# 计算机视觉领域发展现状和关键技术

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。计算机视觉的研究更多的是和其他产业的研究互相推动，互相促进发展。