DEEP PHOTO STYLE TRANSFER

Harshil Goel, Nishant Ravinuthala, Mukul Hase

Abstract

The paper aims to enhance the effects of the deep learning approach to neural style transfer. In particular, the paper augments the style transfer approach to transfer near perfectly, the style of one photo to the other, while keeping the photo as realistic as possible. In the traditional style transfer approach, even when both the input and reference images are photographs, the output still exhibits distortions reminiscent of a painting. The paper's contribution is to constrain the transformation from the input to the output to be locally affine in colorspace, and to express this constraint as a custom fully differentiable energy term. This approach successfully suppresses distortion and yields satisfying photorealistic style transfers in a broad variety of scenarios, including transfer of the time of day, weather, season, and artistic edits.

Introduction

A fairly involved challenge, photographic style transfer seeks to transfer the style of a reference style photo onto another input picture. By appropriately choosing the reference style photo, one can make the input picture look like it has been taken under a different illumination, time of day, or weather, or that it has been artistically retouched with a different intent. Existing techniques are very narrow in scope, and work for very particular cases, In this paper, a deep-learning approach to photographic style transfer is introduced that handles a large variety of image content while accurately transferring the reference style. The paper tries to nullify the painting effects produced by the neural style approach by preventing spatial distortion and constraining the transfer operation to happen only in color space.

Contributions

1. Structure preservation : The kind of transformation the paper tries to seek is where the colorspace is drastically affected but there is no geometric distortion.

2. Semantic accuracy and transfer faithfulness: The work also ensures that the transfer happens between semantically equivalent subregions and within each of them, the mapping is close to uniform. The algorithm preserves the richness of the desired style and prevents spillovers

Method

The Neural Style algorithm is augmented by using two core ideas:

Photorealism Regularization:

This technique is applied on the input image. As the input image is being converted to the output image, since the input image is photorealistic, in order for the output image to be realistic, any term added to the input image if results in a distortion, is penalized which is captured by the following loss function. The aim is to seek an image transform that is locally affine in color space, that is, a function such that for each output patch, there is an affine function that maps the input RGB values onto their output counterparts.

$$\mathcal{L}_m = \sum_{c=1}^{3} V_c[O]^T \mathcal{M}_I V_c[O]$$

Augmented Style Loss with semantic segmentation:

The Gram matrix approach in neural style transfer takes into account the global style of the image and applies it to the output image. That, however, isn't suitable for photo realistic style transfer as we want to avoid spilling of textures to semantically incorrect places as much as possible. Thus, the paper makes use of semantic segmentation masks that other papers like Neural Doodle make use of. These semantic segmentation masks are concatenated with the input image as additional channels and the style loss is updated as follows.

$$\mathcal{L}_{s+}^{\ell} = \sum_{c=1}^{C} \frac{1}{2N_{\ell,c}^2} \sum_{ij} (G_{\ell,c}[O] - G_{\ell,c}[S])_{ij}^2$$

$$F_{\ell,c}[O] = F_\ell[O] M_{\ell,c}[I] \quad F_{\ell,c}[S] = F_\ell[S] M_{\ell,c}[S]$$

where C is the number of channels in the semantic segmentation mask, $M_{l,c}[\cdot]$ denotes the channel c of the segmentation mask in layer l, and $G_{l,c}[\cdot]$ is the Gram matrix corresponding to $F_{l,c}[\cdot]$.

The photorealistic style transfer is finally completed by combining the content, style and deformation loss:

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_{s+}^\ell + \lambda \mathcal{L}_m$$

where L is the total number of convolutional layers and l indicates the lth convolutional layer of the deep neural network. Γ is a weight that controls the style loss. α_l and β_l are the weights to configure layer preferences. λ is a weight that controls the photorealism regularization. L^l_c is the content loss. L^l_{s+} is the augmented style loss. L_m is the photorealism regularization.