A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

Final project: LoRA Training

Dylan O'Neill

Intro to Stable Diffusion



Stable Diffusion

- Stable Diffusion is a deep learning model originally developed to take text based prompts and turn them into images
- This is done by compiling billions of images and feeding them into a large neural network
- This leads to one model being able to generate a wide variety of images from a vast user input



[https://en.wikipedia.org/wiki/Stable_Diffusion#/media/File:Astronaut_Riding_a_Horse_\(SD3.5\).webp](https://en.wikipedia.org/wiki/Stable_Diffusion#/media/File:Astronaut_Riding_a_Horse_(SD3.5).webp)

Stable Diffusion cont.

- There is a cost to this approach: Lack of specificity and large file sizes
- The base SD 1.5 model is 4.85GB (and that's the pruned model, The original is over 8GB!)
- Even generic cases (see on the right) can lead to less than desirable results from merely using the generic model



https://www.reddit.com/r/creepy/comments/14xid3/some_ai_generated_faces/#lightbox



Stable Diffusion cont. cont.

- This led many people in 2022 to write off AI generated images entirely, deducing that the time needed to get a half-decent image wasn't worth it
- It was simple, using only generalised models, you were never going to get good, specific results without models being larger than the available space on people's hard drives
 - Cloud Computing
- Long term doubts that this technology would have staying power

Introducing: LoRAs





Origin of LoRAs

- LoRAs came from a paper published in March 2022 from a team of researchers at Microsoft with the goal of reducing GPU compute needed to augment existing natural language processors (Using an early version of ChatGPT as their base)
- Microsoft wanted to prevent having to retrain an entire model while only making minor changes and tweaks to suit a specific need
 - Cloud computing costs are high
 - Economic Opportunity Loss can be quite large (Lucrative cloud computing market)
- Since then, LoRAs have become the defacto way of making minor additions on top of existing deep learning models
 - Portability - A LoRA trained on one model can be used on many others
 - File size - Most LoRAs are a fraction of the size of large deep learning models



Summary of my project

- I trained a Stable Diffusion LoRA using Kohya_ss to train my model based on my dog, Juneau
- I wanted to see how effective a LoRA could be trained on a non-human subject, while being able to actually see how well a LoRA could actually replicate a specific , singular being while keeping a limited file size

The process behind
training the model





Dataset

- I collected my data from photographs taken by myself and my family
 - Almost all were taken on phones with a vertical camera position
 - Wanted to keep the file size low ⇒ Use a limited number of photos
- I ended up using a mere 16 curated images for my model out of a possible pool of 104
- Tough to curate an image set for a dog
 - Dogs tend to be bad at posing for pictures...
- How do you choose to capture her?
 - Mix of pictures from when she was a puppy and now?
 - All from one specific time frame?

Kohya_ss: How pictures turn to Stable Diffusion LoRAs

- Kohya_ss is a tool to take images and captioned prompts and use them as inputs to help train a LoRA based on a specific Stable Diffusion model
 - In my case, I took the base SD 1.5 as the base model
 - I then took my curated photos and made short prompts to help associate keywords with the photo

(16 photos with captions)
(Real Photo)



juneau, sitting in car, looking at viewer, mouth open, tongue out, ears facing forward, sitting next to car door, sitting next to car window, car window rolled down, car in background, trees in background, foot paw facing up, juneau partially out of frame

Kohya Colabs

- Google Colab is a free cloud service that allows the sharing of Jupyter notebook files that can be run remotely on google's servers
 - Provide Nvidia GPUs for limited amounts of time for free users
- Such a [notebook](#) for kohya_ss exists, and that's what I used to train my LoRA
- Saves all data to your google drive ⇒ Easy to move across multiple PCs

Kohya Colabs

Accessible Google Colab notebooks for Stable Diffusion Lora training, based on the work of [kohya-ss](#) and [Linaqruf](#).

If you need support I now have a public [Discord server](#)

You can support me on [Buy me a coffee](#)

	🇺🇸 English	🇪🇸 Spanish
📁 Dataset Maker	Open in Colab	Abrir en Colab
★ Lora Trainer	Open in Colab	Abrir en Colab
⚡ XL Lora Trainer	Open in Colab	-
Lora making Guide	Click Here	-
Stable Diffusion guide	Click Here	Click Aquí



What the code looks like

- (Switch to my laptop and show the workspace)

What are the actual results?

Preliminary

- Before I trained my LoRA, I got some initial data of trying to recreate my dog using only text-based prompts using SD 1.5



What are the actual results?

Version 1

- Using the same prompt as the previous image, this is the result I got when I used the same prompt with my own LoRA enabled with a weighting of 1



- Small details make this much more similar to my dog than the preliminary image
- Small problem, she does have two heads...

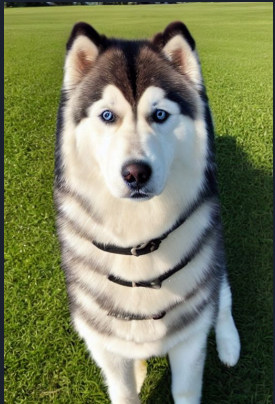
(Real photo)



What are the actual results?

Version 1 Pros and Cons

- Pros:
 - Took only 10 minutes to train on Kohya Colabs
 - Stuck to my 16 picture limit
 - Had only an 18mb file size for a noticeable improvement over standard SD 1.5
- Cons:
 - Had trouble with body shape
 - Eyes are blurry
 - Had that soft, airbrushed look

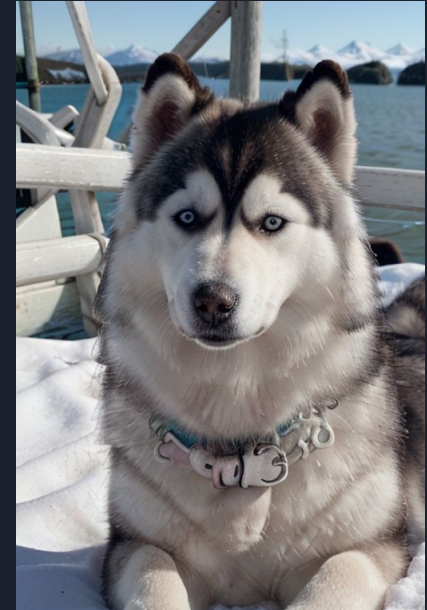


What are the actual results?

Version 1 Addendum

- I also ran version 1 on a model specifically trained on photographs (analogMadness v7)
- Here are the results - Better in some ways, especially in the more complex backgrounds

(Real photo)





What are the actual results?

Version 2 Improvements/Changes

- I changed how I trained Version 2 from Version 1
 - I trained with images rescaled to 768x768 versus 512x512
 - I used AdamW optimiser rather than AdamW 8bit
 - I ran for 10 epochs with 14 repeats each versus 6 epochs with 10 repeats each
 - A batch size of 2 was used for Version 2 rather than a batch size of 3
- The dataset remained the same 16 images with slightly tweaked captions (added more about background detail to help keep the subject in focus)
- This meant that version 2 ended up taking around 50 minutes to train rather than the 10 minutes of version 1
 - However, Version 2 still keeps the small 18MB file size

What are the actual results?

Version 2 w/ AnalogMadness v7



(Real Photo)



What are the actual results?
Version 2 extra





Conclusion

- Main takeaways
 - LoRAs can be an effective tool for making specific beings in Stable Diffusion, and it isn't necessarily any less effective with an animal rather than a human subject
 - Kohya Colab is a really great way to train LoRAs without having a beefy Nvidia GPU of your own
- Things I'd change for next time
 - Sacrifice a bit of file size to add more images and captions
 - Use a more modern stable diffusion version (Either SD 2.1, or Stable Diffusion XL for larger photos and resolution support)
 - I'd train using a more photorealistic model rather than a base, generic model (Thinking how I used Analog Madness v7 as a base my LoRA ran on top of, but didn't think to train with it)
 - Try to crop photos down for a more square resolution (Generally, square resolution photos tended to look better than the photos with a more phone-like aspect ratio)