计算机科学
COMPUTER SCIENCE

语义风格一致的任意图像风格迁移

颜明强, 余鹏飞, 李海燕, 李红松

引用本文

颜明强, 余鹏飞, 李海燕, 李红松.语义风格一致的任意图像风格迁移[J]. 计算机科学, 2023, 50(7): 129-136.

YAN Mingqiang, YU Pengfei, LI Haiyan, LI Hongsong. Arbitrary Image Style Transfer with Consistent Semantic Style [J]. Computer Science, 2023, 50(7): 129-136.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220700008

语义风格一致的任意图像风格迁移

颜明强　余鹏飞　李海燕　李红松
云南大学信息学院　昆明 650000
(mingqyan@163.com)

摘　要　图像风格迁移的目标是通过将目标图像风格迁移到给定的内容图像来合成输出图像。目前已有大量关于图像风格迁移的工作,但这些方法的风格化结果忽略了内容图像不同语义区域的流形分布,同时,大多数方法使用全局统计数据(如 Gram 矩阵或协方差矩阵)来实现风格特征到内容特征的匹配,不可避免地存在内容丢失、风格泄漏和伪影的问题,从而产生不一致的风格化结果。针对以上问题,提出了一个基于自注意力机制的渐进式流形特征映射模块(MFMM-AM),用于协调一致地匹配相关内容和风格流形之间的特征;然后通过在图像特征空间中应用精确直方图匹配(EHM)来实现风格和内容特征图的高阶分布匹配,减少了图像信息的丢失;最后,引入了两个对比性损失,利用大规模风格数据集的外部信息来学习人类感知的风格信息,使风格化图像的色彩分布和纹理图案更加合理。实验结果表明,与现有典型的任意图像风格迁移方法相比,所提网络极大地弥合了人类创作的艺术品和人工智能创作的艺术品之间的鸿沟,可以生成视觉上更加和谐和令人满意的艺术图像。

关键词:图像风格迁移;流形分布;自注意力机制;特征映射;高阶分布匹配
中图法分类号　TP391

Arbitrary Image Style Transfer with Consistent Semantic Style

YAN Mingqiang, YU Pengfei, LI Haiyan and LI Hongsong
School of Information, Yunnan University, Kunming 650000, China

Abstract　The goal of image style transfer is to synthesize an output image by transferring the style of the target image to a given content image. There are a large number of image style transfer works, but the stylization results ignore the manifold distribution of different semantic regions of the content image. At the same time, most methods use global statistics (for example, Gram matrix or covariance matrix) to achieve the matching of style feature to content feature. There are inevitable issues of content loss, style leakage, and the presence of artifacts, resulting in inconsistent stylized results. Aiming at the above problems, a self-attention mechanism-based progressive manifold feature mapping module (MFMM-AM) is proposed to coordinately match features between related content and style manifolds. Exact histogram matching (EHM) is applied to achieve higher-order distribution matching of style and content feature maps, reducing the loss of image information. Finally, two contrastive losses are introduced to learn human beings using the external information of large-scale style datasets perceived style information that makes the color distribution and texture patterns of stylized images more reasonable. Experimental results show that, compared with the existing typical arbitrary image style transfer methods, the proposed network greatly bridges the gap between human-created artworks and AI-created artworks, and can generate visually more harmonious and satisfying artistic images.

Keywords　Image style transfer, Manifold distribution, Self-attention mechanism, Feature mapping, Higher-order distribution matching

1　引言

图像风格迁移指借助某种算法,将参考图像的风格特征移植到目标内容图像中[1]。然而,图像风格是一个比较抽象的艺术概念,除了专门从事艺术研究的专家以外,一般研究者很难对其有一个清晰的概念把握,如何有效提取并迁移风格特征一直是图像风格迁移的难点之一。2015 年,Gatys 等[2]借助卷积神经网络来提取图像浅层语义信息和深层风格信息,开创性的地提出了神经风格迁移工作,吸引了大批研究者的关注,为在计算机辅助下进行图像生成注入大量的动力。目前,大多数方法[2-5]都假设图像风格可以由深度网络特征的全局统计信息来表示,如 Gram 矩阵或协方差矩阵,使用这样

到稿日期:2022-07-01　返修日期:2022-11-17
基金项目:国家自然科学基金(62066046)
This work was supported by the National Natural Science Foundation of China(62066046).
通信作者:余鹏飞(pfyu@ynu.edu.cn)

的全局统计数据从整张图像中捕获风格，并应用于内容图像，而不区分内容图像语义区域。因此，对于包含不同语义区域的内容图像，这样的全局统计信息不足以表示正确风格迁移所需的全部风格。另一种风格迁移方法是基于局部切片的方法[6-10]，其将内容图像的局部特征和风格图像的局部特征进行替换，以产生风格化的输出。然而，风格特征可能会与不具有相似语义的内容特征匹配，从而导致伪影。因此，在许多情况下，基于全局统计的方法和基于局部补丁的方法都不适用。此后也有不少研究者借助生成对抗网络(Generative Adversarial Networks，GANs)[11]进行图像风格迁移及生成，但大量的工作更多的是关注图像的纹理和颜色信息，图像局部以及更深层次的语义信息却没能得到很好的保留，传输质量以及效率都有待改进。之后的研究工作更多的是关注实时的任意图像风格迁移[12-18]，让模型可以接受任何领域的风格图像作为参考图像，输入图像在预先训练好的模型指导下就能实现灵活且高效的风格迁移效果，取得了显著的进步和发展。但该类方法为了满足任意风格转换的灵活性，损失了图像的局部特征和一些空间语义信息质量，如油画、水墨画等艺术作品的笔触和写意等。

总之，现有的图像风格迁移方法借助深度神经网络有了显著的改进，但真实的艺术作品与迁移合成风格化图像之间仍然存在着很大的差距，不可避免地存在以下问题。

(1)如图1所示，以Gatys等[2]基于迭代优化的方法(见图1第一列)、自适应实例规范化(Adaptive Instance Normalization，AdaIN)风格迁移方法[3](见图1第二列)、WCT方法[4](见图1第三列)和线性风格转移(LST)[5]方法(见图1第四列)为例，风格化的结果图像通常会存在色彩不协调(见图1黄色框)和内容细节丢失(见图1绿色框)等问题，这使得它们很容易与真实的艺术品区分开来。

(a)NST (b)AdaIN (c)WCT (d)LST

图1 风格图像局部区域对比(电子版为彩图)

Fig. 1 Local area comparison of style images

(2)目前大多数任意图像风格迁移方法倾向于使用全局统计数据(如Gram矩阵或协方差矩阵)强制输出图像和风格图像具有近似的全局统计数据分布来实现图像风格迁移[2-4]。如图1的第二列所示，人脸、衣服和背景来自不同的语义区域，由于采用相同的风格模式渲染，不同语义区域不能被独立风格化，导致结果混乱，产生了视觉假象。

(3)现有的风格迁移方法通常采用内容损失和风格损失来加强内容到风格的关系和风格到内容的关系，而忽略了风格到风格以及内容到内容的关系[2-5]。具体来说，用相同风格图像渲染的风格化图像应该比用不同风格图像渲染的风格化图像具有更密切的风格关系。同样，基于相同内容图像得到的风格化图像应该比基于不同内容图像得到的风格化图像具有更密切的内容关系。

通过以上分析，本文针对此前的工作仍未能完全解决传输灵活性和迁移结果质量方面的问题，提出了一种新的任意图像风格迁移思路和解决方案。

(1)参考Huo[19]等的观点，内容图像每个语义区域对应一个特征流形，一张完整的内容图像的多个语义区域应遵循多种流形分布。同时，内容区域的特征应该只由那些最相关的风格流形进行风格化，不同语义区域应呈现不同的风格模式。为此，本文方法在模型训练过程中提出了一个基于自注意力机制的流形特征映射模块(Manifold Feature Mapping Module based on self-Attention Mechanism，MFMM-AM)，该模块通过自注意力机制逐步对齐内容流形与风格流形，并使用多个损失函数约束，以实现语义区域之间的一致性。

(2)仅使用一阶和二阶统计量进行特征分布匹配的图像风格迁移方法准确率较低，而使用高阶统计量进行分布匹配在计算资源上是不允许的。本文通过在图像特征空间中应用精确直方图匹配(Exact Histogram Matching，EHM)来实现经验累积分布函数(empirical Cumulative Distribution Function，eCDFs)图像特征的精确匹配，该方法隐式地利用高阶统计量，来捕获更深层次的风格特征，产生更令人满意的风格化结果。

(3)为了减小风格数据集中不同图像在细节上的差异，使它们在笔触、颜色分布、纹理模式、色调等上更符合人类的感知，本文引入了两个具有对比性的损失，即内容对比损失和风格对比损失，在内容或风格相同的情况下，将同一训练批次内基于相同风格图像(或内容图像)获得的风格化图像的风格特征(或内容特征)嵌入拉近，而把其他基于不同内容和风格图像获得的风格化图像特征嵌入推远。通过对比损失来学习大规模风格数据集的外部同类风格和内容信息，进而改进风格化结果，使风格化图像的色彩分布和纹理图案更加和谐。

2 任意图像风格迁移模型

本文方法完整的模型结构图如2所示。

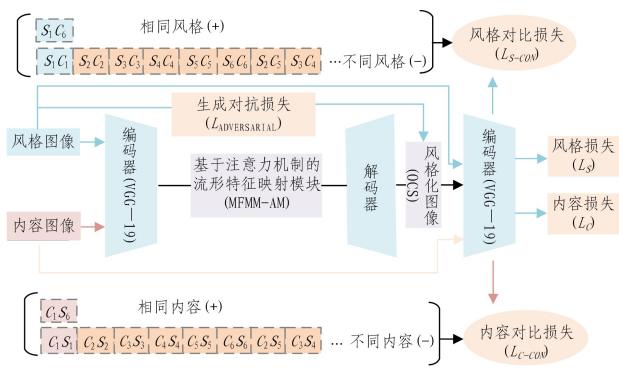


图2 本文网络整体结构

Fig. 2 Overall structure of the proposed network

模型由编码器(Encoder)模块、基于自注意力机制的流形特征映射模块(MFMM-AM)和解码器(Decoder)模块组成。Encoder是一个预训练的VGG-19网络[20]，取VGG-19的前几层(直到ReLU5_1)，用于提取图像特征。MFMM-AM是一个基于自注意力机制、空间感知插值、实例归一化3个子模块构成的流形特征映射网络，它可以灵活地将最接近语义的风格特征与内容特征匹配。Decoder是一个生成网络，用于将编码的语义特征映射转换为风格化的图像，解码器与编码器在结构上是对称的。

2.1 流形特征映射

现有方法忽略了特征的多流形分布，为了解决上述问题，本文提出基于自注意力机制的流形特征映射模块(MFMM-AM)，该模块主要包含自注意力机制、空间感知插值和实例归一化3个子模块。其中自注意力机制子模块能够学习一个感知流形分布的度量参数，协调一致地匹配相关内容和风格流形之间的特征。

首先，基于自注意力机制的流形特征映射模块(MFMM-AM)采用自注意力机制子模块，根据内容特征的空间结构对风格特征进行重新排列，通过对内容特征和风格特征的匹配，将相关的内容流形和风格流形注意特征图对应。然后借助实例归一化子模块，将一个给定内容特征图的内容特征与风格流形注意特征图相匹配，同时保留其语义信息。最后，空间感知插值融合子模块调整相应的流形与自适应权值。空间感知插值可以动态增加对应流形的结构相似性，使自注意力机制子模块更容易匹配它们之间的特征。

图3给出了本文提出的基于自注意力机制的流行特征映射(MFMM-AM)模块。首先使用预训练的VGG网络对内容图像 F_C 和风格图像 F_S 进行编码，得到ReLU5_1层内容特征图和风格特征图。自注意力机制子模块生成自注意力特征图 A_{CS}，重新排列风格特征。自注意力子模块、实例归一化子模块得到风格化特征图 F_{CS1} 和 F_{CS2}，空间感知插值子模块对通道信息进行密集处理，同时推理学习得到自适应权重 γ 和 β，δ 用于在风格化特征图 F_{CS1} 和 F_{CS2} 之间进行插值。

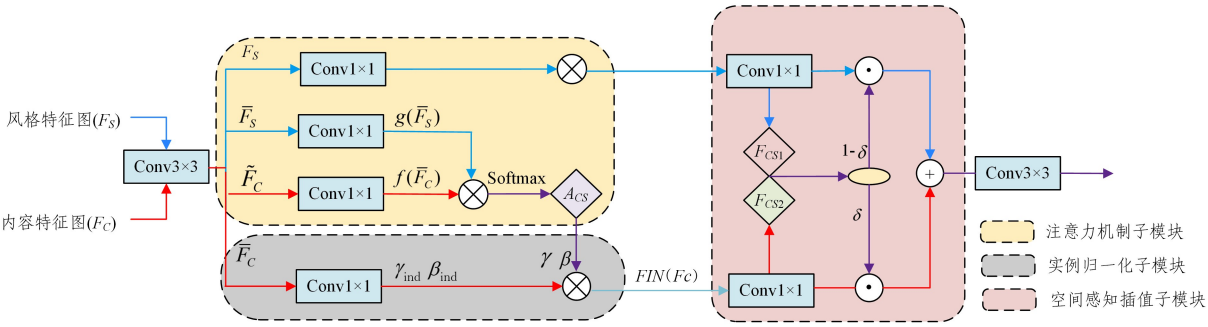


图3 基于自注意力机制的流形特征映射模块

Fig. 3 Manifold feature mapping module based on self-attention mechanism

2.1.1 自注意力机制子模块

自注意力机制子模块用于捕捉内容特征图和风格特征图之间的语义信息，生成基于风格化图像的归一化参数 γ 和 β。

$$A_{CS}=\mathrm{softmax}(f(\bar{F}_C)^{\mathrm{T}}\otimes g(\bar{F}_S)) \tag{1}$$

$$\gamma=\mathrm{ReLU}(w_\gamma\otimes A_{CS}) \tag{2}$$

$$\beta=\mathrm{ReLU}(w_S\otimes A_{CS}) \tag{3}$$

在这里，自注意力网络中 w_γ，w_s，$f(\cdot)$ 和 $g(\cdot)$ 表示用于特征嵌入的1×1卷积层，$\otimes$ 表示卷积运算，γ 和 β 是两个三维的规范化参数($\gamma,\beta\in R^{H\times W\times C}$)。其输入 F_C 和 F_S 为VGG19输出的内容图像和风格图像的特征图，对其先进行归一化处理得到 $\bar{F}_C$ 和 $\bar{F}_S$，然后计算其自注意力，自注意力特征图包含内容特征和风格特征之间的成对相似特征。本文使用自注意力机制子模块来生成规范化参数 γ 和 β，用于对内容特征进行相应元素方向的转换和调整，可以更灵活地将不同层次的颜色和纹理信息传输到内容图像的不同空间区域。

2.1.2 实例归一化子模块

给定任意内容特征图 $F_C\in R^{H\times W\times C}$，需要将内容特征与风格特征进行转换和匹配，同时保留内容图像语义信息。利用归一化操作获得归一化内容特征 $\bar{F}_C$ 后，实例归一化子模块的转换过程如下：

$$FIN(F_C)=\gamma\times(\boldsymbol{\gamma}_{\mathrm{ind}}\times\bar{F}_C+\boldsymbol{\beta}_{\mathrm{ind}})+\beta \tag{4}$$

此处 $\boldsymbol{\gamma}_{\mathrm{ind}}$ 和 $\boldsymbol{\beta}_{\mathrm{ind}}$ 表示与风格无关的参数，是沿通道维数的一维向量(即 $\boldsymbol{\gamma}_{\mathrm{ind}}$，$\boldsymbol{\beta}_{\mathrm{ind}}\in R^C$)。这些参数在风格转换之前转换 $\bar{F}_C$，用来捕获不同风格图像之间的共享转换，仅仅依靠空间自适应归一化并不能得到高质量的风格转换图像，实验结果图像存在局部的伪影信息，还需要对不同位置的风格特征和内容特征进行空间位置的约束。本文在归一化过程中引入自注意力机制子模块，以生成更加可靠的风格化结果图像。

2.1.3 空间感知插值子模块

该子模块将不同尺度的卷积核应用于拼接的特征上，总结多尺度的区域信息，然后进行通道密集运算，自适应地在 F_{CS1} 和 F_{CS2} 之间插入区域信息。

$$F_{CS1}=h(j(F_S)^{\mathrm{T}})\otimes A_{CS} \tag{5}$$

$$F_{CS2}=h(j(FIN(F_C))^{\mathrm{T}})\otimes A_{CS} \tag{6}$$

$$\delta=\frac{1}{n}\sum_{i=1}^{n}\phi_i([F_{CS1},F_{CS2}]) \tag{7}$$

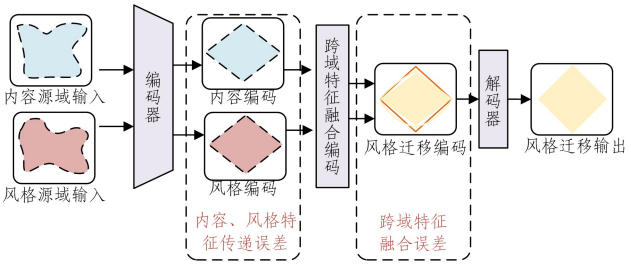
同样，$h(\cdot)$ 和 $j(\cdot)$ 表示用于特征嵌入的1×1卷积层。$\phi_i(\cdot)$ 表示第 i 个卷积核，$[\cdot,\cdot]$ 表示通道级联操作。级联特性可以帮助识别对应的内容和风格通道之间的差异，找出自注意力机制子模块引发的局部不一致性。这个可学习的通道密集运算输出空间匹配权值 $\delta\in R^{H\times W}$，用于插值。

$$F_{CS}=\delta\odot F_{CS1}+(1-\delta)\odot F_{CS2} \tag{8}$$

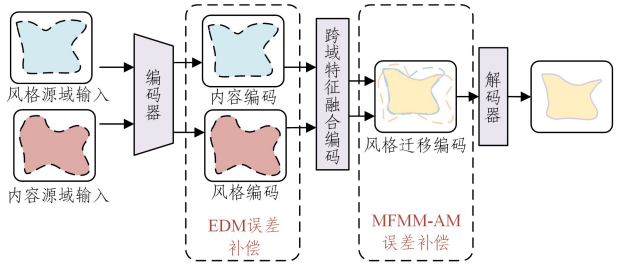
其中，$\odot$表示对应元素相乘，δ用于衡量流形匹配程度。该等式表明，内容流形中的特征只能由最相关的风格流形中的特征进行风格化，任何流形外渲染都被视为不匹配。如果不匹配权值δ相对较高，则风格化结果往往不一致。本节空间感知插值融合的是同一空间中的特征，因此插值后的内容特征不会出现退化，解决了局部失真问题。此外，空间感知插值增加了对应通道之间的亲和力，使自注意力子模块更容易呈现一致的语义区域。最后，对得到的风格化特征图F_{CS}实行高阶分布匹配以进行进一步的细化。

2.2 高阶分布匹配

任意图像风格迁移可以看作是特征分布匹配问题。在特征分布为高斯分布的假设下，传统的特征分布匹配方法[2,21]通常匹配特征的均值和标准差。然而，现实世界数据的特征分布通常太复杂，无法用高斯模型来建模。如图4所示，仅使用均值和标准差等统计量(见图4(a))进行特征分布匹配的准确率较低，存在特征传递误差和跨域特征融合误差。而使用高阶统计量(见图4(b))进行特征分布匹配能避免更多的特征传递误差，基于自注意力机制的流形特征映射模块(MFMM-AM)能减少更多的跨域特征融合误差。



(a)风格迁移输出存在传递误差



(b)MFMM-AM 和 EDM 进行误差补偿

图4 编解码风格迁移网络特征传递

Fig. 4 Encoder-Decoder style transfer network feature transmission

本文通过在图像特征空间中应用精确直方图匹配(EHM)来实现精确匹配图像特征的经验累积分布函数(eCDFs)，间接地利用高阶统计量来实现高阶精确特征分布匹配(EDM)[22]。该方法可以实现更多的风格化特征增强，对图像特征进行精确匹配。为了在模型中实现梯度反向传播，本文通过式(9)来实现EDM。

$$EDM(F_C,F_S)=(F_C)_{v_i}+(F_S)_{k_i}-\langle(F_C)_{v_i}\rangle \tag{9}$$

其中，$\langle\cdot\rangle$表示停止梯度传递操作，v_i和k_i表示内容和风格特征在$\boldsymbol{v}$和$\boldsymbol{k}$两个矢量方向上的一系列排序值。对给定输入数据$F_C\in R^{B\times C\times HW}$和风格数据$F_S\in R^{B\times C\times HW}$，首先将风格和内容特征编码到特征空间，然后在$\boldsymbol{v}$和$\boldsymbol{k}$两个矢量方向上进行排序匹配。其中$B,C,H,W$分别表示批次大小、通道尺寸、高度和宽度。最后得到风格迁移特征I为：

$$I=EDM(E(F_C),E(F_S)) \tag{10}$$

其中，E表示编码器操作。借助基于自注意力机制的流形特征映射模块(MFMM-AM)进行特征映射，将风格特征映射到图像空间，最后训练一个随机初始化的解码器Decoder，得到风格化的图像。

2.3 损失函数

2.3.1 内容损失

内容损失L_C是风格化图像特征$E(D(I))$与风格迁移后的特征I之间的欧氏距离。

$$L_C=\|E(D(I))-I\|_2 \tag{11}$$

其中，E和D分别表示编码器和解码器对风格迁移后的特征I进行编解码过程。

2.3.2 风格损失

风格损失度量风格图像$D(I)$和风格图像F_s的特征分布散度，为了更精确地测量分布散度，我们将风格损失引入为风格化图像的特征$\phi_i(D(I))$与其风格转移目标$EDM(_i(D(I)),\varphi_i(F_S))$之间的欧氏距离之和。

$$L_S=\sum_{i-1}^{L}\|\varphi_i(D(I))-EDM(\varphi_i(D(I)),\varphi_i(F_S))\|_2 \tag{12}$$

此处取$\{\varphi_i\}_{i=1}^{L}$为VGG-19的ReLU1_1，ReLU2_1，ReLU3_1，ReLU4_1和ReLU5_1层。

2.3.3 对比损失

为了弥补人工创作与人工创作之间的巨大差距，我们使用了两种新颖的对比损失[23]——风格对比损失和内容对比损失，以学习现有风格迁移方法忽略的风格化与风格化的关系。当存在一组内容图像C和一组风格图像S，使用对比损失的目标是学习输入的某一张风格图像的风格特征，以及风格数据集S中提取的外部人类感知的风格信息，然后转移到任意内容图像$Ic\in C$，生成新的艺术图像O_{SC}。

为了使表达更清晰，接下来用Ci表示第i个内容图像，Si表示第i个风格图像，$CiSi$表示内容图像Ci和风格图像Si生成的风格化图像。为了在每个训练批次中进行对比学习，我们将一批风格和内容图像按如下方式排列：

取批次batch大小为12，这是一个偶数。然后排序得到一批风格图像[$S1,S2,S3,S4,S5,S6,S6,S5,S4,S3,S2,S1$]，以及一批内容图像[$C1,C2,C3,C4,C5,C6,C1,C2,C3,C4,C5,C6$]，对应的风格化图像结果为[$C1S1,C2S2,C3S3,C4S4,C5S5,C6S6,C1S6,C2S5,C3S4,C4S3,C5S2,C6S1$]。通过这种方式，确保对于每个风格化图像$CiSj$，都可以找到一个风格化图像$CiSx(x\neq j)$与它共享相同的内容，但风格不一样，以及一个风格化图像$CySj(y\neq i)$与它在同一批次中共享相同的风格，但内容不一样。图2以$b=12$为例描述了这个过程。为了关联具有相同风格的风格化图像，对于一个风格化图像$CiSj$，我们选择$CySj(i\neq y)$作为它的正例($CiSj$与$CySj$具有相同的风格)，$CmSn(m\neq i$和$n\neq j)$作为它的反例。

此处 $CmSn$ 表示一系列风格化的图像，而不仅仅是一个图像。综上，风格对比损失表述如下：

$$L_{\text{S-CON}} = -\log\left(\exp\left(l_S\ (CiSj)^{\mathrm{T}} l_S\ (CySj)/\tau\right)/\exp\left(l_S\ (CiSj)^{\mathrm{T}} l_S(CySj)/t\right)+\sum\exp\left(l_S\ (CiSj)^{\mathrm{T}} l_S(CmSn)/t\right)\right) \tag{13}$$

其中，l_S = MFMM-AM($\varphi_i(\cdot)$)，MFMM-AM 为本文提出的基于自注意力机制的流形特征映射模块，φ_i 取 ReLU3_1 层。l_S 用于从风格化图像中获取风格嵌入，τ 是控制推力和拉力的超参数。

类似于风格对比损失，为了将具有相同内容的风格化图像相关联，对于一个风格化图像 $CiSj$，选择 $CiSy(y\neq j)$ 作为它的正例（$CiSj$ 和 $CiSy$ 内容相同），$CmSn(m\neq i, n\neq j)$ 作为它的反例。将内容对比损失表示为：

$$L_{\text{C-CON}} = -\log\left(\exp\left(l_C\ (CiSj)^{\mathrm{T}} l_C\ (CiSy)/t\right)/\exp\left(l_C\ (CiSj)^{\mathrm{T}} l_C(CiSy)/t\right)+\sum\exp\left(l_C\ (CiSj)^{\mathrm{T}} l_C(CmSn)/t\right)\right) \tag{14}$$

其中 l_C = MFMM-AM($\varphi_i(\cdot)$)，与风格损失对应，φ_i 取 ReLU4_1 层。l_C 用于从风格化图像中获取内容嵌入。

2.3.4 生成对抗损失

同时，本文使用 GAN[24] 从风格数据集 S 中学习人类感知的风格信息。GAN 是由两个相互竞争的网络（生成器 G 和鉴别器 D_{im}）组成的生成模型。其中生成器 G 由一个编码器（Encoder）、一个基于自注意力机制的流形特征映射模块（MFMM-AM）和一个解码器（Decoder）组成。生成对抗损失可以表述为：

$$L_{\text{ADVERSARIAL}} = \underset{F_S\sim S}{E}\left[\log\left(D_{im}(F_S)\right)\right]+\underset{F_C\sim C,F_S\sim S}{E}\left[\log\left(1-D_{im}\left(D(\text{MFMM-AM}(E(F_C),E(F_S)))\right)\right)\right] \tag{15}$$

2.3.5 身份损失

与 Park 等[25] 的研究相似，当内容图像和风格图像相同时，本文用身份损失来激励生成器 G 得到一个无差异的生成结果，这样可以在风格化结果中保留更多的内容结构和风格特征。身份损失定义为：

$$L_{\text{IDENTITY}} = \omega_{\text{IDENTITY1}}\left(\|F_{CC}-F_C\|^2+\|F_{SS}-F_S\|^2\right)+\omega_{\text{IDENTITY2}}\left(\|\varphi_i(F_{CC})-\varphi_i(F_C)\|^2+\|\varphi_i(F_{SS})-\varphi_i(F_S)\|^2\right) \tag{16}$$

其中，F_{CC} 和 F_{SS} 分别表示从两个相同内容（或风格）图像合成的输出图像，φ_i 表示编码器中的 ReLU1_1，ReLU2_1，ReLU3_1，ReLU4_1 和 ReLU5_1 层。

2.3.6 损失优化目标

总结前面提到的所有损失，得到本文模型的最终目标：

$$L_{\text{TOTAL}} = \omega_1 L_S + \omega_2 L_C + \omega_3 L_{\text{ADVERSARIAL}} + \omega_4 L_{\text{S-CON}} + \omega_5 L_{\text{C-CON}} + L_{\text{IDENTITY}} \tag{17}$$

其中，$\omega_{\text{IDENTITY1}}$，$\omega_{\text{IDENTITY2}}$，$\omega_1$，$\omega_2$，$\omega_3$，$\omega_4$ 和 ω_5 是为了在损失中达到适当平衡设置的超参数，用于加权组合训练解码器。

3 实验结果及分析

3.1 实验细节

本文实验使用包含 118 287 张实况图像的 MS-COCO 2017[26] 作为内容数据集，使用包含 79 433 张艺术图像的 WikiArt[27] 作为风格数据集。本文使用预训练自图像分类任务的 VGG-19 网络（截至 ReLU5_1 层）作为编码器，编码块由 ReLUn_1 层分隔。解码器结构上与编码器对称，其使用邻近上采样扩大特征图。至于鉴别器，本文采用了 Wang 等[28] 提出的多尺度鉴别器，在 3 个不同的尺度上进行判别并对结果取平均。式(16)和式(17)中的损耗权重分别设置为 $\omega_{\text{IDENTITY1}}=50$，$\omega_{\text{IDENTITY2}}=1$，$\omega_1=1$，$\omega_2=1$，$\omega_3=5$，$\omega_4=0.3$ 和 $\omega_5=0.3$。使用 Adam[29] 作为优化器。在训练期间，本文将 12 对内容和风格图像作为一个批次，较小尺寸的内容和风格图像被重新缩放到 512，然后随机裁剪为 256×256 的尺寸，以进行有效的训练。在测试阶段，网络模型支持处理任意大小的图像。

3.2 网络稳定性

本文实验环境基于操作系统为 Ubuntu20.04 的服务器，服务器版本为 i9 7900，GPU 为一块 Geforce GTX 3090。深度学习框架选择 pytorch1.6，编程环境选择 python3.8，CUDA 驱动版本选择 11.1。如图 5 所示，通过对输入图像的不断学习，然后根据模型输出结果的多个损失值反向调整生成模型和判别模型的网络参数，反复迭代优化，使多个损失函数值逐渐下降，网络模型逐渐稳定，当模型进行 160 000 次迭代优化、训练 71 h 后，模型稳定。

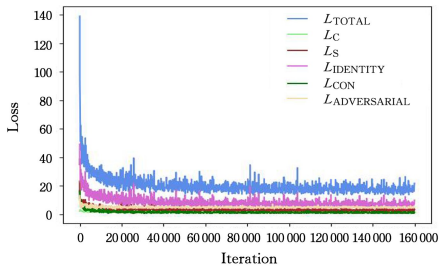


图 5 损失函数训练收敛情况

Fig. 5 Loss function training convergence

3.3 消融实验

为了证明本文实验所用模块的有效性，本文将逐步增加实验模块（MFMM-AM）和高阶分布匹配（EDM）后的实验结果图像进行了比较，如图 6 所示，最左边一列为图像输入，左上角为内容图像，右下角为风格图像。

为了进行公平的比较，每一例的对比模块和参数设置均保持相同。当不包含 MFMM-AM 和 EDM 模块时（见图 6(a)），内容图像只是迁移到风格图像简单的颜色和纹理。当加入 EDM 模块时（见图6(b)），内容图像学习到更多风格图像的颜色和纹理，但缺少不同语义区域的流形特征匹配，内容图像轮廓细节破坏严重。单独加入 MFMM-AM 模块时（见图 6(c)），结果图像迁移效果较好，但与同时加入 MFMM-AM 和 EDM 相比（见图6(d)），结果图像获得了更多风格图像的色彩饱和度和局部几何结构。事实证明，本文方法 MFMM-AM 和 EDM 均能在不同程度上对风格化结果起到优化作用，在同时加入 MFMM-AM 和 EDM 时，获得了最佳的风格化图像。

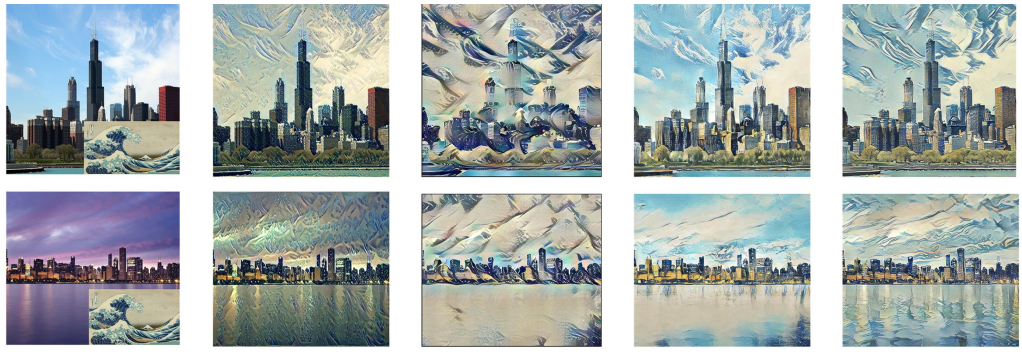


图 6 基于自注意力机制的流形特征映射模块(MFMM-AM)和高阶分布匹配(EDM)对风格化结果影响程度的比较

Fig. 6 Comparison of influence of manifold feature mapping module(MFMM-AM) and higher order distribution matching (EDM) based on attention mechanism on stylized results

3.4 对比实验

为验证本文方法的有效性,将其与目前具有代表性的任意图像风格迁移方法在风格化结果图像上进行了比较,并对本文方法进行评估和分析。

3.4.1 风格化结果比较

如图 7 所示,首先将本文方法与神经风格转移(Gatys)[2]、自适应实例归一化(AdaIN)[3]、WCT[4]方法、线性风格转移(LST)[5]、通过可逆神经流进行无偏图像风格转移的(ArtFlow)[16]方法、自适应注意力归一化(AdaAttN)[17]、用于通用风格迁移的对比相干性保留损失(CCPL)[18]和对比任意风格迁移(CAST)[19]进行风格化结果对比。Gayts 方法整体图像转换效果较差,其虽然能转移部分颜色和纹理,但大部分的风格内容丢失,转换前后区别不明显。AdaIN 直接调整内容全局特征的一、二阶统计分布信息,可以看到风格特征被转移,但存在严重的内容细节丢失(第 2—4 行)。WCT 方法主要针对写实类风格图片风格转换,画作图像风格化效果模糊,斑块明显,有明显的伪影(第 1,3,4 行)。LST 方法分别通过线性投影和单信道相关来修改特征,两者都能得到相对清晰的风格化输出,但风格图像的纹理模式没有自适应捕获,内容细节丢失(第 1,3,4 行)。通过可逆神经流进行无偏图像风格转移的(ArtFlow)方法尽可能保留了内容图像中的内容,但整体上风格化结果图像风格特征不明显。自适应注意力归一化(AdaAttN)取得了较好的风格迁移结果,但相比本文方法,由于其未考虑空间流形对齐,区域风格化混乱(第 4—6 行)。用于通用风格迁移的对比相干性保留损失(CCPL)方法虽然能对部分图像进行风格化,但伪影和区域风格混乱的问题依然存在(第 1 和第 5 行)。对比任意风格迁移(CAST)方法部分风格化结果存在内容丢失问题(第 2 和第 4 行)。如第 11 列所示,本文方法由于使用了流形对齐,可以自适应地将风格特征适当地传递到内容图像的不同内容区域。此外,高阶分布匹配和多种损失函数联合优化使得结果图像的更多细节得到保留。相比之下,本文方法在风格转移和内容结构保留之间取得了较好的权衡。

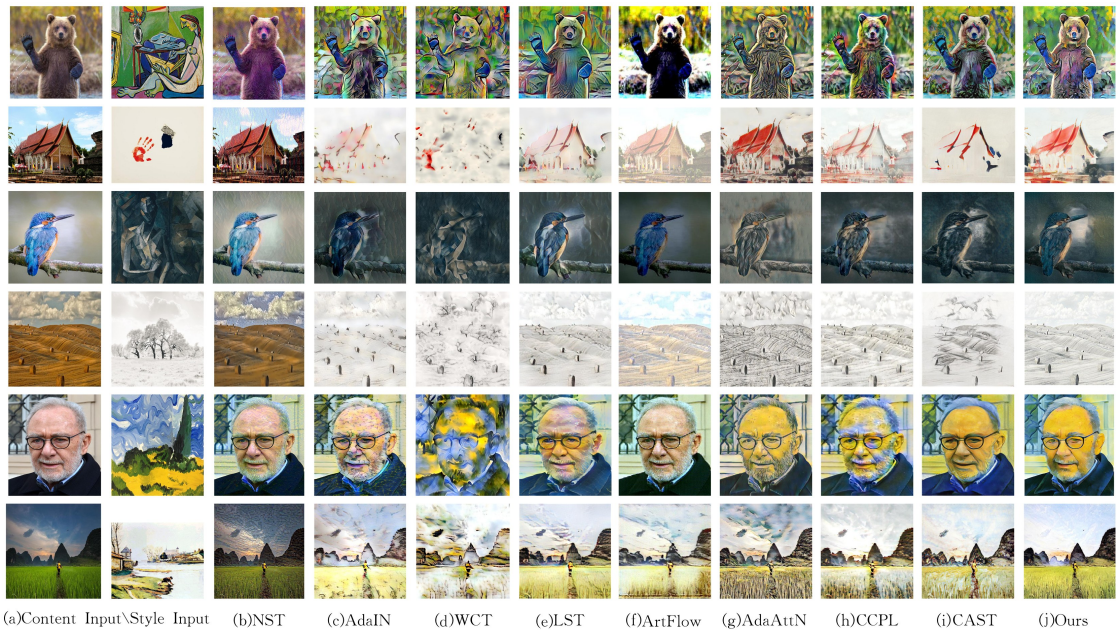


图 7 与其他风格迁移方法的风格化图像比较

Fig. 7 Comparison of stylized images with other style transfer methods

3.4.2 方法评估与分析

(1)客观评价分析

在表 1 中，本文利用单张图片风格化所需平均时间、结构相似度(Structural SIMilarity，SSIM)、峰值信噪比(Peak Signal-to-Noise Ratio，PSNR)和学习感知图像块相似度(Learned Perceptual Image Patch Similarity，LPIPS) 4 个指标参数，将本文方法与上述的风格迁移方法进行定性比较，其中，SSIM 和 PSNR 参数数值越大表示失真越小，LPIPS 参数数值越小表示两张图像越相似。为了保证比较的公平性，我们随机选取 6 对风格图像和内容图像，除了输入图片相同，其他都遵循相应方法的原始设置。从表 1 中的结果可以看出，本文所提方法取得了可与目前先进的迁移方法相媲美的结果。

表 1 不同风格迁移方法间的定量指标比较

Table 1 Comparison of quantitative indicators between different style transfer methods

Method	Time/s	*SSIM*	*PSNR*	*LPIPS*
Gayts	150.32	0.34	9.14	0.31
AdaIN	2.33	0.27	10.07	0.27
WCT	3.98	0.18	8.67	0.25
LST	1.74	0.24	8.96	0.24
ArtFlow	2.45	0.33	10.9	0.23
AdaAttN	1.68	0.31	11.1	**0.21**
CCPL	1.89	0.28	11.7	0.23
CAST	2.77	**0.36**	12.4	0.24
Ours	**1.66**	0.29	**13.9**	**0.21**

(2)主观评价分析

本文使用 25 个内容图像和 25 个样式图像，为每种方法生成总共 625 个样式化结果，然后从 625 个结果中随机抽取 20 个样本，分发给用户。对于每个样本，用户被要求在所有方法中选择他们最喜欢的样式化结果。我们使用了 3 个评价指标，即内容保留、风格质量和整体结果，来得到最少的伪影和扭曲。我们收集了 100 位用户的投票，并绘制了图 8。结果表明，本文方法产生的结果具有更高的一致性和整体性能。

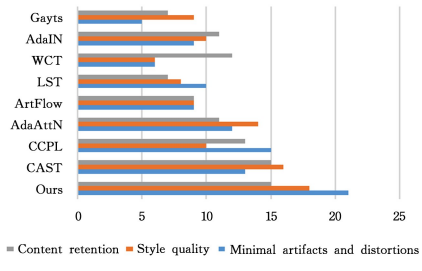


图 8 不同风格迁移方法的用户偏好调查

Fig. 8 User preference survey for different style transfer methods

结束语 为提高图像风格迁移的风格化质量，解决语义风格不一致的问题，本文提出了基于自注意力机制的流形特征映射模块(MFMM-AM)来缓解风格迁移结果色彩不协调和内容细节信息丢失的问题。MFMM-AM 中使用自注意力机制可以捕捉特征的多流形分布，揭示最相关的内容流形和样式流形之间的对应关系；通过执行实例归一化过程，最相关的内容和样式流形之间的结构相似性会增加，从而使自注意力机制更容易在它们之间一致地匹配特征；最后应用空间感知插值自适应地融合相关流形。本文使用多个损失函数联合约束迁移结果，实现多次重复特征流形匹配，从而产生高质量的一致性结果。高阶分布匹配(EDM)用于学习除均值和协方差之外的高阶统计分布信息，减少了特征传递过程中内容和风格信息传递误差。大量实验表明，该方法不仅可以生成视觉上更和谐、更令人满意的艺术图像，而且能提高任意输入的内容图像和风格图像对迁移结果的稳定性和一致性。此外，该方法快速高效，可以从一个新的角度为艺术风格转换任务提供参考。

参 考 文 献

[1] LI W S, ZHAO P, YIN L Z, et al. Regional diversified image style transfer method based on Gaussian sampling [J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(5): 8.

[2] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2414-2423.

[3] YIN W, YIN H, BARAKA K, et al. Dance style transfer with cross-modal transformer [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023: 5058-5067.

[4] LI Y, FANG C, YANG J, et al. Universal style transfer via feature transforms [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 385-395.

[5] LI X, LIU S, KAUTZ J, et al. Learning linear transformations for fast image and video style transfer [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3809-3817.

[6] SAMUTH B, TSCHUMPERLÉ D, RABIN J. A Patch-Based Approach for Artistic Style Transfer via Constrained Multi-Scale Image Matching [C] // 2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022: 3490-3494.

[7] SHENG L, LIN Z, SHAO J, et al. Avatar-net: Multi-scale zero-shot style transfer by feature decoration [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8242-8250.

[8] KıNLı F, ÖZCAN B, KıRAÇ F. Patch-wise contrastive style learning for instagram filter removal [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 578-588.

[9] LIAO J, YAO Y, YUAN L, et al. Visual attribute transfer through deep image analogy [J]. arXiv: 1705.01088, 2017.

[10] GU S, CHEN C, LIAO J, et al. Arbitrary style transfer with deep feature reshuffle [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8222-8231.

[11] KIM H, CHOI Y, KIM J, et al. Exploiting spatial dimensions of latent in gan for real-time image editing [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 852-861.

[12] SANAKOYEU A, KOTOVENKO D, LANG S, et al. A style-aware content loss for real-time hd style transfer [C] // Proceedings of the European Conference on Computer Vision (ECCV).

2018:698-714.
[13] ZHENG X A, MICHAEL WILBER B, CHEN F C, et al. Adversarial training for fast arbitrary style transfer[J]. Computers & Graphics, 2020, 87:1-11.
[14] ZHANG Y, LI M, LI R, et al. Exact feature distribution matching for arbitrary style transfer and domain generalization [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:8035-8045.
[15] AN J, HUANG S, SONG Y, et al. Artflow: Unbiased image style transfer via reversible neural flows[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:862-871.
[16] LIU S, LIN T, HE D, et al. Adaattn: Revisit attention mechanism in arbitrary neural style transfer[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 6649-6658.
[17] WU Z, ZHU Z, DU J, et al. CCPL: Contrastive Coherence Preserving Loss for Versatile Style Transfer[J]. arXiv: 2207.04808, 2022.
[18] ZHANG Y, TANG F, DONG W, et al. Domain Enhanced Arbitrary Image Style Transfer via Contrastive Learning[J]. arXiv: 2205.09542, 2022.
[19] HUO J, JIN S, LI W, et al. Manifold alignment for semantically aligned style transfer [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021:14861-14869.
[20] DING X, ZHANG X, MA N, et al. Repvgg: Making vgg-style convnets great again[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13733-13742.
[21] LI P, ZHAO L, XU D, et al. Optimal transport of deep feature for image style transfer[C]// Proceedings of the 2019 4th International Conference on Multimedia Systems and Signal Processing. 2019:167-171.
[22] ZHANG Y, LI M, LI R, et al. Exact feature distribution matching for arbitrary style transfer and domain generalization [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:8035-8045.
[23] CHEN H, WANG Z, ZHANG H, et al. Artistic Style Transfer with Internal-external Learning and Contrastive Learning[C]// NeurIPS. 2021.
[24] KAMMOUN A, SLAMA R, TABIA H, et al. Generative Adversarial Networks for face generation: A survey[J]. ACM Computing Surveys, 2022, 55(5):1-37.
[25] PARK D Y, LEE K H. Arbitrary style transfer with style-attentional networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:5880-5888.
[26] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]// European Conference on Computer Vision. 2014:740-755.
[27] PHILLIPS F, MACKINTOSH B. Wiki Art Gallery, Inc.: A case for critical thinking[J]. Issues in Accounting Education, 2011, 26(3):593-608.
[28] 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.
[29] LU L H. Simulation physics-informed deep neural network by adaptive Adam optimization method to perform a comparative study of the system[J]. Engineering with Computers, 2022, 38(Suppl 2):1111-1130.

YAN Mingqiang, born in 1996, postgraduate. His main research interests include pattern recognition and image style transfer.

YU Pengfei, born in 1974, Ph.D, associate professor. His main research interests include pattern recognition and image processing.

(责任编辑:何杨)