Path Aggregation Network for Instance Segmentation

CVPR2018——PANet用于对象检测和实例分割（港中文+腾讯优图）

COCO17实例分割第一名，检测第二名

摘要

提出PANet，旨在提升基于候选区域的实例分割框架内的信息流传播，具体来说，通过自上而下的路径增强在较低层中的定位信息流，缩短底层特征和高层特征之间的信息路径，从而增强整个特征层次。同时提出自适应特征池化，连接特征网格和所有特征层次，让每个特征层次的有用信息直接传播到后续的候选区域子网络。另外创建一个分支用于捕获各个候选区域的不同识图，进一步提升mask预测。

介绍

在Mask R-CNN中信息流动可以进一步提升。低层特征有助于识别大型目标，但是低层特征到高层特征的路径太长，这增加了定位信息流动的难度。每个候选区域来自于同一层级特征的池化结果，这个分配是启发式的，这个过程可以进一步提升，因为其他没有使用的特征层次信息可能有助于最终预测，原先的分割预测分支只是在单层特征图上，这失去了收集不同信息的机会。

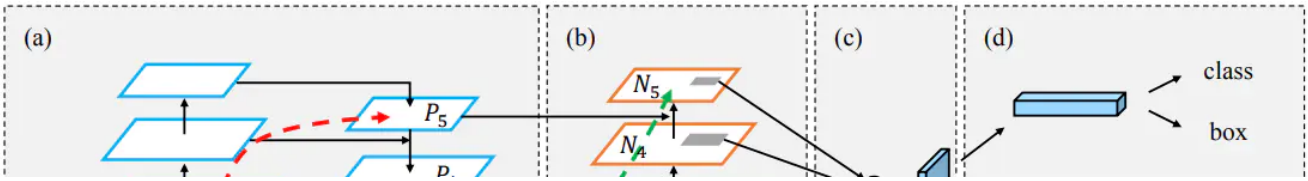
该网络基于FPN和Mask RCNN模型上提出三点创新，显著地提升了模型在物体检测和实例分割网络上的性能：

1、PANet改进主干网络结构，加强了特征金字塔的结构，缩短了高低层特征融合的路径；

2、提出了更加灵活的ROI池化，之前FPN的ROI池化只从高层特征取值，现在在各个尺度的特征上操作；

3、预测分割的时候用一个额外的FC支路来辅助全卷积分割支路的结果。

网络模型如图所示：



模型结构图

(a)代表FPN股骨架网络，(b)代表自下而上的路径增强，(c)自适应特征池化，(d)边框分支，(e)分割预测分支（全连接融合）

最近的工作

实例分割

1、基于候选区域的方法：这样的方法和目标检测有很强的的联系，R-CNN中的候选区域使用Selective Search方法提取，送到模型中提取特征用于分类。Fast/Faster R-CNN 和SPPNet通过池化全局特征图加速处理过程。也有工作是在网络中生成分割掩膜作为候选区域或最终结果。Mask R-CNN也是基于侯选区域的，本文以Mask R-CNN为基础从几个不同方向上做了改进。

2、基于分割：学习特定的设计的变换和对象的边界，实例分割是从预测转换上解码出来的，DIN融合目标检测和语义分割，也有使用图模型来预测实例分割的方法。

多层特征图

来自不同的层的特征图可用于图像识别。SharpMask，LRR融合特征用于精细分割；FCN、U-Net通过skip-connection融合来自低层的信息；TDM，FPN通过横向连接增加了从上到下的路径用于目标检测，TDM是将最高分辨率的特征图融入到池化的特征中，与TDM不同的是，SSD，MS-CNN、FPN将候选区域分配到合适的特征层用于推断，论文以FPN为基准并且大幅增强该结构。

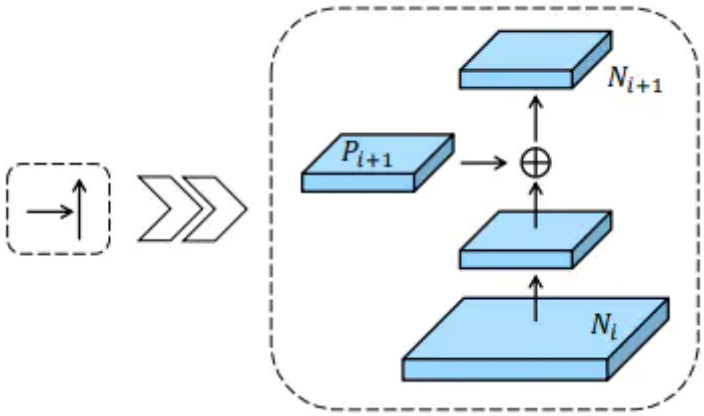
网络模型

Bottom-up Path Augmentation

高层的神经元对应整个目标的响应，低层神经元更可能是被局部图像所激活。这表明需要增强从上而下的传播，获得强语义信息。

PANet主干网络与FPN不同之处在于新构建的N2（P2）——N5卷积，相邻两层之间的细节结构如下图所示，其中的融合操作是逐像素加。每个特征图 N_i 先经过步长为2的3*3卷积降分辨率降为原来的一半，特征图 P_{i+1} 的每个元素和下采样的特征图通过横向连接相加，再经过一个3*3卷积生成

N_{i+1} 。所有的特征图都使用256通道，所有卷积后接ReLU，每个候选区域的特征是从新生成的特征图N2（P2）——N5上池化生成的。



特征融合

构建该新支路的优势在于缩短了底层尺寸大的特征到高层尺寸小的特征之间的距离，让特征融合更加有效。其变化可以参考网络结构图中红色（原FPN特征融合路径）和绿色（PANet特征融合路径）虚线。绿色虚线所跨越的卷积层会更少，10层不到，FPN容和路径跨越了100多层。

Adaptive Feature Pooling

在FPN中，依据候选区域的大小将候选区域分配到不同特征层次。这样小的候选区域分配到低层特征，大的候选区域分配到高层特征，这虽然简单但却很有效，却也可能会产生非最优结果。例如两个具有10个像素差的候选区域可能分配到不同特征层次，但是实际上这两个候选区域非常相似。特征的重要性可能与他们所属的特征层次没有太大关系，高层特征具有大的感受野并捕获了丰富的上下文信息，有利于小目标检测；低层特征具有许多微小细节和高定位精度，有利于大目标检测。也就是说，无论是低层特征还是高层特征都有用。对于每个候选区域，池化来自所有层次的特征，然后融合它们做预测，这称之为自适应特征池化。

对于每个候选区域，将其映射到不同的特征层次，如下图的深灰色区域，使用ROIAlign池化来自不同层次的特征，再使用融合操作融和不同层次的特征。

自适应特征池化

Fully-connected Fusion

全连接层和MLP广泛应用于实例分割中，用于预测mask和生成mask候选区域。FCN同样也能够预测逐像素的mask。 Mask R-CNN使用了一个小型的FCN应用于池化后特征网格用于预测对于的mask。

全连接层和全卷积层相比：

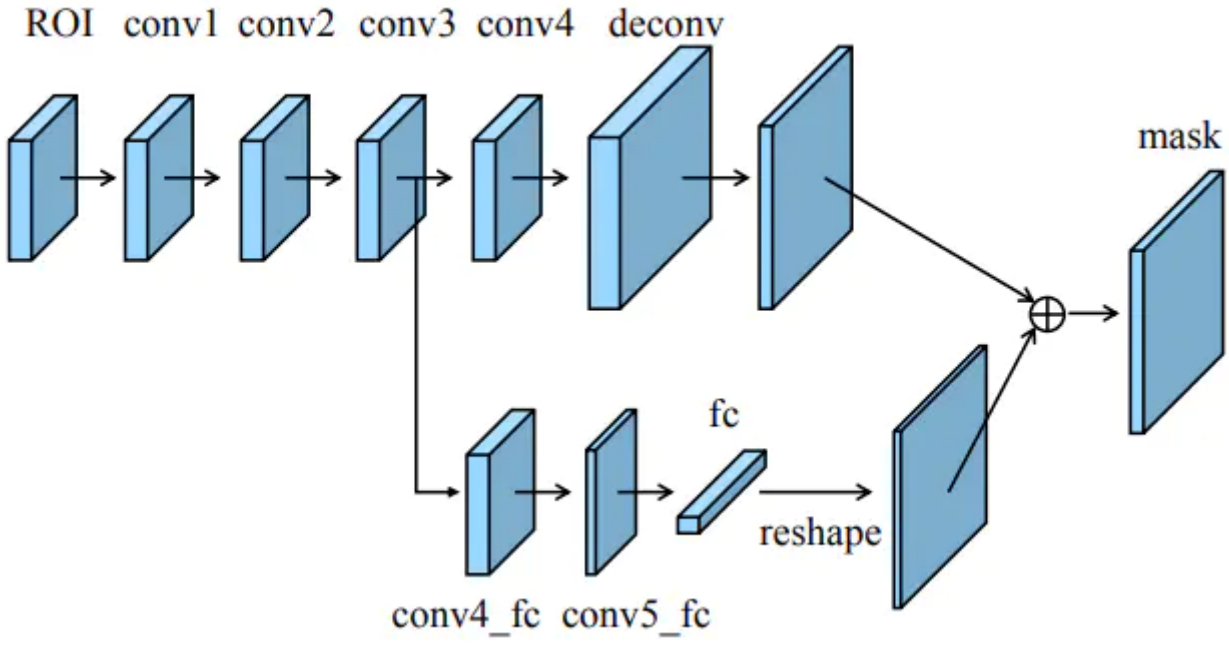
FCN给出了像素级预测，这是基于局部感受野和共享的核参数；

FC是位置敏感的，对于不同空间位置的预测都是通过一组可变参数实现的，FC具有适应不同空间位置的能力。FC对于每个空间位置预测是基于整个候选区域的全局信息，这对于区分不同实例和识别属于同一对象的分离部分很有效。

因此将FCN和FC的预测结果融合可以达到更好的预测。

Mask Prediction Structure

负责预测mask的组件是一个轻量级、易实现的分支，mask分支的输入是每个候选区域融合后的池化特征，如下图：



Mask预测分支

主分支是4个连续的卷积层和一个反卷积，每个卷积层核大小为3×3，通道为256，后面再接一个上采样2倍的deconv。这是用于预测每个类别mask的二进制像素值。

使用一个短路径从conv3连接到FC层，中间过两个卷积层conv4_fc和conv5_fc，conv5_fc卷积层通道数减半以减少计算量。

mask大小设置为28×28，FC层产生784×1×1的向量，故reshape成和FCN预测的mask同样的空间尺寸。再和FCN的输出相加得到最终预测。FC用于预测类别不可知的背景/前景 mask，这不仅效率高，而且允许更多样本训练FC层的参数，因此泛化能力更强。只使用一个FC层的原因是防止隐藏的空间特征坍塌成短特征向量，导致丢失空间信息。

Experiments

Method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	Backbone
Champion 2016 [33]	37.6	59.9	40.4	17.1	41.0	56.0	6×ResNet-101
Mask R-CNN [21]+FPN [35]	35.7	58.0	37.8	15.5	38.1	52.4	ResNet-101
Mask R-CNN [21]+FPN [35]	37.1	60.0	39.4	16.9	39.9	53.5	ResNeXt-101
PANet / PANet [ms-train]	36.6 / 38.2	58.0 / 60.2	39.3 / 41.4	16.3 / 19.1	38.1 / 41.1	53.1 / 52.6	ResNet-50
PANet / PANet [ms-train]	40.0 / **42.0**	62.8 / **65.1**	43.1 / **45.7**	18.8 / **22.4**	42.3 / **44.7**	57.2 / **58.1**	ResNeXt-101

COCO语义分割数据集

Method	AP^{bb}	AP^{bb}_{50}	AP^{bb}_{75}	AP^{bb}_S	AP^{bb}_M	AP^{bb}_L	Backbone
Champion 2016 [27]	41.6	62.3	45.6	24.0	43.9	55.2	2×ResNet-101 + 3×Inception-ResNet-v2
RentinaNet [36]	39.1	59.1	42.3	21.8	42.7	50.2	ResNet-101
Mask R-CNN [21]+FPN [35]	38.2	60.3	41.7	20.1	41.1	50.2	ResNet-101
Mask R-CNN [21]+FPN [35]	39.8	62.3	43.4	22.1	43.2	51.2	ResNeXt-101
PANet / PANet [ms-train]	41.2 / 42.5	60.4 / 62.3	44.4 / 46.4	22.7 / 26.3	44.0 / 47.0	54.6 / 52.3	ResNet-50
PANet / PANet [ms-train]	45.0 / **47.4**	65.0 / **67.2**	48.6 / **51.8**	25.4 / **30.1**	48.6 / **51.7**	59.1 / **60.0**	ResNeXt-101

COCO目标检测数据集