## 1.1物体检测简介

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，即物体的识别+定位。图1.1展示了物体检测的示例。算法成功定位了不同大小和长宽比的人、马、狗、车，并给出了对应的类别名称和置信度。

物体检测作为计算机视觉的核心和根本问题之一，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。

## 1.2研究意义

物体检测是计算机视觉/人工智能的重要组成部分，作为计算机理解人类世界的一个入口，有着广泛的应用：例如，相机拍摄时找到图像中的人脸以完成自动对焦，检测图像中的笑脸以完成自动快门；车载电脑检测道路前方的行人，并对驾驶者做出警示；无人飞行器在空中跟拍下方正在进行滑雪、冲浪等运动的人。这些令人激动的应用场景只是近几年来计算机视觉/人工智能给我们带来众多改变中的冰山一角。晚饭后，家用机器人帮您送上一个苹果并端来一杯热水；平日里，扫地机器人自动将用过的餐盒，饮料瓶等分类收整好，以待丢弃或回收；出门时，汽车能够完全自动地驾驶，避开行人和其他车辆；行人过马路时，红绿灯可以根据路口行人及车辆的多少自动调节红绿灯的转换和时长。这些美好的愿景在不远的将来可能就会在我们的生活中实现。而这一切都需要计算机能够像我们人类一样，从视觉图像中获取到更多的信息和知识。这些信息中关键的一组就是图像中的物体在哪里（where），以及他们是什么（what）-而解析这些信息的技术就是物体检测。

对于人来说，眼睛是人类获取信息的主要媒介。人类通过眼睛获取图像，然后分析周围的物体与环境进而做出判断。对于计算机来说也是一样—从图像中获取信息对于现阶段帮助人类处理繁琐事务的低级人工智能，以及未来真正理解世界的高级人工智能来说都不可缺少的·项重要能力。而要想从图像中获取信息理解世界，如何从中检测并识别物体是必须要解决的问题。目前计算机视觉，尤其是物体检测、识别相关方向的研究正处在高速发展阶段，新方法的出现带来了基础上的变革，一些根本性的难题有了新的进展，展现出了很大的研究潜力：另外一方面一些相对成熟的技术开始被应用在商业领域，展现出了广阔的应用前景。

从研究角度看，检测问题是计算机视觉的根本问题之一，是图像分析和理解的基础。通用物体检测从概念上说分为两个部分—定位与识别。定位即找到物体所在的位置，识别即判断该位置物体的类别。相应的，物体检测从技术上一般也包括两个主要部分，其一是图像分类/识别算法，其目的是判定图像中任何一个窗口中的子图像的类别和置信度。其二是子窗口搜索方法，对于一幅图像，如果在多尺度及长宽比上通过滑动窗口的方法枚举所有可能的子图，即使对于一幅600\*800的图像，可能的子图像个数也会达到百万量级。利用完整的图像分类算法对这个量级的子图分别进行分类，在时间上是不可以接受的。因此，在考虑图像分类算法的基础上提出创新的窗口搜索方法是物体检测领域必须考虑的。物体检测是一个对于精度和效率都要求极高的任务，如何更加准确地对图像进行分类，以及如何在资源有限的前提下，对图像中数量庞大的可能子图进行搜索分类，这两方面的研究不仅在物体检测领域有重要的研究意义，对于计算机视觉的其他方向也有着重要的借鉴意义。

一般来说，计算机视觉对于图像的操作可以分为三个阶段：图像处理，图像分析和图像理解。三个阶段对于抽象类别的概念层层加深。物体检测作为图像分析和理解中的重要基石，对于图像描述、事件检测、情感理解等更加抽象和高层的任务来说都是不可或缺的，而这些任务正是下一代人机交互和高级人工智能的基础。

从应用角度看，物体检测已经表现出广泛的应用需求和前景。笑脸快门、辅助驾驶、图像搜索等技术已经在不经意间默默提升着我们的生活质量，在不远的将米，更加善解人意的人工智能类应用更是值得我们期待。所有这些应用与改变中，基于物体检测的深度图像理解技术都是核心。

笑脸快门是很多相机都已经内置的功能，当镜头中的人露出笑容的时候会自动激发快门，保留下这温馨美丽的一刻。笑脸快门，顾名思义，其中第一步也是关键一步是要从图像中快速准确的找到人脸的位置，即物体检测中的人脸检测，进而才能进行后续的人脸对准，表情分类等操作。这其中人脸检测的准确度和召回率直接决定了整个系统的可用性和鲁棒性。由于人脸检测其需要处理图像中数量庞大的可能位置，是整个框架中最耗时的部分，而笑脸快门一般都出现在相机手机等便携性设备上，计算资源受限的情况下如何高效可靠的检测出人脸是整个系统成功的关键。当前相机中很多已经集成了人脸检测模块，但是由于计算问题，都只能检测跟踪少量的人脸，并且会受到偏光照、遮挡等问题的影响。如何更快更好地检测人脸是业界热切关注的问题。

辅助驾驶旨在给予驾驶者以危险提醒，对于行人、车辆的检测可以有效避免恶性交通事故的发生；对于路牌，交通标示牌的检测与识别则可以帮助车辆更加安全稳定地行驶。这个应用的核心问题是检测的鲁棒性，如何在确保召回率的情况下提高正确率是行人、车辆、路牌检测能真正让人放心使用，乃至代替人实现自动驾驶的关键。相比人类而言，计算机视觉的一个优势是在可见光图像之外还可以利用其它性质的摄像头得到图像以简化视觉问题的难度。例如目前在工业界和学术界都很受关注的深度摄像头，借助深度信息，检测技术在很多条件下被推向了实用。如何利用图像，包括深度摄像头、雷达等广义图像，辅助驾驶乃至自动驾驶，是很有工业前景的一个问题。

遥想不远的将来，随着物体检测技术的发展，笑脸快门不单能检测一张笑脸，还能在班级毕业合照的时候确保留下所有人都睁着眼睛露出笑脸的瞬间：辅助驾驶不单单能提醒驾驶者前方可能的行人，还能让汽车自己驾驶，确保不会跟行人及其他车辆发生碰撞。这些变革都会有物体检测技术在背后的贡献。

综上所述，基于物体检测的图像理解技术正影响着越来越多的人。创新的应用不断推出，在视觉及图像理解层面能更像人类的人工智能也初现端倪并一点点地成长。物体检测技术的发展与创新值得期待。

## 1.3 算法框架

了解了物体检测的概念及意义之后，这一节我们主要介绍物体检测任务的算法框架。如同我们所讨论的，物体检测任务从本质上由识别与定位这两个互相耦合的子问题组成。在物体检测几十年的研究过程中，尽管用来处理识别任务的特征和分类器在不断改进，处理定位任务的基本策略却始终如一对图像的每个子区域逐一进行识别。这一算法思想与我们人类天然的做法在思想上一致。人在观察一幅复杂图像时会浏览图像的不同位置，并依次进行识别。物体检测算法也会浏览图像中不同位置、大小的子图，并依次进行识别。这种检测过程因其滑动浏览的过程被形象地称为滑动窗口（Sliding Window）。

检测框是物体检测的基本处理单元（如图l.2左上的红色框所示）。算法对于框中是否包含物体，包含何种物体进行判断。这里的包含有两重意思，其是指物体在框内，其二要求框是物体的紧致外边框。为了识别不同位置的物体，检测框会遍历图像的所有位置（如图1.2所示）。当图像宽度为w，长度为h时，使用这样一个框进行滑动窗口检测时需要处理约w\*h个位置，当对一个框内子图进行识别的平均计算量为C时，使用一个固定大小的框进行滑动窗l检测的总计算复杂度为O（w\*h\*C）。由于物体的大小及长宽比变化不一，只使用一个大小/长宽比固定的框进行物体检测是不够的。例如，图1.2中，红色的长方形框跟图中狗的大小相近，用来检测狗较为合适，但是它的大小与图中的马相差较大，用来检测马则无法得到满意的结果。因此，在使用滑动窗口检测时，多尺度/长宽比的处理是必须的。多尺度/长宽比检测的其中一种做法是使用不同大小，不同长宽比的框来分别检测物体，如图1.2右下角所示。这种方法被称为分类器金字塔。

使用金字塔的目的是使得检测框能尽量匹配物体的大小变化。为了达到这个目的，除了上述分类器金字塔，图像金字塔也是物体检测中常用的方法。如图1.3所示，图像金字塔，即使用固定大小的分类器，作用在多尺度/多长宽比的图像上。分类器金字塔和图像金字塔都可以达到改变检测框与图像中物体相对大小的作用，因此都可以实现多尺度/多长宽比物体检测的目的。实践中，这两种方法经常被同时使用。

当使用N个不同尺度/长宽比的分类器金字塔/图像金字塔进行滑动窗口检测时，算法总的计算复杂度为



## 1.4主要困难

公式（1.1）是一个重要的公式。我们可以粗略地估算一下，当N=10（算法里的N一般会大于这个值），w=800，h=600时，检测的计算量约为0（480，000×C）。需要着重指出的是，这里的O（C）代表着一个底层图像识别算法的实现，同样需要大量的计算。即使识别一张子图只需要1毫秒，滑动窗口检测一张600×800的图像也需要长达8分钟！这几乎是无法忍受的。巨大的计算复杂度及计算量正是物体检测任务面临的主要困难之一。

除了计算量上的困难，物体检测面临的另一个困难是物体多样性带来的识别上的困难。颜色、角度、光照、大小、遮挡、形变等因素的变化都会增加物体的类内差异，对物体的识别造成困难（图1.4）。其他一些外在因素，比如图像的模糊、噪声以及失真也会对识别提出额外的挑战。

物体检测除了识别之外的另一个维度是定位，定位同样会受到上述因素的干扰。其中影响最大的可能是遮挡，如图1.4所示，当物体被部分遮挡或者大部分遮挡时，如何准确确定物体的边缘是相当困难的问题。当多个物体之间存在遮挡时，如何将他们正确识别并定位成多个物体也是物体识别面对的重要挑战之一。

计算复杂度、识别以及定位都是物体检测面临的困难，更为棘手的是，这些困难往往是耦合在一起的。不能准确定位，就无法精确识别；不能正确识别物体类别，就很难帮助定位物体：计算代价太高，就不得不降低识别和定位精度以向运行时间妥协。如何设计新的算法框架，以减小计算量，提高计算效率，提高模型性能，是物体检测任务研究的主要方向之一。

## 1.5研究现状

现代计算机视觉的研究从20世纪60年代兴起，经历了从线段等图像基本组成部分的检测[4，5]，到人脸[6-10]、行人[11-14]等形状相对稳定的物体的检测，再到对形变剧烈、方差巨大的一般物体的检测[15-18]的发展。随着数据生成、采集难度的降低，计算机计算能力的飞速提高，以及物体检测算法的不断改进，计算机在检测问题上对于颜色、大小、光照、遮挡、角度、形变等变化的鲁棒性不断提高，并一步一步接近实用。

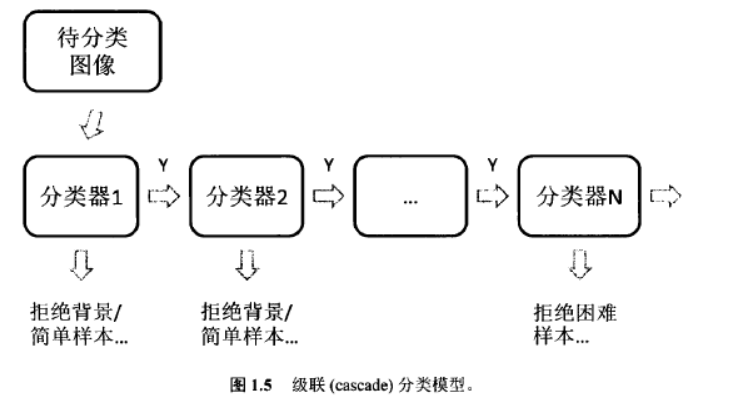
物体检测的基础是特征（对原始图像的抽象表示），从特征角度看，物体检测的发展主要经历了三个阶段。

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。所以通过合理的设计，基于这类特征的检测系统甚至适用于十年前、乃至二十年前的计算机，可见其出众的高效性。这一类特征的缺点也很明显，它们只包含了简单的图像明暗信息，并不具备高层语义上的抽象能力，因此这类特征只有与强大的非线性分类器结合使用，才能完成图像识别的任务。这类特征一般不适用于形变较大的物体。在当今的应用中，这类基于像素的特征依然发挥着重要的作用，尤其是在计算资源高度受限的手机等移动和便携式设备上。

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）。例如对于行人、人脸等一些介于刚体和非刚体之间的物体，这些特征描述子具有很好的描述能力。从l990年LBP特征被提出以来，各种特征描述子被人们设计出来，而与图片相关的模式识别任务大多都用到了这些特征描述子，例如图像分类[21]、物体检测[22-25]、图像分割[26]、图像检索[27]等。与基于图像原始像素的特征相比，这一类手工设计的特征描述子在具有更好不变性的同时一般需要更大的计算量。在现代计算机上，基于这些特征的应用，通过合理的设计和优化，一般可以满足实时计算的需求。但是与基于原始像素的特征类似，手工设计的特征描述子也没有高层语义上的抽象能力，这种语义抽象上的缺失可以直观地理解为，这种特征的近似性与图像内容在人理解中的相似性并不吻合。

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。对于卷积神经网络特征而言，特征空间上的相近性即在一定程度上代表了语义层面上的相近性。特征对图像的描述也从对人来说可能并不重要的细节以及区域相似性的表述推进到了高层语义层面。这一方面是因为卷积神经网络是由大量根据语义层面的类别信息组织起来的数据训练得到，另一方面是因为卷积神经网络的层递结构，使得它在每一层学到了不同语义层次的图像表述。现代卷积神经网络已经在图像分类、场景分类等基于图像的模式识别任务上表现出了强大的能力[29]。具有强表述能力的同时，由于卷积神经网络的多层结构，相比手工设计的特征描述子而言，卷积神经网络特征所需要的计算代价也明显更大。从数字上看，在当前的最好的GPU上，使用现阶段性能较好的卷积神经网络，处理一个子图片的分类任务需要几毫秒（ms）的时间，如何在物体检测任务中高效地使用这种高性能的特征，是对物体检测算法框架的·大挑战。纵向对比这三个阶段，从特征角度看，近二十年来，物体检测领域取得了非常大的进展，尤其是近几年卷积神经网络特征在语义抽象性方面表现出了巨大潜力。但是我们也必须看到，特征发展的另一个特点是特征复杂度的明显提高和抽取所需时间的显著增加，这对于物体检测任务提出了更大的挑战。如何利用卷积神经网络特征的优良抽象表述能力去提升物体检测系统的性能，是一个重要的研究课题，于此同时，设计新颖高效的算法框架以降低卷积神经网络高昂的计算代价，把卷积神经网络物体检测方法推向实用是与前一问题息息相关且同样重要的一个课题。

不止对于卷积神经网络特征，对于任何一种特征而言，将特征、分类器与滑动窗口定位策略简单进行结合的检测框架往往有着不可接受的计算代价。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一种分治的方法一对于待分类图像集合（滑动窗口检测框集合）中较简单的部分使用计算量小、能力弱的分类器进行分类，以降低整体的计算代价；对于较难的部分再附加使用计算量大、能力强的分类器进行再分类。如图1.5所示，这种级联分类器可能由若干级分类器组成，分别处理不同难度的样本。对于物体检测任务，一般多数滑动检测窗口都是背景，以及跟图中物体尺度相差较大的简单样本，因此在级联分类模型中，检测框的平均计算量，一般只相当于第一级或是前几级简单分类器的计算量。级联分类器由于其在提高检测算法效率上的显著作用，被很多检测算法所使用。级联分类模型可以与像素特征[7]、特征描述子[30]、卷积神经网路特征[18]等几乎所有特征相结合。广泛用于人脸检测、行人检测、通用物体检测的提升模型（boosting）就是天然的级联结构，我们会在章节2.3.1中详细讨论。

经典检测框架中的级联分类器中，当特征所提取的图像区域相同时，所提取到的特征也是相同的。这样提取到的特征对于物体的角度等变化不具备鲁棒性，例如在检测人脸时，使用同样的特征提取参数在一张正脸以及一张侧脸上提取到的特征实际上是来自于人脸的不同区域，它们所包含的语义信息自然较少，用于分类时的鲁棒性也就较差。使用这些鲁棒性差的特征进行分类时，级联分类不得不使用更多的级数、更强的分类器，最终提高了检测的难度和计算代价。

在第五章中，我们创新地提出了随着级联框架改变的特征。如图1.6所示，在级联框架中，我们不单使用分类器由简到难地拒绝负样本，还使用回归器由粗到细地生成人脸语义关键点的信息。这些关键点的信息会在特征提取的时候作为参考点，帮助我们提取对人脸角度和表情等更加鲁棒的形状索引特征2.1.1。基于人脸关键点提取的形状索引特征含有更好的分类信息，可以帮助级联分类器更快地拒绝样本，从而提高整体系统的检测效率。

其二是特征之间的计算共享。如图1.2和图1.3所示，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，从概念上讲，每个子区域之间是彼此独立的。但是仔细观察后我们会发现，相邻的区域是有很大的重叠；换言之，每个图像中的像素点都分别属于很多个不同的检测框区域，会被很多个分类器处理并提取特征。从概念上来说，虽然同一个像素点在不同检测框区域的相对位置不同，对特征的共享可能就不同。但是从像素级别来说这种提取方式在某些情况下是冗余的，这种冗余性为我们设计方法提高检测效率提供了潜在的可能性。

特征间计算共享的经典使用是Haar 特征积分图。Haar特征需要计算图像任意矩形框内像素值的和时（图l.7中灰色区域）。利用预先计算的积分图（积分图（x，y）点的值，图像上任意一个以（x1，y1）为左上角，（x2.y2）为右下角的矩形区域内的像素和S可以被快速地计算：



通过积分图的引入，图像上任何一个检测区域框内的任意一个矩形区域像素值之和都可以通过4次取值和加减操作完成。这本质上是共享了不同检测框区域之内对局部像素块的加和操作。这种特征间的计算共享显著加速了Haar特征的计算，为Haar特征的成功应用奠定了基础。Haar特征的细节我们会在章节2.1.1中讨论。

方向梯度直方图特征（HOG）[13]通过将图像中n×n像素（例如，6×6）的信息汇总在一起作为特征的基本单元，并在不同区域间共享基本单元的计算来达到特征间计算共享的日的。这种特征共享加速了基于方向梯度直方图特征的检测算法，但是加大了滑动窗口的移动间隔/步长（stride）（从1个像素变成了n个像素），理论上会损失定位及识别的精准度。有关方向梯度直方图特征的细节我们会在章节2.1.2中讨论。

与方向梯度直方图特征[13]以及尺度不变特征变换（SIFT）特征[19]不同的是，在卷积神经网络中特征与分类器是合二为一共同训练的，特征与分类器的界限相对模糊。使用卷积神经网络仿照基于方向梯度直方图特征检测算法的框架，直接进行多尺度图像金字塔滑动窗口检测的性能并不理想[31]。Girshick el al.[18]提出使用传统区域牛成（region proposal）算法[32]作为级联分类器的第一级，过滤掉绝大多数的滑动窗口检测区域，而只保留约2000个大小、长宽比不定的较可能包含物体的区域。然后使用卷积神经网络作为级联分类器第二级对于这2000个生成区域进行分别识别。由于全连接层的限制，卷积神经网络只能处理固定大小的图像。[18]将2000个图像子区域大小变换成既定大小后，交给网络处理。简单来说，[18]检测一张图像，需要基于卷积神经网络的图像识别算法运行约2000次（如图l.8（左））。将神经网络运行如此多的次数，计算代价无疑是巨大的。

受到前人特征间计算共享方法的启发，我们在第三中研究提出了可以在神经网络检测中将卷积神经网络计算共享的算法框架。如图1.8（右）所示，在检测张图像时，我们只需将网络的卷积层在整张图像上计算一次，这会从本质上减少冗余的计算量，提高卷积神经网络检测算法的效率。我们提出的算法同时还可以使得卷积神经网络支持任意大小的图像输入，显著提高卷积神经网络使用的灵活性。

第三章中提出的算法框架使得检测算法的效率提高了数十倍，我们可以在l/10秒量级的时间内处理2000左右的检测框区域。但是如果使用这样的算法框架和效率来处理与多尺度滑动窗口数目相同量级的检测框区域，仍然是不现实的。为了在单一尺度卷积神经网络中实现多尺度滑动窗口检测算法，我们在第四章中提出了区域生成网络（RPN）框架。在区域生成网络框架中，我打破了分类算法中，对哪个图像区域做识别就在相应图像区域上提取特征（即特征提取区域=分类目标区域）的定式，提出了“锚点”的概念（即图l.9（右）中的分类区域）。在我们的算法中，特征提取区域不再等同于分类目标区域（锚点），对于同一位置不同尺度/长宽比分类目标区域（锚点）的识别将会使用同样（共享）的特征。基于锚点的检测模型实现了多尺度识别之间的特征计算共享，使得卷积神经网络可以摆脱低效的传统区域生成算法，而完整的实现整个物体检测算法框架。

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，比如我们在第一条中所讨论的级联分类就是可以显著提高子区域识别效率的方法之一。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。

第五章中，针对人脸对准（精确定位）问题，我们创新地提出了局部特征学习与全局共享特征学习相结合的学习方法。通过局部提取特征学习单个特征点的回归函数，降低了问题的难度，提升了特征提取的效率。进而将多个局部的二值特征综合起来进行全局所有特征点回归问题的学习，提高了算法的泛化能力与鲁棒性。通过这种学习方法得到的人脸子区域定位器在精度与效率上都表现优秀。

## 1.6研究趋势

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。在物体检测任务上，探索研究卷积神经网络优于传统方法的原因，以及针对卷积神经网络的特点，设计适合的算法框架以提升物体检测精度都会是研究领域的重点。设计更高效合理的算法框架以提升物体检测算法的效率会是学术界和工业界共同关注的问题，从学术界出发提升算法效率是找到问题核心部分的方法之一；从工业界出发，使得技术能够最终实用才是目标。

而在手机端和各种移动设备上，基于卷积神经网络的检测算法由于其计算量太大，所以需要专用的芯片才能实时运行，这也会是软硬件交叉领域的一个研究热点。另一方面，对于人脸检测、行人检测等任务，基于原始像素类特征设计更加精确高效的实时算法依然会是一个被学术和工业领域所关注的方向。

## 1.7主要创新点

基于章节1.4的讨论，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。本文的主要创新点如下：

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。基于这种池化方法的检测系统改变了原来每个子图像分类都需要完全独立计算的情况，使得多区域之间可以共享特征的计算，显著降低了检测系统的计算代价。

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。基于锚点金字塔的检测系统使得多尺度滑动窗口检测可以在一个单一尺度的网络中快速完成，并且将区域生成与子区域识别融于同一个网络之中，提高了整体检测系统的效率。

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。基于所提出特征及学习方法的人脸对准系统在计算效率上有了本质的提升。在手机等移动设备上，即使使用10%的计算资源也可以实现人脸实时对准与人脸视频实时跟踪。

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。基于该框架的人脸检测系统利用人脸对准的定位信息提取特征帮助检测。联合任务中的特征共享提高了该系统的效率与实用性。

## 1.8章节安排

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。具体章节安排如下：

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。

第三章提出针对卷积神经网络检测算法的多区域特征共享策略。该策略使得神经网络检测算法中的不同位置的检测框之间可以共享特征。使得神经网络检测算法的时间从10秒量级降到l秒量级。从根本上降低检测算法计算复杂度的同时，保持了卷积神经网络检测算法的卓越性能。

第四章提出不同于分类器金字塔和图像金字塔的锚点金字塔。使得神经网络检测算法复杂度中有关N的项被消除。神经网络检测算法的时间从l秒量级降到0.1秒量级。在降低检测算法复杂度的同时，取得了更高的检测精度。

第五章分为两部分。第一部分针对人脸检测相关的精细对准任务，提出了局部学习二值特征+全局学习的新方法。该方法提升了人脸对准模型的性能。是当前速度最快的人脸对准算法。第二部分在第一部分的基础上，将人脸对准和人脸分类两个任务合二为一，共享特征，产生了当前最快最好的非神经网络人脸检测算法。

第六章回顾本文内容，总结内容与创新点，并发现其中的不足为将来的作提供方向上的指导。