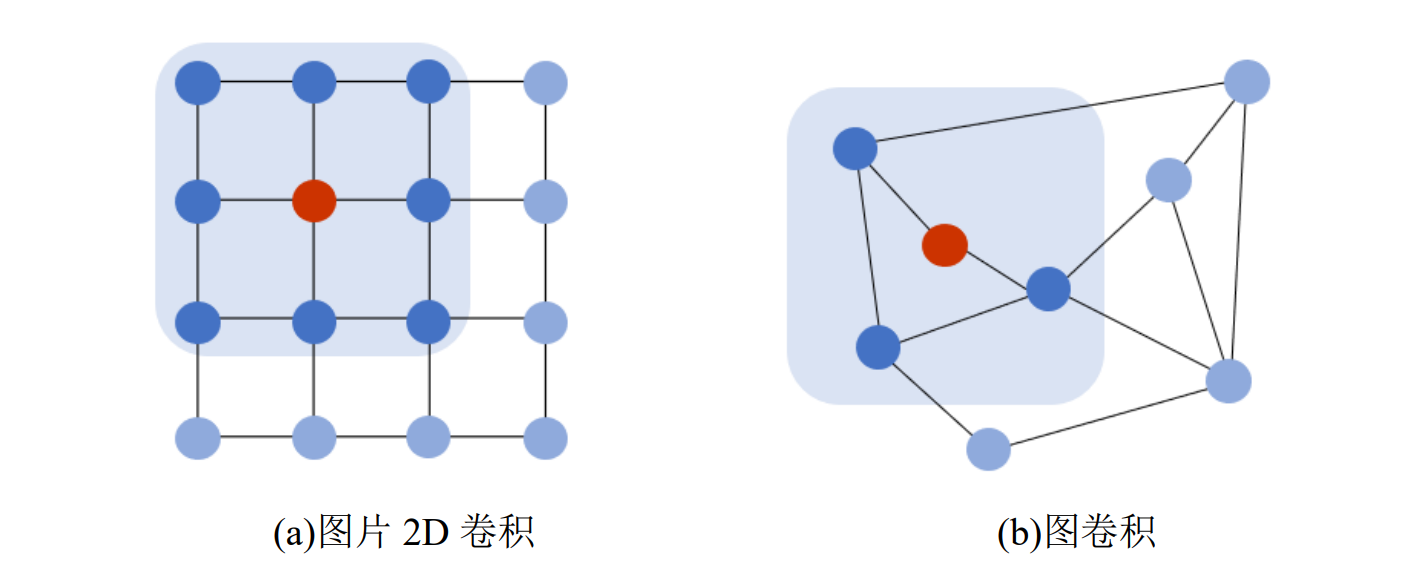
2级标题GCN图卷积网络

在介绍图卷积网络之前，首先来看一下一般的卷积神经网络CNN是怎样工作的。一般CNN处理的是2D的图像或者特征图，它的输出也是2D特征图。假定一个卷积的步长为1，填充为0，卷积核大小为，输入的特征图为，其通道大小为，则在空间位置处单个通道的输出值可以写成：

其中为取样函数，表示以输入的位置得到其邻接值得位置，表示权重函数，其取样位置与对应。

如图x-x所示，将卷积网络得思想推广到图数据结构上[1]，可以得到图神经卷积网络GCN。



图x-x 图卷积示意图

GCN中将特征映射改写为，表示每个节点都有一个向量表示其特征。将取样函数定义为，即，为与的距离，定义距离1为相隔一个节点，使用不同的分区策略，将得到个不同的分区，权重函数可以改写为，则图神经网络公式改写如下：

a

其中为归一化项，等于每个子集中邻接节点的个数，加入这一项是为了在求和时平衡每个子集的输出。

式（GCN理论公式）是GCN的解析式，在计算机中利用解析式解问题是比较困难的，所以实际运用中用多项式对问题进行拟合[2]。接下来对其拟合式推论过程进行大致说明。

用邻接矩阵表示图结构，表示度矩阵。首先， ，假设函数能将A换成具有良好性质的矩阵（如拉普拉斯矩阵），其中（是单位阵）的性质最好，可以进行正交分解，即

s

由频谱变换的思想可得[2] 2

展开得到 3

为可学习参数，将式1和式3带入式2可得：

4

由矩阵分解可得[3] 5

由切比雪夫多项式可得，令k=1（一阶近似），得为了进一步简化参数，令，得，在论文[1]中根据实验结果使用Renormalization trick中令，得GCN得近似公式：

6

在文献[4]里有更加详尽得理论推导和背景介绍，除GCN外还介绍了其他几种图神经网络。

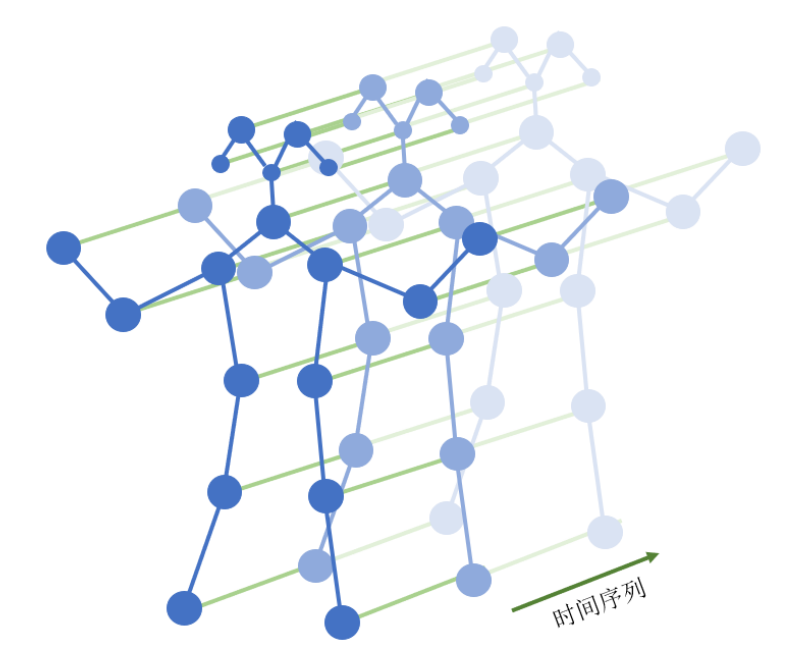
2级标题ST-GCN

3级标题人体骨架序列时空图介绍

人体骨架图通常指的是由关节点和它们之间的连线组成的图像，由于骨架信息表示了丰富的人体运动和姿态语义信息，因此在计算机视觉和机器学习领域中得到了广泛的研究。骨架图中的节点可以通过使用传感器技术（如深度相机）捕获到的人体姿势数据来确定，另外也可以通过一些姿态估计算法来获得人体骨架图，比如openpose[7]，alphapose[8]，poseC3d[9]，HRnet[10]等，这些模型可以再低复杂度场景下实现高鲁棒性和准确性，但是在高复杂度下难以实现多人体的姿态估计，若结合精准的检测器，能够提高在复杂场景的鲁棒性。

为了获得空间域和时间域的语义信息，将每一帧图像经过姿态估测器得到的人体骨架图按照时间顺序排列，可以得到连续帧人体骨架，如图x-x所示，即人体骨架序列时空图。其中蓝色的圆圈表示人体骨架关节点，蓝色圆圈由蓝色线条连接的是空间域人体骨架图，由绿色线条连接的是时间域人体骨架图。

设构建人体骨架序列时空图需要个关节点和帧的骨架序列，用表示帧数，表示关节点的编号，则人体骨架序列时空图可以表示成，其中，，为空间边，表示关节点在骨架内部的连接关系，为时间边，表示关节点在相邻骨架之间的连接。



图x-x人体骨架序列时空图

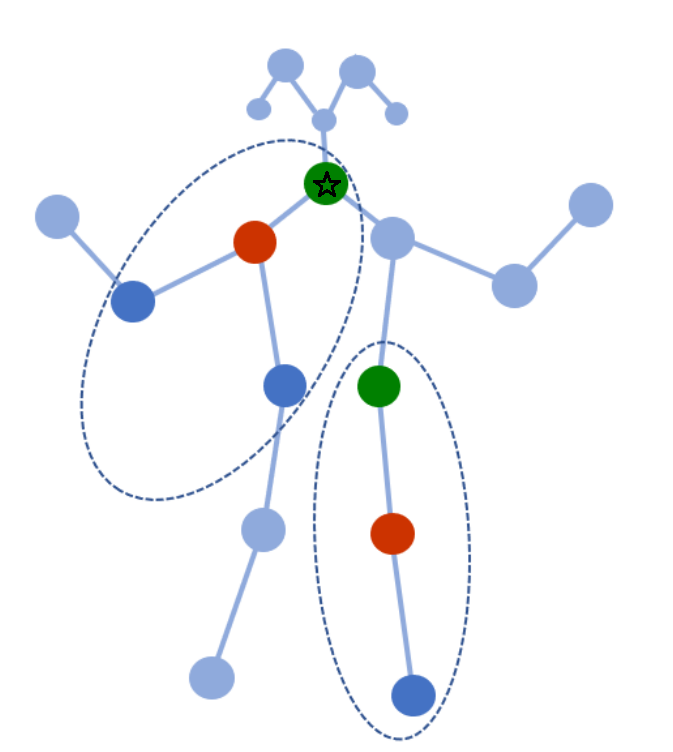
ST-GCN基本原理

ST-GCN为了同时获得时域和空域的语义信息，分别对空间域人体骨架图和时间域人体骨架图进行卷积，最后再将特征信息进行融合分类。

对于空间域骨架图，ST-GCN在GCN的基础上增加了可学习编码（大小为），主要学习骨架图中关节点与关节点联系的强弱，即：

7

表示关节点的邻域的子集划分几何，在ST-GCN中将邻域划分成三个子集，如图x-x所示。第一个子集只包含节点本身，图x-x中用红色圆圈表示。第二个子集由比根节点更接近骨架重心（图x-x中用五角星标出）的邻接点组成，包含了向心运动的特征，图x-x中用绿色圆圈表示。第三个子集由比根节点更远离骨架重心的邻接点组成，包含了离心运动的特征，图x-x中用深蓝色圆圈表示。



ST-GCN邻域子集划分

用表示骨架关节点i到骨架重心的距离，则对应的子集分区函数为[5][6]：

8

在空间图卷积的基础上，构建时间域卷积就比较简单，时间域上卷积使用TCN[11]的卷积思想，首先扩充邻域分区到时间域上，定义分区：

9

其中参数控制的是邻域中的时间范围，可以看作时间域卷积的卷积核大小，因此时间域的分区映射函数可以写为：

10

将上式10带入空间卷积式7中就能得到时间卷积TCN[5]。

3级标题ST-GCN缺点

ST-GCN将GCN利用到骨架序列的工作是开创性的，但是其在benchmark上的表现并不好，从ST-gcn的模型结构和数据结构分析，ST-GCN存在以下几个问题：

1）ST-GCN中使用的人体骨架图是启发式预定义的，仅表示人体的物理结构。因此它不能保证这样的结构对于动作识别任务就是最优的。比如说，对于识别“阅读”和“拍手”两只手的互动动作，这两个动作从骨架序列的角度看来没有太大的区别，若使用骨架序列图这种固定的图结构，就很难学习到两个手掌关节点之间的关联程度。

2）由式st-gcn公式我们可以知道，ST-GCN所有的层的卷积网络都没有改变图的拓扑结构，这是不便于提取特征，因为网络可能会在GCN中学到其他拓扑结构信息，骨架图的拓扑结构应该是独立于数据集之外的，而ST-GCN却不是。

3）ST-GCN中只利用了骨骼图的一类信息，而没有利用人体骨架图的其他类信息，如骨骼长度和骨骼方向[6]。

4）最初的ST-GCN虽然有BN、DropOut等正则化操作，但是ST-GCN表现不佳，可能是因为梯度消失，一些后期工作引入了Residual残差连接[12]和Attention注意力机制[13]。

2级标题2S-AGCN

3级标题2S-AGCN基本原理

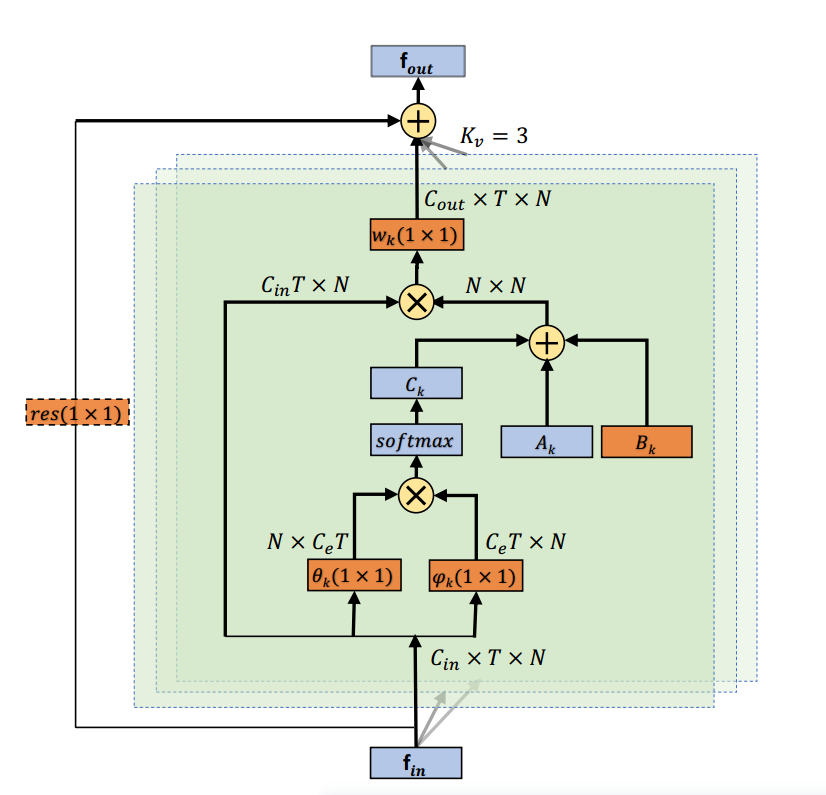
针对ST-GCN的问题，2s-agcn提出了一种自适应的gcn网络，它能够兼顾人体骨架的全局拓扑结构信息，也能够从具体的数据样本中学习到独有的拓扑结构。这两种结构信息都是可学习的，并且对于不同卷积层它们学习的结构信息都是不固定的，以端到端的学习方式，使图的拓扑结构于其他参数一起优化。

具体来说，在公式ST-gcn中，图的拓扑结构由和决定，表示哪些关节点相连，表示这些关节点相关联的强弱，为了使图的拓扑结构更加灵活，2s-agcn将公式ST-gcn改写为：

与公式ST-gcn中的相比没有变化，大小为，表示人体骨架固定的物理结构。

也使大小为的邻接矩阵，只不过中的参数使可学习的（初始化为0），完全根据数据集的特性进行学习。使用这种数据驱动的方式，使模型可以学习完全针对检测任务的特征，并且在不同的卷积层还可以学习特殊拓扑结构新形式。它不仅表明两个关节之间是否存在连接，而且还表明连接的强度。它能在功能上替代。但是公式st-gcn中是与点乘，如果中有一个元素为0，则对应中学习到的任何值都将没有意义，因此ST-GCN不能生成原始人体骨架图中不存在的新连接。所以比更加灵活。

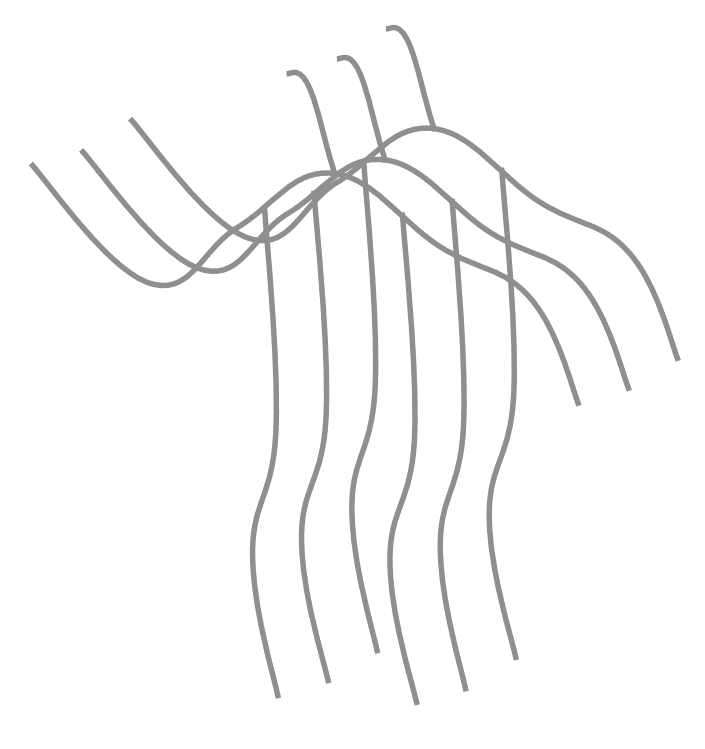
的大小也是N\*N，它可以从每一个样本中学习该样本中哪些节点可以关节点相连，相关联的强度有多大。具体来说，假设输入特征图的大小为（是通道大小，是帧数，节点个数），将特征图输入到两个Embedding Function 和中，这两个函数使用1\*1的卷积，分别输出一个大小的矩阵和一个的矩阵，再把这两个矩阵相乘可以获得大小为N\*N的矩阵，点积可以看作用来测量两个节点的相似性，再使用函数将矩阵中的元素归一化到0到1之间。C\_{k}的具体公式可以表示为，其中表示转置：



如果有时间再画一个st-gcn的操作图，用于对比

3级标题人体骨长信息

2s-agcn使用骨头的长度和骨头的方向作为人体骨架序列的第二种信息，如图x-x，利用这种信息能够更加充分的表示人体的姿态和运动信息。



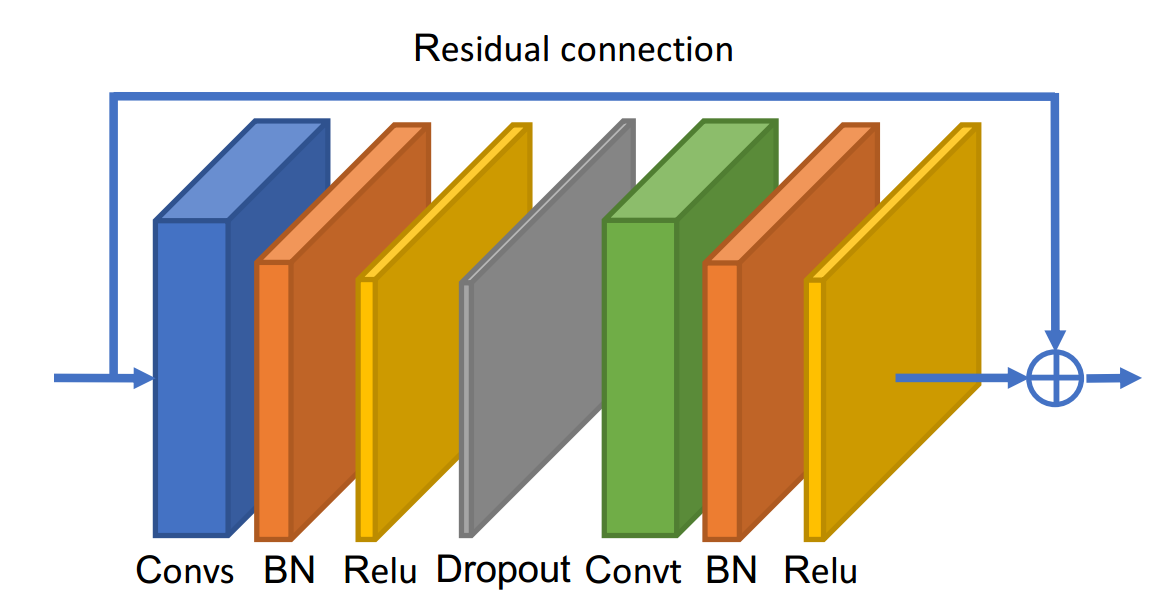
图x-x人体骨长信息

对人体骨长信息不需要定义新的网络模型，只需要将骨头的长度和方向转变成关节点的形式。定义只有相邻的关节点才有骨头连接，以距离重心近的关节点为头节点，距离重心远的为尾节点建立向量，因为人体骨架图是无环的[5]，可以得到个向量，再让重心同时作为头节点和尾节点自我相连，这样就得到个向量，可以按照关节点一样构建人体骨长拓扑图。

双流网络

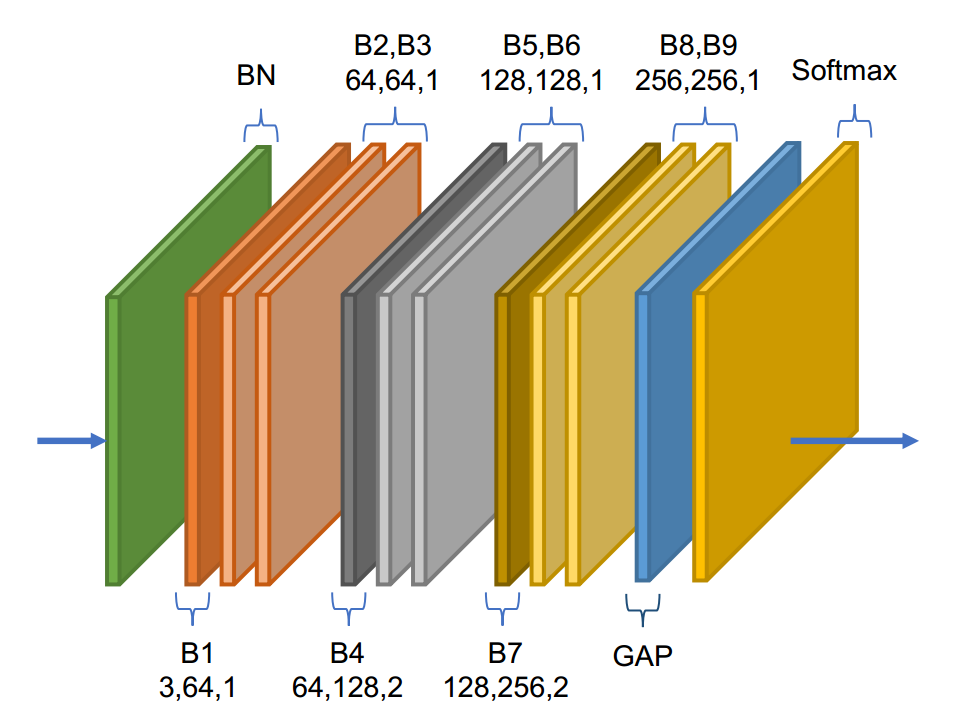
实现细节：

2S-AGCN使用的分区策略和时域卷积与st-gcn的一样，都是将邻域分成三个子集，并使用TCN作为时域卷积。如图x-x所示，2s-agcn网络的基础模块是将时域图卷积和空域图卷积结与Batch Normalization（BN）layer、Relu Layer、Dropout相结合，其中Dropout的丢失率一般设置为0.5，在卷积层后都加上一个BN和Relu层。另外，为了稳定训练过程，会在基础模块中加入残差连接[12]。



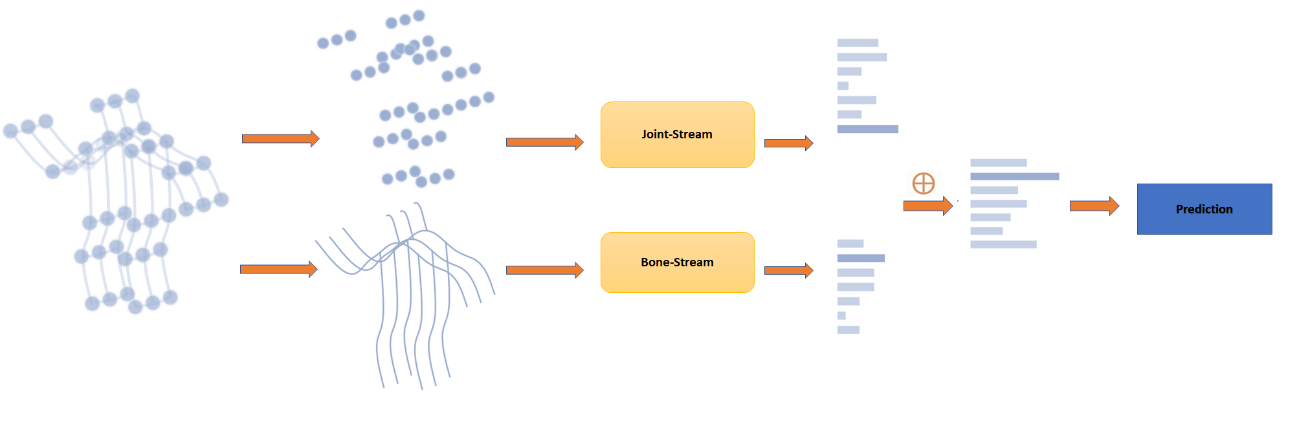
图x-x 自适应图卷积基础模块

如图x-x所示，将上述基础模块重复堆叠9次，再加上就形成了自适应图卷积网络（AGCN），每一个图卷积模块的输出通道数分别是64，64，64，128，128，256，256和输出。为了归一化输入数据，在网络的开头加上一个BN层。最后将输出送入到一个Global Average Pooling（GAP）层，将输出的特征图“压缩”到1维，最后送入和函数中进行分类预测。



图x-x AGCN网络结构

人体骨架图的第二类信息骨长可以通过一定的也变换成为图结构，只需要将骨节点图和骨长图分别输入到Joint-Stream和Bone-Stream（都是AGCN网络，参数不共享）中，最后将两个流的函数输出分数相加，就得到融合后的分数并预测动作标签。



图x-x 2S-AGCN

1. **ChebNet** Thomas N. Kipf, Max Welling. Semi-Supervised Classification with Graph Convolutional Networks[C]. ICLR: International Conference on Learning Representations, 2017: 22-36.
2. M. Defferrard, X. Bresson, P. Vandergheynst. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering[J]. Advances in Neural Information Processing Systems, 2016, Vol 29: 3844-3852.
3. Radu Horaud. A Short Tutorial on Graph Laplacians, Laplacian Embedding, and Spectral Clustering[EB/OL]. Csustan.edu. <https://csustan.csustan.edu/~tom/Clustering/GraphLaplacian-tutorial.pdf>
4. Ameya Daigavane, Balaraman Ravindran, Gaurav Aggarwal. Understanding Convolutions on Graphs[EB/OL]. Distill.pub. <https://distill.pub/2021/understanding-gnns/#polynomial-convolutions>
5. **ST-GCN**Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition[C]. AAAI: Association for the Advancement of Artificial Intelligence, 2018.
6. **2s-AGCN**Lei Shi, Yifan Zhang, Jian Cheng, etc. Two-Stream Adaptive Graph Convolutional Networks for Skeleton-Based Action Recognition[C]. CVPR: Computer Vision and Pattern Recognition, 2019.
7. **openpose** Zhe Cao, Tomas Simon, Shih-En Wei, etc. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields[C]. CVPR: Computer Vision and Pattern Recognition, 2017.
8. **alphapose** Hao-Shu Fang, Jiefeng Li, Hongyang Tang, etc. AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time[EB/OL] Arxiv.org. https://doi.org/10.48550/arXiv.2211.03375
9. **poseC3d** Haodong Duan, Yue Zhao, Kai Chen, etc. Revisiting Skeleton-based Action Recognition[EB/OL]. Arxiv.org. https://doi.org/10.48550/arXiv.2104.13586
10. **HRnet** Ke Sun, Bin Xiao, Dong Liu, etc. Deep High-Resolution Representation Learning for Human Pose Estimation[C]. CVPR: Computer Vision and Pattern Recognition, 2019.
11. **TCN**Shaojie Bai, J. Zico Kolter, Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling[EB/OL]. Arxiv.org. <https://doi.org/10.48550/arXiv.1803.01271>
12. **Residual** Kaiming He，Xiangyu Zhang, Shaoqing Ren, etc. Deep Residual Learning for Image Recognition[C]. CVPR: Computer Vision and Pattern Recognition, 2016.
13. **Attention**Ashish Vaswani, Noam Shazeer, Niki Parmar, etc. Attention is All You Need[EB/OL]. Arxiv.org. https://doi.org/10.48550/arXiv.1706.03762.