使用 CGAN 对图形验证码的优化

（基于 FashionMNIST 数据集）

张知行

清华大学 材料学院

传统的验证码主要由数字或字符构成，在图像识别技术相当成熟的时代，已经难以满足安全性的要求。因此，使用一种对机器较为难以识别但是肉眼可以轻易辨别的验证手段十分必要。为了克服传统验证码的缺陷，我们创新地决定使用衣饰用品图片作为验证码。在此基础上，我们设计了一种可以产生种类明确，易于肉眼区分的图片的方式来适应网络应用的发展。

FashionMNIST[5] 是一个由 28×28 的灰度图片构成的衣饰用品数据集，具有相比于传统的 **MNIST** 更高的辨别模型训练难度。我们在 **FashionMNIST** 数据集上使用 **CGAN** 进行了模型构建与训练，并将生成的图片与已经证明得到较好验证效果的生成模型 **PGGAN**[3] 生成的图片进行对比，说明了我们的模型存在一定的优势并可以更好地应用于当前的验证码图片生成当中。

Introduction

互联网应用的发展十分迅速，几乎所有的服务商都提供互联网服务。验证码是一种区分用户是计算机还是人的公共全自动程序，可以用来有效防止恶意破解密码、刷票、论坛灌水等行为，是一种为了保证网络应用服务手段的有效方式。但是目前大部分的验证码都是使用数字或者字母验证码的阶段。对于数字和字母验证码，目前已经有了有效

的使用技术手段破解的方式[1]。传统的验证码在很大程度上已经难以实现区分人类和机器的目标。因此，使用一种对机器较为难以识别但是肉眼可以轻易辨别的验证手段十分必要。同时，使用现有的数据集来制作的验证手段十分容易通过算法进行识别。因此为了适应网络应用的发展，我们需要大量种类明确，易于肉眼区分但是机器难以分辨的图片。

GAN[2]（Generative Adversarial Networks）是 GoodFellow 在 2014 年的论文中提出的一种思想，要求 Generator 学习将一个已知的高斯分布映射到更高维的空间中去拟合真实图像的分布。简单来说，GAN 是一种将生成模型与判别模型进行结合来实现更好的生成的方式。

但是朴素的 GAN 存在模型过于自由不可控的缺陷。同时，朴素的 GAN 中生成的结果并不含有标签，无法直接用于我们需求的场景当中。因此我们选取 CGAN[4](Conditional Generative Adversarial Nets) 作为我们模型的基础。CGAN 的简要结构如图 1所示。

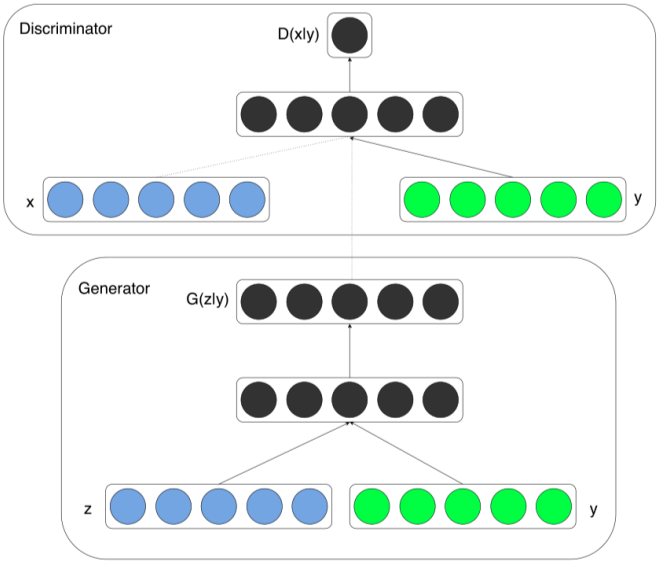

Figure 1: CGAN 基本结构[4]

在此基础上，我们设计了相应的算法并进行了一系列的参数调整来实现尽可能最优的结果。我们对最后生成的结果与渐进式的模型中得到的结果进行了比较，验证了我们的模型的良好表现并说明了我们的方式的良好的应用前景。

Main Objectives

1. 在 CGAN 模型的基础上设计一个适用于 FashionMNIST 数据集的算法并实现。
2. 通过大量的实验对设计的模型进行参数调优。
3. 对 PGGAN 模型中源码进行修改并在相同数据集下进行训练得到最终输出结果。
4. 对两种模型下的表现进行分析并验证我们模型在验证码生成情形下的优势。
5. 使用图像识别程序对我们生成的验证图片和使用现有数据集生成的图片进行判断，验证我们生成的图片的优势性。

Methods & Experiments

基于 [4] 中的设计思想，我们设计了图 2中的网络作为我们的基本模型并在此基础上进行训练与模型调整。我们使用的训练机器为配置了 4 核心 Intel(R) Core(TM) i5-7500 CPU @ 3.40GHz，24G RAM，GTX 1080Ti 的 Ubuntu16.04 极客云服务器，我们的代码编写环境为 Python=3.6。

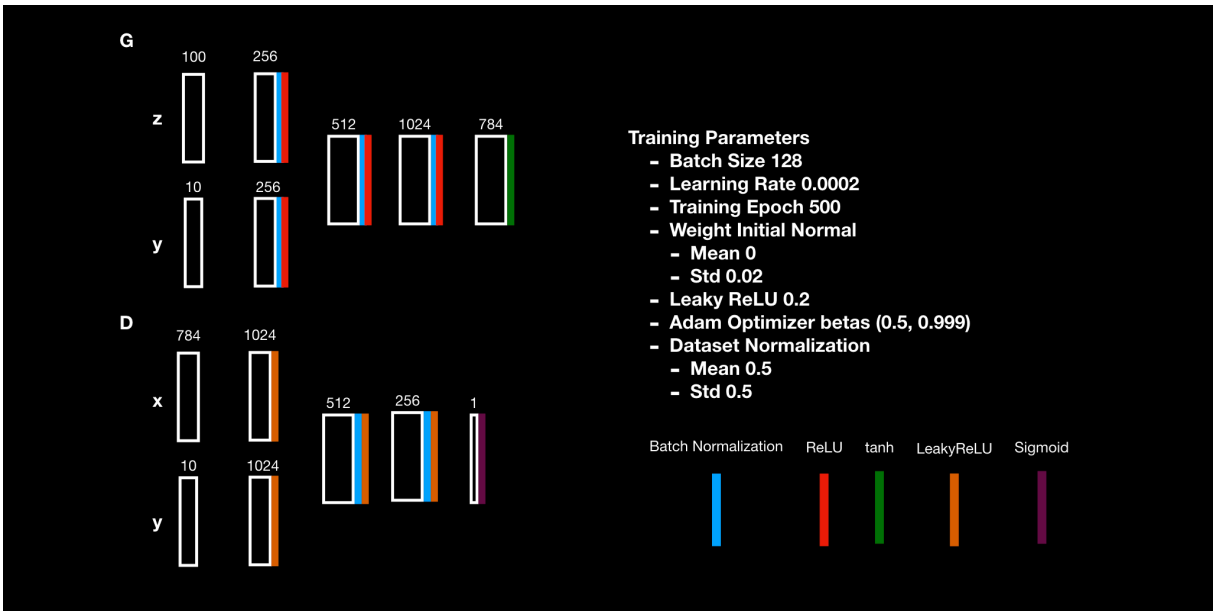

Figure 2: Framework

Experiments & Parameters Modification

1. 基本思路：根据不同情况采取不同的策略灵活调整学习率。

2. 基本策略：我们使用 BCELoss 作为损失函数，以相隔一定间距的损失函数值是否下降为标准，判断损失函数是否已经接近要达到的极小值；已经接近的时候乘一个折合系数。

3. 改进：根据具体在数据集上运行的表现，在不同的迭代次数区间选取不同的判断间距和折合系数。

4. 效果：经过多次调整和尝试，损失值在 50 次迭代之后会下降到 0.15，150 次之后下降到 0.13，200 次之后下降至 0.12，但在 200 步以后直至 500 步损失值都不再有什么变化。

5. 尝试：在损失值长期平稳以后把学习率重新上调，希望摆脱局部最优。

Results

通过大量的实验测试，我们使用 500 次迭代得到的最好结果如图 3，可以看出我们生成的图片可以轻易使用肉眼辨别出各自类别，同时可以看出图片中有足够的噪音可以有效干扰相应图片分类算法对图片分类的判断。

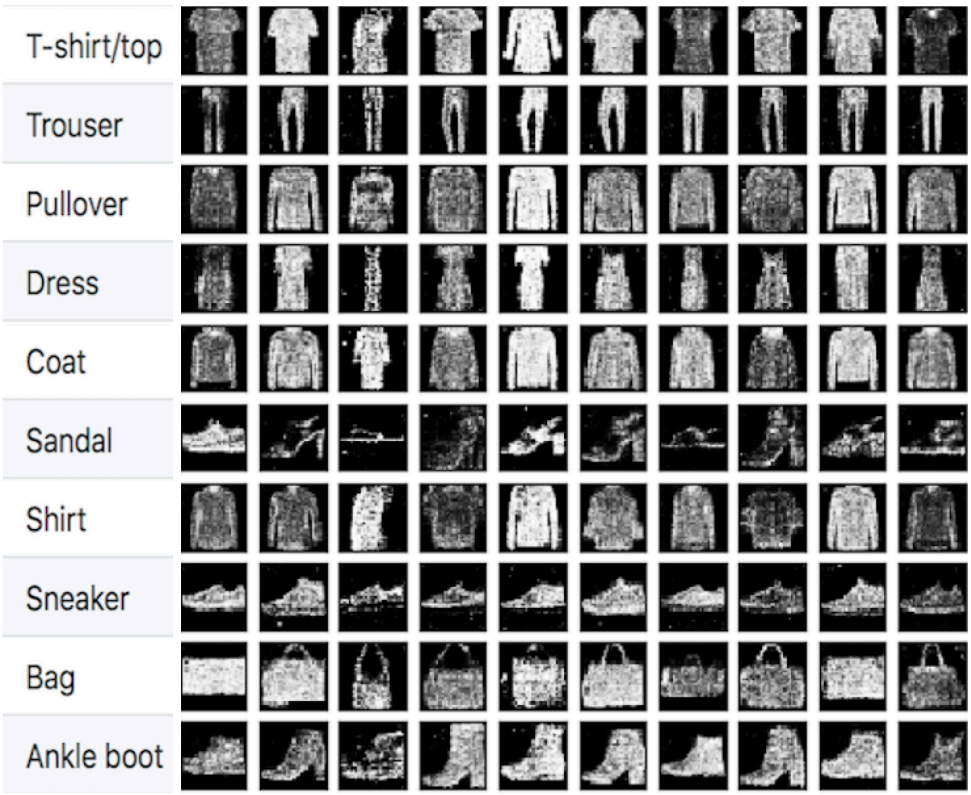

Figure 3: 500 次迭代训练结果

之后，我们对渐进式增长生成对抗网络[3] 文章的源码进行修改使其适应 FashionMNIST 的数据格式，之后使用该程序进行训练，得到生成结果，与我们的结果对比如图 4。

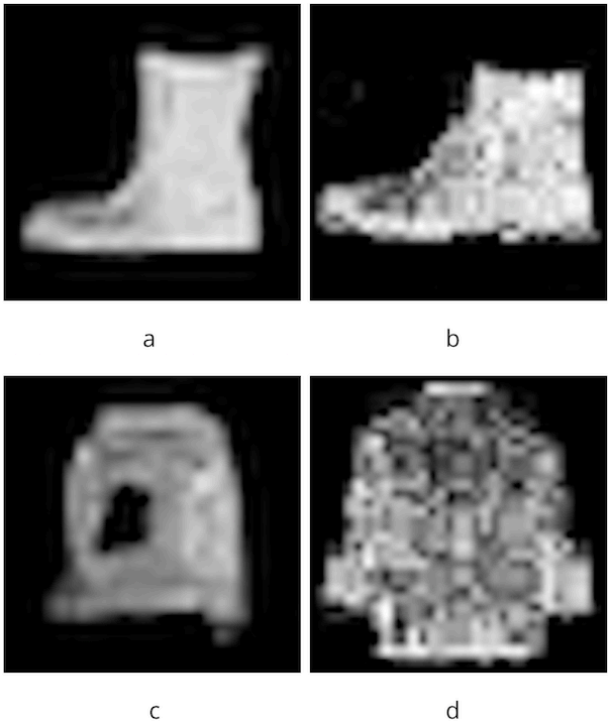

Figure 4: a, c 为使用 PGGAN 生成图片, b, d 为使用 CGAN 生成图片

为了验证我们的方式可以有效地避免图像识别软件对图片的分类识别，我们使用开源的图像识别程序对我们的图片和 MNIST 中的图片进行训练后进行识别，对两种情况下的准确率进行测试比较，识别准确率结果如图 5。可以看出，我们的图像有着很好的优势，在一定程度上干扰了图像识别程序对图像的识别的准确率。由于在我们的实验中使用的是单一的图片进行识别，因此图像识别程序经过一定的训练之后仍然可以实现较高的识别准确率，但是如果将这种方式应用到实际工业中，通过对几张图片进行组合之后构成的图片应该会有更低的准确率，从而实现我们改善图形验证码的目标。

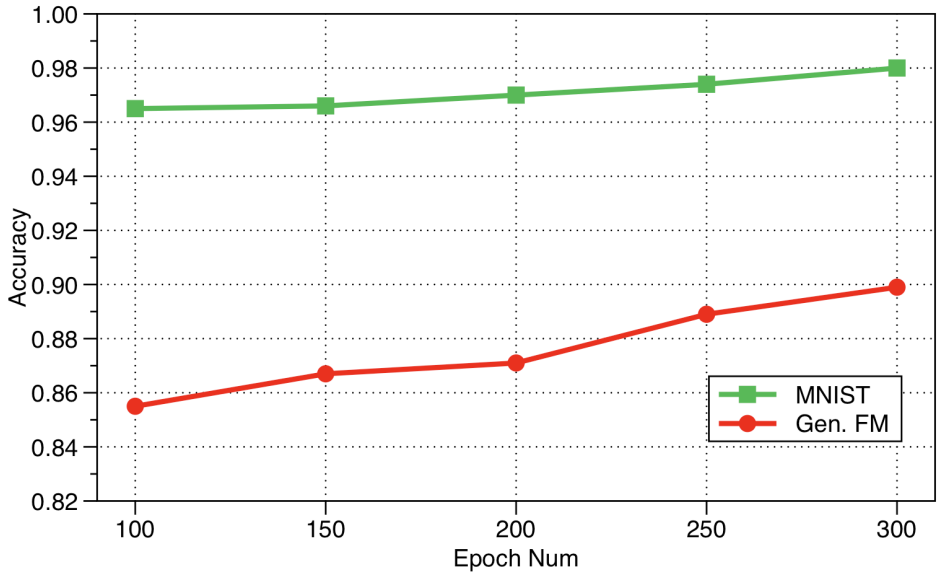

Figure 5: 准确率比较

Conclusions

通过对两种生成方式生成的图片的细节比较，我们得到以下结论：

1. PGGAN[3] 使用从低分辨率图片开始训练的思路，因此对物品轮廓的描述十分清晰，但由于 PGGAN 设计针对高分辨率图片生成，因此在 28×28 情况下对细节的描述并没有突出优势。

2. 我们设计的模型利用实际图片进行训练，噪音较为严重，但是我们的结果具有准确的标签以及较好的细节处理。

3. 因此在我们生成验证码的情形下，我们的模型在结果足够清晰的同时具有实现简单，训练速度较快，生成图片具有明确分类的优势。

4. 通过实验验证，对应相同的图像识别程序，我们的图片可以有效降低识别的准确率。

Forthcoming Research

从实验结果可以看出，我们的模型生成的图片噪点较多，轮廓不够清晰，根据 PGGAN 生成图片的优点，我们考虑可以将渐进增长的训练思路引入我们设计的模型。在我们设计的半监督模型下，从低分辨率开始对数据集进行训练，生成质量更高，更易辨别的验证码图片。同时，我们的图片生成模型并不局限于 FashionMNIST 数据集，也可以推广到更为复杂的数据集上，从而进一步改善生成图片的质量，使其更适用于验证码环境中。

References

[1] Kun Fang, Zhan Bu, and Zheng You Xia. "Segmentation of CAPTCHAs Based on Complex Networks". In: *Artificial Intelligence and Computational Intelligence*. Ed. by Jingsheng Lei et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 735–743. ISBN: 978-3-642-33478-8.

[2] Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[3] Tero Karras et al. "Progressive Growing of GANs for Improved Quality, Stability, and Variation". In: *CoRR* abs/1710.10196 (2017). arXiv: 1710.10196. URL: http://arxiv.org/abs/1710.10196.

[4] Mehdi Mirza and Simon Osindero. "Conditional Generative Adversarial Nets". In: *CoRR* abs/1411.1784 (2014). arXiv: 1411.1784. URL: http://arxiv.org/abs/1411.1784.

[5] Han Xiao, Kashif Rasul, and Roland Vollgraf. "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms". In: *CoRR* abs/1708.07747 (2017). arXiv: 1708.07747. URL: http://arxiv.org/abs/1708.07747.

Personal Summary

在人工智能导论这门课的学习过程中，我们学习了一些经典的人工智能的算法并在课程作业中进行了实践，加深了我们对课堂上学习到的知识的掌握。在大作业的完成中，我们可以开放地进行选题，尝试一些十分新颖的人工智能方面的算法，对这些算法的特性以及实现进行学习并完成自己的项目。同时，最后展示的环节采取了和以往的计算机辅修课程不相同的 POSTER 展示的方式，让我们可以按照自己的兴趣对同学实现的算法进行学习与讨论，让我们认识到人工智能的方法在许多方面的应用以及实现方式，比以往的课堂 ORAL 汇报的方式带来更大的收获。同时，采用小组共同完成一个大作业的方式不但减轻了我们的负担，还允许我们按照自己的兴趣对一个题目中的不同方面进行深入探究。这一系列较为灵活的教学方式给我们带来的很大的收获。

最后，十分感谢老师和助教耐心的回答我们的问题以及十分认真的指导。

Distribution of Work

1. 张知行：主要负责了模型的设计、代码的编写以及实验的方案设计。

2. 钟昊东：主要负责了模型的参数调整以及模型的训练和 POSTER 展示。

3. 朱文轩：主要负责了模型对比方案的文献查找与研究。