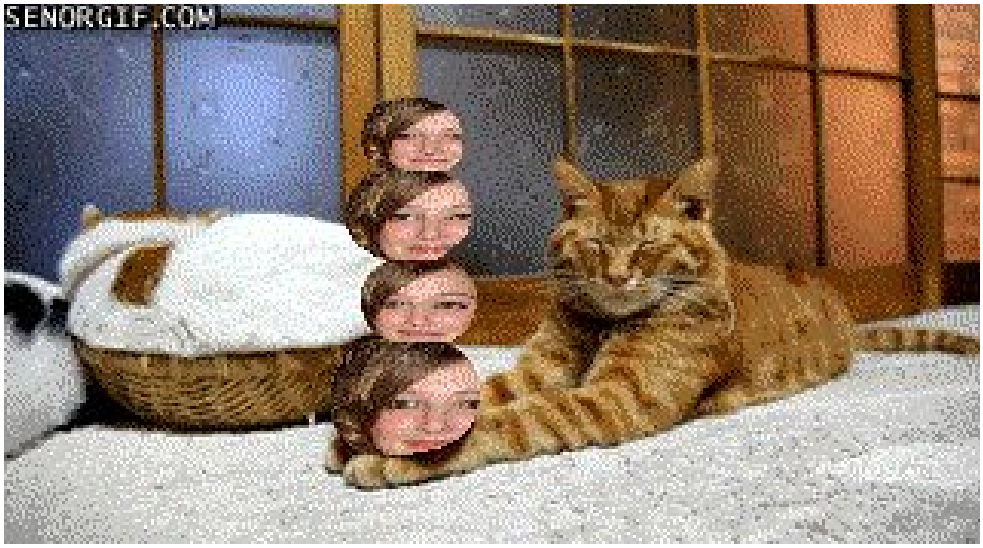
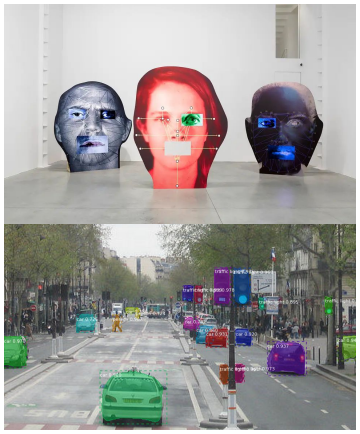
ART AND MACHINE LEARNING
CMU 2019 SPRING
FINAL PROJECT REPORT

EVERYDAY FACES



GROUP 16
Boxiang Lyu / Chuning Yang / Edgar Xi / Tianjun Ma / Zhiyu Bai

CONCEPT



Similar to the concepts behind our work in Project 2, we decide to dig deeper into converting everyday objects to human faces, heavily inspired by the works of Tony Oursler. We want to depict a dystopian world where everyday objects are utilized for surveillance purposes and they arouse a sense of fear and insecurity in every aspect of people's lives. In Project 2, we established a rudimentary pipeline for processing image data. For the final project, we propose many new aspects in which we can improve upon existing methods, which we will discuss in the methods section. Furthermore, we will experiment with new artistic styles for rendering our results, presenting them in a more aesthetically pleasing way.

RESEARCH

As an effort to improve upon existing work, we will utilize some recent innovations in machine learning in order to generate higher quality faces and reduce the amount of human intervention needed. In particular, we will focus on the following models:

1. Pix2pixHD [1]: Introduced T. Wang et al., the model is capable of generating high quality faces, as long as the training images are of high quality. This model will be the most crucial aspect in improving the performance.

Input: Hand-drawn edges					
Output					

2. CelebA-HQ: A by-product of the NVIDIA sponsored paper on Progressive GANs [2], CelebA-HQ used neural networks to transform lower quality CelebA dataset into high dimensional ones, up to 1024x1024.

3. Holistically-Nested Edge Detection [3]: Introduced by S. Xie and Z. Tu, the algorithm uses machine learning as opposed to traditional computer vision algorithms to detect edges in an image. Our original algorithm often produces noisy and blurry edges, and it is our hope that HED can generate refined edges that lead to higher quality faces.

These improved methods, combined with state-of-the-art Mask RCNN structure pretrained on the Microsoft COCO dataset, proved to be extremely powerful, capable of generating aesthetically pleasing works.

IMPROVED METHODS

- Algorithmic Improvements -

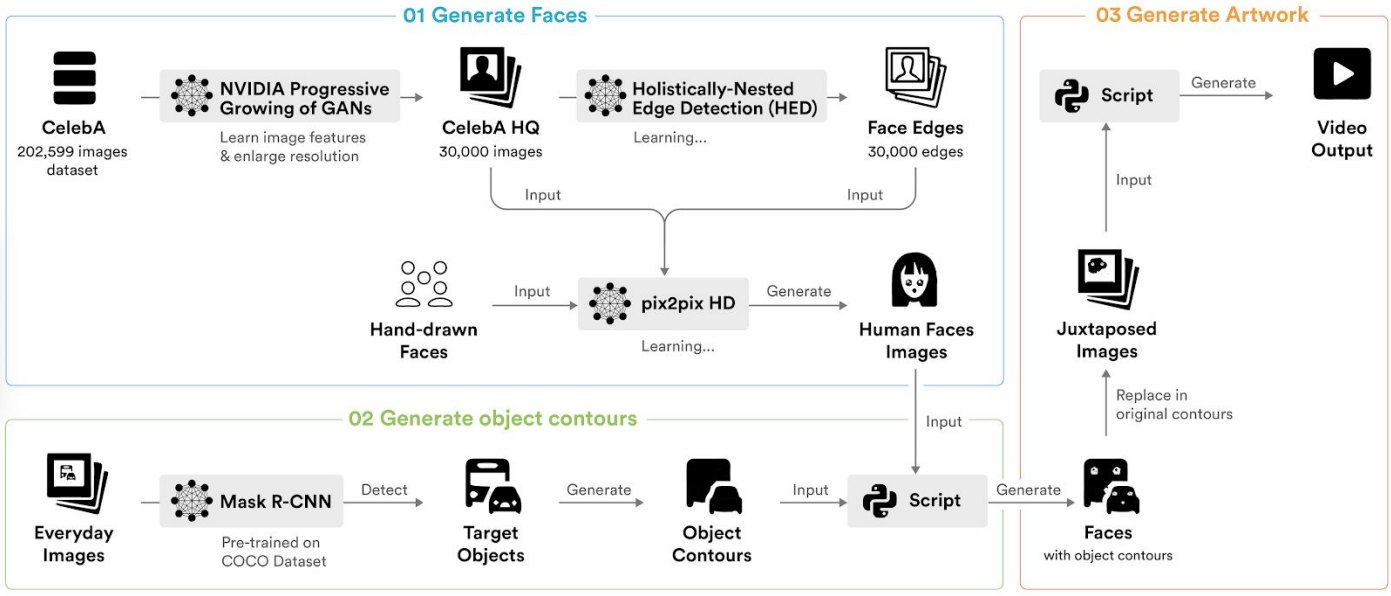
1. Higher Resolution Dataset: Instead of using CelebA, which is often low in resolution (around 200x200), we are going to use CelebA-HQ discussed above.

2. Better Model: We will use pix2pixHD instead of vanilla pix2pix to generate higher quality faces.

3. Better Edge Detection: In the training phase, we originally used a suboptimal algorithm for edge detection (canny). The resulting edges can often look noisy and bring down the quality of the generated faces. We managed to get HED in the original model to work and will use the improved, deep-learning based algorithm to detect edges for the final project.

4. New Procedure for Generating Faces: During our project we discovered that detected edges in the cars often fail to generate meaningful faces and manually drawn ones are used. We propose the following procedure to speed up the process.

PROCESS



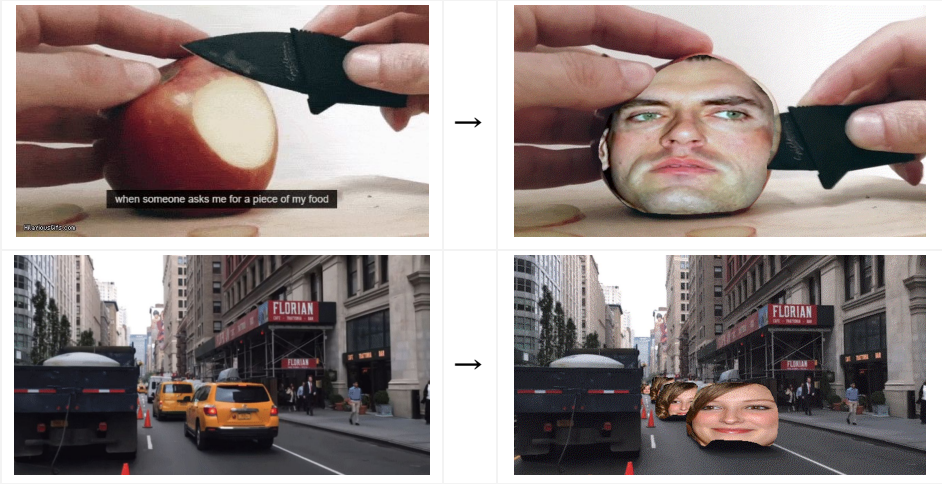
- Ideas for Better Presentation -

1. Incorporation of Short videos: For the final project we will use the proposed data processing

method to generate videos using mask-RCNN. Visual effects will be added to make the result look like old movies, and we will transform the background to black and white.

2. Blurred Edges: In project 2 we discovered that the edges of the generated images may be too sharp, making the edited faces stand out and look unnatural. In the final project we blured the edges to make the image more natural and pleasing.

3. Juxtaposing:



In conclusion, our final project is an extensive overhaul of the concept that we established in project 2: we want to utilize generative models to impose facial components onto everyday objects to create snapshots of a dystopian world. We broke our project down to several components: pix2pix model for human faces, dataset for that model, identification of objects to be replaced by generated faces, and the procedure of that replacement. In our final project, we will improvements upon all four areas by using the HD version of the pix2pix model and dataset, refined edge detection with minimal noise, and python automation for embedding faces in the detected contours rather than drawing them manually. Finally, we will present our result in a more impactful manner by extending the result into audio-visual format. Necessary steps such as edge blurring will also be utilized for the refinement of every frame.

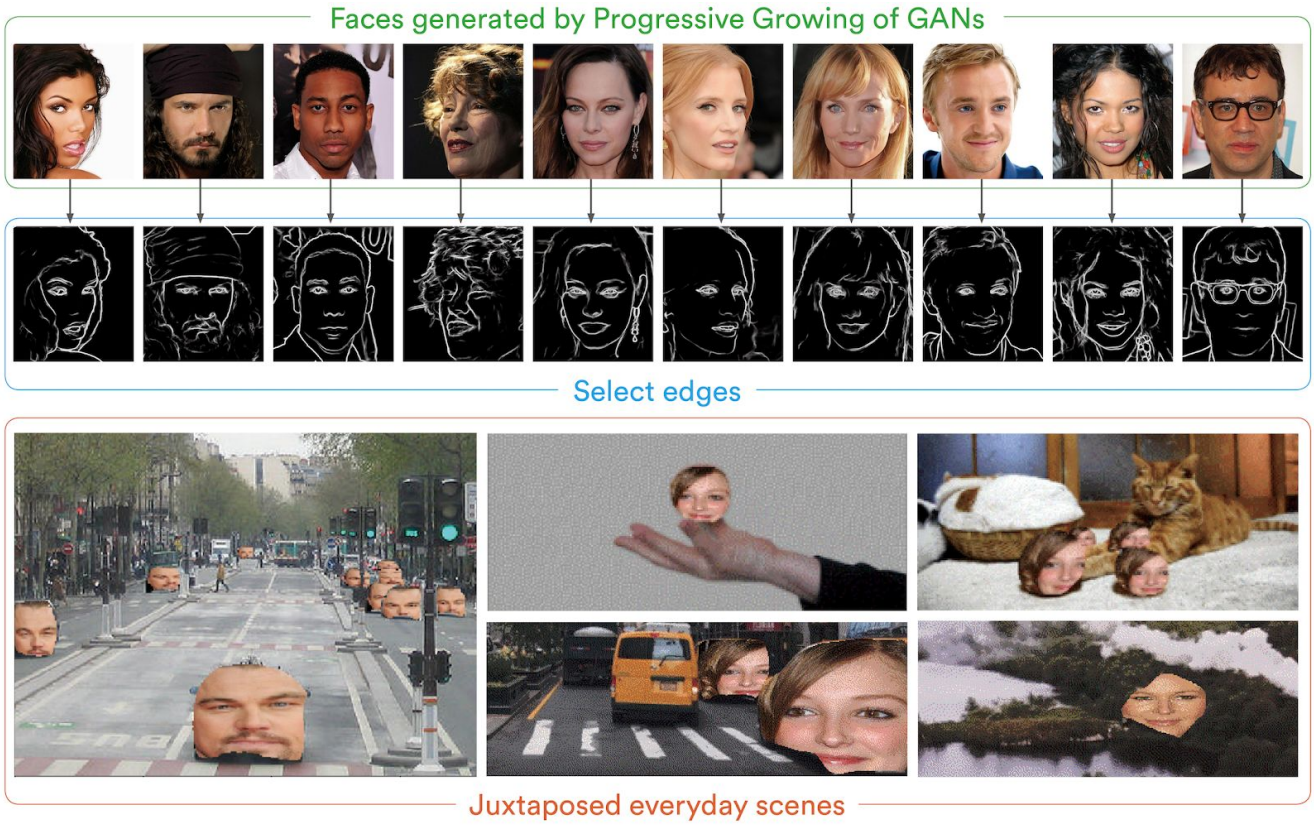
TIMELINE

Mar 26 - Apr 11	Reconstructed CelebA-HQ dataset using the delta values provided by the authors of Progressive Growing of GANs. Used pre-trained HED model to detect the edges from the CelebA-HQ dataset.
Apr 11 - Apr 15	**Train pix2pix HD:** We spent more time training pix2pix because the result is not realistic. Generate faces
Apr 18 - Apr 25	Generate gif

Apr 25 - Apr 30	Generate video

RESULT

We created much more refined images using Progressive Growing of GANS and generated clearer edges using HED. By using higher quality images with Mask-RCNN, we created still images and videos, where the daily objects are replaced with human faces, successfully accomplishing what we set out to do. Our artistic choices also proved to be helpful, making the juxtapositions less jarring and much more pleasant to look at.



Video: https://youtu.be/-c2pjbec-2Y

Images: https://drive.google.com/open?id=1Pr9LWye27JPNckOwy1hYKO5AQCVO_O3T

Code: https://github.com/tianjunm/artml-s19/tree/master/final_project

REFLECTIONS

From an algorithmic perspective, we are impressed with the quality of the edges produced by HED. Originally we thought canny edge detection may be a suitable substitute for HED, but it turns out that HED outperforms canny edge detection significantly. The differences between the edges produced

are striking. In addition, while pix2pixHD is truly capable of generating high quality images, it is extremely costly to train. We used 4 GPUs and spent hundreds of dollars in AWS credit on the training phase. In practice, if high resolution is not necessary, there is little need to use pix2pixHD instead of pix2pix due to the high cost and the long duration of training.

From the generated images, we found that human is extremely sensitive at the task of recognizing human faces. We tend to picture random patterns as part of a human face. So our result gifs with human faces instead of other common objects can attract the viewers' attention in a different way than the original gifs.

From this project, we can see that it is feasible to replace real-world objects with manually created images. This feasibility allows us to quickly produce works similar to the styles of Tony Oursler, further exploring the boundary between human and objects. In a world where humans constantly fear being replaced by machines, we showed that machines can also be easily replaced by humans. The boundary between the two may not be as forbidding as one may have thought.

REFERENCES

[1] T. Wang et al. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In *Arxiv.org* https://arxiv.org/abs/1711.11585
[2] T. Karras et al. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In *Arxiv.org* https://arxiv.org/abs/1710.10196
[3] S. Xie and Z. Tu. Holistically-Nested Edge Detection. In *Arxiv.org* *https://arxiv.org/abs/1504.06375*