LoRA-Enhanced Diffusion-Based Framework for Automated Comic Book Image Generation

Devadarshini P[1], Dr. Mary Louis L [2]
[1] *Student, Department of Data Science, Kumaraguru College of Liberal Arts and Science, Coimbatore, India*
[2] *Head of Department, Department of Mathematics, Kumaraguru College of Liberal Arts and Science, Coimbatore, India*

***Abstract*— The rapid progress of generative artificial intelligence has enabled automated image synthesis from natural language descriptions. However, producing domain-specific artistic content with a consistent visual style remains a challenging task. Comic book illustration demands expressive characters, bold line art, coherent visual themes, and stylistic consistency across generated images. This paper presents a LoRA-enhanced diffusion-based framework for automated comic book image generation. It transforms textual prompts into high-quality comic-style illustrations. The proposed system is built on a latent diffusion architecture and uses Low-Rank Adaptation to efficiently specialize the model for comic-specific visual characteristics, without retraining the entire network. By introducing parameter-efficient adaptations, the framework achieves improved style control, reduced computational overhead, and enhanced consistency compared to conventional diffusion-based generation. The results indicate that integrating diffusion models with LoRA provides an effective and scalable solution for creative content generation and digital storytelling applications.**

Keywords*— Comic Book Generation, Diffusion Models, Low-Rank Adaptation, Text-to-Image Synthesis, Style Consistency, Creative AI

I. INTRODUCTION

Recently, breakthroughs have emerged in the field of generative artificial intelligence. These advances make it possible to automatically generate images from descriptions given in natural language. Within this category, comic book illustration poses challenges. It involves expressive figures, bold line work, and highly visual identities. Manual comic creation is extremely time-consuming and requires significant skill. This has prompted the construction of intelligent systems to help facilitate the process.

Diffusion models have appeared as a prominent line of research in text-to-image synthesis, providing better visual quality and training stability compared to traditional generative models. Through refinement of noise input into structured images based on text guidance, diffusion models can generate high-quality and semantically rich outputs. But diffusion-based models trained on large general-purpose datasets tend not to produce domain-specific artistic styles. When considering comic book illustration as a domain of interest, this becomes a problem of inconsistent characters' appearance, a lack of style consistency, and a produced style that is more focused on photorealistic content compared to comic book illustration.

Fine-tuning large diffusion models for the target specialized creative domain would often require heavy fine-tuning of the model parameters, incurring high computational costs. Nevertheless, parameter-efficient fine-tuning approaches have been developed as alternatives to the above challenge. Low-Rank Adaptation (LoRA) is one of the most popular solutions in this pursuit. It incorporates learnable components in the models to enable the models to learn domain-specific traits with only a slight increase in computational cost.

In this work, an AI system that utilizes LoRA-upgraded diffusion models to create comic book images from text descriptions is proposed. The proposed approach focuses on adapting a diffusion backbone to specialize in generating comic illustrations based on text inputs while adapting to the desired styles at a feasible cost in terms of computations. With LoRA adaptation enabling adaptive comic image generation, this proposed system clearly validates that adapting diffusion models using a reduced number of parameters can be an optimal approach to automatically generate comic illustrations.

II. BACKGROUND

A. Diffusion-Based Image Generation

Diffusion models have emerged as a popular approach for high-resolution image generation due to their robust training and good generation capabilities. The diffusion models work by performing an iterative cleaning operation of the noise samples based on certain additional information, such as text features. The latent diffusion models have also optimized the diffusion step for faster performance.

Text-conditioned diffusion models are very commonly utilized for general-purpose image generation. However, when the diffusion models are trained on large and diverse datasets, they learn about general visual distributions, but they may no longer possess the capability to impose strong visual constraints related to artistic disciplines.

B. Characteristics of Comic-Style Image Generation

Comic book illustrations are quite different from photograph-like picture creation. The comic book style uses dramatic line work, minimalistic shading, proportion manipulation, facial exaggeration, and high contrast colors. Apart from that, repeatability and consistency among various pictures are necessary considerations for comic books so that a story can be made.

General diffusion models tend to generate results with mixed example and photo-realistic characteristics, resulting in inconsistent character identity and loss of stylistic identity. Such challenges make it clear that the required adaptation at the domain level could ensure effective comic-style image generation.

C. Challenges in Domain-Specific Adaptation

The transfer of large diffusion models to specialized fields has generally been associated with the fine-tuning of model parameters. Although it is an efficient method, it is computationally expensive and restricted by scalability in scenarios requiring different styles. These constraints become even more limited in creative fields where rapid experimentation under limited hardware environments is often preferred.

Therefore, there is a need for efficient adaptation techniques that will allow domain specialization without the need for model retraining or the involvement of multiple large model instances.

D. Parameter-Efficient Fine-Tuning with LoRA

Low-Rank Adaptation (LoRA) alleviates these issues by adding learnable low-rank matrices to some layers of an existing model while freezing all model parameters. This further saves on cost and memory during training. LoRA makes it possible to learn styles such as line art, color abstraction, and illustrative elements in diffusion-based image generation. The integration of LoRA in diffusion pipelines offers a sensible starting point for the efficient production of comic-style images. This motivates the proposed approach.

III. RELATED WORK

A. Text-to-Image Generation Using Diffusion Models

Diffusion-based generative models have reached state-of-the-art performance in text-to-image synthesis tasks because they leverage the ability to better align image output with the description given in text form. Latent diffusion models improve scalability and require fewer computations with similar levels of image realism. Such models are the key to most image generation techniques.

Although these models have been highly successful, they have not been able to fulfil the constraints needed for artistic works even after being trained on general datasets. Their results have always been more realistic than abstract, which is not suitable for their application in illustration work, for instance, comic creation.

B. Artistic and Stylized Image Generation

Investigated areas related to the creation of artistic images include neural style translation and text-conditioned generative models. Though style translation tools are successful at transferring texture and color patterns, they do not provide enough semantic expressiveness for creating artistic images from text descriptors. Current generative techniques offer more flexibility but can find it challenging to provide uniform visual style access.

In a comic, for example, consistency in character depiction and style is needed. Currently, methods in use often produce outcomes with diverse qualities in visual elements and a lack of repetition needed in comics.

C. Domain Adaptation and Fine-Tuning of Generative Models

Conventional fine-tuning techniques for adapting the generative models involved updating most of the parameters in the models. Consequently, the models consumed much computational power. More recent approaches for adapting the models with fewer computations required the storage of multiple models corresponding to various styles/domains.
Such issues necessitate the need to apply 'parameter-efficient' approaches that make it possible to specialize in a scalable manner without compromising on performance.

D. Parameter-Efficient Fine-Tuning and LoRA

Methods such as LoRA have recently been introduced for parameter-efficient fine-tuning of large models with low overhead. LoRA allows domain adaptation to be performed with a low rank, which helps in preserving the generalization ability of the base model. Although LoRA has been successfully used in various generative modeling applications, there is less work in the aspect of comic image creation using LoRA. Specifically, there is no in-depth study about the ability of LoRA in cooperation with diffusion models for the generation of style-consistent comic illustrations based on text descriptions.

E. Research Gap

Currently, there is no unified approach that combines diffusion-based text-to-image synthesis and parameter-adaptive adaptation specifically for creating comic book illustrations. This paper will fill that gap with its approach that aims to enhance diffusion using LoRA for efficient comic image synthesis that is coherent with styles.

IV. SYSTEM ARCHITECTURE AND METHODOLOGY

A. System Overview

The proposed system is a LoRA-boosted diffusion-based network intended for comic-style image generation based on natural language descriptions. The network consists of three main components: a text conditioning network, a diffusion-based backbone network, and parameter-efficient LoRA adaptation layers that are designed specifically for comic-style image generation. The system aims to achieve a tradeoff between style consistency, controllability, and efficiency to run on consumer-grade hardware.
Overall, the user input text prompts are converted into embeddings that give their input to the diffusion process, and LoRA adapters are used to adjust specific layers of the diffusion model to include comic stylistics, with the general image synthesis ability retained through the base model being frozen.

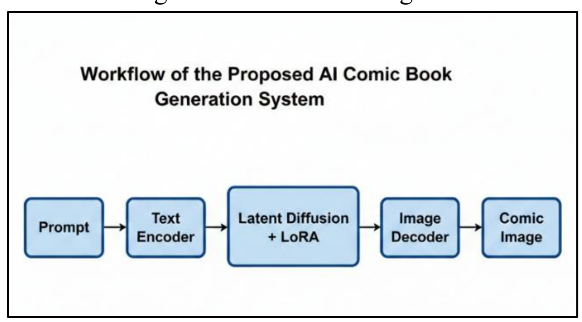

Figure 1. Workflow of the proposed LoRA-enhanced diffusion-based comic book image generation system.

B. Text Encoding and Prompt Conditioning

Natural language prompts undergo initial processing through a text encoder that is linked with a diffusion model. The text encoder converts a description of images into dense vectors to control the process of generating images. To make the style more realistic and comic-like, natural language prompts undergo augmentation with comic-style descriptors like line art, soft shading, and comic-style facial expressions. Negative prompts are used to remove unwanted visual attributes, such as photorealistic textures, too much shading, and complicated background elements that go against the aesthetic nature of comics. This conditioning approach via prompts improves semantic alignment while also enforcing stylistic constraints without needing extra structural inputs.

C. Latent Diffusion Backbone

The underlying task of image generation is achieved by a latent diffusion model, which is pre-trained. Differently, instead of diffusing in the pixel domain, a diffusion process takes place in the latent space, where it consumes relatively lower memory compared to diffusions in the pixel domain. The diffusion process denoises the variables, given the text embedding, until an image representation is synthesized. The base diffusion model is fixed throughout the training and evaluation process. This will ensure robustness to changes in the visual world and prevent degradation in general visual knowledge.

The base diffusion model is fixed to tap its benefits from expressive pretrained models while incurring minimal adaptation cost.

D. LoRA-Based Comic Style Adaptation

Low-Rank Adaptation (LoRA) modules are added to specific attention layers and convolutional layers of the diffusion model. The adapters are made of low-dimensional trainable weight matrices that encode the comicy-specific style elements of bold outlines, simplified shading, comicy-proportions, and high-contrast colors.

In training, only the LoRA parameters are learned, and the base model's weight remains constant. In terms of memory saving and training speed, using LoRA is much better than the entire model fine-tuning. The model's LoRA components can be trained to embody different comic styles or character designs and then loaded dynamically at runtime to avoid redundant base model storage in memory.

E. Training Procedure

The LoRA adapters are trained on a carefully curated dataset that consists of comic-style illustrations ranging from character illustrations to scene compositions and panel style illustrations. The training examples are accompanied by text descriptions that convey semantic as well as aesthetic information. The image data is resized to match the input requirements of the diffusion model.

The objective function that gets minimized is based on standard diffusion loss, and there is a constraint on the updates of LoRA parameters. itsxed-precision training is used to make its memory-efficient and to reduce GPU memory usage during training. LoRA is lightweight; thus, it has fast convergence, and it's easy to try various adaptations of styles.

F. Inference Pipeline

In the inference phase, the model where the inference tasks are performed loads a pre-trained diffusion model, as well as a LoRA layer adapter of a desired comic style. Comic-style images are produced in the inference phase using a diffusion process where the diffusion model performs a latent space denoising procedure.

In addition to the steps, the latent space image is further decoded into an image in the inference phase using a 'pretrained decoder.' Such a modular inference pipeline facilitates fast style transfer, good resource management, and scalability. Furthermore, this work enables flexible and reproducible comic image creation geared towards digital storytelling, concept art, and creative exploration through detached style transfer from a pre-trained base model.

V. EXPERIMENTAL SETUP

A. Dataset Description

The LoRA adapters are trained with a specially prepared dataset of illustrations of a comic style that was taken from publicly available sources of these kinds of illustrations. The dataset contains portraits of characters, action scenes, as well as panel illustrations with general characteristics of comic illustrations, for instance, prominent outlines of figures, simplified shading, proportionally distorted human figures, and highly contrasting colors of figures with backgrounds. While ensuring that stylistic variations are minimal, only illustrations with realistic rendering of figures with sufficient simplicity of backgrounds are left out of consideration.

For every image, a corresponding text description was assigned, which conveyed both semantic information and stylistic characteristics. The resized images were normalized to match the input resolution of the diffusion part of the backbone. The structure of the dataset allows for the learning of comic-style features while being properly aligned with the text and image outputs.

B. Training Configuration

The model makes use of a latent diffusion model that has been pretrained as the base model of the generator; Low-Rank Adaptation (LoRA) layers are utilized to specialize in comic styles. While training the model, all parameters in the base model were set to remain frozen, and only the LoRA parameters were updated.

The training was performed with mixed precision, where training efficiency was enhanced, along with reduced GPU memory requirements. Standard diffusion loss functions were used, where training was done until convergence was achieved for the stylistic features. The lightness of LoRA allowed for rapid experimentation with ease of use of the available hardware.

C. Implementation Details

It was developed using a state-of-the-art deep learning technology stack that utilizes the PyTorch library along with a diffusion model library. It was trained and tested on a single consumer-level GPU. It also features the ability to load the LoRA adapter dynamically so that the comic styles can be changed without the need for a second diffusion model.

Prompt conditioning involved the use of both positive attributes to encourage comic-style painting and negative attributes to discourage photorealistic effects. The prompt approaches were maintained consistently through the experiments for the purpose of facilitating qualitative comparisons.

D. Evaluation Criteria

The evaluation focuses on qualitative and diagnostic quantitative assessment of generated comic images. The primary evaluation criteria include:

- Style Consistency: Visual uniformity across generated images with respect to line art, shading, and color usage
- Prompt Alignment: Correspondence between textual descriptions and generated visual content
- Character Expressiveness: Ability to depict exaggerated expressions and dynamic poses typical of comic art
- Computational Efficiency: Memory usage and inference behaviour resulting from LoRA-based adaptation

Due to the artistic nature of comic illustration, evaluation emphasizes visual coherence and stylistic fidelity rather than photorealistic accuracy.

E. Experimental Protocol

Experiments were performed by producing comic-style images based on a broad range of text inputs that involved characters, activities, and scenarios. The inference settings for the inference task were uniform for all the text inputs. The comic-style image generation using the diffusion model without adaptation and with adaptation using LoRA was qualitatively analyzed to determine the effectiveness of adaptation in comic-style image generation.

Table I. Model and Training Configuration

Component	Description
Base Model	Latent Diffusion Model
Adaptation Method	Low-Rank Adaptation (LoRA)
Trainable Parameters	LoRA adapters only
Precision	Mixed precision (FP16)
Prompt Conditioning	Positive + Negative prompts
Output Style	Comic-style illustration

VI. RESULTS AND ANALYSIS

A. Visual Quality and Comic Style Consistency

The diffusion framework LORA-boosted shows a clear enhancement over the diffusion baseline in terms of the comic-style image features. The generated images portray thick outline drawings, basic shading effects, cartoonish ratio characteristics, and bright contrasting colors appropriate for a comic drawing setting. However, the diffusion baseline output often illustrates realistic texture effects and a variety of render settings that tend not to be ideal and appropriate for a comic artifact.

On different question sets, the LoRA-adjusted model has an overall visual consistency, thereby showing that the adjusted low rank representations learn domainshift features successfully. For the comic generation task, overall visual consistency directly affects the continuity of narratives presented, thereby immersing readers.

B. Prompt Alignment and Semantic Accuracy

The proposed system can effectively translate text-based prompting messages into semantically significant comic illustrations. The semantically significant comic illustrations produced by the system depict necessary attributes of characters and actions as specified in the prompting messages. The use of task-specific prompting message descriptors and negative prompting messages improves semantic alignment.

Although it works well with single-scene and character-related tasks, sometimes narrative descriptions involving complex scenarios lead to partial omissions or simplified descriptions. This could be seen as an extension of the general limitations related to text conditioning image generation, with future needs focusing on story-aware conditioning.

C. Character Expressiveness and Artistic Abstraction

Comic illustration is very expressive in terms of face expressions, character actions, and symbols used. The capability of the LoRA-adaptive model is better in terms of character expression compared to the diffusion model. Character faces are more caricaturized in terms of their expression, as well as the body actions, with higher levels of abstraction.

Nevertheless, replicating character identity with a degree of accuracy across several generations is a tough task. Although stylistic elements are retained, more specific identity features like faces or costumes can be altered in each image, suggesting a restriction in retaining character identity.

D. Impact of LoRA on Computational Efficiency

As updates are restricted to LoRA adapters for training and prediction, this approach saves a substantial amount of computation compared to fine-tuning CNNs. The memory consumed is constant, while the time taken to train is greatly reduced because of the small number of parameters to learn for each of these fine-tuning tasks.

This allows various comic styles to be modeled by different LoRA adapters without having to replicate the underlying diffusion model, thereby further enhancing scalability and flexibility. These requirements are met by the proposed solution, which should be helpful in creative tasks needing rapid iteration and switching among styles.

E. Comparative Observations

Comparison between diffusion-based output at the baseline level and LoRA-improved diffusion-based output reveals that there is serious importance in domain-specific adaptation in artistic creation. Even if the diffusion-based output at the baseline level reveals a realistic image with a lot of details, there is no artistic abstraction in the image that would enable the creation of a comic strip.

F. Limitations Observed in Results

Despite the progress made by the LoRA adaptation approach, there are some drawbacks that still prevail. It may have some disparities in terms of size or location in complex multi-character scenes, along with less control to arrange panels in the current approach. It also needs some approaches to maintain character identity in the case of consistency in comic sequences. These constraints point to various ways in which future improvements could be made, such as embedding persistence, panel conditioning, and structured layout constraints.

While no absolute quantitative benchmark is reported due to the artistic nature of comic illustration, qualitative consistency across prompts was consistently observed.

VII. DISCUSSION

A. Effectiveness of LoRA for Comic-Style Generation

The above experiments demonstrate that the Low-Rank Adaptation technique is an efficient method for adapting diffusion models for comic-style image generation. By learning a small set of adaptational parameters while keeping the learned representations unchanged, the technique has been able to obtain comic-style features such as bold lines, caricature facial expressions, and illustrative abstraction while keeping the primary generative function of the pre-trained diffusion model unchanged. It proves that adaptation for different styles can be achieved independently of common visual knowledge for efficient reuse of generic large-scale generative models.

Compared to full fine-tuning methods, the proposed method achieves a great balance between visual quality and efficiency based on LoRAs. The results obtained in style consistency demonstrate that low-rank adaptations are adequate to represent artistic attributes from the target domain; thus, they are very valuable in scenarios in which style repetition and controllability have greater significance than photorealistic details.

B. Implications for Creative Workflows

The proposed framework has practical implications for digital artists, storytellers, and independent creators. With its ability to help create comic-style images with text-driven inputs while sustaining visual identity, this framework can be used as an ideation assistant for quick idea creation before going into production. This can also be used to jump across styles easily using LoRA adaptors, making it easier for artists or creators to test multiple styles without extensive re-training.

Importantly, the system is meant to work in conjunction with creativity rather than in place of creativity. The images can be perfected, edited, or merged by artists to build a narrative, which puts the system in a supportive role in the pipeline of creativity.

C. Limitations and Challenges

Despite the efficacy of the framework, the model has some limitations that are typical of present diffusion models for generating images from text. It is hard to preserve the character identity within the large set of images generated through the model since the character identity attributes might change even when the style is preserved. It is also difficult to create complex scenarios through the generated characters.

Another limitation exists in the lack of narrative and panel-level conditioning. For instance, although images relate well with semantics in a prompt, modelling dependencies in a narrative sequence, which is needed in a comic book, has yet to be accomplished. Handling these issues will demand going beyond prompt-level conditioning.

D. Comparison with Existing Approaches

Compared with general text-to-image diffusion models, the proposed method shows better appropriateness for comic illustrations, concerning their emphasis on style abstraction and uniformity. Compared with traditional methods of artistic style transfer, the system provides a higher degree of semantic freedom because images are not generated from existing content but from text descriptions.

In contrast to other methods based upon full model fine-tuning or multiple style models, the “parameter-efficient” architecture enables scalable and rapid adaptation. Such a model, therefore, has the potential to serve as a competitor in domains where creative performance and efficiency are required.

E. Directions for Future Research

Future research could investigate ways to maintain the identity of character generations, possibly using character embeddings or memory-based conditioning. The inclusion of panel-level layout constraints and story-conditioning may also help to improve overall story consistency. Further extension to enable multi-panel generation with an interface for editing would bring it even closer to fully automatic comic book generators.

More quantitative assessment and even human perceptual studies could offer a richer understanding of artistic quality and user satisfaction. They offer new and fruitful directions for research in comic assistance with AI.

VIII. APPLICATIONS AND ETHICAL CONSIDERATIONS

A. Applications of Automated Comic Generation

The proposed diffusion framework with the application of LoRA is of use in various realms of creative as well as educational purposes. The most fundamental use of the diffusion framework would be found in the realm of digital comic production. In this context, the diffusion framework can be of use to comic artists as well as writers in the initial stages of their creativity in terms of quickly generating character concepts or scenes using text descriptions.

The framework can further be used for digital storytelling in web comics, graphic novels, and interactive media. With a consistent comic-style image provided by the system, storyboards can be created. In educational institutions, the system can be used for generating learning modules in a pictorial format.

Additionally, the proposed approach applies to game and animation pre-visualization, where comic-style imagery is often used during early design phases. The parameter-efficient nature of LoRA allows multiple artistic styles to be maintained and switched efficiently, making the framework suitable for small studios and independent creators with limited computational resources.

B. Ethical Considerations and Responsible Use

Ethics pertaining to authorship, originality, and use come into play regarding the use of generative AI in creative tasks. As intended, the system should be used for purposes of creativity assistance and not as a replacement for human artists. Images created from the comic generators should be seen as work in progress that could be perfected and put into perspective by the artist.

Provenance of the training data is a significant concern in this moral realm. Though the work uses pre-trained diffusion models and comic-style datasets, there should be a concern for satisfying copyright and licensing issues. Future applications should be transparent in their sources for the training data and not engage in unauthorized copying of recognizable styles of artwork attributed to artists.

Bias and representation fairness are other things to consider. The generative model may end up carrying forward biases inherent in the data it uses to train, which could impact character representation, culture representation, or style. Finally, responsible usage needs to convey clear information about system constraints. While the framework does not promise anything regarding story consistency, character integrity over a sequence of scenes, or publishing requirements, it is important to convey this information to prevent abuse.

IX. CONCLUSION

This paper introduced a LoRA-updated diffusion model framework for automatically creating comic book images from natural language prompts. With the employment of the Low-Rank Adaptation technique, which is a parameter-efficient addition to a latent diffusion model, it was possible to create comic-style images with enhanced stylistic coherence, representational abstraction, and efficiency. This work directly addresses several interesting issues associated with comic book drawing, such as requiring a controlled artistic style with a predictable visual identity, while still being flexible enough to be text-driven.

Experimental results showed that the adapted diffusion model through LoRA adapted well to the task of generating coherent comic-style artworks that closely matched the semantics of the prompt and had higher adherence to the desired style than a general diffusion model. The outcome indicates the efficacy of parameter-level adaptation at learning domain-specific artistic traits from a generative model without having to train the entire model from scratch.

While the current system focuses on single-image generation, the findings establish a foundation for more advanced comic creation pipelines. Future work may extend the framework to support multi-panel layouts, persistent character identity, and narrative-aware conditioning, enabling end-to-end automated comic storytelling. Overall, this work demonstrates that LoRA-enhanced diffusion models offer a practical and scalable solution for AI-assisted comic book illustration and creative digital storytelling.

REFERENCE

[1] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851, 2020.

[2] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-Resolution Image Synthesis with Latent Diffusion Models,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10684–10695, 2022.

[3] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “LoRA: Low-Rank Adaptation of Large Language Models,” in Proceedings of the International Conference on Learning Representations (ICLR), 2022.

[4] L. Zhang, A. Rao, and M. Agrawala, “Adding Conditional Control to Text-to-Image Diffusion Models,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3836–3847, 2023.

[5] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps," Advances in Neural Information Processing Systems, vol. 35, pp. 5775–5787, 2022.
[6] T. Dao, D. Fu, S. Ermon, A. Rudra, and C. Ré, "FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness," Advances in Neural Information Processing Systems, vol. 35, pp. 16344–16359, 2022.
[7] P. Gatys, A. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," Journal of Vision, vol. 16, no. 12, pp. 326–326, 2016.
[8] X. Huang and S. Belongie, "Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1501–1510, 2017.
[9] A. Radford, J. W. Kim, C. Hallacy, et al., "Learning Transferable Visual Models from Natural Language Supervision," in Proceedings of the International Conference on Machine Learning (ICML), pp. 8748–8763, 2021.
[10] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 586–595, 2018.