

# MoFu: Model independent, Fast Tuning of Stable Diffusion concepts.

[Mizuhashi Parsee / yoinked](#)<sup>1,2</sup>

<sup>1</sup>[Touhou AI Experimental Group](#), <sup>2</sup>[M.O.F.U.](#)

[\[Github\]](#)

## 1. Abstract

I present MoFu, a model-independent, fast tuning approach that enhances Stable Diffusion. Compared to other more traditional methods, such as Low Rank adaptation for the model or fine tuning it, MoFu doesn't modify the weights of the main model at all. MoFu seamlessly integrates with Stable Diffusion's text encoder, enabling rapid style/concept addition without modifying or fine-tuning the encoder's weights.

## 2. Methodology

The methodology of MoFu revolves around a simple yet effective process. To achieve the desired results, we begin by comparing natural prompts given to a set of images. This comparison allows us to extract the essential concepts or styles from the text prompts. These identified concepts are then stored in a mixin, creating a compact representation of the desired style information. The mixin is designed to be compatible with Stable Diffusion's architecture and serves as an additive to the text encoder's output.

By adding the mixin with the text encoder's output (the mixin, or MoFu model, can also be multiplied by a weight, in order to make its effect stronger or weaker), MoFu efficiently injects the extracted concepts into the image generation process. This injection enables Stable Diffusion to generate images with the desired style without altering the underlying weights of the main model. As a result, MoFu provides a powerful and flexible solution for style

transfer or concept addition in Stable Diffusion without the need for extensive model modifications or resource-intensive fine-tuning.

## 3. Results

To evaluate the effectiveness of MoFu, I conducted a series of experiments and compared its performance to LoRA and fine-tuning methods. Our results demonstrate that MoFu achieves comparable performance to LoRAs while requiring significantly less training time, taking only around 10-20 seconds on average, primarily due to being CPU-bound. This is in stark contrast to LoRAs, which typically demand several hours to train. However, I also observed that MoFu falls short of fine-tuning, as the latter can achieve even better precision/quality but at the cost of a much longer training.

## 4. Conclusion

In conclusion, MoFu offers an efficient and model-independent solution for adding new styles or concepts to Stable Diffusion without modifying the main model's weights. It achieves comparable results to LoRA while significantly reducing training time, making it a practical choice for rapid adaptation. Though fine-tuning still outperforms MoFu in quality, the trade-off between speed and accuracy makes MoFu a valuable option for various applications. Future work may focus on optimizing the implementation / quality of MoFu.