A Content Transformation Block For Image Style Transfer

RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG

IWR
Interdisciplinary Center
for Scientific Computing

Heidelberg Collaboratory
HCI
for Image Processing

Vision
Computer
group heidelberg

Dmytro Kotovenko Artsiom Sanakoyeu Pingchuan Ma Sabine Lang Björn Ommer

Heidelberg University, HCI/IWR, Germany {first.last}@iwr.uni-heidelberg.de

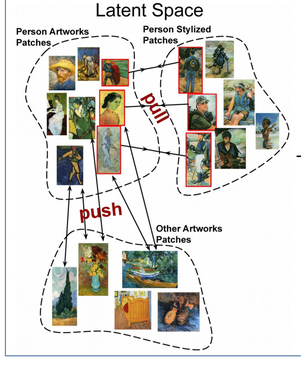
http://bit.ly/CTstyle

Motivation

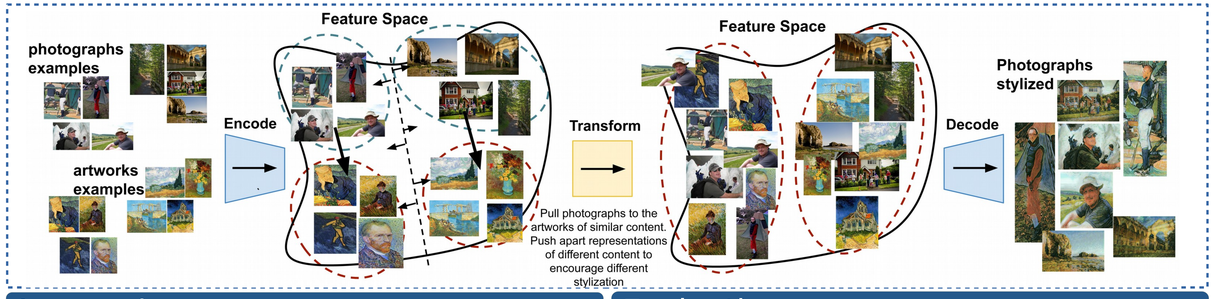
Style Transfer(ST):
- extract the very essence of a style from a single image or a collection of style examples
- translate extracted style on a query image, while preserving content of the later.

Motivation:
- artists alter objects and not just their appearance
- artistic style affects objects of different classes differently
- same content is transformed differently by different artists

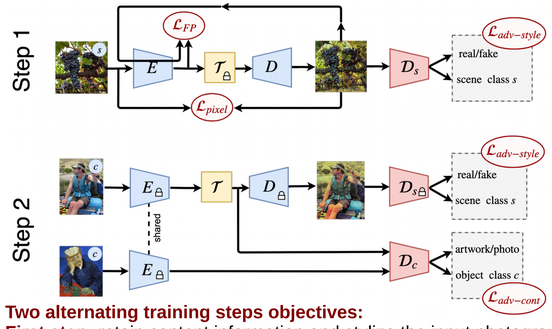
Problems and Contributions:
- existing ST approaches stylize images disregarding their semantics
  →we propose an **object-specific stylization**
- artists transform each object in its specific way
  →we introduce a **content transformation block** which alters objects specifically



- existing ST methods retain content as it appears in the input photograph.
  → We, however, **classify stylized images** during training, thus forcing to retain **what** is present in the image and not **how** it was depicted in the input image.
- there is no method to measure quality of content transformation for different styles
  → we propose **Relative Style Specific Content Distance (RSSCD)**



Approach



Two alternating training steps objectives:
First step: retain content information and stylize the input photograph
Second step: apply the content transformation block to the encoded photographs to:
→ match distributions of encoded artworks and photographs conditioned on class
→ learn to produce convincing stylizations

$$\mathcal{L}_{adv-style} := \mathop{\mathbb{E}}_{(y,s)\sim\textbf{Artwork}}[log(\mathcal{D}_s(y|s))] + \mathop{\mathbb{E}}_{(x,s)\sim\textbf{Photograph}}[log\,(1-\mathcal{D}_s\,(D(\mathcal{T}(E(x)))|s))]$$

$$\mathcal{L}_{pxl} := \mathop{\mathbb{E}}_{x\sim\textbf{Photograph}}[\|D(\mathcal{T}(E(x))) - x\|_2^2].$$

$$\mathcal{L}_{adv-cont} := \mathop{\mathbb{E}}_{(y,c)\sim\textbf{Artwork}}[log(\mathcal{D}_c(E(y)|c))] + \mathop{\mathbb{E}}_{(x,c)\sim\textbf{Photograph}}[log\,(1-\mathcal{D}_c\,(\mathcal{T}(E(x))|c))]$$

$$\mathcal{L}_{FP} := \mathop{\mathbb{E}}_{x\sim\textbf{Photograph}}[\|E(D(\mathcal{T}(E(x)))) - E(x)\|_2^2].$$

Step 1

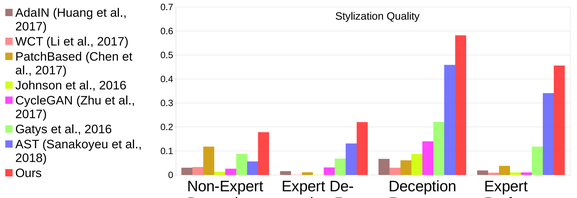
$$\min_{\theta_E,\theta_D}\max_{\theta_{\mathcal{D}_s}} \lambda_{\mathcal{L}_{pxl}}\mathcal{L}_{pxl} + \lambda_{\mathcal{L}_{FP}}\mathcal{L}_{FP} + \lambda_{\mathcal{L}_{adv-style}}\mathcal{L}_{adv-style}$$

Step 2

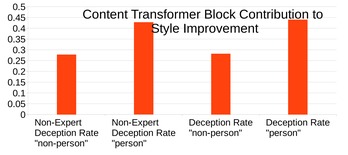
$$\min_{\theta_{\mathcal{T}}}\max_{\theta_{\mathcal{D}_c}} \lambda_{\mathcal{L}_{adv-cont}}\mathcal{L}_{adv-cont} + \lambda_{\mathcal{L}_{adv-style}}\mathcal{L}_{adv-style}$$

Evaluation

Stylization quality
Expert / non-expert deception rate = ability of the stylizations to fool the art historian and non-expert (higher is better).
Expert preference rate = indicates which stylization reflects target style more accurate.

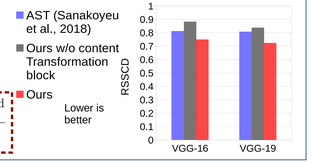


Content transformation block improves stylization for both seen and **unseen** object classes



Relative Style Specific Content Distance (RSSCD):
Measures content-specificity of the Stylization. We evaluate using VGG-16 and VGG-19 networks.

$$\text{RSSCD} := \frac{\text{average distance between a stylized image and the nearest real artwork of the same content}}{\text{average distance between real artworks of different content}}$$



Results