

调研报告：机器人应用与 AI 产业

石睿 (211300024、211300024@smail.nju.edu.cn)

(南京大学 人工智能学院, 南京 210093)

摘要： 本文对机器人行业的背景进行简单介绍，并根据某 pick-and-place 智能机器人的例子展示机器人应用中的 AI 技术，最后给出作者自己对于机器人产业发展的一点思考

关键词： 机器人；机器视觉；深度学习；语音感知；CNN

中图法分类号：TP301 文献标识码：A

1、行业背景

机器人技术是综合了计算机、控制论、信息和传感技术、人工智能、仿生学等多学科而形成的高新技术，集成了多学科的发展成果，代表高技术的发展前沿，是当前科技研究的热点方向。自 1956 年机器人产业诞生后，经过近 60 年发展，机器人已经被广泛应用在装备制造、新材料、生物医药、智慧新能源等多个高新产业。^[2]

60 多年机器人在学术和工业届的发展大致可以分为三个阶段：^[1]

第一阶段：可编程示教再现型机器人。此时人手把着机械手，机器把手只能把应当完成的任务做一遍。

第二阶段：具有一定的感觉和自适应能力的离线编程机器人，可以根据作业对象的状况改变作业内容。

第三阶段：智能机器人。随着人工智能的发展，促进了语音识别，图像识别，自然交互能力的提升。如今人工智能与机器人结合，为机器人的发展插上了腾飞的翅膀。这种机器人带有多种传感器，能够将多种传感器得到的信息进行融合、能够有效的适应变化的环境，具有很强的自适应能力、学习能力和自治功能。



图 1：如今设计智能机器人的基本流程图

机器人的产生是要服务于人们日常工作生活的，依据机器人可以给人们提供的帮助，主要分为以下几种类型，不同类型所需要的人工智能技术的加持也大相径庭。

分类	完成的功能	
工业机器人	DEF: 替代人工完成各类繁重、乏味或有害环境下的体力劳动，多为应用于多关节机械手或多自由度的机器装置 具有一定的感知功能：如视觉、力度感知、位移检测等，从而实现对环境和工作对象自主判断和决策。可以依靠自身的动力能源和控制能力实现工业加工，如：焊接、喷涂、放置、搬运等	
服务机器人 ^[3]	家用服务机器人	DEF: 服务于人们生活的机器人，辅助完成人们生活中一些需要琐碎的、花费时间较长的工作，其类型主要可以分为以下三种类型。 可以自动在房间内完成清理工作如扫地、窗户清理等的只能家用电器。 如扫地机器人、擦窗机器人等。此类服务机器人的市场发展的较为成熟
	娱乐休闲机器人	用于激发使用者学习兴趣、培养综合能力的机器人教学工具或者提供早教等功能的机器人。 如儿童陪读机器人，儿童教育学习机器人
	养老陪护机器人	应用于家庭或者养老院等环境中，帮助部分老年人有人看护的心理需求，并且解决翻身、健康监测以及提供陪伴逗乐、吃药提醒等功能。 如护理机器人等
特殊用途机器人 ^[4]	DEF: 服务于特殊任务的机器人，常常这些任务具有较大的危险性或者需要很高的精度，起到对执行任务人的辅助或直接执行任务。 比如扫雷机器人，地下探测机器人，海洋探测机器人 以及如今应用较广的医疗机器人，可以用于辅助外科医生进行手术（灵活精准，手术时间段、创伤小出血少）或替代、辅助康复治疗师对行动障碍人群进行治疗。	

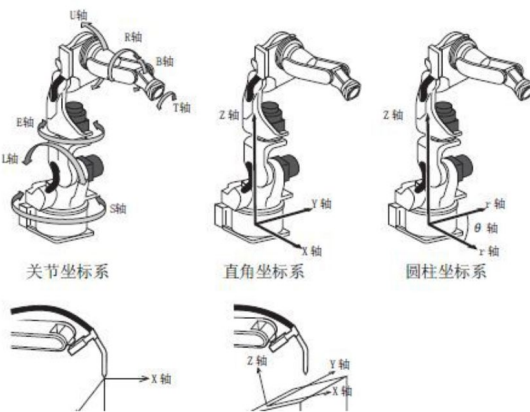


图 2.3.4:左、中、右分别为工业机器人、服务机器人和特殊用途机器人的示例

2、关键技术与应用实例

在一年多的专业学习以及在科沃斯企业的参观之中，如今受人工智能技术加持而蓬勃发展的多为商业机器人（餐饮配送，外卖快递小车等）以及家用服务机器人（扫地机器人等）。他们所共有的所需技术可以总结为如下几点：① **SLAM**，即同步定位与建图（Simultaneous Localization and Mapping）。机器人在未知环境中从一个未知位置开始移动，在移动过程中根据位置和地图进行自身定位，同时在自身定位的基础上建造增量式地图，实现机器人的自主定位和导航。② **环境感知**，即完成多个传感器的融合，综合信息得到全局视角；障碍物识别，是规划机器人行走路线的前提；地图构建，科沃斯的工程师说他们的机器人常常在人们不在家、不在工作状态中的时候进行家庭环境建模，因为人们的生活会使部分物品摆放出现移动，所以地图构建是一件经常要完成更新的事情。③ **运动控制**，即底盘、核心零部件控制单元以及完成核心功能的控制单元，如地板清洁。④ **人机交互**，即通过自然语言处理（NLP）、人脸识别、手势识别等解析人们对机器人发送的指令，并作出相应。

以下通过一个日本安川电气（此公司以运动控制、机器人和系统工程研发和生产为核心）研发的 Motoman（MotoMINI）——一种抓放（pick-and-place）机器人的应用实例对机器人的某些技术进行介绍。^[4]



MOVJ	功能	以关节插补方式向示教位置移动。	
	添加项目	位置数据、基座轴位置数据、工装轴位置数据	画面中不显示
	VJ= (再现速度)	VJ:0.01 ~ 100.00 %	
	PL= (定位等级)	PL:0 ~ 8	
	NWAIT		
	UNTIL 语句		
MOVL	功能	以直线插补方式向示教位置移动。	
	添加项目	位置数据、基座轴位置数据、工装轴位置数据	画面中不显示
	V= (再现速度)、	V:0.1 ~ 1500.0 mm/秒	
	VR= (姿态的再现速度)、	R:1 ~ 9000 °/分	
	VE= (外部轴的再现速度)	VE:0.1 ~ 180.0 °/秒	
	PL= (定位等级)	PL:0 ~ 8	
MOVE	功能	以圆弧插补方式向示教位置移动。	
	添加项目	位置数据、基座轴位置数据、工装轴位置数据	画面中不显示
	V= (再现速度)、	V:0.1 ~ 1500.0 mm/秒	
	VR= (姿态的再现速度)、	R:1 ~ 9000 °/分	
	VE= (外部轴的再现速度)	VE:0.1 ~ 180.0 °/秒	
	PL= (定位等级)	PL:0 ~ 8	

图 5、6：Motoman 的模型样例以及部分可以实现的功能

2.1 机器视觉+深度学习

2.1.1 基于深度学习的机器视觉

机器视觉是一套光机电计结合的复杂设备，涉及光学成像、传感器、视频传输、机械控制、相机控制、图像处理等多种技术，每个环节都会影响到最终的检测结果。机器视觉重点观测目标的特征、尺寸、形态等信息，其目的在于根据判断的结果来控制现场的设备动作。相比计算机视觉更偏向于理论算法的研究，机器视觉更加强调算法的实时性、高效率和高精度。

一个完整的机器视觉由以下几部分组成^[5]:

①一个或多个工业相机（包括可见光、非可见光、多光谱）；②照明光源（包括可见光、非可见光、激光、闪光灯）；③光学系统（包括一个或多个透镜组和滤光片）；④机械部分（为整个机器视觉系统提供安装基准和部件调节功能）；⑤电气部分（主要是供电、走线和数据接口）；⑥硬件部分（包括工控机、嵌入式系统、图像采集卡、传感器、触发器、执行器），对于涉及与机器人协作的机器视觉系统，机器人可被视为执行器；⑦算法；⑧软件系统（包括人机交互界面、与其他类计算机设备的通信模块、远程维护模块、离线更新模块）。

随着深度学习的发展，深度学习也逐渐和机器视觉进行了交叉，其好处有如下两点。**一、更能适应复杂场景和低质量图像。**当生产现场由于条件所限，无法提供理解的成像条件（如外部粉尘、光线、人等的干扰，或者拍摄成像不具有高对比度）时，使用深度学习也可以获得较好处理结果，从而有可能满足实际任务需求。**二、能获得更高的算法处理准确度。**当用户对系统的最终处理结果有很高的精度要求时，如高精度语义分割或目标检测等任务，相比传统机器视觉中通过提高相机分辨率来优化，使用深度学习可以获得更好的效果。

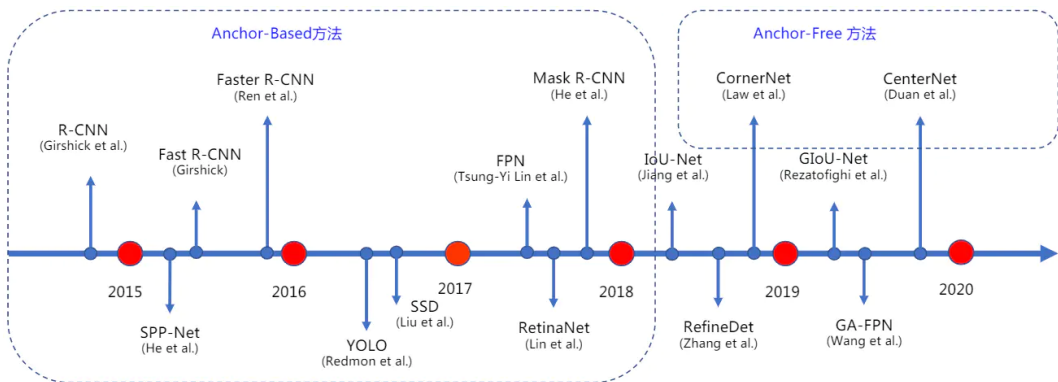


图 7：基于深度学习的目标检测算法发展史

2.1.2 Motoman 的应用

Motoman 作为一种抓放机器人，其本身需要对被抓放的物体进行空间建模。而被抓放的物体作为本身包

含非常丰富信息的图像，也经常被机器人系统用于对物体的感知和识别。安川电气的研发团队将深度学习 DL 应用在计算机视觉上，帮助机器人实现基于图像的物体检测。具体识别和训练的过程如下所述：

离线训练

Step1: **【对训练集中的被抓取物体进行建模】**通过传感器获得被抓取物体的图像信息，根据机器视觉算法对被抓取物体进行建模。识别物体的边缘信息以及空间信息。

Step2: **【训练目标检测函数，Motoman 运用的是一种 Anchor-Based 方法】**对被抓取物体的部分建模信息打上标签，把图片作为输入，把图片中的物品边缘作为锚框，把标签作为输出，到网络中训练识别函数。

目标检测算法中，通常会在输入图像中采样大量区域，然后判断这些区域是否包含所感兴趣的目标，并调整区域边界从而更加准确的预测目标真实的边缘框。

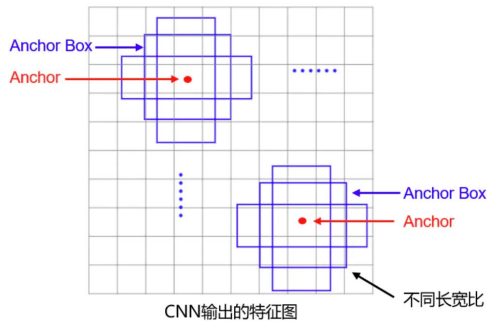


图8：目标检测算法中锚点和锚框（用于网络训练函数）

在线识别

Step1: 对于新的物体重新进行建模，识别 label 和锚点锚框。把这些作为训练的目标检测网络的输入，机器手臂就可以得到对应的目标物体的信息啦。

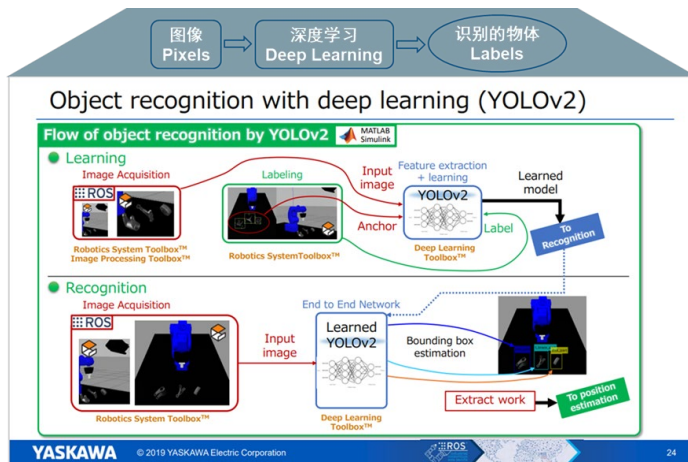


图 9：Motoman 中基于深度学习的机器视觉流程图

2.2 语音感知+卷积神经网络

2.2.1 CNN^[6]

在 Motoman 的语音感知过程中使用了卷积神经网络（CNN），它是一类深度前馈人工神经网络，旨在模拟动物视觉皮层的生物学行为，最常用于分析视觉图像。它由卷积层和池采样层组成，以便网络可以对图像属性进行编码。相比其他浅层或深度神经网络，卷积神经网络需要考量的参数更少，进而可以在特征数量爆炸的情况下（如多通道的图像处理），只选择最重要的一部分特征进行分析。

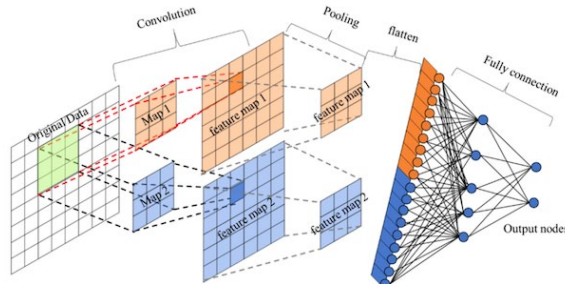


图 10: CNN 示意图

CNN 本质是一个多层感知机（MLP），成功的原因在于其所采用的**局部连接**和**权值共享**的方式：一方面减少了权值的数量使得网络易于优化。另一方面降低了模型的复杂度，也就是减小了过拟合的风险。其中隐含层的卷积层和池采样层是实现卷积神经网络特征提取功能的核心模块。该网络模型通过采用梯度下降法最小化损失函数对网络中的权重参数逐层反向调节，通过频繁的迭代训练提高网络的精度。

卷积神经网络通过三个关键的操作来实现，其一是局部感受野，其二是权值共享，其三是池采样层。

Step1: 局部感受野

在卷积神经网络中，感受野是指特征图上的某个点能看到的输入图像的区域，即特征图上的点是由输入图像中感受野大小区域的计算得到的。神经元感受野的值越大表示其能接触到的原始图像范围就越大，也意味着它可能蕴含更为全局，语义层次更高的特征；值越小则表示其所包含的特征越趋向局部和细节。因此感受野的值可以用来大致判断每一层的抽象层次。如下面左图所示。

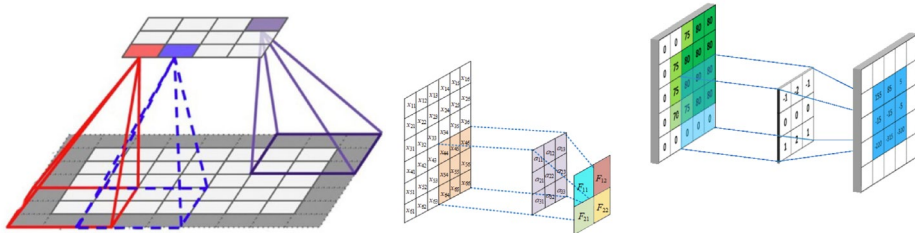


图 11、12: 局部感受野和权值共享示意图

Step2: 权值共享

当给定一张输入图片，用一个卷积核去扫描这张图，卷积核里面的数就叫权重，这张图每个位置是被同样的卷积核扫的，所以权重是一样的。其实就是一张图片使用同一个卷积核内的参数。如上面右图所示

Step3: 池化层

池化层中池化的目标是压缩数据和参数的数量，同时在此过程中也应该遵循以下两个思想：第一，特征不变形：池化操作使模型更加关注是否存在某些特征，而不是特征的具体位置。第二，特征降维：在空间范围内做了维度约减，从而使模型可以抽取更加广泛的特征，也减小了下一层的输入大小，减少参数数量。

2.2.2 Motoman 的应用

在 Motoman 的例子中使用基于 CNN 的语音驱动的控制技术，实现人机交互。具体的实现流程如下所述：

离线学习

Step1: 【对训练集中的语音处理】把训练集中的语音信息处理成(时间,强度,频率)的一个三元组，并且每一个三元组对应的自然语言中的“词”作为标记

Step2: 【CNN 训练语音识别函数】把上述的(时间,强度,频率)三元组作为 CNN 的输入，把自然语言中对应的“词”作为输出进行训练

在线识别

Step1: 对于新的语音进行如上第一步的信息处理，把这些作为训练好的语音识别函数的网络的输入，机器人手臂就可以得到对应的命令信息啦

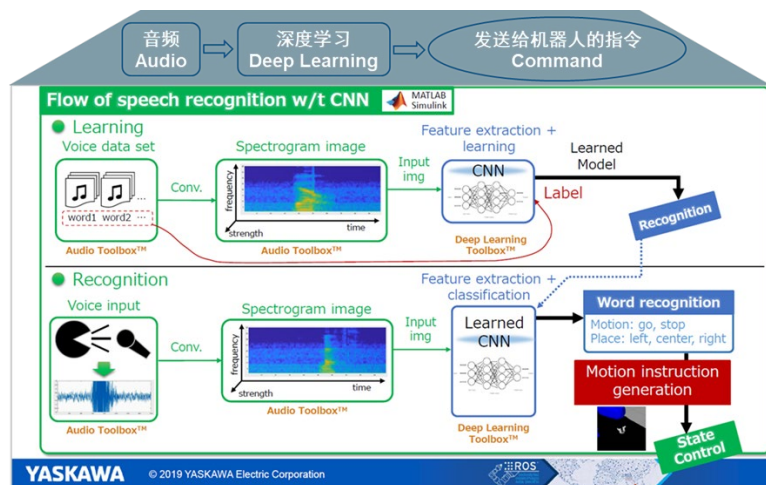


图 13: Motoman 中基于 CNN 的语音感知流程图

3、个人思考

在当年的技术限制下，机器人产生的目的仅仅为了帮助人们完成一些繁琐、重复性、不需要太多技术含量的事情。如今随着各种智能化技术的发展，机器人研发的目的更加的长远，其目标是替代人们完成一系列可以有很大难度但有规律可循的大型工程或项目、和人们协同合作完成工作了。

比如最近发展迅速的**工业机器人与网络融合**，组成复杂性强的生产系统。各种智能算法如蚁群算法、免疫算法等可以逐步应用于机器人应用中，使其具有类人的学习能力，多台机器人协同技术使一套生产解决方案成为可能。

马斯克创办的 Neuralink 公司也让**人类意识控制机器人**火了起来，这一新操作模式利用人们的“思维力”和“意志力”控制机器人的行为。最出名的可谓是脑机接口了，这是一种全新的、不依赖脑内神经和肌肉的交流与控制通道，可实现人（或其他动物）大脑与外部设备的双向直接交互。Natruue 去年发布的文章中有研发团队研发出“意念打字”技术，甚至可以称作“意念写字”技术。该团队通过脑机接口技术，将受试者脑中想象的“笔迹”转为屏幕文本，准确率超 99%，且字符输入速度可达每分钟 90 个……

我在日常生活中也遇到的一些智能机器人，比如汽车中的交互系统、扫地机器人、快递外卖小车车等。从他们完成工作的过程中可以看到，智能机器人完成特定功能已经不成问题，只是灵活性上会有一点不足。但是在准确理解人们下达的命令的过程中会有非常大的偏差，几乎只能识别自然语言中特定的一些词汇。所以尽管机器人能够通过简易的感应方式理解人类语言、图形、身体指令，利用其部分组件免除工人复杂的操作。但现有阶段的人机协作存在较大的安全问题，尽管具有视觉和先进传感器的轻型工业、商业、家用机器人已经被开发出来，但是目前仍然缺乏可靠安全的机器人协作的技术规范。

参考文献及文章：

- [1] 孟庆春,齐勇,张淑军 智能机器人及其发展 中国海洋大学学报
- [2] 全球机器人产业发展现状及趋势 <https://siaf.gymf.com.cn/newslst/industrynews/35476>
- [3] 2022 中国机器人行业蓝皮书 来源：CIC
- [4] 人工智能（AI）应用于机器人的典型场景 <https://m.elecfans.com/article/1883875.html>
- [5] 机器视觉系列概述 <https://blog.csdn.net/oruchimaru0420/article/details/125122598>
- [6] 卷积神经网络详解 <https://blog.csdn.net/jiaoyangwm/article/details/80011656/>