深度神经网络

与神经网络的区别在隐藏层数量变多了（即网络变深了）

从宏观来说，机器学习有两个高潮的发展：浅层学习（Shallow Learning）和深度学习（Deep Learning, DL）

（1）第一个是在20世纪80年代末期，BP算法的发明，掀起了一波机器学习的热潮。因为这种算法可以让模型从大量训练样本中学习到一定的规律，这比基于人工规则的系统效果好。当时的人工网络基本上只含有一层隐含层，是浅层网络。20世纪90年代，SVM、Boost等各种浅层机器学习模型相继出现。

（2）2006年，加拿大多伦多大学的Geoffrey Hinton教授和他的学生发表了一篇文章，其主要观点是：多隐藏层的神经网络具有更好的特征学习能力，从此开启深度学习的浪潮。

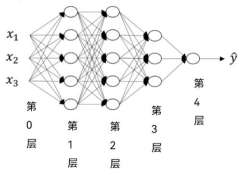
➤ 表示深度神经网络（Deep Neural Network，DNN）的元素

❑ 深度神经网络的层数：$L=4$

❑ 第l层的神经元数：$n^{[l]}$

❑ 第l层的激活函数：$a^{[l]}$

❑ 第l层的权重与偏置：$\boldsymbol{w}^{[l]}$，$\boldsymbol{b}^{[l]}$



➤ 参数与超参数

❑ 参数指与预测函数相关的 $\boldsymbol{w}^{[l]}$，$\boldsymbol{b}^{[l]}$

❑ 超参数指控制最终参数 $\boldsymbol{w}^{[l]}$，$\boldsymbol{b}^{[l]}$取值的其他参数，如隐含层层数与每层神经元数、激活函数，学习率、梯度下降的迭代次数等

负黑盒模型 ⇒不知晓隐层每行神经元在理解和提取什么特征

仅能通过走查参数调控其大致行为

LeNet → AlexNet → ZFNet → GoogleNet → VGG → ResNet → DenseNet → EfficientNet

LeNet：6 hidden，20c90s

AlexNet：8 hidden，2012，ImageNet，调整Alex网络框架

ZFNet：8 hidden，2013

GoogleNet：22 hidden，Inception模块降低模型参数

VGG：19 hidden，2014，加深网络，仍性能关键

ResNet：152 hidden，2015，ImageNet

DenseNet：2016

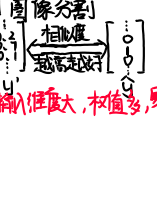
EfficientNet：2019

卷积神经网络

计算机视觉：让机器理解图片/视频中的内容 ⇒ 图像分类、目标检测、图像分割

全连接神经网络处理图像面临问题

3×100×100图像 ⟶ NN ⟶ n类评分向量(n×1维)



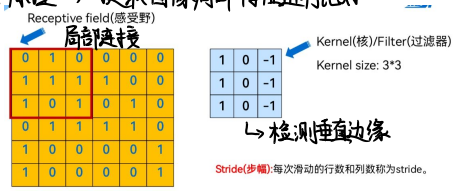
若采用全连接神经网络 输入1维度 3×100×100 ⇒输入维度大，权值多，易过拟合

图像平铺丢失掉像素点在空间联系

对图像移位、缩放和其它形状的变化不敏感

负卷积神经网络

工作原理 ⇒ 提取图像局部特征进行比对



二维卷积运算：将输入域卷积核做互相关运算，从最左上方开始，按从左往右、从上往下的顺序，二维核窗口一次在输入数组上滑动，当窗口滑动到某一位置时，窗口中的输入数组与核数组按元素乘积并求和，得到输出数组中相应位置的元素。

↳ 检测垂直边缘

输入 $n_h \times n_w$，有卷积核大小 $k_h \times k_w$，

padding大小 p_h, p_w，卷积步长 stride

输出大小 $\left(\frac{n_h-k_h+p_h}{stride}+1\right)\times\left(\frac{n_w-k_w+p_w}{stride}+1\right)$

padding：在输入图像四周填充元素以方便提取出输出形状

$p=\frac{k-1}{2}$ 时输入输出大小相同

每个卷积核只与输入数据的局部区域连接 ⇒ 大大减少网络参数，可以提取局部特征

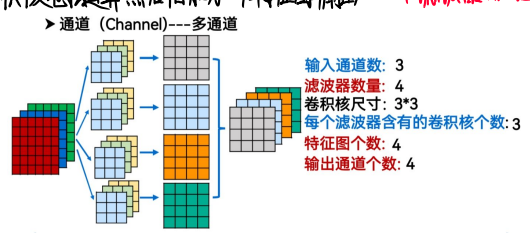
更深的网络可使特征图单个元素感受野更加广阔，从而捕获更大尺寸的特征

通道 channel 彩色图像三通道 ⇒ 多通道卷积 ⇒ 输出层 {黑白：单channel；彩色：三channel}

输入通道数为上层输出通道数，输出通道数取决于滤波器数目 运算时对通道和个卷积核运算然后相加为一个特征图输出

一个滤波器由一组卷积核组成

➤ 通道（Channel）---多通道



需训练参数量由卷积核个数和大小决定而与图片大小无关 ⇒ 每个滤波器还有一个偏置量

例：假设输入的图像的尺寸是100×100×3，卷积层使用10个3×3 ×3的卷积核，这一层需要训练的参数是多少？

$10\times(3\times3\times3+1)=280$

偏置

参数量远小于全连接神经网络

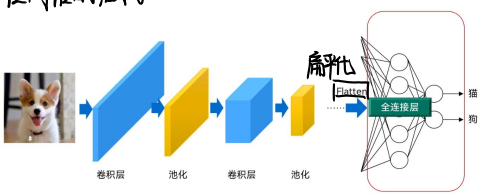
卷积层

参数共享/权值共享：控制参数数目，每个滤波器与上一层局部连接且所有局部连接都使用同样的参数，有效减小参数数目

池化 ⇒ 下采样，采样不改变图像中的物体；可缩小图像尺寸，减少最后连接层参数量，加快计算防止过拟合

最大池化/平均池化

卷积神经网络的结构

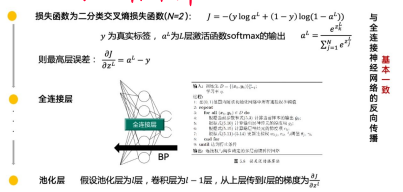


卷积神经网络的反向传播 更新全连接层/卷积层的权重；对池化层采样

CNN与全连接神经网络的反向传播：

➤ 最后的输出层到最后一个卷积层或池化层的误差计算方式是基本相同

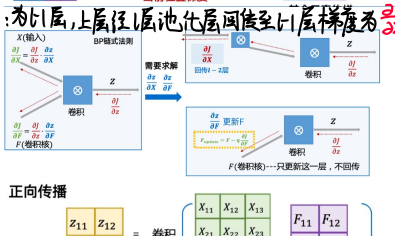
➤ 池化层到卷积层，卷积层到输出层（or池化层）的误差传播方式与全连接网络是大不相同



池化层：l-1层为卷积层，l层为池化层 当前位置梯度传递 卷积层激活函数偏导

$$\frac{\partial J}{\partial z^{l-1}}=\frac{\partial J}{\partial z^{l}}\cdot\frac{\partial z^{l}}{\partial a^{l-1}}\cdot\frac{\partial a^{l-1}}{\partial z^{l-1}}=\text{upsample}\left(\frac{\partial J}{\partial z^{l}}\right)\odot\frac{\partial a^{l-1}}{\partial z^{l-1}}$$

卷积层：为l-1层上层往l层池化层回传至l-1层梯度为 $\frac{\partial J}{\partial z^{l-1}}$，记为 $\frac{\partial J}{\partial z}$

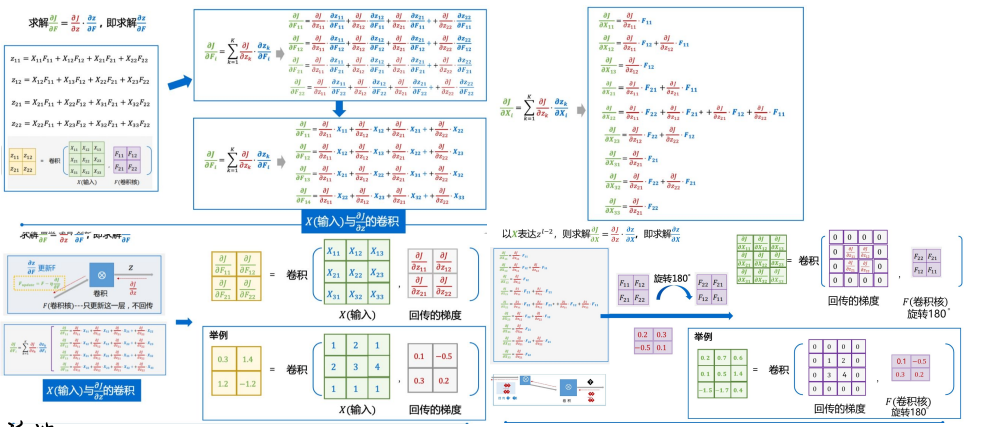


正向传播

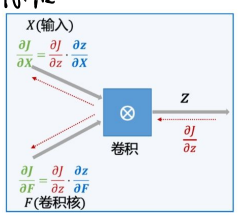
$$\begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} = \text{卷积}\left(\begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{bmatrix}, \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \right)$$

X(输入)　F(卷积核)

$$z_{11}=X_{11}F_{11}+X_{12}F_{12}+X_{21}F_{21}+X_{22}F_{22}$$
$$z_{12}=X_{12}F_{11}+X_{13}F_{12}+X_{22}F_{21}+X_{23}F_{22}$$
$$z_{21}=X_{21}F_{11}+X_{22}F_{12}+X_{31}F_{21}+X_{32}F_{22}$$
$$z_{22}=X_{22}F_{11}+X_{23}F_{12}+X_{32}F_{21}+X_{33}F_{22}$$



原始



$$\frac{\partial z}{\partial F}=\text{卷积}\left(X(\text{输入}),\ \text{回传的梯度}\right)$$
$$\frac{\partial z}{\partial X}=\text{卷积}\left(\text{回传的梯度},\ F(\text{卷积核})\ \text{旋转}180^{\circ}\right)$$

卷积神经网络经典结构 PPT P50-P78