


Research on Artificial Intelligence-Assisted Game Design and Development

Xian Chen ^a

Faculty of Information Technology, Clayton Campus, Monash University, Melbourne, Victoria, Australia

Keywords: Artificial Intelligence, Game Design, Multimodal Generation.


Abstract: With the rapid development of Artificial Intelligence technology, Its application in game design and development become more and more extensive. AI technology not only improve the level of the intelligence in game, but also has a profound influence on content creation. AI enables game developer and designer to generate high-quality visual content and text with lower cost, improve the efficacy and creativity. This research explored the impact of AI development of game design and development procession, especially how creative AI can reduce the cost of graphic in game development and a new way of create text. In this paper, four key technologies will be introduced: cross-model generation, Generation based on Generative Adversarial Network (GAN), Diffusion model-based generation, and Autoregressive model-based generation. Then the paper will discuss game-assisted design methods and concludes with a future outlook of the field. This research not only showcases the innovative application of AI in game development but also provides new ideas and directions for the future development of the gaming industry.

1 INTRODUCTION

There are new production tools be provided for many field when the technology of AI-related rapid development, and quite significant breakthroughs have been made in many areas such as image generation, video generation and so on. At the same time, in the field of game design and development, AI technology is transforming from auxiliary tool to a core driving force for innovation in game content. In traditional game development processes, the attempts at designing content are indispensable and require a significant amount of time and resources, such as scene construction, character design and story planning. With the breakthroughs in GAN models, diffusion models, autoregressive models, and cross-modal generation technologies, the introduction of AI technology can not only significantly improve efficiency and reduce the costs, and also can provide new possibility for game development. Not only that, the gradually improvement of multimodal generation technology has also made the real-time generation of graphics and audio based on user input by AI become the expected direction of game design.

Base on the continuous development of AI technology, its application boundaries in the gaming industry are still expanding. From automated content generation to intelligent interactive experiences, AI is push game development into a new era. This not only reduces the costs and improves production efficiency but also provides game designers and developers with richer creative tools, making the possibilities for games even more expansive. In the future of the game development, AI may not only be an auxiliary tool for game development but could also become an important participant in game creation, potentially even leading to new paradigms in game design.

This article will introduce GAN model, Diffusion model, and Autoregressive model, the application of these models in AI technology, and the influence of game design with the introduction of AI technology. At the same time, three generative models trained by relevant principles and one architecture which uses large language models to simulate human behaviour will be presented.

^a <https://orcid.org/0009-0005-7127-3457>

2 KEY TECHNOLOGY

2.1 Cross-modal Generation

Cross-modal generative refers to the generation of data based on one or multimodal data to another modal data, for example, generate images from text. The difficulty of cross-modal generative lies in the conversion and fusion between different modal data. Lyu, Zheng and Wang (2024) proposed using the Stable Diffusion with pre-train and a generative method called Image Anything incorporates knowledge graphs and handles detailed attributes. This approach allows the model to perform cross-modal generation without additional raining (Lyu, Zheng, & Wang, 2024).

2.2 Generation based on GAN

GAN model is deep learning model which be proposed by Goodfellow et al., this model learns by adversarial interactions between the generator and discriminator within the framework, and produce outputs by this way (Goodfellow et al., 2014). The generator generates the data from the noise as true as possible and this data will be called fake data, the discriminator learns by distinguishing real data and fake data which generate by generator. GAN model train by the adversarial training between generator and discriminator, by this way, model can have ability to generate data similar with real data. Karras et al. purpose a new way called Progressive Growing, data training start from low resolution and lack of detail and increase resolution and details, this method can enhance the quality of generated images and shorten training time (Karras et al., 2017).

2.3 Generation based on Diffusion Models

Diffusion model is an outstanding generative model in the area of image generation and synthesis. The core steps are adding noise to the data and then use another model to remove the noise, thereby improve the ability of generation.

There are two processes called diffusion and reverse process. In diffusion process, the data will be added noise. In the reverse process, another model will remove the noise from the data which was added noise. Ho, Jain and Abbeel proposed the diffusion probabilistic model (Ho, Jain, & Abbeel, 2020). In the same year, Song and Ermon (2020) introduced a new addressed the challenges of learning and sampling in

high-dimensional spaces, improving the model's stability (Song & Ermon, 2020).

2.4 Generation based on Autoregressive Models

The generation based on autoregressive models differs from traditional generative models (such as GANs). This model uses Convolutional Neural Networks to establish relationships between pixels and predict the value of the next pixel by the existing pixels to generate an image. Van den Oord, Kalchbrenner, and Kavukcuoglu proposed the Pixel Recurrent Neural Networks model and applying pixel recurrent neural networks to image generation (Van Den Oord et al., 2016).

3 AI-BASED GAME DESIGN ASSISTANCE

The maturity of creative artificial intelligence in recognition and generation has provided many new tools for game design and development, facilitating the creation of assets and simplifying a large amount of work. However, depending on the form of the game, the role of artificial intelligence varies. Therefore, this paper categorizes game types into two main directions based on the primary modes of interaction between the game and the player: visual-based game design assistance and text-based game design assistance, and introduces them separately.

The visual-based game, which interact with players by images, videos, character models, and other visual elements, there is a typical examples such as most First-Person Shooter (FPS) games.

In FPS games, players need to perform precise actions by what they see in order to receive rewards. This type of game design is typically fast-paced and requires the use of images or other visual stimuli to provide quick responses for players.

Text-based games, which interact with players through dialogue text and written materials such as books, are represented by most mystery and detective games. Players typically need to extract useful information from large blocks of text in order to achieve their goals. These games often have strong interconnectivity, where the elements in the game require a tight and coherent logical relationship. Players must repeatedly verify from multiple directions to avoid confusion and errors.

The three artificial intelligence models introduced below represent different directions of creative AI in the field of visual effect generation.

3.1 Stable Diffusion 3(SD3)

Esser et al. proposed an improved noise sampling technique (Esser et al., 2024), which has been used in SD3 to enhance generation quality. SD3 is a diffusion model focused on image generation, and its main function is to generate relevant images from text. Under traditional flow model training, the computational cost is high. The SD3 model transforms the generation problem into optimizing a new loss function, which allows the model to more quickly reach the optimal solution and improve efficiency.

$$L_w(x_0) = -\frac{1}{2} E_{t \sim u(t), \epsilon \sim N(0,1)} [w_t \lambda_t' \|\epsilon_\theta(z_t, t)\|^2] \quad (1)$$

Function 1 is the loss function of the model at the initial data point x_0 . $E_{t \sim u(t), \epsilon \sim N(0,1)}$ denotes the expectation over a distribution. w_t is a weighting term at time step t . λ_t' is another weighting factor at time step t . $\|\epsilon_\theta(z_t, t) - \epsilon\|^2$ is the squared error between the model's predicted noise $\epsilon_\theta(z_t, t)$ and the actual noise ϵ . [7]

The flow trajectory is simultaneously optimized to enhance the corrected flow model, aiming to improve the training effectiveness, making the model more accurate and efficient. However, at intermediate time steps (when t is close to 0.5), the error tends to increase. Therefore, this method is equivalent to a weighted loss:

$$w_t^\pi = \frac{t}{1-t} \pi(t) \quad (2)$$

Function 2 is a time-step-dependent weighting factor, where $t/(1-t)$ gives more weight when t is larger, allowing the model to focus on recovering data from noise.

It also proposes architecture for text-to-image generation, Multimodal Diffusion Transformer, which can simultaneously handle information from both text and image modalities. Although this model still have some issues, such as: while removing the Text-to-Text Transfer Transformer (T5) text encoder improves the efficiency of image generation, the performance significantly decreases when generating written text without the T5 text encoder. However, the model's expansion trend shows no signs of saturation, and there is still considerable room for

improvement in the future. The SD3 model demonstrates the achievements of artificial intelligence in image generation and represents how AI can provide a large number of uniformly styled images during the game development phase, saving costs and development time after training.

3.2 Movie Gen

Polyak et al. proposed Movie Gen, a model based on GAN training that focuses on video generation (Polyak et al., 2024). Its main feature is generating high-quality videos while maintaining audio consistency. Movie Gen uses a 30B-parameter Transformer model to generate videos from text. The model is first pre-trained on low-resolution images, then jointly pre-trained on high-resolution images and videos, and finally fine-tuned on high-quality videos. This approach allows the production of videos.

By adding conditions to the pre-trained generative model, the model generates personalized videos by referencing images and text prompts. The key feature of Movie Gen is that it uses a 13B-parameter model to generate sound effects and music that are synchronized with the video, based on text prompts.

However, Movie Gen still has many shortcomings. For example, the generated videos may have issues with complex shapes and physics, and during action-intensive scenes, such as tap dancing, or when the visual is obstructed or small, such as footstep sounds, the audio may be out of sync. Additionally, it does not support language generation. Despite these issues, it can still be seen as a significant milestone in greatly reducing game development difficulty and costs.

Game developers can more easily use CG (videos played in games) within their games, offering more personalized and context-sensitive CGs, such as generating corresponding CGs based on the player's character appearance, thus enhancing the player's immersion. It provides an option between CG videos, which are more immersive, and real-time rendering, which is less prone to interference, offering a more balanced solution in real-time rendering and CG playback.

3.3 Genie

Genie is a generative interactive environment proposed by Bruce et al. in 2024 (Bruce et al., 2024). This model is based on a GAN diffusion model and can be considered a foundational world model. The key feature of Genie is its ability to generate interactive virtual environments through text, images,

photos, and even sketches — that is, interactive videos. It is also the first unsupervised and unlabeled training-based generative interactive environment. Genie’s latent action model is used to infer the potential actions between each pair of frames, while the dynamics model predicts the next frame of the video based on the inferred actions and past frames.

The introduction of Genie brings a new game development approach, namely, playable videos. In game development, Genie can easily generate game prototypes or test gameplay mechanics, and even supplement the content of games developed by creators.

3.4 Simulated Agent

A simulated agent refers to an intelligent entity generated based on corresponding data and controlled by artificial intelligence. In the study by Park et al., a language model was trained on in-depth interview data from participants to generate a simulated agent for each volunteer (Park et al., 2024). The agents then answered questions on various topics, such as the prisoner’s dilemma and other game-theoretic problems. By dividing the original accuracy (68.85%) by the participant replication accuracy (81.25%), the average accuracy achieved was 0.85 (Table 1).

Table 1: Performance of Different Agents in Social Survey.

General Social Survey	Accuracy	Normalized Accuracy	Correlation	Normalized Correlation
Participant Replication	81.25%(std=8.11)	1.00 (std=0.00)	0.83 (std=0.30)	1.00 (std=0.00)
Agents w/ Interview	68.85%(std=6.01)	0.85 (std=0.11)	0.66 (std=0.19)	0.83 (std=0.31)
Agents w/ Demog. Info.	57.00%(std=7.45)	0.71 (std=0.11)	0.51 (std=0.19)	0.63 (std=0.26)
Agents w/ Persona Desc.	56.79%(std=7.76)	0.70 (std=0.11)	0.50 (std=0.20)	0.62 (std=0.25)

In game development, non-playable characters (NPCs) can use simulated agents to reduce the workload and complexity of designing character dialogues. This approach allows game characters to

interact more flexibly with players based on their specific situation.

4 FUTURE OUTLOOK

With breakthroughs in AI technology and its increasing maturity, artificial intelligence is expected to significantly reduce time costs and lower the difficulty of acquiring multimodal resources in game design and development. Similarly, the advancement of multimodal generation will enhance the flexibility of game design, providing richer multimodal data to refine generative models. Concurrently, reduced costs will liberate game design objectives from resource constraints, enabling more sophisticated gameplay mechanics. Generative models will also achieve a balance between output quality and real-time performance through iterative usage. Furthermore, these models could operate on personal devices via local deployment, facilitating real-time generation to reduce reliance on server-side infrastructure. Even idle computing power from individual devices could be aggregated to optimize efficiency and utilization.

5 CONCLUSIONS

Artificial intelligence, through multimodal generation, is currently a major influence on game design and development. It can effectively reduce the difficulty of obtaining resources such as text and images, enabling rapid correction in game design and reducing costs. However, existing technology is still immature, and a large amount of training data is needed to standardize the generated resources before they can be produced. Furthermore, the inability to freely initiate multimodal generation from a single type of resource indicates that relying solely on generative models as the primary production tool in game design is unrealistic. The generative models listed in this paper are current models that have already achieved significant results in single-modal generation. They represent the capability of artificial intelligence to learn training data thoroughly and make reasonable predictions in generating corresponding modal resources. With more data training, generative models with more unified and long-term generation effects are not beyond reach.

REFERENCES

- Bruce, J., Dennis, M. D., Edwards, A., Parker-Holder, J., Shi, Y., Hughes, E., ... & Rocktäschel, T. (2024, January). Genie: Generative interactive environments. In Forty-first International Conference on Machine Learning.
- Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., ... & Rombach, R. (2024, July). Scaling rectified flow transformers for high-resolution image synthesis. In Forty-first International Conference on Machine Learning.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851. <https://arxiv.org/abs/1406.2661>
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of GANs for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.
- Lyu, Y., Zheng, X., & Wang, L. (2024). Image anything: Towards reasoning-coherent and training-free multi-modal image generation. *arXiv preprint arXiv:2401.17664*.
- Park, J. S., Zou, C. Q., Shaw, A., Hill, B. M., Cai, C., Morris, M. R., ... & Bernstein, M. S. (2024). Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109*.
- Polyak, A., Zohar, A., Brown, et al. (2024). Movie Gen: A cast of media foundation models. *arXiv preprint arXiv:2410.13720*.
- Song, Y., & Ermon, S. (2020). Improved techniques for training score-based generative models. *Advances in Neural Information Processing Systems*, 33, 12438-12448.
- Van Den Oord, A., Kalchbrenner, N., & Kavukcuoglu, K. (2016, June). Pixel recurrent neural networks. In *International Conference on Machine Learning* (pp. 1747-1756). PMLR.