CN 51-1346/O4 ISSN 1003-501X (印刷版) ISSN 2094-4019 (网络版)

光电工程
Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

基于昇腾处理器的边端人体动作识别算法设计与实现

赵冬冬，赖亮，陈朋，周鸿超，李亦然，梁荣华

引用本文:

赵冬冬，赖亮，陈朋，等. 基于昇腾处理器的边端人体动作识别算法设计与实现[J]. 光电工程，2024， **51**(6): 240072.

Zhao D D, Lai L, Chen P, et al. Design and implementation of edge-based human action recognition algorithm based on ascend processor[J]. *Opto-Electron Eng*, 2024, **51**(6): 240072.

https://doi.org/10.12086/oee.2024.240072

收稿日期: 2024-03-25; 修改日期: 2024-05-23; 录用日期: 2024-05-23

相关论文

基于ZYNQ的轻量化YOLOv5声呐图像目标检测算法及实现
赵冬冬，谢墩翰，陈朋，梁荣华，沈伊，郭新新
光电工程 2024, **51**(1): 230284 doi: 10.12086/oee.2024.230284

基于残差和注意力网络的声呐图像去噪方法
赵冬冬，叶逸飞，陈朋，梁荣华，蔡天诚，郭新新
光电工程 2023, **50**(6): 230017 doi: 10.12086/oee.2023.230017

伪标签细化引导的相机感知无监督行人重识别方法
程思雨，陈莹
光电工程 2023, **50**(12): 230239 doi: 10.12086/oee.2023.230239

多特征聚合的红外-可见光行人重识别
郑海君，葛斌，夏晨星，邬成
光电工程 2023, **50**(7): 230136 doi: 10.12086/oee.2023.230136

更多相关论文见光电期刊集群网站

OEE 光电工程 Opto-Electronic Engineering

http://cn.oejournal.org/oee

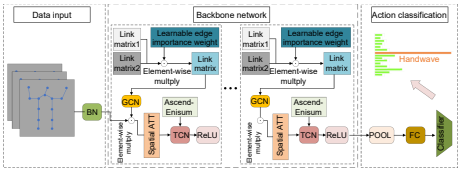
OE_Journal Website

Opto-Electronic Engineering
光 电 工 程

Article
2024 年, 第 51 卷, 第 6 期

DOI: 10.12086/oee.2024.240072

基于昇腾处理器的边端人体动作识别算法设计与实现



赵冬冬，赖　亮，陈　朋*，周鸿超，
李亦然，梁荣华

浙江工业大学计算机科学与技术学院，浙江 杭州 310023

摘要：针对现有的人体动作识别算法精度不足、计算量大、缺少在边端设备上的部署等问题，本文提出一种基于昇腾处理器的边端轻量化人体动作识别时空图卷积算法。通过设计隐性联系骨架连接方法并构建隐性邻接矩阵，结合自然骨架连接邻接矩阵，构造显隐性融合空间图卷积。在时间维度加入空间注意力机制，使模型关注不同帧间关节点位置空间特征，进一步设计时间图卷积，构建时空图卷积。此外设计网络中的 Ascend-Enisum 算子，进行张量融合运算，降低了计算复杂度，使模型轻量化。针对上述改进，在 KTH 数据集上进行实验验证，与经典单流算法 ST-GCN 相比，模型计算量减小了 22.28%，Top-1 精度达到 84.17%，提升了 5%。基于上述算法设计了昇腾 AI 人体动作识别系统，并在边端设备成功部署，可以进行实时人体动作识别。

关键词：边端人体动作识别；昇腾处理器；时空图卷积；轻量化

中图分类号：TP391　　**文献标志码：**A

赵冬冬，赖亮，陈朋，等. 基于昇腾处理器的边端人体动作识别算法设计与实现 [J]. 光电工程，2024，**51**(6): 240072
Zhao D D, Lai L, Chen P, et al. Design and implementation of edge-based human action recognition algorithm based on ascend processor[J]. *Opto-Electron Eng*, 2024, **51**(6): 240072

Design and implementation of edge-based human action recognition algorithm based on ascend processor

Zhao Dongdong, Lai Liang, Chen Peng*, Zhou Hongchao, Li Yiran, Liang Ronghua

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China

Abstract: Aiming at the problems of existing human action recognition algorithms such as insufficient accuracy, large amount of calculation, and lack of deployment on edge devices, this paper proposes an edge-side lightweight human action recognition spatial temporal graph convolutional algorithm based on the Ascend processor. By designing an implicit skeletal connection method and constructing an implicit adjacency matrix, combined with the natural skeletal connection adjacency matrix, we create an explicit-implicit fusion spatial graph convolution. A spatial attention mechanism is added to the temporal dimension, enabling the model to focus on spatial features of joint positions across different frames. Furthermore, we design a temporal graph convolution to construct a

收稿日期：2024-03-25；修回日期：2024-05-23；录用日期：2024-05-23
基金项目：国家自然科学基金面上项目 (62371421)；浙江省领军型创新创业团队项目 (2021R01002)；浙江省自然科学基金重大项目 (LD24F020005)
*通信作者：陈朋，chenpeng@zjut.edu.cn

spatiotemporal graph convolution. Additionally, the Ascend-Enisum operator is designed within the network to perform tensor fusion operations, reducing computational complexity and lightening the model. Experimental validation on the KTH dataset demonstrates that, compared to the classical single-stream ST-GCN algorithm, our model achieves a 22.28% reduction in computational cost while attaining a Top-1 accuracy of 84.17%, representing a 5% improvement. Based on this algorithm, we have designed the Ascend AI human action recognition system, which has been successfully deployed on edge devices for real-time human action recognition.

Keywords: edge human action recognition; ascend processor; spatial temporal graph convolutional; lightweight

1 引　言

近年来，人工智能迅猛发展，智能设备已广泛融入人们日常生活中的各个领域，显著提升了生活的便利性。这种进步为多个领域带来了巨大变革，还促成了运动交互产品的发展，如运动跟踪器、智能运动服以及互动健身车等，这些产品可以协助健康管理、运动训练和安全保护。就这些产品而言，动作识别是关键部分，然而这些产品在对用户进行动作识别时，使用方式大多较为繁琐，需要佩戴特定的传感器等。随着计算机视觉的发展，基于视觉算法的非接触式运动交互产品随之兴起。但是，此类产品对人体动作识别[1]算法有较高要求，不仅要求系统能够理解和分类人体的各种动作，还要能够进行实时分析和反馈。因此在资源受限的边端设备上进行人体动作识别成为研究热点之一。

基于视觉技术的动作识别任务的核心在于利用计算机算法从视频或图像序列中自动识别和解析出人体的运动状态和行为模式。从算法输入的数据模态角度来看，当前主流的识别算法主要包括基于人体轮廓特征、深度图序列、光流、多模态数据和骨架信息这几种动作识别算法。

基于人体轮廓特征的动作识别算法[2]，主要通过提取和分析人体轮廓特征来进行人体动作识别，此类算法极易受到背景、光照、遮挡等因素影响，鲁棒性较差。基于深度图序列的动作识别算法[3-4]，主要通过深度信息来捕捉人体动作行为，由于深度图提供了更丰富的信息，此类算法虽然消除了基于人体轮廓存在的数据模糊等问题，但是深度图序列的数据复杂度高，导致相关动作识别模型需要较大的参数量，不仅使得训练时间长，还对硬件设备有一定要求，限制了其可扩展性。基于光流的动作识别算法[5-6]，通过分析图像序列中像素点的运动模式来识别人体动作，此类算法动态信息捕获能力强，但是存在光照敏感度高、运动边界模糊等问题。基于多模态数据的人体动作识别算法[7-8]，通过多个传感器的数据共同分析和识别人体动作，此类方法存在数据之间难以同步和对齐且模态间数据信息冗余等问题。基于骨架的人体动作识别算法[9-12]，通过从图像序列中提取人体骨架信息进行动作识别，此类方法虽然数据简单，对光照敏感度低，具有较好的鲁棒性并且骨架数据更加贴切地反映了人体动作的物理本质，从而可以更好地表征人体运动的过程。但是目前的基于骨架的算法，存在计算量大、时空信息利用不足，在空间维度其主要关注自然连接的人体骨骼点之间的联系，缺乏对关节点隐性联系的关注，例如跑步时、左手和右脚会有协同作用。针对不同帧之间相似动作关节点的位置缺少关注。同时此类算法缺少完善的系统，难以在边端设备上进行部署。

本文做出了如下贡献：1) 根据人体骨架信息，提出对关节点连接的隐性划分方式。构建一个可学习的邻接矩阵来动态捕捉关节点之间的隐性联系及其联系强度。结合显性特征，设计融合显隐性联系的空间图卷积。2) 针对时空图序列中，关节点位置变化可能对动作识别存在影响，加入空间注意力，使得模型可以更加关注关节点不同帧的空间位置特征，从而解决传统方法对于不同帧间关节点位置空间信息处理不充分的问题。3) 针对 Ascend-STGCN 中 Einsum 算子在进行复杂运算时计算消耗大，且不适配昇腾处理器的问题，本文设计了 Ascend-Einsum 算子，通过进行张量维度融合计算，减小了计算过程中的开销，使模型轻量化，同时解决了 Einsum 算子在昇腾处理器上适配性的问题。最后，结合目标检测算法 YOLOv5 和姿态估计算法 OpenPose，提出一种 YOP-Ascend-STGCN 算法框架，用于实现边端人体动作识别系统。在昇腾处理器上部署后，该算法实现了每秒 26 帧的动作识别速度。

2 相关工作

2.1 目标检测网络

目前，基于深度学习的目标检测算法主要分为两大类：无锚点 (anchor free) 算法和基于锚点 (anchor based) 算法。无锚点算法直接预测目标的关键点或者中心区域，而不依赖预设框，如 CornerNet[13]、CenterNet[14]、FSAF[15]、SAPD[16] 等算法，尽管这类算法简化了模型结构，但是可能会面临检测结果不稳定、语义模糊以及正负样本不均衡等问题。另一方面，基于锚点的算法可以细分为双阶段和单阶段策略[17]。其中，双阶段算法，如 RCNN[18]、Faster RCNN[19]、FPN[20] 等算法，此类算法通过区域提议和分类回归两个阶段来提高检测精度，但往往速度较慢且计算复杂度较高。相比之下，单阶段算法，如 YOLO 系列、SSD[21]、RetinaNet[22] 等算法。此类算法通过直接预测边界框和类别概率，实现了更快的检测速度。在单阶段算法中，YOLO 系列算法具有速度快和结构简单的特点。特别是 YOLOv5[23-24] 算法，它内存占用低、在速度和精度上做出了较好的平衡，更适合在嵌入式设备上运行。

2.2 姿态估计网络

目前，基于深度学习的姿态估计算法主要有 AlphaPose[25]、MobileNet[26]、OpenPose[27] 等。其中 AlphaPose 算法基于 Resnet50、Resnet101 网络模型，这两个网络模型计算量较大，需要较高的计算能力和显存支持。MobileNet 算法虽然速度快，但是在进行骨骼点提取时无法处理遮挡和复杂背景的情况，其鲁棒性较差。而 OpenPose 网络通过使用卷积神经网络 (CNN) 提取出两种关键信息：置信度图 (part confidence maps, PCMs) 和关联性图 (part affinity fields, PAFs)。其中 PCMs 用于获取关键点位置信息和置信度，PAFs 用于预测骨骼点方向信息，得到关键点位置和关键点方向向量。然后将沿着连接两个关键点的线 (肢体段)，计算该线上每个像素的 PAFs 向量与连接向量的点积分，作为两个关键点之间的相关性。最后通过贪心分析算法对人体关键点进行编码，获得人体骨骼点信息。在提取行人骨架时，具有较好的鲁棒性，同时实时性也表现良好，适合在嵌入式式设备上运行。

3 基于昇腾处理器的人体动作识别算法设计

3.1 YOP-Ascend-STGCN 算法设计

图 1 展示了算法的总体框架，一共分为三个部分：YOLO 行人提取模块、OpenPose 骨骼提取模块以及 Ascend-STGCN 动作分析模块。首先，系统从摄像头、视频文件或图像序列中获取输入数据，使用 YOLOv5 模型检测图像中的目标，并获取目标的边界框以及类别标签，对每个检测到的目标类别进行筛选，将每个 Person 目标裁剪出来，形成目标图片序列。其次，将目标图片序列输入 OpenPose 模型提取人体姿态，该模块获取目标的骨骼点信息，如头、肩膀、手肘、膝盖等，然后对目标关键点进行关联连接，获取骨架序列。最后，将骨架序列，输入 Ascend-STGCN 模型对其进行动作识别，分析骨架在不同时间的变化来确定动作类别。

3.2 Ascend-STGCN 动作识别模块算法设计

3.2.1 定义并设计人体关节点

用 YOP-Ascend-STGCN 算法的目标检测模块，快速筛选出带有 person 标签的有效视频帧序列，再利用姿态识别模块，提取关键帧序列中的 14 个人体关节点信息，将这些点作为人体运动特征信息，构建关节点的显性集合：$H=\{V_0,V_1,\cdots,V_{13}\}$，同时，根据关节点可能存在的隐性联系，构建关节点的隐性集合：$P=\{V_4,V_7,V_{10},V_{13}\}$。如图 2 所示，是本文构建的骨架显性关联和隐性关联结构图。0~13 编号分别表示人体 (头、颈、右肩、右肘、右腕、左肩、左肘、左腕、右髋、右膝、右踝、左髋、左膝、左踝)。图中实线是人体骨架显性关联 (人体自然连接)，虚线是隐性关联 (人体非自然连接，四肢之间相对位置)。可以构成两个边集显性边集E_S和隐性边集E_P。其中隐性联系符合人类运动的一般行为逻辑。因此，融合显性联系和隐性联系可以有效解决骨架图无法学习距离较远的关节点的动作之间关联性的问题，从而提升动作识别精度。

3.2.2 时空信息建模

人体骨架时空图涵盖了时间和空间两个关键维度。根据视频帧序列中的人体关节点时空信息，可以构建时空图:$G=(V,E)$，图中包含关节点集 V 和边集 E。定义关节点:$V=\{V_{ti}|i=1,\cdots,N,t=1,\cdots,T\}$，时空图中的边集表示为:$E=\{E_S,E_P,E_T\}$，其中帧内人体骨架

显性连接表示为:$E_S=\{V_{ti}V_{tj}|t=1,\cdots,T,(i,j)\in H\}$，$H$为自然连接的人体关节集合，帧内人体骨架隐性联系：$E_P=\{V_{ti}V_{tj}|t=1,\cdots,T,(i,j)\in P\}$, P表示隐性连接的人体关节点集合，帧间连接:$E_T=\{V_{ti}V_{(t+1)i}|t\in 1,\cdots,N\}$, V_{ti}表示在t时刻关节点V_i对应时空图中的图节点。V_{ti}坐标为：$P_{ti}=\{x(t,i),y(t,i),t\in 1,\cdots,N\}$，表示对每一帧图像所生成的关节点，将相邻两帧人体关节对应点相连。形成骨架时空图，如图 3 所示。

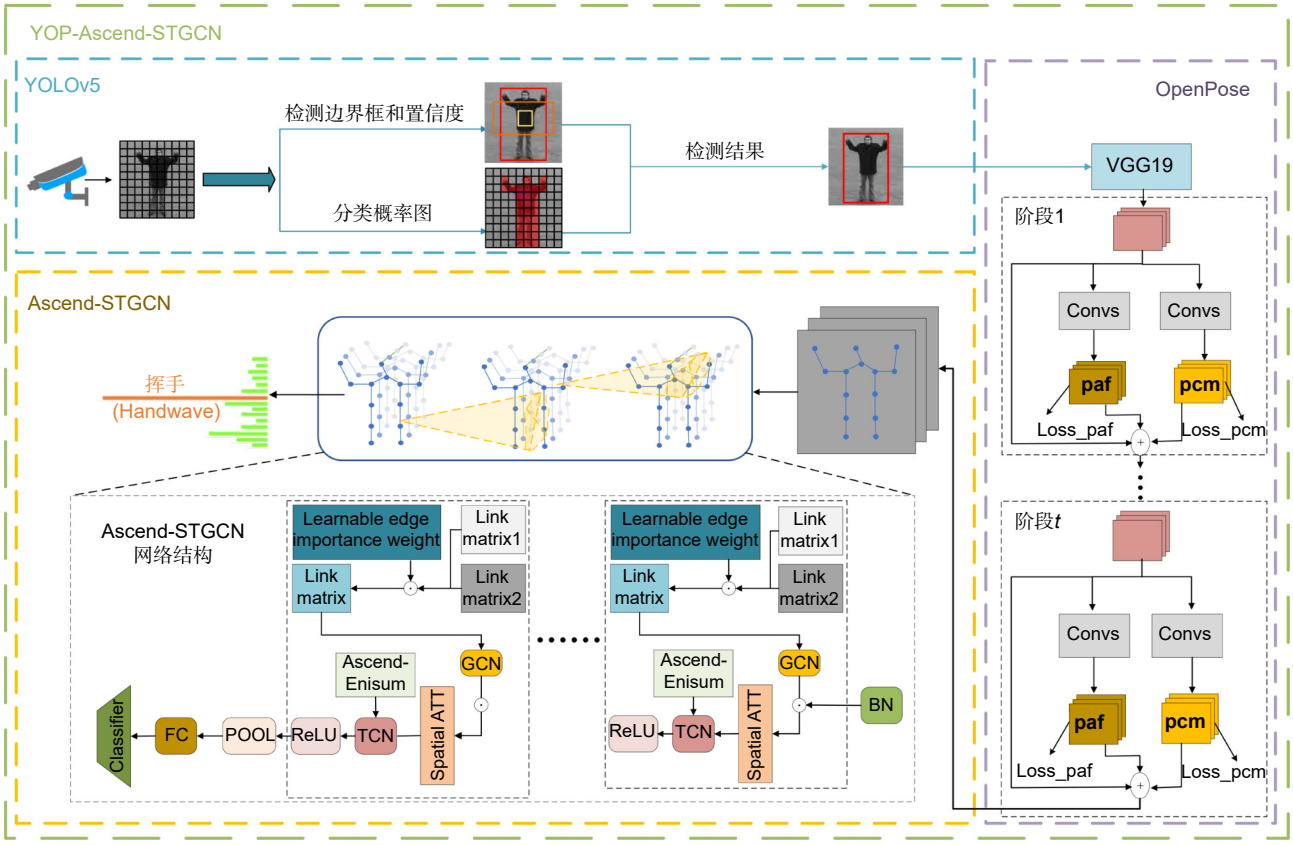


图 1　YOP-Ascend-STGCN 算法总体框架

Fig. 1　Overall framework of the YOP-Ascend-STGCN algorithm

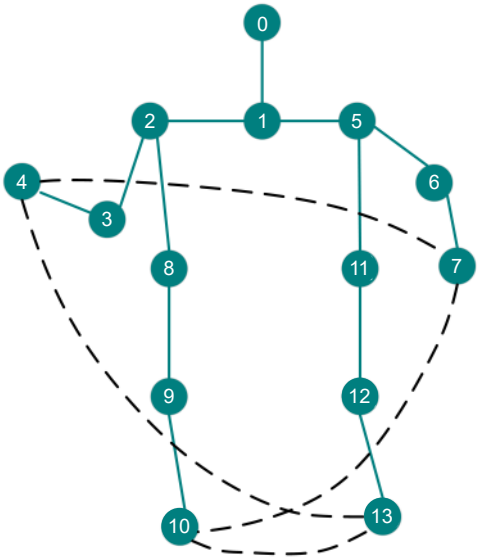


图 2　跑步动作的骨架显性关联和隐性关联结构图

Fig. 2　Skeleton explicit and implicit association structure diagram of the running action

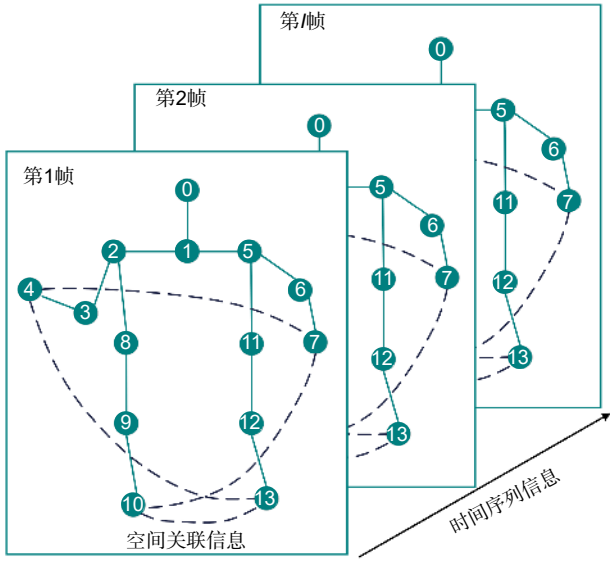


图 3　骨架时空建模图

Fig. 3　Spatial-temporal modeling diagram of the skeleton information

3.2.3 显隐性骨架关联空间图卷积模块设计

在空间图卷积[28]中，通常使用显性骨骼连接进行特征提取与卷积运算。而对于人体动作识别来说，四肢之间协作扮演了至关重要的角色，四肢为动作中的关节点相对位置和移动方式提供了丰富的信息，但是传统的空间卷积对这种隐性联系关注度不够，本文设计融合显隐性关节点联系的空间图卷积加强这方面的关注度。

假设空间显性邻接矩阵为：$\boldsymbol{A}_{\mathrm{S}} \in \mathbb{R}^{v \times v}$，隐性邻接矩阵为：$\boldsymbol{A}_{\mathrm{P}} \in \mathbb{R}^{v \times v}$，$\forall V_m, V_n \in H$存在显性联系，则$\boldsymbol{A}_{\mathrm{S}}(V_m, V_n) = \boldsymbol{A}_{\mathrm{S}}(V_n, V_m) = 1$，否则为 0；同样地，$\forall V_m, V_n \in P$存在隐性联系，则$\boldsymbol{A}_{\mathrm{P}}(V_m, V_n) = \boldsymbol{A}_{\mathrm{P}}(V_n, V_m) = 1$，其余为 0。假设显性邻域为：$B_{\mathrm{S}} = \{V_{ti} | d(V_{ti}, V_{tj}) \leqslant n\}$，存在映射函数:$L_{\mathrm{S}} : B_{\mathrm{S}}(V_{ti}) \to \{0, 1, \cdots, n-1\}$，将子图中的节点映射到子集，隐性邻域为：$B_{\mathrm{P}} = \{V_{ti} | d(V_{ti}, V_{tj}) \leqslant n\}$，存在映射函数：$L_{\mathrm{P}} : B_{\mathrm{P}}(V_{ti}) \to \{0, 1, \cdots, n-1\}$，其中 d 表示最短路径。n 为邻居节点采集范围。当 n 取 3 时，可以将显性邻域和隐性邻域分别划分为三个不同子集$Y_{\mathrm{S}} = \{a, b, c\}$和$Y_{\mathrm{P}} = \{a, b, c\}$。再通过人体运动趋势，向心和离心运动，对关节的邻居节点进行卷积计算。显性和隐性子集划分方式分别如图 4 所示，图 4(a) 表示显性划分方式，图 4(b) 表示隐性划分方式。图 4(a)、4(b) 中关节点 a 表示根节点，b 表示向心节点，c 表示离心节点，黑色×表示重心。其中图 4(b) 为了避免隐性邻接矩阵重复学习特征，例如四肢之间重复连接，设计与根节点最近的一个显性连接节点和另外两个隐性节点作为其邻域。

将显性邻接矩阵$\boldsymbol{A}_{\mathrm{S}}$和隐性邻接矩阵$\boldsymbol{A}_{\mathrm{P}}$分别分解成 3 个子邻接矩阵。即：

$$\boldsymbol{A}_{\mathrm{S}} + \boldsymbol{I}_{\mathrm{S}} = \boldsymbol{A}_{\mathrm{S}}^{a} + \boldsymbol{A}_{\mathrm{S}}^{b} + \boldsymbol{A}_{\mathrm{S}}^{c} = \sum_{y \in Y_{\mathrm{S}}} \boldsymbol{A}_{\mathrm{S}}{}^{y}, \tag{1}$$

$$\boldsymbol{A}_{\mathrm{P}}^{a} + \boldsymbol{A}_{\mathrm{P}}^{b} + \boldsymbol{A}_{\mathrm{P}}^{c} = \sum_{y \in Y_{\mathrm{P}}} \boldsymbol{A}_{\mathrm{P}}{}^{y}, \tag{2}$$

其中：$\boldsymbol{A}_{\mathrm{S}}^{a}$、$\boldsymbol{A}_{\mathrm{S}}^{b}$和$\boldsymbol{A}_{\mathrm{S}}^{c}$，表示显性邻接子矩阵，$\boldsymbol{I}_{\mathrm{S}}$为对角矩阵，表示每个节点的自环，$\boldsymbol{A}_{\mathrm{P}}^{a}$、$\boldsymbol{A}_{\mathrm{P}}^{b}$和$\boldsymbol{A}_{\mathrm{P}}^{c}$，表示隐性邻接子矩阵，依次分别代表根节点、向心运动特征节点和离心运动特征节点。

基于显性关系的图卷积表达公式为

$$\boldsymbol{F}_{\mathrm{S}} = \sum_{y \in Y_{\mathrm{S}}} \boldsymbol{M}_{\mathrm{S}}^{y} \otimes \left((\boldsymbol{D}_{\mathrm{S}}^{y})^{-\frac{1}{2}} \boldsymbol{A}_{\mathrm{S}}{}^{y} (\boldsymbol{D}_{\mathrm{S}}^{y})^{-\frac{1}{2}} \right) \boldsymbol{F}_{\mathrm{S}}(\mathrm{in}) \boldsymbol{W}_{\mathrm{S}}^{y}, \tag{3}$$

其中：$\boldsymbol{F}_{\mathrm{S}}(\mathrm{in})$表示骨架显性联系输入特征，$\boldsymbol{D}_{\mathrm{S}}^{y}$表示 y 子图对应的度矩阵，$\boldsymbol{A}_{\mathrm{S}}^{y}$表示 y 子图对应的邻接矩阵，$\otimes$表示矩阵按元素位逐个相乘，$\boldsymbol{M}_{\mathrm{S}}^{y} \in \mathbb{R}^{n \times n}$表示 y 子图对应的掩膜矩阵，$\boldsymbol{W}_{\mathrm{S}}^{y}$表示 y 子图对应的权重矩阵。$\boldsymbol{F}_{\mathrm{S}}$表示显性联系输出特征。

基于隐性关系的图卷积表达公式为

$$\boldsymbol{F}_{\mathrm{P}} = \sum_{y \in Y_{\mathrm{P}}} M_{\mathrm{P}}^{y} \otimes \left((\boldsymbol{D}_{\mathrm{P}}^{y})^{-\frac{1}{2}} \boldsymbol{A}_{\mathrm{P}}{}^{y} (\boldsymbol{D}_{\mathrm{P}}^{y})^{-\frac{1}{2}} \right) \boldsymbol{F}_{\mathrm{P}}(\mathrm{in}) \boldsymbol{W}_{\mathrm{P}}^{y}, \tag{4}$$

式中变量含义类似式 (3)，特别地，$\boldsymbol{F}_{\mathrm{P}}(\mathrm{in})$表示骨架隐性联系输入特征，$\boldsymbol{F}_{\mathrm{P}}$表示隐性联系输出特征。则融合显性联系和隐性联系的空间图卷积F_{out}可以表示为

$$\boldsymbol{F}_{\mathrm{out}} = \boldsymbol{F}_{\mathrm{S}} + \boldsymbol{F}_{\mathrm{P}}. \tag{5}$$

3.2.4 空间注意力模块及时间图卷积设计

人体动作信息在帧与帧之间存在联系，传统的时间图卷积[29]只是将连续帧的对应点进行连接，然后再进行时间图卷积操作，这样处理可能会忽略关节点

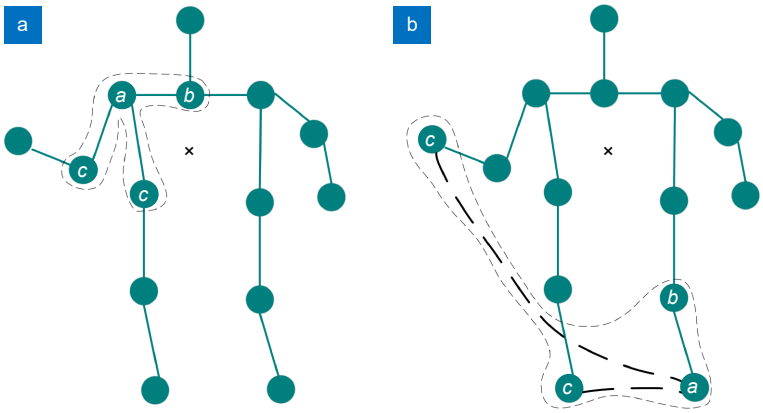


图 4　显性和隐性子集划分方式

Fig. 4　Explicit and implicit subgraph partitioning methods

的细微移动，从而无法准确捕捉不同帧中动作在空间维度上的细微差别。本文提出空间注意力机制，通过注意力权重，对包含重要动作信息的区域进行特征提取，更加关注帧与帧之间关节点位置的细微变化，然后为每个节点分配不同注意力权重使得模型更加关注动作的细微变化，从而提高对相似动作识别的精确度。空间注意力权重的公式为

$$\begin{aligned}\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{F}) &=\sigma\left(f^{7\times7}(Avgpool(\boldsymbol{F});Maxpool(\boldsymbol{F}))\right)\\ &=\sigma\left(f^{7\times7}\left(\boldsymbol{F}^{\mathrm{S}}_{\mathrm{avg}};\boldsymbol{F}^{\mathrm{S}}_{\mathrm{max}}\right)\right)\end{aligned},\quad (6)$$

其中：$\boldsymbol{F}\in\mathbb{R}^{C\times T\times V}$表示输入特征，$\sigma$表示 sigmoid 函数，$\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{F})$表示权重文件。将输入特征在通道维度上分别进行平均池化和最大池化，生成$\boldsymbol{F}_{\mathrm{avg}}\in\mathbb{R}^{C\times T\times V}$和$\boldsymbol{F}_{\mathrm{max}}\in\mathbb{R}^{C\times T\times V}$然后将其拼接，再接一个卷积网络，卷积核大小7×7，生成空间注意力。

将骨架时空建模图中时间维度上相邻的骨架图对应的骨骼点连接起来，图节点就在时间维度上构成了一个时间序列，这个时间序列反映了骨架随时间变化特征。通过设计时间图卷积可以将这些特征进行时空分析。假设关节点在时间维度构建了一个时间序列$\boldsymbol{F}_T(\mathrm{in})$，时间的卷积核大小为$\Gamma$，时间邻域：$C_T=\{V_{\mathrm{P}}||p-t|\leqslant\lfloor\Gamma/2\rfloor\}$，则时间卷积公式可以表示为

$$\boldsymbol{F}_T=\sum_{p\in\boldsymbol{C}_T}\boldsymbol{F}_T(\mathrm{in})\cdot\boldsymbol{W}_T(P),\quad (7)$$

其中：$\boldsymbol{F}_T$表示时间特征信息输出，$\boldsymbol{W}_T(p)$表示权重文件。

3.2.5 Ascend-Einsum 算子设计

昇腾 AI 处理器在边端部署深度学习模型时，需要将模型转化为其适配的 om 格式。在此过程中，有一个关键点是处理 PyTorch 框架中的 Einsum 算子[30]。Einsum 算子通常用于执行张量运算。在进行复杂运算时，需要进额外的开销，并且受限于 Numpy 库的支持范围，对嵌入式设备移植并不友好。

Einsum 算子本质上是对两个张量 $\boldsymbol{x}$ 和 $\boldsymbol{A}$ 进行融合运算的过程。其中$\boldsymbol{x}\in\mathbb{R}^{N\times K\times C\times T\times V}$是一个五维张量，其中 N 表示样本数量，K 表示空间内核大小，C 表示通道数，T 表示时间步长，V 表示关节点数，$\boldsymbol{A}\in\mathbb{R}^{K\times V\times M}$表示是邻接矩阵，将其相乘最后结果只有四个维度。这个运算过程可以用式 (8) 表示：

$$\boldsymbol{out}_{nctv}=\sum_k\sum_v\boldsymbol{x}_{nkctv}\otimes\boldsymbol{A}_{kvm}.\quad (8)$$

为了适应昇腾处理器，本文设计了 Ascend-Einsum 算子，首选将输入张量 $\boldsymbol{x}$ 和 $\boldsymbol{A}$ 进行维度调整，扩展的维度对应值用 1 补全。

$$\boldsymbol{d}_{nkctvj}=\boldsymbol{x}_{nkctv},\quad (9)$$

$$\boldsymbol{e}_{lkghvm}=\boldsymbol{A}_{kvm},\quad (10)$$

其中：n,k,c,t,v,j 分别表示张量 $\boldsymbol{d}$ 的不同维度索引值，l,k,g,h,v,m 分别表示张量 $\boldsymbol{e}$ 不同维度上的索引值。

将张量 $\boldsymbol{d}$，$\boldsymbol{e}$ 进行点逐元素相乘，得到张量 $\boldsymbol{f}$:

$$\boldsymbol{f}_{nkctvm}=\boldsymbol{d}_{nkctvj}\otimes\boldsymbol{e}_{lkghvm},\quad (11)$$

然后分别在维度 1 和 4 上进行求和：

$$\boldsymbol{out}(k)_{nctvm}=\sum\nolimits_k\boldsymbol{f}_{nkctvm},\quad (12)$$

$$\boldsymbol{newout}_{nctv}=\sum\nolimits_v\boldsymbol{out}(k)_{nctvm}.\quad (13)$$

式 (9)~(13) 替代了 Einsum 操作，以实现式 (8) 相同的功能。同时为了适应昇腾上 OpenPose 算法模块的输出，在进行动作识别时，需要将骨骼数据处理成四维形式。原始输入张量 $\boldsymbol{x}$ 的维度为 (N,K,C,T,V)，本文对维度进行转置，然后将 K 和 V 维度合并，从而使张量的维度转换为 (N,C,T,V)，在保持数据完整性的同时为后续卷积处理减少计算复杂度。

3.3 Ascend-STGCN 网络结构设计

上述工作完成了 Ascend-STGCN 网络的主要模块设计包括显隐性邻接矩阵、时空图卷积、空间注意力，Ascend-Enisum 算子。图 5 是本文的 Ascend-STGCN 网络结构，网络一共由三个部分组成，数据输入层，主干网络，动作分类层。

数据输入层：因为 Ascend-STGCN 在不同节点上共享权重文件，需要对输入数据进行批量标准化。

主干网络：主干网络由 9 层时空图卷积组成，其中，每一层主要包括有可学习权重文件、显隐性邻接矩阵 (Link matrix1 和 Link matrix2)、GCN、TCN 以及空间注意力和 Ascend-Enisum 算子。空间图卷积中显性和隐性邻接矩阵通过可学习的权重文件与设计的邻接矩阵进行矩阵运算得到，然后将输出特征加上空间注意力，再将其输入到时间图卷积中，最后通过 ReLU 进行激活。

动作分类层：包含池化、全连接和分类器，将主干网络的输出结果进行全局池化和全连接，最后通过分类器输出识别结果。

4 基于昇腾处理器的边端人体动作识别系统设计

4.1 昇腾边端硬件系统设计

本系统的硬件设计如图 6 所示，基于昇腾处理器的人体动作识别控制主板采用了昇腾 310[31] 芯片，集成 DaVinci AI Core 和 ARM Cortex-A55 核心，搭载于 Atlas 200 AI 加速模块。内置视频处理子系统 (DVPP) 高效处理视频流，进行智能分析和决策。主板设计多种关键功能和接口，如 SD 卡、SSD 模块提供灵活存储，千兆以太网接口确保可靠网络连接，USB3.0 接口支持外部设备扩展，满足各种应用需求，确保系统稳定、可维护和高效性。

4.2 昇腾边端软件系统设计

本系统的软件设计流程如图 7 所示，首先视频元数据输入到系统中，通过 OpenCV 读取视频帧并进行预处理操作，接着将数据输入到昇腾处理器，使用 YOLOv5 识别视频帧中的行人，将行人数据输入到 OpenPose 提取骨架，再利用 Ascend-STGCN 进行动作推理，最后将推理结果保存。

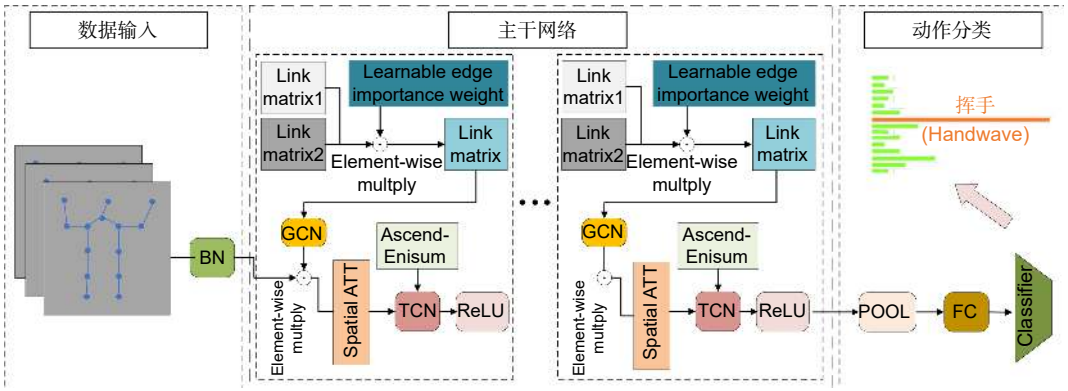


图 5 Ascend-STGCN 网络结构图

Fig. 5 Ascend-STGCN network structure diagram

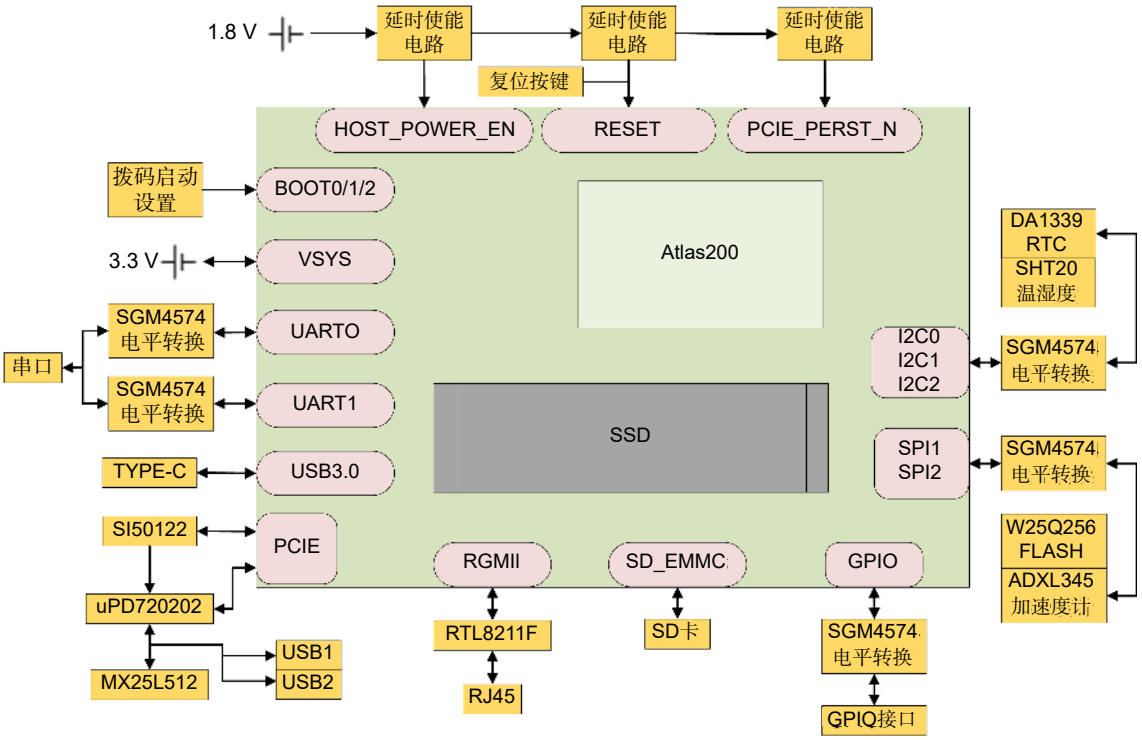


图 6 系统硬件结构设计

Fig. 6 System hardware structure

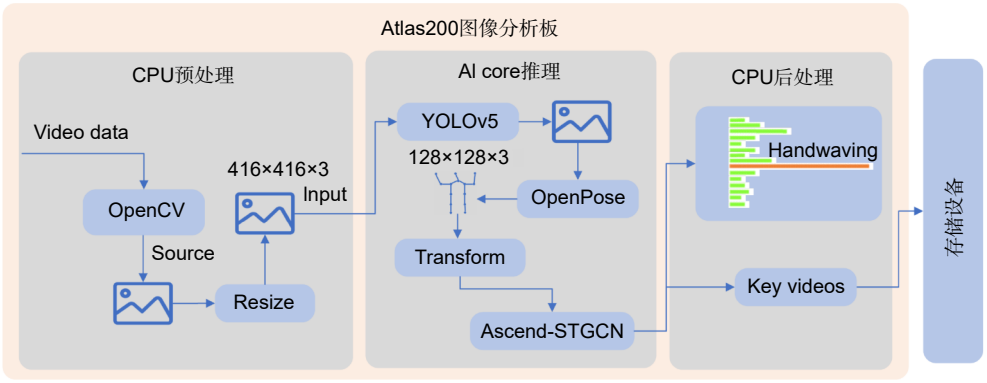


图 7 系统软件流程设计

Fig. 7 System software process

5 实 验

5.1 实验环境

实验训练环境为：Inter(R) Core(TM) Gold i9-10900x CPU@2.30 GHz 3.70 GHz，64 G 运行内存，NVIDIA GeForce RTX 2080ti，Linux，64 位操作系统，Pytorch 深度学习框架。编程语言为 Python，图形处理器加速软件为 CUDA11.0。

5.2 数据集和评价指标

为了验证改进模型的有效性，以及在嵌入式设备上部署的可行性，本文基于 KTH 视频数据[32-33] 进行自建 kinetics-skeleton 格式数据进行对比实验。KTH 数据集包含 6 类人体行为：行走、慢跑、奔跑、拳击、挥手和鼓掌，每类行为由 25 个人在四种不同的场景(室外、伴有尺度变化的室外、伴有衣着变化的室外、室内) 执行多次，相机固定。该数据库共有 2391 个视频样本。视频帧率为 25 f/s，分辨率为 160×120。数据集以 6:2:2 的比例划分为训练集、验证集和测试集。每完成 5 次训练就对模型进行 1 次验证，每完成 5 次训练对模型进行 1 次测试。实验一共进行 50 次训练。最终通过 Softmax 函数对特征进行分类，将最后测试集上的识别准确率 Top-1 作为评价指标。对于测试样本中，模型会对输出的样本进行预测，并输出一个概率分布，表示各个类别的可能性。Top-1 概率就是取这个概率分布中概率最大的值，其对应的类别作为模型的输出类别，如公式 (14) 所示。

$$Top\ 1 = \frac{\sum_{i}^{N} \delta\left(class_i^{\text{real}} = rank\left(class_i^{\text{pred}}\right)\right)}{N}, \tag{14}$$

其中：N 表示样本总数；δ表示判断函数，如果满足条件$\delta = 1$，否则$\delta = 0$；$class_i^{\text{real}}$代表第 i 个样本的真实类别；$rank(class_i^{\text{pred}})$表示第 i 个样本预测概率排名第一的类别。

5.3 数据处理与训练过程

实验通过将 KTH 视频数据裁剪成 5~8 s 的视频片段，创建一个自定义 kinetics-skeleton 格式的行为识别骨骼数据集。首先对裁剪后的视频使用 ffmpeg 进行调整至 128 pixel×128 pixel 的大小，30 f/s 的帧率，然后通过 OpenPose 算法提取视频人体 14 个骨架点的位置信息和置信度。由于 Ascend-STGCN 在不同的节点共享权重文件，所以在不同的节点上保持输入数据规模一致性十分重要，因此需要 OpenPose 算法提取关节点 2 维坐标进行归一化，使得坐标数值范围在[0,1]，这样将每个动作的 RGB 视频处理成骨架点数据的 JSON 文件，将多个 JSON 文件整合在一起，数据集为.npy 文件，动作标签为.pkl 文件。数据集的训练轮次均为 50，考虑到训练所用 GPU 显存的限制，batch size 设为 16，优化策略使用随机梯度下降法，初始学习率设置为 0.01，依次第 10、20、30、40 个轮次衰减 0.1 倍。

5.4 结果分析

5.4.1 对比实验

本节选取经典的单流网络 ST-GCN[34] 和经典的双流网络 AS-GCN[35]、ST-TR[36] 以及热图堆叠网络 PoseConv3D[37] 进行对比实验，分别通过精度，参数量、计算量来评估人体动作识别算法的性能。

将训练集一次性送入网络，并在验证集上进行验证，最后计算整个测试集的动作识别准确率。通过计算一个完整输入、输出时模型所需要的参数量和计算

量来评估模型在嵌入式设备上的性能优劣。本方法在 KTH 数据集上与其他方法的比较结果如表 1 所示。与单流网络 ST-GCN 相比，本方法在识别精确度上提升了 5%，在参数量上减少了 18.26%，在计算量减少了 22.28%。与双流网络 AS-GCN 相比，本方法在识别精确度上提升了 1.09%，在参数量上减少了 57%，在计算量上减少了 57.25%。与 ST-TR 算法相比，ST-TR 算法精度略低于本文的方法，并且参数量也是本文的 3.8 倍，计算量是本文的 3 倍，与 PoseConv3D 方法进行比较，虽然 PoseConv3D 算法精度略高于本文的算法，但是其参数量比本文高 7.3%，而计算量更是本文算法的 3.52 倍。从实验结果上可以看出本文提出的方法，在动作识别整体性能上有较为明显的优势，对于移植到边端设备是更合适的选择。

表 1　KTH 数据集上的对比试验

Table 1　Comparative experiments on the KTH dataset

Algorithm	Top-1/%	Params/M	Flops/G
ST-GCN	79.17	3.68	5.88
AS-GCN	83.08	7.00	10.69
ST-TR	83.83	11.51	13.88
PoseConv3D	**85.67**	3.23	16.1
This paper	84.17	**3.01**	**4.57**

针对每个动作类别，本文进一步进行了实验，比较了 Ascend-STGCN 与 ST-GCN 和 AS-GCN 三种算法，在测试集中每个动作识别精度，结果如图 8 所示。

从上述图表中，可以看出，本文的算法在大多数动作的检测精度优于其他两种算法。尤其是在识别类似动作方面，如识别 jog、run、walk 这三个动作，本文算法优势明显。从而验证了本文提出的空间注意力机制的有效性，它能够学习到相似动作的相似特征，使算法更加关注局部特征细微变化，在区分相似动作时有较大的优势。

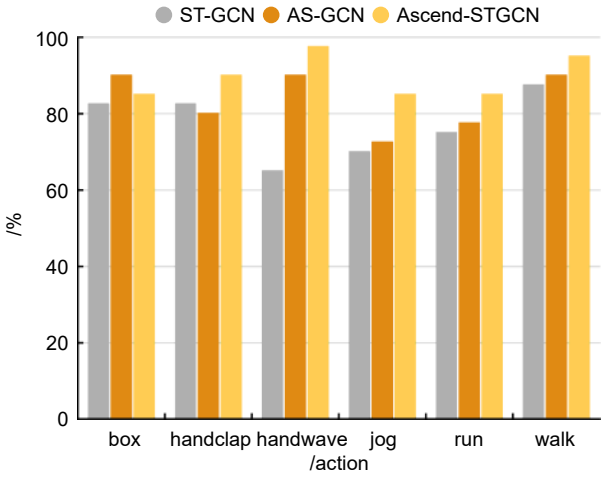


图 8　每种动作对比数据直方图

Fig. 8　Histogram of the comparison data for each action

为了验证算法的真实性可行性，本实验分别将六种动作视频输入到网络中得到输出结果。图 9 展示了拳击、鼓掌、挥手、快走、跑步、行走这六个动作推理结果，其中图 9(a-f) 每张图由四个部分组成，第一

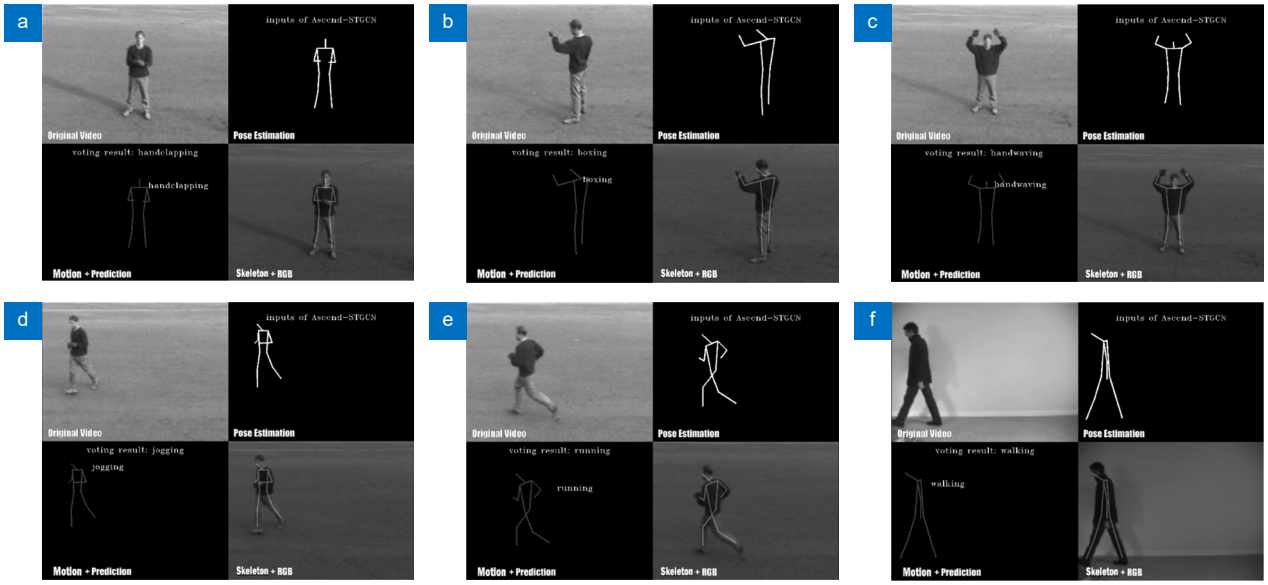


图 9　每种动作推理结果截图

Fig. 9　Screenshots of inference results for each action

个是原始视频；第二个是 OpenPose 提取到的骨架信息；第三个是模型的预测动作，该部分每一帧都会推理一个动作标签，以出现动作次数最多的标签作为推理结果输出；第四个是骨架信息加上原始视频的 RGB 格式。

5.4.2 消融实验

本文 Ascend-STGCN 模型主要由显性关系空间图卷积 (EC)、隐性关系空间图卷积 (IC)、时间图卷积 (TC)、空间注意力 (SA) 以及 Ascend-Enisum 算子 (AE) 五个模块组成。为了说明这些模块融合的有效性，在 KTH 数据集上进行了消融实验。ST-GCN 是经典的人体骨架时空图卷积，采用了 EC+TC 模式，因此将其作为对比模型的基准。

从表 2 可以看出在加入了 Ascend-Enisum 算子、隐性联系图卷积和空间注意力后模型的精度分别提高了 1.67%、3.33% 和 4.16%，证明单独加入这三个模块是有效的。并且在 50 个轮次里，加入了 AE 模块，可以减少模型参数量，同时计算减少了 22.28%，Ascend-STGCN 模型相比于 ST-GCN 精度提升了 5%，证明了融合这五个模块使动作识别效果更佳。

5.5 昇腾处理器上部署实验

本文基于昇腾处理器搭建了人体动作识别系统。将昇腾处理器搭载在摄像头里面，如图 10 所示，图 10(a) 为摄像头外部结构，图 10(b) 是昇腾处理器搭载在摄像头上实物图。

如图 11 所示，将 YOP-Ascend-STGCN 算法部署到搭载昇腾处理器的摄像头上进行实验验证。通过模拟两个人体动作视频，在边端设备上对本文提出的算法进行了实验，图 11(a) 是鼓掌动作实验，图 11(b) 是挥手动作实验。实验表明，在背景较为复杂的条件下，系统也能够进行正确的动作识别任务，并且每秒可以处理 26 帧数据。

6 结　论

本文针对现有的人体动作识别算法精度不足、计算量大、缺少在边端设备上的部署与应用等问题，设计了一种基于昇腾处理器的边端轻量化人体动作识别时空图卷积算法。提出了基于隐性联系的骨架连接方法并构建隐性邻接矩阵；针对不同帧之间关节点位置信息关注不足，提出了空间注意力，加强对不同帧之间骨骼点位置空间特征的关注度；针对 Enisum 算子计算量大且不适配昇腾处理器的问题，设计了 Ascend-Enisum 算子，减小了计算量，使模型轻量化。最后将训练好的模型转换成昇腾适配的格式并结合目

表 2　KTH 数据集上的消融实验

Table 2　Ablation experiments on the KTH dataset

Algorithm	Top-1/%	Params/M	FLOPs/G	Epochs/轮
ST-GCN (EC+TC)	79.17	3.68	5.88	50
EC+TC+AE	81.67	2.93	4.57	50
EC+TC+IC	82.50	3.79	5.88	50
EC+TC+SA	83.33	3.68	5.88	50
Ours (EC+IC+TC+AE+SA)	84.17	3.01	4.57	50

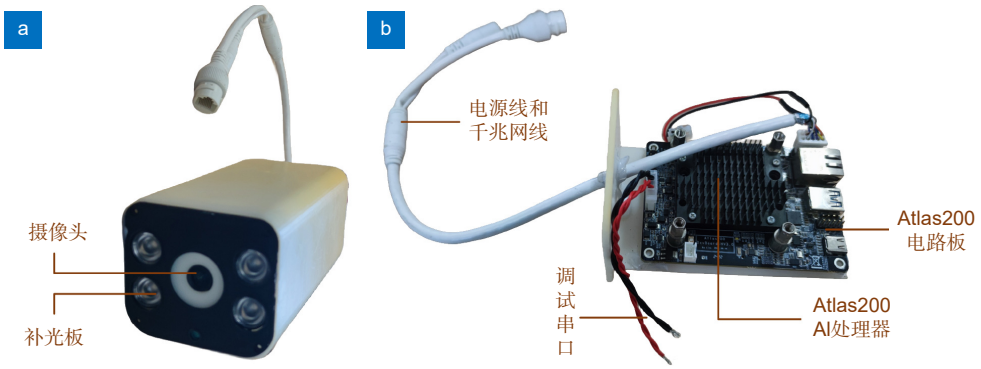


图 10　装置结构图

Fig. 10　Device structure diagram

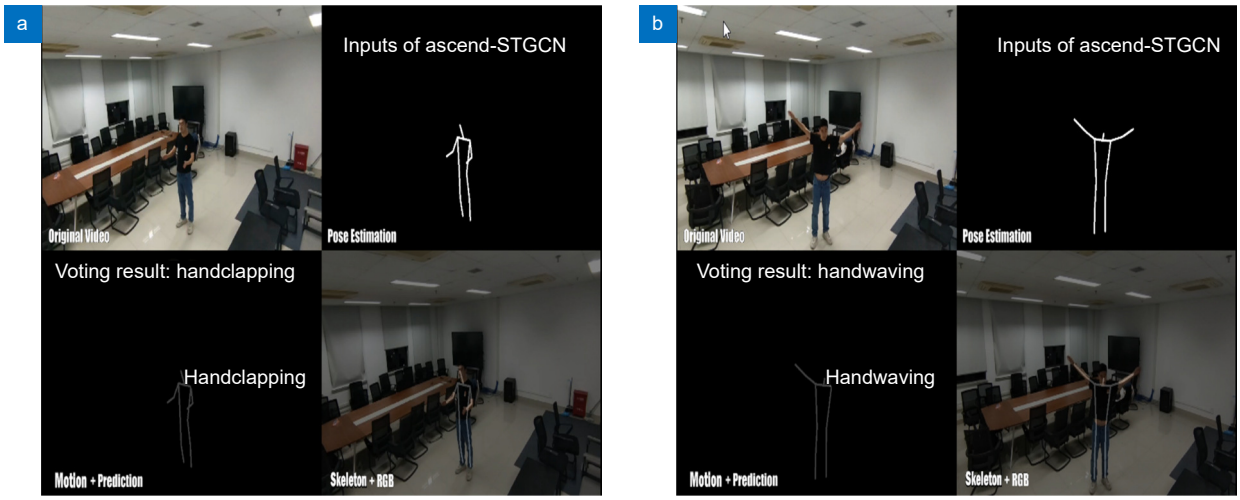


图 11 模拟动作的推理结果截图

Fig. 11 Screenshot of inference results for simulated actions

标检测 YOLOv5 算法和姿态估计 OpenPose 算法，设计 YOP-Ascend-STGCN 算法，将其部署到昇腾处理器上，实现了端到端人体动作检测。实验证明，本文提出的方法在识别类似动作时有较高的精度，可以应用到实时人体动作识别系统中。

参考文献

[1] Li Y D, Xu X P. Human action recognition by decision-making level fusion based on spatial-temporal features[J]. *Acta Opt Sin*, 2018, **38**(8): 0810001.
李艳荻, 徐熙平. 基于空-时域特征决策级融合的人体行为识别算法[J]. 光学学报, 2018, **38**(8): 0810001.

[2] Sun Z H, Ke Q H, Rahmani H, et al. Human action recognition from various data modalities: a review[J]. *IEEE Trans Pattern Anal Mach Intell*, 2023, **45**(3): 3200−3225.

[3] Chen C, Liu K, Kehtarnavaz N. Real-time human action recognition based on depth motion maps[J]. *J Real-Time Image Process*, 2016, **12**(1): 155−163.

[4] Li C K, Hou Y H, Wang P C, et al. Joint distance maps based action recognition with convolutional neural networks[J]. *IEEE Signal Process Lett*, 2017, **24**(5): 624−628.

[5] Kumar S S, John M. Human activity recognition using optical flow based feature set[C]//*2016 IEEE International Carnahan Conference on Security Technology* (*ICCST*), Orlando, 2016: 1–5.
https://doi.org/10.1109/CCST.2016.7815694.

[6] Leng J X, Tan M P, Hu B, et al. Video anomaly detection based on implicit view transformation[J]. *Comput Sci*, 2022, **49**(2): 142−148.
冷佳旭, 谭明圮, 胡波, 等. 基于隐式视角转换的视频异常检测[J]. 计算机科学, 2022, **49**(2): 142−148.

[7] Li G Y, Li C G, Wang W J, et al. Research on multi-feature human pose model recognition based on one-shot learning[J]. *Opto-Electron Eng*, 2021, **48**(2): 200099.
李国友, 李晨光, 王维江, 等. 基于单样本学习的多特征人体姿态模型识别研究[J]. 光电工程, 2021, **48**(2): 200099.

[8] Wu H B, Ma X, Li Y B. Spatiotemporal multimodal learning with 3D CNNs for video action recognition[J]. *IEEE Trans Circuits Syst Video Technol*, 2022, **32**(3): 1250−1261.

[9] Liu Z Y, Zhang H W, Chen Z H, et al. Disentangling and unifying graph convolutions for skeleton-based action recognition[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2020: 140–149.
https://doi.org/10.1109/CVPR42600.2020.00022.

[10] Cheng K, Zhang Y F, He X Y, et al. Skeleton-based action recognition with shift graph convolutional network[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2020: 180–189.
https://doi.org/10.1109/CVPR42600.2020.00026.

[11] Xie Y, Yang R L, Liu G X, et al. Human skeleton action recognition algorithm based on dynamic topological graph[J]. *Comput Sci*, 2022, **49**(2): 62−68.
解宇, 杨瑞玲, 刘公绪, 等. 基于动态拓扑图的人体骨架动作识别算法[J]. 计算机科学, 2022, **49**(2): 62−68.

[12] Liu B L, Zhou S, Dong J F, et al. Research progress in skeleton-based human action recognition[J]. *J Comput-Aided Des Comput Graphics*, 2023, **35**(9): 1299−1322.
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.

[13] Law H, Deng J. CornerNet: detecting objects as paired keypoints[C]//*Proceedings of the 15th European Conference on Computer Vision* (*ECCV*), Munich, 2018: 765–781.
https://doi.org/10.1007/978-3-030-01264-9_45.

[14] Duan K W, Bai S, Xie L X, et al. CenterNet: keypoint triplets for object detection[C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 2019: 6568–6577.
https://doi.org/10.1109/ICCV.2019.00667.

[15] Zhu C C, He Y H, Savvides M. Feature selective anchor-free module for single-shot object detection[C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019: 840–849.
https://doi.org/10.1109/CVPR.2019.00093.

[16] Zhu C C, Chen F Y, Shen Z Q, et al. Soft anchor-point object detection[C]//*Proceedings of the 16th European Conference*

on Computer Vision, Glasgow, 2020: 91–107. https://doi.org/10.1007/978-3-030-58545-7_6.

[17] Ma L, Gou Y T, Lei T, et al. Small object detection based on multi-scale feature fusion using remote sensing images[J]. *Opto-Electron Eng*, 2022, **49**(4): 210363.
马梁, 苟于涛, 雷涛, 等. 基于多尺度特征融合的遥感图像小目标检测[J]. 光电工程, 2022, **49**(4): 210363.

[18] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 2014: 580–587. https://doi.org/10.1109/CVPR.2014.81.

[19] Girshick R. Fast R-CNN[C]//*Proceedings of the 2015 IEEE International Conference on Computer Vision*, Santiago, 2015: 1440–1448. https://doi.org/10.1109/ICCV.2015.169.

[20] Peng H, Wang W Q, Chen L, et al. Few-shot object detection via online inferential calibration[J]. *Opto-Electron Eng*, 2023, **50**(1): 220180.
彭昊, 王婉祺, 陈龙, 等. 在线推断校准的小样本目标检测[J]. 光电工程, 2023, **50**(1): 220180.

[21] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, Amsterdam, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

[22] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, 2017: 2999–3007. https://doi.org/10.1109/ICCV.2017.324.

[23] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, **49**(3): 210372.
陈旭, 彭冬亮, 谷雨. 基于改进 YOLOv5s 的无人机图像实时目标检测[J]. 光电工程, 2022, **49**(3): 210372.

[24] Zhao D D, Xie D H, Chen P, et al. Lightweight YOLOv5 sonar image object detection algorithm and implementation based on ZYNQ[J]. *Opto-Electron Eng*, 2024, **51**(1): 230284.
赵冬冬, 谢墩翰, 陈朋, 等. 基于 ZYNQ 的轻量化 YOLOv5 声呐图像目标检测算法及实现[J]. 光电工程, 2024, **51**(1): 230284.

[25] Fang H S, Xie S Q, Tai Y W, et al. RMPE: regional multi-person pose estimation[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, 2017: 2353–2362. https://doi.org/10.1109/ICCV.2017.256.

[26] Debnath B, O'Brien M, Yamaguchi M, et al. Adapting MobileNets for mobile based upper body pose estimation[C]//*2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance* (*AVSS*), Auckland, 2018: 1–6. https://doi.org/10.1109/AVSS.2018.8639378.

[27] Qiao S, Wang Y L, Li J. Real-time human gesture grading based on OpenPose[C]//*2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* (*CISP-BMEI*), Shanghai, 2017: 1–6. https://doi.org/10.1109/CISP-BMEI.2017.8301910.

[28] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[C]//*Proceedings of the 5th International Conference on Learning Representations*, Toulon, 2017.

[29] Zhao L, Song Y J, Zhang C, et al. T-GCN: a temporal graph convolutional network for traffic prediction[J]. *IEEE Trans Intell Transp Syst*, 2020, **21**(9): 3848−3858.

[30] Klaus J, Blacher M, Giesen J. Compiling tensor expressions into einsum[C]//*Proceedings of the 23rd International Conference on Computational Science*, Prague, 2023: 129–136. https://doi.org/10.1007/978-3-031-36021-3_10.

[31] Zhao D D, Zhou H C, Chen P, et al. Design of forward-looking sonar system for real-time image segmentation with light multiscale attention net[J]. *IEEE Trans Instrum Meas*, 2024, **73**: 4501217.

[32] He Y C, Zhan Y G, Xu G Q, et al. Progress in the application of OpenPose technology for human pose estimation in rehabilitation[J]. *Chin J Rehabil*, 2023, **38**(7): 437−441.
何英春, 詹益镐, 许桂清, 等. 人体姿势估计 OpenPose 技术在康复领域的应用进展[J]. 中国康复, 2023, **38**(7): 437−441.

[33] Qi Y M, Chen S Y, Sun L. Two-stream CNN fall detection based on improved ViBe algorithm[J]. *Comput Eng Des*, 2023, **44**(6): 1812−1819.
戚亚明, 陈树越, 孙磊. 基于改进 ViBe 算法的双流 CNN 跌倒检测[J]. 计算机工程与设计, 2023, **44**(6): 1812−1819.

[34] Yan S J, Xiong Y J, Lin D H. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]//*Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, New Orleans, 2018: 912.

[35] Li M S, Chen S H, Chen X, et al. Actional-structural graph convolutional networks for skeleton-based action recognition[C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019: 3590–3598. https://doi.org/10.1109/CVPR.2019.00371.

[36] Plizzari C, Cannici M, Matteucci M. Skeleton-based action recognition via spatial and temporal transformer networks[J]. *Comput Vision Image Underst*, 2021, **208–209**: 103219. https://doi.org/10.1016/j.cviu.2021.103219.

[37] Duan H D, Zhao Y, Chen K, et al. Revisiting skeleton-based action recognition[C]//*Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 2022: 2959–2968. https://doi.org/10.1109/CVPR52688.2022.00298.

作者简介

赵冬冬 (1990-)，男，博士，副教授，博士生导师，主要研究方向为图像以及信号处理。

E-mail：zhaodd@zjut.edu.cn

赖亮 (1997-)，男，硕士研究生，主要研究方向为人体动作识别、嵌入式系统程序设计。

E-mail：2112112110@zjut.edu.cn

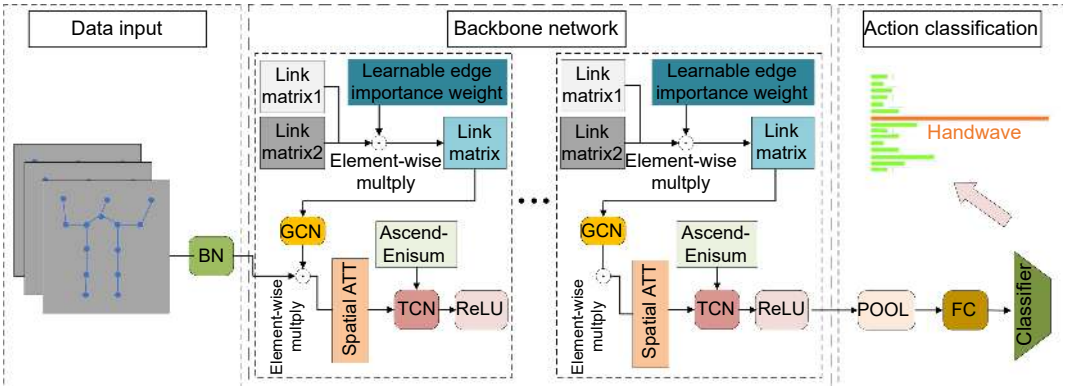
【通信作者】陈朋 (1981-)，男，博士，教授，博士生导师，主要研究方向为图像处理、嵌入式系统设计。

E-mail: chenpeng@zjut.edu.cn

扫描二维码，获取PDF全文

Design and implementation of edge-based human action recognition algorithm based on ascend processor

Zhao Dongdong, Lai Liang, Chen Peng*, Zhou Hongchao, Li Yiran, Liang Ronghua



Ascend-STGCN network structure diagram

Overview: In practical applications, human action recognition has always been an immensely challenging task in the visual domain, demanding not only the system's ability to comprehend and categorize a wide range of human movements but also to perform real-time analysis and provide instant feedback. Presently, most human action recognition algorithms are deployed in the cloud, which poses issues such as network dependency and computational resource wastage compared to edge-based deployment. Therefore, deploying human action recognition on resource-constrained edge devices has emerged as a significant research focus. Addressing the challenges of insufficient accuracy, high computational complexity, and limited deployment on edge devices in current human action recognition algorithms, this paper presents a lightweight spatio-temporal graph convolutional algorithm for human action recognition optimized for the Ascend processor. It introduces an implicit connection-based skeleton approach and constructs an implicit adjacency matrix, which is combined with the explicit skeleton connection adjacency matrix to create a fused spatial graph convolution encompassing both explicit and implicit connections. Meanwhile, to overcome inadequate attention to joint spatial position information across frames, spatial attention is introduced in the temporal dimension, enhancing the focus on spatial features of skeletal points between frames. Furthermore, temporal graph convolution is designed and integrated with spatial graph convolution to form a spatio-temporal graph convolution. To address the computational intensity and incompatibility issues of the Enisum operator with the Ascend processor, an Ascend-Enisum operator is devised, optimizing the computational load and facilitating model lightweighting. The trained model, converted into Ascend-compatible format, is integrated with the YOLOv5 object detection algorithm and the OpenPose pose estimation algorithm to develop an end-to-end YOP-Ascend-STGCN human action recognition system. Experimental deployment on cameras equipped with Ascend processors demonstrates the high accuracy of the proposed method, suitable for real-time human action recognition in edge devices.

Zhao D D, Lai L, Chen P, et al. Design and implementation of edge-based human action recognition algorithm based on ascend processor[J]. *Opto-Electron Eng*, 2024, **51**(6): 240072; DOI: 10.12086/oee.2024.240072

Foundation item: Project supported by National Natural Science Foundation of China (62371421), the Leading Innovation Team of the Zhejiang Province (2021R01002), and the Zhejiang Provincial Natural Science Foundation of China (LD24F020005)

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China
* E-mail: chenpeng@zjut.edu.cn