

DOI:10.19651/j.cnki.emt.2211527

# 基于图卷积网络改进的人体动作识别模型

陶峰 李燕萍 王瑞

(上海大学通信与信息工程学院 上海 200444)

**摘要:** 针对双流自适应图卷积动作识别网络 2S-AGCN 模型忽略了人体动作识别中特征的长距离信息以及通道之间的依赖的缺点,设计了一种双重注意力机制对 2S-AGCN 模型的图卷积模块进行改进,实现精度的提升。双重注意力机制包含了空间注意力机制以及通道注意力机制,其中空间注意力机制有选择性地聚集上下文,通道注意力机制分为两个并行的模块,第 1 部分提高了特征的可辨性,第 2 部分在捕获特征远程依赖的同时,保留了精准的位置信息。结果表明,以双流自适应图卷积动作识别网络 2S-AGCN 模型为基础网络,融入了双重注意力机制模块的模型在 Kinetics 数据集上的 Top-1 和 Top-5 分别提升了 0.6% 和 1.3%,在 NTURGB+D120 数据集的 CS 和 CV 上的 Top-1 分别提升了 1.2% 和 0.5%,在 NTURGB+D 数据集的 CS 和 CV 上的 Top-1 分别提升了 0.2% 和 0.1%。

**关键词:** 动作识别;双流自适应图卷积网络;双重注意力机制;深度学习

**中图分类号:** TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

## Improved human action recognition model based on graph convolutional networks

Tao Feng Li Yanping Wang Rui

(College of Communication and Information Engineering, Shanghai University, Shanghai 200444, China)

**Abstract:** In view of the shortcomings that the 2S-AGCN model of the two-stream adaptive graph convolutional network ignores the long-distance information of features and channel dependence in human motion recognition, a dual attention mechanism is designed to improve the graph convolution module of the 2S-AGCN model to improve the accuracy. The dual attention mechanism includes the spatial attention mechanism and the channel attention mechanism. The spatial attention mechanism selectively focuses on the context. The channel attention mechanism is divided into two parallel modules. The first part improves the distinguishability of features. The second part preserves accurate location information while capturing the remote dependency of features. The results show that the model based on the two-stream adaptive graph convolutional networks 2S-AGCN, which incorporates the dual attention mechanism module, has improved Top-1 and Top-5 on the Kinetics dataset by 0.6 and 1.3 percentage points respectively, Top-1 on the CS and CV of NTURGB+D120 dataset by 1.2 and 0.5 percentage points respectively, and Top-1 on the CS and CV of NTURGB+D dataset by 0.2 and 0.1 percentage points respectively.

**Keywords:** action recognition; two-stream adaptive graph convolutional networks; dual attention mechanism; deep learning

## 0 引言

近年来,随着深度学习<sup>[1]</sup>的发展,图像处理算法逐渐被应用于人们生活之中并极大的方便了人们的日常生活。其中人体动作识别<sup>[2-3]</sup>也是研究人员们重点研究的领域之一。人体动作识别有着丰富的应用场景,可应用于对地铁、机场中的人员越界的监控、对老人和孩子等特定人群的监控以及人机智能交互、虚拟现实和视频监控等领域。人类的动

作为可以从多种模式中被识别出来,如外观、深度、光流以及身体骨架等。传统的方法一般通过梯度直方图,时空点检测,以及光流直方图等来提取图像和时序的特征表示。

从实现方式来分,可以把人体动作识别分为基于传感器和视觉<sup>[4]</sup>两种,基于传感器的动作识别需要佩戴相应的传感器,使用不够灵活、操作复杂。基于视觉的动作识别又可以分为基于单帧图像和视频的识别。基于单帧图像的识别不能有效地获取行为的连贯时间信息,而基于视频的动

作识别能够很好地获取视频中的空间和时间信息,准确率得到提升<sup>[5]</sup>。

目前基于视频的动作识别使用深度学习为主。基于深度学习的动作识别网络结构可以分为两类,第 1 类使用 3D 卷积网络结构对图像的多个通道分别进行卷积<sup>[6]</sup>,将多个通道的信息进行组合,或者通过获取不同时间序列的短时信息并将其叠加来获取长时信息<sup>[7]</sup>。第 2 类使用图卷积网络结构<sup>[8]</sup>,通过预处理工具对图像进行预处理,将视频中的人体关节标注并连接形成骨架图,将骨架图输入图卷积网络进行特征提取分类。

目前,图卷积神经网络被广泛应用于基于关节数据的动作识别。早期应用了图卷积网络的是 Yan 等<sup>[9]</sup>提出的 ST-GCN(spatial temporal graph convolutional networks)网络模型,这是一种端到端卷积共现特征学习框架。特征的学习采用分层的方法,将不同层次的语境信息逐渐聚集在一起,对每个关节的信息进行独立编码,然后将它们组合成空间和时间域的语义表示。

在此基础上,Shi 等<sup>[10]</sup>提出了 2S-AGCN(two-stream adaptive graph convolutional networks)网络模型,这是一种双流自适应图卷积网络,可以同时处理骨骼数据和关节数据进行建模,增加了骨骼数据构建模型的灵活性,以适应各种数据样本。除此之外 Si 等<sup>[11]</sup>提出了一种注意力增强图卷积 LSTM(long-short term memory)网络模型,用来捕获动态时间的不同特征,提取时间特征信息。

针对 2S-AGCN 图卷积网络没有完全利用特征的长距离信息,上下文以及通道之间的依赖关系等问题,本文在基准网络 2S-AGCN 的基础上,设计了一种双重注意力机制模块并融入基准网络的图卷积模块。改进后的图卷积模块在多个数据集上能得到一定的动作识别精确度提升。

### 1 基准模型结构分析

动作识别网络 2S-AGCN<sup>[10]</sup>模型以图卷积网络结构为基础,如图 1 所示,通过多个图卷积模块(BLOCK)堆叠而成,通过批标准化层(batch normalization, BN)标准化输入数据,每个图卷积模块的 3 个数字分别代表输入通道数、输出通道数和步幅,通过全局平均池化层(global average pooling, GAP)将不同样本的特征映射池化为相同大小。随着模块的递增,模块的输出通道数逐渐递增,分别为 64 层、128 层、256 层,通过 Softmax 分类器进行分类预测。

输入数据包含骨骼数据和关节数据两种,如图 2 所示,给定一个样本,首先根据关节的数据计算骨骼的数据,然后将关节数据和骨骼数据分别输入 J 流和 B 流,将两个流的 Softmax 分数相加,得到融合分数并预测动作标签。由于骨骼数据的图形没有循环,因此可以为每个骨骼指定一个唯一的目标关节。这样,骨骼的图形和网络都可以设计成与关节相同的图形和网络,因为每个骨骼都可以绑定到一个唯一的关节。

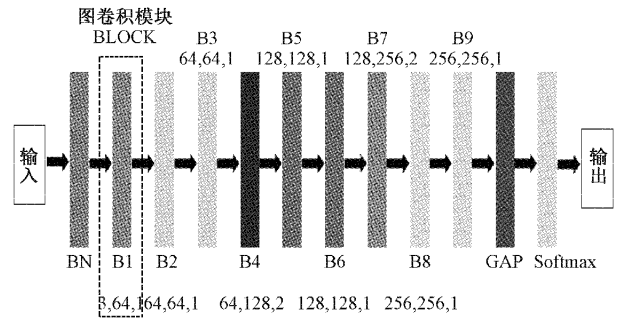


图 1 2S-AGCN 网络结构

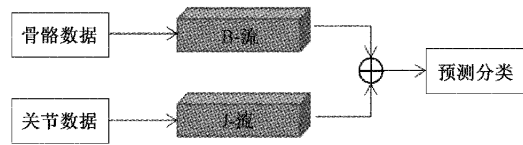


图 2 2S-AGCN 双流结构

关节数据的获取是通过 OpenPose<sup>[12]</sup>网络测量得到。该网络是基于卷积神经网络和监督学习开发的开源库,可以实现人体动作、人脸情绪等姿态估计。采用自上而下的姿态关键点检测算法,网络使用 VGG19 卷积网络提取特征,使用 CNN 网络提取置信度和关联度,通过匹配求出相关性并将同一个人的关节点连接起来。根据关节点的数量,会生成相应通道数量的热图,每一个通道为某个关节的热图分布。

骨骼数据基于关节数据获取得到,每个骨骼都由两个关节绑定,因此定义靠近骨骼重心的关节是源关节,远离重心的关节是目标关节。

视频中一般包含多帧连续的时间序列,每一帧都有一组关节点坐标以及相应的骨骼数据。关节点坐标包含了横坐标,纵坐标,置信度。

### 2 算法设计

本文主要针对基准网络 2S-AGCN 设计了一种双重注意力机制。注意力机制可以分为多种,如软注意力机制,强注意力机制,自注意力机制等。不同的注意力机制侧重的关注点不尽相同。软注意力关注区域和通道,强注意力关注图像中点的延伸与动态变化,自注意力可以提取输入的全局依赖性<sup>[13]</sup>。因此,基于现有的双重注意力机制<sup>[14]</sup>以及协调注意力机制<sup>[15]</sup>,设计了一种基于自注意力机制的双重注意力机制并将其融入基准网络 2S-AGCN 模型的图卷积模块中,该双重注意力机制被用来提取输入数据在空间维度上长距离信息以及通道之间的依赖关系等。

本文设计的双重注意力机制模块使用了两个并行的注意力模块,如图 3 所示,分别为空间注意力模块以及通道注意力模块,然后将两个注意力机制模块的结果进行融合。其中空间注意力模块将更广泛的上下文信息编码到局部特征中,从而增强其表示能力,而通道注意力模块被用来学习

通道之间相互依赖关系等。

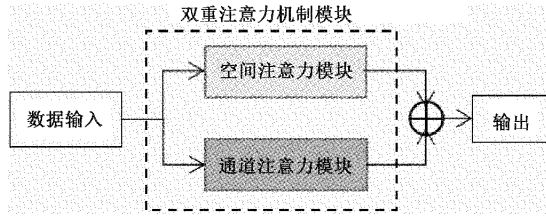


图3 双重注意力机制模块结构

2S-AGCN 模型经过改进后的图卷积模块结构如图4所示,包含有空间图卷积模块、时间图卷积模块、本文经过设计的双重注意力机制模块(dual attention module)以及批标准化层、激活函数层、Dropout层。其中空间图卷积模块用于对由关节数据生成的邻接矩阵图进行图卷积运算,提取空间上的特征信息,时间图卷积模块用于提取数据的时间维度上的特征信息。

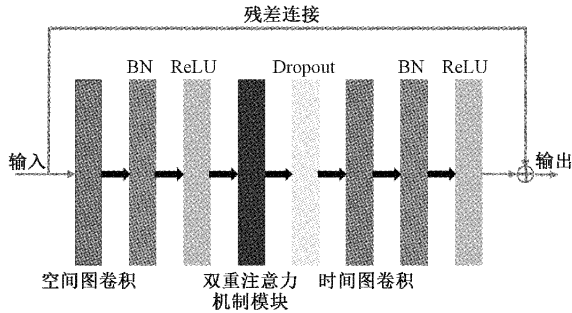


图4 2S-AGCN 网络改进后图卷积模块

## 2.1 空间注意力模块

空间注意力模块(spatial attention, SAT)具体设计如图5所示,给定输入特征,通过 $1 \times 1$ 的卷积层分别压缩提取出两个新的特征映射,得到特征图 $B$ 和 $C$ ,通过重塑(reshape)操作将 $B$ 和 $C$ 的尺寸变换, $C$ 与 $B$ 的转置之间进行矩阵乘法操作并通过Softmax层计算得到空间注意力图 $S$ ,计算公式可以表示为:

$$s_{ji} = \frac{\exp(B_i \cdot C_j)}{\sum_{i=1}^N \exp(B_i \cdot C_j)} \quad (1)$$

其中, $s_{ji}$ 测量了 $i$ 位置对第 $j$ 位置的影响。两个位置的特征表示越相似,它们之间的相关性就越大。同时,给定输入特征 $A$ 经过 $1 \times 1$ 的卷积层得到新的特征映射 $D$ ,通过reshape操作进行尺寸变换。将特征图 $D$ 与空间注意力图 $S$ 的转置相乘并通过reshape操作进行尺寸变换。将输出结果乘以比例参数 $\alpha$ 并对特征进行元素求和运算,获得第5个特征输出 $E$ ,其中 $\alpha$ 初始化为0,通过每一轮的迭代学习逐渐分配更多权重,计算公式可以表示为:

$$E_j = \alpha \sum_{i=1}^N (S_{ji} D_i) + A_j \quad (2)$$

每个位置的结果特征是所有位置特征和原始特征的加

权和。因此,它具有全局上下文视图信息,可以根据空间注意图自适应的聚合长距离上下文信息,建模丰富的上下文依赖关系来学习特征的空间相关性。通过相似的语义特征实现了相互增益。

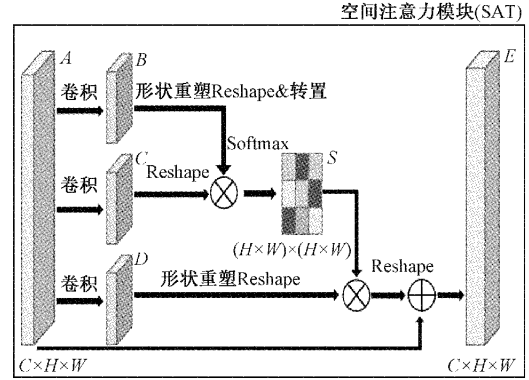


图5 空间注意力模块结构图

## 2.2 通道注意力机制

通道注意力模块分为两部分,如图6所示。对于第一部分通道注意力(channel attention-1, CAT-1)。给定特征 $A$ ,通过reshape操作调整特征 $A$ 尺寸大小,将 $A$ 和 $A$ 的转置进行矩阵乘法操作,通过Softmax层获得通道注意力,计算公式可以表示为:

$$x_{ji} = \frac{\exp(A_i \cdot A_j)}{\sum_{i=1}^c \exp(A_i \cdot A_j)} \quad (3)$$

通道注意力图 $X$ 的计算公式如式(3)所示,其中 $x_{ji}$ 测量第 $i$ 个通道对第 $j$ 个通道的影响。然后将通道注意力图 $X$ 和给定特征 $A$ 的转置之间进行矩阵乘法,通过reshape操作调整输出特征大小。将结果乘以比例参数 $\beta$ ,并对给定特征 $A$ 进行元素求和运算,获得输出特征图。其中 $\beta$ 初始化为0,通过每一轮的迭代学习逐渐分配更多权重,计算公式可以表示为:

$$F_j = \beta \sum_{i=1}^c (x_{ji} A_i) + A_j \quad (4)$$

根据式(4)可以得到,每个通道的特征 $F$ 是所有通道的特征和原始特征的加权和,这模拟了特征映射之间的长期语义依赖关系,有助于提高特征的可区分性和语义一致性。

对于第二部分通道注意力(channel attention-2, CAT-2),给定输入特征 $A$ ,使用两个尺寸分别为 $(H, 1)$ 和 $(1, W)$ 的池化层卷积核沿水平和垂直的方向对每个通道进行编码,如式(5)、(6)所示,得到的特征图 $B$ 和 $C$ 可以分别用公式表达为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq j < W} x_c(h, j) \quad (5)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (6)$$

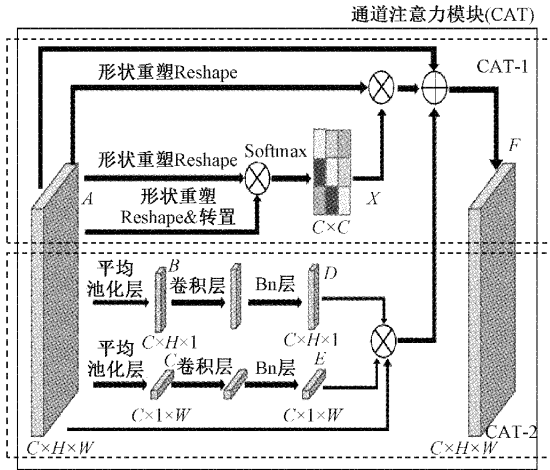


图 6 通道注意力模块结构图

以上两种转换方式将相对的两个空间方向的特征聚合在一起,从而能够沿一个方向捕捉长距离依赖关系,沿另一个方向保持精准的原始位置信息,帮助网络更精确的定位感兴趣的特征。

特征图 B 和 C 后沿空间维度拼接在一起,并通过  $1 \times 1$  的卷积变换,批归一化层以及非线性激活函数,公式表示为:

$$f = \delta(F_1([z^h, z^w])) \quad (7)$$

将结果分别输入两个不共享权值的卷积变换层并通过 Sigmoid 激活函数计算得到特征图 D 和 E,如公式所示:

$$g^h = \sigma(F_h(f^h)) \quad (8)$$

$$g^w = \sigma(F_w(f^w)) \quad (9)$$

最后将特征图 D 和 E 以及原始特征输入 A 进行点乘操作得到输出特征图。这种设计比较简单和轻量,可以充分利用捕获的位置信息,精确的突出感兴趣的部分,与此同时它还能有效地捕获通道与通道之间的关系,与第 1 种通道注意力机制之间互补。

### 2.3 损失函数

动作分类识别使用的损失函数为交叉熵损失函数,在做分类训练的时候,利用该损失函数去衡量网络的实际输出和正确标签的差异,并利用这种差异经过反向传播更新网络参数,其公式可以表示为:

$$Loss = - \sum_{i=1}^n p(x_i) \log(q(x_i)) \quad (10)$$

其中,  $n$  为动作分类的数量,  $p(x_i)$  为符号函数,  $q(x_i)$  为实际网络输出。

## 3 实验结果与分析

### 3.1 数据集

实验使用目前流行且广泛使用的几个数据集。包括 NTU-RGB+D, NTU-RGB+D120 和 Kinetics 数据集。

其中 NTU-RGB+D 数据集<sup>[16]</sup>包含 60 个动作,56 000 个动作片段。由年龄从 10~35 岁的 40 名志愿者拍摄而

成。关节数据通过使用 Kinect 深度传感器检测每个帧的 3D 关节位置得到。按照两种方式划分训练集和测试集,一种是按照人物:CS(cross-subject)来划分,将 40 名对象分为训练组和测试组,另一种是按照拍摄相机 CV(cross-view)来划分,将 1 号摄像机采集的 18 960 个视频样本用于测试,其余的样本用作训练。同样地,NTU-RGB+D120 数据集也同样分为两种方式,一种按照人物 ID,另一种按照拍摄相机,1 号摄像机采集的样本作为训练,其他摄像机拍摄的视频样本作为测试。

Kinetics 数据集是一个包含 300 000 个视频剪辑,400 种分类的数据集。每段视频都含有一个标签,标签内容包括动作类别,视频编号,属于哪个数据集等等。该数据集的关节数据是通过 OpenPose 工具箱提取出每帧图像的 18 个骨骼点得到,训练集包含 240 000 个样本,测试集包含 20 000 个样本。

如图 7 所示,左边为 Kinetics 数据集关节点的分布和连接,编号从 0~17,右边为 NTU-RGB+D 数据集关节点的分布和连接,编号从 1~25。

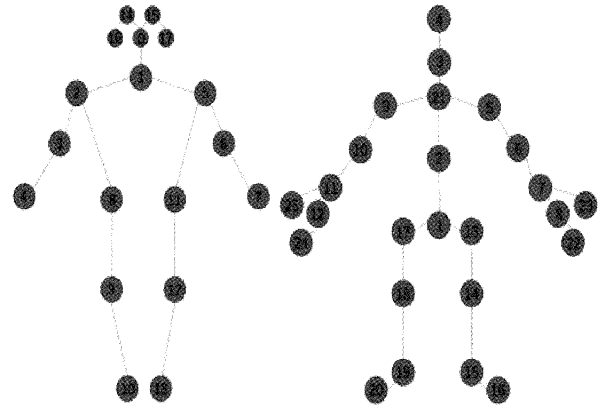


图 7 关节数据图

### 3.2 实验环境与参数设置

实验在配备了 2 个 NVIDIA 3090 GPU 的服务器上训练。训练次数以及学习率根据不同数据集而不同。在 NTU-RGB+D 数据集上,训练轮数设置为 50 次,学习率设为 0.005,在 NTU-RGB+D120 数据集上,训练轮数设置为 60 次,学习率设为 0.002,在 Kinetics 数据集上,训练轮数设置为 50 次,学习率设为 0.002。

### 3.3 实验分析

#### 1) 消融实验

本实验主要用于研究两种通道注意力机制对模型精度提升的影响。在 NTU-RGB+D120 数据集上验证本文设计的通道注意力机制模块的有效性。双重注意力机制包含了两种通道注意力机制模块并且并行连接。分别测量了只使用一种通道注意力机制模块(CAT-1)和同时使用两种通道注意力机制模块(CAT-1+CAT-2),两种情况下的精度提升。以 2S-AGCN 网络为基准模型,将融合了双重注意



力机制模块的网络和原本网络进行对比,如表1所示。

表1 两种通道注意力机制测试对比 %

模型	CS	CV
SAT+CAT-1	83.7	85.2
SAT+CAT-1+CAT-2	84.1	85.4

通过表1的各项数据,可以得到两个结论:(1)使用了两种通道注意力机制模块并以并行方式连接的双重注意力机制模块相较于只使用了一种通道注意力机制的网络得到了更好的精度提升,说明了两种通道注意力机制模块并行连接的有效性。(2)并行连接的通道注意力机制模块相较于以拍摄相机(CV)来划分数据集,在以不同人物(CS)来划分数据集的情况下,提升的精度相对高一点,说明这种注意力机制模块,更擅长识别固定拍摄角度下,不同人物的动作行为。

### 2) 双重注意力机制模块实验

本实验将双重注意力机制模块融入图卷积网络2S-AGCN模型中,通过进行端到端的训练以完成人体动作识别任务的分类,用于验证经过设计的双重注意力机制模块在动作识别任务中的提升效果。以2S-AGCN为基准网络模型,分别测试在不同数据集上的训练结果,为了和原网络进行对比,在相同数据集上训练过程所用的参数,学习率,训练轮数,使用的GPU型号数量全部相同,如表2表示了NTU-RGB+D和NTU-RGB+D120数据集上的测试结果,表3表示了Kinetics数据集上的测试结果。

表2 有无数注意力机制对比 %

模型	NTU	NTU	NTU120	NTU120
	(CS)	(CV)	(CS)	(CV)
2S-AGCN	88.5	95.1	82.9	84.9
本文	88.7	95.2	84.1	85.4

表3 有无数注意力机制对比 %

模型	Top-1	Top-5
2S-AGCN	36.1	58.7
本文	36.7	60.0

根据表2可以看出。在NTU-RGB+D数据集上,以不同情况划分数据集的情况下,CS上提升了0.2%,CV上提升了0.1%,在NTU-RGB+D120数据集上,CS上提升了1.2%,CV上提升了0.5%。可以得出结论:(1)本文设计的双重注意力机制模块对基准网络2S-AGCN网络的识别精度提升有一定的效果,验证了有效性。(2)根据划分数据集方式的不同以及精度提升结果,再次说明了本文设计的双重注意力机制模块更适合检测角度固定,人员不同情况下的动作分类识别精度。(3)相较于NTU-RGB+D数

据集,在拥有更多分类类别的NTU-RGB+D120数据集上,得到了相对更高的精度提升。

根据表3可以看出,在Kinetics数据集中,Top1的识别准确率达到36.7%,相较于原本提升了0.6%,而Top5的准确率达到60.0%,提升了1.3%。可以得到两个结论:(1)本文设计的双重注意力机制模块对于Kinetics这种超大型的数据集有一定程度的提升。(2)对于Kinetics这种更大、更复杂的数据集,除去模型的上下文关系之外,还需要考虑更多时间维度上的特征提取,因此动作识别精度提升幅度不大。

### 3) 对比实验

本实验主要将本文改进的算法模型和使用相同数据集的其他动作识别算法模型进行对比,分别使用NTU-RGB+D120数据集和Kinetics数据集进行对比,如表4和5所示。

表4 NTU-RGB+D120数据集算法对比实验 %

算法	CS	CV
ST-LSTM	55.7	57.9
GCA-LSTM	61.2	63.3
2S-AGCN	82.9	84.9
MS-G3D	86.9	88.4
Ours	84.1	85.4

表5 Kinetics数据集算法对比实验 %

算法	Top-1	Top-5
ST-GCN	30.7	52.8
AS-GCN	34.8	56.5
ST-GR	33.6	56.1
2S-AGCN	36.1	58.7
MS-G3D	38.0	60.9
Ours	36.7	60.0

通过对比可以观察得到,本文将设计的双重注意力机制模块融入基准网络2S-AGCN上后得到的模型在多个数据集上取得了一定的精度提升。并且相较于以前的方法有一定程度的提升。由于设计的双重注意力机制主要考虑了空间注意力机制以及通道注意力机制,并没有考虑到动作识别中时间序列特征提取可能带来的问题,因此无法和最新的考虑了时间维度特征信息的动作识别网络相比,但相较于其他算法仍有一定优势。

## 4 结 论

本文在现有的2S-AGCN网络模型的基础上,使用双重注意力机制改进了原本网络的时空卷积模块,双重注意力机制的核心思想就是使用了两个并行的注意力模块,一个用来建立丰富的上下文关系模型,学习特征的相互依赖

关系,另一个用来建立通道相互依赖关系,从而增强了表示能力,然后将两个注意力机制的结果进行融合。

实验在多个数据集上进行训练和测试,并且分别得到了不同程度的提升,在 NTU-RGB+D 数据中,根据数据集划分方式不同,识别准确率分别达到了 88.7% 和 95.2%。在 NTU-RGB+D120 数据集中,识别准确率分别达到了 84.1%,85.4%,在 Kinetics 数据集中,Top1 准确率达到了 36.7%。

本文的不足在于没有考虑了时间维度上特征信息对动作识别带来的影响,而动作往往是由一段时间内,多个图像组成的,因此接下来需要继续探索时间特征的信息以及其与空间特征信息的关系。

### 参考文献

- [1] 袁首,乔勇军,苏航,等. 基于深度学习的行为识别方法综述[J]. 微电子学与计算机,2022,39(8):1-10.
- [2] 李瑞峰,王亮亮,王珂. 人体动作行为识别研究综述[J]. 模式识别与人工智能,2014,27(1):35-48.
- [3] 朱相华,智敏,殷雁君. 基于 2D CNN 和 Transformer 的人体动作识别[J]. 电子测量技术,2022,45(15):123-129.
- [4] 张华. 模式识别及其在计算机视觉中的实现[J]. 电子技术与软件工程,2014(24):181-182.
- [5] 杨戈,邹武星. 基于深度学习的视频行为分类方法综述[J]. 电子技术应用,2022,48(7):1-7,12.
- [6] 马世拓,班一杰,戴陈至力. 卷积神经网络综述[J]. 现代信息科技,2021,5(2):11-15.
- [7] 罗会兰,童康,孔繁胜. 基于深度学习的视频中人体动作识别进展综述[J]. 电子学报,2019,47(5):1162-1173.
- [8] 王小娟,钟云,金磊,等. 基于骨架的自适应尺度图卷积动作识别[J]. 天津大学学报(自然科学与工程技术版),2022,55(3):306-312.
- [9] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
- [10] SHI L, ZHANG Y, CHENG J, et al. Two-stream adaptive graph convolutional networks for skeleton-based action recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12026-12035.
- [11] SI C, CHEN W, WANG W, et al. An attention enhanced graph convolutional lstm network for skeleton-based action recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1227-1236.
- [12] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7291-7299.
- [13] 高广尚. 深度学习推荐模型中的注意力机制研究综述[J]. 计算机工程与应用,2022,58(9):9-18.
- [14] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3146-3154.
- [15] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [16] SHAHROUDY A, LIU J, NG T T, et al. Ntu RGB+d: A large scale dataset for 3D human activity analysis[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1010-1019.

### 作者简介

陶峰,硕士研究生,主要研究方向为深度学习、动作识别。

E-mail:20721457tf@shu.edu.cn

李燕苹,博士,助理研究员,主要研究方向为智能信息处理、AI教育。

E-mail:yanpingli@shu.edu.cn

王瑞,教授,主要研究方向为通信信号处理、传感器网络、几何代数等。

E-mail:rwang@sh.edu.cn