

基于无人机的人体行为识别说明书

目录

- 1 问题导入..... 2
- 2. 模型介绍..... 3
 - 2.1 TE-GCN..... 3
 - 1. 2. 2 方法说明..... 3
 - 1. 2. 2 网络架构..... 5
 - 2.2 CTR-GCN..... 5
 - 2. 2. 1 特点..... 6
 - 2. 2. 2 方法实现..... 6
 - 2.3 TD-GCN..... 7
 - 2. 3. 1 特点..... 7
 - 2. 3. 2 方法实现..... 8
- 3. 创新性说明..... 9
 - 3.1 Kolmogorov-Arnold Networks (KAN) 9
 - 3. 1. 1 KAN 定义..... 9
 - 3. 1. 2 架构说明..... 9
 - 3.2 粒子群优化 (Particle Swarm Optimization, PSO) 10
 - 3. 2. 1 基本原理..... 10

1 问题导入

随着无人机技术的进步和快速发展，它们在安防领域的应用逐渐拓宽，尤其是在人体行为识别方面。无人机可以搭载多种传感器和摄像设备，实现对复杂场景中的人类行为实时监测和识别，为安全管理和紧急救援提供了新思路。然而，要实现这一目标，仍然面临许多技术挑战和创新机会。首先，高效的人体行为识别算法至关重要。无人机需要在不同的环境中准确识别出多种行为模式，如行走、奔跑、跌倒等，这对算法的实时性和准确性提出了更高要求。此外，如何合理配置无人机的传感器和摄像设备，使其在飞行中采集到清晰、可靠的行为数据，也是提高识别效果的重要因素。无人机的飞行特性还要求这些传感器具有良好的动态跟踪能力，以应对多角度、复杂光照和遮挡等条件。为了推动无人机在安防领域的智能化应用，许多科研与开发活动都集中于将深度学习、计算机视觉和模式识别等技术结合，设计高效的行为识别模型。这不仅为无人机提供了新的数据处理能力，也使其在应急和安防场景中的应用更加灵活、高效。同时，通过各方合作、技术交流，我们可以进一步提升无人机行为识别技术在实际应用中的可靠性和适应性，为构建智能化的社会安全监控系统提供强有力的技术支撑。

基于上述问题，我们通过深度学习、模式识别、图像处理等技术手段，设计并训练高效准确的人体行为识别模型。传统的基于手工特征的方法已经被数据驱动的深度学习方法所取代，这些方法可以直接从原始骨架数据中学习特征。循环神经网络（RNN）和卷积神经网络（CNN）被广泛用于以端到端的方式捕获关节间的空间和时间相关性。双向 RNN（bi-RNNs）和深度 LSTM（deep LSTMs），有效地捕获连续帧之间的时间依赖性。GCNs 通过执行标准卷积操作逐层聚合每个节点的邻域作为其新的隐藏表示，将 CNN 从图像推广到图结构，已经在多个应用中成功使用。对于基于骨架的行为识别任务，通过构建空间图和时间图，GCNs 能够有效捕获关节间的依赖关系。

2. 模型介绍

2.1 TE-GCN

在 TE-GCN 模型中，骨架数据被视为一个图结构，其中关节作为节点，骨骼作为边。这种表示方法使得骨架序列的时间动态性在行为识别任务中扮演了关键角色。

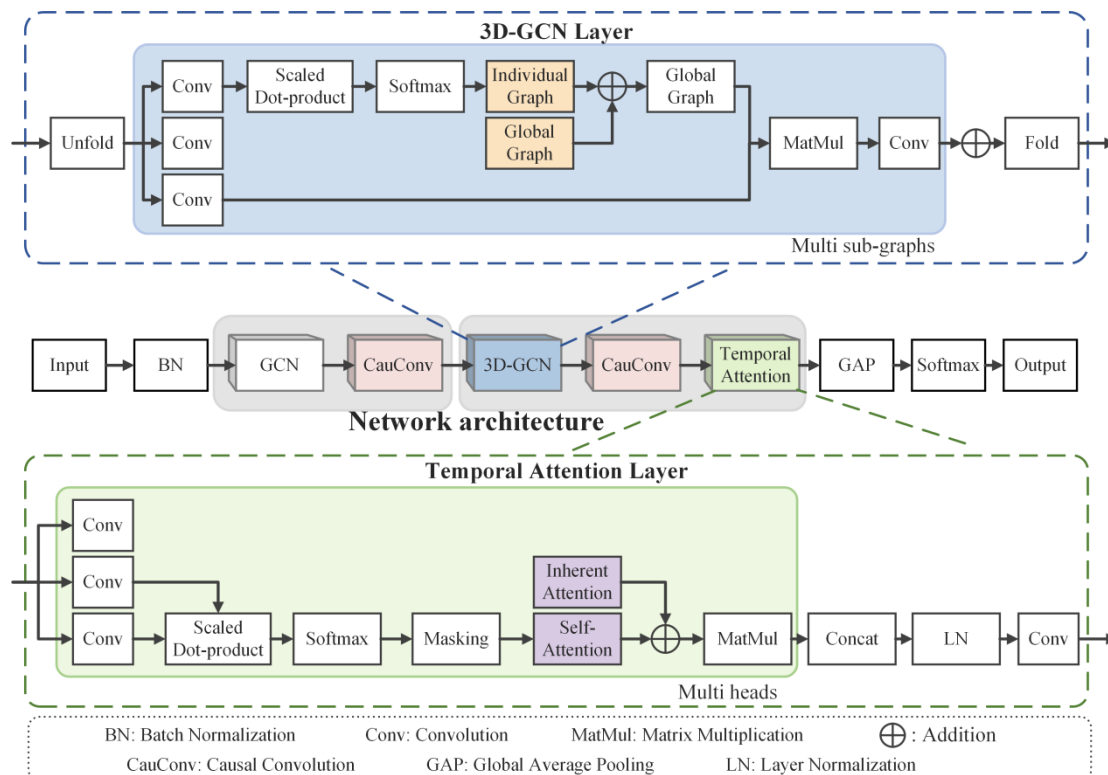


图 2. TE-GCN 的详细架构图。展示了 TE-GCN 模型的结构和不同层之间的连接。

1.2.2 方法说明

传统的基于图卷积网络（GCN）的方法通常通过叠加多层一维局部卷积来提取相邻时间步的时序关系。但是，这种方法在层数增多时可能会遭受信息稀释效应，导致一些关键的时间信息，尤其是那些在时间上不连续但对行为识别至关重要的信息丢失。在人类行为分析领域，关节点之间的依赖性不仅体现在空间维度上，时间维度的重要性也不容忽视。常规的 GCN 方法往往将空间和时间图分别处理，这样的分离处理限制了模型对非连续时间步之间关节点的空间-时间跨依赖性的捕捉能力。对于那些时间跨度较大的行为，如“穿脱夹克”和“跳跃”，标准的二维卷积由于感受野的限制，难以有效捕捉不同时间点之间的长期依赖关系。

为了克服这些挑战，TE-GCN 模型通过引入因果卷积层、跨空间-时间图卷积层和时间注意力层，有效地提高了人体行为识别的准确率，模型如图 2 所示。这些创新使得 TE-GCN 能够充分捕捉关节间复杂的空间-时间依赖性，并有效地识别持续时间较长的人类行为。

(1) 因果卷积 (Causal Convolution)：为了解决标准 2D 卷积中的时间信息泄漏问题，使用因果卷积通过在时间维度上对输入特征图进行零填充，确保卷积操作只涉及当前时间步之前的时间步，保持输入骨架的顺序信息。因果卷积遵循未来信息不泄露，输出长度与输入相同两大原则。因果卷积可以表示为：

$$f_{out} = \text{CauConv}(f_{in})$$

其中， f_{in} 为输入特征图， CauConv 为因果卷积操作， f_{out} 为输出特征图。在时间维度上，对输入特征图进行零填充，使用标准的 2D 卷积操作，通过调整填充和步长参数，使得卷积核只覆盖到当前时间步及其之前的步骤，去除额外维度，以确保输出特征图与输入特征图在时间维度上长度相同。

(2) 空间图卷积网络 (Spatial GCN)：为了捕捉人体骨架数据中关节（节点）之间的空间关系，有效地将图结构数据转化为可以被深度学习模型处理的形式，提高了骨架行为识别的性能。使用一个邻接矩阵 A 来表示全局图，该矩阵定义了人体关节之间的物理连接，即如果两个关节之间有直接的骨骼连接，则邻接矩阵中相应的元素为 1，否则为 0。引入自适应图的概念，通过自注意力机制学习每个样本独特的图结构，使得模型能够更加灵活地捕捉非物理连接关节之间的空间依赖性。

$$f_{out} = \sum_s^S \text{Conv}(f_{in} \cdot A_s)$$

$$f_{out} = \sum_s^S \text{Conv}(f_{in} \cdot (B_s + \alpha C_s))$$

其中， f_{in} 是输入特征图，包含每个关节的特征。 A 是邻接矩阵。 B_s 是全局图的邻接矩阵。 C_s 是基于自注意力机制学习的个体图的邻接矩阵。 α 是平衡全局图和个体图影响的可学习参数。在训练过程中，全局图的元素被参数化并更新，而个体图则为每个样本学习一个独特的结构。

(3) 3D-GCN：将单帧的空间图扩展到一个时间窗口 τ 内的体积图，在这个体积图中，每个关节在时间窗口内都有 τ 个对应的节点，从而允许模型捕捉关节在空间和时间上的关系。将空间图的邻接矩阵 A 扩展到时间窗口 τ ，构建全局图的初始邻接矩阵 \hat{B}_k 。对于个体图，通过滑动窗口操作将输入特征图展开，然后应用图聚合操作：

$$f_{out} = \text{Unflod}(f_{in})$$

$$f_{out} = \sum_s^S \text{Conv}(f_{in} \cdot (\hat{B}_s + \alpha \hat{C}_s))$$

其中， \hat{C}_s 是学习得到的个体图的邻接矩阵。在图聚合操作后，使用折叠操作将额外的维度合并回去，以保持特征图的形状不变： $f_{out} = \text{Unflod}(f_{out})$ 。在训练过程中，全局图和个体图的邻接矩阵会被参数化并更新，以适应不同的数据样本。

(4) 时间注意力 (Temporal Attention)：为了捕获骨架序列数据中不同时间步之间的长期依赖性,并使模型能够关注于重要的时间步骤。在时间注意力层采用了多头注意力机制,使模型在不同的表示子空间中同时关注不同位置的信息。输入特征图 f_{in} 通过 1×1 的 2D 卷积被映射到 $Query(Q)$ 、 $Key(K)$ 和 $Value(V)$ ，通过使用缩放点积操作计算注意力权重矩阵：

$$S = softmax\left(\frac{Q^T \cdot K}{\sqrt{C_\epsilon}}\right)$$

C_ϵ 缩放因子用于防止训练中的梯度消失问题。为了保持时间的因果性,即当前时间步不能依赖于未来的信息,使用一个上三角的因果掩码 M 来屏蔽未来时间步的影响,内在注意力 I 表示一个时间步在网络后续层中对自身的固有影响。将自注意力权重矩阵 S 与因果掩码 M 相乘,然后与内在注意力矩阵 I 相加,最后乘以 $Value(V)$ 得到最终的注意力输出：

$$f_{out} = Cinv(V \cdot (\gamma \odot M \odot S + I))$$

通过这种方式,TE-GCN 模型中的时间注意力层能够提高对时间动态的建模能力,增强模型对骨架行为识别任务中重要时间信息的捕捉能力。

1.2.2 网络架构

TE-GCN 模型的网络架构由总共 10 层组成,这些层被分为两个部分:前 5 层为 CAGCN 层,后 5 层为 TE-GCN 层。CAGCN 层由图卷积网络 (GCN) 层和 9×1 核的因果卷积层构成,而 TE-GCN 层则包含 3D-GCN 层、 9×1 核的因果卷积层以及时间注意力层。这种结构设计使得模型能够有效地捕获骨架数据中的空间和时间特征,并增强对长期依赖性的建模能力。

在网络的第 5 层和第 8 层,引入了步长为 2 的步长步骤,压缩时间维度上的长度 T ,以降低计算复杂度并提取更高层次的特征。为了进一步稳定训练过程并减轻梯度消失或爆炸的问题,网络在起始位置使用了批量归一化 (BN) 层,并且在每个卷积层之后都跟随着 BN 层和 ReLU 激活层,以实现特征的归一化和非线性变换。

此外,为了提高模型的训练稳定性和性能,网络中还加入了残差连接,这有助于直接传递梯度,从而减轻深层网络中常见的梯度传播问题。最终,经过一系列复杂的特征提取和转换后,网络输出的向量被送入 SoftMax 分类器,以进行动作类别的推断。这样的网络设计不仅提高了骨架行为识别的准确性,还增强了模型对不同动作之间细微差别的识别能力。

2.2 CTR-GCN

Channel-wise Topology Refinement Graph Convolution (CTR-GCN) 是一种专门为骨架 (skeleton) 动作识别设计的图卷积网络结构,目的是在动作识别任务中更有效地建模人体关节之间的复杂关系。通过引入基于通道的拓扑结构自适应调整机制,使得模型能够更灵

活地捕捉不同动作中的特征。CTR-GCN 的整体结构设计使其在骨架动作识别任务中表现优异，特别是在动态场景或复杂动作识别中更具优势。

2.2.1 特点

CTR-GCN 通过一系列创新优化了动作识别任务的处理。首先，它采用基于通道的拓扑结构优化，利用通道注意力机制动态调整关节点间的拓扑关系，以适应不同类型的动作。其次，通过自适应拓扑学习模块，模型能够在训练过程中自动调整图结构，自适应地学习节点连接的强弱，从而捕捉关节点间更细致的关系。此外，CTR-GCN 结合了时空图卷积网络的思想，在时间维度和空间维度上联合建模关节点关系，以更好地理解动作的时间动态。最后，为了适应长时间序列数据，CTR-GCN 降低了计算复杂度，使其更适合处理大规模动作识别任务。

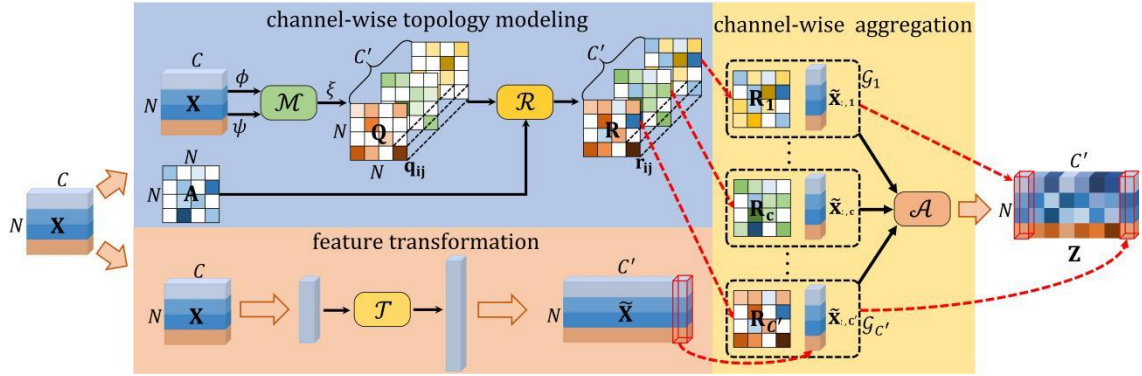


图 1. 所提出的通道级拓扑细化图卷积的框架。信道拓扑建模通过推断的信道特定相关性改进了可训练的共享拓扑。特征转换旨在将输入特征转换为高级表示。最终，通过信道级聚合的方法获得了输出特征。

2.2.2 方法实现

(1) 拓扑共享的图卷积：正规拓扑共享图卷积利用权值 ω 进行特征变换，并通过 a_{ij} 对 v_i 的邻域顶点进行聚合表示，更新其表示，表示为 z_i

$$z_i = \sum_{j \in N(v_i)} a_{ij} x_j W$$

对于静态方法， a_{ij} 被手动定义或设置为可训练的参数。对于动态方法， a_{ij} 通常由基于输入样本的模型生成。CTR-GC 包含三个部分：(1) 由变换函数 $T(\bullet)$ 完成的特征变换；(2) 由相关建模函数 $M(\bullet)$ 和细化函数 $R(\bullet)$ 组成的通道拓扑建模；(3) 通道聚合，由聚合函数 $A(\bullet)$ 完成。特征转换的目的是通过 $T(\bullet)$ 将输入特征转换为高级表示。采用一个简单的线性变换作为拓扑共享图卷积，其表示为：

$$\tilde{X} = \tau(X) = XW$$

其中， W 为权重矩阵。

(2) 通道拓扑建模：邻接矩阵被用作所有通道的共享拓扑，并通过反向传播进行学习。具体来说，首先使用相关建模函数 M 来建模顶点之间的通道级相关性。为了降低计算成本，

在将输入特征发送到 M 之前，利用线性变换 ϕ 和 ψ 来降低特征维数。给定一对顶点 (v_i, v_j) 及其对应的特征 (x_i, x_j) ，设计了两个简单而有效的相关建模函数。

(3) 通道聚合：给定细化的信道级拓扑 R 和高级特征，CTR-GC 以信道级的方式聚合特征。具体来说，CTR-GC 为每个通道构造一个通道图，具有相应的细化拓扑和特征。每个信道图都反映了在某一类型的运动特征下的顶点之间的关系。因此，对每个信道图进行特征聚合，通过将所有信道图的输出特征连接起来，得到最终的输出 Z ，其表示为：

$$Z = A(\tilde{X}, R) = [R_1 \tilde{x}_{:,1} \| R_2 \tilde{x}_{:,2} \| \cdots R_{C'} \tilde{x}_{:,C'}]$$

在整个过程中，信道特定相关性 Q 的推断依赖于输入样本。因此，所提出的 CTR-GC 是一种动态图卷积，它随不同的输入样本自适应变化。

2.3 TD-GCN

TD-GCN (Temporal Decoupling Graph Convolutional Network, 时序解耦图卷积网络) 是为处理动态图数据而设计的一种图卷积网络 (GCN) 的变种，尤其在交通流预测等领域中表现突出。它的核心思想是时空解耦，即将时序依赖和空间结构分开建模，以提高对动态变化数据的处理能力。

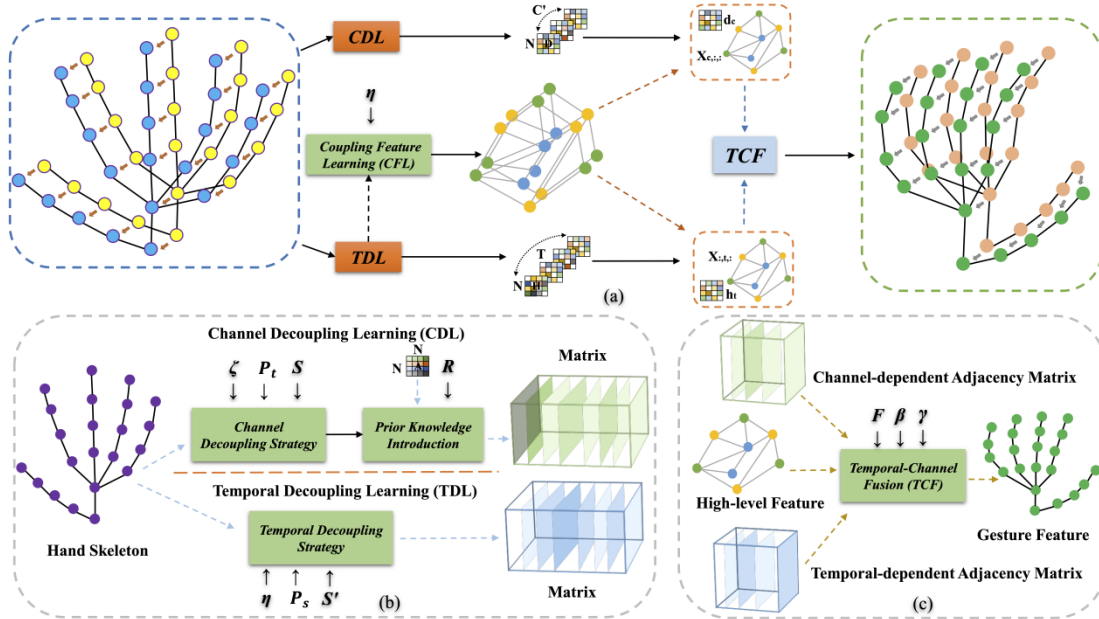


图 3. 时间解耦图卷积的框架。(a)TD-GC 使用耦合特征学习 (CFL) 模块从骨架序列中学习高级特征。(b) 信道解耦学习 (CDL) 模块计算信道相关的邻接矩阵，时间解耦学习 (TDL) 模块计算时间相关的邻接矩阵。(c)利用时间通道融合 (TCF) 模块来融合和更新关节顶点的高级时空特征。

2.3.1 特点

TD-GCN 模型通过时空解耦策略，与传统 GCN 模型不同，它将时序信息和图结构（空间信息）分开处理，允许每个维度独立建模，更有效地捕捉交通流等动态系统中的时空依赖

关系。模型引入了时序卷积模块来直接处理时间序列数据，避免了传统递归神经网络（如 LSTM）在处理长时间序列时可能出现的梯度消失或爆炸问题。同时，空间卷积模块利用图卷积操作捕捉不同路段或区域间的空间依赖性。这种设计使 TD-GCN 能够适应交通数据的时变性，通过时序解耦和独立建模的方法，提升了模型在动态环境中的预测表现。

2.3.2 方法实现

TD-GC 和时间卷积，构建了基于骨架的手势识别任务的 TD-GCN。构建了 ST-GCN 的主干，在图卷积模块中，融合了三个 TD-GC 层的输出，当利用先验知识引入一个预定义的矩阵 \mathbf{a} 时，每一个层都是不同的。在时态卷积模块方面，采用了一个类似于 GoogLeNet 的框架，并设计了一个多分支的时态卷积。所提出的时间解耦图卷积的框架如图 2 所示。其中，TD-GC 通过以下方式更新 N 个顶点的特征：

$$X' = F \left[M(D, \tilde{X}), M(H, \tilde{X}) \right]$$

D 和 H 分别是通道依赖和帧依赖的邻接矩阵，表示手关节之间的关联。 F 是一个在时间和通道维度上聚合关节特征的融合函数，其表述为：

$$F(y, z) = y \cdot \beta + z \cdot \gamma$$

通过乘法运算·为信道特征和时间特征分配自适应权值。符号 β 和 γ 代表两个可学习的参数，它们将在训练过程中进行优化。 M 是一种特殊的矩阵乘法运算，具有串联运算利用提取的高级特征和手关节之间的关联矩阵来更新骨架的特征。

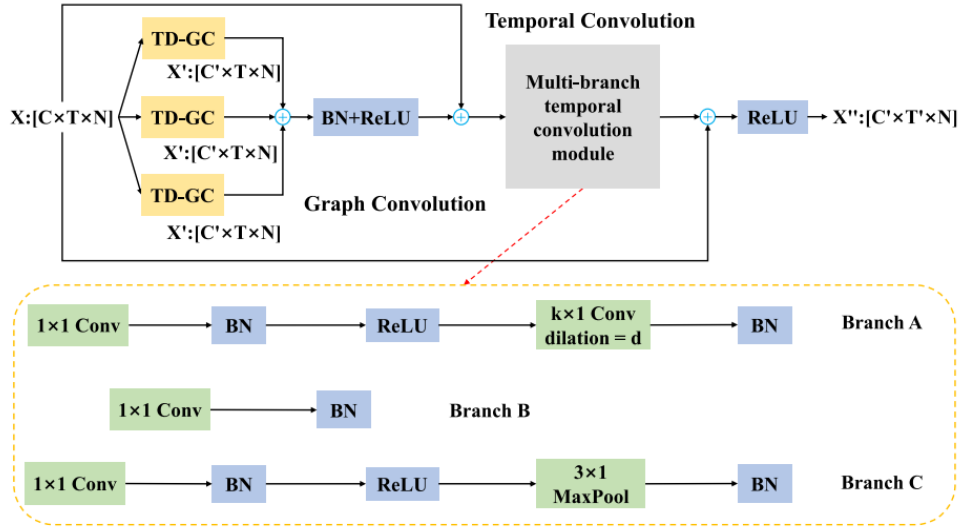


图 4. TD-GCN 的基本块。基本块的图卷积结合了三个 TD-GC 块的输出，每个 TD-GC 块基于人体骨架的物理连接，在 CDL 模块中预定义了三个不同的邻接矩阵 \mathbf{A} 。

如图 4 所示，时间卷积模块包含三种基本分支（即 A、B、C），并将这些分支的结果连接起来得到输出。分支 A 主要由两个卷积层组成，经过归一化和非线性激活后，第一个 1×1

卷积的输出作为第二个 $k \times 1$ 卷积的输入，其中 k 表示卷积核大小。分支 B 是一个简单的具有归一化功能的 1×1 卷积层。分支 C 主要包括卷积层和最大池化层。与分支 A 相似，卷积层和最大池化层之间存在归一化和非线性激活。TD-GCN 总共有 11 个块，前 10 个块是图 4 中所示的基本块，最后一个块包含一个全局平均池和一个用于预测动作类的 softmax 分类器。此外，对于基于骨架的手势识别和动作识别，使用多流结构来融合预测分数。

3. 创新性说明

在本节中，我们简要概述了模型的几个关键创新点。(1) 将 CTRGCN 和 TDGCN 网络的代码移植到 TEGCN 代码框架中。通过这一过程，我们有效地利用了 CTRGCN 的通道拓扑细化图卷积技术以及 TDGCN 的动态图卷积技术。在 pre 后处理阶段，整合不同模型和不同模态的训练数据，实现了优势互补。(2) 采用了 KAN 网络进行最终的分线性变换，这一改进使得模型更加简洁高效，同时提高了性能。(3) 利用粒子群优化算法实现了自动权重搜索，这在验证集（val 集）中帮助我们找到了最优参数。这些创新性操作共同推动了模型性能的显著提升。

3.1 Kolmogorov-Arnold Networks(KAN)

3.1.1 KAN 定义

Kolmogorov-Arnold Networks（简称 KAN）是一种新型神经网络架构，与传统使用固定激活函数的神经网络不同，KAN 在网络的边缘采用可学习的激活函数。这种设计使得 KAN 中每个权重参数都可以被一个单变量函数替换，这些函数通常以样条函数形式参数化，从而提供了极高的灵活性，并能够用更少的参数来模拟复杂的函数，增强了模型的可解释性。

KAN 理论基础来源于多变量连续函数的概念，根据这一理论，任何多变量连续函数 f 都可以表示为有限个单变量连续函数的组合。

$$f(x) = f(x_1, \dots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^n (x_p) \right)$$

该定理允许将复杂的高维函数分解为简单的一维函数组合。通过集中优化这些一维函数而非整个多变量空间，KAN 大幅降低了实现精确建模所需的复杂性和参数数量，并且函数简单易于解释。

3.1.2 架构说明

传统的权重参数在网络的边缘被单变量函数参数所取代。在 KAN 中，每个节点汇总这些函数的输出时不进行任何非线性变换，这与 MLP 中的做法（线性变换后跟非线性激活函数）形成对比。

$$KAN(x) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \dots \circ \Phi_1 \circ \Phi_0)$$

$$MLP(x) = (W_{L-1} \circ \sigma \circ W_{L-2} \circ \sigma \dots \circ W_1 \circ \sigma \circ W_0 \circ \sigma)$$

样条函数构成了 KAN 学习机制的核心部分，它们替代了神经网络中常规的权重参数。样条函数的灵活性允许它们通过改变形状来适当地模拟数据中的复杂关联，进而减少近似误差，这提升了网络从高维数据中学习细致模式的能力。KAN 中样条的通用公式可以用 B-样条来表示：

$$simple(x) = \sum_i c_i B_i(x)$$

这里， $spline(x)$ 表示样条函数。 c_i 是训练期间优化的系数，而是定义在网格上的 B-样条基函数。网格点定义了每个基函数 b_i 活跃并显著影响形状和平滑度的区间。更多的网格意味着更多的控制和更高的精度，同时也意味着需要学习更多的参数。

KAN 网络的最终分类线性变换模块可以显著提升模型的效率和性能。通过将复杂的非线性计算简化为一层线性变换，KAN 在保证分类准确性的前提下，减少了计算资源的消耗。此外，KAN 网络能够有效适应数据的分布，精确提取关键特征，提升了整体的模型泛化能力。与传统的复杂网络结构相比，KAN 在精度和效率之间实现了良好的平衡，因此非常适合在资源受限的环境中部署。

3.2 粒子群优化 (Particle Swarm Optimization, PSO)

粒子群优化算法是一种基于群体智能的优化算法，它通过模拟鸟群或鱼群的社会行为来寻找最优解。PSO 算法以其简单、并行、鲁棒和易于实现的特点，在许多优化问题中得到了广泛应用。

3.2.1 基本原理

PSO 模拟了一个粒子群在搜索空间中的飞行行为。每个粒子代表了问题的一个潜在解，它们在搜索空间中飞行，通过跟踪两个“极值”来更新自己的位置和速度

个体极值 (pBest)：每个粒子所找到的最优解。

全局极值 (gBest)：整个粒子群中所有粒子所找到的最优解。

粒子根据自身的经验（个体极值）和群体的经验（全局极值）来调整飞行方向和速度，以寻找问题的最优解。PSO 算法的主要步骤如图 5 所示。

假设在一个 D 维的目标搜索空间中，有 N 个粒子组成一个群落，其中第 i 个粒子表示为一个 D 维的向量（位置）：

$$X_i = (x_{i1}, x_{i2}, \dots, x_{iD}), i = 1, 2, \dots, N$$

第 i 个粒子自己的“飞行”速度也是一个 D 维的向量，记为：

$$V_i = (v_{i1}, v_{i2}, \dots, v_{iD}), i = 1, 2, \dots, N$$

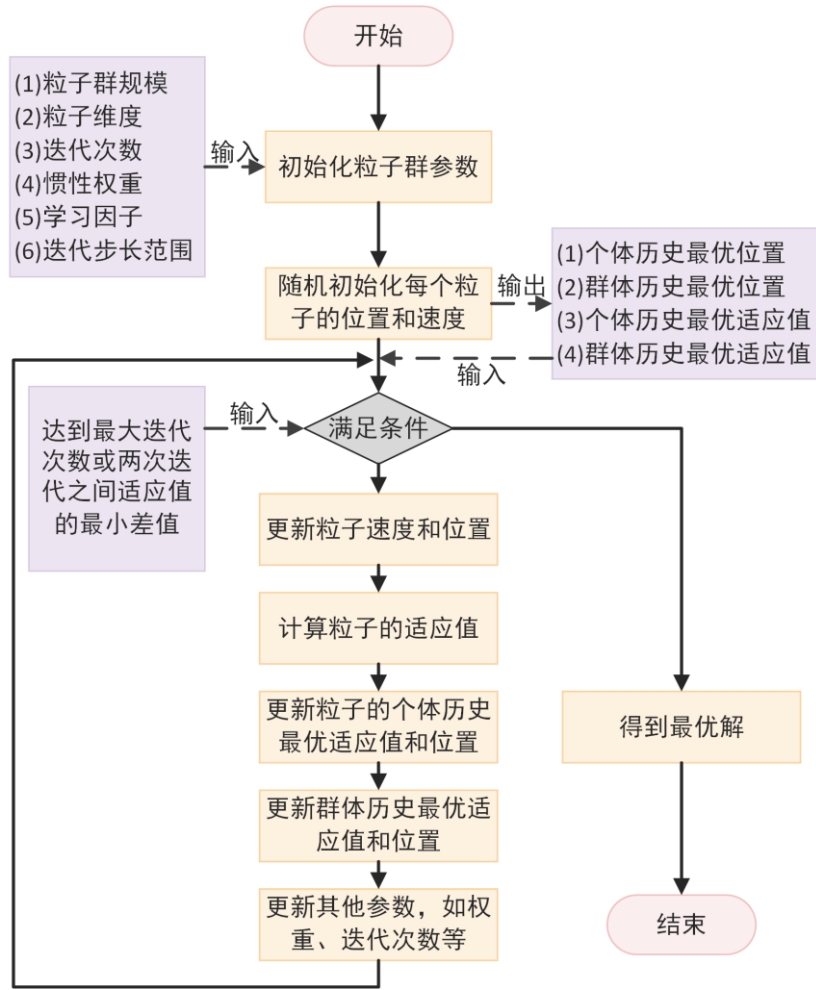


图 5.粒子群算法的流程图

在第 t 代的第 i 个粒子向第 $t+1$ 代进化时, 根据如下式子更新:

$$v_{ij}(t+1) = wv_{ij}(t) + c_1r_1(t)[p_{ij}(t) - x_{ij}(t)] + c_2r_2(t)[p_{gj}(t) - x_{ij}(t)]$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1)$$

第 t 代到第 $t+1$ 代迭代, 既要有个体自己的速度又要往全局处靠拢, $v_{ij}(t+1)$ 表示此次飞行中的距离和大小, 有三项表示, $wv_{ij}(t)$ 表示自己的惯性速度; $[p_{ij}(t) - x_{ij}(t)]$ 表示去他迭代这么多次个体极值飞的大小和方向; $[p_{gj}(t) - x_{ij}(t)]$ 表示去它第 t 代全局最优飞的大小和方向。 $r_1(t)$ 、 $r_2(t)$ 表示 0-1 的随机数, w 为惯性权重, 表示上一代粒子的速度对当代粒子的速度的影响, 值越大, 探索新区域的能力越强, 全局寻优能力越强, 但是局部寻优能力越弱。 c_1 为个体的学习系数, 表示粒子下一步动作来源于自身经验部分所占的权重, c_2 为全体的学习系数, 表示粒子下一步动作来源于其它粒子经验部分所占的权重。