In this paper, we propose a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 30% relative to the previous best result on VOC 2012—achieving a mAP of 53.3%

首次利用卷积神经网络，提高了平均准确率，比以往的方法提高了30%，达到了53.3%的准确率。

Two key(1) one can apply high-capacity convolutional neural networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data is scarce, supervised pre-training for an auxiliary task,followed by domain-speciﬁc ﬁne-tuning, yields a signiﬁcant performance boost

两个关键改进点：(1)我们可以将大容量卷积神经网络(CNNs)应用到自底向上的区域选择中，实现目标的定位和分割(2) 当标记的训练数据不足时，对辅助任务进行有监督的预先训练，然后进行特定于领域的微调，可以显著提高性能

Since we combine region proposals with CNNs, we call our method R-CNN: Regions with CNN features

While R-CNN is agnostic to the particular region proposal method, we use selective search to enable a controlled comparison with priordetection work

将区域建议与CNN结合在一起，故称为R-CNN

J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders.Selective search for object recognition. IJCV, 2013.

The ﬁrst generates category-independent region proposals.These proposals deﬁne the set of candidate detections available to our detector. The second module is a large convolutional neural network that extracts a ﬁxed-length feature vector from each region. The third module is a set of class-speciﬁc linear SVMs

R-CNN模型分为三个部分：1.区域选择，2.卷积神经网络提取特征，3.支持向量机

Two properties make detection efﬁcient. First, all CNN parameters are shared across all categories. Second, the feature vectors computed by the CNN are low-dimensional when compared to other common approaches, such as spatial pyramids with bag-of-visual-word encodings.

使检测高效的两个性质：1.神经网络为所有类别共享权值，2.卷积后的特征向量维度低。