基于TE-GCN的人体行为识别模型

**一、作品简介**

为了充分考虑动作建模中的时间因素，我们运用了一种新型的时间增强图卷积网络TE-GCN。首先，引入因果卷积层以确保每个时间步长中没有未来信息泄漏，并保持输入顺序信息。其次，提出了一个新颖的跨空间-时间图卷积层，该层将自适应图从空间域扩展到时间域，以捕获关节之间的局部跨空间-时间依赖性。第三，设计了一个时间注意力层来增强对长期时间依赖性的建模能力，帮助网络直接聚焦于重要的时间步长。

我们的贡献可以总结如下：

（1）我们引入因果卷积以确保不泄露未来信息，从而保持订单相关动作识别的排序信息。

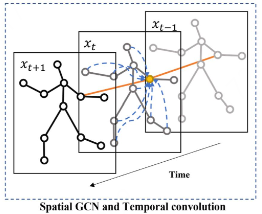
（2）我们提出了一种3DGCN层，以适应性地捕获复杂的局部跨空间和时间的联合依赖关系。

（3）我们提出TA层来捕获长距离的时序依赖性，使模型能够关注重要的时间步而不受接收域限制。

**二、TE-GCN模型介绍**

**（1）GCN基本原理**

现有的GCN方法主要集中在增强空间维度的图结构建模能力，而在时间维度上，则采用标准二维卷积来捕获不同时间步长下同一空间节点之间的相关性，如图所示：



在基于GCN的方法中，空间GCN和二维卷积交替使用以模拟有限时间接收场范围内的跨时空相关性。为了增强时域图结构的建模能力，我们从空间维度扩展了GCN到时间维度，称为3DGCN。这种时空图结构可以直接表达局部跨时空的相关性

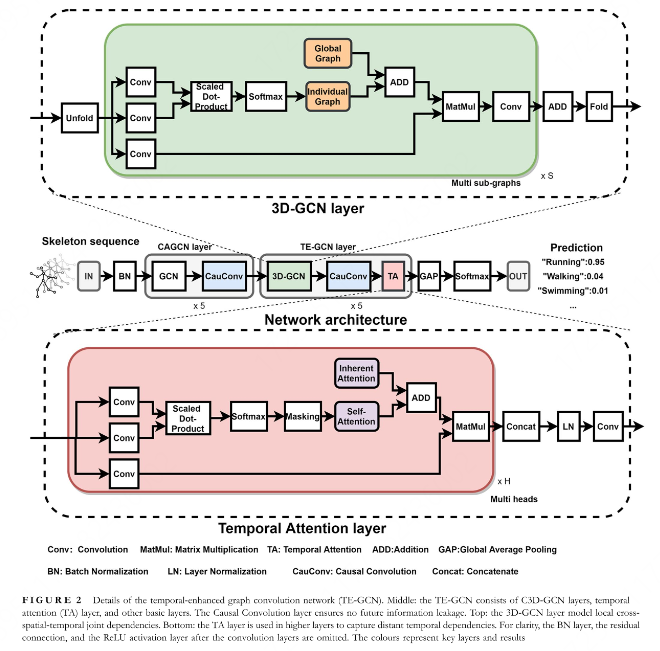
**（2）网络架构**

如图2所示，网络用于动作识别（分类）任务，在该任务中，骨架序列输入到网络中，并输出分类结果。在我们的网络中，包括了五个CAGCN层和五个TE GNC层的十层结构。CAGCN层由一个GCN层和一个具有9×1核的CauConv层组成。TE GNC层由时间窗口t的3DGNC层、具有9×1核的CauConv层以及TA层组成。我们将TE GNC层放置在更高的层次上以捕获更多的语义信息进行动作识别。

第10层的输出通道数为64，64，64，64，128，128，128，256，256和256。在第五层和第八层中，有步长（步长大小是2）来压缩时间长度T。

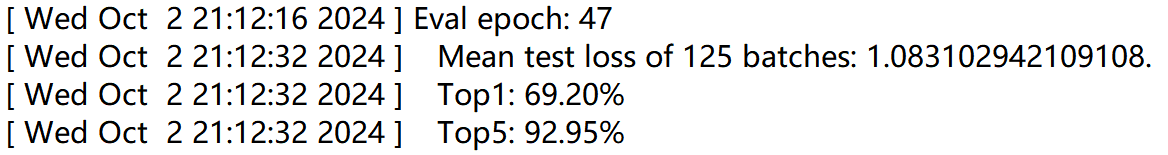
在开始时，使用批归一化（BN）层来标准化输入数据。卷积层后接一个BN层和一个ReLU激活层。残差分支层被添加以稳定训练过程并促进梯度传播。特别地，如果输入输出张量形状不一致，则我们使用额外的1× 1二维卷积确保残差捷径中的元素级加法接收具有相同形状的特征图。最后，这个向量发送到SoftMax分类器进行动作类别的推断。

。



**（3）实施细节**

我们的模型使用随机梯度下降（SGD）(动量为0.9和权重衰减为0.0001)进行训练，并且批量大小为32。我们选择交叉熵损失作为动作分类损失。对于NTURGB + D和UAV人类数据集，我们进行了50个周期的训练，在第30个和第40个周期时将学习率除以10。所有骨骼序列均零填充至T = 300帧。遵循预处理策略，输入数据然后经过归一化和翻译处理。对于赛题数据集，我们进行50个周期的训练。识别准确率可达到60%左右。



**三、参考文献**

1.Cao，Z.等：使用部分亲和力场的实时多人2D姿态估计。在：第30届IEEE计算机视觉与模式识别会议论文集（CVPR），卷2017，页1302-1310 (2017)

Atwood，J。，Towsley，D。：扩散卷积神经网络。在：神经信息处理系统进展（NIPS），第2001-2009页（2016）

3.Niepert，M。，Ahmad，M。，Kutzkov，K。：学习图卷积神经网络。在第33届国际机器学习大会ICML 2016上发表的论文，卷4，页2958-2967（2016）

Kipf，T.N。，Welling，M：基于图卷积网络的半监督分类。在第5届国际表示学习会议上。ICLR 2017——会议轨道论文集，第1-14页（2019）

5.1 Monti，F.等：使用混合模型CNN在图和曲面上的几何深度学习。 在：IEEE计算机视觉与模式识别会议论文集（CVPR），卷2017，第5425-5434页（2017）

Hamilton，W.L。，Ying，R。，Leskovec，J。：在大型图上进行归纳表示学习。在神经信息处理系统（NIPS）中，第2017年12月卷，第1025-1035页（2017）。