Towards Deep Style Transfer: A Content-Aware Perspective

Yi-Lei Chen
https://sites.google.com.tw/site/fallcolor
Pixart Imaging Inc.
Hsinchu, Taiwan

Chiou-Ting Hsu
http://www.cs.nthu.edu.tw/~cthsu/candy.html
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan

Motivation

Recently, it has been shown that one can invert a deep convolutional neural network originally trained for classification tasks to transfer image style. There is, however, a dearth of research on content-aware style transfer. In this paper, we generalize the original neural algorithm [1] for style transfer from two perspectives: *where to transfer* and *what to transfer*. To specify where to transfer, we propose a simple yet effective *masking out* strategy to constrain the transfer layout. To illustrate what to transfer, we define a new style feature by high-order statistics to better characterize content coherency.

Methodology

Given a source image (or content image) $\mathbf{c}$ and a target image (or style image) $\mathbf{s}$, [1] aims to synthesize an image $\mathbf{x}$ which simultaneously shares the visual content of $\mathbf{c}$ and the style representation of $\mathbf{s}$. Specifically, the image rendering was modelled as an optimization problem by minimizing the difference between $\mathbf{c}$ and $\mathbf{x}$ and the difference between $\mathbf{s}$ and $\mathbf{x}$ in terms of content and style features, respectively. The authors characterize both features by the deep convolutional neural network (CNN). The desired image was obtained by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \lambda \sum_{l \in l_c} \left\| \mathbf{F}_{l(\mathbf{x})} - \mathbf{F}_{l(\mathbf{c})} \right\|^2 + \sum_{l \in l_s} \left\| \mathbf{G}_{l(\mathbf{x})} - \mathbf{G}_{l(\mathbf{s})} \right\|^2 + \gamma \Gamma(\mathbf{x}) \,, \quad (1)$$

where the content feature $\mathbf{F}_{l(\mathbf{x})}$ is the layer-wise response and the style feature $\mathbf{G}_{l(\mathbf{x})} = \mathbf{F}_{l(\mathbf{x})}{}^{\mathrm{T}} \mathbf{F}_{l(\mathbf{x})}$ encodes cross-feature dependencies globally.

We formulate the generalized style transfer based on Equation (1) under two additional constraints: *where to transfer* and *what to transfer*. To constrain *where to transfer*, we introduce a diagonal matrix $\mathbf{M}_{l(\mathbf{x})}$, whose $(i,i)^{\text{th}}$ entry m_i $(0 \le m_i \le 1)$ is a soft indictor of feature aggregation, to specify the spatial correspondence. To constrain *what to transfer*, we propose a new feature statistics $\widehat{\mathbf{G}}_{l(\mathbf{x})} = (\mathbf{P}_{l(\mathbf{x})}\mathbf{F}_{l(\mathbf{x})})^{\mathrm{T}}(\mathbf{P}_{l(\mathbf{x})}\mathbf{F}_{l(\mathbf{x})})$, by introducing a high-order convolutional matrix $\mathbf{P}_l$, to better match the style representation. Finally, we propose to embed both two constraints into the style loss of Equation (1) and derive the layer-wise gradient in a general form:

$$\nabla_{\mathbf{F}_{l(\mathbf{x})}} = \sum_{j=1}^{J} \sum_{k=1}^{K} \mathbf{P}_l^{(j)\,\mathrm{T}} \mathbf{M}_{l(\mathbf{x})}^{(k)\,\mathrm{T}} \left(\mathbf{M}_{l(\mathbf{x})}^{(k)} \mathbf{P}_l^{(j)} \mathbf{F}_{l(\mathbf{x})}\right)\left(\widehat{\mathbf{G}}_{l(\mathbf{x})}^{(k)} - \widehat{\mathbf{G}}_{l(\mathbf{s})}^{(k)} / N_l^{(k)}\right),$$

$$\text{where } \widehat{\mathbf{G}}_{l(\mathbf{x})}^{(k)} = \left(\mathbf{M}_{l(\mathbf{x})}^{(k)} \mathbf{P}_l^{(j)} \mathbf{F}_{l(\mathbf{x})}\right)^{\mathrm{T}} \left(\mathbf{M}_{l(\mathbf{x})}^{(k)} \mathbf{P}_l^{(j)} \mathbf{F}_{l(\mathbf{x})}\right). \quad (2)$$

Results

We show an example for real-life photo transfer in Fig. 1. Using the semantic masks estimated by image matting, we successfully transfer the dogs' appearance without either changing background or producing noticeable artifacts. Please refer to our paper for more style transfer results

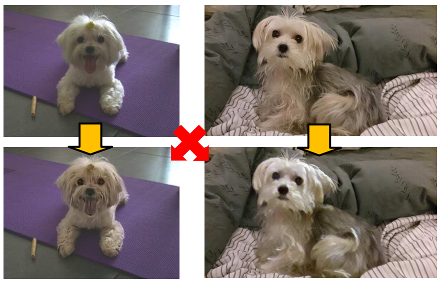

Figure 1: An example of content-aware style transfer. The face appearance of two different breeds of dogs, *Maltese* and *Yorkshire terrier*, are exchanged by our method.

[1] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proc. CVPR*, 2016.