基于深度鉴伪的图像和谐化方法

张琦[1)], 邢冠宇[2)], 董哲镐[3)], 刘艳丽[1)*]

[1)] (四川大学计算机学院 成都 610065)
[2)] (四川大学网络空间安全学院 成都 610065)
[3)] (四川大学建筑与环境学院 成都 610207)
(yanliliu@scu.edu.cn)

摘 要: 在计算机视觉和增强现实领域中, 将前景物体融合到背景场景中并实现图像和谐是一项重要且具有挑战性的任务. 目前和谐化方法大多通过调整图像前景的外观使其与背景相适应来达到视觉上一致性, 和谐化效果的提升空间有限. 为了进一步提升图像和谐化性能, 所提方法通过引入深度鉴伪网络, 利用生成对抗机制将鉴伪网络的结果作为判定指标与现有编码器-解码器和谐化网络构建 GAN 模型, 二者互相博弈以达到鉴伪网络无法识别出和谐重构后的结果为合成图像的目的, 并在鉴伪网络中添加了图像差分模块和图像光照模块, 该部分有效地增强其鉴定性能. 从定性和定量 2 方面在公共的基准数据集 iHarmony4 上进行多组实验, 结果表明, 所提方法均优于对比方法, 其中, 比性能最好的对比方法在 MSE 和 PSNR 评估指标上分别提高了 0.24 和 0.16 dB, 在图像和谐化任务上取得了优异的表现.

关键词: 图像合成; 图像和谐化; 图像鉴伪; 光照图; 差分图
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00345

Image Harmonization Method Based on Deep Forgery Detection

Zhang Qi[1)], Xing Guanyu[2)], Dong Zhehao[3)], and Liu Yanli[1)*]

[1)] (*College of Computer Science, Sichuan University, Chengdu* 610065)
[2)] (*School of Cyber Science and Engineering, Sichuan University, Chengdu* 610065)
[3)] (*College of Architecture and Environment, Sichuan University, Chengdu* 610207)

Abstract: In computer vision and augmented reality fields, it is an important and challenging task to fuse foreground objects into the background scene and achieve image harmonization. Most of the current mainstream harmonization methods adjust the appearance of the image foreground to make it compatible with the background visually, but the room for improving the harmonization effect is limited. In order to further improve the performance, this paper proposes an image harmonization authenticity identification network. Construct a GAN model with the results of the forgery detection network as the judgment indicator and the existing encoder-decoder harmonization network using the generative adversarial mechanism. The two networks compete with each other to achieve the result that the authenticity identification network cannot recognize the reconstruction result of the harmonization network as a synthetic image. The image difference module and image illumination module are added to the counterfeiting identification network, which effectively enhances its identification performance. Qualitative and quantitative experiments were conducted on the public benchmark dataset iHarmony4. The results

收稿日期: 2023-06-18; 修回日期: 2023-12-27. **基金项目**: 国家自然科学基金(62472297, 62172290); 四川省重点研发计划(2023YFS0454). **张琦**(1996—), 男, 博士研究生, 主要研究方向为图像处理、计算机视觉; **邢冠宇**(1985—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为计算机图形学、增强现实等; **董哲镐**(2002—), 男, 在校学生, 主要研究方向为图像处理、计算机视觉; **刘艳丽**(1981—), 女, 博士, 教授, 博士生导师, CCF 会员, 论文通信作者, 主要研究方向为计算机图形学、增强现实、图像/视频处理等.

demonstrate that the proposed method outperforms all comparative methods. In particular, compared to the best-performing comparative method, there were improvements of 0.24 and 0.16 dB in the MSE and PSNR, and this method has achieved excellent performance in image harmonization tasks.

Key words: image composition; image harmonization; image forgery detection; illumination map; difference map

在计算机视觉领域中，图像合成是图像编辑和数据增强中最常用的技术之一. 由于前景图像和背景图像往往在不同的条件下拍摄得到(如不同的时间、相机设置、天气、季节等)，使得图像在视觉上存在不协调、不和谐的缺陷，这种不和谐现象让合成图像缺乏真实感. 图像和谐化[1-5]旨在通过合成图像背景调整前景的外观，使得整幅图像前后区域外观兼容，达到接近真实图像的目的. 与重光照技术[6-7]相比，图像和谐化不需要推断复杂的光照、材料和几何信息，不假设光源，对现实复杂场景的适用性更佳[8]，已经受到了大量学者的关注和研究.

图像和谐化任务不仅需要注意前景和背景之间的光照和色彩的一致性，还需要在合成过程中处理阴影和纹理等方面的细节，以提高合成图像的质量和真实感，是一项重要且具有挑战性的工作. 当前，已有诸多解决该任务的方法. 其中，基于深度学习技术取得了颇为不错的成果，能够较为合理、准确地重构出目标和谐图像. 但大多数方法都集中在单一重构网络卷积层面上的改进，缺少对重构结果的约束和鉴定，使得模型的改进受阻[3-5]；此外，对图像是否和谐的判断也仅站在人眼视觉角度或者计算重构误差值，无法保证和谐化后的图像能达到人眼和机器视觉都不能区分的效果.

为了解决上述问题且更好地实现图像和谐化重构，本文提出基于深度鉴伪网络的图像和谐化方法. 在输入条件为不和谐图像和前景目标掩码的基础上，通过构建和谐化鉴伪网络模型，依据生成对抗博弈机制，将其与轻量级和谐化重构模块 iSSAM[4]搭建生成对抗网络(generative adversarial networks, GAN)[9]，使得模型能够通过对抗博弈学习不和谐图像的特征信息；在鉴伪网络中，依据 HDRNet 模型[7]生成光照图的方式设计图像光照模块，能快速且高效地生成目标图像的光照信息，并将它作为图像光照一致性的重要鉴定依据，确保和谐化后的图像前景和背景光照兼容；还利用图像通道间相关性构建图像差分方法，能够显著地反映出图像不和谐区域的通道特征差异；最后建立新的特征提取模块和融合模块，对光照、差分和原图像的特征进行融合鉴定，有效地增强了模型的检测判别能力，促使和谐化网络实现更加精确、合理重构效果. 在公开的基准数据集 iHarmony4[1] 上的多组定性和定量对比实验表明，本文模型和谐化后的图像在 MSE 和 PSNR 评估指标上取得了最优异的结果.

1　相关工作

图像和谐化流程如图 1 所示. 目前，图像和谐化分为传统方法和基于深度学习的方法 2 类. 传统的图像和谐化方法通过对前景进行颜色变换来匹配前景和背景之间的颜色统计量，达到和谐化的效果. 不同的方法之间的差异在于匹配的具体细节. Sunkavalli 等[10]通过操纵合成图像和目标图像金字塔的不同层次实现前景和背景之间的外观转移；Xue 等[11]进一步考虑合成图像的视觉真实感，对前景与背景之间最能匹配的区域提出通过训练分类器预测每个统计度量的直方图，并调整前景颜色以匹配前景与背景的区域；Lalonde 等[12]用颜

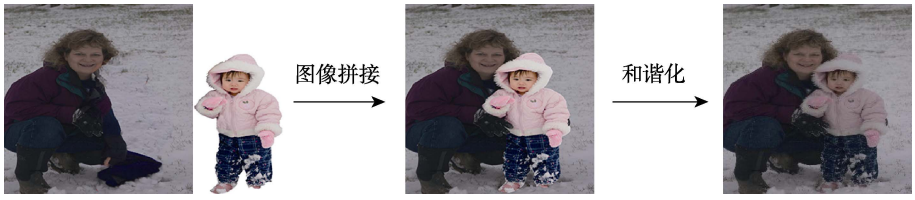


图 1　图像和谐化流程

色簇表示前景和背景的颜色分布，使用运输问题的解决方案匹配前景与背景的颜色簇. 传统的图像和谐化方法整体上通过统计人为手动设计特征调整前景图像的颜色，但是当外观差异较大时会产生不真实的结果；此外，这类方法主要依赖于人为手动设计特征，缺乏灵活性和自适应性.

近年来，基于深度学习的方法被广泛应用于图像和谐化研究中，这类方法利用深度卷积神经网络(convolutional neural network, CNN)强大的特征提取能力，可以从大量的图像数据中学习图像的各层特征信息和纹理，更好地保留图像细节. 目前，基于深度学习的图像和谐化方法结合了注意力机制、颜色变换、风格迁移等技术. Tsai 等[3]提出第 1 个用于图像和谐化的端到端 CNN，并利用辅助语义分割分支增强基础的图像和谐化网络；Cong 等[1]构建一个具有空间分离的注意力块的编码器-解码器模型，并引入一种域验证判别器对前景域和背景域拟合；Ling 等[13]将图像和谐化重新定义为背景到前景的风格迁移问题，引入区域感知的自适应实例归一化将视觉风格从背景迁移到前景；Sofiiuk 等[4]结合预训练的语义分割模型和编码器-解码器架构用于图像和谐化，将高级语义特征插入编码器以提供辅助信息；受传统图像和谐化方法应用颜色变换调整前景外观的启发，Cong 等[2]提出一种通过学习基于 LUT 颜色变换[14]的图像和谐化的方法，将颜色到颜色的变换和像素到像素的变换结合在统一的框架中：与之前的方法相比，尽管这些基于深度学习的方法已经取得了显著的成功，尤其是在模型处理效果和实用性等方面，但仍存在一些不容忽视的缺陷. 例如，多数方法单一地构建和谐化网络，基于卷积网络层面的改进方式有限，在一定程度上限制了提升空间；采用 GAN 的机制，引入域验证判别器强化前域和背景域的区分能力，但判别器的输入和结构比较单一，存在性能提升不明显等问题. 目前，在解决图像不兼容问题的研究中，增加图像鉴伪模块以构造生成对抗模型的工作较为匮乏，进行相关研究有望进一步提高图像和谐化的效果.

2 本文方法

对于一幅不和谐的图像 $\boldsymbol{I}$，利用其前景掩码 $\boldsymbol{M}$ 和对应的背景掩码 $\hat{\boldsymbol{M}}=\boldsymbol{1}-\boldsymbol{M}$，能够有效地区分目标图像的和谐区域(背景区域) $\boldsymbol{I}\cdot\hat{\boldsymbol{M}}$ 和不和谐区域(前景区域) $\boldsymbol{I}\cdot\boldsymbol{M}$，图像和谐化的目的就是将目标图像前景的不和谐区域更加真实地融合到背景区域中. 因此，基于深度学习的和谐化任务本质就是训练神经网络模型，使其能重构图像的不和谐前景区域，让整幅图像尽可能地接近真实的自然图像.

本文方法基于 GAN，其结构如图 2 所示. 网络生成器 G 采用编码器-解码器的构造方式，将目标图像 $\boldsymbol{I}$ 与前景掩码 $\boldsymbol{M}$ 作为输入，经过生成器网络结构后，图像 $\hat{\boldsymbol{I}}$ 为和谐化后的结果. 判别器 D 采用图像鉴伪网络的原理设计，将图像通道相关性和光源一致性作为约束，分别解析出生成器和谐化后图像的光照信息和差分信息，使得在鉴定过程充分考虑图像光照和通道关系的一致性，之后由特征提取和融合网络实现更加优异的和谐化鉴定.

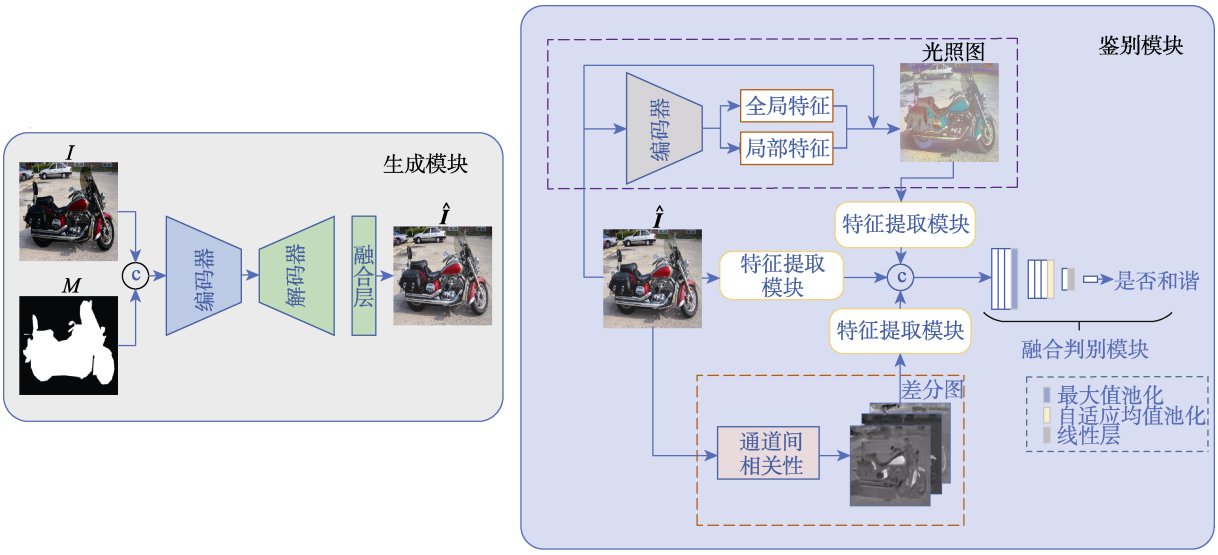


图 2 本文方法结构

将鉴定的结果与生成器结果进行生成对抗博弈，最终促进优化模型的和谐化性能.

2.1　基于编码-解码结构的图像生成器

在编码器-解码器结构的基础上，图像和谐化生成网络可以用多种方式实现. Sofiiuk 等[4]采用添加图像融合层的方式获取图像的混合掩码，并在此基础上评估出目标图像的不同分布区域的注意力程度，该方法能够充分利用输入图像信息改善模型和谐化的真实程度. 本文采用其改进后的 S^2AM(spatial-separated attention module)[5]，即 iS^2AM，将其作为和谐化图像的生成器. iS^2AM 是一个结构较为简易、整体占用空间不大的轻量级模型，且在之前的和谐化任务中取得了颇为不错的效果，因而非常适合作为通用的和谐化模型骨干[2,4]. 生成器的输入为不和谐图像 $\boldsymbol{I}$ 和前景掩码 $\boldsymbol{M}$，将两者进行维度结合，经过网络的编码器、解码器和图像融合层后得到和谐化后的图像. 对生成器和谐化后的结果 $\boldsymbol{G}(\boldsymbol{I})$ 和真实自然图像 $\boldsymbol{I}_{\text{real}}$ 之间建立构造损失函数 $L_{\text{pix}}=\frac{1}{N}\sum_{i}^{N}\left(\boldsymbol{M}\cdot\boldsymbol{G}(\boldsymbol{I}_i)-\boldsymbol{M}\cdot\boldsymbol{I}_{i_\text{real}}\right)^2$. 其中，$\boldsymbol{I}_i$ 表示输入的不和谐图像；$\cdot$ 表示矩阵点乘运算；$\boldsymbol{G}(\cdot)$ 表示生成器网络作用后的结果；N 表示训练样本总数. 训练过程中，最小化损失函数 L_{pix}，使得生成器和谐化后的图像达到逼近真实图像.

2.2　基于深度鉴伪网络的图像判别器

深度鉴伪网络利用神经网络优异的性能，克服了传统鉴别方法诸多的缺点和不足，成为鉴别技术的主流方法[15-16]. 与图像和谐化生成器结果的不确定性和差异性相对应，图像鉴伪网络能够有效地判定其结果的优异，合理地补缺和谐化网络重构效果的缺陷. 本文构造鉴伪网络作为和谐化网络的判别器，引入鉴伪模型的思路与直接构造 GAN 解决和谐化问题的出发点不同，之前的方法中没有直接或间接采用深度鉴伪模型对和谐结果进行鉴定，相关 GAN 方法设计判别器模型的目的是与生成器搭建对抗机制. 此外，这些方法中判别器的实现方式也较为简易，实现细节与深度鉴伪模型大不相同，远没有后者的网络结构复杂.

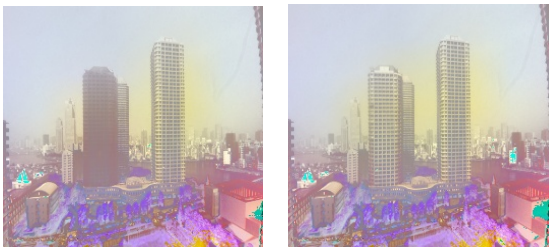
本文鉴伪网络采用 3 个特征提取模块和一个特征融合模块综合鉴定不和谐图像的真伪性，该网络的整体结构如图 2 右侧部分所示. 其中，一个特征提取模块作用于和谐化后的图像；其余 2 个特征提取模块分别提取由图像解析出的光照信息和差分信息特征. 全面结合特征作为鉴别信息，能够有效地保障和谐化后图像前景区域和背景区域的光照一致性和通道相关性，增强模型的判别鉴伪性能. 下面分别介绍鉴伪网络的各个模块.

2.2.1　图像光照模块

在对目标图像进行和谐化操作后，图像的前景区域依据背景区域的光照特征属性进行合适的调整以达到二者兼容；但由于生成器和谐化效果的局限性，无法确切地保证图像光照协调后精确一致，因此对于检验图像和谐化质量的判别器，和谐化后的图像前景区域和背景区域是否能够维持光照一致性，是一个重要的评判因素.

借鉴 HDRNet 模型生成光照图的方式，本文提出图像光照模块，其中对于目标图像，需要兼顾考虑局部特征(如对比度、细节锐度、阴影和高光等)，以及全局特征(如颜色分布、平均亮度和场景类别等). 为了合理、准确地提取图像光照特征，本文从编码器网络生成特征中解析出局部特征和全局特征；进一步，对 2 个特征进行融合和线性预测，最终得到目标图像光照信息. 图像光照模块的整体结构如图 2 右侧部分的上面线框所示，其中，由于生成光照图的对象已固定为生成器和谐化后的结果，因此不再需要下采样和上采样操作，直接由提取特征预测光照图，该模块能够高效、准确地获取图像的光照信息.

光照图上的每个像素值可理解为原图像对应位置的光源估计亮度值. 通常，一个区域内的光源遵守一致性原则，图像中每个邻域内的光源值应该相同或较为接近. 然而，对于不和谐图像，由于前景区域和背景区域的获取条件不同，其前后区域两者的光源本身就存在较大差异，光照图中的像素点值也截然不同，因此，不和谐图像无法保持光照一致性. 图 3 所示为不和谐图像和真实图像的光照图在前景区域上的差异. 图 3a 中，左侧的高楼区域(不和谐区域)与周围物体的光源存在显著的差异，并且边缘的割裂感明显，对比图 3b 真实图像的附近区域，不和谐区域的光照一致性被破坏. 根据以上分析，本文从光照角度出发，在图像



a. 不和谐图像　　b. 真实图像

图 3　图像光照图对比

鉴别网络中增加光照约束, 利用图像光照模块提取光照信息, 并将其作为和谐化鉴定的一个重要判别依据.

2.2.2 图像差分模块

在图像处理中, 绝大多数彩色图像由 RGB 三色通道组成, 每幅图像的通道内部会存在一定相关性[17]. 但对于不和谐图像, 其前景区域的环境条件发生改变或者经历了颜色变换等操作的处理, 图像整体的通道相关性被破坏; 此外, 单一的生成网络没有充分地利用彩色图像通道间相关性, 导致重构后的图像可能会存在色偏的现象或者出现伪影的现象[15].

基于上述通道相关性的性质, 本文提出针对和谐化任务的图像差分模块, 并将其作为另一个重要的和谐化判别依据. 图像通道间差分值为

$$D = I_{c_1} - I_{c_2} \tag{1}$$

其中, I_c 表示每个通道的像素值, $c_2, c_2 \in \{R, G, B\}$. 由式(1)可知, 本文采用图像差分的方式直接量化该部分差异, 能够准确地反映出通道间固有相关性的变化. 图 4 所示为真实图像和不和谐图像在 R-G 通道差分图上的结果.



a. 不和谐图像 b. 真实图像

图 4 图像差分图对比

图 4a 中, 不和谐区域为草原上右侧的羊, 与图 4b 相比, 不和谐区域内的目标图像丢失了很多图像细节, 如腿和头部的纹理信息和角边缘的信息等. 由此可以看出, 差分图能够合理、精确地反映出图像通道间的关联性, 甚至能够进一步放大已被破坏通道相关性图像的差异. 考虑到和谐化网络中缺乏差分约束, 增加差分模块作为限制因素会有效地保障重构过程中的图像相关性, 增强模型的和谐化性能. 为了保持输入维度的统一, 本文分别计算 R-G, B-G 和 R-B 这 3 个图像通道差分结果, 并将它们在通道维度拼接, 构成最终的图像差分信息. 图像差分模块的结构如图 2 右侧部分下面虚线框所示.

2.2.3 特征提取与融合模块

本文提出基于图像卷积网络的特征提取模块和特征融合模块. 特征提取模块采用 VGG-16 模型[18]的前 10 层网络, 凭借 CNN 强大的特征提取能力构建 3 个特征提取模块, 分别对生成器和谐化后的图像、图像光照图和图像差分图提取高维特征; 然后将它们在维度方向拼接输入到融合模块. 特征融合模块的作用是对输入的拼接特征进行融合和鉴定, 其结构是首先经过 3 个 3×3 卷积层, 加一个最大值池化层采样; 然后经过 3 个 3×3 卷积层后, 加一个自适应均值池化层采样; 最后经过 3 个全连接加 Sigmoid 函数. 与原始的 VGG 相比, 本文提出的融合模块将池化层替换为自适应均值池化采样, 不仅减少全连接层的数量, 且有效地减少了模型的参数量. 最后的输出结果作用 Sigmoid 函数, 能够获得最终判别器网络鉴定的图像和谐化程度.

根据鉴伪网络估计的图像和谐化程度结果, 本文设计判别器的对抗损失函数 L_{d1} 和 L_{d2} 分别为

$$L_{\mathrm{d1}}\left(D(\hat{\boldsymbol{I}}), \mathrm{False}\right) = -\ln\left(1 - D(\hat{\boldsymbol{I}})\right),$$

$$L_{\mathrm{d2}}\left(D(\boldsymbol{I}_{\mathrm{real}}), \mathrm{True}\right) = -\ln\left(D(\boldsymbol{I}_{\mathrm{real}})\right).$$

其中, $\hat{\boldsymbol{I}} = G(\boldsymbol{I})$ 表示生成器和谐化后的图像; $D(\cdot)$ 表示判别器作用后的结果; False 和 True 的数值分别为 0 和 1, 分别表示模型重构后的结果不和谐或和谐. 当输入判别器的图像为生成器和谐化后的结果时, 让判别器学习并使其鉴定为不和谐图像; 当输入的图像为真实图像时, 再让判别器学习并将其鉴定为和谐图像. 因此, 本文模型判别器总的损失函数用 L_{D} 表示为 $L_{\mathrm{D}} = \lambda_1 L_{\mathrm{d1}} + \lambda_2 L_{\mathrm{d2}}$. 其中, λ_1 和 λ_2 表示 2 项的权重系数.

与之对应, 根据 GAN 的生成对抗机制, 能够凭借判别器的鉴定结果督促优化生成器, 进一步提高生成器的和谐化性能. 本文设计生成器的对抗损失函数为 $L_{\mathrm{g}}\left(D(G(\boldsymbol{I})), \mathrm{ture}\right) = -\ln\left(D(G(\boldsymbol{I}))\right)$. 其中, $\boldsymbol{I}$ 表示输入的不和谐图像. 通过建立该对抗损失, 强化生成器, 使其和谐化后的结果与真实图像足够逼真, 达到欺骗判别器的目的. 因此, L_{G} 和 L_{D} 的值是相互博弈的关系, 两者此消彼长, 共同促进模型参数的优化.

综合第 2.1 节的生成器构造损失, 本文模型生成器的总损失函数表示为 $L_{\mathrm{G}} = \eta_1 L_{\mathrm{pix}} + \eta_2 L_{\mathrm{g}}$. 其中, η_1 和 η_2 分别表示 2 项权重系数.

3 实验与结果分析

3.1 实验设置

本文实验在 Pytorch1.10.0 环境, 采用 PyCharm

软件, Ubuntu20.04 LTS 操作系统, 在 128 GB 内存, Intel(R) Xeon(R) Bronze 3204 CPU @ 1.90 GHz CPU, NVIDIA RTX 3090 GPU 平台下进行.

在模型训练过程中, 生成器和判别器的优化器均采用 Adam 梯度优化器, 设置梯度学习率为 2×10^{-4}; 损失函数系数分别为 $\lambda_1=1$, $\lambda_2=1$, $\eta_1=10$ 和 $\eta_2=1$; 网络训练的批量大小为 4. 整个训练过程迭代了 182 次, 训练耗时大约为 7 d, 每次迭代的完成后的模型均用实验评估指标在测试集上进行验证.

3.2　实验数据集

本文使用公共基准数据集 iHarmony4[1]作为实验数据集. 该数据集由 HCOCO, HFlickr, HAdobe5k 和 Hday2night 这 4 个子数据集组成, 每个子数据集中的图像包含多种前景物体, 足以确保数据集的样本多样性.

HCOCO 基于 Microsoft COCO 数据集[19]构造, 其中的合成图像由真实图像中提取的背景和前景区域组成, 包含 42 828 对合成图像和真实图像. 通过寻找与前景区域相同类别的另一幅图像, 并将该图像的前景区域迁移颜色信息对目标前景区域外观进行修改, 使得图像前景和背景区域不兼容, 合成的图像即为不和谐图像.

HAdobe5k 数据集[20]包含各种各样的场景、物体和照明条件, 总共有 5 000 幅图像, 每幅都由 5 位摄影师进行润色, 生成 5 幅不同外观的合成图像. Cong 等[1]从初始的 5 000 幅图像中挑选 4 329 幅图像, 每幅图像都包含 1 个人工分割的前景对象掩码, 构建新的 HAdobe5k 子数据集, 该数据集共包含 21 597 对合成图像和真实图像.

Cong 等[1]基于使用爬虫技术获取的 4 833 幅 Flickr 图像构建了 HFlickr 子数据集, 每幅图像中都有 1~2 个分割出来的前景物体. 由于 Flickr 数据集没有分割注释, Cong 等[1]使用预训练的场景解析 ADE20K 模型[21]来获得前景物体的类别. HFlickr 子数据集包含 8 277 对合成图像和真实图像.

Hday2night 数据集[22]是从 AMOS 数据集[23]中收集得到的, 其中的图像由固定的网络摄像头在一天中不同时间拍摄得到, 总共有 101 个不同场景的 8 571 幅图像. Cong 等[1]从中选择了 80 个场景中的 106 幅目标图像, 并从每幅图像中分割出一个前景物体来生成合成图像, 合成图像的构造过程与 HAdobe5k 数据集相同, 背景区域和前景区域分别从同一场景的不同图像中获取. Hday2night 子数据集包含 444 对合成图像和真实图像.

3.3　评估指标

采用均方误差(mean-square error, MSE)和峰值信噪比(peak signal-to-noise ratio, PSNR)对实验结果进行评估, 它们分别定义为

$$\mathrm{MSE}=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\|I(i,j)-K(i,j)\right\|^2,$$

$$\mathrm{PSNR}=10\cdot\lg\left(\frac{M_I^2}{\mathrm{MSE}}\right)=20\cdot\lg\left(\frac{M_I}{\sqrt{\mathrm{MSE}}}\right).$$

其中, I 和 K 代表 2 幅对比图像; i 和 j 分别表示遍历图像的横纵坐标; m 和 n 为图像的长宽像素值; M_I 为该图像数据类型的最大动态范围. MSE 指标通过计算图像每个位置上的像素差异, 能有效评估 2 幅图像的相近程度. PSNR 指标则是根据 MSE 指标值进一步计算获取. 对于图像和谐化重构任务, MSE 值越小, 表示目标图像失真越小; PSNR 值越小, 表示目标图像失真越严重.

3.4　对比实验

将本文模型与图像和谐化研究中的多个模型进行对比实验, 包括经典模型 DoveNet[1]和 RainNet[13]、效果优异模型 iSSAM[4], 以及当前和谐化效果最好的模型 CDTNet[2]. 本文采用这些文献作者上传的预训练模型作为测试对象, 用统一基准的数据测试集进行结果评估.

3.4.1　定量分析

为了精确地评估模型的性能, 分别比较了各个模型在 4 个测试子集上和谐化后的 MSE 和 PSNR 均值, 共有 7 404 对测试图像. 采用量化方法首先对原真实图像分割出前景目标区域; 然后对其进行光照颜色等方面的变换, 进而拼接为一幅不和谐图像; 最终, 和谐化后的结果与原真实图像之间的差异即可反映出模型的和谐化性能. 该方式比人工标注是否和谐更加精准合理. 其中, DoveNet[1], iSSAM[4]和 CDTNet[2]的结果为原文献中计算出的指标值; RainNet[13]模型原文献中只记录了 PSNR 评估指标结果, 本文用该文献作者上传的预训练模型重新测试得到 MSE, 最终的统计结果如表 1 所示.

从表 1 可以看出, 输入图像的评估结果为直接检测数据集中的不和谐图像的评估值, 其 PSNR 和 MSE 都表现得非常差, 与输入的图像存在严重的前景和背景不兼容情况相对应; 与输入图像相比, DoveNet[1]和 RainNet[13]模型的结果有一定程度的改善, MSE 有较大幅度的下降, PSNR 也有一定的提升, 但仍然存在有待提升的空间; iSSAM[4]和 CDTNet[2]模型优化了图像和谐化的结果, 将评估

表 1 5 种模型在 4 个子数据集中的评估结果对比

模型	HCOCO		HAdobe5k		HFlickr		Hday2night		iHarmony4	
	MSE ↓	PSNR/dB ↑	MSE ↓	PSNR/dB ↑	MSE ↓	PSNR/dB ↑	MSE ↓	PSNR/dB ↑	MSE ↓	PSNR/dB ↑
输入图像	69.37	33.94	345.54	28.16	264.35	28.32	109.65	34.01	172.47	31.63
DoveNet[1]	36.72	35.83	52.32	34.34	133.14	30.21	54.05	35.18	52.36	34.75
RainNet[13]	31.12	37.08	56.36	36.22	117.58	31.64	47.24	34.83	44.50	36.12
iSSAM[4]	16.48	39.16	21.88	38.08	69.67	33.56	40.59	37.72	24.44	38.19
CDTNet[2]	16.25	39.15	**20.62**	38.24	68.61	33.55	36.72	37.95	23.75	38.23
本文	**16.08**	**39.27**	21.55	**38.47**	**65.68**	**33.69**	**32.31**	**38.00**	**23.51**	**38.39**

注. 粗体表示最优值; ↓表示该指标的值越小性能越优异, ↑表示该指标的值越大性能越优异.

指标提升到了新高度; 本文模型采用生成对抗博弈机制构造和谐化图像鉴伪模型, 并增加光照和差分约束, 有效地改进图像和谐化的效果, 在 HCOCO, HFlickr 和 Hday2night 这 3 个子数据集上均取得了最好的结果, 在 HAdobe5k 子数据集中, PSNR 值比其他模型更好, MSE 值取得了第 2 名的成绩; 在整个 iHarmony4 数据测试集的 PSNR 和 MSE 均值上, 本文模型取得最优异的成绩, 表现出最佳的性能.

3.4.2 定性分析

分别对各个模型在 4 个子数据集中的和谐化示例图像展示视觉方向和谐化性能, 结果如表 2 所示.

从表 2 可以看出: (1) 输入的不和谐图像在前景区域和背景区域存在十分明显的不兼容问题. 例如, 第 1 行图像在傍晚时间周围环境显得暗淡, 树上的鸟羽毛却十分明亮; 第 5 行图像中, 鸡的背景木框较为明亮, 鸡的阴影区域也十分明显, 表明前景物体处于比较明亮的环境, 但物体本身的表面光照却显得灰暗, 前后景物差异化严重; 第 8 行图像上, 田地中多数植物都为暗绿色, 却有一处田地显示为翠绿色, 与周围环境格格不入. (2) DoveNet[1]和 RainNet[13]模型和谐化后的结果有相对的改善, 但仍然存在较为严重的像素和光照差异; 与 DoveNet[1]和 RainNet[13]模型相比, iSSAM[4]和 CDTNet[2]模型的和谐化效果更加优异, 但与真实图像相比, 在部分区域还存在一些欠缺, 例如, 对比树上白色鸟的羽毛和亮度不和谐的路牌, 输入图像有一定程度的光亮和色差调整, 仍然未达到真实图像应有的效果; 与各个对比模型相比, 本文模型和谐化后的图像中, 前景不和谐区域和背景区域显得更加自然协调, 不存在两者不兼容的视觉效果, 几乎完全融为一体, 和谐化的结果也十分接近目标真实图像, 达到最优异的和谐化性能.

3.5 消融实验

通过多组消融实验, 证明本文模型的有效性. 其中, 采用 iHarmony4 数据集中的 4 个测试子集进行实验, 共有 7 404 对测试图像. 消融实验的 MSE 和 PSNR 均值结果如表 3 所示. 其中, 实验 1 为只采用图像和谐化网络生成器进行测试的评估结果; 实验 2 为在和谐化生成器的基础上只增加判别器骨干网络, 除去光照模块和差分模块再进行测试的评估结果; 实验 3 是完整模型的测试评估结果. 可以看出, 与实验 1 相比, 实验 3 的效果更好, 有效地验证了上文分析的使用生成对抗机制对和谐化结果进行鉴伪方式的正确性; 与实验 2 相比, 在和谐化鉴定过程中, 实验 3 对图像进行光照一致性和通道相关性的约束是十分有用且至关重要的, 再凭借 CNN 强大的特征提取和融合能力, 有效地增强了判别器模型的鉴定能力和生成器模型的和谐化能力. 消融实验结果表明, 本文模型在 4 个子数据集的评估指标上取得最优异的成绩; 同时证明了第 2 节提出的增强图像和谐化性能改进方法的有效性和合理性.

3.6 损失函数参数分析

本文模型的损失函数系数包括 λ_1, λ_2, η_1 和 η_2, 除 η_1 外, 其他参数都是生成器和鉴别器对抗损失的系数, 本文中这 3 个参数统一取值为 1. 参数 η_1 能够反映网络训练过程中构造损失和对抗损失的侧重比, 取值越大, 表示训练过程越依赖构造损失; 反之亦然. 因此, 本文进行 3 组实验, 探讨参数 η_1 的最佳取值. 实验采用 iHarmony4 训练集, 在其余条件全部不改变的情况下, 仅分别设置 η_1 为 1, 10 和 100, 再用 iHarmony4 测试集数据分别评估训练出的模型, 记录 3 组评估指标, 结果如表 4 所示. 可以看出, η_1 值不宜过大也不宜过小, 取值为 10 时能够训练出效果好的目标模型.

表 2　5 种模型和谐化结果对比

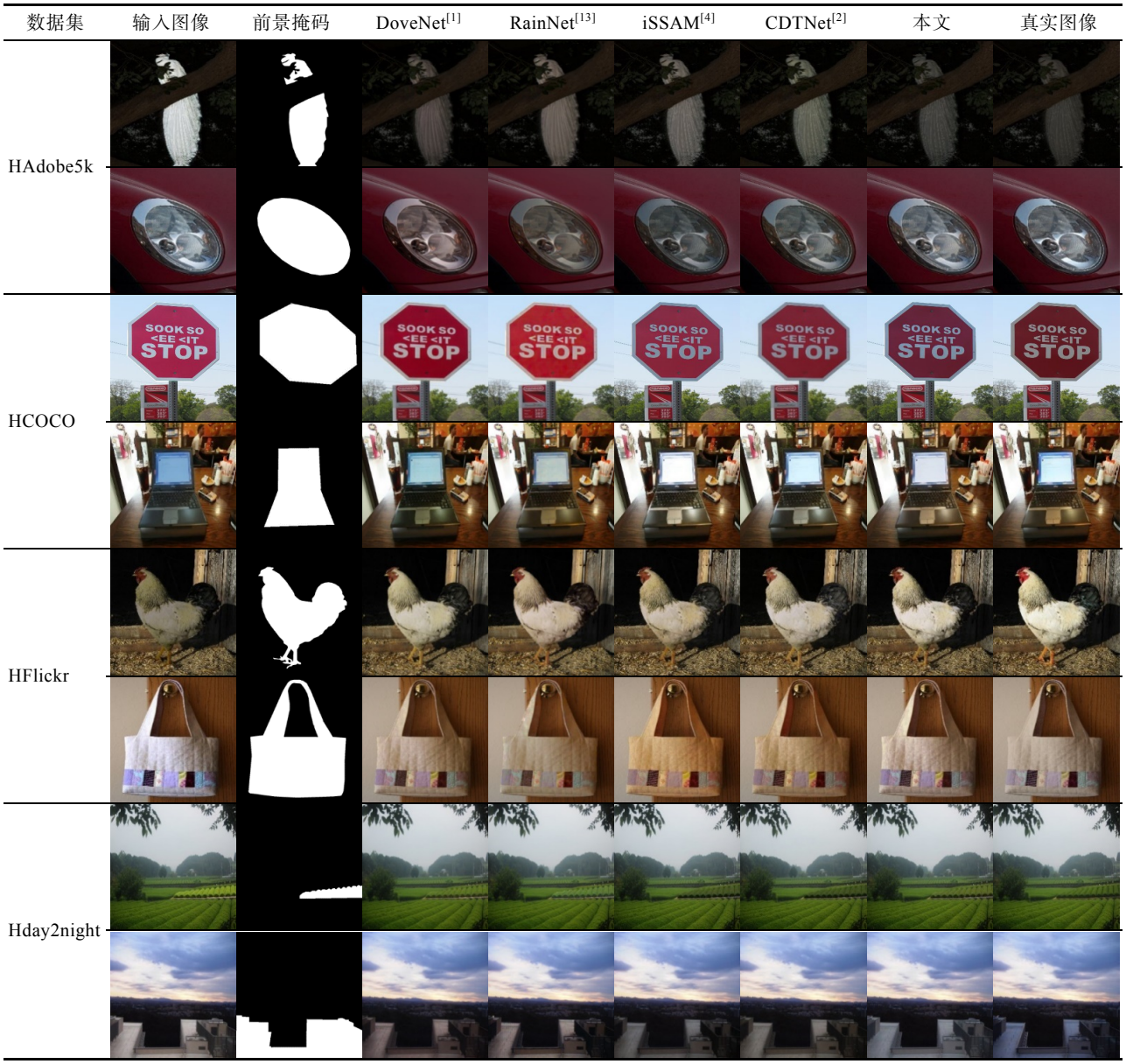


表 3　本文模型消融实验结果对比

实验	生成器	判别器		MSE ↓	PSNR/dB ↑
		骨干	完整		
1	√			24.67	38.09
2	√	√		23.64	38.29
3	√		√	23.51	38.39

注. ↓表示值越小性能越优异, ↑表示值越大性能越优异.

表 4　参数 η_1 对比实验

η_1	MSE ↓	PSNR/dB ↑
1	23.64	38.29
10	23.51	38.39
100	24.56	38.11

注. ↓表示值越小性能越优异, ↑表示值越大性能越优异.

3.7　时间性能分析

由以上分析可知，本文模型在和谐化后的结果精度上取得显著的改进和增强. 考虑增强现实应用的特点，对本文模型的时间性能进行实验. 测试数据集采用 HCOCO 子数据集的测试图像，共有 4 283 幅，其中，测试图像尺寸统一为 256×256 像素，将测试集数据分别经过和谐化网络重构生成，并统计该过程总的消耗时长，计算出处理一幅图像的时间为 0.036 s. 实验结果表明，本文模型在处理速率上具有非常高效的性能.

3.8　局限性分析

本文模型凭借引入鉴别网络和生成对抗机制有效地增强了图像和谐化的能力，取得了十分优

异的和谐化结果. 但是, 该模型也存在弊端和局限性. 如图 5 所示, 图像中的不和谐物体为一个在黑夜行驶的挂满了五颜六色灯光的电车, 由于其外表光照颜色等信息变化复杂, 且周围环境较阴暗, 可借鉴的环境信息较少, 增加了和谐化目标图像的难度. 可以看出, 本文鉴伪网络无法准确地判定车上的每一种颜色的正确性, 出现的检验误差会导致和谐化网络的能力受限, 对比和谐化结果和真实图像的结果, 本文模型只能相对改善输入图像的不和谐程度, 无法精确地还原出电车不同部位的颜色光照信息, 与真实图像存在着较为明显的差异.



a. 输入图像 b. 和谐化结果 c. 真实图像

图 5 本文模型和谐化局限性案例

因此, 对于周围环境信息缺失和目标光照颜色信息复杂的场景图像, 本文模型还存在进一步的提升和改进空间.

4 总 结

针对解决合成图像和谐化的任务, 本文提出一种图像和谐化鉴伪方法, 并与当前轻量级和谐化网络构造生成对抗机制, 达到增强模型和谐化性能的目的. 其中, 在判别器网络中增加了图像光照模块和图像差分模块, 有效地约束和谐化网络结果的光照一致性和通道相关性, 并通过设计的基于 CNN 的特征提取模块和特征融合模块, 有效地提高了模型的和谐化鉴定能力. 对基准数据集 iHarmony4 上的测试集进行和谐化结果评估, 实验结果表明, 本文模型的 MSE 和 PSNR 评估指标均取得了最佳的成绩, 证明了本文模型在图像和谐化任务中表现出优异的性能.

未来, 将本文模型构建为一个通用性的框架, 以适用于任何和谐化网络模型, 并与之建立生成对抗机制, 进而实现改善和强化其性能的目的; 对于和谐化任务, 可以将图像和谐化进一步延伸到视频和谐化、三维图像和谐化等任务中, 实现更多拓展性的研究.

参考文献(References):

[1] Cong W Y, Zhang J F, Niu L, *et al*. DoveNet: deep image harmonization via domain verification[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 8391-8400

[2] Cong W Y, Tao X H, Niu L, *et al*. High-resolution image harmonization via collaborative dual transformations[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 18449-18459

[3] Tsai Y H, Shen X H, Lin Z, *et al*. Deep image harmonization[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 2799-2807

[4] Sofiiuk K, Popenova P, Konushin A. Foreground-aware semantic representations for image harmonization[C] //Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 1619-1628

[5] Cun X, Pun C M. Improving the harmony of the composite image by spatial-separated attention module[J]. IEEE Transactions on Image Processing, 2020, 29: 4759-4771

[6] Debevec P, Hawkins T, Tchou C, *et al*. Acquiring the reflectance field of a human face[C] //Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2000: 145-156

[7] Ren P R, Dong Y, Lin S, *et al*. Image based relighting using neural networks[J]. ACM Transactions on Graphics, 2015, 34(4): Article No.111

[8] Niu L, Cong W, Liu L, *et al*. Making images real again: a comprehensive survey on deep image composition[OL]. [2023-06-18]. https://arxiv.org/abs/2106.14490

[9] Goodfellow I, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144

[10] Sunkavalli K, Johnson M K, Matusik W, *et al*. Multi-scale image harmonization[J]. ACM Transactions on Graphics, 2010, 29(4): Article No.125

[11] Xue S, Agarwala A, Dorsey J, *et al*. Understanding and improving the realism of image composites[J]. ACM Transactions on Graphics, 2012, 31(4): Article No.84

[12] Lalonde J F, Efros A A. Using color compatibility for assessing image realism[C] //Proceedings of the IEEE 11th International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2007: 1-8

[13] Ling J, Xue H, Song L, *et al*. Region-aware adaptive instance normalization for image harmonization[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 9357-9366

[14] Zeng H, Cai J R, Li L D, *et al*. Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 2058-2073

[15] Mo Caiwang. Color image super-resolution based on inter-channel correlation[D]. Nanning: Guangxi University, 2019(in Chinese)
(莫彩网. 基于通道间相关性的彩色图像超分辨率技术研究[D]. 南宁: 广西大学, 2019)

[16] Liang J, Niu L, Zhang L Q. Inharmonious region localization[C] //Proceedings of the IEEE International Conference on Multimedia and Expo. Los Alamitos: IEEE Computer Society Press, 2021: 1-6

[17] Gunturk B K, Altunbasak Y, Mersereau R M. Color plane interpolation using alternating projections[J]. IEEE Transactions on Image Processing, 2002, 11(9): 997-1013

[18] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[OL]. [2023-06-18]. https://arxiv.org/abs/1409.1556

[19] Lin T Y, Maire M, Belongie S, *et al*. Microsoft COCO: common objects in context[C] //Proceedings of the 13th European Conference on Computer Vision. Heidelberg: Springer, 2014: 740-755

[20] Bychkovsky V, Paris S, Chan E, *et al*. Learning photographic global tonal adjustment with a database of input/output image pairs[C] //Proceedings of the CVPR 2011. Los Alamitos: IEEE Computer Society Press, 2011: 97-104

[21] Zhou B B, Zhao H, Puig X, *et al*. Scene parsing through ADE20K dataset[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5122-5130

[22] Zhou H, Sattler T, Jacobs D W. Evaluating local features for day-night matching[C] //Proceedings of the Computer Vision. Heidelberg: Springer, 2016: 724-736

[23] Jacobs N, Roman N, Pless R. Consistent temporal variations in many outdoor scenes[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2007: 1-6