ART AND MACHINE LEARNING
CMU 2022 SPRING
PROJECT 2

Head in the Clouds

Team 3
Kazi Jawad - SCS Stat ML
Richard Zhou - CFA Design
Michael Kim - CFA Design

DESCRIPTION

For our project, we were curious about the application of using machine learning methods to create unique sets of images from quick drawings, specifically portraits of people. To further this concept, we tried to stylize each image by representing them as clouds, hence the name "Head in the Clouds". We explored a wide variety of GANs including Pix2pix, CycleGAN, and style transfer to create the final outputs as well as using Houdini, Cinema4D, and Octane Renderer to generate custom rendered cloud images to feed in as a training dataset.

CONCEPT

For our "Head in the Clouds" concept, we were inspired by the child-like pastime of finding images of faces in objects like food, cars, and buildings. The thought of generating unique images that could be represented as clouds was a concept we all agreed to when thought of, as searching for objects in clouds felt like something we all did growing up. As a final result, we wanted to take a sketch, generate it into a face, and turn the generated face into something that resembled a cloud.

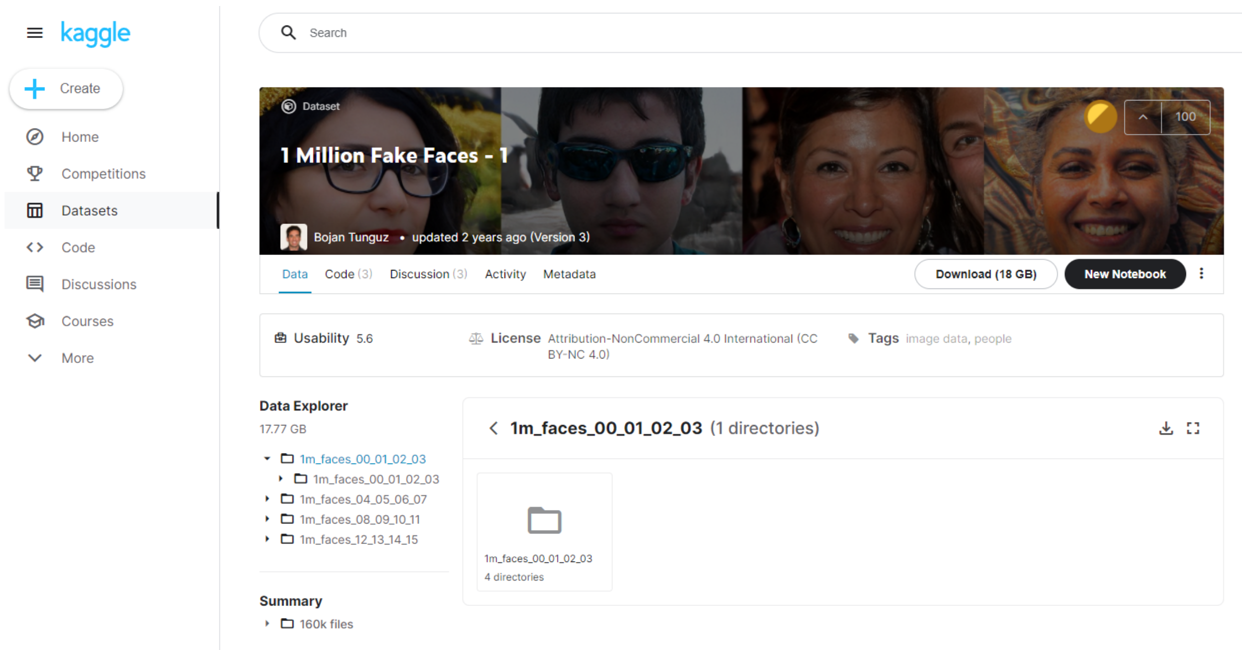
TECHNIQUE

Our initial idea was to generate clouds from quick sketches using Pix2pix and CycleGAN. A Pix2pix model would be trained to generate images of faces from sketches, and then that output would be inputted to a CycleGAN model trained to create unique clouds from faces. Through training a CycleGAN model, we recognized some issues and ended up opting for a style transfer model to replace the CycleGAN part in the end. After comparing and looking for different ways to tweak our system, we ended up returning back to CycleGAN with a modified cloud training dataset.

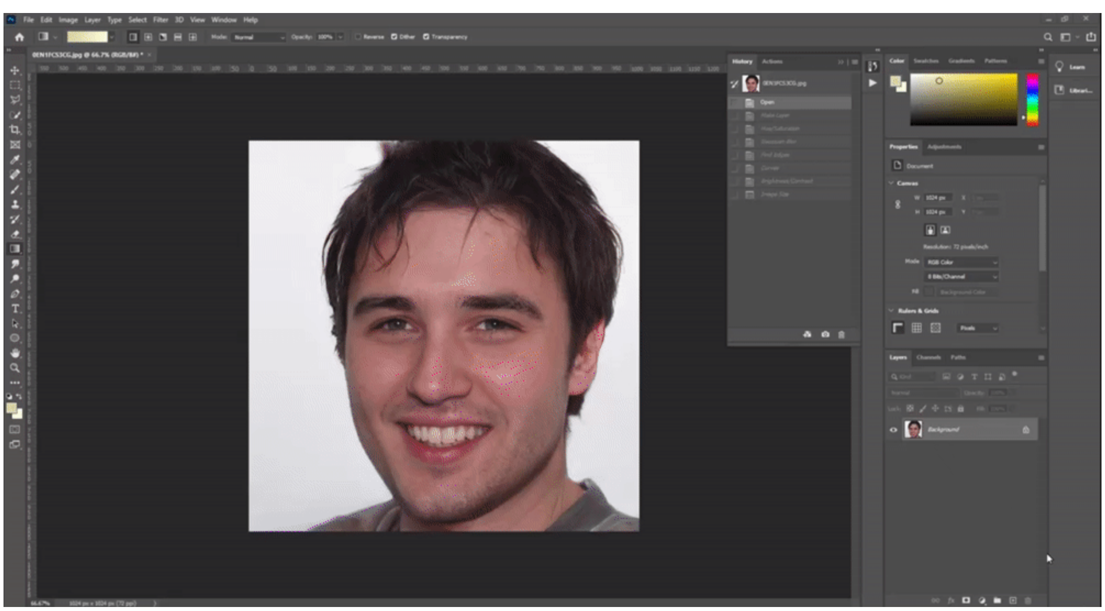
PROCESS

Faces and Drawings Dataset

For the faces dataset, we used a free dataset found on Kaggle that contained numerous fake faces: https://www.kaggle.com/tunguz/1-million-fake-faces



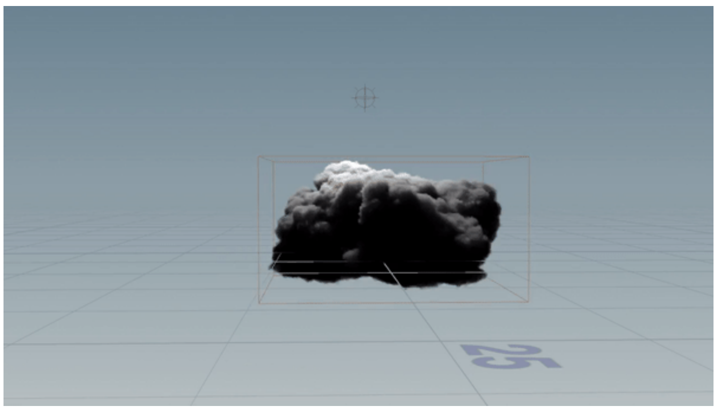
Using Photoshop, we were then able to manipulate these images to create a "drawn" version, and batch exported these as a separate folder. This set ended up including 320 images for faces from the Kaggle dataset, 320 corresponding sketches of faces created by us, and 250 additional test images from Kaggle.

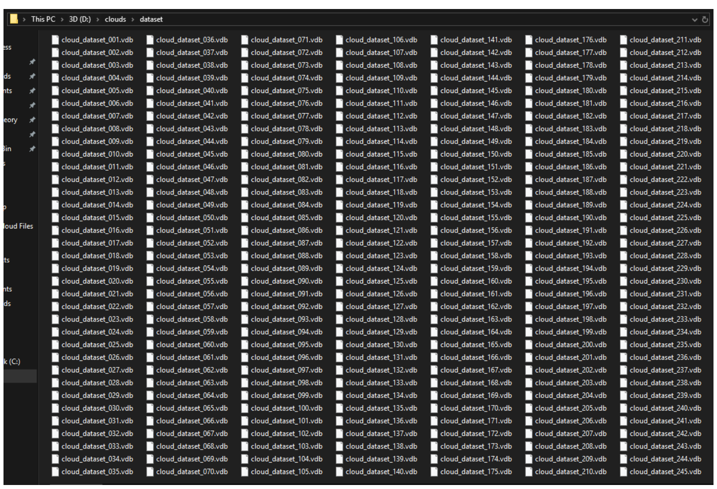


Generating Clouds Dataset

Because we quickly realized that finding hundreds of cloud images that looked similar would be a bit of a problem, we opted on creating a custom dataset workflow by generating the clouds ourselves. So using Houdini, we created a node based system to generate different clouds using a unique "seed" that would influence parameters including:

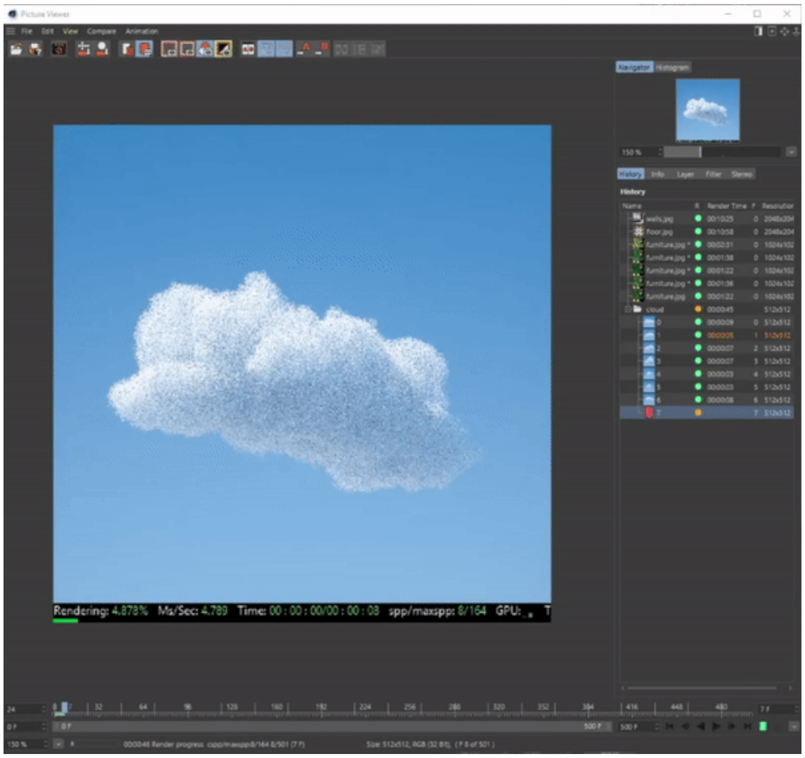
- Cloud Width/Length/Height,
- X/Z Rotation
- Noise Amplitude/Offset
- Random Point Generation
- Amount of Secondary Noise

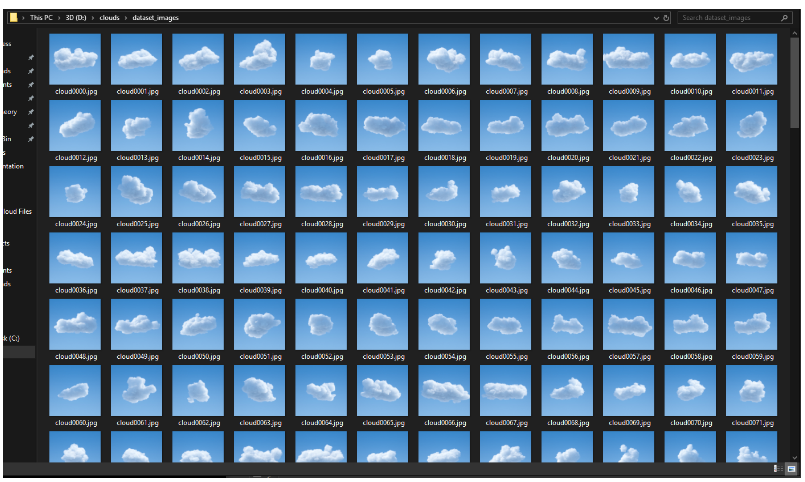




The unique seed was then attached to the timeline so that each frame would be sequenced as a different cloud, and exported as a light VDB volume file suited to be rendered at 512x512px resolution.

We loaded the VDB sequence into Cinema 4D using Octane Renderer's VDB plugin which allowed us to playback each file as part of an animation. This meant that each frame could be treated the same way as Houdini, with an independent frame as a unique cloud. From here, we set up a lighting rig that would light from random angles and rendered out a 400 frame jpg sequence.

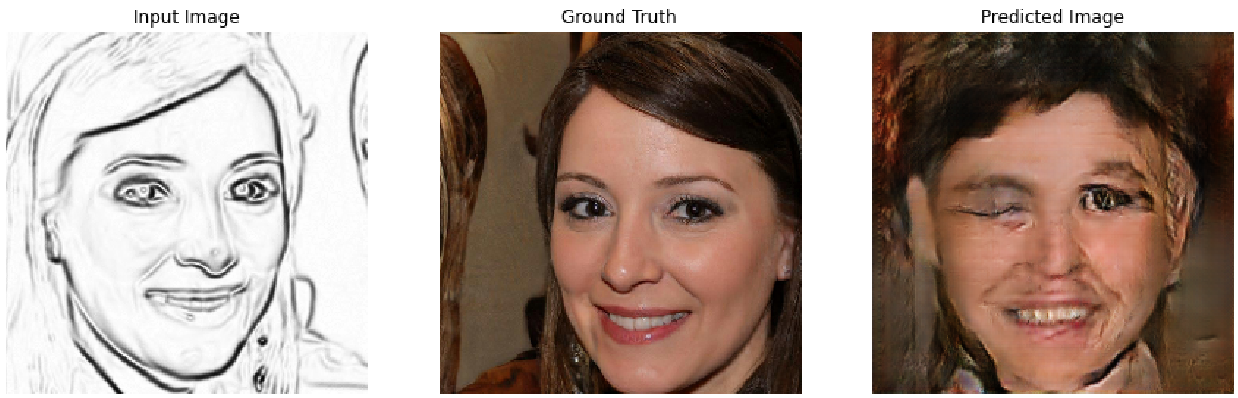

Picture Viewer
File Edit View Compare Animation
Navigator Histogram
History Info Layer Filter Stereo
History
Rendering: 4.878% Ms/Sec: 4.789 Time: 00 : 00 : 00/00 : 00 : 08 spp/maxspp: 8/164 GPU:


This PC › 3D (D:) › clouds › dataset_images
cloud0000.jpg
cloud0001.jpg
cloud0002.jpg
cloud0003.jpg
cloud0004.jpg
cloud0005.jpg
cloud0006.jpg
cloud0007.jpg
cloud0008.jpg
cloud0009.jpg
cloud0010.jpg
cloud0011.jpg

pix2pix

In order to generate faces from sketches, we created a pix2pix model that would take inputs of faces with corresponding drawings to predict the generation of sketches that we could input.

Final Result (40 Epochs):



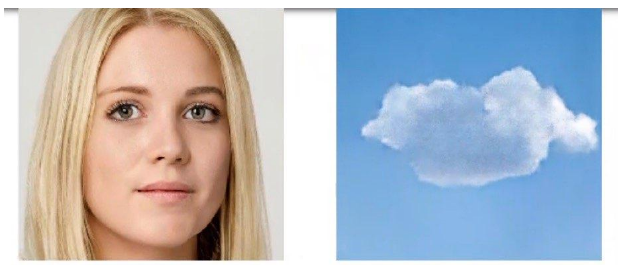
The architecture of the generator followed a modified U-net, while the discriminator used a PatchGAN classifier, as defined in the original pix2pix research paper. This was trained for 40 epochs across 500 sets of sketches and faces for a total of 1000 images per epoch. The training was set up using Anaconda and TensorFlow on a Jupyter Notebook and was trained on one of our personal computers paired with a Geforce RTX 2080 Ti graphics card overnight. With 40 epochs, we were starting to get results that resembled human faces, while not sacrificing too much processing time on a machine.

CycleGAN

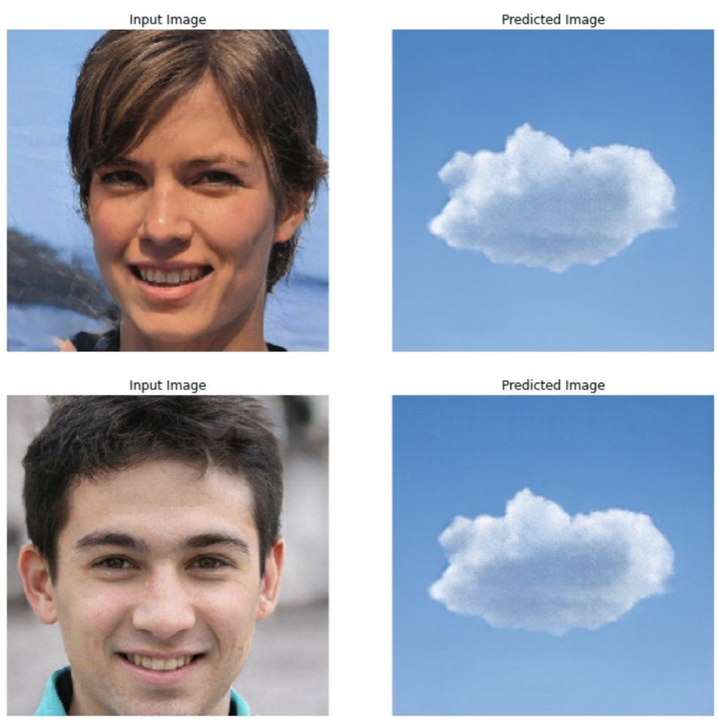
For our CycleGAN training, we input the 500 unique faces images and 400 unique cloud rendered images as the training set over 40 epochs. Additionally, we attempted a longer training period with 60 epochs, but ended up getting very similar results to our first trial.

The architecture of this model's generator and discriminator follows our pix2pix setup to remain consistent. The training was set up using Anaconda and TensorFlow on a Jupyter Notebook and each trial was trained on one of our personal computers paired with a Geforce RTX 2080 Ti graphics card overnight.

Initial Input:



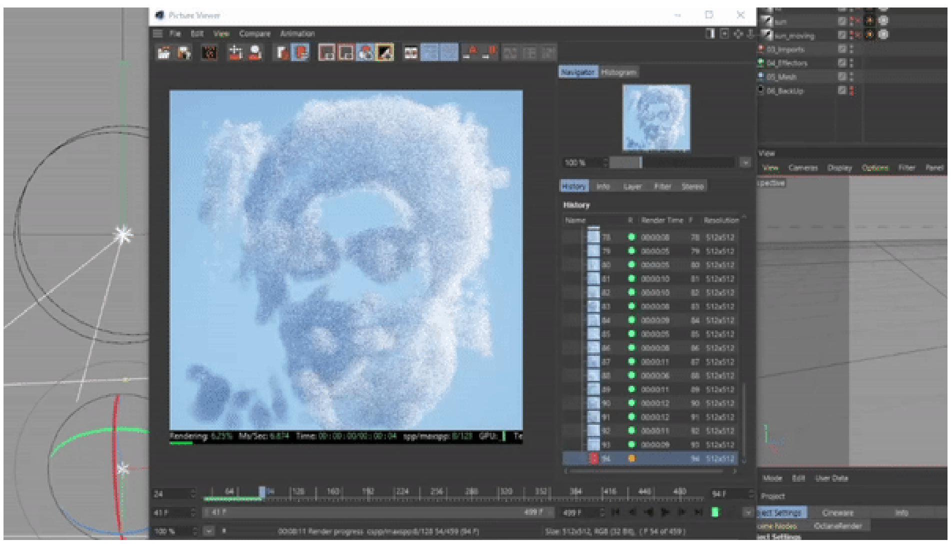
Final Result (40 Epochs):

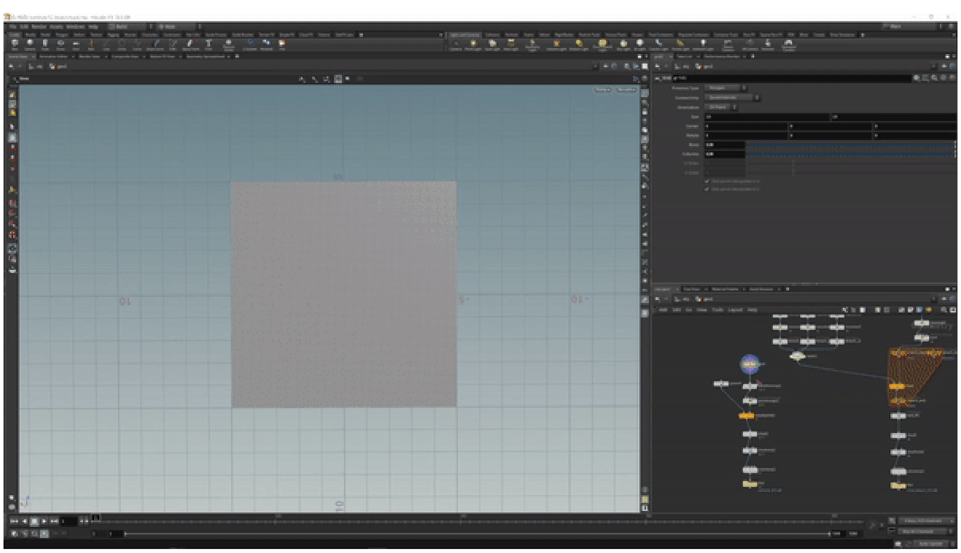


Through training a CycleGAN model, we recognized that there was a low amount of contrast in the cloud images and did not allow for 1-1 input for each face to each cloud input. This ended up creating a set of similarly looking images with new test inputs. Realizing that this was making little progress, we shifted our thinking to using Style Transfer instead.

CycleGAN v3

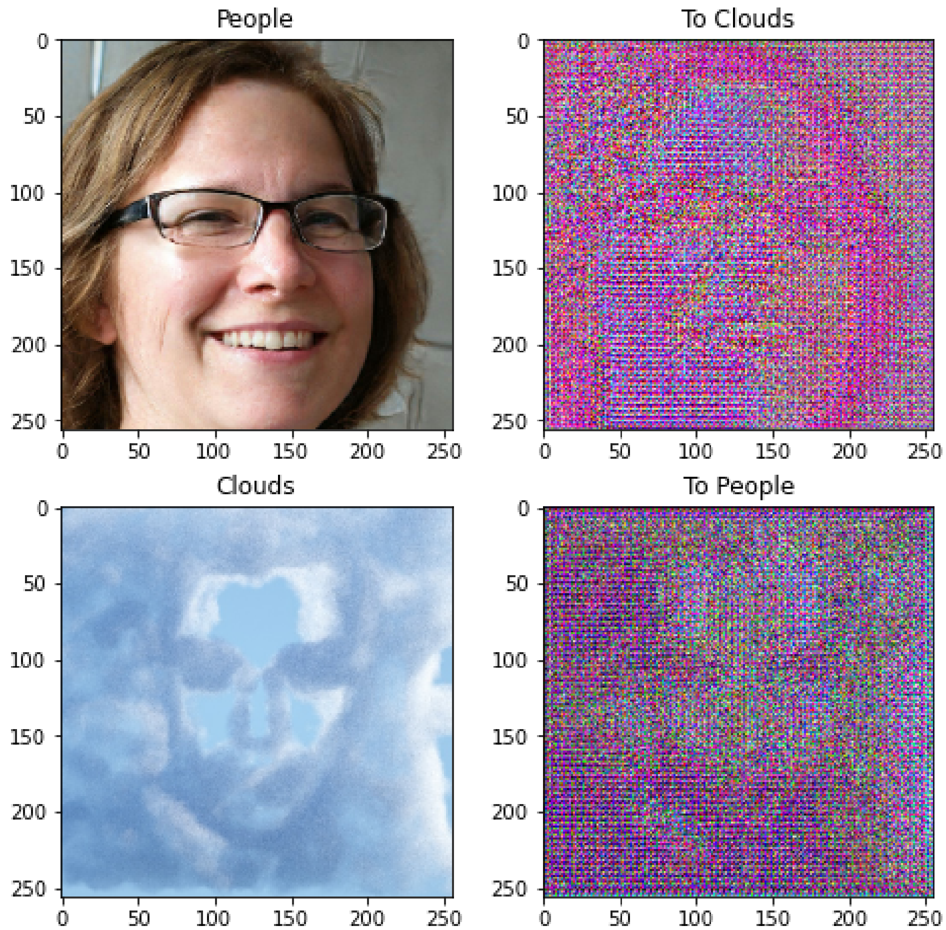
After 2 unsuccessful trials and attempting Style Transfer, we came back to CycleGAN for a last trial. This trial would attempt to use clouds formatted more similarly to the portraits setup, and even use image information from the human dataset to create the cloud training set. Each cloud was custom generated according to the portrait dataset used previously, except turned into a 3D volume geometry that would be rendered in Cinema 4D using Octane Renderer.



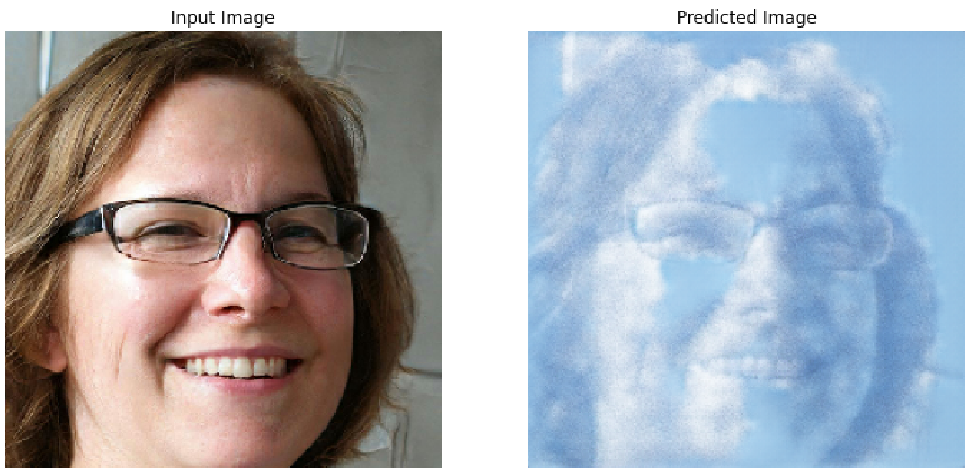


A total set of 500 cloud jpeg images along with the previous 500 faces images were trained over 40 epochs using a RTX 3090 and 2080Ti machine in about 5 hours.

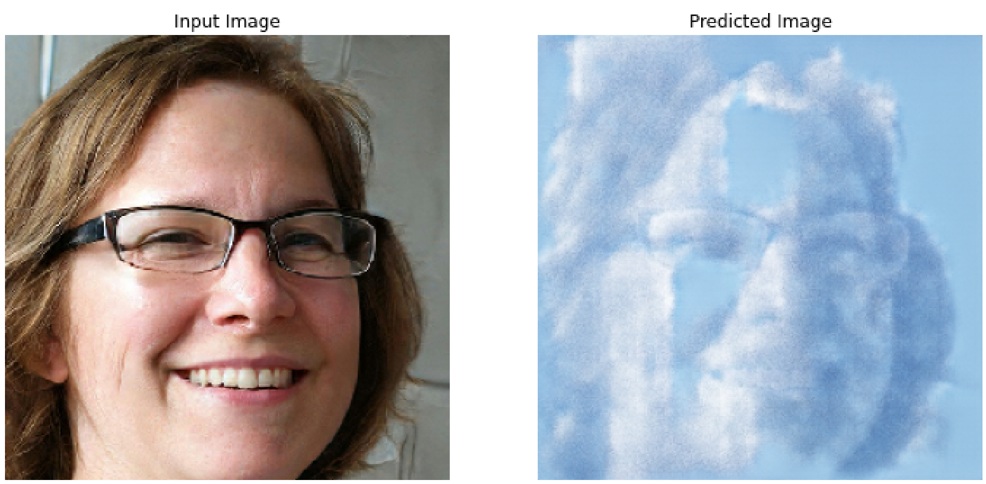
Initial Input:



Final Result (Epoch 30 + 40):



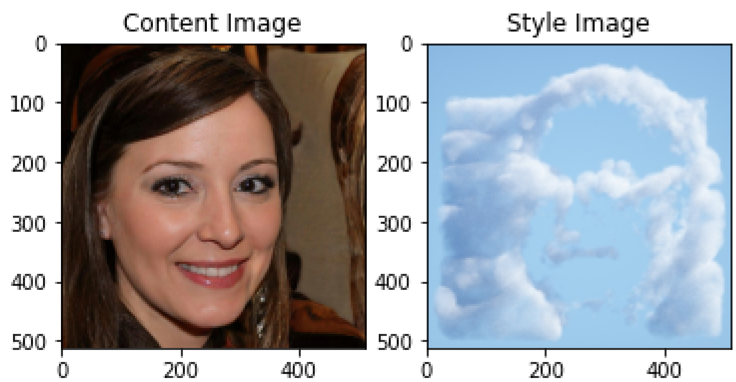

```
Saving checkpoint for epoch 30 at ./checkpoints/cycle_gan\ckpt-6
Time taken for epoch 30 is 373.71489930152893 sec
```




```
Saving checkpoint for epoch 40 at ./checkpoints/cycle_gan\ckpt-8
Time taken for epoch 40 is 374.5382812023163 sec
```

Style Transfer

As a second attempt to create the cloud images from the generated faces, we attempted a neural style transfer. This was trained for 20 epochs across 100 steps per epoch on a pair of face and cloud images. Given the small number of epochs and the lack of color variation in the style image, the result ended up creating incomplete results.



Final Result:

The architecture of the model followed a standard VGG16 with an Adam optimizer configuration, as well as factoring the image variation into the loss calculation. The training was set up using Anaconda and TensorFlow on a Jupyter Notebook and each trial was trained on one of our personal computers.

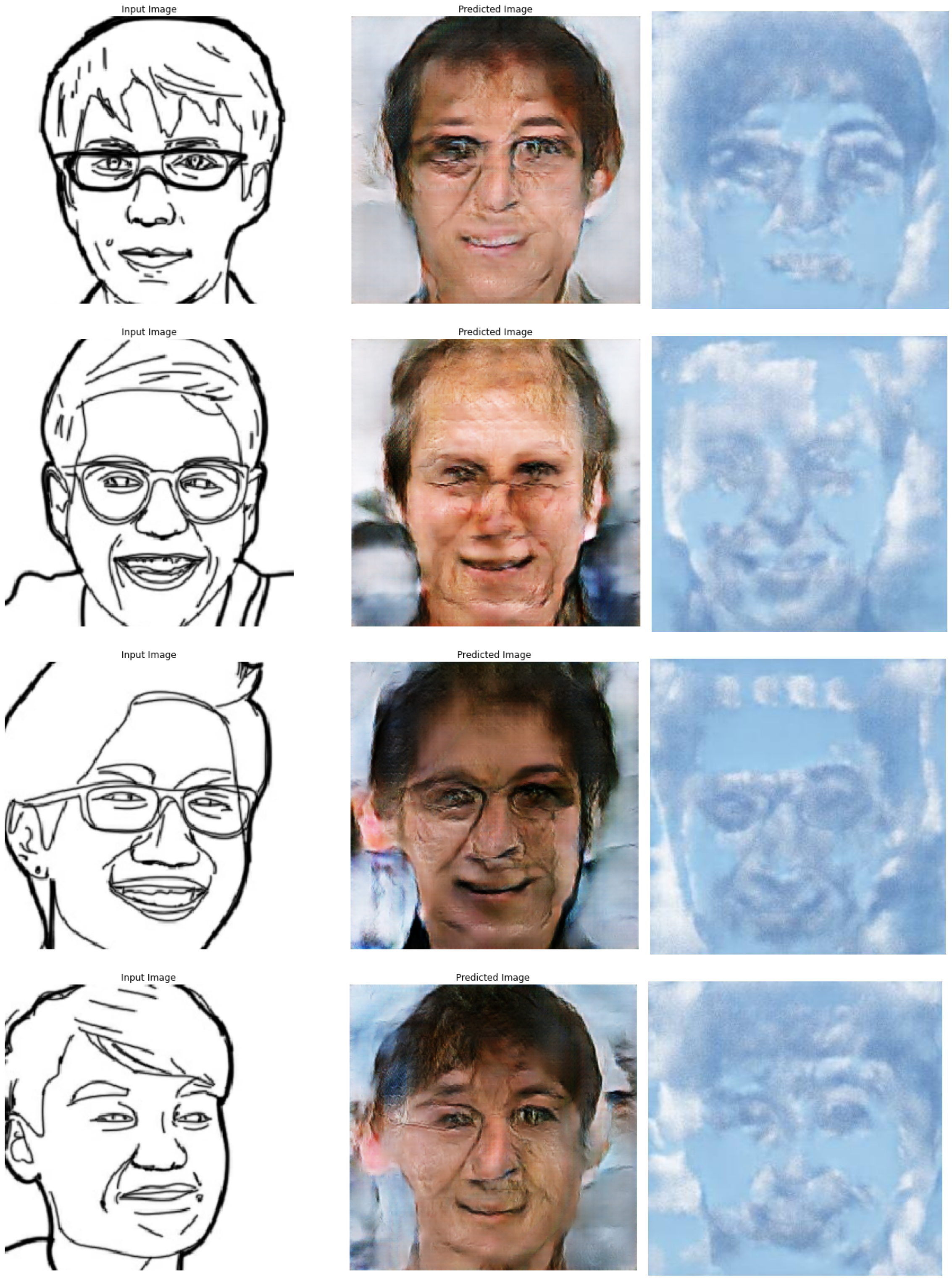
REFLECTION

Through creating this project, we can all agree that the greatest takeaway from it was the fact that machine learning is unpredictable. We spent days creating the concept for the project, but quickly were humbled by the fact that the process would not be as predictable as expected. The clouds generating poorly from CycleGAN was a surprise to all of us, and we had to be flexible with our thinking to try a new GAN instead of continuing to use time with CycleGAN. Expectations were not reality, and we definitely needed more thought in the beginning to plan for this sort of troubleshooting time. Ironically, things worked out in the end coming back to CycleGAN, but it required much more thinking than we had previously had planned for scoping the idea for the project. Style Transfer worked out being ok, but coming back to CycleGAN was also an unexpected route that we came to at the very end of the project preparing for the presentation.

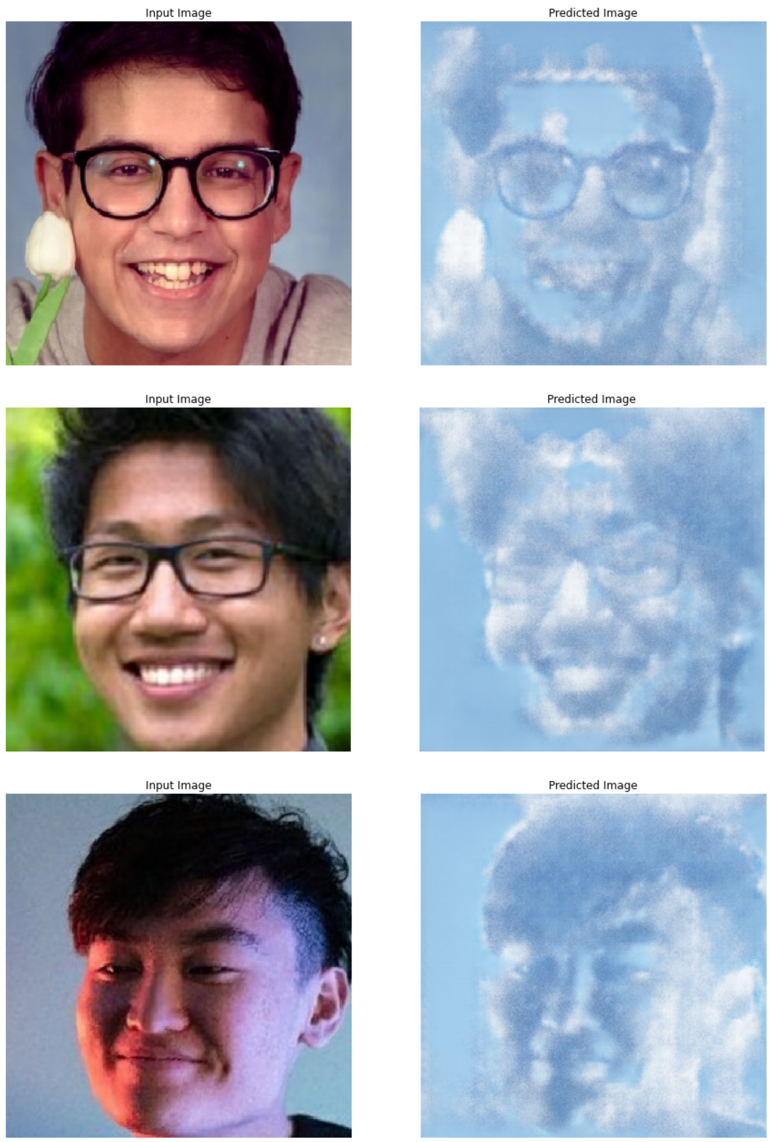
If we had more time, we would have spent it cleaning the training data and possibly searching for more drawings that could have been used as the "thumbnail" or "poster image" for the project. We primarily utilized TensorFlow's eager execution, but with more time we could have refactored it to primarily use TensorFlow's graph pipeline with dataset caching to improve the training time. We realized a lot of AI prototyping involves inferencing and iterations, and we ended up pushing this type of work until the last few days before the presentation.

RESULT

Here are our final results using pix2pix and CycleGAN. From left to right: we inputted custom sketches, generated the face images from pix2pix, and generated the final image style transferred as a cloud.


Input Image
Predicted Image
Input Image
Predicted Image
Input Image
Predicted Image
Input Image
Predicted Image

We also ran CycleGAN on its own against the raw image inputs of our individual faces:



Based on these outputs, it's clear that the pix2pix model likely needed a larger and more varied training set. The majority of the training data used in our pix2pix model had faces directly at the camera. This led to the image generation to be skewed towards faces facing forward. However, it's also clear our new approach to cloud generation significantly improved the final output and the trained CycleGAN was able to match the input very well.

CODE

https://github.com/kazijawad/head-in-the-clouds

Additionally Used Code/Dataset:

- https://www.tensorflow.org/tutorials/generative/cyclegan
- https://www.tensorflow.org/tutorials/generative/pix2pix
- https://www.tensorflow.org/tutorials/generative/style_transfer
- https://www.kaggle.com/tunguz/1-million-fake-faces

More output files, training data, and training checkpoints:
https://drive.google.com/drive/folders/1s8mQXz0PmKKj81FkPTlnAbBCmkrI5KPI?usp=sharing