Art and Machine Learning
CMU 2019 Spring
Project 2

Memory Palace

Mira Murali
Loki Ravi
Izzy Stephen
Yuhan Xiao

Concept

The very concept of ‘forgetting’ underscores the difference between human memory, which is mutable and unreliable, and computer memory, which by virtue of its architecture is pristine and meticulously organized. In the human brain, memories are degraded over one’s life and literally rewritten, with loss, each time they are recalled. Lived experiences, once detailed and poignant, meld together over time until they become a blur. Strange memorization techniques like the ‘memory palace’ take advantage of the brain’s strange power to associate seemingly unrelated concepts, showing that the architecture of human memory is far from straightforward. But is the frailty of human memory a shortcoming, or an essential part of the human experience? Perhaps the distortion the brain inflicts on our memories gives way to creative output and self-reflection. This project imagines an AI with human memory. The program has been trained to restore heavily distorted images to their original state, as if dredging up the face of a long-forgotten friend or a location visited in childhood. It is presented as pairs of the original image with its ‘reconstructed’ version, as well as a collage envisioning an AI’s warped memory palace.

Technique

The main objective of this project is to teach a generative model to explore identities when given an image of a faceless human being. We want to address the following questions through this project:
Could the model dream up interesting features in a face? What is a familiar face to AI: is it simply the most commonly seen facial features put together? Or does AI learn to have preferences? When it sees an image, whatever it may be, does AI construct a face from it?

Model and Representations
We use our own implementation of the pix2pix model to run our experiments with customizations. The pix2pix model learns a pixel to pixel mapping given representation A to a target representation B. We experimented with a few representations:

Input A: Blurred Image	=> Target B: Original HQ image
Input A: (Blurred Image + B/W Sketch)	=> Target B: Original HQ image
Input A: (Blurred Image + B/W Sketch + Age + Gender)	=> Target B: Original HQ image

Dataset
We use Nvidia’s Flickr-Faces-HQ (FFHQ) dataset for training and validation. The dataset consists of 70,000 1024x1024 PNG images. We resize the images to 512x512 to reduce the computational cost of processing and training on such high quality data.
For testing, we collected 5000 non-face images of different categories - Stanford dogs, Stanford cars, Caltech home objects, and MIT indoor scene recognition.

Preprocessing
The blurred image was created by applying median blur 20 times consecutively to the ‘real’ image and pyramidal mean shift filtering to the final output. We also applied Holistically-Nested Edge Detection (HED) to extract the black and white sketch of the original image. We used an age/gender estimator model trained on the IMDB-WIKI faces dataset to estimate the age and gender from the HQ images. Where applicable, the blurred image and B/W sketch were concatenated into a single 4 channel input to form representation A.

Process

We had three ideas for what we wanted the model to achieve by learning the representations which we outline below as our main experiments:

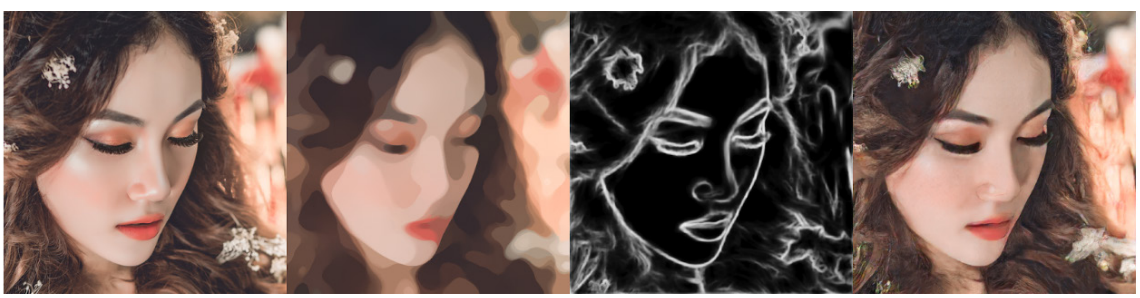
(Blurred Image | B/W Sketch) => HQ Image
We experimented with conditioning the GAN on more than one representation. We could see that the model could effectively learn from more than one representation.

(Blurred Image | B/W Sketch) + Age/Gender => HQ Image
We wanted to control the generation of the image while conditioning on an input representation. By using Age and Gender as additional inputs to the model we hoped to control the facial features generated by the model such as skin texture and hair density. Contemporary experiments have shown that different layers of a well trained generator are specialized in generating different features such as eyebrows and hair color. Since age is a factor that would affect every layer of the generator we decided to feed age and gender as a one-hot vector to every layer of the generator using learned weights to represent its significance to that particular feature. We failed to obtain any results since increasing the number of parameters in every layer of the generator as well as decorrelating the filters with per layer noise significantly increased the training time.

Blurred Image => HQ Image
Our intuition for concatenating the black-and-white sketch of the face as an additional channel was to help the model understand facial features. However, we were curious to find out whether the sketch was actually helping at all. It was possible that the extra information was hindering the model from understanding the connection between the representations. So as an experiment, we removed the additional sketch channel and fed the model only the blurred image as representation A.

Results

Since the model is trained and validated only on faces, an interesting test was to pass in non-face images as representation A and observe what the model generates. Would it dream up a face in an outdoor scene if that was presented as a test image?
In this section, we present the best test results for each of our experiments on the FFHQ dataset as well as the non-face dataset:
(Blurred Image | B/W Sketch) => HQ Image:
We present two of our best results on FFHQ test data when using the the blurred image and the black-and-white sketch of the image. The images presented below are in the following order: (from left to right) Original Image| Blurred Image| B/W Sketch| Generated Image:

(a)


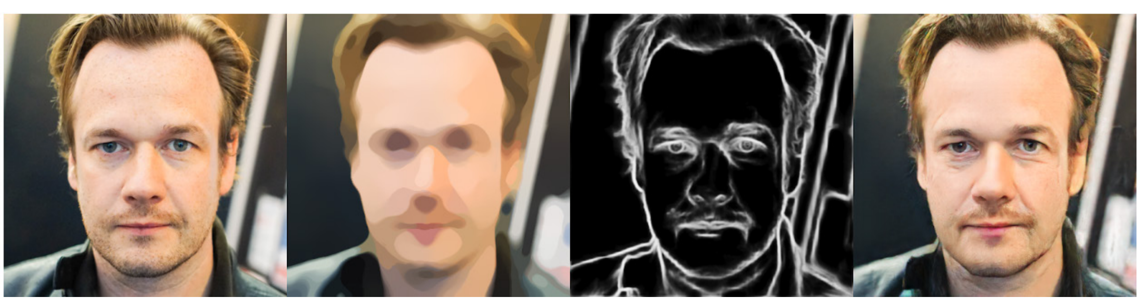
(b)


Fig 1. (a) and 1. (b): Two of our best results on testing the pix2pix model - trained on both blurred images and the edge sketch - on the FFHQ dataset
We observe that the model smoothes out beauty marks and other scars when generating the final image, because the blurred image and the edge sketches together are not descriptive enough to capture small blemishes/marks.

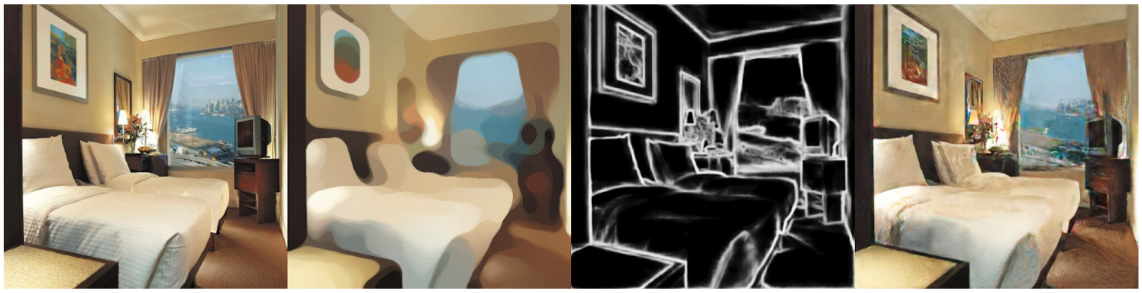


Fig. 2: Our best test result using the trained pix2pix model on the non-face dataset
Blurred Image => HQ Image

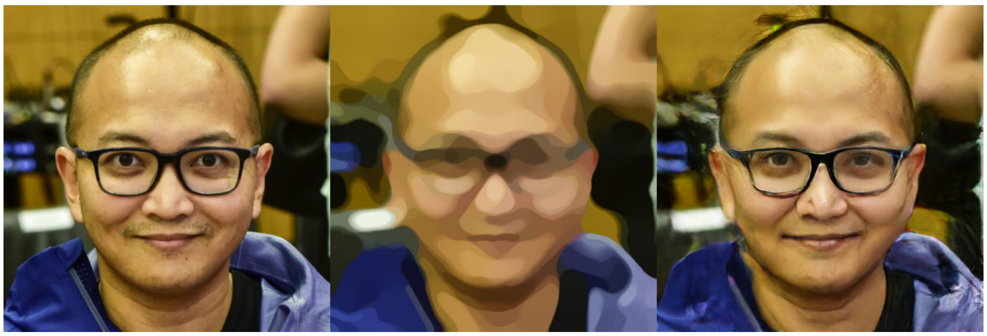


Fig 3: Result of testing the pix2pix model trained using only blurred images on the FFHQ dataset
The figure above is our best testing result on faces for this particular experiment. We see that the model does quite well at generating faces even when it has not been fed the black and white sketch.
However, when testing the model on the non-face dataset, we obtained the following result:

Fig 4: Result of testing the pix2pix model trained using only blurred images on the non-face dataset
The result shows that the model generalizes well to non-face images despite being trained on only face images. There is some level of smoothing in the image. For instance, each type of food item blends together with other similar food items. This is probably because the model was only fed the blurry images. There is also some distortion but we do not see any resemblance to a face in the generated image.
Overall, we conclude that including the edge channel helped the model learn better features and hindered it in

Reflection

We observe that pix2pix is quite good at learning a generalized pixel to pixel mapping between an input representation and an output representation. Stacking multiple input representations allow us to condition on multiple features. This allows for coarse grained control of the generated image. For example, this could potentially allow us to erase scars and other blemishes on the skin. More sophisticated features such as hair color, skin texture, are tightly correlated and require decorrelation of the various layers. This could be achieved by training an additional decorrelation/control model that would allow for input/noise to be inserted into every level of the U-net generator. However, this comes at the cost of exponential increase in training time.

Code

https://github.com/mira-murali/pix2pix-artrepo

Contributions

Artistic meaning, concept, final result compilation: Izzy Stephen
Test dataset collection and curation: Loki Ravi, Yuhan Xiao
Data pre-processing: Mira Murali, Loki Ravi
Training, testing: Mira Murali, Loki Ravi
Presentation and report preparation: All