在前两次分享中，我们介绍了人工智能简介和机器学习的一些内容，本次我们将介绍深度学习的内容。

第 2 页

当前，深度学习在语音识别、图像识别等领域展现出了巨大的优势。

国际会议、期刊上也涌现出大量深度学习的文章，其中卷积神经网络(CNN)被引入很多领域。

很多国内外知名高科技公司都在深度学习领域加大了研究投入。

第 3 页 人脑识别的过程

深度学习是机器学习研究中的一个新的领域，其动机在于建立可以模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据，例如，图像、声音和文本。

总的来说，人的视觉系统的信息处理是分级的。从低级的 V1 区提取边缘特征，再到 V2 区的形状或者目标的部分等，再到更高层，整个目标、目标的行为等。也就是说高层的特征是低层特征的组合，从低层到高层的特征表示越来越抽象，越来越能表现语义或者意图。

而抽象层面越高，存在的可能猜测就越少，就越利于分类。

例如，在人脸识别的过程中，是由像素点组成了边缘，再由边缘构成了目标部分，如眼睛、鼻子、嘴巴等五官，最后再由五官构成了整个人脸。

第 4 页 生物神经元

神经元即神经细胞，是神经系统结构和机能的单位。它的基本作用是接收和传送信息。在人脑中，“神经元”是脑结构中的最小构成节点。单个神经细胞只有**兴奋 & 抑制**两种状态：

第 5 页 人工神经元

在人工神经网络中，一个人工记忆神经元的功能是求得输入向量与权向量的内积后，经一个激活函数得到一个标量结果。人工神经元的各种不同数学模型的主要区别在于采用了不同的变换函数，从而使神经元具有不同的信息处理特性 。

神经元的信息处理特性是决定人工神经网络整体性能的三大要素之一（人工神经网络的三要素，即人工神经元模型、人工神经元的联接方式、人工神经网络的训练与学习），因此变换函数的研究具有重要意义。神经元的变换函数反映了神经元输出与其激活状态之间的关系。最常用的变换函数有以下 4 种形式：阈值型变换函数、非线性变换函数、分段线性变换函数和概率型变换函数。

第 6 页 人工神经网络

要想模拟人脑的能力，单一的神经元是远远不够的，需要通过很多神经元一起协作来完成复杂的功能。这样通过一定的连接方式或信息传递方式进行协作的神经元可以看作是一个网络，就是神经网络。人工神经网络由神经元模型构成，这种由许多神经元组成的信息处理网络具有并行分布结构。**目前常用的神经网络结构有以下三种：**

前馈网络

前馈网络中各个神经元按接受信息的先后分为不同的组。每一组可以看作一个神经层。每一层中的神经元接受前一层神经元的输出，并输出到下一层神经元。整个网络中的信息是朝一个方向传播，没有反向的信息传播，可以用一个有向无环路图表示。

前馈网络可以看作一个函数，通过简单非线性函数的多次复合，实现输入空间到输出空间的复杂映射。这种网络结构简单，易于实现。

反馈网络（具有更强的计算和记忆能力。）

反馈网络中神经元不但可以接收其它神经元的信号，也可以接收自己的反馈信号。和前馈网络相比，反馈网络中的神经元具有记忆功能，在不同的时刻具有不同的状态。反馈神经网络中的信息传播可以是单向或双向传递，因此可用一个有向循环图或无向图来表示。

图网络：

实际应用中很多数据是图结构的数据，比如知识图谱、社交网络、分子（molecular ）网络等。前馈网络和反馈网络很难处理图结构的数据。

图网络是定义在图结构数据上的神经网络。图中每个节点都一个或一组神经元构成。节点之间的连接可以是有向的，也可以是无向的。每个节点可以收到来自相邻节点或自身的信息。

第 7 页 人工神经网络

人工神经网络主要由大量的神经元以及它们之间的有向连接构成。

人工神经网络的三要素：即**人工神经元模型、人工神经元的联接方式、人工神经网络的训练与学习**。因此我们在学习中可以重点考虑以下三方面：

- 神经元的激活规则： 主要是指神经元输入到输出之间的映射关系，一般为非线性函数。
- 网络的拓扑结构： 不同神经元之间的连接关系。
- 学习算法： 通过训练数据来学习神经网络的参数。

第 8 页 人工神经网络与深度学习

总的来说：

- 深度学习的概念源于人工神经网络的研究，是机器学习的分支；
- 机器学习的发展经历了浅层学习和深度学习两次浪潮

- ✓ **浅层学习**，通常包含一层或两层的非线性特征变换，可以看成是具有一层隐含层或者没有隐含层的结构。大多数传统的机器学习和信号处理技术，都是利用浅层结构的架构。例如高斯混合模型(GMMs)、支持向量机(SVMs)等都是浅层结构。
- ✓ **深度学习**，深度只是一个表面意思，代表着这一整套学习机制有着复杂而且深刻的逻辑在里面，并不是说按照一个刻度分为“浅度学习”、“中度学习”、“深度学习”。深层网络的定义也是与时俱进的，从 AlexNet 到 ResNet，一百多层的网络也是随处可见，之前是三层以上已经是深层，近来 5 层之内也仅仅被看做浅层

第 9 页

接下来，我们重点介绍一下卷积神经网络 CNN

第 10 页

卷积神经网络（CNN）是人工神经网络的一种，是多层感知机（MLP）的一个变种模型，它从生物学概念中演化而来。

Hubel(胡贝尔)和 Wiesel(威塞尔)早期对猫的视觉皮层的研究中得知在视觉皮层存在一种细胞的复杂分布，这些细胞对于外界的输入局部很敏感，它们被称为**“感受野”（细胞）**。这些细胞就像一些滤波器一样，够更好地挖掘出自然图像中的目标的空间关系信息。

视觉皮层存在两类相关的细胞，**S 细胞**（Simple Cell）和 **C 细胞**（Complex Cell）。S 细胞在自身的感受野内最大限度地对图像中类似边缘模式的刺激做出响应，而 C 细胞具有更大的感受野，它可以对图像中产生刺激的模式的空间位置进行精准定位。

CNN 的工作原理受其启发，架构上更接近生物学意义上的神经网络，更适合处理二维的图像数据，在人脸识别、医疗影像和无人驾驶等计算机视觉领域应用广泛，是推动深度学习迅速发展最主要的动力之一。

第 11 页

下面我们将介绍卷积神经网络里面几个重要的概念:

- **局部感受野（local receptive fields）**
- **共享权重（shared weights）**
- **卷积(Convolutional)**
- **池化（pooling）**

第 12 页 局部感受野（Local Receptive Fields）

- 图像的空间联系是局部的，就像人通过局部的感受野去感受外界图像一样，**每个神经元只感受局部的图像区域**，然后在更高层，将这些感受不同局部的神经元综合起来就可以得到全局的信息了。

- CNN 中相邻层之间是部分连接，也就是某个神经单元的感知区域来自于上层的部分神经单元。
- 卷积神经网络可以通过局部感知野降低参数数目。一般认为人对外界的认知是从局部到全局的，而图像的空间联系也是局部的像素联系较为紧密，而距离较远的像素相关性则较弱。因而，每个神经元其实没有必要对全局图像进行感知，只需要对局部进行感知，然后在更高层将局部的信息综合起来就得到了全局的信息。
- 视觉皮层的神经元就是局部接受信息的（即这些神经元只响应某些特定区域的刺激）。如下图所示：左图为全连接，右图为局部连接。
- 在上右图中，假如每个神经元只和 10×10 个像素值相连，那么权值数据为 1000000×100 个参数，减少为原来的万分之一。而那 10×10 个像素值对应的 10×10 个参数，其实就相当于卷积操作。

第 13 页 共享权重（Shared Weights）

但其实这样的话参数仍然过多，那么就可以引入权值共享。在上面的局部连接中，每个神经元都对应 100 个参数，一共 1000000 个神经元，如果这 1000000 个神经元的 100 个参数都是相等的，那么参数数目就变为 100 了。

怎么理解权值共享呢？我们可以这 100 个参数（也就是卷积操作）看成是提取特征的方式，该方式与位置无关。

这其中隐含的原理则是：图像的一部分的统计特性与其他部分是一样的。这也意味着我们在这一部分学习的特征也能用在另一部分上，所以对于这个图像上的所有位置，我们都能使用同样的学习特征。更直观一些，当从一个大尺寸图像中随机选取一小块，比如说 8x8 作为样本，并且从这个小块样本中学习到了一些特征，这时我们可以把从这个 8x8 样本中学习到的特征作为探测器，应用到这个图像的任意地方中去。特别是，我们可以用从 8x8 样本中所学习到的特征跟原本的大尺寸图像作卷积，从而对这个大尺寸图像上的任一位置获得一个不同特征的激活值。

第 14 页 卷积层

卷积网络之所以叫做卷积网络，是因为这种前馈网络其中采用了**卷积**的数学操作。在卷积网络之前，一般的网络采用的是**矩阵乘法**的方式，前一层的每一个单元都对下一层每一个单元有影响。

卷积层的作用：提取图片每个小部分里具有的特征

- ✓ 假定有一个尺寸为 6*6 的图像，每一个像素点里都存储着图像的信息。
- ✓ 再定义一个**卷积核**（相当于权重），用来从图像中提取一定的特征。
- ✓ 卷积核与数字矩阵对应位相乘再相加，得到卷积层输出结果。

卷积核的取值在没有以往学习的经验下，可由函数随机生成，再逐步训练调整，当所有的像素点都至少被覆盖一次后，就可以产生一个卷积层的输出。

第 15 页 卷积层具体工作过程：

机器一开始并不知道要识别的部分具有哪些特征，是通过与不同的卷积核相作用得到的输出值，相互比较来判断哪一个卷积核最能表现该图片的特征——比如我们要识别图像中的某种特征（比如曲线），也就是说，这个卷积核要对这种曲线有很高的输出值，对其他形状（比如三角形）则输出较低。卷积层输出值越高，就说明匹配程度越高，越能表现该图片的特征。

卷积层具体工作过程：
比如我们设计的一个卷积核，想要识别出来右边的曲线：

现在我们用上面的卷积核，来识别这一只漫画老鼠

第 16 页

当机器识别到老鼠的屁股的时候，卷积核与真实区域数字矩阵作用后，输出较大：6600

而用同一个卷积核，来识别老鼠的耳朵的时候，输出则很小：0

我们就可以认为：现有的这个卷积核保存着曲线的特征，匹配识别出来了老鼠的屁股是曲线的。我们则还需要其他特征的卷积核，来匹配识别出来老鼠的其他部分。卷积层的作用其实就是通过不断的改变卷积核，来确定能初步表征图片特征的有用的卷积核是哪些，再得到与相应的卷积核相乘后的输出矩阵

第 17 页 特征图（Feature Map）

- 为了提取不同的特征，需要多个滤波器。每种滤波器的参数不一样，表示它提出输入图像的不同特征。这样每种滤波器去卷积图像就得到对图像的不同特征的反映，我们称之为 Feature Map。
- 100 种卷积核就有 100 个 Feature Map。这 100 个 Feature Map 就组成了一层神经元。

第 18 页 池化（pooling）

池化层的输入就是卷积层输出的原数据与相应的卷积核相乘后的输出矩阵

- 原理：根据图像局部相关的原理，图像某个邻域内只需要一个像素点就能表达整个区域的信息，也称为混合、下采样
- 作用：
    - 减少空间信息的大小, 提高了运算效率;
    - 减少空间信息也就意味着减少参数, 降低了 overfit 的风险;
    - 获得空间变换不变性(translation rotation scale invarance,平移旋转缩放的不变性);

池化层的目的：

为了减少训练参数的数量，降低卷积层输出的特征向量的维度

减小过拟合现象，只保留最有用的图片信息，减少噪声的传递

最常见的两种池化层的形式：

最大池化：选取指定区域内最大的一个数来代表整片区域

均值池化：选取指定区域内数值的平均值来代表整片区域

**在 4*4 的数字矩阵里，以步长 2*2 选取区域，比如上左将区域[1,2,3,4]中最大的值 4 池化输出；上右将区域[1,2,3,4]中平均值 5/2 池化输出

第 19 页 全连接层工作原理

- 卷积层和池化层的工作就是**提取特征**，并减少原始图像带来的参数。然而，为了生成最终的输出，需要应用全连接层来生成一个等于我们需要的类的数量的分类器。
- **全连接层的工作原理:**和之前的神经网络学习很类似，我们需要把池化层输出的张量重新切割成一些向量，乘上权重矩阵，加上偏置值，然后对其使用 ReLU 激活函数，用梯度下降法优化参数既可。

第 20 页 CNN 的结构

- CNN 的网络层分为

✓ **卷积层**，C*，特征提取层，得到特征图，目的是使原信号特征增强，并且降低噪音；

✓ **池化层**，S*，特征映射层，将 C*层多个像素变为一个像素，目的是在保留有用信息的同时，尽可能减少数据量

第 21 页 卷积网络结构

- **卷积网络是由卷积层、子采样层、全连接层交叉堆叠而成。**趋向于小卷积、大深度；趋向于全卷积

一般来说 CNN 具有卷积层，池化层和全连接层 FC（正如在常规神经网络中所见）,在池化层之前一般会有个激活函数，我们将堆叠这些层，形成一个完整的架构。

第 22 页 CNN 如何训练

- 网络初始化
- **第一阶段**，向前传播阶段：

  a）从样本集中取一个样本(X，Yp)，将 X 输入网络；

  b）计算相应的实际输出 Op。

- **第二阶段**，向后传播阶段

  a）算实际输出 Op 与相应的理想输出 Yp 的差；

  b）按极小化误差的方法反向传播调整权矩阵。

第 23 页 CNN 的优缺点

- **优点**
  - ✓ 对几何变换、形变、光照具有一定程度的不变性；
  - ✓ 特征提取和模式分类同时进行，并同时在训练中产生；
  - ✓ 局部感受野和权重共享可以减少网络的训练参数，使神经网络结构变得更简单，适应性更强。
- **缺点**
  - ✓ 需要大量具有类标号的训练样本。

第 24 页

近年来，CNN 的应用于多个领域，如：AlphaGo 的围棋大战、 图像生成、画风迁移和生成对抗样本等等

常用的常用的深度学习框架：可以进行 简易和快速的原型设计，自动梯度计算和无缝 CPU 和 GPU 切换。