

FPN目标检测网络



报告人： 石强

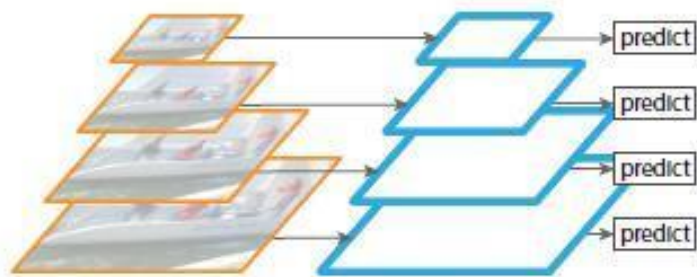


简介

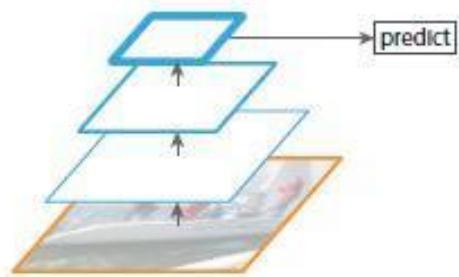
Contents

原来多数的object detection算法都是只采用顶层特征做预测，但我们知道低层的特征语义信息比较少，但是目标位置准确；高层的特征语义信息比较丰富，但是目标位置比较粗略。另外虽然也有些算法采用多尺度特征融合的方式，但是一般是采用融合后的特征做预测，而本文不一样的地方在于预测是在不同特征层独立进行的。

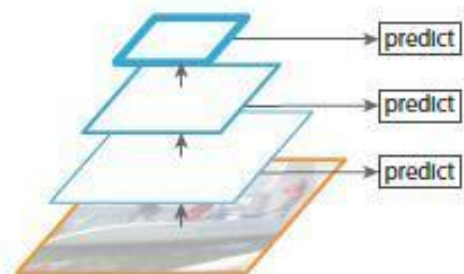
4种利用特征的形式



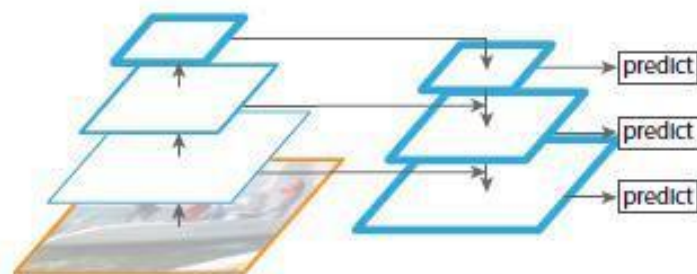
(a) Featurized image pyramid



(b) Single feature map

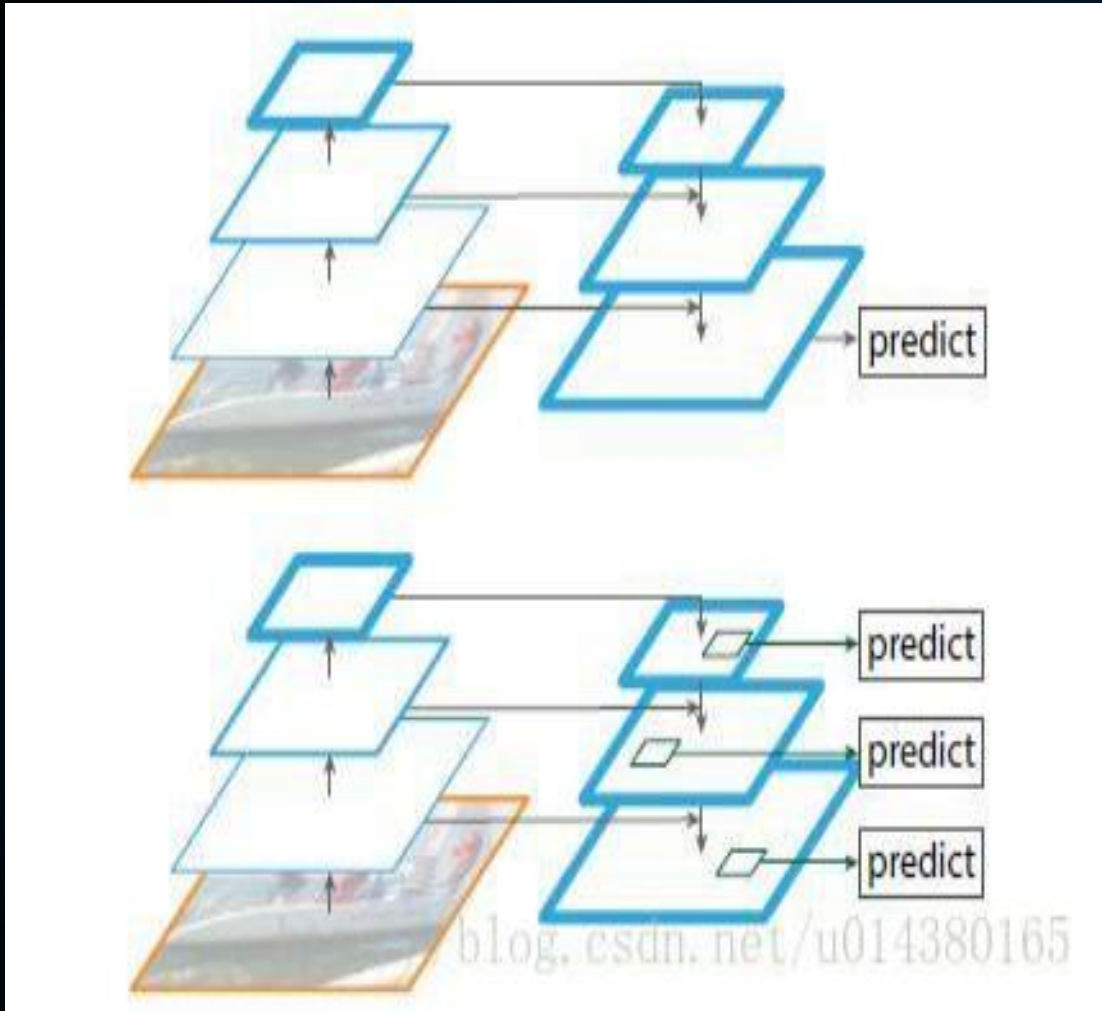


(c) Pyramidal feature hierarchy



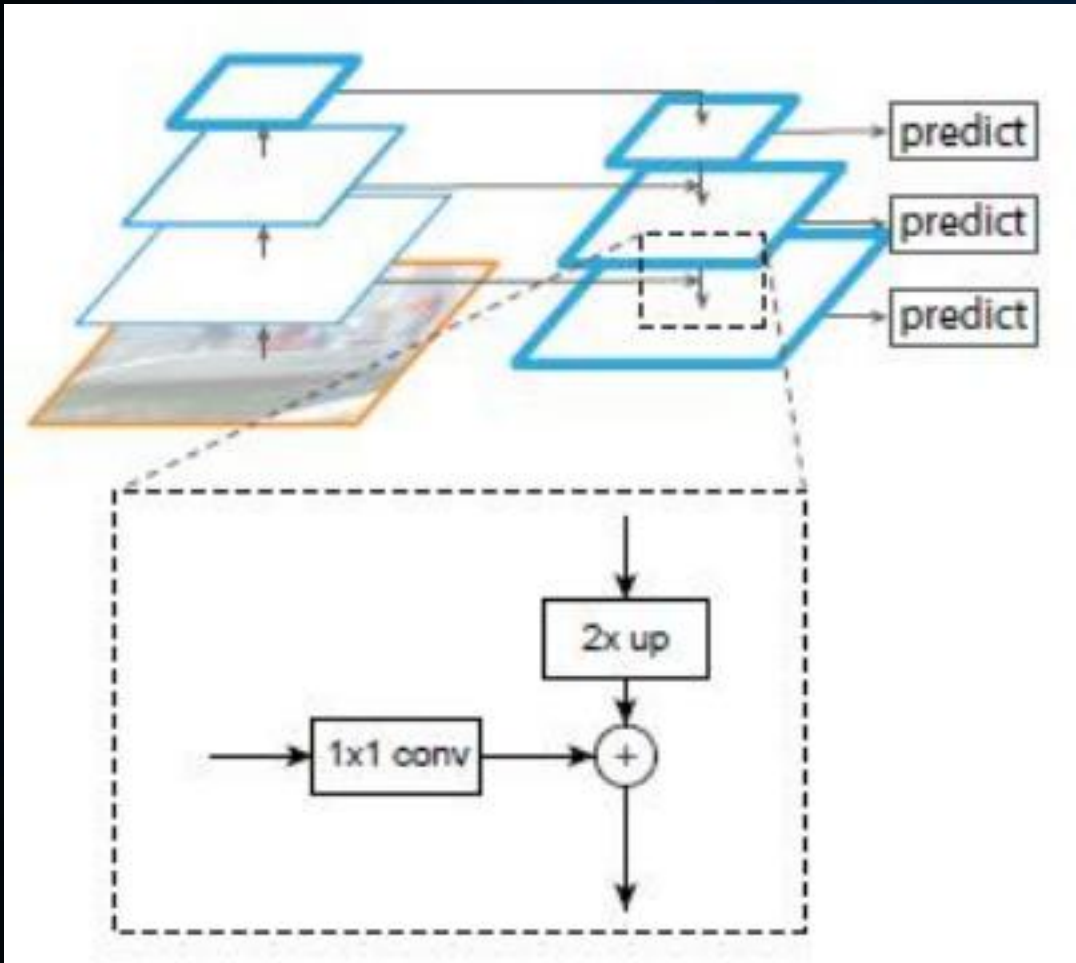
(d) Feature Pyramid Network

a 图像金子塔 b 顶层特征 c 多层特征融合 d 多层特征融合 (底层特征和顶层特征融合)



上一个结构中，带有skip connection的网络结构在预测的时候是在finest level（自顶向下的最后一层）进行的，简单讲就是经过多次上采样并融合特征到最后一步，拿最后一步生成的特征做预测

而下面一个网络结构和上面的类似，区别在于预测是在每一层中独立进行的

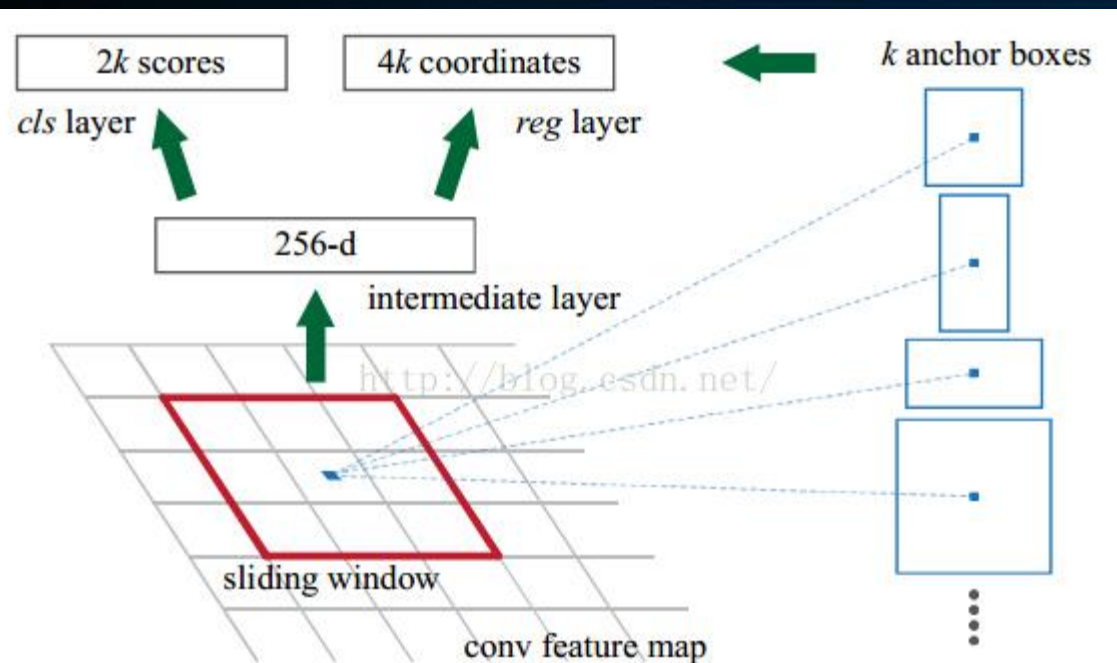


一个自底向上的线路，一个自顶向下的线路，横向连接（lateral connection）。图中放大的区域就是横向连接，这里 1×1 的卷积核的主要作用是减少卷积核的个数，也就是减少了feature map的个数，并不改变feature map的尺寸大小

自底向上其实就是网络的前向过程。在前向过程中，feature map的大小在经过某些层后会改变，而在经过其他一些层的时候不会改变，作者将不改变feature map大小的层归为一个stage，因此每次抽取的特征都是每个stage的最后一个层输出，这样就能构成特征金字塔。

自顶向下的过程采用上采样（upsampling）进行，而横向连接则是将上采样的结果和自底向上生成的相同大小的feature map进行融合（merge）。在融合之后还会再采用 3×3 的卷积核对每个融合结果进行卷积，目的是消除上采样的混叠效应（aliasing effect）

RPN网络



作者一方面将FPN放在RPN网络中用于生成proposal，原来的RPN网络是以主网络的某个卷积层输出的feature map作为输入，简单讲就是只用这一个尺度的feature map。但是现在要将FPN嵌在RPN网络中，生成不同尺度特征并融合作为RPN网络的输入。在每一个scale层，都定义了不同大小的anchor，对于P2, P3, P4, P5, P6这些层，定义anchor的大小为 $32^2, 64^2, 128^2, 256^2, 512^2$ ，另外每个scale层都有3个长宽对比度：1:2, 1:1, 2:1。所以整个特征金字塔有15种

正负样本

如果某个anchor和一个给定的ground truth有最高的IOU或者和任意一个Ground truth的IOU都大于0.7，则是正样本。如果一个anchor和任意一个ground truth的IOU都小于0.3，则为负样本。

假如fpn的rpn的有效性

RPN	feature	# anchors	lateral?	top-down?	AR ¹⁰⁰	AR ^{1k}	AR ^{1k} _s	AR ^{1k} _m	AR ^{1k} _l
(a) baseline on conv4	C ₄	47k			36.1	48.3	32.0	58.7	62.2
(b) baseline on conv5	C ₅	12k			36.3	44.9	25.3	55.5	64.2
(c) FPN	{P _k }	200k	✓	✓	44.0	56.3	44.9	63.4	66.2
Ablation experiments follow:									
(d) bottom-up pyramid	{P _k }	200k	✓		37.4	49.5	30.5	59.9	68.0
(e) top-down pyramid, w/o lateral	{P _k }	200k		✓	34.5	46.1	26.5	57.4	64.7
(f) only finest level	P ₂	750k	✓	✓	38.4	51.3	35.1	59.7	67.6

网络这些结果都是基于ResNet-50。评价标准采用AR，AR表示Average Recall，AR右上角的100表示每张图像有100个anchor，AR的右下角s，m，l表示COCO数据集中object的大小分别是小，中，大。feature列的大括号{}表示每层独立预测。

从 (a) (b) (c) 的对比可以看出FRN的作用确实很明显。另外 (a) 和 (b) 的对比可以看出高层特征并非比低一层的特征有效。

(d) 表示只有横向连接，而没有自顶向下的过程，也就是仅仅对自底向上 (bottom-up) 的每一层结果做一个1*1的横向连接和3*3的卷积得到最终的结果，有点像Fig1的 (b)。从feature列可以看出预测还是分层独立的。作者推测 (d) 的结果并不好的原因在于在自底向上的不同层之间的semantic gaps比较大。

(e) 表示有自顶向下的过程，但是没有横向连接，即向下过程没有融合原来的特征。这样效果也不好的原因在于目标的location特征在经过多次降采样和上采样过程后变得更加不准确。

(f) 采用finest level层做预测（参考Fig2的上面那个结构），即经过多次特征上采样和融合到最后一步生成的特征用于预测，主要是证明金字塔分层独立预测的表达能能力。显然finest level的效果不如FPN好，原因在于PRN网络是一个窗口大小固定的滑动窗口检测器，因此在金字塔的不同层滑动可以增加其对尺度变化的鲁棒性。另外 (f) 有更多的anchor说明增加anchor的数量并不能有效提高准确率。

Faster R-CNN	proposals	feature	head	lateral?	top-down?	AP@0.5	AP	AP _s	AP _m	AP _l
(*) baseline from He <i>et al.</i> [16] [†]	RPN, C_4	C_4	conv5			47.3	26.3	-	-	-
(a) baseline on conv4	RPN, C_4	C_4	conv5			53.1	31.6	13.2	35.6	47.1
(b) baseline on conv5	RPN, C_5	C_5	2fc			51.7	28.0	9.6	31.9	43.1
(c) FPN	RPN, $\{P_k\}$	$\{P_k\}$	2fc	✓	✓	56.9	33.9	17.8	37.7	45.8

method	backbone	competition	image pyramid	test-dev					test-std				
				AP@.5	AP	AP _s	AP _m	AP _l	AP@.5	AP	AP _s	AP _m	AP _l
ours, Faster R-CNN on FPN	ResNet-101	-		59.1	36.2	18.2	39.0	48.2	58.5	35.8	17.5	38.7	47.8
<i>Competition-winning single-model results follow:</i>													
G-RMI [†]	Inception-ResNet	2016		-	34.7	-	-	-	-	-	-	-	-
AttractionNet [‡] [10]	VGG16 + Wide ResNet [§]	2016	✓	53.4	35.7	15.6	38.0	52.7	52.9	35.3	14.7	37.6	51.9
Faster R-CNN +++ [16]	ResNet-101	2015	✓	55.7	34.9	15.6	38.7	50.9	-	-	-	-	-
Multipath [40] (on minival)	VGG-16	2015		49.6	31.5	-	-	-	-	-	-	-	-
ION [‡] [2]	VGG-16	2015		53.4	31.2	12.8	32.9	45.2	52.9	30.7	11.8	32.8	44.8

作者提出的FPN (Feature Pyramid Network) 算法同时利用低层特征高分辨率和高层特征的高语义信息，通过融合这些不同层的特征达到预测的效果。并且预测是在每个融合后的特征层上单独进行的，这和常规的特征融合方式不同。

祝您有个美好的一天

谢谢聆听

