

# 基于主题色板的图像上色方法研究

李锦峰<sup>1</sup>, 裴伟<sup>2</sup>, 朱永英<sup>3</sup>, 鲁明羽<sup>1</sup>, 宋琳<sup>1</sup>

(1.大连海事大学信息科学技术学院, 辽宁 大连 116026)

(2.大连海事大学环境科学与工程学院, 辽宁 大连 116026)

(3.大连海洋大学海洋与土木工程学院, 辽宁 大连 116023)

**[摘要]** 现有基于主题色板的图像上色方法存在着主题不准确、色彩不和谐、美感评价不客观等问题. 鉴于此, 本文提出了一套上色解决方案, 用 Lasso 回归模型对 Mask R-CNN 分割的前景目标提取主题色、WGAN\_gp 对主题色扩展、NIMA 对主题上色方案评价. 在美感评价实验中, 采用本文方案上色后 LPIPS 降低了 37.5%, NIMA 提高了 6.6%, 表明该方案可行有效.

**[关键词]** 图像上色, 图像分割, 主题色板, 目标色板

**[中图分类号]** TP391 **[文献标志码]** A **[文章编号]** 1001-4616(2022)03-0116-07

## Research on Image Coloring Method Based on Theme Palette

Li Jinfeng<sup>1</sup>, Pei Wei<sup>2</sup>, Zhu Yongying<sup>3</sup>, Lu Mingyu<sup>1</sup>, Song Lin<sup>1</sup>

(1.Information Science and Technology College, Dalian Maritime University, Dalian 116026, China)

(2.College of Environmental Sciences and Engineering, Dalian Maritime University, Dalian 116026, China)

(3.College of Ocean and Civil Engineering, Dalian Ocean University, Dalian 116023, China)

**Abstract:** There exist some problems in the image colorizing methods based on theme palette today, such as inaccurate themes, inharmonious colors and biased aesthetic evaluation. In this regard, this paper proposed a set of precise colorizing schemes, which adopts Lasso regression model to extract the theme color from the foreground object segmented by Mask R-CNN, extends the theme color by WGAN\_gp, and uses NIMA to quantify the optimal scheme. In the aesthetic evaluation experiment, LPIPS decreased by 37.5%, and NIMA increased by 6.6% after coloring, indicating that the scheme is feasible and effective.

**Key words:** image coloring, image segmentation, theme palette, target palette

中国传统文化悠远绵长, 包罗万象, 是世界文明最优秀成分之一. 千百年来, 斗转星移, 沧海桑田, 中华民族以服饰、器物、建筑等为载体, 续写、传承着中国传统文化的精髓. 但由于时空、文化变迁, 古人对服饰的色彩运用技法已散漫难寻. 在世界多元文化融合发展的大背景下, 人工智能引领古老智慧与东方审美融合的潮流, 固本清源, 助力中国传统色走向更加缤纷多彩的空间.

图像上色算法为中国传统色的传承打开了一扇门. 图像上色一般分为自动上色<sup>[1-6]</sup>和交互上色<sup>[7-9]</sup>. Zhang 等<sup>[2]</sup>将 U-net 集成到带有辅助分类器的生成对抗网络 AC-GAN 中, 实现漫画草图的上色; Yoo 等<sup>[4]</sup>提出了一种新颖的存储记忆网络 MemoPainter, 可在无类标签的条件下, 实现高质量的图像上色. 以上基于深度神经网络的自动上色方法, 在各自领域都展现出了出色的效果, 但在人和服饰并存的场景中, 上色不准确, 人的皮肤易上错颜色(见图 1(a)、(b)、(c)). 交互上色指的是在上色过程中有人为干预, 例如基于涂鸦的交互上色需用户先提供大量的涂鸦才能完成图像上色<sup>[7]</sup>. 这对于非专业用户来说较困难. 基于主题色板的图像上色是另一种可靠的交互上色, 其需要从图像中获取主题, 然后再根据用户提供的目标色板完成图像上色, 图像的最终颜色和目标色板的颜色一致. 目前有很多研究方法是基于主题

收稿日期: 2021-06-01.

基金项目: 国家自然科学基金项目(61001158、61272369、61370070)、辽宁省自然科学基金项目(2014025003)、辽宁省教育厅科学研究一般项目(L2012270)、大连市科技创新基金项目(2018J12GX043).

通讯作者: 裴伟, 博士, 副教授, 研究方向: 模式识别、人工智能. E-mail: peiwei@dlmu.edu.cn

色板来进行图像上色,例如 Chang 等<sup>[8]</sup>提出了一种改进的 K-means 算法提取图像的主题,允许用户编辑主色板来实现图像的重新上色;Tan 等<sup>[9]</sup>计算图像像素在 5D RGBXY 颜色空间中的凸包顶点,将其作为图像的主题色板,根据优化后的图层分解算法完成图像上色。

在众多的图像上色算法中,基于主题色板的图像上色较符合中华传统服饰的传承上色要求,既还原原有色彩,又符合当代人的审美。但现有的基于主题色板的图像上色,大多采用无监督的方式提取图像主题,提取的图像主题不准确、色彩不和谐、上色目标不明确,易出现张冠李戴的情况(见图 1(d))。

本文针对上述问题,提出一种将图像分割、主题色自动提取、主题色扩增及图像视觉评价有机结合的自动上色方案,实现中华传统服饰在新时代的传承上色。



图 1 图像上色结果

Fig. 1 Image coloring result

# 1 基于主题色板的自动上色模型

针对基于主题色板的图像上色方法存在着主题不准确、色彩不和谐、美感评价不客观等问题,本文提出的自动上色方案如图 2 所示,用 Lasso 回归模型对 Mask R-CNN 分割的前景目标提取主题色、WGAN\_gp 对主题色扩展、NIMA 对主题上色方案评价,得分最高的作为图像上色输出。

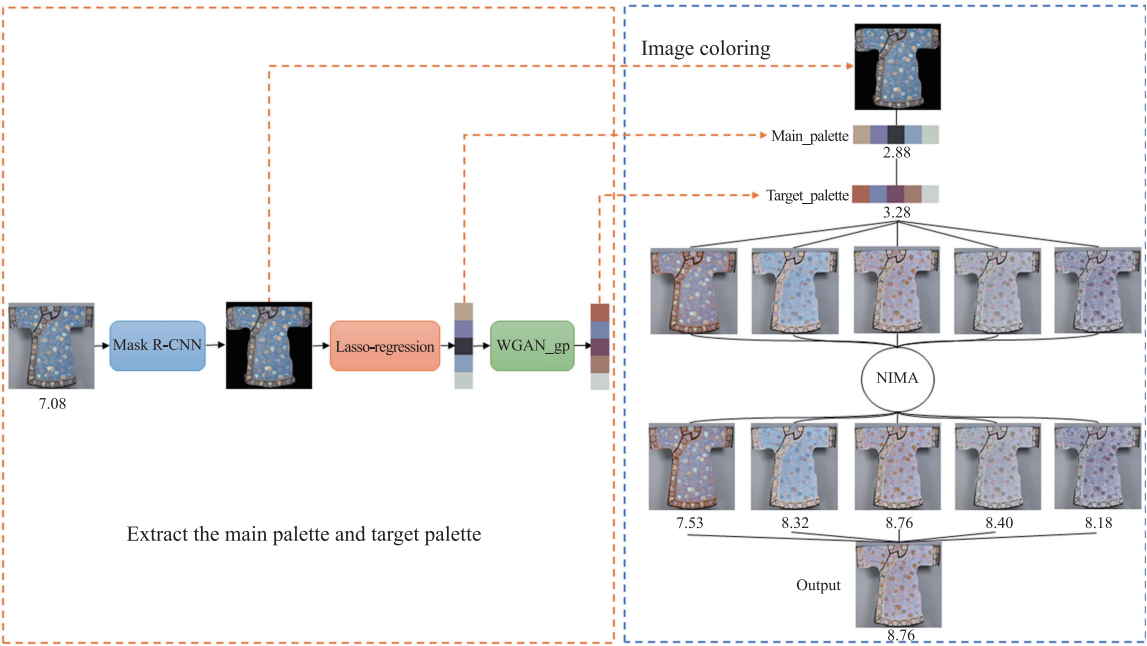


图 2 基于图像主题的图像上色模型结构

Fig. 2 Structure of image coloring model based on image theme

1.1 图像分割

随着人们审美意识的不断提高,人们对于图像上色有了更高的要求,例如单独修改图像中服装颜色,实现多重变装;或者改变环境颜色,呈现不同的情境. 现有的基于主题色板的图像上色,只能对整张图像进行上色,存在上色不准、局部难把控的问题.

Mask R-CNN<sup>[10]</sup>是通用的实例分割网络,分为 2 个分支,第 1 个分支为原始 Faster R-CNN 的结构<sup>[11]</sup>,用于对候选窗口进行分类和窗口坐标回归;第 2 个分支为全卷积网络 FCN<sup>[12]</sup>,用于对每一个实例的掩膜进行预测. 本文根据用户的需求及现实的语义环境,用 Mask R-CNN 进行图像分割,识别出图像中需要上色的区域. 识别出来的区域将作为图像主题的提色区域及图像最终的上色区域.

1.2 图像主题提取

目前,基于主题色板的图像上色方法<sup>[8-9]</sup>大多采用无监督的方式提取图像主题,在复杂场景下无监督的提色效果并不理想. 为了解决这一问题,本文采用了 2013 年 Lin 等提出的 Lasso 回归模型<sup>[13]</sup>. 其主要步骤如下:

Step 1 从候选颜色中随机选取 5 个颜色作为该图像的初始主题色板,并基于图像的显著性、像素覆盖、分割、颜色多样性、颜色纯度及聚类统计这 6 大特征对初始主题色板进行评分.

Step 2 从初始主题色板的第 1 个颜色开始使用候选颜色中的颜色替换,重复此操作,到更换初始主题色板第 1 个颜色不能再增加该初始主题色板的评分为止. 对初始主题色板的第 2、3、4、5 个颜色执行相同的操作.

Step 3 重复 Step 2 的操作,到替换初始主题色板中任意位置颜色不再增加初始主题色板的评分为止. 此时获得的主题色板即为该图像的主题.

1.3 图像评估

一直以来图像上色都缺乏适当的评估指标<sup>[14]</sup>,导致上色之后的图像不能进行精准的评估. 为此,本文引入 NIMA 模型<sup>[15]</sup>. NIMA 使用预训练的 VGG16 网络作为基准网络,在 VGG16 网络的基础上,使用随机初始化的全连接层替换掉 VGG16 网络的最后一层,作为 Softmax 激活的输入,输出图像的评价分布. 网络结构如图 3 所示.

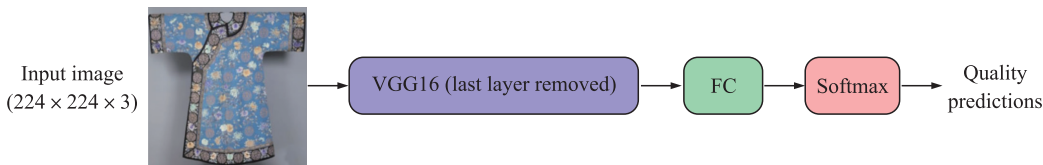


图 3 NIMA 模型网络结构  
Fig. 3 NIMA model network structure

1.4 图像主题扩展

现有基于主题色板的图像上色方法需要用户自己提供目标色板,对于非专业用户而言提供一个色彩和谐的目标色板是非常困难的.

相比于原始 GAN, WGAN\_gp<sup>[16]</sup>使用 Wasserstein 距离和梯度惩罚策略,解决了原始 GAN 中梯度消失、模式崩塌等问题,改善了网络的稳定性和图像的生成质量. 因此,本文使用 WGAN\_gp 增强 Lasso 回归模型提取的图像主题质量,生成配色更加符合视觉审美的目标色板. 基于 WGAN\_gp 设计的图像主题加强网络结构如图 4 所示,该网络由生成器和判别器构成,生成器和判别器以对抗学习的方式来训练,最终生成色彩和谐的目标色板.

1.5 图像上色

在获得图像重新上色的主题色板及目标色板之后,本文采用 2015 年 Chang 等提出的色彩转换方法<sup>[8]</sup>,实现图像中需要上色区域的上色. 该色彩转换方式将单个色彩转换分为了 2 个部分:(1)在 L 通道上的转换;(2)在 A、B 通道上的转换. 其中,在 L 通道的变化由对色板亮度线性插值得到;在 A、B 通道的变化由颜色在 LAB 空间内平移得到. 一组色彩的变换可以看作是若干个单独色彩变换的组合,具体的表现形式如式(1)所示.

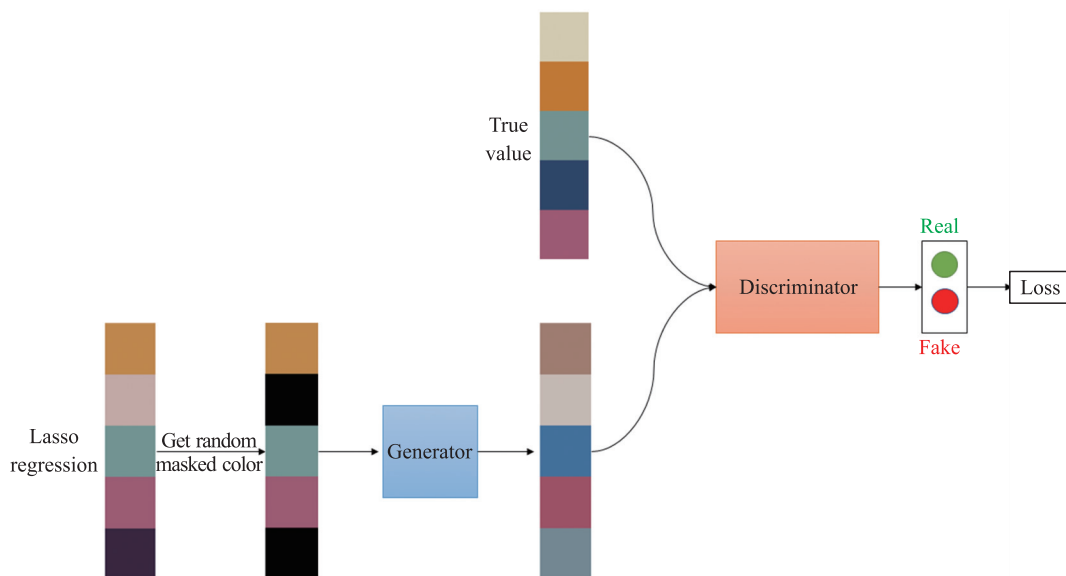


图4 基于 WGAN\_gp 的图像主题加强网络结构

Fig. 4 Image theme enhancement network structure based on WGAN\_gp

$$f(x) = \sum_i^k w_i(x) f_i(x), \quad (1)$$

式中,权重  $w_i(x)$  由径向基函数得到;  $f_i(x)$  为单个色彩变换的结果.

每个目标色板可以生成 120 张不同上色效果的图像,为了选出最优的上色图像,本文使用 NIMA 评分模型对这 120 张图像进行评分,评分最高的图像将作为图像最终的上色结果. 本文的图像上色流程如图 5 所示.

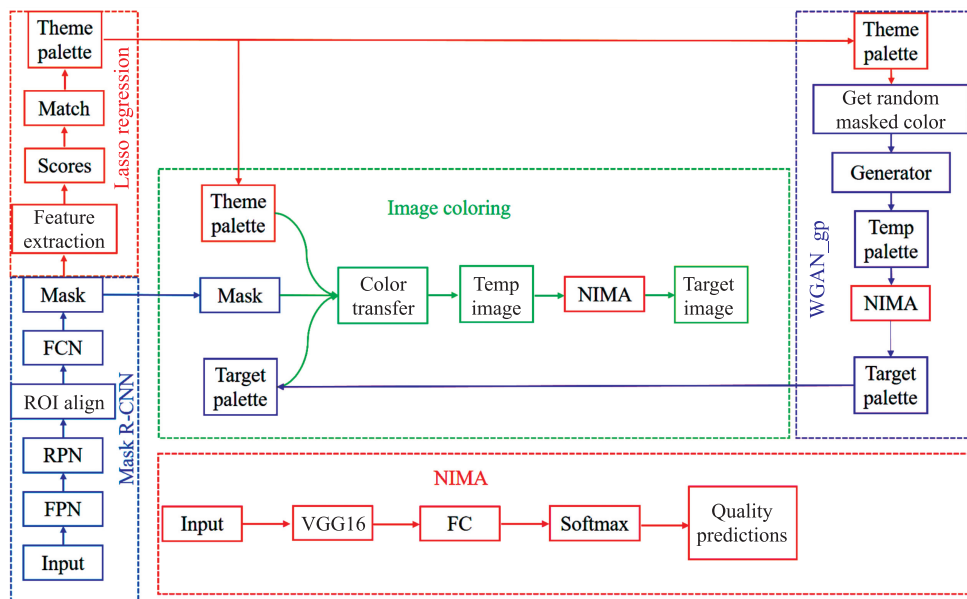


图5 图像上色流程

Fig. 5 Image coloring process

## 2 实验

### 2.1 图像上色的主题色板提取

首先运用 Mask R-CNN 模型获得分割图像,再用 Lasso 回归模型提取图像的主题. 图像分割和图像主题的提取结果如图 6 所示. 提取的结果将作为图像上色的主题色板.



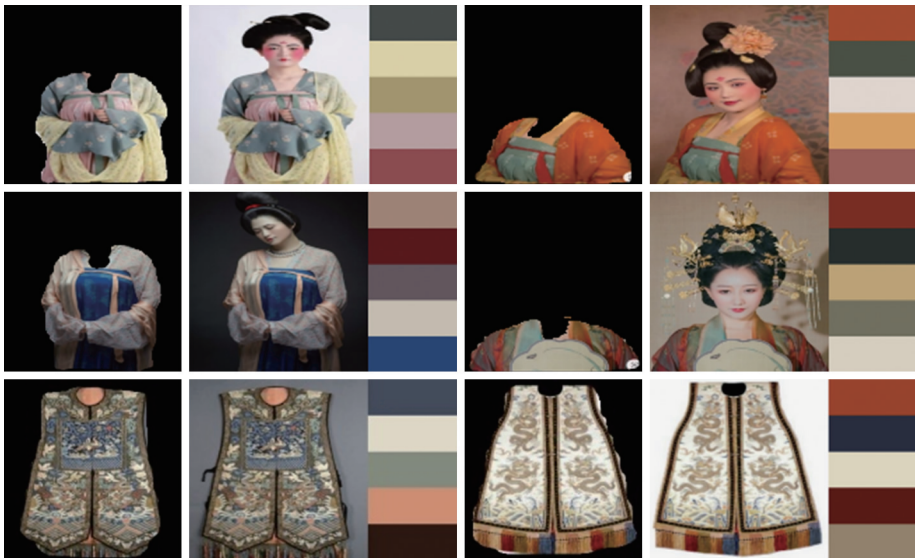


图 6 图像分割和图像主题的提取结果

Fig. 6 Image segmentation and image subject extraction results

为了验证 Lasso 回归模型的提色准确性,本文加入了 KM(K-Means)、CM(C-Means)、Rand 3 组对照实验. KM、CM 采用无监督的方式提取图像主题;Rand 采用随机生成的方式提取图像主题. 表 1 为实验结果,表中的数值为同一数据集下各个方法提取的图像主题与专家提取的图像主题之间的平均距离. 通过对比可以发现 Lasso 回归模型提取的图像主题更接近专家提取的结果.

表 1 KM、CM、Rand 和 Lasso 结果对比

Table 1 Comparison of KM, CM, Rand and Lasso results				
模型	Lasso	KM	CM	Rand
平均距离	19.93	21.20	22.04	23.74

2.2 图像上色的目标色板提取

图像的主题色板在经过 WGAN\_gp 优化后,可以获得 31 种不同的色板. 为了选择最优的色板,本文以 colourlovers 网站上收集到的 1.1 万张带有评分的五色色板为依托,重新训练了 NIMA 评分模型.

为了验证模型的准确度,以用户评分均值 2.95 作为评价的阈值,计算测试集的混淆矩阵(如表 2 所示),训练后模型的准确率为 75.9%. 图 7 为测试集的预测评分分布及真实评分分布,可以看出模型预测的评分分布和真实评分分布十分接近,表明重新训练好的 NIMA 模型可以为五色色板打一个接近人类感知水平的可靠分数.

表 2 测试集的混淆矩阵

Table 2 Confusion matrix of the test set		
Count	Positive	Negative
True	312	257
False	88	92

使用重新训练好的 NIMA 模型对 31 种不同的主题色板进行评分,最终选择评分最高的主题色板作为图像重新上色的目标色板. 目标色板的提取结果如图 8 所示. 同时本文对比了主色板和目标色板的 NIMA 评分,二者的 NIMA 评分均值分别为 2.88、3.23,相比于主色板,目标色板的 NIMA 评分均值提高了 12%,表明目标色板更能得到人们的认可.

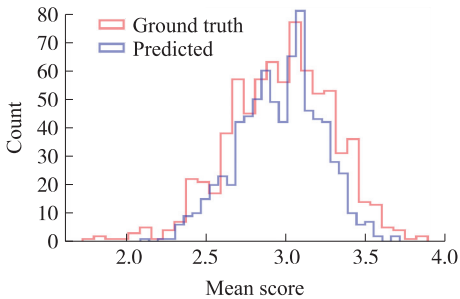


图 7 测试集的真实评分分布与预测评分分布直方图

Fig. 7 Histogram of the true score distribution and the predicted score distribution of the test set

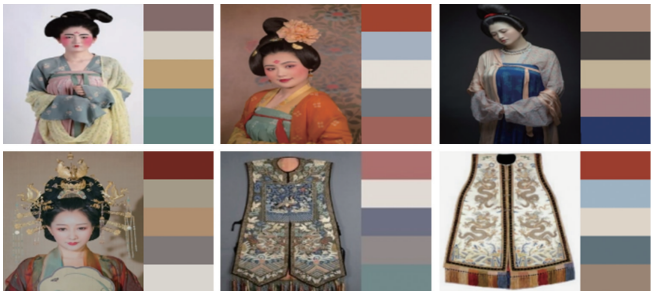


图 8 图像上色的目标色板

Fig. 8 The target palette for coloring the image

2.3 图像上色的效果对比

用本文的图像上色方案进行图像上色,实验结果如图 9、10 所示,其中图 9 为原始的图像,图 10 为重新上色之后的图像. 同时本文对比了图像上色前后的 NIMA 评分,结果显示,重新上色之后,图像 NIMA 评分均值提高了 3%,表明重新上色之后的图像更能得到人们的认可.

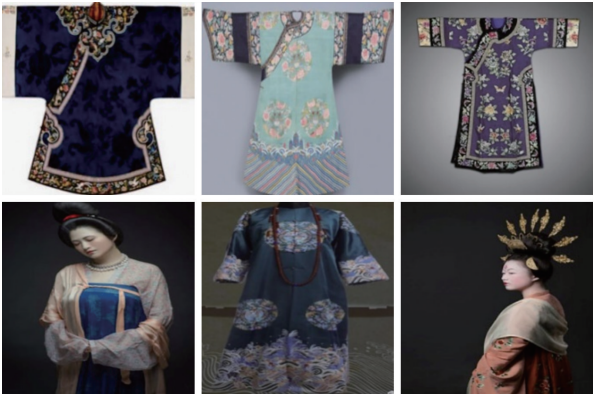


图 9 原始图像  
Fig. 9 The original image

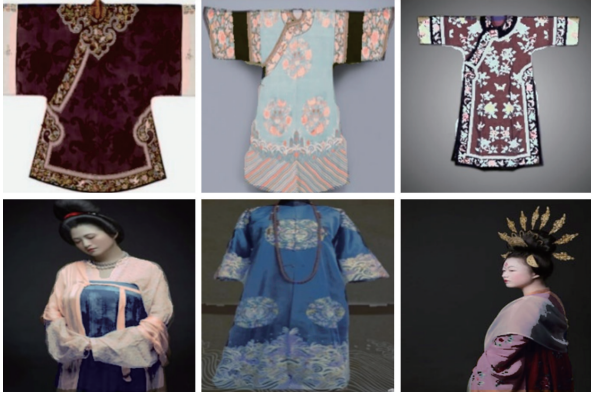


图 10 重新上色之后的图片  
Fig. 10 Image after recoloring

另外,将本文的图像上色方案与 Su 等<sup>[5]</sup>提出的基于实例感知的图像上色模型、Antic<sup>[3]</sup>提出的基于 NOGAN 的图像上色模型、Chang 等<sup>[8]</sup>提出的基于主题色板的图像上色模型进行对比. 相比于 PSNR、SSIM 等传统评估指标,LPIPS<sup>[17]</sup>和 NIMA 更符合人们的审美观,因此使用 NIMA 和 LPIPS 作为评估指标,验证本文图像上色方案的有效性. 由结果(表 3)可知,相比于性能第一的基于实例感知的图像上色模型<sup>[5]</sup>,采用本文方案上色后 LPIPS 降低了 37.5%,NIMA 提高了 6.6%,表明该方案可行有效.

表 3 4 种模型的上色效果指标比较  
Table 3 Comparison of the coloring effect indicators of the four models

模型	指标	
	LPIPS ↓	NIMA ↑
基于主题色板 <sup>[8]</sup>	0.26	7.61
基于 NOGAN <sup>[3]</sup>	0.18	8.16
基于实例感知 <sup>[5]</sup>	0.16	8.17
本文	0.10	8.71

3 结论

现有基于主题色板的图像上色方法存在着主题不准确、色彩不和谐、上色目标不明确、美感评价不客观等问题. 针对这些问题,本文以中华传统服饰上色作为切入点,以 Mask R-CNN、Lasso 回归、WGAN\_gp 和 NIMA 为传承方法,将古老的智慧与新时代东方审美相融合,实现传统服饰在新时代中的传承上色. 其中,Mask R-CNN 用来分割图像,识别出图像中需要上色的传统服饰,解决上色目标不明确的问题;Lasso 回归代替原先无监督的提色方式,使得图像主题色板的颜色代表性得到增强;WGAN\_gp 用来强化主题色板,生成色彩和谐的目标色板;NIMA 能够为传统服饰打一个接近人类感知水平的可靠分数,解决美感评价不客观的问题. 在图像上色实验中,本文模型的 LPIPS 指标和 NIMA 指标都优于其他模型,表明本方案主题提取准确、色彩和谐、评价客观,更符合人们的审美观. 但是使用 Mask R-CNN 从图像中分割出的传统服饰边缘不准确. 以后将尝试利用图像的不连续信息及方向场信息,优化 Mask R-CNN 实例分割的结果,提高图像细粒度的上色效果.

[参考文献]

[1] VITORIA P,RAAD L,BALLESTER C. ChromaGAN:adversarial picture colorization with semantic class distribution[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision(WACV),Snowmass,CO,USA,2020. Piscataway,NJ:IEEE,2020:2445-2454.

[2] ZHANG L,JI Y,LIN X,et al. Style transfer for anime sketches with enhanced residual U-net and auxiliary classifier GAN[C]// 2017 4th IAPR Asian Conference on Pattern Recognition(ACPR),Nanjing,China,2017. Piscataway,NJ:IEEE,2017:506-511.

- [3] ANTIC J. A deep learning based project for colorizing and restoring old images (and video!) [EB/OL]. [2021-05-28]. <https://github.com/jantic/DeOldify>.
- [4] YOO S, BAHNG H, CHUNG S, et al. Coloring with limited data: few-shot colorization via memory augmented networks [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019. Piscataway, NJ: IEEE, 2019: 11283-11292.
- [5] SU J, CHU H, HUANG J. Instance-aware image colorization [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020. Piscataway, NJ: IEEE, 2020: 7968-7977.
- [6] 贾玉福, 胡胜红, 刘文平, 等. 使用条件生成对抗网络的自然图像增强方法 [J]. 南京师大学报(自然科学版), 2019, 42(3): 88-95.
- [7] LEVIN A, LISCHINSKI D, WEISS Y. Colorization using optimization [C]// ACM SIGGRAPH 2004, Los Angeles, California, USA, 2004. New York, NY: ACM, 2004: 689-694.
- [8] CHANG H, FRIED O, LIU Y, et al. Palette-based photo recoloring [J]. ACM transactions on graphics, 2015, 34(4): 131-139.
- [9] TAN J, ECHEVARRIA J, GINGOLD Y. Efficient palette-based decomposition and recoloring of images via RGBXY-space geometry [J]. ACM transactions on graphics, 2018, 37(6): 1-10.
- [10] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017. Piscataway, NJ: IEEE, 2017: 2961-2969.
- [11] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. Advances in neural information processing systems, 2015, 28: 91-99.
- [12] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA, 2015. Piscataway, NJ: IEEE, 2015: 3431-3440.
- [13] LIN S, HANRAHAN P. Modeling how people extract color themes from images [C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, USA, 2013. New York, NY: ACM, 2013: 3101-3110.
- [14] ANWAR S, TAHIR M, LI C, et al. Image colorization: a survey and dataset [EB/OL]. [2020-11-03]. <https://arxiv.org/abs/2008.10774>.
- [15] TALEBI H, MILANFAR P. NIMA: neural image assessment [J]. IEEE transactions on image processing, 2018, 27(8): 3998-4011.
- [16] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein Gans [EB/OL]. [2017-12-25]. <https://arxiv.org/abs/1704.00028>.
- [17] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, USA, 2018. Piscataway, NJ: IEEE, 2018: 586-595.

[责任编辑: 丁 蓉]