一年来 deep learning 与 video 融合论文 347 篇

2018.11.18

1. 伟大的生物和 smal：从视频中恢复动物的形状和运动

作 者 :benjamin biggs, thomas roddick, andrew fitfitgibbon, roberto cipolla

摘要: 我们提出了一个系统，以恢复 3d 形状和运动的各种四足动物从**视频**。该系统包括预测候选二维关节位置的机器**学习**前端、查找运动学上合理的关节对应的离散优化以及适合详细三维模型的能量最小化阶段。图像。为了克服动物运动捕捉训练数据有限的问题，以及生成逼真的综合训练图像的难度，该系统旨在处理轮廓数据。联合候选预测器接受综合生成的轮廓图像的训练，并在测试时使用**深度学习**方法或标准**视频**分割工具从真实数据中提取轮廓。该系统在来自多个物种**的动物视频**上进行了测试，并显示了对三维形状和姿势的精确重建。少

2018 年 11 月 14 日提交;**最初宣布** 2018 年 11 月。

2. loans：使用定位器评估器网络进行弱监控的对象检测

作　者：christian bartz, haojin yang,joseph betge, christoph meinel

文摘: 近年来,**深**部神经网络在目标检测和识别任务中取得了显著的成绩。成功的原因主要在于大规模、完全注释化的数据集的可用性，但创建这样一个数据集是一项复杂且成本高昂的任务。本文提出了一种新的弱监督对象检测方法，简化了采集对象检测器的数据收集过程。我们训练了两个模特的组合，他们以学生和老师的方式一起工作。我们的学生（本地化程序）是**学习**本地化对象的模型，教师（评估员）评估本地化的质量，并向学生提供反馈。学生使用此反馈来**学习**如何本地化对象，因此完全由教师监督，因为我们没有使用标签来训练本地化程序。在我们的实验中，我们证明了我们的模型是非常强大的噪音，并达到了竞争性能相比，最先进的完全监督的方法。我们还展示了基于一些**视频**(例如从 youtube 下载）和人工生成的数据创建新数据集的简单性。少

2018 年 11 月 15 日提交;**v1** 于 2018 年 11 月 14 日提交;**最初宣布** 2018 年 11 月。

3. 面向对象的自主驾驶政策

作者:王德泉,科琳·德文,蔡启志,于费舍尔,特雷弗·达雷尔

摘要: 虽然以端到端的方式**学习**视觉运动技能很吸引人,**但深度**神经网络往往是无法解释的，并以惊人的方式失败。对于机器人任务 (如自动驾驶)，显式表示对象的模型可能对新场景更加可靠，并提供直观的可视化效果。我们描述了一个以对象为中心的模型的分类，它利用了对象实例和端**到端学习**。在大盗汽车 v 模拟器中，我们展示了以对象为中心的模型在与其他车辆和行人的场景中优于对象无关的方法，即使使用不完善的探测器也是如此。我们还通过在伯克利深度驱动视频数据集**上**进行评估，证明我们的架构在现实世界环境中表现良好。少

2018 年 11 月 13 日提交;**最初宣布** 2018 年 11 月。

4. 背景减法的深层神经网络概念：系统回顾与比较评价

作 者 :thierry bouwmans, sajid javed, maryam sultana, soon ki jung

摘要: 传统的神经网络显示了一个强大的框架背景减法的视频获得的静态相机。事实上，众所周知的基于神经网络的 sobs 方法及其变体是长期大规模 cdnet 2012

数据集上的领先方法。近年来，属于**深度学习**方法的卷积神经网络被用于背景初始化、前景检测和**深度学习**特征等方面取得了成功。目前，cdnet 2014 中的顶级电流背景减法是基于**深度**神经网络，与基于多特征或多线索策略的传统无监督方法相比，性能差距较大。此外，自 2016 年以来，布拉汉姆和范德罗根布鲁克在美国有线电视新闻网上发表了他们的第一部作品，适用于背景减法，定期提高业绩。在此背景下，我们首次回顾了新手和专家背景减法中的**深层**神经网络概念，以便分析这一成功并提供进一步的方向。为此，我们首先考察了使用背景初始化、背景减法和**深入学习**的功能的方法。然后，讨论了**深**部神经网络在背景减法中的充分性。最后，在 cdnet 2014 数据集上给出了实验结果。少

2018 年 11 月 13 日提交;**最初宣布** 2018 年 11 月。

5. 广告：一个深度网络，用于检测广告

作　　者：murhaf hossari, soumyabrata dev, matthewnicolson, killian mccabe, atul Nautiyal, clare conran, jian tang, wei xu, françois pitié

摘要: 在线**视频**广告使内容提供商能够提供引人注目的内容，接触越来越多的受众，并从在线媒体中获得额外收入。最近，广告策略旨在在**视频**帧中查找原始广告，并将其替换为新广告。这些策略，俗称产品投放或嵌入式营销，极大地帮助营销机构接触到更广泛的受众。然而，在现有文献中，这种检测候选帧的视频序列中的广告集成的目的，是手动完成的。在本文中，我们提出了一个深度**学习**架构，称为 adnet，自动检测**广告在视频**帧中的存在。我们的方法是首款自动检测**视频**帧中广告存在情况的此类方法，并在公共数据集上实现最先进的结果。少

2018 年 11 月 9 日提交;**最初宣布** 2018 年 11 月。

6. 利用深贝叶斯网络进行手术工作流分析的主动学习

作者:sebastian bodenstedt, dominik rivoir,亚历山大·詹克, martin wagner, sören torge mees, jürgen weitz , stefanie speidel

摘要: 对于计算机辅助手术领域的许多应用，例如提供肿瘤的位置，指定外科医生接下来所需的最可能的工具，或确定手术的剩余持续时间，手术工作流程分析的方法是前提。通常，基于机器**学习**的方法可作为外科工

作流分析的基础。在一般的机器**学习**算法，如卷积神经网络 (cnn)，需要大量的标记数据。虽然数据通常很多，但外科工作流分析中的许多任务都需要领域专家的注释数据，因此很难获得足够数量的注释。利用主动**学习**训练机器**学习**模型的目的是为了减少注释工作。主动**学习**方法根据预测不确定性等一些度量值确定哪些未标记的数据点提供的信息最多。然后，专家们将被要求只对这些数据点进行注释。然后使用新数据对模型进行再培训，并用于选择更多数据进行注释。近年来，利用**深**贝叶斯网络 (dbn) 将主动**学习**应用于 cnn。这些网络使得有可能为预测分配不确定性。本文提出了一种基于 dbn 的主动**学习**方法，适用于基于图像的外科工作流分析任务。此外，通过使用经常性体系结构，我们将该网络扩展到基于**视频**的外科工作流分析。我们通过执行仪器存在检测和手术相位分割来评估这些方法在 cholec80 数据集上的应用。在这里，我们能够证明，使用基于 dbn 的主动**学习**方法来选择下一步要注释的数据点，优于基于随机选择数据点的基线。少

2018 年 11 月 8 日提交;**最初宣布** 2018 年 11 月。

7. 基于变分推理的贝叶斯活动识别

作 者 :ranganath krishnan, mahesh subedar, omesh tikoo

文摘: **深**部神经网络中的不确定性估计对于设计可靠、可靠的人工智能系统至关重要。**用于**识别可疑活动的视频监控等应用程序是使用**深度**神经网络 (dnn) 设计的，但 dnn 不提供不确定性估计。捕获安全和安保关键应用程序中可靠的不确定性估计将有助于建立对 ai 系统的信任。我们的贡献是将贝叶斯**深度学习**框架应用于视觉活动识别应用，并以原则性信心量化模型不确定性。利用变分推理技术训练贝叶斯 dnn 来推断模型参数周围的近似后验分布，并对模型参数的后验进行蒙特卡罗采样，以获得预测分布。结果表明，与传统的 dnn 相比，应用于 dnn 的贝叶斯推理为视觉活动识别任务提供了可靠的置信度。我们还表明，与非贝叶斯基线相比，该方法提高了视觉活动识别精确召回分数 6%。我们通过选择分布内和分布**外视频**样本的子集，在时间点 (mitt) 活动识别数据集上评估模型。少

2018 年 11 月 8 日提交;**最初宣布** 2018 年 11 月。

8. 基于强化学习的视觉跟踪相关滤波器选择

作者:谢延春,肖继民,黄开珠 , jeyarajan thiyagalingam, 姚赵

文摘: 相关滤波器已被证明是视觉跟踪方法的有效工具,特别是在跟踪准确性和速度之间寻求良好平衡的方法。但是，基于相关筛选器的模型很容易受到不准确跟踪结果导致的错误更新的影响。到目前为止，在处理相关筛选器更新问题方面投入的精力很少。本文提出了一种解决相关滤波器更新问题的新方法。在我们的方法中,我们并行更新和维护多个相关滤波器模型，并使用**深层**强化**学习**来选择它们之间的最优相关滤波器模型。为了有效地促进决策过程，我们提出了一个决策网络来处理目标外观建模，它通过数百个具有挑战性的**视频**进行训练，使用近端策略优化和轻量级**学习**网络中。对 otb100 和 otb2013 基准的拟议方法进行的详尽评估表明，该方法足够有效，可以实现平均成功率 62.3 和平均精度分数 81.2，这两种方法都超过了传统的性能。基于相关滤波器的跟踪器。少

2018 年 11 月 7 日提交;**最初宣布** 2018 年 11 月。

9. 利用随机标签记忆进行无监督的预培训

作者:vinaychandran pondenkandath, micree alberti, sammer puran, rof ingold, marcus liwicki

摘要: 我们提出了一种新的方法，通过在随机标记的数据集上预培训最先进的**深度**神经网络来利用大型未标记的数据集。具体来说，我们训练神经网络来记住数据集中所有样本的任意标签，并将这些预先训练的网络作为定期**监督学习的**起点。我们的假设是，网络在随机标签培训过程中**学习**的 "记忆基础设施" 也被证明对传统的监督**学习**是有益的。我们通过比较具有和不具有随机标签预训练的同一网络的结果，测试我们在多个**视频**操作识别数据集 (hmdb51、ucf101、动力学) 上的预训练的有效性。我们的方法在分类精度上有一个改进--从 ucfc-101 的 1.5% 到动力学的 5% 不等，这就需要在这方面进行进一步的研究。少

2018 年 11 月 5 日提交;**最初宣布** 2018 年 11 月。

10. stnet: 用于行动识别的局部和全局时空建模

作者:何东亮,周志超,庄干, 富力,小刘,李延东, 王丽敏, 文世雷

摘要: 尽管对静态图像理解的**深度学习**取得了成功，但视频中时空建模最有效的网络架构是什么，目前还不

清楚。本文与现有的 cnn + rnn 或纯基于三维卷积的方法相比，探索了一种适用于**视频**局部和全局时空建模的新的空间时间网络 (stnet) 体系结构。特别是, stnet 将 n 个连续**的视频**帧堆叠到一个具有 3n 通道的 \ 指超级图像} 中，并在超级图像上应用二维卷积来捕获局部时空关系。为了建立全局时空关系模型，我们在局部时空特征图上应用了时间卷积。具体而言，在 stnet 中提出了一种新的时间 Xception 块。它在**视频**的特征序列上采用了一个单独的通道式和时间上的卷积。在动力学数据集上进行的大量实验表明，我们的框架在动作识别方面优于几种最先进的方法，可以在识别精度和模型复杂性之间达成令人满意的权衡。我们进一步演示了在 ucf101 数据集上倾斜**的视频**表示的泛化性能。少

2018 年 11 月 6 日提交;**v1** 于 2018 年 11 月 5 日提交;**最初宣布** 2018 年 11 月。

11. 水下鱼类深度学习在水能应用中的检测

作者:徐文伟,沙里·马茨纳

摘要: 随着潮汐和安装涡轮机等新技术的发展，海洋和河流的清洁能源正在成为现实，这些涡轮机通过自然流动的水发电。目前正在利用水下**视频**监测这些新技术

对鱼类和其他野生动物的影响。需要水下**视频**自动分析的方法，以降低分析成本，提高准确性。一个**深度学习**模型，yolo，被训练识别水下视频中的鱼使用三个非常不同的数据集记录在现实世界**的**水源位。通过对所有三个数据集的示例进行训练和测试，平均精度（map）得分为 0.5392。为了测试模型能够很好地推广到新数据集，仅使用两个数据集的示例对模型进行了训练，然后对所有三个数据集的示例进行了测试。生成的模型无法识别数据集中不属于训练集的鱼。包含在训练集中的其他两个数据集的 map 分数高于在所有三个数据集中接受培训的模型所取得的分数。这些结果表明，需要不同的方法来产生一个训练有素的模型，该模型可以推广到新的数据集，例如在现实世界中遇到的数据集。少

2018 年 11 月 4 日提交;**最初宣布** 2018 年 11 月。

12. 基于在线软挖掘的深度计量学习及注意事项

作者:王新绍,杨华,艾莉亚尔·科迪罗夫，胡国生,尼尔·罗伯逊

摘要: **深度度量学习**的目的是**学习**一种**深度**嵌入，它可以捕获数据点的语义相似性。考虑到大量训练样本的可用性，已知深度**度量学习**由于大量的琐碎样本而具有

缓慢的收敛性。因此，大多数现有的方法通常采用采样挖掘策略来选择重要的样本，以加快收敛速度并提高性能。在这项工作中，我们确定了样本挖掘方法的两个关键限制，并为它们提供了解决方案。首先，以前的挖掘方法为每个样本分配一个二进制分数，即丢弃或保留它，因此它们只在小型批次中选择相关样本的子集。因此，我们提出了一种新的样品挖掘方法，称为在线软挖掘 (osm)，它为每个样本分配一个连续的分数，以利用微型批次中的所有样本。osm **学习**扩展流形，通过关注更多相似的阳性来保持有用的内部差异。其次，现有的方法很容易受到异常值的影响，因为它们一般都包括在挖掘的子集中。为了解决这个问题，我们引入了 "类感知注意" (caa)，它很少关注异常数据样本。此外，通过 osm 和 caa 的结合，提出了一种新的加权对比损失，以**学习**判别嵌入。在两个细粒度视觉分类数据集和两个**基于视频**的人员重新识别基准上进行的大量实验表明，我们的方法明显优于最先进的方法。少

2018 年 11 月 4 日提交;**最初宣布** 2018 年 11 月。

13. 纹理图像检索中的纹理合成引导深部哈希

作者:ayan kumar bhunia, perla sai raj kishore, pranay mukherjee,abhirup das, partha pratim roy

文摘: 随着互联网上图像和**视频**的大规模爆炸，开发了高效的哈希方法，以方便对类似图像的记忆和时间高效检索。然而，现有的作品都没有使用哈希来解决纹理图像检索问题，这主要是因为缺乏足够大的纹理图像数据库。我们的工作解决了这个问题，开发了一个新的**深度学习**架构，生成输入纹理图像的二进制哈希代码。为此，我们首先预训练纹理合成网络 (tsn)，该网络以纹理补丁作为输入，并通过注入较新的纹理内容输出纹理的放大视图。因此，它表示 tsn 对其中间层中所学到的纹理特定信息进行编码。在下一阶段，第二个网络使用通道级的关注收集 tsn 中间层的多尺度特征表示，并以渐进的方式将其组合到密集的连续表示中，最终转换为二进制哈希代码的个人和对标签信息的帮助。新的扩大纹理补丁也有助于数据扩充，以缓解纹理数据不足的问题，并用于培训网络的第二阶段。在三个公共纹理图像检索数据集上的实验表明，与目前最先进的纹理合成方法相比，我们的纹理合成引导哈希方法具有优越性。少

2018 年 11 月 7 日提交;**v1** 于 2018 年 11 月 4 日提交;**最初宣布** 2018 年 11 月。

14. 通过检测面部扭曲的伪影来曝光深度伪造视频

作者:李月尊,刘思伟

文摘: 在这项工作中，我们描述了一种新的**深度学习**为基础的方法，可以有效地区分 ai 生成的假**视频**(以下简称 deepfake **视频**) 和真实**的视频**。我们的方法是基于目前的 deepfaake 算法只能生成分辨率有限的图像的观察，这些图像需要进一步扭曲，才能与源**视频**中的原始面相匹配。这样的变换在产生的 deep 码视频中留下了独特的伪影，我们表明它们可以被卷积神经网络有效地捕获。我们的方法是在一组 deepfake **视频**上进行评估，以确定其在实践中的有效性。少

2018 年 11 月 1 日提交;**最初宣布** 2018 年 11 月。

15. 用于多目标跟踪的深度关联网络

作者:孙世杰,纳维德·阿赫塔尔,欢生宋,阿杰马尔·米安,穆巴拉克·沙阿

文摘: 多目标跟踪 (mot) 在解决计算机视觉视频分析中的许多基本问题方面发挥着重要作用。大多数 mot 方法采用两个步骤: 对象检测和数据关联。第一步检测**视频**每个帧中感兴趣的对象，第二步建立不同帧中检测到的对象之间的对应关系，以获取其轨迹。近年来，由于**深入学习**，目标检测取得了巨大的进步。但是，用

于跟踪的数据关联仍然依赖于手工制作的约束，如外观、运动、空间接近度、分组等，以计算不同帧中对象之间的相关性。本文利用深度学习的力量进行数据关联跟踪，共同建模对象外观及其在不同帧之间的端到端**方式**的相关性。提出的**深**亲和力网络 (dan)**学习**紧凑型;但在多个抽象级别上检测到的对象的全面功能，并在任意两个帧中对这些特征执行详尽的配对排列，以推断对象的相关性。dan 还考虑了**视频**帧之间出现和消失的多个对象。我们利用生成的高效关联计算将当前帧中的对象与前面的帧深入到前面的帧中进行关联，以便进行可靠的在线跟踪。我们的技术是在流行的多目标跟踪挑战 mot15, mot17 和 ua-detrac 进行评估。12 个评估指标下的全面基准测试表明，我们的方法是领导董事会中应对这些挑战的最佳技术之一。我们工作的开源实现可在 https://github.com/shijieS/SST.git 上查阅。少

2018 年 10 月 28 日提交;**最初宣布** 2018 年 10 月。

16. rgb-db 深度人员重构的一种交叉模态蒸馏网络

作 者 :frank hafner, amran bhuiyan, julian f. p.kooij, eric 格兰杰

摘要: 人的重新识别涉及随着时间的推移识别使用多个分布式传感器捕获的个人。随着能够**学习**视觉识别的判别表示的强大**深度学习**方法的出现，基于不同传感器模式的跨模式人重新识别在许多方面变得可行在自动驾驶、机器人和**视频**监控等领域具有挑战性的应用。虽然提出了一些在红外和 rgb 图像之间重新识别的方法，但地址深度和 rgb 图像很少。除了与遮挡、杂乱、错位以及姿势和照明的变化相关的每种模式都面临的挑战外，由于来自 rgb 和深度图像的数据是异构的，因此在不同模式之间也发生了相当大的变化。本文提出了一种新的多模态蒸馏网络，用于 rgb 和深度传感器之间的鲁棒人重新识别。该方法采用两步优化过程，在模式之间进行监督，从 rgb 和深度模式中提取相似的结构特征，从而产生对公共特征空间的判别映射。我们的实验研究了嵌入空间维数的影响，比较了从深度到 rgb 的转移学习，反之亦然，并与其他最先进的交叉模态重新识别方法进行了比较。利用 biwi 和 RobotPKU 数据集获得的结果表明，该方法能够成功地将描述性结构特征从深度模式转移到 rgb 模式。它可以显著优于最先进的传统方法和深度神经网络，用于 rgb 和**深度**之间的交叉模态传感，不会对计算复杂度产生影响。少

2018 年 11 月 5 日提交;**v1** 于 2018 年 10 月 27 日提交;**最初宣布** 2018 年 10 月。

17. a 个 2-网：双注意力网络

作者：陈云鹏，yannis kalantidis，jishuli，shu 现 chen yan，jiashi feng

摘要：**学习**捕捉远程关系是图像/**视频**识别的基础。现有的有线电视新闻网模型一般依靠增加深度来模拟这种效率很低的关系。在这项工作中，我们提出了 "双重注意力块"，这是一个新颖的组件，它从输入图像/**视频**的整个时空空间聚合和传播信息全局特征，从而使后续的卷积层能够访问从整个空间的功能效率。该组件采用两个步骤的双重关注机制进行设计，第一步通过二阶注意力池将整个空间的要素收集到一个紧凑的集合中，第二步自适应地选择特征并将其分配给每个步骤通过另一个关注的位置。所提出的双注意块易于采用，可以方便地插入现有**的深**部神经网络。我们对图像和**视频**识别任务进行了广泛的烧蚀研究和实验，以评估其性能。在图像识别任务中，配备了双注意力块的 resnet-50 在 imagete-1k 数据集上的 resnet-152 体系结构的性能优于更大的 resnet-152 体系结构，参数数量减少了 40% 以上，flop 更少。在动作识别任务上，我们提出的

模型在动力学和 ucf-101 数据集上获得了最先进的结果，其效率明显高于近期的工作。少

2018 年 10 月 26 日提交;**最初宣布** 2018 年 10 月。

18. 深卷神经网络在视频跟踪质量评价中的应用

作者:roger gomez nieto, eugenio tamura morimitsu

摘要: 监控**视频**通常会在采集和存储过程中出现模糊和曝光失真，这可能会对**视频**分析任务中的自动图像分析结果产生不利影响。本文的目的是部署一种算法，可以自动评估**视频**中曝光失真的存在。在这项工作中，我们设计和构建了一个用于**深度学习**的架构，用于识别**视频**中的扭曲。目标是了解**视频**是否存在曝光扭曲。这种算法可用于增强或恢复图像，或创建对象跟踪器失真感知。少

2018 年 10 月 26 日提交;**最初宣布** 2018 年 10 月。

19. 基于视频的基于视频的人的基于时空注意网络的重新识别

作 者 :shivansh rao, tanila rahman, mrigank rochan, yang wang

文摘: 我们考虑基于**视频**的人重新识别的问题。目标是从不同相机下拍摄的**视频**中识别一个人。本文提出了一种有效的基于时空注意的视频人员再识别模型. 我们的方法根据框架级别的功能为每个帧生成一个焦点分数。**视频**中所有帧的注意分数用于为输入**视频**生成加权特征向量。与大多数使用全局表示的**现有深度学习**方法不同，我们的方法侧重于注意力得分。在两个基准数据集上进行的大量实验表明，我们的方法实现了最先进的性能。这是一份技术报告。少

2018 年 10 月 26 日提交;**最初宣布** 2018 年 10 月。

20. 视频游戏的反向强化学习

作者:aaron tucker, adam g 兹, stuart russell

文摘: **深度强化学习**在一系列**视频**游戏环境中实现了超人的表现，但要求设计师手动指定奖励功能。提供目标行为的演示通常比设计描述该行为的奖励函数更容易。反向增强**学习**(irl) 算法可以从低维连续控制环境中的演示中推断奖励，但在将 irl 应用于高维**视频**游戏方面却很少有工作要做。在我们的 cnn-airl 基线中，我们修改了最先进的对抗性 irl (airl) 算法，使其使用 cns 作为生成器和鉴别器。为了稳定训练，我们规范奖励并增

加鉴别器训练数据集的大小。此外，我们还**学习**了一种低维状态表示，使用针对**视频**游戏环境进行调整的新型自动编码器体系结构。该嵌入作为奖励网络的输入，提高了专家演示的样本效率。我们的方法在简单的捕手**电子**游戏上实现了高水平的性能，大大优于 cnn-airl 基线。我们在恩杜罗·阿塔利赛车比赛中也得分，但与专家表现不匹配，凸显了进一步工作的必要性。少

2018 年 10 月 24 日提交;**最初宣布** 2018 年 10 月。

21. 视频语义分割的 uavid 数据集

作者:叶柳,乔治·沃塞曼,夏桂松,阿尔珀·伊尔马兹,迈克尔·英阳

摘要: **视频**语义分割是近年来计算机视觉研究的热点之一。它是机器人和自动驾驶等许多领域的感知基础。语义分割的快速发展极大地赋予了大规模数据集，特别是**与深度学习**相关的方法。目前，已存在多个复杂城市场景的语义分割数据集，如城市景观和 camvid 数据集。它们一直是比较语义分割方法的标准数据集。本文引入了一种新的高分辨率无人机**视频**语义分割数据集作为补体 uavid。我们的无人机数据集由 30 个**视频**序列组成，可捕获高分辨率图像。总共有 300 张图像被密

集地贴上了 8 个类的标签，用于城市场景理解任务。我们的数据集带来了新的挑战。我们提供了几种**深度学习**基线方法，其中提出的新的多尺度扩展网络通过多尺度特征提取性能最佳。我们还利用 crf 模型在空间和时间域中探讨了序列数据的可用性。少

2018 年 10 月 24 日提交;**最初宣布** 2018 年 10 月。

22. 面部表达式识别中的一致性约束

作者:lisa graziani, Stefano Melacci, marco gori

摘要: 从静态图像或**视频**序列中识别面部表情是一个广泛研究但仍具有挑战性的问题。**通过深度**神经架构或异构模型的集合获得的最新进展表明，集成多个输入表示可以获得最先进的结果。特别是，输入面的外观和形状，或某些面部分的表示形式，通常用于提高识别器的质量。本文研究了卷积神经网络 (cnn) 的应用，旨在构建一种可进一步应用于**视频**序列的静态图像表达式的多功能识别器。我们首先研究了不同面部部位在识别任务中的重要性，重点是外观和形状相关特征。然后，我们将**学习问题**投射到半监督设置中，利用**视频**数据，在那里只有几个帧被监督。培训数据的无监督部分用于加强三种类型的一致性，即时间一致性、面部部分预测之

间的一致性以及外观与形状表示之间的一致性。我们的实验分析表明，相干约束可以提高表达式识别器的质量，从而为有益地利用无监督**视频**序列提供了合适的依据。最后，我们给出了一些例子，其中基于形状的预测器的性能优于外观预测器。少

2018 年 10 月 17 日提交;**最初宣布** 2018 年 10 月。

23. nestdnn：面向资源感知的多租户设备深度学习，实现连续移动视觉

作者:方碧义,小曾,张米

摘要: 智能手机、无人机和增强现实耳机等移动视觉系统正在彻底改变我们的生活。这些系统通常同时运行多个应用程序，由于启动新应用程序、关闭现有应用程序和应用程序优先级更改等事件，它们在运行时的可用资源是动态的。在本文中，我们提出了一个考虑运行时资源动态的 ststdnn 框架，以实现移动视觉系统的资源感知多租户设备深度**学习**。nestdnn 使每个**深度学习**模型都能提供灵活的资源准确性权衡。在运行时，它动态地为每个**深度学习**模型选择最佳的资源精度权衡，以使模型的资源需求与系统的可用运行时资源相适应。在此过程中, nestdnn 有效地利用移动视觉系统中有限

的资源，共同最大限度地提高所有同时运行的应用程序的性能。实验表明，与资源无关的现状方法相比，nestdnn 在推理精度方面提高了 4.2%，视频帧处理率提高了 2.0x,**能耗**降低了 1.7 倍。少

2018 年 10 月 23 日提交;**最初宣布** 2018 年 10 月。

24. 基于层次的跨平台多视图特征学习在场地类别预测中的实现

作者:姜树强,敏威清,梅书欢

摘要: 在这项工作中，我们关注视觉场地类别预测，它可以促进基于位置的服务和个性化的各种应用。考虑到不同媒体平台的互补性，利用不同平台的与市场相关的媒体数据来提高预测性能是合理的。从直觉上看，识别一个场地类别涉及多种语义暗示，特别是对象和场景，因此它们应该共同为场地类别预测做出贡献。此外，这些场馆可以组织在自然的等级结构中，为指导场馆类别估计提供事先的知识。考虑到这些因素，我们提出了一个依赖层次结构的跨平台多视图功能**学习**(hcm-fl)框架，通过利用来自**其他**平台的图像从视频中进行场地类别预测。hcm-fl 包括两个主要组成部分，即跨平台迁移深度**学习**(cptdl) 和具有分层场地结构 (mvfl-hvs)

的多视图特征**学习**。cptdl 能够利用来自其他平台的图像从**视频**中**增强学到**的**深度**网络。具体来说, cptdl 首先使用**视频**训练**了一个深度**网络。这些来自其他平台的图像由学习的网络过滤，这些选定的图像随后被输入到这个被学习的网络中，以增强它。在 imagenet 和地点数据集上使用两种预先训练的网络。然后开发 MVFL-HVS，实现多视图特征融合。它能够嵌入层次结构本体，支持更多的判别关节特征**学习**。我们在来自葡萄树的**视频**和来自 foursqure 的图像上进行实验。这些实验结果证明了我们提出的框架的优点。少

2018 年 10 月 23 日提交;**最初宣布** 2018 年 10 月。

25. 这是在哪里？基于神经网络特征的视频地理定位

作者:萨尔瓦多麦地那,戴竹云,高英凯

摘要: 在这项工作中，我们提出了一种方法，地理定位范围内的区域内的**广泛**区域完全基于帧视觉内容。我们提出的方法是以谷歌街景为参照点，通过传统的图像检索技术来处理视频地理定位**问题**。为了实现这一目标，我们使用从 netvlad 获得的**深度学习**特征来表示图像，因为通过这个特征向量，相似性就是它们的 l2 范数。本文提出了一种基于矢量的方法集帧位化结果，从而

提高**了视频**地理定位结果。通过我们的实验发现的最佳聚合考虑了 netvlad 和 sift 的相似性，以及最相似结果的地理位置密度。为了测试我们提出的方法，我们从匹兹堡市区地区收集了一个新的**视频**数据集，以受益和刺激在这一领域开展更多的工作。我们的系统实现了 90% 的精度，同时对距离原始位置 150 米或两个街区范围内的**视频**进行地理定位。少

2018 年 10 月 22 日提交;**v1** 于 2018 年 10 月 21 日提交;**最初宣布** 2018 年 10 月。

26. 音频：基于音频的活动识别与大规模的声学嵌入从 youtube 视频

作者:梁大伟,爱迪生·托马斯

摘要: 活动传感和识别已被证明在保健和智能家庭应用中至关重要。与传统的活动识别方法，如使用加速度计或陀螺仪，声学方法可以收集丰富的人类活动信息和活动背景，因此更适合识别高级复合活动。然而，基于音频的活动识别在实践中总是受到从个人用户收集地面真相音频数据的繁琐而耗时的过程的影响。本文利用一般视频视频视频剪辑中数以百万计**的**嵌入功能，提出了一种完全不受用户培训数据影响的基于音频的活

动识别新机制。基于过度采样和**深度学习**方法的结合，我们的方案不需要进一步的特征提取或异常值过滤来实现。我们制定了识别 15 种常见家庭相关活动的方案，并评估了其在专用场景和野外脚本场景下的性能。在专门的录音测试中，我们的方案为所有 15 项活动提供了 81.1 的整体精度和 80.0% 的总 f 分。在野外脚本测试中，我们获得了平均前 1 名的分类精度为 64.9，在实际家庭环境中，4 名受试者的平均分类精度为 80.6。本文还讨论了数据集标签与目标活动之间的关联、分割大小的影响以及隐私问题等设计注意事项。少

2018 年 10 月 19 日提交;**最初宣布** 2018 年 10 月。

27. 多行人跟踪中概率数据关联的深部人再识别

作者:王强,王燕，基连 q. 温伯格,马克·坎贝尔

摘要: 我们提出了一种基于视觉的多行人跟踪的数据关联方法，利用**深度**卷积特征根据不同的人的外观来区分他们。**学习**这些重新识别 (重新识别) 特征时，它们对旋转、平移和背景更改等转换是不变的，从而可以一致地识别在场景中移动的行人。我们将重新识别特征集成到用于多人跟踪的通用数据关联似然模型中，通过使用该模型在两个评估**视频**序列中执行跟踪来实验验

证该模型，并检查性能与几种基准方法相比取得了进展。我们的研究结果表明，使用**深度**人员重新身份识别进行数据关联可大大提高对遮挡和路径交叉等挑战的跟踪鲁棒性。少

2018 年 10 月 19 日提交;**最初宣布** 2018 年 10 月。

28. 实现高效的大尺度图形神经网络计算

作者:马凌晓,智阳,苗友山,薛继龙,吴明，周丽东，戴亚飞

文摘: 最近**的深度学习**模型已经超越了低维的常规网格(如图像、**视频**和语音)，扩展到了高维图形结构数据(如社交网络、大脑连接和知识图)。这种演变导致了大型的基于图形的不规则和稀疏模型，这些模型超越了现有的**深度学习**框架的设计范围。此外，这些模型不易在并行硬件 (如 gpu) 上实现高效、大规模的加速。我们介绍了 ngra，这是基于图形的**深度**神经网络 (gnn) 的第一个并行处理框架。ngra 提出了一种新的 saga-nn 模型，用于将**深层**神经网络表示为顶点程序，每个层都处于定义明确的图形操作阶段 (散点图、应用边缘、收集、应用顶点)。该模型不仅允许 gnn 直观地表达，而且还便于映射到有效的数据流表示。ngra 通

过从 gpu 内核或多个 gpu 进行自动图形分区和基于分页的流处理来透明地解决可伸缩性挑战，这些数据仔细考虑数据位置、数据移动和并行重叠处理和数据移动。ngra 通过高度优化的 satter®在 gpu 上聚集操作员，进一步实现了效率，尽管它的稀疏性很大。我们的评估显示，ngra 扩展到现有框架都无法直接处理的大型真实图形，而在计地产流的多基线设计中，即使在较小的尺度上也能实现多达 4 倍的加速。少

2018 年 10 月 19 日提交;**最初宣布** 2018 年 10 月。

29. 基于学习的面向质量驱动的无线视频传输的电源控制

作者:创业, m. cenk gursoy, senem velipasalar

文摘: 本文研究了在总传输功率和最小所需**视频**质量约束下对多个用户的无线**视频**传输。为了在实时**视频**传输中为最终用户提供所需的性能水平，同时有效地利用能源资源，我们假设采用了功率控制。由于存在干扰，确定最优功率控制是一个非凸问题，但可以通过单调优化框架来解决。然而，单调优化是一种迭代算法，通常会产生相当大的计算复杂性，因此不适合实时应用。为了解决这个问题，我们提出了一种基于**学习**的方法，将资源分配算法的输入和输出视为未知的非线性映射，

并使用**深度**神经网络 (dnn) 来**学习**此映射。通过 dnn 进行的这种**学习**映射可以在给定的信道条件下快速提供最佳的功率水平。少

2018 年 10 月 16 日提交;**最初宣布** 2018 年 10 月。

30. 在人的速度：深度强化学习与行动延迟

作者:vlad firoiu, tinaju , josh tenenbaum

摘要: 最近，游戏人工智能的能力发生了爆炸式增长。许多类别的任务，从**电子**游戏到运动控制再到棋类游戏，现在都可以通过相当通用的算法来解决，这些算法基于**深入的学习**和强化**学习**,**可以从这些算法中学习**以最少的先验知识积累经验。然而，这些机器往往不能仅仅通过智力取胜--它们拥有超强的速度和精度，使它们能够以人类永远无法做到的方式行事。为了创造公平的竞争环境，我们将机器的反应时间限制在人的层面上，发现标准**的深度**强化**学习**方法的性能迅速下降。我们提出了一种解决人类感知所激发的动作延迟问题的方法--为代理提供一个环境的神经预测模型，"消除" 其环境中固有的延迟--并展示了其对专业玩家的有效性。超级粉碎兄弟. 梅莱，一个流行的控制台战斗游戏。少

2018 年 10 月 16 日提交;**最初宣布** 2018 年 10 月。

31. 利用预测编码进行时空预测的降门卷积 lstm

作 者 :nelly elsayed, anthony s. maida, magdy bayoumi

文摘: 时空 时空序列预测是**深度学习**中的一个重要问题。我们使用使用卷积长短期存储器 (convlstm) 模块的基于**深度学习**的预测编码框架来研究下一帧视频预测。我们引入了一种新的降门卷积 lstm (rgclstm) 体系结构，它所需的参数预算明显低于可比的 convlstm。我们的降门模型在使用较小的参数预算的同时，实现了与原始卷积 lstm 相同或更好的下一帧预测精度，从而缩短了训练时间。我们在移动的 mnist 和 kititi 数据集上的预测编码架构中测试了我们的减少门模块。我们发现，与标准的卷积 lstm 模型相比，我们的降门模型显著减少了训练参数总数的约 40%，经过的训练时间减少了 25%。这使得我们的模型对硬件实现更具吸引力，尤其是在小型设备上。少

2018 年 10 月 23 日提交;**v1** 于 2018 年 10 月 16 日提交;**最初宣布** 2018 年 10 月。

32. 基于深网的视频活动识别的静态和运动组合特征

作　　者:sameera ramasinghe, jathushanrajasgaran, vinoj jayasundara, kanchana Ramasinghe, ranga rodgo, ajith a. pasqual

摘要: **在深度学习**环境中的视频中的活动识别--或以其他方式--同时使用静态和预先计算的运动组件。将这两个组成部分结合起来，同时减少对**深**部网络的负担，仍然没有得到调查。此外，不清楚个别组成部分的贡献程度是多少，以及如何控制贡献。在这项工作中，我们使用了 cnn 产生的静态特征和运动特征的组合，以运动管的形式。我们提出了三个用于组合静态和运动分量的模式：基于方差比、主成分和 cholesky 分解。基于 cholesky 分解的方法允许控制贡献。通过静态和运动特征的方差分析给出的比率与基于 cholesky 分解方法的实验优化比吻合较好。在使用三个热门数据集测试时，所产生的活动识别系统更好或与现有的最先进技术相当。这些发现还使我们能够根据数据集丰富的运动信息来描述它的特征。少

2018 年 10 月 16 日提交;**最初宣布** 2018 年 10 月。

33. 使用场景名的视觉语义导航

作者：杨晓龙，王晓龙，ali farhadi，abhinav gupta，Roozbeh mottaghi

摘要：人类如何导航到新场景中的目标对象？我们是否使用多年来构建的语义功能原点来高效搜索和导航？例如，为了搜索杯子，我们在咖啡机附近搜索柜子，在冰箱里搜索水果。在这项工作中，我们专注于将语义优先点合并到语义导航任务中。我们建议使用图卷积网络将现有知识纳入一个**深层**强化**学习**框架。代理使用知识图中的特征来预测操作。对于评估，我们使用 ai2-thor 框架。我们的实验展示了语义知识如何显著提高性能。更重要的是，我们在对看不见的场景和对象的泛化方面表现出了改进。补充**视频**可通过以下链接访问：https://youtu.be/otKjuO805dE。少

2018 年 10 月 15 日提交;**最初宣布** 2018 年 10 月。

34. 每个像素计数 + +：通过 3d 整体理解联合学习几何和运动

作者:罗晨旭,杨振恒,王鹏，王洋，徐伟,南华，艾伦·尤尔

摘要：**通过深**卷积网络观看未标记**的视频**，学习在单个帧中估计 3d 几何和连续帧的光学流，是近年来的重要过程。目前最先进的（sota）方法独立处理这些任务。

当前深度估计管道的一个重要假设是，场景不包含运动物体，可以用光流来补充。在本文中，我们建议作为一个整体来处理这两个任务，即共同理解每个像素的三维几何和运动。这也消除了静态场景假设的需要，并在**学习**过程中强制实施固有的几何一致性，从而显著提高了这两项任务的结果。我们将我们的方法称为 "每个像素计数 + +" 或 "epc ++"。具体而言，在训练过程中，给定**视频**中的两个连续帧，我们采用三个并行网络分别预测摄像机运动 (motionnet)、密集深度图 (depknet) 和两个帧之间的每像素光流 (flownet)。在 2012 年 kitti 和 kitti 2015 数据集上进行了全面实验。深度估计、光流估计、气味测量、运动目标分割和场景流估计等五个任务的性能表明，我们的方法优于其他 sota 方法，证明了我们提出的每个模块的有效性方法。少

2018 年 10 月 14 日提交;**最初宣布** 2018 年 10 月。

35. 4d 全景深度地图中的人体通信

作者:李敏,周旺一腾,余景义

摘要: 经济实惠的 3d 全身重建系统的可用性产生了人类形状的自由视点**视频**(fvv)。大多数现有的解产生时间

上不相关的点云或网格与未知的点-顶点对应。单独压缩每个帧是无效的，仍然会产生超大的数据大小。我们提出了一个端到端深度**学习**方案，以建立密集的形状对应，并随后压缩数据。我们的方法使用稀疏的 "全景" 深度地图或 pdm 集，每个模拟一个内视同心马赛克。然后，我们开发了一种基于**学习**的技术来**学习** pdm 上的像素级特征描述符。结果被输入到基于自动编码器的网络中进行压缩。全面的实验证明，我们的解决方案对公共数据集和新捕获的数据集都是可靠和有效的。少

2018 年 10 月 11 日提交;**最初宣布** 2018 年 10 月。

36. 使用高度、颜色和性别在监控视频中检索人员

作者:hiren galiyawala, kenilshah , vandit gajjar, mehul s. raval

摘要: 一个人通常被描述为身高、身材、布色、布类型和性别等属性。这些属性被称为软生物识别。它们在监控**视频**中弥合了人的描述与人的检索之间的语义差距。本文提出了一种基于深度学习的基于深度学习的利用高度、布料颜色和性别进行**人**的检索的线性滤波方法。该方法使用掩码 r-nnn 进行像素化的人分割。它消除背景杂乱，并提供精确的人周围的边界。使用亚历克网

络对颜色和性别模型进行微调，并在 softbiorsurch 数据集上测试该算法。它在具有挑战性的条件下，利用语义查询对人的检索达到了较好的精度。少

2018 年 9 月 24 日提交;最初宣布 2018 年 10 月。

37. Isa2: 从外观的智能速度适应

作 者 :carlos herranz-perdiguero, roberto j. lópez-sastre

摘要: 在本文中，我们从外观上介绍了一个新的问题--智能速度适应 (isa)2).从技术上讲, isa 的目标 2 模型是预测给定的驾驶场景的图像车辆的适当速度。请注意，此问题不同于预测车辆的实际速度。它定义了一个新的回归问题，在考虑到交通情况的情况下，必须直接映射外观信息，以获得对车辆行驶速度的预测。首先，我们发布了一个新问题的新数据集，其中提供了多个驱动**视频**序列，每个帧具有注释的适当速度。然后，我们介绍两个**基于**深**部学习**的 isa2 模型，这些模型被训练在给定测试图像的情况下执行正确速度的最终回归。最后，我们进行了彻底的实验验证，结果显示了建议任务的难度级别。数据集和拟议的模型都将公开提供，以鼓励对这一问题进行急需的进一步研究。少

2018 年 10 月 11 日提交;最初宣布 2018 年 10 月。

38. vipl-hr: 一种用于低受限人脸视频脉冲估计的多模态数据库

作者:牛学松、胡汉、石光山、陈锡林

摘要: 心率 (hr) 是反映人类身体和情感活动的重要生理信号。传统的 hr 测量主要基于接触监视器，这是不方便的，可能会导致对象的不适。近年来，有人提出了从面部**视频**进行远程人力资源估计的方法。但是，大多数现有方法都侧重于控制良好的方案，但它们在受限较小的方案中的泛化能力尚不清楚。同时，由于缺乏大规模的数据库，使得深度表示学习**方法**在远程人力资源估计中的应用受到**了限制**。本文介绍了一个大型多模式 hr 数据库 (名为 vipl-hr)，其中包含 2, 451 种可见光**视频**(vis) 和 752 个近红外 (nir)**视频**，共 107 名受试者。我们的 vipl-hr 数据库还包含各种变化，如头部运动、照明变化和采集设备更改。我们还**学习**了一个深度的 hr 估计器 (称为节奏网络) 与建议的**时空**表示，这在公共领域和我们的 vipl-hr 估计数据库中都取得了很有希望的结果。我们想将 vipl-hr 数据库置于公共领域。少

2018 年 10 月 11 日提交;**最初宣布** 2018 年 10 月。

39. 深度学习时代的显著性预测：一个实证研究

作者:阿里·博尔吉

摘要: 近年来，由于**深度学习**和大规模注释数据的进步，视觉显著性模型的性能有了很大的飞跃。然而，尽管付出了巨大的努力和巨大的突破，但模型在达到人类水平的准确性方面仍然没有达到。在这项工作中，我探讨了该领域的景观，强调新的**深度**显著模型、基准和数据集。通过两个图像基准和两个大型**视频**数据集，对大量的图像和**视频**显著性模型进行了回顾和比较。此外，我还确定了造成模型与人之间差距的因素，并讨论了为建立下一代更强大的显著性模型而需要解决的剩余问题。讨论的一些具体问题包括：目前的模型以何种方式失败，如何补救，从注意力的认知研究**中学到**什么，明确的显著判断与固定有关，如何进行公平的模型比较，以及显著性模型的新应用。少

2018 年 10 月 11 日提交;**v1** 于 2018 年 10 月 8 日提交;**最初宣布** 2018 年 10 月。

40. 通过观看立体声视频，在无人监督的情况下联合学习光流和深度

作者:杨王,杨振恒,王鹏，杨毅，罗晨旭, 徐伟

文摘: 通过深度神经网络通过**观看视频**学习深度和光流**,**近年来取得了显著进展。在本文中，我们通过利用立体**视频**中的基本几何规则，共同解决了这两个任务。具体而言，给出视频中的两个连续立体图像对，我们首先估计三个神经网络的深度、相机自我运动和光流。然后通过对深度和自我运动产生的估计光流和刚性流的比较，将整个场景分解为运动前景和静态背景。我们提出了一种新的一致性损失，让光学流从静态区域更精确的刚性流**中学习**。我们还设计了一个刚性对齐模块，通过使用估计的深度和光流，帮助细化自我运动估计。在 kititi 数据集上的实验表明，我们的结果明显优于其他最先进的算法。源代码可在 http://githu. com/baidu-trem· undthflow

2018 年 10 月 8 日提交;**最初宣布** 2018 年 10 月。

41. sfv: 加强从视频中学习身体技能

作 者 : 薛 斌 鹏 , 金 泽 安 乔 , jitendra malik, pieter abbeel, sergey levine

摘要: 基于运动捕捉的数据驱动角色动画可以产生高度自然主义的行为，如果与物理模拟相结合，可以提供对

物理扰动、环境变化和形态的自然过程反应差异。运动捕获仍然是最流行的运动数据来源，但收集 mocap 数据通常需要大量检测的环境和行为者。在本文中，我们提出了一种方法，使物理模拟字符学习技能的视频(sfv)。我们的方法基于**深度**姿态估计和**深度强化学习**，允许数据驱动的动画利用网络上大量公开的**视频**剪辑，比如 youtube 上的视频剪辑。这样就可以通过查询所需行为**的视频记录来**实现字符控制器的快速、轻松设计。由此产生的控制器对扰动具有鲁棒性，可以适应新的设置，可以执行基本的对象交互，并可以通过增强学习重定向**到**新的形态。我们进一步证明，我们的方法可以通过对从观察到的姿势中初始化的**学习**控制器进行正向模拟，从静止图像中预测潜在的人体运动。我们的框架能够**学习**广泛的动态技能，包括运动、杂技和武术。少

2018 年 10 月 15 日提交;**v1** 于 2018 年 10 月 8 日提交;**最初宣布** 2018 年 10 月。

42. 深层强化学习中的互联网拥塞控制

作 者 :nathan jay, noga h. rotman, p. brighten godfrey, michael schapira, aviv tamar

文摘: 我们提出并研究了一个新的、及时的深度**强化学习**应用领域: 互联网拥塞控制。拥塞控制是调节流量源数据传输速率以高效、公平地分配网络资源的核心网络任务。拥塞控制是计算机网络研究和实践的基础，最近随着实时**视频**、增强和虚拟现实等具有挑战性的互联网应用的出现，拥塞控制一直是人们广泛关注的话题。物联网，和更多。我们在最近引入的面向性能的拥塞控制(pcc) 框架的基础上，将拥塞控制协议设计制定为 rl 任务。我们的 rl 框架为网络从业者，甚至应用程序开发人员提供了机会，以培训基于小型、引导模型或复杂的自定义模型的拥塞控制模型，以满足其本地性能目标的要求。资源和需求的价值。我们提出并讨论了必须克服的挑战，以实现我们控制互联网拥堵的长期愿景。少

2018 年 10 月 7 日提交;**最初宣布** 2018 年 10 月。

43. 深度概率视频压缩

作 者 :jun han, salvator lombardo , christopher schroers, stephan mant

文摘: 提出了一种深度概率**视频**压缩的变分推理方法。我们的模型将变分自动编码器 (vae) 的进步用于顺序

数据，并将其与最近在神经图像压缩方面的工作结合起来。该方法共同**学习**将原始**视频**转换为一个低维表示，以及熵代码这个表示根据一个时间条件的概率模型。我们将潜在空间划分为局部（每个帧）和全局（每个段）变量，并表明训练 vae 来利用这两种表示可以提高速率失真性能。对来自不同复杂性和多样性的公共数据集的小型**视频**进行的评估表明，我们的模型在对通用**视频**内容进行培训时，会产生有竞争力的结果。如果模型在类似的**视频**上进行训练，则可对具有专门内容的**视频**实现极端压缩性能。少

2018 年 10 月 5 日提交;**最初宣布** 2018 年 10 月。

44. 用于从图像中学习视频对象检测器的无监督对抗视觉水平域适应

作者:avisek lahiri, charan reddy , prabir kumar bis 什

摘要: **深部基于学习**的对象检测器需要数千个多样化的边界框和类注释示例。尽管近年来随着多个大型静态图像数据集的发布，图像对象检测器显示出了快速的进展,**但由于**附加注释**的视频**帧稀缺，视频上的目标检测仍然是一个悬而未决的问题。拥有一个强大**的视频**对象探测器是**视频**理解和策划**视频**中的大规模自动注释的

重要组成部分。图像和**视频**之间的域差异使得图像对象探测器的可转移性低于最佳。最常见的解决方案是使用弱监督注释，其中必须标记**视频**帧，以确保没有对象类别。这仍然需要手动工作。在本文中，我们向前迈出了一步，将无监督的对抗性图像到图像的转换概念调整为难以承受静态高质量图像，使其在视觉上无法与一组**视频**帧区分开来。我们假设存在完全注释的静态图像数据集和未批注的**视频**数据集。对象检测器使用原始数据集的批注对对手转换的图像数据集进行训练。在 youtube 对象和 youtube 对象上进行的实验-带有两个当代基线对象探测器的子集数据集显示，这种无监督像素级域适应提高了**视频**帧的泛化性能。直接应用原始图像对象检测器。另外，与最近监督不力的方法相比，我们实现了有竞争力的业绩。本文可作为图像转换在跨域目标检测中的应用。少

2018 年 10 月 4 日提交;**最初宣布** 2018 年 10 月。

45. 超深度：自我监督，超分辨单目深度估计

作者:sudeep pillai, rares ambrus , adren gaidon

摘要: 最近在自我监督的单目深度估计技术接近监督方法的性能，但只在低分辨率下运行。我们表明，高分辨

率是高保真自我监督单目深度预测的关键。在最近单图像超分辨率**深度学习**方法的启发下，我们提出了一个深度超分辨率的亚像素卷积层扩展，准确地综合了它们的高分辨率差距。相应的低分辨率卷积特征。此外，我们还引入了一个可区分的翻转增强层，它可以准确地将预测与图像及其水平翻转版本融合在一起，从而减少了由于遮挡而在视差图中生成的左阴影区域的影响。这两种贡献都比最先进的自我监督深度获得了显著的业绩提升，并对公共 kititi 基准进行了估计。我们方法的**视频**可以在 https://youtu.be/jKNgBeBMx0I 上找到。少

2018 年 10 月 3 日提交;**最初宣布** 2018 年 10 月。

46. 利用循环矩阵训练用于视频分类的紧凑型深层学习模型

作者:亚历山大·阿劳霍, benjamin Negrevergne, yann chevaleyre, jamal atif

摘要: 在现实世界中，模型精度几乎不是唯一需要考虑的因素。大型模型占用更多内存，并且计算密集型更大，这使得它们难以训练和部署，尤其是在移动设备上。本文以线性代数和**深度学习**的交叉点的最新结果为基础，

论证了如何将结构施加到较大的权重矩阵上，以减小模型的大小。我们提出了非常紧凑的视频分类模型的基础上，最先进的**网络**架构，如深包的**框架**, netvlad 和 NetFisherVectors 器。然后，我们使用大型 youtube-8m **视频**分类数据集进行彻底实验。正如我们将展示的，循环 dbof 嵌入实现了一个很好的大小和精度之间的权衡。少

2018 年 10 月 8 日提交;**v1** 于 2018 年 10 月 2 日提交;**最初宣布** 2018 年 10 月。

47. 云追逐者：低计算功率器件的实时深度学习计算机视觉

作者:罗正义,奥斯汀·斯莫尔,利亚姆·杜根,斯蒂芬·莱恩

摘要: 物联网 (iot) 设备、移动电话和机器人系统由于其有限的计算能力，往往被剥夺**了深度学习**算法的强大功能。然而，为了提供应急、家庭援助、监控等时间关键服务，这些设备往往需要对其相机数据进行实时分析。本文试图通过利用云的计算能力，提供一种可行的方法，将基于高性能**深度学习**的计算机视觉算法与低资源、低功耗设备集成。通过将计算工作卸载到云，无需专用硬件就可以在现有的低计算能力设备上启用

深层神经网络。基于树莓派的机器人 "云追逐者" 旨在展示使用云计算执行实时视觉任务的强大功能。此外，为了减少延迟和提高实时性能，提出了将实时**视频**帧传输到云中的压缩算法并进行了评估。少

2018 年 10 月 2 日提交;**最初宣布** 2018 年 10 月。

48. 无监督轨迹分割与多模态手术演示的推广

作者:邵振洲,赵洪发,谢继新，应区,永关,谭金东

文摘: 为了提高机器人辅助微创手术机器人**学习**的手术轨迹分割效率，本文提出了一种利用**视频**和运动学数据的快速无监督方法，然后进行了推广。程序来解决过度分割的问题。采用无**监督深学习**网络--堆叠卷积自动编码器，有效地从**视频**中提取出更多的判别特征。为了进一步提高分割的精度，一方面利用小波变换对**视频**和运动学数据中存在的噪声进行了滤波。另一方面，通过基于预定义的相似度测量识别没有状态转换的相邻段，提高了分割结果。在公共数据集上进行的大量实验表明，与最先进的方法相比，我们的方法在较短的时间内实现了更高的分割精度。少

2018 年 10 月 1 日提交;**最初宣布** 2018 年 10 月。

49. 反思自我驾驶：多任务知识，提高泛化能力，提高事故解释能力

作 者 ： 李 志 浩 ， toshiyuki motoyoshi,kazuma sasaki, Tetsuya ogata,shigeki sugano

摘要: 当前的端到**端**深度**学习**驾驶模型存在两个问题: (1) 训练驾驶数据集多样性有限时，未观察到的驾驶环境的泛化能力较差 (2) 缺乏事故解释能力。驾驶模型不能像预期的那样工作。为了解决这两个问题，我们认为相关的易任务知识是解决困难任务的好处，我们提出了一个新的驱动模型，它由面向 \ textion{see 和认为} 的感知模块和用于 \textit{see 的驱动模块组成。行为，并培训它与多任务感知相关的基础知识和推动知识的逐步。具体分割地图和深度图 (像素级对图像的理解) 被认为是在生成最终控制命令之前解决更容易的驾驶感知问题的 "内容" {0。为困难的驾驶任务。实验结果表明，多任务感知知识对提高泛化和事故解释能力是有效的。在未经训练的 drl 测试中，在未经训练的城市的 frl 测试中完成最困难的导航任务的平均成功率超过了目前的基准方法，在训练有素的天气中，成功率为 15%，在未经训练的天气中，成功率超过 20%。演示**视**

频链接是： https://www.youtube.com/watch ?v=N7ePnnZZwdE 少

2018 年 9 月 28 日提交;**最初宣布** 2018 年 9 月。

50. 学会在野外检测假脸图像

作者:许志忠,李嘉燕,庄一秀

摘要: 虽然生成对抗性网络 (gan) 可以用来生成逼真的图像，但这些技术的不当使用带来了隐藏的问题。例如, gan 可用于为特定人员和不适当的事件生成被篡改**的视频**，从而创建对特定人员有害的图像，甚至可能影响该个人安全。本文将开发一种**深度**伪造鉴别器 (deepfd)，以高效、高效地检测计算机生成的图像。直接**学习**二进制分类器是比较棘手的，因为很难找到常见的判别特征来判断从不同的有机器官产生的假图像。为了解决这一缺陷，我们在寻找不同甘肃的合成图像的典型特征时采取了对比损失，然后将分类器连接起来，检测此类计算机生成的图像。实验结果表明，所提出的 deepfd 成功地检测出了 94.7% 的假图像，这些假图像是由几种最先进的有机器官产生的。少

2018 年 10 月 18 日提交;**v1** 于 2018 年 9 月 24 日提交;**最初宣布** 2018 年 9 月。

51. 通过 hr 光流估计学习视频超分辨率

作者:王龙光,郭玉兰,林泽平,邓新普,魏安

摘要: **视频**超分辨率 (sr) 旨在生成一系列高分辨率 (hr) 帧，这些帧具有来自低分辨率 (lr) 对应方的合理且一致的详细信息。准确对应的生成在**视频** sr 中起着重要的作用。传统**的视频** sr 方法表明，图像和光流的同步 sr 可以提供准确的对应和更好的 sr 结果。然而，lr 光流被用于现有**的基于深度学习**的通信生成方法中。在本文中，我们提出了一个端到端可训练的**视频** sr 框架，以超解析图像和光流。具体而言，我们首先提出了一个光流重建网络 (ofrnet)，以粗化的方式推断 hr 光流。然后，根据 hr 光流进行运动补偿。最后，将补偿 lr 输入输入输入到超分辨率网络 (srnet) 以生成 sr 结果。大量实验表明，hr 光流比 lr 提供更精确的对应，并提高了精度和一致性性能。vid4 和 davs-10 数据集的比较结果表明，我们的框架实现了最先进的性能。少

2018 年 10 月 25 日提交;**v1** 于 2018 年 9 月 23 日提交;**最初宣布** 2018 年 9 月。

52. 在视频帧插值中实现自适应可分离卷积

作者:mart kartašev, carlo rapisarda, dominik fay

摘要: 随着**深**神经网络越来越流行，很多注意力都集中在计算机视觉问题上，这些问题过去是用更传统的方法来解决的。**视频**帧插值是其中一个挑战，已经看到**了新的**研究涉及各种技术的深度学习。本文复制了 niklaus 等人在自适应可分离卷积上的工作，该工作在**视频**帧插值任务上获得了高质量的结果。我们应用了在较小的数据集中训练的相同网络结构，并对各种不同的丢失函数进行了实验，以确定数据稀缺情况下的最佳方法。最好的结果模型仍然能够提供视觉上令人愉快的**视频**，尽管获得较低的评价分数。少

2018 年 9 月 20 日提交;**最初宣布** 2018 年 9 月。

53. 一种快速、准确的人脸检测、识别和验证系统

作者:rajeev ranjan, ankan bansal, jingxiaozheng, h 行 yuxu , joshua glason, boyu lu, anirudh nanduri, jun-cheng chen, carlos d. castillo, rama 切拉帕

摘要: 大型注释数据集的提供和负担得起的计算能力使 cnn 在各种物体检测和识别基准方面的性能有了令人印象深刻的改进。这些，再加上对**深度学习**方法的更好理解，也提高了机器对人脸理解的能力。cn 能够检测人脸，定位面部地标，估计姿势，并识别无约束图像和

视频中的人脸。本文详细介绍了用于无约束人脸识别和验证的**深度学习**管道，该管道在多个基准数据集上实现了最先进的性能。我们提出了一种新型的人脸检测器，**深**金字塔单面检测器 (dpssd)，它是快速和能够检测具有大规模变化的人脸 (特别是微小的脸)。我们给出了自动人脸识别涉及的各个模块的设计细节: 人脸检测、地标定位和对齐以及人脸识别/验证。我们提供了具有挑战性的无约束人脸检测数据集的人脸检测器的评价结果。然后，我们介绍了 iarpa janus 基准 a、b 和 c (ijb-a、jibb-b、ijb-c) 和 janus 挑战集 5 (cs5) 的实验结果。少

2018 年 9 月 20 日提交;**最初宣布** 2018 年 9 月。

54. 学习场景识别的有效 rgb-d 表示

作者:宋新航,姜树强,路易斯·赫兰茨,陈成鹏

摘要: **由于**有了大型数据集 (如 "地点")，深度卷积网络 (cnn) 可以在 rgb 场景识别方面取得令人印象深刻的效果。相比之下，由于本文所涉及的 rgb-d 数据的两个局限性，rgb-d 场景识别仍不发达。第一个限制是缺乏培养**深度学习**模型的深度数据。我们不是通过微调或传输 rgb 特定的功能来解决这一限制，而是提出一

种体系结构和两步培训方法，通过修补程序使用弱监视直接**学习有效的**特定于深度的功能。由此产生的 rgb-d 模型还受益于更互补的多式联运特性。另一个限制是深度传感器范围短（通常为 0.5 米至 5.5 米），导致深度图像无法在 rgb 图像所能捕获的场景中捕获遥远的物体。我们表明，这种限制可以通过使用 rgb-d **视频**来解决，在这些视频中，随着摄像机在场景中的传播，会积累更全面的深度信息。在这种情况下，我们引入 isia rgb-d **视频**数据集，以评估 rgb-d 场景识别与**视频**。我们的**视频**识别架构结合了卷积和递归神经网络（rnn），这些神经网络在三个步骤中接受训练，数据越来越复杂，以**学习**有效的功能（即补丁、帧和序列）。我们的方法可获得最先进的 rgb-d 图像（nyud2 和 sun rgb-d）和**视频**(isia rgb-d) 场景识别性能。少

2018 年 9 月 17 日提交;**最初宣布** 2018 年 9 月。

55. 情绪识别的多模态特征、分类器和融合方法的研究

作者:郑莲,雅丽，陶建华,黄健

摘要: 自动情感识别是一项具有挑战性的任务。在本文中，我们介绍了我们的努力，基于**音频视频**的情感识别在野生（emotiw）2018 年挑战，这要求参与者分配一个

单一的情感标签的视频剪辑从六个普遍的情绪（愤怒，厌恶，恐惧，幸福，悲伤和惊喜）和中立。提出的多模情感识别系统考虑了音频、**视频**和文本信息。除了手工功能外，我们还通过转移**学习从深**中性网络（dnn）中提取瓶颈特征。对时间分类器和非时间分类器进行了评价，以获得最佳的单峰情绪分类结果。然后提取可能性并将其传递到光束搜索融合（bs-fusion）中。我们在 emotiw 2018 挑战中测试我们的方法，我们获得了有希望的结果。与基线系统相比，有了显著改善。我们在测试数据集上实现了 60.34 的精度，仅比获胜者低 1.5%。这表明我们的方法很有竞争力。少

2018 年 9 月 13 日提交;**最初宣布** 2018 年 9 月。

56. 利用卷积神经网络研究不同隐藏层和时代手写数字识别精度变化的研究与观察

作　者：rezoana bente arif, md. abu bakr siddique, mohammad mahmudur rahman khan, mahjabin rahman oishe

摘要: 如今,**深度学习**可以应用于医学、工程等多个领域。在**深度学习**中，卷积神经网络（cnn）广泛应用于模式和序列识别、**视频**分析、自然语言处理、垃圾邮件检测、

主题分类、回归分析、语音识别、图像分类、目标检测、分割、人脸识别、机器人和控制。与其在大型应用中近乎人的水平准确相关的好处，导致近年来美国有线电视新闻网的接受程度越来越高。本文的主要贡献是分析美国有线电视新闻网隐藏层模式对网络整体性能的影响。为了证明这种影响，我们在修改后的国家标准与技术研究所（mnist）数据集上应用了不同层次的神经网络。同时，观察不同数量的隐藏层和时代的网络精度变化，并对它们进行比较和对比。利用随机梯度和反向传播算法对系统进行训练，并采用前馈算法进行测试。少

2018 年 9 月 22 日提交;**v1** 于 2018 年 9 月 17 日提交;**最初宣布** 2018 年 9 月。

57. 零例视频检索的双密集编码

作者:董建峰、李锡荣、徐朝西、舒灵吉、王迅

文摘: 本文提出了零例**视频**检索这一具有挑战性的问题。在这种检索范式中，最终用户通过自然语言文本中描述的临时查询搜索未标记**的视频**，但没有提供可视示例。大多数现有方法都是以概念为基础的，从查询和**视频**中提取相关概念，并据此在两种模式之间建立联系。相反，本文遵循了一种新的无概念、**基于**深度**学习**的编

码趋势。为此，我们提出了一种双**深编码**网络，该网络既适用于视频，也适用于**查询**方面。该网络可以灵活地与现有的通用空间**学习**模块耦合，用于**视频**-文本相似度计算。正如 msr-vtt、trecvid 2016 和 2017 辅助视频搜索这三个基准**的**实验所显示的那样，该方法为零示例**视频**检索建立了新的最新技术。少

2018 年 9 月 17 日提交;**最初宣布** 2018 年 9 月。

58. 从运动中学习结构

作 者 :clément pinard, laure chevlley, antoine manzanera, david filliat

摘要: 这项工作是基于对深度神经网络从一张图像执行深度预测所使用的**质量**指标的质疑，以及对最近发表的关于在不受监督的情况下从视频中学习深度**的**作品的可用性的质疑.为了克服它们的局限性，我们建议以同样的无监督方式从单目**视频**中学习一个深度地图推理系统,该系统以一对图像为输入。该算法实际上是从运动**中学习结构**,而不仅仅是从上下文外观中学习结构。对尺度因子问题进行了明确处理，绝对深度图可以从相机位移幅度中估计出来，这可以很容易地从廉价的外部传感器中测量出来。我们的解决方案在通过微调进

行域变化和适应方面也要强大得多，因为它并不完全依赖于上下文。考虑了两个用例，不稳定的移动相机**视频和**稳定的。这一选择的动机是无人机（无人飞行器）机箱，该用例通常提供可靠的方向测量。我们提供的实验表明，在只有速度才能知道的实际情况下，我们的网络在大多数深度质量测量方面都优于竞争对手。在众所周知的 kitti 数据集上给出了结果，该数据集为我们的第二个用例提供了强大的稳定性，但也包含了非常典型的车内道路背景的移动场景。然后，我们在合成数据集上提供结果，我们认为该数据集更能代表典型的无人机场景。最后，我们提出了两个域适应用例，显示了我们的方法优于单视图深度算法的鲁棒性，这表明它更适合于高度可变的视觉上下文。少

2018 年 10 月 19 日提交;**v1** 于 2018 年 9 月 12 日提交;**最初宣布** 2018 年 9 月。

59. 并行可分离的三维卷积，用于视频和体积数据的理解

作者:felix gonda, dunlai wei, toufiq parag, hanspeter Toufiq

文摘: 对于**视频**和体积数据的理解，三维卷积层被广泛用于**深度学习**，但是，代价是增加计算和训练时间。最

近的工作试图用卷积块取代三维卷积层，例如二维和一维卷积层的结构组合。本文提出了一种新的卷积块--平行可分离三维卷积（pmscn），它应用了 n 2d 的 m 并行流和一个沿不同维度的一维卷积层。我们首先在数学上证明需要并行流（pm）通过张量分解来替换单个 3d 卷积层。然后，我们将现代网络体系结构中常见的连续 3d 卷积图层与多个二维卷积层（cn）共同替换。最后，我们的经验表明，pmscn 适用于不同的主干体系结构，如 resnet、densenet 和 unet，适用于不同的应用，如**视频**动作识别、mri 大脑分割和电子显微镜分割。在所有这三个应用中，我们将最先进模型中的 3d 卷积图层替换为 pmscn，并使测试性能提高约 14%，模型大小和平均值减少 40%。少

2018 年 9 月 11 日提交;**最初宣布** 2018 年 9 月。

60. 用于行动识别的时空映射

作者:宋晓林,兰翠玲,曾文军， 邢俊良,杨景宇,孙晓燕

摘要: **深部学习**模型在图像分类和目标检测等与图像相关的计算机视觉任务中取得了巨大的成功。然而，对于像人的行动识别这样的与**视频**相关的任务，进展还没有那么重要。主要的挑战是缺乏有效和高效的模型来建

模视频中丰富的时间空间信息。我们引入了一个简单而有效的操作，称为时间空间映射（tsm），通过联合分析**视频**的所有帧来捕捉帧的时间演化。我们提出了**一个视频**级 2d 要素表示，通过将所有帧的卷积要素转换为 2d 要素图，称为 videocap。由于每一行都是帧的矢量化特征表示，则对时空特征进行了紧凑的表示，而时间动态演化也很好地嵌入在一起。在视频地图表示的基础上，进一步提出了浅卷积神经网络中的时间注意模型，以有效地利用时空动力学。实验结果表明，该方案在具有挑战性的人类动作基准数据集 hmdb51 上，通过竞争基线方法--时间段网络（tsn），实现了最先进的性能，精度提高了 4.2%。少

2018 年 9 月 10 日提交;**最初宣布** 2018 年 9 月。

61. 用相位代替光流进行动作识别

作 者 :omar hommos，silvia l. pintea，pascal s.m. mettes, jan c. van gemert

摘要: 目前，动作识别最常见的运动表示是光流。光学流是基于粒子跟踪，坚持拉格朗日的动态观点。与拉格朗日的观点不同的是，欧拉动力学模型并不跟踪，而是描述了局部变化。对于**视频**，一种基于欧拉相位的运动

表示，使用复杂的可操纵滤波器，最近已成功地应用于运动放大和**视频**帧插值。在这些作品的启发下，我们在这里提出了在**深层**架构中**学习**欧拉运动表示的方法，以实现行动识别。我们以端到端的方式学习复杂领域中**的**筛选器。我们设计这些复杂的过滤器，以类似于复杂的 gapor 过滤器，通常用于相位信息提取。我们提出了一个基于这些复杂滤波器的相位信息提取模块，该模块可用于任何网络体系结构中提取欧拉表示。我们对我们提出的相位提取模块提取的欧勒运动表示的附加值进行了实验分析，并与现有的基于光流的运动表示进行了比较。少

2018 年 9 月 14 日提交;**v1** 于 2018 年 9 月 10 日提交;**最初宣布** 2018 年 9 月。

62. 深度学习跟踪关联的无人监督人员再识别

作者:李敏贤,朱夏天，龚少刚

摘要: 最存在的角色识别 (重新 id) 方法在每个相机对手动标记的一对配对训练数据上重新进行监督模型**学习**。这导致实际重新身份部署中的可扩展性较差，因为每个相机对的图像正面和负面对都没有详尽的标识标签。在这项工作中，我们通过提出一种无监督的重新身

份深度**学习**方法来解决这个问题，这种方法能够逐步发现和利用来自自动生成的人的基本重新身份判别信息在端到端模型优化中跟踪**视频**中的数据。我们制定了一个跟踪协会无监督深度**学习**(taudl) 框架，其特点是共同**学习**每个相机 (相机内) 跟踪协会 (标签) 和交叉摄像头跟踪相关最大限度地发现最有可能的跨摄像机视图的跟踪关系。大量实验证明，与使用六个人的重新身份基准数据集的最先进的无监督和域适应重新定位方法相比，拟议的 taudl 模型具有优越性。少

2018 年 9 月 8 日提交;**最初宣布** 2018 年 9 月。

63. 利用统计奖励积累改进政策学

作者:邓玉斌,顾玉友,林大华,唐晓欧,陈易露

文摘: **深层**加固..。更多

2018 年 9 月 7 日提交;**最初宣布** 2018 年 9 月。

64. 视图不变行为表示的无监督学习

作者:李俊南,黄永康, 赵琪，莫汉 s. 坎坎哈利

文摘: 最近人类行动识别的成功与**深度学习**方法大多采用了监督学习范式，这需要大量的人工**标记**数据才能

取得良好的效果。但是，标签收集是一个昂贵且耗时的过程。在这项工作中，我们提出了一个无监督的**学习**框架，利用未标记的数据来**学习视频**表示。与以往视频表示学习的工作不同，我们的无监督**学习**任务是使用源视图中的**视频**表示来预测多个目标视图中的 3d 运动。通过**学习**外推交叉视图运动，表示可以捕获对动作有判别性的视图不变运动动力学。此外，我们还提出了一种视对波训练方法，以增强对视图不变特征的**学习**。我们展示**了在**多个数据集中进行操作识别的经验形式的有效性。少

2018 年 9 月 6 日提交;**最初宣布** 2018 年 9 月。

65. 基于运动唾液引导时空传播的无监督视频对象分割

作者:胡元亭,黄嘉斌,亚历山大 g. 施温

摘要: 无监督**视频**分割在从对象识别到压缩的各种应用中发挥着重要作用。然而，到目前为止，快速运动、运动模糊和遮挡带来了重大挑战。为了应对无监督**视频分割的**这些挑战，我们开发了一种新的显著估计技术以及一种基于光流和边缘线索的新的邻域图。我们的方法带来了显著更好的初步前景背景估计，其稳健和准确的跨时间扩散。我们评估了我们提出的算法在具有挑战

性的 davis, segtrack v2 和 fbms-59 数据集上。尽管只使用了在 200 张图像上训练过的标准边缘探测器，但我们的方法在无监督环境中的效果优于基于**深度学习**的方法。我们甚至在 davis 数据集的半监督设置中展示了与基于**深度学习**的方法相当的竞争结果。少

2018 年 9 月 4 日提交;**最初宣布** 2018 年 9 月。

66. 视频匹配：基于视频对象分割的匹配

作者:胡元亭,黄嘉斌,亚历山大 g. 施温

摘要: **视频**对象分割具有挑战性，但在视频分析的各种应用中也很重要。最近的工作制定视频对象分割**作为**一项预测任务，使用**深**网，以实现吸引人的最先进的性能。由于该公式是一项预测任务，因此大多数这些方法都需要在测试期间进行微调，以便**深**网记住给定**视频**中感兴趣对象的外观。但是，微调既耗时又昂贵，因此算法远不是实时的。针对这一问题，我们开发了一种新的基于匹配的**视频**对象分割算法。与基于记忆的分类技术不同，该方法**学习**将提取的特征与提供的模板匹配，而不记住对象的外观。我们验证了该方法在具有挑战性的 DAVIS-16、datvs-17、youtube-对象和 jumpcut 数据集上的有效性和稳健性。大量的结果表明，我们的方法

在不微调的情况下实现了可比的性能，在计算时间上更有利。少

2018 年 9 月 4 日提交;**最初宣布** 2018 年 9 月。

67. 深秋--利用深时空卷积自动编码器进行无创坠落检测

作者:jacob n 报, shehroz s. khan,alex mihailidis

摘要: 人的跌倒很少发生;然而，从健康和安全的角度来看，检测坠落是非常重要的。由于瀑布的罕见性，很难使用监督分类技术来检测它们。此外，在这些高度倾斜的情况下，也很难提取域特定的特征来识别落差。本文提出了一个新的框架，即 \ texit{deepffar}，该框架将坠落检测问题表述为异常检测问题。\ textytit 拟{deepfall} 框架介绍了**使用深度**时空卷积自动编码器使用非侵入性传感方式从正常活动**中学习**空间和时间特征的新方法。我们还提出了一种新的异常评分方法，该方法结合**了视频**序列中帧的重建分数，以检测看不见的坠落。我们在通过非侵入性传感模式、热像仪和深度摄像机收集的三个可公开数据集上测试了 \ texttitumtum·d 普尔·迪普法勒框架，并显示了与传统自动编码器和卷积相比的卓越结果自动编码器方法来识别看不见的瀑布。少

2018 年 9 月 13 日提交;**v1** 于 2018 年 8 月 30 日提交;**最初宣布** 2018 年 9 月。

68. 图像和视频中的目标非线性对抗性摄动

作者:roberto rey-de-castro, herschel rabitz

文摘: 我们介绍了一种针对单个图像或**视频**的对抗摄动学习方法。**学习到的**扰动被发现是稀疏的，同时包含了高水平的特征细节。因此，提取的扰动允许一种形式的对象或动作识别，并提供了深入了解什么特征所研究的**深度**神经网络模型认为什么是重要的，在作出分类决定。从对抗性的角度来看，稀疏的扰动成功地将模型混淆为错误分类，尽管通过目视检查，扰动样本仍然属于同一原始类。这是在一个前瞻性的数据增强方案的角度进行了讨论。稀疏但高质量的扰动也可用于图像或**视频**压缩。少

2018 年 8 月 27 日提交;**最初宣布** 2018 年 9 月。

69. 中网络：一种紧凑型面部视频伪造检测网络

作　者　:darius afchar, vincent nozick, junichi yamagishi, isao echizen

文摘: 本文提出了一种自动检测视频中的人脸篡改的方法，重点介绍了**用于**生成超逼真**的伪造视频**的两种最新技术: deep 机假和 face2face。传统的图像取证技术通常不太适合**视频**，因为压缩会使数据严重退化。因此，本文采用**了深度学习**的方法，提出了两个网络，这两个网络的层数都很低，可以聚焦于图像的介观特性。我们评估现有数据集上的快速网络和我们从在线**视频**构建的数据集。测试证明了一个非常成功的检测率，超过 98% 的深度假和 95% 的脸 2 脸。少

2018 年 9 月 4 日提交;**最初宣布** 2018 年 9 月。

70. 通过对抗性目标和增强转化进行深层强化学习

作者:许淑贤，沈一超, 陈炳宇

文摘: 在过去的几年里,**深度强化学习**已经被证明可以解决**电子**游戏或棋类游戏等复杂状态的问题。智能代理的下一步将能够在任务之间进行概括，并利用以前的经验更快地获得新技能。然而，目前大多数强化**学习**算法即使面对非常相似的目标任务，也往往会遭受灾难性的遗忘。我们的方法使代理能够从单个源任务中概括知识，并在面对新任务时使用半虚拟学习方法促进**学习**进度. 我们在 atari 游戏上对这种方法进行了评估，

这是一种流行**的强化学习**基准，并表明它优于基于预训练和微调的常见基线。少

2018 年 9 月 3 日提交;**最初宣布** 2018 年 9 月。

71. 内部网络：大型多传感器照片逼真的室内场景数据集

作 者 :wenbin li, sajad saeedi, john mccormac, ronald clark, dimos tzoumanikas, chingye, yu 作 huang, rui tang, stefan leutenegger

文摘: 数据集在计算机视觉社区中获得了极大的普及，从**基于深度学习**的方法的培训和评估到同时本地化和映射 (slam) 的基准测试。毫无疑问，合成图像具有巨大的潜力，因为在不繁琐的手工接地图像注释或测量的情况下，可在可获得的数据量方面实现可扩展性。在这里，我们展示了一个数据集，目的是提供更高的照片真实感、更大的比例、更大的可变性，以及与现有数据集相比提供更广泛的用途。我们的数据集利用了数以百万计的专业室内设计以及数百万生产级家具和物品资产的可用性，所有这些都带有精美的几何细节和高分辨率纹理。我们渲染高分辨率和高帧速率**的视频**序列遵循逼真的轨迹，同时支持各种相机类型，并提供惯性测量 。 在 发 布 数 据 集 的 同 时 ， 我 们 将 在

https://interiornetdataset.github.io 提供交互式模拟器软件的可执行程序以及呈现器。为了展示我们数据集的可用性和唯一性，我们展示了稀疏和密集的 slam 算法的基准测试结果。少

2018 年 9 月 3 日提交;**最初宣布** 2018 年 9 月。

72. 论事件边界在光流自我中心活动识别中的作用

作 者 :alejandro cartas, estefania talavera, petia radeva, mariella Dimiccoli

摘要: 事件边界作为**视频**中人类活动的检测、本地化和识别任务的预处理步骤，发挥着至关重要的作用。通常情况下，虽然其固有的主观性，时间边界是手动提供的训练操作识别算法的输入。然而，它们在自我中心光流领域的活动识别作用迄今被忽视。在本文中，我们提供了如何自动计算边界可以影响活动识别结果在新兴领域的自我中心光流的见解。此外，我们收集了一个新的注释数据集，由 15 人通过可穿戴的照片相机获得，我们使用它来显示几个**深度学习**的架构的泛化功能给看不见的用户。少

2018 年 9 月 6 日提交;**v1** 于 2018 年 9 月 2 日提交;**最初宣布** 2018 年 9 月。

73. 利用竞技强化学习在阿塔利游戏之间进行视觉转换

作 者 :akshita mittel, sowmya munukutla , himanshi yadav

文摘: 本文探讨了**利用深层**强化**学习**代理将知识从一个环境转移到另一个环境中的方法。更具体地说，该方法利用异步优势参与者评论家 (a3c) 体系结构，使用在 atari 的源游戏中训练的代理对目标游戏进行泛化。我们建议使用与目标游戏的不同表示形式并行训练的多个代理更新模型，而不是为目标游戏微调预先训练的模型。利用传输对的**视频序列之间**的视觉映射来推导目标游戏的新表示形式;对目标游戏的这些可视表示进行的培训在性能、数据效率和稳定性方面改进了模型更新。为了展示该体系结构的功能, atari 游戏 pongv0 和突破 v0 正在 openai 健身房环境中使用;作为源和目标环境。少

2018 年 9 月 2 日提交;**最初宣布** 2018 年 9 月。

74. 利用深层强化学习进行自然语言搜索

作者:ankitshah, tyler vuong

文摘: 最近在**深度强化学习**方面的成功是让经纪人**学会**如何玩围棋，在没有任何事先了解比赛的情况下击败世界冠军。在这项任务中，代理人必须根据件的位置决定要采取什么行动。最近，人们使用基于自然语言的图像文本描述来探索人员搜索，以实现**视频**监控应用 (s. li 等人. al)。我们看到 (f. et al) 提供了一个端到端方法，用于基于对象的检索，使用**深度强化学习**，而不限制检测到哪些对象。但是，我们相信，对于实际应用 (如人员搜索)，定义特定约束 (识别人员而不是从常规对象检测开始) 将在性能和所需的计算资源方面具有优势。在我们的任务中,**深度强化学习**将通过重塑边界框的大小来定位图像中的人。**深部**具有适当约束**的强化学习将**只查找图像中的相关人员，而不是对图像中的每个对象进行排名的不受约束的方法。对于人员搜索，代理试图在与描述匹配的图像中的人员周围形成一个紧密的边界框。边界框初始化为完整的图像，在每个时间步长上，代理都会根据人员的描述和当前边界的像素值，决定如何更改当前边界框，使其在人员周围有更严格的绑定箱。代理执行操作后，将根据当前边界框和地面真相框的 "联合" (iou) 交叉点 (iou) 给予奖励。一旦代理人认为边界框覆盖了该人，就会表明找到了该人。少

2018 年 9 月 2 日提交;**最初宣布** 2018 年 9 月。

75. arbee：实现对野外情感身体表情的自动识别

作者:余罗,叶建波,雷金纳德 b.亚当斯,小，贾丽，米歇尔·纽曼, 王建波

摘要: 可以说，人类天生就准备好了从微妙的身体动作中理解他人情感表达的能力。如果机器人或计算机能够获得这种功能，许多机器人应用就成为可能。然而，在不受约束的情况下自动识别人类的身体表达是令人生畏的，因为对身体运动和情感表达之间的关系缺乏充分的了解。目前的研究是计算机和信息科学、心理学和统计学之间的一项多学科努力，提出了一种可扩展和可靠的众包方法，用于收集野生感知情感数据，供计算机**学习**识别人类的肢体语言。为此，创建了一个大型且不断增长的附加注释数据集,其中包含 9 876 个身体动作视频剪辑和 13 239 个人类角色，名为 bold (肢体语言数据集)。全面的统计分析从数据集中揭示了许多有趣的见解。开发并评价了一种基于身体运动的情绪表达模型系统，名为 arbee (情感身体表情的自动识别)。我们的特征分析显示了拉班运动分析 (lma) 特征在描述唤醒方面的有效性。我们使用**深层**模型的实验进一步证明了身体表达的可计算性。这项工作中介绍的数据集和发现很可能成为未来在肢体语言理解方面的多种发现

的启动平台，这将使未来的机器人在与人类互动和协作时更加有用。少

2018 年 8 月 28 日提交;**最初宣布** 2018 年 8 月。

76. 基于图的 i-图变换

作者:renata Khasanova, pascal frossard

摘要: **视觉**数据的学习变换不变表示是计算机视觉中的一个重要问题。**深**卷积网络在图像和**视频**分类任务中取得了显著的效果。然而，它们在进行几何变换的图像分类方面只取得了有限的成功。在这项工作中，我们提出了一个新的变换不变图形为基础的网络 (tiglanet)，它**学习基于**图形的特征，这些特征本质上是对等距变换的不变的，例如输入图像的旋转和转换。特别是，图像被表示为图形上的信号，这使得可以用图形谱卷积和动态图池层替换深网络中的经典卷积和池层，这些图谱卷积和动态图池层共同促成了不变性。等距变换。我们的实验表明，与对数据转换非常敏感的经典体系结构相比，测试集的旋转和转换图像具有高性能。我们框架固有的不变性特性提供了关键优势，例如提高了对数据可变性的恢复能力，并在有限的训练集中保持了性能。我们的代码可在线使用。少

2018 年 8 月 21 日提交;**最初宣布** 2018 年 8 月。

77. 无监督视频人员重新识别的深层关联学习

作者:陈燕贝,朱夏田，龚少刚

摘要: **深部学习**方法已经开始主导基于**视频**的人的重新识别 (重新识别) 的研究进展。然而，现有的方法大多考虑监督**学习**，这需要在跨视图对数据上贴标签时进行详尽的人工工作。因此，在实际**视频**监控应用中，它们严重缺乏可扩展性和实用性。在这项工作中，为了解决**视频**人的重新身份任务，我们制定了一个新的**深度**关联**学习**(dal) 方案，这是第一个使用模型中没有任何身份标签的端到端深度**学习**方法初始化和培训。dal 通过端到端方式联合优化两个基于边距的关联损耗，学习了一种**深度**重新识别匹配模型，这有效地限制了每个帧与最匹配的相机内表示的关联和交叉摄像头表示。现有的标准 cnn 可以很容易地在我们的 dal 计划中使用。实验结果表明，我们提出的 dal 在三个基准上明显优于当前最先进的**无监督视频**人重新识别方法: prid 2011、iLIDS-VID 和 mars。少

2018 年 8 月 22 日提交;**最初宣布** 2018 年 8 月。

78. 用于活动识别的深层自适应时间集合

作 者 ： 宋 西 波 ， 张 恩 - 张 恩 ， vijay chandrasekhar, bappaditya manal

文摘: **深**部神经网络最近在人类活动识别方面取得了具有竞争力的精度。但是，仍有改进的余地，尤其是在建模长期的时间重要性和确定**视频**中不同时间段的活动相关性方面。为了解决这个问题，我们提出了一个可学习和可微的模块:**深度**自适应时间池 (datp)。datp 采用一种自注意机制自适应地汇集不同**视频**段的分类分数。具体来说，使用框架级特征，datp 会回归不同时间段的重要性，并为它们生成权重。值得注意的是，仅使用**视频**级标签对 datp 进行培训。除了**视频**级活动类标签外，不需要额外的监督。我们进行了广泛的实验，以研究各种输入功能和不同的重量模型。实验结果表明，datp 可以**学习**为关键**视频**段分配较大的权重。更重要的是，datp 可以提高帧级特征提取器的训练。这是因为相关的临时段在反向传播过程中被分配了较大的权重。总体而言，我们在 ucf101、hmdb51 和动力学数据集上实现了最先进的性能。少

2018 年 8 月 22 日提交;**最初宣布** 2018 年 8 月。

79. 基于深层视频的性能克隆

作者:kfir ab 人, mingyishi, jing liao, dani linchinski, 陈宝全, daniel cohen-or

文摘: 我们提出了一种新**的基于视频**的性能克隆技术。在使用捕获目标演员的外观和动态**的参考视频**训练了一个**深度生成网络**后，我们能够生成**视频**，让这个演员在那里进行其他表演。所有的训练数据和驾驶性能都是以普通**视频**段的方式提供的，没有运动捕捉或深度信息。我们的生成模型是作为一个**有**两个分支的深度神经网络实现的，这两个分支都使用共享权重训练相同的时空条件生成器。一个分支，负责**学习**生成以各种姿势生成目标参与者的外观，使用 \ 指 {一方} 训练数据，从参考**视频**中自行生成。第二个分支使用未配对的数据来改进看不见的姿势序列的时间相干**视频**再现的生成。我们展示了各种有希望的结果，我们的方法能够生成时间连贯的视频，为具有挑战性的**场景**，其中参考和驾驶**视频**包括非常不同的舞蹈表演。补充**视频**: https://youtu.be/JpwsEeqNhhA。少

2018 年 8 月 21 日提交;**最初宣布** 2018 年 8 月。

80. 像素客观性：学习在图像和视频中自动分割通用对象

作者:熊波, suyog dutt jain, kisten gruman

摘要: 我们提出了一个端到端**学习**框架，用于分割图像和**视频**中的通用对象。给定一个新的图像或**视频**，我们的方法为所有 "对象相似" 区域生成像素级掩码--即使是在训练中从未见过的对象类别也是如此。我们将任务表述为一个结构化的预测问题，即将对象背景标签分配给每个像素，并使用**深度**完全卷积网络来实现。当应用于**视频**时，我们的模型进一步集成了一个运动流，网络学习将外观和运动结合起来，并尝试提取所有突出的对象，无论它们是否在移动。在核心模型之外，我们的方法的第二个贡献是如何利用培训注释的不同优势。像素级注释很难获得，但对于培训**深度**网络方法进行分段至关重要。因此，我们提出了利用弱标记数据学习密集前景分割的方法。对于图像，我们将对象类别示例与图像级标签以及具有边界级别注释的相对较少的图像混合在一起的值。对于**视频**，我们展示了如何引导弱注释**视频**以及为图像分割而训练的网络。通过在多个具有挑战性的图像和**视频**分割基准上的实验，我们的方法提供了一致的强大结果，并改进了通用 (看不见) 对象的全自动分割的最先进技术。此外，我们还演示了我们的方法如何有利于图像检索和图像重定向，这两者在给定我们高质量的前景地图时都很繁荣。代码、模型

和 **视** **频** 的 网 址 是 : http://vision.cs.utexas.edu/projects/pixelobjectness/少

2018 年 8 月 11 日提交;**最初宣布** 2018 年 8 月。

81. 前景对象检测的无监督学习

作 者 :ioana croitoru, simion-vlad bogolin, marius leordeanu

摘要: 无监督**学习**是当今计算机视觉中最困难的挑战之一。这项任务具有巨大的实际价值，在人工智能和新兴技术中有许多应用，因为可以以相对较低的成本收集大量未标记**的视频**。本文在检测单个图像中主要前景对象的背景下，讨论了无监督**学习**问题。我们训练学生**深度**网络来预测教师路径的输出，该路径在**视频**或大型图像集合中执行无监督对象发现。我们的方法不同于在无监督对象发现上发布的方法。我们在训练期间移动无监督**学习**阶段，然后在考试时沿着学生路径应用标准的前馈处理。该策略的优点是允许在培训期间增加泛化的可能性，同时在测试中保持快速。我们的无监督**学习**算法可以贯穿几代学生-教师培训。因此，在第一代接受培训的一批学生网络集体创造了下一代的教师。在实验中，我们的方法在**视频**、无监督图像分割和显著性检测

三个当前数据集上获得了最高的结果。在测试时，所提出的系统速度很快，比公布的无监督方法快一到两个数量级。少

2018 年 8 月 14 日提交;**最初宣布** 2018 年 8 月。

82. 面对面：手术室中的匿名视频

作者:evangello flouty, oddseal zisimopoulos , danail stoyanov

摘要: 手术手术室 (or) 中的**视频**采集越来越有可能，并有可能用于计算机辅助干预 (cai)、外科数据科学和智能 or 集成。捕获的**视频**天生携带敏感信息，为了保持患者和临床团队的身份，这些信息不应该完全可见。当手术**视频**流存储在服务器上时，如果在医院外拍摄，**则**必须在存储前匿名播放视频。在本文中，我们描述了如何**将深度学习**模型 "更快的 r-cnn" 用于此目的，并帮助匿名处理在 or 中捕获的**视频**数据。该模型检测并模糊面，以保持匿名性。在测试了现有的人脸检测训练模型后，收集了一个适合手术环境的新数据集，其面部被外科口罩和帽子挡住，以便进行微调，从而在 or 中实现更高的面部检出率。我们还提出了一个时间正则化

内核，以提高召回率。该模型在应用时间平滑前后分别实现了 8.05% 和 93.45 的人脸检测召回率。少

2018 年 8 月 6 日提交;**最初宣布** 2018 年 8 月。

83. 不应憎恨：打击网络仇恨言论

作 者 :binny mathew, hardik tharad,subham Prajwal, prajwal singhania, suan kalyan maity, pawan goyal, animesh mukherje

摘要: 社交媒体中的仇恨内容越来越多。虽然脸谱、推特、谷歌试图采取几个步骤来解决这一仇恨内容，但他们最经常冒着侵犯言论自由的风险。另一方面，反言论为在不丧失言论自由的情况下解决网络仇恨问题提供了有效途径。因此，这些平台的另一种策略可以是促进反言论，作为对仇恨内容的防御。然而，要想成功地推广到这样的反言论，就必须对其在网络世界中的动态有**一个深刻**的了解。缺乏精心策划的数据在很大程度上阻碍了这种理解。在本文中，我们使用 youtube 的评论创建并发布了第一个用于反演讲的数据集。数据包含 9438 手动注释，其中标签指示注释是否为反演讲。这些数据使我们能够首次对反言论的语言结构进行严格的测量研究。这一分析得出了各种有趣的见解，如：反言

论评论收到的不反言论评论得到的喜欢是双倍的，对于某些社区来说，大多数的反言论评论往往是仇恨言论,不同类型的反言论并不都是同样有效的，发布反言论的用户的语言选择与详细的心理语言学分析所揭示的发布反言论的用户有很大的不同。最后，我们构建了一组机器**学习**模型，这些模型能够自动检测到youtube **视频**中的反言论，f1-分数为 0.73。少

2018 年 8 月 13 日提交;**最初宣布** 2018 年 8 月。

84. 深 rnn 开放世界的立体声视频匹配

作者:钟一兰,李洪东,戴玉超

摘要: **深部**基于**学习**的立体匹配方法在不同的基准中取得了巨大的成功，并取得了最高分。然而，与大多数数据驱动的方法一样，现有**的深层**立体匹配网络也存在一些众所周知的缺点，例如需要大量的标记训练数据，而且它们的性能受到以下因素的限制:泛化能力。在本文中，我们提出了一个新的递归神经网络 (rnn)，它以一个连续的 (可能是以前看不到的) 立体视频作为输入，并直接预测一个深度映射在每个帧没有预先训练的过程，而不需要地面真相深度地图作为监督。由于反复出现的性质 (由两个 w 发货-lstm 块提供)，我们的网络

能够记忆和**学习**过去的经验，并修改其内部参数（网络权重），以适应以前看不见或不熟悉的情况环境。这表明了网络的非凡泛化能力，使其适用于 {\ em 开放世界} 环境。我们的方法可以在场景内容、图像统计、照明和季节条件 {\ em 等} 方面的变化而工作。通过大量的实验，我们证明了该方法在不同的场景之间无缝地适应。同样重要的是，在立体声匹配精度方面，它在 kitti 和 Middlebury 立体声等标准基准数据集上的性能优于最先进的深层立体声方法. 少

2018 年 8 月 12 日提交;**最初宣布** 2018 年 8 月。

85. 深层视频彩色传播

作 者 :simone meyer, victor cornillère, abdelaziz djelouah, christopher schroers, markus gross

摘要: 传统**的视频**颜色传播方法依赖于连续**视频**帧之间的某种形式的匹配。使用外观描述符，颜色在空间和时间上传播。但是，这些方法在计算上非常昂贵，并且不能利用场景的语义信息。在这项工作中，我们提出了一个**深入的学习**框架的颜色传播结合本地策略，传播颜色逐帧确保时间稳定，和全局策略，使用语义的颜色传播在更长的范围。我们的评价显示了我们的策略相对于

现有**的视频**和图像颜色传播方法以及神经照片逼真风格传输方法的优势。少

2018 年 8 月 9 日提交;**最初宣布** 2018 年 8 月。

86. 可控制的图像到视频的翻译--以面部表情生成为例

作者:范丽杰,黄文兵,庄甘,黄俊洲,龚博清

文摘: 最近在**深度学习**方面的进步使得利用神经网络生成照片逼真图像成为可能，甚至可以从输入**视频**剪辑中推断**视频**帧。本文研究了图像到视频的翻译，特别是面部表情的视频，既是为了进一步探索这一探索，也是为了促进我们自己对现实应用的**兴趣**。与图像到图像的平移相比，这个问题通过另一个时间维数来挑战**深度**神经网络。此外，它的单个输入图像无法使用大多数依赖于重复模型的现有**视频**生成方法。我们提出了一种用户可控的方法，以便从单个人脸图像生成不同长度的**视频**剪辑。表达式的长度和类型由用户控制。为此，我们设计了一种新的神经网络体系结构，该体系结构可以将用户输入集成到其跳过连接中，并对神经网络的对抗训练方法提出了一些改进建议。实验和用户研究验证了我们的方法的有效性。特别是，我们要强调的是，即使是野外的面部图像（从网上下载和作者自己的照

片)，我们的模型可以生成高品质的面部表情**视频**，其中约 50\% 被亚马逊标记为真实土耳其机械工人。少

2018 年 8 月 8 日提交;**最初宣布** 2018 年 8 月。

87. 何时看：基于视频的人的重新识别的深度暹罗注意力网络

作者:林武,杨旺,高俊斌,薛丽

摘要: **基于视频**的人员重新识别（重新识别）是监控系统中的一个核心应用，在安全方面存在重大问题。由于视觉变化大、帧速率不受控制，视频片段中跨不相交相机视图匹配的人员本质上具有挑战性。人的自我评价有两个关键的步骤，即判别特征**学习**和度量学习。但是，现有的方法独立考虑这两个步骤，并且没有充分利用**视频**中的时间和空间信息。在本文中，我们提出了一个暹罗注意架构，共同**学习**时空**视频**表示及其相似度指标。该网络从每个帧的区域提取局部卷积特征，并通过在测量与另一个行人**视频**的相似性时关注不同的区域来增强其判别能力。注意机制被嵌入到空间门控复发单元中，有选择地传播相关特征，并通过网络记忆其空间依赖关系。该模型实质上**了解**哪些部分（\ 他人强调{哪个}）与匹配的人员相关且与众不同，并在其中给予

更高的重视。提出的 siamese 模型是端到端可培训，共同**学习**可比较的隐藏表示配对行人**视频**及其相似值。在三个基准数据集上进行的大量实验表明，拟议**的深**部网络的每个组件的有效性，同时优于最先进的方法。少

2018 年 10 月 14 日提交;**v1** 于 2018 年 8 月 2 日提交;**最初宣布** 2018 年 8 月。

88. 在无监督时空特征学习中引入可伸缩性

作者:sujoy paul, sourya roy, amit k. roy-chowdhury

文摘: **深层**神经网络是一种高效**的学习**机器，它利用大量手动标记的数据来**学习**判别特征。但是，获取大量的监督数据，尤其是**视频**数据，在各种计算机视觉任务中可能是一项繁琐的工作。这就需要在无人监督的环境中从**视频**中**学习**视觉功能。在本文中，我们提出了一个计算简单，但有效的框架，学习时空特征嵌入从未标记**的视频**。我们训练一个卷积 3d 暹罗网络使用正对**和负对挖掘从视频在一定**的概率假设。对三个数据集的实验结果表明，我们提出的框架能够**学习**可用于相同以及交叉数据集和任务的权重。少

2018 年 8 月 14 日提交;**v1** 于 2018 年 8 月 6 日提交;**最初宣布** 2018 年 8 月。

89. 基于深度学习的物联网和移动边缘计算嵌入式系统的多目标视觉跟踪

作　　者:beatriz blanco-filgueira, daniel garcía-lesta, mauro fernández-sanjurjo, víctor m. brea, paula lópez

文摘: 最先进的**深度学习**方法的计算和内存需求仍然是一个缺点，必须加以解决，使它们在物联网终端节点上有用。特别是，最近的结果描述了使用卷积神经 netwoks (cnn) 进行图像处理的前景，但对于物联网和移动边缘计算应用来说，软件和硬件实现之间的差距已经相当大，因为它们高功耗。该方案执行在 nvidia jetson tx2 开发套件上实现的基于低功耗和实时**深度学习**的多目标视觉跟踪。它包括摄像头和无线连接功能，并为移动和户外应用提供电池供电。使用机载摄像机 detrusc **视频**数据集捕获的具有代表性序列的集合用于举例说明所提出的算法的性能，并便于基准测试。在功耗和帧速率方面的研究结果表明了在嵌入式平台上**进行深度学习**算法的可行性，但还需要对 cnn 的联合算法和硬件设计进行更多的努力。少

2018 年 7 月 31 日提交;最初宣布 2018 年 8 月。

90. 基于 rgb 的网球动作识别，利用深厚的历史长期短期记忆

作者:蔡嘉新,新唐

摘要: 动作识别在计算机视觉中的应用越来越受到 rgb 输入的关注，部分原因是虚拟网球比赛、网球技术、战术分析等运动的身体模拟和统计有潜在的应用.近年来，**基于深度学习**的方法在动作识别方面取得了很有希望的效果。本文提出了采用卷积神经网络表示的三维网球镜头识别加权长期记忆方法。首先，利用预先训练的初始网络从每个**视频**帧中分别提取局部二维卷积神经网络空间表示。然后，引入加权长期记忆解码器，利用时间 t 的输出状态和时间 t-1 的历史嵌入特征，利用评分加权方案生成特征向量。最后，我们使用所采用的 cnn 和加权 lstm 将原始视觉特征映射到一个向量空间，生成视觉序列的时空语义描述，并对动作**视频**内容进行分类。在基准上的实验表明，我们的方法只使用简单的原始 rgb **视频**可以达到更好的性能比最先进的基线网球镜头识别。少

2018 年 9 月 25 日提交;**v1** 于 2018 年 8 月 2 日提交;**最初宣布** 2018 年 8 月。

91. 基于能量的多 gpu 对流层神经网络的优化

作 者 :francisco m. castro, nicolás guil, manuel j. marín-jiménez , jesús pérez-serrano, manuel ujaldón

摘要: **深部学习与学习**(dl) 应用在人工智能领域的发展势头越来越大，特别是在 gpu 展示了加速其具有挑战性的计算需求的卓越技能之后。在此背景下，卷积神经网络 (cnn) 模型是在一系列复杂应用中取得成功的一个有代表性的例子，特别是在数据集上，在这些数据集中，目标可以通过增加的局部特征层次结构来表示语义复杂性。在大多数实际场景中，改进结果的路线图依赖于 cnn 设置，其中涉及蛮力计算，研究人员最近证明 nvidia gpu 是加速的最佳硬件对应方之一。我们的工作是对这些发现的补充，对旗舰图像和**视频**应用部署 cnn 的关键参数进行能源研究: 分别是物体识别和步态识别人。我们根据两个最受欢迎的网络 (resnet/alesnet) 评估四个不同网络的能耗: resnet (167 层)、2d cnn (15 层)、CaffeNet (25 层) 和 resetim (94 层)，使用 64、128 和 256 的批处理大小，然后将这些数据相关联以加速和精确来确定最佳设置。在具有双

maxwell 和双 pascal tian x gpu 的多 gpu 服务器上的实验结果表明，能量与性能相关，与 maxwell 相比，pascal 可能有高达 40% 的增益。较大的批次尺寸可扩展性能提升和节能，但我们必须密切关注准确性，这有时表明我们倾向于小批量。我们期望这项工作为现代 hpc 时代的广泛 cnn 和 dl 应用提供初步指导，其中 gflopsw 比率是主要目标。少

2018 年 8 月 1 日提交;**最初宣布** 2018 年 8 月。

92. 下一代宣传的广告创作体系

作 者 :atul nautiyal , killian mccabe, murhaf hossari, soumyabratadev, matthew nicolson, clare conran, declan mckibben, jian tang, xu wei, 弗朗索瓦·皮蒂

摘要: 随着互联网多媒体数据的迅速普及，为观众制作**的视频**也在迅速上升。这使得观众可以跳过**视频**中的广告中断，使用广告拦截器和 "跳过广告" 按钮-将在线营销和宣传带到摊位。本文演示了一个能够有效地将新广告集成到**视频**序列中的系统。我们使用最先进**的技术**，从**深度**学习和计算摄影测量，有效地检测现有的广告，

并无缝地将新广告集成到**视频**序列。这对有针对性的广告很有帮助，为下一代的宣传铺平了道路。少

2018 年 8 月 1 日提交;**最初宣布** 2018 年 8 月。

93. 通过重新保护减少训练记忆

作者:冯建伟,黄东

摘要: **深部**在现代图像/**视频**数据库上进行训练时，神经网络 (dnn) 需要巨大的 gpu 内存。遗憾的是, gpu 内存始终是有限的，这限制了图像分辨率、批处理大小和**学习**速率，可以进行调整以获得更好的性能。在本文中，我们提出了一种新的方法，称为重新转发，大大减少了培训中的内存使用。我们的方法仅在第一个向前将张量保存在图层的子集上，并执行额外的本地向前 (重新转发过程) 来计算向后所需的缺少张量。总内存成本成为 (1) 图层子集的成本和 (2) 重新转发过程的最大成本的总和。我们提出了用线性或任意优化图实现 dnn 最优内存解的理论和算法。实验表明，重新转发减少了大量的训练记忆在所有流行的 dnn，如亚历克莱克, vgg 网, resnet, densenet 和宗称网。少

2018 年 7 月 31 日提交;**最初宣布** 2018 年 8 月。

94. 学习看力：基于 rgb-point 云时空卷积网络的手术力预测

作 者 :gong gong，x 兴 通 liu，michael peven，mathias unberath，austin reiter

摘要: 机器人手术已被证明在外科手术中提供了明显的优势，然而，主要的局限性之一是获得触觉反馈。由于设计具有精确力反馈的硬件解决方案通常具有挑战性，因此我们建议使用 "视觉线索" 来推断组织变形的力。内窥镜**视频**是一种可自由使用的被动传感器，在任何微创程序已经利用它的意义上。为此，我们采用**深入的学习**方法，从**视频**中推断力，将其作为典型的复杂和昂贵的硬件解决方案的一种极具吸引力的低成本和准确的替代方案。首先，我们使用贴有光力传感器的达芬奇手术系统，在幻象环境中演示我们的方法。其次，我们在体外肝脏器官上验证我们的方法。我们的方法在体外研究中产生 0.814 n 的平均绝对误差，表明它可能是一个有前途的替代硬件为基础的手术力量反馈在内窥镜程序。少

2018 年 7 月 31 日提交;**最初宣布** 2018 年 8 月。

95. 关注是我们所需要的：以对象为中心的锁定注意以自我为中心的活动识别

作者:swathikiran sudhakaran, oswald lanz

文摘: 本文提出了一种用于自我中心活动识别的**端**到端可训练深度神经网络模型。我们的模型是建立在这样的观察基础上的，即自我中心的活动是高度的特点的对象及其位置在**视频**中。在此基础上，我们开发了一个空间关注机制，使网络能够关注包含与所考虑的活动相关的对象的区域。我们学习高度专业化的注意地图的每个帧使用类特定的激活从 cnn 预先培训的通用图像识别，并使用它们的时空编码的视频与卷积 lstm。我们的模型使用原始**视频**级活动类标签在弱监督环境中进行训练。尽管如此，在标准的自我中心活动基准上，我们的模型超过了高达 + 6% 的积分识别精度，这是目前性能最好的方法，利用手工分割和对象定位的强大监督进行培训。我们对网络生成的注意图进行了可视化分析，揭示了网络成功识别视频帧中存在的相关对象，这可以解释**强大**的识别性能。我们还讨论了有关设计选择的广泛的烧蚀分析。少

2018 年 7 月 31 日提交;**最初宣布** 2018 年 7 月。

96. 利用视频中的运动优先性改善人的分割

作者:陈玉婷,张文延，陆海伦,吴廷凡，孙敏

文摘: 尽管在**基于深度学习**的语义分割方面取得了许多进展，但现实世界中由于分布不匹配而导致的性能下降是经常遇到的。最近，一些领域适应和主动**学习**的方法已被提出来，以减轻性能下降。然而，很少注意利用视频中的**信息**，这些信息在大多数相机系统中都是自然捕捉到的。在这项工作中，我们建议利用**视频**中的 "运动优先"，在弱监督的主动**学习**环境中改进人类分割。通过**在视频**中利用光流提取运动信息，我们可以提取出可能与人体段相对应的候选前景运动段 (称为运动优先)。我们建议**学习**一种基于内存网络的政策模型，通过强化**学习**来选择强的候选段 (称为强运动优先)。所选线段具有较高的精度，可直接用于调整模型。在新收集的监控摄像机数据集和公开提供的城市街数据集中，我们提出的方法可提高跨多个场景和模式 (即 rgb 至红外线 (ir)) 的人工分割性能。最后但并非最不重要的是，我们的方法是经验上对现有域适应方法的补充，因此，通过将我们的弱监督主动**学习**方法与域适应方法相结合，可以获得额外的性能提升。少

2018 年 7 月 30 日提交;**最初宣布** 2018 年 7 月。

97. 深层强化学习加强时空自我监督

作者:uta büchler, biagio brattoli, björn ommer

摘要: 卷积神经网络的自我监督**学习**可以利用大量廉价的未标记数据来训练强大的特征表示。作为代理任务，我们共同解决了空间和时间域中视觉数据的排序问题。训练样本的排列，是自我监督的核心，通过排序，到目前为止，从一个固定的预选集随机采样。在**深度强化学习的**基础上，我们提出了适应网络状况的抽样政策，目前正在进行培训。因此，根据预期的更新卷积特征表示的效用，对新的排列进行采样。对无监督和转移**学习**任务的实验评估显示了图像和**视频**分类标准基准和最近邻居检索的竞争性能。少

2018 年 7 月 30 日提交;**最初宣布** 2018 年 7 月。

98. 在实时 mri 中实现声乐轨迹形状动力学的自动语音识别

作者:pramit sahar, praneeth srungarapu, sidney fels

摘要: 声道结构在产生可区分的语音、调节气流和在语音产生中产生不同的共振腔方面发挥着至关重要的作用。它们包含丰富的信息，可以用来更好地理解潜在的

语音产生机制。作为将声道形状几何自动映射到声学的一个步骤，本文采用了有效**的视频**动作识别技术，如长期重复卷积网络 (lrcn) 模型，以识别不同的声音运动。从声道的动态塑造的元音-辅音元音 (vcv) 序列。这样的模型通常结合了基于 cnn 的**深层**分层视觉特征提取器和递归网络，理想的情况下，使网络在时空上足够**深**，以**了解**一个短**的顺序动态**用于**视频**分类任务的视频剪辑。我们使用的数据库包括由 17 扬声器在 vcv 话语过程中形成声带的 2d 实时 mri。讨论了该类算法在各种参数设置和各种分类任务下的比较性能。有趣的是，结果表明，在语音分类的上下文中，模型性能与泛型序列或**视频**分类任务有显著差异。少

2018 年 7 月 29 日提交;**最初宣布** 2018 年 7 月。

99. 视频中语义分割的有效不确定性估计

作者:黄宝玉,许万婷,赵春月,吴廷凡，孙敏

文摘: **深度学习**中的不确定性估计是近年来越来越重要的。如果我们不知道深度学习模型对决策是否确定，就不能将**其** 应用于实际应用。一些文献提出了贝叶斯神经网络，它可以通过蒙特卡洛德普特 (mc 降) 来估计不确定性。然而, mc 辍学需要转发模型 n 的时间，导致

n 次慢。对于自驾游系统等实时应用，需要尽快获得预测和不确定性，使 mc 辍学变得不切实际。在本文中，我们提出了基于区域的时间聚合 (rta) 方法，该方法利用**视频**中的时间信息来模拟采样过程。我们的 rta 方法与提拉米苏主干比 mc 辍学率快 10 倍 (n=5).此外，我们的 rta 方法获得的不确定性估计可与 mc 级和帧级指标的不确定性估计相比较。少

2018 年 7 月 29 日提交;**最初宣布** 2018 年 7 月。

100. 在全定向视频上弥合 vqa 与人类行为之间的差距：一个大规模的数据集和深度学习模型

作者:陈丽,徐麦梅,杜新哲,王祖林

抽象: 全方位**的视频**使球面刺激与 360x180 元。查看范围。同时，只有全向视频的视口区域可以被观察者通过头部运动 (hm) 看到，并且可以通过眼动 (em) 清楚地看到视口内更小的区域。因此，全方位**视频**的主观质量可能与人类行为的 hm 和 em 有关。为了填补主观质量与人类行为之间的空白，本文提出了一种大规模**的全向**视频视觉质量评估 (vqa) 数据集，称为 vqa-ov，收集了 60 个参考序列和 540 个受损序列。我们的 vqa-ov 数据集不仅提供序列的主观质量得分，还提供

受试者的 hm 和 em 数据。通过挖掘我们的数据集，我们发现全方位**视频**的主观质量确实与 hm 和 em 有关。因此，我们开发了一个嵌入 hm 和 em 的**深度** 学习模型，用于全方位**视频**上的目标 vqa。实验结果表明，该模型显著提高了 vqa 在全向视频上的最先进性能。少

2018 年 7 月 28 日提交;**最初宣布** 2018 年 7 月。

101. 赫尔姆霍兹方法：利用感知压缩降低机器学习复杂性

作者:杰拉尔德·弗里德兰,王景康,贾若西,李波

文摘: 本文对多媒体计算和机器**学习**中经常提出的问题提出了一个基本的答案：感知压缩产生的工件是否会导致机器**学习**过程中的错误，如果是，有多少？我们解决这个问题的方法是从物理角度重新解释赫尔姆霍兹自由能公式，以解释使用传感器 (如相机或麦克风) 捕获多媒体数据时的内容和噪声之间的关系。重新解释允许通过将分类器与感知压缩 (如 jpeg 或 mp3) 相结合，对图像、音频和**视频**中包含的噪声进行位测量。我们在 cifar-10 和弗劳恩霍夫的 idmt-smt-听觉效果数据集上的实验表明，在正确的质量水平下，感知压缩实

际上无害，但有助于显著降低机器学习的复杂性程。也就是说，我们的噪声量化方法可以显著加快**深度学习**分类器的训练，同时保持甚至提高整体分类精度。此外，我们的研究结果为**深入学习**成功的原因提供了真知灼见。少

2018 年 7 月 9 日提交;**最初宣布** 2018 年 7 月。

102. 深度: catactc 视频中的手术相位识别

作 者 :odase zisimopoulos, evangello flouty, imanol luengo , petros giataganas, jean nehme, andrechow, danail stoyanov

摘要: 自动手术工作流程分析和理解可以帮助外科医生标准化程序，加强手术后评估和索引，以及介入监测。基于**视频**的计算机辅助介入 (cai) 系统可以通过手术器械的识别来进行工作流估计，同时将其与程序阶段的本体联系起来。在这项工作中，我们采用**了一个深入的学习**范式来检测白内障手术**视频**中的手术器械，这反过来又为手术相位推断提供了一个复发网络，编码了其中阶段步骤的时间方面。阶段分类。我们的型号提供了与手术工具检测和相位识别的最新结果相当的结果，精度分别为 99% 和 78%。少

2018 年 7 月 17 日提交;最初宣布 2018 年 7 月。

103. unet ++: 用于医学图像分割的嵌套 u-net 体系结构

作者: 周宗伟, md mahfuzur rahman siddiquee, nima tajbakhsh, k 元明 liang

摘要: 在本文中，我们提出了一个新的，更强大的医学图像分割体系结构 unet ++。我们的架构本质上是一个深度监控的编码器解码器网络，在该网络中，编码器和解码器子网络通过一系列嵌套的密集跳过路径进行连接。重新设计的跳过路径旨在缩小编码器特征图与解码器子网络之间的语义间隙。我们认为，当来自解码器和编码器网络的要素映射在语义上相似时，优化器将处理更简单的**学习**任务。我们在多个医学图像分割任务中与 u-net 和宽 u-net 架构进行了比较，对其进行了评估: 胸部低剂量 ct 扫描中的结节分割、显微镜图像中的细胞核分割、腹部的肝脏分割 ct 扫描**和息肉分割在**结肠镜检查视频。我们的实验表明,**在深度**监控下, unet ++ 比 u-net 和宽 u-net 分别实现了 3.9 个百分点和 3.4 点的平均 iou 增益。少

2018 年 7 月 18 日提交;最初宣布 2018 年 7 月。

104. 视频游戏流亮点的深度无监督多视图检测

作者:charles ringer, mihalis a. nicoraou

文摘: 我们考虑**了视频**游戏流中的自动高光检测问题。目前，绝大多数游戏的高亮检测系统都是由硬编码游戏事件（例如，分数变化，最终游戏）的发生触发的，而大多数高级工具和技术都是基于通过视觉分析检测高光游戏画面。我们认为，在游戏流媒体的背景下，可能构成亮点的事件不仅取决于游戏画面，还取决于流媒体在游戏时段传达的社交信号（例如，在与观众互动时，或在评论时)和对游戏的反应)。在此基础上，我们提出了一种基于新颖性的高光检测的多视图无监督深度**学习**方法。该方法共同分析游戏画面和社交信号，如玩家的面部表情和讲话，并显示了有希望的结果，生成亮点的流行游戏流，如玩家的战场。少

2018 年 7 月 25 日提交;**最初宣布** 2018 年 7 月。

105. 使用对抗性摄动学习判别性视频表示

作者:王觉,阿诺普·切里安

摘要: 对抗性扰动是类似噪音的模式，可以巧妙地改变数据，同时无法使用精确的分类器。在本文中，我们建

议使用这样的扰动来提高**视频**表示的鲁棒性。为此，为每帧**视频**识别提供了一个训练有素的**深度**模型，我们首先生成适合此模型的对抗性噪声。利用完整**视频**序列中的原始数据功能及其不安的对应，作为两个独立的包，我们开发了一个二进制分类问题,**学习**一组判别超平面-作为一个子空间，这将把两个袋子分开。然后将此子空间用作**视频**的描述符，称为判别子空间池。由于扰动要素属于可能与原始要素混淆的数据类，判别子空间将描述功能空间中更能代表原始数据的部分，从而提供鲁棒性**视频**表示。为了**学习**这些描述符，我们在 stiefel 流形上制定了一个子空间**学习**目标，并采用黎曼优化方法来有效地解决它。我们在多个**视频**数据集上提供实验，并演示最先进的结果。少

2018 年 7 月 26 日提交;**v1** 于 2018 年 7 月 24 日提交;**最初宣布** 2018 年 7 月。

106. 利用因果关系的信息提出的可规划表示

作者:thanard k 鲁 utach, aviv tamar, ge yang , stuart russell, pieter abbeel

摘要: 近年来,**深度**生成模型已被证明是 "想象" 令人信服的高维观测，如图像，音频，甚至**视频**，直接从原始

数据**学习**。在这项工作中，我们问如何想象目标导向的视觉计划--一个合理的观测序列，将一个动态系统从其当前配置过渡到一个理想的目标状态，它后来可以作为控制的参考轨迹。我们专注于具有高维观测（如图像）的系统，并提出了一种自然结合表示学习和规划**的**方法。我们**的框架学习**了顺序观测的生成模型，其中生成过程是由低维规划模型中的过渡和额外的噪声引起的。通过在规划模型中最大限度地增加生成的观测值和转换之间的相互信息，我们获得了一个低维表示形式，可以最好地解释数据的因果性质。我们构建了与高效规划算法兼容的规划模型，并提出了几种基于离散或连续状态的规划模型。最后，为了生成一个可视化的计划，我们将当前和目标观测投影到它们各自的状态中，规划一个轨迹，然后使用生成模型将轨迹转换为一系列观测。我们展示了我们想象绳子操纵的合理视觉计划的方法。少

2018 年 7 月 24 日提交;**最初宣布** 2018 年 7 月。

107. 大眼区闭塞的身份保存面完成

作者:赵亚杰,陈伟凯,邢军，李晓明,泽赫·贝辛格，刘福昌, 左王蒙,杨瑞刚

摘要: 我们提出了一种新的**深度学习**方法，以合成完整的人脸图像在存在的大的眼部区域闭塞。这是由最近阻碍面对面交流的 vrar 显示器激增的推动。与最先进的面部涂装方法不同的是，我们的方法在保持特征的同时，能够忠实地恢复各种头部姿势下的缺失内容，而不是对合成内容的控制，只能处理正面面部姿势。我们的方法的核心是一个新的生成网络，具有专门的约束，以规范合成过程。为了保持身份，我们的网络采用任意的无遮挡图像的目标身份来推断缺失的内容，其高级 cnn 功能作为一个身份，然后再规范发电机的搜索空间。由于输入参考图像可能具有不同的姿态，因此进一步采用了姿态图和新的姿态判别器来监督隐式姿态变换的**学习**。我们的方法能够产生连贯的面部画与一致的身份在**视频**与大变化的头部运动。对合成数据和实际数据的实验表明，我们的方法在合成质量和鲁棒性方面都大大优于最先进的方法。少

2018 年 7 月 23 日提交;**最初宣布** 2018 年 7 月。

108. 相关网：时空多态式深度学习

作者:novanto yudistira, takio k 鲁 ita

摘要: 这封信描述了一个网络，它能够捕获任意时间戳上的时空相关性。该方案是在时空区域上作为一个互补的、扩展的网络运作的。近年来，多模态融合在**深度学习**中得到了广泛的研究。对于动作识别，空间和时间流是**深**卷积神经网络（cnn）的重要组成部分，但减少这两个流的过度拟合和融合的发生仍然是开放的问题。现有的融合方法是平均这两个流。为此，我们提出了一个具有香农融合的相关网络，以**学习**已经接受过培训的cnn。远程**视频**可能包含任意时间的时空相关性。这种相关性可以通过简单的完全连接的图层来捕获，从而形成相关网络。这被认为是对现有网络融合方法的补充。我们评估了我们在 ucf-101 和 hmdb-51 数据集上的方法，由此获得的准确性改进表明了多模态关联的重要性。少

2018 年 10 月 6 日提交;**v1** 于 2018 年 7 月 22 日提交;**最初宣布** 2018 年 7 月。

109. 视频分类的深层判别模型

作者:mohammad tavakolian, abdenour hadid

文摘: 本文提出了一种新的**基于视频**的场景分类深度**学习**方法。我们设计了一个异构**深度**判别模型（hdm），其

参数是通过使用高斯受限玻尔兹曼机器（grbm）以分层方式执行无监督的预训练来初始化的。为了避免相邻帧的冗余，我们提取帧内的时空变化模式，并使用稀疏立方体对称模式（scsp）稀疏地表示它们。然后，使用每个类的**视频**分别训练预初始化的 hddm，以**学习**特定于类的模型。根据所学习的类特定模型中的最小重构误差，采用加权投票策略进行分类。在两个动作识别数据集上对该方法的性能进行了广泛的评价;ucf101 和好莱坞 ii，以及三个动态纹理和动态场景数据集;dytex、yupen 和 maryland。实验结果表明，该方法在所有数据集上都能获得较好的性能。少

2018 年 7 月 22 日提交;**最初宣布** 2018 年 7 月。

110. 用于星际争霸 ii 的异步优势演员-批评剂

作者:basel alghanem, keerthana p g

文摘: **深度**强化**学习**，特别是异步优势 ac-纵食肉算法，已成功地应用于各种**电子**游戏中的超人性能。随着谷歌深度思维和暴雪娱乐公司提出的 pysc2 **学习**环境的发布，星际争霸 ii 是强化**学习**社区面临的新挑战。尽管是几个 ai 开发人员的目标，但很少有人实现了人的级别性能。在这个项目中，我们解释了这种环境的复杂性，

并讨论了我们在环境实验中的结果。我们比较了各种架构，证明了转移学习可以成为需要技能转移的复杂场景中强化学习研究的有效范式。少

2018 年 7 月 21 日提交;最初宣布 2018 年 7 月。

111. 对象检测器的物理对抗示例

作 者 :kevin eykholt, ivan evtimov, earlence fernandes, bo li, amir rahmati, florian tramer, atul prakash, tadayoshi kohno, dawn song

摘要: 深层神经网络 (dnn) 容易受到对抗例--恶意制作的输入的影响，这些输入导致 dnn 做出不正确的预测。最近的研究表明，这些攻击概括到物理域，在各种真实条件下愚弄图像分类器的物理对象上产生扰动。这类攻击对安全关键型网络物理系统中使用的深度学习模型构成了风险。在这项工作中，我们将物理攻击扩展到更具挑战性的对象检测模型，这是一种更广泛的深度学习算法，广泛用于检测场景中的多个对象并对其进行标记。在改进以前对图像分类器的物理攻击的基础上，我们创建了被对象检测模型忽略或错误标记的扰动物理对象。我们实施了一个失踪攻击，在其中我们导致停止标志 "消失" 根据侦探，要么覆盖该标志与敌对停止

标志海报，或添加敌对贴纸上的标志。在受控实验室环境中录制的**视频**中，最先进的 YOLOv2 探测器未能在 85% 以上的**视频**帧中识别出这些对抗性停止标志。在户外实验中，yolo 分别在 72.5 和 63.5 的**视频**帧中被海报和贴纸攻击所愚弄。我们还使用更快的 r-cnn，一个不同的目标检测模型，以证明我们的对抗性扰动的可转移性。创建的海报摄动能够愚弄更快 r-处在 85.9% 的**视频**帧在受控的实验室环境中，和 40.2 的**视频**帧在室外环境中。最后，我们提出了一个新的创造攻击的初步结果，其中无害的物理贴纸愚弄模型，以检测不存在的对象。少

2018 年 10 月 5 日提交;**v1** 于 2018 年 7 月 20 日提交;**最初宣布** 2018 年 7 月。

112. 骨骼运动到彩色地图：一种新的三维动作识别的表示形式和初始剩余网络

作 者 :huy hieu pham, louahdi khoudour, alain crouzil, pablo zegers, sergio a. velastin

文摘: 我们提出了一种新的基于骨架的基于**骨架**的表示三维行动识别在**视频**中使用深卷积神经网络 (d-nn)。两个关键问题已经得到了解决：第一，如何构造一个强大

的表示，很容易捕获从骨架序列的运动的时空演化。其次，如何设计能够有效地从新的表现中**学习**判别特征的 d-nns。为了解决这些任务，提出了一种基于骨架的表示形式，即 spmf (骨架后运动特征)。spmf 是由人类行动的两个最重要的属性构成的: 姿势和它们的动作。因此，它们能够有效地代表复杂的行动。对于**学习**和识别任务，我们根据初始残差网络的理念设计和优化新的 d-nn，以预测 spmf 的操作。我们的方法在两个具有挑战性的数据集上进行了评估，包括 msr action3d 和 ntu-rgb + d. 实验结果表明，该方法超越了最先进的方法，同时需要较少的计算。少

2018 年 7 月 18 日提交;**最初宣布** 2018 年 7 月。

113. 基于深度表达式的着色

作者:何明明,陈东东,廖静,佩德罗德·桑德,陆元

文摘: 我们提出了第一个基于范例的局部着色的**深度学习**方法。给定参考彩色图像，我们的卷积神经网络直接将灰度图像映射到输出彩色图像。我们的端到端着色网络学习如何从大规模数据中选择、传播和预测颜色，而不是像传统的基于示例的方法那样使用手工制作的规则。即使在使用与输入灰度图像无关的参考图像时，该

方法也能实现可靠的推广。更重要的是，与其他基于**学习的**着色方法相比，我们的网络允许用户通过简单地提供不同的参考来实现可定制的结果。为了进一步减少选择参考的人工工作量，系统自动推荐参考与我们提出的图像检索算法，其中考虑语义和亮度信息。只需选择最重要的参考建议，即可完全自动执行着色。我们的方法通过用户研究和与最先进的方法进行有利的定量比较得到验证。此外，我们的方法可以自然地扩展到**视频**着色。我们的代码和型号可供公众免费使用。少

2018 年 7 月 21 日提交;**v1** 于 2018 年 7 月 17 日提交;**最初宣布** 2018 年 7 月。

114. 使用深度确定性策略梯度的双足行走机器人

作者:arun kumar, navneet paul, s n omkar

文摘: 机器**学习**算法在机器人和控制系统领域有几个应用。控制系统社区已经开始对来自监督**学习**、模仿**学习**和强化**学习**等子领域的几种机器**学习**算法表现出兴趣。实现自主控制和智能决策。在许多复杂的控制问题中，稳定的两步行走一直是最具挑战性的问题。在本文中，我们提出了一个架构，设计和模拟平面双足行走机器人 (bwr) 使用一个现实的机器人模拟器，凉亭。机器人

通过学习它的几个试验和错误来**展示**成功的行走行为，而事先不知道自己或世界的动态。采用强化**学习**算法(ddpg) 实现了 bwr 的自动行走。ddpg 是连续动作空间中**学习**控制的算法之一。在仿真训练模型后，观察到，通过适当的形状奖励功能，机器人实现了更快的行走速度，甚至实现了平均速度为 0.83 m/的跑步步态。将双足步行者的步态模式与实际的人类行走模式进行了比较。结果表明，两足行走模式与人类行走模式具有相似的特点。介绍我们实验的**视频**可在 https://goo.gl/NHXKqR.

2018 年 7 月 17 日提交;**v1** 于 2018 年 7 月 16 日提交;**最初宣布** 2018 年 7 月。

115. 利用深层神经网络评估水下视频中的鱼类丰度

作 者 :ranju mandal, rod m. connolly, thomas a. schlacherz , bela stantic

摘要: 海洋生物学家正在迅速采用水下**视频**来评估鱼类的多样性和丰度。人工处理**视频**，供人类分析师量化是耗时和劳动密集型的。**视频**的自动处理可以用来实现目标的成本和时间的方式。其目的是建立一个准确可靠的鱼类探测和识别系统，这对于一个自主的机器人平台

非常重要。然而，这项任务涉及许多挑战（例如复杂的背景、变形、低分辨率和光传播）。**近年来**深度神经网络的发展导致了目标检测和识别在实时场景中的发展。介绍了一种端**到端深度学习**的体系结构，该体系表现优于最先进的方法，也是一种在鱼类评估任务中的首创。由一个名为"更快 r-cnn"的物体探测器引入的区域提案网络（rpn）与三个分类网络相结合，用于检测和识别从偏远的水下**视频**站（ruvs）获得的鱼类物种。从实验中得到的 82.4 (map) 的精度远远高于以前提出的方法。少

2018 年 7 月 16 日提交;**最初宣布** 2018 年 7 月。

116. 自监督学习中的交叉像素光流相似性

作 者 : Aravindh mahendran，james thewlis，andrea vedaldi

文摘: 提出了一种在没有人工监控的情况下**学习**卷积神经图像的新方法。我们以光流的形式使用运动线索来监督静态图像的表示。由于该预测任务中固有的模糊性，训练网络以预测单个图像流量的明显方法可能会遇到不必要的困难。相反，我们提出了一个简单得多的**学习**目标：嵌入像素，使它们的嵌入之间的相似性与它们的

光流向量之间的相似性相匹配。在测试时,**学习的深网络**可以在不访问视频或流量信息的情况下使用,并传输到图像分类、检测和分割等任务。我们的方法大大简化了以前使用运动进行自我监督的尝试，在使用运动暗示的自我监督方面取得了最先进的结果，一般的自我监督有竞争的结果，是最先进的整体状态在语义图像分割的自我监督的预训练中，如标准基准所示。少

2018 年 7 月 15 日提交;**最初宣布** 2018 年 7 月。

117. 深度学习中的目标检测：综述

作者:赵忠秋,郑鹏,徐守涛,吴新东

文摘: 由于目标检测与**视频**分析和图像理解的密切关系,近年来引起了广泛的研究。传统的对象检测方法建立在手工制作的特征和可扩展的浅层体系结构之上。通过构造复杂的组合，将多个低级图像特征与对象探测器和场景分类器的高级上下文结合起来，它们的性能很容易停滞。随着**深度学习**的快速发展，引入了更强大的工具来解决传统体系结构中存在的问题，这些工具能够**学习**语义、高层次、更深层次的功能。这些模型在网络体系结构、训练策略和优化功能等方面的行为不同。本文对基于**深度学习**的目标检测框架进行了综述。我们的

回顾首先简要介绍了**深度学习**的历史及其代表性工具，即卷积神经网络 (cnn)。然后，我们重点介绍了典型的通用对象检测体系结构以及一些修改和有用的技巧，以进一步提高检测性能。由于不同的特定检测任务表现出不同的特征，我们还简要介绍了几个具体的任务，包括突出的目标检测、人脸检测和行人检测。并进行了实验分析，比较了各种方法，得出了一些有意义的结论。最后，提出了几个有希望的方向和任务，作为今后对象检测和相关神经网络**学习**系统工作的指导。少

2018 年 7 月 15 日提交;**最初宣布** 2018 年 7 月。

118. 人的重新识别任务深度学习技术综述

作者:bahram lavi, Ihsan fatan serj,ihsan ullah

文摘: 智能**视频**监控是目前计算机视觉和机器**学习**技术的一个活跃的研究领域。它为监控操作员和法医**视频**调查人员提供了有用的工具。人员重新识别 (preid) 是这些工具之一。它包括识别一个人是否已经在网络中通过相机被观察到。该工具还可用于各种可能的应用，如离线检索所有显示图像被查询的感兴趣的个人的视频序列，以及在线行人跟踪多个摄像机视图。为此，提出了许多提高 preid 性能的技术。在这些系统中，许多研究

人员使用了深度神经网络 (dnn)，因为它们在测试时性能更好，执行速度快。我们的目标是为未来的研究人员提供迄今为止在 piid 上所做的工作。因此，我们总结了用于此任务的最先进的 dnn 模型。给出了每个模型的简要描述及其对一组基准数据集的评估。最后，对这些模型进行了详细比较，然后提出了一些限制，可作为今后研究的指导方针。少

2018 年 7 月 19 日提交;**v1** 于 2018 年 7 月 13 日提交;**最初宣布** 2018 年 7 月。

119. 采样湍流去除网络

作者:伟和泽,春邦楼,乐明雷

文摘: 我们提出了一种深度学习的方法，从湍流变形和时空变化的模糊中恢复一系列湍流扭曲的视频帧。我们设计了一种基于新的数据增强方法的训练策略，而不是在深层网络中要求大量的训练样本大小，以模拟来自相对较小的数据集的湍流。然后介绍了一种子采样方法，以提高所提出的 gan 模型的恢复性能。本文的贡献有三：首先，我们介绍了一种简单而有效的数据扩充算法，对深部网络训练的现实生活中的动荡进行建模；第二，我们首先将 wasserstein gan 与我 1 成功恢复

损坏的**视频**序列的成本;第三，结合子采样算法，筛选出损坏严重的帧，生成质量较好的视频序列。少

2018 年 8 月 13 日提交;**v1** 于 2018 年 7 月 12 日提交;**最初宣布** 2018 年 7 月。

120. 场景 ednet: 场景流估计的深层学习方法

作者:ravi kumar thakur, nehasis mukherjee

文摘: 对 rgb-d **视频**中的场景流进行估计，由于其在机器人技术中的潜在应用，引起了计算机视觉研究人员的极大兴趣。最先进的场景流估计技术通常依赖于帧的场景结构知识和帧之间的对应关系。然而，随着越来越多的 rgb-d 数据从复杂的传感器，如微软 kinect，以及最近在先进的**深度学习**技术领域的进展，引入了一个高效的**深**场景流估计的**学习**技术越来越重要。本文首先介绍了采用深度学习方法直接**估计**场景流的方法，提出了一种**具有编码**器解码器 (ed) 体系结构的完全卷积神经网络。提出的网络场景 ednet 涉及从立体图像序列中估计所有场景点的三维运动矢量。直接估计场景流的训练是利用连续对立体图像和相应的场景流地面真相进行的。建议的体系结构应用于巨大的数据集，并提供有意义的结果。少

2018 年 7 月 9 日提交;**最初宣布** 2018 年 7 月。

121. youtube 的患者教育：从用户生成的视频中了解医学知识的深层学习方法

作者:小刘,张斌,安雅娜·苏萨拉,雷玛·帕德曼

摘要: youtube 提供了一个前所未有的机会，探讨机器**学习**方法如何改善医疗信息的传播。我们提出了一个跨学科的镜头，综合机器**学习**方法与医疗信息学主题，以解决开发一个可扩展的算法解决方案，以评估视频从健康知识和患者教育的视角。我们开发了一种**深度学习**方法来了解 youtube **视频**中编码的医学知识水平。初步结果表明，我们可以从 youtube **视频**中提取医学知识，并根据嵌入的知识对**视频**进行分类，性能令人满意。**深部学习**方法在知识提取、自然语言理解和图像分类方面显示出很大的希望，特别是在以病人为中心的护理和精确医学的时代。少

2018 年 7 月 6 日提交;**最初宣布** 2018 年 7 月。

122. 深度全球互联网络与广义多点重排网络在深层学习中的激活

作者:陈志贤,何平汉

摘要: 最近的进展表明，利用卷积神经网络中的隐藏层神经元结合精心设计的激活函数，可以在计算机视觉领域产生更好的分类效果。本文首先介绍了一种新的**深度学习**体系结构，旨在缓解逐渐消失的问题，即早期的隐藏层神经元可以与最后一个隐藏层直接连接，输入到最后一个隐藏层图层进行分类。然后设计了广义线性整流器函数作为激活函数，通过对参数的训练来逼近任意复杂函数。我们将展示，我们的设计可以在一些对象识别和**视频**动作基准任务中实现类似的性能，在参数明显减少和网络基础设施较浅的情况下，这不仅在在计算负担和内存使用方面进行培训，但也适用于低计算、低内存的移动方案。少

2018 年 6 月 19 日提交;**最初宣布** 2018 年 7 月。

123. nmt-keras：一个非常灵活的工具包，专注于交互式 nmt 和在线学习

作者:alvaro peris, francisco casacuberta

摘要: 我们提出了一个灵活的工具包 nmt-keras，用于培训**深度学习**模型，它特别强调开发神经机器翻译系统的高级应用，如交互预测翻译协议和翻译系统的长期适应通过不断**的学习**。nmt-keras 是基于流行的

keras 库的扩展版本，它在 theano 和 tensorflow 上运行。按照 keras 提供的高级框架，部署和使用最先进的神经机器翻译模型。由于其高度的模块化和灵活性，它也被扩展到解决不同的问题，如图像和**视频**字幕，句子分类和视觉问题回答。少

2018 年 8 月 16 日提交;**v1** 于 2018 年 7 月 9 日提交;**最初宣布** 2018 年 7 月。

124. 深度强化学习的分类视频总结

作者:周开阳,陶翔,安德烈·卡瓦拉罗

摘要: 大多数现有**的视频**摘要方法都是基于监督或非监督**学习**。本文提出了一种基于强化学习的弱监督**方法**，利用易于获取的**视频**级类别标签，并鼓励摘要包含与分类相关的信息和维护类别可识别性。具体而言，我们将**视频**摘要作为一个顺序决策过程，并培训一个带有**深度 q 学习**(dqsn) 的摘要网络。还对配套分类网络进行了培训，为培训 dqsn 提供奖励。利用分类网络，在分类结果的基础上，开发了全局可识别性奖励。关键的是，为了应对长序列强化学习的时间延迟**和稀疏奖励问题**，还提出了一种新的密集排名奖励。在两个基准数

据集上进行的大量实验表明，该方法具有最先进的性能。少

2018 年 9 月 3 日提交;**v1** 于 2018 年 7 月 9 日提交;**最初宣布** 2018 年 7 月。

125. 从最小摄像机视点进行体积性能捕获

作 者 :andrewgilbert, marco volino, john collomosse, adrian hilton

摘要: 我们提出了一个卷积自动编码器，使人类性能的高保真体积重建能够从仅包含一小部分摄像机视图的多视图视频中捕获。我们的方法产生的端到端重建误差与使用更多 (双或更多) 视点计算的概率视觉船体的重中误差相似。我们使用由在广泛的主题和操作的视光多视视频素材数据集上训练的自动编码器隐式学习的**深度**先验.这就为高端体积性能捕获提供了可能性，在这种情况下，在设置和数据中，时间或成本禁止高见证摄像机数量。少

2018 年 7 月 10 日提交;**v1** 于 2018 年 7 月 5 日提交;**最初宣布** 2018 年 7 月。

126. 和：自回归新颖性检测器

作　者:davandabati, angelo porrello, simone calderara, rita cucchiara

文摘: 我们提出了一个无监督的新颖性检测模型。该主题被视为密度估计问题，其中使用**深层**神经网络来**学习**参数函数，最大限度地提高训练样本的概率。这是通过为自动编码器配备一个新的模块来实现的，该模块负责通过自回归实现压缩代码的可能性最大化。我们说明了设计选择和适当的层，以便在处理图像和**视频**输入时执行自回归密度估计。尽管有一个非常通用的公式，我们的模型显示了在不同的单级新颖性检测和**视频**异常检测基准有希望的结果。少

2018 年 7 月 4 日提交;**最初宣布** 2018 年 7 月。

127. 即插即用深部线性嵌入视频帧插值

作　者:anh-duc nguyen, woojae kim, jongyoo kim, sanghoon lee

文摘: 提出了一个考虑**视频**帧插值问题的生成框架。我们的框架，我们称之为**深**局部线性嵌入 (deeplle)，由**一个深**卷积神经网络 (cnn) 供电，而它可以立即使用像传统模型。deeplle 将自动编码的 cnn 设置为一组连续的帧，并在潜在代码上嵌入线性约束，以便通过插

入新的潜在代码来生成新的帧。与当前需要大型数据集培训的**深度学习**范式不同, deeplle 以即插即用和无监督的方式工作，并且能够生成任意数量的帧。彻底的实验表明，如果没有铃声和口哨，我们的方法在目前最先进的型号中具有很强的竞争力。少

2018 年 7 月 4 日提交;**最初宣布** 2018 年 7 月。

128. 模拟飞行形状和模拟平面机械手数据集的介绍

作者:fabio ferreira, jonas rothfuss, eren erdal aksoy, you zhou, tamim asfour

文摘: 我们发布了两个人工数据集，模拟飞行形状和模拟平面机械手，允许测试**视频**处理系统的**学习**能力。特别是，该数据集是指一种工具，可以轻松地评估深度神经网络模型的**理智**，旨在对视频帧序列进行编码、重建或预测.每个数据集由 90000 **视频**组成。"模拟飞行形状" 数据集包括显示两个形状相等 (矩形、三角形和圆形)和大小的对象的场景，其中一个对象接近其对应对象。模拟平面机械手显示了一个 3 自由度的平面机械手，它执行一个拾取和放置任务，在这个任务中，它必须在平方平台上放置一个大小变化的圆圈。与移动 mnist [1]、[2] 等其他广泛使用的数据集不同，两个呈现的数据集

涉及面向目标的任务 (例如，操纵者抓取对象并将其放置在平台上)，而不是显示随机移动。这使得我们的数据集更适合于测试预测能力和通过机器**学习**模型**学习**复杂的运动。本技术文档旨在介绍这两个数据集的使用情况。少

2018 年 7 月 2 日提交;**最初宣布** 2018 年 7 月。

129. 基于外观的 3d 凝视估计与个人校准

作者:erik lindén , jonas sjöstrand,亚历山大 proutiere

摘要: 我们提出了一种将个人校准整合到基于**视频**的凝视估计深度**学习**模型中的**方法**。使用我们的方法，我们表明，通过校准每人六个参数，精度可以提高 2.2 到 2.5。个人参数的数量，每只眼睛三个，与几何模型预测的数量相似。在 miigaze 数据集上进行评估时，我们的估计器的性能优于特定于人的估计。为了提高泛化能力，我们预测三维凝视光线 (凝视的起源和方向)。在现有数据集中，由于所有凝视目标都与相机位于同一平面上，因此 3d 凝视距离确定不足。对合成数据的实验表明，只有带注释的凝视目标，而没有注释的眼睛位置,才有可能获得精确的三维凝视。少

2018 年 7 月 2 日提交;**最初宣布** 2018 年 7 月。

130. 逼真的视频风格传输

作者:michael honke, rahul iyer, Dishant mittal

摘要:光逼真风格转换是利用**深度学习**和优化技术将色彩从一个参考域转移到另一个参考域的技术。在这里，我们展示了一种技术，我们使用它将样式和颜色从参考图像传输到**视频**。

2018 年 7 月 1 日提交;**最初宣布** 2018 年 7 月。

131. 通过视频分析利用服务业有限的资源

作者:郑春鸿,伊约拉 e. Olatunji

摘要: 服务业对许多发达经济体和发展中经济体作出了重大贡献。随着业务活动的迅速扩大，由于资源短缺导致服务反应迟缓，许多服务公司难以保持客户的满意度。在资源短缺和解决方案发生之前就对其进行预测是减少对运营的不利影响的有效方法。然而，就能力和劳动力成本而言，这种积极主动的做法非常昂贵。许多公司陷入生产力难题，因为它们未能找到足够有力的论据来证明新技术的成本是合理的，但却不能不投资于新技术，以与竞争对手相匹配。问题是，是否有创新的解决办法来最大限度地利用现有资源，并大幅减少资

源短缺可能导致但以低成本实现高水平服务质量的影响。这项工作通过对我们在香港国际机场 (hkia) 设计和部署的手推车跟踪系统的实际分析，说明**视频**分析如何帮助实现管理层通过实时满足客户需求的目标使用现有**视频技术**而不是采用新技术，检测和预防他们在服务使用过程中可能遇到的问题。本文介绍了商业**视频**监控系统与**视频**分析深度**学习**算法的集成。结果表明，系统能在面对全部或部分遮挡时提供准确的决策，精度高，显著改善了日常操作。根据设想，这项工作将提高服务业资源管理综合技术的认识，并将其作为实时客户援助的措施。少

2018 年 6 月 30 日提交;**最初宣布** 2018 年 7 月。

132. 基于多视图动态图像的深度视频动作识别

作者:杨晓,陈军,曹志国,周天一岳,香白

摘要: 动态图像是最近出现的能够准确捕捉时间演化的动作表示范式，特别是在**深**卷积神经网络 (cnn) 的背景下。灵感来自于 rgb **视频**的初步成功，我们建议将其扩展到深度域。为了更好地利用深度**视频**的三维特征来利用其性能，提出了多视图动态图像。特别是，原始深度视频将通过围绕 3d 空间中的特定实例旋转虚拟摄像

机，密集**地**投射到不同的成像视点上。然后分别从生成的多视图深度**视频**中提取动态图像，构成多视图动态图像。这样，多视图动态图像中的视域代表性信息就可以比单视图图像涉及更多的视点代表性信息。提出了一种新的 cnn **学习**模型，在多视图动态图像上进行特征**学习**。来自不同视图的动态图像将共享相同的卷积图层，但具有不同的完全连接图层。该模型旨在通过缓解梯度消失来增强浅卷积层的调谐。此外，为了解决空间变化的影响，提出了一种基于更快 r-napn 的行动建议方法。动态图像将仅从行动建议区域中提取。在实验中，我们的方法可以在 3 个具有挑战性的数据集（即 ntu rgb-d、西北-ucla 和 UWA3DII）上实现最先进的性能。少

2018 年 8 月 4 日提交;**v1** 于 2018 年 6 月 29 日提交;**最初宣布** 2018 年 6 月。

133. 通过程序层次生成在深层强化学习中的照明推广

作 者 :niels justesen, ruben rodriguez torrado, phillip bontrager, ahmed khalifa, julian togelius, sebastian risi

文摘: **深度强化学习**(rl) 在多个领域都显示出令人印象深刻的成果，直接从高维感官**流中学习.**然而，当神经

网络在固定的环境中训练时，比如**电子**游戏中的一个级别，它们通常会过度适应，无法推广到新的级别。当 rl 模型过度适用时，即使对环境稍作修改，也会导致代理性能不佳。在本文中，我们探讨了培训过程中程序生成的水平如何增加通用性。我们表明，对于某些游戏，过程级别生成使泛化到相同分布中的新级别。此外，通过根据代理的性能操作级别的难度，可以在数据较少的情况下实现更好的性能。**学习行为**的普遍性也在一组人类设计的层面上进行了评价。我们的研究结果表明，推广到人类设计水平的能力在很大程度上取决于水平发电机的设计。我们应用维数约简和聚类技术来可视化发电机的水平分布，并分析它们在多大程度上可以产生与人类设计的水平相似的水平。少

2018 年 9 月 7 日提交;**v1** 于 2018 年 6 月 27 日提交;**最初宣布** 2018 年 6 月。

134. 社交直播服务中的成人内容：描述异常用户和关系

作者: Lykousas lykousas, Vicenç gómez, constantinos patsakis

摘要: 社会直播流服务（slss）利用了一个新的社会互动水平。这些服务的主要挑战之一是如何发现和防止违

反社区准则的异常行为。在这项工作中，我们重点关注成人内容的生产和消费在两个广泛使用的 slss，即 live. me 和循环 live，其中有数百万用户每天生产大量**的视频**内容。我们使用预先培训的**深度学习**模型来识别成人内容的广播机构。我们的研究结果表明，现有的适度系统在暂停此类用户的账户方面非常无效。我们通过爬网这些平台的社交图形来创建两个大型数据集，我们对这些图形进行分析，以确定成人内容制作者和消费者的特征，并发现它们之间有趣的关系模式，这在这两个网络中都很明显。少

2018 年 6 月 27 日提交;**最初宣布** 2018 年 6 月。

135. 每个像素计数：具有整体 3d 运动理解的无监督几何学习

作者:杨振恒,王鹏,王洋，徐伟,拉姆·内马提亚

摘要: **通过深**卷积网络观看未标记**的视频**来学习估计单个图像中的 3d 几何，是近年来的重要过程。目前最先进的 (sota) 方法是基于刚性结构自运动的**学习**框架，在这种学习框架中，只有三维相机自我运动被建模为几何估计。然而，移动的物体也存在于许多**视频**中，例如在街道场景中移动汽车。本文通过将每个像素的三维

物体运动另外集成到**学习**框架中来解决这种运动问题，该框架提供了全面的三维场景流理解，并有助于单图像几何估计。具体而言，给定**视频**中的两个连续帧，我们采用一个运动网络来预测它们的相对 3d 相机姿势和分割掩码识别运动物体和刚性背景。光流网络用于估计密集的 2d 每像素对应。单个图像深度网络预测这两种图像的深度贴图。四种类型的信息，即 2d 流、相机姿态、分段掩码和深度贴图，被集成到一个可区分的整体三维运动解析器 (hmp) 中，在该解析器中恢复刚性背景和运动物体的每个像素 3d 运动。我们设计了用于训练深度和运动网络的两种类型的三维运动，从而为估计的几何形状进一步减少误差。最后，为了解决单目**视频**的三维运动混乱，我们将立体图像结合到关节训练中。在 kitti 2015 数据集上进行的实验表明，我们估计的几何形状、3d 运动和移动物体掩码不仅受到一致的限制，而且明显优于其他 sota 算法，这证明了我们的方法的优势。少

2018 年 8 月 15 日提交;**v1** 于 2018 年 6 月 27 日提交;**最初宣布** 2018 年 6 月。

136. 基于联合学习时间结构与空间细节的视频绘制

作者:王川,黄海斌,韩晓光,王觉

文摘: 提出了一种新的数据驱动**视频**绘制方法来恢复**视频**帧缺失区域。提出了一种新的**深度学习**体系结构，它包含两个子网络：时间结构推理网络和空间细节恢复网络。时间结构推理网络是建立在三维完全卷积结构的基础上的：考虑到三维卷积的昂贵计算成本，它才**学会**完成低分辨率**的视频**量。低分辨率结果为空间细节恢复网络提供了时间指导，该网络采用 2d 完全卷积网络进行基于图像的绘制，以原始分辨率生成恢复的**视频**帧。这种两步网络设计既确保了每个帧的空间质量，又确保了帧之间的时间一致性。我们的方法以端到端的方式联合培训这两个子网络。我们对三个数据集进行定性和定量评估，证明我们的方法优于以前**基于学习**的**视频**绘制方法。少

2018 年 6 月 21 日提交;**最初宣布** 2018 年 6 月。

137. 网络档案图像数据中人的关系研究

作者:eric müller-budack, kader Pustu-Iren, sebastian diering, ralph ewerth

摘要: 万维网上的多媒体内容正在迅速发展，包含了不同领域许多应用的宝贵信息。因此，自 90 年代中期以来，互联网档案计划已经收集了数十亿的过时的网页。

但是，大量数据很少用适当的元数据标记，并且需要自动方法来启用语义搜索。通常情况下，互联网档案的文本内容被用来提取实体及其可能的关系跨领域，如政治和娱乐，而图像和**视频**内容通常被忽视。本文介绍了一种存储在互联网档案中的网络新闻图像内容中的人识别系统。因此，该系统补充了文本中的实体识别，使研究人员和分析人员能够更准确地跟踪媒体报道和人员关系。基于**深度学习**人脸识别方法，我们建议建立一个系统，自动检测感兴趣的人，并收集样本材料，随后用于在互联网档案的图像数据中识别他们。我们在适当的标准基准数据集上评估人脸识别系统的性能，并通过两个用例演示该方法的可行性。少

2018 年 6 月 21 日提交;**最初宣布** 2018 年 6 月。

138. 头发网：利用卷积神经网络进行单视头发重建

作者:周毅,胡立文,邢军,陈伟凯, 孔汉伟,新通,郝丽

文摘: 我们引入了一种基于**深度学习**的方法，从无约束图像生成完整的 3d 头发几何。我们的方法可以恢复本地链的细节，并具有实时性能。最先进的发型建模技术依靠大型发型集合进行最近的邻居检索，然后进行临时细化。相比之下，我们**的深度学习**方法在存储方面效

率很高，在用 30k 股发发的同时，可以以 1000 倍的速度运行。卷积神经网络以头发图像的二维方向场作为输入，生成均匀分布在参数化二维头皮上的链状特征。我们引入碰撞损失来合成更合理的发型，每条线的可见性也被用作一个权重项，以提高重建精度。我们网络的编码解码器架构自然为发型提供了紧凑而连续的表示形式，使我们能够在发型之间自然插值。我们使用大量的渲染合成头发模型来训练我们的网络。我们的方法可以缩放到真实图像，因为一个中间的二维定向场，从真实图像自动计算，计算出合成和真实头发之间的差异。我们展示了我们的方法的有效性和鲁棒性，在广泛的具有挑战性的真实互联网图片，并显示从**视频**重建的头发序列。少

2018 年 7 月 10 日提交;**v1** 于 2018 年 6 月 19 日提交;**最初宣布** 2018 年 6 月。

139. 重复估计

作者:tom f. h.runia , cees g.m . snoek, arnold w. m. 斯梅尔斯

摘要: 视觉重复在我们的世界里无处不在。它出现在人类活动 (运动、烹饪)、动物行为 (蜜蜂的摇摆舞)、自然

现象（风中的叶子）和城市环境（闪烁的灯光）中。从逼真**的视频**估计视觉重复是具有挑战性的，因为周期运动很少是完全静态和静止的。为了更好地处理现实**的视频**，我们提升了现有工作经常做出的静态和静态假设。我们的时空滤波方法建立在周期运动理论的基础上，有效地处理了各种各样的外观，不需要**学习**。从三维运动开始，我们通过将运动场分解为其基本组成部分，推导出三种周期运动类型。此外，从场的时间动力学中产生了三个时间运动连续性。对于三维运动的二维感知，我们考虑相对于运动的视点;以下是 18 例反复运动感知。为了估计在任何情况下的重复，我们的理论意味着构造微分运动图的混合：梯度，发散和卷曲。我们利用小波滤波器对运动图进行时间上的卷积，以估计重复的动态。我们的方法能够直接从在运动图上密集计算的时间滤波器响应进行空间分割重复运动。为了对我们的权利进行实验验证，我们使用我们的新数据集进行重复估计，用非静态和非平稳的重复运动更好地反映现实。在重复计数的任务上，与**深度学习**的替代方案相比，我们获得了良好的效果。少

2018 年 6 月 18 日提交;**最初宣布** 2018 年 6 月。

140. 深部神经网络与认知体系结构相结合的语义图像检索

作者:阿列克谢·波塔波夫,因诺肯蒂伊·日丹诺夫, oleg scherbakov, nikolai skorobogoko, hugo Potapov, enzo fenoglio

摘要: 多年来，以语义内容进行图像和**视频**检索一直是一项重要而具有挑战性的任务，因为它最终需要弥合符号间的鸿沟。最近在**深度学习**方面取得的成功使人们能够检测到属于许多类的对象，这大大优于传统的计算机视觉技术。但是，仍然无法提供能够执行检索查询的**深度学习**解决方案。我们提出了一个混合解决方案，包括一个深度神经网络的**对象**检测和认知架构的查询执行。具体来说，我们使用 YOLOv2 和 opencog。实现了允许检索包含指定类和指定空间排列的**对象的视频**帧的查询。少

2018 年 6 月 14 日提交;**最初宣布** 2018 年 6 月。

141. 通过解释预测来理解基于修补程序的学习

作 者 :christopher anders, grégoire montavon , wojciech samek, klaus-robert müller

摘要: **深**网络能够**学习**视频数据的高度预测模型 。由于**视频**长度的原因，一个常见的策略是在小**视频**片段上对他们进行训练。我们应用**深度**泰勒/lrp 技术来理解**深**部网络的分类决策，并识别 "边界效应": 分类器主要看输入的边框的趋势。此效果与用于构建**视频**代码段的步骤大小有关，然后我们可以对其进行调整，以提高分类器的准确性，而无需对模型进行再培训。据我们所知，这是首次将**深度**泰勒/lrp 技术应用于任何**视频**分析神经网络。少

2018 年 6 月 11 日提交;**最初宣布** 2018 年 6 月。

142. 玩第一人称射击游戏的任务相关对象发现和分类

作者:junchi liang, abdeslam boularias

摘要: 我们认为**学习**玩第一人称射击游戏 (fps)**电子**游戏的问题，使用原始屏幕图像作为观察和键盘输入作为行动。在这种类型的应用中，观测结果的高维性导致对无模型方法 (如**深** q 网络 (dqn) 及其反复的变量 drqn) 的训练数据的需求令人望而却步。因此，最近的工作重点是**学习**低维表示，这可能会减少对数据的需求。本文提出了一种新的、有效的学习这种表示的方法。从连续帧的光流中检测到连续帧的显著片段，并根据

其特征描述符进行聚集。群集通常对应于不同的已发现对象类别。然后根据最近的群集对在新帧中检测到的段进行分类。因为只有少数类别与给定的任务相关，所以类别的重要性被定义为其发生与代理性能之间的相关性。结果被编码为一个向量，指示框架中的对象及其位置，并用作 drqn 的侧输入。游戏 "末日" 上的实验为这种方法的好处提供了很好的证据。少

2018 年 6 月 17 日提交;**最初宣布** 2018 年 6 月。

143. 生成性抗性网络与视频超解析的感知损失

作者:alice lucas, santiago lopez tapia , rafael molina , Aggelos k. ktsaggelos

摘要: **视频**超分辨率已成为**视频处理中**最关键的问题之一。在**深层学习**文献中，最近的研究表明了利用感性损失来提高各种图像恢复任务的性能的好处;然而，这些还没有被应用于**视频**超分辨率。在这项工作中，我们提出了使用一个非常**深**的残余神经网络 vsrresnet，以执行高质量的**视频**超分辨率。我们表明，与大多数比例因子的 psnr/ssim 指标相比, vsrresnet 超过了目前最先进的 vsr 模型。此外，我们还通过对抗、特征空间和像素空间损耗的凸组合来训练这种体系结构，以获得

vsrres 特点 gan 模型。最后，我们使用 psnr、ssim 和一个新的感知距离度量（感知度量）将生成的 vsr 模型与当前最先进的模型进行了比较。使用后一个指标，我们表明 vsrres 特点 gan 在数量和质量上都优于当前最先进的 sr 模型。少

2018 年 6 月 14 日提交;**最初宣布** 2018 年 6 月。

144. 深度学习解释的交互式分类

作 者 :angel cabrera, fred hohman, jason lin, duen horng chou

摘要: 我们提出了一个互动系统，使用户能够操纵图像，探索**深度学习**图像分类器的鲁棒性和敏感性。使用现代 web 技术运行浏览器内推理，用户可以删除图像功能使用内画算法，并获得新的分类实时，这使得他们可以通过实验修改图像和看看模型是如何反应的我们的系统允许用户比较和对比人类**和机器学习**模型用于分类的图像区域，揭示了从壮观的失败（例如，"水瓶" 图像成为一个 "音乐会 "当删除一个人）令人印象深刻的弹性（例如，一个" 棒球运动员 "的形象仍然正确分类，即使没有手套或基地)。我们在 2018 年计算机视觉和模式识别（cvpr）会议上展示了我们的系统，让观众现场

试 用 。 我 们 的 系 统 是 开 源 的 https://github.com/poloclub/interactive-classification 。 https://youtu.be/llub5GcOF6w **提供视频**演示。少

2018 年 6 月 14 日提交;**最初宣布** 2018 年 6 月。

145. 学习分解和分解表示用于视频预测

作者:谢俊婷,刘炳斌,黄德安，李飞, 胡安·卡洛斯·涅布尔斯

摘要: 我们的目标是在给定一系列输入帧的情况下预测未来**的视频**帧。尽管**有大量的视频**数据，但由于**视频**帧的高维性，这仍然是一项具有挑战性的任务。我们通过提出分解成角预测自动编码器 (ddpae) 来应对这一挑战，该框架将结构化概率模型和**深层**网络结合起来，自动 (i) 分解高维度**视频**，我们的目标是预测成组件，和 (ii) 分离每个组件有低维的时间动力学，更容易预测。关键的是，通过适当指定的**视频**帧生成模型，我们的 ddpae 能够在没有明确监督的情况下同时**了解**潜在的分解和分离。对于移动 mnist 数据集，我们显示 ddpae 能够像直观地一样恢复基础组件 (单个数字) 和分离 (外观和位置)。进一步证明了 ddpae 可以应用于涉及多个对象之间复杂交互的弹跳球数据集，从而

直接从像素预测视频帧，并在没有明确监督的情况下恢复**物理**状态。少

2018 年 10 月 17 日提交;**v1** 于 2018 年 6 月 11 日提交;**最初宣布** 2018 年 6 月。

146. 社会环境中的深层好奇心循环

作者:jonatan barkan, goren gordon

摘要: 在婴儿内在学习动机的启发下，我们开发了一个深刻的好奇心循环 (dcl) 架构，他们重视以其自身为基础的信息丰富的感官渠道。dcl 由一个学习者组成，它试图**学习**代理的状态动作转换的正向模型，以及一个新的增强学习 (rl) 组件，即一个动卷深 q 网络，它使用学习者的预测错误作为奖励。我们的代理的环境是由视觉社交场景组成的，由情景喜剧**视频流组成**，因此学习者和 rl 都被构造为**深层**卷积神经网络。代理人的学习者**学习**预测视觉场景动态的零顺序，从而产生与其社会环境中的变化成正比的内在回报。情景喜剧中这些社会信息变化的来源主要是面孔和手的动作，导致在无人监督的情况下**学习**社会交往特征。人手检测以价值函数为代表，社会互动的光流由策略表示。我们的研

究结果表明，人脸和手的**检测**是嵌入社会环境中的基于好奇心的学习的紧急属性。少

2018 年 6 月 10 日提交;**最初宣布** 2018 年 6 月。

147. 基于加控卷积的自由曲面图像绘制

作者:余嘉辉,林哲,杨继美,沈晓辉,新路,黄晓思

摘要: 我们提出了一个新的**深度学习**图像画系统，用自由形式的面具和输入来完成图像。该系统以从数百万张图像中**获得**的门控卷为基础，无需额外的标签工作。提出的门控卷积解决了将所有输入像素视为有效上卷的香草卷积问题，通过为每个空间位置的每个通道提供可学习的动态特征选择机制来概括部分卷积层。此外，由于自由形式的掩码可能出现在任何形状的图像中的任何地方，为单个矩形面罩设计的全球和局部甘肃组织并不合适。为此，我们还提出了一个新的 gan 损失，名为 sn-patchgan，通过应用光谱归一化判别器在密集的图像补丁。配方简单，训练速度快、稳定。自动图像绘制和用户引导扩展的结果表明，与以前的方法相比，我们的系统产生了更高质量、更灵活的结果。我们展示了我们的系统可以帮助用户快速删除分散注意力的对象、修改图像布局、清除水印、编辑人脸以及以交

互方式在图像中创建新对象。此外,**学习特征**表示的可视化揭示了门控卷积的有效性，并提供了对所建议的神经网络如何在缺失区域填充的解释。更多高分辨率的结果和**视频**材料可在 http://jiahuiyu.com/deepfill2 少

2018 年 6 月 10 日提交;**最初宣布** 2018 年 6 月。

148. 一般视频游戏人工智能的深层强化学习

作者:鲁本·罗德里格斯·托拉多,菲利普·邦特拉格,朱利安·托格柳斯, 刘嘉林， 迭戈·佩雷斯-利巴纳

摘要: 通用**视频**游戏 ai (gvgai) 竞赛及其相关的软件框架提供了一种在大量以特定领域的描述语言编写的游戏上对 ai 算法进行基准测试的方法。虽然比赛引起了人们的极大兴趣，但到目前为止，它一直专注于在线规划，提供了一个向前模型，允许使用蒙特卡洛树搜索等算法。本文介绍了 gvgai 如何与 openai 健身房环境(一种广泛使用的将代理连接到增强**学习**问题的方式)接口。使用这个接口，我们描述了在一些 gvgai 游戏中广泛使用的几种**深层**强化**学习**算法的实现情况。我们进一步分析结果，以提供这些游戏相对于彼此的相对难度的初步指示，以及相对于类似条件下街机**学习**环境中的游戏的相对困难。少

2018 年 6 月 6 日提交;**最初宣布** 2018 年 6 月。

149. y 网：一种用于息肉检测的深卷积神经网络

作 者 :ahmed mohammed, sule yildirim, ivar farup, marius petersen, Øistein hovde

摘要: 结直肠息肉是结肠癌的重要前兆，结肠癌是导致男性和女性癌症死亡的第三大常见原因。这是一种早期发现至关重要的疾病。结肠镜检查通常用于癌症和癌前病理的早期发现。这是一个要求很高的程序，需要专业的医生和护士大量的时间，除了严重的误息肉率的专家。结肠镜检查**视频**中的自动息肉检测已被证明是一个有前途的方法来处理这个问题。{然而，息肉检测是一个具有挑战性的问题，因为有限的培训数据和息肉的外观变化很大。针对这一问题，提出了一种新的**深学习**方法 y-net，该方法由两个编码器网络和一个解码器网络组成。我们提出的 y-net 方法依赖于通过新的同时滑块连接操作有效地使用预先训练和未经训练的模型。每个编码器都接受了解码器上编码器特定**学习**速率的训练。与以往采用手工制作的特征或 2-d 三维卷积神经网络的方法相比，我们的方法优于最先进的息肉检测方法，f1 分数和 13% 的召回率提高了。少

2018 年 6 月 5 日提交;**最初宣布** 2018 年 6 月。

150. 顺序出席、推断、重复：运动对象的生成建模

作 者 :adam r. kosiorek, hyunjik kim, ingmar posner, yee whye teh

摘要: 我们提出了顺序出席，推断，重复 (sqair)，一个可解释的深层生成模型的**视频**移动对象. 它可以可靠地发现和跟踪整个帧序列中的对象，还可以生成当前帧上的未来帧调节，从而模拟对象的预期运动。这是通过显式编码模型的潜在变量中的对象存在、位置和外观来实现的。sqair 保留了其前身 "出席、介绍、重复 (air, elami 等人, 2016 年) 的所有优势，包括在无人监督的情况下**学习**，并解决其缺点。我们使用移动的多 mnist 数据集来显示 air 在检测重叠或部分遮挡对象方面的局限性，并显示 sqair 如何通过利用对象的时间一致性来克服这些限制。最后，我们还将 sqair 应用于真实的行人央视数据，在这些数据中，我们**学习**在没有监督的情况下可靠地检测、跟踪和生成步行行人。少

2018 年 6 月 5 日提交;**最初宣布** 2018 年 6 月。

151. 视频描述：方法、数据集和评估指标的调查

作 者 :nayyer aafaq, syed zulqarnain gilani, wei liu, ajmal mian

摘要: 自动**视频**描述可用于帮助视障人士、人机交互、机器人技术和**视频**索引。在过去的几年里，由于计算机视觉和自然语言处理方面的深度 **学习**取得了前所未有的成功，这一领域的研究兴趣大增。文献中提出了许多方法、数据集和评价措施，呼吁进行全面调查，以便更好地将研究工作的重点放在这一蓬勃发展的方向上。本文通过对包括**深度学习**模式在内的最先进方法的调查，准确地满足了这一需求;比较基准数据集的域、类的数量和存储库大小;并确定各种评价指标的利弊，如 BLEU、rouge、meteor、cider、spice 和大规模毁灭性武器。我们的调查显示,**视频**描述研究要与人类的表现相匹配，还有很长的路要走，造成这种不足的主要原因有两个。首先，现有数据集不能充分代表开放域**视频**和复杂语言结构的多样性。其次，目前的评价措施与人的判断不一致。例如，同**一个视频**可以有非常不同但正确的描述。我们的结论是，在规模、多样性和注释准确性方面，需要改进评价措施和数据集，因为它们直接影响到更好的**视频**描述模型的开发。从算法的角度来看，对描述质量的诊断具有挑战性，因为与所采用的语言模

型自然产生的偏差相比，很难评估视觉特征的贡献程度。少

2018 年 6 月 1 日提交;**最初宣布** 2018 年 6 月。

152. 在无约束视频中进行亲属关系验证的混合规范自编码器监控

作者:naman kohli, daksha yadav,mayank vatsa, richa singh, afzel noore

摘要: 由于组织和标记了大量上传在互联网上的大量**视频**等几个应用程序，识别亲属关系引起了人们的兴趣。目前在亲属关系验证方面的研究主要集中在图像对的亲属关系预测上。在本研究中，我们提出了一个新的深度学习框架，在无约束的视频亲属验证使用一个新的**监督** 混合规范正则化自编码器 (smnah)。这种新的自动编码器公式引入了权重矩阵中的类特定稀疏性。提出的基于 smnae 的三阶段亲缘验证框架利用视频帧中**学习到的**时空表示来验证一对视频中的亲缘关系。为这项研究收集了一个新的亲属**视频**(kivi) 数据库，该数据库由 500 多人组成，因光照、姿势、遮挡、种族和表达而有变化。它包括总共 355 真正的亲属**视频**对超过 250,000 帧。kivi 数据库和六个现有亲属数据库显示了拟议

框架的有效性。在 kivi 数据库中, smnai 的基于**视频**的亲属验证精度为 83.18，比现有算法至少高 3.2%。该算法还在六个公开可用的亲属数据库上进行评估，并与报告的最佳结果进行比较。据观察，拟议的 smbe 在所有数据库上都能产生最佳的结果 。

2018 年 5 月 30 日提交;**最初宣布** 2018 年 5 月。

153. 前瞻：通过在线培训对混沌道路环境的未来关注预测

作者:anil sharma, prabhat kumar

文摘: 本文训练了一个递归神经网络，以**了解**混乱道路环境的动态，并在图像上预测环境的未来。未来的投影可以用来预测一个看不见的环境，例如，在自动驾驶。由于车辆和行人等交通参与者之间的互动，道路环境高度动态和复杂。即使在这种复杂的环境中，无论交通参与者的人数多少，人类司机也能在混乱的道路上安全驾驶。**深度学习**研究的激增表明了神经网络在**学习**人类行为方面的有效性。在同一方向上，我们研究复发性神经网络，以了解行人、车辆 (汽车、卡车、自行车等)，有时还有动物共享的混乱道路环境。我们提出了 \ 强调 {前}，这是一个单向的门载经常性单元 (gru) 网络，

以图像的形式关注环境的项目未来。我们在德里的道路上收集了几个**视频**，包括各种交通参与者、背景和基础设施差异（如三维人行横道），时间在不同的时间。我们以无监督的方式训练 \ 素 {foresee}，我们使用在线培训将帧投影到 0。5 提前几秒钟。我们证明了我们提出的模型比最先进的方法（prednet 和 enc. dec. lstm）表现得更好，最后，我们展示了我们训练的模型推广到一个公共数据集，以便将来进行预测。少

2018 年 5 月 30 日提交;**最初宣布** 2018 年 5 月。

154. 受监督的策略更新

作者:全武荣,张一明,基思·w·罗斯

文摘: 我们提出了一种新的样品效率的方法，称为监督策略更新（spu），用于**深层**强化**学习**。从当前策略生成的数据开始，spu 制定并解决了非参数化近端策略空间中的约束优化问题。然后，使用监督回归，将最佳非参数化策略转换为参数化策略，并从中绘制新的示例。该方法是通用的，因为它适用于离散和连续的操作空间，并且可以处理非参数化优化问题的各种接近约束。我们展示了如何使用此方法解决自然策略梯度和信任区域策略优化（npg/trpo）问题以及近端策略优化（ppo）问

题。spu 实现比 trpo 简单得多。在样本效率方面，我们广泛的实验表明，spu 在 mujoco 模拟机器人任务中优于 trpo，在 atari **视频**游戏任务中的性能优于 ppo。少

2018 年 9 月 28 日提交;**v1** 于 2018 年 5 月 29 日提交;**最初宣布** 2018 年 5 月。

155. 通过观看 youtube 玩艰苦的探索游戏

作者:yusuf aytar, tobias pfaff, david budden, tomle paine, ziyuwang, nando de freitas

文摘: **深度**强化**学习**方法传统上与环境奖励特别稀少的任务作斗争。指导这些领域勘探的一个成功方法是模仿人类演示者提供的轨迹。然而，这些演示通常是在人为的条件下收集的，即可以访问代理人的确切环境设置以及演示者的行动和奖励轨迹。在这里，我们提出了一个两阶段的方法，克服这些限制，依靠嘈杂，不对齐的镜头，而无需访问此类数据。首先，我们**学习**使用在时间和模式（即视觉和声音）中构建的自我监督的目标，将来自多个来源的不对齐视频映射到通用表示。其次，我们在此表示中嵌入了一个 youtube **视频**，以构建一个奖励函数，鼓励代理模仿人类的游戏。这种一枪模仿

的方法让我们的经纪人在声名狼藉的艰难探索游戏蒙特祖马的复仇，陷阱中令人信服地超越了人类水平的表现!和私人眼睛的第一次，即使代理没有向任何环境奖励。少

2018 年 5 月 29 日提交;**最初宣布** 2018 年 5 月。

156. 基于高斯混合的全卷变自动编码器视频异常检测与定位

作者:范亚祥,宫建文,李德仁,邱少华，马丁·莱文

文摘: 我们提出了一种新的端 到端部分监督深度学习方法，只使用正常样本进行**视频**异常检测和定位。激发这项研究的见解是，正常样本可以与高斯混合模型 (gmm) 的至少一个高斯分量相关联，而异常要么不属于任何高斯分量。该方法基于高斯混合变分自编码器，可以**学习**正常样本的特征表示作为高斯混合模型训练使用**深度学习**。编码器解码器结构采用不包含完全连接层的完全卷积网络 (fcn)，以保留输入图像和输出要素图之间的相对空间坐标。根据各高斯混合分量的联合概率，提出了一种基于样本能量的图像测试补丁异常评分方法。采用双流网络框架将外观和运动异常结合起来，对后一帧和动态流图像采用 rgb 帧。我们在两个流行

的基准（ucsd 数据集和 avenue 数据集）上测试我们的方法。实验结果验证了我们的方法与艺术状况相比的优越性。少

2018 年 5 月 28 日提交;**最初宣布** 2018 年 5 月。

157. 一种用于图像集分类的简单黎曼流形网络

作者:王瑞,吴晓军,约瑟夫·基特勒

摘要: 在基于图像集的分类领域，将原始图像集表示为协方差矩阵，取得了相当大的进展，这是黎曼流形中典型的协方差矩阵。具体来说，它是一个对称正定（spd）流形。传统的多流**式学习**方法不可避免地具有较高的计算复杂度或弱的特征表示性能。为了克服这些限制，我们提出了一个非常简单的黎曼流形网络进行图像集分类。在**深度学习**架构的启发下，我们设计了一个完全连接的层，以生成更新颖、更强大的 spd 矩阵。然而，我们利用整流层，以防止输入 spd 矩阵是奇异的。我们还介绍了具有创新目标函数的网络的非线性**学习**。此外，我们还设计了一个池层，以进一步减少输入 spd 矩阵的冗余，并设计了日志映射层，将 spd 流形投影到欧几里得空间。为了**学习**输入层和完全连接层之间的连接权，我们使用双向二维主成分分析（2D)2PCA）算法。

提出的黎曼流形网络 (riemnet) 避免了复杂的计算，可以非常简单、高效地构建和训练。我们还开发了一个深版本的 rimnet，名为 drimnet。提出的 rimnet 和 drimnet 在三个方面进行了评估：基于视频的人脸识别、基于集的对象分类和基于集的单元识别。大量的实验结果表明，我们的方法优于最先进的方法。少

2018 年 5 月 27 日提交;**最初宣布** 2018 年 5 月。

158. 基于监控摄像机的定制深度学习视频分析的部署

作 者 :pratik dual, rohan mahadev, suraj kothawade, kunal dargan, rishabh iyer

文摘: 本文展示了我们定制的**基于**深度**学习**的视频分析系统在以安全、安全、客户分析和流程合规性为重点的各种应用中的有效性。我们描述了我们的**视频**分析系统，包括搜索、汇总、统计和实时警报，并概述了其构建块。这些构建块包括目标检测、跟踪、人脸检测和识别、人脸和人脸子属性分析。在每种情况下，我们都演示了使用部署方案中的数据训练的自定义模型如何提供比现成模型优越得多的精度。为此，我们描述了我们的数据处理和模型培训管道，它可以在快速周转时间内从**视频**中训练和微调模型。最后，由于这些模型大多部署在

现场，因此必须有不需要 gpu 的资源受限模型。我们演示了如何自定义资源受限模型，并将其部署到嵌入式设备上，而不会显著降低准确性。据我们所知，这是第一份在监控**视频分析的**各种实际客户部署场景中全面评估不同深度**学习**模型的工作。通过分享我们的实施细节和从为不同客户部署定制的**深度学习**模型中**获得**的经验，我们希望基于定制的**深度学习 视频**分析广泛融入世界各地的商业产品中。少

2018 年 6 月 27 日提交;**v1** 于 2018 年 5 月 27 日提交;**最初宣布** 2018 年 5 月。

159. 利用多任务学习剩余完全卷积网络从航空图像和视频中分割车辆实例

作者:莫立超,朱晓祥

摘要: 目标检测和语义分割是高分辨率遥感图像对象检索的两个主题，最近通过冲浪深度 **学习**的浪潮，更引人注目的是,卷积神经网络 (cnn)。在本文中，我们感兴趣的是一个新的，更具挑战性的车辆实例分割问题，这需要识别，在像素级，车辆出现的位置，以及将每个像素与一个车辆的物理实例相关联。相比之下，车辆检测和语义分割各只涉及这两个问题中的一个。我们建议通

过语义边界感知的多任务**学习**网络来解决这个问题。更具体地说，我们利用剩余**学习**(resnet) 的哲学来构建一个完全卷积的网络，该网络能够利用从不同残差**学中学习到**的多层次上下文特征表示块。我们从理论上分析和讨论了为什么残差网络可以为像素分割任务生成更好的概率图。然后，在此基础上，提出了一个统一的多任务**学习**网络，它可以同时**学习**两个互补的任务，即分割车辆区域和检测语义边界。后一个子问题有助于区分紧密间隔的车辆，这些车辆通常没有正确地分离到实例中。目前，具有基于像素的车辆提取注释的数据集是 isprs 数据集和 ieee grss dfc2015 数据集，专门用于语义分割。因此，我们为车辆实例分割构建了一个新的、更具挑战性的数据集，称为 "繁忙的停车场无人机**视频**数据集"，并在 http://www.sipeo.bgu.tum.de/download 提供我们的数据集，以便用于衡量未来车辆实例分割算法。少

2018 年 5 月 26 日提交;**最初宣布** 2018 年 5 月。

160. 一种基于光场人脸识别的双深时空角学习框架

作 者 :alireza sepas-moghaddam, mohammad a. haque, paulo lobato Alireza, kamal nasrollahi, thomas b. moeslund, fernando pereira

摘要: 人脸识别由于其应用范围广，受到越来越多的关注，但在生物识别数据特性出现较大变化时，人脸识别仍然具有挑战性。lenslet 光场摄像机最近在捕捉丰富的空间角度信息方面发挥了突出作用，从而为先进的生物识别系统提供了新的可能性。本文提出了一种基于光场人脸识别的双**深**空间**角学习**框架，该框架能够利用卷积表示按顺序**学习**纹理和角动力学;这是一个新的识别框架，从来没有提出过，无论是人脸识别或任何其他视觉识别任务。拟议的双**深学习**框架包括一个长的短期存储器 (lstm) 递归网络，其输入是 vgg-face 描述，使用 vgg-ver-ver-x 零 3 层卷积神经网络 (cnn) 计算.vgg-16 网络使用从完整的光场图像呈现的不同的人脸视点，这些视点被组织为伪**视频**序列。利用 ist-eurecom 光场面数据库进行了一系列全面的实验，以执行各种具有挑战性的识别任务。结果表明，与最先进的框架相比，该框架具有更好的人脸识别性能。少

2018 年 10 月 9 日提交;**v1** 于 2018 年 5 月 25 日提交;**最初宣布** 2018 年 5 月。

161. 立体放大倍率：使用多平面图像的学习视图合成

作者:周廷辉,理查德·塔克,约翰·弗林,格雷厄姆·费夫,诺亚·斯纳夫利

摘要: 视图合成问题--从已知图像中生成场景的新视图--最近引起了人们的关注，部分原因是在虚拟现实和增强现实中的应用引人注目。在本文中，我们探索了一个有趣的视图合成场景：从窄基线立体摄像机（包括 vr 摄像机和现在广泛使用的双镜头相机手机）捕获的图像中推断视图。我们将此问题称为立体声放大，并提出一个**学习**框架，该框架利用我们称之为多平面图像 (mpi) 的新的分层表示形式。我们的方法还使用了一个庞大的新数据源进行**学习**视图外推: youtube 上的在线**视频**。利用从这些**视频**中提取的数据，我们训练一个**深度**网络，从输入立体声图像对预测 mpi。然后，这个推断的 mpi 可以用来合成一系列新的场景视图，包括明显推断超出输入基线的视图。结果表明，该方法与近年来几种视图合成方法进行了较好的比较，并在放大窄基线立体图像中得到了应用。少

2018 年 5 月 24 日提交;**最初宣布** 2018 年 5 月。

162. avid:v-抗性视觉不规则性检测

作　　者:mohammad sabokrou, masoud pourreza, mohsenfayyaz, rahim entezari, mahood fathy, jürgen gall, ehsan adeli

摘要: 实时检测视觉数据中的不规则性在许多潜在的应用中非常宝贵和有用，包括监控、患者监测系统等。随着近年来**深度学习**方法的激增，研究人员针对不同的应用尝试了广泛的方法。但是，对于**视频**中的不规则性或异常检测，培训端到端模型仍然是一个开放的挑战，因为通常不规则性没有明确定义，并且在训练过程中没有足够的不规则样本可供使用。本文的灵感来自于生成对抗网络 (gans) 在无监督或自我监督环境中训练**深层**模型的成功，我们提出了一个端到端**的深度**网络，用于检测和精细定位。**视频**(和图像) 中的不规则性。我们提出的体系结构由两个网络组成，它们在相互竞争的同时进行协作，以发现不规则性。一个网络是像素级的不规则性 in 不见器，另一个网络是拍片级探测器。在一次对抗性的自我监督训练之后，我试图愚弄 d 接受其输入输出为常规 (正常)，这两个网络协作检测并细分任何给定测试**视频**中的不规则性。我们在三个不同数据集上的结果表明，我们的方法可以比最先进的方法和精细分割的不规则性。少

2018 年 7 月 17 日提交;**v1** 于 2018 年 5 月 24 日提交;**最初宣布** 2018 年 5 月。

163. 励磁退出：在深部神经网络中鼓励可塑性

作 者 :andrea zunino, sarah adel pargal, pietro morerio, janming, stan sclaroff, vittorio murino

摘要: 我们提出了一个基于网络预测的证据的深部网络的引导辍学调节器: 在特定路径中发射神经元。在这项工作中，我们利用每个神经元的证据来确定辍学的概率，而不是像标准辍学时那样均匀地随机丢弃神经元。从本质上讲，我们辍学的概率更高，那些神经元在训练时对决策贡献更大。这种方法惩罚与模型预测最相关的高显著性神经元，即那些有更有力证据的神经元。通过丢弃这些高显著性神经元，网络被迫**学习**替代路径，以保持损失最小化，从而导致类似塑料的行为，这也是人类大脑的一个特征。我们展示了更好的泛化能力、更高的网络神经元利用率，以及在四个图像/**视频**识别基准上使用多个指标提高了对网络压缩的恢复能力。少

2018 年 5 月 23 日提交;**最初宣布** 2018 年 5 月。

164. 基于网络优化的智能人口系统年龄估计

作者:许振珍,孙鹏,永港温

摘要: 年龄估计是一项艰巨的任务，需要对面部特征进行自动检测和解释。近年来，卷积神经网络 (cnn) 在基准数据集中的**学习**年龄模式方面取得了显著的改善。然

而，对于 "野外" 的脸（来自**视频**框架或互联网）来说，现有的算法并不像正面和中性的脸那样准确。此外，随着野生老化数据的不断增加，现有**深度学习**平台的计算速度成为另一个关键问题。本文提出了一种高效的年龄估计系统，该系统具有年龄估计算法和**深度学习**系统的联合优化。该系统与城市监控网络配合使用，可为智能人口统计提供年龄组分析。首先，我们构建了一个三层雾计算架构，包括边缘、雾和云层，它直接处理原始**视频**中的年龄估计。其次，优化了基于 cnn 的具有标签分布和 k-l 发散距离的年龄估计算法，并对最新野生老化数据集的模型进行了评价。实验结果表明: 1、系统在无接触的情况下动态采集远距离人口统计数据，实现城市人口分析;和 2。年龄模型培训在不失去训练进度或模型质量的情况下加快了速度。据我们所知，这是第一个在提高智能城市和城市生活效率方面具有潜在应用前景的智能人口统计系统。少

2018 年 5 月 21 日提交;**最初宣布** 2018 年 5 月。

165. 深度强化学习的无监督视频对象分割

作者:vik goel, jameson weng , pascal poupart

文 摘 : We present a new technique for **deep** reinforcement **learning** that automatically detects moving objects and uses the relevant information for action selection. The detection of moving objects is done in an unsupervised way by exploiting structure from motion. Instead of directly **learning** a policy from raw images, the agent first **learns** to detect and segment moving objects by exploiting flow information in **video** sequences. The **learned** representation is then used to focus the policy of the agent on the moving objects. Over time, the agent identifies which objects are critical for decision making and gradually builds a policy based on relevant moving objects. This approach, which we call Motion-Oriented REinforcement **Learning** (MOREL), is demonstrated on a suite of Atari games where the ability to detect moving objects reduces the amount of interaction needed with the environment to obtain a good policy. Furthermore, the resulting policy is more interpretable than policies that directly map images to actions or values with a black box neural network. We can gain

insight into the policy by inspecting the segmentation and motion of each object detected by the agent. This allows practitioners to confirm whether a policy is making decisions based on sensible information. △ Less

Submitted 20 May, 2018; **originally announced** May 2018.

166. Long-term face tracking in the wild using deep learning

Authors: Kunlei Zhang, Elaheh Rashedi, Elaheh Barati, Xue-wen Chen

Abstract: This paper investigates long-term face tracking of a specific person given his/her face image in a single frame as a query in a **video**stream. Through taking advantage of pre-trained… ▽ More

Submitted 19 May, 2018; **originally announced** May 2018.

167. Language Expansion In Text-Based Games

Authors: Ghulam Ahmed Ansari, Sagar J P, Sarath Chandar, Balaraman Ravindran

Abstract: Text-based games are suitable test-beds for designing agents that can **learn** by interaction with the environment in the form of natural language text. Very recently,… ▽ More

Submitted 17 May, 2018; **originally announced** May 2018.

168. Auxiliary Tasks in Multi-task Learning

Authors: Lukas Liebel, Marco Körner

Abstract: …impressive results for certain combinations of tasks, such as single-image depth estimation (SIDE) and semantic segmentation. This is achieved by pushing the network towards **learning** a robust representation that generalizes well to different atomic tasks. We extend this concept by adding auxiliary tasks, which are of minor relevance for the application, to t… ▽ More

Submitted 17 May, 2018; **v1**submitted 16 May, 2018; **originally announced** May 2018.

169. Covariance Pooling For Facial Expression Recognition

Authors: Dinesh Acharya, Zhiwu Huang, Danda Paudel, Luc Van Gool

Abstract: …first employ such kind of manifold networks in conjunction with tradi- tional convolutional networks for spatial pooling within in- dividual image feature maps in an end-to-end **deep**… ▽ More

Submitted 13 May, 2018; **originally announced** May 2018.

170. Exploiting Images for Video Recognition with Hierarchical Generative Adversarial Networks

Authors: Feiwu Yu, Xinxiao Wu, Yuchao Sun, Lixin Duan

Abstract: Existing **deep**… ▽ More

Submitted 11 May, 2018; **originally announced** May 2018.

171. **Learning Optical Flow via Dilated Networks and Occlusion Reasoning**

Authors: Yi Zhu, Shawn Newsam

Abstract: Despite the significant progress that has been made on estimating optical flow recently, most estimation methods, including classical and **deep**… ▽ More

Submitted 7 May, 2018; **originally announced** May 2018.

172. **QARC: Video Quality Aware Rate Control for Real-Time Video Streaming via Deep Reinforcement Learning**

Authors: Tianchi Huang, Rui-Xiao Zhang, Chao Zhou, Lifeng Sun

Abstract: Due to the fluctuation of throughput under various network conditions, how to choose a proper bitrate adaptively for real-time **video**streaming has

become an upcoming and interesting issue. Recent work focuses on providing high… ▽ More

Submitted 27 October, 2018; **v1**submitted 7 May, 2018; **originally announced** May 2018.

173. RMDL: Random Multimodel Deep Learning for Classification

Authors: Kamran Kowsari, Mojtaba Heidarysafa, Donald E. Brown, Kiana Jafari Meimandi, Laura E. Barnes

Abstract: The continually increasing number of complex datasets each year necessitates ever improving machine **learning** methods for robust and accurate categorization of these data. This paper introduces Random Multimodel… ▽ More

Submitted 31 May, 2018; **v1**submitted 3 May, 2018; **originally announced** May 2018.

174. Object and Text-guided Semantics for CNN-based Activity Recognition

Authors: Sungmin Eum, Christopher Reale, Heesung Kwon, Claire Bonial, Clare Voss

Abstract: Many previous methods have demonstrated the importance of considering semantically relevant objects for carrying out **video**-based human activity recognition, yet none of the methods have harvested the power of large text corpora to relate the objects and the activities to be transferred into… ▽ More

Submitted 4 May, 2018; **originally announced** May 2018.

175. A Multi-component CNN-RNN Approach for Dimensional Emotion Recognition in-the-wild

Authors: Dimitrios Kollias, Stefanos Zafeiriou

Abstract: …our approach to the One-Minute Gradual-Emotion Recognition (OMG-Emotion) Challenge, focusing on dimensional emotion recognition through visual analysis of the provided emotion **videos**. The approach is based on a Convolutional and Recurrent (CNN-RNN)… ▽ More

Submitted 12 November, 2018; **v1**submitted 3 May, 2018; **originally announced** May 2018.

176. Dimensional emotion recognition using visual and textual cues

Authors: Pedro M. Ferreira, Diogo Pernes, Kelwin Fernandes, Ana Rebelo, Jaime S. Cardoso

Abstract: …emotion expressions in the two-dimensional emotion representation space (i.e., arousal and valence). The adopted methodology is a weighted ensemble of several models from both **video** and text modalities. For…▽ More

Submitted 3 May, 2018; **originally announced** May 2018.

177. Detection of Unknown Anomalies in Streaming Videos with Generative Energy-based Boltzmann Models

Authors: Hung Vu, Tu Dinh Nguyen, Dinh Phung

Abstract: Abnormal event detection is one of the important objectives in research and practical

applications of **video** surveillance. However, there are still three challenging problems for most anomaly detection systems in practical setting: limited labeled data, ambiguous definition of "abnormal" and expensive feature engineering steps. This paper introduces a… ▽ More

Submitted 29 September, 2018; **v1**submitted 2 May, 2018; **originally announced** May 2018.

178. Blazelt: Fast Exploratory Video Queries using Neural Networks

Authors: Daniel Kang, Peter Bailis, Matei Zaharia

Abstract: As **video** volumes grow, analysts have increasingly turned to **deep learning** to process visual data. While these **deep** networks deliver impressive levels of accuracy, they execute as much as 10x slower than real time (3 fps) on a $8,000 GPU, wh… ▽ More

Submitted 2 May, 2018; **originally announced** May 2018.

179. Delay-Constrained Rate Control for Real-Time Video Streaming with Bounded Neural Network

Authors: Tianchi Huang, Rui-Xiao Zhang, Chao Zhou, Lifeng Sun

Abstract: Rate control is widely adopted during **video** streaming to provide both high… ▽ More

Submitted 2 May, 2018; **originally announced** May 2018.

180. Deep learning approach to Fourier ptychographic microscopy

Authors: Thanh Nguyen, Yujia Xue, Yunzhe Li, Lei Tian, George Nehmetallah

Abstract: Convolutional neural networks (CNNs) have gained tremendous success in solving complex inverse problems. The aim of this work is to develop a novel CNN framework to reconstruct **video** sequence of dynamic live cells captured using a computational microscopy technique, Fourier ptychographic

microscopy (FPM). The unique feature of the FPM is its capability to re… ▽ More

Submitted 30 July, 2018; **v1**submitted 26 April, 2018; **originally announced** May 2018.

181. Staircase Network: structural language identification via hierarchical attentive units

Authors: Trung Ngo Trong, Ville Hautamäki, Kristiina Jokinen

Abstract: …a dependency enforced by, for example, the language family, which affects negatively on classification. The other external information sources (e.g. audio encoding, telephony or **video**speech) can also decrease classification accuracy. In this paper, we attempt to solve these issues by constructing a… ▽ More

Submitted 30 April, 2018; **originally announced** April 2018.

182. Deep Keyframe Detection in Human Action Videos

Authors: Xiang Yan, Syed Zulqarnain Gilani, Hanlin Qin, Mingtao Feng, Liang Zhang, Ajmal Mian

Abstract: Detecting representative frames in **videos** based on human actions is quite challenging because of the combined factors of human pose in action and the background. This paper addresses this problem and formulates the key frame detection as one of finding the… ▽ More

Submitted 26 April, 2018; **originally announced** April 2018.

183. **Driving Policy Transfer via Modularity and Abstraction**

Authors: Matthias Müller, Alexey Dosovitskiy, Bernard Ghanem, Vladlen Koltun

Abstract: …to reality via modularity and abstraction. Our approach is inspired by classic driving systems and aims to combine the benefits of modular architectures and end-to-end **deep**… ▽ More

Submitted 1 July, 2018; **v1**submitted 25 April, 2018; **originally announced** April 2018.

184. Person Identification from Partial Gait Cycle Using Fully Convolutional Neural Network

Authors: Maryam Babaee, Linwei Li, Gerhard Rigoll

Abstract: Gait as a biometric property for person identification plays a key role in **video** surveillance and security applications. In gait recognition, normally, gait feature such as Gait Energy Image (GEI) is extracted from one full gait cycle. However in many circumstances, such a full gait cycle might not be available due to occlusion. Thus, the GEI is not complete… ▽ More

Submitted 23 April, 2018; **originally announced** April 2018.

185. Fully Convolutional Adaptation Networks for Semantic Segmentation

Authors: Yiheng Zhang, Zhaofan Qiu, Ting Yao, Dong Liu, Tao Mei

Abstract: The recent advances in **deep** neural networks have convincingly demonstrated high capability in… ▽ More

Submitted 23 April, 2018; **originally announced** April 2018.

186. To Create What You Tell: Generating Videos from Captions

Authors: Yingwei Pan, Zhaofan Qiu, Ting Yao, Houqiang Li, Tao Mei

Abstract: …multimedia contents everyday and everywhere. While automatic content generation has played a fundamental challenge to multimedia community for decades, recent advances of **deep**… ▽ More

Submitted 23 April, 2018; **originally announced** April 2018.

187. Expert Finding in Community Question Answering: A Review

Authors: Sha Yuan, Yu Zhang, Jie Tang, Juan Bautista Cabotà

Abstract: …classify all the existing solutions into four different categories: matrix factorization based models (MF-based models), gradient boosting tree based models (GBT-based models), **deep**… ▽ More

Submitted 21 April, 2018; **originally announced** April 2018.

188. Motion Fused Frames: Data Level Fusion Strategy for Hand Gesture Recognition

Authors: Okan Köpüklü, Neslihan Köse, Gerhard Rigoll

Abstract: …Fused Frames (MFFs), designed to fuse motion information into static images as better representatives of spatio-temporal states of an action. MFFs can be used as input to any **deep**… ▽ More

Submitted 26 April, 2018; **v1**submitted 19 April, 2018; **originally announced** April 2018.

189. Video Compression through Image Interpolation

Authors: Chao-Yuan Wu, Nayan Singhal, Philipp Krähenbühl

Abstract: An ever increasing amount of our digital communication, media consumption, and content creation revolves around **videos**. We share, watch, and archive many aspects of our lives through them, all of which are powered by strong… ▽ More

Submitted 18 April, 2018; **originally announced** April 2018.

190. Deep Generative Networks For Sequence Prediction

Authors: Markus Beissinger

Abstract: This thesis investigates unsupervised time series representation **learning** for sequence prediction problems, i.e. generating nice-looking input samples given a previous history, for high dimensional input sequences by decoupling the static

input representation from the recurrent sequence representation. We introduce three models based on Generative Stochastic… ▽ More

Submitted 18 April, 2018; **originally announced** April 2018.

191. Learning how to be robust: Deep polynomial regression

Authors: Juan-Manuel Perez-Rua, Tomas Crivelli, Patrick Bouthemy, Patrick Perez

Abstract: …Moreover, the problem is even harder when outliers have strong structure. Departing from problem-tailored heuristics for robust estimation of parametric models, we explore **deep**convolutional neural networks. Our work aims to find a generic approach for training… ▽ More

Submitted 23 May, 2018; **v1**submitted 17 April, 2018; **originally announced** April 2018.

192. PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in Spatiotemporal Predictive Learning

Authors: Yunbo Wang, Zhifeng Gao, Mingsheng Long, Jianmin Wang, Philip S. Yu

Abstract: We present PredRNN++, an improved recurrent network for **video** predictive… ▽ More

Submitted 17 April, 2018; **originally announced** April 2018.

193. PlaneNet: Piece-wise Planar Reconstruction from a Single RGB Image

Authors: Chen Liu, Jimei Yang, Duygu Ceylan, Ersin Yumer, Yasutaka Furukawa

Abstract: This paper proposes a **deep** neural network (DNN) for piece-wise planar depthmap reconstruction from a single RGB image. While DNNs have brought remarkable progress to single-image depth prediction, piece-wise planar depthmap reconstruction requires a structured geometry

representation, and has been a difficult task to master even for DNNs. The proposed end-to… ▽ More

Submitted 17 April, 2018; **originally announced** April 2018.

194. Watch, Listen, and Describe: Globally and Locally Aligned Cross-Modal Attentions for Video Captioning

Authors: Xin Wang, Yuan-Fang Wang, William Yang Wang

Abstract: A major challenge for **video** captioning is to combine audio and visual cues. Existing multi-modal fusion methods have shown encouraging results in… ▽ More

Submitted 15 April, 2018; **originally announced** April 2018.

195. Robust Dual View Deep Agent

Authors: Ibrahim M. Sobh, Nevin M. Darwish

Abstract: Motivated by recent advance of machine **learning**using… ▽ More

Submitted 17 April, 2018; **v1**submitted 13 April, 2018; **originally announced** April 2018.

196. Deep Neural Networks motivated by Partial Differential Equations

Authors: Lars Ruthotto, Eldad Haber

Abstract: …approaches that benefit a vast area of tasks including image segmentation, denoising, registration, and reconstruction. In this paper, we establish a new PDE-interpretation of **deep** convolution neural networks (CNN) that are commonly used for… ▽ More

Submitted 11 April, 2018; **originally announced** April 2018.

197. Learning to Extract a Video Sequence from a Single Motion-Blurred Image

Authors: Meiguang Jin, Givi Meishvili, Paolo Favaro

Abstract: We present a method to extract a **video** sequence from a single motion-blurred image. Motion-blurred images are the result of an averaging process, where instant frames are accumulated over time during the exposure of the sensor. Unfortunately, reversing this process is nontrivial. Firstly, averaging destroys the temporal ordering of the frames. Secondly, the… ▽ More

Submitted 11 April, 2018; **originally announced** April 2018.

198. DeepQoE: A unified Framework for Learning to Predict Video QoE

Authors: Huaizheng Zhang, Han Hu, Guanyu Gao, Yonggang Wen, Kyle Guan

Abstract: Motivated by the prowess of **deep**… ▽ More

Submitted 10 April, 2018; **originally announced** April 2018.

199. Echo-Liquid State Deep Learning for 360◦ Content Transmission and Caching in Wireless VR Networks with Cellular-Connected UAVs

Authors: Mingzhe Chen, Walid Saad, Changchuan Yin

Abstract: In this paper, the problem of content caching and transmission is studied for a wireless virtual reality (VR) network in which unmanned aerial vehicles (UAVs) capture **videos** on live games or sceneries and transmit them to small base stations (SBSs) that service the VR users. However, due to its limited capacity, the wireless network may not be able to meet t… ▽ More

Submitted 9 April, 2018; **originally announced** April 2018.

200. Recurrent Neural Networks for Person Re-identification Revisited

Authors: Jean-Baptiste Boin, Andre Araujo, Bernd Girod

Abstract: The task of person re-identification has recently received rising attention due to the high performance achieved by new methods based on **deep learning**. In particular, in the context of **video**-based re-identification, many state-of-the-art works have explored the use of Recurrent N… ▽ More

Submitted 9 April, 2018; **originally announced** April 2018.

201. 基于学习的视频运动放大倍率

作 者 :tae-hyunoh, ronnachai jaroensri, changil kim, mohamed elgharib, frédo durand, william t.freeman, wojciech matusik

摘要: **视频**运动放大技术使我们能够看到肉眼以前看不到的小动作，例如在风的影响下振动飞机机翼或晃动的建筑物。由于运动较小，因此放大倍率结果容易产生噪音或过度模糊。最先进的技术依赖于手工设计的过滤器来提取可能不是最佳的表示。在本文中，我们寻求学习滤波器直接从**实例使用深层**卷积神经网络。为了使训练更容易，我们仔细设计了一个综合数据集，该数据集

能够很好地捕获小运动，并使用两帧输入进行训练。我们表明,**学习滤波器**在真实**视频**上获得高质量的结果，与以往的方法相比，它们的振铃伪影更少，噪声特性更好。虽然我们的模型没有受过时间滤波器的训练，但我们发现，时间滤波器可以与我们提取的放大倍率一起使用，从而实现基于频率的运动选择。最后，我们分析**了学习过的**筛选器，并表明它们的行为类似于以前作品中使用的衍生滤波器。我们的代码、培训模型和数据集将在线提供。少

2018 年 7 月 31 日提交;**v1** 于 2018 年 4 月 8 日提交;**最初宣布** 2018 年 4 月。

202. 通过观看未标记的视频来学习分离对象声音

作者:高若汉, rofero feris, ken gruman

摘要: 最充分地感知场景需要所有的感官。然而，模拟对象的外观和声音具有挑战性：大多数自然场景和事件包含多个对象，而音频轨道将所有声源混合在一起。我们建议从未标记**的视频**中**学习**视听对象模型，然后利用视觉上下文在新**视频**中执行音频源分离。我们的方法依靠**一个深入**的多实例多标签学习框架来分离映射到单个视觉对象的音频基础，即使不观察到这些对象

的孤立。我们展示了如何利用回收的解开基来引导音频源分离，以获得更好的分离，对象级的声音。我们的工作是首次**学习**音频源分离从大规模的 "在野外"**视频**包含多个音频源每个**视频**。我们在视觉辅助音频源分离和音频去噪方面获得最先进的结果。我们**的视频**结果：http://vision.cs.utexas.edu/projects/separating_object_sounds/少

2018 年 7 月 26 日提交;**v1** 于 2018 年 4 月 5 日提交;**最初宣布** 2018 年 4 月。

203. 无约束人脸验证与识别的晶体损耗与质量聚模

作者:rajeev ranjan, ankan bansal, h 行 yu xu, swami sankaranarayanan, jun-chengchen, carlos d.castillo, rama chellappa

文摘: 近年来，基于**深层**卷积神经网络 (dcnn) 的人脸验证识别系统的性能有了显著提高。一种典型的人脸验证管道包括训练深网络进行主题分类，使其具有软最大值损失，使用倒数第二层输出作为特征描述符，并在给定一对人脸图像的情况下生成余弦相似度评分或**视频**。softmax 损失函数不能优化特征，使正极具有较高的相似性分数，而负对的相似度得分较低，从而导致性

能差距。本文提出了一种新的损耗函数，称为 "水晶损耗"，该函数限制了位于固定半径超球面上的特征。使用现有**的深度学习**框架可以很容易地实现损失。我们表明，将这一简单步骤集成到训练管道中，可显著提高人脸验证和识别系统的性能。我们在具有挑战性的 lfw、ijb-a、ijb-b 和 ijb-c 数据集上实现了最先进的性能，在大范围的虚警率 (10-1 至 10-1) 上实现了人脸验证和识别。少

2018 年 4 月 3 日提交;**最初宣布** 2018 年 4 月。

204. 视频帧插值的相控网

作 者 :simone meyer, abdelaziz djelouah, brian mcwilliams, 亚 历 山 大 · 索 金 基 - 霍农, markusgross, christopher schroers

摘要: 大多数**视频**帧插值方法都需要精确的密集对应来合成帧之间的帧。因此，它们在具有挑战性的场景中表现不佳，例如灯光变化或运动模糊。最近依靠内核来表示运动的**深度学习**方法只能在一定程度上缓解这些问题。在这些情况下，使用基于每个像素的相位的运动表示的方法已被证明工作良好。但是，它们仅适用于有限数量的运动。我们提出了一种新的方法，期网，旨在有

力地处理具有挑战性的情况，同时也应对更大的运动。我们的方法由一个神经网络解码器组成，该解码器直接估计中间帧的相位分解。我们证明，这优于以前在基于相位的方法中使用的手工启发式，也优于最近的**深度学习方法的视频**帧插值具有挑战性的数据集。少

2018 年 4 月 3 日提交;**最初宣布** 2018 年 4 月。

205. 深层外观地图

作者:maxim maximov, tobias ritschel, mario fritz

文摘: 提出了一种**外观的深层**表现，即颜色、表面定位、观看者位置、材料和照明的关系。以往的方法是利用**深度学习**来提取与反射率模型参数（例如 phong）或照明（例如 hdr 环境图）有关的经典外观表示。我们建议将外观本身直接表示为**我们称之为深层**外观映射（dam）的网络。这是一个 4d 泛化的二维反射率地图，其中保持了视图方向固定。首先，我们展示了如何从图像或**视频**帧**中学习** dam，然后在给定新的表面方向和查看器位置的情况下，用于合成外观。其次，我们演示了如何使用另一个网络将图像或**视频**帧映射到 dam 网络以重现这种外观，而无需使用冗长的优化（如随机梯度下降（**学习**到学习）).最后，我们将其推广到外观估

计和分割任务中，在该任务中，我们将显示多个材料的图像映射到多个网络，再现其外观，以及每个像素的分割。少

2018 年 4 月 3 日提交;**最初宣布** 2018 年 4 月。

206. 野外学生参与的预测与定位

作 者 :amjot kaur，aamir 穆斯塔法，love mehta ，abhinav dhall

摘要: 本文介绍了一种新的学生参与检测和本地化数据集。数字革命改变了传统的教学程序，对学生在电子**学习**环境中的参与进行结果分析，将有助于有效的任务完成和**学习**。众所周知的社会线索，参与-脱离接触可以从面部表情，身体运动和凝视模式推断。本文记录了学生对各种刺激**视频**的反应，提取了重要的线索来估计参与程度的变化。在本文中，我们研究了一个受试的行为暗示与他的/她的参与水平的关系，由标签者注释。然后，我们使用**基于深度多实例学习**的框架，将非参与部分定位到刺激**视频**中，这可以为设计大规模开放在线课程 (mooc)**视频**提供有用的见解材料。认识到用户参与领域中缺乏任何公开的数据集，因此创建了一个新的 "野外" 数据集，用于研究主题参与问题。该数据集

包含从 78 个主题中拍摄的 195 个**视频**，记录时间约为 16.5 小时。我们使用不同的分类器提供详细的基线结果，从传统的机器**学习**到**深度学习**方法。执行主题独立分析，以便将其推广到新用户。参与预测问题被建模为弱监督**学习**问题。数据集由不同的标签器手动批注四个级别的独立参与，并报告不同分类器对**视频**标注和预测视频标签之间的相关性研究。此数据集的创建是为了促进各种电子**学习**环境 (如智能教学系统、mooc 和其他设备) 的研究。少

2018 年 6 月 26 日提交;**v1** 于 2018 年 4 月 3 日提交;**最初宣布** 2018 年 4 月。

207. 深度强化学习对无级导航的中心驱动探索

作 者 :oleksii zhelo，k 借助 weizhang，lei tai，ming liu，wolfram burkard

文摘: 本文探讨**了深度强化学习**(drl) 方法学习移动机器人导航策略的探索策略。特别是，我们用好奇心测量的内在奖励信号来增强训练 drl 算法的正常外部奖励。我们在无毛导航设置中测试我们的方法，在这种环境中，自主代理需要在没有环境占用图的情况下导航到通过低成本解决方案 (例如可见光) 可以轻松获取其相

对位置的目标本地化, wi-fi 信号本地化)。我们验证了内在动机对于在具有挑战性的勘探要求的任务中提高 drl 性能至关重要。实验结果表明，我们提出的方法能够更有效地**学习**导航策略，并在以前看不见的环境中具有更好的泛化能力。我们实验结果的**视频**可以在 https://goo.gl/pWbpcF 上找到。少

2018 年 5 月 14 日提交;**v1** 于 2018 年 4 月 2 日提交;**最初宣布** 2018 年 4 月。

208. 学习在没有地图的城市中导航

作者:piotr mirowski, mateichi grimes, mateuz malinowski, karl moritz hermann, keithanderson, denis teplyashin, karen simonyan, koray kavukuguoglu, andrew zisserman, raia hadsell

摘要: 在非结构化环境中导航是智能生物的基本能力，因此对人工智能的研究和发展具有根本的意义。远程导航是一项复杂的认知任务，它依赖于开发空间的内部表示，其基础是可识别的地标和强大的视觉处理，可以同时支持连续的自我定位 ("我在这里") 和表示的目标 ("我要去那里")。在近年来**将深层**强化**学习**应用于迷宫导航问题的研究基础上，提出了一种可在城市规模上

应用的端到端深度**强化学习**方法。认识到成功的导航依赖于一般策略与特定于区域设置的知识的集成，我们提出了一种双重路径体系结构，该体系结构允许封装特定于区域设置的功能，同时仍可将其传输到多个城市.我们提供了一个交互式导航环境，使用 google streetview 的摄影内容和全球覆盖，并演示我们的**学习**方法允许代理学习导航多个城市和穿越到可能公里外的目标目的地。视频总结了我们的研究,并显示了在不同城市环境中以及在转移任务中受过训练的代理商,可在 https://sites.google.com/view/streetlearn 上查阅。少

2018 年 4 月 17 日提交;**v1** 于 2018 年 3 月 31 日提交;**最初宣布** 2018 年 4 月。

209. 通过深度自转移强化学习实现缓存动态速率分配

作者:张正明,郑亚鲁,孟华,黄永明,杨鲁西

文摘: 缓存和速率分配是支持无线网络**视频流的**两种很有前途的方法。然而，现有的费率分配设计并没有充分利用这两种方法的优势。本文研究了支持 cachee 驱动的**视频**速率分配问题。针对这一问题，我们建立了一个数学模型，指出传统的动态规划很难解决这个问题。然

后提出了一种**深度强化的学习**方法来解决这一问题。首先，我们将该问题建模为马尔可夫决策问题。然后提出了一种具有特殊知识转移过程的**深度 q 学习**算法，以找出有效的分配策略。最后，数值结果表明，该解决方案能够有效地保持移动用户在小细胞间移动的高质量用户体验。我们还研究了关键参数的配置对算法性能的影响。少

2018 年 3 月 30 日提交;**最初宣布** 2018 年 3 月。

210. diy 人类行动数据集生成

作 者 :mehran khodabandeh, hamid reza vaezi joze, ilya zharkov, vivek pradeep

文摘: 近年来在应用**深度学习**技术解决标准计算机视觉问题方面取得的成功，已成为研究人员在不同领域提出新的计算机视觉问题的需要。正如以前在该领域所确立的那样，培训数据本身在机器**学习**过程中发挥着重要作用，特别是需要数据的**深度学习**方法。为了解决每一个新问题并获得体面的性能，需要收集大量数据，在许多情况下可能会造成后勤困难。因此，能够生成新数据或扩展现有数据集，无论其规模多么小，以满足当前网络的数据需求，可能是非常宝贵的。在此，我们介绍

了一种将动作**视频**剪辑划分为动作、主题和语境的新方法。每个部分都是单独操作的，并使用我们提出的**视频**生成技术进行重新组合。此外，我们的新型人类骨架轨迹生成与我们提出的**视频**生成技术，使我们能够生成无限的行动识别训练数据。这些技术使我们能够从一个小集合生成**视频**动作剪辑，而无需昂贵且耗时的数据采集。最后，通过对两个小型人体动作识别数据集的大量实验证明，这种新的数据生成技术可以提高当前动作识别神经网络的性能。少

2018 年 3 月 29 日提交;**最初宣布** 2018 年 3 月。

211. 用于地地形识别的深纹理歧管

作者:贾雪,张航,克里斯汀·达纳

文摘: 针对地面地形识别的任务，提出了一种称为深编码池网络 (dep) 的纹理网络。地面地形识别是建立机器人或车辆控制参数以及在室外环境中定位的一项重要任务。dep 的体系结构集成了无秩序纹理细节和局部空间信息, dep 的性能超过了最先进的方法来完成这项任务。gtos 数据库 (由 40 种 40 类室外地形的图像组成) 可在监督下进行识别。为了在现实条件下进行评估，我们使用的测试图像不是来自现有的 gtos 数据集，而是

来自类似地形的手持手机**视频**。这个新的评估数据集 GTOS-mobile，由 31 个地面地形的 81 个**视频**组成，如草、砾石、沥青和沙子。由此产生的网络不仅对 gtos 移动显示了出色的性能，而且对更一般的数据库 (minc 和 dtd) 也表现出了出色的性能。利用从该网络**中获得**的判别特征，我们构建了一个新的纹理流形称为 dep 流形。我们**以**完全受监督的方式学习特征空间中的参数分布，它给出了类之间的距离关系，并提供了一种隐式表示模糊类边界的方法。源代码和数据库是公开的。少

2018 年 4 月 2 日提交;**v1** 于 2018 年 3 月 28 日提交;**最初宣布** 2018 年 3 月。

212. 跟踪网：野外目标跟踪的大型数据集和基准

作 者 :matthias müller, adel bibi, silvio giancola, salman al-subaihi, bernard ghanem

摘要: 尽管在目标跟踪方面有了许多发展，但目前跟踪算法的进一步发展受到小的和大多是饱和的数据集的限制。事实上，由于缺乏专门的大规模跟踪数据集，基于**深度学习**的数据渴望跟踪器目前依赖于对象检测数据集。在这项工作中，我们提出了 trackingnet，第一个大规模的数据集和基准的对象跟踪在野外。我们提供超

过 30k **的视频**，有超过 1400 万密集的边界框注释。我们的数据集涵盖了广泛和多样背景下的各种对象类的选择。通过发布如此大规模的数据集，我们期望**深度**跟踪器能够进一步改进和推广。此外，我们还引入了一个由 500 个新颖**视频**组成的新基准，该基准采用类似于我们的训练数据集的分布进行建模。通过保存测试集的注释并提供在线评估服务器，我们为对象跟踪器的未来开发提供了一个公平的基准。**深度**跟踪器对我们数据集的一小部分进行微调，可在 otb100 上将性能提高 1.6%，在 trackingnet 测试中提高到 1.7%。我们通过评估 20 多个跟踪器，为 trackingnet 提供了一个广泛的基准。我们的研究结果表明，野外的物体跟踪远未得到解决。少

2018 年 3 月 28 日提交;**最初宣布** 2018 年 3 月。

213. 基于判别池的视频表示学习

作者：王觉，阿诺普·切里安，fatih porikli，stephen gould

文摘: 视频中流行的深度**操作**识别模型会生成对短片的独立预测，然后将其集中起来，将动作标签分配给完整**的视频**段。由于并非所有框架都可能成为基本操作的特

征--事实上，许多框架在多个操作中都很常见----将对所有帧具有同等重要性的集合方案可能是不利的。为了解决这个问题，我们提出了判别池，其基础是，在所有短片生成的**深层**特征中，至少有一个是行动的特征。为此，我们学习了一个 (非线性) 超平面，它**将**这个未知但有鉴别力的特征与其他特征区分开来。在大边缘设置中应用多个实例学习，我们将这种分离超平面**的**参数用作整个**视频**段的描述符。由于这些参数与最大边距框架中的支持向量直接相关，因此它们可作为函数集合的鲁棒表示。我们制定了一个联合目标和一个高效的求解器，每个**视频**学习这些超平面，并在超平面上**学习**相应的动作分类器。我们的池方案可在**深入**的框架内进行端到端培训。我们报告了在三个基准数据集上进行的实验的结果，这些数据集跨越了各种挑战，并展示了这些任务中最先进的性能。少

2018 年 3 月 29 日提交;**v1** 于 2018 年 3 月 26 日提交;**最初宣布** 2018 年 3 月。

214. 自我监督深度强化学习与推搡的学习协同作用

作者:曾先生,宋舒兰,斯特凡·韦尔克,约翰尼·李,阿尔贝托·罗德里格斯,托马斯·丰克豪斯

摘要: 熟练的机器人操作受益于非可抓取 (如推) 和可抓取 (如抓取) 行动之间的复杂协同作用: 推送可以帮助重新排列杂乱的物体, 为武器和手指腾出空间;同样, 抓取可以帮助取代物体, 使推杆运动更加精确和无冲突。在这项工作中, 我们证明, 通过无模型深度强化**学习**, 可以从零开始发现和**学习**这些协同作用。我们的方法包括训练两个完全卷积的网络, 从视觉观察映射到行动: 一个推断推送的效用, 以密集的像素为导向的末端效应方向和位置采样, 而另一个则相同的抓握。这两个网络都是在 q **学习**框架内联合培训的, 完全由试错自我监督, 成功的掌握提供奖励。这样, 我们的政策**学会了**推动运动, 使未来能够掌握, 同时**学习**掌握, 可以利用过去的推动。在模拟和实际场景中的拾取实验中, 我们发现我们的系统在充满挑战的杂乱情况下快速**学习**复杂的行为, 并比基线实现更好的把握成功率和拾取效率选择只有几个小时的培训。我们进一步证明了我们的方法能够推广到新的对象。定性结果 (**视频**)、代码、预先培训的模型和模拟环境可 http://vpg.cs.princeton.edu 少

2018 年 9 月 30 日提交;**v1** 于 2018 年 3 月 27 日提交;**最初宣布** 2018 年 3 月。

215. 神经网络超参数的一种有纪律的方法：第 1 部分--学习速率、批处理大小、动量和权重衰减

作者:莱斯利·史密斯

摘要: 尽管**在**过去几年里，深度**学习**在图像、语音和**视频**处理的应用中取得了令人眼花缭乱的成功，但大多数培训都是超参数不理想的，需要不必要的长培训时间。设置超参数仍然是一门需要多年经验才能获得的黑色艺术。本报告提出了几种有效的方法来设置超参数，显著减少训练时间并提高性能。具体而言，本报告演示了如何检查训练验证/测试损失函数，以寻找不合适和过度拟合的细微线索，并提出了向最佳平衡点移动的指导原则。然后讨论了如何提高**学习**速度，加快训练的势头。我们的实验表明，平衡每个数据集和体系结构的各种正则化方式是至关重要的。重量衰减被用作样本调节器，以显示其最佳值如何与**学习**速率和动量紧密耦合。可提供帮助复制此处报告的结果的文件。少

2018 年 4 月 24 日提交;**v1** 于 2018 年 3 月 26 日提交;**最初宣布** 2018 年 3 月。

216. 学习任务型注意转换预测以自我为中心的视频中的注视

作者:黄一飞,蔡敏杰, 李振强,佐藤洋一

摘要: 通过探索依赖自我中心操作任务的注视固定（注意力转换) 的时间转移模式，提出了一种新的自我中心视频凝视预测计算模型。我们的假设是，任务如何以某种方式完成的高级上下文对注意力转换有很大影响，应该在自然动态场景中进行凝视预测建模。具体而言，我们提出了一种基于**深度**神经网络的混合模型，该模型将任务相关的注意力转换与自下而上的显著性预测相结合。特别是，任务依赖注意力转换是通过反复出现的神经网络来**学习**的，以利用凝视固定的时间上下文，例如，在将目光移离抓住的瓶子后看着杯子。对公众自我中心活动数据集的实验表明，我们的模型明显优于最先进的凝视预测方法，能够**学会人类**注意力的有意义的转变。少

2018 年 7 月 20 日提交;**v1** 于 2018 年 3 月 24 日提交;**最初宣布** 2018 年 3 月。

217. 组规范化

作者:吴玉欣,何凯明

文摘 批量归一化（bn）是**深度学习**发展中的一项里程碑技术，使各种网络能够进行培训。但是，沿批处理维

度进行规范化会带来问题----当批处理大小变小时，由于批处理统计估计不准确而导致 bn 的错误会迅速增加。这限制了 bn 用于训练较大的模型，并将功能传输到计算机视觉任务，包括检测、分段和**视频**，这些任务需要小批量受内存消耗的限制。在本文中，我们提出了组归一化 (gn) 作为 bn 的一个简单的替代方法，将通道划分为多个组，并在每个组中计算归一化的均值和方差。gn 的计算与批处理大小无关，其精度在各种批次大小中都是稳定的。在 imagenet 训练的 resnet-50 上，使用 2 的批处理大小时，gn 的误差比 bn 值低 10.6;使用典型的批处理大小时，gn 与 bn 相比较好，并且优于其他规范化变体。此外，gn 可以自然地从预训练转移到微调。在 coco 中的目标检测和分割以及动力学中的视频分类方面，gn 的**性能**可以优于基于 bn 的网络，这表明 gn 可以在各种任务中有效地取代强大的 bn。在现代库中，gn 可以很容易地通过几行代码来实现。少

2018 年 6 月 11 日提交;**v1** 于 2018 年 3 月 22 日提交;**最初宣布** 2018 年 3 月。

218. 多视图多类对象投比的统一框架

作者:池莉,金白,格雷戈里·黑格

摘要: 对象姿态估计的一个核心挑战是确保在复杂的背景杂乱中，大量不同前景对象的准确和鲁棒性能。在本工作中，我们提出了一个可扩展的框架，以便从单个或多个视图准确推断大量对象类的六个自由度（6-dof）姿势。为了**学习**判别姿势特征，我们将三个新功能集成到**一个深**卷积神经网络（cnn）中：一种基于统一的细分结构的分类和姿势回归相结合的推理方案。三个维度(SE(3) 的特殊欧几里得群，通过平铺的类地图将班级前科融合到训练过程中，并使用带有对象掩码的**深度**监控进行额外的正则化。此外，还制定了一个有效的多视图框架，以解决单视图模糊问题。我们表明，该框架持续提高了单视图网络的性能。我们根据三个大规模基准对我们的方法进行评估:ycb-视频、jushcene-50 和 ObjectNet-3D。与目前最先进的方法相比，我们的方法具有竞争力或卓越的性能。少

2018 年 10 月 6 日提交;**v1** 于 2018 年 3 月 21 日提交;**最初宣布** 2018 年 3 月。

219. 利用深层残差网络从骨骼数据中识别人体动作

作 者 :huy-hieu pham, louahdi khoudour, alain crouzil, pablo zegers, sergio a. velastin

摘要: 计算机视觉社区目前正专注于解决真实**视频**中的行动识别问题，其中包含数千个样本，面临诸多挑战。在这一过程中,**深**卷神经网络 (d-nnn) 在各种基于视觉的行动识别系统中发挥了重要作用。最近，在一个名为"剩余网络 (resnet)" 的单一架构中引入了与更传统的 cnn 模型相结合的剩余连接，显示出令人印象深刻的性能和图像识别任务的巨大潜力。本文利用深度传感器提供的骨骼数据，研究并应用**深度**重置网络进行人体动作识别。首先，将骨架序列中人体关节的三维坐标转化为基于图像的表示，并存储为 rgb 图像。这些彩色图像能够从骨架序列中捕获三维运动的时空演化，并可通过 d-nns 有效地**学习**。然后，我们提出了一个新的基于 resnet 的**深度学习**架构，以**学习**从获得的基于颜色的表示特征，并将它们分类为行动类。该方法在三个具有挑战性的基准数据集上进行评估，包括 msr action 3d、kard 和 ntu-rgb + d 数据集。实验结果表明，我们的方法在降低计算资源的同时，实现了所有这些基准的最先进性能。特别是，在 msr action 3d 数据集上，该方法大大超过了以前的方法 3.4%，在 kard 数据集上超过了 0.67，在 ntu-rgb + d 数据集上超过了 2.5%。少

2018 年 3 月 21 日提交;**最初宣布** 2018 年 3 月。

220. 利用深部残余神经网络从骨骼运动中学习和识别人的行为

作 者 :huy-hieu pham, louahdi khoudour, alain crouzil, pablo zegers, sergio a. velastin

摘要: 自动人体动作识别对于**视频**监控、人机界面、**视频**检索等几乎是人工智能系统不可或缺的。尽管取得了很大的进展，但在未知**视频**中识别动作在计算机视觉中仍然是一项具有挑战性的任务。近年来,**深度学习**算法在许多与视觉相关的识别任务中被证明具有巨大的潜力。本文提出了利用**深**部残余神经网络 (resnet) 从 kinect 传感器提供的骨架数据中**学习**和识别人类行为的方法。首先，将体关节坐标转换为三维阵列，并保存在 rgb 图像空间中。设计了五种不同的基于 resnet 的**深层学习**模型，提取图像特征并将其分类。在两个公共**视频**数据集上进行了实验，以识别人类的行动，其中包含了各种挑战。结果表明，与现有方法相比，该方法具有最先进的性能。少

2018 年 3 月 21 日提交;**最初宣布** 2018 年 3 月。

221. 直觉推理的框架与基准

作　　者:ronan riochet, mario ynocente castro, mathieubernard, adam lerer, rob fergus , véronique izard , emmanuel dupoux

摘要: 为了达到人类在复杂视觉任务上的表现，人工系统需要从宏观物体、运动、力量等方面纳入对世界的大量了解。在婴儿直觉物理工作的启发下，我们提出了一个评估框架，通过测试一个特定系统是否能够区分可能与不可能发生的事件的匹配**视频**来诊断给定系统对物理的理解程度。测试要求系统计算整个**视频**的物理合理性得分。它没有偏见，可以测试一系列特定的物理推理技能。然后，我们描述了基准数据集的第一个版本，该版本旨在使用使用游戏引擎构建的**视频**，以无监督的方式**学习**直观的物理。我们描述了两个**深**神经网络基线系统训练与未来的帧预测目标，并测试了可能的和不可能的区分任务。通过对其结果与人类数据的比较，可以对下一帧预测体系结构的潜力和局限性进行新的洞察。少

2018 年 6 月 26 日提交;**v1** 于 2018 年 3 月 20 日提交;**最初宣布** 2018 年 3 月。

222.　一种用于视频帧绘制的时间感知插值网络

作者:孙锡蒙, ryan szeto, jason j. corso

文摘: 我们提出了第一个**深度学习**解决方案的视频帧画，一个具有挑战性的实例，一般**视频**绘制问题的应用在**视频**编辑，操作，和取证.我们的任务没有帧插值和**视频**预测那么模糊，因为我们可以访问时间上下文和对未来的部分一瞥，从而使我们能够客观地评估模型预测的质量。设计了一个由两个模块组成的管道: 双向**视频**预测模块和时间感知帧插值模块。预测模块使用基于卷积 lstm 的编码解码器对缺失的帧进行两个中间预测，一个以前面的帧为条件，另一个以以下帧为条件。插值模块混合中间预测以形成最终结果。具体来说，它利用**视频**预测模块中的时间信息和隐藏激活来解决预测之间的分歧。我们的实验表明，与最先进的**视频**预测方法和许多强大的帧绘制基线相比，我们的方法产生了更准确、更高质量的结果。少

2018 年 11 月 3 日提交;**v1** 于 2018 年 3 月 19 日提交;**最初宣布** 2018 年 3 月。

223. 机器人操纵的可组合深层强化学习

作 者 :tuomas haarnoja, vitchyr pong, aurick zhou, murtaza dalal, pieter abbeel, sergey levine

文摘: 无模型**深层**加固**学习**已被证明在**从电子**游戏到模拟机器人操作和运动等领域表现出良好的性能。但是，当与环境的交互时间有限时，无模型方法的性能很差，就像大多数现实世界中的机器人任务一样。本文研究了如何将使用软 q **学习**训练的最大熵策略应用于现实世界中的机器人操作。软 q **学习**的两个重要特征促进了该方法在实际操作中的应用。首先，软 q **学习**可以通过以表达能量为基础的**模型来表示**的学习策略来学习多模态探索策略。其次，我们表明，用软 q **学习学学习**学学习得出的政策可以组合起来，以创建新的策略，所产生的策略的最优性可以从组合策略之间的差异来限制。这种组合为现实世界的操作提供了一个特别有价值的工具，在这种工具中，通过组合现有技能构建新的策略可以从零开始提供比培训更高的效率。我们的实验评估表明，软 q **学习**比以前的无模型深层强化**学习**方法具有更高的样本效率，并且可以对两者进行组合模拟和现实世界的任务。少

2018 年 3 月 18 日提交;**最初宣布** 2018 年 3 月。

224. 表面：充分利用序列信息进行人脸识别

作者:胡晓宇,黄阳宇,张帆,李瑞瑞,李伟,袁国东

文摘 近年来，深层卷积神经网络（cnn）大大提高了人脸识别（fr）的性能。fr 中几乎所有的 cnn 都在包含大量身份的经过仔细标记的数据集上接受培训。然而，这样高质量的数据集的收集成本非常高，这限制了许多研究人员达到最先进的性能。在本文中，我们提出了一个框架，称为 seqface，学习判别脸的特点。除了传统的身份训练数据集外，设计的 seqface 还可以使用额外的数据集（包括从**视频**中收集的大量面部序列）来培训 cnn。此外，还利用标签平滑正则化（lsr）和新提出的判别序列代理（dsa）损耗，充分利用序列数据，增强**了深**面特征的判别能力。我们的方法在野外标记面(lfw)、youtube 面（ytf）上实现了出色的性能，只需使用单个 resnet 即可。代码和模型可在网上公开使用(https://github.com/huangyangyu/SeqFace)。少

2018 年 3 月 23 日提交;**v1** 于 2018 年 3 月 17 日提交;**最初宣布** 2018 年 3 月。

225. 一种具有工作记忆的可视化推理的数据集和体系结构

作者:杨光宇,加尼切夫，王晓静,乔纳森·舒伦，大卫·苏西略

摘要: 人工智能中的一个令人烦恼的问题是对复杂的、不断变化的视觉刺激（如**视频**分析或游戏中）中发生的事件进行推理。在认知心理学和神经科学的视觉推理和记忆的丰富传统的启发下，我们开发了一个人工的、可配置的视觉问题和答案数据集（cog），用于人类和动物的平行实验。cog 比**一般的视频**分析问题要简单得多，但它解决了许多与视觉和逻辑推理及记忆有关的问题--这些问题对现代**深度学习**来说仍然是一个挑战架构。此外，我们还提出了一个**深入的学习**架构，该架构可在其他诊断 vqa 数据集（即 clear）上具有竞争力，并可轻松设置 cog 数据集。但是，cog 的几个设置会导致**数据**集的学习难度逐渐提高。培训结束后，网络可以将零镜头泛化到许多新任务。对在 cog 上训练的网络架构的初步分析表明，网络以人类可以解释的方式完成任务。少

2018 年 7 月 20 日提交;**v1** 于 2018 年 3 月 16 日提交;**最初宣布** 2018 年 3 月。

226. 乐高：通过观看视频，一次学习几何边缘

作者:杨振恒,王鹏,王洋，徐伟,拉姆·内马提亚

摘要: **通过深**卷积网络观看未标记**的视频**，学习估计单个图像中的 3d 几何是一个重要的问题。本文介绍了管道内部的 "尽可能平滑的 3d (3d-asap)"，该方法实现了边缘和三维场景的联合估计，在精细的详细结构中获得了显著提高的精度。具体来说，我们在此之前定义 3d-asap，要求从图像中以 3d 方式恢复的任何两个点如果没有提供其他提示，则应位于现有的平面上。我们设计了一个无监督的框架，该框架可一次学习边缘和几何 (深度、正常) (乐高)。预测的边缘嵌入到深度和表面正常平滑项中，其中没有边缘的像素被约束以满足前面的值。在我们的框架中，预测的深度、法线和边一直都是一致的。我们在 kititi 上进行实验，以评估我们估计的几何形状和城市标准，以进行边缘评估。我们展示了在所有任务中, i.e.depth、正常和边缘，我们的算法大大优于其他最先进的 (sota) 算法，展示了我们的方法的优点。少

2018 年 3 月 23 日提交;**v1** 于 2018 年 3 月 15 日提交;**最初宣布** 2018 年 3 月。

227. 通过对象嵌入网络学习玩视频游戏

作者:william woof, ke chen

文摘: **深度强化学习**(drl) 已被证明是创建一般**视频**游戏 ai 的有效工具。然而，大多数当前的 drl **视频**游戏代理从游戏的**视频**输出中**学习端到**端，这对许多应用程序来说是多余的，并产生了一些额外的问题。更重要的是，直接处理基于像素的原始**视频**数据与人类玩家所做的工作有很大的不同。本文提出了一种新的方法，使 drl 代理能够直接从对象信息中**学习.**这是通过使用对象嵌入网络 (oen) 获得的，该网络将一组不同长度的对象特征向量压缩为一个表示当前游戏状态的固定长度统一特征向量，并同时实现 drl 。我们通过与 gvg-ai 竞赛中选择的游戏的几种最先进的方法进行比较，来评估我们基于 oens 的 drl 代理。实验结果表明，我们的基于对象的 drl 剂的性能与我们比较研究中使用的方法相当。少

2018 年 5 月 28 日提交;**v1** 于 2018 年 3 月 14 日提交;**最初宣布** 2018 年 3 月。

228. 2017 年 aibirds 竞赛

作者:马修·斯蒂芬森,任俊,葛晓宇,张鹏

摘要: 本文概述了在第 26 届人工智能国际联席会议上举行的第六届 aibirds 竞赛。本次比赛的任务是参与者

开发一个智能代理，可以玩基于物理的益智游戏愤怒的小鸟。这个游戏使用一个复杂的物理引擎，需要代理在有限的环境信息下推理和预测行动的结果。参加这次比赛的代理商被要求在规定的时限内解决大量以前看不见的水平。解决这些层次所需的物理推理和规划与许多现实问题非常相似。今年的比赛中，一些迄今开发的最好的代理商，甚至包括**了一些**新的 ai 技术，如深度强化**学习**。在本文中，我们描述了本次比赛的框架、规则、提交的代理和结果。我们还提供了一些有关相关工作和其他**视频**游戏 ai 竞赛的背景信息，并讨论了未来 aibirds 比赛和代理改进的一些潜在想法。少

2018 年 3 月 14 日提交;**最初宣布** 2018 年 3 月。

229. 压缩视频的多帧质量提升

作者:任阳,徐麦明,王祖林,李天一

文摘: 在过去的几年里，在应用**深度学习**来提高压缩图像/**视频**的质量方面取得了巨大的成功。现有的方法主要侧重于提高单个帧的质量，而忽略了连续帧之间的相似性。本文研究了压缩**视频**帧中存在的质量波动较大的问题，利用相邻的高质量帧（称为多帧质量增强(mfqe)，可以提高低质量帧。因此，本文提出了压缩**视**

频的 mfqe 方法，作为这方面的首次尝试。在我们的方法中，我们首先开发了一个基于支持向量机 (svm) 的检测器，以定位压缩**视频**中的峰值质量帧 (pqf)。然后，设计了一种新的多帧卷积神经网络 (mf-cnn)，以提高压缩**视频**的质量，其中非 pqf 及其最近的两个 pqf 作为输入。mf-cnn 通过运动补偿子网 (mc-子网) 补偿非 pqf 和 pqf 之间的运动。随后，质量增强子网 (qe-子网) 在其最近的 pqf 的帮助下，减少了非 pqf 的压缩伪影。最后，实验验证了我们的 mfqe 方法在推动**压缩视频**质量最先进方面的有效性和通用性。我们的 mfqe 方法的代码可在 https://github.com/ryangBUAA/MFQE.git 少

2018 年 3 月 15 日提交;**v1** 于 2018 年 3 月 13 日提交;**最初宣布** 2018 年 3 月。

230. gonet: 一种半监督的可跟踪性估计深度学习方法

作 者 :noriaki hirose, amir sadeghian, marynel vázquez, patrick goebel, silvio savarese

文摘: 提出了一种半监督深度**学习**方法，用于从鱼眼图像中进行可遍历**性**估计。我们的方法 (gonet) 和建议的扩展利用生成对抗性网络 (gans) 来有效地预测输入

图像中看到的区域对于机器人的遍历是否安全。这些方法是训练与许多正面图像的可穿越的地方，但只是一小群负面图像描绘被封锁和不安全的地区。这使得所提出的方法切实可行。通过简单地通过可遍历的空间操作机器人，可以很容易地收集出积极的例子，而获得负面的例子既耗时又昂贵，而且具有潜在的危险性。通过大量的实验和几个演示，我们证明了所提出的可遍历性估计方法是鲁棒性的，可以推广到看不见的场景。此外，我们还证明了我们的方法是内存高效和快速的，允许在带有单台或立体声鱼眼相机的移动机器人上实时操作。作为我们贡献的一部分，我们将两个新的数据集开源，用于可遍历性估计。这些数据集由来自超过 25 个室内环境的**大约 24 小时视频**组成。我们的方法优于基线方法，可在这些新数据集上进行可遍历性估计。少

2018 年 3 月 8 日提交;**最初宣布** 2018 年 3 月。

231. 分散的顺序自动编码器

作者:李英珍,斯蒂芬·曼特

摘要: 我们提出了一个 vae 架构，用于编码和生成高维序列数据，如**视频**或音频。我们**的深层**生成模型**学习**了数据的潜在表示形式，这些数据被分解成静态和动

态部分，使我们能够从保留的特征中近似地分离出潜在的时间相关特征（动态)时间（内容)。这种架构通过对这一组功能中的任何一组进行调节，使我们对生成内容和动态进行了部分控制。在我们人工生成**的卡通视频**剪辑和语音录制的实验中，我们表明，我们可以通过这样的内容交换将给定序列的内容转换为另一个序列。对于音频，这允许我们将男性扬声器转换为女性扬声器，反之亦然，而对于**视频**，我们可以单独操作形状和动态。此外，我们还给出了随机 rnn 作为潜在状态模型的假设的经验证据，这种假设比确定性序列更有效地压缩和生成长序列，这可能与**视频**中的应用有关压缩。少

2018 年 6 月 12 日提交;**v1** 于 2018 年 3 月 8 日提交;**最初宣布** 2018 年 3 月。

232. 无监督的面部表现学

作者:samyak datta, gaurav sharma, c. v. jawahar

文摘: 我们提出了一个方法，在不受监督的培训 cnn，以**学习**歧视性的脸表示。我们挖掘监督培训数据，指出同一**视频**帧中的多个面必须属于不同的人，而跨多个帧跟踪的同一面必须属于同一个人。我们无需使用任何

手动监控即可从数百个**视频**中获取数百万张面孔对。尽管从**视频**中提取的人脸的空间分辨率低于作为标准监督人脸数据集（如 lfw 和 casia-webface）的一部分提供的人，但前者代表了一个更为现实的设置，例如在监视中大多数检测到的人脸都很小的情况。我们使用从收集到的**视频**帧中提取的相对较低的分辨率面对 cnn 进行培训，并在基准的 lfw 数据集上实现更高的验证精度，参见手动制作的要素（如 lbp），甚至超过了最先进的**深**网络，如 vgg 面，当它们被制成工作与低分辨率的输入图像。少

2018 年 3 月 3 日提交;**最初宣布** 2018 年 3 月。

233. 与相互信息的哈希

作 者 :fatih cakir, kun he , sarah adel argal, stan sclaroff

文摘: 二元矢量嵌入可在高维对象的大型数据库中实现快速最近的邻域检索，并在图像和**视频**检索等许多实际应用中发挥着重要作用。我们研究了在监督环境下**学习**二元矢量嵌入的问题，也称为哈希。提出了一种新的基于优化信息理论量: 互信息的监督哈希方法。我们表明，优化互信息可以减少**学习**的汉明空间中诱导邻域

结构的模糊性，这对获得较高的检索性能至关重要。为此，我们优化了具有微型随机梯度下降的深部神经网络中的互信息，并采用了最大限度、高效地利用现有监控的公式。在包括 imagenet 在内的四个图像检索基准上的实验证实了我们的方法在学习高质量的二进制嵌入以实现最近邻检索方面的有效性。少

2018 年 6 月 24 日提交;**v1** 于 2018 年 3 月 2 日提交;**最初宣布** 2018 年 3 月。

234. 移动网络中 youtube http 自适应流媒体服务的分类流和缓冲区状态

作　　者:dimitrios tsilimantos,theodoros karagkioules, stefan valentin

摘要: 准确的跨层信息对于针对特定应用优化移动网络非常有用。但是，由于广泛采用端到端加密，并且缺乏跨层信令标准，向较低的协议层提供应用层信息变得非常困难。作为一种替代方法，本文提出了一种流量分析解决方案，用于被动估计较低层的 http 自适应流 (has) 应用程序的参数。通过观察 ip 数据包的到达，我们的机器学习系统可以识别视频流，并实时检测 has 客户端的回放缓冲区的状态。我们对 youtube 移

动客户端的实验表明，随机森林即使在链接质量的变化很大的情况下也能实现非常高的精度。由于这种高性能是在 ip 级别上实现的，具有较小的通用功能集，因此我们的方法不需要**深度**数据包检查 (dpi)，而且复杂性较低，并且不会干扰加密。因此，流量分析是监视和管理移动网络中甚至加密的 has 流量的一个功能强大的新工具。少

2018 年 5 月 29 日提交;**v1** 于 2018 年 3 月 1 日提交;**最初宣布** 2018 年 3 月。

235. 视频监控中人脸识别的深度学习体系结构

作者:saman bashbaghi, eric granger, robert sabourin , mostafa parchami

摘要: 用于**视频**监控 (vs) 应用的人脸识别 (fr) 系统试图在分布式摄像机网络上准确检测目标个人的存在。在**基于视频**的 fr 系统中，目标个体的面部模型在注册过程中使用数量有限的参考静止图像或**视频**数据进行了先验设计。由于照明、姿势、比例、遮挡、模糊和相机互操作性的差异，这些面部模型通常不能代表在操作过程中观察到的面部。具体而言，在静止不动的视频 fr 应用程序中，使用在受控条件下使用静止相机捕获的

单个高质量参考静止图像来生成面部模型，以便以后与拍摄的低质量的面孔进行匹配。**摄像机**在不受控制的条件下。当前基于**视频的** fr 系统可以在受控方案上很好地执行，而它们在不受控制的方案中的性能并不令人满意，主要原因是源（注册）和目标（操作）域之间的差异。这一领域的大部分工作都是在不受限制的监控环境中设计强大的基于**视频**的 fr 系统。本章概述了通过深层卷积神经网络（cnn）在**视频**到**视频**的 fr 场景中的最新进展。特别是，文献中提出的基于三重损失函数（例如，互相关联匹配的 cnn、树干分支合奏 cnn 和 haarnet）和监督的自动编码器（例如，规范面）的**深度学习**架构表示美国有线电视新闻网）审查和比较的准确性和计算复杂度。少

2018 年 6 月 27 日提交;**v1** 于 2018 年 2 月 27 日提交;**最初宣布** 2018 年 2 月。

236. 按分区、草和卷曲分列的真实世界重复估计

作者:tom f. h.runia , cees g.m . snoek, arnold w. m. 斯梅尔斯

摘要: 我们考虑的问题是估计**视频**中的重复，比如表演俯卧撑、切瓜或拉小提琴。现有工作在静态和平稳周期

的假设下取得了较好的效果。由于逼真**的视频**很少是完全静态和静止的，通常喜欢的基于傅立利的测量是不合适的。相反，我们采用小波变换来更好地处理非静态和非平稳**视频**动力学。从流场及其微分中，我们推导出三维内在周期性的三种基本运动类型和三种运动连续性。除此之外，3d 周期的 2d 感知还考虑了两个极端的观点。以下是在 2d 中反复感知的 18 例基本案例。在实践中，为了处理重复外观的多样性，我们的理论意味着测量时变流及其在分段前景运动上的差异 (梯度、发散和卷曲)。对于实验，我们引入了新的 quva 重复数据集，通过包含非静态和非静止**视频**来反映实际情况。在**计算视频**重复的任务上，与**深度学习**的替代方案相比，我们获得了良好的效果。少

2018 年 2 月 27 日提交;**最初宣布** 2018 年 2 月。

237. 在新颖性检测中的兼职分类器

作　　者：mohammad sabokrou, mohammad kalooei, mahood fathy,ehsan adeli

摘要: 新颖性检测是识别在某些方面不同于训练观察 (目标类) 的观察的过程。在现实中，新奇的阶层往往是缺席的培训，抽样差或没有很好的定义。因此，一类分

类器可以有效地对此类问题进行建模。然而，由于无法获得新奇类的数据，培训端到端**深度**网络是一项繁琐的任务。本文在生成对抗网络成功的启发下，在无监督和半监督环境中训练**深层**模型，提出了一类分类的端到端体系结构。我们的体系结构由两个**深层**网络组成，每个深度网络都经过相互竞争的训练，同时协作了解目标类中的基本概念，然后对测试样本进行分类。一个网络作为新奇的探测器工作，而另一个网络通过增强输入样本和扭曲异常值来支持它。直觉是，增强的初始化器和扭曲的异常值的可分性远远好于决定原始样本。该框架适用于图像和**视频**中异常和异常点检测的不同相关应用。mnist 和 caltech-256 图像数据集的结果，以及具有挑战性的 ucsd ped2 视频异常检测数据集，表明我们提出的方法有效地**学习**了目标类，优于基线和最先进的方法。少

2018 年 5 月 24 日提交;**v1** 于 2018 年 2 月 25 日提交;**最初宣布** 2018 年 2 月。

238. 深度在线视频稳定

作者:王苗,杨国业,林金坤, 沙米尔, 张松海, 吴少平,胡世民

摘要: **由于**高频抖动，视频稳定技术对于大多数手持**视频**来说是必不可少的。几种基于 2d、2.5 d 和 3d 的稳定技术得到了很好的研究，但据我们所知，还没有提出基于**深度**神经网络的解决方案。其原因主要是培训数据短缺，以及使用神经网络建模问题的挑战。本文利用卷积神经网络 (convnet) 解决**了视频**稳定问题。我们没有处理基于特征匹配的离线整体摄像机路径平滑，而是专注于低延迟实时摄像机路径平滑，而不明确表示摄像机路径。我们的网络称为 stabnet，它在创建更稳定的潜在摄像机路径的同时，逐渐为每个输入的非定常帧学习转换。为了训练网络，我们通过精心设计的手持硬件创建了同步稳定视频对的数据集。实验结果表明，该在线方法 (不使用未来帧) 的性能与传统的离线**视频**稳定方法相比，运行速度快 3 0 倍左右。此外，拟议的 stabnet 能够处理夜间和模糊**的视频**，其中现有的方法在鲁棒功能匹配中失败。少

2018 年 2 月 22 日提交;**最初宣布** 2018 年 2 月。

239. 用于行动识别的学习代表时间特征

作者:ali javidani, ahmad Mahmoudi-Aznaveh

文摘: 本文提出了一种新的**视频**分类方法，旨在有效地识别不同类别的第三人称**视频**。这样做的目的是通过随着时间的推移跟踪光学流元素来跟踪**视频**中的运动。为了有效地对产生的运动时间序列进行分类，我们的想法是让机器学习沿时间维数的时间特征。这是通过训练多通道一维卷积神经网络 (1d-inn) 来实现的。由于 cnn 按层次表示输入数据，因此通过进一步处理较低级别层中的要素来获得高级要素。因此，在时间序列的情况下，从短期时间特征中提取长期时间特征。此外，该方法相对于大多数**基于深度学习**的方法相比的优越性在于，我们只尝试**沿着**时间维度学习具有代表性的时间特征。这显著减少**了学习**参数的数量，从而使我们的方法能够在更小的数据集上进行培训。通过更有效的特征向量表示，该方法可以在两个公共数据集 ucf11 和 jhmdb 上获得最先进的结果。少

2018 年 3 月 14 日提交;**v1** 于 2018 年 2 月 19 日提交;**最初宣布** 2018 年 2 月。

240. 利用乌尔都语视听信息进行唇部阅读深度学习

作者:m faisal, sanaullah manzoor

摘要: 读唇是一项具有挑战性的任务。它不仅需要对基础语言的了解，还需要视觉线索来预测口语。专家需要一定程度的经验和对视觉表达的理解,**学习**解码口语。现在，在**深入学习**的帮助下，可以将嘴唇序列转化为有意义的单词。在嘈杂的环境中，语音识别可以随着视觉信息的增加而增加 [1]。为了证明这一点，在这个项目中，我们尝试了两种不同的**深层学习**模型进行唇读: 第一个用于使用时空卷积神经网络的**视频**序列，双门置复发神经网络和连接器时间分类丢失，其次是将 mfcc 功能输入到 lstm 单元层并输出序列的音频。我们还收集了一个小型视听数据集，以培训和测试我们的模型。我们的目标是整合我们的两个模型，以改善在嘈杂环境中的语音识别少

2018 年 2 月 15 日提交;**最初宣布** 2018 年 2 月。

241. gep-pg: 深度强化学习算法中的解耦探索与开发

作者:cédric colas, olivier sigaud , pierre-yves oudeyer

文摘: 在连续行动领域，像 ddpg 这样的标准**深度强化学习**算法在面临稀疏或欺骗的奖励问题时，会受到低效的探索。相反，进化和发展方法侧重于探索，如新颖性搜索，质量多样性或目标探索过程探索更有力，但效

率较低的微调政策使用梯度下降。在本文中，我们提出了 gep-pg 方法，通过依次结合一个目标探索过程和两个 ddpg 变体，充分利用了这两个领域的优点。我们研究这些组件的**学习**性能及其组合在一个低维度的欺骗奖励问题和更大的半猎豹基准。我们表明 ddpg 在前一种情况下失败，并且 gep-pg 在这两种环境中都优于最佳 ddpg 变体。补充**视频**和讨论可在 http://frama.link/gep_pg 找到，代码在 http://github.com/flowersteam/geppg。少

2018 年 9 月 20 日提交;**v1** 于 2018 年 2 月 14 日提交;**最初宣布** 2018 年 2 月。

242. 张量理解：框架-不可知性高性能机器学习抽象

作者:nicolas vasilache, oleksandr zinenko, theodoros theodoridis, priya goyal, zachary devito, william s. moses, sven verdoolaege, andrew adams, albert 科恩

摘要: **深部**具有卷积和递归网络的**学习**模型现已无处不在，可分析大量的音频、图像、**视频**、文本和图形数据，并在自动翻译、语音到文本、场景理解等方面得到应用，排名用户首选项、广告展示位置等。构建这些网络的竞

争框架，如 ensorflow、chainer、cntk、torch/pytorch、Caffe1/2、mxnet 和 theano，探讨了可用性和表现力、研究或生产方向以及支持的硬件之间的不同权衡。它们在计算运算符的 dag 上运行，包装高性能库，如 cudnn (适用于 nvidia gpu) 或 nnnpack (适用于各种 cpu)，并自动进行内存分配、同步和分发。如果计算不适合现有的高性能库调用 (通常是以高昂的工程成本)，则需要自定义运算符。当研究人员发明新的运营商时，这往往是必要的：这类运营商受到严重的性能处罚，这限制了创新的速度。此外，即使这些框架可以使用现有的运行时调用，它通常也不能为用户的特定网络体系结构和数据集提供最佳性能，缺少操作员之间的优化以及可以执行的优化了解数据的大小和形状。我们的贡献包括 (1) 一种接近深度**学习**数学的语言，叫做 "归纳理解"，(2) 一种多面体实时编译器，用于转换深度**学习**的数学描述 dag 进入具有委派内存管理和同步的 cuda 内核，还提供了优化，如特定大小的运算符融合和专业化，(3) 由自动程序填充的编译缓存。[抽象截止]少

2018 年 6 月 28 日提交;**v1** 于 2018 年 2 月 13 日提交;**最初宣布** 2018 年 2 月。

243. rsdnet: 学习预测在没有手动批注的腹腔镜视频中剩余手术持续时间

作 者 :an 筒 鲁 putra twinanda, gaurav yengera, didier mutter, jacques Marescaux, nicolas padoy

摘要: 目的: 准确的手术持续时间估计是优化 or 规划所必需的，对患者的舒适性和安全性以及资源优化具有重要作用。然而，术前预测手术时间是很有挑战性的，因为它因患者病情、外科医生技能和术中情况的不同而有很大差异。我们提出了一种术中估计剩余手术持续时间的方法，这非常适合在 or 中部署。方法: 我们提出了一个**深度学习**管道，名为 rsdnet，它自动估计术中剩余的手术时间，只使用腹腔镜**视频**的视觉信息。rsdnet 的一个有趣的特点是，它在训练过程中不依赖于任何手动注释。结果: 实验结果表明，该网络明显优于外科手术设施中常用的估计手术持续时间的方法。此外，该方法的通用性证明了通过测试管道两个大型数据集，其中包含不同类型的手术，120 份胆囊切除术和 170 胃旁路**视频**。结论: 创建人工注释需要专家知识，这是一个耗时的过程，特别是考虑到在医院进行的手术类型众多，以及可提供的大量腹腔镜**视频**。由于拟议的

管道并不依赖于人工注释，因此很容易扩展到许多类型的手术。意义: rsdnet 具有卓越的性能和有效扩展到多种手术的能力，可以开发出改进的 or 管理系统。少

2018 年 2 月 9 日提交;**最初宣布** 2018 年 2 月。

244. 学习得分花样滑冰运动视频

作者:徐成明,傅燕薇,张兵，陈子田,姜玉刚,向阳雪

文摘: 本文以学习花样滑冰运动**视频得分**为**目标**。为了解决这一任务，我们提出了一个包含两个互补组件的**深层**体系结构，即自衰减 lstm 和多尺度卷积跳过 lstm。这两个组件可以有效地**了解**每个**视频**中的本地和全局顺序信息。此外，我们提出了一个大规模的花样滑冰运动**视频**数据集--fsv 数据集。此数据集包括 500 个花样滑冰**视频**，平均长度为 2 分 50 秒。每个**视频**由 9 个不同的裁判分分 (即总元素得分 (tes) 和总程序组件得分 (pcs)) 注释。我们提出的模型在 fsv 和 mit 滑板数据集中进行了验证。实验结果表明，我们的模型在**学习**得分花样滑冰**视频**方面是有效的。少

2018 年 7 月 28 日提交;**v1** 于 2018 年 2 月 8 日提交;**最初宣布** 2018 年 2 月。

245. 递归神经网络的有效量化方法

作 者 :md zahangir alom, adam t moody, naoya maruyama, brian c van essen, tarek m. taha

摘要: **深部学习**，特别是递归神经网络 (rnn) 在机器翻译、语言理解和电影框架生成等各种任务中表现出卓越的准确性。然而，这些**深度学习**方法在计算方面非常昂贵。在大多数情况下，图形处理单元 (gpu) 用于大规模实现。同时，提出了在包括现场编程门阵列 (fpga) 和移动平台在内的特殊硬件上部署解决方案的节能 rnn 方法。本文提出了一种有效的递归神经网络 (rnn) 技术的量化方法，包括长期短期存储器 (lstm)、g 平价递归单元 (gru) 和卷积长期短期存储器 (convlstm)。我们实现了不同的量化方法，包括二进制连接 {-1, 1}、三元连接 {-1, 0, 1} 和第四纪连接 {-1,-0.5, 0.5, 1}。对这些建议的方法进行了对不同数据集的评估，以便对 imdb 进行情绪分析，并对移动的 mnist 数据集进行视频帧预测。将实验结果与 lstm、gru 和 convlstm 的全精度版本进行了比较。它们在情绪分析和**视频**帧预测方面都显示出了很有希望的结果。少

2018 年 2 月 7 日提交;**最初宣布** 2018 年 2 月。

246. 社交 ml：社交媒体视频创作者的机器学习

作 者 :tomasz trzcinski, adam bielski, pawel cyrta, matthew zak

摘要: 近年来，社交媒体已成为数十亿用户出版和消费创意内容的主要场所之一。与传统媒体相反，社交媒体使出版商能够以前所未有的规模获得关于其创作工作的几乎即时反馈。这是机器**学习**方法的一个完美用例，可以利用这些海量数据为内容创作者提供鼓舞人心的想法和对其工作的建设性批评。在这项工作中，我们全面概述了我们为第九集团 media 的**视频**创作者开发的机器**学习**授权工具--这是创建包含三个以上视频的主要社交媒体公司之一每月 10 亿次浏览。我们的主要贡献是一套工具，允许创作者利用海量数据来改进他们的创建过程，在出版前评估他们的**视频**，并提高内容质量。这些应用程序包括一个交互式对话机器人，允许访问材料存档，一个基于 web 的应用程序，用于自动选择最佳**的视频**缩略图，以及**深度学习**方法优化标题和预测**视频**受欢迎程度。我们的 aob 测试显示，部署我们的工具可显著增加 12.9 的**平均视频**视图计数。我们的额外贡献是在部署这些工具的过程中收集的一组考虑因素，这些考虑因素可以降低

2018 年 1 月 25 日提交;最初宣布 2018 年 2 月。

247. 每一个微笑都是独一无二的: 以地标为导向的多样化微笑一代

作者: 王伟, xavier alameda-pineda, dan xu, pascal fua, elisa ricci, nicu sebe

摘要: 每一个微笑都是独一无二的: 一个人肯定会以不同的方式微笑 (例如, 闭上眼睛或嘴巴)。给定一个中性面的输入图像, 我们能否生成具有独特特征的多个微笑**视频**? 针对这种一对多视频生成问题, 提出了一种新的**深度学习**体系结构--条件多模网络 (cmm-net)。为了更好地编码面部表情的动态, cmm-net 明确地利用面部地标生成微笑序列。具体来说, 变分自动编码器用于**学习**面部地标嵌入。然后, 条件递归网络利用这种单一的嵌入, 该网络生成一个以特定表达式 (例如自发微笑) 为条件的地标嵌入序列。接下来, 生成的地标嵌入被输入到一个多模式的递归地标生成器中, 生成一组仍与给定的微笑类相关但彼此明显不同的地标序列。最后, 这些地标序列被翻译成面部**视频**。我们的实验结果证明了我们的 cmm-net 在生成多个微笑表达式**的逼真视频**方面的有效性。少

2018 年 3 月 28 日提交;**v1** 于 2018 年 2 月 6 日提交;**最初宣布** 2018 年 2 月。

248. 通过深层强化学习实现共享自治

作者: Siddharth reddy, anca d. dragan, sergey levine

摘要: 在共享自治中，用户输入与半自主控制相结合，以实现共同的目标。目标通常是未知的事前，因此之前的工作使代理能够从用户输入中推断目标并协助完成任务。这类方法倾向于假定对环境动态、用户的策略(给定其目标) 以及用户可能针对的一组可能目标的某种知识组合，这些目标将其应用程序限制在实际场景中。我们提出了一个**深度**强化学习框架，无模型共享**自治,**以提升这些假设。我们使用具有神经网络函数近似**的**人在环强化学习,**学习**从环境观察和用户输入到代理操作值的端到端映射，任务奖励是唯一的形式。监督。这种方法带来了足够严格地遵循用户命令的挑战，以便为用户提供实时操作反馈，从而确保高质量的用户输入，但在用户的操作不理想时也会偏离这些命令。我们通过放弃值低于某个阈值的操作，然后选择最接近用户输入的剩余操作来平衡这两种需求。与用户 (n = 12) 和合成**飞行员进行控制研究** (n = 12)，以及与用户 (n = 4) 进行试点研究，使用户 (n = 4) 飞行一个真正

的四旋翼，展示了我们的算法的能力，以帮助用户执行代理无法执行的实时控制任务通过观察直接访问用户的私人信息，但接收奖励信号和用户输入，两者都取决于用户的意图。代理**学习**帮助用户而不访问此私人信息，并隐式地从用户的输入推断这些信息。本文是一个概念的证明，说明**了深层**强化**学习**的潜力，以实现灵活和实用的辅助系统。少

2018 年 5 月 22 日提交;**v1** 于 2018 年 2 月 5 日提交;**最初宣布** 2018 年 2 月。

249. 利用深度学习进行压缩光场重建

作 者 :mayank gupta, arjun jauhari , kuldeep kulkarni, suren jayasuriya, alyosha molnar, pavan turaga

摘要: 光场成像在计算处理空间和角度尺寸的高采样需求方面受到限制。单发光场摄像机牺牲空间分辨率来采样角视点，通常通过将传入射线多路复用到 2d 传感器阵列上。虽然这种分辨率可以用压缩传感来恢复，但这些迭代解在处理光场方面进展缓慢。我们提出了一种深度**学习**方法，使用一个新的两个分支网络体系结构，包括一个自动编码器和一个 4d cnn，从一个编码的 2d

图像恢复高分辨率的 4d 光场。该网络显著缩短了重建时间，同时在各种光场上实现了 26-32 db 的平均 psnr 值。特别是，与具有等效视觉质量的字典方法相比，重建时间从 35 分钟缩短到 6.7 分钟。这些重建是在低至 8% 的小采样压缩比下进行的，从而可以使用更便宜的编码光场摄像机。我们在合成光场上测试我们的网络重建，模拟从 lytro illum 相机捕获的真实光场的编码测量，以及从自定义 cmos 衍射光场相机拍摄的真实编码图像。压缩光场捕获与**深度学习**相结合，为未来实时光场**视频**采集系统提供了潜力。少

2018 年 2 月 5 日提交;**最初宣布** 2018 年 2 月。

250. 用于监测铁路操作人员班次的人脸识别

作者:s ritika, dattaraj rao

摘要: 火车飞行员是一项非常乏味和压力很大的工作。飞行员必须时刻保持警惕，很容易忘记轮班时间。在美国这样的国家，飞行员被法律授权坚持 8 小时轮班。如果他们的班次超过 8 小时，铁路可能会因为司机过度疲劳而受到处罚。当 8 小时的轮班可能在旅途中结束时，就会出现这个问题。在这种情况下，新司机必须转移到正在运行的换班机车位置。因此，在司机轮班期间对其

进行准确的监控，并确保正确安排班次，对铁路来说非常重要。在这里，我们提出了一个自动化的摄像系统，使用安装在机车驾驶室内的摄像头，以不断记录**视频**源。对这些源进行实时分析，以检测驾驶员的面部,并使用最先进的**深度学习**技术识别驱动程序。其结果是提高了火车飞行员的安全性。摄像机从驾驶室内连续捕获**存储在**车载数据采集设备上的视频。利用先进的计算机视觉和**深度学习**技术，定期对**视频**进行分析，以检测飞行员的存在并识别飞行员。使用基于时间的分析，可以确定该移位处于活动状态的时间。如果此时间超过分配的班次时间，则会向调度发送警报以调整班次时间。少

2018 年 5 月 21 日提交;**v1** 于 2018 年 2 月 5 日提交;**最初宣布** 2018 年 2 月。

251. 通过网络调制实现高效的视频对象分割

作者:杨林杰,王艳兰,熊学汉,杨建超，阿格洛斯 k. 卡卡塔格洛斯

摘要: **视频**对象分割的目标是在整个**视频**序列中分割特定对象，只提供一个带注释的第一帧。最近基于**深度学习**的方法通过使用数百次梯度下降迭代对带注释的帧上的通用分割模型进行微调发现了它的有效性。尽管这

些方法实现了高精度，但微调过程效率低下，无法满足实际应用的要求。我们提出了一种新的方法，它使用单个正向传递来调整分割模型以适应特定对象的外观。具体而言，在目标对象的视觉和空间信息有限的情况下，**学习**了第二个元神经网络，名为调制器，以操作分割网络的中间层。实验表明，我们的方法比微调方法快70times，同时实现了相似的精度。少

2018 年 2 月 4 日提交;**最初宣布** 2018 年 2 月。

252. 机器人在自闭症治疗中的情感和参与感的个性化机器学习

作 者 : Ognjen rudovic, jaeryounglee, miles dai, bjorn schuller, rosalind picard

摘要: 机器人具有很大的潜力，可以为自闭症谱系的儿童提供未来的治疗。然而，现有的机器人缺乏自动感知和响应人类影响的能力，而这对于建立和维持引人入胜的互动是必要的。此外，由于许多自闭症患者有非典型且异常多样的表达情感认知状态的方式，他们的推理挑战变得更加严峻。为了解决自闭症儿童行为暗示的异质性，我们利用**深度学习**的最新进展，制定了一个个性化的机器学习 (ml) 框架，以自动感知儿童机器人辅

助自闭症治疗过程中的情感状态和参与。我们方法的关键是从传统的 ml 范式进行新的转变--我们的个性化 ml 框架不是使用 "一刀切" 的 ml 模型，而是通过利用相关的上下文信息 (人口统计和行为) 为每个孩子优化的评估分数) 和每个孩子的个人特征。我们使用 35 名自闭症儿童 (3-13) 和来自 2 种文化 (亚洲和欧洲) 的多模式音频、**视频**和自主生理数据数据集设计和评估了这一框架，参与了 25 分钟的儿童机器人互动 (~ 500k) 数据点)。我们的实验证实了机器人感知影响和参与的可行性，显示出明显的改进，由于模型的个性化。该方法有可能通过提供更有效的监测和总结治疗进展，改进现有的自闭症疗法。少

2018 年 6 月 18 日提交;**v1** 于 2018 年 2 月 4 日提交;**最初宣布** 2018 年 2 月。

253. 机器人的 vr 护目镜：用于视觉控制的真实的领域适应

作者:张景伟,雷泰,熊玉峰，彭云,刘明，乔什卡·博德克，沃尔夫拉姆·伯加德

文摘: 本文从一个新的角度来处理现实差距，将**模拟**环境中**学习**的深度强化**学习**(drl) 策略转移到现实领域进

行视觉控制任务。我们试图通过将真实世界的图像流转换回合成域来解决这个问题，而不是通过提高模拟器输出的合成图像的视觉保真度来解决这个问题在部署阶段，让机器人有宾至如归的感觉。我们建议将其作为一种轻量级、灵活和高效的视觉控制解决方案，因为 1) 在对 drl 代理进行昂贵的模拟培训期间，不需要额外的传输步骤;2) 受过训练的 drl 代理不会被限制在一个特定的实际环境中部署;3) 政策培训和转移操作是脱钩的，可以并行进行。除此之外，我们还提出了一个简单而有效的换档损失，以限制后续框架之间的一致性，这对于一致的政策产出非常重要。我们验证**了视频**和领域适应艺术风格转换的移位损失，并验证了我们在室内和室外机器人实验中的视觉控制方法。我们的结果视频可在: https://goo.gl/P76TTo。少

2018 年 9 月 11 日提交;**v1** 于 2018 年 2 月 1 日提交;**最初宣布** 2018 年 2 月。

254. 图像 2gif: 利用递归深 q 网络生成电影图

作者:周一平,宋耶鲁,塔玛拉 l. berg

摘要: 给定静止图像，可以想象动态对象在静态背景下的移动情况。这个想法已经以电影的形式实现，在静止

图像中特定物体的运动被重复，给观众一种动画的感觉。在本文中，我们**学习**了计算模型，可以产生电影序列自动给定的单个图像。为了生成电影，我们探索将生成模型与递归神经网络和**深度** q 网络相结合，以增强序列生成的能力。为了启用和评估这些模型，我们使用了两个数据集，一个是合成生成的，另一个包含真实**的视频**生成的电影。定性和定量评估都证明了我们模型在合成数据集和真实数据集上的有效性。少

2018 年 1 月 27 日提交;**最初宣布** 2018 年 1 月。

255. 超能力：通过切割 cnn 预测平滑的追求为基础的注意力，提高自然主义视频的固定预测

作者:mikhail startsev, michael dorr

摘要: 预测注意力是人类和计算机视觉交汇的热门话题，但**视频**显著性预测直到最近才开始从**基于深度学习**的方法中受益。尽管大多数可用**的基于视频**的显著性数据集和模型声称以人类观察者的固定为目标，但它们未能将其与平滑追求 (sp) 区分开来，平稳追求 (sp) 是一种主要的眼动类型，是动态场景感知所独有的。在这项工作中，我们的目标是明确这一区别，为此，我们 (i) 使用算法和手动注释 sp 跟踪和其他眼动为两个公认

的**视频**显著性数据集，(ii) 训练切片卷成神经网络(s-cnn)，用于对固定点或 sp 突出位置进行显著性预测，并 (iii) 评估我们的和 20 多个流行的已发布的显著性模型，用于预测 sp 和固定点，以及另一个人类数据集固定。我们提出的模型在一组独立的**视频**上进行了培训，在所有经过考虑的数据集的 sp 预测任务中，其性能优于最先进的显著性模型。此外，该模型还展示了在预测基于 "经典" 定位的显著性方面的卓越性能。我们的研究结果强调了有选择地进行训练集构建以进行注意力建模的重要性。少

2018 年 1 月 29 日提交;**v1** 于 2018 年 1 月 26 日提交;**最初宣布** 2018 年 1 月。

256. 让我们跳舞：从网络舞蹈视频中学习

作 者 :daniel castro, steven hickson, patsorn sansklоy, bhavishya mittal, sean dai, james hays, irfan essa

文摘: 近年来,**深度**神经网络方法通过聚合每帧分类作为操作识别的基线，自然扩展到**视频**领域。这一领域的大部分工作都来自成像领域，导致对时间数据采取视觉特征重的方法。为了解决这个问题，我们引入了 "让

我们跳舞"，这是一个 1000 个**视频**数据集（并不断增长），由 10 个视觉上重叠的舞蹈类别组成，需要运动进行分类。我们强调了人类运动作为我们工作中的关键区别的重要性，因为正如我们在这项工作中所表明的那样，视觉信息不足以对运动重的类别进行分类。我们使用 ucf-101 成像技术比较数据集的性能，并证明了这一固有的困难。为了分析这些方法，我们提出了我们数据集上的许多最新技术的比较，使用三种不同的表示形式（**视频**、光流和多人姿态数据）。我们讨论了它们各自的运动参数化及其在**学习**在线**舞蹈视频分类中**的价值。最后，我们发布此数据集（及其三种表示形式），供研究社区使用。少

2018 年 1 月 22 日提交;**最初宣布** 2018 年 1 月。

257. 使用生成对抗性网络的视频半监督行动识别

作者:unaiza ahsan, chen sun, irfan essa

文摘: 我们提出了一个使用基因再生对抗性网络的行动识别框架。我们的模型包括使用没有 label 信息的大型**视频**活动数据集培训深层卷积生成敌对网络（dcgan）. 然后，我们使用 gan 模型中的训练鉴别器作为无监督的训练前步骤，并在标记的数据集上微调训

练的鉴别器模型，以识别人类活动。我们确定了良好的网络体系结构和超参数设置，以便将 dcgan 的鉴别器用作训练模型，以**学习用于操作**识别的有用表示。我们仅使用外观信息的半监督框架在两种具有挑战性的**视频**活动数据集上实现了与当前最先进的半监督动作识别方法相比的卓越或可比性能: ucf101 和 hmdb51。少

2018 年 1 月 22 日提交;**最初宣布** 2018 年 1 月。

258. 一种基于多视点接收场的视频监控的面向推理网络

作者:thangarajah akilan

文摘: 前景 (fg) 像素标记在**视频**监控中起着至关重要的作用。最近的工程解决方案试图利用**深度学习**(dl) 模型的有效性，这些模型最初的目标是图像分类，以处理 fg 像素标记。这种策略的一个主要缺点是，在训练样本有限的情况下，缺乏视觉对象的划分。为了解决这个问题，我们引入了一个多视点接受场完全卷积神经网络(mv-fcn)，利用最近的开创性思想，如完全卷积结构、初始模块和残差网络。由此，我们以编码器解码器的方式实现了一个包含核心和两个互补特征流路径的系统。该模型利用了具有三种不同接受场大小的初始模块，在不同尺度上捕获不变性。在编码阶段**学到**的特征通过剩

余连接与解码阶段的适当要素图融合在一起，以实现增强的空间表示。这些多视图接受场和剩余特征连接有望产生高度通用的特征，以实现精确的像素明智的 fg 区域识别。因此，它接受了数据库特定的示例式分割的训练，以预测所需的 fg 对象。11 个基准数据集的对比实验结果验证了该模型在先验和最先进的算法下具有很强的竞争性能。我们还报告说，转移**学习**方法在多大程度上有助于提高我们提出的 mv-fcn 的性能。少

2018 年 1 月 19 日提交;**最初宣布** 2018 年 1 月。

259. 基于欺骗树、遗传规划和神经网络的感知视听质量建模

作者:edip demirbilek, jean-charles grégoire

摘要: 我们的目标是构建基于机器**学习**的模型，直接从从目标质量数据集提取的一组相关参数中预测视听质量。我们使用了 inrs 视听质量数据集的比特流版本，该版本反映**了视频**帧速率、**视频**量化、降噪参数和网络数据包丢失率的当代实时配置。我们利用这个数据集建立了基于随机森林、套袋、**深度学习**和遗传规划方法的比特流感知质量估计模型。我们采取了经验方法，并根据质量数据集使用的要素数量生成了各种模型，从非

常简单到最复杂。随机林和套袋模型在 rmse 和 pearson 相关系数值方面总体上产生了最准确的结果。**深部**基于**学习**和遗传规划的比特流模型也取得了良好的效果，但只有在功能范围有限的情况下才能观察到高性能。我们还得到了每个模型的表观不敏感 rmse 值，并计算了相关系数之间差异的意义。总的来说，我们的结论是，计算比特流信息是值得的，它所需的努力，以生成和帮助建立更准确的模型的实时通信。但是，它仅适用于使用精心选择的功能子集部署正确的算法。本研究期间开发的数据集和工具可用于研究和开发目的。少

2017 年 12 月 5 日提交;**最初宣布** 2018 年 1 月。

260. 原始眼：利用图像特征和眼动分析保留优先的第一人称视觉

作 者 :julian steil, marion koelle , wilko heuten, susanne boll, andrias bulling

摘要: 随着头戴式显示器中的第一人称摄像机越来越普遍，侵犯用户和旁观者隐私的问题也越来越普遍。为了应对这一挑战，我们推出了 privaceye，这是一个概念验证系统，可检测隐私敏感的日常情况，并使用机械快门自动启用和禁用第一人称相机。为了关闭快门，

privaceye 使用端到端深度**学习**模型检测第一人称相机**视频**中的敏感情况. 为了在没有视觉输入的情况下打开快门, privaceye 使用单独的、较小的眼相机来检测用户眼动的变化，以衡量当前情况下 "隐私级别" 的变化。我们根据 17 名参与者日常生活中记录的第一人称视频数据集对 privaceye 进行评估，这些视频带有隐私敏感性级别的注释。我们在定量技术评价以及半结构化访谈的定性见解的基础上讨论概念验证系统的优缺点。少

2018 年 1 月 13 日提交;**最初宣布** 2018 年 1 月。

261. 监控视频中的真实世界异常检测

作者:waqas sultani, chen chen , mubarak shah

摘要: 监控**视频**能够捕获各种逼真的异常。在本文中，我们建议通过利用正常和异常**视频**来**学习**异常。为了避免在培训**视频**中注释异常段或剪辑，这非常耗时，我们建议通过深度多实例排名框架，利用弱标记的训练来**学习**异常**视频**，即训练标签 (异常或正常) 是在**视频**级别，而不是剪辑级别。在我们的方法中，我们将正常和异常**的视频**视为包和**视频**段作为多个实例**学习**(mil) 中的实例，并自动**学习**一个**深度**异常排名模型,预测异常

视频段的异常分数很高。此外，我们还引入了排名损失函数中的稀疏和时间平滑约束，以更好地定位训练过程中的异常。我们还推出了一个新的大规模的第一个数据集的 128 位的 128 位视频。它由 1900 个长而未修剪的现实世界监控**录像**组成，有 13 个现实的反常现象，如战斗、交通事故、入室盗窃、抢劫等以及正常活动。此数据集可用于两个任务。首先，考虑到一个组的所有异常和另一个组的所有正常活动的一般异常检测。第二，用于识别 13 个异常活动中的每一个。实验结果表明，与最先进的方法相比，我们的 mil 异常检测方法在异常检测性能上有了显著的提高。我们提供了最近几个**关于**异常活动识别的深度**学习**基线的结果。这些基线的识别性能较低，这表明我们的数据集非常具有挑战性，并为今后的工作提供了更多的机会。数据集可在以下 http://crcv.ucf.edu/projects/real-world/

2018 年 3 月 31 日提交;**v1** 于 2018 年 1 月 12 日提交;**最初宣布** 2018 年 1 月。

262. 深度搜索：基于内容的图像搜索和检索

作者:tanya piplani, david bamman

摘要: 今天的互联网大多由**包括视频**和图像在内的数字媒体组成。随着像素成为大多数交易在互联网上发生的货币，有一种相对轻松地浏览这片信息海洋的方式变得越来越重要。youtube 每分钟上传 400 个小时的**视频**，在 instagam、facebook 等网站上浏览了数百万张图片。在**深刻学习**和成功领域的最新进步的启发下，我们在图像字幕、机器翻译、单词 2vec、跳过思想等各种问题上取得了进步，我们提出了一种深度语言处理基于**深度学习**模型，允许用户输入要搜索的图像类型的描述，并作为响应，系统检索所有与查询在语义上和上下文上相关的图像。以下各节介绍了两种方法。少

2018 年 1 月 11 日提交;**v1** 于 2018 年 1 月 9 日提交;**最初宣布** 2018 年 1 月。

263. 基于深度学习的视频无监督和半监督异常检测方法综述

作 者 :b ravi kiran, dilip mathew thomas , ranjith parakkal

摘要: **视频**是监控应用的主要信息来源，可大量获得，但在大多数情况下，很少或根本没有监督**学习**的注释。本文综述了最先进的**基于深度学习的视频**异常检测方

法，并根据检测的模型类型和检测标准对其进行了分类。我们还进行了简单的研究，以了解不同的方法，并提供评估标准的时空异常检测。少

2018 年 1 月 30 日提交;**v1** 于 2018 年 1 月 9 日提交;**最初宣布** 2018 年 1 月。

264. remotenet: 大规模家庭监控视频的高效相关运动事件检测

作者:余瑞一,王洪成，蔡世戴维斯

文摘: 本文讨论了在大型家庭监控录像中检测**感兴趣**的物体 (例如人和车辆) 引起的相关运动的问题。传统的方法通常由两个不同的步骤组成，即检测在相机上运行背景减法的运动物体，并过滤出**带有深度的扰民运动事件 (如树木、云、阴影、雨雪、旗)基于学习**的目标检测和跟踪在云上运行。该方法极其缓慢，因此不具有成本效益，并且没有完全利用时空冗余与预先训练的现成对象探测器。为了显著加快相关运动事件检测并提高其性能，我们提出了一种新的相关运动事件检测网络 remotenet，这是一种基于时空注意的 3d 的统一的端到端数据驱动方法在**视频**中联合建模感兴趣的对象的外观和运动。remotenet 在神经网络的一个正向传递

中分析整个**视频**剪辑，以实现显著的加速。同时，它利用家庭监控**视频**的特性，例如，相关运动在空间和时间上稀疏，并利用时空注意模型和参考框架减法增强三维卷网，以鼓励网络，以关注相关的运动对象。实验表明，与基于目标检测的方法相比，我们的方法可以实现可比或事件更好的性能，但在 gpu 设备上的速度会达到 3 到 4 个数量级 (高达 20k 倍)。我们的网络高效、紧凑、重量轻。它可以在 gpu 上的 4-8 毫秒内检测 15s 监控视频剪辑上的相关运动,在模型大小小于 1mb 的 cpu 上检测秒 (0.17-0.39) 的一小部分 (0.17-0.39)。少

2018 年 1 月 6 日提交;**最初宣布** 2018 年 1 月。

265. 三维以太网：一种单级视频车载检测器

作者:李绥昌

文摘: **基于视频**的车辆检测在过去的十年里受到了相当多的关注，并有许多研究结果。更多

2018 年 1 月 15 日提交;**v1** 于 2018 年 1 月 5 日提交;**最初宣布** 2018 年 1 月。

266. 我们从行动识别的深度表示中学到了什么?

作 者 :christoph feichtenhofer, axel pinz, richard p. wildes, andrewzisserman

摘要: 由于**深度**模型的成功导致它们部署在计算机视觉的所有领域，了解这些表示是如何工作的以及它们捕获的内容变得越来越重要。在本文中，我们通过可视化双流模型所**学到**的东西来揭示深层时空表示，以识别**视频**中的动作。我们表明，局部探测器的外观和运动对象出现，形成分布式表示识别人类的行动。主要意见包括以下内容。首先，跨流融合可以**学习**真正的时空特征，而不是简单地分离外观和运动特征。其次，网络可以学习具有高度类特定的本地表示形式，还可以**学习**可为一系列类提供服务的泛型表示形式。第三，在整个网络层次结构中，要素变得更加抽象，并对数据中对所需区别不重要的方面表现出越来越大的不变性 (例如，不同速度的运动模式)。第四，可视化不仅可以用来揭示**学习的**表示，还可以用来揭示训练数据的特性，并解释系统的故障案例。少

2018 年 1 月 4 日提交;**最初宣布** 2018 年 1 月。

267. 以多元化代表性奖励的无监督视频总结深层强化学习

作者:周开阳,余桥,陶翔

摘要: **视频**摘要旨在通过制作多样化、代表性丰富的视频的简短摘要，方便大规模**视频**浏览。本文将**视频**摘要作为一个连续决策过程，并开发了一个**深度**摘要网络(dsn) 来总结**视频**。dsn 预测每个**视频**帧的概率，该概率指示选择帧的可能性，然后根据概率分布采取行动选择帧，形成**视频**摘要。为了培训我们的 dsn，我们提出了一个端到端的强化**学习**框架，在这个框架中，我们设计了一个新的奖励函数，共同考虑到生成摘要的多样性和代表性，而不依赖于标签或用户在所有的互动。在培训过程中，奖励职能判断是多么多样化和代表性的生成的摘要，而 dsn 努力通过**学习**产生更多样化和更有代表性的摘要来获得更高的奖励。由于标签不是必需的，我们的方法可以完全无人监督。在两个基准数据集上进行的大量实验表明，我们的无监督方法不仅优于其他最先进的无监督方法，而且与大多数已发布的监督方法相当，甚至优于大多数已发布的监督方法。少

2018 年 2 月 13 日提交;**v1** 于 2017 年 12 月 29 日提交;**最初宣布** 2018 年 1 月。

268. 学习深入而紧凑的手势识别模型

作者:koustav mullick, anop m. Namboodiri

摘要:我们研究了在**深度学习**框架中从**视频**中开发紧凑、准确的手势识别模型的问题。为此，我们提出了一个联合 3dcnn-lstm 模型，是端到端可培训，并被证明是更适合捕捉动态信息在..。更多

2017 年 12 月 29 日提交;**最初宣布** 2017 年 12 月。

269. 用于快速视频检索的类别掩码深度哈希

作者:徐柳,赵丽丽,丁大军,董亚娇

文摘: 本文提出了一种具有类别掩码的端到端深度哈希框架，用于快速**视频**检索。我们通过充分利用阶级间的多样性和阶级内的身份，以监督的方式培训我们的网络。对分类损失进行了优化，以最大限度地提高类间多样性，同时引入了对，以**学习**具有代表性的类内标识。我们研究了与类别相关的二进制位分布，发现二进制位的有效性与数据类别高度相关，有些位可能会降低某些类别的分类性能。然后，我们设计了带有类别掩码的哈希代码生成方案，以筛选出具有负贡献的位。实验结果表明，该方法在公共数据集的各种评价指标下优于几种先进的方法。少

2018 年 5 月 24 日提交;**v1** 于 2017 年 12 月 22 日提交;**最初宣布** 2017 年 12 月。

270. 铁路轨道开关一次性检测的暹罗神经网络

作者:dattaraj j rao, shruti mittal, s. ritika

摘要: **深部学习**方法已被广泛用于分析视频数据，通过对**图像**帧进行分类和检测对象来提取有价值的信息。我们描述了一种独特的方法，使用视频馈送从移动机车连续监测铁路轨道和检测重要的资产，如开关上的轨道。这里使用的技术被称为暹罗网络，它使用两个相同的网络来**学习**两个图像之间的相似性。在这里，我们将使用一个暹罗网络来持续比较轨道图像，并检测轨道中的任何显著差异。交换机将是不同的图像之一，我们将找到一个映射，明确区分交换机与其他可能的轨道异常。然后将推广同样的方法，以检测铁路轨道上的任何异常。铁路运输是独特的，因为它有轮式车辆，火车由机车拉，在接近每小时 200 英里的高速上在有制导的铁轨上运行．铁路网络上的多个轨道使用一种名为 switch 或道岔。交换机可以手动操作，也可以通过控制中心的命令自动操作，它控制列车在网络的不同轨道上的移动。这些开关的准确位置对铁路来说非常重要，在现场真实了解它们的状态很重要。现代列车使用面向

轨道的高清**摄像机**，不断从轨道上录制**视频**。我们使用暹罗网络并与基准图像进行比较，描述了一种监视跟踪和突出显示异常的方法。少

2017 年 12 月 21 日提交;**最初宣布** 2017 年 12 月。

271. 通过深度网络增强恶劣条件下的视觉识别

作者:刘丁,程宝文,王章阳,张海超,黄晓明

摘要: 在图像采集、传输或存储过程中，由于质量失真的存在，在恶劣条件下的视觉识别是一个非常重要且具有挑战性的问题，具有很高的实用价值。虽然**深**部神经网络分别在低质量图像恢复和高质量图像识别任务技术中得到了广泛的应用，但从低质量的图像。本文提出了一种**基于深度学习**的框架，利用鲁棒的不良训练或其攻击性变体，提高图像和**视频**识别模型在恶劣条件下的性能。强大的不良训练前算法利用预训练的力量，并推广传统的非监督训练前和数据增强方法。我们进一步开发了一种转移**学习**方法，以应对未知不利条件下的真实数据集。对拟议框架进行了一些图像和**视频**识别基准的全面评估，并在各种单一或混合不利条件下获得了显著的性能改进。我们的可视化和分析进一步增加了结果的可解释性。少

2017 年 12 月 20 日提交;**最初宣布** 2017 年 12 月。

272. 基于递归控制的模拟视图不变视觉服务

作 者 :fereshteh sadeghi, 亚 历 山 大 · 托 舍 夫 , eric jang, sergey levine

摘要: 即使在光学扭曲的情况下，人类也非常善于从广泛的角度和角度控制自己的四肢和工具。在机器人技术中，这种能力被称为视觉伺服：主要使用视觉反馈将工具或终点移动到所需位置。在本文中，我们研究了如何在机器人操作场景中自动**学习**视点不变的视觉伺服技巧。为此，我们训练一个**深度**的递归控制器，它可以自动确定哪些动作将机器人手臂的终点移动到所需的对象。该控制器必须解决的问题基本上是不明确的：在观点的严重变化下，可能无法确定单个前馈操作中的操作。相反，我们的视觉伺服系统必须利用过去动作的记忆，从当前的角度了解动作是如何影响机器人运动的，纠正错误，逐渐向目标靠拢。这种能力与大多数视觉伺服方法形成了鲜明的对比，后者要么假定已知的动态，要么需要校准阶段。我们展示了如何使用模拟数据和强化**学习**目标**来学习**这种递归控制器。然后，我们描述如何将生成的模型从控制中分离出来，只适应视觉层，从而将生成的模型转移到现实世界的机器人中。在现实世

界的库卡·iiwa 机器人手臂上，这种适应模型可以从新的角度向以前看不见的物体伺服。有关补充**视频**，请参见 : https://fsadeghi.github.io/Sim2RealViewInvariantServo

2017 年 12 月 20 日提交;**最初宣布** 2017 年 12 月。

273. 深度学习在现代推荐体系中的运用--对近期工作的总结

作者:ayush singhal, pradeep sinha, rakesh pant

摘要: 随着互联网上数字信息量的呈指数级增长，网上商店、在线音乐、视频和图片库、搜索引擎和推荐系统已成为查找相关信息的最便捷方式在很短的时间内。近年来,**深度学习的**进步在语音识别、图像处理和自然语言处理等领域受到了广泛关注。同时，最近的几项研究表明,**深度学习**在推荐系统和信息检索领域也很有用。在这个简短的回顾中，我们介绍了在推荐**领域** 使用各种学习技术变体取得的最新进展。本文从协同系统、基于内容的系统和混合系统三个部分对本文进行了回顾。本文还讨论了**深度学习**集成推荐系统对多个应用领域的贡献。审查最后讨论了在各个领域的推荐系统中**的深**

度学习的影响，以及**深度学习**是否比传统系统有任何重大改进以获得建议。最后，根据推荐系统中**深度学习**的使用现状，我们还提供了未来可能的研究方向。少

2017 年 12 月 20 日提交;**最初宣布** 2017 年 12 月。

274. 基于深度学习的铁路轨道特定交通信号选择

作者:s ritika, shruti mittal, dattaraj rao

摘要: 随着铁路运输行业积极走向自动化，准确定位和库存路边轨道资产，如交通信号，道口，开关，里程等是至关重要的。随着新的积极列车控制 (ptc) 规定的生效，许多铁路安全规则将直接与里程和信号等资产的位置挂钩。将根据列车在路边资产方面的位置执行较新的速度规定。因此，铁路必须有一个关于这些资产的类型和地点的准确数据库。本文讨论了一个真实世界的用例，即从安装在移动机车上的摄像机检测铁路信号并跟踪其位置。该摄像机经过设计，可承受行驶中列车上的环境因素，并以每秒 30 帧左右的速度提供一致的稳定图像。利用先进的图像分析和**深度学习**技术，在这些相机图像中检测到信号，并创建了其位置的数据库。铁路信号与道路信号在形状和轨道放置规则方面有很大差异。由于空间限制和城市地区的交通密度，信号没有

放在轨道的同一侧，多条线路可以平行运行。因此，有必要将检测到的信号与列车运行的轨道联系起来。我们提出了一种方法，将信号与它们所属的特定轨道联系起来，使用安装在引线机车上的前置摄像头的**视频**馈送。一个轨道检测、感兴趣的区域选择、信号检测的管道已经实现，在覆盖 150 公里的线路上，总的精度为 94.7%，有 247 个信号。少

2017 年 12 月 17 日提交;**最初宣布** 2017 年 12 月。

275. 城市场景理解的自我监督相关深度学习

作 者 ： 姜 怀 祖 ， erik leuned-miller， gustav larsson, michael maire, greg shakhnarovich

摘要: 当一个代理在世界中移动时，场景元素的表观运动 (通常) 与它们的深度成反比。**学习**代理自然会将图像模式与它们随时间的位移程度联系起来: 随着代理的移动，遥远的山脉不会移动太多;附近的树移动了很多。物体的出现与它们的运动之间的这种自然关系是关于世界的丰富信息来源。在这项工作中，我们首先训练一个**深**网络，利用全自动监控，从单一图像中预测相对场景深度。相对深度训练图像自动从汽车在场景中移动的简单**视频**中获得，使用的是最近的运动分割技术，而

不是人为提供的标签。这种从单个图像预测相对深度的代理任务会引入网络中的特征，从而在一组下游任务(包括语义分割、联合道路分割和汽车检测以及单目（绝对））中实现了大量改进深度估计，通过从零开始训练的网络。语义分割任务的改进比任何其他自动监督方法所产生的改进都要大。此外，对于单目深度估计，我们的无监督训练前方法甚至优于 imagenet 的监督训练前训练。此外，我们还展示了**在**与各种下游任务相关**的特定视频**中学习预测（无监督）相对深度的好处。我们以无监督的方式适应这些任务中的特定场景，以提高性能。总之，为了语义分割，我们在不使用监督训练前的方法中提供最先进的结果，甚至超过了监督的 imagenet 预训练模型的性能，用于单目深度估计，从而获得结果与最先进的方法相当。少

2018 年 4 月 2 日提交;**v1** 于 2017 年 12 月 13 日提交;**最初宣布** 2017 年 12 月。

276. uv-gan：用于存在不变人脸识别的对抗性紫外地图完成

作者:邓建康,程石阳,薛念南，周玉祥，斯特凡诺斯·扎费里乌

摘要: 最近提出的鲁棒三维人脸对齐方法在三维人脸模型和二维人脸图像之间建立了密集或稀疏的对应关系。这些方法的使用为面部纹理分析带来了新的挑战和机遇。特别是，通过使用拟合模型对图像进行采样，可以创建面部紫外线。不幸的是，由于自我遮挡，这样的紫外线图总是不完整的。本文提出了一个训练**深**卷积神经网络 (dcnn) 的框架，以完成从野外图像中提取的面紫外图像。为此，我们首先通过将 3d 可变形模型 (3DMM) 安装到各种多视图图像和**视频**数据集，以及利用具有 3, 000 多个身份的新 3d 数据集来收集完整的 uv 地图。其次，我们设计了一个精心设计的架构，结合本地和全球对抗 dnn,**学习**一个保存身份的面部紫外线完成模型。我们证明，通过将已完成的 uv 附加到拟合的网格并生成任意姿态的实例，我们可以增加姿势变化，用于训练**深**面识别/验证模型，并在测试过程中最大限度地减少姿势差异,从而获得更好的性能。在受控和野外紫外线数据集上进行的实验证明了我们的对抗性紫外线完成模型的有效性。我们实现了最先进的验证精度,94.05%，根据 cfp 前端轮廓协议，只有在训练期间结合姿势增加，并在测试过程中减少姿势差异。我们将发布第一个野生紫外线数据集 (我们称为 wilduv)，其

中包括来自 1, 892 身份的完整的面部紫外线地图，用于研究目的。少

2017 年 12 月 13 日提交;**最初宣布** 2017 年 12 月。

277. minos：复杂环境下导航的多模态室内模拟器

作者:manolis savva, angel x. chang,亚历克西·多索维茨基, thomas funkhouser, vladlen koltun

摘要: 我们介绍了 minos，这是一个模拟器，旨在支持在复杂的室内环境中开发多感官模型，用于目标导航。模拟器利用复杂 3d 环境的大型数据集，并支持多模式传感器套件的灵活配置。我们使用 minos 对**基于深度学习**的导航方法进行基准测试，分析环境复杂性对导航性能的影响，并对传感器运动学习中的多模态进行控制研究.实验表明，目前的**深层**强化**学习**方法在较大的现实环境中失败。实验还表明，多模态有利于**学习**浏览杂乱的场景。minos 在 http://minosworld.org 向研究界发布开源信息。**视频**显示 minos 可以在 https://youtu.be/c0mL9K64q84 少找到

2017 年 12 月 11 日提交;**最初宣布** 2017 年 12 月。

278. 相机本地化地图的几何感知学习

作 者 :samarth brahmbhatt, jinwei gu, kihwan kim, james hays, jan kautz

摘要: 地图是基于图像的相机定位和可视化 slam 系统中的关键组件: 它们用于在图像之间建立几何约束, 纠正相对姿态估计中的漂移, 并在丢失跟踪后重新定位相机。然而, 地图的确切定义通常是特定于应用程序的, 并且是针对不同场景 (例如 3d 地标、线条、平面、视觉单词袋) 手工制作的。我们建议将地图表示为一种称为 mapnet 的**深度**神经网络, 它可以**学习**数据驱动的地图表示形式。与以前在**学习**地图方面的工作不同, mapnet 除了图像外, 还利用廉价和无处不在的感官输入, 如视觉气味测量和 gps, 并将它们融合在一起, 以便进行相机本地化。这些输入所表示的几何约束传统上用于束调整或后图优化, 在 mapnet 培训中被表述为损失项, 也在推理过程中使用。除了直接提高本地化精度外, 这还允许我们使用来自现场的其他未标记**视频**序列以自我监督的方式更新 mapnet (即地图)。我们还提出了一种新的相机旋转参数化方法, 该参数化更适合**于基于深度学习**的相机姿态回归。室内 7 场景数据集和室外牛津机器人汽车数据集的实验结果显示, 与以往的工作相比, 性能有了显著提高。mapnet 项目网页 https://goo.gl/mRB3Au。少

2018 年 4 月 1 日提交;**v1** 于 2017 年 12 月 9 日提交;**最初宣布** 2017 年 12 月。

279. 电影中学习情感弧线的视听情感分析

作者:朱志强,德布·罗伊

摘要: 故事可以拥有巨大的力量----不仅对娱乐有用，而且可以激活我们的利益，动员我们的行动。一个故事在多大程度上引起了观众的共鸣，这在一定程度上反映在它带到观众的情感旅程上。在本文中，我们使用机器**学习**方法来构造电影中的情感弧线，计算弧线的家族，并证明某些弧线预测观众参与度的能力。该系统适用于好莱坞电影和网页上发现的高质量短裤。我们首先使用**深层**卷积神经网络进行视听情感分析。这些模型在新的和现有的大型数据集上进行了训练，之后它们可用于计算单独的音频和视频情感弧线。然后，我们对从弧线的高点和低点**提取**的 30 秒视频剪辑进行众包注释，以评估系统的微观精度，并根据系统预测之间的极性一致性来测量精度和注释者的评分。这些批注还用于组合音频和视频预测。接下来，我们通过调查情感弧线是否存在 "普遍形状" 来研究电影的宏观特征。特别是，我们开发了一种聚类方法来发现不同类别的情感弧线。最后，我们在一个简短的网络**视频**样本中显示，某些情感

弧线是**视频收到的**评论数量的统计显著预测因子。这些结果表明，我们的方法所**学到**的情感弧线成功地代表了**视频**故事的宏观方面，推动了观众的参与。这样的机器理解可以用来预测观众对**视频**故事的反应，最终提高我们作为说书人相互交流的能力。少

2017 年 12 月 7 日提交;**最初宣布** 2017 年 12 月。

280. 深度缓存：用于移动深视觉的原则缓存

作者:徐梦伟,朱梦泽，刘云欣,李菲克斯·林晓珠,刘玄哲

摘要: 我们提出了 deepcache，这是一种原则缓存设计，用于在连续移动视觉中**进行深度学习**推理。通过利用输入**视频**流中的时间局部性, deepcache 有利于模型执行效率。它解决了移动视觉带来的一个关键挑战:缓存必须在**视频**场景变化下运行，同时在可缓存性、开销和模型准确性损失之间进行权衡。在模型的输入下, deepcache 通过利用**视频的**内部结构来发现视频时间位置，并从视频压缩中借用了**经过验证**的启发式方法;在模型中, deepcache 通过利用模型的内部结构传播可重用结果的区域。值得注意的是, deepcache 避免了将**视频**启发式应用于内部模型，而这些内部不是像素，而是高维、难以解释的数据。我们的 deepcache 实现

了未修改的**深度学习**模型，无需开发人员的手动工作，因此可立即部署到现成的移动设备上。我们的实验表明，deepcache 平均节省了 18% 的推理执行时间，最高可节省 47%。deepcache 平均可将系统能耗降低 20%。少

2018 年 8 月 31 日提交;**v1** 于 2017 年 12 月 1 日提交;**最初宣布** 2017 年 12 月。

281. 面向人工智能的智能城市大规模视频管理：技术、标准和其他

作者:袁灵宇,卢一航,王世奇，高文,永瑞

摘要: **深部在**计算机视觉的一系列任务中，学习取得了巨大的成功。智能**视频**分析可以广泛应用于各种智能城市应用中的**视频**监控，也可以通过这种强大的**深度学习**引擎来驱动。为了在大规模 **视频**分析中实际方便深度神经网络模型，大规模**视频**数据管理仍然面临着前所未有的挑战。**深度**功能编码，而不是**视频**编码，为处理大规模**视频**监控数据提供了一个实用的解决方案。为了在**深度**功能编码的背景下实现互操作性，标准化是当务之急和重要内容。然而，由于**深度学习**算法的爆炸式发展和特征编码的特殊性，在标准化过程中还存在

许多问题。本文提出了面向人工智能的大规模**视频**管理的**未来深层**特征编码标准，并讨论了这些开放问题的现有技术、标准和可能的解决方案。少

2017 年 12 月 4 日提交;**最初宣布** 2017 年 12 月。

282. 长期视觉对象跟踪基准

作者:abhinav moudgil, vineet gandhi

文摘: 本文提出了一种新的长**视频**数据集（称为跟踪长和普罗斯-tlp）和视觉目标跟踪基准。该数据集由来自真实世界场景的 50 个**视频**组成，包括超过 400 分钟的持续时间（676k 帧），使其在每个序列的平均持续时间内超过 20 倍，在覆盖的总持续时间方面超过 8 倍，如与现有的用于可视化跟踪的通用数据集进行比较。拟议的数据集为适当评估长期跟踪性能和培训更好的**深度学习**架构（避免减少增加，这可能无法反映现实的现实世界行为）铺平了道路。我们在 17 个最先进的跟踪器上对数据集进行了基准测试，并根据跟踪精度和运行时速度对其进行了排名。我们进一步提出了深入的定性和定量评价，突出了长期跟踪方面的重要性。我们最有趣的观察是: (a) 现有的短序列基准未能揭示跟踪算法的内在差异，这些差异在跟踪长序列时有所扩大; (b) 大

多数跟踪器的精度突然下降到具有挑战性的长序列，表明在长期跟踪的方向上的研究工作的潜在需要。少

2018 年 3 月 22 日提交;**v1** 于 2017 年 12 月 4 日提交;**最初宣布** 2017 年 12 月。

283. 压缩视频操作识别

作者:吴朝元, manzil zaheer, hahianghu, r. manmatha, 亚历山大 j .smola , philipp krähenbühl

摘要: 事实证明，培训强大的**深层视频**表示比**学习深层**图像表示更具挑战性。这在一定程度上是由于原始**视频**流的巨大大小和高时间冗余;真实而有趣的信号往往被太多不相关的数据淹没。在这种情况下，通过**视频**压缩(使用 h.264、hevc 等)，多余的信息最多可以减少两个数量级，我们建议直接在压缩**视频**上训练一个**深**网络。这种表示具有较高的信息密度，我们发现培训更容易。此外，压缩**视频**中的信号提供免费的 (尽管是嘈杂的)运动信息。我们提出了有效利用它们的新技术。我们的方法比 res3d 快 4.6 倍，比 resnet-152 快 2.7 倍。在操作识别任务上，我们的方法优于 ucf-101、hmdb-51 和 charades 数据集上的所有其他方法。少

2018 年 3 月 29 日提交;**v1** 于 2017 年 12 月 2 日提交;**最初宣布** 2017 年 12 月。

284. 通过深层神经网络预测在线视频的受欢迎程度

作者:岳茂,沈毅,秦刚,蔡龙军

摘要: 预测在线**视频**的受欢迎**程度对于视频**流内容提供商来说很重要。这是一个具有挑战性的问题，原因如下。首先，问题既 "广泛"，又 **"深刻"**。也就是说，它不仅取决于广泛的功能，而且高度非线性和复杂。其次，可能涉及多个竞争对手。本文提出了一种利用多任务**学习** (mtl) 模块和关系网络 (rn) 模块的通用预测模型，在该模型中, mtl 可以减少过度拟合, rn 可以模拟多个竞争对手之间的关系。实验结果表明，该方法显著提高了 rn 和 mtl 模块对电视剧总视图计数的预测精度。少

2017 年 11 月 29 日提交;**v1** 于 2017 年 11 月 29 日提交;**最初宣布** 2017 年 11 月。

285. 自主驾驶中碰撞风险评估的深层预测模型

作 者 :mark strickland, gegereos fainekos, heni ben amor

文摘: 本文研究了一种自主驾驶和辅助驾驶碰撞风险评估的预测方法。**一个深刻**的预测模型被训练来预测来自传统**视频**流的迫在眉睫的事故。特别是，该模型**学习**识别 rgb 图像中的线索，以预测危险的即将出现的情况。与以往的工作不同的是，我们的方法包括: (a) 决策期间的时间信息; (b) 关于环境的多模式信息，以及受控制车辆的本体感知状态和转向行动; (c) 信息关于任务所固有的不确定性。为此，我们讨论了**深度**预测模型，并提出了一个使用贝叶斯卷积 lstm 的实现。在一个简单的仿真环境下进行的实验表明，该方法能够**学会**以合理的精度预测即将发生的事故，特别是在使用多台摄像机作为输入源的情况下。少

2018 年 3 月 29 日提交;**v1** 于 2017 年 11 月 28 日提交; **最初宣布** 2017 年 11 月。

286. 利用伪三维残差网络学习时空表示

作者:邱兆凡,姚婷,陶梅

摘要: 卷积神经网络 (cnn) 已被认为是图像识别问题的一类强大的模型。然而，它不是微不足道的,当使用美国有线电视新闻网学习时空**视频**表示。一些研究表明，执行 3d 卷积是一种很有收获的方法，可以在**视频**中捕

获空间和时间维度。然而，从零开始开发一个非常**深**的 3d cnn 导致昂贵的计算成本和内存需求。一个有效的问题是，为什么不回收现成的 2d 网络为 3d cnn。在本文中，我们设计了一个剩余**学习**框架中的瓶颈构建块的多个变体，方法是通过模拟 3 个 x3 个 x3 个卷积与 1x3 个 x3 个空间域上的卷积滤波器 (相当于 2d cnn) 加上 3 个 x1x1 卷积，以在相邻要素图上及时构造时间连接。此外，我们还提出了一个名为伪三维残差网 (p3d resnet) 的新架构，它利用了块的所有变体，但在不同的 resnet 位置中组合各，遵循的理念是，通过深入增强结构多样性可以提高神经网络的功率。我们的 p3d resnet 在体育-1m 视频分类数据集上实现了明显的改进，而 3d cnn 和基于框架的 2d cnn **分别**提高了 5.3% 和 1.8%。我们进一步研究了我们预先培训的 p3d resnet 在五种不同基准和三种不同任务上制作的**视频**表示的泛化性能，展示了优于几项最先进的基准的性能技术。少

2017 年 11 月 28 日提交;**最初宣布** 2017 年 11 月。

287. 分分补强化学习

作 者 :dibya ghosh, avi singh, aravind rajeswaran, vikash kumar, sergey levine

文摘: 标准无模型深度强化**学习**(rl) 算法为每个试验采样一个新的初始状态，使它们能够优化即使在高度随机环境下也能很好地执行的策略。然而，表现出相当大的初始状态变化的问题通常会产生无模型 rl 的高方差梯度估计，这使得直接策略或值函数优化具有挑战性。在本文中，我们开发了一种新的算法，该算法将初始状态空间划分为 "切片"，并优化了策略集合，每个策略集合位于不同的切片上。合奏逐渐统一为一个单一的政策，可以在整个国家空间取得成功。这种方法，我们称之为分而治之的 rl，能够解决传统的**深度** rl 方法无效的复杂任务。结果表明，在具有挑战性的抓取、操纵和运动任务上，分而治地的 rl 大大优于传统的政策梯度方法，并且超过了以往各种方法的性能。我们的算法**所学习**的**政策视频**可以在 http://bit.ly/dnc-rl 可以查看少

2018 年 4 月 27 日提交;**v1** 于 2017 年 11 月 27 日提交;**最初宣布** 2017 年 11 月。

288. 利用自我监督学习开发无标记内窥镜视频数据的潜力

作 者 :tobias ross, david zimmerer, anant vemuri, fabian isensee, manuel wiesenfarth,sebastian bodenstedt, fabian 两者,菲利

普 · 凯 斯 勒 ， martin wagner ， 击 败 müller, hannes kenngot, stefanie speidel ， annette kop-schneider, klaus Maier-Hein, lena Maier-Hein

摘要: 外科数据科学是一个新的研究领域，旨在观察患者治疗过程的各个方面，以便在正确的时间提供正确的帮助。由于基于**深度学习**的图像自动注释解决方案取得了突破性的成功，算法训练参考注释的可用性正成为该领域的主要瓶颈。本文旨在探讨自我监督**学习**的概念，以解决这一问题。我们的方法是基于这样的假设，即未标记**的视频**数据可用于**学习**目标域的表示形式，从而提高最先进的机器**学习**算法的性能。训练前。该方法的核心是基于目标域原始**内**镜视频数据的辅助任务，用于初始化目标任务的卷积神经网络 (cnn)。本文提出了以基于生成对抗网络 (gan) 的体系结构作为辅助任务对医学图像进行重新着色的建议。该方法的一个变体涉及基于相关域中目标任务的标记数据的第二个预训练步骤。我们使用医疗仪器分割作为目标任务来验证这两个变种。该方法可用于从根本上减少培训 cnn 所涉及的手动注释工作，与从头开始生成注释数据的基线方法相比，我们的方法可探索地将标记图像的数量减少 75%在不牺牲性能的情况下。我们的方法也优于美国有线电视新闻网预培训的替代方法，例如使用目标任

务（在这种情况下：分段）对公开提供的非医疗或医疗数据进行预培训。由于该方法有效地利用了现有的（非）公共和（非）标记数据，因此有可能成为美国有线电视新闻网（前）培训的宝贵工具。少

2018 年 1 月 31 日提交;**v1** 于 2017 年 11 月 27 日提交;**最初宣布** 2017 年 11 月。

289. 深层视频生成、人体动作序列的预测和完成

作者:蔡浩业,白春燕,太原，崔强堂

摘要: 目前**视频**生成的**深度学习**结果有限，而视频预测仅有少量初步结果，**视频**完成方面没有相关的显著结果。这是由于这三个问题所固有的严重的不恰当之处。在本文中，我们重点介绍了人类行动**视频**，并提出了一个通用的两阶段**深度**框架，以生成没有约束或任意数量约束的人体动作**视频**，从而统一解决这三个问题:**视频**生成给定没有输入帧，视频预测给定前几个帧，视频完成给定的第一个和最后一个帧。为了使问题易于处理，在第一阶段，我们训练一个**深层**生成模型，从随机噪声生成一个人的姿势序列。在第二阶段，训练一个骨架到图像的网络，考虑到第一阶段生成的完整的人体姿势序列，该网络用于生成人类动作**视频**。通过引入两阶段策

略，我们避开了原来的问题，同时首次产生了高质量的**视频**生成/预测/持续时间更长的完成结果。我们提出的定量和定性评价，以表明我们的两阶段方法优于最先进的方法，在**视频**生成，预测和**视频**完成。我们的**视频**演示可以在 https://iamacewhite.github.io/supp/index.html 查看少

2017 年 12 月 8 日提交;**v1** 于 2017 年 11 月 23 日提交;**最初宣布** 2017 年 11 月。

290. 基于 rgb-d 的深层学习中的人类运动识别研究综述

作者:pihaowang, wan 青丽, phillip Ogunbona, jun wan, sergio ecalera

摘要: 人体运动识别是以人为中心的研究活动的重要分支之一。近年来，基于 rgb-d 数据的运动识别备受关注。随着人工智能的发展,**深度学习**技术在计算机视觉方面取得了显著的成功。特别是卷积神经网络 (cnn) 在基于图像的任务中取得了巨大的成功，而递归神经网络 (rnn) 则以基于序列的问题而闻名。具体而言，采用基于 cnn 和 rnn 架构的**深度学习**方法对基于 rgb-d 数据的运动识别。本文对基于 rgb-d 的运动识别的最新

进展进行了详细的综述。根据识别采用的方式，审查的方法大致分为四组：基于 rgb、基于深度、基于骨架和基于 rgb + d。作为一项以**深度学习**应用于基于 rgb-d 的运动识别为重点的调查，我们明确讨论了现有技术的优点和局限性。特别是，我们强调了**视频**序列中固有的时空结构信息的编码方法，并讨论了未来研究的潜在方向。少

2018 年 4 月 24 日提交;**v1** 于 2017 年 10 月 31 日提交;**最初宣布** 2017 年 11 月。

291. 由生成对抗性和排序网络对对象发现的监管薄弱

作者:ali diba, vivek sharma, rainer stiefelhagen, luc van gool

文摘: **深度**生成对抗网络 (gan) 最近已被证明是有希望的不同的计算机视觉应用，如图像编辑，合成高分辨率图像,**生成视频**等。这些网络和相应**的学习**方案可以处理各种视觉空间映射。我们用一种新的训练方法和**学习**目标来接近 gans，发现三个案例的多个对象实例：1) 综合一个杂乱场景中特定对象的图片;2) 本地化图像中的不同类别，以进行弱监督对象检测;3) 改进检测管道中的对象丢弃。我们方法的一个关键优势是它**学习**了一

个新的**深度**相似度度量，以区分一个时代的多个对象。我们证明，网络可以作为编码器解码器生成包含物体的图像部分，或作为一个修改后的深 cnn 重新搜索图像的**对象**检测在监督和弱监督方案。我们排名的 gan 提供了一种新的方法来搜索图像中的对象特定的模式。我们对不同的场景进行了实验，并演示了使用 ms-coco 和 pascal voc 数据集进行目标合成和弱监督对象检测和分类的方法性能。少

2018 年 4 月 17 日提交;**v1** 于 2017 年 11 月 22 日提交;**最初宣布** 2017 年 11 月。

292. 科学驱动的创新为移动产品提供动力：云 ai 与智能设备上的设备 ai 解决方案

作者:德广港

摘要: 近年来，移动设备 (如 iphone) 因其给人类生活带来的便利而日益普及。一方面，来自异构信息源的丰富的用户分析和行为数据 (包括每个应用级别、应用交互级别和系统交互级别) 使提供更好的服务 (如推荐、广告) 成为可能) 面向客户，这进一步推动了了解用户行为和提高用户参与度的收入。为了取悦客户，智能个人助理 (如亚马逊亚历克莎，谷歌家居和谷歌 now)

是非常可取的，以提供实时音频，视频和图像识别，自然语言理解，舒适的用户互动接口，满意的推荐和有效的广告定位。本文介绍了我们在移动设备上开展的研究工作，这些研究旨在通过利用统计数据和大数据科学、机器**学习**和**深度学习**，提供更智能、更方便的服务,用户建模和营销技术，在移动设备上带来显著的用户增长和用户参与度和满意度 (和满意度)。开发的新功能建立在云端或设备侧，和谐地协同工作，以提高当前的服务，目的是提高用户的满意度。我们用不同的案例研究从系统和算法的角度来设计这些新功能，通过这些案例研究，人们可以很容易地了解科学驱动的创新如何有助于提供更好的技术服务，带来更多的收入在业务中的发展。同时，这些研究工作也有明显的科学贡献，并在顶级场馆发表，这些场馆对移动 ai 产品发挥着越来越重要的作用。少

2017 年 11 月 20 日提交;**最初宣布** 2017 年 11 月。

293. 基于记忆的视频流深层表示的在线学习

作 者 :fedico pernici, fedico bartoli, matteo bruni, alberto del bimbo

文摘: 我们提出了一种新的在线无监督方法，从**视频**流中进行人脸身份**学习**。该方法利用**深层**的人脸描述符和基于记忆的学习机制，利用视觉数据的**时间**一致性。具体而言，我们引入了一种基于反向近邻的判别特征匹配解决方案和一种特征遗忘策略，该策略检测冗余特征，并在时间进展时适当地丢弃这些特征。结果表明，所提出的**学习**过程是渐近稳定的，可以有效地应用于无约束**视频**流的多面识别和跟踪等相关应用。实验结果表明，该方法利用未来信息的离线方法，在多人脸跟踪任务中取得了可比的效果，在人脸识别方面取得了较好的性能。代码将公开提供。少

2017 年 11 月 17 日提交;**最初宣布** 2017 年 11 月。

294. 麻省理工学院自主车辆技术研究：基于大规模深度学习的驾驶员行为分析与自动化交互

作 者 :lex fridman, daniel e. brown , michael glazer, william angell, spencer Fridman, Benediktjenik, jack terwilliger, julia kindelsberger, li ding, 肖恩·希曼、希拉里·亚伯拉罕、阿莱亚·梅勒、安德鲁·西普利、安东尼·佩蒂纳托、博比·塞普佩特、琳达·安格尔、布鲁斯·梅勒、布莱恩雷默尔

摘要: 对于可预见的未来，人类很可能仍然是驾驶任务的一个组成部分，监控 ai 系统，因为它的性能从刚刚超过 0% 到接近 100% 的驾驶。麻省理工学院自主车辆技术 (mit-avt) 研究的指导目标是: (1) 进行大规模的现实驾驶数据收集，其中包括高清**视频**，以推动**深度学习的发展**基于内部和外部感知系统, (2) 通过将**视频**数据与车辆状态数据、驾驶员特征、心理模型和自我报告的技术经验, (3) 确定如何以拯救生命的方式改进与自动化采用和使用有关的技术和其他因素。为了实现这些目标，我们为长期 (每名司机一年以上) 和中期 (每个司机一个月) 自然驾驶数据，检测了 21 辆特斯拉 s 型和 x 型车、2 辆沃尔沃 s90 车、2 辆越野车 evoque ect6 车和 2 辆凯迪拉克 ct6 车。收集。此外，我们还在不断开发新的方法来分析从检测车队收集到的大规模数据集。记录的数据流包括 imu、gps、can 消息以及驾驶员脸、驾驶室、前进道路和仪表群 (在特定车辆上) 的高清**视频**流。这项研究正在进行中，而且还在不断发展。到目前为止，我们有 99 名参与者, 110, 846 天的参与, 405,807 英里,**和** 550 亿视频帧。本文介绍了研究的设计、数据采集硬件、数据处理以及目前用于从数据中提取可操作知识的计算机视觉算法。少

2018 年 9 月 30 日提交;**v1** 于 2017 年 11 月 19 日提交;**最初宣布** 2017 年 11 月。

295. 基于视觉的铁路轨道深度学习监测

作者:shruti mittal, dattaraj rao

摘要: 计算机视觉的铁路轨道缺陷检测方法在过去已经得到了探索，但由于传统的图像处理方法和**深度学习**分类器都受过训练，完全自动化一直是一个挑战考虑到标记的数据数量有限，从零开始，无法将这种情况很好地概括为现实世界中看到的无限的新场景。最近，机器**学习**模型利用了来自不同但相关领域的知识，取得了越来越大的进展。本文表明，尽管没有类似的域数据，但**转移学习提供**了对其他真实世界对象的模型理解，并使培训生产规模**深度学习**分类器成为可能。不受控制的真实世界数据。我们的模型可有效地检测诸如阳光、松散镇流器和铁路资产 (如开关和信号) 的轨道缺陷。模型通过在不同大陆录制的小时**跟踪视频**进行验证，从而产生不同的天气条件、不同的氛围和环境。还提出了轨道健康指数概念，以监测完整的铁路网。少

2017 年 11 月 17 日提交;**最初宣布** 2017 年 11 月。

296. 利用语义引导生成对抗性网络提高序列画的一致性和正确性

作 者 :avisek lahiri, arnav jain , prabir kumar bis 邮件, pabitra mitra

摘要: 当代图像画的基准方法是基于**深层**生成模型，并专门利用对抗性损失产生现实的重建。但是，由于内部的缺点，这些模型不能直接应用于图像/**视频**序列--重建可能是独立现实的，但是，当可视化为序列时，通常缺乏对原始未损坏的图像/视频序列的保真度序列。其根本原因是，这些方法试图在没有任何基于距离的明显损失的情况下，在自然图像流形附近找到最佳的匹配潜在空间表示。本文提出了一种用于序列画的语义条件生成对抗性松网络 (gan)。条件信息约束 gan 将潜在表示映射到图像流形中的一个点，该点尊重场景的基本姿态和语义。据我们所知，这是第一部同时解决基于绘画的生成模型的一致性和正确性的作品。我们表明，我们的生成模型**学习**分离的姿势和外观信息;这种独立性被我们的模型所利用，以产生高度一致的重建。条件信息还有助于 gan 中的发电机网络产生比最初的 gan 公式更清晰的图像。这有助于实现更有吸引力的绘画表演。虽然是通用的，但我们的算法是针对人脸上的

画画的。当应用于 celeba 和 youtube 面数据集时，该方法比目前的基准有了显著的改进，无论是在定量评价 (峰值信号与噪声比率) 还是在人类视觉评分方面，都是在多样化组合方面。决议和变形。少

2017 年 11 月 17 日提交;**v1** 于 2017 年 11 月 16 日提交;**最初宣布** 2017 年 11 月。

297. 多尺度深部损失函数和生成对抗性比域的帧插值

作 者 :joost van amersfoort, wenzheshi, alejandro acosta, francisco m 萨, jones totz, zhan wang, jose caballero

摘要: 帧插值尝试在给定一个或多个连续**视频**帧的情况下合成中间帧。近年来,**深度学习**方法，特别是卷积神经网络，成功地解决了包括帧插值在内的低、高层计算机视觉问题。在这一领域的研究中，主要有两个方面，即算法效率和重建质量。本文提出了一种用于帧插值 (igan) 的多尺度生成对抗网络。为了最大限度地提高网络的效率，我们提出了一个新的多尺度残差估计模块，其中预测的流量和合成帧是以一种粗糙到精细的方式构建的。为了提高综合**中间视频**帧的质量，我们的网络在不同级别共同监督，具有感知损失功能，包括对抗性

损失和两个内容损失。我们使用 youtube-8m 中的 60fps **视频**集合来评估建议的方法。我们的结果提高了最先进的精度和效率，以及与性能最佳的插值方法相媲美的主观视觉质量。少

2017 年 11 月 16 日提交;**最初宣布** 2017 年 11 月。

298. 端到端视频级表示学习促进行动识别

作者:朱家刚,邹伟,朱正

文摘: 从帧-剪辑级特征**学习**到**视频**级表示构建，近年来动作识别中的**深度学习**方法发展迅速。然而，目前的方法却受到部分观察训练或没有端**到端学习**，或仅限于单时间尺度建模等因素造成的混乱。本文以双流凸网为基础，提出了一种端到端视频级表示学习方法--时间金字塔池 (dtpp) 的**深**网，以解决这些问题。具体来说，首先, rgb 图像和光流堆栈在整个**视频**中的采样量很少。然后利用时间金字塔池层对由空间和时间线索组成的框架级特征进行聚合。最后，该训练模型具有具有多个时间尺度的**紧凑的视频**级表示形式，既具有全局性，又具有序列性。实验结果表明, dtpp 通过 imagenet 预训练或动力学预训练，在 ucf101 和 hmdb51 两个具有挑战性的**视频**操作数据集上实现了最先进的性能。少

2018 年 4 月 21 日提交;**v1** 于 2017 年 11 月 11 日提交;**最初宣布** 2017 年 11 月。

299. 具有边缘感知深度-正常一致性的无监督几何学习

作者:杨振恒,王鹏,徐伟,赵亮，拉玛康特内瓦蒂亚

文摘: 近年来，通过**深**卷积网络（dcn）观看未标记**的视频**来学习重建单个图像中的深度正受到人们的广泛关注。本文介绍了无监督深度估计框架的表面正态表示。我们的估计深度必须与预测的法线兼容，从而产生更可靠的几何结果。具体而言，我们制定了一个边缘感知深度-法线一致性项，并通过在 dcn 内构造一个深度到法线层和一个法线层来解决它。深度到法线层以估计深度为输入，并使用基于相邻像素的交叉生成计算正常方向。然后给出估计的法线，法线到深度层通过局部平面平滑输出正则深度图。这两个图层都是通过感知图像内部的边缘来计算的，以帮助解决深度正常不连续性的问题并保留锐利的边缘。最后，为了训练网络，我们将光度误差和梯度平滑度应用于深度和正常预测。我们在室外（kititi）和室内（nyuv2）数据集上进行了实验，并表明我们的算法大大优于最先进的算法，这证明了我们的方法所带来的好处。少

2017 年 11 月 9 日提交;最初宣布 2017 年 11 月。

300. 视频分类中具有时空关注的双流协作学习

作者:彭玉欣,赵云珍,张俊超

摘要: **视频**分类在**视频**搜索和智能监控等广泛应用中具有十分重要的意义。**视频**自然由静态和运动信息组成,这些信息可以用帧和光流来表示。近年来,研究人员普遍采用**深**网来捕捉静态和运动信息\textbf{\emph{separately}},主要有两个局限性: (1) 忽视时空关注的共存关系,在将其组合建模为**视频**的时空演化的同时,可以提取出判别**视频**特征。(2) 忽略静态信息和运动信息在**视频**中共存的强大互补性,同时应协作**学习**相互提升。针对上述两个局限性,本文提出了具有时空关注的双流协作**学习**方法,该方法由两个模型组成: (1) 时空关注模型: 时空关注模型: 时空级注意强调画面中的突出区域,在视频中注意利用**判别**帧。他们共同**学习**和相互提升,以**学习**判别静态和运动特征,更好的分类性能。(2) 静态运动协作模型: 它不仅实现了静态和运动信息的相互引导,促进了特征**学习**,而且自适应地**学习**了静态和运动流的融合权重,从而加以利用静态信息和运动信息之间的强互补性,促进**视频**分类。在 4 个广泛使用的数据集上进行的实验表明,

与 10 多多种最先进的方法相比，我们的 tclst 方法实现了最佳性能。少

2017 年 11 月 9 日提交;**最初宣布** 2017 年 11 月。

301. 具有知识图的端到端视频分类

作者:方元,王哲，林杰, 路易斯·费尔南多·达罗，金正杰, 曾曾,维杰伊·钱德拉塞卡

文摘: **视频**理解引起了广泛的研究，尤其是最近大规模**视频**基准的提供。本文讨论了多标签**视频**分类问题。我们首先观察到机器和人类**之间的学习**方式存在着巨大的知识差距。也就是说，虽然目前的机器**学习**方法包括**深度**神经网络，主要集中在给定数据的表示上，但人类往往超越手头的数据，利用外部知识做出更好的决定。为了缩小差距，我们建议将外部知识图纳入**视频**分类。特别是，我们将传统的 "无知识" 机器**学习**模型和知识图统一到一个新的端到端框架中。该框架灵活地使用了大多数现有**的视频**分类算法，包括最先进的**深层**模型。最后，我们在最大的公共**视频**数据集 youtube-8m 上进行了广泛的实验。结果在各方面都很有希望，平均精度提高了 2. 9%。少

2017 年 11 月 5 日提交;**最初宣布** 2017 年 11 月。

302. 用于视频分类的卷积漂移网络

作 者 :dillon graham, seyed hamed fatemi langroudi, christopher kanan , dhireesha kudithipudi

摘要: 分析时空数据 (如**视频)**是一项具有挑战性的任务，需要有效地处理视觉和时间信息。卷积神经网络通过转移**学习显示**了作为基线固定特征提取器的前景，这一技术有助于最大限度地降低视觉信息的培训成本。时间信息通常使用手工制作的特征或重复神经网络来处理，但这可能过于具体或过于复杂。构建一个完全可培训的系统，能够在没有手工制作的特征或复杂培训的情况下高效地分析时空数据，这是一个开放的挑战。我们提出了一个新的神经网络架构来应对这一挑战，卷积漂移网络 (cdn)。我们的 cdn 架构将**深**卷积神经网络的视觉特征提取能力与储层计算提供的本质上高效的时间处理相结合。在这篇关于 cdn 的介绍性论文中，我们提供了一个在两个自我中心 (第一人称)**视频**活动数据集上测试的非常简单的基线实现。我们实现**了视频**级活动分类结果与最先进的方法相当。值得注意的是，在这一复杂的时空任务上的性能是通过只在 cdn 中训练一个前馈层而产生的。少

2017 年 11 月 3 日提交;**最初宣布** 2017 年 11 月。

303. 用于跟踪和预测的深部和反向感知模型

作者:亚历山大·兰伯特, amureza shaban, amit raj,镇刘，拜伦靴子

摘要: 我们考虑了**学习**向前模型的问题，这些模型将状态映射到高维图像，并将高维图像映射到机器人中的状态的逆模型。具体来说，我们提出了一个感知模型，用于从**具有深层**网络的状态生成**视频**帧，并为其在跟踪和预测任务中的使用提供了一个框架。结果表明，我们提出的模型大大优于标准的脱色方法和有机组织的图像生成，产生清晰、逼真的图像。我们还开发了一个卷积神经网络模型进行状态估计，并将结果与扩展卡尔曼滤波器进行比较，以估计机器人轨迹。我们在一个真正的机器人系统上验证所有模型。少

2018 年 5 月 19 日提交;**v1** 于 2017 年 10 月 30 日提交;**最初宣布** 2017 年 10 月。

304. 预测全景视频中的头部运动：一种深层强化学习方法

作者:宋玉航,徐麦明,乔明朗,王建义，霍良宇,王祖林

摘要: 全景**视频**通过使人类能够通过头部运动 (hm) 控制视野 (fov), 提供沉浸式和互动体验。因此, hm **在全景视频**上模拟人类注意力方面发挥着关键作用。本文建立了一个收集图像序列主体 hm 的数据库. 从这个数据库中, 我们发现各异的 hm 数据是高度一致的。此外, 我们还发现, **深度**强化**学习**(drl) 可以应用于预测 hm 位置, 通过最大限度地通过代理的作用模仿人的 hm 扫描的回报。根据我们的研究结果, 我们提出了一种基于 drl 的 hm 预测 (dhp) 方法, 该方法包含离线和在线版本, 称为离线 dhp 和在线 dhp。在离线 dhp 中, 运行多个 drl 工作流, 以确定每个全景帧上潜在的 hm 位置。然后, 生成一个潜在的 hm 位置的热图, 称为 hm 映射, 作为离线 dhp 的输出。在在线 dhp 中, 根据目前观察到的 hm 位置, 估计了一个主体的下一个 hm 位置, 该位置是通过**在学习到**的离线 dhp 模型上开发 drl 算法来实现的。最后, 实验验证了该方法在全景**视频** hm 位置的离线和在线预测中的有效性, 以及**所学的**离线 dhp 模型可以提高在线 dhp 的性能。少

2018 年 9 月 20 日提交;**v1** 于 2017 年 10 月 29 日提交; **最初宣布** 2017 年 10 月。

305. 上下文 vp：完全上下文感知视频预测

作 者 :wonmin byeon, qinwang, rupesh kumar srivastava, petros Koumoutsakos

摘要: 基于卷积网络、递归网络及其组合的**视频**预测模型通常会导致模糊的预测。我们确定了文献中没有充分研究过的预测不准确的一个重要因素：盲点，即无法获得所有相关的过去信息，以便准确预测未来。为了解决这个问题，我们引入了一个完全上下文感知的体系结构，该体系结构使用并行多维 lstm 单元捕获每个像素的整个可用过去上下文，并使用混合单元对其进行聚合。我们的模型优于由 20 个反复卷积层组成的强大基线网络，并可在三个具有挑战性的真实**视频**数据集（人 3.6 m、caltech 行人和 ucf-101）上获得最先进的性能，以便下一步进行预测。此外，它这样做的参数比最近提出的几个模型少，并且不依赖**于深层**卷积网络、多尺度体系结构、背景和前景建模的分离、运动**流学习**或对抗性训练。这些结果突出表明，充分了解过去的背景对于**视频**预测至关重要。少

2018 年 9 月 9 日提交;**v1** 于 2017 年 10 月 23 日提交;**最初宣布** 2017 年 10 月。

306. UG^2：评估图像恢复和增强对自动视觉识别的影响的视频基准

作 者 :rosaura g. vidal, sreya banerjee, klemen grm, vitomir 色, walter j. cheeirer

摘要: 图像恢复和增强技术的进步导致了关于如何将这种算法扫描作为预处理步骤来提高自动视觉识别的讨论。原则上，去模糊和超分辨率等技术应该通过在输入图像中不再强调噪声和增加信号来改进。但计算摄影和视觉识别界历史上不同的目标，已经产生了在这方面开展更多工作的巨大需求。为了促进新的研究，我们引入了一个新的基准数据集，称为 UG^2，其中包含三个困难的现实场景：无人机和载人滑翔机拍摄的不受控制的**视频**，以及在地面上拍摄的受控**视频**。超过 160,000 个带注释的帧可用于数百个 imagenet 类，这些帧用于基线实验，用于评估已知和未知图像伪影及其他条件对常见的**深层学习**对象的影响分类方法。此外，通过确定当前的图像恢复和增强技术是否能提高基线分类性能来进行评估。结果显示，算法创新有很大的空间，使此数据集成为前进的有用工具。少

2018 年 2 月 6 日提交;**v1** 于 2017 年 10 月 8 日提交;**最初宣布** 2017 年 10 月。

307. 通过有条件模拟学习进行端到端驾驶

作者:felipe codevilla, matthias müller, antonio lópez, vladlen kortun, 阿列克谢·多索维茨基

摘要: 接受过人类驾驶示范训练的**深度**网络已经**学会了**走道路和避免障碍。然而，通过模仿**学习**训练的驾驶政策在考试时无法控制。在即将到来的十字路口，不能引导接受过端到端仿真专家的车辆进行特定转弯。这限制了此类系统的效用。提出了在高级命令输入的条件下进行模拟**学习**。在考试时,**学习过的**驾驶政策起到司机的作用，处理传感器运动的协调，但继续响应导航命令。我们评估不同的体系结构的条件模仿学习在基于视觉的驱动。我们在城市驾驶的逼真的三维模拟和训练在居民区行驶的半规模机器人卡车上进行实验。这两个系统都基于可视输入驱动，但仍对高级导航命令保持响应。补充**视频**可在 https://youtu.be/cFtnflNe5fM 内观看少

2018 年 3 月 2 日提交;**v1** 于 2017 年 10 月 6 日提交;**最初宣布** 2017 年 10 月。

308. 通过观看 youtube 学习对人的细分

作者:梁晓丹,魏云超,梁林,陈云鹏，沈晓辉,杨建超，严水城

摘要: 人类分割的直觉是，当一个人在**视频**中移动时，视频-上下文 (例如,**外观**和运动线索) 可能会推断整个人体合理的掩码信息。在这一鼓舞下，基于流行的**深层**卷积神经网络 (cnn)，我们探索了一个非常弱的人类分割任务的监督**学习**框架，在这个框架中，只有一个不完美的人类探测器和大量的弱标签的 youtube **视频**。在我们的解决方案中,**视频**上下文引导人掩码推理和基于 cnn 的分段网络**学习**迭代，相互增强，直到没有进一步的改进成果。在第一步中，通过无监督**视频**分割将每个**视频**分解为超体素。然后，利用不完善的人类检测结果和在上一次迭代中训练的美国有线电视新闻网的预测置信度图，利用一元能量通过一元能量对超体素中的超级像素进行了分类，将超体素分为人或非人。在第二步中,**视频**上下文衍生的人类面具被用作训练美国有线电视新闻网的直接标签。在具有挑战性的 pascal voc 2012 语义分割基准上进行的大量实验表明，与以前所有具有对象类或边界框注释的弱监督方法相比，该框架已经取得了卓越的效果。此外，通过使用 pascal voc 2012 中的注释掩码进行扩充，我们的方法在人工分割任务上达到了新的最先进的性能。少

2018 年 2 月 27 日提交;**v1** 于 2017 年 10 月 4 日提交;**最初宣布** 2017 年 10 月。

309. 降低 hevc 的复杂性: 一种深度学习方法

作者:麦旭,李天一,王祖林, 新登,任阳,关振宇

摘要: 高效**视频**编码 (hevc) 比正在进行的 h.264 标准显著降低了比特率, 但代价是极高的编码复杂性。在 hevc 中, 由于对速率失真优化 (rdo) 的暴力搜索, 编码单元 (cu) 的四叉树划分消耗了很大一部分 hevc 编码复杂性。因此, 本文提出了一种**基于**卷积神经网络 (cnn) 和长期和短期记忆 (lstm) 的深度**学习**方法来预测 cu 分区, 以降低内模和模间的 hevc 复杂度。) 网络。首先, 我们建立了一个大型数据库, 包括 hevc 内部和模式间的大量 cu 分区数据。这可以在 cu 分区上**进行深度学习**。其次, 我们以分层 cu 分区映射 (hcpm) 的形式表示整个编码树单元 (ctu) 的 cu 分区。然后, 我们提出了一个早期终止的等级 cnn (et-nnn), 用于**学习**预测 hcpm。因此, 通过将暴力搜索替换为 et-nnn 来决定 cu 分区, 可以大大降低模式内 hevc 的编码复杂度。第三, 提出了一种早期终止层次 lstm (et-lstm) 来**学习** cu 分区的时间相关性。然后, 我们结合 et-lstm 和 et-nnn 预测 cu 分区, 以降低交互模式的 hevc 复杂度。最后, 实验结果表明, 我们的方法在降低 hevc 复杂度方面优于其他最先进的方法。少

2018 年 3 月 22 日提交;**v1** 于 2017 年 9 月 18 日提交;**最初宣布** 2017 年 10 月。

310. 最小迭代动态博弈：在非线性机器人控制任务中的应用

作 者 :olalekan Ogunmolu, nicolas gans, tyler summers

摘要: 多级决策策略在高维状态空间中提供了有用的控制策略，尤其是在复杂的控制任务中。但是，在存在扰动、模型不匹配或模型不确定性的情况下，它们表现出微弱的性能保证。这种脆性限制了它们在高风险情况下的使用。我们介绍了如何量化这些政策的敏感性，以便了解其稳健能力。我们还提出了一个最小迭代动态博弈框架，用于在存在扰动/不确定性的情况下设计可靠的策略。我们在精心设计的**深层**神经网络策略上检验量化假设;然后，我们提出一个最小迭代动态博弈（isg）框架，以提高策略的鲁棒性，在存在对抗干扰的情况下。我们在一个 mecant 轮椅机器人上评估我们的 isg 框架，其目标是找到一个具有空间强大的最优多级策略，以实现给定的目标覆盖任务。该算法简单且适应性强，可用于设计对扰动、模型不匹配或模型不确定性具有鲁棒性的元学习/**深度**策略，直至扰动约束。结果**的视频**出

现在作者的网站上，http://ecs.utdallas.edu/~opo140030/iros18/iros2018.html，而复制我们实验的代码则在 github 上，https://github.com/lakehanne/youbot/tree/rilqg。一个自成一体的环境，复制我们的结果是在码头上，https://hub.docker.com/r/lakehanne/youbotbuntu14/ 少

2018 年 8 月 5 日提交;**v1** 于 2017 年 10 月 2 日提交;**最初宣布** 2017 年 10 月。

311. 双域中的快速卷积稀疏编码

作者:lama affara, bernard ghanem, peter wonka

摘要: 卷积稀疏编码 (csc) 是从图像和**视频**压缩到**深度学习**等许多计算机视觉应用的重要组成部分。我们为 csc 的最新发展提供了两项贡献。首先，我们通过提出一个新的优化框架来解决双域中的问题，从而显著加快计算速度。其次，我们将原始配方扩展到更高的尺寸，以便处理更广泛的输入，如 rgb 图像和**视频**。与目前最先进的 csc 求解器相比，我们的结果速度提高了 20 倍。 少

2018 年 4 月 8 日提交;**v1** 于 2017 年 9 月 27 日提交;**最初宣布** 2017 年 9 月。

312. 基于级联的基于区域的密集连接网络，用于事件检测：一种地震应用

作者:岳武、林友佐 、郑州、 大卫·奇斯·博尔顿、刘基 、保罗·约翰逊

摘要: 时间序列信号的事件自动检测具有广泛的应用，如**视频**监控中的异常事件检测和地球物理数据中的事件检测。传统的检测方法主要通过在数据中使用相似性和相关性来检测事件。这些方法可能效率低下，而且精度较低。近年来，由于计算能力的显著提高，机器**学习**技术使许多科学和工程领域发生了革命性的变化。在本研究中，我们应用了一种**基于深度学习**的方法来检测时间序列地震信号中的事件。然而，从二维物体检测到我们的问题，对类似的想法的直接适应面临着两个挑战。第一个挑战是地震事件的持续时间差别很大;另一个是所产生的建议在时间上是相互关联的。为了应对这些挑战，我们提出了一种新的基于级联区域的卷积神经网络来捕获不同大小的地震事件，同时结合上下文信息来丰富每个建议的特征。为了实现更好的泛化性能，我们使用密集连接的块作为网络的主干。由于一些正事

件没有正确注释，我们进一步将检测问题表述为噪声学习问题. 为了验证我们的检测方法的性能，我们将我们的方法应用于岩石力学实验室的双轴 "地震机器" 生成的地震数据，并在专家的帮助下获取标签。通过数值试验表明，我们的新检测技术具有较高的精度。因此，我们新的**基于深度学习**的检测方法有可能成为在各种应用中从时间序列数据中定位事件的强大工具。少

2017 年 11 月 28 日提交;**v1** 于 2017 年 9 月 12 日提交;**最初宣布** 2017 年 9 月。

313. 稀疏到密集：稀疏深度样本和单个图像的深度预测

作者:马方昌,塞尔塔奇·卡拉曼

摘要: 我们从一组稀疏的深度测量和一个 rgb 图像中考虑了密集深度预测的问题。由于仅从单目图像中进行的深度估计本质上是模糊和不可靠的，为了获得更高的鲁棒性和准确性，我们引入了额外的稀疏深度样本，这些样本要么是用低分辨率的深度传感器获得的，要么是计算的通过可视同步本地化和映射 (slam) 算法。我们建议使用单一**的深回归**网络直接从 rgb-d 原始数据中**学习**，并探讨深度样本数量对预测精度的影响。我们的实验表明，与仅使用 rgb 图像相比，添加 100 个

空间随机深度样本可将 nyu-debth-v2 室内数据集上的预测根均方误差减少 50%。它还将 kititi 数据集上可靠预测的百分比从 59% 提高到 92%。我们展示了该算法的两个应用: 一个是 slam 中的插件模块, 用于将稀疏映射转换为密集映射, 以及用于 lidar 的超分辨率。软件和**视频**演示是公开提供的。少

2018 年 2 月 25 日提交;**v1** 于 2017 年 9 月 21 日提交;**最初宣布** 2017 年 9 月。

314. 采用两相校准程序, 采用电缆驱动机器人进行快速、可靠的自主手术清配

作者:daniel seita , sanjay krishnan, roy fox, stephen mckinley, john canny , ken goldberg

摘要: 使用机器人手术助理 (rsa) (如达·芬奇研究套件 (dvrk)) 进行清创 (去除死亡或患病的组织碎片) 等精确子任务具有挑战性, 因为电缆驱动系统存在固有的非线性。提出并评价了一种新的两相粗化标定方法。在第一阶段 (粗), 我们在末端效应器上放置一个红色校准标记, 并让它随机移动到一组开环轨迹中, 以获得大量相机像素和内部机器人末端执行器配置的样本集。然后, 这些粗糙的数据被用来训练**深**神经网络 (dnn)**来学**

习粗变换偏差。在第二阶段（罚款）中，第一阶段的偏差用于将末端效应器移动到打印工作表上的一小部分特定目标点。对于每个目标，人工操作员通过直接接触(而不是通过远程操作）手动调整末端执行器位置，并记录剩余补偿偏差。然后，这些精细数据被用来训练随机森林（rf），**以学习**精细变换偏差。随后的实验表明，如果不进行校准，位置误差平均为 4.55 mm。第一阶段可以将平均误差减少到 2.14 mm，第一阶段和第二阶段的组合可以将平均误差降低到 1.08 mm。我们将这些结果应用于葡萄干和南瓜籽作为碎片幻影的清创。利用具有标准边缘检测的内窥镜立体摄像机，120 项试验的实验实现了平均成功率 94.5，超过了以前的结果，碎片大得多（8. 9.4%），加速了 2.1 x，减少了每个片段的时间从 15。8 秒到 7.3 秒。源代码、数据和**视频**可在 https://sites.google.com/view/calib-icra/。少

2018 年 2 月 24 日提交;**v1** 于 2017 年 9 月 19 日提交;**最初宣布** 2017 年 9 月。

315. ajile 运动预测：自然人类神经记录和视频的多模态深度学习

作者:nancy xinru wang, ali farhadi, rajesh rao, bindni brunton

摘要: 在大脑和机器之间开发有用的接口是神经工程的一项重大挑战。一个有效的接口不仅有能力解释神经信号，而且有能力预测人类在不久的将来执行动作的意图;在控制良好的实验室实验之外，预测变得更加具有挑战性。本文介绍了我们检测和预测未来自然人体手臂运动的方法，这是大脑计算机接口中从未尝试过的一个关键挑战。介绍了长期 ecog (ajile) 数据集中的新型附加关节;ajile 包括 7 个上肢关节的自动注释姿势，适用于超过 670小时 (超过7200万帧) 的4名患者，以及相应的同时获得的颅内神经记录。ajile 的尺寸和范围大大超过了以前所有的运动数据集和电皮质成像 (ecog)，从而有可能采取**深度学习**的方法来预测运动。我们提出了一个多模态模型，结合**深**卷积神经网络 (cnn) 与长短期记忆 (lstm) 块，利用 ecog 和**视频**模式。我们证明，我们的模型能够在运动启动前检测到运动并预测未来的运动。此外，我们的多模态运动预测模型显示出对输入神经信号模拟消融的弹性。我们认为，考虑到背景的自然神经解码多模态方法对于推进生物电子技术和人类神经科学至关重要。少

2018 年 3 月 1 日提交;**v1** 于 2017 年 9 月 12 日提交;**最初宣布** 2017 年 9 月。

316. 深度防御：在线加速防御对抗性深度学习

作　者：bita darvish rouhani, mohammad samragh, mojan javaheripi, tara javidi , farinaz koushanfar

文摘: 对抗**深度学习** (dl) 的最新进展为恶意攻击开辟了一个基本未探索的表面，危及自主深度学习系统的完整性。随着 dl 在关键和时间敏感的应用中的广泛使用，包括无人驾驶车辆、无人机和**视频**监控系统，在线检测恶意输入至关重要。我们提出 deepfense，这是第一个端到端自动化框架，可同时高效、安全地执行 dl 模型。deepfense 将阻止对抗攻击的目标正式化为一个优化问题，最大限度地减少了 dl 网络跨越的潜在要素空间中很少观察到的区域。为了解决上述最小化问题，训练了一组互补但不分离的模块冗余，以验证输入样本与受害者 dl 模型平行的合法性。deepfense 利用硬件/软件算法的共同设计和自定义加速，在资源受限的设置中实现实时性能。拟议的对策是无人监督的，这意味着没有对抗样本被用来训练模块冗余。我们还提供了一个配套的 api，以降低非经常性工程成本，并确保自动适应各种平台。对 fpga 和 gpu 的广泛评估显示，

性能提高了两个数量级，同时能够在线进行对抗性样本检测。少

2018 年 8 月 20 日提交;**v1** 于 2017 年 9 月 8 日提交;**最初宣布** 2017 年 9 月。

317. 基于三维卷积神经网络的曲棍球视频中的多标签类不平衡动作识别

作 者 :konstantin sozykin, stanislav protasov, adil khan, Protasov hussain, jooyoung lee

文摘: **视频**自动分析是计算机视觉和机器**学习**领域最复杂的问题之一。这项研究的一个重要部分涉及 (人类) 活动识别 (har)，因为人类及其执行的活动产生了大部分**视频**语义。**基于视频**的 har 在各个领域都有应用，但其中最重要和最具挑战性的是体育**视频**中的 har。一些主要问题包括高的类之间和类内差异、大量的班级不平衡、存在的集体行动和单人游戏的行为，以及识别同时的行动，即多标签**学习**问题。考虑到这些挑战和 cnn 最近在解决各种计算机视觉问题方面取得的成功，在这项工作中，我们实现了基于 3d cn 的多标签深 har 系统，用于曲棍球 **中的多标签不平衡动作识别的视频**。我们测试我们的系统在两个不同的情况下: 一个组合

的 K 二进制网络与单个网络的对比 K-输出网络，在公开可用的数据集上。我们还将结果与最初为所选数据集设计的系统进行了比较。实验结果表明，该方法比现有解决方案性能更好。少

2018 年 5 月 3 日提交;**v1** 于 2017 年 9 月 5 日提交;**最初宣布** 2017 年 9 月。

318. 从图像和视频字幕检索的文本中预测视觉特征

作者:董建峰,李希荣, cees g. m. snoek

摘要: 本文试图在一组句子中找到一个最好的描述给定图像或**视频**内容的句子。与现有的作品不同的是，我们依靠一个联合子空间进行图像和**视频**字幕检索，我们建议专门在视觉空间中这样做。除了这种概念上的新颖性之外，我们还提出了 {word2visualvec}，这是一种**深度**神经网络体系结构,**它学习**从文本输入中预测视觉特征表示。示例字幕被编码到基于多尺度句子矢量化的文本嵌入中，并通过简单的多层感知器进一步转移到所选择的**深层**视觉特征中。我们进一步推广了 word2visualvec, 用于视频字幕检索，方法是从文本中预测三维卷积神经网络特征以及视觉音频表示。在 flickr8k、Flickr8k、microsoft **视频**描述数据集和最近 nist

TrecVid 对**视频**字幕检索的挑战上进行的实验详细介绍了 wordvisualvece 的属性、它相对于文本嵌入的好处、多模式查询组合的潜力及其最先进的结果。少

2018 年 7 月 14 日提交;**v1** 于 2017 年 9 月 5 日提交;**最初宣布** 2017 年 9 月。

319. 基于注意的编码解码器解码网络的视频摘要

作者:钟基,熊凯林,庞艳伟,李学龙

文摘: 本文将监控**视频**摘要表述为序列**到序列学习**问题,其中输入为原始**视频**帧序列，输出为键盘序列，从而解决了监控视频摘要问题。我们的关键思想是**学习**一个**带有**注意力机制的深度摘要网络，以模仿人类关键镜头的选择方式。为此，我们提出了一个新的**视频**摘要框架,名为"视频摘要 (avs) 注意编码器解码器网络"，其中编码器使用双向长期短期内存 (bilstm) 对上下文进行编码。输入**视频**帧之间的信息。对于解码器，分别利用加法和乘法目标函数对两个基于注意的 lstm 网络进行了研究。在三个**视频**摘要基准数据集 (如 summe 和 tvsum) 上进行了大量实验。结果表明，提出的基于 avs 的方法相对于最先进的方法具有优越性，在两个数据集上分别有从 0.8% 到 3% 的显著改进。少

2018 年 4 月 15 日提交;**v1** 于 2017 年 8 月 30 日提交;**最初宣布** 2017 年 8 月。

320. 基于深神经网络的广义反应策略学习

作者:edward groshev, maxwell goldstein, aviv tamar , Siddharth srivastava, pieter abbeel

摘要: 我们提出了一种新的**规划**学习方法，在解决一组特定的规划问题时获得的知识被用来在相关但新的问题实例中更快地进行规划。我们证明,**深度**神经网络可以用来**学习**和表示一个 \ 强调 {广义反应策略} (grp), 它将问题实例和状态映射到一个操作，并且**学习的** grp 可以有效地解决大型类具有挑战性的问题实例。与之前在这方面所做的努力不同，我们的方法大大减少**了学习**对手工制作的领域知识或特征选择的依赖。相反, grp 是使用一组成功的执行跟踪从零开始训练的。我们证明了我们的方法也可以用来自动**学习**一个启发式函数, 可以在定向搜索算法中使用。我们使用两个具有挑战性的规划问题领域的大量实验来评估我们的方法，并表明我们的方法有助于**学习**复杂的决策策略和强大的启发式功能，而只需最少的人工输入。我们的搜索**结果**视频可在 gog/hpy4e3 上查阅。少

2018 年 7 月 24 日提交;**v1** 于 2017 年 8 月 24 日提交;**最初宣布** 2017 年 8 月。

321. sim4cv：一种用于计算机视觉应用的逼真模拟器

作 者 :matthias müller, vincent casser, jean lahoud, neil smith, bernard ghanem

摘要: 我们提出了一个照片逼真的训练和评估模拟器 (sim4cv)，在计算机视觉的各个领域有着广泛的应用。该模拟器建在虚幻引擎之上，集成了全功能的物理车辆、无人驾驶飞行器 (uav) 以及不同城市和郊区 3d 环境中的动画人。我们通过两个案例研究演示了模拟器的多功能性：基于 uav 的运动物体自动跟踪和使用监督学习的自动驾驶.模拟器将几种最先进的跟踪算法与基准评估工具和**深度**神经网络 (dnn) 架构完全集成在一起，用于培训车辆自主驾驶。它生成具有自动地面真实注释的合成照片逼真数据集，以轻松扩展现有的真实数据集，并通过使用自动世界生成工具。补充**视频**可以 https://youtu.be/SqAxzsQ7qUU 少

2018 年 3 月 24 日提交;**v1** 于 2017 年 8 月 19 日提交;**最初宣布** 2017 年 8 月。

322. 视频和球面图像的艺术风格传输

作者:manuel ruder, 阿列克谢·多索维茨基, thomas brox

摘要: 以某种艺术风格手动再现形象需要专业艺术家很长时间。单枪匹马地为**视频**序列执行此操作超出了您的想象。我们提出了两种计算方法，将样式从一个图像(例如，一幅画) 转移到一个完整**的视频**序列。在我们的第一种方法中，我们适应了基于能量最小化的原始图像风格传输技术的**视频**。我们引入了新的初始化方法和新的损耗函数，即使在运动大、遮挡强的情况下，也能生成一致、稳定的**风格化视频**序列。我们的第二种方法将**视频**程式化表述为**一个学习**问题。我们提出了**一个深入的**网络架构和培训程序，使我们能够以一致和稳定的方式，几乎实时地风格化**任意长度的视频**。我们表明，所提出的方法在质量和数量上都明显优于更简单的基线。最后，我们提出了一种方法来适应这些方法也 360 度的图像和**视频**，因为他们出现了最近的虚拟现实硬件。少

2018 年 8 月 5 日提交;**v1** 于 2017 年 8 月 13 日提交;**最初宣布** 2017 年 8 月。

323. 基于模型的无模型微调深部加固学习的神经网络动力学

作者:anusha Nagabandi, gregory kahn, ronald s. 恐惧, sergey levine

文摘: 无模型**深度**增强**学习**算法已被证明能够**学习**广泛的机器人技能，但通常需要非常多的样本才能获得良好的性能。原则上，基于模型的算法可以提供更高效的**学习**，但事实证明很难扩展到有表现力的高容量模型，如**深度**神经网络。在本工作中，我们证明了中型神经网络模型实际上可以与模型预测控制 (mpc) 相结合，以实现优秀的样本复杂性在一个基于模型的增强**学习**算法，产生稳定和似是而非的步态，以完成各种复杂的运动任务。我们还建议使用**深度**神经网络动力学模型来初始化无模型学习者，以便将基于模型的方法的采样效率与无模型方法的高任务特性结合起来。我们在 mujoco 运动任务上实证证明，我们仅在随机动作数据上训练的纯基于模型的方法可以遵循具有出色采样效率的任意轨迹，我们的混合算法可以加速无模型**学习**高速基准任务，在游泳者、猎豹、漏斗和蚂蚁剂身上实现 3-5 倍的样品效率提升。**视频**可以在 https://sites.google.com/view/mbmf 中找到少

2017 年 12 月 1 日提交;**v1** 于 2017 年 8 月 8 日提交;**最初宣布** 2017 年 8 月。

324. 对深度学习模型的强大物理世界攻击

作 者 :kevin eykholt, ivan evtimov, earlence fernandes, bo li, amir rahmati, chaowei xiao, atul prakash, tadayoshi kohno, dawn song

摘要: 最近的研究表明，最先进的**深度**神经网络 (dnn) 容易受到对抗的例子，这是由输入中增加的小尺度扰动造成的。鉴于新兴的物理系统在安全危急的情况下使用 dnn，敌对的例子可能会误导这些系统，造成危险的情况。因此，了解物理世界中的对抗性示例是朝着开发弹性**学习**算法迈出的重要一步。我们提出了一种通用的攻击算法，鲁棒物理扰动 (rp2)，在不同的物理条件下产生鲁棒的视觉对抗扰动。利用道路标志分类的实际案例，我们证明了在各种环境下，使用 rp2 生成的对抗例对物理世界中的标准体系结构路标分类器实现了较高的目标错误分类率条件，包括观点。由于目前缺乏标准化的测试方法，我们提出了一个两阶段的评估方法，包括实验室和现场测试的强大的物理对抗示例。使用这种方法，我们评估物理对抗操纵的有效性对真实的对象。在只有黑白贴纸形式的扰动下，我们攻击一个真正的停止标志，导致 100% 的在实验室环境中获得

的图像出现有针对性的分类，在 84.8 中，在移动车辆(字段) 上捕获的**视频**帧有针对性地分类测试)。少

2018 年 4 月 10 日提交;**v1** 于 2017 年 7 月 27 日提交;**最初宣布** 2017 年 7 月。

325. 野外深度图像的头部检测

作者:diego ballotta, guido borghi, roberto vezzani, rita cucchiara

摘要: 头部检测和定位是一项艰巨的任务，也是许多计算机视觉应用 (如**视频**监控、人机交互和人脸分析) 的关键要素。在 rgb 图像上检测人脸所做的大量工作，以及巨大的人脸数据集的可用性，使得在该域上建立了非常有效的系统。但是，由于照明问题，在实际应用中可能需要红外或深度摄像机。本文介绍了一种利用**深度学习**方法的分类能力的深度图像头部检测新方法。除了减少对外部照明的依赖之外，深度图像隐式嵌入了有用的信息来处理目标对象的比例。利用了两个公共数据集：第一个数据集称为 pandora，用于训练具有人脸和非人脸图像的**深层**二进制分类器。第二种是康奈尔大学收集的，用于在不受限制的环境中的日常活动中执

行交叉数据集测试。实验结果表明，该方法克服了最先进的深度图像方法的性能。少

2017 年 11 月 8 日提交;**v1** 于 2017 年 7 月 21 日提交;**最初宣布** 2017 年 7 月。

326. 从观察的模仿：通过语境翻译从原始视频中学习到模仿行为

作 者 :yuxuan liu, abhishek gupta, pieter abbeel, sergey levine

摘要: 模仿**学习**是自治系统在没有明确奖励功能的情况下获得控制政策的有效方法，使用的是专家（通常是人类操作者）的示范提供的监督。但是，标准的模拟**学习**方法假定代理接收可提供给监督学习算法的观测动作元组的示例. 这与人类和动物的模仿方式形成鲜明对比: 我们观察另一个人的行为，然后找出哪些行为会意识到这种行为，补偿观点、环境、物体位置和类型等方面的变化因素。我们将这种模仿**学习**称为 "模仿观察"，提出了一种基于**视频**预测的基于语境翻译和**深层强化学习**的模仿**学习**方法。这就提高了在模拟**学习**中的假设，即演示应包括在相同环境配置中的观测，并支持各种有趣的应用，包括**学习**机器人技能,只需观察人类工具

使用的**视频**，就可以使用工具。我们的实验结果显示了我们的方法在**学习**各种现实世界中的机器人任务的有效性，这些任务模仿了人类演示者的**视频**，包括扫地、拉过杏仁、推送对象以及模拟中的一些任务。少

2018 年 6 月 18 日提交;**v1** 于 2017 年 7 月 11 日提交;**最初宣布** 2017 年 7 月。

327. 深度强化学习对人员重新识别的注意选择

作者:徐兰,王汉晓,龚少刚，朱夏天

摘要: 现有的人员重新识别（重新识别）方法假定提供准确裁剪的人边界框与最小的背景噪声，主要是通过手动裁剪。在实际应用时，如果必须自动检测到人的边界框，则会处理大量图像和视频，从而显著突破这一点。与手动精心裁剪相比，自动检测到的边界框的精度要低得多，而随机的背景杂波量会显著降低人的重新识别匹配精度。在这项工作中，我们开发了一个联合**学习深度**模型，通过加强**学习**背景杂波最小化，优化了任何自动检测的人边界框中的人重新注意力选择。重新生成标签对约束。具体而言，我们制定了一个新的统一的重新 id 架构，称为身份判别注意强化**学习**(ideal)，以准确地选择自动检测边界框中的重新注意力，从而优化

重新 id 性能。我们的模型可以提高重新定义的准确性，可与详尽的人工人工裁剪边界箱具有更多的优势，从身份歧视注意选择，特别是有利于重新的任务超越人类知识。广泛的比较评估显示，在两个自动检测到的重新识别基准 cuhk03 和 mark-1501 上，拟议的 ideal 模型相对于各种最先进的重新识别方法具有重新的优势。少

2018 年 7 月 7 日提交;**v1** 于 2017 年 7 月 10 日提交;**最初宣布** 2017 年 7 月。

328. 基于图像和文本联合嵌入的多媒体语义完整性评估

作 者 :ayush jaiswal, ekraam sabir , wael abdalmageed, Ekraam natarajan

摘要: 现实世界中的多媒体数据通常由多种模式组成，例如图像或带有关联文本的**视频**(例如字幕、用户评论等) 和元数据。这种多式联运数据包容易纵 ，这些模式的子集可能会被更改为歪曲或重新使用数据包 ，可能会有恶意。因此，必须制定方法来评估或验证这些多媒体包的完整性。使用计算机视觉和自然语言处理方法直接比较图像 (或**视频**) 和相关标题，以验证媒体包的完整性，只有在有限的一组对象和场景中才有可能。本文

提出了一种新的**基于深度学习**的方法，利用一组参考多媒体包，对包含图像和字幕的多媒体包的语义完整性进行评估。我们在一个框架中构造了一个图像和字幕**的**联合嵌入，并在参考数据集上进行**深度**多模态表示学习，该框架还提供了图像标题一致性得分 (iccs)。查询媒体包的完整性被评估为查询 iccs 相对于引用数据集的完整性。我们提出了多模型信息管理数据集 (maim)，这是 flickr 媒体包的一个新数据集，我们将其提供给研究社区。我们使用新创建的数据集以及 flickr30k 和 ms coco 数据集对我们建议的方法进行定量评估。引用数据集不包含未操作的被篡改查询包版本。我们的方法能够在 maim、flickr30k 和 ms coco 上分别获得 0.75、0.75 和 0.75 的 f1 分数，用于检测语义上不连贯的媒体包。少

2018 年 6 月 28 日提交;**v1** 于 2017 年 7 月 5 日提交;**最初宣布** 2017 年 7 月。

329. 神经系统 slam：学习使用外部记忆进行探索

作 者：张 景 伟 , 李 泰 , jjochka boedecker, wolfram burkard, ming liu

摘要: 我们为代理提供了一种从传感器数据中**学习**全局地图表示的方法，以帮助他们在新环境中进行探索。为了实现这一点，我们将类似于传统的同时本地化和映射（slam）的过程嵌入到基于软关注的外部内存体系结构寻址中，其中外部内存充当外部内存的内部表示形式。环境。这种结构鼓励了一个完全可微的**深**部神经网络内类似 slam 的行为的进化。我们表明，这种方法可以帮助**强化学习**代理，成功地探索新的环境，其中长期记忆是必不可少的。我们在具有挑战性的电网世界环境和初步的凉亭实验中验证了我们的方法。我们实验的**视频**可以在 https://goo.gl/G2Vu5y 上找到。少

2017 年 11 月 29 日提交;**v1** 于 2017 年 6 月 28 日提交;**最初宣布** 2017 年 6 月。

330. 自然视频序列预测中的分解运动和内容

作 者 :ruben villegas, jmei yang, seunghoon hong, xunyulin, hong lak lee

文摘: 我们提出了一个**深**神经网络来预测自然**视频**序列中的未来帧。为了有效地处理**视频**中像素的复杂演化，我们提出了分解视频中产生动态的两个关键组件的运动和内容。我们的模型是建立在编码器卷积神经网络和

卷积 lstm 的像素级预测，独立捕获图像的空间布局和相应的时间动态。通过独立建模运动和内容，预测下一帧可减少通过识别的运动特征将提取的内容要素转换为下一帧内容，从而简化了预测任务。我们的模型可在多个时间步骤中进行端到端培训，并且自然**无需单独培训即可自然地学习**分解运动和内容。我们使用 kth、weizmann 操作和 ucf-101 数据集评估人类活动**视频**的拟议网络体系结构。与最近的方法相比，我们展示了最先进的性能。据我们所知，这是第一个具有运动和内容分离的端到端可训练网络体系结构，用于模拟自然**视频**中像素级未来预测的时空动力学模型。少

2018 年 1 月 7 日提交;**v1** 于 2017 年 6 月 25 日提交;**最初宣布** 2017 年 6 月。

331. 使用地理标记视频对人类活动进行大规模映射

作者:朱毅,刘森,肖恩·纽卡姆

文摘: 本文是利用地理标记**视频**的视觉内容进行人类活动时空映射的第一项工作。我们利用最近的**一个基于深度学习**的**视频**分析框架，称为隐藏的双流网络，以识别 youtube **视频**中的一系列活动。此框架是高效的，可以实时或更快地运行，这对于识别流**媒体视频**中发生的

事件或减少分析已捕获的**视频**的延迟非常重要。这反过来对于在智能城市应用程序中使用**视频**也很重要。我们进行了一系列实验，以表明我们的方法能够准确地映射空间和时间上的活动。我们还展示了使用视觉内容而不是标签标题的优势。少

2017 年 11 月 28 日提交;**v1** 于 2017 年 6 月 24 日提交;**最初宣布** 2017 年 6 月。

332. 深度监管离散冲击

作者:李启强，孙振南,何然,谭铁牛

文摘: 随着网络上图像和**视频**数据的快速增长，哈希在图像或**视频**搜索方面得到了广泛的研究。从**深度学习**的最新进展中受益,**深度**哈希方法在图像检索方面取得了很有希望的效果。但是，以前**的深层**哈希方法存在一些限制 (例如，语义信息未被充分利用)。本文基于学习的二进制码应该是理想的分类的假设，提出**了**一种**深度**监督离散哈希算法。对标的信息和分类信息都用于在一个流框架中**学习**哈希代码。将最后一层的输出直接限制为二进制码，这在**深度**哈希算法中很少被研究。由于哈希码的离散性质，采用交替最小化方法对目标函数进

行优化。实验结果表明，该方法在基准数据集上的性能优于目前最先进的方法。少

2017 年 11 月 27 日提交;**v1** 于 2017 年 5 月 31 日提交;**最初宣布** 2017 年 5 月。

333. quo vadis，行动识别？一种新的模型与动力学数据集

作者:joao carreira, Carreira zisserman

摘要：由于目前的行动分类数据集 (ucf-101 和 hmdb-51) 中缺乏**视频**，因此很难确定良好的**视频**架构，因为大多数方法在现有的小规模基准上获得了类似的性能。本文根据新的动力学**人体行动视频**数据集，重新评估了最先进的体系结构。动力学有两个数量级以上的数据，每班有 400 个人类动作类和 400 多个剪辑，这些数据是从现实的、具有挑战性的 youtube **视频**中收集的。我们分析了当前体系结构在此数据集上的操作分类任务的效果，以及在对动力学进行预培训后，在较小的基准数据集上提高了多少性能。我们还引入了一种基于 2d convnet 充气的新的双流充气 3d convnet (i3d)：将非常**深**的图像分类凸网的过滤器和池核扩展到 3d，从而实现无缝**学习**时空特征提取从**视频**，同时利用成功

的 imagenet 架构设计，甚至其参数。我们表明，经过动力学的预培训后，i3d 模型大大改进了行动分类方面的最先进技术，在 hmdb-51 上达到 80.9，在 ucf-101 上达到 98.0%。少

2018 年 2 月 12 日提交;**v1** 于 2017 年 5 月 22 日提交;**最初宣布** 2017 年 5 月。

334. 用于操作识别的二阶时间池

作者:anop cherian, stephen gould

摘要: **深部**基于**视频**的**动作**识别学习模型通常会生成短剪辑的特征（由几个帧组成);此类剪辑级别的功能通过计算这些功能的统计信息聚合到**视频**级表示。通常使用零（最大值）或一阶（平均）统计信息。在本文中，我们探讨了使用二阶统计的好处。具体而言，我们提出了一种新的端到端可学习特征聚合方案，称为时间相关池，通过捕获剪辑级 cnn 功能通过**视频**计算。这样的描述符，虽然计算成本很低，但也自然编码多个 cnn 功能的协同激活，从而提供了比他们的一阶对应对行动更丰富的描述。我们还提出了该方案的高阶扩展，方法是在复制内核希尔伯特空间中嵌入 cnn 功能后计算相关性。我们提供关于基准数据集（如 hmd-51 和 ucf-101)、

细粒度数据集（如 mpii 烹饪活动和 jhmdb）以及最近的 kinics-600 的实验。我们的结果证明了高阶池方案的优势，当与手工制作的功能（标准做法）结合使用时，可实现最先进的精度。少

2018 年 8 月 6 日提交;**v1** 于 2017 年 4 月 23 日提交;**最初宣布** 2017 年 4 月。

335. 用于操作检测的预测纠正网络

作者:achal dave, olga russakovsky, deva ramanan

摘要: 虽然**深度**特征**学习**使静态图像理解的技术发生了革命性的变化，但视频处理的技术却不太成立。用于**视频**的体系结构和优化技术在很大程度上是基于静态图像的体系结构和优化技术，这可能是对丰富**视频**信息的充分利用不足。在这项工作中，我们重新思考了基础网络体系结构和随机**学习**范式的时间数据。为此，我们从线性动态系统的经典理论中汲取灵感，用于建模时间序列。通过将这些模型扩展到包括非线性映射，我们推导出一系列新的递归神经网络，这些神经网络依次对未来进行自上而下的预测，然后用自下而上的观测结果纠正这些预测。预测校正网络具有许多理想的特性: (1) 它们可以自适应地将计算集中在预测需要较大校正

的 "令人惊讶" 的帧上, (2) 它们简化**了学习**，因为只有"残差一样"纠正术语需要随着时间的推移**而学习**, (3) 它们自然以分层的方式将输入数据流联系起来，从而产生更可靠的信号，用于在网络的每一层学习。我们对我们的轻量级和可解释框架进行了广泛的分析，并证明了我们的模型在三个具有挑战性的数据集上与双流网络具有竞争力，而不需要计算开销大的光流。少

2017 年 12 月 12 日提交;**v1** 于 2017 年 4 月 12 日提交; **最初宣布** 2017 年 4 月。

336. 人脸识别顶级性能的良好实践：转移深度特征融合

作者:林雄,贾亚什里·卡勒卡尔,赵健,程毅, 徐燕, 冯家志, 苏吉里·普拉纳塔，沈胜梅

摘要: 在过去几年中，不受约束的人脸识别性能评估传统上侧重于影像的野生 (lfw) 数据集中的标记面和**视频**的 youtubefaces (ytf) 数据集。这一领域的惊人进展使这些基准数据集的验证和识别精度饱和。在本文中，我们提出了一个统一的**学习**框架，称为转移**深度**特征融合 (tdff) 针对新的 iarpa j 间 us 基准 a (iibb-a) 的人脸识别数据集发布的 nist 面临的挑战。ijb-a 数据集包括来自 500 个主体的无约束的实际面孔，这些主题具

有完整的姿势和照明变化，比 lfw 和 ytf 数据集困难得多。在迁移**学习**的启发下，我们分别在源域中训练了两个具有两个不同大数据集的先进的**深**卷积神经网络(dcnn)。通过对两种不同 dcn 的互补性的探索，在目标域中进行特征提取后，利用**了深度**特征融合。然后，采用模板特定的线性支持向量机来增强框架的判别。最后，将多个匹配分数对应不同模板合并为最终结果。这个简单的统一框架在 ijb-a 数据集上表现出出色的性能。根据建议的方法，我们已将 ijb-a 结果提交给国家标准和技术研究所 (nist) 进行官方评估。此外，通过引入新数据和高级神经架构，我们的方法在 ijb-a 数据集上的性能大大优于最先进的数据。少

2018 年 2 月 9 日提交;**v1** 于 2017 年 4 月 3 日提交;**最初宣布** 2017 年 4 月。

337. 谁更好？谁是最好的？用于技能确定的配对深度排序

作 者 :hazel doughty, dima damen , walterio mayol-cuevas

摘要: 我们提出了一种方法来评估技能从**视频**，适用于各种任务，从手术到绘制和滚动比萨饼面团。我们将问

题表述为视频收藏的对等（谁更好？）和总体（谁是最好的？）排名，使用监督**的深度**排名。我们提出了一个新的损失函数，当一对**视频**表现出技能的差异时，**它可以学习**有鉴别力的特征，当一对**视频**表现出可比的技能水平时，**我们会学习**共享的特征。结果证明，我们的方法适用于各个任务，四个数据集的正确订购视频对的**百分比**从 70% 到 83% 不等。通过对其参数的敏感性分析，我们证明了我们的方法的鲁棒性。我们认为这项工作是为了实现操作视频集合的自动化组织和视频中的整体通用技能确定. 少

2018 年 3 月 29 日提交;**v1** 于 2017 年 3 月 29 日提交;**最初宣布** 2017 年 3 月。

338. 边界流：一种在没有运动训练的情况下预测边界运动的暹罗网络

作者:彭磊,李福新,西尼萨·托多罗维奇

摘要: 利用**深度学习**，解决**了视频**中的关节目标边界检测和边界运动估计问题，并将边界流估计称为边界流估计。边界流是一种重要的中层视觉提示，边界是物体空间范围的特征，而边界是指物体运动和相互作用的。然而，以前关于运动估计的大多数工作都集中在密集

的物体运动或特征点上，这些点可能不一定存在于边界上。对于边界流估计，我们指定了一个新的完全卷积暹罗网络（fcsn），它共同估计两个连续帧中的对象级边界。两个帧中的边界对应由同一 fcsn 使用一种新的非常规反卷积方法进行预测。最后，利用基于边缘的滤波方法对边界流估计进行了改进。对**视频**中的边界检测、边界流估计和光流估计三个方面进行了评价。在边界检测方面，我们在基准 vsb100 数据集上实现了最先进的性能。在边界流估计方面，我们在 sintel 训练数据集上给出了第一个结果。对于光流估计，我们运行了最近的 pmflow 方法，但在增强输入上与边界流匹配，并在 sintel 基准上实现了显著的性能改进。少

2018 年 4 月 8 日提交;**v1** 于 2017 年 2 月 27 日提交;**最初宣布** 2017 年 2 月。

339. 通过自然语言进行上下文定制的视频摘要

作者:崔金秀,泰贤欧，在苏昆

摘要: 由于视频的高度主观性，长视频的最佳总结因其高度主观性而在不同的人之间有所不同。即使对同一个人来说，最好的总结也可能会随着时间或情绪的变化而改变。本文介绍了通过简单的文本生成自定义**视频摘**

要的任务。首先，我们训练一个**深入**的体系结构，通过渐进和剩余的方式利用丰富的图像标题数据，有效**地学习**视频帧的语义嵌入.给定特定于用户的文本描述，我们的算法能够选择语义相关的**视频**段，并生成临时对齐的**视频**摘要。为了评估我们的文本定制**视频**摘要，我们与利用地面真相信息的基线方法进行了实验比较。尽管基线具有挑战性，但我们的方法仍然能够显示出可比甚至超过性能。我们还表明，我们的方法能够生成语义上多样化的**视频**摘要，只需利用**学习的**视觉嵌入。少

2018 年 3 月 2 日提交;**v1** 于 2017 年 2 月 6 日提交;**最初宣布** 2017 年 2 月。

340. 利用完全卷积网络检测视频显著的目标

作者:王文关,沈建兵,邵玲

文摘: 本文提出了一种**深度学习**模型，有效地检测**视频**中的突出区域。它解决了两个重要问题: (1) 在没有足够大的、像素化的注释视频数据的**情况下进行深度**视频显著性模型训练，以及 (2) 快速**的视频**显著性培训和检测. 提出的**深度视频**显著性网络由两个模块组成，分别用于捕获空间和时间显著性信息。动态显著性模型明确

地结合了静态显著性模型中的显著性估计，直接产生时空显著性推理，而无需耗时的光流计算。我们进一步提出了一种新的数据增强技术，该技术模拟现有的带注释的图像数据集中的**视频**训练数据，使我们的网络能够**学习**不同的显著性信息，并防止过度符合有限的培训**视频**数量。利用我们的合成**视频**数据（150k **视频**序列）和真实**视频**，我们**深入**的**视频**显著性模型成功地学习了空间和时间显著**性**提示，从而产生准确的时空显著性估计。我们在 davis 数据集（. 06 的 mae）和 fbms 数据集（. 07 的 mae）上推进了最先进的数据集，并以更高的速度（2fps，包括所有步骤）进行了改进。少

2017 年 12 月 8 日提交;**v1** 于 2017 年 2 月 2 日提交;**最初宣布** 2017 年 2 月。

341. 使用组表示法理解图像运动

作者:andrew jaegle, stephen phillips, daphne ippolito , kostas danilidis

摘要: 运动是动态环境中代理的一个重要信号，但**学习**表示无标记**视频**中的运动是一个困难且不受约束的问题。提出了一种基于变换基本群性质的运动模型，并利

用该模型对图像运动进行了表示。虽然大多数估计运动的方法都是基于像素级约束的，但我们使用这些群属性来约束运动本身的抽象表示。我们证明了使用此方法训练的**深度**神经网络可以捕获合成二维序列和实际车辆运动序列中的运动，而不需要任何标签。训练以考虑到这些约束的网络隐式地识别不同序列类型中运动的图像特征。在车辆运动的背景下，该方法提取对定位、跟踪和气味测量有用的信息。我们的结果表明，这种表示是有用的学习运动在一般设置中，显式标签是很难获得。少

2018 年 2 月 26 日提交;**v1** 于 2016 年 12 月 1 日提交;**最初宣布** 2016 年 12 月。

342. 人们会喜欢你的形象吗？学习审美空间

作者:katharina schwarz, patrick wieschollek, hendrik p. a. lensch

摘要: 对图像的美观程度进行评级是一件非常复杂的事情，取决于大量不同的视觉因素。以前的作品已经解决了审美评级问题，在一维评级表上排名，例如，结合手工制作的属性。在本文中，我们提出了一个相当通用的方法，以自动映射审美愉悦的所有复杂性到一个 "审美

空间"，以允许一个高度细粒度的分辨率。在细节上，利用**深度学习**，我们的方法直接**学习**到这个高维的特征空间类似于视觉美学的给定图像的编码。除了上述视觉因素外，个人判断的差异对照片的亲和力也有很大影响。如今，在线平台允许用户 "喜欢" 或青睐某些内容与一个单一的点击。为了融合了各种各样的人，我们利用这种多用户协议，并收集了一个包含相关元信息的 380k 图像 (arod) 的大型数据集，并获得评分来评价特定照片在视觉上的赏心悦目。我们在用户研究中验证了我们导出的美学模型。此外，在没有任何额外数据标签或手工制作的功能的情况下，我们在 ava 基准数据集上实现了最先进的准确性。最后，由于我们的方法能够预测任何任意图像或**视频**的审美质量，我们演示了我们在应用中使用照片收藏、在移动设备上捕捉最佳拍摄和美观的关键帧提取的结果从**视频**。少

2017 年 12 月 4 日提交;**v1** 于 2016 年 11 月 16 日提交;**最初宣布** 2016 年 11 月。

343. 基于最小监督的广义判别模型的学习场景特定对象检测器

作者:罗大鹏，曾志鹏,农生，吴翔,龙生伟，周全正，郑军，陈罗

摘要: 由于不同的照明、背景和相机视点，一个对象类可能会显示较大的变化。在不受限制**的视频**环境下，传统的对象检测方法的性能通常更差。为了解决这个问题，许多现代方法为对象检测建立了深层分层外观表示的**模型**。这些方法大多需要对大型手动贴标样本集进行耗时培训。本文提出的框架以一种自下而上的方式解决多场景检测问题，采取了截然不同的方向。首先，通过鼠标在第一个视频帧中标记对象周围的几个边界框，从一个完全**自主**的**学习**过程中获得特定场景的对象。在这里，不需要人工标记的训练数据或通用检测器。其次，这种**学习**过程在不同的监控场景中方便地多次复制，并在不同的相机视点下产生特定的探测器。因此，该框架可用于多场景对象检测应用，且监督最少。显然，初始场景特定检测器由多个边界框初始化，检测性能较差，传统的在线**学习**算法难以改进。因此，我们提出了生成判别模型来划分检测响应空间，并为每个分区分配一个单独的描述符，逐步达到较高的分类精度。提出了一种新的在线渐进优化过程，以优化广义判别模型，并将重点放在硬样本上。对六个**视频**数据集的实验结果表明，我们的方法实现了与鲁棒监督方法的可比性能，并在不同成像条件下优于最先进的**自学习**方法。少

2018 年 3 月 12 日提交;**v1** 于 2016 年 11 月 12 日提交;**最初宣布** 2016 年 11 月。

344. 视频分类和字幕的深度学习

作者:吴祖轩,姚婷,傅艳伟,姜玉刚

摘要: 随着互联网带宽和存储空间的大幅增加,**视频**数据被生成、发布和爆炸传播，成为当今大数据不可或缺的一部分。本文重点回顾了两行旨在通过**深度学习**来激发**视频理解的**研究:**视频**分类和**视频**字幕。虽然**视频**分类集中在根据视频剪辑的语义内容（如人类行为或复杂事件）自动标记**视频**剪辑时,**视频**字幕则试图生成完整而自然的句子,丰富视频分类中的单个标签，以捕捉**视频**中最翔实的动态。此外，我们还对流行的基准和竞赛进行了审查，这些基准和竞赛对于评价这一充满活力的领域的技术进步至关重要。少

2018 年 2 月 22 日提交;**v1** 于 2016 年 9 月 21 日提交;**最初宣布** 2016 年 9 月。

345. 用于动态显著性预测的时空显著性网络

作 者 :cagdas bak, aysun kocak, erkut erdem, Erkut erdem

文摘: 近年来，静止图像的计算显著性模型得到了极大的欢迎。另一方面，视频中的显著预测，得到的社区兴趣相对较少。在此基础上，我们研究了**深度学习**在动态显著性预测中的应用，并提出了所谓的时空显著性网络。我们模型的关键是双流网络的体系结构，在这种结构中，我们研究不同的融合机制，以整合空间和时间信息。我们在 diem 和 ucf-体育数据集上评估我们的模型，并提供与现有最先进模型相比极具竞争力的结果。我们还利用从这些图像中预测的光学流图，对 mit300 数据集的一些静止图像进行了一些实验。结果表明，以这种方式考虑固有的运动信息有助于静态显著性估计。少

2017 年 11 月 15 日提交;**v1** 于 2016 年 7 月 16 日提交;**最初宣布** 2016 年 7 月。

346. 用于视频压缩传感的深完全连接网络

作 者 :michael lliadis, leonidas spinoulas, Aggelos k. kadsaggelos

文摘: 在这项工作中，我们提出了一个**深入的学习**框架，视频压缩传感。与以前的方法相比，该配方可在几秒钟内恢复**视频**帧，并显著提高重建质量。我们的研究首先

学习**视频**序列和相应的测量帧之间的线性映射，结果提供了有希望的结果。然后，我们将线性公式扩展到**深度**完全连接的网络，并利用更深入的体系结构探索性能提升。我们的分析总是由所建议的框架在现有压缩**视频**架构上的适用性所驱动的。对多个**视频**序列进行的大量模拟记录了我们的方法在数量和质量上的优越性。最后，我们的分析为了解数据集大小和图层数量如何影响重建性能提供了深入的见解，同时为将来的调查提出了几点建议。代码可在 github: https://github.com/miliadis/DeepVideoCS 少

2017 年 12 月 16 日提交;**v1** 于 2016 年 3 月 15 日提交;**最初宣布** 2016 年 3 月。

347. 基于正则深度神经网络的视频分类中的特征与类关系开发

作者:姜玉刚,吴祖轩，王军,薛向阳,张世福

文摘: 本文研究了根据高级语义对**视频**进行分类的具有挑战性的问题，如特定人类行为或复杂事件的存在。虽然近年来投入了大量的精力，但大多数现有的作品都使用简单的融合策略将多个**视频**功能结合起来，而忽略了类间语义关系的利用。本文提出了一个新的统一框

架，该框架共同利用特征关系和类关系来提高分类性能。具体而言，这两种类型的关系是通过在**深度**神经网络 (dnn) 的**学习**过程中严格实施规则来估计和利用的。这种规范化的 dnn (rdnn) 可以通过基于 gpu 的实现有效地实现，并具有经济实惠的培训成本。通过对 dnn 进行武装，使其具有更好的利用特征和类关系的能力，使所提出的 rdnn 更适合**于视频**语义的建模。通过广泛的实验评估，我们表明，与几种最先进的方法相比，rdnn 具有卓越的性能。在著名的 hollywood 2 和 columbia 消费者**视频**基准上，我们获得了非常有竞争力的结果：平均精度分别为 6.9% 和 73.5%。此外，为了对我们的 rdnn 进行实质性评估并促进未来对大规模**视频**分类的研究，我们收集并发布了一个新的基准数据集，称为 fcvid，其中包含 91,223 互联网**视频和** 239 手动视频注释的类别。少

2018 年 2 月 21 日提交;**v1** 于 2015 年 2 月 25 日提交;**最初宣布** 2015 年 2 月。