

doi: 10.11835/j.issn.1000-582X.2025.10.008

引用格式:伊廷婧汶, 黄才生, 覃勇, 等. 基于北斗监测数据的高陡边坡变形 Transformer-CNN 预测模型[J]. 重庆大学学报, 2025,48(10): 81-94.



# 基于北斗监测数据的高陡边坡变形 Transformer-CNN 预测模型

伊廷婧汶<sup>1</sup>, 黄才生<sup>2</sup>, 覃勇<sup>2</sup>, 宋治江<sup>3</sup>, 贺小含<sup>2</sup>, 桂镜骑<sup>2</sup>, 王楷<sup>1</sup>

(1. 重庆大学 自动化学院, 重庆 400044; 2. 重庆市西部水资源开发有限公司, 重庆 400000;

3. 四川美术学院 设计学院, 重庆 401331)

**摘要:** 在大型工程的建设过程中常常会出现高陡边坡, 其变形往往会导致坍塌、滑坡等地质灾害, 给人们的生命财产安全带来危害, 高效采集位移数据并建立合适的预测模型对高陡边坡的变形进行准确预测对保障工程的顺利实施和人民的生命财产安全至关重要。研究通过在 Transformer 模型中融合 CNN (convolutional neural networks) 卷积层和 residual 残差结构的方法构建 Transformer-CNN 混合模型, 结合重庆某大型水利工程的项目背景, 使用北斗卫星监测系统的采样数据集, 通过对结果分析, 发现 Transformer-CNN 模型的 MAE (mean absolute error)、MSE (mean square error)、RMSE (root mean square error) 的值较单一模型相比有所降低, 且预测曲线和真实曲线的拟合程度较好, 证明 Transformer-CNN 模型在提高预测精度上的有效性, 为其他工程实施过程中对高陡边坡变形的预测分析提供可行性方案。

**关键词:** Transformer-CNN; 北斗数据集; 时间序列; 位移预测; 高陡边坡变形

中图分类号: TP39

文献标志码: A

文章编号: 1000-582X(2025)10-081-14

## Transformer-CNN prediction model of high and steep slope deformation based on Beidou detection data

YI Tingjingwen<sup>1</sup>, HUANG Caisheng<sup>2</sup>, QIN Yong<sup>2</sup>, SONG Zhijiang<sup>3</sup>,  
HE Xiaohan<sup>2</sup>, GUI Jingqi<sup>2</sup>, WANG Kai<sup>1</sup>

(1. College of Automation, Chongqing University, Chongqing 400044, P. R. China; 2. Chongqing West Water Resources Development Company Limited, Chongqing 400000, P. R. China; 3. College of Design, Sichuan Fine Arts Institute, Chongqing 401331, P. R. China)

**Abstract:** High and steep slopes are common during the construction of large-scale projects, and their deformation often leads to geological hazards, posing significant threats to life and property. Efficiently collecting

收稿日期: 2024-09-11

基金项目: 重庆市水利科技项目(CQSLK-2023028); 重庆市教育委员会科学技术研究计划(KJZD-K202500303)。

Supported by Chongqing Water Conservancy Science and Technology Project (CQSLK-2023028) and Municipal Education Commission Science and Technology Research Plan(KJZD-K202500303).

作者简介: 伊廷婧汶(2004—), 女, 主要从事机器学习、深度学习及多模态学习等人工智能方向的研究, (E-mail) 20223996@stu.cqu.edu.cn。

通信作者: 黄才生, 男, 高级工程师, (E-mail) 523739915@qq.com。

displacement data and developing an accurate predictive model are therefore essential. This study proposes a Transformer-CNN hybrid model that integrates convolutional layers and residual structures into the Transformer architecture. The optimized model is applied to displacement data obtained from the Beidou satellite system in a large water conservancy project in Chongqing. Experimental results indicate that the Transformer-CNN model achieves lower MAE, MSE, and RMSE values compared to single-model approaches, demonstrating superior prediction accuracy. These findings suggest that the proposed model offers a practical solution for predicting and analyzing slope deformation in similar engineering projects.

**Keywords:** Transformer-CNN; Beidou dataset; time series; displacement prediction; slope deformation

在大型水利工程建设过程中,高陡边坡由于受力特点极其复杂,在外部施工等环境因素的影响下,易发生坍塌、滑坡等严重地质灾害,给现场施工带来不利影响,直接威胁人员的生命和财产安全。边坡位移预测也由于其重要的经济价值和社会意义,成为国内外众多学者的热点课题<sup>[1-3]</sup>。

传统进行高陡边坡变形监测的方法很多,国内有:宏观地质监测法、大地精密测量法、近景摄影测量法<sup>[4]</sup>、工程地质分析法、极限平衡分析法、数值分析法等<sup>[5]</sup>;国外主要有:时域反射法<sup>[6]</sup>、三维激光扫描技术等<sup>[7]</sup>。上述方法只能进行单一点的监测,无法得到其他部分的位移,难以反映形体变形的全貌,对时间域和空间域的分辨能力较差<sup>[8]</sup>。在进行数据获取时,传统的人工测量方法是利用全站仪等仪器进行直接测量,具有测量精度高、数据可靠等优点,但人工采集需要专业的仪器和技术人员,耗时费力、成本较高,常常受到降雨、恶劣天气等环境因素的影响,采样频率较低且存在较大安全风险<sup>[9]</sup>。随着北斗全球组网的完成, GPS 测量系统被广泛应用到包括数据获取在内的各个领域<sup>[10]</sup>,能在复杂环境中保证定位和导航的精度,实现 24 h 无间断测量,不受自然条件影响,快速、客观、准确地获取高陡边坡各个测量点的位移数据,有效解决传统的人工测量法存在成本消耗大、安全性差的问题。

随着人工智能和深度学习算法的不断发展,越来越多用于时序处理的智能预测算法在边坡位移预测中得到应用,例如支持向量机回归算法、随机森林算法、梯度提升机算法、循环神经网络算法等<sup>[11]</sup>。其中 Transformer 算法因其并行计算能力强、能够捕捉长距离依赖关系、可扩展性强、泛化能力强等特点,在时序预测领域受到广大研究者欢迎<sup>[12]</sup>。然而,该算法的参数量度大、计算复杂度高、缺乏局部感知能力等缺点导致其对数据的需求相对较高,存在优化过程困难问题<sup>[13]</sup>。而卷积神经网络 CNN(convolutional neural networks)具有层次化结构,通过在不同位置共享参数,提高了计算效率,泛化能力较强,可以较好捕捉局部特征<sup>[14]</sup>。但受到局部感受野的限制和多层叠加的局限,存在难以捕捉长距离依赖关系、缺乏全局理解能力、可扩展性有限的缺点<sup>[15]</sup>。由于边坡位移的时间序列曲线存在复杂的非线性特征,上述传统预测模型精度不满足现行预测要求<sup>[16]</sup>。为进一步提升不同模型的预测性能和精度、弥补它们在高陡边坡变形预测中存在的缺陷,研究将 Transformer 和 CNN 模型进行有机结合,在 Transformer 模型中加入 CNN 卷积层建立 Transformer-CNN 模型,基于中国西部某大型水利工程进行验证。结果表明,Transformer-CNN 混合模型和单一的 CNN、Transformer 模型相比,MAE(mean absolute error)、MSE(mean square error)、RMSE(root mean square error) 的值均有所降低。说明在 Transformer 模型中加入卷积层后可有效提高数据预测的准确度,预测性能较好。该模型有望进一步应用到其他工程中,为高陡边坡位移的精确预测提供可行性方案。

## 1 基于北斗监测数据的 Transformer-CNN 预测模型架构设计

研究采用 Transformer-CNN 算法构建基于北斗数据集的高陡边坡位移预测模型。该模型以 Transformer 架构为基础,在解码器的每一个自注意力层中加入预先构建的 CNN 卷积层和 residual 残差结构。其中,CNN 卷积层由基本的因果卷积层、后续的归一化层和 ReLU(rectified linear unit)激活层构成的(如图 1 所示)。研究通过对数据进行端到端的多次迭代训练优化模型参数,寻找最佳配置。笔者基于北斗数据集进行模型验证并输出高陡边坡的位移预测结果和 MAE、MSE、RMSE 等量化指标的值。为进一步验证 Transformer-CNN 模型的准确性和可行性,依次去掉 Transformer 的注意力层和 CNN 的卷积层,将 Transformer-CNN 模型的预

测结果和单一 CNN 及 Transformer 模型的预测结果进行对比分析。研究思路如下:

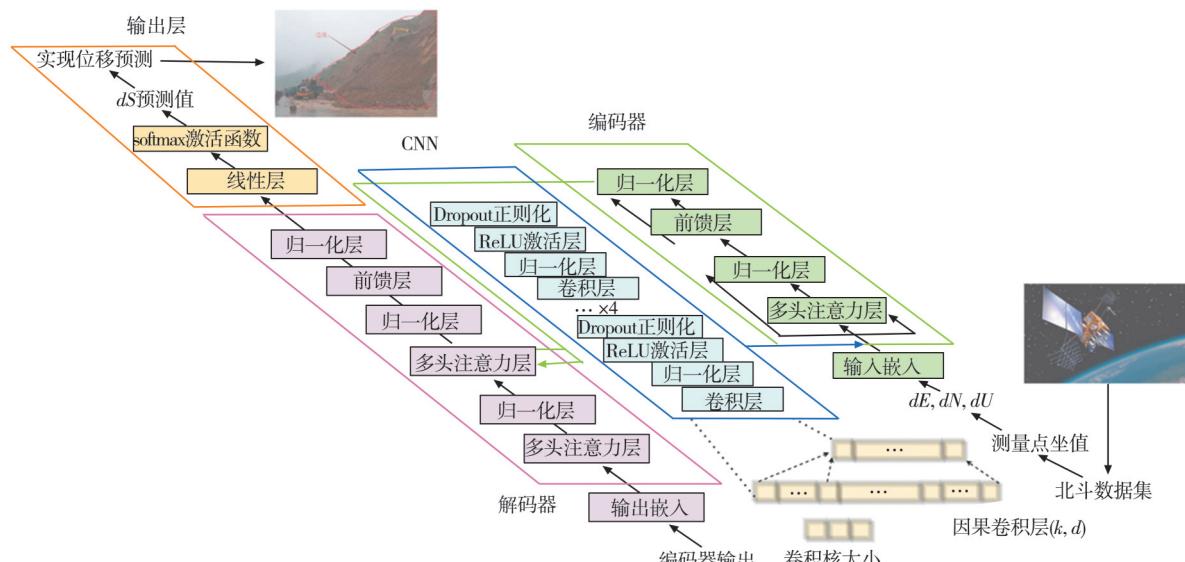


图 1 基于北斗监测数据的 Transformer-CNN 结构图

Fig.1 Structure diagram of Transformer-CNN based on Beidou monitoring data

- 1) 收集金刚沱北斗解算数据, 数据集中包括了北斗卫星采集的 700 多组数据和过去 3~4 年人工采集的数据。
- 2) 鉴于北斗测得的数据存在一些跳变和异常值, 不准确性和波动性较大, 首先对数据进行预处理, 剔除异常值并进行数据归一化。
- 3) 将 Transformer 的自注意力层、解码器层、编码器层和 CNN 的卷积层、residual 残差结构进行融合, 构建 Transformer-CNN 模型。
- 4) 选用 MAE、MSE、RMSE 3 个参数作为量化指标进行消融实验, 将 Transformer-CNN 模型预测高陡边坡变形的结果与其他单一模型进行对比。
- 5) 对比分析各误差评价指标的差异, 评估不同模型预测的有效性和准确性。

## 2 基于 Transformer 架构的高陡边坡变形预测模型

由于使用北斗卫星系统采集的数据存在精度不高、数据波动大的缺点, 存在较多的跳变和异常值, 且采样点的位移受到风速、温度、湿度等其他因素影响, 数据的误差较大。为了在识别异常点的同时进一步提高预测准确性, 采用 Transformer 模型对高陡边坡各个位移点建立预测模型, 快速捕捉数据的全局特征, 完成高陡边坡的位移预测任务。

由 Transformer 进行高陡边坡位移预测的原理可知, 该模型不依赖于递归结构, 通过注自注意力机制捕捉数据的全局特征, 由编码器、解码器和输出层组成, 具体结构如图 2 所示。

### 2.1 编码器

如图 3 所示, 北斗卫星系统采集的数据经由嵌入层后, 维度变为符合 Transformer-CNN 编码器输入的维度。输出数据经过多头注意力层, 该层将多组自注意力层的输出结果进行合并, 得到计算后的输出结果。自注意力层和多头注意力层的结构如图 4 所示。

输入矩阵  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ , 在北斗数据集中,  $\mathbf{X}_1 = dE, \mathbf{X}_2 = dN, \mathbf{X}_3 = dU$ 。其中:  $dE$  表示  $E$  方向的位移差值;  $dN$  表示  $N$  方向的位移差值;  $dU$  表示  $U$  方向的位移差值。 $\mathbf{X}$  与可学习的权重矩阵  $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$  相乘产生查询、键、值矩阵, 通过线性变换得到  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  3 个向量的初始表示

$$\mathbf{Q} = \mathbf{W}_Q \mathbf{X}, \quad (1)$$

$$\mathbf{K} = \mathbf{W}_K \mathbf{X}, \quad (2)$$

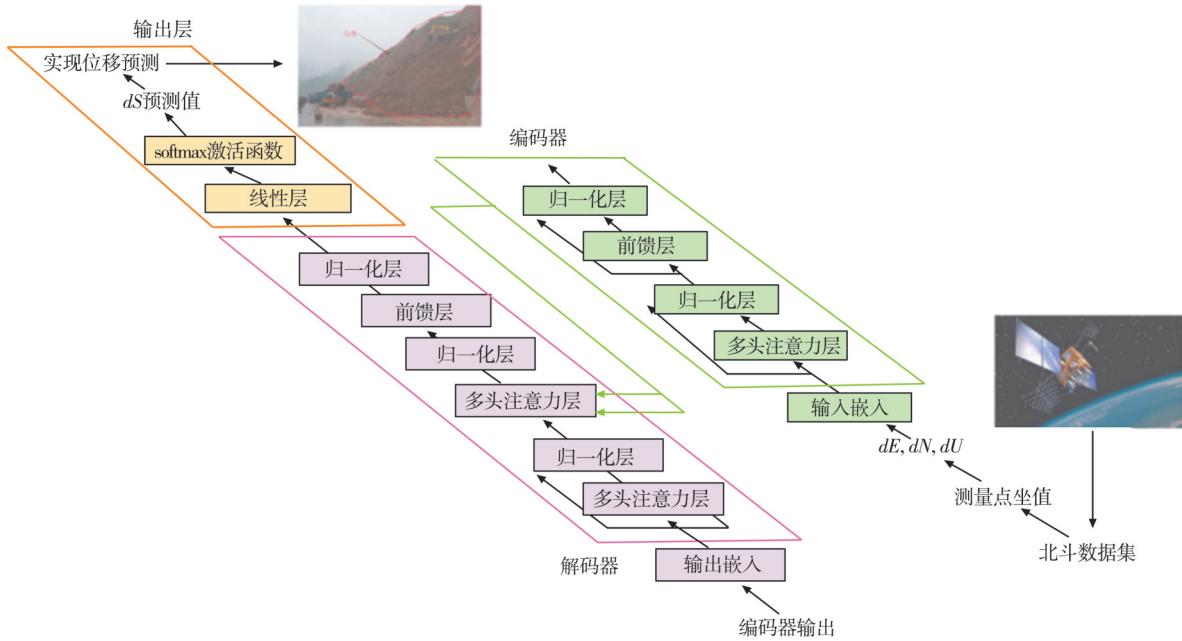


图2 基于北斗监测数据的Transformer结构图

Fig.2 Structure diagram of Transformer based on Beidou monitoring data

$$V = W_V X. \quad (3)$$

通过缩放因子  $\frac{1}{\sqrt{d_k}}$  进行缩放, 其中,  $d_k$  是键向量的维度, 在上述模型中  $d_k$  的值为 3, 应用 softmax 函数来

获取注意力权重。具体计算公式为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}, \quad (4)$$

多头注意力机制将上述过程应用于  $h$  个不同的头, 对于每个头  $h$ , 使用不同的权重矩阵  $\mathbf{W}_h^Q, \mathbf{W}_h^K, \mathbf{W}_h^V$ , 分别与查询、键和值的点积, 具体公式为

$$\text{head}_h = \text{Attention}(\mathbf{Q}\mathbf{W}_h^Q, \mathbf{K}\mathbf{W}_h^K, \mathbf{V}\mathbf{W}_h^V), \quad i=1, \dots, 8. \quad (5)$$

通过前馈层矩阵把计算得到的多个矩阵拼接在一起, 与附加的权重矩阵  $\mathbf{W}^0$  相乘得到最终的输出结果

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)\mathbf{W}^0. \quad (6)$$

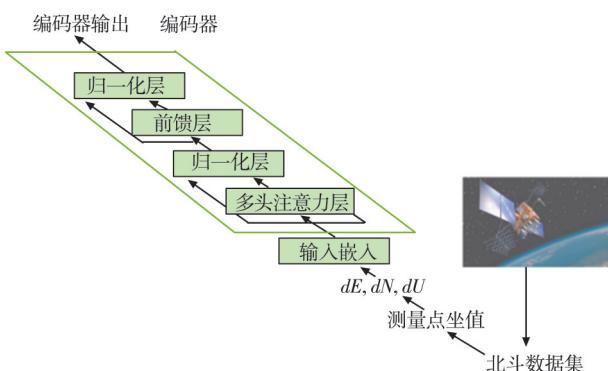


图3 编码器结构

Fig.3 Encoder structure

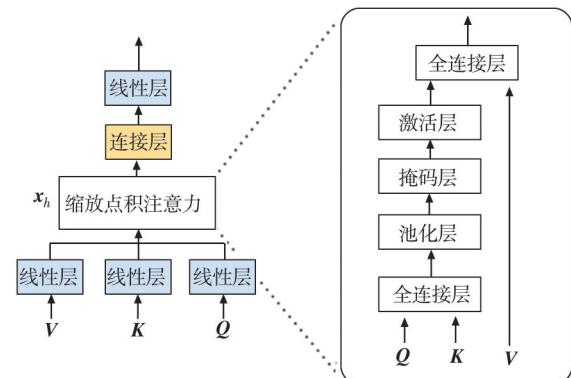


图4 注意力机制

Fig.4 Attention mechanism

归一化层(Add&normalize)是由残差连接 Add 和归一化 Normalize 2部分组成。

1) 残差连接 Add 是指在多头注意力输出数据的基础上加残差块  $X$ , 防止在深度神经网络的训练过程中发生退化, 用于解决多层网络训练问题, 让网络只关注当前差异的部分。

归一化层应用在残差连接之后, 其计算公式如下

自注意力层:

$$\text{LayerNorm}(\mathbf{X} + \text{MultiHeadAttention}(\mathbf{X})), \quad (7)$$

前馈层:

$$\text{LayerNorm}(\mathbf{X} + \text{FeedForward}(\mathbf{X})), \quad (8)$$

式中:  $\mathbf{X}$  表示 Multi-HeadAttention 或者 FeedForward 的输入,  $\text{MultiHeadAttention}(\mathbf{X})$  和  $\text{FeedForward}(\mathbf{X})$  表示输出。

前馈层(FeedForward)是一个 2 层的神经网络, 先线性变换, 再经过 ReLU 非线性, 再次进行线性变换, 具体的计算公式为

$$\text{FFN}(\mathbf{X}) = \text{ReLU}(\mathbf{0}, \mathbf{X}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2. \quad (9)$$

## 2.2 解码器

解码器与编码器的结构类似, 如图 5 所示, 是由多个解码器层堆叠而成, 每一个解码器层均包含了注意力层, 通过内部的 FullAttention 类计算输入序列中各元素之间的注意力权重, 得到输入序列中不同元素对当前元素的重要性, 加权求和后得到最后的结果。

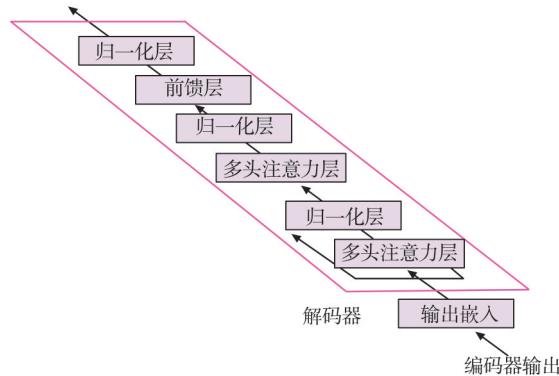


图 5 解码器结构

Fig.5 Decoder structure

与编码器最大的区别在于解码器还包含一个掩码多头注意力层, 该层可对某些值进行掩盖, 使其在参数更新时不产生效果。

## 2.3 输出层

解码器输出的数据经过激活函数层(softmax)和线性层(linear)得到最终的预测结果, 经解码器输出的数据首先经过一次线性变换, 该线性变换层是一个简单的全连接神经网络, 可以把解码组件产生的向量投射到比它大得多的向量里, 然后激活函数(softmax)层把该向量变为概率, 最终输出预测结果, 如图 6 所示。

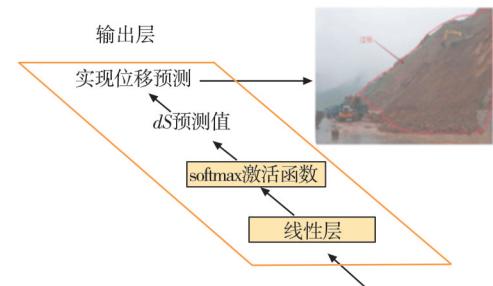


图 6 输出层结构

Fig.6 Output layer structure

## 3 基于 Transformer-CNN 的高陡边坡变形预测模型

北斗数据集经过预处理去掉无关列和较差的数据后, 成为包括  $E$ 、 $N$ 、 $U$ 、 $dE$ 、 $dN$ 、 $dU$  6 列以及采样时间构成的时间序列, 为了克服 Transformer 对数据局部特征的提取能力有限、对数据集依赖过大的缺点, 笔者提出用 CNN 优化 Transformer, 最终构建 Transformer-CNN 模型。

传统用于图像处理的 CNN 网络的卷积核大多是 2 维或 3 维的, 且需要进行多通道信息处理。故对 CNN 模型进行改进, 通过卷积核在时间轴上滑动, 提取时间序列的局部模式, 选用  $1 \times 3$  的卷积核(3 表示时间窗口

的长度)。而 Transformer 模型在小数据集上可能出现过拟合,故在 Transformer 的每一个自注意力层中加入预先构建的 CNN 卷积层和 residual 残差结构用于捕捉局部特征。通过这一预测网络实现数据从端到端的训练,对高陡边坡的位移坐标进行时间序列预测。

### 3.1 可用于一维时序数据处理的 CNN 卷积层

#### 3.1.1 时序预测原理

为了训练 CNN 网络进行预测,训练集由给定时间序列的等大小序列对(输入序列、目标序列)组成。目标序列是相对于各自的输入序列向前移动一定数量输出长度(outputlength)的序列,这意味着输入长度(inputlength)的目标序列包含各自输入序列尾部输入长度-输出长度(inputlength-outputlength)元素作为第 1 个元素,位于输入序列最后一个条目之后的输出长度(outputlength)元素作为它的最后一个元素,使用滑动窗口的方法,重叠的输入目标序列可以创建出一个时间序列,选用处理后数据集中的  $dE$ 、 $dN$ 、 $dU$  作为特征列,选用  $dS$  列作为目标列,如图 7 所示。

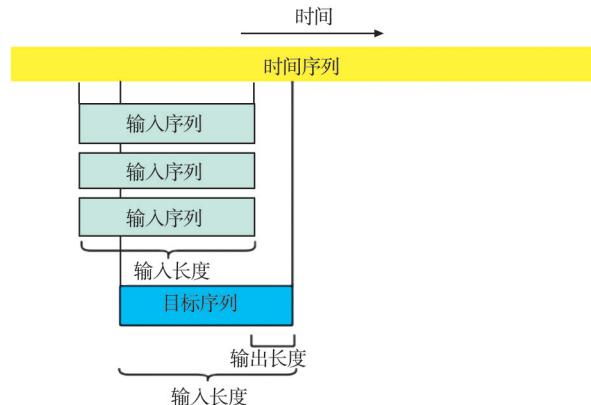


图 7 时序预测原理图

Fig.7 Schematic of time series forecasting

#### 3.1.2 对 CNN 模型进行处理

处理后的 CNN 模型由 4 个 residual block 残差块组成(如图 8 所示)。该残差块把模型由简单的一维因果卷积层改为由相同膨胀因子和残差的 2 层结构。图中  $k$  表示卷积核大小(kernel\_size),  $d$  表示膨胀基数(dilation\_base)( $k \geq d$ )。

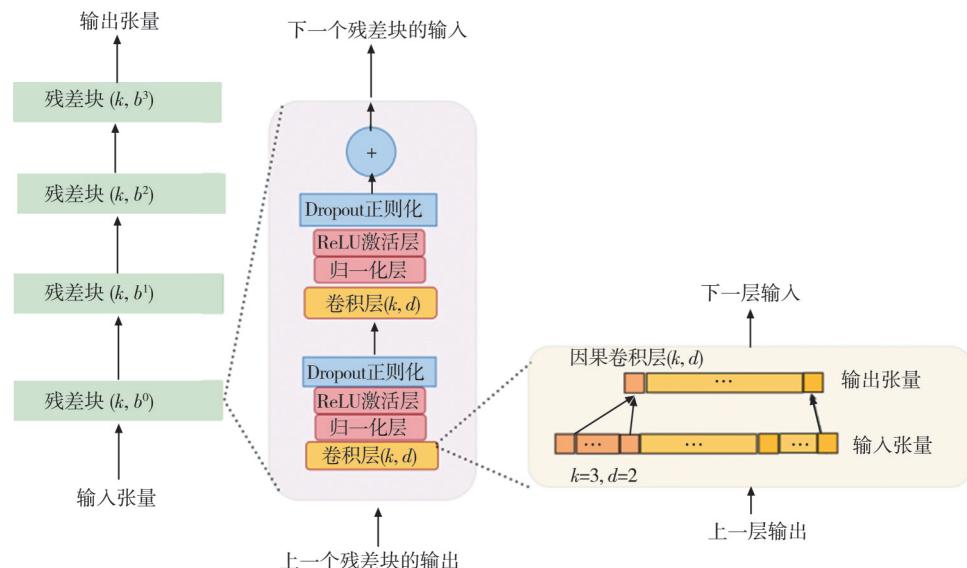


图 8 用于时序预测的 CNN 模型

Fig.8 CNN models for time series prediction

基本的因果卷积层(basic layer( $k, d$ ))原理示意图如图 9 所示。选用北斗数据集中的  $dU, dN, dE$  作为特征列, 选用  $dS$  列作为目标列。要计算输出的一个元素, 需要查看输入的一系列长度为卷积核大小(kernel size)的连续元素, 模型选用了一个大小为 3 的卷积核, 故单个卷积接收一个三维张量作为输入(shape(batchsize, inputlength, inputsize)), 也输出一个三维张量 shape(batchsize, inputlength, outputsize), 通过  $1 \times 3$  的卷积核在时间序列上移动, 获取一系列长度为卷积核大小(kernel\_size)的连续元素, 取该子序列与核向量点积, 得到输出的下一个元素。

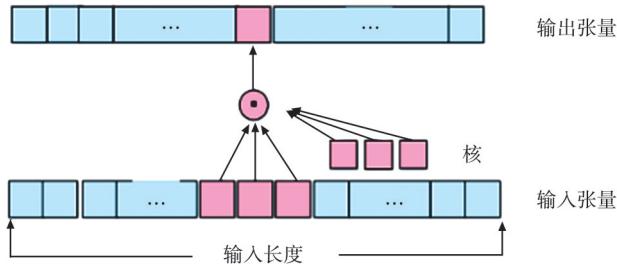


图 9 基本因果卷积层的原理图

Fig.9 Schematic of the basic causal convolutional layer

由于单一的卷积层的输出仅依赖于卷积核大小(kernel\_size)数目的输入, 为了扩大其依赖范围, 通过将多个层叠加在一起对模型进行扩张, 共使用 4 个 basic 层, 此时, CNN 卷积网络的接受野大小为 7, 其原理图如图 10 所示。

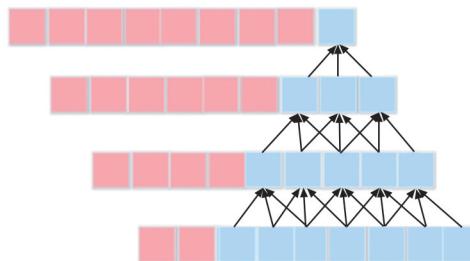


图 10 对模型进行扩张

Fig.10 Model expanding

为进一步提升模型性能, 在每一个卷积层中添加一个  $d=2, k=3$  的归一化层(weight norm), 将 CNN 的单一层转换为残差块。数据经基本卷积层输出后进入归一化层(weight norm), 通过重新参数化网络权重来提高训练的稳定性和速度, 有助于避免梯度消失和梯度爆炸问题。权重归一化的具体步骤为

1) 分解权重: 将每个层的权重矩阵  $W$  分解为 2 个部分: 一个缩放因子  $g$  和一个规范化的权重向量  $v$ 。这里的  $v$  是单位向量, 模为 1。

2) 参数化, 其具体公式为

$$W = g \cdot v \quad (10)$$

式中:  $g$  是权重矩阵中每个向量的缩放因子;  $v$  是对应的单位向量。

3) 训练: 在训练过程中, 只更新缩放因子  $g$  和向量  $v$ , 不直接更新原始的权重矩阵  $W$ 。

为了使构建的 CNN 不仅仅是一个过于复杂的线性回归模型, 在残差块中加入 ReLU 激活函数引入非线性, 数据经归一化层(Weight norm)输出后再输入到激活函数层(ReLU)中, 使网络能够学习更复杂的模式, 激活函数 ReLU 被添加到 2 个卷积层之后的残差块中。为了防止过拟合, 在每个剩余块中通过 dropout 引入正则化, dropout 在模型训练期间, 通过随机忽略一部分神经元, 增强模型的非线性能力和鲁棒性, 提升计算能力和速度, 原理如图 11 所示。

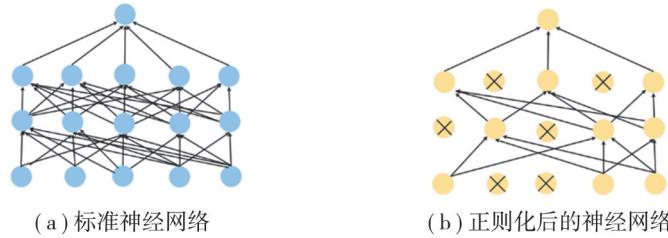


图 11 dropout 原理示意图  
Fig.11 Schematic diagram of dropout principle

最后,在卷积神经网络的末端加入全连接层,用于 CNN 提取特征,并进行回归分析,最终实现对高陡边坡位移的时间序列预测。

### 3.2 融合 CNN 网络结构的 Transformer 架构

将构建好的 CNN 卷积层和 residual 残差结构融入 Transformer 的编码器结构中。北斗卫星系统测得的位置点坐标经过预处理后首先进入 CNN 网络进行卷积操作,实现局部特征提取。输出数据再进入 Transformer 的多头注意力层和归一化层,经前馈层连接后进行归一化。输出的数据进一步进入解码器中进行解码,经线性层和 softmax 激活函数后,得到最终的输出值,实现数据端到端的训练,从而对高陡边坡的位移进行预测。如图 12 所示

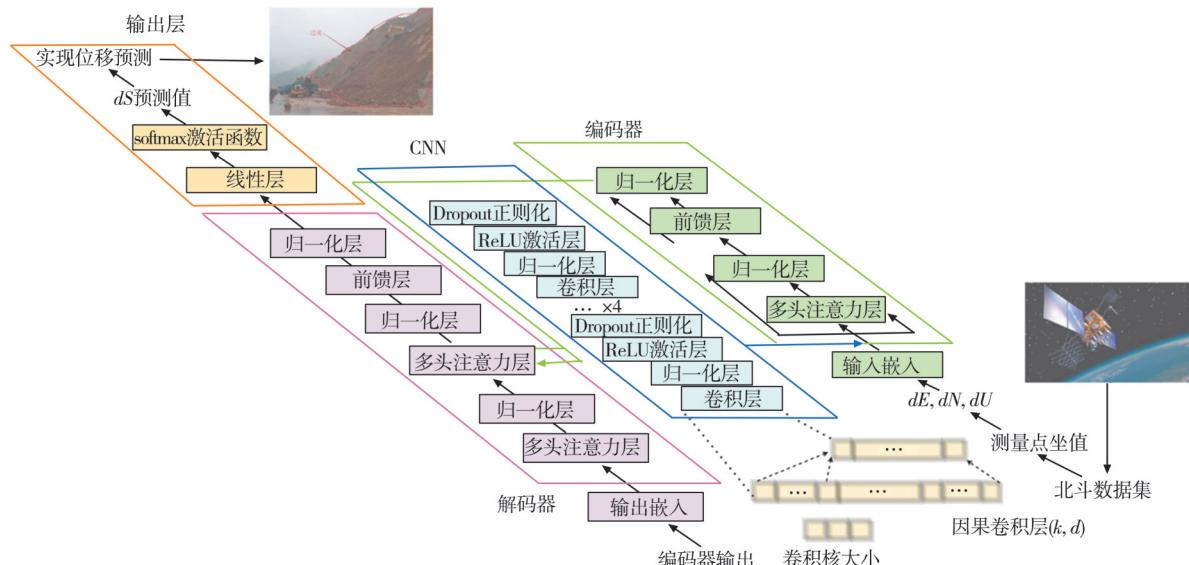


图 12 构建 Transformer-CNN 网络结构  
Fig.12 Construct the Transformer-CNN network structure

## 4 实验与分析

### 4.1 数据来源与数据处理

#### 4.1.1 实验数据来源

本实验使用的数据集来自中国西部某大型水利工程,其中包括了北斗卫星采集的 700 多组数据和人工采集的过去 3~4 年的数据。在过去,人们获取数据主要是通过全站仪进行人工采集,但由于受到人力成本和天气因素的影响,一般每 1~2 周采集一个数据,采样频率较低。采集了现场所有点位(总数超过 20)的  $E$ 、 $N$ 、 $U$  3 个方向的坐标值及  $E$  方向位移、 $N$  方向位移、 $U$  方向位移。

随着科技的不断进步,北斗卫星导航系统实现全球组网,中国大批的工程项目开始使用北斗卫星导航系统进行数据采集,由于其可以进行 24 h 不间断采样,故其采样频率为一天 4 次。以  $E$ 、 $N$ 、 $U$  3 个坐标值确

定一个点的位移,采集了现场共 16 个点的位置坐标(分为 3 个基准点和 13 个检测点)及其  $E$  方向位移、 $N$  方向位移、 $U$  方向位移,用于进一步预测每个点的变形量和变化趋势。但其数据的波动较大,存在一些跳变和异常值,数据的误差较大,部分数据如表 1 所示。

表 1 北斗结算部分数据(处理前)  
Table 1 Beidou solver data(before processing)

时间	解算状态	$E$	$N$	$U$	$dE$	$dN$	$dU$	历元数量	数据质量	可信度	卫星系统
2023-11-08 00 00:12:02	失败	0	0	0				3 046	差	GPS+BDS	
2023-11-08 01 01:08:10	失败	0	0	0				3 405	差	GPS+BDS	
2023-11-08 02 02:08:11	失败	0	0	0				3 704	差	GPS+BDS	
2023-11-08 03 03:08:11	失败	0	0	0				4 063	差	GPS+BDS	
2023-11-08 04 04:08:10	失败	0	0	0				4 423	差	GPS+BDS	
2023-11-08 05 05:08:11	失败	0	0	0				4 782	差	GPS+BDS	
2023-11-08 06 06:08:11	失败	0	0	0				5 141	差	GPS+BDS	
2023-11-08 07 07:08:11	失败	0	0	0				5 499	差	GPS+BDS	
2023-11-08 08 08:08:11	失败	0	0	0				5 858	差	GPS+BDS	
2023-11-08 09 09:08:11	失败	0	0	0				6 217	差	GPS+BDS	
2023-11-08 10 10:08:10	失败	0	0	0				6 576	差	GPS+BDS	
2023-11-09 00 00:14:13	失败	-3 094 379.8	10 714 590.01	6 172 438.85	-0.6	0.7	0	8 546	良	0.711	GPS+BDS
2023-11-10 00 00:14:43	失败	-3 094 379.8	10 714 590	6 172 438.85	0	0.2	-0.3	8 542	优	0.688	GPS+BDS
2023-11-11 00 00:14:41	失败	-3 094 379.8	10 714 590	6 172 438.85	-0.4	0.4	0.2	8 541	优	0.674	GPS+BDS

#### 4.1.2 实验数据预处理

1) 缺失值、异常值处理:由于所得数据中存在解算状态失败的样本以及数据质量较差、可信度较低的样本,故将此类数据删除,同时,将数据中含有 NaN 的行进行删除。

2) 删除无关量:原始数据集包括采样时间、解算状态、 $E$ 、 $N$ 、 $U$ 、 $E$  位移( $dE$ )、 $N$  位移( $dN$ )、 $U$  位移( $dU$ )、历元数量、数据质量、可信度和卫星系统共 11 列数据,但进行数据预测时只需要用到采样时间,以及把  $E$  均值、 $N$  均值、 $U$  均值、 $dE$ 、 $dN$ 、 $dU$  的数据作为目标变量,故对其他列进行删除。

3) 计算 2 个点之间的位移差值,得到  $dE$ 、 $dN$ 、 $dU$  列的数据值,通过  $dE$ 、 $dN$  的值计算得到  $dS$  列的数据值,其计算公式为

$$dS = \sqrt{(dE)^2 + (dN)^2}。 \quad (11)$$

4) 处理时间戳列:把采样的时间列进行分割,提取年、月、日、星期和小时等和时间相关的特征。

5) 数据正则化处理:为了防止过拟合和提高模型的泛化性能,使用 dropout 正则化,随机将网络中的部分神经元输出置 0,以减少神经元之间的依赖性。

6) 在 CNN 网络预测时,使用 MaxMinScaler 函数对数据进行归一化处理,把数据映射到  $[0,1]$  区间内,具体的公式如下

$$X_{std} = \frac{X - X_{min}(\text{axis}=0)}{X_{max}(\text{axis}=0) - X_{min}(\text{axis}=0)}, \quad (12)$$

$$X_{scaled} = X_{std} \cdot (\max - \min) \cdot \min。 \quad (13)$$

7) 划分训练集和测试集,取 80% 的数据作为训练集,剩下 20% 的数据作为测试集。

处理流程图如图 13 所示,处理后的部分数据集如表 2 所示。

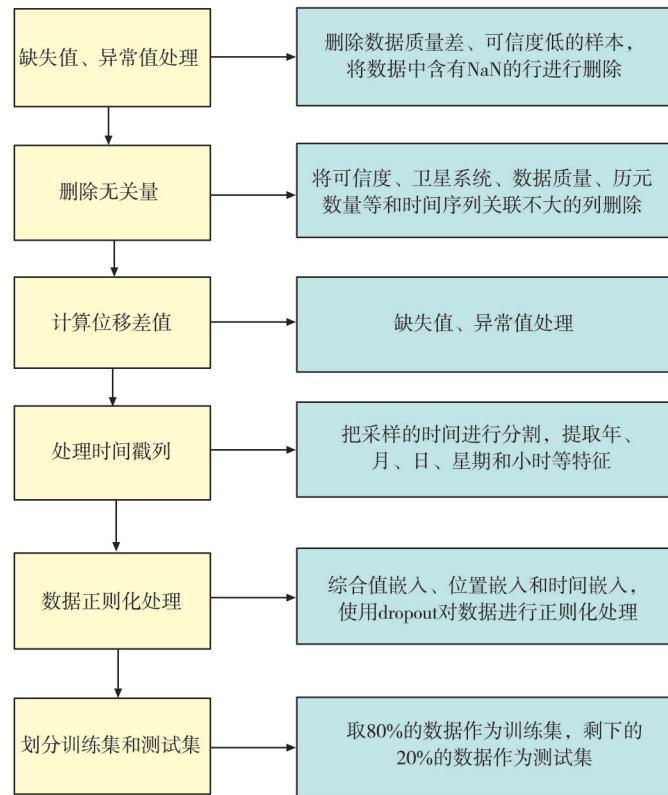


图 13 数据预处理流程图

Fig.13 Flow chart of data preprocessing

表 2 处理后的部分数据集

Table 2 Part of the processed dataset

E均值	N均值	U均值	dE	dN	dU	dS
529 148.34 8	4 937 735.199	334.771 672				
529 148.348 6	4 937 735.198	334.769 772	0.526	-1.192	-1.900	样本1
529 148.348 8	4 937 735.198	334.769 181	0.717	-1.599	-2.491	样本2
529 148.348 6	4 937 735.197	334.768 449	0.556	-1.983	-3.223	2.059 471 998
529.148.348 1	4 937 735.199	334.770 487	0.008	-0.574	-1.185	0.574 055 925
529 148.347 4	4 937 735.198	334.769 959	-0.654	-0.959	-1.713	1.160 774 750
529 148.347 7	4 937 735.198	334.768 884	-0.329	-1.096	-2.788	样本3
529 148.347 8	4 937 735.198	334.768 584	-0.284	-1.230	-3.088	1.144 315 061
529 148.347 5	4 937 735.197	334.766 281	-0.552	-2.282	-5.391	1.262 361 773
529 148.346 9	4 937 735.199	334.767 967	-1.105	-0.692	-3.705	2.347 813 596
529 148.347 0	4 937 735.198	334.765 42	-1.025	-1.696	-6.252	1.303 798 266
529 148.343 6	4 937 735.197	334.764 491	-4.418	-2.097	-7.181	1.983 230 394
529 148.344 6	4 937 735.197	334.764 749	-3.469	-1.943	-6.923	4.890 412 510
529 148.345 3	4 937 735.199	334.765 648	-2.733	-1.927	-6.024	3.976 080 007
529 148.345 1	4 937 735.198	334.765 14	-2.905	-1.274	-6.258	3.172 081 463
529 148.344 8	4 937 735.197	334.766 932	-3.252	-1.952	-4.740	3.792 862 722
529 148.345 0	4 937 735.198	334.762 302	-3.094	-0.979	-9.370	样本12
						3.245 193 160

## 4.2 实验与分析

### 4.2.1 实验参数设置

使用上述北斗数据集对神经网络进行训练,使其学习到输入与输出之间的映射关系。相关参数的设置如表3所示。

表 3 相关参数设置  
Table 3 Relevant parameter settings

相关参数	值
序列长度(seq_len)	64
标签长度(label_len)	64
预测长度(pred_len)	32
编码器输出尺寸(enc_in)	3
解码器输入尺寸(dec_in)	3
输出尺寸(c_out)	3
模型维度(d_moble)	512
注意力头数(n_heads)	8
编码器层数(e_layers)	2
解码器层数(d_layers)	2
前馈网络维度(d_ff)	2 048
滑动平均窗口的大小(moving_avg)	25
注意力因子	1
学习率	0.002
数据张量	32
训练批次大小(batch_size)	32
patience	3
编程语言	python

#### 4.2.2 评价指标

为了公平比较不同模型的表现, 同时保证评价的全面性和准确性, 综合多种评价指标的评价预测性能, 采用了如下多种评价指标:

1) 平均绝对误差(mean absolute error, MAE)<sup>[17]</sup>: 表示预测值和实际值之间的平均绝对误差, 计算公式为

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|_0 \quad (14)$$

2) 均方误差(mean square error, MSE)<sup>[18]</sup>: 量化了误差平方的平均值, 即预测值和实际值之间的差值, 可确保较大的误差对整体误差指标产生不成比例的高影响, 用于评估模型对异常值的敏感程度, 计算公式为

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (15)$$

3) 平均绝对百分比误差(root mean square error, RMSE)<sup>[19]</sup>: 预测值与实际值的误差平方根的均值, 提供与尺度无关的误差度量, 计算公式为

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (16)$$

#### 4.3 实验结果

为进一步定量分析混合模型对预测准确性的提升效果, 分析不同结构在混合模型中所起作用, 去掉 Transformer 模型中的 CNN 卷积层, 变为传统的单一的 Transformer 模型, 再去掉 Transformer 的自注意力机制, 只用改进后可进行时间序列预测的 CNN 模型进行时序预测。使用在 Transformer 基础上进行了改进的变体模型(包括 Reformer、Autoformer、Informer 模型)进行时间序列预测, 其中, Reformer 使用局部敏感哈希代替点积注意力, 采用可逆残差层减少存储需求<sup>[20]</sup>。Autoformer 则是在 Transformer 架构的基础上引入自相关机制和分解架构来处理长期序列预测问题<sup>[21]</sup>。Informer 更加关注减少计算复杂度和提高模型对长序列的学

习能力<sup>[22]</sup>。分别多次使用不同模型对北斗数据集进行时序预测,得到模型达到稳定时的量化指标如表4所示。不同算法的指标对比如图14所示,不同算法的预测曲线如图15所示。

表4 不同模型在训练达到稳定时的量化指标值

Table 4 Quantized index values of different models

算法	MAE	MSE	RMSE
Transformer	0.274	0.166	0.408
CNN	0.787	0.825	0.909
Transformer-CNN	0.245	0.148	0.385
Autoformer	0.434	0.420	0.294
Informer	0.294	0.207	0.455
Reformer	0.409	0.297	0.545

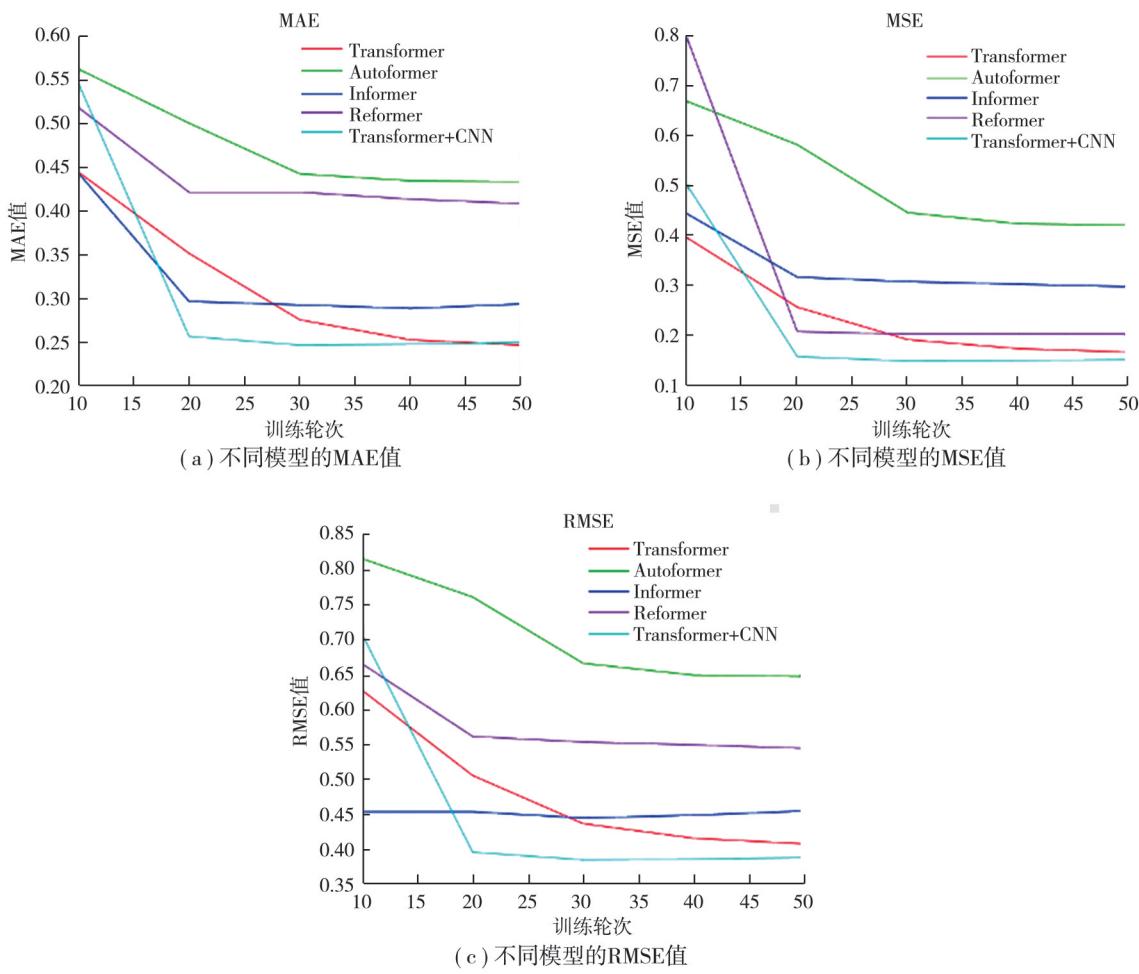


Fig.14 Comparision of parameters of different models

从图14中可以看出,Transformer-CNN模型在迭代20、30、40、50轮次时候MAE的值较单一的Transformer模型相比分别降低了12.88%、14.23%、13.88%、13.19%。MSE的值在迭代20、30、40、50轮次时分别降低了24.15%、26.73%、26.23%、25.24%。RMSE的值在20、30、40、50轮次时分别降低了12.96%、14.25%、14.03%、13.58%。由表4可以看出,Transformer-CNN模型在训练达到稳定时MAE的值比CNN低68.8%、比Transformer低10.58%;MSE的值比CNN低85.69%、比Transformer低10.84%;RMSE的值比CNN低57.64%、

比 Transformer 低 5.63%。由图 15 不同模型的预测曲线对比可以看出, Transformer-CNN 模型的预测结果和真实曲线拟合程度优于其他单一模型, 说明 Transformer-CNN 模型有效提高了预测的准确性。

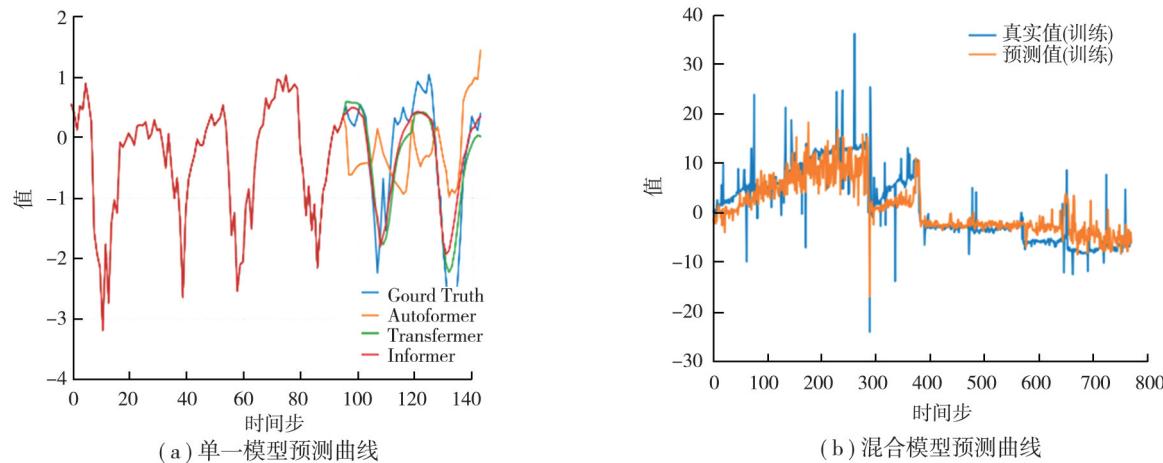


图 15 不同模型的预测曲线

Fig.15 Prediction curves for different models

同时, 在训练过程中发现, 加入 CNN 卷积网络后的 Transformer-CNN 模型在 30 轮后即可达到收敛状态, 而单一的 CNN 模型需要进行上百轮训练方可达到收敛状态, 说明加入卷积层在提高预测精度准确性的同时可节省训练时间和计算资源。

## 5 结语

研究通过在 Transformer 模型中加入 CNN 卷积层构建了 Transformer-CNN 模型用于高陡边坡变形的时间序列预测, 并基于中国西部某大型水利工程的北斗监测数据集进行模型训练和定量分析, 通过将 Transformer-CNN 模型的 MAE、MSE、RMSE 等参数量和 Transformer、CNN、Informer、Autoformer、Reformer 等单一模型进行对比, 发现 Transformer-CNN 模型的 MAE、MSE、RMSE 的值最优。说明在 Transformer 模型中加入卷积层构建的 Transformer-CNN 模型可以在利用 Transformer 时序特征提取优势的同时兼具 CNN 卷积网络提取局部特征的能力, 充分利用了 2 个单一模型的优点, 有效提高数据预测的准确度, 预测性能较好。进一步可以为高陡边坡变形的预测预报、坍塌预警等后续工作提供可靠的科学依据, 为其他工程实施过程中对高陡边坡变形的预测分析提供可行性方案, 为水资源配置工程领域的安全性和可持续性做出积极贡献。

## 参考文献

- [1] 柴章鹏, 黄德昌, 杨麒民, 等. 高陡边坡锚杆格构梁双排脚手架施工技术探究[J]. 价值工程, 2024, 43(27): 144-147.  
Chai Z P, Huang D C, Yang Q M, et al. Research on construction technology of double-row scaffolding in high steep slope anchor rod grillage beams[J]. Value Engineering, 2024, 43(27): 144-147. (in Chinese)
- [2] 李荟, 韩晓飞, 朱万成, 等. 基于多源信息融合的矿山边坡滑坡灾害研究现状与展望[J]. 工矿自动化, 2024, 50(6): 6-15.  
Li H, Han X F, Zhu W C, et al. Current status and prospects of research on landslide disasters in mine slopes based on multi-source information fusion[J]. Journal of Mine Automation, 2024, 50(6): 6-15. (in Chinese)
- [3] 刘礼福, 黄今, 雷国荣, 等. 爆破振动作用下某露天矿高陡边坡稳定性分析[J]. 有色金属(矿山部分), 2021, 73(4): 27-32.  
Liu L F, Huang J, Lei G R, et al. Stability analysis of high and steep slope of an open pit mine under blasting vibration[J]. Non-ferrous Metals (Mining Section), 2021, 73(4): 27-32. (in Chinese)
- [4] 杨一豪, 李学锋, 黄绪兴, 等. 某露天矿边坡稳定性及优化研究[J]. 价值工程, 2024, 43(16): 97-100.  
Yang Y H, Li X F, Huang X X, et al. Research on slope stability of a certain open-pit mine and optimization measures[J]. Value Engineering, 2024, 43(16): 97-100. (in Chinese)

- [5] 房智恒, 王李管, 彭南良, 等. 高陡边坡整体与局部失稳的强度折减及安全度判别分析[J]. 重庆大学学报, 2015, 38(6):8-14.  
Fang Z H, Wang L G, Peng N L, et al. Strength reduction and discriminant analysis of high steep slopes' overall and local instability[J]. Journal of Chongqing University, 2015, 38(6): 8-14. (in Chinese)
- [6] Li S Y, Li H, Ji W B, et al. Optical fiber sensing structure for slope monitoring based on brillouin optical time domain reflectometer[J]. Acta Photonica Sinica, 2021, 50(9): 107-115.
- [7] 陈云敏, 陈赟, 陈仁朋, 等. 滑坡监测 TDR 技术的试验研究[J]. 岩石力学与工程学报, 2004, 23(16): 2748-2755.  
Chen Y M, C Y, Chen R P, et al. Testing study on application of time domain reflectometry to slope monitoring[J]. Chinese Journal of Rock Mechanics and Engineering, 2004, 23(16): 2748-2755. (in Chinese)
- [8] 张占旭, 苏俊辉, 吕光祖, 等. 数字钻孔图像岩体结构面自动化识别方法[J]. 重庆大学学报, 2024, 47(2): 40-50.  
Zhang Z X, Su J H, Lv G Z, et al. Automatic identification of rock structure surface based on digital borehole images[J]. Journal of Chongqing University, 2024, 47(2): 40-50. (in Chinese)
- [9] 邱焕翔, 翟虎, 王冀鹏, 等. 基于无人机航摄的地质踏勘及风险评估[J]. 科学技术与工程, 2024, 24(18):7905-7912.  
Qiu H X, Zhai H, Wang J P, et al. Geological survey and risk assessment based on UAV photogrammetry[J]. Science, Technology and Engineering, 2024, 24(18): 7905-7912. (in Chinese)
- [10] 李烨. 浅析北斗卫星导航系统在测绘行业中的应用[J]. 经纬天地, 2024(3): 19-22.  
Li Y. A preliminary analysis for the application of the Beidou Satellite Navigation System in the surveying and mapping industry[J]. Survey World, 2024(3): 19-22. (in Chinese)
- [11] 刘晖, 朱德康, 许凌杰, 等. 基于改进灰狼算法优化支持向量机的边坡位移预测[J]. 自动化技术与应用, 2024, 43(9): 30-33.  
Liu H, Zhu D K, Xu L J, et al. Slope displacement prediction based on improved grey wolf algorithm and support vector machine[J]. Techniques of Automation and Applications, 2024, 43(9): 30-33. (in Chinese)
- [12] Liu C M, Shen Y, Xiao Q G, et al. PCSformer: pair-wise cross-scale sub-prototypes mining with CNN-transformers for weakly supervised semantic segmentation[J]. Neurocomputing, 2024, 593: 127834.
- [13] Meng W J, Liu S J, Wang H J. AFC-Unet: attention-fused full-scale CNN-transformer unet for medical image segmentation[J]. Biomedical Signal Processing and Control, 2025, 99: 106839.
- [14] 肖海平, 杨旺生, 肖岚, 等. 基于组合预测模型的露天矿高陡边坡滑坡变形研究[J]. 金属矿山, 2014, 43(4): 169-171.  
Xiao H P, Yang W S, Xiao L, et al. Research on landslide deformation of high and steep slope in open-pit mine based on combination prediction model[J]. Metal Mine, 2014, 43(4):169-171. (in Chinese)
- [15] Razavian S A, Azizpour H, Sullivan J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, OH, USA: CVPRW, 2014: 512-519.
- [16] 孙晓云, 段绰, 王明朋, 等. 基于 VMD 模型和 BSA-KELM 模型的高陡边坡位移预测模型研究[J]. 中国矿业, 2022, 31(2): 78-85.  
Sun X Y, Duan C, Wang M M, et al. Study on high and steep slope displacement prediction model based on VMD and BSA-KELM[J]. China Mining Magazine, 2022, 31(2): 78-85. (in Chinese)
- [17] Willmott C J, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance[J]. Climate Research, 2005, 30(1): 79-82.
- [18] Köksoy O. Multiresponse robust design: mean square error (MSE) criterion[J]. Applied Mathematics and Computation, 2006, 175(2): 1716-1729.
- [19] Hodson T O. Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not[J]. Geoscientific Model Development Discussions, 2022, 15(14): 5481-5487.
- [20] Saibabu B B, Kiran T. A novel method for predicting wind speed using data decomposition-based reformer model[J]. Earth Science Informatics, 2024, 17(1):227-249.
- [21] Niu M C, Li Y, Zhu J Y. Optical cable lifespan prediction method based on autoformer[J]. Applied Sciences, 2024, 14(14): 6286.
- [22] Zhang X, Yang K Y, Zheng L M. Transformer fault diagnosis method based on timesnet and informer[J]. Actuators, 2024, 13(2): 74.