

Image Generation with GAN

张益铭 车13 2021010552

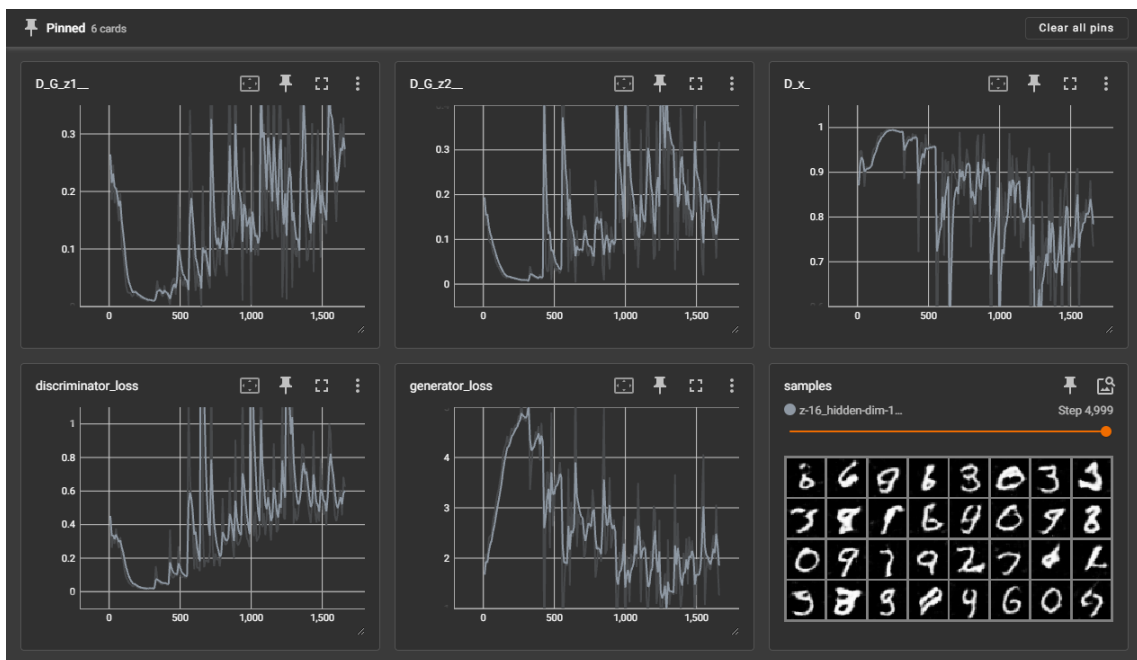
Training

实验参数如下：

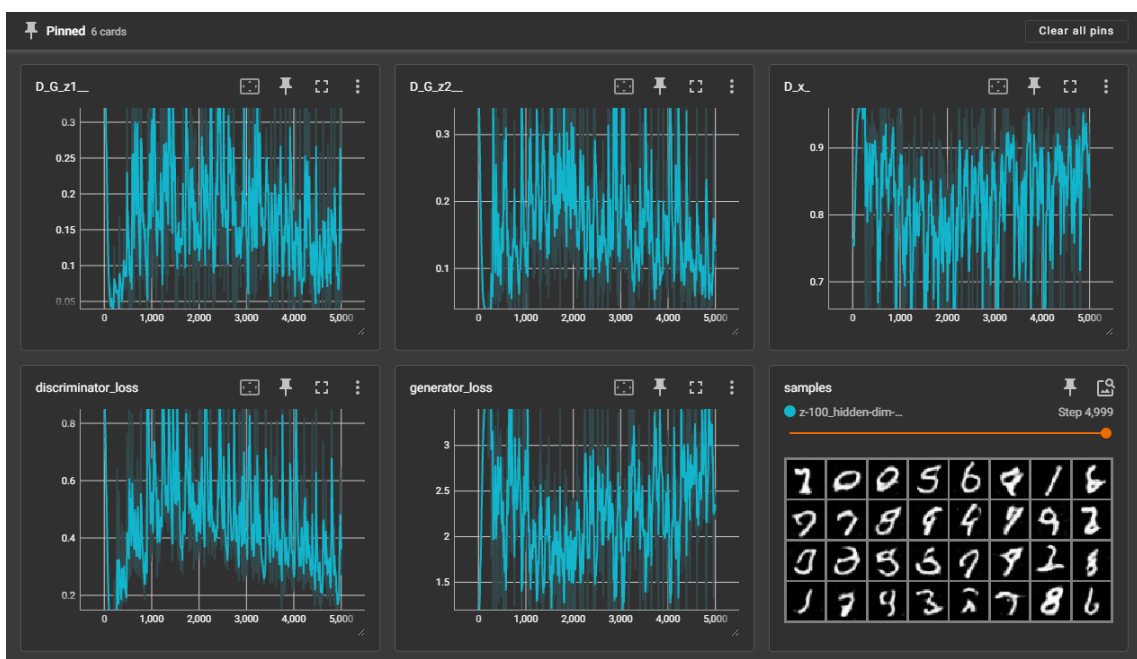
- batch_size: 64
- num_training_steps: 5000

采用不同的 `latent_dim` 和 `hidden_dim`，结果如下：

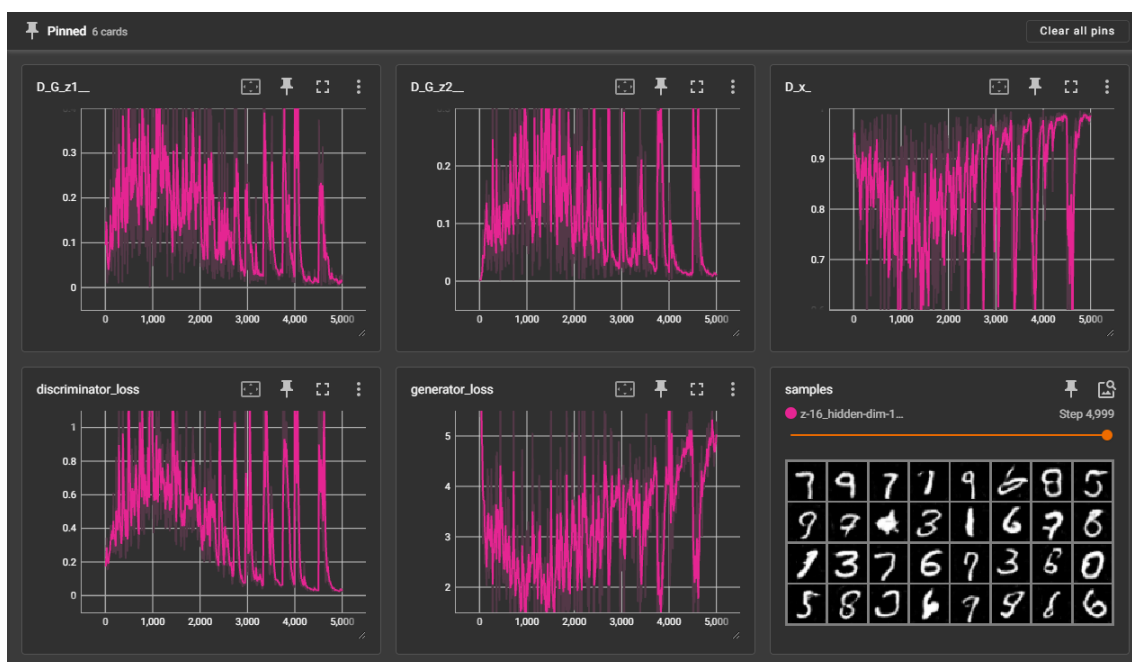
- latent_dim=16, hidden_dim=16



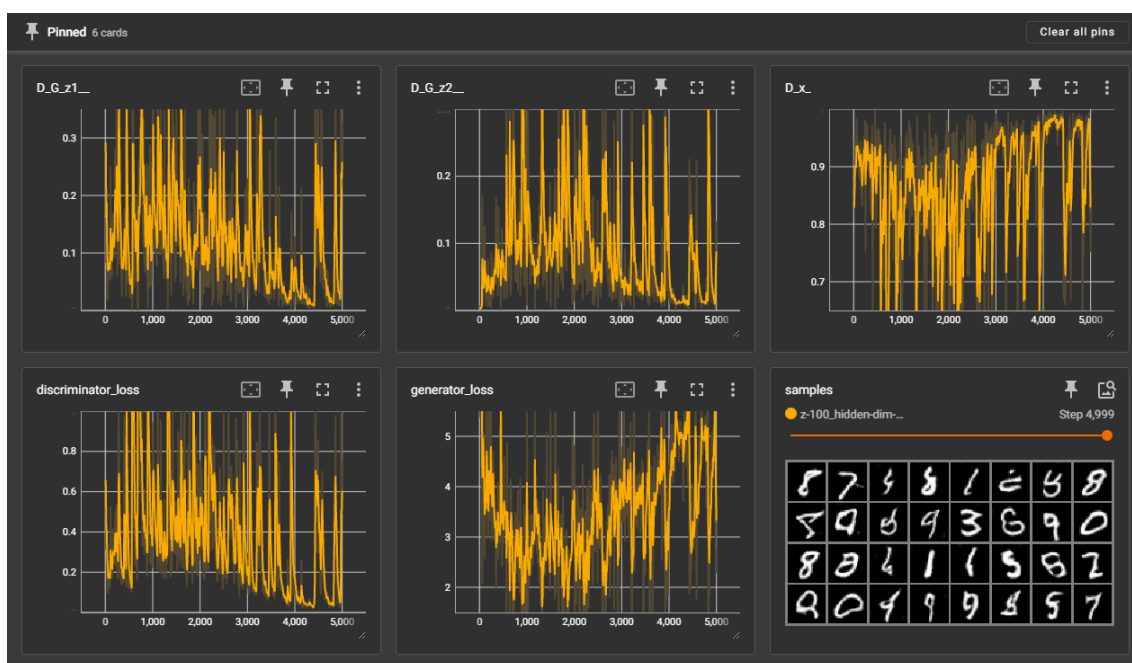
- latent_dim=100, hidden_dim=16



- latent_dim=16, hidden_dim=100



- latent_dim=100, hidden_dim=100



FID score

	latent_dim=16	latent_dim=32	latent_dim=64	latent_dim=100
hidden_dim=16	77.439	83.885	89.062	63.222
hidden_dim=32	76.483	81.280	47.073	64.410
hidden_dim=64	43.285	42.178	30.566	41.588
hidden_dim=100	47.817	50.046	53.034	32.974

latent_dim 和 hidden_dim 的影响

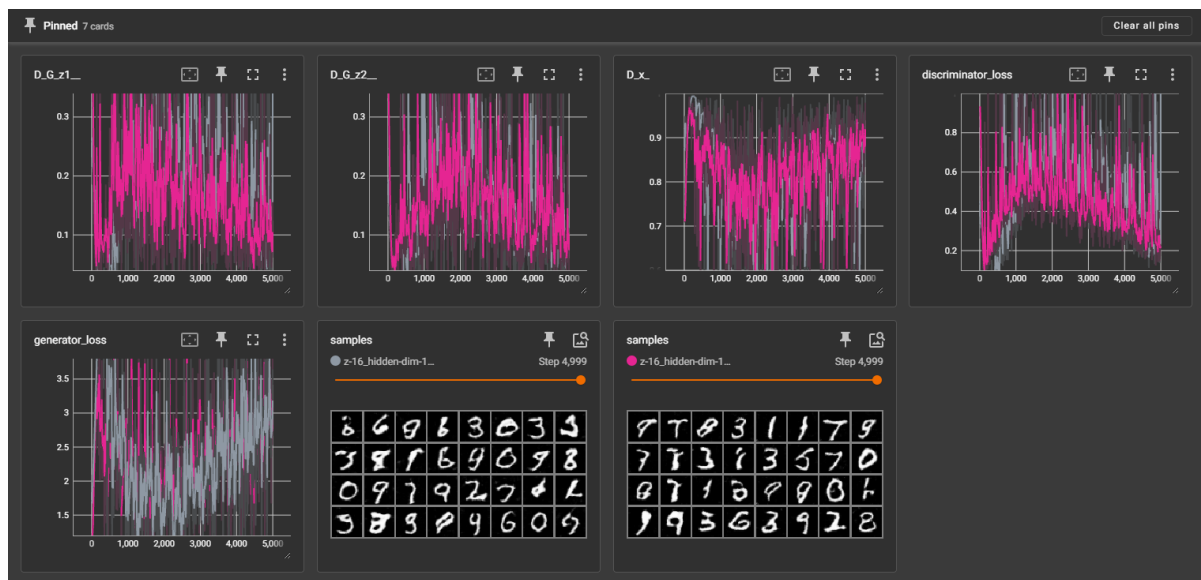
- latent_dim
 - latent_dim 提供了一个紧凑的编码，用于表示生成数据的特征，其的大小决定了生成器可以捕获多少数据的潜在复杂性

- latent_dim 太小，生成器的表达能力不足，可能无法生成具有足够多样性和复杂性的样本
- latent_dim 过大，潜在空间过于稀疏，训练可能变得困难，因为生成器需要从高维空间中找到合适的映射
- hidden_dim
 - hidden_dim 是指生成器和判别器中隐藏层的神经元数量，决定了网络的容量，即模型拟合复杂数据分布的能力
 - 较大的 hidden_dim 提高了网络的表达能力，使其能够捕获更多的特征和细节；同时会增加模型的参数量，从而增加训练时间和显存占用
 - 过大的 hidden_dim 会导致过拟合，尤其是在训练数据有限时
- 较高的 latent_dim 在空间需要隐藏层具有更强的映射能力（即更大的 hidden_dim），以有效利用高维信息

Activation function & batch normalization

均在默认参数（latent_dim=16, hidden_dim=16, batch_size=64, num_training_steps=5000）下进行对比

将 LeakyReLU 替换为 ReLU

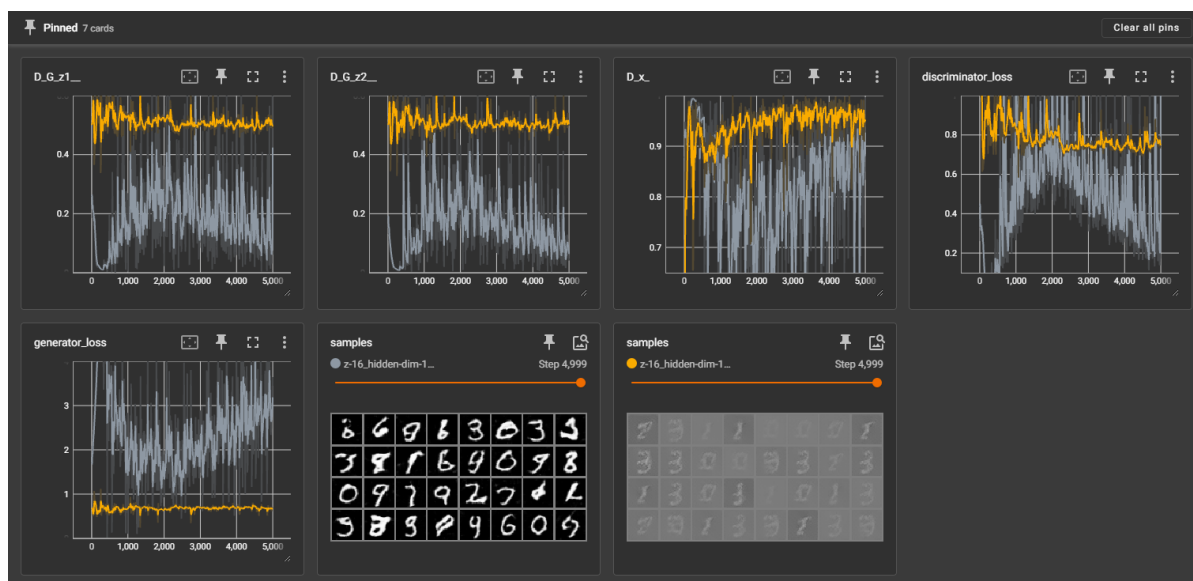


上图中，粉红色为将 LeakyReLU 替换为 ReLU 的结果，FID Score 为 70.004；灰色为默认结果

使用 ReLU 时，生成器损失和判别器损失的波动更剧烈，且整体 loss 较高，而使用 LeakyReLU 时，损失相对稳定，loss 值较低。

因此将 LeakyReLU 替换为 ReLU 会使训练的稳定性降低，损失波动更大，生成样本的质量下降。建议在训练中继续使用 LeakyReLU 或其他改进的激活函数，以避免梯度消失问题并提高模型性能。

删除 Batchnorm

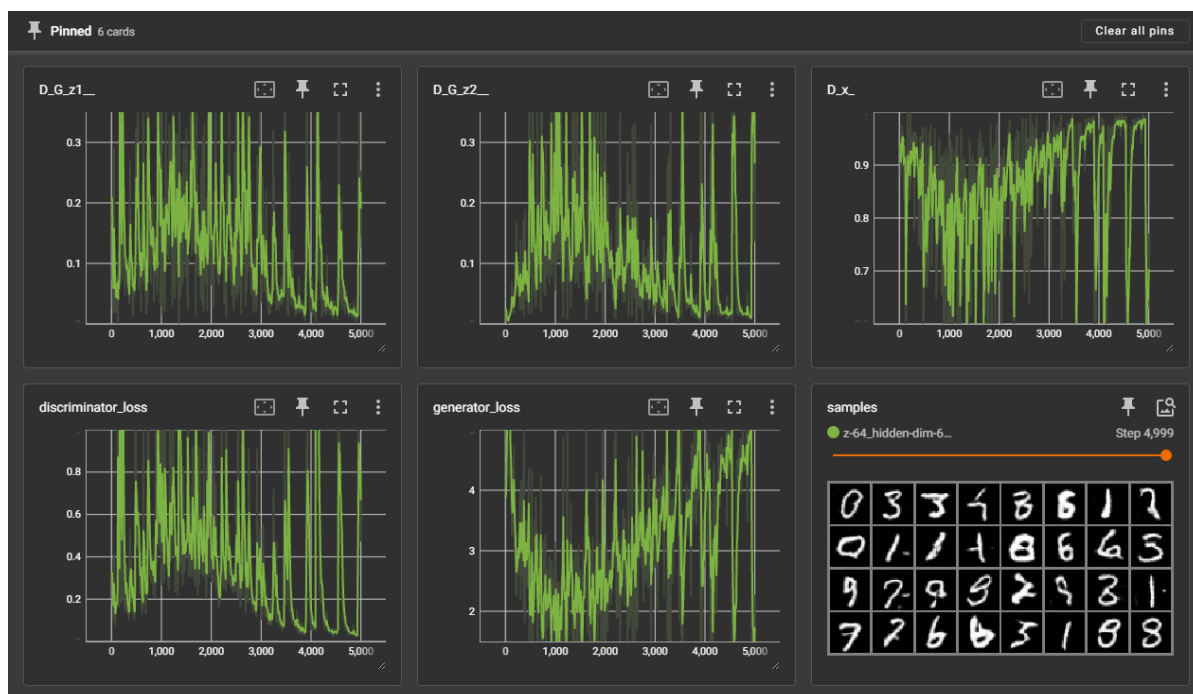


上图中，黄色为删除 Batchnorm 后的结果，FID Score 为 236.592；灰色为默认结果。

删除 BatchNorm 后，判别器的损失显著更高，生成的样本质量较差，许多数字模糊甚至看不清形状，会导致 GAN 的整体性能下降，特别是生成器生成的样本质量显著恶化。

Nash equilibrium

在本次实验中，当 latent_dim=64, hidden_dim=64时，模型具有最小的 FID Score，为30.566，训练结果如下：



从图中可见，本次实验的 GAN 模型并没有达到纳什均衡

- D_G_z1
 - D_G_z1 是判别器对由生成器生成的假样本，记为 $G(z)$ 的输出，通常表示生成器刚开始生成样本时，判别器的判断信心
 - 在上图中，D_G_z1始终处于一个非常低的水平，反映了判别器认为生成样本为假的程度非常高

- D_{G_z2}
 - D_{G_z2} 是判别器对生成器更新后的样本的判断信心
 - 本次实验中, D_{G_z2} 同样处在较低的水平, 表示判别器很确定生成的样本为假
- D_x
 - D_x 是判别器对真实样本 (来自数据集) 判断的输出信心, 反映了判别器是否能正确识别真实样本
 - 本实验中, D_x 接近 1, 表示判别器能够非常确信地将真实样本识别为真

因此可以总结出判别器过强, 生成器难以学习。如果 D_x 、 D_{G_z1} 和 D_{G_z2} 都趋于 0.5, 才说明生成器和判别器达到了纳什均衡。