

文章编号: 1671-7449(2024)03-0298-10

基于SAM分割的交互式局部图像风格迁移方法研究

尚金灿, 张丽红*

(山西大学 物理工程学院, 山西 太原 030006)

摘要: 针对目前局部图像风格迁移目标区域选择困难、迁移灵活性不足、容易出现内容泄露、前景与背景边缘过渡不自然等问题, 提出一种基于任意分割模型(Segment Anything Model, SAM)的交互式局部图像风格迁移方法。首先利用SAM分割网络在用户输入提示的指导下对输入的内容图像进行交互式目标迁移区域提取, 对得到的有效对象掩码进行二值化处理, 以二值化掩码提取全局风格化图像的目标区域作为前景、内容图像作为背景图像进行泊松融合, 实现局部图像风格迁移。为了避免迁移过程中的内容泄露, 全局风格迁移网络采用生成对抗网络架构, 通过多级自适应注意力归一化模块进行风格特征转换, 利用联合损失函数对网络进行综合训练。实验结果表明, 设计的交互式局部图像风格迁移网络能够根据用户提示生成灵活可控的局部迁移结果, 可以对图像中的任意物体进行风格迁移, 迁移结果很好地保留了内容源图像中的内容结构, 避免了内容泄露, 且前景与背景边缘过渡更加自然。

关键词: 局部图像风格迁移; 任意分割模型; 全局风格迁移; 泊松融合; 自适应注意力归一化

中图分类号: TP391 **文献标识码:** A **doi:** 10.3969/j.issn.1671-7449.2024039

引用格式: 尚金灿, 张丽红. 基于SAM分割的交互式局部图像风格迁移方法研究[J]. 测试技术学报, 2024, 38(3): 298-307.

SHANG Jincan, ZHANG Lihong. Research on interactive local image style transfer based on SAM segmentation [J]. Journal of Test and Measurement Technology, 2024, 38(3): 298-307.

Research on Interactive Local Image Style Transfer Based on SAM Segmentation

SHANG Jincan, ZHANG Lihong*

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: Aiming at the problems of the current local image style transfer methods, such as the difficulty in selecting the target region, the lack of transfer flexibility, the easy occurrence of content leakage and unnatural transitions between foreground and background boundaries, an interactive local image style transfer method based on SAM segmentation is proposed in this paper. Firstly, SAM segmentation network is employed to extract the target transfer region of the input content image interactively under the guidance of the user input prompt, and the effective object mask is binarized. The binary mask is used to extract the target region of the global stylized image as the foreground and the content image as the background image for Poisson fusion to realize the local image style transfer. In order to avoid content leakage during the transfer process, the architecture of generative adversarial network is adopted in the global style transfer network. The multi-level adaptive attention normalization module is used for style feature conversion, and the joint loss function is used for comprehensive training of the network. The experimental results show that the interactive local image style

收稿日期: 2023-06-11

基金项目: 山西省研究生创新资助项目(2023SJ012); 山西省高等学校教学改革创新资助项目(J2021086)

作者简介: 尚金灿(1999-), 男, 硕士生, 主要从事深度学习和风格迁移研究。E-mail: 2558622850@qq.com.

* 通信作者: 张丽红(1968-), 女, 教授, 主要从事计算机视觉及机器学习研究。E-mail: lh Zhang@sxu.edu.cn.

transfer network designed in this paper can generate flexible and controllable local transfer results according to user prompts, and can carry out style transfer for any object in the image. The transfer results well preserve the content structure in the content source image, avoid content leakage and the boundaries of foreground and background transit more naturally.

Key words: local image style transfer; segment anything model (SAM); global style transfer; Poisson fusion; adaptive attention normalization

0 引言

近年来,深度学习技术的不断进步推动了图像风格迁移技术的快速发展。图像风格迁移技术旨在将内容图像渲染成具有艺术风格的图像。Gatys等^[1]于2016年将神经网络引入图像风格迁移,提出了基于统计分布的参数化慢速风格迁移算法,开创了基于深度学习的图像风格迁移技术的研究先河。Huang等^[2]于2017年提出自适应实例归一化方法,该方法将内容图像特征的均值和方差与风格图像特征的均值和方差进行对齐,首次实现了实时的任意图像风格迁移。Park等^[3]于2018年提出了一种风格注意网络,该网络利用注意力机制,根据内容图像的语义空间分布灵活地合成局部风格模式。Liu等^[4]于2021年提出了自适应注意力归一化模块(Adaptive Attention Normalization, AdaAttN),该模块利用浅层和深层网络的注意力得分以及特征分布信息,实现了高质量的实时任意图像风格迁移结果。Deng等^[5]于2022年提出了StyTr2网络架构,首次将Transformer架构^[6]应用于图像风格迁移任务中,避免了卷积神经网络中的内容泄露问题。以上方法只能对图像进行全局风格迁移,忽略了图像不同语义区域之

间的差异性。

对图像局部进行风格迁移是目前研究的重点问题,缪永伟等^[7]于2019年提出了基于卷积神经网络的图像局部风格迁移方法,王杨等^[8]于2022年提出了融合视觉注意机制的图像显著性区域风格迁移方法。现有局部图像风格迁移方法容易出现目标区域选择困难、迁移灵活性不足等问题,本文提出一种基于任意分割模型(Segment Anything Model, SAM)^[9]的交互式局部图像风格迁移方法,该方法利用SAM优越的分割性能,能够实现交互式局部图像风格迁移,在用户给定提示的情况下可以对图像中的任意物体进行风格迁移,实现灵活可控的局部图像风格迁移效果。该方法以泊松融合的方式将前景与背景进行融合,合成的图像边缘过渡更加自然。同时,本文设计的全局风格迁移网络很好地保留了内容源图像的内容结构,避免了迁移过程中的内容泄露。

1 系统框架

基于SAM分割的交互式局部图像风格迁移网络的整体架构如图1所示,网络模型由全局风格迁移网络和局部风格迁移网络组成。

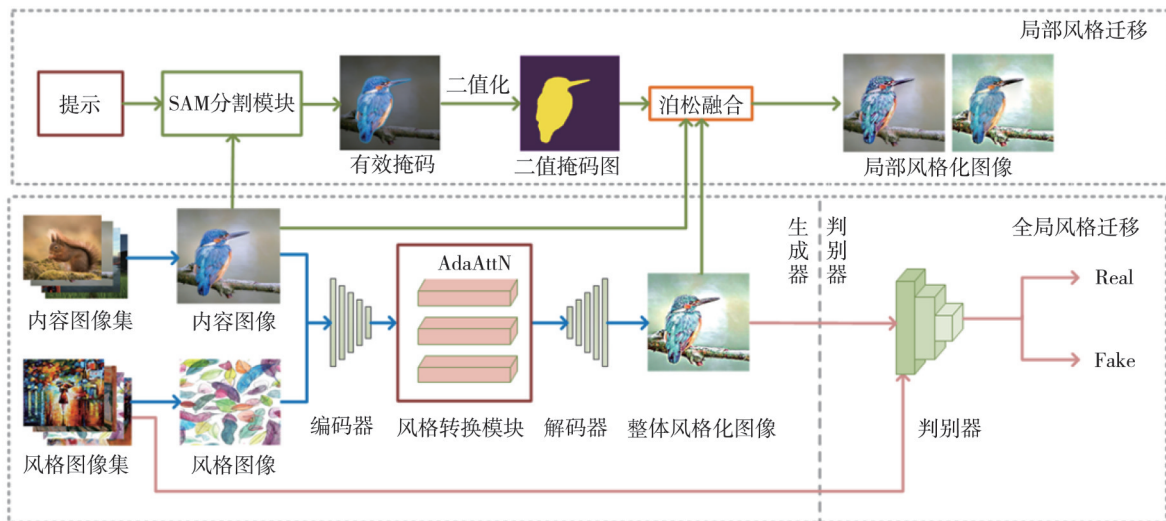


图1 基于SAM分割的交互式局部图像风格迁移网络的整体架构

Fig. 1 Overall architecture of interactive local image style transfer network based on SAM segmentation

全局风格迁移网络采用生成对抗网络架构，生成器由编码器、风格转换模块、解码器组成，该网络以内容图像和风格图像作为输入，编码器采用 ImageNet 数据集上预训练的 VGG19 网络^[10]对图像进行特征提取，风格转换模块采用三级自适应注意力归一化模块将编码器提取的内容特征与风格特征进行统计分布匹配，由解码器生成全局风格化图像。判别器采用多尺度判别器^[11]，通过内容损失函数、风格损失函数、身份保留损失函数、对比一致保留损失函数以及外部对抗损失函数构成的联合损失函数对网络进行综合训练。局部风格迁移网络根据用户给予的提示，利用 SAM 分割模块提取内容图像的有效掩码，对有效掩码进行二值化处理确定风格迁移的目标区域，然后以二值化掩码提取全局风格化图像的目标区域作为前景、内容图像作为背景图像进行泊松融合，实现灵活自然的交互式局部图像风格迁移。

2 基本原理

2.1 SAM 分割模块

SAM 主要包括交互式分割和自动分割两种方式，本文利用其交互式分割的特点在用户输入提示的指导下分割出风格迁移的目标区域。如图 2 所示，用户输入的提示可以是前景点、背景点、粗略的矩形

框、掩码或者文本，通过输入提示生成有效掩码，为实现局部图像风格迁移提供指导。当分割对象存在歧义时，SAM 可以输出多个有效的掩码。SAM 可以在预计算图像嵌入之后即时为任何提示生成分割掩码，从而允许与模型实时交互。

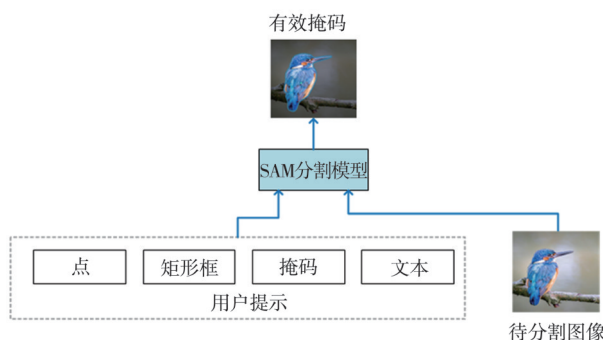


图 2 SAM 分割模型

Fig. 2 SAM segmentation model

SAM 分割模型的结构如图 3 所示，主要包括图像编码器、提示编码器以及掩码解码器 3 个部分。该模型首先将输入图像通过图像编码器转换为图像嵌入，使用掩码（密集提示）卷积后的映射与图像嵌入进行逐元素求和，使用轻量级的提示编码器将点、框或者文本（稀疏提示）转换为提示编码，将得到的图像嵌入与提示编码输送到轻量级掩码解码器中预测分割掩码，生成有效的多个掩码以及每个掩码对应的置信分数。最后使用 focal 损失函数和 dice 损失函数来监督掩码预测。

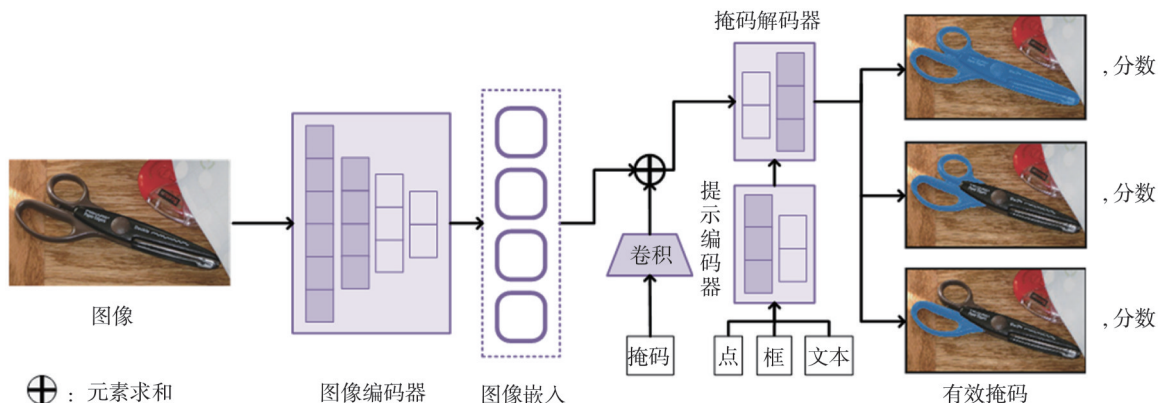


图 3 SAM 分割模型的结构

Fig. 3 The structure of SAM segmentation model

2.2 泊松融合模块

通过直接叠加的方法将目标区域迁移到内容图像源中会导致边缘过渡不自然。泊松融合是一种基于偏微分方程的图像融合算法。在目标图像中选择一个感兴趣区域作为融合区域，通过求解泊松方程

生成融合区域内的像素，最终得到一张新的融合图像。

如图 4 所示，以二值化掩码提取全局风格化图像的目标区域作为源图像 g ， v 为 g 的引导场，内容图像作为目标图像，融合得到结果图像 D ，待融合区域为 Ω ，计算源图像和目标图像的梯度信息 ∇g 和 ∇t ，使用梯度信息 ∇g 代替引导场 v ， $\partial \Omega$

为融合区的边缘部分, t 表示 Ω 区域的函数, t^* 表示目标图像 B 在区域 $D - \Omega$ 的函数。为保证融合区边缘无明显过渡痕迹, 保持 t^* 与 t 在边缘 $\partial\Omega$ 处相等, 为使融合区 Ω 尽可能保持图像 g 的纹理信息, 需要最小化 Ω 区域的梯度与 v 之间的差距, 最终需要求解的方程为

$$\min_t \iint_{\Omega} \|\nabla t - v\|^2 = \min_t \iint_{\Omega} \|\nabla t - \nabla g\|^2, \quad (1)$$

with $t^*|_{\partial\Omega} = t|_{\partial\Omega}$.

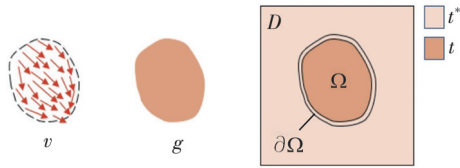


图 4 泊松融合原理

Fig. 4 Poisson fusion principle

表 1 为泊松融合的具体步骤。

表 1 泊松融合步骤

Tab. 1 Poisson fusion steps

泊松融合步骤
Step1: 获得源图像与目标图像, 确定目标迁移区域
Step2: 分别计算源图像与目标图像梯度场
Step3: 求解泊松方程, 得到修正后的源图像
Step4: 将修正后的源图像与目标图像进行融合, 生成最终融合图像

2.3 生成器结构

生成器的输入为内容图像 I_c 与风格图像 I_s , 通过编码器不同层提取出内容图像与风格图像的多

尺度特征。将当前层特征与之前所有层经过下采样后的特征沿着通道维度进行级联得到级联特征, 其表达式为

$$F_*^{1:x} = B_x(F_*^1) \oplus B_x(F_*^2) \oplus \dots \oplus F_*^x, \quad (2)$$

式中: $*$ 可以为 c 或 s, F_c^x 和 F_s^x 分别为编码器 $Relu_x_1$ 层提取的内容特征和风格特征; B_x 为双线性插值层, 用于将特征图下采样至与 F_*^x 具有相同大小; \oplus 表示沿着通道维度进行级联。

如图 5(a) 所示, 风格转换模块采用多层次的策略以充分利用浅层与深层特征, 分别在编码器 $Relu_3_1$ 、 $Relu_4_1$ 和 $Relu_5_1$ 层上集成三级 AdaAttN 模块^[4], 得到多级嵌入特征 F_{cs}^3 、 F_{cs}^4 、 F_{cs}^5 。

AdaAttN 模块综合利用内容和风格图像的浅层与深层特征计算注意力得分, 将注意力加权的风格特征的二阶统计量与均值方差归一化后的内容特征进行匹配, 实现自适应的图像风格迁移。

如图 5(b) 所示, 将级联内容特征 $F_c^{1:x}$ 、级联风格特征 $F_s^{1:x}$ 分别进行通道维度的均值方差归一化计算, 将结果输入到可学习的 1×1 卷积层, 得到注意力机制中的 $Query(Q)$ 和 $Key(K)$ 。将风格特征 F_s^x 输入到可学习的 1×1 卷积层, 得到注意力机制中的 $Value(V)$ 。 Q 、 K 、 V 的表达式为

$$Q = f(Norm(F_c^{1:x})), K = g(Norm(F_s^{1:x})), \quad (3)$$

$$V = h(F_s^x),$$

式中: f 、 g 、 h 为可学习的 1×1 卷积层; $Norm$ 表示通道维度的均值方差归一化。

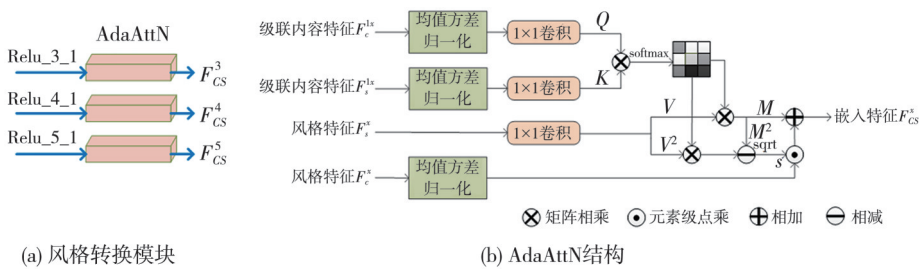


图 5 风格转换模型

Fig. 5 A model for style transfer

将 Q 与 K 进行矩阵相乘, 再计算 $Softmax$ 函数, 得到内容特征与风格特征之间的注意力得分矩阵 A , A 用来衡量内容特征与风格特征之间的相似度, 其表达式为

$$A = Softmax(Q^T \otimes K), \quad (4)$$

式中: \otimes 表示矩阵乘法; T 为矩阵转置。

利用 A 对风格特征 F_s^x 处理得到的 V 进行注意力加权, 注意力加权均值 M 的表达式为

$$M = V \otimes A^T. \quad (5)$$

变量的方差等于其平方的期望减去期望的平方, 所以, 注意力加权标准方差 S 的表达式为

$$S = \sqrt{(V \cdot V) \otimes A^T - M \cdot M}. \quad (6)$$

对均值方差归一化处理后内容特征图的每个位置和通道, 利用 S 做缩放, M 做偏移处理, 得到变换后的嵌入特征 F_{cs}^x , 该过程表达式为

$$F_{cs}^x = S \cdot Norm(F_c^x) + M. \quad (7)$$

将风格转换模块得到的嵌入特征 F_{cs}^3 、 F_{cs}^4 、 F_{cs}^5 作为解码器的输入, 通过解码器合成风格化图像 I_{cs} , 该过程表达式为

$$I_{cs} = Dec(F_{cs}^3, F_{cs}^4, F_{cs}^5), \quad (8)$$

式中: Dec 表示解码器。

3 损失函数设计

3.1 内容损失函数

输入的内容图像 I_c 通过编码器 $Relu_4_1$ 层和 $Relu_5_1$ 层提取出内容特征 F_c^4 和 F_c^5 。生成的风格化图像 I_{cs} 通过编码器 $Relu_4_1$ 层和 $Relu_5_1$ 层提取出风格化特征 F_{cs}^4 和 F_{cs}^5 。计算 F_{cs}^4 和 F_c^4 之间的欧式距离, 与 F_{cs}^5 和 F_c^5 之间的欧式距离之和即为内容损失, 其表达式为

$$L_c = \|F_{cs}^4 - F_c^4\|_2 + \|F_{cs}^5 - F_c^5\|_2. \quad (9)$$

3.2 风格损失函数

输入的风格图像 I_s 与生成的风格化图像 I_{cs} 分别通过编码器提取出风格特征与风格化特征。计算风格特征均值和风格化特征均值之间的欧式距离, 与风格特征方差和风格化特征方差之间的欧式距离之和即为风格损失, 其表达式为

$$L_s = \sum_{i=1}^L \left[\|\mu(\phi_i(I_{cs})) - \mu(\phi_i(I_s))\|_2 + \|\sigma(\phi_i(I_{cs})) - \sigma(\phi_i(I_s))\|_2 \right], \quad (10)$$

式中: ϕ 为编码器输出的特征映射; i 为当前层; μ 为均值; σ 为方差; L 为计算风格损失所用到的总层数, L 取值为 5; 在本文实验中, 使用相同权重的编码器 $Relu_1_1$ 、 $Relu_2_1$ 、 $Relu_3_1$ 、 $Relu_4_1$ 和 $Relu_5_1$ 层来计算风格损失。

3.3 身份保留损失函数

身份保留损失函数可以很好地保持内容图像的结构以及风格图像的风格特征。如图 6 所示, 采用一组相同的内容图像 I_c 或风格图像 I_s 作为输入, 得到相应的输出图像 I_{cc} 或 I_{ss} 。通过计算像素空间内 I_{cc} 和 I_c 之间的欧式距离, 与 I_{ss} 和 I_s 之间的欧式距离之和即为 $L_{identity1}$, 计算特征空间内 I_{cc} 和 I_c 之间的欧式距离, 与 I_{ss} 和 I_s 之间的欧式距离之和

即为 $L_{identity2}$, $L_{identity1}$ 与 $L_{identity2}$ 进行加权求和即为身份保留损失函数 $L_{identity}$, 其表达式为

$$L_{identity} = \lambda_1 L_{identity1} + \lambda_2 L_{identity2},$$

$$L_{identity1} = \|(I_{cc} - I_c)\|_2 + \|(I_{ss} - I_s)\|_2,$$

$$L_{identity2} = \sum_{i=1}^L (\|\phi_i(I_{cc}) - \phi_i(I_c)\|_2 + \|\phi_i(I_{ss}) - \phi_i(I_s)\|_2), \quad (11)$$

式中: λ_1 、 λ_2 为损失权重; I_{cc} 为输入两张相同内容图像生成的结果图像; I_{ss} 为输入两张相同风格图像生成的结果图像; L 为计算 $L_{identity2}$ 所用到的总层数, L 取值为 5; 在本文实验中, 使用编码器 $Relu_1_1$ 、 $Relu_2_1$ 、 $Relu_3_1$ 、 $Relu_4_1$ 和 $Relu_5_1$ 层来计算 $L_{identity2}$ 。

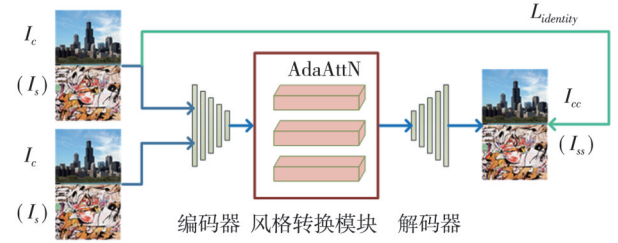


图 6 身份保留损失函数的计算

Fig. 6 Calculation of identity retention loss function

3.4 对比一致保留损失函数

对比一致保留损失函数^[12]采用对比学习的形式, 可以很好地保留内容图像中局部之间的关联性, 减少风格化图像中的局部随机扰动, 提升图像生成质量。对比一致保留损失函数的计算过程如图 7 所示。

具体地, 将合成的风格化图像 I_{cs} 与输入的内容图像 I_c 通过编码器提取不同层的特征图, 从风格化图像 I_{cs} 某一层特征图 F_g 中随机采样非边缘的 N 个向量, 记为中心向量。在中心向量周围采样 8 个邻近向量。将 8 个邻近向量与中心向量分别作差得到 $8 \times N$ 个差向量 d_g 。在内容图像 I_c 对应特征图 F_c 中的相同位置选取 N 个中心向量以及邻近向量, 作差得到 $8 \times N$ 个差向量 d_c 。构建一个两层的多层感知机, 将得到的差向量输入多层感知机投影到隐空间中, 经过标准化后计算 InfoNCE 损失, 每一层对比一致保留损失函数的计算表达式为

$$L_{ccp} = \sum_{m=1}^{8 \times N} -\log \left[\frac{\exp(d_g^m \cdot d_c^m / \tau)}{\exp(d_g^m \cdot d_c^m / \tau) + \sum_{n=1, n \neq m}^{8 \times N} \exp(d_g^m \cdot d_c^n / \tau)} \right], \quad (12)$$

式中: τ 默认取值为 0.07; d_g 和 d_c 分别为风格化图像和内容图像得到的差向量。

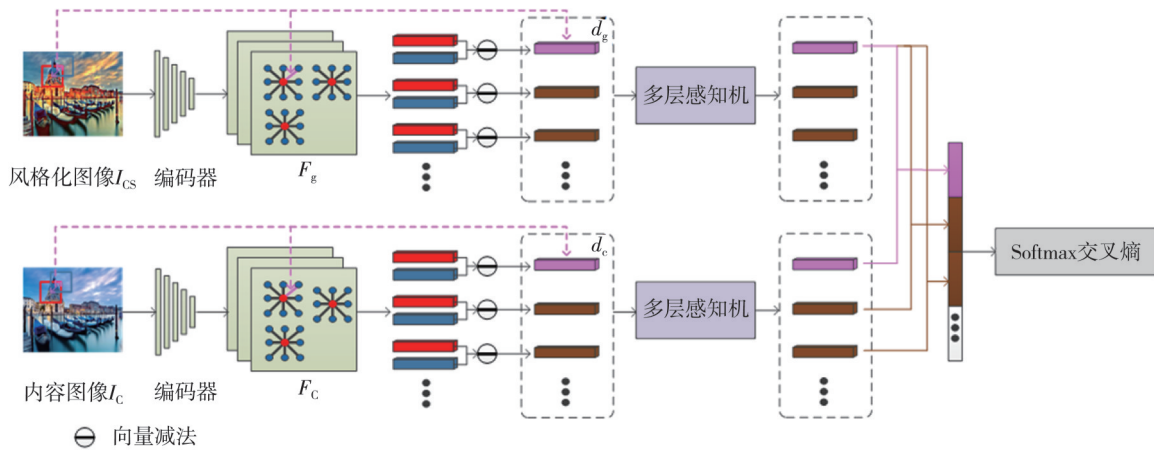


图 7 对比一致保留损失函数的计算

Fig. 7 Calculation of contrastive coherence preserving loss

对不同特征层重复以上所有步骤, 加权求和每一层的损失结果得到总的对比一致保留损失函数值。在本文实验中, 每层采样中心向量个数 N 设置为 64, 在编码器 $Relu_3_1$ 、 $Relu_4_1$ 和 $Relu_5_1$ 层上计算对比一致保留损失函数值。

3.5 外部对抗损失函数

外部对抗损失函数可以学习人类感知的风格信息。将生成器 G 生成的风格化图像 I_{cs} 与从风格图像数据集中随机采样的风格图像 I_s 分别作为假数据和真数据输入到判别器 D 中。在训练阶段, 生成器试图通过生成真实的风格图像来欺骗判别器, 而判别器试图区分生成器生成假的风格图像与真实风格图像。对生成器与判别器进行联合训练, 使得生成器能够学习到人类感知的风格信息, 从而产生与参考风格相似的图像。外部对抗损失函数的表达式为

$$L_{adv} = E[\log(D(I_s))] + E[\log(1 - D(I_{cs}))], \quad (13)$$

式中: E 为函数期望值。

3.6 联合损失函数

联合损失函数是内容损失函数、风格损失函数、身份保留损失函数、对比一致保留损失函数以及外部对抗损失函数的加权求和, 其表达式为 $L = \lambda_c L_c + \lambda_s L_s + L_{identity} + \lambda_{ccp} L_{ccp} + \lambda_{adv} L_{adv}$, (14) 式中: λ_c 、 λ_s 、 λ_{ccp} 、 λ_{adv} 为各损失函数对应的权重。

4 实验结果及分析

4.1 实验设置

在本文实验中, 使用 MS-COCO^[13] 作为内容图

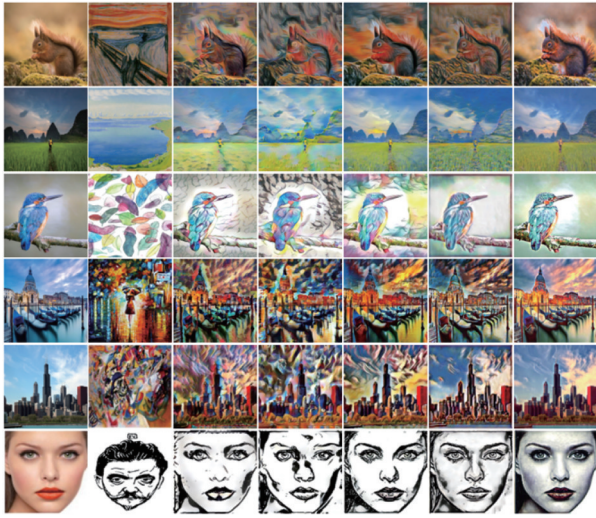
像数据集、WikiArt^[14] 作为风格图像数据集进行训练, 每个数据集取 80 000 张图像。使用 Adam 作为模型优化器, 学习率为 0.000 1, 批量大小为 2 组内容风格图像对。损失函数的权重设置为 $\lambda_c = 1$, $\lambda_s = 10$, $\lambda_1 = 50$, $\lambda_2 = 1$, $\lambda_{ccp} = 5$, $\lambda_{adv} = 5$ 。在训练阶段, 先将每张训练图像大小缩放为 512×512 , 保持图像纵横比, 再随机将图像裁剪为 256×256 。在测试阶段, 网络可支持任意尺寸的图像作为输入。

4.2 全局风格迁移结果评价

4.2.1 全局风格迁移结果主观评价

本文方法与其他方法的全局风格迁移结果对比如图 8 所示, 从左至右依次为内容图像、风格图像、AdaIN 方法、SANet 方法、StyTr2 方法、AdaAttN 方法以及本文方法生成的风格化结果图像。AdaIN 方法直接调整内容特征全局的二阶统计信息, 在风格转换过程中存在着严重的内容细节丢失(与该方法对应所在列的第 1, 4, 6 行)以及混乱的纹理模式(第 4, 5 行)。SANet 方法采用注意力机制, 在深层空间中转移风格特征到内容特征上。该方法会导致内容结构损坏(第 1, 2 行)以及混乱的纹理模式(第 1, 3, 4, 5, 6 行)。StyTr2 方法通过自注意机制学习输入的全局信息, 从而使每一层都能获得对全局信息的理解, 且 Transformer 架构不包含局部性和空间不变性引起的归纳偏差, 相比于前两种方法, StyTr2 方法能够更好地保持内容结构, 但是仍然存在纹理混乱(第 3, 5 行)的问题。AdaAttN 方法综合利用网络的浅层与深层特征, 同时结合全局风格损失函数与局部特征损失函数对网络训练, 在减少混乱的纹理方面是有效的, 但存在部分粗糙的纹理模式(第 6 行)。

与其他方法相比,本文方法生成的风格化图像更好地保留了与内容源图像一致的语义结构,其不规则纹理显著减少,视觉质量得到显著提高。



内容图像 风格图像 AdaIN SANet StyTr2 AdaAttN 本文方法

图8 全局风格化结果对比

Fig. 8 Comparison of global stylization results

4.2.2 全局风格迁移结果客观评价

为了更加客观地评价全局风格迁移结果,采用结构相似性(Structural Similarity, SSIM)、均方误差(Mean Squared Error, MSE)以及峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)3种指标对生成的图像进行评估。

SSIM指标从图像的亮度、对比度及结构3个方面进行评估,利用均值作为亮度的估计,标准差作为对比度的估计,协方差作为结构相似程度的度量。SSIM数值越接近于1表明两幅图像的相似程度越高,其计算表达式为

$$SSIM(c, g) = \frac{(2\mu_c\mu_g + c_1)(2\sigma_{cg} + c_2)}{(\mu_c^2 + \mu_g^2 + c_1)(\sigma_c^2 + \sigma_g^2 + c_2)}, \quad (15)$$

式中: μ_c 、 μ_g 分别为图像 c 、图像 g 的均值; σ_c^2 、 σ_g^2 分别为图像 c 、图像 g 的方差; σ_{cg} 为图像 c 和图像 g 的协方差; c_1 、 c_2 为维持稳定的常数,避免除数为0。

MSE指标是参数预测值与实际值之差平方的期望值,在风格迁移任务中反映风格化图像与输入图像的相似程度。MSE数值越小表明两幅图像的相似程度越高,其计算表达式为

$$MSE(c, g) = \frac{1}{xy} \sum_{i=0}^{x-1} \sum_{j=0}^{y-1} [c(i, j) - g(i, j)]^2, \quad (16)$$

式中: x 为图像的高度; y 为图像的宽度。

PSNR指标用于衡量图像的信噪比,即图像中有效信号与噪声的比值,在风格迁移任务中用

于度量风格化图像相对输入图像的质量。PSNR数值越大表明合成风格化图像的失真越小、质量越高,其计算表达式为

$$PSNR(c, g) = 10 \lg \left(\frac{255^2}{MSE(c, g)} \right). \quad (17)$$

随机选取100个内容风格图像对进行风格化测试,并取其平均值。表2为其他风格迁移方法与本文全局风格迁移方法的指标结果对比。根据表2中的数据可知,本文方法相比其他方法在3个指标上表现出了较好的结果,进一步证明了全局风格迁移网络的有效性。

表2 本文方法与其他方法的评价指标对比

Tab. 2 Comparison of evaluation indexes between the method in this paper and other methods

方法	SSIM	MSE	PSNR
AdaIN ^[2]	0.39	5 418.18	9.50
SANet ^[3]	0.32	7 111.45	9.82
StyTr2 ^[5]	0.49	4 982.12	11.36
AdaAttN ^[4]	0.44	4 402.56	11.83
本文方法	0.63	3 481.37	12.82

4.3 局部风格迁移结果评价

局部风格迁移结果如图9所示,前3行对用户选中的目标区域根据提供的参考风格图像进行风格迁移,后2行对用户选择的不同区域采用不同的参考风格图像进行风格迁移。



内容图像 风格图像 用户提示 二值掩码 局部迁移结果

图9 局部风格迁移结果

Fig. 9 Local style transfer results

观察图 9 可知,本文方法可以根据用户提示精准地捕捉到目标风格迁移区域,对图像中的任意物体进行局部风格迁移,且局部迁移结果与内容源图像的结构保持一致,边缘过渡自然。

4.4 超参数实验

调整损失函数的权值,对损失函数的超参数进行选择。通过大量实验,设置 $\lambda_c = 1, \lambda_s = 10, \lambda_1 = 50, \lambda_2 = 1, \lambda_{adv} = 5$ 。通过调整对比一致保留损失函数的权值 λ_{cp} 为1~10,观察不同参数对风格化图像的影响。

从图 10 中可以看出,随着对比一致保留损失函数权值的增大,风格化图像的清晰度越来越高,其视觉质量越来越接近内容图像源,但呈现出的局部风格特征越来越不明显,为了平衡视觉质量和风格化效果,选择 $\lambda_{cp} = 5$ 。

设置 $\lambda_c = 1, \lambda_s = 10, \lambda_{cp} = 5, \lambda_{adv} = 5, \lambda_2 = 1$ 。通过调整 λ_1 从1到200,观察不同参数对风格化图像的影响。

从图 11 中可以看出,随着 λ_1 取值的变化,风格化图像会出现不同程度的失真,为了使得失真程度最小,选择 $\lambda_1 = 50$ 。

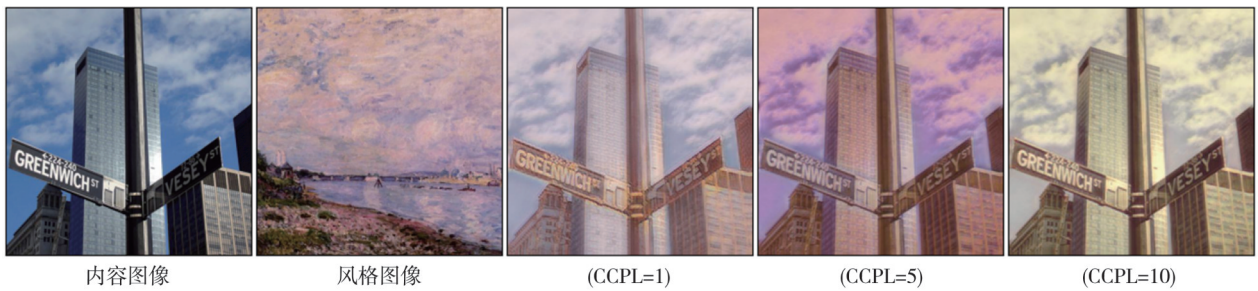


图 10 对比一致保留损失函数权值对比

Fig. 10 Comparison of contrastive coherence preserving loss function weights



图 11 λ_1 的不同权值对比

Fig. 11 Comparison of different weights of λ_1

4.5 消融实验

为了验证全局风格迁移网络所使用损失函数

的有效性,对损失函数进行消融实验,结果如图 12 所示。



图 12 损失函数的消融实验结果

Fig. 12 The ablation experimental results of the loss function

观察图 12 可知, 没有对比一致保留损失和身份保留损失时, 生成的风格化图像存在内容扭曲、局部失真。没有外部对抗损失时, 生成的风格化图像保持了与内容源图像一致的内容结构, 但是风格化特征不明显。完整模型很好地保留了内容源图像的语义结构, 且生成图像的色彩更加鲜艳, 合成的纹理更加干净, 产生的视觉效果质量更佳。

为了验证泊松融合模块的有效性, 将未使用泊松融合模块直接点乘相加与使用泊松融合的方法对比, 得到的实验结果如图 13 所示, 可以发现使用泊松融合后, 生成的图像前景与背景边缘过渡更加平滑, 与内容图像源能够无缝拼接, 图像整体效果更加自然。

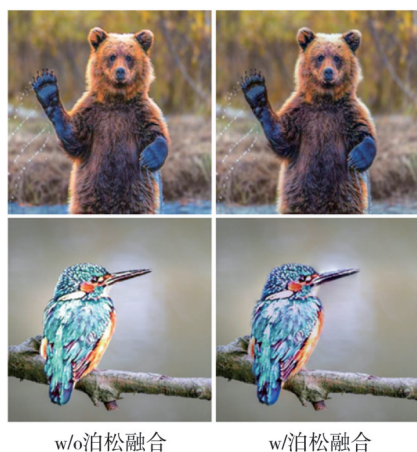


图 13 泊松融合的消融实验结果

Fig. 13 The ablation experimental results of Poisson fusion

4.6 损失函数曲线

如图 14 所示, 本文网络在 MS-COCO 内容图像数据集和 WikiArt 风格图像数据集上迭代了 50 000 次。随着迭代次数增加, 联合损失函数值趋向于稳定, 网络收敛。

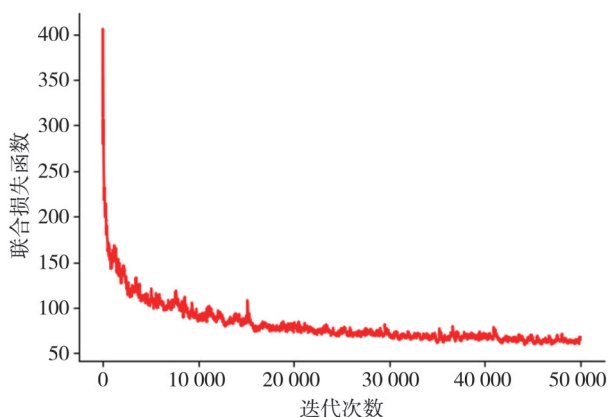


图 14 损失函数曲线

Fig. 14 Loss function curve

5 结论

本文提出了一种基于 SAM 分割的交互式局部图像风格迁移方法。利用 SAM 分割网络得到的二值掩码提取的全局风格化图像作为前景、内容图像作为背景图像进行泊松融合, 该方法能够根据用户提示选择合适的目标风格迁移区域, 可以对图像中的任意物体进行风格迁移, 解决了局部图像风格迁移目标区域选择困难、迁移灵活性不足以及前景与背景边缘过渡不自然的问题。对于全局风格迁移网络, 整体采用生成对抗网络的架构, 通过多级自适应注意力归一化模块进行风格特征转换, 利用设计的联合损失函数对网络进行综合训练, 其结果很好地保留了内容源图像的内容结构, 避免了迁移过程中的内容泄露。

参考文献:

- [1] GATYS L, ECKER A, BETHGE M. A neural algorithm of artistic style[J]. Journal of Vision, 2016, 16(12): 326.
 - [2] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization[C]//2017 Proceedings of the IEEE International Conference on Computer Vision, 2017: 1501-1519.
 - [3] PARK D Y, LEE K H. Arbitrary style transfer with style-attentional networks [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5873-5881.
 - [4] LIU S, LIN T, HE D, et al. Adaattn: revisit attention mechanism in arbitrary neural style transfer[C]//2021 IEEE/CVF International Conference on Computer Vision, 2021: 6629-6638.
 - [5] DENG Y, TANG F, DONG W, et al. StyTr2: image style transfer with transformers [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11316-11326.
 - [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.
 - [7] 缪永伟, 李高怡, 鲍陈, 等. 基于卷积神经网络的图像局部风格迁移[J]. 计算机科学, 2019, 46(9): 259-264.
- MIAO Yongwei, LI Gaoyi, BAO Chen, et al. Image localized style transfer based on convolutional neural network[J]. Computer Science, 2019, 46(9): 259-

264. (in Chinese)
- [8] 王杨, 郁振鑫, 卢嘉. 融合视觉注意机制的图像显著性区域风格迁移方法[J]. 计算机工程与科学, 2022, 44(1): 118-123.
- WANG Yang, YU Zhenxin, LU Jia. An image saliency area style transfer method combining visual attention mechanism[J]. Computer Engineering & Science, 2022, 44(1): 118-123. (in Chinese)
- [9] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything [C]//2023 IEEE/CVF International Conference on Computer Vision, 2023: 3992-4003.
- [10] SIMONYAN, KAREN, ANDREW ZISSERMAN. Very deep convolutional networks for large-scale image recognition [C]//3rd International Conference on Learning Representations, 2015: 1-14.
- [11] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8798-8807.
- [12] WU Z, ZHU Z, DU J, et al. CCPL: contrastive coherence preserving loss for versatile style transfer [C]//Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022: 189-206.
- [13] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//European Conference on Computer Vision, 2014: 740-755.
- [14] PHILLIPS F, MACKINTOSH B. Wiki art gallery, inc. : a case for critical thinking[J]. Issues in Accounting Education, 2011, 26(3): 593-608.