结合 Pix2Pix 生成对抗网络的灰度图像着色方法

李洪安¹⁾, 郑峭雪¹⁾, 张婧¹⁾, 杜卓明^{2)*}, 李占利¹⁾, 康宝生³⁾

¹⁾ (西安科技大学计算机科学与技术学院 西安 710054)

²⁾ (江苏理工学院计算机工程学院 常州 213001)

³⁾ (西北大学信息科学与技术学院 西安 710127)

(an6860@126.com)

摘 要: 针对神经网络在进行图像着色时容易出现物体边界不明确、图像着色质量不高的问题, 提出结合 Pix2Pix 生成对抗网络的灰度图像着色方法. 首先改进 U-Net 结构, 采用 8 个下采样层和 8 个上采样层对图像进行特征提取和颜色预测, 提高网络模型对图像深层次特征的提取能力; 然后使用 L_1 损失和 smooth L_1 损失度量生成图像与真实图像之间的差距, 对比不同损失函数下的图像着色质量; 最后加入梯度惩罚, 在生成图像和真实图像分布之间构造新的数据分布, 对每个输入数据进行梯度惩罚, 改变判别器网络梯度限制方法, 提高网络在训练过程中的稳定性. 在相同实验环境下, 使用 Pix2Pix 模型和 summer2winter 数据进行对比分析. 实验结果表明, 改进后的 U-Net 和使用 smooth L_1 损失作为生成器损失可以生成更好的着色图像; 而 L_1 损失能更好地保持图像结构信息, 使用梯度惩罚可以加速模型的收敛速度, 提高模型稳定性和图像质量; 该方法能更好地学习图像的深层次特征, 减少图像着色模糊现象, 在有效地保持图像结构相似性的同时提高图像着色质量.

关键词: 图像着色; 生成对抗网络; 损失函数; 梯度惩罚

中图分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18596

Pix2Pix-Based Grayscale Image Coloring Method

Li Hong'an¹⁾, Zheng Qiaoxue¹⁾, Zhang Jing¹⁾, Du Zhuoming^{2)*}, Li Zhanli¹⁾, and Kang Baosheng³⁾

¹⁾ (College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054)

²⁾ (Computer Engineering School, Jiangsu University of Technology, Changzhou 213001)

³⁾ (School of Information Science and Technology, Northwest University, Xi'an 710127)

Abstract: In this study, a grayscale image coloring method combining the Pix2Pix model is proposed to solve the problem of unclear object boundaries and low image coloring quality in colorization neural networks. First, an improved U-Net structure, using eight down-sampling and up-sampling layers, is adopted to extract features and predict the image color, which improves the network model's ability to extract deep image features. Second, the coloring image quality is tested under different loss functions, L_1 loss and smooth L_1 loss, to measure the distance between the generated image and ground truth. Finally, gradient penalty is added to improve the network stability of the training process. The gradient of each input data is penalized by constructing a new data distribution between the generated and real image distribution to limit the discriminator gradient. In the same experimental environment, the Pix2Pix model and summer2winter data are

收稿日期: 2020-08-20; 修回日期: 2021-01-17. 基金项目: 国家自然科学基金(61902311); 陕西省自然科学基金基础研究计划(2020JQ-758, 2019JM-162). 李洪安(1978—), 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、可视媒体计算; 郑峭雪(1997—), 男, 硕士研究生, 主要研究方向为可视媒体计算、人工智能; 张婧(1988—), 女, 博士, 讲师, 硕士生导师, 主要研究方向为计算机图形学、可视媒体计算; 杜卓明(1980—), 男, 博士, 副教授, CCF 会员, 论文通讯作者, 主要研究方向为最优化理论、模式识别; 李占利(1964—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究方向为计算机图形学、可视媒体计算; 康宝生(1961—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究方向为计算机辅助几何设计、可视媒体计算.

utilized for comparative analysis. The experiments demonstrate that the improved U-Net using the smooth L_1 loss as generator loss generates better colored images, whereas the L_1 loss better maintains the structural information of the image. Furthermore, the gradient penalty accelerates the model convergence speed, and improves the model stability and image quality. The proposed image coloring method learns deep image features and reduces the image blurs. The model raises the image quality while effectively maintaining the image structure similarity.

Key words: colorization; generative adversarial networks; loss function; gradient penalty

随着科学技术的日益发展,硬件设备的更新换代更加频繁,其质量和速度有着极大提高.在互联网时代,对信息的获取变得更加迫切,数字图像能表达出更加丰富的信息,因此人们对图像的质量要求也不断提高.灰度图像着色作为计算机视觉和数字图像处理技术的一个重要研究分支,在软件方面弥补人们对彩色图像的质量需求.在现实世界中的图像大多包含复杂的物体关系,并在一定程度上存在相似性,因而在图像着色过程中容易造成边界模糊等问题.图像的结构信息蕴含物体的特有属性,根据灰度图像的内容结构信息进行图像着色工作,可在一定程度上提高图像质量.图像着色方法在老旧黑白照片修复、黑白影视作品等图像视频渲染方面有着广泛的应用^[1-3],它按照着色方式的不同可分为 3 类.

(1) 基于颜色标记的图像着色方法.用户给出颜色点或颜色线条指导图像着色. Levin 等^[4]认为,图像在亮度信息相似的相邻像素之间也应该具有相似的颜色,并通过模型优化将标记的颜色扩散到整个图像.该方法为图像着色奠定了基础,但是想要得到良好的着色效果需要大量的颜色标记.为解决需要用户提供大量颜色点的问题,李洪安等^[5]提出一种基于分块特征的交互式图像色彩编辑方法对图像进行分块处理,根据亮度信息和色度信息的线性相关性,使用分块的亮度信息和色度信息计算分块图像的彩色特征,并建立优化模型,最后通过求解优化模型将局部的颜色标记进行扩散,得到着色图像.基于颜色标记的方法可以按照用户意愿对图像着色,但是着色需要人为提供颜色点,对艺术构图能力有一定要求,并且对细小物体的着色容易出现边界模糊问题.

(2) 图像风格迁移方法.使用给定的参考图像对图像进行彩色化.朱黎博等^[6]使用参考图像的颜色对目标灰度图像进行局部上色,然后对上色后的图像像素点进行遍历,将错误着色的像素点颜

色去掉,最后使用 Levin 等^[4]的算法将局部颜色扩散到整个图像.曹丽琴等^[7]在图像的分类结果基础上用参考图像的纹理特征在目标图像上寻找匹配的像素点,然后用局部自适应的权重均值滤波对置信度较高的像素点进行着色,对于置信度较低的像素点采用 Levin 等^[4]的优化算法进行颜色扩散. Li 等^[8]在局部范围上考虑参考图像合适的匹配尺度,然后采用全局优化的方法进行融合,使匹配误差和尺度空间变化达到最小以产生彩色图像.该类图像着色方法有良好的控制性,着色后的灰度图像在色彩上与参考图像基本保持一致,有效地减少用户指导的程度.但参考图像要求尽可能与目标图像在结构和空间上具有相似性,否则会导致颜色迁移偏差,因此对人为选择参考图像的要求较高.

(3) 基于深度学习的图像着色方法.从大规模的彩色图像学习颜色规律,然后将其应用在目标图像中以达到彩色化效果. Zhang 等^[9]通过学习图像的语义相似性来预测用户行为,从而在用户可控的情况下减少工作量,但在使用颜色点进行着色时也会存在误着色. Sangkloy 等^[10]将带有颜色线条的草图输入生成对抗网络,使用编码器和解码器网络结构对图像进行特征提取和上采样操作,使用像素级 L_1 距离损失保持着色图像与颜色标记的一致性. Isola 等^[11]改进传统的生成对抗网络 (generative adversarial networks, GAN),用真实图像代替随机噪声输入,利用 U-Net 提取图像特征信息,使用 L_1 距离损失和 PatchGAN 产生更加接近于真实图像的着色结果.该方法克服了需要人为干预的缺点,但图像的着色效果受图像类别影响较大,对不在训练样本类别中的图像着色效果不理想. He 等^[12]提出将深度学习和参考图像结合的图像着色方法,对目标图像结构一致的参考图像直接进行颜色迁移,否则从大规模的数据中学习颜色对目标图像进行着色,提高了着色可控性.

虽然利用深度学习进行图像着色无需人为参与, 但模型受数据集主要色调影响较大, 导致着色效果一般化, 要获得较好的着色图像仍然依赖大量的图像训练. 在小样本数据学习过程中, 对模型的学习能力、泛化能力和鲁棒性的要求也更高^[13-14]. 因此, 本文提出基于小样本的深度学习图像着色模型, 在 Pix2Pix 网络结构基础上, 使用改进的 GAN 提取亮度图像特征信息并预测颜色信息; 为提高模型的泛化能力, 对模型的训练策略进行调整; 分析 L_1 损失和 smooth L_1 损失对图像着色效果的影响; 引入梯度惩罚增加网络训练的稳定性, 提高图像的着色质量. 对基于深度神经网络的图像着色方面提供改进着色模型, 并通过实验证明提高模型能力方法的可行性.

1 Pix2Pix 图像着色模型

随着深度学习的发展, 使用人工神经网络进行图像处理的技术越来越多, 在基于 ImageNet 数据集的 ILSVRC 等图像分类识别挑战赛中出现大量的神经网络模型^[15-16]. 随后提出的 GAN 模型更是在深度学习领域掀起了一场革命, 其高质量图像的生成能力被广泛应用于数字图像处理中^[17]. 其中, Pix2Pix 模型由 Isola 等^[11]于 2017 年提出, 它由 U-Net 和 PatchGAN 组成, 分别充当 Pix2Pix 模型中的生成器和判别器. 该模型使用户只需提供一个草图便能生成一个与之对应的高质量图像; 对应到图像着色工作中, 网络接收真实图像的亮度信息, 对亮度信息进行特征提取并预测图像颜色值. 对于给定的亮度信息 $x \in \mathbb{R}^{H \times W \times 1}$ (在 Lab 颜色空间下的 L 亮度分量) 和随机噪声 z , 生成器网络 G 生成彩色图像 $y \in \mathbb{R}^{H \times W \times 3}$ (在 Lab 颜色空间下的生成图像), 有 $G: \{x, z\} \rightarrow y$. 判别器网络 D 比较真实图像和生成器生成的图像的相似性概率, 使生成器 G 生成的图像接近真实图像, Pix2Pix 模型的损失函数可以表示为

$$\mathcal{L}(G, D) = E_{x,y} [\ln D(x, y)] + E_{x,y} [\ln(1 - D(y, G(x, z)))].$$

其中, E 表示期望值; x 表示 Lab 颜色空间下的真实亮度分量; y 表示 Lab 颜色空间下的真实图像; z 表示随机噪声. 通过最小化生成器网络损失函数和最大化判别器网络损失函数来产生彩色图像, 即 $G^* = \arg \min_G \max_D \mathcal{L}_{\text{GAN}}(G, D)$. 为使判别器在无法辨别生成图像真假的同时保持生成图像尽可能地接

近真实图像, Pix2Pix 模型添加 L_1 距离损失函数到生成器网络中, 最终模型的优化目标为

$$G^* = \arg \min_G \max_D \mathcal{L}(G, D) + \lambda E_{x,y,z} [\|y - G(x, z)\|_1].$$

其中, E 表示期望值; λ 为 L_1 距离损失的惩罚系数. 通过优化最终的神经网络模型, 使生成器与判别器之间达到纳什平衡, 进而提高模型的着色能力. 值得注意的是, Pix2Pix 模型是在 conditional GAN 的基础上进行的, 所以在理论上网络输入中存在随机噪声, 但在实际的操作中舍弃了随机噪声, 只保留真实输入.

图 1 所示为 Pix2Pix 模型进行图像着色的示意图. 比较图 1a 和图 1b, 可以明显地看出, 该模型能很好地进行图像着色, 能够通过图像的灰度信息预测颜色信息. 从图 1b 和图 1c 可以看出, 着色图像整体色调与真实图像基本一致, 也保持了来自灰度图像的纹理特征. 但是, 图 1b 相对于图 1c 还有一定差距, 色彩鲜艳程度不如真实图像, 在一些细节方面不如真实图像表现细腻, 存在误着色情况.

图 1 Pix2Pix 模型图像着色效果图

2 改进图像着色模型

2.1 网络结构

U-Net^[18]最初被用于医疗图像分割. 随着深度学习的发展, 由于其具有良好的图像压缩和去噪能力, 因此被应用于图像处理中. 与类似的 DeconvNet^[19]和 SegNet^[20]等端到端网络结构相比, U-Net 使用直接特征融合替代池化索引方法, 在下采样和上采样之间没有内存消耗较大的全连接, 因而使用它可以在保持图像结构信息的基础上, 更加快速地进行图像着色. 图像的颜色信息与其结构内容有着极大的相关性, 为充分提取灰度图像中的结构信息, 本文设计如图 2 所示更深层次的 U-Net 结构. 网络中所有的采样层用蓝色表示, 网络输入的亮度信息用灰色表示, 网络输出的色度信息用彩色表示. 图 2 中的数字表示每层特征图像的宽、高和特征通道数.

图 2 U-Net 结构

与经典 U-Net 的 4 层对称网络结构相比, 本文采用 8 个下采样层和 8 个上采样层对图像进行特征提取和颜色预测. 扩大卷积核的感受视野, 由原来 3×3 的卷积核变为 4×4 的卷积核, 设置步长为 2、填充像素为 1 的卷积操作对图像进行深层次特征提取; 同时, 为避免在进行池化操作过程中丢失图像的特征信息, 去掉最大池化操作. U-Net 中的采样层均采用全卷积网络结构, 每个采样层包括卷积、归一化和激活处理. 在下采样模块中的每个激活层采用 Leaky ReLU 函数, 在上采样的前 7 个激活层采用 ReLU 函数, 最后 1 层采用 Tanh 激活函数.

在每次卷积之后, 将图像的尺寸缩小为原来的 $1/2$, 经过 8 个下采样层后, 图像压缩 256 倍, 通道数量变为 512. 为保持图像在分辨率大小上的一致性, 8 个上采样层同样也采用卷积核大小为 4、步长为 2、填充像素为 1 的反卷积操作. 通过跳跃连接将每次上采样后的图像特征与下采样层提取的图像特征进行融合并输出, 使图像的低级结构信息和高级语义特征得到充分利用, 具体网络参数如表 1 所示.

表 1 U-Net 网络结构参数

卷积层	卷积核尺寸	反卷积层	卷积核尺寸	步长
1	[4, 4, 1, 64]	9	[4, 4, 512, 512]	2
2	[4, 4, 64, 128]	10	[4, 4, 1 024, 512]	2
3	[4, 4, 128, 256]	11	[4, 4, 1 024, 512]	2
4	[4, 4, 256, 512]	12	[4, 4, 1 024, 512]	2
5	[4, 4, 512, 512]	13	[4, 4, 1 024, 256]	2
6	[4, 4, 512, 512]	14	[4, 4, 512, 128]	2
7	[4, 4, 512, 512]	15	[4, 4, 256, 64]	2
8	[4, 4, 512, 512]	16	[4, 4, 128, 2]	2

2.2 网络训练策略

因为在进行神经网络模型训练时, 通常以批量数据的形式直接输入网络中, 对于每一批数据具有不同的分布, 为加速模型的收敛速度、提高模型的准确率, 通常采用批量归一化(batch normali-

zation, BN)操作^[21]. 在 Pix2Pix 模型中利用 BN 对数据进行归一化, 使数据的分布服从均值为 0, 方差为 1 的标准正态分布. BN 更加注重对每个 batch 中的数据进行归一化, 使其数据分布一致. 本文的做法与其不同, 而是使用实例归一化(instance normalization, IN)对数据进行归一化. 根据图像在亮度相似的相邻像素之间具有相似的色度信息原理进行图像着色, 着色的结果取决于每个图像的亮度信息^[4]. IN 仅在横向和纵向上对单个图像进行归一化, 更能反映单个图像的数值分布、加速模型收敛速度、保持每个图像之间的独立性.

dropout 技术用于解决神经网络在训练过程中产生的过拟合现象, 由 Krizhevsky 等^[22]于 2017 年提出并应用于全连接层中, 其思想是在每个神经元上增加一个概率输出, 使网络的输出不会过度依赖某些特定的神经元、特征之间的协同作用减弱, 从而达到提高网络泛化能力的作用. 在本文中, 为增强网络的健壮性、提高模型预测图像颜色的质量, 在 U-Net 中的第 4~7 层(下采样层)和第 10~12 层(上采样层)添加 dropout, 并根据经验将 dropout 系数设置为 0.5.

2.3 L_1 损失与 smooth L_1 损失

在保证网络模型能够生成高分辨率的着色图像时, 还要尽可能地让生成的着色图像接近真实图像, Pix2Pix 模型采用 L_1 损失函数作为衡量标准对生成图像和真实图像进行比较. L_1 损失函数梯度值为常数, 在学习率不变的情况下, 损失函数将在稳定值附近波动, 很难收敛到更高的精度. 因此, 在图像着色过程中使用 L_1 损失仍然会产生着色模糊的问题. smooth L_1 结合了 L_2 损失函数和 L_1 损失函数的优点, 对异常值不敏感, 梯度值随差异值不断改变, 既可进一步使模型收敛, 还可以防止梯度爆炸. 因此, 本文采用 smooth L_1 损失函数替代 L_1 损失函数作为对照, 分析不同损失函数对图像着色质量的影响. smooth L_1 损失函数的计算式为

$$S = \begin{cases} \frac{1}{2} [Y(i, j) - f(i, j)]^2, & |Y(i, j) - f(i, j)| < 1 \\ |Y(i, j) - f(i, j)| - \frac{1}{2}, & |Y(i, j) - f(i, j)| \geq 1 \end{cases}$$

其中, Y 表示真实值; f 表示预测值; (i, j) 表示每个像素点. L_1 损失函数在网络训练进行到后期时, 真实值与预测值之间的差距较小, 然而梯度值始终保持不变, 使模型训练不稳定. smooth L_1 损失是对 L_1 损失的改进, 在真实值与预测值之间相差

较小时, 梯度值仍然会发生变化, 且变化不会过大, 相比 L_1 损失更能反映出细节信息. L_1 损失函数和 smooth L_1 损失函数图像如图 3 所示.

图 3 损失函数图像

加入 smooth L_1 损失作为生成器损失的一部分后, 生成器优化目标为

$$G^* = \arg \min_{G,D} \mathcal{L}(G,D) + \lambda S(G).$$

其中, S 表示 smooth L_1 损失函数; λ 表示其系数.

2.4 梯度惩罚

生成器的优化目标是使真实分布和生成分布之间的散度最小化, 在判别器网络达到最优, 并且 2 个分布没有重叠部分或重叠部分可忽略的情况下, 生成器的 JS 散度为 $\ln 2$, 导致梯度为 0, 因此网络模型无法继续更新参数. Arjovsky 等^[23]提出 WGAN, 使用 Wasserstein 距离对真实图像和生成图像进行度量, 即使在 2 个分布没有重叠的情况下, 也可以平滑地反映分布远近. 为满足 Lipschitz 连续性, 使用权重剪枝将判别器网络参数限定在一定范围之内, 根据经验将这个范围设置为 $[-0.01, 0.01]$, 在实际网络训练中大部分参数都落在这个范围的两端, 容易导致梯度消失和梯度爆炸. Gulrajani 等^[24]为解决这个问题, 改变对判别器网络梯度限制的方法, 在生成图像和真实图像分布之间采用线性插值的方法构造新的数据分布对每个输入数据进行梯度惩罚 (gradient penalty, GP), 提高网络模型的收敛速度和稳定性, 新的判别器损失为

$$D^* = \frac{E}{\tilde{x} \sim P_g} [D(\tilde{x})] - \frac{E}{x \sim P_r} [D(x)] + \lambda \frac{E}{\hat{x} \sim P_{\hat{x}}} \left[\left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1 \right)^2 \right].$$

其中, E 表示期望值; $\tilde{x}$ 为服从生成器分布的生成图像; $\hat{x}$ 为使用真实分布和生成分布构造的新图像; x 为服从真实分布的真实图像.

3 实验

3.1 实验数据及环境

使用公开数据集进行实验, 选用 CycleGAN 中的 summer2winter 图像数据集, 其中分别包括 1231

幅和 309 幅分辨率为 $256 \times 256 \times 3$ 的彩色 RGB 训练数据集和测试数据集. 在实验中, 对数据集进行预处理, 即对彩色图像进行颜色空间转换, 将 RGB 颜色空间转换为 Lab 颜色空间, 并提取 L 亮度通道信息和 a , b 色度通道信息, 数据预处理如图 4 所示.

图 4 网络输入

所有实验均在同一台计算机上完成, 使用 64 位 Windows 10 操作系统、处理器为 Intel(R)Core(TM) i7-7 500U CPU @ 2.70 GHz 2.90 GHz、显卡为 NVIDIA GeForce 940MX 的硬件设备, 并在 Python 3.7.3, PyTorch1.1.0, CUDA9.0 的环境下运行.

3.2 实验流程及评估指标

采用交替训练生成器网络 G 和判别器网络 D 的方法, 设定 dropout=0.5, 归一化操作为 IN.

预处理阶段, 将每幅输入图像从 RGB 颜色空间转为 Lab 颜色空间并分离 L 亮度通道作为网络输入图像, a 和 b 色度通道作为目标图像.

训练阶段, 将由网络 G 生成的 a 和 b 颜色分量与原始 L 亮度分量结合, 输入网络 D 与真实图像进行比较. 在训练过程中更新网络 G 和网络 D 的参数, 经过 200 次迭代后得到训练好的网络模型.

测试阶段与训练阶段大致相同, 使用训练好的模型进行测试, 将在测试阶段生成的所有图像从 Lab 颜色空间转换为 RGB 颜色空间并保存. 对于 Pix2Pix 模型, 直接使用本文在网络上公开的源代码进行实验.

为更加客观地反映不同模型产生的着色图像质量, 采用峰值信噪比 (peak signal to noise ratio, PSNR)、结构相似性 (structural similarity, SSIM) 和余弦相似度 (cosine similarity, CS) 对生成图像进行评估. PSNR 评价重构图像接近真实图像的程度, SSIM 评价图像的结构相似性, 弥补 PSNR 不能完全反映出图像质量与人眼视觉效果一致性的缺陷^[25]. CS 对图像像素逐个计算颜色角度误差评估不同模型的生成图像与真实图像之间的差距.

3.3 实验结果及分析

3.3.1 深层次 U-Net 结构

为验证本文改进后的网络模型可以提取图像更加深层次的特征信息并改进图像着色质量, 采

用 L_1 损失函数作为生成器损失, 并设置与 Pix2Pix 模型相同的惩罚系数进行实验, 图 5 所示为使用改进的网络模型与 Pix2Pix 模型生成的着色图像。

图 5 Pix2Pix 与本文方法的彩色化效果

由于改进的 Pix2Pix 模型采用更加深层次的 U-Net 结构, 模型能更好地提取图像中的深层特征信息, 学习到大规模图像中的颜色规律, 对天空的蓝色和云朵的白色进行正确着色. 相比于使用 Pix2Pix 模型的着色结果, 本文模型减少了大面积的黄色误着色区域, 着色结果更加接近真实图像。

图 6 所示为不同模型在训练过程中生成器损失函数的变化图像, 图 6 中蓝色线条代表 Pix2Pix 模型, 黄色线条代表本文着色模型. 可以看出, 在使用与 Pix2Pix 模型对 L_1 损失相同的惩罚系数下, 因为本文使用的深层次 U-Net 结构能提取图像的细节信息, 缩小生成图像与真实图像之间的误差距离, 使在模型训练过程中均方误差(mean square error, MSE)损失和 L_1 损失变化幅度更小, 变化更加稳定, 模型的训练结果更加接近真实值。

图 6 Pix2Pix 与本文方法的损失变化

3.3.2 L_1 和 smooth L_1 函数对图像着色的影响

图 7 显示在使用改进 U-Net 模型的基础上, 使

用 L_1 函数和 smooth L_1 函数进行训练的损失变化图像. 图 7 中 G_MSE 表示生成器均方差损失; G_L_1 和 $G_smooth L_1$ 分别表示生成图像与真实图像之间的 L_1 损失和 smooth L_1 损失; D_real 和 D_fake 分别表示对真实数据和生成图像的判别损失. 可以看出, 使用 smooth L_1 度量真实图像与生成图像之间的差距比 L_1 得到的损失变化更加稳定, 尤其在网络训练后期 smooth L_1 损失变化的幅度更小. 这是由于 smooth L_1 损失函数结合了 L_2 和 L_1 损失函数的特点, 对于样本中的离群点敏感度降低, 对位于样本中心的数据更加敏感, 使得在进行模型训练时更具控制性, 从而模型训练更加稳定。

图 7 使用 L_1 和 smooth L_1 函数的损失变化

但是在图像的着色效果对比中, 图 8b 的着色图像质量比图 8c 更好, 没有大面积的误着色现象. 分析其原因, 模型训练时受生成器网络、判别器网络和学习率等因素影响, 而在使用 smooth L_1 进行训练的过程中, 判别器网络发生不稳定现象, 导致模型训练结果不佳。

图 8 使用 L_1 和 smooth L_1 函数的着色效果

3.3.3 梯度惩罚对模型稳定性的影响

从图 7b 中可以看出, 虽然使用改进着色模型

可以提高图像着色质量,但是在训练过程中模型并不稳定,损失函数存在突变等情况.为验证加入梯度惩罚后能够有效地增加模型稳定性,在上述实验的基础上增加梯度惩罚项进行200次训练,实验结果如图9所示.

图9 使用 L_1 +GP 和 smooth L_1 +GP 模型的着色效果

由于梯度惩罚只对单个样本施加 Lipschitz 限制,使用线性插值的方法在真实图像和生成图像间构造新分布并对其集中区域进行判别,所以有较强的可控性.从图9a和图9b中可以看出,加入梯度惩罚后在一定程度上降低图像的模糊程度,相比图8a和图8b减少了白云部分的误着色,水面的细节信息也更加接近真实图像.加入梯度惩罚后,判别器网络的参数更新将受到约束,使其在一个更加合理的范围内进行调整,从而影响生成器网络的优化,因此各项损失值变化更加稳定,使模型更加容易收敛,损失函数变化如图10所示,其中 D_GP 表示梯度惩罚值.

图10 使用 L_1 +GP 和 smooth L_1 +GP 模型的损失变化

3.3.4 定量评估

表2所示为对不同模型生成的图像使用 PSNR

指标进行评价的结果,可以看出 smooth L_1 +GP 模型在最大和平均 PSNR 中取得分数最高,相比 Pix2Pix 模型分别提高 6.061 dB 和 1.338 dB.改进的着色模型使用 L_1 损失作为优化目标,取得的高质量图像数量占比最多,与 Pix2Pix 模型相比提高了 9.061%.虽然改进后的图像着色模型能更好地学习图像的颜色规律,但由于测试数据集中存在与训练分布相差较大的图像,所以对该类图像的着色效果不好,导致最小 PSNR 降低.从整个 PSNR 表中可以看出,本文提出的改进图像着色模型生成的图像质量更好,提高了图像的整体着色水平.

表2 不同模型及损失函数的 PSNR 指标 dB

网络模型及损失函数	最大值	最小值	平均值	>30/%
Pix2Pix(L_1) ^[11]	29.408	11.085	23.024	0.000
本文(L_1)	34.175	9.691	24.293	9.061
本文(L_1 +GP)	34.437	10.175	24.265	7.119
本文(smooth L_1)	33.845	9.344	24.001	7.443
本文(smooth L_1 +GP)	35.469	9.827	24.362	8.737

注:表中加粗字体表示最佳结果.

对不同模型的 SSIM 评价指标如表3所示,可以看出使用相同损失函数进行评价真实图像与生成图像误差时,加入梯度惩罚后生成图像与真实图像之间的平均 SSIM 有所提高.此外, L_1 +GP 模型在 SSIM 指标上明显优于其他模型,相比 Pix2Pix 模型分别提高 6.001、21.242 和 10.299 个百分点,充分证明了利用 L_1 函数作为生成器损失可以提高生成图像与真实图像的结构相似性.

表3 不同模型及损失函数的 SSIM 指标 %

网络模型及损失函数	最大值	最小值	平均值	>90%
Pix2Pix(L_1) ^[11]	92.238	52.554	82.150	2.912
本文(L_1)	97.703	69.184	91.366	72.815
本文(L_1 +GP)	98.239	73.796	92.379	78.964
本文(smooth L_1)	97.944	68.045	91.510	72.815
本文(smooth L_1 +GP)	97.584	72.786	91.849	76.375

注:表中加粗字体表示最佳结果.

表4所示为不同模型在 CS 指标下的评估结果,可以看出,本文提出的改进图像着色模型所得到的角度误差值更小,即生成颜色向量和真实颜色向量之间的夹角更小,说明模型对灰度图像所对应的像素颜色值预测更加精确,进一步证明

了使用本文改进的网络模型具有更强的图像着色能力.

表 4 不同模型及损失函数的 CS 指标

网络模型及损失函数	最大值	最小值	平均值	中值
Pix2Pix(L_1) ^[11]	30.116	3.192	10.154	9.358
本文(L_1)	30.651	1.252	9.656	8.978
本文(L_1 +GP)	28.988	0.981	9.817	9.096
本文(smooth L_1)	34.658	1.283	10.147	9.394
本文(smooth L_1 +GP)	32.336	2.021	9.735	9.067

注: 表中加粗字体表示最佳结果.

3.3.5 用户评估

量化的评价指标固然能够更加客观地反映出不同图像着色模型的性能好坏,但也存在一定的缺陷.基于深度神经网络的灰度图像着色模型学习图像亮度信息与色彩信息之间的相关性,图像的着色效果是无法预期的.对相同灰度图像进行着色,其着色效果可能不忠实于真实图像,但如果整体的图像着色效果能够达到丰富图像色彩信息、增强图像的表达效果,那么该模型的着色能力应当也是被认可的.因此,增加基于用户评价的图像评价指标,从 Pix2Pix 模型和本文模型所生成的所有着色图像中,随机挑选出一定数量的不同着色图像,在未告知用户每幅着色图像是由哪一个模型生成的情况下,让用户根据自己的主观意向挑选出着色效果更佳图像.

基于用户主观意向的评估方法总共有 50 名参与者,并使用随机算法从 Pix2Pix 模型和本文模型生成的着色图像中选择 30 对图像进行实验.向每名参与者同时展示由 2 个模型分别对相同灰度图像生成的着色效果图像,并在 5 s 的时间之内选出被认为着色效果更加自然、更加符合一般视觉效果的图像,参与者在每次做出自己的判断后都不被告知图像的生成模型.表 5 所示为基于用户研究的图像质量评价结果,图像右侧的数值表示参与者认为该模型产生的着色效果更好的百分比.可以看出,本文提出的图像着色方法在某些场景下得到的着色效果比 Pix2Pix 模型更好,更加符合人眼的视觉感受.

为进一步说明改进的着色方法在保持图像结构、改善图像着色模糊和提高图像着色质量方面的有效性,本文增加了基于深层卷积生成对抗网络(deep convolutional generative adversarial networks, DCGAN)^[26]的图像着色方法作为对照实验,使用

表 5 用户对不同模型着色实验评价对比

Pix2Pix ^[11]		本文	
效果图	认可度/%	效果图	认可度/%
19.67		83.33
20.00		80.00
33.33		66.67
50.00		50.00
57.67		43.33
76.67		23.33

更多的图像进行对比,如图 11 所示.可以看出, Pix2Pix 模型能对图像进行着色,但是在还原真实图像、处理图像细节信息上还有待提高.而 DCGAN 模型得到的图像着色效果不佳,整体的图像色调偏淡,图像的色彩信息不够丰富.本文着色模型能更好地捕捉灰度图像的细节信息和提取更多的特征信息,不仅能够较好地保持图像结构相似性,还可以提高图像着色质量.

4 结 语

本文提出改进的 U-Net 结构,可以在更深层次提取图像的特征信息.使用 smooth L_1 损失可以有效地改善图像着色边界不明确的问题, L_1 损失能较好地保持图像结构信息;通过改变网络的训练策略和引入梯度惩罚的方法,可进一步提高图像的着色质量和网络稳定性.实验表明,本文提出的灰度图像着色方法在提高图像质量和保持图像结构信息方面取得了更好的性能表现,并且证明了该方法的有效性和必要性.然而,本文的实验是针对小样本数据集进行的,着色图像类别有限,在后续研究中将使用更加丰富的图像数据集,以进一步提升模型训练效率.

图 11 不同模型下的不同着色图像

参考文献(References):

- [1] Iizuka S, Simo-Serra E, Ishikawa H. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification[J]. *ACM Transactions on Graphics*, 2016, 35(4): Article No.110
- [2] Zhang R, Isola P, Efros A A. Colorful image colorization[C] // *Proceedings of European Conference on Computer Vision*. Heidelberg: Springer, 2016: 649-666
- [3] Deshpande A, Lu J J, Yeh M C, *et al.* Learning diverse image colorization[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2017: 2877-2885
- [4] Levin A, Lischinski D, Weiss Y. Colorization using optimization[J]. *ACM Transactions on Graphics*, 2004, 23(3): 689-694
- [5] Li Hong'an, Zhang Min, Du Zhuoming, *et al.* Interactive image color editing method based on block feature[J]. *Infrared and Laser Engineering*, 2019, 48(12): 1226003(in Chinese)
(李洪安, 张敏, 杜卓明, 等. 一种基于分块特征的交互式图像色彩编辑方法[J]. *红外与激光工程*, 2019, 48(12): 1226003)
- [6] Zhu Libo, Sun Shaoyuan, Gu Xiaojing, *et al.* Image colorization based on color transfer and propagation[J]. *Journal of Image and Graphics*, 2010, 15(2): 200-205(in Chinese)
(朱黎博, 孙韶媛, 谷小婧, 等. 基于色彩传递与扩展的图像着色算法[J]. *中国图象图形学报*, 2010, 15(2): 200-205)
- [7] Cao Liqin, Shang Yongxing, Liu Tingting, *et al.* Novel image colorization of a local adaptive weighted average filter[J]. *Journal of Image and Graphics*, 2019, 24(8): 1249-1257(in Chinese)
(曹丽琴, 商永星, 刘婷婷, 等. 局部自适应的灰度图像彩色化[J]. *中国图象图形学报*, 2019, 24(8): 1249-1257)
- [8] Li B, Lai Y K, John M, *et al.* Automatic example-based image colorization using location-aware cross-scale matching[J]. *IEEE Transactions on Image Processing*, 2019, 28(9): 4606-4619
- [9] Zhang R, Zhu J Y, Isola P, *et al.* Real-time user-guided image colorization with learned deep priors[J]. *ACM Transactions on Graphics*, 2017, 36(4): Article No.119
- [10] Sangkloy P, Lu J W, Fang C, *et al.* Scribbler: controlling deep image synthesis with sketch and color[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2017: 6836-6845
- [11] Isola P, Zhu J Y, Zhou T H, *et al.* Image-to-image translation with conditional adversarial networks[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2017: 5967-5976
- [12] He M M, Chen D D, Liao J, *et al.* Deep exemplar-based colorization[J]. *ACM Transactions on Graphics*, 2018, 37(4): Article No.47
- [13] Shaham T R, Dekel T, Michaeli T. SinGAN: learning a generative model from a single natural image[C] // *Proceedings of the IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2019: 4569-4579
- [14] Yoo S, Bahng H, Chung S, *et al.* Coloring with limited data: few-shot colorization via memory augmented networks[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2019: 11275-11284

- [15] Hu S M, Liang D, Yang G Y, *et al.* Jittor: a novel deep learning framework with meta-operators and unified graph execution[J]. Science China Information Sciences, 2020, 63(12): Article No.222103
- [16] Russakovsky O, Deng J, Su H, *et al.* ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252
- [17] Goodfellow I, 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
- [18] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C] //Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Heidelberg: Springer, 2015: 234-241
- [19] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 1520-1528
- [20] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495
- [21] Luo P, Ren J M, Peng Z L, *et al.* Differentiable learning-to-normalize via switchable normalization[OL]. [2020-08-20]. <https://arxiv.org/pdf/1806.10779.pdf>
- [22] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90
- [23] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[OL]. [2020-08-20]. <https://arxiv.org/pdf/1701.07875.pdf>
- [24] Gulrajani I, Ahmed F, Arjovsky M, *et al.* Improved training of Wasserstein GANs[C] //Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 5769-5779
- [25] Wang Z, Bovik A C, Sheikh H R, *et al.* Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612
- [26] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[OL]. [2020-08-20]. <https://arxiv.org/pdf/1511.06434.pdf>