目录

[一、卷积神经网络 1](#_Toc59266005)

[1.1卷积层 2](#_Toc59266006)

[1.2池化层 2](#_Toc59266007)

[1.3全连接层 3](#_Toc59266008)

[2基于卷积神经网络的目标检测算法 4](#_Toc59266009)

[2.1 基于 two-stage 的算法 4](#_Toc59266010)

[2.1.1 R-CNN 4](#_Toc59266011)

[2.1.2 Fast R-CNN 4](#_Toc59266012)

[2.2 基于 one-stage 的算法 5](#_Toc59266013)

[2.2.1 YLOL v1 5](#_Toc59266014)

[2.2.2 YOLO 改进算法 5](#_Toc59266015)

[2.3 two-stage 和 one-stage 的对比 5](#_Toc59266016)

[3 结语 6](#_Toc59266017)

[参考文献： 6](#_Toc59266018)

基于卷积神经网络的目标检测综述

摘要：传统的目标检测方法大多基于人工设计的特征，精度和鲁棒性较差。随着计算机性能瓶颈的突破，深度学习技术得到快速发展。通过对图像进行卷积处理，卷积神经网络能够提取优于人手工设计的图像特征。基于卷积神经网络的目标检测算法与传统算法相比，准确性和检验速度均有所提高，鲁棒性更强，在目标检测领域获得了广泛的关注。

关键字：目标检测；卷积神经网络；深度学习；计算机视觉

# 一、卷积神经网络

卷积神经网络的核心思想是通过权值共享和局部感知域的方法减少网络的参数个数，缓解了网络模型过拟合的问题，获得了某种程度的尺度、位移、形变不变性，同时还降低了网络模型的复杂程度。卷积神经网络的基础架构主要包含了五个部分，分别是输入层、卷积层、池化层、全连接层与输出层。其网络结构如图1所示。

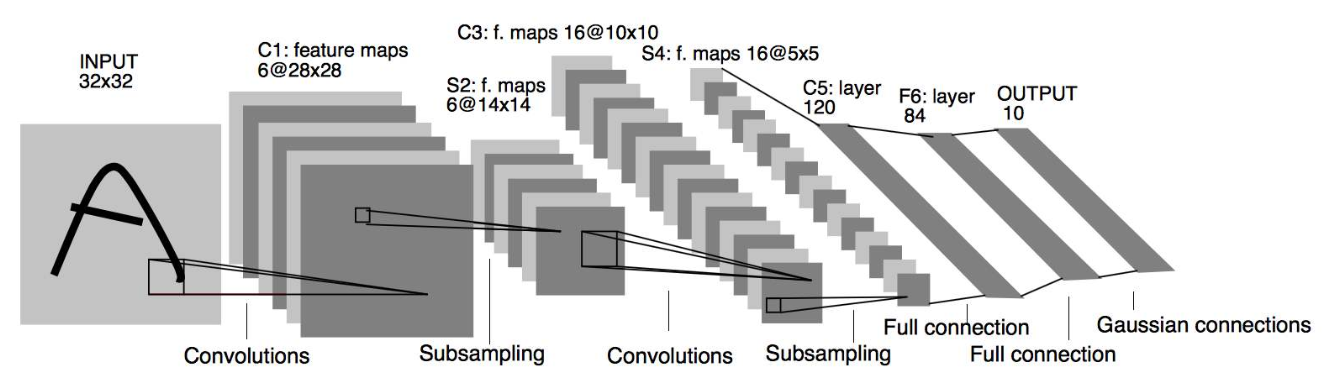


图1 卷积神经网络结构

## 1.1卷积层

卷积层又称为特征提取层，主要作用就是对输入的数据进行特征提取，其内部包含了多个卷积核，组成卷积核的每个元素都对应一个权重系数和一个偏差量，并且卷积核的个数决定了最终提取的数据特征数量。卷积运算的实质是用一个可训练的卷积核去卷积一个输入的图像，然后加一个偏置，得到卷积结果。卷积的详细过程可以理解为使用一个卷积核依次过滤图像的各个区域，从而得到这些区域的特征值。卷积的流程如图2所示。

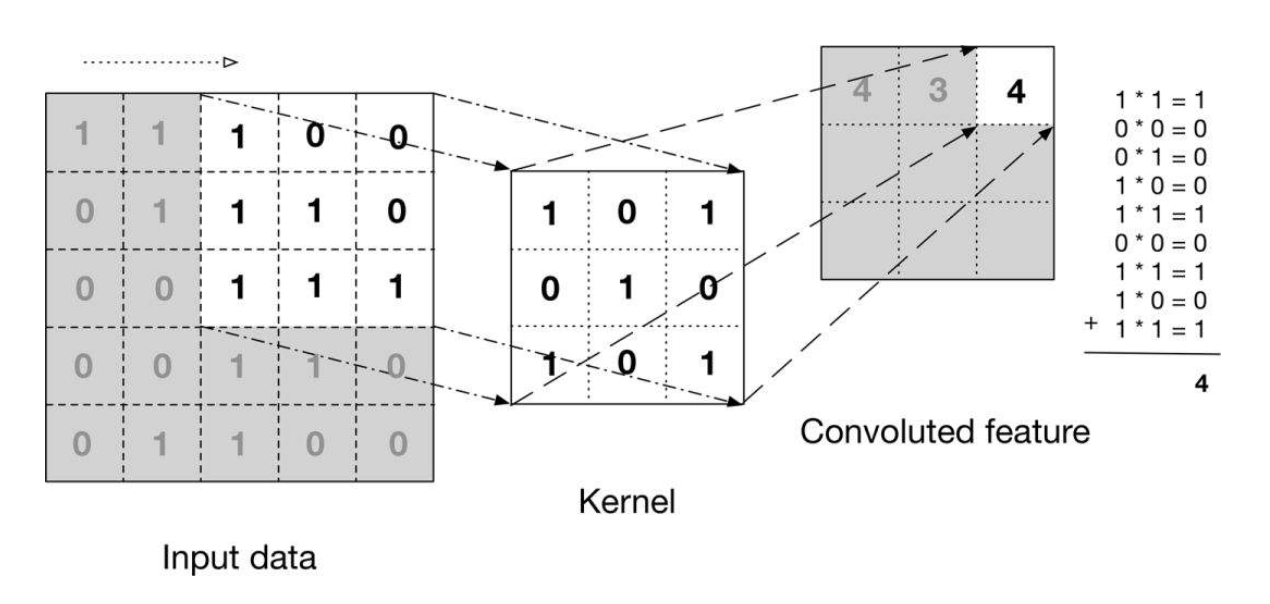


图2 卷积过程

## 1.2池化层

池化操作的主要目的是降维，以降低计算量，并在训练初期提供一些平移不变性。常用的两种池化操作是平均池化和最大值池化。

池化操作就是使用一个固定大小的滑窗在输入上滑动，每次将滑窗内的元素聚合为一个值作为输出。根据聚合方式的不同，可以分为平均池化和最大值池化。最大值池化如图所示。

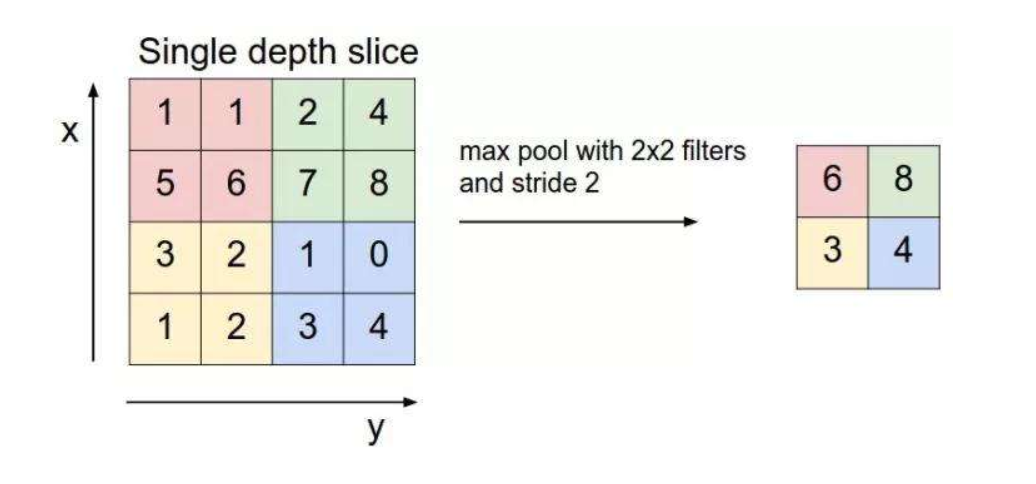


图3 池化

滑窗的大小k和滑动的步长s都会影响最终的输出。通常取,步长与滑窗大小相等，s=2。对于多通道的输入，池化是逐通道进行的，因此不会改变输入的通道数。对于输入，经过上述的池化操作后，输出为，它的长宽将减半。一般来说，对于任意大小的滑动窗口和步长可以使用下式计算。

激活函数是神经网络中的一个十分重要的概念，它的非线性使得神经网络几乎可以任意逼近任何非线性函数。如果不适用激活函数，无论神经网络有多少层，其每一层的输出都是上一层输入的线性组合，这样构成的神经网络仍然是一个线性模型，表达能力有限。

激活函数的选择可以多众多样，一个基本的要求是它们是连续可导的，可以允许在少数点上不可导。常用的激活函数包括S型激活函数和ReLU及其变种等。在隐藏层由于需要对图像特征的提取域降维，为了解决梯度消失的问题，使用了ReLU作为激活函数；在输出层为了避免出现重复分类的现象，使用了softmax作为激活函数。ReLU和softmax函数公式如式所示。

## 1.3全连接层

全连接层在整个卷积神经网络结构中位于池化层的后一层，显而易见其作用是连接经过卷积和降采样后提取出来的所有高级特征，并将最后的输出值送给分类器得出最终的分类结果。全连接层的最终目的是将学到的“分布式特征表示”映射到样本标记空间，其实现过程是通过对输入的特征做加权求和，再加上偏置量，然后通过softmax 激活函数来获得最终的结果，其表达式如式(4)所示。

上式中 表示全连接层中的权重系数;表示通过卷积层和池化层提取出来的特征；表示全连接层的偏置量。

# 2基于卷积神经网络的目标检测算法

卷积神经网络通过学习手动标记特征的数据集来获得目标的特征。目前，基于卷积神经网络的算法大致可分为两种模式，即 two-stage 模式和 one-stage 模式，前者的检测过程分为

两个步骤：首先由算法生成若干个候选框，再通过 CNN 对候选框进行分类；后者则是对目标的类别概率和位置坐标直接回归，相对来说精度有所损失，但速度较 two-stage 模式的算法更快。

## 2.1 基于 two-stage 的算法

### 2.1.1 R-CNN

2014 年 Ross B. Girshick 等 人 在 基 于 CNN 设 计 了 R-CNN模型 , 模型采用了基于 AlexNet 的网络层结构 [11]。在区域推荐（region proposal）上，R-CNN 算法通过选择性搜索（selectivesearch）[12] 来确定候选框，该方法通过选择搜索在图像上确定不同形状和大小的约 2000 个候选框；为了使候选框能够输入到 CNN 中进行特征提取，统一将候选框压缩到 227×227 大小；然后运用 CNN 对候选框进行特征提取；最后使用多个支持向量机（SVM）分类器 [13] 分类输出向量，采用边界回归生成目标区域。整个测试过程如图 1 所示。R-CNN 的应用使得在数据集 PASCAL VOC2012 的结果达到了 53.3%[11]，较之前最佳结果提升了三成。R-CNN 的出现已经成功地将 CNN 应用到目标检测领域，但 R-CNN 也存在着一定问题：在提取候选框时效率低，消耗存储空间较多；在输入网络前需要对候选框进行归一化，有可能导致输入CNN 的信息缺失；所有候选框都要进入 CNN 计算，并且有大量计算重叠。

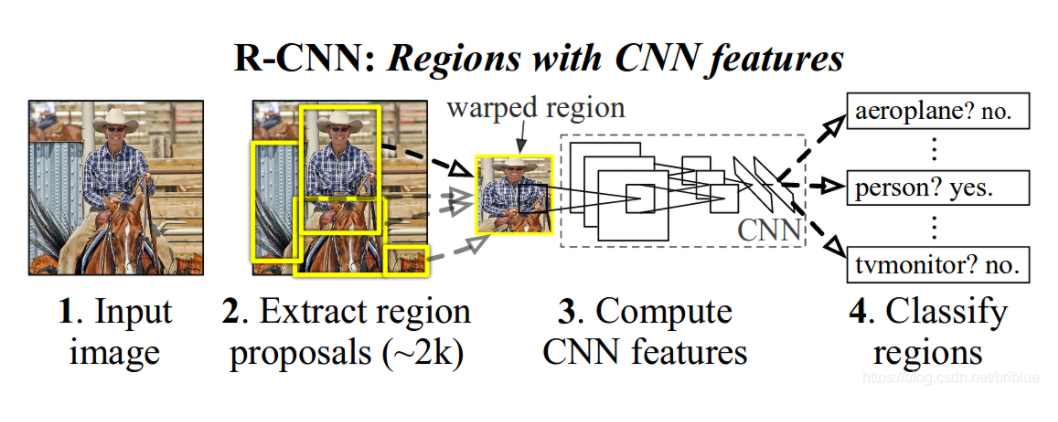


图 1 R-CNN 的测试步骤

### 2.1.2 Fast R-CNN

由于卷积神经网络的全连接层对输入图像的尺寸有着一定的要求，所以需要对 R-CNN 中的候选框进行归一化，无论是采用剪切还是变形，都很难保留图像的完整信息，对于高像素图像，则很容易造成构成缺失和模糊等问题。正是因为这个原因，何凯明等人于 2015 年提出了 SPPNet[14]。SPPNet 解决了 R-CNN 需要重复提取候选区域并使候选框变形的问题，但依然需要分多步骤训练，占用大量磁盘空间。Fast R-CNN 在吸收 SPPNet 优势的同时改进了 R-CNN[15]。Fast R-CNN 仍然使用选择性搜索来确定候选框，但 Fast R-CNN 将整张图片输入到 CNN，在卷积特征层上使用用感兴趣区域（Region of interestpooling，ROI pooling）操作，并从特征图中提取一个特定长度的特征向量；然后将特征向量输入到全连接层，用 softmax 对其进行分类；最后对属于同一特征的候选框进行分类并回归其位置。

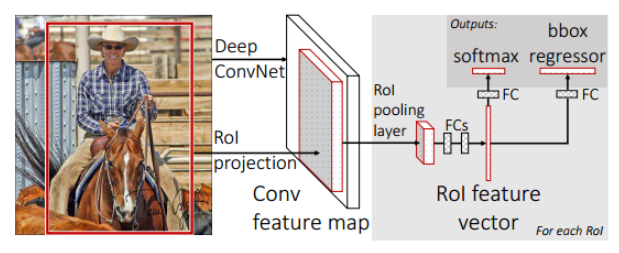


图 2 Fast R-CNN

## 2.2 基于 one-stage 的算法

### 2.2.1 YLOL v1

与基于候选区域的 two-stage 目标检测模型相比，基于回 归方法的 one-stage 目标检测模型免去了候选框提取的步骤， 而是直接在特征图是采取了回归方法。 YOLO 系列算法是一种典型的 one-stage 算法 [17]，它的核 心思想是将完整图片输入网络，在输出层对边界框（bounding box，BBox）所属的种类和 BBox 的坐标进行回归。YOLO v1 的 算法具有一个独立的 CNN 模型，从而实现端到端的检测，其 实现方法为先将图片分为 5×5 个网格单元，再将整张图片输 入 CNN，最后通过整理 CNN 预测结果获得检测的目标。由于 YOLO v1 算法具有统一的框架结构，与 R-CNN 系列算法相比， 其速度更快，训练过程更加简洁。

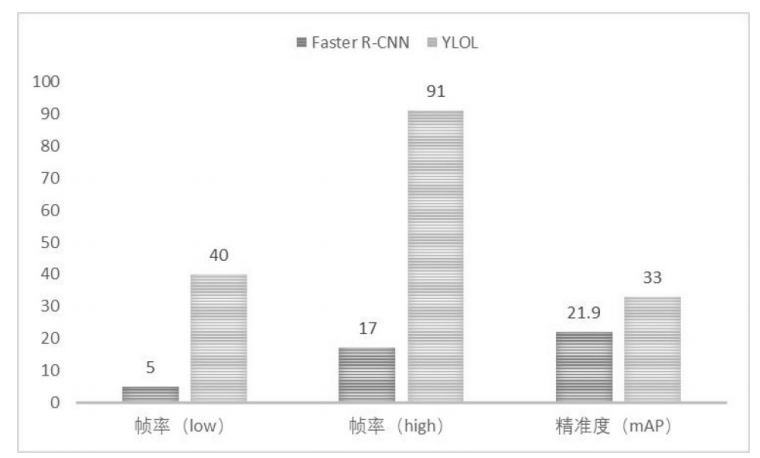
YOLO v1 的效率虽然很高，检测速度很快，但 YOLO v1 的 网格单元只预测两个从属于一个类别的边界框，对于尺度较小 的物体，YOLO 的表现往往不尽人意，同时，YOLO v1 对物体 的形变方面的泛化比率较低，很难定位比例不同寻常的物体， 这也是 YOLO 的一个缺陷。

### 2.2.2 YOLO 改进算法

基于 YOLO v1 存在的问题，已经产生了许多后续的改进算法，如YOLO 9000、YOLO v2、YOLO v3。YOLO-9000 大幅提高了算法可以识别目标的类别数量。YOLO v2 则针对 YOLO v1 的部分内容进行了改进。由于YOLO v1在预训练阶段和检测阶段分别采用了 224×224 和 448×448 的分辨率，导致模型需要适应分辨率的改变。为了解决这个问 题，YOLO v2 将预训练分为了两步。先用 224×224 的输入训练160个循环，再改用448×448的输入训练10个循环。经过微调后，新的模型可以在448×448的图像上顺利过渡。YOLOv2在VOC2007上mAP为78.6%，高于FasterR-CNN[18]。2018年，Redmon等人提出的YOLOv3[19]则在泛化能力上和较小目标的检测上进一步提高，检测速度和精度均有一定的保障。YOLO系列算法是目前一种先进的目标检测算法。因为整个检测框架是一个整体，所以可以端到端地对算法的性能进行优化。同时，因为其优异的性能而被广泛使用，不断产生新的衍生算法。

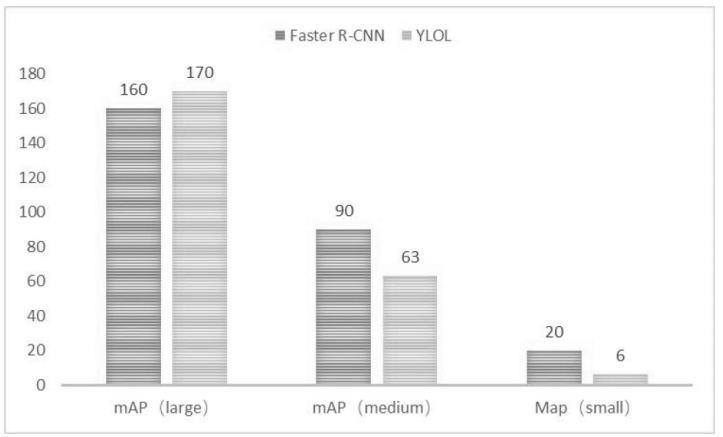
### 2.3two-stage和one-stage的对比

由于结构的差异，two-stage和one-stage算法在检测速度上和精度上也存在着较大的差异。Microsoft COCO数据集是目前计算机视觉领域最权威的数据集之一。作为大型数据集，它Faster R-CNN和YOLOv3分别作为两种结构的代表算法，在Microsoft COCO数据集上的准确度和检验速度如图5所示。



Microsoft COCO上的检测结果

得益于其一体化的网络结构，YOLO算法具有显著高于Faster R-CNN的速度；在准确度上，YOLO略高于Faster R-CNN。但在尺度较小的目标检测中，如图6，在对不同大小的目标进行检测的过程中，由于YOLO每个网格单元只能回归同一种类的两个边界框，导致其对小目标的检测检测精度远低于Faster R-CNN算法。



不同尺度目标的检测结果对比

# 3结语

近年来，目标检测领域高速发展，算法改进方面的研究也在稳步进行，新的成果层出不穷。但整体来说，单一的目标检测算法并不能应用于所有检测环境，各种算法都具有不同的发挥空间。通过将不同的算法进行结合，能获得单一算法所不具备的效率。当前卷积神经网络已成为主流，但是传统算法依然具有一定的价值，能对新的目标检测算法进行启发。

# 参考文献：

[1]丁文锐,刘春蕾,李越,张宝昌.二值卷积神经网络综述[J/OL].航空学报:1-14[2020-12-19].http://kns.cnki.net/kcms/detail/11.1929.V.20200928.1729.010.html.

[2]高卿.基于卷积神经网络的人脸识别研究[D].广西民族大学,2020.

[3]田合雷，丁胜，于长伟，等.基于目标检测及跟踪的视频摘要技术研究[J].计算机科学，2016，43(11)：297-299.

[4]Gu,J.,Wang,Z.,Kuen,J.,Ma,L.,Shahroudy,A.,Shuai,B.,Liu,T.,Wang,X.,Wang,L.,Wang,G.andCai,J.,2015.Recentadvancesinconvolutionalneuralnetworks.ArXivpreprintarXiv:1512.07108.

[5]KrizhevskyA,SutskeverI,HintonG.ImageNetClassificationwithDeepConvolutionalNeuralNetworks[J].Advancesinneuralinformationprocessingsystems,2012,25(2).

[6]Fukushima,K.Neocognitron:Aself-organizingneuralnetworkmodelforamechanismofpatternrecognitionunaffectedbyshiftinposition.Biol.Cybernetics36,193-202(1980).