来自可穿戴式多模态传感器的自主人类活动分类

摘要： 依靠惯性测量单位（IMU）数据或来自提供第三人称视角的静态相机的数据，已有大量的人类活动分类研究工作。使用可穿戴式摄像机提供第一人称视角或以自我为中心的视图的工作相对较少，而将以自我为中心的视频与IMU数据结合起来的方法则更少。仅使用IMU数据限制了可以检测到的活动的多样性和复杂性。例如，可以通过IMU数据检测就座活动，但是不能确定对象是否坐在椅子或沙发上，或者对象在哪里。为了执行细粒度的活动分类，并区分仅使用IMU数据无法区分的活动，我们提出了一种可穿戴式摄像机和IMU的数据使用的自主且健壮的方法。与基于卷积神经网络的方法相比，我们建议采用胶囊网络从以自我为中心的视频数据中获取特征。此外，在以自我为中心的视频和IMU数据上都采用了卷积长短期记忆框架，以捕获动作的时间方面。我们还提出了一种基于遗传算法的方法，可以自动，系统地设置各种网络参数，而不是使用手动设置。已经进行了执行9-和26-标签活动分类的实验，并且该方法使用自主设置的网络参数提供了非常有希望的结果，总体准确度分别为86.6％和77.2％。与仅使用egovision数据和仅使用IMU数据相比，结合了两种模式的拟议方法还提高了准确性。

01: Machine learning methods for classifying human physical activity from on-body accelerometers

摘要：人体可穿戴式传感器的使用在许多学术和工业领域中都很普遍。它们在动态监控和普适计算系统中的应用引起了极大的兴趣。在此，对人体运动的一些定量分析及其自动分类是需要执行的主要计算任务。在本文中，我们讨论了如何使用人体加速度计对人类的身体活动进行分类，其中主要重点是用于此目的的计算算法。特别是，我们激发了我们对基于隐马尔可夫模型（HMM）的分类器的兴趣。通过分析加速度计时间序列的数据集来说明和讨论一个示例。

02: Accelerometry-based classification of human activities using Markov modeling

摘要：加速度计是人体运动传感器的一种流行选择：原因部分在于其提取信息的能力，这些信息可用于自动推断人体所参与的身体活动，此外还具有提供生物力学参数估算器的作用。人体活动的自动分类对于普适计算系统非常有吸引力，而情境感知可以缓解人机交互以及在生物医学中的应用，而可穿戴传感器系统则建议用于长期监控。本文涉及执行分类任务所需的机器学习算法。通过与高斯混合模型（GMM）分类器进行对比研究隐马尔可夫模型（HMM）分类器。HMM 将关于运动动力学的统计信息整合到分类过程中，而不会像GMM那样丢弃先前结果的时间历史记录。通过分析加速度计时间序列的两个数据集，说明并讨论了获得的统计杠杆作用的好处的一个示例。

03: Physical human activity recognition using wearable sensors

摘要：本文介绍了用于从可穿戴惯性传感器数据中识别人类活动的不同分类技术。在这项研究中使用了三个惯性传感器单元，健康受试者在上/下肢的关键部位（胸部，右大腿和左脚踝）佩戴了三个惯性传感器单元。三个主要步骤描述了活动识别过程：传感器的放置，数据预处理和数据分类。四种监督分类技术，即k最近邻（k-NN），支持向量机（SVM），高斯混合模型（GMM）和随机森林（RF），以及三种无监督分类技术，即k-均值，高斯混合模型（GMM）和隐马尔可夫模型（HMM）在正确分类率，F量度，召回率，精确度和特异性方面进行了比较。原始数据和提取的特征分别用作每个分类器的输入。使用基于RF算法的包装方法执行特征选择。根据我们的实验，获得的结果表明，与其他监督分类算法相比，k-NN分类器提供了最佳性能，而HMM分类器是无监督分类算法中提供最佳结果的分类器。这种比较突出显示了在有监督和无监督的情况下哪种方法可以提供更好的性能。应该注意的是，获得的结果仅限于本研究的背景，该研究涉及使用放置在受试者胸部，右胫骨和左脚踝上的三个可穿戴式加速度计对人类日常主要活动的分类。

04: Human activity recognition using thigh angle derived from single thigh mounted IMU data

摘要：准确的人类活动识别是许多领域中一个具有挑战性的研究主题。进行活动识别的常用方法是使用加速度计和/或陀螺仪检测躯干或腿部运动。本文提出了一种基于大腿角度来检测人类活动的新颖方法，该角度是使用单个大腿安装的惯性测量单元（IMU）的数据计算得出的。由于这项工作构成了正在开发的系统的一部分，该系统正在开发中以帮助室内导航中受损的视力，因此在算法开发过程中考虑了室内行人跟踪中常见的活动，例如坐，站和步行。该算法使用简单的信号处理技术，包括峰值检测，过零检测和计时器，以通过融合加速计和陀螺仪计算出的大腿角度来识别活动。这允许将算法植入通用的低端微控制器中。为了减少算法输入参数的数量，假设加速度计y轴与大腿对齐，从而陀螺仪x数据表示大腿向前和向后运动的角速度。该算法在有限数量的样本和理想测试条件下进行的测试中，已显示出站立检测的准确性高于78％，行走的准确性高于92％，并且没有测量到错误。这些结果表明，这种计算量较小的算法在室内行人导航应用中的活动检测中提供了可喜的结果。

05: A tutorial on human activity recognition using body-worn inertial sensors

摘要：在过去的20年中，人类活动识别领域的研究活动不断增加。随着活动识别已经相当成熟，在设计，实施和评估活动识别系统方面的挑战也越来越多。本教程旨在为新手提供有关人类活动识别领域的全面动手介绍。它特别关注使用人体惯性传感器进行活动识别。我们首先讨论人类活动识别与一般模式识别共同面临的主要研究挑战，并确定那些特定于人类活动识别的挑战。然后，我们将活动识别链（ARC）的概念描述为用于设计和评估活动识别系统的通用框架。我们详细介绍了框架的每个组成部分，为相关研究提供了参考，并介绍了活动识别研究社区开发的最佳实践方法。我们以教育示例问题作为结尾，该问题是通过连接到上臂和下臂的惯性传感器识别不同手势的方法。我们说明了如何针对此特定活动识别问题实现此框架的每个组件，并演示了不同实现的比较方式以及它们如何影响总体识别性能。

06: Motion primitive-based human activity recognition using a bag-of-features approach

摘要：使用可穿戴式传感器进行人体活动建模和识别在普及医疗保健中非常重要，其应用包括运动功能，康复和老年护理的定量评估。先前的人类活动识别技术使用“整体运动”模型，其中将连续的传感器流分为具有固定持续时间的窗口，并选择其长度，以便可以从每个窗口中提取每个活动信号中的所有相关信息。在本文中，我们提出了一种基于统计运动原语的框架，用于人类活动的表示和识别。我们的框架基于功能包（BoF），该功能使用原始符号的直方图构建活动模型。我们通过实验验证了基于BoF的框架在识别九种活动类别方面的有效性，并评估了影响框架性能的六个因素。这些因素包括窗口大小，功能选择，构造运动原语的方法，运动词汇量，运动原语分配的加权方案以及学习机器的内核功能。最后，我们证明，与基于非统计字符串匹配的方法相比，基于统计BoF的框架可以实现更好的性能。

07: A study on human activity recognition using accelerometer data from smartphones

摘要：本文介绍了如何使用用户手机生成的加速度数据来识别某些类型的人类身体活动。我们提出了一种识别系统，其中设计了一个新的数字低通滤波器，以便将原始数据中重力加速度的分量与人体加速度的分量隔离开。该系统在真实条件下与多个人类受试者进行的实验中经过培训和测试。使用各种统计功能对几个分类器进行了测试。考虑了数据的高频和低频成分。我们选择了五个分类器，每个分类器都能很好地识别我们的活动，并研究了如何将它们组合为最佳分类器。我们发现，使用平均概率作为融合方法可以达到91.15％的整体准确率。

08: Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition

摘要：传统上，人类活动识别（HAR）任务是使用启发式过程获得的工程特征来解决的。当前的研究表明，深度卷积神经网络适合于从原始传感器输入中自动提取特征。但是，人类活动是由复杂的运动序列组成的，捕获这种时间动态是成功进行HAR的基础。基于时间序列域的递归神经网络的最新成功，我们提出了基于卷积和LSTM递归单元的活动识别的通用深度框架，该框架：（i）适用于多模式可穿戴传感器；（ii）可以自然地进行传感器融合；（iii）在设计特征时不需要专家知识；（iv）明确建模特征激活的时间动态。我们在两个数据集上评估了我们的框架，其中一个已用于公共活动识别挑战中。我们的结果表明，我们的框架在挑战数据集上的竞争性深层非经常性网络的平均表现为4％；胜过某些先前报告的结果最多达9％。我们的结果表明，该框架可以应用于同质传感器模式，但也可以融合多模式传感器以提高性能。我们描述了关键架构超参数对性能的影响，以提供有关其优化的见解。

09: Large-scale video classification with convolutional neural networks

摘要：卷积神经网络（CNN）已建立为图像识别问题的强大模型类别。受这些结果的鼓舞，我们使用属于487个类别的100万个YouTube视频的新数据集，对CNN在大规模视频分类上进行了广泛的经验评估。我们研究了多种在时域上扩展CNN连通性的方法，以利用本地时空信息，并提出了一种多分辨率，偏心的体系结构作为加速培训的一种有希望的方法。与强大的基于特征的基准（55.3％至63.9％）相比，我们最好的时空网络显示出显着的性能改进，而与单帧模型相比（59.3％至60.9％）仅表现出令人惊讶的适度改进。通过对UCF-101动作识别数据集的顶层进行重新训练，我们进一步研究了最佳模型的泛化性能，并观察到与UCF-101基线模型相比，性能有了显着提高（从43.9％增长到63.3％）。

10: Long-term recurrent convolutional networks for visual recognition and description

摘要：基于深度卷积网络的模型主导了最近的图像解释任务。我们研究了周期性的或“暂时性的”模型对于涉及序列，视觉和其他方面的任务是否有效。我们开发了一种新颖的循环卷积体系结构，适用于端到端可训练的大规模视觉学习，并演示了这些模型在基准视频识别任务，图像描述和检索问题以及视频旁白挑战方面的价值。与当前模型假定固定的时空接受域或简单的时间平均以进行顺序处理相比，循环卷积模型“加倍深”，因为它们可以构成空间和时间“层”。当目标概念复杂和/或训练数据有限时，此类模型可能具有优势。当非线性被合并到网络状态更新中时，学习长期依赖性是可能的。长期的RNN模型之所以吸引人，是因为它们可以直接将可变长度的输入（例如，视频帧）映射到可变长度的输出（例如，自然语言文本），并且可以对复杂的时间动态建模。但是可以通过反向传播对其进行优化。我们的循环长期模型直接连接到现代视觉卷积模型，可以共同训练以同时学习时间动态和卷积感知表示。我们的结果表明，与单独定义和/或优化的最新模型相比，此类模型具有明显的优势。

11: Learning spatiotemporal features with 3D convolutional networks

摘要：我们提出了一种简单有效的时空特征学习方法，该方法使用在大规模监督视频数据集上训练的深度3维卷积网络（3D ConvNets）。我们的发现有三点：1）3D ConvNets与2D ConvNets相比更适合时空特征学习； 2）在所有层中均具有小型3x3x3卷积内核的同类架构是3D ConvNets表现最佳的架构之一，以及3）学习的功能，即C3D（卷积3D），具有简单的线性分类器，在4个不同的基准上均胜过最新技术，并且可与其他2个基准上的当前最佳方法相媲美。此外，这些功能还很紧凑：仅10个维度即可在UCF101数据集上达到52.8％的准确度，并且由于ConvNets的快速推断而非常高效地进行计算。最后，它们在概念上非常简单，易于训练和使用。

12: Temporal activity detection in untrimmed videos with recurrent neural networks

摘要：本文探索了使用卷积神经网络和递归神经网络对视频活动进行分类和时间定位的不同方法，并提出了一种实现方法。第一步，使用最先进的3D卷积神经网络从视频帧中提取特征。通过循环神经网络提供此功能，该神经网络以简单灵活的方式解决活动分类和时间定位任务。为了达到最佳性能和学习所提供的视频数据集，已经测试了不同的体系结构和配置。另外，已经研究了经过训练的网络的输出上的不同类型的后处理，以在视频上的活动的时间本地化上获得更好的结果。本文开发的神经网络提供的结果已提交给CVPR的ActivityNet Challenge 2016，使用简单灵活的体系结构即可获得竞争性结果。

13: SST: Single-stream temporal action proposals

摘要：我们的论文提出了一种在未修剪的长视频序列中对人类行为进行时间检测的新方法。我们介绍了单流时间行动提案（SST），这是一种用于生成时间行动提案的新型有效且高效的深度架构。我们的网络可以在非常长的输入视频序列中以单个流连续运行，而无需将输入划分为短的重叠剪辑或临时窗口进行批处理。我们凭经验证明，我们的模型在生成临时动作建议的任务上优于最新技术，同时实现了文献中最快的处理速度。最后，我们证明了将SST建议与现有的动作分类器结合使用可改善最新的时间动作检测性能。

14: UCF101: A dataset of 101 human actions classes from videos in the wild

摘要：我们介绍了UCF101，它是目前最大的人类行为数据集。它包含101个动作类别，超过13000个剪辑和27个小时的视频数据。该数据库由现实的用户上传的视频组成，其中包含摄像机运动和背景混乱。此外，我们使用标准的词袋方法在此新数据集上提供基线动作识别结果，总体性能为44.5％。据我们所知，由于UCF101具有大量的类，大量的剪辑以及此类剪辑的不受约束的性质，因此它目前是最具挑战性的动作数据集。

15: ActivityNet: A large-scale video benchmark for human activity understanding

摘要：尽管在人类动作识别方面有许多数据集方面的努力，但当前的计算机视觉算法在可识别动作的可变性和复杂性方面仍然受到严重限制。这部分是由于当前基准测试的简单性，该基准测试主要关注手动裁剪视频上发生的简单动作和动作。在本文中，我们介绍ActivityNet，这是一种用于了解人类活动的新型大型视频基准。我们的基准旨在涵盖人们日常生活中感兴趣的各种复杂的人类活动。在其当前版本中，ActivityNet提供203个活动类别的样本，每个类别平均137个未修整视频，每个影片1.41个活动实例，总共849个视频小时。我们说明了三个可以使用ActivityNet来比较人类活动理解算法的场景：未修剪的视频分类，修剪的活动分类和活动检测。

16: Health sensing by wearable sensors and mobile phones: A survey

摘要：随着工业化国家人口老龄化的全球趋势，用于帮助老年人或患者的医疗保健的高效信息通信技术（ICT）引起了极大的研究关注。在这些技术中，可穿戴式传感器和手机对健康状态的感知是重要的基础。它监视实时身体状态；存储或将结果发送给远程家庭成员或医生。这样，它既可以帮助人们更多地关注被忽视的现象，例如危险疾病的线索，也可以帮助人们在紧急情况发生时发出恐慌警报。健康感测中有许多关键问题。首先，传感器必须在不影响人的舒适性和安全性的同时提供良好的精度。同时，由于人为穿戴，必须有效地处理由人体运动引起的大量噪声，以减少误报。最后，不同的健康或死亡信号通常需要不同的传感技术和仪器。为了梳理应对这些挑战和多样性的技术优势，本文对使用人体传感器网络和移动电话进行的健康传感技术的现状进行了调查。它按应用目标对相关作品进行分类，包括：i）跌倒检测，ii）步态分析，iii）活动资格，iv）心态感知和v）睡眠感知。它还对相关的传感系统和算法进行了总结和比较，以揭示每个分区的开发路线。

17: A survey on activity detection and classification using wearable sensors

摘要：活动检测和分类对于自主监控人类的应用（包括辅助生活，康复和监视）非常重要。由于可穿戴式传感器的成本不断下降，易于部署和使用，以及能够提供连续监控的能力，因此与固定位置安装的传感器相反，近年来可穿戴式传感器已得到广泛使用。由于现在许多智能手机都配备了各种传感器，例如加速度计，陀螺仪和照相机，因此开发活动监控算法变得更加可行，该算法使用这些传感器中的一个或多个来提高可访问性。我们使用可穿戴式传感器对活动分类进行完整而全面的调查，涵盖了多种传感方式，包括加速度计，陀螺仪，压力传感器以及基于摄像头和深度的系统。我们讨论了通过这种感应方式来解决的活动类型的差异。例如，加速度计，陀螺仪和磁力计系统具有解决全身运动或整体活动的历史，而相机系统则提供了对局部交互作用或个人与对象的交互作用进行分类所需的环境。我们还发现，这些单一的感应方式为处理全球和本地交互类型活动的混合工作奠定了基础。除了传感器的类型和分类的活动类型之外，我们还提供每个可穿戴系统的详细信息，包括人体传感器的位置，所采用的学习方法以及实验设置的范围。我们进一步讨论了针对不同系统在何处执行处理，即本地处理还是远程处理。这是第一批针对不同的可穿戴传感器系统提供此类覆盖范围以进行活动分类的调查。

18: Pixel-level hand detection in ego-centric videos

摘要：我们在以自我为中心的相机环境中解决了像素级手部检测的任务。在以自我为中心的视频中提取手部区域是了解手部对象操作和分析手眼协调性的关键步骤。但是，与手势检测或手势语言识别之类的手检测的传统应用相比，以自我为中心的视频提出了新的挑战，例如照明的快速变化，显着的摄像机运动以及复杂的手部对象操纵。为了量化在这个新领域中的挑战和性能，我们提出了一个完全标记的室内/室外以自我为中心的手部检测基准数据集，其中包含超过2亿个带标记的像素，其中包含在各种照明条件下拍摄的手部图像。使用我们的数据集和一个公开可用的以自我为中心的室内数据集，我们使用各种局部外观特征对检测性能进行了广泛的分析。我们的分析突出了稀疏特征的有效性以及建模全局照明的重要性。我们根据研究结果提出了一种建模策略，并表明我们的模型优于几种基准方法。

19: Discovering important people and objects for egocentric video summarization

摘要：我们开发了一种总结以自我为中心的视频的方法。我们引入了新颖的以自我为中心的功能，以训练可预测重要区域的回归器。通过使用发现的重要区域，我们的方法比通常包含不相关或多余信息的传统方法产生的信息摘要要多得多。

20: First-person activity recognition: What are they doing to me?

摘要：本文讨论了从第一人称视角识别交互级人类活动的问题。目的是使观察者（例如，机器人或可穿戴式摄像机）能够从连续的视频输入中了解“其他人正在执行的活动”。这些包括友好的互动，例如“一个拥抱观察者的人”，以及敌对的互动，例如“冲撞观察者”或“向观察者投掷物体”，其视频涉及由物理互动引起的大量摄像机自我运动。本文研究了整合全局和局部运动信息的多通道内核，并提出了一种新的活动学习/识别方法，该方法明确考虑了第一人称活动视频中显示的时间结构。在我们的实验中，我们不仅显示了分段视频的分类结果，而且证实了我们的新方法能够可靠地检测连续视频中的活动。

21: Learning to recognize objects in egocentric activities

摘要：本文解决了在极弱的监督下从以家庭为中心的家庭活动视频中学习对象模型的问题。对于每个活动序列，我们仅知道其中存在的对象的名称，而对对象的外观或位置一无所知。我们方法的关键是一种强大的，不受监督的自下而上的分割方法，该方法利用以自我为中心的域的结构将每个帧分为手，对象和背景类别。通过使用多实例学习来匹配序列中的对象实例，我们发现并定位了对象的出现。通过转换精炼对象表示，并训练对象级分类器。我们展示了使用弱监督学习产生的模型检测新颖对象实例的令人鼓舞的结果。

22: Fast unsupervised ego-action learning for first-person sports videos

摘要：用于记录动态第一人称视频素材的便携式高质量运动相机（例如，头戴式或头盔式）正在成为许多运动爱好者中的常见物品。我们处理发现第一人称动作类别（我们称为自我动作）的新任务，这对于诸如视频索引和检索之类的任务可能很有用。为了学习自我动作类别，我们研究了基于运动的直方图和无监督学习算法的使用，以快速地对视频内容进行聚类。我们的方法假设一个完全无人监督的场景，其中没有标记的培训视频，视频没有预先划分，自我行动类别的数量未知。在我们提出的框架中，我们表明可以使用堆叠的Dirichlet过程混合模型来自动学习运动直方图密码本和自我行动类别集。我们通过内部和公开YouTube视频对我们的方法进行了定量评估，并展示了针对多种体育类型的可靠的自我行动分类。对比分析表明，在分类准确度和计算速度方面，我们的方法优于其他最新主题模型。初步结果表明，平均而言，一个10分钟的视频序列的分类内容可以在5秒内编入索引。

23: Detecting activities of daily living in first-person camera views

摘要：我们提出了一个新颖的数据集和新颖的算法来解决在第一人称视角下检测日常生活活动（ADL）的问题。我们已经收集了一个由100万人组成的数据集，其中包含数十个人执行无脚本的日常活动。数据集带有活动，对象轨迹，手的位置和交互事件的注释。ADL与典型动作的不同之处在于，它们可能涉及长时间的时间结构（煮茶可能需要几分钟的时间）和复杂的对象交互作用（冰箱的门打开时看上去会有所不同）。我们开发了新颖的表示形式，包括（1）时间金字塔，当对模型进行评分时，该金字塔将众所周知的空间金字塔归纳为近似的时间对应关系；（2）利用对象在交互时看起来不同的事实的复合对象模型。我们进行了广泛的实证评估，并证明我们的新颖表示法比传统方法产生了两倍的改进。我们的分析表明，现实世界中的ADL识别是“关于对象的所有”，尤其是“关于正在与之交互的所有对象”。

24: Object-centric spatio-temporal pyramids for egocentric activity recognition

摘要：以自我为中心的视频中的活动很大程度上是由与相机佩戴者互动的对象定义的，从而使概述对象的表示形式很有用。除了在单个直方图中简单记录每个对象发生的频率之外，时空分箱方法还可以捕获对象的相对布局和顺序。但是，现有方法使用手工制作的装箱方案（例如，均匀分布的分区金字塔），可能无法捕获能最好地区分某些活动的关系。我们建议学习对一组以自我为中心的活动类别有区别的时空划分。我们设计了一种增强方法，该方法可以在随机的候选分区池中自动选择一小组有用的时空金字塔直方图。为了有效地集中候选分区，我们进一步提出了一种“以对象为中心”的切割方案，该方案更喜欢在那些以自我为中心的活动中突出参与的对象附近取样箱边界。这样，我们就将随机化的分区池专门化为以自我为中心的设置，并提高了增强训练的效率。我们的方法可为识别具有挑战性的日常生活提供最先进的准确性。

25: Learning to predict gaze in egocentric video

摘要：我们提出了一个以自我为中心的视频中凝视预测的模型，该模型利用了摄像头佩戴者行为中存在的隐含线索。具体来说，我们从视频中计算出相机佩戴者的头部运动和手部位置，并将它们组合起来以估计眼睛的位置。我们进一步将凝视的动态行为（尤其是注视点）建模为潜在变量，以改善凝视预测。在公众可获得的以自我为中心的视觉数据集上，我们的凝视预测结果大大优于最新算法。此外，我们证明了通过将注视预测插入最新技术中，可以在识别日常动作和分割前景对象方面获得显着的性能提升。

26: Experiments on an RGB-D wearable vision system for egocentric activity recognition

摘要：这项工作从自我中心的角度描述并探索了实现活动识别的新颖步骤。活动识别是计算机视觉中一个广泛研究的话题，但是可穿戴式视觉系统的独特特征带来了新的挑战和机遇。我们评估了一个具有挑战性的新的公开可用数据集，其中包括跨两个室内环境执行一组20多种不同活动的不同用户的轨迹。研究的视觉特征包括紧凑的全局图像描述符，包括GIST和基于新颖皮肤分割的直方图签名，以及用于识别的最新图像表示，包括SIFT单词袋和基于卷积神经网络（CNN）的特征。我们的实验表明，简单而紧凑的功能可提供合理的准确性来获取基本活动信息（在我们的案例中是操纵与非操纵）。但是，对于更细粒度的类别，基于CNN的功能提供了最有希望的结果。未来的步骤包括将具有这些功能的深度信息和时间一致性集成到管道中。

27: Story-driven summarization for egocentric video

摘要：我们提供了一种视频摘要方法，可以发现以自我为中心的视频的故事。给定较长的输入视频，我们的方法选择一小段视频子镜头，以描述基本事件。受文本分析工作的启发，该分析将新闻报道与时间联系在一起，我们定义了基于随机游走的子镜头之间的影响度量，该度量反映了视觉对象如何促进事件的进展。使用此影响度量，我们定义了最佳k-subs热摘要的目标。传统方法优化了摘要的多样性或代表性，而我们的方法则明确说明了一个子事件是如何“导致”另一个子事件的，而关键的是，该事件捕获了简单对象共现之外的事件连通性。结果，我们的摘要提供了更好的故事感。我们将我们的方法应用于从23个独特的相机佩戴者处拍摄的超过12个小时的每日活动视频，并与34个人类受试者的多个基线相比，系统地评估了其质量。

28: Recognition of activities of daily living with egocentric vision: A review

摘要：在环境辅助生活系统中，正在使用基于视频的日常生活活动（ADL）识别，以支持老年人的独立生活。然而，基于位于环境中的照相机的当前系统存在许多问题，例如遮挡和有限的视野。近来，可穿戴式相机已经开始被利用。本文介绍了一种以自我为中心的视觉系统的最新发展状况，该系统遵循以下层次结构来识别ADL：运动，动作和活动级别，其中每个级别提供更高的语义信息并涉及更长的时间范围。当前以自我为中心的视觉文献表明，ADL的识别主要由场景中存在的对象（尤其是与特定任务相关的对象）驱动。但是，尽管事实证明基于对象的方法很受欢迎，但是由于在不受约束的场景中发现的类内差异，对象识别仍然是一个挑战。结果，当前系统的性能远远不能令人满意。

29: Accelerated visual context classification on a low-power smartwatch

摘要：可穿戴式传感器产生的数据在诸如运动和健身的性能增强和培训帮助，对老年人和慢性病管理的持续监控以及游戏和娱乐等方面至关重要。不幸的是，当前市场上的可穿戴设备要么无法执行复杂的功能，要么由于电池寿命短而严重受损。在这项工作中，我们展示了一个基于超低功耗（ULP）异构系统的智能手表平台，该系统由TI MSP430微控制器，PULP可编程并行加速器和一组ULP传感器（包括照相机）组成。嵌入式PULP加速器使基于卷积神经网络的最新上下文分类能够在低于10mW的系统功率范围内应用。我们的方法能够在五种类别的上下文分类中达到较高的准确性（高达84％，五种以上的三种类别达到90％以上的准确性），同时每种分类消耗2.2 mJ的能量，或者每类消耗小于91μJ的超低能耗分类的准确性比偶然性高64％-3.2×。我们的结果表明，相对于MSP430，拟议的异构平台可以在相似的功率范围内提供高达500倍的加速，这将使复杂的计算机视觉算法能够在功率受限的情况下执行。

30: Multi-scale conditional random fields for first-person activity recognition on elders and disabled patients

摘要：我们提出了一种新颖的普及系统，以识别可穿戴设备中人类的日常活动。该系统以名为“智能眼镜”的老花镜形式设计，集成了3轴加速度计和第一人称视角相机。我们的目的是根据对象的视觉和头部运动数据对他们的日常生活活动（ADL）进行分类。这种自我活动识别系统不仅允许看护者跟踪特定的人（例如残疾患者或老年人），而且还具有提醒/警告具有危险状况认知障碍的人的潜力。我们提出以下贡献：从加速度计和视频中提取特征；一种分类算法，将机车（身体运动）和固定活动（无运动或有小运动）集成在一起；一种新颖的多尺度动态图形模型，用于随时间推移进行结构化分类。在本文中，我们在两个大型数据集上收集，训练和验证我们的系统：20小时的老年人ADL数据集和40小时的患者ADL数据集，分别包含12和14种不同的活动。结果表明，与传统的分类方法相比，我们的方法有效地提高了系统性能（F-Measure），平均提高了20％–40％，最高达到84.45％，而老年人的总体准确性为90.04％。此外，我们还对30名不同残疾的患者验证了我们的方法，总体准确率高达77.07％。

31: Situation awareness via sensor-equipped eyeglasses

摘要：新兴的智能手机技术结合了头戴式显示器（HMD）和标准功能，例如接听电话，电子邮件和帮助导航。这为探索网络机器人算法（机器人传感器和人类动力装置）提供了新的机会。为了使这些设备更适应环境条件，用户行为和用户喜好，重要的是要让配备传感器的设备有效地适应并响应用户的活动（例如，禁用电梯中的来电，激活视频记录开车时）。因此，本文提出了一种用于头戴式智能手机的态势感知系统（SAS）。从惯性传感器（加速度计，陀螺仪）和视频数据（相机）收集数据之后，SAS分三个步骤执行活动分类。步骤1将惯性传感器数据转换为与头部方向无关的稳定归一化坐标系。步骤2提取关键特征（统计，物理，GIST）。步骤3对活动进行分类（朴素贝叶斯分类器），对环境进行区分（支持向量机），最后将两个结果进行合并（隐马尔可夫模型）以进行进一步改进。SAS已在配备传感器的眼镜原型上实施，并且在区分20个实际活动时达到了高精度（81.5％）。

32: Temporal segmentation and activity classification from first-person sensing

摘要：将人类运动分解为动作的时间分段对于理解和构建人类动作和活动识别的计算模型至关重要。几个问题加剧了时间分割和人体运动分类的挑战。这些包括人类动作的时间尺度和周期性的巨大变化，表示关节运动的复杂性以及所有可能的运动组合的指数性质。我们通过研究两个不同的问题提供了初步的结果-对正在执行的总体任务进行分类，以及随着时间的推移将各个框架分类为特定操作这一更困难的问题。我们探索通过可穿戴式摄像机和惯性测量单元（IMU）进行的第一人称感知，以在自然环境中将人的运动暂时细分为动作并在烹饪和配方制备的情况下进行活动分类。我们介绍了有监督和无监督的时间分割的基线结果，以及CMU-多式联运活动数据库（CMU-MMAC）中的配方识别。

33: Human activity classification from wearable devices with cameras

摘要：有许多方法可以根据加速度传感器或安装在环境中的摄像机进行人类活动分类。使用自我中心视频的工作相对较少。纯加速度计系统尽管计算效率高，但是可以检测到的活动的多样性和复杂性受到限制。例如，我们可以使用加速度计数据检测到就坐事件，但无法确定用户是否坐在椅子或沙发上，或者用户所处的环境类型。为了检测具有更多细节和上下文的活动，我们提出了一种同时使用从智能手机获得的加速度计和自我视觉数据的稳健且自主的方法。多类支持向量机（SVM）用于通过使用加速度计数据和光流向量对活动进行分类。使用基于聚合通道特征的检测器从摄像机数据中检测场景中的对象。另一个多类SVM用于检测接近的不同对象。然后，采用隐马尔可夫模型（HMM）来检测更复杂的活动。已经对对象进行了坐在椅子上，坐在沙发上和走过门口的活动的实验。所提出的方法实现了95％和89％的整体精度和召回率。

34: Guide to the carnegie mellon University multimodal activity (CMU-MMAC) database

摘要：本文档总结了CMU多模态活动数据库（CMU-MMAC）的技术，过程和数据库组织。CMU-MMAC数据库包含对执行烹饪和食物准备任务的受试者的人类活动的多模式测量。CMU-MMAC数据库是在卡耐基梅隆大学的运动捕捉实验室中收集的。建造了一个厨房，迄今为止，已经记录了五个对象烹饪五个不同的食谱：核仁巧克力饼，比萨饼，三明治，沙拉和炒鸡蛋。记录了以下模式：•视频：（1）三个具有低时间分辨率（30 Hertz）的高空间分辨率（1024×768）彩色摄像机。（2）两个具有高时间分辨率（60 Hertz）的低空间分辨率（640×480）彩色摄像机。（3）一台可穿戴式低空间分辨率（640×480）的低时间分辨率（12 Hertz）相机。•音频：（1）五个平衡麦克风。（2）穿戴式手表。•运动捕捉：具有12个红外MX-40摄像机的Vicon运动捕捉系统。每个摄像机都以120赫兹记录4兆像素分辨率的图像。•五个3轴加速度计和陀螺仪。使用了几台计算机来记录各种模式。使用网络时间协议（NTP）同步了计算机。

35: Dynamic routing between capsules

摘要：胶囊是一组神经元，其活动矢量代表特定类型的实体（例如对象或对象部分）的实例化参数。我们使用活动矢量的长度来表示实体存在的概率，并使用其方向来表示实例化参数。一级的活动胶囊通过转换矩阵对高层胶囊的实例化参数进行预测。当多个预测结果一致时，更高级别的胶囊就会生效。我们表明，经过判别训练的多层胶囊系统在MNIST上达到了最先进的性能，并且在识别高度重叠的数字方面比卷积网络要好得多。为了获得这些结果，我们使用了一种按协议迭代路由的机制：低级胶囊倾向于将其输出发送给活动向量具有大标量积的高级胶囊，而预测来自低级胶囊。

36: Very deep convolutional networks for large-scale image recognition

摘要：在这项工作中，我们研究了卷积网络深度对其在大规模图像识别环境中的准确性的影响。我们的主要贡献是使用具有非常小（3x3）卷积滤波器的体系结构对增加深度的网络进行全面评估，这表明通过将深度推至16-19权重层可以实现对现有技术配置的重大改进。这些发现是我们提交2014年ImageNet挑战赛的基础，我们的团队分别在本地化和分类轨道中获得了第一名和第二名。我们还表明，我们的表示法可以很好地推广到其他数据集，在这些数据集中它们可以达到最新的结果。我们已经公开提供了两个性能最佳的ConvNet模型，以促进对在计算机视觉中使用深层视觉表示的进一步研究。

37: Visualizing and understanding convolutional networks

摘要：大型卷积网络模型最近在ImageNet基准Krizhevsky等人身上展示了令人印象深刻的分类性能。[18]。但是，对于它们为什么表现如此出色或如何进行改进尚无明确的了解。在本文中，我们将探讨这两个问题。我们介绍了一种新颖的可视化技术，可深入了解中间要素层的功能以及分类器的操作。在诊断角色中使用时，这些可视化使我们能够找到在ImageNet分类基准上胜过Krizhevsky等人的模型架构。我们还进行了消融研究，以发现不同模型层对性能的贡献。我们展示了ImageNet模型可以很好地推广到其他数据集：重新训练softmax分类器时，它令人信服地击败了Caltech-101和Caltech-256数据集上的最新结果。

38: Network in network

摘要：我们提出了一种新型的深层网络结构，称为“网络中的网络”（NIN），以增强接受域内局部斑块的模型可分辨性。传统的卷积层使用线性滤波器，然后使用非线性激活函数来扫描输入。取而代之的是，我们构建具有更复杂结构的微神经网络来抽象接受域中的数据。我们用多层感知器实例化微神经网络，该感知器是一种有效的函数逼近器。通过以与CNN类似的方式在输入上滑动微网络来获得特征图。然后将它们送入下一层。深度NIN可以通过堆叠上述结构的多个实现。通过微网络增强的局部建模，我们能够在分类层的特征图上利用全局平均池，与传统的全连接层相比，它更易于解释且不太容易拟合。我们展示了NIF在CIFAR-10和CIFAR-100上的最新分类性能，以及在SVHN和MNIST数据集上的合理性能。

39: Going deeper with convolutions

摘要：我们提出了一种代号为Inception的深度卷积神经网络体系结构，该体系结构在ImageNet大规模视觉识别挑战赛2014（ILSVRC14）中实现了分类和检测的最新技术水平。该体系结构的主要特点是网络内部计算资源的利用率得到提高。通过精心设计，我们在保持计算预算不变的情况下增加了网络的深度和宽度。为了优化质量，架构决策基于Hebbian原则和多尺度处理的直觉。在我们提交的ILSVRC14中使用的一种特定化身叫GoogLeNet，它是一个22层深的网络，其质量在分类和检测的背景下进行评估。

40: Rethinking the inception architecture for computer vision

摘要：卷积网络是用于各种任务的大多数最新计算机视觉解决方案的核心。自2014年以来，非常深的卷积网络开始成为主流，并在各种基准中获得了可观的收益。尽管增加的模型大小和计算成本往往会转化为大多数任务的即时质量提升（只要提供足够的标记数据以进行培训），但计算效率和低参数数量仍然是各种用例（例如移动视觉和大型应用）的促成因素。 -数据方案。在这里，我们正在探索以适当的因式分解卷积和积极的正则化为目标的网络扩展方式，旨在尽可能有效地利用增加的计算量。我们在ILSVRC 2012分类挑战验证集上对我们的方法进行了基准测试，结果表明，与使用最新技术相比，使用网络的单帧评估具有21：2％的top-1和5：6％的top-5误差。每个推论使用50亿个乘法，并且使用少于2500万个参数。通过4种模型的集成和多次作物评估，我们报告了验证集中的3：5％top-5错误和17：3％top-1错误，以及官方测试集中的3：6％top-5错误。

41: Deep residual learning for image recognition

摘要：更深的神经网络更难训练。我们提出了一种残差的学习框架，以简化比以前使用的网络更深入的网络训练。我们显式地将层重新配置为参考层输入学习剩余函数，而不是学习未参考函数。我们提供了全面的经验证据，表明这些残差网络更易于优化，并且可以通过深度的增加而获得准确性。在ImageNet数据集上，我们评估深度最大为152层的残差网络-比VGG网络[40]深8倍，但复杂度仍然较低。这些残留网络的整体在ImageNet测试仪上实现了3.57％的误差。该结果在ILSVRC 2015分类任务中获得第一名。我们还将介绍具有100和1000层的CIFAR-10的分析。表示的深度对于许多视觉识别任务至关重要。仅由于我们的深度表示，我们在COCO对象检测数据集上获得了28％的相对改进。深度残差网是我们向ILSVRC和COCO 2015竞赛1提交的基础，在这些竞赛中，我们还获得了ImageNet检测，ImageNet本地化，COCO检测和COCO分割等任务的第一名。

42: Faster R-CNN: Towards real-time object detection with region proposal networks

摘要：先进的物体检测网络依靠区域提议算法来假设物体的位置。SPPnet和Fast R-CNN等技术的进步减少了这些检测网络的运行时间，使区域提议计算成为瓶颈。在这项工作中，我们介绍了一个区域提议网络（RPN），该区域提议网络与检测网络共享全图像卷积特征，从而实现几乎免费的区域提议。RPN是一种全卷积网络，可同时预测每个位置的对象范围和对象得分。对RPN进行了端到端的培训，以生成高质量的区域建议，Fast R-CNN将其用于检测。通过简单的交替优化，可以训练RPN和Fast R-CNN共享卷积特征。对于非常深的VGG-16模型，我们的检测系统在GPU上具有5fps的帧速率（包括所有步骤），同时在PASCAL VOC 2007（73.2％mAP）和2012上实现了最新的对象检测精度（70.4％mAP），每个图片使用300个投标。可以从https://github.com/ShaoqingRen/faster\_rcnn获得代码。

43: Human activity classification incorporating egocentric video and inertial measurement unit data

摘要：已经提出了许多用于人类活动分类的方法，这些方法要么依赖于惯性测量单位（IMU）数据，要么来自观看对象的静态相机的数据。使用以自我为中心的视频的工作相对较少，而以自我为中心的视频和IMU数据相结合的方法则更少。仅依赖IMU数据的系统在其可检测活动的复杂性方面受到限制。在本文中，我们针对细粒度的活动分类提出了一种健壮且自主的方法，该方法利用来自多个可穿戴传感器模态的数据来区分本质上相似的活动，并且每个传感器都无法达到精确度单独。我们在身体上同时使用以自我为中心的视频和IMU传感器。我们将胶囊网络与卷积长期短期记忆（LSTM）结合使用来分析以自我为中心的视频，并使用LSTM框架来分析IMU数据并捕获动作的时间方面。我们对CMU-MMAC数据集进行了实验，总体召回率和准确率分别为85.8％和86.2％。我们还介绍了单独使用每种传感器形式的结果，这些结果表明，与仅使用自我视觉数据和仅使用IMU数据相比，所提出的方法在准确性上提高了19.47％和39.34％。

44: Deep learning for precipitation nowcasting: A benchmark and A new model

摘要：为了对区域降雨量进行高分辨率预报，降水临近预报已成为从暴雨警告到飞行安全的各种公共服务的重要基础技术。最近，卷积LSTM（ConvLSTM）模型已显示出优于传统的基于光流的降水临近预报方法，这表明深度学习模型具有解决该问题的巨大潜力。然而，基于ConvLSTM的模型中的卷积递归结构是位置不变的，而自然运动和变换（例如，旋转）通常是位置不变的。此外，由于基于深度学习的降水临近预报是一个新兴领域，因此尚未建立明确的评估协议。为了解决这些问题，我们提出了一种新模型和降水临近预报的基准。具体而言，我们超越了ConvLSTM，并提出了可以积极学习循环连接位置可变结构的Trajectory GRU（TrajGRU）模型。此外，我们提供了一个基准，其中包括来自香港天文台的现实世界中的大规模数据集，新的训练损失以及全面的评估协议，以方便将来的研究并评估最新技术水平。

45: Action recognition in the presence of one egocentric and multiple static cameras

摘要：在本文中，我们研究存在单个以自我为中心的相机和多个静态相机的情况下识别人类动作的问题。在静态摄像机中可以更好地呈现某些动作，在静态摄像机中可以看到演员的整个身体和动作的上下文。在以自我为中心的相机中，可以更好地识别其他一些动作，在这些相机中，可以看到手的细微运动和复杂的对象交互作用。在本文中，我们介绍了一个模型，该模型可以通过学习预测每个摄像机在识别每个帧中的动作中的重要性来从两全其美中受益。通过对潜在摄像机重要性变量和动作分类器的联合判别学习，我们的模型在具有挑战性的CMU-MMAC数据集中获得了成功的结果。我们的实验结果表明，根据预期的重要性，学习相机的使用将大大提高。所学习的潜在变量可提供对场景的理解程度，该场景可通过在相机之间进行平滑切换来实现自动摄影，以使每帧中的相关信息量最大化。

46: Very deep convolutional networks for large-scale image recognition

摘要：在这项工作中，我们研究了卷积网络深度对其在大规模图像识别环境中的准确性的影响。我们的主要贡献是使用具有非常小（3x3）卷积滤波器的体系结构对增加深度的网络进行全面评估，这表明通过将深度推至16-19权重层可以实现对现有技术配置的重大改进。这些发现是我们提交2014年ImageNet挑战赛的基础，我们的团队分别在本地化和分类轨道中获得了第一名和第二名。我们还表明，我们的表示法可以很好地推广到其他数据集，在这些数据集中它们可以达到最新的结果。我们已经公开提供了两个性能最佳的ConvNet模型，以促进对在计算机视觉中使用深层视觉表示的进一步研究。