

# intro

---

ST-GCN定义了一个时间卷积策略，通过在相邻连续帧之间连接相同的关节点来构造卷积运算，因此相邻连续帧就构成了一个时间邻域，邻域范围大小是由时间卷积核的大小决定。后续大部分基于ST-GCN的改进工作，主要专注于如何更好地提取空间域特征，而对于提取时间域信息依旧延续着ST-GCN的策略，例如AS-GCN, 2s-AGCN, CTR-GCN 等等。

我们认为这种时间卷积策略有待改进的地方，之前的时间卷积策略的时间感受野范围是固定的，由时间卷积的卷积核大小决定，例如ST-GCN, AS-GCN等每次时间卷积考虑相邻9帧关节点信息，2s-AGCN等考虑相邻5帧信息，时间感受野大小是固定的。CTR-GCN也提出了的多尺度时间建模的概念，但是只是同时考虑了相邻5帧信息和相邻3帧信息的局部信息。（Shift-GCN 的adaptive temporal shift 怎么写还不知道，TSM论文中有提到In-place shift 缺点是空间特征学习能力的丧失，时移后，存储在移位通道中的部分信息将丢失给当前帧。具体需要再看看）

我们认为，人类可以通过适应性地关注具有不同时间尺度的时间片段来区分不同动作，所以在动作序列中不同的时间尺度动作序列反映出不同侧重点的运动信息，甚至是不同的动作类型，他们相互补充，更有利于进行动作识别。（论文：Grouped Spatial-Temporal Aggregation for Efficient Action Recognition, Learning multi-temporal-scale deep information for action recognition, Context-Sensitive Temporal Feature Learning for Gait Recognition）

因此本文将从帧间差，短时，长时，三个时间尺度提取时间域特征。帧间差通过建立相邻两帧特征图之间的点对点差值图，能够将动态关节点和静态关节点区分开来，从而更加关注帧间的动态特征。（AAI2021 paper "Learning Comprehensive Motion Representation for Action Recognition"）

短时特征捕获局部时间上下文线索，这些线索对时间位置敏感，包含了重要的局部细粒度时间信息，有利于对微动模式进行建模。

长期特征代表所有帧的运动特征，揭示了全局动作周期，这些周期对于时间位置是不变的。