Jeffrey Hu
Deep Learning Final Project
Individual Final Report

1. Introduction. An overview of the project and an outline of the shared work.

The project was about generating dog pictures through a variety of methods. We wanted to try something that was discussed in class, but we had no experience tackling, in order to present our procedure towards a challenge. We shared the proposal and thinking work for the project, and discussed GAN, sharing literature and examples for how we could do the GAN. In the proposal, I remember suggesting the possible loss functions that we could use between the generative and discriminative portions.

2. Description of your individual work. Provide some background information on the development of the algorithm and include necessary equations and figures.

When we had finished discussing the model, we felt that a single model may not be enough to compare with the . After seeing how UNet2D could be used as a diffusion model, I had also remembered how someone I knew used UNet to create bounding boxes on very low resolution images to great success, so I thought that it would be a good fit for our project, as the GAN model we had produced took a decent amount of time to train even with low resolution, so we needed something that could handle low resolutions well. Other than that, I was the one who coded (read “borrowed most of the code to implement”) the diffusion model from huggingface, and implemented the sample code for the diffusion model pipeline/model output.

3. Describe the portion of the work that you did on the project in detail. It can be figures, codes, explanation, pre-processing, training, etc.

Throughout the project, I mainly worked on the diffusion model. I researched specifically what it was, as even though I had a superficial idea when midjourney and all the other diffusion art platforms were being popularized in the media during 2021-2022 covid, I realized I didn’t know how the diffusion model actually worked in the context of all the other neural network models that I have been studying so far. The main three takeaways I understood during my research about UNet2D is that many of the models that I’ve studied so far overlap with the diffusion model. The architecture is essentially encoder decoder with convolutional layers for image segmentation, along with skip connections between the contracting and expanding path, and spatial attention blocks. So I added the research on diffusion models to the final paper.

4. Results. Describe the results of your experiments, using figures and tables wherever possible. Include all results (including all figures and tables) in the main body of the report, not in appendices. Provide an explanation of each figure and table that you include. Your discussions in this section will be the most important part of the report.

The figures I used were about either architecture, output of my model, or the cifar10 dataset diffusion model introduced by DDPM. The introduction picture was just to display the diffusion process in the paper, while the architecture helped explain my general model and the function/ability of the UNet2D, as well as help discuss its origin from image segmentation on pixelated biomedical images.

I discuss my findings of my output in the main report, but I'll add them here again.

## UNet2D Diffuser

The loss function used for the model was mean squared error. MSE specifically targets better reconstruction at the pixel level for noise prediction problems, penalizing between the predicted and actual noise values that was added in the forward process of the noise scheduler, and also helps reassure the stability of the model.

Here are some example images from multiple epochs of the model.

Fig 17-1/2, Epoch 0 and 1 of the Diffusion Evaluation

From Epoch 0 and 1, the original noisy and abstract image with which the model must derive a dog picture from is seen. From Epoch 1, a few minor changes can be noticed. Some colors, like in image 1, 7, and 9 (left to right, top to bottom), are darkened, and the quality of the images are sharper than the original templates.

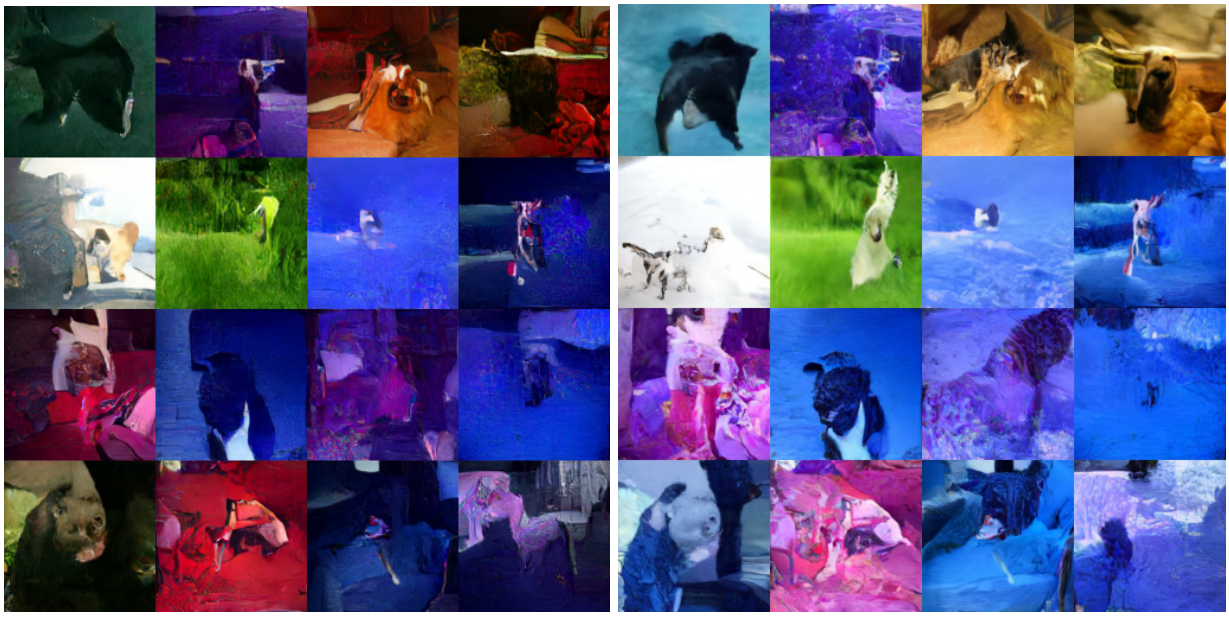

Fig 17-3/4, Epoch 20 and 40 of the Diffusion Evaluation

There is a big change between the first few epochs, and the latter epochs in 20 and 40. Between the first and twentieth epoch, the image begins to focus on the center of the image, as shapes and patterns can be made out. In the 20th epoch, image 10 begins to take the shape of a dog, while image 13 captures the pattern of a dog's nose and tongue. In the 40th epoch, these images are sharpened further, and although the patterns of the dogs are still difficult to see in most images, progress is still noticeable.

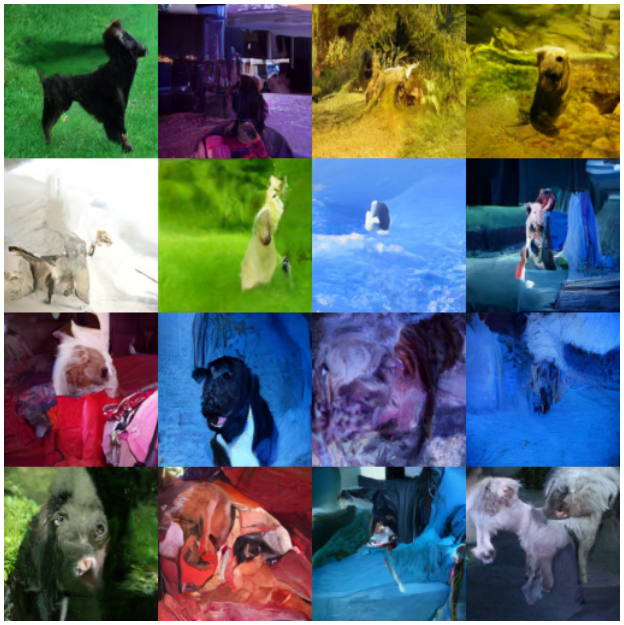

Fig 17-5, Epoch 50

Finally, in the 50th epoch, more colors and backgrounds are sharpened. The first image has changed drastically, as the patterns of legs can now be made out from the original images, behind a green background as opposed to the bluish background at the start. Likewise, the center of each image tends to focus on a black, beige, or white subject, which are more common dog fur colors than the original colors in the image, tinted by the background.

5. Summary and conclusions. Summarize the results you obtained, explain what you have learned, and suggest improvements that could be made in the future.
Future improvements could be made with preprocessing in the GAN model, as we went a little overboard after seeing the dataloader with all the possible preprocessing elements. It's possible that the presentation's streamlit example image of the dog was color blurred because our GAN model mistakenly applied a preprocessing step that color shifted the image before applying the GAN. Also, we should've started from noise to generate our image. Our division of labor could've been better, but it was difficult for each of us to work on the same things especially after we planned to make two models, given our group size. I did also misquote the DDPM paper from memory at the presentation, it's from 2020, not 2006. The first model came out in 2015, so that didn't make any sense. Still, our results showed that even with limited resolution/gpu processing power, GANs and Diffusion models have a

6. Calculate the percentage of the code that you found or copied from the internet. For example, if you used 50 lines of code from the internet and then you modified 10 of lines and added another 15 lines of your own code, the percentage will be 50−10 50+15 ×100.
90% found and copied from HF.

7. References
Code references here (and in python)
https://huggingface.co/docs/diffusers/tutorials/basic_training
https://huggingface.co/docs/diffusers/using-diffusers/conditional_image_generation
https://huggingface.co/docs/diffusers/using-diffusers/img2img
Research references here
Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In Proceedings of the 32nd International Conference on Machine Learning (ICML) (Vol. 37, pp. 2256–2265). PMLR. Retrieved from https://proceedings.mlr.press/v37/sohl-dickstein15.pdf

Stanford Dogs dataset for Fine-Grained Visual Categorization. (n.d.). Vision.stanford.edu. http://vision.stanford.edu/aditya86/ImageNetDogs/

Zhang, J. (2019, October 18). UNet line by line explanation. Medium. https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5