

# 基于参考图语义匹配的花卉线稿工笔效果上色算法

李 媛 陈昭炯 叶东毅  
(福州大学计算机与大数据学院 福州 350108)  
(liyuan057@126.com)

## A Coloring Algorithm for Flower Line Drawings with Meticulous Effect Based on Semantic Matching of Reference Images

Li Yuan, Chen Zhaojiong, and Ye Dongyi  
(College of Computer and Data Science, Fuzhou University, Fuzhou 350108)

**Abstract** The problem of coloring flower line drawings with meticulous effect based on a reference image is addressed. Existing reference-based coloring algorithms for line drawing are difficult to learn and simulate the unique color gradient effect of meticulous flower paintings. Moreover, the reference image in these algorithms is usually required to have similar geometric layout structure to the line drawing, which limits the applicability of the algorithms. Therefore, it is difficult to directly apply existing algorithms to accomplish coloring of line drawings with meticulous effect. On the basis of conditional generative adversarial network(CGAN) framework, a coloring algorithm for flower line drawings with meticulous effect is proposed by means of semantic matching between the reference image and the line drawing. In terms of network structure design, the proposed algorithm uses U-Net as the basis of the generator and designs two additional sub-modules. One is the semantic positioning sub-module. This module pre-trains a semantic segmentation network to generate a semantic label map of the flower line drawing. The label map is encoded as an adaptive instance normalization affine parameter and then introduced into the coloring model to improve the recognition ability of different semantic regions and the accuracy of color positioning. The other is the color coding sub-module. This module extracts the color features of the reference image, and then splices to the first three decoding layers of the generator, in which way, the color information is injected into the color model. Combining this module with semantic location module, our algorithm enhances the learning and simulation of gradient color pattern. In network training stage, the algorithm does not train the model on “original meticulous flower work-flower line drawing” data pairs. Instead, a perturbed version of the original work via such perturbation operations as disturbing the original geometric structure is generated and then “perturbed version-flower line drawing” data pairs are used to train our model, which turns out to reduce the model’s dependence on the spatial geometry layout of the original work and to then improve the applicability of the proposed algorithm. The experimental results show that the proposed algorithm has a correct response to the color semantics of the reference image selected by the user. It is also shown that the introduced structure of semantic positioning module and color

coding module could improve the simulation effect of gradient colors and realize the colorization of the flower line drawing under the guidance of different reference images, as well as diversified coloring results.

**Key words** meticulous flower coloring; semantic matching; conditional generative adversarial network (CGAN); semantic segmentation network; adaptive instance normalization

**摘 要** 研究基于参考图像的花卉线稿图的工笔效果上色问题,现有的基于参考图像的线稿图上色算法对工笔花卉画特有的色彩渐变的特点难以学习和模拟;此外通常还要求参考图像与线稿图具有相似的几何布局结构,这也限制了算法的适用性,故而直接采用现有算法难以实现线稿图的工笔效果上色.基于条件生成对抗网(conditional generative adversarial network, CGAN)框架,提出了一种将参考图像与线稿图进行语义匹配的花卉线稿图工笔效果上色算法 RBSM-CGAN.该算法在网络结构设计方面,以 U 型网络(简称 U-Net)为生成器基础,设计了 2 个附加子模块:1)语义定位子模块.该模块预训练了一个语义分割网络,以生成花卉线稿图的语义标签图,该标签图编码后作为自适应实例归一化的仿射参数引入到上色模型中,提升对不同语义区域的识别能力,进而提高颜色定位的准确性.2)颜色编码子模块.该模块提取参考图像的颜色特征,而后将该特征拼接到生成网络解码层的前 3 层,利用这种方式将颜色信息注入上色模型,与语义定位模块相配合加强算法对渐变色的学习和模拟.另外,算法在网络训练方面改变传统的“工笔花卉原作-花卉线稿图”数据对的训练方式,通过打乱原作的几何结构等摄动操作生成原作摄动图,采用“原作摄动图-花卉线稿图”数据对进行网络训练,降低了模型对原作空间几何结构的依赖性,提升了算法的适用性.实验结果表明:该算法对用户选择的参考图像的颜色语义具有正确的响应,所引入的“语义定位+颜色编码”的结构设计提升了对渐变色的模拟效果,实现了在不同参考图像指导下的花卉线稿图的工笔效果上色,可快速生成多样化的上色结果.

**关键词** 工笔花卉上色;语义匹配;条件生成对抗网络;语义分割网络;自适应实例归一化

**中图法分类号** TP391

工笔花卉画是中国传统绘画形式之一,创作过程通常分成线描图绘画和上色 2 个阶段.首先以线笔勾勒出物体工整细致的外形结构形成线稿图,而后结合主客观因素灵活涂色完成线稿的上彩.其中上彩技法最能表现工笔花卉画的视觉效果和艺术性,传统手工上彩采用薄色多涂的方式,每次蘸取少量精心调制的颜色进行均匀涂抹,重复多次,渲染出对象的明暗关系和层次变化以提升画面的质感.这一阶段耗时耗力,特别是一幅线稿图如果想渲染成不同色调风格的效果,每一次都需要重新调色、重新薄染.本文针对工笔线稿图的自动上色问题进行研究,用户可选定已有的、心仪的工笔彩图作为参考图像,本文系统通过仿真算法自动提取参考图中的色调对输入的花卉线稿图进行颜色渲染,生成与参考图色彩风格相近的工笔彩图.用户通过选定不同的参考图像,就可快速获得线稿图不同的渲染效果.

本文的研究工作属于图像上色方向,图像上色是计算机视觉领域的研究热点之一,当前深度学习的兴起和发展对其产生了很大的促进作用.图像上

色按上色主体可分成灰度图像上色和线稿图上色.基于深度学习框架的灰度图上色<sup>[1-2]</sup>研究取得了较好的效果,上色算法依据灰度图自身携带的亮度、纹理信息识别不同的语义对象,进而渲染出自然真实的图像色彩;而本文的研究对象是线稿图上色,线稿图仅勾勒物体的外形轮廓和主要线条,没有亮度以及明暗信息,且本文的上色目标不是渲染出真实感图像,而是模拟工笔画色彩效果,这就带来了上色过程中对线稿对象语义的辨识和艺术效果处理上的难度.

目前基于深度学习的线稿图上色方法根据交互方式通常分成两大类:

1) 用户通过笔划、颜色文本词、引导图等形式设计配色信息,算法将颜色信息引入到上色模型中完成上色,上色方法通常在交互式场景下完成<sup>[3-4]</sup>.2018 年,Zhang 等人<sup>[5]</sup>通过颜色线条指定线稿图各区域的颜色,用条件生成对抗网络(conditional generative adversarial network, CGAN)对同质区域进行色彩扩展完成上色,算法通过反复调整颜色线条能实现精细化的上色,但也导致需要较多的人机交互;Qiao

等人<sup>[6]</sup>根据文本生成图像,又对图像重新生成文本描述,在保证原始文本和生成文本语义一致性的前提下,约束中间图像和文本不同模态数据之间的语义对齐,因为涉及对不同模态数据的理解,故文本标签在颜色作用域的定位上表现不佳;李媛等人<sup>[7]</sup>针对线稿图工笔花卉的上色设计了一种新的作为人机交互的花色注意力图,对花朵的颜色类别和位置进行刻画,既实现了对颜色的控制,又加强了区块色彩的学习,对工笔画特有的渐变色的模拟效果较好,但对色彩的控制还比较粗糙,仅将色彩分成4种类型.这种上色方式可实现用户设计、模型驱动的半自动艺术创作,比较适用于有一定美术基础的用户,但形态各异的交互设计方式需要用户花更多的精力去适应,且算法难以快速呈现多样化的结果.

2) 基于参考图像的色彩仿真方法<sup>[8-10]</sup>,用户选取已有的彩色图像作为配色样板,算法实现整体色彩风格的迁移,这种方式自动化程度较高,适合一般用户使用.2016年,Gatys等人<sup>[11]</sup>提出一种艺术风格(迁移)的神经算法(a neural algorithm of artistic style, ANAS),当参考图像的艺术风格鲜明奔放时(如印象派油画、波普艺术),迁移的效果较好,但工笔画色彩典雅工丽,使用该方法上色效果不佳;Huang等人<sup>[12]</sup>提出无监督的多模态图像转换模型,将图像空间解耦成内容空间和风格空间,通过保留内容特征,融入不同参考图像的风格特征实现多模态输出,该方法适用于不同的图像转换任务,包括工笔线稿图上色,但由于采用无监督的训练方式,较难控制内容和风格的融合程度,如果参考图像与线稿图的空间结构存在较大差异,可能导致生成结果局部失色甚至缺失对主体对象的颜色渲染;2020年,Zhang等人<sup>[13]</sup>提出基于参考图像的跨域图像转换模型,通过构造一个相关性矩阵,学习2个域图像的密集语义对应关系,根据对应关系在参考图像中找到与源图每个图像块最匹配的区块,之后训练一个转换网络融合图像块特征,以提升生成效果.该方法应用于线稿图上色能较好地将参考图像的整体色调转移到线稿图中,但对边界的定位较粗糙,容易出现上色不足或颜色溢出现象;针对基于参考图像的线稿图上色模型在训练过程中图像数据集不易构造的问题, Lee等人<sup>[14]</sup>提出了对原作进行扭曲、颜色扰动等得到训练彩图,然后分别提取训练彩图和线稿图的卷积特征,在语义层面上寻找二者不同区域最匹配的图像块,从而实现线稿图上色.但是由于线稿图和彩色图携带的信息差异较大,语义匹配时容

易出错,生成图像易出现色彩失真.上述各种方法在线稿图上色方面取得了一定的效果,但是当参考图像和输入的线稿图结构差异较大时,易产生语义信息匹配错误而导致上色错误、溢出等问题;并且由于这些上色过程缺乏针对工笔画特有的渐变色进行模拟的功能设计,因而用于工笔花卉上色的模拟效果不佳.

鉴于上述不足并针对工笔花卉画的染色特点,本文基于CGAN框架,提出了一种将参考图像与线描图进行语义匹配的工笔花卉上色仿真算法(reference-based and semantic matching, RBSM-CGAN).该算法在网络结构设计方面,将生成器设计成3个模块:U-Net模块、线稿语义定位模块和参考图颜色编码模块.其中U-Net为生成器基础,对输入线稿图进行编、解码,生成同尺幅的工笔彩图;线稿语义定位模块是一个预训练的线稿语义分割网络,所生成的语义标签图将信息引导到U-Net的解码器部分,帮助上色模型识别线稿不同的语义区域;参考图颜色编码模块提取参考图的颜色特征导入U-Net,而不对其进行语义分割,参考图像和线稿图之间的空间语义对应关系是通过本文设计的训练方式学习到的,最终完成将参考图像的颜色准确定位到线稿图对应的语义区域.这3个模块设计有助于对参考图像渐变色的捕获、学习和模拟.其次,算法在网络训练方面摒弃了传统的“工笔花卉原作-花卉线稿图”数据对的方式,采用“原作摄动图-花卉线稿图”数据对进行训练,同时还设计了合适的损失函数,2种设计策略降低了模型对原作空间几何结构的依赖性,因而在算法应用时无需特别顾及参考图像的几何构成,参考图的选择就有了更大的自由度,提升了本文算法的适用性.

## 1 算法主要思想及流程图

### 1.1 流程图及分析

本文提出的RBSM-CGAN工笔花卉上色算法的网络训练流程如图1所示,网络由预训练的语义分割网络、生成器以及判别器3个主要部分构成.为了提升对花卉线稿图不同语义区域的识别能力,本文预训练了一个语义分割网络对花卉线稿图进行分割,训练时分割得到的语义标签图连同花卉线稿图以及原作摄动图同时输入到生成器中,在语义标签图的指导下,生成器网络学习花卉线稿图和原作摄动图之间的空间语义对应关系,根据语义将原作

摄动图的颜色信息迁移到花卉线稿图的对应区域,进而完成上色.(花卉线稿图,原作)或(花卉线稿图,

生成图像)构成的二元组作为正/负样本对输入到判别器网络,训练网络识别真假图片.

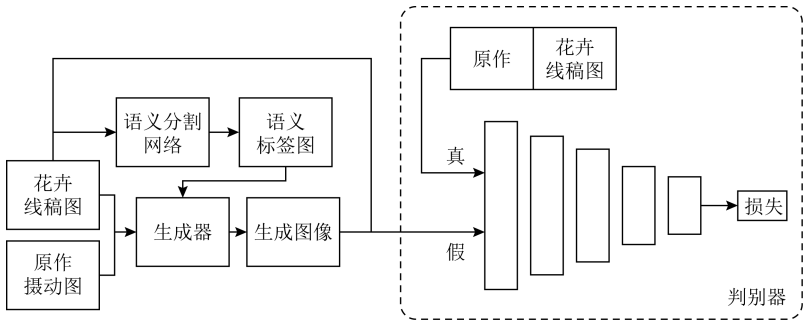


Fig. 1 Flow chart of RBSM-CGAN network in training stage  
图 1 训练阶段 RBSM-CGAN 网络流程图

测试及应用阶段的网络流程如图 2 所示,输入花卉线稿图以及选定的一幅已有的彩色图像作为参考图像,参考图像的布局结构可以不同于线描图,训练好的生成器自动将花卉线稿图渲染成符合参考图像色调信息的工笔彩图,同时参考图像中花、叶、背景等颜色按照语义对应关系自然传递到花卉线稿图的对应位置.通过选取不同的参考图像,可以快速将花卉线稿图渲染成不同色调风格的彩图效果.

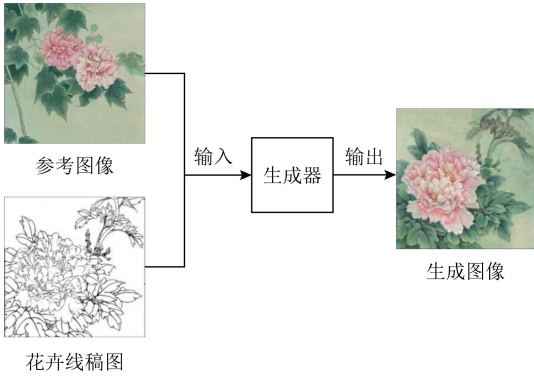


Fig. 2 Flow chart of RBSM-CGAN network in application stage  
图 2 应用阶段 RBSM-CGAN 网络流程图

本文的主要工作是生成器的构造和训练方式的设计,分别在后续第 2 节和第 3 节进行详细论述.下面仅就判别器和损失函数的设计进行分析.

1.2 判别器与损失函数

传统判别器模型普遍采用深度卷积网络,利用网络自主学习表征图像的特征,当网络输入真实图像和生成图像时输出一个概率值,该值表示输入图像属于真实图像的概率.此方法用于评估生成图像整体质量上表现较好,但是工笔花卉画上色仿真在细节

生成上要求较高,比如边缘处颜色处理、渐变色的模拟等.为了加强局部区域的颜色学习,本文采用块生成对抗网络(patch generative adversarial network, PatchGAN)<sup>[15]</sup>对图像局部区域进行评估.具体思路是将图像划分成一系列 70×70 的图像块,判别器网络对每个图像块的生成质量进行评估,最后对所有图像块的判别结果取平均即为整幅图像的判别结果,该结果作为目标函数的损失项更新生成器和判别器的网络参数.

本文的损失函数除了使用  $L_{adv}$  对抗损失约束生成图像符合原作的颜色分布外,还引入  $L_1$  损失,惩罚网络生成图像与原作的颜色差异. $L_1$  损失硬性要求生成图像和原作的像素对齐,可能导致生成图像色彩在空间上不连续,为了消除这一现象,进一步新增了  $L_f$  特征损失,使得生成图像和原作在高层语义上对齐.本文上色网络最终的损失函数  $L_{RBSM-CGAN}$  为:

$$L_{RBSM-CGAN} = \lambda_{adv} L_{adv} + \lambda_1 L_1 + \lambda_f L_f, \quad (1)$$

$$L_{adv} = \min_G \max_D E_{x \sim p_\delta(y)} [\ln D(x, y)] +$$

$$E_{c \sim \text{perturbe}(y), x \sim p_\delta(y)} [\ln(1 - D(x, G(c, x)))], \quad (2)$$

$$L_1 = E_{c \sim \text{perturbe}(y), x \sim p_\delta(x), y \sim p_\delta(y)} [\|y - G(c, x)\|_1], \quad (3)$$

$$L_f = E_{c \sim \text{perturbe}(y), x \sim p_\delta(x), y \sim p_\delta(y)} [\|\phi(y) - \phi(G(c, x))\|_1], \quad (4)$$

其中,  $\lambda_{adv}, \lambda_1, \lambda_f$  分别表示  $L_{adv}$  对抗损失、 $L_1$  损失和  $L_f$  特征损失的正则系数,  $c$  表示原作摄动图,  $x$  表示花卉线稿图,  $y$  表示原作;  $G(c, x)$  表示生成器以花卉线稿图  $x$  和参考图像  $c$  为条件输入时的生成结果,  $\phi(y)$  和  $\phi(G(c, x))$  表示在 ImageNet 数据集<sup>[16]</sup>上的预训练模型 VGG19 对原作  $y$  或生成图像



$G(c, x)$ 提取的中间层特征; $D(x, y)$ 与 $D(x, G(c, x))$ 表示输入图像 $y$ 或 $G(c, x)$ 属于真实图像并且跟花卉线稿图 $x$ 同线条结构的概率。

2 RBSM-CGAN 生成器设计原理及结构

如图 3 所示,生成器网络包括 3 个分支,最上方

的分支是以 U-Net 为基础的网络结构,负责将花卉线稿图转换成工笔彩图;中间分支是语义定位模块,辅助生成器提升对不同语义区域的颜色渲染效果;最下方的分支是颜色学习模块,提取参考图像的颜色信息,将其引入到生成器网络中.3 个分支相辅相成,最终完成将参考图像的颜色迁移到花卉线稿图对应的语义区域。

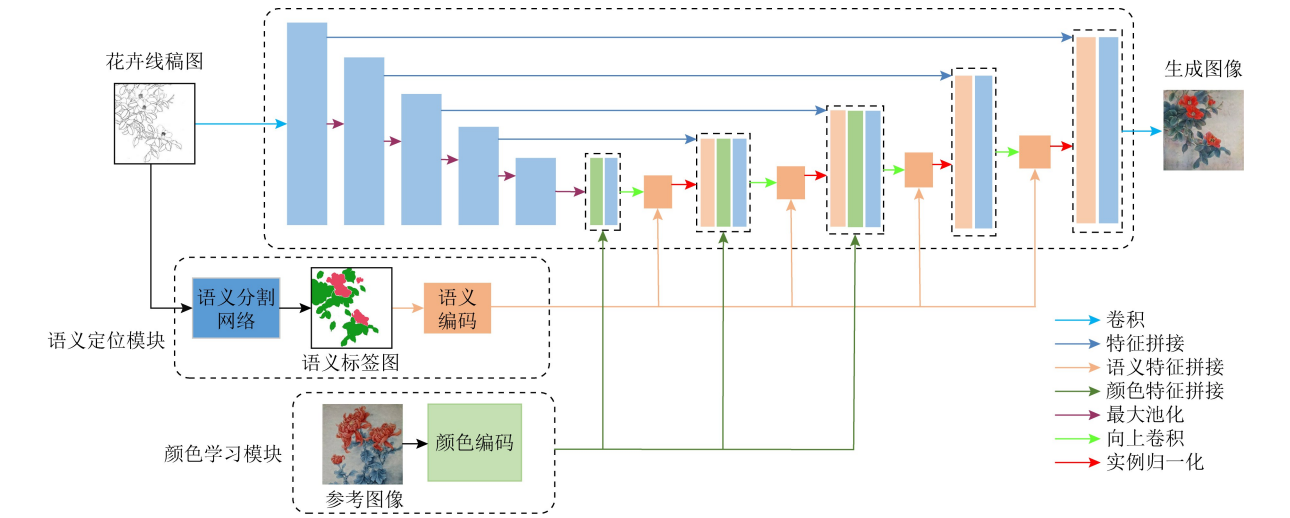


Fig. 3 Architecture of generator of RBSM-CGAN  
图 3 RBSM-CGAN 生成器结构

2.1 基础的 U-Net 结构

现有的“编码-解码”生成器结构<sup>[17]</sup>通过编码网络提取图像特征,又经解码网络重构回原始输入图像空间,可以实现图像的异质转换,包括将花卉线稿图转换成工笔彩图.由于卷积网络经过一系列卷积下采样可能丢失图像的空间特征以及线条信息,导致生成结果颜色平缓、边缘轮廓模糊不清.针对这个问题,本文的生成模型采用 U 型网络(简称 U-Net)<sup>[18]</sup>作为基础框架,U-Net 引入了跨层连接的结构,将编码网络的特征复用到解码网络,如图 3 蓝色箭头所示,该结构加强了特征在不同层之间的传递,缓解了深层特征偏重刻画全局语义可能丢失浅层特征的不足。

如图 3 所示,生成器包含 10 层,卷积核采用 4×4,各层具体参数如表 1 所示.其中 Conv2d 表示 2 维卷积层,Deconv 表示反卷积,Conditional IN 表示条件实例归一化,ReLU 表示修正线性单元,LReLU 表示带泄露修正线性单元,tanh 表示双曲正切函数。

2.2 语义定位模块

工笔花卉上色仿真任务对对象语义区域定位的精确度要求较高,确保颜色准确定位到不同位置至关重要,而颜色定位需要考虑 3 个问题:1)花卉线稿

Table 1 Parameters of the Basic Generator U-Net				
表 1 基础生成器 U-Net 的参数				
基础生成器结构	层	结构	卷积核个数	步长
编码器	层 1	Conv2d;	64	2
	层 2	LReLU, Conv2d, Conditional IN;	128	2
	层 3	LReLU, Conv2d, Conditional IN;	256	2
	层 4,5	LReLU, Conv2d, Conditional IN;	512	2
	层 6	ReLU, Deconv, Conditional IN;	512	2
解码器	层 7	ReLU, Deconv, Conditional IN;	256	2
	层 8	ReLU, Deconv, Conditional IN;	128	2
	层 9	ReLU, Deconv, Conditional IN;	64	2
	层 10	ReLU, Deconv, tanh;	3	1

图不同语义对象的线条结构相似,如果生成网络错误地划分了语义区域进而颜色渲染,可能出现不协调的色块;2)若边界定位不够准确,着色边界较模糊,仿真结果边缘处可能出现颜色溢出或者颜色不足,

降低生成图像的艺术性;3)在网络训练阶段,原作摄动图与线稿图的空间布局虽然不同,但是语义内容上高度相关,测试阶段如果选取与花卉线稿图花型、形态或者布局上差异较大的参考图像,可能无法将颜色信息正确地迁移到线稿图.考虑到 U-Net 在特征提取中表现较好,但是对语义区域的识别不够精确,本模块设计并预训练了一个基于 pix2pix 网络模型<sup>[15]</sup>的语义分割网络,将花卉线稿图划分成 3 个语义区域,分别是花朵、叶子以及其他区域,指导生成网络区分不同对象,进而将参考图像的颜色传递到花卉线稿图的不同位置.

线稿图只有线条信息,不同语义区域难以区分,故本文借助预训练的语义分割网络将线稿图划分成不同语义区域,并将分割结果嵌入到 U-Net 生成器解码网络.而对于参考图像,不同语义对象之间的颜色存在较大差异,颜色即可较好地区分不同语义区,因此不再进行语义分割,使用颜色学习模块编码提取参考图的颜色特征后将其引入到 U-Net 网络.线稿图与参考图像之间的颜色语义关联关系可通过本文设计的训练方式学习获得.参见本文第 3 节.

花卉线稿图的分割结果是一幅与原图相同尺寸、不同语义区域用不同颜色指代的语义标签图,参见图 3.如果直接将语义标签图拼接到生成网络的特征层,生成网络归一化的过程可能丢失语义信息.2017 年 Huang 等人<sup>[9]</sup>提出的自适应实例归一化(adaptive instance normalization, AdaIN)被用在多种类型的图像生成任务中,比如风格迁移.AdaIN 通过改变特征图每个通道上的实例归一化的系数将风格特征引入到网络结构中,又由于实例归一化的系数在每个通道上是不同的,因而可实现风格特征的空间自适应调整.AdaIN 的计算公式为

$$A(x, y) = \sigma(y) \left( \frac{x - \mu(x)}{\sigma(x)} \right) + \mu(y), \quad (5)$$

其中,  $x$  表示内容特征,  $\mu(x)$  和  $\sigma(x)$  分别表示  $x$  的均值与均方差;  $y$  表示风格特征,  $\mu(y)$  和  $\sigma(y)$  分别表示  $y$  的均值与均方差,也即实例归一化中的仿射参数.本文将自适应实例归一化的思想引入到语义定位的网络模块,将语义标签图编码后的特征向量作为自适应实例归一化的仿射参数引入到生成器网络解码层.语义编码网络的结构设计如图 4 所示.相关的计算为:

$$F_{c,h,w} = \alpha_{c,h,w}(\text{Mask}) \left( \frac{f_{c,h,w} - \mu_c}{\sigma_c} \right) + \beta_{c,h,w}(\text{Mask}), \quad (6)$$

$$\mu_c = \frac{1}{H \times W} \sum_{h,w} f_{c,h,w}, \quad (7)$$

$$\sigma_c = \sqrt{\frac{1}{H \times W} \sum_{h,w} (f_{c,h,w})^2 - (\mu_c)^2}, \quad (8)$$

其中, **Mask** 表示语义标签图;  $f_{c,h,w}$  表示自适应实例归一化之前在  $(c, h, w)$  位置的特征输入,  $F_{c,h,w}$  表示表示经过自适应实例归一化后在  $(c, h, w)$  位置的特征输出,  $c \in C, h \in H, w \in W, C, H, W$  分别表示当前特征图  $f$  的通道数、高度和宽度.  $\mu_c$  和  $\sigma_c$  分别表示当前输入特征图  $f$  第  $c$  个通道的均值和标准差.  $\alpha_{c,h,w}(\text{Mask})$  和  $\beta_{c,h,w}(\text{Mask})$  既表示语义标签图经过语义编码后在  $(c, h, w)$  位置的 2 个颜色特征值,同时也是自适应实例归一化的仿射参数.

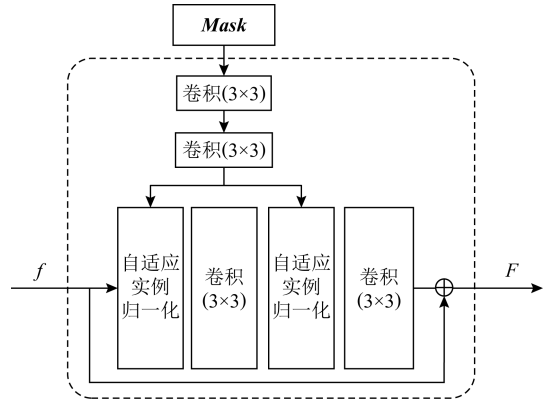


Fig. 4 Semantic coding network

图 4 语义编码网络

该结构简化了上色模型的语义对象识别过程,指导上色模型学习参考图像和花卉线稿图的语义对应关系,提升语义匹配的精确度以及模型渲染与参考图像外形轮廓差异较大的线稿图的能力;同时该语义标签图隐式地引入了内部注意力机制,加强了不同语义区域的颜色学习.

### 2.3 颜色学习模块

在训练过程和测试过程中,彩色参考图像与花卉线稿图同时输入到生成网络.彩色参考图像泛指模型中用到过的提供色彩信息的参考图像.在训练阶段“彩色参考图像”是原作摄动图,在测试过程中“彩色参考图像”是用户选取的工笔彩图. U-Net 编码部分完成花卉线稿图的线条信息提取,在参考图像颜色信息的提取方法上,本文设计的颜色编码网络由 2 层卷积神经网络构成,结构如图 5 所示.该模块用于提取参考图像的颜色特征,并将其拼接到生成器解码网络的前 3 层中(图 3 绿色箭头所示),通过该方式将颜色信息引入上色模型,同时拼接到低

分辨率的特征层可以避免对参考图像颜色布局信息的依赖,提升模型泛化性.

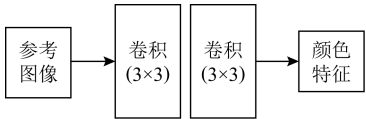


Fig. 5 Color coding network  
图 5 颜色编码网络

在应用阶段用户可以选取任意一幅花型、形态和空间结构与线稿图不同的工笔彩图作为上色参考,算法将颜色信息传递到花卉线稿图对应的语义区域,进而完成工笔花卉线稿图的上色仿真.要实现该过程单纯依靠颜色编码网络是不够的,还必须与训练阶段的策略相结合,具体参见第 3 节.

3 训练过程的处理策略

3.1 网络训练策略

本文的网络训练过程参见图 1,整个流程建立在 CGAN 的训练基础上,其中关键点是生成器训练时配对的彩色图像的构造.成对的“工笔花卉原作-花卉线稿图”可以通过收集及边缘检测算法得到,但是构造反映工笔花卉颜色信息的训练参考图像数据集十分困难,如果将原作直接当成训练参考图像可以节省很多人力物力,但是原作和输入线稿具有完全相同的线条结构,特征表达能力过强,将导致上色网络学到除颜色信息之外的其他信息,比如线条信息等.这样训练的网络在应用时当用户选择的与花卉线稿图结构形式差异很大的参考图像,会出现错误的响应.当训练的网络过度依赖原作,可能导致生成图像出现极端情况,即上色结果图是花卉线稿图和参考图像线性叠加的结果,如图 6(c)所示.

针对这个问题,本文提出训练时对原作进行翻转、错切等几何变换,打乱图像空间结构得到摄动图



Fig. 6 Wrong generation results of using original image in training stage

图 6 训练过程采用原作的错误生成结果

像作为训练阶段的参考图像,既降低线稿图和参考图像结构上的相似性同时又能保留线稿图上色所需的颜色信息.除了几何变换,本文的摄动图还采用基于像素点的圆形扭曲变形操作,公式为

$$q = p - \left( \frac{r^2 - |p - o|^2}{(r^2 - |p - o|^2) + |m - o|} \right)^2 (m - o), \quad (9)$$

其中,  $o$  表示圆心像素点,  $r$  表示圆的半径,  $m$  表示圆心像素点  $o$  移动后的位置,  $p$  表示圆内任意一个像素点.当圆心  $o$  转移到位置  $m$  后,圆内像素点  $p$  由像素点  $q$  的位置转移而来,也就是说像素点  $q$  的颜色决定了变形后像素点  $p$  的颜色.式(9)表明离圆心越远的像素点,对应的偏移量越小,发生形变的程度就越小.

该步骤使得原作摄动图和花卉线稿图线条结构在像素层面也有较大差异,希望上色网络学习到训练参考图像的颜色信息,而摒弃对象的位置、结构等信息.

考虑到原作摄动图由原作变换而来,可保持内容上的相关性,同时在颜色上也具有强关联,为了降低对颜色的强依赖,本文对摄动图进一步做模糊处理,采用  $5 \times 5$  的高斯滤波核处理图像,达到图像色彩模糊化的目的,同时不会过度损失像素点的颜色值,提升模型的鲁棒性.训练用的参考图像最终经由对原作采用翻转、错切、扭曲和模糊化处理后获得,本文称之为原作摄动图,具体步骤如算法 1 所示. 4 个摄动操作与顺序无关,结果如图 7 所示:

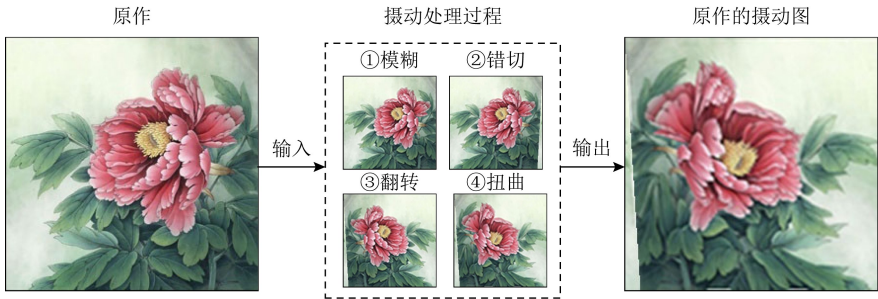


Fig. 7 The generation of perturbed image of the original  
图 7 原作摄动图生成



算法 1. 原作摄动图的生成算法.

输入:工笔花卉原作  $A_1$ ;

输出:工笔花卉原作对应的摄动图  $A_5$ .

记当前输入图像为  $A_1$ , 大小为  $m \times n$ , 摄动处理过程有 4 个步骤:

- ① 模糊.使用  $5 \times 5$  高斯滤波器对图像  $A_1$  滤波得到图像  $A_2$ .
- ② 错切.图像  $A_2$  从上到下,第  $i$  行往左移动  $\text{int}(i/15)$  个像素,并将左侧延展出来的直角三角图像块填充到右侧空白的三角形中,得到图像  $A_3$ .
- ③ 翻转.将图像  $A_3$  左右翻转得到图像  $A_4$ .
- ④ 扭曲.按式(9)对图像  $A_4$  中的像素点进行扭曲变形,进一步打乱图像的空间位置信息,得到图像  $A_5$ ,即工笔花卉原作的摄动图

图.其中,半径  $r=50$ .

同时本文设计的训练方式是将“原作摄动图-原作线稿图”配对输入生成器并要求生成器的输出结果为原作,如图 8 所示,原作摄动图与原作线条的空间结构发生了一定的变化,这就形成了训练彩图与线稿图线条结构的差异性;训练时要求生成器输出的结果是工笔画原作,这一方式就使得摄动图的颜色信息(花朵、叶子、花梗、背景)可以关联到该颜色在线稿图对应的语义位置,同时我们在生成器中引入线稿语义分割模块,可使生成器更准确地定位颜色语义区.在训练完成后,网络应用时,参考图像就类似于摄动图的角色,如图 2 所示,尽管参考图的线条结构可以与输入的线稿图有较大差异,但依然生成器可以较好地捕捉参考图的色彩信息并与输入线稿图相关联.

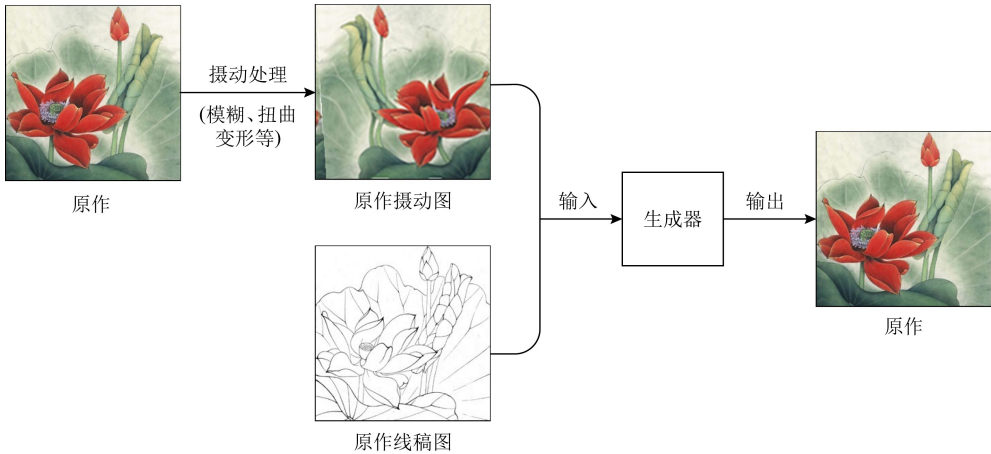


Fig. 8 Flow chart of RBSM-CGAN generator network in training stage  
图 8 训练阶段 RBSM-CGAN 生成器网络流程图

3.2 数据集的收集与处理

本文使用的“工笔花卉原作-花卉线稿图”数据集是从网络上收集获得,类别较丰富,涵盖牡丹、荷花、虞美人、郁金香、菊花等花卉种类.针对数据集中图片分辨率低以及数据不平衡等问题,采取图片筛选、去水印、数据增强等预处理操作;最终构造了一个包括 1 961 幅图片的训练集,其中 259 幅工笔花卉有画家手绘的对应花卉线稿图线稿,其余线稿图则通过扩展高斯差分算法(extended difference of Gaussians, XDoG)<sup>[19]</sup>提取工笔花卉的边缘线条获得.此外,本文上色仿真算法基于卷积神经网络,通过卷积核学习图像特征,为了防止图片大小悬殊引起卷积核参数收敛困难,网络训练不稳定,将数据集中图片的尺寸统一设置成  $256 \times 256$ .

高斯差分(difference of Gaussians, DoG)是基于二阶微分的边缘检测算子,通过计算二阶导过零点得到 1~2 个像素宽的细边缘.DoG 对图像进行不同尺度的高斯滤波得到 2 幅高斯图像,而后计算 2 幅高斯图像的差值得到边缘图,相当于一个带通滤波器,能过滤颜色渐变引起的假边缘,但是也可能过滤掉弱边缘.XDoG 是对 DoG 的改进,通过设置阈值保留差分图中边缘强度低的弱边缘.此外,为了消除噪声的影响,对检测得到的边缘图进行中值滤波,最终得到的检测结果线条较连贯、边缘定位较准确.训练数据集示例如图 9 所示.

本文使用的线稿图中线条的颜色有深有浅,是因为文中收集到的图片分辨率不同.分辨率低的图片携带的信息较少,边缘线条较模糊,对应的线稿图





Fig. 9 The presentation of dataset in this paper  
图 9 本文的数据集示例

线条较细,颜色较浅;反之,线条较粗,颜色较深,再一个原因是部分工笔画能够收集到画家留下的白描图,这类白描图通常线条比较清晰、颜色较深.除此之外,二者没有特殊的含义.为了模型的适应性更强,训练时浅色线稿图也不做增强.

4 实验结果及分析

本文在 Tensorflow 环境下实现花卉线稿图到工笔花卉的转换.生成器和判别器使用 Adam 优化器,学习率设置成 0.000 02,迭代 200 次,网络的训练与测试都在 GPU Tesla P100 上运行.

4.1 工笔花卉上色仿真效果

图 10 是选取了 9 幅不同色调的参考图像对 10 幅不同造型的花卉线稿图的上色结果.线稿图按照花卉类型选取,工笔画作中常见的荷花、牡丹、菊花各选 2 幅,余下 4 幅为虞美人、百合等杂花类;参考图像按照色调的类型选取,有红、橙、粉、绿、蓝、紫等,基本涵盖了工笔画常见的色调类型.图 10(b)粉色和绿色的组合勾勒出春天万物复苏、欣欣向荣的景象;图 10(c)由红色花朵和明黄的背景组成,鲜明的色彩造成视觉冲击;图 10(d)用相近色系的冷色烘托出牡丹的清新别致;图 10(e)深蓝的牡丹别具风格;图 10(f)是宋代工笔芙蓉花,采用石色,颜色厚重;图 10(g)粉白色的玉兰花开的绚烂,在灰绿叶

子的陪衬下独显清冷明艳;图 10(h)使用墨绿色的叶子以及浅绿的背景,衬托出桃红色花朵的鲜艳明亮;图 10(i)在嫩绿色叶子的衬托下,白色花朵愈加显现得干净纯洁;图 10(j)呈现了淡蓝色细长的菊花花瓣,与其他工笔花卉花型相差甚大.对比同一参考图下的不同线稿的上色效果可以发现,本文方法可适用于不同构图、造型的花卉线稿图,能较好地参考图像的色彩风格按照语义迁移到花卉线稿图中,生成图像颜色饱和、层次丰富,渐变色的模拟也有较好的效果.而同一花卉线稿图在不同参考图像的指导下呈现出不同风格的色彩效果,展示了网络输出的多样性.可以观察到,本文方法对参考图像的选取约束较小,如图 10 行 2 荷花线稿图,即使参考图像中的花朵没有花蕊,生成网络可以通过自主学习的方式对荷花线稿图中的花蕊上色.这是上色模型对所学习的数据集特点的一般性模拟效果.

画家笔下不少工笔花卉都有描绘花蕊,本文收集的数据集中也是如此.因此,在训练时,那些有花蕊的训练数据对,其花蕊的特征会被模型所学习到.在模型应用时,如果线稿图有花蕊,即便所选择的参考图中没有花蕊,上色模型也会对花蕊进行上色.实验结果可以看出生成图像颜色合理,能较好地模拟工笔花卉画的色调风格.

4.2 消融实验及分析

为了验证本文设计的算法的有效性,分别对本文提出的 3 个主要策略:语义定位模块、特征损失



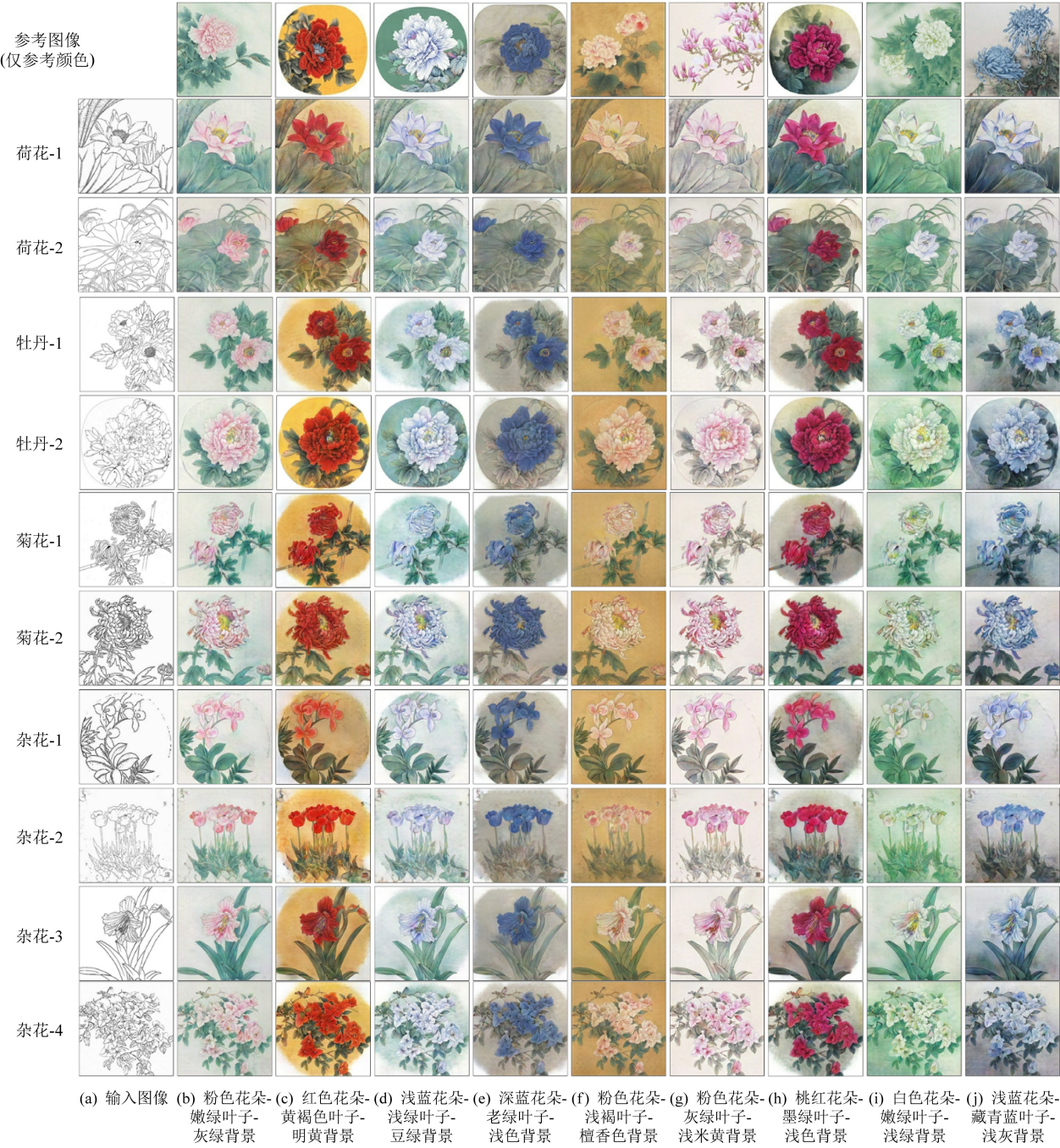


Fig. 10 Coloring results by the proposed algorithm

图 10 本文算法上色结果

函数和摄动图训练方式进行了消融实验及分析.实验结果表明本文算法在语义提取、颜色定位、渐变色模拟效果等方面均表现较优.

4.2.1 语义定位模块的消融测试

本文在生成器网络中引入预训练的语义分割网络,旨在通过语义标签图辅助生成网络识别不同语义区域,提升参考图像和花卉线稿图空间语义匹配精确度.为了验证语义分割模块的有效性,实验对比

了有无语义分割模块的生成效果,结果如图 11 所示.没有语义定位模块的生成网络当参考图像和花卉线稿图的外形轮廓相似时,渲染效果尚可,如图 11(c)行 1 芙蓉花,叶子渲染成黄绿色,背景米黄色,整体颜色风格和参考图像相似;当二者花型或者构图差异较大时,上色模型颜色定位较差,如图 11(c)行 2 牡丹黄色的花朵只有部分上色,且黄色花瓣中混杂了黑色和红色,仿真效果较差;又比如图 11(c)





Fig. 11 Comparison of ablation results with or without semantic segmentation module

图 11 有无语义定位模块的消融实验结果对比

行 3 杂花图叶子中部分区域颜色空洞,缺乏色彩;而图 11(c)行 4 荷花花瓣上色不足,花朵上方又出现了颜色溢出,色彩定位较差.对比本文算法带有语义定位模块的上色结果,整体色彩风格与参考图像保持一致,语义定位较准确,图像失真较少;可以看出语义标签图确实加强了不同区域的颜色学习,生成结果色彩厚重饱满、过渡自然,比较符合真实工笔花卉画的颜色分布和特点.

为了定量分析生成图像的质量,本文采用弗雷歇初始距离(Frechet inception distance,  $FID$ )<sup>[20]</sup>进行客观评价,结果如表 2 所示. $FID$  通过计算生成图像和目标图像隐层空间特征的 Wasserstein-2 距离评估图像的生成质量,分数越低图像越好.由表 2 可见,本文算法生成图像的  $FID$  比未使用语义分割模块的生成图像低,说明语义定位校准模块的有效性.

Table 2  $FID$  Comparison of Images Generated by Different Algorithms

表 2 不同算法生成图像的  $FID$  对比

算法	$FID$
本文算法 RBMS-CGAN	1.00
无语义定位模块的本文算法	1.21
无特征损失 $L_f$ 的本文算法	1.02
ANAS	1.13
MUNIT	1.56
CoCosNet	1.42

4.2.2 特征损失函数消融测试

损失函数可以惩罚生成图像和原作像素点颜色以及语义的差异,提升上色仿真效果.生成模型常用  $L_1$  损失缩小生成图和原作的像素点差异,但是没有语义上的约束,容易导致色彩在空间上不连续,本节

对比了增加特征损失项  $L_f$  前后生成效果的差异,结果如图 12 所示.从图(12)可以看出,没有加特征损失项的生成图像在颜色构成上和参考图像类似,但是色彩分布不协调,如图 12(c)行 1 芙蓉花的花朵左上区域颜色较深,其他区域颜色较浅;图 12(c)行 2 蝴蝶兰有的花朵颜色厚重,有的花朵颜色轻薄,缺乏统一的色调,而本文算法通过加入语义损失,画面

整体颜色柔和统一;图 12(b)行 3 选取的参考图像中包括蓝色的牡丹以及白色的花骨朵儿,没有特征损失的上色模型把蓝色与白色都转移到花卉线稿图的花朵区域,导致蓝色花瓣中混杂了白色,通过加入特征损失,在语义上进行约束,生成蓝色色调的花朵,整体颜色协调.实验表明,加入特征损失后生成网络能更好地模拟真实工笔花卉的颜色分布.

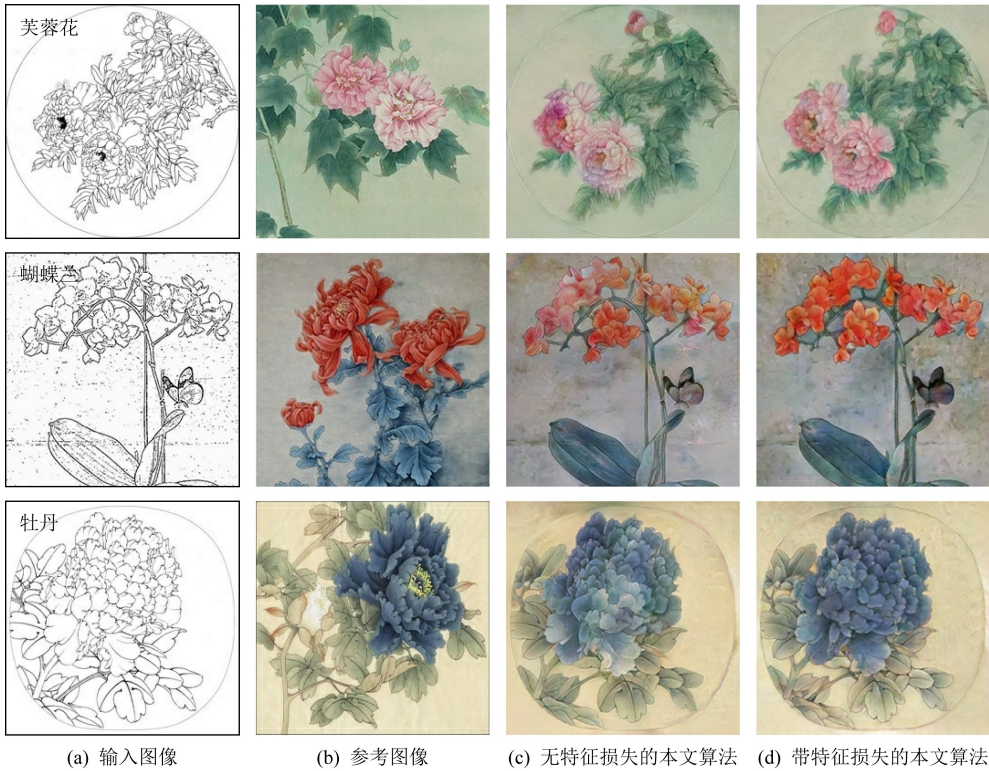


Fig. 12 Comparison of ablation results with or without feature loss

图 12 有无特征损失的消融实验结果对比

为了进一步评估特征损失的有效性,本文采用  $FID$  比较特征损失增加前后的生成效果,结果如表 2 所示.由表 2 可以看出,本文带有特征损失  $L_f$  的算法生成图像的  $FID$  低于未使用特征损失的生成图像.

4.2.3 摄动图训练方式消融实验

为了验证基于原作摄动图的训练方式对提升上色网络泛化能力的效果,本文在训练阶段分别用原作和原作摄动图作为对应花卉线稿图的参考图像,在测试阶段选取与花卉线稿图线稿在花型、布局上有较大差异的参考图像,对比 2 种方法训练的网上色仿真效果,结果如图 13 所示.以原作为参考图像训练的上色网络难以实现花卉线稿图自动上色,上色结果出现极端的情况,即输出结果是花卉线

稿图和参考图像的线性叠加.这种现象出现的原因是训练过程将原作作为参考图像,原作和花卉线稿图的线条结构相同,所以上色模型将原作的颜色按像素点的位置进行点对点传递,就能达到花卉线稿图上色目的,显然这种模型泛化性很差,在实际应用的时候,对于任意选取的一幅参考图像,通过点对点传递参考图像的颜色是无法完成花卉线稿图上色.对比本文训练阶段采用原作摄动图作为参考图像,在测试的时候即使采用与花卉线稿图线条结构差异较大的参考图像,算法也能较准确地将参考图像颜色定位到花卉线稿图的对应语义区域.

4.3 与其他算法的对比结果

ANAS<sup>[11]</sup>是一个基于深度卷积神经网络的艺术风格迁移模型,可将风格图像的颜色和纹理迁移



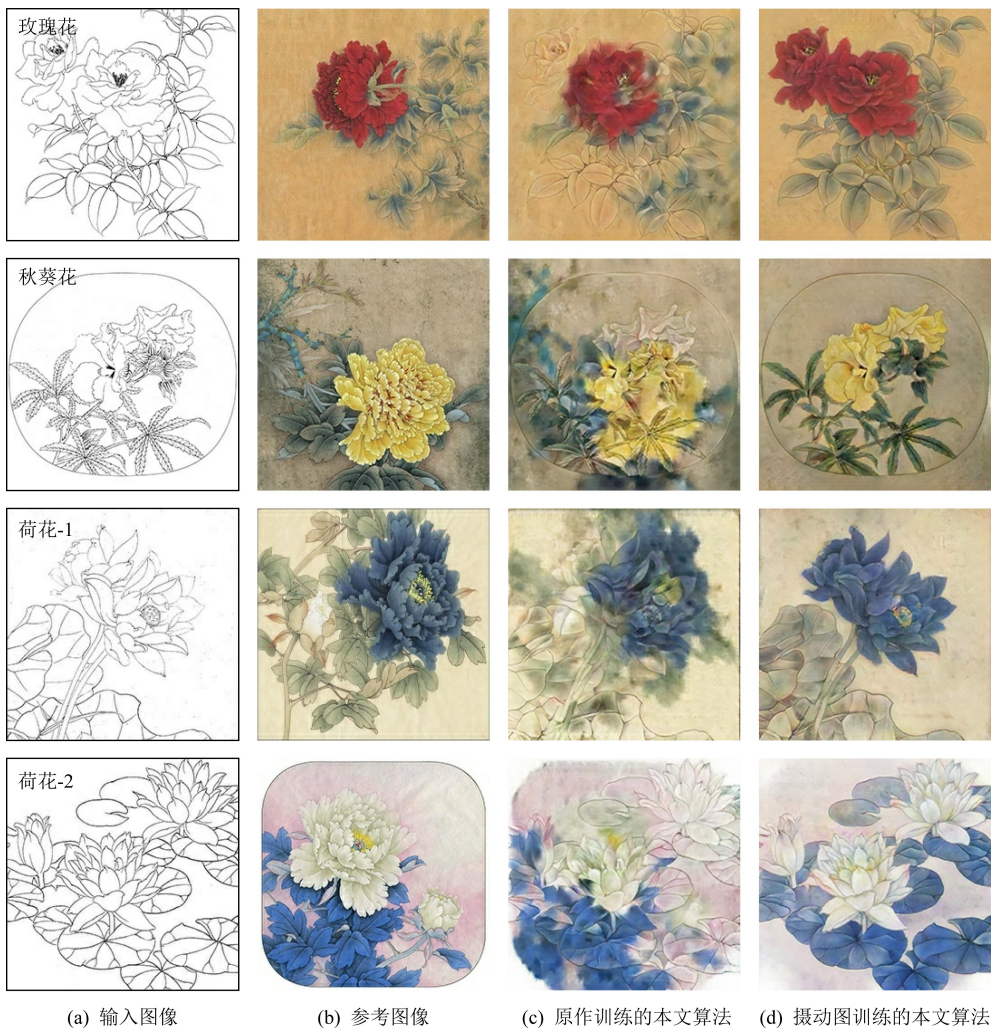


Fig. 13 Generation results of whether uses perturbed images in training stage

图 13 训练过程是否采用原作摄动图的生成结果

到内容图像;Huang 等人<sup>[12]</sup>的算法是通过对隐层空间解耦将参考图像的特征复用到内容图像实现上色的;Zhang 等人<sup>[13]</sup>提出的 CoCosNet 网络模型是基于参考图像的跨域图像转换任务的一个通用框架;本文利用“工笔花卉-花卉线稿图”数据集训练文献[11-13]所提的模型,比较不同模型的工笔花卉上色仿真效果,结果如图 14 所示.

本文选取了 5 种不同花型的线稿图线稿和 5 张参考图像,由图 14 可以看出,ANAS 在迁移参考图像的颜色信息上是可行的,生成结果也反映出参考图像的颜色构成,但是由于工笔花卉颜色典雅工丽,不属于风格强烈鲜明的艺术图像,因此风格迁移效果较差,且颜色定位不够准确.

如图 14(c)行 4 牡丹以及图 14(c)行 5 荷花,花朵区域缺乏颜色渲染,而花卉线稿图的留白区域出现了颜色色块;图 14(c)行 1 百合花、图 14(c)行 2

茶花、图 14(c)行 3 蝴蝶兰仅有部分区域色彩尚合理,大多数区域色彩杂乱,不符合工笔花卉画上色仿真的要求;Huang 等人设计的非成对样本对训练的网络 MUNIT 表现欠佳,偏向于对背景和线条着色;CoCosNet 能较好地模拟参考图像的颜色,但是语义区域定位的准确率较差,导致颜色不足.如图 14(e)行 2 茶花和图 14(e)行 5 荷花花朵区域,几乎没有色彩渲染;有的出现定位错误,如图 14(e)行 4 牡丹草绿色的叶子出现红色的色块,影响生成图像的艺术效果.本文算法能较好地将参考图像的色彩按照语义迁移到花卉线稿图,在颜色模拟上更接近真实工笔花卉效果.

为了更客观地比较本文算法和 ANAS<sup>[11]</sup>,MUNIT<sup>[12]</sup>,CoCosNet<sup>[13]</sup>模型的生成效果,本文采用 FID 进行量化分析,结果如表 2 所示.从表 2 可以看出,本文算法在工笔花卉上色仿真中表现更优.

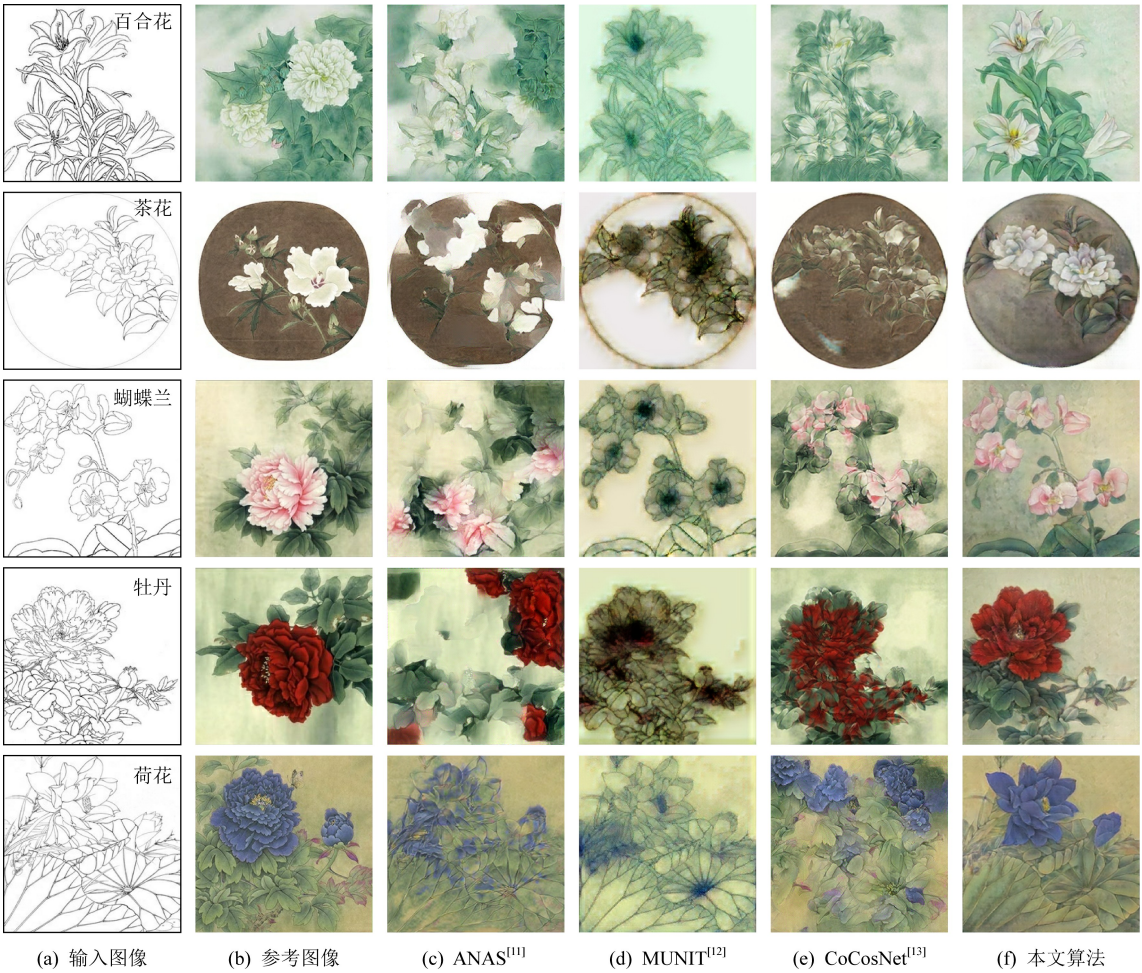


Fig. 14 Comparison of generation results by various algorithms

图 14 不同算法生成效果对比

5 结束语

本文提出的基于语义匹配的工笔花卉上色仿真算法 RBSM-CGAN,实现了花卉线稿图到工笔花卉色图的转换,通过预训练的语义分割网络划分花卉线稿图中的不同语义区域,指导参考图像的颜色定位到花卉线稿图的不同位置,同时,语义标签图隐式地引入了内部注意力机制,加强了不同区域的颜色学习,提升了渐变色的模拟效果.针对训练阶段采用“工笔花卉原作-花卉线稿图”可能导致上色模型过度依赖,本文提出对原作扭曲变形等几何变换得到摄动图像作为参考图像.实验表明,本文算法能较准确地将参考图像的颜色信息传递到任意花型的花卉线稿图,生成具有工笔画效果的图像.后续研究希望增强网络结构,生成特殊肌理的背景;对花蕊等细节部分进行处理,以进一步提升模型的工笔画仿真效

果;生成工笔花卉图的短视频<sup>[21]</sup>,让图动起来;同时会考虑将该算法拓展到其他线稿图上色任务中.

作者贡献声明:李媛提出了算法主要思路和实验方案,完成实验并撰写论文;陈昭炯和叶东毅提出研究的问题及部分算法方案,并就算法和实验过程遇到的问题给出指导意见和修改论文.

参 考 文 献

[1] Cao Yun, Zhou Zhiming, Zhang Weinan, et al. Unsupervised diverse colorization via generative adversarial networks [C] // Proc of the 28th European Conf on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2017: 151-166

[2] Messaoud S, Forsyth D, Schwing A G. Structural consistency and controllability for diverse colorization [C] // Proc of the 15th European Conf on Computer Vision. Berlin: Springer, 2018: 596-612



- [3] Sangkloy P, Lu Jingwan, Fang Chen, et al. Scribbler: Controlling deep image synthesis with sketch and color [C] // Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 5400-5409
- [4] Ci Yuanzheng, Ma Xinzhu, Wang Zhihui, et al. User-guided deep anime line art colorization with conditional adversarial networks [C] // Proc of the 26th ACM Int Conf on Multimedia. New York: ACM, 2018: 1536-1544
- [5] Zhang Lvmin, Li Chengze, Wong Tientsin, et al. Two-stage sketch colorization [J]. ACM Transactions on Graphics, 2018, 37(6): 1-14
- [6] Qiao Tingting, Zhang Jing, Xu Duanqing, et al. Mirrorgan: Learning text-to-image generation by redescription [C] // Proc of the 32nd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 1505-1514
- [7] Li Yuan, Chen Zhaojiong, Ye Dongyi. Interactive meticulous flower coloring algorithm via attention guidance [J]. Pattern Recognition and Artificial Intelligence, 2020, 33(7): 575-587 (in Chinese)  
(李媛, 陈昭炯, 叶东毅. 注意力引导的交互式工笔花卉上色算法[J]. 模式识别与人工智能, 2020, 33(7): 575-587)
- [8] Pang Jiahao, Au O C, Yamashita Y, et al. Self-similarity-based image colorization [C] // Proc of the 21st IEEE Int Conf on Image Processing (ICIP). Piscataway, NJ: IEEE, 2014: 4687-4691
- [9] Huang Xun, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization [C] // Proc of the 16th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 1501-1510
- [10] Lee H Y, Tseng H Y, Huang Jiabin, et al. Diverse image-to-image translation via disentangled representations [C] // Proc of the 15th European Conf on Computer Vision. Berlin: Springer, 2018: 35-51
- [11] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks [C] // Proc of the 29th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2414-2423
- [12] Huang Xun, Liu Mingyu, Belongie S, et al. Multimodal unsupervised image-to-image translation [C] // Proc of the 15th European Conf on Computer Vision. Berlin: Springer, 2018: 172-189
- [13] Zhang Pan, Zhang Bo, Chen Dong, et al. Cross-domain correspondence learning for exemplar-based image translation [C] // Proc of the 33rd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5143-5153
- [14] Lee J, Kim E, Lee Y, et al. Reference-based sketch image colorization using augmented-self reference and dense semantic correspondence [C] // Proc of the 33rd IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 5801-5810
- [15] Isola P, Zhu Junyan, Zhou Tinghui, et al. Image-to-image translation with conditional adversarial networks [C] // Proc of the 30th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 1125-1134
- [16] Deng Jia, Dong Wei, Socher R, et al. ImageNet: A large-scale hierarchical image database [C] // Proc of the 22nd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 248-255
- [17] Yoo D, Kim N, Park S, et al. Pixel-level domain transfer [C] // Proc of the 14th European Conf on Computer Vision. Berlin: Springer, 2016: 517-532
- [18] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation [C] // Proc of the 18th Int Conf on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241
- [19] Winnemöller H, Kyprianidis J E, Olsen S C. XDoG: An extended difference-of-Gaussians compendium including advanced image stylization [J]. Computers & Graphics, 2012, 36(6): 740-753
- [20] Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local Nash equilibrium [C] // Proc of the 31st Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 6626-6637
- [21] Yu Haitao, Yang Xiaoshan, Xu Changsheng. Antagonistic video generation method based on multimodal input [J]. Journal of Computer Research and Development, 2020, 57(7): 1522-1530 (in Chinese)  
(于海涛, 杨小汕, 徐常胜. 基于多模态输入的对抗式视频生成方法[J]. 计算机研究与发展, 2020, 57(7): 1522-1530)



**Li Yuan**, born in 1995. Master. Her main research interest is image processing.

李媛, 1995年生, 硕士. 主要研究方向为图像处理.



**Chen Zhaojiong**, born in 1964. Master, professor. Her main research interests include intelligent image processing and computational intelligence.

陈昭炯, 1964年生, 硕士, 教授. 主要研究方向为智能图像处理和计算智能.



**Ye Dongyi**, born in 1964. PhD, professor. His main research interests include computational intelligence and data mining.

叶东毅, 1964年生, 博士, 教授. 主要研究方向为计算智能和数据挖掘.