Abstract

The cartoon and animation industry has a long history of using hand-drawn techniques to create visually appealing and engaging content. However, the production of traditional cartoons and animations can be labor-intensive and time-consuming, requiring significant human resources to draw and color each frame (key animation, or *genga*, in Japanese). This can limit the overall production capacity of the industry and make it challenging to meet the demand for high-efficiency cartoon editing to deliver new content from existing works. The new contents could be the HD remastering of old classic animations, 3D stereoscopic remakes, editing of the characters and the backgrounds, or a retargeting to mobile-friendly vertical short-video format (e.g., for TikTok.) This project intends to propose a deep learning-based system to implement an automatic analysis and editing pipeline for cartoon animations. The proposed system shall be able to learn from large amounts of data from cartoons and animations to analyze and recognize the characters, subjects, and objects depicted in the content as well as their movements and interactions. With the learned knowledge, the proposed system is further designed to empower automation in multiple anime editing tasks that are currently done manually, such as re-shading, character and background modification, depth analysis, and scene compositing. More importantly, the system does not require any key animations, which will allow direct processing and editing for any arbitrary animations with minimum human effort.

In implementing such a system, several challenges will need to be addressed. First, obtaining sufficient high-quality data from cartoons and animations with fine-grain annotation of the objects and their motions shall be necessary. Second, we propose using a cross-modal transformer model with strong generalization and expression abilities and few-shot learning capabilities to understand and interpret the content of the cartoons and animations in conjunction with their motion information. To the best of our knowledge, there are no existing solutions to this task, due to the stylized and exaggerated nature of cartoons and the lack of relevant data. We believe this approach can outperform traditional key animations in terms of accuracy and speed. More importantly, the few-shot learning ability of transformers should make it effective to ground any arbitrary objects even without prior knowledge. Finally, effectively using the information obtained from the model to achieve downstream tasks is a challenge that has not been researched before. Our goal is to identify the critical, labor-intensive steps in the downstream editing applications such as retargeting, stereoscopic remastering, and text-guided object editing. Then, we will incorporate object and motion information from the previously learned model into deep learning solutions to automate these critical steps to improve overall efficiency and precision in cartoon editing tasks.

In the final stage of this project, we plan to integrate the proposed system into the existing workflow of the cartoon and animation industry. To do this, we shall develop a user interface that allows users to input the desired processing tasks and parameters, and the system will automatically apply the appropriate algorithms and techniques to achieve the desired results. We will also set performance goals for the system in terms of accuracy, speed, and efficiency, and will optimize the system architecture and algorithms to meet these goals. This proposed system is expected to significantly streamline the process of updating and

repurposing existing cartoon and anime works, making it easier and more cost-effective. In addition, the pipeline could help to democratize the creation of high-quality cartoons and animations by enabling individuals and small teams to produce professional-grade content, fostering a more vibrant and diverse anime community. Besides the impact on the cartoon industry and the community, this project should also benefit the research community by advancing state of the art in computer vision, computer graphics, and natural language processing. It should also provide valuable learning opportunities for institute students studying digital media and AI technologies.