

基于机器学习和数字信号处理的漫画创作辅助软件

18324034 林天皓

目录：

起因：第 1 页

引言：第 1-2 页

行业背景调查：第 2-10 页

功能设想：第 10-11 页

软件开发迭代计划：第 11 页

总结 第 12 页

参考资料 第 12-13 页

起因

当今世界中，动漫和漫画占据了文化产业中不可或缺的一部分，具有广阔的市场。但是动漫或漫画的制作工序流程十分复杂，需要大量专业人员长期的工作，具有制作周期长，投资回报率不稳定的特点。带来的后果是从业人员的待遇低下，工作压力大。所以，在当今机器学习与信号处理大发展的时期，采用计算机代替一部分枯燥且繁重的工作，能有效减轻从业人员的负担，提高行业的活力。

引言

动漫和漫画制作是一个十分复杂且带有大量人工主观成分的过程，例如对一般动漫的制作的过程，包括前期的文稿编写，从文稿由导演绘制分镜，再根据分镜分配绘画任务，具体的绘画出一帧一帧的图片，还需要后期对手工绘制的图片进行插值，模糊等等处理。分镜经过处理后，进入合成与配音阶

段，加入各种音效与人声，之后还要经过剪辑，调色等步骤，最终才能制作成片。

经过机器学习技术的不断发展，目前我们已经可以对其中一些较为基础的步骤进行简要地处理。目前机器学习与信号处理在动漫制作上的研究方向包括：自动图片/角色生成，可通过生成虚拟人物的方式，为设计师提供灵感以及非主要人物。照片-动画转换，为摄影师的创作提供新方向，提速动漫中的自然景观绘制。草图-漫画转换，快速转换草图为初稿。根据线稿自动创建阴影，线稿自动上色，加速画师的绘画过程，动漫人物识别，漫画文字提取，创建数据集，为后来的机器学习研究提供材料。自动语音合成，给配音师提供指导，减轻配音负担。自动色调合成，将素材融合入现有背景中，素材便于重复使用，降低制作成本。目前的机器学习技术多数只是对单一声音，画面，语言进行处理，对多种模态进行融合学习的技术目前较少。

以下分方向提供了一些目前的技术参考

行业背景调查

1. 数据集

数据集是进行机器学习训练的基础，在研究过程中，Mohit Iyer Varune[1]等选取了美国十九世纪三十年代到十九世纪五十年代之间评分最高的大约四千本漫画，共越二十万漫画页面，并对其中的对话框特别选出，通过图像识别技术，将图片这种的语句转换为文本。选取了一些页面，对这些漫画的页面进行了分类，将漫画页面分为：无明显文字的图片，做为文字背景的图片，文字和图片描述不同，文字与图片描述对应。同时，对页面与页面之间的关系也有标注，分为五类，瞬间：页面之间经过的时间短。动作：相同的对象在同一场景中的动作。主题：在相同场景中引入了新主题。场景：两个页面之间场景发生了重大的变化。对话：页面由人物之间的对话组成。作者通过这些研究，为进一步探究检测漫画剧情完整性，对漫画中作者选取摄影机位置，漫画生成对抗网络等等做了准备。

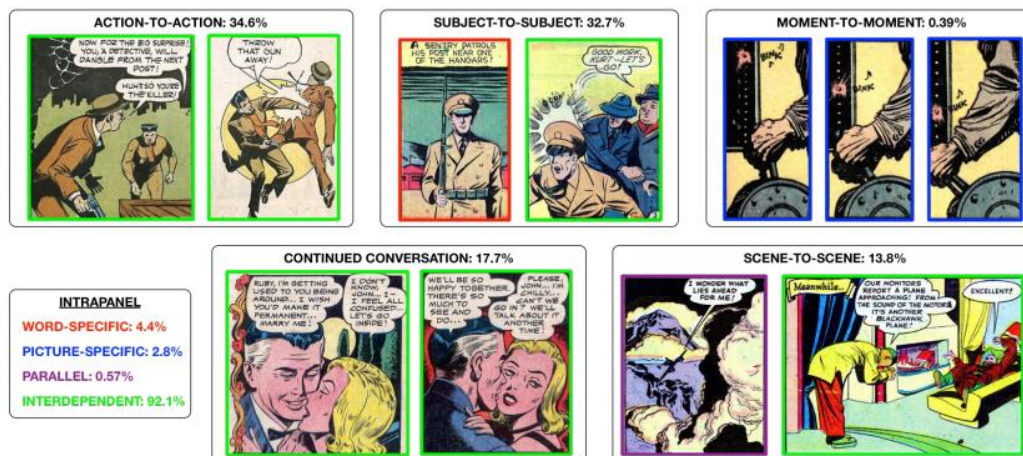


图-分析前后文关系

Lvmin Zhang 等创建了一个漫画分割标注数据集[2]。通过将已有的漫画数据集中的人物与物品通过粗略的模型再通过艺术家的修改标注，或者完全由艺术家手动进行区域标注,再通过标注的数据集训练自己的模型反馈提高进行图像分割的精度，通过人-机-人的工作流程，将分割的效率提高。分割后的图形的局域轮廓，进一步提取画作复杂环境中的人物等。识别出人物后，可以根据艺术家的需求进行进一步的改进颜色，或者对人物背后的环境进行替换，通过与原画作的对比，对画作的光照进行提取，进一步为探究艺术家在创作漫画中的底层原理与计算机创作提供了基础。

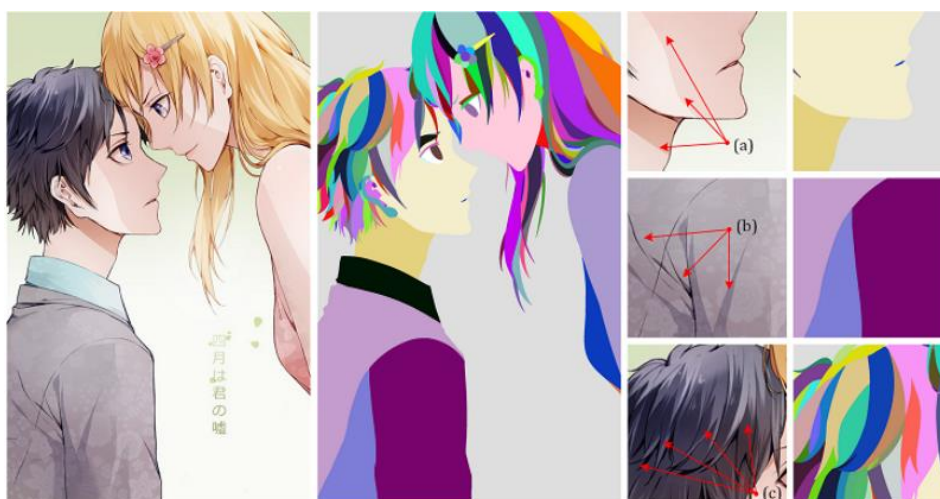


图-图像分割结果



图-从复杂漫画中提取精简人物

2. 姿势-图像转换

目前深度生成模型自动生成图像和视频的方法。这些对于媒体创建工具（例如照片编辑，动画制作和电影制作）很有用。专注于动漫创作，自动角色生成可以激发专家创建新角色，也可以有助于降低绘制动画的成本。自动生成全身角色并为其添加高质量动作，对于制作新角色和绘制动画非常有帮助。Koichi Hamada[3]等通过基于[14]所生成的骨架-图像数据集，通过在每个比例下对全分辨率结构条件图进行下采样以形成多比例条件图。对于每个尺度，生成器从具有结构条件的潜变量生成图像，并且鉴别器基于结构条件来区分生成的图像和真实图像，训练生成对抗网络，解决了传统网络在实现生成高质量的图像和图像结构的一致性上的局限性，使得模型能够生成更高分辨率，详细的人物图像动态画面。同时基于模型，创建了更高分辨率(1024x1024)的数据集。



A Voynov 等[4]使用无监督的方法来研究预先训练好的 GAN 中潜在空间的可解释性。这不同于其他如需要人为标签、预先训练的处理模型或某种形式的自我监督方法，能有效地发现潜在编码的有效

的可解释性移动方向。

通过提出重构分类准确率。作为重构模型中的分类头，其处理的是一个多分类问题。因此，RCA越高，该方向表示的特征和其他特征的解耦度越低，区分性更高。

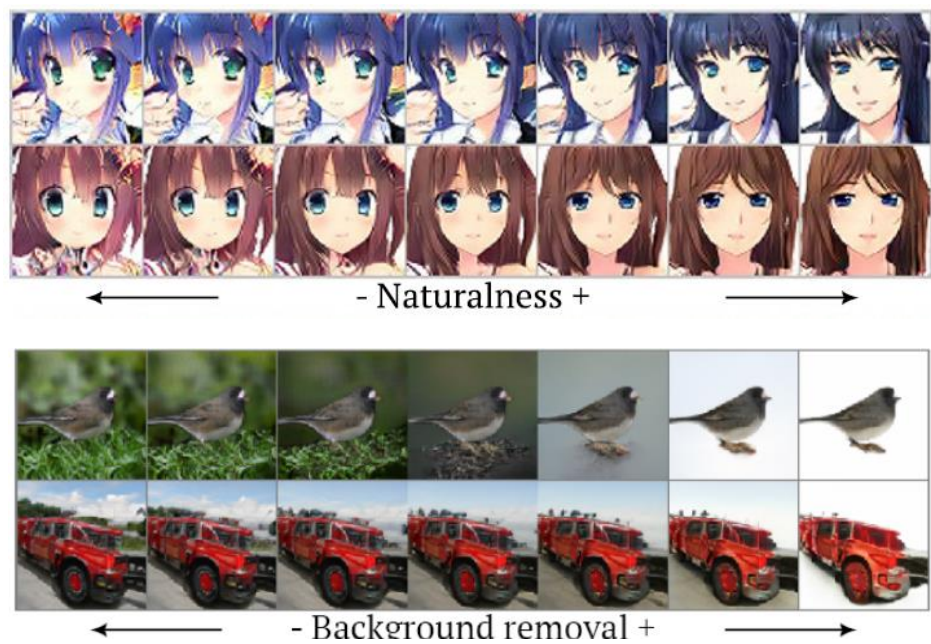
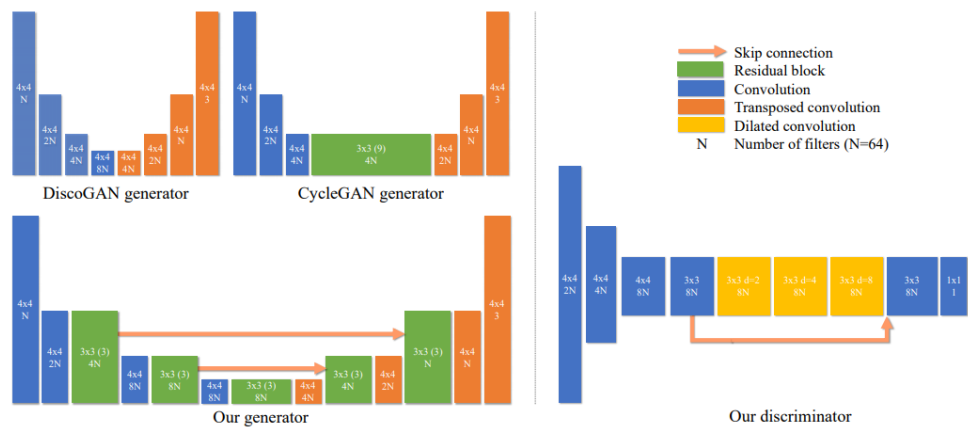


图-正交特征序列图

还能应用与上图中的 背景消除的变换中，为背景消除提供新思路

3. 照片-图像转换

Aaron Gokaslan 等[5]在做图像-图像映射的时候，如果转换前的图像类和转换后的图像类形状差别太大，往往纹理的映射就会出现错乱，在本文中，通过引入跳跃鉴别器，将较远距离的信息并入图像中对于翻译任务很有用，因为鉴别器可以确定区域应适合的位置，可以从更大范围的特征去辨识图像。





Junho Kim[6]等引入了新的注意力辅助分类器，以热力图的形式告诉我们，模型通过哪些像素点得知图片属于某个类别。特征图经过 GAP 处理后每一个特征图包含了不同类别的信息，权重 w 对应分类时的权重。绘制热力图时，提取出所有的权重，往回找到对应的特征图，然后进行加权求和即可。同时引入 AdaLIN

评判机制，具有参数 ρ ，在 IN 较为重要的任务中， ρ 接近 1；在 LN 较为重要的任务中， ρ 接近 0。

AdaLIN 结合了 AdaIN 和 LN 的优点：前者可以保持图像内容，后者更关注于全局信息。

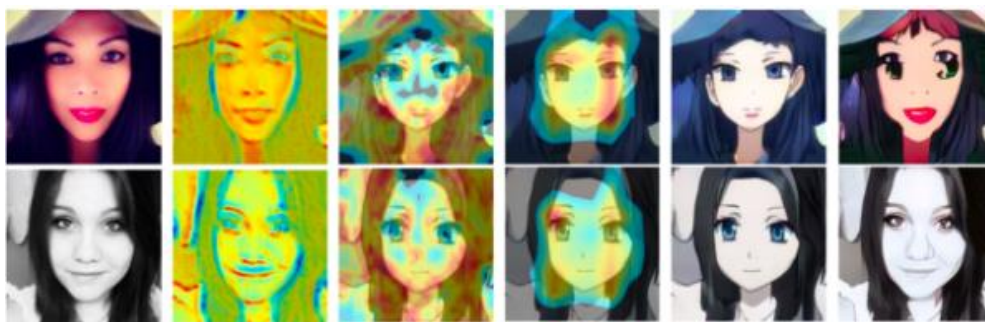


图-热力图展现注意力分类器结果。

CartoonGAN: Yang Chen[7]等一种将现实世界场景的照片转换为卡通风格的图像的生成对抗网络 (GAN) 框架,通过总结了卡通图与现实图中对应的特点：对于区域，卡通风格具有高水平简化和抽象的独特特征，对于区域的边缘来说卡通图像倾向于具有清晰的边缘，平滑的颜色阴影和相对简单的事实

。该对抗网络中，特别注意了鉴定边缘的特征，得以保留清晰的边缘

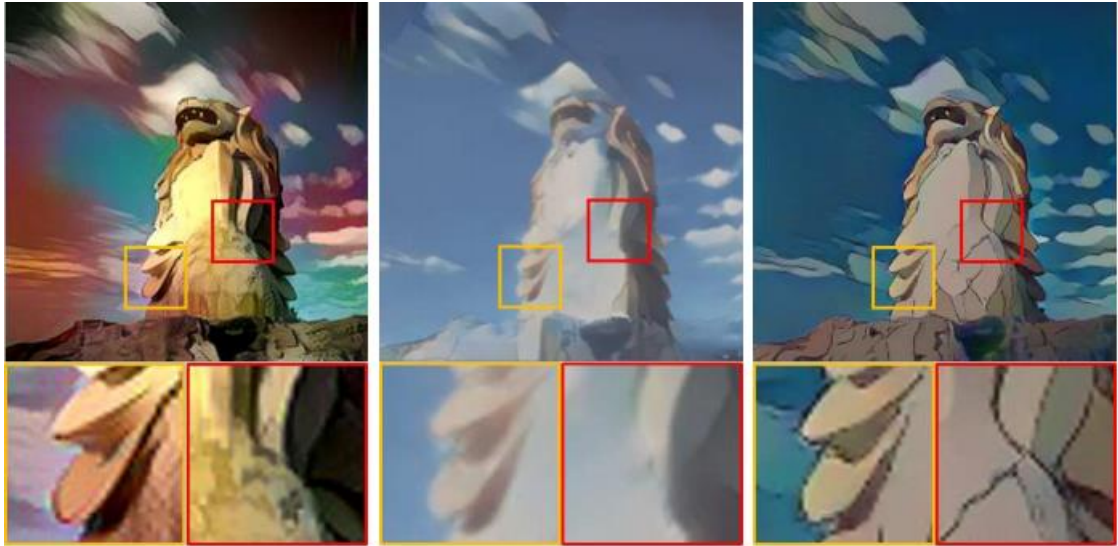


图-边缘线条锐化

4.线稿自动上色

Masashi Aizawae[8]等使用了一种给巩膜着色的方法。在某些样式中，眼睛过于抽象。另外，在灰度线图中，在许多情况下，皮肤和巩膜均以白色表示。因此，不能总是使用现有的自动着色技术来确定边界。结果，巩膜经常被涂成与皮肤相同的颜色，并且线条图中的这些区域和着色结果之间不匹配。本文中通过检测与交集大小训练语义图像分割网络来检测巩膜区域，并提高以前的自动着色方法的准确性。

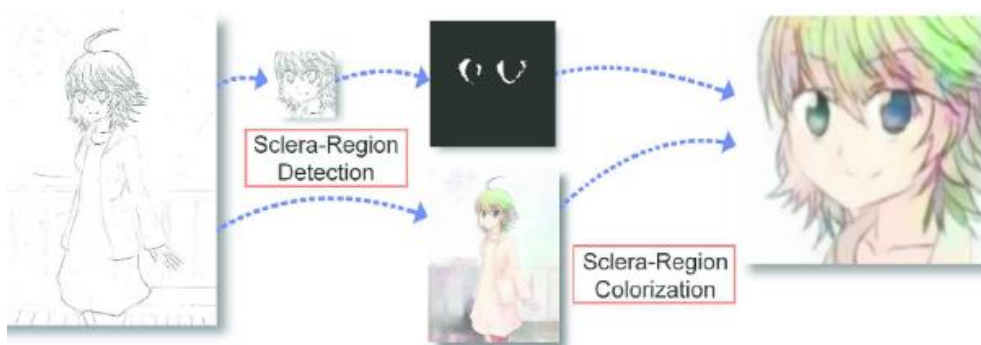


图-巩膜区域检测与着色

两段式上色: Lvmin Zhang[9]等不同与其他的自动上色模型，该模型采用的是点状的指引而不是线状的索引，同时本文将线稿上色的过程分为两个阶段，包括根据点状的提示颜色初步的上色，和检测

初步上色中的缺陷进行精细修整并生成最后的结果图像。在训练初步上色的网络中，生成提示点对应上色图片的数据集中，采用了中心点确定的一个较小的方形区域，并采用区域的平均颜色值。区域的大小从 1×1 至 9×9 的中均匀抽取，同时为了增强颜色的对应性，有 1% 训练图像的颜色是精确对应原来图像的颜色。在训练第二部精确修正的过程中，通过机器对训练集的彩色图片进行了一些破坏，例如人为创造颜色偏移的溢出，填充错误的颜色，类似于打码地将图片遮蔽一部分，用于检测图像中的缺陷区域的数据集。采用这种方法通过两步的工作，可以缩小模型的大小，也可以更灵活的对模型进行调整。



图-两段流程着色

5.精细纹理消除

Yi Ji[10]等使用数字信号处理技术平滑图像。在各种图像处理中，我们需要用到纹理去除，重新纹理化，层着色等等问题。本文提供了一个对图像纹理进行平滑的方法，称为图像外观保存。通过给图中每一个点设置一个期望值，期望值越大，代表这个结构如果被去除，可以使得图片更加“平坦”。有了每个像素对应的期望值，可以通过 0-1 背包的方式，剔除图像中期望值较大的点，通过逐步调整背包”的大小，EAP 方法能逐渐的发现纹理的位置和结构，更好的平滑图像。

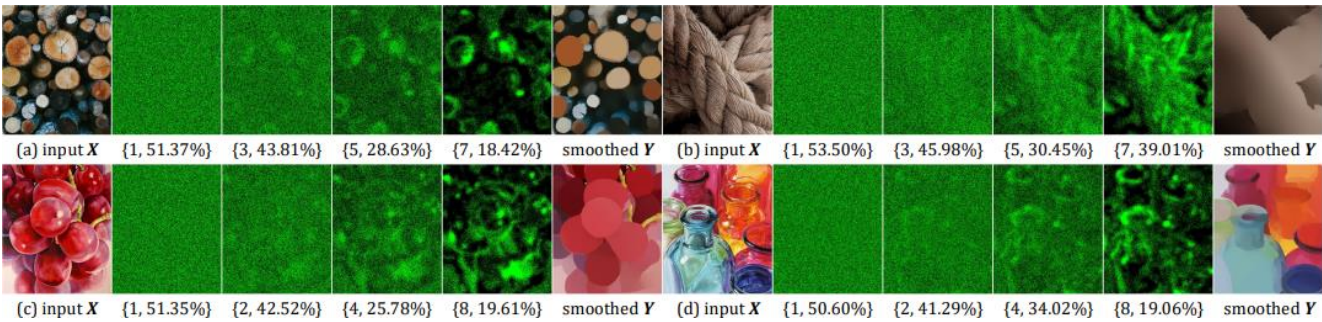
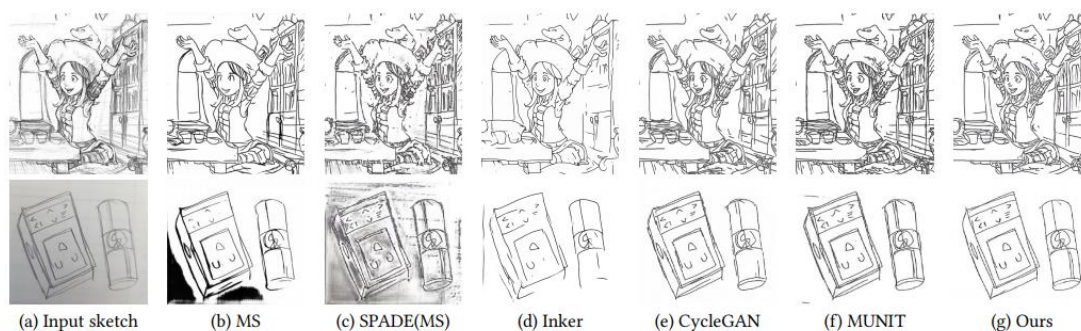


图-图像纹理去除

上图表示不同“背包”的容量大小下的去除纹理的点状态

6. 草稿-线稿转换

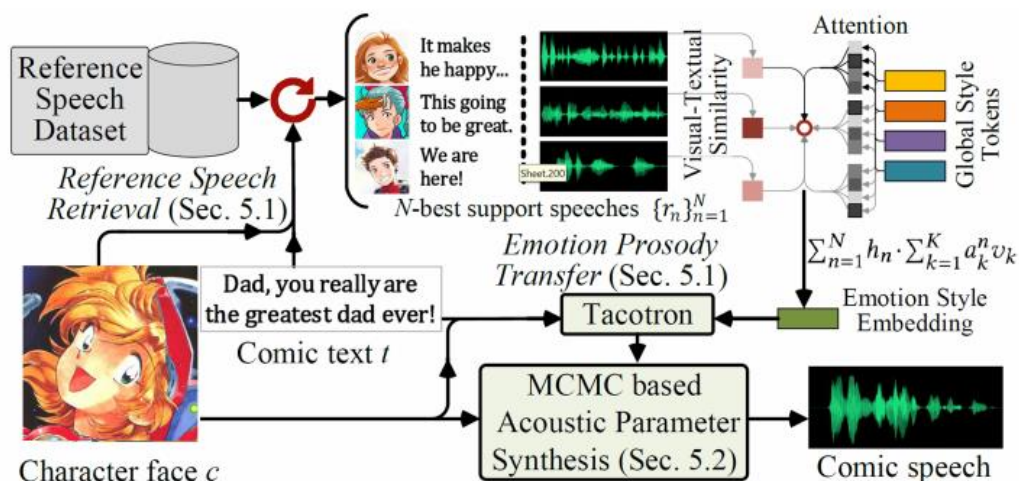
Gayoung Lee, Dohyun[11]等将草图转换清晰的线稿。



草图绘制使艺术家可以快速表达艺术概念，而不必担心细节。草图设计完成后，由于草图笔划的粗糙性和冗余性，美术师需要将草图简化为简洁的线条，这既费时又影响生产率。但通过自动草图简化可以区分要保留的详细表达和需要消除的有噪声的笔触。在实现过程中，通过生成相近的笔触，模拟草图绘制的时候对同一个线条多次描写的情况。通过这样的方法扩充数据，生成草稿与线稿之间的对应数据来进行监督学习。

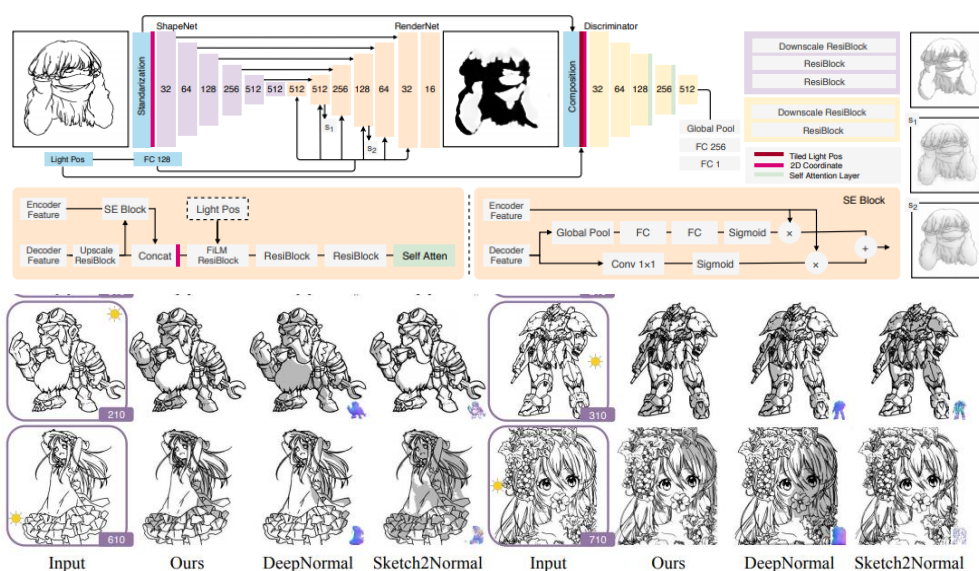
7. 漫画语音合成

Yujia Wang[12]等使用机器学习为漫画生成语音。随着新媒体技术和设备的发展，漫画的格式也已开始发展，还存在新的增强现实（AR）应用程序，这些应用程序通过手机或AR头戴式耳机播放覆盖在漫画图像上的声音和动画。最近的趋势是制作类似于有声读物的有声漫画，并广受欢迎。本文中通过通过建立漫画图像中的场景例如气球，尾巴，文本和人物来识别场景，推断漫画人物的情感和个人属性。同时，通过面部识别技术，辨别人物的年龄。来对应合成文本对应的语音。



8. 线稿添加阴影

在各种艺术媒体和格式中，大多数绘画在应用颜色之前都先使用线条和阴影进行草绘，画家经常花费几个小时来绘制一张吸引人的图片，并多次反复调整角色的形式和结构。对于无法在现实世界中观察到的 2D 草图，绘制阴影尤其具有挑战性，因为没有推理的 3D 参考模型。我们的工作使用深度学习方法来学习一个非线性函数，该函数“理解” 2D 草图所隐含的 3D 空间关系并渲染阴影，给定线条图和照明方向，Schiele, B.[13]等利用模型自动生成图像，并在其中使用详细而准确的硬阴影来增强线条图。数据集的扩充使用了同一个模型旋转多个角度的不同状态。模型的构建分为两部分，第一部分将 2D 草图编码为 3D 结构，用来表示物体的 3D 几何信息。再将编码的 3D 信息传入第二渲染层，用来推理草图的阴影。再将推理的结果合成与草图上完成阴影的绘制。



功能设想

基础功能

1.计算机生成相关：

包括自动(或人工指导)的线稿上色，台词配音合成，照片-漫画转换，3D 动作的自动生成。

通过各种方式，减少艺术家在创作过程中繁琐而又重复的工作，提高工作效率，减轻工作强度，创造一个舒适的工作状态。

2.计算机评价人工作品

包括漫画，配音，动画进阶改良指引，从光照，形态，景物布置的合理性，人物神态等方面辅助艺术家在制作过程中把握作品的基调。

考虑到从业人员的能力发展需求，可以通过计算机评价的方式，帮助刚刚从业的新人发现自己的不足，清除一些限制从业人员的能力发展速度的障碍，促进整体技术水平的提高。

商业化功能

考虑到将一个产品落地需要大量的人力物力，在没有经济支持的情况下，很难做出一款长期稳定发展的软件，所以，考虑通过如下几种方式获得经济来源。

1.相关配套软件高级功能的商业化授权

限制免费用户的使用次数，对于商业化应用采取付费授权的方式，同时给与优先的技术支持。

2.打造一个基于计算机生成艺术品的开发者社区

为了方便行业的从业人员互相交流与发展，提供社区环境不仅可以在其中学习大家的优点，同时也利于行业中人力资源的合理分配，对于 HR，可以通过手续费的方式为行业间的人力资源流通提供中介服务。

3.计算机艺术相关知识的学习与培训

为了行业的长久发展，需要不断的培养行业人才，带来源源不断的新鲜血液，通过合理的环境，打造一批关于计算机艺术的精品课程，使更多有兴趣的学生有方式学习相关知识，从长久促进行业的发展。

软件开发迭代计划

- 1.初步学习计算机图形学，计算机图像\处理，自然语言处理，人工智能等相关技术。
- 2.阅读与学习上述提到的论文，关注 SIGGRAPH,ECCV 等相关会议，学习与复现论文的实现过程，做出初步的模型。
- 3.组建技术团队，进一步学习相关知识，同时自己实现相关算法，为软件的初步上线做准备。
- 4.根据上线结果，与软件使用者的评价，对软件做出进一步的改进，迭代跟新软件功能，并进行商业化运作。

总结

随着计算机计算能力的进一步提高，对硬件资源的充分利用使得各种机器学习的模型在计算机上的训练速度越来越快。许多被运算速度所限制的算法正在逐渐被解决，目前这些模型仅仅在特定的限制下有较好作用。未来，希望技术的发展能简化更多的繁琐工作，为生产力的进一步提高创造条件。

参考资料

- [1]. Mohit Iyer Varun Manjunatha Anupam Guha:The Amazing Mysteries of the Gutter: Drawing Inferences Between Panels in Comic Book Narratives CVPR 2017
- [2]. Lvmin Zhang , Yi Ji , and Chunping Liu :DanbooRegion: An Illustration Region Dataset ECCV 2020
- [3]. Koichi Hamada, Kentaro Tachibana, Tianqi Li:Full-body High-resolution Anime Generation with Progressive Structure-conditional Generative Adversarial Networks
ECCV 2018
- [4]. Unsupervised Discovery of Interpretable Directions in the GAN Latent Space A Voynov, A Babenko -
arXiv preprint arXiv:2002.03754, 2020
- [5]. Aaron Gokaslan, Vivek Ramanujan, Daniel Ritchie: Improving Shape Deformation in Unsupervised Image-to-Image Translation ECCV 2018

- [6]. Junho Kim , Minjae Kim , Hyeonwoo Kang:U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation ICLR 2020
- [7]. Yang Chen, Yu-Kun Lai, Yong-Jin Liu, CartoonGAN: Generative Adversarial Networks for Photo Cartoonization CVPR 2018
- [8]. Masashi Aizawa, Yuichi Sei, Yasuyuki Tahara, Do You Like Sclera? Sclera-region Detection and Colorization for Anime Character Line Drawings IJNDC2019
- [9]. LVMIN ZHANG, CHENGZE LI, TIEN-TSIN WONG:Two-stage Sketch Colorization ACM Trans. Graph. 2018
- [10]. Lvmin Zhang , Chengze Li , Yi Ji:Erasing Appearance Preservation in Optimization-based Smoothing ECCV 2020
- [11]. Gayoung Lee, Dohyun Kim, Youngjoon Yoo:Unpaired Sketch-to-Line Translation via Synthesis of Sketches SIGGRAPH 2019
- [12]. Yujia Wang, Wenguan Wang, Wei Liang:Comic-Guided Speech Synthesis ACM Trans. Graph 2019
- [13]. Qingyuan Zheng , Zhuoru Li , and Adam Bargteil :Learning to Shadow Hand-drawn Sketches CVPR 2020
- [14] Pose Guided Person Image Generation (Ma, L., Sun, Q., Jia, X., Schiele, B., Tuytelaars, T., Gool, L.V.: Pose guided person image generation. In: Proc. of NIPS (2017))