指导教师：杨涛　　提交时间：2016/3/14

CVPR2015 Paper Translation

No: 01
姓名：隆文喜
学号：2013302525
班号：10011303

分层回归神经网络骨架的动作识别

杜勇，王炜，王亮
研究中心的智能感知和计算，CRIPAC
模式识别 Nat'l 实验室，自动化研究所，中国中院院士科学
{yong.du，王伟，wangliang}@nlpr.ia.ac.cn

绪论

人动作可由的轨迹来表示骨骼关节。传统的方法一般建模空间结构和人体骼的时空动态用手工制作的特点和认识到人类活动精心设计的分类。在本文中，考虑到经常性神经网络（RNN）可以模拟长期上下文以及时间序列的信息，我们建议一个终端到终端的分层 RNN 基于骨架行动承认。而不是采取整个骨架作为输入，我们根据人体骨骼分为五个部分划分对人体生理结构，然后分别馈送出他们五子网。作为层的数量增加时，由子网提取的表示是分层稠合到更高的层的输入端。最后陈述骨架序列被送入一个单层感知器，和的时间上累积输出感知是最后的决定。我们与其他五个比较从我们的模型得出深 RNN 架构来验证所建议的网络的效率，并且也比较有三个公开可用的几种其他方法数据集。实验结果表明，我们的模型实现国家的最先进的性能与较高的计算效率。

1.简介

图 1：建议的分级的说明性的草图反复发作的神经网络。整个骨架分分为五部分，其中被送入 5 个双向复发神经网络（烧伤）。作为层的数量增加，由子网中提取的表示是分层融合到更高的层的输入端。 一个完全连接层和 SOFTMAX 层上进行最终表示以动作进行分类。

随着计算机视觉的一个重要分支，动作识别具有广泛的应用范围，例如，智能视频监控，机器人视觉，人机交互，游戏控制，等等[15，36]。传统的研究关于动作识别主要集中在认识行动从 2D 摄像机录制的视频。但实际上，人的行为一般都代表和认可 3D 空间。人体可视为一个铰接系统包括刚性骨骼和关节铰接其中进一步组合成四肢和躯干[31]。人的行动是由这些肢体的运动和躯干其通过人骨骼关节的三维空间[37]的运动表示。目前，可靠关节坐标可以从成本效益来获得使用实时骨架估计算法深度传感器[27，28]。有效的办法应进行调查基于骨架作用的认可。人类骨架的动作识别一般认为是一个时间系列的问题[5,17]，其中身体姿势的特点和他们对动态时间提取到代表一个人的行为。大多数现有的基于骨架动作识别方法明确通过骨骼模型关节的时空动态利用时间金字塔（TPS）[19，31，33]和隐藏马尔可夫模型（HMM）[20，34，35]。TPS 的方法通常由时间窗口的宽度的限制并且只能利用有限的上下文信息。至于隐藏式马尔克夫模型，这是非常困难的，得到的时间比对的序列和相应的发射分布。最近，递归神经网络（人工神经网络）与长 - 短术语存储器（LSTM）[8,10]神经元已被用于动作识别[1，11，16]。所有这些工作只使用单 RNN 层作为序列分类没有基于部分的上进行最终表示以动作进行分类特征提取和层次的融合。

在本文中，建模取深 RNN 充分利用时间序列的长期的上下文信息我们

提出了基于骨架层次 RNN 行为识别。图。图 1 示出所提出的体系结构网络，其中的时间表示低级别的身体部位是由双向模拟复发神经网络（巴恩斯），并组合成的陈述高层部件。

人体大致可以分解成五个部分，例如，两只胳膊，两条腿和一个行李箱，而人类行为是由这些身体部位的动作。特定这个事实，我们把人体骨骼进五相应的
部件和养活他们分为五个双向反复地连接的子网（BRNNs）在第一层中。至运动从邻近的骨架零件建模，我们串联躯干子网与表示与其它四个子网，分别再投入这些级联结果四个 BRNNs 在第三如图所示层。 1.随着类似的操作，在上半身，下半身及的表示全身都在第五和第七层分别获得。到现在为止，我们已经完成了表示骨架序列的学习。最后，完全连接层和一个 SOFTMAX 层所获得的表示进行对动作进行分类。应当指出的是，克服消失梯度问题，当训练 RNN [8,12]，我们采用在最后 BRNN LSTM 元层。

在实验中，我们与其他五个比较深从我们提出的模型衍生 RNN 架构来验证
所建议的网络的有效性，并比较与三个可公开获得的数据集的几种方法。实验结果表明，我们的方法实现国家的最先进的性能与较高的计算效率。我们工作的主要贡献可以概括如下。首先，以我们所知，我们是第一个为骨架的终端到终端的解决方案通过使用分层复发基于动作识别神经网络。其次，与其他五个比较得出
深 RNN 架构，我们验证的有效性所建议的网络的必要的部分，例如，双向网络，LSTM 神经元在最后 RNN 层，分级骨架零件的融合。最后，我们证明
我们提出的模型可以处理基于骨架行动识别非常好没有复杂的预处理。

本文的其余部分安排如下。 在第 2 节，介绍了基于骨架相关工作行为识别。在第 3 节中，我们先回顾一下背景 RNN 和 LSTM，再说明的细节所提出的网络。实验结果与讨论在第 4 节给出最后，我们总结在第 5 节。

2.相关工作

在本节中，我们简要回顾一下现有的文献紧密涉及提出的模型，其中包括三个代表时间动态类的方法由地方特色，连续的状态转换和 RNN。

具有地方特色的方法通过聚类提取关节分为五部分，Wang 等人。 [32]使用的空间和部件的颞字典来表示的动作，它可以捕获人体的空间结构和动作。乔德里等。 [2]编码骨架结构具有空间 - 时间的层次结构，并利用线性动力系统学习动态特性。 Vemulapalli 等。 [31]利用旋转和平移代表在李的身体部位的三维几何关系组，然后采用动态时间规整（DTW）和傅里叶时空金字塔（FTP）建模的时间动力学。相反造型时空演化的特征，Luo 等人。 [19]开发一种新型学习词典
法时空金字塔匹配相结合，保持时间动态。为了表示人类运动及相关对象，王等人。 [33]第一次提取来自各地的外观局部　　占用模式骨骼关节，然后处理它们与 FTP 获取临时结构。扎姆菲尔等。 [38]提出了一种运动姿势描述捕获姿势和骨骼关节。运用 5 关节坐标和它们的时间差异输入，赵和陈[4]一个执行动作识别混合多层感知。在上述方法中，所述本地时空动态一般是在一定代表时间窗口或差量，就不能在全球范围捕捉行动的时间演化。

具有连续状态过渡的途径等吕人。 [20]个人提取和部分组合的局部特征
关节，培养 HMM 模型捕捉动作动力学。基于骨骼关节功能，吴邵[34]采用了深刻的前向神经网络来估算隐藏状

态 HMM 中的发射概率，进而推断动作场面。为了准确计算具有动态集成块两个序列之间的相似性整经，功等。[5]执行这两个时间分割并与结构化的时间序列表示对齐。虽然 HMM 模型能够时空演化的动作，将输入序列已经被分割并对齐，这本身是一个非常困难的任务。

与 RNN 接近 RNN 的结合感知器可以直接归类的序列，没有任何的分割。通过获得与顺序表示 3D 卷积神经网络，Baccouche 等。 [1]提出一个 LSTM-RNN 认识行动。关于光流作为输入的直方图，Grushin 向量等。 [11]使用 LSTM-RNN 的稳健动作识别并取得良好结果在 KTH 数据集。考虑到 LSTM，RNNs 就业在[1]和[11]都只有一个隐藏层，列弗斐尔等单向的。[16]提出了一种双向 LSTM-RNN 一个向前隐层和一个落后隐层手势分类。

以上所有只是工作使用 RNA 作为序列分类而我们提出一个终端到终端的解决方案包括学习功能和顺序分类。考虑到事实上，人的行为是由的议案人体部位，我们使用运行在一个分层的方式。

3.我们的产品型号

为了把我们提出的模型的来龙去脉，我们先回顾回归神经网络（RNN）和长 - 短短期记忆的神经元（LSTM）。然后，我们提出了一个层次双向 RNN 解决骨架的问题基于行为识别。最后，五相关深 RNNs 也有介绍不同的架构。

3.1。RNN 和 LSTM 回顾

RNN 和前馈的主要区别网络是反馈回路存在即产生
展开网络中的经常性的连接。随着复发性结构，RNN 可以模拟上下文信息的时间序列。给定一个输入序列 X =（X 0，...，XT-1），复发性层的隐状态 H =（H0，。，HT-1）和单个的输出隐层 RNN Y =（Y0，。，YT-1），可以得出如下[8，9，10]。

HT= H（Wxhxt+ Whhht-1 + BH）（1）

YT= O（Whoht+ BO）（2）

其中 Wxh，WHH，谁从表示连接权值输入层 X 到隐层 h 时，隐层 Hj 自身和分别隐藏层到输出层 Y，BH 博是两个偏置向量，H（ •）和 O（ •）是在隐藏层和输出层的激活功能。

通常，这是非常困难的训练 RNNs（特别是深运行）与常用的激活功能，例如，双曲正切和 S 型函数，由于消失梯度和错误吹起来的问题[8,12]。为了解决这些问题，长短期记忆（LSTM）架构已经提出[10，13]，它取代了非线性台传统 RNNs。如图 2 示出一个 LSTM 存储器块与单个细胞。它包含一个自我连接存储器小区 c 和三个乘法单元，即输入门我，勿忘门 f 和输出门 0，其可以存储和访问远程上下文信息的时间序列。

存储器单元和所述激活三个门是给出如下：

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3)$$

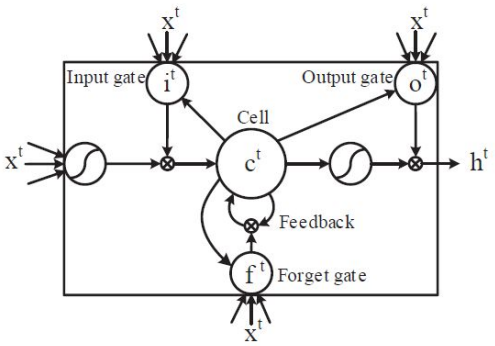
下图：长一个细胞短时记忆块[8]。

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (4)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (6)$$

$$h_t = o_t \tanh(c_t) \quad (7)$$



其中 σ （ •）是 S 形函数，并且所有的矩阵 W¯¯两个单元之间的连接的权重。为了利用每一个过去和未来环境序列中的点，舒斯特尔和 Paliwal[26]提出双向回归神经网络（GRNN），这呈现序列向前和向后两个单独经常隐藏层。这两个隐藏复发层共享相同的输出层。双向复发神经网络示于图。3. 应当指出我们可以很容易地获得 LSTM-RNN 仅仅通过更换非线性单元图。 3 LSTM 块。

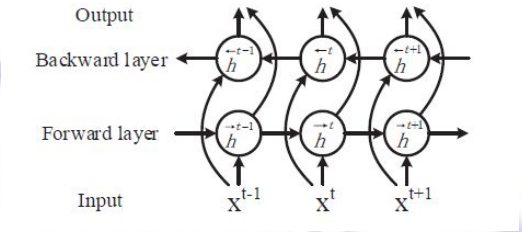


图 3：双向循环神经的体系结构网络[8]。

3.2。分层 RNN 为骨架的行动承认

根据人体的物理结构，人体骨骼可以分解成五个部分，例如，两个臂，两条腿和一条中继线。简单的人的行为是由他们的只是其中的一部分，例如执行，向前冲和脚踢转发主要依靠摆动手臂和腿。有些动作是来自移动上体或下半身，例如，向下弯曲主要 relatesto 上半身。更复杂的动作组成的这五个部分的运动，如跑步和游泳需要协调的胳膊，腿和躯干的运动。这些各个部件和它们的组合的运动是非常必要的。从 RNN 的力量受益访问内容的信息，我们提出了一个分层双向 RNN 基于骨架作用的认可。从传统的建模方法的空间不同结构和时空动态与手工制作的特点并通过精心设计的分类器识别的动作，我们的模型提供了行为识别终端到终端的解决方案。

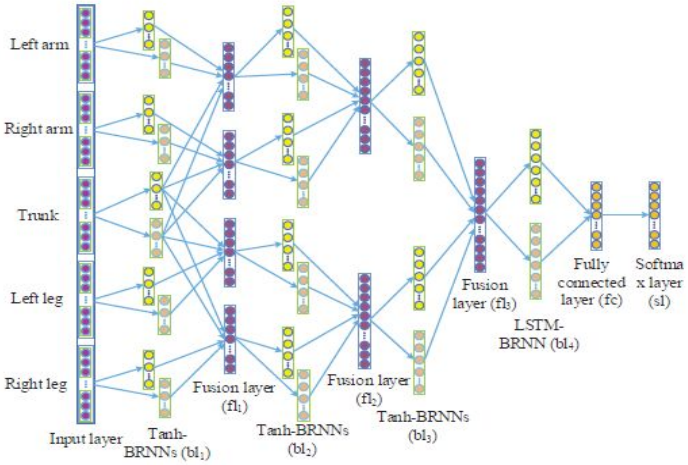


图 4：我们提出的模型的架构。

该模型的框架示于图 4.我们可以看到，我们的模型是由 9 层，即 BL1 - BL4，FL1 - FL3，FC 和 SL，其每一个呈现不同结构，从而在整个起着不同的作用网络。在第一层 BL1，五个骨架部件喂五个相应的双向连接反复地子网（BRNNs）。为了模拟邻近的骨架部件，例如，左手臂，躯干，右手臂，躯干，左下肢，躯干，和右腿，躯干，我们结合的代表性躯干子网与其他四个子网获取在融合层 FL1 四个新表示。如同该层 BL1，这导致 4 表示是分开送入在层 BL2 4 BRNNs。为了模拟全身上下，左臂躯干的表示和右手臂，躯干 BRNNs 进一步结合起来，获得上身的代表性，而交涉左腿，躯干和右腿干线 BRNNs 是组合以获得在下体表示层 FL2。最后，新获得的两个表示被送入在层 BL3 2

BRNNs，并且表示在层 BL3 这两 BRNNs 被再次熔融以表示在该层 FL3 全身。全身表现的 temporaldynamics 进一步建模通过在层 BL4 另一个 BRNN。从视点特征的学习，这些层叠 BRNNs 可以考虑提取骨架的空间和时间特性序列。获得所述骨架的最终特性后序列，完全连接层 FC 和 SOFTMAX 层 SM 执行的行动进行分类。

如第 3.1 节所提到的，LSTM 架构有效克服而消失梯度问题训练 RNN[8，12，13]。但是，我们只是采用 LSTM 在过去经常层神经元(BL4)。前三 BRAIN 层全部采用双曲正切激活功能。 这是一个提高表达能力之间的权衡避免过度拟合。一般地，权重的在一个数 LSTM 块比的数倍的双曲正切神经元。这是很容易过度拟合与有限的培训网络序列。

3.3。 训练

培训该模型包含了一个向前传球和一个落后的通行证。直传：对于在时间 t 第 i 个谷仓层 BLI，鉴于第 q 个输入它智商和双曲正切激活函数，相应前层和后向的第 q 表示层定义如下

$$\overrightarrow{h}_{i_q}^{t} = tanh(W_{I_{i_q}\overrightarrow{h}_{i_q}} I_{i_q}^{t} + W_{\overrightarrow{h}_{i_q}\overrightarrow{h}_{i_q}} \overrightarrow{h}_{i_q}^{t-1} + b_{\overrightarrow{h}_{i_q}}) \quad (8)$$

$$\overleftarrow{h}_{i_q}^{t} = tanh(W_{I_{i_q}\overleftarrow{h}_{i_q}} I_{i_q}^{t} + W_{\overleftarrow{h}_{i_q}\overleftarrow{h}_{i_q}} \overleftarrow{h}_{i_q}^{t+1} + b_{\overleftarrow{h}_{i_q}}) \quad (9)$$

在所有的矩阵 W，向量 b 被相应连接权重和偏见。

对于在时间 t 以下熔融层 FLI，第 p 新级联表示形式的输入第（i +1）个谷仓层 BLI+ 1

$$I_{(i+1)_p}^{t} = \overrightarrow{h}_{i_j}^{t} \oplus \overleftarrow{h}_{i_j}^{t} \oplus \overrightarrow{h}_{i_k}^{t} \oplus \overleftarrow{h}_{i_k}^{t} \quad (10)$$

在哪里表示连接运算符，他们是正向层的隐藏的陈述和在第 i 布兰层的第 j 个部分的向后层，从第 i 层中的第 k 个部分。

$$O^{t} = W_{\overrightarrow{h}_{bl_4}} \overrightarrow{h}_{bl_4}^{t} + W_{\overleftarrow{h}_{bl_4}} \overleftarrow{h}_{bl_4}^{t} \quad (11)$$

最后，该层 fc 的输出被累积跨越 T 框架序列，结果累积{}阿克通过 SOFTMAX 功能标准化获取每个一流的概率 P（CK）：

$$\mathcal{A} = \sum_{t=0}^{T-1} O^{t} \quad (12)$$

$$p(C_k) = \frac{e^{\mathcal{A}_k}}{\sum_{i=0}^{C-1} e^{\mathcal{A}_i}} \quad (13)$$

在这里有人类活动的Ç类。
我们的模型的目标函数是最小化最大似然损耗函数[8]：

$$\mathcal{L}(\Omega) = -\sum_{m=0}^{M-1} \ln \sum_{k=0}^{C-1} \delta(k-r) p(C_k|\Omega_m)$$

在哪里&（ •）是克罗内克函数，r 表示序列&M 接地真相标签。有 M 序列在训练集&。向后传球：我们通过使用反向传播时间（BPTT）算法[8]来获得的衍生物目标函数相对于所有的权重，和最大限度地减少随机梯度下降的目标函数
[8]。

3.4。五比较架构

为了验证所提出的网络的有效性，我们与其他五个不同的架构比较从我们提出的模型的。如前所示，我们提出了一个层次双向 RNN（HBRNNL）基于骨架动作识别（后缀“-L”也就是说，只有在过去经常层由 LSTM 的神经元，其余同样）。为了证明的重要性双向连接，类似的网络单向的连接被提出，这是所谓的分层单向 RNN（HURNN-L）。要验证基于部分的特征提取和层次融合的作用，我们比较了深刻的双向 RNN（DBRNN-L），这是直接叠置若干 RNNs 与整个人类骨骼作为输

入。此外，我们比较深单向 RNN（DURNN-L）不采用两个双向连接和层次融合。为了进一步探讨是否 LSTM 元最后反复层为克服消失有用/在 RNN 爆炸的问题，我们研究另外两个架构 DURNN-T 和 DBRNN-T。这里 DURNN-T 和 DBRNN-T 是类似于网络 DURNN-L 和 DBRNN-L，而在所有层中的双曲正切激活函数。应当指出，所有的六个结构有五个可学习的层，即四个经常隐藏层和一个完全连接层。并在神经元的数目完全连接层等于的动作类别。

4.实验

在本节中，我们评估我们的模型，并与比较其他五种不同的架构和几个最近的工作三地基准数据集：MSR Action3D 数据集[18]伯克利多式联运人类行为数据集（伯克利 MHAD）[22]，以及动作捕捉数据集 HDM05[21]。我们还讨论了过学习问题和计算该模型的效率。

4.1。评估数据集

MSR Action3D 数据集：它是由 Microsoft 生成，超高动力学状深度传感器，它被广泛应用在行动承认。此数据集包括执行 20 个动作 10 科目不受约束的方式为两个或三个，557 与 22077 帧有效样本。所有序列在 15 FPS 被捕获，并且在一个序列中的每个帧包含 20 骨骼关节。骨架的低精度关节坐标和部分片段在某些缺序列使这个数据非常具有挑战性。

伯克利 MHAD：它是由一个多峰采集中捕获系统，其中光学运动捕捉系统用于捕获活性 LED 标记的三维位置以 480 赫兹的频率。该数据集包含 65912 受试者进行 5 对 11 的动作的顺序每个动作的重复。在序列的每一帧，有 35 关节准确地提取根据 3D 标记轨迹。

动作捕捉数据集 HDM05：它是由一个捕获光学基于标记的技术的 120 的频率
赫兹，其中包含 130 行动 2337 系列进行 5 非职业演员，和 31 节中的每个帧。据我们所知，该数据集是目前最大深度序列数据库提供的骨架行动识别联合坐标。如上述[4]，这些 130 动作的部分样品应分为同一类别，例如慢跑空气和慢跑开始从地面都是一样的动作开始，慢跑 2 步和慢跑 4 个步骤属于相同“慢跑”动作。
样品组合之后，动作减少到 65 的类别。

4.2。数据预处理和参数设置

在我们提出的模型中，所有的人体骨骼关节主要分为五个部分，即两条胳膊，两条腿和一个中继线，这是在如图 3 所示。 5.我们可以看到有 4 接头的 MSR Action3D 数据集的每一个部分。 对于伯克利 MHAD 和 HDM05 数据集，联合号
的胳膊，腿和躯干列示如下：7，7,7 和 7，5，7。鉴于人类的行为是独立于它的绝对空间位置，我们正常化骨骼关节的统一坐标系。坐标系统的原点定义如下

$$\mathcal{O} = (\mathcal{J}_{hip\_center} + \mathcal{J}_{hip\_left} + \mathcal{J}_{hip\_right})/3 \qquad (15)$$

表 1：我们提出的模型的参数设置和五款相比在三个评价数据集。该 DU.T 短为 DURANT，其余同样。的 LLI 表示第 i 可学习层（HBR NL BLI）。

Layer	MSR Action3D						Berkeley MHAD & HDM05					
	DU.T	DB.T	DU.L	DB.L	HU.L	HB.L	DU.T	DB.T	DU.L	DB.L	HU.L	HB.L
$LL_1(bl_1)$	80	40	80	40	30 × 5	15 × 2 × 5	90	60	80	40	40 × 5	15 × 2 × 5
$LL_2(bl_1)$	120	80	120	80	60 × 4	30 × 2 × 4	180	120	160	80	80 × 4	30 × 2 × 4
$LL_3(bl_3)$	240	120	180	100	90 × 2	60 × 2 × 2	240	120	180	100	100 × 2	60 × 2 × 2
$LL_4(bl_4)$	120	80	80	40	80 × 1	40 × 2 × 1	120	60	90	60	90 × 1	60 × 2 × 1

为了提高信号的原始数据的信噪比，我们采用一个简单 Savitzky - 格雷平滑滤波器[25]预处理数据。该滤波器被设计如下
F1 = （ -3xi-2 +12xi-1 +17xi+12xi+1-3xi+2）/ 35（16）
其中 xi 表示骨骼关节中的第 i 个坐标帧，以及网络连接表示滤波结果。

考虑到骨架关节的运动轨迹平滑变化，我们从序列采样帧定的时间间隔，以减少计算成本。有取样伯克利 MHAD 数据每 16 帧每 4 帧的 HDM05 数据集。我们不从 MSR Action3D 数据，由于有限的样本框帧速率（15 FPS）和平均长度（小于 40 帧）。

图 1 示出的我们的提议的参数设置模型和三种评价的五个车型相比，数据集。表中的每个值表示的数目在相应的层中使用的神经元，例如，数 30×5（LL1，HU.L）意味着每个单向子网在 HURNN-L 的第一可学习的层具有 30 个神经元，并且×2×5（LL1，HB.L）的数量 15 表示每个双向子网中 HBRNN-L 的第一 BRNN 层（BL1）拥有 15×2 个神经元。对同一数据集这六大网络有大致相同的数量的权重。

应当指出，所有的比较的结果在三个数据集的方法是从它们的相应文件。

4.3。实验结果与分析

MSR Action3D 数据集：虽然有几种验证在[24]总结了该数据集的方法，我们按照在[18]中提供的标准协议。在这个标准中协议，该数据集被分成三个操作将 AS1，AS2 和 AS3。受试者 1，3，5，7 的样品中，9 是用来为同时的受试者 2，4，6，8 的样品训练，10 是用于测试。我们比较了模型 HBRNNL 李等人。 [18] Chen 等。 [3]，Gowayyed 等。 [6]，Vemulapalli 等。 [31]等五种变型架构 DURNN-T，DB RNN-T ， DURNN-L ， DBRNN-L，HURNNL。实验结果示于表。 2.我们可以看到我们的建议 HBRNN-L 达到最佳的平均精度和优于在四种方法[3，6，18，31]用手工制作的特点和派生的两个演出车型 HURNN-L 和 DBRNN-L 是有希望的。它应当指出的是，虽然 Chen 等。 [3]和 Vemulapalli
等。 [31]达到动作集最佳的性能 AS1 和 AS3，分别，我们的 HBRNN-L 性能优于它们相对于平均准确性。此外，HBRNN-L 执行一贯好这三个动作集，这表明 HBRNN-L 是更健壮的各种数据。

图 5：人类骨骼关节分成五个份在这三个数据集。

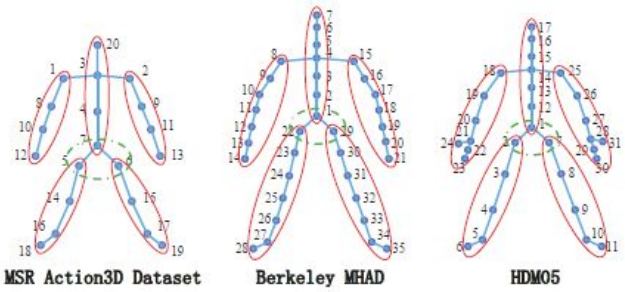
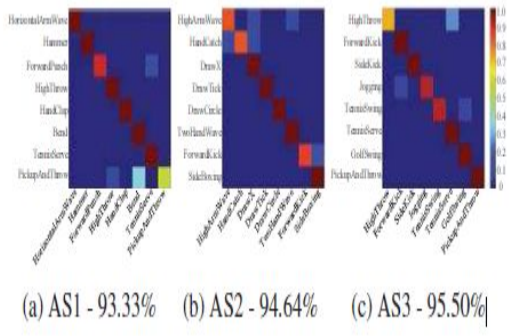


表 2：在 MSR Action3D 实验结果数据集。

Method	AS1	AS2	AS3	Ave.
Li *et al.*, 2010 [18]	72.9	71.9	79.2	74.7
Chen *et al.*, 2013 [3]	**96.2**	83.2	92.0	90.47
Gowayyed *et al.*, 2013 [6]	92.39	90.18	91.43	91.26
Vemulapalli *et al.*, 2014 [31]	95.29	83.87	**98.22**	92.46
DURNN-T	75.24	75.00	81.08	77.11
DBRNN-T	81.90	80.36	88.29	83.52
DURNN-L	87.62	91.96	90.01	89.86
DBRNN-L	88.57	93.75	95.50	92.61
HURNN-L	92.38	93.75	94.59	93.57
HBRNN-L	93.33	**94.64**	95.50	**94.49**

这 HBR NL 获得较高的平均准确的事实比 HURNN-L ， DBR-L 和 DURNN-L，证明双向连接和分层的重要性特征提取。所有 LSTM 神经元网络在过去的复发层（后缀“-L”），均优于他们与正切激活功能的相应网络（带后缀“-T”），验证 LSTM 的有效性神经元所提出的网络。



(a) AS1 - 93.33% (b) AS2 - 94.64% (c) AS3 - 95.50%

图 6：MSR Action3D HBR NL 的混乱矩阵数据集。

这三个行动组的混乱矩阵的图。6.我们可以看到，错误分类主要发生在几个非常类似的行动。例如在图 6A，动作“PickupAndThrow”往往是错误分类为“弯曲”，而行动“则转发冲“被错误分类为”TennisServe“。其实，“PickupAndThrow”只是多了一个“扔”的举动比“德”和“扔”经常移动持有的序列中几帧。因此它是非常难以区分这两种动作。的行动“ForwardPunch 和“TennisServe”分享序列中的一个很大的重叠。他们区别也与只关节坐标非常具有挑战性。

伯克利 MHAD：我们按照实验方案这个数据集[22]提出的。的 384 序列前 7 个科目用于而 275 序列训练最后 5 受试者的用于测试。我们比较我们提出的模型 Ofli 等。[23]，Vantigodi 等。[30]，Vantigodi 等。[29]，Kapsouras 等。[14]，
乔德里等。[2]，以及 DURNN-T，DB RNN-T，DURNN-L，DBRNN-L，HURNN-L。实验结果示于标签。3，我们可以看到，HBRNN-L 达到 100％的准确率以简单的预处理并且执行比这五个衍生 RNN 架构更好，这证明了该模式的优点再来一次。同时，六 RNN 架构获得更高的精确度比 Ofli 等。[23]，Vantigodi 等。[30]，Vantigodi 等。[29]，Kapsouras 等。[14]，和可比较的结果与乔德等。[2]，这意味着我们的提议模型提供了一个有效的终端到终端的解决方案在建模动作序列时空动态。

HDM05：我们遵循中提出的实验方案[4]，并在此数据集进行了 10 倍交叉验证。我们比较我们提出的模型与 Cho 和陈[4]其他五个架构 DURNN-T，DB RNN-T，DURNNL，DBRNN-L，HURNN-L。实验结果是在显示选项卡。4.该模型 HBRNN-L 成于 96.92％的国家的最先进的精度为 0.50 斯坦 - 准偏差。派生模型 HURNN-LDBRNN-L 和 DURNN-L 也取得了优异的成绩。

表 3：对伯克利 MHAD 实验结果。

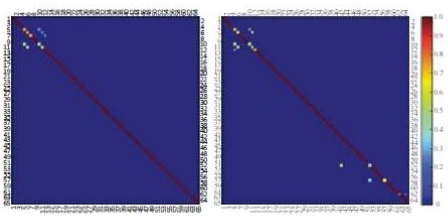
Method	Acc.(%)	Method	Acc.(%)
Ofli *et al.*, 2014 [23]	95.37	DURNN-T	98.55
Vantigodi *et al.*, 2013 [30]	96.06	DBRNN-T	99.27
Vantigodi *et al.*, 2014 [29]	97.58	DURNN-L	98.55
Kapsouras *et al.*, 2014 [14]	98.18	DBRNN-L	99.64
Chaudhry *et al.*, 2013 [2]	99.27	HURNN-L	99.64
Chaudhry *et al.*, 2013 [2]	100	**HBRNN-L**	**100**

表 4：在 HDM05 实验结果。

Method	Ave.(%)	Std.
Cho and Chen, 2013 [4]	95.59	0.76
DURNN-T	94.63	1.16
DBRNN-T	94.79	1.11
DURNN-L	96.62	0.53
DBRNN-L	96.70	0.51
HURNN-L	96.70	0.41
HBRNN-L	**96.92**	**0.50**

图 7：HBR NL 的两种典型的混乱矩阵在 HDM05 数据集。在水平的数字和

垂直轴对应于动作类别[4]。



(a) 97.52% (b) 96.04%

10 倍交叉验证的两个典型的混淆矩阵从 HBRNN-L 示于图 7.我们可以看到我们的模型在大多数的行动表现良好。错误分类主要来自以下类：“5-depositHighR”,“6-depositLowR”,“7-depositMiddleR“,”10-grabHighR“,”11-grabLowR“,和“12 grabMiddleR”。进一步检查“抢”和“定金”相关骨架序列，我们发现这两个动作类别共享相似的空间和时间变化。两者可以分解成三个子行动顺序为：伸出一只手，抢夺或沉积的东西了，往后退了手。抓取和沉积之间的细微差异东西使得难以区分这两种种动作。应当指出的是，虽然原始 130 动作减少到 65 类，还有几个混乱的类别，例如，“39 sitDownChair”和“42-sitDownTable”，它应属于同样的动作。无行动的情况下，例如，认识到椅子并从它们的外观表，这是非常难以区分这些行动只是从骨架流。

4.4。讨论

过度拟合的问题：实验表明，该模型后缀“-L”很容易，而其他与过度拟合
后缀“-T”在训练中总是 underfit。它可以是通过使用正切激活功能梯度消失问题在所有图层。为了克服过拟合在我们提出的 HBRNN-L 的问题和其他衍生的网状后缀“-L”，我们采用像加策略输入噪声，重量噪声和早期停止[7,8]。在我们的实践中，我们发现，增加的重量噪音更比添加所述输入噪声有效，而 commonlyused 辍学策略[39]在这里不起作用。对于欠拟合后缀“-T”型号的问题，我们使用通过调整学习率和增加再培训策略各级输入噪声和重量的噪音。计算效率：我们采取伯克利 MHAD 数据集为一个例子来说明的效率 HBRNN-L。用 C ++在 3.2GHz 的 CPU 执行系统中，我们花费 50 多岁的每个时期由 384 序列（每个序列平均 127ms）在训练中。后约 30 时代，我们可以得到一个精度大于 98％。在测试过程中，它需要每序列 52.46 毫秒（约每个序列 234 帧）。应当提到的是
HURNN-L，它实现了与相当的性能 HBRNN-L 运行得更快，更适合于在线应用程序。

五，结论和未来工作

在本文中，我们提出了一个终端到终端的分层复发神经网络基于骨架作用的认可。我们首先将人体骨骼分为五个部分，然后将它们喂到五个子网。作为层的数量增加，在子网表示是分层稠合到更高的层的输入端。感知器是在骨架序列的教训进行交涉以获得最终识别结果。试验三公开可用的数据集，结果表明，所建议的网络的有效性。

正如我们分析了 HDM05 和 MSR Action3D 数据集，所述类似人类的行为是非常困难的是刚刚从骨骼关节尊贵。在未来，我们会考虑更多的功能合并到建议
分层回归神经网络，例如，对象外观。

确认

这项工作是由国家基础研究的支持中国的计划（2012CB316300）和国家自然科学中国（61175003，61135002 科学基金，61202328，61420106015，U1435221）。

参考：

[1] M. Baccouche，F Mamalet，沃尔夫，C・加西亚，以及 A. Baskurt。连续深度学习人类行为承认。在人类行为的理解，29-页 39.施普林格，2011 年 1，2

[2] R.乔杜里 F. Ofli，G Kurillo，R. Bajcsy 和 R.比达尔。仿生动态 3D 歧视性的骨骼特征为人类行为的认可。在计算机上的 IEEE 会议视觉与模式识别研讨会，471-478 页。IEEE，2013 年 2,7

[3]陈 C.，刘 K.和 N. Kehtarnavaz。实时的人类行为基于深度运动图的认可。杂志实时图像处理，1-9 页，2013 年 6

[4] K. Cho 和 X.陈。分类和可视化运动拍摄使用深层神经网络的序列。科尔，
ABS / 1306.3874，2013 年 2，5，7

[5] D.功，G Medioni 和 X.赵。结构化时间序列分析人的行为分割和识别。
IEEE 交易模式分析与机器智能，VOL。36，NO。 7，2014 年 1，2

[6] M. A. Gowayyed，M. Torki，M. E.侯赛因和 M.埃尔 - 萨班。面向位移（HOD）的直方图：描述轨迹人体关节的行为识别。在国际人工智能，网页联席会议 1351 年至 1357 年。AAAI 出版社，2013 年 6

[7] A.坟墓。实用变推断神经网络。在神经信息处理系统的进步，2348 年至 2356 年页，2011 年 8

[8] A.坟墓。复发性监督序列标注神经网络，体积 385.施普林格，2012 1，2，3，4，5,8

[9] A.格雷夫斯和 N. Jaitly。朝向端至端的语音识别反复发作的神经网络。在国际会议机器学习，页 1764 至 1772 年，2014 年 3

[10] A.格雷夫斯 A.穆罕默德和 G 韩丁。语音识别深递归神经网络。在 IEEE 国际声学会议，语音和信号处理，6645-6649 页。 IEEE，2013 年 1，3

[11] A. Grushin 向量，D. D. Monner，J. A.的 Reggia 和 A.米什拉。强大的通过长短期记忆人体动作识别。在神经网络，网页际联席会议 1-8。 IEEE，2013 年 1，2

[12] S. Hochreiter， Y Bengio， P Frasconi 和 J.施米德休伯。梯度流动复发网：学习困难长期相关性，2001 年 2，3，4

[13] S. Hochreiter 和 J.施米德休。长短期记忆。神经计算，9（8）：1735 至 1780 年，1997 年 3，4

[14]一 Kapsouras 和 N.斯尼古拉迪斯。在动作识别使用 dynemes 转发 differ-动作捕捉数据分配办法表示。 J.可见。共同事业。形象代表，25（6）：1432 年至 1445 年 8 月到 2014 年 7

[15] J。Koutn'ık，J 施米德休伯和 F.戈麦斯。深演变基于视觉的增强非监督卷积网络学习。在遗传与进化会议计算，541-548 页。 ACM，2014 年 1

[16] G.列斐伏尔，S. Berlemont，F Mamalet 和 C.加西亚。Blstm-RNN 基于 3D 手势分类。在人工神经网络和机器学习，381-388 页。施普林格，2013 年 1，3

[17] K. Li 和 Y.赋。通过发现人类活动的预测时间序列模式。在模式 IEEE 交易
分析和机器智能，VOL。 36，NO。 8，2014 年 1

[18]李 W.，Z.张和 Z 柳。基于一个动作识别 3D 点的包。在 IEEE 计算机协会会议计算机视觉与模式识别研讨会，页 9-14。 IEEE，2010 年 5，6

[19] J.罗，王 W.和 H.齐。集团稀疏和几何约束的字典学习从动作识别深度贴图。在计算机上的 IEEE 国际会议愿景，1809 至 1816 年的网页。 IEEE，2013 年 1，2

[20] F. LV 和 R. Nevatia。识别和 3-分割利用 HMM 和多类 adaboost 的 D 人力的行动。在计算机视觉欧洲会议，359-372 页。施普林格，2006 1,2

[21] M.穆勒，T.罗德，M.克劳森，埃伯哈特 B.，B。克鲁格，和 A.韦伯。文档动作捕捉数据库 hdm05。技术报告 CG-2007-2，Universität 大学波恩，2007 年 6 月五号

[22] F. Ofli，R.乔杜里，G Kurillo，R.比达尔和 R. Bajcsy。伯克利 mhad：一个全面的多式联运人的行动数据库。在 IEEE 研讨会上计算机中的应用愿景，53-60 页。 IEEE，2013 年 5，7

[23] F. Ofli，R.乔杜里，G Kurillo，

R.比达尔和 R. Bajcsy。最翔实的关节（smij）序列：一个新的表示人类骨骼行为识别。杂志视觉传达和图像表示，25（1）：24-
38，2014 年 7

[24] J. R.帕迪拉-L'opez，A. A.沙拉维和 F Fl'orez-Revuelta。所采用比较验证测试的讨论使用 MSR action3d 人体动作识别方法数据集。科尔，ABS / 1407.7390，2014 年 6

[25] A. Savitzky 和 M.J。格雷。平滑和分化通过简化最小二乘程序的数据。分析化学，36（8）：1627 至 1639 年，1964 年 6

[26] M.舒斯特尔和 K. K. Paliwal。双向复发神经网络。 IEEE TRANSACTIONS ON 信号处理 45（11）：2673 年至 2681 年，1997 年 3

[27] J.肖顿，T.夏普，A Kipman，菲茨吉本 A.，M Finocchio，A.布雷克，M.库克和 R.摩尔。实时人提出从单一的深度图像部分的识别。通讯 116-124，2013 年 1：ACM，56（1）条

[28] J.汤普森，耆那教 A.，Y LeCun 和 C. Bregler。联合训练卷积网络和用于人类的图形模型的姿态估计。预印本的 arXiv 的 arXiv：1406.2984，2014 年 1

[29] S. Vantigodi 和 V. B.拉达克里希南。动作识别使用元认知 RBF 神经网络运动捕捉数据分类。在智能传感器，传感器网络与信息加工，2014 年第九届 IEEE 国际会议在 1-6 页。 IEEE，2014 年 7

[30] S. Vantigodi 和 R.巴布文卡塔斯。实时的人类行为认可动作捕捉数据。在计算机视觉，模式识别，图像处理和图形，第四次全国会议上，1-4 页。 IEEE，2013 年 7

[31] R. Vemulapalli，F Arrate 酒店，和 R. Chellappa。人类行动承认在谎言中表示 3D 骨架为点组。在计算机视觉和模式识别 IEEE 会议识别，588-595 页。 IEEE，2014 年 1，2，6

[32] C.王某，王勇，和 A. L. Yuille。一种方法来 posebased 行为识别。在计算机上的 IEEE 会议视觉与模式识别，915-922 页。 IEEE，2013。2

[33] J.Wang，Z.刘 Y.Wu 和 J.元。矿业 actionlet 合奏对于深度摄像机动作识别。在 IEEE 会议计算机视觉与模式识别，页 1290-1297。 IEEE，2012 年 1，2

[34] D.吴邵 L.。利用分层参数网络基于骨骼关节的动作分割和承认。在计算机上的 IEEE 国际会议愿景，2014 年 1，2

[35] L. 霞，C.-C. 陈和 J. AGGARWAL。查看不变的人使用 3D 关节的直方图动作识别。在 IEEE 计算机协会会议计算机视觉和模式识别研讨会，20-27 页。 IEEE，2012 年 1

[36]十阳和 Y 天。对于行为识别超级法线使用深度序列。IEEE 会议计算机视觉与模式识别。 1

[37] M.烨，张问，王 L.，朱军，杨河和 J.胆。一个调查从深度数据人体运动分析。在 Timeof-飞行和深度成像。传感器，算法与应用，149-187 页。施普林格，2013 年 1

[38] M。Zanfir，M Leordeanu 和 C. Sminchisescu。移动姿势：低延迟的高效 3D 运动学描述行为识别和检测。在 IEEE 国际会议计算机视觉，2752 年至 2759 年的网页。 IEEE，2013。2

[39] W.萨伦巴，一 Sutskever 和 O. Vinyals。递归神经网络正规化。科尔，ABS / 1409.2329，2014 年 8 月