浙江理工大学学报,第51卷,第4期,2024年7月
Journal of Zhejiang Sci-Tech University
DOI:10.3969/j.issn.1673-3851(n).2024.04.013

基于美学梯度法的人工智能风格化绘画系统

钟梓锐,梁玲琳
(浙江理工大学艺术与设计学院,杭州 310018)

摘　要:当前Stable diffusion等人工智能绘画模型在绘画时难以直接控制图像风格,同时风格模型训练仅针对单种风格。针对该问题,提出了一种基于美学梯度法的人工智能风格化绘画系统,以实现多种图像风格的控制和融合,并提供更加便捷的图像创作体验。收集并分析网络用户数据,结合问卷得到用户对图像风格的感性需求;根据感性需求收集各风格图像数据得到对应的风格图像训练集。使用梯度下降算法计算风格化文本编码器的权重,实现生成图像风格化的效果。通过可用性测试对比用户对该系统与传统人工智能绘画系统产出图像的风格满意程度,结果表明:人工智能风格化绘画系统的平均满意度相较传统人工智能绘画系统提升23%,表明人工智能风格化绘画系统在图像风格生成上具有更好的效果,可满足用户对图像风格的需求。该人工智能风格化绘画系统可以更便捷地实现图像风格调整,允许用户直观选择不同风格的权重,便捷使用一种或多种风格,能够有效满足用户对图像风格设计的需求。

关键词:人工智能绘画模型;Stable diffusion;美学梯度法;感性需求;风格化
中图分类号:TP18　　**文献标志码**:A　　**文章编号**:1673-3851 (2024) 04-0537-11
引文格式:钟梓锐,梁玲琳. 基于美学梯度法的人工智能风格化绘画系统[J]. 浙江理工大学学报(自然科学),2024,51(4):537-547.
Reference Format: ZHONG Zirui, LIANG Linglin. An artificial intelligence stylized painting system based on the aesthetic gradient method[J]. Journal of Zhejiang Sci-Tech University,2024,51(4):537-547.

An artificial intelligence stylized painting system based on the aesthetic gradient method

ZHONG Zirui, LIANG Linglin
(School of Art and Design, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: At present it is difficult for artificial intelligence painting models such as Stable diffusion to directly control image style in painting. At the same time, current style model training is focused on a single style. To address this issue, an artificial intelligence stylized painting system based on the aesthetic gradient method was proposed. It aimed to achieve control and integration of multiple image styles, and to provide a more convenient image creation experience. It collected and analyzed network user data and employed a questionnaire-based approach to obtain the user's perceptual needs for image style. Furthermore, it collected the data of each style image according to the perceptual requirements to obtain the corresponding style image training set. It also used the gradient descent algorithm to calculate the weights of the stylized text encoder to achieve the effect of generating image stylization. A usability test was conducted to compare user satisfaction with the image styles produced by the traditional artificial intelligence painting system and the artificial intelligence stylized painting system. The results show that the average satisfaction of the latter is 23% higher than that of the former, indicating that artificial intelligence stylized painting system has better effects in image style generation and can effectively meet users' needs for image styles. This artificial intelligence stylized painting system can realize image style

收稿日期:2023-11-17　　网络出版日期:2024-05-10
基金项目:国家社会科学基金青年项目(22CXW024)
作者简介:钟梓锐(1999—),男,广州人,硕士研究生,主要从事人工智能绘画方面的研究。
通信作者:梁玲琳,E-mail:lianglinglin916@126.com

adjustment more easily, allow users to intuitively choose the weight of different styles and easily use one or more styles, and can effectively meet users′ needs for image style design.

Key words: artificial intelligence painting model; Stable diffusion; aesthetic gradient; emotional needs; stylization

0 引 言

Stable diffusion模型是2022年发布的深度学习文本生成图像模型,具有生成的图像质量高、运行速度快以及内存占用小的优点[1],在目标检测[2]、产品设计[3]和视频编辑[4]等场景具有广阔的应用前景。Stable diffusion模型分为两个部分,分别是Contrastive language-image pretraining(CILP)多模态预训练模型和U-Net。CLIP是一种预训练神经网络模型,通过对比学习将图像和文本联系起来[5]。U-Net是一种用于图像分割的卷积神经网络架构,由Ronneberger等[6]提出。U-Net的名称来源于其U形的网络结构,由编码器(Encoder)、解码器(Decoder)和跳跃连接(Skip connections)三部分组成。使用时,用户输入的每一个单词由CLIP中的分词器(Tokenizer)转换成文本标记(Token),每个文本标记是768维的向量。CLIP是预训练模型,因此每个文本标记的嵌入向量都是固定的,嵌入向量经过文本转换器后输入到U-Net中,完成最终的绘画[1]。用户对图像的风格进行控制时,需要在文本描述中加入风格形容词汇,如“抽象主义”和“卡通”等。然而,使用通过文本输入的方式控制图像风格的方式仍存在挑战,其中主要原因是输入的风格词汇的嵌入向量之间虽不相同但十分接近,缺乏直接控制图像的风格的方法[7]。

为更直接地指导图像生成的风格,Gal等[8]提出了文本反转方法,将用户提供的同一种风格的3~5张图像,变为嵌入空间中的一个新文本标记。这些文本标记可以组合成自然语言句子,以直观的方式引导个性化创作。Ruiz等[9]提出了一种“个性化”文本到图像扩散模型的方法,通过输入特定风格的少量图像,并对预训练的文本与图像模型进行微调,将文本标记与风格绑定。风格被嵌入到模型的输出域后,文本标记就可以用来在不同场景中合成既定风格的图像。Gallego[10]提出了美学梯度法(Aesthetic gradient),这是一种通过相同风格的图像数据集来个性化编辑CLIP的方法。该方法使用图像数据集训练得到美学嵌入,结合美学嵌入对CLIP中文本编码器的权重做梯度下降计算,得到风格调整后的文本嵌入。上述研究通过相同风格的图像向模型提供风格信息,模型无需解析自然语言描述,而是直接基于模型训练来调整生成的风格,可以减少模型理解文字描述不准确所带来的风格偏差。上述研究的应用场景均为单种图像风格的训练,且应用场景大多集中在单一图像风格的训练上,无法满足用户对多样风格或风格融合的需求。另外,在进行训练前需要确定常见用户风格需求,但以往研究多采用感性工学方法来分析用户需求。

随着互联网技术的发展,网络上存在着大量的用户数据,国内外研究人员通过收集、分析网络上的用户感性意象数据,得到用户的感性需求[11]。例如:Ma等[12]通过收集网络日志、搜索历史和交易数据等用户行为记录,构建了数据层、语义层和应用层的用户需求三层概念模型,通过该模型能将语义信息和用户需求进行匹配。Shi等[13]通过网络信息提取与感性形容词相关的产品关键特征,使用感性意象词问卷与语义差分法来评估产品的特征,生成描述关键特征和相应感性形容词之间关系的强关联规则集。Wang等[14]爬取产品的评论数据,基于自然语言处理技术构建词向量,实现感性图像的参数化表达;提取满足用户偏好的产品方案,量化产品形态与感性形象之间的关系,并根据各参数权重计算针对用户感性需求的产品设计方案的优先排序。除了通过网络用户数据分析用户需求的研究外,还有将用户感性需求数据用于指导图像生成的研究。如Li等[15]使用网络爬虫从现有的文献和网络评价中收集感性词,并通过语义聚类分析将收集到的意思相反的词聚类成簇,再根据感性词的数量和总频率之和对所有聚类进行排序,选择得分前6的感性聚类作为代表性的感性意象词汇;随后发放由产品图像、感性词汇和语义差分量表组成的问卷,并将得到的问卷数据用作图像生成神经网络的训练数据集。实验验证结果表明,训练后的神经网络生成的产品概念图像效果优秀。

现有人工智能绘画模型缺乏直接控制图像风格的方法,同时风格模型训练方法具有局限性。针对这一问题,本文提出了一种基于美学梯度法的人工智能风格化绘画系统。首先通过爬虫与自然语言处

理技术收集并分析大量网络用户评论数据,结合问卷量化用户对图像风格的感性需求;其次,根据感性需求收集各风格图像数据,得到对应的风格图像训练集;再次,使用美学梯度法训练得到各个风格的美学嵌入模型,然后通过对人工智能模型的文本编码器做多重梯度下降计算,使 CLIP 能够输出适应多个风格特征的文本嵌入,实现对最终输出的图像的风格的控制;最后进行该系统的可用性测试,对比用户对传统人工智能绘画系统与对人工智能风格化绘画系统产出的图像风格满意程度,以验证风格化人工智能绘画系统在图像风格生成上是否具有更好的效果。

1　系统设计

1.1　系统组成

人工智能风格化绘画系统由 4 个部分组成,分别是文本编码器、风格模块、图像信息生成器和图像解码器。以生成 512×512 像素的图像为例,系统的整体流程如图 1 所示,分为以下步骤:首先,文本编码器将用户输入的文本描述转化为一个向量特征,通常表示为 c。其次,风格模块基于用户输入的风格参数将向量特征转化为适应用户输入的风格的向量特征。再次,图像信息生成器接收这个向量特征,并将其转换为一个信息数组,其维度为 4×64×64 维。这个数组包含了关于图像内容和风格的关键信息。最后,图像解码器将这个信息数组解码并渲染成最终的图像,其维度为 3×512×512 维,其中 3 表示颜色通道数(红、绿、蓝),512×512 表示图像的宽度和高度。

该系统各个组成部分相互配合,实现了从用户的文本输入到最终图像输出的完整生成过程。用户输入文本描述,决定了图像的内容。用户输入与 5 种画面风格相关的参数。这些参数为正负整数,决定了风格的倾向;参数的大小决定了美学梯度法中梯度下降的步长 ε。用户输入所需的图像长宽后,图像信息生成器根据该长宽生成的信息数组大小,决定最终输出图像的像素大小。用户输入的参数和文本嵌入将决定系统产生的图像结果。

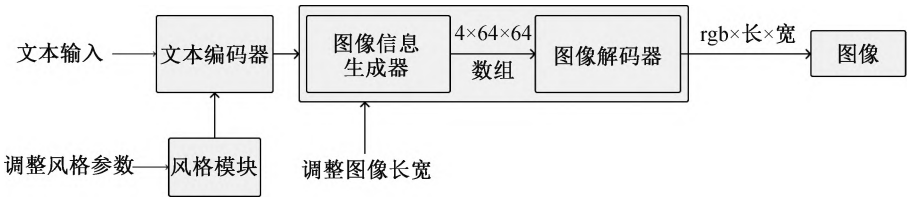


图 1　绘画系统的整体流程示意图

系统设计与用户体验流程示意图如图 2 所示,由感性词汇收集、风格嵌入模型训练和人工智能绘画 3 部分组成。第 1 部分,通过收集并分析网络数据获取用户感性需求并总结,得到 5 组代表性感性意向词汇:古代的-未来的、西方的-东方的、画面明亮的-画面暗淡的、写实的-动漫的、复杂的-简单的。收集图像样本,并根据这 5 组词汇对图像样本集进行分类和贴标,得到各风格训练图像集,保证用户的风格选择的多样性和全面性。第 2 部分,使用图像训练集进行美学嵌入模型训练,根据美学嵌入模型计算得出风格化文本编码器的权重并应用于绘画,最终确保系统产出图像的风格与用户需求具有一致性。第 3 部分,用户通过操作界面输入文本描述和图像风格参数,经过风格化文本编码器的处理后经过图像信息生成器与图像解码器,最终生成图像。

1.2　网络评论数据的收集与处理

1.2.1　初步图像风格形容词收集

与使用问卷获得用户感性意向的方法相比,通过网络评论获取的数据具有量大、快捷、时效性与客观性强等优点[16]。本文采用网络爬虫对微博、贴吧的相关话题与讨论进行爬取。使用网络爬虫爬取用户微博内容,搜索“AI 绘画”“图像风格”“画面风格”等相关词语,返回微博 ID、用户名、用户 ID 和文本等数据;使用 Python selenium 库对 Midjourney、AI 绘画等相关贴吧的帖子进行爬取,返回主帖内容与评论内容,每周重复爬取并去除重复数据。共收集人工智能绘画与图像风格相关的微博、帖子与评论 4434 条。由于初步爬取的内容有大量不连续的干扰信息[17],对初步数据进行数据清理,去除用户名等数据,保留主要内容文本;使用 Python jieba 库对所有内容文本进行自然语言处理,将段落句子切割成词汇并进行统计;将切割错误的词语如“波普艺术”和“赛博朋克”等加入到 jieba 中文词汇语库中,重新进行切割;去除与图像风格无关的词汇,得到形容词词汇 94 个,并合并意义相近的词汇,如“二次元”和“动漫”、“国风”和“汉服”等;对结果进行排序,得到初步的图像风格感性意象形容词 64 个。初步图像风格感性形容词词频见表 1。

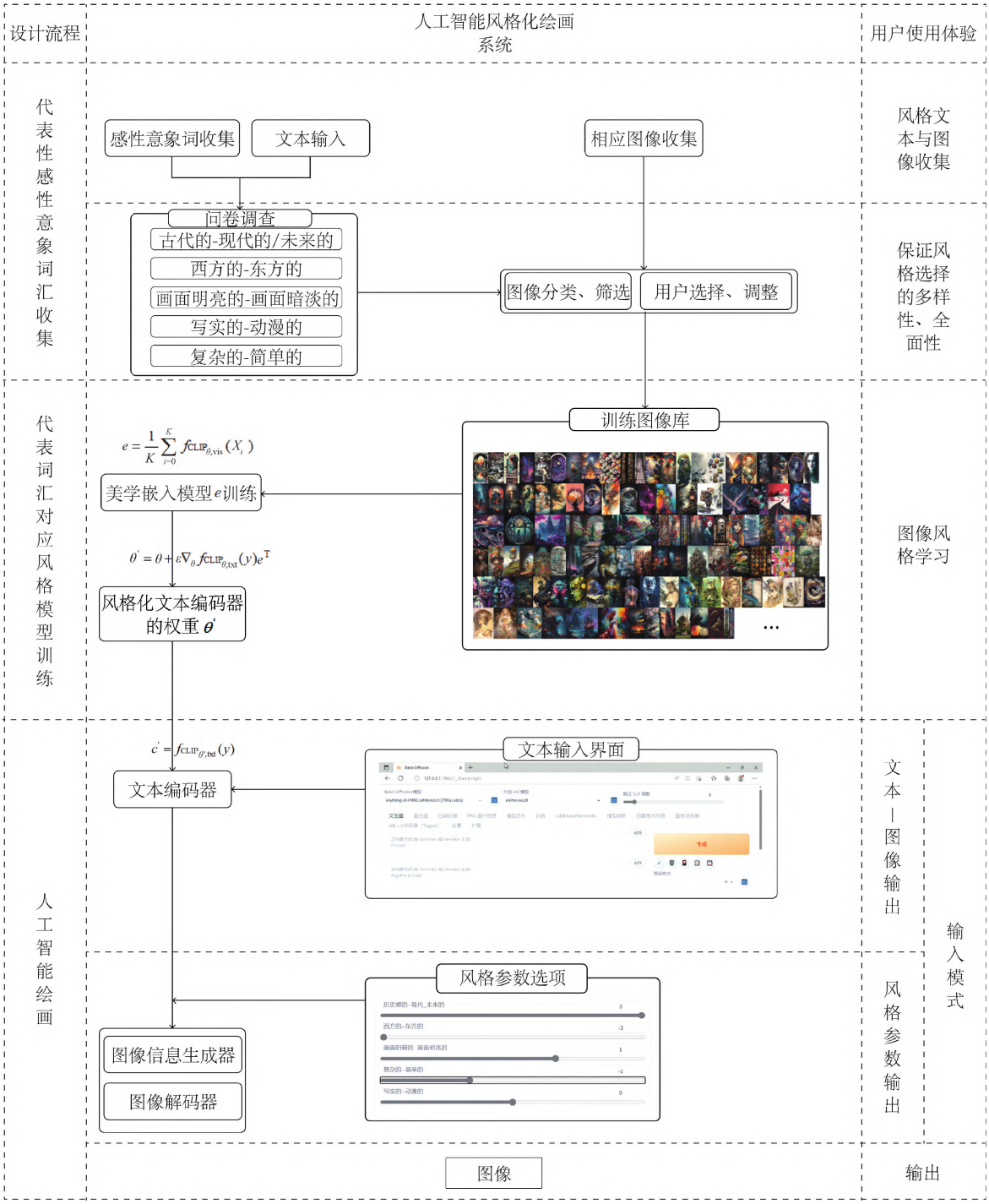


图 2 系统设计与用户体验流程示意图

为保证提取词汇能够准确代表用户的真实诉求,需要对用户评论数据的分词结果和词频排序进行筛选[16]。根据词频排序选择代表性词汇时,高频词阈值的选取决定了词频分析法的结果,对整个分析研究有着重要的影响[18],因此本文选取 Donohue[19]提出的高频词低频词分界公式对感性意象形容词进行筛选,该公式可以表示为:

$$T=\frac{1}{2}\times(-1+\sqrt{1+8\times I_1}) \tag{1}$$

其中:T 标识词频阈值,I_1 表示出现 1 次的词汇数量。对初步图像风格感性意象形容词进行词频统计,根据高低频词界分公式,统计分词后 $I_1=64$,因此 $T\approx 10.81$,阈值为 11,可以得到高频形容词 54 个。

1.2.2 图像风格感性意象形容词问卷调研

为进一步确定用户需求感性词汇,使用调研问卷的方法对感性词汇作进一步筛选。为了获取更加

精准有效的信息,让问卷填写人更加清晰地理解各个图像风格的意义,需在调研问卷中对初步收集的感性意象代表词作出解释。本文基于表 1 的初步图像风格感性形容词,收集人工智能绘画案例,并找到与研究初步得到的 54 个感性词汇相对应的示例图像,用于制作调查问卷,调查问卷示例如图 3。

表 1　初步图像风格感性形容词词频表

代表性词汇	词频/次	代表性词汇	词频/次	代表性词汇	词频/次	代表性词汇	词频/次
画面明亮的	97	文艺复兴风格	18	渐变的	3	通感的	6
科幻的	64	赛博朋克风格	130	和谐的	15	平面设计的	17
未来的	70	非凡的	15	平静的	26	吸引人的	81
动漫的	2157	吸引的	12	性感的	12	统计图的	46
写实的	2641	精致的	56	复杂的	44	极繁主义的	14
东方的	267	波普艺术风格	128	奢侈的	13	幻觉的	6
画面暗淡的	90	色彩缤纷的	30	浮世绘风格	11	地图风格	13
西方的	34	平面的	9	优雅的	32	极简主义的	9
详细的	23	迷惑的	100	骇人的	12	30 年代的	134
简单的	45	有气质的	11	嬉皮士风格	25	20 年代的	51
现实的	15	荒谬的	28	嘈杂的	5	成熟的	102
现代主义的	54	梦幻的	12	浪漫的	8	混乱的	12
哥特风格	11	华丽的	74	神圣的	11	80 年代的	209
夸张的	53	杂乱的	12	无明显风格	14	包豪斯风格的	4
美丽的	486	优雅的	32	奢侈的	5	90 年代的	201
蒸汽朋克	89	怪异的	23	混乱的	9	梦境的	12

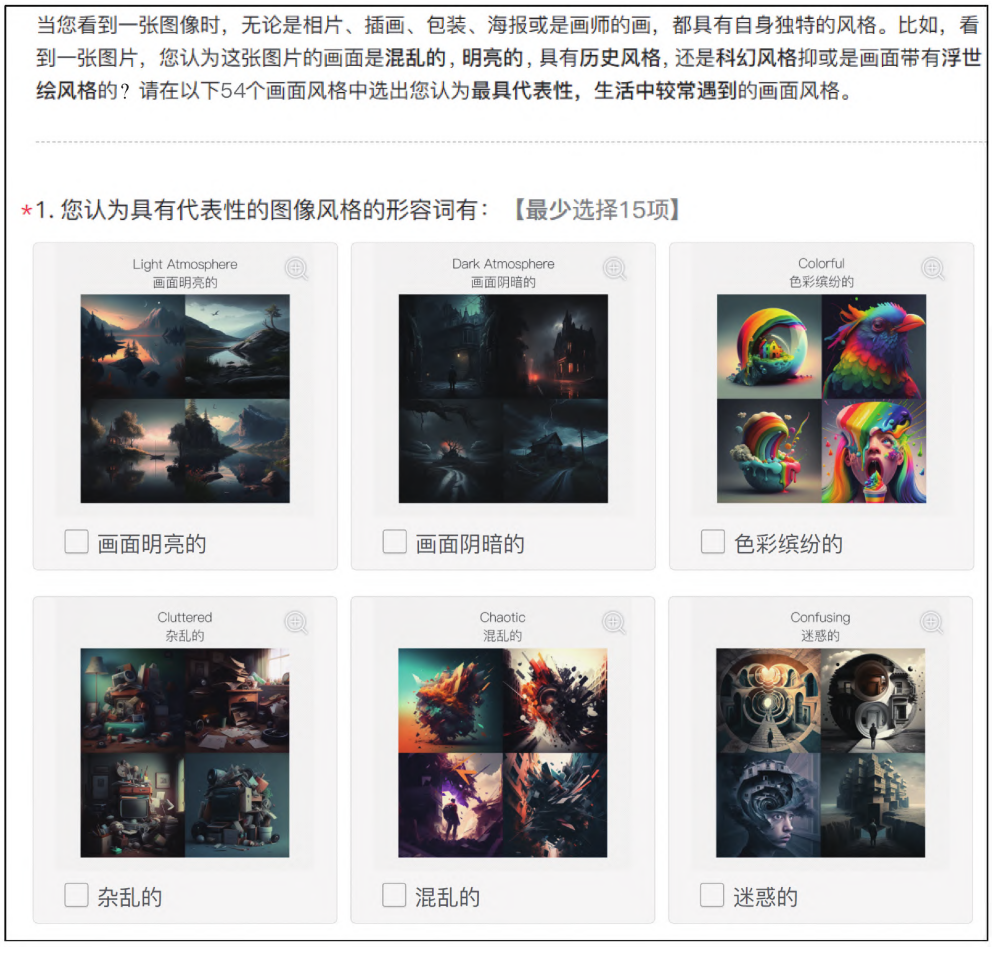


图 3　代表性感性风格形容词调查问卷示例

发放“代表性图像风格形容词调查问卷”,填写人选择其认为最具代表性的和最为常见的图像风格形容词,最终收回有效问卷101份。统计并分析结果,删除意义相反的词语,得到最终的代表性词汇选取频率排序,见表2。最终选取频率排名前5的代表性感性形容词,并将其进行反义词配对,最终得到代表性图像风格感性意象形容词对5对,即古代的-现代的/未来的、西方的-东方的、画面明亮的-画面暗淡的、写实的-动漫的、复杂的-简单的,用于收集训练图像与图像风格训练。

表2 代表性词汇及其频率

代表性词汇	词频/次	代表性词汇	词频/次	代表性词汇	词频/次	代表性词汇	词频/次
画面明亮的	75	文艺复兴风格	32	渐变的	27	通感的	21
科幻的	74	赛博朋克风格	32	和谐的	26	平面设计的	20
未来的	70	非凡的	31	平静的	26	吸引人的	20
动漫的	68	吸引的	31	性感的	25	统计图的	20
写实的	67	精致的	31	复杂的	25	极繁主义的	19
东方的	67	波普艺术风格	31	奢侈的	25	幻觉的	19
画面暗淡的	64	色彩缤纷的	30	浮世绘风格	24	地图风格	18
西方的	63	平面的	28	优雅的	24	极简主义的	17
详细的	59	迷惑的	28	骇人的	24	30年代的	17
简单的	57	有气质的	28	嬉皮士风格	24	20年代的	17
现实的	56	荒谬的	28	嘈杂的	24	成熟的	17
现代主义的	54	梦幻的	28	浪漫的	22	混乱的	15
哥特风格	41	华丽的	27	神圣的	22	80年代的	15
夸张的	38	杂乱的	27				

1.3 人工智能绘画样本收集

代表性图像风格形容词对体现了用户对图像风格的感性需求,为完成从感性需求到产出图像的转化,需要对模型进行各风格形容词相对应的训练。为获得相对应风格的训练数据,在Discord的Stable diffusion频道上与各绘画网站上收集人工智能绘画的图像样本,去除掉分辨率低以及图像长宽比例失衡的样本,得到部分初步样本1495个。最后对样本贴标得到各个风格的训练集,其中“画面暗淡的”风格的图像训练集示例图像如图4所示。

图4 “画面暗淡的”画面风格训练集示例图像

1.4 训练风格嵌入模型

1.4.1 美学梯度法

美学梯度方法通过来自一组相同风格的图像数据集的自定义美学模型来个性化编辑CLIP,将图像生成过程交给用户。该方法支持对单个美学风格进行训练并应用于图像生成过程中。该方法原理如下:

绘画时,Stable diffusion模型通过基于CLIP的文本编码器将用户的文本输入转化为一个文本嵌入,用公式可以表示为:

$$c = f_{\mathrm{CLIP}_{\theta,\mathrm{txt}}}(y) \tag{2}$$

其中:c为嵌入模型;y为用户输入的文本;θ为文本编码器的权重。

使用美学梯度法时,通过对图像训练集进行训练得到e,e为美学嵌入模型,用公式可以表示为:

$$e = \frac{1}{K}\sum_{i=0}^{K} f_{\mathrm{CLIP}_{\theta,\mathrm{vis}}}(X_i) \tag{3}$$

其中:X_i 为美学风格的训练集中的第 i 个元素;K 为该集合中图像的数量。最后对文本编码器的权重做梯度下降计算,更新文本编码器的权重 θ',用公式可以表示为:

$$\theta' = \theta + \varepsilon \nabla_\theta f_{\mathrm{CLIP}_{\theta,\mathrm{txt}}}(y) e^{\mathrm{T}} \tag{4}$$

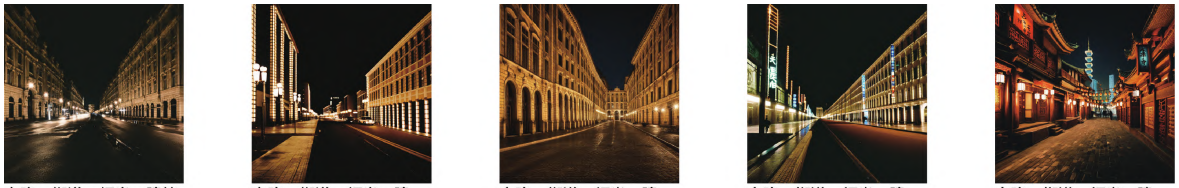
其中:ε 为用户定义的步长。最后将 θ' 应用到文本编码器得到经过美学风格调整后的文本嵌入 c',用公式可以表示为:

$$c' = f_{\mathrm{CLIP}_{\theta',\mathrm{txt}}}(y) \tag{5}$$

由于感性风格形容词有 5 对,需要对美学梯度方法进行改良,对文本编码器的权重做多重梯度下降计算[20]。

1.4.2　风格嵌入模型效果

使用原人工智能绘画模型,将画面风格作为描述关键词输入绘画;使用对应风格模型的人工智能绘画模型绘画。对比二者产出图像,各风格对比与各风格融合示例图像如图 5 所示。图 5(a)中:第 1 张图像基本的内容描述产出图像,输入为"夜晚,街道,灯光,建筑";第 2 张图像为加入了带权重风格描述产出图像,输入为"夜晚,街道,灯光,建筑,西方 * 1.5",其中" * 1.5"表示该描述的权重为 1.5,默认为 1;第 3 张图像为基本的内容描述加风格模型产出图像,输入为"夜晚,街道,灯光,建筑"加西方的风格模型。图 5(b)—(c)中其余内容描述与输入同图 5(a)。图 5(d)为各种风格融合生成图像与原图像的对比。用户只需要若干张相似风格的图像就可以根据需要训练自己的风格模型,并不局限于本文设定的 10 种风格。



(a) "西方的"和"东方的"风格

(b) "现代的/未来的"和"古代的"风格

(c) "写实的"和"动漫的"风格

(d) 多种风格融合

图 5　模型最终效果各风格对比与各风格融合示例图像

2 可用性测试

本文为分析人工智能风格化绘画系统的可用性,设计了测试实验,因变量为人工智能风格化绘画系统可用性,包括效率、满意度和有效性三个方面。针对用户使用人工智能绘画系统的过程设计具体的操作任务,共设置两组任务,分别是使用传统人工智能绘画系统进行绘画的对照组任务,以及使用人工智能风格化绘画系统进行绘画的实验组任务。完成两组任务后,测试者填写SUS问卷与满意度问卷。

2.1 测试材料

2.1.1 控制测试变量

为控制测试变量,需对风格嵌入模型进行测试,获得测试最佳美学风格迭代步数。使用同一文本描述与同一随机种子对10个感性风格进行测试,获得每个感性风格在不同美学风格权重数值与不同美学风格迭代步数下的结果。设置美学风格权重为0.8,测试不同美学风格迭代步数对最终产出图像效果的影响。

如图6(a)所示,在美学风格权重为0.8时,使用文本描述为“cat”,风格选择为“东方的”,美学风格迭代步数为5时,产出图像与东方风格较为契合;如图6(b)所示,而当美学风格迭代步数为10时,产出图像结果与需求描述偏离。图像内容变为东方风格的混乱的建筑,原因是训练图像内容上彼此差异较大,部分训练图像为东方风格的建筑而部分训练内容为东方风格的人像。如图6(c)所示,在美学风格权重为0.8时,使用文本描述为“cat”,风格选择为“复杂的”,美学风格迭代步数为6时,产出图像较为复杂;如图6(d)所示,当美学风格迭代步数为8时,画面与描述有一定关系;如图6(e)所示,当美学风格迭代步数为10时,产出图像结果与需求描述完全偏离。

“cat”,“东方的”,迭代步数:5。

(a) 迭代步数为5的“东方的”风格

“cat”,“东方的”,迭代步数:10。

(b) 迭代步数为10的“东方的”风格

“cat”,“复杂的”,迭代步数:6。

(c) 迭代步数为6的“复杂的”风格

“cat”,“复杂的”,迭代步数:8。

(d) 迭代步数为8的“复杂的”风格

“cat”,“复杂的”,迭代步数:10。

(e) 迭代步数为10的“复杂的”风格

图6 “东方的”“复杂的”风格不同迭代步数测试结果示例图像

测试使用的美学嵌入模型为用户使用时选择的风格,使用的美学风格权重为所选择风格的最佳权重。使用的美学风格迭代步数为用户所选择的风格参数相对应的迭代步数。经过测试得到每个风格对应的最佳权重与最佳迭代步数见表3。

2.1.2 测试设置

本文的实验材料为由开源的Stable diffusion webui改进得到的带有风格感性需求参数调整功能的人工智能风格化绘画模型。由于设备性能限制,将人工智能绘画风格化模型部署在腾讯云服务器,方便测试者通过公网IP访问。随机招募浙江理工大学、浙江大学和北京师范大学-香港浸会大学联合国际学院的在校本科生与研究生,测试者年龄为20~25岁,共28人。

表 3　各图像风格最佳权重与最佳迭代步数表

风格名称	最佳权重	最佳迭代步数/步	风格名称	最佳权重	最佳迭代步数/步
画面明亮的	0.80	8	画面暗淡的	0.55	6
古代的	0.40	6	现代的/未来的	0.60	6
简单的	0.80	7	复杂的	0.40	8
西方的	0.55	6	东方的	0.80	6
写实的	0.70	7	动漫的	0.80	8

2.2　测试流程

本实验开始前,研究人员向测试者介绍实验的基本内容,辅助测试者观看实验流程讲解视频。本文实验需要测试者完成一个对照组任务和一个实验组任务。

对照组任务的具体操作为:a)想象并确定需要绘制的画面;b)向研究人员使用自然语言描述需求画面,用于后续打分;c)在研究人员的帮助下将自然语言描述转化为 AI 模型的正向提示词与反向提示词,用于控制图像生成的内容和主体;d)调整被绘制图像的宽度和高度,用于控制图像的分辨率;e)点击生成按钮,等待 AI 模型进行运算与绘制;f)根据最终的人工智能绘画结果与需求画面对比,并对绘画结果内容满意度打分。由于设备性能限制原因,图像分辨率被限制在 512×512 像素,迭代步数为 20 步,绘制运算总时间大约在 5 min 左右。

实验组任务的具体操作为:a)想象并确定需要绘制的画面;b)向研究人员使用自然语言描述需求画面,用于后续打分;c)在研究人员的帮助下将自然语言描述转化为 AI 模型的正向提示词与反向提示词,用于控制图像生成的内容和主体,与对照组任务保持一致;d)调整被绘制图像的宽度和高度,用于控制图像的分辨率,保持与对照组任务一致;e)点击生成按钮,等待 AI 模型进行运算与绘制;f)尝试调整风格形容词参数,用于控制图像的整体风格;g)根据最终的人工智能绘画结果与需求画面对比,并对绘画结果风格满意度打分。由于设备性能限制原因,图像分辨率被限制在 512×512 像素大小,迭代步数为 20 步,绘制运算总时间大约在 5 min 左右。

两个任务完成后,测试者填写 SUS 系统可用性量表与图像内容风格满意度调查问卷,填写完成后试验结束,具体实验及任务流程见图 7。

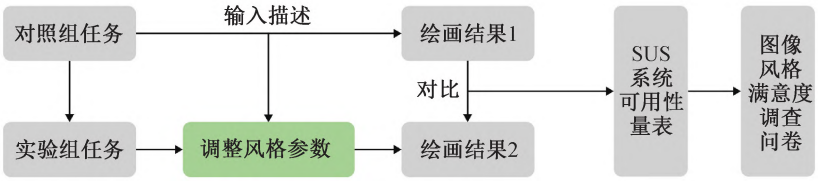


图 7　具体实验及任务流程图

2.3　测试结果

本文采用 SUS 系统可用性量表获取系统可用性,并采用图像内容风格满意度调查问卷获取用户对人工智能绘画系统的主观满意度。SUS 量表为五级的 Likert 量表,共有 10 个对系统的态度问题,其中:1、3、5、7、9 为正向问题,2、4、6、8、10 为反向问题。用户完成任务 1 和任务 2 后分别选择对语句的认可程度。量表第 4 题和第 10 题测量了系统的易于学习性,其余 8 题测量了系统可用性,最后由整体的 SUS 分数反映了总体的满意度[21]。

对任务 1 和任务 2 的两项满意度进行 Cronbach α 信度分析,结果如表 4 所示。结果显示:风格满意度数据信度系数值为 0.826,大于 0.8,表明研究数据信度质量高;两个任务的校正项总计相关性均大于 0.4,表明两个任务之间的满意度具有良好的相关关系。内容满意度数据信度系数值为 0.953,大于 0.8,表明研究数据信度质量高;两个任务的校正项总计相关性均大于 0.4,表明两个任务之间的满意度具有良好的相关关系。

表 4　Cronbach 信度参数

项名	校正项总计相关性(CITC)	Cronbach α 系数
任务 1 风格满意度 任务 2 风格满意度	0.726	0.826
任务 1 内容满意度 任务 2 内容满意度	0.925	0.953

测试结束后将测试者填写的量表选项通过计算转换为分数,转换后得到最终 SUS 得分,见表 5。最终得到无风格化模型版本 SUS 平均分为 73.750,易学性平均分为 65.189,可用性平均分为 75.893;风格化模型版本 SUS 的平均分为 67.250,

易学性平均分为51.333,可用性平均分为69.214。无风格化模型版本SUS得分的方差为135.491,易学性得分的方差为584.343,可用性得分的方差为112.205;风格化模型版本SUS得分的方差为78.205,易学性得分的方差为139.198,可用性得分的方差为67.485。

表5 SUS量表分数统计结果

问卷指标	任务	平均值	方差
SUS得分	任务1	73.750	135.491
	任务2	67.250	78.205
易学性得分	任务1	65.189	584.343
	任务2	51.333	139.198
可用性得分	任务1	75.893	112.205
	任务2	69.214	67.485

本文采用图像内容风格满意度调查问卷获取用户对人工智能绘画效果的主观满意度,量表指标为用户在两次任务产出图像中分别对图像内容与图像风格的满意程度,以及任务2中对5对感性形容词所对应的图像风格效果的主观评价;量表为五级的Likert量表。由被填写的量表得到测试者不同任务中对图像内容与风格主观满意度对比表,见表6。任务1图像内容满意程度平均分3.429,标准差为1.387;任务2图像内容满意程度平均分3.357,标准差为1.172。而任务1图像内风格满意程度平均分3.643,标准差为0.718,任务2图像风格满意程度平均分4.214,标准差为0.558。

表6 不同任务主观满意度对比结果

问卷指标	任务	平均值	标准差
对图像内容的满意程度	任务1	3.429	1.387
	任务2	3.357	1.172
对图像风格的满意程度	任务1	3.643	0.718
	任务2	4.214	0.558

利用配对t检验去研究实验结果的差异性,配对t检验结果如表7所示。由表7可以知,两组配对数据均呈现出差异性($P<0.05$),任务1风格满意度和任务2风格满意度之间呈现出0.01水平的显著差异性($t=-6.000$,$P=0.000$)。

利用配对t检验去研究实验结果的差异性,配对t检验结果如表8所示。由表8可以知,两组配对数据均没有呈现出差异性($P>0.05$)。

表7 风格满意度配对t分析检验结果

配对(平均值±标准差)		差值	t	P
任务1风格满意度	任务2风格满意度			
3.64±0.73	4.21±0.57	−0.57	−6.000	0.000**

注:"**"代表$P<0.001$。

表8 内容满意度配对t分析检验结果

配对(平均值±标准差)		差值	t	P
任务1内容满意度	任务2内容满意度			
3.43±1.00	3.36±1.19	0.07	0.812	0.424

由于数据结果不符合严格正态分布特质,配对t检验准确性下降。对数据结果进一步做非参数检验。由于满意度分数超过2组,使用Kruskal-Wallis检验,分析结果如表9所示。分析结果表明,不同任务1内容满意度样本对于任务2内容满意度全部均呈现出显著性差异,不同任务1风格满意度样本对于任务2风格满意度全部均呈现出显著性差异。对比差异可知,任务1内容满意度的平均值(3.43),高于任务2内容满意度的平均值(3.36)。测试者对风格化的图像内容的满意程度相较于原版的图像内容降低,表明使用风格化模型后的模型图像风格对用户的内容需求符合度稍有降低。而任务1风格满意度的平均值(3.64),明显低于任务2风格满意度的平均值(4.21)。测试者对风格化的图像风格的满意程度相较于原版的图像内容有所提升,表明使用风格化模型后更加符合用户风格需求。

表9 非参数分析检验结果

项名	Kruskal-Wallis检验统计量H值	P
风格满意度	15.016	0.001**
内容满意度	24.396	0.000**

注:"**"表示$P<0.001$。

3 结 论

本文提出了一种基于美学梯度法的人工智能风格化绘画系统,该系统通过训练美学风格模型可以满足用户多样化的风格图像需求。针对风格模型训练均针对单种风格的问题,该系统通过收集网络用户数据,量化用户对图像风格的感性需求;收集图像数据并贴标,并通过该数据训练出符合用户感性需求的多个风格嵌入模型。同时该系统使用多重梯度

下降算法与美学梯度法相结合,实现了融合多种风格的效果。系统可用性测试结果表明,相较于传统的人工智能绘画系统,人工智能风格化绘画系统在用户风格满意度上得到了提升,但在满足用户图像内容需求方面的效果下降。

本文量化用户感性需求用于人工智能绘画的风格训练,优化美学梯度法,使其能够同时使用多个风格嵌入模型,为满足用户个性化需求和提升图像生成质量提供了创新的思路。但由于在训练过程中图像样本分类由研究者自行完成,导致训练集具有较强的主观性,从而降低了后续实验的客观性。此外,研究中使用了 5 种风格融合,但在应用 5 种以上风格时效果开始明显下降。如何保证实验的客观性,有待进一步研究。

参考文献:

[1] Ramesh A, Dhariwal P, Nichol A, et al. Hierarchical text-conditional image generation with CLIP latents[EB/OL]. (2022-04-13)[2023-03-06]. https://arxiv.org/abs/2204.06125.

[2] Jian Y N, Yu F X, Singh S, et al. Stable diffusion for aerial object detection[EB/OL]. (2023-11-21)[2023-11-30]. https://arxiv.org/abs/2311.12345.

[3] Kuang Z Y, Zhang J X, Huang Y Y, et al. Advancing urban renewal: an automated approach to generating historical arcade facades with stable diffusion models [EB/OL]. (2023-11-20)[2023-11-30]. https://arxiv.org/abs/2204.06125.

[4] Chang D, Shi Y, Gao Q, et al. MagicDance: Realistic human dance video generation with motions & facial expressions transfer[EB/OL]. (2023-11-18)[2023-11-30]. https://arxiv.org/abs/2311.12052.

[5] Luo H S, Ji L, Zhong M, et al. CLIP4Clip: An empirical study of CLIP for end to end video clip retrieval and captioning[J]. Neurocomputing, 2022, 508(C): 293-304.

[6] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation [C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.

[7] Borji A. Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2 [EB/OL]. (2023-6-5)[2023-11-30]. https://arxiv.org/abs/2208.01618.

[8] Gal R, Alaluf Y, Atzmon Y, et al. An image is worth one word: Personalizing text-to-image generation using textual inversion[EB/OL]. (2023-8-2)[2023-11-30]. https://arxiv.org/abs/2210.00586.

[9] Ruiz N, Li Y Z, Jampani V, et al. DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation [C] // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada. IEEE, 2023: 22500-22510.

[10] Gallego V. Personalizing text-to-image generation via aesthetic gradients[EB/OL]. (2023-9-25)[2023-11-30]. https://arxiv.org/abs/2209.12330.

[11] 丁满,程语,黄晓光,等.感性工学设计方法研究现状与进展[J].机械设计,2020,37(1):121-127.

[12] Ma F C, Chen Y, Zhao Y M. Research on the organization of user needs information in the big data environment[J]. The Electronic Library, 2017, 35(1): 36-49.

[13] Shi F Q, Sun S Q, Xu J. Employing rough sets and association rule mining in KANSEI knowledge extraction[J]. Information Sciences: an International Journal, 2012, 196: 118-128.

[14] Wang T X. A novel approach of integrating natural language processing techniques with fuzzy TOPSIS for product evaluation[J]. Symmetry, 2022, 14(1): 120.

[15] Li X, Su J N, Zhang Z P, et al. Product innovation concept generation based on deep learning and Kansei engineering[J]. Journal of Engineering Design, 2021, 32(10): 559-589.

[16] 江亚红,许占民,董鑫.基于网络评论的产品感性设计研究[J].包装工程,2023, 44(S1):285-291.

[17] 林丽,张云鹃,牛亚峰, 等.基于网络评价数据的产品感性意象无偏差设计方法[J].东南大学学报(自然科学版),2020, 50(1): 26-32.

[18] 刘奕杉,王玉琳,李明鑫.词频分析法中高频词阈值界定方法适用性的实证分析[J].数字图书馆论坛, 2017(9): 42-49.

[19] Donohue J C. Understanding Scientific literatures: A Bibliometric Approach [M]. Cambridge: The MIT Press,1973:49-50.

[20] Sener O, Koltun V. Multi-task learning as multi-objective optimization[C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada. ACM, 2018: 525-536.

[21] Brooke J. SUS: A Quick and Dirty Usability Scale [M]. London: Taylor & Francis Ltd, 1996: 20-23.

(责任编辑:康　锋)