Final Evaluation

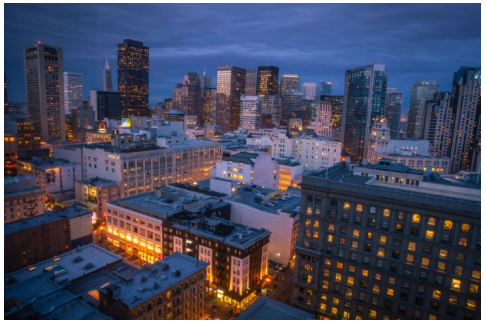
Photorealistic Style Transfer via Wavelet Transforms

Team : Framed

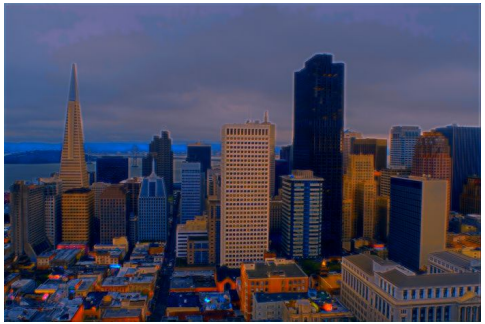
Objective

- Given a pair of images - S : Style image and C : Content image, perform photorealistic style transfer
- Leveraging wavelet transform to overcome limitations of spatial distortions, and introduction of unrealistic artifacts in the final image.
- The paper proposes to perform this via an end-to-end photorealistic style transfer model that allows to remove the additional post-processing steps

Content Image (C)



Style Image (S)



Style Transfer Output

Method

Overview of the method

- ➔ Use the **VGG19** model's feature extraction layers as the encoder and its corresponding mirror as the decoder.
- ➔ Replace the max pooling and unpooling layers with **Haar Wavelet Pooling**.
- ➔ Apply **WCT** after each scale (conv1 X, conv2_X, conv3 X and conv4 X). This is referred to as **progressive stylisation**.

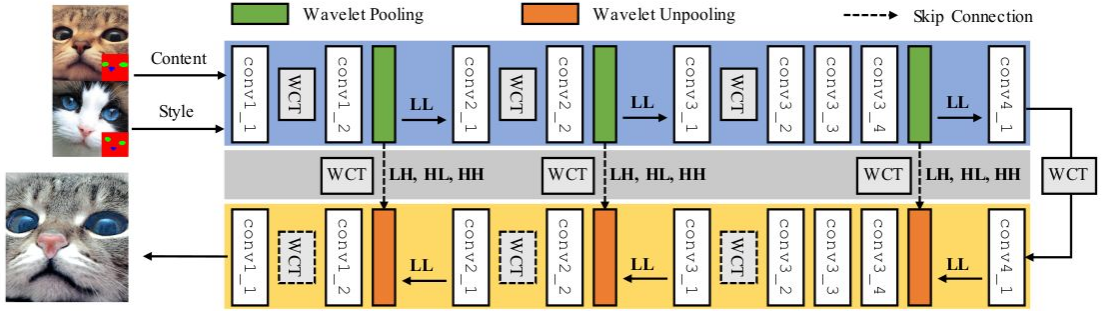


Figure 11: Overview of the proposed progressive stylization. For the encoder, we perform WCT on the output of `convX_1` layer and skip connections. For the decoder, we apply WCT on the output of `convX_2` layer, which is optional.

VGG-19 Model

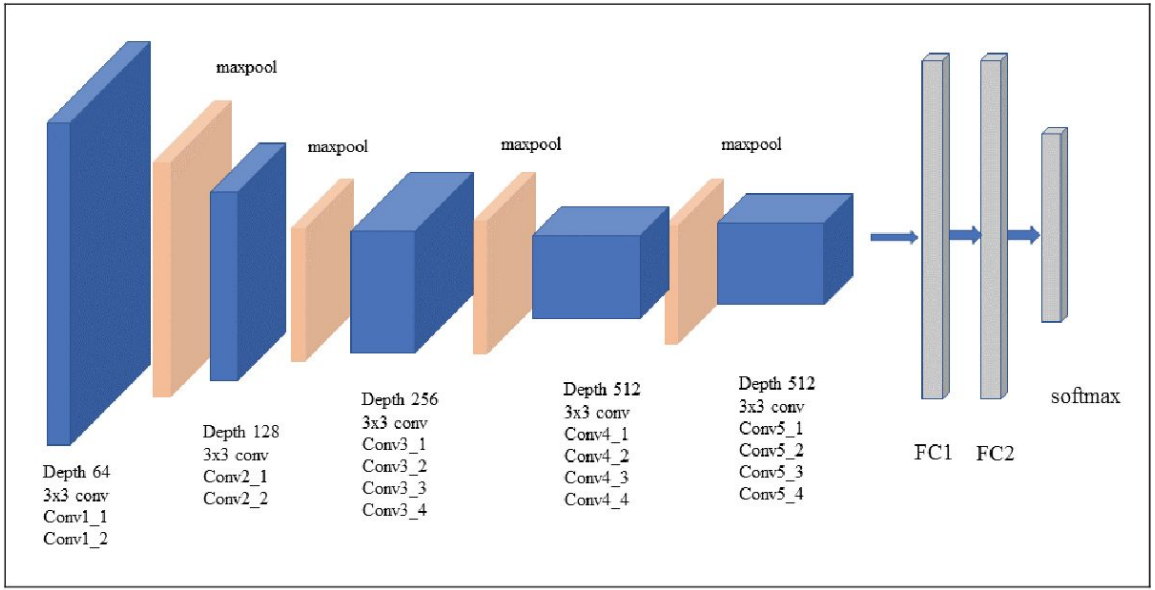


Fig. 3. VGG-19 network architecture

- For our use case, we use the layers upto conv4_1 for our encoder. We use the mirror of this model as our decoder.

Haar Wavelet Pooling

- Haar wavelet pooling has four kernels, { LL^T LH^T HL^T HH^T}
- The low (L) and high (H) pass filters are-

$$L^\top = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \end{bmatrix}, \quad H^\top = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 \end{bmatrix}.$$

- Here, the low-pass filter captures smooth surface and texture while the highpass filters extract vertical, horizontal, and diagonal edgelike information
- One important property of the wavelet pooling is that the original signal can be exactly reconstructed by mirroring its operation
- With this favorable property, the proposed model can stylize an image with minimal information loss and noise amplification.

WCT Procedure

- We first extract the features for both the content (c) and style (s) images. Let us denote this as f_c and f_s.
- We now apply WCT (Whitening and Coloring Transforms).
- We take the output (denoted by f_{cs}) of WCT and pass it to the next layer of the architecture

Whitening and Color Transforms

Whitening Transform

- We first center f_c by subtracting its mean vector.
- Then we transform f_c linearly such that the feature maps are uncorrelated.
- $$\hat{f}_c = E_c \, D_c^{-\frac{1}{2}} \, E_c^{\top} \, f_c$$
- D_c is a diagonal matrix with the eigenvalues of the covariance matrix. E_c is the corresponding orthogonal matrix of eigenvectors.

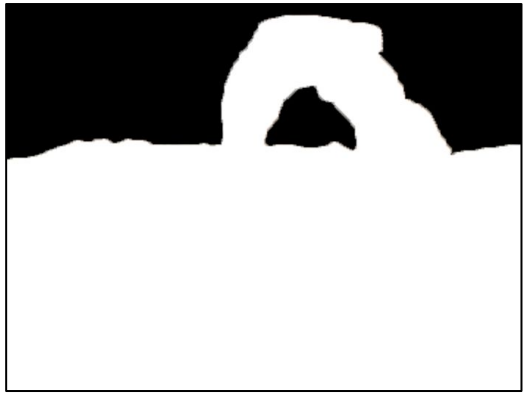
Color Transform

- We first center f_s by subtracting its mean vector.
- We now carry out the coloring transform as follows to obtain f_{cs}.
- $$\hat{f}_{cs} = E_s \, D_s^{\frac{1}{2}} \, E_s^{\top} \, \hat{f}_c$$
- D_s is a diagonal matrix with the eigenvalues of the covariance matrix. E_s is the corresponding orthogonal matrix of eigenvectors.

WCT Core & WCT Core Segment

In WCT Core, we apply WCT, take the output and pass it to the next layer of the architecture without using a semantic segmentation mask.

In WCT Core Segment, we use semantic segmentation of the inputs to avoid content-mismatch problem (difference in content between input and reference images could result in undesirable transfers between unrelated content), which improves the photorealism of the results.

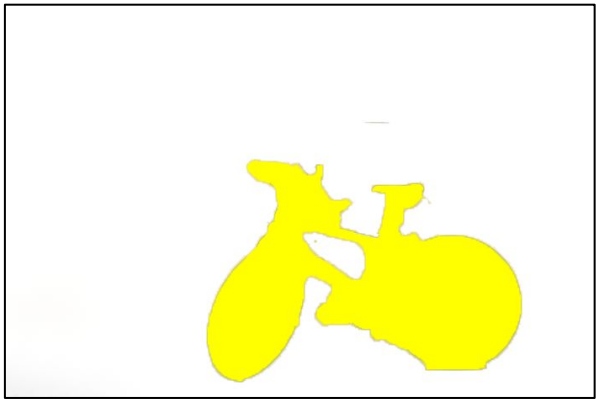


WCT2 with semantic segmentation

In WCT2, to apply semantic segmentation map to the artistic style transfer methods, we followed the spatial control techniques proposed by the authors [1, 2, 3] respectively. Artistic style transfer methods generate undesired distortions and artifacts and often fail to maintain the structural information despite the spatial control with segmentation maps. In comparison, because of wavelet corrected transfer, WCT2 prevents unrealistic artifacts and preserve the structure information such as edges.

[1] https://arxiv.org/abs/1703.07511
[2] https://arxiv.org/abs/1703.06868
[3] https://arxiv.org/abs/1705.08086

Content Image



Content Segmentation

Style Image

With Segmentation

Without Segmentation

Unpooling Options

The paper suggests different unpooling options in the model architecture, such as - *summation, split pooling, learnable, and concatenation* based pooling. Further, the **summation** and **concatenation** based methods are analysed in the paper :

- Concatenation (*CAT5)* pooling outputs look much better than summation based

- *CAT5* unpooling performs channel-wise concatenation of the four feature components from the corresponding scale plus feature output before the wavelet pooling

- Summation based is the conventional transposed convolution and summation

- *CAT5* enables the network to learn the weighted sum of components

Summation Pooling

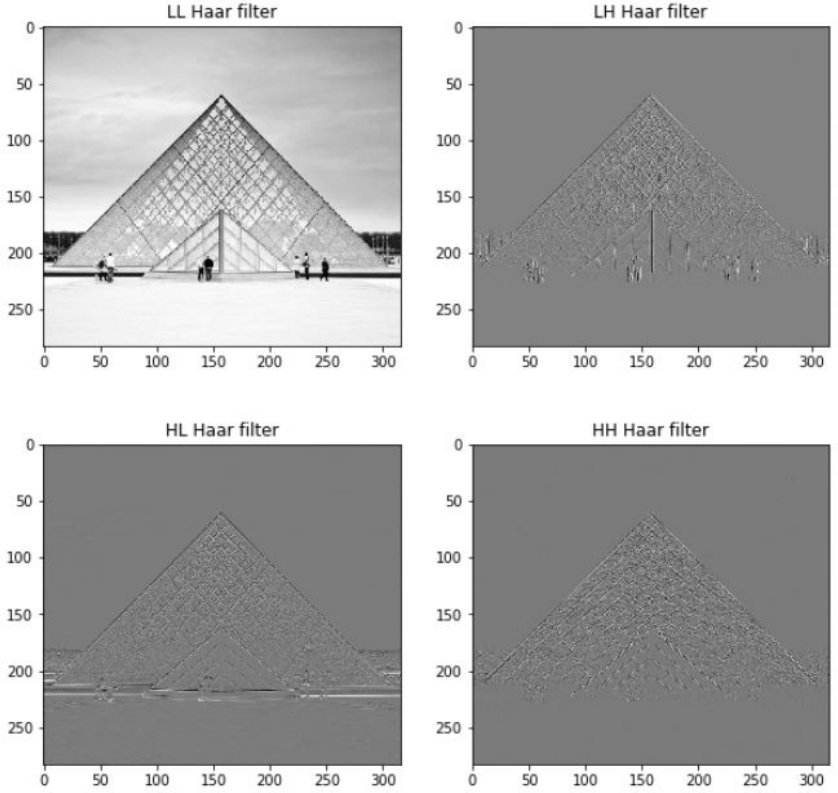
Concatenation Pooling

Experiments

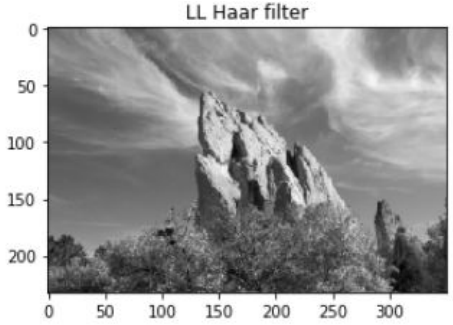
Haar-wavelet

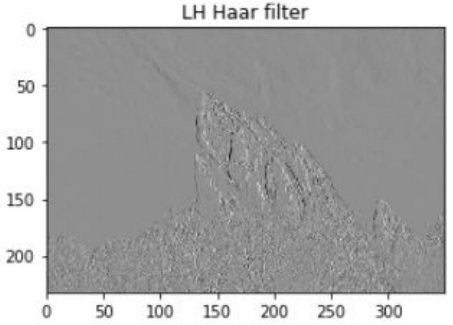
- To test the working of the Haar wavelet convolution filters, we generate the 4 kernels - LL, LH, HL, HH as mentioned in the paper.
- We apply the wavelet filters to two experiment images and show our observations in the next slides.
- It is observed that :
    - The **LL filter convolution has little to no change on the image**. This is because it is a low pass filter and captures only the lower frequencies in the image
    - The HH filter on the other hand captures the **higher frequency, intricate edges of the** image as we can see in the case of the Louvre museum image's glass edges

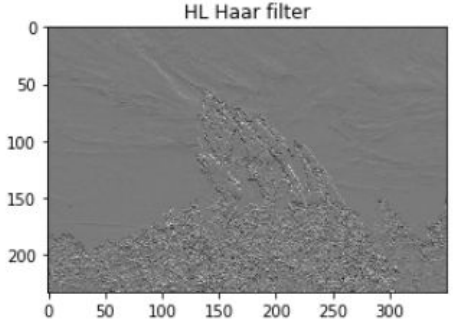
Experiment Image : 1

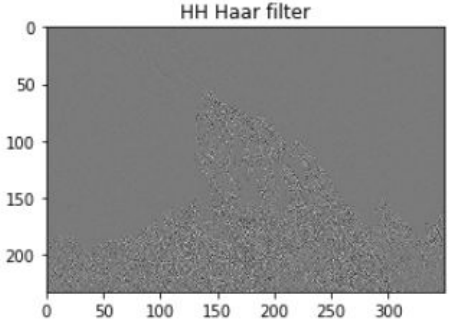


Experiment Image : 2









Results

Result Image : 1

Style Image

Content Image

Result Image

Effect of Skip Connections

WCT (with skip)

WCT (LL only)

- The low frequency component captures **smooth surface** and texture while the high frequency components detect edges.
- More specifically, it implies that applying WCT to LL of the encoder affects overall texture or surface while applying WCT to the high frequency components (i.e., LH, HL, HH) stylize edges.

Result Image : 2

Style Image

Content Image

Result Image

Result Image : 3

Style Image

Content Image

Result Image

Result Image : 4

Style Image

Content Image

Result Image

Result Image : 5

Style Image

Content Image

Result Image

Result Image : 6

Style Image

Content Image

Result Image