



(12) 发明专利

(10) 授权公告号 CN 119251069 B

(45) 授权公告日 2025. 04. 15

(21) 申请号 202411769124.6

(22) 申请日 2024.12.04

(65) 同一申请的已公布的文献号

申请公布号 CN 119251069 A

(43) 申请公布日 2025.01.03

(73) 专利权人 浙江大学计算机创新技术研究院

地址 311200 浙江省杭州市萧山区利一路

188号天人大厦33楼

(72) 发明人 夏瑞晨 张占杰 方科彬 罗尚虎

罗大伟 赵瑶 祝栗

(74) 专利代理机构 杭州求是专利事务有限公

司 33200

专利代理师 林超

(51) Int. Cl.

G06T 5/50 (2006.01)

G06T 5/60 (2024.01)

G06V 10/774 (2022.01)

G06V 10/80 (2022.01)

G06V 10/82 (2022.01)

(56) 对比文件

CN 115995035 A, 2023.04.21

CN 118014822 A, 2024.05.10

审查员 王璐

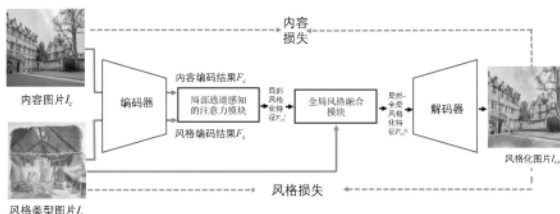
权利要求书2页 说明书8页 附图3页

(54) 发明名称

基于局部和全局风格学习的图片高质量任意风格迁移方法

(57) 摘要

本发明公开了一种基于局部和全局风格学习的图片高质量任意风格迁移方法。方法包括：获取若干内容图片和不同风格类型图片并构建为训练集；建立包含局部通道感知的注意力模块和全局风格融合模块的风格迁移网络，并构建基于补丁的风格迁移损失函数；将训练集输入风格迁移网络中进行训练直至基于补丁的风格迁移损失函数收敛；将待风格迁移的内容图片和目标风格类型图片输入训练完成的风格迁移网络中进行处理后输出风格化图片并在显示器进行显示，实现图片高质量风格迁移。本发明方法能够更好的从局部风格模式和全局风格分布的角度学习，进行任意风格迁移的同时生成高质量的风格化图片。



1.一种基于局部和全局风格学习的图片高质量任意风格迁移方法,其特征在于,包括:

S1:获取若干内容图片和不同风格类型图片并共同构建为训练集;

S2:建立基于局部和全局风格学习的风格迁移网络,风格迁移网络中包含局部通道感知的注意力模块LCWA和全局风格融合模块GSFT,同时构建风格迁移网络的基于补丁的风格迁移损失函数;将训练集输入风格迁移网络中进行训练,直至基于补丁的风格迁移损失函数收敛,获得训练完成的风格迁移网络;

S3:将待风格迁移的内容图片和目标风格类型图片共同输入至训练完成的风格迁移网络中进行处理,处理完成后输出风格化图片并在显示器进行显示,实现图片高质量风格迁移;

所述的步骤S2中,基于局部和全局风格学习的风格迁移网络包括依次连接的视觉VGG模型的编码器、局部通道感知的注意力模块LCWA、全局风格融合模块GSFT和视觉VGG模型的解码器,风格迁移网络的输入为一张内容图片 I_c 和一张风格类型图片 I_s ,依次经过编码器和局部通道感知的注意力模块LCWA处理后获得局部风格化特征 F_{cs}^l ,将局部风格化特征 F_{cs}^l 和风格类型图片 I_s 共同输入至全局风格融合模块GSFT中进行处理,依次经全局风格融合模块GSFT和解码器处理后将风格类型图片的风格转移至内容图片,输出最终的风格化图片 I_{cs} ;

所述的局部通道感知的注意力模块LCWA包括三个归一化层Norm、三个卷积层conv和Softmax激活函数,内容图片 I_c 和风格类型图片 I_s 经处理后获得内容编码结果 F_c 和风格编码结果 F_s ,将内容编码结果 F_c 分别输入至第一和第三归一化层Norm中进行处理,将风格编码结果 F_s 分别输入至第二归一化层Norm和第三卷积层conv中进行处理,第一和第二归一化层Norm处理后的结果分别输入至第一和第二卷积层conv中进行处理,将第一和第二卷积层conv处理后的结果相乘后输入Softmax激活函数中处理后输出注意力图A,第三卷积层conv处理后的结果V和注意力图A相乘后获得第一相乘结果,第三卷积层conv处理后的结果V的平方 V^2 和注意力图A相乘后获得第二相乘结果,第二相乘结果和第一相乘结果的均值M的平方 M^2 相减后再开根号获得第一相乘结果的标准差S,第三归一化层Norm处理后的结果和第一相乘结果的标准差S点乘后再和第一相乘结果的均值M相加,最终输出局部风格化特征 F_{cs}^l 作为局部通道感知的注意力模块LCWA的输出。

2.根据权利要求1所述的基于局部和全局风格学习的图片高质量任意风格迁移方法,其特征在于:所述的全局风格融合模块GSFT为基于深度神经网络Transformer的模块,全局风格融合模块GSFT的输入为风格类型图片 I_s 以及由局部通道感知的注意力模块LCWA处理后获得的局部风格化特征 F_{cs}^l ,风格类型图片 I_s 输入深度神经网络Transformer中进行处理,首先将风格类型图片 I_s 划分为若干个块Patch并构成风格序列 Z_s ,然后根据预定义的深度神经网络Transformer的索引Q、键K和值V对风格序列 Z_s 进行编码,然后再输入第一多头注意力模块中进行处理后输出第一注意力结果 F_s^g ,将第一注意力结果 F_s^g 依次输入第四归一化层Norm和第一前馈神经网络FFN中处理后再经过键K和值V进行编码,然后再输入第二多头注意力模块中进行处理,第二多头注意力模块处理后的结果和经索引Q编码后的局部风格化特征 F_{cs}^l 共同依次进行残差连接Add和第五归一化层Norm处理后分别输入第二和第三前馈神经网络FFN中进行处理后分别输出偏置 b_s 和权重 w_s 作为深度神经网络Transformer的输出结果,权重 w_s 和局部风格化特征 F_{cs}^l 进行点乘后再与偏置 b_s 相加后输出局部-全局风格化特征 F_{cs}^{lg} 作为全局风格融合模块GSFT的输出。

3. 根据权利要求1所述的基于局部和全局风格学习的图片高质量任意风格迁移方法, 其特征在于: 所述的基于补丁的风格迁移损失函数包括视觉VGG模型的内容损失和风格损失, 并构建基于补丁的Gram损失算法加入风格损失中从而建立改进风格损失, 使用内容损失对内容图片 I_c 和风格化图片 I_{cs} 的内容相似度进行约束, 使用改进风格损失对风格类型图片 I_s 和风格化图片 I_{cs} 的风格相似度进行约束。

4. 根据权利要求3所述的基于局部和全局风格学习的图片高质量任意风格迁移方法, 其特征在于: 所述的基于补丁的Gram损失算法包括补丁操作和Gram矩阵损失函数 L_{ls} , 补丁操作过程中首先将风格类型图片 I_s 和风格化图片 I_{cs} 均进行随机裁剪, 然后进行图像增强操作后获得若干风格类型补丁和风格化补丁, 根据各个风格类型补丁和风格化补丁输入Gram矩阵损失函数 L_{ls} 中获得Gram矩阵损失值以进行风格迁移网络的训练。

5. 根据权利要求4所述的基于局部和全局风格学习的图片高质量任意风格迁移方法, 其特征在于: 所述的Gram矩阵损失函数 L_{ls} 如下:

$$L_{ls} = (\sum_i^N R(l_{patch}^i, C(\hat{I}_s^i, I_s), \tau)) / N$$

$$R(l_{patch}^i, C(\hat{I}_s^i, I_s), \tau) = l_{patch}^i, \text{ 当 } C(\hat{I}_s^i, I_s) \geq \tau$$

$$R(l_{patch}^i, C(\hat{I}_s^i, I_s), \tau) = 0, \text{ 其它}$$

$$C(\hat{I}_s^i, I_s) = G(E_I(\hat{I}_s^i)) \cdot G(E_I(I_s)) / (|G(E_I(\hat{I}_s^i))| \cdot |G(E_I(I_s))|)$$

其中, N 表示补丁的总数量; $R()$ 和 τ 分别表示阈值函数及其函数参数; l_{patch}^i 表示第 i 个风格类型补丁和风格化补丁的Gram矩阵的余弦相似度; $C()$ 表示两个张量之间的余弦相似度; $\hat{I}_s^i$ 表示风格类型图片 I_s 的第 i 个风格类型补丁; $G()$ 表示裁剪CLIP空间中的第三层的Gram矩阵; $E_I()$ 表示裁剪CLIP空间的图像编码器。

6. 根据权利要求5所述的基于局部和全局风格学习的图片高质量任意风格迁移方法, 其特征在于: 所述的第 i 个风格类型补丁和风格化补丁的Gram矩阵的余弦相似度 l_{patch}^i 如下:

$$l_{patch}^i = \|G(E_I(aug(\hat{I}_{cs}^i))) - G(E_I(aug(\hat{I}_s^i)))\|_2$$

其中, $\hat{I}_{cs}^i$ 表示风格化图片 I_{cs} 的第 i 个风格化补丁; $aug()$ 表示随机几何增强; $\|\cdot\|_2$ 表示二范数。

7. 一种电子设备, 其特征在于, 包括: 相互耦接的存储器和处理器, 其中, 所述存储器存储有程序数据, 所述处理器调用所述程序数据以执行如权利要求1-6中任一项所述的方法。

8. 一种计算机可读存储介质, 其上存储有程序数据, 其特征在于, 所述程序数据被处理器执行时实现如权利要求1-6中任一项所述的方法。

基于局部和全局风格学习的图片高质量任意风格迁移方法

技术领域

[0001] 本发明涉及了一种图片高质量任意风格迁移方法,涉及计算机视觉和深度学习领域,具体涉及一种基于局部和全局风格学习的图片高质量任意风格迁移方法。

背景技术

[0002] 风格迁移可以在给定任意的内容和风格形象的情况下创造出新的艺术形象。现有的风格迁移方法主要可分为:局部补丁交换方法和全局样式转换方法。现有的一些方法提出在特征空间中交换局部的内容补丁和其最相似的风格补丁,例如使用注意力的方式将风格特征融入到内容特征中,并且提出了在浅层空间和深层空间计算基于注意力的归一化方法,然而,基于局部补丁交换方法可能会不可避免地从风格图片中引入一些语义信息,并且风格化图片的全局风格分布可能会偏离风格图片。基于全局样式转换方法主要在于从欧几里得距离调整内容特征的风格信息去匹配风格图片,如自适应实例归一化的方法去调整内容图片的风格分布,以及通过白化-上色的过程将风格特征迁移到内容特征上,然而简单的使用全局样式转换方法不能学习风格图片的局部细节。

发明内容

[0003] 为了解决背景技术中存在的问题,本发明所提供的一种基于局部和全局风格学习的图片高质量任意风格迁移方法。本发明要解决的问题是如何捕获局部细节纹理,避免从样式图像中引入不需要的语义信息,平衡局部样式模式和全局样式分布,可以同时考虑局部风格模式和全局风格分布,从而生成高质量的风格化图像。

[0004] 本发明采用的技术方案是:

[0005] 本发明的基于局部和全局风格学习的图片高质量任意风格迁移方法,包括:

[0006] S1:获取若干内容图片和不同风格类型图片并共同构建为训练集。

[0007] S2:建立基于局部和全局风格学习的风格迁移网络,风格迁移网络中包含局部通道感知的注意力模块LCWA(Local Channel-Wise Attention)和全局风格融合模块GSFT(Global Style Fusion Transformer),同时构建风格迁移网络的基于补丁的风格迁移损失函数;将训练集输入风格迁移网络中进行训练,直至基于补丁的风格迁移损失函数收敛,获得训练完成的风格迁移网络。

[0008] S3:将待风格迁移的内容图片和目标风格类型图片共同输入至训练完成的风格迁移网络中进行处理,处理完成后输出风格化图片并在显示器进行显示,实现图片高质量风格迁移。

[0009] 所述的步骤S2中,基于局部和全局风格学习的风格迁移网络包括依次连接的视觉VGG(Visual Geometry Group)模型的编码器、局部通道感知的注意力模块LCWA、全局风格融合模块GSFT和视觉VGG模型的解码器,风格迁移网络的输入为一张内容图片 I_c 和一张风格类型图片 I_s ,依次经过编码器和局部通道感知的注意力模块LCWA处理后获得局部风格化特征 F_{cs}^l ,将局部风格化特征 F_{cs}^l 和风格类型图片 I_s 共同输入至全局风格融合模块GSFT中进

行处理,依次经全局风格融合模块GSFT和解码器处理后将风格类型图片的风格转移至内容图片,输出最终的风格化图片 I_{cs} 。

[0010] 所述的局部通道感知的注意力模块LCWA包括三个归一化层Norm、三个卷积层conv和Softmax激活函数,内容图片 I_c 和风格类型图片 I_s 经处理后获得内容编码结果 F_c 和风格编码结果 F_s ,将内容编码结果 F_c 分别输入至第一和第三归一化层Norm中进行处理,将风格编码结果 F_s 分别输入至第二归一化层Norm和第三卷积层conv中进行处理,第一和第二归一化层Norm处理后的结果分别输入至第一和第二卷积层conv中进行处理,将第一和第二卷积层conv处理后的结果相乘后输入Softmax激活函数中处理后输出注意力图A,第一卷积层conv处理后输出的特征中的第一列结果 Xw^1_β 、第二列结果 Xw^2_β 和第三列结果 Xw^3_β 和第二卷积层conv处理后输出的特征中的第一行结果 Xw^1_α 、第二行结果 Xw^2_α 和第三行结果 Xw^3_α 相乘;第三卷积层conv处理后的结果V和注意力图A相乘后获得第一相乘结果,第三卷积层conv处理后的结果V的平方 V^2 和注意力图A相乘后获得第二相乘结果,第二相乘结果和第一相乘结果的均值M的平方 M^2 相减后再开根号获得第一相乘结果的标准差S,第三归一化层Norm处理后的结果和第一相乘结果的标准差S点乘后再和第一相乘结果的均值M相加,最终输出局部风格化特征 F_{cs}^l 作为局部通道感知的注意力模块LCWA的输出。

[0011] 所述的全局风格融合模块GSFT为基于深度神经网络Transformer的模块,全局风格融合模块GSFT的输入为风格类型图片 I_s 以及由局部通道感知的注意力模块LCWA处理后获得的局部风格化特征 F_{cs}^l ,风格类型图片 I_s 输入深度神经网络Transformer中进行处理,首先将风格类型图片 I_s 划分为若干个块Patch并构成风格序列 Z_s ,然后根据预定义的深度神经网络Transformer的索引Q、键K和值V对风格序列 Z_s 进行编码,然后再输入第一多头注意力模块中进行处理后输出第一注意力结果 F_s^g ,将第一注意力结果 F_s^g 依次输入第四归一化层Norm和第一前馈神经网络FFN(Feed Forward Neural Network)中处理后再经过键K和值V进行编码,然后再输入第二多头注意力模块中进行处理,第二多头注意力模块处理后的结果和经索引Q编码后的局部风格化特征 F_{cs}^l 共同依次进行残差连接Add和第五归一化层Norm处理后分别输入第二和第三前馈神经网络FFN中进行处理后分别输出偏置 b_s 和权重 w_s 作为深度神经网络Transformer的输出结果, $b_s \in R^{b \times c \times 1 \times 1}$, $w_s \in R^{b \times c \times h \times w}$, b 、 c 、 h 和 w 分别表示在训练过程中输入一个批次的数量,特征的通道数,特征的高度和宽度,权重 w_s 和局部风格化特征 F_{cs}^l 进行点乘后再与偏置 b_s 相加后输出局部-全局风格化特征 F_{cs}^{lg} 作为全局风格融合模块GSFT的输出。

[0012] 所述的基于补丁的风格迁移损失函数包括视觉VGG模型的内容损失和风格损失,并构建基于补丁的Gram损失(Patch-wise Gram Loss)算法加入风格损失中从而建立改进风格损失,使用内容损失对内容图片 I_c 和风格化图片 I_{cs} 的内容相似度进行约束,使用改进风格损失对风格类型图片 I_s 和风格化图片 I_{cs} 的风格相似度进行约束。内容损失为内容感知损失函数 L_c ,原始的风格损失包括由AdaIN提出的二阶的风格损失函数 L_{gs} 和对抗损失函数 L_{adv} 。

[0013] 所述的基于补丁的Gram损失算法包括补丁操作和Gram矩阵损失函数 L_{ls} ,补丁操作过程中首先将风格类型图片 I_s 和风格化图片 I_{cs} 均进行随机裁剪,然后进行图像增强操作后获得若干风格类型补丁和风格化补丁,根据各个风格类型补丁和风格化补丁输入Gram矩阵损失函数 L_{ls} 中获得Gram矩阵损失值以进行风格迁移网络的训练。

[0014] 所述的Gram矩阵损失函数 L_{ls} 如下:

[0015] $L_{ls} = (\sum_i^N R(l_{patch}^i, C(\hat{I}_s^i, I_s), \tau)) / N$

[0016] $R(l_{patch}^i, C(\hat{I}_s^i, I_s), \tau) = l_{patch}^i$, 当 $C(\hat{I}_s^i, I_s) \geq \tau$

[0017] $R(l_{patch}^i, C(\hat{I}_s^i, I_s), \tau) = 0$, 其它

[0018] $C(\hat{I}_s^i, I_s) = G(E_I(\hat{I}_s^i)) \cdot G(E_I(I_s)) / (|G(E_I(\hat{I}_s^i))| \cdot |G(E_I(I_s))|)$

[0019] 其中, N 表示补丁的总数量; $R()$ 和 τ 分别表示阈值函数及其函数参数; l_{patch}^i 表示第 i 个风格类型补丁和风格化补丁的Gram矩阵的余弦相似度; $C()$ 表示两个张量之间的余弦相似度; $\hat{I}_s^i$ 表示风格类型图片 I_s 的第 i 个风格类型补丁, 作为区分符号用于区分原始图片及其补丁; $G()$ 表示裁剪CLIP空间中的第三层的Gram矩阵; $E_I()$ 表示裁剪CLIP空间的图像编码器。

[0020] 所述的第 i 个风格类型补丁和风格化补丁的Gram矩阵的余弦相似度 l_{patch}^i 如下:

[0021] $l_{patch}^i = \|G(E_I(aug(\hat{I}_{cs}^i))) - G(E_I(aug(\hat{I}_s^i)))\|_2$

[0022] 其中, $\hat{I}_{cs}^i$ 表示风格化图片 I_{cs} 的第 i 个风格化补丁; $aug()$ 表示随机几何增强; $\|_2$ 表示二范数。

[0023] 本发明的电子设备, 包括: 相互耦接的存储器和处理器, 其中, 所述存储器存储有程序数据, 所述处理器调用所述程序数据以执行如上述所述的方法。

[0024] 本发明的计算机可读存储介质, 其上存储有程序数据, 所述程序数据被处理器执行时实现如上述所述的方法。

[0025] 现有的基于注意力的任意风格迁移方法采用的是对全局风格信号中学习风格信息的注意机制, 忽略了不同通道的内容和风格信号与全局的差异。本发明提出基于局部和全局的风格迁移方法, 能够更好的从局部风格模式和全局风格分布的角度学习, 方法通过局部通道感知的注意力模块LCWA去从风格图片中学习局部细节的纹理, 通过基于Transformer的全局风格融合模块GSFT进一步学习全局的风格信息, 去除风格图片中一些不想要的信息, 最后提出基于补丁的风格迁移损失函数去平衡局部的风格模式和全局的风格分布。

[0026] 本发明的有益效果是:

[0027] 现有的基于注意力的任意风格迁移方法采用的是对全局风格信号中学习风格信息的注意机制, 忽略了不同通道的内容和风格信号与全局的差异。本发明采用基于局部和全局风格学习的高质量任意风格迁移方法, 能够更好的从局部风格模式和全局风格分布的角度学习, 进行任意风格迁移的同时生成高质量的风格化图片。

附图说明

[0028] 图1为本发明方法的结构框图;

[0029] 图2为本发明所提出的局部通道感知的注意力模块LCWA示意图;

[0030] 图3为本发明所提出的基于Transformer的全局风格融合模块GSFT示意图;

[0031] 图4为本发明所提出的基于补丁的Gram损失算法示意图;

[0032] 图5为本发明实施例中与现有方法在图片风格迁移上的效果对比图, 其中, 图5的(a)为六张不同的风格类型图片, 图5的(b)为六张不同的内容图片, 图5的(c)为使用本发明方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得风格化

图片,图5的(d)为使用基于扩散模型并用风格和内容表达的图像翻译DiffuseIT(Diffusion-based image translation using disentangled style and content representation)方法将不同的风格类型图片的风格分别迁移至不同的内容图片后获得的六张风格化图片,图5的(e)为使用基于比对一致性保持的通用风格迁移CCPL(contrastive coherence preserving loss for versatile style transfer)方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得的风格化图片,图5的(f)为使用带有风格注意力网络的任意风格迁移SANet(Arbitrary style transfer with style-attentional networks)方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得的风格化图片,图5的(g)为使用基于transformer的图片风格迁移StyTr2(Image style transfer with transformers)方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得的风格化图片,图5的(h)为使用重新风格迁移中的注意力机制AdaAttN(Revisit attention mechanism in arbitrary neural style transfer)方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得的风格化图片,图5的(i)为使用基于内外部学习和对比学习的艺术风格迁移IEST(Artistic style transfer with internal-external learning and contrastive learning)方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得的风格化图片,图5的(j)为使用学习线性转换用于快速图像和视频风格迁移Linear(Learning linear transformations for fast image and video style transfer)方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得的风格化图片,图5的(k)为使用无偏的图像风格迁移通过可逆的神经流ArtFlow(Unbiased image style transfer via reversible neural flows)方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得的风格化图片;

[0033] 图6为本发明实施例中与现有方法在视频风格迁移上的效果对比图,其中,图6的(a)为输入的内容图片和风格类型图片及其相邻两帧之间差的热力图,图6的(b)为使用本发明方法处理获得的风格化图片及其相邻两帧之间差的热力图,图6的(c)为使用基于扩散模型并用风格和内容表达的图像翻译DiffuseIT方法处理获得的风格化图片及其相邻两帧之间差的热力图,图6的(d)为使用基于比对一致性保持的通用风格迁移CCPL方法处理获得的风格化图片及其相邻两帧之间差的热力图,图6的(e)为使用带有风格注意力网络的任意风格迁移SANet方法处理获得的风格化图片及其相邻两帧之间差的热力图,图6的(f)为使用基于transformer的图片风格迁移StyTr2方法处理获得的风格化图片及其相邻两帧之间差的热力图,图6的(g)为使用重新风格迁移中的注意力机制AdaAttN方法处理获得的风格化图片及其相邻两帧之间差的热力图,图6的(h)为使用基于内外部学习和对比学习的艺术风格迁移IEST方法处理获得的风格化图片及其相邻两帧之间差的热力图,图6的(i)为使用学习线性转换用于快速图像和视频风格迁移Linear方法处理获得的风格化图片及其相邻两帧之间差的热力图,图6的(j)为使用无偏的图像风格迁移通过可逆的神经流ArtFlow方法处理获得的风格化图片及其相邻两帧之间差的热力图。

具体实施方式

[0034] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行描述,显

然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0035] 本发明的基于局部和全局风格学习的图片高质量任意风格迁移方法具体如下:

[0036] 如图1所示,建立基于局部和全局风格学习的风格迁移网络,基于局部和全局风格学习的风格迁移网络包括依次连接的视觉VGG模型的编码器、局部通道感知的注意力模块LCWA、全局风格融合模块GSFT和视觉VGG模型的解码器,风格迁移网络的输入为一张内容图片 I_c 和一张风格类型图片 I_s ,依次经过编码器和局部通道感知的注意力模块LCWA处理后获得局部风格化特征 F_{cs}^1 ,将局部风格化特征 F_{cs}^1 和风格类型图片 I_s 共同输入至全局风格融合模块GSFT中进行处理,依次经全局风格融合模块GSFT和解码器处理后将风格类型图片的风格转移至内容图片,输出最终的风格化图片 I_{cs} 。

[0037] 如图2所示,局部通道感知的注意力模块LCWA包括三个归一化层Norm、三个 1×1 卷积层conv和Softmax激活函数,内容图片 I_c 和风格类型图片 I_s 经处理后获得内容编码结果 F_c 和风格编码结果 F_s ,将内容编码结果 F_c 分别输入至第一和第三归一化层Norm中进行处理,将风格编码结果 F_s 分别输入至第二归一化层Norm和第三 1×1 卷积层conv中进行处理,第一和第二归一化层Norm处理后的结果分别输入至第一和第二 1×1 卷积层conv中进行处理,将第一和第二 1×1 卷积层conv处理后的结果相乘后输入Softmax激活函数中处理后输出注意力图A,第一 1×1 卷积层conv处理后输出的特征中的第一列结果 Xw_β^1 、第二列结果 Xw_β^2 和第三列结果 Xw_β^3 和第二 1×1 卷积层conv处理后输出的特征中的第一行结果 Xw_α^1 、第二行结果 Xw_α^2 和第三行结果 Xw_α^3 相乘;第三 1×1 卷积层conv处理后的结果V和注意力图A相乘后获得第一相乘结果,第三 1×1 卷积层conv处理后的结果V的平方 V^2 和注意力图A相乘后获得第二相乘结果,第二相乘结果和第一相乘结果的均值M的平方 M^2 相减后再开根号获得第一相乘结果的标准差S,第三归一化层Norm处理后的结果和第一相乘结果的标准差S点乘后再和第一相乘结果的均值M相加,最终输出局部风格化特征 F_{cs}^1 作为局部通道感知的注意力模块LCWA的输出。

[0038] 目前的任意风格迁移中常使用视觉VGG-19模型来提取显著特征来识别的,而使用视觉VGG-19模型提取风格特征并将其融合到内容特征中是有效的,它忽略了不同渠道之间不同局部模式的多样性,视觉VGG-19模型在获取注意力图时,并不会考虑视觉VGG-19模型提取的内容特征Q和风格特征K在不同通道的重要性,会简单地为每个通道统一的分配权重;本发明利用内容和风格可学习参数 W_{con} 和 W_{sty} 进行扩展,为特征的每个通道重新分配权重, $W_{con}, W_{sty} \in R^{B \times 512 \times 512}$,B表示输入一次性送入网络的数量,具体实施时B=8,能够动态地给风格特征的不同通道分配不同的权重,通过通道的角度更加关注局部的风格特征。局部通道感知的注意力模块LCWA能够感知局部的风格模式,并将其融入到内容特征中,在不同的风格特征通道之间学习局部风格模式,可以关注不同的渠道,避免忽视重要的地方风格模式。

[0039] 如图3所示,全局风格融合模块GSFT为基于深度神经网络Transformer的模块,全局风格融合模块GSFT的输入为风格类型图片 I_s 以及由局部通道感知的注意力模块LCWA处理后获得的局部风格化特征 F_{cs}^1 ,风格类型图片 I_s 输入深度神经网络Transformer中进行处理,首先将风格类型图片 I_s 划分为若干个块Patch并构成风格序列 Z_s ,然后根据预定义的深

度神经网络Transformer的索引Q、键K和值V对风格序列 Z_s 进行编码,然后再输入第一多头注意力模块中进行处理后输出第一注意力结果 F_s^g ,将第一注意力结果 F_s^g 依次输入第四归一化层Norm和第一前馈神经网络FFN中处理后再经过键K和值V进行编码,然后再输入第二多头注意力模块中进行处理,第二多头注意力模块处理后的结果和经索引Q编码后的局部风格化特征 F_{cs}^l 共同依次进行残差连接Add和第五归一化层Norm处理后分别输入第二和第三前馈神经网络FFN中进行处理后分别输出偏置 b_s 和权重 w_s 作为深度神经网络Transformer的输出结果, $b_s \in R^{b \times c \times 1 \times 1}$, $w_s \in R^{b \times c \times h \times w}$, b 、 c 、 h 和 w 分别表示在训练过程中输入一个批次的数量,特征的通道数,特征的高度和宽度权重 w_s 和局部风格化特征 F_{cs}^l 进行点乘后再与偏置 b_s 相加后输出局部-全局风格化特征 F_{cs}^{lg} 作为全局风格融合模块GSFT的输出。

[0040] 使用局部通道感知的注意力模块LCWA仅能学习到局部的风格特征,而无法学习到全局的风格特征,为此,本发明利用深度神经网络Transformer学习全局风格信息,主要是通过将风格类型图片 I_s 进行划分为一个风格序列 Z_s ,然后利用多头注意力机制学习风格信息的长范围依赖,并产生权重 w_s 和偏置 b_s ,用权重 w_s 和偏置 b_s 重新处理学习到的局部风格化特征 F_{cs}^l ,最终获得局部-全局的风格化特征 F_{cs}^{lg} 作为解码器的输入。

[0041] 基于Transformer的全局风格融合模块GSFT更加专注于探索如何将全局的风格特征与局部风格化特征相融合,能够使用Transformer将学习到的全局风格特征重新将其融入到局部的风格化特征中,进一步保持风格化图片与风格图片之间的风格一致性。

[0042] 如图1和图4所示,同时构建风格迁移网络的基于补丁的风格迁移损失函数,包括视觉VGG模型的内容损失和风格损失,并构建基于补丁的Gram损失算法加入风格损失中从而建立改进风格损失,使用内容损失对内容图片 I_c 和风格化图片 I_{cs} 的内容相似度进行约束,使用改进风格损失对风格类型图片 I_s 和风格化图片 I_{cs} 的风格相似度进行约束。内容损失为内容感知损失函数 L_c ,原始的风格损失包括由AdaIN提出的二阶的风格损失函数 L_{gs} 和对抗损失函数 L_{adv} 。

[0043] 基于补丁的Gram损失算法包括补丁操作和Gram矩阵损失函数 L_{ls} ,补丁操作过程中首先将风格类型图片 I_s 和风格化图片 I_{cs} 均进行随机裁剪,然后进行图像增强操作后获得若干风格类型补丁和风格化补丁,根据各个风格类型补丁和风格化补丁输入Gram矩阵损失函数 L_{ls} 中获得Gram矩阵损失值以进行风格迁移网络的训练,Gram矩阵损失函数 L_{ls} 如下:

$$[0044] \quad L_{ls} = (\sum_i^N R(I_{patch}^i, C(\hat{I}_s^i, I_s), \tau)) / N$$

$$[0045] \quad R(I_{patch}^i, C(\hat{I}_s^i, I_s), \tau) = l_{patch}^i, \text{当 } C(\hat{I}_s^i, I_s) \geq \tau$$

$$[0046] \quad R(I_{patch}^i, C(\hat{I}_s^i, I_s), \tau) = 0, \text{其它}$$

$$[0047] \quad C(\hat{I}_s^i, I_s) = G(E_I(\hat{I}_s^i)) \cdot G(E_I(I_s)) / (|G(E_I(\hat{I}_s^i))| \cdot |G(E_I(I_s))|)$$

$$[0048] \quad l_{patch}^i = \|G(E_I(aug(\hat{I}_{cs}^i))) - G(E_I(aug(\hat{I}_s^i)))\|_2$$

[0049] 其中, N 表示补丁的总数量; $R()$ 和 τ 分别表示阈值函数及其函数参数,具体实施时 $\tau=0.75$; l_{patch}^i 表示第 i 个风格类型补丁和风格化补丁的Gram矩阵的余弦相似度; $C()$ 表示两个张量之间的余弦相似度; $\hat{I}_s^i$ 表示风格类型图片 I_s 的第 i 个风格类型补丁,作为区分符号用于区分原始图片及其补丁; $G()$ 表示裁剪CLIP空间中的第三层的Gram矩阵; $E_I()$ 表示裁剪CLIP空间的图像编码器; $\hat{I}_{cs}^i$ 表示风格化图片 I_{cs} 的第 i 个风格化补丁; $aug()$ 表示随机几何增强; $\| \cdot \|_2$ 表示二范数。

[0050] 目前使用的风格损失函数是以全局的角度约束内容和风格之间的相似度,而忽略

了局部的风格信息,本发明的基于补丁的Gram损失算法,通过在风格化模式之间正则化局部风格模式来提高风格化图像的质量,从风格类型图片 I_s 和风格化图片 I_{cs} 中随机抽取相同数量的风格补丁,然后将风格化图片 I_{cs} 中提取到的补丁和风格类型图片 I_s 的补丁在裁剪CLIP空间中并计算Gram矩阵,并且约束其欧几里得距离。

[0051] 为了学习风格图片的局部风格信息,本发明提出了基于补丁的Gram损失算法,具体上,是通过对风格类型图片 I_s 和风格化图片 I_{cs} 分别随机进行裁剪,并且将裁减后的图片进行相同方式的翻转等图像增强操作,然而由于每个裁减得到的风格类型图片 I_s 和风格化图片 I_{cs} 并不能够完全代表本来图片的全局风格,因此去掉了一些和原图相似度低的补丁,具体是计算每个风格补丁的Gram矩阵与风格图像之间的余弦相似度 l_{patch}^i 。

[0052] 将内容数据集MS-COCO和风格类型数据集WikiArt共同构建为训练集,将训练集输入风格迁移网络中进行训练,在训练过程中,所有图片被调整为 512×512 尺寸,然后随机裁剪变成了 256×256 的固定分辨率,同时采用Adam优化器学习率设置为0.0005,设置了批量大小和迭代次数分别为8和160000,所有训练在NVIDIA RTX 3090 (24GB) GPU上使用Pytorch框架搭建。训练直至基于补丁的风格迁移损失函数收敛达到最小,获得训练完成的风格迁移网络。

[0053] 最后将待风格迁移的内容图片和目标风格类型图片共同输入至训练完成的风格迁移网络中进行处理,处理完成后输出风格化图片并在显示器进行显示,实现图片高质量风格迁移。如图5的(a)、图5的(b)、图5的(c)、图5的(d)、图5的(e)、图5的(f)、图5的(g)、图5的(h)、图5的(i)、图5的(j)和图5的(k)所示,分别为六种不同的风格类型图片,六种不同的内容图片,以及由本发明方法、基于扩散模型并用风格和内容表达的图像翻译DiffuseIT方法、基于比对一致性保持的通用风格迁移CCPL方法、带有风格注意力网络的任意风格迁移SANet方法、基于transformer的图片风格迁移StyTr2方法、重新风格迁移中的注意力机制AdaAttN方法、基于内外部学习和比对学习的艺术风格迁移IEST方法、学习线性转换用于快速图像和视频风格迁移Linear方法和无偏的图像风格迁移通过可逆的神经流ArtFlow方法将六张不同的风格类型图片的风格分别迁移至六张不同的内容图片后获得的风格化图片,本发明在具体实施时与各方法在图片风格迁移上进行效果对比。如图6的(a)、图6的(b)、图6的(c)、图6的(d)、图6的(e)、图6的(f)、图6的(g)、图6的(h)、图6的(i)和图6的(j)所示,本发明在具体实施时从视频中提取视频帧并与各方法在视频风格迁移上进行效果对比,其中,图6的(a)中的第一张图片为输入图片,其中的大图为内容图片,小图为风格类型图片,根据各图片相邻两帧之间差的热力图,从而计算获得相似度LPIPS(Learned Perceptual Image Patch Similarity)指标以评估生成图像与真实图像之间感知相似度,本发明方法和其它各方法处理获得的风格化图片和内容图像的相似度LPIPS指标依次为0.3105、0.4206、0.3679、0.3273、0.3792、0.3253、0.3485、0.3371和0.3343,本发明方法处理获得的风格化图片和内容图像的相似度LPIPS指标最低为0.3105,效果最好,可见本发明方法能够更好的从局部风格模式和全局风格分布的角度学习,进行任意风格迁移的同时生成高质量的风格化图片。

[0054] 现有的大多数基于注意力的方法都是对VGG模型提取的特征进行处理。本发明提出的局部通道感知的注意力模块LCWA将局部样式模式融合到内容特征中,从而获得局部风格化特征的方法。然而,仅仅依靠局部通道感知的注意力模块LCWA来关注风格图像中局部

图案的有限区域,可能会导致与全球风格分布的偏差,为此,本发明提出基于Transformer的全局风格融合模块GSFT来解决这个问题,利用一个Transformer来学习全局风格分布,并动态地生成全局风格权重和偏差,以拉出与之对齐的局部风格化特征。考虑到目前广泛使用的风格损失是基于二阶全局风格统计,这可能导致本发明提出的框架倾向于学习更多的全局风格分布,为此,本发明提出基于补丁的风格迁移损失函数,通过在风格化模式之间正则化局部风格模式来提高风格化图像的质量。

[0055] 以上所述实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0056] 以上所述实施例仅表达了本发明的几种实施方式,其描述较为具体和详细,但不能因此而理解为对发明范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变形和改进,这些都属于本发明的保护范围。因此,本发明的保护范围应以所附权利要求为准。

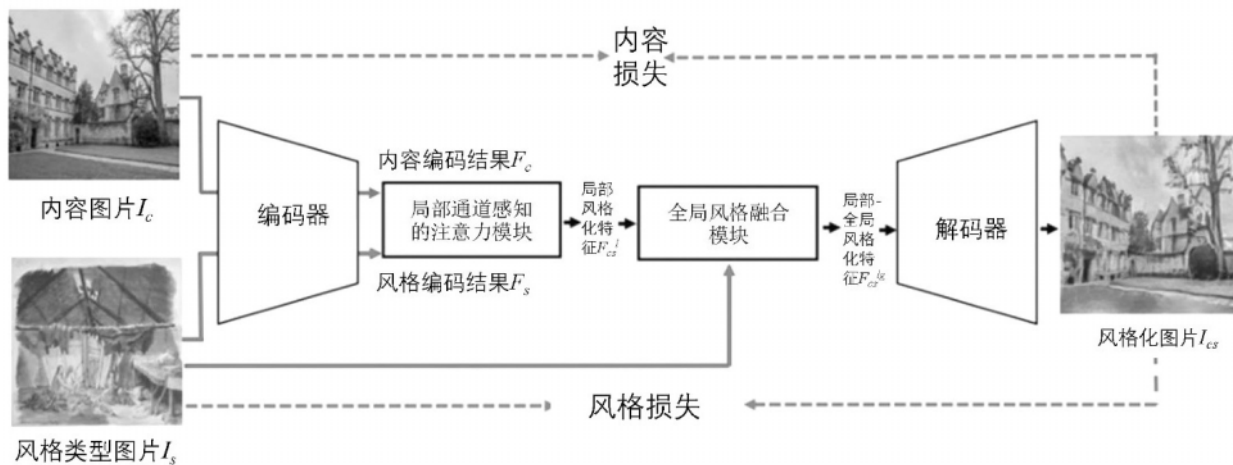


图 1

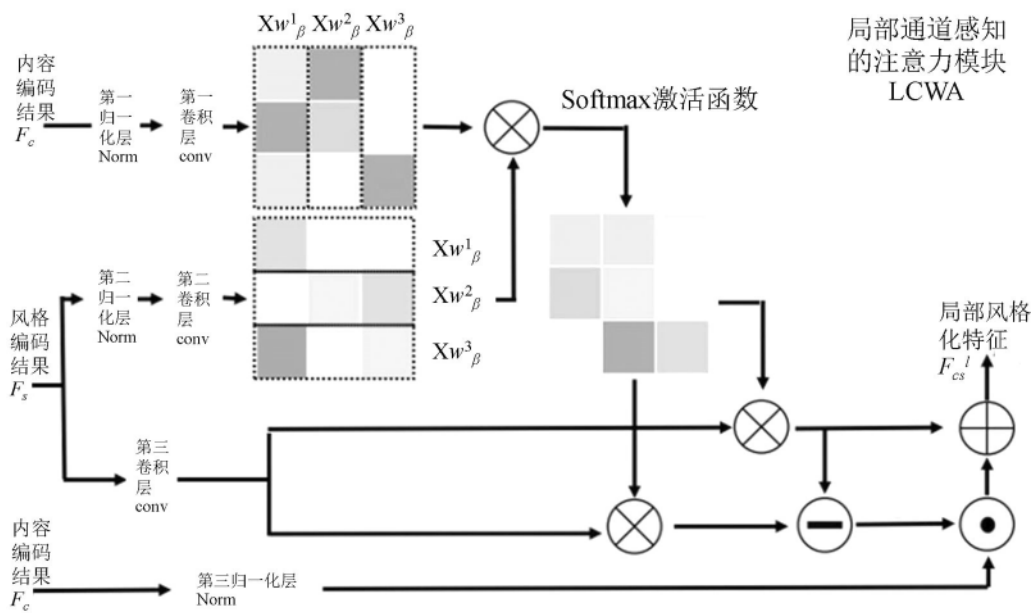


图 2

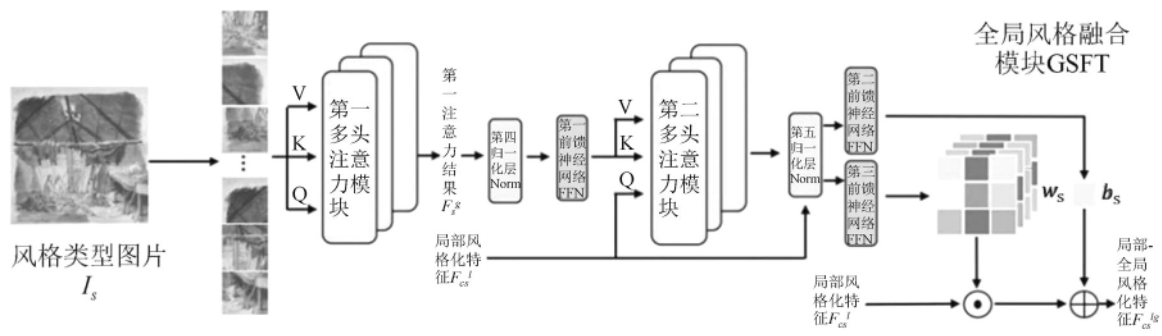


图 3



图 4



图 5

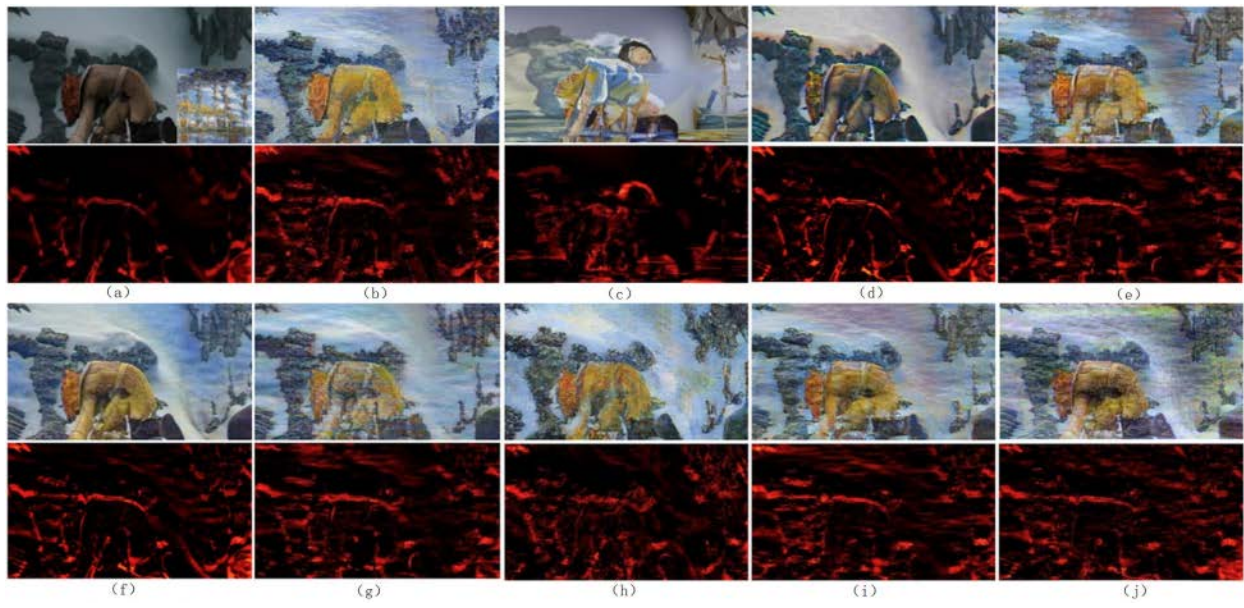


图 6