一位算法工程师从30+场秋招面试中总结出的超强面经——目标检测篇（含答案）

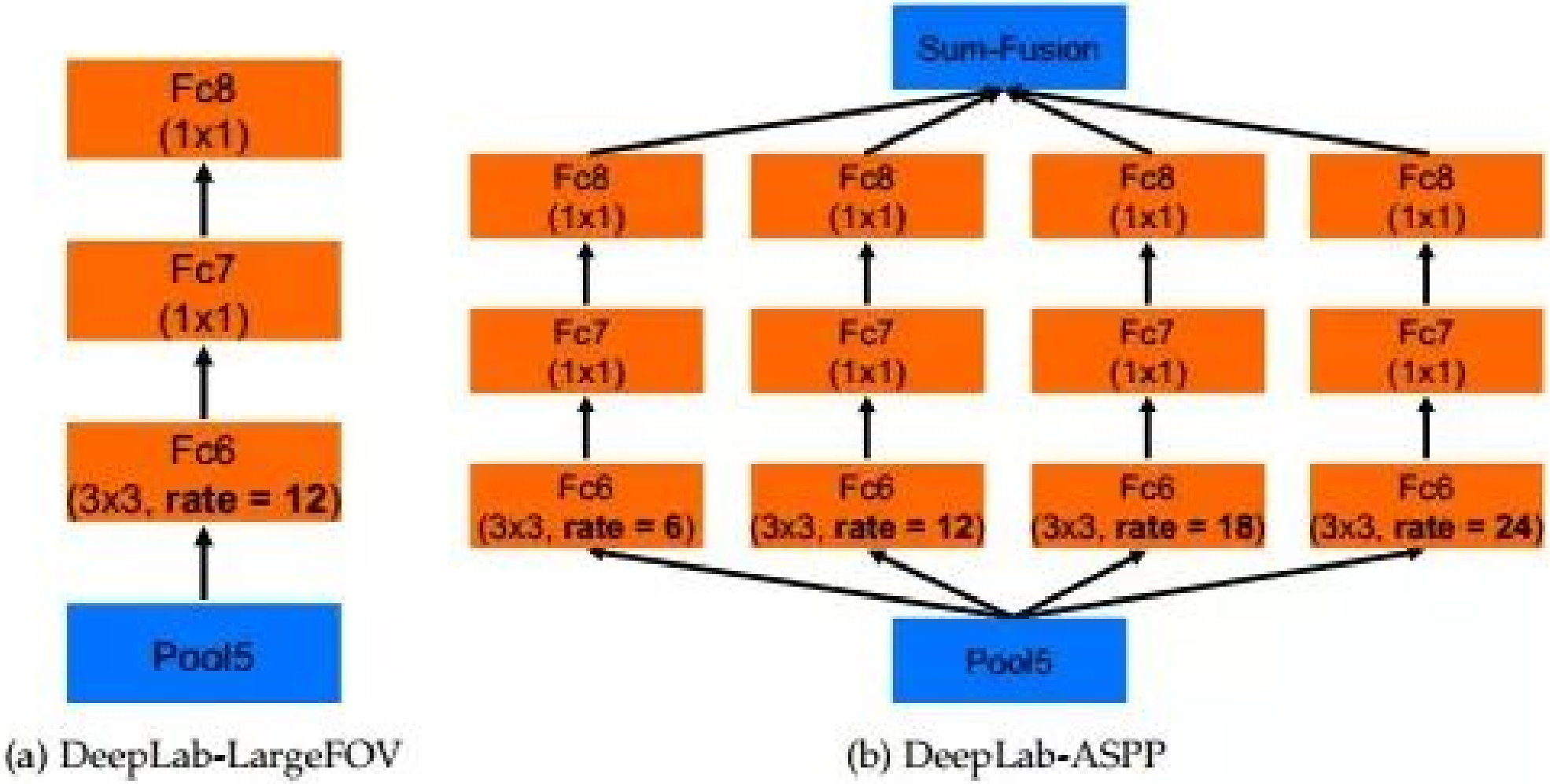
一.deeplab系列

1.简述Deeplab v1网络

DeepLab是结合了深度卷积神经网络（DCNNs）和概率图模型（DenseCRFs）的方法。在实验中发现DCNNs做语义分割时精准度不够的问题，根本原因是DCNNs的高级特征的平移不变性（即高层次特征映射，根源在于重复的池化和下采样）。针对信号下采样或池化降低分辨率，DeepLab是采用的atrous（带孔）算法扩展感受野，获取更多的上下文信息。另外，DeepLab 采用完全连接的条件随机场（CRF）提高模型捕获细节的能力。论文模型基于 VGG16，在 Titan GPU 上运行速度达到了 8FPS，全连接 CRF 平均推断需要 0.5s ，在 PASCAL VOC-2012 达到 71.6% IOU accuracy。

2.简述Deeplab v2网络

DeepLabv2 是相对于 DeepLabv1 基础上的优化。DeepLabv1 在三个方向努力解决，但是问题依然存在：特征分辨率的降低、物体存在多尺度，DCNN 的平移不变性。因 DCNN 连续池化和下采样造成分辨率降低，DeepLabv2 在最后几个最大池化层中去除下采样，取而代之的是使用空洞卷积，以更高的采样密度计算特征映射。物体存在多尺度的问题，DeepLabv1 中是用多个 MLP 结合多尺度特征解决，虽然可以提供系统的性能，但是增加特征计算量和存储空间。论文受到 Spatial Pyramid Pooling (SPP) 的启发，提出了一个类似的结构，在给定的输入上以不同采样率的空洞卷积并行采样，相当于以多个比例捕捉图像的上下文，称为 ASPP (atrous spatial pyramid pooling) 模块。



相比于DeepLab v1，deeplab v2在之前的基础上做了三个方面的贡献：一是使用Atrous Convolution 代替原来上采样的方法，比之前得到更高像素的score map，并且增加了感受野的大小；二是使用ASPP 代替原来对图像做预处理resize 的方法，使得输入图片可以具有任意尺度，而不影响神经网络中全连接层的输入大小；三是使用全连接的CRF，利用低层的细节信息对分类的局部特征进行优化。

论文模型基于 ResNet，在 NVidia Titan X GPU 上运行速度达到了 8FPS，全连接 CRF 平均推断需要 0.5s ，在耗时方面和 DeepLabv1 无差异，但在 PASCAL VOC-2012 达到 79.7 mIOU。

3.简述Deeplab v3网络相比于之前的v1和v2网络有哪些改进

①重新讨论了空洞卷积的使用，这让我们在级联模块和空间金字塔池化的框架下，能够获取更大的感受野从而获取多尺度信息。②改进了ASPP模块：由不同采样率的空洞卷积和BN层组成，我们尝试以级联或并行的方式布局模块。③讨论了一个重要问题：使用大采样率的3×3的空洞卷积，因为图像边界响应无法捕捉远距离信息，会退化为1×1的卷积, 我们建议将图像级特征融合到ASPP模块中。④阐述了训练细节并分享了训练经验。

介绍deeplabv3,画出backbone

DeepLab V3将空洞卷积应用在了级联模块，并且改进了ASPP模块。backbone还是resnet 101. 增强ASPP模块，复制resnet最后的block级联起来，加入BN。没有使用CRFs新的ASPP模块包括：一个1×1卷积和3个3×3的空洞卷积(采样率为(6,12,18))，每个卷积核都有256个且都有BN层；包含图像级特征image-level features(即全局平均池化Global Avearge Pooling)；所有分支得到的结果concate起来通过1×1卷积之后得到最终结果。

DeepLab V3采用atrous convolution的上采样滤波器提取稠密特征映射和去捕获大范围的上下文信息。具体来说，编码多尺度信息，提出的级联模块逐步翻倍的atrous rates，提出的atrous spatial pyramid pooling模块增强图像级的特征，探讨了多采样率和有效视场下的滤波器特性。实验结果表明，该模型在Pascalvoc 2012语义图像分割基准上比以前的DeppLab版本有了明显的改进，并取得了与其他先进模型相当的性能。

DeepLab V3的改进主要包括以下几方面：1）提出了更通用的框架，适用于任何网络

2）复制了ResNet最后的block，并级联起来3）在ASPP中使用BN层4）去掉了CRF。

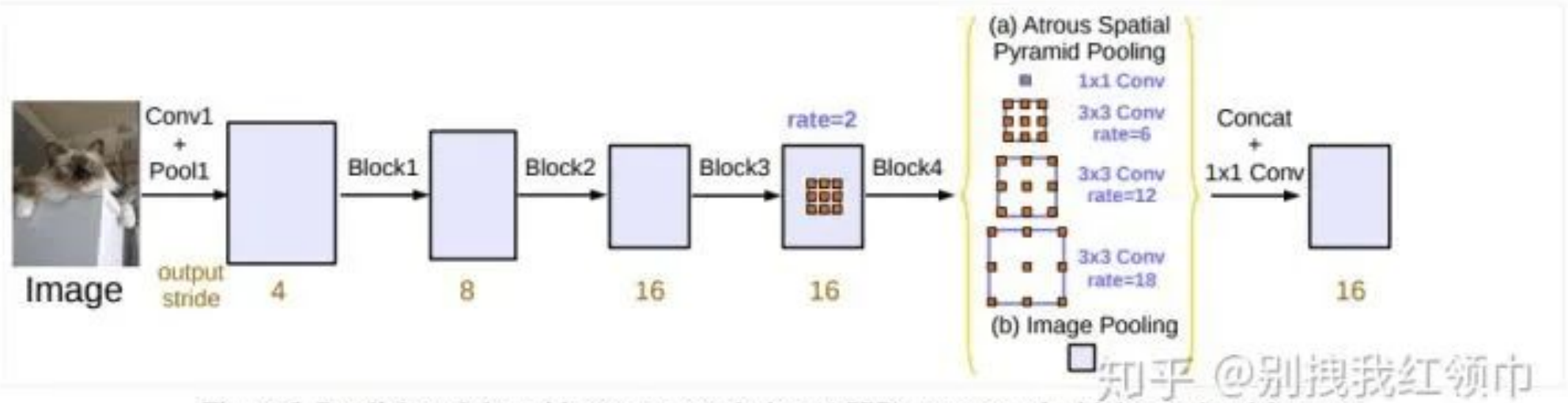


Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

deeplabv3的损失函数交叉熵损失函数

4.deeplabv3+系列

DeepLabv3 +采用编码器解码器结构扩展了DeepLabv3。编码器模块通过在多个尺度上应用atrous卷积来编码多尺度上下文信息，而简单但有效的解码器模块沿着对象边界细化分割结果。DeepLabv3+模型中使用ResNet-101作为网络主干。探索Xecption模型, 将depthwise separable convolution应用到ASPP和解码模块上(更快,更稳定)。

5.条件随机场(CRF)后处理的目的

CRF使像素级别的类别标签的多类别输出与底层图像信息（如像素间的相互关系）有关，这种结合尤其重要，这也是关注于局部细节的CNN所未能考虑到的。CRF 将图像中每个像素点所属的类别都看作一个变量xi，然后考虑任意两个变量之间的关系，建立一个完全图。

6.简要阐述一下UNet网络

UNet网络可以简单看为先下采样，经过不同程度的卷积，学习了深层次的特征，再经过上采样回复为原图大小，上采样用反卷积实现。输出类别数量的特征图，最后使用激活函数softmax将特征图转换为概率图，针对某个像素点，如输出是[0.1，0.9]，则判定这个像素点是第二类的概率更大。

网络结构可以看成3个部分：

①下采样：网络的红色箭头部分，池化实现

②上采样：网络的绿色箭头部分，反卷积实现

③最后层的softmax：在网络结构中，最后输出两张fiture maps后，其实在最后还要做一次softmax，将其转换为概率图。

7. 简述encode和decode思想

将一个input信息编码到一个压缩空间中 将一个压缩空间向量解码到一个原始空间中。

8. FCN与CNN最大的区别？

卷积层不再与FC层相连,而是加入一个全局池化层。

FCN是如何取代FC层的 FCN用卷积层代替FC层。

9.分割出来的结果通常会有不连续的情况，怎么处理？开运算闭运算

设定阈值，去掉阈值较小的连通集，和较小的空洞。

开运算 = 先腐蚀运算，再膨胀运算（看上去把细微连在一起的两块目标分开了）

开运算总结：（1）开运算能够除去孤立的小点，毛刺和小桥，而总的位置和形状不便。

（2）开运算是一个基于几何运算的滤波器。（3）结构元素大小的不同将导致滤波效果的不同。（4）不同的结构元素的选择导致了不同的分割，即提取出不同的特征。

闭运算 = 先膨胀运算，再腐蚀运算（看上去将两个细微连接的图块封闭在一起）

闭运算总结：（1）闭运算能够填平小湖（即小孔），弥合小裂缝，而总的位置和形状不变。

（2）闭运算是通过填充图像的凹角来滤波图像的。（3）结构元素大小的不同将导致滤波效果的不同。（4）不同结构元素的选择导致了不同的分割。

10. 简单阐述一下mIOU,写出mIOU的计算公式

mIoU值是一个衡量图像分割精度的重要指标。mIoU可解释为平均交并比，即在每个类别上计算IoU值（即真正样本数量/（真正样本数量+假负样本数量+假正样本数量））。

11.空洞卷积的具体实现

Dilated convolution就是为了在不是用pooling操作损失信息也能增加感受野。空洞卷积：在3*3卷积核中间填充0，有两种实现方式，第一，卷积核填充0，第二，输入等间隔采样。空洞卷积的rate，代表传统卷积核的相邻之间插入rate-1个空洞数。当rate=1时，相当于传统的卷积核。扩张卷积具有更大的感受野。

Pytorch实现过程如下：

13.简要阐述一下图像分割中常用的Loss

①Log loss

对于二分类而言，对数损失函数如下公式所示：

$$-\frac{1}{N}\sum_{i=1}^{N}(y_i \log p_i + (1-y_i)\log(1-p_i))$$

其中，y_i为输入实例x_i的真实类别, p_i为预测输入实例 x_i属于类别 1的概率.对所有样本的对数损失表示对每个样本的对数损失的平均值,对于完美的分类器, 对数损失为 0。

此loss function每一次梯度的回传对每一个类别具有相同的关注度！所以极易受到类别不平衡的影响。

②WCE Loss

带权重的交叉熵loss — Weighted cross-entropy (WCE)

R为标准的分割图，其中r_n为label 分割图中的某一个像素的GT。P为预测的概率图，p_n为像素的预测概率值，背景像素图的概率值就为1-P。

只有两个类别的带权重的交叉熵为：

$$WCE = -\frac{1}{N}\sum_{n=1}^{N} wr_n \log(p_n) + (1-r_n)\log(1-p_n)$$

w为权重，$w = \frac{N-\sum_n p_n}{\sum_n p_n}$缺点是需要人为的调整困难样本的权重，增加调参难度。

③Focal loss

能否使网络主动学习困难样本呢？focal loss的提出是在目标检测领域，为了解决正负样本比例严重失衡的问题。是由log loss改进而来的，为了与log loss进行对比，公式如下：

$$-\frac{1}{N}\sum_{i=1}^{N}\left(\alpha y_i (1-p_i)^{\gamma}\log p_i + (1-\alpha)(1-y_i)p_i^{Y}\log(1-p_i)\right)$$

说白了就多了一个$(1-p_i)^{\gamma}$，loss随样本概率的大小如下图所示：

其基本思想就是，对于类别极度不均衡的情况下，网络如果在log loss下会倾向于只预测负样本，并且负样本的预测概率p_i也会非常的高，回传的梯度也很大。但是如果添加$(1-p_i)^y$则会使预测概率大的样本得到的loss变小，而预测概率小的样本，loss变得大，从而加强对正样本的关注度。

可以改善目标不均衡的现象，对此情况比 binary_crossentropy 要好很多。

目前在图像分割上只是适应于二分类。需要添加额外的两个全局参数alpha和gamma，对于调参不方便。

④Dice loss

dice loss的提出是在V-net中，其中的一段原因描述是在感兴趣的解剖结构仅占据扫描的非常小的区域，从而使学习过程陷入损失函数的局部最小值。所以要加大前景区域的权重。

Dice 可以理解为是两个轮廓区域的相似程度，用A、B表示两个轮廓区域所包含的点集，定义为：

$$DSC(A,B)=2\frac{|A\cap B|}{|A|+|B|}$$

其次Dice也可以表示为：$DSC=\frac{2TP}{2TP+FN+FP}$其中TP，FP，FN分别是真阳性、假阳性、假阴性的个数。

二分类dice loss:

$$DL_2=1-\frac{\sum_{n=1}^{N}p_nr_n+\epsilon}{\sum_{n=1}^{N}p_n+r_n+\boldsymbol{\epsilon}}-\frac{\sum_{n=1}^{N}(1-p_n)(1-r_n)+\boldsymbol{\epsilon}}{\sum_{n=1}^{N}2-p_n-r_n+\boldsymbol{\epsilon}}$$

结论：

1.有时使用dice loss会使训练曲线有时不可信，而且dice loss好的模型并不一定在其他的评价标准上效果更好，例如mean surface distance 或者是Hausdorff surface distance。不可信的原因是梯度，对于softmax或者是log loss其梯度简化而言为p-t，t为目标值，p为预测值。而dice loss为$\frac{2t^2}{(p+t)^2}$，如果p，t过小则会导致梯度变化剧烈，导致训练困难。

2.属于直接在评价标准上进行优化。

3.不均衡的场景下的确好使。

⑤IOU loss

可类比DICE LOSS，也是直接针对评价标准进行优化。

定义如下：

$IOU = \frac{|A \cap B|}{|A|+|B|-|A \cap B|}$

在图像分割领域评价标准IOU实际上$IOU = \frac{TP}{TP+FP+FN}$ ，而TP，FP，FN分别是真阳性、假阳性、假阴性的个数。

而作为loss function，定义 $IOU = \frac{I(X)}{U(X)}$其中，

$$I(X) = X * Y$$
$$U(X) = X + Y - X * Y$$

X为预测值而Y为真实标签。

IOU loss的缺点同DICE loss是相类似的，训练曲线可能并不可信，训练的过程也可能并不稳定，有时不如使用softmax loss等的曲线有直观性，通常而言softmax loss得到的loss下降曲线较为平滑。

参考文献