SAILENV

Learning in Virtual Visual Environments Made Simple

WHY VIRTUAL ENVIRONMENTS?

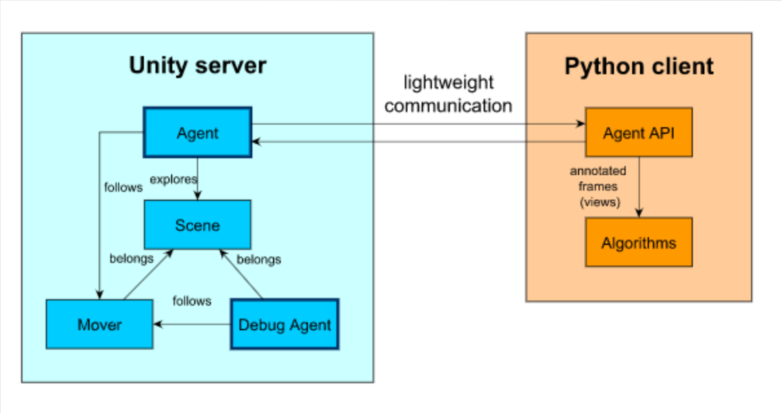
- Simulation of real-world settings with 3D graphics engine
- Perform experiments too costly in real-world settings
- Automatic and precise annotation
  - Bounding boxes, semantic segmentation, motion information, etc…
  - Little to no need of human intervention for data collection
- High degree of control on experimental settings
  - Lighting and weather conditions, image resolution, etc…

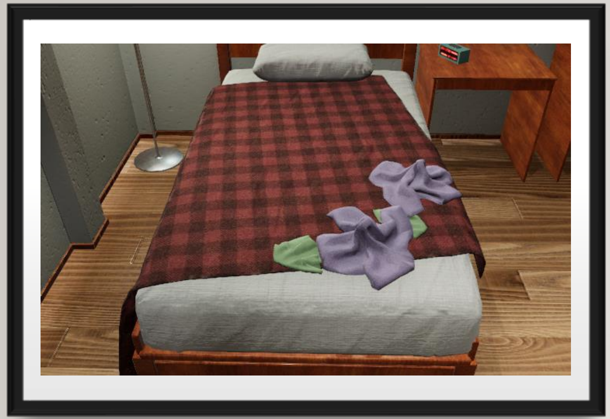
EXISTING VIRTUAL ENVIRONMENTS

Platform	Photoreal	Depth	OptFlow	LightNet	OS
DeepMindLab		✓		n.a.	Unix
Habitat	✓	✓		n.a.	Unix
AI2-THOR	✓	✓			Unix
SAILenv	✓	✓	✓	✓	Win+Unix

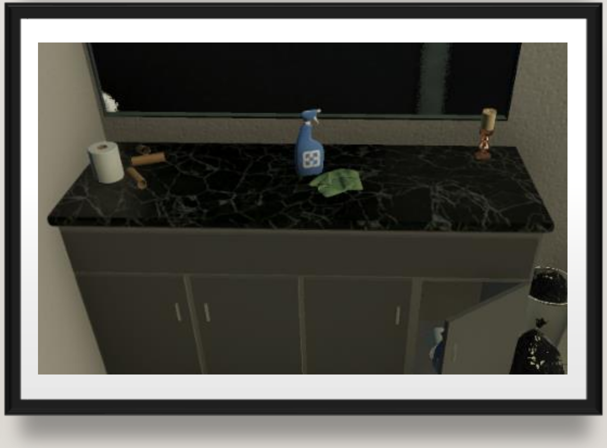
SAILENV ARCHITECTURE

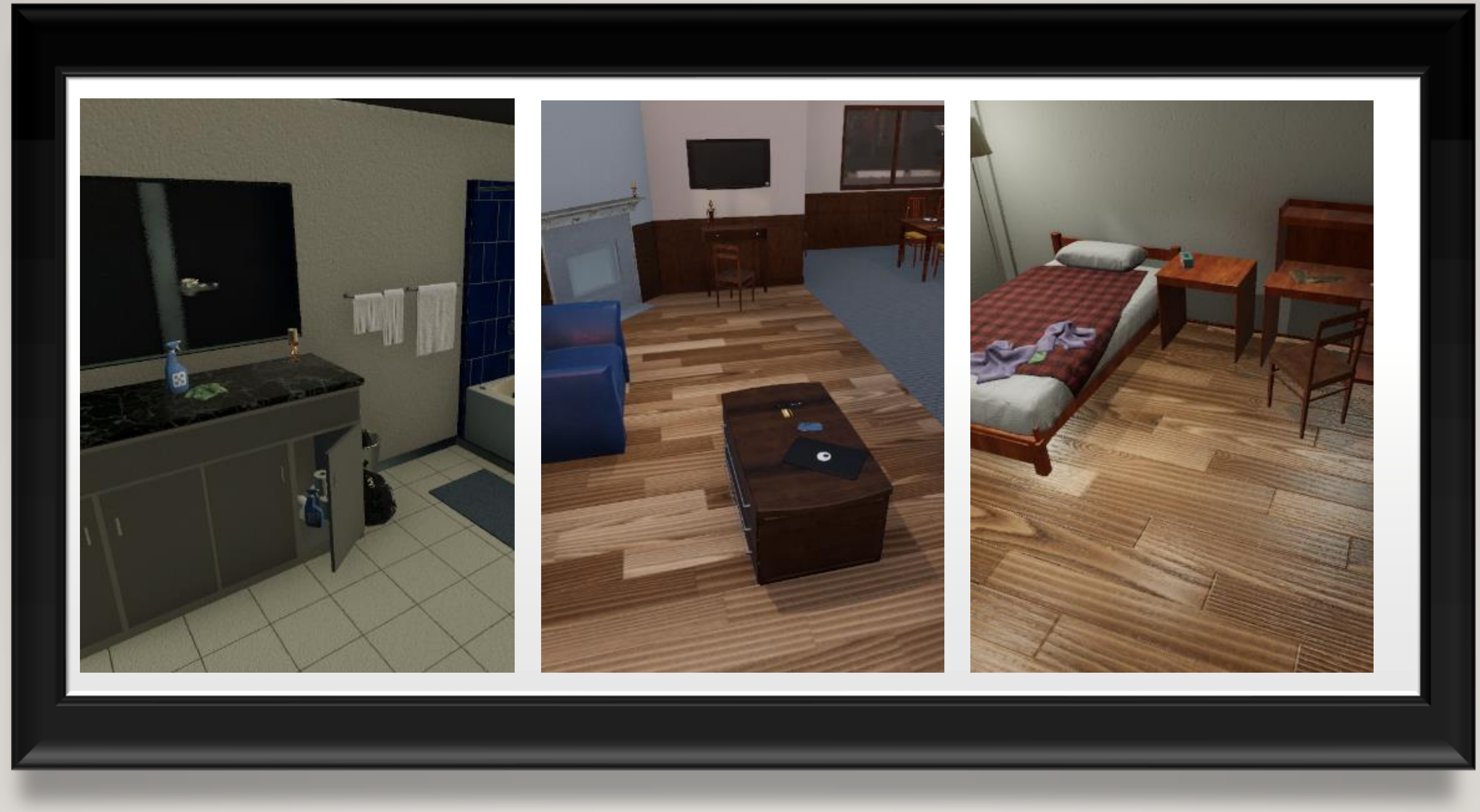
- Client-server architecture
  - Virtual Environment: server
  - Agent API: client
- Unity Server
  - Physics Simulation
  - Real-Time rendering
  - Data generation and annotation
  - Lightweight Network Protocol
- Python Client
  - Lightweight, cross-platform API
  - High-level commands for the Server
  - Exposes views to common ML Frameworks





OBJECT LIBRARY





READY-TO-USE DOMESTIC SCENES

MOVING AGENT IN THE SCENE

- Agent has three ways of moving in the scene

1. Python commands to define custom moving criteria
   - Simple functions for changing position and orientation
2. Following a track included in the scene
   - Track is created by the scene designer
   - Can be changed through the Unity Editor
   - Cannot be changed at runtime
3. Through keyboard and mouse in FPS-like fashion

MOVING OBJECTS IN THE SCENE

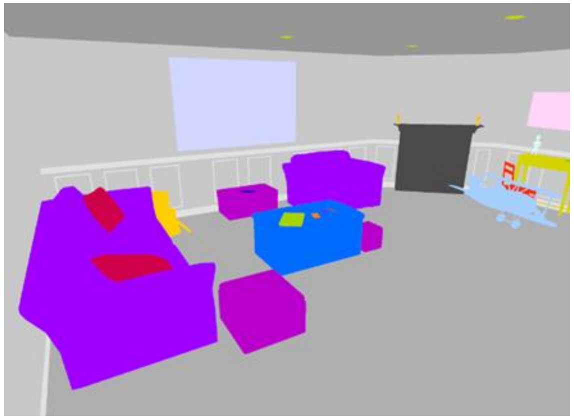
- Movements are simulated through Unity Physics Engine
- The movement behavior is scripted with C#

ENVIRONMENT VIEWS

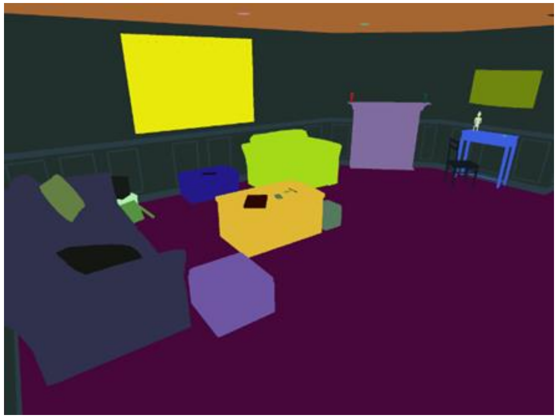
- SAILenv generates views of the environment in real-time
- Every view is taken from the Agent POV
- Each view yields pixel-wise information on the environment
  - *Main:* HxWx3 – RGB view in OpenCV format
  - *Category:* HxWx1 – category ID of the object
  - *Object:* HxWx3 – unique object ID
  - *Flow:* HxWx2 – optical flow of the pixel w.r.t. the Agent
  - *Depth:* HxWx1 – depth of the pixel w.r.t. the Agent

CATEGORY AND INSTANCE SEGMENTATION

- Categories can be quickly customized
  - Through Unity Editor
- Object ID is automatically generated
  - Guaranteed to be unique



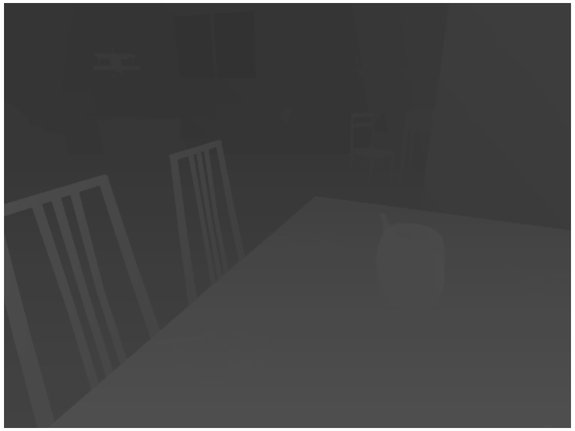
Category View



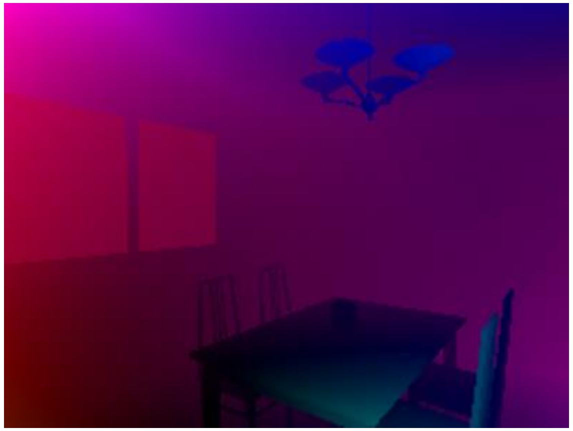
Instance View

DEPTH AND OPTICAL FLOW

- Depth intensity is proportional to vicinity w.r.t. the Agent position
- Optical Flow is the velocity in px per frame of the pixel

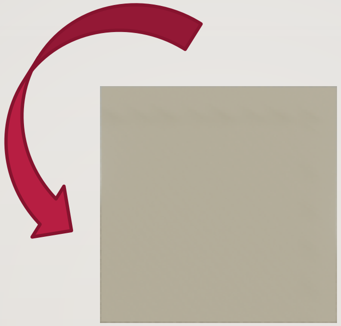


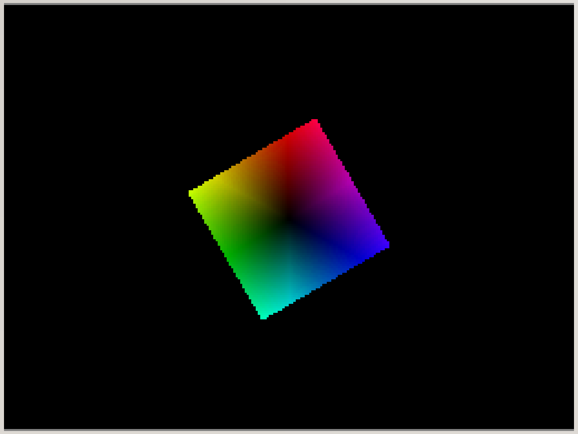
Depth View



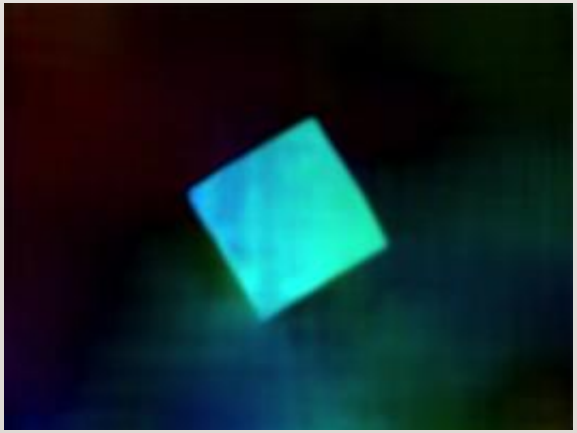
Optical Flow View

OPTICAL FLOW COMPARISON

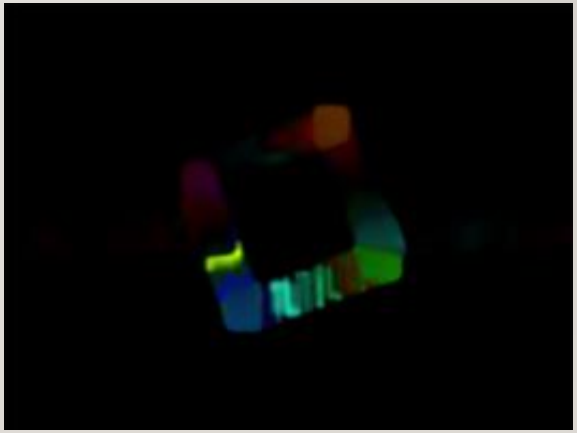




SAILenv



LiteFlowNet

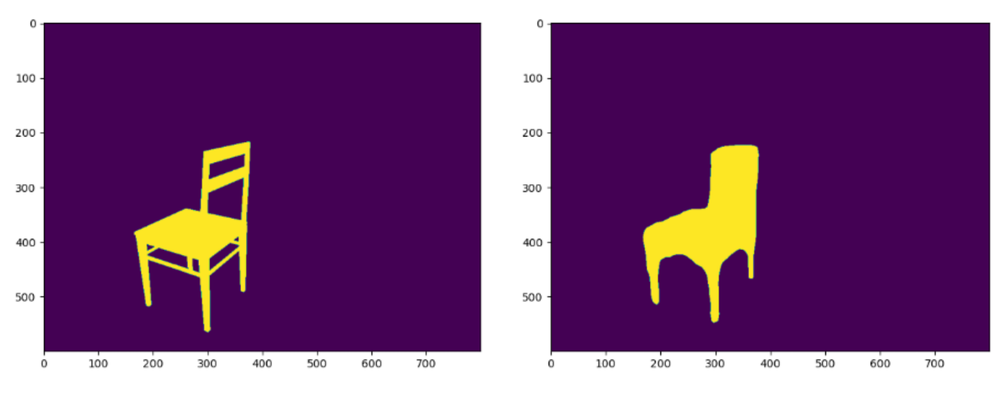


OpenCV

PHOTOREALISM EVALUATION

- Can a state-of-the-art object detector recognize objects in SAILenv?
- We tested with Mask R-CNN trained on COCO-train2017
- We focused on categories from the COCO dataset
- We measured the IoU between predictions and ground truth from SAILenv
- Mask R-CNN robustly detects a large portion of objects
- Some problems arise from occlusions and labeling criteria

DETECTION ERRORS



Ground Truth

Prediction

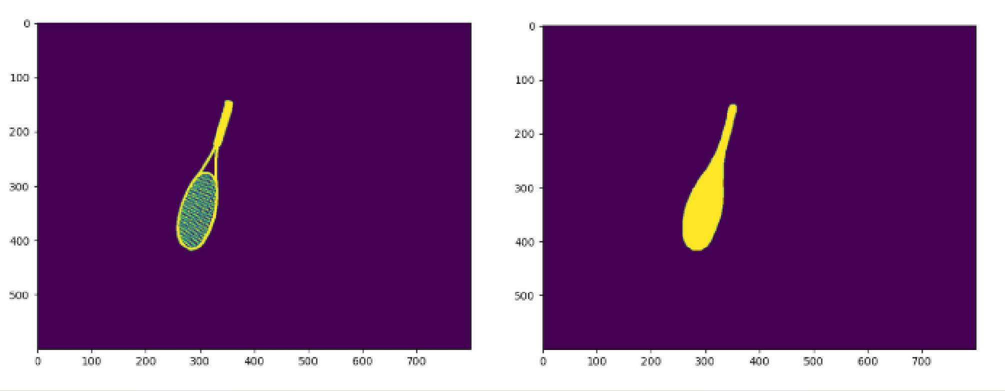
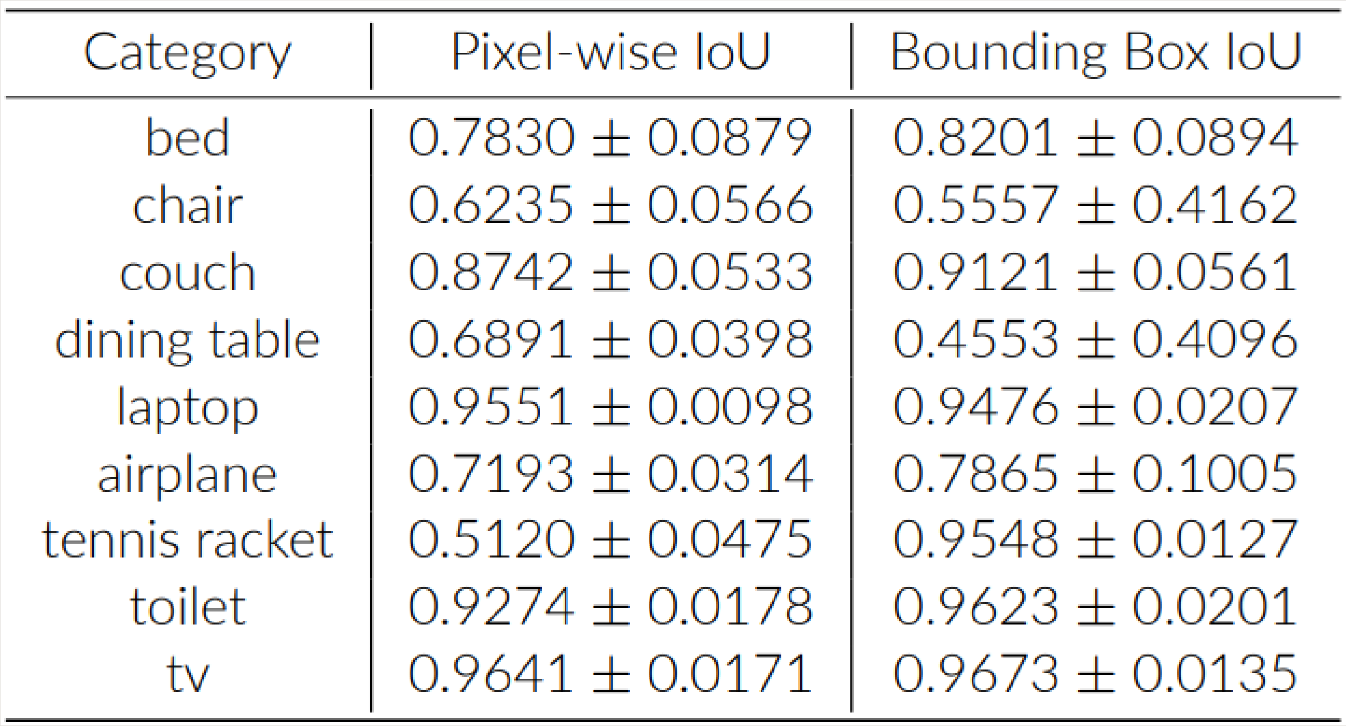


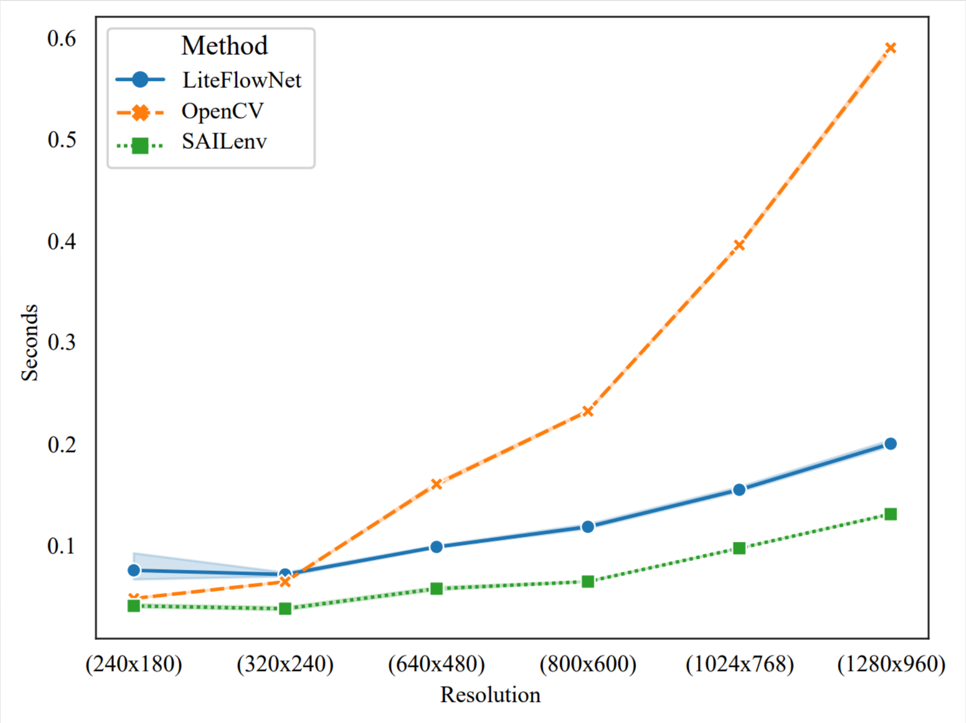
PHOTO-REALISM EVALUATION WITH MASK R-CNN (COCO)



Category	Pixel-wise IoU	Bounding Box IoU
bed	0.7830 ± 0.0879	0.8201 ± 0.0894
chair	0.6235 ± 0.0566	0.5557 ± 0.4162
couch	0.8742 ± 0.0533	0.9121 ± 0.0561
dining table	0.6891 ± 0.0398	0.4553 ± 0.4096
laptop	0.9551 ± 0.0098	0.9476 ± 0.0207
airplane	0.7193 ± 0.0314	0.7865 ± 0.1005
tennis racket	0.5120 ± 0.0475	0.9548 ± 0.0127
toilet	0.9274 ± 0.0178	0.9623 ± 0.0201
tv	0.9641 ± 0.0171	0.9673 ± 0.0135

OPTICAL FLOW EVALUATION

- As seen before, motion estimation is highly accurate
- What is the computational burden of motion estimation?
- We compared with OpenCV and FlowNetLite



CONCLUSIONS

- We presented SAILenv, a platform based on Unity Engine
- Platform which makes it easy to create, run and get data from realistic 3D Virtual Environments
- Vision-related algorithms can be efficiently evaluated
- To the best of our knowledge, SAILenv is the first platform which yields motion information
- We believe it is a good entry point for researchers interested in 3D Virtual Environments

TEAM AND LINKS

- Team members:
  - Enrico Meloni
  - Luca Pasqualini
  - Matteo Tiezzi
  - Stefano Melacci
  - Marco Gori
- Official project page: http://sailab.diism.unisi.it/sailenv/
- arXiv pre-print: https://arxiv.org/abs/2007.08224
- GitHub: https://github.com/sailab-code/sailenv

THANK YOU FOR LISTENING