|  |
| --- |
| 电子科技大学  **计算机专业类课程** |
| **实验报告** |
| **课程名称：人工智能综合实验II**  **学院专业：计算机科学与工程学院**  **学生姓名：崔若晨**  **学　　号：2021060902008**  **指导教师：石小爽** |
| **日　　期： 2023 年 9 月 19 日** |
|  |
| 电子科技大学计算机学院实验中心 |

**电 子 科 技 大 学**

**实 验 报 告**

**实验四**

# 一、实验室名称：

电子科技大学清水河校区主楼A2-413-1

# 二、实验项目名称：

人工智能实验II-4：基于自动编码器的手写数字生成

# 三、实验原理：

自动编码器是无监督学习方法中的一种结构。它通过编码器提取输入数据的隐含特征，再通过解码器根据隐含特征重构输入，实现了一种自监督学习的方法。其中，编码器和解码器通常为一个非线性映射函数，隐含特征往往比输入数据更加紧凑。因此，自动编码器能够实现数据降维、压缩和数据隐含分布的投影。随着深度神经网络的广泛应用，人们尝试使用神经网络搭建编码器和解码器，并取得了广泛成功。本次实验基于神经网络搭建自动编码器，以探究其在隐变量映射和图像降维与还原任务中体现出的各种特性。

# 四、实验目的：

通过使用 PyTorch 深度学习框架，搭建神经网络，实现自动编码器以进行手写数字生成等任务，并在此过程中掌握：

1. PyTorch 中全连接网络、卷积网络等的实现方法；
2. 自动编码器的基本框架和设计思路；
3. 探究自动编码器在图像降维和还原中体现的具体功能。

扩展掌握：

1. 自动编码器的隐变量可视化分析；
2. 变分自编码器的实现方式；
3. 对抗生成网络的实现方式。

# 五、实验内容：

* 1. 问题描述

设计自动编码器（AutoEncoder）生成MNIST和CIFAR数据集图像

* 1. 算法的概要设计与分析

AutoEncoder：将数据集中的数据输入Encoder中进行编码，然后将Encoder的结果输入的Decoder中进行解码，生成图片。Loss为生成图片和原图片基于loss函数的差异，然后通过反向传播算法更新网络参数。

VAE：将数据集中的数据输入Encoder编码为k个均值和k个方差，然后将k个对应的正态分布输入的Decoder中进行解码，生成图片。Loss为生成图片和原图片基于loss函数的差异加上0.1权重的先验分布和变分后验分布的klloss，然后通过反向传播算法更新网络参数。

GAN：由生成器和判别器“左右互博”。我们把伪造的图片和真实的图片交给判别器，并训练判别器辨认哪些图片是伪造的，哪些是真实的。生成器则恰恰相反，它根据判别器的判断来调整自己，尽可能去欺骗判别器。具体操作时：根据随机噪声，生成器生成一组伪造图片。把伪造图片和真实图片交给判别器，并告诉它真实标签，训练进行二分类，判断图片是“真”还是“假”。生成器再生成一组图片，同样交给判别器，但这时我们训练去干扰的分类结果。

MS-SSIM算法：用于衡量两幅图像之间的差异。该算法相似度的测量可由三种对比模块组成，分别为：亮度，对比度，结构。在本实验中用MS-SSIM算法来比较自编码器生成图片和原始图片之间的差异从而来评价生成图片的质量。

1. 核心算法的详细设计与实现

AutoEncoder:如图 1所示，AutoEncoder分为encoder和decoder两部分。Encoder对数据进行编码,通常先提取特征，然后对特征进行编码。常用的特征提取方式有线性层网络（MLP）和卷积神经网络（CNN），特征提取后通常用全连接层（FC）对特征进行编码。Decode根据Encoder决定解码方式。如果采用线性层网络，则Decoder也用线性层网络进行解码。如果Encoder采用卷积网络，则Decoder通过反卷积（ConvTranspose）对特征进行解码。

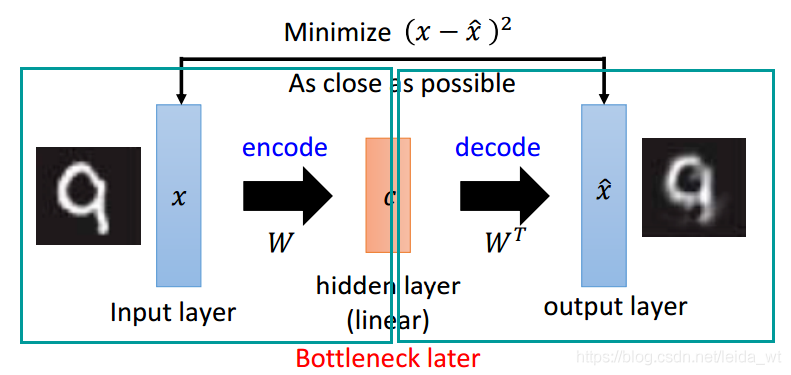


图 1 AutoEncoder结构图

反卷积操作如图 2所示。首先，对输入的feature map进行padding操作，得到新的feature map；然后，随机初始化一定尺寸的卷积核；最后，用随机初始化的一定尺寸的卷积核在新的feature map上进行卷积操作。当s=1时，对于原feature map不进行插值操作，只进行padding操作，这里的padding是卷积操作时设置的padding值，经过padding之后，再经过卷积核进行卷积操作就得到了输出结果。当s>1时，进行插值操作，feature map的元素之间被分隔开了，间隔(s-1)个插入元素0。插值之后的操作同s=1的操作一致。

|  |  |
| --- | --- |
| * 1. S=1 | * 1. S>1 |

图 2 反卷积（ConvTranspose）示意图

Decoder最后要使用Sigmoid函数，将输出固定在0~1之间，这是因为输出的是图像，其像素点大小需要在0~1之间。损失为Decoder生成的图片和真实图片之间的差异，常见的损失函数有MSEloss,binary\_cross\_entropy等。计算出损失后，使用反向传播算法同时对Encoder的参数和Decoder的参数进行更新。

VAE：变分自编码器。它能够学习一个函数/模型，使得输出数据的分布尽可能的逼近原始数据分布。相比AE，VAE不再把分布映射到固定值或变量上，而是映射到一个分布上。如图 3，网络模型中的bottle neck被分解成了两个向量：均值向量方差向量。VAE的Encoder用来计算均值和方差，均值和方差是用来描述正态分布的数字特征，不同的和取值对应不同位置、不同形状的正态分布。这里的均值和方差并不是确定的统计量而是两个函数，是由神经网络训练得到的。其中后验分布是先验分布的近似；是服从标准多元高斯分布的标准化随机变量，是服从高斯分布的随机变量，称为潜在变量或隐变量；是似然分布，即知道变量的分布后根据来估计的分布。当解码时，我们将从每个潜在分布中随机采样，生成一个向量作为解码器模型的输入。由于有了每个属性的概率分布，我们可以简单地从分布中抽取任何值来生成一个新的输出，而且自带过渡效果。

VAE是一种无监督的生成模型，其理论基础是建立在GMM高斯混合模型之上。GMM认为任何一个数据的分布，都可以看作是若干高斯分布的叠加。是由一个标准正态分布所产生的向量，的每一维度都代表了一种属性，且的每一个属性的分布都服从高维高斯分布，通过每一个属性的高维高斯分布来拟合输入样本的真实分布、对于每个采样，会有两个函数和，分别对应到高斯分布的均值和方差，然后在积分域上所有的高斯分布的累加就得到

原始分布的近似，即，整个网络就是把样本经过

降维表达成，再由根据低维特征重构原始输入，对应了一个完整的Encoder-Decoder过程。

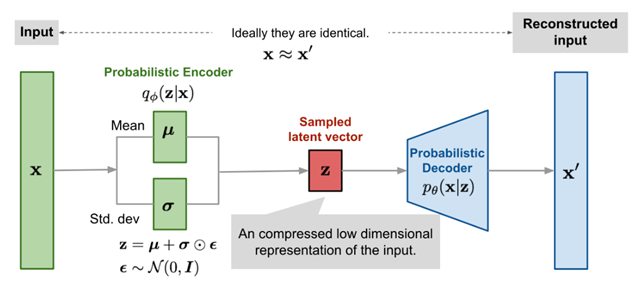


图 3 VAE结构图

我们的目的就是求解的分布。由于是已知的，未知而，于是我们真正要求解的是和两个函数的表达式。通常非常复杂，因而引入两个神经网络进行求解：第一个神经网络是Decoder，用于求解和两个函数，等价于求解。第二个神经网络是Encoder，用于求解，可以表示任何的分布。为了让尽可能接近，我们引入KL散度。KL散度是两个分布之间差异的度量其公式为：



通过最小化KL散度，使两个分布尽可能地相似。因而模型的损失函数除了binary\_cross\_entropy之外，还要加上KL散度，这样最小化损失函数既能是让生成的图片跟真实图片接近又能使生成的分布与原分布尽可能接近，从而提高模型的鲁棒性。

MS-SSIM算法：用于衡量两幅图像之间的差异。该算法相似度的测量可由三种对比模块组成，分别为：亮度，对比度，结构。亮度对比函数为：



其中，为平均灰度，常数是为了避免接近接近0时造成系统的不稳定。通常取

对比度函数为：



其中，为灰度标准差，常数是为了避免接近接近0时造成系统的不稳定。通常取

# 六、实验器材（设备、元器件）：

硬件平台：PC端

开发环境：操作系统Windows 11

测试环境：Anaconda 3，Python 3.11，PyTorch 2.01

# 七、实验步骤：

注：该红色部分的文字在最终版需要删除

网络框架结构搭建介绍

**给出网络设计及核心代码实现，如：**

* 1. **网络类型（全连接或卷积？）；**
  2. **网络每一层的输入输出维度；**
  3. **网络每一层的可学习参数维度等。**

训练执行流程介绍

# 八、实验数据及结果分析：

1. 实验数据

MNIST数据集一共有70000张图片，其中60000张是训练集，10000张是测试集。每张图片是2828的0-9的手写数字图片组成。每个图片是黑底白字的形式，黑底用0表示，白字用0-1之间的浮点数表示，越接近1，颜色越白。如图 4所示。

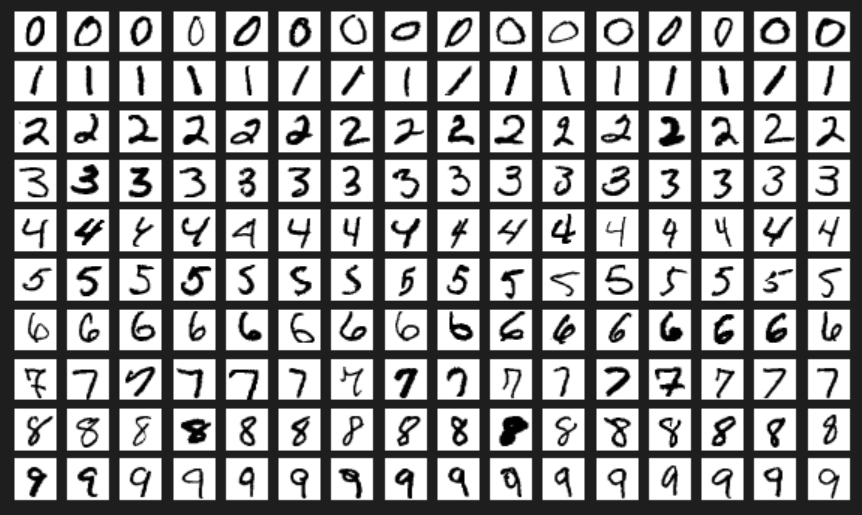


图 4 MNIST数据集

CIFAR-10 是一个包含60000张图片的数据集。如图 5 CIFAR10数据集，其中每张照片为32\*32的彩色照片，每个像素点包括RGB三个数值，数值范围 0 ~ 255。所有照片分属10个不同的类别，分别是 'airplane', 'automobile', 'bird', 'cat', 'deer', 'dog', 'frog', 'horse', 'ship', 'truck'。其中训练集有50000张图片，测试集有10000张图片。

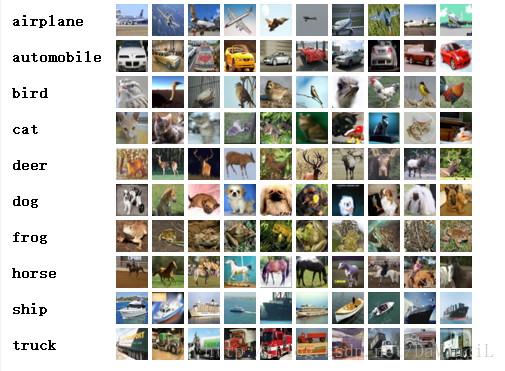


图 5 CIFAR10数据集

2. 结果分析

如图 6为使用线性层AutoEncoder生成结果和loss变化曲线。使用了Adam优化器，学习率为5e-4，一共训练了10000次，最终的loss为0.1215。

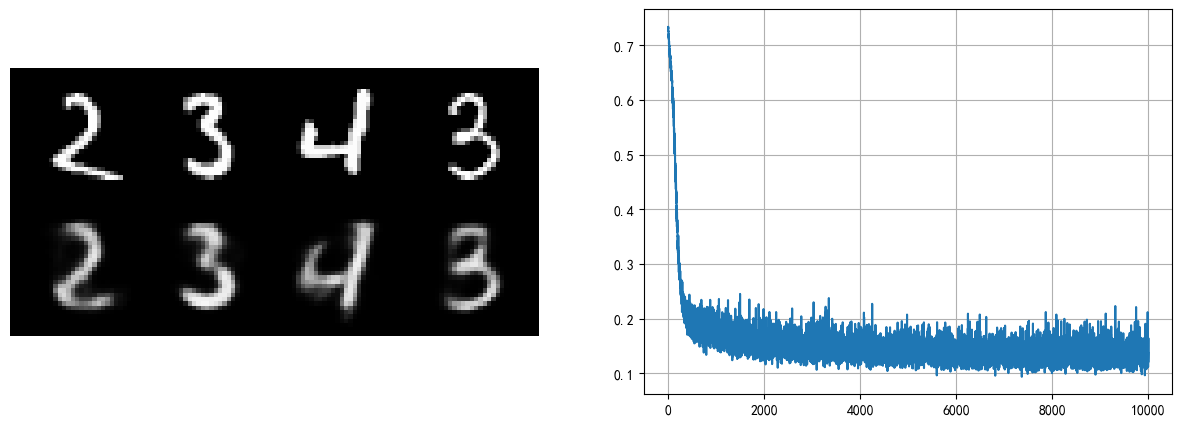


图 6 线性层AutoEncoder生成结果及loss曲线

图 7为使用卷积层AutoEncoder生成结果和loss变化曲线。使用了Adam优化器，学习率为1e-3，一共训练了1000次，最终的loss为0.0675，相比于线性层AutoEncoder,卷积层AutoEncoder生成图片效果更好，最终的loss也更低，并且从loss图中可以看出，卷积层AutoEncoder的loss下降更快，后面的波动也相对更平缓一些。

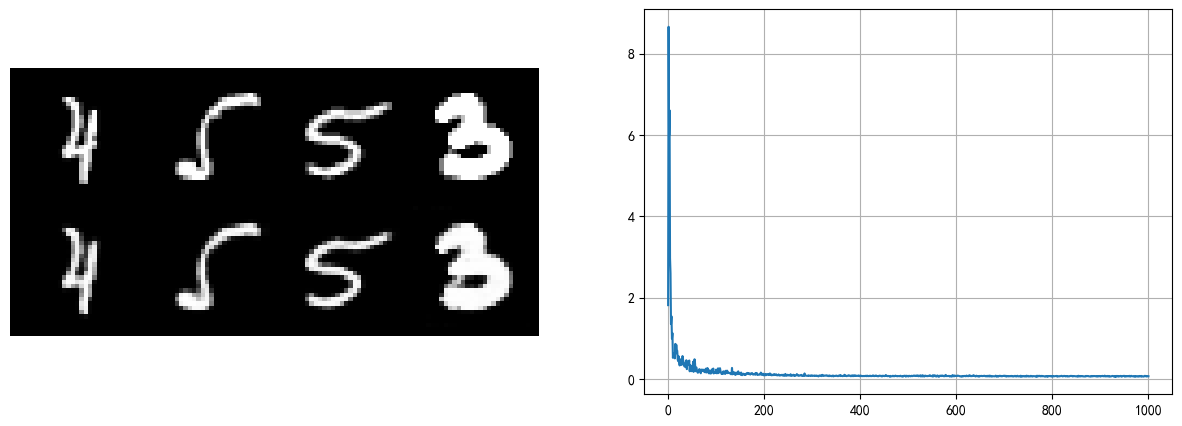


图 7 卷积层AutoEncoder生成结果及loss曲线

上面两个实验均使用了binary\_cross\_entropy作为损失函数，图 8为使用MSELoss后，卷积层AutoEncoder生成结果和loss变化曲线。仍然使用Adam优化器，学习率为1e-3，一共训练了1000次，最终的loss为0.0027。相比于使用binary\_cross\_entropy作为损失函数的AutoEncoder，使用MSELoss生成图片效果更好，最终的loss也更低。这主要AutoEncoder通过生成图片与训练图片之间的loss来更新网络参数。binary\_cross\_entropy更适合于分类任务，也就是它能更好的区分两个样本是否属于同一类。而MSELoss通过L2范数计算生成图片和训练图片之间的差异。显然，直接计算二者的差异比判断二者是否属于同一类的要求和精度会更高，因而使用MSELoss能够使网络参数更快，更好地更新，训练的效果也更好。

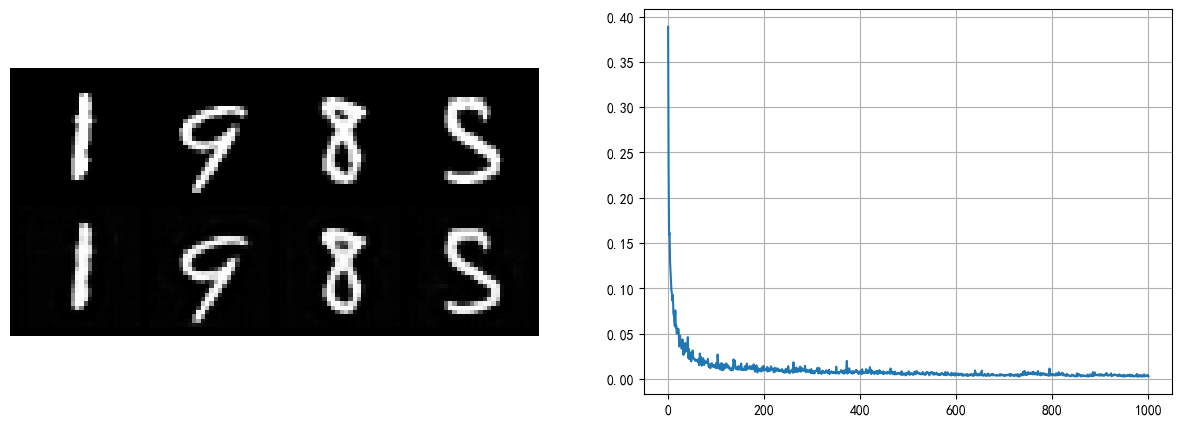


图 8 使用MSELoss卷积层AutoEncoder生成结果及loss曲线

图 9为卷积层AutoEncoder在CIFAR10生成结果及loss曲线，用Adam优化器，学习率为2e-4，一共训练了10000次，最终loss为0.0015。损失函数使用MSELoss。相比于MNIST数据集，CIFAR10的图片是3通道的，因而训练难度更大，生成的图片效果也比MNSIT数据集要差。

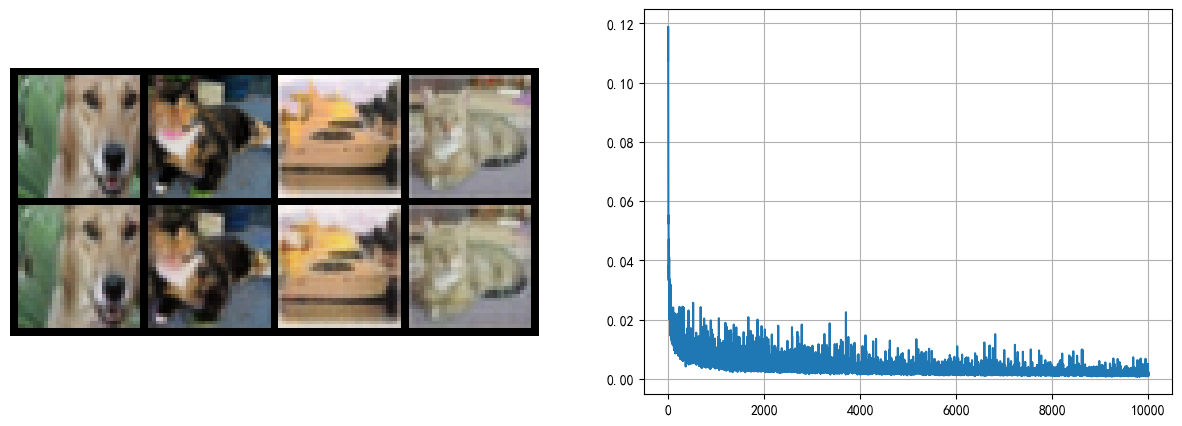


图 9 卷积层AutoEncoder在CIFAR10生成结果及loss曲线

将测试集的所有样本都输入到AutoEncoder中，对其生成的结果作t-SNE,并进行可视化，如图 10所示。a)为线性层AutoEncoder的结果，b)为卷积层AutoEncoder的结果，对比两个子图可以发现使用卷积层AutoEncoder相同类别样本间的距离更近，不同类别样本间的距离更远。这说明卷积层AutoEncoder对数据集特征提取的效果更好，与上文的结论吻合，及卷积层AutoEncoder的效果要好于线性层。

|  |  |
| --- | --- |
| a)线性层AutoEncoder | b)卷积层AutoEncoder |

图 10 AutoEncoder的t-SNE结果

图 11为AutoEncoder生成的10个类别手写数字。a)图为线性层AutoEncoder的结果，b)图为卷积层AutoEncoder的结果，可以直观的看出，卷积层生成的图片清晰图要高于线性层。这一结果与上文结果十分吻合。

|  |  |
| --- | --- |
| a)线性层AutoEncodere | b)卷积层AutoEncoder |

图 11 AutoEncoder生成的手写数字

图 12为VAE生成结果及loss曲线，使用了Adam优化器，学习率为1e-3，一共训练了20万次。最终的loss为0.1173。可以看到VAE生成的图片质量不如AutoEncoder,最终的loss也比AutoEncoder要大。这是因为VAE的Encoder将图片编码为k个均值和k个方差。VAE假设数据服从正太分布，因而将k个正态分布输入到VAE的Decoder中，正态分布的均值和方差对应Encoder的编码结果。相比与直接作反运算的AutoEncoder，VAE的训练难度更大，因而其结果也比AutoEncoder要差。

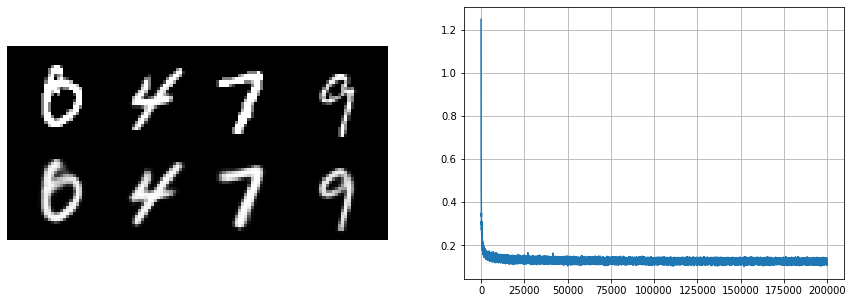


图 12 VAE生成结果及loss曲线

既然Decoder的输入通过采样得来，同时我们已经促使我们的输入所处的分布为‘K’个标准正态分布的混合。因此，我们直接从标准正态分布中随机采样 ‘K’个值，就能当作编码 输入解码器了。图 13为VAE生成结果，此处取K=10。可以看到生成的10个样本基本可以辨识出具体数字，训练出来的VAE效果较优。

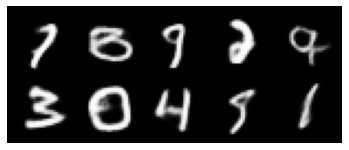


图 13 VAE生成结果

图 14为GAN生成结果及loss曲线。正中间的图为判别器（discriminator）的loss变化曲线，右图为生成器（generator）的loss变化曲线。两个loss都有较大幅度的振荡，这正是生成器和判别器“左右互博”的结果。最理想的结果为生成器的结果让判别器无法辨别，即生成器生成的图片，判别器有0.5的概率预测为真实的图片，有0.5的概率预测为假的图片。此时生成器的loss大概为左右，而判别器的loss大概为生成器的2倍，这是因为由TP和TF两部分构成。生成的图片和抽样的图片完全不一样，这是因为生成器只提取了样本的尺寸（size），我们并没有将抽样图片像素信息传入给生成器，因而生成器生成的图片与抽取的样本并无关系。生成器生成了图片“1978”，较清晰，较真实，效果较好，较难辨别真伪。不过，随着训练次数的增加，生成器生成的种类明显减少，且更频繁生成构成较为简单的数字（如1,7等），这大概是训练的时候loss陷入了局部极小点，生成器发现生成1,7这样简单的数字能更好地逃过判别器的辨识。

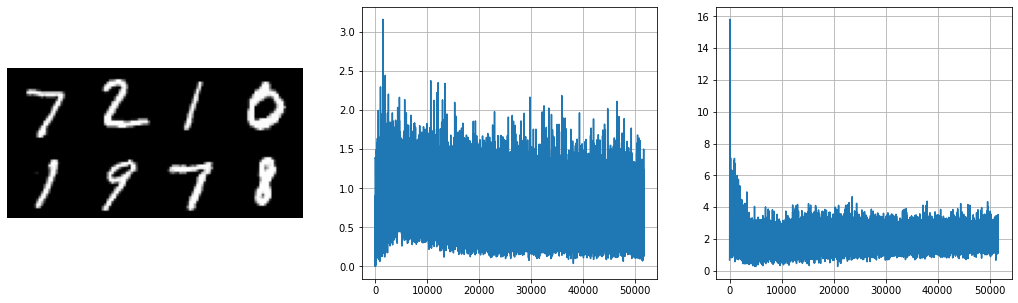


图 14 GAN生成结果及loss曲线

# 九、总结及心得体会：

# 十、对本实验过程及方法、手段的改进建议及展望：

**报告评分：**

**指导教师签字：**