

# 基于骨架数据的列车司机异常行为检测

朱高伟

兰州交通大学电子与信息工程学院, 甘肃 兰州

收稿日期: 2024年6月4日; 录用日期: 2024年7月8日; 发布日期: 2024年7月15日

## 摘要

针对列车司机异常行为检测时存在准确率以及效率低的问题, 在时空图卷积网络(Spatial Temporal Graph Convolutional Network, ST-GCN)行为检测模型的基础上, 提出一种基于骨架数据的列车司机异常行为检测(Spatiotemporal Graph Attention and Multi-Time Scale Temporal Convolutional Network, ST-GAT)模型。通过利用骨架数据进行建模, 并引入图注意力网络(Graph Attention Network, GAT)模块, 利用动态关注机制以提升ST-GCN模型对空间特征的提取能力。通过提出的多时间尺度时域卷积网络(Multi-Scale Temporal Convolutional Network, MS-TCN)模块进行不同时间尺度下的时序提取特征, 扩大时域卷积网络(Temporal Convolutional Network, TCN)模块感受野范围, 克服原始TCN模块不灵活、无法检测不同时间尺度的问题, 提高了模型的准确率, 同时使用交叉熵损失函数克服了模型训练过程收敛速度慢的问题。实验结果表明, 本文方法在测试集上比ST-GCN模型准确率提高8.8%, FLOPS提高2%。因此, 所提出的方法在提高列车司机异常行为检测的准确率和效率方面表现出较好的性能。

## 关键词

骨架数据, 列车司机, 异常行为检测, 图注意力网络, 多尺度时域卷积网络

# Detection of Abnormal Behavior of Train Drivers Based on Skeleton Data

Gaowei Zhu

School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou Gansu

Received: Jun. 4<sup>th</sup>, 2024; accepted: Jul. 8<sup>th</sup>, 2024; published: Jul. 15<sup>th</sup>, 2024

## Abstract

Aiming at the challenges of low detection accuracy and efficiency in train driver abnormal beha-

rior detection, this study proposes a model based on skeleton data, named Spatiotemporal Graph Attention and Multi-Time Scale Temporal Convolutional Network (ST-GAT), built upon the Spatial Temporal Graph Convolutional Network (ST-GCN) behavior detection model. By leveraging skeleton data for modeling, the model introduces a Graph Attention Network (GAT) module with a dynamic attention mechanism to enhance the ST-GCN model's extraction of spatial features. Additionally, the model incorporates a Multi-Scale Temporal Convolutional Network (MS-TCN) module to extract temporal features at different scales, addressing the flexibility and scale limitations of the Temporal Convolutional Network (TCN) module in ST-GCN. To improve training efficiency, the model adopts the cross-entropy loss function to expedite convergence. Experimental results demonstrate an 8.8% improvement in accuracy and a 2% increase in FLOPS on the test set compared to the ST-GCN model. Therefore, the proposed method exhibits favorable performance in enhancing both the accuracy and efficiency of abnormal behavior detection.

## Keywords

**Skeletal Data, Train Driver, Abnormal Behavior Detection, Graph Attention, Multi- Scale Temporal Convolutional Network**

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

列车司机的规范驾驶操作对于保障列车运行安全至关重要，异常的驾驶行为，诸如接打电话、吸烟、玩手机等，不仅对列车运行安全构成威胁，而且已成为造成重大交通事故的主要因素之一[1]。因此，对列车司机的驾驶行为进行异常行为检测具有至关重要的意义。

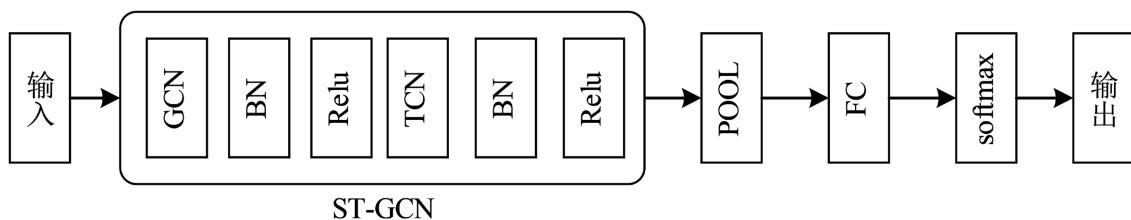
列车司机异常行为的检测方法主要有传统检测方法和基于深度学习检测方法。传统检测方法主要依赖监测驾驶员的生理信号和车辆传感器来识别司机的异常状态，包括心电图[2]、脑电图[3]和眼电图[4]等生理信号检测，以及车辆运动状态(如加速度、刹车等)信息[5]，上述方法在处理复杂环境下的异常行为检测时准确性较低[6]。随着深度学习方法在计算机视觉应用方面的迅猛发展，基于视频序列的深度学习行为检测方法逐渐成为主流，主要包括基于图像序列[7]、基于双流信息融合[8] [9]和基于人体骨架序列[10]等方法，基于图像序列的方法易受到光照和遮挡等因素的干扰[11]，而基于双流信息融合的方法则面临数据复杂性和处理难度较大的挑战[12]，人体骨架序列作为一种拓扑表示[13]方式，在面对复杂环境、不同人体尺度和运动速度等变化时具有数据结构简单、计算高效以及数据获取准确等优势[14] [15]。因此，骨架数据被越来越多用于行为检测。Li 等[16]设计了一种新的骨架数据表示，并通过卷积神经网络(Convolutional Neural Network, CNN)进行行为检测，解决了直接采用原始驾驶图像进行异常行为检测存在的特征不显著和数据量庞大的问题；Yin 等[17]将 2D 人体姿态估计网络与残差网络(Residual Network, ResNet)融合，实现了司机异常行为检测；Morais 等[18]利用动态骨架特征将骨骼运动分解为全局身体姿态和局部身体姿态，实现了离群点检测性能，减少了模型对骨架数据中单一关节点的过度关注；Song 等[19]提出了一个端到端的空间和时间注意力模型，通过长短期记忆网络(Long Short-Term Memory, LSTM)和递归神经网络(Recurrent Neural Network, RNN)选择性地关注每个帧间骨架上的变化关节点信息，并对不同帧间的骨架数据给予不同的关注度，实现了结合时空信息对司机异常行为检测；Yan 等[20]设计了一种骨架序列的动态骨架模型，符合人体关节连接的自然性和时间上的连续性，并将图卷积模块(Graph

Convolutional Network, GCN)和时域卷积模块(Temporal Convolutional Network, TCN)应用到网络模型中,提出了 ST-GCN 模型。

综上所述,在异常行为检测领域,传统方法大多直接采用复杂的图像信息进行检测,面临数据复杂和处理难度较大问题[21];基于深度学习的方法取得了一定的成就,但在处理列车司机驾驶过程中的光照、遮挡等噪声问题时,仍然面临时空信息难以充分利用的挑战[22]。因此,本文在 ST-GCN 模型的基础上提出了基于时空图注意力和多时间尺度时域卷积网络(ST-GAT)的列车司机异常行为检测模型。首先,利用姿态估计算法[23]提取骨架数据,并通过构建骨架时空图[20]实现对连续帧图像中骨架数据的时空信息整合。在模型的设计上,针对 GCN 模块存在的空间特征提取能力不足问题,引入了 GAT 模块[24],通过动态图注意力机制提升模型对空间信息的提取能力。此外,为解决 TCN 模块感受野不灵活、无法检测不同时间尺度的问题,本文设计了一种 MS-TCN 模块,以提取不同时间尺度下的时序特征,从而扩大模型的感受野,提高了不同时间尺度下骨架序列的信息融合能力。最后引入交叉熵损失函数加快模型收敛速度。

## 2. ST-GCN 行为检测网络

ST-GCN 是一种时空信息检测模型,专为基于骨架数据的行为识别任务而设计,通过 GCN 和 TCN 的结合,可以有效捕获骨架数据中的时空关系[19],将人体骨架动作视作图结构,其中每个节点代表身体的一个关键点。在 ST-GCN 的架构中,图卷积负责提取骨架节点之间的空间特征,而时间卷积则用于捕捉这些节点随时间的动态变化。这种结构使得 ST-GCN 能够更加高效和准确地理解和分析人体动作[10],能够直接从骨架数据中学习复杂的时空特征[25],而无需依赖于人工设计的特征提取方法。ST-GCN 的网络结构如图 1 所示。



**Figure 1.** ST-GCN network architecture

**图 1.** ST-GCN 网络结构

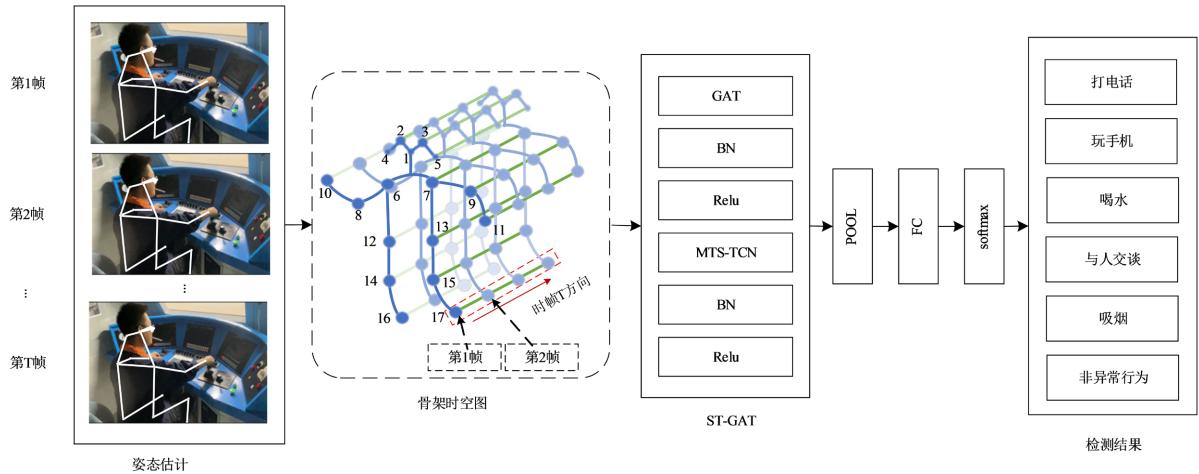
ST-GCN 网络结构包含 GCN 模块和 TCN 模块,每个模块之后都进行了归一化和激活函数处理,能够消除数据间的差异和提高模型的泛化能力,然后通过池化操作对特征进行下采样,并输入全连接层进行特征整合,最后通过 softmax 分类器对输入的骨架序列进行行为分类。

## 3. 基于 ST-GAT 的列车司机异常行为检测模型

基于 ST-GAT 的列车司机异常行为检测模型如图 2 所示。采用时空特征信息相结合的检测方法,骨架时空图中的空间特征信息采用 GAT 模块进行提取,时序特征信息利用 MS-TCN 模块进行提取,通过综合时空特征信息能够更全面地理解骨架数据的时空关系[19],从而更有效地进行异常行为检测。

图 2 中,首先从列车司机的监控视频中采集异常行为的视频帧,预处理后利用姿态估计算法从视频帧中提取列车司机的骨架数据;然后,以骨架数据为基础构建骨架时空图,该图不仅描绘了列车司机的动作轨迹,还整合了连续图像帧中的时空信息;此后,将构建的时空图数据输入 ST-GAT 网络模型,通过 GAT 模块对时空图的空间特征进行提取,并利用 MS-TCN 模块对不同时间尺度下时空图进行时序特

特征提取，经过归一化处理后，把提取到的时空特征进行元素相加；最后，将元素相加得到的特征向量通过池化和全连接处理后输入 softmax 分类器，最终输出列车司机异常行为的检测结果。



**Figure 2.** Train driver abnormal behavior detection model based on ST-GAT  
**图 2.** 基于 ST-GAT 的列车司机异常行为检测模型

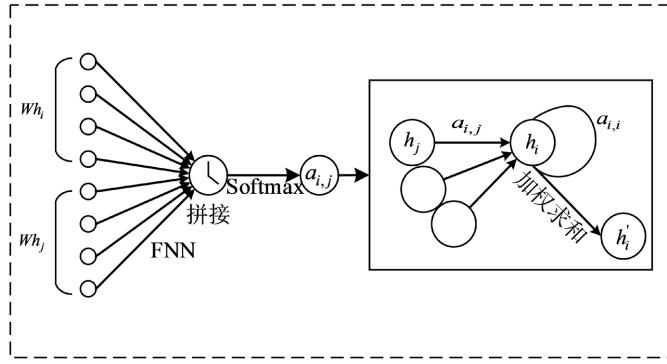
### 3.1. 骨架时空图

首先利用姿态估计算法从视频帧中获取列车司机的骨架数据，所得骨架关节点数据通常是以每帧中每个人体关节点的二维坐标来表示，通过将这些关节点连接起来形成单帧的人体骨架数据[11]，可视化后如图 2 中骨架时空图的第一帧所示。骨架时空图第一帧中 1~17 各关节点依次为鼻子、左眼、右眼、左耳、右耳、左肩、右肩、左肘、右肘、左腕、右腕、左髋、右髋、左膝、右膝、左踝、右踝。

首先构建一个具有  $N$  个节点和  $T$  帧的无向时空图  $G = (V, E)$ ，其中， $V$  表示骨架时空图中所有关节点集合， $E$  表示骨架时空图中所有骨架边集合。关节点集  $V = \{v_{ti} | t = 1, \dots, T, i = 1, \dots, N\}$ ，其中  $v_{ti}$  表示第  $t$  帧上第  $i$  个关节点。边集  $E$  是由两个子集构成，第一个子集  $E_S$  描述帧内相邻节点的骨架边连接，表示为  $E_S = \{v_{ti}v_{tj} | (i, j) \in H\}$ ，其中  $v_{tj}$  表示  $t$  帧上第  $j$  个关节点；第二个子集描述帧间对应关节点的骨架边连接，表示为  $E_F = \{v_{ti}v_{(t+1)i}\}$ ，其中  $v_{(t+1)i}$  表示第  $t+1$  帧上第  $i$  个关节点。所以在构建骨架时空图过程中，不仅要连接帧内相邻关节点，还要连接相邻帧之间的对应关节点，以形成包含空间和时间信息的骨架时空图，如图 2 中骨架时空图所示。

### 3.2. GAT 模块

传统异常行为检测方法通常采用固定权重方式进行关节点特征提取，难以充分挖掘不同节点之间的空间信息[11]。同时，实际复杂环境中存在多样的光照、遮挡等噪声，对骨架数据的准确性造成干扰[22]。为解决上述问题，引入了 GAT 模块，通过动态关注机制使模型能够在不同关节点之间的关系上实现动态聚焦。GAT 模块通过学习动态关注权重，提升模型适应不同节点之间空间关系的灵活性，从而更准确地捕捉骨架数据的运动状态和姿势。GAT 模块在空间信息提取方面具有显著优势，相较于固定权重的关节点特征提取，其动态关注机制使模型更灵活地适应不同节点之间的空间关系，有助于更准确地理解骨架数据的运动特征。此外，GAT 模块在处理骨架时空图数据时表现出一定的鲁棒性，通过动态调整对局部信息的关注度，有效减轻光照、遮挡等噪声对模型训练的不利影响。这为模型在复杂环境中更为可靠地执行异常行为检测任务提供了重要支持。GAT 模块的结构如图 3 所示。



**Figure 3.** GAT module  
**图 3.** GAT 模块

GAT 模块构建过程如下：

首先设置给定节点为  $v_i$ ， $v_j$  是节点  $v_i$  的邻居节点， $h = \{h_1, h_2, \dots, h_N\}$  表示输入的一组节点特征， $h' = \{h'_1, h'_2, \dots, h'_N\}$  表示输出一组新的节点特征。

1) 对节点  $v_i$ ，逐个计算它的邻居节点  $v_j$  和它自己的相似系数  $e_{i,j}$  为：

$$e_{i,j} = a([Wh_i \| Wh_j]) \quad (1)$$

式中： $a$  是一个通过单层前馈神经网络实现的映射函数； $[Wh_i \| Wh_j]$  表示对节点  $v_i$  和邻居节点  $v_j$  的特征进行拼接； $W$  是一个可学习的参数矩阵。

2) 为了使系数在不同节点之间进行比较，使用 Softmax 函数对注意力系数进行归一化，得到注意力系数  $\alpha_{i,j}$  为

$$\alpha_{i,j} = \text{softmax}(e_{i,j}) \quad (2)$$

3) 根据计算得到的注意力系数  $\alpha_{i,j}$ ，对邻居节点的特征按照注意力系数进行加权求和，得到节点  $v_i$  的新特征表示  $h'_i$  为：

$$h'_i = \text{ReLU}\left(\sum_{j \in p_i} \alpha_{i,j} Wh_j\right) \quad (3)$$

4) 为了增强模型表达能力，引入多头注意力机制，对于每个节点  $v_i$ ，计算多个注意力头结果，最后将它们拼接起来得到最终的特征表示  $h'_i(K)$  为：

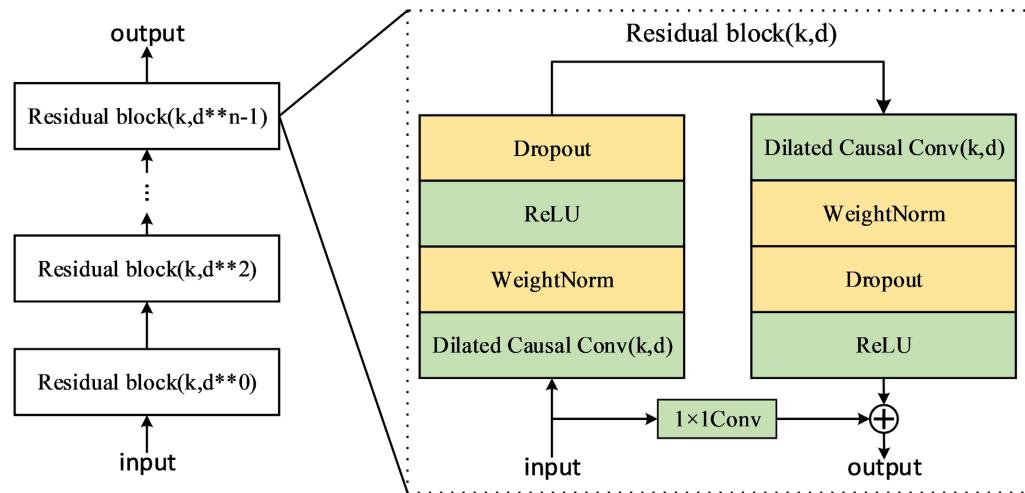
$$h'_i(K) = \left\| \text{ReLU}\left(\sum_{j \in p_i} \alpha_{ij}^k W^k h_j\right) \right\|_{k=1}^K \quad (4)$$

式中： $k$  表示注意力头的数量； $W^k$  是第  $k$  个注意力头的参数矩阵。

### 3.3. MS-TCN 模块

目前，单一时间尺度的 TCN 模块的感受野大小是通过单一卷积核的大小和层数来确定的，导致感受野不灵活[10]，难以适应不同时间尺度下的行为检测。为了解决这个问题，引入 MS-TCN 模块，该模块通过使用两个具有不同卷积核大小的 TCN 模块来处理输入序列，有助于模型更全面地捕捉不同尺度下的时序特征，从而更好地理解行为的动态变化。由于两个 TCN 模块可能导致冗余问题，引入了残差结构以减少冗余并缓解梯度消失问题。

MS-TCN 模块结构如图 4 所示。其中  $k$  表示卷积核大小， $d$  表示膨胀底数。

**Figure 4.** MS-TCN module**图 4.** MS-TCN 模块

假设该模块的输入序列为  $x_1, x_2, \dots, x_T$ , 其中  $T$  是序列长度,  $x_t$  是时刻  $t$  的输入向量。使用残差结构, 将输入  $x_t$  通过残差时空卷积层, 得到输出  $y_t$ 。残差时空卷积层的计算可表示为:

$$y_t = x_t + F(x_{t-k+1:t}) \quad (5)$$

式中  $k$  是卷积核的大小,  $F$  表示卷积操作及激活函数。

上述的残差结构通过  $x_t$  和  $F(x_{t-k+1:t})$  的加法操作, 实现了在保留原始输入信息的同时引入更多的非线性特征。这种结构在深层网络中有助于减少特征冗余和缓解梯度消失问题, 从而提升模型训练效果。这种上述设计使得 MS-TCN 能够更全面地理解行为的动态变化, 提升了模型对时序信息的表达能力。

### 3.4. 损失函数优化

ST-GCN 网络模型采用 Huber 损失函数, 存在训练过程对异常不敏感和收敛速度慢问题。Huber 损失函数是根据预测值和真实值之间的差异, 选择平方误差还是线性误差, 能够更好地平衡对离群值和普通值的拟合, 更适合回归任务[26]。由于列车司机异常行为检测是分类任务, 模型的输出是类别的概率分布, Huber 损失函数缺乏对概率分布的解释, 并且其梯度相对平滑, 导致训练速度较慢[27]。根据模型的任务类型和模型输出的概率分布, 使用交叉熵损失函数更符合分类任务的需求[28]。另外, 由于异常行为标签是以一种类别的形式表示[29], 交叉熵损失函数的梯度信息更为尖锐, 在反向传播时有更好的梯度传播, 能够更快地收敛和更稳定的训练。本文模型交叉熵损失函数表示如下:

$$\text{Loss} = -\sum_i p(i) \log(q(i)) \quad (6)$$

式中  $i$  表示类别的索引,  $p(i)$  是实际类别  $i$  的样本的真实概率(通常为 0 或 1),  $q(i)$  是模型预测类别为  $i$  的样本的概率。整个损失函数的目标是最小化实际概率分布和模型预测概率分布之间的交叉熵, 从而使模型更好地适应多分类任务。

## 4. 实验结果与分析

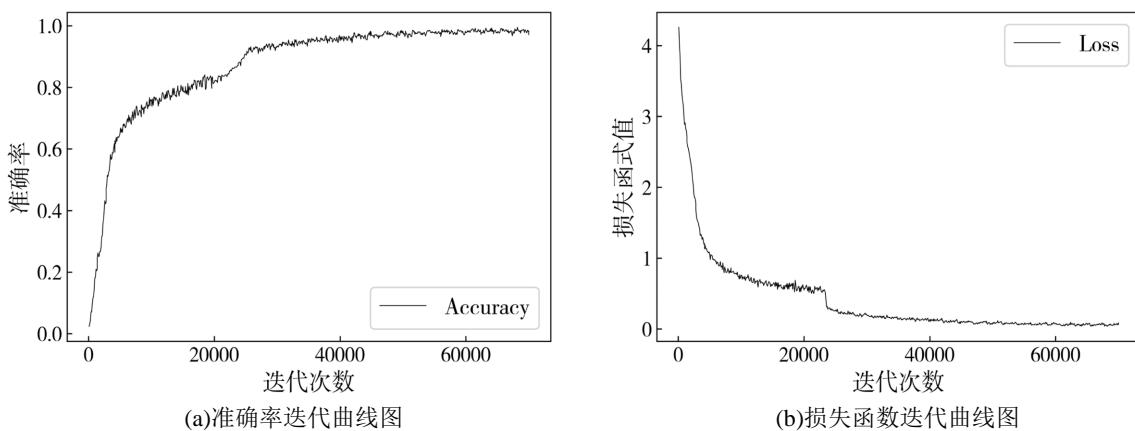
### 4.1. 实验环境与数据集

实验平台采用的显卡为 NVIDIA GeForce RTX 3060, 处理器为 i7-9700k, 内存 16GB, 使用 GPU 加速。

由于目前没有公开的列车司机异常行为检测数据集，本实验采用自采数据集，在列车司机模拟试验平台上利用正前方摄像头采集了 5500 个视频片段，涵盖 5 种异常行为，每个异常行为至少有 1000 个视频片段，每个片段持续约 5 s。本实验标记的异常行为包括接打电话、玩手机、喝水、与人交谈、吸烟等。在自采数据集标注方面，人工标注是解决该问题的主要方式之一。本实验使用可视化的图像标注工具 LabelImg，通过可视化操作对自采数据集中司机异常动作进行标注，每隔 10 帧随机抽取一帧关键帧，共计标记 20,000 张视频帧图像，并通过姿态估计模型 HRNet-W32 获取骨架数据。本文模型的训练，将标注好的自采数据集的 70% 作为训练集，30% 作为测试集。

#### 4.2. 训练的准确率和损失迭代曲线

为了有效地评估模型的训练效果，对训练模型的准确率和损失函数迭代曲线图进行了细致分析。首先，通过图 5(a)的准确率迭代曲线观察，在迭代次数达到 70,000 次后，准确率达到了峰值 0.982，这一结果表明模型在经过足够训练后能够有效地捕捉和学习输入数据的特征。其次，通过图 5(b)中的损失函数迭代曲线分析，损失函数的趋势表明模型在训练过程中逐渐减小预测误差，通过反复迭代逐步学习数据的特征。



**Figure 5.** Training iteration curve chart  
**图 5.** 训练迭代曲线图

#### 4.3. 实验性能指标与分析

为了客观评价本文模型的性能，采用准确率与 FLOPs 进行对比实验分析。实验对比结果如表 1 所示。

**Table 1.** Comparison of detection indicators for abnormal behavior of train drivers under different methods  
**表 1.** 不同方法下列车司机异常行为检测指标对比

方法	FLOPS	准确率/%
文献[16]	3.37	83.48
文献[19]	5.97	85.36
ST-GCN	16.20	89.47
本文方法	16.52	98.27

由表 1 可见，文献[16]的 FLOPs 和准确率相对较低，主要因为其方法直接将原始关节点坐标输入 CNN 进行标签预测。这种处理方式在捕捉时空信息方面效率较低，无法有效捕捉骨架序列中的复杂时空关系，

导致准确率相对较低。另外，文献[19]的方法虽然在 FLOPS 上略高，但由于未对骨架序列进行充分处理和相应的建模，因此在处理复杂时空关系时也存在一定局限不足。相比之下，ST-GCN 模型引入了骨架时空图和图卷积、时域卷积模块，相较于文献[16]和文献[19]，能够更全面地捕捉骨架序列的时空信息。因此，ST-GCN 在 FLOPS 和准确率上都取得了显著提升，凸显了骨架时空图建模和时空卷积对于提高检测性能的积极影响。而本文方法准确率和 FLOPS 均高于其他 3 种方法，这是因为采用图注意力机制能更好地提取骨架数据的空间信息，同时利用不同卷积核大小的 TCN 去增强模型对不同时间尺度语义信息的融合，提高了模型的准确率。为了进一步提高模型的处理速度，本文还引用了交叉熵损失函数，在保持高准确率的同时，提高了模型的 FLOPS。

#### 4.4. 模型消融实验

为直观反映出本文方法采用 GAT 模块、MS-TCN 模块以及优化损失函数等改进点对异常行为检测模型性能提升的作用，采用消融实验进行验证，通过在基准模型上进行训练以及测试，采用准确率作为评价指标，验证其有效性。所有消融实验均采用相同的实验环境，并以相同的数据集进行训练与测试。消融实验如表 2 所示。

消融实验表明，在 ST-GCN 基础上引入 GAT 模块后，通过图注意力卷积增强了各关节点之间空间信息的提取能力，将网络模型的准确率由 89.60% 提升到 94.65%；再引入 MS-TCN 模块，扩大模型的感受野，提高了不同时间下特征信息的捕捉能力，模型的准确率由 94.65% 提升到 96.91%；在训练过程中选择更适合分类任务的交叉熵损失函数进行优化，最终将模型的准确提升到 98.27%。本文方法相比于原 ST-GCN 网络模型平均准确率提升了 8.8%。

**Table 2.** Ablation experiment

**表 2. 消融实验**

基准模型	GAT 模块	MS-TCN 模块	交叉熵损失函数	准确率/%
√				89.60
√	√			94.65
√	√	√		96.91
√	√	√	√	98.27

#### 5. 结束语

针对列车驾驶场景下司机异常行为检测过程存在检测准确率以及效率低的问题，本文提出了一种基于骨架数据的列车司机异常行为检测方法。本文方法创新点如下：在列车驾驶场景下，利用骨架数据建模，更好理解列车司机异常行为的复杂关系和时序动态变化。在 ST-GCN 的基础上，引入 GAT 模块、设计 MS-TCN 模块以及优化损失函数，提高模型检测的准确率和检测效率。相较于 ST-GCN 模型，本文方法在准确率上取得了显著的提高，综合全类平均准确率达到了 98.27%，同时 FLOPS 提高了 2%。实验结果表明：本文方法能够有效识别列车司机的多种异常行为，包括打电话、看手机、喝水、与人交流和抽烟等异常行为，具有较好性能。未来的研究将进一步拓展该模型的适用范围，并结合其他传感器数据进行更全面的异常行为识别和预测，进一步提升列车运行的安全性。

#### 参考文献

- [1] 秦斌斌, 彭良康, 卢向明, 等. 司机分心驾驶检测研究进展[J]. 计算机应用, 2021, 41(8): 2330-2337.
- [2] 肖致明, 刘文, 杨建芳, 等. 心理状态对火车司机安全绩效影响综述[J]. 中国安全科学学报, 2022, 32(S2): 13-18.

- [3] 王洪涛, 殷浩钧, 陈创泉, 等. 基于脑电信号的驾驶疲劳检测综述[J]. 华中科技大学学报(自然科学版), 2022, 50(11): 54-65.
- [4] Liu, L., Ji, Y.F., Gao, Y., et al. (2021) A Novel Fatigue Driving State Recognition and Warning Method Based on EEG and EOG Signals. *Journal of Healthcare Engineering*, 2021, Article ID: 7799793. <https://doi.org/10.1155/2021/7799793>
- [5] 惠飞, 郭静, 贾硕, 等. 基于双向长短记忆网络的异常驾驶行为检测[J]. 计算机工程与应用, 2020, 56(24): 116-122.
- [6] 李小平, 白超. 一种基于多模态信息融合的火车司机疲劳驾驶检测方法[J]. 铁道学报, 2022, 44(6): 56-65.
- [7] 栗志磊, 李俊, 施智平, 等. 用于视频行为识别的高效二维时序建模网络[J]. 计算机工程与应用, 2023, 59(3): 127-134.
- [8] 张冰冰, 葛疏雨, 王旗龙, 等. 基于多阶信息融合的行为识别方法研究[J]. 自动化学报, 2021, 47(3): 609-619.
- [9] Dai, C., Liu, X. and Lai, J. (2020) Human Action Recognition Using Two-Stream Attention Based LSTM Networks. *Applied Soft Computing*, 86, Article 105820. <https://doi.org/10.1016/j.asoc.2019.105820>
- [10] 甄昊宇, 张德. 结合自适应图卷积与时态建模的骨架动作识别[J]. 计算机工程与应用, 2023, 59(18): 137-144.
- [11] 孙晓虎, 余阿祥, 申栩林, 等. 混合注意力机制的异常行为识别[J]. 计算机工程与应用, 2023, 59(5): 140-147.
- [12] Zhao, Y., Man, K.L., Smith, J., et al. (2020) Improved Two-Stream Model for Human Action Recognition. *EURASIP Journal on Image and Video Processing*, 2020, Article No. 24. <https://doi.org/10.1186/s13640-020-00501-x>
- [13] Zhang, H.B., Zhang, Y.X., Zhong, B., et al. (2019) A Comprehensive Survey of Vision-Based Human Action Recognition Methods. *Sensors*, 19, Article 1005. <https://doi.org/10.3390/s19051005>
- [14] 张晓平, 纪佳慧, 王力, 等. 基于视频的人体异常行为识别与检测方法综述[J]. 控制与决策, 2022, 37(1): 14-27.
- [15] Yu, C., Xiao, B., Gao, C., et al. (2021) Lite-HRNet: A Lightweight High-Resolution Network. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, 20-25 June 2021, 10440-10450. <https://doi.org/10.1109/CVPR46437.2021.01030>
- [16] Li, C., Zhong, Q., Xie, D., et al. (2017) Skeleton-Based Action Recognition with Convolutional Neural Networks. 2017 IEEE International Conference on Multimedia & Expo Workshops(ICMEW), Hong Kong, 10-14 July 2017, 597-600.
- [17] 尹智帅, 钟恕, 聂琳真, 等. 基于人体姿态估计的分心驾驶行为检测[J]. 中国公路学报, 2022, 35(6): 312-323.
- [18] Moraes, R., Le, V., Tran, T., et al. (2019) Learning Regularity in Skeleton Trajectories for Anomaly Detection in Videos. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, 15-20 June 2019, 11996-12004. <https://doi.org/10.1109/CVPR.2019.01227>
- [19] Song, S., Lan, C., Xing, J.L., et al. (2017) An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, San Francisco, 4-9 February 2017, 4263-4270. <https://doi.org/10.1609/aaai.v31i1.11212>
- [20] Yan, S.J., Xiong, Y.J. and Lin, D.H. (2018) Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, New Orleans, 2-7 February 2018, 7444-7452. <https://doi.org/10.1609/aaai.v32i1.12328>
- [21] 刘宝龙, 周森, 董建锋, 等. 基于骨架的人体动作识别技术研究进展[J]. 计算机辅助设计与图形学学报, 2023, 35(9): 1299-1322.
- [22] 钱慧芳, 易剑平, 付云虎. 基于深度学习的人体动作识别综述[J]. 计算机科学与探索, 2021, 15(3): 438-455.
- [23] Sun, K., Xiao, B., Liu, D., et al. (2019) Deep High-Resolution Representation Learning for Human Pose Estimation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, 15-20 June 2019, 5693-5703. <https://doi.org/10.1109/CVPR.2019.00584>
- [24] Velickovic, P., Cucurull, G., Casanova, A., et al. (2017) Graph Attention Networks. arXiv:1710.10903.
- [25] 邓森磊, 高振东, 李磊, 等. 基于深度学习的人体行为识别综述[J]. 计算机工程与应用, 2022, 58(13): 14-26.
- [26] 肖振久, 孔祥旭, 宗佳旭, 等. 自适应聚焦损失的图像目标检测算法[J]. 计算机工程与应用, 2021, 57(23): 185-192.
- [27] 孟宪法, 刘方, 李广, 等. 卷积神经网络压缩中的知识蒸馏技术综述[J]. 计算机科学与探索, 2021, 15(10): 1812-1829.
- [28] 张硕, 叶勤, 史婧, 等. 改进 RangeNet++损失函数的车载点云小目标语义分割方法[J]. 计算机辅助设计与图形学学报, 2021, 33(5): 704-711.
- [29] 刘炎, 熊雨露, 褚菲, 等. 基于多源异构信息自适应融合的镁砂熔炼过程运行状态评价[J]. 仪器仪表学报, 2023, 44(9): 322-337.