有了GAN和VAE的阅读量后，对比GAN和VAE都有各自的优缺点。GAN利用生成对抗网络在图片的生成质量上是很好的，图片清晰特征明显。但是在GAN的训练过程中容易发生崩溃，以及训练时梯度消失情况的发生。生成对抗网络的博弈理论只是单纯的让G生成的图片骗过D，这个会让G钻空子一旦骗过了D不论图像的合不合理就作为输出。以上原因导致GAN生成的图片有的时候看起来有点“乱来”，像图1就是GAN训练的时候出现的问题。

图1.GAN训练失败图

VAE是利用已有图片在encoder下编码生成潜在向量，这个向量在服从高斯分布的情况下很好的保留了原图像的特征在decoder解码后得到的图片会更加的合理与准确。但是图像在训练的时候损失函数只能用 MSE 之类的粗略误差衡量，这就导致生成的图片不能很好地保留原图像的清晰度，就会使得图片看上去有点模糊。优点就是VAE生成的图片合理，学习到的潜在向量可以很好地还原出图像，训练不会出现崩溃的状况。

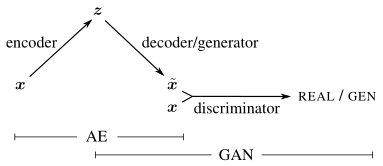
这周我主要的任务是看了VAE与GAN结合的论文，VAE+GAN的目的就是结合VAE和GAN的优点让生成图像在合理的前提下做到模型的稳定与图像的质量上的保证。单纯的[VAE+GAN](https://arxiv.org/pdf/1512.09300.pdf)(Autoencoding beyond pixels using a learned similarity metric)在2016年便提出来了。论文结合了VAE和GAN在无监督学习条件下同时训练了encoder，generator和discriminator来达到图像的生成。

图2.VAE+GAN训练模型

由此模型我们可以看到输入的为样本数据，经过encoder编码后生成一包含特征的潜在向量z，这里的z服从高斯分布，z通过一个generator或者decoder生成图像数据。这里为什么用或呢？因为在后面论文描述的损失函数中要结合decoder与generator两部分，其实这里的其实有两部分，一部分是来自decoder通过高斯分布向量得到，还有就是由generator根据z生成得到的。最后将生成的样本与原始数据送往鉴别器判断真伪，达到VAE与GAN的结合。

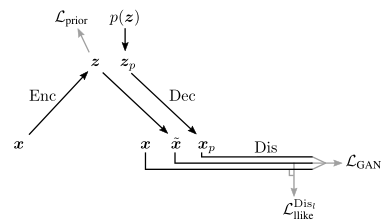
当然这种单纯的结合是不合理的，一个很大的难点就是对训练损失函数的参数的控制上。在具备VAE的理论基础上我们已经推导了损失函数公式，对于由原始数据通过encoder的到服从高斯分布z我们已经很清楚了，这中间的损失就是生成的z与期望分布的z之间的差异的KL距离问题。由z到则可分为两部分，一块是通过生成器的损失部分，这一块可以直接用VAE的损失函数中的解码带来的误差，另一块是带入GAN中的decoder的损失部分，我们用图3帮助说明。

图3.模型损失计算带入模型

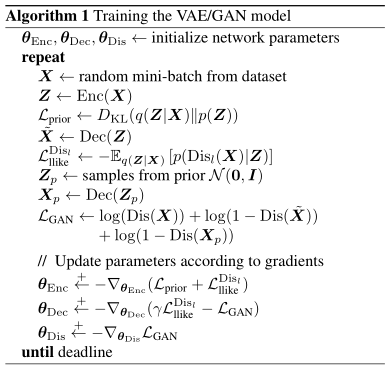
最后根据模型更新损失函数的各个参数达到模型的训练。

图4.训练算法

考虑到报告篇幅，我就不对公式展开分析了，有了这几周的公式梳理我对此类文章中的公式已经基本上可以自行理解了。附录中将附上GAN模型的损失函数的公式推导，有助于对此类文章的公式理解。当然这里我感觉一些论文中的公式改进可能很大一部分来源于在实验中不断的尝试加入与修改函数表达式，在这篇论文中就可以看到这一特征。

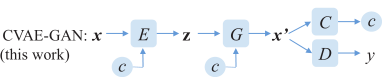
本周还对[CVAE+GAN](https://arxiv.org/pdf/1703.10155.pdf)(CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training)论文做了研读。这篇文章出自中科大研究团队，文章质量很高但是可以很清晰感觉到文章是中国人写的，不谈这个，文章的创新部分还是很有启发的。此片文章在VAE+GAN的基础上新加了一个分类器和类别标签，加入标签的好处就在于可以实现one-to-many mapping problems，论文的模型框架由图5所示。

图5.CVAE+GAN模型结构

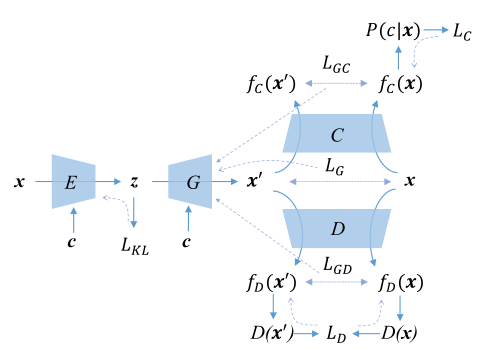
文章的基本原理与上一篇的VAE+GAN很相似，唯独就是加入了标签c，这样在损失函数的这一块将有一个大的改动。当然这也是一个模型最重要的部分。为什么说这篇文章有创新呢就是在损失函数这里引入了新的计算公式，文章称之为均值特征匹配（Mean Feature Matching）。文章中提及正是加入了均值特征匹配才保证CVAE+GAN的结合不再naïve（文中两次提及）。损失函数的部分的解释就要由图6，CVAE+GAN详细模型阐述了。

图6.CVAE+GAN详细训练模型结构

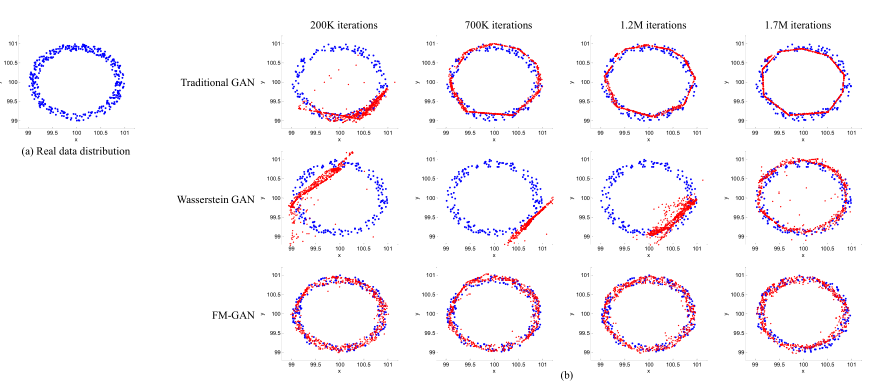
这里的损失计算的考虑很全面，估计是实验一步步下来的总结的吧。模型在generator和generator与classification之间以及generator与discriminator之间都采用了均值特征匹配的方法来计算的损失值。文章为了证实自己的均值特征匹配损失模型（FM-GAN）的优越性还特意做了Toy Example来演示，通过对比确实可以看出在一定程度上效果很不错，图7为实验结果。

图7.Toy Example下不同模型实验结果

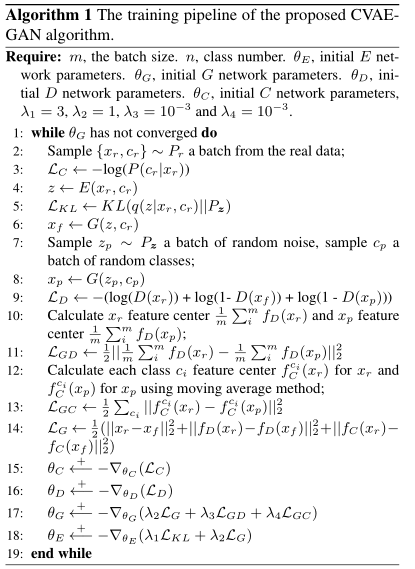
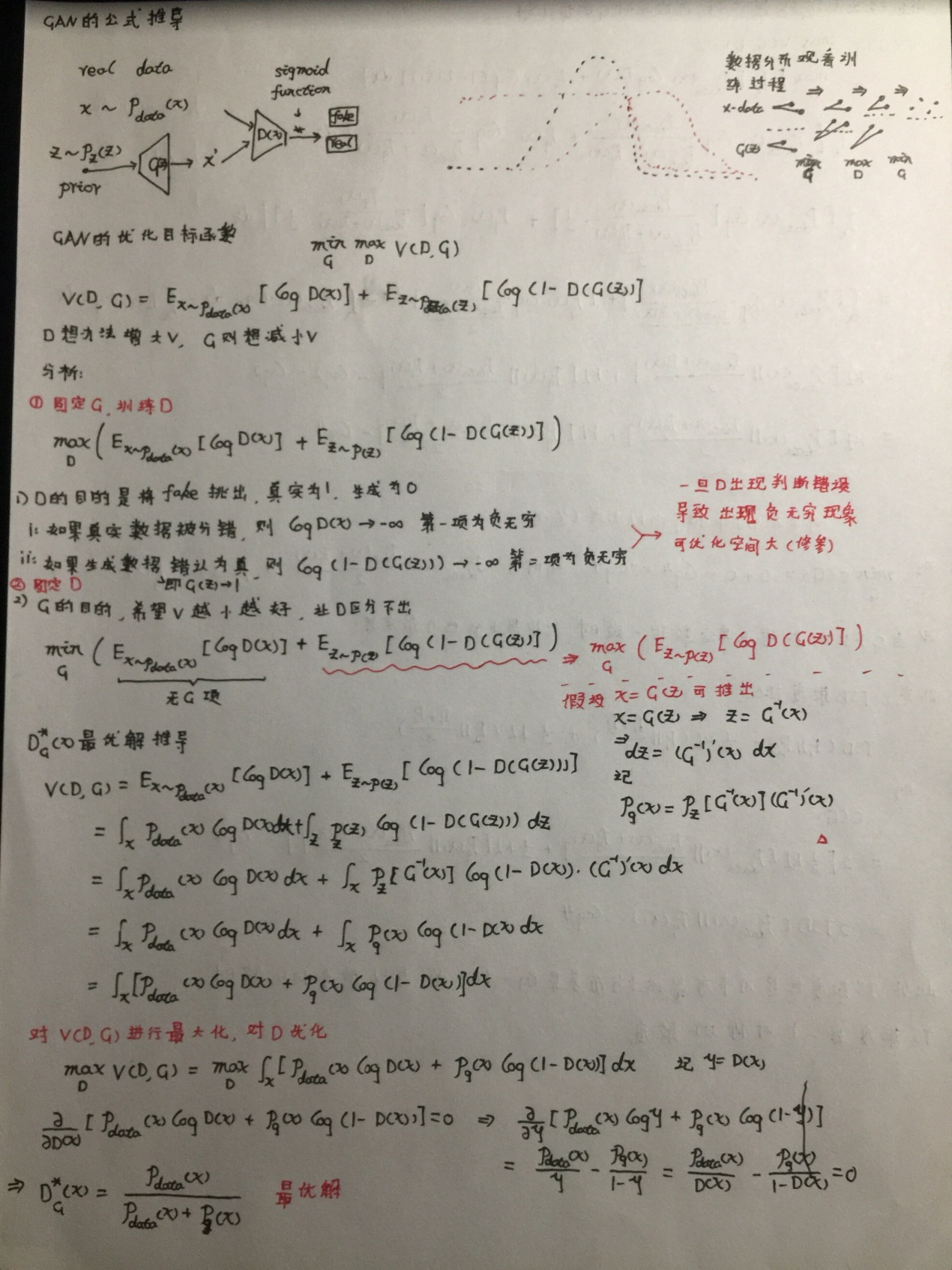
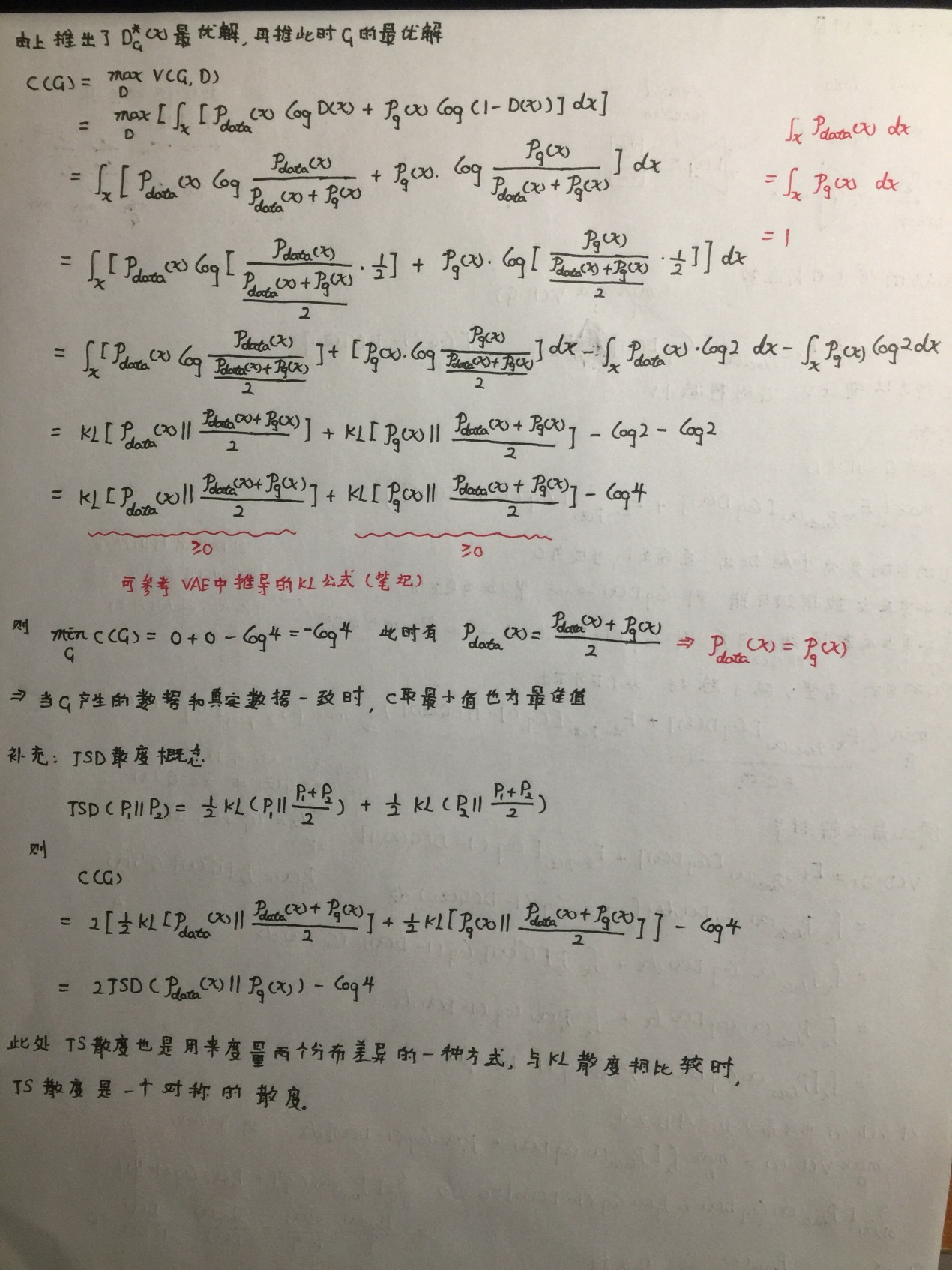
但我感觉在200K iterations下FM-GAN的结果就不错了，这样会不会导致一定程度上的过拟合的产生，这个我后续有时间再去验证吧。最后把CVAE-GAN的算法图8做展示就不详细说明了，可以结合图6做理解。

图8.CVAE+GAN模型训练算法

小结：本周对VAE+GAN的综合应用做了相应的研读，VAE+GAN的训练模型在一定程度上确实实现了将VAE和GAN的优点结合的效果，但是也并不是完全的集两家之所长。训练的过程中也不能完全避免模型的崩溃，其中的原因我感觉存在对抗网络就会有几率出现崩溃。在CVAE+GAN中的一个有心之举就是对各个损失函数的缺失下对结果的影响做了分析，这个确实蛮好的，可以看到具体的损失值在结果上的影响程度，可能在一定程度上有助于调参。本周还对GAN中的损失函数公式做了理论推导，将在附录中展示，有时候公式往往是一篇论文的关键之处，所以还是搞得清楚一些比较好。

下一周我会对之前的各种GAN模型从各个方面做一个比较，从中找出GAN的比较合理的模型和训练方法。再有就是对GAN模型的程序要有自己的一个思维框架出来。这个任务的过程估计不是一周可以完成的，合理安排时间也是比较重要的。

**附录**

****