

OSB3D: An Open-Source Benchmark for Automated Testing and Validation of Complex 3D Game Environments

Introduction

Digital gaming has evolved from a fringe pastime to a mainstream, multi-billion dollar industry, reaching 2.7B players in 2020 with an annual revenue of \$159B (\approx SEK1.6T). Moreover, the gaming industry in Sweden is thriving and generating a global revenue of SEK 58B (2021) with double-digit growth annually. Modern video games have grown in size and complexity and with that the need to test them. However, thorough coverage is often not feasible using manual human play testing. Automated Testing Agents (ATA) systems have emerged as a remedy to this challenge but still fall short of human performance in finding bugs. ATA is far from solved, and as with any emerging technology, it is challenging to find environments that resemble modern games to be able to do research and test the algorithms. The standard benchmark environments that algorithms are evaluated on are often too simplistic and do not sufficiently resemble modern games. For this reason, researchers who want to do research in this field often need to construct their own environment for testing, or rely on benchmarks that are less relevant to modern games that are far more diverse and complex. Access to a more realistic environment will also be usable for other challenges, including complex 3D navigation, motion planning, predictions in dynamic environments, predictive models for finding bugs and issues, sample efficiency exploration, and other opportunities for testing and benchmarking AI. This would be valuable also to non-gaming related industries, including automotive, robotics, and transportation.

Purpose and expected outcome

The proposed project aims to create a public suite of game-testing environments. Researchers and game-industry practitioners can use the suite to test and evaluate ATA and related systems. The benchmark and code will be completely open-source, and it will be available to anyone who wants to modify it, providing a public sandbox that researchers can use based on their needs. The environments will resemble a subset of an actual video game in which different types of known bugs and common issues will be included. The goal of the project is to deliver lightweight and simplified versions of real game scenes to advance ATA and related systems. To ensure the project's feasibility, we will initially focus on creating one scalable and fast-to-simulate game scene. If time permits, we will add additional scenes to the suite. We will also provide a set of baseline Machine Learning (ML) based algorithms and results for researchers to use as a comparison for their new methods. By making the environment and baselines code open-access, we encourage a broader range of researchers and practitioners to contribute to the advancement of ATA and related systems by creating new game scenes, based on the one we will deliver for their needs. The intention is to organize a challenge at NeurIPS. While standard environments for academic research on reinforcement learning agents are helpful in testing basic concepts, they are too simplistic and lack real-world complexity, making our suite potentially valuable for a wider range of applications beyond game testing.

Connection to ongoing WASP projects

This project is partially inspired by an ongoing WASP project at WARA Media and Language, namely the GENE Challenge 2022, which has contributed to the benchmarking culture in the gesture-generation field by standardizing the evaluation practices. Intelligent agents operating in a complex 3D environment strongly relate this project to WARA Public Safety, where autonomous robotic systems work together with humans to find and save for example missing people at sea. In field robotics, a major focus of WARA Public Safety, it is important to develop and verify functionality and behaviors in complex simulated environments before and during the deployment of autonomous robotic systems. Sample efficient exploration and bug finding is crucial in this domain as well.

Participating researchers and industry partners

- **ReaL** – Linköping University: The Reasoning and Learning (ReaL) lab conducts AI research on the integration of novel reasoning and learning methods for large-scale spatiotemporal applications. This includes 3D exploration and dynamic motion planning for autonomous systems, time-series generative adversarial networks (GANs), and multi-agent reinforcement learning. ReaL has extensive experience building and using simulation-supported AI development infrastructure, including the ELLIIT-funded AIOps platform. Participants: Fredrik Heintz (Professor), Mattias Tiger (PostDoc), and Daniel de Leng (Research Engineer).
- **SEED** – Electronic Arts: An advanced R&D group that has worked on game testing for many years and works as an intermediary between academia and production by publishing in the field, and with close collaboration with game studios. They are currently a part of the WASP/WARA Media and Language. They also represent EA, one of the world's biggest game publishers. Participants: Alessandro Sestini (Research Engineer, EA), Linus Gisslén (Technical Director, EA), Konrad Tollmar (Research Director, EA, Associate Professor, KTH).
- **King** – AI Center of Excellence at King has worked on game playtesting for a few years and published a few articles in this domain at conferences, including IEEE CoG and ICMLA. King has a WASP-affiliated industrial Ph.D. student working on the topic of applying generalization and scalability of reinforcement learning for game playtesting. Participants: Sahar Asadi (Director of AI Labs at King), Sara Karimi (Industrial PhD Student at KTH and King, AI/ML Engineer at King), Björn Brinne (Sr Director of AI Center of Excellence at King).

Resources available and cost coverage

The main cost will be hiring a research engineer at LiU to build the environment and do the qualitative and quantitative studies. A cost of a maximum of 100KSEK will also be allocated to infrastructure such as computers, software, etc, of the total sum of 1 MSEK.

SEED – EA will provide expertise in game testing as an industry partner; they will implement game testing ML base-lines, they also will provide support on the specification of the game environment. Furthermore, they will write specifications and partner with the hired

research engineers to implement the environment as they have built several in-house testing environments previously, e.g. see video for an example:

▶ **Imitation Learning to Inform the Design of Computer Games** . The game environments for this project will be developed using Unity 3D game engine, which are well-suited for creating interactive environments due to their robust capabilities and user-friendly design. The SEED team has previous experience using these tools, which have consistently delivered satisfying results. King will provide expertise in game playtesting as an industry partner; they will also provide expertise in the area of reinforcement learning for playtesting, evaluation frameworks and metrics. King will contribute in writing the specifications and will provide input on the design of the environment, architecture, and evaluation frameworks.

Goals and objectives

The short- and long-term main objectives include: developing and publishing an open-source suite composed of one or more game-testing scenes; developing and publishing environments to be used for academic research; publishing a benchmark paper that presents the environments, the taxonomy of bugs, and baselines; organizing and hosting a challenge at a gaming-related or machine learning research conference. In the long term, we aspire that the developed environment will become a part of the main benchmark suite used by both academia and industry in the evaluation of ATAs, and also to establish itself as a complement to the existing benchmark environments in reinforcement learning.

Timeline with milestones

Below is a preliminary and comprehensive timeline. We identified four milestones: (M1) Design a prototype environment with one game/scene, (M2) Implement a baseline in the environment, (M3) Quantitative and Qualitative analysis of the designed environment, and (M4) Publishing a scientific paper.

Phase 1:

- 1st month: Literature and current technologies studies, deliverable(s): document describing scientific landscape.
- 2nd month: Environment design, deliverable(s): a prototype environment to be evaluated.
- 5th months: Environment with at least 1 game/scene. *1st milestone.*

Phase 2:

- 7th months: Implementation and testing of baseline. *2nd milestone.*

Phase 3:

- 8th month: Quantitative analysis design and study.
- 9th month: Qualitative (with professional game developers) design study. *3rd milestone.*

Phase 4:

- 10th month: Improve the environment based on findings from the quantitative and qualitative studies (see above). If time permits, increase the number of environments.
- 11th month: writing a position paper about the new benchmark and eventually about designing/hosting a challenge. *4th milestone.*
- 12th month: end of the project.

Appendix



Letter of support from Electronic Arts

This is a letter of support for the "An Open-Source Benchmark for Automated Testing and Validation of Complex 3D Game Environments" proposal. This proposal is of high interest to Electronic Arts (EA) as it aligns well with our industry's challenges that need to be tackled. Automated game testing is becoming a more and more important tool for game production, but tools to evaluate methods and algorithms for this purpose are lacking. It would be highly beneficial for a company like EA to have access to this kind of environment for testing our methods on. We also see a lot of overlap between gaming industry and robotics where sample efficient exploration in complex and large-scale 3D environments, dynamics environments, and path planning is important.

Additionally, by making the simulated environment open source, they will be inviting the community to contribute and collaborate on its development. This will not only help to improve the quality and capabilities of the environment, but will enable swedish researcher to be in a leading position in this area.

Not only can this project give us the opportunity to test our methods, but it can also bring more researchers into the area so that they get the opportunity to develop and research new algorithms within this field. In the long run, this has the potential to have a large positive impact on the industry as a whole.

Konrad Tollmar

2022-12-20, Stockholm

Dr Konrad Tollmar, Research Director, SEED - Electronic Arts
ktollmar@ea.com



sahar.asadi@king.com
Sveavägen 44, 111 34 Stockholm
20 December 2022

Letter of Support

To the WASP AI Bridge Project Selection Committee

This is a letter of support for the "An Open-Source Benchmark for Automated Testing and Validation of Complex 3D Environments" proposal. This proposal is of high interest to King Digital Entertainment (Midasplayer AB) as it aligns well with our industry's challenges that need to be tackled. Automated game testing is becoming a more and more important tool for game production, but tools to evaluate methods and algorithms for this purpose are lacking. It would be highly beneficial for a company like King Digital Entertainment to have access to this kind of environment to test our research methods on. We also see an overlap between the gaming industry and robotics where reinforcement learning methods and sample efficient exploration techniques in large scale complex dynamic environments are important. Not only can this project allow us to test our methods, but it can also bring more researchers into the area so that they can develop and research new algorithms within this field. In the long run, this has the potential to have a large positive impact on the industry as a whole.

King Digital Entertainment is an internationally renowned game developer behind games such as Candy Crush Saga. King employs hundreds of developers, and its games are played by tens of millions of people every day. To ensure such a large-scale company can continue to operate and grow, we invest in state of the art AI research. One of our research focus areas is to apply AI to the game development process, including building AI tools for playtesting content creation and level design. We have an industrial PhD student in collaboration KTH Royal Institute of Technology as part of WASP that works on reinforcement learning and imitation learning for game playtesting.

Yours sincerely,

Sahar Asadi

Dr Sahar Asadi
Director, King AI Labs



UMEÅ UNIVERSITY

Umeå University
Department of Computing Science
90187 Umeå, Sweden

This is a letter of support for the *An Open-Source Benchmark for Automated Testing and Validation of Complex 3D Game Environments* proposal. I think this is a valuable and important effort that will have a significant impact on the broader community. The ability to test and iterate on algorithms and tools in a simulated environment is crucial for any industry, but especially for the fast-paced and constantly evolving world of gaming. By open-sourcing the environment, they will not only be providing a valuable tool for game developers, but also for researchers, educators, and others who can benefit from the ability to simulate and test various algorithms and scenarios in a controlled setting.

Additionally, by making the simulated environment open source, they will be inviting the community to contribute and collaborate on its development. This will not only help to improve the quality and capabilities of the environment but will enable Swedish researchers to be in a leading position in this area.

I believe the proposed project will benefit Swedish and international researchers so that they can do relevant research that benefits various Swedish industries. In the long run, it will increase the use of machine learning in game production, and very likely also related industries that can make use of simulated environments when testing their machine learning algorithms and tools.

I wholeheartedly support your efforts in open-sourcing this simulated environment and am excited to see the positive impact it will have on the gaming industry and broader industries.

Yours faithfully,

A handwritten signature in black ink, reading "J. Björklund".

Johanna Björklund
Associate Professor, Umeå University