Localized Multi-Style Neural Style Transfer

Haper Liang *Meiyi Song* *Tongda Shen* *Zining Chen*

BROWN

CSCI 1470

INTRODUCTION

Have you ever wondered what you'll look like in a ship sailing away into the vast unknown of uncharted waters in search of a foreign land, or in a Spiderman movie dangling from skyscrapers and swiftly navigating between buildings?

To achieve this, this project leverages Convolutional Neural Networks to create a model that transfers visual styles to faces. While traditional neural style transfer methods often apply a single style globally to an image, our model allows for the fusion and transfer of multiple artistic styles to human faces specifically.
The project achieved the following goals:

1. Implement the paper A Neural Algorithm of Artistic Style (citation)
2. Implemented multi-style transfer and partial face transfer separately
3. Combined multi-style transfer and partial face transfer
4. Performed ablation experiments by comparing model performance with different content layers

METHODS

Dataset:
Style images and human face images are obtained from Kaggle: "Art Images: Drawing/Painting/Sculptures/Engravings" and "Human Faces". The preprocessing step involves converting the image to have (1) RGB channels, (2) pixels are between 0-1 (zero-centered) (3) resizing so that the longer dimension does not exceed the maximum dimension allowed.

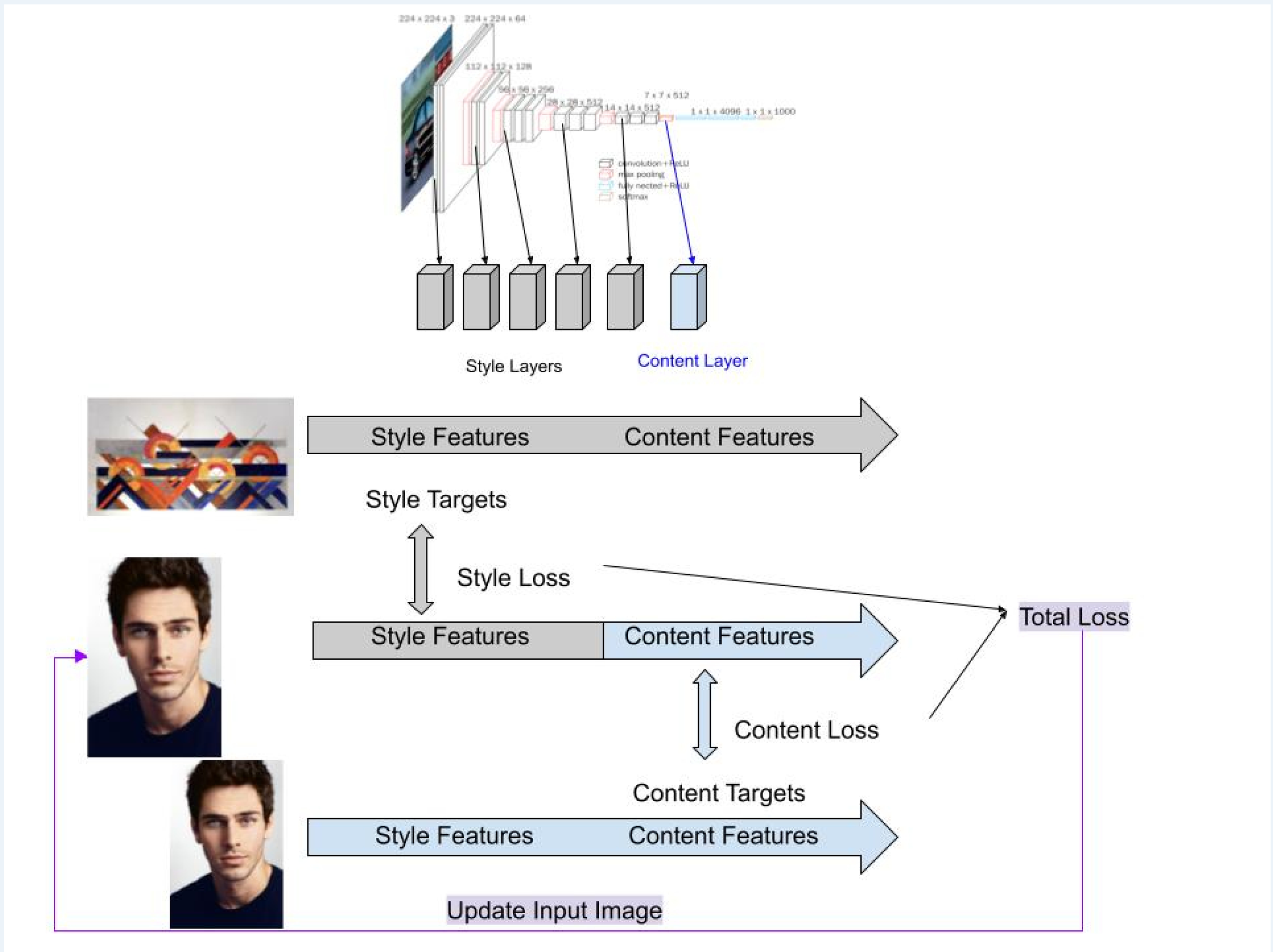
Model Architecture:
We imported VGG19 layers as used in the paper, using 5 convolution layers('conv1_1', 'conv2_1', 'conv3_1', 'conv4_1' and 'conv5_1') and 1 convolution layer('conv5_2') to store and represent features of the style images and content images . The style of the images are further preserved through calculation using Gram matrix. We excluded fully connected layers in both content and style representations.

Training:
Loss is calculated using mean squared error between the style and content feature representation of the generated image and original image and then summed. We trained our model using an Adams optimizer with a learning rate of 0.02 for all epochs. The model is trained on a Tesla T4 GPU for 5 epochs.

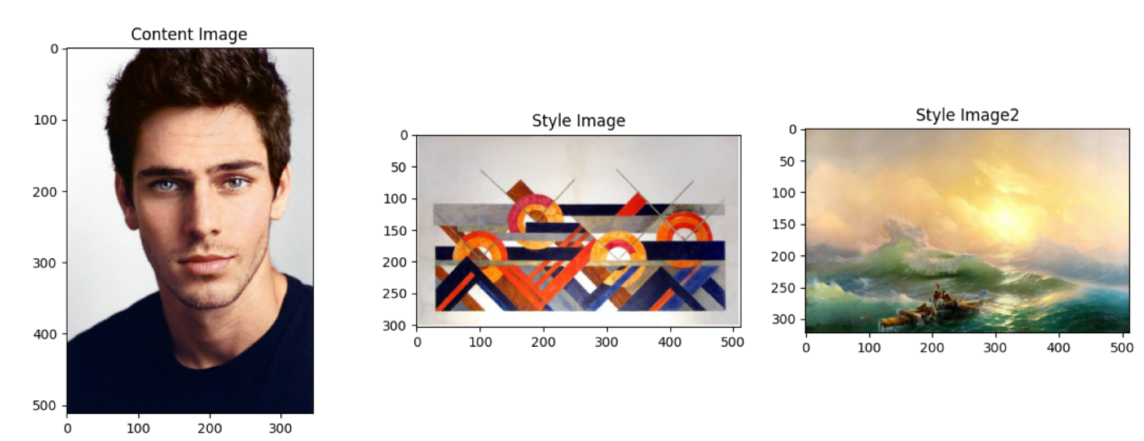
Multi-style Transfer:
The loss function in single style transfer is composed of content loss and style loss. The content loss guarantees that the generated image doesn't diverge too far, while the style loss measures a stylistic difference. We alter the loss function to incorporate a weighted sum over style images. Therefore, we can account for the relative percentage of each style we want in the final image.

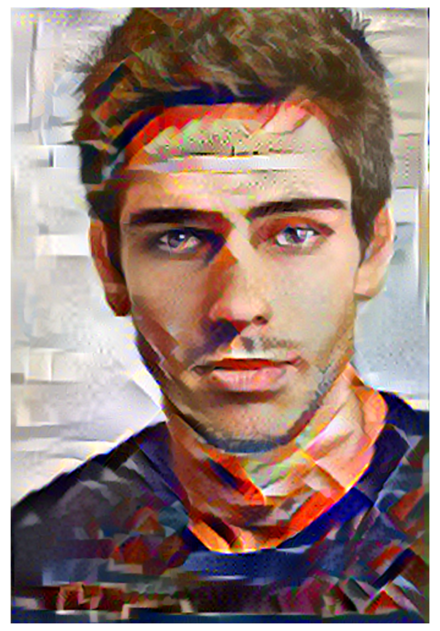
Partial Face Transfer:
The goal of partial transfer was to generate an output that selectively transfers the style of the facial area of an input content image. We believe that this could serve as a method to achieve a novel and unique art style or to aid users in their process of creation. Using the output from the model trained on multi style transfer, we added the MTCNN module to detect faces. We then store the coordinates of the detected faces and style transfer the the Original content image. The recognized facial area was then selected via stored coordinates. The original content image was preserved and stylized facial area after style transfer is pasted onto the original image to achieve style transfer specifically on the face.

Model Architecture

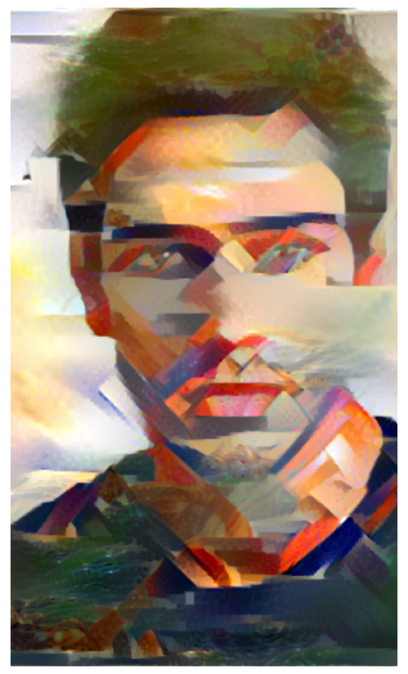


Results

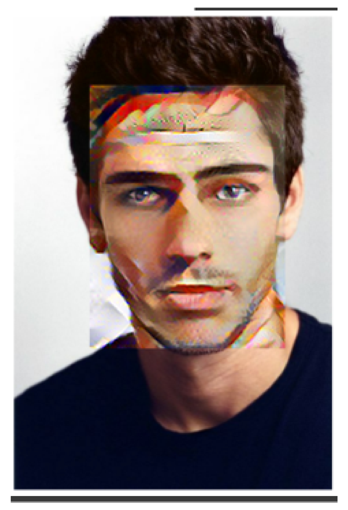




Single Style Transfer



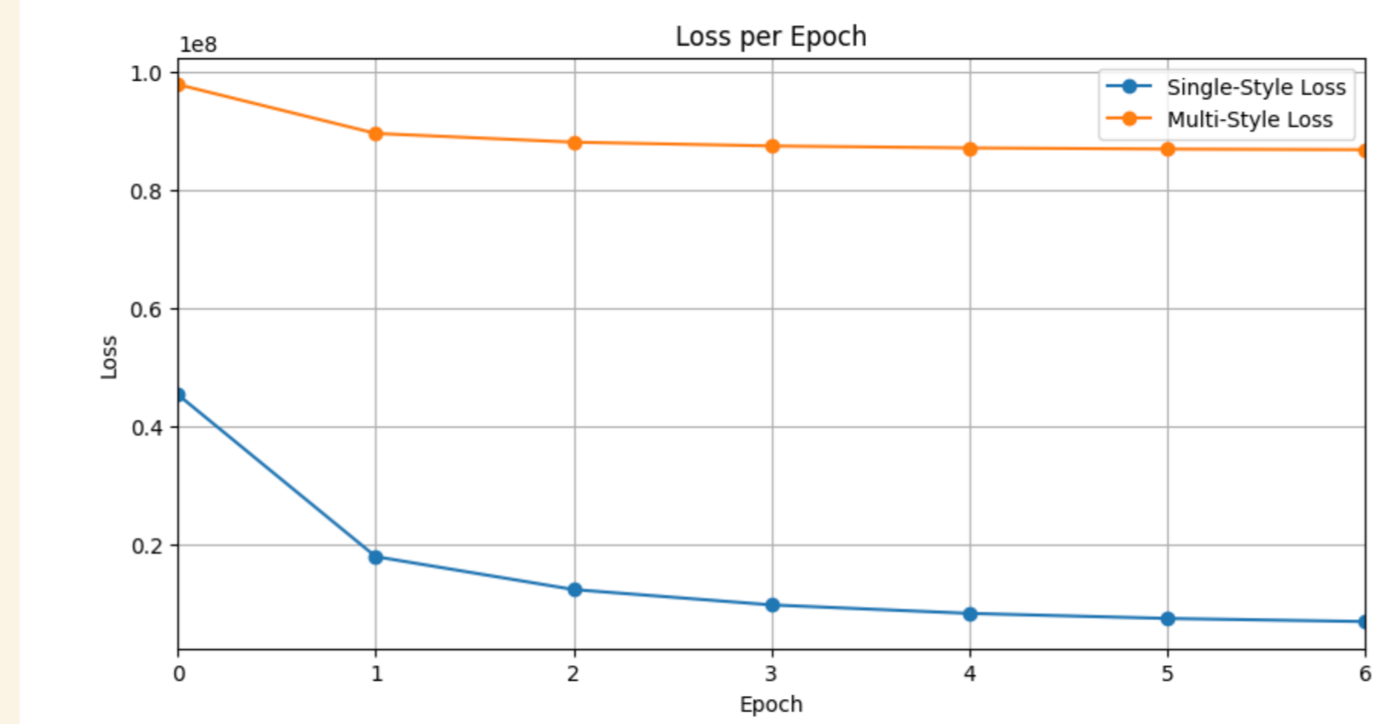
Multi Style Transfer



Single Style Face Transfer

Multi Style Face Transfer

multi style face transfer via mediapipe



single style and Multi Style transfer loss curve

CONCLUSION

Loss:
After training 7 epochs, the single style transfer model reached a loss of around 6.95e+6 while the multi-style transfer model reached a loss of 8.63e+7.

Different Layers:
Ablation experiments reveal that the deeper the content layers are, the more stylized the outputs. The shallower they are, the more that the stylized image is near to perfect or no change. This might be due to the fact that deeper layers in the CNN receive less detailed pixel information and preserve the high-level content of the image due to downsampling. Therefore, if the content image is shallow, the style transfers were made on smaller scales and thus not as evident. Training loss using different content layers is included in the paper.

Face Recognition:
Both methods of face recognition work successfully to recognize the face.

Future Directions

We would like to implement the L-BFGS optimizer as used in the paper as our current implementation uses an Adams optimizer. The current partial face transfer does not include part of the forehead, and we would like to expand on the current method to include the entirety of the upper face.

Also, although our model achieves good transfer results for human faces, we could further improve it on animal faces.

Original Paper

Neural Style Transfer: A Review by Yongcheng Jing, Yezhou Yang, *Member, IEEE,* Zunlei Feng, Jingwen Ye, Yizhou Yu, *Senior Member, IEEE,* and Mingli Song, *Senior Member, IEEE*