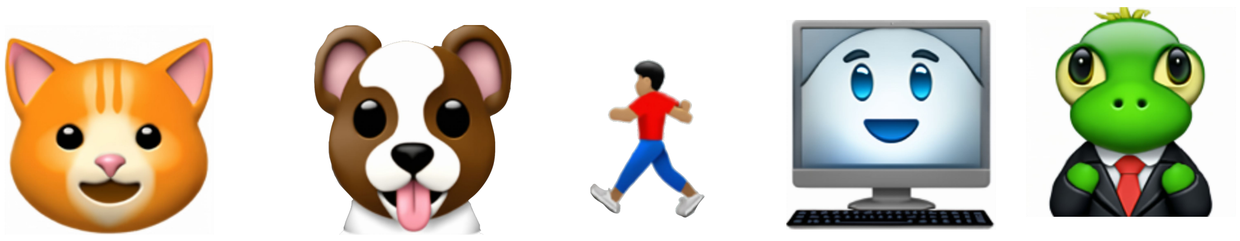
EmojiGEN: Custom Emoji Generation

Elizabeth Eck, Patrick William Hunt II



Our project looks to build existing work of generating emojis using diffusion models and bring this work to open-source. We experimented with various methods, and had a lot of fun in the process

Fine-Tuning

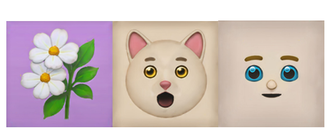
- First, we collected a dataset of emojis in [caption, image] pairs by using and modifying scripts from Evan Zhou
- We then fine-tuned SD-XL, SD-1-5, SD-3, and SD-3.5 models on this dataset using Dreambooth LoRA from the diffusers repository
- One major challenge was formatting data, and which data strategy to use (class vs. instance,) especially with large amounts of data
- We found that SD-3.5 was the best model, with SD-XL second.

Before fine tune (SDXL, SD3 prompted to generate “[X] emoji”)

SD3.5:

Quality, but occasionally confused semantically (see Prompt Augementation!)

SDXL:

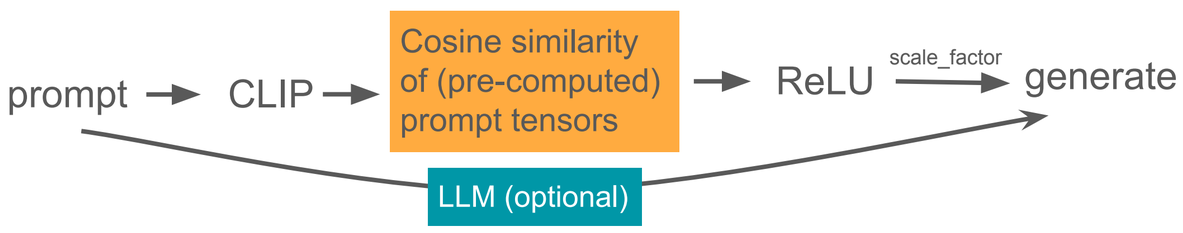


Issues training, very inconsistent but occasionally great results

Even in bad runs, style was learned (SDXL):

Retrieval Augmented Generation

- For most prompts, there will be a similar emoji
- By creating a RAG pipeline, we can efficiently grab the most similar image from our dataset of 1900 emojis
- We pre-compute vector embeddings of all the emoji captions. Then at inference, we can compute the CLIP score of the input prompt (new as of April 2025), fine the most similar, and pass that image as reference



- Unfortunately, we were having a lot of trouble with the IP-Adapter packages and have not yet achieved inference
- However, we want to experiment with the transforming the similarity score to a scale factor.
  - Scale factor is [0,1] with a default of 0.5
  - Want to weight more similar images stronger so ideally transform to a range like [0.3, 0.7], or potentially a non-linear function

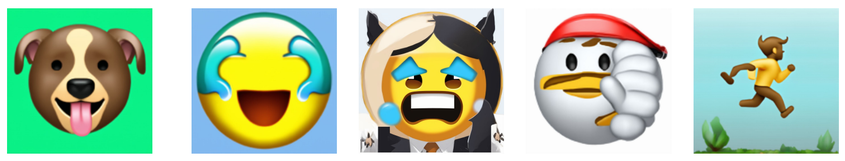
Additional Tools

We created a CLI for this project.

After cloning the repo and minimal setup, you can run “emoji-gen [YOUR_PROMPT]” to generate emojis. We built out the emoji-dev CLI as well, which sets up things like the inference server, fine-tuning, and preparing data.

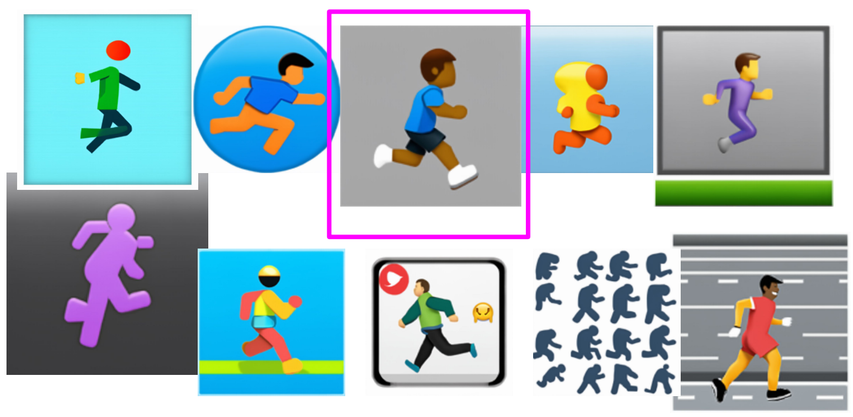
Prompt Augmentation

While our fine-tunes learned the style of emojis, they often produced very sloppy results, as pictured:



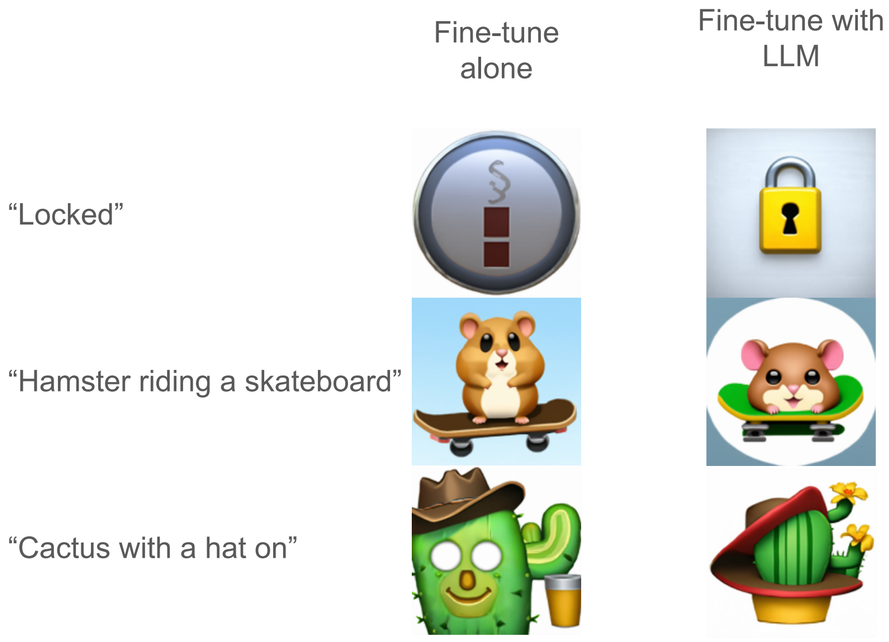
Empirically, we found that more descriptive prompts led to better looking results. We then tried to automate this process. We used a small LLM (Google Flan T5 small) to make the prompt more descriptive, with success.

Here is the prompt “a person running” with 10 different system prompts:



Then, we chose the prompt (too long for this poster) that yielded the best and most consistent results.

We found that this approach could actually have mixed results, but led to generally better emojis. Prompts that were vague got dramatically better, but prompts that were more straightforward seemed to lose style, but remained coherent. In particular, it seemed important to include information about the colors of the emoji, structure of the desired output, and examples of output in the system prompt.



References

[1] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021. 2

[2] Ziqing Li, Yuwei Zhang, Ziyu Zhang, Songyang Liu, Yuning Jiang, Chen Qian, and Chen Change Loy. Ip-adapter: Text-to-image diffusion models with image prompting. *arXiv preprint arXiv:2308.06721*, 2023. 1

[3] Han Liu, Mengdi Zhan, and Xiaoyuan Hu. Emotigan: Emoji art generation using generative adversarial networks. CS229 Final Project, Stanford University, 2017. https://cs229.stanford.edu/proj2017/final-reports/5244346.pdf. 1

[4] David Podell, Zach English, Kunal Lacey, Andreas Blattmann, Theodor Dockhorn, Jan Müller, Joseph Penna, and Robin Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*, 2023. 2

[5] Alec Radford, Jong Wook Kim, Luke Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pam Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*, 2021. 2

[6] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. *arXiv preprint arXiv:2112.10752*, 2022. 2

[7] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. *arXiv preprint arXiv:2208.12242*, 2022. 2

[8] Evan Zhou. Open-genmoji: An open-source emoji generation project. https://github.com/EvanZhouDev/open-genmoji, 2023. Accessed: 2025-05-26. 2