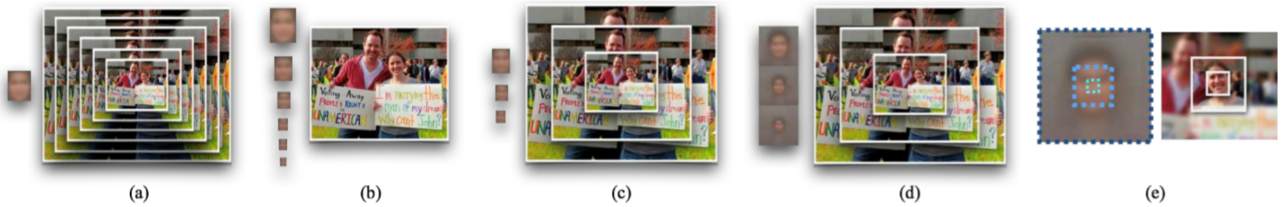
Finding Tiny Faces --Peiyun Hu, Deva Ramanan 读后感

传统检测方法对于微小人脸的检测不足

对于ROI为例的bounding选取而言，如何选择一个合适的bounding-size是很难的。

改进

利用粗糙图像金字塔来获取大规模的变化，定义了多个尺度混合检测器



(a)传统的方法，在非常离散的图像上构建单模板金字塔

(b)为检测不同分辨率下的可用线索，对单个图像建立多个不同标准检测器

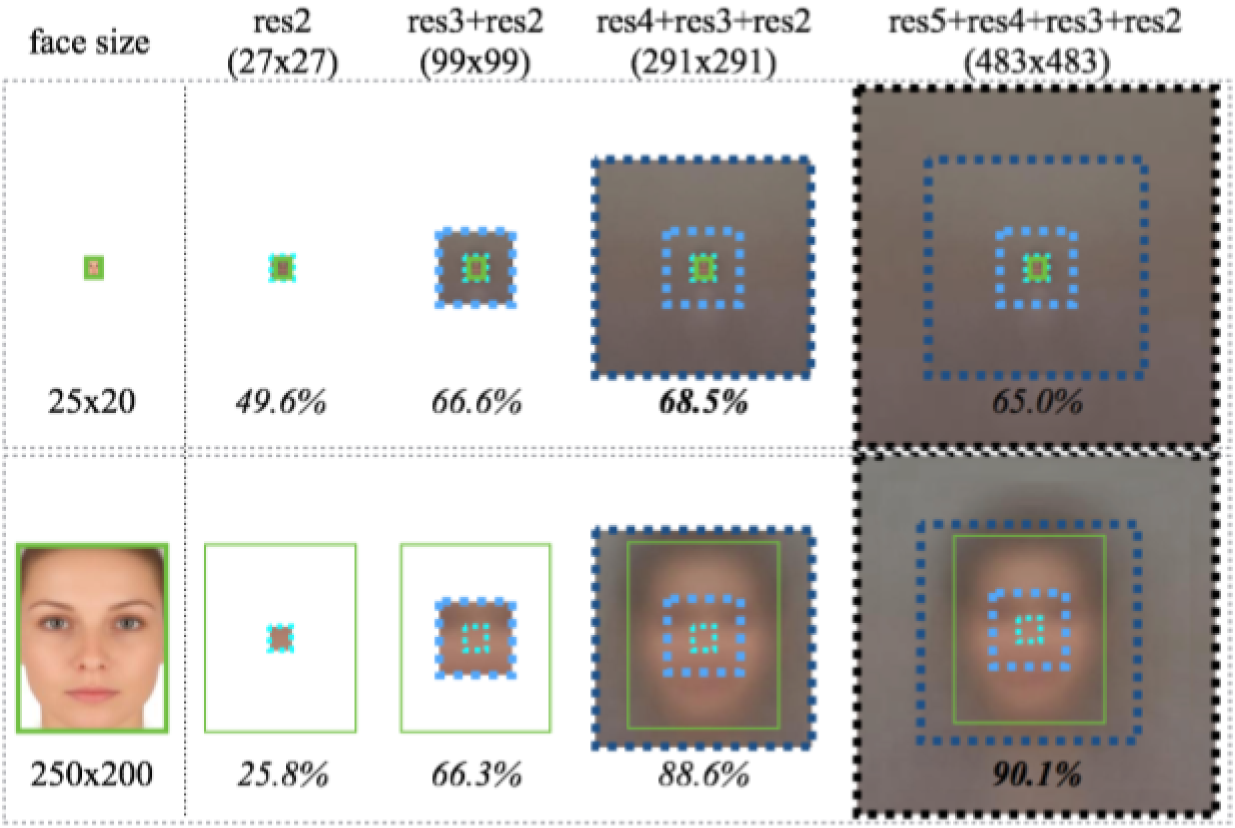
(c)结合a和b，构建一个粗糙的图像金字塔

(d)为提高微小人脸的检测，添加一个能够有效接受所有标量并实现固定尺寸的上下文模型

(e)从深度模型的多个层中提取特征的模板（后文的foveal descriptor）

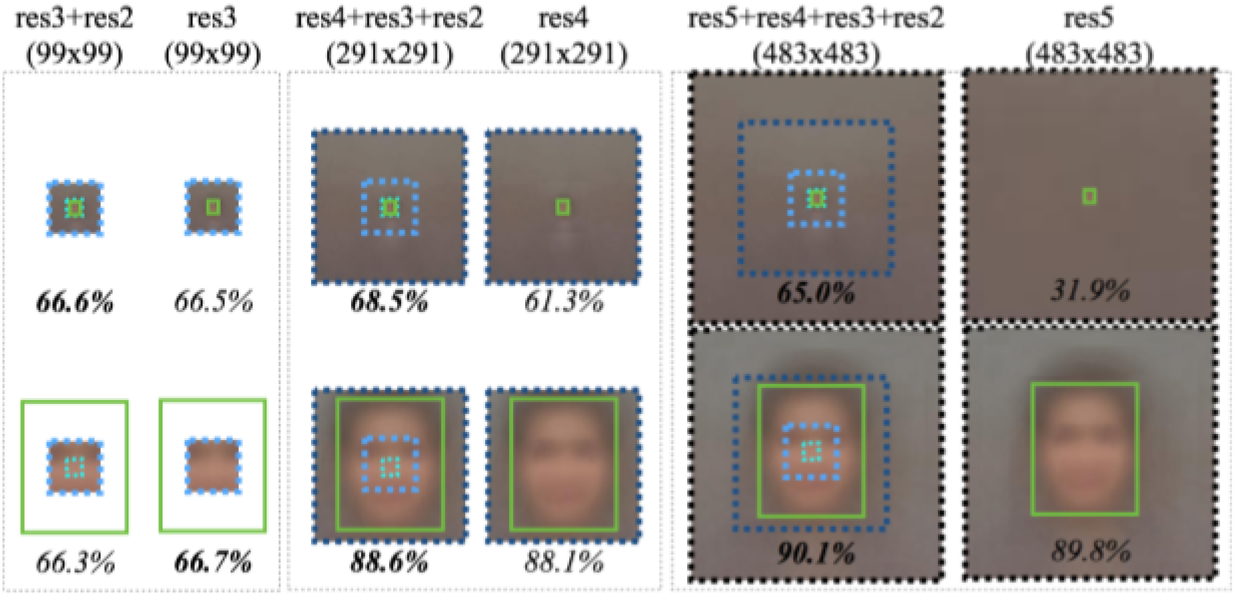
证明foveal descriptor能捕获大感受野(RF)的高分辨率的细节以及低分率的线索特征

上下文的作用



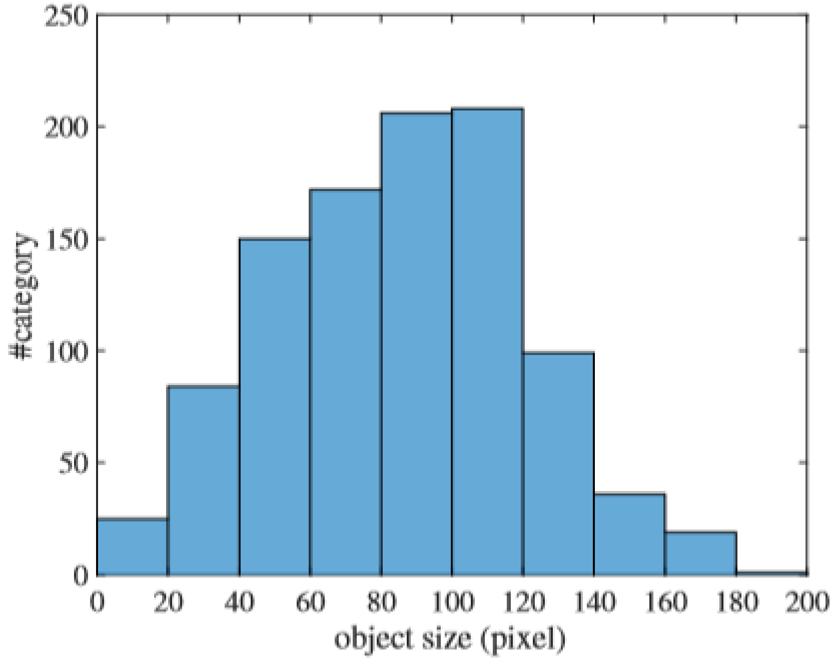
1. 更小的模板(感受野更小)能提高人脸的检测准确度，相对于大尺度人脸，小脸更加明显
2. 感受野超过300px的情况下，准确度下降。

foveal descripor的作用



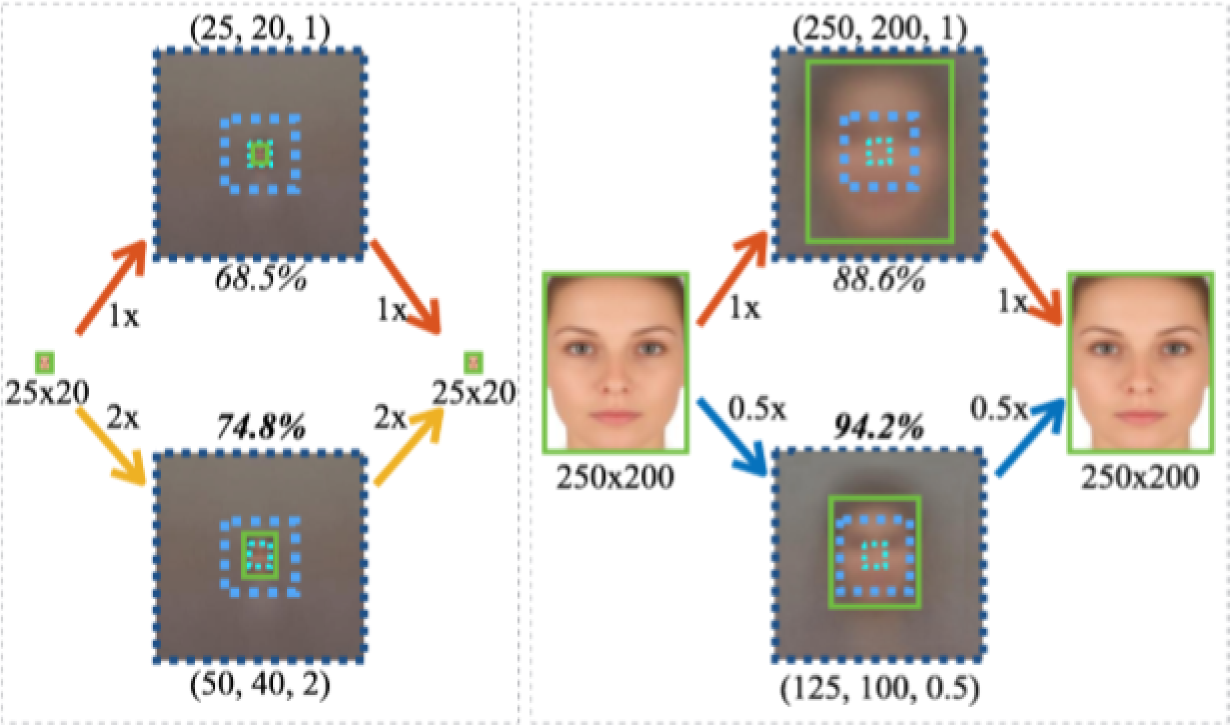
有foveal descripor对于微小人脸的检测有很大作用，对于大尺寸人脸的检测没有太大作用。

关于训练集的问题



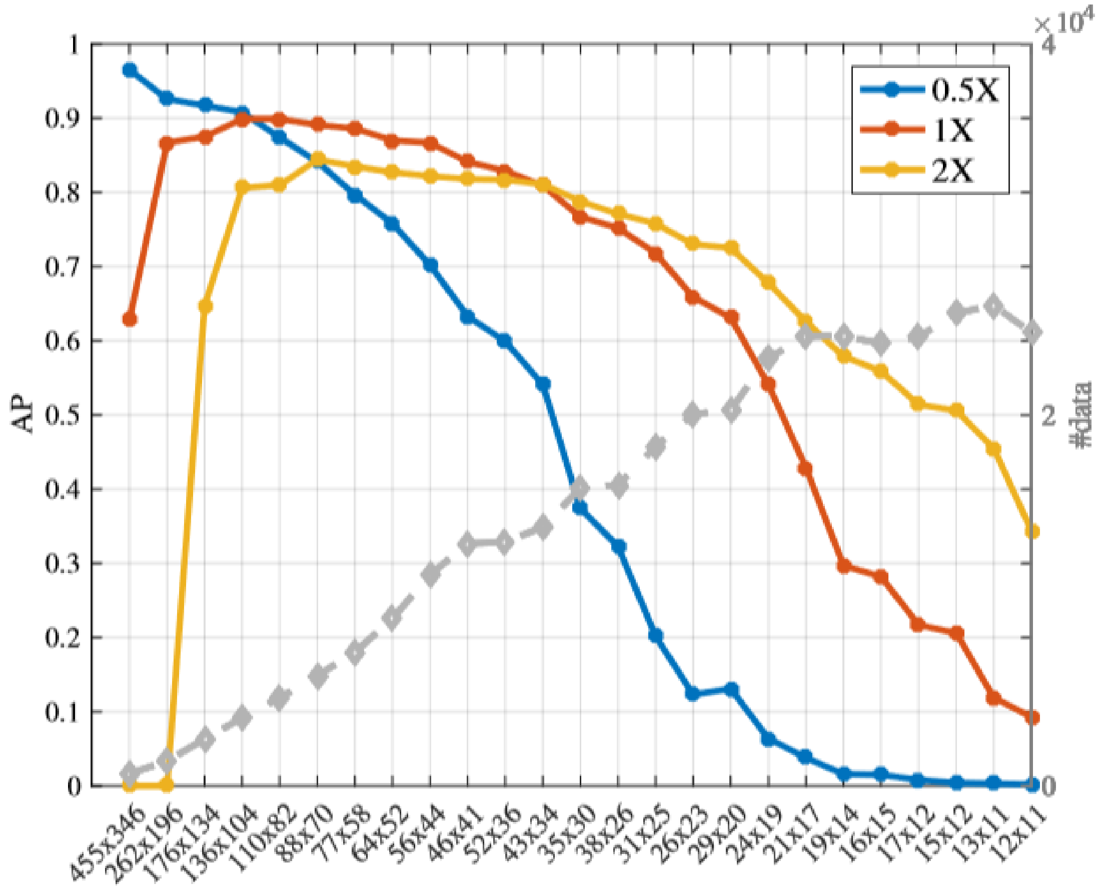
由于ImageNet中前景物体尺寸多数属于中等尺寸(40-140),所以对于模型的预训练对于中尺度物体更敏感,于是导致后续的使用中等大小的模板使得准确率明显升高

图像分辨率(Resolution)



由于训练集的前景物体大小分布问题，发现中等尺寸大小的模板的准确度明显升高，即使模板大小大于物体大小。

这里作者定义了Jaccard distance $d(s_i, s_j) = 1 - J(s_i, s_j)$用来剔除多余模板选取，对于不同尺寸下的物体模板选取有如下的分析。



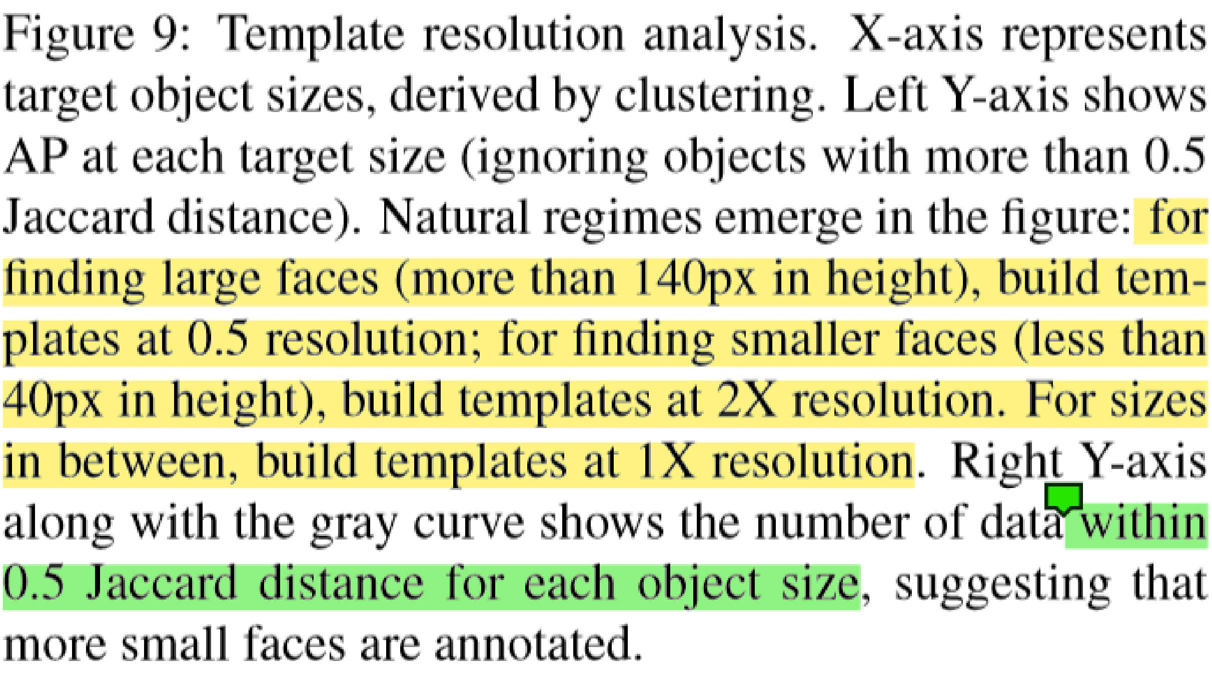


Figure 9: Template resolution analysis. X-axis represents target object sizes, derived by clustering. Left Y-axis shows AP at each target size (ignoring objects with more than 0.5 Jaccard distance). Natural regimes emerge in the figure: for finding large faces (more than 140px in height), build templates at 0.5 resolution; for finding smaller faces (less than 40px in height), build templates at 2X resolution. For sizes in between, build templates at 1X resolution. Right Y-axis along with the gray curve shows the number of data within 0.5 Jaccard distance for each object size, suggesting that more small faces are annotated.

检测大目标（高度大于140px），使用0.5X分辨率的模板。要检测小目标（高度小于40像素）使用2X分辨率的模板。否则，使用相同的（1X）分辨率的模板。 尺寸越小，图像的模板重合概率越高。

优化部分

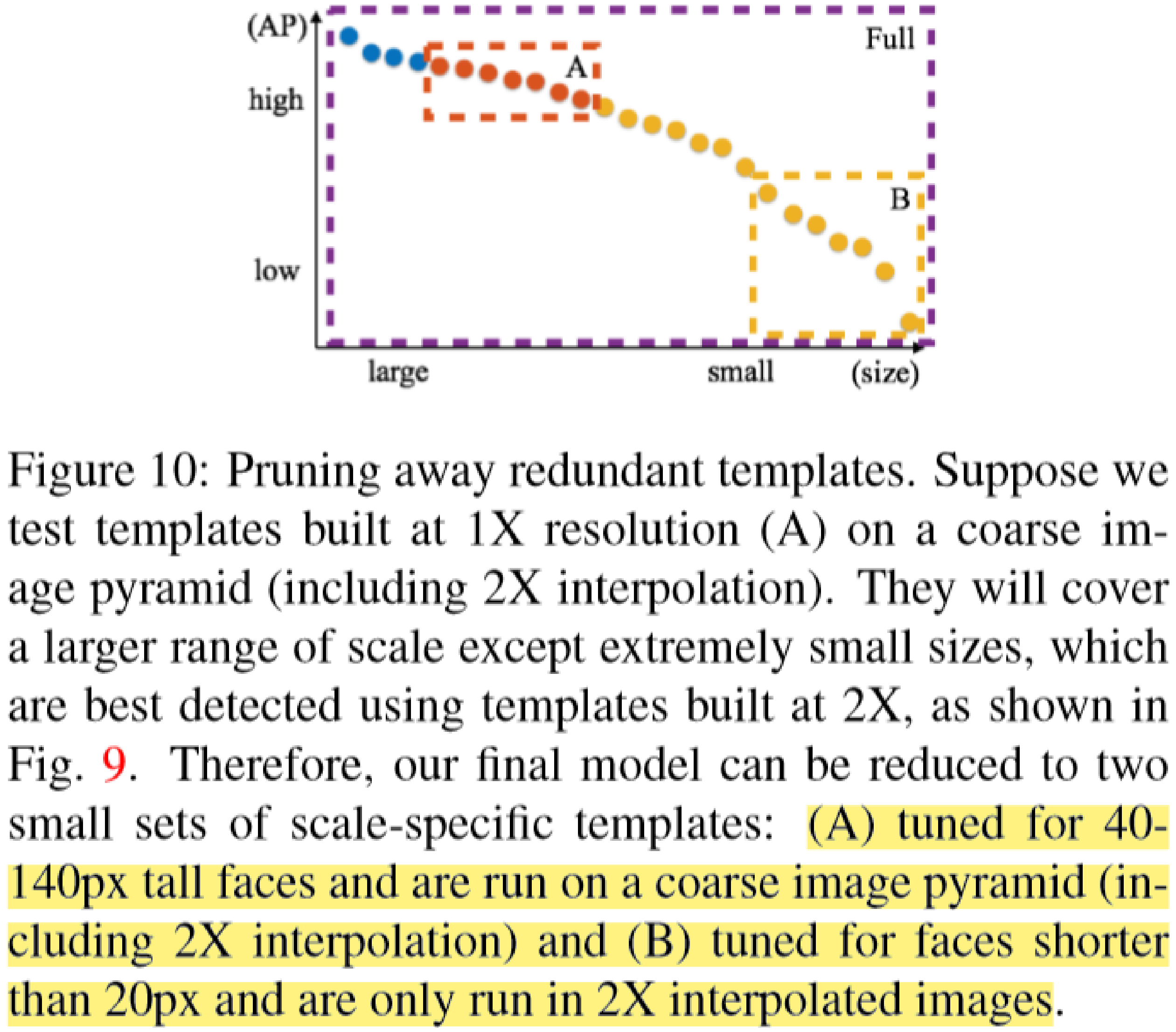
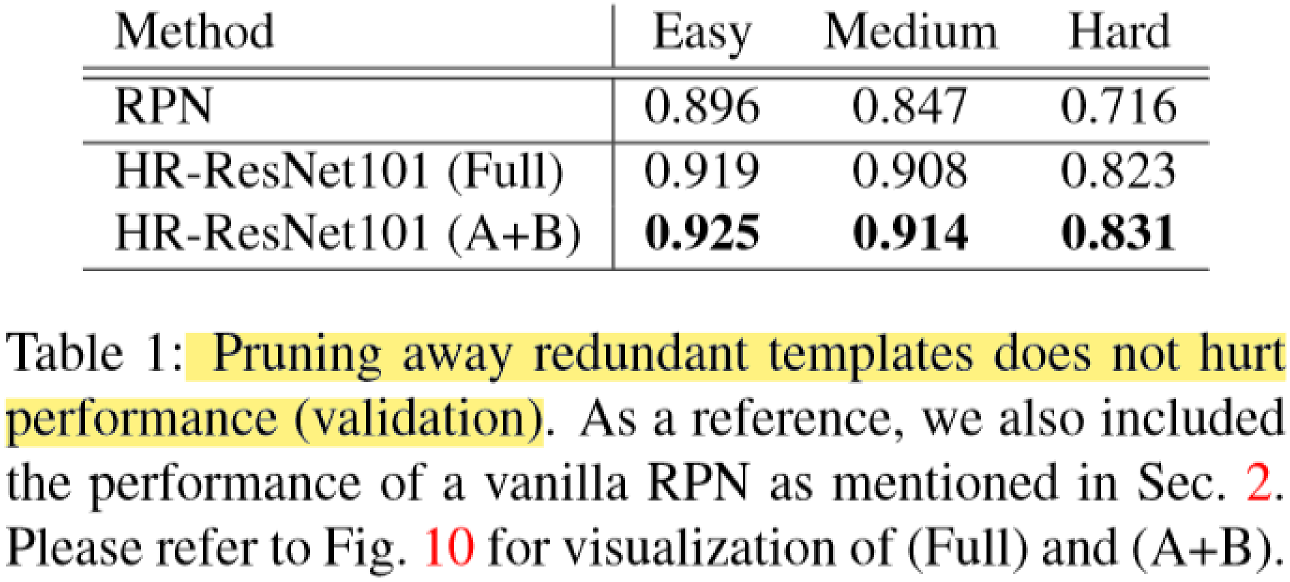


Figure 10: Pruning away redundant templates. Suppose we test templates built at 1X resolution (A) on a coarse image pyramid (including 2X interpolation). They will cover a larger range of scale except extremely small sizes, which are best detected using templates built at 2X, as shown in Fig. 9. Therefore, our final model can be reduced to two small sets of scale-specific templates: (A) tuned for 40-140px tall faces and are run on a coarse image pyramid (including 2X interpolation) and (B) tuned for faces shorter than 20px and are only run in 2X interpolated images.

对于A部分(40-140px)在粗糙图像金字塔包括2X插值进行卷积。对于B部分(>=20px)仅在2X插值进行卷积

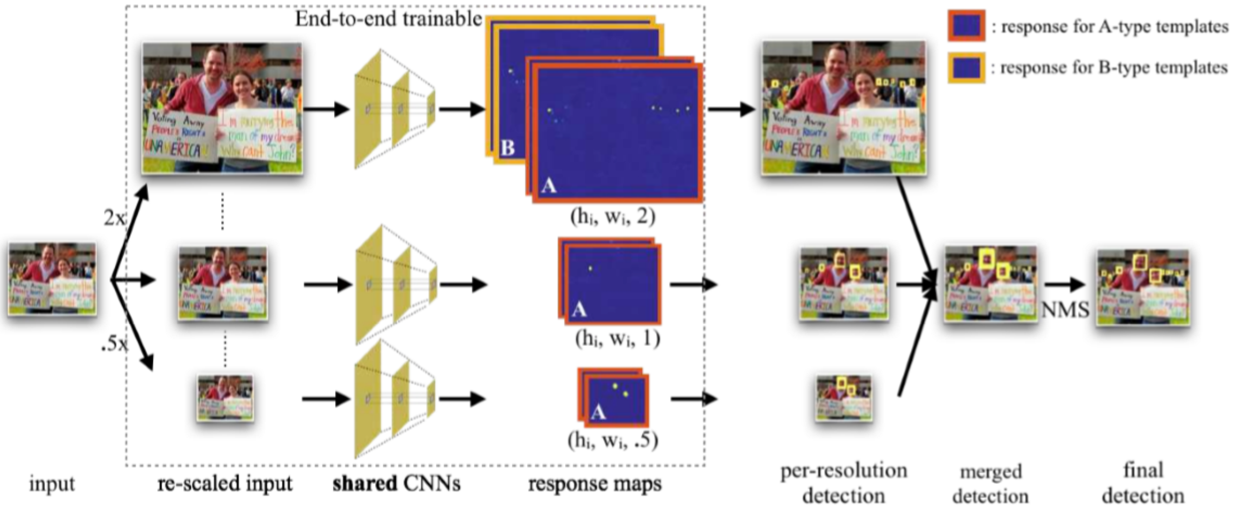


Method	Easy	Medium	Hard
RPN	0.896	0.847	0.716
HR-ResNet101 (Full)	0.919	0.908	0.823
HR-ResNet101 (A+B)	**0.925**	**0.914**	**0.831**

Table 1: Pruning away redundant templates does not hurt performance (validation). As a reference, we also included the performance of a vanilla RPN as mentioned in Sec. 2. Please refer to Fig. 10 for visualization of (Full) and (A+B).

作者证明了多余的模板对结果没有影响，同时比较了只进行A,B两个和全部大小尺寸的精度比较

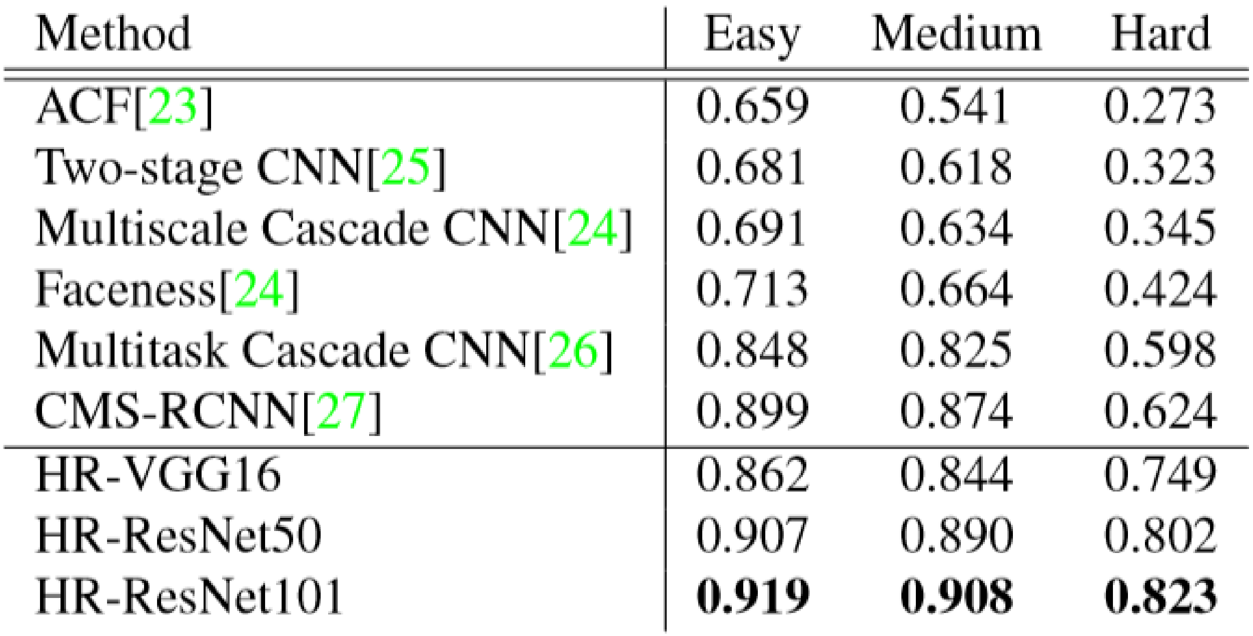
模型结构



1. 创建输入图像的粗糙图像金字塔，包括2X插值
2. 将缩放图像放入共享CNN中，在上述选择条件下进行检测和回归
3. 将三个检测的结果合并，并利用NMS确定位置

其中虚线部分为可训练部分

训练结果



Method	Easy	Medium	Hard
ACF[23]	0.659	0.541	0.273
Two-stage CNN[25]	0.681	0.618	0.323
Multiscale Cascade CNN[24]	0.691	0.634	0.345
Faceness[24]	0.713	0.664	0.424
Multitask Cascade CNN[26]	0.848	0.825	0.598
CMS-RCNN[27]	0.899	0.874	0.624
HR-VGG16	0.862	0.844	0.749
HR-ResNet50	0.907	0.890	0.802
HR-ResNet101	**0.919**	**0.908**	**0.823**