

基于 CNN 深度学习的机器人抓取位置检测方法

申燕萍^{1,2}

(1. 常州工业职业技术学院 信息工程学院, 江苏 常州 213164;

2. 江苏大学 计算机科学与通信工程学院, 江苏 镇江 212000)

摘要: 针对传统检测方法受到复杂环境和人工干预影响而导致检测精度低的问题, 提出了基于 CNN 深度学习的机器人抓取位置检测方法; 根据 CNN 基本结构, 研究基于 CNN 深度学习检测原理; 按照切线斜率方向划分机器人抓取位置模板点, 计算模板匹配距离, 得到机器模板上匹配点到边缘坐标图像点中最近的距离; 保持纵横坐标变量不变, 观察映射图上坐标灰度值及匹配度函数分布情况; 引入 GA 求解匹配方法, 根据匹配流程, 寻找最优解; 分析彩色图像、深度图像的可抓取位置和不可抓取位置信息, 并将其转化为符合 CNN 深度学习的数据格式, 完成信息预处理; 根据机器人抓取作业示意图, 设计具体检测流程, 并显示检测结果, 由此完成机器人抓取位置检测; 由实验结果可知, 该方法检测精度最高可达到 0.988, 能够应用到实际机器人抓取相关任务之中。

关键词: CNN 深度学习; 机器人; 抓取位置; 检测

Detection Method of Robot Grabbing Position Based on CNN Deep Learning

Shen Yanping^{1,2}

(1. School of Information Engineering, Changzhou Institute of Industry Technology, Changzhou 213164, China;

2. College of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212000, China)

Abstract: Aiming at the problem that traditional detection methods are affected by complex environments and human intervention, which leads to low detection accuracy, a CNN deep learning-based robot grabbing position detection method is proposed. According to the basic structure of CNN, the research is based on CNN deep learning detection principle. Divide the robot's grab position template points according to the tangent slope direction, calculate the template matching distance, and get the closest distance from the matching point on the machine template to the edge coordinate image point. Keep the horizontal and vertical coordinate variables unchanged, and observe the distribution of coordinate gray values and matching degree functions on the map. Introduce GA to solve the matching method, and find the optimal solution according to the matching process. Analyze the captureable and non-grabbable position information of the color image and depth image, and convert it into a data format conforming to CNN deep learning to complete the information preprocessing. According to the schematic diagram of the robot grabbing operation, a specific detection process is designed, and the detection result is displayed, thereby completing the robot grabbing position detection. From the experimental results, it can be known that the detection accuracy of this method can reach 0.988, and it can be applied to the tasks related to actual robot grabbing.

Keywords: CNN deep learning; robot; grab position; detection

0 引言

随着当代计算机与网络通信技术快速发展, 智能家居系统受到人们的密切关注, 在智能家居系统中, 机器人逐渐进入人们日常生活^[1]。机器人抓取研究被广泛应用到军事及工业等领域并作为机器人领域重点研究方向。但是实际的机器人抓取却只能完成提前设定好的任务, 一旦遇到复杂的设定环境, 抓取任务就会失败^[2]。为了满足实际需要, 对这方面的要求更加严格, 保证它不仅局限于预先

设定程序, 而且对不同需要抓取的目标在非稳定的环境下进行合理的抓取, 以达到更加准确的抓取目的^[3]。因此, 该研究是具有实际意义与价值的。智能机器人的抓取研究由于抓取的形态、位置等因素变得相对困难, 采用了一种可以直接搜索有效位置的方法, 该方法是通过在编码器的改进下学习目标位置的多模态特征学习, 采用卷积神经网络模型学习目标, 获取最优抓取位置, 并预测出目标位置处的二维图像坐标系下的坐标以及旋转角度。采用上述两种方法进行扫描并提取目标物体所标记的信息时, 受到复杂环境及人为干预影响, 导致检测精度较低。针对该问题, 提出了基于 CNN 深度学习的机器人抓取位置检测方法, 无需人工干预, 通过监督学习, 可获取目标物体在机器人坐

收稿日期: 2020-01-08; 修回日期: 2020-02-13。

作者简介: 申燕萍(1979-), 女, 江苏常州人, 硕士, 讲师, 主要从事软件技术、数据挖掘、人工智能方向的研究。

标系下的姿态, 进而完成抓取动作。

1 基于 CNN 深度学习检测原理

在深度学习目标检测领域中, 大都采用卷积神经网络模型进行检测, 该检测模型以其独特的结构设计被广泛应用。CNN 采用局部连接与权值共享这两种设计结构, 使系统具有类似生物视觉功能, 能够对图像数据进行挖掘和语义上的分析, 大幅度地减少了网络模型参数规模的复杂性, 提高了检测精准性^[4]。同时 CNN 中的特殊层结构具有高度不变性的优点, 在平移、倾斜或任意其他形式的变形情况下被设计应用, 这也是卷积神经网络更适合应用于图像特征学习与表达主要原因, 相比于其他神经网络方法检测可靠性更高^[5]。

CNN 模型的种类较多, 但是其基本结构的设计方式大致相同, CNN 模型由输入层、卷积层、池化层、全连接层、输出层组成。

1) 输入层: 输入层的作用是取得有效待测数据, 它作为 CNN 的起始层, 进行图像数据的获取和预处理。输入层只有一个数据输出操作, 并对这个输出数据进行图像随机剪裁、尺度缩放、去均值化或归一化等预处理, 并把数据图像的特征图作为网络结构层, CNN 把每一层输出的图像数据信息或含有某种语义特征信息都称为特征图^[6-8]。

2) 卷积层: 卷积层是整个 CNN 的核心层, 利用特殊的卷积核核心结构, 对输出的所有数据进行卷积, 并把卷积结果以特征图方式输出, 完成数据特征提取。CNN 的卷积层利用卷积核进行数据输出提取的独特卷积方式, 具有提取精准、效率高的优点, 是 CNN 深度神经网络重要结构^[9]。标准结构如图 1 所示。

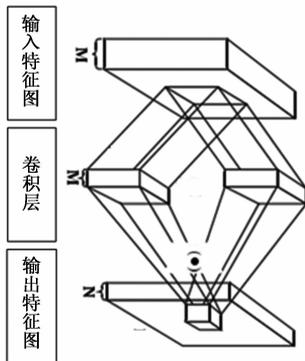


图 1 卷积层的卷积操作

通过卷积层的卷积操作流程可知, 卷积层将局部空间上感受到的信息和特征维度上的信息进行聚合处理, 形成一个完整信息聚合体, 在卷积层经过多层卷积操作后, 得到特征图^[10]。CNN 通过输入特征图, 便可对从全局里面得到的图像做相应的描述和提取。为了实现不同分层的特征提取和融合, 需要学习了解每个卷积核的不同的目标特征, 同时输入多层卷积信息^[11]。

3) 池化层: 池化层在 CNN 中主要是负责对输入的数

据样本进行采样操作。池化层通过对输入特征图的各个空间维度上的特征进行切片划分, 并将划分的特征图设置为一个毫无交点的区域, 然后池化每个所对区域的特征^[12]。

每个区域的特征图像在它的池化过程中, 需要选取尺寸为 2×2 , 但如果有一些非常大的图, 也可以采用的尺寸, 但是这种过大尺寸会造成在池化过程中信息的大量流失。普遍采用平均值和最大值两种池化方法, 过程如图 2 所示。

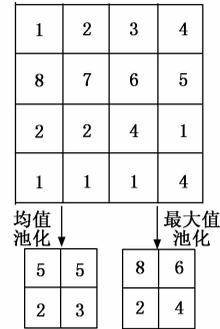


图 2 池化处理

根据均值池化和最大值池化示意图在每个区域的特征图像按照的尺寸进行池化时, 均值池化是将每个区域内的元素进行均值计算, 并作为输出特征值, 或者是把不同区域的相同元素加起来作为输出特征值。每个元素在不同区域内的最大值作为输出特征值, 这两种不同的池化方式能够正确有效地提取不同特征值。

4) 全连接层: 全连接层位于池化层和输出层之间, 主要负责全连接对最后一个池化和输出层的特征图。目前全连接层在 CNN 中已经逐渐被表面全局均值池化层所替代, 可以得到比全连接层更好的效果。全连接层的工作原理与单层前馈神经网络相同, 通过将输入的高维度特征图采用卷积的方法降为一维, 或者直接将具有特征矩阵的元素排列成一维向量, 这两种降维方法都是为了便于输出层的分类与处理。

5) 输出层: 输出层是 CNN 的结尾层, 通过对全连接层得到的一维特征向量, 采用 Logistic 回归或 SVM 进行分类, 或者采用 Softmax 回归等概率形式完成特征向量的输出。输出层基本采用 Softmax 回归的输出方式, 利用逻辑回归的推广原理, 能够处理不同形式的分类问题。

2 机器人抓取位置检测

2.1 抓取位置目标匹配

将机器人抓取位置模板点按照切线斜率方向进行角度划分, 结合距离匹配定义, 得到按照倾角分层的模板匹配距离:

$$d = \frac{1}{M} \sum_{i=1}^M F_m(x_i) \quad (1)$$

式 (1) 中, M 为匹配点总数; x_i 为匹配点投影在机器上的坐标值; m_i 为机器模板点根据切线斜率方向的第 m 幅距离映射图上的取值; $F_m(x_i)$ 为映射图上坐标灰度值。根据上述公式可得到机器模板上匹配点到边缘坐标图像点

中最近的距离。

为研究按倾角分层的距离对映射图上坐标灰度值的影响，需保持横纵坐标变量保持不变，并观察匹配度函数分布情况，结果如图 3 所示。

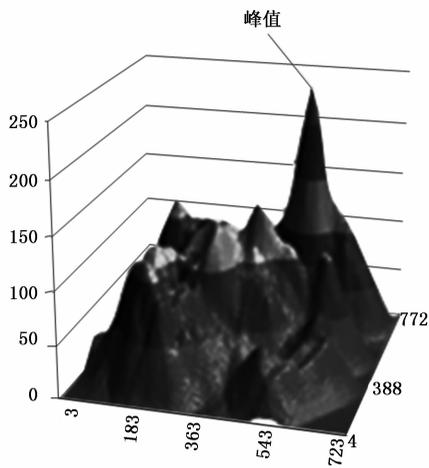


图 3 匹配度函数分布

由图 3 可看出，按倾角分层方法峰值较为明显，能够提高匹配度函数灵敏性。引入 GA 求解匹配方法，匹配步骤如图 4 所示。

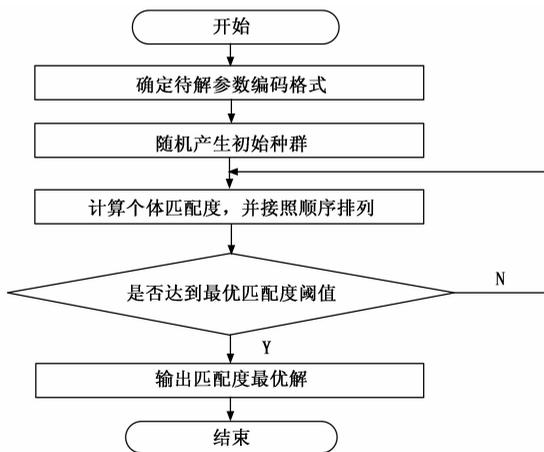


图 4 匹配最优解流程图

GA 作为 CNN 深度学习的随机搜索算法，能够使种群更加靠近全局最优值，一旦搜索接近峰值时，个体占据排序在最前端。当 GA 引进一段时间后，以其初始搜索结果为基准，寻找最优解。GA 求解方法只能对一个变量进行优化，而相似度函数包含了多个变量，这些变量对于图像来说是具有独立意义的，因此，使用该方法能够在初值附近匹配到最优解。

2.2 信息获取与处理

将 CNN 深度学习方法应用到机器人抓取位置检测之中，应先将抓取位置作为深度学习对象，并将目标可抓取位置视为 CNN 深度学习样本数据。抓取位置检测过程可描述为：依据目标物体大小选择不同方向矩形框，并在目标

物体上提取抓取位置，并将抓取位置均匀输入到 CNN 深度学习中进行预测。

抓取位置的 CNN 深度学习结合了多个模态特征信息，其中包括通道彩色图像信息、深度图像信息。图 5 分别列出了两种图像可抓取位置和不可抓取位置信息。

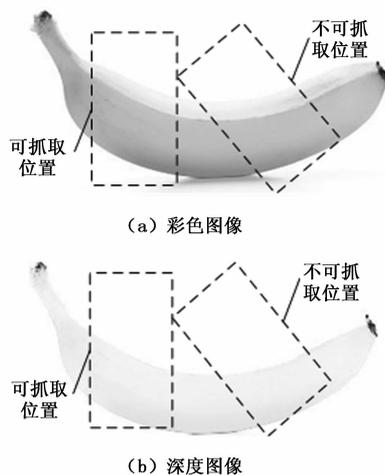


图 5 抓取位置信息

将抓取位置信息进行预处理，将其转化为符合 CNN 深度学习的数据格式，以此提高 CNN 深度学习隐含层提取特征的能力。对于数据的预处理，需先将标记区域对应的模态数据统一转化为单模态的特征图尺寸向量；然后将处理后得到的特征向量输入到多模态特征预训练过程之中；最后将每层训练目标最小化函数代入训练过程之中，由此完成信息预处理。

2.3 检测流程设计

机器人抓取作业示意图如图 6 所示。



图 6 机器人抓取作业示意图

利用机器人视觉系统分析给定的场景彩色图像，推断出目标物体最优抓取状态。机器人抓取过程中主要包括抓取点确定和姿态预估两个阶段。采用 CNN 深度学习法设计具体检测流程，如下所示：

- step1: 输入处理后的抓取位置信息；
- step2: 以射频网为基础，利用边界框及其分数对信息进行分类研究，实现抓取定位角度的粗估计；
- step3: 经过信息筛选与排序，获取边界框顶点坐标；
- step4: 簇估计角度，大约定位抓取位置，再细估计角度；

step5: 将粗估计与细估计角度值相互融合, 获取抓取角度;

step6: 根据边界框顶点坐标, 抓取中心坐标;

step7: 输出检测结果。

2.4 抓取位置检测结果显示

使用旋转矩形框表示目标物体最优抓取位置, 在检测该位置时, 需将包含目标物体的图像输入到 CNN 深度学习过程之中, 经过训练后, 选定机器人最合适的抓取位置坐标和所涉及图像的平面旋转角。

抓取位置检测结果显示如图 7 所示。

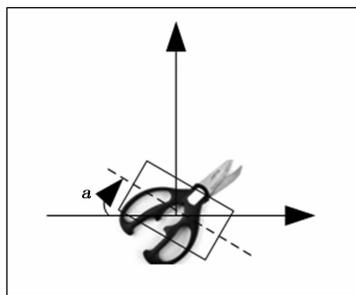


图 7 抓取位置检测结果显示

机器人的抓取方向根据其习惯, 定为矩形框长边方向, 表示矩形框相对图像坐标系横轴旋转角度, 箭头表示矩形框旋转方向。根据该内容, 可得到高精度抓取位置检测结果, 利用 CNN 深度学习方法, 获取清晰显示结果。

3 实验验证

为了验证基于 CNN 深度学习的机器人抓取位置检测方法有效性, 实验环境为 Win7 操作系统, 该系统训练环境为 Matlab R2015a。

3.1 实验准备

3.1.1 数据分析

选用 Canon VC-C50 型号的单目标相机, 具有 320 × 240 分辨率, 在该相机下获取的目标物体图像如图 8 所示。



图 8 样本示例

数据训练所需的基本数据为 50 张不同场景中的图像, 抓取角度为 0°, 无需人工标注边界框就可获取抓取位置, 通过旋转图像将数据扩充到 200 张, 具体扩充方法如下所示:

围绕边界框中心点, 扩充边长为原矩形框边长之和的正方形; 以 5°间隔将图形围绕圆心旋转一周, 能够保证抓取的位置在图形中央, 方便特征提取, 能够省去修改边界框的繁琐步骤; 为了减少第 1 阶段定位耗时, 需将图片合并缩减为 0°、45°、90°、135°。

3.1.2 角度归一化处理

因样本数据规模较小, 内部层数多且复杂, 使用微软数据集训练相关数据作为初始值, 并在此基础上进行微调处理。采用动量法, 在大量噪声环境下对角度进行归一化处理。

3.2 机器人抓取模式设置

通过机器训练, 使其能够精准分辨出目标, 这是机械视觉学习最简单有效的模式。机器人会从多个角度确定目标位置, 然后将所有位置边缘点形成一个整体坐标, 形成三维立体图。

在实验过程中, 机器人成功抓起物体, 如图 9 所示。



图 9 机器人抓取模式

先建立 3D 模型, 再经过一系列分析, 可以判断物体当时所摆放的位置, 进而完成一套抓取动作。

该过程中机器人抓取训练的 2 个阶段时间分别为 15 小时和 1 小时, 其训练曲线如图 10 所示。

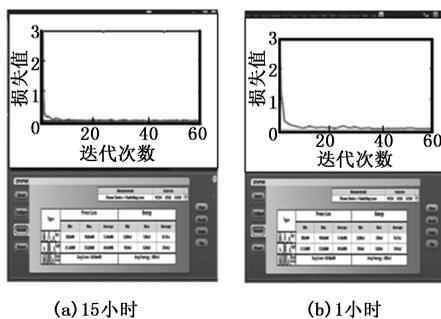


图 10 训练曲线

由图 10 可知: 两个时间阶段下的曲线都趋近于收敛状态。

3.3 实验结果与分析

根据上述内容可知, 机器人实际抓取的物体是杯子和

球，两个物体分别放在桌子上和地上。分别采用传统的改进自动编码器多模态特征学习、卷积神经网络模型学习方法和基于 CNN 深度学习方法分别在训练时间为 15 小时和 1 小时情况下进行对比分析。其中 1 小时无人工干预，环境简单；15 小时受到人工干预，环境复杂。

1) 15 小时：在 15 小时情况下，将 3 种方法的检测精度进行对比分析，结果如图 11 所示。

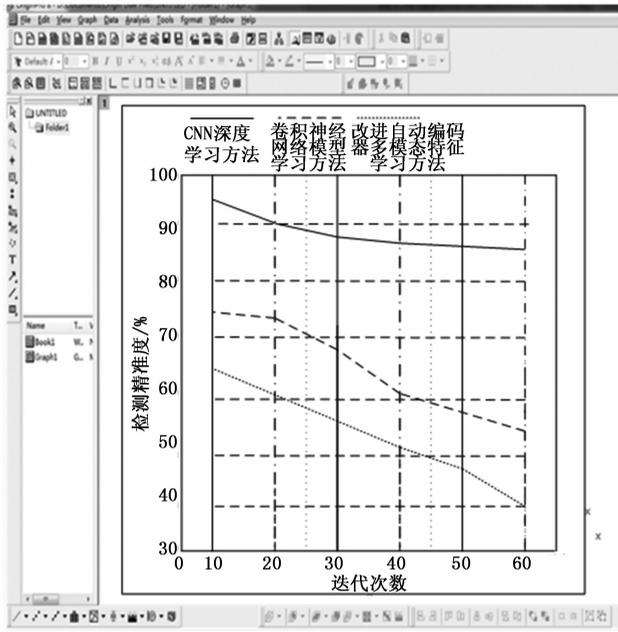


图 11 15 小时情况下 3 种方法检测精度对比分析

由图 11 可知：采用 CNN 深度学习方法检测精度始终维持在 85% 以上；采用卷积神经网络模型学习方法检测精度始终维持在 50% 以上；采用改进自动编码器多模态特征学习方法检测精度始终维持在 40% 以上。由此可知，在 15 小时情况下，采用 CNN 深度学习方法检测精度较高。

2) 1 小时：在 1 小时情况下，将 3 种方法的检测精度进行对比分析，结果如表 1 所示。

表 1 1 小时情况下 3 种方法检测精度对比分析

迭代次数/次	改进自动编码器多模态特征学习方法	卷积神经网络模型学习方法	CNN 深度学习学习方法
10	0.410	0.652	0.985
20	0.409	0.610	0.988
30	0.387	0.588	0.972
40	0.352	0.548	0.975
50	0.331	0.525	0.981
60	0.308	0.517	0.979

由表 1 可知：采用 CNN 深度学习方法检测精度较高，在迭代次数为 20 次时，最高检测精度为 0.988。而采用其余两种方法，都比该方法检测精度要低，当迭代

次数为 10 次时，卷积神经网络模型学习方法最高为 0.652，最低也高于 50%，而改进自动编码器多模态特征学习方法最高检测精度为 0.410，最低为 0.308。由此可知，在 1 小时情况下，采用 CNN 深度学习方法检测精度较高。

4 结束语

针对传统方法存在的问题，提出了基于 CNN 深度学习的机器人抓取位置检测方法，并对该方法进行了实验验证。在实验中，通过与改进自动编码器多模态特征学习方法、卷积神经网络模型学习方法作对比，实验结果表明基于 CNN 深度学习方法能够较好提取目标位置特征，具有精准检测效果。相比于以往方法，既提高了检测方法实时性，又增强了姿态角检测精度，大大减轻了机器人内存消耗。

在未来研究项目中，使用基于 CNN 深度学习检测方法是进一步完善传统方法的网络结构，提高机器人抓取位置检测精度，并将该方法应用到机器人抓取相关任务之中。针对 CNN 深度学习领域轻量化研究，较为繁琐，该领域有着较长的研究周期，很难应用到实际中，所以，权值轻量化的结合需应用到抓取目标位置中，接下来的研究方向可以将重点放在该领域。

参考文献：

- [1] 丁斗建, 赵晓林, 王长根, 等. 基于视觉的机器人自主定位与障碍物检测方法 [J]. 计算机应用, 2019, 39 (6): 1849-1854.
- [2] 郭瑞峰, 张文辉, 彭战奎. 关节型码垛机器人自主避障的方法研究 [J]. 机械设计与制造, 2019, 5 (7): 244-248.
- [3] 王涛, 郭金亮. 基于图像处理的机器人精确抓取的设计与实现 [J]. 制造技术与机床, 2018, 5 (12): 47-51.
- [4] 王肖锋, 张明路, 刘军. 基于增量式双向主成分分析的机器人感知学习方法研究 [J]. 电子与信息学报, 2018, 40 (3): 618-625.
- [5] 王战玺, 李树军, 赵璐, 等. 移动机器人铣削制孔系统基准检测 [J]. 南京航空航天大学学报, 2019, 7 (3): 281-287.
- [6] 袁飞. 线性二次型球式滚动机器人运动稳定性研究与测试分析 [J]. 中国工程机械学报, 2018 (5): 0462-0466.
- [7] 刘汉伟, 曹维清, 王永娟. 基于非结构基本组成分析的自主抓取方法 [J]. 机器人, 2019, 41 (5): 583-590.
- [8] 王营华, 宋光明, 刘盛松, 等. 一种视觉引导的作业型飞行器机器人设计 [J]. 机器人, 2019, 41 (3): 353-361.
- [9] 刘自文, 赵亮, 于鹏, 等. 柔性外骨骼手的抓取力控制方法 [J]. 机器人, 2019, 41 (4): 483-492.
- [10] 刘理, 王耀南, 万智, 等. 基于序列 2D 结构化线特征的桥梁检测机器人末端位姿估计方法 [J]. 电子测量与仪器学报, 2018, 10 (8): 59-68.
- [11] 陈友东, 刘嘉蕾, 胡澜晓. 一种基于高斯过程混合模型的机械臂抓取方法 [J]. 机器人, 2019, 41 (3): 343-352.
- [12] 张潇, 张秋菊. 仿人手掌的机器人变掌机构设计与分析 [J]. 工程设计学报, 2019, 10 (4): 385-394.