

# 控制与决策

Control and Decision

## 一种基于触觉信息的脉冲图注意力残差卷积物体检测算法

吴培良, 林为梁, 毛秉毅, 陈雯柏, 高国伟

引用本文:

吴培良, 林为梁, 毛秉毅, 陈雯柏, 高国伟. 一种基于触觉信息的脉冲图注意力残差卷积物体检测算法[J]. *控制与决策*, 2023, 38(9): 2537–2544.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.2244>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

#### 结合注意力机制的循环神经网络复述识别模型

Recurrent neural networks based paraphrase identification model combined with attention mechanism

控制与决策. 2021, 36(1): 152–158 <https://doi.org/10.13195/j.kzyjc.2019.0638>

#### 复杂背景下全景视频运动小目标检测算法

Panoramic video motion small target detection algorithm in complex background

控制与决策. 2021, 36(1): 249–256 <https://doi.org/10.13195/j.kzyjc.2019.0686>

#### 基于双分支特征融合的场景文本检测方法

A scene text detection based on dual-path feature fusion

控制与决策. 2021, 36(9): 2179–2186 <https://doi.org/10.13195/j.kzyjc.2020.0002>

#### 基于小波变换与差分变异BSO-BP算法的大坝变形预测

Dam deformation prediction based on wavelet transform and differential mutation BSO-BP algorithm

控制与决策. 2021, 36(7): 1611–1618 <https://doi.org/10.13195/j.kzyjc.2019.1431>

# 一种基于触觉信息的脉冲图注意力残差卷积物体检测算法

吴培良<sup>1,2†</sup>, 林为梁<sup>1,2</sup>, 毛秉毅<sup>1,2</sup>, 陈雯柏<sup>3</sup>, 高国伟<sup>3</sup>

(1. 燕山大学 信息科学与工程学院, 河北 秦皇岛 066000; 2. 河北省计算机虚拟技术与系统集成重点实验室, 河北 秦皇岛 066000; 3. 北京信息科技大学 自动化学院, 北京 100000)

**摘要:** 触觉传感器(柔性电子皮肤)在机器人进行人机交互和工具操作时发挥着重要作用,如何有效利用触觉信息进行物体检测是当前研究的主要瓶颈. 鉴于此,提出一种脉冲图残差卷积神经网络SNN-Atten-ResGCN的物体检测算法. 首先使用图残差网络ResGCN模型训练触觉时间序列的表征信息,通过引入深度学习模型中的注意力机制拟合触觉数据图结构的局部特征;然后对重构的触觉图形输入由3个LIF神经元和2个FC全连接层组成的SNN脉冲神经网络训练得到特征向量;最后投票层Vote解码网络特征并检测物体类别. 在EvTouch-Objects和EvTouch-Containers两个家庭常见物体触觉数据集上进行对比实验,实验结果表明,所提出方法在保证模型迭代效率的同时,对各种不同的家庭工具对象和容器对象的检测准确率、精度、召回率和 $F_1$ -score均有提升.

**关键词:** 触觉感知; 物体检测; 脉冲神经网络; 图神经网络; 注意力机制

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.2244

**引用格式:** 吴培良,林为梁,毛秉毅,等. 一种基于触觉信息的脉冲图注意力残差卷积物体检测算法[J]. 控制与决策, 2023, 38(9): 2537-2544.

## An object detection algorithm based on convolution of attention residuals of pulse graph based on tactile information

WU Pei-liang<sup>1,2†</sup>, LIN Wei-liang<sup>1,2</sup>, MAO Bing-yi<sup>1,2</sup>, CHEN Wen-bai<sup>3</sup>, GAO Guo-wei<sup>3</sup>

(1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066000, China; 2. The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066000, China; 3. School of Automation, Beijing Information Science and Technology University, Beijing 100000, China)

**Abstract:** Tactile sensors (flexible electronic skin) play an important role in robot human-computer interaction and tool operation. How to effectively use tactile information for object detection is the main bottleneck of current research. Therefore, a pulse graph convolution neural network, SNN-Atten-ResGCN, is proposed for object detection. Firstly, the graph residual network ResGCN model is used to train the representation of tactile time series. Secondly, the attention mechanism in the deep learning model is introduced to fit the local features of the graphic structure of tactile data. Thirdly, the reconstructed tactile graphics are input, and the SNN pulse neural network composed of three LIF neurons and two FC full connection layers is trained to obtain the feature vector. Finally, vote is utilized to decode the network feature components and discriminate object category. Comparative experiments are carried out on the tactile datasets of EvTouch-Objects and EvTouch-Containers. The experimental results show that the proposed method ensures the model iteration efficiency, and improves the accuracy, precision, recall rate and  $F_1$ -score of various household objects and container are improved.

**Keywords:** tactile perception; object detection; impulse neural network; graph neural network; attention mechanism

## 0 引言

触觉作为人体的重要感觉器官,因其能够准确感知目标的形状、重量、冷热、柔软度、光滑度等信息,被

认为是仅次于视觉的重要信息来源. David Julius和Ardem Patapoutian因发现人体温度感知机制以及触觉机制,获得了2021诺贝尔生理学或医学奖. 在服务

收稿日期: 2021-12-26; 录用日期: 2022-04-27.

基金项目: 国家重点研发计划项目(2018YFB1308300); 国家自然科学基金项目(62276028,U20A20167); 北京市自然科学基金项目(4202026); 河北省自然科学基金项目(F202103079).

责任编辑: 张文安.

†通讯作者. E-mail: peiliangwu@gmail.com.

\*本文附带电子附录文件,可登录本刊官网该文“资源附件”区自行下载阅览.

机器人领域,相较于视觉和测距传感器,触觉传感器通过与工具、环境和服务对象的物理接触,有效提升了机器人提供自然人机交互和共融式宜人服务的能力。此外,在智能穿戴和元宇宙的研究领域,触觉传感器也已成为感知外部环境的核心基础元件。在触觉信息处理方面,基于触觉信息的物体检测方法已经成为研究热点。早期的触觉信息物体检测大多依赖于触觉传感器、机械夹具等硬件系统设计,不同种类触觉传感器的触觉数据生成原理和采集方式往往表现出较大的差异性,导致相应的物体检测算法存在各自的局限性。基于机器学习,尤其是深度学习的方法能够有效提取用于物体检测的触觉特征从而减小硬件系统差距,但由于服务场景中物体种类多样且形态各异,目前算法仍存在检测准确率和精度偏低的问题。

针对上述问题,本文提出基于触觉信息的脉冲图注意力残差卷积物体检测算法。将图注意力机制引入图残差网络Res-GCN,优化聚合邻接节点信息,学习邻接节点的权重特征,实现对邻接节点的加权聚合。在保证模型迭代效率下,通过重构触觉图形序列提升物体检测算法的准确率和精度等关键指标。

## 1 国内外研究现状

本文工作参考了基于触觉信息物体检测方法的最新进展,通过结合深度学习理论进行触觉信息数据分析,能够有效完成对目标物体的感知<sup>[1-2]</sup>,触觉数据信息特征的有效性已经在一些服务机器人物体检测与操作问题上得到验证<sup>[3-5]</sup>。

Chris<sup>[6]</sup>设计了一种配备触觉感知的机器人手指,用于辨识埋藏在水、米或沙子中的物体。该手指在遇到阻力时可灵活转动并灵敏感知到被埋藏物体的详细形状,但是这种尖型手指难以很好地适用于日常物体的检测。

Cui等<sup>[7]</sup>研制了一种高密度柔性触觉手套,该手套传感器阵列由807个独立的压阻传感单元组成。利用PCA降维算法对抓取物体过程中的形状和姿态轨迹进行解码,然后设计了基于卷积神经网络的触觉物体检测算法,但该触觉手套对触觉传感器的柔性度和分布密度要求较高。

Taunayazov等<sup>[8]</sup>使用iCub<sup>[9]</sup>仿真平台收集了23个物体的触觉信息,基于卷积神经网络和循环神经网络构建了触觉表征的物体纹理分类检测模型。Wang等<sup>[10]</sup>提出一种基于触觉信息序列的卷积神经网络,使用电容性柔性触觉电子皮肤收集压力数据,为确保在垂直方向上收集数据的稳定性,通过在电子皮肤上施加恒定质量载荷并拖动柔性电子皮肤产生水平滑

动现象,输出触觉压力信息特征的同时保留滑动过程中的时空信息,该方法需要控制触觉传感器做出滑动运动,一定程度上限制了电子皮肤的适用范围。

Bok等<sup>[11]</sup>开发了一种机器人指尖模型,可以通过直接接触区分物体的纹理和形状。该机器人指尖模型由具有多模态(力反馈和温度)传感功能的触觉传感器组成,能够根据硬度、摩擦系数、粗糙度和热导率4种触觉模式实现对16个样本的正确检测,但其普适性有待进一步验证。

Chun等<sup>[12]</sup>构建了一种人工神经网络触觉手指系统,该系统使用基于粒子聚合物复合传感器信号转换系统模拟人类触觉识别过程。复合传感器有选择地响应压力和振动,产生类似人体触觉神经元的输出信号模式,最后将输出信号与深度学习技术相结合对目标物体进行检测识别,该方法需要转换传感器信息,未能直接提取触觉特征。

Gu等<sup>[13]</sup>提出了基于异步事件驱动触觉传感器taxel<sup>[14-15]</sup>的触觉物体检测方法,触觉传感器taxel异步感知环境中的变化,生成触觉脉冲数据,与标准的基于同步时间的传感器<sup>[16-17]</sup>相比,事件驱动传感器具有更高的效能、更好的可扩展性以及更低的延迟,但仍存在检测准确率和精度偏低的问题。

综上所述,目前针对基于触觉信息的物体检测研究侧重于触觉信号采集<sup>[18]</sup>和触觉传感阵列设计<sup>[19-20]</sup>。本文主要工作在于改进现有触觉信息的分析算法以提高检测准确率和精度,将注意力机制与图残差卷积网络相结合,通过处理触觉传感器采集的局部触觉信息特征生成触觉图形序列,实现对邻近节点特征的加权聚合,并使用脉冲神经网络拟合触觉图形序列来提升物体检测准确率和精度。

## 2 模型整体框架

针对文献[14]中由39个触觉传感器所组成阵列采集得到的家庭常见物体触觉信息,构建一种基于新型脉冲图注意力残差卷积网络(SNN-Atten-ResGCN)的物体检测算法。所构建的SNN-Atten-ResGCN网络的主要组件包括LIF神经元模型、图注意力残差网络Atten-ResGCN、脉冲神经网络SNN以及最终用于物体检测的投票层Vote。SNN-Atten-ResGCN网络模型结构如图1所示。

由图1直观可见:该模型的第1部分为输入序列,根据欧氏距离连接触觉传感器触点构建的触觉图形数据;第2部分为经过训练调整的图注意力残差网络Atten-ResGCN,用于拟合触觉图形数据中节点与节点之间的特征;第3部分为经过训练调整的脉冲神经

网络SNN,用于根据输入时间序列的图形特征计算得出最终的特征向量;第4部分为Vote投票层,用于对特征向量进行分类,得到单个物体检测结果。

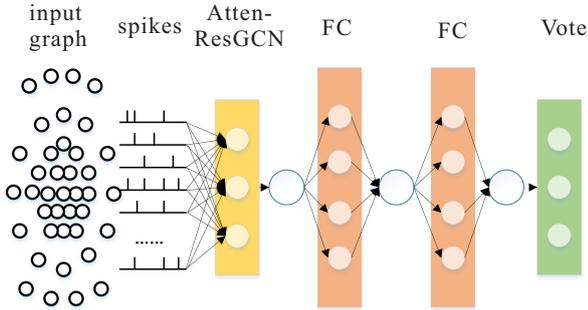


图1 SNN-Atten-ResGCN模型网络结构

### 3 脉冲图注意力残差卷积网络

#### 3.1 图注意力残差卷积网络

图卷积神经网络旨在通过聚集邻域顶点的特征提取更丰富的本顶点特征,GCN通过将每个顶点 $v$ 与特征向量相关联来表示顶点的信息,因此可以通过连接所有顶点的特征表示一个图 $G$ 的特征,即

$$h_G = [h_{v_1}, h_{v_2}, \dots, h_{v_N}]^T \in R^{N \times D}, \quad (1)$$

其中 $N$ 为集合 $V$ 的基数,即图 $G$ 的顶点数。

图卷积运算分为两部分,一是聚合运算,聚集本节点的相邻特征信息和相邻顶点的特征信息;二是更新运算,根据上一步聚合运算得到的特征信息更新下一节点的信息。上述图卷积运算也称为GCN的消息传递机制。GCN图卷积在数学上定义为

$$h_{v_l} = \sum_{u_l \in N(v_l)} \frac{1}{\sqrt{\deg(v_l)}\sqrt{\deg(u_l)}} (\Theta h_{u_{l-1}}), \quad (2)$$

其中 $\deg(v_l)$ 和 $\deg(u_l)$ 分别为节点 $v$ 和其邻接节点 $u$ 的度。邻接节点 $u$ 的特征 $h_{u_{l-1}}$ 先与权重矩阵 $\Theta$ 进行变换,按节点的度归一化处理后进行线性组合。

基于GCN的消息传递机制,第 $l$ 层顶点输入到第 $l+1$ 层顶点的传播方式定义为

$$g_{l+1} = F(g_l, w_l). \quad (3)$$

其中: $g_{l+1}$ 为第 $l$ 层输出信息,即第 $l+1$ 层的输入信息; $g_l$ 为第 $l$ 层的输入信息。根据图卷积运算引入聚合函数和更新函数后,式(3)可转化为

$$g_{l+1} = \text{Update}(\text{Aggregate}(g_l, w_l^{\text{agg}}), w_l^{\text{update}}), \quad (4)$$

其中 $w_l^{\text{agg}}$ 和 $w_l^{\text{update}}$ 分别为聚合函数和更新函数的可学习线性变换权重矩阵。聚合函数用于编译来自邻域顶点的节点信息,而更新函数对聚合信息执行非线性变换以计算新的顶点信息。对于 $l+1$ 层的某一节点 $v_{l+1} \in V_{l+1}$ ,其顶点特征 $h_{v_{l+1}}$ 计算如下:

$$h_{v_{l+1}} = \phi(h_{v_l}, \rho(\{h_{u_l} | u_l \in N(v_l)\}), h_{v_l}, w_\rho), w_\phi). \quad (5)$$

其中: $\rho$ 为顶点特征聚合函数, $h_{v_l}$ 和 $h_{v_{l+1}}$ 分别为第 $l$ 层和第 $l+1$ 层的顶点特征, $N(v_l)$ 为在第 $l$ 层顶点 $v$ 相邻节点的集合, $h_{u_l}$ 为第 $l$ 层顶点 $v$ 相邻节点的特征, $w_\rho$ 和 $w_\phi$ 为上面提到的聚合函数和更新函数的权重矩阵。

相较于CNN,图卷积网络GCN在处理非规则数据的过程中具有较大优势,但现有GCN仍存在梯度消失(vanishing gradient)和过度平滑(over smoothing)等问题,导致在数据样本偏小的场景下局域特征都收敛到相同的值上。针对该问题,本文通过引入残差-连接结构改进ResGCN,并与图注意力机制结合构建了Atten-ResGCN图注意力残差网络模型。如图2所示,Atten-ResGCN图注意力残差网络模型由ResGCN残差块和Attention注意力层构成,输入触觉图形数据经由ResGCN残差块处理后输入到Attention注意力层。ResGCN残差块的主干网络部分为TAG-Conv自适应图卷积和LayerNorm标准归一化层,残差连接部分为GEN-Conv广义通用图卷积和Message Normalization消息规范化层。

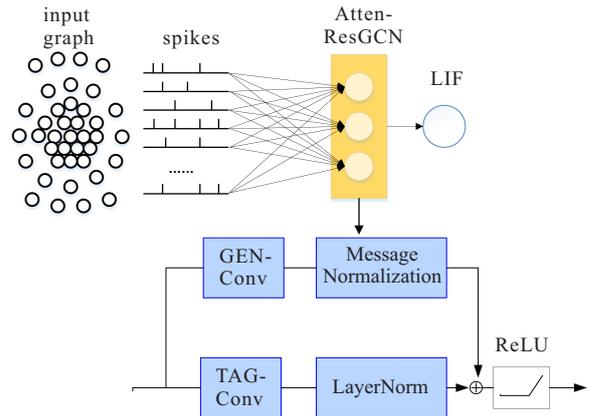


图2 Atten-ResGCN图注意力残差网络模型

与传统的图卷积GCN相比,在主干网络上采用TAG-Conv自适应图卷积,能够更好地匹配输入图形的拓扑,拟合顶点 $v$ 与相邻节点的特征。TAG-Conv自适应图卷积定义如下:

$$H_l = \sum_{c=1}^C G_{c,l} * h_c + b_l. \quad (6)$$

其中: $H_l$ 为第 $l$ 层所有节点的输出特征,即第 $l+1$ 层的输入特征,是第 $h_c$ 层中某一节点 $c$ 的输入特征; $C$ 为总节点数; $b_l$ 为第 $l$ 层的偏置向量;\*为卷积算子定义图卷积操作; $G_{c,l}$ 为第 $l$ 层的图形滤波器卷积核。自适应图卷积使用图的归一化邻接矩阵 $A$ ,定义如下:

$$G_{c,l} = \sum_{k=0}^K g_{c,f,k} A^k. \quad (7)$$

通过借鉴残差神经网络中残差-连接的思想,引

入图残差卷积神经网络ResGCN,传播方式为

$$\begin{aligned} g_{l+1} &= H(g_l, w_l) = \\ F(g_l, w_l) + R(g_l, w_l) &= g_l^{\text{res}} + g_l. \end{aligned} \quad (8)$$

其中 $H$ 为定义的从 $l$ 层到 $l+1$ 层的传播映射.将常规的图卷积映射 $H$ 和残差图卷积映射 $R$ 线性组合拟合结果, $g_l^{\text{res}}$ 为第 $l$ 层的图残差映射输出结果, $g_l$ 为第 $l$ 层的图卷积映射输出结果.

残差连接卷积部分采用基于图卷积神经网络消息传递机制的广义通用图卷积GEN-Conv.与目前国内外研究常使用的均值聚合函数(mean aggregator)或注意力聚合函数(attention aggregator)不同的是,本文所设计的图卷积神经网络消息传递机制的聚合函数为幂均值聚合函数,聚合函数定义如下:

$$\text{PowerMean\_Agg}_p(\cdot) = \left( \frac{1}{|N(v)|} \sum_{u \in N(v)} h_u^p \right)^{\frac{1}{p}}. \quad (9)$$

其中: $h_u^p$ 为对当前节点 $v$ 的邻接顶点 $u$ 的节点特征取 $p$ 次方, $N(v)$ 为顶点 $v$ 的邻接节点集合对顶点 $v$ 的邻接节点的累加特征结果均值后再开 $p$ 次方.

经过GEN-Conv处理过的图形特征会经过消息规范化层(message normalization),规范化层在训练深层神经网络方面能够显著提高网络聚合性能.消息规范化层结合多层感知机MLP对顶点特征和邻接节点特征聚合结果更新 $m_{v_l}$ ,计算公式如下:

$$h_{v_{l+1}} = \text{MLP} \left( h_{v_l} + s \frac{\|h_{v_l}\|_2 m_{v_l}}{\|m_{v_l}\|_2} \right). \quad (10)$$

其中: $s$ 为设定的更新比例因子; $m_{v_l}$ 为由聚合函数PowerMean\_Agg $_p$ 聚集的顶点 $v$ 相邻接节点的特征; $\|h_{v_l}\|_2$ 和 $\|m_{v_l}\|_2$ 分别为顶点特征和邻接节点特征的 $v_l$ 向量范数,即欧氏范数.

基于上述消息传递框架构建的图卷积神经网络GCN能够通过聚合函数和更新函数拟合顶点的相邻节点特征.在一张连通图中,由于节点周围用于聚合的邻居节点重合度较高,每个节点训练得到的特征表示基本一致,导致节点特性趋于消失.因此在图网络的实际应用中,不同节点间的影响权重往往是变化的.为了实现动态适应性,本文通过引入注意力机制构建一个用于描述节点之间重要性的网络层Attention注意力层,以学习不同相邻节点之间的可训练权重.

在Atten-ResGCN模型结构中,任意两个节点 $(i, j)$ 之间的注意力值 $f_{ij}$ 为

$$f_{ij} = \text{LeakyReLU}(W^T[wh_i \| wh_j]). \quad (11)$$

其中: $h_i$ 和 $h_j$ 分别为顶点与顶点的特征向量拼接再映射成为一个标量, $w$ 为投影映射矩阵.

在聚合邻接节点特征时,对每个节点所有邻接节点注意力值使用PowerMean\_Agg $_p$ 函数进行归一化,归一化后的注意力聚合系数作为注意力得分 $s_{ij}$ ,有

$$s_{ij} = \text{soft max}(v_{ij}) = \exp(v_{ij}) / \sum_{k \in N} \exp(v_{ik}). \quad (12)$$

注意力机制模型如图3所示.设当前节点为 $v$ , $s_v$ 为除当前节点 $v$ 外其余节点的注意力得分,对注意力得分 $s_v$ 进行排序生成有序集合 $N^{(d)}(v) = \{u_1, u_2, \dots, u_n\}$ ,从集合 $N^{(d)}(v)$ 中选取节点 $v$ 的 $k$ 个邻接节点,重构触觉图形.

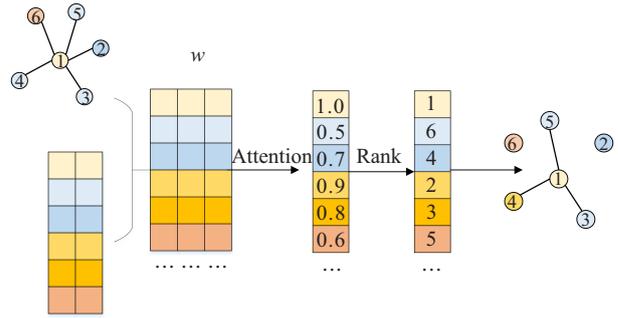


图3 注意力机制

Attention注意力层允许图形结构在每一层发生变化,从而较好地拟合物体的图形表示.引入注意力机制对顶点间的边进行重计算以动态改变感受野,缓解较深的网络带来过度平滑的问题.

### 3.2 SNN脉冲神经网络

相较于使用高精度浮点数据的卷积神经网络,脉冲神经网络能够减少将脉冲数据转换成矩阵的环节,因此可以有效提高模型迭代效率.本节将Atten-ResGCN处理得到的触觉图形数据作为SNN脉冲神经网络的输入.如图4所示,SNN主要由3个LIF神经元和2个全连接层组成<sup>[14]</sup>.

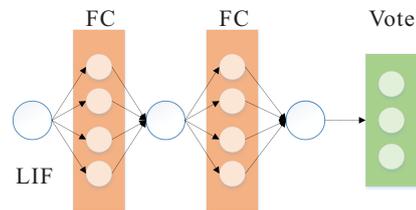


图4 SNN脉冲神经网络结构

设FC全连接层的输入层向量为 $X = [x_0, x_1, \dots, x_n]^T$ ,输出特征为 $H = [h_0, h_1, \dots, h_m]^T$ .将输入从 $n$ 维转换成 $m$ 维需要左乘 $m \times n$ 的矩阵 $W$ ,继而与偏置 $B = [b_0, b_1, \dots, b_m]^T$ 进行线性求和.

考虑到非线性激活函数ReLU难以有效适配脉冲神经网络来处理触觉序列信息,使用LIF神经元模型作为脉冲神经网络的激活函数.设 $I(t)$ 为神经元的输入信号, $o(t)$ 为神经元的输出信号, $u_T$ 为神经元

的膜电位,当神经元的膜电位超过触发阈值时,LIF神经元将生成脉冲,之后膜电位将被重置为 $u_R$ .

LIF模型的电回路组成为由电流 $I$ 驱动的电阻 $R$ 和电容 $C$ ,如果驱动电流 $I(t)$ 消失,则整个电容器的电压重置为 $u_R$ .根据电流守恒定律,将驱动电流分为

$$I(t) = I_R + I_C. \quad (13)$$

第2个流过电容 $C$ 的电流 $I = C \frac{du}{dt}$ ,引入时间常数 $\tau = RC$ ,可得LIF激活函数为

$$\tau \frac{du(t)}{dt} = -u(t) + \sum_i w_i x_i. \quad (14)$$

其中 $u(t)$ 为 $t$ 时刻神经元内膜电位 $\sum_i w_i x_i$ , $\sum_i w_i x_i$ 为前一层神经元输入的加权和.将式(14)转换为

$$u(t+1) = \left(1 - \frac{dt}{\tau}\right)u(t) + \frac{dt}{\tau} \sum_i w_i x_i. \quad (15)$$

进一步令 $\beta = \left(1 - \frac{dt}{\tau}\right)$ , $w'_i = \frac{dt}{\tau} \sum_i w_i x_i$ ,可得

$$u(t+1) = \beta u(t) + \sum_i w'_i x_i. \quad (16)$$

其中: $\beta = (1 - dt/\tau)$ 定义为一个衰减因子, $w'_i$ 为结合了 $dt/\tau$ 的权重.由此LIF神经元模型激活函数描述为

$$f_{LIF}(u) = \text{fire spike} \ \& \ u(t) \leftarrow u_R, \ u(t) \geq u_T. \quad (17)$$

Vote投票层用于解码网络输出,每个输出的分类标签首先与投票层中的一个神经元关联,预测的类别是与指定时间窗内具有最大平均投票数(对应于脉冲数)的神经元相关联的类别.

模型损失函数定义为

$$L = \left\| y - \frac{1}{T} \sum_{t=1}^T U o^t \right\|^2. \quad (18)$$

其中: $U$ 为投票矩阵,将每个投票神经元连接到特定类别的投票矩阵; $o^t$ 为时间 $t$ 的最后一层输出特征.该函数捕获给定时间窗口内标签特征 $y$ 与平均投票结果之间的均方误差.

### 3.3 基于脉冲图注意力残差卷积网络的物体检测

综合上面设计的图注意力残差卷积网络和脉冲神经网络,提出一种新型脉冲图注意力残差卷积网络SNN-Atten-ResGCN,并将其应用于基于触觉信息的物体检测任务.

针对节点的空间分布,由式(19)计算用于描述两节点间邻接关系的欧氏距离,作为模型SNN-Atten-ResGCN的原始输入触觉数据,有

$$d(v_i, v_j) = \|v_i - v_j\|_2. \quad (19)$$

SNN-Atten-ResGCN在接收到原始输入触觉数据后,首先通过图注意力机制和图残差卷积网络拟合

不规则局部触觉信息结构特征,选择性筛选正相关的触觉信息特征;然后使用脉冲神经网络拟合触觉图形序列来提升物体检测准确率和精度.

基于SNN-Atten-ResGCN网络模型的物体检测算法伪代码描述如下.

input: 39个触觉节点数据及触觉节点间边关系;

output: 物体检测准确率、精度、召回率、 $F_1$ -score.

step 1: 根据39个触觉节点数据以及触觉节点间边关系构建原始触觉图形数据.

step 2: 将step 1生成的原始触觉图形数据输入到Atten-ResGCN网络层中,由式(6)和(7)先后在主干网络上通过TAG-Conv层和LayerNorm层得到图卷积映射,由式(9)和(10)先后在残差链接上通过GEN-Conv和Message Normalization得到图残差映射,线性组合得到触觉节点特征图.

step 3: 针对step 2得到的触觉节点特征图,基于式(12)计算各个节点的注意力得分,对触觉图形进行重构.

step 4: SNN接收到step 3的重构后触觉图形,通过3层LIF神经元和2层FC全连接层得到物体检测特征向量.

step 5: 针对step 4得到的物体检测特征向量,Vote投票层通过式(18)减小均方误差,得到具有最大平均脉冲数的类别作为物体检测标签.

step 6: 重复step 1~step 5,得到数据集所有样本的物体检测标签,计算物体检测混淆矩阵并据此计算准确率、精度、召回率和 $F_1$ -score.

## 4 实验及结果分析

实验主要目的是设计对比实验和消融实验,评估所提出基于SNN-Atten-ResGCN的触觉信息物体检测性能.将本文方法与脉冲图卷积神经网络SNN-GNN、脉冲自适应图卷积网络SNN-TAG-GCN等4种深度学习模型进行对比.消融实验中,本文方法与缺失残差连接部分、缺失图卷积主干部分等SNN-Atten-ResGCN的4种变体模型进行对比.

### 4.1 实验数据集

实验选用目前较为完备的家庭常见物体数据集EvTouch-Objects和EvTouch-Containers.

EvTouch-Objects: 数据集包括来自36个对象类别的触觉数据,26个来自YCB家庭工具数据集的对象,其余10个对象是可形变对象,作为对YCB数据集的一种补充.在数据收集过程中,机器人抓手抓住物体,将其从桌子上抬起20 cm,然后放回桌子上.收集从举起物体到释放物体这段时间内的数据( $\approx 5$  s).

对于每个对象类别收集20个样本,共720个样本.

EvTouch-Containers: 该数据集包括4个容器的触觉数据: 一个咖啡罐、一个塑料苏打瓶、一个豆奶盒和一个金属金枪鱼罐. 容器的最大重量分别为250 g、400 g、300 g和140 g. 每个容器分别装有0%、25%、50%、75%、100%容器容量的物体,产生20个类别. 在数据收集过程中,机器人抓手抓住每个容器将其抬高桌子5 cm. 利用抓取物体期间收集的数据,将其举起并保持一段时间(约6.5 s),共300个样本.

#### 4.2 模型超参数

实验中将本文构建的模型与以下4种深度学习模型性能进行比较. 所有方法均使用LIF神经元,共享相同的SNN脉冲神经网络结构和相同的超参数,具体参数如表1所示.

表1 深度学习模型超参数

参数	值
膜电位阈值 $u_T$	0.5
膜电位重置阈值 $u_R$	0
Batch size	1
膜电位衰减因子 $\beta$	0.2
学习率	0.001
矩形函数参数 $a$	0.5

#### 4.3 实验评估指标

针对基于触觉信息的物体检测问题,通过混淆矩阵计算准确率accuracy、精度precision、召回率recall和 $F_1$ -score综合评估本文方法的检测性能,各项指标如下:

1) 准确率accuracy: 对于给定的测试数据集,模型检测正确的样本数与测试数据集样本数之比,即检测正确的概率.

2) 精度precision: 设给定检测物体类别为正样本,其余类别为负样本,检测正确的正样本个数占模型检测物体类别为正样本个数的比例. 精度precision计算方式如下:

$$\text{precision} = \frac{N_{TP}}{N_{TP} + N_{FP}}. \quad (20)$$

其中:  $N_{TP}$  为检测结果判定正样本,实际上是正样本的数量;  $N_{FP}$  为检测结果判定正样本,实际上是负样本的数量.

3) 召回率recall: 检测正确的正样本个数占真实正样本个数的比例. 召回率recall计算方式如下:

$$\text{recall} = \frac{N_{TP}}{N_{TP} + N_{FN}}. \quad (21)$$

其中  $N_{FN}$  为检测结果判定负样本,实际上是正样本的数量.

4)  $F_1$ -score. 统计学中用来衡量二分类模型精确

度的一种指标,同时兼顾了检测模型的精度和召回率,是模型精度和召回率的一种加权平均. 首先根据混淆矩阵计算出每个类别的精度precision和召回率recall,再计算每个类别的 $F_1$ -score $_i$ ,对每个类别的 $F_1$ -score $_i$ 求和取平均值,有

$$F_1\text{-score}_i = 2 \frac{\text{recall}_i \times \text{precision}_i}{\text{recall}_i + \text{precision}_i}, \quad (22)$$

$$F_1\text{-score} = \sum_n^i F_1\text{-score}_i / n. \quad (23)$$

#### 4.4 对比实验及结果分析

将上述两个数据集分为一个训练集(80%)和一个测试集(20%),使用Adam优化器对训练数据数据集上的每个模型进行100次迭代,比较图网络模型GNN、TAG-GCN、TAG-ResGCN、Atten-GCN和Atten-ResGCN在EvTouch-Objects和EvTouch-Containers数据集上与SNN结合的性能表现.

通过对比实验得出5种模型在EvTouch-Objects和EvTouch-Containers数据集上物体检测平均准确率accuracy如表2所示. 可见,本文方法在EvTouch-Objects和EvTouch-Containers数据上的准确率相较于当前最优模型TAG-GCN提高了约2.4%和3.4%. 5种模型的训练损失和测试损失在开始阶段都迅速降低,然后逐渐收敛. 与其他4种模型相比,SNN-Atten-ResGCN模型具有更低的trainloss和testloss.

表2 5种模型检测准确率accuracy

模型	EvTouch-Objects	EvTouch-Containers
SNN-GNN	85.14	58.83
SNN-TAG-GCN	89.44	64.17
SNN-TAG-ResGCN	88.76	63.51
SNN-Atten-GCN	86.47	61.79
SNN-Atten-ResGCN	91.61	66.34

同时,为了验证本文模型相较于其他4种模型在迭代效率上的性能,对比两种数据集上一次迭代的平均耗时如表3所示(单位: s). 由表3可见,5种模型的耗时差距不大,效率差距均在0.5%左右. 另外,采用本文方法分别计算EvTouch-Objects对象数据集和EvTouch-Containers对象数据集的分类混淆矩阵,通过式(20)和(21)计算分类模型的精度和召回率.

表3 5种模型一次迭代平均耗时

模型	EvTouch-Objects	EvTouch-Containers
SNN-GNN	57.625	61.382
SNN-TAG-GCN	58.563	62.655
SNN-TAG-ResGCN	57.931	62.068
SNN-Atten-GCN	57.863	61.824
SNN-Atten-ResGCN	58.934	62.789

为便于直观对比,表4和表5列出了SNN-GNN、SNN-TAG-GCN、SNN-TAG-ResGCN、SNN-Atten-GCN和SNN-Atten-ResGCN等5种模型在数据集EvTouch-Objects和EvTouch-Containers上的精度和召回率.进一步地,通过式(22)和(23)计算得到各模型的 $F_1$ -score,具体如表6所示.

表4 5种模型精度 precision

模型	EvTouch-Objects	EvTouch-Containers
SNN-GNN	0.8402	0.5850
SNN-TAG-GCN	0.8889	0.6361
SNN-TAG-ResGCN	0.8819	0.6193
SNN-Atten-GCN	0.8681	0.5867
SNN-Atten-ResGCN	0.9097	0.6700

表5 5种模型集召回率 recall

模型	EvTouch-Objects	EvTouch-Containers
SNN-GNN	0.8922	0.6130
SNN-TAG-GCN	0.9332	0.9332
SNN-TAG-ResGCN	0.9297	0.6545
SNN-Atten-GCN	0.9156	0.6191
SNN-Atten-ResGCN	0.9522	0.7020

表6 5种模型 $F_1$ -score

模型	EvTouch-Objects	EvTouch-Containers
SNN-GNN	0.8424	0.5836
SNN-TAG-GCN	0.8938	0.6400
SNN-TAG-ResGCN	0.8874	0.6228
SNN-Atten-GCN	0.8724	0.5803
SNN-Atten-ResGCN	0.9159	0.6770

由表4~表6可见,无论是分类模型的精度和召回率,还是衡量多分类模型精确度的指标 $F_1$ -score,SNN-Atten-ResGCN模型都明显优于其他4种模型.从EvTouch-Objects数据集上的表现看,精度precision提高了约2.3%,召回率recall提高了约2.0%, $F_1$ -score提高了约2.4%.在EvTouch-Containers数据集上的表现看,精度precision提高了约5.0%,召回率recall提高了约5.2%, $F_1$ -score提高了约5.4%.

综合各组实验对比数据可见,本文方法在保证模型迭代效率的同时,对不同家庭工具对象和容器对象的准确率、精度、召回率和 $F_1$ -score均有较明显提升.

#### 4.5 消融实验及结果分析

设计消融实验进一步验证所提出SNN-Atten-ResGCN方法中4个主要部分的作用,即Attention注意力层、残差连接部分Residual Block、图卷积部分GCN和脉冲神经网络部分SNN.因此,消融实验设置的5组模型为:1)SNN-Atten-ResGCN:完整的图注意力残差卷积网络;2)SNN-Atten-ResGCN<sub>w/oRB</sub>:缺失残差连接部分;3)SNN-Atten-ResGCN<sub>w/oGCN</sub>:缺失图卷积主干部分;4)SNN-Atten-ResGCN<sub>w/oAtten</sub>:缺失注

意力部分;5)SNN-Atten-ResGCN<sub>w/oSNN</sub>:缺失脉冲神经网络部分.

由表7~表10列出的两个触觉信息数据集上的消融实验表现看,SNN-Atten-ResGCN中缺失上述4个主要部分之一时,衡量多分类模型性能的准确率、精度、召回率和 $F_1$ -score等性能指标均出现不同程度的下降.针对消融实验结果具体分析如下.

1)缺失注意力层的模型性能下降最大,原因是图注意力机制对每个图节点赋予不同的权重,捕获节点主要特征,弱化次要特征,优化聚合邻接节点信息,实现对邻接节点的加权聚合,丢失图注意力导致无法提取主要特征造成性能指标下降幅度最大.

2)缺失图卷积主干部分的模型性能下降略小于缺失注意力层,原因是残差连接部分的广义通用图卷积和消息规范化层可以弥补部分由于主干网络的缺失造成性能下降.

3)脉冲神经网络部分的主要作用是解码触觉图形信息,生成物体检测标签类别,缺失此部分造成的性能下降略高于缺失图残差连接.

4)缺失图残差连接部分性能下降最小,残差连接的作用是防止深度图网络层数的加深引发梯度爆炸、消失等现象,但本文提出的图注意力残差卷积网络需要保证模型迭代效率没有选择加深图残差网络结构,因此缺失图残差连接部分对物件检测准确率和精度等指标影响最小.

表7 消融实验准确率 accuracy

模型	EvTouch-Objects	EvTouch-Containers
SNN-Atten-ResGCN	91.61	66.34
SNN-Atten-ResGCN <sub>w/oRB</sub>	89.10	63.37
SNN-Atten-ResGCN <sub>w/oGCN</sub>	88.15	62.28
SNN-Atten-ResGCN <sub>w/oAtten</sub>	86.89	60.89
SNN-Atten-ResGCN <sub>w/oSNN</sub>	88.67	62.64

表8 消融实验精度 precision

模型	EvTouch-Objects	EvTouch-Containers
SNN-Atten-ResGCN	0.9097	0.6700
SNN-Atten-ResGCN <sub>w/oRB</sub>	0.8896	0.6310
SNN-Atten-ResGCN <sub>w/oGCN</sub>	0.8782	0.6189
SNN-Atten-ResGCN <sub>w/oAtten</sub>	0.8514	0.6024
SNN-Atten-ResGCN <sub>w/oSNN</sub>	0.8893	0.6267

表9 消融实验召回率 recall

模型	EvTouch-Objects	EvTouch-Containers
SNN-Atten-ResGCN	0.9522	0.7020
SNN-Atten-ResGCN <sub>w/oRB</sub>	0.9324	0.6875
SNN-Atten-ResGCN <sub>w/oGCN</sub>	0.9189	0.6634
SNN-Atten-ResGCN <sub>w/oAtten</sub>	0.8991	0.6507
SNN-Atten-ResGCN <sub>w/oSNN</sub>	0.9178	0.6639

表 10 消融实验  $F_1$ -score

模型	EvTouch-Objects	EvTouch-Containers
SNN-Atten-ResGCN	0.9159	0.6770
SNN-Atten-ResGCN <sub>w/oRB</sub>	0.9005	0.6380
SNN-Atten-ResGCN <sub>w/oGCN</sub>	0.8867	0.6203
SNN-Atten-ResGCN <sub>w/oAtten</sub>	0.8646	0.6101
SNN-Atten-ResGCN <sub>w/oSNN</sub>	0.8931	0.6237

## 5 结论

本文提出了一种基于触觉信息的脉冲图注意力残差卷积物体检测算法,引入图注意力机制构建图注意力残差卷积网络 Atten-ResGCN,并与脉冲神经网络模型 SNN 结合组成脉冲图注意力残差卷积网络 SNN-Atten-ResGCN。在 EvTouch-Objects 和 EvTouch-Containers 两个触觉数据集上的实验结果表明,本文方法对家庭常见物体检测准确率和精度均明显优于现有算法,为后续基于触觉信息的自然交互和共融式宜人服务奠定了基础。

### 参考文献(References)

- [1] Cheng L, Liu Z Y. Flexible tactile sensing technology and its application in medical rehabilitation robots[J]. Control and Decision, 2022, 37(6): 1409-1432.
- [2] He Z W, Huang H, Wu Y D, et al. Consistent scale normalization for object perception[J]. Applied Intelligence, 2021, 51(7): 4490-4502.
- [3] Dai Y, Xue Y, Zhang J X. Human-inspired haptic perception and control in robot-assisted milling surgery[J]. IEEE Transactions on Haptics, 2021, 14(2): 359-370.
- [4] Meng X R, Song Y. Multimodal haptic interaction technology and its application[J]. Journal of Physics: Conference Series, 2020, 1650(3): 032138.
- [5] Herzig N, He L, Maiolino P, et al. Conditioned haptic perception for 3D localization of nodules in soft tissue palpation with a variable stiffness probe[J]. PLoS One, 2020, 15(8): e0237379.
- [6] Chris Vavra. Robotic finger uses tactile sensing to find buried items[J]. Control Engineering, 2021, 68(6): 53-59.
- [7] Cui Z J, Gao T S, Wang J, et al. Tactile glove-decode and classify the human grasping process[C]. The 40th Chinese Control Conference. Shanghai, 2021: 8004-8009.
- [8] Taunyazov T, Sng W, See H H, et al. Event-driven visual-tactile sensing and learning for robots[J]. Robotics: Science and Systems, 2020, 53(4): 73-78.
- [9] Koptev M, Figueroa N, Billard A. Real-time self-collision avoidance in joint space for humanoid robots[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 1240-1247.
- [10] Wang Y F, He B, Zhou Y M, et al. Surface texture recognition network based on flexible electronic skin[C].

The 40th Chinese Control Conference. Shanghai, 2021: 458-463.

- [11] Bok B G, Jang J S, Kim M S. Texture identification of objects using a robot fingertip module with multimodal tactile sensing capability[J]. Applied Sciences, 2021, 11(11): 52-56.
- [12] Chun S, Kim J S, Yoo Y, et al. An artificial neural tactile sensing system[J]. Nature Electronics, 2021, 4(6): 429-438.
- [13] Gu F Q, Sng W, Taunyazov T, et al. TactileSGNet: A spiking graph neural network for event-based tactile object recognition[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, 2020: 9876-9882.
- [14] Tulyakov S, Gehrig D, Georgoulis S, et al. Time lens: event-based video frame interpolation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 16150-16159.
- [15] Lee W W, Tan Y J, Yao H, et al. A neuro-inspired artificial peripheral nervous system for scalable electronic skins[J]. Science Robotics, 2019, 4(32): eaax2198.
- [16] Gehrig D, Loquercio A, Derpanis K, et al. End-to-end learning of representations for asynchronous event-based data[C]. IEEE/CVF International Conference on Computer Vision. Seoul, 2019: 5632-5642.
- [17] Mitrokhin A, Ye C X, Fermüller C, et al. EV-IMO: Motion segmentation dataset and learning pipeline for event cameras[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems. Macao, 2019: 6105-6112.
- [18] Gerstner W. Time structure of the activity in neural network models[J]. Physical Review E, 1995, 51(1): 738-758.
- [19] Wu Y J, Deng L, Li G Q, et al. Direct training for spiking neural networks: Faster, larger, better[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 1311-1318.
- [20] Shrestha S, Orchard G. SLAYER: Spike layer error reassignment in time[J]. Proceedings of Advances in Neural Information Processing Systems, 2018, 1(1): 1412-1421.

### 作者简介

吴培良(1981—),男,教授,博士生导师,从事智能信息处理、智能服务机器人等研究, E-mail: peiliangwu@gmail.com;

林为梁(1997—),男,硕士生,从事触觉信息处理、智能服务机器人等研究, E-mail: 447593787@qq.com;

毛秉毅(1964—),男,副研究员,博士,从事智能服务机器人、智能信息处理等研究, E-mail: ysdxmby@163.com;

陈雯柏(1975—),男,教授,博士生导师,从事智能信息处理、智能服务机器人等研究, E-mail: chenwb@bistu.edu.cn;

高国伟(1963—),男,教授,博士生导师,从事智能信息处理、柔性触觉传感器等研究, E-mail: ggw@bistu.edu.cn.