**卷积位姿机**

**Convolutional Pose Machines**

**Shih-En Wei、Varun Ramakrishna（arXiv 2016）**

**一、科学问题**

**1.1 本文所涉及科学问题**

卷积位姿机

**1.2 同行专家如何解决**

**图像结构模型，其中身体各部分之间的空间相关性表示为一个运动先验与相连的肢体耦合的树形结构图形模型。较大的部分（对应于四肢而不是关节）通常具有识别性图像结构，更容易检测。因此有助于解释较小，更难检测的部分的位置**

**1.3 本文所解决的问题**

隐式建立结构化预测任务（如关节姿态估计）中变量之间的长期依赖关系的模型，通过设计一个由卷积网络组成的顺序体系结构来实现这一点。该卷积网络直接在前几个阶段的信念图上运行，从而为零件位置生成越来越精确的估计，而无需进行明确的图形模型样式推断

**1.4 本文解决方案效果**

幸运的是，姿势机器的顺序预测框架提供了一种自然的方法来训练我们解决这个问题的深层架构。姿势机的每个阶段都经过训练，可以重复生成每个部分位置的信念图。我们鼓励网络通过在每个阶段t的输出上定义一个损失函数来重复地达到这种表示，该函数使每个部分的预测和理想信念图之间的l2距离最小化。

**二、研究内容**

**2.1 理论与方法介绍**

通过提供一个自然的学习目标函数，加强中间监督，从而补充反向传播梯度和调节学习过程，解决的训练过程中梯度消失的特征性困难。(i)training from scratch using a global loss function that enforces intermediate supervision (ii) stage-wise; where each stage is trained in a feed-forward fashion and stacked (iii)as same as (i) but initialized with weights from (ii), and (iv)as same as (i) but with no intermediate supervision.

**2.2 验证分析与实验效果**

（In summary, our method improves the accuracy in all parts, over all precisions, across all view angles, and is the first one achieving such high accuracy without any pre-training from other data, or post-inference parsing with hand-design priors or initialization of such a structured prediction task as in[28, 39].）

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

水平有限

**3.2 后续研究重点**

在计算机视觉的多个领域，如语义图像标注、单幅图像深度预测、目标检测和未来工作中，都会出现变量间的空间依赖问题。 将涉及到将我们的架构扩展到这些问题。Our approach achieves state of the art accuracy on all primary benchmarks, however we do observe failure cases mainly when multiple people are in close proximity. Handling multiple people in a single end-to-end architecture is also a challenging problem and an interesting avenue for future work

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：2D human pose estimation: New benchmark and state of the art analysis.

**（2）作者介绍**：M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele.

**（3）摘要**: 在过去的几年中，人体姿态估计取得了重大进展。然而，当前的数据集在其对整体构成估计挑战的覆盖范围方面是有限的。不过，这些都是评估、培训和比较不同模型的常用来源。本文介绍了一个新的基准“MPII人体姿势”，它在多样性和难度方面取得了显著的进步，我们认为这对人体模型的未来发展是必要的。这一综合数据集是使用800多种人类活动的既定分类法收集的。收集到的图像比以前的数据集涵盖了更广泛的人类活动，包括各种娱乐、职业和家庭活动，并从更广泛的视角捕捉人们。我们提供了一套丰富的标签，包括身体关节的位置、完整的三维躯干和头部方向、关节和身体部位的遮挡标签以及活动标签。对于每个图像，我们提供相邻的视频帧，以便于使用运动信息。鉴于这些丰富的注释，我们对主要的人体姿势估计方法进行了详细的分析，并对这些方法的成功和失败获得了见解。

**4.2 相关论文二**

**（1）题目**：Pictorial structures revisited: People detection and articulated pose estimation.

**（2）作者介绍**：M. Andriluka, S. Roth, and B. Schiele

**（3）摘要**：非刚性目标检测和关节姿态估计是计算机视觉中两个相关且具有挑战性的问题。多年来，人们提出了许多模型，并经常解决不同的特殊情况，如行人检测或电视画面中的上身姿势估计。本文认为这种专门化可能是不必要的，并提出了一种基于图形结构框架的通用方法。我们表明，正确选择外观和空间建模的组件对于模型的一般适用性和整体性能至关重要。利用密集采样的形状上下文描述符和经过鉴别训练的Adaboost分类器对车身部件的外观进行建模。此外，在生成模型中，我们将每个分类器的归一化边缘解释为似然性。零件间的非高斯关系在零件间连接的坐标系中表示为高斯关系。每一部分的边缘后方是通过信念传播推断出来的。我们证明了这种模型同样适用于检测和姿态估计任务，在最近提出的三个数据集上优于最新技术。

**4.3 相关论文三**

**（1）题目**：Monocular 3D pose estimation and tracking by detection

**（2）作者介绍**：M. Andriluka, S. Roth, and B. Schiele.

**（3）摘要**：从单眼图像序列中自动恢复三维人体姿态是一个具有挑战性和重要意义的研究课题，具有广泛的应用前景。虽然目前的方法能够在受控环境中恢复单人的3D姿势，但它们受到现实场景的严重挑战，如拥挤的街道场景。为了解决这个问题，我们提出了一个基于一些最新进展的三阶段过程。第一阶段从单帧获得人的二维关节和视角的初步估计。第二阶段允许基于检测跟踪的帧之间的早期数据关联。这两个阶段成功地将可用的二维图像证据积累到短图像序列（=tracklets）上二维肢体位置的稳健估计中。第三个阶段和最后一个阶段使用基于tracklet的估计作为可靠的图像观测，以可靠地恢复3D姿态。我们在Humaneva II基准上展示了最先进的性能，并展示了我们的方法在现实街道条件下立体跟踪的适用性。



**基于视觉的机器人抓取从目标定位、姿态估计、抓取检测到运动规划：综述**

**Vision-based Robotic Grasping from Object Localization, Pose Estimation, Grasp**

**Detection to Motion Planning: A Review**

**（Guoguang Du , Kai Wang and Shiguo Lian**

**Cloudminds Technologies）**

**一、科学问题**

**1.1 本文所涉及科学问题**

**目标定位：目标检测和分割**

**位姿估计：采用基于RGB-based and RGBD-based methods**

**抓取检测：采用传统方法和基于深度学习的方法**

**运动规划：分析方法，模仿学习方法，强化学习方法**

**此外，还对相关的数据集进行了总结，并对每一项任务的最新方法进行了比较。**

**1.2 同行专家如何解决**

**目标定位：2D object detection （传统的二维目标检测方法依赖于模板匹配，采用手工设计的描述符，研究人员根据描述子训练分类器，如神经网络、支持向量机或adaboost来进行目标检测。）**

**3D object detection：(3D neural networks, such as PointNet, PointCNN)**

**位姿估计: 6D Pose Estimation(correspondence, template, voting, and regression respectively)**

**抓取检测：针对已知对象的经验方法 分析法**

**运动规划：DMP-based methods Imitating learning Reinforcement learning to move to the grasp point Reinforcement learning to grasp the objec**

**1.3 本文所解决的问题**

第2章将介绍机器人抓取的简要历史。第三章将代表性作品分为八种方法，并对其进行了详细论述。接下来，与四个任务、评估指标和比较相关的数据集。第4章介绍了最先进的方法。最后，第5章总结了挑战和未来的方向

**1.4 本文解决方案效果**

（使用本文的方法可以达到怎样的效果，一般在总结中会提到）

**二、研究内容**

**2.1 理论与方法介绍**

**PointNet, PointCNN**

**2.2 验证分析与实验效果**

（论文中的实验分析和实验效果）

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

**目标定位：与传统手工制作的描述符方法相比，基于深度学习的方法获得了结果。然而，需要大量的培训数据和概括。 经过训练的模型的离子能力仍然具有挑战性。**

**位姿估计：这些姿态估计方法的主要挑战在于需要精确的三维模型才能得到精确的结果，而在大多数情况下，要获得精确的三维数字模型是很困难的。这就导致了基于深度学习的方法的建议，这些方法在推广到新对象方面仍然缺乏准确性。**

**抓取检测：对于端到端抓取检测方法，计算出的抓取点可能不是全局最优的，因为图像中只有目标的一部分可用。同时，他们只考虑了地球度量信息，而其他重要的因素，如物质和重量是缺失的。**

**路径规划：通过仿真生成的数据显示效率有限，尽管已经提出了一些域适应方法来提高深度机器人抓握。同时，在试图抓住物体时，要避开障碍物仍然具有挑战性。特别是当需要考虑机器人臂的结构时，很难。根据不同任务定义数学抓取模型**

**3.2 后续研究重点**

**未来辅助机器人抓取的方向包括四个方面：数据采集、语义感知、把握决策和数据培训。**

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：End-to-end real-time 3d oriented object bounding box detection from lidar point cloud

**（2）作者介绍**：Waleed Ali, Sherif Abdelkarim, Mahmoud Zidan, Mohamed Zahran, and Ahmad El Sallab.

**（3）摘要**: 三维目标检测与分类是自动驾驶（AD）中的一项关键任务。利用激光雷达传感器对周围环境进行三维点云重构，而实时检测三维物体边界盒的任务仍然是一个很强的算法挑战。本文以二维透视图像空间中一次回归元结构的成功为基础，对其进行扩展，从激光雷达点云生成定向三维对象边界框。我们的主要贡献是扩展了约洛V2的损失函数，将横摆角、笛卡尔坐标中的三维盒子中心和盒子高度作为一个直接回归问题。此公式可实现实时性能，这对于自动驾驶至关重要。我们的结果显示了Kitti基准上有希望的数据，在TitanxGPU上实现了实时性能（40fps）。

**4.2 相关论文二**

**（1）题目**：Gen eralization of human grasping for multi-fingered robot

hands.

**（2）作者介绍**：Heni Ben Amor, Oliver Kroemer, Ulrich Hillenbrand, Gerhard Neumann, and Jan Peters.

**（3）摘要**：多指机器人抓取是一个具有挑战性的问题，很难用手编程序来解决。本文提出了一种模拟学习方法，在人体演示的基础上，对掌握技能进行学习和概括。为此，我们将抓取运动综合的任务分为三个部分：（1）从人类演示中学习有效的抓取表示；（2）将接触点弯曲到新物体上；（3）优化和执行抓取和抓取运动。我们学习了不同抓持类型的低维潜在抓持空间，这为动态运动原语的新扩展奠定了基础。这些潜在空间动态运动原语被用来综合整个伸展和抓握运动。我们在一个真正的仿人机器人上评估了我们的方法。实验结果表明，该方法具有较强的鲁棒性和通用性。  
**4.3 相关论文三**

**（1）题目**：Detection and localization of grasps with deep neural networks

**（2）作者介绍**：Fu-Jen Chu, Ruinian Xu, and Patricio A.Vela. Deep grasp

**（3）摘要**：提出了一种基于深度学习的机器人操纵位置预测方法。我们考虑一个更现实的情况，即一个场景中不能有一个或多个对象。通过将抓取配置回归转化为具有零假设竞争的分类问题，以RGB-D图像输入的深度神经网络预测单个看不见对象上的多个抓取候选对象，并在单个镜头中预测多个新对象上的抓取候选对象。我们使用我们的框架对不同的场景进行了大量的实验，包括没有对象、单个对象和多个对象。我们与康奈尔数据集的最新方法进行了比较，结果表明，在图像分割和对象分割方面，我们分别可以达到96.0%和96.1%的精确度。