

LADDER: Level Analysis Dataset for Difficulty Evaluation and Ranking

Yao Jean-eudes Adjanohoun, Yannick Francillette , Hugo Tremblay , Bruno Bouchard, 

dept. of computer science and mathematics

Université du Québec à Chicoutimi

Chicoutimi, Canada

yjadjanoho@etu.uqac.ca, yannick_francillette@uqac.ca, hugo_tremblay2@uqac.ca, bruno.bouchard@uqac.ca

Abstract—Standardized datasets are fundamental to scientific research. While fields like natural language processing and computer vision have widely accepted datasets that drive progress, video game research still lacks such resources, particularly for studying level difficulty. Existing studies rely on isolated, custom datasets, limiting cross-study comparisons and hindering the development of generalizable models. To bridge this gap, we introduce LADDER, a novel dataset specifically designed to analyze and evaluate level difficulty in video games. Unlike previous datasets that primarily focus on physiological and behavioral player data, LADDER integrates objective performance metrics (e.g., health lost, number of attempts before success), level characteristics (e.g., number of danger zones, object placement), and perceived difficulty ratings across multiple platformer games. This dataset enables researchers to establish benchmarks, enhance collaboration across disciplines, and improve study reproducibility. LADDER provides a standardized foundation for investigating the relationship between game design elements and player experience. By facilitating difficulty assessment and level balancing, it supports advancements in game design, player modeling, and adaptive gameplay systems. We present an overview of existing datasets, describe the methodology behind LADDER’s construction, and showcase its potential through preliminary analyses. The dataset is freely available online, offering a valuable resource for the scientific community to develop more engaging and accessible gaming experiences.

Index Terms—Video Game, Level of Difficulty, Standardized Dataset, Platformer Games

I. INTRODUCTION

Standardized datasets have become an essential pillar of scientific research, fostering reproducibility, comparability, and innovation across disciplines [1]. They enable researchers to validate findings, test new hypotheses, and develop benchmark methodologies, ensuring that progress is built on a consistent foundation. In fields such as natural language processing or computer vision, the availability of widely accepted datasets (e.g., ImageNet, COCO) has driven unprecedented advances by creating a common ground for evaluation [2].

In the context of video game research [3], however, the lack of standardized datasets remains a critical barrier [4]. Studies on game difficulty, player behavior, and level design often rely on custom, isolated datasets tailored to specific experiments. This fragmentation hampers cross-study comparisons and slows the development of generalizable models or insights [5]. A standardized dataset addressing level difficulty, for example, would allow researchers to:

- Establish baselines: By providing a shared benchmark, such datasets can facilitate the comparison of algorithms or models for assessing difficulty.
- Encourage collaboration: A shared resource can unite researchers from different domains (e.g., AI, psychology, design) and foster interdisciplinary studies.
- Promote transparency: Open datasets make it easier to replicate studies, assess the robustness of findings, and identify potential biases in experimental designs.

The creation of such a dataset tailored to level difficulty would not only fill this gap but also pave the way for standardized methodologies in video game research, much like ImageNet has done for computer vision [2]. It represents a necessary step toward advancing our understanding of how players interact with games and how design influences engagement, performance, and satisfaction.

In this paper, we aim to bridge this gap by providing researchers with a tool to explore and refine concepts of difficulty in a reproducible, scalable, and insightful manner. More specifically, we focus on analyzing the difficulty of individual video game levels. The challenge posed by each level plays a critical role in shaping the player experience, directly influencing engagement, satisfaction, and retention [5]. Despite its importance, “level difficulty” remains a subjective concept, highly dependent on the interplay between level design and player skill [6]. Quantifying difficulty is essential not only for designing balanced and engaging games but also for advancing research in game design, player modeling, and adaptive gameplay systems [7]. However, the lack of standardized datasets dedicated to evaluating level difficulty presents a major challenge for both researchers and practitioners.

Existing datasets [7]–[12] primarily focus on player data collected during gameplay, such as eye movements, blood pressure or stress levels. A comprehensive dataset that integrates objective player performance metrics (e.g., health lost, number of attempts before success, bullets used), level characteristics (e.g., number of danger zones, object placement), and perceived difficulty across various games and genres is therefore crucial.

To address this issue, we introduce LADDER, a novel dataset specifically designed to analyze and evaluate level difficulty in platformer video games. This dataset provides a robust foundation for studying the relationship between game design

elements and player behavior, facilitating the development of more advanced tools for difficulty assessment and level balancing. Ladder is freely available online and can be used by all the scientific community¹. We first review existing datasets, then describe the methodology used to construct LADDER, highlight its key features, and demonstrate its potential through a series of preliminary analyses. By offering a standardized resource, we aim to support future research and innovation in game design, ultimately contributing to more engaging and accessible gaming experiences for players of all skill levels.

II. RELATED WORK

In the scientific literature, very few standardized datasets have been proposed to investigate difficulty in video games [7]–[12]. Most existing datasets focus on player-centric elements collected during gameplay sessions [6], [8], [12], such as blood pressure, eye position, and other physiological or behavioral metrics. This highlights a clear need for datasets emphasizing game-centric elements that measure the intrinsic difficulty of the challenges posed by the game itself.

Intrinsic (or game-centric) difficulty refers to the features intentionally designed to create challenges of varying complexity. These include, for example, the topology of a level, the number, type, and positioning of enemies, the use of timers, balancing mechanics, and dynamic elements such as moving platforms [13].

One of the first and most well-known dataset is the one of Yannakakis and his team named Platformer Experience Dataset (PED) [8], which has been published in 2015. PED is an open-access game experience corpus providing a multimodal dataset of user data collected from *Super Mario Bros* players. Designed to capture player experiences, the database integrates context-based information (i.e., game content) with behavioral and visual recordings of platform game players. Additionally, it includes demographic data and self-reported annotations of player experience in the form of ratings and rankings. This dataset represents a significant contribution and can be used to evaluate the difficulty of *Super Mario Bros* levels. However, it primarily focuses on elements related to player-centric difficulty and did not use the original levels of *Super Mario Bros*, but rather custom levels created by fans. This limits its usefulness for analyzing game-centric difficulty, and the use of custom levels instead of standard ones reduces reproducibility.

Later, Barman et al. [14] introduced GamingVideoSET, a dataset comprising twenty-four uncompressed raw gaming videos, each 30 seconds long, with 1080p resolution and a frame rate of 30 fps. While this dataset was primarily designed for research on gaming video quality assessment and the effects of video encoding techniques, it was not explicitly created for difficulty assessment. Although it could potentially be repurposed for evaluating game difficulty, its utility is significantly limited by the lack of annotations, absence of

player performance data, and missing descriptions of the levels featured in the videos.

Svoren et al. [9] introduced Toadstool: A Dataset for Training Emotionally Intelligent Machines Playing *Super Mario Bros*. The dataset comprises video, sensor, and demographic data collected from ten participants during gameplay. Sensor data was gathered using the Empatica E4 wristband, a device that provides high-quality measurements and is classified as a medical-grade tool. While this dataset can be used to analyze the difficulty of a level, it primarily focuses on player-oriented data and lacks information about game-centric elements, limiting its applicability for evaluating intrinsic game difficulty.

Recently, in 2024, D’Angelo et al. [11] introduced PlayMyData, a curated dataset comprising 99,864 multi-platform games gathered from the IGDB website. Using a dedicated API, they collected relevant metadata for each game, including descriptions, genres, ratings, gameplay video URLs, and screenshots. Additionally, they enriched PlayMyData with completion time estimates for each game by mining data from the HLTB (HowLongToBeat) website. This dataset is one of the most comprehensive currently available. However, it was not specifically designed to analyze game difficulty, and it lacks detailed data for individual game levels, limiting its utility for this particular purpose.

In summary, the lack of standardized datasets dedicated to evaluating game-centric difficulty presents a significant challenge for both researchers and practitioners [13]. The few datasets currently available [7]–[12] predominantly focus on player-centric elements. There is a pressing need for a comprehensive dataset that integrates player performance, level characteristics, and perceived difficulty across a variety of games and genres.

III. LADDER: A NEW LEVEL ANALYSIS DATASET FOR DIFFICULTY EVALUATION AND RANKING

Our project aims to develop a comprehensive and structured dataset containing player performance data for a series of platformer video games, and the information concerning each level. The primary games selected for data collection are *Super Mario Bros*¹, *Super Meat Boy*², and *Mega Man*¹.

The goal is to gather precise and detailed information on the levels of each game and on playing sessions containing player actions, for instance the time spent on each level, the number of attempts required to complete an objective, the number of death, the number of jumps, the areas where players encounter the most difficulty, and the strategies employed to overcome obstacles.

Our study primarily rely on freely available online data. It also incorporate data extracted from in-game performance capture tools and post-session questionnaires to gather player feedback on perceived difficulty. In the following sections, we present methodology form acquiring data, the structure of the dataset and its contain of the dataset. Thereafter, we show

²https://en.wikipedia.org/wiki/Super_Mario_Bros.

³https://en.wikipedia.org/wiki/Mega_Man

⁴https://en.wikipedia.org/wiki/Super_Meat_Boy

¹<https://github.com/ajeaneudes/ladder>



Fig. 1: *Super Mario Bros*: in game screenshot²



Fig. 3: *Megaman (original)*: in game screenshot⁴



Fig. 2: *Super Meat Boy*: in game screenshot³

how the dataset can be used to process and analyzed game difficulty using statistical and machine learning techniques to uncover correlations between player performance and game design elements.

A. Data Sources for Dataset Creation

We used different freely available sources to get the necessary data that allowed us to construct the dataset. Here a list of the data source:

- 1. **Speedrun.com** : The site is a community-driven platform dedicated to speedrunning, where players from around the world share their performances and records across various video games. The dedicated *Super Mario Bros* (SMB1) page features leaderboards, speedrun categories, guides, and run videos, serving as a key reference point for enthusiasts looking to optimize their times and compete with others. Speedrun.com offers data from several popular games, including *Super Mario Bros*, *Super Mario 64*, *Mega Man*, and *Super Meat Boy*.
- 2. **Kaggle**: Kaggle.com is an online platform that serves as a hub for data science and machine learning communi-

ties. It hosts datasets related to player performance in *Super Mario Maker*. Notably, the "SMMnet - Super Mario Maker" dataset offers temporal data on player interactions within the game. Accompanying this dataset are analyses such as "Data Analysis - Super Mario Maker," which explores player performance metrics, and "Super Mario Maker Top 10 Players," focusing on top player statistics. These resources provide insights into player behaviors, level completion times, and other performance indicators within *Super Mario Maker*.

- 3. **Archive.org**: Archive.org, also known as the Internet Archive, is a non-profit digital library that provides free access to a vast collection of digitized content, including websites, books, audio recordings, videos, and software. On archive, we found datasets containing detailed information on 100 levels of *Super Mario Maker 2*, focusing on difficulty metrics, player performance, and engagement patterns.
- 4. **Super Mario Bros. Gameplay Dataset**: This dataset was created by Raphael C.P. and is freely available on GitHub². It contains valuable information about each *Super Mario Bros* level, including a complete image of every level and 280 recorded play sessions. These sessions can be analyzed frame by frame at 60 frames per second.
- 5. **Platformer Experience Dataset (PED)** [8]: As cited in the Related Work section, The Platformer Experience Dataset (PED) is the first open-access game experience corpus, containing multiple modalities of user data from *Super Mario Bros* players. It includes demographic in-

²<https://github.com/rafaelcp/smbdataset>

formation, gameplay data, and self-reported player experiences, such as engagement, frustration, and challenge levels. Additionally, the dataset offers visual recordings of players during gameplay, capturing facial expressions and head movements. While the dataset is player-oriented, it also contains valuable game-centric data that can be compiled and merged with data from other sources.

B. Data Acquisition Protocol

The process of creating a standardized dataset for studying difficulty in platform game levels involves several key steps:

Data Manipulation with Python The first stage utilizes Python, a powerful language for data analysis. The Pandas library is employed for data manipulation and analysis, enabling efficient handling of structured data. NumPy is used for numerical computation, facilitating mathematical operations on large datasets. Data is processed from various formats, including CSV and JSON, allowing for flexibility in working with diverse file types. At the end, all the compiled data of the Ladder dataset is given in the form of a database and each level data could be downloaded in the form of a .csv file.

Web Scraping and Data Collection To gather relevant data, web scraping techniques are applied. HTTP requests are made using the Requests library to fetch data from websites. For parsing HTML content, we used the library *BeautifulSoup*³, providing a simple way to extract structured data from web pages. In cases where dynamic content needs to be captured, *Selenium*⁴ is employed to interact with websites and retrieve the necessary data. Additionally, web APIs are managed to access structured data from online sources. For instance, in the case of *Speedrun.com*, we used the Streamlit applications⁵.

Databases The Ladder dataset is organized and stored in databases for efficient querying and management. Fundamental SQL operations such as SELECT, INSERT, UPDATE, and DELETE are used to interact with relational databases. The SQLite database⁶, in combination with Python, offers a lightweight solution for storing and retrieving the dataset.

Data Cleaning and Preprocessing Before using the dataset for difficulty analysis, the dataset undergoes a rigorous cleaning and preprocessing stage. Pandas⁷ is used for cleaning, ensuring that the data is accurate and usable. Missing values and outliers are addressed to improve data quality. Finally, the data is normalized and standardized to ensure consistency and comparability, which is crucial for further analysis and modeling.

This systematic approach ensures that the dataset is reliable, consistent, and ready for detailed analysis, providing valuable insights into the difficulty levels of platform game levels.

C. Structure of the LADDER Dataset

The dataset comprises a large number of elements that have been extracted and compiled from various sources in accordance with the acquisition protocol. For the *Super Mario Bros* game, the dataset includes the following elements:

- An image (PNG) of the map for each of the 32 levels (256×240, 8-bit).
- Data from 386,947 game episodes covering all levels of the original *Super Mario Bros*.
- Number of jumps and number of run actions (run, right, left).
- Total number of frames.
- User scores for each game episode.
- For each of the 32 levels: minimum score, maximum score, mean score, mean number of jumps, etc.
- For each of the 32 levels: number of enemies, types of enemies, number of bricks, number of holes, etc.
- For 280 specific game episodes, additional data is available:
 - A detailed log of each action performed in the episode with a timestamp (jump, click, start, select, direction).
 - Frame-by-frame recording of the episode at 60 frames per second (a total of 737,134 images).
 - A tag indicating whether the episode resulted in a failure or a success.

The data related to *Super Mario Bros* is the most complete and extensive in the dataset. Naturally, it is easier to find gaming session data for this game compared to the others. Additionally, we provide data for 50 *Super Mario 64* game sessions and 18 *Super Mario World* sessions.

For *Mega Man*, we have data from 26 game sessions, capturing the same types of information as for the 280 Mario episodes (jump, click, start, select, direction, frame-by-frame recording, etc.). Similarly, for *Super Meat Boy*, the dataset currently includes 23 game sessions. All data for these two games were collected from speedrun.com.

To extract the data, we developed a script that analyzes video recordings of gameplay sessions. While the number of sessions for these two games is still limited, this is just the beginning. We plan to expand the dataset significantly to improve our analysis.

IV. ANALYZING DIFFICULTY USING THE LADDER DATASET

As stated in the introduction, our goal is to provide researchers with a tool that offers standardized and comparable data, enabling them to explore and refine concepts of difficulty in a reproducible, scalable, and insightful manner. Our work, like that of many researchers in the field [15]–[21], focuses on analyzing the difficulty of individual video game levels to develop tools that assist game designers. The difficulty of each level is a key factor in shaping the player experience, directly influencing engagement, satisfaction, and retention [5]. Accurately assessing difficulty is crucial not only for

³<https://speedrundata.streamlit.app/>

⁴<https://pypi.org/project/beautifulsoup4/>

⁵<https://speedrundata.streamlit.app/>

⁶<https://www.sqlite.org/>

⁷<https://pandas.pydata.org/>

crafting well-balanced and compelling games but also for driving advancements in game design, player modeling, and adaptive gameplay systems [7].

In this section, we provide few examples of how the dataset can be used to assess and compare methods for analyzing game difficulty. As a first example, consider a team developing a machine learning algorithm to estimate the difficulty of platformer game levels based on players performance. This team could easily leverage our dataset to test their approach and determine the effective difficulty of each level using the player data recorded across 386,947 game episodes. For instance, they could generate a heat map for each level, visualizing difficulty intensity based on player performance metrics. Figure 4 illustrates such an analysis, presenting a basic heat map computation for the original *Super Mario Bros* levels. This heat map is derived solely from the number of failures and retries recorded in each game episode, highlighting areas of higher difficulty.

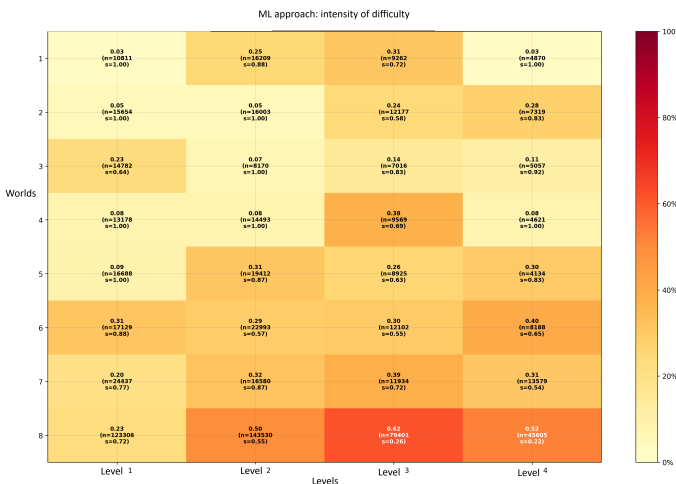


Fig. 4: Calculated intensity of difficulty for each Super Mario Bros level using ML technique with linear regression

As a second example, we recently introduced a novel formal model designed to assist game designers in creating platformer-type games [13]. Our model is entirely game-centric, enabling the automated assessment of executive and motor difficulty within a level by analyzing its interactive elements—specifically, those that produce collisions. It identifies static hazards (e.g., pits, spikes) and simulates dynamic threats posed by enemy movements using a pheromone-based algorithm [22]. This approach eliminates the need for extensive playtesting by providing an objective measure of motor skill difficulty. It does not take into account the players performance. To illustrate the usefulness of our dataset, we also computed a comparative heat maps for each original *Super Mario Bros* level using this non-ML approach, leveraging the same dataset. The results, shown in Figure 5, show how our model assesses difficulty without relying on machine learning.

We observe that the results of both approaches differ. However, the advantage of using a standardized dataset like

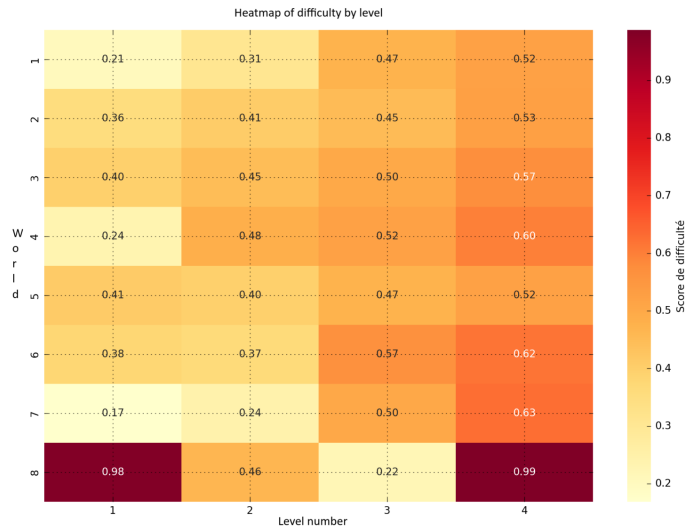


Fig. 5: Calculated intensity of difficulty for each Super Mario Bros level using the non-ML game-centric approach presented in [13]

LADDER is that it allows for an objective and consistent comparison of both results on a common factual basis.

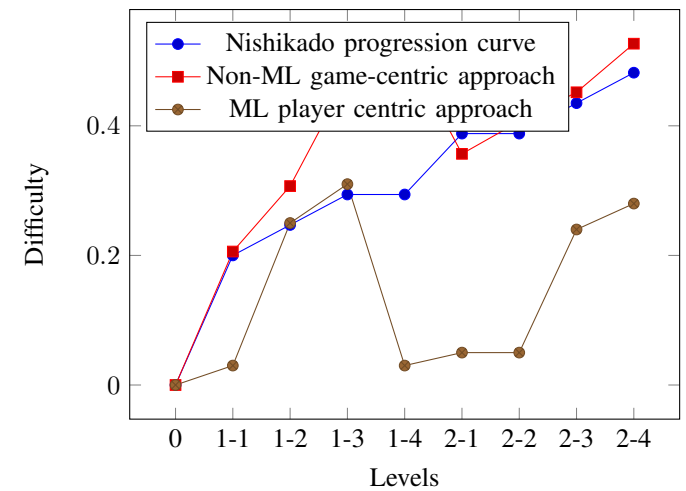


Fig. 6: Nishikado curve and computed difficulty curves.

As a comparative example, Figure 6 presents a comparison of the computed difficulty for each level using the ML-based approach based on player performance, our non-ML game-centric approach, and the designer-intended difficulty progression of the game, following the Nishikado curve [23]. As shown, for the first eight levels of the game, our non-ML approach exhibits a good correlation of 0.63 with the Nishikado curve. For the second part of the game, our approach give less coherent results because the elements of late games (advanced enemies like Lakitu, moving platforms, etc.) as not yet been included in our model. In contrast, the ML-based difficulty evaluation, derived from player performance, appears significantly less correlated, with a correlation factor of 0.27.

A possible explanation is that players attempting speedrun challenges tend to be highly skilled, making them unrepresentative of mainstream players. Additionally, the difficulty in the specific context of speedrunning does not always align with the difficulty of simply completing a level. Ultimately, these results provide a solid scientific foundation for assessing the effectiveness of the proposed difficulty measures based on a standardized dataset.

V. CONCLUSION AND FUTURE WORK

In this paper, we introduced LADDER, a standardized dataset designed to analyze and evaluate level difficulty in video games. By providing a structured and reproducible framework for studying the relationship between game design elements and player behavior, LADDER addresses a critical gap in video game research [4]. Its provide a solid foundation for studying game difficulty and, more specifically, the difficulty of levels in a game. Our preliminary analyses demonstrate its potential for fostering new insights into difficulty assessment, level balancing, and player experience modeling. Ladder is freely available online and can be used by all the scientific community.

Despite these contributions, several avenues remain open for future research. First, while LADDER integrates objective performance metrics and level characteristics, expanding it to include richer physiological and behavioral data (e.g., real-time stress responses, cognitive load measures) could further enhance difficulty modeling. Additionally, increasing the dataset's diversity by incorporating more game genres and player demographics would improve its generalizability and applicability across different gaming contexts.

Another promising direction involves leveraging LADDER to develop models for automated difficulty estimation and adaptive game design. These tools could be incorporated as plug-in inside common game engines such as Unreal 5⁸. Game designers would then be able to automatically assess the difficulty of a level and understand the impact on difficulty, in real time, of a design choice.

Finally, fostering collaboration within the research community is essential to refining and expanding LADDER. We encourage researchers and developers to contribute additional data, refine methodologies, and explore novel applications, ultimately pushing the boundaries of standardized game difficulty analysis. By establishing a shared foundation for difficulty assessment, we hope LADDER will inspire further innovation in game design, leading to more engaging, balanced, and inclusive gaming experiences.

REFERENCES

- [1] N. Pessanha Santos, "The expansion of data science: Dataset standardization," *Standards*, vol. 3, no. 4, pp. 400–410, 2023.
- [2] H. F. Castro, J. S. Cardoso, and M. T. Andrade, "A Systematic Survey of ML Datasets for Prime CV Research Areas—Media and Metadata," *Data*, vol. 6, no. 2, pp. 1–85, January 2021.

- [3] P. D. Paraschos and D. E. Koulouriotis, "Game difficulty adaptation and experience personalization: A literature review," *International Journal of Human–Computer Interaction*, vol. 39, no. 1, pp. 1–22, 2023.
- [4] E. J. Pretty, R. Guarese, H. M. Fayek, and F. Zambetta, "Replicability and transparency for the creation of public human user video game datasets," in *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2023, pp. 74–81.
- [5] Y. F. H. T. B. B. S. L. M. R. J. Linard, "Automated difficulty assessment model for platformer games: A comprehensive approach," in *2023 IEEE CTSoc Gaming, Entertainment and Media conf.*, 2023, pp. 1–6.
- [6] A. Sarkar and S. Cooper, "Inferring and comparing game difficulty curves using player-vs-level match data," in *2019 IEEE Conference on Games (CoG)*, 2019, pp. 1–4.
- [7] P. B. Jeppe Theiss Kristensen, "Difficulty modelling in mobile puzzle games: An empirical study on different methods to combine player analytics and simulated data," *International Journal of Computer Games Technology*, p. 14 pages, 2024.
- [8] K. Karpouzis, G. N. Yannakakis, N. Shaker, and S. Asteriadis, "The platformer experience dataset," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015, pp. 712–718.
- [9] H. Svoren, V. Thambawita, P. Halvorsen, P. Jakobsen, E. Garcia-Ceja, F. M. Noori, H. L. Hammer, M. Lux, M. A. Riegler, and S. A. Hicks, "Toadstool: a dataset for training emotional intelligent machines playing super mario bros," in *Proceedings of the 11th ACM Multimedia Systems Conference*, ser. MMSys '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 309–314.
- [10] X. Zhao, S. Liu, X. Li, G. Li, and X. Xu, "Video coding tool analysis and dataset for gaming content," in *2021 Picture Coding Symposium (PCS)*, 2021, pp. 1–5.
- [11] A. D'Angelo, C. Di Sipio, C. Politowski, and R. Rubei, "Playmydata: a curated dataset of multi-platform video games," in *Proceedings of the 21st International Conference on Mining Software Repositories*, ser. MSR '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 525–529.
- [12] H. H. R. Yin, L.; Zhang, "Game difficulty prediction based on facial cues and game performance," *Applied Sciences*, vol. 14, p. 8778, 2024.
- [13] Y. Francillette, H. Tremblay, B. Bouchard, and B.-A. Menelas, "A comprehensive model of automated evaluation of difficulty in platformer games," *ACM Games*, vol. 3, no. 1, Jan. 2025. [Online]. Available: <https://doi.org/10.1145/3705013>
- [14] N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini, and S. Möller, "Gamingvideaset: A dataset for gaming video streaming applications," in *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*, 2018, pp. 1–6.
- [15] F. Mourato and M. Santos, "Measuring difficulty in platform videogames," 01 2010, pp. 173–180.
- [16] S. J. Alexander, J.t. and A. Oikonomou, "An investigation of the effects of game difficulty on player enjoyment," *Entertainment Computing*, vol. 4(1), pp. 53–62, 2013.
- [17] V. Palban, "Managing difficulty in games," *Game Design Development*, pp. 1–4, 2021.
- [18] J. T. Kristensen, A. Valdivia, and P. Burelli, "Statistical modelling of level difficulty in puzzle games," in *2021 IEEE Conference on Games (CoG)*. IEEE Press, 2021, p. 1–8.
- [19] J. Knorr and C. Vaz de Carvalho, "Using dynamic difficulty adjustment to improve the experience and train fps gamers," in *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)*, ser. TEEM'21. New York, NY, USA: Association for Computing Machinery, 2021, p. 195–200.
- [20] J. Zhang, "Directly controlling the perceived difficulty of a shooting game by the addition of fake enemy bullets," in *CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '21. New York, NY, USA: Association for Computing Machinery, 2021.
- [21] N. Fisher and A. K. Kulshreshth, "Exploring dynamic difficulty adjustment methods for video games," *Virtual Worlds*, vol. 3, no. 2, pp. 230–255, 2024.
- [22] B.-F. Florea, O. Grigore, and M. Datcu, "Pheromone averaging exploration algorithm," in *2015 International Conference on Advanced Robotics (ICAR)*, 2015, pp. 617–622.
- [23] P. Holleman, *Reverse Design: Super Mario World*. CRC Press, 2018.

⁸<https://www.unrealengine.com/en-US/unreal-engine-5>