

深度学习在图像理解和识别的应用

机器学习(Machine Learning, ML)是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。它是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域,它主要使用归纳、综合而不是演绎。而图像理解是为机器学习的预处理步骤,也可以视机器学习为实现图像处理高级形式的实现手段。

为什么需要机器学习?就当今的人脸识别功能,就需要机器通过数据处理来分析人脸图像,模拟人工的识别图像。而很久以前,这些图像数据都需要大多数人力通过底层特征来提取。

深度学习是机器学习目前最火的算法,是通过大量的简单神经元组成,每层的神经元接收更低层的神经元的输入,通过输入与输出之间的非线性关系,将低层特征组合成更高层的抽象表示,并发现观测数据的分布式特征,本质上还是原来的神经网络。因此通过深度学习,我们可以使用一个线性函数来判定图片:定义线性分类对样本分类;定义误差函数评价分类函数的结果的误差;找到优化函数的方法。

卷积神经网络是人工神经网络的一种,也是深度学习的常

用方法之一，已成为当前语音分析和图像识别领域的研究热点。它的权值共享网络结构使之更类似于生物神经网络，降低了网络模型的复杂度，减少了权值的数量。该优点在网络的输入是多维图像时表现的更为明显，使图像可以直接作为网络的输入，避免了传统识别算法中复杂的特征提取和数据重建过程。卷积网络是为识别二维形状而特殊设计的一个多层感知器，这种网络结构对平移、比例缩放、倾斜或者其他形式的变形具有高度不变性。卷积神经网络需要大量卷积盒子，每个盒子提取不同特征，比如横向、纵向、颜色等。其中，池化层是把图像缩小，4K 图片缩小到原来的一半，但是内容不会改变。

通过深度学习可以更好的得到数据的特征，同时由于模型的层次、参数很多，**capacity** 足够，因此，模型有能力表示大规模数据，所以对于图像、语音这种特征不明显（需要手工设计且很多没有直观物理含义）的问题，能够在大规模训练数据上取得更好的效果。当然，深度学习本事也不是完美的，也不是解决世间任何 **ML** 问题的利器，不应该被放大到一个无所不能的长度。

深度学习目前仍有大量工作需要研究。重点还是从机器学习的领域借鉴一些可以在深度学习使用的方法，特别是降维领域。如果深度学习能够有重大突破，那必定能给人工智能领域带来质的飞跃。将来机器一定也能变得如同人一样拥有判断、处理的能力。

