实验5-生成式对抗网络实现

符兴 7203610316

**1 实验环境**

Ubuntu 20.04，Python 3.9， PyTorch1.13，Cuda11.7，TensorBoard；

|  |
| --- |
| 1 |
| 图1 Python环境 |

**2 实验过程**

**2.1 模型构建**

|  |
| --- |
|  |
| 图2 判别器模型 |

在本次实验中，需要基于Pytorch实现GAN、WGAN、WGAN-GP。这几个模型有两个基本的模型，一个是判别器模型，一个是生成器模型；判别器模型输入的是一个点坐标，输出的是该点坐标是否真实的概率；在WGAN和WGAN-GP中，去掉了sigmoid()，因为判别器要拟合的是Wasserstein距离，它不是一个0或1的分类问题，而是回归问题，取值不限于0到1。

|  |
| --- |
|  |
| 图3 生成器模型 |

生成器模型是一个生成图片的网络，它接收一个随机的噪声z，通过这个噪声生成图片。在训练过程中，生成器G的目标就是尽量生成真实的图片去欺骗判别器D。而D的目标就是尽可能把G生成的图片和真实的图片分别开来。这样，G和D构成了一个动态的“博弈过程”。

在最理想的状态下，G可以生成“以假乱真”的图片G(z)，对于D而言，它难以判定G生成的图片究竟是不是真实的，因此D(G(z)) = 0.5。

**2.2 模型训练**

GAN、WGAN、WGAN-GP三个模型在结构上没有太大的区别，三者的本质区别在于Loss的计算上。

|  |
| --- |
|  |
| 图4 GAN的Loss |

在GAN中，先训练判别器D，判别器D是希望V(G, D)越大越好，所以是加上梯度。第二步训练生成器G时，V(G, D)越小越好，所以是减去梯度。整个训练过程交替进行。

|  |
| --- |
|  |
| 图5 WGAN的Loss |

WGAN和GAN的区别在于，WGAN的Loss计算都不再去log，用Wasserstein距离代替JS散度；Wasserstein距离相比KL散度、JS散度的优越性在于，即便两个分布没有重叠，Wasserstein距离仍然能够反映它们的远近。除此之外，还限制了判别器的参数大小，将其限制在预设的区间内。

|  |
| --- |
|  |
| 图6 WGAN-GP的Gradient Penalty |

WGAN在处理Lipschitz限制条件时直接采用了权重截断的方式，这样会限制模型参数固定在一个范围之内，超出这个范围要么取最大值要么取最小值，但是随着层数加深可能会出现梯度消失和梯度爆炸的现。WGAN-GP在WGAN的基础上，通过Gradient Penalty的方式，设置一个额外的loss项来实现梯度与截断阈值之间的联系。

**2.3 隐空间语义方向搜索**

GAN 中的生成器通常以随机采样的隐向量z作为输入，生成高保真图像。通过改变隐向量z，我们可以改变输出图像。然而，为了改变输出图像中某些特定的属性（如面部表情、姿势、肤色等），我们需要知道隐向量z的特定移动方向。在本次实验中，SeFa是一种无监督的方式，无需数据采样和模型训练就能找出这些方向向量来改变输出图像中的属性。

SeFa对图像的操作，可以看做是将d维潜在空间中的对应向量z沿着n的方向进行移动。

而GAN还会将z映射到另一个m维空间的y。

最终，将这一问题转化为：

根据拉格朗日法，求出的特征向量，找出k个使图像进行变换的信息，即选择特征值最大的k个特征向量作为方向。然后，给定一定的步长，将其信息加入生成器的输入向量z中进行训练，得到特定的图像语义变换。

**3 实验结果**

**3.1 不同网络结构、不同优化器实验结果对比**

|  |
| --- |
|  |
| 图7 GAN-Adam的训练结果 |

|  |
| --- |
|  |
| 图8 GAN-SGD的训练结果 |

|  |
| --- |
|  |
| 图9 GAN-RMSprop的训练结果 |

|  |
| --- |
|  |
| 图10 WGAN-RMSprop的训练结果 |

|  |
| --- |
|  |
| 图11 WGAN-GP-RMSprop的训练结果 |

对比上面的实验结果可以发现，WGAN、WGAN-GP模型的收敛速度较快于GAN网络；除此之外，在训练GAN时不同的优化器也具有明显的差异；SGD优化器收敛速度最慢，需要接近2000个Epoch才收敛；Adam收敛速度次之，RMSprop收敛速度最快，效果也是最好的。

**3.2 隐空间语义方向搜索实验结果**

在运行程序后，保存了五张效果图，可以发现找到的语义方向分别为：性别、脸部朝向、表情等。

**4 心得与体会**

在进行本次实验后，我深刻体会到了其在计算机视觉领域的巨大潜力。在实验中，我深入了解了GAN、WGAN、WGAN-GP等模型的基本原理和工作机制。

在这个过程中也遇到了很多问题，比如学习率过大，由于Loss的不稳定导致结果非常丑陋等等。