

Computer Vision—计算机视觉

(一)

(本文为总体概述，后续具体深入。原因的话，记录分享和本人水平成正比嘛~~~算是刚入门~~~)

总的来说，CV（计算机视觉）领域因为深度学习的引入，方兴未艾、蓬勃发展……

先贴一个某深圳互联网科技公司该方向实习生招聘职责|要求：

该岗位主要职责包括但不限于：

- 1、负责图像/视频相关算法的研究与开发，包括深度学习以及常用机器学习方法在机器视觉中的应用；
- 2、负责计算机视觉相关的技术研发工作，包括但不限于：图像检测、图像分类、图像分割、图像跟踪、视频语义分析、人脸识别与分析、车辆与人员的检测识别与跟踪、图像/视频搜索、页面分析与自动合成、OCR等算法与系统研发领域；
- 3、负责深度学习、图像理解、机器学习等前沿技术的研发储备和平台建设，结合未来实际应用场景，提供技术解决方案。

（其实各大公司此方向要求大多如此）

传统 CV

传统 CV 说白了就是**特征+分类器**。

单张图片—>大量图片—>视频。

单张图片

依赖于图片的像素矩阵表示，有了矩阵，很多方法如二值化，阈值化，色彩均化，滤波（模糊/光滑），形态学开集和闭集，以及联通区域划分，图像金字塔等都可以做。这些处理方法在实际当中应用广泛，例如许多图片应用的滤镜/增强/变形效果，以及图片压缩。

大量图片

图片分类以及基于内容的检索等实际需求。传统机器视觉的方法或者说套路是，先针对问题和对特征的具体要求（例如希望特征具有旋转不变性等）设计一些特征抽取方法，有了特征之后，就能拿去喂给一些机器学习算法做分类等其他后续工作。

那么特征提取就是重点，如颜色直方图，只是一个简单的统计描述，而其他常用的特征，如 Harris 角点，FAST 角点，图像梯度以及 HOG, LBP(local binary pattern), SIFT 特征以及其变体 SUFT 和 ORB, haar 等。经过精心设计，并且涉及一些诸如利用积分图像来优化计算的技巧，是传统计算机视觉的重要成果。

有了特征之后，就能解决：边缘检测/轮廓提取，图像分割，图片分类，人脸识别，图片拼接(image stitching)，视觉测距（包括单目视觉测距(Monocular Visual Odometry)和立体视觉测距(Stereo Visual Odometry)）等经典问题。

视频

视频无非就是连续（帧）的图片，图片处理技术就自然而然地扩展到了视频上来（一般的话，图片是一帧数据，视频处理就是在图片处理的外面加一个大循环，不断的处理一帧又一帧的图片就行。当然图片或视频的获取方式有不少，USB、HDMI 或者是直接从内存读取等等，不同方式下图片和视频读取的方式也有差别，但大同小异），但是视频具有动态特征也有其特殊的对待，如均值漂移、GMM（混合高斯模型）等背景建模的方法，以及利用光流法等实现物体的跟踪。

（附：传统方法使用滑动窗口的框架，把一张图分解成几百万个不同位置不同尺度的子窗口，针对每一个窗口使用分类器判断是否包含目标物体。传统方法针对不同的类别的物体，一般会设计不同的特征和分类算法，比如人脸检测的经典算法是 Harr 特征+Adaboosting 分类器；行人检测的经典算法是 HOG(histogram of gradients) + Support Vector Machine；一般性物体的检测的话是 HOG 的特征加上 DPM(deformable part model)的算法)

之后，随着人工智能技术的崛起，深度学习大潮席卷了包括计算机视觉和自然语言处理在内的诸多领域，许多在之前无法有效解决的问题，如图像语义分析、图片/视频内容描述、图片/视频问答等开始得到解决。许多新的问题，如图片生成（GAN）、图片风格迁移、图像预测、图像|视频搜索、OCR 等开始出现，并在深度学习的火炬下显示出无限可能，而传统的方法，那些人工设计特征的时代，似乎正在渐行渐远。

基于机器学习|深度学习的 CV

没错，这是 CV 领域以后的发展方向，包括但不限于：图像检测、图像分类、图像分割、图像跟踪、视频语义分析、人脸识别与分析、车辆与人员的检测识别与跟踪、图像/视频搜索、页面分析与自动合成、OCR 等算法与系统研发领域。

最主要的原因是，**精度**，深度学习可以做到传统方法无法企及的精度。另外，深度学习其他原因|优点：**通用性强，特征迁移能力强，工程框架统一等。**

深度学习技术框架是一颗树形结构：

训练平台是树根，如 tensorflow、theano 等；

模型是树干，是深度学习的重点。典型成果有 AlexNet、VGGNet、GoogleNet、ResNet 等。学术界—怎么提高模型精度，工业界—还要考虑怎么把模型做得更快，更小。

(核心) 任务|方向是树枝，**检测、识别、分割、特征点定位、序列学习**等5大任务。任何计算机视觉的具体应用都可以由这五个任务组合而成，如人脸识别，涉及到人脸检测、特征点定位，特征提取&验证。这就包含了检测、识别、特征点定位三个部分。

下图给出：方向—网络对应（常用）

图像识别：

Alexnet, VGGnet, GoogleNet, ResNet, RetinaNet

目标检测：

Fast-rcnn, faster-rcnn, Yolo, Retina-Net

图像分割：

FCN, Mask-Rcnn

目标跟踪：

GOTURN, ECO

图像生成：

GAN, WGAN

光流：

FlowNet

视频分割：

Segnet

以上为借鉴各路大神观点的总结，以下为个人不成熟见解：

- 1、 CV+机器|深度学习还是十分有趣的，应用也广，像你从摄像头前经过，瞬间就可以知道你的性别、年龄、情绪等等；路上行人、车辆|车牌检测，哪里有老人跌倒，哪里发生车祸，犯人逃到哪里，不需要人为通知，第一时间就可以知道并处理；目标跟踪可以一直监控你；风格迁移-滤镜，拍照神器；图像生成（GAN），比如可以用来直播换脸，想象一下你正在看的 20 多岁的少女直播镜头前其实是一个东北大汉……像自动驾驶（视频分割、图像预测等等）、机器人视觉啥的就更不用说了（相信未来自动驾驶可以用视觉完全替代掉激光雷达，不然，车怕是卖不出去，雷达贵~~~）
- 2、 CV+深度学习的内容确实太多（废话，钱给的多），学习、入门感觉难度不小，但是还是有方法的。比如：实践+理论，具体就是大致了解整个框架情况后，直接找个具体实例，像人脸识别、行人|车辆检测都不错，仅用传统方法和传统+深度学习的都需要学习。这些例子里就是在用那些看上去很牛X的理论，各种图像预处理、特征提取|降维、训练|分类器、后处理，各种像检测|识别|分割|特征点定位等这些方法、数据集、模型、框架平台等等，通过 1-2 个这样具体的例子，大致看懂功能——复现——读源码|理解原理，里面程序有搞不懂的就去查，直到搞懂。然后基本就入门有感觉了，卯足劲儿再去深入学习，后面内容还多着呢。（建议还是不要一开始就抓着一堆看着就头疼的理论读来读去，效率比较低。其实现在这一行，大多都是这样学比较好吧。还有，想到一定要做到！）
- 3、 本人也是刚入门吧，对这一块儿很感兴趣，还有很长路要走，滚去学习了……

（推荐一些大神更详细的文章|资源，很不错：）

1. <https://www.leiphone.com/news/201712/97BM1bEQ6DXEv8ke.html>
2. <https://zhuanlan.zhihu.com/p/21533690>
3. <https://blog.csdn.net/x1kz18nkbqg/article/details/78700447>