

## 1. 多发性硬化症病变分割的多分支卷积神经网络

**作者:**shahab aslani, michael dayan , loredana storelli, masimo fippi, vittorio murino, maria a rocca, diego sona

**抽象:** 本文提出了一种从多模态脑磁共振图像中分割多发性硬化症 (ms) 病变的自动化方法。我们的方法是基于深端对端卷积神经网络 (cnn) 的基于切片的三维体积数据的分割。拟议的 cnn 包括一个多分支向下采样路径, 使网络能够分别对多个模式的切片进行编码。提出了多尺度特征融合块, 将网络不同阶段不同模式的特征图结合起来。然后, 提出了多尺度特征上采样块, 以提高不同分辨率的特征地图组合, 利用病变形状和位置的信息。我们使用每个 3d 模式的正交平面方向训练和测试我们的模型, 以利用各个方向的上下文信息。拟议的管道在两个不同的数据集上进行评估, 一个是包括 37 名 ms 患者的私人数据集, 另一个是由 14 名 ms 患者组成的公开数据集, 称为 wbi 2015 纵向 ms 病变分割挑战数据集。考虑到 isbi 的挑战, 在提交时, 我们的方法是性能最好的解决方案之一。在私有数据集上, 使用与 isbi 挑战相同的性能指标数组, 与其他公开可用的工具相比, 该方法在 ms 病变分割方面有了很高的改进。少

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

## 2. 用渔民区分性 3ncn cnn 成像自动评估心脏 cine 磁共振成像中的全左心室覆盖

**作者:**le zhang, ali gooya, marco pereanez, bo dong, stefan k. piechnik , stefan neubauer , stefen e.petersen, alejandro f. frangi

**摘要:** 心脏磁共振 (cmr) 图像在心血管疾病的诊断成像中起着越来越大的作用。从基部到先端的左心室 (lv) 的全覆盖是 cmr 图像质量的基本标准, 也是准确测量心脏体积和功能评估所必需的。lv 覆盖不完全是通过目视检查确定的, 这很耗时, 通常在评估大型成像队列时进行追溯性检查。本文提出了一种利用渔民判别三维卷积神经网络 (cnn) 自动确定 cmr 图像 lv 覆盖的新方法。与我们以前采用 2d cnn 的方法不同, 这种方法利用 cmr 卷中的空间上下文信息, 提取更具代表性的高级特征, 并增强基线的判别能力 2d cnn 学习框架, 从而实现卓越的检测精度。提出了一个两阶段框架, 以确定在 cmr 体积测量中缺失的基部和顶端切片。首先, fd3d cnn 从 cmr 堆栈中提取高级功能。然后使用这些图像表示来检测缺失的基部和顶端切片。与传统的 3d cnn 策略相比, 拟议的 fd3d cnn 最大限度地减少了类内散射, 并最大限度地提高了类间的分散。我们在英国生物库研究的 5000 多项独立体积 cmr 扫描中进行了广泛的实验验证, 实现了缺失的基础切片检测的低错误率 (4.9\%/4.6%)。该方法也可用于评估其他类型 cmr 图像数据的 lv 覆盖范围。少

2018 年 11 月 8 日提交;v1 于 2018 年 11 月 6 日提交;最初宣布 2018 年 11 月。

### 3. 同质空间上的等价 cnn 的一般理论

作者:taco cohen, mario geiger, maurice weiler

**摘要:** 群等变卷积神经网络 (g-cnn) 最近已成为一个非常有效的模型类, 用于在已知对称性的背景下从信号中学习。对于二维和三维欧几里得空间、图形和球体上的信号, 提出了各种各样的等变量层, 很难看到所有这些方法是如何相关的, 以及它们是如何推广的。本文提出了一个相当通用的等变卷积网络理论。卷积特征空间被描述为同质基空间上的字段, 如平面  $\mathbb{R}^2$ 、球体  $S^2$  或图表  $S \text{ mathcal{Gn}}$ 。该理论能够根据所有现有的 gn-cnn 的对称性、基空间和场类型 (例如标量、矢量或张量场等) 对其进行系统分类。除了这种分类外, 我们还使用 mecey 理论来证明具有等变核的卷积是这些领域之间最一般的等变量映射类, 从而将 g-nnn 确立为等价网络的一类通用类。该理论还解释了如何对等变核的空间进行参数化以进行学习, 从而简化了 g-nn 在新空间和对称性方面的开发。最后, 该理论引入了丰富的几何语义来学习特征空间, 从而提高了深部网络的可解释性, 并与数学和物理学的中心思想建立了联系。少

2018 年 11 月 5 日提交;最初宣布 2018 年 11 月。

#### 4. spnet: 深 3d 对象分类和使用立体投影的检索

作者:mohsen yavartanoo, eu young kim, kyoung mu lee

**文摘:** 我们提出了一个有效的立体投影神经网络 (spnet) 的学习表示的 3d 对象. 我们首先使用立体投影将三维输入体积转换为二维平面图像。然后, 我们提出了一个浅二维卷积神经网络 (cnn) 来估计对象类别, 然后是视图集合, 它结合了来自对象多个视图的响应, 以进一步增强预测。具体而言, 该方法包括四个阶段: (1) 3d 对象的立体投影, (2) 特定于视图的特征学习, (3) 视图选择和 (4) 视图集成。该方法可与最先进的方法相比较, 同时具有极低的 gpu 内存和网络参数。尽管其轻巧性,但三维目标分类和形状检索实验证明了该方法的高性能。少

2018 年 11 月 5 日提交;最初宣布 2018 年 11 月。

#### 5. 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+ mn 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;**<sup>v1</sup> 于 2018 年 11 月 5 日提交;**最初宣布** 2018 年 11 月。

## 6. 利用卷积神经网络在肺计算机层析成像中的假阳性约

**作者:**[gorkem polat](#), [ugur halici](#) , [yesim serinagaoglu dogrusoz](#)

**摘要:** 最近的研究表明, 与传统的胸部 x 线摄影相比, 使用每年低剂量计算机断层扫描 (ct) 进行肺癌筛查可将肺癌死亡率降低 20%。因此, ct 肺筛查已开始在世界各地广泛使用。然而, 分析这些图像是放射科医生的一个沉重负担。在本研究中, 我们提出了一个新的和简单的框架, 分析 ct 肺检查使用卷积神经网络 (cnn), 并减少误报。我们的框架表明, 与传统方法相比, 即使是非复杂的体系结构也非常强大, 可以对 3d 结核数据进行分类。我们也使用不同的融合, 以显示他们的力量和对总分的影响。3d cn

比 2d cn 更受欢迎, 因为数据是 3d 的,而 2d 卷积操作可能会导致信息丢失。为了克服班级不平衡, 使用了小批量。拟议的框架已根据 luna16 挑战评估得到验证, 得分为 0.786, 即 7 个预定义假阳性 (fp) 点的平均灵敏度值。少

2018 年 11 月 4 日提交;最初宣布 2018 年 11 月。

## 7. 基于 cnn 的心脏 mr 运动文物自动检测

作者:lkay oksuz, bram ruij 绒什, esther puyol-anton, james clough, gastao cruz, aurelien bustin, claudia prieto, rene botnar, daniel rueckert, julia a. schnabel , andrew p. king

文摘: 医学图像的良好质量是后续图像分析管道成功的先决条件。因此, 医学图像的质量评估是一项重要活动, 对于大型人口研究, 如英国生物银行 (ukbb), 人工识别文物, 如意外运动造成的文物是繁琐和耗时的。因此, 迫切需要自动图像质量评估技术。本文提出了一种自动检测心脏磁共振 (cine) cine 图像中存在运动相关文物的方法。我们比较了两种深度学习架构, 对质量差的 cmr 图像进行了分类: 1) 三维时空卷积神经网络 (3d-tnn), 2) 长期经常卷积网络 (lrcn)。虽然在实际的临床设置运动文物是常见的, 高品质的成像 ukbb, 其中包括跨部门的人口数据的志愿者谁不一定有健康问题造成一个高度不平衡的分类问题。由于与相对较少的运动文物图像相比, 高质量图像的数量较多, 我们提出了一种基于 k 空间合成人工制品创建的新的数据增强方案。我们还研究

了一种基于合成人工制品严重程度的预定课程的学习方法。我们在英国生物库数据集的一个子集上评估我们的管道, 该数据集由 3510 cmr 图像组成。lrcn 架构的性能优于 3d-cnn 架构, 能够在不到 1 毫秒的高召回率下检测到 2d+ 时间短轴图像.我们将我们的方法与一系列最先进的质量评估方法进行比较。新的数据扩充和课程学习方法都提高了分类性能, 在 0.89 的情况下实现了整体面积。少

2018 年 10 月 29 日提交;最初宣布 2018 年 10 月。

## 8. 改进细粒度对象类别三维路径数据集的注释

作者:王亚明、谭晓、杨毅、李子玉、刘晓丽、周峰、拉里·s·戴维斯

**摘要:** 现有的对象类别的 3d 姿势数据集仅限于泛型对象类型, 并且缺乏细粒度信息。在本工作中, 我们引入了一个新的大型数据集, 其中包括 409 细粒度类别和 31 881 张图像, 并带有精确的 3d 姿态注释。具体而言, 我们通过从 shapenet 中为每个子类别查找特定的 3d 模型, 并通过以下方式手动对每个 2d 图像进行注释, 来增强三个**现有的细粒度**目标识别数据集 (stanfordcars、comprcar 和 fgvc-飞机)。调整全套 7 个连续透视参数。由于细粒度形状允许 3d 模型更好地适应图像, 因此我们通过从人工注释中初始化和对姿势参数进行局部搜索来进一步提高注释质量,



目的是最大限度地提高 iou 从最先进的深卷积神经网络 (cnn) 估计的投影掩码和分割参考之间。我们提供了注释的完整统计数据, 并进行了定性和定量比较, 表明我们的数据集可以成为研究 **三维姿态估计的补充来源**。数据集可以在 <http://users.umiacs.umd.edu/~wym/3dpose.html> 下载。少

**2018 年 10 月 18 日提交;最初宣布** 2018 年 10 月。

## 9. 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)**——现代视频理解中的核心内核。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 (且随着时间的推移而变化) 更大维 度。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, 并提高了性能/瓦, 比基线 **3d cnn cnn** 提高了 5.1 x (平均 4 倍) 加速器, 其面积开销为 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。

## 10. 多视点 x 射线 r-cnn

作者: [jan-martin o. steitz](#), [faraz saeedan](#) , [stefan roth](#)

**摘要:** 作为航空安全筛选的一部分, 我们在检测随身行李中违禁物体的基础上, 开发了一种基于 **cnn** 的多视点 x 射线图像数据目标检测方法。我们的贡献有两个方面。首先, 我们引入了一个新的多视图池层来执行从每个视图中提取的 **2d cnn** 特征的三维聚合。为此, 我们的池层利用成像系统的已知几何形状来确保要素聚合的几何一致性。其次, 我们引入了基于更快 **r-cnn** 的端到端可训练多视图检测管道, 该管道利用这些聚合的多视图功能, 推导出区域建议并在 **3d** 中执行最终分类。与单视图检测相比, 我们的方法显示出显著的精度提升, 而比在每个视图中执行单视图检测更有效。少

2018 年 10 月 4 日提交;最初宣布 2018 年 10 月。

## 11. 使用 rgb-d 相机进行态势识别: 探索三维人体建模和深度学习 方法

**作者:** [mohamed el amine elforaici](#), [ismail chaaraoui](#), [wassim bouachir](#), [youssef ouakrim](#), [neila mezghani](#)

**摘要:** rgb-d 传感器的出现为有效解决复杂的人工视觉问题提供了新的可能性。人体姿势识别是这些计算机视觉问题之一, 具有环境辅助生活和智能医疗系统等广泛的应用。在此背景下, 本文提出了利用 rgb-d 相机设计自动姿态识别系统的新方法和思路。更具体地说, 我们介绍了两种受监督的方法, 以学习和识别人类的姿势使用的主要类型的视觉数据提供了一个 rgb-d 相机。第一种方法是基于从二维图像中提取的卷积特征。卷积神经网络(cnn)受过训练, 可以使用 rgb 上的转移学习和深度图像来识别人类的姿势。其次, 我们建议利用三维空间中的身体关节配置来模拟姿势。然后通过 svm 对基于骨架的特征进行分类来实现态势识别。为了评估建议的方法, 我们创建了一个具有挑战性的姿态识别数据集, 在采集条件方面具有相当大的差异。实验结果表明, 两种方法在识别人体姿势方面具有可比的性能和较高的精度, 在深度图像上应用时, 基于 cnn 的方法略微优于这两种方法。此外, 这两种方法对尺度和方向变化等几个摄动因素表现出很高的鲁棒性。少

2018年10月12日提交;v1于2018年9月29日提交;最初宣布2018年10月。

## 12. 深卷积神经网络与手工制作特征相结合的预测肺结核恶性肿瘤

作者:李树龙,徐盘潘,李斌,陈丽元,周志国,洪霞,段英英,迈克尔·福克特,马建华,姜志强,王静

**摘要:** 为了预测具有高灵敏度和特异性的肺结节恶性肿瘤,我们提出了一种融合算法,该算法将手工制作的特征(hf)结合到三维深层卷积神经网络(cnn)输出层学到的特征中。首先,我们提取了29个手工制作的特征,包括9个强度特征、8个几何特征和12个基于灰度共现矩阵(glcmm)的纹理特征,这些特征来自13个方向。然后,我们训练3d cnn修改从三个最先进的2dcnn架构(亚历克网,vgg-16网络和多作物网),以提取美国有线电视新闻网的功能学到的输出层。对于每个3d cnn,美国有线电视新闻网的功能结合29个手工制作的特征被用作支持向量机(svm)的输入,再加上顺序正向特征选择(sfs)方法,以选择最佳的特征子集,构造分类器。融合算法充分利用了手工制作的功能和在输出层学习到的最高级别的cnn功能。它可以克服手工制作的特征的缺点,可能不能充分反映一个特定病变的独特特征结合在一起,通过结合美国有线电视新闻网的内在特征。同时,在手工制作特征互补的基础上,也降低了对cnn的大规模注释数据集的需求。患者群体包括从lidc/idri数据库中提取的431个恶性结节和795

个良性结节。对于每个被调查的 **cnn** 架构, 提出的融合算法在所有竞争分类模型中实现了最高的 auc、准确性、灵敏度和特异性得分。少

2018 年 9 月 7 日提交;最初宣布 2018 年 9 月。

### 13. 利用深卷积神经网络进行 4dct 通气成像

作者:钟运城,叶夫根尼·维诺格拉茨基,李元·陈,尼克·迈丘克,理查德·卡斯蒂略,爱德华·卡斯蒂略,托马斯·格雷罗,史蒂夫·江,王静

**摘要:** 目的: 功能成像已成为肺癌治疗规划和评价的重要工具。与传统方法相比, 如核医学通气灌注 (vq)、正电子发射断层扫描 (pet)、单光子发射计算机断层扫描 (spect) 或磁共振成像 (mri) 等, 这些方法使用造影剂形成 2d 或三维功能图像, 从 4dct 肺图像获得的通气成像是方便和具有成本效益的, 因为它在放射治疗规划可用。目前从 4dct 肺图像中获取通气图像的方法涉及变形图像配准 (dir) 和基于密度变化的算法 (dir/hu);因此, 所产生的通风图像对 dir 算法的选择非常敏感。方法: 提出一种基于深卷积神经网络 (**cnn**) 的方法, 在没有显式 dir 的情况下直接从 4dct 中提取通气图像, 从而提高通风图像的一致性和准确性。采用该方法对肺癌患者共 82 套 4dct 和通气图像进行了研究。结果: 预测图像与测试数据的标签图像相当。在 10 倍交叉验证中, 相似

指数和相关系数的平均值分别为  $0.883 \pm 0.034$  和  $0.883 \pm 0.034$ 。结论: 深 **cnn** 能在没有明确变形图像配准的情况下, 从 4dct 中生成通气影像, 减少相关的不确定性。少

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

#### 14. 利用机器学习加速可拉伸石墨烯 kirigami 的搜索和设计

**作者 :** paul z. hanakata, ekin d. cubuk , david k. campbell, harold s . park

**摘要:** 将 kirigami 风格的切割制作成板材已被证明是设计具有变质特性的可拉伸材料的有效方法, 在这种材料中, 2d 形状可以转换为复杂的 3d 形状。但是, 找到最佳解决方案并不简单, 因为可能的切割模式的数量会随着系统大小的增加而呈指数级增长。在这里, 我们报告了如何使用机器学习 (ml) 来近似目标属性, 如屈服应力和屈服应变, 作为切割模式的函数。我们的方法能够快速发现通过分子动力学 (md) 模拟验证的具有极高伸缩性的 kirigami 设计。我们发现, 在视觉任务中常用的卷积神经网络 (**cnn**) 可用于回归, 以达到接近 md 模拟精度的精度。然后, 这种方法可用于搜索最佳设计, 最大限度地提高弹性伸缩性, 在一个大的设计空间中, 只有 1000 个训练样本。或  $4 \times 10^6$  候选设计。这个例子展示了 ml 在寻找最佳 kirigami 设计的力量和潜力, 在这种情况下, 纯粹的 md 或基于实验的方法需要的迭代的

一小部分，而在这种情况下，对可管理物理的了解是未知的或可用的。少

2018 年 8 月 18 日提交;最初宣布 2018 年 8 月。

## 15. 神经身体拟合: 深层学习与基于模型的人体姿势和形状估计的统

—

作者:mohamed omran, christoph lassner, gerard pons-moll , peter v.gehler, bernt schiele

**摘要:** 即使对于高度参数化的深度学习模型，直接预测三维车身姿势和形状仍然是一个挑战。从二维图像空间到预测空间的映射是困难的: 透视歧义使得损失函数变得嘈杂，训练数据稀缺。本文提出了一种新的方法（神经体拟合 (nbf)）。它在 **cnn** 中集成了一个统计体模型，利用可靠的自下而上的语义身体部分分割和可靠的自上而下的身体模型约束。nbf 是完全可区分的，可以使用 **2d** 和 **3d** 注释进行训练。在详细的实验中，我们分析了模型的组件如何影响性能，特别是使用部分分段作为显式中间表示，并为 **3d** 人体姿势提供了一个强大、高效的训练框架从 **2d** 图像中估计，并在标准基准上获得有竞争力的结果。代码将以 [http://github.com/mohomran/neural\\_body\\_fitting](http://github.com/mohomran/neural_body_fitting) 的方式提供

2018 年 8 月 17 日提交;最初宣布 2018 年 8 月。

## 16. 基于平行框架的三维曲面的卷积神经网络

作者:潘浩,刘士林,刘洋,新通

**文摘:** 我们将卷积神经网络 (cnn) 扩展到嵌入在三维欧几里得空间中的曲面上, 这些曲面被离散为不规则网格, 并被广泛用于表示几何数据在计算机视觉和图形。我们定义了曲面域切线空间上的表面卷积, 其中卷积具有两个理想的属性: 1) 在投影到切线空间时, 表面域信号的畸变在局部最小; 2) 平移等方差属性通过将切线空间与保留度量值的规范并行传输对齐, 在本地保持。在计算方面, 我们依赖于曲面上的一个平行  $n$  方向帧场, 该帧场最大限度地减少了场变化, 因此尽可能兼容并近似于并行传输。在配有平行框架的切线空间上, 表面卷积的计算成为标准的常规。框架具有旋转对称性, 通过构造平行框架引起的曲面覆盖空间并将特征图相应地分组为  $n$  组来消除其歧义; 卷积是在覆盖空间的  $n$  分支上计算的, 并使用各自的要素图进行计算, 同时共享内核权重。为了在共享核权重的同时处理离散网格的不规则点, 我们做了卷积半离散函数, 即卷积核是多项式函数, 它们与离散曲面点的卷积变成了采样和加权求和。池和解共操作是沿着通过简化构建的网格层次结构计算的。所呈现的**表面 cnn** 允许在网格上进行有效的深度学习。我们表明, 对于分类、分割和非刚性配准任务, 仅使用原始输入信号的表面 **cnn** 比以前使用复杂输入功能的模型具有卓越的性能。少

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



## 17. 从结构脑图像预测年龄的领域引导下的 cnn 体系结构

作者:[pascal sturmfels](#), [saige rutherford](#), [mike angstadt](#), [mark peterson](#), [chandra sripada](#), [jenna wiens](#)

**摘要:** 考虑到卷积神经网络 (cnn) 应用于自然图像的广泛成功, 研究人员已经开始将其应用于神经成像数据。然而, 到目前为止, 对专门针对神经成像数据的新的 cnn 架构的探索有限。最近的几项作品未能利用大脑的三维结构, 而是将大脑视为一组独立的 2d 切片。利用三维卷积的方法依赖于为自然二维图像中的目标识别任务而开发的体系结构。这样的架构对神经成像可能无法保持的输入做出假设。例如, 现有的体系结构假定大脑中的模式表现出平移不变性。然而, 大脑中的模式可能有不同的含义, 这取决于它在大脑中的位置。有必要探索适合大脑图像的新架构。我们提出了两个简单的修改现有的美国有线电视新闻网架构的大脑图像结构。应用于大脑年龄预测的任务, 我们的网络实现了 1.4 年的平均绝对误差 (mae), 训练速度比 cnn 基线快 30%, 达到了 1.6 岁的 mae。我们的研究表明, 从开发自然图像模型中吸取的经验教训可能不会直接转移到神经成像任务中。相反, 在这一领域的模式开发方面仍有大量未探讨的问题, 其答案可能不同于传统智慧。少

2018 年 8 月 11 日提交;最初宣布 2018 年 8 月。

## 18. 三维深度卷积: 减少 3d 视觉任务中的模型参数

作者:叶荣天,刘方宇,张立强

**摘要:** 与 2d 卷积操作相比, 标准的 3d 卷积操作所需的内存和计算成本要大得多。这一事实阻碍了深度神经网络在许多 3d 视觉任务中的发展。本文研究了在三维场景中应用深度可分离卷积的可能性, 并介绍了三维深度卷积的应用。三维深度卷积将单个标准的 3d 卷积拆分为两个独立的步骤, 这将大大减少 3d 卷积中具有多个数量级的参数数量。我们在流行的 cnn 架构上进行了三维深度卷积实验, 并将其与类似的结构称为伪三维卷积进行了比较。结果表明, 在三维深度卷积的情况下, 可以利用更多的轻量化神经网络来完成三维视觉任务, 如分类和重建, 同时还能提供可比较的性能。少

2018 年 8 月 4 日提交;最初宣布 2018 年 8 月。

## 19. 基于距离矩阵的三维人体姿态估计中的遮挡关节恢复

作者:郭翔,戴玉超

**摘要:** 尽管卷积神经网络在单图像三维人体姿态估计方面取得了新的进展, 但处理高遮挡场景等真实场景仍面临着挑战。本文提出利用欧几里得距离矩阵 (edm) 来解决单图像三维人体姿态估计的问题。具体而言, 我们提出了两种基于 edm 的方法, 它可以有效地处理二维图像中的遮挡关节。第一种方法是基于 2d 到 2d 距离矩阵回归, 通过一个简单的 cnn 架构实现。第二种方法是基

于稀疏编码以及一个学习过完整的字典。在人脸 3.6 m 数据集上的实验表明,这两种方法在恢复闭塞观测方面具有优异的性能,并证明了具有遮挡关节的三维人体姿态估计的准确性提高。少

2018 年 7 月 29 日提交;最初宣布 2018 年 7 月。

## 20. 基于暹罗 cnn 的多模态遥感数据的变化检测

作者:张振超,乔治·沃塞曼,马库斯·格尔克, [devis tuia](#), [michael ying yang](#)

摘要: 城市环境地形变化的检测一直是城市规划和监测的重要任务。实际上,遥感数据往往以不同的方式和不同的时间提供。多模数据之间的变化检测可能非常具有挑战性,因为数据表现出不同的特征。考虑到三维激光扫描点云和不同时代的二维图像,本文提出了一个检测建筑物和树木变化的框架。首先,将 2d 和 3d 数据分别转换为图像补丁。随后,一个暹罗美国有线电视新闻网被用来检测这两个时代之间的候选变化。最后,根据候选对象的更改对基于补丁的更改进行分组和验证。对城市数据的实验表明,该模型可以正确地对 86.4%——补丁对进行分类。少

2018 年 7 月 25 日提交;最初宣布 2018 年 7 月。

## 21. 基于远程二维语境的肿瘤分割三维卷积神经网络

作者 :[pawel mlynarski](#), [hervé delingette](#), [antonio 刑事](#), [nicolas ayache](#)

**文摘:** 针对多序列 mr 图像中肿瘤分割的挑战性任务, 提出了一种有效的深层学习方法。近年来, 卷积神经网络 (cnn) 在医学成像的各种识别任务中取得了最先进的性能。由于 cnn 的计算成本相当大, mri 等大容量通常由子卷处理, 例如切片 (轴向、冠状、矢状) 或小型 3d 补丁。本文介绍了一种基于 cnn 的模型, 该模型有效地结合了短距离三维上下文和远程 2d 上下文的优点。为了克服神经网络体系结构特定选择的局限性, 我们还建议通过体外投票策略将多个级联 2d-3d 模型的输出合并。此外, 我们还提出了一种由单独的子网处理不同 mr 序列的网络体系结构, 以便对 mr 序列缺失问题更加稳健。最后, 介绍了一种简单有效的大型 cnn 模型训练算法。我们根据 brats 2017 挑战的公共基准对恶性脑肿瘤的多类分割任务进行了评估。我们的方法取得了良好的性能, 并产生准确的分割, 中位骰子分数为 0.918 (全肿瘤), 0.918 (肿瘤核心) 和 0.918 (增强核)。我们的方法可以自然地应用于各种任务涉及病变或器官的分割。少

2018 年 7 月 23 日提交;最初宣布 2018 年 7 月。

## 22. 使用 2d 卷积网络将三维形状作为多层测地图进行学习

**作者:** [kripasindhu sarkar](#), [basavaraj h 普奇霍利](#), [kiran varanasi](#), [didier stricker](#)

**文摘:** 我们提出了一个新的全局表示的 3d 形状, 适合 2d cnn 的应用。我们将 3d 形状表示为多层高度图 (mlh), 在每个网格位置,

我们存储多个高度贴图实例，从而表示隐藏在多层遮挡后面的 3d 形状细节。为合并视图相关信息提供了一种新的视图合并方法。mlh 描述符)，从多个视图。由于使用 2d cnn 的能力，与基于体素的输入相比，我们的方法在输入分辨率方面具有很高的内存效率。结合 mlh 描述符和我们的多视图合并，我们在建模网络数据集上实现了最先进的分类结果。少

2018 年 7 月 26 日提交;v1 于 2018 年 7 月 23 日提交;最初宣布 2018 年 7 月。

## 23. 卷积神经网络在医学图像分割中的应用时机不确定性估计

作者:王国泰,李文奇,迈克尔·阿尔特森, 简·德普斯特特, 塞巴斯蒂安·乌尔塞林,汤姆·韦考特伦

**摘要:** 尽管在医学图像分割方面具有最先进的性能，但深度卷积神经网络 (cnn) 很少提供有关其分割输出的不确定性估计，例如模型 (认知) 和基于图像的输出 (不确定性)。在本工作中，我们分析了这些不同类型的不确定性为基于 cnn 的 2d 和 3d 医学图像分割任务。此外，我们还提出了一个基于测试时间增强的高度不确定性来分析输入图像的不同变换对分割输出的影响。测试时间扩充以前曾被用来提高分割精度，但不是在一致的数学框架内制定的。因此，我们还提出了一个测试时间增强的理论公式，其中预测的分布是通过蒙特卡罗模拟估计的，在一个涉及图像转换的图

像采集模型中的参数的先验分布和噪音。我们将提出的算法不确定性与模型不确定性进行了比较和结合。从二维和三维磁共振图像 (mri) 对胎儿大脑和脑瘤进行分割的实验表明, 1) 基于时间增强的高度不确定性提供了比单独计算基于测试时间的模型不确定性, 并有助于减少过度自信的不正确预测, 2) 我们的测试时间增强优于单预测基线和基于下拉的多个预测。少

2018 年 7 月 20 日提交;v1 于 2018 年 7 月 19 日提交;最初宣布 2018 年 7 月。

## 24. 基于草图的三维形状检索的语义保留对抗性学习的深度跨模适应

作者:陈嘉新,易芳

文摘: 由于 2d 草图和 3d 形状之间存在较大的跨模式差异, 因此通过草图检索 3d 形状是一项极具挑战性的任务。针对这一问题, 本文提出了一个新的框架, 以学习一个判别的深层跨模适应模型。具体而言, 我们首先采用两个度量网络, 遵循两个深卷积神经网络 (cnn), 以学习基于重要感知度量学习方法的建模特定判别特征。随后, 我们明确引入了一个多模变换网络来补偿两种模式之间的差异, 它可以将二维草图的特征传输到三维形状的特征空间。我们开发了一种基于对抗学习的方法来训练转换模型, 同时增强两种模式的数据分布之间的整体相关性, 并通过最小化最小化来

缓解局部语义差异。跨方式均值差异项。在 shrec 2013 和 shrec 2014 数据集上的实验结果清楚地表明, 与最先进的方法相比, 我们提出的模型具有卓越的检索性能。少

2018 年 7 月 4 日提交;最初宣布 2018 年 7 月。

## 25. 基于三维上下文增强的基于区域的卷积神经网络, 用于端到端病变检测

作者:ke yan, mohammadhadi bagheri, ronald m. summers

**摘要:** 从计算机断层扫描 (ct) 扫描中检测病变是一个重要但困难的问题, 因为非病变和真正的病变可能会出现类似的情况。众所周知, 3d 上下文在这一差异化任务中很有帮助。然而, 现有的卷积神经网络 (cnn) 的端到端检测框架大多是为二维图像而设计的。本文提出了基于 cnn(3dce) 的三维上下文增强 cnn (3dce), 通过聚合二维图像的特征图, 有效地整合三维上下文信息。3dce 在培训和推理方面易于培训和端到端。开发了一种通用病变检测器, 利用深度病变数据集在一个算法中检测各种病变。这一具有挑战性的任务的实验结果证明了 3dce 的有效性。

我 们 在  
[https://github.com/rsummers11/CADLab/tree/master/lesion\\_detector\\_3DCE](https://github.com/rsummers11/CADLab/tree/master/lesion_detector_3DCE) 发布了 3dce 代码。少



2018年7月29日提交;v1于2018年6月25日提交;最初宣布2018年6月。

## 26. 面具形态: 一种基于深度学习的人体形态重建--从二元面膜图像看

作者:季忠平,肖琪,王一刚,徐刚,杜鹏, 吴青

**摘要:** 3d 内容创建被称为计算机图形最基本的任务之一。在过去的几十年里,许多来自二维图像或曲线的三维建模算法已经开发出来。设计人员可以从前面、侧面和顶部视图对齐一些概念图像或绘制一些提示曲线,然后在自动或手动构建 3d 模型时使用它们作为参考。然而,据我们所知,没有任何研究以类似的方式对三维人体重建进行调查。本文提出了一种基于三维人体形状的深度学习重建方法。设计了一个新的基于 cnn 的回归网络,它的两个分支分别对应于正面和侧面视图,用于从 2d 掩模图像中估计三维人体形状。我们分别训练我们的网络,以分离功能描述符,这些特征描述符从不同的视图中编码身体参数,并将它们融合在一起,以估计准确的人体形状。此外,为了克服这方面所需的培训数据不足的问题,我们提出了一些三维人体形状的显著数据增强方案,可用于促进对这一主题的进一步研究。大量的实验结果表明,利用我们的算法可以有效地实现视觉逼真和准确的重建。只需要二进制面膜图像,我们的方法可以帮助用户快速创建自己的数字化

身,也可以很容易地创建数字人体的 3d 游戏,虚拟现实,在线时尚购物。少

2018 年 6 月 22 日提交;最初宣布 2018 年 6 月。

## 27.用于动作分类的时空信道相关网络

作者 :ali diba, mohsen fayyaz, vivek sharma, m. mahdi arzani, rahman yousefzadeh, juergen gall, luc van gool

文摘: 本文的工作是由三维卷积神经网络 (cnn) 是否足够时空相关性的问题所驱动的? 大多数传统的 3d 网络使用局部时空特征。我们引入了一个新的块, 它模拟 3dcnn 频道之间在时间和空间特征方面的相关性。这个新的块可以作为一个剩余单位添加到 3d cnn 的不同部分。我们把我们的小说块命名为 "时空信道相关性" (stc)。通过将此块嵌入到当前最先进的体系结构 (如 resnext 和 resnet) 中, 我们将动力学数据集的性能提高了 2-3%。我们的实验表明, 在当前最先进的体系结构中添加 stc 块优于 hmdb51、ucf101 和动力学数据集上最先进的方法。在培训 3dcnn 的另一个问题是如何从零开始训练他们与一个巨大的标记数据集, 以获得合理的性能。因此,在 2d cn 中学到的知识被完全忽视。这项工作的另一个贡献是一个简单而有效的技术, 将知识从预先训练的 2d cnn 转移到随机初始化的 3d cnn, 以实现稳定的重量初始化。这使我们能够显著减少 3d cnn 的训练样本数量。因此, 通过微调该网络, 我们击败了 3d cnn 中通用方法和最新方法的

性能, 这些方法在大型视频数据集 (例如 sports-1m) 上接受了培训, 并对目标数据集 (例如 hmdb5cf101) 进行了微调。少

2018 年 6 月 25 日提交;v1 于 2018 年 6 月 19 日提交;最初宣布 2018 年 6 月。

## 28. 使用基于修补程序的迭代网络实现快速多重地标定位

作者: 李元伟, amir alansary, juan j.cerrolaza, bisesh khalal, matthewsinclair, jacqueline matthew, chandni gupta, caroline knight, bernhard kainz, 丹尼尔·鲁克特

**摘要:** 我们提出了一个新的基于修补程序的迭代网络 (pin), 用于在 3d 医疗卷中快速准确地定位地标。pin 利用卷积神经网络 (cnn) 来了解图像贴片和解剖地标位置之间的空间关系。在推断过程中, 补丁被反复传递给美国有线电视新闻网,直到估计的地标位置收敛到真正的地标位置。pin 在计算上是有效的, 因为推理阶段只是以迭代的方式选择性地采样少量补丁, 而不是在卷中的每个位置进行密集采样。我们的方法采用多任务学习框架, 将回归和分类结合起来, 以提高本地化的准确性。我们通过使用主成分分析扩展 pin 以定位多个地标, 主成分分析对地标之间的全球解剖关系进行建模。我们使用 72 张胎儿筛查检查的 3d 超声图像对 pin 进行了评估。pin 在数量上实现了 5.59 mm 的平均地标定位误差和 0.44 s 的运行时间, 以预测每卷 10 个地标。从质量上, 从预测的地标位置衍生的解剖二维标准扫描平面在视觉上

与临床地面真相相似。源代码可在 <https://github.com/yuanwei1989/landmark-detection> 公开。

少

**2018 年 10 月 6 日提交;**v1 于 2018 年 6 月 18 日提交;**最初宣布** 2018 年 6 月。

## 29. 从单个深度图像中完成语义场景的视图体积网络

**作者:**郭玉晓,新通

**文摘:** 我们引入了一个视量卷积神经网络 (wnet), 用于从单个深度图像推断体积**三维场景**的占用率和语义标签。wnet 将 **2d** 视图 cnn 和 **3d** 卷 cnn 连接在一起, 具有可微的投影层. 在给定单个 rgb-d 图像的情况下, 我们的方法利用 **2d** 视图 **cnn** 从输入深度图像中提取详细的几何特征, 然后通过投影图层根据输入深度映射将特征投影到 **3d** 卷中。之后, 我们学习场景的 **3d** 上下文信息与 **3d** 卷 **cnn** 计算结果体积占用和语义标签。wnet 结合了 **2d** 和 **3d** 表示, 有效地降低了计算成本, 实现了从多通道高分辨率输入中提取特征, 从而显著提高了结果精度。我们验证了我们的方法, 并在合成 song 和真正的 nyu 数据集上演示了它的效率和有效性。少

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

### 30. 细粒度对象类别的三维姿态估计

作者:王亚明、谭晓、杨毅、刘小丽、丁二瑞、周峰、拉里·s·戴维斯

**摘要:** 现有对象姿势估计数据集与泛型对象类型相关, 到目前为止还没有细粒度对象类别的数据集。在本工作中, 我们引入了一个新的大型数据集, 以基准细粒度对象的姿态估计, 这得益于最近 2d 和 **3d** 细粒度数据的可用性。具体来说, 我们通过为每个子类别找到细粒度的 3d cad 模型, 并使用 **3d** 姿势手动在图像中对每个对象进行注释, 从而增强了两个流行的细粒度识别数据集 (stanfordcars 和 comprcar)。我们表明, 有了足够的训练数据, 就可以单独使用二维外观信息来估计具有连续参数的完整透视模型。我们通过基于 faster/掩码 r-inn 的框架来实现这一目标。这超出了以前关于分类级姿态估计的工作, 它只估计离散连续视点角度或恢复旋转矩阵往往在关键点的帮助下。此外, 利用细粒度的 **3d** 模型, 我们将一个名为定位字段的密集 **3d** 表示形式集成到基于 **cnn** 的姿势估计框架中, 以进一步提高性能。新的数据集可 [www.umiacs.umd.edu/~wym/3dpose.html](http://www.umiacs.umd.edu/~wym/3dpose.html)

**2018 年 11 月 7 日提交;**v1 于 2018 年 6 月 11 日提交;**最初宣布** 2018 年 6 月。

### 31. 基于多目标卷积神经网络的看不见机器人检测与联合估计转移学习

作者 :justinas miseikis, inka brijack , saeed yahyanejad, kirreglte , ole jakob elle, jim torresen

**摘要:** 使用深度学习技术的一个重大问题是可用于培训的数据有限。有一些数据集可用于项目识别和分类或自驾游等热门问题,但对于工业机器人领域来说,却非常有限。在前面的工作中,我们训练了一个多目标卷积神经网络 (cnn), 只需使用二维图像就能识别图像中的机器人身体, 并估计关节的三维位置, 但它仅限于一系列机器人由通用机器人 (ur) 生产。在这项工作中, 我们扩展了我们的方法, 以工作与一个新的机器人手臂-库卡 lbr iwa, 它有一个显着不同的外观和一个额外的关节。然而, 我们没有再次收集大型数据集, 而是收集了一些较小的数据集, 每个数据集包含几百帧, 并在美国有线电视新闻网训练的 ur 机器人上使用传输学习技术, 使其适应具有不同形状的新机器人和视觉特征。我们已经证明, 转移学习不仅适用于这一领域, 而且需要更小的精心准备的培训数据集, 培训速度显著加快, 达到与原始方法相似的精度, 甚至在某些方面进行了改进。少

2018 年 6 月 11 日提交;v1 于 2018 年 5 月 30 日提交;最初宣布 2018 年 5 月。

### 32. 卷积神经网络对脊髓和髓内多发性硬化症病变的自动分割

**作者:**charley gros, benjamin de leener, atef badji, josefina maranzano, dominique eden, sara m.dupont, jason talbott, ren zolquong, yaou liu,tobias granberg, russell ouellette, yasuhiko tachibana, masaakihori, kouhei kamiya, lydia chougar, leszekstawiarz , jan hillert, elise bannier, anne kerbrat, gilles edan, pierre labauge, v 地 iniecallot, jean pelletier, bertrand audoin, henitsoa razoanandrianina, et al .(27)未显示其他作者)

**摘要:** 多发性硬化症 (ms) 患者的脊髓经常受到萎缩和病变的影响。从 mri 数据中分割脊髓和病变提供了损伤的测量方法, 这些措施是 ms 诊断、预后和纵向监测的关键标准。大吞吐量分析管道的效率。跨多站点脊髓数据进行可靠和可靠的分割具有挑战性, 因为与采集参数和图像伪影相关的可变性很大。本研究的目的是开发一个全自动框架, 对图像参数和临床条件的变异性都具有鲁棒性, 用于从常规 mri 数据中分割脊髓和髓内 ms 病变。这项多部位研究包括了对 1, 042 名受试者 (459 名健康对照组、471 名 ms 患者和 112 名其他脊柱疾病患者) 的扫描 (n=30)。数据跨越三个对比 (t1-、t2-和 t2\* 加权), 共 1, 943 卷。提出的脐带和病变自动分割方法是基于两个卷积神经网络 (cnn) 序列。为了处理与体积其他部分相比脊髓和/或病变体素的很小比例, 第一个带有 2d 扩张凸形的 cnn 检测脊髓中心线, 然后是第二个 3nn cnn 与 3d 将脊髓和或病变分割在一起的卷积。与手动分割相比, 我们基于 cnn 的方法显示, propseg (一种最先进的脊髓分割方法) 的中位 dice 为 95%, 而 propseg 的中位 dice 为 88%。关于 ms 数据上的病变分割, 我们的框架提供了 60% 的 dice, 相对



体积差-15%，和平均检测灵敏度和精度分别为 83% 和 77%。拟议的框架是开源的，可在脊髓工具箱中随时获得。少

**2018 年 9 月 11 日提交;**v1 于 2018 年 5 月 16 日提交;**最初宣布** 2018 年 5 月。

### 33. 通过深度学习，利用广域荧光显微镜进行快速三维细胞跟踪

**作者:**[刘干](#),[乔惠惠](#),[吴嘉民](#),[王浩谦](#),[鲁芳](#),[戴琼海](#)

**摘要:** 高速三维跟踪细胞在许多生物医学应用中继续受到广泛关注，例如监测免疫细胞迁移和观察流动血管中的肿瘤转移。在这里，我们提出了一种基于深卷积神经网络（cnn）的方法，从传统广域荧光显微镜捕获的单个 2d 图像中检索荧光体的三维位置，而不使用任何图像。硬件修改。所报告的方法通过两个级联将具有挑战性的三维定位从一个不合适的基于模型的拟合问题，特别是在样本密集、信噪比低的情况下，转化为可解的多标签分类问题 cnn，深度学习技术比其他算法有很大的优势。与传统的核拟合方法相比，该方法在更大的轴向范围内实现了更准确、更鲁棒的多目标定位，并通过三维仿真和实验结果进行了验证分布荧光珠。此外，在体内 3d 跟踪斑马鱼的多个血细胞在 100 fps 进一步验证了我们的框架的可行性。少

**2018 年 5 月 15 日提交;**v1 于 2018 年 5 月 14 日提交;**最初宣布** 2018 年 5 月。

### 34. 在自动收费系统中, 采用稀疏 sfm 深度的深 2.5 d 车辆分类

作者: [georg waltner](#), [michael maurer](#), [thomas holzmann](#), [patrick ruprecht](#), [michael opitz](#), [horst possegger](#), [friedrich Fraundorfer](#), [horst bischof](#)

**摘要:** 自动收费系统依赖于对过往车辆进行适当的分类。当用于分类的图像只覆盖车辆的部分时, 这一点尤其困难。获取有关整车的信息。我们重建车辆作为 3d 对象, 并利用这些额外的信息在一个卷积神经网络 (cnn)。然而, 在使用深层网络进行 3d 对象分类时, 需要大量密集的 3d 模型来获得良好的精度, 由于内存需求, 这些模型往往无法进行处理。因此, 在我们的方法中, 我们使用重建的点、线或两者将 3d 对象重新投影到图像平面上。我们利用这个稀疏深度之前在一个辅助网络分支, 在训练中充当调节器。我们表明, 与真实世界数据集上的 2d 分类相比, 这种辅助规则化有助于提高精度。此外, 由于网络的设计, 在测试时只需要二维摄像机图像进行分类, 就可以在便携式计算机视觉系统中使用。少

2018 年 5 月 11 日提交;v1 于 2018 年 5 月 9 日提交;最初宣布 2018 年 5 月。

### 35. 增强人体行为识别的视觉属性增强三维卷积神经网络

作者:[王云峰](#),[周文刚](#),[张启林](#),[李厚强](#)

**摘要:** 单个视频帧中的可视属性（如特征对象和场景的存在）为视频中的操作识别提供了大量信息。以单个**二维**视频帧为输入, 与目前使用时空滤波器的**三维 cnn** 相比, 使用更复杂的卷积神经网络可以有效、高效地进行视觉属性提取, 由于 **2d cn** 中的参数较少。本文提出将视觉属性（包括检测、编码和分类）集成到多流 **3d cnn** 中, 用于修剪视频中的动作识别, 并对视觉属性进行了增强. **3d cnn** (a3d) 框架。可视化属性管道包括对象检测网络、属性编码网络和分类网络。我们提出的 a3d 框架在 hmdb51 和 ucf101 数据集上实现了最先进的性能。少

**2018 年 5 月 8 日提交;最初宣布** 2018 年 5 月。

### 36. 用黎曼几何计算 cnn 损耗和测距

**作者:** Hajnalhou, [nina miolane](#), [bisesk khalil](#), [matthew c. h. lee](#), [amir alansary](#), [steven mcdonagh](#), [jo v. hajnal](#), [daniel rueckert](#), [ben glocker](#), [bernhard kainz](#)

**摘要:** 姿态估计, 即预测与固定坐标框架有关的**三维**刚性变换,  $SE(3)$ , 是医学图像分析中普遍存在的问题, 应用包括: 图像刚性配准、解剖标准平面检测、跟踪和设备/摄像机姿态估计。深度学习方法通常用将旋转和平移分开的表示来参数化一个姿势。由于通常可用的框架不提供计算流形损失的方法, 回归通常是在旋转和平移的参数化上独立使用  $L_2$  范数进行的, 这是线性空间的一个度量, 不需要考虑到  $SE(3)$  的李群结构。本文提出了一种一般

的黎曼公式的姿态估计问题。我们建议直接在  $SE(3)$  上训练**美国有线电视新闻网**，配备左不变黎曼度量，将对定义姿势的平移和旋转的预测结合起来。在每个训练步骤中，地面真相和预测姿态都是流形的元素，在这些元素中，损失被计算为黎曼测地距离。然后，我们计算优化方向，方法是相对于流形  $SE(3)$  切线空间上的预测姿态反向传播梯度，并更新网络权重。我们通过将损失函数的性能与流行和最常用的现有方法（如基于图像的定位和基于强度的 **2d/3d** 注册）进行比较，彻底评估其有效性。我们还表明，在损耗函数中用于权衡旋转和转换之间的贡献的超参数可以从数据集中进行内部计算，以实现更大的性能优势。少

**2018 年 7 月 17 日提交**;v1 于 2018 年 5 月 2 日提交;**最初宣布** 2018 年 5 月。

### 37.通过潜在的 2.5 d 热图回归进行手的姿势估计

**作者:**[umar iqbal](#), [pavlo molchanov](#), [thomas breuel](#), [juergen gall](#), [jan kautz](#)

**摘要:** 估计手的**三维**姿势是人机交互的重要组成部分。随着计算机视觉的最新发展，使用深度或多视图传感器估计 **3d** 姿态变得更加容易，然而，从单个 rgb 图像中回归姿势就不那么简单了。主要的困难来自于 **3d** 姿势需要某种形式的深度估计，而只给出一个 rgb 图像时，深度估计是不明确的。本文提出了一种新的方法，从一个单一的眼睛图像的手的**姿势**估计通过一个新的 2.5 d 姿

势表示。我们的新表示估计构成一个缩放系数, 如果事先给出手的大小, 可以另外估计。我们隐式地学习深度地图和热图分布与一个新的 **cnn** 架构。我们的系统在多个具有挑战性的数据集上实现了对 **2d** 和 **3d** 手部姿势的最先进的估计, 并存在严重的遮挡。  
少

2018 年 4 月 25 日提交;最初宣布 2018 年 4 月。

### 38. 使深部热图对三维物体姿态估计的局部遮挡具有鲁棒性

作者:[markus oberweger](#), [mahdi rad](#), [vincent lepetit](#)

**文摘:** 介绍了一种在大遮挡条件下对单色图像进行鲁棒、准确三维物体姿态估计的新方法。按照最近的方法, 我们首先预测与目标对象相关的**三维点**的**二维**投影, 然后使用几何方法从这些对应计算三维姿态。不幸的是, 正如我们的实验结果所显示的, 使用普通的 **cnn** 或卷积姿势机预测这些 **2d** 投影对部分闭塞非常敏感, 即使这些方法是用部分训练的被遮挡的示例。我们的解决方案是独立预测来自多个小补丁的热图, 并积累结果以获得准确而可靠的预测。培训随后变得具有挑战性, 因为具有相似外观但在对象上的不同位置的补丁对应于不同的热图。然而, 我们提供了一个简单而有效的解决办法来处理这种模糊之处。我们展示了我们的方法在两个具有挑战性的数据集上优于现有方法: 遮挡的线 mod 数据集和 ycb 视频数据集, 这两个数据集都展示了具有高度遮

挡对象的杂乱场景。项目网站：  
<https://www.tugraz.at/institute/icg/research/team-lepetit/research-projects/robust-object-pose-estimation/>

2018年7月26日提交;v1于2018年4月11日提交;最初宣布2018年4月。

### 39. 用于医学图像分析的旋转-转换协积卷积网络

作者:erik j bekkers, maxime w lafarge, mitko veta , koen aj Eppenhof, josien pw pluim, remco du 在内

**抽象:** 我们提出了一个框架的旋转和翻译共变深度学习使用  $se(2)$  组卷。特殊欧几里得运动群的群产物  $se(2)$  描述了两个 roto 翻译的串联是如何导致净 roto 翻译的。我们将这种几何结构编码为卷积神经网络 (cnn), 通过  $se(2)$  组卷积层, 适合标准的 2d cnn 框架, 并允许一般处理旋转输入样本, 而无需增加数据。我们将引入三个层: 提升层, 将 2d (矢量值) 图像提升到  $se(2)$ -图像, 即其域为的 3d (矢量值) 数据  $se(2)$ ; 一个群卷积层, 从和到一个  $se(2)$ -形象; 和投影层。  $se(2)$ -图像到 2d 图像。提升和分组卷积层是  $se(2)$  协变 (输出转子与输入转换)。最终的投影层, 旋转时的最大强度投影, 使完整的 cnn 旋转不变。我们在组织病理学、视网膜成像和电子显微镜方面存在三个不同的问题, 在提出的 cnn 组中, 可以实现最先进的性能, 而不需要通过旋转和增加数据来增加数据。与标准 cn 相比, 标准的 cn 确实依赖于增加。少

2018 年 6 月 11 日提交;v1 于 2018 年 4 月 10 日提交;最初宣布 2018 年 4 月。

#### 40. 地板网: 从 3d 扫描重建平面图的统一框架

作者:陈柳, 吴嘉业,古河康高

**摘要:** 这项室内测绘研究的最终目标是通过在口袋里的智能手机穿过一栋房子, 自动重建平面图。本文提出了一种新型的深层神经网络—forenet, 解决了这一问题。挑战在于处理跨越大型 3d 空间的 rgb-d 流。floornet 通过三个神经网络分支有效地处理数据: 1) 具有 3d 点的 pointnet, 利用 3d 信息;2) cnn 具有自上而下的 2d 点密度图像, 增强了局部空间推理;3)美国有线电视新闻网与 rgb 图像, 利用完整的图像信息。floornet 在分支机构之间交换中间功能, 以利用所有体系结构中的最佳功能。我们已经建立了一个基准, 为 155 家住宅或公寓提供了 rgb-d 视频流与谷歌探戈手机, 并注释完整的平面图信息。我们的定性和定量评价表明, 三个分支的融合有效地提高了重建质量。我们希望, 该文件和基准将是解决具有挑战性的矢量图形重建问题的一个重要步骤。

代 码 和 数 据 可 在  
<https://github.com/art-programmer/FloorNet>。少

2018 年 3 月 30 日提交;最初宣布 2018 年 4 月。

#### 41. 诱导表示之间的相互作用 (及其在等价神经网络理论中的应用)



作者:taco s. cohen, mario geiger, maurice weiler

**文摘:** 组等变量和可操纵卷积神经网络 (规则和可操纵的  $g$ -nnn) 最近已成为一个非常有效的模型类, 用于学习信号数据, 如 2d 和 3d 图像、视频和其他数据在那里存在对称性。在几何方面, 常规的  $g$ -nnn 代表标量字段 ("要素通道") 方面的数据, 而可操纵的  $g$ -cnn 也可以使用矢量或张量字段 ("胶囊") 来表示数据。在代数术语中, 常规  $g$ -nnn 中的特征空间根据  $g$  组的正则表示进行变换, 而可操纵  $g$ -nnn 中的特征空间则根据  $g$  的一般诱导表示进行变换。为了使网络成为等价的,  $g$ -cnn 中的每一层都需要在与其输入和输出空间相关的感应表示之间交织在一起。本文提出了一个关于欧几里得空间或球体等齐次空间的  $gn$ -cn 的一般数学框架。我们用基本的方法表明, 如果并且只有当输入和输出特征空间根据诱导表示进行转换时, 等变量网络的层才是卷积的。这一结果是根据 g. w. mcmey 关于诱导表示的抽象理论得出的, 将 gnn 确立为一类通用的等变量网络体系结构, 并概括了 kondor & amp; trivedi 最近在规则表示之间的交织。少

2018年3月30日提交;v1于2018年3月28日提交;最初宣布2018年3月。

## 42. 用于立体匹配的级联多尺度多维卷积神经网络

作者:陆海华,徐海,张丽,赵勇

**文摘** 卷积神经网络 (cnn) 已被证明比传统的立体估计算法性能更好。许多工作集中在像素匹配成本计算上, 这是许多最先进的启动算法的重要组成部分。但是, 这些体系结构仅限于小的和单一的接受领域, 并使用传统方法进行成本聚合, 甚至忽略成本聚合。不同的是, 我们把它们都考虑进去了。首先, 我们提出了一种新的多尺度匹配成本计算子网, 其中两种不同大小的接受场是并行实现的。这样, 网络就能充分利用这两种变种, 平衡接受场的增加和细节的损失之间的权衡。此外, 我们还证明了**我们包含二维卷积和三维卷积操作的多维聚合子网**可以为估计准确的初始视差提供丰富的上下文和语义信息。最后, 在具有挑战性的立体声基准 kititi 上进行的实验表明, 即使没有任何额外的后处理, 该方法也能获得有竞争力的结果。少

**2018 年 4 月 17 日提交;**v1 于 2018 年 3 月 26 日提交;**最初宣布** 2018 年 3 月。

#### 43. 二维和三维几何通信的分层度量学习与匹配

**作者:**[mohammed e. fathy](#), [quoc-huy tran](#), [m. zeeshan zia](#), [paul vernaza](#), [manmohan chandraker](#)

**摘要:** 兴趣点描述符推动了计算机视觉中几乎所有问题的进展。深度神经网络的最新进展使特定于任务的学习描述符能够在许多问题上优于手工制作的描述符。我们证明, 常用的度量学习方法并不能最佳地利用在卷积神经网络 (cnn) 中学习的特征层次结构,

特别是在应用于几何特征匹配任务时。虽然适用于**美国有线电视新闻网**最深处的公制损耗，但无论任务如何，往往都会产生理想的特征，但事实上，不断增长的接受场以及条纹效应会导致更浅的特征在高精度匹配时表现得更好任务。我们利用这一见解以及在要素层次结构的多个级别进行明确监督，以实现更好的规范化，从而在几何匹配任务的上下文中了解更有效的描述符。此外，我们建议在 **cnn** 的不同层使用激活图，作为通常用于匹配任务的多分辨率图像金字塔的有效和有原则的替代品。我们提出了具体的 **cnn** 架构，利用这些想法，并评估他们在多个数据集上的 **2d** 和 **3d** 几何匹配以及光流，展示了最先进的结果和泛化跨越数据。少

**2018 年 8 月 1 日提交;****v1** 于 2018 年 3 月 19 日提交;**最初宣布** 2018 年 3 月。

#### 44. mo2cap2: 使用安装在顶层的鱼眼摄像机进行实时移动 3d 运动捕捉

**作者:**徐伟鹏, [Avishek chatterjee](#), [michael zollhoefer](#), [helge rhodin](#), [pascal fua](#), [hans-peter seidel](#), [christian theobalt](#)

**文摘:** 我们提出了在广泛的无约束日常活动中，以自我为中心的三维人体姿态估计的第一种实时方法。此设置具有一组独特的挑战，例如硬件设置的移动性，以及长捕获会话的健壮性，并可从跟踪失败中快速恢复。我们基于一种新颖的轻量级设置来应对这

些挑战, 该设置将标准棒球帽转换为基于单个盖帽安装的鱼眼相机的高质量姿态估计设备。从捕获的以自我为中心的实时流来看, 我们基于美国有线电视新闻网的 3d 姿态估计方法在消费者级别的 gpu 上以 60hz 的速度运行。除了新的硬件设置, 我们的其他主要贡献是: 1) 一个大的地面真相训练语料库自上而下的鱼眼图像和 2) 一个新的解开三维姿态估计方法, 其中包括自我中心的独特属性考虑到的观点。正如我们的评估所显示的, 我们实现了比现有基线更低的 3d 关节误差以及更好的 2d 叠加。少

2018 年 3 月 15 日提交;最初宣布 2018 年 3 月。

#### 45. 三维面部表情识别的精确定位和深度学习

作者:asm jan, huasongding, hhying meng, liming chen, huibin li

文摘: 有意义的面部部位可以传达面部动作单元检测和表情预测的关键线索。纹理三维人脸扫描可以提供详细的三维几何形状和 2d 纹理外观线索的脸, 这有利于人脸表情识别 (fer)。然而, 准确的面部部位提取以及融合是一项具有挑战性的任务。本文设计了一种基于精确的面部零件提取和面部零件深度特征融合的三维 fer 系统。特别是, 每个纹理三维人脸扫描首先表示为 2d 纹理贴图和具有一对一密集对应的深度贴图。然后, 利用一种新的四级过程提取纹理贴图和深度图的面部部分, 包括面部地标定位、面部旋转校正、面部大小调整、面部零件边界盒提取和后处理程序。

最后, 分别从纹理贴图和深度图中学习了所有面部部位的深度融合卷积神经网络 (cnn) 特征, 并利用非线性支持向量机进行表达式预测。在 BU-3DFE 数据库上进行了实验, 证明了在相同环境下对不同的面部部位、纹理和深度提示进行梳理并报告最先进的结果的有效性。少

2018 年 3 月 4 日提交;最初宣布 2018 年 3 月。

#### 46. 利用深度学习实时心血管 mr 与时空文物抑制--先天性心脏病概念的证明

作者: [Steeden hauptmann](#), [simon arridge](#), [felix lucka](#) , [vivek muthurangu](#), [jennifer a.steeden](#)

**摘要:** 目的: 实时评估心室体积需要很高的加速度因子。剩余卷积神经网络 (cnn) 已显示出消除数据采样不足所造成的伪影的潜力。在本研究中, 我们研究了不同径向采样模式对**美国有线电视新闻网**准确性的影响。我们还获得了先天性心脏病 (chd) 患者的实际实时低采样放射数据, 并将 cnn 重建与压缩传感 (cs) 进行了比较。方法: 开发了 3d (2d 加时间) cnn 架构, 并使用 2276 黄金标准配对 **3d** 数据集进行训练, 具有 14x 径向采样不足。使用 169 套以前看不见的**三维**"合成" 测试数据集, 测试了四种抽样方案。在 10 名新患者 (122 **例三维**数据集) 中获得了实际实时微小的黄金角 (tga) 径向 ssfp 数据, 并利用 **3d cnn** 和 cs 算法进行了重建;把握。结果: 采样模式对图像质量和心脏结构的准确可

视化具有重要意义。对于实际的实时数据, **cnn** 的总体重建时间(包括创建别名图像)比巨猿生存方案快 5 倍以上。此外, **cnn** 图像质量和双心室体积的准确性被观察到优于巨猿生存在相同的原始数据中。结论: 本文已经证明了在临床环境下使用**三维 cnn** 对实时径向数据进行深度去锯齿的潜力。利用实时数据与 **cnn** 重建的心室体积的临床测量与金标准、心脏门控、bh 技术在统计学上没有显著差异。少

**2018 年 6 月 14 日提交**;v1 于 2018 年 3 月 14 日提交;**最初宣布** 2018 年 3 月。

#### 47.从体积 oct 数据中进行测重的深层学习方法

**作者:**[nils gessert](#), [matthias schlüter](#),[亚历山大 schlaefer](#)

**摘要:** 跟踪仪器的姿势是图像引导手术中的一个核心问题。对于微观场景, 光学相干断层扫描 (oct) 越来越多地被用作成像方式。oct 由于其千分尺范围分辨率和体积场, 适用于精确的姿态估计。然而, 由于散斑噪声和反射伪影以及图像的**三维**性质, oct 图像处理具有挑战性。我们使用一个新的基于深度学习的跟踪框架来处理 oct 体积数据中的姿态估计问题。为此, 我们设计了一个新的**三维**卷积神经网络 (**cnn**) 架构, 直接预测 oct 卷中的一个标记几何的 6d 姿态。我们使用六足机器人自动获取标记的数据点, 我们使用它来训练 **3d cnn** 架构进行多输出回归。我们使用此

设置提供了对卷的基于深度学习的姿态估计的深入分析。具体来说, 我们证明, 利用量信息进行姿势估计比依赖具有深度信息的 **2d** 表示要高。为支持这一观察, 我们提供了定量和定性的结果, **3d cnn** 有效地利用了标记对象的深度结构。关于深度学习方面, 我们利用 **2d** 深度学习社区的见解, 提出了**高效的 3d cnn** 设计原则。特别是, 我们将感知 **3d** 作为一种新的体系结构呈现, 它最适合我们的应用程序。我们表明, 我们的深度学习方法达到错误, 在我们的地面真相标签的分辨率。我们实现了平均误差的平均值 $\sim 89 \pm 9.3 \mu\text{m}$  和  $\sim 0.096 \pm 0.072^\circ$  分别用于位置和定向学习。少

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

#### 48. 基于几何的增量分割实现快速、准确的语义映射

**作者** :yoshikatsu nakajima , keisuke tateno, fedico tombari, hido saito

**摘要:** 我们提出了一种高效且可扩展的方法, 用于实时增量构建密集的**语义注释 3d** 地图。该方法将类概率分配给 **3d** 地图的每个区域, 而不是每个元素 (例如, surfel 和 voxel), 该映射是通过强大的 slam 框架构建的, 并使用基于几何的分割方法进行增量分割。与所有其他方法不同的是, 我们的方法能够以超过 30hz 的速度运行, 同时执行所有处理组件, 包括 slam、分段、**2d** 识别和更新每个分段标签的类概率。每个传入帧, 由于我们的框架的计



算密集型阶段的高效率。通过利用专门设计的 **cnn** 来提高帧化分割结果, 还可以实现较高的精度。我们通过在精度和计算效率方面与最先进的方法进行比较, 并通过对时间和空间复杂性的分析, 验证了 nyuv2 数据集上的方法。少

**2018 年 3 月 7 日提交;最初宣布** 2018 年 3 月。

#### 49. 否认: 完全卷积密集形状回归野外

作者 : [riza alp guler](#), [yu xiang zhou](#), [george trigeorgis](#), [epameinondas](#) 安东纳科斯, [patrick snape](#), [stefanos zafeiriou](#), [iasonas kokkinos](#)

文摘: 在这项工作中, 我们使用深度学习建立**密集的对应**之间的三维对象模型和图像 "在野外"。我们引入 "densereg", 这是一个完全卷积的神经网络 (**f-nn**), 它在每个前景像素上密集地在单个前馈传递中回归一对  $u-v$  模板坐标。为了训练电雷格, 我们结合**三维变形模型拟合**和**二维地标注释**构造了一个监控信号。我们根据三维变形模型的内在  $u-v$  坐标来定义回归任务, 该模型在训练时与图像实例相对应。许多其他与对象相关的任务(例如部件分割、地标性本地化)被证明是这一任务的副产品, 并因其引入而得到很大改善。我们将语义分割思想与回归网络相结合, 形成了"量化回归" 体系结构, 首先通过分类获得了位置的量化估计, 并对其进行了细化, 从而获得了高精度的回归结果。通过回归的残余。我们表明, 这样的网络可以提高现有最先进的姿势估计系统

的性能。首先, 我们展示了我们的系统可以作为统计变形模型的初始化, 以及一个层级架构的元素, 共同定位地标和估计密集的对应。我们还表明, 所获得的密集通信可以作为 "特权信息" 的来源, 补充和扩展纯地标级注释, 加快和改进姿势估计网络的训练。我们在具有挑战性的 300w 基准上报告了最先进的性能, 用于面部地标定位, 以及用于人体姿态估计的 mpii 和 lsp 数据集。少

2018 年 3 月 11 日提交;[v1](#) 于 2018 年 3 月 5 日提交;[最初宣布](#) 2018 年 3 月。

## 50. 三维打印中视觉对象分类的图像数据集

作者:[李红佳](#),[马晓龙](#),[阿迪塔·辛格·拉托雷](#), 李哲, 安启元, [陈松](#),  
[徐文耀](#),[王延志](#)

**摘要:** 添加剂制造 (am) 的快速发展, 也被称为 3d 打印, 带来了潜在的风险和安全问题以及显著的好处。为了提高三维打印过程的安全水平, 本研究旨在利用深度学习检测和识别非法部件。在这项工作中, 我们收集了一个数据集的 618, 340 2d 图像 (每个图像 28x28) 的 10 类, 包括枪和其他非枪支对象, 对应于原来的 3d 模型的投影结果。为了验证数据集, 我们训练了一个卷积神经网络 (cnn) 的火炮分类模型, 该模型可以达到 98.16% 的分类精度。少

2018 年 3 月 22 日提交;v1 于 2018 年 2 月 15 日提交;最初宣布 2018 年 3 月。

## 51. 一种用于车辆三维检测的通用管道

作者:du, marcelo h. ang jr., sertac karaman, daniela rus

**摘要:** 自动驾驶需要对环境中的车辆和其他物体进行 3d 感知。目前的许多方法都支持 2d 车辆检测。本文提出了一种采用任何二维检测网络的柔性管道, 并将其与三维点云融合在一起, 以最小的二维检测网络变化生成三维信息。为了识别三维盒子, 提出了一种基于通用汽车模型和分数图的有效模型拟合算法。提出了一种两级卷积神经网络 (cnn) 来细化检测到的三维盒。使用两个不同的 2d 检测网络在 kitti 数据集上测试此管道。基于这两个网络的三维检测结果相似, 显示了所建议管道的灵活性。结果在三维检测算法中排名第二, 表明其在三维检测方面的能力。少

2018 年 2 月 12 日提交;最初宣布 2018 年 3 月。

## 52. 基于二维和三维卷积神经网络的 t1 加权 mri 的雄性盆腔合成 ct 生成

作者:jie fu, yliyong , kamal singh, dan ruan, daniel a. low, john h. lewis

**文摘** 为了实现磁共振 (mr) 纯放射治疗, 需要使用一种方法来估计合成 ct (sct) 生成电子密度图和患者定位参考图像。我们研究了二维和三维卷积神经网络 (**cnn**) 的方法, 以生成男性骨盆 sct 使用 t1 加权 mr 图像。对 20 例前列腺癌患者进行了 cts 和 t1 加权 mr 图像的回顾性研究。提出的 **2d cnn** 模型包含 27 个卷积层, 并从 segnet 进行了修改, 以获得更好的性能。还开发了**美国有线电视新闻网**模型的**三维**版本。这两个 **cnn** 模型都是从零开始训练, 将 t1 加权 mr 图像的强度映射到 ct hounsfield 单元 (hu) 值。每个 ct 都是在一个五-双交叉验证框架中生成的, 并与相应的 ct 进行了比较, 使用了体素平均绝对误差 (mae) 和骰子相似系数 (dsc)、召回和骨结构的精度。威尔科克森签署排名测试进行了评估这两个模型之间的差异。所有患者的 mae 平均值为  $40.5 \pm 5.4$  hu 和  $37.6 \pm 5.1$  hu 分别用于 **2d** 和 **3d cnn** 模型。骨结构的 dsc、召回和精度为  $0.81 \pm 0.04$ ,  $0.04 \pm 0.04$  和  $0.04 \pm 0.09$  为 **2dcnn** 模型和  $0.09 \pm 0.04$ ,  $0.04 \pm 0.04$  和  $0.04 \pm$  分别为 **3d cnn** 模型  $0.08$ 。除召回外, 威尔科克森签阶测试的 p 值不到 0.05, 召回为 0.6。**2d** 和 **3d cnn** 模型为 20 名使用 t1 加权 mr 图像的患者生成了准确的盆腔 sct。评价指标和统计测试表明,**三维**模型能够生成具有更好的 mae、骨 dsc 和骨精度的 sCTs。今后将对两种模型的剂量计算和患者定位的准确性进行测试和比较。少

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

### 53. 基于多模态传感器的大型环境语义三维映射

作者:钟民正,跆拳道,金白公园

**摘要:** 语义三维映射是机器人技术中最重要的领域之一,已在机器人导航、监控和虚拟现实等许多应用中得到应用。一般情况下,语义三维映射主要由三维重构和语义分割组成。随着这些技术的发展,近年来语义三维映射有了很大的进步。此外,需要在3d映射中提供语义信息以执行高级任务的机器人应用程序数量也有所增加,并发表了许多关于语义3d映射的研究。现有的方法使用相机进行3d重建和语义分割。但是,这不适用于大规模环境,并且具有计算复杂度高的缺点。为了解决这个问题,我们提出了一个基于多模态传感器的语义三维映射系统,该系统使用的是与摄像机相结合的三维激光雷达。在本研究中,我们建立了一个三维地图,估计气味测量基于全球定位系统(gps)和惯性测量单位(imu),并使用最新的2d卷积神经网络(cnn)进行语义分割。为了构建语义三维地图,我们采用坐标变换和贝叶斯更新方案将三维地图与语义信息结合起来。为了改进语义三维地图,我们提出了一个三维细化过程,以纠正错误分割的体素,并删除3d地图中移动车辆的痕迹。通过对具有挑战性序列的实验,我们证明了我们的方法在精度和交叉交联(iou)方面优于最先进的方法。因此,我们的方法可用于需要3d地图中语义信息的各种应用程序。少

2018年2月28日提交;最初宣布 2018年2月。

## 54. 基于卷积神经网络的同时交通标志检测与边界估计

作者: [hee seok lee](#), [kang kim](#)

**文摘:** 提出了一种利用卷积神经网络 (cnn) 同时估计交通标志位置和精确边界的新型交通标志检测系统。在智能车辆导航系统中, 估计交通标志的精确边界是很重要的, 因为在这些系统中, 交通标志可以作为道路环境的**三维**地标。以前的交通标志检测系统, 包括最近基于**美国有线电视新闻网**的方法, 只提供交通标志的边界框作为输出, 因此需要轮廓估计或图像分割等额外的过程来获得精确的标志边界。在这项工作中, 交通标志的边界估计被表述为一个**二维**姿态和形状类预测问题, 这可以通过一个 **cnn** 有效地解决。利用输入图像中目标交通标志的**预测二维姿态**和形状类, 将相应模板符号图像的边界投影到输入图像平面上, 估计目标符号的实际边界。通过将边界估计问题表述为**基于 cnn 的姿态和形状**预测任务, 我们的方法是端到端可训练的, 比其他依赖等高线估计的边界估计方法更具有鲁棒性和小目标。图像分割。该方法的结构优化提供了一种准确的交通符号边界估计, 在计算上也是有效的, 在低功耗移动平台上的检测帧速率高于每秒 7 帧。少

2018 年 2 月 27 日提交;最初宣布 2018 年 2 月。

## 55. 通过从 2d 训练网络进行传输学习, 实现低剂量 ct 的三维卷积物解码器网络

作者:[山红明](#),[张毅](#),[杨庆松](#),[乌韦克鲁格](#),[曼努迪普·卡拉](#), [孙玲](#), [文祥从](#), [葛王](#)

**摘要:** 低剂量计算机断层扫描 (ct) 已引起医学成像领域的主要关注, 因为 ct 相关的 x 射线辐射对患者的健康有风险。然而, ct 辐射剂量的降低会影响信噪比, 并可能影响图像质量和诊断性能。近年来, 基于深度学习的算法在低剂量 ct 去噪方面取得了很好的效果, 特别是卷积神经网络 (cnn) 和生成对抗网络 (gan)。本文介绍了一种基于承包路径的卷积解码器 (cpce) 网络, 该网络采用 2d 和 3d 配置, 用于低剂量 ct 去噪。我们的方法的一个新特点是, 可以通过扩展训练有素的 2d cnn, 然后微调以合并相邻切片中的 3d 空间信息, 从而直接获得初始 3d cpce 去噪模型。基于从 2d 到 3d 的迁移学习, 3d 网络比从零开始训练的网络收敛得更快, 并实现更好的去噪性能。通过对 cpce 与最近发布的基于模拟 mayo 数据集和真实 mgh 数据集的方法的比较, 证明了三维 cpce 去噪模型具有更好的性能, 抑制了图像噪声, 保留了细微的结构。少

**2018 年 4 月 29 日提交;**v1 于 2018 年 2 月 15 日提交;**最初宣布** 2018 年 2 月。

## 56. 基于图像的深三维人体姿态估计合成

作者:[grégory roez](#), [cordelia schmid](#)



**文摘:** 本文讨论了野外**三维**人体姿态估计的问题。一个重大的挑战是缺乏训练数据, 即用**三维**姿势注释的人类 2d 图像. 这些数据对于培训最先进的 **cnn** 架构是必要的。在这里, 我们提出了一个解决方案, 以生成大量的逼真的合成图像的人类与 **3d** 姿态注释。我们引入了一种基于图像的合成引擎, 该引擎使用**三维**运动捕获数据, 利用 **2d** 人体姿态注释人为地增强了真实图像的数据集。给定候选 **3d** 姿态, 我们的算法为每个关节选择一个图像, 其**二维**姿势在本地与**投影的 3d 姿势**相匹配。然后将选定的图像组合在一起, 通过以运动受限的方式拼接本地图像补丁来生成新的合成图像。由此产生的图像被用来训练一个**端到端的美国有线电视新闻网**, 用于全身**三维姿势**估计。我们将训练数据分组到大量的姿势类中, 并将姿势估计作为一个 K-方式分类问题。这种方法只有在像我们这样的大型训练装置中才是可行的。我们的方法在受控环境中的 **3d** 姿态估计方面优于大多数已发表的作品 (人人 3.6 m), 并显示了现实世界图像 (lsp) 的有希望的结果。这表明, 在人工图像上训练的 **cnn** 很好地概括了真实图像。与从更经典的渲染引擎生成的数据相比, 我们的合成图像不需要任何域适应或微调阶段。

少

**2018 年 2 月 12 日提交;最初宣布** 2018 年 2 月。

## 57. 基于图像的局部化的全帧场景坐标回归

**作者:**李晓天, [juha ylioinas](#), [juho kannala](#)

**摘要:** 基于图像的定位, 即相机的重新定位, 是计算机视觉和机器人技术中的一个基本问题, 它是指从图像中估计相机的姿态。最近最先进的方法使用基于学习的方法, 如随机森林 (rf) 和卷积神经网络 (cnn), 为图像中的每个像素回归其在场景的世界坐标框架中的相应位置, 并解决通过基于 ransacs 的优化方案使用预测的对应进行最终姿势。在本文中, 我们建议以全帧的方式执行场景坐标回归, 以提高测试时间的计算效率, 更重要的是, 在回归过程中添加更多的全局上下文来改进鲁棒性。为此, 我们采用了一种完全卷积的编码器-解码器神经网络体系结构, 该体系结构接受整个图像作为输入, 并对图像中的所有像素生成场景坐标预测。然而, 使用更多的全局上下文很容易过度拟合。为了缓解此问题, 我们建议使用数据扩充来生成更多用于培训的数据。除了二维图像空间中的数据增强外, 我们还对三维空间中的数据进行了扩充。我们在公开的 7 场景数据集上评估我们的方法, 实验表明, 它具有更好的场景坐标预测, 并在本地化方面实现了最先进的结果, 提高了对最坚硬帧 (例如, 重复的结构)。少

**2018 年 6 月 25 日提交;v1 于 2018 年 2 月 9 日提交;最初宣布 2018 年 2 月。**

## 58. 一种空间映射算法, 在深部学习结构分类中的应用

**作者:**thomas corcoran, rafael zamora-reendiz,新联 liu, silvia crivelli

**摘要:** 基于卷积神经网络 (cnn) 的机器学习系统在二维特征提取和图像识别任务方面取得了突破性进展。虽然目前正在开展大量工作, 将 cnn 技术应用于涉及复杂 3d 数据的领域, 但这些努力的成功在一定程度上受到数据表示技术限制的限制。大多数现有的方法依赖于低分辨率的 3d 模型、3d 空间中的范围的战略限制, 或应用有损投影技术来允许使用 2d cnn.为了解决这个问题, 我们提出了一种映射算法, 通过将三维空间填充曲线的遍历映射到相应的 2d 和 1d 曲线的遍历, 将三维结构转换为 2d 和 1d 数据网格。我们探索 2d 和 1d cnn 在使用我们的方法编码的数据上训练的性能, 而不是在流行的基准数据集的原始 3d 数据上操作的可比体积 cnn 的性能。我们的实验表明, 通过我们的方法生成的 3d 数据的 2d 和 1d 表示以 learnable 学习的形式保留了相当一部分 3d 数据的特征. 此外, 我们还证明了我们将 3d 数据编码为低维表示的方法可以降低 cnn 的训练时间成本, 增加原始 3d 模型渲染分辨率, 并支持与纯体积方法相比, 数据通道的数量增加。本演示是在结构生物学分类任务的背景下完成的, 我们在 ras 蛋白家族中两个同源分支的示例上对 3d、2d 和 1d cnn 进行了训练。本文的重要贡献是引入了一种降维方法, 该方法可以简化在以复杂结构数据为特征的领域应用强大的深度学习工具的问题。少

**2018 年 2 月 22 日提交;**v1 于 2018 年 2 月 7 日提交;**最初宣布** 2018 年 2 月。

## 59. 利用主动形状模型和卷积神经网络进行精确的脑部提取

作者:nguyen ho minh duy, nguyen manh duy , mai thanh nhat truong, pham the bao, nguyen thanh binh

**摘要:** 在大多数神经成像处理系统中, 大脑提取或颅骨剥离是一个基本的过程。这一过程的表现对神经影像学分析的成功产生了至关重要的影响。经过几年的研究和开发, 大脑提取仍然是一个具有挑战性的问题。本文提出了一种在磁共振成像 (mri) 扫描中进行颅骨剥离的有效方法, 称为 **asm-nnn**。我们的系统是主动形状模型 (asm) 和卷积神经网络 (**cnn**) 的结合, 充分利用这两种方法取得了显著的效果。我们不是使用**三维**结构, 而是在矢状平面上处理**二维**图像序列。首先, 我们将图像分成不同的组, 以便在每个组中, 大脑边界的形状和结构都有类似的外观。这样就可以为每个组开发精确的算法, 以产生高性能的分割结果。其次, 利用每个组的先验知识, 使用改进后的 asm 版本来检测图像中的大脑边界。最后,应用 **cnn** 和后处理方法, 如条件随机场、高斯过程和一些特殊的规则来细化 asm 产生的分割轮廓。我们比较了 **asm-nnn** 与**五种**最先进的、可公开使用的方法的最新版本, 即 bet、bse、3dss、robex 和 beast。该评估是使用三个公共数据集 ibsr、lpba 和 oasis 进行的。实验结果表明, 该方法优于五种最先进的算法, 在所有实验中都大大优于其他方法。少

**2018 年 2 月 5 日提交;最初宣布** 2018 年 2 月。

## 60. 深度药物-靶结合亲和力预测

作者:[hakime ztürk](#), [elif ozkirimli](#), [arzucan zgür](#)

**文摘:** 新的药物靶向 (dt) 相互作用的识别是药物发现过程的重要组成部分。为预测 dt 交互而提出的大多数计算方法都集中在二进制分类上, 其目标是确定 dt 对是否相互作用。然而, 蛋白质配体相互作用假设了一个连续的结合强度值, 也被称为结合亲和力和预测这个值仍然是一个挑战。dt 知识库中可用的亲和力数据的增加允许使用高级学习技术 (如深度学习体系结构) 来预测绑定亲和力。在本研究中, 我们提出了一个基于深度学习的模型, 该模型仅使用目标和药物的序列信息来预测 dt 相互作用结合亲和力。研究重点研究 dt 结合亲和力预测的少数研究要么使用蛋白质配体配合物的三维结构, 要么使用化合物的二维特征。本文采用的一种新方法是使用卷积神经网络 (cnn) 对蛋白质序列和化合物一维表示进行建模。结果表明, 所提出的基于深度学习的基于深度学习的模型是一种有效的药物靶标结合亲和力预测方法。通过 cnn 构建药物和目标的高级表示的模型在我们的一个较大的基准数据集中实现了最佳的协调指数 (ci) 性能, 其性能优于 kronrls 算法和 simboost, dt 结合亲和力预测的最先进方法。少

**2018 年 6 月 5 日提交;**v1 于 2018 年 1 月 30 日提交;**最初宣布** 2018 年 1 月。

## 61. 球面 cn

作者:[taco s. cohen](#), [mario geiger](#), [jonas koehler](#), [max welling](#)

**摘要:** 卷积神经网络 (cnn) 已成为涉及二维平面图像学习问题的首选方法。然而, 最近感兴趣的一些问题产生了对能够分析球面图像的模型的需求。这方面的例子包括无人机、机器人和自主汽车的全方位视觉、分子回归问题以及全球天气和气候建模。将卷积网络简单地应用于球面信号的平面投影注定要失败, 因为这种投影所产生的空间变化扭曲将使平移权重共享无效。本文介绍了构建球形 cn 的构建模块. 我们提出了一个定义的球面互相相关, 这是既表达和旋转等变量。球面相关满足广义傅立叶定理, 它允许我们使用广义 (非交换) 快速傅立叶变换 (fft) 算法有效地计算它。我们证明了球面 cnn 在三维模型识别和雾化能量回归中的计算效率、数值精度和有效性。少

2018 年 2 月 25 日提交;v1 于 2018 年 1 月 30 日提交;最初宣布 2018 年 1 月。

## 62. 堆叠式滤波器定流用于面向硬件的深部对流层神经网络加速

作者:[高月超](#), [刘念红](#), [张生](#)

**摘要:** 为了解决深卷积神经网络 (cnn) 面向硬件的加速度的内存和计算资源限制问题, 我们提出了一个计算流、堆叠滤波器固定

流 (sfs) 和相应的数据编码格式, 相对索引压缩稀疏过滤格式 (csf), 以充分利用数据稀疏性, 并简化执行时的数据处理。我们还提出了一个三维单指令多数据 (**3d-simd**) 处理器架构, 以说明如何利用 sfs 流和 csf 格式加速深度 **cnn**。与最先进的结果 (han 等人, 2016b) 相比, 我们的方法在减少 alex net 所需的存储空间方面实现了 1.11x 的改进, 在减少 squeezenet 所需存储的情况下实现了 1.11x 的改进, 而不会在 imagenet 上失去准确性数据。此外, 利用这些方法, 可以节省处理不规则稀疏数据访问的逻辑芯片面积。与 dvas、vision 等领域的 **2d-simd** 处理器结构相比, 我们的方法在 alex net 深度压缩数据上实现了约 3.65 x 处理元件 (pe) 阵列利用率的提高 (从 26.4% 提高到 96.5)。

少

**2018 年 2 月 5 日提交;v1** 于 2018 年 1 月 23 日提交;**最初宣布** 2018 年 1 月。

### 63. 基于拼接的脑磁共振成像组织分割全卷神经网络的定量分析

**作者:** jose bernal, kaisar kushibar, mariano cabezas, sergi valverde, amau oliver , xavier lladó

**摘要:** 磁共振成像 (mri) 中准确的脑组织分割引起了医生和研究人员的注意, 因为组织体积的变化有助于诊断和监测神经疾病。多年来, 已经设计了几个建议, 包括传统的机器学习策略以及卷积神经网络 (**cnn**) 方法。特别是, 在本文中, 我们分析了产生密



集预测的深度学习方法的一组分组。这个分支在文献中被称为 "**全有线电视新闻网(fcn)**", 令人感兴趣, 因为这些架构可以在比 **cnn** 更短的时间内处理输入卷, 并且可以对本地空间依赖关系进行编码, 因为多个体素是同时分类的。我们的研究重点是了解类似文学的方法的体系结构优势和弱点。因此, 我们实施了八个 fcn 架构的灵感来自于强大的最先进的方法大脑分割相关的任务。我们使用 ibsr18、miccai2012 和 igeseg2017 数据集对它们进行评估, 因为它们包含婴儿和成人数据, 并展示了不同的体素间距、图像质量、扫描次数和可用的成像方式。讨论从三个方面进行: **2d** 和 **3d** 方法之间的比较、多种模式和重叠作为培训和测试模型抽样策略的重要性。为了鼓励其他研究人员探索评估框架, 可以从我们的研究网站下载公开版本。少

**2018 年 2 月 19 日提交;v1** 于 2018 年 1 月 19 日提交;**最初宣布** 2018 年 1 月。

## 64. 中间概念的深度监督

**作者** :chili, m. zeeshan zia, quoc-huy tran, 项宇, gregory d.hager, manmohan chandraker

**摘要**: 最近的场景解释数据驱动方法主要是将推理作为一种端到端的黑盒映射, 通常由卷积神经网络 (**cnn**) 执行。然而, 几十年来关于人的视觉和机器视觉中的感知组织的工作表明, 通常有一个中间表示是固有的推理任务, 并提供了必要的结构, 以提高泛

化。在这项工作中，我们探索了一种方法，注入现有域结构到神经网络培训中，方法是通过监督 **cnn** 的隐藏层，其中通常在实践中没有观察到的中间概念。我们制定了一个概率框架，将这些概念形式化，并通过这种深度监督方法预测改进的泛化。这种方法的一个优点是，我们只能从杂乱场景的合成 cad 渲染中进行训练，在这种渲染中可以提取概念值，但将结果应用于真实图像。我们的实施实现了 **2d/3d** 关键点定位和图像分类的最先进性能，包括 kitti、pascal voc、pascal 3d+、宜家和 cibar100。我们提供了更多的证据，证明我们的方法优于其他形式的监督，如多任务网络。少

**2018 年 7 月 20 日提交**;v1 于 2018 年 1 月 8 日提交;**最初宣布** 2018 年 1 月。

## 65. 超声中盆底提升机间隙的自正归一化神经网络自动分割方法

**作者:**埃斯特·邦马蒂, [yipenghu](#), [nikhil sindhwani](#), [hans peter dietz](#), [jan d ' hooge](#), [dean barratt](#), [jan depreest](#), [tom Vercauteren](#)

**文摘:** 在超声中分割提升体间隙可以提取对盆底疾病评估非常重要的生物鉴别技术。在这项工作中，我们提出了一个全自动的方法，使用卷积神经网络 (**cnn**) 来概述从**三维**超声体积提取的**二维**图像中的提升体间隙。特别是，我们的方法使用最近开发的缩放指数线性单位 (**selu**) 作为一个非线性自正激活函数，这在医

学成像中首次应用于 **cnn**。selu 具有无参数和独立于小型批处理等重要优点,这可能有助于克服训练过程中的内存限制。在 valsalva、收缩和休息期间,一个由 35 名患者提供 91 张图像的数据集,都贴有 3 个操作人员的标签,用于在一个病人的交叉验证中进行培训和评估。结果显示, dice 相似系数中值为 0.90,四分位间范围为 0.90,与三个运算符的性能相当 (williams 的指数为 1.03),并且在不需要批处理规范化的情况下优于 u-net 体系结构。与以往的半自动方法相比,所提出的全自动方法在分割盆底提升机间隙时达到了等效精度。少

**2017 年 12 月 18 日提交;最初宣布** 2017 年 12 月。

## 66. 美国有线电视新闻网使用 velodyne lidar 进行 imu 辅助测理估计

**作者:**martin velas, m 奇勒·斯帕内尔, m 奇勒·赫拉迪斯,亚当·赫鲁特

**文摘:**介绍了一种利用**三维** lidar 扫描的卷积神经网络进行气味测量的新方法。原始稀疏数据被编码为 **2d** 矩阵,用于训练拟议的网络和预测。与最先进的 loam 方法相比,我们的网络在平移运动参数估计方面表现出了更好的精度,同时实现了实时性能。结合 imu 支持,实现了高质量的气味估计和 lidar 数据配准。此外,我们还提出了替代 **cnn** 培训预测旋转运动参数,同时取得的结果

也可与最先进的。该方法可以在气味估计中取代车轮编码器，或在全球导航卫星系统信号缺席时（例如在室内映射期间）补充缺失的 gps 数据。我们的解决方案带来了实时性能和精度，可用于在线预览地图结果和实时验证地图完整性。少

2017 年 12 月 18 日提交;最初宣布 2017 年 12 月。

## 67.对时空特征学习的再思考：视频分类中的速度精度权衡

作者:谢素宁,孙晨,黄冠明, 杜卓文, 凯文·墨菲

文摘: 尽管在采用卷积神经网络 (cnn) 的带动下, 视频分析取得了稳步进展, 但相对改善的幅度不如二维静态图像分类。存在空间 (图像) 特征表示、时间信息表示和模型计算复杂度三个主要挑战。carreira 和 zisserman 最近显示,从 2d 网络膨胀并在 imagenet 上预先培训的 3d cnn 可能是空间和时间表示学习的一种有前途的方法。然而, 至于模型/计算的复杂性, 3d cnn 比 2d cnn 昂贵得多, 并且容易过度匹配。通过对关键网络设计选择的系统探索, 建立有效高效的视频分类系统, 寻求速度和准确性之间的平衡。特别是, 我们表明, 有可能用低成本的 2d 卷积来取代许多三维卷积。令人惊讶的是, 当替换网络底部的 3d 卷卷时, 最好的结果 (速度和精度) 都是可以实现的, 这表明在高级语义特征上进行时间表示学习更有用。我们的结论概括了具有非常不同属性的数据集。当与其他几个具有成本效益的设计相结合, 包

括可分离的时空卷积和特征门控, 我们的系统产生了一个有效的视频分类系统, 产生非常有竞争力的几个行动的结果分类基准 (动力学、某种东西、ucf101 和 hmdb), 以及两个操作检测 (本地化) 基准 (jhmdb 和 UCF101-24)。少

**2018年7月26日提交;**v1于2017年12月13日提交;**最初宣布** 2017年12月。

## 68. 基于深度的三维手姿势估计: 从当前的成就到未来的目标

**作者 :**shanxinyuan , guillermo garcia–hernando, bjorn stenger, gyeongsikmoon, ju yong chang, kyoung mu lee, pavlo molchanov, jan kautz, sinahonari, 葛俊松,袁俊松,陈兴浩, 王桂金, 范阳, 凯秋山, 杨武, 清福湾, 梅桑马达迪, 塞尔希奥 Escalera, shile li, Dongheui lee, iason oikonomidis, antonis argyros, tae–kyun kim

**文摘:** 在本文中, 我们试图回答两个问题: 从深度图像的**三维手模估计**的当前状态是什么? 而且, 接下来需要应对的挑战是什么? 继 "百万挑战中的手" (him2017) 之后, 我们将研究三个任务的十大最先进的方法: 单帧 **3d** 姿势估计、**3d** 手部跟踪和对象交互过程中的手姿势估计。我们分析了不同的 **cnn** 结构在手的形状, 关节可见性, 视点和发音分布的性能。我们的研究结果包括: (1) 孤立的 **3d** 手部姿态估计在  $[70, 120]$  度的视点范围内实现了较低的均值误差 (10 毫米), 但对于极端视点, 它远未得到解决;(2) **三维体积表示性能**优于**二维 cnn**,更好地捕捉深度数据的空间结

构;(3) 判别方法对看不见的手形仍有很差的概括;(4) 虽然关节遮挡对大多数方法都是一个挑战, 但明确的结构约束建模可以显著缩小可见关节和遮挡关节上的误差之间的差距。少

**2018年3月29日提交**;v1于2017年12月11日提交;**最初宣布** 2017年12月。

## 69. spp-网络: 具有合成视图的深绝对对向回归

**作者:**[pulak purkait](#), [chengzhao](#) , [christopher zach](#)

**摘要:** 基于图像的定位是计算机视觉中的重要问题之一, 因为它广泛应用于机器人技术、增强现实和自主系统。文献中描述了一套丰富的方法, 如何对二维图像进行几何登记. 近年来, 基于深度(和卷积) 前馈网络(cnn) 的方法在姿态回归中越来越流行。然而, 这些基于美国有线电视新闻网的方法尽管速度快, 内存高效, 但仍然不如基于几何的方法准确。在本文中, 我们设计了一个基于稀疏特征描述符的深层神经网络体系结构来估计图像的绝对姿态。我们选择使用稀疏功能描述符有两个主要优点: 第一, 我们的网络明显小于此任务文献中提出的cnn ——从而使我们的方法更高效、更具可扩展性。其次——更重要的是, 稀疏功能的使用允许使用合成视点来增强训练数据, 从而大大改进了泛化性能, 使其具有看不见的姿势。因此, 我们提出的方法旨在将这两个世界中最好的——基于特征的定位和基于cnn的姿态回归——结合起来,

在绝对姿态估计中实现最先进的性能。为了支持我们的方法, 对建议的体系结构进行了详细分析, 并对现有数据集进行了严格的评估。少

2017 年 12 月 9 日提交;最初宣布 2017 年 12 月。

## 70.用于多智能体 2d/3d 医学图像注册的膨胀 fcn

作者: 顺苗、 [sebastien piat](#)、 [peter fischer](#)、  
[ahmet tuysuzoglu](#)、 [phillip mewes](#)、 [tommaso mansi](#)、 [rui liao](#)

**抽象:** 2d/3d 图像配准以对齐 3d 体积和二维 x 射线图像是一个具有挑战性的问题, 因为它的复杂性和 2d x 射线图像中呈现的各种伪影.本文提出了一种具有自动注意机制的多智能体系统, 用于鲁棒高效的 2d/3d 图像配准.具体而言, 对单个代理进行了扩展的全卷网络 (fcn) 的培训, 以便通过观察本地区域在马尔可夫决策过程 (mdp) 中进行注册, 然后根据多个代理机构的建议采取最后行动。由相应的置信度加权。本文的贡献有三个方面。首先, 我们将 2d/3d 注册表述为 mdp, 并对 x 射线成像系统进行适当定义的观察、操作和奖励.其次, 为了处理二维 x 射线图像中的各种伪影, 通过基于 fcn 的结构有效地利用了多个局部代理, 提出了一种自动注意机制, 以支持具有更可靠视觉线索的区域的建议。第三, 建议扩大基于 fcn 的培训机制, 以大幅降低模拟登记环境的自由度, 并与标准有线电视新闻网相比, 大幅提高培训效率一



个数量级——基于训练方法。我们证明了该方法对脊柱锥束计算机断层扫描数据具有较低的信噪比和微创脊柱手术的数据都具有较高的鲁棒性, 在这种手术中, 由于金属而出现了严重的图像伪影和遮挡。螺丝和导丝, 性能超过其他最先进的方法(基于单一代理和优化为基础) 的优势很大。少

2017 年 11 月 21 日提交;最初宣布 2017 年 12 月。

## 71. 基于深部神经网络的信息流建模

作者: [ahmad chaddad](#), [behnaz naisiri](#), [marco pedersoli](#), [eric granger](#), [christian desrosiers](#), [matthew toews](#)

文摘: 本文提出了一种具有原则的信息理论分析, 对深层神经网络结构(如卷积神经网络)的分类进行了分析。卷积滤波器的输出被建模为基于对象类  $c$  和网络滤波器组  $f$  的随机变量  $y$ 。条件熵 (cent)  $h(y|c, f)$  在理论和实验中被证明是一个高度紧凑和类信息的代码, 可以从整个现有的 **cnn** 的滤波器输出计算, 并用于获得比原来的**美国有线电视新闻网**本身更高的分类结果。实验证明了在两个独立的 **cnn** 分类上下文中进行 cent 特征分析的有效性。1) 在阿尔茨海默氏症 (ad) 引起的神经退行性变和三维磁共振图像 (mri) 体积的自然老化的分类中, 3 个分支特征导致全脑 ad 分类的 aucuco:94.6%, 最高报告的准确性为使用的公共 oasis 数据集比为这项任务所训练的原始 **cnn** 的软最大输出高 12%。2) 在从 2d 照片的视觉对象分类的背景下, 基于在现有

**cnn** 中识别的一小部分 **cent** 特征的转移学习会导致 **auc** 值与 1000 功能的最大软输出相媲美。对以前看不见的对象类别进行分类时, 可以使用原始网络。一般信息理论分析解释了**美国有线电视新闻网**最近的各种设计成功, 例如密集连接的 **cnn** 架构, 并为未来的深度学习研究方向提供了见解。少

**2017 年 11 月 29 日提交;最初宣布** 2017 年 12 月。

## 72. 近距离观察用于行动识别的时空卷积

**作者:** [du tran](#), [heng wang](#), [lorenzo torresani](#), [jamie ray](#), [yann lemun](#), [manohar paluri](#)

**文摘:** 本文讨论了几种形式的时空卷积进行视频分析, 并研究了它们对动作识别的影响。我们的动机源于这样的观察,即 **2dcnn** 应用于视频的各个帧在行动识别方面仍然是扎实的表现。在这项工作中, 我们经验证明了 **3d cnn** 比 **2d cnn** 的准确性优势的框架内的剩余学习。此外, 我们还表明,将**三维卷积滤波器**分解为单独的空间和时间组件, 在精度方面具有显著的优势。我们的实证研究导致了一个新的时空卷积块 "**R(2+1)D**" 的设计, 它产生了 **cnn** , 实现的结果与最先进的 **spatiotemporal**, **动力学**, **ucf101** 和 **hmdb51**。少

**2018 年 4 月 11 日提交;v1** 于 2017 年 11 月 30 日提交;**最初宣布** 2017 年 11 月。

### 73. 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) 已被认为是一类强大的图像识别问题模型。然而, 它不是微不足道的, 当使用..。更多

2017 年 11 月 28 日提交;最初宣布 2017 年 11 月。

### 74. 三维-a 网: 三维深度密集描述符与抗疟疾网络的体积形状

作者:任梦伟,牛亮,易芳

摘要: 最近, 研究人员将注意力转移到学习的 3d 形状描述符上, 从手工制作描述符转向更好地解决 3d 物体固有的变形和结构变化这一具有挑战性的问题。三维几何数据通常被转换为具有规则格式的 3d voxel 网格, 以便更好地输入到深层神经网络体系结构中。然而, 将三维卷积网直接应用于三维体积数据的计算难易程度严重限制了处理 3d 的效率 (即缓慢处理) 和有效性 (即不满意的精度) 几何数据。本文以一种新型的敌对网络 (3d-a-netes) 设计为动力, 开发了一种新型的 3d 深密集形状描述符 (3d-d 届时), 以解决高效高效 3d 这一具有挑战性的问题。体积数据处理。我们开发了三维体积数据的二维多层密集表示 (mdr) 的新定义, 以提取简洁但几何信息丰富的形状描述和新的对抗性网络设计, 共同训练一组卷积神经网络 (cnn)、递归神经

网络 (rnn) 和对抗判别器。更具体地说, 生成器网络生成 3d 形状特征, 鼓励使用正确的类标签对同一类别的样本进行聚类, 而鉴别器网络则通过分配这些特征来阻止聚类误导敌对类标签。通过解决 cnn 直接应用于 3d 体积数据的计算效率所带来的挑战, 3d-a-net 可以学习高质量的 3d-dsdd, 这证明了它的优越性在 3d 形状分类和检索方面的性能比其他最先进的技术有很大的优势。少

2017 年 11 月 27 日提交;最初宣布 2017 年 11 月。

## 75.spatiemportal 3n 是否可以追溯 2d cnn 和 imagenet 的历史?

作者:[koso hara](#), [hihikatsu kataoka](#), [yutaka satoh](#)

文摘: 本研究的目的是确定当前的视频数据集是否有足够的数据来训练非常深的卷积神经网络 (kernels) 与时空三维 (3d) 内核。近年来, 3d cnn 在动作识别领域的性能水平有了显著提高。然而, 到目前为止, 传统的研究只探索了相对较浅的 3d 架构。我们研究了各种 3d cnn 的架构, 从相对较浅的架构到非常深的架构上的当前视频数据集。根据这些实验的结果, 可以得出以下结论: (一) resnet-18 培训对 ucf-101、hmdb-51 和 activitynet 进行了显著的过度拟合, 但对动力学没有。(ii) 动力学数据集具有足够的数据来训练深层 3d cnn, 并可训练多达 152 个 resnet 图层, 有趣

的是, 类似于 imagenet 上的 **2d** 重置器。resnex-101 在动力学测试集中实现了 784% 的平均精度。(iii) 预训练的简单 **3d** 架构优于复杂的 **2d** 架构, 预培训的 renet-101 在 ucf-101 和 hmdb-51 上分别达到 94.5 和 70.2。在 imagenet 上训练的 **2dcnn** 的使用在图像的各种任务方面取得了重大进展。我们相信, 使用深度 **3d cnn** 和动力学将追溯 **2d cnn** 和 imagenet 的成功历史, 并刺激视频的计算机视觉的进步。本研究中使用的代码和预先培训的模型是公开提供的。

<https://github.com/kenshohara/3d-> 电 阻

---

2018 年 4 月 1 日提交;v1 于 2017 年 11 月 27 日提交;最初宣布 2017 年 11 月。

## 76. 实时无缝单次射击 6d 物体姿势预测

作者:bugra tekin, sudipta n. sinha, pascal fua

**摘要:** 我们提出了一种单次方法, 用于同时检测 rgb 图像中的对象并预测其 6d 姿态, 而无需多个阶段或必须检查多个假设。与最近为这项任务提出的单次拍摄技术 (kehl 等人, iccv17) 只预测大约 6d 姿势, 然后必须细化, 我们的技术足够准确, 不需要额外的后处理。因此, 它的速度要快得多—在泰坦 x (pascal) gpu 上 50 fps—更适合实时处理。我们方法的关键组成部分是一个新的 **cnn** 架构, 灵感来自 yolo 网络设计, 直接预测对象的 **3d** 边界框

的投影顶点的 **2d** 图像位置。然后使用 pnp 算法估计对象的 6d 姿态。对于 linemod 和 occlusion 数据集上的单个对象和多个对象姿态估计, 当所有方法都在没有后处理的情况下使用时, 我们的方法大大优于**其他基于 cnn** 的方法。在后处理过程中, 可以使用姿势细化步骤来提高现有方法的准确性, 但在 10 fps 或更少的情况下, 它们比我们的方法慢很多。少

**2018 年 4 月 18 日提交;**v1 于 2017 年 11 月 23 日提交;**最初宣布** 2017 年 11 月。

## 77.sgpn: 用于 3d 点云实例细分的相似组建议网络

**作者:**王伟月,于, 黄秋贵 , 乌尔里希·纽曼

**文摘:** 我们介绍了相似性组建议网络 (sgpn), 这是一个简单直观的深度学习框架, 用于在点云上进行 **3d** 对象实例分割。sgpn 使用单个网络来预测点分组建议, 并为每个方案对应一个语义类, 我们可以直接从中提取实例分段结果。sgpn 的有效性的重要意义在于它以相似矩阵的形式对**三维**实例分割结果进行了新的表示, 该矩阵指示嵌入特征空间中每一对点之间的相似性, 从而产生准确的结果。对每个点的分组建议。据我们所知, sgpn 是第一个在点云上学习**三维**实例感知语义分割的**框架**。在各种**三维**场景上的实验结果表明了该方法在**三维**实例分割方面的有效性, 并对 sgpn 提高**三维**目标检测和语义分割结果的能力进行了评价。我

们还通过无缝地将 **2d cnn** 功能集成到框架中来展示其灵活性, 以提高性能。少

2017 年 11 月 23 日提交;最初宣布 2017 年 11 月。

## 78. 三维各向异性混合网络: 将卷积特征从二维图像转换为三维各向异性卷

作者:刘思琪,徐大光, s.kevin zhou, thomas mertelmeier, julia wicklein, anna jerebko, sasa grbic, olivier pauly, weidongcai, dorin 科马尼奇

**文摘** 虽然深卷积神经网络 (**cnn**) 已成功地应用于**二维**图像分析, 但将其应用于**三维**各向异性体积仍面临挑战, 尤其是在片内分辨率要高得多的情况下比切片之间的分辨率, 当 **3d** 卷的数量相对较小。一方面, 直接学习**美国有线电视新闻网**与 **3d** 卷积内核受到缺乏数据, 并可能最终导致不太泛化;gpu 内存不足会限制模型大小或表示能力。另一方面, 将具有泛化特征的 **2dcnn** 应用于**二维**切片而忽略切片间的信息。由于 **lstm** 学习的困难, 将 **2d** 网络与 **lstm** 耦合以进一步处理片间信息并不是最佳的。为了克服上述挑战, 我们提出了一个**三维**各向异性混合网络 (**ah-net**), 将从**二维**图像中学习到的卷积特征传输到**三维**各向异性体积。这样的转移继承了所需的强泛化能力, 用于切片内信息, 同时自然地利用切片间信息进行更有效的建模。焦点丢失进一步用于更有效的端到端学习。我们实验的**建议三维** **ah-net** 在两个不同的医学图

像分析任务, 即病变检测从数字乳房肿瘤合成体积, 肝脏和肝脏肿瘤分割从一个计算机断层扫描体积, 并获得最先进的结果。少

2017 年 12 月 3 日提交;v1 于 2017 年 11 月 23 日提交;最初宣布 2017 年 11 月。

## 79. 时间三维网: 视频分类的新架构和转移学习

作者: [ali diba](#), [mohsen fayyaz](#), [vivek sharma](#), [amir hossein karami](#), [mohammad mahdi arzani](#), [rahman yousefzadeh](#), [luc van gool](#)

**摘要:** 本文的工作是由如何利用视频中可用的时间线索进行准确分类, 特别是人类行动识别的问题所驱动的? 到目前为止, 视觉社区已经把重点放在具有固定时间卷积核深度的时空方法上。我们引入了一个新的时间层, 对可变时间卷积核深度进行建模。我们在我们提出的 **3d cnn** 中嵌入了这个新的时间层。我们通过 **3d** 过滤器和池核扩展了 **densenet** 体系结构 (通常是 **2d**)。我们命名我们提出的视频卷积网络 '时间 **3d** ConvNet'~(T3D) 和它的新时间层 '时间过渡层'~(ttl)。我们的实验表明, t3d 在 **hmdb51**、**ucf101** 和动力学数据集上的性能优于目前最先进的方法。培训 **3d convnet** 的另一个问题是如何从零开始使用一个巨大的标记数据集对它们进行培训, 以获得合理的性能。因此, 在 **2d** 凸网中学到的知识被完全忽略。这项工作的另一个贡献是一个简单而有效的技术, 将知识从预先训练的 **2d cnn** 转移到随机初



始化的 3d cnn, 以实现稳定的重量初始化。这使我们能够显著减少 3d cnn 的训练样本数量。因此, 通过对该网络进行微调, 我们击败了 3d cnn 中通用方法和最新方法的性能, 这些方法在大型视频数据集 (例如 sports-1m) 上进行了训练, 并在目标数据集 (例如 hmdb5ucf101) 上进行了优化。t3d 代码将被释放少

2017 年 11 月 22 日提交;最初宣布 2017 年 11 月。

## 80. 基于二维图像的物体分类和三维姿态估计的卷积网络

作者:[Siddharth mahendran](#), [haider ali](#) , [rene vidal](#)

**摘要:** 目前基于 cnn 的恢复图像中对象的三维姿态的算法假定对图像中的对象类别及其二维定位都有了解。本文放宽了这些约束之一, 并提出从假定已知二维定位的图像出发, 解决关节对象类别和三维姿态估计的问题。我们为此任务设计了一个新的体系结构, 该体系结构由子任务之间共享的要素网络、在要素网络之上构建的对象分类网络以及一系列与类别相关的姿势回归网络组成。我们还介绍了合适的丢失函数和新架构的训练方法。关于具有挑战性的 pascal 3d+ 数据集的实验显示了联合分类和姿态估计任务中最先进的性能。此外, 我们在联合任务上的表现与最先进的方法在更简单的三维姿态估计和已知对象类别任务上的表现相当。

少

2018 年 7 月 20 日提交;v1 于 2017 年 11 月 20 日提交;最初宣布 2017 年 11 月。

## 81. v2v-posenet:voxr 到 voxel 预测网络, 用于从单一深度图进行精确的三维手和人的姿态估计

作者:景植月,朱勇昌,孔穆李

**摘要:** 现有的基于深度学习的 3d 手和人体姿态估计方法大多基于一个通用框架, 该框架采用 2d 深度图, 并直接回归关键点的 3d 坐标, 例如手或人体关节, 通过二维卷积神经网络 (cnn)。这种方法的第一个弱点是在二维深度图中存在透视失真。虽然深度贴图本质上是三维数据, 但以前的许多方法都将深度贴图视为二维图像, 可以通过从 3d 到 2d 空间的投影来扭曲实际对象的形状。这迫使网络执行透视失真不变估计。传统方法的第二个缺点是, 直接从二维图像中回归三维坐标是高度非线性映射, 这给学习过程带来了困难。为了克服这些弱点, 我们首先将 3d 手和人的姿势估计问题从一个深度映射转换为体素到体素的预测, 该预测使用 3d 体素网格, 并估计每个关键点的每个体素的可能性。我们将我们的模型设计为 3d cnn, 在实时运行的同时提供准确的估计。我们的系统在几乎所有公开的 3d 手和人体姿势估计数据集中的性能优于以前的方法, 并在拿手 2017 框架的 3d 手写体估计挑战中排名第一。该代码可 [https://github.com/mks0601/V2V-PoseNet\\_RELEASE](https://github.com/mks0601/V2V-PoseNet_RELEASE)。少

2018年8月16日提交;v1于2017年11月20日提交;最初宣布2017年11月。

## 82. 从数以百万计的 3d 扫描中学习大规模 3d 人脸识别

作者:syed zulqamain gilani, ajmal mian

**摘要:** 在数以百万计的面部图像上训练的深层网络被认为在人脸识别方面接近人类水平的性能。然而, 开放的世界面临的承认仍然是一个挑战。尽管 3d 人脸识别比 2d 识别具有固有的优势, 但由于无法获得大型培训和大型测试数据集, 它并没有从深度学习的最新发展中受益。由于现有的 3d 人脸数据集的图库尺寸较小, 识别精度已经饱和。与 2d 照片不同, 3d 人脸扫描不能来自网络, 这对深度 3d 人脸识别网络和数据集的开发造成了瓶颈。在此背景下, 我们提出了一种生成大量标记的 3d 人脸身份及其用于训练的多个实例的方法, 以及用于合并最具挑战性的**现有 3d** 数据集进行测试的协议。我们还提出了第一个专门为 3d 人脸识别设计的深 **cnn** 模型, 并对 1 亿点 3dd 面部扫描进行了 1000k 身份的**培训**。我们的测试数据集包括 1, 853个**标识**, 在库中进行一次 3d 扫描, 并将另一个 31k 扫描作为探头, 比现有的扫描大几个数量级。如果不对此数据集进行微调, 我们的网络的性能已经超过了最先进的人脸识别的 10% 以上。我们在图库上对网络进行微调, 以执行端到端大规模 3d 人脸识别, 从而进一步提高准确性。最后, 我们展示了我们的方法对开放世界人脸识别问题的有效性。少

2018 年 7 月 5 日提交;v1 于 2017 年 11 月 16 日提交;最初宣布 2017 年 11 月。

### 83. 一种基于三维 2d 的道路自由空间探测联合方法

作者 :suvam pata, pranjali maheshwari, shashank yadav, chetan arora, subhassssbanerjee

文摘: 本文讨论了自动驾驶背景下的道路分割和自由空间检测问题。传统方法要么使用三维 (3d) 线索, 如从激光雷达、radar 或立体摄像机获得的点云, 要么使用二维 (2d) 线索 (如车道标记、道路边界和物体检测)。典型的 3d 点云没有足够的分辨率来检测高度的细微差异, 如道路和路面之间的差异。基于图像的 2d 线索失败时, 遇到不均匀的道路纹理, 如由于阴影, 坑洞, 车道标记或道路恢复。我们提出了一种结合二维和三维线索的自由道路空间探测技术。特别是, 我们使用基于 cnn 的道路分割从 2d 图像和平面盒拟合稀疏深度数据从 slam 作为前提, 以制定一个能量最小化使用条件随机场 (crf) 的道路像素分类。虽然美国有线电视新闻网学习道路纹理, 不受深度边界的影响, 3d 信息有助于克服基于纹理的分类失败。最后, 利用单目 slam 的三维深度数据对所获得的道路分割进行了检测, 用于导航目的。我们在 kitti 气味测量数据集、camvid 数据集以及我们拍摄的视频上的实验验证了所提出的方法相就是最先进的方法的优越性。少

2018 年 1 月 15 日提交;v1 于 2017 年 11 月 6 日提交;最初宣布 2017 年 11 月。

#### 84. rgb 深度摄像机网络中的实时无标记多人 3d 姿态估计

作者:marco carraro, matteo munaro, jeff burke , Emanuele menegatti

文摘: 本文提出了一种新的系统来估计和跟踪多人在标定 rgb 深度摄像网络中的三维姿态。每个人的多视图 3d 姿态由一个中心节点计算, 该节点接收来自网络每个摄像机的单视图结果。每个单视图结果都是通过使用 cnn 进行二维姿态估计并通过传感器深度将结果的骨架扩展到三维来计算的。该系统是无标记的、多人的, 独立于背景, 不对人的外表和初始姿势做出任何假设。该系统提供实时结果, 因此非常适合需要用户交互的应用程序。实验结果表明, 这项工作在不同情况下对基线多视图方法的有效性。为了在这项工作的基础上促进研究和应用, 我们在 openprtrack 中发布了源代码, 这是一个面向 rgb-d 人员跟踪的开源项目。少

2017 年 10 月 17 日提交;最初宣布 2017 年 10 月。

#### 85. 交互式医学图像分割, 采用深度学习和特定于图像的微调

作者 :guotaiwang, wuqili , maria a. zuluaga, rosalind pratt, preal a.patel, michael aertsen, tom doel, anna l.david, jan depreest, sebastien ouselin, tom Vercauteren

**文摘** 卷积神经网络 (cnn) 在自动医学图像分割方面取得了最先进的性能。然而, 它们没有表现出足够准确和稳健的结果供临床使用。此外, 由于缺乏特定于图像的适应, 以及缺乏对以前看不见的对象类的通用性, 它们受到限制。为了解决这些问题, 我们提出了一个新的基于深度学习的交互式分割框架, 将 cnn 集成到一个边界框和基于涂鸦的分割管道中。我们建议特定于图像的微调, 使 cnn 模型适应特定的测试图像, 可以是无监督的 (没有额外的用户交互) 或监督 (与额外的涂鸦)。我们还提出了一个加权损失函数, 考虑网络和基于交互的不确定性的微调。我们将这一框架应用于两种应用: 胎儿 mr 切片中多个器官的**二维**分割, 其中只有两种类型的器官被注释为训练; 不同 mr 序列中的脑肿瘤核 (排除水肿) 和全脑肿瘤 (包括水肿) 的**三维**分割, 其中只对一个 mr 序列中的肿瘤核进行注释进行训练。实验结果表明, 1) 我们的模型比最先进的 cnn 更鲁分以前看不见的物体; 2) 利用所提出的加权损失函数进行图像特定微调, 显著提高分割精度; 与传统的交互式分割方法相比, 我们的方法能以更少的用户交互和更少的用户时间获得准确的结果。少

**2017 年 10 月 11 日提交;最初宣布** 2017 年 10 月。

## 86. 野外视频情感分类的时间多模态融合

**作者:** [valentin vielzeuf](#), [stphane pateux](#), [Stéphane jurie](#)

**摘要:** 本文讨论了情绪分类的问题。这项任务包括预测情绪标签 (采取在一组可能的标签中) 最好地描述在短视频剪辑中包含的情绪。本文在标准框架的基础上——通过监督分类器用来推断标签的视听特征来描述视频——研究了几个新的方向。首先, 提出了一种改进的基于二维和三维并行神经网络的人脸描述符。其次, 本文探讨了几种融合方法, 即时间和多式联运, 包括一种新的层次方法, 结合特征和分数。此外, 我们仔细审查了管道的不同阶段, 并设计了适合这项任务的**美国有线电视新闻网**架构;这一点很重要, 因为与问题的难度相比, 训练集的规模很小, 使得泛化变得困难。在 2017 年的《狂野挑战中》中, 获得的模型排名第 4 位, 准确率为 38.8%。少

**2017 年 9 月 21 日提交;最初宣布** 2017 年 9 月。

## 87.从面向任意的 2d 图像看规范协调空间中的三维重构

**作者 :** Hajnalhou, bheesh khalal, amir alansary, steven mcdonagh, alice davidson david, mary rutherford, jo v. hajnal, daniel rueckert, ben glocker, bernhard kainz

**摘要:** 有限的捕获范围, 以及为基于优化的 **2d/3d** 图像配准方法提供高质量初始化的要求, 会显著降低 **3d** 图像重建和运动的性能补偿管道。具有挑战性的临床成像场景, 其中包含重要的主题运动, 如胎子宫成像, 使 **3d** 图像和体积重建过程复杂化。本文针对学习的规范地图集坐标系统, 提出了一种基于学习的图像配



准方法, 能够预测任意定向**二维**图像切片的**三维**刚性变换。只有图像切片强度信息用于执行配准和规范对齐, 不需要空间变换初始化。为了找到图像转换, 我们利用卷积神经网络 (**cnn**) 架构来学习**能够将二维图像切片映射到三维规范地图集空间的回归函数**。我们对我们的方法在模拟磁共振成像 (mri)、胎儿大脑成像和合成运动中的有效性进行了定量评估, 并进一步展示了实际胎儿 mri 数据的定性结果, 在这些数据中, 我们的方法是集成的到一个完整的重建和运动补偿管道。我们基于学习的注册在模拟数据上实现了 7 毫米的平均空间预测误差, 并为胎龄约为 20 周的大量移动的胎儿提供了质量改进的重建。我们的模型为 **2d/3d** 注册初始化问题提供了通用且计算效率高的解决方案, 适用于实时方案。少

**2018 年 1 月 23 日提交**;v1 于 2017 年 9 月 19 日提交;**最初宣布** 2017 年 9 月。

## 88. 利用条件对抗性网络改进异质人脸识别

**作者:**张武明,舒志新, dimitris samaras, llimingchen

**摘要:** 彩色图像和深度图像之间的异构人脸识别是现实世界应用中非常需要的容量, 在这些应用中, 形状信息被视为仅仅涉及画廊。本文提出了一种跨模态深度学习方法, 作为应对这一挑战的有效有效的解决方法。具体来说, 我们从学习两个卷积神经网络 (**cnn**) 开始, 分别提取 **2d** 和 **2.5 d** 面特征。一旦训练, 他们可以



作为预先培训的模型, 为另一个双向 **cnn** 探索**颜色**和深度之间的相关性部分的异构匹配。与大多数传统的交叉模态方法相比, 我们的方法还能利用条件生成对抗性网 (cgan) 对单色图像进行精确的深度图像重建, 并进一步提高了识别性能。融合多模态匹配结果。通过对基准 frgc **2d/3d** 人脸数据库的定性和定量实验, 证明了所提出的管道在异构人脸识别方面优于最先进的性能, 并确保极高的在线舞台。少

**2017 年 9 月 13 日提交**;v1 于 2017 年 9 月 8 日提交;**最初宣布** 2017 年 9 月。

## 89. 美国有线电视新闻网关于 **velodyne lidar** 数据中的快速地面分割

**作者:**[martin velas](#), [m 奇勒·斯帕内尔](#), [m 奇勒·赫拉迪斯](#),[亚当·赫鲁特](#)

**文摘:** 本文提出了一种新的在 **velodyne** 点云中进行地面分割的方法。我们提出了一种编码稀疏的 **3d** 数据从 **velodyne** 传感器适合训练卷积神经网络 (**cnn**)。这种通用方法用于将稀疏点云分割为地面和非地面点。**lidar** 数据表示为多通道**二维信号**, 其中水平轴与旋转角度相对应, 垂直轴与索引通道 (即激光束) 相对应。使用我们准备的手动注释数据集对相对较浅的 **cnn** (即 3–5 卷积层) 的多个拓扑进行训练和评估。结果表明, 与最先进的方法相比,

张等人在速度方面有了显著的提高, 在精度方面也有了小幅的提高。少

2017 年 9 月 7 日提交;最初宣布 2017 年 9 月。

## 90. 三维组件重构和 cnn 交叉姿态识别的快速地标定位

作者:谢格琳,许,谢鸿成,谢成华

文摘: 提出了两种交叉姿态人脸识别方法, 一种是基于人脸成分的三维重建, 另一种是基于深卷神经网络 (cnn)。与大多数考虑整体人脸的 3d 方法不同, 建议的方法考虑 3d 面部组件。它将 2d 图格面分割成组件, 重新构造每个组件的 3d 曲面, 并通过组件特征识别探头面。分割是基于定位的地标定位的分层算法, 结合了更快的 r-inn 的人脸检测和减少树结构化模型的地标定位。基于美国有线电视新闻网的方法的核心部分是修订 vgg 网络。我们研究了训练集中不同设置的性能, 包括三维重建的综合数据、野外数据库中的真实数据, 以及这两种类型的数据的组合。我们研究了网络作为分类器或设计为特征提取器时的性能。在广泛的实验中对两种识别方法和快速地标定位进行了评价, 并与最先进的方法进行了比较, 以证明其有效性。少

2017 年 8 月 31 日提交;最初宣布 2017 年 8 月。

## 91. 端到端逆渲染的深层结构

作者:[shima kamyab](#), [ali ghodsi](#), [s. zohreh azimifar](#)

**摘要:** 以 3d 格式进行反向渲染, 表示为恢复给定 2d 输入图像的  
场景的 3d 属性, 通常使用基于单个视图图像的 3d 模式 (3DMM)  
方法完成。这些模型将每个面作为从训练数据中提取的一些基础  
向量的加权组合。本文提出了一种通过训练自动编码器网络和卷  
积神经网络 (cnn) 来计算系数和基向量的深层框架。这样做的目  
的是找到一个共同的原因, 可以用深度网络映射到三维结构和相  
应的二维图像。实证结果验证了深层框架在从人脸合成数据集上  
相应的二维图像中寻找精确的人脸三维形状的力量。少

2017 年 8 月 25 日提交;最初宣布 2017 年 8 月。

## 92. 利用三维残差网络学习时空特征, 实现动作识别

作者:[koso hara](#), [hihikatsu kataoka](#), [yutaka satoh](#)

**摘要:** 具有时空三维核的卷积神经网络 (..。更多

2017 年 8 月 25 日提交;最初宣布 2017 年 8 月。

## 93. 脸网: 为无地标面对齐做一个案例

作者:[f 肛 juchang](#), [anh tuan tran](#), [tal h 斯纳](#), [iacopo masi](#), [ram nevatia](#), [gerard medioni](#)

**摘要:** 我们展示了一个简单的卷积神经网络 (cnn) 如何能够直接  
从图像强度中训练出精确而有力地回归 6 度自由度 (6DoF) 三维

头部姿势。我们进一步解释了如何使用这个 faceposenet (fpn) 对齐 2d 和 3d 中的人脸, 作为这些任务的显式面部地标检测的替代方法。我们声称, 在许多情况下, 测量地标探测器精度的标准方法在比较不同的面部排列时可能会产生误导。相反, 我们通过评估 fpn 与现有方法相比如何影响 ijb-a 和 ijb-b 基准的人脸识别精度: 使用相同的识别管道, 但改变人脸对齐方法。我们的研究结果表明: (a) 在 300w 基准上测量的具有里程碑意义的检测精度并不一定意味着更好的人脸识别精度。(b) 我们的 fpn 在两个基准上提供卓越的 2d 和 3d 人脸对齐功能。最后, (c), fpn 以相对精确的地标探测器的计算成本的一小部分对齐。因此, 在许多方面, fpn 是一种比使用面部地标探测器更快、更精确的人脸对齐方法。少

2017 年 8 月 31 日提交;v1 于 2017 年 8 月 24 日提交;最初宣布 2017 年 8 月。

#### 94. 基于 cn 的实时密集人脸重建与逆影照片逼真的人脸图像

作者:郭玉东,张建勇, 蔡建飞,姜博义,郑建民

文摘 基于 cnn 的人脸重建具有卷积神经网络 (cnn) 的强大功能, 近年来在 从二维人脸图像重建详细的人脸形态方面表现出了很有希望的性能。基于美国有线电视新闻网的方法的成功依赖于大量的标记数据。最先进的合成此类数据使用粗糙的可变形的人

脸模型, 但这很难产生详细的照片逼真图像的人脸(皱纹)。本文提出了一种新的人脸数据生成方法。具体来说, 我们基于逆渲染呈现大量具有不同属性的照片逼真人脸图像。此外, 我们还通过将不同比例的细节从一张图像传输到另一张图像来构建一个精细的人脸图像数据集。我们还通过模拟真实视频数据的分布来构造大量的视频类型相邻帧对。利用这些构造良好的数据集, 我们提出了一个由三个卷积网络组成的粗化学习框架。这些网络经过训练, 可从单目视频以及单个图像进行实时详细的三维人脸重建。广泛的实验结果表明, 我们的框架可以产生高质量的重建, 但与最先进的计算时间相比, 计算时间要短得多。此外, 由于数据的多样性, 我们的方法对姿势、表达和照明都是鲁棒性的。少

2018 年 5 月 15 日提交;v1 于 2017 年 8 月 2 日提交;最初宣布 2017 年 8 月。

## 95. 2d-3d 全卷神经网络用于心脏磁共振系统的分割

作者:jay patravali, shubhamjain , sasank chilamkurthy

文摘: 本文利用深卷积神经网络(cnn)开发了一种用于全自动心脏磁共振图像分割的二维和三维分割管道。我们的模型采用 acd 挑战 2017 年数据集进行端到端培训, 其中包括 100 项研究, 每项研究都包含端部舒张和端系统阶段的心脏 mr 图像。我们表明, 我们的两个分割模型在距离指标方面都实现了接近最先进的性能

得分,并在临床参数方面具有令人信服的准确性。通过引入一种新的骰子损失函数及其与交叉熵损失的组合,进行了比较分析。通过探索不同的网络结构和综合实验,我们讨论了获得最优模型性能的几个关键见解,这也是这一挑战主题的核心。少

2017 年 7 月 31 日提交;最初宣布 2017 年 7 月。

## 96. 实现精确的无标记人的形状和随时所需的随时所需时间的估计

作者:黄英豪,费德里卡·博戈, christoph lassner, angjoo kanazawa, peter v.gehler, ijaz akhter, michael j. black

**摘要:** 现有的无标记运动捕获方法通常假定已知的背景、静态摄像机和序列特定的运动原点,这会缩小其应用场景。在这里,我们提出了一个全自动的方法,给出**多视图**视频,估计三维人体运动和身体形状。我们以最近的 `smplify \cite{cemo2016}` 保持} 为基本方法,并以多种方式对其进行扩展。首先,我们将主体与在多视图图像中检测到的 **2d** 特征相匹配。其次,我们使用 **cnn** 方法对每个图像中的人进行分割,并将**三维**身体模型与轮廓相匹配,以进一步提高准确性。第三,我们利用一个通用的和强大的 `dct` 时间之前,处理左右交换问题,有时由**二维**姿势估计。标准基准测试显示,我们的结果可与最先进的状态相媲美,还提供了逼真的 **3d** 形状头像。我们还演示了在 `humaneva` 和挑战舞蹈序列从 `youtube` 在单目的情况下的准确结果。少

2018年4月30日提交;v1于2017年7月24日提交;最初宣布2017年7月。

## 97. 深度地理: 一种用于医学图像分割的深度交互式大地测量框架

作者: [guotaiwang](#), [maria a. zuluaga](#), [w 恩 qili](#), [rosalind pratt](#), [preal a.patel](#), [michael aertsen](#), [tom doel](#), [anna l.david](#), [jan deprest](#), [sebastien ouselin](#), [tom Vercauteren](#)

**摘要:** 准确的医学图像分割对于诊断、手术规划和许多其他应用至关重要。卷积神经网络 (cnn) 已成为最先进的自动分割方法。然而, 全自动结果可能仍需完善, 以变得足够准确和坚固, 供临床使用。我们提出了一种基于深度学习的交互式分割方法, 以提高自动 cnn 获得的结果, 减少用户在细化过程中的交互, 从而获得更高的准确性。我们使用一个 cnn 来获得一个初始的自动分割, 在这个过程中, 用户交互被添加到指示错误的分割上。另一个 cnn 作为输入用户与初始分割的交互, 并给出了一个完善的结果。我们建议通过测地线距离变换将用户交互与 cnn 结合起来, 并提出一个能提供更好的密度预测的分辨率保护网络。此外, 我们还将用户交互作为硬约束集成到可反传播的条件随机字段中。在胎儿 mri 二维胎盘分割的背景下, 从 flair 图像中验证了该框架的有效性。实验结果表明, 与传统的交互方法相比, 该方法与自动 cnn 相比取得了较大的改进, 获得了可比、更高的精度, 用户干预较少, 时间更短。少

2017 年 9 月 19 日提交;v1 于 2017 年 7 月 3 日提交;最初宣布 2017 年 7 月。

## 98. feasknet: 用于三维形状分析的特征驱动的图形卷积

作者:nitika verma, edmond boyer , jakob verbeek

**文摘** 卷积神经网络 (cnn) 对二维图像中的视觉识别产生了巨大的影响, 目前在最先进的方法中无处不在。但是, cnn 并不容易扩展到不代表常规网格的数据, 例如 3d 形状网格或其他图形结构数据, 传统的本地卷积运算符并不直接应用于这些数据。为了解决这个问题, 我们提出了一个新的图形卷积算子来建立滤波器权重和具有任意连接的图形邻域之间的对应关系。我们的方法的关键新颖之处在于, 这些对应是根据网络学习的特征动态计算的, 而不是像以前的工作那样依赖于图形上的预定义静态坐标。我们获得了出色的实验结果, 与以前最先进的形状对应结果相比, 有了显著的改进。这表明, 我们的方法可以从原始输入坐标中学习有效的形状表示, 而不依赖于形状描述符。少

2018 年 3 月 28 日提交;v1 于 2017 年 6 月 16 日提交;最初宣布 2017 年 6 月。

## 99. 三维人体姿态估计局部体积预测的整体平面预测

作者:京西月,朱勇昌,苏玉民, 孔穆李



**文摘:** 我们提出了一种新的方法, 从一个单一的深度图三维人体姿态估计。近年来, 卷积神经网络 (cnn) 已成为计算机视觉中的一个强大范式。许多计算机视觉任务都受益于 cnn, 但是, 传统的方法直接从图像中回归 3d 身体关节位置并不能显著提高性能。相反, 我们从三维占用度网格中将问题表述为估计关键体关节的每个体素的可能性。我们认为, 与从深度图到三维关节坐标的回归学习相比, 使用三维卷积学习从体积输入到体积输出的映射可以持续提高精度。我们提出了一种两阶段的方法来减少体积表示和三维卷积引起的计算开销: 整体二维预测和局部三维预测。在第一阶段, Planimetric 网络 (p-net) 估计整体二维空间中每个身体关节的每个像素的可能性。在第二阶段, 体积网络 (v-net) 估计第一阶段二维估计周围局部 3d 空间中每个体关节的每个体节理的可能性, 从而有效地降低了计算成本。在公开可用的数据集中, 我们的模型的性能大大优于现有方法。少

**2017 年 7 月 8 日提交;**v1 于 2017 年 6 月 15 日提交;**最初宣布** 2017 年 6 月。

## 100. 深度素描 2 面: 一种基于深度学习的三维人脸和漫画建模素描系统

**作者:**韩晓光,高昌,于一洲

**文摘:** 人脸建模在可视化计算领域受到了广泛的关注。有许多场景, 包括卡通人物, 社交媒体的化身, 3d 人脸漫画, 以及与面部相关的艺术和设计, 低成本的交互式面部建模是一种流行的方法, 尤其是在业余用户。本文提出了一种基于深度学习的三维人脸和漫画建模素描系统。该系统具有高效的绘制界面, 允许用户绘制代表面部特征轮廓的手绘不精确但富有表现力的 2d 线。设计了一种基于 cnn 的深回归网络, 用于从二维草图中推断三维人脸模型。我们的网络融合了 cnn 和基于形状的输入草图的特征, 并有两个完全连接的层的独立分支, 为双线性面表示生成独立的系数子集。我们的系统还支持基于手势的交互, 供用户进一步操作初始人脸模型。用户研究和数值结果都表明, 我们的素描系统可以帮助用户快速有效地创建人脸模型。构建了一个显著扩展的人脸数据库, 具有不同的身份、表达和夸张程度, 以促进人脸建模技术的进一步研究和评价。少

**2017 年 6 月 7 日提交;最初宣布** 2017 年 6 月。

## 101. 深度投影三维语义分割

**作者 :** [felix järemo lawin](#), [martin d 德尔 扬](#), [patrik tosteberg](#), [goutam bhat](#), [fahad shahbaz khan](#), [michael felsberg](#)

**摘要:** 3d 点云的语义分割是一个具有挑战性的问题, 适用于众多的真实世界应用。虽然深度学习彻底改变了图像语义分割领域,

但其对点云数据的影响迄今有限。最近的尝试, 基于 **3d** 深度学习方法 (**3d-cnn**), 已经取得了低于预期的结果。此类方法需要对基础点云数据进行体外化, 从而降低空间分辨率并增加内存消耗。此外, **3d-cnn** 受到注释数据集有限的影响。在本文中, 我们提出了一个替代框架, 以避免 **3d-cnn** 的限制。我们首先将点云投影到一组合成的 2d 图像上, 而不是直接在 **3d** 中解决问题。然后将这些图像用作 2d-cnn 的输入, 该 **cnn** 是为语义分割而设计的。最后, 将得到的预测分数重新投影到点云, 以获得分割结果。我们进一步研究了多流网络体系结构中多种模式 (如颜色、深度和表面法线) 的影响。在最近的语义 3d 数据集上进行了实验。与以前的最佳方法相比, 我们的方法实现了 7.9% 的相对收益, 从而创造了一个新的最新技术。少

提交于 2017 年 5 月 9 日;最初宣布 2017 年 5 月。

## 102.vnect: 使用单个 rgb 摄像机进行实时三维人体姿态估计

作者 :dushyant mehta, srinath sridhar , oleksandr sotnychenko, helgerhodin, mohammad shafiei, hans-peter seidel, wi peng xu, dan casas, christian theobalt

摘要: 我们提出了第一个实时方法, 以稳定的、时间一致的方式使用单个 rgb 相机捕获人类的完整的全球 **3d** 骨骼姿势。我们的方法结合了一种新的卷积神经网络 (**cnn**) 基于姿态回归与运动骨架拟合。我们新颖的全卷形姿势配方可实时同时重度回归 **2d** 和

**3d** 关节位置, 不需要紧密裁剪的输入帧。一种实时运动骨架拟合方法利用 **cnn** 输出在相干运动骨架的基础上生成时间稳定的**三维**全局姿态重构。这使得我们的方法成为第一个可用于实时应用(如**3d**字符控制)的单目 **rgb** 方法——迄今为止, 这类应用的唯一单目方法采用了专门的 **rgb-d** 摄像机。我们的方法的精度与最佳的**离线 3d** 单目 **rgb** 姿态估计方法在定量上是一致的。我们的结果在质量上是可比的, 有时比单目 **rgb-d** 方法的结果要好, 如 **kinect**。然而, 我们表明, 我们的方法比 **rgb-d** 解决方案更广泛地适用, 即适用于户外场景、社区视频和低质量商品 **rgb** 摄像机。少

**2017 年 5 月 3 日提交;最初宣布** 2017 年 5 月。

### 103. 基于 3d cnn 的多任务学习对肺结节的风险分层

**作者:**[sarfaraz hussein](#), [kunlincao](#), [qi song](#), [ulas bagci](#)

**摘要:** 肺结节的危险分层是肺癌诊断中最重要的任务。对坚固和准确的结核特性的任何改进都有助于确定癌症阶段、预后和改进治疗规划。在本研究中, 我们提出了一个基于**三维**卷积神经网络(**cnn**)的结核表征策略。通过完全的**3d**方法, 我们利用 **ct** 扫描中的体积信息, 否则将在传统的 **2dcnn** 方法中丢失。为了满足**美国有线电视新闻网**大量培训数据的需求, 我们采取转移学习的方式, 以获得高度歧视的特征。此外, 我们还获得了六个高级节点属

性的任务相关特征表示, 并通过多任务学习 (mtl) 框架融合这些互补信息。最后, 我们建议纳入放射科医生之间的潜在分歧, 同时在图形的规律性稀疏多任务学习中评分不同的结核属性。我们评估了我们提出的方法, 在包括 1018 扫描在内的最大的公开肺结核数据集上, 并获得了最先进的结果, 以降低恶性肿瘤评分。少

2017 年 4 月 27 日提交;最初宣布 2017 年 4 月。

#### 104. 利用深卷积神经网络分割 mr 图像中的股骨近端

作者: [cem m. deniz](#), [siyuu 巷](#), [spencer hallyburton](#), [arakua welbeck](#), [stephen honig](#), [kyunghyuncho](#), [gregory chang](#)

**摘要:** 磁共振成像 (mri) 已被提出作为一个免费的方法来测量骨骼质量和评估骨折风险。然而, 手动分割骨磁共振图像是耗时的, 限制了 mri 测量在临床实践中的使用。本文的目的是提出一种基于深卷积神经网络 (cnn) 的股骨近端自动分割方法。这项研究得到了机构审查委员会的批准, 并获得了所有科目的书面知情同意。专家对 86 例研究对象股骨近端的体积结构 mr 图像数据集进行了人工分割。我们通过训练两个不同的 cnn 架构和多个初始要素图和图层来进行实验, 并使用四倍的手动分割金标准测试它们的分割性能交叉验证。股骨近端自动分割达到 0.94 的高骰子相似性评分 $\pm 0.05$  精度 =  $0.95 \pm 0.02$ , 召回 =  $0.02 \pm 0.08$  使用基于 3d 卷积的 cnn 架构, 超过 2d cnn 的性能。cnn 提供的高分割精度有

可能有助于将骨质量的结构 mri 测量应用到治疗骨质疏松症的临床实践中。少

2018年3月20日提交;v1于2017年4月20日提交;最初宣布2017年4月。

## 105.使用深层神经网络进行端到端 3d 人脸重建

作者:[dou penfei](#), [shishir k.shah](#), [ioannis a. kakadiaris](#)

**摘要:** 单目三维人脸形状重建从单一的二维人脸图像已成为一个活跃的研究领域, 因为它的广泛应用。在深度神经网络 (dnn) 成功的启发下, 我们提出了一种基于 dnn 的方法, 用于从单个二维图像进行端到端 3d 失败重建 (uh-e2far)。与最近使用 rgb 图像和初始 3d 面部形状渲染以迭代方式重建和细化 3d 人脸的作品不同, 我们的 dnn 模型是端到端的, 因此也是复杂的 3d 渲染过程是可以避免的。此外, 我们在 dnn 架构中集成了两个组件, 即多任务损失函数和融合卷积神经网络 (cnn), 以改善面部表情重建。利用多任务损失函数, 将三维人脸重建分为中性三维人脸形状重建和表达性三维人脸造型重建。中性的 3d 面部形状是类特定的。因此, 更高的图层特征非常有用。相比较而言, 富有表现力的 3d 面部形状更倾向于低层或中间层特征。利用融合-cnn, 融合和转换来自不同中间层的特征, 以预测三维表情面部形状。通过

大量的实验, 我们证明了我们的端到端框架在提高三维人脸重建精度方面的优越性。少

2017 年 4 月 17 日提交;最初宣布 2017 年 4 月。

## 106.hand3d: 基于三维神经网络的手姿势估计

作者:邓晓明,杨少明,张银达,谭平,梁昌,王红安

文摘: 提出了一种新的三维神经网络结构, 用于从单个深度图像进行三维手模估计. 与以前的作品不同的是, 我们主要在 2d 深度图像域上运行, 需要中间或后期从 3d 空间引入监控, 我们将深度图转换为 3d 体积表示, 并将其输入到一个 3d 卷积神经网络(cnn), 直接产生的姿势在 3d 不需要进一步的过程。我们的系统不需要地面真值参考点进行初始化, 我们的网络架构自然集成了 3d 空间中的本地功能和全局环境。为了增加训练数据手摆空间的覆盖范围, 我们通过从现有的真实图像数据集中传输手摆来渲染合成深度图像。我们在两个公共基准上评估我们的算法, 并获得最先进的性能。合成的手摆数据集将可用。少

2017 年 4 月 7 日提交;最初宣布 2017 年 4 月。

## 107. 从偏心 rgb-d 传感器遮挡下的实时手部跟踪

作者 :franziska mueller, dushyant mehta , oleksandr sotnychenko, srinath sridhar, dan casas, christian theobalt

**摘要:** 我们提出了一种在杂乱的真实环境中移动自我中心 rgb-d 摄像机的实时、可靠和准确的手部姿态估计方法。对于从自我中心视点成像的杂乱场景中的手对象交互, 现有方法通常失败, 这在虚拟或增强现实应用中很常见。我们的方法使用两个后来应用的卷积神经网络 (cnn) 来本地化手和渐变 3d 关节位置。手部定位是通过**使用 cnn 来估计手中心在输入中的 2d 位置**, 即使在杂波和遮挡的情况下也是如此。局部手部位置, 连同相应的输入深度值, 用于生成归一化裁剪图像, 输入到第二个 **cnn 实时回归相对 3d 手关节位置**。为了提高精度、鲁棒性和时间稳定性, 我们使用运动姿态跟踪能量来细化姿态估计。为了训练 **cnn**, 我们引入了一个新的真实数据集, 它使用合并的现实方法来捕获和合成大量在杂乱场景中自然手部相互作用的附加注释数据。通过定量和定性评价, 表明我们的方法对物体的自遮挡和遮挡具有鲁棒性, 特别是在移动自我中心的角度。少

**2017 年 10 月 5 日提交;**v1 于 2017 年 4 月 7 日提交;**最初宣布** 2017 年 4 月。

**108.bb8: 一种可扩展、准确、坚固到局部遮挡的方法, 用于预测具有挑战性的物体的 3d 路径, 而无需使用深度**

**作者:**[mahdi rad](#), [vincent lepetit](#)



**文摘:** 本文介绍了一种仅从彩色图像中进行**三维**目标检测和姿态估计的新方法。我们首先使用分割来检测 **2d** 中感兴趣的对象, 即使存在部分遮挡和杂乱的背景也是如此。与最近基于拼接的方法相比, 我们依靠一种 "整体" 方法: 我们将一个卷积神经网络 (**cnn**) 应用于被检测到的物体, 该网络以角的**二维**投影的形式预测它们的**三维**姿势。他们的 **3d** 边界框。然而, 这不足以处理来自最近 t-less 数据集的对象: 这些对象表现出旋转对称的轴, 并且在两个不同的姿势下, 这种物体的两个图像的相似性使**训练美国有线电视新闻网**具有挑战性。我们通过限制用于训练的姿势范围来解决这个问题, 并引入分类器来识别姿势在运行时的范围, 然后再对其进行估计。我们还使用可选的附加步骤来细化预测的姿势。我们将 linemod 数据集上的最先进技术从正确注册的 rgb 帧的 73.7 提高到 99.3%。我们也是第一个仅使用彩色图像在遮挡数据集上报告结果的公司。在 t-less 数据集的多个序列上, 我们平均获得 54% 的帧通过 pose 6d 标准, 而在同时使用颜色和深度的相同序列上, 我们获得了 67% 的最先进帧。完整的方法也是可扩展的, 因为可以同时针对多个对象训练单个网络。少

**2018年3月26日提交;v1** 于 2017年3月31日提交;**最初宣布** 2017年3月。

## 109. 深层 3d 人脸识别

作者:[kim donghyun](#), [matthias hernandez](#), [junmoo choi](#), [gerard medioni](#)

**文摘:** 提出了一种利用深层卷积神经网络 (dcnn) 和**三维**增强技术的**三维**人脸识别算法。利用深部神经网络的代表性和大规模标记训练数据的使用,**提高了二维**人脸识别算法的性能。与**二维**人脸识别相比, 由于缺乏大规模的**三维**人脸数据集, 三维人脸识别的训练识别深度特征非常困难. 本文表明, 通过对 **cnn** 进行相对较少的**3d**人脸扫描微调, 从**美国有线电视新闻网**接受 **2d**人脸图像训练的迁移学习可以有效地实现三维人脸识别.我们还提出了一种**3d**人脸增强技术, 它从一个**3d**人脸扫描中合成了许多不同的面部表情。我们提出的方法在 busporus、bo-3dfe 和 **3d-tec** 数据集上显示了出色的识别结果, 而无需使用手工制作的功能。使用我们的深层功能进行的 **3d** 标识也可以很好地扩展大型数据库。少

**2017 年 3 月 30 日提交;最初宣布** 2017 年 3 月。

## 110. [点击这里](#): 人的本地化关键点作为观点估计的指南

作者:[ryan szeto](#), [jason j. corso](#)

**摘要:** 我们激励并解决了单目视点估计任务的人与人的变化, 在这个变量中, 一个语义对象关键点的位置和类在测试时可用。为了利用关键点信息, 我们设计了一个卷积神经网络, 称为点击这

里 **cnn (ch-inn)**, 它集成了关键点信息与来自处理图像的层的激活。它将关键点信息转换为 **2d** 地图, 可用于更重地称重图像某些部分的要素。这些空间特征的加权和与全局图像特征相结合, 为预测图层提供相关信息。为了训练我们的网络, 我们在数千个 cad 模型上收集了一个新的 **3d** 关键点注释数据集, 并利用 **2d** 关键点信息综合呈现了数百万张图像。在 pascal **3d+** 的测试实例中, 我们的模型实现了 90.7 的平均类精度, 而最先进的基线仅获得了 85.7% 的平均类精度, 这证明我们的论点是人为循环推理的。少

**2017 年 8 月 4 日提交;**v1 于 2017 年 3 月 28 日提交;**最初宣布** 2017 年 3 月。

## 11. 通过自适应加权学习进行唾液引导视频分类

作者:[赵云珍](#),[彭玉欣](#)

**摘要:** 视频分类在许多实际应用中都很有效, 最近的深度学习大大提高了视频分类的准确性。然而, 现有的作品往往不加区别地模拟视频帧, 但从运动的角度来看, 视频帧可以自然分解为突出和非突出的区域。显著区域和非突出区域应使用不同的网络进行建模, 因为前者既呈现外观信息, 也呈现运动信息, 后者则呈现静态背景信息。为了解决这一问题, 本文首先用没有监督的光流预测视频的显著性。然后, 对两个 **3d cnn** 流进行了原始帧和

突出区域上的光流的单独训练, 另一个 **2d cnn** 对非突出区域的原始帧进行了训练。由于这三个流对每个类扮演不同的角色, 因此对每个类的每个流的权重都进行了自适应学习。实验结果表明, 唾液引导建模和自适应加权学习可以相互强化, 取得了最先进的结果。少

**2017年3月25日提交**;v1于2017年3月23日提交;**最初宣布** 2017年3月。

## 112. 通过直接体积 cnn 回归从单个图像中重建大姿态三维人脸

**作者** :aaron s. jackson, adrian bulat, vasileios Argyriou, georgios tzimiropoulos

**文摘:** 三维人脸重建是一个非常困难的基本计算机视觉问题。目前的系统通常假定有多个面部图像(有时来自同一主题)作为输入, 并且必须解决一些方法上的挑战, 例如在较大的面部姿势、表情和照明不均匀。一般来说, 这些方法需要复杂和低效的管道来构建和拟合模型。在这项工作中, 我们建议通过在由 **2d** 图像和 **3d** 面部模型或扫描组成的适当数据集上训练卷积神经网络 (**cnn**) 来解决其中的许多限制。我们的 **cnn** 只处理一个 **2d** 面部图像, 不需要精确的对齐, 也不需要图像之间建立密集的对对应关系, 适用于任意的面部姿势和表情, 可以用来重建整个**三维**面部几何(包括面部的不可见部分)绕过**三维**可变形模型的构造(在训练过程中)和拟合(在测试期间)。我们通过一个简单的 **cnn** 架构来实

现这一点, 该架构从一个 **2d** 图像中直接回归三维人脸几何的体积表示。我们还演示了如何将面部地标定位的相关任务纳入建议的框架, 并帮助提高重建质量, 特别是对于大型姿势和面部表情的情况。测试代码将与预先培训的模型一起在线提供 <http://aaronsplace.co.uk/papers/jackson2017recon> 少

**2017 年 9 月 8 日提交;v1** 于 2017 年 3 月 22 日提交;**最初宣布** 2017 年 3 月。

### 113. 磁共振成像中脑提取的自动上下文卷积神经网络 (自动网络)

**作者:** [seyed sadegh mohseni Sadegh](#), [deniz erdogmus](#), [ali Gholipour](#)

**摘要:** 在许多神经图像分析管道中, 大脑提取或全脑分割是重要的第一步。因此, 大脑提取的准确性和鲁棒性对于整个大脑分析过程的准确性至关重要。为了在本研究中设计一种基于学习、几何无关和无注册的大脑提取工具, 我们提出了一种基于自动上下文卷积神经网络 (**cnn**) 的技术, 其中固有的局部和全局图像特征是通过不同窗口大小的 **2d** 补丁学习的。在这个架构中, 三个不同方向 (轴向、冠状和矢状) 的三个并行二维卷积路径隐式学习 **3d** 图像信息, 而不需要计算开销大的 **3d** 图像信息卷积。网络生成的后概率图与原始图像补丁一起被迭代用作上下文信息, 以了解大脑的局部形状和连接性, 从而从非脑组织中提取。我们从算法中获得的大脑提取结果优于最近在两个公开可用的基准数据集

(lpba40 和 oasis) 上报告的结果, 在这两个数据集中, 我们得到了 97.42% 的 dice 重叠系数和分别为 95.40。此外, 我们还评估了我们的算法在重建的胎儿脑磁共振成像 (mri) 数据集中提取任意定向胎儿大脑的挑战性问题中的性能。在此应用中, 我们的算法比其他方法 (dice 系数: 95.98) 表现要好得多, 在这种方法中, 由于 mri 中胎儿大脑的非标准定位和几何形状, 其他方法的性能较差。我们基于 cnn 的方法可以在具有挑战性的应用中提供准确的、与几何无关的大脑提取。少

2017 年 6 月 19 日提交;v1 于 2017 年 3 月 6 日提交;最初宣布 2017 年 3 月。

#### 114. 任意主体运动存在下的切片到卷转换的预测

作者 :Hajnalhou, amir alansary, steven mcdonagh, alice davidson, mary rutherford, jo v.hajnal, daniel rueckert, ben glocker, bernhard kainz

文摘: 本文旨在解决基于强度的 2d/3d 配准中的一个基本问题, 该问题涉及有限的捕获范围和最先进的图像配准方法非常好的初始化需求。我们提出了一种回归方法, 该方法学习预测来自 3d 卷的任意 2d 图像切片的旋转和平移, 该方法涉及一个学习的规范地图集坐标系。为此, 我们利用卷积神经网络 (cnn) 学习高度复杂的回归函数, 该函数将二维图像切片映射到其在 3d 空间中的正确位置和方向。我们的方法在具有挑战性的成像场景中具有

吸引力, 在这种情况下, 重要的主题运动会使 2d 切片数据中 3d 卷的重建性能复杂化. 我们广泛地评估我们的方法的有效性, 定量地模拟 mri 大脑数据与极端随机运动。我们进一步展示了胎儿 mri 的定性结果, 我们的方法被集成到一个完整的重建和运动补偿管道。利用 **cnn** 回归方法, 我们在模拟数据上获得了 7 毫米的平均预测误差, 并在以前的方法失败的情况下, 对非常年轻的胎儿的图像进行了令人信服的重建质量。我们进一步讨论了计算机断层扫描和 x 射线投影的应用。我们的方法是 **2d/3d** 初始化问题的通用解决方案。它具有计算效率, 每片的预测时间为几毫秒, 适用于实时方案。少

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

## 115. 向合成人类学习

**作者** :gül varol, javier romero, xavier martin , naureen mahood, michael j.black, ivan laptev, cordelia schmid

**摘要**: 从图像和视频估计人体姿势、形状和运动是许多应用中的基本挑战。**2d** 人体姿态估计的最新进展使用大量的人工标记训练数据来学习卷积神经网络 (**cnn**)。此类数据的获取非常耗时, 也难以扩展。此外,**手动**标记 3d 姿势, 深度和运动是不切实际的。在这项工作中, 我们提出了 surreal (合成 hUmans foR 真实任务): 一个新的大规模数据集与综合生成, 但现实的图像的人从三维序

列的人类运动捕捉数据。我们生成了 600 多万帧, 以及地面真相姿势、深度地图和分段掩码。我们表明, 在我们的合成数据集上训练的 **cnn** 允许在真实的 **rgb** 图像中进行精确的人体深度估计和人的部分分割。我们的结果和新的数据集为使用廉价和大规模的合成数据推进人员分析开辟了新的可能性。少

**2018 年 1 月 19 日提交;****v1** 于 2017 年 1 月 5 日提交;**最初宣布** 2017 年 1 月。

## 116. 从深部提升: 从单个图像中进行的卷积三维姿态估计

**作者:**[denis tome](#), [chris russell](#), [lourdes agapito](#)

**文摘:** 针对单幅原始 **rgb** 图像的**三维**人体姿态估计问题, 提出了一个统一的公式, 该公式共同为**二维联合估计**和**三维姿态重建**提供了改进的原因。我们采用一种集成的方法, 将 **3d** 人类姿势的概率知识与多级 **cnn** 架构结合起来, 并利用可信的 **3d** 地标位置的知识来完善对更好的 **2d** 的搜索位置。整个过程是经过端到端训练的, 非常高效, 在人的人的 3.6 m 上获得最先进的结果, 优于以前在 **2d** 和 **3d** 错误上的方法。少

**2017 年 10 月 11 日提交;****v1** 于 2017 年 1 月 1 日提交;**最初宣布** 2017 年 1 月。



## 117. 动态动作识别: 一种用于临时组织联合定位数据的卷积神经网络模型

作者 :[adhavan jayabalan](#), [harish karunakaran](#) , [shravan murlidharan](#), [tesia shizume](#)

**摘要:** 动机: 在视频中识别人的行为是一项具有挑战性的任务, 在各个领域都有应用。以前在这一领域的作品要么使用过 2d 相机的图像, 要么使用三维相机的图像。很少有人使用这样的观点, 即人类的行为可以很容易地识别出关节在 3d 空间中的运动, 而是使用递归神经网络 (rnn) 进行建模。卷积神经网络 (cnn) 能够识别数据中的复杂模式, 使其适合于检测人类行为。因此, 我们模拟了一个美国有线电视新闻网, 它可以利用联合数据预测人类的活动。此外, 使用联合数据表示的优点是比图像或视频表示的维数更低。这使得我们的模型比 mn 模型更简单、更快速。在本研究中, 我们开发了一个六层卷积网络, 它将 15x961x4 形式的每个输入特征向量简化为一维二值向量, 从而给我们提供了预测的活动。结果: 我们的模型能够正确识别高达 87% 的精度度的活动。联合数据取自康奈尔大学的活动数据, 这些数据每天都有说话、放松、吃饭、做饭等活动。

2016 年 12 月 20 日提交;最初宣布 2016 年 12 月。

## 118. 基于形状概念的深度监控——遮挡感知三维对象分析

作者:chili, m. zeeshan zia, quoc-huy tran, 项宇, gregory d.hager, manmohan chandraker

**摘要:** 单目三维对象解析在各种场景中非常可取, 包括遮挡推理和整体场景解释。我们提出了一个深层卷积神经网络 (cnn) 架构, 在二维图像和 3d 空间中对语义部分进行本地化, 同时推断它们的可见性状态, 给出一个 rgb 图像。我们的关键见解是利用领域知识, 通过深入监管网络的隐藏层来规范网络, 以便依次推断与最终任务相关的中间概念。为了获得具有地面真实 3d 形状和相关概念的所需数量的训练数据, 我们渲染了三维对象 cad 模型, 以生成大规模合成数据, 并模拟具有挑战性的对象之间的遮挡配置。我们仅在合成数据上对网络进行培训, 并展示了最新的真实图像基准性能, 包括扩展版本的 kitti、pascal voc、pascal 3d+ 和 ikea, 用于 2d 和 3d 关键点定位和实例分割。实证结果证明了我们的深度监督方案的效用, 证明了知识从合成数据到真实图像的有效转移, 与标准的端到端培训相比, 不那么过度。少

2017 年 4 月 20 日提交;v1 于 2016 年 12 月 8 日提交;最初宣布 2016 年 12 月。

## 119. 同步光谱 cnn 用于 3d 形状分割

作者:李毅,郝苏,郭兴文,列奥尼达斯·吉巴斯

**摘要:** 本文研究了以形状图表示的**三维**模型的语义标注问题。功能视图被用来表示图形上的本地化信息, 因此, 零件段或关键点等注释只不过是 0-1 指示器顶点函数。与**二维**网格图像相比, 形状图是不规则的非同构数据结构。为了能够利用卷积神经网络对顶点函数进行预测, 我们采用光谱 **kernels** 方法, 通过对图形拉普拉斯特征基跨越的光谱域中的核进行参数化来实现权重共享。在此设置下, 我们的网络名为 syncspicnn, 努力克服两个关键挑战: 如何在图形的不同部分为单个形状共享系数和进行多尺度分析, 以及如何跨相关但不同的信息共享信息形状, 可能由非常不同的图形表示。针对这些目标, 我们引入了膨胀卷积核的光谱参数化和光谱变压器网络。我们通过实验测试了我们的 syncspecnn 的各种任务, 包括 **3d** 形状零件分割和 **3d** 关键点预测。在所有基准数据集上都实现了最先进的性能。少

**2016 年 12 月 2 日提交;最初宣布** 2016 年 12 月。

## 120. 利用改进的 cnn 监控技术对野外人工剂量进行单目三维估计

**作者:** [dushyant mehta](#), [helge rhodin](#), [dancasas](#), [pascal fua](#), [oleksandr sotnychenko](#), [weipeng hu](#), [christian theobalt](#)

**摘要:** 我们提出了一种基于 **cnn** 的 **3d** 人体姿势估计方法, 从单个 **rgb** 图像, 解决有限的模型只训练在明显有限的公开可用的 **3d** 姿势数据的概率有限的问题。仅使用现有的 **3d** 姿势数据和 **2d** 姿势数据, 我们通过传输学习到的特征, 在既定的基准上显示出

最先进的性能, 同时还将其推广到野外场景中。我们进一步介绍了一个新的训练集, 为人体姿态估计从单目图像的真实人类, 有地面真相捕获与多相机无标记的运动捕捉系统。它补充了现有的语料库, 在姿势、外表、服装、遮挡和观点方面有了更大的多样性, 并扩大了扩充的范围。我们还提供了一个涵盖室外和室内场景的新基准, 并展示了我们的 **3d** 姿势数据集显示出比现有的附加注释数据更好的野外性能, 这与从**二维**姿势数据。总之, 我们认为, 使用转换学习的表示与算法和数据贡献是至关重要的一般 **3d** 身体姿态估计。少

**2017 年 10 月 4 日提交;**v1 于 2016 年 11 月 29 日提交;**最初宣布** 2016 年 11 月。

## 121. 基于距离矩阵回归的单个图像的三维人体姿态估计

作者:[francesc Moreno-Noguer](#)

**抽象:** 本文从一张图像中讨论了**三维**人体姿态估计问题。我们遵循一个标准的**两步**管道, 首先检测二维位置  $n$  身体关节, 然后使用这些观测来推断 **3d** 姿势。第一步, 我们使用了最近的**基于美国有线电视新闻网**的探测器。对于第二步, 大多数现有方法执行  $2n$  至  $3n$  笛卡尔关节坐标的回归。我们表明, 可以通过使用  $n \times n$  将问题表述为**二维至三维**距离矩阵回归。为了学习这样的回归器, 我们利用简单的神经网络架构, 通过构造, 加强预测矩阵的正性和对

称性。该方法还具有自然处理缺失观测的优势，并允许假设未观察到的关节的位置。在 humaneva 和 human3.6 m 数据集上的量化结果表明，与最先进的数据集相比，性能有稳定的提高。利用人本组织在人学 3.6 m 上学到的回归器，对 lsp 数据集的野外图像进行定性评价，揭示了很有希望的泛化结果。少

2016 年 11 月 28 日提交;最初宣布 2016 年 11 月。

## 122. 利用混合模型 cnn 在图形和流形上进行几何深度学习

作者:[fedico monti](#), [davide boscaini](#), [jonathan masci](#), [Emanuele rodolá](#), [jan svoboda](#), [michael m. bronstein](#)

**摘要:** 深度学习在语音识别、自然语言处理和计算机视觉等多个领域取得了显著的性能突破。特别是，卷积神经网络（cnn）架构目前在各种图像分析任务（如对象检测和识别）上产生最先进的性能。到目前为止，大多数深度学习研究都集中在处理一维、二维或三维欧几结构数据，如声学信号、图像或视频。最近，人们对几何深度学习的兴趣越来越大，试图将深度学习方法推广到非欧几里得结构数据（如图形和流形），并从网络分析领域获得各种应用，计算社会科学，或计算机图形。在本文中，我们提出了一个统一的框架，允许将 cnn 架构推广到非欧几里得域（图和流形），并学习局部、固定和组合任务特定的特征。我们表明，以前在文献中提出的各种非欧几里得有线电视新闻网方法可以被视为我们框架的特殊实例。我们从图像、图形和三维形状分析的角度对所提出的

标准任务方法进行了测试, 并表明该方法的性能始终优于以往的方法。少

**2016 年 12 月 6 日提交;**v1 于 2016 年 11 月 25 日提交;**最初宣布** 2016 年 11 月。

## 123. 深时间线性编码网络

**作者:**ali diba, vivek sharma, luc van gool

**摘要:** 美国有线电视新闻网为代表人类行动而对整个视频中的特征进行编码的问题很少得到解决。相反,美国有线电视新闻网的工作重点是融合空间和时间网络的方法, 但这些方法通常仅限于处理较短的序列。我们提出了一个新的视频表示, 称为时间线性编码 (tle) 和嵌入在 **cnn** 作为一个新的层, 捕捉整个视频的外观和运动。它通过端到端学习将这些聚合信息编码为可靠的视频功能表示形式。tle 的优点是: (a) 它们将整个视频编码为紧凑的特征表示形式, 学习语义和判别性特征空间;(二) 它们适用于各种网络, 例如 **2d** 和 **3d cnn**,用于视频分类;(c) 它们以更有表现力的方式模拟特征交互, 而不丢失信息。我们对两个具有挑战性的人类行动数据集进行了实验: hmdb51 和 ucf101。实验表明, tle 在这两个数据集上的性能优于当前最先进的方法。少

**2016 年 11 月 21 日提交;**最初宣布 2016 年 11 月。

## 124. 从单一图像重建通风口: 道路场景理解的形状优先

作者: [j. krishna murthy](#), [g. v. sai krishna](#) , [falak chhaya](#), [k. Chhaya krishna](#)

**摘要:** 我们提出了一种在自动驾驶的背景下, 从单一 (rgb) 图像重建车辆的方法。尽管问题似乎不恰当, 但我们演示了有关如何使用车辆的 **3d** 形状投影到图像的知识, 以推理反向过程, 即形状 (反) 项目如何从 **2d** 到 **3d**。我们用 \ 强调 {个形状的原始} 对这些知识进行编码, 这些知识是在一个小的键点注释数据集上学习的。然后, 我们制定了一个形状感知调整问题, 该问题使用学习的形状原点从图像中恢复查询对象的 **3d** 姿态和形状。对于形状表示和推断, 我们利用最近的成功, 卷积神经网络 (**cnn**) 的任务对象和关键点定位, 并训练一个新的级联全卷化架构本地化车辆 \ 强调 {在图像中的关键点}。然后, 形状感知调整会强劲地恢复形状 (检测到的关键点的 **3d** 位置), 同时填充被遮挡的关键点。为了解决由于错误检测到的关键点而产生的估计错误, 我们使用迭代重新加权最小二乘 (**irls**) 方案进行鲁棒优化, 并作为每个预测关键点的副产品描述噪声模型。我们评估我们的自动驾驶基准的方法, 并提出优于现有的单目, 以及立体声的方法。少

**2016 年 9 月 29 日提交;最初宣布** 2016 年 9 月。

## 125. 学习相机的观点, 使用 **cnn** 提高 **3d** 身体姿势估计



**作者:**mona Fathollahi ghezelghieh, rangachar kasturi, sudeep sarkar

**摘要:** 这项工作的目的是从一个单一的 rgb 图像估计三维人体姿势。提取包含身体部位空间关系及其相对深度的图像表示, 在精确三维姿态重建中起着至关重要的作用。本文首次表明, 在不明确使用透视几何数学模型的情况下, 摄像机视点与二维关节-c 相结合, 显著提高了三维姿态精度。为此, 我们训练了一个深层的卷积神经网络 (cnn) 来学习分类相机的观点。为了使网络对图像中主体的服装和身体形状具有鲁棒性, 我们利用三维计算机渲染来合成额外的训练图像。我们在最大的 3d 姿势估计基准标记 (human3.6 米) 上测试我们的框架, 与不使用人体部件分割的最先进方法相比, 可实现高达 20% 的误差降低。少

2016 年 9 月 18 日提交;最初宣布 2016 年 9 月。

## 126. 语义融合: 基于融合神经网络的密集三维语义映射

**作者:**john mccormac, ankur handa, andrew davison , stefan leutenegger

**摘要:** 使用视觉传感进行的更强大、更准确和更详细的映射已被证明是移动机器人在各种应用中的一个有利因素。对于机器人智能和直观的用户交互的下一个级别, 地图需要超越几何和外观——它们需要包含语义。我们通过结合卷积神经网络 (cnn) 和最先进的密集同时定位和映射 (slam) 系统 elasticfusion 系统来应对这



一挑战, 该系统提供了室内框架之间的长期密集对应rgb-d 视频, 即使在环路扫描轨迹中也是如此。这些对应允许**美国有线电视新闻网**从多个视点的语义预测被概率地融合到地图中。这不仅会生成一个有用的语义 3d 映射, 而且我们还在 nyuv2 数据集上显示, 融合多个预测可使多个预测的**语义**标记比基线单帧预测得到改进。我们还表明, 对于预测视点变化较大的较小重建数据集, 对单帧分割的改进也会增加。我们的系统足够高效, 可以在约 25hz 的帧速率下进行实时交互使用。

**2016 年 9 月 28 日提交**;v1 于 2016 年 9 月 16 日提交;**最初宣布** 2016 年 9 月。

## 127. 基于二维姿态信息的卷积神经网络的三维人体姿态估计

**作者:**[成兴公园](#),[黄继熙](#), [野君夸克](#)

**文摘:** 卷积神经网络 (cnn) 在二维人体姿态估计方面取得了成功, 但**三维**人体姿态估计还没有得到深入的研究。在本文中, 我们解决了**三维**人体姿势估计任务与端到端学习使用 **cnn**。一个关节和其他关节之间的相对 3d 位置是通过 **cnn** 学习的。该方法以两种新颖的思路提高了**美国有线电视新闻网**的性能。首先, 我们添加了**二维**姿态信息, 通过将**二维**姿态估计结果与图像中的特征串联起来, 从图像中估计三维姿态。其次, 我们发现, 通过结合相对位置相对于多个关节的信息, 而不仅仅是一个根关节, 可以获得更

准确的**三维姿势**。实验结果表明, 该方法在人类 3.6 米数据集上达到了与最先进的方法相当的性能。少

**2016 年 9 月 7 日提交**;v1 于 2016 年 8 月 10 日提交;**最初宣布** 2016 年 8 月。

## 128. 保持它 smpl: 从单个图像的三维人体姿态和形状的自动估计

**作者:**[fedica bogo](#), [angjoo kanazawa](#), [christoph lassner](#), [peter gehler](#), [javier romero](#), [michael j. black](#)

**文摘:** 我们描述了第一种从单个无约束图像自动估计人体三维姿势及其三维形状的方法。我们估计了一个完整的 **3d** 网格, 并显示, 仅 **2d** 关节携带了大量有关身体形状的信息。这个问题是具有挑战性的, 因为人体的复杂性, 发音, 遮挡, 服装, 照明, 以及固有的模糊推断 **3d** 从 **2d**。为了解决这个问题, 我们首先使用最近公布的**基于 cnn 的方法** deepcut 来预测 (自下而上) 的 **2d** 人体关节位置。然后, 我们将最近公布的统计体型模型 (自上而下) 拟合 (自上而下) 到 **2d** 关节中, 称为 smpl。我们通过最小化一个目标函数来消除投影的 **3d** 模型关节和检测到的 **2d** 关节之间的误差来实现这一目标。由于 smpl 捕获了整个人群中人类形状的相关性, 因此我们能够将其有力地与很少的数据相匹配。我们进一步利用 **3d** 模型来防止导致相互渗透的解决方案。我们在利兹体育、humaneva 和 human3.6 m 数据集上评估我们的方法 smplify, 在最先进的领域中显示出卓越的姿势准确性。少

2016 年 7 月 27 日提交;最初宣布 2016 年 7 月。

## 129.融合网: 使用多个数据表示的 3d 对象分类

作者:[vishakh hegde](#), [reza zadeh](#)

**摘要:** 高质量的 3d 目标识别是许多视觉和机器人系统的重要组成部分。我们使用两种数据表示来解决对象识别问题, 以在普林斯顿 modelnet 挑战上取得领先的结果。两种表示形式: 1. 体积表示: 3d 对象在空间上被离散为二进制体素-1 如果体素被占用, 0 否则。2. 像素表示: 3d 对象表示为一组投影的 2d 像素图像。目前领先的提交模型网络挑战使用卷积神经网络 (cnn) 的像素表示。然而, 我们偏离了这一趋势, 另外, 使用体积 cn 来弥合上述两种表述的效率之间的差距。我们将这两种表示组合在一起, 利用它们来学习新功能, 这将产生一个比孤立使用这两种表示的分类器好得多的分类器。为了做到这一点, 我们介绍了新的体积 cnn (v-cnn) 架构。少

2016 年 11 月 26 日提交;v1 于 2016 年 7 月 19 日提交;最初宣布 2016 年 7 月。

## 130.基于地图引导的野外三维姿态估计数据增强

作者:[grégory roez](#), [cordelia schmid](#)

**文摘:** 本文讨论了野外**三维**人体姿态估计的问题。一个重大的挑战是缺乏训练数据, 即用**三维**姿势注释的人类 2d 图像. 这些数据对于培训最先进的 **cnn** 架构是必要的。在这里, 我们提出了一个解决方案, 以生成大量的逼真的合成图像的人类与 **3d** 姿态注释。我们引入了一个基于图像的合成引擎, 该引擎使用 3d 运动捕捉 (mocap) 数据, 利用 **2d** 人体姿态注释人为地增强了真实图像的数据集。给定**候选 3d** 姿态, 我们的算法为每个关节选择一个**图像, 其二维姿势在本地与投影的 3d 姿势相匹配**。然后将选定的图像组合在一起, 通过以运动受限的方式拼接本地图像补丁来生成新的合成图像。由此产生的图像被用来训练一个**端到端的美国有线电视新闻网**, 用于全身**三维姿势**估计。我们将训练数据分组到大量的姿势类中, 并将姿态估计作为  $k$  路分类问题来处理。这种方法只有在像我们这样的大型训练装置中才是可行的。我们的方法在受控环境中的**三维**姿态估计方面优于最新技术 (人人 3.6 m), 并显示了野外图像 (lsp) 的良好效果。这表明, 在人工图像上训练的 **cnn** 很好地概括了真实图像。少

**2016 年 10 月 28 日提交;**v1 于 2016 年 7 月 7 日提交;**最初宣布** 2016 年 7 月。

### 131. 一种用于从单个 rgb 图像估计细尺度深度图的双流网络

**作者:**李军,莱因哈德·克莱恩,姚安吉拉

**摘要:** 从单个 rgb 图像估计深度是一个不恰当的、本质上不明确的问题。最先进的深度学习方法现在可以估计准确的 2d 深度地图,但当地图投影到 3d 时,它们缺乏局部细节,并且经常被严重扭曲。我们提出了一个快速训练的两流**美国有线电视新闻网**,预测深度和深度梯度,然后融合在一起,形成一个准确和详细的深度地图。我们还定义了一个新的集损失在多个图像;通过规范一组常见图像之间的估计,网络比竞争方法更不容易过度拟合,并获得更好的精度。在纽约大学深度 v2 数据集上的实验表明,我们的深度预测与最先进的预测具有竞争力,并可实现忠实的 3d 预测。少

**2017 年 12 月 3 日提交;**v1 于 2016 年 7 月 3 日提交;**最初宣布** 2016 年 7 月。

### 132.v-net: 用于体积医学图像分割的全卷神经网络

**作者:**[fausto milletari](#), [nassir navab](#), [seyed-ahmad 艾哈迈迪](#)

**文摘:** 卷积神经网络 (cnn) 是近年来从计算机视觉和医学图像分析两个领域应用的。尽管大多数方法很受欢迎,但它们只能处理 2d 图像,而临床实践中使用的大多数医学数据都是由 3d 卷组成的。在本工作中,我们提出了一种基于体积、完全卷积、神经网络的**三维**图像分割方法。我们的 cnn 接受了描述前列腺的 mri 卷的端到端培训,并学习同时预测整个卷的分割。我们引入了一个新的目标函数,我们在训练过程中优化,基于 dice 系数。这样,

我们就可以处理前景和背景体素数量之间存在严重不平衡的情况。为了应对可用于训练的注释卷数量有限的问题，我们利用随机非线性变换和直方图匹配来增强数据。我们在实验评估中表明，我们的方法在具有挑战性的测试数据上取得了良好的性能，而只需要以前其他方法所需的处理时间的一小部分。少

2016 年 6 月 15 日提交;最初宣布 2016 年 6 月。

### 133. 密集体积与体积血管边界检测

作者 :[jameson merkow](#), [david kriegman](#), [alison marsden](#), [Zhuowen tu](#)

文摘: 在这项工作中，我们提出了一个新的三维卷积神经网络 (cnn) 架构称为 i2ii-3d 预测边界位置的体积数据。我们的精细到精细、深度监管的框架解决了 3d 边界检测的三个关键问题: (1) 高效、全面、端到端体积标签培训和预测 (2) 精确的体素级预测，以捕获精细尺度医学数据中流行的结构和 (3) 定向多尺度、多层次的特征学习。我们评估我们对由 93 个医学图像体积组成的数据集的方法进行评估，该数据集具有广泛的解剖区域和血管结构。在此过程中，我们还引入了 hed-3d，这是最先进的 2d 边缘探测器 (hed) 的 3d 扩展。我们展示了我们的深度学习方法的表现，目前最先进的 3d 血管边界检测 (结构化林 3d)，在很大的范围内，以及 hed 应用于切片，和 hed-3d，同时成功本地化精细结构。

通过我们的方法, 边界检测在典型的  $512 \times 512 \times 512$  卷上大约需要一分钟。少

2016 年 5 月 26 日提交;最初宣布 2016 年 5 月。

### 134. fpnn: 三维数据的现场探测神经网络

作者:李阳燕, [soeren pirk](#), [hao su](#), [charles r.qi](#) , [leonidas j. guibas](#)

**文摘:** 建立三维数据的判别表示一直是计算机图形学和计算机视觉研究的重要课题。卷积神经网络 (cnn) 已证明在 2d 图像上运行, 在各种任务中取得了巨大的成功。将卷积运算符提升到 3d (3dcnn) 似乎是一个合理且有希望的下一步。不幸的是, 3d cnn 的计算复杂度在体素分辨率方面呈小程度增长。此外, 由于大多数 3d 几何表示是基于边界的, 因此被占领区域不会随离散化的大小而增加, 从而导致计算的浪费。在这项工作中, 我们将 3d 空间表示为体积场, 并提出了一种新的设计, 该设计采用现场探测滤波器从这些空间中有效地提取特征。每个现场探测滤波器都是一组探测点——感知空间的传感器。我们的学习算法不仅优化了与探测点相关的权重, 还优化了探测点的位置, 从而变形了探测滤波器的形状, 并自适应地将其分布在 3d 空间中。优化的探测点 "智能" 地感知 3d 空间, 而不是盲目地在整个域上运行。我们表明, 在 3d 对象识别基准数据集的分类任务上, 现场探测的效率明显高于 3DCNNs, 同时提供最先进的性能。少



2016年10月24日提交;v1于2016年5月20日提交;最初宣布2016年5月。

### 135. 综合训练图像, 提高人体三维姿势估计

作者:陈文正,王欢,李阳燕,苏浩,王振华,长河图,丹妮·利辛斯基,丹尼尔·科恩-奥勒,陈宝泉

**摘要:** 从单个图像进行人类三维姿态估计是一项具有挑战性的任务,具有众多的应用价值。卷积神经网络(cnn)最近通过对人群采购收集的二维注释图像进行训练,从一张图像中实现了较好的二维姿态估计任务。这表明,直接估计三维姿势也可以取得类似的成功。但是,3d姿势很难注释,并且缺少合适的注释训练图像会阻碍端到端解决方案的尝试。为了解决这个问题,我们选择自动合成训练图像与地面真相姿态注释。我们的工作沿着这条道路进行系统的研究。我们发现,姿势空间覆盖和纹理多样性是合成训练数据有效性的关键成分。我们提出了一个全自动的,可扩展的方法,采样人类的姿势空间,以指导合成过程,并从真实的图像提取服装纹理。此外,我们还探索了域适应,以弥合合成训练图像和实际测试照片之间的差距。我们证明,cnn训练与我们的合成图像比那些培训的真实照片在3d姿势估计任务。少

于2017年1月5日提交;v1于2016年4月10日提交;最初宣布2016年4月。



### 136.marr 重新显示: 通过表面正常预测进行 2d-3d 对齐

作者:[aayush bansal](#), [bryan russell](#), [abhinav gupta](#)

**摘要:** 我们引入了一种方法, 利用表面正常预测以及外观提示, 从大型 cad 对象库中检索 2d 静止图像中所描绘对象的 3d 模型。我们的方法成功的关键是能够为所描绘场景中的对象恢复精确的曲面法线。我们介绍了一种基于预先训练的牛津 vgg 卷积神经网络 (cnn) 的滑块网络模型, 用于表面法线预测。与以往的方法相比, 我们的模型在 nyuv2 rgb-d 数据集上实现了最先进的精度, 可进行表面法线预测, 并恢复精细对象细节。此外, 我们还在输入图像上开发了一个双流网络, 并预测了表面法线, 共同学习了 cad 模型检索的姿态和风格。在使用预测曲面法线时, 我们的双流网络在姿势预测任务中使用从 rgb-d 图像计算的曲面法线匹配先前的工作, 并在使用 rgb-d 输入时达到了最先进的状态。最后, 我们的双流网络允许我们检索 cad 模型, 以便与基线方法相比, 更好地匹配所描绘对象的样式和姿势。少

2016 年 4 月 5 日提交;最初宣布 2016 年 4 月。

### 137. 后-cnn: 磁共振成像和超声对深脑区分割的深度学习

作者 :[fausto milletari](#), [seyed-ahahahahati](#), [christine kroll](#), [annika plate](#), [verena rozanski](#), [juliana maio 驻](#), [jones levin](#), [olaf dietrich](#), [birgitertl-wagner](#), [kai bötzel](#), [nassir navab](#)

**文摘:** 在本文中, 我们提出了一种利用卷积神经网络 (cnn) 的抽象功能来执行分割的新方法。我们的方法是基于 hough 投票, 这是一个策略, 允许全自动本地化和分割的解剖的利益。这种方法不仅使用 cnn 的分类结果, 而且还通过利用网络最深处产生的功能来实现投票。结果表明, 该方法具有鲁棒性强、多区域、灵活等鲁棒性, 易于适应不同的分割方式。为了展示 cnn 在医学图像分析中的能力和行为, 我们根据最先进的技术对六种不同网络架构的性能进行了系统的研究。标准, 在各种情况下。我们评估不同数量的训练数据和不同的数据维数 (2d、2.5 d 和 3d) 对最终结果的影响。我们在 mri 和经颅美国卷上显示的结果分别描述了 26 个区域的基底神经节和中脑。少

**2016 年 1 月 31 日提交;**v1 于 2016 年 1 月 26 日提交;**最初宣布** 2016 年 1 月。

### 138. 从真实视图到渲染视图的深度表达式 2d-3d 检测

**作者:**francisco m 萨, bryan russell, mathieu aubry

**文摘:** 本文提出了一种用于二维和三维样本检测的端到端卷积神经网络 (cnn)。我们证明, 能够调整自然图像的特征, 以便更好地与 cad 呈现视图的特征保持一致, 这对我们的技术的成功至关重要。我们证明了适应可以通过在自然图像上合成纹理对象模型的呈现视图来学习。我们的方法可以自然地融入美国有线电视新

闻网的检测管道, 并将深度学习的最新进展的准确性和速度优势扩展到 **2d-3d** 样本检测.我们将我们的方法应用于两个任务: 实例检测（我们在宜家数据集上进行评估）和对象类别检测（在 pascal voc 数据集的子集上的 "椅子" 检测方面优于 aubry 等人。少

**2016 年 4 月 18 日提交**;v1 于 2015 年 12 月 8 日提交;**最初宣布** 2015 年 12 月。

### 139. 基于因子时空卷积网络的人的行为识别

作者:[林孙](#),[奎佳](#),[杨迪燕](#),[伯特拉·石](#)

**摘要:** 视频序列中的人类行为是三维时空信号, 描述了相关人类和物体的视觉外观和运动动力学。在卷积神经网络 (**cnn**) 成功进行图像分类的启发下, 最近尝试学习 **3d cnn** 来识别视频中的人类行为。然而, 部分由于 3d 卷积内核训练的高度复杂性和对大量训练视频的需求, 只报告了有限的成功。这引发了我们在本文中研究的一种能够更有效地处理**三维**信号的新的深层体系结构。具体而言, 我们建议将时空卷积网络 (fstcn) 分解原始三维卷积内核学习, 作为在低层 (称为空间) 学习**二维**空间内核的顺序过程卷积层), 然后在上层学习一维时间内核 (称为时间卷积层)。我们引入了一种新的变换和置换算子, 使 fstcn 中的因子分解成为可能。此外, 为了解决序列对齐问题, 我们提出了一种基于从给定动

作视频序列中采样多个视频剪辑的有效训练和推理策略。我们在两个常用的基准数据集 (ucf-101 和 hmdb-51) 上测试了 fstcn。在不使用辅助训练视频来提高性能的情况下, fstcn 的性能优于现有的基于 cnn 的方法, 并与使用辅助训练视频的最新方法实现了可比的性能。少

2015 年 10 月 2 日提交;最初宣布 2015 年 10 月。

#### 140. 并行多维 lstm 在快速生物医学体积图像分割中的应用

作者 :[marijn f. stollenga](#), [wonmin byeon](#), [marcus liwicki](#) ,  
[juergen schmidhuber](#)

摘要: 卷积神经网络 (cnn) 可以在 2d 图像或 3d 视频之间移动, 以分割它们。它们具有固定的输入大小, 通常只感知要归类为前景或背景的像素的小局部上下文。相反, 多维重复域 (md-rnn) 可以在几个像素中通过几个扫描来感知每个像素的整个时空上下文, 尤其是当 mn 是一个长期的短期存储器 (lstm) 时。然而, 尽管有这些理论上的优势, 但与 cnn 不同的是, 以前的 md-lstm 变体很难在 gpu 上并行化。在这里, 我们重新排列了传统的立方体计算顺序在 md-lstm 在金字塔式的方式。由此产生的 pyramid-lstm 很容易并行化, 特别是对于 3d 数据, 如成堆的大脑切片图像。pyramid-lstm 在 mrbrains13 上获得了最著名的像素明智的大脑图像分割结果 (以及在 es-isbi12 上的竞争结果)。

少

2015 年 6 月 24 日提交;最初宣布 2015 年 6 月。

## 141. cnn 的渲染: 使用 cnn 训练的 3n 图像的视点估计

作者:[郝苏](#),[夏尔·齐](#),[李阳燕](#),[列奥尼达斯·吉巴斯](#)

**摘要:** 二维图像中的对象视点估计是计算机视觉中的一项重要任务。然而, 有两个问题阻碍了它的进展: 缺乏带有观点注释的培训数据, 以及缺乏强大的功能。在 3d 模型日益普及的启发下, 我们提出了一个框架, 通过将基于渲染的图像合成和 cnn 相结合来解决这两个问题。我们认为, 三维模型有可能产生大量的高变化图像, 这些图像可以被具有高学习能力的深度美国有线电视新闻网很好地利用。为了实现这一目标, 我们提出了一个可扩展的和抗过度的图像合成管道, 以及一个新的 cnn 专门为观点估计任务量身定制。实验表明, 在 pascal 3d+ 基准测试中, 管道的视点估计可以明显优于最先进的方法。少

2015 年 5 月 21 日提交;最初宣布 2015 年 5 月。

## 142. 稀疏的 3d 卷积神经网络

作者:[本·格雷厄姆](#)

**摘要:** ..... 网络设计用于处理稀疏三维输入数据。我们生活的世界是三维的, 所以有大量的潜在应用, 包括三维物体识别和时空物体的分析。在追求效率的过程中, 我们尝试了...

2015 年 8 月 25 日提交;v1 于 2015 年 5 月 12 日提交;最初宣布 2015 年 5 月。

### 143. 用于三维形状识别的多视景卷积神经网络

作者: [hang su](#), [subhransu maji](#), [evangelos kalogerakis](#), [erik lebed-miller](#)

**摘要:** 计算机视觉中的一个长期问题是 3d 形状的表达以进行识别: 3d 形状是否应使用在其本机 3d 格式 (如体素网格或多边形网格) 上操作的描述符来表示, 或它们是否可以用基于视图的描述符有效地表示? 我们在学习从 2d 图像上呈现的视图集合中识别 3d 形状的上下文中解决了这个问题。我们首先提出了一个标准的 cnn 架构, 训练它可以独立地识别形状的呈现视图, 并表明即使从单个视图中也可以以远远高于使用最先进的精度识别 3d 形状 3d 形状描述符。当提供形状的多个视图时, 识别率会进一步提高。此外, 我们还推出了一个新颖的 cnn 架构, 将来自 3d 形状多个视图的信息组合到一个紧凑的形状描述符中, 提供更好的识别性能。同样的架构可以用来准确地识别人类手绘的形状素描。我们的结论是, 2d 视图的集合可以提供非常丰富的 3d 形状识别信息, 并适合新兴的 cnn 架构及其衍生产品。少

2015 年 9 月 27 日提交;v1 于 2015 年 5 月 5 日提交;最初宣布 2015 年 5 月。

## 144. 基于草图的基于卷积神经网络的三维形状检索

作者:王芳,乐康,李毅

**摘要:** 从二维人体草图中检索 3d 模型在图形、图像检索和计算机视觉领域受到了相当大的关注。几乎总是处于最先进的状态, 对于 3d 模型, 可以计算出大量的 "最佳视图", 希望查询草图与使用预定义特征的 3d 模型的 2d 投影之一相匹配. 我们认为, 这两个阶段的方法 (视图选择-匹配) 是务实的, 但也有问题, 因为 "最佳视图" 是主观的和模糊的, 这使得匹配输入模糊。这种精确匹配的特性进一步使得手动选择功能变得很困难。我们建议使用简约的方法来定义我们的视图, 并为草图和视图学习特征, 而不是依赖于难以捉摸的 "最佳视图" 概念和手工制作的特征。具体来说, 我们将整个数据集的视图数量大幅减少到只有两个预定义的方向。然后, 我们学习了两个暹罗卷积神经网络 (cnn), 一个用于视图, 另一个用于草图。损失函数是在域内定义的, 也是在跨域相似性上定义的。我们在三个基准数据集上的实验表明, 我们的方法明显优于最先进的方法, 并且在所有常规指标中的性能都优于最先进的方法。少

2015 年 4 月 14 日提交;最初宣布 2015 年 4 月。

## 145. 深度 cnn 了解对象的内容是什么?

作者:彭兴超、孙宝辰、卡里姆·阿里、凯特·萨延科

**摘要:** 深层卷积神经网络学习极其强大的图像表示, 但大部分的能量隐藏在数以百万计的深层参数中。这些参数究竟代表什么? 最近的工作已开始分析[美国有线电视新闻网](#)的陈述, 发现, 例如, 它们对 fischer 等人 (2014 年) 的一些二维转换是不变的, 但被特定类型的图像噪声 nguyen 等人 (2014 年) 所迷惑。在这部作品中, 我们更深入地探讨: 对于由三维形状、姿势和照片性引起的对象类变化, cnn 是如何不变的? 少

2015 年 4 月 9 日提交;最初宣布 2015 年 4 月。

#### 146. 一种利用深卷积神经网络随机组检测淋巴结的 2.5 d 新表示

**作者:** holger r. roth, le lu , ari seff, kevin m.cherry, joanne hoffman, shijunwang, jiamin liu, evrim turkbey, ronald m.summers

**抽象:** 自动淋巴结 (ln) 检测是一项重要的临床诊断任务, 但由于计算机断层扫描 (ct) 中周围结构的对比度较低, 且其尺寸、姿势、形状和分布稀疏的位置各不相同, 因此具有很大的挑战性。最先进的研究表明, 在每卷 3.1 假阳性 (fp/vol.) 时, 灵敏度为 52.9%, 而在每个体积 (fp/vol.) 的情况下, 对纵隔 ln 的灵敏度为 60.9, 而在 3d haar 特征上, 则为 6.1 fp/vol., 灵敏度为 60.9。在本文中, 我们首先操作一个初步的候选生成阶段, 以高 fp 水平 (每个病人 40 个) 为代价实现 100% 的敏感性, 以收获兴趣量 (voi)。因此, 我们的 2.5 d 方法通过缩放 2d 重新格式化的正



交视图  $n$  次, 通过缩放、随机平移和相对于 voi 质心坐标的旋转, 分解任何 3d voi。然后利用这些随机视图训练深卷积神经网络 (cnn) 分类器。在测试中, cnn 被用来为所有  $n$  随机视图分配 ln 概率, 这些随机视图可以简单地平均 (作为一组) 来计算每个 voi 的最终分类概率。我们在两个数据集中验证了该方法: 90 ct 体积, 388 纵隔 ln 和 86 例 595 例腹腔 ln。我们在 3 fp/vol. 和 84%/90% 的灵敏度分别在 6 fp/vol. 在纵隔和腹部, 这大大改善了以前的最先进的工作。少

2014 年 6 月 6 日提交;最初宣布 2014 年 6 月。