致谢：

首先要感谢我的指导老师张信明教授，从毕业设计题目的选定，到中期检查时候的方向的微调和后期的论文完成过程中，张老师的悉心教导令我受益匪浅。

其次要感谢中国科大超级计算中心，在前期的工作中，由于自己的笔记本性能不足，在学习过程中实践难度很大，超算中心提供了平台的支持。

还要感谢我的好朋友好同学，前期的实践不仅仅依靠超算中心，也依靠他，借用他的笔记本进行了不少运算。

更要感谢的是我的父母，养育之恩终身难以回报，更是提供了资金更换笔记本电脑，让我得以在自己的计算机上继续毕业设计，免去了超算中心的排队和借用同学电脑的不便之处。

最后感谢这个城市，这所学校，还有身边的同学、朋友，提供了一个合适的平台环境，让我学习成长。

摘要：

1. 绪论
   1. 背景
   2. 对抗生成网络
   3. 本文章节安排

本文对内容做如下安排：

第一章 绪论。介绍对抗生成网络的相关背景以及本人所做毕设的方向。

第二章 对抗生成网络基础。从神经网络入手，介绍神经网络的思想和部分算法，接着是卷积神经网络的算法原理，以及生成对抗网络的基础。

第三章 Wasserstein GAN的关键技术。首先介绍的是本人学习调研的一些生成对抗网络的改进思想，其次着重介绍Wasserstein GAN的改进方法，包括Wasserstein距离的优势和Wasserstein GAN的优势。

第四章 Wasserstein GAN在图像生成上的改进。本章所述为毕设中期的工作，是在图像生成方面对Wasserstein GAN的修改。

第五章 Improved Training of Wasserstein GANs 所做改进。本章所述是提出Wasserstein GAN的作者Martin Arjovsky于三月31日提出的gradient penalty改进。以及结合本人毕业设计之后的网络设计方案。

1. 对抗生成网络基础
   1. 神经网络基础

神经网络是一种模拟生物神经元结构和功能的模型，典型的神经网络模型包含结构、激活函数和学习规则三部分，其作用是通过简单的矩阵运算和非线性的激活运算来对数据进行处理。

下面所述的是一个简单的神经网络。



图1-1 神经网络示意图

如图1-1所示，这是一个简单的神经网络，包含两个输入变量、，三个中间变量、、，一个输出变量。其中所在为输入层，所在为隐藏层，所在为输出层。和被称为权重。计算方式为：

， (1.1)

例如：

(1.2)

为了表达和计算上的方便，用矩阵来表示如上运算过程：

 (1.3)

在实际的运用中，往往还会加上一个偏置项bias，则1.1、1.2、1.3式转变为如下：

， (1.4)

(1.5)

.  (1.6)

对于图1-1，没有表现出偏置项，在实现中，偏置项有多种方式可以添加进去，一种是如式1.6加上一个向量，还可以将偏置项视作1\*w，将偏置项的加法运算放入矩阵相乘中，式1.6转化为：

 (1.7)

其中。这种方式在计算上省去了一步向量相加，但是多出了一步向量的拼接操作，将拼接为了。如图1-2：



图1-2 神经网络示意图，带偏置项bias

在实际的运算中，拼接操作比向量相加消耗更多的计算。于是出现了又一种计算方法。考虑到每次拼接的开销，不妨考虑只在输入增加一个值为1的神经元，之后的除了最后一层的每一层，都附加一个神经元，但是该神经元的值由之前的计算得到，如图1-3：



图1-3 神经网络示意图，转变后的偏置项

相对应的计算公式为：

.  (1.8)

其中，作为隐藏层的偏置项参与运算。

神经网络里面一个举足轻重的反向传播算法back propagation，其作用在于调整网络使之成为我们所需要的状态，简单的来说就是复合函数的链式求导，用梯度下降的方法更新网络，有兴趣的可以参考！！！

* 1. 卷积神经网络

在后文中用到了另一个重要的网络层是卷积层，熟悉图像处理的人知道，卷积操作在图像中十分常见，对图像的平滑、模糊、锐化、边缘检测、浮雕等操作都可以通过特定的卷积算子来实现。下面简略的介绍卷积在图像中的操作方式。

卷积在图像中的操作方式往往是一个2D卷积，对于一张图像和一个卷积核，如图1-4所示，左边是一个5x5的图像，右边是一个3x3的卷积核。



图1-4 图像卷积示意(1)

有一个很形象的说法，将卷积核视为窗口，在图像上滑动，得到卷积后的结果。举个例子，在图1-4中，将卷积核置于图像的左上角，那么运算方式为，每个像素点的值乘以卷积核上与像素位置对应的值，最后累加得到结果。



图1-5 图像卷积示意(2)

运算的式子为：



然后将卷积核向右移动一个，如图1-6：



图1-6 图像卷积示意

对应的运算为：



将卷积核在图像上全部滑动一遍后，得到一个3x3的新图像。这里图像变小了，若是想维持图像大小不变，则在原图像四周补上数据，补充的数据的方式有多种，最简单的是用0填充，卷积操作连边缘都扫过，则可以得到和原图一样大小的图像。上述的卷积的滑动向左和向下的步长都是1，实际上步长还可以为更大的值，比如，步长为2，边缘补上数据，则卷积后的图像大小的长和宽都是原图像的一半。

在实际的运算中，卷积运算会转换成矩阵相乘来实现，首先对原图像和卷积核都做展开，上述的例子则可以转变成：



对结果改变形状成3x3的就得到了最后的结果。

那么，卷积的好处是什么，第一点就是节省空间，若是将一个5x5的图像通过全连接的方式转变成3x3的图像，所谓的全连接，是指上下两层每两个结点都有关联，那么需要权重25x9=225个，如果使用卷积，那么只需要9个，为了充分的获取图像的特征，可以用更多的卷积核来处理同一个图像，即使是这样，也比全连接节约了很多空间。第二点，图像具体局部性，我们在关注图像的时候，通过某一小块区域得到一个信息，这个信息和距离该区域很远的区域关系并不密切，卷积操作很好的关注了局部性的特征。

简略的提一下反卷积操作，反卷积往往用来数据的扩充，卷积和反卷积的输入输出过程刚好相反，卷积的前向计算过程是反卷积的反向传播过程，卷积的反向传播过程是卷积的前向计算过程。

* 1. 生成对抗网络

三年前蒙特利尔大学 Ian Goodfellow 等学者提出“生成对抗网络“(Generative Adversarial Networks，GANs)的概念，去年在学界、业界，人们对GANs的关注突然增加。

什么是生成对抗网络，Ian Goodfellow给出的描述是“生成对抗网络是一种生成模型（Generative Model），其背后基本思想是从训练库里获取很多训练样本，从而学习这些训练案例生成的概率分布。”

对抗生成网络包含两个网络，一个是判别器Discriminator，一个是生成器Generator。生成器将输入的噪声生成和样本一样格式的数据，判别器则是要区分真实数据和生成器生成的伪造数据。传统的GAN通过如下的方式训练：

判别器接收一组真实数据real data，进行计算之后的结果和1进行比较，计算差距，记为判别器对真实数据的损失函数dis\_loss\_real。生成器生成一组伪造数据fake data，传入判别器进行计算，得到的结果和0进行比较，计算差距，记为判别器对伪造数据的损失函数dis\_loss\_fake；得到的结果和1进行比较，计算差距，记为生成器的损失函数。直观上的理解则是判别器想尽可能的区分真实数据和伪造数据，所以想得到对真实数据判定为1、对伪造数据判定为0的结果；生成器则是希望自己生成的数据被认定为真实数据1。



图1-7 生成对抗网络示意

由上述的描述可以得知，生成器需要联合判别器一起才能训练生成器，判别器也需要生成器给提供fake data。训练生成器的时候，判别网络不做更新；训练判别器的时候，生成网络不做更新。两者交替训练，达到生成器生成的数据可以以假乱真的地步。

关于生成对抗网络，已经有很多的改进的版本，比如ali-bi gan、boundary equilibrium gan、conditional gan、infogan等等，它们大都是在损失函数的计算方式和网络结构上做了一些改动，本质上没有改变GAN的性质。

1. Wasserstein GAN的关键技术
   1. 以往的生成对抗网络的改进方案

先看对损失函数的改进方案。考虑原始的GAN，判别器最后的输出是一个数据对应一个输出值，对这个值做sigmoid激活，结果和0或者1比较计算交叉熵，在tensorflow中，使用的是sigmoid\_cross\_entropy\_with\_logits函数，传入的两个参数，logits是判别器输出的结果（不需要sigmoid激活，因为这个函数自带sigmoid激活），labels是同等规模的0或者1。

在调研中，发现对此的改进主要有三个方向，一种是使用sigmoid激活之后用其他函数来处理，第二种是对输出的结果直接做处理，第三种，引入其他的度量标准，修改损失函数。

第一种以ali bigan为例，sigmoid激活后的结果在0~1之间，用log函数可以拉开差距，设sigmoid激活后的结果为X，判别器用的是log(X\_real)+log(1-X\_fake)，生成器用的是log(X\_fake)+log(1-X\_real)，对结果计算均值之后取负号，最小化损失函数，则可以起到对抗训练的效果，判别器想让X\_real向1靠近，让X\_fake向0靠近，生成器则做相反的事情。在Least Square GAN中，对判别器输出的sigmoid激活后的结果，做的操作是：判别器对真实结果减去1再平方，对伪造结果直接平方，生成器对伪造结果减去1再平方，最后的思想还是在于判别器让真实数据的结果逼近1，伪造数据的结果逼近0，生成器让伪造的结果逼近1。

第二种，对输出的结果直接处理的，有Auxiliary Classifier GAN，真实数据是有不同的标签的，对标签进行独热码编码，将判别器输出的结果进行softmax之后，再与编码后的标签计算交叉熵。判别器对真实数据和伪造数据以及对应的标签，都希望得到正确的分类，但是，这样一来无法区分真实数据和伪造数据，于是还需要最小化伪造数据经过判别器之后的结果，这就要求判别器有两组输出，一组是分类，长度为独热码的长度，另一组是一个值，经sigmoid激活之后和前面用相似的方法，判别器需要让伪造数据的第二种输出尽可能接近0，生成器则是希望分类正确并且输出接近1。

至于第三种，方法比较杂乱，有像MaGAN那样给一个阈值m，来对fake data的判断结果做限定，也有像Mode Regularized GAN那样给出通过比较真实数据和伪造数据的差距，来优化网络的。因为方法繁多，并且还使用了其他的优化，在此不做详细说明。

除了改进损失函数的计算方法，另一大类是对网络做修改，并且两种方法往往结合使用。在修改网络的优化中，不得不提的是Conditional GAN，即条件生成对抗网络。一大特点就是将数据的标签也引入到判别器和生成器中，与Auxiliary Classifier GAN中使用标签的方法不同的是，Condition GAN在数据的输入中就加入标签，甚至会在网络的内部层次中将标签数据拼接上去，达到引入数据和控制网络方向的作用。举一个具体的例子，在手写数字的生成中，对于判别器，输入的是图像和标签，对标签往往会进行独热码标签，一方面是加强标签的作用，另一方面是便于数据的拼接操作。在图像和标签的共同作用下，判别器把真实数据判定为1。对于生成器，噪声和标签作为输入，得到的是图像，再将图像和对应的标签放入判别器，训练生成器让判断结果接近1，训练判别器则让结果接近0。DCGAN在手写数字生成上效果很好，生成的数字肉眼几乎无法判断是真是假，其中有个优化就是引入了标签。

* 1. Wasserstein距离的优势

Wasserstein距离又叫Earth-Mover（EM）距离，定义如下：

 (3.1)

看起来很复杂的定义需要解释，表示和组合的所有可能的联合分布的集合，意味着中，每一个分布的边缘分布都是和。对于每一个可能的联合分布来说，可以从中采样得到一个真实样本和一个生成样本，并计算这对样本的距离，因此可以计算该联合分布下样本对距离的期望值。在所有可能的联合分布中对这个期望取下界，定义成Wasserstein距离。

上面的说法确实晦涩难懂，关于EM距离，有个形象的说明，在这个“路径规划”下，把这堆“沙土”挪到位置所需要的消耗看作，那么则是“最优路径规划”下的“最小消耗”，故此成为Earth-Mover（推土机）距离。

在原始的对抗生成网络中，当判别器近似最优的时候，最小化生成器的损失函数等价于最小化和之间的JS散度。问题是，和之间几乎不可能有不可忽略的重叠，所以JS散度为常数log2，对常数的求导有个显然的结果，这样以来生成器的梯度近似为0了，出现梯度消失问题，无法继续有效的训练下去。

原始的GAN不稳定，原因是要么判别器训练的太好，生成器梯度消失严重，生成器的loss降不下去；要么是判别器训练的不够，生成器梯度不准确，导致梯度下降的方向乱跑。故此需要判别器不好不坏的，才能让生成器有个稳定并且不消失的梯度，找到这个不好不坏的判别器很困难，训练的前期和后期都可能不一样，也就导致设想很好的GAN训练起来很困难。

那么Wasserstein距离的优势在于，相比JS散度，即使和分布没有重叠，它都可以反应它们的远近。并且，相比JS散度的突变，Wasserstein距离是平滑的，能够提供梯度。

尽管Wasserstein距离很有优势，但是实际的计算中，没法直接求解，作者用一个已有的定理将Wasserstein距离转变为了如下形式：

 (3.2)

式(3.2)表示，在保证函数的Lipschitz常数（存在常熟，使得在定义域内，连续函数，任取，满足）不超过K的条件下，对所有可能满足条件的函数取到的上界，然后再除以。式(3.2)变成：

 (3.3)

利用神经网络强大的拟合能力，找到一系列的足以高度近似公式3.2要求的不是难事。至于，对权重做范围的限制，即令所有的参数不超过范围[-c,c]即可保证关于输入样本的导数也不会超过某个范围。在tensorflow中，使用clip\_by\_value函数对权重进行限制。

* 1. Wasserstein GAN的优势

之前提到的，原始GAN的问题有：判别器越好，生成器梯度消失越严重。最小化生成器的损失函数loss，会等价于最小化一个不合理的距离衡量，导致两个问题，梯度不稳定和多样性不足。这里的多样性指的是和真实数据相比，生成的结果可能偏向于某一种或者少量的几种，例如手写数字的生成中，原始的GAN在没有标签控制生成方向的时候，结果会向着特征简单的方向生成，比如生成的很多都是1，而那些特征复杂的数字，生成的就很少。

关于手写数字mnist的数据集，已经有效果较好的深度卷积生成对抗网络DCGAN来生成数据，但是由于网络相对复杂，训练的耗时较长。在下一章将会讲到本人在一些网络结构上进行的实验以及优化。

Wasserstein GAN与原始的GAN相比，只修改了四点：

* 判别器的最后一层去掉sigmoid
* 生成器和判别器的loss不取log
* 每次更新判别器的参数之后把它们的绝对值截断到不超过一个固定常熟c
* 不要用基于动量的优化算法（包括momentum和Adam），推荐使用RMSProp，SGD也行。

那么Wasserstein GAN的优势在于什么呢，在实验中发现，对于已有的较为成熟有效的生成对抗网络模型中，在不改变网络结构的情况下，使用Wasserstein GAN可以达到相同的效果。但是，较为成熟的生成对抗网络模型中，使用了很多的优化方法，诸如归一化等操作。比如DCGAN中，使用了反卷积和归一化，并且引入标签来补充数据，去掉这些，模型将会无法正常工作，但是，使用Wasserstein GAN，即使很简单的全链接网络，没有归一化，简单的leaky relu激活，就可以生成较好的图片。此外，在没有标签参与的情况下，多数GAN都出现了多样性不足的情况，Wasserstein GAN却避开了这一点。此外，Wasserstein GAN的损失函数可以指示训练进度，传统的GAN的损失函数上下抖动的较为剧烈，无法作为指示。

1. Wasserstein GAN在图像生成上的改进
   1. 深度卷积神经网络识别验证码

正如程序员的第一课是hello world，神经网络的第一课是手写数字识别，使用的数据集是mnist，本人在实践了caffe训练并识别mnist数据集之后，修改网络结构，做学校教务系统验证码的识别，并移植到tensorflow上，因为python语言的灵活性，以下全部用tensorflow作为框架来描述。

1. 网络构建

首先先描述数据集的特征，验证码图片从<http://mis.teach.ustc.edu.cn/randomImage.do>这个地址抓取，这个地址是原先学校教务系统本科生的地址，现在已经弃用，先抓了一百张，每一张验证码由四个字母组成，图像大小是20x80的，水平四等分之后差不多能把验证码分隔开。验证码包含数字和大写字母，除了数字0、1和字母I、O，也就是一共32种字符，用python将验证码转化为灰度图做阈值分割并且四等分之后，手动给分割后的验证码分类，并制作标签，用python转化成mat文件格式（matlab的文件格式）。

C:\Users\monk\AppData\Local\Microsoft\Windows\INetCache\Content.Word\tmp.jpg

图4-1 验证码原图示例

C:\Users\monk\AppData\Local\Microsoft\Windows\INetCache\Content.Word\1.jpg C:\Users\monk\AppData\Local\Microsoft\Windows\INetCache\Content.Word\2.jpg C:\Users\monk\AppData\Local\Microsoft\Windows\INetCache\Content.Word\3.jpg



图4-2 验证码初步处理并做分割后的图片

参考lenet对手写数字做分类的网络模型，我自己的网络模型如下：

1. 数据输入：20\*20\*1的图像，20\*20是图像尺寸，1表示图像是单通道的灰度图
2. 第一层卷积层，卷积核大小是5\*5，卷积核个数20，卷积步长长和宽两个方向上都为1，卷积的结果加上偏置项，偏置项的长度和卷积核个数相同。一张图像经过20个卷积核之后，得到20张图像，每一张图像都加上各自的偏置项，一张图的所有像素点的数据所加的值相同。卷积的设置是不填充边界的，因此卷积之后的结果规模是16\*16\*20。
3. 第二层激活层，对数据进行relu激活，即只保留正数部分，负数部分置为0，数据规模不变化。
4. 第三层池化层，使用最大化池化，池化的核大小为2\*2，步长为2，意味着图像缩小为原来的四分之一，并且每个像素点对应的是原来4个像素点中的最大值。此时数据规模是8\*8\*20。
5. 第四层卷积层，卷积核大小仍是5\*5，卷积核个数50，步长为1，得到的结果仍旧加上偏置项。每个核都对20个从上一层接收的8\*8的图像做处理，将20张图像按权重线性累加得到当前卷积核输出的结果，累加操作由tensorflow框架内部实现。卷积图像不填充边界，得到的数据结果规模是4\*4\*50。
6. 第五层激活层，relu激活，数据规模仍是4\*4\*50。
7. 第六层池化层，池化方式和第四层相同，输出的数据规模结果是2\*2\*50。
8. 第七层全连接层，将第六层得到的数据转化为一个规模为2\*2\*50=200的一维向量，叉乘一个200\*500的矩阵，加上长为500的偏置项，得到长为500的一维向量。
9. 第八层激活层，relu激活，只保留正数，负数置为0。输出的数据规模是一维的长500的向量。
10. 第九层全连接层，上层结果叉乘500\*32的矩阵之后加上长32的偏置项，最后得到长32的一维向量。
11. 第十层softmax层，对上层结果做softmax回归，得到类似独热码的结果，其中一个数较大，接近1，其他数较小，接近0，数据都在[0,1]之间。
12. 接下来计算损失函数，将进行独热码编码后的真实标签与取对数后的上一层结果相乘并累加，取相反数作为损失函数。
13. 以上是训练的网络，训练过程中计算准确率，将第十层的结果中最大的数的索引和独热码编码后的真实标签中为1的数的索引进行比较，统计相等的个数。
14. 关键程序实现

训练模型的输入为图像数据和标签。以下按照本节a）中所述展示关键代码。

1. x = tf.placeholder(tf.float32, [**None**, 400], **"x"**)

y = tf.placeholder(tf.float32, [**None**, 32], **"y"**)

h0 = tf.reshape(x, shape=[-1, 20, 20, 1], name=**"reshape"**)

1. h1 = tf.nn.conv2d(h0, filter=W1, strides=[1, 1, 1, 1], padding=**"VALID"**) + B1
2. h2 = tf.nn.relu(h1)
3. h3 = tf.nn.max\_pool(h2, ksize=[1, 2, 2, 1], strides=[1, 2, 2, 1], padding=**"VALID"**)
4. h4 = tf.nn.conv2d(h3, filter=W2, strides=[1, 1, 1, 1], padding=**"VALID"**) + B2
5. h5 = tf.nn.relu(h4)
6. h6 = tf.nn.max\_pool(h5, ksize=[1, 2, 2, 1], strides=[1, 2, 2, 1], padding=**"VALID"**)
7. h7 = tf.reshape(h6, shape=[-1, 2 \* 2 \* self.dim\_2], name=**"reshape"**)

h8 = tf.matmul(h7, W3) + B3

1. h9 = tf.nn.relu(h8)
2. h10 = tf.matmul(h9, W4) + B4
3. y\_ = tf.nn.softmax(h10)
4. loss = -tf.reduce\_sum(y \* tf.log(y\_))
5. accuarcy = tf.reduce\_mean(tf.cast(tf.equal(tf.argmax(y\_, axis=1), tf.argmax(y, axis=1)), tf.float32))

训练使用AdamOptimizer算法梯度下降最小化loss。

train\_operate = tf.train.AdamOptimizer(learning\_rate).minimize(loss)

训练过程如下

sess.run(train\_operate, feed\_dict={x: input\_x, y: input\_y})

1. 实验结果

使用tensorflow自带的工具输出了loss和准确率：



图4-3 训练过程中的损失函数loss和准确率

可见，本次的训练结果非常好，损失函数接近0，准确率近乎为1。

保存模型并另抓取验证码图片测试：



图4-4 抓取的验证码图片和识别结果

实际上，在实验中，最初获取的数据较少，约400张，进行手工的分类，用手动分类之后的结果做训练，准确率大概为92%。抓取更多的数据进行分类，再人工检查，修正一些错误，用新的数据进行训练，准确率继续上升。在准确率达到97%左右的时候，用模型去检测数据集，又找出了之前没有发现的数据分类的错误。直到准确率达到近乎100%，已经超过人的识别准确度。

* 1. 深度卷积网络生成验证码

1. Wasserstein GAN的关键技术
2. Wasserstein GAN的关键技术
3. Wasserstein GAN的关键技术