Unifying QD Process of Neural Cellular Automata into Deep Neural Networks for PCG Level Generation

[1] **Musawenkosi Gumpu,** [2] **Steven James,** [3] **Benjamin Rosman,**

[1] 2326254@students.wits.ac.za, [2] steven.james@wits.ac.za, [3] benjros@gmail.com
University of the Witwatersrand
School of Computer Science
and Applied Mathematics

Abstract

This study investigates the integration of Neural Cellular Automata (NCA) and Conditional Deep Generative Adversarial Network (CDGAN) models for procedural content generation in gaming. We replicate the NCA model to establish a baseline for quality diversity (QD) and adapt the CDGAN architecture to a maze format, exploring conditioning strategies for level generation. Our evaluations reveal insights into the diverse generation capabilities of both models and their impact on playability, reliability, and diversity metrics. Results show trade-offs between rapid generation and diverse level creation. Despite limitations in model training and convergence, this research provides crucial insights into optimizing GAN architectures for enhanced rapid, high-quality level generation.

Introduction

The pursuit of effective Procedural Content Generation (PCG) stands as a cornerstone in automated content creation, notably within the gaming sphere, where the demand for endlessly diverse and continuous game levels remains insatiable (Liu *et al.* 2021). PCG has revolutionized asset creation, significantly reducing time and costs while fostering diversity across various domains.

Amid the challenges necessitating real-time continuous asset generation, recent advancements, such as Reinforcement Learning (RL) models, have surfaced (Beukman *et al.* 2022). These models showcase impressive abilities to generalize across diverse level sizes without necessitating retraining, promising rapid content generation. Yet, their lengthy training periods and limitations in personalization pose significant challenges (Muir and James 2022). Combining evolutionary search (ES) and behaviour cloning offers a prospective solution by swiftly generating diverse levels, thereby augmenting the overall diversity (Muir and James 2022).

In addressing the crucial aspect of quality diversity (QD), Earle *et al.* (2021) introduces Neural Cellular Automata (NCA). This model delves into the level space of the domain, crafting an archive of diverse level generators through unsupervised learning, fostering the creation of a rich variety of levels.

Banerjee and Chen (2021) harnesses the potential of Conditional Generative Adversarial Networks (CGANs), adapting these generative models to level generation through a Conditional Deep Convolutional Generative Adversarial Network (CDGAN) architecture. This innovation stitches level boundaries, creating a continuity that enhances the overall coherence of generated levels.

Our primary objective is to explore the fusion of the vast array of diverse levels produced by the NCA within a CDGAN architecture. We aspire to ascertain the feasibility of creating a PCG model capable of emulating the quality diversity achieved by the NCA, while offering the added advantage of personalized level generation. This endeavour aims to develop a real-time, rapid level generator proficient in continuously producing a diverse set of levels according to specific criteria.

Related Work

Reinforcement Learning in Procedural Content Generation

Recent strides in RL models within PCG have garnered attention for transforming content generation tasks into Markov Decision Processes (MDP) (Liu *et al.* 2021). These models iteratively select actions aiming to maximize future content quality. However, content generation necessitates adapting inputs to suitable forms, posing a significant challenge (Liu *et al.* 2021).

Reinforcement Learning and Evolutionary Search

To address challenges like slow generation time and limited content diversity, a combination of Evolutionary Search (ES) and behaviour cloning had been applied to PCGRL models (Beukman *et al.* 2022). This combined approach aims to swiftly create new content without necessitating costly training or complex reward functions. The advantages include rapid generation without expensive training, enabling diverse content creation in real time. However, limitations in personalization and tailoring content to specific preferences or nuanced scenarios persist (Muir and James 2022).

Quality Diversity and Neural Cellular Automata

Exploration into diverse level generators in PCG has led to the pursuit of Quality Diversity (QD) approaches (Earle *et al.* 2021). This work focuses on generating diverse Neural Cellular Automata (NCA) level collections for video games. Utilizing Covariance Matrix Adaptation MAP-Elites (CMA-ME), the approach creates a variety of NCA level generators based on specific aesthetic or functional criteria. While efficient in generating diverse levels, NCAs' reliance on local changes for updates results in longer processing times, presenting challenges in real-time level generation (Earle *et al.* 2021).

Conditional Deep Generative Adversarial Networks

The adaptation of Conditional Generative Adversarial Networks (CGANs) to PCG, specifically in game-level generation, has introduced Conditional Deep Convolutional GANs (CDGANs) (Mirza and Osindero 2014; Banerjee and Chen 2021). These models condition the generator on previous frames, facilitating coherent level progression. However, challenges exist in bridging the information gap used for training the GAN and searching for noise vectors. Evaluation metrics based on player performance during gameplay could enhance level tailoring for better player experiences.

Methodology

Proposal and Research Objective

Our primary aim is to integrate the extensive levels generated by the Neural Cellular Automata (NCA) archive model into the training framework of a Conditional Deep Convolutional Generative Adversarial Network (CDGAN) (Earle *et al.* 2022; Banerjee and Chen 2021). This approach aims to combine the strengths of these distinctive models, creating a framework capable of preserving and potentially improving the quality diversity of levels generated, which is inherent in a GAN architecture (Liu *et al.* 2021). Simultaneously, this integration will also consequently expedite the level generation process, facilitating the seamless creation of coherent, continuous levels at real-time.

At the heart of our investigation lies the exploration of whether distilling a Quality Diversity (QD) model, as achieved by the NCA, into a Deep Neural Network (DNN) within the CDGAN architecture, can inherit the QD attributes demonstrated by the original model, particularly the behavioural characteristics like path length and level symmetry evaluated by Earle *et al.* (2021). This exploration seeks to confirm whether the distilled NCA model within a DNN framework retains the high QD and exploration depth inherent in the NCA's level generators.

Although the CDGAN developed by Banerjee and Chen (2021) excels in generating scrolling levels, our research aims to extend the CDGAN to create full game levels rather than just scrolling segments. This expansion aligns with our objective of upholding the depth of exploration within the level space and enabling the creation of personalized, diverse complete game levels, enabled by the conditioning provided by the CDGAN architecture.

The success of this integration allows us to acquire a comprehensive representation of quality diversity within the CDGAN's latent space. This is ensured by the encoding of essential level generation features within the generator model's latent space (Asperti and Tonelli 2023). Furthermore, this advancement holds the potential to enhance personalization, allowing conditioned targets to be defined by the feature representation obtained through Latent Space Exploration (LSE) without necessitating the initial use of handcrafted features (Fernandes *et al.* 2020).

Objective of Level Generator

We'll assess the performance of the level generator using a binary maze level. We'll compare the quality diversity of the NCA archive against the CDGAN level generator, employing evaluation metrics proposed by Earle *et al.* (2021). These metrics form an objective function o, a combined weighted measure of the following metrics:

- Playability $v \in [-\inf, 0]$

  This aspect evaluates how well the generated levels adhere to specific game constraints. In the context of the maze domain, it checks whether a single connected region in each level has been generated. A perfectly valid playable level will score v as 0.
- Reliability $r \in [-50, 0]$

  This measures the level of consistency generated in a batch of levels. In the maze domain, it assesses how consistently the generator produces levels of a specific length. The reliability penalty r is the batch's standard deviation along these measures.
- Diversity $d \in [0, 1]$

  This metric examines the variation among the generated levels to discourage the NCA generator from focusing solely on producing an optimal level.

Such that the objective o function is:

$$o = v + max(0, r + 10d)$$

Experimental Setup

Neural Cellular Automata Model Replication

We replicated the Neural Cellular Automata (NCA) model sourced from (Earle *et al.* 2022). Using the optimal hyperparameters to train the NCA model allowed us to obtain the baseline NCA archive. This process required the model to be trained over 50,000 iterations, taking three days to generate the archive of level generators.

The fitness function proposed by (Earle *et al.* 2021) was employed to evaluate the generated models, acting as a benchmark for assessing the quality diversity of the baseline model, evaluating aspects such as playability, reliability, and diversity, essential in evaluating the quality diversity of the generated levels.

Adapting the Conditional Deep Generative Adversarial Network

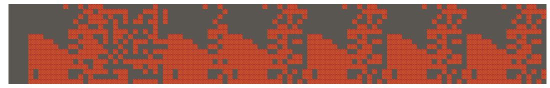
The CDGAN model, as presented in (Banerjee and Chen 2021), was adapted to accommodate a 16x16 maze format. Conditioning channels were put in place to encompass empty spaces and blocks, adjusting the model to the requirements of the experiment. To enable generalization capabilities, adjustments were implemented by introducing additional noise to the conditioned levels. Specifically, the first four columns of the levels were replaced with noise from a Gaussian distribution. This augmentation aimed to mitigate overfitting concerns and enhance the model's adaptability in generating new levels when conditioned on the entire level structure.

Alongside producing a single level, we also considered including the strip of levels from the original model in which the CDGAN as separate segments. This does come at a minor cost in the delay in instant level generation as we investigate the potential benefits of keeping pre-existing level segments. The comparing these methodologies helped clarify the most effective approach in enhancing level generation diversity and quality within the CDGAN framework.

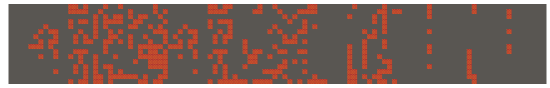
Furthermore, the CDGAN models were trained using the identical levels generated by the NCA model, to maintain consistency across the experimental comparison. Ensuring uniformity, the models were trained using their original architecture parameters, and the outcomes were averaged over four runs. The training iterations extended over 50,000 cycles, a duration mirroring that of the NCA model, also approximately spanning three days.

A supplementary investigation was conducted to explore the conditioning of levels entirely on noise derived from a Gaussian distribution. This exploration aimed to discern novel patterns in level generation, potentially uncovering unique insights into the impact of diverse conditioning strategies on level diversity and quality.

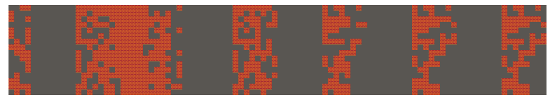
Replicating the model settings and maintaining a consistent evaluation metric, whilst exploring various techniques for ensuring quality diversity is maintained, helped enabling an adequate comparison between the NCA and CDGAN models.



(a) Partial Level Conditioning Scrolling



(b) Noise Level Conditioning



(c) Improved Partial Level Conditioning

Figure 1: CGAN Conditioning Type

Results and Discussion

Scrollable Levels vs Singular Level Generation

As depicted in Figure 1, the initial sections of the generated levels illustrate the effects of different conditioning approaches. Conditioning on noise alone, as seen in the first two sections, within the original CGAN architecture tailored for scrollable level generation, presents challenges. It's apparent that relying solely on noise conditioning may lead to non-convergent output, depleting the generated level into a minimal and uninteresting state. Our primary concern is to ensure the production of suitable levels in real-time. Our approach of partial level conditioning ensures convergence within a few iterations. Enhanced partial conditioning, specifically conditioning on the optimal number of columns (4 columns), results in swift convergence, requiring only a single iteration to generate a suitable and complete level.

NCA and CDGAN Evaluation Comparison

The table 1 presents a comparison between the NCA archive in figure 2 and CDGAN - namely the partial level conditioning of 3 - models, offering insights into various Quality Diversity (QD) metrics. One notable aspect is the broad variance observed in the NCA's fitness, which can be justified by the multiple generators producing a spectrum of level qualities. Visual analysis of path length, symmetry, and diversity in NCA levels further illustrates this. Conversely, the CDGAN showcased signs of potential overfitting, evidenced by fixed characteristics in the generated levels, influencing its overall fitness. Despite slightly lower fitness than the NCA, a single CDGAN generator demonstrated efficiency, swiftly producing high-quality levels compared to the iterative NCA process. Furthermore, while CDGAN levels exhibited slightly lower playability and reliability metrics, they displayed consistent performance across iterations.

Noisy vs Partial Level Conditioning

Conditioning noisy levels in the CDGAN resulted in reduced playability, consequently affecting the overall fitness score, as observed by figure 4. However, limitations were encountered due to the lack of control over personalized level types when employing noise conditioning, making conclusive analysis challenging.

Limitations and Future Work

Our study encountered several limitations that offer directions for future investigations. Time constraints played a significant role, limiting the duration for running the CDGAN model to generate levels. This is due to the known training times of GANs, particularly the CDGAN model's depth, leading to incomplete training on the entire dataset. This limitation potentially contributed to overfitting within the CDGAN model.

Looking ahead, there are promising avenues for further research. Matching the quality diversity exhibited by the NCA archive while enabling swift level generation remains an enticing prospect. Achieving this alignment in the future could pave the way for increased personalization through latent space exploration, as demonstrated by Fernandes *et al.* (2020).

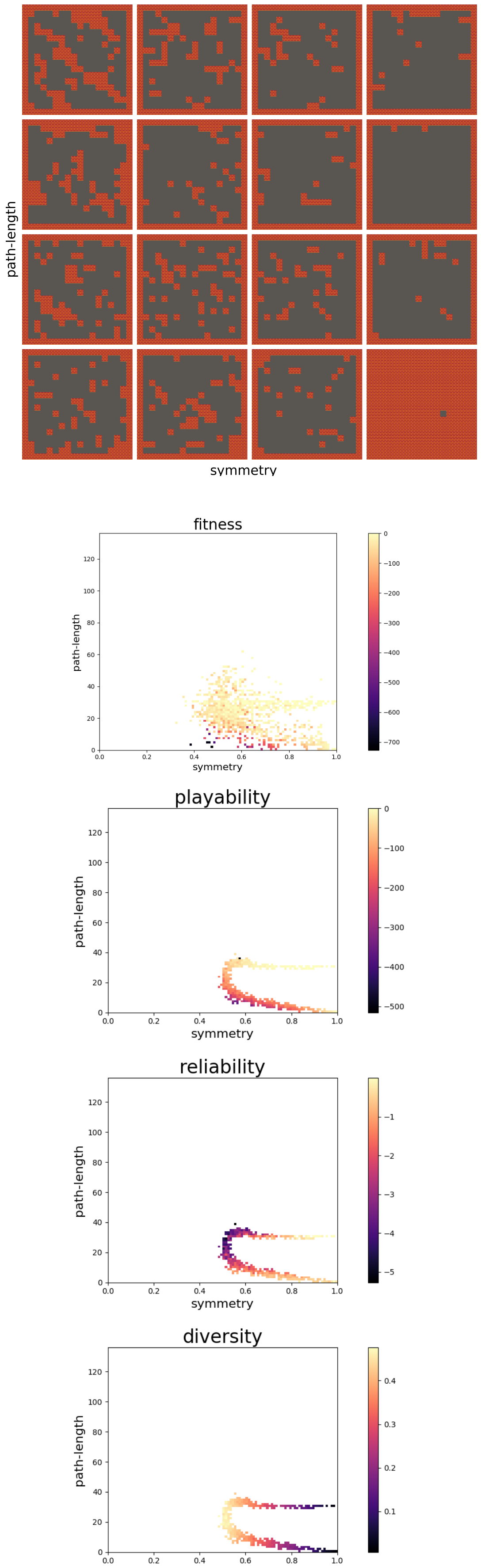


Figure 2: NCA Evaluation

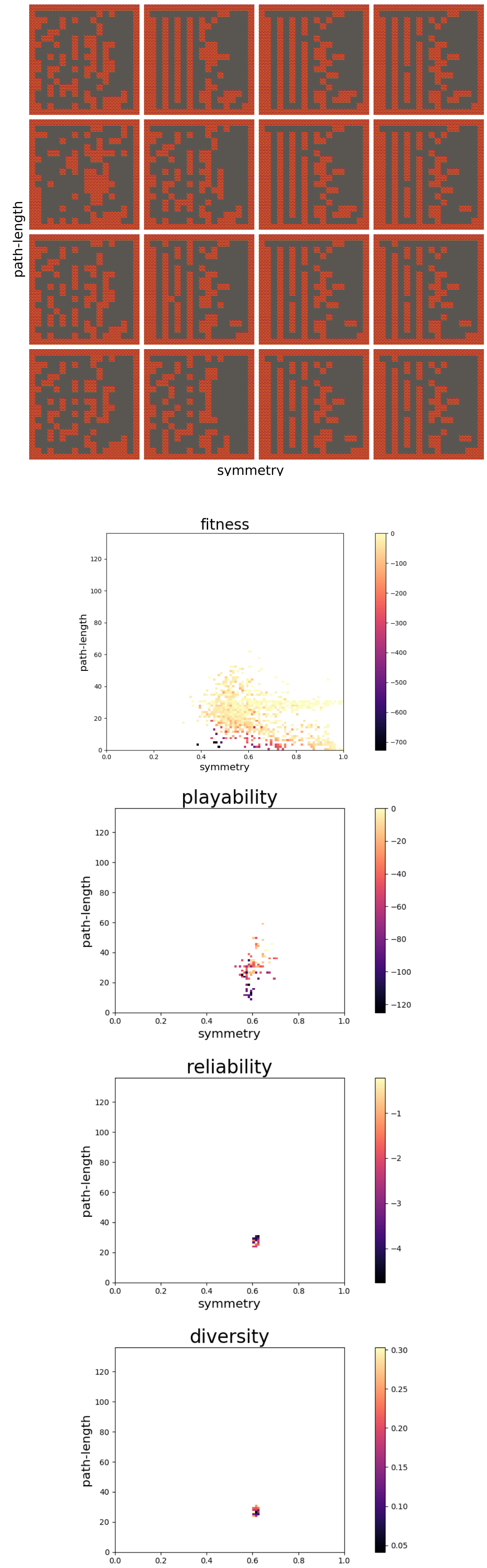


Figure 3: CGAN Evaluation

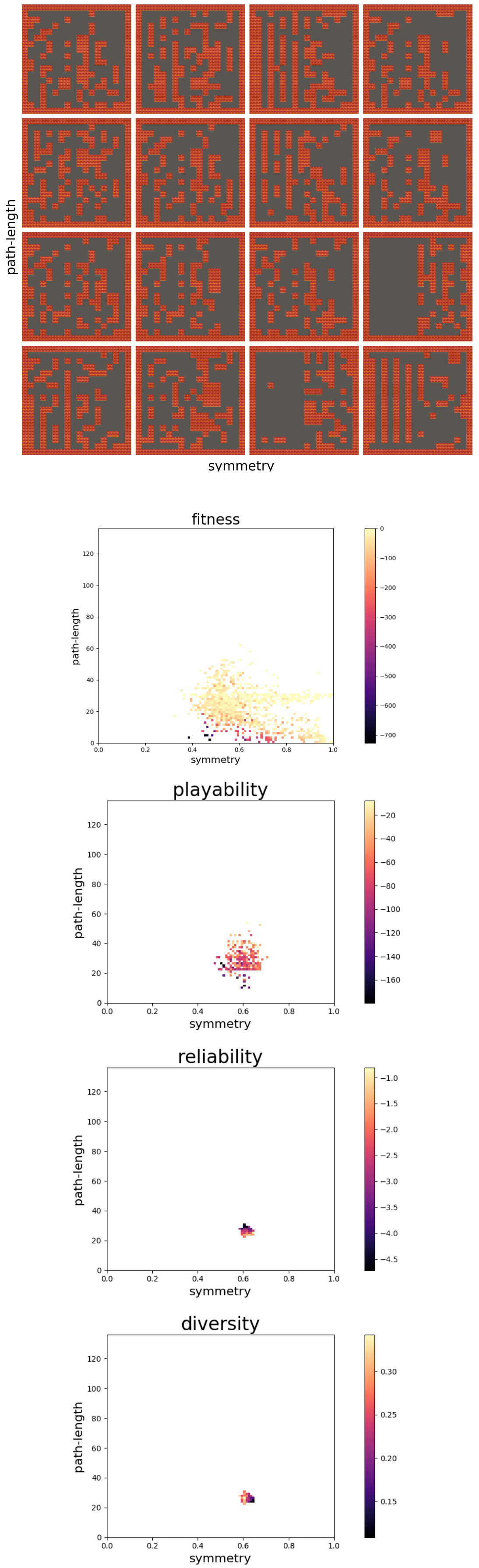


Figure 4: CGAN Noisy Conditioning Evaluation

Moreover, there are unexplored areas that do need attention. Optimizing the hyperparameters of the CDGAN architecture presents an opportunity for refining and enhancement of level generation quality and diversity. Delving deeper into the intricate design of GAN models within the realm of continuous, rapid, and high-quality diverse level generation promises to unveil new possibilities and insights.

Conclusion

In conclusion, our investigation demonstrates the potential distillation of NCA archives into CDGAN models for level generation in gaming. In replicating the NCA model and carefully adapting the CDGAN architecture, we discover the limitations and strengths of both approaches. The exploration of conditioning strategies, though hindered by model overfitting and training durations, sheds light for future research. The findings highlight the balance between speed and diversity in level generation, offering avenues for refining GAN architectures, optimizing hyperparameters, and making use of latent space exploration for personalized content generation. This work lays the foundation for further advancements in continuous, diverse, and rapid level generation, promising enhanced gaming experiences through tailored content creation.

Appendix

Tabular QD Results of NCA and CDGAN Models

Metric	**Fitness**	**Playability**	**Reliability**	**Diversity**
NCA	-106.0 ± 88.4	-107.3 ± 85.9	-1.8 ± 1.2	0.3 ± 0.12
CDGAN_P	-71.3 ± 9.1	-70.5 ± 7.1	-2.6 ± 1.3	0.1 ± 0.07
CDGAN_N	-70.5 ± 7.1	-70.5 ± 7.1	-2.6 ± 1.3	0.1 ± 0.07

Table 1: Comparison of QD metrics between NCA and CDGAN

Code

Code is made available on the following repository

Adjusted NCA code by Earle *et al.* (2022).

Adjusted CDGAN by Banerjee and Chen (2021).

Acknowledgements

Many thanks to Steven James and Benjamin Rosman for their invaluable guidance and unwavering support throughout this research task. Their expertise and assistance were instrumental in overcoming the challenges encountered.

We also would like to appreciate Richard Klein for his interesting insights into Computer Vision models. His profound knowledge and guidance significantly contributed to fortifying our confidence in the methodology adopted.

Their contributions have been indispensable to the successful realization of this research endeavour, and their support appreciated.

References

Andrea Asperti and Valerio Tonelli. Comparing the latent space of generative models. *Neural Computing and Applications*, 35:3155–3172, 2 2023.

Sourish Banerjee and David Chen. *GAN-Based Game Level Generation*, 2021.

Michael Beukman, Christopher W Cleghorn, and Steven James. Procedural content generation using neuroevolution and novelty search for diverse video game levels. pages 1028–1037. ACM, 7 2022.

Sam Earle, Justin Snider, Matthew C Fontaine, Stefanos Nikolaidis, and Julian Togelius. Illuminating diverse neural cellular automata for level generation. *arXiv (Cornell University)*, 3 2021.

Sam Earle, Justin Snider, Matthew C Fontaine, Stefanos Nikolaidis, and Julian Togelius. Illuminating diverse neural cellular automata for level generation. pages 68–76, 2022.

Paulo Fernandes, João Correia, and Penousal Machado. *Evolutionary Latent Space Exploration of Generative Adversarial Networks*, 2020.

Jialin Liu, Sam Snodgrass, Ahmed Khalifa, Sebastian Risi, Georgios N. Yannakakis, and Julian Togelius. Deep learning for procedural content generation. *Neural Computing and Applications*, 33:19–37, 1 2021.

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. 11 2014.

Nicholas Muir and Steven James. Combining evolutionary search with behaviour cloning for procedurally generated content. pages 77–64, 3 2022.