

Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
E-mail: jzus@zju.edu.cn



SmartPaint: 基于生成式对抗神经网络的人机协同绘画系统*

孙凌云^{†1,2,3}, 陈培^{†1,2,3}, 向为^{††1,2,3}, 陈鹏^{1,2,3}, 高晔玥^{1,2,3}, 张克俊^{1,3}

¹浙江省设计智能与数字创意研究重点实验室, 中国杭州, 310027

²浙江大学计算机辅助设计与图形学国家重点实验室, 中国杭州, 310058

³阿里巴巴-浙江大学前沿技术联合研究中心, 中国杭州, 311100

[†]E-mail: sunly@zju.edu.cn; chenpei@zju.edu.cn; wxiang@zju.edu.cn

投稿2019-7-30; 录用2019-12-8; crosscheck 2019-12-24

摘要:当前人工智能在模仿和大批量生产设计作品中扮演重要角色(如电商广告),而在与用户合作创作时表现欠佳。人们有能力使用草图表达创意想法,但缺乏专业绘画技巧完成精美画作。已有的人工智能方法无法基于用户输入草图的语义输出具有艺术美感的画作。本文开发了一种基于生成式对抗神经网络的人机协作绘画系统SmartPaint,支持人机合作创作动漫风景画作。该系统使用动漫图像数据及其相应语义标注图、边缘检测图训练生成式对抗神经网络。通过此种方式,该系统能够同时理解动漫风格以及风景图像中物体的语义和空间关系。在使用中,用户输入草图作为语义标注图,系统自动为其合成边缘图;根据合成的边缘图生成具有恰当风格纹理的画作,从而稳定地处理多样化草图。实验证明该系统可有效满足用户创作需求,生成高质量动漫风格画作。

关键词:协同绘画;深度学习;图像生成

本文译自Sun LY, Chen P, Xiang W, et al., 2019. SmartPaint: a co-creative drawing system based on generative adversarial networks. *Front Inform Technol Electron Eng*, 20(12):1644-1656.

<https://doi.org/10.1631/FITEE.1900386>

中图分类号: TP391

1 引言

人工智能技术的发展为机器提供了更多参与创造性任务的机会,如音乐创作(Roberts et al., 2017)、文本创作(Bowman et al., 2016)以及平面设计(Zhao et al., 2018)。由于人工智能具有模仿和生产大规模设计作品的的能力,因此被广泛应用于批量化自动设计领域,例如生成电商广告(Zhang et al., 2017)。然而,如何促进人工智能与人的合作从而支持多样化的创意表达仍有待探索。绘画是人们探索

想法的重要过程。为创作一幅满意的画作,艺术家通过组织画作各部分元素来表达想象和意图。此过程需要具备专业的知识和高超的绘画技法。因此,与人合作绘画时,计算机需要掌握特定类型画作的领域知识,理解用户绘画输入的语义,从而基于用户输入得到具有特定艺术风格的生动画作。

在理解用户输入方面,已有研究采用草图识别技术识别用户绘制的内容,据此提供实时交互反馈。例如提供建议性指导(Dixon et al., 2010)、相似草图推荐(Lee et al., 2011)以及即兴反馈(Davis et al., 2016a)等。但这些研究专注于学习手绘草图的语义,无法完成草图到画作的转换。因此,它们无法同时完成对用户输入的理解和艺术化。另一类研究将用户输入增强为具有艺术美感的画作,例如将名家作品风格迁移到普通照片上(Gatys et al., 2016)以及

[‡] 通讯作者

* 本文得到以下项目的资助: 科技创新2030——“新一代人工智能”重大项目(No. 2018AAA0100703), 国家自然科学基金(No. 61672451), 浙江省重点研发计划(No. 2019C03137), 中国博士后基金(No. 2018M630658)和浙江大学—新加坡科技设计大学创新、设计、创业联盟(黄廷芳慈善基金会赞助)资助项目

ORCID: 孙凌云, <http://orcid.org/0000-0002-5561-0493>; 向为, <https://orcid.org/0000-0003-2058-5379>

基于边缘检测图和离散颜色块生成图像(Liu et al., 2018)。这些系统的用户可以轻易地从照片或者边缘检测图得到漂亮的画作,但是难以处理多变的用户输入草图,限制了用户的创造与表达。

画作的语义和风格密不可分。画作中各物体的纹理、颜色等与其语义密切相关。本文开发了SmartPaint系统,支持人机合作创作动漫风格的风景画作。使用7234对动漫风景图像及其相应的语义标注图和边缘检测图对一个生成式对抗神经网络(GAN)进行训练。训练完成后,计算机能够同时理解风景图像各物体的语义和空间关系,以及动漫风格中独特的纹理和色彩。在训练网络时加入边缘检测图作为输入提升图像生成质量。用户只需绘制草图作为语义标注图,SmartPaint即可根据草图语义自动为其合成边缘图。用户绘制的草图和系统合成的边缘图作为输入传入到训练完成的GAN中,在几秒内即可生成具有适当颜色和纹理的动漫风格画作。这种合作方式使SmartPaint支持绘画新手和专家自由表达创意想法、获得具有艺术感的画作。实验证明本系统可以成功生成高质量的动漫风格画作。Fig. 1是系统生成的画作示例。

本文贡献如下:第一,开发了基于GAN的人机协同绘画系统SmartPaint,使得用户在自由表达想法的同时高质量输出画作。系统同时学习特定画作的语义和风格,因此可以同时完成对用户输入的理解和增强。第二,提出了基于用户草图语义合成相应边缘图的方法,提高生成画作的质量,并消除用户输入数据和计算机学习数据间的不一致性。第三,提出了四条可参考的人机协作系统的设计规则。

2 相关研究

计算机作为人的合作伙伴,需要在共同完成创意作品时作出的独立贡献(Davis, 2013)。对于绘画

任务,计算机需要理解用户的绘画输入、响应用户的输入以及帮助用户将想法变成精美画作。

2.1 用户输入理解

绘画系统理解用户绘画意图后的行动对于用户才具有意义(Davis et al., 2016b)。已有研究主要采用草图识别算法理解用户手绘草图,并提供绘画指导和建议。EyeSeeYou (Cummmings et al., 2012)结合草图识别算法和单眼绘画技法为用户提供六步绘画指导,帮助用户精确绘制眼睛。ICanDraw system (Dixon et al., 2010)通过识别用户草图提供纠正性的反馈指导,帮助用户绘画人脸。ShadowDraw (Lee et al., 2011)在用户绘制的草图下实时显示阴影图像,引导用户自由绘制物体。Draw Apprentice (Davis et al., 2015, 2016a)对用户输入草图进行分组和分类,并在用户绘制的最新物体旁绘画相关物体,与用户共同创作抽象艺术。Karimi et al. (2018)提出的系统通过识别用户草图,匹配与其结构相似但又不属于同一类别的草图,激发用户创作灵感。Sketch-RNN (Ha and Eck, 2017)通过循环神经网络对草图数据建模,根据用户未完成的草图生成绘画。Sketch2Photo (Chen et al., 2009)和Photosketcher (Eitz et al., 2011)基于用户草图从图像数据库中检索图像块拼接出具有真实感的图像。

这些方法通过识别用户输入,教授和建议用户如何绘画,或者为用户返回满足条件的图像,帮助用户更细致地表达想法。与此同时,创作者也需要具备美化能力的同伴,将创作者自由表达想法的草图转换为美化后的作品。此外,上述方法仅能识别单个笔触或单个物体,没有建立对整幅画作的理解,无法支持用户创作具有丰富内容的画作。



Fig. 1 本文系统生成的画作示例,每个示例均包含用户输入的草图和系统生成的画作

2.2 草图转化画作

传统绘画工具采用物理虚拟笔刷建模和笔触渲染的方法将用户输入转换为具有特定艺术风格的画作。Chu and Tai (2004) 设计了一款虚拟水墨画笔模型, 通过与纸张进行动态交互实现水墨画创作。Ning et al. (2011) 提出了基于轮廓的方法, 将用户输入渲染成水墨画。这些基于规则的方法通常对有限的笔触风格进行预定义, 并在绘制过程中设置诸多限制。因此, 它们只能应用于特定场景, 不适用于处理绘画过程中的开放性问题。

近年来深度学习技术已经成功完成一系列图像合成任务, 可在短时间内输出具有丰富细节和多样艺术风格的图像。Gatys et al. (2016) 等人首先提出了神经风格迁移(Neural style transfer, NST)方法, 使用卷积神经网络将油画风格迁移到普通照片上。后续研究将该方法扩展到了照片风格迁移(Luan et al., 2017)、肖像风格迁移(Selim et al., 2016)、字体风格迁移(Atarsaikh et al., 2017)以及视频风格迁移(Huang et al., 2017)等任务上。本文工作主要受Neural Doodle方法(Champandard, 2016)启发, 利用风格迁移算法将用户涂鸦转换为艺术画作。Neural doodle 通过将用户草图作为语义分割图, 实现相应区域的语义风格迁移, 从而提供了不同于照片输入的新型交互形式。但是, 如同其他NST方法, 它同样没有建立对画作语义和风格的理解, 仅仅只迁移图像低层次的纹理和颜色。因此, NST 方法的效果均严重依赖于选取的风格图像, 只有特定的风格图像才能获得好的迁移效果。

由Goodfellow et al. (2014)提出的GANs被广泛应用于自动图像生成领域。一个GAN 网络包含两个子网络: 生成器网络和判别器网络。生成器旨在生成让判别器无法判别真假的图像, 判别器旨在区分传入的图像是真是假。生成器和判别器通过解决下述极小极大优化问题训练获得:

$$\min_G \max_D V(G, D) = E_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + E_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (1)$$

公式中 $\mathbf{x}$ 表示从真实图像分布 $P_{\text{data}}(\mathbf{x})$ 中随机取样的样本, $\mathbf{z}$ 表示从正态分布中取样的随机向量。

目前已有众多基于GAN 的方法用来实现从草图快速生成图像。Zhu et al. (2016) 提出的交互

式GAN (iGAN) 实现了由用户操作的实时图像编辑。Portenier et al. (2018)和Dekel et al. (2018) 开发了基于草图输入的人脸图像编辑系统。Isola et al. (2017) 提出的“pix2pix”框架可从草图或语义标注图生成真实感的鞋子、衣服、建筑、街道等图像。后续研究通过额外添加输入限制控制生成效果。Xian et al. (2018) 通过输入草图和纹理块控制图像生成结果。Sangkloy et al. (2017) 开发的Scribbler系统允许用户为图像中的物体指定颜色。Zhu et al. (2017) 提出, 在训练数据难以获取时, 非成对图像转换算法比成对图像转换算法更为可取。Li et al. (2018) 则进一步提出堆叠的循环一致对抗网络, 将图像翻译任务分解成多阶段子任务来提升转换质量。但是, 以上基于草图输入的图像生成方法要求用户输入的草图需要具备与训练草图相似的线条特征, 这会对用户创作造成阻碍。

动漫是与日常生活息息相关的艺术表现形式。尽管许多人都渴望创作属于自己的动漫故事, 但这不仅需要特定的艺术技巧, 还极其耗时耗力。现有的许多方法可帮助漫画家减少冗余工作, 如手稿线条简化(Simo-Serra et al., 2016)、动漫图像自动上色(Ci et al., 2018)、照片动漫风格化(Chen et al., 2018)等。

以上方法减少了获取高质量图像所需的复杂度和时间, 但需要保持用户输入的格式与训练数据相同(例如照片或与训练数据线条特征完全一致的草图), 没有试图理解用户想要表达的内容。因此, 用户依然无法通过这些方法来创作具有表达性的、个性化的画作。本文方法同样基于图像翻译算法, 通过让计算机同时学习画作的语义和风格, 并且基于对用户草图的理解合成边缘图用于生成, 消除了用户输入和计算机理解的隔阂, 解决了同时理解和风格化用户输入的问题。

3 系统概述与实现

3.1 系统概述

SmartPaint 系统支持计算机与人共同创作动漫风景画作。用户绘制草图表达创意想法, 计算机将用户绘制的草图转换为动漫风格的画作。Fig. 2 展示了系统整体框架。人机合作通过三个功能模块来

实现：画作生成器、边缘合成器和参考推荐器。

1. 画作生成器

由条件生成式对抗神经网络(Mirza and Osindero, 2014) 组成, 该网络由动漫图像和相应的语义标注图和边缘检测图训练获得。其中, 语义标注图提供动漫图像每个区域的语义信息, 边缘检测图提供每个区域物体的细节信息。将这两者结合作为网络输入, 能够提升计算机生成画作的质量。

2. 边缘合成器

为简化绘画任务, 用户仅需使用不同颜色的画笔绘制草图作为语义标注图, 边缘合成器将根据草图的语义自动合成相应的边缘图。用户绘制的草图和合成的边缘图将一起输入训练好的GAN 中以生成画作。

3. 参考推荐器

在创作过程中提供实时视觉反馈和指导, 能够激发用户创作灵感。给定任意草图输入, SmartPaint实时返回数据库中八张与其最相似的语义标注图给用户作为绘画参考。

Fig. 3a展示了系统主要界面, 包括三个部分: 绘画工具(区域1), 人机共同绘画的画布(区域2), 参考推荐显示(区域3)。绘画工具区包含绘画应用

中常见的工具, 例如画笔。用户可以使用八种不同颜色的画笔分别绘制天空、山峦、树木、草地、河流、岩石、道路和房屋。系统包含两种模式: 用于绘制和修改草图的绘画模式和展示生成结果的生成模式。若用户对生成的画作结果不满意, 可返回绘画模式进行修改。系统使用基于Python 语言的网页框架开发。Fig. 3b 展示了系统的使用情况。文章接下来的部分将详细介绍三个主要功能的实现。

3.2 系统实现

3.2.1 画作生成器

1. 训练数据

包含日式动漫图像及其对应的语义标注图和边缘检测图。所有图像大小均被裁减为 1024×512 像素。为了得到具有相似风格的动漫图像数据, 从宫崎骏绘制的动漫电影或与其画风相似的动漫电影中提取视频帧。本文用到的动漫电影和连续剧列举在附录中。用于训练的图像数据仅限于风景图像, 在数据清洗阶段剔除有人物或明显人造物的图像。为了标注动漫图像的语义, 将数据中包含的物体类别分为九类: 天空、山峦、树木、草地、河流、岩石、

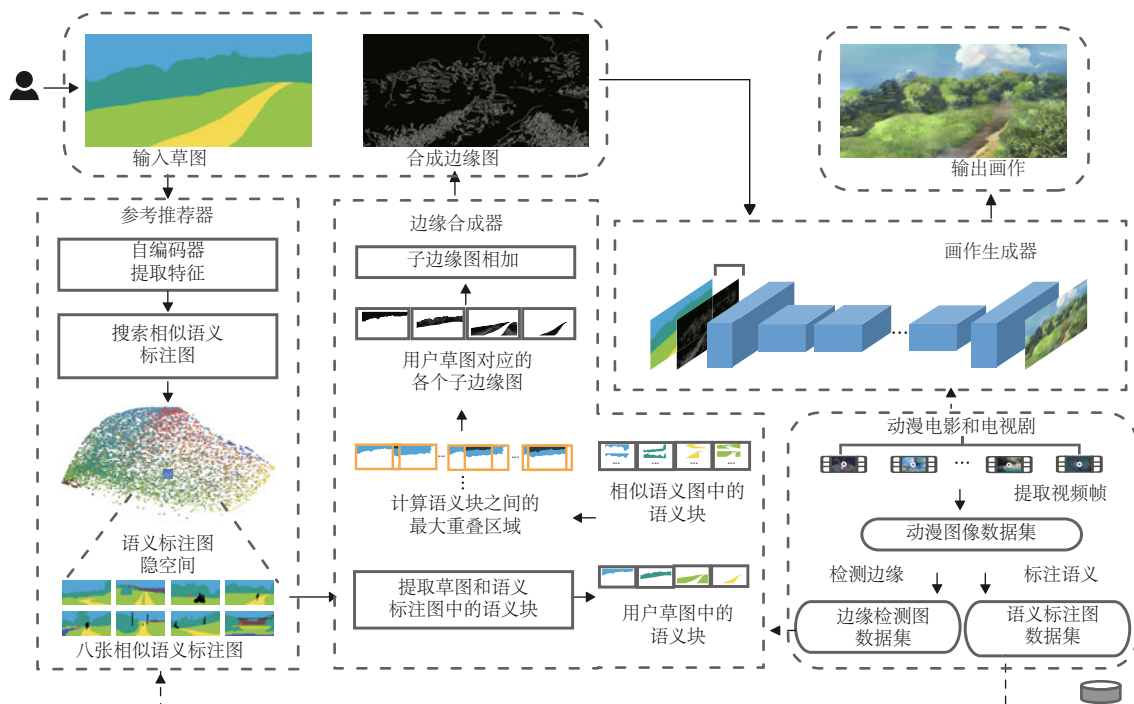


Fig. 2 系统框架

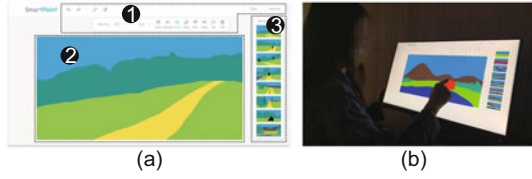


Fig. 3 系统界面和使用情况。(a) 系统界面 (区域1: 包含各种绘画工具, 区域2: 画布, 区域3: 展示推荐图像); (b) 用户使用系统情况

房屋、道路和其他(用黑色标注)。数据集中的每张图像都由人工标注, 每种类别用不同的颜色进行标注。然后, 使用Canny 边缘检测算法(Canny, 1987)提取每张动漫图像的边缘轮廓。最终的数据集包含7234 组数据, 每组包含一张动漫图像、一张语义标注图和一张边缘检测图, 如图Fig. 4 所示。

2. 画作生成网络

基于Wang et al. (2018)提出的pix2pixHD 网络结构。相对pix2pix 框架, pix2pixHD使用 D_1 、 D_2 和 D_3 三个鉴别器作用于不同图像尺寸, 支持更高分辨率和更高精度的图像生成。Pix2pixHD 框架包含两个生成器网络, 本文仅使用其中的全局生成网络用于训练画作生成网络。在本文任务中, 生成器 G 的目标是将语义标注图和边缘检测图翻译成动漫图像, 鉴别器 D 的目标是区分真实图像和翻译生成的图像。边缘检测图以one-hot向量的形式连接在语义标注图之后, 输入到生成器网络中。Pix2pixHD 的损失函数包括对抗损失 $\mathcal{L}_{GAN}$ 和基于判别器的特征匹配损失 $\mathcal{L}_{FM}$ 。已有研究(Güçlütürk et al., 2016) 表明, 在损失函数中加入VGG感知损失 $\mathcal{L}_{con}$ 可以提高图像生成效果。同时, 实验证明VGG 感知损失有利于本文生成任务。最终, 画作生成器网络的损失函数定义如下:

$$\min_G \left(\max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{GAN}(G, D_k) \right) + \lambda \left(\sum_{k=1,2,3} \mathcal{L}_{FM}(G, D_k) + \sum_{k=1,2,3} \mathcal{L}_{con}(G, D_k) \right), \quad (2)$$

其中 λ 用于控制损失函数子项的权重。画作生成网络在单张GTX 1080 Ti GPU 上训练200 轮, 训练时间约为一周。

3.2.2 参考推荐器

根据用户绘制草图实时推荐相似的语义标注图, 引导用户创作。实现该功能时, 需要计算语义标注

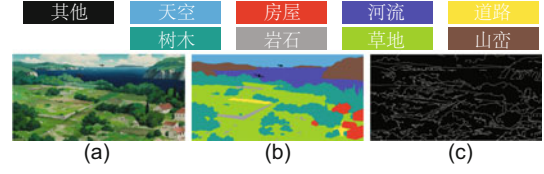


Fig. 4 训练数据集中的数据组形式

图的特征来比较它们之间的相似性。由于草图的多样性, 草图的底层特征(例如物体形状)无法准确代表草图语义。

为解决该问题, 训练一个自编码器提取语义标注图的高层特征, 用于比较语义标注图之间的相似性。自编码器结构如图Fig. 5 所示, 使用多层卷积层提取语义标注图特征。为保证特征提取的速度和准确性, 将语义标注图的大小压缩至 512×256 像素。所有语义标注图在传入自编码器之前都被处理成单通道的灰度图。自编码器输出4096 维的向量作为语义标注图特征。然后, 使用主成分分析(Principal component analysis, PCA) 降维方法将4096 维特征向量降维成20 维特征向量, 定义两个20维向量之间的欧几里得距离为两个语义标注图的差异度。用户任意绘制一笔, 系统将当前草图传入训练好的自编码器中, 并重复上述操作。该过程可获得八张与用户当前绘制草图最相似的语义标注图。当用户绘制的草图没有填满整张画布时, 系统将使用距离空白像素点最近的像素点颜色填充空白像素, 具体填充算法如算法1所述。

Algorithm 1 用户草图空白像素填充算法

- 1: 输入: 用户绘制草图 I
- 2: 输出: 无空白像素的草图 I_{result}
- 3: $blanks \leftarrow I$ 中的空白像素
- 4: **for each** p in $blanks$ **do**
- 5: 分别搜索距离 p 上方, 下方, 左方, 右方最近的有颜色的像素点, 其与 p 的距离分别为 top_{dis} , $down_{dis}$, $left_{dis}$ 和 $right_{dis}$, 像素颜色分别为 top_{col} , $down_{col}$, $left_{col}$ 和 $right_{col}$
- 6: $p_{col} \leftarrow \min(top_{dis}, down_{dis}, left_{dis}, right_{dis})$ 对应的颜色
- 7: **end for**
- 8: **return** I_{result}

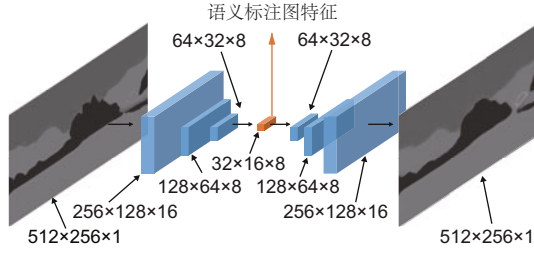


Fig. 5 自编码器结构

3.2.3 边缘合成器

合成边缘图需要考虑动漫线条特征、草图场景语义以及草图中每个物体的形状。训练好的画作生成器仅当输入的边缘图与训练数据具有相同线条特征时才能输出好结果。使用生成网络生成边缘图的方法无法满足该限制，因为生成的边缘图无法避免地存在边缘线条模糊和线条粗细不一等问题。另一种方案是根据用户草图搜索最相似的照片，使用照片的边缘检测图作为用户草图的边缘图。然而，不同于具有规律线条特征的动漫图像，照片的边缘线条曲线不够平滑和清晰(Chen et al., 2016)。因此，本文提出一种半参数化方法，使用数据集中的边缘图作为素材，为用户草图合成新的边缘图。这种方法还可以为处于不同场景中的物体合成不同的边缘。例如，用户可以使用“树”的笔刷在一个场景中绘制单棵大树，在另一场景中绘制成片森林，这两者应具有不同的边缘图。首先，使用参考推荐器，从数据集中获取与新绘制草图相似的八张语义标注图。合成过程具体如下：

第一，向边缘合成器输入用户草图、八张相似语义标注图和相应的边缘图。

第二，使用形状上下文算法(Belongie et al., 2001)分别提取用户草图 T_{label} 和八张相似语义图 S_{label} 的语义块。

第三，对于用户草图中的每个语义块 t ，分别计算其与 ss 中每个语义块的最大重叠区域 moa_s 。其中 ss 代表 S_{label} 的语义块集合， ss 中的每个语义块与 t 具有相同语义。

第四，选取 moa_s 中的最大值， s 对应的边缘图则用于计算语义块 t 的子边缘图。使用相同方法为用户草图中的每个语义块计算子边缘图，将这些子边缘图相加得到最终的边缘图。边缘图合成的伪代码如算法2所示。

Algorithm 2 边缘图合成算法

```

1: 输入：目标语义标注图 $T_{label}$ （用户输入草图），八张相似的语义标注图 $S_{label}$ 及其相应的边缘检测图 $S_{edge}$ 
2: 输出：为 $T_{label}$ 合成的边缘图SE
3:  $Nse \leftarrow$  用户草图包含的语义数目
   // 语义标注图均以矩阵形式输入；矩阵中的值取 $[1, Nse]$ 中的整数，不同值代表不同语义
4: 使用形状上下文算法基于语义提取 $T_{label}$ 和 $S_{label}$ 中的语义块
5: for  $se \leftarrow 1$  to  $Nse$  do
6:    $ts \leftarrow T_{label}$ 中像素值为 $se$ 组成的语义块集合
7:    $ss \leftarrow S_{label}$ 中像素值为 $se$ 组成的语义块集合
8:   for  $t$  in  $ts$  do
9:      $M, N \leftarrow t$ 的高和宽
10:    for  $s$  in  $ss$  do
11:      计算 $t$ 和 $s$ 之间的最大像素重叠区域 $moa_s$ ； $s$ 相对 $t$ 的偏移量为 $(w, h)$ 
12:    end for
13:     $result\_area \leftarrow (s, w, h)$ ,  $s$ 使 $moa_s$ 取得最大值
   // 为 $t$ 计算子边缘图 $iSE$ 
14:     $mask \leftarrow$  高为 $M$ ，宽为 $N$ 的01矩阵，与 $t$ 对应位置取值为1，其余位置取值为0
15:     $iSE \leftarrow result\_area$ 对应的子边缘图
16:     $SE \leftarrow SE + iSE \odot mask$ 
17:  end for
18: end for
19: return SE

```

4 实验

4.1 与原始pix2pixHD算法比较

采用定性分析实验比较本文算法和原始pix2pixHD算法的效果。显示了本文方法在生成各类场景图像上的有效性。用本文训练数据中的动漫图像和相应的语义标注图训练原始pix2pixHD算法，同样训练200轮。绘制具有不同场景语义的草图，分别传入两个方法进行处理。原始pix2pixHD算法生成的图像十分模糊，且图像中各物体几乎没有纹理，仅仅只有颜色区别。相反，本文方法生成的图像不仅具有逼真的动漫风格和细节(Fig. 6d)，而且能够根据不同布局的草图生成不同的图像。例如，将大面积的树诠释为森林，小面积的树诠释为灌木丛。

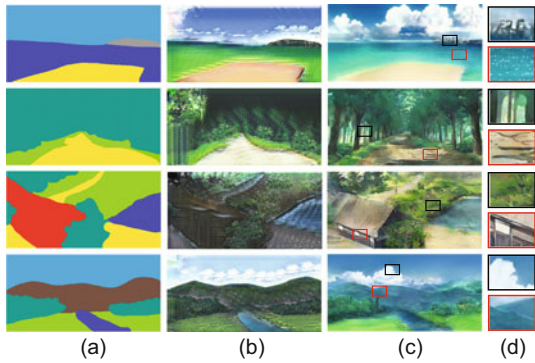


Fig. 6 本文算法与原始pix2pixHD算法效果比较。给定输入草图 (a), pix2pixHD与本文方法分别生成图像 (b) 和 (c), 最右一列 (d) 展示了放大的图像块 (c)

4.2 系统评估

SmartPaint 系统经公展示, 受到超过300名用户体验, 部分用户受邀参与了用户实验。通过两个用户实验对系统实用性进行了评估, 实验目的分别为: (1) 评估SmartPaint 在与绘画新手和专家合作创作时的表现; (2) 评估SmartPaint 在真实设计场景中是否能支持设计师创作。评估准则参照Benedetti et al. (2014) 提出的为新手开发创意支持工具的9条准则以及Oh et al. (2018) 提出的评估人和人工智能系统创作的15条准则。本文使用的10条评估准则强调人机协同系统应该具有: (1) 易用性(2)友好性(3)可交流性(4)灵活性(5)及时性(6)可预测性(7)个性化(8)满意性(9)有趣性(10)有效性。

4.2.1 用户实验一

共20名用户(10名绘画专家、10名绘画新手)参与实验一, 包含两部分: 第一, 用户使用本系统进行自由绘画, 没有主题限制。第二, 用户根据特

定文字描述进行作画。每幅画作的创作时间不超过三分钟。完成绘画后, 用户填写与评估准则相关的调查问卷, 问卷采用李克特五点评分法。最后, 采访用户在问卷中评分的原因以及对系统的其他看法。

4.2.2 实验一结果

实验一结果显示用户对本系统整体满意。每项评分结果的平均分都超过3分, 其中五项的平均分超过4分(易用性、友好性、灵活性、及时性和有趣性)。参与者表示涂抹“颜色块”的绘画方式极大降低了绘画难度。即使在实验中设置时间限制, 绘画专家和新手都创作出了精美画作。用户自由绘画的结果如图Fig. 7所示, 用户按照文字描述的绘画结果如图Fig. 8所示。在两种情况下, 专家创作的画作都比新手创作的画作整体质量更高。原因可能在于本系统的训练图像是从电影数据中提取的专家画作, 专家即使在绘制草图时, 也会考虑更好的布局关系等元素。本系统支持多样化的场景生成, 例如田园、海边、森林等场景。专家和新手都创作出具有创新场景的画作, 例如, “海中梦幻岛”(Fig. 7d)和“空中城堡”(Fig. 7h)。及时的视觉反馈鼓励用户“快速尝试各种可能性”, 并且“以极低的代价预览创意结果”。用户评价系统生成的画作是“梦幻的”、“和谐的”、“有层次感”、“像大师画作”。不过, 专家和新手都希望SmartPaint 能够支持更多风格和场景类型的画作创作。此外, 用户表示SmartPaint常常返回一些没有预料到的结果(可预测性平均分3.4, 标准差0.821), 让人难以理解它的逻辑。对于这种不确定性, 用户看法不一。有的用户认为不确定性的

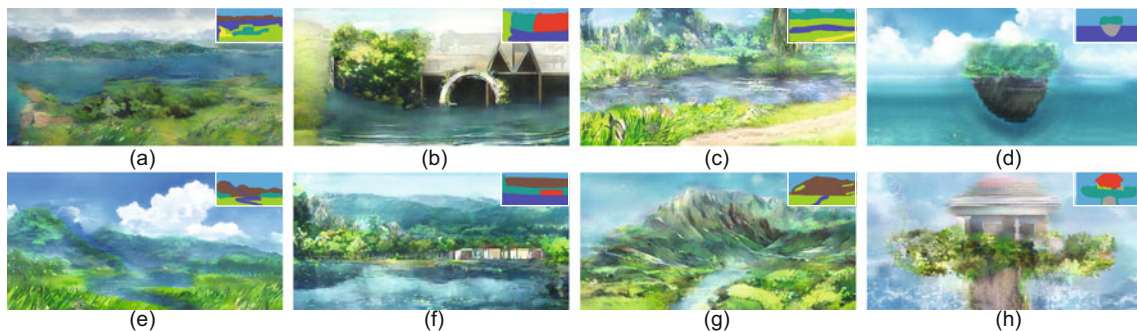


Fig. 7 用户实验一得到的画作结果, 用户在没有主题限制下自由创作这些画作。画作(a) ~ (d) 由新手创作, 画作(e) ~ (h) 由专家创作。每个案例包含用户输入草图和生成的画作

存在使得与Smartpaint的合作是灵感碰撞的过程,并且能够激发他们的创作灵感;也有用户表示不可控性使他们常常无法创作出想要的效果。最后,三名用户表示与SmartPaint的共同绘画过程中成就感较低,因为他们认为Smartpaint是对画作效果的主要贡献者。

4.2.3 用户实验二

实验二旨在探索系统是否能在真实的设计场景中辅助设计师。一位具有丰富商业插画经验、创作过上千幅插画作品的专业插画师被邀请参与实验二。实验二共耗时四小时。首先,向插画师介绍系统功能,其通过半小时熟悉和了解SmartPaint。然后,插画师使用本系统创作五幅插画。他首先使用Smartpaint完成初稿,然后将初稿导入Photoshop中进行优化,优化操作包括“增强边缘”、“修改细节”、增加物体元素等。

4.2.4 实验二结果

该插画师认为SmartPaint系统对人与AI合作的探索非常有价值,并惊讶于SmartPaint的

能力。图9展示了插画师使用系统创作的画作原图和和在Photoshop中优化后的画作。他表示因为SmartPaint无法充分表现动漫图像的典型特征,其生成的画作质量远达不到专业插画的质量要求。例如,在动漫图像中,“天空”的呈现往往是从上往下颜色减淡,但SmartPaint生成的图像中缺乏此特征。他指出SmartPaint目前只支持中规中矩、布局简单的画作创作,无法支持用户创作出具有独特视角、视觉冲击力强的画作。因此,他认为SmartPaint应该学习更多图像数据,否则无法支持成熟设计师完成设计工作。

在和系统合作绘画的过程中,该插画师控制创作过程的意愿比实验一中的被试者更强。他认为系统常常返回无法预测的结果,一方面能使他摆脱部分固定思维习惯,另一方面也使他的创作过程充满过多不确定性,因此他将与SmartPaint的合作形容为“试探的过程”。他希望能对画作进行多维控制,例如能够指定画作的留白区域、画作的情感基调以及画作中每个物体的颜色和透视关系等。此外,插画师表示,相对于为他提供创作灵感,他更喜欢让计算机代替他完成设计中枯燥无味的工作,让他能将更多精力集中在创新上。

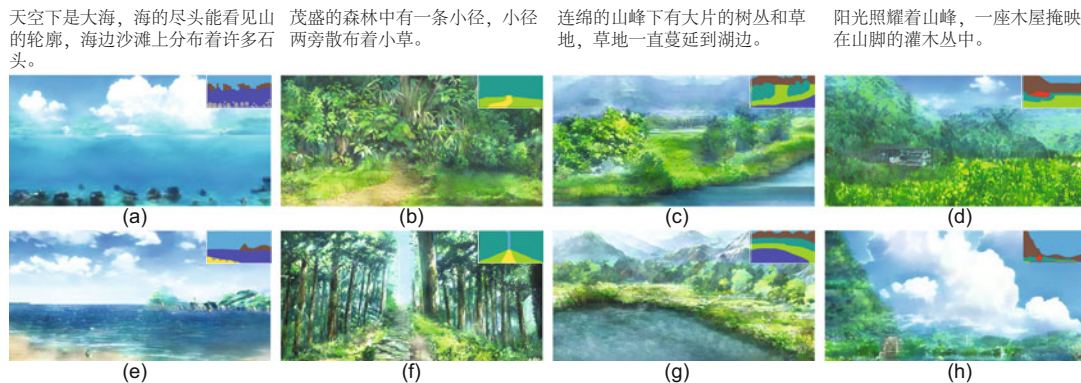


Fig. 8 用户根据指定文字描述创作的画作, (a)与(e), (b)与(f), (c)与(g), (d)与(h) 分别对应相同文字描述。画作(a) ~ (d) 由新手创作, 画作(e) ~ (h) 由专家创作。每个案例包含用户输入草图和生成的画作

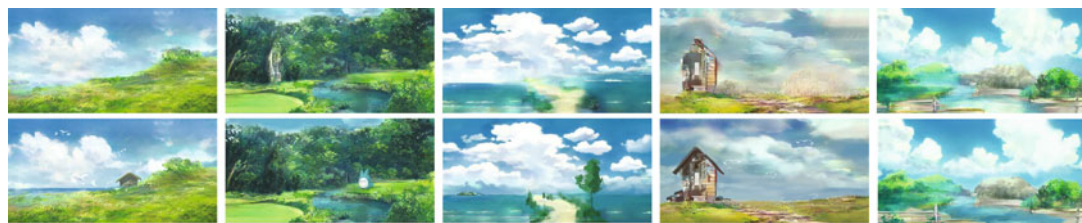


Fig. 9 插画师和系统合作创作的画作。第一排: 系统生成的原始画作。第二排: 插画师在Photoshop中优化后的画作。每幅画作的优化时间分别为 (从左到右): 10分钟、20分钟、25分钟、15分钟和15分钟

5 讨论

本文方法允许机器学习大量具有特定风格的图像以及它们相应的语义标注图和边缘检测图，从而生成出具有特定风格和丰富内容细节的图像。与已有方法相比，本文方法能够生成更加符合特定风格特征的图像。本文的工作目标是让人机通过简单的绘画输入合作绘画。为此，添加动漫图像的边缘检测图作为网络输入；并训练一个自编码器提取语义标注图的高层语义特征，在生成画作时根据用户草图语义自动合成边缘图。该方法进一步改进了生成图像的质量并保证了用户绘画的简单和自由。

本文方法适用于绘制对形状要求不高的物体，对于有特定形状的物体（例如房屋）生成效果较差。系统的不可控性和不可预测性使用户在绘制过程中常常感到困惑，因此本系统更适用于设计过程的早期探索阶段。此外，由于网络使用日式动漫电影中的视频帧数据训练，生成的图像具有较相似的艺术特征。构造足够数量的精细标注数据可将本文方法扩展到其他风格图像的生成。相对于其他基于模版的设计辅助工具，本系统支持个性化的创新内容生成，其简单易用的交互方式使得新手和专家都能创造独特的动漫故事。系统的快速生成能力，也为自动化创意设计提供了新的可能性，例如自动生成创意广告和游戏场景等。

用户实验结果提供了一些人机协同系统可参考的设计准则。整体来说，用户希望在与AI合作的过程中占据主导地位，在AI的协作下高效快速地达成创意目标。因此，第一，人机协同系统应该为人和计算机提供更多维度的交流渠道。与人类伙伴合作时，会通过例如肢体动作、草图、语言描述等多种方式表达自身想法。同样，在与计算机合作时，用户也倾向于使用多通道的方式表达想法。例如，用户希望通过绘制时笔触的轻重告知计算机生成纹理的密集程度。只有当用户向计算机表达的需求足够清晰时，计算机才能生成出符合用户期望的作品。第二，让用户理解AI的逻辑。如用户实验结果所示，一方面，计算机生成结果的不确定性带给用户惊喜，激发用户创意灵感。另一方面，该不确定性也带给用户创作困惑和阻碍。用户需要理解计算机的逻辑从而更好地与其协商。第三，提升用户在创作过程中的成就感。用户实验结果显示，即使用户在与AI

合作的过程中感到愉悦，却只能收获非常低的成就感。AI自动生成的能力使其能在短时间内输出令人惊奇的图像结果，同时也使用户在创作过程中缺乏参与机会。因此，需要有效平衡AI的自动化程度和用户参与度。第四，让AI适应人的绘画或设计习惯。例如，用户的绘画习惯使他们在绘画过程中通常不会填满画布，但只有当用户完全涂满画布时，系统才能最准确地捕捉到用户的想法。人机协同的系统应该为用户提供灵活、自然的交互方式，让用户能够以熟悉的表达习惯进行创作。

6 总结

本文提出了一个人机协同创作动漫风格风景画作的系统，用户无需具备专业知识和技能即可在短时间内创作出精美画作。系统通过学习大量的动漫图像，掌握动漫风景这一独特类型画作的领域知识，从而能够理解用户输入并生成高质量画作。通过算法对比实验测试了系统的生成质量。实验结果显示，相较于原始pix2pixHD方法，本文方法能够生成出更加逼真的动漫图像。通过两个用户实验评估系统实用性，结果显示SmartPaint能够帮助新手用户和专家用户得到满意画作，还能帮助设计师减少设计过程中的冗余步骤，让设计师能够更加容易获得创作灵感。本文总结出四条开发人机协同系统的准则：（1）增加用户和AI的交流通道，（2）改善用户对AI逻辑的理解，（3）提高用户对创意产出的成就感，（4）让AI适应人的创作习惯。

深度学习技术的学习和生成能力为人和人工智能的创意合作带来新的可能性，如何有效结合人类智能和机器智能仍是值得进一步探索的问题。

遵守伦理准则声明

作者声明发表这篇论文没有利益冲突。

References

- Atarsaikhan G, Iwana BK, Narusawa A, et al., 2017. Neural font style transfer. Proc 14th IAPR Int Conf on Document Analysis and Recognition, p.51-56. <https://doi.org/10.1109/ICDAR.2017.328>
- Belongie S, Malik J, Puzicha J, 2001. Shape context: a new descriptor for shape matching and object recognition. Proc 13th Int Conf on Neural Information Processing Systems, p.798-804.

其余文献及附录从略