融合特征位置编码和误差修正的唐卡图像风格迁移模型

时新月[1)], 胡文瑾[2)*], 乔浪[1)], 康文东[1)]

[1)] (语言与文化计算教育部重点实验室　兰州　730030)
[2)] (西北民族大学数学与计算机科学学院　兰州　730030)
(wenjin_zhm@126.com)

摘　要: 唐卡是藏族文化中一种独具特色的艺术风格, 通过对唐卡图像进行二次创作能够帮助人们了解唐卡风格特征, 促进文化传承与保护. 针对目前唐卡图像风格迁移技术生成新图像中, 生成图像存在伪影、局部图像模糊和细节信息处理不到位等问题, 提出一种融合特征位置编码和误差修正的唐卡风格迁移模型 FPC-EI. 利用二维相对位置编码获取内容序列与风格序列之间的对应位置信息, 实现图像特征之间的对齐和匹配; 坐标注意力机制模块强化利用二维相对位置编码增强编码器对于唐卡以及内容图像纹理细节信息的捕获能力; 密集深度反投影网络使用采样层模块不断地进行上下采样, 进行误差修正, 指导模型提高生成图像质量. 在自建的唐卡数据集和任意图像数据集上进行模型训练的实验结果表明, 与 AdaIN, ArtFlow, AdaAttN 等风格迁移模型相比, FPC-EI 的 SSIM 指标平均提升 32%, LPIPS 和 MSE 指标平均下降 7.3%和 2.7%, 可以有效地保留内容结构、特征融合和图像局部细节方面特征.

关键词: 唐卡图像; 风格迁移; 位置编码; 注意力机制; 深度反投影网络
中图法分类号: TP391.41　　**DOI:** 10.3724/SP.J.1089.2023-00068

A Style Transfer Model for Thangka Images Based on Fusion of Feature Position Coding and Error Correction

Shi Xinyue[1)], Hu Wenjin[2)*], Qiao Lang[1)], and Kang Wendong[1)]

[1)] (*Key Laboratory of Linguistic and Cultural Computing Ministry of Education, Lanzhou*　730030)
[2)] (*School of Mathematics and Computer Science, Northwest Minzu University, Lanzhou*　730030)

Abstract: Thangka is a unique art style in Tibetan culture, and the secondary creation of thangka images can help people understand the stylistic characteristics of thangka, and promote the cultural heritage and protection. Aiming at the problems of artifacts, local image blurring and inappropriate processing of detail information in the new image generated by the current thangka image style migration technology, we propose a thangka style migration model FPC-EI, which integrates the feature position encoding and error correction. The two-dimensional relative position encoding is used to obtain the corresponding position information between the content sequences and the style sequences, and to realize the alignment and matching of the image features; the coordinate attention mechanism module reinforces the use of two-dimensional relative position encoding to get the position information between content sequences and style sequences, and realizes the alignment and matching of image features. The coordinate attention mechanism module enhances the ability of the encoder to capture the texture details of the Tangkas and the content images by utilizing 2D

收稿日期: 2023-05-04; 修回日期: 2024-01-11. **基金项目**: 国家自然科学基金(62061042, 61862057); 国家民委创新团队计划(2018[98]号); 中央高校建设世界一流大学(学科)和特色发展引导专项资金. **时新月**(1998—), 女, 硕士, 助教, 主要研究方向为计算机视觉、图像风格迁移; **胡文瑾**(1981—), 女, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为计算机视觉、图像处理; **乔浪**(1997—), 男, 硕士, 助教, 主要研究方向为计算机视觉、图像文本描述; **康文东**(1998—), 男, 硕士, 主要研究方向为计算机视觉、图像文本描述.

relative position coding, and the dense deep inverse projection network uses a module that continuously samples up and down the sampling layer to correct the error and guide the model to improve the quality of the generated images. The experimental results of model training on the self-constructed thangka dataset and arbitrary image dataset show that compared with the style migration models such as AdaIN, ArtFlow, AdaAttN, etc., the SSIM index of FPC-EI is improved by 32% on average, and the LPIPS and MSE indexes are reduced by 7.3% and 2.7% on average, and the structure of the content, the fusion of the features and the local details of the image can be effectively preserved. The FPC-EI can effectively retain the content structure, feature fusion and local details of the image.

Key words: Thangka images; style transfer; position coding; attention mechanism; depth back projection network

唐卡图像蕴含极为丰富的历史、文化和艺术价值, 它不仅是中国传统文化的重要组成部分, 也是中华民族的宝贵财富. 唐卡绘画构图饱满, 结构上主题鲜明, 线条流畅细密富有变化, 具有极强的立体感与质感; 此外, 色彩的运用也是唐卡绘画的独特之处之一, 它常用明亮且厚重的色彩表现极强的主观性, 给观者带来强烈的视觉冲击和心灵震撼. 久而久之, 唐卡绘画在世界美术领域已自成体系, 独树一帜. 随着人工智能技术的蓬勃兴起, 使用现代科学技术对唐卡图像特征进行提取及再利用, 将唐卡艺术风格与其他派别图像结合, 为艺术作品的二次创作提供了新的思路, 不仅有助于保护和传承唐卡艺术作为文化遗产的独特价值, 而且还可以为产品和品牌提供独特的表现能力. 综上所述, 唐卡图像风格迁移研究对于跨文化交流和理解以及文化产业的多元化发展具有重要的意义.

图像风格迁移是提取风格图像中的绘画风格, 再将其转移到内容图像中, 该技术在保留内容的同时可以生成一幅新绘画风格的图像. 传统的图像风格迁移方法只提取目标图像的低层次特征, 忽略了高层语义信息, 模型泛化性差. 为解决该问题, Gatys 等[1-4]利用深度神经网络分别提取内容特征和风格特征, 通过图像重建生成风格化图像; 2014 年, Goodfellow 等[5]首次提出利用生成对抗网络(generative adversarial networks, GAN)进行风格迁移并受到了广泛关注; Li 等[6]通过 GAN 中的判别网络训练一个前馈网络来取代图像块的匹配过程, 加快了模型的运算速度; Zhu 等[7]在 GAN 的基础上提出循环一致对抗网络(cycle GAN, CycleGAN), 使用 2 个生成器和 2 个判别器实现风格域与内容域之间的转换, 并利用循环一致性损失函数确保生成图像的质量. 然而, CycleGAN 只能生成特定风格图像, 结果缺乏多样性. 基于白化和着色变换(whitening and coloring transforms, WCT)[8]模型利用内容图像的协方差与给定风格图像的特征协方差进行匹配, 因而不需要对每个单独的风格图像进行学习. 尽管现有工作可以提高生成图像风格的多样性, 但在一定程度上限制了模型的泛化性. 为此, Wang 等[9]提出可实现多样化任意风格迁移模型, 使用正交的随机噪声矩阵扰动图像的深度特征映射, 增加图像风格的多样性; 受条件实例归一化(conditional instance normalization, CIN)模型[10]的启发, Huang 等[11]提出自适应实例归一化(adaptive instance normalization, AdaIN), 通过在特征空间中传递信道的均值和方差统计量进行风格迁移, 首次实现了实时的任意风格迁移. 尽管基于归一化的模型在实现风格多样化上有很大的提升, 但仍缺乏将不同纹理特征转移到内容图像中不同空间位置的能力. Park 等[12]发现, 基于注意力机制的风格迁移模型可以根据内容特征重新排列风格特征的过程, 来学习内容特征和风格特征之间的语义相关性, 并获取关键特征信息进行传递, 在保留内容结构的同时可实时合成高质量的图像. 这方面的典型代表有: 多线程注意力机制(attention aware multi-stroke, AAMS)[13]使用自注意力机制并交换样式以直接匹配内容结构; 自注意力网络(self-attention networks, SANet)[12]根据内容图像的语义空间分布对风格模式进行整合; 多适应网络(multi-adaptation network, MANet)[14]使用自注意力模块对内容和风格特征进行分离, 再使用交叉注意操作, 根据内容分布对风格特征进行重新组织. 然而, 这些方法有时会使输出图像中存在不必要的特征信息, 出现内容表示错误. 为了解决该问题, 迭代误差反馈[15]通过自校正模型将误差特征进行反馈, 通过扩展特征提取器的表达能力逐步修

正输出结果. Song 等[16]提出误差转换网络(error transition network, ETNet), 通过将误差引入前向网络并对齐误差特征与当前风格化结果, 预测当前风格化的错误, 再对其进行更新得到结果.

目前, 基于 Transformer 的图像风格迁移算法在计算机视觉中得到了广泛关注. 这是因为在自注意力机制的帮助下, 基于 Transformer 的图像风格迁移模型[13]可以很容易地学习输入全局信息, 从而在每一层都可以获得对输入的整体理解; 其次, 该模型具有较强的特征表示能力, 可以避免特征提取过程中细节的丢失, 最大限度地保留生成图像的结构. 但是, 该模型提取图像特征不充分, 使得最终生成图像出现内容图像模糊、伪影的现象. Deng 等[17]提出基于 Transformer 的图像风格迁移(image style transfer with Transformers, StyTr^2)模型, 由 Transformer 内容编码器、Transformer 风格编码器和 Transformer 解码器 3 部分组成. 首先将内容图像和风格图像分别组成一组序列; 其次将添加了内容感知位置编码的内容图像序列传送到 Transformer 内容编码器中, 风格序列传入 Transformer 风格编码器中, 通过 2 个编码器获取内容图像以及风格图像特征信息; 最后将其传送到多层 Transformer 解码器中进行特征融合, 获得风格迁移图像.

尽管以上模型在图像风格迁移中获得了良好的效果, 但对唐卡图像进行风格迁移时发现, 最终得到的结果并不理想. 分析原因如下: (1) 唐卡图像内容结构复杂、线条流畅且细致, 现有模型对唐卡图像特征提取不充分, 故在生成图像中唐卡风格不明显; (2) 唐卡图像的低层次特征也表达着至关重要的细节信息, 现有的编码器对唐卡进行细节信息提取时易丢失这部分关键信息, 导致生成图像出现局部内容模糊、信息丢失、伪影的问题; (3) 唐卡图像色彩鲜明, 具有较强的视觉冲击力, 现有的特征融合任务在进行特征传递过程中不能有效地调节风格特征, 使其与内容结构相融合时效果不佳, 最终出现伪影或内容结构被破坏的问题. 针对以上问题, 本文在 StyTr^2 的基础上, 提出融合特征位置编码和误差修正的唐卡风格迁移模型(a style transfer model for Thangka images based on fusion of feature position coding and error correction, FPC-EI).

唐卡具有丰富且浓厚的色彩和复杂的绘画工艺, 为了更好地获取风格图像的色彩特征以及绘画风格, 本文设计了突出特征位置信息的编码器. 利用二维相对位置编码(2D relative position encoding, IRPE)[18]对内容图像的语义信息和唐卡风格特征进行对齐和匹配, 解决特征提取不充分、图像风格迁移后的唐卡风格不明显的问题. 但是, 由于编码器在特征提取过程中仍会丢失图像低层次特征信息, 如唐卡图像的局部色彩和纹理细节, 内容图像的局部结构和边缘特征, 因此引入坐标注意力(coordinate attention, CoordAtt)机制[19], 强化利用 IRPE 使得编码器在提取关键特征时能有效地捕获低层次特征信息, 解决了生成图像中存在局部内容模糊、伪影的问题.

唐卡图像纹理复杂, 容易导致生成图像中出现边缘纹理不对齐、画质扭曲的现象, 因此添加密集深度反投影网络(dense deep back-projection networks, D-DBPN)[20]. 通过不断地上下采样, 利用迭代误差反馈修改采样结果; 然后累积每个上采样阶段的自校正特征实现纹理对齐, 创建高分辨率图像, 解决了生成图像中部分内容结构被破坏的问题, 进一步提高生成图像的质量.

1　FPC-EI 模型

FPC-EI 模型以 StyTr^2 为基本框架, 由基于特征位置信息的编码器、Transformer 解码器和 D-DBPN 组成, 其结构如图 1 所示.

(1) 对内容图像和唐卡风格图像进行分割, 使用线性投影得到图像序列, 并将其输入到基于特征位置信息的编码器中进行特征提取. 由于编码器采用 IRPE 方式, 因此可以获得内容序列与唐卡

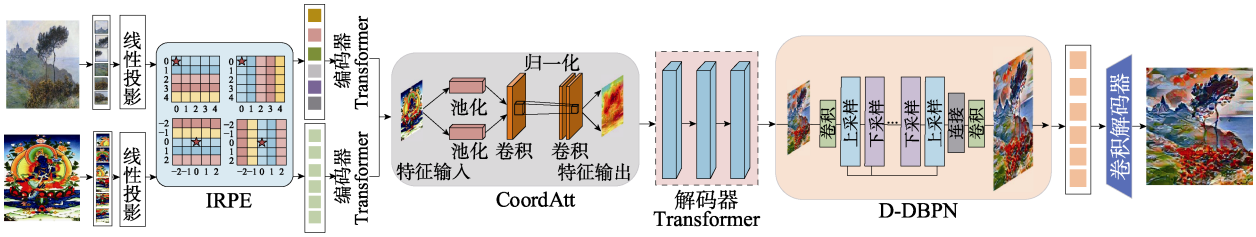


图 1　FPC-EI 模型结构

序列之间的对应位置信息；考虑编码器对唐卡和内容图像进行细节信息提取时易丢失图像边缘、纹理等低层次特征，采用 CoordAtt 机制更精确地获取低层次的特征信息，可有效地减少模型特征提取时出现的细节信息丢失问题. (2) 经过特征位置编码后，采用多层 Transformer 解码器对内容图像融合唐卡风格特征生成新的图像. (3) 为了有效地调节风格特征，避免生成图像内容结构被破坏的问题，采用 D-DBPN 对解码器生成的图像多次上下采样，获得误差预测反馈用以修改采样结果，实现生成图像的误差修正；通过逐步修正输出结果，进一步提高风格迁移图像的质量. (4) 使用卷积解码器对图像细化，获得最终的输出结果.

1.1 突出特征位置信息的编码器

唐卡绘画作品构图严谨，结构上极善于描绘细节、笔触自然流畅、富有极强的律动与视觉张力；在色彩上，喜好用厚重、单纯的原色，形成了色彩饱满、对比强烈的艺术效果，已经成为中国传统文化中独具特色的艺术形式. StyTr^2 模型中的编码器在进行风格迁移时，存在唐卡局部风格特征获取不足的问题，原因是编码器对图像序列进行编码计算时，无法确保内容图像序列与唐卡风格序列之间的对应关系，也会丢失唐卡低层次特征信息，导致生成图像存在模糊、伪影的现象. 综上所述，如何解决生成图像语义信息与风格特征不匹配的问题，增强模型对于唐卡风格的捕获能力，是设计编码器时需要着重考虑的问题. 为此，本文设计了突出特征位置信息的编码器，其结构如图 2 所示. 唐卡色彩丰富且厚重、具有极为精细的纹理特征，为了帮助模型充分地提取图像特征，编码器利用 IRPE 获取内容序列和风格序列之间的对应位置信息，使得模型在内容结构中可根据唐卡的绘画风格实现特征之间的对齐和匹配；另外，图像低层次特征也表示着至关重要的信息，如唐卡边缘细节、内容局部结构语义信息等. 为了解决特征提取时低层次特征信息丢失的问题，在编码器中增加 CoordAtt 机制，针对 IRPE 获得图像序列之间的位置编码进行强化利用，使其更加精确地获取图像低层次特征位置信息. 通过这种方式，编码器能够更好地捕获唐卡风格以及内容语义信息，实现特征对齐和匹配，确保编码器在特征提取过程中充分提取图像特征，使得模型能够更好地生成具有唐卡风格的图像. 突出特征位置信息的编码器将内容序列和唐卡序列作为输入，实现过程如图 2 所示.

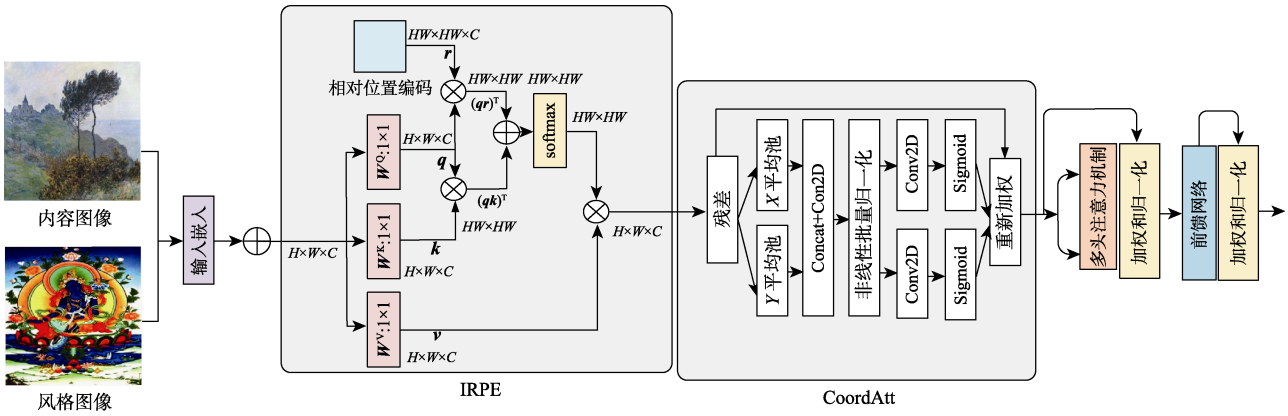


图 2 突出特征位置信息的编码器结构

1.1.1 IRPE

在特征提取过程中，编码器无法获取图像序列之间的对应关系，不能实现特征对齐与匹配，导致最终生成图像唐卡风格特征不明显. 为了解决以上问题，在编码器中利用 IRPE，通过定向相对距离建模，以及查询值和相对位置嵌入之间的相互作用，获取图像序列之间的相对位置信息，完成相似区域块之间的匹配，实现特征对齐. 该位置编码包括上下文模式、分段索引和交叉法.

(1) 上下文模式. 在每个位置处添加一个可学习的位置向量 $\boldsymbol{b}_{ij}=(\boldsymbol{x}_i\boldsymbol{W}^{\mathrm{Q}})\boldsymbol{r}_{ij}^{\mathrm{T}}$ 实现位置编码. 其中，$\boldsymbol{b}_{ij}\in\mathbb{R}$；$\boldsymbol{r}_{ij}\in\mathbb{R}^{d_z}$ 是一个可训练向量，与查询(Q)嵌入相互作用. 在上下文模式下，$\boldsymbol{b}_{ij}$ 有多个变体，例如，在键(K)上操作的相对位置编码可以表示为 $\boldsymbol{b}_{ij}=(\boldsymbol{x}_i\boldsymbol{W}^{\mathrm{Q}})(\boldsymbol{r}_{ij}^{\mathrm{K}})^{\mathrm{T}}+(\boldsymbol{x}_j\boldsymbol{W}^{\mathrm{K}})(\boldsymbol{r}_{ij}^{\mathrm{Q}})^{\mathrm{T}}$. 其中，$\boldsymbol{r}_{ij}^{\mathrm{K}},\boldsymbol{r}_{ij}^{\mathrm{Q}}\in\mathbb{R}^{d_z}$ 都是可学习向量. 如图 2 中 IRPE 所示.

上下文模式中的位置向量不仅包含当前位置信息，还包括邻域信息，因此能更好地捕捉序列之间的关系，并对序列之间相对特征位置关系进行

编码.

(2) 分段索引. 相对位置编码中的索引函数无法捕捉远距离相对位置的上下文信息. 本文采用分段索引函数 $\boldsymbol{g}(x):\mathbb{R}\to\{y\in\mathbb{Z}\mid -\beta\leqslant y\leqslant\beta\}$ 用于索引上下文信息到图像序列相应编码的相对距离, 并根据相对距离分配注意力, 实现序列之间对应特征位置关系的匹配, 表示为

$$\boldsymbol{g}(x)=\begin{cases}[x], & |x|\leqslant\alpha\\ \operatorname{sign}(x)\times\\ \min\left(\beta,\left[\alpha+\arcsin^{-1}\theta\dfrac{\ln(|x|/\alpha)}{\ln(|\gamma|/\alpha)}(\beta-\alpha)\right]\right), & |x|>\alpha\end{cases}.$$

其中, $[\cdot]$ 是一个整数运算; $\operatorname{sign}(\cdot)$ 确定数字的符号, 正输入返回 1, 负输入返回−1, 否则返回 0; α 决定分段点; β 控制在 $[-\beta,\beta]$ 的输出; γ 调节对数部分的曲率.

分段索引函数将相对位置划分为多个区间, 并将其映射到不同的编码空间, 可以提高模型在特征提取时的稳定性和准确性.

(3) 交叉法. 首先计算水平方向和垂直方向上的位置编码, 然后对其进行汇总, 可有效地利用像素的方向信息, 提高风格迁移图像的质量. 计算公式为

$$\boldsymbol{r}_{ij}=\boldsymbol{P}^{\tilde{x}}_{\boldsymbol{I}^{\tilde{x}}(i,j)}+\boldsymbol{P}^{\tilde{y}}_{\boldsymbol{I}^{\tilde{y}}(i,j)},$$
$$\boldsymbol{I}^{\tilde{x}}(i,j)=\boldsymbol{g}(\tilde{x}_i-\tilde{x}_j),$$
$$\boldsymbol{I}^{\tilde{y}}(i,j)=\boldsymbol{g}(\tilde{y}_i-\tilde{y}_j).$$

其中, $\boldsymbol{P}^{\tilde{x}}_{\boldsymbol{I}(i,j)}$ 和 $\boldsymbol{P}^{\tilde{y}}_{\boldsymbol{I}(i,j)}$ 是上下文模式下的可学习向量; x 轴和 y 轴上相同的偏移量共享相同的编码, 如图 1 中 IRPE 所示.

通过比较特征点之间的相对位置关系, 交叉法可以更加准确地捕捉到图像序列之间的特征信息, 增强感兴趣对象的特征表示, 实现唐卡图像中代表风格的色彩、纹理信息与内容图像中语义信息的对齐.

1.1.2 CoordAtt 机制

IRPE 能够帮助编码器实现图像序列之间特征对齐和匹配, 在风格迁移任务中起到重要作用. 但由于唐卡内容结构复杂, 即使有 IRPE 的辅助, 编码器仍然会在特征提取过程中丢失图像低层次特征信息. 为解决以上问题, 使得最终生成图像内容清晰且减少伪影, 本文利用注意力机制强调或选择目标处理对象的重要信息, 抓住关键特征信息进行传递. 因此, 在风格迁移模型中引入 CoordAtt 机制, 通过强化利用 IRPE 的位置编码获得坐标信息调整每个特征点的权重, 增加具有图像低层次特征的位置信息点, 帮助模型更好地提取图像纹理细节信息.

CoordAtt 模块分为坐标信息嵌入和 CoordAtt 生成 2 个步骤, 如图 2 中 CoordAtt 所示. 如果直接将全局信息压缩到信道描述符中, 难以保存图像位置信息. 为了使注意力块精确地获取图像的位置信息, 在空间上实现图像特征之间远程交互, 本文将全局池化进行分解[19], 第 c 通道相关的输出为

$$\boldsymbol{z}_c=\frac{1}{H\times W}\sum_{j=1}^{H}\sum_{j=1}^{W}\boldsymbol{x}_c(i,j).$$

其中, 输入 $\boldsymbol{x}_c$ 来自具有固定内核大小的卷积层, 可以被认为是局部描述符的集合. 给定输入 $\boldsymbol{x}$, 使用池化核的 2 个空间范围 $(H,1)$, $(1,W)$ 分别沿水平方向和垂直方向对每个通道进行编码. 因此, 第 c 通道在高度 h 处的输出可表示为

$$\boldsymbol{z}^h_c(h)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}\boldsymbol{x}_c(i,j)\tag{1}$$

类似地, 第 c 通道在宽度 ω 处的输出可以写成

$$\boldsymbol{z}^\omega_c(h)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}\boldsymbol{x}_c(j,\omega)\tag{2}$$

以上方式使注意力捕捉到了一个空间方向上的长期依赖关系, 在获取风格图像的风格特征与内容图像的语义信息的同时, 保留另一个空间方向上的精确位置信息.

上述处理过程可获得精确的位置信息, 接下来生成 CoordAtt. 首先将式(1)(2)中生成的聚合特征映射连接起来, 发送到共享的1×1卷积变换函数 $\boldsymbol{F}_1$ 中[19], 得到 $\boldsymbol{f}=\delta\left(\boldsymbol{F}_1\left(\left[\boldsymbol{z}^h,\boldsymbol{z}^\omega\right]\right)\right)$. 其中, $[\cdot,\cdot]$ 表示沿空间维数的拼接运算; δ 表示非线性激活函数 Sigmoid; $\boldsymbol{f}\in\mathbb{R}^{C/r\times(H+W)}$ 表示对水平方向和垂直方向的空间信息进行编码的中间特征图. 把 $\boldsymbol{f}$ 沿着空间维度分为 2 个单独的张量 $\boldsymbol{f}^h\in\mathbb{R}^{C/r\times H}$ 和 $\boldsymbol{f}^\omega\in\mathbb{R}^{C/r\times W}$; 将另外 2 个1×1卷积变换 $\boldsymbol{f}^h$ 和 $\boldsymbol{f}^\omega$ 变换为与输入 $\boldsymbol{X}$ 通道数相同的张量, 得到

$$\boldsymbol{g}^h=\sigma\left(\boldsymbol{F}_h\left(\boldsymbol{f}^h\right)\right),$$
$$\boldsymbol{g}^\omega=\sigma\left(\boldsymbol{F}_\omega\left(\boldsymbol{f}^\omega\right)\right).$$

其中, σ 是 Sigmoid 函数. 将输出 $\boldsymbol{g}^h$ 和 $\boldsymbol{g}^\omega$ 展开, 分别用作注意权重, 则坐标注意块 $\boldsymbol{Y}$ 的输出为

$$\boldsymbol{y}_c(i,j)=\boldsymbol{x}_c(i,j)\times\boldsymbol{g}^h_c(i)\times\boldsymbol{g}^\omega_c(j).$$

因此, 沿水平方向和垂直方向的注意力同时应

用于输入张量, 图 2 中 CoordAtt 的每个元素都反映了目标对象是否存在于相应的行和列中. 该过程最大化地获取了低层次关键特征信息, 解决了编码器在特征提取过程中由于缺少边缘、纹理细节信息以及内容语义信息而导致生成图像中内容模糊的问题.

1.2 D-DBPN

上述网络通过特征位置信息的编码和重新校准机制, 较好地解决了唐卡局部风格特征提取, 以及内容图像与唐卡风格序列之间特征对齐不足的问题, 能够更好地保留风格图像的纹理、细节信息, 得到的风格迁移图像较为清晰; 但却出现图像分辨率过低、色彩溢出、部分内容结构被破坏的现象. 为了进一步提高生成图像的质量, 使得模型最终生成图像线条流畅、纹理对齐, 避免出现内容结构被破坏, 导致色彩溢出、边缘羽化、唐卡特征不明显的问题. 受到误差修正机制的启发, 本文在模型中添加 D-DBPN, 运用多级上下采样层增强生成图像超分辨率特征, 并根据不同层次的误差反馈修正采样结果, 提高内容结构完整性. 最后通过自校正特征得到最终的唐卡风格图像. D-DBPN 分为初始特征提取、密集投影和重建 3 个阶段, 其结构如图 1 中 D-DBPN 所示.

(1) 初始特征提取. 首先使用 $\mathrm{Conv}(3,n_0)$ 从输入的生成图像中构建初始低分辨率特征图 L^0, 然后使用 $\mathrm{Conv}(1,n_R)$ 将维数从 n_0 减小到 n_R; 其中, n_0 表示初始 LR 特征提取中使用的滤波器数量, R 表示每个投影单元中使用的滤波器数量.

(2) 密集投影. 在初始特征提取之后是一系列密集投影单元, 如图 3 所示, $[L^1,L^2,\cdots,L^{t-1}]$ 和 $[H^1,H^2,\cdots,H^t]$ 分别表示上下投影单元, 前面所有特征映射被连接作为输入, 其自身的特征映射被用作所有后续单元的输入.

(3) 重建. 将目标高分辨率(high resolution, HR)图像重构为 $\boldsymbol{I}^{\mathrm{sr}}=\boldsymbol{f}_{\mathrm{Rec}}\left([H^1,H^2,\cdots,H^t]\right)$, 其中, $\boldsymbol{f}_{\mathrm{Rec}}$ 使用 $\mathrm{Conv}(3,3)$ 进行重构, $[H^1,H^2,\cdots,H^t]$ 是每个投影单元生成的特征图的拼接.

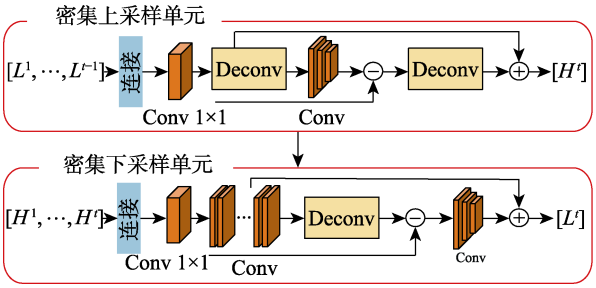


图 3 密集投影单元

通过将风格迁移模型中每个深度误差进行反馈, 用以修改采样结果, 然后累积每次采样阶段的自校正特征创建超分辨率图像. 在该网络中对生成图像进行不断地迭代, 通过调整图像中伪影或者分辨率过低的部分提高生成图像质量, 确保内容结构完整性.

1.3 损失函数

风格迁移模型生成的图像应该最大化地保持内容图像的结构和唐卡图像的风格. 因此, 本文采用 2 个不同的感知损失函数衡量输出图像 I_{o} 和输入内容图像 I_{c} 之间的内容差异, 以及风格差异[17]. 内容感知损失 L_{c} 定义为

$$L_{\mathrm{c}}=\frac{1}{N_l}\sum_{i=0}^{N_l}\left\|\phi_i\left(I_{\mathrm{o}}\right)-\phi_i\left(I_{\mathrm{c}}\right)\right\|_2 \tag{3}$$

其中, ϕ_i 表示从预训练 VGG 模型的第 i 层提取的特征; N_l 表示层数. 风格感知损失 L_{s} 定义为

$$L_{\mathrm{s}}=\frac{1}{N_l}\sum_{i=0}^{N_l}\left\|\mu\left(\phi_i\left(I_{\mathrm{o}}\right)\right)-\mu\left(\phi_i\left(I_{\mathrm{s}}\right)\right)\right\|_2+\left\|\sigma\left(\phi_i\left(I_{\mathrm{o}}\right)\right)-\sigma\left(\phi_i\left(I_{\mathrm{s}}\right)\right)\right\|_2 \tag{4}$$

其中, $\mu(\cdot)$ 和 $\sigma(\cdot)$ 分别表示提取特征的均值和方差. 通过最小化函数, 对整个网络进行优化, 将式(3)(4)代入, 得到 $L=\lambda_{\mathrm{c}}L_{\mathrm{c}}+\lambda_{\mathrm{s}}L_{\mathrm{s}}$.

2 实验结果与分析

2.1 实验环境参数与数据集

本文实验中, 硬件环境为 12 核 CPU, 40 GB 内存; 使用 NVIDIA GeForce RTX 3060 GPU 进行加速处理; 软件环境为 Ubuntu 18 操作系统, Pytorch 1.10.0 深度学习框架, Python 3.6 编程语言, CUDA 10.1. 输入图像尺寸为 256×256 像素, 设置 batch_size 为 8, 模型迭代次数设置为 2 000, 采用 Base-Optimizer 优化器进行优化, 学习率为 5×10^{-3}.

采用本文设计的唐卡图像风格迁移数据集 TangKaTransSet 进行实验, 其中包括 5 000 幅唐卡风格图像, 5 000 幅印象派内容图像, 如图 4 所示. 为了验证模型对任意风格图像的迁移效果, 本文设计了任意图像风格迁移数据集 TransSet, 其中包括漫画、油画和水彩画风格图像, 以及印象派、后印象派和素描图像内容图像, 如图 5 所示.

2.2 对比实验

2.2.1 主观评价

将 FPC-EI 与 AdaIN[10], WCT[8], ArtFlow[21],

图 4　TangKaTransSet 数据集部分示例

图 5　TransSet 数据集部分示例

CAST[22], AdaAttN[23]和 StyTr^2[17]进行实验，采用主观评价方法评估生成图像质量.

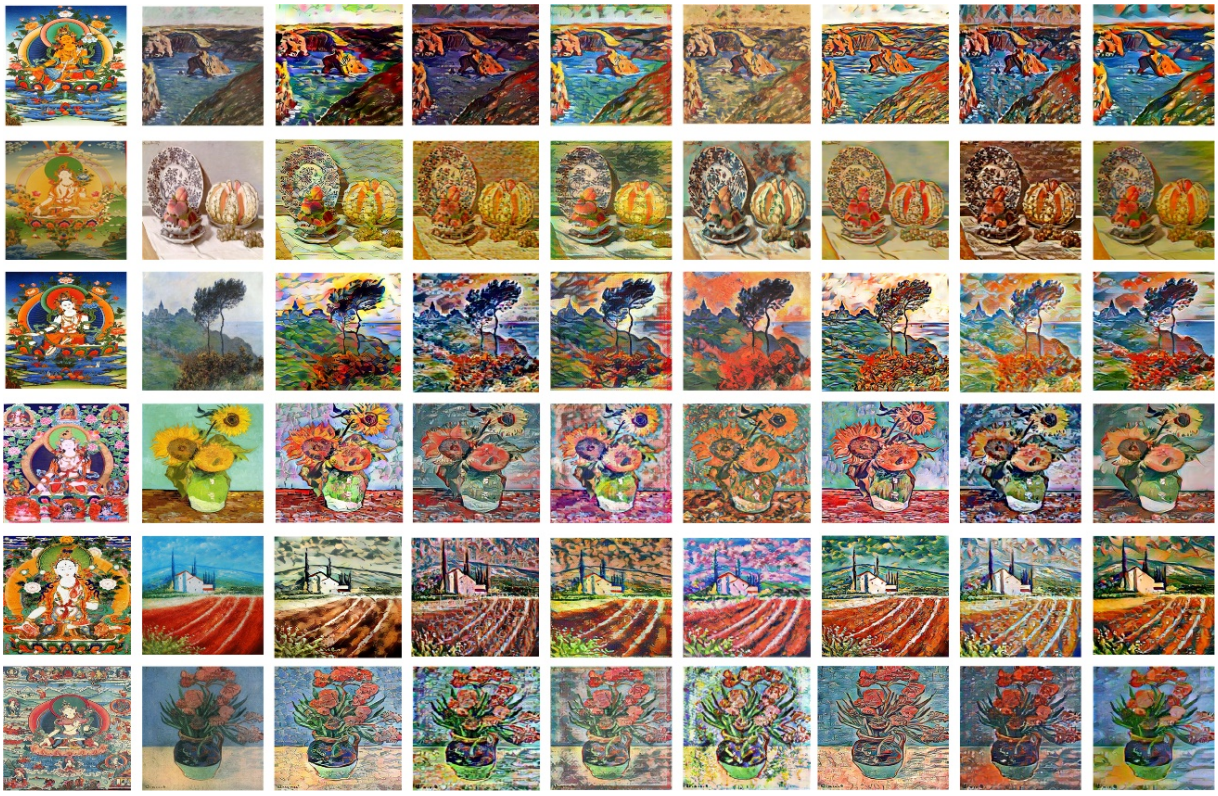
(1) 在 TangKaTransSet 数据集上验证唐卡风格迁移效果，结果如图 6 所示. 可以看出，与其他模型相比, FPC-EI 生成图像色彩饱满，突显了唐卡色彩丰富、艳丽的特征；很好地保留了内容图像的纹理细节信息，如第 3 行中山的边缘、云彩等特征，提高了内容结构的完整性. 因此, FPC-EI 对唐卡风格迁移得到了较高质量的生成图像.

(2) 在 TransSet 数据集上验证任意图像风格迁移效果，如结果图 7 所示. 可以看出，与其他模型相比, FPC-EI 可以很好地保留内容结构和纹理细节，生成图像更加完整且清晰，如第 4 行中，FPC-EI 实现了内容和风格图像之间的特征对齐和匹配，避免出现色彩溢出导致画面整体抽象的问题；但也存在一些不足，如第 2 行中出现了边缘模糊、纹理不清晰的现象.

主观评价实验中，使用 25 幅内容图像和 30 幅风格图像，共生成 750 幅效果图，并邀请 50 名参与者进行评价. 针对每个参与者，随机选择 50 幅生成图像，并将对比模型生成的图像随机并排展示，要求参与者从风格一致性、内容保存和整体质量 3 个方面进行评价；最终，从 50 名参与者中收集 2 500 张选票，图 8 所示为每种模型的投票百分比. 可以看出，与其他对比模型相比, FPC-EI 生成图像的质量最好，能够更好地保留图像的内容和风格特征.

2.2.2　客观评价

采用结构相似性(structural similarity, SSIM)[24]评估生成图像与参考图像之间相似度，比较亮度、对比度和结构 3 个指标, SSIM 值越大，则图像越相



a. 风格图像　b. 内容图像　c. AdaIN[10]　d. WCT[8]　e. ArtFlow[21]　f. CAST[22]　g. AdaAttN[23]　h. StyTr^2[17]　i. FPC-EI

图 6　TangKaTransSet 数据集上 7 种模型实验结果对比

a. 风格图像 b. 内容图像 c. AdaIN[10] d. WCT[8] e. ArtFlow[21] f. CAST[22] g. AdaAttN[23] h. StyTr2[17] i. FPC-EI

图 7 TransSet 数据集上 7 种模型实验结果对比

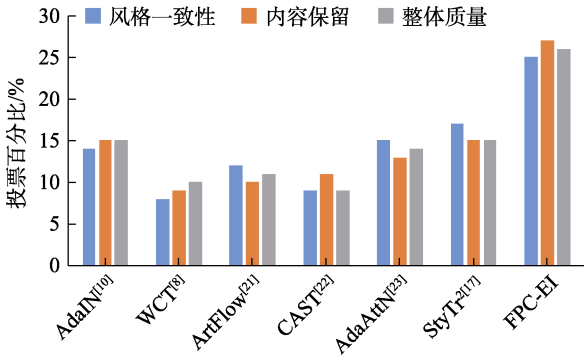


图 8 7 种模型主观评价结果对比

似; 采用学习感知图像块相似性(learned perceptual image patch similarity, LPIPS)[25]评估生成图像的多样性, LPIPS 值越小, 则生成图像具有真实性和多样性; 采用均方误差(mean square error, MSE)[26]评估生成图像与参考图像之间像素级别上的差异程度, MSE 值越小, 则生成图像失真程度越小. 将 FPC-EI 与 AdaIN[10], WCT[8], ArtFlow[21], CAST[22], AdaAttN[23]和 StyTr2[17]进行实验, 3 个评价指标结果如表 1 所示, 为了确保客观评价与主观评价一致, 对每种模型使用相同测试数据, 且不进行数据预处理. 可以看出, 3 组实验中, FPC-EI 的 SSIM 取得最高值, 说明该模型生成图像风格与唐卡风格更相似; LPIPS 取得最小值, 说明 FPC-EI 在迁移过程中更好地保留了内容图像结构, 且生成图像具有唐卡艺术风格; MSE 取得最小值, 说明 FPC-EI 生成的图像不易失真且质量更高.

表 1 7 种模型下 3 个评价指标结果对比

模型	SSIM	LPIPS	MSE
AdaIN[10]	0.46	0.52	0.070
WCT[8]	0.38	0.58	0.081
ArtFlow[21]	0.24	0.45	0.096
CAST[22]	0.18	0.48	0.084
AdaAttN[23]	0.28	0.54	0.085
StyTr2[17]	0.58	0.51	0.065
FPC-EI	**0.68**	**0.44**	**0.053**

注. 粗体表示最优值.

2.3 消融实验

2.3.1 IRPE

通过消融实验, 直观地验证 IRPE 在特征提取时特征信息保留方面产生的积极影响. 实验分为 StyTr2 模型的风格化和添加 IRPE 的风格化 2 组, 每组各有 200 幅风格迁移效果图. 这 2 组除了模型编码器不同之外, 其他参数均相同. 可以看出, 图 9a 中 2 幅图像的红色框中, 模型未能调节天空的局部特征, 导致最终生成图像全局风格与局部风格存在明显的不一致, 如天空中出现伪影, 天空与

山的边缘不清晰；图 9a 中 2 幅图像的蓝色框中，山的纹理信息未能与内容结构相匹配，出现模糊的现象；图 9b 中 2 幅图像中，IRPE 获得了序列之间的对应位置信息，可以实现特征之间的对齐和匹配，较好地保证了风格特征与语义内容的一致性. 实验结果表明，IRPE 可以帮助编码器获取特征之间的对应关系，对于唐卡图像风格迁移研究具有积极影响.

a. StyTr^{2}[17]　　b. IRPE

图 9　IRPE 消融实验对比

2.3.2　CoordAtt 机制

通过消融实验，直观地验证 CoordAtt 机制在特征传递过程中能够帮助模型获取低层次特征信息方面产生的积极影响. 实验分为 StyTr^{2} 模型的风格化和添加 CoordAtt 的风格化 2 组，每组各有 200 幅风格迁移效果图. 这 2 组除了特征传递过程不同外，其他参数设置相同. 可以看出，图 10a 中 2 幅图像的蓝色框中，岩石的横截面出现了一些不属于个体本身的风格元素；图 10a 中 2 幅图像的红色框中，天空和云彩出现了一些伪影，丢失纹理细节信息，从而使整幅图像变得抽象；图 10b 中 2 幅图像中，在 IRPE 的基础上，CoordAtt 机制可以帮助编码器捕获低层次特征的位置信息，更好地识别目标对象，提升特征传递的实时性.

a. StyTr^{2}[17]　　b. CoordAtt

图 10　CoordAtt 机制消融实验对比

2.3.3　D-DBPN

通过消融实验，直观地验证 D-DBPN 在图像生成质量方面产生的积极影响. 实验分为 StyTr^{2} 模型的风格化和添加 D-DBPN 的风格化 2 组，每组有 200 幅风格迁移效果图. 这 2 组除了误差迭代操作不同外，其他参数都相同. 可以看出，图 11a 中 2 幅图像的红色框中，在对山峰进行风格迁移时丢失了一些低层次关键信息，导致出现局部特征模糊的现象；图 11a 中 2 幅图像的蓝色框中，天空部分出现一些伪影；图 11b 中 2 幅图像中，D-DBPN 可以通过将生成图像的误差不断迭代反馈并指导模型实现更好的效果，提高生成图像质量. 实验结果表明，D-DBPN 对于唐卡风格迁移具有积极作用.

a. StyTr^{2}[17]　　b. D-DBPN

图 11　D-DBPN 消融实验对比

3　结　语

本文提出一种融合特征位置编码和误差修正的唐卡图像风格迁移模型 FPC-EI，用于生成以唐卡绘画风格为主的新图像. 首先设计突出特征位置信息的编码器，利用 IRPE 获取特征序列之间的对应位置信息，实现特征对齐和匹配，并融入 CooedAtt 机制，使得编码器在特征提取时选择性地强调低层次特征信息并进行提取；其次通过多层 Transformer 解码器实现特征融合；为了有效地调节风格特征，在模型添加了 D-DBPN，通过不断迭代实现误差修正，确保最终生成图像的质量. 实验结果表明，与现有方法相比，FPC-EI 实现了唐卡艺术风格与内容语义结构之间的匹配，消除了迁移过程中的误差，有效地增强了生成图像纹理信息以及唐卡艺术风格，获得质量更高的风格迁移图像；并通过主观和客观评价，验证了该模型的有效性. 但是，该模型仍然存在一些待解决的问题：虽然在本文数据集上都有不错的表现，但其生成的图像仍会存在一些伪影. 未来，可以结合边缘细化、强调特征等方法，进一步提高模型特征融合的能力.

参考文献(References):

[1] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style[J]. Journal of Vision, 2016, 16(12): Article No.326

[2] Gatys L A, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks[C] //Proceedings of the 28th International Conference on Neural Information Processing Systems. New York: ACM Press, 2015: 262-270

[3] 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. Los Alamitos: IEEE Computer Society Press, 2016: 2414-2423

[4] Gatys L A, Ecker A S, Bethge M, *et al*. Controlling perceptual factors in neural style transfer[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 3730-3738

[5] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets[C] //Proceedings of the 27th International Conference on Neural Information Processing Systems. New York: ACM Press, 2014: 2672-2680

[6] Li C, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks[C] //Proceedings of the 14th European Conference on Computer Vision. Heidelberg: Springer, 2016: 702-716

[7] Zhu J Y, Park T, Isola P, *et al*. Unpaired image-to-image translation using cycle-consistent adversarial networks[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 2242-2251

[8] Li Y J, Fang C, Yang J M, *et al*. Universal style transfer via feature transforms[C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 385-395

[9] Wang Z Z, Zhao L, Chen H B, *et al*. Diversified arbitrary style transfer via deep feature perturbation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 7786-7795

[10] Dumoulin V, Shlens J, Kudlur M. A learned representation for artistic style[OL]. [2023-05-04]. https://arxiv.org/abs/1610. 07629

[11] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 1510-1519

[12] 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. Los Alamitos: IEEE Computer Society Press, 2019: 5873-5881

[13] Yao Y, Ren J Q, Xie X S, *et al*. Attention-aware multi-stroke style transfer [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 1467-1475

[14] Deng Y Y, Tang F, Dong W M, *et al*. Arbitrary style transfer via multi-adaptation network[C] //Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM Press, 2020: 2719-2727

[15] Carrera J, Agrawal P, Fragkiadaki K, *et al*. Human pose estimation with iterative error feedback[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 4733-4742

[16] Song C J, Wu Z J, Zhou Y, *et al*. ETNet: error transition network for arbitrary style transfer[OL]. [2023-05-04]. https://arxiv.org/abs/1910.12056

[17] Deng Y Y, Tang F, Dong W M, *et al*. StyTr^2: image style transfer with transformers[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 11316-11326

[18] Wu K, Peng H W, Chen M H, *et al*. Rethinking and improving relative position encoding for vision transformer[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 10013-10021

[19] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 13708-13717

[20] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 1664-1673

[21] An J, Huang S Y, Song Y B, *et al*. ArtFlow: unbiased image style transfer via reversible neural flows[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 862-871

[22] Zhang Y X, Tang F, Dong W M, *et al*. Domain enhanced arbitrary image style transfer via contrastive learning[C] //Proceedings of the ACM SIGGRAPH Conference Proceedings. New York: ACM Press, 2022: Article No.12

[23] Liu S H, Lin T W, He D L, *et al*. AdaAttN: revisit attention mechanism in arbitrary neural style transfer[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 6629-6638

[24] Yao J C, Liu G Z. Improved SSIM IQA of contrast distortion based on the contrast sensitivity characteristics of HVS[J]. IET Image Processing, 2018, 12(6): 872-879

[25] Zhang R, Isola P, Efros A A, *et al*. The unreasonable effectiveness of deep features as a perceptual metric[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 586-595

[26] Han Han, Zhuo Li, Zhang Jing, *et al*. Summary of no-reference image quality assessment based on deep learning[J]. Measurement & Control Technology, 2022, 41(4): 1-10(in Chinese)
(韩翰, 卓力, 张菁, 等. 基于深度学习的无参考图像质量评价综述[J]. 测控技术, 2022, 41(4): 1-10)