Abstract

In recent years, image generation technology has evolved from simple encoder-decoder structures to adversarial generative networks and diffusion models. With advancements in technology, both the quality and controllability of generated images have improved, enabling the production of high-quality and diverse images.

In the fields of artistic creation and visual design, image generation technology reduces the time and effort required for production while also providing significant support. Among these advancements, research on shading generation from sketches has gained increasing attention. However, most existing methods for shading generation from sketches assume a single-light-source environment and are not capable of handling shading under multiple light sources. One major challenge is the difficulty in appropriately capturing the interactions between multiple light sources. Furthermore, due to the insufficient recognition of sketch structures, incorrect shading is often generated in empty regions both inside and outside the sketched objects.

To address these issues, this study proposes a diffusion-based shading generation method that considers multiple light sources in a virtual 3D space. By incorporating mask generation using the Segment Anything Model (SAM), the proposed method clarifies the regions where shading should be applied, thereby suppressing unintended shading generation outside the target areas. Additionally, a light embedding module is designed to integrate light source information into the diffusion process. By utilizing ConvNeXtBlock, this method ensures that light source information is appropriately embedded, thereby improving the accuracy of shading generation under multiple light sources.

Experimental results demonstrate that, in the case of a single light source, the proposed method reduces the total area of incorrectly generated shading in empty regions by 54.8% compared to the existing method, ShadeSketch. Furthermore, the similarity between the generated shading images and ground truth images improves by 3.4% over ShadeSketch. The proposed method is also adaptable to multiple light sources. In experiments involving two light sources, the average SSIM reached 0.897, indicating that the method successfully generates shading that reflects multiple light sources.

As a future direction, improving the encoding format of light sources remains an important challenge. The current method projects light source information onto predefined labels and controls image generation through label embedding. To enhance generalization and scalability, further investigation into encoding methods that incorporate additional light source attributes, such as type and intensity, is required.