招新题解 4:卷积神经网络理论

1.概念解释

1.卷积的数学公式、平移不变性、局部性

- 卷积的数学公式：指卷积运算在数学上的表达式，通常用来抓取图像对应的矩阵中的特征从而处理图像。我认为w定义了周围像素值对输出值产生的影响，a，b衡量了作用的距离。在深度学习中，卷积层实际上使用交叉相关运算（与卷积对称）来提取图像特征。
- • 二维交叉相关

$$y_{i,j} = \sum_{a=1}^{h} \sum_{b=1}^{w} w_{a,b} x_{i+a,j+b}$$

• 二维卷积

$$y_{i,j} = \sum_{a=1}^{h} \sum_{b=1}^{w} w_{-a,-b} x_{i+a,j+b}$$

• 由于对称性，在实际使用中没有区别

- 平移不变性：指当输入图像的某个特征发生平移时，卷积操作的输出不会发生变化，仍能识别出此特征。这是通过4维w的前两维不变，后两维只和相对位置有关实现的。这使得CNN能识别图像中不同位置的相同特征。
- 局部性：指的是卷积操作通过局部区域的卷积核进行特征提取，而不是全局性地处理整个图像。这是靠 a,b >定值时令w=0控制的

2.互相关运算，卷积核，1*1卷积核，输入输出通道，池化层，汇聚层，卷积层

- 互相关运算：这与卷积运算对称，但没有翻转卷积核。是卷积神经网络中常用的一个运算操作。
- 卷积核：是卷积中用于扫描输入图像的小矩阵，在一个循环中不变，随循环更行，提取特征。
- 1*1卷积核：这是一种特殊的卷积核，其尺寸为*1*1，不识别空间信息，只提取特征，用于融合、调整通道数，可以相当于一个全连接层

- 输入输出通道：指的是在卷积层中，输入数据和输出数据的除行列外的另一个维度，例如RGB就是红绿蓝3个通道。
- 池化层：通常用于每卷积层之后，分为max和avg，池化操作能减少图像的尺寸和特征图的大小，但不改变通道数，可以解决卷积层对空间位置的过度敏感，也可降低后续全连接层的计算复杂度，防止过拟合。
- 汇聚层：与池化层类似（这俩是一个东西吧）
- 卷积层：卷积神经网络中的核心层，参数包含核、填充、步长的大小和通道数，用于提取图像特征。

3.特征映射、感受野、填充、步幅、多输入输出通道

- 特征映射：是卷积操作的输出，表示从输入数据中提取的特征。
- 感受野：指每个输出特征所受输入区域的影响的范围大小。感受野越大，模型能感知的范围也越广。
- 填充：在卷积操作时，通常会对输入数据的边缘进行填充，可以更好的识别边缘特征，降低高和宽减小的速度，使神经网络更深。
- 步幅：是卷积核在输入数据上每次移动的步长，大的步长可以降低卷积层对位置的敏感。因为是做除法，可很大程度上影响输出特征图的尺寸，因此用于快速降低如今像素过高的网络图片
- 多输入输出通道：指的是在卷积层中，可以有多个输入通道（如RGB图像的3个颜色通道）和多个输出通道，不同的输出是对图像不同特征的分类识别。

2. 挑战问答环节

1.适合, 原因有：

- 首先分析共同点：股票走势图和猫猫照片的共同点在于它们都包含局部特征。图像中的局部特征（猫猫的胡须，耳朵）与时间序列中的局部特征（如价格的涨跌）在结构具有相似性和重复性，因此卷积层可以学习、处理这两种类型的数据。
- 像大量卖出时估价的跌停、稳定的上涨等（甚至了解了投资小知识）时间上的局部性，卷积核可以通过在时间轴上滑动的窗口来识别这些特征，类似于CNN在图像中识别纹理等空间特征。
- 平移不变性：分析预测股票时，卷积核可以在时间维度上共享参数，这减少了模型复杂度，使特征无论出现在时间序列的哪个位置，都能被检测到。
- 高效计算：卷积计算高效，适用于股票这种信息量大的数据（训练、测试数据也很丰富）

2.建模题
注：[] 表示向下取等

- 高：$h_o = [(h + 2p_h - k_h + s_h)/s_h]$
- 宽：$w_o = [(w + 2p_w - k_w + s_w)/s_w]$

- 通道数：c_o
- 输出尺寸：$h_o * w_o * c_o$

3.1×1卷积核的神秘作用是什么?

> 上文：不识别空间信息，只提取特征，用于融合、调整通道数，可以相当于一个全连接层

- 1×1卷积可以改变特征图的通道数，而不改变空间尺寸。用来调整网络层的通道数，以便于后续降低通道数，再进行大尺寸卷积，大幅减少参数和计算成本的同时增加神经网络深度。可谓是兼顾了减少计算量和提高模型复杂度的好模块
- 它可以在通道之间进行线性组合，依据通道间的相关性，增强或弱化某些特征，来提高整体的特征提取能力。
- 1×1卷积后通常跟随激活函数（如ReLU），在NiN中用来增加模型的非线性，增强拟合能力。

4.其他pooling

- L2池化：计算池化区域内特征的L2范数（x的平方之和开根），像是平均池化层和最大池化层的结合体，放大了较大值的影响，同时抑制较小值。
- 随机池化：是一种正则化池化方法，根据特征值的概率分布随机选择一个值输出，防止过拟合。
- 注意力池化：这个名字一下让我想到注意力机制，大概是每个特征都有一个注意力分数，通过线性回归得到一个权重，输出的是权重和特征值的加权平均，优点在于自主学习特征的重要性，更好的实现了平移不变性。
- 双线性池化：计算外积，把两个向量变为矩阵再降维处理（具体不太懂），能捕捉局部特征之间的关联。

4. 拓展挑战

对于任意矩阵 A（m*n）它的奇异值分解是: $A = U\Sigma V^T$
奇异值 量化了在对应方向上变换的"强度"。它告诉你，沿着右奇异向量 方向的输入，会被拉伸倍，然后映射到左奇异向量指明的输出方向上。

4.BatchNorm与Dropout的恩怨情仇

1. Dropout在训练时随机丢弃子网络，导致后续BatchNorm层计算出的均值和方差在每次迭代中都会剧烈波动，因为数据的统计特性被随机改变了。BN的本职工作是稳定数据分布，但Dropout却在不断破坏这种稳定性。可以将 Dropout>BatchNorm 的顺序改为 BatchNorm>Dropout，这样BN处理的是相对稳定的分布，受Dropout的影响较小。
2. ResNet成功的关键

- 残差结构：残差指输出函数与目标函数之间的差值，通过将残差与输入值叠加就能使输出不断靠近目标。将残差的概念引入CNN，让神经网络学习残差映射再于输入叠加，使训练超深网络成为可能
- 批量归一化：可以使小批量的输入数据分布在一个均值方差相对稳定的状态。这使反向传播时底层无需随顶层频繁调整，加速了训练收敛，避免了梯度消失
- 结构：使用了 1x1 > 3x3 > 1x1 的结构。中间的 3x3 卷积负责计算，两端的 1x1 卷积负责调整通道数先降维再升维，这大大减少了计算量和参数量，使得网络既能更深，又不会太慢

5. NiN网络性能分析

- vgg中的全连接层（具体是紧跟卷积层后的第一个全连接）参数个数为输入通道数*输出通道数*高*宽*（*可能是512*409677-vgg）
- NiN利用了1*1卷积层参数为 $c_o * c_i * k * k(k = 1)$,参数相对少许多；使用AvgPool代替了全连接层，pooling参数为0

1.显存大户：主要是中间激活值（模型的参数也有贡献）

- 中间激活值（输入输出的特征图）：pytorch为完成反向传播会在前向传播时保存下每一层的输入特征图和输出特征图，对于底部靠近数据的卷积层而言，图片像素很高，批量大小(B) × 通道数(C) × 高度(H) × 宽度(W)是很大的计算量
- （不是主要的）接在3*3卷积层后的第一和第二个1*1通道数分别为256和384（蛮大的），参数虽然相较vgg显著减少但仍占了显存

2.计算瓶颈：接在3*3卷积层后的第二个1*1卷积层的输入通道和输出通道都很大（256*384）

- 计算瓶颈是在向前传播过程中，卷积操作带来的计算量最大的层。
- 最后一个卷积层由于通道数多,kernel=3*3,计算量也比较大

3.宽带利用率

- 指内存带宽的利用效率，就是计算单元在等待数据从内存中加载的时间与计算时间的比例
  1x1卷积参数量少(毕竟k*k=1)，计算密度大（对每个位置都要进行输入通道数（比如256）次乘加运算），那么对于一张图片还要乘上高宽和输出通道的大小，但由于1x1卷积输入的像素（所有通道通用）和权重值（所有像素通用）都可以重复使用，不需要像大卷积核那样进行大量的内存访问，所以其计算效率较高。

6. CNN进化史

纵观整个CNN的发展历史，我认为模型的演变总体是朝着自主性越来越高，计算效率越来越高，准确度越来越好的大方向发展的。其历史背景离不开互联网的兴起与GPU带来的算力增长

Alexnet

- Lenet早在1990年代就已经引入了神经网络的概念，具备了卷积层，池化层和全连接层的架构，但直到2012年提出Alexnet才引发了深度学习热潮，这是因为在这20年间，数据（像素）的增加是百倍的，而算力的增长却是以万为单位的，这才让Alexnet这个参数量是 Lenet 1000倍的深度网络有了用武之地
- Alexnet的改进：
  1.使用ReLU激活函数 > 能有效缓解梯度消失问题，使网络可以更深，训练速度有所提升
  2.引入Dropout > 防止过拟合，模型更稳定
  3.用GPU训练，将训练时间从数月缩短到数天 > 网络结构更深，通道数增加，参数多
  4.由AvgPool变为MaxPool > 对特征位置微小变化的敏感度下降，平移不变性增强；可以确保最大激活值对应的显著特征能够传递到下一层；计算步骤变简单了
- 相比于MLP：还记得在03题中，我自己选择了数据中的相关特征让MLP学习，这对我这种没辍过学的人确实不太容易，自从有了CNN，我终于解放了大脑！MLP需要人工提取特征而CNN则可以自主学习特征，进行一体化训练，这是计算机方法论的进步，毕竟机器才是最懂机器的嘛

VGGNet

- VGGNet历史背景：在AlexNet证明了深度卷积网络的有效性后，做更大更深的网络成为大趋势
- VGGNet的改进：
  1.统一的小卷积核：使用连续的3×3卷积核替代大的卷积核，深但窄的窗口使学习到的细节更多，不停的高宽减半通道加倍...
  2.模块化思想：使用了重复的vgg块串联来构建更深的网络，方便又整洁，有了vgg11，16，19
- 改进效果：小卷积核大大减少了参数数量；更多卷积层带来更多ReLU激活函数模型非线性增强，拟合能力up；架构的模式化为后续研究提供了清晰的深度网络构建结构

NiN

- 历史背景：随着vggnet的提出，网络深度大大增加，vgg中的全连接层（具体是紧跟卷积层后的第一个全连接）参数个数为输入通道数*输出通道数*高*宽（512*409677-vgg），数以万计的参数带来的是巨大的内存与算力消耗以及过拟合的风险。而GPU计算速度极快但内存相对较慢的特点使全连接层的计算效率较差，如何改或替换它成了研究重点
- NiN核心改进：

1. 1×1卷积：大量使用1×1卷积来先升高（学习）后降低（分类）通道数和引入ReLU增强非线性；1*1卷积层参数为 $c_o * c_i * k * k(k = 1)$,参数相对少许多，计算效率高（解释见上文）
2. 全局平均池化：用全局平均池化替代全连接层，AvgPool参数为0，与softmax用于完成分类任务

- 改进效果：创新性的使用1*1卷积层，为GooLeNet，ResNet奠定了基础。参数大幅减少计算量下降计算效率提高；防止过拟合

GoogLeNet 多路径并行的开创者

随着网络的迭代，模型各方面性能不断提升，而其背后的数学原理却不为人知。哪条路径更好是未知的，作为成熟的大人要学会不做选择，统统拿来。goolenet开创了多路径并行的Inception模块，并联了1*1+不同窗口大小的卷积层*、*NiNblock和1*1卷积层，并通过炼丹为不同路径分配不同通道数，使模型可以按一定权重捕获不同尺度的特征信息，拟合能力大大提升的同时1*1卷积层使通道数快速下降，参数总量也有所减少，证明了极深网络的可行性

ResNet

随着网络深度的不断增加，模型出现了梯度消失、网络退化等问题，前者是由于越靠近最优点的导数越小，链式求导的连乘使底部梯度容易趋近0，更深的网络反而有更高的误差。解决优化困难成为CNN性能提高的关键。残差指输出函数与目标函数之间的差值，通过将残差与输入值叠加就能使输出不断靠近目标。将残差的概念引入CNN，让神经网络学习残差再于输入相加，就能解决网络退化的问题。ResNet引入跳跃连接，让网络学习残差映射；加入BatchNorm层，它对小批量的数据进行归一化，使该层的输入数据分布在一个均值方差相对稳定的状态。这使底层无需频繁调整，加速了训练收敛，避免了梯度消失。ResNet真正实现了网络的深度拓展，不再需要专门的初始化技巧和处处需要调整的学习率，简化了网络训练