

从标准化流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: 通过不同分布之间的坐标、概率转换实例, 深入浅出地介绍了标准化流的基本原理。从“随机变量函数的分布”的角度, 阐述概率转换的实质, 引出了概率转换的缩放因子雅可比行列式。把训练集当成从总体中的抽样, 得出标准化流实质上是通过抽样调查, 来对总体的数字特征进行统计推断, 从而利用最大似然估计法建立损失函数。介绍了标准化流是如何巧妙解决高维矩阵行列式计算、神经网络可逆变换的两大应用挑战。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, 基于 TensorFlow Probability 库中的 Glow API, 构建和训练标准化流。模型能够顺利将桥型数据集的复杂分布变换为标准正态分布, 从得到的隐空间采样, 能够生成不同于训练数据集的新桥型。

关键词: 生成式人工智能; 桥型创新; 标准化流; 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 generative flow

Hongjun Zhang

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

Abstract: Through examples of coordinate and probability transformation between different distributions, the basic principle of normalizing flow is introduced in a simple and concise manner. From the perspective of “the distribution of random variable function”, the essence of probability transformation is explained, and the scaling factor Jacobian determinant of probability transformation is introduced. Treating the dataset as a sample from the population, obtaining normalizing flow is essentially through sampling surveys to statistically infer the numerical features of the population, and then the loss function is established by using the maximum likelihood estimation method. This article introduces how normalizing flow cleverly solves the two major application challenges of high-dimensional matrix determinant calculation and neural network reversible transformation. Using symmetric structured image dataset of three-span beam bridge, arch bridge, cable-stayed bridge and suspension bridge, constructing and training normalizing flow based on the Glow API in the TensorFlow Probability library. The model can smoothly transform the complex distribution of the bridge dataset into a standard normal distribution, and from the obtained latent space sampling, it can generate new bridge types that are different from the training dataset.

Keywords: generative artificial intelligence; bridge-type innovation; generative flow; latent space; deep learning

0 引言

桥型创新主要有两种类别: ①结构力学形式的创新。创造出有别于梁式桥、拱式桥、斜拉桥、悬索桥四种基本结构形式之外的新桥型, 从而实现跨度上的突破, 或者能够减少材料消耗量、增加承载能力等。要实现这个目标, 除了结构理论突破之外, 还受制于材料、计算手段、施工技术等等外在条件; ②外观造型的创新。绝大多数桥梁仅以跨越障碍的功能为主, 平淡无奇, 无造型要求。但是某些城市、景区、特殊场合的少数桥梁, 承载着审美要求。此时桥梁设计师需要将结构与艺术造型有机地结合起来, 做到功能与形式的统一。

古往今来, 桥型创新一直是桥梁设计师面临的重大课题, 是一项既有吸引力又有挑战性的艰巨任务。而生成式人工智能, 将会是桥梁工程师增强自身能力的有力工具。

作者之前论文^[1-3]使用的变分自编码器 (Variational Autoencoder, VAE)、生成对抗网络 (Generative Adversarial Network, GAN) 和像素卷积神经网络 (Pixel Convolutional Neural Network, PixelCNN), 它们各自代表了一种各不相同的方式, 来解决图像样本空间 $p(x)$ 建模的难题。图像样本空间是极为复杂的高维空间, 空间容量是天文数字, 而只有很少的样本具有实际意义, 直接建模难以操作。变分自编码器、生成对抗网络, 是将样本空间压缩变换到一个低维且维度相互独立的隐空间。像素卷积神经网络则是维持样本空间的维度数目不变, 通过捕获相邻维度之间的依赖关系, 以条件概率乘积来近似表示联合概率分布。

标准化流模型 (Normalizing Flow, NF)^[4-7]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一样),

它通过构造一种可逆的变换，将复杂的样本空间联合概率分布 $p(x)$ 变换为简单通用的隐空间联合概率分布 $p(z)$ ，然后从隐空间中采样 z ，经过逆变换生成样本空间 x 。

本文采用 Glow (Generative Flow) 标准化流模型，基于之前同样的数据集，进一步尝试桥型创新（本文数据集与源代码开源地址 <https://github.com/QQ583304953/Bridge-Flow>）。

1 标准化流模型简介

1.1 概述

1. VAE 和 GAN 只是在隐式的学习数据空间的分布，并没有给出一个解析形式的概率分布函数。而自回归模型和标准化流有清晰的概率分布函数。但是自回归模型存在明显的弊端，首先是需要逐点生成，并行计算难度大，而且每个变量的生成只依赖其他变量中的一部分，精度先天性不足，而标准化流模型不存在这样的弊端。

标准化流模型是一个可逆过程，它将任意复杂的数据分布无损变换为简单的基本分布（这里采用标准正态分布），反之亦然。利用已知的高维标准正态分布概率密度函数的解析式，倒算出样本空间的联合概率分布。其构思巧妙，巧夺天工。标准化流模型实质上是通过抽样调查，来对总体的数字特征进行统计推断，其推断吻合度主要决于数据集的质量。

标准化流模型主要有 NICE、RealNVP、Glow、FFJORD 等几种架构^[4-7]。

2. 标准化流隐空间的各个维度是相互独立的，空间维度数目=像素数目。采样时，通过一系列层计算，将隐空间坐标变换到样本空间坐标，从而生成有意义的图像。

虽然维度独立这点上，与朴素贝叶斯算法类似，但是朴素贝叶斯算法缺少分布变换步骤，因此标准化流与朴素贝叶斯算法，有本质上的不同。

1.2 分布变换实例

理论上，只要经过不断地变换，标准正态分布可以拟合任意复杂的分布，而任意复杂的分布也可以变换为标准正态分布。变换实例如下：

1. 首先生成服从二维标准正态分布的 5000 个样本点。此时 2 个维度是相互独立的。

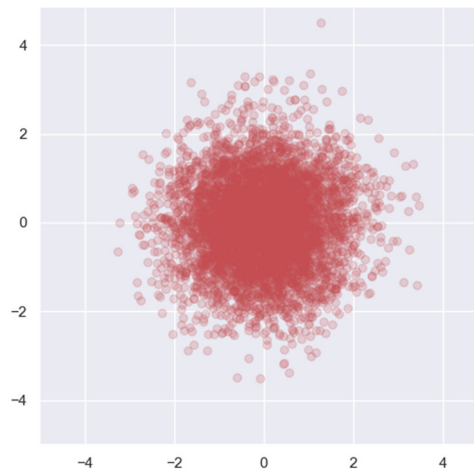


图1 服从二维标准正态分布的 5000 个样本点

Fig.1 5000 sample points that follow two-dimensional standard normal distribution

2. 接着水平轴伸缩变换，变换矩阵为 $\begin{pmatrix} 10 & 0 \\ 0 & 1 \end{pmatrix}$ 。此时 2 个维度仍然是相互独立的。

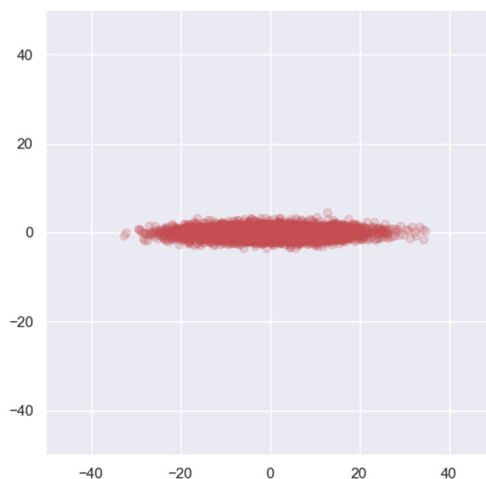


图 2 水平轴伸缩变换的结果

Fig.2 The result of stretching transformation of horizontal axis

3. 接着进行旋转变换，变换矩阵为 $\begin{pmatrix} \cos 75^\circ & -\sin 75^\circ \\ \sin 75^\circ & \cos 75^\circ \end{pmatrix}$ 。此时 2 个维度不再独立，存在线性关系，这是因为旋转变换后，每个维度都含有另一个维度的信息（如 $x' = x \cos 75^\circ - y \sin 75^\circ$ ）。

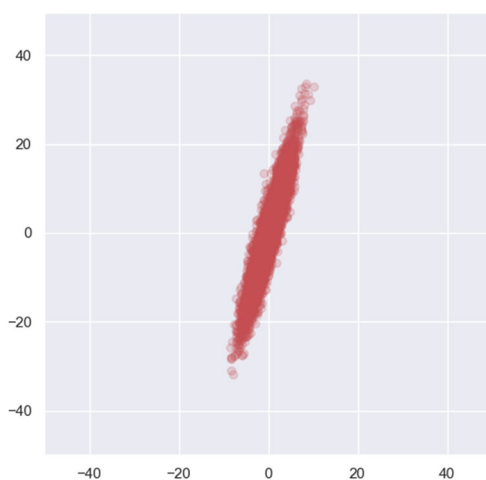


图 3 旋转变换的结果

Fig.3 The result of rotation transformation

4. 接着进行非线性切变变换，变换矩阵为 $\begin{pmatrix} 1 & x_i/40 \\ 0 & 1 \end{pmatrix}$ 。此时 2 个维度存在非线性关系。

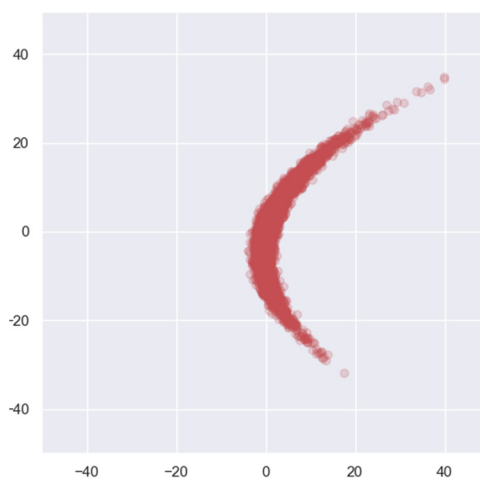


图 4 非线性切变变换的结果

Fig.4 The result of nonlinear shear transformation

可见，经过以上 1~4 步骤，将标准正态分布转换成了一个复杂的分布。

5. 上图复杂的分布也可以转换为标准正态分布。先进行非线性切变变换，变换矩阵为 $\begin{pmatrix} 1 & x_i/40 \\ 0 & 1 \end{pmatrix}$ 的逆矩阵。此步消除了 2 个维度非线性关系，回到之前的线性关系。

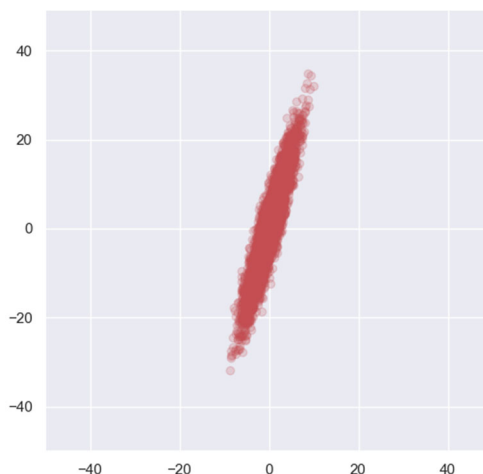


图 5 非线性切变逆变换的结果

Fig.5 The result of inverse nonlinear shear transformation

6. 接着进行旋转变换，变换矩阵为 $\begin{pmatrix} \cos 75^\circ & -\sin 75^\circ \\ \sin 75^\circ & \cos 75^\circ \end{pmatrix}$ 的逆矩阵。此步消除了 2 个维度线性关系，恢复之前的独立关系。

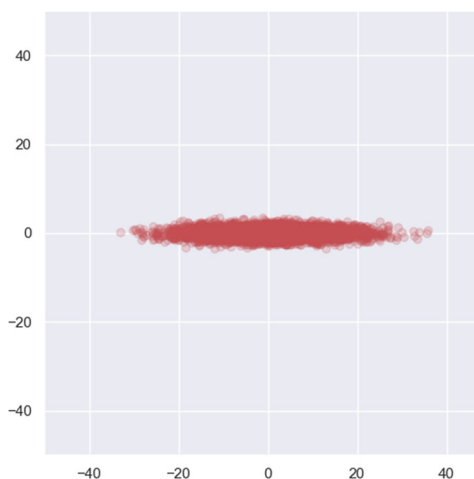


图 6 旋转逆变换的结果

Fig.6 The result of inverse rotation transformation

7. 接着水平轴伸缩变换，变换矩阵为 $\begin{pmatrix} 10 & 0 \\ 0 & 1 \end{pmatrix}$ 的逆矩阵，结果为标准正态分布。

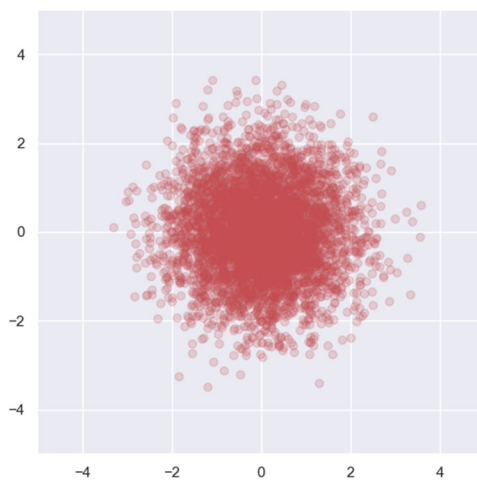


图 7 水平轴伸缩逆变换的结果

Fig.7 The result of inverse stretching transformation of horizontal axis

可见，经过以上 5~7 步骤，将一个复杂的分布转换成了标准正态分布。

1.3 随机变量函数的分布与雅可比行列式

标准化流维持样本空间的维度数目不变，因此概率分布的求解与转换是建模的关键问题。而这个关键问题实质上就是求解“随机变量函数的分布”。以下均采用微分的方法分析，当非线性变换时以直代曲。

1. 先举一个 1 维的实例。假设一个圆盘，其直径 D 在 $[5, 6]$ 上均匀分布，即 $p(D)=1$ ，求解圆盘面积 A 的概率分布密度 $p(A)$ 。

解：为了保证变换前后的概率相同，于是 $p(D)*dD = p(A)*dA$ ，即 $p(A) = p(D)*\left|\frac{dD}{dA}\right|$ ，如果取对数则为 $\ln p(A) = \ln p(D) + \ln \left|\frac{dD}{dA}\right|$ 。 $A = \pi/4 * D^2$ ， $A \in [25\pi/4, 9\pi]$ ， $D = |\pm\sqrt{4A/\pi}|$ ， $\left|\frac{dD}{dA}\right| = (\sqrt{4A/\pi})' = 1/\sqrt{A * \pi}$ ，故 $p(A) = 1 * 1/\sqrt{A * \pi} = 1/\sqrt{A * \pi}$ 。

$\left|\frac{dA}{dD}\right| = p(D) / p(A)$ ，可以认为 $\left|\frac{dA}{dD}\right|$ 是概率分布密度的缩放因子。

2. 再以二维为例。假设一个变换为 $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{cases} ax_1 + bx_2 \\ cx_1 + dx_2 \end{cases}$ ，为了保证变换前后的概率相同，于是

$p(y) * dA_y = p(x) * dA_x$ ，概率分布密度的缩放因子 $\left|\frac{dA_y}{dA_x}\right| = p(x) / p(y)$ 。

根据行列式的几何意义可知，行列式 $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 就是线性变换的图形面积伸缩因子。因此 $\left|\frac{dA_y}{dA_x}\right| =$

$$\left| \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right|。$$

故 $p(x) = p(y) * \left| \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right|$ 。

而 $\frac{\partial y_1}{\partial x_1} = \frac{\partial(ax_1 + bx_2)}{\partial x_1} = a$ ，推广到向量对向量求导，得 Jacobian 矩阵： $J = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 。

即： $p(x) = p(y) * |\det J|$ 。如果取对数则为 $\ln p(x) = \ln p(y) + \ln |\det J|$ 。

其它维度数目时同理。

1.4 最大似然估计法

最大似然估计法的基本思想是利用“概率最大的事件最可能出现”这一直观想法，选择合适的总体参数 $\hat{\theta}$ ，使样本集似然函数 $L(\theta)$ 达到最大值，并把 $\hat{\theta}$ 作为总体参数 θ 的估计值。

例如：抛掷一枚质地可能不均匀的硬币 4 次，结果为正面、反面向上各是 2 次，求硬币正面向上的先验概率 p 是多少？

该题粗略分析如下：先验概率 p 可能是 $0 \sim 1$ 之间的任意值，但是不同的值，本题结果的联合分布律 $L(p)$ 不同。比如 $p=0.4$ ，则本题结果的联合分布律 $L(p=0.4) = p(x_1) * p(x_2) * p(x_3) * p(x_4) = \frac{4}{10} * \frac{4}{10} * \frac{6}{10} * \frac{6}{10}$ 。

$\frac{6}{10} = \frac{576}{10000}$ ； $p=0.5$ ，则本题结果的 $L(p=0.5) = \frac{5}{10} * \frac{5}{10} * \frac{5}{10} * \frac{5}{10} = \frac{625}{10000}$ ； $p=0.6$ ，则本题结果的 $L(p=0.6) = \frac{6}{10} * \frac{4}{10} * \frac{4}{10} * \frac{4}{10} = \frac{576}{10000}$ 。

很明显， $p=0.5$ 作为总体参数的估计值较为合适。严密求解过程略。

为了计算方便，似然函数一般取对数值， $\ln L(\theta) = \sum_{i=1}^n \ln p(x_i)$ 。当总体为连续型随机变量时， $p(x_i)$

为概率密度。

1.5 损失函数

复杂的分布，如果直接对样本空间的联合概率分布建模，难度巨大。而我们能够写出高维正态分布概率密度函数的解析式，利用随机变量函数的分布，倒算出样本空间的联合概率分布。

标准化流使用分布变换的方法，对样本空间间接建模。通过计算样本在标准正态分布空间的概率密度 $p(y)$ 、神经网络变换时的概率密度缩放因子 $|\det J|$ ，间接求出样本空间的概率密度 $p(x)$ ，即 $\ln p(x) = \ln p(y) + \ln |\det J|$ 。

训练数据集实质上就是从总体中的抽样，而最大似然估计法就是通过抽样来对总体参数进行统计推断。所以将似然函数 $\ln L(\theta)$ 最大化，就能得到合理的总体参数 θ 的估计值。也即损失函数 $\text{Loss} = -\ln L(\theta) = -\sum_{i=1}^n \ln p(x_i)$ 。

1.6 实践时的挑战

标准化流在实践时面临两个挑战：①计算高维矩阵行列式的算力开销过大；②常规的神经网络计算都是单向传播的，无法做到输入、输出的无损可逆变换。

这里以 RealNVP 模型为例^[5,8]，展示如何构建一个神经网络，它把复杂分布变换为简单高斯分布，行列式易于求解、具有可逆性。总体流程如下：

1. 训练时：首先通过掩码，将训练集样本的坐标划分为两个部分 $X_{1:d}$ 、 $X_{d+1:D}$ 。让 $X_{1:d}$ 通过耦合层算出缩放因子 s 、平移因子 t ，用这两个因子对 $X_{d+1:D}$ 进行坐标变换 ($z = \frac{x-t}{s}$)，得到隐空间坐标 $Z_{d+1:D}$ 。而隐空间坐标 $Z_{1:d}$ 直接等于 $X_{1:d}$ 。以最简单的二维情况为例， $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ -\frac{t}{s} + \frac{1}{s} * x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ c & \frac{1}{s} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ ，注意这是下三角矩阵。同时耦合层是成双堆叠的，故交替使用掩码模式，从而将不变的那部分坐标在下一层中更新。

如此非常规操作的原因，是为了让每步坐标变换的矩阵，成为下三角矩阵，这让行列式等于对角线元素的乘积，这很容易计算。而矩阵乘积的行列式是行列式的乘积，故耦合层堆叠不会增加行列式计算难度，从而解决了第一个挑战。

2. 采样时：首先通过掩码，将隐空间采样坐标划分为两个部分 $Z_{1:d}$ 、 $Z_{d+1:D}$ 。让 $Z_{d+1:D}$ 通过耦合层算出缩放因子 s 、平移因子 t （注意训练时 $Z_{d+1:D} = X_{d+1:D}$ ，而耦合层权重参数是来自训练阶段的，故这里的 s 、 t 数值必然与训练时一致），用这两个因子对 $Z_{1:d}$ 进行坐标变换 ($x = z * s + t$)，得到样本空间坐标 $X_{1:d}$ 。而 $X_{d+1:D}$ 直接等于 $Z_{d+1:D}$ 。在多个层中交替使用掩码模式，将不变的一部分坐标在下一层中更新。这里层的顺序，与训练时是完全相反的。两个函数复合之逆就是其逆的复合，故解决了第二个挑战。

1.7 Glow 模型

Glow 在 2018 年的 NeurIPS 发表，它是首批展示标准化流能够生成高质量样本、有意义的隐空间的模型之一。关键步骤是用可逆的 1×1 卷积层取代了交替掩码设置。例如，在将 RealNVP 应用于图像的情况下，每个步骤后都会翻转通道的顺序，以确保网络有机会转换所有输入。在 Glow 中，取而代之的是应用 1×1 卷积，它有效地作为一种通用方法来产生模型所需的通道的任何排列。作者表明，即使有了这种添加，整个分布仍然易于处理，行列式和逆变换易于大规模计算^[8]。

2 从 Glow 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^[1-3]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。

对于 512x128 像素的图像，标准化流隐空间维度数目为 65536 维。因此需要足够多的样本，才能取得较好的学习结果。这里采取两个措施，来加大数据集样本数量：①用 OpenCV 对原数据集的图像进行插值，从而将图片总量由 9600 张增加到 108416 张；②使用数据增强，即在每张图像中增加高斯噪声。（注：这两个措施，治标不治本，唯有增加桥型种类才是更好的抽样。）



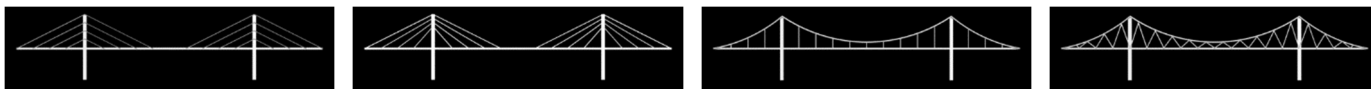


图 8 各桥型立面灰度图

Fig.8 Grayscale image of each bridge facade

每个子类桥型 13552 张各不相同的图片，整个数据集共 108416 张图片。

2.2 Glow 模型构建和训练

1. 基于 TensorFlow Probability 库中的 Glow API，构建和训练标准化流。具体代码如下：

```
#来自https://tensorflow.google.cn/probability/api_docs/python/tfp/distributions/TransformedDistribution
flow = tfd.TransformedDistribution(
    distribution= tfd.MultivariateNormalDiag( tf.zeros((tf.math.reduce_prod(128, 512, 1), )), #多维高斯分布的均值
                                              tf.ones((tf.math.reduce_prod(128, 512, 1), )), #多维高斯分布的标准差
    bijector = tfb.Glow(output_shape=(128, 512, 1),
                        num_glow_blocks = 6, #下采样的次数，这必须平分图像高H和宽W，否则双射体将不可逆。
                        num_steps_per_block=8, #指定在空间层次的每个级别上要包含多少仿射耦合和1x1卷积层。
                        coupling_bijector_fn=GlowDefaultNetwork, #取自定义的
                        exit_bijector_fn=tfb.GlowDefaultExitNetwork, ), ) #取默认的
```

图 9 Glow 模型参数设置

Fig.9 Parameter Settings of Glow

2. GlowDefaultNetwork 模块，神经元数目默认 400，对硬件要求过高，故需要降低该数目。因该参数无法通过形参传递，故复制源代码重写：

```
#来自https://tensorflow.google.cn/probability/api_docs/python/tfp/bijectors/GlowDefaultNetwork
#2024.01.07唯一修改之处: num_hidden=400改为num_hidden=num_hidden
class GlowDefaultNetwork(tfk.Sequential):
    def __init__(self, input_shape, num_hidden=num_hidden, kernel_shape=3):
        this_nchan = input_shape[-1] * 2
        conv_last = functools.partial(
            tfkl.Conv2D,
            padding='same',
            kernel_initializer=tf.initializers.zeros(),
            bias_initializer=tf.initializers.zeros())
        super(GlowDefaultNetwork, self).__init__([
            tfkl.Input(shape=input_shape),
            tfkl.Conv2D(num_hidden, kernel_shape, padding='same',
                        kernel_initializer=tf.initializers.he_normal(),
                        activation='relu'),
            tfkl.Conv2D(num_hidden, 1, padding='same',
                        kernel_initializer=tf.initializers.he_normal(),
                        activation='relu'),
            conv_last(this_nchan, kernel_shape)
        ])
    ])
```

图 10 GlowDefaultNetwork 模块的参数设置

Fig.10 Parameter Settings of GlowDefaultNetwork

3. 为了消除损失 loss 出现 inf、nan 的问题，需要将优化器 Adam 的学习率取 0.0001，且进行权重梯度剪裁 (clipnorm=1)。

4. tfd.TransformedDistribution 模块提供了 "log_prob()" 方法，输出样本在样本空间的概率密度，用户可以使用该值直接建立损失函数，从而让缺乏高等数学背景知识的用户也能够快速建立自己的标准化流模型。

5. 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，生成桥型图像来自多个事先保存的权重模型。

2.3 隐空间采样探索新桥型

使用 tfd.TransformedDistribution 模块提供了 "sample()" 方法采样，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 5 种技术可行的新桥型：

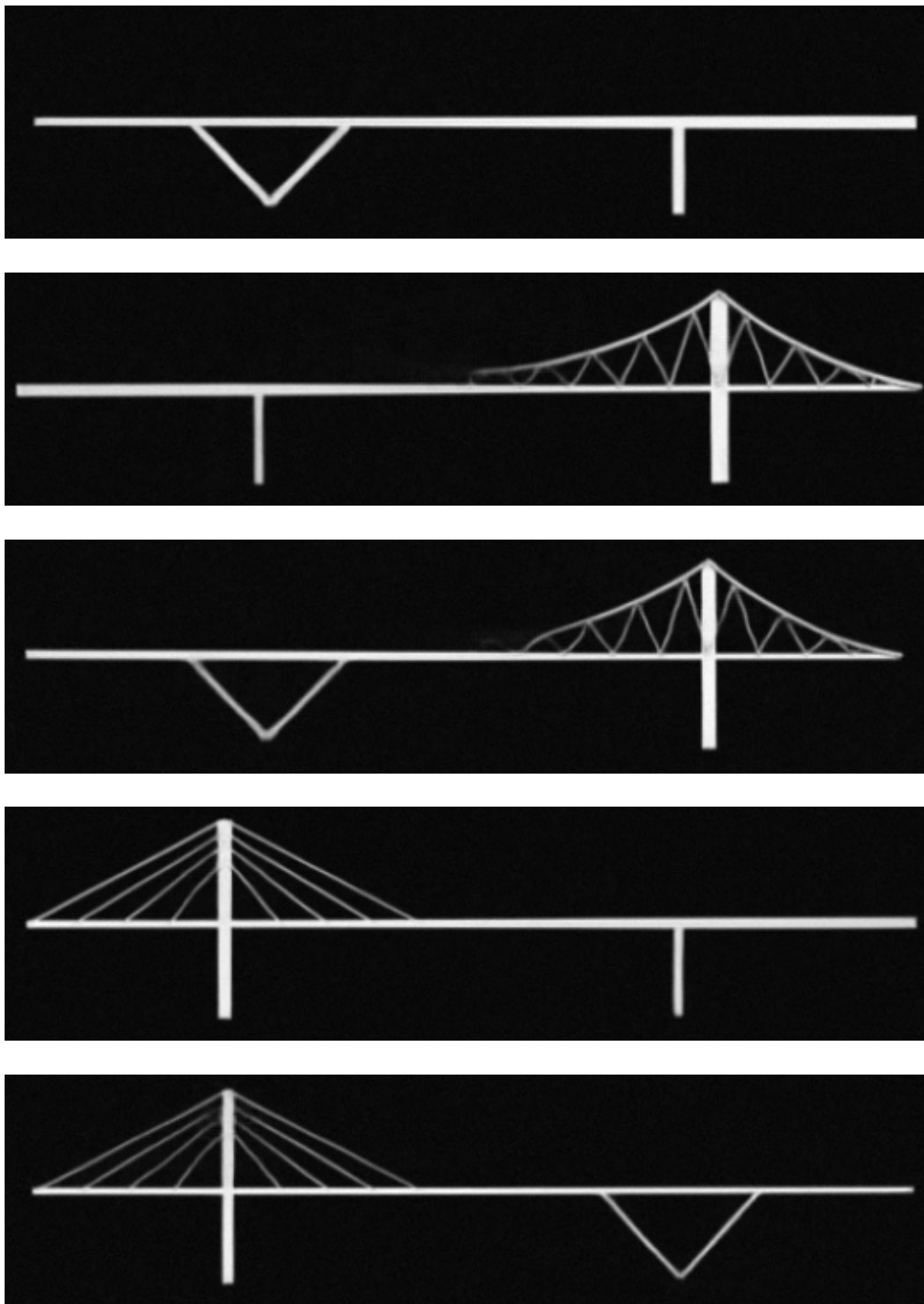


图 11 5 种技术可行的新桥型

Fig.11 Five 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。

2.4 结果分析

数据集的桥型全是对称结构，而标准化流能够生成不对称的桥型，且并不是简单的叠合，而是将不同结构部件进行有机搭配，这点与生成对抗网络类似。

3 结语

标准化流与生成对抗网络类似，比变分自编码器更有创造力，能够在人类原创桥型的基础上，将不同结构部件进行有机搭配，创造生成新桥型，一定程度上具有类似人类的原创能力，它能够打开想象空间，给予人类启发。

参考文献

- [1] 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>.
- [2] 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>.

[3] 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>.

[4] Laurent Dinh, David Krueger, Yoshua Bengio. NICE Non-linear Independent Components Estimation[J].arXiv preprint,2014,arXiv: 1410.8516.

[5] Laurent Dinh, Jascha Sohl-Dickstein, Samy Bengio. Density Estimation Using Real NVP[J].arXiv preprint,2016,arXiv: 1605.08803.

[6] Diederik P. Kingma, Prafulla Dhariwal. Glow: Generative Flow with Invertible 1x1 Convolutions[J].arXiv preprint,2018,arXiv: 1807.03039.

[7] Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, David Duvenaud. FFJORD: Free-Form Continuous Dynamics for Scalable Reversible Generative Models[J].arXiv preprint,2018,arXiv: 1810.01367.

[8] David Foster. Generative Deep Learning[M].2nd Edition. America: O' Reilly,2023.

[9] Tfp.bijectors.Glow [EB/OL]. https://tensorflow.google.cn/probability/api_docs/python/tfp/bijectors/Glow, 2023-11-21.