**丰富的特征层次结构用于精确的目标检测和语义分割**

**Rich feature hierarchies for accurate object detection and semantic segmentation**

Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik

CVPR'14

**一、科学问题**

**1.1 本文所涉及科学问题**

首次将神经网络用于进行目标检测和语义分割

**1.2 同行专家如何解决**

在作者提出该方法以前，大家普遍都是才用 构造特征+SVM的方法

**1.3 本文所解决的问题**

1将大型的卷积神经网络用于目标检测和语义分割

2提出了迁移学习的概念

**1.4 本文解决方案效果**

对于之前VOC 2012的最佳结果，平均精度（MAP）提高了30%以上，实现了53.3%的MAP。

**二、研究内容**

**2.1 理论与方法介绍**

该系统主要由三部分组成

1 采用selective search进行Region proposals

2 将每个region proposal填充为227\*227，输入卷积神经网络中进行特征提取

3 利用SVM对region proposal的特征向量进行分类并训练边框回归

该作者还提出了迁移学习的概念，先在ImageNet上进行权重训练，然后在用自己的数据进行微调，避免数据少容易产生过度拟合的问题。

**2.2 验证分析与实验效果**

对于之前VOC 2012的最佳结果，平均精度（MAP）提高了30%以上，实现了53.3%的MAP。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

对每一个region proposal都进行单独的特征提取，并没有进行共享运算，训练和检测的速度都比较慢。

在训练的过程中模型都权重都写入磁盘中，占用内存比较大。

**3.2 后续研究重点**

对模型的速度和精确度进一步的提高。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：Fast R-CNN

**（2）作者介绍**：Ross Girshick Microsoft Research

**（3）摘要**: 本文提出了一种基于快速区域卷积网络的目标检测方法。快速R-CNN建立在以前的工作基础上，利用深度卷积网络有效地对目标提案进行分类。与之前的工作相比，Fast R-CNN采用了一些创新技术来提高培训和测试速度，同时也提高了检测精度。快速的R-CNN训练非常深的VGG16网络，比CNN快9倍，测试时快213倍，在Pascal VOC 2012上获得更高的地图。与SPPNET相比，快速R-CNN可以更快地训练VGG16 3×更快，测试10×更快，并且更准确。

**4.2 相关论文二**

**（1）题目**：Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

**（2）作者介绍**：Shaoqing Ren∗ Kaiming He Ross Girshick Jian Sun

**（3）摘要**：最先进的物体检测网络依靠区域建议算法来假设物体的位置。SPPNET[7]和Fast R-CNN[5]等技术的进步减少了这些检测网络的运行时间，使区域建议计算成为一个瓶颈。在这项工作中，我们引入了一个区域建议网络（RPN），它与检测网络共享全图像卷积特性，从而实现了几乎免费的区域建议。RPN是一个完全卷积的网络，它同时预测每个位置的对象边界和对象性得分。RPN经过端到端的培训，生成高质量的区域建议，由快速R-CNN用于检测。通过简单的交替优化，RPN和快速的R-CNN可以被训练来共享卷积特征。对于非常深的VGG-16型号[19]，我们的检测系统在GPU上的帧速率为5fps（包括所有步骤），同时在Pascal VOC 2007（73.2%地图）和2012（70.4%地图）上使用300个图像方案实现最先进的目标检测精度。

**4.3 相关论文三**

**（1）题目**：Object Detection in 20 Years: A Survey

**（2）作者介绍**：Zhengxia Zou, Zhenwei Shi, Member, IEEE, Yuhong Guo, and Jieping Ye, Senior Member, IEEE

**（3）摘要**：目标检测作为计算机视觉中最基本、最具挑战性的问题之一，近年来受到了广泛的关注。它在过去二十年的发展可以看作是计算机视觉历史的缩影。如果我们把今天的物体探测看作是一种在深度学习的力量下的技术美学，那么把时钟倒转20年，我们将见证冷武器时代的智慧。本文从技术发展的角度对400多篇对象检测论文进行了广泛的回顾，跨越了四分之一世纪的时间（从90年代到2019年）。本文涵盖了许多主题，包括历史上的里程碑探测器、检测数据集、度量、检测系统的基本构建块、加速技术和最新的检测方法。本文还综述了行人检测、人脸检测、文本检测等重要的检测应用，并对其面临的挑战和近年来的技术进步进行了深入分析。

**Fast R-CNN**

Ross Girshick Microsoft Research

**一、科学问题**

**1.1 本文所涉及科学问题**

R-CNN在目标检测上精度和性能有了大幅度的提升，但还是存在很多问题。多阶段（训练CNN，训练SVM，训练bb回归器）；训练时空间和时间代价高（对于SVM和bb回归器需要把每张图像的每个候选区特征通过CNN提取出来存到磁盘）；预测阶段很慢（因为要对每张图像的每个候选区域提取特征）。

**1.2 同行专家如何解决**

SPPnet的改进：R-CNN预测慢是因为对于每个候选区域要过一次CNN而不共享计算，而SPPnet则是使用共享计算来加速。如下第一个图所示，SPP输入整张图，计算一个feature map，在feature map上找到候选区对应的一个子图，然后对子图做一个金字塔池化得到一个固定长度的特征向量（下图中得到的是4x4+2x2+1个特征）。 尽管如此，SPPnet仍然有多阶段以及特征写入磁盘的缺点，而且因为SPPnet的特征是从conv5的feature map上池化而来的，所以只fine-tuning金字塔池化层之后的层（这限制了深层网络的准确性）。

**1.3 本文所解决的问题**

Fast R-CNN修正了R-CNN和SPPnet（Spatial pyramid pooling）的缺点，提升了速度和准确率。它是一个单阶段的算法，使用了multi-task loss，可以对整个网络进行更新，而且不需要把特征存到磁盘上。

**1.4 本文解决方案效果**

在Pascal VOC 2012[7]上获得最高精度，map为66%（而R-CNN为62%）。

**二、研究内容**

**2.1 理论与方法介绍**

输入为一整张图以及一系列的（selective search生成的）候选区域（映射到conv5的feature map上得到ROI），ROI意为Region of Interest。像SPPnet一样对ROI进行池化，只不过这里是单水平的金字塔池化（如下第一个图是三水平），比如分成7x7个子图对每个子图取最大得到长度为49个ROI特征。池化后经过FC，然后分成两个分支，一个分支用softmax做分类，一个分支用bb回归器做定位，使用multi-task loss进行训练。

**2.2 验证分析与实验效果**

在Pascal VOC 2012[7]上获得最高精度，map为66%（而R-CNN为62%）。

**三、论文存在问题及后续研究重点**

**3.1 论文存在问题**

Fast R-CNN的瓶颈在于生成候选区域（Selective Search的）的方法非常耗时

**3.2 后续研究重点**

对模型的精度和性能进一步提升。

**四、该问题相关研究成果**

**4.1 相关论文一**

**（1）题目**：Fast R-CNN

**（2）作者介绍**：Ross Girshick Microsoft Research

**（3）摘要**: 本文提出了一种基于快速区域卷积网络的目标检测方法。快速R-CNN建立在以前的工作基础上，利用深度卷积网络有效地对目标提案进行分类。与之前的工作相比，Fast R-CNN采用了一些创新技术来提高培训和测试速度，同时也提高了检测精度。快速的R-CNN训练非常深的VGG16网络，比CNN快9倍，测试时快213倍，在Pascal VOC 2012上获得更高的地图。与SPPNET相比，快速R-CNN可以更快地训练VGG16 3×更快，测试10×更快，并且更准确。

**4.2 相关论文二**

**（1）题目**：Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

**（2）作者介绍**：Shaoqing Ren∗ Kaiming He Ross Girshick Jian Sun

**（3）摘要**：最先进的物体检测网络依靠区域建议算法来假设物体的位置。SPPNET[7]和Fast R-CNN[5]等技术的进步减少了这些检测网络的运行时间，使区域建议计算成为一个瓶颈。在这项工作中，我们引入了一个区域建议网络（RPN），它与检测网络共享全图像卷积特性，从而实现了几乎免费的区域建议。RPN是一个完全卷积的网络，它同时预测每个位置的对象边界和对象性得分。RPN经过端到端的培训，生成高质量的区域建议，由快速R-CNN用于检测。通过简单的交替优化，RPN和快速的R-CNN可以被训练来共享卷积特征。对于非常深的VGG-16型号[19]，我们的检测系统在GPU上的帧速率为5fps（包括所有步骤），同时在Pascal VOC 2007（73.2%地图）和2012（70.4%地图）上使用300个图像方案实现最先进的目标检测精度。

**4.3 相关论文三**

**（1）题目**：Object Detection in 20 Years: A Survey

**（2）作者介绍**：Zhengxia Zou, Zhenwei Shi, Member, IEEE, Yuhong Guo, and Jieping Ye, Senior Member, IEEE

**（3）摘要**：目标检测作为计算机视觉中最基本、最具挑战性的问题之一，近年来受到了广泛的关注。它在过去二十年的发展可以看作是计算机视觉历史的缩影。如果我们把今天的物体探测看作是一种在深度学习的力量下的技术美学，那么把时钟倒转20年，我们将见证冷武器时代的智慧。本文从技术发展的角度对400多篇对象检测论文进行了广泛的回顾，跨越了四分之一世纪的时间（从90年代到2019年）。本文涵盖了许多主题，包括历史上的里程碑探测器、检测数据集、度量、检测系统的基本构建块、加速技术和最新的检测方法。本文还综述了行人检测、人脸检测、文本检测等重要的检测应用，并对其面临的挑战和近年来的技术进步进行了深入分析。