

Controle de Estilo Artístico em Modelos Generativos: Uma Análise Comparativa entre Engenharia de Prompt e LoRA

Gabriel Inagaki Marcelino

¹Departamento de Ciências da Computação e Estatística – Universidade Estadual Paulista (UNESP)
Caixa Postal 15054-000 – São José do Rio Preto – SP – Brazil

Abstract. *Image generation using artificial intelligence models has gained prominence due to its potential to produce visually rich content from textual descriptions. In this work, we compare two distinct approaches for artistic style control: prompt engineering and fine-tuning with Low-Rank Adaptation (LoRA), both applied to the generation of images in the impressionist style of Claude Monet using the Stable Diffusion model. A dataset of Monet's real paintings was collected through web scraping using the BeautifulSoup library. Through a controlled experiment, we evaluate style fidelity, consistency across generations, and the visual quality of the resulting images using automatic metrics such as CLIPScore and FID, alongside qualitative analysis. Results indicate that fine-tuning with LoRA achieves higher style fidelity and consistency, while prompt engineering offers a faster, less resource-intensive alternative with satisfactory outcomes for exploratory use cases.*

Resumo. *A geração de imagens com modelos de inteligência artificial tem se destacado pelo potencial de produzir conteúdos visualmente ricos a partir de descrições textuais. Neste trabalho, comparamos duas abordagens distintas para o controle de estilo artístico: a engenharia de prompts (prompt engineering) e o fine-tuning com Low-Rank Adaptation (LoRA), ambas aplicadas à geração de imagens no estilo impressionista de Claude Monet utilizando o modelo Stable Diffusion. A base de dados com pinturas reais de Monet foi coletada via web scraping, utilizando a biblioteca BeautifulSoup. Por meio de um experimento controlado, avaliamos a fidelidade ao estilo, a consistência entre gerações e a qualidade visual das imagens resultantes, utilizando métricas automáticas (CLIPScore e FID) e análise qualitativa. Os resultados indicam que o fine-tuning com LoRA proporciona maior fidelidade e consistência estilística, embora a engenharia de prompts apresente uma solução mais rápida e menos custosa computacionalmente, com resultados ainda satisfatórios para aplicações exploratórias.*

1. Introdução

A geração de imagens por meio de modelos de inteligência artificial teve grandes avanços nos últimos anos, especialmente com o advento de modelos de difusão como o Stable Diffusion [Rombach et al. 2022], permitindo transformar descrições textuais em imagens visualmente coerentes. Controlar o estilo visual das imagens geradas pelos modelos provavelmente é um desafio dada a natureza de “caixa preta” de tais tecnologias.

Duas abordagens se destacam para essa tarefa: a engenharia de *prompts*, que explora descrições textuais detalhadas para guiar o modelo, e o *fine-tuning*, que adapta os pesos do modelo para especializá-lo em um domínio visual específico. A técnica de *Low-Rank Adaptation* (LoRA) tem se consolidado como uma alternativa eficiente para esse *fine-tuning*, exigindo menos recursos computacionais e dados de treinamento [Hu et al. 2021].

Este trabalho investiga e compara essas duas abordagens no contexto da geração de imagens no estilo impressionista de Claude Monet. Utilizaremos o modelo Stable Diffusion como base para ambos os métodos e propomos um experimento com *prompts* padronizados, uma base de dados real das obras de Monet, e métricas de avaliação objetivas (CLIPScore [Hessel et al. 2021], FID [Heusel et al. 2017]) e subjetivas (análise visual qualitativa). O objetivo é analisar em que medida a engenharia de *prompts* pode se aproximar do resultado obtido com o *fine-tuning* especializado, considerando fidelidade ao estilo, consistência e esforço computacional.

2. Materiais e Método

Esta seção apresenta duas abordagens distintas para o controle de estilo na geração de imagens com modelos de difusão: a engenharia de *prompts* e o *fine-tuning* com Low-Rank Adaptation (LoRA)[Hu et al. 2021], ambas aplicadas ao estilo impressionista do pintor Claude Monet. O modelo base utilizado foi o Stable Diffusion v1.4[Rombach et al. 2022], amplamente adotado em tarefas de geração de imagens condicionadas por texto. A metodologia está dividida em: preparação dos dados, adaptação do modelo e avaliação por geração.

2.1. Materiais Utilizados

Esta subseção descreve os recursos computacionais, *frameworks*, ferramentas e bases de dados utilizados no desenvolvimento do projeto.

- O pré-processamento e a etapa de engenharia de *prompts* foram realizados em um computador pessoal.
- A base de imagens foi construída a partir da galeria online Claude Monet Gallery [cmg 2025], com apoio de scripts em Python para baixar e organizar.
- O *fine-tuning* com LoRA foi executado em uma GPU RTX 4090 24GB disponível no servidor do laboratório LIDIA¹.

Tabela 1. Especificações das máquinas utilizadas

Ambiente	Processador	GPU
Computador pessoal	Intel i5-9300H (8) @ 4.10GHz	NVIDIA GeForce GTX 1650 ²
Servidor LIDIA	Intel i9-13900K (32) @ 5.80GHz	NVIDIA RTX 4090 24GB

Fonte: Autoria própria

2.2. Frameworks e Ferramentas

A linguagem utilizada foi Python 3.10. As principais bibliotecas e *frameworks* empregados foram:

- **PyTorch** [Paszke et al. 2019]: para construção e treinamento dos modelos.
- **Diffusers** [von Platen et al. 2022]: para fine-tuning LoRA sobre o Stable Diffusion.
- **Datasets** [Lhoest et al. 2021]: para manipulação e particionamento da base de dados.
- **Accelerate** [Gugger et al. 2022]: para facilitar o gerenciamento de dispositivos e paralelismo.
- **BeautifulSoup** [Richardson 2023]: para extração automatizada de dados (web scraping) das páginas com obras de Monet.
- **CLIPScore e FID**: como métricas automáticas de avaliação.

¹LIDIA - Laboratório de Inovação e Desenvolvimento em Inteligência Artificial do IBILCE/UNESP

2.3. Base de Dados

A base de dados foi composta por 1.956 imagens de pinturas de Claude Monet, extraídas da Claude Monet Gallery [cmg 2025]. Cada imagem foi redimensionada para 512×512 pixels, com padding preto nas bordas para preservar a proporção original.

As legendas (captions) foram geradas automaticamente utilizando o modelo BLIP (*Bootstrapping Language-Image Pretraining*) da HuggingFace: Salesforce/blip-image-captioning-base. Após a geração, as legendas passaram por uma etapa de filtragem para remoção de expressões comuns como “*a painting of*”, “*an artwork of*” e menções diretas ao autor. O objetivo foi permitir que o modelo aprendesse o estilo de Monet a partir das características visuais, sem depender de pistas explícitas no *prompt*.

2.4. Engenharia de *Prompts*

Na primeira abordagem, utilizou-se o modelo base Stable Diffusion sem modificações. Foram criados *prompts* textuais que evocam o estilo impressionista de Monet, com expressões como: “*soft brushstrokes*”, “*light reflections on water*”, “*pastel tones*”, “*impressionist style*”, “*in the style of Claude Monet*”, entre outros. As imagens geradas foram utilizadas para análise qualitativa e quantitativa comparativa.

2.5. Fine-tuning com LoRA

Na segunda abordagem, o modelo foi adaptado utilizando Low-Rank Adaptation (LoRA) [Hu et al. 2021]. O *fine-tuning* foi aplicado apenas sobre o módulo UNet da arquitetura Stable Diffusion, mantendo congelado o codificador textual (CLIP). Os pares imagem-legenda (caption) foram utilizados como supervisão textual. O treinamento foi realizado por 10 épocas, com *batch size* de 1 e *learning rate* de 1e-4, com duração total aproximada de 1,5 hora no servidor LIDIA.

2.6. Método

A Figura 1 apresenta o fluxograma geral do método. Inicialmente, realiza-se a coleta e o pré-processamento dos dados. Em seguida, o modelo é refinado com LoRA. Imagens são geradas tanto com o modelo adaptado quanto com o modelo base. Por fim, as imagens são avaliadas a partir de *prompts* padronizados.

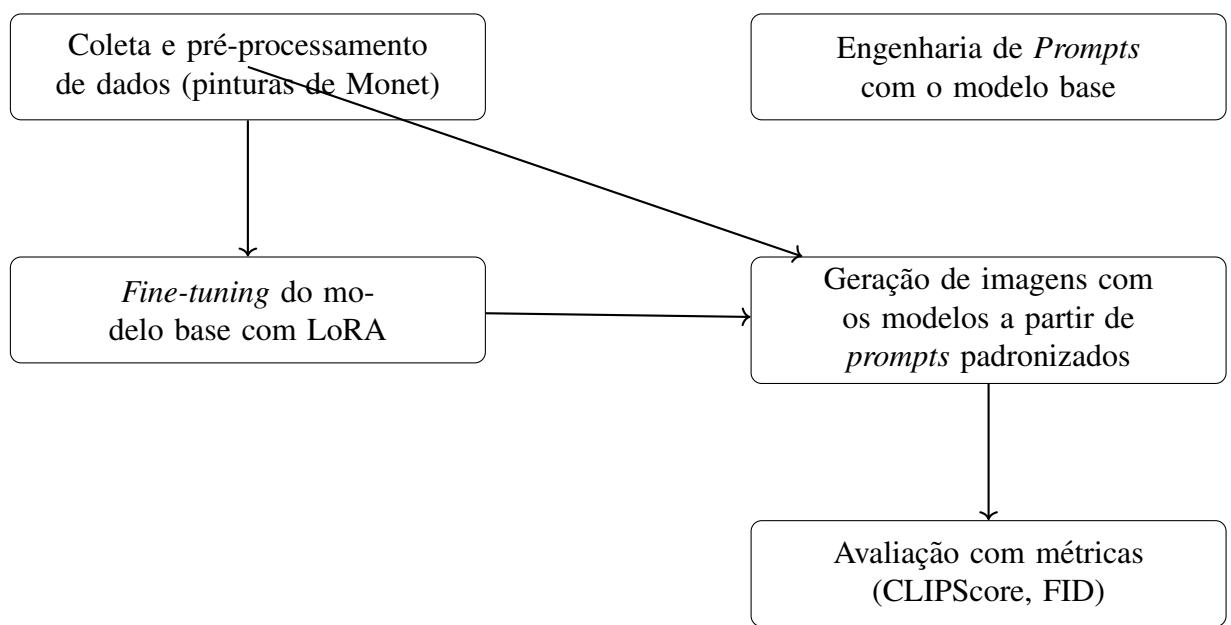


Figura 1. Fluxograma da metodologia. A engenharia de *prompts* ocorre independentemente dos dados; já o *fine-tuning* utiliza para treinamento as imagens coletadas.

Fonte: Autoria própria

3. Experimentos e Resultados

Nesta seção são apresentados os experimentos realizados para comparar as abordagens de controle de estilo por engenharia de *prompts* e por *fine-tuning* com LoRA. São descritas as configurações utilizadas, os critérios de avaliação e os resultados obtidos com base em métricas objetivas e análise qualitativa.

3.1. Configuração Experimental

Foram utilizados os *prompts* com leves diferenças para os dois modelos, levando em conta que o modelo para engenharia de *prompts* é generalista, ele precisa de entradas detalhadas e específicas para gerar as imagens de forma esperada, por exemplo “*A field of blooming wildflowers under a cloudy sky, painted in the impressionist style of Monet*”, enquanto, por outro lado, o ajuste com LoRA do outro modelo foi feito com o propósito de o modelo não precisar de especificações sobre o estilo, mas sim gerar sempre nesse estilo, com a entrada equivalente sendo “*A field of blooming wildflowers under a cloudy sky*”, omitindo detalhes sobre o estilo. A Tabela 2 mostra cada um dos *prompts* utilizados.

Os *prompts* foram utilizados para gerar 10 imagens distintas em cada modelo, com 20 imagens geradas no total, variando a semente aleatória para promover diversidade. As imagens geradas foram então avaliadas segundo critérios de fidelidade ao estilo impressionista e consistência visual.

3.2. Métricas de Avaliação

A avaliação objetiva foi conduzida por meio das seguintes métricas:

- **CLIPScore** [Hessel et al. 2021]: mede a compatibilidade semântica entre o prompt textual e a imagem gerada, baseado nas embeddings do modelo CLIP.
- **FID (*Fréchet Inception Distance*)** [Heusel et al. 2017]: avalia a distância entre as distribuições de características extraídas das imagens geradas e reais, usando o Inception v3. Quanto menor o valor, maior a similaridade.

Além disso, foi realizada uma análise qualitativa visual das imagens, observando aspectos como pinceladas, uso de cor e composição, características típicas do estilo de Claude Monet.

3.3. Resultados Quantitativos

A Tabela 3 apresenta os valores de CLIPScore para os dois modelos (base e *fine-tuned* com LoRA), considerando dois cenários de avaliação: com o *prompt* original utilizado na geração da imagem e com um *prompt* padronizado contendo descrições estilísticas associadas a Claude Monet. A média de CLIPScore usando os *prompts* originais foi de 30.04 para o modelo base e 32.95 para o modelo LoRA, indicando que o *fine-tuning* contribuiu para uma maior fidelidade semântica às descrições fornecidas, mesmo sem menções explícitas ao estilo impressionista durante a geração. Encontra-se uma visualização gráfica desses dados na Figura 2

Ao se utilizar o prompt estilizado com termos como “*impressionist style*” e “*in the style of Claude Monet*”, os resultados se invertem: o modelo base obteve uma média superior (34.81) em relação ao modelo LoRA (31.23). Esse resultado indica que o modelo base responde melhor quando há instruções textuais explícitas sobre o estilo de Monet,

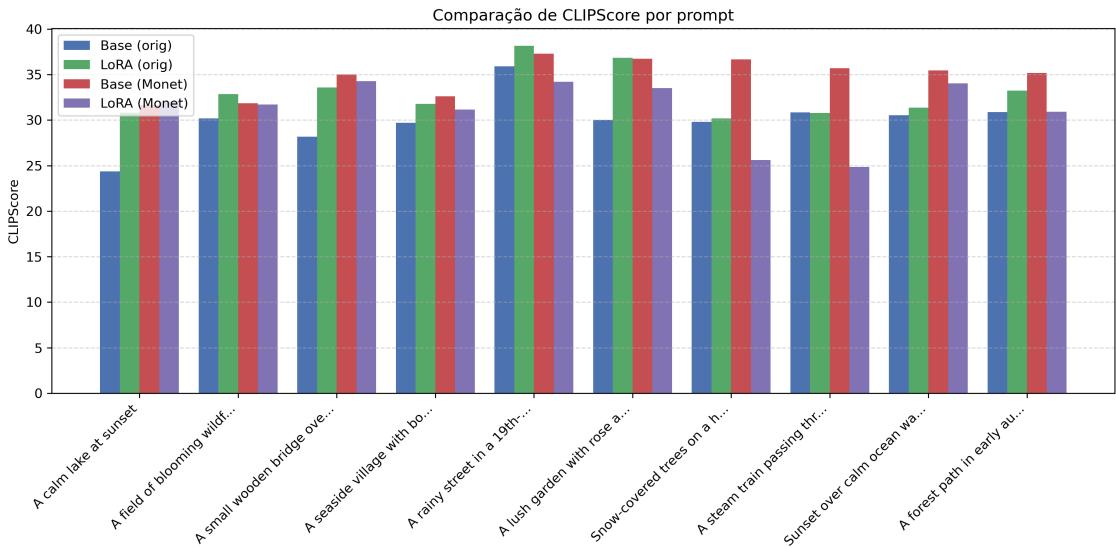


Figura 2. Representação gráfica dos CLIPScores

Fonte: Autoria própria

enquanto o modelo LoRA foi treinado para expressar o estilo aprendendo diretamente das imagens, sem depender de instruções textuais. Ou seja, o modelo LoRA gera imagens mais coerentes com *prompts* neutros, mas não necessariamente se alinha melhor com descrições textuais ricas em estilo, já que esse viés já está internalizado em sua representação visual.

Além disso, a métrica FID entre os dois conjuntos de imagens foi de 305.10, indicando uma grande diferença visual entre as imagens geradas pelos modelos. Este valor reforça que o *fine-tuning* com LoRA resultou em imagens com características visuais significativamente distintas — e mais próximas ao estilo de Monet — em comparação ao modelo base. Em conjunto, os resultados apontam que o *fine-tuning* com LoRA não apenas melhora a fidelidade semântica em *prompts* genéricos, mas também altera substancialmente o espaço visual do modelo para capturar o estilo-alvo, mesmo sem instruções explícitas.

3.4. Análise Qualitativa

As Figuras 3 e 4 apresentam uma comparação visual entre amostras geradas pelas duas abordagens. Observa-se que o modelo *fine-tuned* com LoRA demonstra dificuldade em manter o estilo de Monet em alguns exemplos, como o do *prompt* “*A steam train passing through a countryside station*”, onde ele adota um estilo mais próximo do realismo. Isso reflete a ausência de trens no *dataset* de pinturas de Monet, resultando na falta de representação de objetos do tipo, fazendo com que o modelo opte pela representação já aprendida em sua versão base.

As imagens geradas pelo modelo base, por outro lado, representam uma visão uniformizada do estilo impressionista de Monet, podendo ser explicado pelo fato de hoje o autor ser um grande representante do movimento impressionista, quando na verdade as pinturas do autor incluíam outros estilos variados não necessariamente impressionistas, melhor representados na coluna das imagens geradas com o modelo ajustado com LoRA

As imagens geradas pela abordagem de engenharia de *prompts* demonstram maior capacidade de generalização do estilo, refletindo a flexibilidade da técnica, embora limitada pela falta de especialização do modelo. A abordagem com LoRA, por outro lado, denota maior fidelidade ao estilo do autor, porém com uma notável dificuldade em generalizar o seu estilo para aplicações novas.

Tabela 2. Prompts para comparação entre os modelos com e sem fine-tuning

Prompt (modelo com LoRA)	Prompt (modelo base com engenharia)
<i>A calm lake at sunset</i>	<i>A calm lake at sunset, in the style of Claude Monet, with soft brushstrokes and warm impressionist tones</i>
<i>A field of blooming wildflowers under a cloudy sky</i>	<i>A field of blooming wildflowers under a cloudy sky, painted in the impressionist style of Monet</i>
<i>A small wooden bridge over a quiet river</i>	<i>A small wooden bridge over a quiet river; Claude Monet style, with light reflections and vibrant pastels</i>
<i>A seaside village with boats on the shore</i>	<i>A seaside village with boats on the shore, impressionist painting, in Claude Monet's style</i>
<i>A rainy street in a 19th-century European town</i>	<i>A rainy 19th-century European street, painted with dabs of color and soft outlines, in Claude Monet's style</i>
<i>A lush garden with rose arches and green hedges</i>	<i>A lush garden with rose arches and green hedges, in the style of Claude Monet, impressionist garden scene</i>
<i>Snow-covered trees on a hill in winter</i>	<i>Snow-covered hill with trees, impressionist painting, soft brushwork, Claude Monet winter style</i>
<i>A steam train passing through a countryside station</i>	<i>A steam train in a countryside station, impressionist style, inspired by Monet</i>
<i>Sunset over calm ocean waves with orange glow</i>	<i>Sunset over ocean, impressionist style with glowing pastels, Claude Monet inspired</i>
<i>A forest path in early autumn</i>	<i>A forest path in early autumn, impressionist colors and diffuse lighting, Monet style</i>

Fonte: Autoria própria

Tabela 3. CLIPScore com prompt original e com estilo Monet para os dois modelos

Prompt	Base (orig)	LoRA (orig)	Base (Monet)	LoRA (Monet)
A calm lake at sunset	24.34	30.81	31.62	32.02
A field of blooming wild-flowers under a cloudy sky	30.17	32.85	31.83	31.71
A small wooden bridge over a quiet river	28.18	33.59	35.01	34.28
A seaside village with boats on the shore	29.71	31.79	32.60	31.17
A rainy street in a 19th-century European town	35.91	38.15	37.27	34.21
A lush garden with rose arches and green hedges	30.02	36.83	36.74	33.52
Snow-covered trees on a hill in winter	29.80	30.17	36.67	25.61
A steam train passing through a countryside station	30.84	30.78	35.69	24.84
Sunset over calm ocean waves with orange glow	30.54	31.35	35.46	34.02
A forest path in early autumn	30.87	33.22	35.19	30.92
Média	30.04	32.95	34.81	31.23

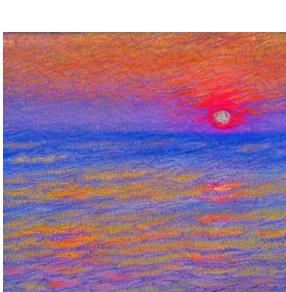
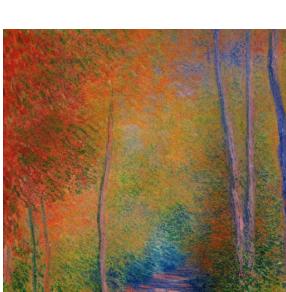
Fonte: Autoria própria

Figura 3. Análise qualitativa (Parte 1): imagens geradas por ambos os modelos para cada *prompt*

Prompt	Base	LoRA
A calm lake at sunset		
A field of blooming wild-flowers under a cloudy sky		
A small wooden bridge over a quiet river		
A seaside village with boats on the shore		
A rainy street in a 19th-century European town		

Fonte: Autoria própria

Figura 4. Análise qualitativa (Parte 2): imagens geradas por ambos os modelos para cada *prompt*

<i>Prompt</i>	Base	LoRA
A lush garden with rose arches and green hedges		
Snow-covered trees on a hill in winter		
A steam train passing through a countryside station		
Sunset over calm ocean waves with orange glow		
A forest path in early autumn		

Fonte: Autoria própria

4. Conclusão

Neste trabalho, foram comparadas duas abordagens para controle de estilo em geração de imagens: engenharia de *prompts* aplicada a um modelo generalista e *fine-tuning* com LoRA para especialização do modelo no estilo impressionista de Claude Monet. Os experimentos realizados demonstraram que a técnica de *fine-tuning* com LoRA consegue internalizar o estilo desejado, permitindo a geração de imagens coerentes com *prompts* genéricos, sem necessidade de menções explícitas ao estilo. Por outro lado, a engenharia de *prompts* apresentou melhor desempenho quando o estilo era especificado diretamente na descrição.

As avaliações quantitativas e qualitativas indicam que o *fine-tuning* com LoRA altera significativamente o espaço visual do modelo, ainda que apresente limitações em generalizar para elementos pouco representados nas pinturas. A engenharia de *prompts*, apesar da sua maior flexibilidade, é menos consistente em expressar o estilo de forma automática, exigindo descrição detalhada das características desejadas.

Assim, é possível que a combinação dessas abordagens seja promissora: o *fine-tuning* com LoRA para a internalização do estilo, aliado à engenharia de *prompts* para ajustes finos e maior controle. Futuras pesquisas podem explorar a integração destas técnicas, criando modelos mais eficientes em representação de estilo.

Referências

- (2025). The complete works. <https://claudemonetgallery.org/the-complete-works.html>. Accessed: 2025-06-06.
- Gugger, S., Debut, L., Wolf, T., Schmid, P., Mueller, Z., Mangrulkar, S., Sun, M., and Bossan, B. (2022). Accelerate: Training and inference at scale made simple, efficient and adaptable. <https://github.com/huggingface/accelerate>.
- Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., and Choi, Y. (2021). Clipscore: A reference-free evaluation metric for image captioning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7514–7528.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, 30. Introdução do FID.
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. (2021). Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Lhoest, Q., Villanova del Moral, A., Jernite, Y., Thakur, A., von Platen, P., Ma, C., Drame, J., Plu, J., Patil, S., Chaumond, J., Sanh, V., and Wolf, T. (2021). Datasets: A community library for natural language processing. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 175–184, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, pages 8024–8035.
- Richardson, L. (2023). Beautiful soup documentation. Accessed: 2025-06-09.
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. CompVis/stable-diffusion-v1-4.
- von Platen, P., Davaadorj, M., Rasul, K., and Wolf, T. (2022). Diffusers: State-of-the-art diffusion models. Hugging Face.