

从变分自编码器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: 尝试利用生成式人工智能技术生成新桥型。采用 3dsMax 动画软件渲染构件宽度变化的桥梁立面灰度图片、接着 OpenCV 模块对图片进行适量的几何变换(旋转、水平缩放、竖向缩放), 获得三跨梁式桥、拱式桥、斜拉桥、悬索桥图像数据集。基于 Python 编程语言、TensorFlow 及 Keras 深度学习平台框架, 构建和训练变分自编码器, 得到便于向量运算的低维桥型隐空间, 实践发现从隐空间中采样能够生成新的组合桥型。变分自编码器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。生成式人工智能技术能够协助桥梁设计师进行桥型创新、可以作为虚拟助手。

关键词: 生成式人工智能; 桥型创新; 变分自编码器; 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 variational autoencoder

ZHANG Hong-jun

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

Abstract: Try to generate new bridge types using generative artificial intelligence technology. The grayscale images of the bridge facade with the change of component width was rendered by 3dsMax animation software, and then the OpenCV module performed an appropriate amount of geometric transformation (rotation, horizontal scale, vertical scale) to obtain the image dataset of three-span beam bridge, arch bridge, cable-stayed bridge and suspension bridge. Based on Python programming language, TensorFlow and Keras deep learning platform framework, variational autoencoder was constructed and trained, and low-dimensional bridge-type latent space that is convenient for vector operations was obtained. Variational autoencoder can combine two bridge types on the basis of the original of human into one that is a new bridge type. Generative artificial intelligence technology can assist bridge designers in bridge-type innovation, and can be used as copilot.

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

1 引言

桥梁结构类型有梁式桥、拱式桥、斜拉桥、悬索桥四种。这四种桥型是否就是全部? 是否还存在未被人类发现的新桥型? 能否找到一种方法来协助人类桥梁设计师进行桥型创新? 本文利用生成式人工智能技术, 尝试去回答。

自上世纪中期人工智能技术诞生以来, 人们一直尝试将该技术应用到桥梁工程领域。专家系统(Expert System)上世纪曾在桥梁领域出现过^[1], 通过人机问答方式帮助桥梁工程师选用合适的工程技术, 但是因人类专家群体对诸多技术问题本身就有分歧、专家系统维护困难等因素而没落; 姜绍飞基于神经网络进行结构优化与损伤检测^[2]; 舒昕探析了人工智能技术在桥梁管理与养护中的应用^[3]; Sophia V. Kuhn 采用聚类、决策树等算法对系杆拱桥初步设计进行优化^[4]; Vera M. Balmer 采用条件变分自编码器对人行天桥的概念设计进行优化^[5]。但是国内外采用人工智能技术进行桥梁类型创新的研究尚无。

从隐空间(Latent Space)中采样生成全新图像或者在隐空间中向量运算编辑现有图像, 这是目前最流行也是最成功的创造性人工智能应用。本文采用该领域中最基础、最广为人知的变分自编码器(VAE)架构, 来尝试桥型创新(本文数据集与源代码开源地址 <https://github.com/QQ583304953/Bridge-VAE>)。

2 隐空间和变分自编码器的简介

2.1 概述

生成式人工智能技术已在绘画、文学等领域, 帮助人类进行创作, 如 Stable Diffusion、ChatGPT 等应用。其内核算法架构是 LSTM、DeepDream、神经风格迁移、VAE、GAN、Transformer 等。该技术能够对图像、语言等的隐空间进行学习, 然后从隐空间中采样, 创造出与模型在训练数据中所见到的作品具有相似特征的新作品^[6]。

2.2 隐空间

隐空间是用来学习数据的内在特征，简化数据的表达（降低数据的维度），通过对数据进行高度概括以便找出统计规律^[7]。以一个正交直桥的平面投影图数据集为例（图 1）：

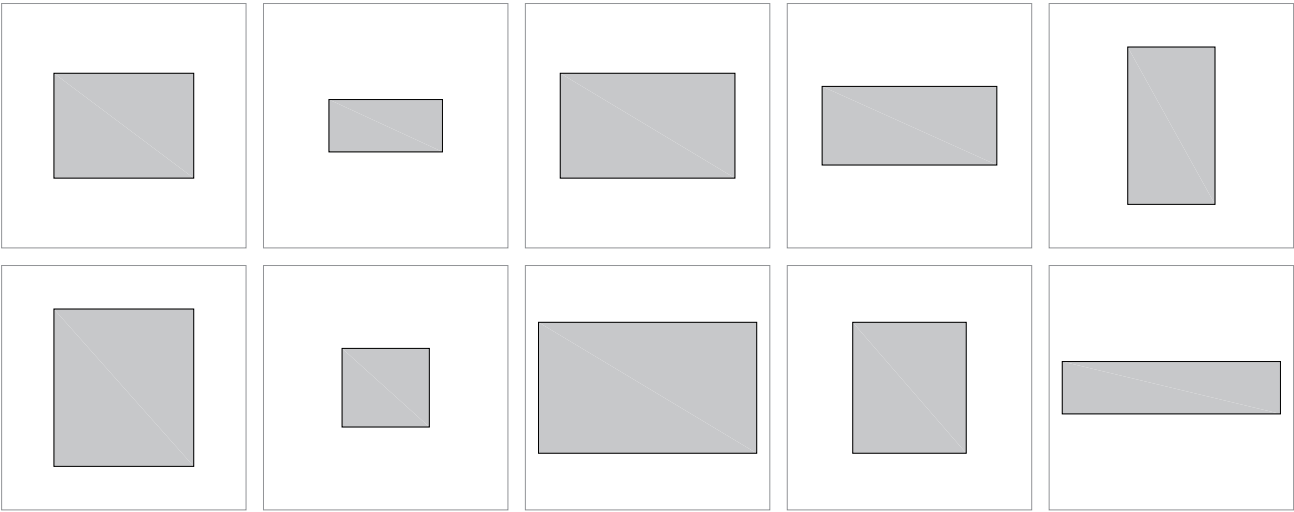


图 1 桥梁平面投影图数据集
Fig. 1 Plane projection dataset of bridge

对人类来说，很明显有两个特征（对应数学模型的两个维度），即桥梁长度 L 和桥梁宽度 B ，可以表示数据集中的每一个样本。给定桥梁长度 L 和桥梁宽度 B ，我们就可以绘制相应的桥梁平面投影图，即使这副图不在已知的数据集中。

对于机器学习而言，它首先需要确定桥梁长度 L 和桥梁宽度 B 是最能描述该数据集的两个隐空间维度，然后学习映射函数 f ，该函数可以将隐空间（向量空间）中一个点坐标（桥梁长度 L ，桥梁宽度 B ），映射成桥梁平面投影图。桥梁平面投影图的隐空间及从隐空间采样生成过程如图所示（图 2）：

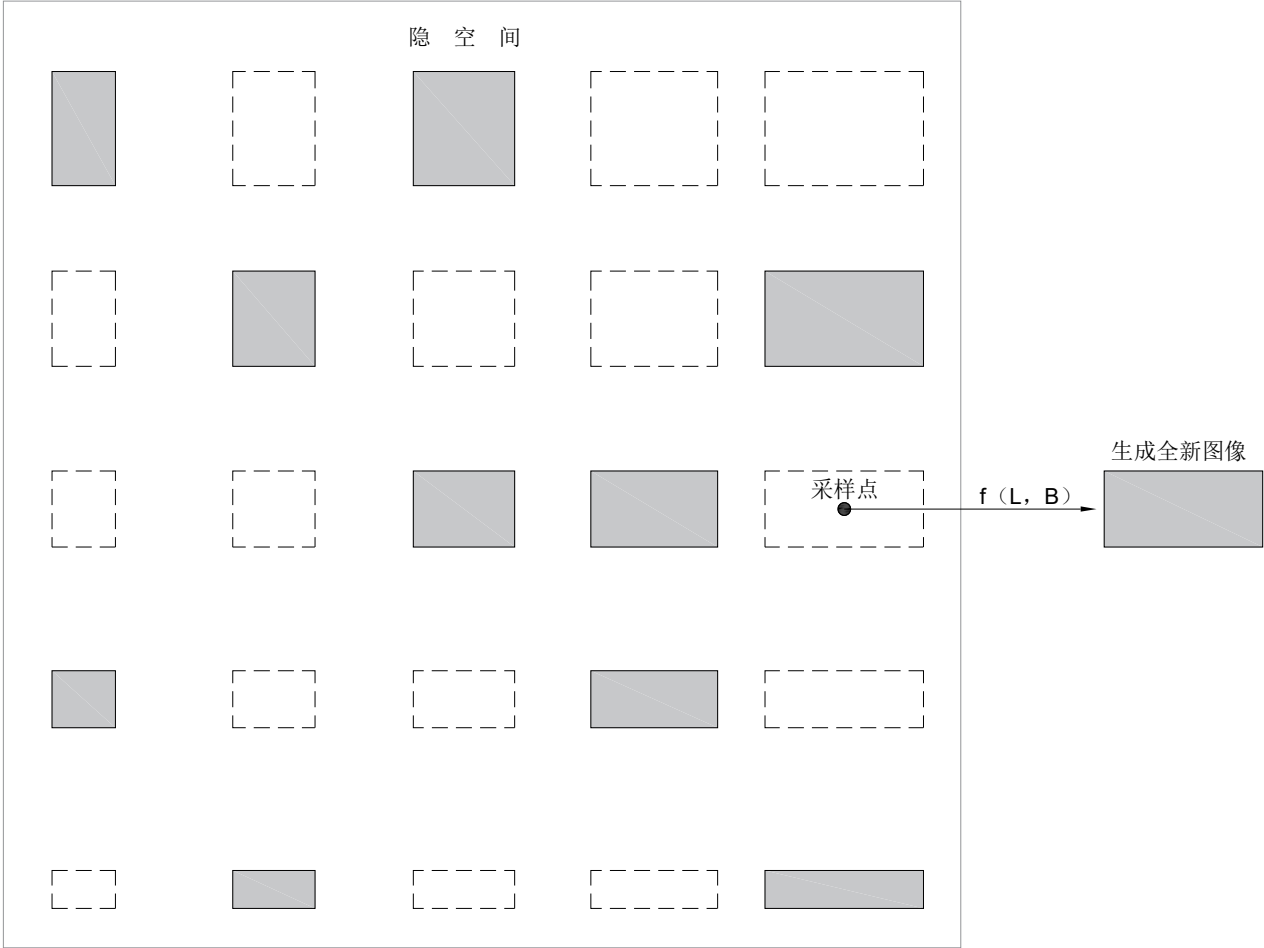


图2 隐空间及从隐空间采样示意图

Fig.2 Schematic diagram of latent space and sampling from latent space

2.3 变分自编码器

变分自编码器具有算力开销低、容易训练、隐空间结构连续、便于向量运算等优点，从桥型创新尝试的角度，其是最适合的算法。

先简要介绍一下自编码器（AutoEncoder）：它包括一个编码器（Encoder）和一个解码器（Decoder）。其架构图如图所示（图3）：

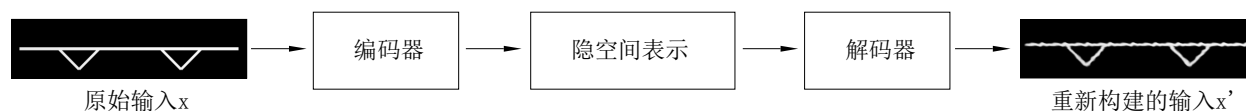


图3 自编码器架构图

Fig.3 Architecture diagram of autoencoder

变分自编码器就是在自编码器结构上，给编码结果（隐空间的表示）添加合适的随机性（噪声）。训练时，变分自编码器的编码器不是将原始输入 x 压缩成隐空间中的固定编码，而是将输入转换成隐空间中的统计分布（正态分布，用平均值和方差表示），然后变分自编码器的解码器使用统计分布从隐空间中随机采样，并将这个带有随机性的采样，解码成重新构建的输入 x' [6]。

变分自编码器通过两个损失函数来进行训练[6]：一个是重构损失（Reconstruction Loss），它迫使解码后的样本与初始输入尽可能相同；另一个是正则化损失（Regularization Loss），它是样本在隐空间中的统计分布与标准正态分布之差。因为编码的结果是统计分布，所以隐空间中任意两个相邻的点都会被解码为高度相似的结果，故隐空间存在连续性。隐空间的低维特性，迫使其每个维度都表示数据集一个有意义的变化轴，这使得其具有良好的结构，非常适合通过向量运算来进行操作。

3 从变分自编码器隐空间中生成新桥型的尝试

3.1 自建数据集

受成本制约，作者难以收集到足够多的所需真实桥梁照片，故采用 3dsmax 动画软件、opencv 模块自建桥型图像数据集。

为了降低本次尝试的难度，每种桥型仅自建两种子类（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）。

具体步骤为：①在 3dsMax 中建立桥梁模型，将构件宽度变化设置为动画（如梁高由1m渐变到4m），每种子桥型渲染16张桥梁立面灰度图片（图4），512x128像素、png格式；②接着采用 OpenCV 模块对图片进行适量的几何变换，依次为旋转（角度为-0.3、0、0.3度三种）、水平缩放（比例在1~1.05内插五种）、竖向缩放（比例在1~1.1内插五种），为了降低训练难度未采用平移几何变换。

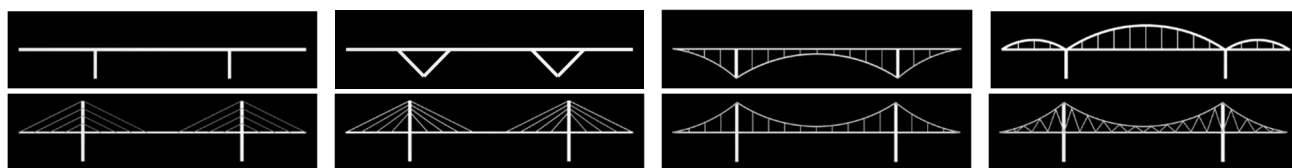


图4 各桥型立面灰度图

Fig.4 Grayscale image of each bridge facade

最终每个子类桥型得到 $16 \times 3 \times 5 \times 5 = 1200$ 张各不相同的图片，整个数据集共 9600 张图片。

3.2 变分自编码器构建

基于 Python 编程语言、TensorFlow 及 Keras 深度学习平台框架，构建和训练变分自编码器。经过测试，8 维隐空间能够取得较为理想的结果。

（1）编码器的架构：

编码器的工作是获取输入图像，将其映射成隐空间中的 8 维正态分布。

先为图像创建一个输入层，然后依次传递给 5 个 Conv2D 层（激活函数 relu），最后展平并连接到两个并列 Dense 层（线性变换），输出结果是正态分布的均值和方差（自然对数值）。模型摘要见下表：

表1 编码器模型摘要

Tab.1 model summary of encoder

Layer (type)	Output Shape	Param #	Connected to
--------------	--------------	---------	--------------

input_1 (InputLayer)	(None, 128, 512, 1)	0	
conv2d (Conv2D)	(None, 64, 256, 64)	640	input_1[0][0]
batch_normalization	(None, 64, 256, 64)	256	conv2d[0][0]
activation (Activation)	(None, 64, 256, 64)	0	batch_normalization[0][0]
dropout (Dropout)	(None, 64, 256, 64)	0	activation[0][0]
conv2d_1 (Conv2D)	(None, 32, 128, 128)	73856	dropout[0][0]
batch_normalization_1	(None, 32, 128, 128)	512	conv2d_1[0][0]
activation_1 (Activation)	(None, 32, 128, 128)	0	batch_normalization_1[0][0]
dropout_1 (Dropout)	(None, 32, 128, 128)	0	activation_1[0][0]
conv2d_2 (Conv2D)	(None, 16, 64, 128)	147584	dropout_1[0][0]
batch_normalization_2	(None, 16, 64, 128)	512	conv2d_2[0][0]
activation_2 (Activation)	(None, 16, 64, 128)	0	batch_normalization_2[0][0]
dropout_2 (Dropout)	(None, 16, 64, 128)	0	activation_2[0][0]
conv2d_3 (Conv2D)	(None, 8, 32, 128)	147584	dropout_2[0][0]
batch_normalization_3	(None, 8, 32, 128)	512	conv2d_3[0][0]
activation_3 (Activation)	(None, 8, 32, 128)	0	batch_normalization_3[0][0]
dropout_3 (Dropout)	(None, 8, 32, 128)	0	activation_3[0][0]
conv2d_4 (Conv2D)	(None, 4, 16, 128)	147584	dropout_3[0][0]
batch_normalization_4	(None, 4, 16, 128)	512	conv2d_4[0][0]
activation_4 (Activation)	(None, 4, 16, 128)	0	batch_normalization_4[0][0]
dropout_4 (Dropout)	(None, 4, 16, 128)	0	activation_4[0][0]
flatten (Flatten)	(None, 8192)	0	dropout_4[0][0]
dense (Dense)	(None, 8)	65544	flatten[0][0]
dense_1 (Dense)	(None, 8)	65544	flatten[0][0]
Total params: 650640			
Trainable params: 649488			
Non-trainable params: 1152			

（2）训练时隐空间样本随机采样坐标计算

每个输入样本被编码器转换为隐空间中的 8 维正态分布。这里为了方便讨论，取某一维为例，计算样本在该维度的随机采样坐标。根据数理统计原理，随机变量的标准化值=（原始值-平均值）/标准差，故原始值=平均值+标准差*标准化值，具体计算公式如下：

$$z = z_mean + \exp(0.5 * z_log_var) * \epsilon$$

式中：z 为随机变量的原始值，这里是隐空间样本随机采样坐标；z_mean 为随机变量的平均值，这里是编码器输出的平均值；z_log_var 为随机变量的方差（自然对数值），这里是编码器输出的方差；epsilon 随机变量的标准化值，这里代表采样的随机性。

计算示例：z_mean=1.72，z_log_var=-4.27，epsilon=3（标准正态分布取值 99.74%落在[-3, 3]区间），则 $z = 1.72 + \exp(-0.5 * 4.27) * 3 = 2.07$ 。

（3）解码器的架构：

解码器的工作是将隐空间中的坐标点映射成图像。

先创建一个坐标点输入层和 Dense 层，然后依次传递给 5 个 Conv2D Transpose 层（前 4 层激活函数 relu、最后一层激活函数 sigmoid），输出结果是该坐标点对应的重构图像。模型摘要见下表：

表 2 解码器模型摘要

Tab. 2 model summary of decoder

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	(None, 8)	0
dense_2 (Dense)	(None, 8192)	73728
reshape (Reshape)	(None, 4, 16, 128)	0
conv2d_transpose	(None, 8, 32, 128)	147584

batch_normalization_5	(None, 8, 32, 128)	512
activation_5 (Activation)	(None, 8, 32, 128)	0
dropout_5 (Dropout)	(None, 8, 32, 128)	0
conv2d_transpose_1	(None, 16, 64, 128)	147584
batch_normalization_6	(None, 16, 64, 128)	512
activation_6 (Activation)	(None, 16, 64, 128)	0
dropout_6 (Dropout)	(None, 16, 64, 128)	0
conv2d_transpose_2	(None, 32, 128, 128)	147584
batch_normalization_7	(None, 32, 128, 128)	512
activation_7 (Activation)	(None, 32, 128, 128)	0
dropout_7 (Dropout)	(None, 32, 128, 128)	0
conv2d_transpose_3	(None, 64, 256, 64)	73792
batch_normalization_8	(None, 64, 256, 64)	256
activation_8 (Activation)	(None, 64, 256, 64)	0
dropout_8 (Dropout)	(None, 64, 256, 64)	0
conv2d_transpose_4	(None, 128, 512, 1)	577
Total params: 592641		
Trainable params: 591745		
Non-trainable params: 896		

3.3 损失函数

(1) 重构损失

采用二元交叉熵来计算输入样本和解码重构图像之间的差异^[8]。

$$\text{reconstruction_loss} = 1/n * \sum \{y * \ln(1/y') + (1-y) * \ln[1/(1-y')]\}$$

式中：reconstruction_loss 为单个样本重构损失；y 为原始图像像素值；y' 为解码重构图像像素值；n 为像素数目；为了避免分母为 0，分母需加上微小的值（2e-07），式中未表达。

计算示例：y=[0.0, 0.1, 0.9, 1.0]，y'=[0.0, 0.9, 0.99, 0.1]，则

$$\text{reconstruction_loss} = 1/4 * (0 + 2.083 + 0.470 + 2.303) = 1.214。$$

(2) 正则化损失

采用 KL 散度（相对熵）来计算输入样本在隐空间中的统计分布与标准正态分布的差异^[7]。

$$\text{kl_loss} = 1/m * \sum \{-0.5 * [1 + z_log_var - \text{square}(z_mean) - \exp(z_log_var)]\}$$

式中：kl_loss 为单个样本正则化损失；z_mean 为编码器输出的平均值；z_log_var 为编码器输出的方差（自然对数值）；m 为隐空间维度数目。

计算示例：z_mean=[4.5, 3.3]，z_log_var=[-3.3, -3.7]，则 kl_loss=1/2*（11.3+6.8）=9.1。

(3) 变分自编码器损失函数是重构损失与正则化损失之和，多个样本取损失均值。

$$\text{total_loss} = \text{reconstruction_loss} + \text{kl_loss} * \text{coefficient}$$

式中：total_loss 为变分自编码器损失函数；coefficient 为损失权重相对系数。

如果重构损失占比过大，则训练样本编码统计分布会偏离标准正态分布过远，隐空间连续性差；如果正则化损失占比过大，则解码重构图像不够清晰。

3.4 训练

采用 RMSProp 优化器更新神经网络参数。数据集标签字典为：{'Arch Bottom_bear': 0, 'Arch Top_bear': 1, 'Beam Three_span': 2, 'Beam V_type': 3, 'Cable Fan_shaped': 4, 'Cable Harp_shaped': 5, 'Suspension Diagonal_sling': 6, 'Suspension Vertical_sling': 7}。

训练刚开始，神经网络参数默认初始值为很小的随机值，故解码图像像素值接近于 0（灰度图像 0 对应黑色，1 对应白色），重构损失 reconstruction_loss=3000*15/（128*512）=0.7 左右（每个输入样本约有 3000 个像素值接近 1，其它为 0），然后随着神经网络参数更新，重构损失会逐步降低。同理，训练刚开始均值、方差为 0，故正则化损失为 0，然后随着神经网络参数的更新，正则化损失会迅速增大成为控制要素，优化器将抑制它，使其逐步减少。

经过调试，损失权重相对系数 coefficient 取 0.001 较为合适，解码图像清晰、训练样本在隐空间分布合

理。由于无法同时显示 3 个以上维度，故这里采用一维直方图（图 5，图中光滑曲线为对比用标准正态分布曲线）、二维散点图（图 6），来展示全部训练样本在隐空间中的分布情况。

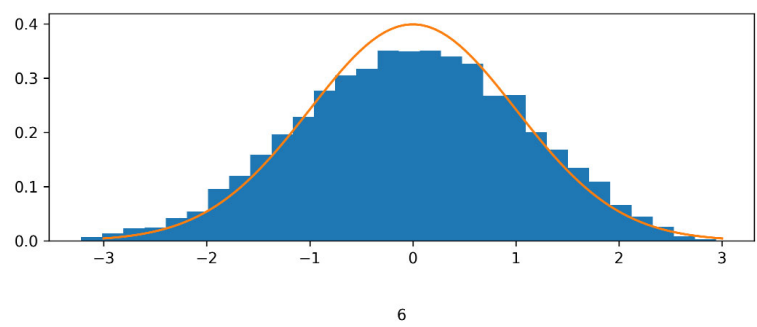


图 5 全部样本在第 6 维上的频率分布直方图

Fig.5 Frequency distribution histogram of all samples in the 6th dimension

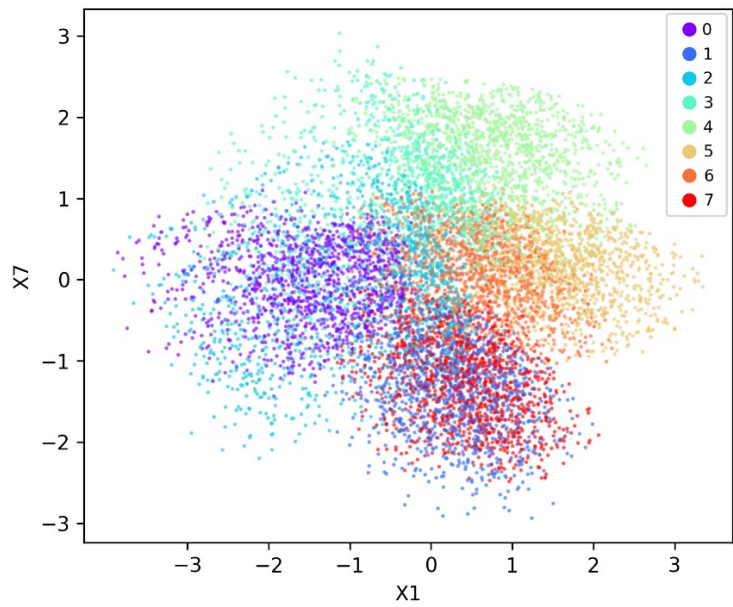


图 6 全部样本在 1、7 维度平面上的散点分布图

Fig.6 Scatter distribution map of all samples in the 1 and 7 dimensional planes

3.5 隐空间采样探索新桥型

想要细致了解 8 个维度的隐空间，理论上可以按照设定的密集空间网格从隐空间中采样。如果每个维度从-4 至+4 等距取 9 个点，那么采样总数目约为 4000 多万张图像（9 的 8 次方），可见如此多的采样，难以实际操作。

本次实践采样思路如下：

- （1）从隐空间的训练样本分布边界外围采样
- 分 4 次采样，每次每个维度仅取 2 个点（坐标分别为[-100, +100]、[-5, +5]、[-4, +4]、[-3, +3]），每次采样数目为 256 张图像（2 的 8 次方）。某次采样结果如下（图 7）：

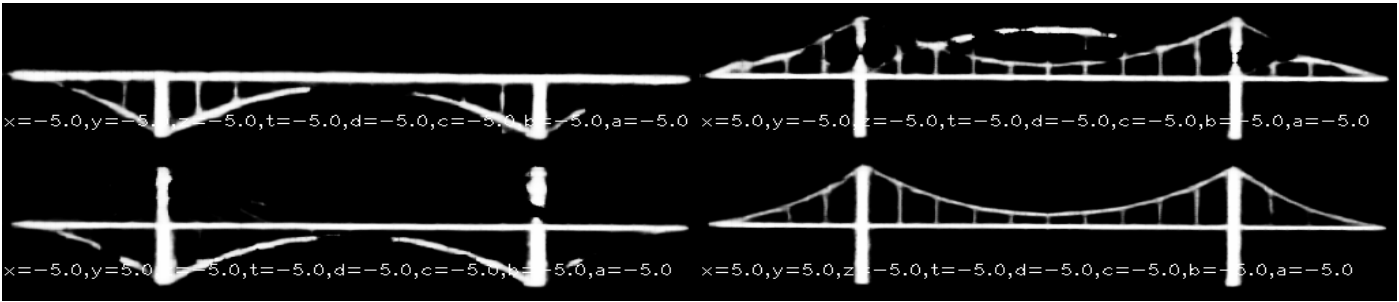


图 7 在样本分布边界外围采样生成的桥型立面图

Fig.7 Bridge facade generated by sampling outside the sample distribution boundary

经过人工探索，未能生成原创性新桥型，但是存在数种与训练样本不同的新组合桥型。因为采样点远离训练样本，故图像质量偏低。

(2) 从隐空间的训练样本分布边界内部采样

首先计算单个桥型所有样本集合的中心坐标, 然后绘制两个桥型之间的渐变图像。

八种桥型中任意选两种, 共有 28 种组合。经过人工探索, 未能生成原创性新桥型, 但是存在数种与训练样本不同的新组合桥型。人工基于工程结构思维筛选, 得到 5 种技术可行的新组合桥型 (图 8)。

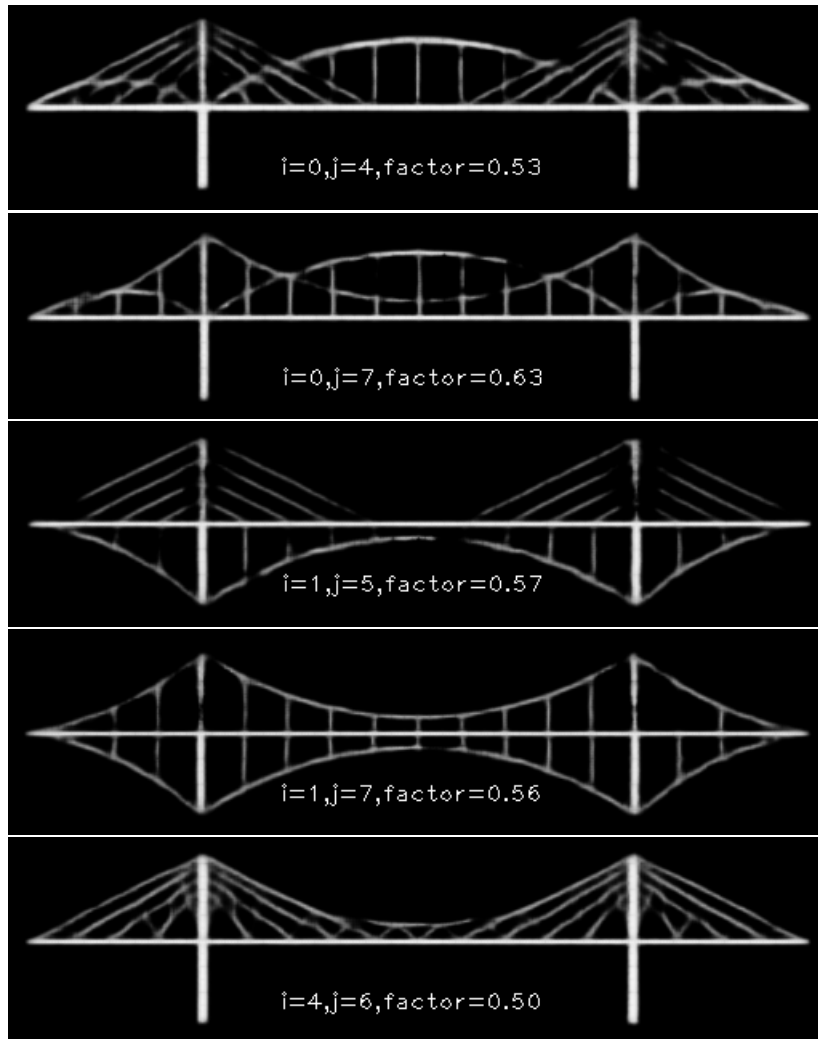


图 8 5 种技术可行的新组合桥型

Fig. 8 Five new combined bridge types with feasible technology

以上部分组合桥型受力复杂或结构不尽合理, 故未必适用于特大跨桥梁, 但可用于城市的景观桥梁。某些组合桥型, 如斜拉-悬索组合体系桥, 桥梁界已经构思和实践过^[9]。

4 结语

变分自编码器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。原创能力是人类的优势, 而组合创造是人工智能的优势, 生成式人工智能拥有巨大潜力, 其应该在桥梁设计领域占有一席之地, 可以作为桥梁设计师的副驾驶。

数据决定了机器学习的上限^[10], 而本次训练集的子桥型仅有 8 种, 这严重制约了变分自编码器的创新能力。如果增加数据集的规模, 甚至采用三维桥型, 那么低维隐空间将趋向于采用基本工程结构单元去描述训练样本, 这样更易生成组合桥型 (甚至生成新的原创桥型)。如果采用规模更大的神经网络模型, 不排除会产生类似大模型的涌现能力 (Emergent abilities)。

参考文献

- [1] Tommaso Pagnoni. 专家系统在桥梁工程中的应用[J]. 国外桥梁, 1990, (3): 21-28.
- [2] 姜绍飞. 基于神经网络的结构优化与损伤检测[M]. 北京: 科学出版社, 2002.
- [3] 舒昕, 沈翔, 李晓行等. 人工智能在桥梁管理与养护中的应用探析[J]. 电脑知识与技术, 2020, (13): 268-269.
- [4] Sophia V.Kuhn, Rafael Bischof, Georgios Klonaris 等. NTAB0: Design Priors for AI-Augmented Generative Design of Network Tied-Arch-Bridges[C]. Forum Bauinformatik, 2022: 437-444
- [5] Vera M.Balmer, Sophia V.Kuhn, Rafael Bischof 等. Design Space Exploration and Explanation via Conditional

Variational Autoencoders in Meta-model-based Conceptual Design of Pedestrian Bridges[J]. arXiv preprint, 2022, arXiv:2211.16406.

[6] 弗朗索瓦·肖莱. Python 深度学习[M]. 第二版. 北京:人民邮电出版社, 2023.

[7] 大卫·福斯特. 生成式深度学习[M]. 北京:中国电力出版社, 2021.

[8] 雷明. 机器学习的数学[M]. 北京:中国电力出版社, 2020.

[9] 周念先. 桥梁方案比选[M]. 上海:同济大学出版社, 1997.

[10] 伊恩·古德费洛. 深度学习[M]. 北京:人民邮电出版社, 2017.