

引言

现有单路Transformer的**图像风格迁移**方法难以兼顾内容保真度和风格相似性, 本文通过**双路视觉Transformer架构**捕获图像长距离依赖关系以建模全局上下文信息, 并引入**Transformer风格参数提取器**对图像风格进行离散化表示, 最终在解码架构中将内容嵌入配准对齐至风格特征, 从而实现高质量任意图像风格迁移

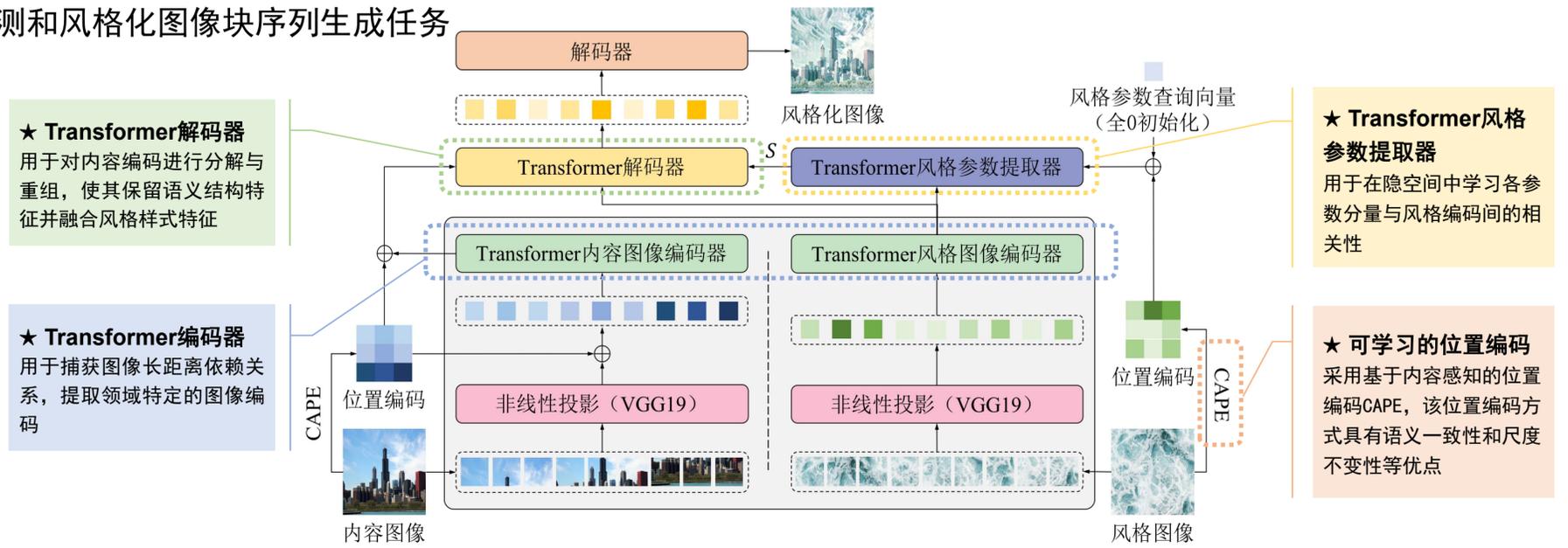
主要贡献

- 针对内容图像域和风格图像域进行**独立编码**, 捕获图像长距离依赖关系以建模全局上下文信息, 兼具较高的内容保真度与风格还原度
- 引入了一个**Transformer风格参数提取器**, 将图像风格离散化表示为一组关键特征, 提升了风格表征刻画的丰富性和准确性
- 在Transformer图像解码阶段**显式度量风格分布特征**, 基于预测得到的风格特征分布参数反归一化网络中间嵌入, 从而基于特征分布匹配实现特定风格渲染



方法

本文提出了一种基于双路视觉Transformer的图像风格迁移方法**Bi-Trans**, 将图像风格迁移划分为风格参数预测和风格化图像块序列生成任务



实验

数据集

本文使用**MSCOCO**数据集和**WikiArt**数据集提供训练所需的内容图像及风格图像

定量对比

感知损失 包括内容损失 \mathcal{L}_{con} 和基于均值方差或Gram矩阵的风格损失 $\mathcal{L}_{sty-\mu, \sigma}$ / $\mathcal{L}_{sty-Gram}$:

$$\mathcal{L}_{con} = \frac{1}{|\Psi_c|} \sum_{l \in \Psi_c} \frac{1}{H^l W^l} \sum_{i=1}^{H^l} \sum_{j=1}^{W^l} (\mathcal{F}_{cs,ij}^l - \mathcal{F}_{c,ij}^l)^2$$

$$\mathcal{L}_{sty-\mu, \sigma} = \frac{1}{|\Psi_s|} \sum_{l \in \Psi_s} (\mu_{cs}^l - \mu_s^l)^2 + (\sigma_{cs}^l - \sigma_s^l)^2$$

$$\mathcal{L}_{sty-Gram} = \frac{1}{|\Psi_s|} \sum_{l \in \Psi_s} (\mu_{cs}^l - \mu_s^l)^2 + (\sigma_{cs}^l - \sigma_s^l)^2$$

消融实验

消融(a): 移除Transformer风格参数提取器

消融(b): 将采用VGG19网络进行图像块划分和编码的方式替换为简单线性投影方式

消融(c): 本文将CNN解码器替换为由全连接层、ReLU激活与上采样层构成的多层感知机

消融(d): 移除两个恒等损失项

方法	$\mathcal{L}_c \downarrow$	$\mathcal{L}_s-\mu, \sigma \downarrow$	\mathcal{L}_s-Gram 矩阵 \downarrow
StyTr ²	0.79	1.35	9.86
S2WAT	0.93	2.64	13.47
STTR	0.79	3.86	27.58
Ghiasi	0.94	2.81	12.51
AdaIN	0.91	1.39	11.63
SANet	0.97	2.88	16.18
本文方法	0.69	1.12	7.31

定量对比

方法	$\mathcal{L}_c \downarrow$	$\mathcal{L}_s-\mu, \sigma \downarrow$	\mathcal{L}_s-Gram 矩阵 \downarrow
消融(a)	0.70	1.02	3.15
消融(b)	0.64	1.37	4.71
消融(c)	1.88	11.36	18.61
消融(d)	1.93	6.75	16.05
完整方法	0.69	0.94	3.02

消融实验

定性对比

