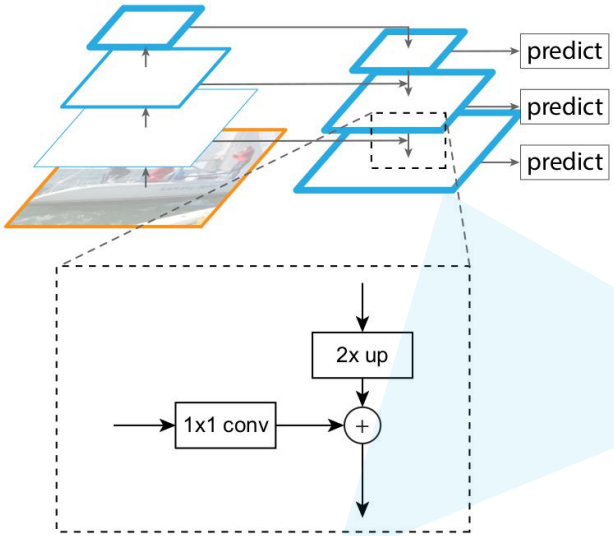
Feature Pyramid Networks

Feature Pyramid Networks for Object Detection

Tsung-Yi Lin[1,2], Piotr Dollár[1], Ross Girshick[1],
Kaiming He[1], Bharath Hariharan[1], and Serge Belongie[2]

[1]Facebook AI Research (FAIR)
[2]Cornell University and Cornell Tech

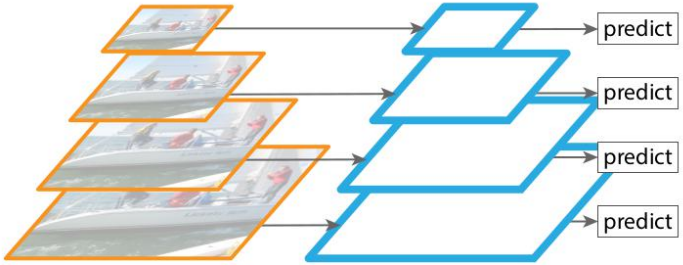
针对目标检测任务
cocoAP提升2.3个点
pascalAP提升3.8个点



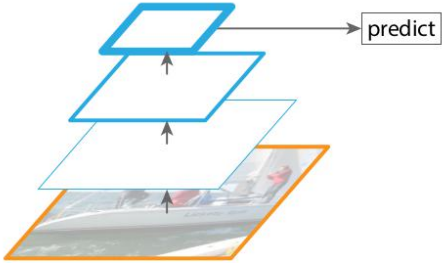
2016
Computer Vision and Pattern Recognition

https://arxiv.org/abs/1612.03144

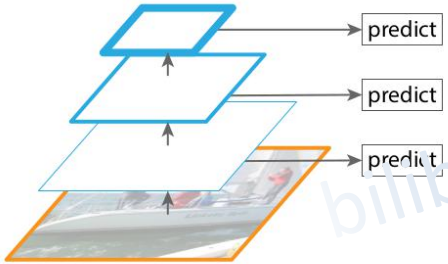
Feature Pyramid Networks



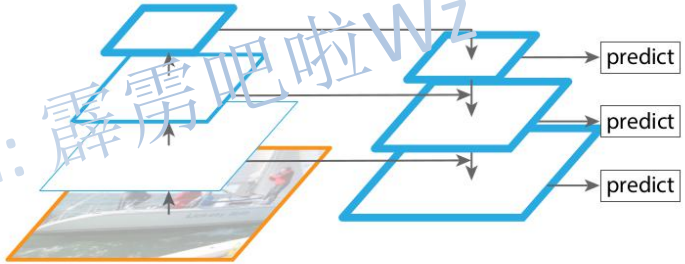
(a) Featurized image pyramid



(b) Single feature map



(c) Pyramidal feature hierarchy



(d) Feature Pyramid Network

Feature Pyramid Networks

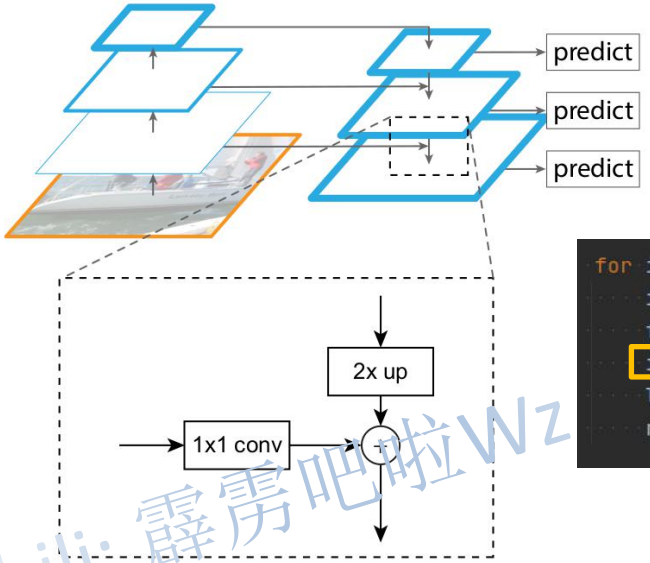
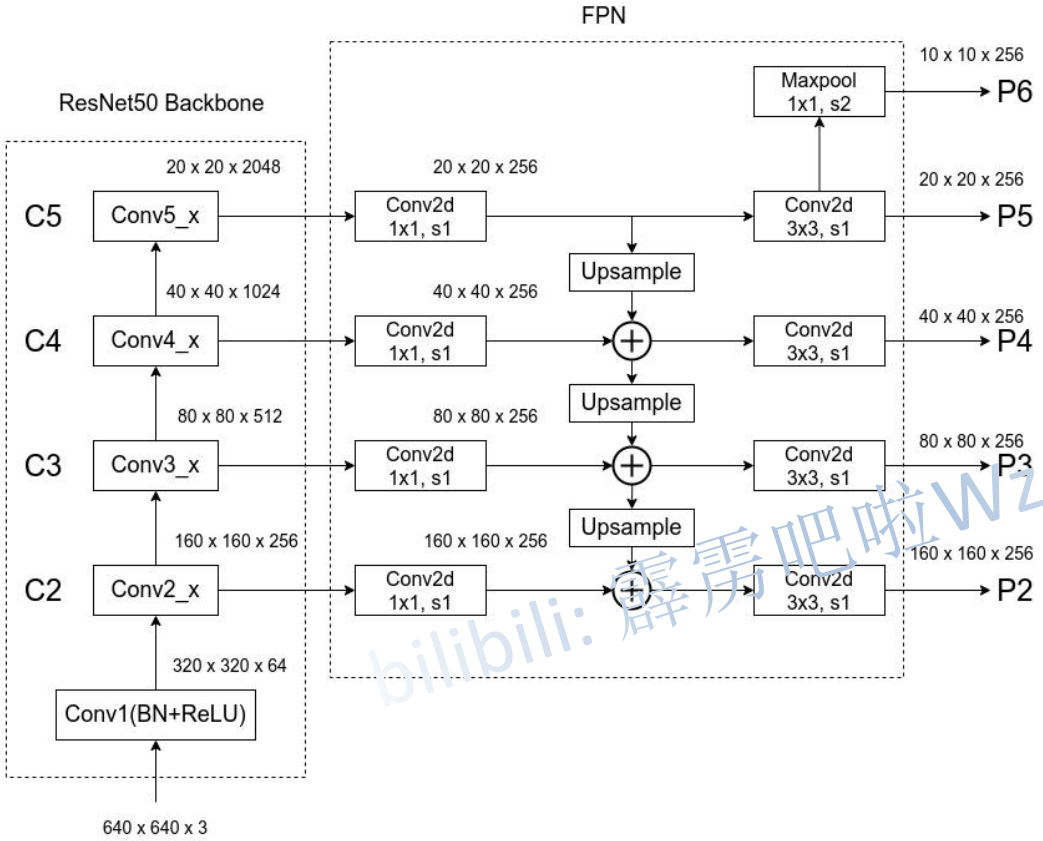


Figure 3. A building block illustrating the lateral connection and the top-down pathway, merged by addition.

Fig. 3 shows the building block that constructs our top-down feature maps. With a coarser-resolution feature map, we upsample the spatial resolution by a factor of 2 (using nearest neighbor upsampling for simplicity). The upsam-

```
for idx in range(len(x) - 2, -1, -1):
    inner_lateral = self.get_result_from_inner_blocks(x[idx], idx)
    feat_shape = inner_lateral.shape[-2:]
    inner_top_down = F.interpolate(last_inner, size=feat_shape, mode="nearest")
    last_inner = inner_lateral + inner_top_down
    results.insert(0, self.get_result_from_layer_blocks(last_inner, idx))
```

Feature Pyramid Networks



注意：P6只用于RPN部分，不在Fast-RCNN部分使用

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 $\{32^2, 64^2, 128^2, 256^2, 512^2\}$ pixels on $\{P_2, P_3, P_4, P_5, P_6\}$ respectively.[1] 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.

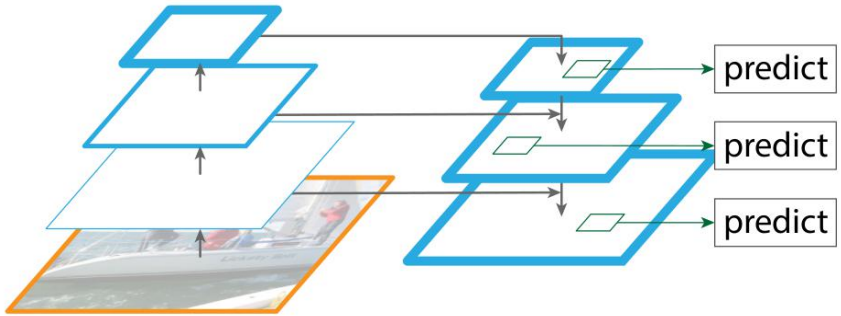
注意：针对不同的预测特征层，RPN和Fast RCNN的权重共享

Feature Pyramid Networks

We view our feature pyramid as if it were produced from an image pyramid. Thus we can adapt the assignment strategy of region-based detectors [15, 11] in the case when they are run on image pyramids. Formally, we assign an RoI of width w and height h (on the input image to the network) to the level P_k of our feature pyramid by:

$$k = \lfloor k_0 + \log_2(\sqrt{wh}/224) \rfloor. \quad (1)$$

Here 224 is the canonical ImageNet pre-training size, and k_0 is the target level on which an RoI with $w \times h = 224^2$ should be mapped into. Analogous to the ResNet-based Faster R-CNN system [16] that uses C_4 as the single-scale feature map, we set k_0 to 4. Intuitively, Eqn. (1) means that if the RoI's scale becomes smaller (say, 1/2 of 224), it should be mapped into a finer-resolution level (say, $k = 3$).



沟通方式

1.github

https://github.com/WZMIAOMIAO/deep-learning-for-image-processing

2.bilibili

https://space.bilibili.com/18161609/channel/index

3.CSDN

https://blog.csdn.net/qq_37541097/article/details/103482003