

Artificial Neural Network HW4

刘泓尊 2018011446 计 84

Department of Computer Science, Tsinghua University

2020 年 11 月 3 日

目录

1 Network Structure and Hyperparameters	1
2 Experiment Result	1
2.1 FID Overview	1
2.2 Tensorboard Curves	2
2.2.1 VAE	2
2.2.2 GAN	3
2.3 Interpolation(Extrapolation) in Latent Space	3
2.3.1 VAE	4
2.3.2 DCGAN	4
3 Question Answering	5
4 Summary	6

1 Network Structure and Hyperparameters

VAE: 网络结构默认, $lr = 0.001$, $batch_size = 32$, 测试了 $latent_dim = 3, 10, 64, 100$ 的情况。
 $num_train_steps = 25000$

GAN: 网络结构默认, $lr = 0.0002$, $\beta_1(adam) = 0.5$, $batch_size = 64$, $latent_dim = 100$,
 $generator_hidden_dim = 16$, $discriminator_hidden_dim = 16$. $num_train_steps = 25000$

2 Experiment Result

2.1 FID Overview

Model/Best	FID
VAE-3	112.661
VAE-10	46.601
VAE-64	25.929
VAE-100	25.411
GAN-100	57.070

2.2 Tensorboard Curves

2.2.1 VAE

下面几幅图展示了 VAE-3, VAE-10, VAE-100 的 Train/Valid Loss Curve, KL Div, 和 recon Loss. 因为 $latent_dim = 100$ 和 $latent_dim = 64$ 的曲线几乎重合, 这里就不予显示 $latent_dim = 64$ 了。

不同颜色曲线的含义在图中有标注。

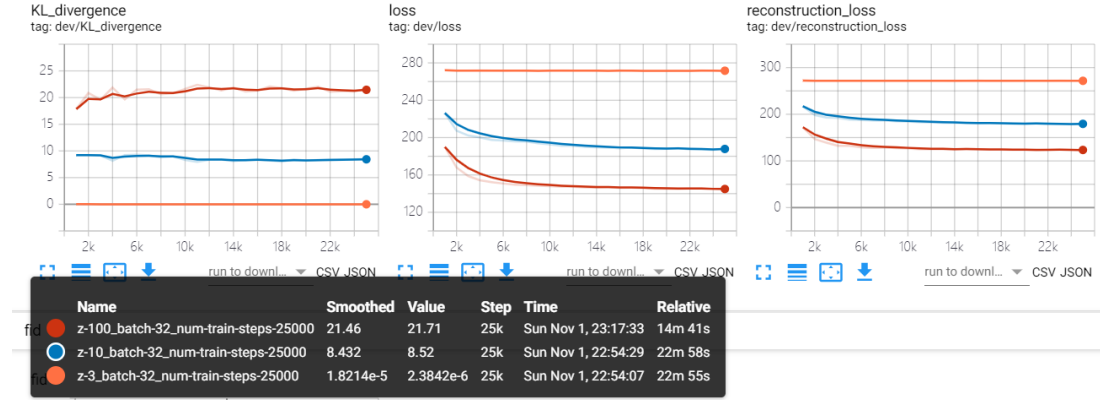


图 1: VAE-3/10/100 的 Dev Loss, Recon Loss & KL div

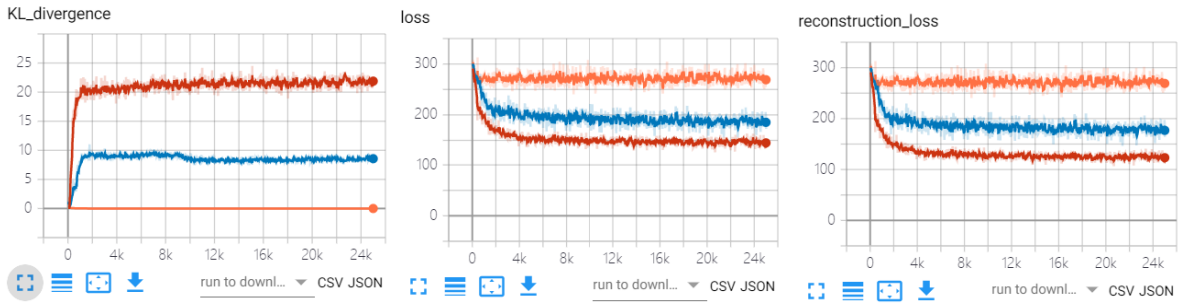


图 2: VAE-3/10/100 的 Train Loss, Recon Loss & KL Div



图 3: VAE-3 @ step25000: Dev Set Image, Reconstruction Image, Generated Image

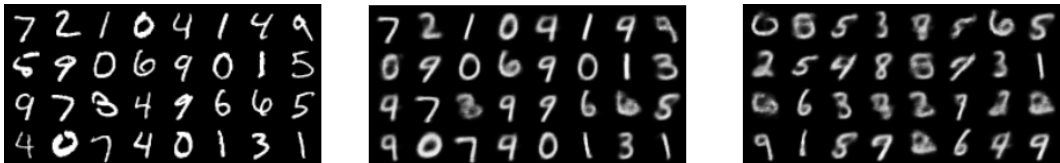


图 4: VAE-10 @ step25000: Dev Set Image, Reconstruction Image, Generated Image

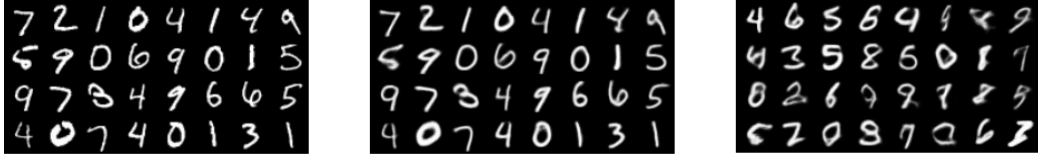


图 5: VAE-100 @ step25000: Dev Set Image, Reconstruction Image, Generated Image

2.2.2 GAN

下面几幅图展示了 GAN-100 的 Loss Curve 和 Discriminator 的输出，以及最后一个 step 生成的图像。



图 6: GAN-100 的 Loss Curve

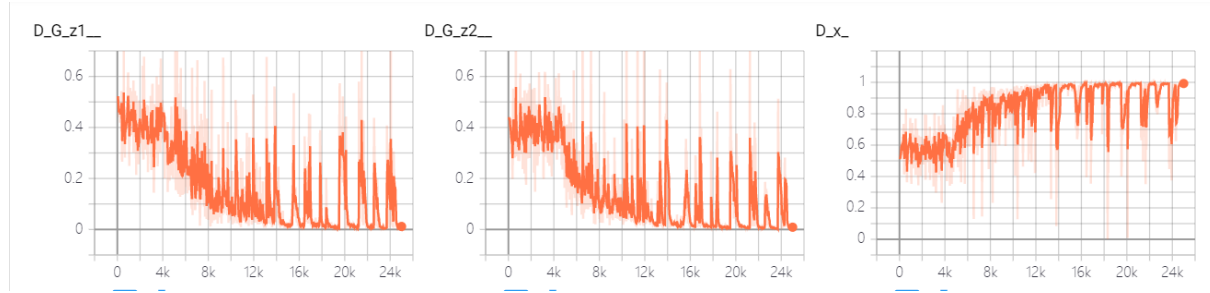


图 7: GAN-100 Discriminator Output

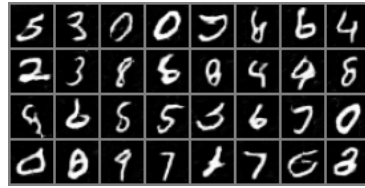


图 8: Images generated by GAN-100 @ step 25000

2.3 Interpolation(Extrapolation) in Latent Space

下面是对 VAE 和 GAN 在 latent space 进行二维插值生成的图像。对于每个二维图像阵，设矩阵为 $K \times K$ ，我随机生成了 4 个 latent variable: z_1, z_2, z_3, z_4 ，对于位置 (x, y) 的图像，其 latent variable $z(x, y)$ 为：

$$\begin{aligned} z(x, 0) &= z_1 + \frac{x}{K}(z_2 - z_1) \\ z(x, K) &= z_3 + \frac{x}{K}(z_4 - z_3) \\ z(x, y) &= z(x, 0) + \frac{y}{K}(z(x, K) - z(x, 0)) \end{aligned} \quad (1)$$

其中 $\frac{x}{K} \in (-0.1, 0.2)$, $\frac{y}{K} \in (-0.1, 0.2)$, $K = 14$

2.3.1 VAE

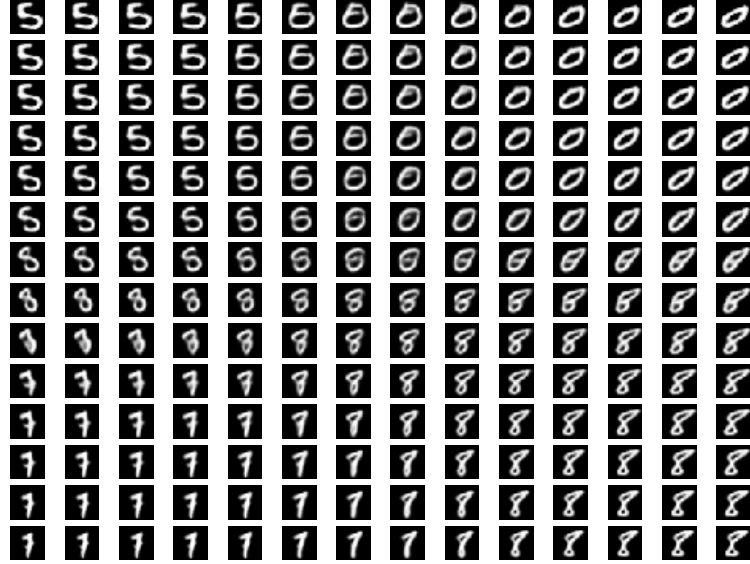


图 9: VAE-100 Interpolation Result

2.3.2 DCGAN

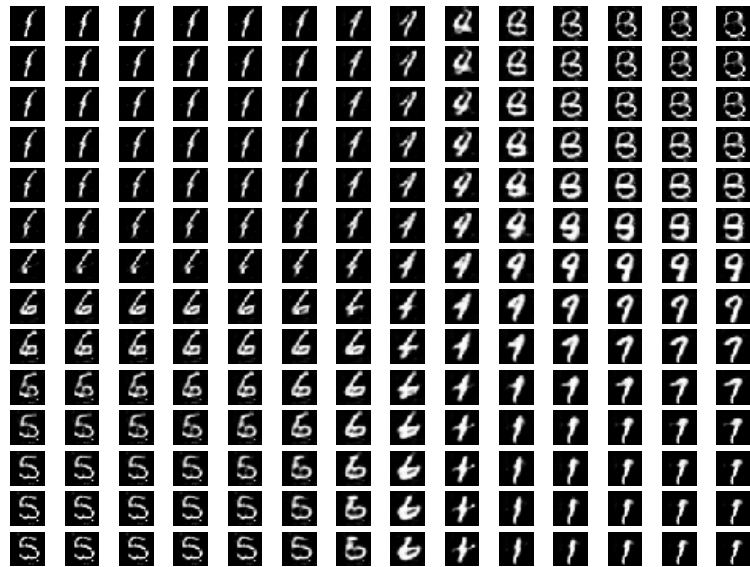


图 10: GAN-100 Interpolation Result

3 Question Answering

Discuss how the dimension of the latent variable influences the performance of VAEs?

从 VAE-3/10/100 生成的图像来看, latent variable 的维度对模型效果有很大影响。

VAE-3 几乎不能获取很有效的多模态特征, 重构的图像和生成的图像一直非常单一且糟糕, 当然, 因为 VAE-3 隐变量维度较低, 其 KL Loss 几乎在一开始便降到了 0, Reconstruction Loss 更是难以收敛。

VAE-10 已经可以建模基本的特征向量, 只是生成的图像还比较模糊。从图4来看, 隐变量维度为 10 已经有了比较好的重构能力。其 FID Score 也降到了可以接受的水平。从训练曲线2来看, 模型已经可以收敛。

VAE-100 的效果最好, 其 FID Score 可以降低到 25 左右。从图5来看, 模型已经有了很好的重构能力, 能够较清晰地重构图像, 并且生成的图像也比 VAE-10 更加清晰。从图2来看, VAE-100 的重构损失可以降到更低, KL Div 会更高。

Has your GAN converged to the Nash equilibrium? See the training curves and explain why.

还没有收敛到纳什均衡。

从图6和图7来看, GAN 的训练很不稳定, 在优化 Generator 阶段, Discriminator 的判别结果会很快上升, 对于生成图像的判别接近了 0.5, 而对真实图像的判别结果可以下降到 0.6 左右。但是在优化 Discriminator 阶段, Discriminator 很快便能分辨生成图像和真实图像, 此时 Discriminator 可能学习到了一个阶跃状的判别函数。因为 Discriminator 在每阶段优化之后依然可以较好地判别真假, 所以生成的分布和实际分布还有差异, 不能说收敛到了纳什均衡点。

Comparing the generated images from DCGAN and VAE-100. Which ones are blurry and which ones are sharp?

VAE-100 生成的图像更模糊。

从训练目标来看, VAE 的 KL Div 要求 Generator 生成的分布向高斯分布看齐, 这可能会和数据集的分布有偏差, 丢失了很多细节信息, 所以生成的图像更模糊。而 GAN 借助 Discriminator 让 Generator 拟合数据集分布, 所以能够捕获细节。如果在 GAN 的训练时采用 L2 Loss, 也相当于用高斯分布拟合, 同样会导致数据的模糊。

但是 VAE-100 能生成更好的形状, DCGAN 生成的图像却经常扭曲, 生成的数字 1 经常带有 7 的特征。可以看到在 MNIST 的数据集上也存在广义上的 Mode Collapse 问题。

Interpolation: Discuss the performance of the generated images. Which one learns better latent representations, DCGAN or VAE-100?

图10和图9展示了 VAE-100 和 GAN 的二维插值结果, 因为模型从连续的隐空间建模, 所以生成的图像也有连续性, 并且维持了较好的质量。VAE-100 的连续性更平滑。

从 VAE-100 的插值结果10和 GAN 的插值结果9来看, VAE-100 学习到了更好的隐变量表示。VAE-100 隐空间插值得到的图像也依然是合理的数字, 但是 GAN 的两个隐变量之间的过渡存在很多扭曲, 有的甚至无法辨认是什么数字。

4 Summary

本次实验我熟悉了 2 种主流的生成模型，并了解了它们的区别和联系。在训练 DCGAN 时，我一开始在 Generator 的 ConvTranspose2d 层使用了 bias, 但是模型生成的图片较差，FID Score 也仅达到了 80 多，所以可能参数量的增加并不一定能带来效果的提升，模型的结构很重要。

感谢老师和助教的悉心指导！