基于无人机的人体行为识别说明书

一、项目概述

本项目旨在设计并训练高效准确的人体行为识别模型，我们的核心模型是**动态边缘图卷积网络（DE-GCN）**，并在其框架内提出了新的**多模态融合**方案。该方案包含关节-骨骼（jbf）、关节-速度（jmf）以及关节-骨骼-速度（jbmf）等模态组合。此外，还融入了 **TeGCN、STTFormer 和 SkateFormer 模型**。对于初始数据方面，我们基于传统的关节（joint）、骨骼（bone）、运动（motion）三种基础模态，引入了**角度（angle）模态**，以更好地捕捉骨架中各关节的角度变化，从而增强模型对复杂行为的辨识能力。通过**多模态**与**多模型**的**联合融合**，有效补足了单一模态或模型的不足，提高了骨架数据的多维度特征表达能力，增强了模型的鲁棒性和精度。

二、模型及创新

1. 基于动态边缘图卷积网络（DE-GCN）的多流训练

对于该模型的基本模块如图 1 所示，我们发现该模型采用了双分支结构，而单流训练则是将相同的模态数据输入到两个分支中进行训练。在输入层面，我们基于传统的关节（joint）、骨骼（bone）、运动（motion）三种基础模态，引入了**角度（angle）模态**作为单流输入。

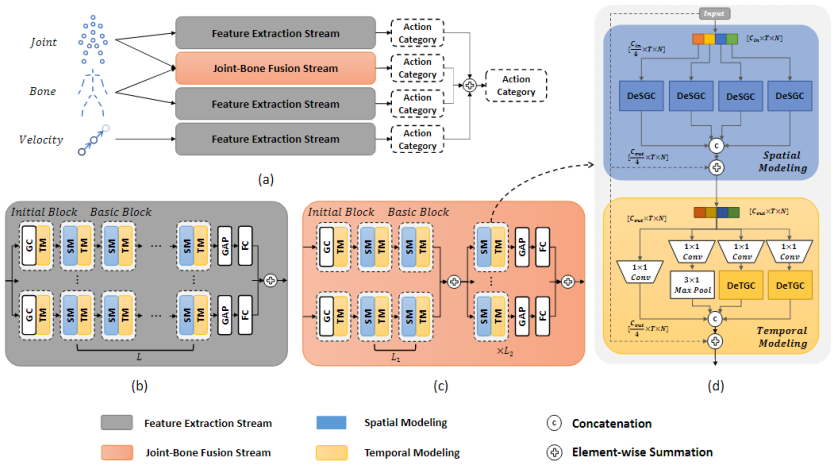

图 2.1.1 DEGCN 的基础模块及单流训练方法

基于这一发现，我们设计了**多流融合**的创新结构，并以图 2.1 的基本架构为基础构建了包含多个特征流的**多模态融合网络**，如双模态融合训练和三模态融合训练。具体而言，我增加了四种融合流：关节-骨骼融合流、关节-速度融合流、骨骼-速度融合流以及关节-骨骼-速度融合流。

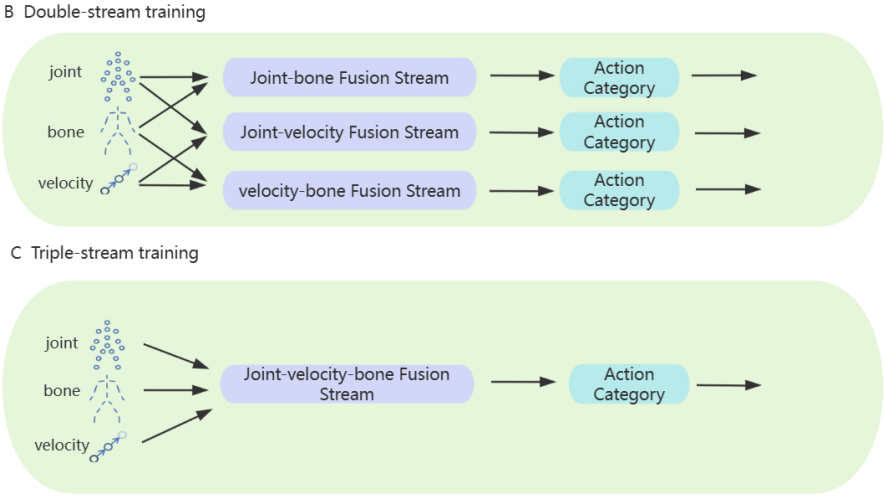

图 2.1.2 DEGCN 多模态训练方法示意图

图 2.1 中的"Feature Extraction Stream"（特征提取流）在单模态流中主要提取关节或骨骼的空间特征，而在多模态融合的创新流中，我将不同模态之间的信息相互结合，增强模型对动作细节的理解能力。

通过**多流设计**，模型可以在每个流中从不同模态的特征中提取有效信息，同时通过各流的组合进一步丰富模型的表达能力。相较于单流方法，该多模态融合结构显著提高了对复杂动作的识别性能，并在空间和时间维度上达到了更加精细和全面的特征建模。

2. TeGCN

TE-GCN（Temporal Enhanced Graph Convolutional Network）是一种针对基于骨架的动作识别任务设计的图卷积网络，主要通过**增强时序信息**的建模能力来提高对动态动作的识别效果。传统的图卷积网络（GCN）通常更注重空间结构的建模，例如骨架关节之间的空间连接关系，而 TE-GCN 通过将时序增强模块整合到 GCN 中，使模型能够在时间维度上捕捉更复杂的动作特征。

TE-GCN 的关键特性：

时序建模增强，空间-时间特征融合，骨架图结构

TE-GCN 的创新点：

1) **时序特征强化：**与普通的 GCN 模型相比，TE-GCN 在时间建模上有明显优势，它不仅在空间上进行卷积操作，还在时间维度上增强了特征提取，使得模型对复杂动作序列中的时序特征具有更高的敏感度。
2) **多层次特征融合：**TE-GCN 通常采用多层网络结构，每一层都结合空间和时间卷积，从低层次到高层次逐步抽象动作特征，从而更好地适应不同复杂度的动作模式。

总体而言，TE-GCN 是一种在骨架动作识别任务中加入时序建模增强的图卷积网络，通过空间和时间特征的协同建模，有效提高了复杂动态动作识别的性能。

3. SkateFormer

SkateFormer 是一种专门用于基于骨架的动作识别任务的模型，它利用了**自注意力机制和变换器架构，**以更加高效地捕捉复杂动作的空间和时序关系。该模型的设计灵感来自于变换器在自然语言处理和计算机视觉任务中的成功应用，特别是在需要处理序列数据的场景中展现了其优势。

SkateFormer 关键特性：

基于自注意力的空间-时间建模， 无需手工设计图结构， 高效的变换器架构

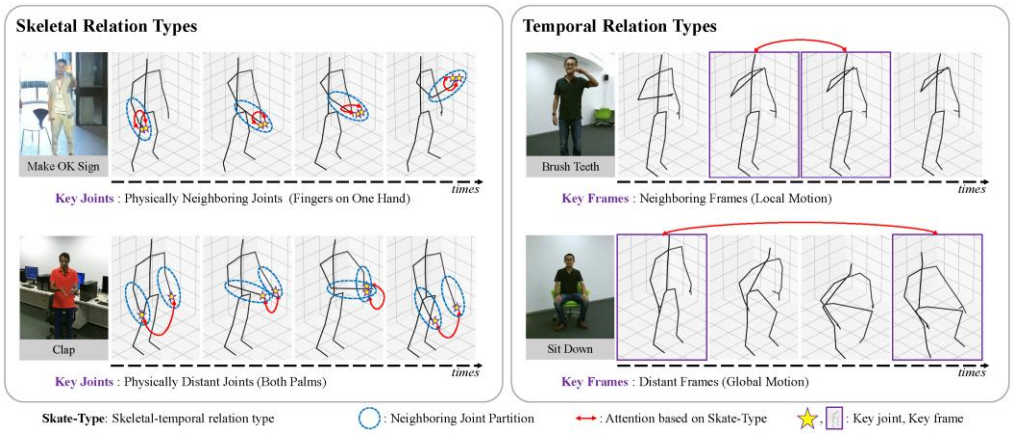

图 2.2 SkateFormer 的分区特定注意力策略

SkateFormer 的核心是将骨架数据表示为图结构，通过 Transformer 架构来处理空间-时间关系。具体而言，骨架关节（如人的肘部、膝盖等）被视为节点，关节之间的骨骼连接视为边。通过 Transformer 的自注意力机制，SkateFormer 可以有效地学习骨架关节之间的关系，不仅限于相邻的关节，还能够在整个骨架

中捕捉远距离的依赖性。这种方法允许模型关注动作序列中最显著的特征，同时对每一个时刻的空间关系进行更灵活的建模。

4. SttFormer

STTFormer（Spatial-Temporal Transformer）是一种专门用于人体行为识别的模型，它通过将空间-时间变换器结构应用到骨架数据中，有效地在空间和时间两个维度上进行独立建模。该模型设计的核心在于分别建模骨架数据中的空间依赖（不同关节点之间的相互关系）和时间依赖（动作在时间上的连续变化），从而在不依赖预设图结构的情况下，自主学习动作的关键特征。

STTFormer 模型采用分离的空间和时间注意力机制。在空间注意力模块上，通过自注意力机制在骨架的空间结构上进行建模，捕捉骨架中关节间的空间关系。这一模块不依赖预设的骨架图结构，而是基于数据动态学习不同关节点之间的关系。在时间注意力模块上，专注于建模动作在时间维度上的演变过程。通过自注意力机制捕捉关键帧之间的动态变化，从而增强对动作顺序和时序依赖的建模能力。

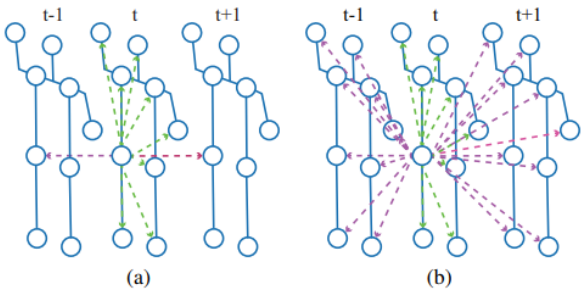

图 2.3 两种时空注意力机制

三、项目创新总结

1. 引入角度（angle）模态：

我们基于传统的关节（joint）、骨骼（bone）、运动（motion）三种基础模态，引入了**角度（angle）模态**，以更好地捕捉骨架中各关节的角度变化，从而增强模型对复杂行为的辨识能力。

2. 多模态融合：

我们通过关节-骨骼（jbf）、关节-速度（jmf）以及关节-骨骼-速度（jbmf）等模态组合，弥补单一模态的不足，减少信息丢失，提高鲁棒性和准确性。通过

融合空间、时间等多维度信息，模型可以更好地适应不同场景，增强泛化能力，进而在多种任务中表现出更高的决策精度和稳定性。

3. 多模型融合：

DEGCN、TEGCN、SkateFormer 和 STTFormer 是多种用于时序数据处理、人体行为识别、视频分析等任务的模型，每个模型都有其独特的优势和特点。将这些模型融合能够**互补**它们在不同方面的优势。例如，DEGCN 和 TEGCN 可以有效捕捉图结构中的时空依赖，SkateFormer 和 STTFormer 则能够利用 Transformer 的强大建模能力进一步优化空间和时间特征的提取。多模型融合能够综合这些优点，提高模型对时空依赖的建模能力，增强对复杂任务的处理效果。同时具有更好的泛化能力和鲁棒性，尤其在多模态数据或复杂环境中表现更佳。

四、消融实验

在表 1 中，分别在基准模型 DE-GCN 上对 Jbf、Jmf 和 Jbmf 在验证集上进行消融实验。

表 1 准确率对比

方法	Acc(%)
DE-GCN(baseline)	46.40
DE-GCN+Jbf	47.50
DE-GCN+Jmf	47.30
DE-GCN+Jbmf	47.20
DE-GCN+Jbf+Jmf	**47.80**
DE-GCN+Jbf+Jmf+JbmF	47.45

从表 1 可以看出，在我们提出的方法中 DE-GCN 和 Jbf 以及 Jmf 进行融合表现效果最好，与基础模型 DE-GCN 在验证集上提高了 1.4%。然而，加入 Jbmf 流是会表现比较差。

五、对比实验

通过在 DE-GCN 模型的基础上，增加 angle 模态、使用多模态组合，同时分别与 TeGCN、SkateFormer 和 STTFormer 两个模型进行多模型融合，对比结果如表 2 所示。

表 2 各模型、模态融合结果

模型	模态	准确率(%)	单模型融合(%)	多模型融合(%)
DE-GCN	joint	46.40	48.95	50.25
	Jbf	45.55		
	jmf	45.50		
	angle	42.95		
SkateFormer	joint	43.35	46.15	
	bone	42.40		
STTFormer	joint	41.80	47.5	
	motion	35.60		
TeGCN	Joint_bone	43.15	43.15	

六、总结

针对人体行为识别挑战，本项目提出了一种新颖的多模态与多模型集成解决方案。首要创新在于引入了 angle（角度）模态，该模态专注于解析骨架关节间的角度动态，进而增强了模型对复杂人体动作的深度理解。在此基础上，我们以 DE-GCN 模型为起点，通过融合 joint（关节）、bone（骨骼）及 motion（运动）等多种模态信息，极大地丰富了骨架数据的表达维度，从而显著提升了动作识别的精确度。

此外，项目还引入了 Skate-Former 模型，该模型凭借创新的分区骨骼-时间自注意力机制，优化了骨骼动态与时间序列的建模，使模型能够聚焦于关节间的关键动态特征。同时，STTFormer 模型通过采用时空元组编码与 Transformer 架构，精准捕捉了连续帧间关节的相互关联，进一步强化了模型对动作的辨识能力。

本项目的核心亮点在于对骨架数据的多模态集成处理，特别是 angle 模态的引入，为处理复杂的人体动作提供了新的视角。同时，通过巧妙地结合 DE-GCN、Skate-Former 与 STTFormer 等多个模型，本项目实现了模型性能的显著提升和鲁棒性的增强。这一综合方法有效克服了传统技术在捕捉关节依赖、处理复杂动作及优化计算效率方面的不足。

实验结果显示，本项目所提出的解决方案实现了识别准确率的显著提升，充分验证了框架的有效性。