

140 | 什么是计算机视觉？

2018-08-24 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:11 大小 3.29M



在之前的一系列分享中，我们详细讲解了人工智能核心技术中的搜索、推荐系统、计算广告以及自然语言处理和文本处理技术。从今天开始，我们来分享专栏里人工智能核心技术模块的最后一部分内容：[计算机视觉技术](#)。

可以说，计算机视觉技术是人工智能技术的核心方向，特别是深度学习技术在计算机视觉中的应用，在最近五六年的人工智能浪潮中担当了先锋者的角色。甚至可以说，如果没有深度学习技术在过去几年对计算机视觉一些核心领域的推动和促进，就很可能没有这一波的人工智能技术浪潮。

我们可以这么来看待人工智能技术，它可以说是利用计算机技术来对人的感官，例如视觉、听觉、触觉以及思维进行模拟，从而建立起逻辑推断等智能才具备的能力。其中，计算机视觉技术无疑是至关重要的，也是非常困难的。

今天我会带你先来看看究竟什么是计算机视觉，以及这个方向的研究都需要解决哪些核心问题。

计算机视觉的定义

关于计算机视觉（Computer Vision, CV），有两种人们普遍接受的定义。

第一种定义认为，计算机视觉是从数字图像（Digital Images）中提取信息。这些信息可以是图像中的物品识别（Identification）、导航系统的位置测量（Space Measurement）以及增强现实（Augmented Reality）的应用。

计算机视觉的第二种定义主要是从应用的角度出发，认为计算机视觉是为了构建可以理解数字图像内容的算法，从而有多种应用。

那到底什么是计算机视觉呢？主要解决哪些问题？我们可以拿人类视觉的主要功能来做类比，就比较容易理解了。

当人类面对一个现实中的场景时，我们有一个感官器官来收集信号，那就是“眼睛”。眼睛收集的原始信号转换为人可以处理的信息之后，这些信息就来到了“大脑”这一个人类信息处理中心，进行分析和处理。

这个过程中最主要的一个处理模块，就是对信号产生“语义”（Semantic）解释或者进行逻辑上的理解。比如，当我们看到一个公园的一角以后，需要识别这个场景里的桥梁、水、树等物体，并且在头脑中形成这些物体的概念。可以说，这就是人类视觉系统的一个简单的框架，眼睛收集信息，大脑处理信息。

那么，在整体的框架上，计算机视觉其实就是希望模仿人类的视觉系统构架。输入依然是一个现实中的场景，但是我们需要借助其他的**感知仪器**（Sensing Device）来从中**获取原始信息**。最常见的感知仪器包括照相机、摄像机以及现在广泛普及的手机摄像头。从这些感知仪器中获取了最初级的信息之后，计算机视觉的“大脑”就是计算机。这里的“计算机”其实是指**计算机算法，通过算法理解原始数据，构建语义信息**。

这么理解起来，计算机视觉技术好像挺简单的。就像 1966 年，麻省理工大学的一个本科生想做这样一个暑期项目，并且认为这个项目可以在一个暑假里解决。这或许就是计算机视觉的一个起源了。但是，令人感慨的是，计算机视觉绝不是可以在一个假期内解决的项目，整

整半个多世纪已经过去了，计算机视觉依然有很多值得挑战的课题，也依然还在高速发展中。

计算机视觉的领域特点

了解了计算机视觉的定义之后，我们来进一步聊一聊这个领域的一些特点。

首先，**计算机视觉是一个“跨学科领域”**。正如刚才所说，对人类视觉的研究给计算机视觉带来了重要的启发。那这里就涉及到生物领域的研究，包括对人的眼睛以及视觉神经的研究。一方面，我们需要感知器来从现实世界中获取信息。那么，对于感知仪器来说，设备越是精确，就越能完整地捕捉外界世界的信息。这里就涉及到物理，特别是光学的研究。另一方面，人脑是处理所有信号并且形成语义概念的器官，理解人脑的信息处理机制就会对计算机视觉的发展有重要的作用，这就涉及到脑科学和认知科学等领域。

除此以外，计算机视觉毕竟是一个和计算机结合得很紧密的学科方向。因此，要想设计高效的计算机视觉算法，就必须和计算机科学的很多其他方向结合并借鉴，例如信息检索、计算机体系结构、机器学习等。

计算机视觉的另外一个特点，就是这个领域包含了很多非常深刻的困难问题。我们说，从计算机视觉被当作一个暑假项目到现在，五十多年已经过去了，这个领域依然在蓬勃发展着。时至今日，我们依然不能说计算机视觉是一个已经被完全解决的问题。

那计算机视觉任务“难”在哪里呢？我认为根本原因在于计算机视觉算法处理的输入，也就是数字化了的图像信息和我们需要理解的语义信息之间存在巨大的鸿沟。举例来说，一个 200 乘以 200 的 RGB 图像其实就是一个由 12 万个数字组成的矩阵，但是这个矩阵可能代表一个非常复杂的图像。从数字到具体的图像中的物体，再到去理解这个图像的语义，这中间有很长的距离。

一直以来，计算机视觉也在尝试去构造和逼近一些人类视觉系统的特点，但是困难重重。比如，人类视觉系统的反应很快。有实验表明，从一幅普通场景的图像中，人类只需要 150 毫秒就能够识别出里面的物体。另外，人类视觉系统的复杂性还来自于对世界认知的理解。例如，人可以依靠过去的记忆或者经验，还可以依靠其他外界知识，来对图像中的物体进行判断。这些都是计算机视觉系统难以企及的。

当然，在经历了半个世纪的研究之后，也有不少学者提出怀疑的观点，计算机视觉研究是否要对人类视觉系统进行完全的模仿呢？一种观点是，计算机视觉系统并不需要亦步亦趋地完

全照搬人类视觉系统，这可能也并不是一条切实有效的道路。有一种观点认为，计算机视觉系统可以从人类视觉或者其他领域得到灵感，但是究竟应该如何搭建一个有效的系统，还是需要开辟新的研究道路。

计算机视觉的应用

计算机视觉技术的领用非常广泛，可以说是深入到了普通人生活的方方面面。在这些应用中，除了我们日常比较容易接触到的，例如面部识别、光学字符识别（OCR）、电影特效、视觉搜索以外，还包括最近几年飞速兴起的自动驾驶、自动无人商店、虚拟现实、增强现实等等。

可以说计算机视觉的应用任务领域众多。近几年都受到深度学习的影响，绝大多数领域都得到了高速发展，但是依然需要领域知识来构建更加有效的模型。

小结

今天我和你讲了计算机视觉技术的一个，是我们计算机视觉基础知识系列的第一篇，帮助你对计算机视觉有一个最基本的概念性的了解。

一起来回顾下要点：第一，我们聊了什么是计算机视觉；第二，我们讲了计算机视觉的特点；第三，我们简要提及了一些计算机视觉的应用。

最后，给你留一个思考题，我们说计算机视觉的核心挑战是从数字到语义的理解，那么理解图像数据有什么特殊的地方吗？

欢迎你给我留言，和我一起讨论。


AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 139 | 数据科学团队怎么选择产品和项目？

下一篇 141 | 掌握计算机视觉任务的基础模型和操作

精选留言 (3)

 写留言



风的轨迹

2018-08-24

 1

我谈谈我的看法——相对于文本，图像的理解不管采用什么样的数据展现形式，其信息量要比文本大很多。深度学习之所以在图像领域取得很好的效果，主要也是因为深度学习经历了从低级到高级多次提取特征，未来如何多维度从图像挖掘出更多的信息，用于分析是理解图像的难点



周平

2019-01-24



单幅图像的数据量一般很大，数字化成0或1，非常抽象

展开 ▾



Sine: 极客...

2018-08-24



老师，初学者小白有个疑问：当公司需要某些功能时，是直接用现成的开源代码，再根据实际情况稍作修改？还是直接从头设计，独立开发呢？

比如中文分词，市面上有清华的THULAC，还有我个人使用感觉效果最好的HanLP，还有一大堆分词器，效果和速度都不同，如果是使用开源代码，哪有该如何选择呢？ ...

展开 ∨