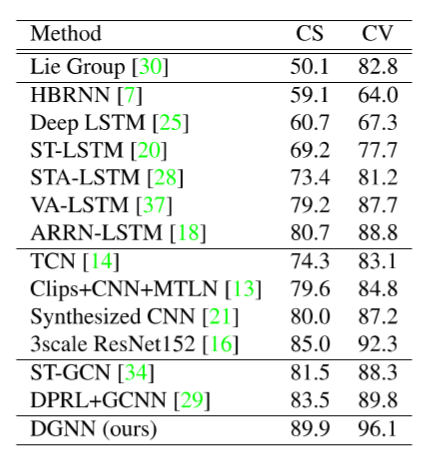
**基于深度学习的人体姿态分类：**

由于我们从开发板上获得的数据直接就是人体的关节、骨头信息，因此使用图神经网络来建模这一问题是相当直观的，并且从SOTA的角度，图神经网络也应该是合理的。我们对于图的定义如图1所示。我们定义关节为图的节点，骨头为图的边。边是有向的，我们定义人体胸部正中为根节点，那么边的方向总是由距离根节点更近的节点指向距离根节点比较远的节点。由于边的指向性，我们可以进一步定义指向一个节点的边为，而离开一个节点的边为，并定义，分别为指向顶点vi的边集合与指出顶点vi的边集合。

表1 2020年NTU数据集上各种深度学习算法的正确率



表注：CS是以受试者作为训练集测试集划分标准；CV是以视角作为训练集测试集划分标准。DGNN算法正确率对于我们这个问题应该是合适的。

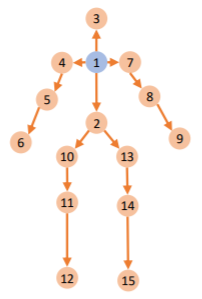
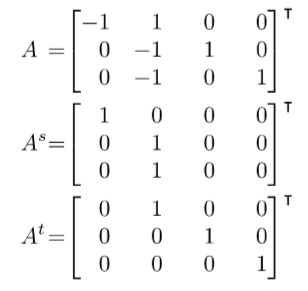


图1 人体骨架的图建模方法

在以上抽象的描述之后我们需要具体化这一张图，假设一张图上有Nv个节点，Ne条边，那么使用邻接矩阵A可以表示为Aij，i=1…Nv, j=1…Ne，如果节点vi是节点边ej的起点，则Aij记为1，终点则记为-1，否则记为0。定义矩阵As=A>0, At=A<0为了后续运算方便。



在充分定义了这样一张图之后，我们希望利用这些图信息来完成动作的分类任务。这就引出了使用directed graph neural network（DGNN）这种网络架构。首先我们需要引入DGN block，它包含了两个更新函数与，同时包含两个聚合函数与，论文给出的聚合方法是平均池化。基于这两个函数以及原有图信息的更新方法如下四个式子所示。其中对于符号需要额外说明的是，第三、四个式子的左值代表的是经过一次图更新之后的节点和边，用“‘”来表示；而他们的右值中[·]代表向量的合并运算。这四个式子以顺序的形式执行，即对于所有节点，首先通过指向这些节点的边的特征来计算出该节点的第一个特征，再通过指出这些节点的边的特征来计算出该节点的第一个特征（式子1，2）；接着，使用该节点的两个特征与以及该节点的原始特征共同计算该节点更新后特征。在对于所有的节点都进行了一次更新之后，就可以利用所有更新后的节点的特征以及边在上一次执行算法后的特征，共同更新出新的边特征。

进一步的，可以将上述的图的矩阵表示与特征推导过程相结合，当我们的聚合函数取到的是平均池化时，在计算过程中其相当于一个单位矩阵（忽略常数项），因此不会在公式中显式标出，而节点和边的特征向量可以用矩阵来表示，即我们定义是一个CT×Nv的矩阵，是一个CT×Ne的矩阵，这里CT是特征的维度数。那么式（1）~（4）可以表示为式（5）~（6）。然而，此时模型仍然是线性的，在此基础上进一步加入激活函数和Batch normalization可以提供非线性以及防止梯度消失或者梯度爆炸。

经过以上的建模过程，我们已经可以很好的对于一帧的信息进行编码。为了进一步利用时间上的信息，我们可以在时间上对于每一个节点与边的特征向量进行一维卷积。最后再通过一个全连接层，即可完成分类任务。同时，为了充分使用时间顺序信息，我们可以进一步的构造特征向量我们让这两个特征向量上文中的节点和边的更新操作进入另一个相同架构的网络，最后也可以通过全连接层输出概率。最后，基于模型融合的思想，将两个模型的概率求和，再以最大的概率输出类标签即作为输出值。

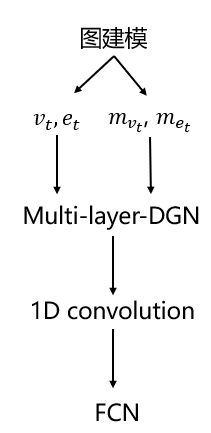


图2 图卷积网络流程图

**需要解决的潜在问题：**

我们同样发现上述的过程在实际实现的过程中会遇到很大的问题。一是NTU数据集对于提取的骨架关键点与我们相机的骨架关键点提取结果不一样。如图3所示，左图是NTU数据集的骨架提取结果，而右图是我们相机的骨架提取结果。可以看出有以下四点比较显著的差异，一是我们相机在肢体只有12个点，而NTU数据集上也有24个点；第二，我们相机时常会出现漏点的情况，比如在图2中，我们的相机就漏掉了左手手腕的点；然而开源的数据集上不存在漏点的可能性；第三，在骨架的提取上，开源数据集的背部是一根线，而我们提取到的数据则是一个四边形；第四，开源数据集中坐标实际是三维的（画图时投影到了二维），然而我们的相机采集的数据却只有二维。基于以上的多种差异，我们计划要首先对开源数据集的点进行采样，更进一步说是将开源数据集适配到我们的采样点上，这样才能在开源数据集上进行大规模的预训练。基于以上的大量问题，目前与王老师组的协商结果是，最终在站立、坐下、躺下、走路这几个类别的识别正确率有70%就相当优秀了

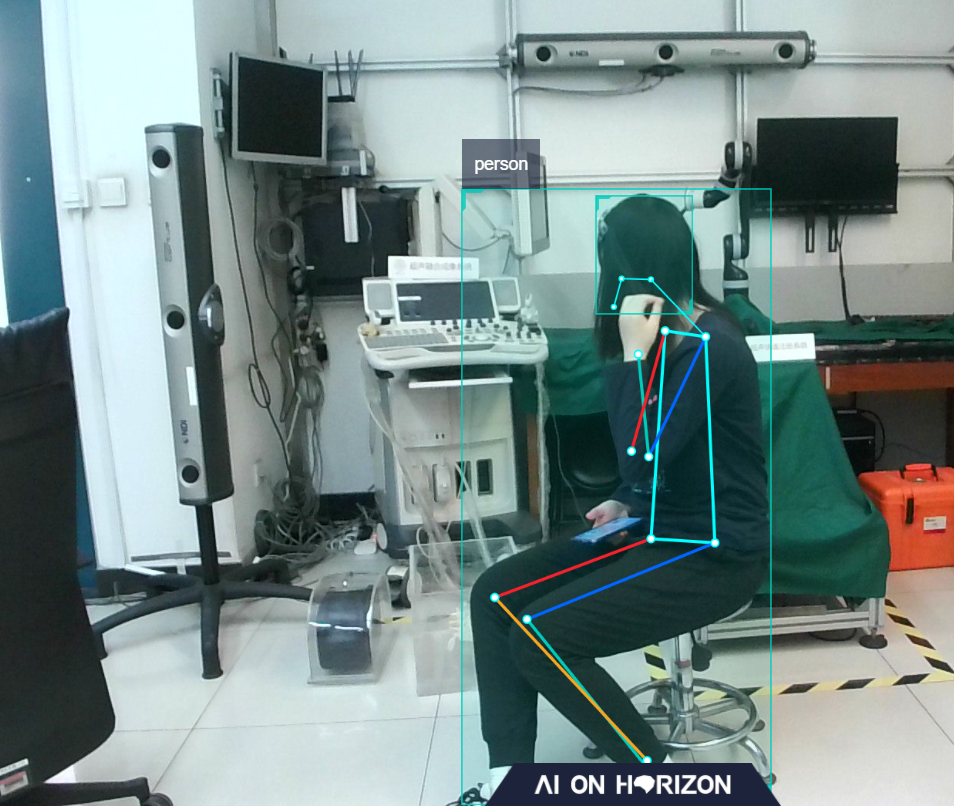
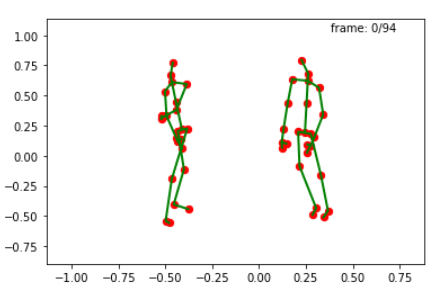


图3 骨架提取结果对比

图注：左图时NTU数据集上的骨架提取结果，右图是我们开发板的骨架提取结果

[1] Shi L , Zhang Y , Cheng J , et al. Skeleton-Based Action Recognition With Directed Graph Neural Networks[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[2] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, Gang Wang, "NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis", IEEE Conference on Computer Vision and Pattern Recognition, 2016

[3] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, Alex C. Kot, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019. [PDF] [bibtex].