Ensuring Visual Coherency in LDMs

Generative AI models are demonstrating strong performance in various domains. Models such as Stable Diffusion, trained using billions of images, are capable of generating highly realistic images based on text prompts or input images. Using either multiple text descriptions or images as input, completely different visual styles can be combined. However, the results are not always visually pleasing or coherent.

A common approach is to experiment with different prompts or input images until the desired visual result is achieved. This is a slow and manual method that potentially wastes significant computing resources. Instead of generating images and then assessing their visual coherence by inspection, this thesis is focused on automatically assessing the likelihood of pleasing visual results from the input alone. Specifically, we will focus on assessing the compatibility of inputs for the open-source image generation model Stable Diffusion.

For this, different text or image inputs are encoded into latent space and analyzed regarding their compatibility. A method needs to be developed that assesses the compatibility of the different inputs. Different metrics such as cosine similarity, FID, or Image Aesthetic Assessment (IAA) methods such as the CLIP score can be used as a starting point. After assigning the input an aesthetic score, the same method should then be used to identify potential changes of the input to increase the predicted aesthetic. The goal of this work is to allow for designing inputs that contain the desired visual concepts, while maximizing the likelihood of visually pleasing and coherent outputs.

This thesis will be conducted externally at Sureel Inc., a startup specializing in secure and legal generative AI content.

Prerequisites

Requirements: Experience with Python and machine learning

Contact

christopher@sureel.ai

Advisors

Eckehard Steinbach
Christopher Kuhn (Sureel Inc.)