



# 神经网络求解系统生物学中刚性问题的研究\*

张艳玲, 王梦收, 洪柳

中山大学数学学院, 广东 广州 510275

**摘要:** 在系统生物学的研究中, 由于所研究问题的复杂性和多尺度性, 经常会遇到刚性方程的求解. 而近年来, 神经网络和深度学习的发展为上述问题提供了新的解决思路和方法. 本研究以经典的Belousov-Zhabotinsky (B-Z) 反应和Van der Pol (VdP) 方程为例, 对四类非时序神经网络, 包括全连接网络、残差网络、改进的残差网络和深度混合卷积网络, 以及三类时序神经网络, 包括循环神经网络(RNN)、长短时记忆网络(LSTM)、注意力机制进行了系统比较. 实验结果表明: 时序神经网络应用于刚性问题的求解精度和计算时间都大幅优于非时序神经网络, 而四类非时序神经网络之间的表现并无显著差异. 此外还将常微分神经网络(ODE-Net)应用于上述刚性问题, 并观察到在极短的计算时间内, 该方法能够达到极高的精度. 本研究为应用神经网络解决系统生物学中各类刚性问题提供了参考和指导.

**关键词:** 系统生物学; 刚性微分方程; 神经网络; 常微分神经网络

**中图分类号:** O241.81 **文献标志码:** A **文章编号:** 2097-0137(2024)06-0265-10

## On solving stiff differential equations in system biology with neural networks

ZHANG Yanling, WANG Mengshou, HONG Liu

School of Mathematics, Sun Yat-sen University, Guangzhou 510275, China

**Abstract:** Stiff differential equations are very common in system biology, due to the intrinsic complexity and multi-scaling nature of the systems under study. In recent years, a variety of neural-network-based methods suitable for solving stiff differential equations have been proposed. In this study, the performance of four non-temporal neural networks, including fully connected networks, residual networks, improved residual networks, and deep mixed convolutional networks, as well as other three temporal neural networks, including recurrent neural networks (RNN), long short-term memory networks (LSTM), and attention mechanisms, are compared systematically with respect to the stiff Belousov-Zhabotinsky (B-Z) reaction and Van der Pol (VdP) equations. Extensive numerical results indicate that for solving stiff problems the accuracy of temporal neural networks is much higher than that of the non-temporal neural networks, the running time of the former is also shorter. Meanwhile, among the four types of non-temporal neural networks, no significant difference is observed. Finally, it is found that the neural ordinary differential equations (ODE-Net) can achieve extremely high accuracy within very little computational time when applying to stiff ordinary differential equations. This study provides insightful guidance for using neural networks to solve stiff differential equations in system biology.

**Key words:** system biology; stiff differential equations; neural networks; ODE-Net

\* 收稿日期: 2024-04-12 录用日期: 2024-05-29 网络首发日期: 2024-08-27

基金项目: 国家重点研发计划(2023YFC2308702);  
广东省基础与应用基础研究基金(2023A1515010157)

作者简介: 张艳玲(2000年生),女;研究方向:应用数学;E-mail: zhangyaling29@mail2.sysu.edu.cn  
通信作者: 洪柳(1981年生),男;研究方向:应用数学;E-mail: hongliu@mail.sysu.edu.cn



ZR20240116

在系统生物学中, 由于所研究问题的复杂性和多尺度性, 我们经常会遇到许多反应量级相差很大的过程, 从而使得系统中各个物质浓度变化速率差异过大, 即所谓“刚性”问题(肖爱国等, 1999). 比如在经典的带负反馈的基因转录和翻译网络中, 蛋白质作为调控因子结合和解离的速率一般会远远大于基因转录和翻译的速率, 从而使得基因时时刻刻处在被有效调控的状态(Thomas et al., 2012). 与此类似还有酵母细胞中的糖代谢网络通路(Fonseca et al., 2011), 基于实验数据所构建的多尺度模型中, 典型反应速率常数相差四个数量级以上, 从而使得酵母细胞“获得了”对各种糖源的不同响应速率. 此外各种带振荡效应的生物学系统, 如协作结合或非协作结合的双基因开关(Gardner et al., 2000)、生物节律振子(Asgari-Targhi et al., 2019)、钙离子振荡诱导的细胞分化(Sneyd et al., 2017)等, 也都很明显地展示出多个不同时间尺度的存在. 显然刚性问题的存在, 无论对于复杂系统生物学模型的数值求解, 还是基于数据的系统生物学模型反演或者未知参数的确定都带来了很大的挑战.

对于系统生物学研究中最常见一类常微分方程的初值问题, 我们采用Shampine和Gear给出的经典定义(曹学年, 2001).

**定义 1** 给定如下一般的常微分方程组

$$\frac{dY}{dt} = AY + C,$$

其中  $Y, C \in \mathbb{R}^n \times 1$ . 系数矩阵  $A \in \mathbb{R}^{n \times n}$  对应的特征值分别为  $\lambda_1, \lambda_2, \dots, \lambda_n$ . 其刚性比定义为:

$$S = \frac{\max_{1 \leq i \leq n} \{ |\operatorname{Re}(\lambda_i)| \}}{\min_{1 \leq j \leq n} \{ |\operatorname{Re}(\lambda_j)| \}},$$

其中分母不取零. 若  $S \gg 1$ , 则称该常微分方程是刚性的或病态的.

袁兆鼎等(1987)、Hairer et al.(1996)和肖爱国等(1999)总结了上述刚性常微分方程的多种数值求解方法, 如向后差分法、指数拟合法、线性多步法、显式和隐式的龙格库塔法、Rosenblock方法等, 以及各种方法的稳定性和收敛性. 步入 21 世纪, 相关研究主要集中在对单步法和多步法新公式的构造和对算法程序的设计上(冯伯培, 1982; 吴新元, 1999; 曹学年, 2001; 石滔, 2008). 其中邹积麟等(2001)利用结构力学的思想推导出更适用于程序化计算的复合结构解法, 刘晓岑等(2012)在已有的线性多步法公式的基础上作加权平均, 得出了一系列新的隐式线性多步法公式. 总体而言, 这些算法往往通过减小积分步长来换取解的精度, 但会使计算量大大增加, 而且算法的稳定性和收敛性等性质也需要被着重考虑.

近年来, 随着以深度神经网络和强化学习为代表的机器学习领域的蓬勃发展, 相关理论方法也被广泛应用于刚性方程求解、参数推断、模型反演等方面. 例如, 刚性物理信息神经网络(Stiff-PINN)(Ji et al., 2021)以物理信息神经网络(PINNs)为基础, 增加了准稳态假设来减小系统的刚性. 但准稳态假设技巧本身并不能完全消除系统(尤其是复杂系统)的刚性, 也无法处理含有多个不同快尺度的系统. Kim et al. (2021)对化学和系统生物学中的刚性问题进行了讨论, 提出了缓解刚性系统中尺度分离挑战的技术. 并且通过罗伯逊问题和空气污染问题表明, 使用具有修正激活函数的深度网络、适当缩放网络输出以及损失函数, 以及稳定梯度计算是实现刚性问题学习的关键所在. De Florio et al.(2022)将 PINNs 与功能连接理论和极限学习机相结合, 形成了所谓的极限功能连接理论, 仅使用单层神经网络就能够高效而稳健地求解包括 Belousov-Zhabotinsky 反应在内的若干化学动力学基准问题. Fabiani et al.(2023)提出了一种基于高斯核和 PINNs 的随机投影数值方法, 用于求解非线性刚性常微分方程和一阶微分代数方程组的初值问题. 为了处理刚性和陡峭的梯度, 提出了自适应步长方案, 并使用连续方法为牛顿迭代提供良好的初始猜测. 多尺度深度神经网络(MscaleDNNs)(Liu et al., 2020)利用频域径向缩放的思想, 将微分方程解的高频逼近问题转化为了低频函数的学习问题, 在求解刚性泊松-玻尔兹曼方程的例子得到了比传统的全连接神经网络更优的结果.

系统生物学模型中往往包含大量未知参数, 需要基于有限的实验数据进行推断. Yazdani et al.(2020)引入了一种新的深度学习方法, 即基于物理信息神经网络的系统信息神经网络. 其具有可靠和鲁棒的参数推断算法, 并能够预测系统中动态的隐藏特征. 此外, 由多步法和深度神经网络混合得到的多步神经网络(MultiNN)(Raissi et al., 2018)可以从给定数据中识别非线性动力系统. 但遇到非均匀时间步长的数据时,

该方法效果依旧有限. 针对从观察数据中发现控制物理和生物系统的数学方程这一基本挑战, Ahmadi et al. (2024)提出了物理信息框架 - AI-Aristotle. 它结合了极限功能连接理论领域分解、物理信息神经网络与符号回归技术, 用于系统生物学中的参数估计和缺失物理识别问题.

尽管神经网络在求解刚性微分方程方面取得了显著进展, 但其具体构建在很大程度上受到所研究问题的影响. 大部分的网络往往在处理特定问题时表现出色, 而在其他类型刚性问题的处理上则可能效果不佳. 这就引发了对通用性的迫切需求. 在本研究中, 我们将通过两类经典刚性常微分方程的求解来系统比较不同网络架构的性能, 深入了解各种网络在不同情境下的表现, 揭示它们在处理刚性问题方面的优势和劣势, 从而更好地理解它们的适用范围, 并为构建更加通用、灵活的神经网络提供指导. 我们期望能为神经网络在系统生物学领域的广泛应用打开新的局面, 并为未来的研究提供更灵活而强大的计算工具.

## 1 准备知识

我们用于测试对比的神经网络架构包括四类非时序神经网络 - 全连接网络、残差网络、改进的残差网络和深度混合卷积网络, 以及三类主流时序神经网络 - 循环神经网络、长短时记忆网络、注意力机制. 具体列举如下:

全连接神经网络(Zhang et al., 2021)(FCNN, 又被称为多层感知机 MLP)作为深度学习神经网络的基础形式, 其结构如图 1(a1)所示. 有普遍逼近理论保证, FCNN 的宽度越大其表示能力越强.

He et al. (2016)发现, 随着网络的加深, 训练精度会达到饱和, 然后迅速退化, 但这种现象并不是由拟合引起的. 于是他们引入了一种深度残差学习框架——残差网络(ResNet), 在原本网络的基础上添加了一个新的跳跃连接, 如图 1(a2)所示.

2019年, 南开大学程明明团队对 ResNet 进行了改进, 提出了残差网络的变种 Res2Net, 用于处理多尺度问题(Gao et al., 2021). 不同于其他通过分层来表示多尺度特征的方法, Res2Net 内部的单个残差块中就存在等级制的残差连接(如图 1(a3)), 从而增加了每层网络的感受野, 这让它能更精细地表示多尺度特征.

深度混合卷积网络(MixNet)最早是 Tan et al. (2019)在图像识别研究中所提出. 他发现在卷积网络中使用较大的核(如  $5 \times 5$  或  $7 \times 7$ )并不总能提高模型的精度, 特别是非常大的核(超过  $9 \times 9$ )可能会造成精度的急剧减小. 基于此观察, 作者提出了在单一卷积层中混合不同大小卷积核的想法, 如图 1(a4)所示.

循环神经网络的研究始于二十世纪 80-90 年代, 并在 21 世纪初发展为深度学习算法的关键组成部分之一(Schmidhuber, 2015). 循环神经网络因其在学习序列的非线性特征方面的优势(邱锡鹏, 2020)在自然语言处理等领域得到广泛应用. 该模型能够记忆历史序列信息并将其应用到当前输出的计算中. 其核心结构是循环单元, 这些单元通过链式连接递归地处理序列(如图 1(b1)).

长短时记忆网络(LSTM)是最早被引入的 RNN 门控算法, 其核心单元包含输入门、遗忘门和输出门三个主要部分. 与传统 RNN 通过递归计算来学习系统状态不同, LSTM 通过这三个门控实现内部自循环(Self-Loop)(Goodfellow et al., 2016), 从而更新系统动态. 具体而言, 输入门控制当前时间步的输入和上一时间步的系统状态对内部状态的更新, 遗忘门控制上一时间步的内部状态对当前时间步内部状态的更新, 输出门则控制内部状态对系统状态的更新. 详情可见图 1(b2).

深度学习领域中, 注意力机制主要被用于从海量信息中筛选出对当前任务目标最为重要的内容. 自注意力机制的基本思想在于, 序列数据中的每个元素(不仅仅是相邻位置的元素)都有机会与序列中的其他元素建立关联. 该机制可以通过计算得到的元素间的相对重要性自适应地捕捉元素间的长程依赖关系, 从而实现了对于序列数据的全局关联性建模, 如图 1(b3)所示.

针对上述各类神经网络, 我们将重点评测如下两个经典刚性问题. 第 1 个例子 B-Z 反应是由两位前苏联科学家 Belousov 和 Zhabotinsky 所发现的一个著名的化学振荡反应. 经过几十年的发展, B-Z 反应的高度理想化模型(Field et al., 1974; Tyson et al., 1980; Györgyi et al., 1992)表示为下面无量纲化后的微分方程组

$$\begin{cases} \dot{x} = (x + y - cx^2 - xy)/a, \\ \dot{y} = -y - xy + 2hz, \\ \dot{z} = (x - z)/b, \end{cases} \quad (1)$$

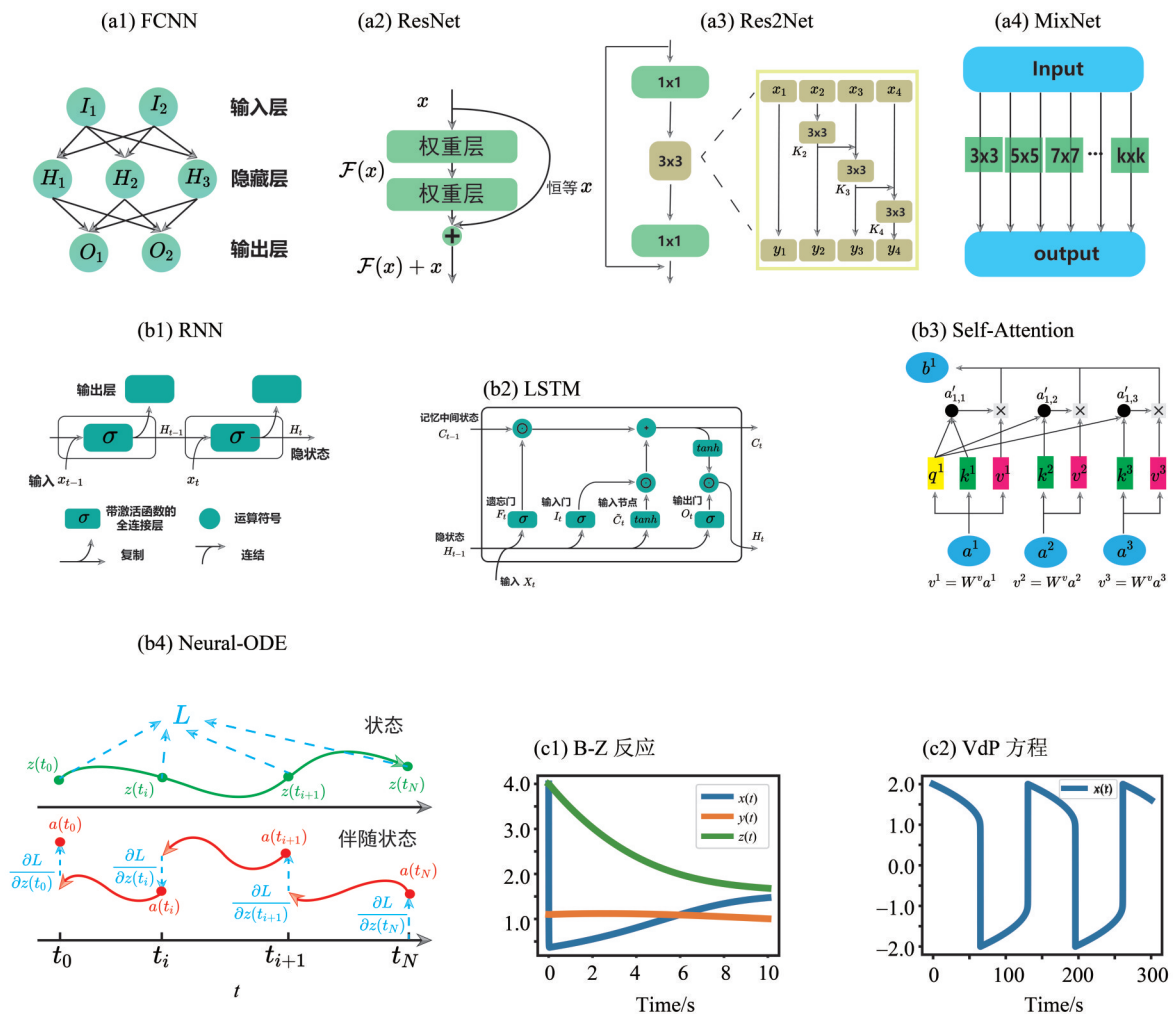


图 1 各类神经网络基本结构及 B-Z 反应和 VdP 方程

Fig. 1 Structures of various neural networks and illustrations of B-Z reaction and VdP equation

其中  $a, b, c$  和  $h$  为常值参数, 其具体取值将影响此常微分方程组刚性的强弱. 在后续评测中, 我们将使用具有化学背景的参数 (Field et al., 1974; Troy et al., 1977)  $h = 1/2, a = 1/77.27, b = 1/0.161, c = -8.375 \times 10^{-6}$ , 初值条件为  $x(0) = 4, y(0) = 1.1, z(0) = 4$ . 此时方程组的刚性比为  $1.934 \times 10^3 \gg 1$ . 这个例子最重要的一个特点是: 当系统的刚性比很大时, 会出现明显的初始层, 即方程的解一开始会在极短的时间尺度内发生巨大的改变.

第 2 个测试例子 VdP 方程最初由荷兰物理学家 Van der Pol 在 1926 年提出, 用于描述在真空管放大器中所观察到的极限环振荡现象 (Van der Pol, 1926). VdP 方程刻画的是非线性阻尼下的非保守振荡系统, 采用如下二阶微分方程表示:

$$\ddot{x} - \mu(1 - x^2)\dot{x} + x = 0, \tag{2}$$

其中  $x$  是位置,  $t$  是时间,  $\mu$  是一个指示非线性和阻尼强度的标量参数.  $\mu$  的不同取值会影响系统的特性. 当  $\mu = 0$  时, 此方程退化为普通的简谐振动方程, 系统满足能量守恒; 当  $\mu < 0$  时, 系统表现为阻尼振荡, 最终振幅逐渐衰减为 0, 系统趋于稳定; 当  $\mu > 0$  时, 方程呈现自激振荡, 所有初始条件下都会收敛到全局唯一的极限环. 特别地,  $\mu$  增大到一定程度时会变为刚性问题. 当参数设置  $x(0) = 2, x'(0) = 0, \mu = 80$  时, 方程刚性比为  $5.76 \times 10^4 \gg 1$ . 这个例子的复杂性主要体现在: 当系统刚性比较大时, 会呈现出显著的快慢时间尺度分离.

## 2 结果和讨论

下面我们将以B-Z反应和VdP方程为例, 深入评测不同神经网络架构在刚性方程求解逼近方面的性能优劣. 由于网络深度、参数量等都会影响网络的性能, 为了保持测评结果的可靠性, 应当尽量使得相关参数设置接近. 在后续实验部分, 我们将在保持相同网络层数的同时, 也尽可能使得自由参数个数维持在同一个数量级.

数据集方面, 对于B-Z反应取样范围为前10 s, 时间步长为 $10^{-4}$  s, 此时数据点的数量为 $10^5$ , 训练数据集的图像如图1(c1)所示. 对于VdP方程取样范围为前300 s, 时间步长为 $10^{-3}$  s, 此时数据点的数量为 $3 \times 10^5$ , 数据集的图像如图1(c2).

### 2.1 B-Z反应上的非时序神经网络

在此训练设置中, 按批量大小为100划分训练批次, 输入数据为时间 $t$ , 标签为对应的方程组的解 $x(t), y(t), z(t)$ . 我们重复训练5次, 进行了200个轮次的迭代(epoch), 以确保模型能够更充分地学习到数据的潜在特征和模式. 损失函数采用了回归模型中常用的均方误差(MSE)损失. 为引入非线性特性, 我们使用了ReLU作为激活函数, 并采用Adam方法进行优化. 具体更新过程如下: 首先对于每个批次, 计算网络输出与标签之间的损失函数值; 然后利用反向传播(BP)算法计算损失函数值对网络参数的梯度; 最后利用预先设置的优化方法更新网络参数, 记录损失函数变化情况及运行用时, 报告实验结果的均值, 见表1.

表1 非时序深度网络在B-Z反应和VdP方程上的参数设置和训练结果汇总<sup>1)</sup>

Table 1 Summary on parameter settings and training results of non-temporal neural networks for the B-Z reaction and VdP equation

项目	B-Z反应				VdP方程			
	ResNet	Res2Net	MixNet	FCNN	ResNet	Res2Net	MixNet	FCNN
非时序网络	ResNet	Res2Net	MixNet	FCNN	ResNet	Res2Net	MixNet	FCNN
参数量/ $10^5$	4.51	4.89	4.65	4.53	0.96	0.93	1.03	0.66
学习率/ $10^{-3}$	1.0	50	0.5	1.0	1.0	0.5	0.5	0.5
平均损失	$1.24 \times 10^{-2}$	$1.48 \times 10^{-2}$	$1.23 \times 10^{-2}$	$8.75 \times 10^{-4}$	$6.21 \times 10^{-2}$	$6.31 \times 10^{-2}$	$6.96 \times 10^{-2}$	$5.67 \times 10^{-2}$
平时时间/s	$1.56 \times 10^3$	$2.71 \times 10^3$	$6.43 \times 10^3$	$1.37 \times 10^3$	$7.08 \times 10^3$	$1.62 \times 10^4$	$1.30 \times 10^4$	$5.35 \times 10^3$

1) 激活函数选取ReLU, 此外, 网络中还包含批量归一化(Batch Normalization)层.

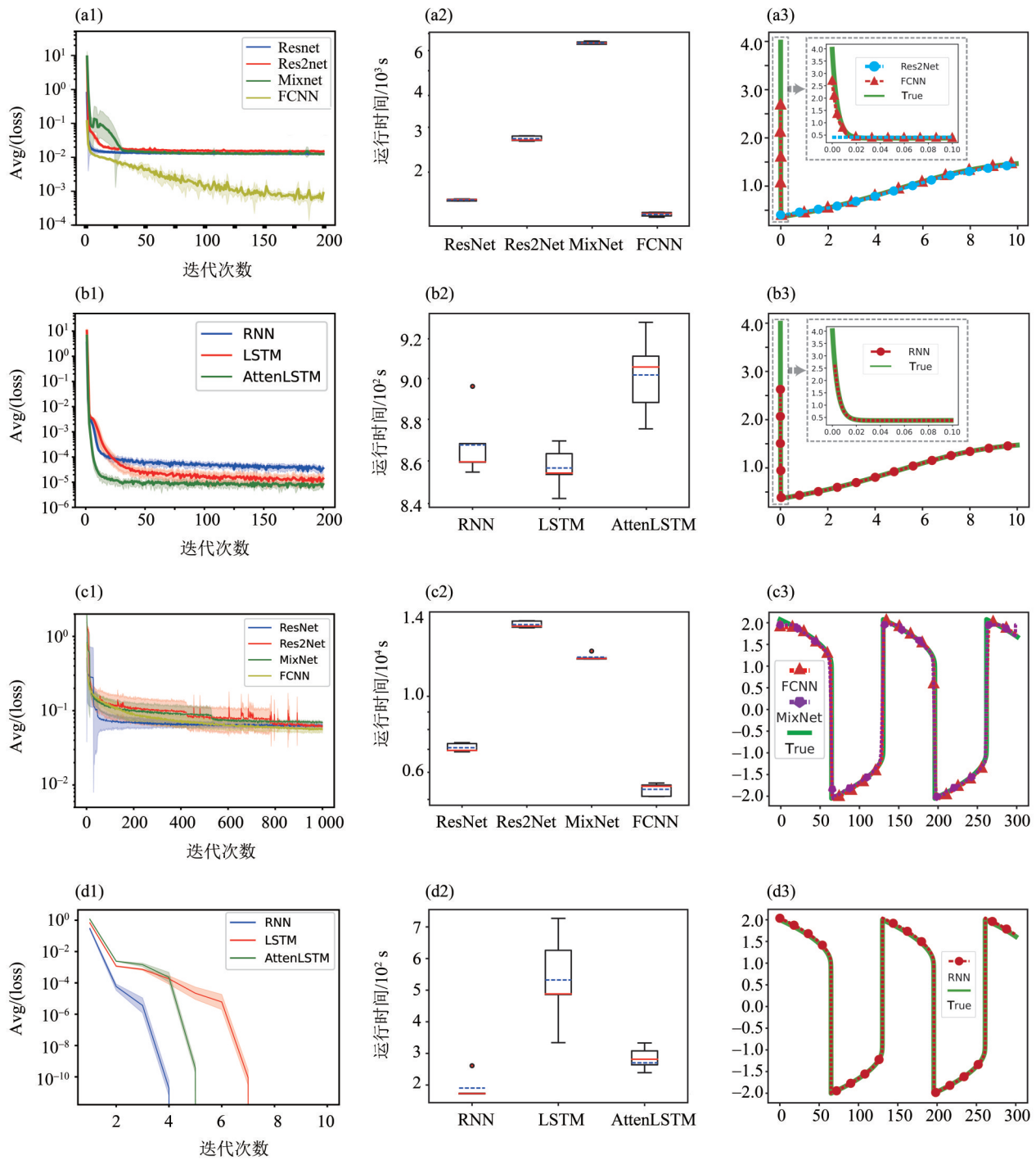
在表1和图2(a1)所展示的网络架构下, 4种深度神经网络的表现显示出了较大的差异性. 在训练结束时网络收敛到的损失函数值方面, 全连接神经网络FCNN收敛到了最低的 $\sim 10^{-3}$ 数量级, 优于其他3种含卷积层的模型.

通过观察4种网络每次运行时损失值的下降过程发现, 残差网络ResNet和残差网络的变种Res2Net在约第10次迭代时损失值即稳定在1附近, 随着迭代次数不断增加, loss值在0到1之间波动, 不再下降(不管学习率增大还是减小, 均为该情况). 混合卷积网络MixNet在前25个epoch的误差限较宽, 训练后期的损失函数值变化情况与残差类网络相似, 说明参数的随机初始化对混合卷积网络MixNet训练前期的影响较为明显. 与之相反, 全连接神经网络FCNN前期误差限很窄, 后期呈现波动下降的趋势. 综合来看, 在逼近效果方面, 全连接神经网络FCNN最好, 其他3种网络的指标相近, 但残差网络ResNet比另两种网络略好.

从箱线图2(a2)来看, 残差网络的变种Res2Net、混合卷积网络MixNet和全连接网络FCNN的训练时间分别约为残差网络ResNet的1.74倍、4.13倍和0.88倍. 4种网络所对应的箱体均很窄, 说明参数随机初始化几乎不影响训练时长的稳定性.

从图2(a3)来看, FCNN相对于残差网络的变种Res2Net能学到最初快变化的部分, 从这里反应了FCNN更优的原因.

综合来看, 对于此例全连接神经网络FCNN的训练耗时最短、同时精度最高. 其他3种神经网络精度相近, 混合卷积网络MixNet训练时间最长, 实际应用中若非其能达到较高精度, 则需要慎重使用.



(a,b) 图对应于B-Z反应的结果, (c,d)图对应于VdP方程的结果. (a1~d1)中实线是损失函数值的平均变化曲线, 上下的填充区域代表10次训练结果的误差带( $\mu - \text{std}$ ,  $\mu + \text{std}$ ), 其中 $\mu$ 表示损失值的均值,  $\text{std}$ 表示损失值的标准差

图2 训练损失演化、运行时间箱线图 and 数值解对比图

Fig. 2 Diagrams for the training loss function, running time and numerical solutions of various neural networks

### 2.2 VdP 方程上的非时序神经网络

训练设置如下: 按照批量大小为300的标准划分训练批次, 输入为 $t$ , 输出为 $x(t)$ . 每个模型重复训练5次, 每次1000个迭代轮次. 损失函数选择均方误差(MSE)损失, 优化器选择Adam优化器. 在训练过程中, 对每个批次计算损失, 然后进行反向传播计算梯度, 并更新网络参数. 将每次迭代中所有批次(共800个)的损失加总并保存, 每训练10个迭代轮次, 输出对应的损失. 在本小节中, 我们将继续比较残差网络

ResNet、残差网络的变种 Res2Net、混合卷积 MixNet 和全连接网络 FCNN 的性能. 网络的主要结构设置与 B-Z 反应算例相似, 模型参数及运行结果如表 1 所示.

这里最后得到整个训练集上的损失函数收敛到  $10^{-2}$  量级, 且四种非时序网络的收敛情况无明显差异. 从图 2(c1) 可以看到, 除了残差网络 ResNet 在训练初期误差限分布较广外, 几种网络的损失下降路径都比较一致.

几种模型所需的训练时间有明显差别, 全连接网络训练速度最快, 残差网络的变种训练速度最慢, Res2Net、MixNet 和 FCNN 所用时间分别为 ResNet 的约 2.3 倍、1.8 倍和 0.8 倍. 观察图 2(c2) 运行时间, 整体训练时间的分布较为平均, 但混合卷积网络 MixNet 的箱线图出现了异常点, 略有不稳定.

### 2.3 B-Z 反应上的时序神经网络

本节将比较 3 种时序神经网络模型: 循环神经网络 RNN、长短时记忆网络 LSTM 以及含有自注意力机制 Self-Attention 的 LSTM 模型(后文简记为 AttenLSTM 模型)在相似设置下对 B-Z 算例数据集的学习效果.

网络设置: 循环神经网络 RNN 的设置包含两个 RNN 层, 而长短时记忆网络 LSTM 则是一个 LSTM 层加上一个全连接层. 带注意力机制的 LSTM 模型 AttenLSTM 中, 数据在进入 LSTM 层后, 会通过一个包含线性变换、激活函数 Tanh 和 Softmax 的注意力组件, 最后经由全连接层输出.

训练设置: 与 B-Z 反应上的非时序网络相同, 优化器和损失函数分别为 Adam 优化器和均方误差 (MSE) 损失, 迭代 200 次, 重复运行 5 轮. 具体的模型参数与训练指标见表 2. 在相同的层数、隐藏层大小设置下, 将循环神经网络 RNN 中的原始神经元换成 LSTM 单元后, 模型参数量变为原来的约 4 倍, 以注意力模块替代第 2 层 LSTM 层得到的 AttenLSTM 模型的参数量则会降为原 LSTM 模型的 40% 左右.

表 2 时序深度网络在 B-Z 反应和 VdP 方程上的参数设置和训练结果汇总

Table 2 Summary on parameter settings and training results of temporal neural networks for the B-Z reaction and VdP equation

项目	B-Z 反应				VdP 方程			
	RNN	LSTM	AttenLSTM	ODE-Net	RNN	LSTM	AttenLSTM	ODE-Net
参数量/ $10^4$	1.29	5.11	2.21	0.46	1.27	5.05	2.14	0.45
学习率/ $10^{-5}$	5.0	2.0	5.0	5.0	0.5	0.5	0.5	0.5
平均损失	$3.56 \times 10^{-5}$	$1.39 \times 10^{-5}$	$8.22 \times 10^{-6}$	$2.87 \times 10^{-9}$	$5.30 \times 10^{-10}$	$5.49 \times 10^{-10}$	$7.12 \times 10^{-10}$	$1.09 \times 10^{-10}$
平时时间/s	867.6	856.7	901.8	8.78	190.4	532.2	288.8	-

-表示由于运行较快导致时间无法准确测出.

从表 2 可以发现 RNN 的精度最差, LSTM 相比于 RNN 在训练时间减少的情况下, 精度反而更高, 而 AttenLSTM 相对于 LSTM 在不明显增加训练时间的情况下, 有效提高了精度. 因此, AttenLSTM 能在较短的训练时间内达到相对较高的精度, 在此例子上是推荐使用的.

从图 2(b1) 损失函数的收敛情况来看, LSTM 相较于 RNN 在最初收敛时较慢, 但最终的损失会更低; 而带有 LSTM 模块和注意力机制模块的 AttenLSTM 能够收敛的更快且最终的损失更低. 这说明了 LSTM 模块和注意力机制模块对于求解刚性问题相比于普通的 RNN 有显著的优势.

从箱线图 2(b2) 来看, LSTM 模型对应的箱体窄且中位数线与平均数线接近, 说明该模型训练时间在三者之间最稳定. AttenLSTM 模型对应的箱体最宽, 中位数(红线)靠近下四分位线, 但没有出现异常值点, 说明其运行时间的分布范围较大但整体分布较为均匀. RNN 模型对应的箱体出现了一个异常点, 该异常点拉高了整体的均值, 说明参数的随机初始化可能会给 RNN 模型的训练效率造成一定影响. 从图 2(b3) 来看, 即使是最差的 RNN 也能学到最初快变化的部分, 这里也能体现时序网络的优势.

总的来说, 综合运行时间和回归精度, AttenLSTM 相比于 LSTM 和 RNN 在 B-Z 反应求解方面具有明显的优势.

## 2.4 VdP 方程上的时序神经网络

网络设置: RNN 和 LSTM 的设置方法类似, 都是以若干层 RNN/LSTM 层连全连接层. 注意力机制实现如下: 线性层对输入作线性变换、经过激活函数 Tanh 后与可学习参数相乘、经过 Softmax 函数得到注意力权重.

训练设置: 优化器和损失函数分别用的 Adam 和 MSE. 以前 20 个  $x(t)$  作为输入, 以当前时刻的  $x(t)$  作为标签. 重复训练 5 次. 为了避免过拟合, 当批次的损失值小于  $10^{-6}$  后会停止本次训练. 运行后发现最多在 7 个迭代轮次就能跳出, 最终将迭代轮次设为 10. 其他设置及训练结果见表 2.

从表 2 可以看到, 以上模型在约 5 次迭代后损失都能降到  $10^{-10}$  数量级, 因此能较快达到阈值跳出训练, 计算用时都在 10 min 内. 总体而言, 在此例子上, 3 种时序神经网络的运行时间和精度都比较接近, 都能够达到令人满意的求解效果.

## 2.5 ODE-Net 求解刚性问题

神经常微分方程(Neural-ODE)是在 2018 年的 NIPS 会议上被提出的(Chen et al., 2018), 对于残差块  $z_{i+1} = z_i + f(z_i, \theta)$ , 如果将网络看作是连续的, 残差项  $z_{i+1} - z_i$  将接近于隐藏层对  $t$  的导数. 也即可以用神经网络对隐层状态的变化进行模拟, 之后使用常微分方程求解器(ODE Solver)来积分. 该方法极大地推动了神经网络与微分方程相结合的研究, 并为时序数据建模提供了新的思路.

具体而言, 假设某个未知过程的演化遵循如下一般常微分方程组:

$$\frac{dz}{dt} = f(z(t), t). \quad (3)$$

观察到该过程轨迹上的一些带有噪声的数据点  $\{(z_0, t_0), (z_1, t_1), \dots, (z_M, t_M)\}$ , 我们希望找到参数化的动力学函数  $\hat{f}(z, t; \theta)$  (如系数为  $\theta$  的多项式展开) 作为真实动力学函数  $f(z, t)$  在某种意义下的最佳近似.

给定轨迹上任意两个相邻的观测值点  $(z_0, t_0), (z_1, t_1)$ , 我们基于参数化的动力学函数对系统从初始值  $(z_0, t_0)$  进行由  $t_0$  到  $t_1$  的演化(比如常微分方程求解器), 目标是 minimized 更新状态  $(\hat{z}_1, t_1)$  与观测值  $z_1$  的差异:

$$L(\hat{z}_1 - z_1) = L\left(\int_{t_0}^{t_1} \hat{f}(z(t), t; \theta) dt + z_0 - z_1\right) = L\left(\text{ODESolver}(z_0, t_0, t_1, \hat{f}(z(t), t; \theta)) + z_0 - z_1\right). \quad (4)$$

为了优化损失函数  $L$ , 需要计算其关于初值  $z_0$  和参数  $\theta$  的梯度, 具体分析可参考 Chen et al.(2018) 或示意蛋白的生长模型等. ODE-Net 展现出其在处理非均匀时间序列数据和刚性问题数据方面的优良性能, 并具备了较强的抗噪能力. 在此作为对比, 我们将 ODE-Net 同样使用在刚性问题 B-Z 反应和 VdP 方程.

从表 2 不难看到, 对于 B-Z 反应, ODE-Net 在复杂性远低于其他 3 种时序网络的情况下, 损失函数能达到  $10^{-9}$  量级, 远远优于其他时序或非时序网络, 并且运行时间也是最快的. 对于 VdP 方程, ODE-Net 的损失函数能达到  $10^{-10}$  量级, 并且运行时间几乎不能被检测到. 这说明 ODE-Net 在处理刚性问题时能在花费极少时间下取得极高的精度.

## 3 结 论

刚性问题在系统生物学研究中极其常见, 多种可被用于刚性方程(组)求解的神经网络方法被提出, 但目前缺乏这些神经网络之间的细致对比, 也没有在应用于刚性问题时普遍适用的选择准则. 针对上述问题, 本文基于 B-Z 反应和 VdP 方程两个经典算例, 系统测试并对比了多种主流非时序和时序神经网络(ResNet、Res2Net、MixNet、FCNN 和 RNN、LSTM、AttenLSTM)的求解效果和运行时间.

总体而言, 在本文测试例子上, 无论是训练时间还是在训练数据集上的精度, 时序神经网络的表现都远远优于非时序神经网络, 因此它们更适用于刚性问题的求解. 而对于 4 类非时序神经网络: ResNet、Res2Net、MixNet 和 FCNN, 它们的表现之间并无显著差异.

如果是关注于基于数据的刚性方程模型的导出, 那么还可以利用神经常微分方程, 如 ODE-Net 进行求解. 在本文所测试的两个例子中, ODE-Net 都可以达到非常令人满意的计算精度, 并且计算所需时间也远远小于各类非时序或时序神经网络.

总之, 本研究中我们对多种主流的神经网络在刚性方程求解性能方面进行了较为全面细致的对比. 我们的结果不仅可以为实际应用中神经网络架构的选择提供依据, 而且为进一步优化已有神经网络结构, 设计适用于系统生物学中各类刚性问题新的神经网络架构提供启发和参考.

### 参考文献:

- 曹学年, 2001. 刚性微分方程的并行 Rosenbrock 方法[D]. 绵阳: 中国工程物理研究院.
- 冯伯培, 1982. 解一般或刚性常微分方程初值问题的 Gear 方法[J]. 数值计算与计算机应用, 3(1): 12-23.
- 刘晓岑, 刘冬兵, 2012. 用加权平均方法构造新的隐式线性多步法公式[J]. 计算数学, 34(3): 309-316.
- 邱锡鹏, 2020. 神经网络与深度学习[M]. 北京: 机械工业出版社.
- 石滔, 2008. 求解刚性常微分方程的一类多步方法[D]. 武汉: 华中科技大学.
- 吴新元, 1999. 病态线性代数方程组的一种刚性问题数值解法[J]. 计算数学, 21(2): 157-162.
- 肖爱国, 李寿佛, 符鸿源, 等. 非线性刚性微分方程算法理论的发展[J]. 自然科学进展: 国家重点实验室通讯, 1999, 9(12): 1065-1072.
- 袁兆鼎, 费景高, 刘德贵, 1987. 刚性常微分方程初值问题的数值解法[M]. 北京: 科学出版社.
- 邹积麟, 钱稼茹, 2001. 病态方程的复合结构解法[J]. 清华大学学报: 自然科学版, 41(S1): 231-234.
- AHMADI DARYAKENARI N, De FLORIO M, SHUKLA K, et al, 2024. AI-Aristotle: A physics-informed framework for systems biology gray-box identification[J]. PLoS Comput Biol, 20(3): e1011916.
- ASGARI-TARGHI A, KLERMAN E B, 2019. Mathematical modeling of circadian rhythms[J]. Wiley Interdiscip Rev Syst Biol Med, 11(2): e1439.
- CHEN R T Q, RUBANOVA Y, BETTENCOURT J, et al, 2018. Neural ordinary differential equations [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems: 6572-6583.
- De FLORIO M, SCHIASSI E, FURFARO R, 2022. Physics-informed neural networks and functional interpolation for stiff chemical kinetics[J]. Chaos, 32(6): 063107.
- FABIANI G, GALARIS E, RUSSO L, et al, 2023. Parsimonious physics-informed random projection neural networks for initial value problems of ODEs and index-1 DAEs[J]. Chaos, 33(4): 043128.
- FIELD R J, NOYES R M, 1974. Oscillations in chemical systems. V. Quantitative explanation of band migration in the Belousov-Zhabotinskii reaction[J]. J Amer Chem Soc, 96: 2001-2006.
- FONSECA L L, SÁNCHEZ C, SANTOS H, et al, 2011. Complex coordination of multi-scale cellular responses to environmental stress[J]. Mol Biosyst, 7(3): 731-741.
- GAO S H, CHENG M M, ZHAO K, et al, 2021. Res2Net: A new multi-scale backbone architecture[J]. IEEE Trans Pattern Anal Mach Intell, 43(2): 652-662.
- GARDNER T S, CANTOR C R, COLLINS J J, 2000. Construction of a genetic toggle switch in Escherichia coli[J]. Nature, 403(6767): 339-342.
- GOODFELLOW I, BENGIO Y, COURVILLE A, 2016. Deep learning[M]. Vol 1, Cambridge: MIT Press.
- GYÖRGYI L, FIELD R J, 1992. A three-variable model of deterministic chaos in the Belousov-Zhabotinsky reaction[J]. Nature, 355: 808-810.
- HAIRER E, WANNER G, 1996. Solving ordinary differential equations II: Stiff and differential-algebraic problems [M]. Berlin: Springer.
- HE K, ZHANG X, REN S, et al, 2016. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): 770-778.
- HU P, YANG W, ZHU Y, et al, 2022. Revealing hidden dynamics from time-series data by ODENet[J]. J Comput Phys, 461: 111203.
- JI W, QIU W, SHI Z, et al, 2021. Stiff-PINN: Physics-informed neural network for stiff chemical kinetics[J]. J Phys Chem A, 125(36): 8098-8106.
- KIM S, JI W, DENG S, et al, 2021. Stiff neural ordinary differential equations[J]. Chaos, 31(9): 093122.
- LIU Z, CAI W, XU Z Q J, 2020. Multi-scale deep neural network (MscaleDNN) for solving Poisson-Boltzmann equation in com-

- plex domains[J]. *Commun Comput Phys*, 28(5): 1970–2001.
- RAISSI M, PERDIKARIS P, KARNIADAKIS G E, 2018. Multistep neural networks for data-driven discovery of nonlinear dynamical systems[EB/OL]. arXiv: 1801.01236. <http://arxiv.org/abs/1801.01236>.
- SCHMIDHUBER J, 2015. Deep learning in neural networks: An overview[J]. *Neural Netw*, 61: 85–117.
- SNEYD J, HAN J M, WANG L, et al, 2017. On the dynamical structure of calcium oscillations[J]. *Proc Natl Acad Sci*, 114(7): 1456–1461.
- TAN M, LE Q V, 2019. MixConv: Mixed depthwise convolutional kernels[EB/OL]. arXiv: 1907.09595. <http://arxiv.org/abs/1907.09595>.
- THOMAS P, STRAUBE A V, GRIMA R, 2012. The slow-scale linear noise approximation: An accurate, reduced stochastic description of biochemical networks under timescale separation conditions[J]. *BMC Syst Biol*, 6: 39.
- TROY W C, FIELD R J, 1977. The amplification before decay of perturbations around stable states in a model of the Zhabotinskii reaction[J]. *SIAM J Appl Math*, 32(2): 306–322.
- TYSON J J, FIFE P C, 1980. Target patterns in a realistic model of the Belousov-Zhabotinskii reaction[J]. *J Chem Phys*, 73(5): 2224–2237.
- Van der POL B, 1926. On relaxation-oscillations[J]. *Lond Edinb Dublin Philos Mag J Sci*, 2(11): 978–992.
- YANG W, PENG L, ZHU Y, et al, 2020. When machine learning meets multiscale modeling in chemical reactions[J]. *J Chem Phys*, 153(9): 094117.
- YAZDANI A, LU L, RAISSI M, et al, 2020. Systems biology informed deep learning for inferring parameters and hidden dynamics[J]. *PLoS Comput Biol*, 16(11): e1007575.
- ZHANG A, LIPTON Z C, LI M, et al, 2021. Dive into deep learning[EB/OL]. arXiv: 2106.11342. <http://arxiv.org/abs/2106.11342>.

(责任编辑 冯兆永)