智能感知认知实践实践项目：图像生成

涂宇清 522030910152

1 引言

变分自编码器（VAE）是一种生成模型，能够学习数据的潜在分布并生成新样本。它是无监督学习的有力工具，广泛应用于图像生成、数据压缩和表示学习等领域。在扩散模型（Diffusion Models）兴起后，VAE 仍然是生成式建模领域最为流行的网络结构之一。

VAE 基于变分推断框架，采用经典的编码—解码器架构。其中，编码器将输入数据映射到潜在空间，解码器则从潜在表示重构输入数据。VAE 的核心思想是通过最大化数据似然的证据下界（ELBO）来学习潜在空间的分布。为使 ELBO 的优化可行，还引入了重参数化技巧，将随机采样过程与模型参数优化解耦。

在本项目中，我们聚焦于 VAE 在图像生成中的应用，这是 VAE 的典型应用场景之一。我们将实现一个简单的 VAE 模型，并在 MNIST 数据集上进行训练，以生成手写数字。实验中将探究潜在空间结构如何影响生成效果，以及如何通过最小化重构误差来提升生成图像的质量。

2 变分自编码器（VAE）原理

VAE（Variational Autoencoder）是一种基于概率图模型的生成模型，其引入潜在变量 z，使得观测数据 x 的生成过程可被刻画为联合分布 $p_\theta(x, z)$。

2.1 模型框架

- **先验分布**：通常假设潜在变量的先验为标准正态分布

$$p(z) = \mathcal{N}(z; 0, I).$$

- **似然模型（解码器）**：用参数 θ 的神经网络刻画

$$p_\theta(x \mid z).$$

- **近似后验（编码器）**：用参数 ϕ 的神经网络拟合后验

$$q_\phi(z \mid x) = \mathcal{N}\big(z;\, \mu_\phi(x),\, \mathrm{diag}(\sigma_\phi^2(x))\big).$$

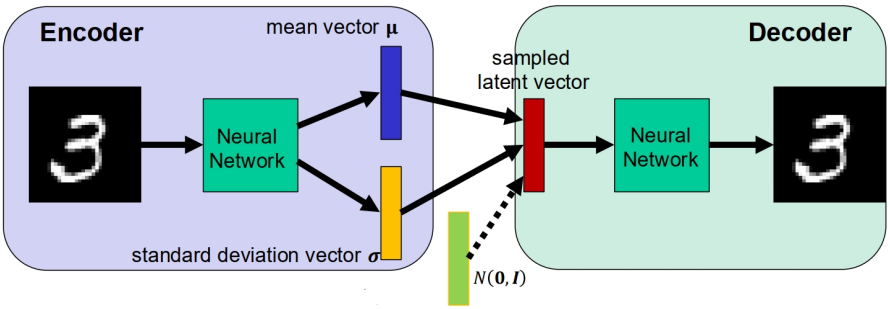


图 2.1　VAE 结构图

2.2 证据下界（ELBO）

由于真实后验 $p_\theta(z \mid x)$ 通常不可解，我们引入变分下界（ELBO）：

$$\log p_\theta(x) \;\geq\; \mathbb{E}_{q_\phi(z|x)}\big[\log p_\theta(x \mid z)\big] - D_{KL}\big(q_\phi(z \mid x) \,\|\, p(z)\big).$$

最大化该下界即可近似最大化数据似然。

2.3 重参数化技巧

为了对 ELBO 进行基于梯度的优化，我们采用重参数化技巧将随机采样写作可导的确定性变换：

$$z = \mu_\phi(x) + \sigma_\phi(x) \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I).$$

这样，梯度可直接反传至 ϕ。

2.4 训练目标

完整的优化目标（对所有样本求期望）可写为

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{x \sim \mathcal{D}}\Big[-\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x \mid z)] + D_{KL}\big(q_\phi(z \mid x) \,\|\, p(z)\big)\Big].$$

其中第一项对应重构误差（Reconstruction Loss），第二项为正则化的 KL 散度。

3 实验过程

在本节中，我们将实现 VAE 模型并将其应用于图像生成，并在 MNIST 数据集上训练模型，使其能够生成手写数字，并研究潜在空间如何影响生成过程。

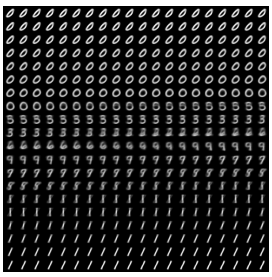
3.1 实验设置

我们使用 MNIST 数据集来训练 VAE 模型。该数据集由 6 万张训练图像和 1 万张测试图像组成，其中测试图像仅用于生成样本和评估模型性能。

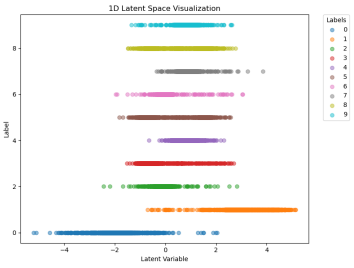
在模型设置中，我们将 28×28 灰度图像平展为 784 维矢量。编码器和解码器都被实现为多层感知器（MLP），具有维度为 400 的隐藏层。使用 Adam 优化器，以 10^{-4} 的学习率训练模型 300 个 epoch。

3.2 消融实验

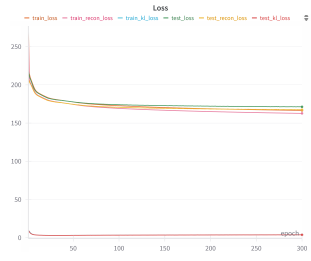
我们设置潜在空间的维度 $z \in \{1, 2, 3, 16, 32\}$，实验结果如图3.1、3.2、3.3、3.4、3.5所示。每个图包含三个子图，分别展示了生成的图像、潜在空间分布和训练损失曲线。



(a) z=1 的输出

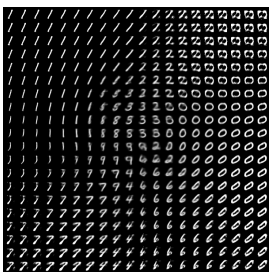


(b) z=1 的潜在空间分布

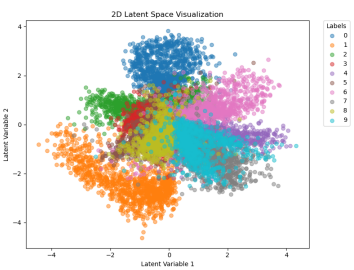


(c) z=1 的损失曲线

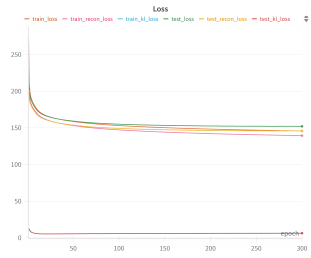
图 3.1　z=1 的实验结果



(a) z=2 的输出

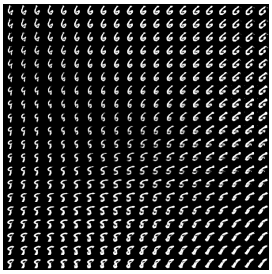


(b) z=2 的潜在空间分布

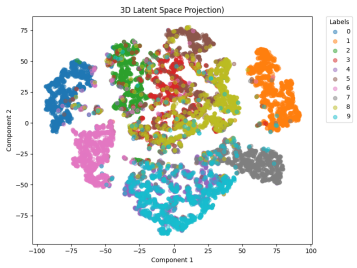


(c) z=2 的损失曲线

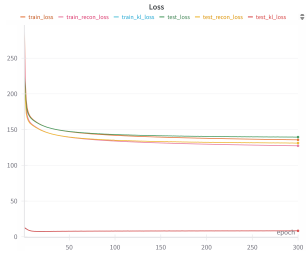
图 3.2　z=2 的实验结果



(a) z=3 的输出

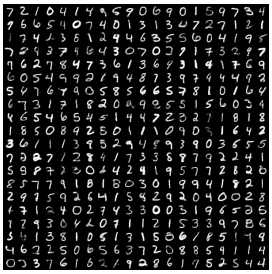


(b) z=3 的潜在空间分布

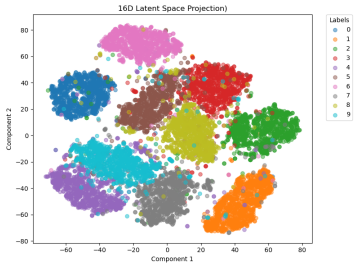


(c) z=3 的损失曲线

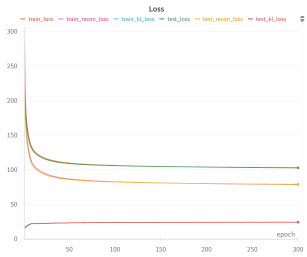
图 3.3　z=3 的实验结果



(a) z=16 的输出

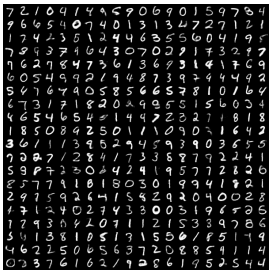


(b) z=16 的潜在空间分布

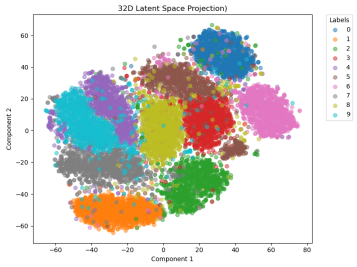


(c) z=16 的损失曲线

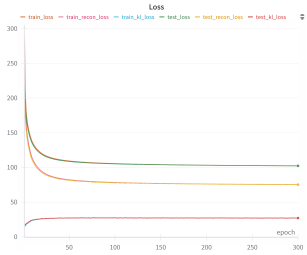
图 3.4　z=16 的实验结果



(a) z=32 的输出



(b) z=32 的潜在空间分布



(c) z=32 的损失曲线

图 3.5　z=32 的实验结果

从图中可以看出，随着潜在空间维度的增大，模型生成的手写数字质量更高，不同标签在潜在空间中的分布更加分离，重构误差逐渐减小。

当潜在空间维度为 1 时，重构误差较高，生成的图像较为模糊，并且在潜在空间分布中不同标签明显混叠，因此一些数字无法清晰生成。当潜在空间维度设为 2 或 3 时，生成图像的质量和多样性有所提升，但重构误差仍不够理想，潜在空间中不同标签之间依然存在重叠。将潜在空间维度增大至 16 与 32 时，模型能够生成高质量且多样性强的图像，重构误差显著降低。通过使用 t-SNE 对高维潜在空间向低维进行降维处理，可以观察到潜在向量分布彼此分离良好。

4 总结

本项目聚焦于变分自编码器（VAE）在图像生成任务中的应用，通过理论分析与实验验证，深入探究了 VAE 的工作原理及其性能影响因素。实验表明，潜在空间维度是决定生成图像质量与多样性的关键因素。低维潜在空间限制了模型对数据复杂分布的捕捉能力，导致生成图像模糊且标签混叠；而随着潜在空间维度的提升，模型能够更精准地重构输入并生成高质量图像，潜在空间的分布也愈发清晰分离。这验证了 VAE 在合理配置下可有效利用概率生成模型特性，平衡数据拟合与正则化。未来可进一步优化 VAE 架构或结合其他生成模型优势，以突破其在图像生成任务中的局限，拓展其在复杂数据集上的应用潜力。

5 代码及复现方式

详情见GitHub。