动作识别实验的总结

最近一周工作总结

最近一周由于中期答辩的临近，在同时增补论文内容的同时准备了答辩材料的填写。在实验方面，我参考了开源程序[1]和他人文献[2]后，先后完成了朴素贝叶斯最近邻分类和隐马尔可夫模型的编写，动作识别准确率达到了较高标准。

在论文撰写上，由于相关经验不足，摘要不能够达到老师的“言简意赅”和“切中要点”的要求。因此，需要在接下来的时间内认真构思摘要的撰写。

实验大致方法介绍

1) 使用的数据集

实验使用 MSR Action3D 数据集，其内部分类与样本信息如下表所示。由于该数据集由多个动作执行者录制，因此采用跨目标验证的方式。一种典型的跨目标验证的方式如表 1 所示：

表 1 本次实验所使用数据集的相关信息

数 据 集	类别数	样本数	特 性
MSR Action3D	20	567	10 个演员，每类动作每个演员执行 2~3 次；提供 20 个关节点的 3D 坐标数据、深度图像与 RGB 图像；视频序列为无背景的纯人体运动目标

2) 模型测试方法

对于动作识别精确度的判断，目前主要采用跨目标验证和交叉验证的方法。

跨目标验证的思想是：训练样本与测试样本分别来自不同动作执行者的动作序列。此方法便是为了解决上一节提到的，同类型的动作不同动作执行者模型评价问题。

交叉验证是用来验证分类器性能的一种常用统计分析方法，基本思想是按照一定的划分方式将原始数据集进行分组，一部分作为训练集，另一部分作为验证集。首先用训练集对分类器进行训练，再利用验证集来测试训练得到的模型。评

价分类器的性能指标将使用验证集中的测试结果得出。

跨目标验证实验分多批进行，每批测试使用不同的动作执行者作为测试集与训练集，正如表 2 所示。在所有批次测试结束后，对正确率和混淆矩阵求均值，作为最终的模型评价指标。在我的实验中，总共有 8 批这样的训练。

表 2 每批训练的动作执行者划分表

分组序号	训练动作执行者序号	测试动作执行者序号
1	3、1、10、5、2	4、6、7、8、9
2	10、5、2、4、3	1、6、7、8、9
3	3、5、6、2、7	1、4、8、9、10
……	……	……

3) 模型评价方法

在实验和论文中，主要使用正确率和混淆矩阵对模型进行评价。以下给出正确率和混淆矩阵的定义。

对一批二分类样本进行分类后，对于每一个样本，其分类结果必然属于表 3 四种情况之一，模型分类的准确率如公式 1 所示。

表 3 二分类结果可能出现的情况

真正例（True Positive，简称 TP）	将一个正例正确判断成一个正例
伪正例（False Positive，简称 FP）	将一个反例错误判断为一个正例
真反例（True Negtive，简称 TN）	将一个反例正确判断为一个反例
伪反例（False Negtive，简称 FN）	将一个正例错误判断为一个反例

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

将四种情况以如表 4 表 4 分类准确率的二维分布表二维表格形式表示，便可以清晰地表示出模型分类性能，以及哪些类更加容易混淆。

表 4 分类准确率的二维分布表

		预测分类	
		0	1
实际分类	0	TN	FP
	1	FN	TP

除去标签，$\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$便是二分类的混淆矩阵定义。对于M分类问题，混淆矩阵为一个M × M的矩阵。在本文中，以关节的空间绝对位置作为骨架关节特征，利用朴素贝叶斯最近邻（NBNN）分类器对 MSR Action3D 数据集进行动作分类后的结果如图 1 所示：

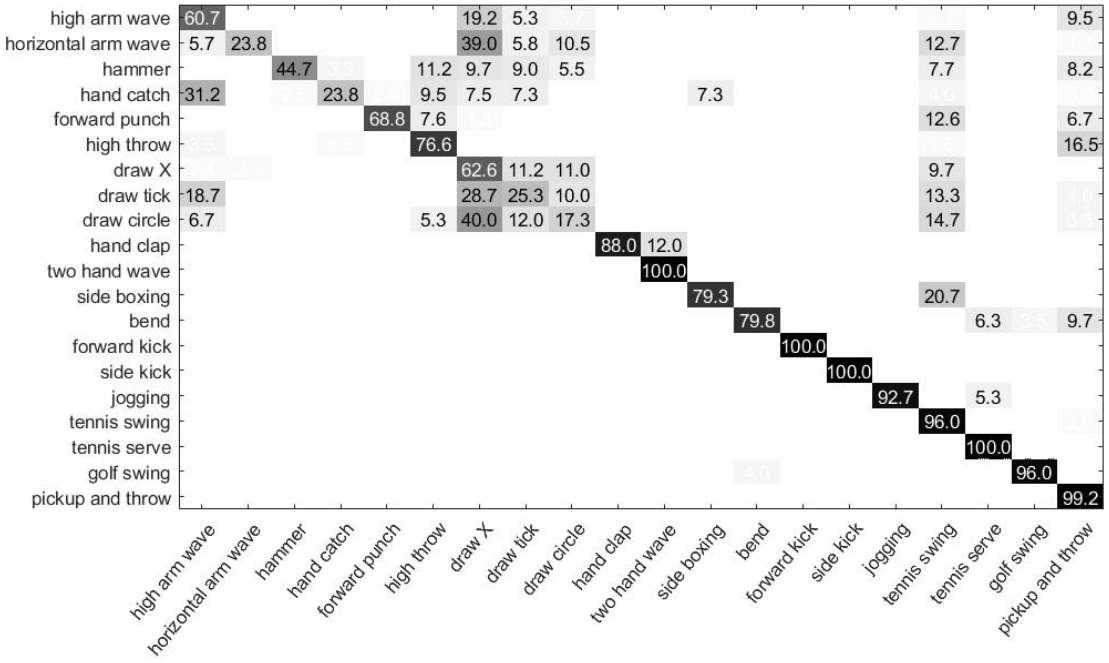

图 1 评估动作分类模型的混淆矩阵示意图

由此可见，使用关节绝对位置和 NBNN 分类器对动作进行分类时，high throw、draw X、draw tick、draw circle、tennis swing 和 pickup and throw 具有较强的易混淆性，需要使用更具有识别力的骨架关节特

实验原理介绍

1) NBNN 分类器原理

当前用于动作识别的深度图像数据集种类繁多，且形式不一。例如，本文使用的 MSR Action3D 数据集，每个样本最多只有 50 帧，而小样本只有 20 多帧。对于不同的样本，需要进行样本大小的标准化。为了使样本总数不减少，过小的样本不能直接丢弃。在此，我提出两种简单的办法。

在文献[2]中提到的 NBNN 分类器使用的是类似于最邻近方法对视频序列进行分类，即找出与待分类样本v^*“距离”最近的样本v_C，C即为样本v_C所属的类

别。对每一个可能的分类C求出“距离”后，找到“距离”最近的样本v^*。该样本所属的类别便是待分类样本的类别。该距离dist的算法如下：

$$\mathrm{dist}(v_C, v^*) = \sum_{i=1}^{M} \|d_i - NN_c(d_i)\|^2 \tag{2}$$

下面开始介绍“距离”的计算方法：M为每个视频样本所拥有的帧数，d_i为视频序列每一帧的描述符，$NN_c(d_i)$指的是描述符d_i在C类内的最近邻。d_i即是姿态特征的抽象表示，本节中使用特征关节f_{norm}作视频每帧的描述符。在利用公式 2 求出所有可能的分类C的“距离”后，可得待分类视频样本 v^*所属的分类$C^* = \underset{C}{\mathrm{argmin}}\ \mathrm{dist}(v_C, v^*)$。

2) 隐马尔可夫模型原理

使用隐马尔可夫模型的动作分类大致包含三个过程，即特征抽取、向量量化和离散隐马尔可夫建模。特征抽取使用线性判别分析法（Linear Discriminant Analysis，简称 LDA），也叫做 Fisher 线性判别(Fisher Linear Discriminant，简称 FLD)。线性鉴别分析的基本思想是将高维的模式样本投影到最佳鉴别矢量空间，以达到抽取分类信息和压缩特征空间维数的效果，投影后保证模式样本在新的子空间有最大的类间距离和最小的类内距离，即模式在该空间中有最佳的可分离性。向量量化则使用 K 聚类算法，将训练集中的每一个视频样本中的每一帧的特征描述符量化为视觉词。

我们将三元组$\lambda = \{A, B, \pi\}$视作一个 HMM 模型H_i，对向量量化后的视频特征训练M个 HMM 模型。对于一个输出序列，其分类方法如公式 3 所示：

$$\mathrm{decision} = \underset{i=1,2,\ldots,M}{\mathrm{argmax}}\{\Pr(O|H_i)\} \tag{3}$$

实验结果总结

表 5 不同分类器与不同骨架关节特征搭配的正确率

模型名称	关节绝对位置	关节相对位置	特征关节
朴素贝叶斯最近邻	0.7251	0.7802	0.8408
隐马尔可夫模型	0.3189	0.3396	0.2855

对不同分类器和不同骨架关节特征进行分类后，可以得到其正确率如表 5 所示。表中隐马尔可夫模型的实验结果很不理想，初步猜测是模型参数（如状态数与输出数等）设置的不够合理。可尝试将状态数加大，而输出数是对每一帧动作视频的特征描述符聚类后的类数，应将聚类后的类数减少。

中期答辩后的工作安排

1) 隐马尔可夫模型的评估

前面已经在第一节介绍过的跨目标测试可以科学的评估一个动作识别模型的性能。在动作识别分类器评价上，应该与已经完成的 NBNN 分类器进行对比。通过使用不同骨架关节特征，找出其中最适合使用隐马尔可夫模型的特征作为每一帧的描述符。

2) 三维模型特征的对比

三维模型特征与骨架关节特征略有不同，在骨架的基础上，三维模型包含了与人体交互的物体的空间信息，因此三维模型可以在动作识别上更加具有识别力。获取深度的三维空间信息则是较为关键的第一步。

假设(u, v)图像中的二维坐标点，z_c为(u, v)的深度，其对应的空间坐标点为(x_w, y_w, z_w)。从二维坐标到空间坐标的转化公式如下：

$$\begin{cases} x_w = z_c \cdot (u - u_0) \cdot C_x \\ y_w = z_c \cdot (v - v_0) \cdot C_y \\ z_w = z_c \end{cases} \tag{3}$$

其中u_0、v_0、C_x和C_y均为传感器的内部参数。此时，人体和环境中的物体均以点云的形式表现出来，人体和环境物体的点云相互作用形成的特征被称为局部占用模式（Local Occupancy Patterns，简称 LOP）。将提取出的局部占用模式和骨架关节特征按照时间有序组合后，使用已经实现的分类器对特征的识别力进行评估。

3) 论文撰写的工作

进一步精简摘要的表述，同时补充英文摘要。同时对论文相关课题研究调研的部分，存在表述不准确和逻辑表述关系不密切的问题，需要进一步筛查和改善。在完善前面的任务后，需要将工作的结果以图形的方式在论文中整合。

由于本课题涉及多种深度动作数据集，属于对他人工作成果的引用。因此，在论文提到不同数据集名称和相关信息时，应该标注参考文献与数据集出处。在对论文内容进一步扩充的同时，应检查文献的标注是否准确以及是否存在遗漏的情况。

引用文献

[1] https://github.com/kailiuXD/STWP

[2] Wang J , Liu Z , Wu Y , et al. Learning Actionlet Ensemble for 3D Human Action Recognition[J]. IEEE Transactions on Software Engineering, 2013, 36(5):914-927.

[3] Xia L , Chen C C , Aggarwal J K . View invariant human action recognition using histograms of 3D joints[C]// Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE, 2012.

[4] Wang J , Liu Z , Chorowski J , et al. Robust 3D Action Recognition with Random Occupancy Patterns[M]// Computer Vision – ECCV 2012. 2012.