

太极拳动作分类论文结构

更新：2022 年 5 月 4 日

目录

1	Abstract	2
2	Introduction	2
3	Related Work	2
4	Method	3
5	Experiment	3
5.1	Pre-trained Model	4
5.2	linear protocol evaluation 中学习率的选取 (<i>code : exp0</i>)	4
5.3	linear protocol evaluation 中太极拳数据预处理对分类结果的影响 (<i>code : exp1</i>)	5
5.4	五通道数据的通用性和有效性 (<i>code : exp2</i>)	6
5.5	验证预训练可以学到人体运动特征 (<i>code : exp4</i>)	6
5.6	linear protocol evaluation 中不同训练配置的分类结果 (<i>code : exp3</i>)	8
5.7	linear protocol evaluation 中多模态数据特征融合 (<i>code : exp6 + exp8 + tools/kplb_fusion.py</i>)	8
5.8	finetune evaluation 中学习率的选择 (<i>code : exp5</i>)	8
5.9	finetune evaluation 中模型预训练的作用 (<i>code : exp5/without - pre</i>)	9

1 Abstract

基于骨架的人类动作识别在近年来获得了充分的研究。这些研究通常使用大规模数据集对人类日常动作行为进行分类或者分割。我们尝试使用深度学习对小规模的太极拳数据集进行动作分类，先前的工作使用空间 Transformer 网络对太极拳动作进行分类，但是在小规模训练集上表现较差。这篇文章利用人体运动学 (antropometric kinematics) 的先验信息，通过预训练模型提取人类运动的共有特征，辅助提升太极拳的分类性能；我们还将人体的关节点分成躯干和四肢五部分，提升数据的利用效率和通用性；我们将多模态的数据在特征维进行融合，进一步提升了分类准确率。Linear Protocol Evaluation 和 Finetune Evaluation 实验证明了，这种预训练的方式确实能够捕捉到不同类运动之间的共有特征，并且在不同训练配置下的分类结果可以验证该方法在小规模太极拳数据集分类任务中具有稳定且精确的分类能力，效果优于之前的方法，缓解过拟合现象。

关键词：Action Recognition, Small-scale Dataset, Tai Chi Action, Transfer Learning

2 Introduction

- 动作数据集的介绍，包括数据格式和数据采集方式等。
- 深度学习应用于动作识别的各大类方法和流行趋势。
- 太极拳研究的意义，介绍太极拳的动作特点和与普通动作的区别。
- 之前对太极拳的研究方向和研究内容。
- 本篇论文的研究思路图1和实验流程以及创新点。
 1. 将关节点按照人体结构进行分类，不同数据集可以通用且高效地共享动作特征。
 2. 利用人体运动学的先验信息，提取运动的共有特征，提升在小数据集样本上的分类能力。
 3. 将多模态的数据在特征维上进行融合，进一步提升分类性能。
 4. 太极拳动作分类任务性能大幅提升，state of the art。

3 Related Work

- 基于骨架人体动作数据集，选择使用 NTU60 数据集作为预训练数据集的原因
- 动作识别研究方法的介绍，选择使用 PoseC3D 的原因。

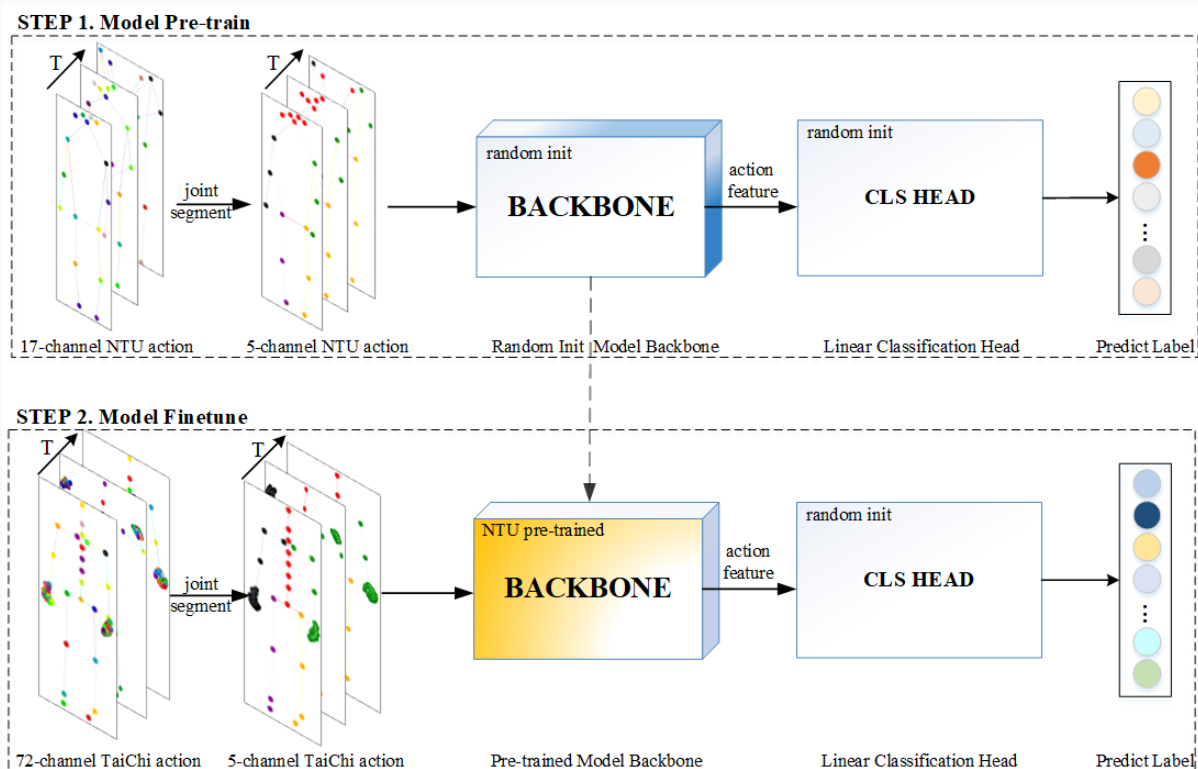


图 1: Tai Chi Model Framework

- Linear Protocol 和 Finetune 两种验证方式介绍 (这部分查找资料，现有的参考文献看的较少)，两部分在文章中应用验证不同的内容。Linear Protocol 验证模型学到了人类动作间的共同特征，Finetune 验证预训练模型帮助提升分类性能
- 与太极拳相关的任务，与其余任务的不同之处。

4 Method

- 两类数据集从骨架关节坐标到 3D Heatmap Volumes 的运算过程。
- 模型 backbone 具体内容介绍，附上一个 SlowFast 中 Slow Pathway 的流程图。
- 两种数据融合方式的介绍。
- Linear Protocol 和 Finetune 中的训练策略。

5 Experiment

通过对比实验验证创新点。实验分为 linear protocol evaluation 和 finetune 两部分。准备工作利用 NTU 数据对模型进行预训练^{5.1}；确定两部分的网络学习率^{5.2}和^{5.8}；并且探寻合适的数据预处理的方式^{5.3}。在 linear protocol 中首先验证了将关节点分成 5 通道

的有效性，与 17 通道的分类结果进行了对比实验5.4；其次验证了这种预训练方式能够需要人体运动学特征，应用于新的分类场景中5.5；最后展示了不同训练配置的分类结果5.6和多模态数据融合对分类效果的提升5.7。在 finetune 中验证了预训练模型对于模型分类效果的提升5.9，和在小规模训练集条件下训练结果的稳定性5.10。

5.1 Pre-trained Model

使用 PoseC3D 作为分类模型，利用 2D 骨架数据的 NTU60 数据集对模型进行预训练，2D 骨架数据是利用算法提取数据而非深度相机采集，数据稳定简洁，具有 17 个节点，图2所示。这里设计了四种预训练模型：只包含 keypoint 或 limb 的 5 通道预训练模型 5part-kp(`code : 5part_ntu60_xsub_kp_0331`), 5part-lb(`code : 5part_ntu60_xsub_limb_0412`)；包含 keypoint 的 17 通道预训练模型 17part-kp(`code : 17part_ntu60_xsub_kp_0401`)；包含 keypoint 和 limb 的 34 通道预训练模型 34part-kp+lb(`code : 34part_ntu60_xsub_kp + limb_0402`)；包含 keypoint 和 lb 的 10 通道预训练模型 10part-kp+lb(`code : 10part_ntu60_xsub_kp + limb`)，利用 NTU60 的 cross-subject 实验进行预训练并且验证分类结果，预训练模型分类结果表1所示，不同配置的模型分类结果在 93% 左右。

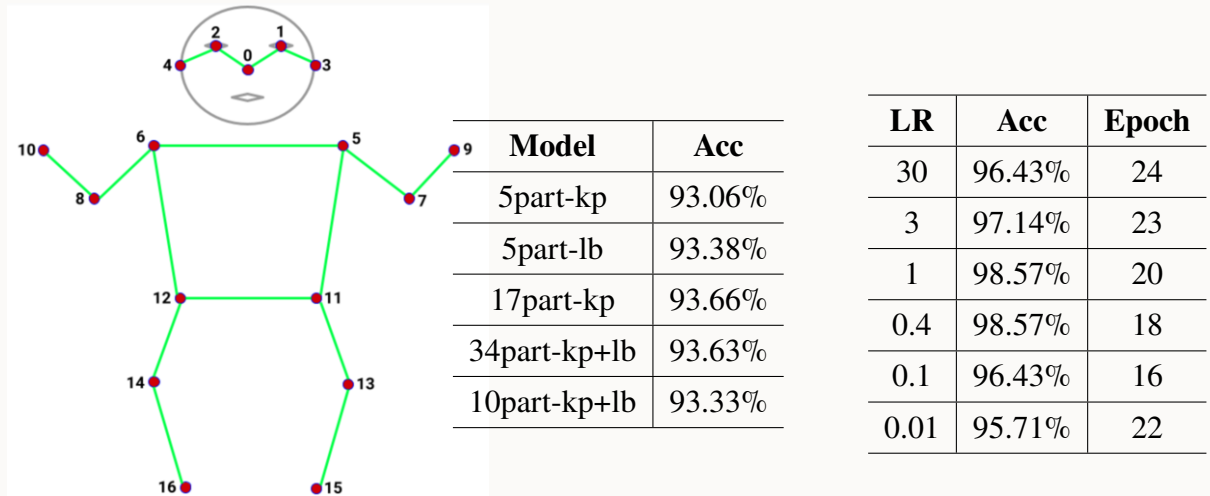


图 2: NTU17 节点示意图

表 1: NTU60 Xsub 分类结果

表 2: Linear Protocol LR Grid Search

5.2 linear protocol evaluation 中学习率的选取 (`code : exp0`)

在无监督任务中，MoCo 通过 grid search 发现最优的初始学习率为 30，我们也通过 grid search 探寻太极拳任务中，linear protocol evaluation 的最优初始学习率。这里利用 5part-kp 的主干网络作为动作特征提取器，选择太极拳 30% 的样本做训练集，70% 的样

本做测试集，训练最后的线性分类器，不同学习率下的分类准确率和达到最高分类效果的 epoch 值如表2所示，选择合适的初始学习率 $lr = 0.4$

5.3 linear protocol evaluation 中太极拳数据预处理对分类结果的影响 (code : exp1)

在 linear protocol 中，太极拳训练过程不改变主干网络的权重，数据预处理帮助模型的分类器对提取到的动作特征进行分类，找到一种最优的分类方式。

数据预处理的方式包含三维坐标绕 y 轴旋转 (Rotate) 操作，三维坐标的剪切 (Shear) 操作，坐标变换形式如公式1,2所示，其中 $\alpha \in [-60^\circ, 60^\circ]$ ， $a \sim U[-0.5, 0.5]$

$$\begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix} = \begin{pmatrix} \cos\varphi & 0 & \sin\varphi \\ 0 & 1 & 0 \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (1) \quad \begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} = \begin{pmatrix} 1 & a_{12} & a_{13} \\ a_{21} & 1 & a_{23} \\ a_{31} & a_{32} & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (2)$$

其余的预处理包括人体左右关节的随机交换 (Flip)，视频尺寸的随机剪裁 (Crop)，以及数据集的扩增 (AUG)。实验均在 10% 样本训练集 90% 样本测试集的条件下进行，对于 rotate 和 shear, Flip 和 Crop, WSWR 代表 with shear with rotate, NSNR 代表 no shear no rotate。对于数据扩增，对训练集样本分别扩增 2,4,8 倍，测试模型分类器的分类性能如表3所示。

Cfg	Acc	Cfg	Acc	Cfg	Acc
NSNR	88.89%	NFNC	81.67%	TEST9AUG2	88.33%
NSWR	82.78%	NFWC	82.78%	TEST9AUG4	88.89%
WSNR	86.67%	WFNC	81.67%	TEST9AUG8	86.11%
WSWR	81.11%	WFWC	88.89%		

表 3: Linear Protocol Preprocess Results

由于训练只改变最后分类器的权值，数据预处理帮助分类器对提取到的特征进行分类，可以看出 Rotate 和 Shear 操作不能帮助分类器更好的收敛，但 Flip 和 Shear 结合能帮助分类器更好地分类，可以得到左右关节和尺寸的变换对于分类器收敛起到重要作用，最后数据扩增对分类效果影响不大，可能是由于在训练时，一个 epoch 中训练集数据被重复 20 遍，包含了数据扩增的效果，所以后续的实验选择在 NSNR+WFWC, AUG=4 的条件下进行。

5.4 五通道数据的通用性和有效性 (*code : exp2*)

在数据处理中，将 NTU 数据集和太极拳数据集的骨架数据都分成了躯干和四肢五部分，每一部分作为一个通道输入到网络中，在其余参数不变的情况下，对于 NTU60-Xsub 分类任务，均到了 93% 以上的分类效果。由于不同类型的数据集关节点定义不同，如果直接进行关节点匹配，有可能出现关节点缺失或者不完全利用的情形。比如 17 个关节点的 NTU 数据集包含丰富的面部信息但是四肢信息较少，而太极拳动作有丰富的四肢关节点信息但缺少面部信息，如图3(a)所示，所以骨架数据分成躯干和四肢五部分可以更好地利用数据间的共同特征。

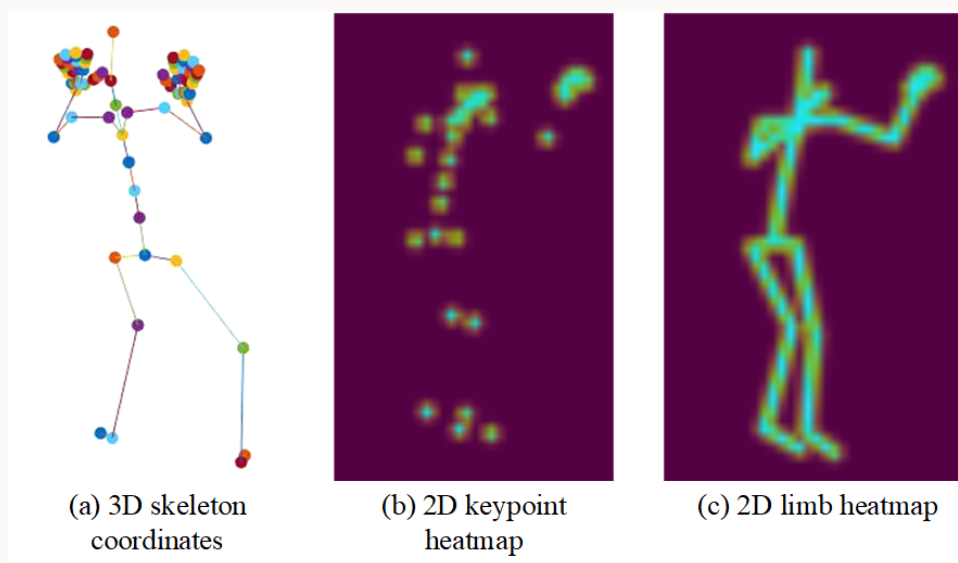


图 3: 3D joint coordinates and 2D heatmaps of Tai Chi action

我们分别利用 5 通道的预训练模型和 17 通道的预训练模型，针对 90% 测试集的太极拳分类任务进行 linear protocol evaluation，并对主干网络经过 GlobalAveragePooling3D 之后的特征进行 t-SNE 可视化并比较分类效果。从图4,5中可以看出，5 通道的 t-SNE 特征具有更紧密的分布和较少孤立点，并且有 4% 分类结果的提升。

5.5 验证预训练可以学到人体运动特征 (*code : exp4*)

我们认为预训练可以学到人体的运动特征，我们使用预训练的权重和随机初始化的权重，针对 90% 测试集的太极拳分类任务进行 linear protocol evaluation，并对特征进行 t-SNE 可视化。从图6,7可以看出，随机初始化的权重没有对数据进行很好的区分并且分类失败。

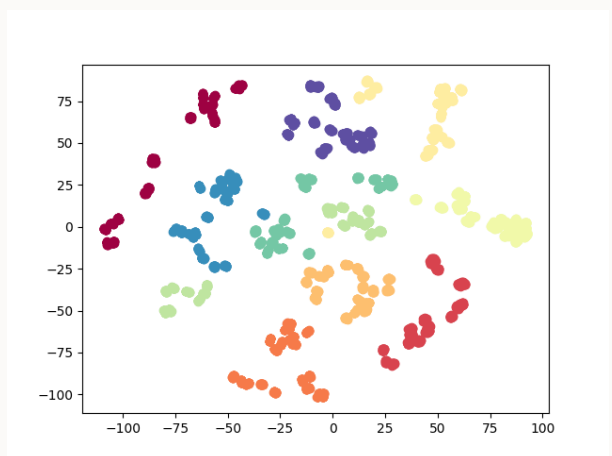


图 4: 5 通道分类准确率 88.89%

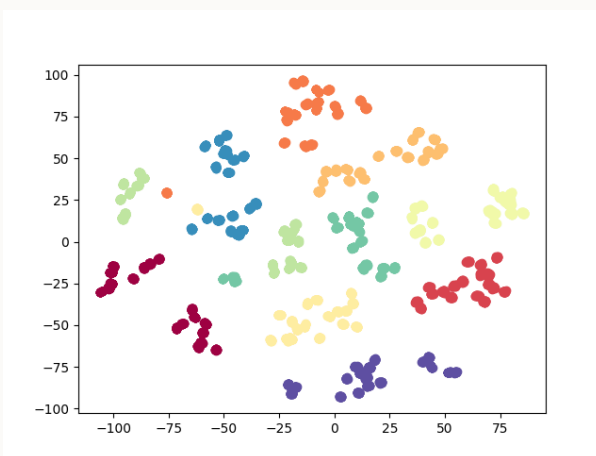


图 5: 17 通道分类准确率 85%

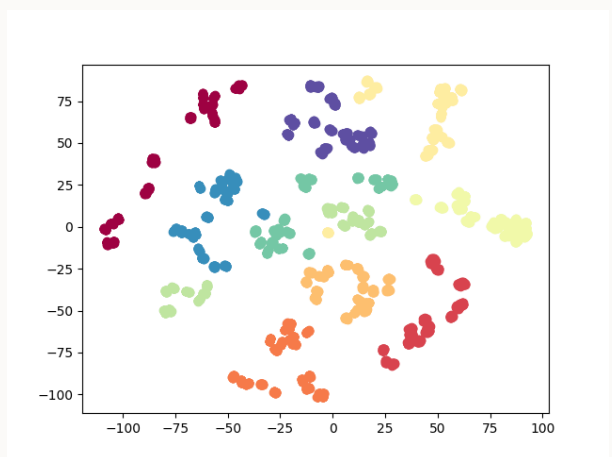


图 6: 5 通道预训练模型分类准确率 88.89%

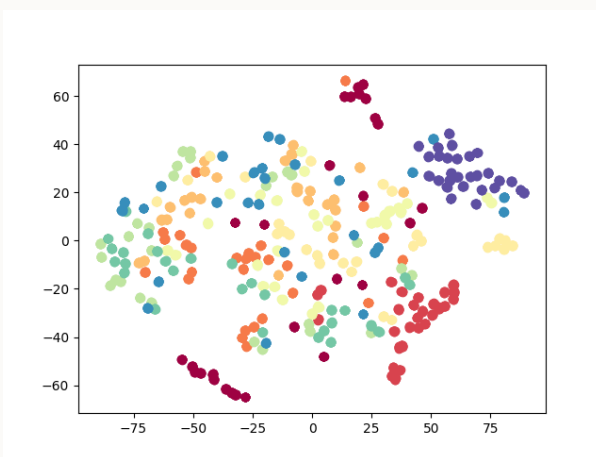


图 7: 5 通道随机初始化模型准确率 10%

5.6 linear protocol evaluation 中不同训练配置的分类结果 (*code : exp3*)

在同一个的 5 关节点通道的主干网络作为预训练模型的基础上，针对 90% 到 30% 测试集的太极拳分类任务进行 linear protocol evaluation, 其分类效果如表4所示，可以看出增广系数对太极拳分类效果影响较小，分类结果随着测试集比例的降低而升高，说明了更多不同组样本可以帮助分类器更好的收敛，并且 70% 以下测试集样本就可以达到 10 类太极拳动作的准确分类。

Cfg	Acc	Cfg	Acc
TEST9AUG2	88.33%	TEST9AUG4	88.89%
TEST7AUG2	98.57%	TEST7AUG4	98.57%
TEST5AUG2	99.00%	TEST5AUG4	98.00%
TEST3AUG2	100.00%	TEST3AUG4	100.00%

表 4: Linear Protocol Results across Configs

5.7 linear protocol evaluation 中多模态数据特征融合 (*code : exp6 + exp8 + tools/kplb_fusion.py*)

关节点 keypoint 和骨骼 limb 数据都是骨架数据的常用数据类型，为了更好地利用两种类型的数据，提出将两种数据在通道维度上融合 (early fusion) 或者分别进行训练在最后的分类层进行融合 (late fusion) 两种融合方法。太极拳 limb 数据分类结果，包含 10 通道的 early fusion 分类结果，late fusion 的分类结果如表5所示，可以看出 late fusion 能得到比单独的 keypoint 或 limb 都出色的分类结果而 early fusion 分类表现不佳。

Cfg	Keypoint	Limb	Early Fusion	Late Fusion
TEST9AUG2	88.33%	87.78%	78.33%	90.56%
TEST9AUG4	88.89%	87.22%	80.56%	89.45%
TEST9AUG8	86.11%	88.33%	77.22%	88.33%

表 5: Linear Protocol Results across Fusion Methods

5.8 finetune evaluation 中学习率的选择 (*code : exp5*)

利用预 5 关节点通道的预训练作为太极拳动作分类的初始化权重，利用 10% 的太极拳训练集数据对模型进行微调，测试 90% 太极拳测试数据的分类结果，通过 Grid

Search 选择合适的学习率，不同学习率下的分类准确率和达到最高分类效果的 epoch 值如表6所示，选择合适的初始学习率 $lr = 0.01$

LR	Acc	Epoch
1	89.45%	43
0.1	96.67%	28
0.01	100.00%	19

表 6: Finetune LR Grid Search

5.9 finetune evaluation 中模型预训练的作用 (`code : exp5/without - pre`)

验证对模型进行预训练可以提升模型在小样本数据集上的分类精度，我们利用随机初始化的模型对太极拳数据集进行分类，得到的分类结果和特征可视化与预训练模型的对比如图8,9所示。可以看出预训练模型的提取到的特征更有区分度而随机初始化的模型提取到的特征区分度不强，收到训练集样本数量的限制，过拟合严重。

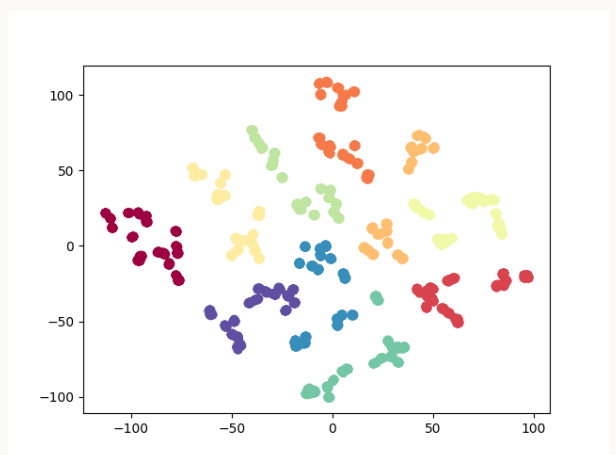


图 8: 预训练模型分类准确率 100.00%

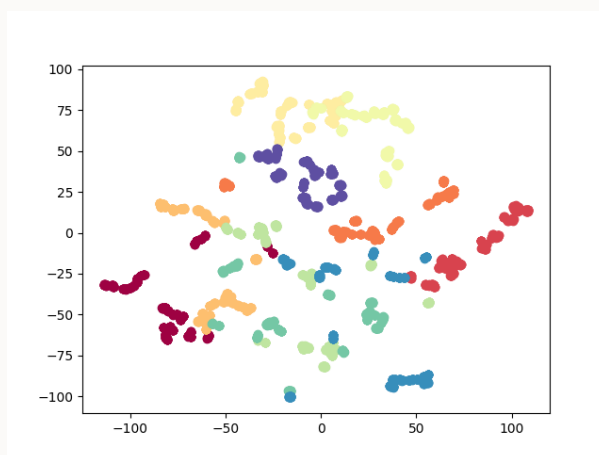


图 9: 随机初始化模型分类准确率 73.89%

5.10 finetune evaluation 中训练结果的稳定性验证 (`code : exp7`)

为了验证 finetune 模式下太极拳分类结果的稳定性，我们在 10% 样本作为训练集，90% 样本作为测试集的基础上，对于不同的数据预处理配置进行实验，查看分类的准确率，如表7所示，通过实验结果可以看出所有的数据预处理配置下都有很好的分类效果，并且数据预处理对于提升训练结果有着正面的作用，证明了预训练模型和数据预处理的有效性。

Cfg	Acc	Cfg	Acc	Cfg	Acc
NSNR	100.00%	NFNC	98.33%	TEST9AUG2	100.00%
NSWR	100.00%	NFWC	98.33%	TEST9AUG4	100.00%
WSNR	100.00%	WFNC	100.00%	TEST9AUG8	100.00%
WSWR	98.33%	WFWC	100.00%		

表 7: Finetune Preprocess Results Under TEST9