



## (12) 发明专利

(10) 授权公告号 CN 118710552 B

(45) 授权公告日 2025. 06. 20

(21) 申请号 202410843431.8

G06T 5/90 (2024.01)

(22) 申请日 2024.06.27

G06V 10/764 (2022.01)

(65) 同一申请的已公布的文献号

G06T 7/11 (2017.01)

申请公布号 CN 118710552 A

G06T 5/60 (2024.01)

G06V 10/82 (2022.01)

(43) 申请公布日 2024.09.27

(56) 对比文件

(73) 专利权人 西藏民族大学

CN 113449135 A, 2021.09.28

地址 712082 陕西省咸阳市渭城区文汇东路6号

CN 118196113 A, 2024.06.14

审查员 郑岩

(72) 发明人 樊瑶 柏劲威 赵芷蔚

(74) 专利代理机构 北京万新知识产权代理有限公司 16195

专利代理师 甘善甜

(51) Int. Cl.

G06T 5/77 (2024.01)

G06V 10/774 (2022.01)

权利要求书5页 说明书20页 附图7页

## (54) 发明名称

一种唐卡图像修复方法、系统及存储介质

## (57) 摘要

本发明提供一种唐卡图像修复方法、系统及存储介质,涉及数字图像处理技术领域,通过将输入图像分割成非重叠子区域并进行非线性变换,有效隔离局部信息,同时引入矢量量化码本以更好地捕获和保留图像结构信息和细节,并行CSWin分辨率Transformer模块通过十字形窗口和局部增强位置编码加强上下文建模能力;新颖的多尺度特征引导模块通过不同尺度的局部知识适应性地学习非缺损区域的特征信息;CDCT模型在多个数据集上的实验显示其修复结果具有竞争力;通过采用SSIM、PSNR和综合质量评价指标QI,实现了修复质量的显著提升。



1. 一种唐卡图像修复方法,其特征在于,具体步骤包括:

步骤S1、收集包含破损部分和缺失部分的唐卡图像,并对收集的唐卡图像进行预处理,将预处理后的唐卡图像构建为图像数据集;

步骤S2、构建基于Transformer模型的编码器-解码器结构,并用于共同学习一个离散码本,将图像数据集内部的唐卡图像输入至构建的编码器内部,利用编码器将输入的唐卡图像划分为固定大小的非重叠子区域,并将子区域通过非线性变换映射到连续的潜在空间表示,得到特征向量;

步骤S3、引入矢量量化码本,对编码器输出的连续潜在特征向量进行矢量量化,所述离散码本采用聚类算法构建,码本中的每个向量代表图像中的一块子区域的潜在空间表示,最后将得到的离散码本作为码本先验知识;

步骤S4、构建并行CSWin分辨率Transformer模块,该模块采用十字形窗口以及局部增强位置编码的设计,将步骤S3中特征向量作为输入,并对该特征向量添加额外的可学习位置嵌入以保留空间信息,随后将特征向量沿空间维度扁平化,从而获得该模块的最终输入,以预测下一个索引的概率分布;

步骤S5、利用步骤S4中并行CSWin分辨率Transformer模块来准确推断缺失token的索引,并通过这些索引从步骤S3中得到的离散码本中找到对应的离散向量,用于图像修复,完成一次修复尝试后,系统将进入迭代循环;

步骤S6,在每次迭代后,采集生成的修复图像,并获取该修复图像与参考图像的结构相似性指数、该修复图像与参考图像的峰值信噪比PSNR,并对结构相似性指数、峰值信噪比PSNR进行分析处理,生成综合质量评价指数QI,该指数用于对图像质量、修复效果进行评价,并生成相应的学习率二次调整策略;

步骤S7,在自适应学习率调整的基础上,设计多尺度特征引导模块,该模块利用非破损区域的特征,以促进生成区域与未损坏区域在结构和纹理上的一致性,从而提升修复结果的质量和保真度;

步骤S8、在修复过程中,根据步骤S6生成的综合质量评价指数QI和学习率调整策略,动态调整模型的学习率,以优化修复效果;

步骤S9、完成所有迭代后,对最终修复的唐卡图像进行后处理,包括但不限于图像增强、颜色校正和细节优化,以提高修复图像的质量和视觉效果,最终输出修复后的唐卡图像。

2. 根据权利要求1所述的一种唐卡图像修复方法,其特征在于:在共享码本学习阶段,系统架构包含三个核心组件,依次为码本编码器E、码本解码器G以及一个含有K个离散编码的码本 $C = \{c_k\}_{k=0}^K$ ;

当处理输入图像 $I_i \in \mathbb{R}^{H \times W \times 3}$ 时,首先由码本编码器E将图像 $I_i$ 转换为一个高维空间中的潜在表示 $Z$ , $I_i \in \mathbb{R}^{H \times W \times 3}$ 表示一个具有高度H、宽度W和3个颜色通道的图像,H和W分别代表图像的垂直和水平尺寸;H和W用于描述图像的实际尺寸;

即 $Z = E(x) \in \mathbb{R}^{m \times n \times d}$ ,这里d代表构成这个潜在向量的维度数量;其中, $m \times n$ 是潜在表示中的空间分辨率,且m和n分别表示空间维度的高度和宽度;

将此空间潜在表示 $Z$ 中的每一个向量量化至码本中最邻近的码词 $c_k$ ,从而得到以下向量量化的输出 $Z_c$ 和相应的代码令牌序列 $s \in \{0, \dots, N-1\}^{m' \cdot n'}$ ,并标记为公式(1):

$$Z_c^{(i,j)} = \mathbf{q}(Z) = \arg \min_{c_k \in C} \|Z^{(i,j)} - c_k\|$$

$$s^{(i,j)} = \arg \min_k \|Z^{(i,j)} - c_k\|_2$$

其中,每个元素 $Z_c^{(i,j)}$ 是 $Z^{(i,j)}$ 在码本 $C$ 中找到的最接近的码词;量化操作通过计算距离 $\|Z^{(i,j)} - c_k\|$ 并选择最小距离对应的码词来实现;

$Z_c^{(i,j)}$ 是向量量化后的输出,表示在位置 $i$ ,分辨率 $j$ 处的潜在表示 $Z^{(i,j)}$ 被量化到码本 $C$ 中最接近的码词;

$\mathbf{q}(Z)$ 是向量量化的操作,将潜在表示 $Z$ 中的每个向量映射到码本 $C$ 中最接近的码词;

$\arg \min_{c_k \in C} \|Z^{(i,j)} - c_k\|$ 是在码本 $C$ 中找到与 $Z^{(i,j)}$ 距离最小的码词 $c_k$ 的操作; $\arg \min$ 表示找到使距离 $\|Z^{(i,j)} - c_k\|$ 最小的码词 $c_k$ 的索引;

$s^{(i,j)}$ 是量化后得到的代码令牌序列中的一个元素,表示在位置 $i$ ,分辨率 $j$ 处的潜在表示 $Z^{(i,j)}$ 被量化到的码词的索引;

$\arg \min_k \|Z^{(i,j)} - c_k\|_2$ 是在码本 $C$ 中找到与 $Z^{(i,j)}$ 欧氏距离最小的码词 $c_k$ 的操作, $\arg \min$ 表示找到使欧氏距离 $\|Z^{(i,j)} - c_k\|_2$ 最小的码词 $c_k$ 的索引;

$c_k \in C$ 表示码本 $C$ 中的每个码词 $c_k$ 是一个预定义的向量,用于量化潜在表示 $Z$ ;

$Z^{(i,j)}$ 是潜在表示 $Z$ 在位置 $i$ ,分辨率 $j$ 处的向量;

$\|Z^{(i,j)} - c_k\|$ 是 $Z^{(i,j)}$ 与码词 $c_k$ 之间的距离,是欧氏距离或其他距离度量;

$\|Z^{(i,j)} - c_k\|_2$ 是 $Z^{(i,j)}$ 与码词 $c_k$ 之间的欧氏距离;

$s \in \{0, \dots, N-1\}^{m' \cdot n'}$ 是量化后得到的代码令牌序列,其中 $m' \cdot n'$ 表示潜在表示 $Z$ 的空间大小,即 $Z$ 中包含的向量总数,这里的 $m' \cdot n'$ 指潜在表示 $Z$ 中的元素数量,每个元素对应一个潜在向量, $N$ 是码本 $C$ 中码词的数量;

随后,解码器 $G$ 在给定 $Z_c$ 的情况下重建高质量图像 $I_{\text{rec}}$ ;形成的 $m' \cdot n'$ 码令牌序列 $s$ 代表了新的潜在离散表征,

即当 $s^{(i,j)} = k$ 时,  $Z_c^{(i,j)} = c_k$ ;整体重构 $I_{\text{rec}} \approx I_t$ 公式化为以下,并标记为公式(2):

$$I_{\text{rec}} = G(Z_c) = G(\mathbf{q}(E(I)))$$

编码器执行映射操作,将尺寸为 $H \times W$ 的图像数据转化成为 $H/m \times W/n$ 尺度的离散编码形式;并通过重建损失使码本和模型进行端到端的训练。

3. 根据权利要求2所述的一种唐卡图像修复方法,其特征在于:所述码本编码器 $E$ 设计点包括共享码本学习阶段的编码器设计成以非重叠补丁方式,并通过多个线性残差层处理输入图像;

基于码本先验的图像修复阶段,包括对特征向量添加额外的可学习位置嵌入以保留空间信息,随后将特征向量沿空间维度扁平化,从而获得该模块的最终输入;

模型使用并行CSWinTransformer块,其中,每个块由并行的多头自注意力块和交叉形窗口注意力块、前馈层构成;使PCT模块结合多头和交叉形窗口,交叉形窗口注意力块在线性投影值 $V$ 上添加位置编码机制LePE以增强局部归纳偏差;

所述交叉形窗口注意力将通道拆分成水平和垂直条纹,一半的头部捕获水平条纹注意力,另一半的头部捕获垂直条纹注意力;

所述的PCT模块的输出经过一个线性层,并使用Softmax函数,被映射成一个概率分布,该分布对应于码本e中K个潜在向量的概率分布。

4.根据权利要求3所述的一种唐卡图像修复方法,其特征在于:训练所述PCT模块以预测下一个索引的概率分布 $p(s_i | s_{<i>_i})$ ;使训练目标等于最小化数据表示的负对数似然性;

所述在每次迭代中,评估生成图像的质量,使用SSIM和PSNR质量评价指标;

梯度信息收集,在每次迭代中,收集生成器和判别器的梯度信息 $\nabla L$ ,其中L表示损失函数;

学习率调整:

使用Adam优化器,其学习率调整公式为:

$$lr_{t+1} = lr_t \cdot \sqrt{1 - \beta_2^t} / (1 - \beta_1^t)$$

其中, $lr_t$ 是第t次迭代的学习率, $\beta_1$ 、 $\beta_2$ 是Adam优化器的超参数,分别设置为0.9和0.95;

为了进一步自适应调整学习率,引入梯度变化率 $grad_{var_t}$ 来动态调整学习率:

$$lr_{t+1} = lr_t \cdot (1 + \alpha_3 \cdot grad_{var_t})$$

其中, $grad_{var_t}$ 是第t次迭代中梯度方差, $\alpha_3$ 是一个调整因子,用于控制梯度变化对学习率的影响程度;

计算梯度方差,以反映梯度的稳定性:

$$grad_{var_t} = \frac{1}{N_2} \sum_{i=1}^{N_2} (g_{t,i} - \mu_t)^2$$

其中, $g_{t,i}$ 是第t次迭代中第i个参数的梯度, $N_2$ 是参数总数, $\mu_t$ 是梯度的平均值。

5.根据权利要求4所述的一种唐卡图像修复方法,其特征在于:在每次迭代后,采集生成的修复图像,并获取该修复图像与参考图像的结构相似性指数、峰值信噪比PSNR,并对结构相似性指数、峰值信噪比PSNR进行分析处理,生成综合质量评价指标QI,该指数用于对图像质量、修复效果进行评价,并生成相应的学习率二次调整策略;

在每次迭代后,采集生成的修复图像,并计算以下质量参数:

结构相似性指数SSIM如下:

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

其中, $SSIM(x, y)$ 中x和y分别是参考图像和修复图像的局部窗口, $\mu_x$ 、 $\mu_y$ 是均值, $\sigma_x^2$ 、 $\sigma_y^2$ 是方差, $\sigma_{xy}$ 是协方差, $c_1$ 、 $c_2$ 是常数,用于稳定计算;

峰值信噪比PSNR如下:

$$PSNR(x, y) = 10 \cdot \lg \left( \frac{MAX_I^2}{MSE(x, y)} \right)$$

其中, $PSNR(x, y)$ 中x和y分别是参考图像和修复图像的局部窗口, $MAX_I$ 是图像像素的最大值, $MSE(x, y)$ 是均方误差;

综合SSIM和PSNR,生成一个综合质量评价指标QI,计算公式如下:

$$QI = \frac{\sum_{i=1}^4 \omega_i \cdot f(\text{Metric}_{i''})}{\sum_{i=1}^4 \omega_i} \cdot g(\text{NIQE})$$

参数解释,  $\omega_i$  为权重因子,用于平衡不同质量评估指标的影响,其中  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$ 、 $\omega_4$  分别对应SSIM、PSNR、FSIM和NIQE的权重;

$f(\text{Metric}_{i''})$  为复杂函数,  $i'' \in \{1, 2, 3, 4\}$ , 其中  $\text{Metric}_{i''}$  中  $i''$  分别取值1、2、3、4时, 分别代表SSIM、PSNR、FSIM、NIQE, 用于对每个质量评估指标进行非线性变换, 公式如下:

$$f(\text{SSIM}) = \log(1 + \text{SSIM})$$

$$f(\text{PSNR}) = \exp(-\text{PSNR}/100)$$

$$f(\text{FSIM}) = \sqrt{\text{FSIM}}$$

$$f(\text{NIQE}) = \frac{1}{1 + \text{NIQE}}$$

$g(\text{NIQE})$  为归一化函数, 用于调整NIQE的影响,  $g(x) = \frac{1}{2}(\tanh(x) + 1) = \frac{1}{2} \left( \frac{e^x - e^{-x}}{e^x + e^{-x}} + 1 \right)$ ;  $x$

$\in \text{NIQE}$ ;

设定QI的值域范围为(0, 1), 当QI接近1时, 表示图像质量接近原始图像, 修复效果好; 当QI接近0时, 表示图像质量较差, 修复效果不佳。

6. 根据权利要求5所述的一种唐卡图像修复方法, 其特征在于: 当QI值增加时, 表示图像修复质量提高, 图像更接近原始图像的视觉和结构特征; 反之, QI值减少表示修复效果不佳, 需要调整模型参数或训练策略;

根据这些指标的变化情况, 动态调整生成器和判别器的学习率, 如果质量评价指标提升缓慢或下降, 增加学习率以探索新的参数空间; 如果质量评价指标稳定提升, 则保持或适度降低学习率, 以稳定训练; 具体包括以下内容:

根据生成的综合质量评价指标QI, 对学习率进行二次调整:

$$\text{lr}_{t+1} = \text{lr}_t \cdot (1 + \beta_5 \cdot (\text{QI}_t - \text{QI}_{\text{target}}))$$

其中,  $\text{QI}_t$  是第  $t$  次迭代时的综合质量评价指标,  $\text{QI}_{\text{target}}$  是目标质量评价指标,  $\beta_5$  是调整因子, 用于控制质量评价指标对学习率的影响;

当  $\text{QI}_t$  取值在区间一 (0, 0.3) 时, 图像质量较差, 需要增加学习率以探索新参数, 快速改善图像修复效果, 阈值设定为0.2, 低于此值时, 紧急增加学习率以期达成显著改进;

当  $\text{QI}_t$  取值在区间二 [0.3, 0.7) 时, 图像质量有改善空间; 采用调整策略, 维持或轻微增加学习率以稳步提升图像质量, 阈值为0.5, 以保持训练的稳定性 and 连续改进;

当  $\text{QI}_t$  取值在区间三 [0.7, 1) 时, 表示图像质量接近理想, 在此区间, 降低学习率以稳定训练并防止过拟合; 设定阈值为0.85, 超过此值时进一步降低学习率, 确保质量的持续优化和稳定性。

7. 根据权利要求6所述的一种唐卡图像修复方法, 其特征在于: 所述设计多尺度特征引导模块, 该模块利用非破损区域的特征, 以促进生成区域与未损坏区域在结构和纹理上的

一致性,从而提升修复结果的质量和保真度;具体包括以下内容:

假设输入图像是具有掩码 $m$ 的掩码输入 $Y$ ,该模块将掩码图像输入表示为多层特征映射,在多尺度特征引导模块中注入基于大核的卷积,

使用LKA结构,该结构使用扩张率为 $d$ 的 $\frac{k}{d} \times \frac{k}{d}$ 深度卷积提取局部特征,接着,通过一个 $(2d-1) \times (2d-1)$ 深度膨胀卷积捕获长距离依赖关系,最后,通过 $1 \times 1$ 逐点卷积整合信息并调整通道数,增强通道间的交互;

在LKA模块之后加入前馈网络2,所述前馈网络2由RMS归一化、 $3 \times 3$ 卷积、Swish激活函数、 $3 \times 3$ 卷积以及Dropout组成。

8.一种唐卡图像修复系统,其特征在于:所述系统用于执行权利要求1-7任意一项所述的方法。

9.一种存储介质,其上存储有计算机程序,其特征在于:所述计算机程序被处理器执行时实现权利要求1至7中任一项所述唐卡图像修复方法的步骤。

## 一种唐卡图像修复方法、系统及存储介质

### 技术领域

[0001] 本发明涉及数字图像处理技术领域,具体为一种唐卡图像修复方法、系统及存储介质。

### 背景技术

[0002] 在数字图像处理领域,图像修复技术一直是研究的热点,其目的在于通过计算方法修复受损的图像内容,以便恢复图像的完整性和美观性;这一领域起始于20世纪末,早期主要依赖简单的插值算法,如最近邻插值和双线性插值;进入21世纪,随着计算机视觉和机器学习的快速发展,图像修复技术得到了显著的提升;尤其是深度学习的兴起,使得基于神经网络的图像修复方法,如卷积神经网络(CNN)和生成对抗网络(GAN),逐渐成为主流,它们能够学习大量数据的复杂映射,实现高度复杂的图像修复效果;然而,传统方法和早期深度学习方法在处理具有丰富纹理和复杂结构的图像,然而,与自然图像不同,唐卡中佛像的服饰或者背景中的花蔓、云朵、山水的图案复杂精美,细节丰富,面对这类唐卡图像时,这类方法预测的先验信息的准确性会大大降低,导致在填充内部纹理时出现大范围的修复错误;如唐卡等艺术品图像时,仍然面临着巨大挑战;

[0003] 现有技术的不足主要体现在,对高度结构化和精细详细的图像内容修复上的局限性;尤其是唐卡图像,通常包含精细的线条和复杂的图案,这些内容一旦破坏,使用简单的纹理复制或基本的学习模型难以达到令人满意的修复效果;此外,现有技术在修复过程中往往忽略了图像的全局一致性和局部细节的恢复,导致修复后的图像在视觉上无法与原图无缝对接;更为关键的是,大多数方法未能有效利用图像内在的结构信息,无法保持图像修复过程中的结构完整性,尤其是在存在大面积损坏时,这一问题尤为突出;而生成对抗网络(GAN)在为修复图像提供逼真纹理方面表现出色,但其训练稳定性和模式坍塌仍然是需要解决的问题;此外,质量评价指标在现有技术中多为单一指标评价,缺乏对图像质量的多维度综合评估。

[0004] 在所述背景技术部分公开的上述信息仅用于加强对本公开的背景的理解,因此它可以包括不构成对本领域普通技术人员已知的现有技术的信息。

### 发明内容

[0005] 本发明的目的在于提供一种唐卡图像修复方法、系统及存储介质,以解决上述背景技术中提出的问题。

[0006] 为实现上述目的,本发明提供如下技术方案:

[0007] 一种唐卡图像修复方法,具体步骤包括:

[0008] 步骤S1、收集包含破损部分和缺失部分的唐卡图像,并对收集的唐卡图像进行预处理,将预处理后的唐卡图像构建为图像数据集;

[0009] 步骤S2、构建基于Transformer模型的编码器-解码器结构,并用于共同学习一个离散码本,将图像数据集内部的唐卡图像输入至构建的编码器内部,利用编码器将输入的

唐卡图像划分为固定大小的非重叠子区域,并将子区域通过非线性变换映射到连续的潜在空间表示,得到特征向量;

[0010] 步骤S3、引入矢量量化码本,对编码器输出的连续潜在特征向量进行矢量量化,所述离散码本采用聚类算法构建,码本中的每个向量代表图像中的一块子区域的潜在空间表示,最后将得到的离散码本作为码本先验知识;

[0011] 步骤S4,构建并行CSWin分辨率Transformer模块,该模块采用十字形窗口以及局部增强位置编码的设计,将步骤S3中特征向量作为输入,并对该特征向量添加额外的可学习位置嵌入以保留空间信息,随后将特征向量沿空间维度扁平化,从而获得该模块的最终输入,以预测下一个索引的概率分布;

[0012] 步骤S5、利用步骤S4中并行CSWin分辨率Transformer模块来准确推断缺失token的索引,并通过这些索引从步骤S3中得到的离散码本中找到对应的离散向量,用于图像修复,完成一次修复尝试后,系统将进入迭代循环;

[0013] 步骤S6,在每次迭代后,采集生成的修复图像,并获取该修复图像与参考图像的结构相似性指数、该修复图像与参考图像的峰值信噪比PSNR,并对结构相似性指数、峰值信噪比PSNR进行分析处理,生成综合质量评价指标QI,该指数用于对图像质量、修复效果进行评价,并生成相应的学习率二次调整策略;

[0014] 步骤S7,在自适应学习率调整的基础上,设计多尺度特征引导模块,该模块利用非破损区域的特征,以促进生成区域与未损坏区域在结构和纹理上的一致性,从而提升修复结果的质量和保真度;

[0015] 步骤S8、在修复过程中,根据步骤S6生成的综合质量评价指标QI和学习率调整策略,动态调整模型的学习率,以优化修复效果;

[0016] 步骤S9、完成所有迭代后,对最终修复的唐卡图像进行后处理,包括但不限于图像增强、颜色校正和细节优化,以提高修复图像的质量和视觉效果,最终输出修复后的唐卡图像。

[0017] 进一步地,在每次迭代后,采集生成的修复图像,并获取该修复图像与参考图像的结构相似性指数、峰值信噪比PSNR,并对结构相似性指数、峰值信噪比PSNR进行分析处理,生成综合质量评价指标QI,该指数用于对图像质量、修复效果进行评价,并生成相应的学习率二次调整策略;

[0018] 在每次迭代后,采集生成的修复图像,并计算以下质量参数:

[0019] 结构相似性指数SSIM如下:

$$[0020] \quad SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

[0021] 其中,x和y分别是参考图像和修复图像的局部窗口, $\mu_x$ 、 $\mu_y$ 是均值, $\sigma_x^2$ 、 $\sigma_y^2$ 是方差, $\sigma_{xy}$ 是协方差, $c_1$ 、 $c_2$ 是常数,用于稳定计算;

[0022] 峰值信噪比PSNR如下:

$$[0023] \quad PSNR(x, y) = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE(x, y)} \right)$$



[0024] 其中,  $\text{MAX}_I$  是图像像素的最大值,  $\text{MSE}(x, y)$  是均方误差;

[0025] 综合SSIM和PSNR, 生成一个综合质量评价指标QI, 计算公式如下:

$$[0026] \quad \text{QI} = \frac{\sum_{i=1}^4 \omega_i \cdot f(\text{Metric}_{i''})}{\sum_{i=1}^4 \omega_i} \cdot g(\text{NIQE})$$

[0027] 参数解释,  $\omega_i$  为权重因子, 用于平衡不同质量评估指标的影响, 其中  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$ 、 $\omega_4$  分别对应SSIM、PSNR、FSIM和NIQE的权重;

[0028]  $f(\text{Metric}_{i''})$  为复杂函数,  $i'' \in \{1, 2, 3, 4\}$ , 其中  $\text{Metric}_{i''}$  中  $i''$  分别取值1、2、3、4时, 分别代表SSIM、PSNR、FSIM、NIQE, 用于对每个质量评估指标进行非线性变换, 公式如下:

[0029]  $f(\text{SSIM}) = \log(1 + \text{SSIM})$

[0030]  $f(\text{PSNR}) = \exp(-\text{PSNR}/100)$

$$f(\text{FSIM}) = \sqrt{\text{FSIM}}$$

[0031]

$$f(\text{NIQE}) = \frac{1}{1 + \text{NIQE}}$$

[0032]  $g(\text{NIQE})$  为归一化函数, 用于调整NIQE的影响,

$$g(x) = \frac{1}{2}(\tanh(x) + 1) = \frac{1}{2} \left( \frac{e^x - e^{-x}}{e^x + e^{-x}} + 1 \right); x \in \text{NIQE};$$

[0033] 设定QI的值域范围为(0, 1), 当QI接近1时, 表示图像质量接近原始图像, 修复效果良好; 当QI接近0时, 表示图像质量较差, 修复效果不佳。

[0034] 进一步地, 当QI值增加时, 表示图像修复质量提高, 图像更接近原始图像的视觉和结构特征; 反之, QI值减少表示修复效果不佳, 需要调整模型参数或训练策略;

[0035] 根据这些指标的变化情况, 动态调整生成器和判别器的学习率, 如果质量评价指标提升缓慢或下降, 增加学习率以探索新的参数空间; 如果质量评价指标稳定提升, 则保持或适度降低学习率, 以稳定训练; 具体包括以下内容:

[0036] 根据生成的质量评价指标QI, 对学习率进行二次调整:

[0037]  $\text{lr}_{t+1} = \text{lr}_t \cdot (1 + \beta_5 \cdot (\text{QI}_t - \text{QI}_{\text{target}}))$

[0038] 其中,  $\text{QI}_t$  是第t次迭代时的质量评价指标,  $\text{QI}_{\text{target}}$  是目标质量评价指标,  $\beta_5$  是调整因子, 用于控制质量评价指标对学习率的影响;

[0039] 当 $\text{QI}_t$ 取值在区间一(0, 0.3)时, 图像质量较差, 需要增加学习率以探索新参数, 快速改善图像修复效果, 阈值设定为0.2, 低于此值时, 紧急增加学习率以期达成显著改进;

[0040] 当 $\text{QI}_t$ 取值在区间二[0.3, 0.7)时, 图像质量有改善空间; 采用调整策略, 维持或轻微增加学习率以稳步提升图像质量, 阈值为0.5, 以保持训练的稳定性 and 连续改进;

[0041] 当 $\text{QI}_t$ 取值在区间三[0.7, 1)时, 表示图像质量接近理想, 在此区间, 降低学习率以稳定训练并防止过拟合; 设定阈值为0.85, 超过此值时进一步降低学习率, 确保质量的持续优化和稳定性。

[0042] 一种唐卡图像修复系统, 所述系统用于执行所述的方法。

[0043] 一种存储介质, 其上存储有计算机程序, 所述计算机程序被处理器执行时实现所

述唐卡图像修复方法的步骤。

[0044] 与现有技术相比,本发明的有益效果是:

[0045] 1、设计了一种新颖的码本学习框架,其中,编码器将输入图像划分为固定大小的非重叠子区域(patch),再非线性变换为潜在特征向量,确保了局部信息的有效隔离;引入矢量量化码本更有效地捕捉和保留图像的结构信息和细节,从而在重建过程中产生更逼真的结果;

[0046] 2、设计一个并行CSWinTransformer模块,十字形窗口以及局部增强位置编码的设计,加强了上下文建模能力的同时降低了计算成本,提高索引预测的准确度;

[0047] 3、创新地设计了一个多尺度特征引导模块,其中不同尺度的LKA利用局部信息和通道上的适应性,更好地学习非缺损区域的特征信息;

[0048] 4、CDCT模型在Celeba-HQ、Places2、自制的唐卡数据集上与现有的前沿方法进行了大量实验;定性和定量实验表明,CDCT模型的修复结果具有竞争力;为唐卡图像等文化遗产的修复工作开辟了新颖的技术路径;

[0049] 5、采用了SSIM和PSNR作为质量评价指标,并引入综合质量评价指标QI,通过对图像修复结果的多维度评价,实现学习率的二次调整,进一步提升了修复质量;

[0050] 综上,在码本学习阶段,设计改进了一种基于矢量量化码本的网络框架,来离散化编码输入图像的中间特征,得到上下文丰富的离散码本;第二阶段,提出一种基于十字形窗口的并行Transformer模块,能够在有限的计算成本下准确预测图像缺失区域的索引组合;此外,提出了一种多尺度特征引导模块,将未破损区域的特征和码本中的纹理特征逐步融合,从而更好地保留未破损区域的局部细节。

## 附图说明

[0051] 图1为本发明整体方法流程示意图;

[0052] 图2为本发明的CDCT模型总体框架示意图;

[0053] 图3为本发明的并行CSWin分辨率Transformer模块示意图;

[0054] 图4为本发明的多尺度特征引导模块示意图;

[0055] 图5为本发明的损失值随迭代次数的变化曲线示意图;

[0056] 图6为本发明在Celeba-HQ数据集上的定性比较结果;

[0057] 图7为本发明在places2数据集上的定性比较结果;

[0058] 图8为本发明在自制的唐卡数据集上的定性比较结果;

[0059] 图9为本发明的模型各组件视觉效果分析示意图。

## 具体实施方式

[0060] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,对本发明进一步详细说明。

[0061] 需要说明的是,除非另外定义,本发明使用的技术术语或者科学术语应当为本发明所属领域内具有一般技能的人士所理解的通常意义。本发明中使用的“第一”“第二”以及类似的词语并不表示任何顺序、数量或者重要性,而只是用来区分不同的组成部分。“包括”或者“包含”等类似的词语意指出现该词前面的元件或者物件涵盖出现在该词后面列举的

元件或者物件及其等同,而不排除其他元件或者物件。“连接”或者“相连”等类似的词语并非限定于物理的或者机械的连接,而是可以包括电性的连接,不管是直接的还是间接的。“上”“下”“左”“右”等仅用于表示相对位置关系,当被描述对象的绝对位置改变后,则该相对位置关系也可能相应地改变。

[0062] 实施例一:

[0063] 请参阅图1至图9,本发明提供一种技术方案:

[0064] 一种唐卡图像修复方法,具体步骤包括:

[0065] 步骤S1、收集包含破损部分和缺失部分的唐卡图像,并对收集的唐卡图像进行预处理,将预处理后的唐卡图像构建为图像数据集;

[0066] 步骤S2、构建基于Transformer模型的编码器-解码器结构,并用于共同学习一个离散码本,将图像数据集内部的唐卡图像输入至构建的编码器内部,利用编码器将输入的唐卡图像划分为固定大小的非重叠子区域,并将子区域通过非线性变换映射到连续的潜在空间表示,得到特征向量;

[0067] 步骤S3、引入矢量量化码本,对编码器输出的连续潜在特征向量进行矢量量化,所述离散码本采用聚类算法构建,码本中的每个向量代表图像中的一块子区域的潜在空间表示,最后将得到的离散码本作为码本先验知识;

[0068] 步骤S4、构建并行CSWin分辨率Transformer模块,该模块采用十字形窗口以及局部增强位置编码(LePE)的设计,将步骤S3中特征向量作为输入,并对该特征向量添加额外的可学习位置嵌入以保留空间信息,随后将特征向量沿空间维度扁平化,从而获得该模块的最终输入,以预测下一个可能索引的概率分布;

[0069] 步骤S5、利用步骤S4中并行CSWin分辨率Transformer模块来准确推断缺失token的索引,并通过这些索引从步骤S3中得到的离散码本中找到对应的离散向量,用于图像修复,完成一次修复尝试后,系统将进入迭代循环;

[0070] 步骤S6,在每次迭代后,采集生成的修复图像,并获取该修复图像与参考图像的结构相似性指数、该修复图像与参考图像的峰值信噪比PSNR,并对结构相似性指数、峰值信噪比PSNR进行分析处理,生成综合质量评价指标QI,该指数用于对图像质量、修复效果进行评价,并生成相应的学习率二次调整策略;

[0071] 步骤S7,在自适应学习率调整的基础上,设计多尺度特征引导模块,该模块利用非破损区域的特征,以促进生成区域与未损坏区域在结构和纹理上的一致性,从而提升修复结果的质量和保真度;

[0072] 步骤S8、在修复过程中,根据步骤S6生成的综合质量评价指标QI和学习率调整策略,动态调整模型的学习率,以优化修复效果;

[0073] 步骤S9、完成所有迭代后,对最终修复的唐卡图像进行后处理,包括但不限于图像增强、颜色校正和细节优化,以提高修复图像的质量和视觉效果,最终输出修复后的唐卡图像。

[0074] 实施例二:

[0075] 在实施例一的基础上进一步说明,在共享码本学习阶段,系统架构包含三个核心组件:码本编码器E、码本解码器G以及一个含有K个离散编码的码本 $\mathbf{C} = \{\mathbf{c}_k\}_{k=0}^K$ ;这是一个包含K个离散编码的集合,每个 $\mathbf{c}_k$ 是一个码词,代表一种特定的特征或模式;

[0076] 当处理输入图像  $I_t \in \mathbb{R}^{H \times W \times 3}$  时, 首先由码本编码器E将图像  $I_t$  转换为一个高维空间中的潜在表示  $Z$ ,  $I_t \in \mathbb{R}^{H \times W \times 3}$  表示一个具有高度H、宽度W和3个颜色通道的图像;

[0077] 即  $Z = E(x) \in \mathbb{R}^{m \times n \times d}$ , 这里d代表构成这个潜在向量的维度数量; $m \times n$ 是潜在表示中的空间分辨率;

[0078] 随后, 采用逐元素量化操作  $q(\cdot)$ , 这是一个函数, 用于将潜在表示Z中的每个向量量化到码本C中最接近的码词, 这个操作是逐元素进行的, 即对Z中的每个元素  $(i, j)$  分别进行量化;

[0079] 将此空间潜在表示Z中的每一个向量量化至码本中最邻近的码词  $c_k$ , 从而得到向量量化的输出  $Z_c$  和相应的代码令牌序列  $s \in \{0, \dots, N-1\}^{m' \cdot n'} ::$

$$\begin{aligned} Z_c^{(i,j)} &= \mathbf{q}(Z) = \arg \min_{c_k \in C} \|Z^{(i,j)} - c_k\| \\ [0080] \quad s^{(i,j)} &= \arg \min_k \|Z^{(i,j)} - c_k\|_2 \end{aligned} \quad (1)$$

[0081] 其中, 每个元素  $Z_c^{(i,j)}$  是  $Z^{(i,j)}$  在码本C中找到的最接近的码词; 量化操作通过计算距离  $\|Z^{(i,j)} - c_k\|$  并选择最小距离对应的码词来实现;

[0082] 随后, 解码器G在给定  $Z_c$  的情况下重建高质量图像  $I_{rec}$ ; 形成的  $m' \cdot n'$  码令牌序列  $s$  代表了新的潜在离散表征, 明确指示了学习到的码本中各个位置的码字索引, 即当  $s^{(i,j)} = k$  时,  $Z_c^{(i,j)} = c_k$ ; 整体重构  $I_{rec} \approx I_t$  公式化为:

$$[0083] \quad I_{rec} = G(Z_c) = G(q(E(I))) \quad (2)$$

[0084] 编码器执行映射操作, 将尺寸为  $H \times W$  的图像数据转化成为  $H/m \times W/n$  尺度的离散编码形式, 其中参数  $m, n$  标识了下采样比例;

[0085] 这一过程实质上是 将图像  $I_t$  内每个  $m \times n$  区域的信息凝聚为单一编码单元, 因此, 当提及  $Z_c$  中的任意编码元素时,  $m \times n$  同样象征着该编码在原始图像  $I_t$  空间上的对应覆盖范围;

[0086] 通过重建损失使码本和模型进行端到端的训练, 本申请采用了四种图像级别的重建损失, 所述重建损失包括L1损失  $L_1$ , 感知损失  $L_{per}$ 、对抗损失  $L_{adv}$  和风格损失  $L_{style}$ ;

[0087] 具体损失函数定义如下:

$$[0088] \quad \begin{cases} L_1 = \|I_t - I_{rec}\|_1 \\ L_{per} = \|\Phi(I_t) - \Phi(I_{rec})\|_2^2 \\ L_{adv} = [\log D(I_t) + \log(1 - D(I_{rec}))] \\ L_{style} = E \left[ \sum_k \frac{1}{M_k} \|G_{\phi_k}(I_t) - G_{\phi_k}(I_{rec})\|_1 \right] \end{cases} \quad (3)$$

[0089] 其中,  $\Phi$  指的是VGG19网络中的特征提取器, 由于在更新码本项时, 图像级别的损失约束不足,

[0090]  $I_t$  表示目标图像, 即希望模型生成的图像能够接近参考图像;

[0091]  $I_{rec}$  表示重建图像,  $\|I_t - I_{rec}\|_1$  表示目标图像和重建图像之间的L1距离, 即两幅图像每个对应像素值之差的绝对值之和;

[0092]  $\Phi$  指的是VGG19网络中的特征提取器；

[0093]  $\Phi(I_t)$  和  $\Phi(I_{rec})$  分别表示使用预训练的神经网络提取的目标图像和重建图像的特征表示；

[0094]  $\|\Phi(I_t) - \Phi(I_{rec})\|_2^2$  表示这两组特征表示之间的欧氏距离的平方,用于衡量两幅图像在感知上的差异；

[0095]  $D(I_t)$  和  $D(I_{rec})$  分别表示判别器对目标图像和重建图像的判别结果,其中D是判别器网络；

[0096]  $\log D(I_t)$  和  $\log(1 - D(I_{rec}))$  分别表示判别器正确识别目标图像和错误识别重建图像的概率的对数；

[0097]  $G_{\phi_k}(I_t)$  和  $G_{\phi_k}(I_{rec})$  分别表示目标图像和重建图像在第k个特征通道上的Gram矩阵,其中  $\phi_k$  表示第k个特征通道的参数；

[0098]  $E \left[ \sum_k \frac{1}{M_k} \|G_{\phi_k}(I_t) - G_{\phi_k}(I_{rec})\|_1 \right]$  表示所有特征通道上Gram矩阵之间的L1距离的期望值,用于衡量两幅图像在风格上的差异,其中  $M_k$  是第k个特征通道的元素数量；

[0099] 因此本申请还采用中间代码级别的损失  $L_{quantize}$  来减小码本C和嵌入的输入特征Z之间的差异；

$$[0100] \quad L_{quantize} = \|sg(Z) - Z_c\|_2^2 + \beta \|Z - sg(Z_c)\|_2^2 \quad (4)$$

[0101] 其中,  $sg(\cdot)$  指代停止梯度运算符,而参数  $\beta$  设为0.25,旨在均衡编码器与码本更新速度之间的权重；

[0102] 针对公式(1)所示的特征量化过程不具备可导性,采取了一种直接传递策略,即在反向传播期间将梯度从解码环节镜像至编码环节,以保障反向传播的实施；

[0103] 为了全面指导码本的先验知识学习,将综合的损失函数  $L_{codebook}$  作为优化目标,来驱动整个端到端训练流程；

$$[0104] \quad L_{codebook} = L_1 + L_{per} + L_{quantize} + \lambda_{adv} \cdot L_{adv} + L_{style} \quad (5)$$

[0105] 其中,在申请的实验中,  $\lambda_{adv}$  设置为0.8；

[0106] 虽然更多的码本项可以简化重建,冗余元素会在后续的代码预测中引起歧义,因此,本申请CDCT方法将码本的项数N设置为1024,这足以实现精确的图像重建；此外,码本维度d被设置为256。

[0107] 实施例三：

[0108] 在实施例二的基础上进一步说明,所述码本编码器E设计点具体包括：

[0109] 传统的基于CNN的编码器以滑动窗口的方式利用若干卷积核来处理输入图像,这不适合于图像修复,因为它们会在掩蔽区域和未掩蔽区域之间引入干扰,因此,共享码本学习阶段的编码器设计成以非重叠补丁方式,并通过多个线性残差层处理输入图像；

[0110] 具体来说,Token表示是使用线性残差结构在8个块中提取的,每个区块中,包括两组GELU激活函数、线性层和残差连接；

[0111] 首先,对输入图片进行unfold操作将尺寸变为  $(3 \times m \times n, L)$ , L指块的数量,再通过调整层将特征变换为  $(L, d)$  的尺寸；

[0112] 然后在每个块中,对输入特征进行256至128维度,128至256维度的变换;经过八个线性残差层的特征提取后,通过fold操作得到潜在表示Z;

[0113] 获得了 $r=H/n=W/m=32$ 的大压缩比,这使得在第二阶段中的全局建模具有很好的抗退化鲁棒性和易于处理的计算成本;

[0114] 本申请的解码器G由3个转置卷积和1个卷积层组成,用于上采样;反卷积核的大小为 $4 \times 4$ ,表示反卷积核的宽度和高度都是4;步幅为2,表示反卷积核在输入图像上的滑动步长为2;

[0115] 填充大小为1,表示在输入图像的边缘填充一个像素,以保持输出的空间尺寸;首先通过三个转置卷积将维度为 $256 \times 32 \times 32$ 的特征上采样到 $64 \times 256 \times 256$ ,再通过1个卷积核尺寸为 $3 \times 3$ 、反射填充参数为1、步长为1的卷积将输出调整为 $256 \times 256 \times 3$ ,得到重建的图像。

[0116] 实施例四:

[0117] 在实施例三的基础上进一步说明,所述基于码本先验的图像修复阶段,具体包括以下内容:

[0118] 在现有的用于图像修复和补全的Transformer架构中,量化像素的索引既用作输入又用作预测目标;虽然这种利用上下文索引预测缺失索引的策略提高了计算效率,但这种Transformer的输入类型会出现严重的信息丢失问题,不利于索引序列预测;因此,本申请设计的并行CSWinTransformer模块(PCT)直接将

[0119] codebook编码器的特征向量 $\hat{Z}$ 作为输入,有助于在降低信息损失的同时做出更精确的预测;

[0120] 并行CSWinTransformer模块如图3所示,

[0121] 对特征向量 $\hat{Z} \in \mathbb{R}^{(H \times W) \times C}$ 添加额外的可学习位置嵌入以保留空间信息,随后将特征向量沿空间维度扁平化,从而获得该模块的最终输入;

[0122] 模型使用12个并行CSWinTransformer块,其中,每个块由并行的多头自注意力块和交叉形窗口注意力块、前馈层(Feedforward1)构成;

[0123] 自注意力的头数设置为8,不同于常见的Transformer模块,PCT模块结合多头和交叉形窗口,大大减少了计算量,达到了更好的修复效果,另外,交叉形窗口注意力块在线性投影值V上添加位置编码机制LePE以增强局部归纳偏差;

[0124] 值得注意,PCT模块中的交叉形窗口注意力和完全自注意力从不同的感受野进行训练,并通过残差连接在一起,因此,标准自我注意力块不会受到CSWin注意力块的影响,前馈层1中Swish函数能够在保留ReLU的非线性特性的同时更好地平滑梯度;

[0125] 十字形窗口以及局部增强位置编码的设计如下:

[0126] 与轴向注意力不同,所述交叉形窗口注意力将通道拆分成水平和垂直条纹,一半的头部捕获水平条纹注意力,另一半的头部捕获垂直条纹注意力;

[0127] 具体以水平条纹自注意力为例,将特征矩阵S等间隔划分为一系列宽度为b且互不重叠的水平条纹段 $[S^1, \dots, S^N]$ ,其中 $N=H/b$ ,每个条纹段均包含b列、W行的元素;此外,灵活调整b这一超参数,以在学习能力与计算成本间取得平衡,假设每个头部(head)对应的查询、键和值向量的维度均为d,则每个头部所处理的水平条纹自注意力输出定义为如下表达

式:

$$[0128] \quad S = [S^1, S^2, \dots, S^N],$$

$$[0129] \quad Y^i = \text{Attention}(S^i W^Q, S^i W^K, S^i W^V), (6)$$

$$[0130] \quad \text{Attention}^H(S) = [Y^1, Y^2, \dots, Y^N]$$

[0131] 其中  $S^i \in \mathbb{R}^{(b \times W) \times C}$ ,  $i = 1, \dots, N$ ,  $W^Q \in \mathbb{R}^{C \times d}$ ,  $W^K \in \mathbb{R}^{C \times d}$ ,  $W^V \in \mathbb{R}^{C \times d}$  分别代表各个头部对输入特征矩阵进行线性变换后得到的查询矩阵、键矩阵以及值矩阵;

[0132]  $S = [S^1, S^2, \dots, S^N]$  表示特征矩阵  $S$  被等间隔划分为一系列宽度为  $b$  且互不重叠的水平条纹段, 每个条纹段  $S^i$  包含  $b$  列和  $W$  行的元素;

[0133]  $N = H/b$  表示  $N$  是水平条纹段的数量, 等于特征矩阵的高度  $H$  除以每个条纹段的宽度  $b$ ;

[0134]  $S^i \in \mathbb{R}^{(b \times W) \times C}$  表示每个条纹段  $S^i$  是一个  $b \times W$  的矩阵, 其中  $b$  是列数,  $W$  是行数,  $C$  是特征维度;

[0135]  $W^Q \in \mathbb{R}^{C \times d}$ ,  $W^K \in \mathbb{R}^{C \times d}$ ,  $W^V \in \mathbb{R}^{C \times d}$  表示这些是线性变换的权重矩阵, 用于将输入特征矩阵  $S^i$  转换为查询矩阵 (Query)、键矩阵 (Key) 和值矩阵 (Value),  $C$  是输入特征的维度,  $d$  是每个头部 (head) 的维度;

[0136]  $Y^i = \text{Attention}(S^i W^Q, S^i W^K, S^i W^V)$  表示这是注意力机制的计算过程, 其中  $S^i W^Q$ ,  $S^i W^K$  分别是查询、键和值的线性变换结果, 注意力机制计算出的输出  $Y^i$  是第  $i$  个水平条纹段的注意力输出;

[0137]  $\text{Attention}^H(S) = [Y^1, Y^2, \dots, Y^N]$  表示这是所有水平条纹段的注意力输出结果的集合, 表示为水平方向的注意力输出;

[0138]  $\text{Attention}^V(S)$  表示这是针对垂直条纹区域执行的局部自注意力操作的输出结果, 表示为垂直方向的注意力输出;

[0139] 与之类似地, 针对垂直条纹区域执行的局部自注意力操作也能够相应地推导得出, 每个头部对应的输出结果则采用  $\text{Attention}^V(S)$  来表示;

[0140] 所述的PCT模块的输出经过一个线性层, 并使用Softmax函数, 被映射成一个概率分布, 该分布对应于码本  $e$  中  $K$  个潜在向量的概率分布; 即图像块对应码本中特征的概率;

[0141] 为了量化模型预测与类别标签之间的一致程度, 训练PCT模块以预测下一个索引的概率分布  $p(s_i | s_{<i})$ ; 使训练目标等于最小化数据表示的负对数似然性:

$$[0142] \quad L_{\text{Transformer}} = E_{x' \sim p(x')} [-\log p(s)] \quad (7)$$

[0143] 其中,  $p(s) = \prod_i p(s_i | s_{<i})$ ;

[0144]  $p(s_i | s_{<i})$  是一个条件概率分布, 表示在已知序列中索引小于  $i$  的所有元素  $s_{<i}$  的条件下, 预测序列中第  $i$  个索引  $s_i$  的概率分布, 这个概率分布是由PCT模块在训练过程中学习得到的;

[0145]  $L_{\text{Transformer}}$  是Transformer模型的损失函数, 用于量化模型预测与真实类别标签之间的一致程度, 在训练过程中, 目标是使这个损失函数最小化;

[0146]  $E_{x \sim p(x')}$  是期望值的符号, 表示对从数据分布  $p(x')$  中抽取的样本  $x'$  进行平均, 在这里, 它表示对所有可能的数据样本  $x'$  计算损失函数的平均值;

[0147]  $-\log p(s)$  是负对数似然性, 用于计算模型预测的概率分布  $p(s)$  与真实标签  $s$  之间的差异, 负对数似然性是一个常用的损失函数, 用于优化概率模型;



[0148]  $p(s) = \prod_i p(s_i | s_{<i})$  是序列中所有索引的概率分布的乘积,它表示模型对整个序列的预测概率,其中每个索引 $s_i$ 的概率是基于之前所有索引 $s_{<i}$ 的条件概率;

[0149] 所述在每次迭代中,评估生成图像的质量,使用SSIM和PSNR质量评价指标;

[0150] 梯度信息收集,在每次迭代中,收集生成器和判别器的梯度信息 $\nabla L$ ,其中L表示损失函数;

[0151] 学习率调整:

[0152] 使用Adam优化器,其学习率调整公式为:

$$[0153] \quad lr_{t+1} = lr_t \cdot \sqrt{1 - \beta_2^t} / (1 - \beta_1^t)$$

[0154] 其中, $lr_t$ 是第t次迭代的学习率, $\beta_1$ 、 $\beta_2$ 是Adam优化器的超参数,分别设置为0.9和0.95;

[0155] 为了进一步自适应调整学习率,引入梯度变化率 $grad_{\text{var}}$ 来动态调整学习率:

$$[0156] \quad lr_{t+1} = lr_t \cdot (1 + \alpha_3 \cdot grad_{\text{var}})$$

[0157] 其中, $grad_{\text{var}}$ 是第t次迭代中梯度方差, $\alpha_3$ 是一个调整因子,用于控制梯度变化对学习率的影响程度;

[0158] 梯度方差计算:

[0159] 计算梯度方差,以反映梯度的稳定性:

$$[0160] \quad grad_{\text{var}} = \frac{1}{N_2} \sum_{i=1}^{N_2} (g_{t,i} - \mu_t)^2$$

[0161] 其中, $g_{t,i}$ 是第t次迭代中第i个参数的梯度, $N_2$ 是参数总数, $\mu_t$ 是梯度的平均值;

[0162] 学习率更新,根据上述公式更新学习率,并在下一次迭代中使用新的学习率进行参数更新;本方案通过引入梯度方差来动态调整学习率,相较于传统的固定学习率或简单的学习率衰减策略,能够更精细地控制训练过程,提高模型对复杂图像修复任务的适应性和效果;

[0163] 在每次迭代后,采集生成的修复图像,并获取该修复图像与参考图像的结构相似性指数、峰值信噪比PSNR,并对结构相似性指数、峰值信噪比PSNR进行分析处理,生成综合质量评价指数QI,该指数用于对图像质量、修复效果进行评价,并生成相应的学习率二次调整策略;具体包括以下内容:

[0164] 在每次迭代后,采集生成的修复图像,并计算以下质量参数:

[0165] 结构相似性指数SSIM如下:

$$[0166] \quad SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

[0167] 其中,x和y分别是参考图像和修复图像的局部窗口, $\mu_x$ 、 $\mu_y$ 是均值, $\sigma_x^2$ 、 $\sigma_y^2$ 是方差, $\sigma_{xy}$ 是协方差, $c_1$ 、 $c_2$ 是常数,用于稳定计算;

[0168] 峰值信噪比PSNR如下:

$$[0169] \quad PSNR(x, y) = 10 \cdot \lg \left( \frac{MAX_I^2}{MSE(x, y)} \right)$$



[0170] 其中,  $\text{MAX}_I$  是图像像素的最大值,  $\text{MSE}(x, y)$  是均方误差;

[0171] 综合SSIM和PSNR, 生成一个综合质量评价指标QI, 计算公式如下:

$$[0172] \quad \text{QI} = \frac{\sum_{i=1}^4 \omega_i \cdot f(\text{Metric}_i)}{\sum_{i=1}^4 \omega_i} \cdot g(\text{NIQE})$$

[0173] 参数解释,  $\omega_i$  为权重因子, 用于平衡不同质量评估指标的影响, 其中  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$ 、 $\omega_4$  分别对应SSIM、PSNR、FSIM和NIQE的权重, 分别决定了每个评估指标对整体QI值的影响程度, 且  $\sum_{i=1}^4 \omega_i = 1$ ;

[0174] NIQE是一种无参考的图像质量评估方法, 它基于自然场景统计特性NSS的偏差来评估图像质量; 其数学表达式可以简化表示为:

$$[0175] \quad \text{NIQE} = \sqrt{(\mathbf{v} - \mathbf{v}_0)^T \Sigma^{-1} (\mathbf{v} - \mathbf{v}_0)}$$

[0176] 其中,  $\mathbf{v}$  表示测试图像的特征向量,  $\mathbf{v}_0$  是从参考图像库中提取的特征向量的平均值,  $\Sigma$  是特征向量的协方差矩阵,  $\Sigma^{-1}$  是协方差矩阵的逆矩阵; 这个公式计算了测试图像的特征向量与参考图像特征向量之间的马氏距离;

[0177] FSIM表示FeatureSimilarityIndex; FSIM是一种基于图像特征的相似性指数, 用来评估两幅图像之间的相似度; 其数学表达式简化表示为:

$$[0178] \quad \text{FSIM} = \frac{\sum_{x,y} S_L(x,y) \cdot S_C(x,y) \cdot S_p(x,y)}{\sum_{x,y} 1}$$

[0179] 其中,  $S_L(x, y)$  是位置  $(x, y)$  处的亮度相似性,  $S_C(x, y)$  是位置  $(x, y)$  处的对比度相似性,  $S_p(x, y)$  是位置  $(x, y)$  处的相位一致性; 这个公式通过综合考虑图像的亮度、对比度和相位一致性等特征, 计算出一个综合的相似性度量值;

[0180]  $f(\text{Metric}_i)$  为复杂函数,  $i \in \{1, 2, 3, 4\}$ , 其中  $\text{Metric}_i$  中  $i$  分别取值1、2、3、4时, 分别代表SSIM、PSNR、FSIM、NIQE, 用于对每个质量评估指标进行非线性变换, 公式如下:

$$[0181] \quad f(\text{SSIM}) = \log(1 + \text{SSIM})$$

$$[0182] \quad f(\text{PSNR}) = \exp(-\text{PSNR}/100)$$

$$f(\text{FSIM}) = \sqrt{\text{FSIM}}$$

$$[0183] \quad f(\text{NIQE}) = \frac{1}{1 + \text{NIQE}}$$

[0184]  $g(\text{NIQE})$  为归一化函数, 用于调整NIQE的影响,

$$g(x) = \frac{1}{2}(\tanh(x) + 1) = \frac{1}{2} \left( \frac{e^x - e^{-x}}{e^x + e^{-x}} + 1 \right), x \in \text{NIQE};$$

[0185] 设定QI的值域范围为  $(0, 1)$ , 当QI接近1时, 表示图像质量接近原始图像, 修复效果

良好;当QI接近0时,表示图像质量较差,修复效果不佳;

[0186] 辅助公式:

[0187]  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ , 用于将NIQE的值压缩到(0,1)范围内;

[0188] 当QI值增加时,表示图像修复质量提高,图像更接近原始图像的视觉和结构特征;反之,QI值减少表示修复效果不佳,需要调整模型参数或训练策略;

[0189] 本公式通过引入复杂的非线性变换函数和归一化函数,综合考虑了多个图像质量评估指标的影响,并通过权重因子动态调整各指标的重要性,实现了对图像修复质量的全面评估;此设计不仅提高了评估的准确性,还增强了模型对不同图像修复任务的适应性;

[0190] 根据这些指标的变化情况,动态调整生成器和判别器的学习率,如果质量评价指标提升缓慢或下降,增加学习率以探索新的参数空间;如果质量评价指标稳定提升,则保持或适度降低学习率,以稳定训练;具体包括以下内容:

[0191] 学习率二次调整:

[0192] 根据生成的综合质量评价指数QI,对学习率进行二次调整:

[0193]  $lr_{t+1} = lr_t \cdot (1 + \beta_5 \cdot (QI_t - QI_{target}))$

[0194] 其中,QI<sub>t</sub>是第t次迭代时的综合质量评价指数,QI<sub>target</sub>是目标质量评价指数, $\beta_5$ 是调整因子,用于控制质量评价指数对学习率的影响;

[0195] 将QI<sub>t</sub>的值域(0,1)划分为三个区间,依次为区间一(0,0.3);区间二[0.3,0.7];区间三[0.7,1);

[0196] 区间一(0,0.3),图像质量较差,需要增加学习率以探索新参数,快速改善图像修复效果,阈值设定为0.2,低于此值时,紧急增加学习率以期达成显著改进;

[0197] 量化内容描述:此区间图像修复质量较差,修复效果不佳;此时,SSIM和PSNR指标均显示显著的下降,分别下降了30%和25%;FSIM和NIQE也显示出图像质量的严重退化,FSIM下降了20%,NIQE增加了40%;且SSIM和PSNR在连续三次迭代中提升不超过5%或出现下降,应增加学习率;这种情况下,应显著增加学习率,增加50%,以探索新的参数空间,尝试改善图像修复效果;

[0198] 判断标准和规则:设定阈值为0.2,当QI<sub>t</sub>值低于0.2时,启动紧急调整机制;具体规则如下:如果QI<sub>t</sub>值连续三次迭代低于0.2,则将学习率增加50%,并重新评估SSIM和PSNR的变化;如果SSIM和PSNR在接下来的五次迭代中没有显著改善,改善不超过10%,则进一步增加学习率至75%;

[0199] 交互规则描述:在QI<sub>t</sub>值低于0.3的区间内,各参数的交互变化如下:

[0200] 当SSIM下降30%时,PSNR相应下降25%,FSIM下降20%,NIQE增加40%;这种参数间的负相关变化表明图像质量的整体退化;学习率的增加50%旨在通过参数空间的探索,寻找能够改善图像质量的新参数组合;

[0201] 区间二:[0.3,0.7),图像质量有改善空间;采用调整策略,维持或轻微增加学习率以稳步提升图像质量,阈值为0.5,以保持训练的稳定性和连续改进;

[0202] 量化内容描述,在QI<sub>t</sub>值位于[0.3,0.7)的区间内,图像修复质量有所改善,但仍未达到理想状态;SSIM和PSNR指标显示轻微改善,分别提升了10%和15%;FSIM保持稳定,NIQE下降10%,此时,调整学习率,增加10%以维持训练的稳定性,并继续观察指标变化;

[0203] 判断标准和规则,设定阈值为0.5,当QIt值在0.5附近波动时,采取保守策略;具体规则如下:如果QIt值连续五次迭代在0.45至0.55之间波动,则保持当前学习率不变;如果QIt值连续五次迭代低于0.45或高于0.55,则相应地增加或减少学习率5%;

[0204] 交互规则描述,在QIt值位于[0.3,0.7)的区间内,各参数的交互变化如下:当SSIM提升10%时,PSNR提升15%,FSIM保持不变,NIQE下降10%;这种参数间的正相关变化表明图像质量的逐步改善;学习率的适度增加10%旨在稳定当前的改善趋势,避免过度调整导致训练不稳定;

[0205] 区间三[0.7,1),表示图像质量接近理想,在此区间,降低学习率以稳定训练并防止过拟合;设定阈值为0.85,超过此值时进一步降低学习率,确保质量的持续优化和稳定性;

[0206] 量化内容描述:

[0207] 在QIt值位于[0.7,1)的区间内,图像修复质量接近或达到理想状态。SSIM和PSNR指标显示显著提升,分别提升了20%和25%。FSIM和NIQE也显示出图像质量的显著改善,FSIM提升15%,NIQE下降30%。此时,应降低学习率,减少10%,以稳定训练并防止过拟合;

[0208] 判断标准和规则,设定阈值为0.85,当QIt值超过0.85时,启动稳定策略;具体规则如下:如果QIt值连续三次迭代超过0.85,则将学习率降低10%,并继续监测SSIM和PSNR的变化,如果SSIM和PSNR在接下来的五次迭代中保持稳定或继续提升,则进一步降低学习率至15%;

[0209] 交互规则描述,在QIt值高于0.7的区间内,各参数的交互变化如下:当SSIM提升20%时,PSNR提升25%,FSIM提升15%,NIQE下降30%;这种参数间的正相关变化表明图像质量的显著改善;学习率降低15%,旨在稳定当前的高质量修复状态,防止因学习率过高导致的过拟合。

[0210] 实施例五:

[0211] 在实施例四的基础上进一步说明,所述设计多尺度特征引导模块,充分利用非破坏区域的特征,以促进生成区域与未损坏区域在结构和纹理上的协调一致,提升修复结果的质量和保真度;具体包括以下内容:

[0212] 如图4所示,设计多尺度特征引导模块,旨在保留图像非掩码区域的细节;假设输入图像是具有掩码m的掩码输入Y,该模块将掩码图像输入表示为多层特征映射,而不是将其压缩为单层特征;

[0213] 在多尺度特征引导模块中注入基于大核的卷积,旨在集CNN操作与注意力机制的优势于一体;

[0214] 具体来说,使用LKA(LargeKernelAttention)结构,该结构使用扩张率为d的 $\frac{k}{d} \times \frac{k}{d}$ 深度卷积(DW-Conv)提取局部特征,接着,通过一个 $(2d-1) \times (2d-1)$ 深度膨胀卷积(DW-D-Conv)捕获长距离依赖关系,最后,通过 $1 \times 1$ 逐点卷积整合信息并调整通道数,增强通道间的交互;

[0215] 由于LKA专注于借助宽广感受野优化被遮挡区域的特征表达,有助于在频域中对规则纹理的全局学习;此外,为了确保LKA的泛化能力,还在LKA模块之后加入前馈网络2;具体来说,所述前馈网络2由RMS归一化、 $3 \times 3$ 卷积、Swish激活函数、 $3 \times 3$ 卷积以及Dropout组成;

[0216] 其中,使用RMSNorm归一化函数,以提高训练稳定性;Swish函数能够在保留ReLU的非线性特性的同时更好地平滑梯度,也能解决ReLU函数的不是零中心的且在负数部分的梯度为零的问题。

[0217] 实施例六:

[0218] 在实施例五的基础上进一步说明,本实验是在三个不同的数据集上对本文的模型进行了训练和评估:Celeba-HQ是CelebA数据集的一个扩展版本,包含了高质量、高分辨率的人脸图像,选择了27000张用于训练的图像,3000张用于测试和验证,选取了20个场景类别用于实验,其中90000张图像用于训练,10000张图像被用于定量评估;自制的西藏唐卡数据集,包含佛教唐卡、显密唐卡与家庭唐卡等,其中2500张用于训练,500张用于测试和验证,具体如表1所示;

[0219] 表1Celeba、Facade和自制的唐卡数据集的设置

[0220]	Dataset	Trainingset	Validationse	Testset
[0221]			t	
	Celeba-HQ	27000	1800	1200
	Places2	90000	6000	4000
	唐卡	2500	300	200

[0222] 为了定量比较,本实施例使用了各种图像质量指标,包括传统的峰值信噪比(PSNR),结构相似性指数(SSIM),平均绝对误差(MAE)以及最新的基于特征的学习感知图像块相似性(LPIPS);

[0223] 实施细节如下:

[0224] 对于第一阶段的共享码本学习阶段,本文方法使用了Adam优化器( $\beta_1=0, \beta_2=0.9$ )进行优化,批次大小为16;

[0225] 第二阶段的基于码本先验的图像修复阶段,使用Adam( $\beta_1=0.9, \beta_2=0.95$ )进行优化,批次大小为4,将两阶段的学习率分别设置为 $2e-4$ 和 $3e-4$ ,并采用余弦调度器进行衰减;本实施例方法使用PyTorch框架实现,并使用1个NVIDIA3090GPU进行训练;

[0226] 所有对比模型都在Celeba-HQ和Places2数据集中进行了比较,本申请还在自制的唐卡数据集上重新训练了EC,CTSDG,ICT,PUT和MAT,以进一步讨论修复效果;

[0227] 由图5能够看出,本文模型的第一阶段网络在Places2数据集上的训练情况,随着训练的进行Quantize损失和Adv损失短暂上升,随后在震荡中趋于平稳右侧的L1损失、Perceptual损失和Style损失通过持续训练和调优,本文模型可以逐渐降低损失值并提高生成图像的质量。

[0228] 实施例七:

[0229] 在实施例六的基础上进一步说明,图6、图7和图8展示了从Celeba-HQ、Places2和自制的唐卡数据集中随机选择的测试图像的修复结果的视觉比较;

[0230] 在Celeba数据集上对本文模型与现有先进方法进行了比较,如图6所示,EC和CTSDG这两种修复方法,在面对大面积缺损图像时,结构预测不完整,导致修复结果存在大

面积失真；

[0231] 在图6(b)和(c)中第3行和第5行的人物脸颊和眼睛存在缺失；

[0232] ICT利用Transformer重建视觉先验,修复结果的整体结构较为合理,但图像修复细节上不够完善；

[0233] 如图6(d)的第一行和第二行修复的镜框出现形变,人眼不对称;MAT是一种掩码引导transformer的大面积缺损修复模型,对于图像中的小块缺失区域,处理效果不理想；

[0234] 如图6(e)中第4行、第5行的修复结果中人物的头发和眼睛不符合人脸特征；

[0235] PUT中P-VQVAE编码器将原分辨率的图像以非重叠的方式转换为潜在特征,避免了信息交叉影响,但对语义特征理解不充分；

[0236] 图6(f)中最后两行没有将填充区域与周围像素进行很好的融合修复,帽子的颜色不协调,眼镜的细节不够完善；

[0237] 与上述方法相比,本文算法在结合矢量量化的思想的同时引入了并行CSWinTransformer模块和多尺度特征引导模块,展现出边缘清晰、色彩过渡自然的修复效果,即使在破损严重的区域,修复的内容语义合理,没有出现不协调或突兀的局部；

[0238] 图7展示了各模型在places2数据集上的修复效果,EC、CTSDG由于无法捕获远距离特征,产生模糊且边界不一致的伪影,ICT由于下采样过程中大量信息丢失,修复的马腿存在缺陷；

[0239] MAT和PUT的修复结果出现语义不一致、颜色差异的现象；

[0240] 如图7(e)的第三行在草原上生成了一个柜子；

[0241] 图7(f)的第四行去除人物后的背景修复,生成的礁石不自然；

[0242] 本文方法通过共享码本学习,避免图像信息丢失,从而得到更丰富的语义信息,实现高保真的图像修复；

[0243] 图8为针对各类唐卡破损区域的修复对比图;EC算法在面对大面积缺损时,由于受制于较小的感受野,使得修复结果出现大范围的纹理模糊,且无法重构出图像结构；

[0244] CTSDG算法在应对人物局部缺失区域时,能够凭借边缘信息的优势重构出人物的基本外形轮廓,然而,在材质特性和微观细节层面上,其恢复程度并不理想；

[0245] 从图8(e)可以看出MAT算法在缺损面积较大时仍然展现出较强的修复能力,修复出了前两种算法在第二行和第四行的眼部区域无法修复的结果,但眼部位置还不合理,脸部存在畸变；

[0246] 从视觉上看出这五幅图通过本文算法的修复后,无论是结构连贯性还是纹理细节的精确性都与原图保持一致;因此,能够验证本文算法更适合于唐卡这类纹理复杂、色彩丰富的图片；

[0247] 由于不同个体的感受和评判标准存在差异,通过具体的数值对比能更准确地反映出细微的优劣差距,使得研究结果更具可验证性和重现性;选取了PSNR、SSIM、MAE和LPIPS四种评价标准,针对Celeba-HQ、Places2以及自制唐卡数据集上进行了实验；

[0248] 所有测试图像统一设定为 $256 \times 256$ 分辨率,并在这些图像中施加相同遮罩比例的不规则掩码;对比研究了EC、CTSDG、ICT、MAT、PUT等现有主流算法与本文所提出的算法模型,在此基础上统计得到了各项评价指标的具体数值,如表2所示；

[0249] 分析表2中的结果得出结论,在Places2场景数据集及自制唐卡数据集中,相较于

其他算法,本文算法无论是在像素级别还是结构层面,相似度方面均表现出显著优势;个别情况下,客观评价指标与直观视觉观察存在差异,这恰好证实了仅依赖单一客观或主观评价手段来衡量图像修复质量的局限性,同时,这也有力证明了本文结合两种评价方法进行综合评估的合理性与必要性;

[0250] 以下表2在具有不同掩码比率的三种数据集上,本文算法与EC、CTSDG、ICT、MAT、PUT的客观定量比较;

[0251] 表2

[0252]

Meth od	Dataset	PSNR ↑			SSIM ↑			LPIPS ↓			MAE ↓		
		20-4 0%	40-6 0%	Rand om	20-4 0%	40-6 0%	Rand om	20-4 0%	40-6 0%	Rand om	20-4 0%	40-6 0%	Rand om
EC	Celeba- HQ	26.6 11	21.8 75	21.50 4	0.92 8	0.81 8	0.799	0.06 85	0.14 2	0.144 8	0.02 16	0.04 81	0.036 6
CTS DG		28.3 83	24.2 44	23.21 6	0.95 1	0.87 4	0.845	0.06 09	0.14 01	0.135 4	0.01 65	0.03 53	0.028 5
ICT		27.7 81	23.2 65	22.74 2	0.94 8	0.85 1	0.864	0.05 13	0.10 87	0.106 6	0.02 12	0.04 25	0.031 9
MAT		29.7 37	25.1 58	24.65 7	0.96 2	0.90 6	0.89	0.05 25	0.09 33	0.102 1	0.01 74	0.03 35	0.025 9
PUT		28.2 32	23.6 03	24.59 9	0.95 4	0.87 1	0.898	0.04 15	0.10 28	0.094	0.01 76	0.03 84	0.025 4
本算 法		30.3 16	25.8 07	24.05 6	0.96 4	0.91 1	0.9	0.03 68	0.07 08	0.096 8	0.01 79	0.03 24	0.028 9
EC	Places2	23.2 48	19.2 36	20.77 8	0.84 3	0.66 1	0.742	0.13 17	0.24 74	0.179 2	0.03 1	0.06 36	0.046 7
CTS DG		19.2 62	19.0 55	22.20 6	0.79 8	0.68 7	0.786	0.17 34	0.24 3	0.192 2	0.05 46	0.06 28	0.038 8
ICT		22.7 94	18.2 57	20.09 7	0.83 4	0.62 4	0.725	0.11 7	0.22 47	0.172 9	0.03 57	0.07 37	0.053 7
MAT		23.9 82	18.8 82	21.42 1	0.86 1	0.67 2	0.77	0.14 85	0.25 53	0.195 5	0.03 26	0.06 9	0.048
PUT		22.8 3	17.9 53	20.91 4	0.84 3	0.65 3	0.783	0.08 08	0.21 14	0.141 9	0.02 76	0.06 66	0.042 6
本算 法		23.9 71	19.5 82	20.60 5	0.87 2	0.74 2	0.765	0.08 52	0.16 53	0.145 7	0.02 7	0.04 71	0.045 6
EC	唐卡	22.8 8	20.6 15	20.37 8	0.78	0.69 5	0.683	0.13 87	0.22 18	0.224 1	0.03 49	0.04 54	0.050 5
CTS DG		24.1 8	21.8 74	21.61 5	0.83 2	0.76 7	0.754	0.10 76	0.17 77	0.187 9	0.02 82	0.03 66	0.042
ICT		23.7 87	22.0 73	21.17 1	0.85 1	0.81 2	0.799	0.09 56	0.13 33	0.149 7	0.03 93	0.04 3	0.049 5
MAT		23.1 11	21.1 97	20.73 1	0.80 6	0.74 9	0.732	0.09 33	0.12 93	0.150 6	0.03 16	0.03 89	0.046 8
PUT		27.8 58	26.4 6	25.57 6	0.92 8	0.91 9	0.905	0.04 02	0.04 67	0.063 5	0.01 94	0.02 22	0.026 7
本算 法		30.0 64	29.2 6	27.81 1	0.96 2	0.95 6	0.947	0.02 09	0.02 46	0.035 2	0.01 81	0.01 89	0.022 6

[0253] 3.4消融研究：

[0254] 为了验证本申请所提出的方法各关键部件的有效性,本申请在自制唐卡数据集上

对其进行一系列的消融实验;主要包括如下实验:

[0255] (b) 本文CDCT模型的Encoder部分使用同等尺寸的Conv层来替换Linear层,

[0256] (c) 并行CSWinTransformer模块采用数量相同的标准Transformer模块来替换,

[0257] (d) 将PCT模块中标准自注意力和CSWin注意力之间的并行结构改为串行,

[0258] (e) 去除多尺度特征引导模块,

[0259] (f) 将多尺度特征引导模块中的LKA结构替换为Conv层,

[0260] (g) 实现本申请的完整网络结构;

[0261] 表3展示了不同组件消融研究的客观评价结果;变体1和变体2使用源自VQGAN的编码器以及标准Transformer模块,使得信息压缩过度以及局部细节利用不充分,影响了模型的表现;变体3是将PCT模块内部结构的并行方式改为顺序方式,指标小幅下降;

[0262] 变体4和变体5证明了多尺度特征引导模块在充分利用非掩码区域特征的同时,保持对潜在表征的译码能力;加入线性残差编码器模块、并行CSWinTransformer模块和多尺度特征引导模块的完整模型,相对于其他替换组件来说,PSNR值和SSIM值平均提升1.741dB和0.038,LPIPS值以及MAE值平均下降0.0221和0.0053;这表明这些改进模块对修复结果的质量有积极影响;

[0263] 以下表3为本文方法在自制唐卡数据集上的定量消融分析:

[0264] 表3

[0265]

类型	组件	PSNR/dB ↑	SSIM ↑	LPIPS ↓	MAE ↓
变体 1	+卷积编码器	25.792	0.901	0.0583	0.0335
变体 2	+标准Transformer 模块	25.938	0.904	0.0560	0.0254
变体 3	+串行CSWinTransformer 模块	27.318	0.924	0.0403	0.0225



	块				
变体 4	—多尺度特征引导模块	21.453	0.781	0.1227	0.0643
变体 5	+卷积引导模块	26.650	0.913	0.0509	0.0234
Our	本文完整结构	28.165	0.948	0.0292 <sub>9</sub>	0.020

[0267] 图9给出了本文模型各组件的可视化结果；

[0268] 由图9 (b) 所示,变体1破损区域与周围区域缺乏一致性,人物胳膊、面部以及胸前的肤色存在明暗色差;变体2和变体3能够看出人物手中的串珠大小不均,出现伪影问题;如图9 (e) 所示,去除多尺度特征引导模块后,修复结果的局部有效信息减少,人物手指受到周围蓝色背景的影响,图像眼睛的边缘结构存在扭曲,过渡不自然;

[0269] 如图9 (g) 所示,验证了本文所提出的CDCT算法在处理色彩复杂图像问题上的有效性和优越性,使用该方法获得更为逼真、更具合理性的修复效果;

[0270] 在第一阶段网络中,提出的模型将连续特征嵌入到一个具有有限大小的离散空间中,即k个码向量;本实施例进行了一项消融研究,以了解码本中的码向量个数(k)对模型性能的影响;表4展示了在唐卡数据集上码本大小为1024时,产生更好的结果,改善重建质量更有效,而不是码本向量越大数据压缩越合理;

[0271] 表4不同码本大小对模型性能的影响:

码本大小 (k)	PSNR/dB↑	SSIM↑	LPIPS↓	MAE↓
512	26.033	0.839	0.0491	0.0250
1024	27.868	0.889	0.0311	0.0208
2048	26.889	0.868	0.0414	0.0216

[0273] 本实施例通过5组实验来确定Attentionhead和embeddingdimension最佳的超参数设置;

[0274] 当PCT模块的注意力头部设置为8,嵌入维度设置为512时,模型能够更好地捕捉输入序列中的长距离依赖关系,四种评价指标显著提高,同时避免嵌入维度过高增加模型计算负担,如表5所示;

[0275] 表5PCT模块不同超参数组合的性能

Heads	Embeddingdms	Params (M)	PSNR/dB↑	SSIM↑	LPIPS↓	MAE↓
4	512	53.05	26.684	0.892	0.0420	0.0218
8	512	53.05	27.752	0.908	0.0302	0.0200
16	512	53.05	26.158	0.874	0.0471	0.0231
8	256	20.95	23.487	0.822	0.0884	0.0292

8	768	106.13	26.174	0.875	0.0470	0.0230
---	-----	--------	--------	-------	--------	--------

[0277] 本实施例提出了一种离散码本和Transformer协同的图像修复方法,它具有多个新的设计特点;

[0278] 首先,利用线性编码器代替卷积下采样,特征块之间独立编码,避免信息交叉影响,不同于常规修复模型,本实施例使用codebook来离散编码模型中间特征;其次,为了避免Transformer中的信息损失,输入到Transformer中的不是离散的token,即索引,而是encoder出来的特征;

[0279] 同时,离散的token只用作Transformer的输出;另外,并行CSwinTransformer模块的设计,提升token预测的准确度也降低的参数量;随后,在解码器上加入了一个额外的多尺度特征引导模块,能够更好地保留非缺损区域的局部细节,并从编码器的量化输出中恢复细节;

[0280] 通过在多个具有代表性的任务上的广泛实验,验证了CDCT方法既能处理色彩多样、语义丰富的唐卡图像,也能有效修复自然图像中的各种缺损;通过深入的消融研究,展示了所提模型设计的有效性;旨在精准识别并修复唐卡图像中的残缺部分,以此作为优化图像复原工作的新方向。

[0281] 实施例八:

[0282] 一种唐卡图像修复系统,所述系统用于执行所述的方法。

[0283] 实施例九:

[0284] 一种存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现所述唐卡图像修复方法的步骤。

[0285] 上述公式均是去量纲取其数值计算,公式是由采集大量数据进行软件模拟得到最近真实情况的一个公式,公式中的预设参数由本领域的技术人员根据实际情况进行设置。

[0286] 上述实施例,可以全部或部分地通过软件、硬件、固件或其他任意组合来实现。当使用软件实现时,上述实施例可以全部或部分地以计算机程序产品的形式实现。本领域技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够通过电子硬件,或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方法来执行,取决于技术方案的特定应用和设计约束条件。

[0287] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,既可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0288] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。

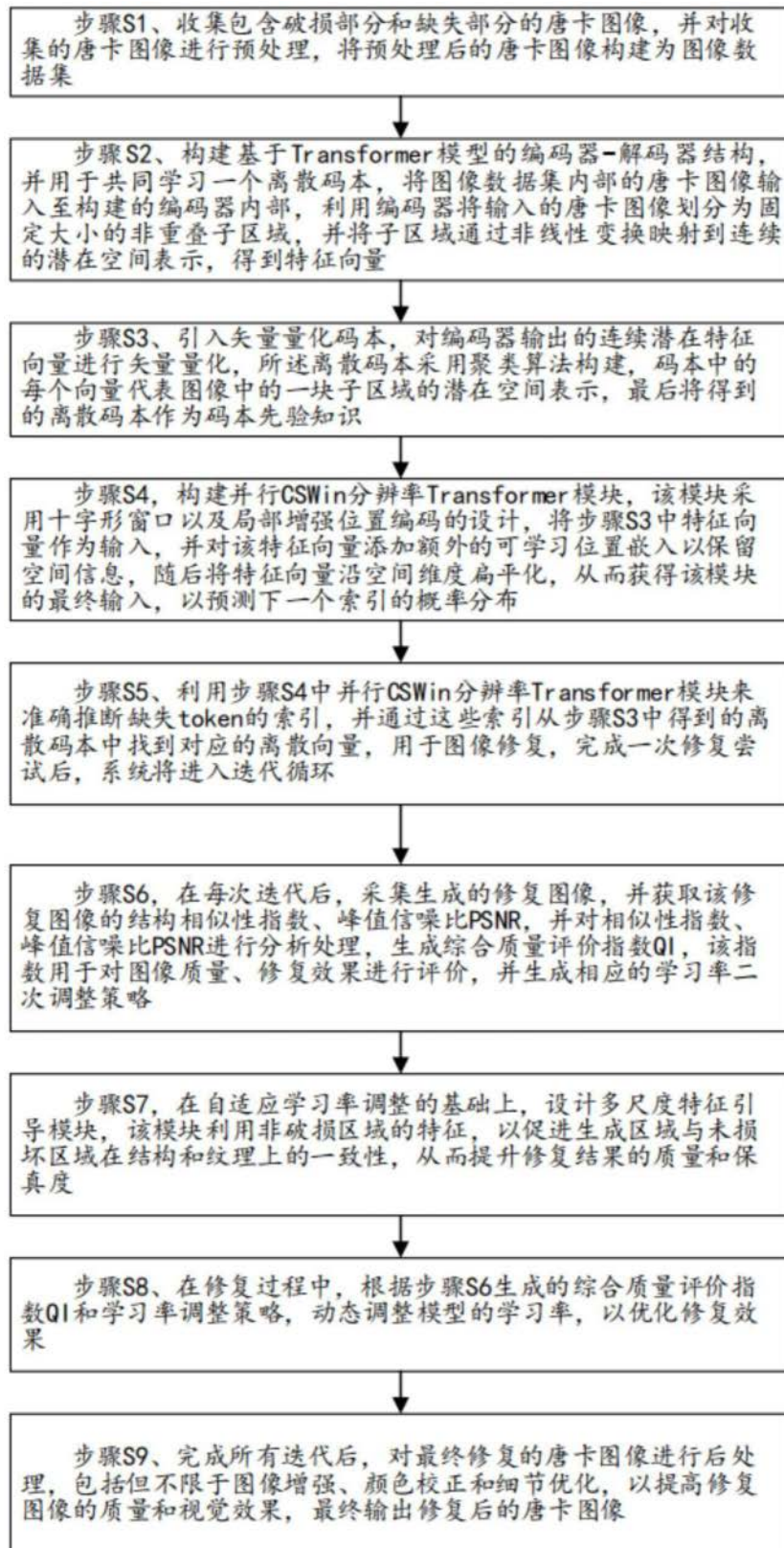


图1

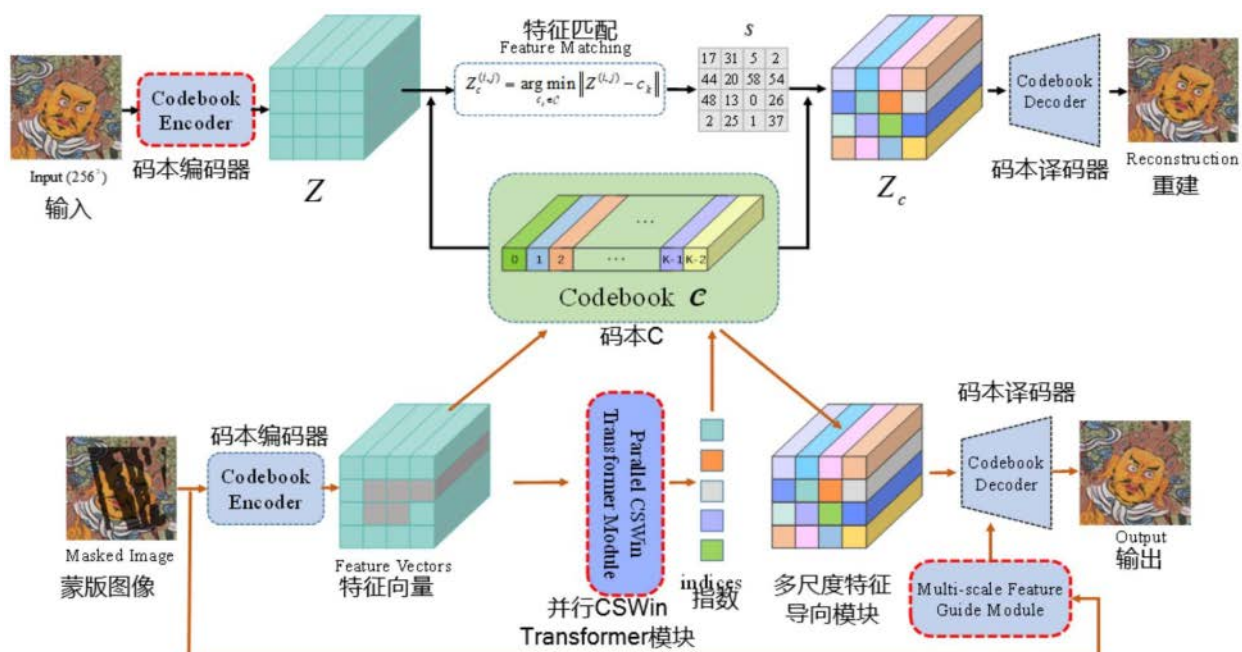


图2

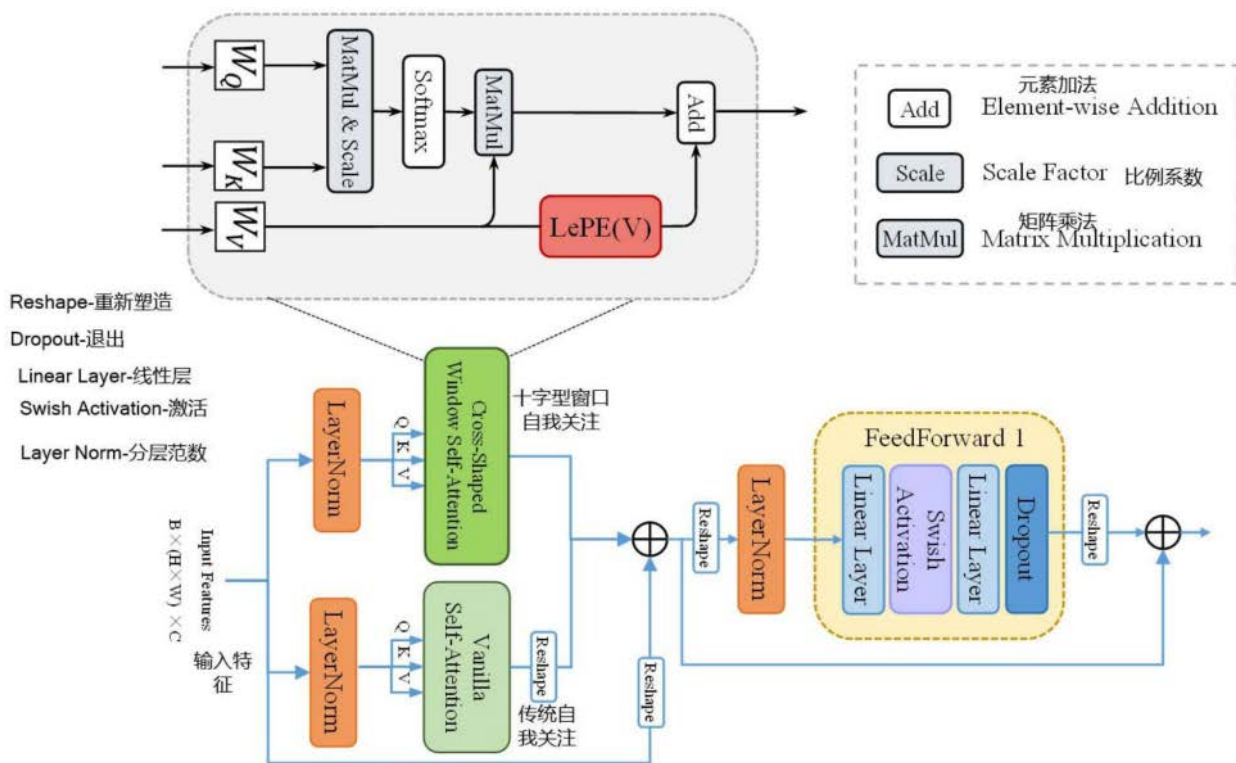


图3

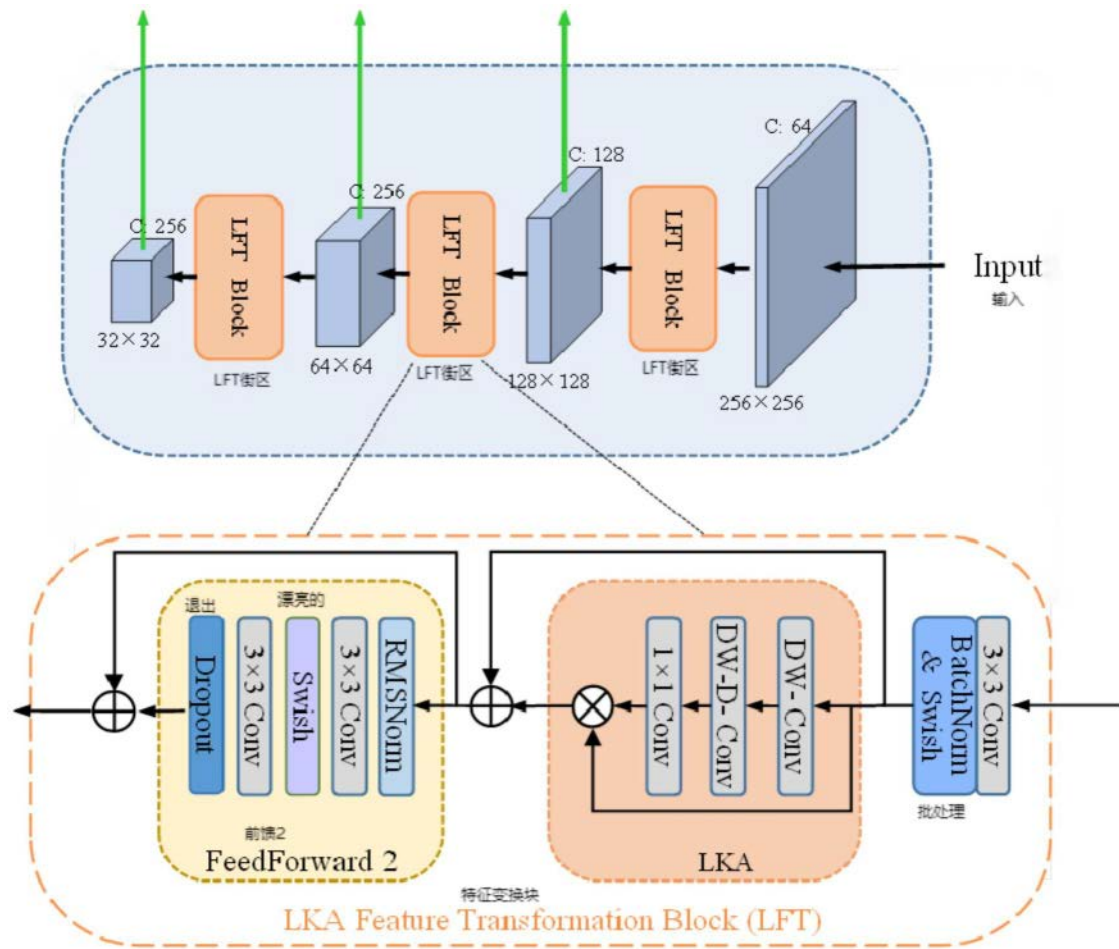


图4



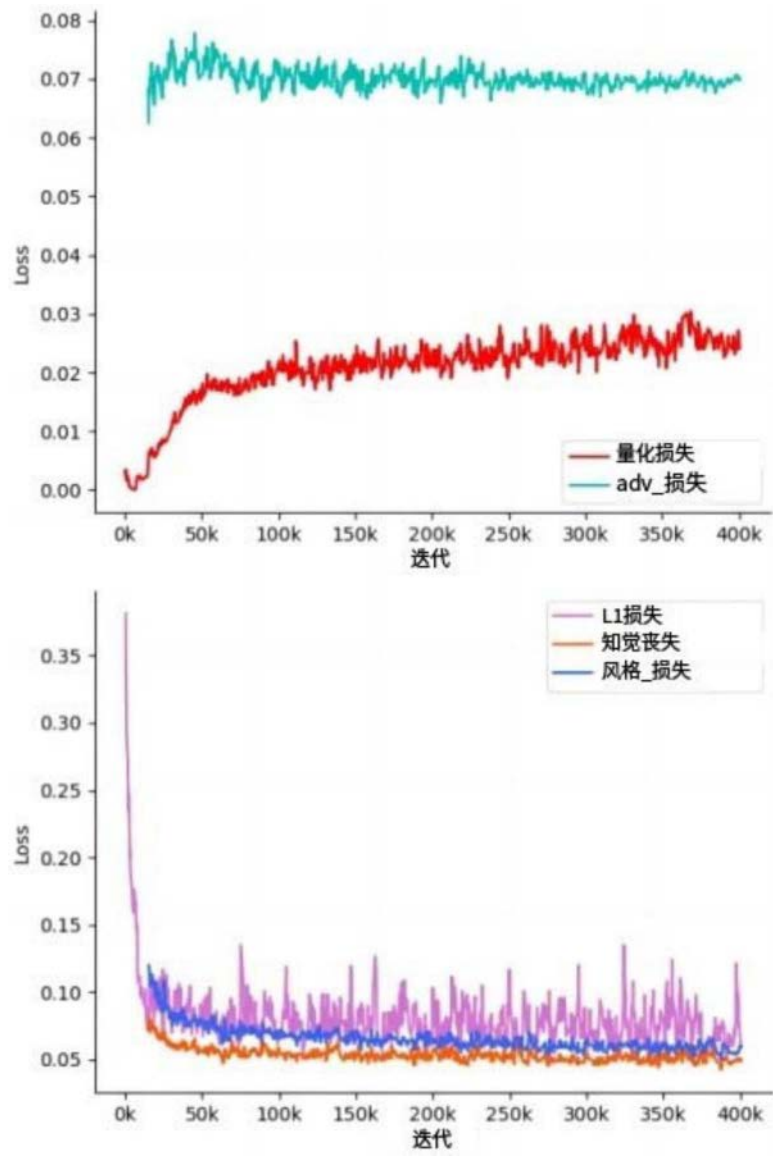


图5



图6

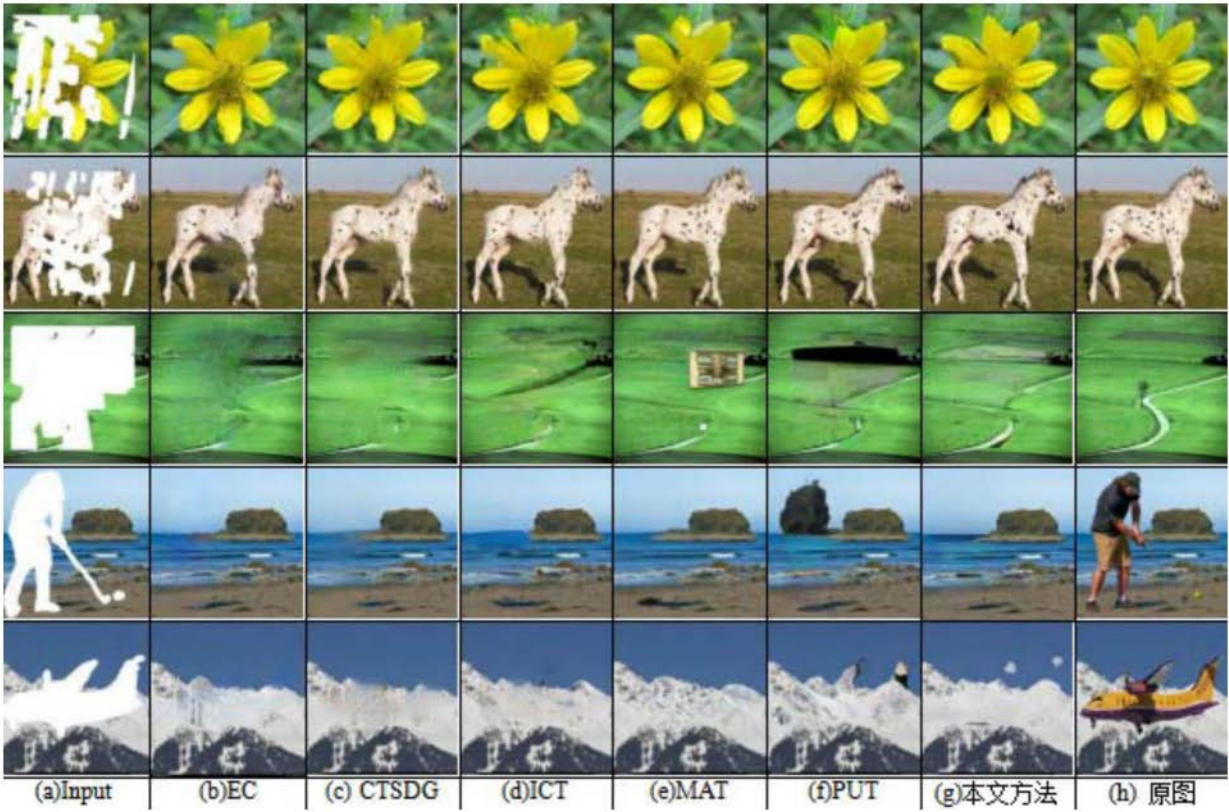


图7



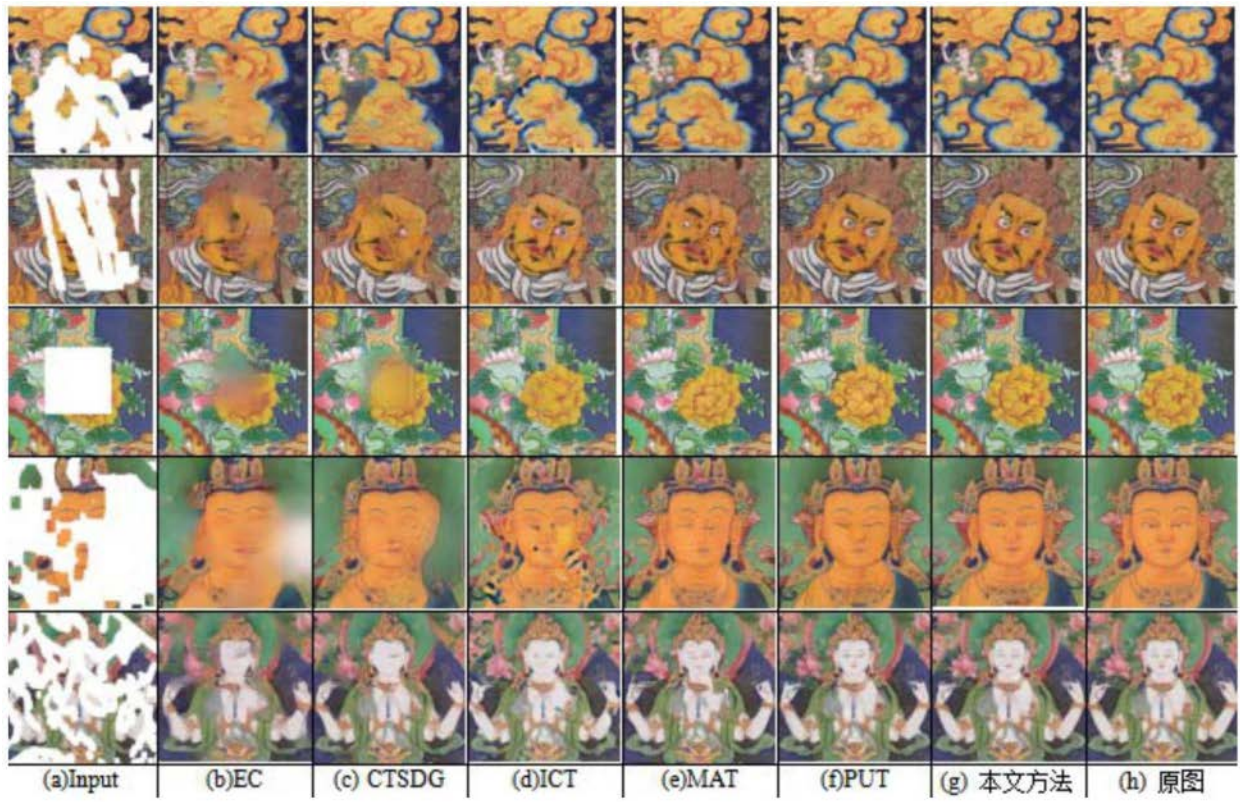


图8

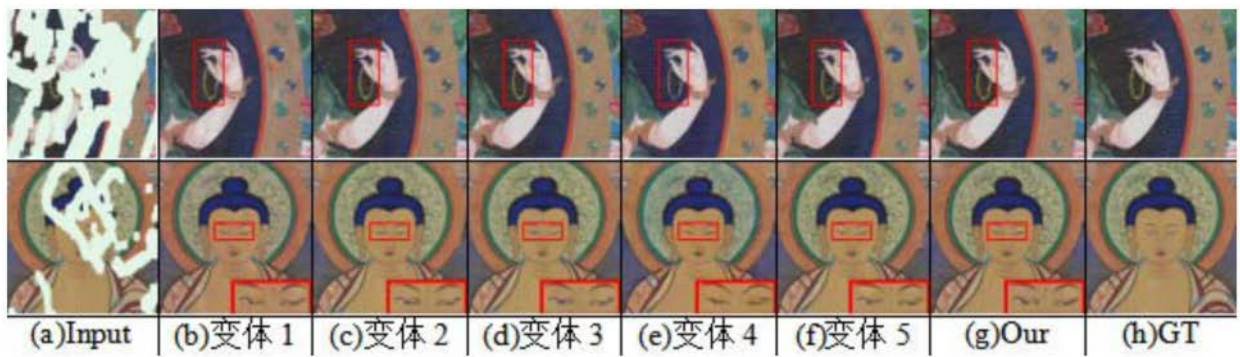


图9