Research Areas
Research Areas

• Computer Vision:

- Vanishing Points
- Reshading
- 3D Segmentation
- Autoencoding
- Scene Classification

Top 5 prediction: television_room, hotel_room, bedroom, living_room, home_theater

Top 5 prediction: home_theater, sliding_door, patio, terrace, studio_couch, day,
- Colorization
- Semantic Segmentation
Research Areas

• Computer Vision:

• Perception-for-Robotics:
Research Areas

• Computer Vision:

• Machine Learning:

• Perception-for-Robotics:

Cross-Task Consistent Learning

Baseline Learning
Vision $\leftrightarrow$ Action

Predator

Prey

M. Bank et al., 2015.
M. Land. 2002
Vision $\leftrightarrow$ Action

A

B

P_1

M. Bank et al., 2015.
M. Land. 2002
Vision $\leftrightarrow$ 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”.

---

**Animal Eye, M. Land**
Vision In-the-loop

Agent in the World

Sensory Observation

Perception

Control

Actions
Perceptual Robotic Simulators

Perceptual Robotics

Local planning ("go to the target")


Stair climb
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

- RGB Image
- Curvature
- Van. Points
- (Re)Shading
- Frags. (Flow)
- Tex. Edges
- Occ. Edges
- 2D Segm.
- 2.5D Segm.
- Panoptic Seg.
- Semant. Seg.
- Sf. Normals
- Point Match.
- Cam. Intrins.
- Cam. Extrin.
- 2D Keypts.
- 3D Keypts.
- Z-Depth
- Eucl. Depth
- Instance Seg.
- Classification
- Z-Depth
- Eucl. Depth
- Instance Seg.
- Classification
OmniData, ICCV2021.

- RGB
- Semantics
- Surface Normals
- 3D Keypoints
- Occlusion Edges
OmniData, ICCV2021.

Task: Depth Estimation

Ours

MiDaS

Ours

MiDaS

Ours

MiDaS

[Ranftl et al. TPAMI 2020]

[Ranftl et al. 2021]
OmniData, ICCV2021.

Task : Depth Estimation

RGB

Depth

Surface normal extracted from depth

Ours  MiDaS  Ours  MiDaS  Ours  MiDaS  Ours  MiDaS  Ours  MiDaS

[Ranftl et al. TPAMI 2020]
[Ranftl et al. 2021]
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.

Omnidata, ICCV2021.
3D Common Corruptions CVPR
**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

Camera
- FOV: 75
- modality: RGB
- pose: (3.8, 4.2, 7.2, 0, -10, 55)
- resolution: 1024x1024

Object
- class: bed
- color: blue, brown
- material: wood, fabric
- area: 2.2 m²
- shape: prism rectangular
- action affordance: sit on, lay on

Space
- class: living room
- shape: prism rectangular
- size: (6.5, 4.9, 3.5)
- illumination: [18 ceiling lights, 3 spotlights, 11 windows, 2 lamps]
- floor number: 3
- function: residential
- shape: prism rectangular
- area: 13.8m²

Building
- 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:

- Task Grouping:
  - "Which Tasks Should Be Learned Together in Multi-task Learning?", ICML20.
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, Bachmann, Mizrahi, Atanov, Zamir. ECCV 2022
MultiMAE: Multi-Modal Multi-Task Masked Autoencoders, Bachmann, Mizrahi, Atanov, Zamir. ECCV 2022
MultiMAE: Multi-Modal Multi-Task Masked Autoencoders, Bachmann, Mizrahi, Atanov, Zamir. ECCV 2022
**MultiMAE**

Multi-Modal Multi-Task Masked Autoencoders.

<table>
<thead>
<tr>
<th>Depth input</th>
<th>Masked RGB input (Hue +0°)</th>
<th>RGB prediction</th>
<th>Original RGB reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Depth input" /></td>
<td><img src="image2.png" alt="Masked RGB input" /></td>
<td><img src="image3.png" alt="RGB prediction" /></td>
<td><img src="image4.png" alt="Original RGB reference" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Depth input" /></td>
<td><img src="image6.png" alt="Masked RGB input" /></td>
<td><img src="image7.png" alt="RGB prediction" /></td>
<td><img src="image8.png" alt="Original RGB reference" /></td>
</tr>
<tr>
<td><img src="image9.png" alt="Depth input" /></td>
<td><img src="image10.png" alt="Masked RGB input" /></td>
<td><img src="image11.png" alt="RGB prediction" /></td>
<td><img src="image12.png" alt="Original RGB reference" /></td>
</tr>
<tr>
<td><img src="image13.png" alt="Depth input" /></td>
<td><img src="image14.png" alt="Masked RGB input" /></td>
<td><img src="image15.png" alt="RGB prediction" /></td>
<td><img src="image16.png" alt="Original RGB reference" /></td>
</tr>
</tbody>
</table>
MultiMAE
Multi-Modal Multi-Task Masked Autoencoders.

(a) masked visual pretraining
(b) learning motor control
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)

Standard Task Supervision

Fix a task

Task Discovery (this paper)

Search over tasks

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\textsuperscript{1,2}, Ehsan Pajouheshgar\textsuperscript{1}, Jessica Y. Bo\textsuperscript{1}, Roman Bachmann\textsuperscript{1}, Amit Haim Bermano\textsuperscript{2}, Daniel Cohen-Or\textsuperscript{2}, Amir Zamir\textsuperscript{1}, Ariel Shamir\textsuperscript{3}

\textsuperscript{1}Swiss Federal Institute of Technology (EPFL), \textsuperscript{2}Tel Aviv University, \textsuperscript{3}Reichman University

SIGGRAPH 2022 (Best Paper Award)

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/]