# 计算机视觉领域发展现状和关键技术

姓名：王子贤

学号：15030140049

根据百度百科的解释，计算机视觉是一门研究如何使[机器](https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8/2275865)“看”的科学，更进一步的说，就是是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等[机器视觉](https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8%E8%A7%86%E8%A7%89/7414484)，并进一步做图形处理，使电脑处理成为更适合人眼观察或传送给仪器检测的图像。作为一个科学学科，计算机视觉研究相关的理论和技术，试图建立能够从图像或者多维数据中获取‘信息’的[人工智能](https://baike.baidu.com/item/%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD/9180)系统。这里所 指的信息指Shannon定义的，可以用来帮助做一个“决定”的信息。因为感知可以看作是从感官信号中提 取信息，所以计算机视觉也可以看作是研究如何使人工系统从图像或多维数据中“感知”的科学。

从这里来看，我们可以总结成一句话，计算机视觉是一门科学。一门科学的形成和发展是不开的社会的发展以及当前学术研究的突破。计算机视觉就是在生物和计算机这两个主要的领域之间都发展到一定阶段，以及由社会生产力的需求驱动下诞生出来新的研究领域。目的是利用计算机实现人眼视觉功能，从而提高社会生产力。

这里我们可以简单地将人眼视觉功能分为获取，处理和理解图像信息。那么这个过程具体是怎样的呢？

第一是获取，获取就是通过光的反射形成颜色，而物体的颜色和形状就是人类大脑需要理解的重要两个点。

第二就是处理，处理是将图像映射在视网膜上面同时将这两种信息传递到大脑中。大脑会根据传递的信息提取物体中有用的内容，比如说，轮廓和颜色。接下来大脑会开始搜寻记忆中的存储内容看有没有类似长的像的物体，然后识别出来。或者根据物体的描述内容以及其他的物体，推测出来目标。

第三就是理解，当我们的大脑处理完图像以后，就会思考图像表达的意图，通过一个个识别出来的物体，大脑可以结合语言神经系统，把图像描述出来这是什么。

当然，这只是一个很粗略人类理解图像的这么一个过程。实际中，还要涉及到神经系统的脉冲信号传递以及2D图像重建3D物体等等一系列复杂过程。计算机视觉的目标就是实现人类的视觉系统的功能，并且最终超过人类视觉系统。我们可以抽象来看这个视觉系统的作用，就是对信息的加工和转换。计算机视觉研究的内容可以不必要完全模仿人类大脑的视觉系统的运作。而是拥有自己的一套运作过程和结构。

人类的视觉系统有多复杂，计算机视觉研究的内容就有多么浩瀚无垠。比如现阶段研究的方向比较成熟的有：图像处理、模式识别或图像识别、景物分析、图象理解、空间形状的描述、几何建模以及认识过程等。研究的方法和手段数目也是年年递增，当然大多数方法都是在经典的算法上进行部分的调整，然后变的更加完善以及精准适应生活场景中，比如svm，贝叶斯算法，K-近邻算法，决策树等等。现阶段最流行的算法当然是深度学习。但是深度学习在工程领域上，虽然非常的流行，但是因为缺少理论的强大支持，很多科研人员对于深度学习这个方法仍在探索中。

方法虽然繁多，那么运用到要解决的场景，现阶段又做到什么程度了呢？

对于图像识别，例如人脸识别的正确率已经远远超过人眼的识别，在物体的检测上，只要数据量足够大，已经可以达到超过人眼的识别率。从2009年开始，每年都会有imagenet比赛，这个比赛就是比拼物体识别和检测。这个比赛在2017年举办完最后一场便宣布这是最后一届。因为现阶段的深度学习的网络模型识别错误率已经明显低于人类。错误率低至2.2%。这一领域已经达到非常成熟的阶段了。

解决了人类视觉系统处理的第一个步骤和第二个步骤的识别过程，那么下一个研究的热点就是图片的分类。切割一张图片。也就是将一张图片的内容提取出里面的所有物体。这一难度非常大，以为前者比赛是在图像标注很好的情况，就能做到而后者则是在现实生产环境中，拥有非常多的数据资源，有对的有错的。如何能做到正确的分类，这计算机视觉进入到图像理解的重要阶段。于是诞生新的方法，比如自步学习（Self-paced Learning）和课程学习（curriculum learning）。这是今年新提出来的概念。具体来说，是先学习简单物体的分类，然后再逐渐增加难度，慢慢过渡到更高难度和更专业性的物体。最终达到所有物体的分类和切割。

这一领域目前仍在探索中，相信在现阶段成熟的计算资源和丰富的识别经验，图像处理的将会很快进入下一步就是图像的精准理解。

在我们生活中，已经有许多机器具备识别功能，并且借助它们的识别给我们带来更多的方便和舒适的体验。比如手机刷脸解锁，让手机开的更快。刷脸进入火车站，极大提高了火车站人的流动。洗衣机的识别并分类衣物，让衣服清洗的更加容易。车牌号的识别，不放走一个违法车牌。病变细胞的发现，让老花眼的医生摘下老花镜，更加及时的提供病人治疗。家禽的识别，更科学的合理养殖家禽，吃放心的鸡鸭鱼肉。总之，计算机视觉突破的发展，也给工业界带来数不尽的财富和节省数不尽的人力。人民的生活水平质量更加提高。

成熟的产业已经在加快推广各个城市和乡村。在13亿以上的人口大国，智能化城市将有助于我们的政府的管理和发现问题。那么推进无人驾驶汽车，城市大脑，以及医疗设备自动化都是需要更精准的图像识别来推测机器下一步的动作。

无人驾驶在国内属于探索中，这一产业和美国相比，属于比较接近。并且有望超越。在百度的Apollo和DuerOS两个项目的开源后，国内无人驾驶立刻进入无人车创业时代，比较著名的有，景驰，Momenta，商汤，Pony.ai等等。我国的政府也是加大财政力度投资这类公司，以及放宽政策提供这类公司试验场地。希望可以加速发展无人车的研发，在这一领域一举超过美国。这样就可以拿下人工智能的重要一个场景。无人驾驶的最大挑战仍然是图像识别。在我国农村山路崎岖不平，大城市车水马龙。要想识别出来当前车面前遇到的情况，并快速反应作出调整。难度非常大。需要不断完善算法和耗时耗力的测试精准度。同时也要求计算的芯片需要高性能高可用。这也带动了我国自研芯片的浪潮。

纵观计算机视觉发展到如今，已经不再是单纯一门理论研究，而是更多不同学科的交叉推进各个领域的研究。同样的，计算机视觉的研究可以更多的和其他产业合作。理论如果不应用于工业，那么就无法产生价值，就没有研究的意义。

现阶段，我国各大学校实验室都在视觉系统的建设贡献自己的一份力。

视觉系统的建设主要分为三个方面，计算理论层次，表达与算法层次和硬件实现层次。

计算理论层次要回答系统各个部分的计算目的与计算策略，亦即各部分的输入输出是什么，之间的关系是什么变换或者具有何种约束。Marr 对视觉系统的总的输入输出关系规定了一个总的目标，即输入二维图像，输出是由二维图像"重建"（reconstruction）出来的三维物体的位置与形状。Marr 认为，视觉系统的任务是对环境中三维物体进行识别、定位与运动分析，但这仅仅是一种对视觉行为（behavior）目的性定义，而非计算理论层次上的目的性定义。三维物体千差万别，应存在一种计算层次上的一般性"目的"描述，达到了这一"目的"，则不管是什么具体的物体，视觉任务均可完成。Marr 认为，这一"目的"，就是通过视觉系统，重建三维物体的形状和位置。而且若每个时刻都能做到这一点，那么运动分析也可以做到。对视觉系统的各个层次与模块，Marr 也初步给出了计算理论层次上的目标。

对于表达与算法层次，视觉系统的研究应给出各部分（模块）的输入、输出和内部的信息表达，以及实现计算理论所规定的目标的算法。算法与表达有关，不同的表达方式，完成同一计算的算法会不同，但 Marr 认为，算法与表达是比计算理论低一层次的问题，不同的表达与算法，在计算理论层次是可以相同的。

最后一个硬件层次，是要回答"如何用硬件实现以上的算法"。

区分上述三个不同层次，对于深刻理解计算机视觉与生物视觉系统，以及它们之间的关系是大有裨益的，例如，人的视觉系统与目前的计算机视觉系统在"硬件实现"层次上是截然不同的，前者是极为复杂的神经网络，而后者是目前使用的计算机，但它们可能在计算理论层次上具有完全相同的功能。

目前的计算机视觉的研究工作主要集中在前两个层次上，即计算理论，表达于算法层次，对于硬件实现，目前只有比较成熟的部分，如低层次处理中的噪声去除，边缘抽取；对于简单二维物体识别及简单场景下的视觉方法，已有专门芯片或其它并行处理体系结构方面的研究与试验产品；从系统上构造一般的视觉系统，虽有一些尝试，但一般并不成功。

视觉是一个信息处理过程。这个过程根据外部世界的图像产生对观察者有用的描述。这些描述依次由许多不同但固定的、每个都记录了外界的某方面特征的表象（representation）所构成或组合而成。一种新的表象之所以提高了一步是因为新的表象表达了某种信息，而这种信息将便于对信息作进一步解释。按这种逻辑来思考可得到这样的结论：即在对数据作进一步解释以前我们需要关于被观察物体的某些信息，这就是所谓的本征图像。

然而，数据进入我们的眼睛是要以光线为媒介的。灰度图像中至少包含关于照明情况、观察者相对于物体位置的信息。因此，按 Marr 的方法首先要解决的问题是如何把这些因素分解开。他认为低层视觉（即视觉处理的第一阶段）的目的就是要分清哪些变化是由哪些因素引起的。大体上来说这个过程要经过两个步骤来完成：第一步是获得表示图像中变化和结构的表象。这包括检测灰度的变化、表示和分析局部的几何结构、以及检测照明的效应等处理。第一步得到的结果被称为初始简图（primal sketch）的表象；第二步对初始简图进行一系列运算得到能反映可见表面几何特征的表象，这种表象被称为二维半（2.5 Dimension，意味着部分的，不完整的三维信息描述）简图或本征图像。"计算"的语言讲，就是重建三维物体在观察者为中心的坐标系下的三维形状和位置，当人眼或摄像机观察周围环境时，观察者对环境中的三维物体最初是以自身的坐标系来描述的。另外，我们只能观察到周围环境物体的一部分（另一部分是物体的背面或被其它的物体所遮挡的部分）。这样，重建的结果是以观察者坐标系下描述的部分三维物体形状，称为二维半描述。这一阶段包括由立体视觉运算提取深度信息，根据灰度影调、纹理等信息恢复表面方向，由运动视觉运算获取表面形状和空间关系信息等。

这些运算的结果都集成到本征图像这个中间表象层次。因为这个中间表象已经从原始的图像中去除了许多的多义性，是纯粹地表示了物体表面的特征，其中包括光照、反射率、方向、距离等。根据本征图像表示的这些信息可以可靠地把图像分成有明确含义的区域（这称为分割），从而可得到比线条、区域、形状等更为高层的描述。这个层次的处理称为中层视觉处理（intermediate processing）。

二维半描述是不够的，事实上，从各种不同角度去观察物体，观察到的形状都是不完整的。人脑中不可能存有同一物体的从所有可能的观察角度看到的物体形象，以用来与所谓的物体的二维半描述进行匹配与比较。因此，二维半描述必须进一步处理以得到物体的完整的三维描述，而且必须是物体本身某一固定坐标系下的描述，这一阶段称为第三阶段，即三维阶段。Marr 视觉理论中的这一阶段的表象层次是三维模型，它适用于物体的识别。这个层次的处理涉及物体，并且要依靠和应用与领域有关的先验知识来构成对景物的描述，因此被称为高层视觉处理。

显然，工业界对于视觉方面是要求高，且需求多。因为要在实际生活中，克服许许多多的干扰和不确定因素，创造出能稳定运行的视觉系统，其难度不亚于古埃及人造金字塔，秦始皇造长城，以及古巴比伦造空中花园。

但是在目前国家的大力支持下，视觉的领域的进步将会是飞速，现如今已有非常多优秀的公司，也就是AI独角兽，如旷视科技，商汤，大疆无人机等等。其中旷视今年在视觉方面的比赛coco是一口气拿下7个冠军。相信在这些优秀公司驱动下，工业界的视觉发展将会不可限量。