Research

Generative AI: Diffusion Models

Suyog Avhad, Shivaditya Baghel, Saurav Chavan, Kunal Nikam
Department of Information Technology, The Bombay Salesian Society's
Don Bosco Institute of Technology , Mumbai-400070

2023 - 2024

Abstract

In this project, we show how to fine-tune Stable Diffusion for generating domain-specific images, focusing on the untrained Rick and Morty dataset from Hugging Face. By integrating Low-Rank Adaptation (LoRA), our approach gives more efficient optimization, faster convergence, and improved generalization, thereby significantly reducing training time. Fine-tuning Stable Diffusion with LoRA improves adaptability to diverse datasets, making it a suitable asset. By using LoRA's adaptive learning rate strategy, our model flexibly adjusts to the nuances of the Rick and Morty dataset, resulting in detailed and contextual image generation. Our model can help animators and content creators use Generative AI to generate captivating images with minimal computational overhead. Our work not only shows new improvements in Text-to-Image Generation but further research into it can revolutionize animation by enabling creators to visualize ideas with ease and fidelity.

Methodology/Architectural Design

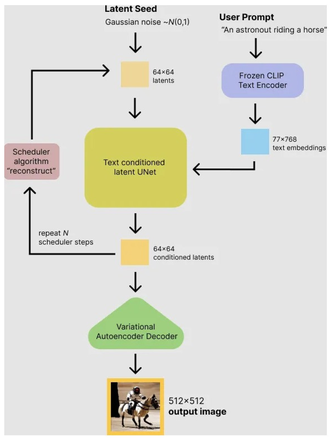

Fig 1. Inference for Stable Diffusion.
[6] Mishra, O. “Stable Diffusion Explained.” Medium. 2022.

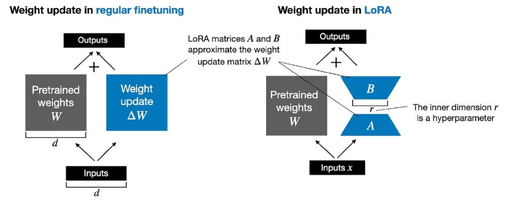

Fig 2. LoRA
[7] Raschka, S. “Practical Tips for Fine-Tuning LLMs.” Sebastian Raschka's Newsletter. 2022.

Results and Conclusion

Prompt: Rick and Morty in a swimming pool

Output :

Prompt: Rick and Morty in space

Output :

Conclusion: In our study, we conclude that Cosine learning scheduler with increased training steps, generates optimal results. Our study underscores the importance of fine-tuning models like Stable Diffusion 2.0 with specific datasets to unlock their full potential in image generation tasks.

Background Information

In recent years, cutting-edge image generation tools like Stable Diffusion and DALL-E are revolutionizing the field with their unique abilities. Stable Diffusion helps in generating high-quality images using diffusion processes, while DALL-E is used in generating images from text descriptions. Other new tools like ReCo, GLIGEN, and SceneComposer offer help in various tasks like multi-object generation and attention map enhancement. In our study we found that Low-Rank Adaptation (LoRA) has improved diffusion models further, making it faster and more adaptable. With LoRA, models can adjust their learning rates dynamically, improving image quality and making them more useful for tasks like text-to-image generation and medical imaging. These advancements in tools is pushing the boundaries of generative modeling, creating new possibilities in creative content generation.

Significance

Low Rank Adaptation (LoRA) Technique:
Weight Update:

$$W_{updated} = W + \Delta W$$

Problem:
Computing Δ*W* with 7B parameters is computationally and memory-intensive

LoRA Solution:
Compose ΔW into lower-rank matrices A and B:
ΔW = AB

Importance of Rank r in LoRA:

- Memory Saving

1. For Δ*W* with 10,000 rows and 20,000 columns (200M parameters),
2. Choosing *r*=8 reduces parameters to 240,000,
3. Saving about 830x memory.

References

[1] Hoogeboom, Emiel, et al. ”Stable diffusion models for image generation.” Advances in Neural Information Processing Systems. 2021.

[2] Tay, Yi, et al. ”Long-range arena: A benchmark for efficient transformers.” arXiv preprint arXiv:2109.01940. 2021.

[3] Brown, Tom B., et al. ”Language models are few-shot learners.” Advances in Neural Information Processing Systems. 2020.

[4] Goodfellow, Ian, et al. ”Generative adversarial nets.” Advances in Neural Information Processing Systems. 2014.

[5] Kingma, Diederik P., and Max Welling. ”Auto-encoding variational bayes.” International Conference on Learning Representations. 2014.

[6] Mishra, O. “Stable Diffusion Explained.” Medium. 2022.

[7] Raschka, S. “Practical Tips for Fine-Tuning LLMs.” Sebastian Raschka's Newsletter. 2022.