Controlling Lighting with LoRA Adaptors

Lars de Wolf
lars.de.wolf@student.uva.nl
l.de.wolf@student.vu.nl

Nimi Barazani
nimrod.barazani@student.uva.nl

Yigit Can Ozkaya
yigit.ozkaya@student.uva.nl
y.c.oezkaya@student.vu.nl

Martijn van Raaphorst
martijn.van.raaphorst@student.uva.nl

Simon Hüsgen
simon.huesgen@student.uva.nl

Loong Li
loong.li2@student.uva.nl

Abstract

Advancements in image generation using diffusion models have improved the ability to produce high-resolution images. However, controlling specific attributes, such as lighting, with their complex effects on shadows and reflections, remains challenging. This report investigates the use of Low-rank Adaptors (LoRA) concept sliders to control lighting conditions in generated images. Building on previous research, we explore the potential of these sliders to control attributes like light source direction and lens flare correction. Utilizing paired datasets and a pre-trained diffusion model, our experiments reveal mainly the limitations of textual and visual sliders. Textual sliders excel in managing straightforward attributes but fall short with abstract concepts such as light direction. Visual sliders offer some promise in adjusting lighting attributes but encounter challenges in achieving consistent results. Our study highlights the need for advanced training methods and comprehensive datasets to improve control over complex visual features in diffusion models.

1. Introduction

Development in image generation has been progressing quickly over the past few years. Diffusion models have been playing a big role in this development because they can generate high-resolution images of varying quality relatively consistently. However, using only text prompts, it can be challenging to adjust certain attributes in the generated image, such as a person's age or the intensity of the weather [3]. This can become a problem for artists who wish to use diffusion models [2].

In a recent paper [6], the authors propose the use of concept sliders, which are both interpretable and allow for the desired control of detail. They use the method to achieve this by making what they call concept sliders by applying Low-rank adaptors (LoRA) [9] to their diffusion model. This approach shows that they can control details like weather, age, styles, and expression. Though this is a major improvement in control over a normal diffusion model, the concepts demonstrated in this paper are still relatively simple, focusing mainly on human attributes and, to a small extent, on the natural and physical conditions of the image.

Recognizing the method's potential, we aim to extend the use of these concept sliders in this report. We look into investigating whether more complex and nuanced attributes can be effectively controlled using the LoRA method, with concepts focusing on the lighting conditions of the generated image.

Concept ideas Our investigation started by developing a number of concept ideas that were more complex than the original paper but still achievable within the limited time window we had for this project. We came up with four main concepts.

- *Perspective Control*: This concept potentially controls the perspective of a scene by training the model to multi-view images of the same scene in preferably the same lighting. The goal is to enable the model to control and adjust the perspective of a scene.
- *Time Control*: This idea aims to control the time of day depicted in an image, which would be reflected through changes in shadows, brightness, and other related attributes. This concept would be trained using images captured at different intervals during a timelapse.
- *Flare Control*: This concept focuses on managing image flare by training the model on images of the same scene with varying intensities of flare caused by a light source. This would allow the model to correct the flares' presence and intensity.
- *Lighting Source*: Similar to the time control slider, this idea aims to control lighting conditions, including shadows or brightness, under different lighting directions.

Given the complexity of these concepts, we decided to focus on two concepts: the *Lighting Source* concept and the

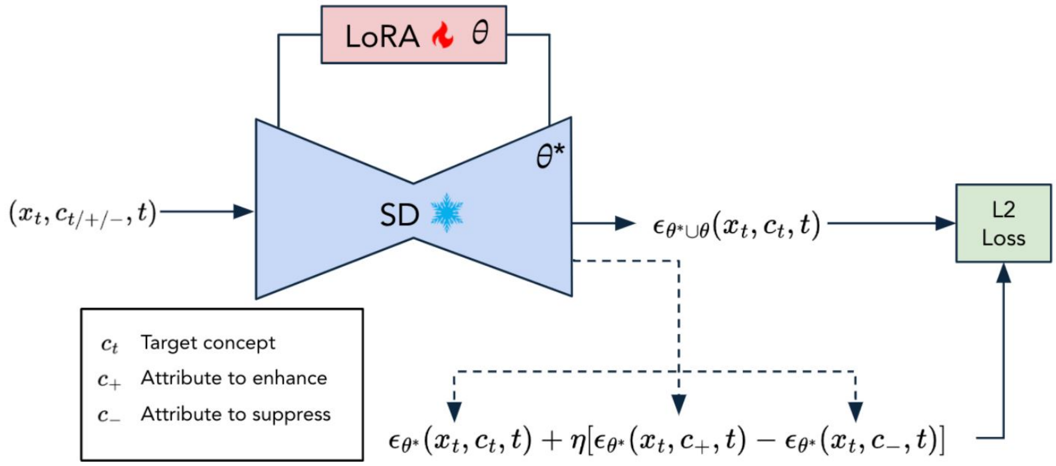

Figure 1. LoRA adaptors concept sliders training framework

Flare Control concept. Focusing on these two concepts, we investigate the possibility of using LoRA control sliders to manage the physical effects of different lighting conditions in an image, aiming to enhance the ability of diffusion models to handle lighting variations. Specifically, the Lighting Source concept aims to control the direction and intensity of lighting. Meanwhile, the Flare Control concept seeks to manage the presence and severity of lens flares, ensuring that the generated images maintain visual clarity and realism.

Our study provides insights into using LoRA concept sliders to control the lighting attributes of generated images in diffusion models. Disentangling more complex features like light direction and flares using text sliders is particularly challenging due to the difficulty of defining them through text. Similarly, visual sliders perform poorly on more complex concepts such as light direction and flares. Features such as brightness often become entangled with attributes like image detail and sharpness, complicating the control process. Additionally, we believe that training on natural images, out of the distribution of the images generated by the diffusion model, might also impact the inability to control the lighting of the image and that other methods with additional conditioning information might be needed to control the lighting in the generation.

2. Related Work

2.1. Diffusion models

Diffusion models are generative models that use the concept of reversing a diffusion process to generate data [8, 19]. In the initial forward diffusion process, noise is incrementally added to the data, transforming it from an organized state, the original image, x_0 to a state of complete Gaussian noise, e.g. at time t the noised image is modeled as follows:

$$x_t = \sqrt{1-\beta_t}x_0 + \sqrt{\beta_t}\epsilon$$

where ϵ is a gaussian noise. The training objective is to minimize the difference between predicted and true noise

$$\mathrm{E}_{x_0 \sim \mathcal{D}, \epsilon \sim \mathcal{N}(0,1)} \left[\|\epsilon - \epsilon_\theta(x_t, c, t)\|^2 \right]$$

where $\epsilon_\theta(x_t, c, t)$ is the predicted noise at time step t conditioned on c. By iteratively optimizing this loss over multiple timesteps and data samples, the model learns to denoise accurately at each stage, effectively reversing the diffusion process and enabling high-quality data generation from pure noise. Recent developments in stable diffusion models have significantly advanced this field, allowing the generation of high-quality images. For example, Stable Diffusion 3, handles multi-subject prompts [18], and DALL-E 3 works well even for detailed image descriptions [1]. However, even with advanced models, to precisely control specific image details such as lighting, a descriptive prompt to describe that is challenging. Following the original work [6], we will use Stable Diffusion XL [17] for our experiments.

2.2. Concept control

Diffusion models can generate high-quality images. However, these models do not provide detailed control over the generated image. In one stream of research, concept control was explored, for example, through the adaption of latent representation in a generative model [10], by optimizing an additional neural network to add spatial controls to condition the diffusion model [22], or by fine-tuning subset of the model parameters [14]. Although these methods can control a concept, they lack the ability to add precise control over it and adjust the strength of the concept. In a recent work [6], concept sliders are introduced to tackle this problem. These concept sliders are based on Low-Rank adaption (LoRA) and enable concept-targeted image control. The concept sliders learn low-rank parameter directions that increase or decrease the expression of specific attributes when conditioned on a target concept. This can be done by using either textual sliders or visual sliders. For textual sliders, the goal is to obtain θ that modifies the likelihood of attributes c_+ and c_- in image X when conditioned on c_t. The process is illustrated in figure 1. The visual sliders learn to capture the visual concept by contrasting image pairs. This allows more abstract concepts to be used since they only have to be illustrated and not described.

2.3. Lighting control in generative models

Lighting control in generative models has seen significant advancements but remains challenging. Traditional relighting methods often relied on classical approaches such as image-based rendering or shape from shading [5, 16]. Generally, relighting of images has advanced with neural methods from image to image style transfer [11, 23], learned relighting function [21] to the use of volumetric representation and NeRF-like methods to render a scene at different lighting conditions [20]. As mentioned, conditioned generative models have enabled some control over features, but explicit control over lighting remains a largely unexplored area. Several methods for generative models [12, 15, 22] have provided some control over generated image features. Notably, in LightIt [13] the authors developed a method for explicit illumination control for image generation through the estimation of shading of the image from a single image to condition a trainable module that can control the diffusion model.

3. Method

In our research, we aim to understand the learnability of complex concepts with LoRa sliders for diffusion models. In this work, we use the concept slider approach established by Gandikota et al[6], and introduce novel visual and textual concepts that describe more complex transformations. In the original work, LoRa sliders for concepts such as age, eye size and eyebrow shape were developed, which mainly focus on human attributes. Unlike the original work, we aim to implement more abstract concept sliders that modify the lighting conditions in a scene, for example, brightness or light direction. The following section outlines the methodology used in the development of these more complex concept sliders, starting with the model definition.

3.1. Model Definition

We utilized a pre-trained diffusion model, Stable Diffusion XL [17], as the base model for our experiments. This architecture involves a two-stage pipeline, where a 2.6 billion parameter-sized Unet generates the initial noisy latents. These are then refined and pushed through a variational autoencoder (VAE), resulting in a full-resolution image. This model allows for high-quality image generation and will serve as the backbone for our image generation process. We denote this base model as θ_0 and train for each concept c_i a set of low-rank parameters $\Delta\theta$. These parameters are defined as a set of weight matrices, where each matrix $w_i \in \mathbb{R}^{d\times k}$ can be obtained by multiplying two lower-rank matrices B & A, where B $\in \mathbb{R}^{d\times r}$ and A $\in \mathbb{R}^{r\times k}$ [9]. B and A represent our trainable parameters, and their product ($\Delta\theta$) is merged with θ_0, which is left completely unchanged during training and inference. Additionally, we scale $\Delta\theta$ by α during inference, resulting in an increase or decrease of the concept c_i infers. More formally, we compute for each concept c_i low-rank parameters $\Delta\theta$ and scale and merge these with the base model θ_0. This gives us θ^*, where

$$\theta^* = \theta_0 + \alpha\Delta\theta = \theta_0 + \alpha BA \tag{1}$$

3.2. LoRA Training

In the following section, we briefly discuss the process for training LoRa concept sliders based on text prompts. Additionally, we also provide the training loss for training visual concept sliders. We closely follow the method detailed by Gandikota et al.[6]

3.2.1 Text Sliders

Our objective is to obtain a model θ^*, which when conditioned on target concept c_t, increases the likelihood of positive attributes c_+ and decreases the likelihood of negative attributes c_- on image X. In this case, the target, as well as positive and negative attributes are specified as text prompts. We achieve this by modifying the distribution of the base model θ_0 conditioned on the target concept with the attribute-conditioned distribution.

$$\theta^*(X|c_t) = \theta_0(X|c_t)\left(\frac{\theta_0(c_+|X)}{\theta_0(c_-|X)}\right)^\eta \tag{2}$$

By taking the logarithm of the base distribution and the attribute-conditioned distribution, we can rewrite the objective function as an addition and subtraction of the objective probabilities. These individual probabilities are converted to scores by taking the gradient, resulting in the following:

$$\nabla\theta_0(X|c_t) + \eta(\nabla\theta_0(c_+|X) - \nabla\theta_0(c_-|X)) \tag{3}$$

As discussed in 2.1, diffusion models operate by partially denoising Gaussian noise in image X. The objective function (3) is used in training the denoising process, by turning each score into a denoising prediction ϵ(X, c_t, t). Thus the objective becomes to denoise Gaussian noise at timestep t given the conditioning. In addition to this transformation, so-called preservation concepts were introduced. These concepts aim to preserve attributes that are preferred to keep constant during the optimization. Therefore we condition on the positive and negative attributes for each concept p in the preservation set P, resulting in the final objective function.

$$\begin{aligned}\epsilon_{\theta^*}(X, c_t, t) =& \epsilon_{\theta_0}(X, c_t, t) + \\ & \eta \sum_{p \in P} (\epsilon_{\theta_0}(X, c_+, t) - \epsilon_{\theta_0}(X, c_-, t))\end{aligned} \tag{4}$$

3.2.2 Visual Sliders

We will now discuss the training of visual concept sliders, which requires a paired dataset where the effect of the concepts is applied across various concept scales. The concept is then learned by capturing the contrast between the same images of different scales. Similar to training text sliders, we aim to increase/decrease the positive and negative attributes c_-, c_+, given target concept c_i. However, in this case, given scales X, Y where X < Y, we aim to train the LoRa parameters such that scale X is more directed towards the negative attributes and Y is more directed towards the positive attributes. This gives us the following loss function:

$$\|\epsilon_{\theta_-}(x_t^A, `', t) - \epsilon\|^2 + \|\epsilon_{\theta_+}(x_t^B, `', t) - \epsilon\|^2 \tag{5}$$

4. Experiments

In this section, we describe the experiments we conducted to achieve control over lighting conditions in image generation.

4.1. Datasets

The training of visual concept sliders using LoRA starts by creating paired image datasets, which should capture the desired effect of the concepts across various scales. The intensity of the concept should vary incrementally along the scale, enabling the model to progressively learn and refine its understanding of the concept. The quality of the training data will significantly impact the effectiveness of the concept slider. Therefore,

In our experiments to achieve control over lighting conditions in image generation, we utilized two datasets: the Multi-PIE Face dataset [7] and the Flare7K dataset [4]. These datasets were selected for their ability to capture the required lighting effects and conditions we set to test. The first is the ability to control lighting source direction, and the latter is to control flare effects and correct them.

Multi-PIE Face Dataset The Multi-PIE Face Dataset contains images of faces with varying lighting from different perspectives and expressions. For our experiments, we focused on images where the subject's pose is centered, ensuring consistency in the training data. This dataset provides 498 groups or pairs of images, each pair featuring the same subject under different lighting conditions. There are 20 distinct illumination settings in total. This variety in lighting conditions allows to model how changes in illumination affect the appearance of the subject's face. In figure 2, a subset of such groups of images is displayed.

Figure 2. Examples from the Multi-PIE Face Dataset: This figure displays images from the Multi-PIE Face Dataset, with different lighting conditions applied to centered posed faces. Each row shows a different subject, while the columns present the same subject under 20 different lighting settings.

Flare7K Flare7K is a specialized dataset designed for removing nighttime lens flares from images. It is created based on real-world observations of nighttime flares and includes 7,000 flare patterns, with 5,000 scattering flares and 2,000 reflective flares. The dataset features 25 types of scattering flares and 10 types of reflective flares, providing a wide variety of flare effects. It also includes paired images, with each flare-corrupted image matched to a flare-free version. This setup allows easier training. Additionally, the dataset comes with detailed annotations, such as the positions of light sources, glare, reflective flares, and streaks, which are often missing in other datasets. In figure 3, you can see some example images from Flare7K.

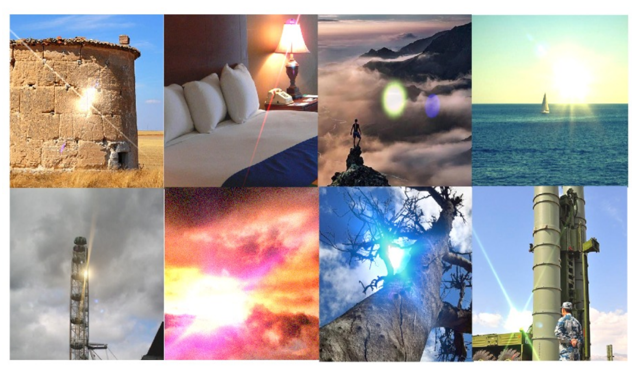

Figure 3. Examples from the Flare7k Face Dataset

4.2. Training Details

Our approach builds upon the methodology and public implementation provided in the original paper. For the visual sliders, we conducted training for 2500 epochs, and for the text sliders, we trained for 1000 epochs. We utilized a batch size of 1 throughout our experiment. For the results presented in Figure 6, we only focused on two lighting directions, light from the right and light from the left. For the results presented in Figure 7, we choose two other lighting directions, light in front of the subject and light from the back. The textual sliders and some visual sliders were used with relevant prompts that are provided in Supplementary (Section 7 and Section 8). We evaluated our approach on a pre-trained model, Stable Diffusion XL [17], a high-resolution 1024-pixel model. Similar to the method's original paper [6], the LoRA adaptor multipliers are set to 0 for inference at the initial diffusion steps to maintain structure and semantics. We activated the LoRA adaptor for the remaining steps. This parameter can be tuned at inference time.

4.3. Results

Our experiments aimed to evaluate the effectiveness of textual and visual concept sliders using LoRA adaptors for brightness, light direction, and flare removal. We applied visual and textual sliders to assess their performance in manipulating these attributes.

4.3.1 Textual Concept Sliders

As a preliminary step, we reproduced two text sliders, the age, and the smiling textual sliders, that were suggested and presented in [6]. In addition, as a first step toward light control, we tested a brightness control slider, and a textual slider was tested to see if descriptive prompts could effectively control brightness levels in generated images. The results were promising, as the sliders successfully influenced brightness. Other attributes in the image remained consistent while traversing the sliders, and moving the sliders in the negative direction can demonstrate an opposing effect on the brightness. The generated images for these 3 sliders are presented in Figure 4.

The results were promising for the capability of textual sliders to handle specific attributes, so we applied the same method to modify light direction. However, the textual sliders could not achieve any substantial difference for this as can be seen in Figure 5. Changes in light direction were either too subtle or non-existent, indicating that the model could not accurately interpret and apply the prompts to manipulate lighting since this concept is probably hard to describe with text and, thus, also hard to learn.

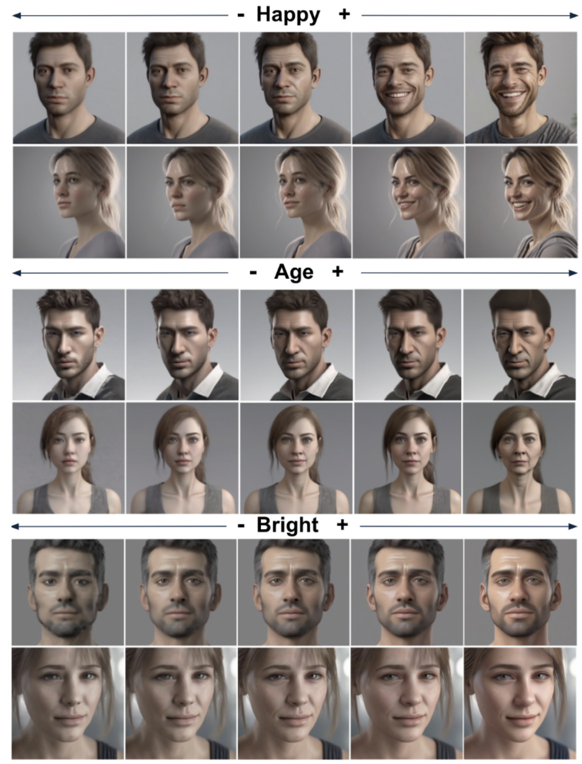

Figure 4. Generated images for brightness, age, and facial expression with textual concept sliders.

Figure 5. Generated images for light direction with textual sliders. Direction changes from left to right, respectively, from minus to plus.

4.3.2 Visual Concept Sliders

For both concepts, the flare and light direction, the visual sliders were designed to modify the concepts by using contrasting image pairs.

Similarly to the experiments with textual sliders, we tested a brightness slider as a first step toward our light direction slider. However, as seen in Figure 7, the visual sliders struggled to achieve the desired brightness adjustments for the concept of brightness. The changes observed were not consistent or significant enough to be deemed successful. One of the reasons for that is brightness of an image may get entangled with image detail and sharpness.

There were some changes for the light direction experiment, seen in Figure 6, such as slight adjustments in shadows and reflections over the objects' eyes. However, these changes were minimal and often lacked coherence. The visual sliders did not convincingly alter the perceived direction of light, indicating that this method also faced significant challenges in handling this complex attribute

Interpreting our flare concept slider results, we observe the removal of flares to some extent. Light rays are reduced and in some cases, the brightness of the light source is slightly dimmed. However, as visible in 8, the flare concept slider is not able to completely remove lens flares from images. Our concept slider is able to modify camera flares (e.g., the first image pair). However, remnants of the original lens flare still remain after modification. Additionally, we observe that the model is not capable of reducing multiple flares and, in some cases, will struggle to keep other attributes consistent.

Figure 6. Generated images for light direction with visual concept sliders.

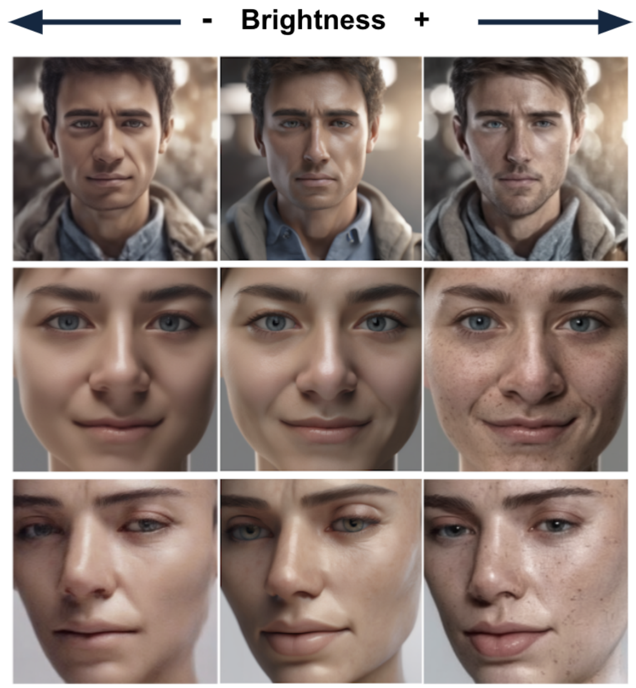

Figure 7. Generated images for brightness with visual concept sliders.

5. Discussion

The research presented in this paper extends the application of Concept Sliders for diffusion models. Specifically, LoRA adaptors that are able to control and adjust visual attributes in image generation. Therefore, our study contributes to the understanding and offers new insights into the limitations of diffusion models utilizing LoRA for precise image control.

Our experiments focused on both textual and visual sliders, exposing strength and weaknesses in handling of complex visual concepts.

Using textual sliders we demonstrated their ability to control clear and distinct features such as age and facial features. These successful experiments can be attributed to the clear and textual conditioning which closely aligns to the feature representations in the training data. When examining more abstract attributes, such as light direction, its performance plunged. A lack of direct textual correlations of these attributes in the training data and the difficulty to explain such correlations well via text, makes the model unable to learn and manipulate these well.

Similarly, visual sliders, designed to capture these attributes through contrasting pairs of images, have shown competence in manipulating attributes with limited overall image impacts, such as eye and eyebrow size, as presented in [6], suggesting that conditioning diffusion models based on images can enhance control over specific attributes. When experimenting with more complex features,

Figure 8. Visual concept sliders for removing lens flare.

such as light direction and brightness, the same challenges as with the textual sliders appeared. The visual sliders achieve some correct adjustments of visual attributes, such as moving shadows, most likely due to their advantage of capturing contrasting visual attributes through image pairs. However, most of the generated images had barely any noticeable changes or strictly non-cohesive ones. More specifically, the model often confused brightness with other elements like image detail and sharpness. These abstract attributes seem impossible for the model to disentangle, suggesting that the current implementation of the concept sliders is unable to fully capture complex interplay of various image attributes.

Although the visual sliders provide slightly better results on the complex interleaved image attributes, both are unable to generate good results for those.

Our findings, therefore, suggest that the ability of these Concept Sliders depends on the distinctiveness and clarity of the visual attributes in question. This indicates the need for more sophisticated training strategies or model architectures so that intertwined attributes can be separated. In addition, we experimented with existing real-world or synthetic datasets, meaning out-of-distribution images from those generated by the diffusion models. Using edited generated images by the model could improve the results. Accordingly, the current applicability of Concept Sliders is mainly practical for facilitating more nuanced control over attributes that are well represented and can be distinctly modeled within the training data. Thus, our research also highlights the importance of dataset composition in training diffusion model concept sliders.

As we discussed before, the main limitations in our study were related to the model's difficulty in handling complex interleaved visual attributes. Concepts like light direction and brightness often led to mixed results where adjusting one attribute unintentionally affected other aspects of the image, like sharpness and detail. This shows that we need more sophisticated training methods or model designs to better separate these attributes. Additionally, we tried to rely on high-quality, diverse datasets to capture the full range of each attribute. If the datasets were incomplete or biased, the sliders would be less effective.

6. Conclusion

While this study has demonstrated that Concept Sliders using LoRA adaptors offer a promising path for enhanced control for diffusion models, it also revealed significant challenges that remain. We showed that using very low amount of image pairs or only using textual descriptions, specific image attributes can be controlled. This ability to manipulate image attributes in diffusion models directly with an efficient fine-tuning via LoRA adaptors represents a significant advancement in the field of image generation. However, to fully reach its potential the challenge of disentangling complex attributes in latent space needs to be addressed.

Future research could explore methods to achieve the disentanglement of complex image attributes. This could include the integration of more advanced techniques such as ControlNet, conditioning on shadings, or similar methods to improve the separation in latent space. Additionally, expanding the scope and complexity of usable datasets for this task is essential.

References

[1] James Betker, Gabriel Goh, Li Jing, † TimBrooks, Jianfeng Wang, Linjie Li, † LongOuyang, † JuntangZhuang, † JoyceLee, † YufeiGuo, † WesamManassra, † PrafullaDhariwal, † CaseyChu, † YunxinJiao, and Aditya Ramesh. Improving image generation with better captions. 2

[2] Victor Boutin, Thomas Fel, Lakshya Singhal, Rishav Mukherji, Akash Nagaraj, Julien Colin, and Thomas Serre. Diffusion models as artists: Are we closing the gap between humans and machines?, 2023. 1

[3] Pu Cao, Feng Zhou, Qing Song, and Lu Yang. Controllable generation with text-to-image diffusion models: A survey, 2024. 1

[4] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological nighttime flare removal dataset. *Advances in Neural Information Processing Systems*, 35:3926–3937, 2022. 4

[5] Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Jorge Lopez Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. Multiview intrinsic images of outdoors scenes with an application to relighting. *ACM Trans. Graph.*, 34(5), 2015. 3

[6] Rohit Gandikota, Joanna Materzynska, Tingrui Zhou, Antonio Torralba, and David Bau. Concept sliders: Lora adaptors for precise control in diffusion models. *arXiv preprint arXiv:2311.12092*, 2023. 1, 2, 3, 5, 6

[7] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-pie. *Image and vision computing*, 28(5):807–813, 2010. 4

[8] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. 2

[9] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 1, 3

[10] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls, 2020. 2

[11] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review, 2018. 3

[12] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks, 2019. 3

[13] Peter Kocsis, Julien Philip, Kalyan Sunkavalli, Matthias Nießner, and Yannick Hold-Geoffroy. Lightit: Illumination modeling and control for diffusion models, 2024. 3

[14] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-concept customization of text-to-image diffusion, 2023. 2

[15] Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, Ying Shan, and Xiaohu Qie. T2i-adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models, 2023. 3

[16] Julien Philip, Michaël Gharbi, Tinghui Zhou, Alexei A. Efros, and George Drettakis. Multi-view relighting using a geometry-aware network. *ACM Trans. Graph.*, 38(4), 2019. 3

[17] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis, 2023. 2, 3, 5

[18] K. C. Sabreena Basheer. Stability ai introduces stable diffusion 3: Next-gen ai imagery. *Analytics Vidhya*, 2024. 2

[19] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics, 2015. 2

[20] Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. Neural fields meet explicit geometric representation for inverse rendering of urban scenes, 2023. 3

[21] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. Deep image-based relighting from optimal sparse samples. *ACM Trans. Graph.*, 37(4), 2018. 3

[22] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models, 2023. 2, 3

[23] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David Jacobs. Deep single-image portrait relighting. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7193–7201, 2019. 3

Controlling Lighting with LoRA Adaptors

Supplementary Material

7. Textual Sliders Prompts

7.1. Brightness

- **Target**:"person face"
- **Positive**:"person face, bright, well-lit, sunlit, vibrant colors, illuminated"
- **Unconditional**:"person face, dark, dimly lit, shadows, desaturated colors, obfuscate"
- **Neutral**:"style, realistic, sharp, sharpness, texture, color"

7.2. Age

- **Target**:"male person, female person"
- **Positive**:"male person, female person, very old"
- **Unconditional**:"male person, female person, very young"
- **Neutral**:"male person, female person"

7.3. Happiness

- **Target**:"male person, female person"
- **Positive**:"male person, female person, very happy, smiling"
- **Unconditional**:"male person, female person, very sad"
- **Neutral**:"male person, female person"

7.4. Light Direction

- **Target**:"male person face, female person face"
- **Positive**:"male person face with light directed from the right and shadow on left side, female person face with light directed from the right and shadow on left side"
- **Unconditional**:"male person face with light directed from the left and shadow on the right side, female person face with light directed from the left and shadow on the right side"
- **Neutral**:"style, realistic, sharp, sharpness, texture, color"

8. Visual Sliders Prompts

8.1. Light Direction

- **Target**:"person face"
- **Positive**:"person face with light directed from the right and shadow on left side"
- **Unconditional**:"person face with light directed from the left and shadow on the right side"
- **Neutral**:"style, realistic, sharp, sharpness, texture, color"

8.2. Flare Removal

- **Target**:"image"
- **Positive**:"image with no lens flare"
- **Unconditional**:"image with lens flare"
- **Neutral**:"image"