

基于深度学习的 2d 3d 课题科研前沿进展（不含 Convolutional neural network 关键词）

2018.11.10 方建勇

1. 建议: 1811. 03208[[pdf](#),其他] Cs。简历

基于综合数据的树状结构深层语义实例分割

作者:[kerry halupka](#), [rail gamavi](#), [stephen moore](#)

抽象: 类似树的结构, 如血管, 往往表达复杂的非常精细的规模, 需要高分辨率的网格来充分描述它们的形状。这种稀疏形态可以交替地表示为中心线点的位置, 但从这种类型的深度学习数据中学习是具有挑战性的, 因为它是无序的, 并且排列不变。在这项工作中, 我们提出了一个深度神经网络, 直接消耗沿分支结构的中心线的无序点, 以确定在单次拍摄中表示的结构的拓扑结构。我们方法的关键是使用一种新的多任务损失函数, 从而实现任意复杂分支结构的实例分割。我们仅使用综合生成的数据对网络进行培训, 利用域随机化来促进向真实 2d 和 3d 数据的传输。结果表明, 我们的网络能够可靠地提取分支位置、分叉和端点的有意义的信息, 并为分支结构中的语义实例分割建立了新的基准。少

2018 年 11 月 7 日提交;最初宣布 2018 年 11 月。

评论:接受 2019 年 wacv

2. 第 1811. 01549[[pdf](#),其他] Cs。简历

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:尽管成功的深度..。更多

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

3. 第 1810.10342[[pdf](#)] Cs。简历

从眼底照片中预测光学相干性断层扫描学衍生的糖尿病黄斑水肿等级

作者:[avinash varadarajan](#), [pinal bavishi](#), [paisan raumviboonsuk](#), [peranut chotcomwongse](#), [subhashini venugopalan](#), [arunachalam narayanaswamy](#), [豪尔赫·夸德罗斯](#), [kuniyoshikanai](#), [george bresnick](#), [mongkol tadarati](#), [Mongkol silpa-archa](#), [jirawut limwattanayingyong](#), [variya nganthavee](#), [joe ledsam](#), [pearse a keane](#), [格雷格·s·科拉多](#),[莉莉·彭莉莉](#),[戴尔·r·韦伯斯特](#)

摘要: 糖尿病眼病是可预防失明的最快原因之一。随着血管内皮生长因子 (vegf) 疗法的出现, 检测中源性糖尿病黄斑水肿变得越来越重要。然而, 中心参与糖尿病黄斑水肿是通过光学相干断层扫描 (oct) 诊断的, 由于成本和工作流程的限制, 在筛查地点通常无法获得。相反, 筛选程序依赖于检测硬渗出物作为颜色眼底照片上 dme 的代名词, 通常会导致高假阳性或假阴性电话。为了提高 dme 筛选的准确性, 我们训练了一个深度学习模型, 使用色底照片来预测从 oct 考试获得的 dme 等级。我们的 "oct-dme" 模型的 auc 为 0.89 (95% ci:0.87-0.91), 在 80% 的特异性下, 灵敏度为 85%。相比之下, 三位视网膜专家具有相似的敏感性 (82-85), 但只有一半的特异性

(45-50, $p < 0.001$ 与模型的比较)。oct-dme 模型的阳性预测值 (ppv) 为 61% (95% ci:56-66%), 约为视网膜专家 36-38 的两倍。此外, 我们还使用了显著性和其他技术来检查模型是如何进行预测的。深度学习算法能够做出临床相关的预测, 通常需要从简单的 2d 图像中获得复杂的 3d 成像设备, 这与许多其他应用有着广泛的相关性。医学成像。少

2018 年 10 月 18 日提交;最初宣布 2018 年 10 月。

4. 第 181010337[[pdf](#),[其他](#)] Cs。简历

投影故障: 基于光的对深度学习分类器的对抗攻击

作者:[nicole nicole](#), [robert j 斯珀](#)

摘要:这项工作展示了一个物理攻击的深..。更多

2018 年 10 月 16 日提交;最初宣布 2018 年 10 月。

5. 第: 1810.0733[[pdf](#),[其他](#)] Cs。简历

[多伊](#) 10.1007/978-3-030-00889-5 _ 26

学习最优深度投影 18f-fdg pet 成像在帕金森病综合征早期鉴别诊断中的应用

作者:[shubham kumar](#) , [abhijit guharoy](#), [ping wu](#), [sailesh conjeti](#), [r. s.anand](#), [jian wang](#), [igor yakushev](#), [stefan förster](#), [markus schwaiger](#), [黄成成](#) , [阿克塞尔·罗金格](#), [左传涛](#),[石光宇](#)

摘要: 几种帕金森病综合征疾病在早期出现类似症状, 到目前为止还没有客观广泛使用的诊断方法得到批准。正电子发射断层扫描 (pet) 18f-fdg 被证明能够评估早期神经功能障碍的系统核病和战术病。基于张力分解 (tf) 的方法已被应用于识别特征代谢模式的鉴别诊断。然而, 这些传统的降维策略假定数据内部的线性或多线性关系, 因此不足以区分各种帕金森病综合征之间的非线性代谢差异。本文提出了一种深度投影神经网络 (dpnn), 用于识别典型的代谢模式, 用于帕金森病综合征的早期鉴别诊断。我们从现有的 tf 方法中汲取灵感。网络由 (i) 压缩部分组成: 它使用深度网络来学习 3d 扫描的最佳 2d 投影, 以及 (ii) 分类部分: 将 2d 投影映射到标签。可以使用剩余的未标记数据集对压缩部分进行预训练。此外, 由于分类部分在这些二维投影上运行, 因此可以使用有限的标记数据对端进行端到端训练, 而不是 3d 方法。我们表明, 与现有的最先进和合理的基线相比, dpnn 更有效。少

2018 年 10 月 11 日提交;最初宣布 2018 年 10 月。

评论:8 页, 3 个数字, 会议, miccai dlmia, 2018 年

类:l.2.10;l.2.4;l.4.10;l.2。1

日记本参考:kumar, shubham, et al. dlmia, springer, cham, 2018。227-235

6. 建议: 1810-05206[[pdf](#),[其他](#)] Cs。铭

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](#)

文摘: 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, 在现实世界中, 3d 对象被投影到 2d 通过拍摄不同的学习 (识别) 任务的域。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对受害者进行错误分类对象作为任意恶意目标，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。

2018 年 10 月 11 日提交;最初宣布 2018 年 10 月。

7. **新建: 1810.0088**[pdf,其他] Cs。简历

带有级联 unet 的胶质瘤分割

作者:[dmitry lachinov](#), [evgeny vasiliev](#), [vadim turlapov](#)

文摘 mri 分析在脑肿瘤诊断和治疗中处于中心位置，因此其精确评价至关重要。但是，它的 **3d** 特性带来了一些挑战，因此通常会对 **2d** 投影进行分析，从而降低复杂性，但增加偏差。另一方面，耗时的**三维**评估，如分割，能够提供一些有价值的空间特征的精确估计，让我们了解疾病的过程最近的研究，重点是与传统的计算机视觉算法相比，**深度学习**方法具有卓越的分割任务。但尽管如此，这仍然是一个具有挑战性的问题。本文提出了脑肿瘤自动分割的深级联方法。与最近的对象检测方法类似，我们的实现基于神经网络;我们提出了对 **3d** unet 体系结构的修改和增强策略，以有效地处理多模态 mri 输入，此外，我们还引入了利用从相同拓扑模型中获得的上下文来提高分割质量的方法在缩小规模的数据上运行。我们评估了 brats 2018 数据集的介绍方法，并讨论了结果。

2018 年 10 月 9 日提交;最初宣布 2018 年 10 月。

8. **第 xiv: 180 0909061**[pdf,其他] Cs。简历

稀疏到连续: 使用占用率图增强单目深度估计

作者:[nicolas rosa](#), [vitor guizilini](#), [valdir grassi jr](#)

文摘: 本文讨论了单图像深度估计 (side) 问题，重点研究了**提高深度神经网络**预测的准确性。在监督学习场景中，预测的质量与指导优化过程的培训标签有着内在的联系。对于室内场景，基于结构的基于光线的深度传感器 (例如 kinect) 能够提供密集的深度图，尽管是短程的深度图。另一方面，对于室外场景，lidar 仍然被认为是标准传感器，它相对提供了更少的测量，特别是在更远的地区。本文介绍了一种基于希尔伯特地图框架的深度图致密化新技术，而不是修改神经网络结构来处理稀疏深度图。根据 lidar 扫描的 **3d** 点生成连续占用率图，并将生成的重建曲面投影到具有任意分辨率的 **2d** 深度图中。在 kititi 数据集的各种子集上进行的实验表明，在培训方法中没有引入额外信息的情况下，拟议的 "从量到连续" 技术所产生的改进。

2018 年 9 月 24 日提交;最初宣布 2018 年 9 月。

评论:8 页, 4 个数字, 期刊

9. **特别报告: 189.07921**[pdf,其他] Cs。简历

基于多模态深度监控的三维人的姿态估计

作者:[周坤](#), [蔡金庙](#), [姚丽](#), [石玉龙](#), [韩晓光](#), [江念娟](#), [奎佳](#), [江波路](#)

文摘: 本文提出了一种新的**基于深度学习的框架**, 从单一图像中推断**三维人体姿势**。具体而言, 制定了两阶段办法。首先利用具有两个分支的生成器分别提取显式和隐式深度信息。在培训过程中, 还采用了对抗计划, 以进一步提高业绩。在第二步中, 用一个广泛使用的**估计器**生成的估计二维关节的隐式和显式深度信息被一起输入到一个深的**三维姿态回归器**中, 用于最终的姿势生成。我们的方法在 **eccv2018 3d 人体剂量估计挑战**中达到了 58.68 mm 的 mpjpe。少

2018 年 9 月 20 日提交;最初宣布 2018 年 9 月。

10. 第 [xiv:1809.07016](#)[pdf,其他] Cs. 铬

生成 3d 对抗点云

作者:[冲翔](#),[齐晓明](#),[李波](#)

摘要: 机器学习模型, 特别是**深度神经网络 (dnn)** 已成功地应用于各种应用。然而, dnn 是众所周知的容易受到敌对例子, 这些实例是精心打造的实例, 旨在使学习模型做出不正确的预测。近年来, 对**二维图像**、自然语言和音频数据集的对抗示例进行了广泛的研究, 而 **3d** 模型的鲁棒性尚未得到探索。鉴于 **3d** 模型在自动驾驶中的广泛的安全关键应用, 例如激光雷达数据的 pointnet, 了解 **3d** 模型在各种对抗攻击下的脆弱性非常重要。由于点云数据的特殊格式, 在点云空间中生成对抗示例具有挑战性。在这项工作中, 我们提出了新的算法来生成对 pointnet 的敌对点云, pointnet 是处理点云数据的最广泛使用的模型。我们主要在点云上提出两种类型的攻击: 不明显的对抗点云和物理攻击的可制造的对抗点群。对于不明显的点云, 我们建议要么转移现有的点, 要么添加新的点, 以产生 "不明显" 的扰动。对于对抗点集群, 我们建议生成少量明确的 "可制造的对抗点集群", 这些集群是显而易见的, 但却是有意义的集群。这些对抗点聚类的目标是通过 **3d** 打印合成对象并将其粘附到原始对象上来实现 "物理攻击"。此外, 我们还针对不同的攻击提出了 7 个摄动测量指标, 并进行了广泛的实验, 以评估在 modelnet40 数据集上提出的算法。总体而言, 我们的攻击算法实现了所有目标攻击的 100% 攻击成功率。少

2018 年 9 月 19 日提交;最初宣布 2018 年 9 月。

11. 第: [1809.06121](#)[pdf,其他] Cs. Lg

基于 naf 强化学习的生物力学模拟中的肌肉激发估计

作者:[amir h. abdi](#), [pramit saha](#), [praneeth srungarapu](#), [sidney fels](#)

摘要: 运动控制是一组时变肌肉兴奋, 产生所需的运动的生物力学系统。肌肉兴奋不能直接从活体实验对象中测量。另一种方法是使用逆运动驱动模拟来估计肌肉的激活。在本文中, 我们提出了一种**深层强化学习方法**来估计模拟生物力学系统中的肌肉兴奋。在这里, 我们引入了一个定制的奖励函数, 它激励更快地跟踪目标运动的点对点。此外, 我们还部署了两种新技术, 即基于插曲的硬更新和双缓冲区体验重播, 以避免反馈训练循环。该方法在 6 个至 24 段轴向肌肉的四个模拟**二维和三维**环境中进行了测试。结果表明, 模型能够在近 10 万次模拟动作后, 学会给定运动的肌肉兴奋。此外, 通过实验达到目标的点对点的根均方误差还不到运动域长度的 1%。我们的强化学习方法与传统的动态方法相去甚远, 因为肌肉控制是由一组分布的神经元功能派生的。这可以为神经活动解释这一现象开辟路径。少

2018 年 9 月 17 日提交;最初宣布 2018 年 9 月。

评论:miocai 2018 会议, 9 页, 3 个数字

12. 第 [xiv:1809.05955](#)[pdf,其他] Cs. 简历

从单个二维荧光图像为机器人辅助血管内主动脉修复的三维路径规划

作者:郑建青,周晓云,西莉亚里加, 杨光忠

摘要: 目前的术中导航标准要求插入的设备和主动脉分支之间需要三维对齐。导航通常通过二维透视图象, 缺乏解剖信息, 导致手术时间延长和辐射照射。本文介绍了一种基于腹部主动脉瘤 (aaa) 单二维透视图象的实时三维机器人路径规划框架。提出了一种图形匹配方法来建立三维术前与二维术中 aaa 骨架之间的对应关系, 然后通过骨架变形和正则化来记录这两个骨架。骨架长度和平滑度。此外,深度学习还用于从计算机断层扫描 (ct) 扫描中分割3d术前 aaa, 以促进框架自动化。模拟、幻象和患者 aaa 数据集已被用来验证所建议的框架。在幻象设置中实现了2mm的三维距离误差。在准确性、鲁棒性和时间效率方面也取得了性能优势。所有代码都将是开源的。少

2018年9月16日提交;最初宣布2018年9月。

13. 第 xiv:1809.05068[pdf,其他] Cs. 简历

用于单视图3d完成和重建的学习形状优先级

作者:吴家军,张成凯,张秀明, 张志东,威廉·t·弗里曼,约书亚·泰南鲍姆

摘要: 单视图 3d 形状完成或重建的问题具有挑战性, 因为在解释观察的许多可能形状中, 大多数是不可信的, 并且不符合自然物体。最近在这一领域的研究利用了深层卷积网络的表现力来解决这一问题。事实上, 还有另一个经常被忽略的模糊层次: 在似是而非的形状中, 仍然有多个形状同样适合2d图像;也就是说, 在给定单视图输入的情况下, 地面真相大白形状是不确定的。现有的完全监督的方法无法解决此问题, 通常会产生具有光滑表面但没有精细细节的模糊均值形状。在本文中, 我们提出 shapehd, 通过将深层生成模型与对抗学习的形状优先相结合, 推动单视图形状完成和重建的极限。博学的前科起到了调节的作用, 只有当模型的输出不现实时, 而不是在它偏离地面真理的情况下, 才会惩罚模型。因此, 我们的设计克服了上述两种程度的模糊。实验表明, shapehd 在多个真实数据集中的形状完成和形状重建方面都优于最先进的领域。少

2018年9月13日提交;最初宣布2018年9月。

评论:eccv 2018。前两位作者对这项工作作出了同样的贡献。项目页:

<http://shapehd.csail.mit.edu/>

14. 第 xiv:1809.04096[pdf,其他] Cs. 简历

并行可分离的三维卷积, 用于视频和体积数据的理解

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

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

2018年9月11日提交;最初宣布2018年9月。

15. 第 [xiv:1809.01019](#)[pdf,其他] Cs. 简历

利用深层可视化描述符实现分层高效本地化

作者:[pa牌-edouard sarlin](#), [frédéric debraine](#), [marcin Dymczyk](#), [roand siegwart](#), [cesar cadena](#)

摘要: 尽管在大型和不断变化的环境中运行, 但许多机器人应用仍需要精确的姿态估计。这可以通过视觉定位来解决, 使用预先计算的**环境 3d** 模型。然后, 姿势估计相当于使用本地描述符查找查询图像中的**2d** 关键点和模型中的**3d** 点之间的对应。但是, 在机器人平台上, 计算能力通常是有限的, 这使得这项任务在大规模环境中具有挑战性。二进制特征描述符显著加快了这种**2d-3d** 匹配的速度, 并在机器人社区中变得很受欢迎, 但也严重损害了感知锯齿的鲁棒性以及视点、照明和场景的变化结构。在这项工作中, 我们建议利用**深度学习**的最新进展来执行有效的分层本地化。我们首先使用学习过的图像范围的全局描述符在地图级别进行本地化, 然后从仅在候选位置计算的**2d-3d** 匹配中估计精确的姿势。这限制了本地搜索, 从而可以有效地利用通常在资源受限的设备上被忽略的强大的非二进制描述符。我们的方法可实现最先进的本地化性能, 同时在流行的移动平台上实时运行, 为机器人研究带来新的前景。少

2018 年 9 月 18 日提交;v1 于 2018 年 9 月 4 日提交;最初宣布 2018 年 9 月。

评论:[cdri 2018](#) 相机就绪 (修复拼写错误和更新引文)

16. 第 [09iv:1800.005553](#)[pdf,其他] Cs. 简历

基于多视点通信的单目图像目标位置估计

作者:[jogendra nath kundu](#), [rahul m. v.](#), [aditya ganeshan](#), [r. venkatesh babu](#)

摘要: 了解**二维**图像中对象的几何形状和姿态是广泛的现实世界应用的基本必要条件。在深度神经网络的推动下, 最近的方法给物体姿态估计带来了显著的改进。然而, 由于关键点/附加注释的真实图像稀缺, 它们受到影响, 因此无法有效地利用物体的**三维**结构信息。在本文中, 我们提出了一种利用内部物体的几何规律性进行姿态估计的数据高效方法。首先, 我们从简单的**二维 rgb** 图像中学习对象部分的一个不变的局部描述符。然后, 这些描述符以及从固定**3d** 模板模型的呈现中获得的关键点被用来为给定的单目真实图像生成关键点对应映射。最后, 一个姿态估计网络利用这些对应映射预测物体的**三维**姿态。此管道进一步扩展到多视图方法, 该方法吸收从**3d** 模板模型的多个视图生成的对应集中的关键点信息。多视图信息的融合显著提高了系统的几何理解, 进而提高了姿态估计性能。此外, 利用负责**学习**姿势不变关键点描述符的对应框架, 也可以有效地缓解数据稀缺问题。这使我们的方法能够在多个真实图像视点估计数据集 (如 [pascal3d+](#) 和 [objectnet3d](#)) 上实现最先进的性能。为了鼓励可重复的研究, 我们发布了我们建议的方法的代码。少

2018 年 9 月 3 日提交;最初宣布 2018 年 9 月。

评论:接受 [eccv-w](#);代码可在此 [http url: https://github.com/val-iisc/pose_estimation](http://github.com/val-iisc/pose_estimation)

17. 第 [1808.09351](#)[pdf,其他] Cs. 简历

通过反向图形进行三维感知场景操作

作者:[姚顺宇](#), [徐子明](#), [朱俊燕](#), [吴嘉军](#), [安东尼奥·托拉尔巴](#), [威廉·弗里曼](#), [约书亚 b. 泰南鲍姆](#)

摘要: 我们的目标是获得一个可解释的, 表达和解开的场景表示, 其中包含为每个对象的全面的结构和纹理信息。以前通过神经网络学习到的场景表示通常是不可解释的, 仅

限于单个对象, 或缺乏 3d 知识。在本工作中, 我们提出了 3d 场景去渲染网络 (3d-sdn), 通过将语义、几何和外观的解角表示集成到一个深层生成模型中来解决上述问题。我们的场景编码器执行反向图形, 将场景转换为结构化的对象式表示形式。我们的解码器有两个组件: 可微形状呈现器和神经纹理生成器。语义、几何和外观的分离支持 3d 感知场景操作, 例如, 在保持一致形状和纹理的同时自由旋转和移动对象, 并在不影响其形状和纹理的情况下更改对象外观形状。实验表明, 基于 3d-sdn 的编辑方案优于二维编辑方案。少

2018 年 10 月 11 日提交;v1 于 2018 年 8 月 28 日提交;最初宣布 2018 年 8 月。

评论:NIPS

18. 第 1808. 08578[pdf,其他] Cs。简历

基于形状约束的多任务深度学习方法的肝脏图像三维双室分割

作者:金明段, ghalib bello, jo schlemper, wenjia bai, timothy j w d 排

es, carlo biffi, antonio de marvao, georgia doumou, declan p ' regan, 丹尼尔·鲁克特

摘要: 深度学习在肝脏磁共振 (cmr) 图像分割方面取得了最先进的性能。然而, 大多数方法都专注于学习用于分割的图像强度特征, 而解剖形状的结合却较少受到关注。本文将多任务深度学习方法与地图集传播相结合, 开发了一种用于短轴 cmr 体积图像的基于形状约束的双室分割管道。该管道首先采用完全卷积网络 (fcn), 同时学习分割和地标本地化任务。拟议 fcn 的体系结构采用 2.5 d 表示, 从而结合了二维 fcn 网络的计算优势和在不影响分割精度的情况下解决三维空间一致性问题的能力。此外, 细化步骤旨在明确地强制实施形状约束并提高分割质量。此步骤对于克服图像人工制品 (例如, 由于不同的呼吸保持位置和较大的切片厚度) 是有效的, 这将阻止创建具有解剖意义的 3d 肝脏形状。由于网络能够推断地标, 拟议的管道是完全自动化的, 然后在管道下游使用这些地标来初始化地图集传播。我们验证了 1831 个健康受试者和 649 肺动脉高压患者的肝管。对这两个数据集进行的大量数值实验表明, 尽管输入 cmr 体积中的人工制品, 但我们提出的方法是鲁棒性强的, 能够生成准确、高分辨率和解剖光滑的双室三维模型。少

2018 年 8 月 28 日提交;v1 于 2018 年 8 月 26 日提交;最初宣布 2018 年 8 月。

19. 第 xiv:1808. 05942[pdf,其他] Cs。简历

神经身体拟合: 深度学习与基于模型的人体姿势和形状估计的统一

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 (神经身体拟合 (nbf))。它在 cnn 中集成了一个统计身体模型, 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 的方式提供

2018 年 8 月 17 日提交;最初宣布 2018 年 8 月。

评论:3dv 2018

20. 第 1808. 05205[[pdf](#),其他] Cs. 简历

多伊 [10.1016/j.media.2018.07.010](#)

利用分割网络构建数据非常有限的医学图像分类器

作者: [ken c. l. wong](#), [tanveer Syeda-Mahmood](#), [mehdi moradi](#)

摘要: 深度学习在医学图像分析方面显示出了很好的结果, 但由于缺乏非常大的注释数据集, 限制了其全部潜力。尽管使用 imagenet 预先训练的分类模型进行迁移学习可以缓解这一问题, 但受限的图像大小和模型复杂性可能会导致不必要的计算成本增加和性能下降。由于许多常见的形态特征通常由器官的不同分类任务共享, 如果我们能提取出这样的特征, 用有限的样本来改进分类, 那就非常有益了。因此, 在课程学习理念的启发下, 提出了一种利用分割网络特征构建医学图像分类器的策略。通过使用对类似数据进行预训练的分割网络作为分类任务, 机器可以先学习更简单的形状和结构概念, 然后再处理通常涉及更复杂的实际分类问题概念。利用我们提出的三维三类脑肿瘤类型分类问题框架, 我们在 191 个测试样本和 91 个训练样本中获得了 82% 的准确率。在应用于 2d 九级心脏语义级别分类问题时, 我们在 263 个测试样本和 108 个训练样本中达到了 86% 的准确率。介绍了与 imagenet 预训练分类器和从头开始训练的分类器的比较。少

2018 年 8 月 15 日提交;最初宣布 2018 年 8 月。

评论:本文已被医学图像分析所接受

日记本参考:医学图像分析 49 (2018) 105-116

21. 第 xiv: 1808. 02084[[pdf](#),其他] Cs. 简历

基于混合表示的场景合成的深层生成建模

作者: [张泽伟](#), [杨振培](#), [马崇阳](#), [罗林杰](#), [亚历山大·胡特](#), [艾蒂安·武加](#), [黄启兴](#)

文摘: 提出了一种适用于室内环境的深层生成场景建模技术。我们的目标是使用前馈神经网络训练生成模型, 该神经网络将先前的分布 (例如正态分布) 映射到室内场景中主要对象的分布。我们引入了一种 3d 对象排列表示形式, 根据对象的大小和形状属性对对象的位置和方向进行建模。此外, 我们的场景表示适用于从数据库中选择的具有不同倍数 (重复计数) 的 3d 对象。我们展示了一种有原则的方法来训练这种模型, 方法是将三维对象排列表示和基于 2d 图像的二维图像表示的鉴别损耗结合起来。我们展示了场景表示的有效性和基准数据集上的深度学习方法。我们还展示了该生成模型在场景插值和场景完成中的应用。少

2018 年 8 月 6 日提交;最初宣布 2018 年 8 月。

22. 第 1808. 01134[[pdf](#),其他] Cs. 简历

ispa-net: 迭代语义对齐网络

作者: [jogendra nath kundu](#), [aditya ganeshan](#), [rahul m.v.](#), [aditya prakash](#), [r. venkatesh babu](#)

文摘: 从单目二维图像中理解和提取物体的三维信息是计算机视觉中的一个基本问题。在三维对象姿态估计任务中, 最近基于数据驱动的深度神经网络方法缺乏具有三维关键点和姿态注释的真实图像。从人类认知中汲取灵感, 注释者以三维 cad 模型为结构参考, 获取真实图像的基本真理观点;我们提出了一个迭代语义 pose 对齐网络, 称为 imsa-net。我们的方法侧重于利用语义三维结构的规律性, 通过预测给定图像之间的视点差异来解决细粒度姿态估计的任务。这种基于图像比较的方法还缓解了数据稀缺性

问题,从而提高了所提出的新对象类别的可扩展性,只需最少的注释。细粒度物体姿态估计也得益于输入图像对的学习空间描述符的对应关系。所提出的姿势对齐框架可以利用在线渲染设置以及后差的非统一纸盒分类的有效性来改进其在连续迭代中的初始姿态估计。这使得 ispa-net 能够在各种真实图像视点估计数据集上实现最先进的性能。此外,我们还展示了该方法在多个应用中的有效性。首先,我们展示了活动对象视点定位的结果,以从只考虑单个图像作为姿态参考的类似姿态中捕获图像。其次,我们证明了学习的语义对应只使用每个对象类的一个部分注释的 3d 模板模型来执行无监督的部分分割传输的能力。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 8 月 3 日提交;最初宣布 2018 年 8 月。

评论:2018 年在 acmmm 会议上接受。代码可在 <https://github.com/val-iisc/iSPA-Net>

23. 第 1807. 11929 条[[pdf](#),其他] Cs. 简历

以自我为中心的空间记忆

作者:[张敏](#),[马景德](#), 严世成, [林惠伟](#),[赵琪](#), [冯家志](#)

摘要: 以自我为中心的空间记忆 (esm) 定义了一个存储系统, 该系统从自我中心的角度对有关环境的空间信息进行编码、存储、识别和回忆。我们介绍了一种用于 esm 建模的集成深度神经网络体系结构。它学习估计世界的占用状态, 并逐步构建自上而下的 2d 全球地图, 从自我中心的观点在空间扩展的环境中。在探索过程中, 我们提出的 esm 模型使用重复神经网络更新了基于局部观测的全局地图的信念。它还通过新的外部存储器增强了本地映射, 以便在大型环境中进行长期探索, 对被访问位置的潜在表示形式进行编码和存储, 从而使代理能够执行位置识别, 从而实现循环关闭。我们提出的 esm 网络在以下几个方面做出了贡献: (1) 在没有特征工程的情况下, 我们的模型以端到端的方式有效地预测基于自我中心视图的自由空间;(2) 不同于其他基于深度学习的映射系统, esmn 处理连续动作和状态, 这对于实际应用中的机器人控制至关重要。在实验中, 我们通过与几个竞争基线的比较, 证明了它在三维虚拟迷宫和逼真室内环境中具有准确、稳健的全局映射能力。少

2018 年 7 月 31 日提交;最初宣布 2018 年 7 月。

评论:2018 年接受 8 页, 6 位数字

24. 第 xiv: 180. 7.09064[[pdf](#),其他] Cs. 简历

漫画商店: 个性化和逼真的漫画素描

作者:[韩晓光](#),[侯康成](#),[董渡](#), 邱玉达,[于一洲](#),[周坤](#),[崔树光](#)

摘要: 本文提出了一种交互式个性化、逼真的面部漫画素描系统。输入人脸的图像, 用户可以通过操纵其面部特征曲线来创建漫画照片。我们的系统首先根据编辑的草图对恢复的 3d 人脸模型进行夸张, 通过为每个顶点的拉普拉斯分配一个缩放因子来实现。为了构造二维草图与顶点缩放场之间的映射, 提出了一种新的深度学习体系结构。利用所获得的三维合成模型, 生成了两个图像, 一个是在底层三维网格变形的指导下应用二维翘曲获得的, 另一个是通过重新渲染变形的三维纹理获得的。模型。然后将这两个图像无缝集成, 以生成我们的最终输出。由于网格的严重拉伸, 渲染的纹理是模糊的外观。采用深度学习方法来推断缺失的细节, 以增强这些模糊的区域。此外, 还发明了一种照明操作, 以进一步提高结果的光现实主义。定量和定性实验结果验证了我们的素描系统的有效性和我们提出的技术相对于现有方法的优越性。少

2018 年 7 月 24 日提交;最初宣布 2018 年 7 月。

评论:向 [ieee tvcg](#) 提交 12 页, 16 位数字

25. 建议: 1807. 06742[[pdf](#),[其他](#)] Cs。简历

三维全球卷积对抗网络 \ 用于前列腺 mr 体积分割

作者:[贾浩哲](#),[宋洋](#),[张东豪](#),[黄恒](#), 冯大干, 迈克尔·富勒姆, 夏勇, [蔡伟东](#)

摘要: 已经开发出先进的深度学习方法, 以 **2d** 或 **3d** 完全卷积的方式进行前列腺磁共振体积分割。然而, 由于前列腺体积的大量空间信息在切片分割过程中被丢弃, 二维方法的分割性能往往有限; **3d** 方法也有改进的余地, 因为它们使用各向同性内核来执行 **3d** 卷积, 而大多数前列腺 mr 卷具有各向异性的空间分辨率。此外, 完全卷积的结构方法在定位问题上取得了较好的性能, 但忽略了分割任务的每体素分类。本文提出了一种**三维全球卷积对抗网络 (3d gca-net)** 来解决高效前列腺 mr 体积分割问题。我们首先设计了一个 **3d resnet** 编码器, 从前列腺扫描中提取**三维特征**, 然后开发了由多尺度 **3d** 全局卷积块和**三维边界细化块**组成的解码器, 以同时解决体积分割的分类和定位问题。此外, 我们将编解码器分割网络与训练短语中的敌对网络结合起来, 以加强远程空间预测的连续性。在整个模型中, 我们使用各向异性卷积处理来更好地学习前列腺磁共振扫描的特征。我们在两个公共前列腺 mr 数据集上评估了我们的 **3d gca-net** 模型, 并获得了最先进的性能。少

2018 年 7 月 17 日提交;最初宣布 2018 年 7 月。

评论:9 页, 3 个数字, 2 个表, 16 个参考

26. 第 [xiv:18006.11137](#)[[pdf](#),[其他](#)] Cs。简历

基于深度学习的基于实例的三维生物医学图像的弱注释分割

作者:[赵卓](#),[林阳](#),[郝正](#), 伊恩·h·古尔德纳,[张思远](#),[陈先生](#)

摘要: **三维图像**的实例分割是生物医学图像分析的一项基本任务。而。。。更多

2018 年 6 月 28 日提交;最初宣布 2018 年 6 月。

评论:2018 年接受 miccai

27. 第 [xiv:1806.10556](#)[[pdf](#),[其他](#)] Cs。简历

每个像素计数: 具有整体 3d 运动理解的无监督几何学习

作者:[杨振恒](#),[王鹏](#),[王洋](#), [徐伟](#),[拉姆·内马提亚](#)

摘要: 通过深卷积网络观看未标记的视频来学习估计单个图像中的 **3d** 几何, 是近年来的重要过程。目前最先进的 (sota) 方法是基于刚性结构自运动的学习框架, 在这种学习框架中, 只有**三维相机自我运动**被建模为几何估计。然而, 移动的物体也存在于许多视频中, 例如在街道场景中移动汽车。本文通过将每个像素的**三维物体运动**另外集成到学习框架中来解决这种运动问题, 该框架提供了全面的**三维场景流**理解, 并有助于单图像几何估计。具体而言, 给定视频中的两个连续帧, 我们采用一个运动网络来预测它们的相对 **3d** 相机姿势和分割掩码识别运动物体和刚性背景。光流网络用于估计密集的 **2d** 每像素对应。单个图像深度网络预测这两种图像的深度贴图。四种类型的信息, 即 **2d** 流、相机姿态、分段掩码和深度贴图, 集成到一个可区分的整体**三维运动解析器 (hmp)** 中, 其中用于刚性背景和移动的**每个像素 3d** 运动对象被恢复。我们设计了用于训练深度和运动网络的**两种类型的三维运动**, 从而为估计的几何形状进一步减少误差。最后, 为了解决单目视频的**三维运动混乱**, 我们将立体图像结合到关节训练中。在 kitti 2015 数据集上进行的实验表明, 我们估计的几何形状、**3d** 运动和移动物体掩码不仅受到一致的限制, 而且明显优于其他 sota 算法, 这证明了我们的方法的优势。少

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

评论:eccv18 的提交

28. 第 1806.08741[[pdf](#),其他] Cs。简历

基于通用和并行方法的可扩展简单线性迭代聚类 (sslic)

作者:[bradley c. lowekamp](#), [david t. chen](#), [ziv yaniv](#), [terry s. yoo](#)

摘要: 超级像素算法已被证明是图像分割和后续处理的一个有用的初始步骤, 通过将昂贵的每像素基元的使用替换为更高级别的抽象、超级像素来降低计算复杂性。它们已成功地应用于传统的图像分析和深度学习。在本文中, 我们提出了一个简单的线性迭代聚类 (slc) 超像素算法的广义实现, 该算法已推广到 n 维标量和多通道图像。此外, 标准迭代实现被并行的多线程实现所取代。我们描述了实现的详细信息, 并使用强尺度公式分析其可伸缩性。使用 **3d** 图像、可见人体冷冻数据集和来自同一数据集的 **2d** 图像进行定量评估。结果显示, 即使使用大量超过可用内核物理数量 (超线程) 的线程, 运行时也会获得良好的可伸缩性。少

2018 年 7 月 23 日提交;v1 于 2018 年 6 月 22 日提交;最初宣布 2018 年 6 月。

评论:提交给《国际观察杂志》(itk) 的手稿

29. 第 1806.08485[[pdf](#),其他] Cs。Gr

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。设计了一种新的基于 $cnmn$ 的回归网络, 该回归网络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。

评论:11 页

30. 第 [xiv:1806.08482](#)[[pdf](#),其他] Cs。简历

基于联合学习时间结构与空间细节的视频绘制

作者:[王川](#),[黄海斌](#),[韩晓光](#),[王觉](#)

文摘: 提出了一种新的数据驱动视频绘制方法来恢复视频帧缺失区域。提出了一种新的深度学习体系结构, 它包含两个子网络: 时间结构推理网络和空间细节恢复网络。时间结构推理网络是建立在 **三维** 完全卷积结构的基础上的: 考虑到 **三维** 卷积的昂贵计算成本, 它才学会完成低分辨率的视频量。低分辨率结果为空间细节恢复网络提供了时间指导, 该网络采用 **2d** 完全卷积网络进行基于图像的绘制, 以原始分辨率生成恢复的视频帧。这种两步网络设计既确保了每个帧的空间质量, 又确保了帧之间的时间一致性。我

们的方法以端到端的方式联合培训这两个子网络。我们对三个数据集进行定性和定量评估, 证明我们的方法优于以前基于学习的视频绘制方法。少

2018 年 6 月 21 日提交;最初宣布 2018 年 6 月。

评论:13 页, 5 个数字

31. 第 [xiv:1806.08472](#)[pdf,其他] Cs. 简历

一种高分辨率人脸边缘化的高保真度假设不变模型的研究

作者:[曹杰](#),[胡一波](#),[张洪文](#), [何然](#),[孙振南](#)

摘要: 面正面化是指从给定的轮廓合成人脸的正面视图的过程。由于野外的自遮挡和外观失真, 恢复忠实的结果并将纹理细节保存在高分辨率中是极具挑战性的。本文提出了一种高保真度假设不变模型 (hf-pim) 来产生摄影和身份保存的结果。hf-pim 通过一个新的纹理扭曲过程将轮廓正面化, 并利用密集的对应该场绑定 2d 和 3d 表面空间。我们将翘曲的前提分解为密集的对应该场估计和人脸纹理图的恢复, 这两个问题都得到了深度网络的很好解决。与这些依赖三维数据的重建方法不同, 我们还提出了抗性残差字典学习(ardl) 来监督仅使用单目图像的面部纹理图的恢复。在受控和不受控制的环境下进行的详尽实验表明, 该方法不仅提高了后不变人脸识别的性能, 而且显著提高了高分辨率的正面识别外观。少

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

评论:将于 2018 年 NIPS 出现

32. 第 [xiv:1806.08294](#)[pdf,其他] Cs. 简历

从全景图像的布局与几何和深度学习

作者:[clara Fernandez-Labrador](#) , [Perez-Yus perez-yus](#), [gonzalo lopez-nolas](#), [jose j. guerrero](#)

文摘: 本文提出了一种新的从 360 度全景图像中恢复室内场景的新方法。有了这样的图像, 就能同时看到所有场景, 从而能够恢复封闭的几何形状。我们的方法战略性地将几何推理 (线条和消失点) 提供的精度与深度学习技术 (边缘和普通地图) 实现的更高级别的数据抽象和模式识别结合起来。因此, 我们提取结构角, 从中我们产生布局假设的房间假设假设曼哈顿世界。选择最佳布局模式, 在简单的房间 (箱型) 和复杂形状的房间 (四面墙以上) 上都能实现良好的性能。该方法的实验是在两个公共数据集中进行的, 即 sun360 和斯坦福 (2d-3d-s), 展示了通过将几何和深度学习以及我们的建议的有效性方面的最新技术。少

2018 年 6 月 21 日提交;最初宣布 2018 年 6 月。

评论:8 页, 12 个数字, 2 个表

33. 第 [xiv:1806.06984](#)[pdf,其他] Cs. 简历

重复估计

作者:[tom f. h. runia](#) , [cees g. m. snoek](#), [arnold w. m. 斯梅尔斯](#)

摘要: 视觉重复在我们的世界里无处不在。它出现在人类活动 (运动、烹饪)、动物行为 (蜜蜂的摇摆舞)、自然现象 (风中的叶子) 和城市环境 (闪烁的灯光) 中。从逼真的视频估计视觉重复是具有挑战性的, 因为周期运动很少是完全静态和静止的。为了更好地处理现实的视频, 我们提升了现有工作经常做出的静态和静态假设。我们的时空滤波方法建立在周期运动理论的基础上, 有效地处理了各种各样的外观, 不需要学习。从三维运动开始, 我们通过将运动场分解为其基本组成部分, 推导出三种周期运动类型。此外,

从场的时间动力学中产生了三个时间运动连续性。对于三维运动的二维感知, 我们考虑相对于运动的视点; 以下是 18 例反复运动感知。为了估计在任何情况下的重复, 我们的理论意味着构造微分运动图的混合: 梯度, 发散和卷曲。我们利用小波滤波器对运动图进行时间上的卷积, 以估计重复的动态。我们的方法能够直接从在运动图上密集计算的时间滤波器响应进行空间分割重复运动。为了对我们的权利进行实验验证, 我们使用我们的新数据集进行重复估计, 用非静态和非平稳的重复运动更好地反映现实。在重复计数的任务上, 与深度学习的替代方案相比, 我们获得了良好的效果。少

2018 年 6 月 18 日提交; 最初宣布 2018 年 6 月。

34. 第 xiv:1806.05767[[pdf](#),[其他](#)] 反渗透委员会

运动规划网络

作者: [ahmed h. qureshi](#), [mayur j. byny](#), [michael c. yip](#)

摘要: 快速高效的运动规划算法对于许多最先进的机器人应用 (如自驾游) 至关重要。现有的运动规划方法 (如 rt^* 、 a^* 和 d^*) 变得无效, 因为它们的计算复杂性随着运动规划问题的维数而呈指数级增长。为了解决这个问题, 我们提出了一种基于神经网络的新规划算法, 该算法生成端到端无碰撞路径, 而不考虑障碍物的几何形状。该方法称为 `mpnet` (运动规划网络), 包括直接从点云测量对给定工作空间进行编码的收缩自动编码器和采用工作空间编码的深度前馈神经网络, 启动和目标配置, 并生成端到端可行的运动轨迹, 供机器人遵循。我们评估 `mpnet` 的多个规划问题, 如规划点质量机器人, 刚体, 和 7 dof 巴克斯特机器人机器人在各种 2d 和 3d 环境中。结果表明, `mpnet` 不仅在所有二维和三维环境中具有一致的计算效率, 而且对完全看不见的环境也表现出显著的泛化效果。结果表明, `mpnet` 的计算时间始终小于 1 秒, 明显低于现有的最先进的运动规划算法。此外, 通过转移学习, 在一种场景 (如室内生活场所) 中接受培训的 `mpnet` 还可以在数据量较小的情况下迅速适应新的场景 (如工厂车间)。少

2018 年 6 月 14 日提交; 最初宣布 2018 年 6 月。

35. 第 xiv:1806.03891[[pdf](#),[其他](#)] Cs. 简历

基于深度的 6d 对象姿势和人群场景中的联合注册的多任务深部网络

作者: [juil sock](#), [kwang in kim](#), [caner sahin](#), [tae-kyun kim](#)

摘要: 在挑选垃圾的场景中, 感兴趣对象的多个实例随机堆叠在一个堆中, 因此, 这些实例本质上受到挑战的影响: 严重的遮挡、杂乱和类似的干扰。但是, 大多数现有方法都适用于单个独立对象实例, 而最近的一些方法将人群场景作为后细化来处理, 这些方法考虑到多个对象关系。本文通过深度神经网络中的多任务学习, 在深度选择场景中, 讨论了多实例恢复 6d 姿态的深度模式。我们的架构共同学习多个子任务: 2d 检测、深度和单个对象的三维姿态估计; 和多个对象的联合注册。对于训练数据生成, 物理模拟中的 3d 对象模型生成物理上合理的对象姿态配置的深度图像, 从而生成不同的遮挡模式以供学习。我们采用了最先进的目标检测器, 并通过网络进一步估计 2d 偏移量, 以优化不对齐的 2d 检测。深度和三维姿态估计器的设计是为了产生多个假设, 每次检测。这使得联合注册网络能够学习遮挡模式, 并删除物理上不可信的姿势假设。我们将我们的体系结构应用于合成 (我们自己的和硅烷数据集) 和真实 (公共取金数据集) 数据, 显示它的平均精度明显优于最先进的方法 15-31。少

2018 年 6 月 11 日提交; 最初宣布 2018 年 6 月。

36. 第 xiv:1806.01759[[pdf](#),[其他](#)] Cs. 简历

多伊 [10.114/3272127.3275110](#)

非统一采样点云学习的蒙特卡罗卷积

作者: [pedro hermosilla](#), [tobiasritschel](#), [pere-pau vázquez](#), [Álvar vinacua](#), [timo 罗平斯基](#)

摘要: 深度学习系统广泛使用卷积操作来处理输入数据。尽管卷积对于结构化数据 (如 2d 图像或 3d 卷) 有明确的定义, 但对于其他数据类型 (如稀疏点云), 情况并非如此。以前的技术已经发展到限制条件下的卷积近似。遗憾的是, 它们的适用性有限, 不能用于一般点云。我们提出了一种学习非均匀采样点云的卷积的有效方法, 因为它们是用现代采集技术获得的。学习有四个关键的新事物: 第一, 将卷积内核本身表示为多层感知器; 第二, 将卷积表述为蒙特卡罗积分问题, 第三, 利用这一概念将不同层次的多个样本中的信息结合起来; 第四, 将泊松磁盘采样作为分层点云学习的可扩展手段。所有这些贡献的关键思想是保证从蒙特卡罗的角度充分考虑底层的非均匀样本分布函数。为了使建议的概念适用于实际任务, 我们还提出了一个有效的实现, 显著减少了培训过程中所需的 gpu 内存。通过在分层网络架构中采用我们的方法, 我们可以在已建立的点云分割、分类和正常估计基准上超越大多数最先进的网络。此外, 与大多数现有方法不同的是, 我们还展示了我们的方法在采样变化方面的鲁棒性, 即使仅使用一致采样数据进行训练也是如此。为了支持这些概念的直接应用, 我们提供了这些层的现成使用的 tensorflow 实现, <https://github.com/viscom-ulm/MCCNN> 少

2018 年 9 月 25 日提交; v1 于 2018 年 6 月 5 日提交; 最初宣布 2018 年 6 月。

评论: 关于图形的交易 (2018 年亚洲统计展的提案)

37. 第 1805.511504[[pdf](#), [ps](#), [其他](#)] Cs. 简历

利用生成对抗性的网络从计算机层析成像图像中捕获变异性

作者: [umair javaid](#), [john a. lee](#)

摘要: 随着深度学习(dl) 技术 (尤其是生成性对抗网络 (gans) 的出现, 数据的增强和生成正在快速发展, 这些领域最近引起了人们的极大兴趣。然而, dl 技术对数据要求很高, 由于医疗数据不易访问, 它们受到数据不足的影响。为了解决此限制, 使用了不同的数据增强技术。在这里, 我们提出了一种新的无监督数据驱动的数据增强方法, 它可以使用简单的 gan 生成二维计算机断层扫描(ct) 图像。生成的 ct 图像具有良好的全局和局部特征, 可用于增强训练数据集, 以便进行有效学习。在这项概念验证研究中, 我们表明我们提出的使用 gans 的解决方案能够捕获一些全局和局部 ct 变量。我们的网络能够生成逼真的 ct 图像, 我们的目标是通过将其缩放到更高的分辨率, 并有可能从 2d 到 3d 进一步增强其输出。少

2018 年 5 月 29 日提交; 最初宣布 2018 年 5 月。

日记本参考: 欧洲人工神经网络、计算智能和机器学习研讨会 (esann) 论文集, 第 4403-408 页, 2018 年 4 月 25 日至 27 日

38. 第 1805.01026[[pdf](#), [其他](#)] Cs. 简历

用黎曼几何计算 cnn 损耗和测距

作者: [Hajnal hou](#), [nina miolane](#), [bisesh khanal](#), [matthew c. h. lee](#), [amir alansary](#), [steven mcdonagh](#), [jo v. hajnal](#), [daniel rüeckert](#), [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。

39. 第 1804. 01110[[pdf](#),[其他](#)] Cs. 简历

三维人体姿态估计的无监督几何感知表示

作者:[helge rhodin](#), [mathieu salzmänn](#), [pascal fua](#)

文摘: 现代**三维**人体姿态估计技术依赖于深网络, 这需要大量的训练数据。虽然弱监督方法需要较少的监督, 通过使用 **2d** 姿势或没有注释的多视图图像, 他们仍然需要足够大的样本集与 **3d** 注释, 以学习成功。在本文中, 我们提出要通过学习一个几何感知身体表示从多视图图像没有注释, 以克服这个问题。为此, 我们使用编码器解码器, 从一个角度预测一个图像, 从另一个角度给出一个图像。由于此表示对 **3d** 几何进行编码, 因此在半监督环境中使用它可以更轻松地学习从它到 **3d** 人体姿势的映射。正如我们的实验所证明的, 在相同数量的标记数据下, 我们的方法明显优于完全监督的方法, 并且比其他半监督方法改进, 同时使用的标记数据只有 1%。少

2018 年 4 月 3 日提交;最初宣布 2018 年 4 月。

40. 第 1803. 11404[[pdf](#),[其他](#)] Cs. 简历

交叉模态深变手的剂量估计

作者:[adrian 刺激](#), [jie song](#), [seonwook 公园](#), [otmar hilliges](#)

摘要: 人工手以复杂和高维的方式移动, 使仅从图像中估计 **3d** 手的姿势配置成为一项具有挑战性的任务。在这项工作中, 我们提出了一个方法来学习一个统计手模型代表一个交叉模态训练的潜在空间, 通过生成的深层神经网络。我们从 vae 框架的变分下界推导出一个目标函数, 并共同优化由此产生的交叉模态 kl 发散和后验重建目标, 自然承认一个训练系统, 从而导致一个连贯的潜在跨多个模式的空间, 如 rgb 图像、**2d** 关键点检测或 **3d** 手配置。此外, 它还提供了一种使用半监督的直接方法。这个潜在空间可以直接用于从 rgb 图像中估计 **3d** 手的姿势, 在不同的设置中超过了最先进的。此外, 我们还证明了我们提出的方法可以在不改变深度图像的情况下使用, 并与专门的方法相比。最后, 该模型是完全生成的, 可以综合一致的手配置对跨模式。我们在 rgb 和深度数据集上对我们的方法进行了评估, 并对潜在空间进行了定性分析。少

2018 年 3 月 30 日提交;最初宣布 2018 年 3 月。

41. 第 1803. 10932[[pdf](#),[其他](#)] Cs. 简历

学习三维对象重建的自由形式变形

作者:[dominic jack](#), [jhony k. pontes](#), [sridha sridharan](#), [clinton fookes](#), [sareh shirazi](#), [fredic maire](#), [anders eriksson](#)

摘要: 以准确、高效和紧凑的方式在深度学习框架中代表 3d 形状仍然是一个开放的挑战。大多数现有工作通过使用基于体素的表示来解决此问题。虽然这些方法通过将二维卷积推广到 3d 设置, 从计算机视觉的进步中受益匪浅, 但它们也有几个相当大的缺点。随着分辨率的提高, 体素编码的计算复杂度逐渐增大, 从而将这种表示限制为低分辨率的 3d 重建。为了解决这一问题, 提出了点云表示。尽管点云比体素表示更有效, 因为它们只覆盖曲面而不是体积, 但它们不会对有关点之间关系的详细几何信息进行编码。本文提出了一种从单个图像中学习三维重建任务的自由变形 (ffd) 方法。通过学习变形从高质量网格中采样的点, 我们训练的模型可用于生成任意密集的点云或具有细粒度几何的网格。我们根据合成数据和真实世界数据评估我们提出的框架, 并在点云和体积指标上实现最先进的结果。此外, 我们还定性地论证了它在三维语义分割标签传输中的适用性。少

2018 年 3 月 29 日提交;最初宣布 2018 年 3 月。

评论:16 页, 7 个数字, 3 个表

42. 第 1803.08181 [pdf,其他] 反渗透委员会

天网: 利用三维空间变压器网络进行自我监控的外部校准

作者:ganesh iyer, karnik ram r., j. krishna murthy, k. madhava krishna

摘要: 3d lidars 和 2d 摄像机越来越多地在传感器钻机中与对方一起使用, 以完成感知任务。但是, 在使用这些传感器收集有意义的数据之前, 需要准确校准它们的外部特性(和本质), 因为传感器钻机的性能对这些校准参数极为敏感。绝大多数现有的校准技术都需要大量的数据和校准目标以及人力, 这严重影响了它们在大型生产系统中的适用性。我们解决了与 calibre 网的这一差距: 一个自我监督的深网络, 能够实时估计 3d lidar 和 2d 摄像机之间的 6-dof 刚体变换。calibre 网络减少了对校准目标的需求, 从而显著节省了校准工作的成本。在训练过程中, 网络只输入一个 lidar 点云、相应的单目图像和摄像机校准矩阵 k。在列车运行时间, 我们不进行直接监督 (例如, 我们不会直接回归校准参数)。相反, 我们训练网络来预测校准参数, 以最大限度地提高输入图像和点云的几何和光度一致性。calibre 网络学习迭代解决潜在的几何问题, 并准确预测各种错误校准的外部校准参数, 而无需再培训或域适应。项目页面托管在 <https://epiception.github.io/CalibNet>

2018 年 3 月 21 日提交;最初宣布 2018 年 3 月。

评论:提交给 iee 智能机器人和系统国际会议 (iros) 2018

43. 第 xiv:1808071 [pdf,其他] Cs. 简历

基于零特征值的深部网络无征训练

作者:郑当,光墨义,胡银林, 王飞, 帕斯卡尔·福亚,马蒂厄·萨尔兹曼

文摘: 许多经典的计算机视觉问题, 如基本矩阵计算和从三维到二维对应的姿态估计, 可以通过找到与矩阵的最小或零特征值相对应的特征向量来解决表示线性系统。将此纳入深度学习框架将使我们能够显式地编码已知的几何概念, 而不是让网络隐式地从数据中学习这些概念。但是, 在网络中执行特征定位需要区分此操作的能力。遗憾的是, 这虽然在理论上是可行的, 但在实际优化过程中引入了数值不稳定性。本文介绍了一种无特征值的深网训练方法, 其损耗取决于与网络预测的矩阵的零特征值相对应的特征向量。我们在几个任务上证明了这一点, 包括关键点匹配和三维姿态估计, 我们的方法比特征的显式微分要可靠得多, 它具有更好的收敛特性, 并产生了最先进的结果两个任务的结果。少

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

评论:25 页

44. 第 18002286[[pdf](#),[其他](#)] 反渗透委员会

利用密集的 3d 流的密集三维绘图学习单目视觉测深

作者:[赵成](#),[孙丽](#),[普拉克](#),[汤姆·杜克特](#),[鲁斯塔姆·斯托尔金](#)

文摘: 本文介绍了一种全深度学习单目 slam 的方法, 该方法可以利用神经网络进行同步定位, 用于学习视觉测量 (l-vo) 和密集的三维绘图。密集的二维流和深度图像是由子网络从单目图像中生成的, 然后由 l-vo 网络中的三维流关联层用于生成密集的三维流。在这种三维流动的情况下, 双流 l-vo 网络可以预测 6dof 的相对姿态, 进而重建车辆轨迹。为了了解运动方向之间的相关性, 在损耗函数中采用了双变量高斯建模。l-vo 网络的总体性能为 2.68, 平均平移误差为 2.68, 在 kitti 测速基准上, 平均旋转误差的总体性能为 0.0143 分。此外, 还充分利用学习到的深度生成密集的 3d 地图。因此, 实现了整个视觉 slam 系统, 即学习单目测深与密集的 3d 映射相结合。少

2018 年 7 月 25 日提交;v1 于 2018 年 3 月 6 日提交;最初宣布 2018 年 3 月。

评论:智能机器人与系统国际会议 (iros 2018)

日记本参考:智能机器人与系统国际会议 (iros 2018)

45. 第: 1802.09975[[pdf](#),[其他](#)] Cs。简历

基于深度学习检测和 pmbm 滤波的单摄像机 3d 多目标跟踪

作者:[samuel scheidegger](#),[joachim benjaminsson](#),[emil rosenberg](#),[amrit krishnan](#),[karl granstro](#)

摘要: 单目摄像机是汽车行业中最常用的自主车辆传感器之一。使用单目摄像机的一个主要缺点是, 它只在二维图像平面上进行观测, 不能直接测量与物体的距离。在本文中, 我们的目标是通过开发一个多目标跟踪算法来填补这一空白, 该算法以图像为输入, 并在世界坐标系中产生检测到的物体的轨迹。我们通过使用经过训练的深度神经网络来检测和估计单个输入图像与对象的距离来解决这个问题。从一系列图像中检测到的图像被输入到最先进的泊松多伯努利混合跟踪滤波器中。学习的检测器和 pmbm 滤波器的组合导致了一种仅使用单摄像机图像作为输入来实现 3d 跟踪的算法。使用公开的 kitti 对象跟踪数据集, 在 3d 世界坐标和二维图像坐标中评估算法的性能。该算法显示了准确跟踪对象、正确处理数据关联的能力, 即使图像中的对象有很大的重叠, 也是 kitti 对象跟踪基准上性能最好的算法之一。此外, 该算法还很有效, 平均每秒运行近 20 帧。少

2018 年 2 月 27 日提交;最初宣布 2018 年 2 月。

评论:8 页, 2 个数字, 相关视频, 请参阅 <https://goo.gl/AoydgW>

46. 第 1802.09971[[pdf](#),[其他](#)] Cs。简历

按分区、草和卷曲分列的真实世界重复估计

作者:[tom f. h. runia](#),[cees g. m. snoek](#),[arnold w. m. 斯梅尔斯](#)

摘要: 我们考虑的问题是估计视频中的重复, 比如表演俯卧撑、切瓜或拉小提琴。现有工作在静态和平稳周期的假设下取得了较好的效果。由于逼真的视频很少是完全静态和静止的, 通常喜欢的基于傅立利的测量是不合适的。相反, 我们采用小波变换来更好地处理非静态和非平稳视频动力学。从流场及其微分中, 我们推导出三维内在周期性的三种基本运动类型和三种运动连续性。除此之外, 3d 周期的 2d 感知还考虑了两个极端

的观点。以下是在 **2d** 中反复感知的 18 例基本案例。在实践中, 为了处理重复外观的多样性, 我们的理论意味着测量时变流及其在分段前景运动上的差异 (梯度、发散和卷曲)。对于实验, 我们引入了新的 quva 重复数据集, 通过包含非静态和非静止视频来反映实际情况。在计算视频重复的任务上, 与深度学习的替代方案相比, 我们获得了良好的效果。少

2018 年 2 月 27 日提交;最初宣布 2018 年 2 月。

47. [建议: 180008824](#)[pdf,其他] 反渗透委员会

基于深层强化学习的室内机器人视觉导航

作者:mo kaichun, haxiangli, zhe lin, joon-young lee

文摘: 深度强化学习(drl) 展示了其在学习机器人视觉导航无模型导航策略方面的潜力。然而, 在训练中, 数据需求算法依赖于大量的导航轨迹。支持训练此类机器人导航算法的现有数据集包括 **3d** 合成场景或重建场景。合成数据与真实世界场景的域间隙存在影响, 而从 **3d** 重建场景中呈现的视觉输入则存在不需要的漏洞和伪影。本文提出了一个在现实世界中收集的新数据集, 以方便基于 drl 的视觉导航研究。我们的数据集包括真实场景的 **3d** 重建以及密集拍摄的真实 **2d** 图像。它为机器人在密集的网络位置提供高质量的视觉输入, 并具有真实的场景复杂性。进一步研究和测试了一种最新的基于 drl 的导航算法, 并提出了提高其对场景中看不见的测试目标的通用性的尝试和思考。少

2018 年 2 月 24 日提交;最初宣布 2018 年 2 月。

48. [建议: 1802.0 01500](#)[pdf,其他] Cs. 简历

多伊 [10.1109/ICCVW.2017.90](#)

点云三维语义分割的空间语境探索

作者:francis engelmann, theodora kontogianni, 亚历山大·赫尔曼斯, 巴斯蒂安·莱贝

摘要: 深部在过去的几年里,学习方法在语义分割领域取得了巨大的进步。但是, 大多数当前的方法都在 **2d** 图像空间中运行。非结构化三维点云的直接语义分割仍是一个研究热点。最近提出的 pointnet 体系结构向前迈出了有趣的一步, 因为它可以在非结构化点云上运行, 从而实现令人鼓舞的分段结果。但是, 它将输入点细分为块的网格, 并单独处理每个此类块。在本文中, 我们研究了如何扩展这样的体系结构以纳入更大规模的空间环境的问题。我们建立在 pointnet 的基础上, 并提出了两个扩展, 扩大了接受领域在 **3d** 场景。我们评估了关于具有挑战性的室内和室外数据集的拟议策略, 并在这两种情况下显示了改进的结果。少

2018 年 2 月 5 日提交;最初宣布 2018 年 2 月。

49. [xiv:1801. 09128](#)[pdf,其他] Cs. 简历

对 **3d** 重构中的纠错方法

作者:michael tanner, stefan saftescu, alex bewley , paul newman

摘要: 密集重建通常包含以前使用高质量传感器和调节输出的工作已减少的错误。然而, 错误仍然存在。本文提出了一种机器学习技术来识别三维网格中的误差。除了简单地识别误差外, 我们的方法还在查看场景时量化深度估计误差的大小和方向。这使我们能够提高重建精度。我们训练一个具有两个 **3d** 网络的合适的深网络架构: 高质量的激光重建和更高质量的立体图像重建。该网络预测了与高质量重建相比的质量较低的重建中的误差量, 只通过其输入来查看前者。通过修正从三维模型中提取的二维 (**2d**) 反向

深度图像, 对该方法进行了评估, 并表明该方法将这些深度重建的质量提高了 10% 的相对 rmse。少

2018 年 1 月 27 日提交;最初宣布 2018 年 1 月。

评论:2018 年机器人与自动化国际会议 (icra) 被接受

50. **建议: 1801. 1.08599**[pdf,其他] Cs。简历

深度逻辑 mos: 基于深度学习图的胰腺肿瘤 ct 扫描三维分割

作者:郭志辉,张玲, 乐乐, 穆罕默德·巴格赫里, 罗纳德·萨默斯, 米兰·松卡, 姚建华

文摘: 本文报道了将深层上下文学习中的边界信息与 logismos 分层的多层最优图形图像分割相结合, 实现三维肿瘤分割的深度逻辑对象和曲面。准确可靠的肿瘤分割对肿瘤生长分析和治疗选择至关重要。一个完全卷积的网络 (fcn) unet 首先使用以肿瘤为中心的三个相邻的 2d 补丁进行训练, 为每个 2d 补丁提供上下文 unet 分割和概率图。然后利用高斯混合模型 (gmm) 和形态学运算对 unet 分割进行细化。利用改进后的 unet 分割方法提供初始形状边界, 以建立分割图。图形的每个节点的成本由 unet 概率映射确定。最后, 采用最大流算法求全局最优解, 从而获得最终的分割。为了进行评估, 我们将该方法应用于胰腺肿瘤的分割, 对 51 例 ct 扫描的数据集进行了分析, 其中 30 次扫描用于训练, 21 次用于检测。与上下文 unet 和或 logismos 相比, 深度逻辑、dsc (dsc) 和相对体积差 (rvd) 分别达到 83.2 \pm 7.8% 和 18.6% \pm 17.4%, 两者均有显著改善 ($p < 0.05$)。少

2018 年 1 月 25 日提交;最初宣布 2018 年 1 月。

评论:4 页, 3 个数字

51. **建议: 1801. 05671**[pdf,其他] 反渗透委员会

多伊 10.114/3171221

人机交互的仿人机器人的紧凑型实时避碰

作者:dong hai phuong nguyen, matej hoffmann, alessandro roncone, ugo patacini, giorgio metta

摘要: 随着机器人离开工厂, 进入控制较少的领域, 可能与人类共享空间, 安全至关重要, 对车身表面和周边环境的多模式意识至关重要。我们从人类的个人空间表示中汲取灵感, 在一个人形机器人上提出了一个框架, 该框架动态地维护着这样一个保护性的安全区域, 由以下主要组成部分组成: (i) 人类 2d 关键点估计管道采用基于深度学习的算法, 扩展到基于视差的三维;(ii) 围绕机器人身体部位的分布式个人空间表示;(iii) 一个到达控制器, 它包含了进入机器人安全区域的所有障碍物。试点实验表明, 机器人和人体部位之间保持了有效的安全边际。建议的解决方案是灵活和多才多艺的, 因为单个机器人和人体部位周围的安全区域可以有选择地进行调制--在这里, 我们展示了更强的避免人头相比, 其余的身体。我们的系统是实时工作的, 是独立的, 没有外部感官设备, 只使用机载摄像头。少

2018 年 1 月 17 日提交;最初宣布 2018 年 1 月。

日记本参考:hri "18:2018 acm/ieee 人机互动国际会议, 2018 年 3 月 5 日至 8 日, 美国伊利诺伊州芝加哥

52. **建议: 1801. 05297**[pdf,其他] 反渗透委员会

利用深度学习进行证据占用网格增强

作者:sascha wirges, felix hartenbach, christoph 史迪勒

摘要: 详细的环境表示是自动化车辆的重要组成部分。使用单一范围传感器扫描, 数据往往过于稀疏, 容易被遮挡。因此, 我们提出了一种从单一视图增加占用度网格地图的方法, 其类似于使用深度学习从不同视图获取的证据占用图。为了实现这一目标, 我们估计后续范围传感器测量之间的运动, 并在广泛的后处理步骤中创建一个证据 3d 体素贴图。在这个体素映射中, 我们用证据理论显式地建模不确定性, 并使用组合规则创建二维投影。作为神经网络的输入, 我们使用多层网格映射, 该网格由三个特征检测、传输和强度组成, 每个要素都用于地面和非地面测量。最后, 对不同的网络体系结构进行了定量和定性的评价, 表明不同的网络体系结构能够实时准确地推断证据度量。少

2018 年 4 月 18 日提交;v1 于 2018 年 1 月 16 日提交;最初宣布 2018 年 1 月。

评论:7 页, 5 个数字

53. 成果: 180003431[[pdf](#),[其他](#)] Cs。简历

从二维组织学图像中推断第三个空间维度

作者: [maxime w. lafarge](#), [josien p.w. pluim](#), [koen a. j. epenhof](#), [pim moeskops](#), [mitko veta](#)

摘要: 组织学图像是通过经过染色的组织标本传输光而获得的, 以便产生对比度。这一过程产生了具有三维结构的样品的二维图像。在本文中, 我们提出了一种方法来推断污渍是如何分布在垂直于幻灯片表面的方向上的, 以获得组织的三维表示。这种推断是通过在约束下分解染色浓度图来实现的, 以确保原始二维图像的逼真分解和重建。我们的研究表明, 在训练深度学习模型时, 可以生成逼真的三维图像, 使这种方法成为数据增强的潜在工具。少

2018 年 1 月 10 日提交;最初宣布 2018 年 1 月。

评论:ieee 生物医学成像国际研讨会 (isbi), 2018

54. xiv:1801. 01949[[pdf](#),[其他](#)] Cs。简历

人脸闪烁: 一种基于光反射的安全活动检测协议

作者: [唐迪](#), [周哲](#), [张银谦](#), [张克环](#)

摘要: 人脸认证系统越来越普遍, 尤其是随着深度学习技术的快速发展。然而, 人类面部信息很容易被捕获和复制, 这使得面部身份验证系统容易受到各种攻击。活动检测是防止此类攻击的重要防御技术, 但现有的解决方案并不能提供明确而有力的安全保障, 特别是在时间方面。为了克服这些限制, 我们提出了一种名为 "人脸闪烁" 的新的活动检测协议, 该协议大大提高了在人脸身份验证系统上成功发起攻击的门槛。通过在屏幕上随机闪烁设计良好的图片并分析反射光, 我们的协议利用了人脸的物理特征: 以光速进行反射处理、独特的文本特征和不均匀的 3d 形状。我们的协议与屏幕和数码相机的工作机制合作, 能够检测到攻击过程中留下的微妙痕迹。为了演示人脸闪烁的有效性, 我们实现了一个原型, 并对从实际场景中收集的大型数据集进行了彻底的评估。结果表明, 我们的时间验证能够有效地检测合法身份验证与恶意案例之间的时间差距。我们的人脸验证还可以准确地区分二维平面和 3d 对象。我们的活动检测系统的整体精度为 98.8%, 并在不同的情况下对其鲁棒性进行了评估。在最坏的情况下, 我们的系统的精度下降到仍然很高的 97.3%。少

2018 年 8 月 22 日提交;v1 于 2018 年 1 月 5 日提交;最初宣布 2018 年 1 月。

55. 建议: 1801. 01258[[pdf](#),[其他](#)] Cs。简历

9 视图双能 ct 行李扫描器的深度学习改造

作者:韩义夫,康经固,叶钟哲

文摘: 在国土和运输安全应用中,二维 x 射线爆炸物探测系统 (eds) 得到了广泛的应用,但在识别隐藏物体的三维形状方面存在局限性。在各种类型的三维计算机断层扫描 (ct) 系统中,为了解决这一问题,本文对使用固定 x 射线源和探测器的固定 ct 感兴趣。然而,由于投影视图的数量有限,解析重建算法产生了严重的条纹伪影。在稀疏视图 ct 重建深度学习方法近期成功的启发下,本文提出了一种新的图像和形图域深度学习体系结构,用于三维重建。稀疏视图测量。该算法已被测试的实际数据的原型 9 视双能量固定 ct eds 随身行李扫描仪开发的 gemss 医疗系统,韩国,这证实了优越的重建性能比现有的方法。少

2018 年 1 月 4 日提交;最初宣布 2018 年 1 月。

56. 第 171207262[pdf,其他] Cs。简历

折网: 点云自动编码器通过深网变形

作者:杨耀清,陈峰,沈一如,董田

摘要: 最近直接处理点集中点的深度网络 (例如 pointnet) 对于点云上的监督学习任务 (如分类和分段) 来说是最先进的。在这项工作中,提出了一种新的端到端深自动编码器,以解决点云上的无监督学习挑战。在编码器方面,实施了基于图形的增强,以促进 pointnet 之上的局部结构。然后,一种新的基于折叠的解码器将规范的 2d 网格变形到点云的底层 3d 对象表面上,即使对于结构精细的对象,也能实现较低的重建误差。该解码器仅使用完全连接神经网络的解码器的 7% 左右的参数,但却得到了比基准更高的线性 svm 分类精度的判别表示。此外,理论上,所提出的解码器结构是一个通用的体系结构,能够从 2d 网格重建任意点云。我们的代码可

<http://www.merl.com/research/license#FoldingNet>

2018 年 4 月 2 日提交;v1 于 2017 年 12 月 19 日提交;最初宣布 2017 年 12 月。

评论:作为 cvpr18 的焦点文件接受

57. 第 171207999[pdf,其他] Cs。简历

面部 uv 地图的多属性鲁棒分量分析

作者:stylianos moschoglou, evangelos ververas, yannis panagakis, mihalis nicoraou, stefanos zafeiriou

文摘: 近年来,由于大规模三维人脸模型的收集,以及深度学习的出现,在 3d 人脸对齐 "野外" 领域取得了重大进展。也就是说,在二维人脸图像和三维人脸模型之间建立稀疏或密集的三维对应的方法很多。三维人脸对齐的应用带来了新的挑战和研究方向,特别是在面部纹理图像的分析方面。特别是,纹理不再受到翘曲效果的影响 (使用 2d 人脸对齐方法时发生)。然而,由于面部图像通常是在任意记录条件下拍摄的,因此会观察到大量缺失的信息和总异常值 (例如,由于自我遮挡,或被试戴眼镜)。鉴于为面部分析任务开发了许多附加说明的数据库,显然需要开发成分分析技术,以缓解上述挑战引起的问题。本文提出了一种新的成分分析技术,该技术适用于包含大量缺失信息和异常值的面部 uv 地图,同时还融合了来自各种属性 (如年龄和身份) 的知识。我们评估了所提出的多属性鲁棒组件分析 (ma-rca) 的问题,如 uv 完成和年龄进展,其中建议的方法优于比较技术。最后,我们证明了 ma-rca 方法的强大功能足以作为各种应用 (如照明传输) 的深度学习系统的训练提供弱注释。少

2017 年 12 月 15 日提交;最初宣布 2017 年 12 月。

58. 第 171204101[pdf,其他] Cs。Lg

多伊 10.114/3167132.3 167165

以外部知识为助推的深层强化学习

作者: nicolas bougie, ryutaro icise

文摘 最近在深度强化学习方面的改进解决了许多 2d 领域的问题, 如 atari 游戏。然而, 在复杂的 3d 环境中, 需要大量的学习插曲, 这些学习插曲可能过于耗时, 甚至不可能, 尤其是在真实场景中。我们提出了一个新的架构, 结合外部知识和深度强化学习只使用视觉输入。我们系统的一个关键概念是通过添加环境特征信息和结合两个决策源来增加图像输入。我们从 microsoft malmo 平台评估我们的方法在部分可观察的 3d 环境中的性能。与单一的强化学习模型相比, 实验评价具有更高的性能和更快的学习速度。少

2017 年 12 月 11 日提交;最初宣布 2017 年 12 月。

类: l.2。6

59. 第 1712.03866[pdf,其他] Cs。简历

在野外使用单一 rgb 帧进行实时 3d 手部姿态估计

作者: paschalis panteleris, iason oikonomidis, antonis argyros

文摘: 我们提出了一种使用单一商品 rgb 相机实时估计一个或多个人的手的整个三维姿势的方法。最近在这一领域的工作在使用 rgb-d 投入方面取得了令人印象深刻的进展。然而, 自引入 rgb-d 传感器以来, 单目颜色输入的情况进展甚微。我们利用深度学习的最新进展, 将它们与生成手姿势估计技术的力量相结合, 在不受限制的情况下实现实时单目 3d 手部姿态估计。更具体地说, 考虑到 rgb 图像和相关的摄像机校准信息, 我们采用最先进的检测器来本地化手。考虑到图像中的一只手, 我们运行了预先训练的 opose 手网络, 以估计手关节的 2d 位置。最后, 非线性最小二乘最小化将手的三维模型与估计的二维关节位置相匹配, 恢复三维手部姿势。广泛的实验结果提供了与最新技术的比较, 以及对该方法在野外的定性评估。少

2017 年 12 月 11 日提交;最初宣布 2017 年 12 月。

60. 第: 1712.01600[pdf,其他] Cs。简历

具有丰富光谱内容的遥感图像语义分割的深度学习

作者: a hamida, a. Benoît, p. lambert, l klein, camar, n. audebert, s.lefèvre

摘要: 随着遥感采集技术的快速发展, 有必要扩大和改进处理工具, 以应对观测到的数据量和丰富度的增加。在遥感领域流行的技术中, 深度学习越来越受到人们的兴趣, 但取决于培训数据的质量。因此, 本文提出了近年来精细或粗放土地覆被语义分割估计的深度学习方法。对各种二维体系结构进行了测试, 并引入了一种新的三维模型, 以联合处理数据的空间和光谱尺寸。这样的网络集可以比较不同的光谱融合方案。此外, 我们还评估了在培训和测试网络时使用 "嘈杂的地面真相" (即过时和低空间分辨率标签) 的情况。少

2017 年 12 月 5 日提交;最初宣布 2017 年 12 月。

评论: ieee 国际地球科学与遥感研讨会, 2017 年 7 月, 美国沃思堡

61. 第 1711. 09312[pdf, ps,其他] Cs。简历

通过对抗学习从单个图像中进行无监督的 3d 重建

作者: 王凌静, 易芳

摘要: 最近在深度学习方面取得的进展为从单个 2d 图像中学习高质量的 3d 模型提供了新的机会, 并对大规模数据集进行了充分的培训。然而, 可用的图像量和 3d 模型之间的显著不平衡, 以及标记的 2d 图像数据 (即图像与其相应的 3d 模型之间的手动注释对) 的可用性有限, 在实践中, 对最受监督的深度学习方法的训练产生了严重影响。在对抗网络的新设计的推动下, 我们开发了一个无监督的学习范式, 从一个 2d 图像重建三维模型, 该模型不需要手动注释成对输入图像及其关联的 3d 模型。特别是, 该范例从通过自动编码器训练适应网络和对抗性损失开始, 将未配对的 2d 合成图像域与真实世界图像域嵌入到一个共享的潜在向量空间中。然后, 我们结合训练一个三维脱色网络, 将潜在的向量空间与嵌入过程一起转化为三维对象空间。我们的实验验证了我们的网络在处理从单个二维图像生成 3d 体积物体方面的鲁棒性和卓越性能。少

2017 年 11 月 25 日提交;最初宣布 2017 年 11 月。

62. 第: 1711.08588[[pdf](#),[其他](#)] Cs。简历

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 2dcnn 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。

63. 第: 1711.08488[[pdf](#),[其他](#)] Cs。简历

基于 rgb-d 数据的三维对象检测

作者:[夏尔·齐](#),[刘伟](#),[吴晨霞](#),[苏浩](#),[列奥尼达·吉巴斯](#)

文摘: 在本工作中, 我们研究了从 rgb-d 数据在室内和室外场景中的三维目标检测。虽然以前的方法侧重于图像或 3d 体素, 通常掩盖了自然的 3d 模式和 3d 数据的不变性, 但我们通过弹出 rgb-d 扫描直接在原始点云上操作。然而, 这种方法的一个关键挑战是如何有效地定位大型场景点云中的对象 (区域建议)。我们的方法不是仅仅依靠 3d 建议, 而是利用成熟的 2d 对象探测器和高级 3d 深度学习来实现对象定位, 从而实现效率和高召回率。甚至小的物体。我们的方法直接从原始点云中学习中受益, 即使在强遮挡或点非常稀疏的情况下, 也能够精确估计 3d 边界框。通过对 kitti 和 sun rgb-d 3d 检测基准的评估, 我们的方法在具有实时性能的同时, 以卓越的边缘性能超过了最先进的技术。少

2018 年 4 月 12 日提交;v1 于 2017 年 11 月 22 日提交;最初宣布 2017 年 11 月。

评论:15 页, 12 个数字, 14 张表格

64. 第: 1711.07888[[pdf](#),[其他](#)] Cs。简历

银器: 从剪影中学习的单视图和多视图重建

作者:[olivia wiles](#), [andrew zisserman](#)

文摘: 本文的目的是从单个和多个图像中理解三维形状。为此, 我们引入了一种新的深度学习体系结构和损耗函数 silnet, 它可以以与订单无关的方式处理多个视图. 该架构是完全卷积的, 对于训练, 我们使用轮廓预测的代理任务, 而不是像最近的工作中的目标那样直接学习从 2d 图像到 3d 形状的映射。我们证明, 使用 silnet 体系结构可以概括视图的数量--例如, 在 2 个视图上训练的 silnet 可以在测试时与 3 或 4 个视图一起使用;和性能提高与更多的视图。我们引入了两个新的合成数据集: 一个是用于预训练的模糊对象数据集, 另一个是具有挑战性和现实的雕塑数据集;并在这些数据集中证明 silnet 确实学会了 3d 形状。最后, 我们展示了 silnet 超过了 shapenet 基准数据集上的最新技术, 并使用 silnet 生成雕塑数据集的新视图。少

2017 年 11 月 21 日提交;最初宣布 2017 年 11 月。

评论:bmvc 2017;最佳海报

65. 第: 1711.07479[[pdf](#),[其他](#)] 反渗透委员会

教机器阅读地图与深度强化学习

作者:[gino brunner](#), [oliver richter](#), [yuyi wang](#), [roger wattenhofer](#)

摘要: 使用 2d 地图在复杂的 3d 环境中导航的能力是相当显著的, 对许多人来说甚至是困难的。本地化和导航也是机器人等领域的一个重要问题, 近年来成为深度强化学习界关注的焦点。在本文中, 我们教一个强化学习代理阅读地图, 以找到最短的途径, 从一个随机迷宫, 它从来没有见过。我们的系统结合了几种最先进的方法, 如 a3c, 并结合了新的元素, 如反复定位单元。我们的代理学习本地化自己基于 3d 第一人称图像和近似方向角度。该代理很好地推广到了较大的迷宫, 表明它学会了有用的本地化和导航功能。少

2017 年 11 月 20 日提交;最初宣布 2017 年 11 月。

评论:第 32 届阿拉伯人工智能会议通过论文, 阿拉伯人工智能会议 2018 年, 美国路易斯安那州新奥尔良

类:l.2.0;l.2.6;l.2.9;l.2.10

66. 第 xiv:1711. 05942[[pdf](#),[其他](#)] Cs. 简历

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。虽然三维人脸识别比 2d 识别有固有的优势, 但由于没有大型培训和大型测试, 它并没有从深度学习的最新发展中受益数据。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。

评论:11 页

日记本参考:ieee 计算机视觉和模式识别会议, 2018

67. **决议: 1711.00253**[pdf,其他] Cs. 简历

地标化的结构感知全卷网络的对抗性学习

作者:余晨,沈春华,陈浩,魏秀贤, 刘灵桥,杨健

摘要: 单目图像中的地标姿态估计由于其重要的应用, 在计算机视觉中得到了很大的努力。当输入图像严重遮挡 (例如, 不良相机视图) 导致的情况下, 这仍然是一项具有挑战性的任务。在这种情况下, 可能会产生生物上不可信的姿势预测。相反, 人类视觉能够通过利用地标点互联的几何约束来预测姿势。为了解决这个问题, 通过结合关于姿势组件结构的先验, 我们提出了一种新的结构意识完全卷积网络, 在深部网络的训练中含蓄地考虑到这些先验。明确了解这些约束通常具有挑战性。相反, 在人类如何识别不可信的姿势的启发下, 我们设计了鉴别器来区分真实的姿势和虚假的姿势 (如生物上不可信的姿势)。如果姿势生成器 g 生成的结果, 鉴别器不能区分与实际的结果, 网络成功地学习了优先点。网络的培训遵循有条件生成对抗性网络 (gans) 的策略。对该网络的有效性进行了三个与假设相关的任务:二维单人姿态估计、二维面部地标估计和三维单人姿态估计。该方法的性能大大优于最先进的方法, 几乎总是生成合理的姿势预测, 表明使用 gans 的结构的内隐学习是有用的。少

2018 年 11 月 2 日提交;v1 于 2017 年 11 月 1 日提交;最初宣布 2017 年 11 月。

评论:17 页。arxiv:1705.00389 的扩展版。由 ieee trans 进行审查。模式分析和机器智能。2018 年 10 月小修订

68. **第: 1710.04934**[pdf, ps,其他] Cs. 简历

radnet: 放射科医生水平精度使用深度学习在 ct 扫描中检测

作者:monika grewal, muktabh mayank srivastava, pulkit kumar , shrikrishna varadarajan

摘要: 我们描述了一种从计算机断层扫描 (ct) 中自动检测脑出血的深度学习方法。我们的模型模拟了放射科医生在现实世界中分析三维 ct 扫描所遵循的程序。与放射科医生类似, 该模型筛选通过 2d 横截面切片, 同时密切关注潜在的出血区域。此外, 该模型还利用相邻切片中的 3d 上下文来改进每个切片的预测, 并随后聚合切片级别的预测, 以便在 ct 级别提供诊断。我们将我们提出的方法称为 "重复注意 densenet (radnet)", 因为它采用了原始的 densenet 体系结构, 同时增加了用于切片级预测和用于合并 3d 的递归神经网络层的注意组件上下文。radnet 的实际性能是根据三名高级放射科医生对 77 例脑 t 进行的独立分析确定的。radnet 在 ct 水平上的出血预测准确率为 81.82, 与放射科医生相当。此外, radnet 实现了比三名放射科医生中的两位更高的召回率, 这是很了不起的。少

2018 年 1 月 3 日提交;v1 于 2017 年 10 月 13 日提交;最初宣布 2017 年 10 月。

评论:2018 年被 ieee 生物医学成像研讨会 (isbi) 作为会议论文接受

69. **第 xiv: 170 09.06532**[pdf,其他] Cs. 简历

当三维辅助 2d 人脸识别与深度学习时: 一种用于可能不变的人脸识别的扩展 ur2d

作者:项旭,邓鹏飞,哈 a.le , ioannis a. kakadiaris

摘要: 大多数人脸识别作品都集中在特定的模块上或展示了一个研究理念。本文提出了一个具有鲁棒性的不变三维辅助二维人脸识别系统 (ur2d), 该系统能够对高达 90 的变

化进行鲁棒性。利用深度学习技术。介绍了 ur2d 的体系结构和接口, 并对各模块进行了详细的介绍。在 uhdb31 和 ijb-a 上进行了大量实验, 表明 ur2d 在 uhdb31 数据集上的性能优于现有的 2d 人脸识别系统, 如 vgg-face、facenet 和商用现成软件 (cots), 至少占 9%, 而在 uhdb31 数据集上的性能至少优于商用现成软件。在面对面识别任务中, 平均在 ijb-a 数据集上。通过比较模板匹配的 rank-1 精度得分, ur2d 还在 ijb-a 数据集上实现了 85% 的最先进性能。它通过提供一个三维辅助的 2d 人脸识别系统来填补一个空白, 该系统使用深度学习技术, 与 2d 人脸识别系统具有兼容的结果。少

2017 年 9 月 19 日提交;最初宣布 2017 年 9 月。

评论:提交给关于野生生物识别、图像和视觉计算的特刊

70. 第 xiv:1709.05665[pdf,其他] Cs。简历

机器人辅助视网膜手术立体视觉的工具地标自动检测

作者:thomas probst, kevis-kokitsi maninis, ajad chhatkuli, mouloud ourak, emmanuel vanderpoorten, luc van gool

摘要: 计算机视觉和机器人技术正越来越多地应用于医疗干预。特别是在需要极其精确的干预措施中, 它们可以有所作为。其中一种应用是机器人辅助视网膜显微手术。在最近的工作中, 这种干预是在立体显微镜下进行的, 使用的是机器人控制的手术工具。然而, 计算机视觉和机器人技术的互补性尚未得到充分利用。为了提高机器人的控制能力, 我们对解剖的三维重建和使用立体显微镜的自动工具定位感兴趣。在本文中, 我们首次使用单个管道解决了这一问题, 从未校准的摄像机开始, 在视网膜显微手术中达到公制三维重建和注册。我们的方法的关键成分是: (a) 手术工具地标检测, 和 (b) 3d 重建与立体显微镜, 使用检测到的地标。为了解决前者, 我们提出了一种新的深度学习方法, 以高于实时的速度检测和识别高清图像中的关键点。我们使用检测到的 2d 关键点及其从机器人传感器获得的相应三维坐标, 使用仿射投影模型校准立体显微镜。我们设计了一个在线 3d 重建管道, 利用平滑度约束, 并执行机器人到相机的注册。整个管道在开放天空猪眼序列上得到了广泛的验证。介绍了所有步骤的定量和定性结果。少

2017 年 11 月 20 日提交;v1 于 2017 年 9 月 17 日提交;最初宣布 2017 年 9 月。

评论:接受机器人和自动化信件 (ra-l)。项目页:

<http://www.vision.ee.ethz.ch/~kmaninis/keypoints2stereo/index.html>

71. 第: 1709.03485[pdf,其他] Cs。简历

多伊 10.1016/j.cmpb.2018.01.025

niftenet: 医学影像学的深度学习平台

作者:eli gibbon, wwqili, carole sudre, lucas fidon, dzhoshkun i.shakir, guotai wang, zach Eaton-Rosen, robert gray, tom doel, yipeng hu, tom whyntie, parashkev nachev, marc modat, dean c.barratt, sébastien ourselin, m.豪尔赫·卡多佐, tom Vercauteren

文摘: 医学图像分析和计算机辅助干预问题越来越多地得到以深度学习为基础的解决方案的解决。建立的深度学习平台是灵活的, 但不提供特定的医学图像分析功能, 并为这一应用程序进行调整需要大量的实施工作。因此, 许多研究小组之间出现了大量的工作重复和不兼容的基础设施。本工作为医学成像中的深度学习提供了开源 niftenet 平台。niftenet 的目标是加快和简化这些解决方案的开发, 并提供一个共同的机制, 传播研究成果, 供社区使用、调整和发展。niftenet 为包括分割、回归、图像生成和表示学习

应用在内的一系列医学成像应用提供了模块化的深度学习管道。niftenet 管道的组件,包括数据加载、数据扩充、网络架构、损失函数和评估指标,都是根据医学图像分析和计算机辅助的特性量身定制的,并利用这些特性干预。niftenet 建立在 tensorflow 上,默认情况下支持二维和三维图像和计算图的紧张板可视化。我们提出了 3 说明医学图像分析应用建立使用 niftenet: (1) 从计算机断层扫描多个腹部器官的分割;(2) 图像回归,预测来自脑磁共振图像的计算机断层扫描衰减图;(3) 为指定的解剖姿势生成模拟超声图像。niftenet 使研究人员能够快速开发和分发用于细分、回归、图像生成和表示学习应用程序的深度学习解决方案,或将平台扩展到新的应用。少

2017 年 10 月 16 日提交;v1 于 2017 年 9 月 11 日提交;最初宣布 2017 年 9 月。

评论:李文奇和伊莱·吉布森为这项工作做出了同样的贡献。豪尔赫·卡多佐和汤姆·韦考特伦对这项工作作出了同样的贡献。26 页, 6 个数字;更新包括其他应用程序、更新的作者列表和日志提交的格式

72. 第 1708. 09300[[pdf](#),[其他](#)] Cs。简历

用于脑物质分割的纹理与结构相结合的散网混合深度学习网络 (ts-shdl)

作者:[amarjot singh](#), [devamuyhazarika](#), [aniruddha bhattacharya](#)

文摘 由于灰色和白质区域之间的不规则边界, mr 图像中大脑物质分割的自动化是一项具有挑战性的任务。此外, mr 图像中强度不均匀的存在使问题更加复杂。在本文中,我们提出了一个纹理和容器结合版本的散网混合深度学习网络 (ts-shdl) 提取分层不变的中级特征,使用的渔民向量编码和条件随机字段 (crf) 来执行所需的分段。通过广泛的实验和与从合成麦吉尔脑网拍摄的几次二维 mri 扫描以及真正 3d 的 mrbrains 数据集的最先进的方法的比较,对拟议网络的性能进行了评估核磁共振扫描。除了与最先进的网络相比,ts-shdl 网络具有比先进的性能更高的优点,还介绍了 ts-shdl 网络相对于监督深度学习网络的优势。少

2017 年 8 月 30 日提交;最初宣布 2017 年 8 月。

评论:出席 ieee 计算机视觉研讨会 (iccvw) 2017 年国际会议

73. 第: 1707.04682[[pdf](#),[其他](#)] Cs。简历

反思重定向: 从单一图像中关闭具有感知形状重建的循环

作者:[瑞珠](#),[哈德·基亚尼·加卢加希](#),[王朝阳](#),[西蒙·卢西](#)

文摘:计算机视觉中出现的一个问题是从一张图像中重建物体的三维形状和姿态。到目前为止,这个问题已经通过应用规范来解决。更多

2017 年 7 月 26 日提交;v1 于 2017 年 7 月 14 日提交;最初宣布 2017 年 7 月。

评论:第一个子

74. 特别报告: 1707. 03742[[pdf](#),[其他](#)] Cs。Hc

大规模多视三维手持式数据集

作者:[francisco gomez-donooso](#), [sergio ors-escolano](#), [miguel Cazorla](#)

摘要: 关节级的准确手脸估计在人机交互、用户接口和虚拟现实应用中有多用途。然而,目前这并不是一个已解决的问题。新的深度学习技术可以在这个问题上有很大的改进,但他们需要大量的附加说明的数据。迄今为止发布的手持式数据集提出了一些问题,使其无法在深度学习方法上使用,例如数量很少的样本、高级抽象注释或由深度映射组成的样本。在本工作中,我们介绍了一个多视图的手摆数据集,在这个数据集中,我们为每一个人提供了手的彩色图像和不同类型的注释,即边界框以及手部关节上的 2d 和

3d 位置。此外,我们还介绍了一种简单而准确的深度学习架构,用于实时鲁棒的 **2d** 手部姿态估计。少

2017 年 7 月 18 日提交;v1 于 2017 年 7 月 12 日提交;最初宣布 2017 年 7 月。

75. 第 077.0. 03237[[pdf](#),[其他](#)] Cs。简历

[多伊](#) [10.1007/978-3-319-6558-9_28](#)

用于高度不平衡分段的深度学习损耗函数的广义 **dice** 重叠

作者:[carole h sudre](#), [wwqili](#), [tom Vercauteren](#), [sébastien ourselin](#), [m](#) 豪尔赫·卡多佐

摘要:近年来,深度学习已被证明是图像分析的有力工具,目前已被广泛用于对**二维**和**三维**医学图像进行分割。深度学习分割框架不仅依赖于网络体系结构的选择,还依赖于损失函数的选择。当分割过程针对罕见的观察时,候选标签之间很可能会出现严重的类不平衡,从而导致性能不理想。为了缓解这一问题,提出了加权交叉熵函数、灵敏度函数或 **dice** 损失函数等策略。在本工作中,我们研究了这些丢失函数的行为及其在 **2d** 和 **3d** 分割任务中存在不同标签不平衡率的情况下对学习速率调优的敏感性。我们还建议使用广义 **dice** 重叠的类重新平衡属性 (用于分段评估的已知度量),作为不平衡任务的可靠且准确的深度学习丢失函数。少

2017 年 7 月 14 日提交;v1 于 2017 年 7 月 11 日提交;最初宣布 2017 年 7 月。

76. 第 077.0. 02747[[pdf](#),[其他](#)] Cs。Lg

多种行为的稳健模仿

作者:[王子宇](#),[乔什·梅雷尔](#),[斯科特·里德](#), [格雷格·韦恩](#), [南多·德弗雷塔斯](#), [尼古拉斯·里斯](#)

摘要:深度生成模型最近在电机控制的模仿学习中显示出了巨大的希望。如果有足够的数据,即使是有监督的方法也可以进行一次性模仿学习;然而,当代理轨迹偏离演示时,它们很容易出现级联故障。与纯粹监督的方法相比,生成性抗冒名顶替学习 (gail) 可以从更少的演示中学习更强大的控制器,但本质上是寻找模式,更难以训练。在本文中,我们展示了如何将这两种方法的有利方面结合起来。我们模型的基础是一种新的基于学习语义策略嵌入的演示轨迹的变分自动编码器。我们表明,这些嵌入可以在 9 dof **jaco** 机器人手臂上学习,以完成任务,然后顺利插值,从而顺利插值到达行为。利用这些策略表示,我们开发了一个新版本的 gail,它 (1) 比纯粹监督的控制器更强大,特别是很少演示,(2) 避免模式崩溃,在 gail 本身捕捉许多不同的行为时不。我们展示了我们在 **mujoco** 物理环境中的 **2d** 双足和 62 dof **3d** 人形演示中学习不同步态的方法。少

2017 年 7 月 14 日提交;v1 于 2017 年 7 月 10 日提交;最初宣布 2017 年 7 月。

77. 第 [xiv:170.6.07036](#)[[pdf](#),[其他](#)] Cs。简历

学习高效点云生成,实现密集 **3d** 对象重建

作者:[林振轩](#), [陈光](#), [西蒙·卢西](#)

摘要:传统的**三维**物体生成建模方法使用具有**三维**卷积操作的深网络来学习体积预测,这与经典的**二维**网络直接类似。但是,这些方法在计算上是浪费的,试图预测 **3d** 形状,其中信息仅在曲面上丰富。本文提出了一种新的**三维**生成建模框架,以密集点云的形式有效地生成物体形状。我们使用**二维**卷积运算从多个角度对**三维**结构进行预测,并将几何推理与**二维**投影优化结合起来。我们引入了伪呈现器,这是一个可微分的模块,用于

近似真实的渲染运算,以合成新的深度图进行优化。单图像三维目标重构任务的实验结果表明,在形状相似度和预测密度方面,我们的性能优于最先进的方法。少

2017年6月21日提交;最初宣布2017年6月。

78. 第 [xiv:1706.03142](#)[pdf, ps, 其他] Cs. 简历

从非各向同性三维电子显微镜中深入学习各向同性超分辨率

作者:larissa heinrich, john a. bogovic, stepan saalfeld

文摘:从非各向同性电子显微镜(em)生成三维各向同性超分辨率(sr)的最复杂的现有方法是以学习词典为基础的。不幸的是,现有的方法都没有产生实际令人满意的结果。对于2d自然图像,最近开发的使用深度学习的超分辨率方法已经被证明明显优于以前的技术状态。我们已经将其中一个最成功的架构(fsmnnn)改编为3d超分辨率,并将其性能与以前未用于生成超分辨率的3d u-net 体系结构进行了比较。我们从聚焦离子束铣削扫描em(fib-sem)中对这两种体系结构进行了人为缩小各向同性地面真理的训练,并测试了各种超参数设置的性能。我们的研究表明,这两种体系结构都能成功地从非各向同性em中生成三维各向同性超分辨率,u-net 的性能持续较好。我们提出了几个有希望的实际应用方向。少

提交于2017年6月9日;最初宣布2017年6月。

79. 第 [xiv:1706.02042](#)[pdf, 其他] Cs. Gr

多伊 [10.114/30729299.3073629](#)

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。

评论:12页,16个数字,将出现在2017年号图

80. 第 [07:1705.10904](#)[pdf, 其他] Cs. 简历

弱监管的3d重建与对抗约束

作者:junyoung gwak, christopher b. choy, animesh garg, manmohan chandraker, silvio savarese

文摘:通过使用深层神经网络,在监督下的3d重建取得了重大进展。但是,这种性能的提高需要对2d/3d数据进行大规模注释。本文探讨了廉价的二维监控作为昂贵的三维cad注释的一种替代方法。具体来说,我们使用前景掩码作为弱监督通过光线跟踪池层,使透视投影和反向传播。此外,由于从掩码进行3d重建是一个不正确的问题,我

们建议将 3d 重建限制在与掩模观测相匹配的未标记的逼真 3d 形状的流形上。我们证明,学习此约束优化问题的逻辑屏障解决方案类似于 gan 目标,从而能够使用现有工具来训练 gans。我们对各种数据集中的流形约束重构进行了评价和分析,实现了合成图像和真实图像的单视图和多视图重建。少

2017 年 10 月 4 日提交;v1 于 2017 年 5 月 30 日提交;最初宣布 2017 年 5 月。

81. 第 1705.07750[[pdf](#),[其他](#)] Cs. 简历

quo vadis, 行动识别? 一种新的模型与动力学数据集

作者:[joao carreira](#), [Carreira zisserman](#)

摘要: 由于目前的行动分类数据集 (ucf-101 和 hmdb-51) 中缺乏视频, 因此很难确定良好的视频架构, 因为大多数方法在现有的小规模基准上获得了类似的性能。本文根据新的动力学人体行动视频数据集, 重新评估了最先进的体系结构。动力学有两个数量级以上的数据, 每班有 400 个人类动作类和 400 多个剪辑, 这些数据是从现实的、具有挑战性的 youtube 视频中收集的。我们分析了当前体系结构在此数据集上的操作分类任务的效果, 以及在对动力学进行预培训后, 在较小的基准数据集上提高了多少性能。我们还引入了一种新的双流充气 3d 凸网(i3d), 它基于 2d convnet 充气: 将非常深的图像分类的滤镜和池核扩展到 3d, 从而有可能从视频中学习无缝时空要素提取器, 同时利用成功的 imagenet 体系结构设计, 甚至其参数。我们表明, 经过动力学的预培训后, i3d 模型大大改进了行动分类方面的最先进技术, 在 hmdb-51 上达到 80.9, 在 ucf-101 上达到 98.0%。少

2018 年 2 月 12 日提交;v1 于 2017 年 5 月 22 日提交;最初宣布 2017 年 5 月。

评论:删除了对从未公开过的小型动力学数据集的引用, 并重复了整个动力学数据集上的所有实验

82. 第 1705.03428[[pdf](#),[其他](#)] Cs. 简历

深度投影三维语义分割

作者:[felix järema 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-nm), 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 在很大程度上受到附加注释数据集可用性的限制的影响。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。

评论:提交至 2017 年 caip

83. 第 07iv:170 5.002883[[pdf](#),[其他](#)] Cs. 简历

一种基于单一图像的三维人体姿态估计的双源方法

作者:umar iqbal, andraas doering, hashim yasin , björn krüger, Juergen weber ,
juergen gall

抽象: 在这项工作中, 我们解决了从单个图像的三维人体姿势估计的具有挑战性的问题。最近的方法学习深层神经网络来直接从图像中回归 **3d** 姿势。然而, 这种方法的一个主要挑战是收集培训数据。具体来说, 收集大量包含使用精确 **3d** 姿势注释的无约束图像的训练数据是不可行的。因此, 我们建议使用两个独立的培训来源。第一个源由精确的 **3d** 运动捕获数据组成, 第二个源由带有注释**二维姿态**的无约束图像组成。为了集成这两个来源, 我们提出了一种双源方法, 将**二维姿态**估计与高效的**三维姿态**检索相结合。为此, 我们首先将运动捕获数据转换为归一化的**二维姿态**空间, 并分别从图像数据中学习**二维姿态**估计模型。在推理过程中, 我们估计**二维姿态**, 并有效地检索最近的**三维姿态**。然后, 我们共同估计从**三维姿势**空间到图像的映射, 并重建**三维姿势**。我们对所提出的方法进行了全面的评估, 并通过实验证明了我们的方法的有效性, 即使这两个来源的骨架结构有很大的不同。少

2017 年 9 月 6 日提交;v1 于 2017 年 5 月 8 日提交;最初宣布 2017 年 5 月。

评论:在计算机视觉和图像理解中进行考虑。cvpr-2016 论文扩展版, 第 xiv:1509.0773

84. 第 xiv:17004.0754[[pdf](#),[其他](#)] Cs. 简历

三生物医学分割的联合序列学习与交叉模态卷积

作者:曾冠伦,林延良, 许文森,黄忠阳

摘要: 深部卷积神经网络等学习模型在**三维生物医学分割**中得到了广泛的应用, 并取得了最先进的性能。然而, 他们中的大多数人往往采用单一模式或将多种模式堆叠为不同的输入通道。为了更好地利用多模, 我们提出了一种具有跨模卷积层的**深层编码解码器**结构, 以纳入不同的 mri 数据模式。此外, 我们利用卷积 lstm 对**二维切片**序列进行建模, 并以端到端的方式**共同学习**多模和卷积 lstm。为了避免与某些标签融合, 我们采用了重新加权方案和两阶段培训来处理标签不平衡。brats-2015 的实验结果表明, 该方法优于最先进的生物医学分割方法。少

Submitted 25 April, 2017; originally announced April 2017.

Comments: CVPR 2017

85. [arXiv:1704.02447](#) [[pdf](#), [other](#)] cs.CV

Towards 3D Human Pose Estimation in the Wild: a Weakly-supervised Approach

Authors: Xingyi Zhou, Qixing Huang, Xiao Sun, Xiangyang Xue, Yichen Wei

Abstract: In this paper, we study the task of **3D**human pose estimation in the wild. This task is challenging due to lack of training data, as existing datasets are either in the wild images with **2D** pose or in the lab images with **3D** pose. We propose a weakly-supervised transfer **learning** method that uses mixed **2D** and **3D** labels in a unified **deep**neural network that presents two-stage cascaded structure. Our network augments a state-of-the-art **2D** pose estimation sub-network with a **3D** depth regression sub-network. Unlike previous two stage approaches that train the two sub-networks sequentially and separately, our training is end-to-end and fully exploits the correlation between the **2D** pose and depth estimation sub-tasks. The **deep**features are better learnt through shared representations. In doing so, the **3D** pose labels in controlled lab environments are transferred to in the

wild images. In addition, we introduce a **3D** geometric constraint to regularize the **3D** pose prediction, which is effective in the absence of ground truth depth labels. Our method achieves competitive results on both **2D** and **3D** benchmarks. \triangle Less

Submitted 30 July, 2017; v1 submitted 8 April, 2017; originally announced April 2017.

Comments: Accepted to ICCV 2017

86. [arXiv:1704.01880](#) [[pdf](#), [other](#)] [cs.CV](#)

A Convolution Tree with Deconvolution Branches: Exploiting Geometric Relationships for Single Shot Keypoint Detection

Authors: [Amit Kumar](#), [Rama Chellappa](#)

Abstract: Recently, **Deep** Convolution Networks (DCNNs) have been applied to the task of face alignment and have shown potential for **learning** improved feature representations. Although deeper layers can capture abstract concepts like pose, it is difficult to capture the geometric relationships among the keypoints in DCNNs. In this paper, we propose a novel convolution-deconvolution network for facial keypoint detection. Our model predicts the **2D** locations of the keypoints and their individual visibility along with **3D** head pose, while exploiting the spatial relationships among different keypoints. Different from existing approaches of modeling these relationships, we propose learnable transform functions which captures the relationships between keypoints at feature level. However, due to extensive variations in pose, not all of these relationships act at once, and hence we propose, a pose-based routing function which implicitly models the active relationships. Both transform functions and the routing function are implemented through convolutions in a multi-task framework. Our approach presents a single-shot keypoint detection method, making it different from many existing cascade regression-based methods. We also show that **learning** these relationships significantly improve the accuracy of keypoint detections for in-the-wild face images from challenging datasets such as AFW and AFLW. \triangle Less

Submitted 6 April, 2017; originally announced April 2017.

87. [arXiv:1703.09831](#) [[pdf](#), [other](#)] [cs.CL](#)

A Deep Compositional Framework for Human-like Language Acquisition in Virtual Environment

Authors: [Haonan Yu](#), [Haichao Zhang](#), [Wei Xu](#)

Abstract: We tackle a task where an agent **learns** to navigate in a **2D** maze-like environment called XWORLD. In each session, the agent perceives a sequence of raw-pixel frames, a natural language command issued by a teacher, and a set of rewards. The agent **learns** the teacher's language from scratch in a grounded and compositional manner, such that after training it is able to correctly execute zero-shot commands: 1) the combination of words in the command never appeared before, and/or 2) the command contains new object concepts that are **learned** from another task but never **learned** from navigation. Our **deep** framework for the agent is trained end to end: it **learns** simultaneously the visual representations of the

environment, the syntax and semantics of the language, and the action module that outputs actions. The zero-shot **learning** capability of our framework results from its compositionality and modularity with parameter tying. We visualize the intermediate outputs of the framework, demonstrating that the agent truly understands how to solve the problem. We believe that our results provide some preliminary insights on how to train an agent with similar abilities in a **3D** environment. △ Less

Submitted 19 May, 2017; v1 submitted 28 March, 2017; originally announced March 2017.

88. [arXiv:1703.08025](#) [[pdf](#), [ps](#), [other](#)] [cs.CV](#)

Saliency-guided video classification via adaptively weighted learning

Authors: [Yunzhen Zhao](#), [Yuxin Peng](#)

Abstract: Video classification is productive in many practical applications, and the recent **deep learning** has greatly improved its accuracy. However, existing works often model video frames indiscriminately, but from the view of motion, video frames can be decomposed into salient and non-salient areas naturally. Salient and non-salient areas should be modeled with different networks, for the former present both appearance and motion information, and the latter present static background information. To address this problem, in this paper, video saliency is predicted by optical flow without supervision firstly. Then two streams of **3D** CNN are trained individually for raw frames and optical flow on salient areas, and another **2D** CNN is trained for raw frames on non-salient areas. For the reason that these three streams play different roles for each class, the weights of each stream are adaptively **learned** for each class. Experimental results show that saliency-guided modeling and adaptively weighted **learning** can reinforce each other, and we achieve the state-of-the-art results. △ Less

Submitted 25 March, 2017; v1 submitted 23 March, 2017; originally announced March 2017.

Comments: 6 pages, 1 figure, accepted by ICME 2017

89. [arXiv:1703.07475](#) [[pdf](#), [other](#)] [cs.CV](#)

PKU-MMD: A Large Scale Benchmark for Continuous Multi-Modal Human Action Understanding

Authors: [Chunhui Liu](#), [Yueyu Hu](#), [Yanghao Li](#), [Sijie Song](#), [Jiaying Liu](#)

Abstract: Despite the fact that many **3D** human activity benchmarks being proposed, most existing action datasets focus on the action recognition tasks for the segmented videos. There is a lack of standard large-scale benchmarks, especially for current popular data-hungry **deep learning** based methods. In this paper, we introduce a new large scale benchmark (PKU-MMD) for continuous multi-modality **3D** human action understanding and cover a wide range of complex human activities with well annotated information. PKU-MMD contains 1076 long video sequences in 51 action categories, performed by 66 subjects in three camera views. It contains almost 20,000 action instances and 5.4 million frames in total. Our dataset also provides multi-modality data sources, including RGB, depth, Infrared

Radiation and Skeleton. With different modalities, we conduct extensive experiments on our dataset in terms of two scenarios and evaluate different methods by various metrics, including a new proposed evaluation protocol **2D-AP**. We believe this large-scale dataset will benefit future researches on action detection for the community. \triangle Less

Submitted 27 March, 2017; v1 submitted 21 March, 2017; originally announced March 2017.

Comments: 10 pages

90. [arXiv:1703.04699](#) [[pdf](#), [other](#)] [cs.CV](#)

[doi:10.1109/ICAR.2017.8023499](#)

A fully end-to-end deep learning approach for real-time simultaneous 3D reconstruction and material recognition

Authors: [Cheng Zhao](#), [Li Sun](#), [Rustam Stolkin](#)

Abstract: This paper addresses the problem of simultaneous **3D** reconstruction and material recognition and segmentation. Enabling robots to recognise different materials (concrete, metal etc.) in a scene is important for many tasks, e.g. robotic interventions in nuclear decommissioning. Previous work on **3D** semantic reconstruction has predominantly focused on recognition of everyday domestic objects (tables, chairs etc.), whereas previous work on material recognition has largely been confined to single **2D** images without any **3D** reconstruction. Meanwhile, most **3D** semantic reconstruction methods rely on computationally expensive post-processing, using Fully-Connected Conditional Random Fields (CRFs), to achieve consistent segmentations. In contrast, we propose a **deep learning** method which performs **3D** reconstruction while simultaneously recognising different types of materials and labelling them at the pixel level. Unlike previous methods, we propose a fully end-to-end approach, which does not require hand-crafted features or CRF post-processing. Instead, we use only **learned** features, and the CRF segmentation constraints are incorporated inside the fully end-to-end **learned** system. We present the results of experiments, in which we trained our system to perform real-time **3D** semantic reconstruction for 23 different materials in a real-world application. The run-time performance of the system can be boosted to around 10Hz, using a conventional GPU, which is enough to achieve real-time semantic reconstruction using a 30fps RGB-D camera. To the best of our knowledge, this work is the first real-time end-to-end system for simultaneous **3D** reconstruction and material recognition. \triangle Less

Submitted 14 March, 2017; originally announced March 2017.

Comments: 8 pages, 7 figures, 4 tables

91. [arXiv:1612.05872](#) [[pdf](#), [other](#)] [cs.CV](#)

3D Shape Induction from 2D Views of Multiple Objects

Authors: [Matheus Gadelha](#), [Subhransu Maji](#), [Rui Wang](#)

Abstract: In this paper we investigate the problem of inducing a distribution over three-dimensional structures given two-dimensional views of multiple objects taken

from unknown viewpoints. Our approach called "projective generative adversarial networks" (PrGANs) trains a **deep** generative model of **3D** shapes whose projections match the distributions of the input **2D** views. The addition of a projection module allows us to infer the underlying **3D** shape distribution without using any **3D**, viewpoint information, or annotation during the **learning** phase. We show that our approach produces **3D** shapes of comparable quality to GANs trained on **3D** data for a number of shape categories including chairs, airplanes, and cars. Experiments also show that the disentangled representation of **2D** shapes into geometry and viewpoint leads to a good generative model of **2D** shapes. The key advantage is that our model allows us to predict **3D**, viewpoint, and generate novel views from an input image in a completely unsupervised manner. △ Less

Submitted 18 December, 2016; originally announced December 2016.

Comments: Submitted to CVPR 2017

92. [arXiv:1612.00380](#) [[pdf](#), [other](#)] [cs.AI](#)

Playing Doom with SLAM-Augmented Deep Reinforcement Learning

Authors: [Shehroze Bhatti](#), [Alban Desmaison](#), [Ondrej Miksik](#), [Nantas Nardelli](#), [N. Siddharth](#), [Philip H. S. Torr](#)

Abstract: A number of recent approaches to policy **learning** in **2D** game domains have been successful going directly from raw input images to actions. However when employed in complex **3D** environments, they typically suffer from challenges related to partial observability, combinatorial exploration spaces, path planning, and a scarcity of rewarding scenarios. Inspired from prior work in human cognition that indicates how humans employ a variety of semantic concepts and abstractions (object categories, localisation, etc.) to reason about the world, we build an agent-model that incorporates such abstractions into its policy-**learning** framework. We augment the raw image input to a **Deep Q-Learning** Network (DQN), by adding details of objects and structural elements encountered, along with the agent's localisation. The different components are automatically extracted and composed into a topological representation using on-the-fly object detection and **3D**-scene reconstruction. We evaluate the efficacy of our approach in Doom, a **3D** first-person combat game that exhibits a number of challenges discussed, and show that our augmented framework consistently **learns** better, more effective policies. △ Less

Submitted 1 December, 2016; originally announced December 2016.

93. [arXiv:1611.10010](#) [[pdf](#), [other](#)] [cs.CV](#)

Deep Cuboid Detection: Beyond 2D Bounding Boxes

Authors: [Debidatta Dwibedi](#), [Tomasz Malisiewicz](#), [Vijay Badrinarayanan](#), [Andrew Rabinovich](#)

Abstract: We present a **Deep** Cuboid Detector which takes a consumer-quality RGB image of a cluttered scene and localizes all **3D** cuboids (box-like objects). Contrary to classical approaches which fit a **3D** model from low-level cues like corners, edges, and vanishing points, we propose an end-to-end **deeplearning** system to detect cuboids across many semantic

categories (e.g., ovens, shipping boxes, and furniture). We localize cuboids with a **2D** bounding box, and simultaneously localize the cuboid's corners, effectively producing a **3D** interpretation of box-like objects. We refine keypoints by pooling convolutional features iteratively, improving the baseline method significantly. Our **deep learning** cuboid detector is trained in an end-to-end fashion and is suitable for real-time applications in augmented reality (AR) and robotics. \triangle Less
Submitted 30 November, 2016; originally announced November 2016.

94. [arXiv:1611.05708](#) [[pdf](#), [other](#)] [cs.CV](#)

Learning to Fuse 2D and 3D Image Cues for Monocular Body Pose Estimation

Authors: [Bugra Tekin](#), [Pablo Márquez-Neila](#), [Mathieu Salzmann](#), [Pascal Fua](#)

Abstract: Most recent approaches to monocular **3D** human pose estimation rely on **Deep Learning**. They typically involve regressing from an image to either **3D** joint coordinates directly or **2D** joint locations from which **3D** coordinates are inferred. Both approaches have their strengths and weaknesses and we therefore propose a novel architecture designed to deliver the best of both worlds by performing both simultaneously and fusing the information along the way. At the heart of our framework is a trainable fusion scheme that **learns** how to fuse the information optimally instead of being hand-designed. This yields significant improvements upon the state-of-the-art on standard **3D** human pose estimation benchmarks. \triangle Less

Submitted 10 April, 2017; v1 submitted 17 November, 2016; originally announced November 2016.

95. [arXiv:1611.04144](#) [[pdf](#), [other](#)] [cs.CV](#)

Semi-Dense 3D Semantic Mapping from Monocular SLAM

Authors: [Xuanpeng Li](#), [Rachid Belaroussi](#)

Abstract: The bundle of geometry and appearance in computer vision has proven to be a promising solution for robots across a wide variety of applications. Stereo cameras and RGB-D sensors are widely used to realise fast **3D** reconstruction and trajectory tracking in a dense way. However, they lack flexibility of seamless switch between different scaled environments, i.e., indoor and outdoor scenes. In addition, semantic information are still hard to acquire in a **3D** mapping. We address this challenge by combining the state-of-art **deeplearning** method and semi-dense Simultaneous Localisation and Mapping (SLAM) based on video stream from a monocular camera. In our approach, **2D** semantic information are transferred to **3D** mapping via correspondence between connective Keyframes with spatial consistency. There is no need to obtain a semantic segmentation for each frame in a sequence, so that it could achieve a reasonable computation time. We evaluate our method on indoor/outdoor datasets and lead to an improvement in the **2D** semantic labelling over baseline single frame predictions. \triangle Less

Submitted 13 November, 2016; originally announced November 2016.

96. [arXiv:1610.09157](#) [[pdf](#), [other](#)] [cs.CV](#)

[doi:10.1038/srep46479](#)

Towards automatic pulmonary nodule management in lung cancer screening with deep learning

Authors: [Francesco Ciompi](#), [Kaman Chung](#), [Sarah J. van Riel](#), [Arnaud Arindra Adiyoso Setio](#), [Paul K. Gerke](#), [Colin Jacobs](#), [Ernst Th. Scholten](#), [Cornelia Schaefer-Prokop](#), [Mathilde M. W. Wille](#), [Alfonso Marchiano](#), [Ugo Pastorino](#), [Mathias Prokop](#), [Bram van Ginneken](#)

Abstract: The introduction of lung cancer screening programs will produce an unprecedented amount of chest CT scans in the near future, which radiologists will have to read in order to decide on a patient follow-up strategy. According to the current guidelines, the workup of screen-detected nodules strongly relies on nodule size and nodule type. In this paper, we present a **deep learning** system based on multi-stream multi-scale convolutional networks, which automatically classifies all nodule types relevant for nodule workup. The system processes raw CT data containing a nodule without the need for any additional information such as nodule segmentation or nodule size and **learns** a representation of **3D** data by analyzing an arbitrary number of **2D** views of a given nodule. The **deeplearning** system was trained with data from the Italian MILD screening trial and validated on an independent set of data from the Danish DLCST screening trial. We analyze the advantage of processing nodules at multiple scales with a multi-stream convolutional network architecture, and we show that the proposed **deep learning** system achieves performance at classifying nodule type that surpasses the one of classical machine **learning** approaches and is within the inter-observer variability among four experienced human observers. △ Less

Submitted 23 May, 2017; v1 submitted 28 October, 2016; originally announced October 2016.

Comments: Published on Scientific Reports

Journal ref: Sci. Rep. 7, 46479; (2017)

97. [arXiv:1609.09545](#) [[pdf](#), [other](#)] [cs.CV](#)

[doi](#)[10.1007/978-3-319-48881-3_43](#)

Two-stage Convolutional Part Heatmap Regression for the 1st 3D Face Alignment in the Wild (3DFAW) Challenge

Authors: [Adrian Bulat](#), [Georgios Tzimiropoulos](#)

Abstract: This paper describes our submission to the 1st **3D** Face Alignment in the Wild (3DFAW) Challenge. Our method builds upon the idea of convolutional part heatmap regression [1], extending it for **3D** face alignment. Our method decomposes the problem into two parts: (a) X,Y (**2D**) estimation and (b) Z (depth) estimation. At the first stage, our method estimates the X,Y coordinates of the facial landmarks by producing a set of **2D** heatmaps, one for each landmark, using convolutional part heatmap regression. Then, these heatmaps, alongside the input RGB image, are used as input to a very **deep** subnetwork trained via residual **learning** for regressing the Z coordinate. Our method ranked 1st in the 3DFAW Challenge, surpassing the second best result by more than 22%. △ Less

Submitted 29 September, 2016; originally announced September 2016.

Comments: Winner of 3D Face Alignment in the Wild (3DFAW) Challenge, ECCV 2016

98. [arXiv:1609.05521](#) [[pdf](#), [other](#)] [cs.AI](#)

Playing FPS Games with Deep Reinforcement Learning

Authors: [Guillaume Lample](#), [Devendra Singh Chaplot](#)

Abstract: Advances in **deep** reinforcement **learning** have allowed autonomous agents to perform well on Atari games, often outperforming humans, using only raw pixels to make their decisions. However, most of these games take place in **2D** environments that are fully observable to the agent. In this paper, we present the first architecture to tackle **3D** environments in first-person shooter games, that involve partially observable states. Typically, **deep** reinforcement **learning** methods only utilize visual input for training. We present a method to augment these models to exploit game feature information such as the presence of enemies or items, during the training phase. Our model is trained to simultaneously **learn** these features along with minimizing a Q-**learning** objective, which is shown to dramatically improve the training speed and performance of our agent. Our architecture is also modularized to allow different models to be independently trained for different phases of the game. We show that the proposed architecture substantially outperforms built-in AI agents of the game as well as humans in deathmatch scenarios. \triangle Less

Submitted 29 January, 2018; v1 submitted 18 September, 2016; originally announced September 2016.

Comments: The authors contributed equally to this work

99. [arXiv:1608.05895](#) [[pdf](#), [other](#)] [cs.CV](#)

VoxResNet: Deep Voxelwise Residual Networks for Volumetric Brain Segmentation

Authors: [Hao Chen](#), [Qi Dou](#), [Lequan Yu](#), [Pheng-Ann Heng](#)

Abstract: Recently **deep** residual **learning** with residual units for training very **deep** neural networks advanced the state-of-the-art performance on **2D** image recognition tasks, e.g., object detection and segmentation. However, how to fully leverage contextual representations for recognition tasks from volumetric data has not been well studied, especially in the field of medical image computing, where a majority of image modalities are in volumetric format. In this paper we explore the **deep** residual **learning** on the task of volumetric brain segmentation. There are at least two main contributions in our work. First, we propose a **deep** voxelwise residual network, referred as VoxResNet, which borrows the spirit of **deep** residual **learning** in **2D** image recognition tasks, and is extended into a **3D** variant for handling volumetric data. Second, an auto-context version of VoxResNet is proposed by seamlessly integrating the low-level image appearance features, implicit shape information and high-level context together for further improving the volumetric segmentation performance. Extensive experiments on the

challenging benchmark of brain segmentation from magnetic resonance (MR) images corroborated the efficacy of our proposed method in dealing with volumetric data. We believe this work unravels the potential of **3D deeplearning** to advance the recognition performance on volumetric image segmentation. \triangle Less

Submitted 21 August, 2016; originally announced August 2016.

100. [arXiv:1607.03597](#) [[pdf](#), [other](#)] [cs.CV](#)

Accelerating Eulerian Fluid Simulation With Convolutional Networks

Authors: [Jonathan Tompson](#), [Kristofer Schlachter](#), [Pablo Sprechmann](#), [Ken Perlin](#)

Abstract: Efficient simulation of the Navier-Stokes equations for fluid flow is a long standing problem in applied mathematics, for which state-of-the-art methods require large compute resources. In this work, we propose a data-driven approach that leverages the approximation power of **deep-learning** with the precision of standard solvers to obtain fast and highly realistic simulations. Our method solves the incompressible Euler equations using the standard operator splitting method, in which a large sparse linear system with many free parameters must be solved. We use a Convolutional Network with a highly tailored architecture, trained using a novel unsupervised **learning** framework to solve the linear system. We present real-time **2D** and **3D** simulations that outperform recently proposed data-driven methods; the obtained results are realistic and show good generalization properties. \triangle Less

Submitted 22 June, 2017; v1 submitted 13 July, 2016; originally announced July 2016.

Comments: Significant revision

101. [arXiv:1607.02504](#) [[pdf](#), [other](#)] [cs.CV](#)

Fast Predictive Image Registration

Authors: [Xiao Yang](#), [Roland Kwitt](#), [Marc Niethammer](#)

Abstract: We present a method to predict image deformations based on patch-wise image appearance. Specifically, we design a patch-based **deep** encoder-decoder network which **learns** the pixel/voxel-wise mapping between image appearance and registration parameters. Our approach can predict general deformation parameterizations, however, we focus on the large deformation diffeomorphic metric mapping (LDDMM) registration model. By predicting the LDDMM momentum-parameterization we retain the desirable theoretical properties of LDDMM, while reducing computation time by orders of magnitude: combined with patch pruning, we achieve a 1500x/66x speed up compared to GPU-based optimization for **2D/3D** image registration. Our approach has better prediction accuracy than predicting deformation or velocity fields and results in diffeomorphic transformations. Additionally, we create a Bayesian probabilistic version of our network, which allows evaluation of deformation field uncertainty through Monte Carlo sampling using dropout at test time. We show that deformation uncertainty highlights areas of ambiguous deformations. We test our method on the OASIS brain image dataset in **2D** and **3D**. \triangle Less

Submitted 8 July, 2016; originally announced July 2016.

102. [arXiv:1607.00730](#) [[pdf](#), [other](#)] [cs.CV](#)

A Two-Streamed Network for Estimating Fine-Scaled Depth Maps from Single RGB Images

Authors: [Jun Li](#), [Reinhard Klein](#), [Angela Yao](#)

Abstract: Estimating depth from a single RGB image is an ill-posed and inherently ambiguous problem. State-of-the-art **deep learning** methods can now estimate accurate **2D** depth maps, but when the maps are projected into **3D**, they lack local detail and are often highly distorted. We propose a fast-to-train two-streamed CNN that predicts depth and depth gradients, which are then fused together into an accurate and detailed depth map. We also define a novel set loss over multiple images; by regularizing the estimation between a common set of images, the network is less prone to over-fitting and achieves better accuracy than competing methods. Experiments on the NYU Depth v2 dataset shows that our depth predictions are competitive with state-of-the-art and lead to faithful **3D** projections. \triangle Less

Submitted 3 December, 2017; v1 submitted 3 July, 2016; originally announced July 2016.

103. [arXiv:1607.00662](#) [[pdf](#), [other](#)] [cs.CV](#)

Unsupervised Learning of 3D Structure from Images

Authors: [Danilo Jimenez Rezende](#), [S. M. Ali Eslami](#), [Shakir Mohamed](#), [Peter Battaglia](#), [Max Jaderberg](#), [Nicolas Heess](#)

Abstract: A key goal of computer vision is to recover the underlying **3D** structure from **2D** observations of the world. In this paper we **learn** strong **deep** generative models of **3D** structures, and recover these structures from **3D** and **2D** images via probabilistic inference. We demonstrate high-quality samples and report log-likelihoods on several datasets, including ShapeNet [2], and establish the first benchmarks in the literature. We also show how these models and their inference networks can be trained end-to-end from **2D** images. This demonstrates for the first time the feasibility of **learning** to infer **3D** representations of the world in a purely unsupervised manner. \triangle Less

Submitted 19 June, 2018; v1 submitted 3 July, 2016; originally announced July 2016.

Comments: Appears in Advances in Neural Information Processing Systems 29 (NIPS 2016)

104. [arXiv:1606.02382](#) [[pdf](#), [other](#)] [cs.CV](#)

Deep Learning Convolutional Networks for Multiphoton Microscopy Vasculature Segmentation

Authors: [Petteri Teikari](#), [Marc Santos](#), [Charissa Poon](#), [Kullervo Hynynen](#)

Abstract: Recently there has been an increasing trend to use **deep learning** frameworks for both **2D** consumer images and for **3D** medical images. However, there has been little effort to use **deep** frameworks for volumetric

vascular segmentation. We wanted to address this by providing a freely available dataset of 12 annotated two-photon vasculature microscopy stacks. We demonstrated the use of **deep learning** framework consisting both **2D** and **3D** convolutional filters (ConvNet). Our hybrid **2D-3D** architecture produced promising segmentation result. We derived the architectures from Lee et al. who used the ZNN framework initially designed for electron microscope image segmentation. We hope that by sharing our volumetric vasculature datasets, we will inspire other researchers to experiment with vasculature dataset and improve the used network architectures. \triangle Less

Submitted 7 June, 2016; originally announced June 2016.

Comments: 23 pages, 10 figures

ACM Class: I.2.6; I.5.1; I.5.4; I.4.6

105. [arXiv:1604.05865](#) [[pdf](#), [other](#)] [cs.CV](#)

[doi:10.1016/j.patcog.2017.04.017](#)

Estimating 3D Trajectories from 2D Projections via Disjunctive Factored Four-Way Conditional Restricted Boltzmann Machines

Authors: [Decebal Constantin Mocanu](#), [Haitham Bou Ammar](#), [Luis Puig](#), [Eric Eaton](#), [Antonio Liotta](#)

Abstract: Estimation, recognition, and near-future prediction of **3D** trajectories based on their two dimensional projections available from one camera source is an exceptionally difficult problem due to uncertainty in the trajectories and environment, high dimensionality of the specific trajectory states, lack of enough labeled data and so on. In this article, we propose a solution to solve this problem based on a novel **deep learning** model dubbed Disjunctive Factored Four-Way Conditional Restricted Boltzmann Machine (DFFW-CRBM). Our method improves state-of-the-art **deep learning** techniques for high dimensional time-series modeling by introducing a novel tensor factorization capable of driving forth order Boltzmann machines to considerably lower energy levels, at no computational costs. DFFW-CRBMs are capable of accurately estimating, recognizing, and performing near-future prediction of three-dimensional trajectories from their **2D** projections while requiring limited amount of labeled data. We evaluate our method on both simulated and real-world data, showing its effectiveness in predicting and classifying complex ball trajectories and human activities. \triangle Less

Submitted 29 April, 2017; v1 submitted 20 April, 2016; originally announced April 2016.

Comments: Pattern Recognition, ISSN [0031-3203](#), Elsevier, 2017

106. [arXiv:1603.03958](#) [[pdf](#), [other](#)] [cs.CV](#)

Template Adaptation for Face Verification and Identification

Authors: [Nate Crosswhite](#), [Jeffrey Byrne](#), [Omkar M. Parkhi](#), [Chris Stauffer](#), [Qiong Cao](#), [Andrew Zisserman](#)

Abstract: Face recognition performance evaluation has traditionally focused on one-to-one verification, popularized by the Labeled Faces in the Wild dataset for

imagery and the YouTubeFaces dataset for videos. In contrast, the newly released IJB-A face recognition dataset unifies evaluation of one-to-many face identification with one-to-one face verification over templates, or sets of imagery and videos for a subject. In this paper, we study the problem of template adaptation, a form of transfer **learning** to the set of media in a template. Extensive performance evaluations on IJB-A show a surprising result, that perhaps the simplest method of template adaptation, combining **deep** convolutional network features with template specific linear SVMs, outperforms the state-of-the-art by a wide margin. We study the effects of template size, negative set construction and classifier fusion on performance, then compare template adaptation to convolutional networks with metric **learning**, **2D** and **3D** alignment. Our unexpected conclusion is that these other methods, when combined with template adaptation, all achieve nearly the same top performance on IJB-A for template-based face verification and identification. \triangle Less

Submitted 5 April, 2016; v1 submitted 12 March, 2016; originally announced March 2016.

107. [arXiv:1511.02300](#) [[pdf](#), [other](#)] [cs.CV](#)

Deep Sliding Shapes for Amodal 3D Object Detection in RGB-D Images

Authors: [Shuran Song](#), [Jianxiong Xiao](#)

Abstract: We focus on the task of amodal **3D** object detection in RGB-D images, which aims to produce a **3D** bounding box of an object in metric form at its full extent. We introduce **Deep Sliding Shapes**, a **3D**ConvNet formulation that takes a **3D** volumetric scene from a RGB-D image as input and outputs **3D** object bounding boxes. In our approach, we propose the first **3D** Region Proposal Network (RPN) to **learn** objectness from geometric shapes and the first joint Object Recognition Network (ORN) to extract geometric features in **3D** and color features in **2D**. In particular, we handle objects of various sizes by training an amodal RPN at two different scales and an ORN to regress **3D** bounding boxes. Experiments show that our algorithm outperforms the state-of-the-art by 13.8 in mAP and is 200x faster than the original Sliding Shapes. All source code and pre-trained models will be available at GitHub. \triangle Less

Submitted 9 March, 2016; v1 submitted 6 November, 2015; originally announced November 2015.

108. [arXiv:1506.06448](#) [[pdf](#), [other](#)] [cs.CV](#)

DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation

Authors: [Holger R. Roth](#), [Le Lu](#), [Amal Farag](#), [Hoo-Chang Shin](#), [Jiamin Liu](#), [Evrin Turkbey](#), [Ronald M. Summers](#)

Abstract: Automatic organ segmentation is an important yet challenging problem for medical image analysis. The pancreas is an abdominal organ with very high anatomical variability. This inhibits previous segmentation methods from achieving high accuracies, especially compared to other organs such as the liver, heart or

kidneys. In this paper, we present a probabilistic bottom-up approach for pancreas segmentation in abdominal computed tomography (CT) scans, using multi-level **deep** convolutional networks (ConvNets). We propose and evaluate several variations of **deep**ConvNets in the context of hierarchical, coarse-to-fine classification on image patches and regions, i.e. superpixels. We first present a dense labeling of local image patches via P—ConvNet and nearest neighbor fusion. Then we describe a regional ConvNet (R1—ConvNet) that samples a set of bounding boxes around each image superpixel at different scales of contexts in a "zoom-out" fashion. Our ConvNets **learn** to assign class probabilities for each superpixel region of being pancreas. Last, we study a stacked R2—ConvNet leveraging the joint space of CT intensities and the P—ConvNet dense probability maps. Both **3D** Gaussian smoothing and **2D** conditional random fields are exploited as structured predictions for post-processing. We evaluate on CT images of 82 patients in 4-fold cross-validation. We achieve a Dice Similarity Coefficient of $83.6 \pm 6.3\%$ in training and $71.8 \pm 10.7\%$ in testing. \triangle Less

Submitted 21 June, 2015; originally announced June 2015.

Comments: To be presented at MICCAI 2015 - 18th International Conference on Medical Computing and Computer Assisted Interventions, Munich, Germany

109. [arXiv:1505.05641](#) [[pdf](#), [other](#)] **cs.CV**

Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views

Authors: [Hao Su](#), [Charles R. Qi](#), [Yangyan Li](#), [Leonidas Guibas](#)

Abstract: Object viewpoint estimation from **2D** images is an essential task in computer vision. However, two issues hinder its progress: scarcity of training data with viewpoint annotations, and a lack of powerful features. Inspired by the growing availability of **3D** models, we propose a framework to address both issues by combining render-based image synthesis and CNNs. We believe that **3D** models have the potential in generating a large number of images of high variation, which can be well exploited by **deep** CNN with a high **learning** capacity. Towards this goal, we propose a scalable and overfit-resistant image synthesis pipeline, together with a novel CNN specifically tailored for the viewpoint estimation task. Experimentally, we show that the viewpoint estimation from our pipeline can significantly outperform state-of-the-art methods on PASCAL **3D+** benchmark. \triangle Less

Submitted 21 May, 2015; originally announced May 2015.

110. [arXiv:1505.02000](#) [[pdf](#), [other](#)] **cs.LG**

Deep Learning for Medical Image Segmentation

Authors: [Matthew Lai](#)

Abstract: This report provides an overview of the current state of the art **deep learning** architectures and optimisation techniques, and uses the ADNI hippocampus MRI dataset as an example to compare the effectiveness and efficiency of different convolutional architectures on the task of patch-based

3-dimensional hippocampal segmentation, which is important in the diagnosis of Alzheimer's Disease. We found that a slightly unconventional "stacked **2D**" approach provides much better classification performance than simple **2D** patches without requiring significantly more computational power. We also examined the popular "tri-planar" approach used in some recently published studies, and found that it provides much better results than the **2D** approaches, but also with a moderate increase in computational power requirement. Finally, we evaluated a full **3D** convolutional architecture, and found that it provides marginally better results than the tri-planar approach, but at the cost of a very significant increase in computational power requirement. △ Less

Submitted 8 May, 2015; originally announced May 2015.

111. [arXiv:1412.0767](#) [[pdf](#), [other](#)] [cs.CV](#)

Learning Spatiotemporal Features with 3D Convolutional Networks

Authors: [Du Tran](#), [Lubomir Bourdev](#), [Rob Fergus](#), [Lorenzo Torresani](#), [Manohar Paluri](#)

Abstract: We propose a simple, yet effective approach for spatiotemporal feature **learning** using **deep** 3-dimensional convolutional networks (**3D**ConvNets) trained on a large scale supervised video dataset. Our findings are three-fold: 1) **3D** ConvNets are more suitable for spatiotemporal feature **learning** compared to **2D** ConvNets; 2) A homogeneous architecture with small 3x3x3 convolution kernels in all layers is among the best performing architectures for **3D** ConvNets; and 3) Our **learned** features, namely C3D (Convolutional **3D**), with a simple linear classifier outperform state-of-the-art methods on 4 different benchmarks and are comparable with current best methods on the other 2 benchmarks. In addition, the features are compact: achieving 52.8% accuracy on UCF101 dataset with only 10 dimensions and also very efficient to compute due to the fast inference of ConvNets. Finally, they are conceptually very simple and easy to train and use. △ Less

Submitted 6 October, 2015; v1 submitted 1 December, 2014; originally announced December 2014.

112. [arXiv:1409.7164](#) [[pdf](#), [other](#)] [cs.CV](#)

Deep Learning Representation using Autoencoder for 3D Shape Retrieval

Authors: [Zhuotun Zhu](#), [Xinggang Wang](#), [Song Bai](#), [Cong Yao](#), [Xiang Bai](#)

Abstract: We study the problem of how to build a **deep learning** representation for **3D** shape. **Deeplearning** has shown to be very effective in variety of visual applications, such as image classification and object detection. However, it has not been successfully applied to **3D** shape recognition. This is because **3D** shape has complex structure in **3D**space and there are limited number of **3D** shapes for feature **learning**. To address these problems, we project **3D** shapes into **2D** space and use autoencoder for feature **learning** on the **2D** images. High accuracy **3D** shape retrieval performance is obtained by aggregating the features **learned** on **2D** images. In addition, we show the proposed **deeplearning** feature is complementary to conventional local image

descriptors. By combining the global **deeplearning** representation and the local descriptor representation, our method can obtain the state-of-the-art performance on **3D** shape retrieval benchmarks. [△ Less](#)

Submitted 25 September, 2014; originally announced September 2014.

Comments: 6 pages, 7 figures, 2014ICSPAC

113. [arXiv:1406.6947](#) [[pdf](#), [other](#)] [cs.CV](#)

Deep Learning Multi-View Representation for Face Recognition

Authors: [Zhenyao Zhu](#), [Ping Luo](#), [Xiaogang Wang](#), [Xiaoou Tang](#)

Abstract: Various factors, such as identities, views (poses), and illuminations, are coupled in face images. Disentangling the identity and view representations is a major challenge in face recognition. Existing face recognition systems either use handcrafted features or **learn** features discriminatively to improve recognition accuracy. This is different from the behavior of human brain. Intriguingly, even without accessing **3D** data, human not only can recognize face identity, but can also imagine face images of a person under different viewpoints given a single **2D** image, making face perception in the brain robust to view changes. In this sense, human brain has **learned** and encoded **3D** face models from **2D** images. To take into account this instinct, this paper proposes a novel **deep** neural net, named multi-view perceptron (MVP), which can untangle the identity and view features, and infer a full spectrum of multi-view images in the meanwhile, given a single **2D** face image. The identity features of MVP achieve superior performance on the MultiPIE dataset. MVP is also capable to interpolate and predict images under viewpoints that are unobserved in the training data. [△ Less](#)

Submitted 26 June, 2014; originally announced June 2014.