Fashion Style Transfer using CNN

Mishahal Salim Mohammed Palakuniyil[1], Atulya Shetty[2],

Jacob Jeevan John[3], Yashkumar Makwana[4]

[1,2,3,4]*Department of Computer Science, University of Alabama at Birmingham*

Abstract— In the past, style transfers have been widely used to impose an artistic style on arbitrary images. Here, we try to use style transfer with the help of Convolutional Neural Networks (CNN) to impose the artistic style of fashion wears, giving us a neural network algorithm to create artistic fashion images.

I. Introduction

Imposing images from one onto another can be done using various algorithms, but a texture imposing problem is considered to be more complex. Synthesizing photorealistic textures into the given input source pixel can be done using a large number of non-parametric algorithms. There are some algorithms that achieve extraordinary results, but they only use a low-level image to mention it to the texture transfer to get the target image as the output. Although, a style transfer algorithm would be able to jot down the detailed semantics of the target image and then notify the texture transfer to change the semantic content obtained in creating the final output image of the targeted image[1]. However, recent developments of computer vision systems in the field of Deep Convolutional Neural Network, the system has learned to draw out high-level semantic image content from the natural images. If the Convolutional neural network is trained with the proper amount of data on tasks such as recognizing the object, it will learn to draw out high-level image content which will be applied throughout the dataset, which includes texture recognition, style classification, and visual processing tasks too.

In this work, we use the Convolutional Neural Network to show the learning done by the network to draw out high-level content, using it independently and to manipulate natural images' final output contents. We use a pre-trained CNN model to extract the style layers from artistic images and transfer them onto fashion accessories.

II. Convolutional neural network

Convolutional Neural Networks (CNN) are a type of neural network that uses the linear mathematical operation, convolution. It is an operation which is performed on two functions where the output is how one of the functions is modified by others. Convolutional neural networks are basically networks that use this operation instead of matrix multiplication in at least one of the layers.[2] The architecture of this neural network is similar to that of any other neural network; it consist of an input and output layer with several hidden layers. The hidden layers basically are made of convolutional layers. For image, the network is able to classify and reconstruct parts of the images on the basis of style and content.[1] Figure 1 below shows how the image would be divided into layers on these basis.

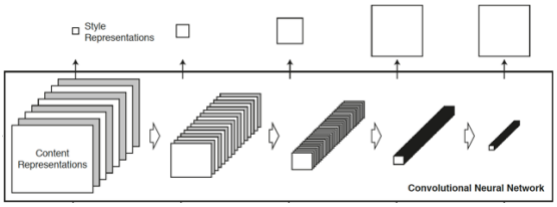

Figure 1: CNN Image Representation.

We use the VGG16 model of Convolutional Neural Network in this study. VGG16 is a CNN architecture model that was from the University of Oxford which was used to test the ImageNet (dataset with 14 million images) resulting in about 90% accuracy. It uses multiple 3x3 kernel-sized filters and was trained using NVIDIA Titan Black GPUs for weeks.[3] Figure 2 shows the architecture of VGG16 and has the 3x3 filter specifications mentioned as well.

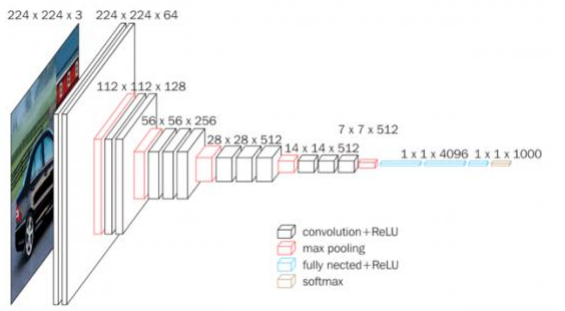

Figure 2: VGG16 Architecture

The layers of VGG has an ideal set of weights for our study and it makes it much less complex for the model to perform classification. This model also helps in extraction of the style layers which can then be applied to the arbitrary image to give us the style imposed image as output.

III. Data Used

The content images were basic stock fashion wear images that are most commonly found on the internet. For the style images, we went through various artistic images and handpicked images that we thought would go well with the model and the content images. We also deployed the VGG16 CNN which has been pre-trained on ImageNet using NVIDIA Titan Black GPUs for weeks; ImageNet is a dataset with about 14 million images belonging to 10,000 classes.

IV. Style Transfer Method

We performed our study in Python using Jupyter notebook. We used OpenCV and TensorFlow packages to process the images; resizing the input style and content images to be the same size, grabbing the dimensions and casting them onto the right shape to feed into the neural net. We used TensorFlow to load the VGG16, a CNN that is 19 layers deep.

Once everything is loaded and the input is given, we distinguish and separate the layers based on the style (i.e. color, shape, texture) and on the content (i.e. fashion wear) which makes it easier to work on the separated layers. The next task was to minimize the style loss and content loss. This can also be considered as an optimization problem where we minimize the distance between the input and output images (content loss) and the distance between style and output images (style loss)[4]. We minimize the content loss to maintain the content of the images and the style loss so the the new style is maintained. We used Adam Optimizer to get adaptive learning rates so that we could converge faster while training

We then converted our final model to a python script and with the help of a python library, argparser, we set up the script arguments. This enabled us to run the model from the command line which made it easier to test with different hyperparameters that were passed in as arguments to the model.

We started with a learning rate of 0.02 and trained for 100 to 1000 iterations. We found better results when training for 1000 iterations compared to the other.

V. Results

As shown in the figures below are the results of our study. As seen, the fashion wear images is the content images and the art images are the style images; the bigger images are the result images.

Some of the combinations used do not produce results as expected. Figure 3 has the result that trained for 100 iterations while the others are ones trained for 1000 iterations.

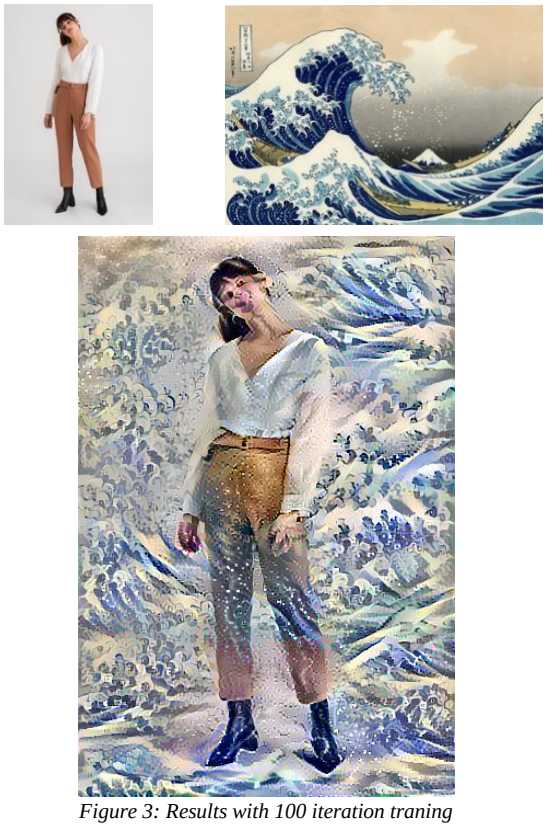

Figure 3: Results with 100 iteration traning

Figure 4: Results with apparel to apparel combination

Figure 5: Results with apparel to art combination (not imposed well)

Figure 6: Results with apparel to art combination

Figure 7: Results with apparel to landscape combination

Figure 8: Results with apparel to cartoon combination (not as expected)

VI. Discussion And Future work

The model could have achieved with deeper with hyperparameters and training the model for longer iterations. However, we were limited by time and constrainted by computing resouces (Cheaha Supercomputer).

For future work, we would like to rerun the model experimenting with semantic segmentation and markov random field blending[5] for constraining the style transfer on a localized area on the image. We had used global style transfer in this study.

VII. Conclusion

Our goal was to come up with a novel approach to impose different styles from artistic or abstract images on fashion wear images using CNN. As evident in the results, the vanilla neural style transfer model is not suitable for our application since it is unconstrained and applies style to the entire image, and results in disorted/artistic output rather than a more realistic output image.

VIII. References

[1] A Neural Algorithm of Artistic Style

[2] Convolutional Neural Networks, Wikipedia

[3] VGG16- CNN for Classification and Detection

[4] Style Transfer – Styling images with CNN

[5] Localized Style Transfer

[6] Neural Style Transfer [Code reference]