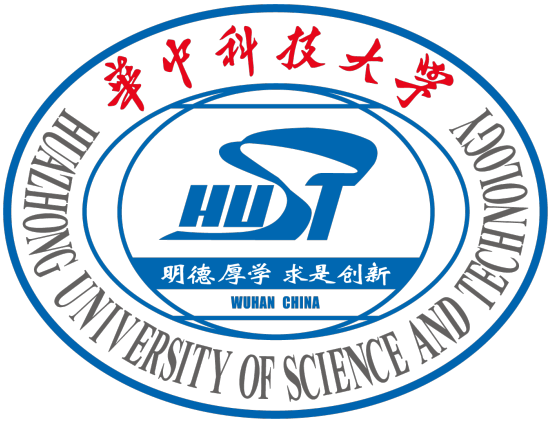
基于无人机的人体行为识别算法说明书



队伍名称：**HUSTEEE**

团队成员：王宏伟, 牛君元, 李香荟

华中科技大学

2024 年 11 月 12 日

目录

摘要

近年来，基于骨架的人体行为识别在智能视觉领域引起了广泛关注，尤其是在更贴近现实的无人机视角下。由于无人机拍摄涉及多样的视角和尺度变化，现有模型的性能在实际场景中往往不佳。我们希望从无人机视角以及骨架行为数据处理和特征提取几个方面入手，对现有的模型、算法进行优化，实现识别精度的提高。

在我们设计的算法框架中，我们针对模拟无人机视角进行了特定的数据增强设计，增强了模型在复杂场景下的鲁棒性。同时，我们创新性地加入了 **JBF（Joint-Bone Fusion）** 模块，将关节和骨骼模态进行融合，弥补了传统训练方式中关节与骨骼模态相关性的不足，从而更好地捕捉了骨架数据中的关联信息。此外，针对数据集规模有限且动作类型种类多难区分的问题，我们使用 **GAP(Generative Action Description Prompts)**，引入大语言模型生成的额外语义信息设计语义头，以丰富数据特征表达，提升模型对动作的理解能力。针对官方提供的数据集多实体这一特点，在识别多实体交互动作方面，我们通过引入 **CHASE（Convex Hull Adaptive Shift）** 模块，优化了主干网络结构，使其更好地适应实体骨架间的分布差异。最终，我们结合 GCN 和 Transformer 的优势，利用双分支网络，充分融合两者的优点，实现了更强的识别效果。

一、 数据预处理

1.1 数据增强的动机

我们通过对骨骼数据的方向向量进行随机缩放、旋转、平移、形变和倾斜等增强操作，模拟无人机视角下的多样化观察角度与距离变化。增强过程首先将关节点数据转换为骨骼方向向量，再在方向向量上应用随机变换，最后还原为关节点坐标。通过组合不同的增强操作，生成多样化的数据样本，使模型在训练中适应不同的无人机视角，提升对动态和复杂视角下的鲁棒性和泛化能力。

在实验过程中并不是这些增强都有正面的效果，根据不同模型和不同数据流，我们采用了不同数量和种类的增强。

1.2 算法说明

1.2.1 骨架缩放

对骨骼长度进行随机缩放，模拟无人机视角下的距离变化。缩放在指定范围内生成随机比例，对称骨骼对保持一致缩放，确保整体结构对称性。

假设骨骼关节点数据的方向向量和长度分别为 $\mathbf{d}_i$ 和 l_i，其中 i 表示第 i 根骨骼。通过以下步骤实现随机缩放：

1. **生成缩放因子**：

在指定范围内生成随机缩放因子 α_i：

$$\alpha_i = 1 + \text{uniform}(-\theta, \theta) \tag{1}$$

其中 θ 是缩放幅度的范围，$\text{uniform}(-\theta, \theta)$ 表示在 $[-\theta, \theta]$ 范围内的均匀分布随机数。

2. **对称性保持**：

对称骨骼对 (i, j) 共享相同的缩放因子，以保证左右对称性：

$$\alpha_i = \alpha_j \tag{2}$$

3. **应用缩放**：

将缩放因子 α_i 应用于每个骨骼的长度 l_i：

$$l_i' = \alpha_i \cdot l_i \tag{3}$$

其中 l_i' 表示缩放后的骨骼长度。

4. **骨骼方向的更新**：

最终，每个骨骼的增强后的方向向量为：

$$\mathbf{b}_i = l_i' \cdot \mathbf{d}_i \tag{4}$$

该过程对所有骨骼方向向量应用，从而实现骨骼长度的随机缩放效果。

1.2.2 骨架旋转

对于骨骼旋转增强，我们对所有骨骼的方向向量进行整体旋转，以模拟无人机视角下不同的拍摄角度变化。假设骨骼方向向量为 $\mathbf{d}_i$，其中 i 表示第 i 根骨骼，旋转过程分为以下几个步骤：

1. **生成旋转角度**：

在指定范围内生成一个随机旋转角度 θ，例如在 $[-\theta_{\text{max}}, \theta_{\text{max}}]$ 范围内：

$$\theta = \text{uniform}(-\theta_{\text{max}}, \theta_{\text{max}}) \tag{5}$$

其中 $\text{uniform}(-\theta_{\text{max}}, \theta_{\text{max}})$ 表示在 $[-\theta_{\text{max}}, \theta_{\text{max}}]$ 范围内的均匀分布随机数。

2. **构建旋转矩阵**：

假设旋转发生在 Z 轴（即水平平面旋转），对应的旋转矩阵 $\mathbf{R}_z$ 为：

$$\mathbf{R}_z = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

3. **应用旋转**：

对于每个骨骼的方向向量 $\mathbf{d}_i$，应用旋转矩阵 $\mathbf{R}_z$ 进行旋转变换，得到旋转后的方向向量 $\mathbf{d}_i'$：

$$\mathbf{d}_i' = \mathbf{R}_z \cdot \mathbf{d}_i \tag{7}$$

这样，每个骨骼的方向向量在空间中围绕 Z 轴旋转 θ 角度。

4. **恢复骨骼长度并重构位置**：

旋转后的骨骼方向 $\mathbf{d}_i'$ 保持了原始骨骼的长度 l_i，因此每个旋转后的骨骼向量可以表示为：

$$\mathbf{b}_i' = l_i \cdot \mathbf{d}_i' \tag{8}$$

通过旋转后的骨骼向量 $\mathbf{b}_i'$ 可以重构所有关节点的位置，实现增强后的骨骼姿态。

该过程对所有骨骼的方向向量应用统一的旋转变换，从而实现模拟不同拍摄角度下的骨骼旋转效果。

1.2.3 骨架平移

在骨骼数据增强中，平移操作用于模拟不同位置下的骨骼偏移情况，使得模型更具鲁棒性。假设骨骼数据中的关节点坐标为 $\mathbf{p}_i$，其中 i 表示第 i 个关节点，平移过程如下：

1. **生成平移向量**：

在每个坐标轴方向上生成一个随机平移量 $\Delta x, \Delta y, \Delta z$，模拟三维空间中的随机平移。平移向量 $\mathbf{t}$ 表示为：

$$\mathbf{t} = \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \tag{9}$$

其中 $\Delta x, \Delta y, \Delta z$ 是在预设范围 $[-t_{\text{max}}, t_{\text{max}}]$ 内的均匀分布随机数，例如：

$$\Delta x, \Delta y, \Delta z \sim \text{uniform}(-t_{\text{max}}, t_{\text{max}}) \tag{10}$$

2. **应用平移**：

将平移向量 $\mathbf{t}$ 应用到每个关节点坐标 $\mathbf{p}_i$，得到平移后的关节点位置 $\mathbf{p}'_i$：

$$\mathbf{p}'_i = \mathbf{p}_i + \mathbf{t} \tag{11}$$

其中，$\mathbf{p}'_i$ 表示在平移后的新位置下，第 i 个关节点的坐标。

3. **保持骨架结构不变**：

平移操作应用于所有关节点，骨架结构的相对位置保持不变。因此，平移后的骨架数据可以用来模拟无人机在不同位置观察相同姿态的效果。

通过这种平移增强操作，骨架数据可以在三维空间中模拟不同的相对位置，有助于模型在训练过程中学习到更广泛的位置变化特征。

1.2.4 骨架倾斜

倾斜操作用于模拟拍摄视角的倾斜，增加模型对不同观察角度的鲁棒性。假设骨骼数据中的关节点坐标为 $\mathbf{p}_i$，其中 i 表示第 i 个关节点。倾斜过程如下：

1. **生成倾斜角度**：

在指定范围内生成一个随机倾斜角度 α，例如在 $[-\alpha_{\text{max}}, \alpha_{\text{max}}]$ 范围内：

$$\alpha = \text{uniform}(-\alpha_{\text{max}}, \alpha_{\text{max}}) \tag{12}$$

其中 $\text{uniform}(-\alpha_{\text{max}}, \alpha_{\text{max}})$ 表示在 $[-\alpha_{\text{max}}, \alpha_{\text{max}}]$ 范围内的均匀分布随机数。

2. **构建倾斜矩阵**：

以 X 轴倾斜为例（即在 YZ 平面上倾斜），倾斜矩阵 $\mathbf{T}_x$ 可表示为：

$$\mathbf{T}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \tan(\alpha) \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

同样，若要以 Y 轴为基准（在 XZ 平面上倾斜），则倾斜矩阵 $\mathbf{T}_y$ 为：

$$\mathbf{T}_y = \begin{bmatrix} 1 & 0 & 0 \\ \tan(\alpha) & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

3. **应用倾斜变换**：

对于每个关节点 $\mathbf{p}_i$，将倾斜矩阵 $\mathbf{T}$ 应用到其坐标上，得到倾斜后的关节点坐标 $\mathbf{p}'_i$：

$$\mathbf{p}'_i = \mathbf{T} \cdot \mathbf{p}_i \tag{15}$$

其中，$\mathbf{T}$ 可以是 $\mathbf{T}_x$ 或 $\mathbf{T}_y$，具体取决于倾斜方向。

4. **保持骨架结构的一致性**：

倾斜操作应用于所有关节点，以模拟整个骨架在倾斜角度下的姿态变化。该操作保持关节点的相对距离，但在空间中表现出倾斜效果。

通过倾斜增强操作，可以模拟不同的倾斜视角，提升模型在训练时适应各种观察角度的能力。

1.2.5 骨架形变

形变操作用于模拟拍摄角度的空间扭曲或形变效果，帮助模型适应不同观察视角下的空间变形。假设骨骼数据中的关节点坐标为 $\mathbf{p}_i$，其中 i 表示第 i 个关节点。形变过程如下：

1. **生成形变因子**：

在指定范围内生成形变因子 s_x 和 s_y，以模拟在不同轴方向上的形变变形。形变因子通常在 $[-s_{\text{max}}, s_{\text{max}}]$ 范围内随机生成，例如：

$$s_x, s_y \sim \text{uniform}(-s_{\text{max}}, s_{\text{max}}) \tag{16}$$

2. **构建形变矩阵**：

假设形变操作发生在 XY 平面，即沿 X 轴和 Y 轴方向的变形。对应的形变矩阵 $\mathbf{S}$ 表示为：

$$\mathbf{S} = \begin{bmatrix} 1 & s_x & 0 \\ s_y & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{17}$$

形变矩阵 $\mathbf{S}$ 会将 X 和 Y 轴上的关节点进行相对位置的拉伸或压缩。

3. **应用形变变换**：

将形变矩阵 $\mathbf{S}$ 应用到每个关节点的坐标 $\mathbf{p}_i$，得到形变后的关节点坐标 $\mathbf{p}'_i$：

$$\mathbf{p}'_i = \mathbf{S} \cdot \mathbf{p}_i \tag{18}$$

其中，$\mathbf{p}_i'$ 表示第 i 个关节点在形变变形后的新坐标位置。

4. **保持骨架结构的一致性**：

形变操作应用于所有关节点，以模拟整个骨架在特定方向上的形变。该操作保持关节点的连通性，但在空间中产生非对称的变形效果。

通过形变增强操作，可以模拟拍摄视角引起的形变效果，从而增加模型对不同视角和拍摄角度的适应能力。

二、 特征提取流的创新

2.1 采用 JBF 特征流的动机

骨骼模态通过人体的语义连接来表示，与关节一起是构成骨骼图不可或缺的线索。然而，大多数现有方法忽略了关节和骨骼这两种模态之间的相关性，而是将它们分开处理。所以，我们采用在特征层面对两个模态进行融合的方式，充分挖掘关节与骨骼的特征相关性，从而提升动作识别的精度。

2.2 特征流算法设计

我们采用关节 - 骨骼融合流的方式。具体操作是，首先为每个模态（关节和骨骼）通过由三个基本块（即 L1=3）组成的流提取判别性表示。然后通过元素 - 求和并结合一个批归一化层将它们融合。最后，融合后的输出通过七个基本块（即 L2=7），以多分支形式（如图1）学习相关表示，这与特征提取流类似。

每个分支包含一个初始块和九个基本块，后面接着一个全局平均池化层（GAP）和一个全连接层 (FC)。基本块：由一个常规的图卷积（GCN）模块和时间建模（TCN）模块组成。

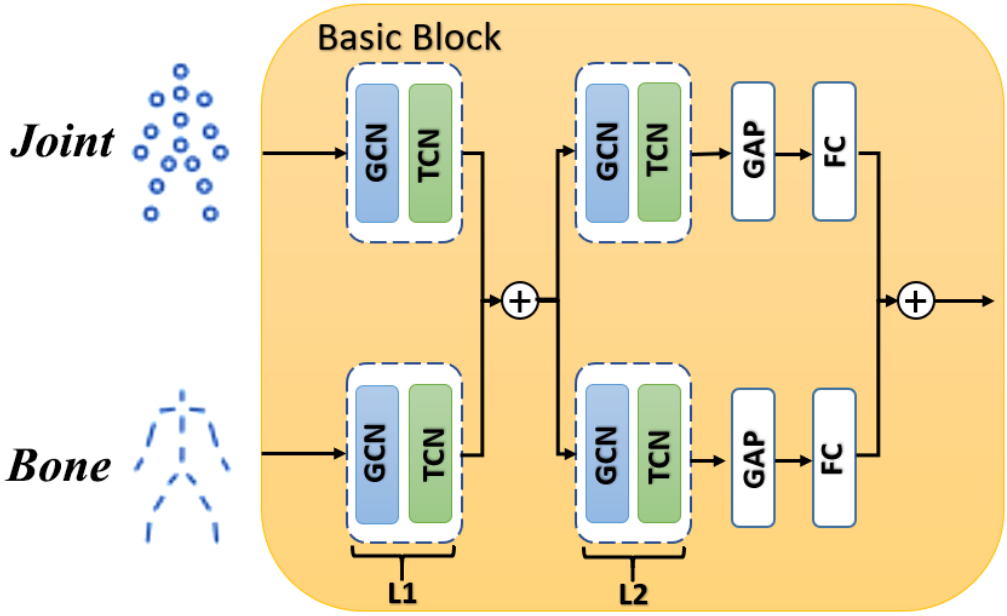


图 1　关节骨骼融合流

为了更好的挖掘关节与骨骼之间的相关性，我们在 GCN 模块添加了注意力机制（如图2）。

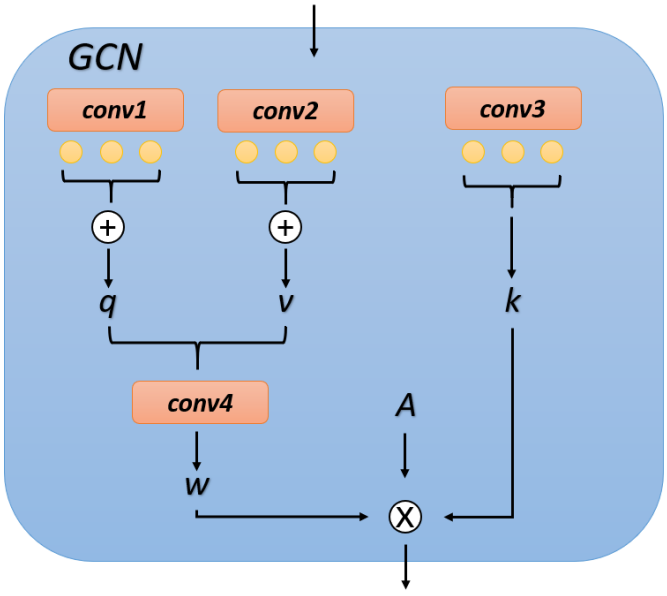


图 2　注意力机制 GCN 模块

1. **特征提取**

- conv1 和 conv2 从输入 x 中提取特征，用于构建查询向量 q 和键向量 k。具体来说：

  - `conv1(x)` 生成查询特征 q，并通过在时序维度上进行平均，得到一个具有位置维度的特征张量。
  - `conv2(x)` 生成键特征 k，并同样在时序维度上进行平均。
- `conv3(x)` 从输入 x 中生成值特征 v,，保留表示不同尺度和头数的特征。

2. **注意力权重计算**

   `conv4(x)` 利用查询特征 q 和键特征 k 计算注意力权重图。

3. **应用注意力图**

   - 在得到各节点的注意力权重 `weights` 后，将它与 α 权重相乘，并添加邻接矩阵 A，生成最终的图结构。
   - 使用爱因斯坦求和简写方式将注意力权重 `weights` 应用于值特征 v：通过矩阵乘法生成最终输出。

2.3 实验效果

我们在赛题官方的验证集上在 CTRGCN 基线下进行了准确率评估。JBF 特征流的效果优于单独使用 joint 和 bone 模态，并在 3s 下较 2s 提升准确率 1.25%。

特征流	Acc. (%)
Joint	42.60
Bone	42.25
Joint+Bone(2s)	45.30
JBF	**44.35**
Joint+Bone+JBF(3s)	**46.55**

表 1　特征流比较结果

三、 结合生成式动作描述的模型优化

3.1 优化思路

当前的骨架动作识别方法通常被形式化为独热编码分类任务，并没有充分利用动作之间的语义关系。例如，“make victory sign” 和 “thumb up” 是手势的两个动作，它们的主要区别在于手的动作。此信息与操作类的分类 one-hot 编码无关，但可以从操作描述中揭示。因此，在训练中使用动作描述可能有益于表征学习。

在此次比赛官方赛题数据所给出的 155 个动作标签上，各类动作的细粒度分类难度高，而引入文本信息可以拉开类间差距缩小类内差距。

在赛题项目中，我们尝试使用了一种基于骨骼数据结合生成式动作描述（**G**enerative **A**ction-description **P**rompts（GAP））的动作识别方法。更具体地说，我们采用预先训练的大规模语言模型作为知识引擎，自动生成身体部位动作的文本描述，并利用文本编码器为不同的身体部位生成特征向量，并监督骨骼编码器进行动作表示学习，从而使用多模态训练方案。

3.2 具体工作

本次无人机行为识别算法赛所提供的数据是单模态的数据，受到最近图像和语言多模态训练成功的启发，我们在基线模型上结合 GAP 方法，利用以语言提示的形式生成类别级人类动作描述。动作的语言定义包含丰富的先验知识。例如，不同的动作聚焦于不同身体部位的运动：“划圈” 和 “挥动网球拍” 描述手臂的运动；“跳跃” 和 “侧踢” 依赖于脚和腿的运动。一些动作描述了多个身体部位的互动，例如，“戴帽子” 和 “穿鞋” 涉及手和头、手和脚的动作。这些关于动作的先验知识可以为表示学习提供细粒度的指导。此外，为了解决收集人类动作提示的繁琐工作，我们求助于大型语言模型（LLM），例如 GPT-4o，以便高效地自动生成提示。

我们的优化工作主要集中在以下几个方面：

· 生成式动作描述生成：利用大语言模型（GPT-4o）生成动作的文本描述，为每个动作的特征学习提供指导。具体来说，根据赛题官方提供的动作数据名称，我们为每个动作设计了动作的细粒度身体部位描述，使模型能够更精确地捕捉动作特征。

· 多模态训练框架搭建：我们设计并实现了一个由骨架编码器和文本编码器组成的双编码器框架。骨架编码器负责将骨架坐标输入转换为局部和全局特征表示，而文本编码器则将生成的动作描述转换为文本特征。在训练过程中，通过多部分对比损失将骨架特征和文本特征对齐，强化模型的动作表示能力。

具体而言，我们使用了一种新的训练范式，通过在训练中利用动作描述，来辅助骨架信息，通过两个编码器（骨架编码器、文本编码器）的结合——骨架编码器部分：生成全局平均特征和部分平均特征；文本编码器部分：首先通过大语言模型生成适配标签的文本，再通过文本编码器生成全局和部分特征。最后实现将骨架编码器的特征和文本编码器的特征进行对齐，通过文本编码器的特征辅助骨架编码器特征，使得骨架编码器能够学习到更加丰富和准确的特征。

在**图3**中，我们将我们使用的框架（b）和（c）与传统的单编码器基于骨架的动作识别框架（a）进行了比较。在我们的框架中，使用了一种多模态训练方案，其中包含一个骨架编码器和一个文本编码器。多部分对比损失（对于（b）为单一对比损失）用

于对齐文本部分特征和骨架部分特征，并且交叉熵损失应用于全局特征上，增加对比损失能帮助模型更好的学习动作特征。

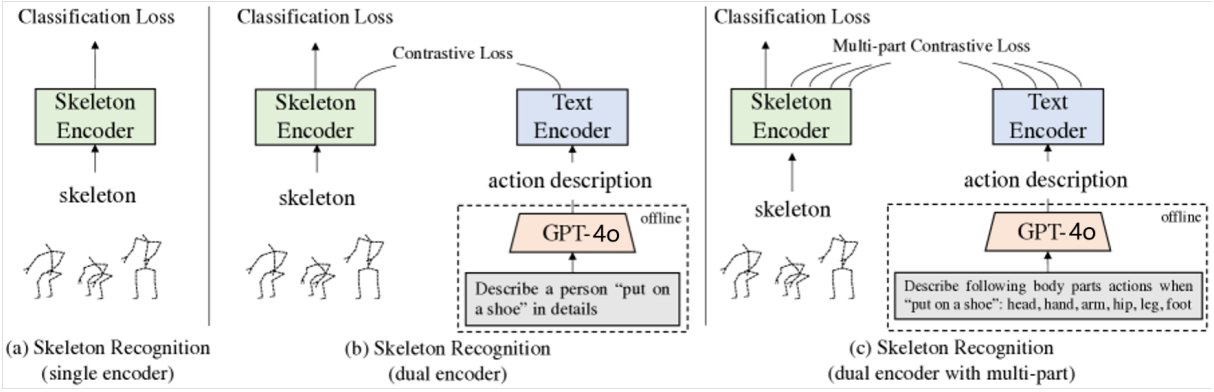


图 3　GAP 方法模型框架图

3.3 实验效果

实验结果表明，结合了 GAP 的模型在国赛数据集上的准确率相较基线模型提高了 1.0%-1.3%，展现了细粒度描述提示对无人机视角下复杂动作识别的提升效果。而且由于 GAP 方法引入了动作细节描述，模型在应对数据集中视角变化和背景噪声时表现更为稳定，特别是在动态背景和视角偏移较大的场景中。在计算效率方面，GAP 方法仅在训练阶段引入文本描述编码器，在推理阶段模型仍然只依赖骨架编码器的全局特征，未增加额外的推理计算成本。

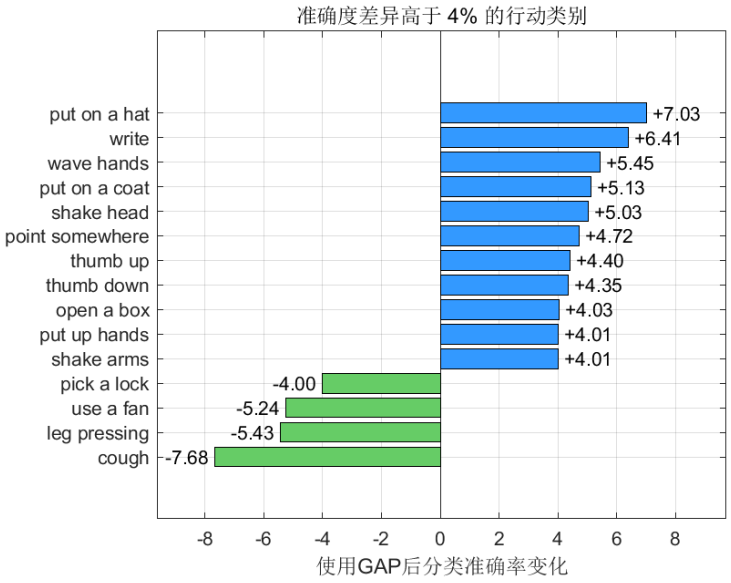


图 4　准确率变化效果

四、 多实体行为数据的模型优化

4.1 引入 CHASE 的动机

赛题数据集的场景下有 1-2 个目标实体，而基于骨骼的多实体动作识别是一项具有挑战性的任务，旨在识别涉及多个不同实体的交互式动作或群体活动。由于实体骨架之间固有的分布差异，现有的个体模型往往无法完成这项任务，从而导致主干优化不理想。为此，我们引入了一种基于 Convex Hull Adaptive Shift 的多实体动作识别方法（CHASE），该方法可以减少实体间的分布差距并消除了后续主干的偏差。

4.2 模块说明

CHASE 包括一个可学习的参数化网络和一个辅助目标。参数化网络通过两个关键组件实现了骨架序列的合理、样本自适应重新定位。首先，隐式凸包约束自适应偏移确保坐标系的新原点位于骨架凸包内。其次，Coefficient Learning Block 提供了从骨架序列到凸组合中特定系数的映射的轻量级参数化。此外，为了指导该网络的优化以实现差异最小化，在目标损失中加入 Mini-batch Pair-wise Maximum Mean Discrepancy 作为附加目标。CHASE 作为一种样本自适应归一化方法，以减轻实体间分布差异，从而减少数据偏差并提高后续分类器的多实体动作识别性能。

CHASE 作为一个即插即用的模块，其中的权重系数学习模块（Coefficient Learning Block）（如图5）在骨架识别模型框架中充当 Neck 的部分，该结构由三个 $1 \times 1 \times 1$ 的三维卷积层组成，分别用于特征压缩、激活和组合。首先，输入特征经过一个卷积层，随后通过 Squeeze 操作进一步压缩通道信息，再经过第二个 $1 \times 1 \times 1$ 卷积层和激活函数（如 ReLU）增加非线性表达能力，最后通过第三个 $1 \times 1 \times 1$ 卷积层生成权重系数，用于骨骼特征的加权融合。

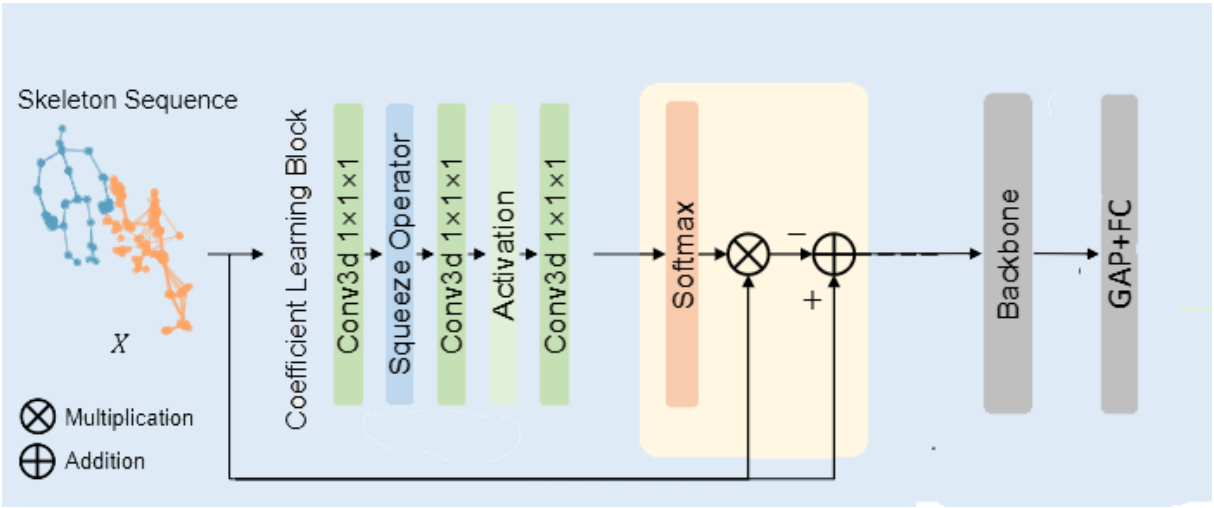


图 5　权重系数学习模块

权重系数学习模块的作用在于学习骨骼特征的权重系数，以实现特征的自适应加权。具体来说，这个模块可以根据输入骨骼序列的特征重要性动态地调整各个特征的权重，从而在后续处理中对关键特征给予更高的关注，忽略或减弱不重要的信息。这种自适应加权机制能够增强模型对不同姿态和动态变化的识别能力，提高对重要关节点或关键骨骼信息的捕捉能力，最终提升模型在动作识别或行为分析等任务中的性能。

4.3 实验效果

在赛题验证集上对比 Joint+Bone 两个模态融合的 CTRGCN 和 InfoGCN, 加入 CHASE 模块后识别准确率分别增加了 0.85% 和 0.75%。

Baseline	Acc. (%)
CTRGCN(2s)	45.30
InfoGCN(2s)	44.95
CTRGCN(2s)+**CHASE**	**46.15**
InfoGCN(2s)+**CHASE**	**45.70**

表 2　特征流比较结果

五、 整体模型架构——GCN 与 Transformer 集成学习

5.1 模型研究动机

本次比赛着眼于无人机视角下的人体行为识别。然而，由于无人机视角的独特性，数据采集角度常常存在较大变化，这对行为识别系统提出了更高的要求。骨架数据因其对光照、背景干扰的鲁棒性，在该场景中展现出显著优势。然而，传统的行为识别模型难以应对无人机视角的复杂背景和运动模糊，准确率和泛化能力较弱。

基于此，在此次比赛中我们团队使用图卷积网络 (GCN) 和 Transformer 结合的双分支结构为基础模型，以提升骨架数据的特征表达能力。在基础模型中，利用 GCN 捕捉骨架数据的空间结构特征，利用 Transformer 处理时间序列信息，从而形成多层次的特征表示，实现了在复杂视角和环境下的准确行为识别。

5.2 基础模型构建

在基础模型中，我们采用了两个并行的分支，分别为 GCN 分支和 Transformer 分支。两个分支协同处理骨架数据，以实现空间和时间特征的全面提取，其中：

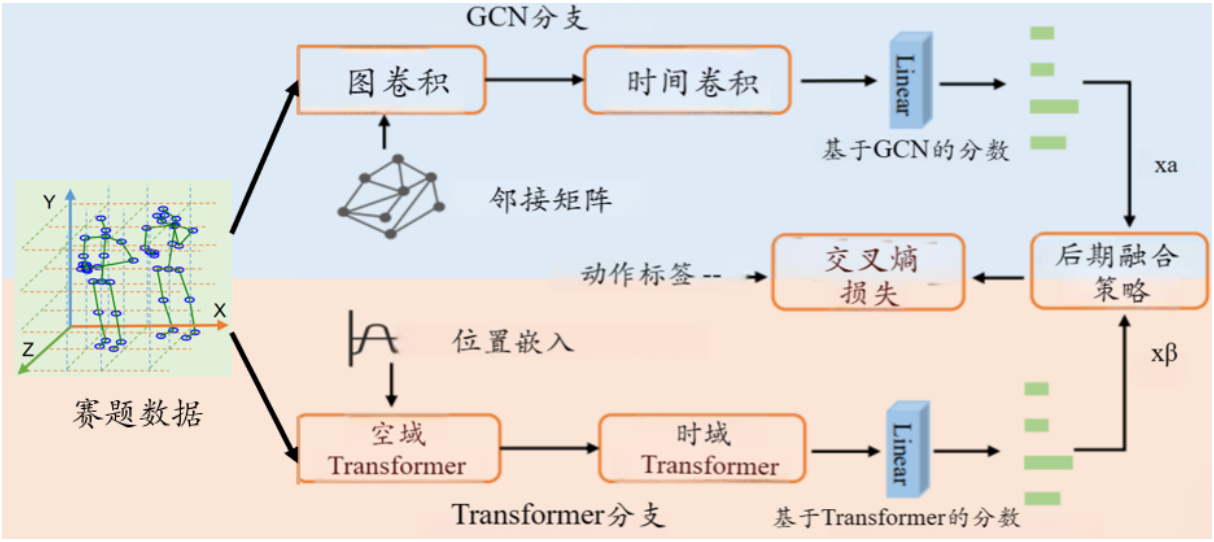


图 6　整体模型框架

GCN 分支利用图卷积网络处理骨架数据的空间特征。将骨架数据被建模为图结构，其中关节被视为图的节点，骨骼连接被视为边。GCN 通过邻接矩阵 A 和度矩阵 D 来表征各关节之间的空间依赖关系。在图卷积层中，骨架特征的更新公式如下：

$$H^{(l+1)} = \sigma\left(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{19}$$

其中，$H^{(l)}$ 表示第 l 层的特征矩阵，$W^{(l)}$ 为可学习的权重矩阵，σ 为激活函数。通过逐层的图卷积操作，模型逐步捕捉骨架的局部和全局结构特征。此外，为了在时间维度上建模动作序列的动态变化，GCN 分支还结合了时间卷积层，以捕捉动作发生的时间顺序。该分支的最终输出通过线性层映射为基于 GCN 的分数，并乘以加权系数 α，然后输入融合模块。

Transformer 分支利用 Transformer 模型处理骨架数据的时间序列信息。Transformer 通过自注意力机制捕捉骨架数据的全局依赖，能够有效处理序列中的长时依赖关系，从而生成时间特征。对于每一帧的骨架数据，首先通过位置编码来表示时间步长，然后在空间 Transformer 中进行处理。空间 Transformer 通过自注意力机制计算输入特征 X 的全局依赖关系，其计算公式为：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{20}$$

其中，$Q = XW^Q, K = XW^K, V = XW^V$ 分别表示查询、键和值的线性变换，W^Q、W^K、W^V 为可学习的权重矩阵，d_k 为缩放因子，用于稳定训练。随后，时间 Transformer 进一步编码空间特征中的时间依赖。该分支的最终输出通过线性层映射为基于 Transformer 的分数，并乘以系数 β，然后与 GCN 分支的输出在融合模块中组合。

在模型的最后阶段，采用后期融合策略。将 GCN 分支和 Transformer 分支的特征进行加权求和，生成最终的分类分数。融合公式为：

$$S = \alpha \cdot S_{\text{GCN}} + \beta \cdot S_{\text{Transformer}} \tag{21}$$

其中，S_{GCN} 和 $S_{\text{Transformer}}$ 分别表示 GCN 和 Transformer 分支的输出分数，α 和 β 为加权系数，用于平衡两者的影响。最终的分类结果通过交叉熵损失函数进行优化，以提高模型的识别准确率。