第十四届中国研究生电子设计竞赛 技术论文

论文题目：

基于深度学习的水果识别分类

Fruit recognition and classification based on deep learning

参赛单位：华中师范大学物理科学与技术学院

队伍名称：仨女一男

指导老师：张新晨

参赛队员：詹鹏杰 占妮 田新春 汪燎原

完成时间：2019年5月28号

# 摘 要

图像识别作为模式识别领域中一种典型的应用，如何快速地识别图像一直是全世界学者研究的话题。水果识别也一直是图像识别技术的一种典型的应用。本文主要利用卷积神经网络中的Mask R-CNN模型来进行水果识别，对水果进行检测分类。目前我们常见的水果识别仍以人工检测分类为主，花费大量的人力财力，且存在很多人为因素的影响导致检测效率低、分类错误高；简单的机器检测虽然解放了一定的劳动，但是还需要人来协助，在一定程度上也会受人为因素干扰。本文是在深度学习框架下利用图形处理技术来实现自动识别分类，用来提高水果检测分类的准确率和效率。

**关键词：**图像识别；深度学习；水果识别

# Abstract

Image recognition is a typical application in the field of pattern recognition. How to quickly recognize images has always been a topic for scholars all over the world. Fruit recognition has also been a typical application of image recognition technology. This paper mainly uses Mask R-CNN model of convolutional neural network to recognize fruits and detect and classify fruits. At present, our common fruit recognition still focuses on manual detection and classification, which costs a lot of manpower and financial resources, and there are many human factors that lead to low detection efficiency and high classification errors. Simple machine detection has liberated a certain amount of labor, but it still needs human assistance, and to a certain extent, it will be interfered by human factors. This paper uses graph processing technology to realize automatic recognition and classification under the framework of in-depth learning, which is used to improve the accuracy and efficiency of fruit detection and classification.

**Key words:** image recognition; deep learning; fruit recognition

**目 录**

摘 要Ⅰ

AbstractⅡ

第一章 绪论1

1.1 课题研究的背景及意义1

1.2 作品难点与创新1

第二章 原理分析3

2.1 深度学习3

2.2 深度学习基本方法3

2.2.1 卷积神经网络3

2.2.2 卷积神经网络的基本网络结构4

2.2.3 Mask R-CNN4

第三章 软件设计6

3.1 特征提取网络结构设计6

3.2 候选窗口分类器结构设计6

3.3 数据集选取7

第四章 实验结果及分析8

4.1 对比模型介绍8

4.2 模型性能对比8

4.3 同种水果的检测9

4.4 不同种水果的检测10

4.5 实时视频检测11

4.6 轻量模型移植到Android平台检测12

4.7 结果分析12

第五章 总结与展望13

参考文献14

# 第一章 绪论

## 1.1 课题研究的背景及意义

我国是农业生产大国，土地面积居于世界前列，这必然使得我国成为世界上水果产量排前列的国家。总管目前形势，深度学习已成为当前发展潮流，因此水果识别技术在我国当有广阔的发展前景。

目前我国现有的水果检测方法仍然是以人工检测分类为主，通过人眼对水果尺寸的大小、果皮颜色、果皮纹路等明显特征的判断来判别水果的种类。这种判别的方法需要依靠大量的人力财力，并且准确率与效率还极具不稳定性；此外，检测分类的结果容易受到人们主观因素的影响，比如人眼视力、情绪状况、身体状况等因素。基于深度学习框架下的图像自动识别技术可以检测到人眼看不到的东西，弥补了人眼可见范围有限的缺点；还可以不受主观因素的影响，不知疲倦地工作，并且能保持较高的准确率和效率。这是人工检测达不到的高度，极具发展前景。

深度学习是机器学习的一个主要分支，它来源于对人工神经网络的研究，由Hinton等人于2006年提出。它是由包含复杂结构或由多重非线性变换构成的多个处理层对数据进行高层抽象的算法。早在上世纪80年代卷积神经网络就已经用于商业用途。当Lecun等人在1998年发明LeNet网络时，它被认为是深度学习最初的网络模型。目前深度学习的应用场景涵盖图像识别、无人驾驶、自然语言处理、智能搜索引擎以及机器人技术等，深度学习对这些领域产生了巨大、深远的影响。本文采用的是在深度学习的框架下基于卷积神经网络的方法来对水果进行识别分类。

## 1.2 作品难点与创新

随着经济水平的不断改善，人们追求更加健康的生活方式，水果的地位受到了大众的肯定，人们对水果的消费需求量逐年上升。传统的人工分拣识别存在人均工作量大、分级标准有效执行困难的情况。人在工作量大的状态下容易受情绪、疲劳等因素的影响，从而产生较大的主观性认知偏差进而使得分拣效率低下，还容易出现漏拣和误拣等问题。而简单的机械式水果分拣识别系统存在容易因与水果表面接触损坏水果品质的缺点。我们的作品针对以上不足做了以下改进：

（1）基于目前的现状我们希望采用非接触式的方法，利用摄像头来模拟人的眼睛，借由计算机的算法来实现决策，使得识别更加准确快捷。我们在深度学习的框架下，利用卷积神经网络算法来训练数据集，从而提高水果识别的准确性。

（2）在卷积网络的选取上，我们没有采用常规的卷积神经网络算法，而是选取了Mask R-CNN算法，并在Mask R-CNN算法的基础上进行了一些改进：1)设计特征提取的网络结构，通过网络层数的减少提高了算法的效率；2)对候选窗口分类器结构进行设计，采用双线性插值减小了特征提取感兴趣区域时的误差，提高了水果检测的准确率。

（3）相比于传统物品识别，Mask R-CNN可以准确提取物品轮廓，对被遮蔽和阻挡的物品检测性能更强。

（4）如果将模型轻量化，有望迁移到Android移动平台如手机等设备进行识别检测。

该作品在实施的过程中也存在着一些困难。当水果堆叠在一起时，被遮蔽的水果，检测起来就会具有一定的难度，无法保证所有水果都被检测识别出来。

# 第二章 原理分析

## 2.1 深度学习

在实际的识别问题中，我们要考虑的是如何对识别对象进行有效地表示，用具有更高层次的特征代替初始的像素。在一幅图像中，我们很难在像素级别的特征中发现区分性的有用信息，我们需要的是具有结构性的特征来进行区分，只有这样，学习算法才能发挥作用。传统算法所提取的纹理、几何等较高层次的特征在分类识别任务中表现出了极大的优势，但是这些传统方法提取的特征的好坏在很大程度上依靠经验和运气，这就造成结果的不稳定性。研究人员开始思考能否让计算机模仿人的视觉机理，通过无监督的训练方式从目标中自动学习本质的特征。而深度学习作为一种无监督与监督相结合的自动特征提取算法，通过构建深层网络来实现对数据的分布式表示，并从中自动地学习数据中丰富的内在信息，提取出其中更具表达力和推广性的特征。

深度学习的思想最早是受到人脑分层运算的启发而提出的。生理学家研究发现人的大脑计算是分多层进行的，比如人的视觉系统在处理图像上就是分级的，对原始图像进行边缘特征的提取，通过对这些低级的特征进行组合获得高层次的局部特征，通过对上一级的特征进行组合获得更高层次的目标及行为，这一层层的组合使得特征变得越来越抽象，语义信息更加明显。2006年Hinton首次提出深度学习的概念，通过构建类似于人脑的多层网络模型来实现对目标的分层抽象表达，摆脱了人工提取特征的诸多限制，让机器能够自动地完成对特征的提取和选择，减少了人为因素的影响，使深度学习开始广泛地应用于模式识别、计算机视觉、语音识别等领域。

深度学习的本质就是深层网络借助于非线性信息处理机制，通过监督与无监督相结合的训练方式来实现对特征的提取和转换，借助于分布式特征表示来实现对样本间复杂数据关系的拟合。相对于浅层学习模型不能对复杂函数进行有效表示和在处理复杂分类问题时泛化性能不足的局限性，深度学习能够实现对复杂函数的逼近并对观测样本进行拟合，从而在数据中学习到概念化的本质特征。另外，这种深层网络结构能在一定程度上弱化前一层网络所提取的错误特征，并且可以用较少的参数对复杂函数进行表示，使得网络计算的复杂度更加紧凑，从而具有较高的效率和效果。

## 2.2 深度学习基本方法

### 2.2.1 卷积神经网络

卷积神经网络（Convolution Neural Network，CNN）作为目前计算机视觉领域中应用最广的深度学习模型，它最早是一种由日本学者Fukushima仿照生物视觉系统的机理提出的识别算法，后来经过LeCun、Simard和Steinkraus等人的改进和简化，识别性能进一步提升。CNN的模型与BP神经网络的模型类似，都选用了前向传播计算输出值、反向传播调整权和偏置的方法；但它们两者之间的最大差异是CNN中相邻两层的神经元之间采用部分连接的方式，而不是BP神经网络中采用的全连接方式。另外，CNN作为一种深层神经网络，也不会像传统的深层神经网络一样存在不可计算的问题。CNN能够直接对二维图像进行特征提取，利用BP算法对网络中的参数进行优化，相比于一般的传统神经网络，CNN在图像识别方面显示出较大的优势：(1)直接将二维图像作为输入避免了复杂的预处理过程；(2)特征提取与模式分类同时进行，将特征提取、模式分类结合在一起，通过不断优化来获得所需参数，在输出层给出结果；(3)借助于稀疏连接和权值共享的方式使得网络中需要训练的参数数目大大降低，泛化性能进一步增强。

### 2.2.2 卷积神经网络的基本网络结构

(1)卷积层：卷积层的目的是对输入的图像进行特征提取，如下图所示，卷积操作即对图像和卷积核做内积运算。绿色的5\*5单元格表示一幅图片，其中3\*3黄色的单元格为卷积核，假设做步长为1的卷积操作，那么卷积核每次向右移动一个格子。卷积核内每个单元格的右下角为权重，3\*3的卷积核每移动一次，将图像上的像素和对应的权重相乘再将所有结果相加得到对应结果，从图中可以看出，经过一次卷积操作图像从5\*5变成了3\*3。

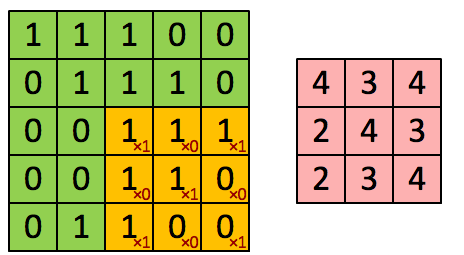
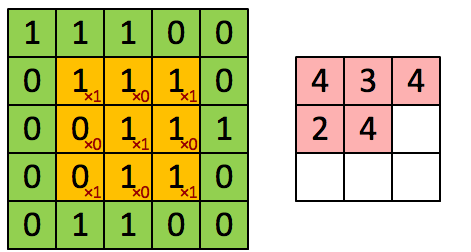
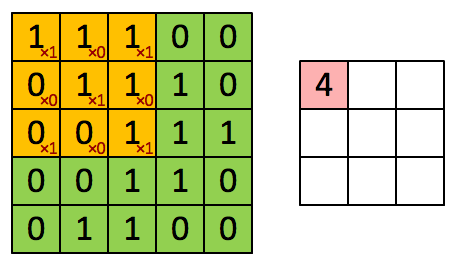


图2-1

(2)池化层：通过卷积层提取特征，可以直接将提取到的特征输入到分类器中，但是提取到的特征数量依旧非常庞大。因此通过池化层对特征图像进行压缩，以达到简化网络计算的复杂度并提取主要特征[1]-[2]。池化层通常有以下几个作用：

1）保持某种不变性，包括平移、旋转、尺度；

2）防止过拟合现象的发生；

3）保留主要的特征，同时减少下一层的参数和计算量。

(3)激活函数层：激活函数的目的是通过将非线性因素加入神经网络，以此来更好地解决较为复杂的问题。图像的分类明显是非线性的，而在对图像进行卷积操作时是将像素值乘以卷积核的权值并求和，这个操作显然是线性的，因此需要非线性映射。经典的激活函数有Sigmoid函数、Tanh函数、ReLU函数。

(4)全连接层：全连接层在整个卷积神经网络中起到“分类器”的作用。前面提到的各层是将原始数据映射到特征空间，全连接层则将分布式特征映射到样本标记空间[3]-[4]，从而达到分类的目的。在实际的应用过程中，可以通过卷积操作实现全连接层，对于前一层是卷积层的全连接层，可以通过大小为前一层宽和高组成的卷积核进行全局卷积实现。而对于前一层是全连接的全连接层，可以通过单位大小的卷积核进行卷积操作实现。

### 2.2.3 Mask R-CNN

Mask R-CNN是一个实例分割模型，它不仅可以识别出图像中各个物体的种类和具体的位置信息，还能对物体进行像素级别的轮廓描述，即实例分割。实例分割不仅可以对同一类别感兴趣的区域进行分割，还可以对不同类别同时进行分割。所以相比于Faster R-CNN，Mask R-CNN的主要不同在于其用像素级别的轮廓描述标出目标物体在图像中的位置。

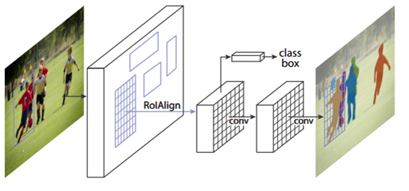


图2-2

Faster R-CNN对于每个目标对象都有目标分类标签和候选窗口两个输出。Mask R-CNN为了进行像素级别的实例分割，在此基础上增加了指示对象像素位置的二进制掩模（mask）。此输出需要更精细的位置信息，为此Mask R-CNN在Faster R-CNN上添加一个分支网络——全卷积网络FCN。

# 第三章 模型设计

## 3.1 特征提取网络结构设计

传统的算法通常是通过加宽或者加深网络的方式来提高准确率，但此做法会导致超参数数量增多，使得网络设计变得困难，而且计算量也随之增多。因此ResNeXt被提出，其网络结构简明，可以在不增加参数的情况下提高准确率，并且还因此缩小了超参数的数量。所以本文将ResNeXt-101-FPN[6]-[7]网络结构应用在Mask R-CNN中，在保证精度的同时减少计算量（一个101层的ResNeXt网络，和200层的ResNet准确度相似，但其计算量只有后者的一半）并降低复杂度。低层中有细节和语义，ResNeXt-101-FPN网络结构可以将高层中的特征传入低层中去，所以对特征提取的性能影响较大。因此本文选用ResNeXt-101-FPN网络结构，并在此基础上将收集到的数据集作为训练集进行训练。

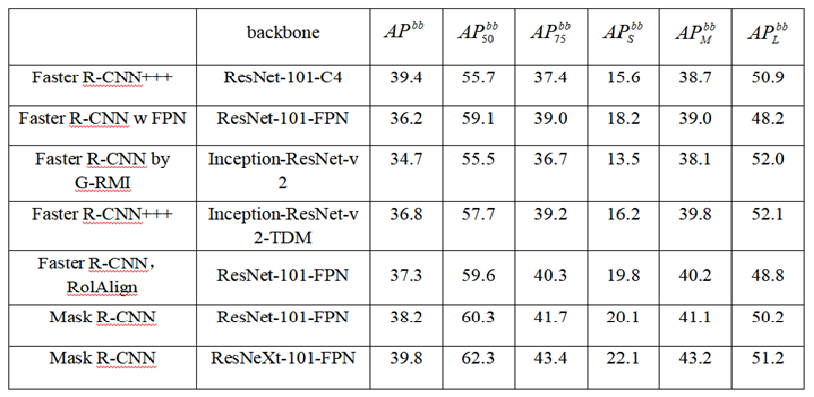


图3-1

## 3.2 候选窗口分类器结构设计

Faster R-CNN采用RoI Pooling层完成量化的操作，但是因为Faster R-CNN的结构中包含了全连接层，所以量化操作需要在全连接层前完成。通过过量化操作后，即可处理任意分辨率的图像。但是这种方法不能使输入图像的像素和输出图像的像素保持一致，虽然对分类影响不会太大，但是对像实例分割这样像素级的操作会有很大的影响。

为了解决这一问题，Mask R-CNN采用RoIAlign替代RoI Pooling。通过对比RoI Pooling和ROIAlign可以发现，RoI Pooling在进行将原图映射到特征图时，进行了对应的取整，使得存在了误差。而此误差会使得特征图在反向映射回原图的时候在经过最大池化层后，特征与原图之间空间存在明显偏差。因此，Mask R-CNN的ROIAlign层去除了所有的量化过程，目标检测支路依旧和RoI Pooling相同，而目标分割支路则采用14\*14的新尺寸，并且从原图到特征图的ROI映射使用双线性插值，不取整，以降低误差，经过池化后再对应回原图的准确性也更高。本文设计的结构采用双线性插值来减小特征提取感兴趣区域时的误差。

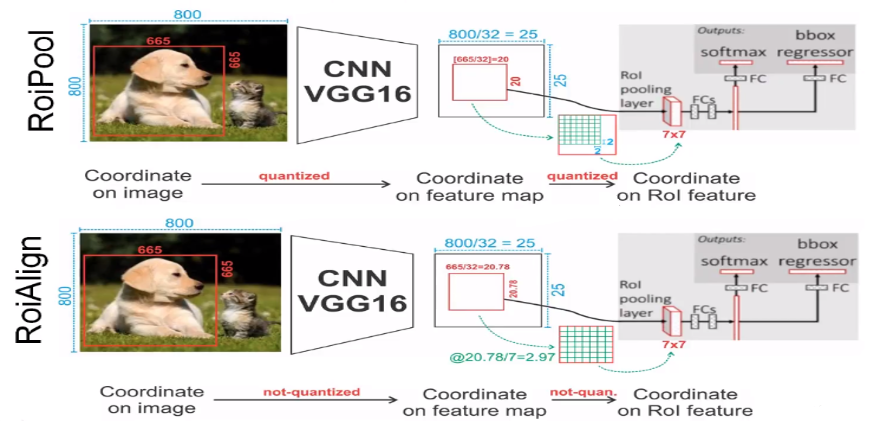


图3-2

## 3.3 数据集选取

首先，选取1000张香蕉图片、1000张苹果图片、1000张橙子图片共3000张图片作为训练集用来训练模型；其次，选取300张包含苹果、橙子、香蕉的图片，其中共有421个苹果、386个橙子和529个香蕉，共1336个水果作为测试集进行测试。

# 第四章 实验结果及分析

## 4.1 对比模型介绍

在这里，我们采用MobileNet-v1作为对比模型。MobileNet-v1是Google在2017年四月针对手机等嵌入式设备提出的一种轻量级的深层神经网络,用于移动和嵌入式视觉应用。MobileNet-v1是基于一个流线型的架构，它使用深度可分离的卷积来构建轻量级的深层神经网络，降低计算量。这种模型的优点是：模型较小，检测速度快，可移植到移动设备等平台。但是，这种模型同样存在缺点：由于网络模型比较小，因此检测准确度不够高

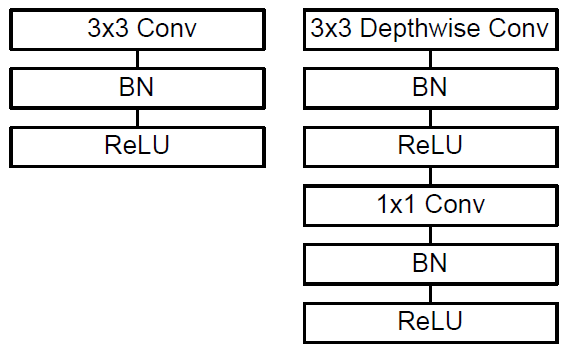


图4-1

## 4.2 模型性能对比

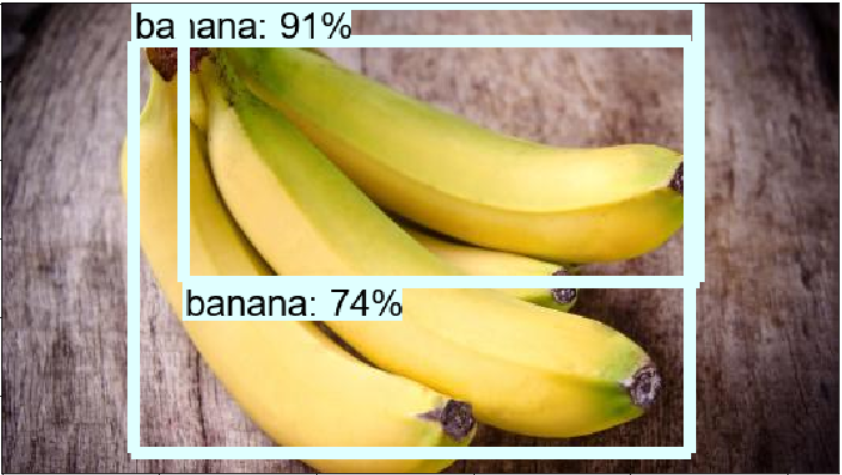
接下来，我们选取300张包含苹果、橙子、香蕉的图片，其中共有421个苹果，386个橙子和529个香蕉，共1336个水果作为测试集，分别对我们使用的Mask R-CNN以及作为对比模型的MobileNet-v1进行测试，比较实验结果，如下表：

表4-1

|  |  |  |
| --- | --- | --- |
| 模型种类 | MobileNet-v1 | Mask R-CNN |
| 测试水果数 | 1136 | 1336 |
| 正确检测水果数 | 1081 | 1268 |
| 错误检测水果数 | 26 | 9 |
| 未检测出水果数 | 229 | 59 |
| 准确率 | 80.09% | 94.9% |
| 召回率 | 97.9% | 99.3% |

下面是我们针对各种不同情况的检测。

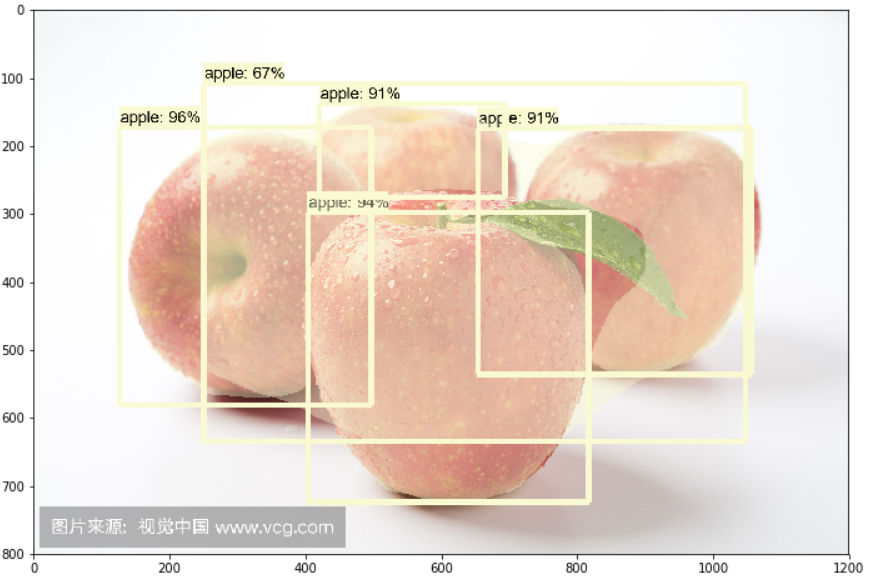
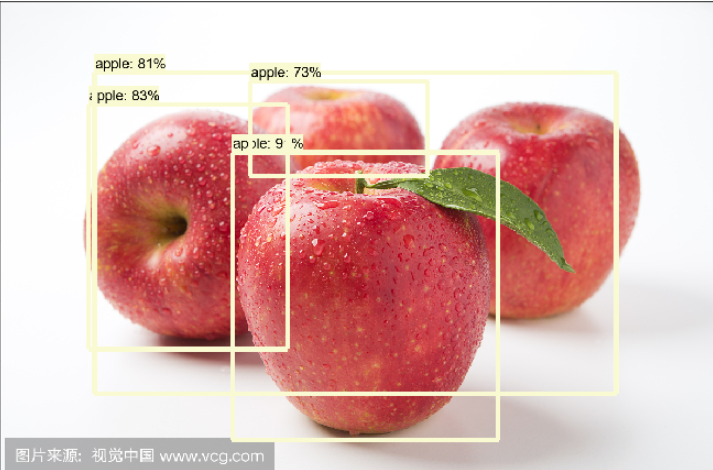
## 4.3 同种水果的检测



（a）MobileNet-v1 （b）Mask R-CNN

图4-2

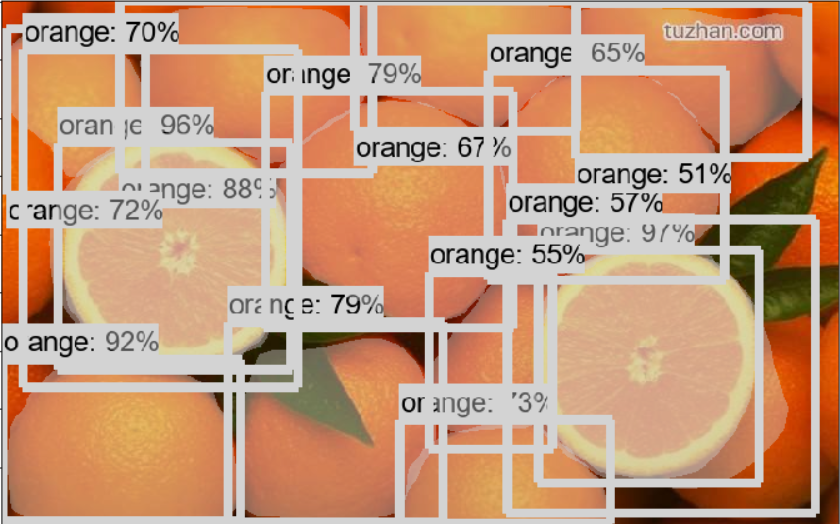
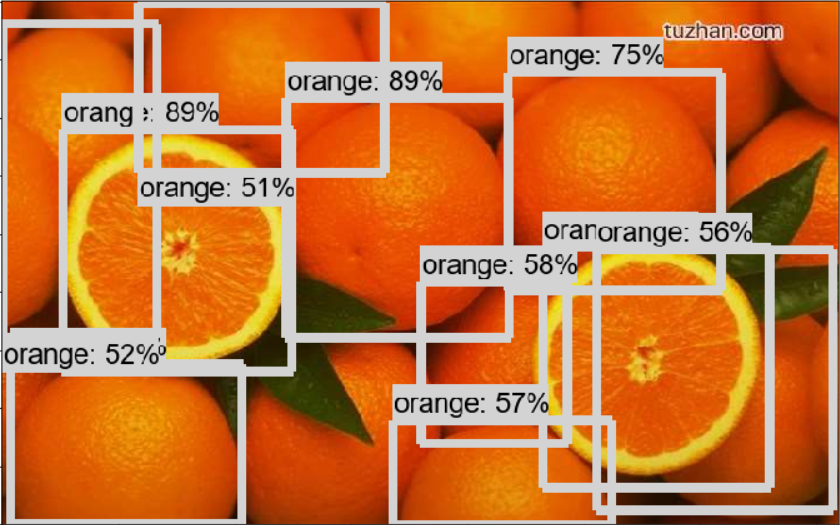
我们可以看到，上图是将香蕉放在一起的场景，从识别框上可以看到，MobileNet-v1只识别出了部分露出面积多的香蕉；而Mask R-CNN识别出的香蕉数目更多，识别效果显然更好。框上的百分比表示：判断为该水果的概率。



（a）MobileNet-v1 （b）Mask R-CNN

图4-3

上图是将苹果放在一起的场景，从识别框上可以看到，MobileNet-v1识别出了左边的三个苹果，最右边的苹果没有明确识别出来；而Mask R-CNN显然将四个苹果都识别出来了，识别效果更好。



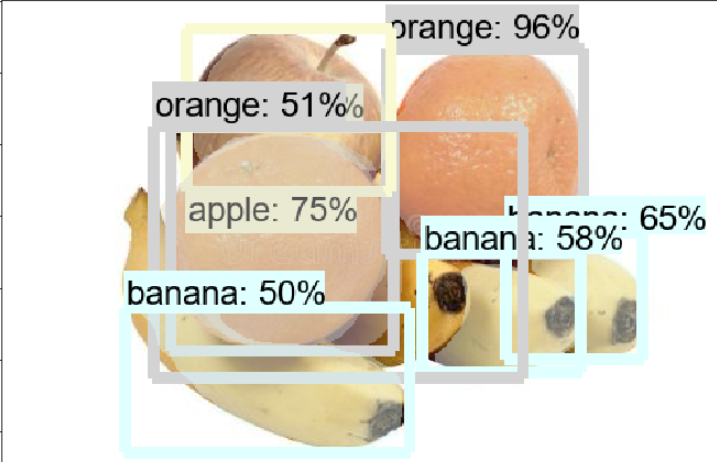
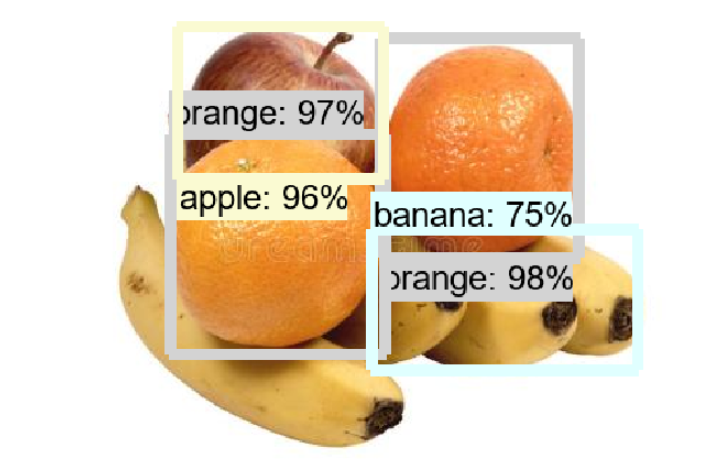
（a）MobileNet-v1 （b）Mask R-CNN

图4-4

上图是将放在一起的场景，MobileNet-v1识别出了上面露出部分较多的几个橘子，对于下面露出部分少的橘子没有明确识别出来；而Mask R-CNN除了上面的橘子，下面的部分橘子也识别了出来，识别效果更好。

由上图的三种情况我们可以看出：对于将同种水果放在一起进行识别的情况，Mask R-CNN可以得到更好的检测结果，在水果被严重遮蔽的情况下，可能会无法识别出这部分水果。

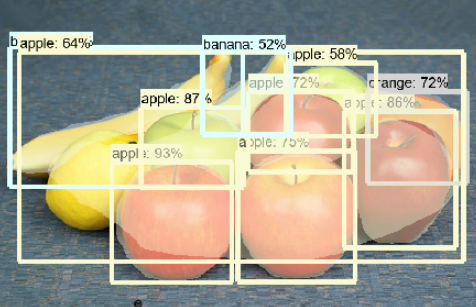
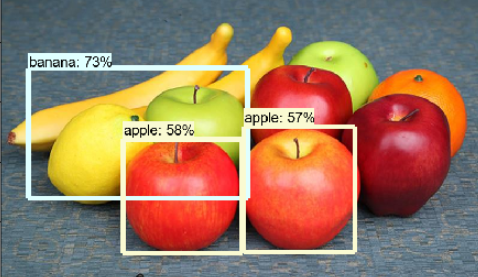
## 4.4 不同种水果的检测



（a）MobileNet-v1 （b）Mask R-CNN

图4-5

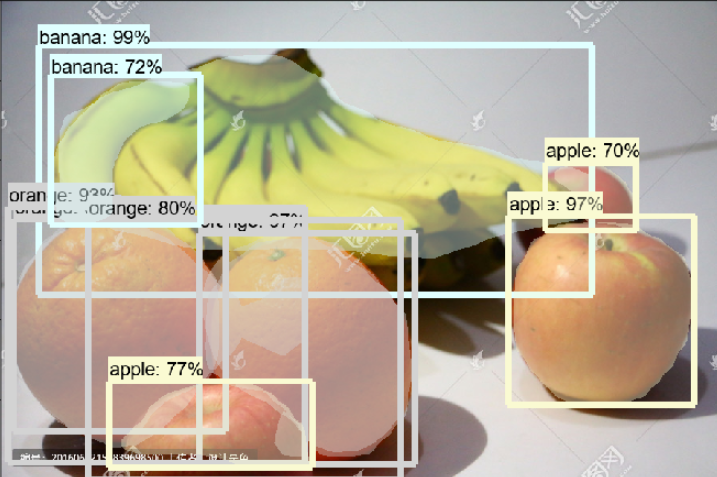
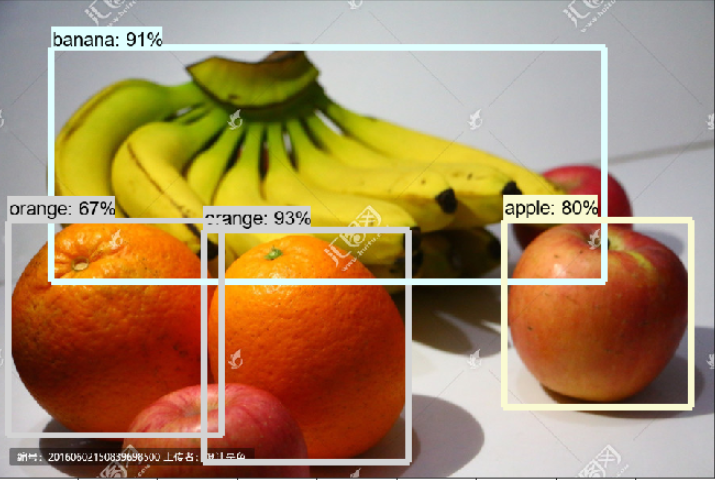
上图是将苹果、橘子和香蕉放在一起的场景，MobileNet-v1识别出了苹果、橘子，并将下面的三根香蕉作为整体识别出来，但是未检测出最左边的香蕉；而Mask R-CNN除了上面的苹果和橘子，最左边的香蕉也识别了出来，右边三根香蕉的识别情况也较前者要好。



（a）MobileNet-v1 （b）Mask R-CNN

图4-6

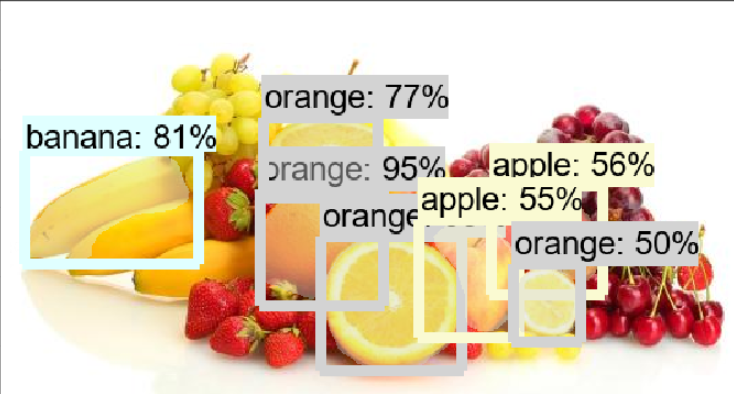
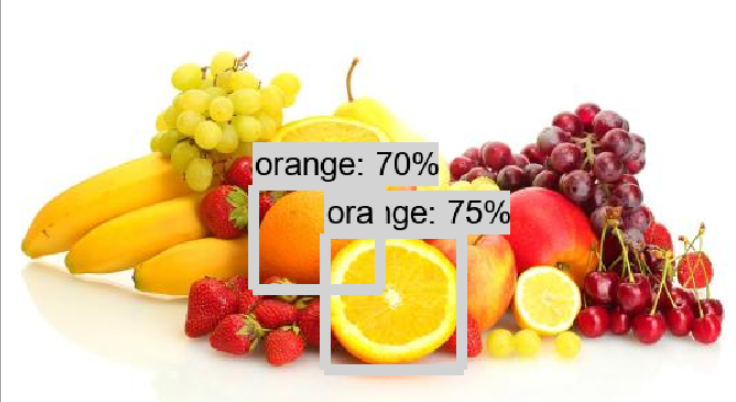
上图也是将苹果、橘子、香蕉和柠檬放在一起的场景，MobileNet-v1识别出了前面的两个苹果，并将后面的两根香蕉作为整体识别出来，但是未检测出后面的苹果和橘子；而Mask R-CNN除了前面的苹果，后面的苹果、橘子和香蕉也识别了出来，识别效果更好。由于我们在训练时只针对苹果、香蕉、橘子三种水果进行训练，因此此时均没有将柠檬识别出来。



（a）MobileNet-v1 （b）Mask R-CNN

图4-7

上图也是将苹果、橘子和香蕉放在一起的场景，MobileNet-v1识别出了露出面积大的橘子、苹果，并将后面的香蕉作为整体识别出来，但是未检测出露出面积小的苹果；而Mask R-CNN除了露出面积大的橘子、苹果，露出面积小的苹果也识别了出来，识别效果更好。



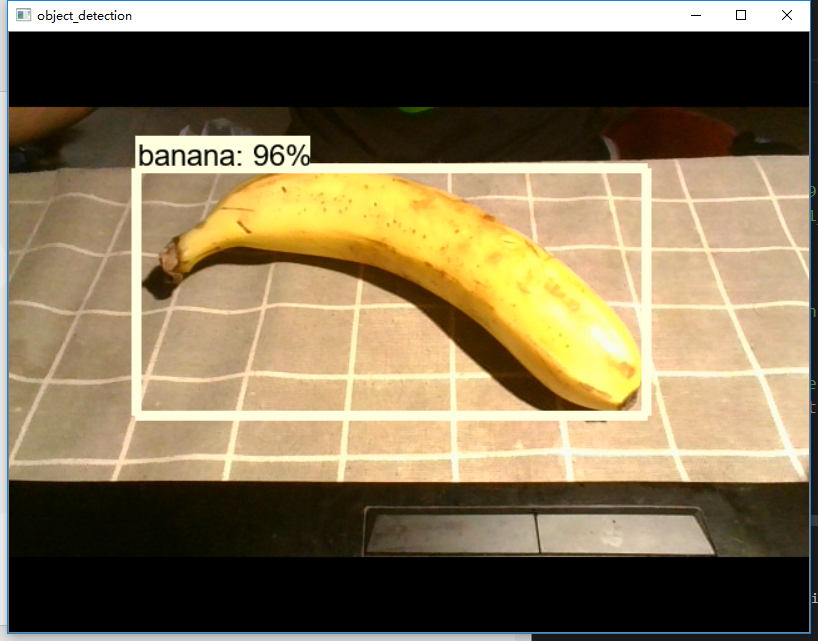
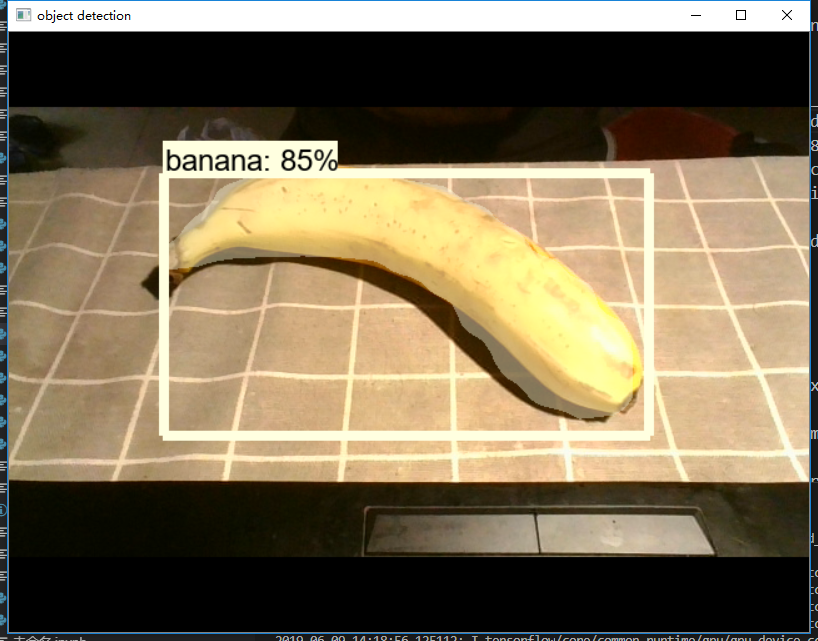
（a）MobileNet-v1 （b）Mask R-CNN

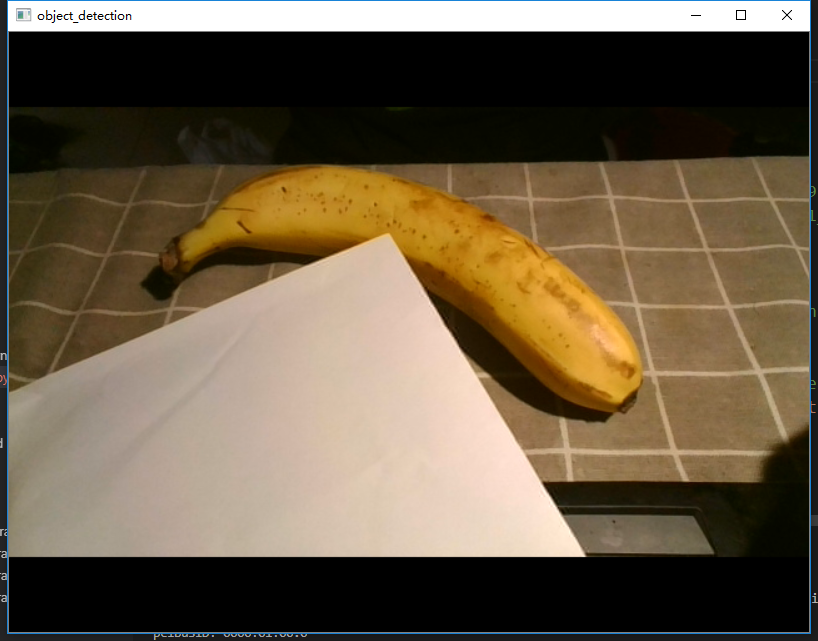
图4-8

上图是将香蕉、苹果、橘子、草莓、葡萄、樱桃等多种水果放在一起的场景，MobileNet-v1只识别出了前面的两个橘子，后面的大量水果没有识别出来；而Mask R-CNN显然识别出了更多的水果，识别效果更好。由于我们在训练时只针对苹果、香蕉、橘子三种水果进行训练，因此此时同样并未将其他水果识别出来。

## 4.5 实时视频检测

我们打开摄像头，来对这两种模型进行实时检测：

（a）MobileNet-v1 （b）Mask R-CNN

图4-9

在亮度较亮，且无遮蔽的情况下，两种模型都识别出了水果，但当光线较暗且存在遮蔽时，MobileNet-v1识别失败，而Mask R-CNN识别成功，识别效果更佳。

## 4.6 轻量模型移植到Android平台检测

我们尝试将google提供的MobileNet-v1轻量模型移植到Android平台上进行检测，发现识别效果比较不错，如图所示，可以成功检测出水瓶、键盘以及鼠标，因此接下来考虑优化精简Mask R-CNN模型使其可以移植到Android平台上进行检测。

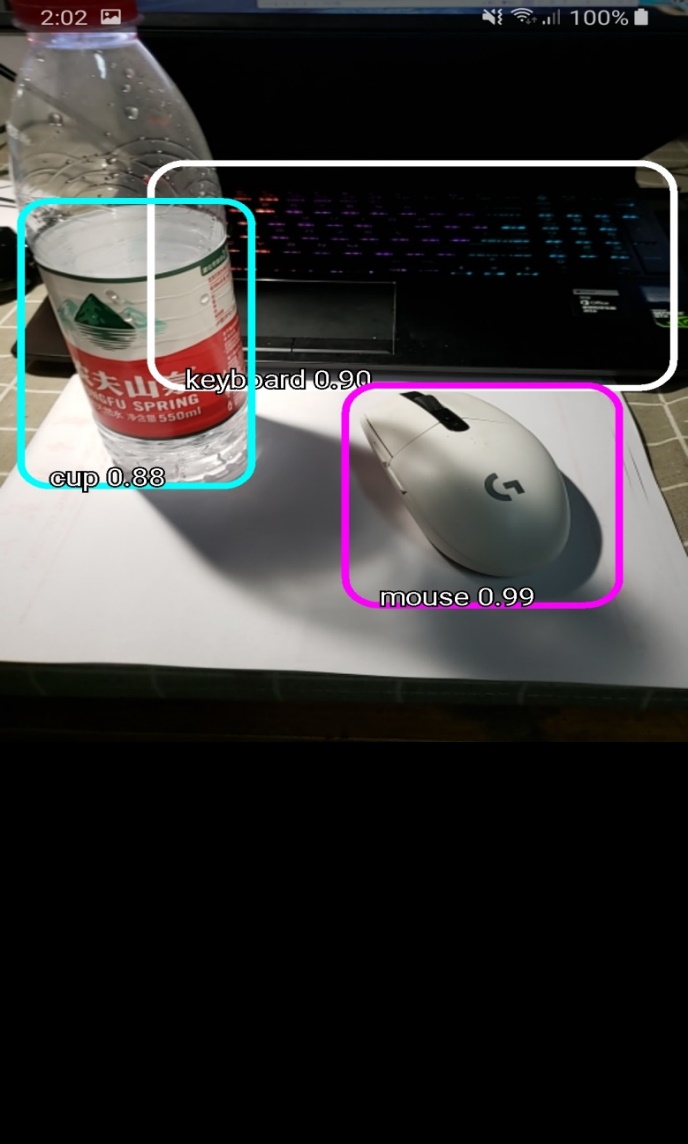


图4-10

## 4.7 结果分析

以上的实验结果我们可以得到：不论是同种水果还是不同种水果放在一起的场景，又不论是上传图片或是开启摄像头进行检测，对于水果的分类识别Mask R-CNN都可以得到更好的结果。在被遮蔽以及光线较暗的情况下，Mask R-CNN也同样可以将水果正确检测出来。在将轻量模型移植到Android平台上进行检测时，同样也可以得到满意的效果。由于我们在训练时只针对苹果、香蕉、橘子三种水果进行训练，因此我们暂时还无法识别其他水果。

# 第五章 总结与展望

伴随着科学技术的发展，人们的生活水平大幅度提升，对于生活质量的要求也越来越高，其中，人们对于食品方面的需求也日益强烈。水果作为其中具有营养价值的一部分，自然受到人们的关注。

我国传统的水果检测方法仍然以人工检测分类方法为主，这种检测方法依靠大量的人力财力，准确率与效率也具有不稳定性，容易受到某些人为因素的影响，所以采用传统方法恐怕难以在科技发展迅速的当下顺应时代的潮流。相比传统方法，基于深度学习框架下的图像自动识别技术可以弥补人眼可见范围有限的缺点，并保持较高的准确率和效率。因此，我们尝试将基于深度学习框架下的图像识别技术应用到这方面中去，试着寻找到一种解决问题的方法。本实验依据实际的场景，研究了基于卷积神经网络的水果检测算法。本实验的主要工作如下：

（1）本实验分析概括了课题研究的背景及意义，作品难点与创新，并对实验的原理进行分析，对深度学习的基本概念与卷积神经网络的原理和结构进行了阐述。此外，针对实际应用场景，为获取适合的目标检测算法，研究了目前主流的基于卷积神经网络的两种目标检测算法模型：Faster R-CNN和Mask R-CNN，并对它们的性能进行了对比和分析，确定了选择效果更好的Mask R-CNN作为本实验水果检测的算法。

（2）在对Mask R-CNN算法原理研究的基础上，对其结构进行了改进：设计特征提取的网络结构，通过减少网络层数提高了算法的效率；对候选窗口分类器结构进行设计，采用双线性插值减小了特征提取感兴趣区域时的误差，提高了车辆检测的准确率。

（3）选用数据集对改进后的算法进行了训练，并进行了实际效果测试。采用MobileNet-v1作为对比模型，同改进后的Mask R-CNN进行效果比较。在多种场景下，Mask R-CNN均具有较好的检测效果。

当然我们的结果目前还只是简单的实现，未来我们还可以进行相应的改进发展。首先，我们可以使用更大的数据集进行训练，实现更多种类的水果识别；其次，由于目前关于水果成熟度的数据集尚且有限，我们还无法实现对水果成熟度的识别，如果未来有相关水果成熟度的数据集，我们就可以进一步实现对水果成熟度的识别；另外，我们还可以进一步优化模型结构，从而实现在移动平台的迁移，以适用于更多的应用场景。

# 参考文献

[1] 韩鹏承,胡西川.基于深度卷积神经网络的街景门牌号识别方法[J].现代计算机.2018.

[2] 刘婷婷,朱文东,刘广一.基于深度学习的文本分类研究进展[J].电力信息与通信技术,2018.

[3] 钱丰,梅剑平,潘荣胜.深度学习在汽车制造物流规划工作中的应用[J].物流技术,2017.

[4] 徐一丁,杜慧敏,毛智礼,张丽果,顾文宁.基于卷积神经网络的工件识别算法[J].组合机床与自动化加工技术,2018.

[5] He K, Gkioxari G, Dollar P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis &Machine Intelligence, 2017, PP(99):1-1.

[6] [Saining Xie](https://arxiv.org/search/cs?searchtype=author&query=Xie,+S), [Ross Girshick](https://arxiv.org/search/cs?searchtype=author&query=Girshick,+R), [Piotr Dollár](https://arxiv.org/search/cs?searchtype=author&query=Doll%C3%A1r,+P), [Zhuowen Tu](https://arxiv.org/search/cs?searchtype=author&query=Tu,+Z), [Kaiming He](https://arxiv.org/search/cs?searchtype=author&query=He,+K).Aggregated Residual Transformations for Deep Neural Networks[J].2016.

[7] Lin T Y, Dollár P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017:936-944.