计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(3):118−125 [doi: 10.15888/j.cnki.csa.009449]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

保留细节特征的图像任意风格迁移①

蒋亨畅, 张笃振

(江苏师范大学 计算机科学与技术学院, 徐州 221116)
通信作者: 张笃振, E-mail: zhduzhen@jsnu.edu.cn

摘　要: 一些主流的图像任意风格迁移模型在保持内容图像的显著性信息和细节特征方面依然有局限性, 生成的图像往往具有内容模糊、细节失真等问题. 针对以上问题, 本文提出一种可以有效保留内容图像细节特征的图像任意风格迁移模型. 模型包括灵活地融合从编码器提取到的浅层至深层的多层级图像特征; 提出一种新的特征融合模块, 该模块可以高质量地融合内容特征和风格特征. 此外, 还提出一个新的损失函数, 该损失函数可以很好地保持内容和风格全局结构, 消除伪影. 实验结果表明, 本文提出的图像任意风格迁移模型可以很好地平衡风格和内容, 保留内容图像完整的语义信息和细节特征, 生成视觉效果更好的风格化图像.

关键词: 图像任意风格迁移; 保留细节特征; 多层级图像特征; 特征融合; 损失函数; 注意力机制

引用格式: 蒋亨畅,张笃振.保留细节特征的图像任意风格迁移.计算机系统应用,2024,33(3):118–125. http://www.c-s-a.org.cn/1003-3254/9449.html

Image Arbitrary Style Transfer with Preserving Detailed Features

JIANG Heng-Chang, ZHANG Du-Zhen

(School of Computer Science and Technology, Jiangsu Normal University, Xuzhou 221116, China)

Abstract: Some mainstream image arbitrary style transfer models still have limitations in maintaining the saliency information and detailed features of content images, resulting in problems such as content blurring and loss of details in the generated images. To solve the problems, this study proposes an arbitrary style transfer model that can effectively preserve the detailed features of content images. The model includes flexible fusing shallow to deep multi-layer image features extracted from the encoder. A new feature fusion is proposed, which allows for a high-quality fusion of content features and style features. In addition, a new loss function is proposed, which can well preserve the global structure of content and style and eliminate artifacts. The experimental results show that the proposed image arbitrary style transfer model can effectively balance style and content, preserve the complete semantic information and detailed features of the content image, and generate stylized images with better visual effects.

Key words: image arbitrary style transfer; preserving detailed features; multi-layer image features; feature fusion; loss function; attention mechanism

图像风格迁移是指给定一幅图像作为内容图像和另一幅图像作为风格图像, 通过某种算法或者模型将内容图像的内容语义特征和风格图像的风格特征进行融合并生成一幅新的图像, 新的图像既具有内容图像的内容语义信息, 又可以使其风格和风格图像的保持一致. 图像风格迁移已经应用于多个领域, 包括基于风格迁移的趣味手机应用 Prisma、Ostagram 软件等; 美颜软件中面部特征变换编辑; 动漫、游戏、影视制作

① 基金项目: 江苏省高等学校自然科学研究面上项目 (19KJB520032); 江苏师范大学博士学位教师科研支持项目 (20XSRS018); 江苏省研究生科研与实践创新计划 (KYCX22_2859)
收稿时间: 2023-08-25; 修改时间: 2023-10-09, 2023-10-25; 采用时间: 2023-11-15; csa 在线出版时间: 2024-01-18
CNKI 网络首发时间: 2024-01-19

中的图像上色等.

传统的风格迁移方法需要先对图像建立合适的数学模型, 再手工提取图像的底层特征, 然后进行纹理合成并绘制图像, 但是该过程费时且生成的图像质量也不理想. 近年来随着人工智能的发展, 图像风格迁移技术也不断发展, 特别是随着卷积神经网络 (convolutional neural network, CNN)[1]的成功应用, 它已成为进行图像风格迁移任务的主流方法. Gatys 等人[2]首先将 CNN 应用于图像风格迁移中, 发现可以通过使用预先训练好的 VGG (visual geometry group)[3]网络从图像中提取出深度特征, 然后使用提取到的特征构造 Gram 矩阵来表示风格图像的风格特征, 并提出一种图像重建算法实现风格迁移. 然而该算法效率较低, 十分耗时. 为了解决这个问题, Johnson 等人[4]和 Ulyanov 等人[5]均提出训练一个前馈卷积神经网络实现快速风格迁移, 将计算负担转移到模型训练的阶段. 这种方法虽然解决了生成图像效率低下的问题, 但是只能迁移一种风格图像, 且需要特定的风格图像数据集, 当需要迁移其他风格时, 就要重新收集该风格的图像来训练模型, 而收集这些风格图像也存在着困难. Li 等人[6]提出一种简单而有效的方法实现了任意风格迁移, 方法的关键是嵌入一对特征变换, 即白化和着色. 白化操作是去掉图像的风格特征, 而着色是重新给图像迁移风格, 提出的方法不需要训练参数, 因此具有通用性.

深入研究当前的一些任意风格迁移模型时, 可以发现: 模型通常是基于更高抽象层次的深度 CNN 特征, 过度依赖于深度图像特征而忽略底层细节; 这些方法不能协调内容图像和风格图像的视觉分布, 以及局部和全局的内容样式, 导致全局和局部内容结构不平衡; 虽然实现了风格迁移的多样性, 但是生成的图像效果却不尽如人意, 无法保持完整的内容图像的结构和细节特征, 生成的风格化图像的视觉效果尚有较大的改善余地. 为此, 本文提出一种简单而有效的方法来解决这些问题, 本文主要工作如下.

(1) 提出一个可以有效保留内容图像细节特征的任意风格迁移模型. 为了充分地保留内容图像的语义信息和细节特征以及更好地迁移风格特征, 模型融合从编码器提取到的浅层至深层的多层级图像特征.

(2) 提出一种新的特征融合模块, 该模块利用内容特征和风格特征的语义特征来引导特征进行融合. 模块可以很好地融合风格和内容特征的全局特征和局部特征, 有效地、灵活地进行内容特征和风格特征的融合.

(3) 提出一个新的损失函数进行模型的优化和训练, 该损失函数可以保持内容和风格结构的一致性, 消除伪影.

(4) 实验表明, 本文提出的任意风格迁移模型能够保持内容图像的显著性结构和局部细节特征, 并灵活地整合内容特征和风格特征, 生成的图像具有良好的视觉效果.

1 相关工作

1.1 任意风格迁移

许多现有的图像任意风格迁移研究已经取得了较好的效果. Huang 等人[7]发现图像特征图的平均值和方差可以用来表示该图像的风格, 因此, 提出一种新颖的方法, 即 AdaIN (adaptive instance normalization). 方法将内容特征的平均值和方差与风格特征的进行对齐操作实现图像的风格迁移. Park 等人[8]将非局部自注意力机制应用到风格迁移模型中, 提出了一种基于非局部自注意力的图像任意风格迁移模型即 SANet (style-attention network), 可以灵活地进行风格迁移. Liu 等人[9]提出一种自适应注意力规范化网络, 从内容和风格图像的浅层和深层特征中学习空间注意力, 在每个特征的基础上自适应地执行注意力规范化. Huo 等人[10]假设相同语义区域的特征形成一个流形, 遵循流形分布, 基于这一假设提出了一个基于流形对齐的风格迁移模型. An 等人[11]提出可逆神经流和无偏图像特征迁移模块来防止通用风格迁移过程中的内容泄漏. Wu 等人[12]认为全局不一致性造成局部不一致性, 并设计了一个适用于局部补丁的通用对比一致性保持损失函数. Wang 等人[13]提出了一种基于自适应通道网络的任意风格迁移模型, 可以灵活地选择特定的通道进行风格迁移.

然而 AdaIN[7]虽然可以迁移纹理和颜色, 但是无法保留局部特征, 容易造成局部失真. SANet[8]只使用编码深层特征进行风格迁移, 因此不可避免地丢失一些重要信息和细节特征, 容易导致生成的图形扭曲变形. AdaAttN[9]生成的图像局部容易出现不符合或者重复的风格, 且细节特征无法保留. ArtFlow[11]解决了内容泄露的问题, 但是无法平衡全局和局部风格. CCPL[12]提出的模型可以较好地迁移风格, 但是只使用编码器提取到的深层特征进行风格迁移, 因此不可避免地丢失一些细节特征. 因此, 本文提出一种任意风格迁移模型, 模型灵活地融合编码器提取到的浅层和深层图像特征, 并提出一种新的特征融合模块, 该模块可以根据

图像的语义空间分布很好地融合风格和内容特征的全局特征和局部特征. 同时, 提出一个新的损失函数进行模型的优化和训练. 该模型可以有效地实现图像任意风格迁移, 既能保持内容图像的细节特征, 又能平衡全局和局部风格, 生成的图像具有良好的视觉效果.

1.2 注意力机制

注意力机制在许多计算机视觉任务中都取得较好的效果, 包括图像分类、目标检测、语义分割、视频理解、图像生成、三维视觉、多模态任务和自监督学习等[14]. 自注意力机制首先由 Vaswani 等人[15]提出, 并迅速在自然语言处理领域取得了巨大的进展. Wang 等人[16]率先在计算机视觉中引入自注意力, 提出了一种新颖的非局部网络, 在视频理解和目标检测方面取得巨大的成功. Fu 等人[17]在图像分割中引入非局部自注意力, 提出了一种双重注意力, 以提高在语义分割场景下的特征表示. He 等人[18]提出一种 ACM (adaptive context module) 注意力, 基于 ACM 注意力提出一种自适应金字塔上下文网络. 注意力也在风格迁移领域中得到了广泛的应用, SANet[8]是第 1 个基于自注意力的风格迁移模型. Deng 等人[19]提出了由多个基于注意力机制自适应模块组成的风格迁移模型. 杨玥等人[20]引入交叉注意力作为特征融合的方法. Li 等人[21]提出了一种新的注意力风格迁移网络, 该网络依赖于最优传输来计算注意力权重图来得到内容和风格图像之间的相似性, 较好地实现了风格迁移. 在特征融合中引入注意力, 可以有效地对齐内容和风格特征并保留更多的语义信息.

2 本文方法

2.1 模型结构

本文提出的图像任意风格迁移模型是端到端的模型, 它由编码器 (VGG19)、特征融合模块、解码器组成. 图 1(a) 展示了整个模型的框架. 使用一个预训练好的 VGG19 作为编码器用于提取图像的深度特征, 使用与 VGG19 网络具有对称结构的网络作为解码器. 为保留更多的浅层的细节特征和深层的语义信息, 本文使用 VGG19 的 ReLU_1_1, ReLU_2_1, ReLU_3_1, ReLU_4_1 和 ReLU_5_1 这 5 层提取到的图像特征作为输入. 假设一幅内容图像 I_C 和任意一幅风格图像 I_S, 使用 VGG19 提取它们的深度特征, 得到不同层级的内容图像特征和风格图像特征分别表示为F_C^i和F_S^i. 表达式为:

$$F_C^i = E^i(I_C) \tag{1}$$

$$F_S^i = E^i(I_S) \tag{2}$$

其中, i 表示不同层, E 表示 VGG19 中该层的特征图.

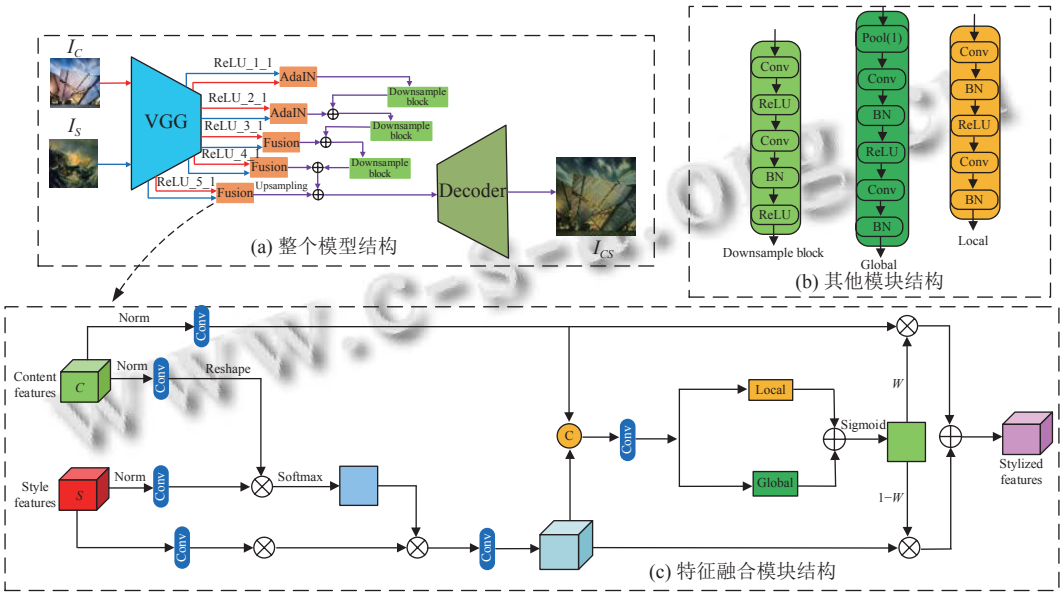

图 1 模型结构

在使用预训练好的 VGG19 提取到内容图像和风格图像特征后, 输入到特征融合模块中, 得到融合之后的风格化特征. 由于计算条件的限制以及为了利用图像特征的二阶统计量, 浅层 ReLU_1_1 和 ReLU_2_1 使用 AdaIN[7]进行特征融合得到风格化特征, 深层 ReLU_3_1, ReLU_4_1 和 ReLU_5_1 层使用本文提出的特征融合模块进行融合得到风格化特征. 表达式为:

$$F_{CS}^i = Fuse(F_C^i, F_S^i) \tag{3}$$

其中, i 表示不同层, *Fuse* 表示特征融合.

为了融合不同层的图像特征, 本文将不同层级的风格化特征相加融合. 因此, 提出一个下采样模块 (结构如图 1(b)), 浅层的风格化特征经过下采样模块后和深层的风格化特征相加. 表达式为:

$$F_{CS\_C}^{r\_2\_1} = F_{CS}^{r\_2\_1} + Down(F_{CS}^{r\_1\_1}) \tag{4}$$

$$F_{CS\_C}^{r\_3\_1} = F_{CS}^{r\_3\_1} + Down(F_{CS\_C}^{r\_2\_1}) \tag{5}$$

$$F_{CS\_C}^{r\_4\_1} = F_{CS}^{r\_4\_1} + Down(F_{CS\_C}^{r\_3\_1}) \tag{6}$$

其中, *Down* 表示下采样模块, *r*_1_1, *r*_2_1, *r*_3_1 和 *r*_4_1 分别表示 ReLU_1_1, ReLU_2_1, ReLU_3_1 和 ReLU_4_1 层.

对 $F_{CS}^{r\_5\_1}$ 进行上采样操作, 然后与 $F_{CS\_C}^{r\_4\_1}$ 相加后输入到解码器中得到最终的风格化图像, 表达式为:

$$I_{CS} = Decoder(F_{CS\_C}^{r\_4\_1} + Up(F_{CS}^{r\_5\_1})) \tag{7}$$

其中, *Decoder* 表示编码器, *Up* 表示上采样.

2.2 特征融合模块

本文提出一种新的基于注意力机制的特征融合模块, 模块可以根据内容的语义空间分布, 有效地、灵活地进行内容特征和风格特征的融合. 模块结构如图 1(c). 内容图像特征 F_C 和风格图像特征 F_S 分别先正则化然后再经过卷积后得到特征 $\bar{F}_C$ 和 $\bar{F}_S$, $\bar{F}_C$ 的转置和 $\bar{F}_S$ 进行矩阵相乘后使用 Softmax 函数得到内容图像特征和风格图像特征的相似性注意力图, 再将风格图像特征经过一个卷积后与该注意力图相乘得到融合特征 F_{CS}. 然后正则化后的 F_C 再经过一个卷积生成特征 $\bar{F}_{CC}$, $\bar{F}_{CC}$ 和 F_{CS} 在通道维度进行拼接融合后输入到一个卷积中调整通道数后, 分别经过全局分支提取全局信息和局部分支提取局部特征 (结构如图 1(b)) 并相加后使用 Sigmoid 函数得到注意力图 W, $\bar{F}_{CC}$ 和 F_{CS} 分别与 W 和 $(1–W)$ 相乘后相加得到最终的风格化特征.

2.3 损失函数

本文提出的网络使用 3 个损失函数进行模型训练, 分别为内容损失函数、风格损失函数、一致性损失函数. 同时, 使用预训练好的 VGG19 网络提取图像特征计算损失函数. 因此, 总损失函数表达式为:

$$L_{\text{all}} = \lambda_C L_C + \lambda_S L_S + L_{\text{CON}} \tag{8}$$

其中, 内容、风格和一致性的损失函数表示分别为 L_C、L_S 和 L_{CON}; λ_C 和 λ_S 是不同损失的权重.

2.3.1 内容损失函数

与 AdaIN[7]类似, 分别计算 VGG19 的 ReLU_4_1 层和 ReLU_5_1 层输出的风格化图像特征和内容图像特征之间的欧氏距离之差后相加. 具体表达式为:

$$L_C = \left\| E(I_{CS}^{r\_4\_1}) - E(I_C^{r\_4\_1}) \right\|_2 + \left\| E(I_{CS}^{r\_5\_1}) - E(I_C^{r\_5\_1}) \right\|_2 \tag{9}$$

其中, E 表示用于计算损失的编码器中该层的特征图.

2.3.2 风格损失函数

与 AdaIN[7]类似, 分别计算 VGG19 不同层输出的风格化图像特征的均值、方差和风格图像特征的均值、方差之间的欧氏距离之差后相加, 具体表达式为:

$$S(x,y) = \|\mu(x) - \mu(y)\|_2 + \|\sigma(x) - \sigma(y)\|_2 \tag{10}$$

$$\begin{aligned} L_S &= \sum_{i=1}^{N} S(E^i(I_{CS}), E^i(I_S)) \\ &= \sum_{i=1}^{N} \left\| \mu(E^i(I_{CS})) - \mu(E^i(I_S)) \right\|_2 \\ &\quad + \sum_{i=1}^{N} \left\| \sigma(E^i(I_{CS})) - \sigma(E^i(I_S)) \right\|_2 \end{aligned} \tag{11}$$

其中, μ 表示特征的平均值; σ 表示特征的方差; i 表示不同层; N 表示计算所用的层, 使用权重相同的 ReLU_1_1, ReLU_2_1, ReLU_3_1, ReLU_4_1 和 ReLU_5_1 层.

2.3.3 一致性损失函数

本文提出一种新的损失函数, 即一致性损失函数. 该损失采用相同的图像作为输入, 因此该损失函数更关注图像本身内容结构的保留和风格的一致性而不是风格的改变. 所以, 可以较好地保留内容图像的结构和风格图像的风格. 具体表达式如下:

$$\begin{aligned} L_{\text{CON}} &= \lambda_1(\|I_{CC} - I_C\|_2 + S(I_{SS}, I_S)) \\ &\quad + \lambda_2 \sum_{i=1}^{N} \left(\left\| E^i(I_{CC}) - E^i(I_C) \right\|_2 + S(E^i(I_{SS}), E^i(I_S)) \right) \end{aligned} \tag{12}$$

其中, I_{CC} 表示两个相同的内容图像生成的输出图像; I_{SS} 表示两个相同的风格图像生成的输出图像; i 表示不同层; N 表示计算所用的层, 使用权重相同 ReLU_1_1, ReLU_2_1, ReLU_3_1, ReLU_4_1 和 ReLU_5_1 层; λ_1 和 λ_2 表示权重.

3 实验

3.1 实验设置

训练模型时, 使用 MS-COCO[22]作为内容图像数

据集, WikiArt[23]作为风格图像数据集, 每个数据集包含 80 000 幅图像. 在训练过程中, 损失函数的权重分别设置为 λ_C=1, λ_S=3, λ_1=1, λ_2=50. 使用 Adam 优化器, 设置学习率为 0.000 1, 一个批次为 5 幅内容图像和风格图像, 把输入的图像随机裁切为 256×256 像素大小. 测试模型时, 支持任意大小的图片作为输入.

3.2 视觉效果对比

为了评价本文提出的模型, 展示了与目前主流模型结果的比较, 包括 AdaIN[7]、SANet[8]、AdaAttN[9]、ArtFlow[11]、CCPL[12]. 图 2 给出了各种模型生成的风格化图片. AdaIN 只需要将内容特征和风格特征的均值和方差进行对齐, 但是由于该方法过于简化, 其生成的图像质量受到影响, 第 3 列所展示的第 2 个女生、熊、花瓣图像质量明显较差. SANet 提出一种风格注意力网络进行内容和风格特征的融合. 但是内容图像的显著特征会被扭曲且无法保留图像细节语义, 第 4 列的所有图像有明显失真和内容扭曲. AdaAttN 提出一种基于注意力的自适应规范化网络. 但是 AdaAttN 模型上生成的风格图像出现了风格不符合的或者重复风格的内容, 且没法保留内容图像的细节特征, 比如第 5 列城市上方的天空、女人的脸部、熊的身体、花瓣的花蕊. ArtFlow 模型虽然有效防止了内容泄露, 保持了内容图像的整体结构, 但是细节上还有些欠缺, 比如第 6 列城市上方天空云彩的丢失, 第 2 个女生的面部有些扭曲, 麦田里麦子的缺失, 以及花瓣整体细节的缺失. CCPL 模型生成的图像整体风格化效果较好, 但是风格化图像的内容图像细节保留不完整, 比如第 7 列第 2 个女生的嘴巴和眼睛、熊的身体、花瓣的花蕊. 与上述方法不同的是, 本文提出的模型可以进一步保留内容的显著信息和细节纹理. 此外, 提出的方法可以通过学习内容与风格特征之间的关系, 有效地整合内容和风格, 使生成的结果不仅保留了内容的语义信息, 而且还包含了丰富的风格图像的颜色和纹理.

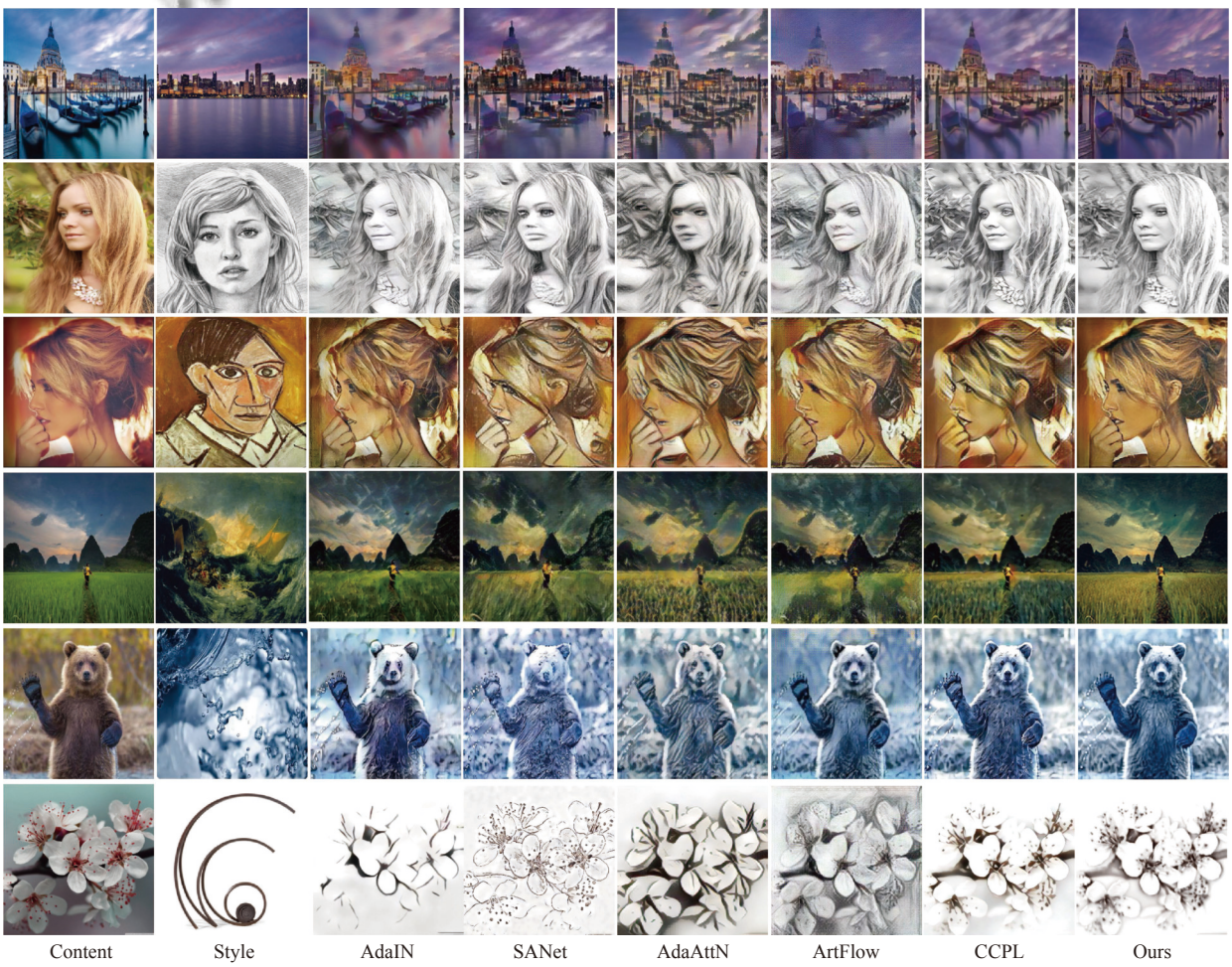

图 2 视觉对比

3.3 用户调查

用户的直观感受也是评价风格迁移质量的重要指标, 本文对 108 名参与者进行投票调查. 随机选取 10 张内容图片和 10 张风格图片, 并用上面提到的 6 种模型生成风格化图片. 108 名参与者从每种风格化图片中选择出自己认为最佳的图片, 共收集到 1 077 张投票. 从图 3 中可以看出, 本文提出的模型生成的风格化图片比其他的模型生成的更具有吸引力.

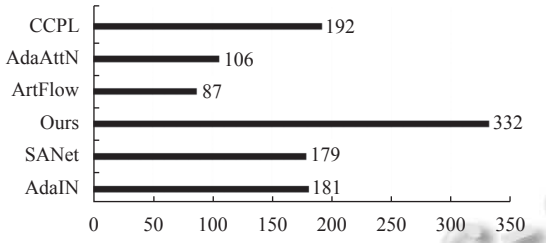

图 3 用户投票

3.4 定量指标对比

本文使用 SSIM (structure similarity index measure) 和 PSNR (peak signal-to-noise ratio) 来衡量内容图像和生成的风格化图像之间的结构相似性. 为了准确计算 SSIM 和 PSNR, 计算测试集的内容图像和模型生成的风格图像之间的 SSIM 和 PSNR, 并取平均值, SSIM 和 PSNR 的数值越大代表图像的质量越高. 结果如表 1. 可以看出, 本文模型的 SSIM 和 PSNR 值最高.

表 1 本文模型和其他模型的 SSIM 和 PSNR 对比

Model	SSIM	PSNR (dB)
AdaIN	0.659	13.587
SANet	0.536	12.427
AdaAttN	0.632	13.843
ArtFlow	0.572	12.801
CCPL	0.680	13.773
Ours	**0.701**	**14.381**

3.5 消融实验

为了验证本文中所提出的方法的有效性, 分别设计了特征融合模块、多层级特征、一致性损失函数的消融实验, 从定量和视觉对比两个方面进行有效性的验证.

3.5.1 特征融合模块消融实验

为了验证特征融合模块的有效性, 设计了相关的消融实验. 模型 a 是使用 AdaIN[7]替换本文提出的特征融合模块, 其他设置保持一致. 如图 4(a) 是模型 a 生成的风格化图像. 从实验结果可以看出, 本文提出的模块更好地保留内容图像的结构和细节信息以及风格图像的风格, 例如花朵的花蕊、女人人像的面部等细节都得到了较好的保留. 在测试集上计算风格化图像和内容图像之间的 SSIM 和 PSNR 值, 并取平均值. 从表 2 可以看出两个值都有小幅度的下降. 因此, 本文提出的特征融合模块在模型中是比较重要的模块.

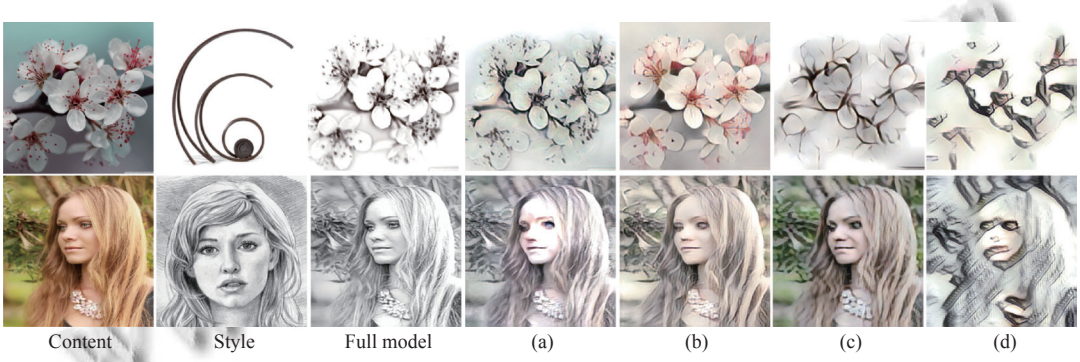

图 4 消融实验

3.5.2 多层级特征消融实验

为了验证使用多层级特征的有效性, 设计了 2 个消融实验. 模型 b 是去掉 ReLU_1_1 和 ReLU_2_1 层, 其他设置保持一致. 模型 c 是与 SANet[8]一样只使用 ReLU_4_1 和 ReLU_5_1 层提取出的深层特征, 其他设置保持一致. 如图 4(b) 和图 4(c) 图像是模型 b 和模型 c 生成的风格化图像. 从实验结果可以看出, 产生的图片的细节没法保留完整, 比如图 4(b) 和图 4(c) 中女人的眼睛和嘴巴的扭曲、失真, 花瓣细节的缺失, 并且相比于去掉 3 层特征的模型 c 生成的图像, 只去掉 2 层特征的模型 b 生成的图像也保留了更多的细节. 在测试集上计算风格化图像和内容图像之间的 SSIM 和 PSNR 值, 并取平均值. 从表 2 可以看出两个值都有一定幅度的下降. 因此, 模型中使用提取到的多层特征是有效的.

3.5.3 一致性损失函数消融实验

为了验证提出的一致性损失函数的有效性, 设计

了相关的消融实验. 模型 d 是训练模型时去掉该损失函数, 其他设置保持一致. 如图 4(d) 是模型 d 生成的风格化图像. 从实验结果可以很明显地看出, 去掉一致性损失函数后产生的图片无法保持全局结构的完整性, 而加上该损失函数后完整的网络可以避免这些问题. 在测试集上计算风格化图像和内容图像之间的 SSIM 和 PSNR 值, 并取平均值. 从表 2 可以看出两个值有大幅度的下降. 因此, 一致性损失函数在模型训练过程中是比较重要的.

表 2 消融实验的 SSIM 和 PSNR 对比

Model	SSIM	PSNR (dB)
Full	**0.701**	**14.381**
a	0.631	13.775
b	0.564	13.492
c	0.551	13.236
d	0.402	11.732

4 结束语

本文提出一个可以有效保留内容图像细节特征的图像任意风格迁移模型, 模型中包括融合编码器提取到的多层级图像特征, 确保在风格迁移过程中保留内容图像的重要语义信息和细节特征; 提出一个新的特征融合模块, 可以高质量地融合内容特征和风格特征. 此外, 在训练过程中, 本文还提出一个新的损失函数来训练模型. 实验结果表明, 本文提出的迁移模型可以保留内容图像的细节特征, 较好地融合内容特征和风格特征, 生成高质量的具有比较好的观感的风格化图像.

参考文献

1 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: ACM, 2012. 1097–1105. [doi: 10.5555/2999134.2999257]

2 Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 2414–2423.

3 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2015.

4 Johnson J, Alahi A, Li FF. Perceptual losses for real-time style transfer and super-resolution. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 694–711.

5 Ulyanov D, Lebedev V, Vedaldi A, *et al*. Texture networks: Feed-forward synthesis of textures and stylized images. Proceedings of the 33rd International Conference on Machine Learning. New York: ACM, 2016. 1349–1357.

6 Li YJ, Fang C, Yang JM, *et al*. Universal style transfer via feature transforms. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017. 385–395.

7 Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 1510–1519.

8 Park DY, Lee KH. Arbitrary style transfer with style-attentional networks. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 5873–5881.

9 Liu SH, Lin TW, He DL, *et al*. AdaAttN: Revisit attention mechanism in arbitrary neural style transfer. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 6629–6638.

10 Huo J, Jin SY, Li WB, *et al*. Manifold alignment for semantically aligned style transfer. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 14841–14849.

11 An J, Huang SY, Song YB, *et al*. ArtFlow: Unbiased image style transfer via reversible neural flows. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 862–871.

12 Wu ZJ, Zhu Z, Du JP, *et al*. CCPL: Contrastive coherence preserving loss for versatile style transfer. Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022. 189–206.

13 Wang YZ, Geng YL. Arbitrary style transfer with adaptive channel network. Proceedings of the 28th International Conference on Multimedia Modeling. Phu Quoc: Springer, 2022. 481–492.

14 Guo MH, Xu TX, Liu JJ, *et al*. Attention mechanisms in computer vision: A survey. Computational Visual Media, 2022, 8(3): 331–368. [doi: 10.1007/s41095-022-0271-y]

15 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017. 6000–6010.

16 Wang XL, Girshick R, Gupta A, *et al*. Non-local neural networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7794–7803.

17 Fu J, Liu J, Tian HJ, *et al*. Dual attention network for scene segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3141–3149.

18 He JJ, Deng ZY, Zhou L, *et al*. Adaptive pyramid context network for semantic segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: 2019. 7511–7520.

19 Deng YY, Tang F, Dong WM, *et al*. Arbitrary style transfer via multi-adaptation network. Proceedings of the 28th ACM International Conference on Multimedia. Seattle: ACM, 2020. 2719–2727.

20 杨玥, 冯涛, 梁虹, 等. 融合交叉注意力机制的图像任意风格迁移. 计算机科学, 2022, 49(S1): 345–352, 396. [doi: 10.11896/jsjkx.210700236]

21 Li J, Wu LW, Xu D, *et al*. Arbitrary style transfer with attentional networks via unbalanced optimal transport. IET Image Processing, 2022, 16(7): 1778–1792. [doi: 10.1049/ipr2.12403]

22 Lin TY, Maire M, Belongie S, *et al*. Microsoft COCO: Common objects in context. Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014. 740–755.

23 Phillips F, Mackintosh B. Wiki art gallery, Inc.: A case for critical thinking. Issues in Accounting Education, 2011, 26(3): 593–608. [doi: 10.2308/iace-50038]

(校对责编: 孙君艳)