

结合特征聚类和深度残差收缩网络的壁画风格迁移

吴萌^{1*}, 高怡宁¹, 王佳²

¹西安建筑科技大学信息与控制工程学院, 陕西 西安 710055;

²陕西历史博物馆, 陕西 西安 710061

摘要 古代壁画为存世孤本且艺术风格独特。然而, 由于可参考数据稀少, 缺失信息重建的效果不佳。如何产生足够可参考样本是提升壁画修复效果的关键节点。为了解决此问题, 提出一种改进的风格迁移模型来生成高质量的壁画图像, 达到扩充数据的目的。首先, 在输入图像的低维空间进行聚类以保持图像的结构完整性。其次, 引入基于注意机制和软阈值函数的残差收缩模块去除图像中冗余信息, 有效保留纹理细节信息。最后, 将内容聚类集合与匹配到的风格聚类集合进行实时的特征转换, 得到任意风格的迁移图像。实验结果表明, 与其他常见的风格迁移方法相比, 所提方法能够生成自然清晰的壁画图像, 并且在客观指标峰值信噪比和结构相似度上均取得较好的结果。此外, 在基于生成对抗网络的壁画数字生成实验中, 验证了所提方法较传统增广方法的有效性。

关键词 风格迁移; 数据增广; 聚类匹配; 深度残差收缩网络; 数字生成

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP222583

Mural Style Transfer with Feature Clustering and Deep Residual Shrinkage Network

Wu Meng^{1*}, Gao Yining¹, Wang Jia²

¹School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, Shaanxi, China;

²Shaanxi History Museum, Xi'an 710061, Shaanxi, China

Abstract Ancient murals are unique surviving copies and unique in artistic style. However, the effect of reconstructing missing information is poor as reference data are scarce. The key to improving the reconstruction process of murals is producing enough reference samples. In this study, an improved style transfer method is proposed to generate high-quality mural images. First, clustering is performed in the low-dimensional space of the input images to maintain their structural integrity. Second, a residual shrinkage module based on the attention mechanism and soft threshold function is introduced to remove redundant information in the images and effectively retain texture details. Finally, the content and matched style clustering sets are converted into real-time features to obtain migration images of any style. The experimental results show that the proposed method can generate natural and clear mural images as well as achieve better results in terms of peak signal-to-noise ratio and structural similarity than other common style transfer methods. Moreover, in an experiment of mural digitization based on a generative adversarial network, the superiority of the proposed method over the conventional augmentation method is verified.

Key words style transfer; data augmentation; clustering matching; deep residual shrinkage network; digital generation

1 引言

墓室壁画具有极高的考古价值和艺术价值, 由于

大部分壁画的原址无法作为展出地点, 通常会采取分块揭取的特殊方式^[1]搬迁至博物馆内保存与展览。但是, 该方式很可能会损伤壁画, 致使壁画块间信息缺

收稿日期: 2022-09-20; 修回日期: 2022-11-13; 录用日期: 2022-11-24; 网络首发日期: 2023-01-04

基金项目: 国家自然科学基金(61701388)

通信作者: *wumeng@xauat.edu.cn

失,给相关的研究造成一定的障碍。为了重建壁画块间丢失的信息,通常采用图像复原技术根据壁画图像中的已知内容向外扩展纹理信息。一般来说,训练一个理想的深度神经网络需要大量数据。然而,收集完整的原始壁画非常耗时且难度大,这使得创建理想的大规模壁画数据集的任务难以完成。因此,以现有的少量数据集为基础对样本空间进行扩充,增加壁画生成网络训练数据的多样性十分重要。

目前,常用的图像扩充方法包括传统的转换方法和基于深度学习的方法。前者主要通过旋转、裁剪或向原始图像添加随机噪声等方法来生成新样本^[2]。这些方法只改变图像呈现方式而不是其内容,无法显著提高壁画生成模型的泛化能力。后者利用深度神经网络的学习能力来产生不同于原始数据的新数据。例如: Xiao 等^[3]结合风格迁移网络和模拟样本生成大量遥感船舶图像,一定程度上提升遥感图像分类任务的精度; Zhang 等^[4]利用风格迁移技术合成一个风格化视网膜眼底数据集,提高血管分割性能。为了增加训练数据的多样性,本实验组构建基于风格迁移网络的图像生成模型来为壁画数字生成研究创建更多的样本。

风格迁移是指在不改变语义内容的情况下,将一幅图像的艺术风格应用于另一幅图像的过程,主要是为了将原始图像用参考图像的色彩和纹理表示。 Gatys 等^[5]率先利用卷积神经网络对图像的内容和风格特征进行分离,实现风格迁移。该方法需要对图像进行迭代优化,速度较慢。 Johnson 等^[6]通过迭代优化模型来训练特定风格的前向残差网络,极大地提升了迁移速

度,但受限于固定数量的风格。 Huang 等^[7]提出一种结合自适应实例归一化层(AdaIN)和迭代优化的实时风格转换方法,通过调整内容图像的全局特征统计以实现任意风格的迁移。但是,壁画整体色彩单调,利用全局特征统计表示风格的方法不可避免地会扭曲局部风格特征,导致纹理重复。 Park 等^[8]提出风格注意力网络,平衡全局和局部风格模式,但可能会出现内容结构失真的问题。 Zhang 等^[9]认为图像风格特征遵循多重分布,提出使用 graph cut 的多模态表示方法,这种方法成功实现了风格与内容特征的匹配,但忽略了内容图像的结构完整性,通常伴随着大量噪声,导致生成图像出现伪影、扭曲等不符合人类视觉的部分。

基于上述分析,本文提出一种改进的风格迁移模型来生成高质量的壁画图像。该方法针对壁画迁移过程中因模型忽略局部风格特征导致的边缘光晕及伪影问题,在输入图像的低维 LAB 颜色空间进行聚类分析,提取结构特征以更好地模拟壁画的特征分布。其次,为了更好地辨别输入图像背景冗余信息,消除图像中的噪声,引入深度残差收缩模块以提高有效特征的权重,更好地捕捉图像的细节和纹理,进一步提高生成图像的质量。此外,使用结构相似性对生成图像进行评估,并将扩充后的数据集应用于壁画分块外延数字生成技术研究中,以证明所提网络的有效性。

2 风格迁移网络架构

所提风格迁移网络采用编码器-解码器架构,如图 1 所示,算法具体流程如图 2 所示。

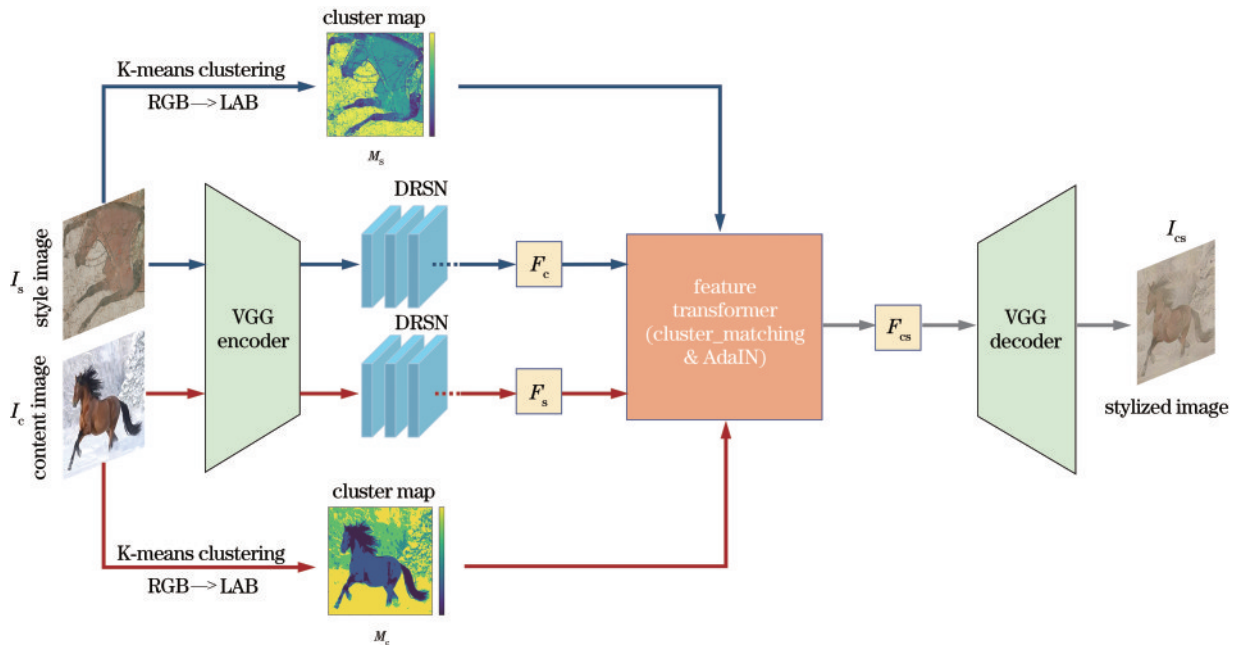


图 1 网络架构

Fig. 1 Network framework

Algorithm 1 Mural style transfer algorithm

Input: A pair of content image I_c and style image I_s
Output: A stylized image I_{cs}

- 1: for I_c and I_s do
- 2: Extract content features F_c and style features F_s using pre-trained VGG19
- 3: Filtering significant features and removing noise using the DRSN module
- 4: In the CIELAB color space K-means clustering to obtain M_c and M_s
- 5: for eSach cluster in M_c do
- 6: Clustering center norm indicates the cluster features
- 7: Set coverage ratio to calculate the matching result set for the clusters of M_s
- 8: Stylization F_c with F_s to obtain F_{cs} using AdaIN
- 9: end for
- 10: Sum over all F_{cs}
- 11: Feed F_{cs} to decoder
- 12: end for

图 2 算法流程

Fig. 2 Algorithm flow

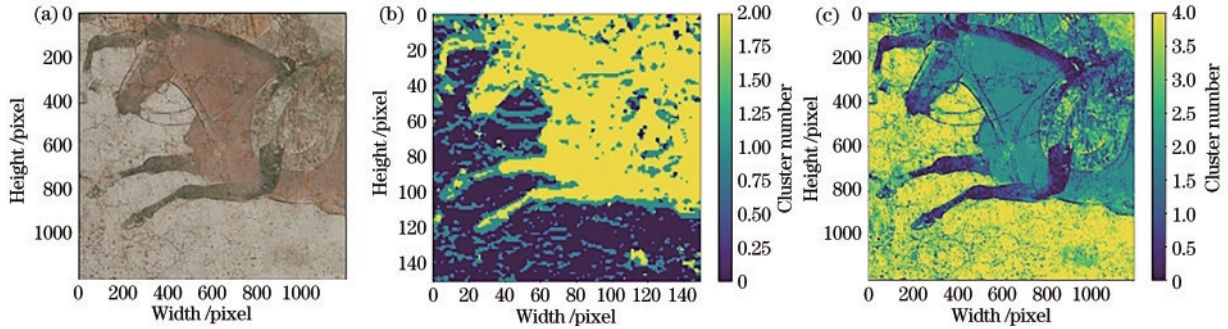


图 3 集群图对比。(a)输入图像;(b) VGG19空间;(c) RGB空间

Fig. 3 Cluster graph comparison. (a) Input image; (b) VGG19 space; (c) RGB space

K-means 需要指定聚类的个数作为超参数,本研究在确定该参数时采用以面积为基准、自动确定聚类个数的算法。目的是使各聚类的面积比例尽可能地分布在最小比例与最大比例的范围,算法流程如图 4

Algorithm 2 According to the area to determine the clustering number of K-means algorithm

Input: Input image, maximum number of clusters (K_{max}), maximum proportion (P_{max}), minimum proportion (P_{min})
Output: The cluster graph corresponding to the input image

- 1: K-means clustering is performed starting from $K = 2$
- 2: while the proportion of the largest cluster $\geq P_{max}$ or the proportion of the smallest cluster $\geq P_{min}$ do
- 3: if $K > K_{max}$ or the proportion of the second small cluster $\leq P_{min}$ then
- 4: return the cluster graph at $K - 1$
- 5: $K = K + 1$, clustering by K-means
- 6: return the cluster graph

图 4 算法流程

Fig. 4 Algorithm flow

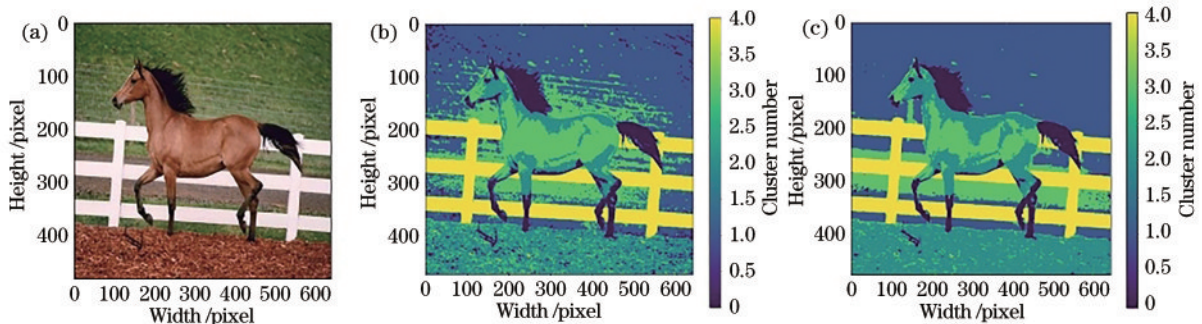


图 5 颜色空间对比。(a)输入图像;(b) RGB空间;(c) LAB空间

Fig. 5 Color space comparison. (a) Input image; (b) RGB space; (c) LAB space

3 所提方法

3.1 构建集群图

通过 K-means 可以对 VGG19^[10] 中 Conv4_1 层提取的图像特征进行聚类,构建出不同风格模式的集群图^[9]。然而该方式忽略了特征图是高维且低分辨率的,导致聚类图中的内容结构明显缺失。因此,本研究直接在输入图像的低维空间上构建包含结构信息的集群图。图 3 给出了 1 个示例,图 3(a)是风格图像,图 3(b)是在 VGG19 的 Conv4_1 下的聚类映射,图 3(c)是该图像在低维 RGB 空间的映射。结果表明,低维空间的聚类结果能够更好地保留输入图像的结构信息。

所示。该算法需要 3 个参数作为输入,分别为最大聚类数目 K_{max} (default 为 5)、最大比例 p_{max} (default 为 0.7) 以及最小比例 p_{min} (default 为 0.1)。

3.2 颜色空间转换

颜色空间的选择对图像处理的结果有很大的影响,因此选择正确的颜色空间对于这项任务至关重要。CIELAB 色彩空间是国际照明委员会(CIE)于 1931 年制定的比色标准,其主要优点之一是它是基于人类对色彩感知的均匀性而提出的,比 RGB 更适合人类视觉。本研究在同一图像的不同颜色空间上进行 K-means 特征聚类的结果如图 5 所示,结果表明,LAB 空间可以更准确地区分马匹与背景,描述图像的结构信息。

3.3 深度残差收缩网络

图 6 所示的残差收缩模块(RSB)为改进后的残差模块。深度残差收缩网络是深度残差网络、注意力机制和软阈值函数的集成,常用于加强深度神经网络从含噪声信号中提取有用特征的能力,消除图像的冗余信息^[11]。在使用 VGG19 提取图像特征之后,进一步通过 3 个残差收缩模块构成深度残差收缩网络(DRSN),自适应地对特征图进行噪声的滤除。此外,调整卷积、批处理归一化(BN)层和 ReLU 激活函数层的顺序。其中,BN 层和 ReLU 激活函数层位于卷积层之前,使得网络更容易训练^[12]。

该体系结构在经典残差网络的输出之前嵌入一个子网络,其工作原理为使用注意力机制识别到不重要的特征,并通过软阈值函数将其设为零^[11]:

$$y = \begin{cases} x - \tau, & x > \tau \\ 0, & -\tau \leq x \leq \tau \\ x + \tau, & x < -\tau \end{cases} \quad (1)$$

式中: x 表示输入特征; y 表示输出特征; τ 表示阈值。软阈值化将接近零的特征设置为零,将绝对值大于这个阈值的特征朝着零的方向收缩,过程如图 7(a)所示。图 7(b)为输出对输入的导数,可以看到,该软阈值化函数的导数不是 0 就是 1,可以很好地防止梯度消失和梯度爆炸问题:

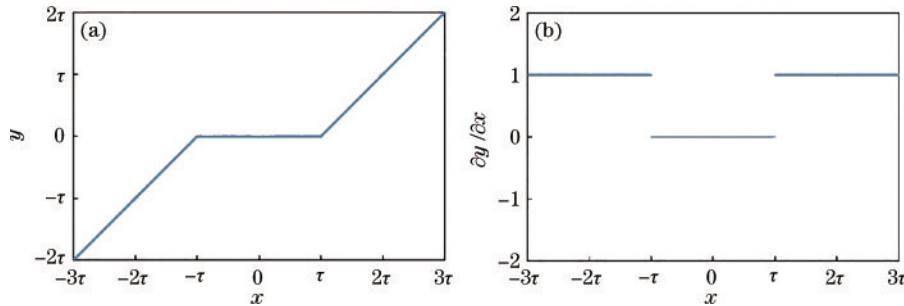


图 7 软阈值函数及其导数。(a)软阈值函数;(b)软阈值函数导数

Fig. 7 Soft threshold function and its derivative. (a) Soft threshold function; (b) soft threshold function derivative

在传统的去噪方法中,设置一个合适的阈值是困难的。针对这个问题,DRSN可以在深度结构中自动获取该值,以避免人工操作带来的错误。首先将输入的特征图取绝对值,然后应用全局平均池化(GAP)将特征降维成一维向量并输入一个小型的全连接网络中。这个全连接网络以 Sigmoid 函数作为最后一层,将输出归一化到 0~1 之间,获得特定通道的尺度值。Sigmoid 函数可表示为

$$\alpha = \frac{1}{1 + \exp(-z)}, \quad (3)$$

式中: z 为 RSB 中 1×1 卷积的输入; α 为输出的缩放参数。最后将其与输入特征图的特定通道平均值相乘,得到每一个通道的阈值 $\tau = \alpha \cdot \text{average}|x|$ 。这样不仅可以保证阈值为正,而且不会太大。

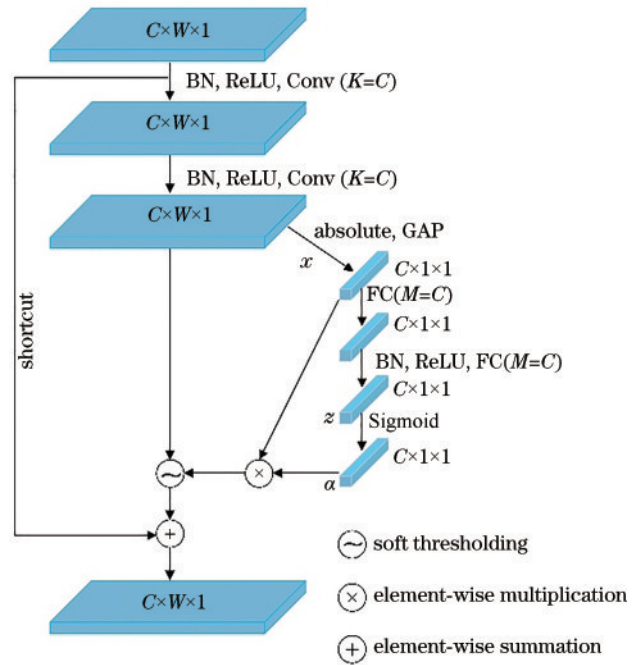


图 6 RSB 结构

Fig. 6 RSB structure

$$\frac{\partial y}{\partial x} = \begin{cases} 1, & x > \tau \\ 0, & |x| \leq \tau \\ -1, & x < -\tau \end{cases} \quad (2)$$

通过这种方式,每个样本都有着自己独特的阈值。因此,在一定程度上,该方式可以理解成一种特殊的注意力机制:结合样本自身特点,对特征通道进行加权调整,提取重要的特征。

3.4 特征转换

对于特征转换,前提是内容集群图与风格集群图的匹配。假设得到的内容集群图和风格集群图分别为 M_c 和 M_s 。接下来,以内容集群中心的范数为中心,如果在覆盖比例 $[0, 1]$ 内存在风格集群中心的范数,则在对应的集群之间建立匹配关系,最终得到内容图像与风格图像的匹配集合为 $\{F_c^k, F_s^k (k = 1, \dots, K)\}$ 。

根据得到的内容和风格图像匹配集,使用每个集合中内容特征的均值和标准差将其转换为对应的风格特征。该过程由 AdaIN^[7]实现,其生成的目标特征映

射 F_{cs}^k 为

$$F_{cs}^k = \text{AdaIN}(F_c^k, F_s^k), \quad (4)$$

$$\text{AdaIN}(x, y) = \sigma(y) \cdot \frac{x - \mu(x)}{\sigma(x)} + \mu(y), \quad (5)$$

式中: F_c^k 和 F_s^k 分别是匹配的内容和风格特征; $\mu(x)$ 和 $\sigma(x)$ 分别代表 x 的均值和标准差, 对于 y , 同样成立。因此, 式(5)可以理解为先将原内容图像减去自身均值再除以自身标准差, 再乘上风格图像的标准差并加上均值, 最后使得内容图像和风格图像的风格一致。

经过风格转换, 将 F_{cs}^k 与内容特征 F_c^k 按照式(6)混合后再输入至解码器。

$$F_{cs}^k = \alpha_k F_{cs}^k + (1 - \alpha_k) F_c^k, \quad (6)$$

式中: 控制融合比例的超参数 $\alpha_k \in [0, 1]$, 默认设定为 1。混合所有特征后, 得到最终生成图像的特征, 最后将 F_{cs} 输入解码器重构输出生成图像。

$$F_{cs} = F_{cs}^1 \cup F_{cs}^2 \cup \dots \cup F_{cs}^K. \quad (7)$$

3.5 感知损失

图 1 中解码器设计成与编码器对称的结构。与文献[6, 7]类似, 使用预训练的 VGG19 计算损失函数来训练解码器:

$$L_{\text{total}} = L_{\text{content}} + L_{\text{style}}, \quad (8)$$

式中: L_{content} 为内容损失; L_{style} 为风格损失。文献[5, 6]研究表明使用网络较高层提取的特征重构图像时, 可以捕获图像的内容和整体空间结构, 但对颜色、纹理和确切的形状没有太大约束, 而对于较低层特征的重建只是复制了原始图像的像素值。因此, 网络高层中的特征响应被称为内容表示。提取 VGG19 中 Conv_4_1 层的特征来计算内容损失 L_{content} :

$$L_{\text{content}} = \|\phi_{4_1}(I_c) - \phi_{4_1}(I_{cs})\|_2, \quad (9)$$

式中: $\phi_{4_1}(\cdot)$ 提取 VGG19 中 Conv_4_1 层的特征; I_c 和 I_{cs} 分别为内容图像和生成图像。

风格损失 L_{style} 主要衡量风格化后的图像 I_{cs} 和风格图 I_s 在风格上的相近程度, 因为 AdaIN(式 5) 特征转

换只传递了风格特征的平均值和标准差, 所以风格损失只匹配这些统计数据, 通过计算均值、标准差之间欧氏距离的和得到 L_{style} :

$$L_{\text{style}} = \sum_{i=1}^4 \left\{ \left\| \mu[\phi_{i,1}(I_s)] - \mu[\phi_{i,1}(I_{cs})] \right\|_2 \right\} + \sum_{i=1}^4 \left\{ \left\| \sigma[\phi_{i,1}(I_s)] - \sigma[\phi_{i,1}(I_{cs})] \right\|_2 \right\}, \quad (10)$$

式中: $\phi_{i,1}(\cdot)$ 提取 VGG19 中 Conv_i_1 层的特征; 利用 $\mu(\cdot)$ 和 $\sigma(\cdot)$ 计算内容特征和风格特征的均值和标准差。

4 实验结果

4.1 实验设置与数据集

使用 PyTorch 框架构建网络模型, 实验设备为搭载 Windows 10 系统的台式计算机, CPU 为 Intel I9 9900K 5.0 GHz, RAM 为 64 GB, 使用 NVIDIA RTX 2080Ti 进行 GPU 加速。在训练过程中将图像尺寸调整为 512×512 作为输入, 使用 Adam^[13] 作为优化函数, 学习率为 1×10^{-5} 。

实验采用 Microsoft COCO^[14] 作为内容图像数据集, 唐章怀太子墓的马球图作为风格图像数据集训练网络模型。唐章怀太子墓马球图使用瑞士的 Sinar P2 型大画幅技术相机和 Sinar 75LV 数字后背进行数据采集。将采集到的高清壁画分割成 $256 \text{ pixel} \times 256 \text{ pixel}$ 的壁画分块, 进行数据的预处理并筛选出 744 张纹理细节较为丰富的壁画样本作为风格数据集。

4.2 消融实验

4.2.1 卷积神经网络特征重构

为了可视化 VGG19 网络每一层包含的关于输入图像的信息, 采用文献[15]的方法从 VGG19 不同层中提取特征并重构图像。如图 8 所示, 由深层次提取出的特征偏向于表现图像结构信息, 而没有过多保留颜色及纹理。因此, 提取 VGG19 中 Conv_4_1 层的特征来计算内容损失, 最终使输出图像在感知上与目标图像相似, 但不会强迫它们完全匹配。

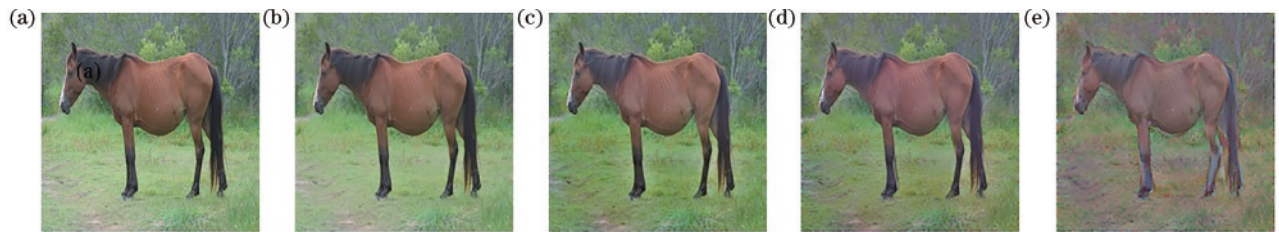


图 8 不同网络层次的图像表示。(a)原始图像;(b) Conv_1_1;(c) Conv_2_1;(d) Conv_3_1;(e) Conv_4_1

Fig. 8 Image representation of different network levels. (a) Original image; (b) Conv_1_1; (c) Conv_2_1; (d) Conv_3_1; (e) Conv_4_1

4.2.2 RSB 有效性

将所提方法与不使用 RSB 的方法应用于一组内容图像和壁画风格图像中, 并采用 class activation mapping(CAM)^[16] 可视化获得生成图像的注意力热力图。图 9(a) 显示了原始内容图像和相应的 CAM 图,

图 9(b)、(c) 显示了添加模块前后生成的图像和 CAM 比较结果。相比之下, 图 9(c) 的 CAM 与内容图像的 CAM 更相似, 即 RSB 可以针对内容图像的重要特征进行转换, 保留内容图像的更多细节纹理, 最终生成更高质量的风格化图像。

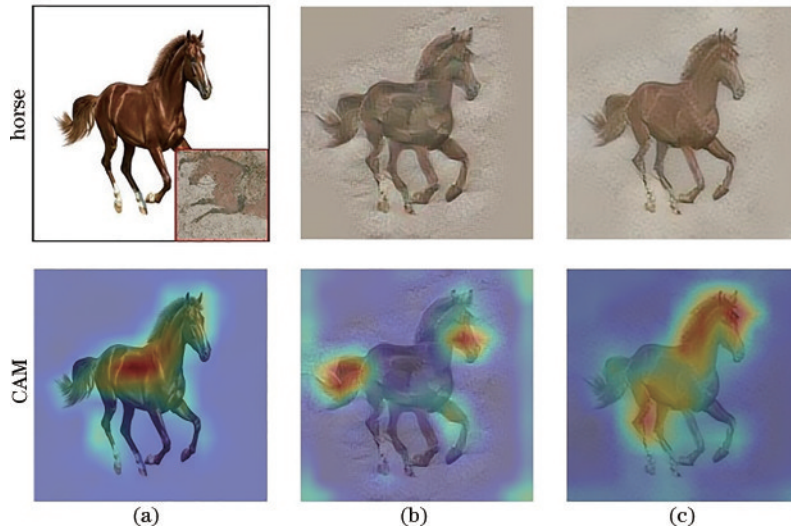


图 9 RSB 有效性验证。(a)内容图像;(b)不使用RSB;(c)使用RSB

Fig. 9 RSB validity verification. (a) Content image; (b) without RSB; (c) with RSB

4.2.3 聚类数目

根据 Algorithm 2 可知,聚类数目是由最大聚类数目、最大比例以及最小比例共同控制的。为了探究不同的聚类数目对风格迁移效果的影响,将输入图像直接划分成指定的聚类个数,而不是观察上述 3 个超参数对聚类个数的影响。图 10 中,随着聚类数目的增加,迁移后图像的颜色分布会发生变化。当 $K=1$ 时,相当于对真实的马匹图像和壁画图像的特征进行全局

处理,忽略了输入图像的结构特征,缺乏灵活性,图像颜色不鲜明。当 $K=3$ 时,能够实现真实图像和壁画图像相似特征的匹配,提升了生成图像色彩、纹理等特征,增强了图像对比度。当 $K=5$ 时,匹配效果比 $K=3$ 时更好,图中马匹不再受背景色彩的影响,更好地保留了输入图像的色彩特征,进一步提升了迁移效果。当 K 值从 5 增加到 10 时,迁移后的图像几乎没有差异,即不再随聚类数量的增加而产生更多的颜色变化。

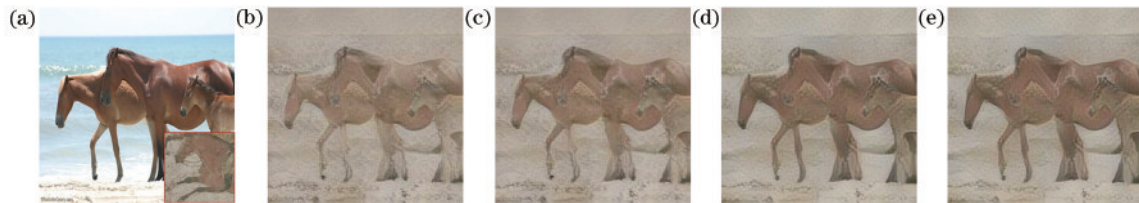


图 10 不同聚类数目风格迁移效果。(a)输入图像;(b) $K=1$;(c) $K=3$;(d) $K=5$;(e) $K=10$

Fig. 10 The effect of style transfer with different cluster numbers. (a) Input image; (b) $K=1$; (c) $K=3$; (d) $K=5$; (e) $K=10$

4.2.4 融合系数

融合系数 α_k 是控制生成图像与内容图像融合程度的参数。本小节研究了融合系数 α_k 对风格化的影

响,结果如图 11 所示。可以看出, α_k 值越大,生成的图像融合效果越好,整体质量越高。为了得到更好的样式迁移效果,最终设置 $\alpha_k=1.0$ 。

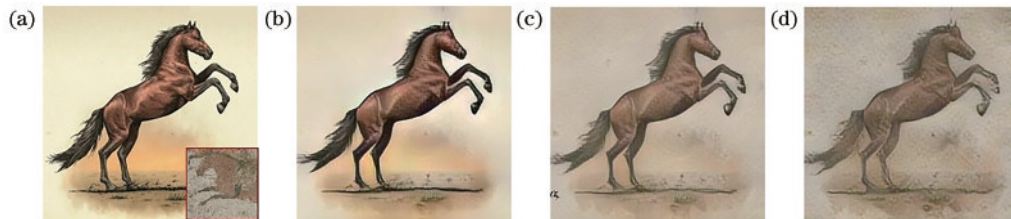


图 11 不同融合系数风格迁移效果。(a)输入图像;(b) $\alpha_k=0.2$;(c) $\alpha_k=0.6$;(d) $\alpha_k=1.0$

Fig. 11 The effect of style transfer with different fusion coefficients. (a) Input image; (b) $\alpha_k=0.2$; (c) $\alpha_k=0.6$; (d) $\alpha_k=1.0$

4.2.5 覆盖比例

覆盖比例是集群图匹配过程中控制内容和风格聚类之间对应程度的超参数。随着值的增加,内容集群将会匹配更多的风格集群,即在风格转换过程中可以

混合更多的风格,因此该值可以根据风格图像的色彩丰富程度来确定。图 12 中,由于壁画的背景颜色较单一,随着覆盖比例的增加,人与背景的融合变得不那么生动。因此,将覆盖比例设置为 0.25。

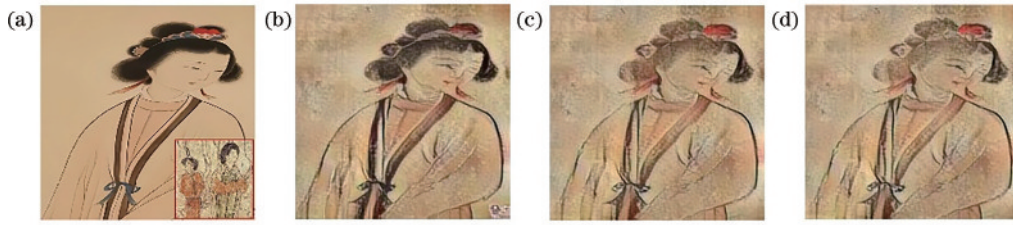


图 12 不同覆盖比例风格迁移效果。(a)输入图像;(b)覆盖比例为 0.25;(c)覆盖比例为 0.5;(d)覆盖比例为 1.0

Fig. 12 The effect of style transfer with different coverage ratios. (a) Input image; (b) coverage ratio is 0.25; (c) coverage ratio is 0.5; (d) coverage ratio is 1.0

4.3 实验结果与分析

4.3.1 定性分析

为了评估所提改进模型的有效性,使用相同的数

据集进行训练和测试。图 13 为所提方法与其他风格迁移方法的效果对比,包括 Gatys 等^[5]的方法、AdaIN^[7]、SANet^[8]和 MST^[9]。



图 13 不同风格迁移方法效果对比。(a)内容图像;(b)风格图像;(c)文献[5]方法;(d) AdaIN;(e) SANet;(f) MST;(g)所提方法
Fig. 13 Comparison of different style transfer methods. (a) Content images; (b) style images; (c) method of reference [5]; (d) AdaIN; (e) SANet; (f) MST; (g) proposed method

从图 13 的对比结果可以看出,Gatys 等^[5]的方法使用迭代图像优化获得了较清晰的转换结果,但由于时间开销大,不适合批量数据扩充。AdaIN^[7]忽略了内容和风格之间的局部相关性,这可能会在图像边缘产生一些光晕,影响视觉效果[如图 13(d)的 picture1]。SANet^[8]采用了注意力机制,减少了纹理的重复,但受风格图像的过分影响,背景的噪声较明显[图 13(e)的 picture3]。MST^[9]虽然遵循了风格特征的分布,但它在图像的高维空间进行特征聚类,扭曲了内容图像的结构,生成的图像产生明显形变。相比之下,所提方法取得了更好的效果。通过在低维空间进行聚类,强化了特征之间相似性的联系,减少了点状伪影现象。此外,引入的残差收缩模块关注图像中显著的特征,减少

了噪声等冗余信息,提升了生成图像的质量。

4.3.2 定量分析

采用结构相似性(M_{SSIM})、峰值信噪比(R_{PSNR})以及均方差(E_{MSE})等 3 种指标^[17-19]对风格迁移模型生成的风格化图像质量进行定量分析。

M_{SSIM} 一般从相关性损失、亮度、对比度等方面衡量两幅图像的相似程度。其值越接近于 1,代表两张图像越相似,即迁移过程中内容结构扭曲程度越小:

$$M_{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (11)$$

式中: μ_x 、 μ_y 分别表示图像 x 和 y 的均值; σ_x^2 、 σ_y^2 分别表示图像 x 和 y 的方差; σ_{xy} 代表图像 x 和 y 协方差; c_1 、 c_2

是维持稳定的常数。

E_{MSE} 是参数估计值与参数真值之差平方的期望值,在风格迁移算法中指的是迁移后图像像素值与原始像素值之差平方和的均值:

$$E_{MSE}(f, g) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [f(i, j) - g(i, j)]^2. \quad (12)$$

R_{PSNR} 通过 E_{MSE} 度量图像相对于生成图像的质量。PSNR 值越高,模型生成的图像质量越可接受:

$$R_{PSNR}(f, g) = 10 \lg \left[\frac{255^2}{E_{MSE}(f, g)} \right]. \quad (13)$$

分别用 Gatys 等^[5]的方法、AdaIN^[7]、SANet^[8]和 MST^[9]以及所提方法对图 13 中的 4 幅内容图像进行风格化测试,并对每次实验结果的 M_{SSIM} 、 R_{PSNR} 以及 E_{MSE} 进行统计,结果如表 1 所示。

表 1 风格化测试定量统计结果

Table 1 Quantitative statistical results of stylized test

Image	Method	M_{SSIM}	R_{PSNR}	$E_{MSE}/10^3$
picture1	Method of reference [5]	0.56	9.30	7.65
	AdaIN	0.58	9.03	8.13
	SANet	0.52	7.38	7.15
	MST	0.45	9.20	7.82
	Proposed method	0.64	11.25	4.88
picture2	Method of reference [5]	0.48	12.81	3.40
	AdaIN	0.51	13.19	3.12
	SANet	0.37	11.88	4.21
	MST	0.39	11.62	4.48
	Proposed method	0.53	14.69	2.21
picture3	Method of reference [5]	0.53	15.64	1.76
	AdaIN	0.48	15.25	1.94
	SANet	0.34	14.36	2.38
	MST	0.34	13.67	2.79
	Proposed method	0.55	16.10	1.59
picture4	Method of reference [5]	0.43	15.78	1.24
	AdaIN	0.38	17.85	1.15
	SANet	0.28	16.45	1.47
	MST	0.34	16.71	1.38
	Proposed method	0.45	18.25	1.13

根据表 1 中的数据可知,在 4 幅内容图像的测试中,所提方法在 3 个指标上都取得了较好的结果,进一步证明了其有效性。一方面由于在低维空间进行聚类可以保留结构的完整性,另一方面引入的深度残差收缩网络强化了图像中细节纹理等重要信息并弱化了噪声等无关信息。因此,生成图像的质量有所提升,风格迁移效果较好。

4.3.3 速度比较

随机选择 5 张内容图像、5 张风格图像,使用不同方法进行风格迁移,结果如表 2 所示。

表 2 不同方法的平均生成时间比较

Table 2 Comparison of average generation time of different methods

Method	Running time /s
AdaIN	0.40
SANet	0.89
MST	4.77
Proposed method	1.65

由表 2 可得,所提方法的平均风格化时间在 2 s 以内,可以实现实时风格转换。由于多模态方法 MST 在 512 维的高维空间上进行聚类,因此计算时间接近 5 s。所提方法在三维 CIELAB 颜色空间中进行聚类,虽然计算时间相较于 AdaIN、SANet 略有增加,但是只要加上特征聚类,就可以有效地转换风格,提升迁移图像的质量。

4.4 数据增强性能

为了进一步验证所提数据扩充方法的有效性,设置了以 Outpainting 为核心的墓室壁画分块外延信息数字生成对比实验,实验方案是使用训练后的生成对抗网络模型^[20]扩展掩码图像的左右方向,以生成掩码图像中缺失的壁画信息。一共制备了 4 个数据集:744 张真实预修复壁画图像^[1],记为 real;5208 张所提方法生成图像,记为 synthetic;real+synthetic 共 5952 张图像;real+s. a. (传统增广方法)共 5952 张图像。其中,传统增广方法包括水平翻转、旋转 90°、旋转 180°、旋转 270°、对比度、颜色增强、亮度增强等 7 种方法。每种方法分别对 744 张原始数据集进行变换,不同传统数据增广方法的对比结果如图 14 所示。

壁画外延信息生成的结果如图 15 所示。其中,real 数据集训练的生成模型存在明显的纹理缺陷,如图 15(c)所示。对于 real+s. a. 数据集,经过传统数据增广的生成模型的纹理生成能力明显增强,但在掩码位置仍然存在伪影残留,如图 15(d)所示。经过所提风格迁移方法扩充数据集后,纹理生成能力得到一定程度的提升。相比之下,在 real+synthetic 上训练的生成模型性能明显提高,能够学习更多的纹理细节信息,提高纹理生成精度并平滑外延边界,如图 15(f)所示。

此外,将所提改进的风格迁移方法与其他风格迁移技术构成的数据集进行了比较。每个方法都合成了 5208 张图像,在数字生成实验上进行测试。为客观评价生成效果,采用 M_{SSIM} 、 R_{PSNR} 以及 E_{MSE} 作为评价指标,结果如表 3 所示。

由表 3 可得,所提方法较其他风格迁移方法的 M_{SSIM} 提高了 0.03~0.08, R_{PSNR} 提高了 0.3~1.48, E_{MSE} 降低了 $0.37 \times 10^3 \sim 1.21 \times 10^3$ 。该结果验证了所提风格迁移方法数据增强性能的有效性。

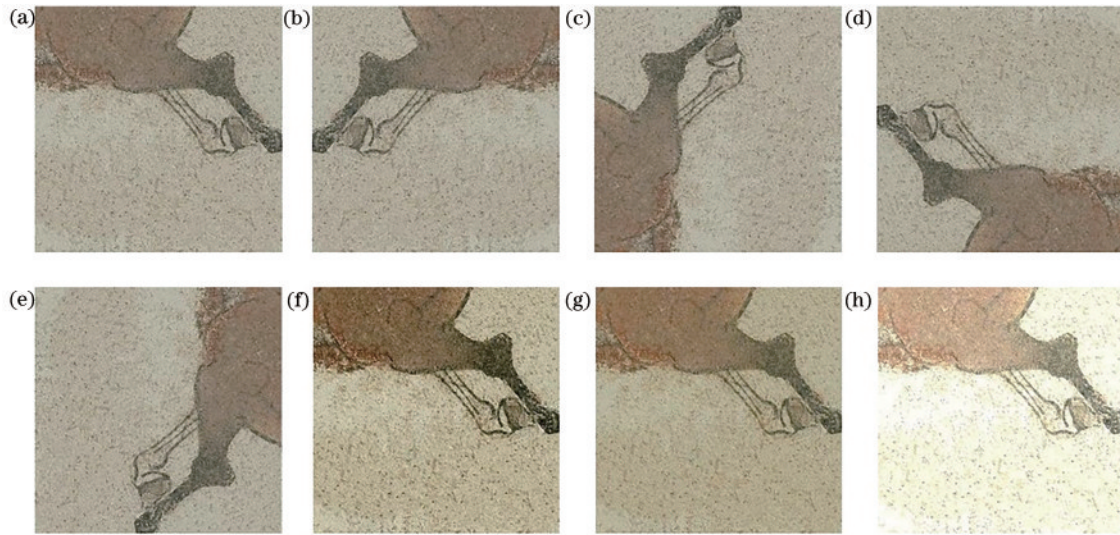


图 14 不同数据增广方法对比结果。(a)原始图像;(b)水平翻转;(c)旋转 90°;(d)旋转 180°;(e)旋转 270°;(f)对比度;(g)颜色增强;(h)亮度增强

Fig. 14 Comparison results of different data augmentation methods. (a) Original image; (b) flip horizontal; (c) rotate 90°; (d) rotate 180°; (e) rotate 270°; (f) contrast; (g) color enhancement; (h) brightness enhancement

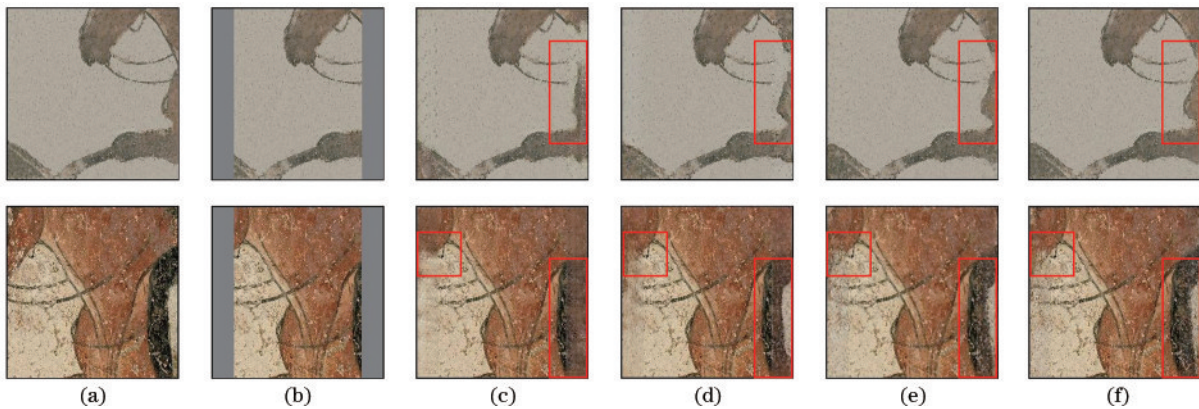


图 15 不同数据集生成效果。(a)原始壁画;(b)掩码图像;(c) real;(d) real+s. a.; (e) synthetic; (f) real+synthetic

Fig. 15 Effect of different dataset generation. (a) Original murals; (b) masked images; (c) real; (d) real+s. a.; (e) synthetic; (f) real+synthetic

表 3 壁画数字生成结果

Table 3 Mural digital generation results

Method	M_{SSIM}	R_{PSNR}	$E_{MSE} / 10^3$
AdaIN	0.43	18.76	2.68
SANet	0.41	18.22	2.94
MST	0.38	17.78	3.52
Proposed method	0.49	19.26	2.31

5 结 论

针对壁画数字生成研究中数据集欠缺的问题,提出一种基于风格迁移的样本扩充方法。相比常见的风格迁移方法,该方法解决了风格转换效果差、生成图像质量低的问题。通过在图像的低维空间上使用聚类匹配实现壁画特征的多模态分布,有效地获取了风格特征。同时,残差收缩模块的引入去除了图像的噪声,提高了生成图像的质量。结果表明,所构建模型生成的

图像在主观感知和客观评价方面都有更好的表现。尽管缺乏真实的壁画图像,使用风格迁移生成的图像作为训练集仍然可以提高壁画数字生成模型的性能,进一步扩展了模拟图像在深度学习中的潜在用途。所提算法的目的是在不改变内容图像形状的情况下将壁画的风格转移到内容图像中。接下来会进一步改进算法,使其能够迁移壁画中物体的姿态等特征。

参 考 文 献

- [1] 申婧妮,王慧琴,吴萌,等. MCA 分解的唐墓室壁画修复算法[J]. 计算机科学与探索, 2017, 11(11): 1826-1836.
Shen J N, Wang H Q, Wu M, et al. Tang dynasty tomb murals inpainting algorithm of MCA decomposition[J]. Journal of Frontiers of Computer Science and Technology, 2017, 11(11): 1826-1836.
- [2] Raitoharju J, Riabchenko E, Meissner K, et al. Data

- enrichment in fine-grained classification of aquatic macroinvertebrates[C]//2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI), December 4, 2016, Cancun, Mexico. New York: IEEE Press, 2016: 43-48.
- [3] Xiao Q, Liu B, Li Z, et al. Progressive data augmentation method for remote sensing ship image classification based on imaging simulation system and neural style transfer[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 9176-9186.
- [4] Zhang Y P, Wang Q, Hu B L. MinimalGAN: diverse medical image synthesis for data augmentation using minimal training data[J]. *Applied Intelligence*, 2022: 1-18.
- [5] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2414-2423.
- [6] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9906: 694-711.
- [7] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 1510-1519.
- [8] Park D Y, Lee K H. Arbitrary style transfer with style-attentional networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 5873-5881.
- [9] Zhang Y L, Fang C, Wang Y L, et al. Multimodal style transfer via graph cuts[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 5942-5950.
- [10] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2022-08-06]. <https://arxiv.org/abs/1409.1556>.
- [11] Zhao M H, Zhong S S, Fu X Y, et al. Deep residual shrinkage networks for fault diagnosis[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(7): 4681-4690.
- [12] He K M, Zhang X Y, Ren S Q, et al. Identity mappings in deep residual networks[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9908: 630-645.
- [13] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22)[2022-08-06]. <https://arxiv.org/abs/1412.6980>.
- [14] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. *Computer vision-ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8693: 740-755.
- [15] Mahendran A, Vedaldi A. Understanding deep image representations by inverting them[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 5188-5196.
- [16] Zhou B L, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2921-2929.
- [17] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training GANs[C]//NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. New York: ACM Press, 2016: 2234-2242.
- [18] Che T, Li Y R, Jacob A P, et al. Mode regularized generative adversarial networks[EB/OL]. (2016-12-07)[2022-08-06]. <https://arxiv.org/abs/1612.02136>.
- [19] Heusel M, Ramsauer H, Unterthiner T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. New York: ACM Press, 2017: 6629-6640.
- [20] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[EB/OL]. (2014-06-10)[2022-05-06]. <https://arxiv.org/abs/1406.2661>.