

目标识别前沿论文最新进展

2018.11.06 方建勇

提示：采用手机 safari 微软翻译技术

1. 时间序列分类的转移学习

作者: [hassan ismail fawaz](#), [germain forestier](#), [jonathan weber](#), [lhassane idumghar](#), [pierre-alain muller](#)

摘要: 深度神经网络的迁移学习是首先在源数据集上训练基础网络，然后将学习的要素（网络的权重）传输到第二个网络以在目标数据集上进行训练的过程。这一思想已被证明可以提高深度神经网络在**图像识别**和对象定位等许多计算机视觉任务中的泛化能力。除了这些应用外，深卷神经网络（cnn）最近也在时间序列分类（tsc）社区中受到欢迎。然而，与**图像识别**问题不同的是，tsc任务的转移学习技术尚未得到深入研究。这令人惊讶，因为如果从预先训练的神经网络对模型进行微调，而不是从零开始训练，tsc 深度学习模型的准确性可能会得到提高。在本文中，我们通过研究如何为海训方案任务转移深度 cnn 来填补这一空白。为了评估转移学习的潜力，我们使用 ucr 归档文件进行了广泛的实验，该归档文件是包含 85 个数据集的最大的公开可用 tsc 基准。对于归档文件中的每个数据集，我们预先训练了一个模型，然后对其进行了其他数据集的微调，从而产生 7140 种不同的深层神经网络。这些实验表明，根据用于传输的数据集，转移学习可以改

进或降低模型的预测。因此, 为了预测给定目标数据集的最佳源数据集, 我们提出了一种基于动态时间扭曲来测量数据集间相似性的新方法。我们介绍了我们的方法如何指导传输, 以选择最佳的源数据集, 从而提高 85 个数据集中的 71 个数据集的准确性。
少

2018 年 11 月 5 日提交;最初宣布 2018 年 11 月。

2. fedml: 了解预处理噪声过滤对对抗性机器学习的影响

作者: [faiq khalid](#), [muhammad abdullah hanif](#), [semeen rehman](#) ,
[junaid qadir](#), [muhammad shafique](#)

摘要: 基于深度神经网络 (dnn) 的机器学习 (ml) 算法最近已成为领先的 ml 范式, 特别是在分类任务方面, 因为它们具有从大型数据集高效学习的卓越能力。然而, 发现了一些众所周知的攻击, 如数据集中毒、对抗示例和网络操作 (通过添加恶意节点), 这让人们完全关注基于 dnn 的 ml 系统缺乏安全性。特别是, 恶意行为者可能会利用这些众所周知的攻击, 通过只对环境参数进行轻微但系统的操作, 从而导致随机/有针对性的错误分类, 或导致预测信心的变化, 推理数据, 或数据采集块。但是, 以前的大多数对抗攻击都没有考虑到通常与 ml 推理模块集成的预处理噪声滤波器。我们在这项工作中的贡献是要显示, 这是一个重大的遗漏, 因为这些噪音过滤器会使现有的大部分攻击失效, 而这些攻击基本上是依赖引入对抗性噪音。除此之外, 我们还通过提出一

种新的预处理噪声滤波感知逆转录酶 ml 攻击称为 fad-ml 攻击来扩展了最先进的技术。为了证明拟议方法的有效性, 我们利用为 "德国交通标志识别基准" (gtsrb) 培训的 "vggnet" dnn 生成了一个敌对攻击图像, 尽管该数据集没有视觉噪声, 甚至在预处理噪声滤波器的存在下也会导致分类器的分类错误。少

2018 年 11 月 4 日提交;最初宣布 2018 年 11 月。

3. isa4ml: 训练数据--对自主车辆机器学习模块的不知情安全攻击

作者 :[faiq khalid](#), [muhammad abdullah hanif](#), [semeen rehman](#), [muhammad shafique](#)

摘要: 由于大数据分析能力, 机器学习 (ml) 算法在自主车辆的多个应用中越来越流行。然而, ml 算法具有固有的安全漏洞, 这增加了对强大 ml 算法的需求。最近, 各种组已经证明, 如何利用 ml 中的漏洞来执行多个安全攻击, 以减少信心和随机/有针对性的错误分类, 通过使用数据操作技术。这些传统的数据操作技术, 特别是在训练阶段, 引入了随机视觉噪声。但是, 在攻击或测试过程中, 可以通过噪声检测过滤或人在环检测到这种视觉噪声。本文提出了一种利用训练后的深部神经网络 (dnn) 的反向传播特性自动生成 "潜移默化攻击" 的新方法。与最先进的推理攻击不同, 我们的方法不需要了解攻击图像生成过程中的训练数据集。为了说明所建议方法的有效性, 我们提出了一个案例研究, 用于交通标志检测在一个自主驾驶用例。我们部署了针对德国交通标

志识别基准 (gtsrb) 数据集的最先进的 vggnet dnn 培训。我们的实验结果表明, 在主观测试 (即视觉感知) 和客观测试 (即相关性和结构相似性指数没有任何明显变化) 中, 所产生的攻击都是无法察觉的, 但仍然会得到影响成功的错误分类攻击。少

2018 年 11 月 2 日提交;最初宣布 2018 年 11 月。

4. 变形: 基于 3d cnn 的视频理解的灵活加速

作者:kartik hegde, rohit Agrawal, yulun yao, christopher w. fletcher

文摘 在过去几年中, 卷积神经网络 (cnn) 的使用和加速器的设计都出现了爆炸式增长, 使 cnn 推理成为现实。在架构界, 最大的努力是针对美国有线电视新闻网的图像识别推理。与之密切相关的视频识别问题作为加速器目标受到的关注要少得多。这令人惊讶, 因为视频识别比图像识别需要更多的计算, 预计视频流量将在未来几年成为互联网流量的大多数。本文通过提供了一种设计空间探索和灵活的结构来加速三维卷积神经网络 (3d cnn)——现代视频的核心内核——填补了视频识别算法和硬件进步之间的空白。理解。与用于图像识别的 (2d) cnn 相比, 高效加速 3d cnn 带来了重大的工程挑战, 因为它们的内存占用空间大 (且随时间变化), 且尺寸更高。为了应对这些挑战, 我们设计了一种名为 "变形" 的新型加速器, 它可以根据每个目标 3d cnn 的每一层的需要, 自适应地支持不同的空间和时间平铺策略。我们在变形硬

件的同时设计了一个软件基础架构,以找到合适的参数来控制硬件。在最先进的 3d cnn 上进行评估后, morph 实现了高达 3.4 倍(平均 2.5 倍)的能耗,并比基线 3d cnn 加速器提高了 5.1 x (平均 4 倍)的性能,面积开销为 5%。与 eyeriss 相比, morph 在 3d cnn 上进一步实现了 15.9x 的平均能量降低。少

2018 年 10 月 16 日提交;最初宣布 2018 年 10 月。

5. 基于两级特征转移的深卷积神经网络特征表示分析--弥漫性肺病分类分析

作者 :aiga suzuki, hidenori sakanashi, shoji kido, hayaru shouno

摘要: 迁移学习是一种机器学习技术,旨在通过使用从其他学习任务中获得的预先训练的参数来提高泛化性能。对于**图像识别**任务,以前的许多研究都报告说,当将迁移学习应用于深层神经网络时,尽管训练数据有限,但性能仍在提高。提出了一种以纹理医学**图像识别**为重点的两阶段特征传递学习方法。在该方法中,连续训练了大量的自然**图像**、一些纹理**图像**和**目标图像**。将该方法应用于弥漫性肺疾病纹理 x 线计算机断层扫描**图像**的分类工作。在我们的实验中,与零学习和传统的单级特征传输相比,两阶段特征转移实现了最佳性能。我们还研究了**基于大小**的目标数据集的鲁棒性。与其他两种学习方法相比,两阶段特征传输表现出更好的鲁棒性。此外,我们还使用可视化方法分析了从 dddd 图像

输入中获得的每个要素传输模型的特征表示。我们表明, 两阶段特征转移既获得了荒漠化管的边缘特征, 也获得了荒漠化管的纹理特征, 这在传统的单级特征转移模型中并不存在。少

2018 年 10 月 15 日提交;最初宣布 2018 年 10 月。

6. 3d 网格中逼真的对抗示例

作者:杨大伟,肖朝伟,李波, 邓佳, 刘明燕

摘要: 深部神经网络 (dnn) 等高表现力模型已广泛应用于各种应用, 并取得了越来越大的成功。然而, 最近的研究表明, 这种机器学习模式似乎不利于敌对例子。到目前为止, 对 2d 图像的对抗性示例已经进行了大量的探索, 而很少有作品能够理解现实世界中存在的 3d 对象的漏洞, 在这些对象中, 3d 对象通过不同的照片拍摄投影到 2d 域学习 (识别) 任务。在本文中, 我们通过操作对象的给定三维网格表示的形状和纹理来考虑实际场景中的对抗行为。我们的目标是使用逼真的呈现器将优化的 "敌对网格" 投影到 2d, 并且仍然能够误导不同的机器学习模型。大量实验表明, 通过对三维网格的形状或纹理产生不明显的三维对抗摄动, 相应的投影二维实例要么导致分类器将受害者对象错误地归类为任意恶意目标, 或从对象探测器隐藏场景中的任何目标对象。我们进行人类研究, 以表明我们优化的对抗 3d 摄动对于人类视觉系统是非常不明显的。除了给定 3d 网格的细微摄动外, 我们还建议合

成一个逼真的 3d 网格, 并将其放入模拟类似渲染条件的场景中, 从而攻击不同的机器学习模型。提供了对各种 3d 呈现器和网格脆弱区域之间的可转移性的深入分析, 以帮助更好地理解现实世界中的对抗行为。少

2018 年 10 月 11 日提交;最初宣布 2018 年 10 月。

7. 超越外观-将真实图像映射到无监督 cid 识别的几何域

作者: [benjamin planche](#), [sergey zakharov](#), [ziyan wu](#), [andoras hutter](#), [harald kosch](#), [Slobodan ilic](#)

摘要: 虽然卷积神经网络在计算机视觉领域占主导地位, 但人们通常无法获得训练所需的大量领域相关数据。因此, 使用可用的合成样本沿域适应方案为**目标域**准备算法变得很普遍。针对这一问题, 我们从不同的角度引入了一个管道, 将看不见的**目标样本**映射到用于训练特定任务方法的合成域中。对数据进行去噪, 只保留这些**识别**算法所熟悉的功能, 我们的解决方案大大提高了它们的性能。由于这种映射比相反的映射更容易学习 (即学习生成现实的功能来增加源样本), 我们演示了如何完全在增强合成数据上训练我们的整个解决方案, 并且仍然比训练过的方法表现更好与域相关的信息 (如 3d 模型的真实**图像**或逼真纹理)。应用我们的方法从无纹理 cad 数据**进行对象识别**, 我们提出了一个自定义生成网络, 充分利用纯几何信息来学习可靠的特征, 并为看不见的颜色实现更精细的映射图像。少

2018 年 10 月 9 日提交;最初宣布 2018 年 10 月。

8. 一种混合神经网络框架及其在雷达自动目标识别中的应用

作者:[张哲](#),[陈翔](#),[智田](#)

摘要: 深神经网络 (dnn) 已在各种信号处理 (sp) 问题中得到应用。大多数工作要么直接采用 dnn 作为黑盒方法来执行某些 sp 任务, 而不考虑信号模型的任何已知属性, 要么将预定义的 sp 运算符插入 dnn 作为附加数据处理阶段。本文提出了一种新的混合神经网络框架, 其中一个或多个 sp 层以连贯的方式插入 dnn 体系结构, 以提高网络能力和特征提取效率。这些 sp 图层经过正确设计, 可以很好地利用可用的模型和数据的属性。混合神经网络训练算法旨在通过同时优化 dnn 的权重和 sp 运算符的未知调优参数, 将 sp 层积极地纳入学习目标。针对雷达自动目标识别(atr) 问题, 对所提出的混合神经网络进行了测试。通过 5, 000 张雷达 atr 训练图像, 实现了 96% 的高验证精度。与普通 dnn 相比, 混合神经网络可以显著减少所需的训练数据量, 提高学习成绩。少

2018 年 9 月 27 日提交;最初宣布 2018 年 9 月。

9. 通过拼写学习阅读: 实现无监督文本识别

作者:[ankush gupta](#), [andrea vedaldi](#), [andrew zisserman](#)

文摘: 本工作提出了一种在不使用任何配对监控数据的情况下进行可视化**文本识别**的方法。我们将**文本识别**任务表述为将给定**文本图像**预测的字符串的条件分布与从**目标**语料库中采样的词汇有效字符串对齐。这样就可以完全自动化和无监督地从行级**文本图像**和未配对的**文本字符串**示例中进行学习,从而无需对大型数据集进行调整。我们对拟议方法的各个方面进行了详细分析,即—(1) 训练序列长度对收敛的影响, (2) 字符频率与学习顺序之间的关系, 以及 (3) 演示我们的**识别网络**对任意长度输入的泛化能力。最后, 我们在合成生成的**文本图像**和真实印刷书籍的扫描**图像**上展示了出色的**文本识别**精度, 没有标记的训练示例。少

2018 年 9 月 23 日提交;最初宣布 2018 年 9 月。

10. 适用于陷阱图像中单个和触摸飞蛾计数的支持向量机 (svm) 识别方法

作者 : [mohamed chafik bakkay](#), [silvie chambon](#), [hatem a.rashwan](#), [christian lubat](#), [sébastien barsotti](#)

文摘: 本文旨在开发一种在实际条件下从陷阱**图像**中识别飞蛾的自动**算法**。此方法使用我们以前的工作进行检测 [1], 并介绍了一个合适的分类步骤。更准确地说, Curviness 分类器是用多尺度描述符——曲率直方图 (hcs) 进行训练的。该描述符对光照变化具有鲁棒性, 能够在多尺度上检测和描述**目标**昆虫的外部 and 内部轮廓。所提出的分类方法可以用一小套**图像**进行训练。定量评价表明,

该方法能够比最先进的方法对昆虫进行更高的分类，准确率 (9.5.8%)。少

2018 年 9 月 18 日提交;最初宣布 2018 年 9 月。

11. 开放式一组对抗示例

作者:郑哲东,郑亮,胡志兰,杨毅

文摘: 近期 最近作品中的对抗例子针对的是封闭的集合**识别**系统, 其中的培训和测试课程是相同的。但是, 在实际场景中, 测试类可能与培训类 (称为开放集**识别**) 重叠的情况有限 (如果有)。据我们所知, 社区并没有**针对**这一实际环境的对抗实例的具体设计。可以说, 新的设置在两个方面损害了传统的封闭集攻击方法。首先, 闭集攻击方法也是基于分类和**分类**目标, 但开放集问题提出了不同的任务, 即 \ \ 强调。闭集**识别**的生成机制与开放式**识别**的目的不同是不可取的。其次, 由于查询**图像**通常是一个看不见的类, 从培训课程中预测其类别是不合理的, 从而导致下的对抗梯度。在这项工作中, 我们将开放集**识别**视为一种检索任务, 并提出了一种新的方法, 即对立面定向特征攻击 (odfa), 以生成对抗示例/查询。当使用被攻击的示例作为查询时, 我们的目标是尽可能低排名真正的匹配项。在解决闭集攻击方法的两个局限性时, odfa 直接处理用于检索的特征。这样做的目的是将对抗式查询的特征从原来特征的相反方向推开。简单地说, odfa 导致 recall @

k 和 map 的下降幅度大于两个开放集识别数据集 (\-指明 i. e.,} markart-1501 和 cub-200-2011) 上的闭集攻击方法。我们还证明了 odfa 的攻击性能并不明显优于闭合集识别(cifar-10)下的最先进的方法, 这表明它对开放集问题的特异性。少

2018 年 9 月 7 日提交;最初宣布 2018 年 9 月。

12. 图像和视频中的目标非线性对抗性扰动

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扰动学习方法。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。

13. midef: 基于多模型的神经网络对抗防御实例

作者:siwakorn srisakaokul, zexuan ch 中生, y 声 zhang, wei yang, tao zie

摘要: 尽管神经网络模型被广泛用于图像识别和分类等许多应用领域, 但它被发现容易受到对抗示例的影响: 给出了一个模型和一个正确分类的示例: 该模型是一个新的例子, 通过在给定的例子上应用小摄动 (人类无法察觉), 使模型错误地分类新的例子。在实际应用中, 对抗示例可能会给安全或安保带来潜在风险。近年来, 在采用脆弱模型的情况下, 防御方法, 如对抗性训练和防御性蒸馏, 改进了模型, 使其更强大, 以抵御对抗的例子。但是, 基于改进的模型, 攻击者仍然可以生成对抗示例来成功攻击模型。为了解决这种限制, 我们根据多样性的设计原则, 提出了一种新的防御方法, 名为 muldef。给定目标模型 (作为种子模型) 和要防御的攻击方法, muldef 与种子模型一起构造其他模型 (来自种子模型), 以形成一系列模型, 从而使模型相互补充。实现鲁棒性多样性 (即一个模型的对抗性例子通常不会成为其他模型的对抗示例), 同时对正常示例保持大致相同的准确性。在运行时, 给定一个输入示例, muldef 将就要应用于给定示例的族中随机选择一个模型。模型家族的鲁棒性多样性和从家族中随机选择模型共同降低了攻击的成功率。我们的评估结果表明, 在白盒和黑盒攻击场景中, muldef 在对抗示例上的精度分别提高了 35–50% 和 2–10%。

2018 年 8 月 31 日提交; 最初宣布 2018 年 9 月。

14. 食品计算综述

作者:关伟清,姜树强,刘林虎,荣瑞,拉梅什·杰恩

摘要: 食物对人类生活非常重要, 对人类的体验也是必不可少的。与食品相关的研究可以支持多种应用和服务, 如引导人类行为、改善人类健康和了解烹饪文化。随着社交网络、移动网络和物联网 (iot) 的快速发展, 人们通常会上传、共享和录制食品图像、食谱、烹饪视频和食品日记, 从而产生大规模的食品数据。大规模的粮食数据提供了丰富的食品知识, 可以帮助解决人类社会的许多核心问题。因此, 现在是将与食品计算相关的几个不同问题分组的时候了。食品计算从不同来源获取和分析异质食品数据, 用于食品的感知、**识别**、检索、推荐和监控。在食品计算中, 采用计算方法来解决医学、生物学、美食和农学领域与食品有关的问题。大规模的食品数据和最近在计算机科学方面的突破正在改变我们分析食品数据的方式。因此, 在食品领域开展了大量工作, **针对**不同的面向粮食的任务和应用。然而, 很少有系统的审查, 这些审查很好地塑造了这一领域, 并全面和深入地总结了这一领域目前的努力或详细的未决问题。在本文中, 我们将食品计算正式化, 并对各种新出现的概念、方法和任务进行了全面的概述。我们总结了食品计算的主要挑战和未来方向。这是首次**针对**食品领域计算技术研究的综合调查, 还提供了一系列研究和技术, 使从事不同食品相关工作的研究人员和从业人员受益领域。少

2018 年 9 月 9 日提交;v1 于 2018 年 8 月 21 日提交;最初宣布 2018 年 8 月。

15. 复杂 d 体系结构中高性能深度学习卷积的解剖

作者 :[evangelos georganas](#), [sasikanth avancha](#), [kunal banerjee](#), [dhiraj kalamkar](#), [greg henry](#), [hans pabst](#), [亚历山大 heanecke](#)

摘要: 卷积层在许多类型的深层神经网络中都很普遍, 包括卷积神经网络 (cnn), 它为**图像识别**、神经机器翻译和语音等任务提供了最先进的结果**认可**。卷积操作的计算开销导致了实现的激增, 包括矩阵乘法公式和主要针对 gpu 的直接卷积。本文介绍了 x86 体系结构的直接卷积内核, 特别是 x86 和 xeonphi 系统, 这些内核是通过动态编译方法实现的。我们基于 jit 的实现显示了接近理论峰值性能的性能, 具体取决于当前的设置和 cpu 体系结构。我们还演示了如何将这些 jit 优化的内核集成到轻量级的多节点图形执行模型中。这说明, 在 cpu 上执行最先进的**图像识别**任务时, 单节点和多节点运行可产生高效率和**高图像吞吐量**。
少

2018 年 8 月 20 日提交;v1 于 2018 年 8 月 16 日提交;最初宣布 2018 年 8 月。

16. 利用 elo 评级系统测量合成孔径雷达图像中的人体评估复杂性

作者:brian reinhardt, isaac gerg, daniel brown, joonho park

文摘: 合成孔径雷达数据的目标**自动识别**性能在很大程度上取决于波束形成图像的复杂性。有几种机制可以促成这一点, 包括不需要的车辆动态、现场的测深以及自然和人造杂物的存在。为了了解环境复杂性对**图像**感知的影响, 研究人员采取了植根于信息论或启发式的方法。尽管作出了这些努力, 但对复杂性的量化衡量与由此产生的现象学无关。通过使用主题专家 (sme), 我们为**一组**解释了基本现象学的图像推导出一个复杂性度量。本工作的目的是了解合成孔径雷达图像中几种常见的信息理论和启发式措施与中小企业感知复杂性的关系。为了实现这一目标, 从跨越多个环境的高频 sas 数据集中裁剪了一个 10 米 x 10 米的**图像**。中小企业提出了对**图像**的对, 从中可以对相对**图像**的复杂性进行评价。然后, 这些比较被转换为所需的顺序排名使用的方法首先开发的 a. elo 建立排名的棋手。elo 方法在整个数据集中生成了一个合理的排名排序。然后将启发式和信息理论指标与从中得出的**图像**等级进行比较。相关性最高的指标是与空间信息有关的指标, 例如像素强度的变化, r 平方值约为 0.9。然而, 这一协议取决于测量空间变化的尺度。还将为许多其他措施提供结果, 包括稀缺性、**图像**压缩和熵。少

2018 年 8 月 15 日提交;最初宣布 2018 年 8 月。

17. 利用层次语义嵌入的细粒表示学习与识别

作者:陈天水,吴文喜,高月芳,黎东,罗晓南,梁林

摘要: 对象类别本质上形成了具有不同级别的概念抽象的层次结构,特别是对于细粒度的类别。例如,鸟类(aves)可以根据等级、家族、属和物种的四级等级进行分类。该层次结构对不同层次的不同类别之间的丰富相关性进行编码,可以有效地规范语义空间,从而减少预测的模糊性。然而,以往对细粒度**图像识别**的研究主要集中在某一层次的类别上,通常忽略了这一相关信息。在本文中,我们通过开发一种新的层次语义嵌入(hse)框架,同时预测层次结构中不同层次的类别,并将这些结构化相关信息集成到深层神经网络中。具体来说,hse 框架按顺序预测层次结构中每个级别的类别分数向量,从最高到最低。在每个级别,它将更高级别的预测分数向量作为学习更细粒度的特征表示的先验知识。在训练过程中,还采用较高水平的预测分数向量,将其作为相应子类别的**软目标**,对标签预测进行规范。为了评估拟议的框架,我们组织了具有四级类别层次结构的 200 种 caltech-ucsd 鸟类物种,并构建了一个大型蝴蝶数据集,该数据集还涵盖四个级别的类别。对这两个数据集和新发布的 vegrofre 数据集进行的大量实验证明了我们的 hse 框架优于基线方法和现有的竞争对手。少

2018 年 8 月 13 日提交;最初宣布 2018 年 8 月。

18. 雅典娜 wfi 科学产品模块

作者: david n. burrows, steven allen, marshall bautz, esra bulbul, julia erdley, abraham d. falcone, stanislav fort, catherine e. grant, sven herrmann, jamie kennea, robert klar, ralph kraft, adam mantz, eric d. miller, paul nulsén, steve persyn, pragati pradhan, dan wilkins

摘要: 科学产品模块 (spm) 是美国对雅典娜广域成像仪的贡献, 是一个能力很强的二级 cpu, 对科学数据流执行特殊处理。spm 将可以访问已接受的 x 射线事件和被机载事件识别处理拒绝的事件。它将包括两个软件模块。瞬态分析模块将对科学图像进行机载处理, 以识别和描述主要目标的可变性, 并在视场中检测偶然的瞬态 x 射线源。背景分析模块将对潜在的背景事件进行更复杂的标记, 并改进背景特征, 利用未被遥测到地面的数据, 提供改进的背景图和光谱。我们给出了 spm 硬件的初步设计, 并对正在开发的软件算法进行了简要概述。少

2018 年 8 月 8 日提交;最初宣布 2018 年 8 月。

19. 浅谈小型训练装置的学习视听情感识别

作者: valentin vielzeuf, corentin kervadec, stéphane pateux, 亚历克西斯·莱切维, Stéphane jurie

文摘: 本文提出了一种用于视听情感识别的轻量级、精确的深层神经模型。为了设计这个模型, 作者遵循了简单的理念, 极大地限制了从目标数据集中学习的参数数量, 总是选择最简单的收入方法: i) 转移学习和低维空间嵌入允许减少表示的维度。ii) 等时间信息

由一个简单的每个帧的分数选择过程处理, 在不同时间内平均。iii) 还提出了一种简单的帧选择机制来对序列的**图像**进行加权。(四) 不同模式的融合是在预测水平(后期错觉)进行的。我们还强调了 afew 数据集固有的挑战和模型选择的难度, 仅有 383 个验证均衡。提出的实时情感分类器在 afew 的测试装置上达到了 606.64% 的最先进精度, 在 2018 年的《野生挑战》中排名第 4 位。少

2018 年 8 月 8 日提交;最初宣布 2018 年 8 月。

20. 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 perez-serrano, manuel ujaldón

摘要: 深度学习 (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程序中取得成功的一个具有代表性的例子, 特别是在数据集上, 在这些数据集中, 目标可以通过本地特征的层次结构来表示.增加语义复杂性。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 resnetim (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。

21. 用于在线多目标跟踪的时空 ksvd 词典学习

作者: [huynh manh](#), [gita alaghband](#)

文摘: 本文提出了一种新的空间判别 ksvd 字典算法 (stksvd), 用于在线多目标跟踪中的目标外观学习. 与其他分类/识别任务 (如人脸、图像识别) 不同, 学习目标在在线多目标跟踪中的出现受到诸如图像识别/发音变化、背景场景或其他目标的部分遮挡、背景变化 (人体探测边界框涵盖人体部分和场景的一部分) 等。然而, 我们观察到, 这些变化发生的空间和时间动态逐渐发生。我们通过一种新的 stksvd 外观学习算法来描述目标样本之间的时空信息, 以更好地区分稀疏代码、线性分类器参数, 并最大限度地

减少单个帧的重建误差优化系统。我们的外观学习算法和跟踪框架在两阶段关联的每个阶段采用两种不同的计算外观相似性分数的方法：第一阶段的线性分类器和第二阶段的最小残差。使用2dmot2015数据集及其公共聚合通道特征（acf）进行所有比较的测试结果表明，我们的方法优于现有的相关学习方法。少

2018年7月5日提交;最初宣布 2018年7月。

22. 通过在线领域适应在野外进行调整

作者: [Massimiliano Mancini](#), [hakan karaoguz](#), [elisa ricci](#), [p. Jensfelt](#), [barbara caputo](#)

摘要: 技术发展要求提高机器人的感知和行动能力。除其他技能外, 还需要能够适应工作条件任何可能变化的视觉系统。由于这些条件是不可预测的, 我们需要能够评估视觉识别算法的泛化和鲁棒性的基准。在这项工作中, 我们关注的是在不受约束的情况下的机器人装配。作为第一个贡献, 我们为装配任务提供了一个新的视觉数据集。与标准对象识别数据集不同, 我们提供在相机、照明和背景更改的各种条件下获取的相同对象的图像。这种新颖的数据集允许测试机器人视觉识别算法对一系列不同领域移动的鲁棒性, 无论是隔离还是统一。我们的第二个贡献是基于批控归一化层的一种新的深度模型在线适应算法, 它允许模型不断适应当前的工作条件。与标准的域适应算法不同, 它在训练时不需要目标域的任何图像。我们对该算法在所建议的数据集上的性能进行

了基准测试, 显示了该算法在标准体系结构的性能与离线适应给定目标域的对应体系结构之间的差距的能力。少

2018 年 7 月 3 日提交;最初宣布 2018 年 7 月。

23. 基于 mcs 2018 对白盒脸识别系统对抗攻击的动量多元输入快速梯度符号法 (m-d2-fgsm) 攻击方法的评价

作者:[md ashraful alam milton](#)

摘要: 卷积神经网络是近年来在分类、分割和检测等各种计算机视觉任务上成功的关键工具。卷积神经网络在这些任务中实现了最先进的性能, 每天都在推动计算机视觉和人工智能的极限。然而, 对计算机视觉系统的对抗攻击正在威胁其在现实生活和安全关键应用中的应用。在必要的情况下, 找到敌对性的例子对于发现容易受到攻击的模型和采取保障措施以克服对抗攻击是很重要的。在这方面, mcs 2018 对白装人脸识别挑战的攻击旨在促进寻找新的对抗性攻击技术及其生成对抗实例的有效性的研究。在这一挑战中, 攻击的性质是对黑盒神经网络的**有针对性的**攻击, 我们对黑块的内部结构一无所知。攻击者必须修改一组由一个人的**五个图像**组成的图像, 以便神经网络错误地将它们归类为**目标图像**, 而目标图像是另一个人的一组**五个图像**。在本次比赛中, 我们采用动量等输入迭代快速梯度签名方法 (m-d2-fgsm) 对黑匣子人脸识别系统进行了对抗攻击。我们在 mcs 2018 对黑盒面部识

别挑战的反攻击中测试了我们的方法, 并发现了竞争结果。我们的解决方案得到了 1.404 的验证分数, 比基线分数 1.407 要好, 在领导板 132 支球队中排名第 14 位。通过从源**图像**中找到改进的特征提取、精心选择的超参数、找到改进的黑匣子替代模型和更好的优化方法, 可以进一步改进。少

2018 年 6 月 23 日提交;最初宣布 2018 年 6 月。

24. 零射击学习的全局语义一致性

作者:范武,启田,关继宏,周水庚

摘要: 在**图像识别**中, 在许多情况下, 训练样本不能覆盖所有**目标类**。零拍摄学习 (zsl) 利用类语义信息对训练集中没有相应样本的看不见的类别的样本进行分类。在本文中, 我们提出了一个端到端框架, 称为全球语义一致性网络 (简称 gsc-net), 它完全利用有已看到类和看不见类的语义信息, 以支持有效的零镜头学习。我们还采用了软标签嵌入丢失, 以进一步利用类之间的语义关系。为了使 gsc-net 适应更实用的环境, 通用零镜头学习 (gzsl) 引入了一种参数化的新颖性检测机制。我们的方法通过三个可视化属性数据集在 zsl 和 gzsl 任务上实现了最先进的性能, 从而验证了所建议框架的有效性和优势。少

2018 年 6 月 22 日提交;最初宣布 2018 年 6 月。

25. 资源节约型神经架构师

作者:周艳琪, siavash ebrahimi, sercan k.ark, haonan yu, hairongliu, greg diamos

摘要: 神经体系结构搜索 (nas) 是一个费力的过程。以前在自动化 nas 目标方面的工作主要是提高准确性, 但缺乏对计算资源使用的考虑。我们提出了资源节约型神经架构师 (rena), 这是一种利用网络嵌入强化学习的高效资源约束的 nas。rena 使用策略网络来处理网络嵌入以生成新的配置。我们演示了 rena 在图像识别和关键字识别 (kws) 问题上的问题。rena 可以找到即使在资源紧张的情况下也能实现高性能的新型架构。对于 cifar10, 当计算强度大于 100 flops 字节时, 它实现了 2.95 的测试误差, 当模型大小小于 3m 参数时, 测试误差达到 3.87。对于 google 语音命令数据集, rena 在没有资源限制的情况下实现了最先进的精度, 并且在有限的资源约束下优于优化的体系结构。少

2018 年 6 月 12 日提交;最初宣布 2018 年 6 月。

26. 视觉问题回答的交叉数据集适应

作者:赵伟伦,胡和祥, 费沙

摘要: 我们研究了用于视觉问题回答 (可视化 qa) 的交叉数据集适应问题。我们的目标是在源数据集上训练可视 qa 模型, 但将其应用于另一个目标数据集。类似于用于可视识别的域适应, 当

目标数据集没有足够数量的标记数据来学习 "域内" 模型时, 此设置很有吸引力。关键的挑战是, 这两个数据集的构造不同, 导致图像、问题或答案上的交叉数据集不匹配。提出了一种新的域适应算法, 克服了这一困难。我们的方法通过转换目标数据集中数据的特征表示来减少统计分布的差异。此外, 它最大限度地提高了使用在源数据集上训练的可视 qa 模型正确回答问题 (在目标数据集中) 的可能性。我们对该方法在几种流行的可视 qa 数据集中的适应效果进行了实证研究。我们表明, 在没有适应和其他几种适应方法的情况下, 所提出的方法改进了基线。我们对适应的影响主要是有效的进行了定量和定性的分析。少

2018 年 6 月 10 日提交;最初宣布 2018 年 6 月。

27. 开放式领域适应的学习表象

作者 :mahsa baktashmotlagh, masoud faraki, tom drummond, mathieu salzmann

文摘: 近年来, 视觉识别领域适应取得了很大进展。然而, 大多数现有方法在所谓的闭集方案中工作, 假设目标图像所描述的类与源域所描述的类完全相同。在本文中, 我们讨论了更具挑战性但更现实的开放式域适应案例, 其中目标数据中可能存在新的未知类。虽然在无人监督的情况下, 我们不能期望能够识别每个特定的新类, 但我们的目标是自动检测哪些样本属于这些新类, 并将它们从识别过程中丢弃。为此, 我们依赖于一种直觉, 即描述已知

类的源和**目标**样本可以由共享子空间生成，而来自未知类的**目标**样本来自不同的私有子空间。因此，我们引入了一个框架，将数据分解为共享部分和私有部分，同时鼓励共享表示是歧视性的。我们在标准基准上的实验证明，我们的方法明显优于开放式域适应领域的最先进方法。少

2018 年 5 月 30 日提交;最初宣布 2018 年 5 月。

28. 基于层次贝叶斯模型的上下文开发

作者:[christopher a. george](#), [pranab banerjee](#) , [kendra e. moore](#)

文摘: 我们考虑了如何改进自动目标识别的问题，将天真的传感器级分类决策与 "直觉" 或上下文融合在一起，以一种数学上有原则的方式。这是一种与上下文的许多定义兼容的通用方法，但为了具体而将上下文视为图像中的共存。特别是，我们考虑包含在不同置信度级别上标识的多个对象的**图像**。我们学习每个上下文中的共现模式，然后将这些模式用作分层贝叶斯模型的超参数。其结果是，通过这些读数与上下文融合，可以显著改进低置信度传感器分类决策。我们进一步使用超级优先来解决多个上下文可能合适的情况。我们还考虑贝叶斯网络，这是层次贝叶斯模型的替代方案，它在计算上更有效，但假定上下文和传感器读数是不相关的。少

2018 年 5 月 30 日提交;最初宣布 2018 年 5 月。

29. 使用级联超分辨率和身份原点的面部幻觉

作者 : [klemen grm](#), [simon dobrišek](#), [walter j. cheeirer](#),
[vitomirš truc](#)

文摘: 本文讨论了在高放大倍率下, 来自不对齐低分辨率输入的高分辨率面部**图像**的幻觉问题。我们探讨卷积神经网络 (cnn) 的问题, 并提出了一个新的 (深) 面对幻觉模型, 将身份优先纳入学习过程。该模型由两个主要部分组成: (一) 级联超分辨率网络, 用于提升低分辨率**图像**; (ii) 一组人脸识别模型, 作为超分辨率网络的身份**优先级**。培训。与通常依赖于单个模型进行升级 (即使是放大倍率较大) 的竞争超分辨率方法不同, 我们的网络使用多个 sr 模型的级联, 这些模型会逐步升级低分辨率**图像**使用步骤 $2\times$ 。这一特性使我们能够以不同的分辨率应用监控信号 (目标外观), 并在多尺度上纳入标识约束。我们的模型能够扩展 (非常) 低分辨率**图像**, 在无约束条件下捕获, 并产生视觉上令人信服的结果。我们在大型面部**图像**数据集上严格评估了建议的模型, 并报告了与最先进的模型相比的卓越性能。少

2018 年 5 月 28 日提交;最初宣布 2018 年 5 月。

30. 利用句法在自然图像中的地面引用表达式

作者 : [volkan cirik](#), [taylor berg-kirkpatrick](#), [louis-philippe momand](#)

文摘: 我们引入了接地网, 这是一种用于引用表达式识别的神经网络——在自然语言表达式所指的对象的图像中定位 (或接地) 的任务。我们对这一任务的方法首先是依靠输入引用表达式的句法分析来为计算图的结构提供信息。给定一个输入表达式的解析树, 我们显式地将树中存在的句法成分和关系映射到一个由神经模块组成的组合图, 该图定义了我们执行本地化的体系结构。这种基于同步的方法有助于将表达式中提到的**目标**对象和辅助支持对象本地化。因此, 粗暴网比以前的方法更易于解释: 我们可以 (1) 确定引用表达式的哪个短语指向**图像**中的哪个对象, (2) 跟踪**目标**对象的本地化是如何由网络。我们通过在 google ref 数据集上引入一组新的注释来评估支持对象的本地化, 从而根据经验研究此属性。我们的实验表明, 地面网络在识别支撑物体方面达到了最先进的精度, 同时在**目标**对象的定位方面保持了可比的性能。

少

2018 年 5 月 26 日提交;最初宣布 2018 年 5 月。

31. 使用正则化超位图匹配的无监督域适应

作者: [debasmit das](#), [c. s. george lee](#)

摘要: 域适应 (da) 解决了训练 (源) 和测试 (**目标**) 数据分布之间差异的实际**图像**分类问题。我们提出了一种不受监督的 da 方法, 该方法只考虑目标域中是否存在无标签数据。我们的方法主

要是查找源域和目标域的样本之间的匹配项。通过将源域和目标域视为超图, 并使用图之间的一阶、二阶和三阶相似性进行类正则化的超图匹配, 得到了匹配。我们还开发了一种计算效率高的算法, 首先选择样本的子集来构造图形, 然后开发基于条件梯度和交替方向的图形匹配自定义优化例程乘法器方法。这使得所提出的方法得到了广泛的应用。我们还对标准对象识别数据集进行了一组实验, 以验证我们的框架相对于最先进方法的有效性。少

2018 年 5 月 22 日提交;最初宣布 2018 年 5 月。

32. 基于基于层的两流关系网络, 用于视频中的操作识别

作者: [王伟](#), [张金进](#), [四晨阳](#), [王亮](#)

文摘: 近年来, 与传统的基于外观的方法相比, 基于位置的动作识别性能更好, 越来越受到人们的关注。但是, 仍有两个问题需要进一步解决。首先, 现有的基于位置的方法通常识别人类的行为, 捕获的三维人类姿势在实际场景中很难获得。其次, 很少有基于位置的方法在识别人与对象交互行为时对与动作相关的对象进行建模, 其中对象在其中发挥着重要作用。为了解决上述问题, 我们提出了一个基于组的双流关系网络 (psrn) 进行动作识别。在 psrn 中, 一个流模拟直接从原始视频中提取的目标 2d 人体姿势序列的时间动态, 另一个流从随机采样的视频帧中模拟与动作相关的对象。最重要的是, 我们不像以前那样在类乐谱中融合两个

流, 而是提出了一个对象前关系网络来模拟人类姿势和与操作相关的对象之间的关系。我们根据两个具有挑战性的基准 (即 jhmdb 和 pennaction) 对拟议的 psrn 进行评估。实验结果表明, 我们的 psrn 在次 jhmdb (80.2) 和 penaction (981%) 上获得了最先进的性能。我们的工作通过结合从原始视频和**图像**外观中提取的 2d 人的姿势,为动作识别打开了新的大门。少

2018 年 5 月 22 日提交;最初宣布 2018 年 5 月。

33. 基于移动平台的手语识别迁移学习的优化

作者:[dhruv rathi](#)

文摘: 本研究的目的是对美国手语(asl) 的**识别**系统进行实验、迭代和推荐。这是一个具有挑战性的问题, 也是一个有趣的问题, 如果解决, 将带来社会和技术方面的飞跃。本文提出了一种基于移动平台的 asl 实时识别器, 使其具有更多的可访问性, 并提供了易用性。实现的技术是将 asl 中字母的手手势的新数据的转移学习基于各种预先训练的高端模型, 并优化在移动平台上运行的最佳模型, 同时考虑到在以下过程中相同的各种限制优化。所使用的数据包括 27, 455 张图像, 其中包括 24 个 asl 字母。优化后的模型在内存高效的移动应用程序上运行时, 提供了 9003% 的**准确识别**精度, 平均**识别**时间为 2.42 秒。与以往的研究相比, 这种方法确保了在**准确性**和**识别**时间上的相当大的区分。少

2018 年 5 月 17 日提交;最初宣布 2018 年 5 月。

34. 视觉识别中的深度注意力结构表征学习

作者:[krishna kanth nakka](#), [mathieu salzmann](#)

摘要: 结构化表示, 如 "单词袋"、"vlad" 和 "费舍尔矢量", 已被证明非常有效地处理复杂的视觉识别任务。因此, 它们最近被纳入了深层架构。然而, 所产生的深层结构化表示学习策略虽然有效, 但通常会从整个**图像**中聚合本地特征, 而忽略了这样一个事实, 即在复杂的**识别**任务中, 某些区域提供了更多比其他人更有鉴别力的信息。在本文中, 我们引入了一个注意结构化表示学习框架, 它将特定于**图像**的注意机制集成到聚合过程中。我们的框架学习以端到端的方式共同预测**图像**类标签和注意图, 而无需目标标签以外的任何其他监督。正如我们的实验所证明的那样, 这始终优于不那么注重的结构化表示学习, 并在标准场景**识别**和细粒度分类基准方面产生最先进的结果。少

2018 年 5 月 14 日提交;最初宣布 2018 年 5 月。

35. 胡福: 基于神经网络的硬件和软件协同攻击框架

作者:[李文硕](#), [于金成](#), [宁雪飞](#), [王鹏军](#), [齐伟](#), 王玉华, 杨华忠

文摘: 近年来, 深度学习 (dl), 特别是卷积神经网络 (cnn) 发展迅速, 应用于**图像**分类、人脸识别、**图像**分割和人的检测等许多

任务。基于 dl 的型号由于其卓越的性能，在许多领域具有广泛的应用，其中一些领域对安全至关重要，例如智能监控和自动驾驶。由于云计算的延迟和隐私问题，嵌入式加速器在这些安全关键领域很受欢迎。但是，嵌入式 dl 系统的鲁棒性可能会因为将硬件软件 trojans 插入加速器和神经网络模型而受到损害，因为加速器和部署工具（或神经网络模型）通常由第三方提供公司。幸运的是，插入硬件特洛伊木马只能实现不灵活的攻击，这意味着硬件特洛伊木马程序可以很容易地打破整个系统或交换两个输出，但不能使美国有线电视新闻网识别未知图片为目标。尽管插入软件特洛伊木马程序具有更多的攻击自由，但它通常需要篡改输入图像，这对攻击者来说并不容易。因此，在本文中，我们提出了一个硬件-软件协作攻击框架，以注入隐藏的神经网络特洛伊木马程序，它作为后门，无需操纵输入图像，并且对不同的场景是灵活的。我们测试了图像分类和人脸识别任务的攻击框架，并分别在 cifar10 和 youtube 人脸上获得了 92.6 和 100% 的攻击成功率，同时保持了与中的无攻击模型几乎相同的准确性。正常模式。此外，我们还展示了一个特定的攻击场景，在这种情况下，人脸识别系统受到攻击，并给出了一个具体的错误答案。少

2018 年 5 月 14 日提交;最初宣布 2018 年 5 月。

36. 利用分层生成对抗性抗网络开发视频识别图像

作者:余飞武,吴新晓,孙玉超,段立新

文摘: 现有的**视频识别**深度学习方法通常需要大量的标记视频进行训练。但对于一个新的任务, 视频往往是未标记的, 对它们进行注释也是耗时和耗费时间的。我们试图利用现有的完全标记的图像来帮助识别这些视频, 而不是人为的**注释**。但是, 由于域移动和异构特征表示的问题, 在图像上训练的分类器的性能可能会显著降低, 以**执行视频识别**任务。本文提出了一种新的方法, 称为分层生成性抗性网络 (higan), 通过从**图像**(即源域) 传输知识来增强视频 (即**目标域**) 的**识别能力**。higan 模型由一个 \ 强调 {低级别} 条件 gan 和一个 \ 强调 {一个高级} 条件 gan 组成。通过利用这些两级对抗学习, 我们的方法能够学习源**图像**和**目标**视频的域不变特征表示。对两个具有挑战性的**视频识别**数据集 (即 ucf101 和 hmdb51) 进行的综合实验表明, 与现有的最先进的领域适应方法相比, 该方法是有效的。少

2018 年 5 月 11 日提交;最初宣布 2018 年 5 月。

37. 监视人脸识别挑战

作者:[程志义](#),[朱夏天](#), [龚少刚](#)

文摘 人脸识别(fr) 是计算机视觉中研究最广泛的问题之一。由于最近引入了更大规模的 fr 挑战, 特别是社交媒体**网络图像**有限, 例如专业摄影记者拍摄的名人面孔高分辨率照片, 在 fr 方面取得了重大进展.然而, 在不受约束和低分辨率的**监控图像**中, 更具

挑战性的 fr 在很大程度上仍然没有得到充分研究。为了促进更多的研究开发 fr 模型,有效和稳健的低分辨率监控面部**图像**,我们引入了一个新的监控人脸**识别**挑战,我们称之为 qmul-survface 基准。这个新的基准是最大的,更重要的是唯一真正的监控 fr 基准,以我们最好的知识,低分辨率的**图像**不是通过人工向下采样的本地高分辨率**图像**合成。这一挑战包含 443 507 张在广阔的空间和时间内,在现实世界中不合作的监视场景中捕获的 15 573 个不同身份的**图像**。因此,它提出了一个极具挑战性的 fr 基准。与现有基准相比,我们使用五个具有代表性的深度学习人脸**识别**模型来衡量 fr 在这一挑战上的性能。我们表明,在实际的法医场景中,解决调查不足的监控 fr 问题,目前的艺术状况还远远不能令人满意。在开放的环境中,人脸**识别**通常比较困难,这在监控场景中很常见,因为大量非**目标**人(干扰物)出现了开放间隔的场景。这显然是为了在新的监控 fr 挑战中,性能最好的中心在 megface 基准上深度学习 fr 模型现在只能以 10% 的虚警率实现 13.2% 的成功率(在 rank-20)。少

2018年8月29日提交;v1于2018年4月25日提交;**最初宣布** 2018年4月。

38. 排序相对主成分属性网络模型 (rel-pcanet) 在全包发展指数估计中的应用

作者:[anwar irmatov](#), [elnura lmatov](#)

摘要: 2018 年, 在达沃斯世界经济论坛上, 提出了一个新的国家经济业绩指标, 名为包容性发展指数, 由 12 项指标组成。新的指标意味着, 各国可能需要实现结构性改革, 以改善经济扩张和社会包容业绩。这就是为什么 idi 计算方法必须有很强的统计和数学基础, 以便结果对公共目的是准确和透明的。在目前的工作中, 我们提出了一种新的 idi 估计方法——排序相对主成分属性网络模型 (rel-pcanet)。该模型以 relarm 和 ranknet 原理为基础, 结合了 pca 的元素、应用于**图像识别**的技术和学习对机制进行排序。此外, 我们还确定了**估计目标**概率矩阵的新方法, 以反映各国包容性发展的动态变化。实证研究证明, rel-pcanet 保证了可靠和稳健的分数和排名, 因此建议实际实施。少

2018 年 4 月 16 日提交;最初宣布 2018 年 4 月。

39. 毒蛙!神经网络上的目标清洁标签中毒攻击

作者 : [ali shafahi](#), [w. ronnyhuang](#), [mahyar najibi](#), [octavian suciu](#), [christoph studer](#), [tudor dumitras](#), [tomgoldstein](#)

摘要: 数据中毒是对机器学习模型的一种对抗攻击, 攻击者在训练集中添加示例, 以便在测试时操纵模型的行为。本文探讨了一类广泛的对神经网络的中毒攻击。拟议的攻击使用 "清洁标签"; 它们不要求攻击者对训练数据的标签有任何控制。它们也成为**目标**;它们控制分类器在特定测试实例上的行为, 而不会显著降低分类器在其他实例上的性能。例如, 攻击者可以将看似无害的**图像**

(标记正确) 添加到人脸识别引擎的训练集中, 并在测试时控制所选人员的身份。因为攻击者不需要控制标记功能, 所以只需将毒药放到联机并等待数据收集机器人对其进行刮, 就可以将其输入到训练集中。提出了一种基于优化的中毒合成方法, 并表明在使用转移学习时, 只有一个毒图可以控制分类器的行为。对于完整的端到端培训, 我们提出了一个 "水印" 策略, 使中毒可靠使用多个 (~ 50) 中毒培训实例。我们通过从 cifar 数据集生成中毒青蛙图像并使用它们来操作图像分类器来演示我们的方法。少

2018 年 4 月 2 日提交;最初宣布 2018 年 4 月。

40. 针对人工神经网络的无剪切攻击

作者:[boussad addad](#), [jerome kdjabachian](#) , [christophe meyer](#)

摘要: 在过去的几年里, 由于人工深层神经网络在计算机视觉、自然语言处理、语音识别、恶意软件检测等。然而, 它们极易受到容易制作的对抗性例子的影响。许多调查都指出了这一事实, 并提出了不同的方法来制造攻击, 同时对原始数据增加有限的扰动。到目前为止, 已知的最可靠的方法是所谓的 c & w 攻击 [1]。尽管如此, 一个被称为功能挤压加上整体防御的对策表明, 这些攻击的大部分可以被摧毁 [6]。在本文中, 我们提出了一种新的方法, 我们称之为中心初始攻击 (cia), 其优点是双重的: 首先, 它通过构造确保最大扰动小于事先固定的阈值, 而不降低的裁剪

过程攻击的质量。其次，它对最近引入的防御措施（如功能压缩、jpeg 编码，甚至针对投票组合的防御）都是稳健的。虽然它的应用并不局限于**图像**，但我们使用 imagenet 数据集上当前最好的五个分类器来说明这一点，其中两个是针对攻击的强健进行对手重新训练。在任何像素上的固定最大扰动仅为 1.5%，大约 80% 的攻击（**有针对性**）愚弄投票合奏防御，当扰动只有 6% 时，几乎 100% 的攻击。虽然这表明了抵御中情局攻击是多么困难，但本文的最后一节给出了一些限制其影响的指导方针。少

2018 年 3 月 28 日提交;v1 于 2018 年 3 月 26 日提交;**最初宣布** 2018 年 3 月。

41. tbd: 深度神经网络培训的标杆与分析

作者：朱宏宇，[mohamed akrou](#), [bojian zheng](#), [andrew pelegris](#), [amar phanishayee](#), [bianca schroeder](#), [gennady pekhimenko](#)

摘要：近年来，深部神经网络（dnn）的普及，对有效地进行 dnn 相关计算产生了很大的研究兴趣。然而，主要重点通常非常狭窄，仅限于（一）推断——即如何有效执行已经训练过的模型，（二）**图像**分类网络作为评价的主要基准。我们在这项工作中的主要目标是打破这种短视的观点，通过（i）提出一个新的 dnn 培训基准，称为 tbd（tbd 是 dnn 的培训基准的简称），使用一套具有代表性的 dnn 模型，涵盖广泛的机器学习应用：**图像**分类、机器

翻译、语音识别、目标检测、对抗网络、强化学习, 以及 (ii) 通过对培训这些不同应用进行广泛的性能分析。三种跨不同硬件配置 (单 gpu、多 gpu 和多机) 的主要深度学习框架 (tensorflow、mxnet、cntk)。tbd 目前涵盖六个主要应用领域和八个不同的最先进的模型。我们为这些模型提供了一个新的性能分析工具链, 该工具结合了**现有**性能分析工具的目标使用、仔细选择新的和现有的度量标准和方法来分析结果以及 dnn 训练的领域特定特征。我们还在所有三个主要框架中构建了一组用于内存分析的新工具; 急需的工具, 最终可以阐明 dnn 培训中不同数据结构 (权重、激活、渐变、工作区) 消耗了多少内存。通过使用我们的工具和方法, 我们就 dnn 培训的未来研究和优化应集中在哪里提出了一些重要的意见和建议。少

2018 年 4 月 13 日提交;v1 于 2018 年 3 月 16 日提交;**最初宣布** 2018 年 3 月。

42. 视频的稀疏对抗摄动

作者:魏兴兴,朱军,苏航

文摘: 尽管对深度神经网络 (dnn) 的对抗性样本进行了大量的静态图像研究, 但从未对其在视频中的扩展进行过探索。与**图像**相比, 攻击视频不仅需要考虑空间线索, 还需要考虑时间线索。此外, 为了提高隐蔽性并降低计算成本, 应在尽可能少的帧上添加扰动,

即对抗扰动在时间上稀疏。这进一步激发了扰动的传播,这意味着在当前帧上添加的扰动可以通过它们的时间相互作用传递到下一个帧。因此,这些帧不需要(或很少)额外的扰动来对它们进行错误分类。为此,我们提出了一种基于 $l_{2,1}$ 范数的优化算法来计算视频的稀疏对抗扰动。我们选择行动识别作为目标任务,并选择网络与 $\text{cnn} + \text{mn}$ 架构作为威胁模型,以验证我们的方法。由于传播,我们可以计算缩短版本视频上的扰动,然后使它们适应长版本视频来愚弄 dnn 。ucf101 数据集上的实验结果表明,即使视频中只有一个帧被扰动,则是愚弄率仍然可以达到 15.9.7%。少

2018 年 3 月 7 日提交;最初宣布 2018 年 3 月。

43. 野生的联合像素和功能级域适应

作者:luan tran, kihyuk sohn, xiu; xia 示, manmohan chandraker

摘要: 深域适应的最新发展使得知识从标记的源域转移到中间特征或输入像素级别的未标记目标域。我们建议,优势可以通过结合它们,以不同的见解的形式,导致一个新的设计和互补的属性,导致更好的性能。在特征层面,在领域对抗神经网络中的半监督学习洞察的启发下,我们以领域对抗熵最小化的形式提出了一种新的正则化方法。接下来,我们假设来自计算机视觉的洞察更适合在像素级进行注入,并具体解决了跨不同语义级别进行适应的关键挑战。特别是,我们使用基于广义外观流的 3d 几何和图像合

成来保留更高级别的姿势变换之间的标识, 同时使用属性条件的 biclegan 将单个源转换为在较低级别的属性 (如照明) 中不同的 **多个目标图像**。我们验证了一个新的问题, 汽车识别在没有标记的 **监控图像**使用标记的图像从网络上, 处理明确指定的, 通过像素级和隐式的可命名的变化因素, 通过特征级适应的未指定因素。广泛的实验获得了最先进的结果, 通过我们提出的域适应方法证明了补充特征和像素级信息的有效性。少

2018 年 2 月 28 日提交;最初宣布 2018 年 3 月。

44. 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](#)

文摘: 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, 在静止视频到视频的 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视环境中设计可靠的基于视频的 fr 系统。本章概述了通过深层卷积神经网络(cnn)在静态视频到视频的 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, 交叉相关匹配 cnn、树干分支合奏 cnn 和 haarnet)和监督自动编码器(如规范的人脸表示 cnn)的深度学习架构是在准确性和计算复杂度方面进行了审查和比较。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。

45. 资源约束下行业级人脸识别系统的无查询攻击

作者:唐迪,王晓峰,张克环

摘要: 要对基于深度神经网络(dnn)的人脸识别(fr)系统发起黑匣子攻击, 需要建立具有 \textit{一位替换} 模型来模拟**目标模型**, 因此从替换中发现了敌对示例模型也可能误导**目标模型**。在最近的研究中, 通过查询**目标模型**来获取用于训练替代模型的数据, 从而实现了这样的无纹理量。然而, 像 fr 执法系统这样的现实目标, 对手来说却不太容易接近。为了攻击这样的系统, 需要一个质量与**目标模型**相似的替代模型来识别它们的共同缺陷。这很难, 因为对手往往没有足够的资源来训练这样一个强

大的模型（需要数亿的**图像**和 gpu 的房间来训练一个商业的 fr 系统）。然而，我们在研究中发现，资源受限的对手仍然可以有效地逼近目标模型识别**具有识别具**（特定）的个体的能力，方法是在其他模型上训练具有 "表达性"（特定）的替代模型。那些受害者的图像，其身份的攻击者希望覆盖或冒充。这是由我们发现的一个新属性（称为 "结纹理 {近局部线性}" (nll)）实现的，该属性模拟了理想的 dnn 模型生成图像表示（嵌入）的**观察**，该图像表示形式之间的距离是真实的描述了人类对输入**图像**之间差异的感知。通过围绕受害者的**图像**模拟此属性，我们将黑匣子模拟攻击的可转移性提高了近 50%。特别是，我们成功地攻击了一个商业系统训练了超过 2000 万**张图像**，使用 400 万**张图像**和超过一半的训练时间，但在模拟攻击中实现了 62% 的可转移性，在躲闪攻击中实现了 89%。少

2018 年 8 月 22 日提交;v1 于 2018 年 2 月 13 日提交;**最初宣布** 2018 年 2 月。

46. 利用多任务深度学习进行三维姿态估计和动作识别

作者:[diogo c. luvizon](#), [davidpicard](#) , [hedi tabia](#)

摘要: 动作识别和人体姿态估计密切相关，但这两个问题一般都是作为文献中的不同任务来处理的。在本文的工作中，我们提出了一个多任务框架，用于从静止**图像**和从视频序列的人类动作识

别中进行联合二维和三维姿态估计。我们证明了一个单一的架构可以用来有效地解决这两个问题，仍然达到最先进的结果。此外，我们还证明，从端到端进行优化可显著提高比分离学习的准确性。可以无缝地同时使用来自不同类别的数据来训练建议的体系结构。在四个数据集（mpii、human3.6 m、penn action 和 ntu）上报告的结果证明了我们的方法在目标任务上的有效性。少

2018 年 3 月 21 日提交;v1 于 2018 年 2 月 26 日提交;最初宣布 2018 年 2 月。

47. 利用形状和运动检测恶劣天气预报中的共状云

作者:郑新业,叶建波,陈玉坤,斯蒂芬·维斯塔尔,贾丽,何塞·pibera-fernandez, michael a. steinberg, james z.wang

摘要: 气象学家使用卫星图像中的云的形状和运动作为几种主要类型的严重风暴的指标。卫星虚拟数据在空间和时间上的分辨率都越来越高,这使得人类无法充分利用其预测中的数据。需要自动卫星图像分析方法,以便在发现与风暴有关的云模式后立即找到这些模式。我们提出了一种基于机器学习和模式识别的方法来检测卫星图像中的“逗号形”云,这些云是与旋风配方密切相关的特定云分布模式。为了使用目标运动模式检测区域,我们的方法在由形状和运动敏感特征表示的手动注释云示例上进行了培训。不同尺度的滑动窗口被用来确保密集的云被捕获,我们实施有效

的选择规则来缩小这些滑动窗口之间的兴趣区域。最后，我们对一个保留带注释的逗号形云数据集的方法进行了评估，并将结果与恶劣天气数据库中记录的风暴事件进行了交叉匹配。我们的方法的有效性和准确性表明，在协助气象学家天气预报的潜力很大。少

2018 年 6 月 5 日提交;v1 于 2018 年 2 月 24 日提交;最初宣布 2018 年 2 月。

48. 判别标签一致域适应

作者:罗凌坤,陈立明, 李英路,胡世强

摘要: 域适应 (da) 是一种转移学习，其目的是在源和目标之间的数据分布不匹配的情况下，从源数据中学习目标数据的有效预测。本文提出了一种新的跨域视觉识别无监督 da 方法，该方法同时优化了理论上建立的误差约束的三个术语。具体而言，建议的 da 方法迭代搜索一个潜在的共享特征子空间，其中不仅源域和目标域之间的数据分布差异会像最先进的 da 方法一样减少，但也增加了班级间的距离，以促进歧视性学习。此外，所提出的 da 方法从共享子空间中实现的特征中稀疏地回归类标签，同时最大限度地减少源数据上的预测误差，并确保源和目标之间的标签一致性。还对数据异常值进行了核算，以进一步避免负面知识转移。全面的实验和深入分析验证了所提出的 da 方法的有效性，

该方法在标准 da 基准（即 12 个跨域图像分类任务）上的性能始终优于最先进的 da 方法。少

2018 年 2 月 21 日提交;最初宣布 2018 年 2 月。

49. 博物馆展览识别领域改编及超越的挑战

作者:piotr koniusz, yusuf tas, h 洪 guangzhang, mehrtash harandi, fatih porikli, rui zhang

摘要: 本文探讨了一个开放的艺术品识别问题, 并提出了一个新的数据集, 称为开放博物馆识别挑战 (开放博物馆识别挑战)。它包含在几个博物馆的 10 个不同展览空间拍摄的展品照片, 这些展览空间展示了绘画、钟表、雕塑、玻璃器皿、文物、科学展品、自然历史作品、陶器、陶器、工具和本土工艺品。open mic 的目标是通过为精度达到 90% 的著名办公室数据集提供测试台补充, 促进领域适应、自我中心识别和少镜头学习方面的研究。为了形成我们的数据集, 我们用手机和可穿戴相机分别捕捉到了每幅艺术品的一些图像, 形成了源数据和目标数据的分裂。为了实现可靠的基线, 我们建立在最近的一种方法的基础上, 该方法对齐源和目标 cnn 流的每个类散点矩阵 [15]。此外, 我们利用这种表示的肯定性质, 通过使用端到端布雷格曼分歧和黎曼度量。我们提出了基线, 如培训/培训评价/培训/培训评价的组合集涵盖 866 个展览身份。由于每个展览都提出了独特的挑战, 例如照明质量、

运动模糊、遮挡、杂乱、视点和音阶变化、旋转、眩光、透明度、非平面、剪裁, 我们分解这些因素的结果。少

2018 年 2 月 4 日提交;最初宣布 2018 年 2 月。

50. 从合成孔径雷达图像中深度学习端到端自动目标识别

作者:[古河秀俊](#)

文摘: 合成孔径雷达 (sar) 自动**目标识别**(atr) 的标准体系结构包括检测、判别和分类三个阶段。近年来, 人们提出了 sar atr 的卷积神经网络 (cnn), 但大多将从 sar 图像中提取的**目标芯片**中提取的**目标芯片**中的目标类分类, 作为 sar atr 第三阶段的分类。在本报告中, 我们提出了一个新的 cnn 端到端 atr 从 sar 图像。cnn 命名的验证支持网络 (versnet) 端到端执行 sar atr 的所有三个阶段。versnet 输入具有多个类和多个**目标**的任意大小的 **sar 图像**, 并输出表示每个检测到的目标的位置、类和姿势的 **sar atr 图像**. 本报告描述了 versnet 的评估结果, 该评估结果训练输出了所有 12 个类的分数: 10 个**目标类**、**1 个目标前班**和 1 个后台类, 每个像素使用移动和固定**目标公共数据集**。少

2018 年 1 月 25 日提交;最初宣布 2018 年 1 月。

51. 合成普遍对抗摄动的通用无数据目标

作者:[konda reddy mopuri](#), [aditya ganeshan](#), [r. venkatesh babu](#)

摘要: 机器学习模型容易受到对抗扰动的影响: 对输入的微小变化可能会导致输出的大变化。研究还表明, 存在着称为普遍对抗摄动的输入无关摄动, 它可以改变**大多数**数据样本的目标模型推断。但是, 现有的制作通用扰动的方法是 (i) 特定于任务的, (二) 需要来自训练数据分布的样本, 以及 (iii) 执行复杂的优化。此外, 由于数据依赖性, 制作的扰动的可骗能力与可用的训练数据成正比。在本文中, 我们提出了一个新的, 通用的和无数数据的方法来制定普遍的对抗性扰动。独立于基础任务, 我们的目标通过破坏多层提取的要素来实现愚弄。因此, 该目标可推广到在目标识别、语义分割和深度估计等多种视觉任务中形成**图像**无关扰动。在黑盒攻击场景的实际设置中 (当攻击者无法访问**目标**模型及其训练数据时), 我们表明我们的目标优于数据依赖的目标, 以愚弄学习的模型。此外, 通过利用与数据分布相关的简单前提, 我们的目标显著提高了制作的摄动的愚弄能力。通过我们的目标实现的大量愚弄率强调, 目前的深度学习模型现在面临着更大的风险, 因为我们的目标在多个任务中进行概括, 而不需要训练数据来制作扰动。为了鼓励可重复的研究, 我们发布了我们提出的算法的代码。

少

2018 年 7 月 24 日提交;v1 于 2018 年 1 月 24 日提交;**最初宣布** 2018 年 1 月。

52. rgb-d 场景识别的深度 cnn: 从零开始学习比从 RGB-D 转移要好

作者:宋新航,路易斯·赫兰茨,姜树强

摘要: 由于卷积神经网络 (cnn) 和大型场景数据集, rgb 图像的场景识别得到了广泛的研究, 并达到了非常显著的识别水平。相反, 当前的 rgb-d 场景数据要有限得多, 因此通常会利用 rgb 大型数据集, 方法是传输预训练的 rgb cnn 模型, 并使用目标 rgb-d 数据集进行微调。然而, 我们表明, 这种方法有很难达到底层的局限性, 这是学习模型特定特征的关键。相反, 我们专注于底层, 并提出了一种替代策略, 以学习局部弱监督训练与图像相结合的深度特征, 然后进行全局微调。这种策略能够在深度图像有限的情况下学习非常有鉴别力的特定于深度的特征, 而不需要诉诸地方——美国有线电视新闻网。此外, 我们还提出了一个改进的 cnn 架构, 以进一步匹配模型的复杂性和可用的数据量。对于 rgb-d 场景识别, 深度和 rgb 特征通过在公共空间中投影并进一步倾斜多层分类器进行组合, 该分类器在端到端网络中进行了联合优化。我们的框架仅在深度和联合 rgb-d 数据上实现了 nyu2 和 sun rgb-d 的最先进精度。少

2018 年 1 月 21 日提交;最初宣布 2018 年 1 月。

53. 生成传感: 转换不可靠的传感器数据以实现可靠识别

作者:[lina karam](#), [tejas borkar](#), [yu cao](#), [junseok chae](#)

文摘: 本文介绍了一种集低端传感器与计算智能于一体的深度学习生成传感框架, 实现了与高端传感器相当的高识别精度。提出的生成传感框架旨在将低端、低质量的传感器数据转化为质量更高的传感器数据, 实现分类精度。低端数据可以转化为同一模式的更高质量数据, 也可以转化为另一种方式的数据。与现有的图像生成方法不同, 该框架以判别模型和目标为基础, 最大限度地提高识别精度, 而不是相似度度量。这是通过在深度神经网络(dnn) 中引入选择性特征再生来实现的。拟议的生成传感将从本质上将低质量的传感器数据转换为高质量的信息, 以实现可靠的感知。结果说明了拟议框架的性能。少

2018 年 1 月 8 日提交;最初宣布 2018 年 1 月。

54. 你听到了吗? 自动语音识别的对抗实例

作者:[moustafa alzantot](#), [bharathan balaji](#),[mani srivastava](#)

摘要: 语音是人与人之间常见而有效的沟通方式, 智能手机和家庭集线器等现代消费设备配备了基于深度学习的精确自动语音识别功能, 可实现人与人之间的自然互动人类和机器。最近, 研究人员展示了对机器学习模型的强大攻击, 这些攻击可能会愚弄他们产生不正确的结果。然而, 以前几乎所有关于对抗性攻击的研究都集中在图像识别和物体检测模型上。在这篇简短的论文中, 我

们首次展示了针对语音分类模型的对抗性攻击。我们的算法通过添加小背景噪声来执行**有针对性的**攻击, 成功率为 87%, 而无需了解基础模型参数和体系结构。我们的攻击只改变音频剪辑样本子集中最不重要的位, 噪声不会改变 89% 的人类听众对音频剪辑的感知, 在我们的人类研究中评估。少

2018 年 1 月 2 日提交;最初宣布 2018 年 1 月。

55. 对抗性识别: 最先进的人脸识别中的神经网络攻击

作者:[mahood sharif](#), [sruti bhagavatula](#), [lujo bauer](#) , [michael k. reiter](#)

摘要: 在本文中, 我们证明了错误分类攻击基于深度神经网络 (dnn) 的面部**识别**系统比以前演示的更危险, 即使在这种情况下, 对手只能操纵她的物理(与直接操作图像输入到 dnn)。具体而言, 我们展示了如何创建眼镜, 当佩戴时, 可以成功地进行**有针对性** (模拟) 或非针对性 (躲闪) 攻击, 同时改进以前在三个方面中的一个或多个方面的工作: (i) 对观察者不显眼, 我们通过用户研究进行测试;(二) 对拟议防御的攻击的稳健性;(iii) 将眼镜创建与佩戴眼镜的主体脱钩的意义上的可扩展性, 即通过创建 "通用" 眼镜集, 从而便于错误分类。这些改进的核心是**对抗性网**, 我们建议自动生成物理可实现的攻击伪影 (这里是眼镜)。少

2017 年 12 月 31 日提交;最初宣布 2018 年 1 月。

56. 论反向传播在语音纹理生成和语音转换中的应用

作者: [jan chorowski](#), [ron j. weiss](#), [rif a. saurous](#), [samy bengio](#)

摘要: 在最近依靠反向向网络输入进行神经网络**图像**生成工作的启发下, 我们提出了一种基于近似机制的语音纹理合成和语音转换的概念验证系统: 语音**识别**神经网络学习的表示的反演, 以及不同源和**目标**话语之间神经元激活的匹配统计。与**图像**纹理合成和神经风格传输类似, 该系统通过优化输入波形样本的成本函数来工作。为此, 我们使用了一个可微的熔融滤波器库特征提取管道, 并构建了一个卷积的 ctc 语音**识别**网络。我们的系统能够从非常有限的**目标**扬声器数据中提取扬声器特征, 只需几秒钟, 就可以用来产生逼真的言语或用不同的声音重建话语。少

2018 年 3 月 8 日提交;v1 于 2017 年 12 月 22 日提交;**最初宣布** 2017 年 12 月。

57. 用于 word 识别和索引脚本检测的跨语言框架

作者 : [ayan kumar bhunia](#), [partha pratimroy](#), [akash mohta](#), [umapada pal](#)

摘要: 手写的单词**识别**和发现低资源脚本是困难的, 因为没有足够的培训数据, 收集这类脚本的数据往往费用很高。本文提出了一个新的跨语言**平台**, 用于手写单词识别和发现这种低资源脚本, 在这种平台上, 使用足够大的可用脚本数据集 (被视为源脚本)

进行培训, 并测试是在其他脚本上完成的(被视为**目标脚本**)。由于脚本之间手写可变性的复杂性, 使用一个源脚本进行培训并使用另一个脚本进行测试以获得合理的结果在手写域中并不容易。当源字符和**目标**字符出现在草书**字图像**中时, 也很难在它们之间进行映射。拟议的 indic 跨语言框架利用大量数据集资源进行培训, 并将其用于识别和发现其他**目标脚本**的文本, 因为这些脚本没有足够的培训数据。由于印度脚本主要是写在 3 个区域, 即上部、中间和下部, 我们使用区域化字符(或组件)映射来实现高效的学习目的。我们的跨语言框架的性能取决于源脚本和**目标脚本**之间的相似程度。因此, 我们设计了一个基于熵的脚本相似度评分, 利用源到目标字符映射, 这将提供跨语言转录的可行性。我们已经测试了我们的方法在三个印度脚本, 即孟加拉语, devanagari 和古鲁木希, 并报告了相应的结果。少

2018 年 1 月 28 日提交;v1 于 2017 年 12 月 19 日提交;**最初宣布** 2017 年 12 月。

58. 堆叠去噪自动编码器和转移学习用于免疫金粒子检测与识别

作者:ricardo gamelas sousa,豪尔赫 m. santos, luís m.silva, luis a. 亚历山大, tiago esteves, sara rocha, paulo monjardino, joaquim marques de sa, franciscofigueiredo, pedro quelhas

文摘: 本文提出了一种检测免疫金颗粒的系统 and 用于识别这些免疫型颗粒的**转移学习** (tl) 框架。免疫金粒子是一种高放大方法的一部分, 用于在亚细胞水平上选择性定位生物分子, 只有通过电子显微镜才能看到。细胞壁中免疫金颗粒的数量允许评估其成分差异, 为分析不同植物的质量提供了一个工具。对于它的量化, 你需要一个费尽周万尽的手动标记 (或注释) 的**图像**包含数百个粒子。本文提出的系统可以显著利用这一人工任务的负担。对于粒子检测, 我们使用与 sda 耦合的 log 滤波器。为了提高**识别**能力, 我们还研究了 tl 设置在免疫原识别中的适用性。tl 在包含不同大小粒子的其他数据集 (**目标问题**) 上重用源问题的学习模型。该系统是为解决玉米细胞的一个特殊问题而开发的, 即确定胚乳转移细胞中细胞壁生长物的组成。这个新颖的数据集以及复制我们实验的代码是公开的。我们确定, 仅 log 检测器就能达到 f 测量精度的 84% 以上。与基准模型相比, 利用 tl 开发免疫型**识别**也提供了卓越的性能, 使准确率提高了 10\%。少

2017 年 12 月 7 日提交;最初宣布 2017 年 12 月。

59. 基于计算记忆的深部神经网络的混合精度训练

作者: [nandakumar s. r.](#), [manuel le gallo](#) , [irem boybat](#), [bipin rajendran](#), [abu sebastian](#) , [evangelos Eleftheriou](#)

摘要: 深度神经网络通过在解决**图像**和语音**识别**等许多实际问题方面提供前所未有的类似人类的性能, 彻底改变了机器学习领域。

然而, 大型 dnn 的培训是一项计算密集型任务, 这就需要开发针对此应用程序的新计算体系结构。在横杆阵列中组织电阻记忆装置的计算内存单元可用于在本地存储其电导状态的突触权。昂贵的乘法累积操作可以使用基什霍夫的电路定律以非冯·诺依曼的方式在适当的地方执行。然而, 一个关键的挑战仍然是无法在重量更新过程中以可靠的方式改变设备的电导状态。我们提出了一个混合精度的架构, 结合了计算记忆单元存储突触权重与数字处理单元和额外的内存单元积累重量更新在高精度。新的体系结构提供了与浮点实现相当的分类精度, 而不受与新兴电阻存储器的非理想重量更新特性相关的挑战的限制。利用相变存储器器件的非线性随机模型实现计算记忆单元的两层神经网络, 在 mnist 手写数字分类问题上实现了 97.40% 的测试精度。少

2017 年 12 月 4 日提交;最初宣布 2017 年 12 月。

60. 深度学习缩放是可预测的, 经验上

作者:joel hestness, sharannarang , newsha ardalani, gregory dianos, heewoojun, hassan kianinejad, md. mostofa ali patware, yang yang , yangqizhou

摘要: 深度学习 (dl) 创造了有影响力的进步, 遵循了一个良性的方法: 模型体系结构搜索、创建大型培训数据集和扩展计算。人们普遍认为, 不断增长的培训集和模型应该会提高准确性, 产生更好的产品。随着 dl 应用程序域的增长, 我们希望更深入地了解

训练集大小、计算规模和模型精度改进之间的关系，以推进最先进的技术。本文提出了随着训练集的增长，泛化误差和模型尺寸增长的大规模经验表征。我们介绍了一种测量和测试四个机器学习领域的方法：机器翻译、语言建模、**图像处理**和**语音识别**。我们的经验结果表明，功率定律泛化误差跨越广泛的因素，导致功率定律指数——学习曲线的“陡峭”——尚未解释的理论工作。此外，模型改进只会改变误差，但似乎不会影响功率定律指数。我们还表明，模型大小尺度随着数据大小的提前升华。这些扩展关系对深度学习研究、实践和系统具有重要影响。它们可以帮助模型调试、设置准确性目标以及有关数据集增长的决策。它们还可以指导计算系统的设计，并强调持续计算缩放的重要性。少

2017 年 12 月 1 日提交;最初宣布 2017 年 12 月。

61. 跨域和任务的可转换表示的标签高效学习

作者:[罗泽伦](#),[邹玉良](#),[朱迪·霍夫曼](#),[李飞飞](#)

摘要: 我们提出了一个框架，学习一个代表可以转让不同的领域和任务在标签高效的方式。我们的方法与领域对抗损失作斗争，并推广使用基于度量学习的方法将嵌入到新任务中。我们的模型同时在**目标域**中的标记源数据和无标记或稀疏标记的数据上进行优化。我们的方法在新域中的新类上显示了引人注目的结果，即使每个类只有几个标记的示例可用，也会优于流行的微调方法。

此外，我们还展示了我们的学习任务框架从**图像对象识别**到**视频动作识别**的有效性。少

2017 年 11 月 30 日提交;最初宣布 2017 年 12 月。

62. 典型化对象的可扩展对象检测

作者 :aayush garg, thilo will, william darling, willi richert, clemens marschner

摘要: 继卷积神经网络和单片模型架构最近取得突破之后，最先进的目标检测模型可以可靠、准确地扩展到多达数千个类的领域。然而，当扩展到数万，或者最终扩展到数百万个或数十亿的独特对象时，情况很快就会崩溃。此外，边界框培训的端到端模型需要大量的培训数据。尽管——使用层次结构的一些技巧——有时可以扩展到数千个类，但干净**图像**注释的人工要求很快就会失控。本文提出了一种针对存在原型**图像**的品牌标识和其他程式化对象的两层对象检测方法。它可以扩展到大量的唯一类。我们的第一层是美国有线电视新闻网从单次拍摄多盒探测器系列的模型，学习提出一些程式化的对象可能会出现的区域。然后，对针对手头检索任务的图像索引运行建议的**边界框**的内容。拟议的体系结构可扩展到大量对象类，允许在无需再培训的情况下连续添加新类，并在风格化的对象检测任务（如徽标**识别**）上表现出最先进的质量。少

2017年11月29日提交;v1于2017年11月27日提交;最初宣布2017年11月。

63. 基于微机和深度学习方法的人机微多普勒分类

作者:sherif abdulatif, qian wei, fady aziz, bernhard kleiner, urs schneider

抽象: 雷达传感器可用于分析由于微运动而引起的频率变化, 这些变化的范围和速度维度都被确定为微多普勒 (M-d) 和微范围 (M-r), 分别。不同的移动目标将具有唯一的M-d 和M-r可用于目标分类的 r 签名。这种分类可用于步态识别、安全和监控等多个领域。本文将一种 25 ghz fmcw 单输入单输出 (siso) 雷达应用于工业安全中, 用于人机实时识别。由于实时约束, 对联合距离多普勒 (r-d) 映射进行了直接分析, 解决了分类问题。此外, 还对传统的经典学习方法与手工提取特征、集成分类器和深度学习方法进行了比较。对于集成分类器, 重构的范围和速度剖面直接传递给集成树, 如梯度提升和随机林, 而不进行特征提取。最后, 利用深卷积神经网络 (dcnn), 将原始 r-d 图像直接输入所构建的网络。dcnn 在一张 r-d 地图上显示出 99% 的精度在识别机器人中的卓越性能。少

2018年2月26日提交;v1于2017年11月24日提交;最初宣布2017年11月。

64. 基于大脑皮层流模型的识别系统的有效实现

作者:[ahmad w. bitar](#), [mohammad m.mansour](#), [ali chehab](#)

文摘: 本文提出了一种基于视觉皮层原始 hmax 模型的**识别系统**的有效实现方法。提出了**各种优化**, 以提高 hmax 模型的所谓 s1、c1 和 s2 层的准确性。在 s1 层, 从**图像**中消除所有不重要的信息, 如照明和表达式变化。然后, 在空间域中使用 64 个可分离的 g 局长滤波器对每个**图像**进行卷曲处理。在 c1 层, 利用最小尺度值, 使用加嵌入空间嵌入到最大比例值中。在 s2 层, 使用 medoid (pam) 聚类算法以更有效的方式生成原型。在 caltech101 数据库上评估了这些优化在准确性和计算复杂度方面的影响, 并与使用支持向量机 (svm) 和最近邻居 (nn) 分类器的基准性能进行了比较。结果表明, 在计算复杂度也降低的情况下, 我们的模型在 s1 层的精度有了显著的提高, 提高了 10% 以上。c1 和 s2 层的近似精度略有提高。少

2017 年 11 月 21 日提交;最初宣布 2017 年 11 月。

65. 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**定位**以客户的身份, 进一步推动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, 舒适的用户交互界面、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, 其目的是通过利用统计和大数据科学、机器学习和深度学习、用户建模和营销技术, 提供更智能、更方便的服务。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)。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。

66. 利用逆合成孔径雷达实现飞机目标自动识别的多雷达方法

作者:carlos pena-caballero, elifaleth cantu, jesus rodriguez, adolfo gonzáales, osvaldo castellanos, angel cantu , meegan 海峡, jae son, dengchul kim

文摘: 随着雷达技术的进步, 利用合成孔径雷达 (sar) 和逆 sar (isar) 进行自动**目标识别**(atr) 已成为一个活跃的研究领域。sar isar 是一种雷达技术, 用于生成**目标**的二维高分辨率**图像**。与使用卷积神经网络 (cnn) 来解决这个问题的其他类似实验不同, 我们使用了一种不寻常的方法, 从而提高了性能并缩短了训练时间。我们的美国有线电视新闻网使用模拟生成的复杂值来训练网络; 此外, 我们采用多雷达方法来提高培训和测试过程的准确性, 从而比其他在 sarisar 上的论文具有更高的精度。我们用我们开发的雷达模拟器生成了我们的数据集, 其中包括 7 个不同的飞机模型, 称为 radarpixel; 它是一个使用 matlab 和 java 编程实现的 windows gui 程序, 模拟器能够准确地复制真正的 sarsisar 配置。我们的目标是利用我们的多雷达技术, 确定检测和分类**目标**所需的雷达的最佳数量。少

2018 年 3 月 12 日提交;v1 于 2017 年 11 月 13 日提交;**最初宣布** 2017 年 11 月。

67. 深部神经网络的全迁移学习及其在脓毒症分类中的应用

作者 :yoshihide sawada, yoshikuni sato, toru nakada, kei ujimoto, nobuhiro hayashi

摘要: 在本文中, 我们提出了一种深度神经网络 (dnn) 的转移学习方法。深度学习在许多应用中得到了广泛的应用。然而, 在没有大量培训数据的情况下, 应用深度学习是有问题的。解决这一问题的常规方法之一是 dnn 的迁移学习。在**图像识别**领域, 除输出层外, 最先进的 dnn 重用源域数据参数的转移学习方法。但是, 当**目标域**数据量明显较小时, 此方法可能会导致分类性能较差。为了解决这个问题, 我们提出了一种称为 "全转移深度学习" 的方法, 该方法允许传输 dnn 的所有参数。利用这种方法, 我们可以通过源域知识来计算源标签和**目标标签**之间的关系。我们将我们的方法应用于实际的二维电泳**图像**~ (2-de **图像**) 分类, 以确定一个人是否患有脓毒症;首次尝试将分类方法应用于蛋白质组学的 2-de **图像**, 作为基因组学之外的延伸, 引起了相当大的关注。结果表明, 我们提出的方法优于传统的 dnn 转移学习方法。

2017 年 11 月 13 日提交;最初宣布 2017 年 11 月。

68. 不受约束的阿拉伯语脚本场景文本和视频文本识别

作者:[mohit jain](#), [minesh mathew](#), [c. v. jawahar](#)

摘要: 为阿拉伯语构建强大的识别器一直是一项挑战。我们展示了端到端可培训的 cnn-rnn 混合架构在识别视频和自然场景中的阿拉伯语文本方面的有效性。我们在两个公开的视频文本数据集 (alif 和 actiiv) 上的性能优于以前的最先进数据。对于**场景文本**

识别任务，我们引入了一个新的阿拉伯场景文本数据集，并建立了基线结果。对于阿拉伯语这样的脚本，开发强效识别器的一个主要挑战是缺少大量注释数据。我们通过合成来自大量阿拉伯语单词和短语的数百万阿拉伯文本**图像**来克服这一问题。我们的实现是建立在这里介绍的模型之上的 [37]，这被证明对英语场景**文本识别**相当有效。该模型遵循无分段序列序列转录方法。网络将一系列卷积特征从输入**图像**转录到一系列**目标**标签。这就消除了将输入**图像**分割为组成特征的需要，这对于阿拉伯语脚本来说往往是很困难的。此外，mn 对上下文依赖关系建模的能力会产生卓越的**识别**结果。少

2017 年 11 月 7 日提交;最初宣布 2017 年 11 月。

69. 剩余网络与 rnn-ctc 相结合的手写数字字符串识别

作者:[詹洪建](#),[王庆清](#),[陆岳](#)

摘要: 递归神经网络 (rnn) 和连接器时间分类 (ctc) 在许多序列标记任务中都取得了成功，具有很强的处理输入和**目标**标签对齐的问题的能力。未知。剩余网络是一种新的卷积神经网络结构，在各种计算机视觉任务中运行良好。本文利用上述体系结构，创建了一个新的手写数字字符串**识别**网络。首先，我们设计了一个剩余网络来从输入**图像**中提取特征，然后我们使用一个 mn 来建模特征序列中的上下文信息并预测**识别**结果。在这个网络的顶端，

采用了一个标准的 ctc 来计算损失并得出最终结果。这三个部分构成了一个端到端可培训网络。所提出的新架构在 orand-car-a 和 orand-car-b 上的性能最高,识别率分别为 89.75 和 91.14。此外,在字符串长度更长的生成的验证码数据集上进行的实验显示了所建议的网络处理长字符串的潜力。少

提交于 2017 年 10 月 9 日;最初宣布 2017 年 10 月。

70.更深入、更广泛、更专业的领域泛化

作者:李大新,杨永新,宋一哲,蒂莫西·m·霍斯帕莱斯

摘要: 域泛化的问题是从多个训练域中学习,并提取一个领域无关的模型,然后将其应用到看不见的域中。在存在具有明显特征但用于培训的数据稀疏的目标域的情况下,域泛化(dg)具有明显的动机。例如,素描图像中的识别,明显比照片更抽象、更罕见。然而,dg 方法主要是根据仅照片基准进行评估的,重点是缓解数据集的偏差,因为在这些基准中,领域显著性和数据稀疏的问题都可以是最小的。我们认为,这些基准过于简单,并表明简单的深度学习基线在这些基准上表现得惊人的好。本文主要贡献了两个方面:第一,在深度学习方法的有利领域移位鲁棒性的基础上,开发了一种用于端到端 dg 学习的低等级参数化 cnn 模型。其次,我们开发了一个 dg 基准数据集,涵盖了照片,素描,卡通和绘画领域。这与现有基准相比,更具实际相关性,也更大(更大

的领域转移)。结果表明, 我们的方法优于现有的 dg 替代品, 我们的数据集为推动未来的研究提供了更重要的 dg 挑战。少

提交于 2017 年 10 月 9 日;最初宣布 2017 年 10 月。

71. 实例分类问题中的多实例活动学习的包级聚合

作者:mard-andrécarbonneau, eric granger, ghyslain gagnon

摘要: 越来越多的应用, 如视频监控和医学图像分析, 需要从大量弱注释数据中建立培训识别系统, 而一些与领域专家进行有针对性的互动以改进培训过程。在这种情况下, 主动学习 (al) 可以通过查询专家以提供信息最丰富的实例的标签来降低培训分类器的标签成本。本文重点研究了在多实例学习 (mil) 中的自动分类问题, 即数据被排列成集合, 称为袋子, 这些集合被弱标记。大多数 al 方法都集中在单个实例学习问题上。这些方法不适用于 mil 问题, 因为它们不能解释数据的包结构。本文针对多实例主动学习 (mial), 提出了一种新的实例信息聚类的方法。方法基于分类器不确定性标识信息最丰富的实例, 并识别包含最多信息的查询包。另一种建议的方法称为基于 \textit{textit} 拟的聚合采样, 它在实例空间中分层地对数据进行群集。通过考虑袋子标签、推断实例标签以及在集群中仍需发现的标签比例来评估实例的信息性。在使用来自多个应用领域的基准数据进行的大量实验中, 这两种方法的性能都明显优于参考方法。结果表明, 使用适当的策略来解

决策问题会显著减少实现与单个实例方法相同的性能级别所需的查询数量。少

2017 年 10 月 6 日提交;最初宣布 2017 年 10 月。

72. 基于 mtj 的神经网络中的功率优化

作者: [ankit mondal](#) , [ankur srivastava](#)

文摘: 人工神经网络 (anns) 在模式识别和图像分类等任务中得到了广泛的应用。然而, 使用传统二进制算术单元的 ann 硬件实现是计算成本高、耗能大、面积大的开销。随机计算 (sc) 是一种新兴的范式, 它用简单的逻辑电路取代了这些传统单元, 特别适用于容错应用。自旋电子设备, 如磁性隧道连接点 (mtj), 能够在内存和逻辑电路中取代 cmos。在这项工作中, 我们提出了一个节能使用 mtj, 表现出概率切换行为, 作为随机数字发生器 (sng), 这构成了我们在 sc 域中的神经网络实现的基础。此外, nn 的误差弹性目标应用程序允许我们引入近似计算, 在这个框架中, 计算的准确性被权衡, 从而大幅降低功耗。我们建议在基于 mtj 的 nn 实现中, 以 mtj-sng 的特性带来的方式近似突触权重, 以实现能效。由于问题公式的凸性, 我们设计了一种算法, 该算法能够在单层 nn 中的给定误差公差范围内以最佳方式执行这种近似。然后, 我们使用该算法, 并开发了一种启发式方法来逼近多层 nn。为了说明该方法的有效性, 在标准分类问题上, 功耗降

低了 43%，精度损失不到 1%，该算法实现了 26% 的精度损失。
少

2017 年 8 月 17 日提交;最初宣布 2017 年 9 月。

73.关注: 实现自然图像中的准确文本识别

作者:郑展,范白, 徐云禄,郑刚, 浦先普,周水庚

摘要: 场景文本识别由于其应用的多样性, 一直是计算机视觉领域的研究热点。最先进的技术是基于注意的编码器解码器框架, 它以纯粹数据驱动的方式学习输入**图像**和输出序列之间的映射。然而, 我们观察到, 现有的基于注意的方法在复杂和低质量的**图像**上表现不佳。一个主要原因是, 现有的方法无法在特征区域和此类**图像**的**目标**之间获得准确的对齐。我们称之为 "注意力漂移"。针对这一问题, 本文提出了利用聚焦注意机制自动吸引注意力的 fan (聚焦注意网络的缩写) 方法。fan 由两个主要组件组成: 一个是负责识别现有方法中的字符**目标**的关注网络 (an), 以及一个负责通过评估 an 是否支付来调整关注的焦点网络 (fn)正确注意**图像**中的**目标**区域。此外, 与现有方法不同的是, 我们采用了基于 resnet 的网络来丰富场景**文本图像**的深层表示。在各种基准 (包括 iit5k、svt 和 icdar 数据集) 上进行的大量实验表明, fan 方法的性能大大优于现有的方法。少

2017 年 10 月 17 日提交;v1 于 2017 年 9 月 6 日提交;最初宣布 2017 年 9 月。

74. 条件差分自体编码器的逼真面部表情合成

作者:周玉谦,伯特拉姆·施伯特拉姆

文摘: 从单面图像中合成逼真的人脸表情可以广泛应用于人脸识别、情感识别或娱乐的数据增强等方面。这个问题具有挑战性,部分原因是缺乏标记的面部表情数据,使得算法很难消除因身份和表情而产生的变化而产生的歧义。本文提出了用于面部表情合成的条件差分对抗性自动编码器 cdaae。cdaae 采取一个以前看不见的人的面部图像,并产生一个图像的人的脸与目标情绪或面部动作单位标签。cdaae 将前馈路径添加到自动编码器结构中,将编码器的低级特征连接到解码器相应级别的功能。它通过学习生成同一人的图像的低级特征之间的差异,但具有不同的面部表情,从而处理因身份和面部表情而产生的模糊变化的问题。通过组合和插值训练集中的面部表情单位,可以利用 cdaae 结构生成新的表达式。我们的实验结果表明,在为看不见的对象生成面部表情时, cdaae 比以前的方法更忠实地保存身份信息。这在使用小型数据库进行培训时尤其有利。少

2017 年 8 月 30 日提交;最初宣布 2017 年 8 月。

75. sar 图像中目标分类的深度学习: 数据增强与翻译不变性

作者:古河秀俊

文摘: 本文研究了合成孔径雷达 (sar) 图像自动 **目标识别**卷积神经网络 (nn) 的平移不变性。特别是 sar atr 的 cnn 平移不变性代表了对从 sar **图像**中提取的**目标芯片**的不对齐的鲁棒性。为了了解 cnn 的平移不变性, 我们训练了 cnn, 在有数据和无数据增加的条件下, 将**目标芯片**从 mstar 分为十类, 然后可视化 cn。根据我们的研究结果, 即使我们使用深残差网络, cnn 的平移不变性, 而无需使用 mstar **目标芯片**等对齐图像进行数据增强, 也不会那么大。平移不变性的一个更重要的因素是增强训练数据的使用。此外, 我们使用增强训练数据的美国有线电视新闻网实现了 99.6% 的最先进的分类精度。这些结果显示了特定于域的数据扩充的重要性。少

2017 年 8 月 25 日提交;最初宣布 2017 年 8 月。

76. 视频识别中的网络图像注意力传递

作者:李俊南,黄永康, 赵琪, 莫汉·坎坎哈里

摘要: 培训基于深度学习的视频分类器进行**动作识别**需要大量的标记视频。标签过程是劳动密集型和耗时的。另一方面, 用户每天都会将大量弱标签的**图片**上传到互联网。为了利用丰富和高度多样化的 web **映像集**, 一种可扩展的方法是对这些**图像**进行爬网, 以训练基于深度学习的分类器, 如卷积神经网络 (cnn)。但是,

由于域移位问题, 训练深分类器的 web 图像的性能在直接部署到视频时往往会降低。解决此问题的一种方法是在视频上微调受过训练的模型, 但仍需要足够数量的注释视频。在本文的工作中, 我们提出了一种新的方法来将知识从图像域转移到视频域。该方法可以在训练数据数量有限的情况下适应目标域 (即视频数据)。我们的方法使用 cnn 的类判别空间注意力图将视频帧映射到一个低维要素空间。我们设计了一个新的 siamese energy net 结构, 通过共同优化两个损失函数来学习注意力图上的能量函数, 从而使与地面真理概念相对应的注意力图具有更高的能量。我们对两个具有挑战性的视频识别数据集 (即 tvhi 和 ucf101) 进行了广泛的实验, 并展示了我们所建议的方法的有效性。少

2017 年 8 月 2 日提交;最初宣布 2017 年 8 月。

77. 基于多方面感知双向 lstm 递归神经网络的 sar 目标识别

作者: 范章, 陈虎, 尹强, 李伟, 李恒超, 文红

文摘: 深度学习的出色模式识别性能为合成孔径雷达 (sar) 自动目标识别 (atr) 带来了新的活力。然而, 目前基于深度学习的 atr 解决方案存在一个局限性, 即每个学习过程只处理一个 sar 图像, 即学习静态散射信息, 而忽略空间变化信息。显然, 引入空间变化散射信息的多面联合识别应提高分类的准确性和鲁棒性。本文提出了一种新的多方面感知方法, 通过基于空间变化散射信息学习

的双向长期短期存储器 (lstm) 递归神经网络来实现这一思想。具体来说, 我们首先选择不同的方面图像来生成多方面**空间变化**的**图像序列**。然后, 逐步实现 gabor 滤波器和三补丁局部二进制模式 (tplbp), 以提取一个全面的空间特征, 然后利用多层感知器 (mlp) 网络进行维数约简。最后, 我们设计了一个双向 lstm 递归神经网络, 以进一步集成软最大分类器来学习多方面特征, 实现**目标识别**。实验结果表明, 该方法对 10 类**识别**的精度达到 99.9%。此外, 其抗噪声、抗混淆性能也优于传统的深度学习方法。少

2017 年 7 月 25 日提交;最初宣布 2017 年 7 月。

78. 优化: 具有贝叶斯优化的卷积网的联合微调与压缩

作者:[frederick tong](#), [srikanth muralidharan](#), [greg mori](#)

文摘: 在处理专用**图像域**中的新**视觉识别**问题时, 常见的策略是从预先训练的深层神经网络开始, 并根据专门的领域对其进行微调。如果**目标域**覆盖的视觉空间小于用于预培训的源域 (例如 imagenet), 则微调网络很可能会被过度参数化。但是, 将网络修剪作为后处理步骤来减少内存需求有缺点: 微调和修剪是独立执行的;修剪参数设置一次, 随着时间的推移无法适应;而最先进的修剪方法具有高度参数化的特性, 因此无法手动搜索深层网络的修剪参数空间, 从而导致粗糙的近似。我们提出了一个有原则的方法, 共同微调和压缩一个预先训练的卷积网络, 克服这些限制。

在两个专用**图像域**（遥感**图像**和描述纹理）上的实验证明了该方法的有效性。少

2017 年 7 月 28 日提交;最初宣布 2017 年 7 月。

79.共聚焦激光内窥镜图像上的斑块肿瘤检测——一种跨现场鲁棒性评估

作者 :marc aubreville, miguel goncalves, christian knnifer , nicolai oetter, tobias wuerfl, helmut neumann, florian stelzle, christopher bohr, andreas maier

文摘: 卷积神经网络 (cnn) 等深度学习技术为**图像识别**提供了强大的方法, 近年来已被应用于共聚焦激光内镜 (cle) 的癌症自动检测领域**图像**。cle 是一种 (亚) 表面显微**成像**技术, 可达到 1000x 的放大倍率, 因此适用于体内结构组织分析。在这项工作中, 我们的目的是评估一个优先开发的深度学习为基础的算法的前景,**旨在**识别口腔鳞状细胞癌, 并将其推广到进一步的鳞状细胞的解剖位置头部和颈部的癌症。我们将该算法应用于从 5 例经组织学证实的鳞状细胞癌患者声带区域获得的**图像**, 并可能是临床正常的侧向声带健康控制**图像**。我们发现, 在口腔数据上训练的网络在声带数据上的应用时, 准确率达到了 99.45%, 曲线下区域 (auc) 值达到了 0.955。与最先进的数据相比, 我们获得了非常相似的结果, 但使用的是在完全分离的数据集上训练的算法。连接这两个数据集在交叉验证方面得到了进一步的改进, 准确率

为 90.81, auc 为 0.970。在本研究中, 我们首次了解到利用 **cle 图像**进行口腔癌鉴别的深度学习机制, 可应用于头部和颈部领域的其他学科。本研究显示了算法方法的前景, 以推广良好的头部和颈部的其他恶性实体, 无论解剖位置, 并进一步以检查器独立的方式。少

2017 年 7 月 25 日提交;最初宣布 2017 年 7 月。

80. 快速功能傻瓜: 一种独立于数据的通用对抗摄动方法

作者:[konda reddy mopuri](#), [utsav garg](#), [r. venkatesh babu](#)

文摘: 最先进的物体识别卷积神经网络 (cnn) 被图像不可知的摄动 (称为普遍对抗摄动) 所愚弄。还观察到, 这些扰动分布在对同一目标数据进行训练的多个网络中。但是, 这些算法需要训练数据, 对 cnn 进行训练, 并通过复杂的优化计算对抗摄动。这些方法的愚弄性能与可用培训数据的数量成正比。这使得它们不适合实际攻击, 因为攻击者访问训练数据是不合理的。本文首次提出了一种新的数据无关方法, 用于生成一系列接受目标识别训练的 cnn 的**图像**不可知摄动。我们进一步表明, 这些扰动可以通过在相同或不同数据上训练的多个网络体系结构进行转移。在没有数据的情况下, 我们的方法通过愚弄在多层学习到的特征, 有效地生成普遍的对抗扰动, 从而导致 cnn 的错误分类。实验表明, 在

没有任何训练数据的情况下生成的所建议的通用扰动具有令人印象深刻的愚弄率和令人惊讶的可转移性。少

2017 年 7 月 18 日提交;最初宣布 2017 年 7 月。

81. 通过分层卷积特征进行稳健的视觉跟踪

作者:马超,黄家斌, 杨晓康,杨明轩

文摘 本文提出利用深卷积神经网络的丰富层次特征, 提高视觉跟踪的准确性和鲁棒性。在目标识别数据集上训练的深层神经网络由多个卷积层组成。这些图层使用不同级别的抽象对目标外观进行编码。例如, 最后一个卷积层的输出对目标的语义信息进行编码, 这种表示对显著的外观变化是不变的。但是, 它们的空间分辨率过于粗糙, 无法精确地定位目标。相反, 早期卷积图层中的要素提供了更精确的本地化, 但对外观更改的不变程度较低。我们将卷积层的层次特征解释为图像金字塔表示的非线性对应, 并显式利用这些多层次的抽象来表示目标对象。具体来说, 我们学习每个卷积层的输出上的自适应相关筛选器, 以对目标外观进行编码。我们推断每一层的最大响应, 以粗到细的方式定位目标。为了进一步处理规模估计和从严重遮挡或视线外移动引起的跟踪故障中重新检测目标对象的问题, 我们保守地学习了另一个相关滤波器, 该筛选器保持了长期内存目标外观, 作为一个判别分类器。我们将分类器应用于两种类型的对象建议: (1) 小步长的建议, 并紧紧

围绕规模估计的估计位置;(2) 具有较大步长的建议, 并可在整个图像中进行目标重新检测。大规模基准数据集的大量实验结果表明, 该算法在最先进的跟踪方法下具有较好的效果。少

2018 年 8 月 11 日提交;v1 于 2017 年 7 月 12 日提交;最初宣布 2017 年 7 月。

82. 人脸图像的自动质量预测

作者:[lacey best-roden](#), [anil k. jain](#)

文摘: 人脸图像质量可以定义为衡量人脸图像对人脸自动识别的效用。在这项工作中, 我们提出 (和比较) 两种方法, 自动人脸图像质量的基础上, 目标人脸质量值从 (i) 人对人脸图像质量的评估 (比赛无关), 和 (ii) 质量值计算从相似性分数 (取决于媒人)。利用利用深卷积神经网络 (convnet) 提取的人脸特征的支持向量回归模型来预测人脸图像的质量。在 lfw 和 ijb-a 两个无约束人脸图像数据库上对该方法进行了评价, 这两个数据库都包含具有多种质量因素的人脸变化。对拟议的人脸图像质量测量方法的评价表明, 通过使用建议的人脸质量来选择人脸的子集, 我们能够将两个人脸匹配器 (一个 cots 匹配器和一个 convnet 匹配器) 在 1% fmr 时的 fmr 减少至少 13% 图像和视频帧匹配模板 (即, 每个主题的多个面) 在 ijb-a 协议中。据我们所知, 这是在

设计无约束人脸质量预测器时首次利用人脸**图像**质量的评估, 证明该预测器在跨数据库评估中是有效的。少

2017 年 6 月 29 日提交;最初宣布 2017 年 6 月。

83. 什么是我的: 预培训 cnn 有限培训声纳 atr

作者:[john mckay](#), [isaac gerg](#), [vishal monga](#), [raghu raj](#)

文摘: 在声纳图像中查找地雷是一个重大问题, 与航海军事和商业努力密切相关。不幸的是, 由于缺乏庞大的 sonar **图像**数据集, 自动**目标识别**(atr) 算法无法从其他计算机视觉领域中看到的一些相同的进步中获得。也就是说, 卷积神经网络 (cnn) 的繁荣已经能够取得令人难以置信的结果-----甚至超过人类行为者-----对许多 sonar atr 的实践者来说并不是一条容易可行的路线。我们展示了将 cnn 纳入声纳 atr 的一个途径的力量: 转移学习。我们首先展示了一个简单、灵活的美国有线电视新闻网功能提取策略如何能够很好地获得令人印象深刻的结果, 如果不是最先进的结果的话。其次, 我们提出了一种方法, 利用强大的转移学习方法在提供的合成孔径雷达数据集中进行多实例**目标检测**和识别。少

2017 年 6 月 29 日提交;最初宣布 2017 年 6 月。

84. 深神经网络是否受到拥挤的影响?

作者:anna volokitin, gemma roig, tomaso poggio

摘要: 拥挤是人类遭受的一种视觉效果, 在这种效果中, 当其他物体 (称为侧翼) 被放置在它附近时, 可以孤立地识别的物体就不能再被识别。在这项工作中, 我们研究了拥挤的人为神经网络的对象识别的影响。我们分析了标准的深卷积神经网络 (dcnn) 以及新版本的 dhnn, 即 1) 多尺度和 2), 卷积滤波器的大小根据偏心 wrt 到固定中心的大小而变化。这样的网络, 我们称之为偏心依赖, 是灵长类视觉皮层前馈路径的计算模型。我们的研究结果表明, 隔离在目标对象上训练的偏心相关模型可以在侧翼存在的情况下识别这些目标, 如果目标靠近图像的中心, 而 dcnn 不能。此外, 对于所有经过测试的网络, 当在目标上单独训练时, 我们发现网络的识别精度降低的距离越近, 侧翼与目标的距离越高, 侧翼就越多。我们发现, 目标和侧翼之间的视觉相似性也起着一定的作用, 在网络的早期层集中会导致更多的拥挤。此外, 我们还表明, 将侧翼纳入训练组的图像并不能在拥挤的情况下提高性能。

少

2017 年 6 月 26 日提交;最初宣布 2017 年 6 月。

85. 通过贝叶斯脉冲修正稀疏分类实现鲁棒声纳 atr

作者:john mckay, vishal monga , raghu g. raj

摘要: 在 在过去几十年中, 声纳**成像**有了巨大的进步, 部分原因是合成孔径雷达声纳 (sas) 的进步。先进的分类技术现在可以用于声纳自动**目标识别**(atr) 来定位地雷和其他威胁物体。其中最有前途的方法是稀疏重建分类 (src), 它对噪声、模糊和遮挡表现出了令人印象深刻的弹性。我们提出了一个连贯的战略, 以扩大 src 的声纳 atr, 保持 src 的鲁棒性, 同时也能够处理具有不同几何排列, 麻烦的瑞利噪声, 和不可避免的背景杂乱的目标。我们的方法, 构成修正稀疏 (pcs), 结合字典学习方案, 结合字典学习方案, 结合字典学习方案, 结合字典学习方案, 将尖峰和板坯概率分布的新解释, 用于类特定的判别, 作为贝叶斯之前的贝叶斯提取。此外, pcs 还提供了异常检测的可能性, 以避免在不需要额外培训的情况下对训练集中的测试对象进行错误标识。使用美国海军地面作战中心提供的数据库显示了令人信服的结果。少

2017 年 6 月 26 日提交;最初宣布 2017 年 6 月。

86. 动态纹理合成的两流卷积网络

作者:[matthew tesfaldet](#), [marcus a. rubaker](#) , [konstantinos g. derpanis](#)

文摘: 我们介绍了一种用于动态纹理合成的双流模型。我们的模型基于预先训练的卷积网络 (convnet), **它针对**两个独立的任务: (i) **对象识别**和 (ii) 光学流预测。在给定输入动态纹理的情况下, 来自**对象识别** convnet 的滤波器响应的统计信息封装了输入纹理

的每帧外观,而来自光学流 convnet 模型的滤波器响应的统计信息则是动态的。为了生成新的纹理,对随机初始化的输入序列进行了优化,以匹配示例纹理的每个流中的要素统计信息。在最近图像风格转换工作的启发下,在双流模型的支持下,我们还应用合成方法将一个纹理的纹理外观与另一个纹理的动态结合起来,生成完全新颖的动态纹理。我们的方法表明,我们的方法产生新的,高质量的样本,既匹配框架外观和时间演变的输入纹理。最后,我们通过对用户的深入研究,定量地评估了我们的纹理合成方法。

少

2018 年 4 月 12 日提交;v1 于 2017 年 6 月 21 日提交;**最初宣布** 2017 年 6 月。

87. 旋转不变性神经网络

作者:李世元

摘要: 旋转不变性和平移不变性在**图像识别**任务中具有重要的价值。本文提出了一种新的卷积神经网络 (cnn) 体系结构,称为循环卷积层,以实现二维符号**识别**中的旋转不变性。我们还可以通过网络获得二维符号的位置和方向,以实现**对多个非重叠目标的检测目的**。最后但并非最不重要的是,在某些情况下,这种架构可以使用这些不变性实现一次性学习。少

2017 年 6 月 17 日提交;**最初宣布** 2017 年 6 月。

88. 使用字幕生成器进行深层图像表示

作者:konda reddy mopuri, vishal b.athreya , r. venkatesh babu

摘要: 深度学习利用大量标记数据来学习功能强大的模型。当目标数据集较小时, 通常的做法是使用预先训练的模型执行迁移学习, 以学习新的任务特定表示形式。然而, 在训练期间, 预先培训的图像识别 cnn 提供的图像信息有限, 这只是标签。场景检索等任务受到这种薄弱监督中学到的功能的影响, 需要更严格的监督, 以更好地了解图像的内容。本文利用从字幕生成模型中获得的特征, 学习新的任务特定图像表示。特别是, 我们考虑了最先进的字幕系统 "显示和告诉" {snt-pami-2016} 和密集的区域描述模型 densecap ~ \ecece{denseicap-cvpr-2016}。我们表明, 由于在培训过程中提供了更丰富的监督, 字幕系统学到的功能比 cnn 的功能更好。此外, 我们还训练了一个具有修改后的配对损耗的暹罗网络, 以融合由 ~ \cite{nnt-pami-2016} 和 ~ \cite{ine{denseap-cpro-2016}} 学习的图像表示。实验表明, 所提出的融合利用了各个特征的互补性质, 在基准数据集上产生了最先进的检索结果。少

2017 年 5 月 25 日提交;最初宣布 2017 年 5 月。

89. 试论判别式机器学习模型

作者:郭文波,张凯轩, 林林, 黄瑞, 邢新宇

文摘: 通常不可能理解机器学习模型是如何做出决定的。虽然最近的研究提出了各种技术方法, 为学习模型如何做出个人决策提供了一些线索, 但它们无法为用户提供作为一个完整实体检查学习模型的能力。在这项工作中, 我们提出了一个新的技术方法, 以增强贝叶斯回归混合模型与多个弹性网。利用增强型混合模型, 通过全局逼近提取目标模型的解释。为了证明我们的方法的效用, 我们在涵盖文本挖掘和**图像识别**任务的不同学习模型上对其进行了评估。我们的研究结果表明, 该方法不仅在解释单个决策方面优于最先进的技术, 而且还为用户提供了发现学习模型的漏洞的能力。少

2017 年 5 月 23 日提交;最初宣布 2017 年 5 月。

90. 通用 3d 可穿戴指纹目标: 推进指纹读取器评估

作者: [joshua j. engelsma](#), [sunpreet s.arora](#), [anil k.jain](#), [nicolas g. paolter jr](#)

文摘: 我们介绍了高保真通用 3d 指纹**目标**的设计和制造, 这些目标可以在各种指纹传感技术上**进行成像**, 即电容、接触光学和非接触式光学。通用 3d 指纹**目标**首次使指纹读取器的可重复和可控评估, 而且还能够进行指纹读取器互操作性研究。指纹读取器互操作性是指指纹识别系统对不同类型的指纹读取器所获得的**图像**的变化有多大的鲁棒性。为了构建通用的 3d 指纹**目标**, 我们采用了一个成型和铸造框架, 包括 (i) 指纹**图像**的数字映射到负模,

(ii) cad 建模一个脚手架系统, 以保持负模具, (三)使用高分辨率 3d 打印机制造模具和脚手架系统, (iv) 生产或混合具有与人体手指相似的电气、光学和机械性能的材料, 以及 (v) 制造 3d 指纹目标使用可控铸造。我们使用 piv 和附录 f 认证的光学(接触和非接触式)和电容式指纹读取器进行的实验证明了通用三维指纹目标对于受控和可重复指纹读取器的有用性评估以及指纹读取器互操作性研究。少

2017 年 5 月 22 日提交;最初宣布 2017 年 5 月。

91. 具有听觉刺激的图像的跨模式编纂: 一种针对视障人士的语言

作者:takahisa kishino, sun zhe , roberto marchisio, ruggero miceletto

文摘: 在本研究中, 我们描述了一种方法, 以实现视觉图像认知在更广泛的意义上, 通过听觉通道的交叉模态刺激。建立了一种从二维图像转换为声音的原始算法, 并在多个主题上进行了测试。我们的研究表明, 研究对象能够分辨出与不同测试几何形状相对应的 95% 不同声音的精度。此外, 在简单图像上进行简短的学习后, 通过听到声音, 在 16 个复杂和从未受过训练的图像中能够识别出一个单一主题的主题。识别率取决于图像特征, 在 90% 的情况下, 研究对象的表现要好于随机选择。这项研究有助于理解交叉模态感知, 并有助于实现使用声学信号帮助视障人士识别物体和提高导航能力的系统

2017 年 5 月 15 日提交;最初宣布 2017 年 5 月。

92. 多尺度空间加权局部直方图在 $O(1)$

作者 :[mahdiah poostchi](#), [ali shafiekhani](#), [kannappan palaniappan](#), [guna seetharaman](#)

摘要: 考虑到其位置的加权像素贡献是许多基本图像处理任务的一个关键特征, 包括滤波、对象建模和距离匹配。提出了几种将空间信息结合起来的技術, 以提高检测、跟踪和识别系统的准确性, 并以速度为代价提高其性能。但是, 如何利用积分直方图在恒定时间内有效地显式加权局部直方图尚不清楚。本文提出了一种利用加权积分直方图 (swih) 进行快速搜索的一种新算法, 用于在恒定时间内精确计算多尺度空间加权局部直方图。我们应用空间加权积分直方图方法进行快速跟踪, 并获得了更准确、更稳健的目标定位结果, 与普通直方图进行了比较。少

提交于 2017 年 5 月 9 日;最初宣布 2017 年 5 月。

93. 反式深信念网络的近似计算

作者:[徐晓静](#),[斯林霍伊达斯](#),[肯·克鲁兹-德尔加多](#)

摘要: 概率生成神经网络在图像分类、语音等许多应用中都很有用。

更多

2017 年 5 月 6 日提交;v1 于 2017 年 4 月 13 日提交;最初宣布 2017 年 4 月。

94. 集识别的质量感知网络

作者:刘宇,严俊杰,欧阳万里

文摘: 本文针对集识别问题, 学习两个图像集之间的度量。每个集合中的图像属于相同的标识。由于一组图像可以是互补的, 因此希望它们能在实际应用中获得更高的精度。但是, 每个样品的质量都不能得到保证, 质量差的样品会影响指标。本文提出了质量感知网络 (qan) 来解决这一问题, 在培训阶段没有明确提供此类信息的情况下, 可以自动了解每个样本的质量。网络有两个分支, 其中第一个分支提取每个样本的外观特征嵌入, 另一个分支预测每个样本的质量得分。然后聚合集合中所有示例的特征和质量分数, 以生成最终的特征嵌入。我们表明, 在只考虑到集级别标识注释的情况下, 可以以端到端方式训练这两个分支。对该机制梯度分布的分析表明, 网络学到的质量有利于集对集识别, 简化了网络所需的分布。在人脸验证和人的重新识别方面的实验表明, 所提出的 qan 具有优势。源代码和网络结构可在 <https://github.com/sciencefans/Quality-Aware-Network> 下载。少

2017 年 4 月 11 日提交;最初宣布 2017 年 4 月。

95. 语义分割和对象检测的对抗性示例

作者: [谢子航](#), [王建宇](#), [张志辉](#), [周玉音](#), [谢玲西](#), [朱尼尔](#)

摘要: 事实已经证明, 在图像分类上, 深部网络通常存在对抗性的例子, 即添加了视觉上无法察觉的扰动的自然图像. 本文将对抗示例扩展到语义分割和对象检测, 这是困难得多。我们的观察是, 分割和检测都是基于对图像上的多个目标进行分类 (例如, 基本目标是分割中的像素或接受场, 以及检测中的对象建议), 激励我们在生成对抗扰动的一组像素/建议上优化损失函数。基于这一想法, 我们提出了一种名为 "密集对抗生成" (dag) 的新算法, 该算法生成大量的对抗示例, 并应用于各种最先进的深度网络进行分割和检测。我们还发现, 对抗扰动可以跨网络传输, 具有不同的训练数据, 基于不同的体系结构, 甚至用于不同的识别任务。特别是, 具有相同体系结构的网络之间的可转移性比其他情况下更重要。此外, 总结异构扰动往往会带来更好的传输性能, 从而为黑匣子对抗攻击提供了一种有效的方法。少

2017 年 7 月 21 日提交;v1 于 2017 年 3 月 24 日提交;最初宣布 2017 年 3 月。

96. 图像中侵犯人权行为的检测: 卷积神经网络能有所帮助吗?

作者: [grigorios kalliatakis](#), [shoaib ehsan](#), [maria fasli](#), [ales leonardis](#), [juergen gall](#), [klaus d. mcdonal-maier](#)

摘要: 在建立了图像、视频、语音和音频处理的性能基准后, 深卷积网络是近年来**图像识别**任务取得最大进展的核心。这就提出了一个问题, 即针对这些引人注目的深度架构, 不尝试通过数字**图像**承认侵犯人权行为, 是否有任何好处。在这一视角下, 我们引入了一个新的、以人权为中心的新的、抽样良好的数据集, 称为 "人权理解" (hrun)。我们在一个共同点上进行严格的评估, 将这一数据集与不同的最先进的深卷建筑结合起来, 以实现**对侵犯人权行为的承认**。在 hrun 数据集上的实验结果表明, 性能最好的 cnn 架构可以实现高达 88.10% 的平均平均精度。此外, 我们的实验表明, 增加训练样本的大小对于在使用非常深的网络时主要提高平均精度至关重要。少

2017 年 3 月 16 日提交;v1 于 2017 年 3 月 12 日提交;**最初宣布** 2017 年 3 月。

97. 从深层面模板中重建人脸图像

作者: [广康麦](#), [曹凯](#), [pong c. yuen.](#), [anil k. jain](#)

摘要: 最先进的人脸**识别**系统基于深度 (卷积) 神经网络。因此, 必须确定从深层网络派生的人脸模板在多大程度上可以倒置, 以获得原始的人脸**图像**。本文研究了一种基于模板重建攻击的最先进的人脸**识别**系统的脆弱性。我们提出了一个邻域去卷积神经网络 (\text{untnet}) 来重建面部**图像**从他们的深层模板。在我

们的实验中, 我们假设没有关于目标主体和深度网络的知识。为了训练具有齐子的 nbnet} 重建模型, 我们增强了两个基准人脸数据集 (vgg-face 和多 pie), 并使用人脸生成器合成了大量图像。使用 i 型 (将重建后的图像与用于生成深层模板的原始人脸图像进行比较) 和 ii 型 (将重建后的图像与深模板进行比较) 和第二类图像进行了评估。不同的人脸形象的同一主题) 攻击。考虑到从 \textit{nnets} 重建的图像, 我们表明, 为了进行验证, 我们在 i 型 (类型 ii) 攻击 @ far 为 0.1% 的情况下, 在 lfw 上实现了 95.20 (50.5%) 的 tar。此外, 从分区模板 xtext{f} (:text{fb}) 中重建的图像的 96.58% (92.84%) 可以从颜色 feret 中的分区 \textit{fo} 中标识出来。我们的研究表明, 需要在人脸识别系统中保护深度模板。少

2018 年 4 月 28 日提交;v1 于 2017 年 3 月 2 日提交;最初宣布 2017 年 3 月。

98. 利用重新描述挖掘对认知障碍和阿尔茨海默病患者的临床和生物学特性进行比较

作者:matej mihelč ić , goran š imić , mirjana babić leko, nada lavrač , saš o džeroski, tomlavš mt

摘要: 我们使用重新描述挖掘来查找可解释的规则, 这些规则揭示了那些决定因素之间的联系, 这些决定因素提供了有关阿尔茨海默氏症 (ad) 的见解。我们将 clus-rm 重构挖掘算法扩展到基

于约束的重构挖掘 (cbm) 设置, 从而实现对特定的用户约束关联进行**有针对性的**探索的几种模式。重新描述挖掘能够找到临床和生物属性的特定结构, 描述不同规模、同质性和认知障碍水平的许多研究对象群体。我们证实了一些先前已知的发现。然而, 在某些情况下, 与属性: 睾酮, 成像属性异常的空间模式识别早期 ad, 以及瘦素和血管生成素-2 的水平, 在血浆中, 我们证实之前有争议的发现或提供了更多的信息, 这些变量及其与 ad 发病机制的关联。对 adni 数据进行重新描述挖掘的结果是发现了一个基本未知的属性: 妊娠相关蛋白-a (app-a), 我们发现它与 ad 中的认知障碍高度相关。在 app-a 与不同临床试验之间发现了统计学意义上的相关性 ($p < 0.01$)。这一发现的高度重要性在于, papp-a 是一种金属蛋白酶, 已知它能裂解胰岛素样生长因子结合蛋白。由于它还与 a 不整合蛋白和金属蛋白酶家族的酶共享相似的底物, 这些酶在非淀粉样变途径中对生理上的裂解淀粉样体前体蛋白 (app) 起到 α 分泌酶的作用, 因此它可以直接参与在疾病过程中, app 的代谢非常早期。因此, 进一步的研究应更深入地探讨 papp-a 在反倾销发展中的作用。少

2017年11月14日提交;v1于2017年2月20日提交;**最初宣布** 2017年2月。

99. 前瞻性探地雷达埋地目标分类的特征分类方法的较大比较

作者: [joseph a. camilo](#) , [leslie m. collins](#), [jordan m. malof](#)

摘要: 前瞻性探地雷达 (flgpr) 最近被研究为一种遥感方式, 用于探测**埋地**目标 (例如地雷)。在此背景下, 将原始 flgpr 数据成**图像**, 然后应用计算机算法对地下**埋地**目标进行自动检测。大多数现有算法都是受监督的, 这意味着它们被训练来区分标记的**目标**图像和非**目标**图像, 通常基于从图像中提取的特征。为此提出了大量的特点, 但迄今尚不清楚哪些是最有效的。这项工作的第一个目标是利用大量 flgpr 数据集合上的现有功能, 对检测性能进行全面比较。还考虑了处理每个要素所产生的决策的融合。本工作的第二个目标是从**对象识别文献**中研究两种现代特征学习方法: 用于 flgpr 处理的视语袋和 fisher 向量。结果表明, 新的特征学习方法优于现有方法。结果还表明, 现有功能和新功能之间的融合几乎不会带来额外的性能改进。少

于 2017 年 2 月 9 日提交;**最初宣布** 2017 年 2 月。

100.胸部 x 线照片多级分割的完全卷积体系结构

作者:[阿列克谢·诺维科夫](#)、[季米特里奥斯·列尼斯](#)、[大卫·梅杰](#)、[吉里·赫拉德乌伊夫卡](#)、[玛丽亚·温默](#)、[katja bühler](#)

文摘 深层卷积神经网络在**图像**分类和**识别**任务方面的成功, 在非常多样化的环境中得到了新的应用, 包括医学**成像**领域。本文研究并提出了用于胸部 x 线摄影解剖器官 (即肺、锁骨和心脏) 的多级分割的神经网络结构。我们通过融合卷积网络中最近的概念

并使其适应 cxr 中的分段问题, 解决了一些开放的挑战, 包括模型过度拟合、减少参数数量和处理 cxr 中严重不平衡的数据。我们展示了我们的架构结合了延迟亚采样, 指数线性单位, 高度限制性正则化和大量的高分辨率低级抽象特征优于最先进的方法, 在所有被考虑的器官, 以及人类在肺和心脏上的观察者这些模型使用具有三个目标类的多类配置, 并在公开的 jsrt 数据库上进行培训和测试, 其中包括 247 个 x 射线图像, 可在 scr 数据库中使用地面真相掩码。我们最好的表现模型, 训练与损失函数的 dice 系数, 达到平均贾卡德重叠分数 99.09% 的肺, 868.7% 的锁骨和 882% 的心脏。这种结构的表现优于人类观察者对肺和心脏的观察结果。少

Submitted 13 February, 2018; **submitted** 30 January, 2017; **originally announced** January 2017.

101. 基于互信息的正向特征选择方法的理论基础

作者 :francisco macedo , m. rosário oliveira, antónio pacheco, rui valadas

摘要: 特征选择问题在微阵列分析、临床预测、文本分类、图像分类和人脸识别、多标签学习和互联网分类等各种应用中都出现了交通。在各类方法中, 基于互信息的前瞻性特征选择方法已成为人们非常流行的方法, 并在实践中得到了广泛的应用。然而, 由于基于特定的数据集和分类器, 对这些方法的比较评价受到限制。在本文中, 我们开发了一个理论框架, 允许评估的方法, 基于他

们的理论性质。我们的框架是基于这些方法试图近似的**目标目标**函数的性质, 以及根据它们对类解释的贡献对特征进行新的分类; 我们推导出**目标目标**函数的上限和下限, 并将这些边界与特征类型联系起来。然后, 我们描述了这些方法所采用的近似类型, 并分析了这些近似如何处理**目标目标**函数的良好性质。此外, 我们还开发了一个分布设置, 旨在说明这些方法的各种缺陷, 并提供了几个错误特征选择的示例。根据我们的工作, 我们清楚地确定了应该避免的方法, 以及目前性能最好的方法。少

2018 年 2 月 14 日提交;v1 于 2017 年 1 月 26 日提交;**最初宣布** 2017 年 1 月。

102. 用于路边监视的无线电层析成像

作者: [christopher r. anderson](#), [richard k.martin](#) , [t. owens walker](#), [ryan w.thomas](#)

摘要: 无线电层析**成像** (rti) 最近被提议通过无线电波跟踪物体的位置, 而不需要物体传输或接收无线电信号。该位置是通过推断哪些方数阻碍密集无线传感器网络中的无线电链路子集而提取的。针对路边监控方案, 本文提出了 rti 的各种建模和算法改进方法。其中包括使用更有物理动机的重量矩阵、减少因噪声观测而产生的负面 (形而上学) 数据的方法, 以及将移动车辆的车架组合到单个**图像**中的方法。在测量数据集中, 所提出的方法可显

示成像（对人在环目标 识别有用）和自动目标识别方面的改进。

少

2016 年 12 月 14 日提交;最初宣布 2017 年 1 月。

103. 噪声语音识别的不变表示

作者 :d 三国 riy serdyuk, kartik audhkhasi, philémon brakel, bhuvana ramabhadran,samuel thomas, yeshua bengio

摘要: 现代自动语音识别(asr) 系统需要在环境、扬声器、通道和记录条件产生的声学变异性下具有鲁棒性。在现代基于神经网络的 asr 系统中, 确保这种对可变性的鲁棒性是一项挑战, 尤其是在训练中看不到所有类型的可变性的情况下。我们试图通过鼓励神经网络声学模型来学习不变特征表示来解决这个问题。我们使用了最近关于使用生成对抗性网络**生成图像**的研究的思想和扩展基于对抗梯度的培训的领域适应思想。ganin 等人最近的一项工作提出, 利用主**目标**分类网络的中间表示对图像域适应**进行对抗训练**, 以提高域分类器的性能通过一个单独的神经网络。我们的工作重点是研究神经结构, 这些结构产生了 asr 噪声条件的表示不变。我们评估了 aurora-4 任务的拟议体系结构, 这是噪声鲁棒性 asr 的流行基准。我们表明, 我们的方法概括比标准的多条件训练更好, 特别是在训练中只看到少数噪声类别的情况下。

少

2016 年 11 月 27 日提交;最初宣布 2016 年 12 月。

104. 使用预先培训的模型从弱标记数据自动构建新域的人脸数据集

作者:[丁胜勇](#),[吴俊宇](#), [徐伟](#), [赵洪阳](#)

摘要:训练数据在人脸识别系统中至关重要。但是,为特定域的大规模人脸数据标记非常繁琐。本文提出了一种利用海量弱标记数据自动、增量地构造数据集的方法。更多

2016 年 11 月 24 日提交;最初宣布 2016 年 11 月。

105. 基于形态的星系图像数据库查询

作者:[廖尔·沙米尔](#)

摘要:罕见形态的星系具有最高的科学意义,因为它们承载着关于过去、现在和未来宇宙的重要信息。一旦一个罕见的星系被发现,更有效地研究它需要一组类似形态的星系,允许概括和统计分析,而这是不能做的 $n=1$ 。数字天空测量生成的数据库可以包含大量的星系图像,因此,一旦发现一个罕见的感兴趣的星系,数据库中也有可能存在更多相同形态的实例。然而,当研究人员在数据库中发现某一星系的罕见形态时,在寻找类似形态的星系时,几乎不可能手动挖掘数据库。在这里,我们提出了一种计算机方法,可以自动搜索星系图像的数据库,并识别在形态上类似于特定用户定义的查询星系的星系。也就是说,研究人员提供了一个

感兴趣星系的**图像**，模式**识别**系统自动返回视觉上与**目标**星系相似的星系列表。该算法使用一组全面的描述符，使其能够支持不同类型的星系，而且不限于有限的已知形态集。虽然返回星系的列表既不干净也不完整，但它包含了更高频率的星系形态，提供了大量的数据减少。这种算法可以集成到自主数字天空测量的数据管理系统中，如大型天气测量望远镜 (lsst)，其中数据库中的星系数量非常多。该方法的源代码可在 <http://vfacstaff.ltu.edu/lshamir/downloads/udat>。少

2016 年 11 月 19 日提交;最初宣布 2016 年 11 月。

106. 有效领域转移深网络的图像可信度分析

作者:金志伟,曹娟,罗洁波,张永东

摘要: 大量的假**图像**传播在今天的社交媒体上，并可能严重危及在线内容的信誉向公众。本文采用深层网络来学习明显的假**图像**相关特征。与真实的**图像**相比，假**图像**往往是引人注目的和视觉上惊人的。与传统的视觉**识别**任务相比，理解这些心理触发的假**图像**视觉模式是极具挑战性的。传统的一般**图像**分类数据集（如 imagenet set）是为对象级别的特征学习而设计的，但不适合学习**图像**可信度分析所需的超特征。为了克服假**图像**训练样本的不足，我们首先构建了一个与这一任务间接相关的大型辅助数据集。此辅助数据集包含从社交媒体自动收集的 60 万个弱标记的假图

像和真实**图像**。通过类似 adaboos 的传输学习算法，我们训练了一个 cnn 模型，在**目标训练集**中有几个实例，在收集的辅助集中有 60 万**张图像**。这种学习算法能够利用辅助集的知识，逐步将其转移到**目标任务**中。在一个真实的测试集中进行的实验表明，我们提出的域名转移 cnn 模型优于几个竞争基线。它在一般 imagenet 集的基础上，通过迁移学习方法得到了超强的结果。此外，案例研究表明，我们提出的方法揭示了一些有趣的模式，以区分假和真实的**图像**。少

2016 年 11 月 16 日提交;最初宣布 2016 年 11 月。

107. 用于在杂乱场景中进行端到端主动对象识别的重复三维注意网络

作者:[刘敏](#),[石一飞](#),[郑林涛](#),[徐凯](#)

摘要: 主动视觉本质上是以注意力为导向的: 代理选择观察视图以最佳方式处理视觉任务, 同时改进其对被观察场景的内部表示。在基于单个 rgb **图像**的 2d 视觉任务中, 基于注意的模型最近取得成功的启发下, 我们建议使用注意机制解决基于主视图的主动对象**识别**问题, 方法是开发端到端的经常性 3d 关注网络。该体系结构由一个递归神经网络 (rnn) 组成, 存储和更新内部表示, 以及两级空间变压器单元, 指导两级关注。我们的模型采用 3d 形状数据库训练, 能够迭代地关注**针对**感兴趣对象的最佳视图, 以识

别该对象, 并将重点放在每个视图中的对象上, 以消除背景混乱。为了实现三维视图的选择, 我们推导出了一个可微用于反向传播训练的三维空间变压器网络, 实现的收敛速度必须快于大多数现有基于注意力的模型所采用的增强学习。实验表明, 该方法在杂乱场景中优于最先进的方法。少

2016 年 10 月 13 日提交;最初宣布 2016 年 10 月。

108. 多实例学习卷积神经网络的对象识别

作者: 苗 孙, [tony x. han](#), [ming-changliu](#), [ahmad khodayari-rostamabad](#)

摘要: 卷积神经网络 (cnn) 在计算机视觉、语音识别和自然语言处理等领域的成功应用。对于对象识别, cnn 可能会受到其严格的标签要求和隐含的假设的限制,即图像应该是目标-对象主导的最佳解决方案。然而, 标签程序, 需要布局目标对象的位置, 是非常繁琐的, 使高质量的大规模数据集昂贵得令人望而却步。在深度网络存在训练数据不足的情况下, 数据扩充方案得到了广泛的应用。通过数据扩充生成的所有图像都共享相同的标签, 这可能是有问题的, 因为并非所有的数据扩容方法都是标签保留。本文提出了一个弱监督的 cnn 框架, 名为多实例学习卷积神经网络 (milnnn) 来解决这个问题。我们应用 milnnn 框架进行对象识别, 并在三个基准数据集上报告最先进的性能: cifar10、cifar100 和 ilsvrc2015 分类数据集。少

2016 年 10 月 10 日提交;最初宣布 2016 年 10 月。

109. 帕洛马尔瞬态工厂调查中的小近地小行星: 实时条纹探测系统

作者: adam waszczak, thomas a.prince, rass laher, frank masci, brian bue, umaa rebbapragada, tombarlow, jason surace, george helou, shrinivas 库尔卡尔尼

摘要: 在 1–100 米大小范围内的近地小行星 (nea) 估计为或数量是或目前已编制了 15, 000 个 nea, 其中大部分在 0.5–10 公里的范围内。10–100 米大小的近地天体的撞击在统计上不会危及生命, 但可能会造成重大的区域损害, 而 1–10 米大小的近地天体相对于地球的速度较低, 是空间飞行任务的主要目标。我们描述了一个实时 nea 发现系统的实现和初步结果, 该系统专门用于检测帕洛马尔瞬态工厂 (ptf) 图像中的小、高角速度 (视条纹) 近地天体。ptf 是一个 1.2 米的光圈, 7.3–deg² 光学视场测量主要用于在 60 秒曝光中发现星系外瞬变 (如超新星) 或 20.5 视觉震级。我们的实时 nea 发现管道使用机器学习的分类器来过滤大量的假阳性条纹检测, 允许人类扫描仪在夜间有效和远程识别真正的小行星条纹。在识别有条纹的 nea 检测 (通常在发现曝光后一小时内) 后, 扫描仪将触发相同望远镜的跟踪, 并将观测结果发布到小行星中心进行全球确认。我们描述了我们最初确认的十个发现, 所有的小近天体都经过了距离地球 0.3–15 的月球距离。最后, 我们推导出有用的标度定律, 用于比较不同测量的链路近地天体

检测能力, 作为其硬件和测量模式特征的函数。这项工作最直接地告知对 zwicky 瞬态设施 (ztf 计划在 2017 年成功 ptf) 的条纹检测能力的估计, 该设施将在 47-deg 上应用 ptf 目前的分辨率和灵敏度 2 场。少

2016 年 9 月 26 日提交;最初宣布 2016 年 9 月。

110. 使用字符草图和重心转移的无语言字符识别

作者:masoud nostrati, fak 尾 h rahimi, ronak karimi

文摘: 在本研究中, 我们提出了一种字符识别的启发式方法。为此, 从包含要识别的字符的图像中构造草图。此草图包含最重要的图像像素, 它们是原始图像的代表。这些点是图像逐像素匹配中最可能的点, 适合 目标图像。此外, 还利用一种称为重力移动的技术来处理字符的伸长问题。将素描和重力技术结合起来的結果导致了一种无语言的字符识别方法。该方法可以作为特征提取算法, 独立实现实时使用或结合其他分类器。低复杂性和可接受的性能是这种方法最令人印象深刻的功能, 使其能够简单地移动和电池有限的计算设备中实现。结果表明, 在最佳情况下, 获得 86% 的准确率, 在最坏的情况下, 28% 的识别字符是准确的。少

2016 年 8 月 3 日提交;最初宣布 2016 年 8 月。

111. 用于鲁棒细粒场景识别的语义聚类

作者:maian george, mandar dixit, gábor zogg, nuno vasconcelos

摘要: 在域泛化中, 从一个或多个源域中获得的知识被转移到看不见的**目标域**。在本文中, 我们提出了一种新的细粒度场景**识别**的域泛化方法。我们首先提出一个语义场景描述符, 共同捕获细粒度场景之间的细微差异, 同时对跨域的不同对象配置保持鲁棒性。我们对场景中对象的发生模式进行建模, 捕捉每个场景中每个对象的信息和可分性。然后, 我们将此类事件转换为每个场景**图像**的场景概率。其次, 我们认为场景**图像**属于隐藏的语义主题, 可以通过聚类我们的语义描述符来发现。为了评估所提出的方法, 我们提出了一个新的细粒度场景数据集在跨域设置。在建议的数据集和三个基准场景数据集上进行的大量实验表明了所提出的细粒度场景传输方法的有效性, 我们的性能优于最先进的场景**识别**和域泛化方法。少

2016 年 7 月 26 日提交;最初宣布 2016 年 7 月。

112. 计算颜色不变的深层结构-输出回归学习

作者:钱延林,陈克,乔尼-kristian kamarainen, jarno nikkanen, jiri matas

摘要: 计算颜色的恒久性是计算机视觉中一个基本而活跃的问题, 它可以表述为回归问题。要学习颜色恒常的强大回归, 获取有意义的图像特征并捕获输出变量之间的潜在相关性起着至关重要

的作用。在本文中，我们引入了一个新的深层结构-输出回归学习框架，以同时实现这两个目标。通过借用最初为视觉识别而设计的深层卷积神经网络 (cnn) 的力量，该框架可以自动发现在不同光照条件下进行白平衡的强大特征，并学习超越特征和目标之间的基础关系的多输出回归器，用于查找目标变量的不同维度的复杂相互依存关系。在两个公共基准上的实验表明，与最先进的方法相比，我们的方法实现了竞争性能。少

2016 年 8 月 11 日提交;v1 于 2016 年 7 月 13 日提交;最初宣布 2016 年 7 月。

113. 无监督域适应的深度重构分类网络

作者:muhammad Kleijn, w. bastiaan Bastiaan, mémjie zhang, david balduzzi, wen li

文摘: 本文提出了一种新的基于深度学习的视觉目标识别无监督域适应算法。具体而言，我们设计了一个名为 "深度重建分类网络" (drcn) 的新模型，该模型共同学习两个任务的共享编码表示: i) 标记源数据的监督分类，以及 ii) 无监督重建未标记的目标数据。这样，学到的表示不仅保留了可分性，而且还从目标域中编码了有用的信息。我们的新 drcn 模型可以通过类似地利用反向传播作为标准神经网络进行优化。我们评估 drcn 在一系列跨域对象识别任务上的性能，与以前最先进的算法相比，drcn 提供了相当大的改进（精度高达 8%）。有趣的是，我们还观察到 drcn 的

重建管道将源域中的**图像**转换为外观类似于**目标数据集的图像**。这表明 drcn 的性能是由于构造了一个单一的复合表示形式, 该表示形式对**有关目标图像**结构和源**图像**分类的信息进行编码。最后, 我们提供了一个形式化的分析, 以证明算法的目标在域适应上下文中的合理性。少

2016 年 8 月 1 日提交;v1 于 2016 年 7 月 12 日提交;**最初宣布** 2016 年 7 月。

114. 弱标记语义视频对象分割的半监控域适应

作者:王慧玲,塔帕尼雷科,拉塞·连苏,王廷怀,朱哈·卡胡宁

摘要: 在给定大型标记数据集的许多高级计算机视觉任务中, 深层卷积神经网络 (cnn) 取得了巨大的成功。然而, 对于视频语义对象分割来说, 标签稀缺的领域, 有效地利用 cnn 在训练数据有限的情况下的表示能力仍然是一个挑战。简单地借用现有的预先训练的 cnn **图像识别**模型进行视频分割任务, 就会严重影响视频分割任务的性能。我们提出了一种半监督的方法, 利用从 cnn 学到的语义证据和内在结构, 将从标记**图像**数据训练的 cnn **图像识别**模型应用于**目标域**视频数据。通过显式建模和补偿从源域到**目标域**的域转换, 该方法针对自然中外观、形状和遮挡的变化, 建立了一种鲁棒的语义对象分割方法。视频。我们在具有挑战性的

数据集上进行了广泛的实验，证明了与最先进的方法相比，我们的方法具有卓越的性能。少

2016 年 6 月 7 日提交;最初宣布 2016 年 6 月。

115. 通过概念和属性进行操作分类

作者: [amir rosenfeld](#), [shimon ullman](#)

摘要: 自然图像中的类倾向于遵循长尾分布。当罕见课程的培训示例不足时，这是有问题的。复合类强调了这种效果，涉及几个概念的结合，例如出现在操作识别数据集中的概念。在本文中，我们建议通过学习如何利用现成的常见视觉概念来解决这个问题。我们检测图像中存在的显著概念，并使用它们来推断目标标签，而不是直接使用视觉特征，将视觉和自然语言处理中的工具结合起来。我们验证了我们的方法，最近引入的 hico 数据集达到了 31.54% 的 map 和 standfort-40 action 数据集，其中建议的方法优于通过直接视觉特征获得的方法，获得了 83.12 的精度。此外，该方法为每个类提供了一个语义上有意义的关键字列表和将其与其组成概念相关联的相关图像区域。少

2018 年 3 月 6 日提交;v1 于 2016 年 5 月 25 日提交;最初宣布 2016 年 5 月。

116. 自动检测无线电地图中的扩展源: 到访的进展情况

作者: s. riggi, a. ingallinera, p. leto, f. cavallaro, f. bufano, f. schillir, c. trigilio, g. umana, c. s. buemi, r. p. norris

摘要: 自动源提取和参数化是下一代无线电干涉仪测量的一个关键挑战, 例如使用平方公里阵列 (ska) 及其前体进行的测量。本文提出了一种新的算法, 称为 caesar (紧凑型 and 扩展源自动识别), 用于检测和参数化无线电干涉测量地图中的扩展源。它基于预滤波阶段, 允许图像去噪, 紧凑的源抑制和增强漫反射, 其次是一个自适应的超像素聚类阶段, 用于最终的源分割。参数化阶段为后处理分析提供了源通量信息和广泛的形态学估计。我们在模块化软件库中开发了 caesar, 包括用于局部背景估计和图像过滤的不同方法, 以及紧凑和漫反射源提取的替代算法。该方法被应用于澳大利亚望远镜紧凑型阵列 (atca) 在 scorpio 项目内收集的真实无线电连续体数据, 该项目是 askap+emu 调查的开创者。在存在紧凑源、成像文物和来自银河平面的漫射发射的情况下, 对不同测试领域的源重建能力进行了研究, 并与现有算法进行了比较。与人为分析相比, 所设计的算法能够检测已知的目标源和漫反射发射区域, 在考虑的领域上优于替代方法。少

2016 年 5 月 6 日提交;最初宣布 2016 年 5 月。

117. 规则分解空间中多维数据的分层建模在图像分析中的应用

作者: 奥利维尔·盖伊

文摘: 最后一篇文章展示了**图像**分析中分层建模技术的逐步引入。第一章论述了模式**识别**在工业应用领域开展的首批工作。第二章重点介绍了这些技术在卫星图像中的使用情况, 以及开发卫星数据存档系统的问题, 以便在数字地理中使用这些技术。第三章是基于平面**图像**分析的人脸**识别**和部分隐藏模式的**识别**。本出版物的结尾是对未来与现行视频编码标准有关的静止图像或移动图像的自我描述编码系统的描述。与前几篇文献一样, 它将在附件**算法中找到基于图像**分析的算法, 这些算法基于两种互补的方法: 人工视觉的工业应用基于边界的方法;-基于区域的**卫星图像**分析方法。少

2016 年 5 月 4 日提交;最初宣布 2016 年 5 月。

118. 在极端压缩率下的纯相位压缩滤波

作者 : david pastor-calle, anna pastuszczak, m 奇 哈尔·mikolajczyk, ra 夸尔·kotynski

文摘: 我们介绍了一种重建**压缩测量图像**与纯相位滤波器之间相关性的有效方法。该方法基于相位滤波的两个性质: 这种滤波是一个一元循环变换, 它产生的相关平面通常是稀疏的。由于这些特性, 仅相位滤波器与压缩传感框架完全兼容。此外, 采用纯相位滤波作为压缩矩阵时, 基于激光的恢复算法速度非常快。该方法可以看作是基于相关的模式**识别**技术的推广, 在此直接应用于非

自适应采集的压缩数据。在测量时，不需要事先了解将扫描数据的目标对象。我们表明，以极高的压缩率测量的**图像**可能仍然包含足够的信息，用于**目标分类**和本地化，即使压缩率足够高，视觉识别重建**图像**中的**目标**不再可能。该方法已应用于从单像素相机获得的高度低采样测量，并基于随机选择的 walsh-hadamard 模式进行采样。少

2016 年 9 月 29 日提交;v1 于 2016 年 4 月 26 日提交;**最初宣布** 2016 年 4 月。

119. 典性在对象分类中的作用: 提高卷积神经网络的泛化能力

作者:[babak 萨利赫](#), [ahmed elgammal](#), [jacob feldman](#)

文摘: 深人工神经网络在计算机视觉领域的不同任务中取得了显著的进展。然而，这些模型的实证分析及其失败案例的调查最近受到了关注。在这项工作中，我们表明，深度学习模型不能概括为与训练**图像**有很大不同的非典型**图像**。这与人脑视觉系统的卓越泛化能力形成鲜明对比。我们关注的是卷积神经网络 (cnn) 作为**目标识别**和分类领域的最新模型;对这一问题进行了更详细的研究，并假设训练 cnn 模型会受到非结构化损失最小化的影响。我们提出了计算模型，通过考虑训练**图像**的典型外观来提高 cnn 的泛化能力。通过大量的实验，我们表明，涉及典性测量可以大大

提高一组**新图像**的分类结果。更重要的是,在不对**目标图像集**进行微调的情况下,实现了这一重大改进。少

2016 年 2 月 9 日提交;最初宣布 2016 年 2 月。

120. 几何鲁棒声纳 atr 的本地化词典设计

作者:john mckay, vishal monca, raghu raj

摘要: 声纳**图像**捕获的进步为自动目标**识别**(atr) 提供了强大的分类方案的大门。最近的工作特别看到了稀疏重建为基础的分类(src) 应用于声纳 atr, 它提供了令人信服的准确率,即使在存在噪声和模糊。然而,现有的基于稀疏的声纳 atr 技术假定测试**图像**显示的几何姿态与训练集是一致的。这项工作解决了与训练相比处理不一致的测试声纳**图像**的突出开放挑战。我们开发了一种新的基于块的本地化字典设计,它可以实现几何,即姿势鲁棒性。此外,还采用了字典学习方法,以提高性能和效率。提出的具有局部姿势管理(lpm)的 src 由于其识别声纳图像中背景杂乱的能力,优于最先进的 sift 特征和 svm **方法**。少

2016 年 1 月 13 日提交;最初宣布 2016 年 1 月。

121. 声纳 atr 的判别稀疏性

作者:john mckay, raghu raj, vishal mona, jason isaacs

摘要: 声纳**图像**捕获的进步使研究人员能够应用复杂的目标识别算法, 以便在地雷等**图像**中定位感兴趣的**目标**。尽管在这一领域取得了进展, 但现代声纳自动**目标识别**(atr) 方法对实际情况中预期的噪声量缺乏鲁棒性, 处理物理模糊的能力。**图像**捕获, 以及在相对较少的训练样本中脱颖而出的能力。我们通过将现代基于稀疏的技术与由每个班级的培训组成的字典相适应来应对这些挑战。我们开发了新的判别 (而不是生成) 稀疏表示, 可以帮助声纳成像中自动对**目标**进行分类。利用海军地面作战中心 (nswc) 的模拟 sas 数据集, 即使在训练样本中噪声和稀疏性较大的情况下, 我们也获得了具有吸引力的多级问题分类率。少

2016 年 1 月 1 日提交;最初宣布 2016 年 1 月。

122. 视觉识别的中级表示

作者:莫因·纳比

摘要: 视觉识别是人工智能的基本挑战之一, 人工智能的目标是理解视觉数据的语义。特别是采用中级表示, 改变了视觉识别的范式。中级**图像**/视频表示包括发现和训练一组中级视觉模式 (例如, 零件和属性), 并表示利用它们的**给定图像**/视频。利用视觉现象的运动和外观信息, 可以从**图像**和视频中提取中层模式。本文的目标是将中级表示方法应用于不同的高级视觉识别任务, 即 (i) 图像理解和 (ii) 视频理解。在**图像**理解的情况下, 我们关注的是

对象检测/识别任务。我们研究发现和学习一组用于表示对象类别图像的**中级补丁**。我们专门以一种可分类感知的杂草监管的方式使用判别补丁。此外，我们还研究了通过使用基于子分类的模型来消除数据集偏差所提供的结果。少

2015 年 12 月 22 日提交;最初宣布 2015 年 12 月。

123. 预先训练的神经网络可以检测解剖吗？

作者:[vado menkovski](#), [zharko Aleksovski](#), [axel saalbach](#), [h 下 es nickisch](#)

文摘: 卷积神经网络在标记训练数据丰富的计算机视觉和语音识别任务中表现出突出的经验成果。在医学**成像**中，有各种各样的**可能的成像**方式和对比，其中注释数据通常非常稀缺。我们提出了应对这一挑战的两种办法。在具有丰富数据的不同域中预训练的网络被用作特征提取器，而后续分类器则在较小的**目标数据集**上进行训练;和一个深刻的建筑训练与沉重的扩增，并配备了复杂的正则化方法。我们在 x 射线**图像**的语料库上测试了方法，设计了一个解剖检测系统。少

2015 年 12 月 18 日提交;最初宣布 2015 年 12 月。

124. 即兴的高目标检测和操作

作者:[abhishek maity](#)

文摘: 在突出的主题识别的情况下, 计算机算法一直严重依赖扫描图像从左上角到右下角系统, 并在试图定位感兴趣的对象时施加暴力。因此, 这个过程被证明是相当耗时的。本文讨论了解决上述问题的一种新方法和简单的方法。本文通过分割映射实现了一种目标操作和检测方法, 这将有助于去饱和化, 换句话说, 有助于消除图像的背景。利用贾卡德指数对著名的地面真相目标箱技术进行性能评价。少

2015 年 11 月 10 日提交;最初宣布 2015 年 11 月。

125. 用于面部表情识别和动作单元检测的可变状态潜在条件随机域

作者: [robert walecki](#), [Ognjen rudovic](#), [vladimir pavlovic](#) , [maja pantic](#)

摘要: 从视频中自动识别面部表情和检测面部动作单元 (au) 关键取决于其动态建模。这些动态的特点是时间相 (固定偏移) 和情绪表达强度和 au 的变化, 其外观在不同的目标对象之间可能会有很大的差异, 使识别/检测任务非常具有挑战性。最先进的潜在条件随机字段 (l-crf) 框架允许一个人通过潜在的状态有效地编码这些动态, 这解释了情绪表达的时间一致性和其强度之间的有序关系水平, 这些潜在的状态通常被认为是无序的 (名义上的) 或完全有序的 (序数)。然而, 这种做法往往限制性太强。例如, 在 au 检测的情况下, 目标是区分此 au 处于活动状态或非活动状态的图像序列的段。虽然包含目标 au 激活的序列段可以最好用

序数潜在状态来描述, 但非活动段最好用无序(名义)潜在状态来描述, 因为不能对它们的序列段做出任何假设。基础结构(因为它们可以包含中性面或非目标 au 的激活)。为了解决这个问题, 我们提出了可变状态 l-crf (vsl-crf) 模型, 该模型自动选择目标 图像序列的最佳潜在状态。为了减少模型过度拟合的名义或序数的潜在状态, 我们提出了一个新的图-拉普拉斯正则化的潜在状态。我们在三个公共表达数据库上的实验表明, 与传统的 l-crf 和其他相关的最新模型相比, 该模型具有更好的泛化性能。少

2015 年 10 月 13 日提交;最初宣布 2015 年 10 月。

126. 原子网络: 一种用于结构药物发现生物活性预测的深卷进化神经网络

作者: [ighar wallach](#), [michael dzamba](#), [abraham heifets](#)

摘要: 深层卷积神经网络由一个深部神经网络(dnn)子类组成, 其结构受限, 利用它们所建模的域的空间和时间结构。卷积网络通过将简单的局部特征分层组合到复杂的模型中, 在语音和图像识别等领域实现了最佳的预测性能。尽管 dnn 已被用于 qsar 的药物发现和基于配体的生物活性预测, 但这些模型都没有从这一强大的卷积结构中受益。本文介绍了原子网络, 这是第一个基于结构的深层卷积神经网络, 用于预测小分子在药物发现应用中的生物活性。我们展示了如何将特征局部和层次组成的卷积概念

应用于生物活性和化学相互作用的建模。与现有的 dnn 技术相比, 我们表明, 原子能网将局部卷积滤波器应用于结构**目标**信息, 成功地预测了以前不知道的**目标的新**活性分子调制 器。最后, 我们展示了原子网在一组不同的基准上的表现大大优于以前的对接方法, 在 dude 基准中的 57.8% 上实现了超过 0.9 的非盟委员会。少

2015 年 10 月 9 日提交;最初宣布 2015 年 10 月。

127. 上下文感知区域神经码的深层属性

作者:[罗建伟](#),[李建国](#), 王军, 蒋志刚,[陈玉龙](#)

文摘: 近年来, 许多研究将卷积神经网络模型 (cnn) 的中间层输出作为不同视觉**识别**任务的特征。虽然在一些实证研究中取得了可喜的成果, 但这种类型的表现仍然受到众所周知的语义差距问题的影响。本文从三个方面提出了所谓的深层属性框架来缓解这一问题。首先, 我们引入对象区域建议作为表示**目标图像**的媒体, 并从区域建议中提取特征。其次, 我们研究了所有地区提案的不同 cnn 层的综合特征。聚合可生成**输入图像**的整体而紧凑的表示形式。结果表明, 软最大层输出的跨区域最大池比所有其他层都要高。由于软最大层直接对应于语义概念, 因此这种表示称为"深层属性"。第三, 我们观察到, 只有一小部分生成的区域由对象建议算法与分类**目标**相关。因此, 我们引入上下文感知区域细化

算法来挑选上下文区域并构建上下文感知分类器。我们将提出的深度属性框架应用于各种视觉任务。对图像分类、细粒度识别和可视化实例检索这三种视觉识别任务的标准基准进行了广泛的实验。结果表明，深度属性方法获得了最先进的结果，并且以显著的优势超越了现有的对等方法，尽管有些基准与预先训练的 cnn 模型几乎没有概念重叠。少

2015 年 9 月 8 日提交;最初宣布 2015 年 9 月。

128. 使用基于附加的编码器描述多媒体内容——解码器网络

作者:赵京云,阿伦·考维尔, [yeshua bengio](#)

摘要: 虽然深度神经网络最初主要用于分类任务，但它们在结构化输出问题领域正在迅速扩展，在结构化输出问题领域，观察到的目标由多个随机变量组成，这些变量具有丰富的联合分布，给定输入。本文重点介绍了输入结构也丰富且输入和输出结构在某种程度上是相关的情况。我们描述了学习在输入中注意不同位置、输出的每个元素的系统，用于各种任务：机器翻译、图像标题生成、视频剪辑描述和语音识别。所有这些系统都基于一组共享的构建块：门控递归神经网络和卷积神经网络，以及训练有素的关注机制。我们报告了这些系统的实验结果，显示了令人印象深刻的良好性能和注意机制的优势。少

2015 年 7 月 3 日提交;最初宣布 2015 年 7 月。

129. 基于航拍图像序列的车辆自动跟踪和识别

作者:[奥格涅仁·阿兰杰洛维奇](#)

文摘: 本文从航空图像序列出发, 研究了车辆**自动跟踪和识别**的问题。在现有文献成功的推动下, 重点是使用线性外观子空间来描述多视图对象的外观, 并突出其作为实际系统的一部分应用所涉及的挑战。描述了一个工作解决方案, 其中包括数据提取和规范化的步骤。在对真实数据的实验中, 拟议的方法取得了很有希望的结果, 正确**识别率**很高, 很少出现有意义的错误 (第二类错误, 即真正相似的目标有时与之相混淆)。彼此)。讨论了今后研究的方向和可能改进的方法。少

2015 年 6 月 23 日提交;最初宣布 2015 年 6 月。

130. 预测未标记视频中的视觉表示

作者:[carl vondrick](#), [hamed pirsiavash](#), [antonio torralba](#)

摘要: 在操作和对象开始或出现之前对其进行预测是计算机视觉中的一个难题, 因为有几个实际应用程序。这项任务具有挑战性, 部分原因是它需要利用世界上难以写下来的广泛知识。我们相信, 有效学习这些知识的一个有希望的资源是通过现成的未标记视频。我们提出了一个框架, 利用时间结构在未标记的视频, 以学习预测人类的行动和对象。我们方法背后的关键思想是, 我们可以训

练深度网络来预测**未来图像**的视觉表现。可视化表示是一个很有前途的预测**目标**,因为它们在比自动计算的像素更高的语义级别上对**图像**进行编码。然后,我们将**识别算法应用于**预测表示对象和操作。我们在两个数据集中实验验证了这一想法,预测了未来一秒的行动,并在未来 5 秒钟内进行了目标。少

2016 年 11 月 29 日提交;v1 于 2015 年 4 月 29 日提交;**最初宣布** 2015 年 4 月。

131. 学习自由朴素贝叶斯最近的邻里域名改编

作者:[faraz saeedan](#), [barbara caputo](#)

摘要: 到目前为止,对象分类算法还不能达到在现实世界中可靠工作所需的鲁棒性和通用性水平。即使是我们所训练的最强大的卷积神经网络,在对来自不同数据库的数据进行训练和测试时,也无法令人满意地执行。此问题在文献中称为域自适应和/或数据集偏差,是由于数据集合之间的分布不匹配造成的。解决这一问题的方法从最大边距分类器到学习如何修改特征并获得更强大的表示形式。最近的研究表明,通过**将**问题引入图像到类**识别**框架中,域适应问题得到了显著缓解 \cite{danbnn}。在这里,我们遵循这种方法,并展示了一个非常简单的,学习免费的朴素贝叶斯最近邻居 (nbnn) 基于域适应算法如何可以显著缓解源和**目标**数据之间的分布不匹配,尤其是当类的数量和源的数量增长。关于

文献中使用的标准基准的实验表明, 我们的方法 (a) 在小规模问题上与目前的最新水平相竞争, (b) 随着类和来源数量的增长, 实现了目前的最新水平, 而且数量很少计算要求。少

2015 年 3 月 26 日提交;最初宣布 2015 年 3 月。

132. 通过近似的重载函数获得单图像超分辨率

作者:邓良健,郭伟宏,黄廷珠

摘要: 图像超分辨率是提高图像分辨率的过程。广泛应用于医学成像、卫星成像、目标识别等领域。本文进行了连续建模, 假设未知图像强度函数是在连续域上定义的, 属于具有冗余基础的空间。我们提出了一种基于观测的单图像超分辨率迭代模型:图像由平滑分量和非光滑分量组成, 我们使用两类近似的重信度函数 (ahf) 来实现分别代表他们。由于非光滑成分的稀疏性,我 1 模型被使用。此外, 我们还将所提出的迭代模型应用于图像补丁, 以减少计算和存储。与现有的一些竞争方法进行了比较, 表明了该方法的有效性。少

2015 年 3 月 12 日提交;最初宣布 2015 年 3 月。

133. 基于时间金字塔的行动识别卷积神经网络

作者:王元洲,曹元洲,沈春华, 刘灵桥,沈恒涛

摘要: 在卷积神经网络 (cnn) 在**图像**分类方面取得成功的鼓舞下, 近年来在将 cnn 应用于基于视频的动作**识别**问题方面付出了大量的努力。一个挑战是视频包含不同数量的帧, 这与 cnn 的标准输入格式不兼容。现有的方法可以直接采样固定数量的帧, 也可以通过引入 3d 来绕过这个问题在时空域中进行卷积的卷积层。为了解决这个问题, 我们在这里提出了一个新的网络结构, 它允许任意数量的帧作为网络输入。我们解决方案的关键是引入一个由编码层和时间金字塔池层组成的模块。编码层将以前图层的激活映射到适合于池的要素向量, 而时间金字塔池图层将多个帧级激活转换为固定长度的视频级表示形式。此外, 我们还采用了将外观信息和运动信息结合起来的特征串联层。与帧采样策略相比, 我们的方法避免了缺失任何重要帧的风险。与需要大量视频数据集进行网络训练的三维卷积方法相比, 我们的模型可以在一个小的**目标数据集**上学习, 因为我们可以利用现成的**图像级** cnn 来获取模型参数初始化。在两个具有挑战性的数据集 (hollywood 2 和 hmdb51) 上进行的实验表明, 我们的方法比最先进的方法具有卓越的性能, 同时需要更少的训练数据。少

2015 年 4 月 16 日提交;v1 于 2015 年 3 月 4 日提交;**最初宣布** 2015 年 3 月。

134. 基于卷积神经网络的判别式显著度图在线跟踪

作者:seunghoonhong, tackgeunyou, suha kwak, bohyung han

文摘: 我们提出了一种在线视觉跟踪算法, 利用卷积神经网络 (cnn) 学习判别显著性图。考虑到 cnn 在离线时对大型**图像**存储库进行了预先训练, 我们的算法将网络隐藏层的输出作为特征描述符, 因为它们在各种一般视觉中**表现出出色的表示性能**识别问题。这些功能用于使用在线支持向量机 (svm) 学习判别目标外观模型。此外, 我们还在支持向量机的指导下, 通过反传播 cnn 特征构造目标特定的显著性图, 并根据用显著性映射等价构造的外观模型, 获得**每个帧**的最终跟踪结果。由于显著性地图能有效地实现**目标**的空间配置, 提高了**目标定位**的准确性, 使我们能够实现像素级**目标**分割。我们通过在具有挑战性的基准上进行的广泛实验来验证我们的跟踪算法的有效性, 在这种实验中, 我们的方法与最先进的跟踪算法相比, 展示了出色的性能。少

2015 年 2 月 24 日提交;最初宣布 2015 年 2 月。

135. 一种在杂声场景中的对称零件检测框架

作者 : tom lee, sanja fidler, alex levinstein, ristian minchesescu, sven dickinson

摘要: 在计算机视觉的发展过程中, 对称性的作用从一开始就起起就越来越重要。起初, 它在支持自下而上的索引方面表现得很突出, 但由于形状被外观所让位,**识别被识别**出现在的位置, 它就失去了青睐。在**以目标对象**的形式出现强势先验的情况下, 感知群所提供的较弱的前科的作用大大削弱。然而, 随着字段从大型

数据库返回识别问题，在杂乱场景中构成对象的部分的自下而上的恢复对于识别对象至关重要。长期以来，中轴群落一直利用对称的普遍规律作为将闭合轮廓分解为内侧部分的基础。然而，今天的识别系统面临着杂乱的场景，假设存在封闭的轮廓，即图形-地面分割已经解决，使许多中间轴社区的工作不适用。在本文中，我们回顾了以前在 lee 等人 (2013 年)、levinshtein 等人 (2009 年、2013 年) 中报告的计算框架，该框架将中间轴的代表能力与在杂乱场景中恢复和分组对象部件的需要架起了桥梁。我们的框架植根于这样一种想法，即一个最大刻板，一个中轴的组成部分，可以建模为图像中的一个紧凑型超级像素。我们对杂乱场景图像的方法进行了评价。少

2015 年 2 月 5 日提交;最初宣布 2015 年 2 月。

136. 可转换性多视点零射击学习

作者:傅艳薇,蒂莫西·霍斯帕达莱斯,陶香,龚少刚

摘要: 大多数现有的零镜头学习方法都是通过在带注释的辅助数据集和具有不同类且没有注释的目标数据集之间共享的中级语义表示来利用迁移学习。从辅助数据集中学习从低级要素空间到语义表示空间的投影，并在不适应目标数据集的情况下应用。在本文中，我们确定了这些方法的两个固有限制。首先，由于有不相交和潜在不相关的类，从辅助数据表域中获得的投影函数在直接应

用于目标数据集域时是有偏差的。我们将这一问题称为投影域偏移问题, 并提出了一种新的框架——多视图嵌入来解决。第二个限制是原型稀疏问题, 它指的是这样一个事实, 即对于每个目标类, 只有一个原型可用于零镜头学习给定语义表示。为解决这一问题, 提出了一种新的异构多视图超图标签传播方法, 用于传感器嵌入空间中的零镜头学习。它有效地利用了不同语义表示所提供的互补信息, 并以连贯的方式利用了多表示空间的多种结构。通过大量实验证明, 所提出的方法 (1) 纠正了辅助域和目标域之间的投影偏移, (2) 显著利用了多重语义表示的互补性, (3) 在三个图像和视频基准数据集上, 实现零拍摄和 n 拍摄识别的现有方法优于现有方法, 并且 (4) 支持新的跨视图注释任务。少

2015 年 3 月 2 日提交;v1 于 2015 年 1 月 19 日提交;最初宣布 2015 年 1 月。

137. 关于 ddos 攻击相关的最小切割问题

作者:qi uan, haadi jafarian, ehab al-shaer, jjinhui xu

摘要: 本文研究了经典最小切割问题的两个重要扩展, 即 $\{\text{em 连接保持最小切割 (cpmc)}\}$ 问题和 $\{\text{em 阈值最小切割 (tmc)}\}$ 问题, 它们在大规模 ddos 攻击中有着重要的应用。在 cpmc 问题中, 寻求最小削减, 以将源与目标节点分开, 同时保持源与其伙伴节点之间的连接。cpmc 问题在应急响应、图像处理、模式

识别和医学等许多其他领域也有重要应用。在 tmc 问题中, 寻求最小削减, 将目标节点与伙伴节点的阈值数量隔离开来。tmc 问题是网络抑制问题的一个重要特例, 在网络安全中有着重要的应用。我们表明, 一般的 cpmc 问题不能在 l . 不, 不, 不 G_n 除非 $nP=P$ 具有准多项式算法。我们还证明了平面图中两组 cpmc 问题的一个特例可以在多项式时间内求解。这一结果的必然结果是, 平面图中的网络分流问题 P , 是以前打开的问题。结果表明, 阈值最小节点切割 (tmnc) 问题可以近似在比范围内 $\alpha(n^{1/2})$ 阈值最小边缘切割问题 (tmec) 可以近似在比率范围内 $\alpha(\log 2n)$. 我们还回答了另一个长期存在的开放问题: 网络抑制问题的硬度和网络阻截问题。我们表明, 它们都不能在任何恒定比范围内近似。除非 $nP \notin \text{OBP}$ 我们 $\text{me}(2n)$. 少

2015 年 4 月 17 日提交;v1 于 2014 年 12 月 10 日提交;最初宣布 2014 年 12 月。

138. 多实例强化学习, 实现图像中高效的弱监控检测

作者: [stefan mathe](#) , [cristian smisesescu](#)

摘要: 最先进的视觉识别和检测系统越来越依赖于大量的培训数据和复杂的分类器。因此, 手动对数据集进行批注和将运行时间保持在实际应用程序可接受的水平变得越来越昂贵。在本文中, 我们提出了解决这些问题的两种解决办法。首先, 我们引入了一

种弱监督、基于分段的方法, 从弱监控信号中学习精确的检测器和**图像**分类器, 只提供对**目标**定位的近似约束。我们用人类视觉搜索模式作为训练信号, 说明了我们的系统在静态**图像**中的动作检测问题 (pascal voc action 2012)。其次, 从人类视觉系统的搜索和固定工作原理出发, 采用强化学习技术训练高效的检测模型。我们的顺序方法是弱监督和一般的 (它不需要眼球运动), 找到任何给定的检测置信度函数的最佳搜索策略, 并实现类似于详尽的滑动窗口搜索的性能, 其一小部分计算成本。少

2014 年 11 月 29 日提交;最初宣布 2014 年 12 月。

139. 用于视觉识别和描述的长期递归卷积网络

作者 :[jeff donahue](#), [lisa anne hendricks](#), [marcus rohrbach](#), [subhashini venugopalan](#) , [sergio guadarrama](#), [kate saenko](#), [trevor darrell](#)

摘要: 基于深卷积网络的模型主导了近期的**图像**解释任务;我们调查模型是否也是经常性的, 或 "时间深度", 是有效的任务涉及序列, 视觉和其他。我们开发了一种适用于大规模视觉学习的新型递归卷积架构, 该架构是端到端可持续的, 并展示了这些模型在基准视频**识别**任务、图像描述和**图像**描述方面的价值。检索问题和视频旁白的挑战。与目前假设序列处理为固定时空接受场或简单时间平均的模型不同, 反复卷积模型是 "双重深" 的, 因为它们可以在时空 "层" 中构成。当**目标**概念复杂, 训练数据有限时,

这类模型可能具有优势。当非线性被整合到网络状态更新中时, 可以学习长期依赖关系。长期的 mn 模型很有吸引力, 因为它们可以直接将可变长度输入 (例如视频帧) 映射到可变长度输出 (例如自然语言文本), 并可以模拟复杂的时间动态;但它们可以通过反向传播进行优化。我们反复出现的长期模型直接连接到现代视觉凸网模型, 并可以共同培训, 同时学习时间动力学和卷积感知表示。我们的结果表明, 与单独定义和优化的最先进的识别或生成模型相比, 此类模型具有明显的优势。少

2016 年 5 月 31 日提交;v1 于 2014 年 11 月 17 日提交;**最初宣布** 2014 年 11 月。

140. 重力透镜类星体的数据挖掘

作者:adriano agnello, brandon c. kelly, tommaso treu, phillip j. marshall

文摘: 引力透镜 (gl) 类星体比未被带光的类星体更明亮, 并产生具有独特形态特征的**图像**。过去的搜索和**目标**选择算法, 特别是斯隆准镜头搜索 (sqls), 都依赖于基本的形态标准, 这些标准被应用于明亮的、光谱上确认的类星体样本。sqls 技术不足以搜索新的调查 (例如 des、ps1、lsst), 因为光谱信息不容易获得, 而且大数据量要求目标/候选选择的纯度更高. 我们对机器学习技术进行了系统的探索, 并证明了两步策略可以非常有效。在第一步中, 我们使用目录级别的信息 (GriZ+ wise 大小, 第二个时刻),

使用人工神经网络预先选择**目标**。然后使用像素逐像素模式识别算法（渐变引导树）检查接受的目标，以形成最终的候选项。此过程的结果可用于进一步细化更简单的 sqls 算法，有一个两倍（或三倍）获得纯度和相同（或 80%）在**目标**选择阶段的完整性，或纯度 70%和完整性的 60%在候选人选择步骤之后。中更简单的光度搜索 GriZ+ 基于颜色切割的 wise 将提供样品 7。%纯度或更低。我们的技术非常快，因为可以在 {几个} cpu 小时内从第三阶段实验（例如 des 目录数据库）中获得候选列表。这些技术可以很容易地扩展到第四阶段的实验，如 lsst，并添加时域信息。少

2014 年 10 月 16 日提交;最初宣布 2014 年 10 月。

141. 用于对象识别的域自适应神经网络

作者: [muhammad Kleijn](#), [w. bastiaan Bastiaan](#), [mwjie zhang](#)

文摘: 提出了一个简单的神经网络模型来处理**目标识别**中的域自适应问题。我们的模型将最大平均差异 (mmd) 度量作为监督学习中的正则化，以减少源域和**目标域**在潜在空间中的分布不匹配。实验表明，mmd 正则化是在特定**图像数据集**的 surf 特征和原始**图像像素**上提供良好域适应模型的有效工具。我们还表明，我们提出的模型，在之前进行去噪自动编码器预训练，在相同的数据集中实现了比最近的基准模型更好的性能。这项工作代表了在神经网络背景下首次对 mmd 测量进行研究。少

2014 年 9 月 21 日提交;最初宣布 2014 年 9 月。

142. 利用深卷积神经网络的稀缺训练数据实现视频识别的转移学习

作者:苏玉川,赵子轩,叶春燕,黄信富,许文森

摘要: 无约束视频识别和深度卷积网络 (dcn) 是近年来计算机视觉中的两个活跃课题。在这项工作中,我们将 dcns 应用为基于框架的视频识别识别器。然而,我们的初步研究表明,具有完整地面真相的视频语料库通常规模不大,多样性不够,无法学习一个稳健的模型。直接在视频数据集上训练的网络严重过度拟合,在测试集上的识别率很低。同样的培训样本问题限制了在广泛的计算机视觉问题上使用深度模型,因为在这些问题上很难获得培训数据。为了克服这个问题,我们进行了从图像到视频的学习,利用弱标记图像语料库中的知识进行视频识别。图像语料库有助于学习自然图像的重要视觉模式,而这些模式则被只在视频语料库上训练的模型所忽视。因此,所产生的网络具有更好的通用性和较好的识别率。我们展示了通过将学习从图像转移到视频的方式,我们可以学习一个只有 4k 视频的基于框架的识别器。由于图像语料库的标记较弱,因此整个学习过程只需要 4k 注释实例,这远远低于以前作品所需的百万比例图像数据集。同样的方法也可应用于只有稀缺的培训数据的其他视觉识别任务,并提高了 dcns 在各种计算机视觉问题中的适用性。考虑到目标问题和数据的性质,我们的实验还揭示了元参数与 dcns 性能之间的相关

性。这些结果为未来的研究带来了元参数选择的启发式选择，而不依赖于耗时的元参数搜索。少

2015年6月15日提交;v1于2014年9月14日提交;最初宣布2014年9月。

143. 用于域适应的分层自适应结构支持向量机

作者:徐交龙, 塞巴斯蒂安·拉莫斯, david vazquez, antonio m. lopez

摘要: 分类中的一个关键主题是训练（源）域中的数据分布与测试（目标）域中的数据分布不同时产生的精度损失。这被认为是许多计算机视觉任务（如图像分类、对象检测和对象类别识别）的一个非常相关的问题。在本文中，我们提出了一种新的域适应方法，它利用了分层适应树中的多个目标域（或子域）。其核心理念是利用共同考虑的目标域的共性和差异。考虑到结构 svm (ssvm) 分类器的相关性，我们将我们的想法应用到自适应 svm (a-ssvm) 中，它只需要目标域样本和现有的源域分类器来执行所需的适应。总之，我们将我们的建议称为分层 a-ssvm (ha-ssvm)。作为概念的证明，我们使用 ha-ssvm 进行行人检测和对象类别识别。在前者中，我们将 ha-ssvm 应用于基于变形的部分模型 (dpm)，而在后一种方法下，将 ha-ssvm 应用于多类别分类器。在这两种情况下，我们都展示了 ha-ssvm 如何有效地提高了忽略目标数据结构的适应策略的检测/识别精度。由于

目标数据的子域并不总是先验已知的，因此我们展示了 ha-ssvm 如何将子域结构发现纳入对象类别识别。少

2014 年 8 月 22 日提交;最初宣布 2014 年 8 月。

144. 激光振动测量的模态不敏感极限，光谱还原需要超对称性

作者:michael c. kobold

摘要: 实验室测量表明，在少数情况下，利用遥感对物体振动特征进行识别和监测是有限的。这项工作提供了两个必要条件，以推断在规定的范围内确定实际目标;未能识别频谱被证明是罕见的。激光回波的模态调制利用慢扫正弦 (sss) 和小偏转多模态 (mm) 分析生成足够的光谱 id 数据集。使用这些完全不同的计算结果可实际去除用于物体 id 的遥感问题、光谱 "还原" (sr) 返回。光学回归为非**成像**检测和分类提供了结构模式 id。使用大的光斑尺寸来完全绘制振动物体的计算为使用点大小作为变量的实验室测量中发现的 sr 提供了洞察。比较 sss 和 mm 近似的**非成像**计算表明，振动矩形板具有空间集成 (非**成像**) 返回，在低频振动模式之间有很大差异。这两种方法的数据聚类是 id 的必要条件。sss 和 mm 理论描述了模态**识别**能力的信号处理物理，以及纯一维模态**目标**和超对称方板如何阻碍分类。少

2014 年 8 月 21 日提交;v1 于 2014 年 8 月 10 日提交;最初宣布 2014 年 8 月。

145. 2009 年朝圣视频录制中的单个目标跟踪分析

作者:mohamed h. dridi

文摘: 本文介绍了在聚集了数千人的高密度人群场景中跟踪单个目标的多种方法（手动、半自动和自动）。关于个体运动的必要数据和许多其他物理信息可以通过不同的方式从连续的**图像**序列中提取出来，包括光流和块运动估计。块匹配方法是跟踪运动物体的著名方法之一。这种估计主题运动的方法需要一个比较窗口的规范，它确定估计的比例。在这项工作中，我们提出了一个实时方法，行人识别和跟踪序列的高分辨率**图像**获得的固定（高清晰度）相机在不同的地方在圣地**清真寺**在麦加。目标是估计行人速度与局部密度的函数有关。下面一节介绍了基于视频序列跟踪移动行人的数据。通过被评估的系统每个行人的时空坐标在 tawaf 仪式期间建立。说明并非常精确地记录了朝圣者的速度，这是马塔夫地区当地密度的函数。少

2014 年 8 月 7 日提交;v1 于 2014 年 7 月 8 日提交;**最初宣布** 2014 年 7 月。

146. 拓展草核家族：一种嵌入的视角

作者 :mehrtash t. harandi , mathieu salzmnn, sadeep jayasumana, richard hartley, hantong li

摘要: 将视频和**图像**集建模为线性子空间已被证明对许多视觉**识别**任务有益。然而, 它也带来了挑战, 因为线性子空间不服从欧几里得几何, 但在于一种特殊类型的黎曼流形称为格拉斯曼尼安。为了利用为欧几里得空间(例如支持向量机)开发的技术和子空间, 最近的几项研究提出了利用一个正定核将格拉斯曼尼亚嵌入到希尔伯特空间中的建议。不幸的是, 只有两个格拉斯曼核是已知的, 其中没有一个——正如我们将展示的——是通用的, 这限制了它们任意逼近**目标**函数的能力。在这里, 我们介绍了几个正定的草核, 包括宇宙核, 并展示了它们在分类、聚类、稀疏编码和哈希等各种任务中优于以前已知的内核。少

2014 年 7 月 4 日提交;最初宣布 2014 年 7 月。

147. 通用凸网表示的可转换性因素

作者 :[hossein azizpour](#), [ali sharif razan](#), [josefine sullivan](#), [atsuto maki](#), [stefan carlsson](#)

摘要: 有证据表明, 卷积网络(convnet)是视觉**识别**任务中最有效的表示学习方法。在常见的方案中, convnet 在较大的标记数据集(源)上进行训练, 而在网络的某一层激活训练过的网络的前馈单元, 则用作相对具有相对特性的任务的**输入图像**的通用表示形式。较小的训练集(**目标**)。最近的研究表明, 这种形式的表示转移适用于广泛的**目标**视觉**识别**任务。本文介绍并研究了影响这种表示的可转移性的几个因素。它包括源 convnet 的训练参数,

如其体系结构、训练数据的分布等, 以及特征提取的参数, 如训练的 convnet 层、维数约简等。然后, 通过对这些因素的优化, 我们表明, 在各种 (17) 视觉识别任务上可以实现显著的改进。我们进一步表明, 这些视觉识别任务可以根据它们与源任务的距离来明确排序, 从而使任务的执行与源任务之间的距离之间的相关性。观察。少

2015 年 7 月 15 日提交;v1 于 2014 年 6 月 22 日提交;最初宣布 2014 年 6 月。

148. 图像拼接技术综述

作者:dushyant vaghela , kabildev naina 教授

文摘: 图片说明马赛克是一种将同一场景的多个图像构造成更大图像的方法。图像拼接的输出将是两个输入图像的联合。图像拼接算法用于获得拼接图像。图片说明拼接处理基本上分为 5 个阶段。其中包括;特征点提取,图像注册, 主声像计算, 扭曲和混合, 如果图像。各种角点检测算法被用于特征提取。此角落可生成高效且信息丰富的输出马赛克图像。图像拼接广泛应用于三维图像、医学成像、计算机视觉、卫星数据和军用自动目标识别等方面。少

2014 年 5 月 11 日提交;最初宣布 2014 年 5 月。

149. 深卷活化特征的多尺度无秩序池

作者:[龚云超](#),[王立伟](#),[郭瑞奇](#),[斯韦特拉娜·拉兹布尼克](#)

摘要: 深层卷积神经网络 (cnn) 已经显示出它们作为**识别**的普遍表示的希望。然而, 全球 cnn 激活缺乏几何不变性, 这限制了它们对高度可变场景的分类和匹配的鲁棒性。为了在不降低 cnn 激活的判别能力的情况下提高其不变性, 本文提出了一种简单而有效的方案——多尺度无阶汇集 (mop-cnn)。该方案提取 cnn 在多个比例级别上的本地补丁激活, 在每个级别分别执行这些激活的无顺序 vlad 池, 并将结果连接在一起。由此产生的 mop-cnn 表示可以作为监督或非监督**识别**任务的通用功能, 从**图像分类**到实例级检索;它的性能始终优于 cnn 的全球激活, 而无需为特定**目标数据集**对预测图层进行任何联合培训。从绝对意义上说, 它在具有挑战性的 SUN397 和 mit 室内场景分类数据集上实现了最先进的结果, 并在 ilsvrc2012/2013 分类和 inria 假日检索数据集上获得了竞争结果。少

2014 年 9 月 8 日提交;**v1** 于 2014 年 3 月 7 日提交;**最初宣布** 2014 年 3 月。

150. 用于广域跟踪的平移变焦相机的连续定位和映射

作者:[giuseppe lisanti](#), [iacopo masi](#), [fedico pernici](#), [alberto del bimbo](#)

摘要: 泛倾斜变焦 (ptz) 摄像机功能强大, 可支持远场场景中的目标识别和**识别**。然而, ptz 摄像机在实际环境中的有效使用变得更加复杂, 因为需要连续的在线摄像机校准, 并且无法使用摄像机执行器提供的绝对平移、倾斜和缩放位置值, 因为不使用与视频流同步。因此, 必须直接从帧的视觉内容中提取准确的校准。此外, 大尺度和突然的规模变化, 场景背景的变化, 由于相机的操作和相机运动补偿的需要, 使**目标跟踪**与这些相机极具挑战性。在本文中, 我们提出了一个解决方案, 提供了连续在线校准 ptz 相机, 这是强大的快速相机运动, 环境的变化, 由于照明或移动的物体和规模超过成千上万的地标。该方法直接推导了**目标**在三维世界平面中的位置与二维**图像**中相应的比例和位置之间的关系, 并允许对具有高稳定性的多个**目标**进行实时跟踪即使在远距离和任何缩放水平的精度。少

2015 年 3 月 23 日提交;v1 于 2014 年 1 月 25 日提交;**最初宣布** 2014 年 1 月。