**Deep Learning for Generic Object Detection A Survey阅读摘录**

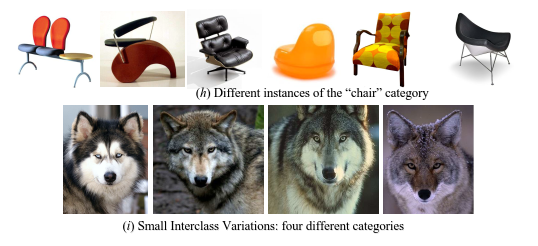
1.Object detection can be grouped into one of two types [69,240]: detection of speciﬁc instance and detection of speciﬁc cat-egories. The ﬁrst type aims at detecting instances of a particular object (such as Donald Trump’s face, the Pentagon building, or my dog Penny), whereas the goal of the second type is to detect different instances of predeﬁned object categories (for example humans, cars, bicycles, and dogs).

**目标检测可以分为两种类型 [69, 240]**：特定实例检测和特定类别检测。前者的目标是检测一个特定目标的实例（比如唐纳德·特朗普的脸、五角大楼建筑或我的狗 Penny），而后者的目标则是检测预定义目标类别的不同实例（比如人、车、自行车和狗）。

2.Generic object detection (i.e., generic object category detection), also called object class detection [240] or object category detection, is defined as follows. Given an image, the goal of generic object detection is to determine whether or not there are instances of objects from many predefined categories and, if present, to return the spatial location and extent of each instance. It places greater emphasis on detecting a broad range of natural categories, as opposed to specific object category detection where only a narrower predefined category of interest (e.g., faces, pedestrians, or cars) may be present. Although thousands of objects occupy the visual world in which we live, currently the research community is primarily interested in the localization of highly structured objects (e.g., cars, faces, bicycles and airplanes) and articulated (e.g., humans, cows and horses) rather than unstructured scenes (such as sky, grass and cloud).

**目标检测的阐述：**一般对象检测(即(通用对象类别检测)，又称对象类检测[240]或对象类别检测，定义如下:对于给定的图像，通用对象检测的目标是确定是否存在来自许多预定义类别的对象实例，如果存在，则返回每个实例的空间位置和范围。它更强调检测范围广泛的自然类别，就像op-构成的特定对象类别检测，其中可能只存在较窄的预定义感兴趣的类别(例如，人脸、行人或汽车)。尽管成千上万的对象占据我们生活的视觉世界,目前研究社区主要感兴趣高度结构化对象的定位(如汽车、脸,自行车和飞机)和铰接(例如,人,牛和马)而非结构化场景(如天空、草地和云)。

3.Challenge



①There is an astonishing variation in what is meant to be a single object class (h). In contrast, the four images in (i) appear very similar, but in fact are from four different object classes

①一个单独的对象类(h)的含义有一个惊人的变化。相比之下，(i)中的四个图像看起来非常相似，但实际上是来自四个不同的对象类

②A further challenge is that of scalability: A detector should be able to handle unseen objects, unknown situations, and rapidly increasing image data. For example, the scale of ILSVRC [179] is already imposing limits on the manual annotations that are feasible to obtain. As the number of images and the number of categories grow even larger, it may become impossible to annotate them manually, forcing algorithms to rely more on weakly supervised training data

②另一个挑战是可伸缩性:检测器应该能够处理不可见的对象、未知的情况和快速增长的图像数据。例如，ILSVRC[179]的规模已经对可以获得的手工注释施加了限制。随着图像的数量和类别的数量越来越多，对它们进行人工注释可能变得不可能，迫使算法更多地依赖于弱监督的训练数据

4.DCNN应用于目标检测算法的两类

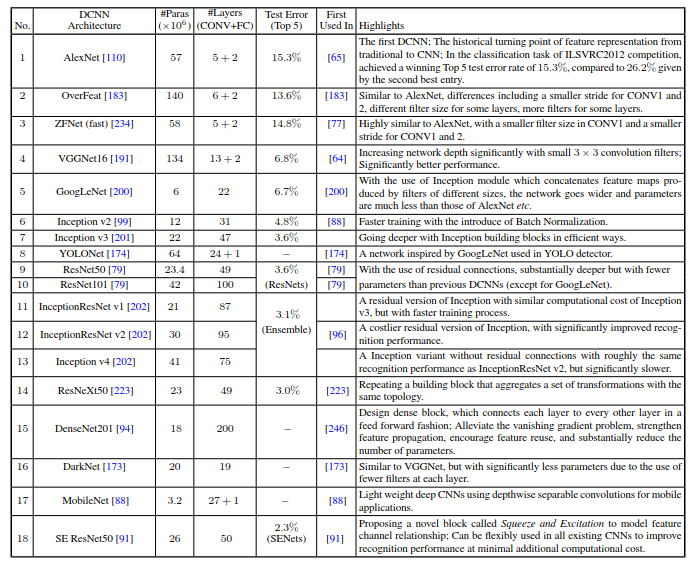
①Region Based (Two Stage Framework)

RCNN、SPPNet、Fast RCNN、Faster RCNN、RFCN、Mask RCNN

②Uniﬁed Pipeline (One Stage Pipeline)

OverFeat、YOLO、YOLOv2、SSD

5.Table2,各种backbone network（基础网络结构）的性能对比



6.the later (or higher) layers have a large receptive ﬁeld and strong semantics, and are the most robust to variations such as object pose, illumination and part deformation, but the resolution is low and the geometric details are lost. On the contrary,the earlier (or lower) layers have a small receptive ﬁeld and rich geometric details, but the resolution is high and is much less sensitive to semantics

卷积神经网络的深层，具有更强的语义信息，但是，几何形状信息相对就比较缺乏了。相反的，浅层具有弱语义，强几何学信息。

multiscale object detection（多次度目标检测）分类:

1) Detecting with combined features of multiple CNN layers [75,

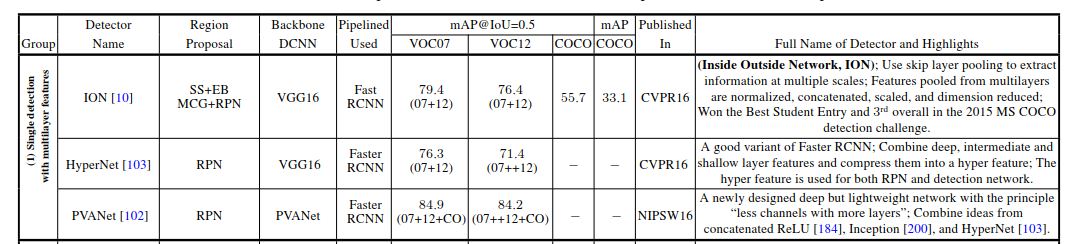
103, 10];

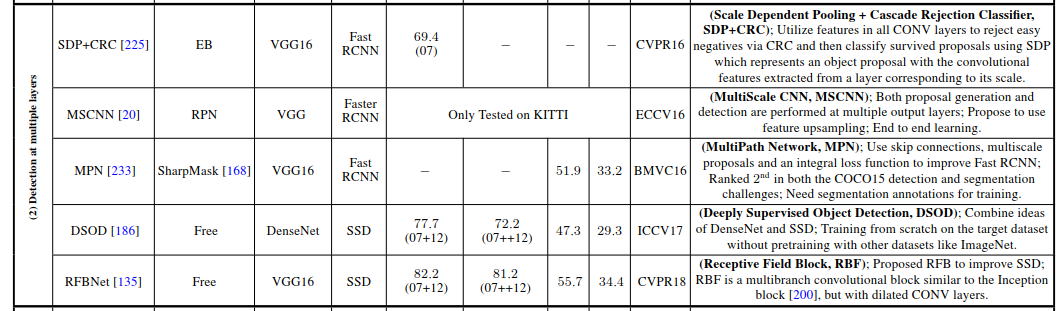
2) Detecting at multiple CNN layers;

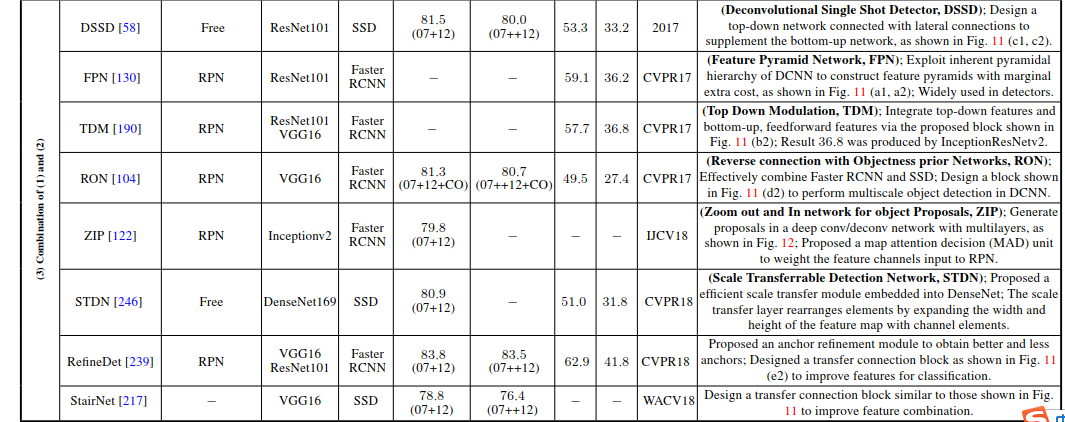
3) Combinations of the above two methods [58, 130, 190, 104,

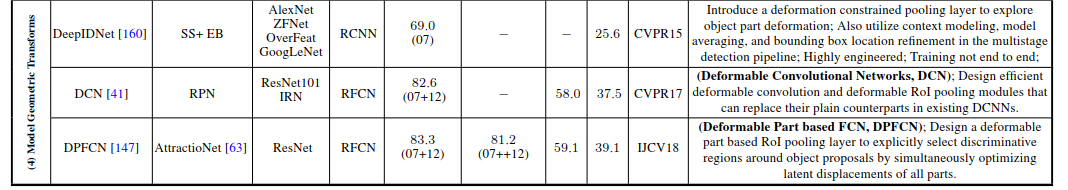
246, 239].

7. Table3,目标检测基于多尺度（multiscale object detection）的改进情况









8.In the physical world visual objects occur in particula environments and usually coexist with other related objects,and there is strong psychological evidence [13, 9] that context plays an essential role in human object recognition.It is recognized that proper modeling of context helps object detection and recognition [203,155, 27, 26, 47, 59]

现实世界的物体之间是有关联的，人根据关联信息识别物体，所提机器也可以学习context信息，来进行目标检测。

9. Context can broadly be grouped into one of three categories [13, 59]:

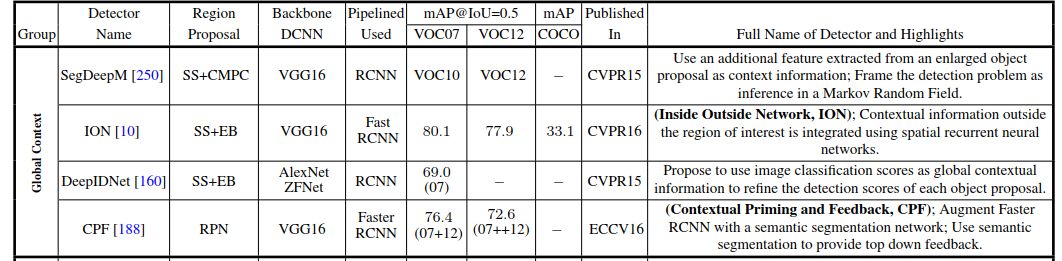
1) Semantic context: The likelihood of an object to be found in some scenes but not in others;

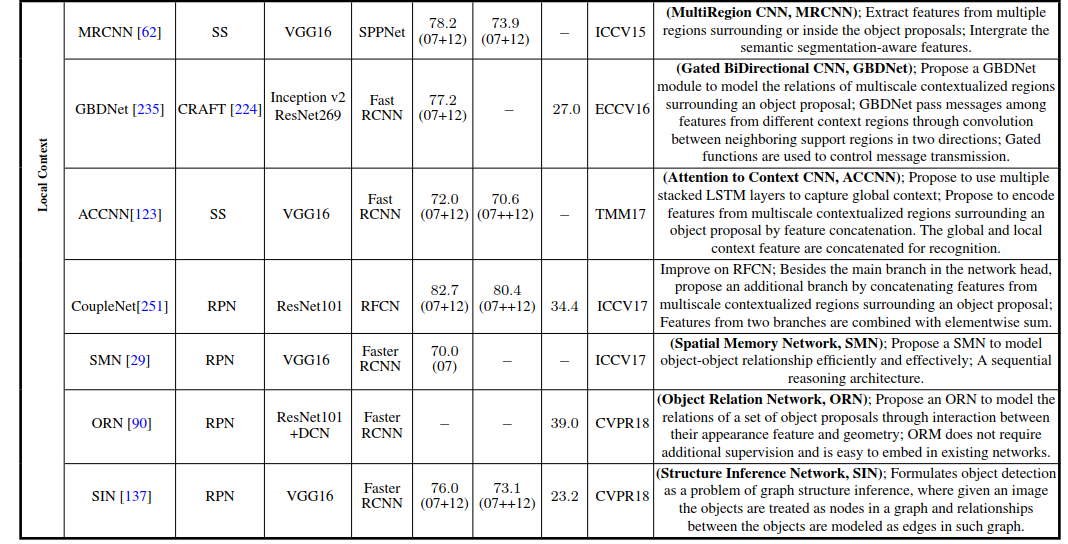
2) Spatial context: Tthe likelihood of ﬁnding an object in some position and not others with respect to other objects in the scene;

3) Scale context: Objects have a limited set of sizes relative to other objects in the scene.

语义context，空间context，规模context

10.Table4，Global context和Local context方面的研究情况





11. Therefore, a good detection proposal should have the following characteristics:

检测提议（detection proposal）通常被用作一个预处理步骤，以通过限制需要检测器评估的区域的数量来降低计算的复杂性。因此，一个优良的检测提议应当具备以下特点：

1)High recall, which can be achieved with only a few proposals;

高召回率，只使用少量提议就可以实现；

2)The proposals match the objects as accurately as possible;

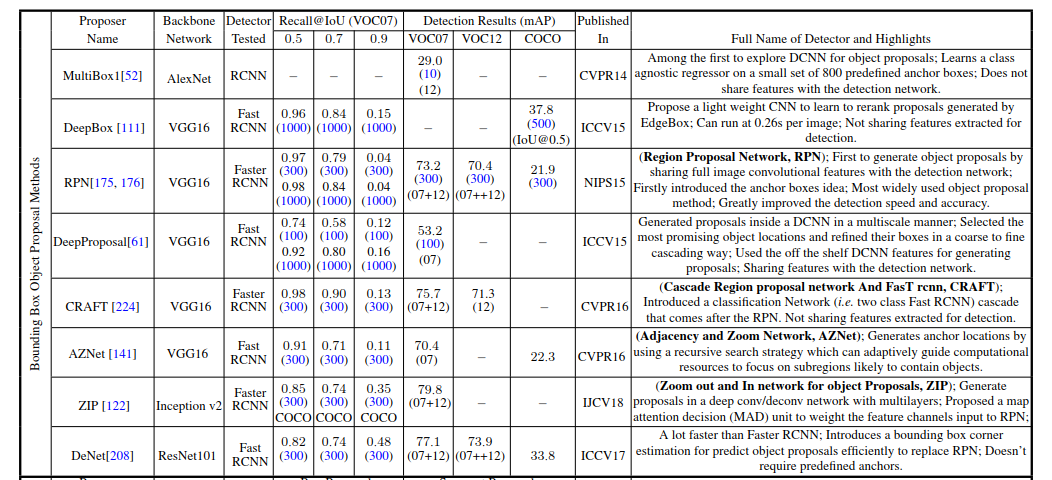
提议尽可能准确地匹配目标；

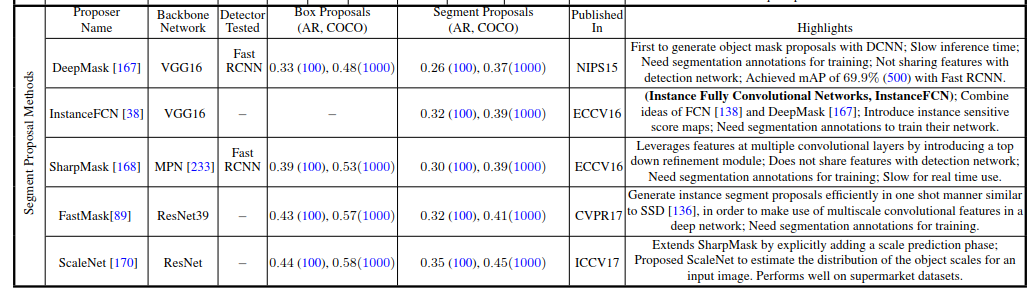
3)High efficiency.

高效率

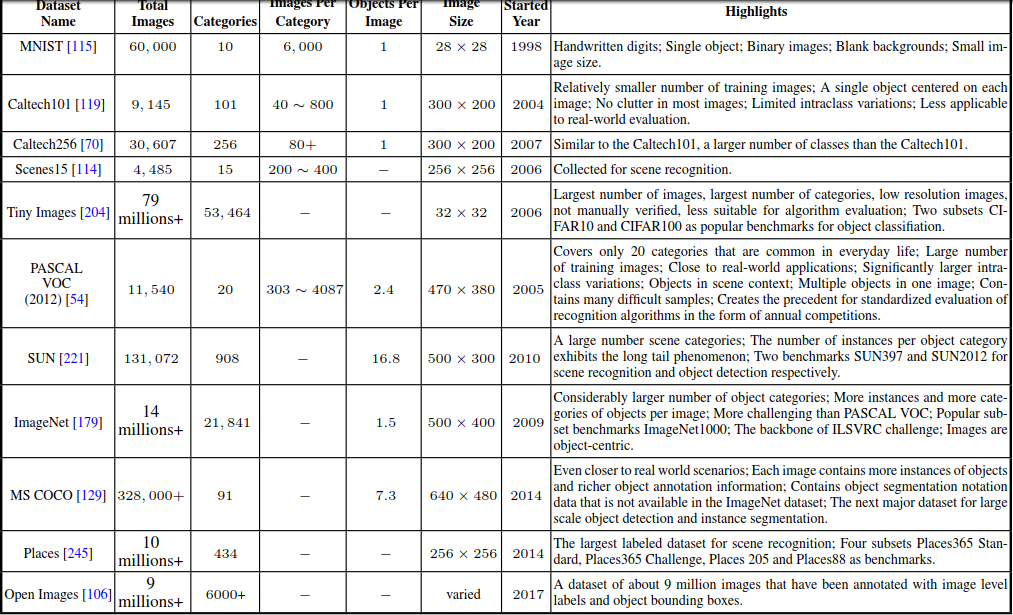
12.区域建议方法分类：Bounding Box Proposal Methods和Object Segment Proposal Methods

13. Table 5使用DCNN 的目标提议方法汇总





14.Table7目标检测的常用数据集



15. There are three criteria for evaluating the performance of detection algorithms: detection speed (Frames Per Second, FPS), precision,and recall. The most commonly used metric is Average Precision(AP), derived from precision and recall. AP is usually evaluated in a category speciﬁc manner, i.e., computed for each object category separately. In generic object detection, detectors are usually tested in terms of detecting a number of object categories. To compare performance over all object categories, the mean AP (mAP) averaged over all object categories is adopted as the ﬁnal measure of performance.

正确率（Accuracy）、精确率（Precision）和召回率（Recall）的区别：

|  |  |  |
| --- | --- | --- |
| 真实  情况 | 预测结果 | |
| 正例 | 反例 |
| 正例 | TP（真正例） | FN（假反例） |
| 反例 | FP（假正例） | TN（真反例） |

A（Accuracy）= （TP+TN）/（TP+TN+FP+FN）

P（Precision）= TP / (TP+FP)

R（Recall）= TP / (TP+FN)

16. Much work remains to be done, which we see focused on the following eight domains:

我们可以看到研究者的关注重点主要集中在以下八个领域

Open World Learning

开放世界学习

Better and More Efﬁcient Detection Frameworks

更好更高效的检测框架

Compact and Efﬁcient Deep CNN Features

紧凑高效的深度 CNN 特征

Robust Object Representations

稳健的目标表征

Context Reasoning

形境推理

Object Instance Segmentation

目标实例分割

Weakly Supervised or Unsupervised Learning

弱监督或无监督学习

3D Object Detection

三维目标检测

17.Table9用于评估目标检测器的常用指标汇总

