Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU34230

Generative AI in 2D Animation: Applications, Challenges, and Future Prospects

Miaoyu Wang

Golden Apple Jin Cheng No.1 Secondary School, Chengdu, China
wangmiaoyu111@outlook.com

Abstract. Generative AI has been developing quickly in recent years and is able to influence different stages of 2D animation production. This paper reviews the current applications of generative AI in 2D animation from the perspective of computer graphics, its main challenges, and future prospects. The paper discusses how generative AI supports pre-production visual development, which includes concept art, character design, storyboard, and layout generation. It also reviews its applications in motion generation and frame synthesis, which include pose-guided animation, in-betweening, and animation sequence generation. The paper also analyzes its role in visual refinement and end-to-end generation, which includes colorization, style control, and integrated animation production systems. This paper identifies three challenges: limited controllability and interpretability, weak temporal and stylistic consistency, and limited applicability in real animation workflows; it outlines three future directions: more controllable generation frameworks, improved long-range consistency and style preservation, and better integration into artist-centered production pipelines. Overall, generative AI shows strong potential as a helpful tool in 2D animation production, but it still cannot fully replace human artists in the whole workflows.

Keywords: Computer graphics, diffusion models, generative AI

1. Introduction

As not only a form of art but also a large-scale industry with a highly complex and sophisticated production structure, animation, unlike many other visual arts, tends to be labor-intensive and often requires collaboration among numerous teams, companies, and even countries [1]. Usually, the whole process of producing an industrialized animation includes concept development, storyboarding, modeling, rigging, in-betweening, compositing, and post-production. In this case, the workflow usually involves numerous specialized stages, so the process becomes time-consuming and costly [1, 2]. Moreover, the animation industry is becoming increasingly globalized. Production is separated across different regions according to labor cost, technical skill, policy support, and market conditions. The scale and coordination demands constantly increase [1, 3]. As a result, animation studios are often faced with pressure from high labor expenses, long production cycles, and complicated management requirements. These structural characteristics help explain why the animation industry has continuously pursued more efficient production methods and technological support [1-3].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU34230

At the same time, in recent years, generative AI, especially the diffusion-based models, for video has developed rapidly [4]. Compared with earlier generations, the visual quality, controllability, and text-guided content creation of recent video generation have been largely improved [4]. This rapid development has shown the potential power of generative AI in visual media production. The possibility of expanding generative AI to the huge animation industry has drawn increasing attention from animation artists and researchers. Recent studies suggest that such technologies may offer practical support for animation production and may gradually take over some traditionally manual procedures in the pipeline [5]. As these models continue to improve, generative AI is increasingly being viewed as a promising technical direction for reshaping animation workflows and improving production efficiency [4, 5].

Therefore, it is necessary to have a review of generative AI in animation. Although recent research has presented growing interest in applying AI to animation, the related findings are separated and scattered. As a result, there is still a lack of a clear overview of where generative AI can be applied in animation, where it has already shown practical value, and what future potential it may hold for the industry. Carrying deep concern for this thriving artistic and cultural industry, people hope to see AI inject new vitality into it, free creators from repetitive manual labor, and promote greater creativity and diversity. A summary of these can organize and help researchers and artists better understand the direction of development of generative AI in animation production.

2. The application of generative AI in 2D animation

Generative AI can assist with specific tasks or combine several steps in the whole system. It has begun to have an influence on different stages in 2D animation production. AI can participate in the process that includes visual development, motion generation, and post-production refinement. So this section will talk about some applications of Generative AI: pre-production visual development, motion generation and frame synthesis, and visual refinement and end-to-end generation.

2.1. Pre-production visual development

At the first pre-production stage, Generative AI helps people visualize descriptions and concepts. It can help in developing concepts, designing characters, drafting storyboards, and planning layout. According to recent studies, Midjourney and Stable Diffusion, a type of text-to-image model, can generate as follows: many character images to present different appearances, haircuts, body size, or dressing; settings like where the character was from, what temperament the character has, and how the character can involve the story; and atmosphere designs such as the landscapes scenes or interior designs [5]. Moreover, AI-assisted systems for character design help team members communicate more efficiently because they have already had previous concept input that provides visual feedback to refine details [6]. All of these developments can reduce labor and time costs of early exploration.

Furthermore, generative AI can also support drafting storyboards. Make-A-Storyboard, a framework combining large language models with image generation methods, lets story texts become continuous story scenes and shots. This improves character consistency and scene correlation [7].

In the Layout generation mentioned before, the application of generative AI is still limited. As the stage after storyboarding in traditional animation production, it requires artists to guarantee that the perspective and spatial arrangement are consistent. But AI is gradually extending to this field [5, 7].

DOI: 10.54254/2755-2721/2026.GU34230

2.2. Motion generation and frame synthesis

In traditional workflows of 2D animation production, motion generation often includes stages as follows: key animation to decide the most important gesture and position, motion refinement that improves the quality of line and final movement of all the objects, and in-betweening to fill all the gaps between any key animation. It is extremely costly because it requires a large number of laborers and artists to draw pictures, where audiences may not even pay much attention, in the final animation. In this case, generative AI is valuable because it helps transform drafts or static images into animated pictures. This can reduce the labor required to fill the gap between key poses and motion.

One important direction for generative AI is that it can generate animation by gesture's tendency and images to indicate and predict the next step of motion. When artists already have a character design sketch, which indicates the characters' appearance and special facial expression, or keyframes, which present the main motion and gestures of each shot, they can use models as follows to explore more movement possibilities. Animate Anyone shows that a static character image can be animated consistently and continuously by gesture control. Even though it is not developed only for 2D animation, the model of Animate Anyone is one of the practical ways for generating fluent motion in animation [8]. Moreover, AniSora can turn images into videos based on the information in the prompt and in the images, blend different frames that improve the fluency of the video or create the sense of motion, and use image patches to guide animations' motion. It shows that AI can both pose the characters' gestures and help to create short animated scenes [9].

The application of generative AI in the stage of in-betweening is the most practical and helpful one in recent animation production, for in-betweening requires relatively little of artists' skills and animation understanding but more of a huge number of laborers to finish this repetitive and massive work. For example, ToonCrafter focuses on exaggerated and nonlinear motion and dis-occlusion, which is valuable in producing stylized animation [10]. Rather than only realistic human videos, AnyMoLe also assists motion of in-betweening to combine intermediate motion for animation characters [11]. In addition, if sparse line-art is input, AniDoc, working as a video diffusion model, can automatically connect these sequences [12].

To sum up, these models indicate that generative AI is beginning to assist directly in the motion-production stage of 2D animation.

2.3. Visual refinement and end-to-end generation

Visual refinement mainly relates to colorization, cleanup, stylistic unification, and compositing. One important application of generative AI is in this stage is automatic colorization. AnimeDiffusion presents that diffusion models can assist anime line drawing colorization. Artists are able to quickly fill color to line art while maintaining reference-based visual coherence [13]. In addition, AI can also contribute to refining and consistent style. VToonify, for example, shows that neural style transfer can generate consistent stylized video outputs, suggesting its potential value [14].

In addition to replacing or assisting with some of the single steps in 2D animation production, some recent systems can combine multiple stages together. For instance, ToonComposer and AniDoc can combine in-betweening and colorization into a single stage after keyframes, which means that artists only need to draft line-art keyframes, reducing the need for operations in the later production process [9, 15]. Moreover, AniSora can transform images into videos and insert frames to help with end-to-end animation generation [10].

DOI: 10.54254/2755-2721/2026.GU34230

3. Discussion

However, even though generative AI has potential in 2D animation production, it still faces several limitations. The application of generative AI in 2D animation production is confronted with challenges that limit its use and the directions for future research that may be developed, so in this section, there are three major issues: controllability and interpretability, temporal and stylistic consistency, and workflow applicability, and outlines the corresponding future prospects.

3.1. Challenges

3.1.1. Limited controllability and interpretability

Limited controllability is a major challenge in 2D animation production. Although current models can generate plausible motion, artists often cannot precisely control how to transform the pose, what could be local details, or how to compose shots. The output is difficult to predict and revise. This problem is significant in animation that requires fine-grained and consistent control on character performance and visual planning [11, 16].

3.1.2. Weak temporal and stylistic consistency

Another major problem of generative AI in 2D animation is weak temporal and stylistic consistency. Many models can make single frames look good. But they often fail to keep textures, character looks, and visual style stable across long sequences. Prior work has identified flickering, low-level texture instability, and limited global temporal coherence as common problems in video generation and translation systems. This problem is more serious in 2D animation. In this field, lines, colors, and stylized motion need to stay coherent from frame to frame.

3.1.3. Limited applicability in real animation workflows

A further challenge is the limited applicability of many generative AI systems in real animation workflows. Current models can make strong visual results. But they often do not fit real production needs. Artists need to edit results, keep frames stable over time, control parts of the image, and use these tools in existing pipelines. Some previous studies of cel-animation and films have also shown that artists need to edit finely to refine the motion and increase the consistency of scenes, for industrial animation is a task for both visual aesthetics and workflow design, which means that in a large-scale cooperative project, all the crews need to well adjust to the application of AI in their work and smoothly integrate it into their work.

3.2. Future prospects

3.2.1. More controllable and interpretable generation frameworks

To address this limitation, future research is likely to focus on more controllable and interpretable generation frameworks.

One suggestion for future research to address this limitation is that they can focus on more controllable and interpretable generation frameworks. For example, while using generative AI, clearer guidance helps a lot. Artists can input as much as possible, like different pose control to clearly define the specific differences of objects and characters, reference images that present the

DOI: 10.54254/2755-2721/2026.GU34230

small features in detail, or other multi-condition inputs to indicate slight differences or important details. These methods can help generate motion match artistic intention better.

Another promising direction is that generative AI is able to present intermediate steps during the whole workflow, which means that artists are enabled to monitor, intervene, and edit in the middle stages during the generation process. These may help artists predict results and edit them more easily. These developments would help generative AI function less as a black-box generator and more as a reliable tool in animation production [11-18].

3.2.2. Improved long-range consistency and style preservation

To overcome this challenge, future research will likely focus on stronger long-range consistency modeling and more robust style preservation.

Existing studies already suggest several directions. For instance, local-global temporal strategies can be understood as keeping an eye on the details while also considering the big picture. Local includes such as the instantaneous small actions like a character blinking their eyes or a finger slightly shaking, and global includes such as the entire shot being pulled from a distance to a close-up, or a character moving from the left side of the frame to the right. Explicit spatial-temporal correspondence means that the AI can clearly know where this point in the frame was in the previous frame and where it should go in the next frame. A world-consistent generation framework enables the AI to always remember the basic laws of this world, such as gravity and occlusion relationships, and ensures that the animation remains stable and not clumsy from beginning to end [19-21]. For stylized animation, another important direction is that AI should be developed with the ability to preserve line and paint visual features, and it also needs to prevent flicker or drift. This can help generate stable outputs over time with a coherent visual style [14, 22].

3.2.3. Better integration into artist-centered production pipelines

For better integration into the artist-centered production pipelines, future research should make generative AI that is applied in 2D animation production no longer a fully automatic black-box generator. A more useful direction is to let artists correct results step by step, edit them during the process, and connect AI tools with existing animation practices. Recent work already points in this direction: AniDoc is designed to integrate with anime production pipelines; interactive systems such as MagicQuill show how human intention can remain central during the editing process. This shift may make AI more usable in professional animation settings.

4. Conclusion

In conclusion, generative AI has begun to be a effective and practical tool in several stages of 2D animation production. It can support pre-production visual development, motion generation, and visual refinement to help artists work faster and reduce repetitive work. However, it still has clear limits. Current generative AI often lacks strong control, stable long-sequence generation, and fitting real animation workflows well, so generative AI cannot fully replace human artists in professional 2D animation currently. In the future, researchers can focus more on the direction of controllability, consistency, and workflow integration. These areas will be important for the wider use of AI in animation production.

DOI: 10.54254/2755-2721/2026.GU34230

References

[1] Westcott, T. (2011). An overview of the global animation industry. Creative Industries Journal, 3(3), 253–259.
[2] Cole, A. (2008). Distant neighbors: The new geography of animated film production in Europe. Regional Studies, 42(6), 891–904.
[3] Yoon, H. (2017). Globalization of the animation industry: Multi-scalar linkages of six animation production centers. International Journal of Cultural Policy, 23(5), 634–651.
[4] Xing, Z., Feng, Q., Chen, H., Dai, Q., Hu, H., Xu, H., Wu, Z., & Jiang, Y.-G. (2024). A survey on video diffusion models. ACM Computing Surveys, 57(2), Article 41, 1–42.
[5] Tang, Y., Guo, J., Liu, P., Wang, Z., Hua, H., Zhong, J.-X., Xiao, Y., Huang, C., Song, L., Liang, S., Song, Y., He, L., Bi, J., Feng, M., Li, X., Zhang, Z., & Xu, C. (2025). Generative AI for cel-animation: A survey. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) (pp. 3837–3850).
[6] Ling, L., et al. (2024). Sketchar: Supporting character design and illustration prototyping using generative AI. Proceedings of the ACM on Human-Computer Interaction. https: //doi.org/10.1145/3677102
[7] Su, S., et al. (2023). Make-A-Storyboard: A general framework for storyboard with disentangled and merged control. arXiv.
[8] Hu, L., Gao, X., Zhang, P., Sun, K., Zhang, B., & Bo, L. (2024). Animate anyone: Consistent and controllable image-to-video synthesis for character animation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8153–8163).
[9] Jiang, Y., Xu, B., Yang, S., Yin, M., Liu, J., Xu, C., Wang, S., Wu, Y., Zhu, B., Zhang, X., Zheng, X., Xu, J., Zhang, Y., Hou, J., & Sun, H. (2025). Exploring the frontiers of animation video generation in the Sora era: Method, dataset, and benchmark. In Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI) (Paper 141).
[10] Xing, J., Liu, H., Xia, M., Zhang, Y., Wang, X., Shan, Y., & Wong, T.-T. (2024). ToonCrafter: Generative cartoon interpolation. ACM Transactions on Graphics, 43(6), Article 300, 1–11.
[11] Yun, K., Hong, S., Kim, C., & Noh, J. (2025). AnyMoLe: Any character motion in-betweening leveraging video diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[12] Meng, Y., Ouyang, H., Wang, H., Wang, Q., Wang, W., Cheng, K. L., Liu, Z., Shen, Y., & Qu, H. (2025). AniDoc: Animation creation made easier. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 18187–18197).
[13] Cao, Y., Meng, X., Mok, P. Y., Liu, X., Lee, T.-Y., & Li, P. (2024). AnimeDiffusion: Anime diffusion colorization. IEEE Transactions on Visualization and Computer Graphics, 30(8), 5311–5323.
[14] Yang, S., Jiang, L., Liu, Z., & Loy, C. C. (2022). VToonify: Controllable high-resolution portrait video style transfer. ACM Transactions on Graphics, 41(6), Article 246, 1–13.
[15] Li, L., et al. (2025). ToonComposer: Streamlining cartoon production with generative post-keyframing. arXiv.
[16] Ma, Y., et al. (2025). Controllable video generation: A survey. arXiv preprint arXiv: 2507.16869.
[17] Wang, et al. (2024). Follow-your-pose v2: Multiple-condition guided character animation. arXiv preprint arXiv: 2406.03035.
[18] Zhang, C., et al. (2024). VAST 1.0: A unified framework for controllable and consistent video generation. arXiv preprint arXiv: 2412.16677.
[19] Zhou, S., et al. (2024). Upscale-A-video: Temporal-consistent diffusion model for real-world video super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[20] Yang, S., et al. (2024). FRESCO: Spatial-temporal correspondence for zero-shot video translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[21] Ren, X., et al. (2025). GEN3C: 3D-informed world-consistent video generation with precise camera control. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
[22] Gong, Y., et al. (2025). FlowStyler: Artistic video stylization via transformation fields transports. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).