



# 计算机科学

COMPUTER SCIENCE

## 结合绘画先验的线稿上色方法

窦智, 王宁, 王世杰, 王智慧, 李豪杰

引用本文

窦智, 王宁, 王世杰, 王智慧, 李豪杰. [结合绘画先验的线稿上色方法](#)[J]. 计算机科学, 2022, 49(4): 195-202.

DOU Zhi, WANG Ning, WANG Shi-jie, WANG Zhi-hui, LI Hao-jie. [Sketch Colorization Method with Drawing Prior](#)[J]. Computer Science, 2022, 49(4): 195-202.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [基于时空自适应图卷积神经网络的脑电信号情绪识别](#)

EEG Emotion Recognition Based on Spatiotemporal Self-Adaptive Graph Convolutional Neural Network  
计算机科学, 2022, 49(4): 30-36. <https://doi.org/10.11896/jsjcx.210900200>

### [大数据驱动的社会经济地位分析研究综述](#)

Big Data-driven Based Socioeconomic Status Analysis:A Survey  
计算机科学, 2022, 49(4): 80-87. <https://doi.org/10.11896/jsjcx.211100014>

### [图神经网络在 Text-to-SQL 解析中的技术研究](#)

Technical Research of Graph Neural Network for Text-to-SQL Parsing  
计算机科学, 2022, 49(4): 110-115. <https://doi.org/10.11896/jsjcx.210200173>

### [一种基于深度学习的供热策略优化方法](#)

Heating Strategy Optimization Method Based on Deep Learning  
计算机科学, 2022, 49(4): 263-268. <https://doi.org/10.11896/jsjcx.210300155>

### [GSO:基于图神经网络的深度学习计算图子图替换优化框架](#)

GSO:A GNN-based Deep Learning Computation Graph Substitutions Optimization Framework  
计算机科学, 2022, 49(3): 86-91. <https://doi.org/10.11896/jsjcx.210700199>

# 结合绘画先验的线稿上色方法

窦智 王宁 王世杰 王智慧 李豪杰

大连理工大学软件学院 辽宁 大连 116000

(931647107@qq.com)

**摘要** 自动线稿上色近年来已成为计算机视觉领域的研究热点之一。现有方法旨在通过改进网络架构或图像生成流程来提升上色的质量,但生成结果多存在色相集中、饱和度和明度分布不合理等现象。为此,提出一种结合绘画先验的线稿上色方法。该方法受插画师上色过程的启发,学习其广泛使用的绘画先验知识(如色相变化、饱和度对比和明暗对比)来提升自动线稿上色的质量。具体来讲,该方法在 HSV 色彩空间上增加了像素级损失,引导网络生成异常纹理较少的结果。同时,提出的三项启发式损失函数分别引入了色相变化、饱和度对比和明暗对比等绘画先验,引导网络生成具有合理色彩组成的上色结果。在真实线稿构建的测试数据集上,采用 FID 和 MOS 两项指标对所提方法和现有方法从生成结果与真实数据的分布相似度及视觉质量方面进行了比较。实验结果表明,相比性能第二的模型,所提方法的 FID 指标降低了 21.00, MOS 指标提高了 0.96,因此所提线稿上色方法有效提升了自动线稿上色的视觉质量。

**关键词:** 自动线稿上色;深度学习;对抗生成网络(GAN);绘画先验;HSV 色彩空间

中图法分类号 TP391

## Sketch Colorization Method with Drawing Prior

DOU Zhi, WANG Ning, WANG Shi-jie, WANG Zhi-hui and LI Hao-jie

College of Software Technology, Dalian University of Technology, Dalian, Liaoning 116000, China

**Abstract** Automatic sketch colorization has become an important research topic in computer vision. Previous methods intent to improve the colorization quality with advanced network architecture or innovative pipeline. However, they usually generate results with concentrated hue, unreasonable saturation and gray distribution. To alleviate these problems, this paper proposes a sketch colorization method with drawing priors. Inspired by the actual coloring process, this method learns the widely used drawing priors (such as hue variation, saturation contrast, and gray contrast) to improve the quality of automatic sketch colorization. Specifically, it incorporates pixel-level loss in the HSV color space to gain more natural results with less artifacts. Meanwhile, three heuristic loss functions that introduce the drawing priors such as hue variation, saturation and gray contrast are used to train our method to generate results with harmonious color composition. We compare our method with current state-of-the-art methods on test dataset constructed by real sketch images. Fréchet inception distance (FID) and mean opinion score (MOS) are adopted to measure the similarity between the distribution of real and generated images and the visual quality, respectively. Compared to the second-best method, the experimental results show that the FID of our method decreases by 21.00 and the MOS increases by 0.96, respectively. All the experimental results prove that the proposed method effectively improves the visual quality of automatic sketch colorization.

**Keywords** Automatic sketch colorization, Deep learning, Generative adversarial networks (GAN), Drawing prior, HSV color space

## 1 引言

动漫图像因其夸张的色彩对比和鲜明的人物设计,受到广大观众的喜爱。而在动漫图像的绘制过程中,线稿上色是其中最为重要的步骤之一,但其工序复杂且存在大量冗余劳动。因此,如何有效减少动漫图像上色过程中的重复性工作

是亟待解决的问题。自动线稿上色方法通过让用户提供或不提供少量的色彩提示,来实现自动为线稿生成相应的上色结果,为动漫线稿上色提供了一种快速且灵活的交互方式,能够极大地降低动漫图像绘制的成本与周期。

在自动线稿上色领域,传统的交互式上色方法<sup>[1-2]</sup>多基于纹理定位和区域分割的方式进行色彩区域填充,但因其无法

到稿日期:2021-03-12 返修日期:2021-07-03

基金项目:国家自然科学基金(61772108,61932020,61976038)

This work was supported by the National Natural Science Foundation of China(61772108,61932020,61976038).

通信作者:李豪杰(hjli@dlut.edu.cn)

自动生成阴影,画面缺少立体感,纹理细节不够丰富。近年来,深度学习技术特别是生成对抗网络<sup>[3]</sup>(Generative adversarial network, GAN)在图像生成领域包括自动线稿上色方面获得广泛的研究,并取得了重要进展。Taizan<sup>[4-6]</sup>提出了用户易于操作的 PaintsChainer 系列,针对特定风格的线稿,此类方法可得到不错的上色结果,但多数情况下其生成结果仍包含颜色溢出及异常纹理的问题。针对此问题,Style2Paints<sup>[7]</sup>利用双阶段生成流程对其进行优化;Ci 等<sup>[8]</sup>提出利用局部特征来增强网络泛化性能;MemoPainter<sup>[9]</sup>则使用内存扩展网络检索已有的数据特征,辅助线稿生成上色结果。相比传统方法,以上方法的上色质量大幅提升,但在明暗对比度、色彩丰富度等画面色彩组成上还可进一步优化。

本文认为,现有方法产生的结果色彩组成较差的一个关键原因是这些方法仅专注于在 RGB 色彩空间(红、绿、蓝)进行网络约束和学习,而忽略了在实际上色过程中插画师应用的重要绘画先验知识。插画师通常将绘画先验在 HSV 色彩空间中进行表达,原因是 HSV 色彩空间是根据色彩直观特性构建的颜色模型,符合人眼对颜色的认知。如文献<sup>[10]</sup>所述,该空间包含色相(H)、饱和度(S)和明度(V)3个通道,而常用的绘画先验也与这3个通道属性相对应,因此,插画师更容易在 HSV 色彩空间中进行颜色的选取,并在画面中表达绘画先验。而利用加法叠加表示颜色的 RGB 色彩空间不如 HSV 色彩空间直观,难以选取颜色表达绘画先验。

受此启发,本文在 HSV 色彩空间设计了启发式损失函数来对绘画先验建模,引导网络学习插画师常用的绘画先验,如色相变化、饱和度对比及明暗对比<sup>[11-13]</sup>。本文希望通过引入色相变化先验得到色彩丰富且和谐的图像;引入饱和度对比先验以确保图像的焦点突出;引入明暗对比先验以保证画面的层次分明。

本文主要贡献可总结为:1)受插画师实际上色过程的启发,首次利用 HSV 色彩空间作为辅助信息,对线稿上色方法进行监督训练,有效缓解了现有方法上色结果中色彩组成不合理的问题,量化及可视化实验结果表明所提方法相比现有方法能得到视觉质量更佳(高 MOS)且接近真实插画图像(低 FID)的上色结果。2)提出了像素级的色彩空间损失函数和启发式的损失函数,从 HSV 色彩空间对生成器进行监督,以此来学习绘画先验,能够得到具有合理色彩组成的上色结果。

## 2 相关工作

### 2.1 绘画先验

在实际创作过程中,画面的和谐感常受多方面因素的影响。为达到和谐美观的效果,插画师们会利用他们对画面的理解,即绘画先验进行创作,帮助观众们区分画面中的层次关系,引导观众关注焦点区域。在上色中常用的绘画先验有色相变化、饱和度对比及明暗对比等。如 James<sup>[13]</sup>所指出,色相变化可营造不同的气氛,增加图像趣味性,如图 1(a)为经典的冷暖色相对比的示例图像,以暖色为前景、冷色为背景,可营造出和谐的氛围。饱和度对比可突出画面的焦点,如图 1(b)中的消防栓和广告牌添加饱和度较高的黄色后就容易吸引观众的注意力。明暗对比同样具有保证画面和谐统一的

效果,如图 1(c)所示,暗色置于前景,亮色置于背景,可将图像的层次区分出来,保证画面的和谐感。由此可以看出,绘画先验广泛蕴含于画面的色彩组成中,且与 HSV 色彩空间的属性对应。因此,本文尝试将绘画先验引入到自动线稿上色任务,通过在 HSV 色彩空间设计合适的损失函数来引导网络学习上述常见的绘画先验,从而提升自动线稿上色的质量。



图 1 绘画先验解释<sup>[13]</sup>(电子版为彩色)

Fig. 1 Explanation of drawing priors<sup>[13]</sup>

### 2.2 非深度学习的上色方法

在自动上色方法研究领域初期,交互式上色方法<sup>[14-15]</sup>基于灰度图中具有相似灰度值的邻近像素应有相似颜色等假设进行颜色传播填充。但此类方法多针对真实黑白照片上色,不适用于抽象且缺乏灰度信息的线稿。此后 Qu 等<sup>[1]</sup>提出针对不同纹理制作特定的模板,并将用户选取的颜色传播到与线稿中纹理相似的区域的方法。虽然此方法将自动上色扩展到了线稿领域,但含有重复纹理的线稿在实际应用中十分少见,因此该方法的局限性很大。此后 LazyBrush<sup>[2]</sup>采用陷阱球(trapped ball)的方式分割出线稿的不同区域,然后应用用户指定的颜色对不同区域进行色彩填充。此方法最大的问题在于,每块线稿区域仅有单一的色彩,无法生成阴影及特有的动漫图像纹理,难以达到较好的视觉效果。

### 2.3 基于深度学习的上色方法

近年来,基于深度学习的方法已深入应用于图像生成领域。GAN<sup>[3]</sup>的提出促使配对图像转换任务<sup>[16-18]</sup>和非配对图像转换任务<sup>[19-20]</sup>产生了很多高质量的研究成果,这也为研发高性能的自动线稿上色方法提供了很大的可能性。Furusawa 等<sup>[21]</sup>提出了提取参考图像色块特征进行上色的方法。Frans<sup>[22]</sup>通过串联颜色和阴影生成网络实现了可交互的线稿上色方法。Auto-painter<sup>[23]</sup>使用条件生成对抗网络<sup>[24]</sup>引入对抗损失,提升了上色质量。Tag2Pix<sup>[25]</sup>则将线稿和色彩标签作为输入,引入特征图通道选择机制进行上色。然而上述方法常会导致异常纹理和颜色错误的问题。

商用系统 PaintsChainer 系列<sup>[4-6]</sup>提供了方便用户交互的自动线稿上色方法,它读取线稿输入并根据用户提供的少量

色彩信息自动生成上色结果。尽管此方法可得到具有动漫纹理和颜色渐变的上色结果,但 Tanpopo<sup>[5]</sup>存在颜色扩散及生成颜色饱和度过低的问题;Satsuki<sup>[6]</sup>会出现色彩溢出和生成异常纹理的问题;Canna<sup>[4]</sup>虽然提高了生成结果的清晰度,但线条和颜色会有失真的情况。相比 PaintsChainer 系列,Ci 等<sup>[8]</sup>提出了一种用户可交互的线稿上色方法,该方法利用局部特征提升了网络的泛化性能,可更准确地捕获真实彩色动漫图像的分布,并以此生成具有合理阴影及纹理的上色结果。

尽管上述基于深度学习的方法相比非深度学习方法能产生具有动漫图像纹理和颜色渐变的上色结果,但仍存在饱和度不足、明暗对比不合理及色相分布集中的问题。而本文提出的方法使用启发式的损失函数,考虑了实际上色过程中绘画先验的重要性,能够得到视觉效果更佳的上色结果。

### 3 本文方法

本文方法的目的是通过学习绘画先验来提升自动线稿

上色的质量。为此,本文参考 Ci 等<sup>[8]</sup>的方法构建了算法的网络,并在此基础上,首次尝试通过 HSV 色彩空间约束和三项启发式损失函数学习绘画先验,引导网络训练。

#### 3.1 方法流程

图 2 给出了本文方法的整体流程图,算法中生成器和判别器的网络架构参考 Ci 等<sup>[8]</sup>的方法进行构建,Ci 等<sup>[8]</sup>的方法是条件 WGAN-GP 框架<sup>[26]</sup>的算法,包含生成器  $G$  和判别器  $D$ 。生成器  $G$  学习从线稿到上色结果(RGB 色彩空间)的映射,判别器  $D$  则对输入的图像进行分类,判断输入图像是真实的还是生成的。通过对抗生成的方式训练  $G$ ,可以使其生成愈发接近真实的上色结果。生成器  $G$  以随机色彩提示  $H$ 、线稿  $X$  和相应的局部特征  $F(X)$  作为输入,输出彩色图像  $Y_g$ 。判别器  $D$  计算  $F(X)-Y_g$  图像对和  $F(X)-Y$  图像对的对抗损失  $L_{adv}$ 。特征提取器  $V$  用来获得  $Y_g$  和  $Y$  之间的视觉感知损失  $L_{featXC}$ ,以减少生成图像与真实图像在特征层面的重构误差。

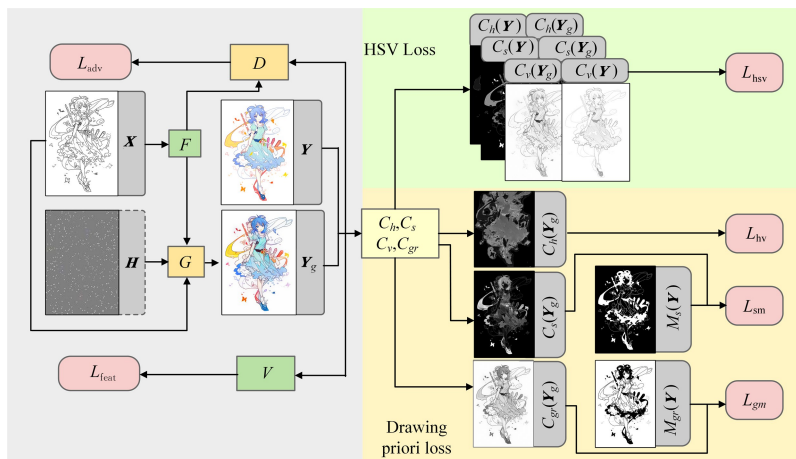


图 2 本文方法的整体流程

Fig. 2 Overview of our method

在 Ci 等<sup>[8]</sup>方法的基础上,本文方法利用转换函数  $C_h$ 、 $C_s$  和  $C_v$  将生成器的输出结果  $Y_g$  和真实训练彩图  $Y$  从 RGB 图像转换为 H,S,V 三通道图像,并在此基础上引入 L1 形式的 HSV 损失  $L_{hsv}$ 。通过这种在 HSV 色彩空间上加入直接约束的方式,可鼓励生成器得到稳定的、具有较少异常纹理的生成结果,同时为后续引入绘画先验损失奠定了基础。

本文方法在生成结果的 H 通道上计算色相变化损失  $L_{hv}$  以引入色相变化先验,使得生成结果  $Y_g$  的色彩变得丰富。对于饱和度对比先验  $L_{sm}$ ,本文使用生成结果  $Y_g$  的 S 通道,根据由真实图像  $Y$  获得的饱和度掩模  $M_s(Y)$  划分出的高低饱和区域来增大区域间差值,以此学习饱和度对比先验。对于明暗对比损失  $L_{gm}$ ,采用 RGB 到灰度图的转换函数  $C_{gr}$  来获取  $Y$  与  $Y_g$  的灰度图,并按照与  $L_{sm}$  相似的方式,根据  $Y_g$  的灰度图和对应的灰度掩模  $M_{gr}(Y)$  计算明暗对比损失,引导生成器学习明暗对比先验。值得注意的是,本文并未直接在 V 通道中计算明暗对比损失,原因是图像的明暗信息不完全包含在 V 通道中,它还受颜色的色相和饱和度的影响,如图 3 所示,在明度值 V 设置不变的情况下,色相及饱和度的变化均会影响颜色的明暗程度。

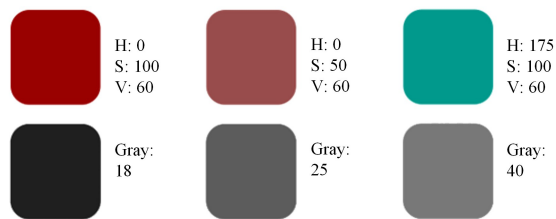


图 3 V 通道与灰度值的关系

Fig. 3 Relationship of V channel and gray value

#### 3.2 网络结构

本文方法中的生成器和判别器的网络架构参考 Ci 等<sup>[8]</sup>的方法进行构建。该方法中生成器  $G$  为 U-Net 结构<sup>[16]</sup>,如图 4 所示, $G$  将灰度线稿图像  $X$  和随机颜色提示  $H$  作为输入,输出生成的彩色图像  $Y_g$ 。其中随机颜色提示  $H$  根据 4 倍下采样的真实彩色图像  $Y$ ,采用文献[8]提出的方式获取。生成器  $G$  堆叠了由 ResNeXt 残差块<sup>[27]</sup>构建的子网络来增强网络的拟合能力,图 4 中的  $b$  表示此子网络中堆叠的残差块个数, $c$  表示各卷积层输出的特征图通道数。在解码器部分,该生成器采用 Yu 等<sup>[28]</sup>的构建原则增加膨胀卷积,并使用 Shi 等<sup>[29]</sup>提出的 pixel-shuffle 卷积层,预防普通反卷积出现的棋盘格

伪影。为缓解因使用生成数据训练网络而产生的过拟合问题,生成器  $G$  采用局部特征  $F(\mathbf{X})$  作为条件输入,此

局部特征为预训练的 Illustrator2Vec<sup>[30]</sup> 网络提取的第六层卷积特征。

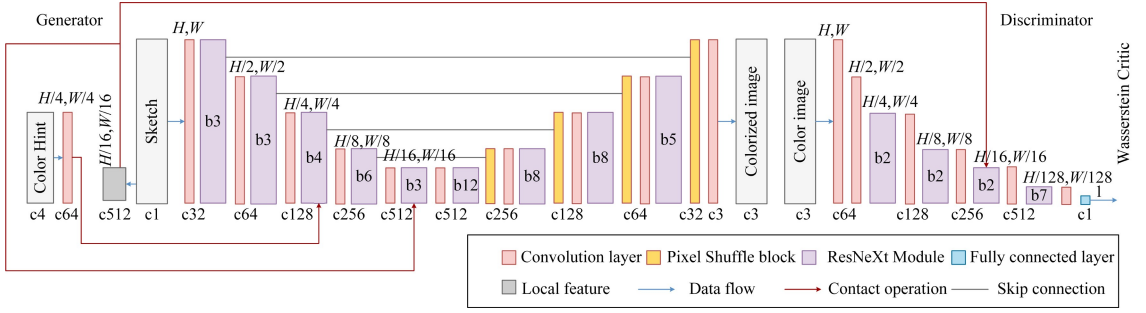


图4 生成器与判别器的网络结构

Fig. 4 Network architectures of generator and discriminator

判别器  $D$  用来分类真实图像和生成图像,其架构按照 SRGAN<sup>[31]</sup> 的配置进行构建,并在此基础上堆叠  $G$  中应用的 ResNeXt 子网络,帮助其处理高分辨率彩色图像。与  $G$  相同,  $D$  也将  $F$  提取的局部特征  $F(\mathbf{X})$  作为条件输入,与  $G$  组成条件生成对抗网络<sup>[24]</sup>。在整体的网络架构中,除  $G$  的最后一层卷积层采用  $\tanh$  激活函数外,其余卷积层均使用 LeakyReLU 激活函数,斜率设置为 0.2。

### 3.3 损失函数

如 3.1 节所述,本文方法用到了对抗损失  $L_{adv}$  和视觉感知损失  $L_{feat}$ ,同时本文还提出了 HSV 色彩空间约束  $L_{hsv}$  以及三项用来引入绘画先验的启发式损失函数  $L_{hv}$ ,  $L_{sm}$ ,  $L_{gm}$ ,这三项损失总体由  $L_{dp}$  表示。因此,本文方法使用的完整的损失函数  $L_{total}$  如式(1)所示:

$$L_{total} = L_{adv} + L_{hsv} + L_{dp} + L_{feat} \quad (1)$$

#### 3.3.1 对抗损失函数

本文使用的对抗损失函数与 Ci 等<sup>[8]</sup> 的方法一致,是以局部特征  $F(\mathbf{X})$  作为条件输入的 WGAN-GP 形式的对抗损失,其鼓励生成器  $G$  输出真假难辨的上色结果,具体公式如下:

$$L_{adv} = L_{ad} + L_{gp} + L_{drift} \quad (2)$$

其共由 3 项构成,分别为对抗项、梯度惩罚项,以及为了保证训练稳定性而提出的梯度偏移损失项<sup>[32]</sup>,其公式分别如式(3)一式(5)所示:

$$\min_G \max_D L_{ad} = E_{Y \sim P_{rgb}} [D(Y, F(\mathbf{X}))] - E_{Y_g \sim P_g} [D(Y_g, F(\mathbf{X}))] \quad (3)$$

$$L_{gp} = \lambda_{gp} E_{Y \sim P_i} [\| \nabla_{\hat{Y}} D(Y, F(\mathbf{X})) \|_2 - 1]^2 \quad (4)$$

$$L_{drift} = \lambda_d E_{Y \sim P_{rgb}} [D(Y, F(\mathbf{X}))]^2 \quad (5)$$

其中,  $\mathbf{Y}_g$  表示由  $G$  获得的生成图像,它是  $G(\mathbf{X}, F(\mathbf{X}), \mathbf{H})$  的简要表示形式。  $\mathbf{Y}$  则是用于训练的 RGB 色彩空间中的真实彩色图像。同时,式(4)中计算梯度惩罚项用到的分布  $P_i$  为生成数据分布  $P_g$  和真实数据分布  $P_{rgb}$  之间插值点的分布。式(4)中  $\lambda_{gp}$  取值为 10,式(5)中  $\lambda_{drift}$  取值为  $1 \times 10^{-3}$ 。

#### 3.3.2 HSV 损失函数

如前文所述,本文方法拟学习线稿上色过程中常用的绘画先验知识,为此需要采用插画师常用的 HSV 色彩空间来约束生成的图像。通过在 HSV 色彩空间加入直接的像素级

约束可驱动生成器输出稳定的、具有更少异常纹理的上色结果,同时也为后续引入绘画先验损失奠定了基础。HSV 损失定义为:

$$L_{hsv} = \sum_{\theta \in (h,s,v)} E_{Y \sim P_{rgb}} [\| C_\theta(Y_g) - C_\theta(Y) \|_1] \quad (6)$$

其中,  $C_\theta$  是 RGB 色彩空间到 HSV 色彩空间的转换函数。  $\theta$  是选择 HSV 色彩空间通道的变量,具体可选择  $h, s, v$  这三个值。

#### 3.3.3 绘画先验损失函数

为保证生成的图像能学习到绘画先验知识,如丰富的色相变化、合理的饱和度及明暗对比等,本文设计了应用于线稿上色任务的损失函数——绘画先验损失函数  $L_{dp}$ ,其公式如式(7)所示,共包含色相变化损失  $L_{hv}$ 、饱和度对比损失  $L_{sm}$ 、明暗对比损失  $L_{gm}$  这三项损失函数。其中  $\lambda_{hv}, \lambda_{sm}, \lambda_{gm}$  分别为 3 项损失的权重超参,经参数调整及对训练过程的观察,本文最终选择  $\lambda_{hv} = 0.2, \lambda_{sm} = 0.2, \lambda_{gm} = 0.3$  以获取最稳定的生成结果。

$$L_{dp} = \lambda_{hv} L_{hv} + \lambda_{sm} L_{sm} + \lambda_{gm} L_{gm} \quad (7)$$

色相在 HSV 色彩空间中是用于描述颜色类别的属性。如 2.1 节所述,色相间的差异能营造不同画面气氛,增加图像趣味性。因此,图像中的色相应呈现分散式分布来增强画面色彩的丰富程度,故色相变化损失  $L_{hv}$  采用增大图像 H 通道方差的方式进行计算,如式(8)所示:

$$L_{hv} = -\frac{1}{m} \sum_{i=1}^m ((C_h(\mathbf{Y}_g))(i)) - \frac{1}{m} \sum_{i=1}^m ((C_h(\mathbf{Y}_g))(i))^2 \quad (8)$$

其中,  $(C_h(\mathbf{Y}_g))(i)$  是  $C_h(\mathbf{Y}_g)$  上的第  $i$  个像素,  $m$  是图像高度和宽度的乘积。

本文旨在使使用  $L_{dp}$  中的饱和度对比损失项  $L_{sm}$  引导方法学习饱和度对比先验。当饱和度对比在图像中形成后,画面中的焦点会突出出来,整个画面将变得和谐,  $L_{sm}$  的公式定义为:

$$L_{sm} = \frac{1}{m_1} \sum_{M_i(\mathbf{Y})(j,k)=1} \max(0, df(C_s(\mathbf{Y}_g), \alpha_1)(j,k)) - \frac{1}{m_2} \sum_{M_i(\mathbf{Y})(j,k)=0} \min(0, df(C_s(\mathbf{Y}_g), \alpha_2)(j,k)) \quad (9)$$

其中,  $M_i(\mathbf{Y})$  是由实彩色图像  $\mathbf{Y}$  的饱和度通道通过 Otsu 算法<sup>[33]</sup> 获取的二进制掩模,饱和度及灰度掩膜的示例如图 5 所示。

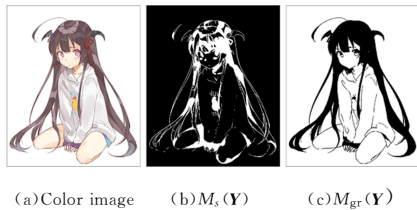


图 5 彩色图像及其饱和度和灰度掩模

Fig. 5 Real color image and its corresponding saturation and gray mask

$M_s(\mathbf{Y})$ 可划分出 $\mathbf{Y}$ 中高饱和度及低饱和区域,从而引导网络学习正确的饱和度拉伸位置,其中白色和黑色像素的总数分别由 $m_1$ 和 $m_2$ 表示。 $(j,k)$ 为选取图像中坐标编号为 $j,k$ 的像素点的操作。获取到 $M_s(\mathbf{Y})$ 掩模之后, $L_{sm}$ 可以根据它来增大生成图像的高低饱和区域与均值间的差值,以此引导网络学习饱和度对比。式(9)中用于计算图像各像素点与均值间差值图的函数 $df$ 定义为:

$$df(Z, \alpha) = \frac{1}{m} \sum_{i=1}^m Z(i) + \alpha - Z \quad (10)$$

其中, $Z$ 为输入的图像, $i$ 为输入图像中的每个像素点的编号, $m$ 为图像中全部像素点的个数, $\alpha$ 为控制图像对比度差距的偏置。对于所有实验,本文设置 $\alpha_1$ 为0.06, $\alpha_2$ 为一0.06。当由式(10)获得各像素点与均值间的差异图后,在式(9)中, $L_{sm}$ 将根据差异图采用比较函数 $max$ 及 $min$ 选取出高饱和度及低饱和区域中存在损失的像素点,并降低这些点的损失,以学习图像的饱和度对比。

为引入明暗对比的先验知识,本文采用与 $L_{sm}$ 类似的方法计算 $L_{gm}$ 。如3.1节所述,明暗对比先验在灰度信息层面引入,而非在HSV的V通道中引入,原因为HSV色彩空间的V通道无法完整地表示图像的灰度信息,因此在V通道中应用明暗对比约束将存在误差。 $L_{gm}$ 的定义为:

$$L_{gm} = \frac{1}{m_3} \sum_{M_{gr}(\mathbf{Y})(j,k)=1} \max(0, df(C_{gr}(Y_g), \alpha_1)(j,k)) - \frac{1}{m_4} \sum_{M_{gr}(\mathbf{Y})(j,k)=0} \min(0, df(C_{gr}(Y_g), \alpha_2)(j,k)) \quad (11)$$

其中, $C_{gr}$ 是从RGB图像到灰度图像的转换函数。 $M_{gr}(\mathbf{Y})$ 是与 $M_s(\mathbf{Y})$ 类似的二进制掩模,用来区分图像的高明度和低明度区域,通过Otsu算法<sup>[33]</sup>从真实彩图 $\mathbf{Y}$ 的灰度图中产生; $m_3$ 与 $m_4$ 为 $M_{gr}(\mathbf{Y})$ 中白色和黑色像素的数量。 $L_{gm}$ 可辅助生成的图像具有合理的明暗分布,从而提升生成图像的视觉效果。

### 3.3.4 视觉感知损失函数

本文同样应用了Justin等<sup>[34]</sup>提出的视觉感知损失函数,来减少生成器 $G$ 的输出结果 $\mathbf{Y}_g$ 和真实彩色图像 $\mathbf{Y}$ 之间的特征重构误差,其计算方法如式(12)所示:

$$L_{feat} = \frac{1}{chw} \|V(\mathbf{Y}_g) - V(\mathbf{Y})\|_2^2 \quad (12)$$

其中,函数 $V$ 表示预训练的VGG16网络的第四层卷积层的输出。使用 $L_{feat}$ 可以让网络的生成结果在特征层面上更接近真实的彩色图像。

## 4 实验

### 4.1 数据集

本文构建了3组数据集来训练和验证所提出的方法,其中训练数据集包含10950张网络上公开的彩色动漫图像,

线稿提取数据集包含596组真实线稿和对应彩色图像的图像对,测试数据集包含1279组真实的线稿。本文首先使用线稿提取数据集训练了U-Net结构<sup>[16]</sup>的全卷积网络作为线稿提取器,然后应用此线稿提取器和XDoG算法<sup>[35]</sup>来对训练数据集中的彩色图像生成线稿,制作生成线稿与真实彩色图像的训练对。XDoG算法的参数设置为: $\varphi=1 \times 10^{20}$ , $\tau=0.999997$ , $\kappa=1.0001$ , $\sigma$ 选择0.6,0.7,0.8,以此模拟不同插画师在绘制线稿时的笔触宽度。对于XDoG线稿,本文应用了随机灰度调整公式 $\mathbf{X}=1-\lambda(1-\mathbf{X})$ 来获取灰度线稿,这里 $\lambda$ 为0~1范围内的随机小数。本文对于线稿提取器得到的灰度线稿不使用此后处理操作,生成的线稿示例如图6所示。



图 6 训练集生成线稿示例

Fig. 6 Example of synthetic sketch of our training dataset

此后,本文将所有训练图像对的短边调整为512像素,再随机裁剪出 $512 \times 512$ 的图像块,并对其进行随机翻转,作为最终的训练数据。因单张彩色图像共产生4种线稿,所以在实际训练时,将从4种线稿中随机选择一种与真实彩图配对,然后进行训练。

除生成线稿数据外,本文还需为训练数据集中的彩色图像准备饱和度及明暗对比掩模数据,用于方法的训练。具体操作为:在这些彩色图像的饱和度通道及灰度图像上采用Otsu算法<sup>[33]</sup>进行二值化,并将此结果作为饱和度及明暗对比掩模数据。在训练过程中,采用与线稿相同的变换裁剪方式来保证掩模数据与彩色及线稿图像的像素级对应。

### 4.2 训练流程

本文方法采用PyTorch框架实现,所有网络的训练和实验测试都是在NVIDIA Titan Xp GPU上完成。由于方法的最终损失函数有较多的约束项,采用完整的损失函数直接从头训练将难以收敛。为解决此问题,本文将训练分为初始化阶段和标准阶段。在初始化阶段,使用 $L_{adv}$ 和 $L_{feat}$ 训练网络,迭代次数为 $1.2 \times 10^5$ ;再在标准阶段,采用完整的损失函数利用绘画先验引导网络进行训练,迭代次数为 $1.2 \times 10^5$ 。网络使用Adam优化器进行训练,学习率设置为0.0001。Adam优化器的动量参数为 $\beta_1=0.5$ , $\beta_2=0.9$ ,训练时的批量大小为4,生成器 $G$ 和判别器 $D$ 在整个训练过程中交替更新。在测试阶段中,只需为训练好的生成器提供线稿图像及可选的颜色提示图便可完成线稿的自动上色。

### 4.3 消融实验

本节通过消融实验,分析了本文方法中损失函数的作用,实验结果如图7所示。可以看出,所有损失项都是协助本文方法提升自动线稿上色效果的必要因素。其中HSV损失 $L_{hsv}$ 对于生成稳定结果至关重要。由于HSV色彩空间是引入绘画先验的基础,因此,在没有对HSV色彩空间进行直接约束的情况下训练方法,本文提出的其他绘画先验损失函数会导致方法生成的结果严重偏离真实彩色图像,如图7(c)所示,其生成的

结果将模糊不清,具有更多异常纹理。使用色相变化损失  $L_{hv}$  会鼓励生成器得到色彩丰富的图像。如图 7(d)所示,如去除  $L_{hv}$ ,生成结果的色相易集中分布,出现整体偏红的效果,这证明提出的色相变化损失有助于生成结果具有更加丰富的颜色。如图 7(e)所示,  $L_{sm}$  可让图像中的焦点区域具有适当的饱和度,

使画面更加吸引观众。而去除  $L_{sm}$ 后,结果中出现的高饱和区域不是眼睛、头发或精致的装饰品这些通常被关注的区域,导致视觉效果较差。与  $L_{sm}$ 相似,明暗对比损失  $L_{gm}$ 对调整生成图像的灰度分布和使画面变得和谐有很大的作用。如图 7(f)所示,去除  $L_{gm}$ ,结果中会出现很多破坏画面和谐性的深色区域。



图 7 消融实验结果

Fig. 7 Ablation study

#### 4.4 量化比较实验

为证明利用了 HSV 监督信息的本文方法比仅使用 RGB 监督信息的现有方法得到的上色结果的视觉质量更高,现采用量化方式将所提方法与目前的 4 种先进方法进行比较,对比方法分别为 PaintsChainer 系列的 Canna<sup>[4]</sup>, Tanpopo<sup>[5]</sup> 和 Satsuki<sup>[6]</sup>,以及本文的基础工作 Ci 等<sup>[8]</sup>的方法。为公平起见,在相同实验设置下,采用提出的训练数据集对 Ci 等<sup>[8]</sup>的方法进行训练。

由于 PaintsChainer 的模型并不是开源的,实验中很难为所有方法提供完全一致的用户颜色提示。因此,本文采用无提示的上色任务在测试数据集上对所有的方法进行比较,部分比较结果如图 8 所示。可以看到,PaintsChainer 系列<sup>[4-6]</sup>及 Ci 等<sup>[8]</sup>的方法产生的结果易倾向于单一色相,相比之下,本文方法的结果具有更多色彩,且在饱和度及明暗分布方面,本文方法的结果更加和谐,便于观众区分和理解画面层次。而对比较方法的结果并不具备合理的饱和度及明暗分布,视觉效果较差。



图 8 自动上色对比结果

Fig. 8 Comparison of automatic colorization

本文采用两项量化指标衡量各方法的性能。首先是 FID<sup>[36]</sup> (Fréchet Inception Distance),它是一项在 GAN 方法中广泛使用的评估指标。本文采用 FID 来评估训练数据集中真实彩色图像与方法生成图像间的分布相似度,同时本文还计算了线稿提取数据集和训练数据集中彩色图像的 FID 分数作为参考。实验获得的 FID 分数记录在表 1 中。结果表明,本文方法得到了最低的 FID,证明本文方法相比其他方法能够生成更接近真实数据分布的上色结果。

表 1 FID 及 MOS 分数的量化比较

Table 1 Quantitative comparison of FID and MOS

Algorithm	FID	MOS
Sketch extractor dataset	53.296 0	—
PaintsChainer Canna	130.132 4	3.028 2
PaintsChainer Tanpopo	131.967 3	2.080 9
PaintsChainer Satsuki	147.195 0	2.229 0
Ci. et al	89.369 4	3.225 5
Our Method	68.365 1	4.182 7

为评价各方法的上色质量,本文采用了 MOS (Mean

Opinion Score)量化指标。评估流程如下:共邀请 20 位打分人员对方法上色结果进行评分,分数从 1(最差)到 5(最佳)不等。每个打分者需对 55 张线稿经 5 种方法上色后的版本进行评分,共 275 个实例。此后,计算各方法得分的平均值作为 MOS 指标,表 1 列出了该实验的结果。可以看出,本文方法优于其他方法,取得了最高的 MOS,证明了利用 HSV 监督信息学习绘画先验能够得到视觉效果更佳的上色结果。

#### 4.5 有颜色提示的上色比较

针对具有特定色彩设定的线稿,无颜色提示的上色结果

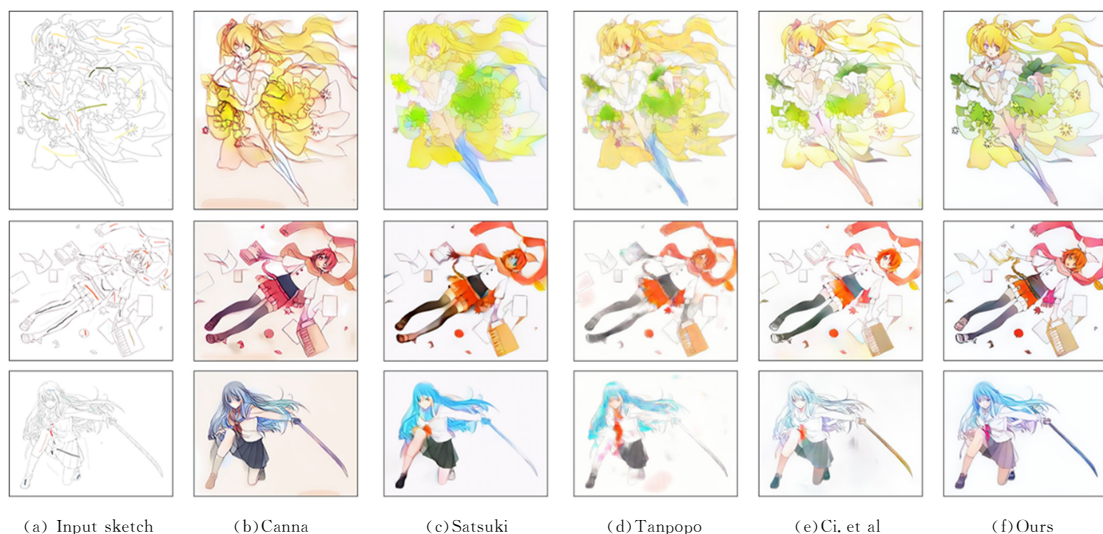


图 9 有用户提示的上色任务对比结果

Fig. 9 Comparison of colorization with user hint

**结束语** 本文受到实际绘画过程的启发,提出了一种绘画先验引导的线稿上色方法,该方法相比现有方法可得到视觉效果更好的上色结果。本文方法的主要贡献为:通过引入像素级 HSV 色彩空间约束来生成稳定的上色结果,并在此基础上设计启发式的色相变化损失、饱和度对比损失和明暗对比损失来学习插画师常用的绘画先验知识,以此引导生成器的训练,提升自动线稿上色的质量。通过使用 FID 和 MOS 指标评估生成的结果,证明了本文提出的利用 HSV 监督信息的方法有效提升了自动线稿上色的质量。

由于本文为首次尝试在自动线稿上色任务中引入绘画先验的工作,所设计的启发式损失仅考虑了广泛应用的绘画先验,方法仍具有一定的局限性。因此探索其他泛用性更高的引入绘画先验的方式,将是接下来的研究重点。

#### 参 考 文 献

[1] QU Y G, WONG T T, HENG P A. Manga colorization [J]. ACM Transactions on Graphics, 2006, 25(3): 1214-1220.  
 [2] SYKORA D, DINGLIANA J, COLLINS S. LazyBrush: Flexible Painting Tool for Hand-drawn Cartoons[J]. Computer Graphics Forum, 2009, 28(2): 599-608.  
 [3] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[C]// Advances in Neural Information Processing Systems. 2014: 2672-2680.

通常不能满足需求,而本文方法可在用户给出划线或点状颜色提示的情况下为线稿上色,得到符合特定色彩及风格设定的上色结果。具体操作为:为网络输入线稿的同时提供用户指定的颜色划线或点图,对其进行 4 倍下采样作为网络的颜色提示输入。如图 9 所示,与其他现有方法相比,本文方法在细节生成方面表现得更好,如眼睛的高光、阴影及布料的纹理等。同时,本文方法相比其他方法能更好地保存原有线稿,并在此基础上可区分出线稿中的合适区域,根据颜色提示分配合理的颜色。综上所述,本文方法在可交互上色中的色彩控制上表现更加出色。

[4] TAIZAN. Paintschainer canna[EB/OL]. [https://petalica-paint.pixiv.dev/index\\_en.html](https://petalica-paint.pixiv.dev/index_en.html).  
 [5] TAIZAN. Paintschainer tanpopo[EB/OL]. [https://petalica-paint.pixiv.dev/index\\_en.html](https://petalica-paint.pixiv.dev/index_en.html).  
 [6] TAIZAN. Paintschainer satsuki [EB/OL]. [https://petalica-paint.pixiv.dev/index\\_en.html](https://petalica-paint.pixiv.dev/index_en.html).  
 [7] ZHANG L, LI C, WONG T T, et al. Two-stage sketch colorization[J]. ACM Transactions on Graphics, 2018, 37(6): 1-14.  
 [8] CI Y Z, MA X Z, WANG Z H, et al. User-guided deep anime line art colorization with conditional adversarial networks [C] // ACM Multimedia Conference on Multimedia Conference. 2018: 1536-1544.  
 [9] YOO S J, BAHNG H J, CHUNG S H, et al. Coloring with limited data: Few-shot colorization via memory augmented networks [C] // IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA, 2019: 11283-11292.  
 [10] SMITH A R. Color gamut transform pairs[C] // Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1978. Atlanta, GA, USA, 1978: 12-19.  
 [11] KIM E J, SUK H J. Hue extraction and Tone match: Generating a Theme Color to Enhance the Emotional Quality of an Image[J/OL]. ACM Siggraph. <https://dl.acm.org/doi/abs/10.1145/2787626.2787657>.  
 [12] LYNCH D K, LIVINGSTON W. Color and light in nature[J].

- Optometry and Vision, 2001, 74(6):1342-1343.
- [13] JAMES G. Color and Light: A Guide for the Realist Painter [M]. Andrews Memeel Publishing, 2010.
- [14] HUANG Y C, TUNG Y S, CHEN J C, et al. An adaptive edge detection based colorization algorithm and its applications[C]// Proceedings of the 13th Annual ACM International Conference on Multimedia. ACM, 2005: 351-354.
- [15] LEVIN A, LISCHINSKI D, WEISS Y. Colorization using optimization[J]. ACM Transactions on Graphics, 2004, 23: 689-694.
- [16] ISOLA P, ZHU J Y, ZHOU T H, et al. Image-to-Image Translation with Conditional Adversarial Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, 2017: 5967-5976.
- [17] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional gans[C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, 2018: 8798-8807.
- [18] CHEN W, HAYS J. SketchyGAN: Towards Diverse and Realistic Sketch to Image Synthesis[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2018. Salt Lake City, UT, USA: 9416-9425.
- [19] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C] // IEEE International Conference on Computer Vision (ICCV 2017). Venice, Italy, 2017: 2242-2251.
- [20] YI Z, ZHANG H, TAN P. Dualgan: unsupervised dual learning for image-to-image translation[C]// IEEE International Conference on Computer Vision (ICCV 2017). Venice, Italy, 2017: 2868-2876.
- [21] FURUSAWA C, HIROSHIBA K, OGAKI K, et al. Comi-colorization: Semi-Automatic Manga Colorization [J]. SIGGRAPH Asia 2017 Technical Briefs, 2017, 12: 1-4.
- [22] FRANS K. Outline colorization through tandem adversarial networks[J]. arXiv: 1704. 08834, 2017.
- [23] LIU Y F, QIN Z C, WAN T, et al. Auto-painter: cartoon image generation from sketch by using conditional wasserstein generative adversarial networks [J]. Neurocomputing, 2018, 311: 78-87.
- [24] MEHDI M, SIMON O. Conditional Generative Adversarial Nets[J]. arXiv: 1411. 1784, 2014.
- [25] HYUNSU K, HO Y J, EUNHYEOK P, et al. Tag2Pix: Line Art Colorization Using Text Tag With SECat and Changing Loss[C]// 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South), 2019: 9055-9064.
- [26] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein gans[C]// Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, 2017. Long Beach, CA, USA: 5767-5777.
- [27] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated Residual Transformations for Deep Neural Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017: 5987-5995.
- [28] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017: 5987-5995.
- [29] SHI W, CABALLERO J, FERENC H, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 1874-1883.
- [30] SAITO M, MATSUI Y. Illustration2Vec: a semantic vector representation of illustrations[C]// SIGGRAPH Asia 2015 Technical Briefs. Kobe, Japan, 2015, 5: 1-4.
- [31] LEDIG C, THEIS L, HUSZAR F, et al. Cunningham A and Acosta A. Photo-realistic single image super-resolution using a generative adversarial network[C]// IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2016: 105-114.
- [32] KARRAS T, AILA T, LAINE S, et al. Progressive growing of gans for improved quality, stability, and variation [J]. arXiv: 1710. 10196.
- [33] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.
- [34] JUSTIN J, ALEXANDRE A, LI F F. Perceptual losses for real-time style transfer and super-resolution[C]// Computer Vision (ECCV 2016). 14th European Conference. Amsterdam, The Netherlands, 2016: 694-711.
- [35] WINNEMOELLER H, KYPRIANIDIS J E, OLSEN S C. Xdog: an extended difference-of-gaussians compendium including advanced image stylization [J]. Computers and Graphics, 2012, 36(6): 740-753.
- [36] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[C]// Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, 2017. Long Beach, CA, USA: 6626-6637.



**DOU Zhi**, born in 1996, postgraduate. His main research interests include computer vision and image generation.



**LI Hao-jie**, born in 1972, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include computer vision and image processing.