

World of Bugs: A Platform for Automated Bug Detection in 3D Video Games

1st Benedict Wilkins

Computer Science

Royal Holloway University of London

London, UK

0000-0002-9107-2901

2nd Kostas Stathis

Computer Science

Royal Holloway University of London

London, UK

0000-0002-9946-4037

Abstract—We present World of Bugs (WOB), an open platform that aims to support Automated Bug Detection (ABD) research in video games. We discuss some open problems in ABD and how they relate to the platform’s design, arguing that learning-based solutions are required if further progress is to be made. The platform’s key feature is a growing collection of common video game bugs that may be used for training and evaluating ABD approaches.

Index Terms—Video Games, Automated Testing, Automated Bug Detection (ABD), Machine Learning, ABD Platform

I. INTRODUCTION

Video games are uniquely positioned in the space of software. They are designed to be an immersive experience for the user with a focus on narrative, interactivity and rich sensory presentation. However, the level of immersion a player experiences is drastically diminished if bugs are discovered during play. Although the industry invests significant effort in testing games before launch, it is still common for at least some bugs to slip through. One of the major contributing factors is the relative size and complexity of modern video games. It is becoming impractical to manually test every interaction and explore everything that a game has to offer even with a dedicated team of testers.

With recent developments in automated game playing, software agents may offer a solution to the testing automation problem. Automated bug detection (ABD) is an emerging field of study that is receiving increasing attention in the video games research community [1, 4, 5, 12, 13, 10, 14]. The focus of ABD research is on designing software agents capable of (a) *exploration*, playing the game in order to discover bugs, and (b) *identification*, to determine whether a particular situation contains a bug.

As is common in an emerging field, the literature is somewhat fractured when it comes to testing and comparing approaches to these problems. Without direct ties to industry it is difficult to obtain a development version of a game, meaning that researchers need to construct their own. Ironically, as bugs are hard to come by, the experimental setup cost is high. This has discouraged progress on certain problems associated with ABD, particularly on methods for identifying some of the more challenging bugs that video games exhibit.

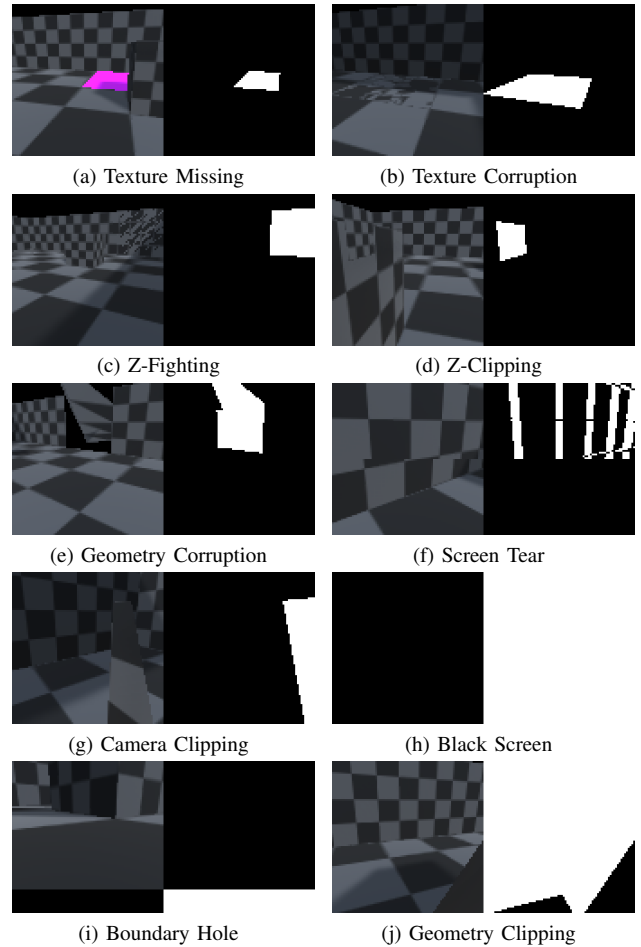


Fig. 1: Some of the bugs implemented in WOB. Each agent observation (left) contains a bug which is labelled with a mask (right).

With the aim of supporting further work in all aspects of ABD we have developed World of Bugs (WOB), an open experimentation platform built with the Unity3D engine. WOB provides a means of training and testing ABD agents in both exploration and bug identification. The core contribution of this work is in a growing collection of common video game

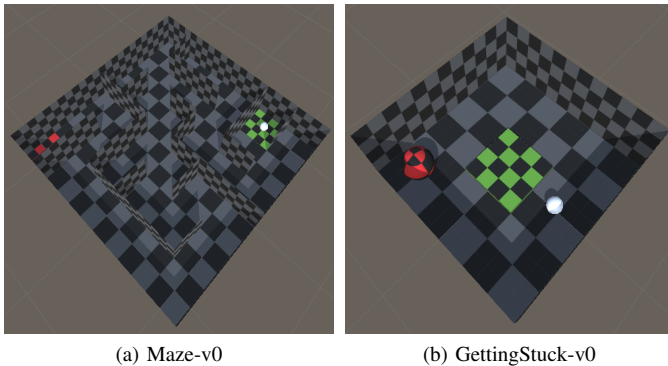


Fig. 2: Example environments built in to WOB. The agent appears as a white sphere. In (a) the agent is able to see through some walls by clipping the camera, constituting an *invalid information access* bug. In (b) the agent can get stuck at the bottom of an elevator shaft after taking a particular sequence of actions. Both environments also include various perceptual bugs.

bugs [7, 8, 9] implemented in the platform, their integration with the abstractions provided by Unity’s ML-Agents package [6] and support for learning-based bug identification. The platform also provides a collection of pre-built baseline game environments, agents and datasets which researchers can use to conduct experiments.

II. WORLD OF BUGS (WOB)

We have focused our attention on logical and perceptual bugs in 3D first person games. Logical bugs include crashes, getting stuck, performing invalid actions, out-of-bounds and invalid information access bugs, among others. Perceptual bugs manifest visually to the player (see Fig. 1) and include geometry and texture corruption, clipping issues, shadow-acne, and other rendering related issues. This setup provides a substantial challenge but is not completely out of reach of modern machine learning algorithms for both the exploration and identification problems.

Of the approaches to ABD many focus on developing explorative agents. Works such as [2, 4, 14] focus on learning agents capable of finding simple logical bugs (e.g. getting stuck, or out-of-bounds) by extensive exploration of the game with rules or guards for identification. However, the sorts of bugs that game testers are looking for are often not easily identified with such methods. High-level behavioural bugs such as those related to progression, narrative, or involve complex interactions or game AI are difficult to identify with rules, and it is not yet clear how they may be tackled in general.

For perceptual bugs and the more complex logical bugs, which have scarcely been explored in conjunction with exploration, vision and learning based methods are a promising direction [13, 9]. Progress on these bugs is an essential next step in the development of more general ABD agents that do

not rely on hand-crafted rules for identification. The primary purpose of WOB is to support work in this area, but it also has the potential to be extended to support research into the more challenging behavioural bugs.

Details on the implementation of bugs, the tools for developing agents and environments, and some examples are given in the next sections. The platform is available as a Unity package on the open source unity package registry OpenUPM, and on GitHub. Links and further details can be found [here](https://benedictwilkins.github.io/world-of-bugs/)¹.

A. Agents

An agent in WOB is created by default with three sensors: a main camera, which renders the view of the scene as the player would see it, a bug mask camera, which renders a mask over the scene showing in which regions there is a bug present, and a sensor which records various environment/agent properties such as the agent’s position and rotation. The bug mask acts as a label for the agents observation and is instrumental in enabling machine learning to be applied to the bug identification problem. In an attempt to standardise the comparison of explorative agents, WOB also implements a set of common actions including movement, jumping, view changes and simple event-based interaction, which can be configured during setup. Agent behaviours, goals and bug identification models may be specified in Python or in C# with the support of ML-Agents and associated Python packages.

WOB provides a simple navigational agent that uses Unity’s built in navigation system inspired by [11]. The agent’s behaviour is to walk and look around, picking random points in the scene to navigate to. The agent can be used in simpler environments where extensive exploration is not necessary to ease experimentation with approaches to bug identification. Additionally WOB provides player controlled agents that may be used for debugging new environments or bug implementations, or for generating training data. The intention is that with time the community will provide a collection of ABD agents that may serve as comparisons for further work.

B. Bugs

Bugs are implemented as collections of scripts and shaders, they can be added to the scene and enabled/disabled from our Python API. Newly implemented bugs will be automatically registered with the API when added to the scene. Implementing a bug requires two considerations: how to manifest the bug in the main camera and where to mask in the mask camera. The first is straightforward, and often is just a matter of writing a script to modify some properties of the environment (e.g. deleting a texture for a missing texture bug). The second is more challenging, and for this reason we have provided some scripts and shaders to ease the process. Every bug has a tag, a unique identifier which determines its colour in the bug mask. Bug tags can be used to label bugs that are associated with particular objects in the environment, such as missing textures, corrupted geometry and bad animations. Once tagged,

¹<https://benedictwilkins.github.io/world-of-bugs/>

these objects will be rendered in the bug mask. Bugs such as screen tearing, flickering, and freezing, are slightly more difficult to label but typically involve a post processing step in the rendering pipeline to directly modify the final bug mask. For details on how this may be implemented refer to the supplementary documentation. The bug mask shader also contains code for rendering bugs that are not associated with a particular object and do not effect the whole scene, a good example is camera clipping (seeing through walls). It will automatically render the back-side geometry as part of the mask, that is, if the agent can see inside a geometry, the inside faces will be rendered in the mask. Additionally, it will render the sky box below a certain height, this may be used to label so-called *map-hole* or *boundary bug* (see Fig. 1).

The choice to label bugs as part of a mask over the players observation (the final rendered screen) has been made with generality and extensibility in mind. All bugs manifest at this level in some form². In cases where the bug includes a temporal aspect, for example with a miss-behaving NPC, the mask can be extended through time. A mask provides the greatest flexibility for labelling newly implemented bugs and gives more information than a single scalar label.

Labelling a newly implemented bug is an important design choice. The most appropriate mask is one that gives the most information about what has gone wrong. In many cases this corresponds to masking the afflicted game object, for example in the case of a misbehaving NPC the simplest choice may be to highlight its body. If a cut-scene fails to show at a given point things are less clear, perhaps the whole observation should be masked. For bugs such as the so-called *invalid information access* bug [8] in which a player gains access to information they are not supposed to have, the source (or a proxy indicator) of the information can be masked. A common example that is already supported by WOB is *camera clipping* (looking through a wall). This bug may enable the player to see the structure or exit of an upcoming maze (see Fig. 2.a). In such a case, backside faces may be visible and act as a proxy indicator.

The label mask can be used in a number of ways, for example, to test different inductive biases for learning models, for bug segmentation or classification, or simply in model evaluation. The platform is meant only as a means to experiment with different approaches to ABD. Practical deployment of these models is currently beyond the scope of this work, but is an important consideration going forward.

C. Environments

At this time WOB has a number of built-in environments, details for each can be found in the platform documentation. They provide a stable testing ground for ABD approaches and have been designed either to exhibit particular bugs (see Fig. 2), or simplify certain aspects of the ABD problem. For example, one environment consists of a single static

room in which the agent is able to move and look around. This environment should generally be used to test approaches to identifying perceptual bugs as exploration is trivial. In addition to the built-in environments currently available, new game environments can be built for testing specific kinds of behaviours or to support new kinds of bugs. The tools are available to make this process relatively straight forward. The current set of implemented bugs have been designed to be as generic as possible and can be applied directly to any newly implemented 3D environments.

D. Python API

To better manage experiments and open the platform up to the Python machine learning ecosystem, WOB exposes a Python API which makes heavy use of ML-Agents and OpenAI gym [3]. In addition to the interface provided by these packages, the WOB API allows users to enable and disable bugs, change agent behaviours and control other aspects of the environment and agents at run-time.

E. Datasets

To our knowledge there are currently no publicly available datasets that contain examples of video game bugs that are suitable for training/testing learning-based approaches to bug identification³. As it is often easier to work with a static dataset when training learning models, WOB has explicit support for generating datasets given an environment and a game playing agent. We have generated a collection of datasets using the built-in agents and environments which we hope will serve as baselines for bug identification research. They consist of large (up to 500k) collections of labelled observations and actions. Links to datasets can be found in the platform documentation.

III. LIMITATIONS

The focus of WOB is currently on visual issues, we have made no consideration of audio related problems. While not explicitly supported, many of the design decisions that have gone into visual issues could be reused for audio. Although ML-Agents support parallel agents running in the same instance of Unity. This is not currently supported by WOB as our Python API makes use of OpenAI gym which supports only single agent interactions. It is however possible to create multiple instances of any WOB environment running as separate processes to enable parallel training.

IV. CONCLUSIONS & FUTURE WORK

The bugs implemented thus far are a small portion of the many diverse bugs that video games exhibit. There are plans to grow the list of available bugs and move beyond simpler logical and perceptual bugs into more challenging areas such as progression and narrative. For this to happen however, richer game environments are required. The game environments currently provided are simple when compared with games that ABD might eventually be applied to in practice. We hope that over time the community will provide some more interesting

²Audio related issues are not considered in this work, but a similar idea could be applied there.

³[13] provides a dataset, however the bugs are artificial and unrealistic.

examples that may exhibit a wider variety of bugs. ABD has great potential in addressing some key issues in testing modern video games, but to do so requires a deeper exploration of both game playing and bug identification. WOB provides researchers with the tools required to make progress on some of the key problems in ABD.

REFERENCES

- [1] Sinan Ariyurek, Aysu Betin-Can, and Elif Surer. “Automated Video Game Testing Using Synthetic and Humanlike Agents”. In: *IEEE Transactions on Games* 13.1 (Mar. 2021), pp. 50–67. ISSN: 24751510. DOI: [10.1109/TG.2019.2947597](https://doi.org/10.1109/TG.2019.2947597). arXiv: [1906.00317](https://arxiv.org/abs/1906.00317).
- [2] Joakim Bergdahl et al. “Augmenting Automated Game Testing with Deep Reinforcement Learning”. In: *IEEE Conference on Computational Intelligence and Games, CIG*. Vol. 2020-Augus. IEEE Computer Society, Aug. 2020, pp. 600–603. ISBN: 9781728145334. DOI: [10.1109/CoG47356.2020.9231552](https://doi.org/10.1109/CoG47356.2020.9231552). arXiv: [2103.15819](https://arxiv.org/abs/2103.15819).
- [3] Greg Brockman et al. *OpenAI Gym*. 2016. eprint: [arXiv: 1606.01540](https://arxiv.org/abs/1606.01540).
- [4] Camilo Gordillo et al. “Improving Playtesting Coverage via Curiosity Driven Reinforcement Learning Agents”. In: (Mar. 2021). DOI: [10.1109/cog52621.2021.9619048](https://doi.org/10.1109/cog52621.2021.9619048). arXiv: [2103.13798](https://arxiv.org/abs/2103.13798).
- [5] Stefan Freyr Gudmundsson et al. “Human-Like Playtesting with Deep Learning”. In: *IEEE Conference on Computational Intelligence and Games, CIG*. Vol. 2018-Augus. IEEE Computer Society, Oct. 2018. ISBN: 9781538643594. DOI: [10.1109/CIG.2018.8490442](https://doi.org/10.1109/CIG.2018.8490442).
- [6] Arthur Juliani et al. “Unity: A General Platform for Intelligent Agents”. In: (Sept. 2018). ISSN: 2331-8422. arXiv: [1809.02627](https://arxiv.org/abs/1809.02627).
- [7] L. Levy and J. Novak. *Game Development Essentials: Game QA & Testing*. 2009.
- [8] Chris Lewis, Jim Whitehead, and Noah Wardrip-Fruin. “What went wrong: A taxonomy of video game bugs”. In: *FDG 2010 - Proceedings of the 5th International Conference on the Foundations of Digital Games*. ACM Press, 2010, pp. 108–115. ISBN: 9781605589374. DOI: [10.1145/1822348.1822363](https://doi.org/10.1145/1822348.1822363).
- [9] Alfredo Nantes, Ross Brown, and Frederic Maire. “Neural network-based detection of virtual environment anomalies”. In: *Neural Computing and Applications* 23.6 (2013), pp. 1711–1728. ISSN: 09410643. DOI: [10.1007/s00521-012-1132-x](https://doi.org/10.1007/s00521-012-1132-x).
- [10] Johannes Pfau, Jan David Smeddinck, and Rainer Malaka. “Automated game testing with ICARUS: Intelligent completion of adventure riddles via unsupervised solving”. In: *CHI PLAY 2017*. Association for Computing Machinery, Inc, Oct. 2017, pp. 153–163. ISBN: 9781450351119. DOI: [10.1145/3130859.3131439](https://doi.org/10.1145/3130859.3131439).
- [11] I.S.W.B. Prasetya et al. “Navigation and exploration in 3D-game automated play testing”. In: *Proceedings of the 11th ACM SIGSOFT International Workshop on Automating Test Case Design, Selection, and Evaluation*, pp. 3–9. ISBN: 9781450381017. DOI: [10.1145/3412452.3423570](https://doi.org/10.1145/3412452.3423570).
- [12] “Video Game Automated Testing Approaches: An Assessment Framework”. In: *IEEE Transactions on Games* (2020), undefined–undefined. ISSN: 24751510. DOI: [10.1109/TG.2020.3032796](https://doi.org/10.1109/TG.2020.3032796).
- [13] Benedict Wilkins, Chris Watkins, and Kostas Stathis. “A Metric Learning Approach to Anomaly Detection in Video Games”. In: *IEEE Conference on Games*. Vol. 2020-Augus. 2020, pp. 604–607. ISBN: 9781728145334. DOI: [10.1109/CoG47356.2020.9231700](https://doi.org/10.1109/CoG47356.2020.9231700). arXiv: [2005.10211](https://arxiv.org/abs/2005.10211).
- [14] Yan Zheng et al. “Wuji: Automatic Online Combat Game Testing Using Evolutionary Deep Reinforcement Learning”. In: *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, Nov. 2019, pp. 772–784. ISBN: 978-1-72812-508-4. DOI: [10.1109/ASE.2019.00077](https://doi.org/10.1109/ASE.2019.00077).