# 引言

物体检测，一种使计算机能够在图像中自动找到既定类别的物体，并判断物体的类别、位置、大小及置信度的技术。简单来说，即包括物体的识别和定位。作为计算机视觉的核心和根本问题之一，同时作为人工智能的重要组成部分，是场景理解、图像描述、智能监控、图像内容检索、自动驾驶、事件检测、增强现实等任务的基础。

物体检测的基础是特征（对原始图像的抽象表示），从特征角度看，物体检测的发展主要经历了三个阶段。第一阶段是基于图像原始像素特征或简单扩展像素特征的阶段。其中有代表性的是2001年Viola和Jones提出的基于像素块的harr特征[7]。这一类特征的优点是所需计算量小、提取速度非常快，一般只需几个机器指令就可以完成一个特征的提取。第二阶段是基于特征描述子的阶段。其中有代表性的是尺度不变特征变换（SIFT）[19]，方向梯度直方图（HOG）[13]，局部二值模式（LBP）[20]等。通过精巧的设计以及量化、池化、归一化等操作，这一类特征对于光照、角度、尺度的变化具有了一定的不变性，可以更好地表示一些形状变化较小的模式（pattern）。第三阶段是基于卷积神经网络特征的阶段。卷积神经网络特征即卷积神经网络的某些中间层对于图像的表述[28]。相比于传统的原始像素特征和手工设计的特征描述子，卷积神经网络特征在语义抽象能力方面有着颠覆性的提升。

如何设计高效的检测框架，以减少检测系统整体的计算代价并提高检测性能是物体检测发展中的主要问题之一。在物体检测发展的历史上，提升物体检测效率的策略大体可以分为三类：其一是级联分类模型。级联分类模型本质上是对于简单物体和较复杂物体分别进行分类和计算。其二是特征之间的计算共享。分类器对于滑动窗口中的每个检测框区域分别进行特征提取并使用分类器进行分类。其三是提高子区域识别的效率。提高子区域识别的效率的方法有很多，除了级联分类，对特征子来说设计学习更加高效的特征[33]，对神经网络来说设计更加高效的网络结构[34-37]，以及使用更有效率的分类器[38]、加速的特征计算[39]等都是被广泛研究的方向。

基于以上分析，在个人电脑端以及云服务端，考虑到卷积神经网络在高层语义抽象以及对于大形变物体表示方面已经展现出的强大潜力，基于卷积神经网络的物体检测，尤其是通用物体检测必然是这个领域未来一段时间内的热点之一。

物体检测任务存在很多本质上的困难，本文的工作重点在于寻找设计能提高检测系统效率与精度的算法框架。本文的主要创新点有：1.提出了针对神经网络的空间金字塔池化操作，解决了神经网络分类器只能处理固定大小图像的问题；2.提出了“锚点”以及锚点金字塔的概念，解决了神经网络模拟多尺度滑动窗口检测时计算量过大的问题；3.提出了局部二值特征，以及先局部特征学习后全局共享特征学习的学习方法，解决了人脸对准中特征描述效率低的问题；4.提出了基于二值特征的联合人脸检测与对准框架，解决了人脸检测特征定位描述能力差的问题。

本文主要关注于新的物体检测算法框架，以减小物体检测的计算量，提高计算效率，提高模型性能。第二章介绍相关工作。从物体检测的特征，基本组件以及框架三个层次介绍物体检测历史上的著名方法以及与本文相关的工作。第三章到第五章主要介绍本文的主要工作。其中第三章和第四章主要着眼于多尺度/位置检测框之间的特征共享。第五章主要着眼于检测框内的特征共享。