湖北汽车工业学院



**商务数据分析课程报告**

**专 业： 计算机科学与技术**

**班 级: 计算机172**

**姓 名： 张嘉惠**

**成 绩：**

**完成日期: 2020年6月20日**

**任课教师: 龚家元**

**深度学习框架下的目标检测**

# 一、深度学习框架概述

在深度学习初始阶段，每个深度学习研究者都需要编写大量的重复代码。为了提高工作效率，这些研究者就将这些代码写成了一个框架放到网上以便所有研究者一起使用。接着，网上就出现了不同的框架。随着时间的推移，最为好用的几个框架被大量的人使用从而流行了起来。全世界最为流行的深度学习框架有PaddlePaddle、TensorFlow、Caffe、Theano、MX Net、Torch和PyTorch。

深度学习框架是一种界面、库或工具，它使我们无需深入了解底层算法的细节的情况下，能够更容易、更快速构建深度学习模型。深度学习框架利用预先构建和优化好的组件集合定义模型，为模型的实现提供了一种清晰而简洁的方法。利用恰当的框架来快速构建模型，而无需编写数百行代码。

一个良好的深度学习框架具备以下关键特征：

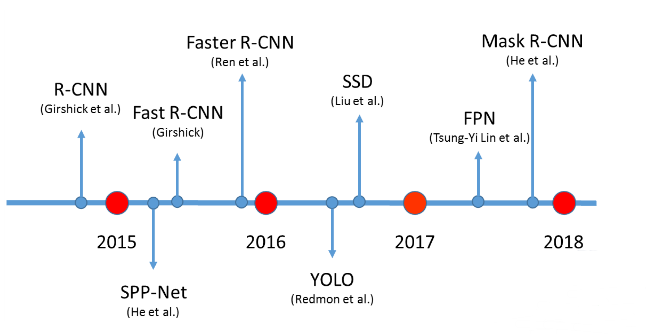
1. 针对性能进行优化
2. 易于理解和编码
3. 良好的社区支持
4. 并行化进程以减少计算
5. 自动计算简便

深度学习框架的出现降低了入门的门槛，我们不需要从复杂的神经网络开始编代码，可以根据需要选择已有的模型，通过训练得到模型参数，也可以在已有模型的基础上增加自己的layer，或者是在顶端选择自己需要的分类器和优化算法（比如常用的梯度下降法）。当然也正因如此，没有什么框架是完美的，就像一套积木里可能没有你需要的那一种积木，所以不同的框架适用的领域不完全一致。

# 二、目标检测

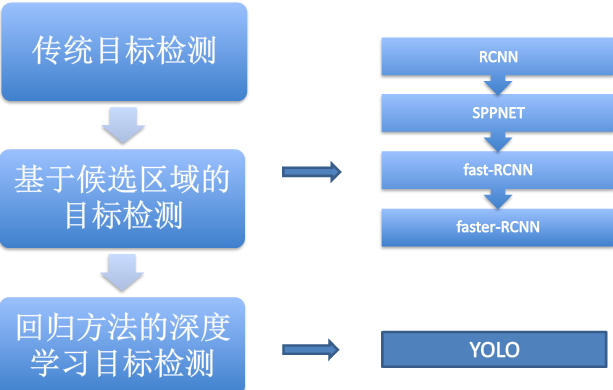
目标定位在实际生活中的应用很广泛，比如，智能剪切(通过定位目标所在的位置,识别需要图片从哪里剪切) ，或者进行常规的目标提取以便进一步处理。结合图像分类技术，它不仅仅可以定位目标，还能对该物体分类。**在迭代处理定位和图片分类问题时，我们最终还是需要对多个目标进行同时检测和分类。**目标检测是在图片中对可变数量的目标进行查找和分类。其中重要的区别是“可变”这一部分。和图像分类问题不同的是，由于每一张图片待检测目标的数量不一，目标检测的输出长度是可变的。

## 2.1 目标检测概况及基本概念



目标检测是将目标从图像中提取出来。运动目标检测是从不断变化的序列图像中进行目标检测，分为静态背景下的运动检测和动态背景下的运动检测。两种主要的思路是：依赖于目标的先验知识，需要提前为运动目标建模，然后在图像序列中实时找到匹配的目标；不依赖于先验知识，直接从图像序列中检测到运动目标，并进行目标识别，最终跟踪感兴趣的运动目标；在目标检测和追踪需要注意运动目标的表示，例如对其视觉特征进行建模，并采用相似性度量，对帧图像进行匹配，在追踪过程中需要处理大量冗余信息，采用搜索算法缩小比较范围。目前跟踪分类主要基于主动轮廓的跟踪、基于特征、区域、模型的跟踪等。

## 2.2基于候选区域的目标检测



**RCNN**

RCNN的思想是利用SS算法提取候选框，然后利用CNN提取特征，最后利用SVM进行识别。与在大量区域上工作不同，RCNN算法提出在图像中创建多个边界框，检查这些边框中是否含有目标物体。RCNN使用选择性搜索来从一张图片中提取这些边框。首先，明确什么是选择性搜索，以及它是如何辨别不同区域的。组成目标物体通常有四个要素：变化尺度、颜色、结构（材质）、所占面积。选择性搜索会确定物体在图片中的这些特征，然后基于这些特征突出不同区域。

存在的缺点：计算量依然很大，其中还有不少其实是重复计算；SVM模型还是线性模型，在标注数据不缺的时候显然不是最好的选择；训练测试分为多步。

**Fast RCNN**

在Fast RCNN中，将图片输入到CNN中，会相应地生成传统特征映射。利用这些映射，就能提取出感兴趣区域。之后，我们使用一个Rol池化层将所有提出的区域重新修正到合适的尺寸，以输入到完全连接的网络中。

Fast RCNN和RCNN相比，训练时间从84小时减少为9.5小时，测试时间从47秒减少为0.32秒。Fast RCNN不足：Region proposal耗时，大部分时间用来提Region proposal。

**Faster RCNN**

Faster RCNN的思想可以简单地看做“区域生成网络RPNs + Fast R-CNN”的系统，用区域生成网络代替Fast RCNN中的Selective Search方法。Faster RCNN是Fast RCNN的优化版本，二者主要的不同在于感兴趣区域的生成方法，Fast RCNN使用的是选择性搜索，而Faster RCNN用的是Region Proposal网络（RPN）。RPN将图像特征映射作为输入，生成一系列object proposals，每个都带有相应的分数。将CNN中得来的特征映射输入到Faster RCNN中，然后将其传递到Region Proposal Network中。RPN会在这些特征映射上使用一个滑动窗口，每个窗口会生成具有不同形状和尺寸的k个anchor box。

存在的问题：算法需要让图像经过多个步骤才能提取出所有目标；由于有多个步骤嵌套，系统的表现常常取决于前面步骤的表现水平



### 2.3回归方法的深度学习目标检测

YOLO 是 You Only Look Once 的缩写。它是一种使用深度卷积神经网络学得的特征来检测对象的目标检测器。YOLO算法的核心思想：利用整张图作为网络的输入，直接在输出层回归bounding box的位置和bounding box所属的类别。YOLO对相互靠的很近的物体，还有很小的群体检测效果不好，这是因为一个网络中只预测了两个框，并且只属于一类。对测试图像中，同一类物体出现的新的不常见的长宽比和其他情况泛化能力偏弱。由于损失函数的问题，定位误差是影响检测效果的主要原因。尤其是大小物体的处理上，还有待加强。

直接对bounding box回归,并不是计算候选和gt之间的 Δ ,并且不构造 hard negatives

速度快，将检测过程变为一个回归问题，简化了很多计算过程（Faster R-CNN也是回归）

全局观，不像R-CNN那样只利用region信息，YOLO会利用整体信息

泛化能力强，尤其训练和测试数据不一致的时候可以表现比较好的泛化性

准去率略低，YOLO是准确率和速度之间的tradeoff

# 三、在人工智能的应用

人工智能从学术理论研究到生产应用的产品化开发过程中通常会涉及到多个不同的步骤和工具，这使得人工智能开发依赖的环境安装、部署、测试以及不断迭代改进准确性和性能调优的工作变得非常繁琐耗时也非常复杂。为了简化、加速和优化这个过程，学界和业界都作了很多的努力，开发并完善了多个基础的平台和通用工具，也被称会机器学习框架或深度学习框架。

目标检测作为图像处理和计算机视觉领域中的经典课题,在交通监控、图像检索、人机交互等方面有着广泛的应用。它旨在一个静态图像(或动态视频)中检测出人们感兴趣的目标对象。传统的目标检测算法中特征提取和分类决策分开进行，对特征选取的要求就更加严格，在面对复杂场景的时候很难得到理想效果。自提出深度学习理论，越来越多的研究人员发现在目标检测领域应用深度学习，可以有效提高检测效果和性能，于是深度学习在实时视频的目标检测开始获得大规模的应用。

目标检测领域有几个主要的难点。首先,视觉场景中广泛分布着许多小物体,这些小目标的存在给检测算法的敏捷性、可靠性带来了挑战;其次,检测精度和速度存在着此消彼长的关系,兼顾这两项性能尚需突破许多技术瓶颈;最后,规模庞大的模型参数是制约深度网络片上装载的重要原因,在保证检测精度的同时压缩模型大小是一个有意义并且亟待解决的问题。

# 四、总结与体会

如果把深度学习比作一座城，框架则是这座城中的水路系统，而基于拓扑图的计算恰似城中水的流动，这种流动赋予了这座城以生命。一个优雅的框架在保证城市繁荣的同时，还能让这座城市散发出独特的魅力，使它保持永恒的生命力。总的来说深度学习框架提供了一系列的深度学习的组件（对于通用的算法，里面会有实现），当需要使用新的算法的时候就需要用户自己去定义，然后调用深度学习框架的函数接口使用用户自定义的新算法。

目标检测（Object Detection）是计算机视觉领域的一个基本问题，在非受控自然场景中快速准确地定位和识别特定目标是许多人工智能应用场景的重要功能基础。近年来，随着深度学习（Deep Learning）技术的快速发展，基于卷积神经网络（Convolutional Neural Network，CNN）的目标检测算法获得了广泛关注和广泛研究，出现了很多性能优良、简洁高效的网络结构，使算法的大规模实际应用成为可能。