Universal Style Transfer via Feature Transforms

Team : **Mandelbrot**

Team Members : Krishna Mahesh Teja
Sai Krishna Charan Dara
Sai Deva Harsha Annam
Venkata Susheel Voora

Motivation

- Headshot portraits are commonly taken in professional settings, and a lot of time and effort is spent in styling these photographs, through various editing methods.
- This project uses a combination of various techniques to style a Content image ‘C” using a professionally taken Style image ‘S’.
- Two different kind of approaches exist for this problem in the literature
    - Optimization based methods that use covariance matrix to formulate the correlation between features. They can handle arbitrary values with a good visual quality but have high computational costs.
    - Feed forward based methods that work for a fixed number of styles with a compromised visual quality but with a good computational efficiency.
- Objective is to develop a universal style transfer approach with a decent visual quality and efficiency.

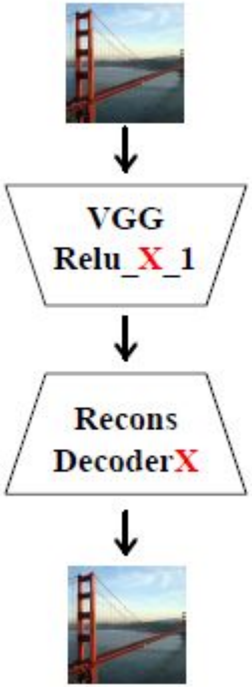
Approach

We design a universal style transfer pipeline in 3 steps -

- Reconstruction
  - Encoder and Decoder
- Single level Stylization
  - Whitening and Coloring Transform
- Multi level Stylization

In each intermediate layer, our objective is to transform the extracted content features such that they exhibit the same statistical characteristics as the style features of the same layer.

Reconstruction

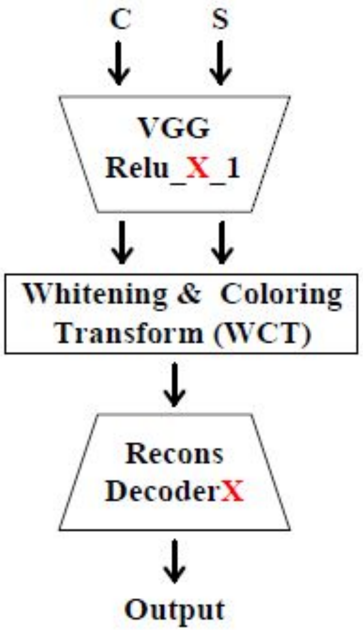


- We use the VGG-19 network as the feature extractor (encoder), and train a symmetric decoder to invert the VGG-19 features to the original image.
- To evaluate with features extracted at different layers, we select feature maps at five layers of the VGG-19, i.e., Relu_X_1 (X=1,2,3,4,5), and train five decoders accordingly.
- Loss function used is

  $$L = \|I_o - I_i\|_2^2 + \lambda\|\Phi(I_o) - \Phi(I_i)\|_2^2$$

  where ϕ is the VGG encoder that extracts the Relu_X_1 features.
- Once trained, encoder and decoder are fixed through all the experiments.

Single level stylization



- Given a pair of content image I_c and style image I_s, we first extract their vectorized VGG feature maps f_c and f_s at a certain layer (e.g., Relu_4_1), where H_c, W_c (H_s, W_s) are the height and width of the content (style) feature, and C is the number of channels.
- The decoder will reconstruct the original image I_c if f_c is directly fed into it.
- The goal of WCT is to directly transform the f_c to match the covariance matrix of f_s

Tentative plan

- Mid eval - Completing the code for single level stylization using WCT
- Final eval - Extending the code for multi level stylization