ITM Web of Conferences **78**, 04009 (2025) https://doi.org/10.1051/itmconf/20257804009
CSEIT 2025

Application Scenario Analysis of Styleshot Based Game Style Generation

Chuangu Wang

School of Computer Science and Technology, East China Normal University, Shanghai, China

Abstract. The continuous improvement of hardware and technology has made a huge leap in the development of games, and the beauty and variety of game graphics have also gone to the next level. However, as the volume of games increases, there is a need for a project that can migrate different game styles to the characters or scenes that we want to generate, so as to ensure that the generated images fit the original styles and increase the efficiency of the work. The aim of this work is to select the most suitable project for migrating styles in game development by comparing the advantage zones of various style migration projects. With the help of comparative experiments, literature reading, and dataset analysis, it is concluded that the style migration project named StyleShot can meet the requirements well with its style-aware encoder, content fusion encoder, and a well-organized style dataset named StyleGallery, which improves the expressiveness and generalization of the style representation.

1 Introduction

As a gaming enthusiast, would you choose to reminisce your childhood in the age-old but still timeless pixel-art style games, or feel the Japanese secondary culture in the many anime-inspired secondary games on the market, or immerse yourself in another time and space to experience the game content in 3A masterpieces that are gradually breaking through the limits of picture quality? Known by young people as the “ninth art,” gaming is growing rapidly in this day and age.

As a result of this growth, the volume of games is increasing and the art styles are diversifying. Players want to choose games that fit their preferences. At the same time, for games that are gradually breaking through the sub-genre wall, only one style is no longer enough to satisfy the appetite of many game designers. For example, in this year's release of the Hazelight Studios developed by the “Split Fiction” will be with the advancement of the plot, the style of the game screen constantly switching, the main line in the sci-fi style and fantasy style switching back and forth, and the branch line is a young child animation style, pencil drawing style and other colorful styles [1].

To meet this demand, more art designers are needed. In addition to the core characters and scenes of a game, there are also a lot of performance scenes and NPCs that need to be designed. Therefore, through experimental comparisons to find a project to replace the art engineers to carry out some of the more repetitive and less important work, and then

10245102426@stu.ecnu.sdu.cn

improve the efficiency of the work. In this work, this research is based on StyleShot, Dream Booth and other projects for comparative experiments [2], reference to several literature, analyzed a number of datasets involved, and came to the conclusion that the project based on StyleShot is able to better fulfill the above requirements and has a better development prospect in the future game development, and did the analysis of the application of the game scenarios in this regard.

In many of today's style migration-driven generation tasks, there are inefficiencies in the face of multi-style generation scenarios, especially in the aspect of training models for different styles individually, which can take extra time in training. In addition, the datasets of most projects for different styles cannot cover most styles. This study found that StyleShot has obvious advantages in the above aspects. StyleShot has a pre-trained style-aware encoder and a content fusion encoder, as well as a balanced style dataset (StyleGallery) that integrates datasets such as JourneyDB [3], WikiArt [4] and so on, which can be useful in the aspects of character design, scene generation, and dynamic filters in game development. generation, and dynamic filters in game development.

2 Advantages

2.1 Efficient

StyleShot is able to generate characters and scenes directly based on reference styles, eliminating the need to train separate models for each style like most projects. For games with multiple styles, the application of StyleShot can greatly improve the efficiency. The running time cost of StyleShot and other SOTA style transfer methods[1](e.g., StyleDrop [5], DreamBooth) is analyzed through an experiment, which sets the inference step of the diffusion-based method to 50, and refers to part of the experimental data of the experiment conducted by StyleDrop to obtain the cost of the other SOTA style transfer methods (e.g., StyleDrop, DreamBooth) [5]. As shown in Table 1, SOTA style transfer methods (e.g., StyleDrop, DreamBooth) need to be fine-tuned for each style during testing, which is computationally expensive, whereas StyleShot can directly generate the target styles without spending extra time on training.

Table 1: Running time cost between StyleShot and other style transfer methods (StyleDrop, DreamBooth)

Project	StyleShot	DreamBooth	StyleDrop
Training	-	371s	302s
Inference	5s	5s	7s

2.2 Excellent style generation performance

In the process of game development, character design is always an indispensable part, and a good character design can inject the soul of the game itself. Through comparative tests, this study can find that whether it is text-driven generation or style migration, StyleShot's character generation results are far ahead of other projects, so in games that rely on characterization, the results made by StyleShot can meet the requirements very well.

To explain why StyleShot has such excellent style generation results, it is necessary to compare the style-aware encoder mentioned repeatedly in the previous section with Contrastive Language-Image Pre-Training (CLIP) [6].

The core of CLIP is to establish semantic alignment of images with text, and thus its encoder focuses on extracting global semantic features (e.g., object class, scene type). Its processing method is to perform single-scale image chunking and extract image features through a uniform depth transformer layer, which leads to the consequence of mixing semantic and stylistic information, making decoupling difficult (This process is visible in Figure 2). Meanwhile, as shown in figure 1, according to Raphi Kang et al, it is proved that CLIP can not accomplish both represent basic descriptions and image content, represent attribute binding, represent spatial location and relationships, represent negation any two of the four tasks at the same time, making CLIP's latent space fall short in handling complex visual-textual interaction tasks [6].

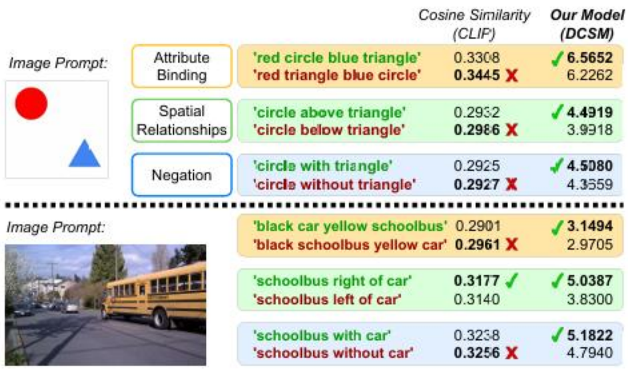

Figure 1: The semantics of textual semantic cues are not accurately reflected by CLIP score [6].

An experiment is set up here in which paintings by Monet, representing the Impressionists, are fed into the CLIP encoder to visualize their attentional heat maps [1]. Impressionist paintings are characterized by their focus on light and shadow expression techniques, without very clear shadows or prominent or flat painted contour lines, thus it is a great challenge for the encoder to recognize local details when generating the attention heat map. The experimental results in Figure 3 will show that the attention of CLIP will focus more on the more semantically significant regions and less on the style elements. When dealing with some abstract painting styles, it may result in focusing on some “strokes that look like the shape of an object”, for example, a curve may be recognized as a cloud, while ignoring such style features as color gradient and stroke details.

Styleshot is able to avoid this problem by analyzing its style encoder, and it can be found that two key techniques play a significant role.

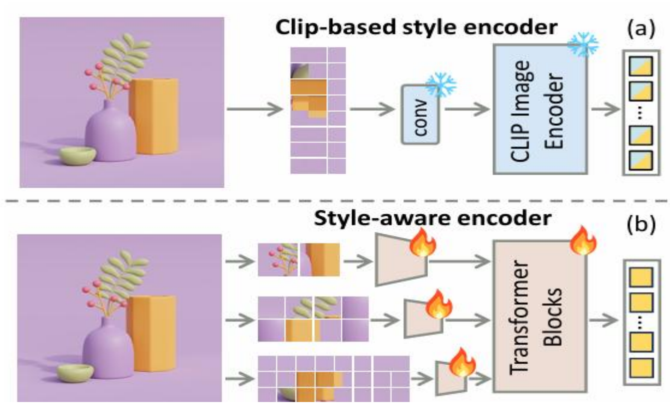

Figure 2: Styleshot's style-aware encoder vs. CLIP encoder for processing image details [1].

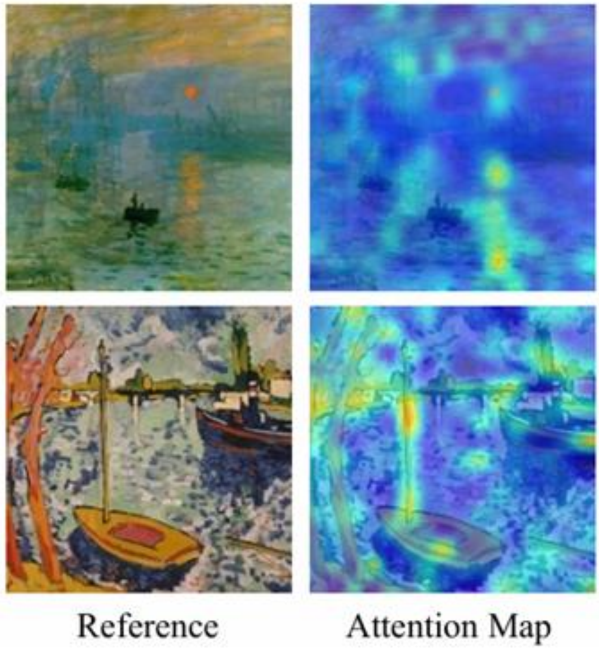

Figure 3: Attentional heat maps of CLIP encoders in a particular style of image [1].

The first one is multi-scale patch segmentation. Different from CLIP which focuses on the extraction of global semantics, StyleShot is able to divide the image into local regions of different scales to capture low-order styles (e.g. colors, strokes, etc.) and higher-order styles (e.g. layouts, textures, etc.) respectively. The extraction of low-order styles ensures the similarity of the overall style, which sets the overall style performance, while the extraction of high-order styles polishes the details of the generated image, avoiding the situation of “only visible from afar, but not close to the eye”, and meeting the needs of the game industry, which has increasingly high requirements for details.

The second is the Mixture-of-Expert (MoE) [7], whose most notable feature is its ability to efficiently scale the model parameters while conserving computational resources as much as possible. MoE consists of a gate and a pool of experts, through which experts can be dynamically selected to process the input data. Specifically in StyleShot, StyleShot utilizes this feature to assign different Resblocks to different scales of the above segmented patches for style extraction. Therefore, in the process of game development, MoE can assign exclusive generation models for different content types (terrain, props, etc.) to improve generation efficiency and quality.

2.3 Diversified Data Sets

In addition to the encoder and other reasons mentioned above, StyleShot can perform well in a variety of styles, and there is also a good improvement in the dataset, which solves the previous dataset in the face of multi-style tasks because of the uneven distribution of styles, resulting in easy to over-fit the high-frequency styles of the pain point. The effect of style migration is closely related to the style distribution of the training dataset, and the dataset based style migration project is like the foundation based building, which affects the overall effect of the style migration “high-rise”. StyleShot includes datasets such as MidJourney [8], WIKIART [4], and LAION-Aesthetics [9], and also incorporates rarer painting styles (e.g., pointillism and ink painting, etc.), combining the strengths of these datasets while avoiding the pain points that exist to make this dataset more diverse.

In addition to the advantage of an even distribution of styles, there is another advantage that provides great convenience for developers in the usage phase. 99.7% of the images in StyleGallery are labeled with style descriptions [1], which paves the way for developers in the following ways.

The first and most direct benefit is the semantics-driven style generation, where the developer defines the style through tags and generates the desired result through text-to-image generation techniques. In addition to this, it improves the efficiency of the development process by eliminating the need to conceptualize image details from scratch and focusing only on defining the target style, making the creative process more efficient.

Secondly, the large number of fine-grained painting styles from WIKIART enables precise and innovative style migration. Developers can freely choose between the “heavy brushstrokes of the oil painting style” and the “delicate light and shadow of the stippling style”, making the details of the game screen more precise. In addition, it is also possible to generate new images by combining multiple styles, such as combining the cyberpunk style, which has been very popular in recent years, with various places of interest around the world to create unique visual effects.

2.4 De-stylization

StyleShot, as a style migration project, as is shown in Figure 4, is also excellent in de-styling, avoiding style migration to interfere with game interaction elements. It can better satisfy different players to customize the function without losing the image content. Such a result is the result of the aforementioned content style encoder, which can realize the de-stylization process, using HED edge detection to extract the structural information of the screen, combined with a style-aware encoder to strip the original style of the image [1]. Based on these features, it can be tried to be used in dynamic game filters.

De-stylization emphasizes the separation of style information (e.g., artistic style features) from content information (e.g., structures, objects, etc., in the image). This is similar to the design concept of game filters, where the “style” of the game screen is adjusted independently (e.g., turning the screen into a watercolor effect) without changing the core content (e.g., the structure of the characters and scenes).

This separation supports real-time filter application, allowing players or developers to quickly switch or overlay different visual styles without affecting the game's content. If for any reason you must use mixed units, the units used for each quantity in an equation must be stated.

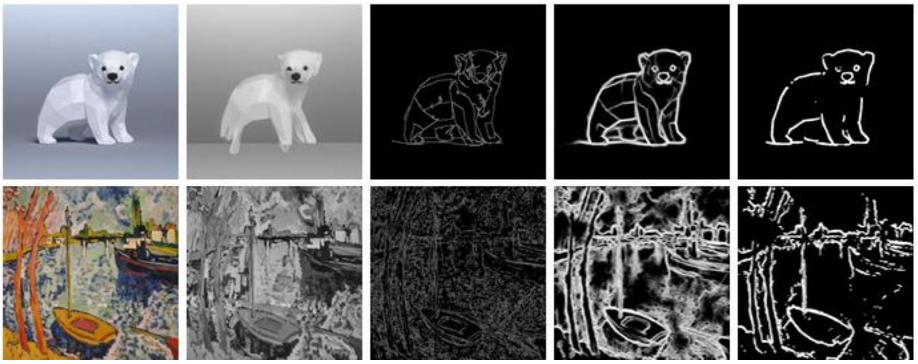

Figure 4: Performance of individual projects on the de-stylization task [1].

3 Challenges and Prospects

3.1 The Challenge

StyleShot relies on deep learning models for style migration, and its current operational efficiency can already support higher-quality offline generation, but further optimization is still needed to achieve real-time style transformation in games. For example, how to complete complex style extraction and content fusion under limited GPU resources is a key challenge for real-time applications.

Visual styles in games usually need to be adapted to different scenarios, e.g., from 3D rendering to 2D cartoon, or even minimalist abstract styles. how StyleShot can further expand its style libraries and maintain the ability to adapt to extreme styles is the task it faces.

Players expect to be able to adjust the visual style of the game in real time, which requires the development of flexible user interfaces and algorithms that allow players to customize their own visual experience through simple actions. This places higher demands on StyleShot's user experience design and real-time algorithms.

3.2 The Prospect

Combining StyleShot's style-aware encoder and content fusion encoder, an AI-based real-time filter system can be developed to provide players with dynamic switchable multi-style visual experiences, such as automatically applying darker tones in horror scenarios or switching to pastel colors in fairy tale worlds.

StyleShot supports game developers to quickly design and generate diverse art styles, reducing art production costs. For example, the StyleShot library can be used to quickly generate different artists' styles for character design, scene layout, or special effects [10].

4 Conclusion

Uttam Kokil conducted a survey using the AttrakDiff questionnaire and found that game participants found high visual aesthetic quality more appealing. That's why this study chose StyleShot that excel in terms of visual effects and more. This study focuses on the potential advantages and application value of StyleShot in game development by comparing multiple style migration projects. StyleShot demonstrates excellent performance in generating images with rich and broad stylistic expressions based on its style-aware encoder, content fusion encoder, and the well-organized StyleGallery style dataset. Compared to other approaches, StyleShot is able to accomplish the task of generating multi-style images quickly and efficiently without the need for style fine-tuning during testing, significantly improving development efficiency.

In addition, StyleShot's de-stylization technique effectively separates the style and content information of an image, providing powerful support for the development of dynamic game filters. Combined with the HED edge detection algorithm and multi-scale style extraction technology, StyleShot has unique advantages in detail processing and style consistency in complex scenes. Its style dataset StyleGallery not only covers a wide range of art styles, but also improves the expressive ability of style migration through fine annotation and balanced distribution, which provides great convenience for game developers in the fields of character design, scene generation and dynamic special effects.

In the future, with the advancement of real-time performance optimization technology, StyleShot is expected to play a greater role in real-time style migration and personalized visual experience in games, bringing more immersive and diversified gaming experiences to players and developers. Overall, StyleShot's combination of artistic creativity and technical feasibility lays a solid foundation for next-generation game development tools and demonstrates its broad development prospects in the game industry.

References

1. Gao, J., Liu, Y., Sun, Y., Tang, Y., Zeng, Y., Chen, K., and Zhao, C.: ‘StyleShot: A Snapshot on Any Style,’ arXiv preprint arXiv:2407.01414, 2025
2. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., and Aberman, K.: ‘DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation,’ in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2023, pp. 22500-22510
3. Sun, K., Pan, J., Ge, Y., Li, H., Duan, H., et al.: ‘Journeydb: A benchmark for generative image understanding,’ Advances in Neural Information Processing Systems, vol. 36, pp. 49659–49678, 2023
4. Phillips, F., and Mackintosh, B., ‘Wiki art gallery, inc.: A case for critical thinking,’ Issues in Accounting Education, vol. 26, no. 3, pp. 593–608, 2011
5. Sohn, K., Ruiz, N., Lee, K., Chin, D. C., Blok, I., Chang, H., Barber, J., Jiang, L., Entis, G., Li, Y., et al.: ‘Styledrop: Text-to-image generation in any style’, arXiv preprint arXiv:2306.00983, 2023
6. Kang, R., Song, Y., Gkioxari, G., and Perona, P.: ‘Is CLIP ideal? No. Can we fix it? Yes!’, arXiv preprint arXiv:2503.08723, 2025
7. Cai, W., Jiang, J., Wang, F., Tang, J., Kim, S., and Huang, J.: ‘A Survey on Mixture of Experts in Large Language Models,’ IEEE Transactions on Knowledge and Data Engineering, pp. 1–20, 2025
8. Jaruga-Rozdolska, A.: ‘Artificial intelligence as part of future practices in the architect’s work: MidJourney generative tool as part of a process of creating an architectural form,’ Architectus, no. 3 (71), pp. 95–104, 2022
9. Burger, L. J.: ‘Laion: Image data, ai, and dispossession.’ MA New Media & Digital Culture. 2023
10. Kokil, U.: ‘The Impact of Visual Aesthetic Quality on User Engagement during Gameplay,’ in ACHI 2018: The Eleventh International Conference on Advances in Computer-Human Interactions, 2018, pp. 159–164