Interactive Storyboarding for Rapid Visual Story Generation

Sihyeon Jo[1], Zhenyuan Yuan[2], and Seong-Woo Kim[1]
[1]Seoul National University, Seoul, South Korea
[2]Penn State University, PA, United States
sihyeonjo@snu.ac.kr, snwoo@snu.ac.kr

2022 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) | 978-1-6654-6434-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCE-ASIA57006.2022.9954679

Abstract

Artificial Intelligence (AI) technologies have impacted almost every domain and its systems, including the entertainment industry. Although AI-based systems are expected to offer significant benefits in making content, it is still challenging to build a real-world AI application that can effectively contribute to content production. In this paper, we present a novel approach for developing a storyboard; a sequence of images displayed for pre-visualizing a motion picture, animation, motion graphic, or interactive media. We implement a prototype system, Gennie, that can interact with users and suggest AI-generated sketches for each scene of the storyboard.

Keywords: Human-AI Interaction, multimodal embedding, large-scale pre-trained model

1. Introduction

AI technologies are transforming the way we approach real-world tasks done by humans. Recent years have seen a surge in the research field of deep learning, where massive parameters are tuned to generalize on carrying out a particular task. For example, with the understanding of images, deep learning models have surpassed that of humans in several vision tasks [1, 2]. Besides, we are witnessing the possibility of AI applications for story generation [3], music composition [4], drawing [5], and so on. However, successful uses of deep learning algorithms in creative areas are raising the bar for required sensibleness and specificity, which are far below those of humans [6].

The storyboard is a sequence of drawings, typically with some directions and dialogue, that represents the shots planned for story products. The storyboard creation step is crucial in that storyboards serve as a visual road map during the story product development period. Storyboard creation is a difficult task even for professional artists, let alone novices, to simultaneously consider vital components of the storyboard such as subject, background, and point of view.

This paper represents user studies and visualizes the co-creation process of a storyboard. Given the text descriptions specified by the users, Gennie represents multiple draft sketches that match the story, and the users can get some inspiration or utilize the sketches for their story. Since draft sketches offered by Gennie have the flexibility and lack detail, users can utilize the sketches as blueprints for their sake as sources for the composition of objects or the final look of visual scenes.

2. Related work

Advance in AI technologies has opened up the possibilities of human-AI co-creation for drawing [7, 8], creative writing [9, 10], music composition [11], and video games [12]. For example, AI can create a half-sketched picture [13], write the next paragraph of the story [14], or add images to the design mood board. The key challenge in this range of previous tasks was to develop collaborative AI agents that could coordinate tasks based on users' goals and behaviors. To this end, some systems were designed to generate outputs according to the surrounding context of human-generated content, and some systems utilized user feedback to better match AI behavior to user intentions [15, 16].

AI-based systems offer potential benefits in making artworks or content [17, 18]; however, few of the promising applications of AI were produced without the proper engagement of humans. Instead, humans can collaborate with AI agents to achieve users' creative goals by getting some inspiration [19], gaining practical support in the progress, or enjoying the co-creation process itself. To integrate AI into the already-complicated human workflow, bringing the human-centered design philosophy into the computational interaction research is crucial. To deeply understand the user experience of human-AI co-creation in generating a storyboard, Gennie is implemented.

3. Proposed system

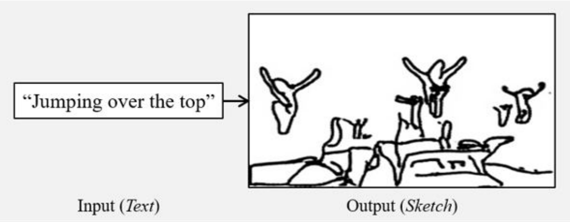


Fig. 1. The proposed system's inputs and outputs.

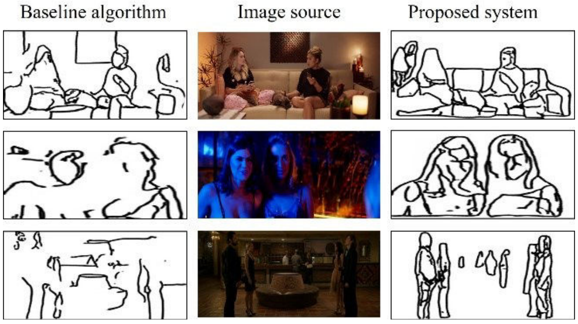


Fig. 2. Effects of segmentation masks in producing sketch results.

We develop an AI collaborator Gennie, and propose a co-creation framework for storyboard generation. Gennie consists of knowledge preparation, story-to-sketches retrieval, and user interface. This section first introduces how to collect and process the data for preparing a scene knowledge database and describes the story-to-sketches model.

We have prepared a knowledge base to provide sketches suitable for the user's creative goals. Inspired by the fact that movies provide abundant sources of scene knowledge, which can be utilized to compose storyboards, we collected about 300,000 captured images from the 7083 movie trailers and processed them with deep learning models to extract scene knowledge. Object detection and semantic segmentation modules based on Convolutional Neural Networks (CNNs) are applied to spot objects which could play essential roles in the plot. The image-to-sketch style transfer model composed of Generative Adversarial Networks (GANs) is adopted to generate sketches from images of segmented objects.

Before translating to the sketch, we post-processed an image with the segmentation masks. By extracting the overlapped regions between segmentation masks and the original image, we could clearly distill key object-grounded regions without uninformative objects or backgrounds. Moreover, we could automatically filter out noisy images with this process. For example, trailer scenes that only contains text descriptions, or just blank images which can be captured at scene change moments. After all, 25% of total images were considered as noises and removed.

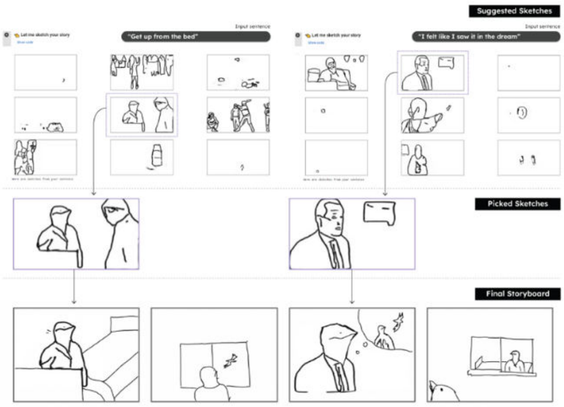


Fig. 3. A series of scenes generated with Gennie.

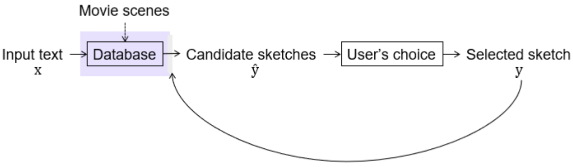


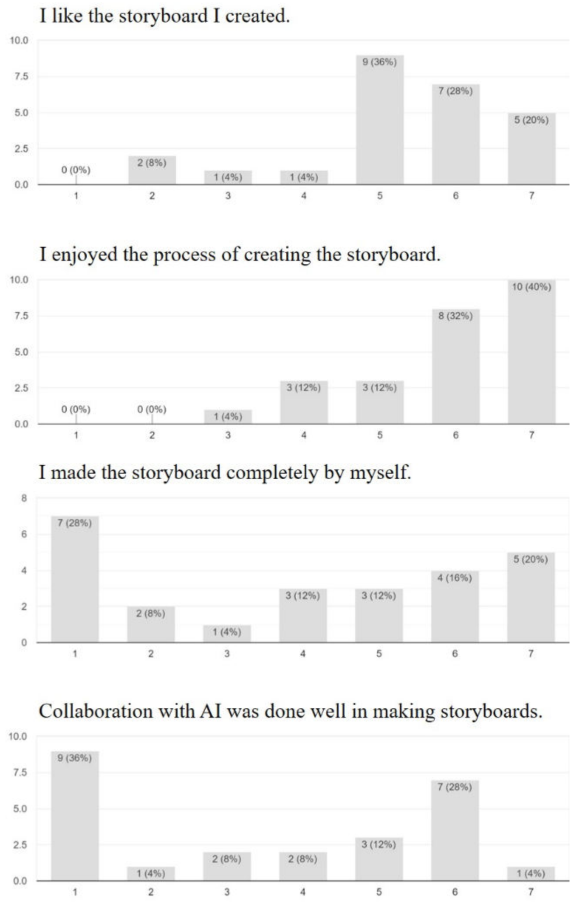
Fig. 4. A feedback loop to improve the proposed system.

Then we generated the contour drawings from refined images. We exploited pix2pix [20] as an Image-to-sketch translation model. Different from conventional edge or boundary detection algorithms, pix2pix predicts the salient contours in images and outputs as a familiar style which resembles human drawing sketches. We followed suggested methods from Photo-Sketching model [21]. As a result, we succeed in achieving plausible sketch outputs as shown in Fig. 2.

A large-scale pre-trained language model is employed to generate relevant sentences to users' inputs, while another large-scale pre-trained text encoder is utilized to match the sentences to scenes using the similarity scores calculated in the text-image co-embedding vector space. Given the text descriptions specified by the users, Gennie represents multiple draft sketches that match the story, and the users can get some inspiration or utilize the sketches for their visual story. Since draft sketches offered by Gennie have the flexibility and lack of detail, users can utilize the sketches as blueprints to suggest the final look of visual stories or adapt concepts as needed.

When a user inputs an input sentence, GPT-2 [22] generates a related sentence. This process is to suggest various ideas related to it from a single input. After generating several related sentences from the input sentence, the CLIP model [23] finds a co-embedding space between text and image. At this time, sketches are drawn for the calculated closest images and presented to the user. By utilizing the above Gennie engine, users can freely use the sketch presented by Gennie for creating their own stories.

4. Results


I like the storyboard I created.

I enjoyed the process of creating the storyboard.

I made the storyboard completely by myself.

Collaboration with AI was done well in making storyboards.

Fig. 5. The evalution of participants on Gennie.

The user evaluation is conducted by 25 participants after creating three storyboards with Gennie. Fig. 5. shows that the participants evaluate the storyboard creation process with Gennie as a satisfying experience. The proposed system has the potential to better engage the artists with the results for the questions in the first and second while the graphs in the third and fourth indicate that the co-creation process requires some time to adapt.

Users had mainly two goals leveraging the interactive storyboarding system: generating a new topic and drawing a specific scene. The users employed different strategies to achieve each goal. In order to achieve their task goal of creating a free-topic storyboard, participants used Gennie in different manners.

First, users utilize Gennie as a trigger to compose a storyline. Inspired by the proposed sketches, participants came up with interesting subjects and stories. Results from Gennie were used as various triggers for participants' ideations in this case. Once participants succeeded in getting some clues to start a story from Gennie, they went on their way to create storyboards without Gennie.

Second, participants used Gennie as a tool to sketch out scenes to represent their stories. When participants could not specify how they should sketch the scenes from their topics, they interacted with Gennie to get insights into how to fill their stories. Even when Gennie created unexpected results, participants did not ignore the results; they reflected the sketches on their stories.

The two behaviors above were not clearly divided, as participants used Gennie for diverse needs. When the users anticipated Gennie's outputs depict high-level concepts, participants used the engine as a casual partner in the ideation process. In contrast, when participants expected more specific results from Gennie, users added more specific input.

Drawing a storyboard is crucial to planning any form of visual narrative. The composition of the objects in the frame and the point of view created by the angle can significantly enhance or alter how the viewer understands the story. The AI agent Gennie is developed in this regard to help extend the limits of an individual's imagination. Gennie is not only a novel invention but also a potential partner in collaboration with the user.

5. Conclusion

In this work, we focused on storyboard generation, which has multi-modal characteristics for visual storytelling. We developed an AI system for storyboard co-creation with users. Several deep learning models are effectively incorporated to implement a user-friendly and practical system. Several implications for AI systems for storyboard co-creation are discussed.

Acknowledgements

This work was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0008747) supervised by the Korea Institute for Advancement of Technology (KIAT), and in part by the National Research Foundation of Korea (NRF) through the Ministry of Science and ICT under Grant 2021R1A2C1093957. The Institute of Engineering Research at Seoul National University provided research facilities for this work.

References

[1] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[2] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[3] Jain, Parag, et al. "Story generation from sequence of independent short descriptions." arXiv preprint arXiv:1707.05501. 2017.

[4] Jaques, Natasha, et al. "Generating music by fine-tuning recurrent neural networks with reinforcement learning." 2016.

[5] Xu, Peng, et al. "Deep learning for free-hand sketch: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022.

[6] Adiwardana, Daniel, et al. "Towards a human-like open-domain chatbot." arXiv preprint arXiv:2001.09977. 2020.

[7] Sun, Lingyun, et al. "SmartPaint: a co-creative drawing system based on generative adversarial networks." Frontiers of Information Technology & Electronic Engineering 20.12: 1644-1656. 2019.

[8] Davis, Nicholas Mark, et al. "Co-creative drawing agent with object recognition." Twelfth artificial intelligence and interactive digital entertainment conference. 2016.

[9] Bensaid, Eden, et al. "Fairytailor: A multimodal generative framework for storytelling." arXiv preprint arXiv:2108.04324. 2021.

[10] Ratawal, Yamini, et al. "PoemAI: Text Generator Assistant for Writers." Intelligent Sustainable Systems. Springer, 575-584. 2022.

[11] Louie, Ryan, et al. "Novice-AI music co-creation via AI-steering tools for deep generative models." Proceedings of the 2020 CHI conference on human factors in computing systems. 2020.

[12] Kim, Seung Wook, et al. "Learning to simulate dynamic environments with gamegan." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[13] Li, Mengtian, et al. "Photo-sketching: Inferring contour drawings from images." 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019.

[14] Rashkin, Hannah, et al. "Plotmachines: Outline-conditioned generation with dynamic plot state tracking." arXiv preprint arXiv:2004.14967. 2020.

[15] Karimi, Pegah, et al. "Relating cognitive models of design creativity to the similarity of sketches generated by an ai partner." Proceedings of the 2019 on Creativity and Cognition. 259-270. 2019.

[16] Kim, Kyungsun, Jeongyun Heo, and Sanghoon Jeong. "Tool or Partner: The Designer's Perception of an AI-Style Generating Service." International Conference on Human-Computer Interaction. Springer, Cham, 2021.

[17] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[18] Park, Taesung, et al. "Semantic image synthesis with spatially-adaptive normalization." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[19] Sbai, Othman, et al. "Design: Design inspiration from generative networks." Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018.

[20] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[21] Li, Mengtian, et al. "Photo-sketching: Inferring contour drawings from images." 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019.

[22] Radford, Alec, et al. "Language models are unsupervised multitask learners." OpenAI blog 1.8: 9. 2019.

[23] Li, Manling, et al. "Clip-event: Connecting text and images with event structures." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.