SAILenv: Learning in Virtual Visual Environments Made Simple

Enrico Meloni [1,2] Luca Pasqualini [2] Matteo Tiezzi [2] Marco Gori[2,3] Stefano Melacci[2]

[1]DINFO, University of Florence, Florence, Italy [2]Department of Information Engineering and Mathematics, University of Siena, Siena, Italy [3]Universitè Côte d'Azur, Nice, France

SAILab Siena Artificial Intelligence Lab

ICPR 2020

Introduction

- ▷ Training and test data are usually collected from the operational real environment and then used for learning neural models.
- ▷ 3D Simulations can be used to generate artificial visual settings substituting an expensive data collection from real-world.
- ▷ Generated environments must be highly photorealistic to reduce the effort needed to compensate the gap between real and virtual.
- ▷ Many of the available platforms handle navigation-related or ad-hoc tasks not meant for easy customization.
- ▷ Not all Machine Learning researchers have robust skills in creating 3D scenes, discouraging the use of complex 3D environments.

Comparison

Platform	Photoreal	Depth	OptFlow	LightNet	OS
DeepMind Lab [1]		✓		n.a.	Unix
Habitat [3]	✓	✓		n.a.	Unix
AI2-THOR [2]	✓	✓			Unix
SAILenv	✓	✓	✓	✓	Win+Unix

Comparison of the main features of SAILenv with other popular platforms. *LightNet* refers to lightweight communication over the network (n.a. means network communication is not directly provided).

Contributions

- ▷ SAILenv, a novel Unity-based platform designed for ease of use and customization of visual recognition tasks in 3D scenes.
- ▷ SAILenv yields pixel-wise semantic and instance segmentation, depth and optical flow.
- ▷ SAILenv, to the best of our knowledge, is the first platform to yield real-time motion information.
- ▷ SAILenv comes with an extensive library of ready-to-use objects and 3D scenes that can be easily customized and extended.

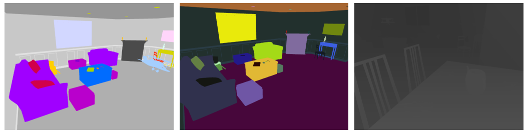
Resources

Project Page: `https://sailab.diism.unisi.it/sailenv/`

Photo-realistic Objects and Scenes

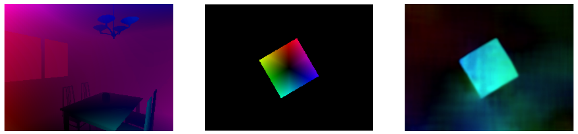
Objects are highly detailed with photorealistic PBR textures and predefined Scenes give a noticeable head start in learning the tool.

Pixel-Wise Annotations



SAILenv yields real-time pixel-wise semantic and instance segmentation as well as the depth of pixels in the scene.

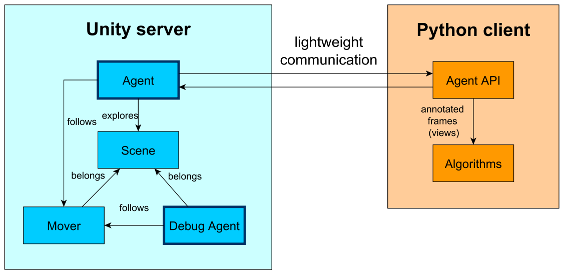
Real-Time Optical Flow



SAILenv LiteFlowNet

SAILenv yields real-time and precise motion information. By comparison, while SAILenv correctly manages to describe the rotation of the cube, LiteFlowNet can only partially infer it from the images.

Interfacing with SAILenv



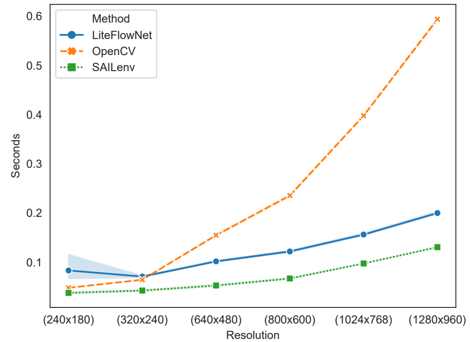
SAILenv can be interfaced with most Machine Learning frameworks through a few simple lines of code.

Photo-realism Evaluation

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

We tested the performance of a known neural model, Mask R-CNN trained on MS COCO Dataset, to evaluate the degree of photo-realism of objects and scenes included in SAILenv. The differences between pixel-wise and bounding box IoU is due to the fact that SAILenv yields more fine-grained annotations w.r.t. to MS COCO.

Optical Flow Performance



We compare the computational overhead of motion estimation with the popular OpenCV implementation of Farneback's algorithm, and the neural-based LiteFLowNet, proving greater scalability to higher resolutions.

References & Acknowledgements

This work was partly supported by the PRIN 2017 project RexLearn, funded by the Italian Ministry of Education, University and Research (grant no. 2017TWNMH2)

[1] Charles Beattie, Joel Z Leibo, Denis Teplyashin, Tom Ward, Marcus Wainwright, Heinrich Küttler, Andrew Lefrancq, Simon Green, Víctor Valdés, Amir Sadik, et al. Deepmind lab. *arXiv preprint arXiv:1612.03801*, 2016.

[2] Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Daniel Gordon, Yuke Zhu, Abhinav Gupta, and Ali Farhadi. Ai2-thor: An interactive 3d environment for visual ai. *arXiv*, 2017.

[3] Manolis Savva*, Abhishek Kadian*, Oleksandr Maksymets*, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A Platform for Embodied AI Research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.