许多机器学习方法已被应用于人类活动识别。然而，这一领域仍面临许多技术挑战。其中一些挑战与计算机视觉和自然语言处理等其他模式识别领域共享，而另一些挑战是基于传感器的活动识别所特有的，需要专门的方法用于实际应用。这里列出了一些活动识别社区应该应对的挑战。

Annotation Scarcity

无监督学习主要用于探索性数据分析，发现数据之间的模式

近年来，深度生成模型如深度信念网络(DBNs)和自动编码器已经成为无监督学习的主流。dbn和自动编码器由多层隐藏单元组成。它们在从大量数据中提取特征和发现模式方面很有用。此外，与判别模型相比，深层生成模型对过度拟合问题具有更强的鲁棒性。因此，研究人员倾向于将其作为特征提取的过程来利用未标记的数据，因为收集未标记的活动数据集既简单又便宜

深度模型的生成式预训练指导判别训练获得更好的泛化解。因此，以无监督的方式对大规模未标记的数据集进行深度网络的预训练变得非常普遍。整个识别过程可以分为两个部分。首先，将输入数据输入到特征提取器(通常是深度生成模型)中进行预处理，以提取特征。然后加入一个顶层分类器或其他分类器，用带标签的数据进行训练，并以监督的方式进行分类。在监督训练过程中，可以对特征提取器中的权值进行微调。

尽管深度生成模型在人类活动识别的无监督学习中取得了成功，但由于无监督学习没有能够识别出活动的真实标签而没有能够呈现出地面真实的标签样本，因此无监督学习仍然不能独立地承担活动识别任务。因此，将上述方法视为半监督学习，即利用标记数据和未标记数据对神经网络进行训练

由于获取标记数据的困难，半监督学习已成为活动识别的一个新趋势

半监督方法需要较少的标记数据和大量的标记数据进行训练。

在自学习方法中，首先用少量的标记数据训练弱分类器。该分类器用于未标记样本的分类。对置信度高的样本进行标记，并加入标记集对分类器进行再训练。

在联合训练中，使用多个分类器，每个分类器使用一个单独的训练数据视图进行训练。同样地，分类器选择未标记的样本，通过信心评分或多数投票将其添加到标记集

为了减轻标记的负担，主动学习的目标是为注释者选择信息量最大的未标记实例，并使用这些数据标记和改进分类器，以便尽可能少地进行人工监督。

Heterogeneity

许多最先进的人类活动识别方法都假设训练数据和测试数据是独立且同分布的

传感器的异质性可分为三类。

第一个是用户的异质性，它是由不同的人执行不同的活动时产生的不同的运动模式。

第二个异质性与时间有关。在动态流环境中，活动的数据分布随时间而变化，并且可能出现新的活动。

第三类是传感器的异质性。用于人类活动识别的传感器通常是敏感的。一个小变化传感器会对传感器数据造成很大的干扰。可能导致传感器异构的因素包括传感器实例、类型、位置和环境中的布局。

Concept Drif： 它表示训练域和测试域(或源域和目标域)之间的分布偏移。

Concept Evolution： 流数据中出现的新活动

Open-Set： 开放集问题是当前的一个热门话题。在此之前，大多数最先进的工作是针对“封闭集”问题，其中训练集和测试集包含相同的活动集。开放集也源于我们在最初的训练阶段无法收集到足够种类的活动。

用于活动识别的传感器包括可穿戴传感器和环境传感器。由于传感器的敏感性，传感器的微小变化可能导致传感器收集或传输的数据发生重大变化。

Interpretability of Deep Learning Models in Sensory Data

人类活动的传感器数据是高维的、不可读的。在一个时间窗口内，一个数据样本可能包含来自测试体的多个位置(例如，手腕、脚踝)的不同模式(例如，加速度、角速度)

然而，只有少数几种模式来自特定的位置有助于确定某些活动。例如，当人处于水平(磁力)状态时，说谎是可分辨的，而人的踝关节向前和向上的加速度可以识别楼梯。不相关模态会引入噪声，降低识别性能。此外，模式的重要性随时间而变化。例如，在帕金森病检测系统中，异常只出现在短时间内，而不是整个时间窗口

例如，一个深度学习模型识别用户正在行走，我们倾向于知道哪个模态哪个时间间隔是决定因素。因此，深度学习方法的可解释性已成为人类活动识别的新趋势

可解释的深度学习方法的基本思想是自动确定输入数据的每个部分的重要性，通过忽略不重要的部分，聚焦于突出的部分来达到高精度。在

软”指的是可微的。软注意将从0到1的权重分配给输入的每个元素

软注意在注意层中使用softmax函数来计算权重，因此整个模型是一个完全可微的确定性机制，其中梯度可以通过软注意机制传播到网络的其他部分

硬注意力： 是否参加一部分投入或不参加。它分配给输入部分的权值不是0就是1，所以这个问题是不可微的。这个过程涉及到对要处理的部分进行一系列的选择。例如，模型关注输入的一部分以获取信息，并根据已知信息决定下一步去哪里。

例如，模型关注输入的一部分以获取信息，并根据已知信息决定下一步去哪里。选择可以通过神经网络输出。然而，由于没有地面真值来指示正确的选择策略，所以硬注意应该表示为一个随机过程。这就是深度强化学习的用用之处。深度强化学习解决了深度学习中的选择问题，允许模型在选择策略空间中传播梯度

通过深度强化学习，硬注意力可以用softmax函数训练，用反向传播训练标准梯度下降

Data Segmentation

由于原始传感器数据是通过连续的流信号来表示的，因此通常使用一个固定大小的窗口将原始传感器数据序列分割成段作为输入到模型中进行活动识别

理想情况下，一个分区的数据段只处理一个活动，因此对模型进行建模可以预测单个窗口中所有样本的单个标签。然而，当活动转换发生在窗口中间时，一个窗口中的示例可能并不总是共享相同的标签。因此，一种优化的分割方法是提高活动识别精度的关键。

一种直观的方式是经验地尝试各种固定的窗口大小。 它增加了过渡发生在窗户中间的可能性。相反，小窗口无法提供足够的信息。鉴于这个问题。有人用一种分层信号分割方法，该方法最初使用一个大的窗口大小，然后逐渐缩小分割范围，直到子窗口中只有一个活动。缩小准则是两个连续的窗口有不同的标签或分类器的置信度小于阈值。

**5. 复合行为（Composite Activities）**  
大多数人类行为识别任务是基于简单的活动，比如走路和坐着。然而，记录人类日常活动更有意义的方式是由一系列简单行为组成的复合行为。例如，「洗手」可以表示为 {打开水龙头，皂洗，搓手，关掉水龙头}。由于复合行为不仅仅需要识别人体活动，还需要识别周遭环境的信息，因此比识别简单行为更有挑战性。

**7. 并行行为（Concurrent Activity）**  
在真实的场景中，除了按顺序逐个执行每个行为之外，一个人可以同时进行多个行为，这被称为并行行为。例如，一个人可以在看电视时打电话。从传感器的角度来看，一段数据可能对应多个行为。因此，并行行为识别可以抽象为一个多标签任务（multi-label task）。  
**8. 多人行为（Multi-occupant Activity）**  
生活和工作空间通常由多个人居住，因此，设计解决多人行为问题的方案具有重要的现实意义。主要有两种类型的多人行为：1. 平行行为（Parallel Activity），比如一个人在吃饭，另一个人在看电视。2. 合作行为（Collaborative Activity），多个居住者合作执行相同的活动，如两个受试者打乒乓球。对于平行行为的识别，当只有可穿戴传感器时，可将其划分为多个单人行为识别任务，并采取传统的解决方案；当使用环境或对象传感器时，数据关联映射到多人的感知信号是主要挑战，并且随着空间中人数的增加而变得更加困难。合作行为通常包括人与人之间的交互并使用各种器械，因此，上下文和对象使用信息在设计识别解决方案中起着至关重要的作用。

Computation Cost

虽然深度学习模型在基于传感器的人类行为识别中显示出了主导作用，它们通常是资源密集型的。例如早期的 DCNN 架构，AlexNet，它有 5 个 CNN 层和 3 个全连接层，处理 61M 参数（249MB 内存），执行 1.5B 高精度操作进行预测。对于不可移植的应用程序，我们通常使用图形处理单元（GPU）加速计算。然而，GPU 是非常昂贵和耗电的，所以不适合用于移动设备上的实时应用程序。目前的研究已经证明了通过引入额外的层和节点来加深神经网络是一种关键的提高模型性能的方法，但是这样不可避免地增加了计算复杂度。因此，如何解决计算量大的问题实现实时性是一个非常重要和具有挑战性的课题。

考虑到深度神经网络在特征提取方面比浅层神经网络更有效，将人工和深度特征相结合是降低计算成本的潜在解决方案

手工制作的特征和神经网络的结合是在移动设备上实现实时活动识别的潜在方案。

作者结合了光谱特征，只有一个CNN层和两个完全连接的人类活动识别层。通过对四个基准数据集的评估，该混合体系结构显示出与最先进方法相比较的识别精度

Privacy

类行为识别的主要应用是对人类行为的监测，因此传感器需要不断地捕捉用户的活动。由于执行行为的方式因用户而异，所以对手可以通过时间序列传感器数据推断用户的敏感信息，如年龄。具体来说，对于深度学习技术而言，其黑箱特性可能会在无意中暴露出用户的鉴别特征。研究表明，即使 CNN 只接受针对行为分类的交叉熵损失训练，所获得的 CNN 特征仍然具有很强的用户识别能力。因此，解决深度学习模型的隐私泄露问题至关重要。简单的逻辑回归器在使用基本为活动提取的CNN特征时，其用户分类准确率高达84.7%，而相同的分类器对原始传感器数据的用户分类准确率仅为35.2%

借鉴计算机视觉社区的图像样式转换思想，一次性保护所有的私有信息。作者从两个方面创造性地看待原始传感器信号:“风格”方面，描述了用户如何执行一项活动，并受到用户的相同信息，如年龄、体重、性别、身高等的影响;“内容”方面，描述用户执行的活动。他们提出对原始传感器数据进行变换，使“内容”不变，但“风格”类似于随机噪声。因此，该方法具有保护所有敏感inf的潜力

提出扰乱目标函数的传统深auto-encoder执行ϵ-differential隐私。除了隐私保护特征提取层,还开发了一个ϵ-differential隐私保护softmax层分类或预测。与上述方法不同，该方法提供了理论上的隐私保证和误差范围

1. 作者想用统计学变量来创建feature vector