# Image stylization based on deep convolutional neural network

*Guangrui Ding* grding@bu.edu
*Tom Panenko* tompan@bu.edu
*Zhenghao Sun* szh1007@bu.edu
*Shangzhou Yin* syin10@bu.edu

(a) content image (b) style image (c) result image

Figure 1. An example of style transfer performed on (a) and (b), producing (c).

## 1. Task

Our goal is to apply modern deep learning techniques to perform neural style transfer, which is shown in Figure 1. While several algorithms exist that perform neural style transfer, they use techniques that were the state of the art in 2015-2017, which could be improved with the newly developed models(GAN and ResNet), upon in terms of result quality and training/processing time.

## 2. Related Work

We have read the results of some relevant literature as a basis for our project. In ***Image Style Transfer Using Convolutional Neural Networks***[1], Leon A. Gatys mentions that the style transfer from one image to another can be considered as a process of texture migration. In his model, we can obtain the feature spectrum between different layers of these convolutional neural networks, and the texture of an image can be represented by the relationship between these different layers.

In the same year, in his second paper, ***A Neural Algorithm of Artistic Style***[2], another creative idea was introduced: we can freely separate and reconstruct the content and style of an image because the features of the image can be extracted from the neural network to represent the texture of the image, and each layer in a deep convolutional neural network does not retain the same information. In a deep convolutional neural network, the higher layers retain semantic information and the lower layers learn information such as texture. Using this feature, the content and style of the image can be well separated.

In 2016, Justin Johnson, Alexandre Alahi and Feifei Li from Stanford University published the paper ***Perceptual Losses for Real-Time Style Transfer and Super-Resolution***[3], which improves on the approach used by Gatys by proposing a fast stylization method that can generate stylized images in just a few seconds on the GPU. The network structure proposed in this paper adds an image transformation network, which is essentially the last deep residual convolutional network (Resnet), and a loss network using a pre-trained ***VGG-16***[4] network. The two are connected and the training set images are input for training, and the total loss value is minimized after going through the whole system, and the parameters in the image transformation network are continuously updated and optimized until the final ideal model is obtained. After the training is completed, it only needs to input an original image into the model to get the stylized image output immediately.[5]

Then, due to the rise of **generative adversarial networks(GAN)**, researchers found GAN to be excellent for this task. StyleGAN network is more advanced than the GAN network, which introduces significant modification to the generative model. StyleGAN has been developed to the third version, which can handle video generation with better textural detail recovery [6]. Meanwhile, U-net architecture, which is widely used in biomedical images, contains encoding and decoding parts. It can extract micro and macro features within the image, which might also be helpful for our project[7].

## 3. Approach

We seek to implement style transfer processing of images by deep convolutional neural networks. The input of the whole migration system is two images, one providing the style and one providing the content, and the output is the result of combining the provided style and content. In order to achieve a good separation of image content and style, we intend to use a classical deep convolutional neural network VGG-16 to extract the texture information and content information of the image, and use the mathematical properties of the gram matrix to calculate the style loss function to approximate the style of the image. After obtaining the content loss function and style loss function, the back propagation method is used to make the gradient of the loss function decrease as the number of iterations increases. Finally, by comparing and testing different optimization methods (Newton's method, Newton-like method, Monte-Carlo method), a suitable optimizer is chosen so that the loss function can converge quickly and the desired stylized image is obtained [8].

### 4. Dataset and Metric

**Datasets**: For stylization purposes, we only need an input image and a reference style image. We intend to use a portion of the LSUN dataset for this purpose. Additionally, we will use photos we have taken ourselves to evaluate the performance of our methods.
**Metrics**:We will utilize the Mean Square Error (MSE) metric to calculate the content similarities between the input image, also the reference image for style similarities, and the generated image. Additionally, the speed of the training process is a crucial aspect of our evaluation. We aim to reduce the time and cost involved in the training process without compromising the similarities between the reference and generated images. In addition to the above, we also need to evaluate the content distortion in the generated image compared to the reference image. To achieve this, we will use the Frechet Inception Distance (FID) metric, which measures the content-wise similarity between the two components. FID serves as a quantitative indicator that determines how closely the texture of the generated image matches the raw input image. By using both the MSE and FID metrics, we can obtain a comprehensive understanding of the quality of images generated by the generative model. Our primary objective is to create high-quality images with minimal content distortion and a reduced training time, making our method efficient and effective.

### 5. Approximate Timeline

| **Task** | **Deadline** |
|---|---|
| Research and read related papers about Image Stylization | 02/28/23 |
| Project Proposal | 02/28/23 |
| **Data Collection and Metric Defining** | |
| Collect several Reference Style Image and Input Images | 03/05/23 |
| **Develop Deep Learning Models** | |
| Research Deep Learning Models such as VGG, ResNet, GAN | 03/08/23 |
| Develop Options for Deep Learning Models | 03/15/23 |
| Research Optimization Methods | 03/20/23 |
| Test Each Model with different optimizers and Compare the loss with the Baseline Model | 04/01/23 |
| Project Status Report | 04/06/23 |
| Revise/Improve the Models based on the Loss | 04/15/23 |
| Develop GUI and Present the Output Image | 04/20/23 |
| **Final Step** | |
| Presentation Slide | 04/27/23 |
| Project Presentation | 04/27/23 |
| Final Project and Code | 05/05/23 |

### References


1）Leon A. Gatys, Alexander S. Ecker, Matthias Bethge: “Texture Synthesis Using Convolutional Neural Networks”, 2015; arXiv:1505.07376.

2）Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style[J]. arXiv preprint arXiv:1508.06576, 2015.

3）Johnson J, Alahi A, Fei-Fei L . Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. Springer, Cham, 2016.

4）Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

5）Fujun Luan, Sylvain Paris, Eli Shechtman, Kavita Bala: “Deep Photo Style Transfer”, 2017; arXiv:1703.07511.

6）T. Wei et al., "E2Style: Improve the Efficiency and Effectiveness of StyleGAN Inversion," in IEEE Transactions on Image Processing, vol. 31, pp. 3267-3280, 2022, doi: 10.1109/TIP.2022.3167305.

7）M. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Quebec City, QC, Canada, Sep. 2015, pp. 234-241.

8）Jing Y, Yang Y, Feng Z, et al. Neural style transfer: A review[J]. IEEE transactions on visualization and computer graphics, 2019, 26(11): 3365-3385.