

表一：毕业论文（设计）开题报告
Form 1: Research Proposal of Graduation Thesis (Design)

论文（设计）题目 Thesis (Design) Title: 草图到图像自动生成算法的研究与实现
<p>（简述选题的目的、思路、方法、相关支持条件及进度安排等） （ Please briefly state the research objective, research methodology, research procedure and research schedule in this part. ）</p> <p>一、研究背景</p> <p>作为人类接受信息的主要形式，以数字图像为载体的视觉信息占比达80%。在数字图像的采集、传输、处理等各个环节都可能导致图像品质的下降，这直接影响了提取图像所含信息的质量。而对于监督学习，数据集的标注成本高，所以有效的图像增强技术一直是计算机视觉领域重要的研究课题；另一方面，对于不同风格和不同领域下的图像，并不像自然图像那样容易获取，比如游戏和动漫中的人物角色图像，需要专业人员使用渲染软件进行人工制作，成本较高；而一些画家的作品的艺术风格，也不是一朝一夕可以模仿和创作。对于以上问题，一种流行的解决方案就是图像生成。在机器学习中，生成模型可以用来直接对数据建模，也可以用来建立变量间的条件概率分布。在 GAN 提出之前有影响力的生成模型主要有 Autoregressive model（自回归模型）、VAE（变分自编码器）、GLOW（基于流的方法），但都存在计算开销较大或者生成的图像较为模糊等问题，而 GAN 则利用博弈不断的优化生成器和判别器从而使得生成的图像与真实图像在分布上越来越相近，传统概率生成模型一般都需要进行马可夫链式的采样和推断，而 GAN 避免了这个计算复杂度特别高的过程，直接进行采样和推断，从而提高了 GAN 的应用效率。</p>

二、研究现状

生成模型是建立服从标准的正态分布的输入与真实数据分布之间的转换，但其难题就是判断生成分布与真实分布的相似度，因为我们只知道两者的采样结果，不知道它们的分布表达式。传统的生成模型基于生成图像的重构损失进行训练，如使用 MSE（均方误差）作为重构误差，但会造成重构图像较为模糊的问题，因为重构的图像在概率分布上只是与真实图像的平均值较为接近，而与每一个单独的图像相距较远。一种解决办法就是利用 CNN 提取重构图像与真实图像的特征，从而比较它们的相似性以计算损失，这个方法曾经在由低分辨率图像生成超分辨率的图像问题中使用[1]，但一些低频的信息容易在卷积中丢失。而 GAN 则利用博弈论中的零和博弈，将生成问题视作判别器和生成器这两个网络的对抗和博弈：生成器从给定噪声中（一般是指均匀分布或者正态分布）产生合成图像，判别器分辨生成器的输出和真实图像。由此，两个网络在对抗中进步，在进步后继续对抗，由生成网络得到的图像也就越来越完美，逼近真实数据，从而可以生成想要得到的图像。

现在的图像生成主要有几种应用，一种是生成动画或者游戏中的虚拟人物角色的面部照片[2-3]。众所周知，游戏开发和动画制作成本很高，并且需要雇佣许多制作艺术家来完成相对常规的任务。但通过 GAN 就可以自动生成动画角色并为其上色。第二种是风格迁移的任务，即可以将图像从一个领域（如真实的风景）转换为另一个领域（莫奈绘画或梵高）[4]，上述都是基于无监督学习进行 GAN 的训练。而 cGAN 则可以引入图片作为监督信息，将条件输入添加到随机噪声中，从而控制完成图像与图像之间的转换问题，如把轮廓图转化成真实图片，把 mask 转化成真实图，把黑白图转化成真实图等[5]。cGAN 的应用还包括文字到图像的转换[6]，通过循环文字编码器（recurrent text encoder）生成的文字属性，来作为条件对 GAN 进行训练。这使得 GAN 能够在输入的文字描述和输出图像之间建立相关性联系。

但这种跨域问题具有一对一的特点，即生成的图像较为严格地遵循参照的条件，一旦条件较差，或者条件难以得到，真实图片的生成就会出现一些问题。比如人手绘制的草图，由于相比边缘提取后的真实图片缺少许多细节信息，生成的图片就不自然，趋向于与输入草图像素级别的映射，只是进行

了上色。试想一下，当我们某时某刻灵机一现，脑中迸发出一幅简单的素描，或是随手描绘映入眼中的惊艳物品的草图后，能够通过特定算法转化成真实完整的图片，这无论是对于交流还是回忆，都有极好的帮助。当我们向他人描述一件物品的时候，仅仅通过语言来表达的话，是比较单一并且可能存在理解障碍的，而此时最好的交流方式便是通过图像来表达，但是实物照片资源并不容易获得的情况下，只能通过草图来传述，而此时若是使用草图到图像自动生成算法的话便能够极大地提升双方交流的流畅度和理解度。它可以在获得一定自由度的同时并保持与输入条件的相似性。

三、技术路线

此问题的解决方法虽然是属于监督学习的范围，但不能将输入条件作为强条件。这会导致一对一的映射，所以需要将其作为一种弱约束来设计损失函数，比如将生成的图像再用逆映射生成回去应该与输入的图像尽可能接近。在转换中使用两个生成器和两个判别器，两个生成器和进行相反的转换，试图在转换周期后保留输入图像。

由于是监督学习，构造数据集也是一个重点，尤其是获得人手绘制的草图，需要应用特定场景下数字图像下的算法。

四、进度安排

时间段	时 间 长 度	工作安排
2019.12-2019.1	1 个月	阅读有关论文，提出最终的解决方案，确定 具体使用到的技术，以及了解相关实现。
2020.1-2020.2	1 个月	进行初步的实验，并发现存在的问题。
2020.2-2020.3	1 个月	得到初步的结果，改进实验的算法，与导师讨论，解决遇到的问题。
2020.3-2020.4	1 个月	与其他模型进行量化比较，完成论文定稿。

五、参考文献

- [1] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi: "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". In: CVPR(2017)
- [2] Yanghua Jin, Jiakai Zhang, Minjun Li, Yingtao Tian, Huachun Zhu, Zhihao Fang : "Towards the Automatic Anime Characters Creation with Generative Adversarial Networks". In: CVPR(2017)
- [3] Tianyang Shi (1), Yi Yuan (1), Changjie Fan (1), Zhengxia Zou (2), Zhenwei Shi (3), Yong Liu (4) ((1) NetEase Fuxi AI Lab, (2) University of Michigan, (3) Beihang University, (4) Zhejiang University): "Face-to-Parameter Translation for Game Character Auto-Creation". In: ICCV(2019)
- [4] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros: "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks". In: ICCV(2017)
- [5] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros: "Image-to-Image Translation with Conditional Adversarial Networks". In: CVPR(2017)
- [6] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaolei Huang, Xiaogang Wang, Dimitris Metaxas: "ICCV 2017 StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks". In: ICCV(2017)

Student Signature:

Date:

指导教师意见

Comments from Supervisor:

1.同意开题

2.修改后开题

3.重新开题

1.Approved()

2. Approved after Revision ()

3. Disapproved()

Supervisor Signature:

Date: