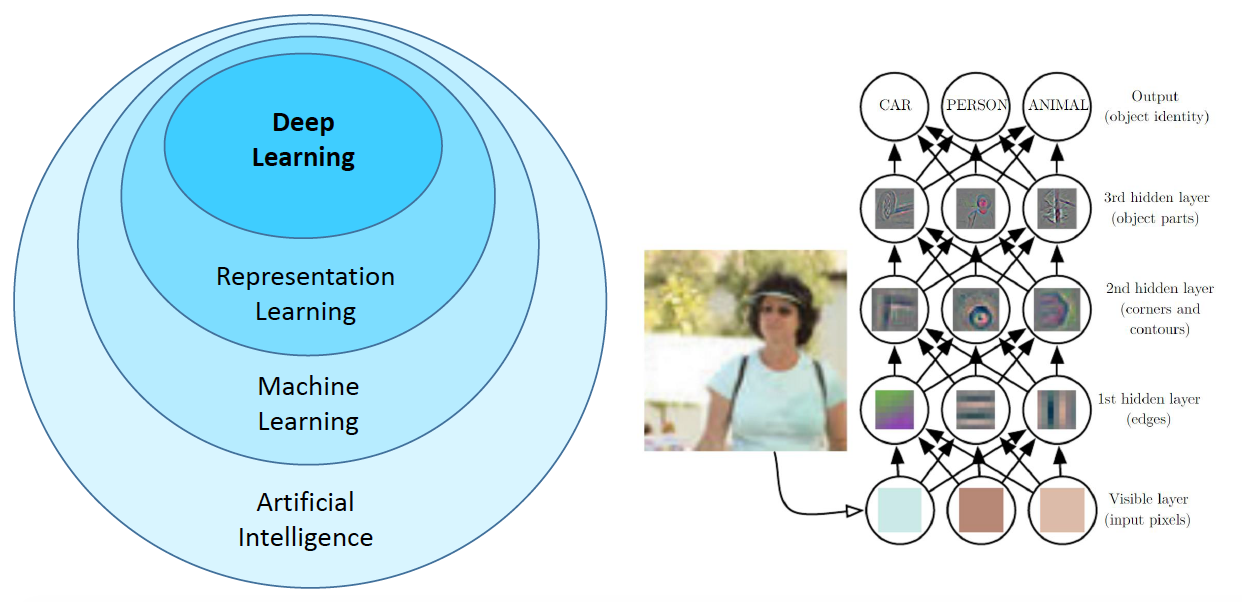
这次session主要是给大家介绍一些有意思的深度学习应用，由于时间问题，不会涉及到很多理论知识，所以大家可以轻松的听，感兴趣的话，可以带着好奇和疑问进一步去查阅资料或者之后跟我讨论。

首先会简单的介绍一下深度学习，然后主要是介绍它在计算机视觉以及自然语言处理方面的应用，最后跟大家分享一份行业报告。

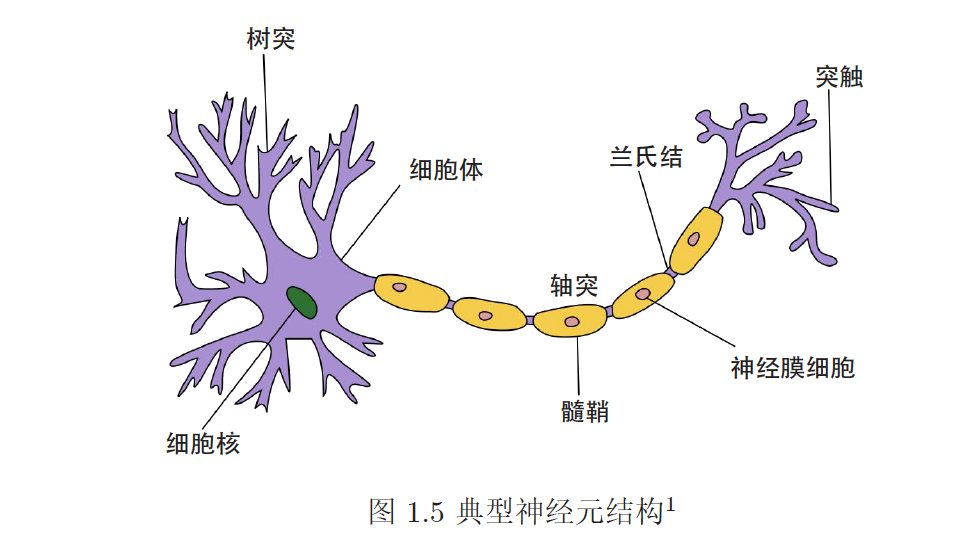


我们常常听到说AI，深度学习，机器学习，有人知道它们的关系是怎样的吗？

深度学习其实只是AI的一部分，但最近十来年的快速发展，也使它成为非常重要的一部分。它属于表征学习，表征学习就是能自动的将输入信息转化为有效的特征，并提高最终模型的性能。“深度”的目的是多层的特征转换，把原始数据

变成为更高层次、更抽象的表示。它跟传统的机器学习最大的区别是不需要手工的设计特征，而是通过网络训练一层一层的自动提取。

深度学习，也称为深度神经网络。我们都知道人类的智能行为都和大脑活动有关。人脑神经系统是一个非常复杂的组织，包含近860 亿个神经元，每个神经元有上千个突触和其它神经元相连接。这些神经元和它们之间的连接形成巨大的复杂网络，其中神经连接的总长度可达数千公里。



受到人脑神经系统的启发，早期的神经科学家构造了一种模仿人脑神经系统的数学模型，称为人工神经网络，简称神经网络。人工神经网络与生物神经元类似，由多个节点（人工神经元）互相连接而成，可以用来对数据之间的复杂关系进行建模。不同节点之间的连接被赋予了不同的权重，每个权重代表了一个节点对另一个节点的影响大小。来自其他节点的信息经过其相应的权重综合计算，输入到一个激活函数中可以使当前节点得到一个新的活性值（兴奋或抑制）。

其实，人工神经网络模型早在1943年就被提出了（1943 年∼ 1969 年，是神经网络发展的第一个高潮期），到1969 年∼ 1983 年，为神经网络发展的第一个低谷期（缺陷：第一个是感知器无法处理“异或”回路问题；第二个是当时的计算机无法支持处理大型神经网络所需要计算能力）。在此期间，神经网络的研究处于长年停滞及低潮状态。第三阶段为1983 年∼1995 年，为神经网络发展的第二个高潮期。这个时期中，反向传播算法重新激发了人们对神经网络的兴趣。第四个阶段为1995∼2006年，在此期间，支持向量机和其他更简单的方法（例如线性分类器）在机器学习领域的流行度逐渐超过了神经网络。从2006年，Hinton等人采用“预训练+ 精调”的方式有效地解决深度神经网络训练难的问题。到2012年，深度神经网络在语音识别和图像分类等任务上取得了巨大的成功，以神经网络为基础的深度学习迅速崛起。随着大规模并行计算以及GPU 设备的普及，计算机的计算能力得以大幅提高。此外，可供机器学习的数据规模也越来越大。所以深度学习迎来了第三次高潮。

目前比较流行的深度学习开发框架主要是TensorFlow和Pytorch，TensorFlow是Google开发的，PyTorch是Facebook、 NVIDIA、 Twitter 等公司开发维护的深度学习框架。目前的趋势是是用PyTorch的越来越多，因为与TensorFlow静态计算图相比，PyTorch的计算图是动态的，可以根据需要实时改变计算图。

计算机视觉

目前，卷积神经网络已经成为计算机视觉领域的主流模型，卷积神经网络是受生物学上感受野机制启发而提出的。AlexNet是第一个现代深度卷积网络模型，可以说是深度学习技术在图像分类上真正突破的开端，在AlexNet 之后，出现了很多优秀的卷积网络，可以达到上千层。

目标识别

说到目标识别，大家会想到哪些具体的应用？安检的人证匹配，长泰的人脸识别机器人，还有比如交通方面的车型识别，安防方面的人群计数，跨境追踪等等都是目标识别的具体应用。这些都离我们很近其实。

高分辨率重建

不知道大家最近有去电影院看《海上钢琴师》吗？它是21年前的电影通过4K修复技术再次上映。这个4K修复技术其实就采用了先进的图像超分辨率技术。超分辨率重建是将一系列低分辨率的图像来得到一幅高分辨率的图像过程。在医疗图像分析、生物特征识别、视频监控与安全等实际场景中也有着广泛的应用。

图像描述

Image Caption，就是从图片中自动生成一段描述性文字，有点类似于我们小时候做过的“看图说话”，十分有趣。对于人来说，Image Caption是简单而自然的一件事，但对于机器来说，这项任务却充满了挑战性。原因在于机器不仅要能检测出图像中的物体，而且要理解物体之间的相互关系，最后还要用合理的语言表达出来。这里其实有两部分，一部分是前面我们说的目标识别，另一部分就是要将词组组成句子，涉及到自然语言处理。

实时的姿态估计

姿态估计的目标是基于图像／视频重建人的关节和肢干（定位人体解剖学上的关键点（如肘部、腕部等）或部位）。姿态估计在[机器人视觉](https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8%E4%BA%BA%E8%A7%86%E8%A7%89/3321008" \t "_blank)（如机械臂抓取）

、AR/VR领域、动作跟踪等很多领域都有应用。（DeepIM：2018年新的深度神经网络对物体的6D姿态（3D位置和3D方向））

机器人

首先我们来看一段视频，思考一下可能会用到我们前面提到过的哪些深度学习技术？这里涉及到一个概念叫机器人视觉，是指不仅要把视觉信息作为输入，而且还要对这些信息进行处理，进而提取出有用的信息提供给机器人。机器人要拥有3D视觉，能处理三维空间里的三维物体问题，并且具有复杂算法，支撑机器人对位置、动作、轨迹等复杂信息的捕捉，这必须要依赖人工智能和深度学习来完成。

自动驾驶

同样先看视频。我们看到首先，自动驾驶要能检测与驾驶相关的各种目标，包括标志线，交通标志，信号灯，其他车辆，行人，障碍物等等。将这些目标检测出来之后，通过一种方法将其合并到最终的决策。

自动驾驶任务当然非常具有挑战性。场景分割，目标识别这些计算机视觉任务本身在现实应用中还存在一些难题。

图像自动着色

在图像着色任务中,我们的目标是在给定灰度输入图像的情况下生成彩色图像。

以假乱真

风格迁移

艺术作品的风格迁移到真实图片中去

给定内容图像（第一行左边图像）以及风格图像（左下角图像）可以生成特定风格下的混合图像。网络多次运算后，人眼很难判断出该图像是否为梵高或者毕加索的真迹。

Deep Dream

谷歌用来

大多时候我们是根据给定的数据和标签，去训练和调整网络的参数。不过也有时候，我们是固定网络的参数，根据某个损失函数调整输入数据，例如在图像风格迁移里，根据内容损失函数和风格损失函数调整合成的图片。

对于常见的图片分类模型，输入一张图片，网络中的每个tensor会输出相应的响应值，值越大说明这个tensor越“喜欢”这张图片。比如输入一张狗的图片，网络中用于识别和分类狗的tensor就会输出较大的响应值。把优化目标设为最大化某个tensor的响应值，以此来调整输入图片，这就是DeepDream的原理。随着tensor所在的层数变深，DeepDream优化出来的图形也更加复杂。

自然语言处理

卷积神经网络主要用于图像处理，它的信息传递是单向的，难以处理时序数据。而像语言，语音，视频分析等，有个很重要的特征就是时序数据。因此，当处理与时序数据相关的问题时，就需要一种能力更强的模型。循环神经网络是具有短期记忆能力的神经网络，但会存在梯度爆炸问题和记忆容量问题，也就是说随着记忆单元存储的内容越来越多，其丢失的信息也越来越多。而长短期记忆（Long Short-Term Memory，LSTM）网络可以改善这个问题，在RNN的基础上引入门控机制来控制信息的累积速度，包括有选择地加入新的信息，并有选择地遗忘之前累积的信息。