Creating Visual Cognitive Illusions

潘致辰
2300011090
2300011090@stu.pku.edu.cn

王子骁
2300012802
2300012802@stu.pku.edu.cn

邹宜轩
2300012967
2300012967@stu.pku.edu.cn

1 Abstract

Visual cognitive illusions utilize high-level human cognitive processing, leading to the same image being interpreted differently through specific transformations. This project is based on the existing Diffusion-Illusions framework. It has successfully reproduced the key visual results shown in the paper and generated result images using our own prompts. Furthermore, this project has extended the original framework to support image prompts as well as mixed image-text prompts; it has added illusion types, including camouflage illusions and the kaleidoscope effect in rotation overlay illusions; and it has added some functions to visualize the result images as GIFs.

2 Introduction

Creating visual cognitive illusions aims to generate "illusion images" that can cause viewers to perceive differently based on given textual or pictorial prompts, under different perspectives and organizational arrangements.

Through this task, we can better understand the human cognitive mechanism for images. Additionally, based on the images and text provided by users, we can process them into cognitive illusion images to create impressive visual effects.

This project has extended the Diffusion-Illusions framework to support multiple types of prompts, including images and text. It has added illusion types, such as camouflage illusions (where one image is disguised within another) and the kaleidoscope effect in rotation overlay illusions (supporting rotation at various angles). Additionally, it has introduced the functionality to visualize the result images as GIFs.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

3 Related Work

Works related to this project include the Diffusion-Illusions model[1]and the Illusion Diffusion HQ model[2].

Compared with the Diffusion-Illusions model, this project is an improvement based on it, supporting more types of prompts and illusions, as well as visualization of the result images.

The Illusion Diffusion HQ model only supports camouflage illusions and only supports a combination of one image and one text as prompt. This project supports more types of illusions and also supports using only texts or only images as prompts.

4 Data

This project does not use a dataset for training, but directly utilizes the large model stable-diffusion-v1-4 provided in the code repository.

5 Methods

5.1 Image Prompts

Our implementation is as follows:

For the input image prompt, we convert it into a tensor and pass it as a parameter to the train function of the StableDiffusion class. During the diffusion process, we also encode the image prompt into the latent space through a VAE to obtain prompt-latents. Then, during the noise addition process, we calculate the residual between the latents of the predicted image and the prompt-latents, and add this residual to the original noise. This noise then goes through a denoising process. As a result, in the latent space, the feature representation of the resulting image will tend towards the prompt-latents, so the final output image will have the characteristics of the image prompt in terms of visual effects.

At the same time, since we do not change the original handling of the embedding of the text prompt during the noise addition and denoising process, when the input text guidance is effective, we can achieve a mixed prompt of text and image. The generated images that follow prove that our method has achieved good results.

5.2 Camouflage Image

This is a new type of illusionary image we have constructed.

Our goal is as follows: Given two sets of prompts (each set consisting of a text prompt and an image prompt, with one of the two prompt types possibly missing), we aim to generate a single image that visually incorporates the information from both sets of prompts. For example, the first set of prompts is ("man dancing", Cai Xukun dancing image.png), and the second set of prompts is ("beautiful scenery with mountains", landscape image with mountains.png). We hope that the generated image will appear as a landscape, but the contours of the mountains will resemble the silhouette of Cai Xukun dancing.

Our method:
During the Stable Diffusion process, we encode both sets of image prompts into the latent space. In a single noise addition and denoising process, we perform noise addition, denoising, and backpropagation for each set of prompts successively. Each set of prompts is handled in the same way as described in section 5.1 for both text and image prompts. In this manner, during the diffusion process, the image latent will continuously approach the implicit features of both sets of image and text prompts in the latent space, thereby achieving the effect of integrating these four pieces of information in the output image.

5.3 Kaleidoscope Effect

The kaleidoscope effect is reflected in the rotation overlay illusion type. We have further achieved the illusion generated by rotating 45 degrees, in addition to the original 90-degree rotation.

These modifications are in the $rotation\_overlays\_for\_colab.ipynb$ file. A new global variable IS_ROTATE_45 has been added to determine whether to rotate by 45 degrees.The corresponding rotation angles in $learnable\_image\_w, x, y, z$ should be changed accordingly.

5.4 Visualize Results as GIFs

We have visualized the result images for each illusion type (except Parker Puzzle) as GIFs, in order to better demonstrate the effects of the illusions.

These modifications are in the *ipynb* files for each illusion. A corresponding generate gif function is added at the end of the original code. The framework for each illusion type is essentially similar. It uses frames to store each frame of the GIF, and finally saves it using imageio to obtain the corresponding GIF. The processing details for each frame of each illusion type, such as rotation and translation, need to be determined according to the specific type.

6 Experiments

For Step 1, the replication of the paper's result images and the result images generated using our own prompts have already been demonstrated in the milestone. Therefore, the following mainly presents the experimental results of the new features we have added.

The following experimental result images are mostly available in GIF format in the supplementary material.

6.1 Image Prompts

The following image is the result of the hidden character using image prompts.

The images used as prompts will be provided in the supplementary material.

图 1: Hidden Character using image prompts

6.2 Camouflage Image

The following are two camouflage illusion images generated with image and text prompts.

The images used as prompts will be provided in the supplementary material while text prompt is "a beautiful scenery with mountains" and "man dancing".

图 2: Camouflage Image 1

图 3: Camouflage Image 2

6.3 Kaleidoscope Effect

The following image is the result of the 45-degree rotation overlay.

example prompt: 'miku froggo lipstick pyramids'

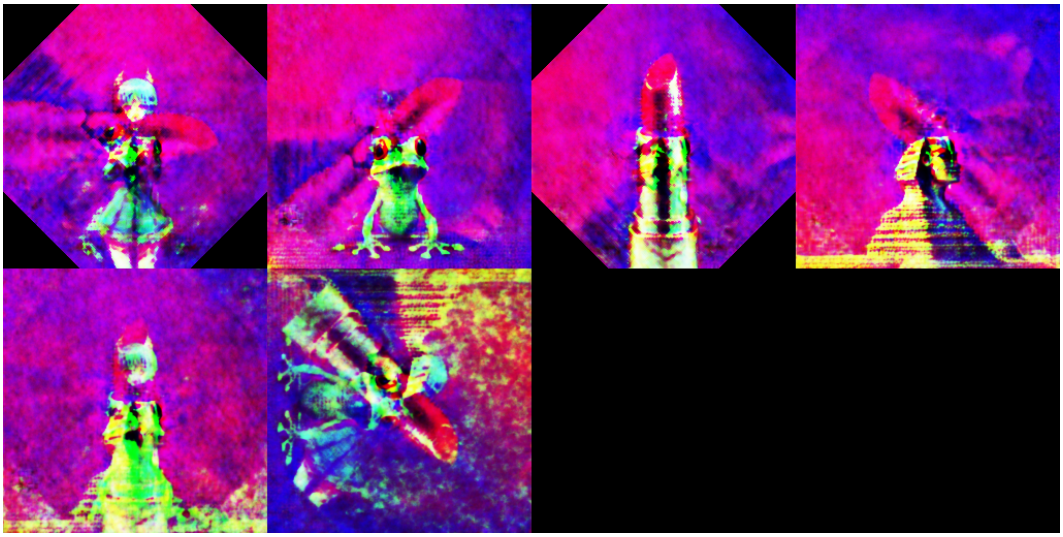

图 4: 45-degree Rotation Overlay

7 Conclusion

In this experiment, we get on hand with the method of stable diffusion illusion.We explored various fascinating illusion kinds and learned about the principle of stable

diffusion with condition prompts. Specifically, we learned how to generate prompt-related images by encoding prompts to latents or embeddings.We also learned methods to encode images and operate with the stable diffusion pipiline, which is very useful. For future extensions , we think that we can enrich the kind of angles in Kaleidoscope Effect, we can also try to increase the numbers of prompts.For future application, we think that it would be a good idea to develop a UI interface for our project, which allows users to modify their own text and image prompt and chose their willing illusion effect.

8 Reference

[1] Burgert, Ryan, et al. "Diffusion illusions: Hiding images in plain sight." ACM SIGGRAPH 2024 Conference Papers. 2024. [Project Page] [Codebase]

[2] Illusion Diffusion Illusion Diffusion