

# 计算机视觉面试复习笔记（三）：深度学习基础面试常见问题总结

极市平台 2021-10-21 22:00:00 手机阅读 𐄂

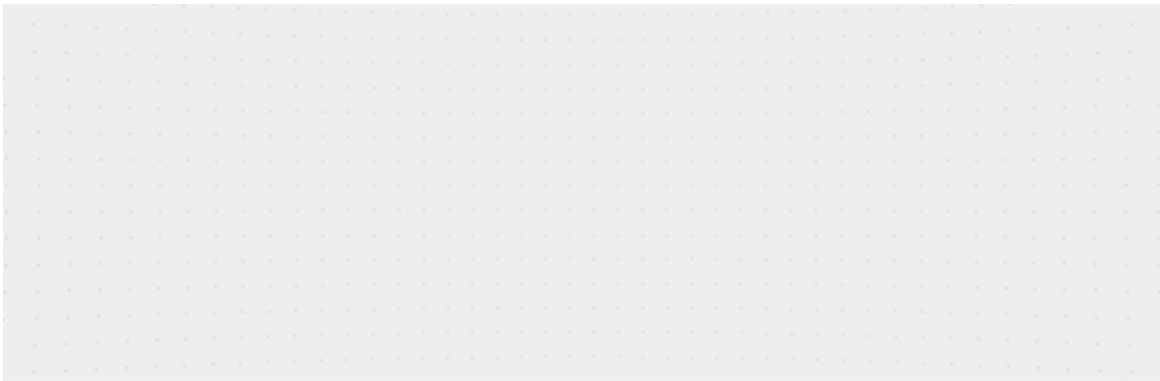
以下文章来源于AI约读社，作者那年当上博士前



AI约读社

一个热衷于AI学习和分享的平台，期待与您相遇。

↑ 点击蓝字 关注极市平台



作者 | 南山  
来源 | AI约读社  
编辑 | 极市平台

## 极市导读

本文介绍深度学习基础面经知识点的第三部分深度神经网络面临的挑战。 >>加入极市CV技术交流群，走在计算机视觉的最前沿

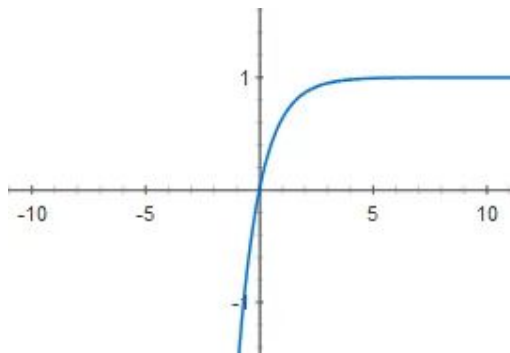
## 三、深度学习基础面试常见问题总结

### 1. 为什么需要做特征归一化、标准化？

- 1. 使不同量纲的特征处于同一数值量级，减少方差大的特征的影响，使模型更准确。
- 2. 加快学习算法的收敛速度。

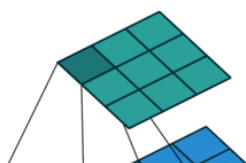
## 2. 常用的归一化和标准化的方法有哪些？

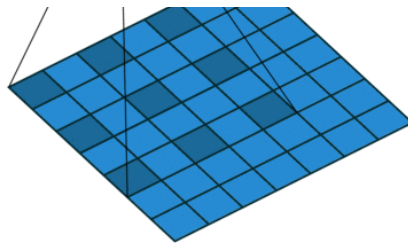
1. 线性归一化（min-max标准化）
2.  $x' = (x - \min(x)) / (\max(x) - \min(x))$ ，其中max是样本数据的最大值，min是样本数据的最小值
3. 适用于数值比较集中的情况，可使用经验值常量来代替max，min
4. 标准差归一化（z-score 0均值标准化）
5.  $x' = (x - \mu) / \sigma$ ，其中 $\mu$ 为所有样本的均值， $\sigma$ 为所有样本的标准差
6. 经过处理后符合标准正态分布，即均值为0，标准差为1
7. 非线性归一化
8. 使用非线性函数log、指数、正切等，如 $y = 1 - e^{-x}$ ，在 $x \in [0, 6]$ 变化较明显，用在数据分化比较大的场景



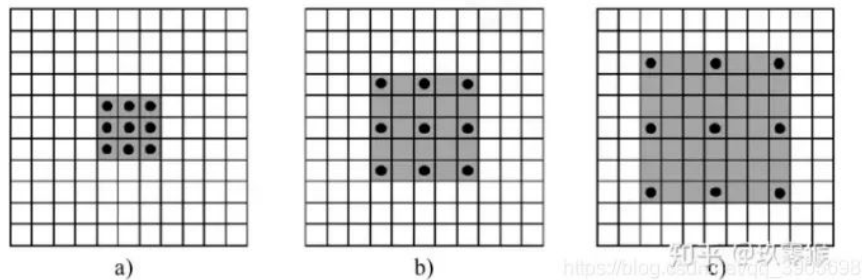
## 3. 介绍一下空洞卷积的原理和作用

空洞卷积也叫做膨胀卷积、扩张卷积，最初的提出是为了解决图像分割在用下采样（池化、卷积）增加感受野时带来的特征图缩小，后再上采样回去时造成的精度上的损失。空洞卷积通过引入了一个扩张率的超参数，该参数定义了卷积核处理数据时各值的间距。





以在增加感受野的同时保持特征图的尺寸不变,从而代替下采样和上采样, 通过调整扩张率得到不同的感受野大小:



a是普通的卷积过程(dilation rate = 1),卷积后的感受野为3

b是dilation rate = 2的空洞卷积,卷积后的感受野为5

c是dilation rate = 3的空洞卷积,卷积后的感受野为8可以这么说,普通卷积是空洞卷积的一种特殊情况

#### 4. 为什么线性回归使用mse作为损失函数?

在使用线性回归的时候的基本假设是噪声服从正态分布, 当噪声符合正态分布 $N(0, \sigma^2)$ 时, 因变量则符合正态分布 $N(ax(i)+b, \sigma^2)$ , 其中预测函数 $y=ax(i)+b$ 。这个结论可以由正态分布的概率密度函数得到。也就是说当噪声符合正态分布时, 其因变量必然也符合正态分布。因此, 我们使用mse的时候实际上是假设y服从正态分布的。

#### 5. 怎么判断模型是否过拟合, 有哪些防止过拟合的策略?

在构建模型的过程中, 通常会划分训练集、测试集。

当模型在训练集上精度很高, 在测试集上精度很差时, 模型过拟合; 当模型在训练集和测试集上精度都很差时, 模型欠拟合。

预防过拟合策略:

- (1) 增加训练数据: 获取更多数据, 也可以使用图像增强、增样等;
- (2) 使用合适的模型: 适当减少网络的层数、降低网络参数量;
- (3) Dropout: 随机抑制网络中一部分神经元, 使的每次训练都有一批神经元不参与模型训练;

- (4) L1、L2正则化：训练时限制权值的大小，增加惩罚机制，使得网络更稀疏；
- (5) 数据清洗：去除问题数据、错误标签和噪声数据；
- (6) 限制网络训练时间：在训练时将训练集和验证集损失分别输出，当训练集损失持续下降，而验证集损失不再下降时，网络就开始出现过拟合现象，此时就可以停止训练了；
- (7) 在网络中使用BN层（Batch Normalization）也可以一定程度上防止过拟合（原理不介绍，感兴趣的同学可以去了解）。

## 6. 除了SGD和Adam之外，你还知道哪些优化算法？

主要有三大类：

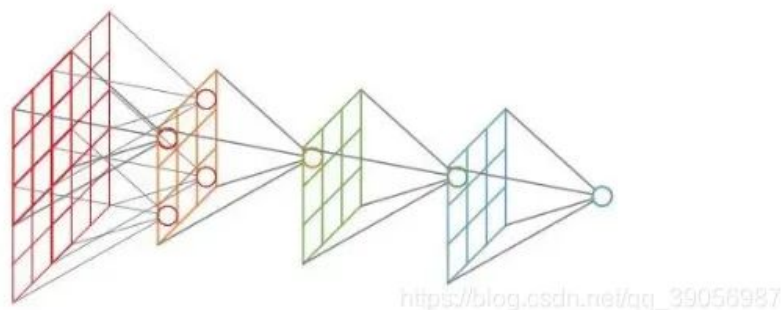
1. 基本梯度下降法，包括 GD，BGD，SGD；
2. 动量优化法，包括 Momentum，NAG 等；
3. 自适应学习率优化法，包括 Adam，AdaGrad，RMSProp 等

## 7. 阐述一下感受野的概念

感受野指的是卷积神经网络每一层输出的特征图上每个像素点映射回输入图像上的区域的大小，神经元感受野的范围越大表示其接触到的原始图像范围就越大，也就意味着它能学习更为全局，语义层次更高的特征信息，相反，范围越小则表示其所包含的特征越趋向局部和细节。因此感受野的范围可以用来大致判断每一层的抽象层次，并且我们可以很明显地知道网络越深，神经元的感受野越大。

卷积层的感受野大小与其之前层的卷积核尺寸和步长有关，与padding无关。

计算公式为： $F_{j-1} = K_j + (F_j - 1) * S_j$  (最后一层特征图的感受野大小是其计算卷积核大小)



## 8. 下采样的作用是什么？通常有哪些方式？

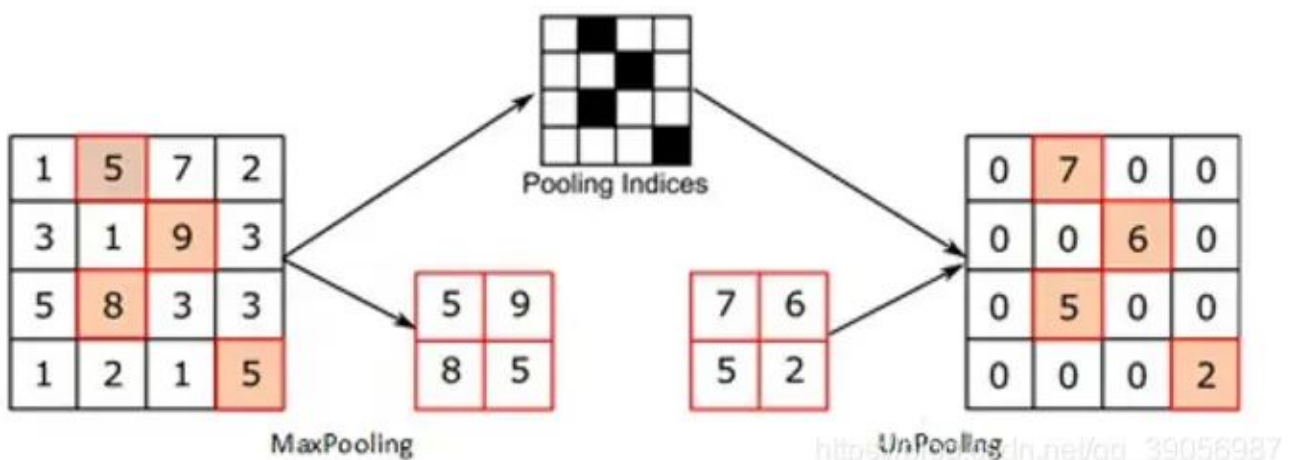
下采样层有两个作用，一是减少计算量，防止过拟合；二是增大感受野，使得后面的卷积核能够学到更加全局的信息。下采样的方式主要有两种：

- 采用stride为2的池化层，如Max-pooling和Average-pooling，目前通常使用Max-pooling，因为他计算简单而且能够更好的保留纹理特征；
- 采用stride为2的卷积层，下采样的过程是一个信息损失的过程，而池化层是不可学习的，用stride为2的可学习卷积层来代替pooling可以得到更好的效果，当然同时也增加了一定的计算量。

## 9. 上采样的原理和常用方式

在卷积神经网络中，由于输入图像通过卷积神经网络(CNN)提取特征后，输出的尺寸往往会变小，而有时我们需要将图像恢复到原来的尺寸以便进行进一步的计算(如图像的语义分割)，这个使图像由小分辨率映射到大分辨率的操作，叫做上采样，它的实现一般有三种方式：

- 插值，一般使用的是双线性插值，因为效果最好，虽然计算上比其他插值方式复杂，但是对于卷积计算可以说不值一提，其他插值方式还有最近邻插值、三线性插值等；
- 转置卷积又或是说反卷积，通过对输入feature map间隔填充0，再进行标准的卷积计算，可以使得输出feature map的尺寸比输入更大；
- Max Unpooling，在对称的max pooling位置记录最大值的索引位置，然后在unpooling阶段时将对应的值放置到原先最大值位置，其余位置补0；



## 10. 模型的参数量指的是什么？怎么计算？

参数量指的是网络中可学习变量的数量，包括卷积核的权重weights，批归一化（BN）的缩放系数 $\gamma$ ，偏移系数 $\beta$ ，有些没有BN的层可能有偏置bias，这些都是可学习的参数，即在模型训练开始前被赋予初值，在训练过程根据链式法则不断迭代更新，整个模型的参数量主要是由卷积核的权重weights的数量决定，参数量越大，则该结构对平台运行的内存要求越高。

参数量的计算方式：

$K_h \times K_w \times C_{in} \times C_{out}$  （Conv卷积网络）

$C_{in} \times C_{out}$  （FC全连接网络）

## 11. 模型的FLOPs（计算量）指的是什么？怎么计算？

神经网络的前向推理过程基本上都是乘累加计算，所以它的计算量也是指的前向推理过程中乘加运算的次数，通常用FLOPs来表示，即floating point operations(浮点运算数)。计算量越大，在同一平台上模型运行延时越长，尤其是在移动端/嵌入式这种资源受限的平台上想要达到实时性的要求就必须要求模型的计算量尽可能地低，但这个不是严格成正比关系，也跟具体算子的计算密集程度(即计算时间与IO时间占比)和该算子底层优化的程度有关。

FLOPs的计算方式：

$K_h \times K_w \times C_{in} \times C_{out} \times H \times W = \text{params} \times H \times W$  （Conv卷积网络）

$C_{in} \times C_{out}$  （FC全连接网络）

计算量 = 输出的feature map \* 当前层filter 即  $(H \times W \times C_{out}) \times (K \times K \times C_{in})$

## 12. 深度可分离卷积的概念和作用

深度可分离卷积将传统的卷积分两步进行，分别是depthwise和pointwise。首先按照通道进行计算按位相乘的计算，深度可分离卷积中的卷积核都是单通道的，输出不能改变feature map的通道数，此时通道数不变；然后依然得到将第一步的结果，使用1\*1的卷积核进行传统的卷积运算，此时通道数可以进行改变。



Focus模块，将W、H信息集中到通道空间，输入通道扩充了4倍，作用是可以使信息不丢失的情况下提高计算力。具体操作为把一张图片每隔一个像素拿到一个值，类似于邻近下采样，这样我们就拿到了4张图，4张图片互补，长的差不多，但信息没有丢失，拼接起来相当于RGB模式下变为12个通道，通道多少对计算量影响不大，但图像缩小，大大减少了计算量。

### 计算量的前后对比：

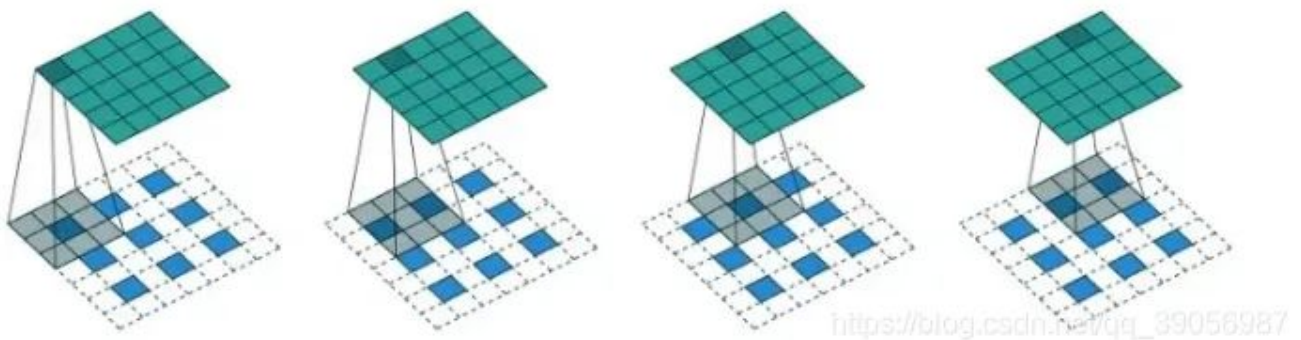
$$K_h \times K_w \times C_{in} \times C_{out} \times H \times W$$

$$\text{变成了 } K_h \times K_w \times C_{in} \times H \times W + 1 \times 1 \times C_{in} \times C_{out} \times H \times W$$

通过深度可分离卷积，当卷积核大小为3时，深度可分离卷积比传统卷积少8到9倍的计算量。

## 13. 转置卷积的原理

转置卷积又称反卷积(Deconvolution)，它和空洞卷积的思路正好相反，是为上采样而生，也应用于语义分割当中，而且他的计算也和空洞卷积正好相反，先对输入的feature map间隔补0，卷积核不变，然后使用标准的卷积进行计算，得到更大尺寸的feature map。

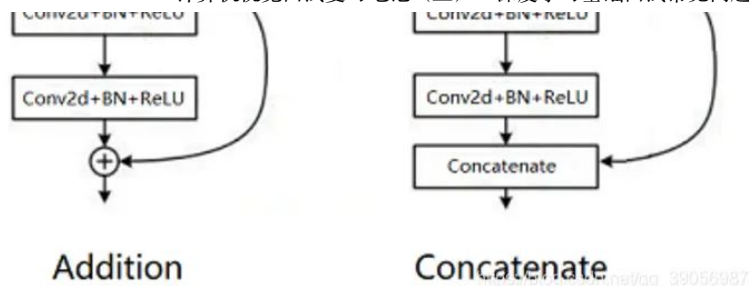


## 14. 神经网络中Addition / Concatenate区别是什么？

Addition和Concatenate分支操作统称为shortcut，Addition是在ResNet中提出，两个相同维度的feature map相同位置点的值直接相加，得到新的相同维度feature map，这个操作可以融合之前的特征，增加信息的表达，Concatenate操作是在Inception中首次使用，被DenseNet发扬光大，和addition不同的是，它只要求两个feature map的HW相同，通道数可以不同，然后两个feature map在通道上直接拼接，得到一个更大的feature map，它保留了一些原始的特征，增加了特征的数量，使得有效的信息流继续向后传递。

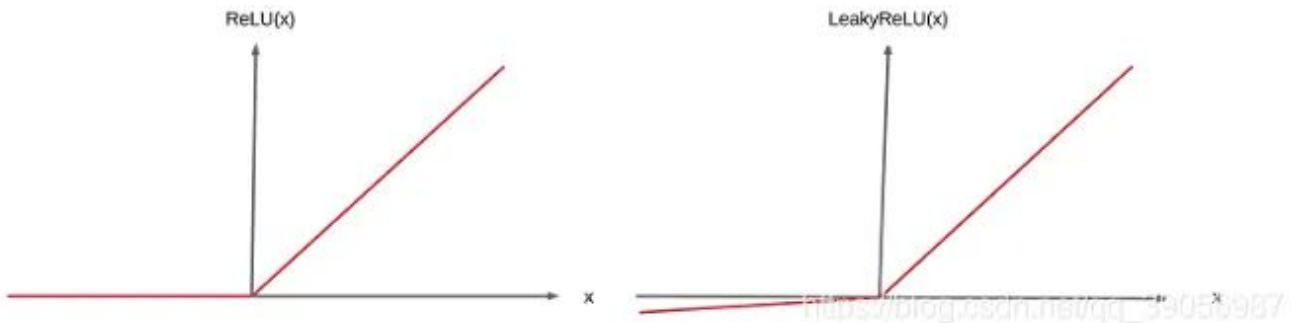






### 15. 你知道哪些常用的激活函数？

**ReLU系列：** ReLU、ReLU6和leaky ReLU；ReLU6与ReLU相比也只是在正向部分多了个阈值，大于6的值等于6，而leaky ReLU和ReLU正向部分一样，都是大于0等于原始值，但负向部分却是等于原始值的1/10，浮点运算的话乘个0.1就好了。



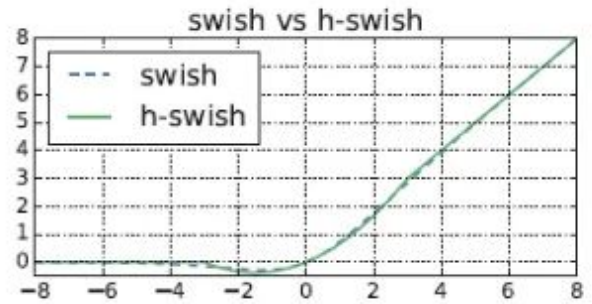
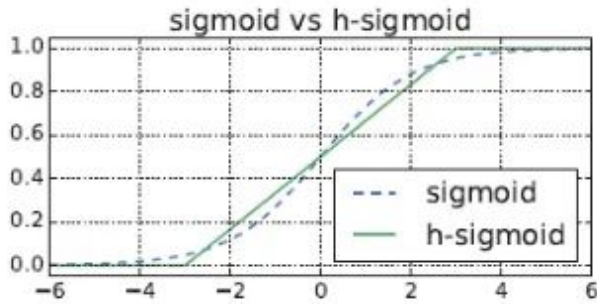
- ReLU函数的优点：
- 解决了梯度消失的问题；
- 计算速度和收敛速度非常快；
- ReLU函数的缺点：
- 低维特征向高维转换时会部分丢失；
- 均值为非零；

**Sigmoid系列：** Sigmoid、swish、h-sigmoid、h-swish；sigmoid对低性能的硬件来说非常不友好，因为涉及到大量的exp指数运算和除法运算，于是有研究人员针对此专门设计了近似的硬件友好的函数h-sigmoid和h-swish函数，这里的h指的就是hardware的意思：



$$\text{Hsigmoid}(x) = \max(0, \min(1, \frac{x+1}{2}))$$

$$\text{Hswish} = x \frac{\text{ReLU6}(x+3)}{6}$$



Sigmoid函数的优点：

1. 输出为 0 到 1 之间的连续实值，此输出范围和概率范围一致，因此可以用概率的方式解释输出；
2. 将线性函数转变为非线性函数；

Sigmoid函数的缺点：

1. 幂运算相对来讲比较耗时；
2. 输出均值为非 0；容易出现梯度消失的问题；

**tanh**

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

tanh函数的优点：

Tanh 函数的导数比 Sigmoid 函数导数值更大、梯度变化更快，在训练过程中收敛速度更快；使得输出均值为 0，可以提高训练的效率；将线性函数转变为非线性函数；

tanh函数的缺点：幂运算相对来讲比较耗时；容易出现梯度消失；

## 16. 神经网络中1\*1卷积有什么作用？

降维，减少计算量；在ResNet模块中，先通过1\*1卷积对通道数进行降通道，再送入3\*3的卷积中，能够有效的减少神经网络的参数量和计算量；

升维；用最少的参数拓宽网络通道，通常在轻量级的网络中会用到，经过深度可分离卷积后，使用 $1 \times 1$ 卷积核增加通道的数量，例如mobilenet、shufflenet等；

实现跨通道的交互和信息整合；增强通道层面上特征融合的信息，在feature map尺度不变的情况下，实现通道升维、降维操作其实就是通道间信息的线性组合变化，也就是通道的信息交互整合的过程；

$1 \times 1$ 卷积核可以在保持feature map尺度（不损失分辨率）不变的情况下，大幅增加非线性特性（利用后接的非线性激活函数）。卷积神经网络中用 $1 \times 1$ 卷积有什么作用或者好处呢？

## 17. BN(Batch Normalization)的原理和作用是什么？

将一个batch的数据变换到均值为0、方差为1的正态分布上，从而使数据分布一致，每层的梯度不会随着网络结构的加深发生太大变化，从而避免发生梯度消失或者梯度爆炸，能够加快模型收敛，同时还有防止过拟合的效果。

### 实现过程

计算训练阶段mini\_batch数量激活函数前结果的均值和方差，然后对其进行归一化，最后对其进行缩放和平移。

### 作用

限制参数对隐层数据分布的影响，使其始终保持均值为0，方差为1的分布；

削弱了前层参数和后层参数之间的联系，使得当前层稍稍独立于其他层，加快收敛速度；有轻微的正则化效果。

## 18. 随机梯度下降相比全局梯度下降好处是什么？

当处理大量数据时，比如SSD或者faster-rcnn等目标检测算法，每个样本都有大量候选框参与训练，这时使用随机梯度下降法能够加快梯度的计算；

每次只随机选取一个样本来更新模型参数，因此每次的学习是非常快速的，并且可以进行在线更新。

## 22. 如果在网络初始化时给网络赋予0的权重，这个网络能正常训练嘛？

不能，因为初始化权重是0，每次传入的不同数据得到的结果是相同的。网络无法更新

## 19. 无监督学习方法有哪些？

强化学习、K-means 聚类、自编码、受限波尔兹曼机

## 20. 增大感受野的方法？

空洞卷积、池化操作、较大卷积核尺寸的卷积操作

## 21. 神经网络的正则化方法？/过拟合的解决方法？

数据增强(镜像对称、随机裁剪、旋转图像、剪切图像、局部弯曲图像、色彩转换)

early stopping(比较训练损失和验证损失曲线，验证损失最小即为最优迭代次数)

L2正则化(权重参数的平方和)

L1正则化(权重参数的绝对值之和)

dropout 正则化(设置keep\_pro参数随机让当前层神经元失活)

## 22. 梯度消失和梯度爆炸的原因是什么？

原因：激活函数的选择。

梯度消失：令 $\text{bias}=0$ ，则神经网络的输出结果等于各层权重参数的积再与输入数据集相乘，若参数值较小时，则权重参数呈指数级减小。

梯度爆炸：令 $\text{bias}=0$ ，则神经网络的输出结果等于各层权重参数的积再与输入数据集相乘，若参数值较大时，则权重参数呈指数级增长。

## 23. 深度学习为什么在计算机视觉领域这么好？

以目标检测为例，传统的计算机视觉方法需首先基于经验手动设计特征，然后使用分类器分类，这两个过程都是分开的。而深度学习里的卷积网络可实现对局部区域信息的提取，获得更高级的特征，当神经网络层数越多时，提取的特征会更抽象，将更有助于分类，同时神经网络将提取特征和分类融合在一个结构中。

## 24. 为什么神经网络种常用relu作为激活函数？

在前向传播和反向传播过程中，ReLU相比于Sigmoid等激活函数计算量小；

在反向传播过程中，Sigmoid函数存在饱和区，若激活值进入饱和区，则其梯度更新值非常小，导致出现梯度消失的现象。而ReLU没有饱和区，可避免此问题；

ReLU可令部分神经元输出为0，造成网络的稀疏性，减少前后层参数对当前层参数的影响，提升了模型的泛化性能；

## 25. 卷积层和全连接层的区别是什么？

卷积层是局部连接，所以提取的是局部信息；全连接层是全局连接，所以提取的是全局信息；

当卷积层的局部连接是全局连接时，全连接层是卷积层的特例；

## 26. 什么是正则化？L1正则化和L2正则化有什么区别？

所谓的正则化，就是在原来 Loss Function 的基础上，加了一些正则化项，或者叫做模型复杂度惩罚项，正则化机器学习中一种常用的技术，其主要目的是控制模型复杂度，减小过拟合。

两者的区别：

L1范式：它的优良性质是能产生稀疏性，导致  $W$  中许多项变成零。稀疏的解除了计算量上的好处之外，更重要的是更具有“可解释性”。

L2范式：使得模型的解偏向于范数较小的  $W$ ，通过限制  $W$  范数的大小实现了对模型空间的限制，从而在一定程度上避免了过拟合。不过岭回归并不具有产生稀疏解的能力，得到的系数仍然需要数据中的所有特征才能计算预测结果，从计算量上来说并没有得到改观。

## 27. 常用的模型压缩方式有哪些？

使用轻量型的特征提取网络，例如：mobileNet、shuffleNet、GhostNet系列等等；通道、层剪枝；模型蒸馏；模型量化。

如果觉得有用，就请分享到朋友圈吧！



极市平台

为计算机视觉开发者提供全流程算法开发训练平台，以及大咖技术分享、社区交流、竞...  
765篇原创内容

公众号

△点击卡片关注极市平台，获取最新CV干货

公众号后台回复“CVPR21检测”获取CVPR2021目标检测论文下载~

## 极市干货

神经网络：视觉神经网络模型优秀开源工作：timm库使用方法和最新代码解读

技术综述：综述：神经网络中 Normalization 的发展历程 | CNN轻量化模型及其设计原则综述

算法技巧（trick）：8点PyTorch提速技巧汇总 | 图像分类算法优化技巧



# CV技术社群邀请函 #



△长按添加极市小助手

添加极市小助手微信（ID：cvmart4）

备注：姓名-学校/公司-研究方向-城市（如：小极-北大-目标检测-深圳）

即可申请加入极市目标检测/图像分割/工业检测/人脸/医学影像/3D/SLAM/自动驾驶/超分辨率/姿态估计/ReID/GAN/图像增强/OCR/视频理解等技术交流群

每月大咖直播分享、真实项目需求对接、求职内推、算法竞赛、干货资讯汇总、与 10000+ 来自港科大、北大、清华、中科院、CMU、腾讯、百度等名校名企视觉开发者互动交流~

觉得有用麻烦给个在看啦~ ○

喜欢此内容的人还喜欢

我是如何从头开始写一篇顶级论文的

计算机视觉与机器学习



产品渲染练习（前半段）

小斯设计工作室



改进的双流卷积神经网络模型

人工智能感知信息处理算法研究院

