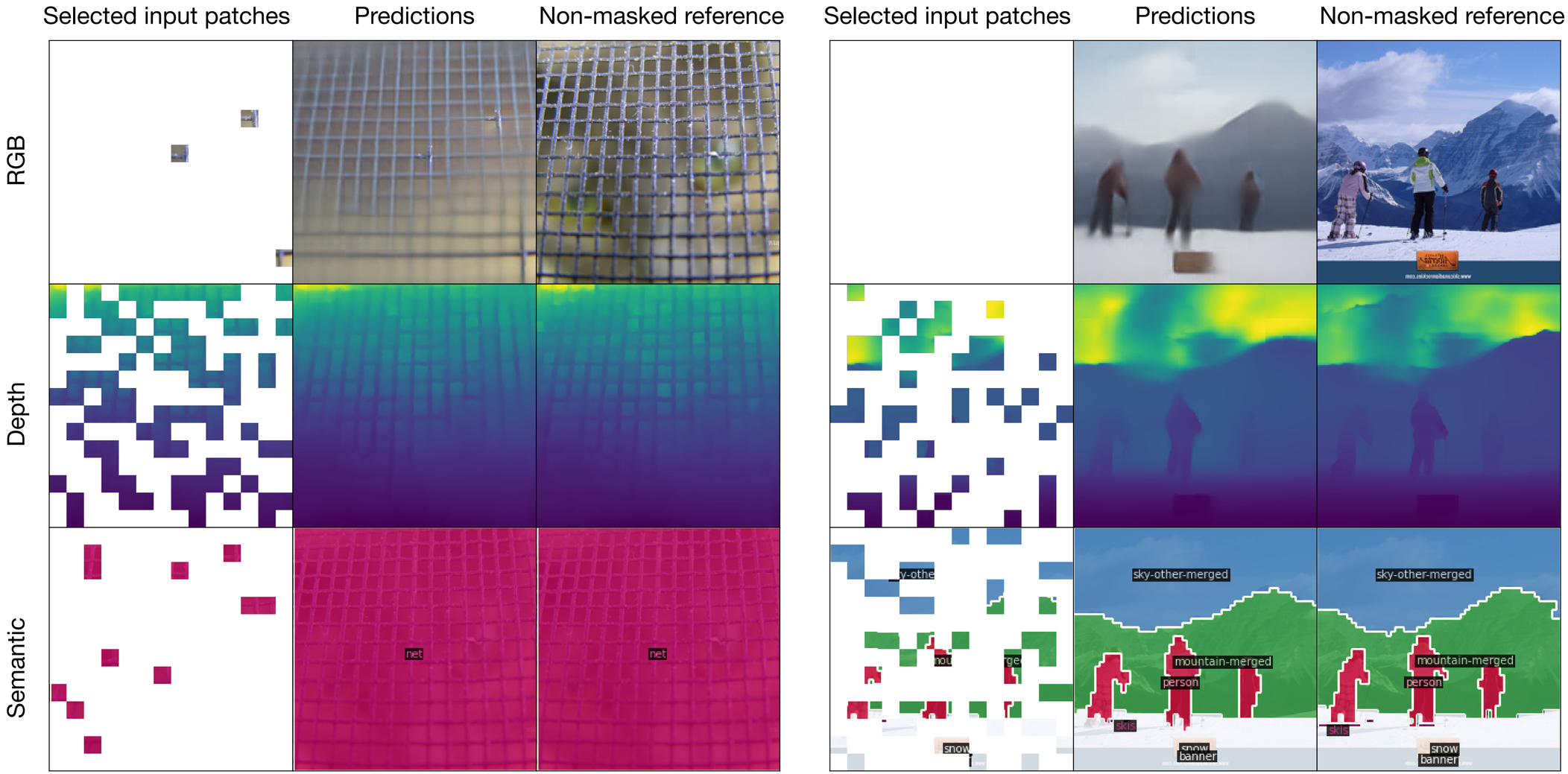
[Masked Autoencoders Are Scalable Vision Learners. He et al. 2021]

MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.



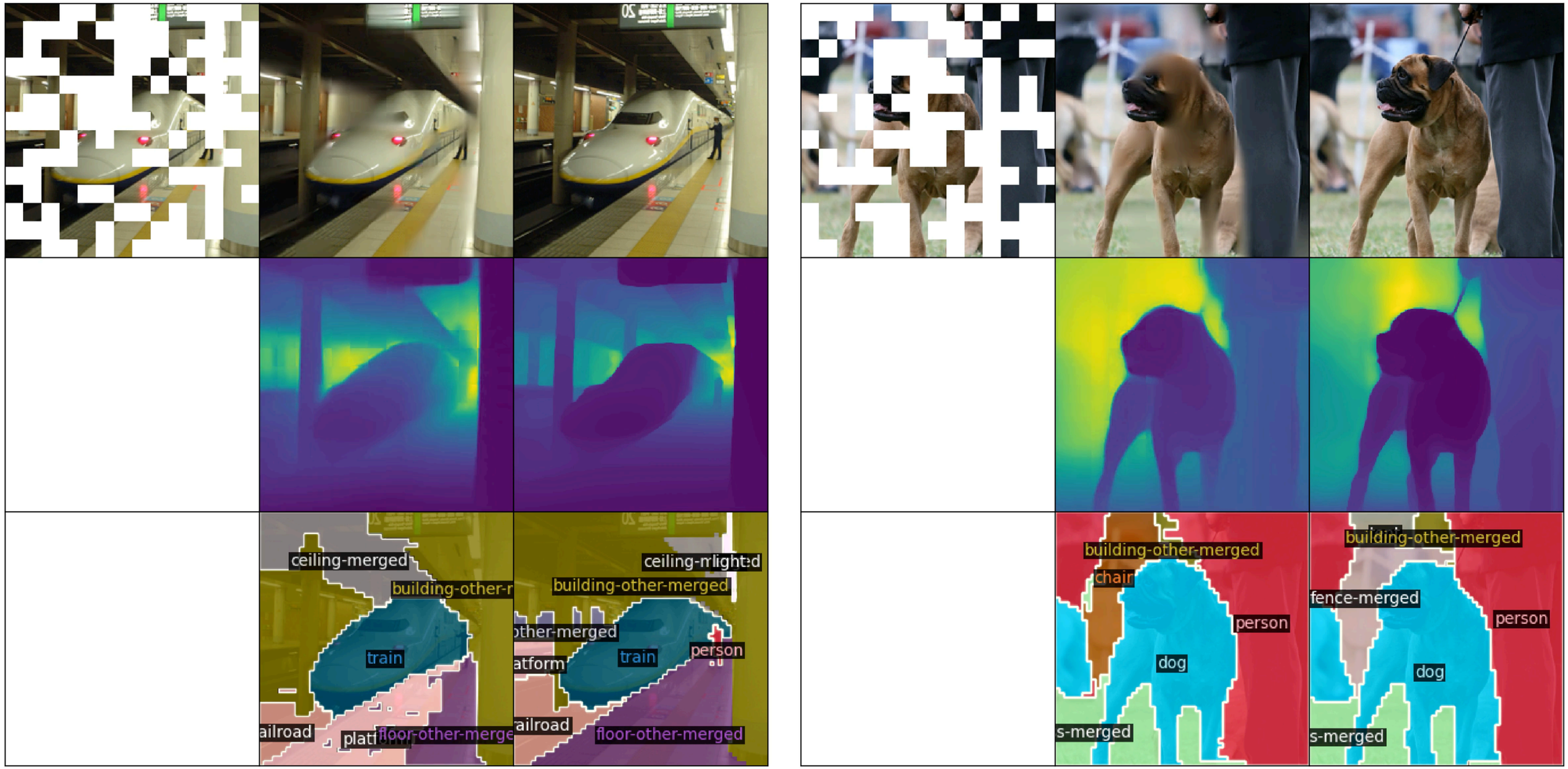
MultiMAE: Multi-Modal Multi-Task Masked Autoencoders, Bachmann, Mizrahi, Atanov, Zamir. ECCV 2022

MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.



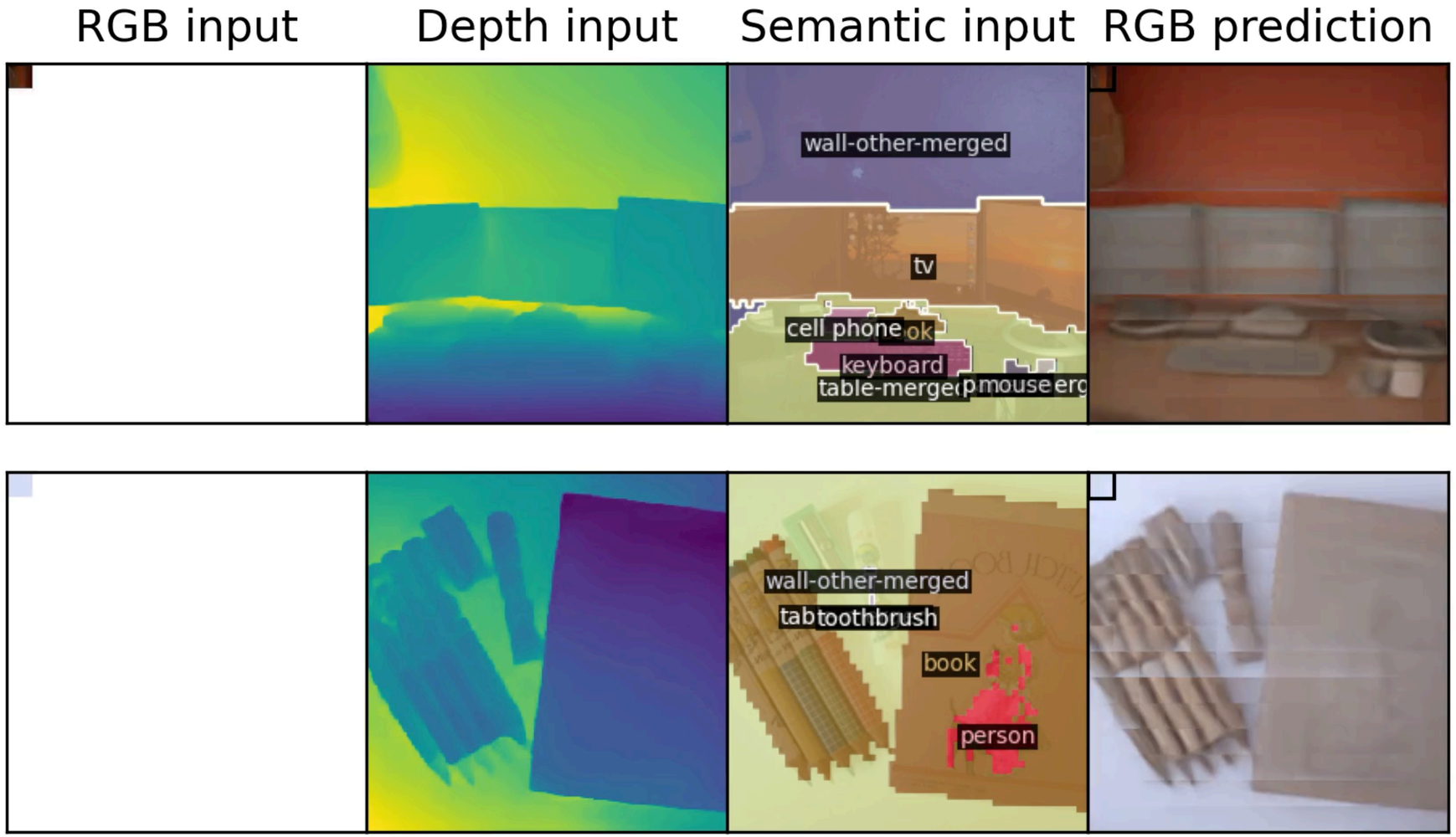
MultiMAE: Multi-Modal Multi-Task Masked Autoencoders, Bachmann, Mizrahi, Atanov, Zamir. ECCV 2022

MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.



MultiMAE: Multi-Modal Multi-Task Masked Autoencoders, Bachmann, Mizrahi, Atanov, Zamir. ECCV 2022

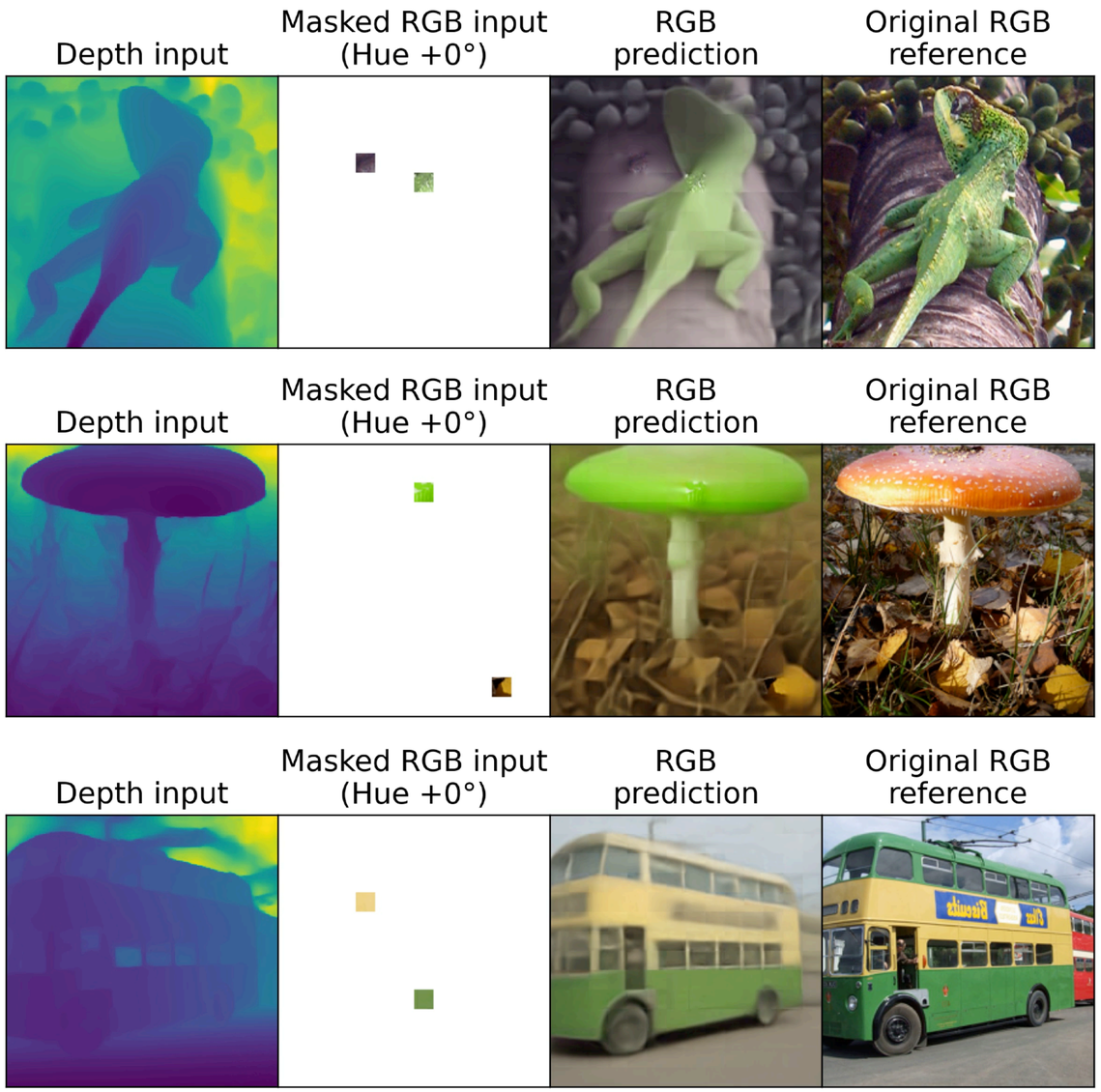
MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.



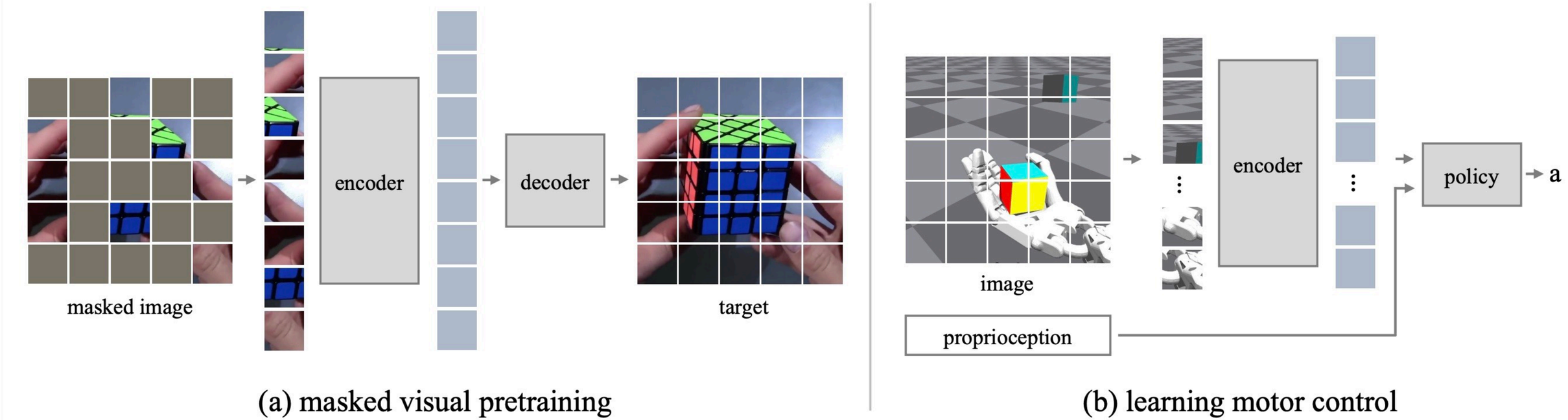
MultiMAE: Multi-Modal Multi-Task Masked Autoencoders, Bachmann, Mizrahi, Atanov, Zamir. ECCV 2022

MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.

Zamir



MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.

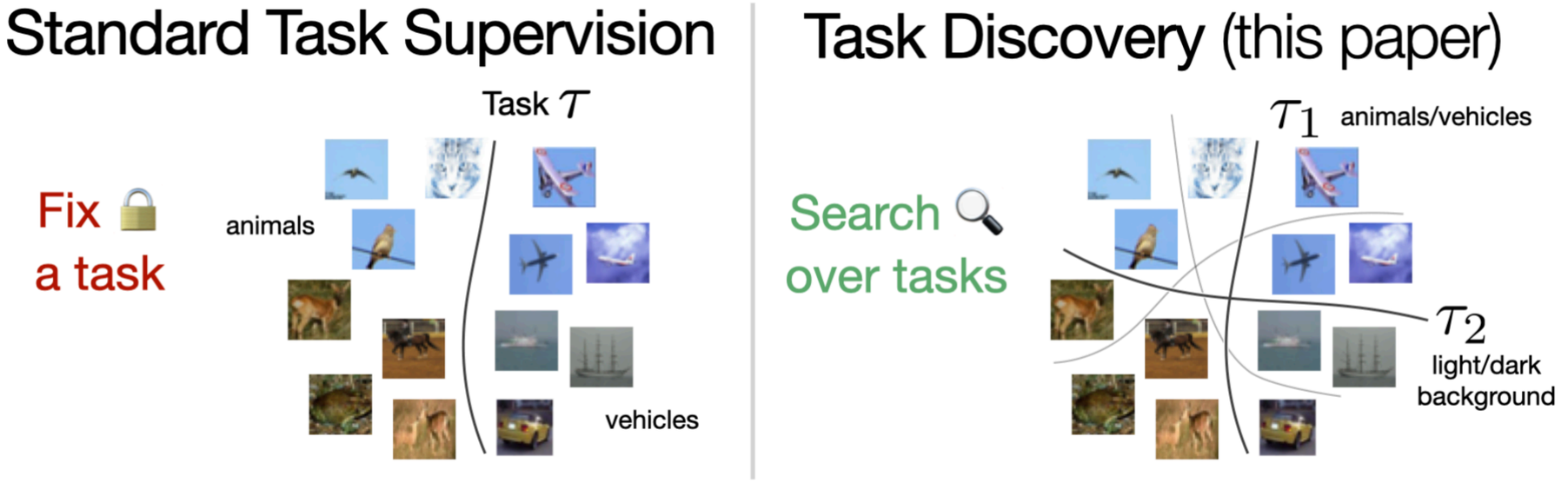


- Masked Visual Pre-training for Motor Control, Xiao, Radosavovic, Darrell, Malik. 2022
- MultiMAE: Multi-Modal Multi-Task Masked Autoencoders, Bachmann, Mizrahi, Atanov, Zamir. ECCV 2022

🔍 Task Discovery: Finding the Tasks that Neural Networks Generalize on

Andrei Atanov, Andrei Filatov, Teresa Yeo, Ajay Sohmshetty, Amir Zamir
Swiss Federal Institute of Technology in Lausanne (EPFL)

Paper Code



In contrast to a common supervised learning approach, in **task discovery**, we fix an architecture and search in the space of tasks to find ones the network generalizes on.

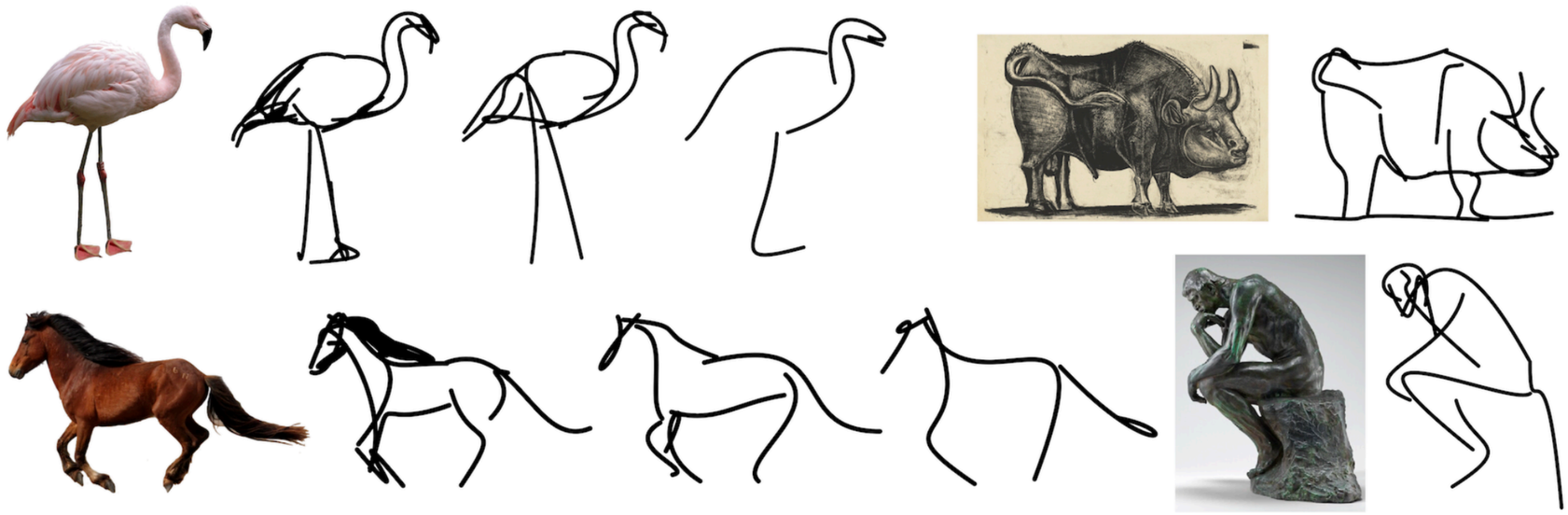
- https://taskdiscovery.epfl.ch/

CLIPasso: Semantically-Aware Object Sketching

Yael Vinker[1,2], Ehsan Pajouheshgar[1], Jessica Y. Bo[1], Roman Bachmann[1],
Amit Haim Bermano[2], Daniel Cohen-Or[2], Amir Zamir[1], Ariel Shamir[3]
[1]Swiss Federal Institute of Technology (EPFL), [2]Tel Aviv University, [3]Reichman University

SIGGRAPH 2022 (Best Paper Award)

arXiv Code Colab Demo



- https://taskdiscovery.epfl.ch/

Some Project Links

- Task Discovery: [https://taskdiscovery.epfl.ch/]
- MultiMAE: [https://multimae.epfl.ch/]
- CLIPasso: [https://clipasso.github.io/clipasso/]
- Omnidata [https://omnidata.vision/]
- 3D Common Corruptions [https://3dcommoncorruptions.epfl.ch/]
- Cross-Task Consistency [https://consistency.epfl.ch/]
- Taskonomy [http://taskonomy.vision/]
- Gibson Environment [http://gibson.vision/]
- Mid-Level Vision for Robotics [http://perceptual.actor/][sites.google.com/view/mid-level-representations/]
- Side-Tuning [http://sidetuning.berkeley.edu/]
- Which tasks should be learned together? [http://taskgrouping.stanford.edu/]
- 3D scene graph [http://3dscenegraph.stanford.edu/]

EDIC Seminar 2022

Amir Zamir, VILAB.

06.09.2022

École
polytechnique
fédérale
de Lausanne