

# 从基于能量的模型隐空间中生成新桥型的尝试

张洪俊

(万世先行数智交通科技有限公司, 南京 210016)

**摘要:** 使用基于能量的模型进行桥型创新。采用博弈的思想阐述损失函数, 逻辑清晰, 公式简洁明了, 从而避免采用最大似然估计法解释损失函数, 也不需要蒙特卡洛方法求解归一化分母。假设桥型总体服从玻尔兹曼分布, 构建神经网络来表示能量函数。使用 Langevin dynamics 技术生成一个低能量分的新样本, 从而建立基于能量的桥型生成式模型。采用对称结构的三跨梁式桥、拱式桥、斜拉桥、悬索桥图像数据集, 训练能量函数, 使其能够正确地计算出真实样本、虚假样本的能量值。从隐空间中采样, 运用梯度下降算法, 让能量函数将采样点变换为低能量分数的样本, 从而生成不同于训练数据集的新桥型。由于本次尝试存在训练不稳定、训练缓慢的问题, 导致生成的新桥型可能性稀少、生成的图像清晰度偏低。

**关键词:** 生成式人工智能; 桥型创新; 基于能量的模型; 隐空间; 深度学习

**中图分类号:** U448.2; TP181

**文献标志码:** A

## An attempt to generate new bridge types from latent space of energy-based model

Hongjun Zhang

Wanshi Antecedence Digital Intelligence Traffic Technology Co., Ltd, Nanjing, 210016, China

**Abstract:** Use energy-based model for bridge-type innovation. The loss function is explained by the game theory, the logic is clear and the formula is simple and clear. Thus avoid the use of maximum likelihood estimation to explain the loss function and eliminate the need for Monte Carlo methods to solve the normalized denominator. Assuming that the bridge-type population follows a Boltzmann distribution, a neural network is constructed to represent the energy function. Use Langevin dynamics technology to generate a new sample with low energy value, thus a generative model of bridge-type based on energy is established. Train energy function on symmetric structured image dataset of three span beam bridge, arch bridge, cable-stayed bridge, and suspension bridge to accurately calculate the energy values of real and fake samples. Sampling from latent space, using gradient descent algorithm, the energy function transforms the sampling points into low energy score samples, thereby generating new bridge types different from the dataset. Due to unstable and slow training in this attempt, the possibility of generating new bridge types is rare and the image definition of generated images is low.

**Keywords:** generative artificial intelligence; bridge-type innovation; energy-based model; latent space; deep learning

### 0 引言

常见的创新方法有组合法、类比法和列举法等, 它们都曾推动过桥梁工程的技术进步。例如: ①系杆拱桥就是梁桥与拱桥的组合; ②法国园艺家约瑟夫·莫尼尔, 观察到植物根系能够提高泥土的抗裂强度, 通过类比思维发明了钢筋混凝土; ③桥梁采用强度更高的混凝土、钢材来建造, 代替原始的木材、石料, 这可归为属性列举法。

仅使用桥型图片数据集, 训练神经网络模型, 那么只能实现桥型的几何形态组合创新, 无法实现类比法、列举法等创新。而使用多种结构与事物的图片数据集, 或者它们的预训练模型, 甚至多模态模型(Multimodal Models), 则能够带来全新的视角, 从而实现类比法、列举法等创新, 但这对算法、数据、硬件等有更高的要求。

这里使用的基于能量的模型(Energy-Based Model, EBM)<sup>[1-4]</sup>, 与作者之前论文<sup>[5-8]</sup>使用的变分自编码器(Variational Autoencoder, VAE)、生成对抗网络(Generative Adversarial Network, GAN)、像素卷积神经网络(Pixel Convolutional Neural Network, PixelCNN)和标准化流(Normalizing Flow, NF)一样, 都是仅使用桥型图片数据集, 训练神经网络模型, 实现桥型的几何形态组合创新。

基于能量的生成式模型, 也是维持样本空间的维度数目不变(这点与像素卷积神经网络、标准化流一样), 它构建神经网络来表示能量函数, 通过训练使其能够正确地计算出真实样本、虚假样本的能量值(真实样本为低能量分

数，虚假样本为高能量分数)。然后从隐空间中采样，运用梯度下降算法，让能量函数将采样点变换为低能量分数的样本，从而得到逼真的生成图像。

本文建立基于能量的模型，采用之前同样的桥型图片数据集，进一步尝试桥型的几何形态组合创新（本文数据集与源代码开源地址 <https://github.com/QQ583304953/Bridge-EBM>）。

## 1 基于能量的模型简介

### 1.1 概述

基于能量的模型核心思想借鉴自统计物理学的势场中粒子分布概率模型（玻尔兹曼分布律）。当有保守外力（如重力场）作用时，理想气体粒子不再均匀分布，空间不同位置处粒子密度不同，粒子分布规律用玻尔兹曼分布律描述。

基于能量的模型，最早在机器学习中的应用是 1985 年的玻尔兹曼机（Boltzmann Machine），其主要思想是用能量来学习未知的概率分布。对于一个给定的数据集，如果不知道其分布形式，无法直接写出似然函数，是非常难学习的。而各种概率分布之间是可以相互转换的（随机变量函数的分布）。通过适当变换，可将数据集未知的概率分布变换为玻尔兹曼分布。所以基于能量的模型，是学习数据集未知概率分布的一种方法。

### 1.2 玻尔兹曼分布

以理想气体在重力场中为例。气体粒子受到两种作用：粒子的热运动使得它们在空间趋于均匀分布，而重力作用则使它们趋于向地面降落。当两种作用达到平衡时，气体的粒子数密度和压强都将随高度的增加而减小。现实中地面处空气密度远大于高空的。

1. 气体的压强为：

$$p=nkT \quad (1)$$

式中：p 为气体的压强，n 为粒子数密度，k 为物理常数，T 为温度（平衡态下气体的温度处处相同）。

2. 在气体中截取一竖直柱体，根据流体静力学，静止气体中高度 h 处的压强 p 与高度为 h+dh 处的压强 p' 之差为：

$$dp=p' - p=-\rho gdh=-nmgdh \quad (2)$$

式中：ρ 为气体的质量密度，g 为重力加速度，m 为单个粒子质量。

3. 因气体内各处温度相同，故这个压强差只能是由粒子数密度 n 不同引起的，即：

$$dp=kTdn \quad (3)$$

由（1）、（2）式，得：

$$-nmgdh = kTdn \quad (4)$$

$$\frac{1}{n}dn = -\frac{mg}{kT}dh \quad (5)$$

$$\int \frac{1}{n}dn = -\frac{mg}{kT} \int dh \quad (6)$$

$$\ln|n| = -\frac{mg}{kT}h + C \quad (7)$$

$$n = e^C e^{-\frac{mgh}{kT}} \quad (8)$$

$$= n_0 e^{-\frac{\varepsilon}{kT}} \quad (9)$$

式中：n<sub>0</sub> 为 h=0 时的粒子数密度，ε 为粒子的重力势能 mgh。

4. 公式（8）、（9）表明，在势场中的粒子总是优先占据势能较低的位置。在空间中，有能量高的区域和能量低的区域，能量越低，粒子出现的概率就越大。

粒子在空间的分布规律用玻尔兹曼分布律描述：

$$p(x) = \frac{e^{-E(x)}}{Z} \quad (10)$$

式中：x 为空间位置，E(x) 是能量函数（能量由空间位置决定），Z 是归一化分母  $\int e^{-E(x)}dx$ （积分区间为整个空间）。

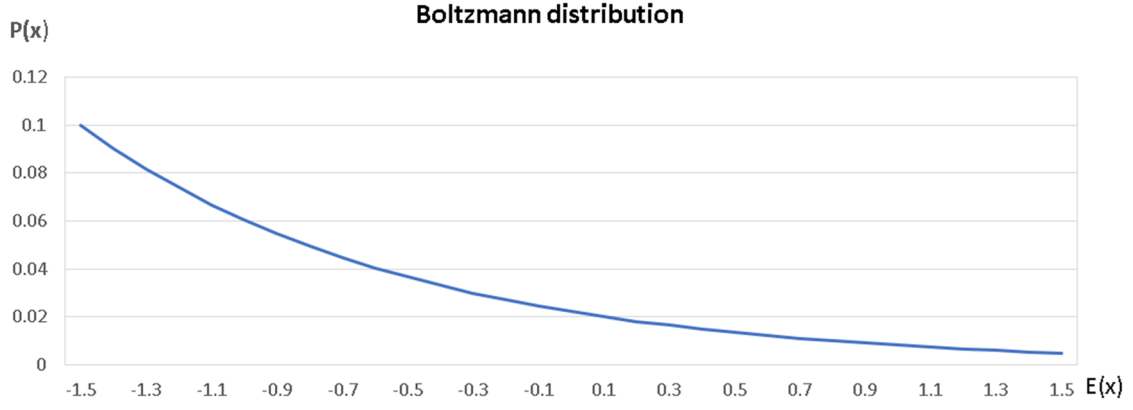


图1 玻尔兹曼分布

Fig.1 Boltzmann distribution

当一个事件的概率用玻尔兹曼分布来表达时，式（10）中  $x$  为事件。

基于能量的模型中的样本、样本特征，对应式（10）中的粒子、粒子空间位置。

### 1.3 样本空间和隐空间

图像的样本空间是高维的像素空间，维度之间存在复杂的关联。隐空间是人为设定的，是样本空间经过神经网络变换后的假设空间。隐空间与样本空间存在映射关系。

基于能量的模型的隐空间维度是相互独立的，维度数目与样本空间相同，维度一一对应。隐空间坐标是生成样本的初始种子。隐空间坐标值可以取  $0 \sim 1$  浮点值，样本空间坐标值一般为  $0 \sim 255$  整数值。

### 1.4 损失函数

1. 参考文献[9]采用最大似然估计法建立损失函数，蒙特卡洛方法近似求解归一化分母  $\int e^{-E(x)} dx$ 。细节见附录 A。

该方法逻辑上不够清晰，公式形式上也不够直观。

2. 本文采用博弈的思想阐述损失函数，逻辑清晰，公式简洁明了。具体如下：

（1）样本空间中大部分样本点没有实际意义，或者说概率趋于 0。而数据集样本均来自于实践，有实际意义，其对应的样本点，有明确的发生概率。假设总体服从玻尔兹曼分布，那么数据集必定位于能量低的区域。

（2）那么将数据集样本（真实样本）输入能量函数，应该输出低能量值。而随机图像、与数据集特征完全不同的图像（虚假样本），应该输出高能量值。故构造损失函数为：

$$\text{Loss}(\theta) = \mathbb{E}[E_{\theta}(x)] - \mathbb{E}[E_{\theta}(y)] \quad (11)$$

式中： $\theta$  为能量函数神经网络权重参数， $\mathbb{E}[E_{\theta}(x)]$  为数据集样本的能量均值， $\mathbb{E}[E_{\theta}(y)]$  为虚假样本的能量均值。

（3）如果虚假样本直接取数据集特征完全不同的图像，那么随着训练的进展，损失值会趋于负无穷，神经网络仅需要学会根据数据集个别特征分类即可，这显然没有什么意义。

因此虚假图像改由能量函数生成，如此虚假图像将含有上一轮神经网络学会的数据集特征。这迫使下一轮训练时，神经网络学会更多的数据集特征。如此迭代，直到神经网络掌握了数据集全部特征，损失值趋于 0。而此时能量函数生成的虚假图像，会非常逼真。

### 1.5 能量函数生成样本

1. 给定一个图像输入，能量函数输出一个能量值标量。如何才能利用该函数来生成一个低能量分的新样本呢？方法是使用一种称为 Langevin dynamics 的技术<sup>[3]</sup>。首先从隐空间随机点出发，然后计算能量函数输出相对于图像输入的梯度，接着在负梯度方向上对图像输入进行小幅度更新，再输给能量函数。如此多次迭代，随机点将逐渐转化为拥有训练集样本特征的图像。整个过程保持能量函数权重参数不变。为了避免陷入局部最小值，在输给能量函数之前，添加少许噪音到图像输入中<sup>[10]</sup>。

$$x_{\text{new}} = x_{\text{old}} - \alpha * \nabla_x E(x) \quad (12)$$

式中： $x_{\text{new}}$  为更新后的图像， $x_{\text{old}}$  为更新前的图像， $\alpha$  为步长（学习率）， $E(x)$  为能量函数（能量由图像特征决定）。

2. 样本的能量示意图，可以用山谷、山脊交错的地形来模拟描述。

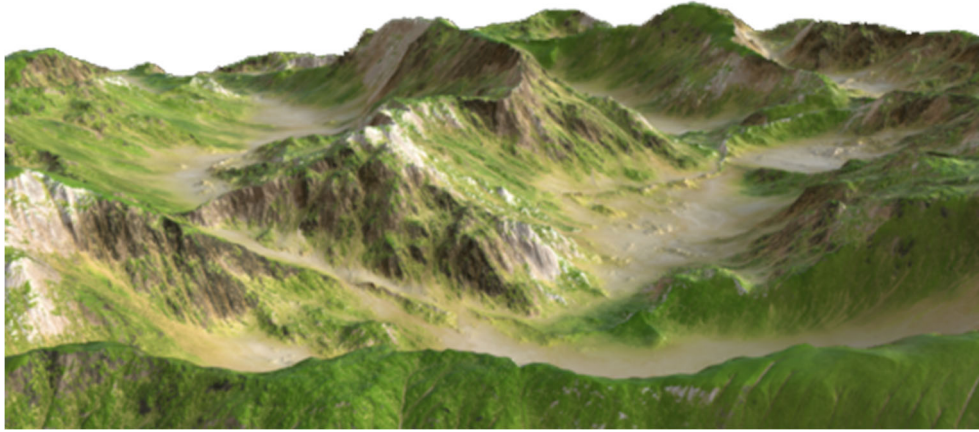


图 2 样本的能量示意图

Fig. 2 Energy diagram of the samples

XY 水平面设想为样本所有维度、Z 轴设想为能量值。特征相似的样本会聚集在一起。数据集样本分布在各个山谷底部（能量低的区域），而山脊处（能量高的区域）为随机噪音或者不相关的图像。隐空间采样生成的样本，越是接近于山谷，能量越低，也越是类似于数据集样本，清晰度也越高；而接近于半山腰位置的样本，往往具有两种标签样本的特征，有可能实现几何形态组合，但是清晰度、现实意义难以保证，有很大的偶然性。

## 2 从基于能量的模型隐空间中生成新桥型的尝试

### 2.1 数据集

采用作者之前论文<sup>[5-8]</sup>的数据集，即每种桥型两种子类（分别为等截面梁式桥、V 形墩刚构梁式桥、上承式拱式桥、下承式拱式桥、竖琴式斜拉桥、扇式斜拉桥、竖吊杆悬索桥、斜吊杆悬索桥），且均为三跨（梁式桥为 80+140+80m，其它桥型均为 67+166+67m），结构对称。

本模型调试难度较大，故将图片尺寸由 512x128 减少为 192x48 像素（缺点是清晰度降低了许多）。

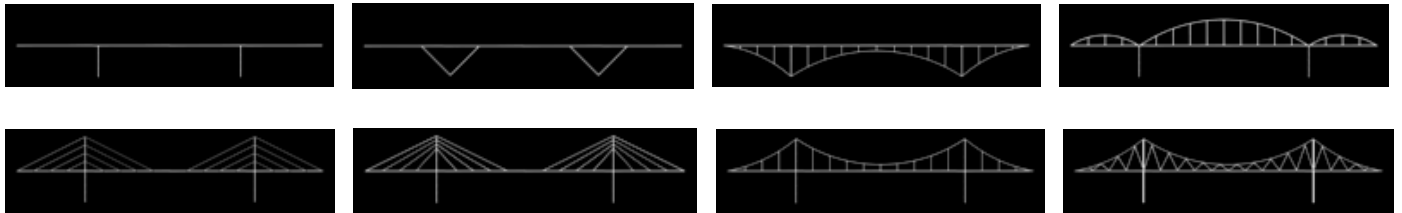


图 3 各桥型立面灰度图

Fig. 3 Grayscale image of each bridge facade

每个子类桥型 1200 张各不相同的图片，整个数据集共 9600 张图片。

### 2.2 模型构建和训练

基于 Python 编程语言、TensorFlow 及 Keras 深度学习平台框架，构建和训练基于能量的模型<sup>[10]</sup>。

#### 1. 能量函数的架构

能量函数的工作计算输入图像的能量值。

先为图像创建一个输入层，然后依次传递给 4 个 Conv2D 层（激活函数 swish），最后展平并依次连接到两个 Dense 层（激活函数 swish 与线性变换），输出结果是标量。模型摘要见下表：

表 1 能量函数模型摘要

Tab.1 model summary of energy function

| Layer (type)                              | Output Shape      | Param # |
|---|-------------------|---------|
| input_1(InputLayer)                       | [(None,48,192,1)] | 0       |
| conv2d(Conv2D)                            | (None,24,96,32)   | 832     |
| batch_normalization(BatchNormalization)   | (None,24,96,32)   | 128     |
| activation(Activation)                    | (None,24,96,32)   | 0       |
| conv2d_1(Conv2D)                          | (None,12,48,64)   | 18496   |
| batch_normalization_1(BatchNormalization) | (None,12,48,64)   | 256     |
| activation_1(Activation)                  | (None,12,48,64)   | 0       |
| conv2d_2(Conv2D)                          | (None,6,24,128)   | 73856   |

|   |                 |        |
|---|-----------------|--------|
| batch_normalization_2(BatchNormalization) | (None,6,24,128) | 512    |
| activation_2(Activation)                  | (None,6,24,128) | 0      |
| conv2d_3(Conv2D)                          | (None,3,12,128) | 147584 |
| batch_normalization_3(BatchNormalization) | (None,3,12,128) | 512    |
| activation_3(Activation)                  | (None,3,12,128) | 0      |
| flatten(Flatten)                          | (None,4608)     | 0      |
| dense(Dense)                              | (None,128)      | 589952 |
| dense_1(Dense)                            | (None,1)        | 129    |
| Total params: 832,257                     |                 |        |
| Trainable params: 831,553                 |                 |        |
| Non-trainable params: 704                 |                 |        |

## 2. 损失函数

能量模型不容易训练，如果超参数没有得到很好的调整，往往会发散。故除了真、假样本间的能量值差异对比散度损失外(Constractive Divergence, CD)，再增加正则化损失，以利于训练。

正则化损失为真、假样本的能量值的平方和的均值，同时乘以一个权重因子（取 0.1 左右）。如此真、假样本的能量值趋向于各为负、正，且绝对值基本相同。

## 3. 训练

从模型中采用生成虚假样本，需要进行大量迭代计算。有一个训练技巧可以显著降低采样成本：使用采样缓冲区。将之前批次的虚假样本存储在缓冲区中，并将它们重新用作下一个批次的采样生成起点，这样更少的步骤就能得到较高质量的样本。为了不完全依赖以前的样本，每个批次有 5% 的样本是从头采样生成的。

训练过程的损失变化曲线见下图（不含前 3 轮结果）：

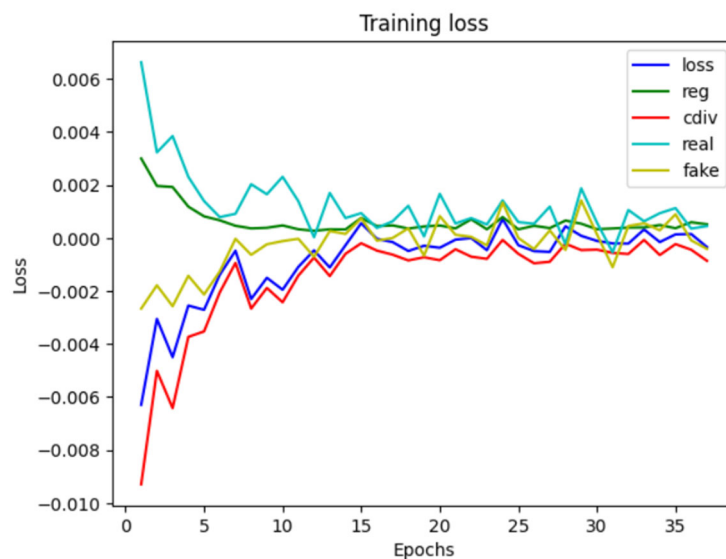


图 4 训练损失曲线

Fig.4 Training loss curve

其中：loss 为总损失均值，reg 为正则化损失均值，cddiv 为对比散度损失均值，real 为数据集样本的能量均值，fake 为虚假样本的能量均值。

本次尝试存在训练不稳定、训练缓慢的问题，导致生成的新桥型可能性稀少、生成的图像清晰度偏低。受硬件条件制约，不可能充分优化，所以只能边测试参数边采样与筛选。

## 4. 由噪音到图像的生成过程

为了深入了解图像在迭代过程中如何变化，可以查看中间样本。

下图展示了一个隐空间随机点，是如何一步步变换为图像的。

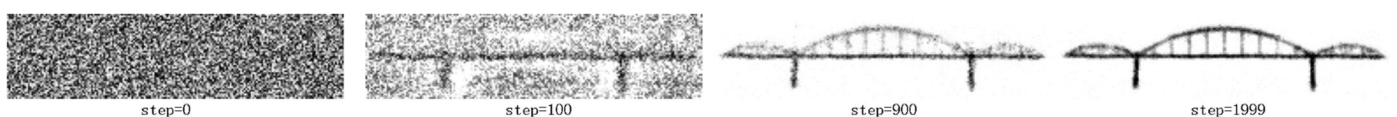


图 5 由噪音到图像的生成过程

Fig.5 The process of generating images from noise

我们看到，虽然起点是从噪声开始，但经过 900 步就获得了基本的形状，到第 1999 步时，形状变得清晰。

### 2.3 隐空间采样探索新桥型

虽然训练时保存了多个模型，但是生成图像质量清晰度过低。而训练时采样缓冲区中的图像清晰度较高，故从中挑选，然后人工基于工程结构思维筛选，得到了与训练集完全不同的 3 种技术可行的新桥型：

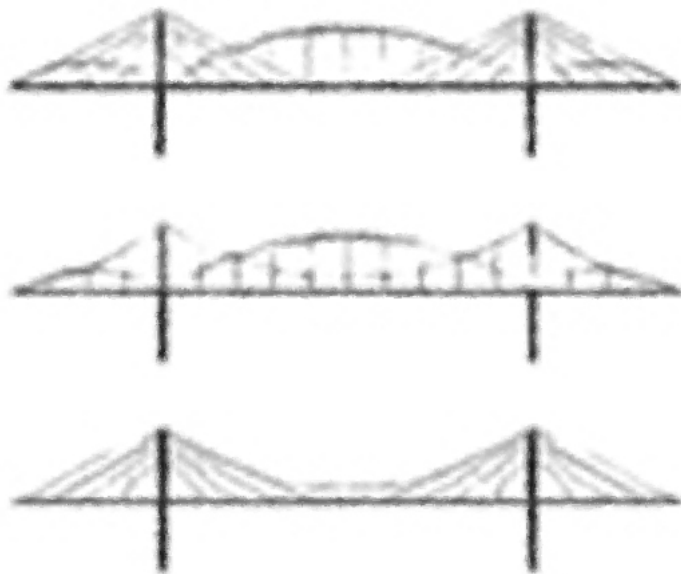


图 6 3 种技术可行的新桥型

Fig.6 Three new bridge types with feasible technology

这里新桥型是指数据集中未曾出现过，而神经网络根据算法创作出的，这代表了模型具有创新能力。

## 3 结语

基于能量的模型能够在人类原创桥型的基础上，将两种桥型合为一体，组合创造生成新桥型。

受硬件条件制约，本次模型参数未充分优化，导致生成的新桥型可能性稀少、生成的图像清晰度偏低。

采用博弈的思想阐述损失函数，逻辑清晰，公式简洁明了，从而避免采用最大似然估计法解释损失函数，也不需要蒙特卡洛方法求解归一化分母，这极大地方便普通人掌握该算法。

### 参考文献

- [1] Yann LeCun, Sumit Chopra, Raia Hadsell, et al. A Tutorial on Energy-Based Learning[C]. Predicting Structured Data, 2006.
- [2] Junbo Zhao, Michael Mathieu, Yann LeCun. ENERGY-BASED GENERATIVE ADVERSARIAL NETWORKS[J]. arXiv preprint, 2016, arXiv: 1609.03126.
- [3] Yilun Du, Igor Mordatch. Implicit Generation and Modeling with Energy-Based Models[J]. arXiv preprint, 2019, arXiv: 1903.08689.
- [4] Bo Pang, Tian Han, Erik Nijkamp, et al. Learning Latent Space Energy-Based Prior Model[C]. NeurIPS, 2020.
- [5] 张洪俊. 从变分自编码器隐空间中生成新桥型的尝试 [EB/OL]. 北京：中国科技论文在线 [2023-11-06]. <http://www.paper.edu.cn/releasepaper/content/202311-5>.
- [6] 张洪俊. 从生成对抗网络隐空间中生成新桥型的尝试 [EB/OL]. 北京：中国科技论文在线 [2023-12-25]. <http://www.paper.edu.cn/releasepaper/content/202312-73>.
- [7] 张洪俊. 从像素卷积神经网络隐空间中生成新桥型的尝试 [EB/OL]. 北京：中国科技论文在线 [2024-01-08]. <http://www.paper.edu.cn/releasepaper/content/202401-1>.
- [8] 张洪俊. 从标准化流隐空间中生成新桥型的尝试 [EB/OL]. 北京：中国科技论文在线 [2024-01-25]. <http://www.paper.edu.cn/releasepaper/content/202401-64>.
- [9] Oliver Woodford. Notes on Contrastive Divergence[EB/OL]. <https://www.robots.ox.ac.uk/~ojw/files/NotesOnCD.pdf>.
- [10] David Foster. Generative Deep Learning[M]. 2nd Edition. America: O' Reilly, 2023.



## 附录 A 第 1.4.1 节的细节

### A.1 损失函数

数据集是从总体中的抽样，故用数据集样本 $x_i$ 训练神经网络（能量函数）时，可以采用最大似然估计法建立损失函数<sup>[9]</sup>：

$$\text{Loss}(\theta) = -\frac{1}{n} \sum_{i=1}^n \ln p_{\theta}(x_i) \quad (13)$$

$$= -\mathbb{E}[\ln p_{\theta}(x_i)] \quad (14)$$

$$= \mathbb{E}[E_{\theta}(x_i)] + \ln z_{\theta} \quad (15)$$

因为我们是生成式模型，目标是让能量函数具有生成逼真图像的能力，因此归一化分母 $z_{\theta}$ 用生成图像的能量值计算，如此才能实现目标。

为了简化计算，先进行化简，上式两边对 $\theta$ 求导：

$$\nabla_{\theta} \text{Loss}(\theta) = \mathbb{E}[\nabla E_{\theta}(x_i)] + \nabla \ln z_{\theta} \quad (16)$$

$$= \mathbb{E}[\nabla E_{\theta}(x_i)] + \frac{\nabla z_{\theta}}{z_{\theta}} \quad (17)$$

$$= \mathbb{E}[\nabla E_{\theta}(x_i)] - \frac{1}{z_{\theta}} \int e^{-E_{\theta}(x)} \nabla E_{\theta}(x) dx \quad (18)$$

$$= \mathbb{E}[\nabla E_{\theta}(x_i)] - \int \frac{e^{-E_{\theta}(x)}}{z_{\theta}} \nabla E_{\theta}(x) dx \quad (19)$$

$$= \mathbb{E}[\nabla E_{\theta}(x_i)] - \int p_{\theta}(x) \nabla E_{\theta}(x) dx \quad (20)$$

$$= \mathbb{E}[\nabla E_{\theta}(x_i)] - \mathbb{E}[\nabla E_{\theta}(x)] \quad (21)$$

$$= \nabla \{ \mathbb{E}[E_{\theta}(x_i)] - \mathbb{E}[E_{\theta}(x)] \} \quad (22)$$

注：上式为了简洁， $\nabla_{\theta}$ 简写为 $\nabla$ 。

上式两边对 $\theta$ 积分，忽略常数项得：

$$\text{Loss}(\theta) = \mathbb{E}[E_{\theta}(x_i)] - \mathbb{E}[E_{\theta}(x)] \quad (23)$$

式中： $\mathbb{E}[E_{\theta}(x_i)]$ 为数据集样本 $x_i$ 的能量均值， $\mathbb{E}[E_{\theta}(x)]$ 为模型采样生成样本的能量均值。可以采用蒙特卡洛方法近似求解 $\mathbb{E}[E_{\theta}(x)]$ 。

### A.2 蒙特卡罗方法

蒙特卡洛方法是统计模拟方法，通过大量采样来近似求解期望值、面积、积分等问题。例如：当一枚硬币正面向上的先验概率未知时，可以通过多次抛掷它，然后统计其正面向上的频率，来近似估计先验概率。

可以在隐空间中随机采样一批数据点，然后运用梯度下降算法，让能量函数将其变换为近似服从总体分布的样本 $x_j$ ，从而近似求解 $\mathbb{E}[E_{\theta}(x)]$ 。即：

$$\mathbb{E}[E_{\theta}(x)] \approx \frac{1}{J} \sum_{j=1}^J E_{\theta}(x_j) \quad (24)$$