

Deep generative models（深度生成模型）主要是用来建模数据的分布，并能生成与训练数据相似的新样本。我们可以从四类主流模型来讲它们的原理：**VAE**、**GAN**、**Diffusion Models** 和 **Autoregressive Models (AR)**。


1. VAE (Variational Autoencoder)

核心思想：概率编码 + 重建

- 结构：编码器（Encoder）+ 解码器（Decoder）
- 工作方式：
 - 编码器把输入 x 映射到一个隐变量 z 的分布（不是点，而是分布，比如高斯）。
 - 解码器根据 z 的样本来生成 \hat{x} 。
 - 训练目标：最大化生成概率 $p(x)$ ，但直接算太难 → 用变分推断近似 $\log p(x)$
- 损失函数：

$$ELBO = \mathbb{E}_{q(z|x)}[\log p(x|z)] - \text{KL}(q(z|x)||p(z)) \quad (1)$$

- 第一项：重建损失
- 第二项：正则化，使隐变量分布不偏离标准正态分布

 关键词：概率建模、连续潜变量、重参数技巧

2. GAN (Generative Adversarial Network)

核心思想：对抗博弈，生成器骗过判别器

- 结构：生成器（G）+ 判别器（D）
- 工作方式：
 - G 从随机噪声 $z \sim \mathcal{N}(0, I)$ 生成数据 $G(z)$
 - D 判断输入是真实样本还是 G 生成的
 - 两者玩“零和博弈”：G 尽量生成以假乱真的数据，D 尽量辨认真假

Generator network: try to fool the discriminator by generating real-looking images

Discriminator network: try to distinguish between real and fake images

- 损失函数（经典 GAN）

$$\min_{\theta_g} \max_{\theta_d} [\mathbb{E}_{x \sim p_{\text{data}}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z)))] \quad (2)$$

注意这里：

$$\min_{\theta_g} \max_{\theta_d} [\mathbb{E}_{x \sim p_{\text{data}}} \log \underbrace{D_{\theta_d}(x)}_{\substack{\text{Discriminator output} \\ \text{for real data } x}} + \mathbb{E}_{z \sim p(z)} \log(1 - \underbrace{D_{\theta_d}(G_{\theta_g}(z))}_{\substack{\text{Discriminator output for} \\ \text{generated } G(z)}})] \quad (3)$$

Discriminator (θ_d) wants to **maximize objective such that $D(x)$ is close to 1 (real) and $D(G(z))$ is close to 0 (fake)**

Generator (θ_g) wants to **minimize objective such that $D(G(z))$ is close to 1** (discriminator is fooled into thinking generated $G(z)$ is real)

Training: Alternate between

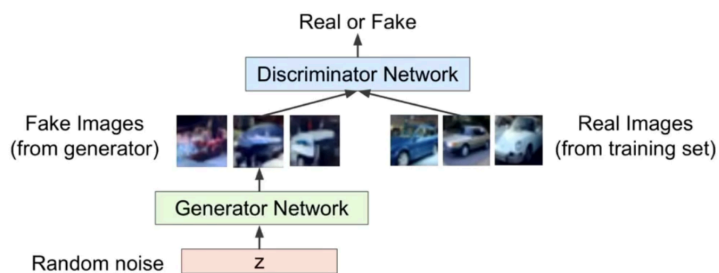
1. Gradient ascent on discriminator

$$\max_{\theta_d} [\mathbb{E}_{x \sim p_{\text{data}}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z)))] \quad (4)$$

2. Gradient descent on generator

$$\min_{\theta_g} \mathbb{E}_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))) \quad (5)$$

📌 关键词：对抗训练、不显式建模数据分布



3. Diffusion Models (扩散模型)

核心思想：正向添加噪声、反向学习去噪

- 过程：
 1. 正向过程（前向扩散）：逐步给图像添加高斯噪声，最后变成纯噪声（可看作马尔科夫链）
 2. 反向过程（生成）：训练一个网络来一步步“去噪”，恢复原图
- 损失函数（常见为 DDPM）：

$$\mathbb{E}_{x, \epsilon, t} [\|\epsilon - \epsilon_{\theta}(x_t, t)\|^2] \quad (6)$$

- x_t 是加噪后的图像
- ϵ_{θ} 是预测噪声的网络

📌 关键词：马尔科夫链、逐步采样、稳定但慢


4. Autoregressive Models (自回归模型)

核心思想：链式建模联合分布

- 基本方法：将联合概率分布分解为条件概率的乘积：

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i | x_{<i}) \quad (7)$$

- 每次生成一个 token / 像素，条件是前面已经生成的
- 应用示例：
 - 文本：GPT、Transformer LM
 - 图像：PixelRNN, PixelCNN

 关键词：精确采样、高质量输出、一次一个、生成速度慢

总结：

模型类型	优点	缺点	核心思想
VAE	有概率解释，训练稳定	样本质量略低	变分推断，隐变量建模
GAN	样本质量高	训练不稳定，模式崩溃	对抗博弈
Diffusion	生成稳定，质量高	生成慢	正向加噪+反向去噪
AR	精确建模，适合序列	生成慢	条件概率链式生成