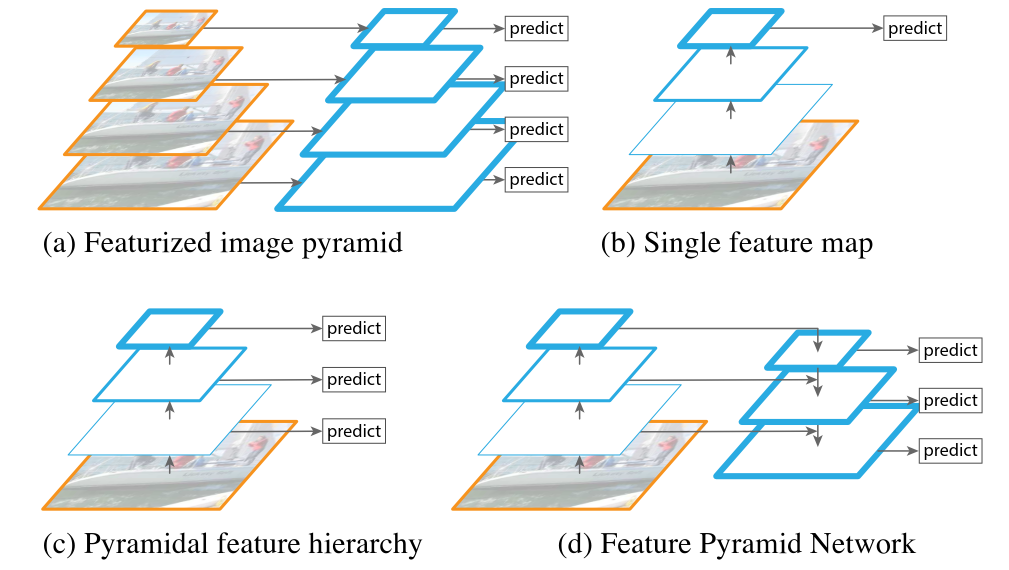
1.Using FPN in a basic Faster R-CNN system, our method achieves state-of-the-art single-model results on the COCO detection benchmark without bells and whistles, surpassing all existing single-model entries including those from the COCO 2016 challenge winners

基于Faster R-CNN，构建了FPN，不吹牛的通过实验在COCO数据集获得领先效果，超过所有COCO2016比赛的胜者。

2.几种常见的金字塔结构



(a) Using an image pyramid to build a feature pyramid.Features are computed on each of the image scales independently,which is slow. (b) Recent detection systems have opted to use only single scale features for faster detection. (c) An alternative is to reuse the pyramidal feature hierarchy computed by a ConvNet as if it were a featurized image pyramid. (d) Our proposed Feature Pyramid Network (FPN) is fast like (b) and (c), but more accurate.In this ﬁgure, feature maps are indicate by blue outlines and thicker outlines denote semantically stronger features

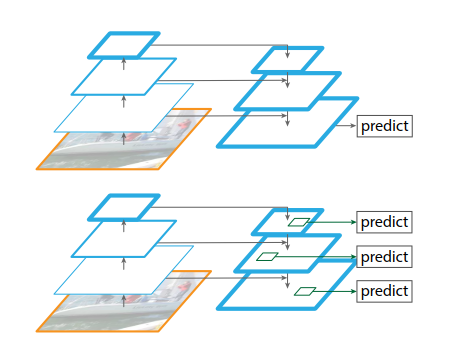
(a)使用图像金字塔构建特征金字塔。每个图像尺度上的特征都是独立计算的，速度很慢。(b)最近的探测系统只选择使用单一比额表的特点来加速探测。(c)另一种办法是重复使用由ConvNet计算的金字塔特征层次结构，就象它是一个特征化的图像金字塔一样。(d)我们提出的特征金字塔网络(Feature Pyramid Network, FPN)与(b)和(c)一样快，但更准确。在这个图中，特征图用蓝色轮廓表示，较粗的轮廓表示语义上较强的特征

3.The Single Shot Detector (SSD) [22] is one of the ﬁrst attempts at using a ConvNet’s pyramidal feature hierarchy as if it were a featurized image pyramid (Fig. 1(c)). Ideally,the SSD-style pyramid would reuse the multi-scale feature maps from different layers computed in the forward pass and thus come free of cost. But to avoid using low-level features SSD foregoes reusing already computed layers and instead builds the pyramid starting from high up in the network (e.g., conv4 3 of VGG nets [36]) and then by adding several new layers. Thus it misses the opportunity to reuse the higher-resolution maps of the feature hierarchy. We show that these are important for detecting small objects

(SSD)[22]是使用ConvNet s的特征层次金子塔的首批尝试之一(图1(c))。理想情况下，ssd样式的金字塔将重用在前向遍历中计算的来自不同层的多尺度特征映射，因此无需成本。但是为了避免使用底层特性，SSD放弃了重用已经计算好的层，而是从网络的高处开始构建金字塔(例如，VGG nets[36]中的conv4 3)，然后添加几个新层。因此，它错过了重用特性层次结构的高分辨率映射的机会。我们证明这些对于探测小物体是很重要的

4.Top: a top-down architecture with skip connections,where predictions are made on the ﬁnest level (e.g., [28]). Bottom:our model that has a similar structure but leverages it as a feature pyramid, with predictions made independently at all levels

Top:一个带有跳过连接的自顶向下架构，其中预测是在最好的级别上进行的(例如，[28])。底部:我们的模型有一个类似的结构，但是利用它作为一个特征金字塔，在所有级别独立地进行预测



5.In ablation experiments, we ﬁnd that for bounding box proposals, FPN signiﬁcantly increases the Average Recall (AR) by 8.0 points; for object detection, it improves the COCO-style Average Precision (AP) by 2.3 points and PASCAL-style AP by 3.8 points, over a strong single-scale baseline of Faster R-CNN on ResNets

FPN将平均召回率提升了8%，在COCO数据集上的平均准确率提升了2.3%，在PASCAL数据集上的平均准确率提升了3.8%.

6. Feature Pyramid Networks(网络是如何构成的)

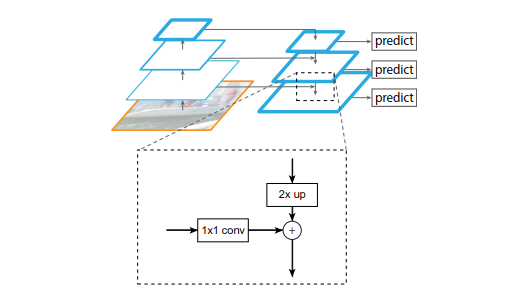


Figure 3. A building block illustrating the lateral connection and the top-down pathway, merged by addition

一个展示横向连接和自顶向下路径的构建块，通过加法合并

7.网络组合方式

**Bottom-up pathway：**

The bottom-up pathway is the feedforward computation of the backbone ConvNet, which computes a feature hierarchy consisting of feature maps at several scales. We do not include conv1 into the pyramid due to its large memory footprint.

取conv2，conv3，conv4，conv5作为金字塔型的一层，不选用第一层。

**Top-down pathway and lateral connections：**

The bottom-up feature map is of lower-level semantics, but its activations are more accurately localized as it was subsampled fewer times.

8.FPN的应用

**Feature Pyramid Networks for RPN**

We adapt RPN by replacing the single-scale feature map with our FPN. We attach a head of the same design (3×3 conv and two sibling 1×1 convs) to each level on our feature pyramid. Because the head slides densely over all locations in all pyramid levels, it is not necessary to have multi-scale anchors on a speciﬁc level. Instead, we assign anchors of a single scale to each level. Formally, we deﬁne the anchors to have areas of {322, 642, 1282, 2562, 5122} pixels on {P2, P3, P4, P5, P6} respectively. As in [29] we also use anchors of multiple aspect ratios {1:2, 1:1, 2:1} at each level. So in total there are 15 anchors over the pyramid.

an anchor is assigneda positive label if it has the highest IoU for a given ground-truth box or an IoU over 0.7 with any ground-truth box,and a negative label if it has IoU lower than 0.3 for all ground-truth boxes.

每一个anchor，如果对某个ground-truth的IoU>0.7则打上正标签，如果对于所有ground-truth的IoU<0.3则打上负标签。

**Feature Pyramid Networks for Fast R-CNN**

Fast R-CNN is most commonly performed on a single-scale feature map. To use it with our FPN, we need to assign RoIs of different scales to the pyramid levels.

So unlike[16], we simply adopt RoI pooling to extract 7×7 features,and attach two hidden 1,024-d fully-connected (fc) layers(each followed by ReLU) before the ﬁnal classiﬁcation and bounding box regression layers.

跟Fast R-CNN不同的是，我们在RoI pooling之后，连接了两个全连接层。

9.实验对比

