

Title of Project

English: Instance-aware Cartoon Stylization of Photo and Videos

Chinese: 基於實例感知的相片與視頻卡通風格化

Abstract of Research Comprehensible to a non-specialist

Cartoon is one of the most widely used and popular non-photorealistic media forms. Starting from the 19th century, cartoon has been gaining wide popularity due to its iconic exaggerated drawing style and its freedom in storytelling and expressing the artists' ideas. Recently with the development of the mobile web and social media, people tend to share their thoughts and ideas with more creativity and uniqueness. For example, after Spiderman's movie release, the audiences may want to represent their own life as comic strips and wish to have their own life stories illustrated in the style of the *Spiderman* comic by converting their captured photos and videos into comic frames. On the other side, professional cartoon and animation productions also have a specific need to create cartoon frames based on existing captured photorealistic scenes and characters. Unfortunately, current approaches to render natural photos and videos into cartoon styles usually require expertise in computer graphics pipelines, knowledge of art, and intensive human labor, which limits the further growth of these cartoon-related applications.

This project intends to propose a fully automatic framework to tackle the cartoon stylization task. The proposed framework can apply visually aesthetic cartoon styles to a diverse set of user-specified input, including both photographs and videos. The realization of the framework meets several challenges: firstly, because of the free-from sketch basis and the arbitrary shading styles in cartoon production, the characters, objects, and scenes must be rendered with the unique exaggerated and abstract cartoons styles. However, simple image filtering cannot achieve this task because it can hardly understand the image content and apply cartoon-style image editing. Besides, the conversion should still retain the original image content so that people can still recognize the content after stylization. Yet, unfortunately, no existing style transfer nor conditional image generation frameworks can solve this challenging problem. Finally, when processing video contents, the framework should also take care of the temporal coherence to make the result looks smooth and pleasant during playback.

Our preliminary approach to achieve cartoon stylization is constructing an unsupervised cross-modal understanding and synthesis between natural contents and cartoons with an *instance-level awareness* through deep learning techniques. Specifically, the term *instance* means the stylization is on an object-level basis. The instance-wise conversion also matches the real-life cartoon stylization pipeline by the artists. To build up our framework, we first extract object instances from raw photo images and cartoons and facilitate a *feature alignment network to learn a cross-modal feature alignment with an awareness of the underlying object category* (e.g., humans, cats, moving objects, etc.) After the input instance-level features are aligned, we propose to synthesize the features into stylized cartoon instances with a *cartoon synthesis network*. Afterward, we introduce a *canvas refinement network* to composite the individual objects into a single canvas. The network learns to enhance visual harmony among these converted instances and (optionally) proceeds to output temporal-smooth video sequences.

After the framework development, we shall evaluate the expressivity and the visual quality of

our outputs by a comprehensive subjective study from both casual users and experienced artists. We shall build software toolkits and API interfaces for tighter integration with existing creative software studios and mobile applications. We believe the proposed framework helps provide a more effortless and innovative way for the public to express their inspirations and lifestyles. Moreover, we also expect the proposed framework to benefit the whole cartoon and animation industry by cutting off the tedious visual design and asset preparation procedures. This project's potential outcome, including the deep learning algorithms, interactive software, and academic publications, will help motivate the computer graphics and deep learning community. At the same time, we hope that this project can continually contribute to developing students' and teachers' skills in digital entertainment and AI technologies of the institute.