使用CGAN对图形验证码的优化
基于*FashionMNIST*数据集

张知行 钟昊东 朱文轩
清华大学 材料学院

Abstract

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。因此我们需要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[2]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很好地验证效果的生成模型PGGAN[3]生成的图片进行对比，说明了我们实现的模型的优势，并通过比较说明了我们的模型可以很好地应用于当前的验证码图片生成当中。

Introduction

机器学习的模型大体上可以分为两类，生成模型与判别模型。其中，判别模型要求输入变量x，通过某种模型来预测$p(y|x)$。但是，生成模型是给定某些隐含信息，从而生成随机数据。这样的学习方式的困难之处在于，人们对生成结果的期望是一种无法使用数学公理化定义的范式。但是对于这种难以公理化的任务来说，判别模型已经有了很好地解决方式。这种将生成模型与判别模型进行结合来进行更好的生成的方式就是GAN[2]。GAN（Generative Adversarial Networks）是GoodFellow在2014年的论文中提出的一种思想，要求使用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。

GAN具有很好地应用前景，从目前的文献来看，GAN 在图像上的应用主要是往图像修改方向发展。涉及的图像修改方向主要包括：单图像超分辨率、交互式图像生成、图像编辑、图像到图像的翻译等。GAN在NLP方向上也有着很好地应用前景。

但是朴素的GAN存在以下几点问题：收敛性难以得到保障，模型崩溃以及模型过于自由不可控。同时，朴素的GAN中生成的结果并不含有标签，无法直接用于我们的场景当中。因此我们选取经过一定的修改的CGAN[4]作为我们的基本模型。CGAN的简要结构如图1所示。

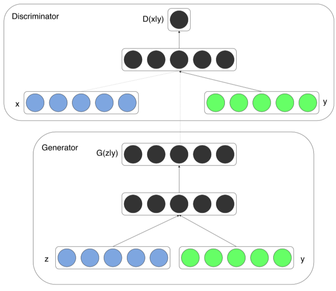

Figure 1: CGAN基本结构[4]

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。

Methods & Experiments

基于[4]中的设计思想，我们设计了2中的网络作为我们的基本模型并在此基础上进行训练。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，24GRAM，GTX 1080Ti的Ubuntu16.04极客云服务器，我们的代码编写环境为Python=3.6。

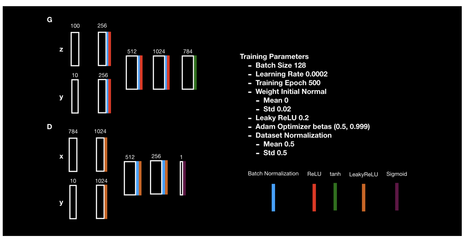

Figure 2: 网络设计及具体参数

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

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

我们对渐进式增长生成对抗网络[3]文章实现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。

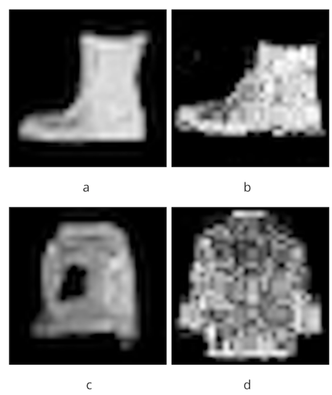

Figure 4: a, c 为使用PGGAN生成图片, b, d 为使用CGAN生成图片

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，但是由于我们引入了图片标签进行训练，因此在细节的处理上仍然有着较好的表现
3. 对于我们生成验证码的情形，我们的模型在结果清晰的同事具有实现简单，训练速度较快，生成图片具有明确分类的优势。

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，我们考虑可以将逐渐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。

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.