

结合全局语义优化的对抗性灰度图像彩色化

万园园, 王雨青, 张晓宁, 李荅群, 陈小林

引用本文:

万园园, 王雨青, 张晓宁, 等. 结合全局语义优化的对抗性灰度图像彩色化[J]. 液晶与显示, 2021, 36(9): 1305–1313.

WAN Yuan-yuan, WANG Yu-qing, ZHANG Xiao-ning, et al. Adversarial grayscale image colorization combined with global semantic optimization[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(9): 1305–1313.

在线阅读 View online: <https://doi.org/10.37188/CJLCD.2021-0012>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度卷积生成对抗网络的图像识别算法

Image recognition algorithms based on deep convolution generative adversarial network

液晶与显示. 2020, 35(4): 383–388 <https://doi.org/10.3788/YJYXS20203504.0383>

基于密集连接生成对抗网络的图像颜色迁移

Image color transfer with dense connections generative adversarial networks

液晶与显示. 2020, 35(3): 272–280 <https://doi.org/10.3788/YJYXS20203503.0272>

基于局部期望最大化注意力的图像降噪

Local expectation-maximization attention network for image denoising

液晶与显示. 2020, 35(4): 350–359 <https://doi.org/10.3788/YJYXS20203504.0350>

本征图像分解的稀疏表示彩色图像去噪算法

Color image denoising algorithm based on intrinsic image decomposition and sparse representation

液晶与显示. 2019, 34(11): 1104–1114 <https://doi.org/10.3788/YJYXS20193411.1104>

融合多尺度特征及注意力机制的医学图像检索

Medical image retrieval with multiscale features and attention mechanisms

液晶与显示. 2021, 36(8): 1174–1185 <https://doi.org/10.37188/CJLCD.2020-0248>

文章编号:1007-2780(2021)09-1305-09

结合全局语义优化的 对抗性灰度图像彩色化

万园园^{1,2}, 王雨青¹, 张晓宁¹, 李荅群¹, 陈小林^{1*}

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

摘要:针对当前灰度图像彩色化算法容易出现色彩枯燥、颜色溢出和图像细节损失等问题,本文提出一种结合全局语义优化的对抗性灰度图像彩色化算法。其中,生成网络采用自主改进的 U-Net 网络。一方面,改进的 U-Net 网络利用多层卷积对输入图像进行逐步下采样,在获取多尺度层级特征和全局特征的同时,在跳跃连接中将全局特征和多尺度层级特征进行自适应融合,从而有效增强算法对全局语义信息的理解能力并缓解颜色溢出现象;另一方面,改进的 U-Net 网络在上采样过程中融合通道注意力模块,使得在提取卷积特征时能够有效抑制噪声并降低特征冗余性。判别网络主要采用全卷积结构,通过反向传播误差以达到优化生成网络的目的。此外,本文算法的损失函数将 WGAN-GP 网络的优化思想和颜色损失相结合,从而解决传统生成对抗网络训练时出现的梯度消失和模式崩溃等问题。本文算法在 Place365 测试集上所获取的峰值信噪比、结构相似性和信息熵指标分别为 24.455 dB、0.943 和 7.489。实验结果表明,本文算法能够缓解颜色溢出,且细节保持和颜色饱和度方面都具有一定优势。

关键词:全局特征;通道注意力模块;WGAN-GP;图像彩色化

中图分类号:TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2021-0012

Adversarial grayscale image colorization combined with global semantic optimization

WAN Yuan-yuan^{1,2}, WANG Yu-qing¹, ZHANG Xiao-ning¹,
LI Da-qun¹, CHEN Xiao-lin^{1*}(1. *Changchun Institute of Optics, Fine Mechanics and Physics,*
*Chinese Academy of Sciences, Changchun 130033, China;*2. *University of Chinese Academy of Sciences, Beijing 100049, China*)

Abstract: The current colorization algorithms generally suffer from color boring, color bleeding and detail loss. In order to address these problems, this paper proposes a adversarial colorization algorithm combined with global semantic optimization. The generator adopts the improved U-Net network. On the one hand, the improved U-Net gradually obtains multi-scale features and global feature by multiple convolution during the downsampling process. At the same time, the global feature and

收稿日期:2021-01-12;修订日期:2021-01-28.

基金项目:国家重点研发计划(No.2018YFC0308100)

Supported by National Key Research and Development Project (No.2018YFC0308100)

* 通信联系人, E-mail: cl.hong@163.com

multi-scale features get adaptive fused in the skip connections, thus it can effectively enhance algorithm's ability to understand global semantic and ease the color bleeding. On the other hand, the improved U-Net network integrates the channel attention module in the upsampling process, which can effectively suppress the noise and reduce the feature redundancy when the convolutional features are extracted. The discriminator mainly adopts full convolution structure and achieves the purpose of optimizing the generator by reverse error transfer. In addition, the loss function combines the optimization idea of WGAN-GP with color loss to solve the gradient disappearance and mode collapse problems in the training of traditional GAN. The proposed algorithm is tested on Place365 test set, and the PSNR, SSIM and IE indexes are 24.455 dB, 0.943 and 7.489, respectively. The test results show that the proposed algorithm can alleviate color bleeding, and has certain advantages in detail retention and color saturation.

Key words: global feature; channel attention module; WGAN-GP; image colorization

1 引言

图像彩色化实质就是为灰度图像的每一个像素分配颜色^[1]。该技术在机器视觉、图像修复和影视制作等方面具有重要的应用价值。现有的彩色化方法有传统彩色化方法和基于深度学习的彩色化方法。

传统的彩色化方法可分成两种:基于局部颜色扩展的方法^[1-4]和基于参考图的颜色传递的方法^[1-4]。其中,基于局部颜色扩展的方法需要人为在灰度图像上进行颜色涂鸦,需要的用户操作量较大,在颜色扩散过程中易出现边缘误扩散等问题。基于参考图的颜色传递的方法需要选择一张与待彩色化灰度图相似的彩色图进行颜色迁移,虽然此类方法减小了用户操作量,但是对参考图的依赖程度高。由于传统算法受人工干预影响较大,且效率不高,在这个图像海量化的信息时代已经不再适用。

目前,基于深度学习的彩色化算法已经能实现自动彩色化。如 Cheng^[1]等人利用卷积神经网络来提取灰度图特征,使用联合双边滤波进行后续处理,实现了较好的彩色化效果。Iizuka^[2]等人利用双流结构,融合灰度图像的局部特征和全局特征,并添加标签分类来实现颜色预测,但生成图像的颜色饱和度较低且细节损失较多。Zhang^[3]等人将图像彩色化问题转化成对像素点颜色的分类问题,将饱和度高的颜色赋予高权重来获得鲜艳的颜色,但是容易造成颜色溢出的问题。Larsson^[4]等人使用 VGG 网络来解释场景

的语义组成和对象的定位,但仍存在细节损失等问题。

近年来,生成对抗网络(Generative Adversarial Network, GAN)因其特殊的网络结构和训练机制,在超分辨重建^[5]、图像合成^[6]与风格转换^[7]等领域都取得了巨大的成功。受此启发,GAN 网络也被用来解决图像彩色化的问题。Isola 等^[8]人提出使用 U-Net 与 GAN 相结合的方法,在一定程度上提升了彩色化效果。后来 Nazeri 等^[9]利用条件生成对抗网络^[8](conditional Generative Adversarial Network, cGAN)来还原彩色化的过程。Cao 等^[10]也通过在 cGAN 的多个卷积层中添加噪声来增加彩色化效果的多样性,但彩色化后颜色溢出的现象仍未改善。ChromaGAN^[11]是 Patricia 等人提出的一种最新的对抗性彩色化算法,他们把图像场景分类引入 GAN 网络,通过对颜色信息和类别分布的感知来实现灰度图像彩色化,可改善场景彩色化错误的现象。由于以上基于 GAN 的彩色化方法均利用相对较浅的低维特征,对图像的全局语义信息(例如场景的布局)理解不足,从而易导致颜色溢出等现象,且在特征提取时存在一定程度的细节损失。

针对以上问题,本文提出一种结合全局语义优化的生成对抗彩色化算法。通过对输入图像进行下采样获取多层特征和全局特征,并在跳跃连接中将得到的全局特征分别与多尺度层级特征融合来提升对图像整体语义信息的理解能力,且在上采样过程中引入通道注意力机制来抑制噪声和降低冗余特征的权重,增强有用特征的学习能

力^[13]。同时搭建判别网络,动态化地评价生成图像的质量,进一步提升颜色的丰富程度。本文的损失函数在传统颜色损失的基础上引入带梯度惩罚的 Wasserstein 生成对抗网络^[13-14] (Wasserstein Generative Adversarial Network with Gradient Penalty, WGAN-GP) 的优化思想,用 Wasserstein 距离^[15-16] 度量真假样本,以解决传统 GAN 在训练过程中出现的梯度消失和模式崩溃等问题。实验证明,相比于目前主流算法,本文算法的彩色化效果有一定提升。

2 全局语义优化的对抗性彩色化算法

由于传统的 RGB 模型并不能直观地反映亮

度信息,因此先把图像转换成 Lab 模型。生成网络的输入为亮度 L 通道。生成网络主要包括 3 部分:下采样、特征融合和上采样。首先下采样提取多尺度层级特征和全局特征,然后在跳跃连接中将得到的全局特征与每个尺度的层级特征融合,接着在上采样过程中重构图像,最终预测输出三通道 Lab 的彩色图。之后将生成的彩色图和真实的彩色图输入判别网络中进行识别。生成网络和判别网络在动态博弈中不断增强性能,最终达到生成高质量图像的目的。

2.1 生成网络

本文的生成网络在 U-Net^[18] 网络的基础上进行改进,加深了网络深度并引入自适应特征融合模块和注意力机制。具体结构如图 1 所示。

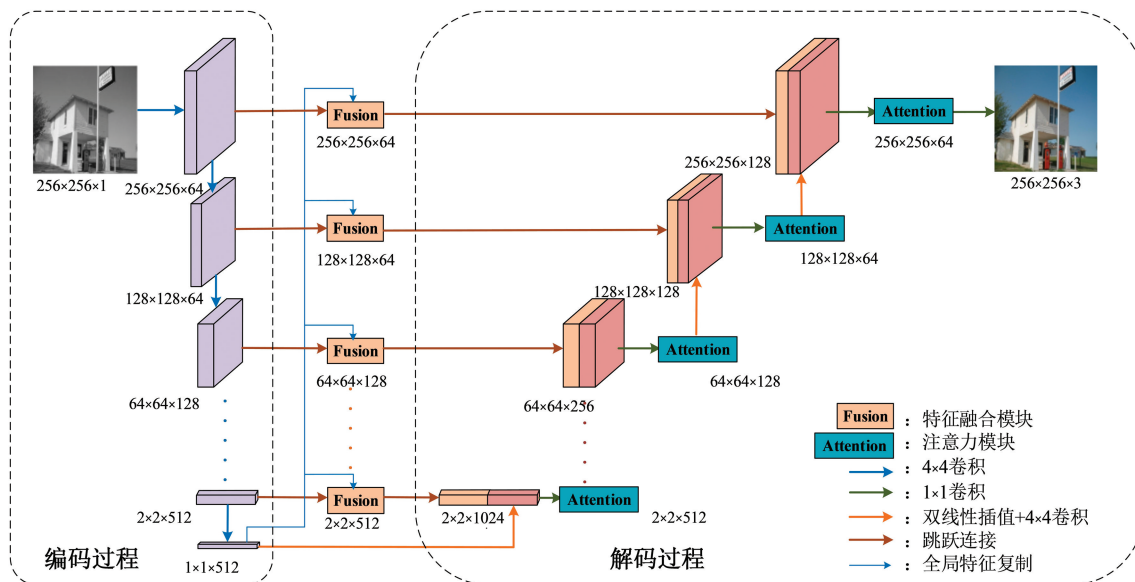


图 1 生成网络结构

Fig.1 Structure of generator

其中,为了减少最大池化带来的信息丢失,本文采用卷积核为 4,步幅为 2,填充为 1 的卷积层逐步下采样,依次获得不同尺度的层级特征和全局特征。本文在跳跃连接添加特征融合模块分别将全局特征和多尺度层级特征进行自适应融合。由于层级特征提取的是感受野相对较小的局部信息,而全局特征具有从整个图像的感受野中提取的高维信息(如场景布局 and 类型等),因此全局特征可作为一种先验信息来增强各个尺度的层级特征^[2,17],而且这两种特征的融合形成了对图像信息的高效表达,有利于生成颜色自然且细节丰富

的图像。

与此同时,在上采样中将得到的全局特征逐渐恢复高分辨率来构建图像。由于转置卷积易产生棋盘格噪声,本文采用双线性插值和 4×4 的卷积层组合的方式代替转置卷积来扩大图像尺寸。随后,按照 U-Net 的思想,在每个尺度下将已增强的低维层级特征通过跳跃连接与上采样的高维特征拼接融合,共享低维特征的空间信息和细节信息^[15,17]。为了提升算法特征学习的性能,添加了通道注意力机制^[18]对融合之后的特征进行通道加权,抑制噪声同时更关注与任务相关的特征

通道。

2.1.1 自适应特征融合模块

由于提取的层级特征感受野较小,主要包含局部特征信息。一般来说,全局特征反应图像的整体结构,例如场景布局和类型等^[17]。局部特征更加注重图像的细节信息,反应图像区域内更详细的变化。两者融合可以有效获取丰富的特征信息,并且能增强算法全局语义信息的理解能力。因此,本文在不同尺度上通过特征融合模块动态的将全局特征和包含局部信息的层级特征进行融合,如图 2 所示。

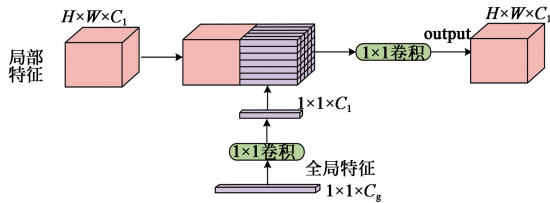


图 2 特征融合模块结构

Fig.2 Structure of feature fusion block

假设层级特征图 f_l 的尺寸为 $H \times W \times C_l$, 全局特征图 f_g 的尺寸为 $1 \times 1 \times C_g$ 。首先将全局特征 f_g 经过 1×1 的卷积层进行调整, 该卷积层可自适应地从全局特征中提取最有用的信息来增强通道数为 C_l 的层级特征。然后把卷积之后的全局特征复制 $H \times W$ 次, 使得到的全局特征尺度与待融合的层级特征尺度一致。最后将全局特征与层级特征进行拼接融合, 输出的特征尺寸为 $H \times W \times C_l$ 。融合公式如下:

$$f_{out} = \delta \left(\mathbf{b} + \mathbf{W} \begin{bmatrix} f_l \\ f_g \end{bmatrix} \right), \quad (1)$$

其中 $\mathbf{W}$ 是一个 $C_l \times 2 C_l$ 的权重矩阵, 而 $\mathbf{b}$ 是一个维度为 C_l 的向量, $\mathbf{W}$ 和 $\mathbf{b}$ 都是网络的可学习参数。

2.1.2 注意力模块

U-Net 结构的优势就是将下采样的多尺度特征通过跳跃连接在上采样过程中进行整合, 使得语义信息丰富的高维特征和包含细节信息的低维特征互补, 增强算法的性能。但是融合之后的特征层也包含了噪声响应和冗余特征信息, 会对图像彩色化造成干扰。因此, 为了能高效地从特征模块中获取有用特征信息, 引入了通道注意力机制使网络自动学习各个特征通道的相互依赖关

系, 抑制噪声并增强有用特征通道的权重。注意力机制模块结构如图 3 所示。

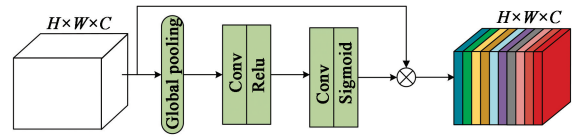


图 3 注意力模块结构

Fig.3 Structure of attention block

输入和输出的特征图尺寸均为 $H \times W \times C$, 首先对输入的特征图进行全局平均池化, 得到一个 $1 \times 1 \times C$ 的特征图。从某种程度来说, 该特征图具有全局感受野, 可看作是对各个通道的描述。全局平均池化公式如下:

$$Z_c = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j), \quad (2)$$

其中, $u_c(i, j)$ 表示输入特征图的第 c 个通道内位置为 (i, j) 的像素值, Z 为全局平均池化后的特征图。随后将特征图 Z 经过两个级联的卷积层和激活层来学习各个特征通道的相关性。第一个卷积层对特征通道生成尺度为 $1 \times 1 \times C/16$ 的特征图, 第二个卷积层再把特征图扩展回 $1 \times 1 \times C$, 经过 sigmoid 函数激活输出得到每个特征通道的重要性。最后得到的特征权重通过乘法加权在输入的特征上, 从而达到抑制噪声和增强与任务相关的特征通道权重的目的^[18], 在一定程度上提升了算法的性能。

2.2 判别网络

判别网络的目标是尽可能地区分生成图像和真实图像, 本文判别网络由 5 层全卷积层构成, 详细结构如表 1 所示。

表 1 判别网络结构

Tab.1 Structure of discriminator

卷积层	结构	步长	通道数
Conv1	4×4	2	64
Conv2	4×4	2	128
Conv3	4×4	2	256
Conv4	4×4	2	512
Conv5	4×4	1	1

判别网络前 4 层采用的是步长为 2 的 4×4 卷积层, 同时每次下采样之后特征图的通道数加

倍,每个卷积层后采用 Leaky Relu 激活函数。判别网络的最后一层将特征图映射成单通道输出。判别网络的输入是生成网络生成的彩色图和与之对应的真实彩色图,经过判别网络处理之后返回输入图像是真或假的概率值。判别网络对生成图像进行评估判别,通过优化损失函数,反向传播误差调整网络结构参数,以进一步改善生成网络的彩色化效果。

2.3 目标损失函数的设计

传统 GAN 大多基于 JS 散度进行优化,而 JS 散度易发散从而导致产生梯度消失的现象,不利于生成网络进一步的学习。WGAN-GP 网络采用了 Wasserstein 距离^[14]作为判别真假图像的依据,且在损失函数中添加了梯度惩罚使得判别网络满足 1-Lipschitz 限制,以解决传统 GAN 在训练过程中产生的梯度消失和模式崩溃等问题。

本文将 WGAN-GP 的优化思想与颜色损失相结合来设计算法的损失函数。生成网络的损失包括颜色损失和对抗损失两部分,具体表达式如下:

$$L_g(G, D) = L_c(G) + \lambda_1 L_{adv}(G, D), \quad (3)$$

其中: L_c 表示颜色损失, L_{adv} 表示对抗损失。由于 $L_{adv} : L_c = 1\ 000 : 1$, 通常情况下, 当对抗损失为主导时, 会使得网络权重被带偏, 因此需要设置 λ_1 用于平衡两类损失。考虑到算法的主要任务是学习颜色信息, 实验设置 λ_1 为 $1e-4$, 使得颜色损失占据主导地位。

对抗损失的具体定义如式 4:

$$L_{adv}(G, D) = -E_{\hat{x} \sim p_g} [D(G(\hat{x}))], \quad (4)$$

P_g 表示生成彩色图的数据分布, 颜色损失 L_c 表达式如式(5)所示, 它是计算真实彩色图与生成彩色图的 L_1 距离得到的。相比于 L_2 损失, L_1 损失能较少生成图像细节模糊的现象。其中 P_r 表示真实彩色图的数据分布。

$$L_c(G) = E_{(L, a_r, b_r) \sim p_r} \left[\|G(L) - (L, a_r, b_r)\|_1 \right], \quad (5)$$

其次, 判别网络的损失函数引入了 Wasserstein 距离^[14], 公式表达如式(6)所示。

$$L_d(G, D) = E_{\hat{x} \sim p_g} [D(G(\hat{x}))] - E_{x \sim p_r} [D(x)] + \lambda L_{gp}, \quad (6)$$

公式前两项表示真实数据分布与生成数据分布之间的 Wasserstein 距离。权重 λ 取 10。 L_{gp} 表示梯度惩罚项, 公式定义为:

$$L_{gp} = E_{\tilde{x} \sim p_{\tilde{x}}} [(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2]. \quad (7)$$

添加梯度惩罚的目的是使得判别网络满足 Lipschitz 限制。其中 $P_{\tilde{x}}$ 表示惩罚项的采样分布, 获取 $P_{\tilde{x}}$ 的方法为: 在真实数据分布 P_r 和生成数据分布 P_g 中各进行一次采样, 然后在两点连线处进行一次采样。采样公式表达为:

$$\tilde{x} = \epsilon x + (1 - \epsilon) \hat{x}, \quad (8)$$

ϵ 为从 $[0, 1]$ 的正态分布中选取的随机数, $x \sim p_r$, $\hat{x} \sim p_g$ 。

3 实验结果与分析

3.1 实验环境与设置

本实验在 Pytorch 深度学习框架下实现, 利用 TITAN RTX 加速训练。选用 Place365 场景数据集对网络进行端到端的训练, 该数据集包含 365 个不同场景的 180 多万张图片。训练之前先将数据集里少量灰度图和场景灰暗的图片滤除以保证算法能更好地学习颜色信息, 处理之后最终选取 1 687 424 张图片作为训练集, 10 000 张图片为测试集。

首先将输入图片颜色空间从 RGB 转化成 Lab, 将 L 通道作为生成网络的输入。由于网络存在尺度为 $1 \times 1 \times C$ 的全局特征, 需要固定输入尺寸的大小, 实验选用 256×256 的输入图像。算法采用 Adam 算法优化损失函数, 初始学习率 $l_r = 0.000\ 1$, $\beta_1 = 0.5$, $\beta_2 = 0.999$, 权重衰减为 0.005。每批次大小为 32, 共训练 10 个周期, 训练期间每迭代 1 000 个批次进行一次验证输出, 观察算法的彩色化效果。

3.2 评价指标

为了评估算法的彩色化效果, 本文分别从定性和定量两个方面进行分析。定性分析主要从视觉效果上进行评价, 定量分析采用峰值信噪比 (Peak Signal to Noise Ratio, PSNR)、结构相似性 (Structural Similarity, SSIM) 和信息熵 (Information Entropy, IE) 作为评价指标。峰值信噪比是一种使用广泛的客观评价指标, 其数值越大表示失真越小。结构相似性是衡量两幅图像相似度的指标, 而信息熵表示的是图像包含信息的丰富程度, 信息熵越大, 图像信息越丰富, 图像质量

越好。

3.3 不同算法的结果对比

为了证明本文算法的有效性,将本文算法与现有彩色化算法^[2-4,11]进行比较。实验结果如图 4 所示。

首先针对颜色溢出的问题,例如第 1 行中,由于缺乏对图像全局语义的理解,文献[2,4]算法倾向于把跑道赋予天空的蓝色,文献[3,11]算法稍好一些,但也将原本红色的跑道混合了少量蓝色,而本文算法较好地跑道的颜色恢复出来。在第 2 行,文献[3]算法生成了更接近真实图像饱和度高的红色,而鲜艳红色也违和地出现在轮胎上,文

献[2,4,11]算法得到的颜色饱和度较低且颜色不连续,本文算法在保持颜色饱和度较高的同时,也没有出现颜色溢出的现象。在第 3 行,文献[3]生成了紫色和黄色混合的花朵,呈现不自然的彩色化效果,文献[2]花瓣的颜色饱和度较低且分布不均匀,文献[4]的颜色鲜艳但是没有还原花蕊的颜色信息,文献[11]突出花蕊的细节信息,而有部分花瓣混合少量绿色。相较之下,本文算法生成的花朵颜色鲜艳且突出花蕊细节信息,在一定程度上能够减少颜色溢出。

针对细节损失的问题,在第 4 行,本文算法还原了泳池颜色这个细节信息,而且周围的草地树

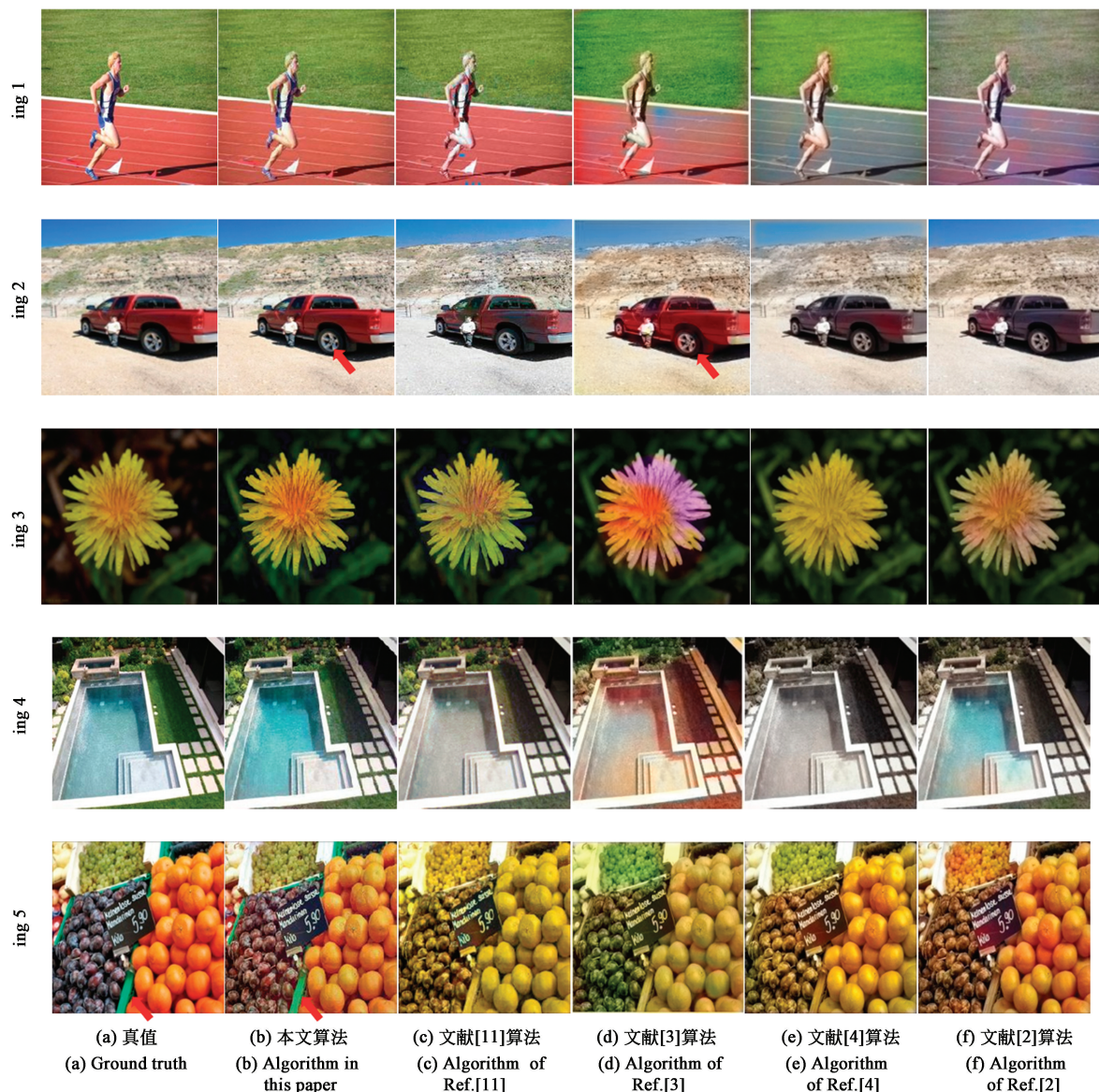


图 4 不同彩色化算法结果对比

Fig.4 Result comparison of different colorization algorithms

木也呈现较好的彩色化效果。其他算法均没有给泳池赋予恰当的颜色,泳池周围的草地等也整体呈现灰褐色。在第5行中,其他算法都忽略水果筐的颜色细节,且对水果的颜色恢复较差,而本文算法不仅还原了水果筐的颜色细节,而且图像整体颜色也更接近真实图像。

通过以上对比可知,本文算法在保持较饱和颜色信息的同时,一定程度上也能够缓解颜色溢出和细节损失等问题。

同时,为了从客观的角度证明本文算法的优越性。采用3.2介绍的3种评价指标,对多种彩色化算法结果进行定量评估,对比结果如表2所示。表中计算的结果是对于所有测试集图像的指标平均值。可以观察到,对于相同的测试图片,本文算法具有更好的评价指标,证明本文算法生成的彩色图相比于其他几种算法彩色化效果更佳。

表2 不同算法结果对比

Tab.2 Result comparison of different algorithms

算法	PSNR/dB	SSIM	IE
文献[12]	24.143	0.938	7.447
文献[2]	23.974	0.935	7.435
文献[4]	23.022	0.924	7.394
文献[3]	22.869	0.917	7.440
本文算法	24.455	0.943	7.489

以上各种算法处理 256×256 尺寸大小的图像平均所需时间如表3所示。

表3 不同算法处理时间

Tab.3 Different algorithms processing time

算法	时间/s
文献[12]	0.1
文献[2]	0.4
文献[4]	0.8
文献[3]	0.3
本文算法	0.2

3.4 消融实验

本实验采用不同的网络结构进行训练,来验证GAN结构和各模块对算法性能的影响。(1)原始U-Net+GAN;(2)U-Net+GF+GAN;

GAN的生成网络在U-Net基础上添加全局特征优化模块;(3)本文算法:U-Net+GF+Attention+GAN,生成网络继续添加通道注意力模块;(4)本文算法去除GAN:U-Net+GF+Attention,单独训练本文算法的生成网络G。前3种生成网络复杂度依次升高,可分别验证全局特征优化和通道注意力的作用,最后只训练本文算法的生成网络G,来验证GAN结构对彩色化效果的影响。

由表4可知,添加全局特征优化模块后,相比于原始的U-Net结构,各项指标均有所提升,说明全局特征与多尺度层级特征的融合有利于生成质量更高的图像。通道注意力模块的添加进一步增强算法的性能。而只训练本文算法的生成网络,评价指标均有所下降,证明GAN结构对图像彩色化有一定提升效果。具体效果如图5所示。

表4 各模块作用验证

Tab.4 Effect validation of each module

算法	PSNR/dB	SSIM	IE
U-Net+GAN	22.077	0.918	7.469
U-Net+GF+GAN	23.696	0.934	7.486
本文算法	24.455	0.943	7.489
本文算法去除GAN	22.148	0.921	7.442

原始U-Net结构的生成网络容易出现颜色溢出的现象,如图5(a)把房子的屋顶和部分墙面都赋予了天空的蓝色,且忽略了房子旁边的指示牌的颜色信息。添加全局特征优化模块也相当于引入了全局语义优化,如图5(b)屋顶和墙面的蓝色去除,图像的边界更加清晰,指示牌也还原了较浅的红色。说明全局特征优化模块一定程度上增强了算法对图像的整体信息的理解,有利于减少颜色溢出的现象。接着再添加完通道注意力模块后(图5(c)),指示牌和门等细节的颜色信息都有少许的加深。由此可知,全局特征优化模块与通道注意力模块对算法性能都有不同程度的提升效果。而在本文算法的基础上去除GAN结构训练得到的效果图(图5(d)),虽然缓解了颜色溢出的问题,但是颜色的饱和度和图像质量都不如本文算法的效果图(图5(e))。这说明GAN网络有利于生成彩色化效果更好的图像。

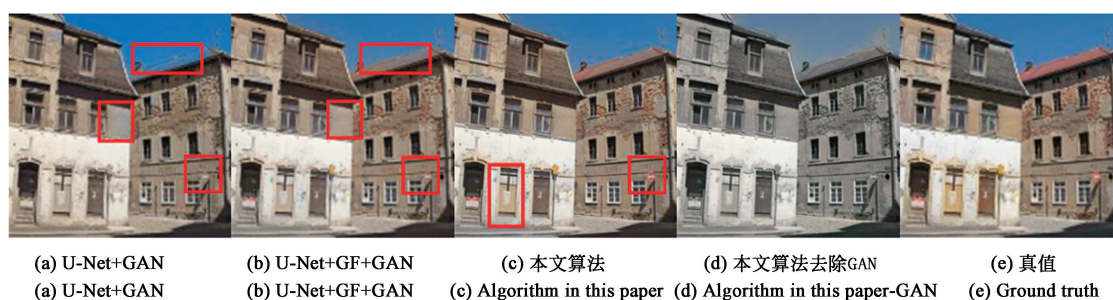


图 5 不同网络结构的结果对比

Fig.5 Result comparison of different network structures

4 结 论

本文提出的彩色化算法是基于一种对抗策略,将全局特征与多尺度层级特征融合来捕捉图像的全局语义信息,并且引入通道注意力模块,通过抑制噪声和增强与任务相关的特征通道的权重,以提高算法的性能。同时本算法的损失函数还在 WGAN-GP 的优化基础上添加颜色损失,便

于网络训练。实验结果表明,该算法在主观视觉上 and 评价指标上都取得了较好的效果。其中在 Place365 测试集上 PSNR 和 SSIM 指标分别达到 24.455 dB 和 0.943。而且通过消融实验证明了全局特征与多尺度层级特征融合模块、注意力模块和 GAN 网络结构对算法性能的提升都起到了实质性的作用。相比于以往算法,本文算法在全局语义信息理解、图像细节保持和颜色饱和度方面有较好的优势。

参 考 文 献:

- [1] CHENG Z Z, YANG Q X, SHENG B. Deep colorization [C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 415-423.
- [2] 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): 110.
- [3] ZHANG R, ISOLA P, EFROS A A. Colorful image colorization [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016: 649-666.
- [4] LARSSON G, MAIRE M, SHAKHNAPOVICH G. Learning representations for automatic colorization [C]//*Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016: 577-593.
- [5] WANG X T, YU K, WU S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks [C]//*Proceedings of European Conference on Computer Vision*. Munich: Springer, 2018: 63-79.
- [6] 余思泉,韩志,唐延东,等.基于对抗生成网络的纹理合成方法[J].*红外与激光工程*, 2018, 47(2): 0203005.
YU S Q, HAN Z, TANG Y D, et al. Texture synthesis method based on generative adversarial networks [J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203005. (in Chinese)
- [7] 杜振龙,沈海洋,宋国美,等.基于改进 CycleGAN 的图像风格迁移[J].*光学精密工程*, 2019, 27(8): 1836-1844.
DU Z L, SHEN H Y, SONG G M, et al. Image style transfer based on improved CycleGAN [J]. *Optics and Precision Engineering*, 2019, 27(8): 1836-1844. (in Chinese)
- [8] ISOLA P, ZHU J Y, ZHOU T H, et al. Image-to-image translation with conditional adversarial networks [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 5967-5976.
- [9] NAZERI K, NG E, EBRAHIMI M. Image colorization with generative adversarial networks [J]. *arXiv*: 1803.05400, 2018.

- [10] CAO Y, ZHOU Z M, ZHANG W N, *et al.* Unsupervised diverse colorization via generative adversarial networks [C]//*Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Skopje: Springer, 2017: 151-166.
- [11] VITORIA P, RAAD L, BALLESTER C. ChromaGAN: adversarial picture colorization with semantic classdistribution [C]//*Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision*. Snowmass: IEEE, 2020: 2434-2443.
- [12] ZHANG S H, LI H, KONG W H. Object counting method based on dual attention network [J]. *IET Image Processing*, 2020, 14(8): 1621-1627.
- [13] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks [C]//*Proceedings of the 34th International Conference on Machine Learning*. Sydney: PMLR, 2017: 214-223.
- [14] VITORIA P, SINTES J, BALLESTER C. Semantic image inpainting through improved wasserstein generative adversarial networks [C]//*Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. Funchal: SciTePress, 2019: 249-260.
- [15] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//*Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer, 2015: 234-241.
- [16] 张良,车进,杨琦.多粒度特征融合的行人再识别研究[J].*液晶与显示*,2020,35(6):555-563.
ZHANG L, CHE J, YANG Q. Multi-granularity feature fusion for person re-identification [J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(6): 555-563. (in Chinese)
- [17] BAKER N, LU H J, ERLIKHMAN G, *et al.* Local features and global shape information in object classification by deep convolutional neural networks [J]. *Vision Research*, 2020, 172: 46-61.
- [18] 张海涛,张梦.引入通道注意力机制的SSD目标检测算法[J].*计算机工程*,2020,46(8):264-270.
ZHANG H T, ZHANG M. SSD target detection algorithm with channel attention mechanism [J]. *Computer Engineering*, 2020, 46(8): 264-270. (in Chinese)

作者简介:



万园园(1996—),女,江西南昌人,硕士研究生,2016年于南昌大学获得学士学位,主要从事计算机视觉及机器学习方面的研究。E-mail:1123362384@qq.com



陈小林(1980—),男,吉林长春人,博士,副研究员,2010年于中国科学院长春光学精密机械与物理研究所获得博士学位,研究方向为实时视频图像处理,嵌入式处理器设计,景象模拟仿真嵌入式实时处理器设计。E-mail:cl.hong@163.com