Image Transformations

Abstract

Visualizing image transformations can be costly and time-consuming. Applying different styles (modern, minimalistic, contemporary, etc.) before starting the remodeling work helps align with the owner's vision. My method, utilizing autoencoder architecture with block training, high-frequency residual skip connections, and bottleneck feature aggregation, achieves photorealistic style transfer for images.

Introduction

Deep Neural Networks have been applied to object and face recognition. In 2015, Gatys et.al applied this to the realm of fine art [1]. They took natural images and stylized them with famous artworks by extracting the content representation and style representation of each image. In 2017, I introduced an autoencoder approach to the task [2]. I expanded these works to style transferring of images in general. The input to my algorithm is an image used as the content and another image used as the style. My final model uses an autoencoder approach, like that of Li et al., to output a generated image with the new style.

Dataset

To train the autoencoder I will use the MSCOCO [3] dataset that was used in a few of the Whiten-Color Transform (WCT)papers. In the training set of this dataset there are 118,288 images, 5000 images in the validation set, 40670 images in the test set. We also used the ADE20K [4] dataset for my initial semantic segmentation approach (described in experiments below) which has 25,574 training images and 2,000 images in the validation set. This dataset contains the semantic segmentation of each scene.

Research Question

1. How does the utilization of an autoencoder architecture with block training, high-frequency residual skip connections, and bottleneck feature aggregation contribute to achieving photorealistic style transfer for images?
2. What are the computational and memory requirements of training and deploying the autoencoder model for image style transfer, considering the large size of the datasets used?
3. How effective are the techniques adopted from PhotoWCT and PhotoNet in refining the autoencoder model and enhancing the quality of style-transferred images?
4. What are the limitations and potential challenges associated with the proposed approach, particularly in terms of scalability and applicability to different styles and content images?

Approach

I will utilize an autoencoder architecture for style transfer. Specifically, for my final model, I will adopt the baseline of WCT (Whitening and Coloring Transform). Here's a breakdown of my method:

Feature Extraction and Reconstruction: I will employ the VGG-19 network as the feature extractor (encoder), and a symmetric decoder will be trained on images from the MSCOCO dataset to invert the VGG-19 features and reconstruct the content image. Both my encoder and decoder will be trained to minimize pixel reconstruction loss and feature loss.

Whitening and Coloring Transformations (WCT): Once trained, my encoder and decoder will be fixed. To perform style transfer, both the content and style images will be fed into the encoder to extract vectorized feature maps. I will apply WCT transformations to match the covariance matrix of the featured content with that of the featurized style. This will involve whitening to maintain global structure and coloring to introduce the style from the style image. The transformed features will then be decoded to reconstruct the image with the new style.

I will further refine my autoencoder by incorporating techniques from PhotoWCT and PhotoNet. Additionally, I will conduct further research to explore additional methods for improvement.

References

[1] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. 2015. https://arxiv.org/abs/1508.06576

[2] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin lu, and Ming-Hsuan Yang. Universal style transfer via feature transformations. 2017. https://arxiv.org/abs/1705.08086

[3] https://cocodataset.org/#download

[4] https://groups.csail.mit.edu/vision/datasets/ADE20K/