

密级: _____



中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

基于 Adaboost 级联架构的通信资源受限条件下 航拍目标识别系统研制

作者姓名: 王忻雷

指导教师: 韩振军 副教授

学位类别: 工程硕士

学科专业: 计算机技术

研究所 : 中国科学院大学电子电气与通信工程学院

二〇一六年五月

The Research on Adaboost Based Object Recognition
Cascaded System in Aerial Images under Limited
Communication Resources Conditions

By
Xinlei Wang

A Thesis Submitted to
The University of Chinese Academy of Sciences
In partial fulfillment of the requirement
For the degree of
Master of Computer Technology

School of Electronics, Electrical and Communication Engineering
University of Chinese Academy of Sciences

May, 2016

中国科学院大学直属院系 研究生学位论文原创性声明

本人郑重声明:所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知,除文中已经注明引用的内容外,本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体,均已在文中以明确方式标明或致谢。

作者签名:

日 期:

中国科学院大学直属院系 学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定,即中国科学院有权保留送交学位论文的副本,允许该论文被查阅,可以公布该论文的全部或部分内容,可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密的学位论文在解密后适用本声明。

作者签名:

日 期:

导师签名:

日 期:

摘要

近年来，目标识别研究由于其在社会生产生活中的巨大应用潜力获得了快速的发展。目前，目标识别系统已经广泛应用于智能交通管理、辅助驾驶系统、医学图像分析等领域。但是，在一些如航拍检测等灵活性要求较高的场景下，受到灵活性等因素的限制，目标识别系统还无法得到大规模应用。近年来无线网络获得了高速发展，目标识别系统与无线网络相结合可以很好的增大系统的灵活性。同样，通过目标识别系统向嵌入式平台移植也能减小灵活性对目标识别系统带来的约束。

然而，无线网络与目标识别技术相结合以及目标识别系统向嵌入式平台移植带来了一些传统目标识别系统尚未遇到的新问题。无线网络与目标识别技术相结合受到通信资源的限制，目标识别系统向嵌入式平台的移植受到嵌入式平台计算资源的限制。本文以航拍检测为实验背景，借助嵌入式平台增加系统灵活性的同时，研制了一套基于 Adaboost 级联架构的通信资源受限条件下的目标识别系统。

本文的主要工作内容如下：

1) 设计并提出了一套通信资源受限下的目标识别系统框架，即动态级联目标识别系统框架。本文从目前传统的目标识别系统入手，逐步对现有方案进行改进并分析其可行性，最终论证了新型目标识别系统框架的合理性和有效性；

2) 以无人机航拍检测为实验背景，实现软件模拟的动态级联目标识别系统。搭建系统运行的软硬件平台，设计系统内置的目标识别算法并与其他的目标识别算法进行可行性对比分析，通过分析软件模拟目标识别系统的检测结果来证明框架设计的合理性和有效性。同时，在航模与沙盘模拟飞机场的场景下，将软件模拟系统升级为准实物的目标识别系统，完成目标识别系统向嵌入式平台移植的工作，并初步对系统框架进行了验证。

关键词：无线网络，网络约束，目标识别，航拍检测

Abstract

In recent years, the research onto object recognition in the computer community which is relying on computer vision access to rapid development due to its great potential in the production of social life. Object recognition currently has been widely applied in intelligent traffic management, driver assistance system, medical image analysis, etc. Its applications are limited in some demanding highly flexibility scenarios, for example aerial detection. Its combination with wireless networks and the aid of embedded platform improve its adaptability.

However, this combination with wireless networks and the transplantation of embedded platform introduce some intractable problems. The combined system of wireless network and object Recognition technology is limited by communication resources, and the transplantation to embedded devices is limited by computing resources. In this paper, we propose an efficient dynamic cascaded object system framework to solve the problem of wireless network resource constraints to object recognition system. The main contributions of this paper are as follows:

- 1) Propose theoretically an object recognition system framework involving wireless network, dynamic cascaded target recognition system framework. We improve the conventional object recognition system, analyze its advantages and disadvantages, and justify the proposed recognition system framework.
- 2) Implement real dynamic cascaded object recognition system built on the simulated one for the unmanned aerial detection task. We build the software and hardware platform, design a built-in object recognition algorithm, conduct comparable experiments with other related algorithms and analysis its necessarily for the whole recognition system. Besides, with the sand table and the model aircraft to simulate airport scenarios, we implement the whole system built on the simulated one, accomplish the work on the transplantation of embedded platform, and validate the developed system.

Key words: Wireless Network, Network Constraints, Object Identification, Aerial Detection

目录

摘 要	I
Abstract.....	III
图目录	VII
表目录	IX
第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	3
1.3 本文研究内容	6
1.4 本文的组织结构	7
第 2 章 目标识别系统的相关技术研究	9
2.1 Haar-like 特征	9
2.2 决策树	9
2.3 Adaboost 算法	12
2.4 卷积神经网	13
2.5 本章小结	15
第 3 章 无线网络约束下的目标识别系统框架	17
3.1 传统离线式目标识别系统	17
3.2 传统离线式目标识别系统改进方案	18
3.2.1 改进方案一：全部在线式目标识别系统在服务器完成	18
3.2.2 改进方案二：全部在线式目标识别系统在前端完成	22
3.3 新型动态级联目标识别系统	24
3.4 本章小结	26
第 4 章 动态级联目标识别系统	27
4.1 前端目标识别系统设计方案	27
4.1.1 前端系统硬件平台	27
4.1.2 目标识别算法	29
4.1.3 网络数据传输模块	30
4.2 服务器端目标识别系统设计方案	31
4.3 模拟目标识别系统的工程实现	32
4.4 模拟系统的实验结果	35
4.4.1 模型的准确率和召回率	35
4.4.2 模型各级完成检测后节省的带宽百分比	41
4.4.3 前端模型各级完成检测后需要的计算量	44
4.4.4 通信资源与计算资源相互转换情况	45
4.4.5 可调的级数	46
4.5 模拟系统升级为实际系统	47

4.5.1 前端目标识别系统.....	47
4.5.2 服务器端目标识别系统.....	51
4.6 实际目标识别系统的检测结果	51
4.6.1 计算量与级数之间的关系.....	52
4.6.2 可调的级数.....	53
4.6.3 实际系统检测结果演示.....	54
4.7 本章小结	55
第 5 章 结论与展望.....	57
参 考 文 献	59
致 谢	63
个人简介	65

图目录

图 1-1 图像分割算法结果示意图 ¹²	4
图 1-2 Selective Search 提取候选窗结果 ¹²	5
图 2-1 Haar-like 特征模板	9
图 2-2 决策树结构	10
图 2-3 神经元结构图	14
图 2-4 卷积神经网结构示意图 ³⁸	14
图 3-1 传统系统到新型系统改进方案	17
图 3-2 传统的目标识别系统流程图	17
图 3-3 改进方案一流程图	19
图 3-4 路径衰减	21
图 3-5 移动环境下的链路质量变化样本 ⁴⁵	22
图 3-6 静止环境下的链路质量变化样本 ⁴⁶	22
图 3-7 改进方案二流程图	23
图 3-8 无线网络约束下的目标识别系统	25
图 4-1 Jetson TK1 板卡实体图 ⁴	28
图 4-2 Adaboost 框架流程图	29
图 4-3 DPM 各窗口得分情况	30
图 4-4 TCP 连接的三次握手协议	31
图 4-5 新型动态级联目标识别系统框架	34
图 4-6 前端目标识别系统各级性能	36
图 4-7 召回率与级数的关系	37
图 4-8 部分丢失的目标	37
图 4-9 复杂场景和多角度下的识别情况	38
图 4-10 经过前端处理后在服务器端用 CNN 检测后的性能统计	39
图 4-11 未经过前端处理直接在服务器上进行识别的结果	40

图 4-12 前端目标识别系统检测结果演示	41
图 4-13 级联系统的检测结果演示	41
图 4-14 原高清图像与服务器端收到的不同带宽下的图像	43
图 4-15 Adaboost 模型各级与带宽百分比之间的关系	44
图 4-16 前端系统各级与计算量之间的关系	45
图 4-17 计算资源与通信资源相互转换结果展示	46
图 4-18 Adaboost 模型各级可调级数	47
图 4-19 BitRate 示意图	49
图 4-20 实际系统的硬件实物图	50
图 4-21 航模与机场模拟真实机场场景	50
图 4-22 动态级联目标识别结果演示	52
图 4-23 可调的级数	54
图 4-24 角度、尺度变化以及模糊情况对目标识别系统的影响	55

表目录

表 2-1 信息增益算法与信息增益比算法	11
表 2-2 C4.5 算法流程	11
表 2-3 Adaboost 算法流程	12
表 4-1 带宽与级数之间的关系	51
表 4-2 模拟系统与实际系统窗口数对比	53

第1章 绪论

1.1 研究背景与意义

目标识别技术作为计算机视觉领域的一个重要分支在社会生产生活中有着巨大的应用潜力,得到了很多国内外专家学者的重视。经过几十年的研究和发展,目标识别技术在很多领域已经在一定程度上得到了应用,如:智能交通管理、辅助驾驶系统、医学图像分析等等。但是当前投入实际应用的目标识别系统的应用场景一般复杂程度较小且位置固定,监测场景一般变化不大,相对来说比较简单。如车牌识别系统一般是将摄像头安置在固定地点,对特定位置下的车牌进行识别。而且在这些领域中,人们追求的是目标识别系统内置模型的精确度的提高,较少考虑系统自身的体积、功耗以及系统内置目标识别模型的复杂度等等。在一些如航拍检测等灵活性要求较高的复杂场景下,目前传统的目标识别系统尚难以投入使用。因此,研究一个能够在灵活性要求较高的场景下高效运行的目标识别系统框架具有非常重要的意义。

近年来,通信网络已在世界范围内普及,无线网络也在全世界范围内高速发展。由于无线设备相比于有线设备具有较高的灵活性,通过目标识别系统与无线网络技术相结合可以有效解决现有目标识别系统灵活性不足的问题。目前,在高灵活性场景下(如:航拍检测),很多专家学者已经做了相关研究,但是主要精力集中于提高检测的性能而非改进系统的结构。而且这类工作通常都是以科学的研究为目的,只是定期将数据采集到研究中心进行分析。要实现在这类高灵活性场景下实时的目标识别,实现目标识别技术与无线网络技术相结合具有非常大的应用价值。然而,无线网络最显著的特点是带宽资源不够充分、稳定性不高,将无线网络与目标识别系统相结合势必会带来许多传统目标识别系统未曾遇到的新问题,其中,最主要的就是通信资源受限问题。目前绝大多数无人机测控系统仍然满足不了大数据量稳定传输的需求。因此,如何解决通信资源受限的问题是目标识别系统与无线网络相结合的一大难题。

另一个提高目标识别系统灵活性的方案是将目标识别系统由服务器移植到嵌入式平台。由于模型的性能与计算资源一般成正比关系,高性能的模型一般需要比较充分的计算资源来支持,因此传统的目标识别模型一般运行在研究中心的服务器上,而对计算资源要求较低的目标识别模型在性能方面往往不是很好。

随着硬件平台的发展，嵌入式平台的计算资源越来越充分，嵌入式平台相比于服务器具有更高的灵活性，如摩尔定律所言，当价格不变时，集成电路上可容纳的元器件的数目，约每隔 18-24 个月就会翻一倍，性能也将提升一倍。虽然就目前硬件水平而言，在嵌入式平台上运行精确的目标识别模型在计算资源方面还有不足，但是运行一些对计算资源要求不高的目标识别算法已经可行。随着硬件水平的逐步提高，在嵌入式平台上运行目标识别模型将是未来发展的趋势。

然而，目标识别技术与无线网络相结合以及目标识别系统向嵌入式设备移植带来了一些传统目标识别系统尚未遇到的新问题，例如，无线网络与目标识别技术相结合受到通信资源的限制，目标识别系统向嵌入式平台的移植受到嵌入式平台计算资源的限制。因此，在通信资源受限的情况下，寻找一种新型的目标识别系统框架来解决目标识别技术与通信网络相结合的问题至关重要。

无人机技术近年来发展迅猛，得到了广泛的关注。无人机技术在很多方面相比于过去有了很大提升，如：飞行的稳定性提高、功耗大幅降低、续航时间显著增强、造价逐步降低，已经在民用、军用多个领域投入使用。国内的大疆创新科技有限公司^[1]研制的无人机已经应用于业余摄像和专业数据采集等方面，美国政府投资研制的全球鹰^[2]无人机已经在美国对阿富汗和伊拉克战争中执行过任务。无人机本身体积较小，不需要现场的人为操作，在军方执行一些任务时可以有效避免人员伤亡。在民用方面，使用无人机可以在多个视角获取图像来满足人们的需求。同时，无人机本身的具备了非常高的灵活性，以无人机作为系统硬件平台的载体对以后系统的应用具备更大应用价值。

航拍检测^[3]是近年来比较热门的研究方向，可以通过无人机对机场、交通枢纽等公共场所进行监测以避免恐怖袭击等事故的发生，也可以通过无人机航拍检测来完成监测军事基地或探查敌情等任务。航拍检测是灵活性较高场景的典型代表，在航拍检测场景下实现新型目标识别系统框架具有较大的应用价值。

本文以无人机航拍检测为实验背景，为当前目标识别技术所面临的通信资源受限的问题提出一种新型的目标识别框架，即基于 Adaboost 级联架构的通信资源受限下的航拍目标识别系统。并且分别用软件和 Jetson TK1^[4]嵌入式平台搭建模拟系统和试验系统，通过分析实验结果证明目标识别框架的合理性和有效性。

1.2 国内外研究现状

关于通信资源受限和目标识别模型移植到嵌入式平台这两个问题,国内外很多专家学者已经对这两个问题做出了相关的研究。

随着硬件平台性能的高速增长,将目标识别模型移植到嵌入式设备成为了一个非常具有研究价值的方向。2005 年和 2007 年,卡耐基梅隆大学的 Anthony Rowe 等人分别推出了 CMUcam2^[5] 嵌入式目标识别系统平台和 CMUcam3^[6] 嵌入式目标识别系统平台。后者是一个开源的嵌入式计算机视觉平台,具有重量轻、低功耗等特点,保持了很好的灵活性,并且很好的减少了硬件平台对平台载体带来的载重和功耗等方面的压力。在 CMUcam3 平台中内置了背景建模中的帧差法^[7]来检测监测场景中的运动目标,该算法通过对图像序列中相邻两帧或三帧图像采用基于图像的时间差分来提取出图像中的运动区域。但是,该方法因为要求检测场景的背景是静止的,这很大程度上限制了该平台的应用范围。同一年,Anthony Rowe 等人提出了一个应用在无线传感网的图像处理框架 FireFly Mosaic^[8],该框架在无线传感网络的每一个节点上配置一个 CMUcam3 平台作为图像处理单元,在无线传感网中,带宽资源问题是各个节点进行通信的一大难题。Anthony Rowe 等人通过设计数据传输的策略来缓解带宽资源的压力,在 FireFly Mosaic 框架中,所有节点都只可以与相邻的所有节点相通信。还有很多专家学者对通信网络与目标识别技术相结合的问题和目标识别系统移植到嵌入式平台问题做出了研究。斯坦福大学的 Stephan Hengstler 和 Hamid Aghajan 设计的 MeshEye^[9]是一个低功耗的无线分布式视频监控系统。加利福尼亚大学洛杉矶分校的 Mohammad Rahimi 和 Rick Baer 开发的微型相机 Cyclop^[10],使用 8 位的微处理器和 FPGA 来捕获和处理图像。Onur Atan 等人^[11]以移动设备上的目标识别系统为应用背景,致力于在无线网络信道不稳定情况下的数据传输问题和多节点之间数据流调度问题。

本文在提出抽象的目标识别框架后,将框架在具体的航拍检测场景下实现。由于近年来卫星遥感等技术的发展逐步被人们所关注,航拍检测也成为目标检测的一个重要研究方向。在本文实现的目标识别系统中,虽然通过无人机而非通过卫星获取图像,但是二者拍摄得到图像具有非常相似的特点。航拍检测中运用的目标检测的方法,在基本原理方面与传统目标检测一致,只是根据实际情况对模型和特征等做了一系列改进,因此有必要了解传统目标检测技术的发展。关于目标检测技术,主要包括窗口的提取、抽取特征和分类器评分。

在检测一幅图像之前，需要预先提取出可能是目标的区域作为目标识别算法的输入，即提取窗口的过程。目前比较常用的提取窗口的方法主要分为两类，一类是穷举法，穷举法首先将图像通过升采样和降采样做不同尺度的变换，在每个尺度之下，通过逐行逐列的扫描将所有可能是目标的区域都作为窗口输入到后续的目标识别模型当中。穷举法的优点是不会在窗口提取的过程中漏掉任何一个目标，但是其缺点也非常明显，就是其会产生数目非常庞大的窗口并且会造成非常多的窗口重叠，造成信息冗余并且也给后面特征提取和分类器评分的工作带来了非常大的时间开销和空间开销，通常一幅一百万像素图像将会产生数十万甚至上百万个候选框。正因为穷举法信息量的冗余以及庞大的窗口数目，许多研究人员开始设法减少窗口数量和信息量冗余。第二类提取窗口的方法是通过算法（如图像分割）来提取感兴趣的区域。其中，运用比较广泛的 Selective Search^[12]就是通过这种思路实现的。Selective Search 首先用 Felzenszwalb and Huttenlocher 在 2004 年提出的图像分割算法进行图像分割^[13]，该算法是基于图的图像分割算法，该算法将图像中每个像素作为一个节点，通过最小生成树方法对节点进行合并，算法的输出结果如图 1-1 所示。图像分割算法将图像分割成为很多小区域，Selective Search 将这些小区域作为输入，通过颜色、纹理、尺度、形状的相似度对这些小区域进行合并，这些作为输入的区域以及它们的合并结果作为最后 Selective Search 最终的候选区域，工作人员可以根据实际场景选取一组或几组标准进行区域合并，Selective Search 结果如图 1-2 所示。相比于穷举法，Selective Search 能够大幅降低窗口的数目和信息冗余，提取到的窗口数目一般是穷举法提取到窗口数目的几十之一。



图 1-1 图像分割算法结果示意图^[12]

图 1-2 Selective Search 提取候选窗结果^[12]

特征提取是目标检测中非常重要的一部分。特征提取主要作用是将图像中的信息量转换成其他更加便于分类器检测的形式。目前很多特征已经得到了广泛的应用，如：Haar-like^[14]、HOG(Histogram of Oriented Gradient)^[15, 16]、LBP (Local Binary Patterns) ^[17]等特征以及深度学习等可以自动从图像中提取特征的方法。Haar-like 特征由 Papageorgiou 和 Poggio 等人提出^[18]，在文章中采用 Haar 小波描述行人、人脸和车辆。Haar-like 通过描述图像的灰度变化情况来识别目标物体，它能够抓取灰度变化显著的区域。HOG 由 Dalal 和 Triggs 于 2005 年提出，并首先应用于人体检测。HOG 特征借鉴了 SIFT(Scale Invariant Feature Transform) ^[19]特征点中运用梯度方向直方图表示目标的思想，用梯度直方图来描述图像特征。HOG 的基本原理是：首先将原图像转化为灰度图像并对其进行颜色空间标准化等一系列操作，对图像中的每个像素求梯度，之后将图像划分成 cell 并对 cell 中的梯度进行统计得到梯度直方图特征。将相邻几个 cell 组成 block 得到该 block 的 HOG 特征描述子，将 block 特征描述子串联成一个长向量就得到了最终的特征描述子。LBP 特征可以表述纹理，对单调灰度变化有敏感性。当背景比较复杂，特别是有干扰边缘时，HOG 特征将受到很大影响，而此时 LBP 特征可以滤除背景噪声。深度学习^[20, 21]是在研究传统 BP 神经网络的过程中提出的，它的基本原理是通过样本的基本信息学习目标的底层表示，再通过对底层表示的学习，形成目标的高层抽象表示。

分类器的评分最终决定该窗口是否为目标。在分类器的设计方面，主要可以分为两种方法：基于概率的方法（Probabilistic Method）^[22]、基于判别的方法（Discriminative Method）^[23, 24]，目标检测研究的初期，研究者们多采用基于概率的方法，如模板匹配，Fisher 判别等。后来，由于判别算法训练简单，检测效果好，人们逐渐地转向判别算法。目前比较常用的方法如 SVM (Support Vector Machine, 支撑向量机)，Adaboost 算法都属于基于判别的方法。SVM^[23-25]算法是

二分类模型，其主要思想为获得特征空间上的间隔最大的线性分类器，最终可转化为一个凸二次规划问题的求解。考虑到样本的非线性分布，SVM 方法中提出使用核理论将样本投影到高维空间中，并将核理论中的内积形式运用到优化模型的对偶规划中，进而在高维空间中求解线性分类器。DPM（Deformable Parts Model，可变形部件模型）^[26-28]可以看成是 HOG 的扩展，大体思路与 HOG 一致。先计算梯度方向直方图，然后用 SVM 训练得到物体的梯度模型。Adaboost^[29] 算法是 Boosting^[29-31] 算法中典型的代表算法之一，从结构方面来讲，Adaboost 设计多个弱分类器，将多个弱分类器级联起来，前一级弱分类器的输出作为后一级弱分类器的输入。Adaboost 通过每一个弱分类器检测时的错误率得到他们的权值，最终结果由各个弱分类器加权投票得到。

1.3 本文研究内容

本课题的研究目标是在保持系统灵活性的前提下，针对受到无线网络约束或者是通信资源受限的情况，提出一种有效解决资源受限问题的目标识别系统框架。在通信资源受限的情况下，目标识别系统的主要矛盾有两点：第一，带宽不足及信道稳定性不高与待传输数据量过大之间的矛盾。针对这种情况，主要的解决方案是将一部分目标识别系统移植到无人机端（无人机搭载的嵌入式平台），只发送经过前端目标识别系统处理后的识别结果。第二，由于目标识别模型向嵌入式平台移植导致计算资源不足，造成目标识别模型工作需要的计算资源与嵌入式设备较低的硬件配置形成了矛盾。本文为了缓解计算资源的压力，降低前端目标识别系统对识别精度的要求，保证召回率（召回率：检测到的正确的目标/目标总数）的前提下减少待传输的数据量，在前端与服务器端分别搭载一套目标识别系统共同合作来完成全部的目标识别过程。

本文主要研究内容如下：

1)设计一种新型目标识别系统框架。从理论上比较各个方案的区别和优劣，分析各个方案的可行性，确立最终的框架方案。

在航拍检测场景下将抽象的框架具体化，设计具体的实现方案。主要包括两点：第一，搭建系统的硬件平台和软件平台。第二，设计目标识别系统内置的各个算法模型，并且模拟分析各个算法模型性能的优劣。

2)根据指定方案实现模拟系统并验证框架的有效性。以 SDL 高清航拍数据集为数据来源，通过 PC 机与服务器级联来模拟整个动态级联目标识别系统。测

试并分析新型动态级联目标识别系统的各项指标，验证新型系统的合理性和有效性。

3) 在嵌入式平台和服务器上搭建实际的动态级联目标识别系统，以信号衰减器来模拟带宽实际的变化，定制沙盘和航模来搭建模拟的小型飞机场，通过高清摄像头实时“航拍”数据，使得整个系统运行在与实际情况最逼真的航拍场景之下。

1.4 本文的组织结构

第一章，绪论。介绍本文研究的问题背景和意义、当前研究领域的研究状况的综述、论文的研究任务以及组织结构。

第二章，相关技术研究。本章是承上启下的章节，主要是对本文使用的相关技术的介绍，包括 Adaboost 模型、DPM 模型、卷积神经网等内容。

第三章，通信资源受限下的目标识别系统框架。主要通过分析传统目标识别系统以及各个改进方案的优劣来提出新型动态级联目标识别系统。确定了框架后，通过对比回将各个模型内置于本文场景下的性能来确定系统最终内置的模型。

第四章，动态级联架构目标识别系统。实现模拟系统，通过在 PC 机上模拟系统来分析新型系统各个指标，进而验证本文提出的框架是否合理。将前端目标识别系统移植到嵌入式平台，模拟系统升级为实际系统。

第五章，总结与展望。主要对本文的研究工作进行总结以及对下一步工作的开展进行展望。

第2章 目标识别系统的相关技术研究

第1章简要介绍本论文的问题背景和意义、当前相关领域的研究综述、本文的主要研究内容和本文结构安排。本章是第一章的延伸，将介绍第三章、第四章研究工作中使用到的算法和相关技术。

2.1 Haar-like 特征

在本文实现的动态级联系统当中，我们使用 Haar-like 特征作为检测图像的特征。Haar-like 特征分为三类：边缘特征、线性特征、中心特征。下面图 2-1 以边缘特征为例来说明 Haar-like 特征的计算方法。图中的特征模板有白色和黑色两种矩形区域，该模板的特征值为白色矩形区域的像素和减去黑色矩形区域的像素之和。Haar-like 特征值反映了图像的灰度变化情况。

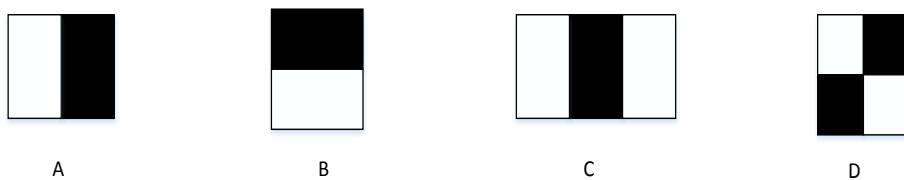


图 2-1 Haar-like 特征模板

对于图 2-1 所示的 A, B 和 D 特征模板，特征数值计算公式为： $v = \text{Sum 白} - \text{Sum 黑}$ ，而对于 C 来说，计算公式如下： $v = \text{Sum 白} - 2 * \text{Sum 黑}$ 。通过改变特征模板的大小和位置在子窗口内穷举出大量的特征。计算 Haar-like 特征，通过积分图可以极大地节省计算量。积分图的主要思想是构建一个数组保存在内存中，数组中的元素为从起点开始到各个像素点所形成的矩形区域像素之和。在特征提取过程中，计算某个区域的像素和时可以直接索引数组的元素，因此，积分图加快了计算速度。

2.2 决策树

在本文中，使用 Adaboost 作为前端系统的目标识别算法。Adaboost 是级联结构，由多个弱分类器级联构成一个强分类器，而决策树方法是 Adaboost 中构造弱分类器常用的算法。构造决策树的方法有很多，比较常用的是 ID3、C4.5 和 CART 方法。在 1966 年，Hunt 等人^[32]提出了第一个构造决策树的算法 CLS，后

来的许多决策树算法都可以看作是对 CLS 算法的改进与更新。Quinlan 于 1986 年提出了 ID3 (Iterative Diehotomize) 算法^[33], 以信息嫡作为选取分裂属性的度量标准。1993 年, Quiulan 提出了 C4.5 算法^[34], 使用信息增益率替代信息增益来选择属性。分类和回归树 (CART, Classification and Regression Tree) 算法^[35]是由 Breiman 等于 1984 年提出的一种经典决策树算法, 不仅可以用于传统的分类, 还可以用作回归。由于 ID3 算法与 C4.5 算法非常相似, 下面以 C4.5 算法为例来介绍构建决策树算法。

决策树是一个树状模型, 训练时从根结点开始自上而下进行结点分裂, 在当前结点对所有属性进行测试, 选取最优属性将数据集合空间分割成多块, 然后对子结点递归调用以上方法, 最后构建出决策树模型, 每个叶结点是带有分类标签的数据分割。决策树模型如图 2-2 所示。

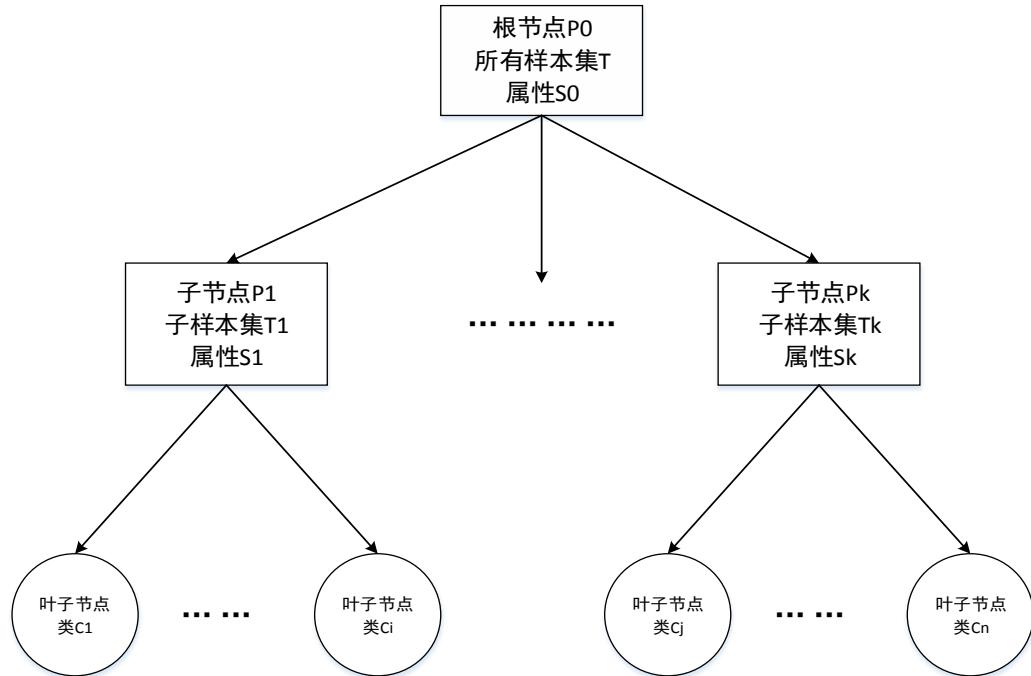


图 2-2 决策树结构

不同的算法生成决策树的方式有所不同。下面以 C4.5 例来介绍决策树生成的方法, 了解 C4.5 算法必须要了解信息增益以及信息增益比的概念。信息增益表示由于特征 A 而使得数据集 D 的分类的不确定性减少的程度。但是使用信息增益作为划分训练数据集的特征, 存在偏向于选择取值较多特征的问题, 信息增益比对这一问题进行了校正。信息增益和信息增益比的算法如表 2-1 所示。

表 2-1 信息增益算法与信息增益比算法

信息增益算法流程如下：

输入：训练数据集 D , 特征集 A

输出：特征 A 对训练数据集 D 的信息增益

1. 计算数据集 D 的经验熵 $H(D)$

$$H(D) = \sum_{k=1}^K \frac{|C_k|}{|D|} \log_2 \frac{|C_k|}{|D|} \quad (1)$$

2. 计算特征 A 对数据集 D 的经验条件熵 $H(D|A)$

$$H(D|A) = \sum_{i=1}^n \frac{|D_i|}{|D|} H(D_i) = - \sum_{i=1}^n \frac{|D_i|}{|D|} \sum_{k=1}^K \frac{|D_{ik}|}{|D_i|} \log_2 \frac{|D_{ik}|}{|D_i|} \quad (2)$$

3. 计算信息增益 $g(D, A)$

$$g(D, A) = H(D) - H(D|A) \quad (3)$$

4. 计算信息增益比 $g_R(D, A)$

$$g_R(D, A) = \frac{g(D, A)}{H_A(D)} \quad (4)$$

其中，

$$H_A(D) = \sum_{i=1}^n \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|}, \quad n \text{是特征 } A \text{ 取值的个数}$$

ID3 算法通过计算各个特征的信息增益来确定用哪个特征作为当前节点的分类标准，以信息增益作为划分训练数据集的特征，存在偏向于选择取值较多的特征的问题。为了解决这个问题，C4.5 算法通过计算各个特征的信息增益比来确定用哪个特征作为当前节点的分类标准，这是 ID3 算法同 C4.5 算法的主要不同。C4.5 生成决策树的方法如表 2-2 所示。

表 2-2 C4.5 算法流程

C4.5 算法流程如下：

输入：训练数据集 D , 特征集 A , 阈值 ϵ

输出：决策树 T

1. 如果 D 中所有实例属于同一类 C_k , 则置 T 为单节点树，并将 C_k 作为该节点的类，返回 T ;

2. 如果 $A = \emptyset$, 则置 T 为单节点树，并将 D 中实例数最大的类 C_k 作为该节点的类，返回 T ;

3.否则，按照表 2-1 所示方法计算 \mathbf{A} 中各个特征对 \mathbf{D} 的信息增益比，选择信息增益比最大的特征 \mathbf{A}_g ；

4.如果 \mathbf{A}_g 的信息增益比小于阈值 ϵ ，则 T 为单节点树，并将 \mathbf{D} 中实例数最大的类 \mathbf{C}_k 作为该节点的类，返回 T ；

5.否则，对 \mathbf{A}_g 的每一个可能值 a_i ，依 $\mathbf{A}_g = a_i$ 将 D 分割为子集若干非空 D_i ，将 D_i 中实例数最大的类作为标记，构建子结点，由结点及其子结点构成树 T ，返回树 T ；

6.对结点 i ，以 D_i 为训练集，以 $\mathbf{A} - \{\mathbf{A}_g\}$ 为特征集，递归的调用步 1~步 5，得到子树 T_i ，返回 T_i ；

决策树模型简单直观，容易理解，计算效率非常高；决策树方法能够有效的抑制训练样本噪声。但是自上而下的逐层判别可能会过于适应噪声，从而导致过度拟合。

2.3 Adaboost 算法

本文在前端目标识别系统中使用 Adaboost 作为目标识别算法。Adaboost 是 Boosting 方法的一种，其本身是由多个弱分类器级联构成的强分类器。在保证召回率的前提下，通过改变 Adaboost 运行的级数可以改变检测结果，这与前端目标识别系统需要动态改变需要传送的数据量的需求相一致。Adaboost^[29, 36, 37]是一种迭代算法，是 Boosting 方法中的典型算法之一。算法本身是通过改变数据的分布实现的，它根据每次训练集中的每个样本的分类是否正确，以及上次的总体分类的准确率，来确定每个样本的权值，分类器会减小分类正确样本的权值并增加分类错误样本的权值，这样，分类器会越来越重视那些分错的样本。分类器会将修改权值的新数据送给下层分类器进行训练，然后将每次训练得到的弱分类器串联起来，作为最后的决策分类器。

表 2-3 Adaboost 算法流程

Adaboost 算法流程如下：

输入：训练数据集 $T=\{(x_1, y_1), (x_i, y_i), \dots, (x_n, y_n)\}$ ，其中， $x_i \in \mathcal{X} \subseteq \mathbb{R}^n$ ，

$y_i \in \mathcal{Y}=\{-1, +1\}$ ；

输出：最终分类器： $G(x)$

1. 初始化训练数据的权值分布

$$\mathbf{D}_1=(w_{11}, \dots, w_{1i}, \dots, w_{1N}), w_{1i} = \frac{1}{N}, i = 1, 2, \dots, N \quad (1)$$

2. 对于 $m=1, \dots, M$:

a). 使用具有权值分布 \mathbf{D}_m 的训练数据集学习, 得到基本分类器:

$$\mathbf{G}_m(x): \mathcal{X} \rightarrow \{-1, +1\} \quad (2)$$

b). 计算 $\mathbf{G}_m(x)$ 在训练数据集上的分类误差率

$$e_m = P(\mathbf{G}_m(x) \neq y_i) = \sum_{i=1}^N w_{mi} I(\mathbf{G}_m(x_i) \neq y_i) \quad (3)$$

c) 计算 $\mathbf{G}_m(x)$ 的系数:

$$\alpha_m = \frac{1}{2} \log \left\{ \frac{1-e_m}{e_m} \right\} \quad (4)$$

这里的对数是自然对数

d) 更新训练数据集的权值分布:

$$\mathbf{D}_{m+1} = (\mathbf{w}_{m+1,1}, \dots, \mathbf{w}_{m+1,i}, \dots, \mathbf{w}_{m+1,N}) \quad (5)$$

$$\mathbf{w}_{m+1,i} = \frac{\mathbf{w}_{mi}}{Z_m} e^{(-\alpha_m y_i \mathbf{G}_m(x_i))}, \quad i = 1, 2, \dots, N \quad (6)$$

其中 Z_m 是规范化因子

$$Z_m = \sum_{i=1}^N \mathbf{w}_{mi} e^{(-\alpha_m y_i \mathbf{G}_m(x_i))}, \quad i = 1, 2, \dots, N \quad (7)$$

3. 构建基本分类器的线性组合

$$f(x) = \sum_{m=1}^M \alpha_m \mathbf{G}_m(x) \quad (8)$$

得到最后的分类器:

$$G(x) = \text{sign}(f(x)) = \text{sign}(\sum_{m=1}^M \alpha_m \mathbf{G}_m(x)) \quad (9)$$

2.4 卷积神经网

本文在服务器端目标识别系统中使用的目标识别模型是 Fast CNN (Fast Convolutional Neural Networks, 快速卷积神经网)。在服务器端目标识别系统中, 目标识别算法需要在保证检测速度的同时保证准确率。CNN 是目前目标检测领域准确率非常高的一种模型, 在此对 CNN 和 Fast CNN 做一些简单的介绍。

CNN (Convolutional Neural Networks, 卷积神经网)^[38-41]是近年来广泛应用于模式识别、图像处理领域的一种高效的算法。神经元卷积神经网最基本的组成单元, 如图 2-3 所示。

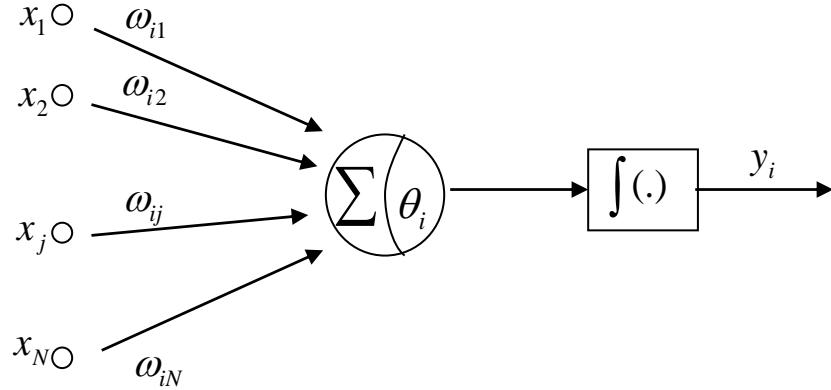
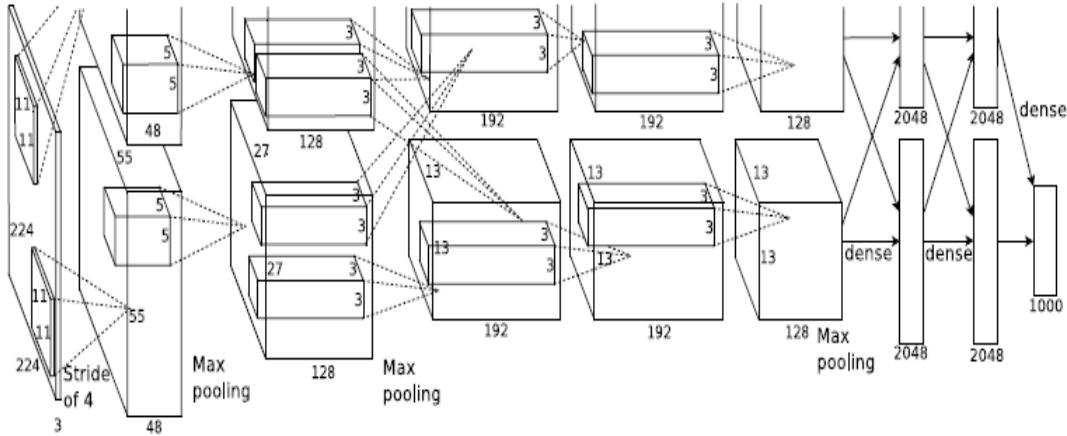


图 2-3 神经元结构图

神经元的模型如下：

$$y_i = \mathcal{F}(\sum_i \omega_{ij} - \theta_i) \quad (2-1)$$

式中， θ_i 偏移量， $f(\cdot)$ 是激活函数，激活函数比较常用的有 Sigmoid 函数等。卷积神经网络是一个多层的神经网络，每层由多个二维矩阵组成，而每个平面由多个独立神经元组成。

图 2-4 卷积神经网结构示意图^[38]

卷积神经网分为输入层、卷积层、pooling 层、全连接层、输出层。

输入层：图 2-4 第一层为输入层，卷积神经网以原图像作为模型的输入。

卷积层：卷积层的输入是上一层的输出。在卷积层中，将卷积模板与输入数据进行卷积得到输出结果。

Pooling 层：池化。池化是利用图像局部相关性的原理，对图像进行子抽样，可以减少数据处理量同时保留有用信息，有缩小隐藏节点对于卷积特征维数的

作用。

全连接层：全连接层将二维数据转化为一维数据，通过加权求和得到最终结果。

输出层：由全连接层加权求和得来，输出结果为类别。

RCNN^[42, 43]是目标检测领域非常有效的一个模型。RCNN 通过 Selective Search 提取候选框，通过 CNN 提取特征，最后用 SVM 作为分类器。Fast CNN^[42, 43]可以看做是 RCNN 进一步加速，很大程度上降低了 RCNN 的时间开销和空间开销。

2.5 本章小结

本章对于本文中涉及到的主要的相关技术做了介绍。第一节介绍了 Haar-like 特征的基本原理，Haar-like 特征是前端目标识别系统提取的特征。第二节以 C4.5 为构建决策树的算法对决策树做了详细的介绍，决策树算法是 Adaboost 算法中构建弱分类器的算法。第三节介绍了 Adaboost 算法，Adaboost 算法是前端目标识别系统使用的目标识别算法。第四节介绍了卷积神经网，卷积神经网是服务器端目标识别系统使用的目标识别算法。通过以上相关技术的介绍，为后续无线网络约束下目标识别系统的研究进行了铺垫。

第3章 无线网络约束下的目标识别系统框架

上一章介绍了与本文研究工作相关的技术，本章将会详细分析本文所提出的动态级联目标识别系统框架的设计思路。在航拍检测等高灵活性的场景下，传统目标识别系统还处于一种离线式的机制。针对目前传统目标识别系统的不足，本文与实际情况相结合，提出一系列改进方案（如图 3-1 所示）并做了可行性分析。最终，本文提出了一种灵活的在线式的无线网络目标识别系统框架。下面，我们通过分析传统离线式目标识别系统和两种中间改进方案的目标识别系统来引入新型在线式的动态级联目标识别系统并介绍其优势。

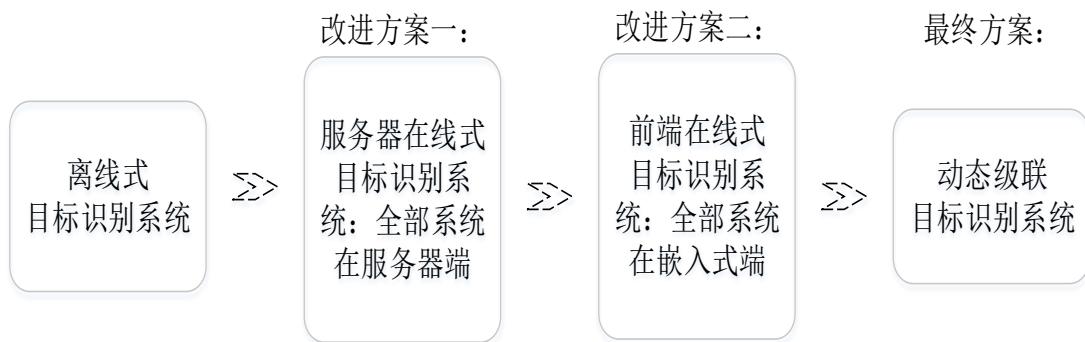


图 3-1 传统系统到新型系统改进方案

3.1 传统离线式目标识别系统

目前，市场上应用的传统目标识别系统的流程图如图 3-2 所示。传统目标识别系统采用离线式的机制，即：数据获取阶段与目标识别阶段是分离的。具体来说，将采集到的数据保存到放置在监测现场的存储设备当中，工作人员定期将存储设备中的数据拷贝到服务器，用于后期目标识别工作。数据获取和目标识别分成了两个阶段分别完成。

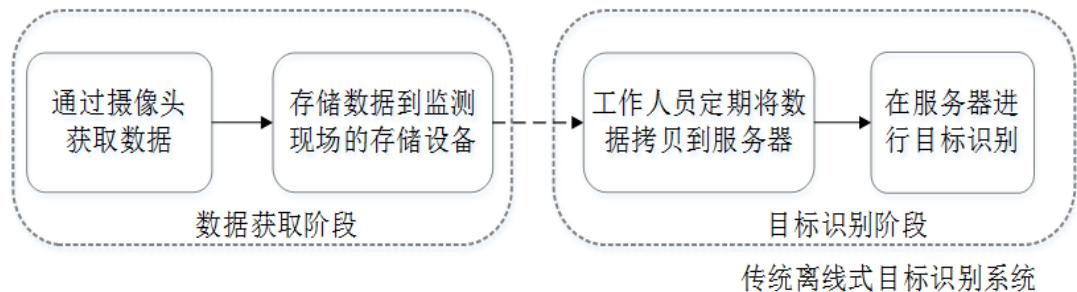


图 3-2 传统的目标识别系统流程图

通过传统目标识别系统的检测过程可以得知它本身具备以下几点不足：

1) 离线式。传统的目标识别系统是一个非智能化的系统，其最显著的特点在于它离线式的机制，即只是单纯将采集到的数据保存下来，数据的获取和处理不是同时进行的。这导致它在许多需要及时反馈信息的场景之下无法应用。

2) 人工辅助。由于传统目标识别系统的数据采集与目标识别两部分工作不是自动进行的，而是需要人工干预，所以在智能化、自动化方面，传统目标识别系统具有明显的不足。

综上所述，传统目标识别系统由于其离线式的机制，在很多需要实时给出反馈信息的场景下难以得到应用，针对上述问题，本文以无人机实时监测飞机场作为研究背景，对传统目标识别系统做了一系列的改进，最终得出一种能够更好适应更多场景的在线式动态级联目标识别系统。

3.2 传统离线式目标识别系统改进方案

传统目标识别系统由于其离线式的机制导致系统不能够实时的反馈信息。因此，本文将目标识别系统与无线网络相结合提出改进方案一。

3.2.1 改进方案一：全部在线式目标识别系统在服务器完成

由前文可知，传统目标识别系统由于其离线式机制难以在需要及时反馈信息的场景下应用。所以，将离线式机制转换为在线式机制是新型系统的首要目标，即：前端获取图像后立即发送给服务器上的目标识别系统，目标识别系统在收到下一幅图像之前完成本幅图像的目标识别工作，循环执行上述过程。上述过程与传统目标识别系统离线式机制的不同在于数据获取阶段与目标识别阶段不再是分离开来的，而是同步进行的，这样就将传统目标识别系统离线式的机制转换成了在线式的机制。具体做法是：在监测现场安置摄像头来获取监测图像，将获取到的数据实时通过通信网络发送给服务器上的目标识别系统，目标识别系统对图像实时处理，如果发现异常目标立即发出警报。这样就解决了需要实时反馈信息的需求，扩大了目标识别系统的应用场景范围，改进方案一如图 3-3 所示。

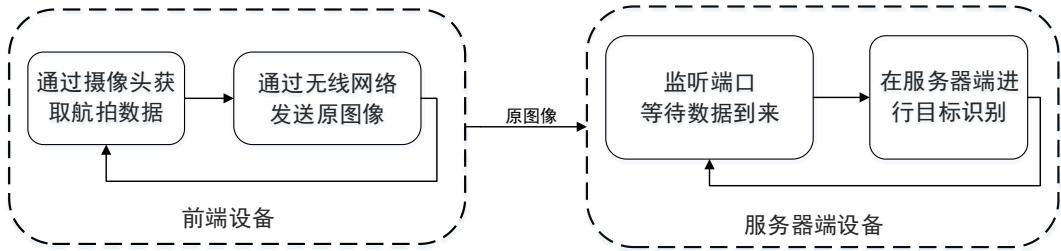


图 3-3 改进方案一流程图

在这种解决方案中，前端系统（监测现场的设备）到服务器端系统（研究中心的设备）的数据传输问题是整个系统实现的瓶颈。在数据传输媒介的选择上，由于部分应用场景下存在现场有线网络布线困难，设备繁琐等问题，甚至在一些应用场景下（比如本文的无人机航拍场景）无法布置有线网络。所以通过无线网络传输数据在很多情况下是一个较优的选择。而无线网络最显著的两个特点是带宽受限和稳定性不好。传送整幅高清图像超过了整个系统所能承受的最大限度。

具体来说，在带宽受限方面，主要体现为传输数据速度慢，信道容量不足。目前，国内外大型高端无人机往往使用卫星通信来达到数据传输的目的，如美国的“全球鹰”无人机等，其通信带宽需要 560Mbps 来支持其操控和侦查图像的实时回传。日本军方在采购到全球鹰无人侦察机后发现其卫星资源无法提供足够带宽供全球鹰系统使用，极大的降低了该系统的使用效能。即使是美军也同样面临卫星资源有限，即同时可部署的全球鹰系统数目有限，限制了其作战效能。

在国内，由于卫星信道资源有限，我国无人机主要采用射频通信作为无人机测控通信的主要技术手段。然而无论是卫星信道或是地面基站式射频信道，其通信资源都受到香农定理与信号衰落的约束。其示意图在图 3-4 中给出，随着通信距离的增加，信噪比（图像的信噪比等于信号与噪声的功率谱之比，是衡量信道质量的一个重要指标）会有所下降，根据香农定理(公式 3-1)，信噪比越小，信道容量就越小，支持的最大吞吐量就越小。

$$R_{max} = W \times \log_2(1 + S/N) \quad (3-1)$$

其中， R_{max} 表示信道容量， W 表示信道带宽， S/N 表示信噪比。

$$P_{rss} = P_{tx} - PL(d) + X_{\sigma} \quad (3-2)$$

式中， P_{rss} 为接收端信号强度， P_{tx} 为发射设备功率， $PL(d)$ 为路损分量：

$$PL(d) = PL(d_0) + 10n \log_{10} \left(\frac{d}{d_0} \right) \quad (3-3)$$

该模型通常被称为对数路径损耗模型。其中， d 为发射节点和接收节点之间的距离，可由相对导航定位系统直接估计得出，也可由 GPS 输出的绝对地理位置经地球曲率模型解算其欧式距离。

公式（3-3）中 n 为路径损耗指数， n 可以表征路径损耗随距离增加而增大的快慢程度，与周围的环境相关，在单一作战场景下通常可以被认为是恒定值。 d_0 是参考距离， $PL(d_0)$ 表示参考距离为 d_0 时的路径损耗。该模型描述了理想信道环境，如空基数据链通信场景，信号强度随收发设备间距离增长而衰减的关系。考虑射频系统在小时间尺度上热噪声系数的恒定性，容易得出网络内特定链路质量与链路两端节点间距离的关系符合单调但是非线性的映射。这个映射中对数关系导致的非线性成分使得最小跳数等传统路由选择算法在无线网络中性能严重降低。总结而言，地理距离最近的原则能够保证局部链路质量最优，但是无法保证全局路由端到端的质量最优。

公式（3-2）中， X_σ 为表示信道衰落随环境变化影响的随机分布向量，描述了由于周边环境，如山体、植被、建筑等导致的信号反射和在接收端汇聚导致的多径效应，导致的等效信号强度下降效应。在慢衰落等场景下， X_σ 可以被描述为一个均值为 0，标准差 σ 的高斯分布随机变量， σ 值越大模型的不确定性越大。陆基战术数据链场景下的信道更适合被描述为同时存在直视信道和多径信道的莱斯（Rician）分布，其接收信号包络的概率密度函数(PDF)为：

$$P(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2+s^2}{2\sigma^2}\right) I_0\left(\frac{rs}{\sigma^2}\right) \quad (3-4)$$

式中： r 为接收到的信号幅度包络； σ^2 为多径分量的方差； s^2 为主信号的功率，可由公式（3-4）描述； I_0 为第一类 0 阶修正贝塞尔函数。

图 3-4 表示随着通信距离的增加信噪比的衰减程度。横轴表示距离，db 定义为有用信号与所有信号（包括干、扰信号）的比值。由图可见，随着距离的增加，信噪比越来越小，由香农定理知，信道容量也会随着信噪比的减小而减小。所以，随着通信距离的增加，信道支持的最大吞吐量就会减少，导致传送原图像非常困难。由于前端系统距离基站距离比较远，一般在可视范围之外，当通信距离逐渐增加时，信噪比和信道容量会越来越小，难以满足直接传送高清原图像的需求。此外，射频信号在不同地表环境（如建筑物，山体等）的作用下，通常还将受到衰落效应（即 fading）的影响，产生如图 3-4 所示的信号质量波动，从而进一步影响信道容量。

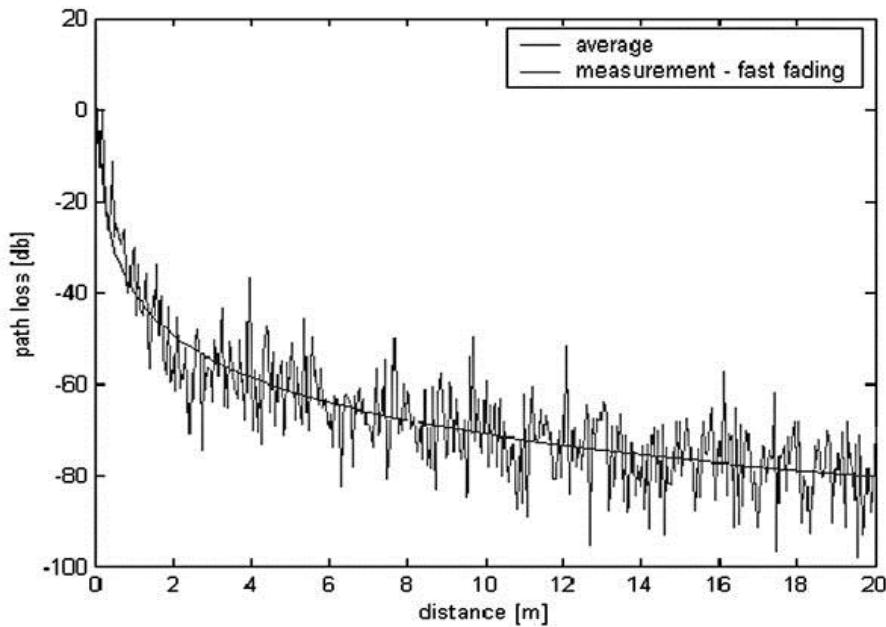
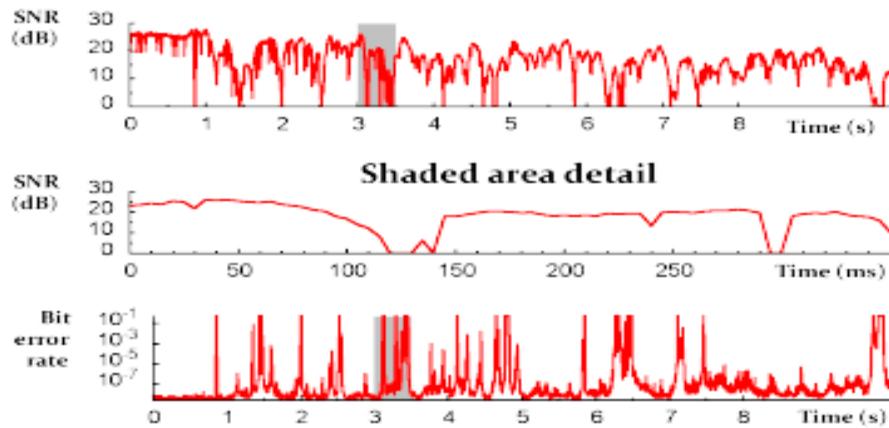
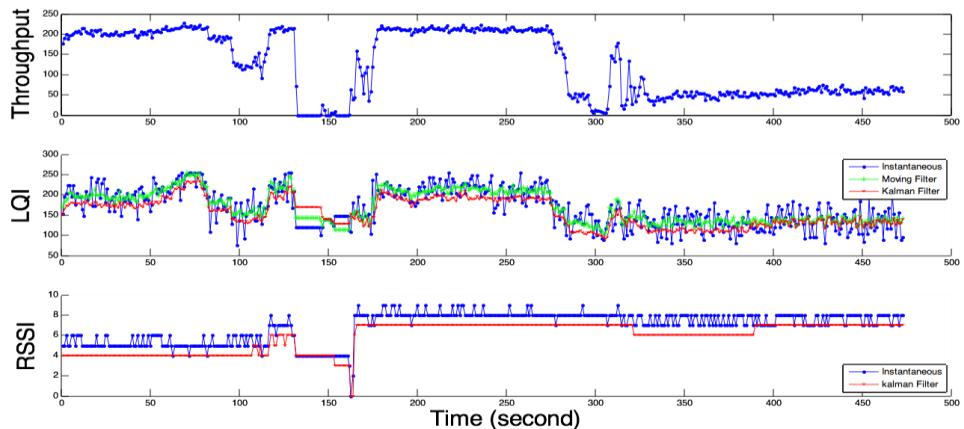


图 3-4 路径衰减

该效应在图 3-5^[44, 45]与图 3-6^[46]中给出两个典型的信道环境变化实测图样。其中图 3-5 为美国麻省理工大学利用 IEEE802.11 设备在移动环境测量所得，图 3-6 为在英国 UCL 大学利用 IEEE802.15.4 设备在教学楼环境下固定部署测量所得。图 3-5 表示移动环境下的链路质量，无人机与服务器之间的无线信道正是一个移动环境下的链路。图 3-6 前两幅图横轴为时间，纵轴为信噪比。由图可见，移动环境下的通信链路状况非常不稳定。图 3-5 中的第三幅图表示错误率与时间的关系，可见，在较长时间内，移动链路的错误率同样稳定性不好。图 3-6 表示静止状态下链路质量变化样本，在无人机悬停时，无人机与服务器间的无线信道是一个静止的链路，由图 3-6 可见，即使在静止状态下，由于信道受到周边环境的影响，仍然是一个不稳定的信道。对于直接传送原图像的需求来说，需要通信链路在较长时间维持在一个比较稳定的状态，这与无线信道的不稳定性相矛盾。

图 3-5 移动环境下的链路质量变化样本^[45]图 3-6 静止环境下的链路质量变化样本^[46]

综上所述,由于有线网络的布线困难和无线网络带宽受限和信道不稳定等特点,数据的传输成为了限制整个系统性能的主要问题,直接由前端设备向服务器端设备传送原图像的方案并不可行。在改进方案一的基础上,可行的改进方案是避免数据传输(方案二)或者减少需要传输的数据量,从而减少数据传输对无线网络带宽带来的压力,并且短时间内维持一个稳定的信道环境相对更加可行。

3.2.2 改进方案二: 全部在线式目标识别系统在前端完成

根据前文分析,实时传送高清航拍原图像回服务器不具备可行性,那么,另一种解决方案是将全部目标识别系统装载在监测现场(前端),通过高清摄像头获取航拍图像之后,直接传送给无人机上搭载的嵌入式设备,运行在嵌入式设备上的目标识别系统接收到图像后,直接在监测现场完成目标识别的全部工作,将

识别的结果通过无线网络传送到服务器端。这种方案能够避免改进方案一中数据传输给系统带来的限制，改进方案二如图 3-7 所示。

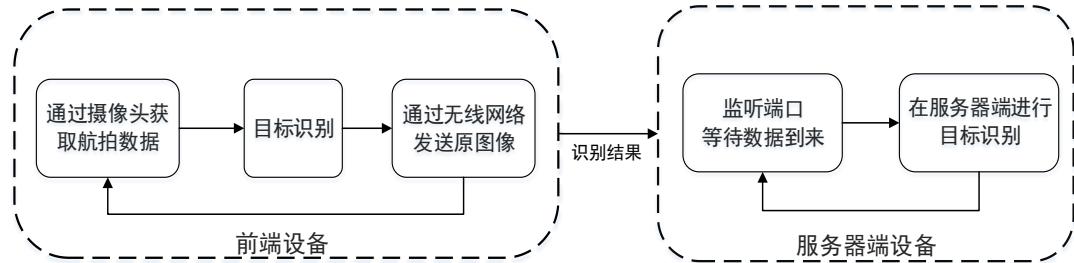


图 3-7 改进方案二流程图

受到 MPEG (Moving Picture Experts Group, 动态图像专家组) 制定的动态图像压缩编码方式的启发，我们提出了改进方案二。MPEG 压缩方法的原理是：如果两幅图像有相同的内容，则第一幅如果已经完整传送了整幅图像的信息，那第二幅只传送两幅图像之间的差别即可。然而，在本文的无人机航拍检测的背景下，由于航拍图像的角度变化非常大，导致图像之间相似的内容不多，使用 MPEG 方法效果并不明显。因此，改进方案二通过目标识别系统来识别出图像中含有信息量的部分，进而在保证信息量不丢失的情况下减少需要传送到研究中心的数据量。

从应用角度说，前端目标识别系统更注重便捷性，因此，前端系统需要根据现场情况实时改变自身位置角度等。例如本文的无人机航拍场景，如果无人机载重过多的话，势必会对无人机的灵活性以及续航时间造成负面影响。目前，广泛意义上的的无人机载重在 20kg 以内，如比较知名的大疆创新科技有限公司生产的无人机产品，在全球属于领先行列，载重一般在 10kg 以内。在目前硬件水平条件下，通过无人机搭载大型的设备进行空中目标识别并不合适，并且大型设备对无人机能源的消耗比较多，严重缩短了无人机单次作业的周期。所以，无人机搭载灵活轻便、造价适宜的嵌入式设备是比较好的选择。

如果能够将目标识别算法完全运行在嵌入式设备之上，那么新型在线式的目标识别系统将能够得到广泛的应用。然而嵌入式设备的运算能力有限，性能精确、复杂度高的模型难以在嵌入式上运行。比如 CNN (Convolution Neural Network, 卷积神经网)，AP (Average precision, 平均准确率) 能够达到 80% 以上，然而运行时需要占用非常大的内存，存储模型亦需要非常大的存储空间。在运算能力方面，根据本文调研，在 NVIDIA Tesla K20 (2688cores, 显卡核心数目)，3.95T

FLOP (Floating Point Operations Per Second, 每秒所执行的浮点运算次数) 上用 CNN 检测一幅图像需耗时 1s 左右, 而目前业内领先水平的嵌入式板卡, 即本文采用的 NVIDIA 生产的 Jetson Tegra K1 嵌入式板卡, 有 2G 内存, GPU 配置也仅仅是 192 核, FLOP 在 300G 左右, 与服务器硬件配置有数量级上的差别。CNN 在服务器较好的硬件资源配置下, 检测一幅图像需要十几秒。因此, CNN 这类复杂的目标识别模型运行在嵌入式设备上, 就目前而言, 技术还不成熟。所以, 在目前国内外硬件水平的限制之下, 嵌入式设备计算资源不够充分, 全部目标识别过程在前端完成还不可行。

3.3 新型动态级联目标识别系统

前两种解决方案由于数据传输问题和当前硬件设备的限制导致其实现的可行性不高。针对上述问题, 本文将目标识别系统整体分成前端目标识别系统和服务器端目标识别系统两部分, 在前端(监测现场)和服务器端各自搭载一套目标识别系统, 前端目标识别系统的输出作为服务器端目标识别系统的输入, 在无人机航拍场景之下, 二者共同构成一套空地级联的目标识别系统。但是由于前端目标识别系统与服务器目标识别系统在整个系统扮演的角色不同, 对二者的要求和选择也有所不同。前端目标识别系统主要的作用是在保证不丢失信息量的前提下, 减少需要通过无线网络传输的数据量。因此在前端需要配置一个对计算资源消耗较低并且能够保证召回率的目标识别模型, 把前端目标识别系统的识别结果通过无线网络传送给服务器, 作为服务器端目标识别系统的输入。经过前端目标识别系统完成目标识别后, 识别结果去除掉了前端目标识别系统认为一定不包含目标的那部分图像, 识别结果所保留的数据量要远小于原图, 只需要将剩余的这部分数据传回服务器端, 这样就在计算资源受到约束的情况下满足了通信资源对整个系统的约束, 在不丢失信息量的同时减小了需要通过无线网络传送的数据量, 使得在线式目标识别系统具备了可行性。由于网络带宽状况本身是实时变化的, 带宽好与带宽不好时传输数据量应当是有所区别的, 所以要求前端系统能够根据带宽状况实时自动改变识别结果, 即传输的数据量应该是根据需求是实时变化的, 使得整个系统在不同的情况之下都能够确定好最合适的数据量来通过无线网络传送。

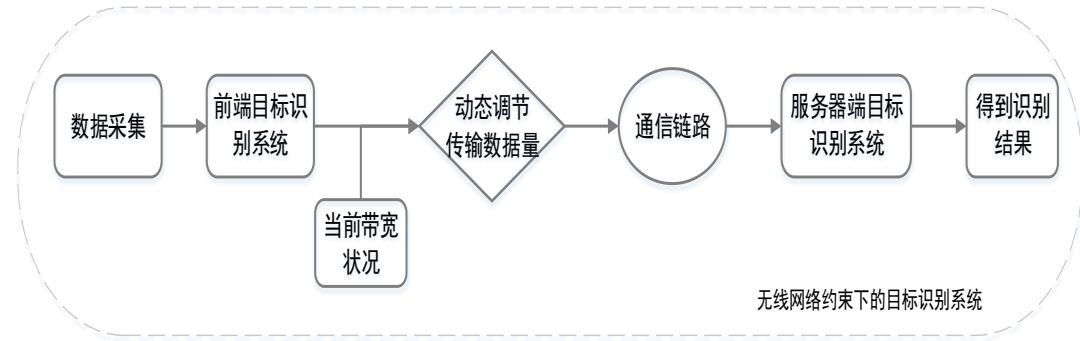


图 3-8 无线网络约束下的目标识别系统

图 3-8 表示为提出的在线式动态级联目标识别系统框架，其各部分介绍如下：

- 1) 数据获取。系统通过布置在现场（前端）的高清摄像头实时获取数据。将获取到的数据作为前端目标识别系统的输入。
- 2) 前端目标识别系统。前端目标识别系统是新型目标识别系统的关键部分，用于第一步的目标检测工作，承担着实现计算资源与通信资源相互转换的任务。前端目标识别系统应该具备以下特点。首先，由于通信资源是实时变化的，前端目标识别系统的模型应该能够根据带宽的情况自动改变模型的检测结果以达到改变传输数据量的目的，带宽资源充分时，多传送一些数据，带宽资源不好时，少传送一些数据。能够通过增加计算量来增加模型识别精度，进而实现了通信资源与计算资源相互转换的机制，有效地解决了在线式系统资源受限的问题。其次，前端系统计算资源不应消耗过多，检测时间不应过长。由于无线网络通信带宽是实时变化的，它与前端系统距离基站的远近、干扰物等等一些客观因素关系非常大。在本文应用场景下测试，将新型目标识别系统应用于本文的无人机航拍场景之下时，级联系统经过前端目标识别系统后，传输数据量下降 20%~80%，并且能够在不同带宽下保证信息量不丢失。

- 3) 服务器端目标识别系统。服务器端目标识别系统最终决定检测到的物体是否为目标。在服务器端，软硬件资源充分，不必考虑计算资源方面的限制，因此在服务器端目标识别系统可以采用复杂度较高，性能较好的模型来完成最终检测的任务。

与传统目标识别系统相比，新型系统具有明显优势：

- 1) 智能化，在线式。新型智能化无线网络目标识别系统从采集数据到得到监测结果的整个过程，不需要人工干预，它将整个目标识别系统一分为二，由前

端目标识别系统与服务器端目标识别系统两部分构成，前端系统先对原图像进行检测，将检测结果通过无线网络发送给服务器端目标识别系统进行下一步识别工作，有效地解决了计算资源与通信资源不足的问题。

2) 动态性。由于网络带宽状况本身是实时变化的，带宽好时与带宽不好时传输数据量应有所区别。所以，前端目标识别系统需要能够按照带宽的状况来动态改变传输的数据量，最大化地利用了通信资源与计算资源，并且使得二者可以相互转化，使得资源分配更为合理。

3) 准确性。新型系统在保证减少通过无线网络传送数据量的同时，也保证了信息量不会在前端系统目标识别的过程中丢失，使得整个系统的性能不会因为级联而下降。

4) 灵活性、便捷性。新型系统解决了传统系统不易携带、检测现场不易布置系统等问题，通过无线网络与目标识别相结合，极大地增加了系统的灵活性。

3.4 本章小结

本章通过分析传统目标识别系统现存的问题，对现有的传统目标识别系统提出了两种改进方案。第一种改进方案是将全部目标识别工作在服务器端完成，前端系统直接将原图像传送给服务器目标识别系统。这一方案受到了通信资源的限制，无线网络的带宽不足和稳定性不足的特点使得这一改进方案在实际情况中难以得到实施。而改进方案二将全部目标识别工作移植到嵌入式平台，嵌入式平台的软硬件环境不足以支撑整个目标识别系统正常运行。最终本文提出了动态级联目标识别系统框架，通过级联机制缓解了通信资源和计算资源的压力，并且根据实时的带宽状况动态调整需要传送的数据量，使得系统能够运行在不同的情况之下。

第4章 动态级联目标识别系统

上一章详细介绍了动态级联目标识别系统框架的设计过程。着重分析了传统目标识别系统的弊端，提出了两种改进方案并分析其可行性，最终提出了动态级联目标识别系统框架。本章是上一章提出的动态级联目标识别系统框架在无人机航拍检测场景内进行具体的实现。主要内容包括动态级联目标识别系统具体的软硬件平台介绍及目标识别系统中各个备选模型的优劣的分析。目标识别系统主要包括前端（无人机端）目标识别系统和服务器端目标识别系统。前端目标识别系统的识别结果通过无线网络发送给服务器端目标识别系统作为其输入，服务器端目标识别系统的识别结果作为整个系统的最终结果。

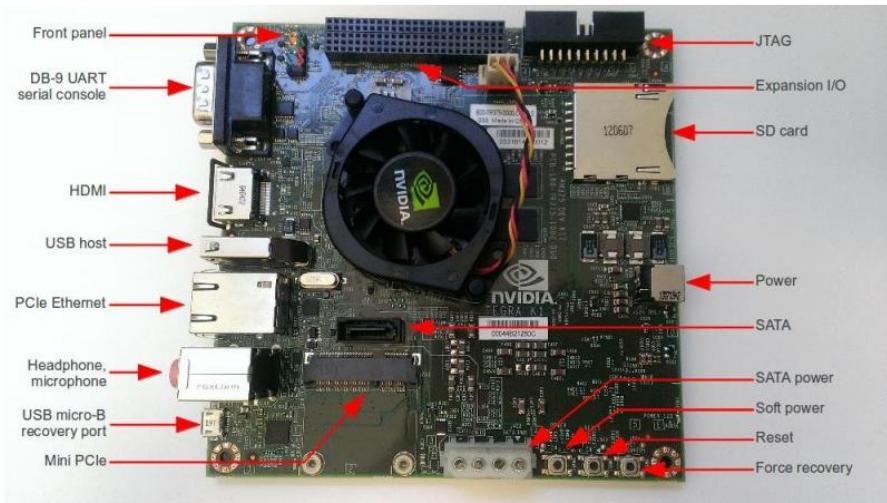
4.1 前端目标识别系统设计方案

前端目标识别系统具体包括系统硬件平台、内置目标识别算法，网络数据传输模块三个部分。

在整个目标识别系统中，我们传送的是图像而非视频流，虽然从理论上来说，传输视频流只传输每一帧图像中与前一帧图像不同的内容，比每秒传输 24 帧图像节省很多数据量。但是实际情况当中，在航拍场景下由于航拍角度的变化会导致视频流中各帧图像相差较大，导致压缩后数据量不会有明显的缩小。并且在航拍检测场景下我们并不需要每秒传送 24 帧图像那么多，只需要以合适的频率来向服务器端发送数据即可，从应用角度看，每秒传送一次数据就能够达到监测机场的需求。所以，本文采用图像而非视频流作为目标识别系统的输入数据。

4.1.1 前端系统硬件平台

在前端硬件平台搭建方面，本文选择在无人机上搭载一套具有较高处理能力的嵌入式板卡，并为其配备高清摄像头用于采集图像。在嵌入式板卡的选择方面，本文选择 NVIDIA 生产的 Jetson TK1^[4] 嵌入式开发板，如图 4-1 所示。

图 4-1 Jetson TK1 板卡实体图^[4]

Jetson TK1 是 NVIDIA 生产的嵌入式 Linux 开发平台，能够代表当前国内外嵌入式板卡的一流配置。该板卡轻巧便捷（127mm x 127mm），可以有效减少无人机的负载并且减少了由于载重对能源带来的消耗，间接延长了续航时间。Jetson TK1 一大特点在于集成 CPU+GPU+ISP 在一块芯片之内，GPU 采用 NVIDIA 的 Kepler GPU，CPU 采用 ARM quad-core Cortex-A15 CPU。Jetson TK1 拥有 2GB 内存，16G 的 eMMC 内嵌式存储器，4MB 的 ROM，拥有 USB2.0、USB3.0 接口，另外，Jetson TK1 需要 12v 电源供电。

Jetson TK1 最大的优势在于板卡内置 192 核 GPU 并且支持 CUDA，这样就可以并发处理多窗口多级算法，大幅度提升系统性能，缩短检测时间。同时 Jetson TK1 由于小巧轻便，能够较合适的将其放置在无人机上，由于电源占无人机载重很大比例，Jetson TK1 由于其低功耗的特点，很好的缓解了无人机的电源携带问题。

Jetson TK1 预装有 L4T(Linux4Tegra)^[47]操作系统，L4T 操作系统是 NVIDIA 专门为 Jetson TK1 配置的操作系统，该系统在很大程度上为开发人员屏蔽了底层硬件问题，使得开发人员可以像在正常 Linux 操作系统下开发程序一样方便。Jetson TK1 外接无线网卡，通过 Wi-Fi 从 Jetson TK1 把数据传送回服务器端。

4.1.2 目标识别算法

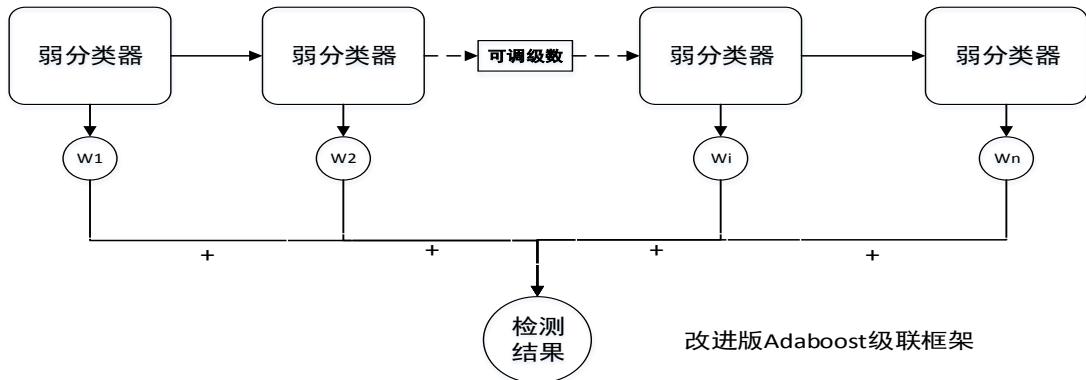


图 4-2 Adaboost 框架流程图

在本文中,选用 Adaboost 算法作为前端目标识别系统的模型。选用 Adaboost 算法主要考虑到以下两点:

1) 前端模型需要满足计算资源的约束。目前一些复杂模型如 CNN、DPM 等对硬件设备的内存、运算能力要求比较高,而在能够满足本文无人机航拍场景对算法性能要求的大前提下, Adaboost 算法本身能够很流畅地在嵌入式平台上运行。

2) 前端模型需要能够保证信息量不丢失的前提下动态调节需要传送的数据量。首先,在能够改变数据量的模型当中, Adaboost 和 DPM 等都具备这样的特点,而 CNN 则无法动态改变传输的数据量,CNN 网络的中间结果并不随着层数而变化,无法通过运行 CNN 不同数目的网络层数来实现数据量的缩小,因此,CNN 不满足本文对前端目标识别系统的要求。Adaboost 本身的级联结构使得其可以通过加载级数的多少来改变传输的数据量,在保证信息量不丢失的情况下,通过计算资源与通信资源的相互转换来减少系统对其各自的压力。而 DPM 则可以通过动态改变模型中的阈值来实现数据量的改变,经过实验统计,Adaboost 运行达到一定级数之后,此后的每一级的召回率都在 0.8 以上,因此不必担心因为改变级数造成信息量的丢失,而 DPM 则不满足系统对召回率的要求,虽然 DPM 模型整体召回率在 0.9 以上,但是,在通过改变阈值实现数据量改变的过程当中,没有办法保证系统的召回率,一旦信息量丢失,即使服务器端模型再精确,也没有办法找回丢失的信息量。如图 4-3 所示,红色线条表示目标区域得分,蓝色线条表示非目标区域得分。由图可见,蓝色线条与红色线条没有一个明显的分界线,如果通过分数的阈值来控制数据量的改变没有办法保证模型的

召回率，因此使用 DPM 作为前端目标识别系统的目地识别算法存在较大限制。

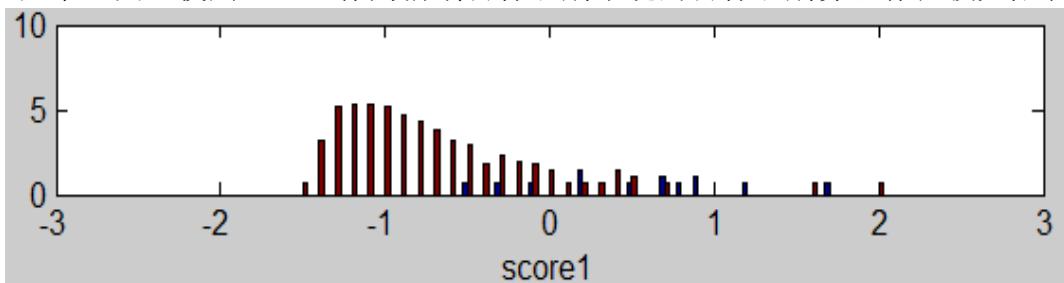


图 4-3 DPM 各窗口得分情况

4.1.3 网络数据传输模块

在网络传输方面，本文采用 Wi-Fi 作为前端系统和目标识别系统之间传输数据的方式，采用 Socket 套接字作为网络传输数据的工程上实现的方法。服务器端使用阻塞式套接字实时监控事先绑定好的端口，一旦有数据通过指定端口传送进来，会自动接收数据并且将接收到的数据发送给目标识别模块并在本地存储副本。图像检测完成后，套接字重新进入等待接收数据的模式。

在数据传输协议方面，有 TCP (Transmission Control Protocol 传输控制协议) 传输控制协议和 UDP (User Datagram Protocol) 传输控制协议可供选择，本文采用 TCP 数据报传输协议。TCP 是一种面向连接的、可靠的、基于字节流的传输层通信协议，TCP 在进行连接的时候需要接收端与发送端进行三次握手，在连接终止时需要四次握手，二者原理相似，下面着重介绍 TCP 连接时的三次握手协议，如图 4-4 所示。

TCP 三次握手的过程如下：

- 1) 客户端发送 SYN (SEQ=x) 报文给服务器端，进入 SYN_SEND 状态。
- 2) 服务器端收到 SYN 报文，回应一个 SYN (SEQ=y) ACK(ACK=x+1) 报文，进入 SYN_RECV 状态。
- 3) 客户端收到服务器端的 SYN 报文，回应一个 ACK(ACK=y+1) 报文，进入 Established 状态。三次握手完成，TCP 客户端和服务器端成功地建立连接，可以开始传输数据了。

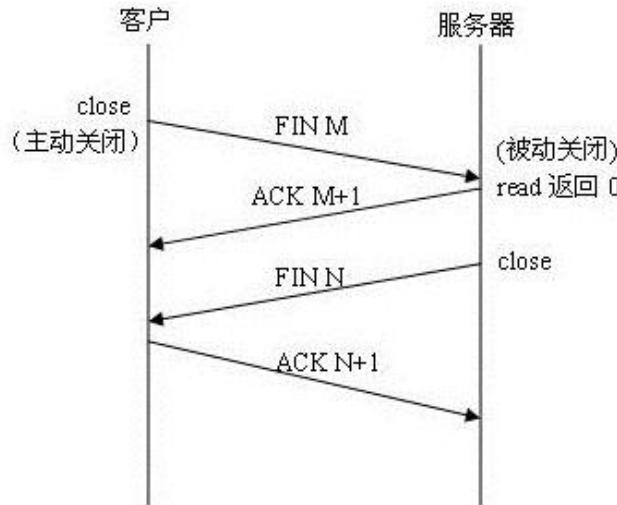


图 4-4 TCP 连接的三次握手协议

而 UDP 是一种不可靠的信息传送服务，是一种典型的客户/服务器的网络应用模式，UDP 报文没有可靠性保证、顺序保证和流量控制字段等，可靠性较差。根据本文动态级联目标识别系统框架的主要思想，首先要保证的是从前端目标识别系统到服务器端目标识别系统的数据传输过程不丢失信息量，因此，TCP 相对来说更适合作为本文级联目标识别系统的数据传输方式。

4.2 服务器端目标识别系统设计方案

在整个空地级联目标识别系统当中，服务器端是目标识别系统的第二部分，它在整个框架中起到的是最终确定是否是目标物体的作用。在实际情况中，由于它与前端目标识别系统在整个框架中扮演的角色不同，所以本文对它要求的侧重点也有所不同，总结来说有以下特点：

1) 约束较少

由于服务器端目标识别系统放置在实验室，计算资源充裕，没有软硬件方面的限制。在这样的前提之下，可以选择运算复杂、精确度高的模型来进行目标识别。服务器端的主要的约束在于受到帧的实时性约束。因此对算法的检测时间有一定的约束。

2) 算法精确

由于判定一个物体最终是否为目标是由服务器端目标识别系统来决定的，所以它对于算法精确性要求较高。目前准确率和召回率等性能指标比较好的模型有 DPM、CNN 等等，算法具体的选择问题还需要相互比较和实践。服务器端算

法并不是本文关注的重点，本文用 CNN 来作为服务器端目标识别算法，用来比较经过级联的系统与传统目标识别系统的区别。

4.3 模拟目标识别系统的工程实现

目前，模拟系统在工程上基本实现了整个系统所需的全部功能。其中包括：

1) 通过摄像头获取实时数据。

2) 解析保存模型的 xml (可扩展标记语言，Extensible Markup Language) 文件。

3) Adaboost 目标识别系统模型的加载。

4) Haar 特征提取。

5) Adaboost 目标识别系统检测过程。

6) 无线网络收发数据。

7) 实验结果的统计功能。

本文在 PC 机使用软件来模拟前端目标识别系统。服务器端目标识别系统目前暂时用 CNN 来做性能对比，用来说明新型框架的优势。在整个的代码系统中，工程上全部采用 C++ 程序设计语言来实现系统功能的各个模块。其中有部分依赖于 OpenCV 中有关图像处理的算法库。Adaboost 模型的训练模块，采用 OpenCV 中自带的 Adaboost 模型代码来训练。我们开发了前端 Adaboost 目标检测部分和服务器端的 Adaboost 目标检测部分，网络传输模块，并且添加了统计各级窗口数、计算量、计算资源与通信资源的转换情况等功能。为了更好的检验本框架在性能方面的合理性和有效性，在进行结果统计和性能分析时用 CNN 来作为对照。

图 4-5 表示新型动态级联系统的具体实现的流程框图，左半部分表示前端目标识别系统运行的进程，右半部分表示服务器端目标识别系统运行的进程。二者和网络传输模块共同构成了整个目标识别系统。首先，通过摄像头实时获取航拍数据作为前端系统的输入（目前模拟从视频流中每 5 秒抓取一幅图像），获取数据之后，前端系统从 xml 文件中加载 Adaboost 模型，通过带宽检测模块获取当前带宽状况，目前暂时用随机数来模拟带宽，更好的采用相关设备模拟带宽状况也是本文下一步的工作，获取到当前带宽情况后，根据带宽情况确定模型实际用于检测的级数，带宽状况良好，则少运行几级，带宽状况不好，则多运行几级。一般情况下会有一些级没有实际运行到，如图 4-5 左下部分。目标检测完毕后将检测结果存储成 jpg 图像，在前端系统和服务器端系统双方确认好可以收发数据

后，将 jpg 图像通过无线网络以二进制流的形式通过流套接字从前端系统发送给服务器端系统。服务器系统在加载模型完毕后，通过 Socket 套接字实时监控指定的端口，进程进入阻塞状态，服务器端系统通过 TCP 传输协议接收到数据后，立即将获取到的数据输入到服务器端目标识别系统当中，服务器端目标识别系统对接收到的数据做最后的目标识别处理工作，并得到最终的识别结果。识别完成后，立即进入阻塞状态，等待下一组数据的到来。

图 4-5 所示的系统架构可以非常容易的升级到实际部署系统。程序本身在开发过程当中已经设计好升级成实际系统需要开发的模块，比如实时读取数据模块等等。因此，整个系统只需稍加改动就可以移植到其他平台。这套模拟系统很容易就可以升级成实际系统。

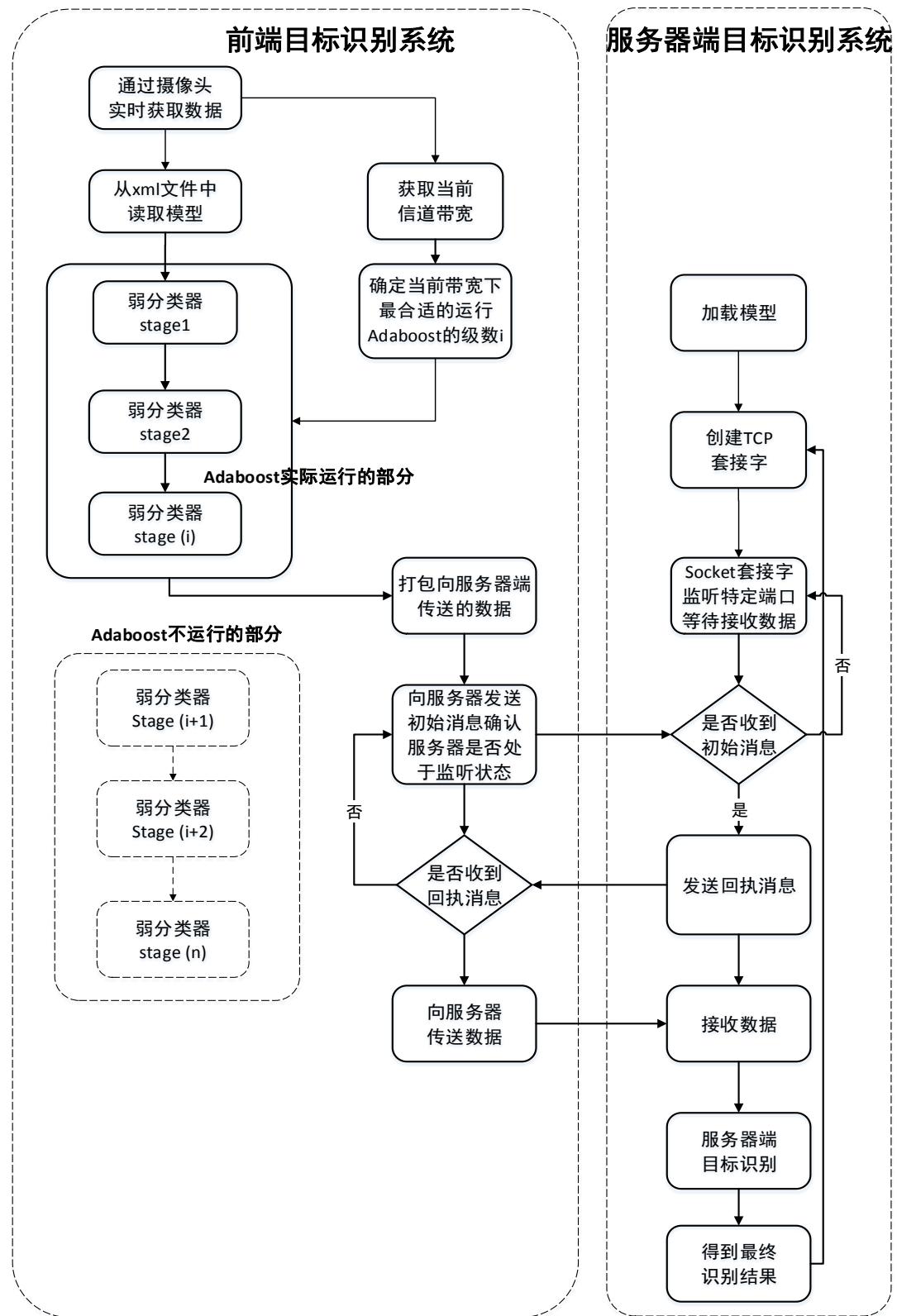


图 4-5 新型动态级联目标识别系统框架

4.4 模拟系统的实验结果

对整个系统影响比较大的主要有以下指标：

- 1) 前端系统模型各级的准确率和召回率；
- 2) 前端模型各级完成检测后节省带宽百分比；
- 3) 前端模型各级完成检测后需要的计算量；
- 4) 计算资源与带宽资源相互转换结果；
- 5) 可调的级数。

4.4.1 模型的准确率和召回率

本文采用的数据集是由实验室制作的高清航拍数据集，使用了数据集中的以飞机为样本的部分。数据集以飞机为样本的部分共有 1000 幅飞机图像，正样本 7482 个，负样本在以飞机场为背景（不包含飞机）的 1000 幅图像中随机抽取。在本文中所做实验当中，为了降低 Haar 特征不具有旋转不变性所带来的影响，我们将图像中的每个目标做了 90 度、180 度、270 度旋转，旋转过后，共有 29928 个正样本。

下面对比三种情况下系统的性能：

- 1) 前端目标识别系统；
- 2) 前端识别系统与服务器识别系统级联；
- 3) 服务器目标识别系统。

此次前端目标识别系统所用的模型是由 12000 个正样本，12000 个负样本，每级弱分类器最大个数为 30 个，训练得到一个 61 级的 Adaboost 模型，下面是经过统计的各级性能指标：

如图 4-6 所示，图中横轴表示召回率，纵轴表示准确率，图中显示了一个 61 级的模型里面从第 36 级到第 61 级中随机选取 5 级的性能。前端目标识别系统作为整个空地级联目标识别系统的第一部分，重点在于不丢失信息量的同时尽量减少需要传送的数据量，所以其召回率对于整个系统全局来说，相对于准确率更加重要。如图所示，各级召回率基本在 80% 以上，而各级准确率在 0.9~1，整体的 AP (Average Precision) 在 40%~50% 之间，级数越高，整体性能越好。由于 Haar 特征不具有旋转不变性以及 Adaboost 模型本身的性能不足以在航拍场景下达到非常高的性能，其检测结果的性能还不足以作为整个系统的最终输出。而对于整个空地级联目标识别系统而言，前端目标识别系统最大的意义在于在

保证信息量不丢失的情况下，最大限度地节省带宽，因此本文将召回率作为评测前端目标识别系统性能好坏的主要指标，从图中可以看出，系统各级的召回率基本在 80% 以上，基本能够达到要求。

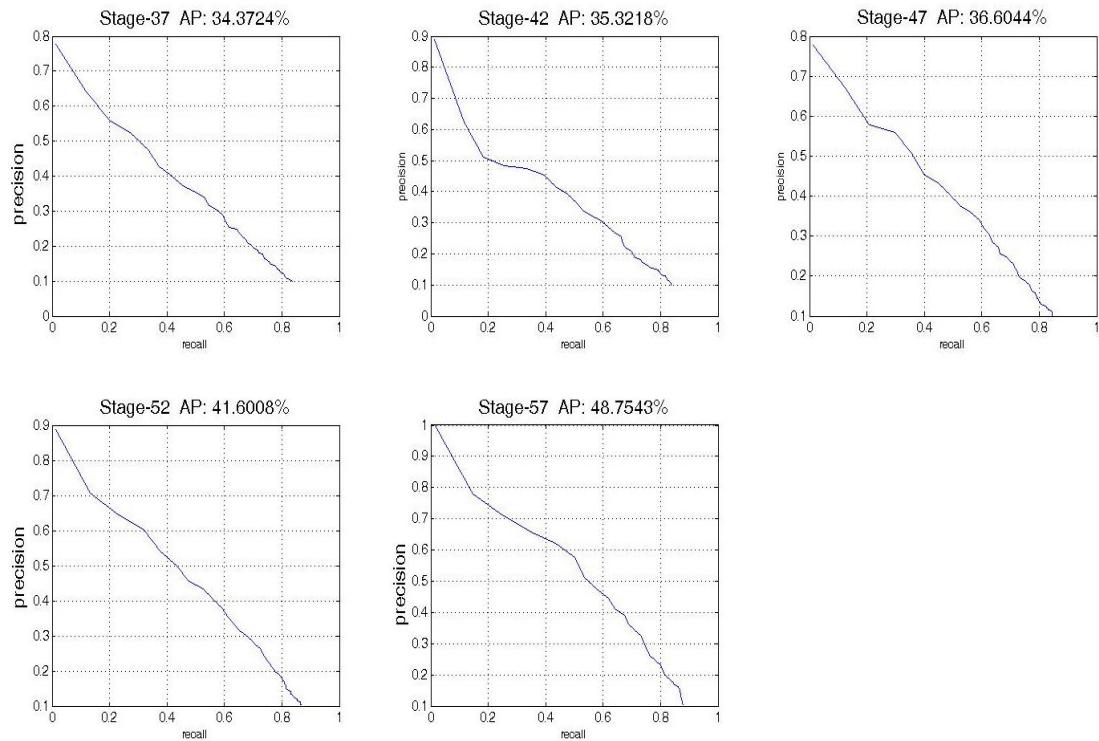


图 4-6 前端目标识别系统各级性能

同时，本文用前端系统对一百幅图像进行了目标检测，随着级数增加，有部分目标丢失，根据标定信息，100 幅图像中实际共包含 561 个样本，在第 46 级以前，前端系统保留了所有目标，在第 60 级时，在 561 个目标中丢失 6 个目标，召回率为 99% 左右，在带宽状况比较差的时候，会有部分目标丢失，而经过下文（图 4-7）统计分析，即使目前性能最好的模型，也无法保证能检测到丢失的 6 个样本。如图 4-7 所示，横轴表示级数，纵轴表示召回率。

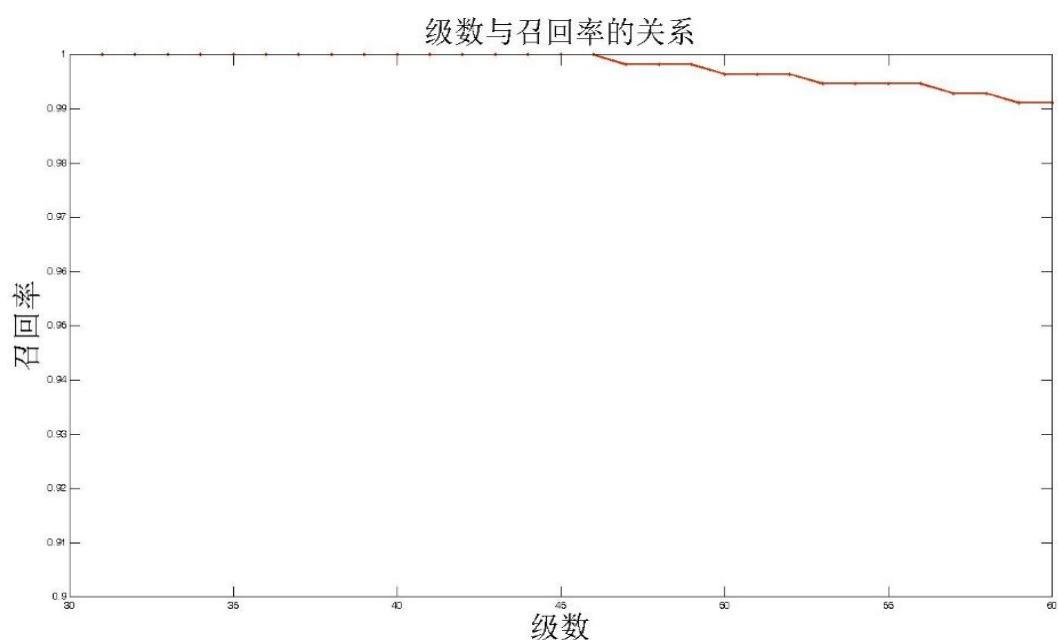


图 4-7 召回率与级数的关系

根据实验结果，丢失的目标主要是一些在图像边缘的目标、机身颜色与普通飞机不一致的目标或者是背景非常复杂时的目标，如图 4-8 所示。

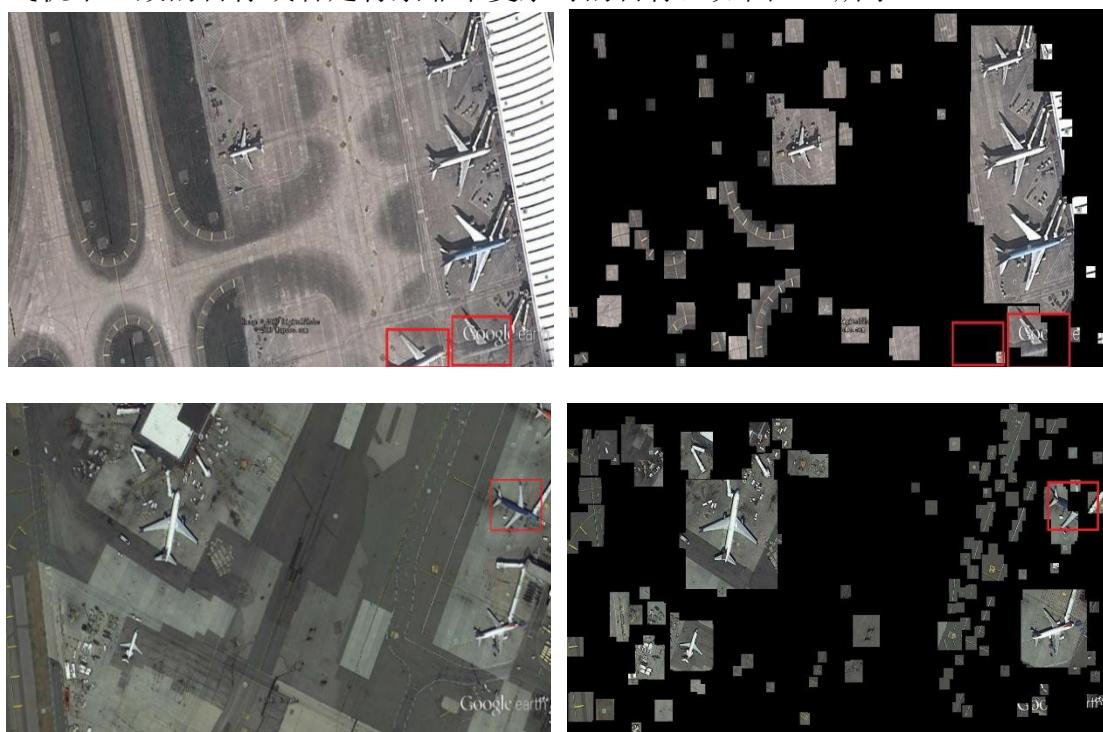


图 4-8 部分丢失的目标

对训练样本进行旋转后，前端系统对不同角度的目标检测具有一定的鲁棒性，同时，本文对训练背景的挑选非常多样化，能够将绝大部分复杂背景下的目标识别出来，如图 4-9 所示。

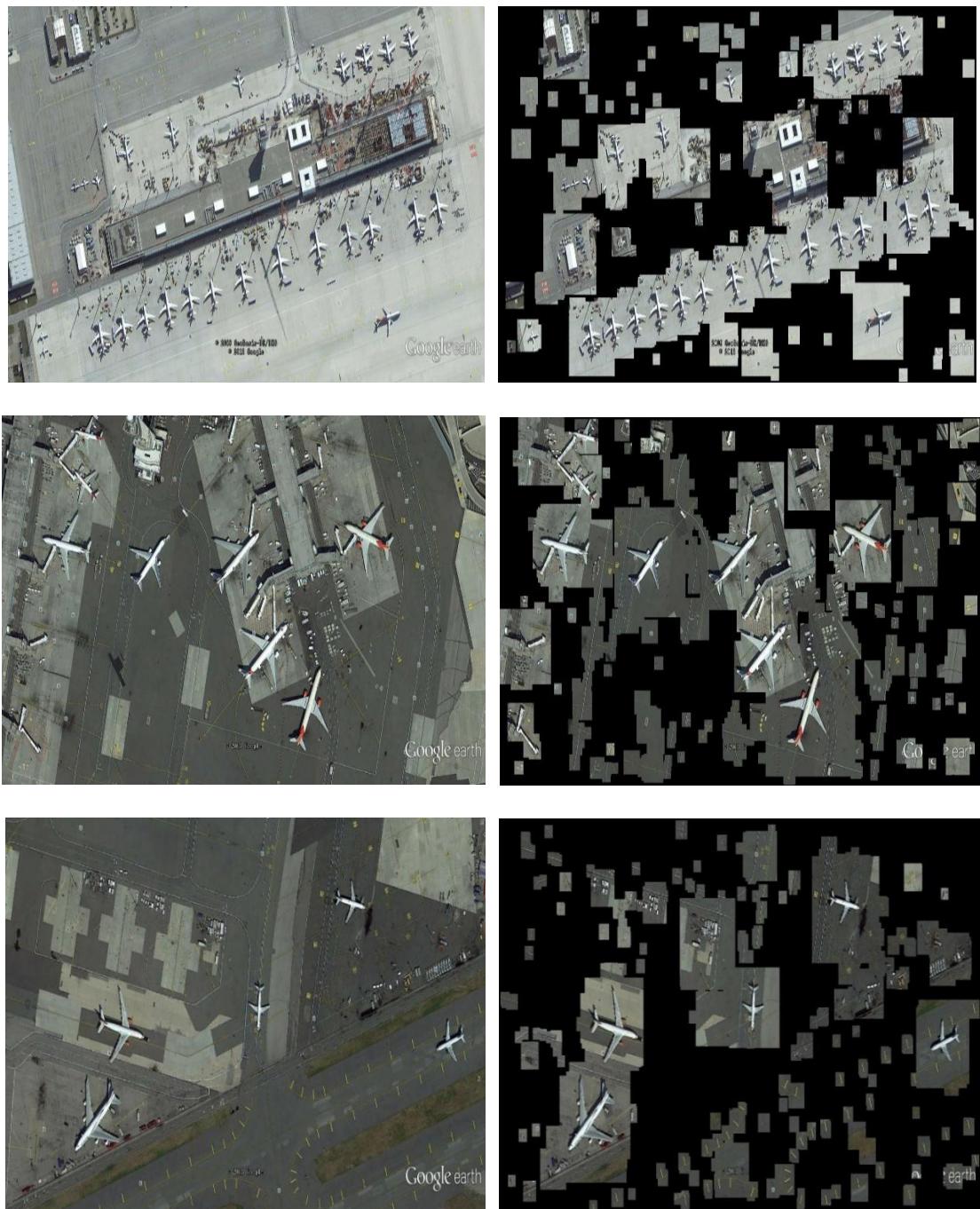


图 4-9 复杂场景和多角度下的识别情况

虽然前端目标识别系统识别出了绝大部分目标，保证了召回率。但是，前端

目标识别系统的检测结果存在很多误检，单纯将前端模型的结果作为最终的检测结果则不可取，在检测性能上不能满足要求。同时又由于前端计算资源的限制，复杂且性能较好的模型在前端系统难以运行。这也进一步说明了改进方案二的不可行性和系统级联的重要性。

图 4-10 表示前端识别系统与服务器端识别系统级联后得出的结果。将前端发送回来的图像通过 CNN 进行第二步目标检测，得到的 AP 在 68% 左右。这个结果说明级联系统要比全部目标识别系统都放置在前端要好，在有效节省带宽的情况下，提升了整个系统的性能，并且经过前端目标识别系统处理后，需要通过无线网络发送的数据量大大减小，极大地缓解了传输数据量大给带宽造成压力，节省的数据量会在后文进行比较。

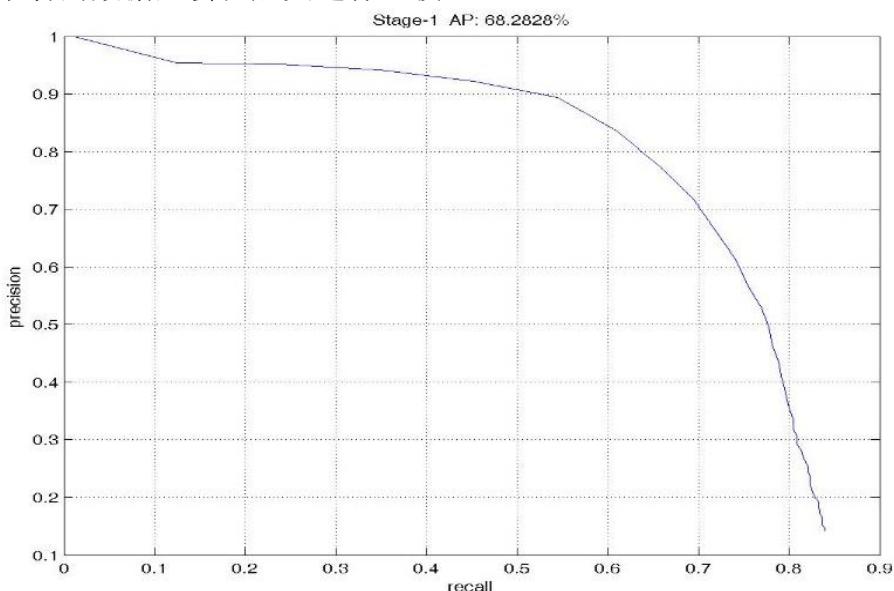


图 4-10 经过前端处理后在服务器端用 CNN 检测后的性能统计

图 4-11 表示未经过前端目标识别系统处理，直接检测原图得到的结果。系统的 AP 在 79% 左右。可见，经过级联后的识别结果要略差于直接输入原图得到的识别结果，但是性能上的差别不是很大，基本在本文的预期范围之内。造成性能下降的可能原因是图像经过前端系统处理后，变成有很多“黑块”的图像，与模型训练时的图像背景有所不同，模型挖掘的难例没有起到作用甚至起到了相反的作用，导致了模型的性能略有下降。对服务器端系统的检测造成了一定的影响，可行的解决办法之一是直接使用带黑色区域的图像进行训练，这也是下一步的工作内容之一。另一个可能的原因是前端目标识别检测系统检测结果丢失了部

分信息量。前文对前端系统的召回率做过统计（如图 4-10），统计召回率的方法是根据标定信息，对比在目标样本中是否存在“黑块”，如果目标区域中的“黑块”面积超过了 30%，则认为前端系统经过检测后丢失了这个目标。由前文统计得知，在 561 个目标样本中，丢失了 6 个目标样本。但是即使目前最为精确的目标识别算法召回率同样达不到百分之百，丢失的信息量控制在 5% 以内，基本符合了本文对前端系统的要求。

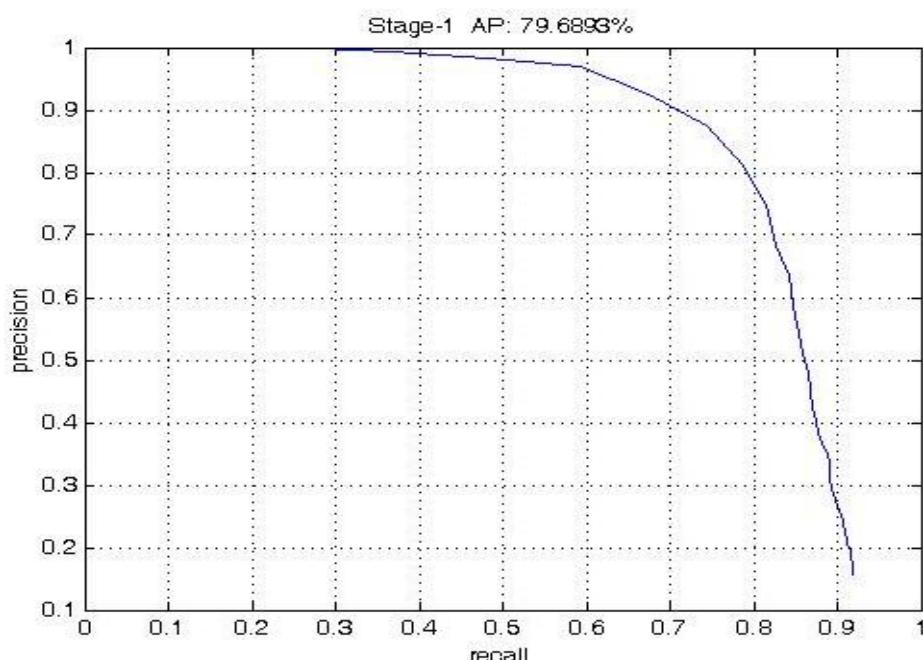


图 4-11 未经过前端处理直接在服务器上进行识别的结果

图 4-12 直观展示了前端目标识别系统的检测结果，由图可见，尽管由于 Haar 特征不具有旋转不变性和 Adaboost 模型自身性能的限制，导致了图中有很多虚框，然而图像中包含有目标的部分全部都检测到了，召回率基本达到系统的识别要求。将前端目标识别检测结果作为服务器端目标识别系统的输入，进行更进一步的精确目标检测。



图 4-12 前端目标识别系统检测结果演示

图 4-13 是经过级联后的服务器端检测结果的演示。服务器端的目标识别算法暂时采用 CNN 作为目标识别算法，是经过前端目标识别系统（Adaboost）检测过后通过无线网络发送回服务器端后经过服务器端目标识别系统(CNN)进行检测得到的结果。二者在整体框架中是一个级联结构，前端系统的输出经过网络传输后，作为服务器系统的输入，经过服务器目标识别系统的检测，得到最终结果。

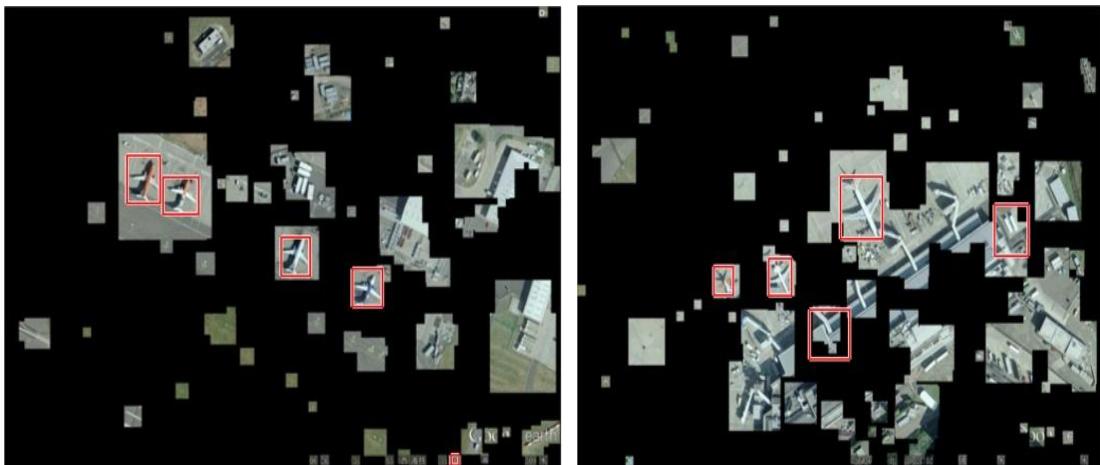


图 4-13 级联系统的检测结果演示

4.4.2 模型各级完成检测后节省的带宽百分比

下面来比较系统在上述性能时能节省的需要传送的数据量。图 4-14 为原高清图像与服务器端目标识别系统收到的不同带宽下的图像。图 4-14(a)表示通过摄像头获取的高清航拍原图像，图 4-14(b)到图 4-14(g)分别表示前端目标识别系统运行在 35 级，40 级，45 级，50 级，55 级，60 级后经过无线网络传输到达服

务器端目标识别系统的图像，即表示的是系统运行在不同的带宽状况时，服务器端目标识别系统收到的图像数据。图中黑色的部分表示经检测后确认不属于目标的部分，有图像的部分表示可能包含目标的区域，是需要通过网络由前端系统传送给服务器端的部分。由图可见，图像在经过前端目标识别系统后，不同的级数需要传输的数据量会有较大的差异，系统能够在省去 20%~80% 带宽的前提下，保证有信息量的部分不丢失。运行 35 级时，图像中只有很少一部分有“黑块”，需要传送原图像 85.4129% 的数据量。此后，每经过 5 级，需要传送的数据量会下降 10% 左右，运行的级数越多，减少的数据量越小，从 35 级到 40 级的过程中，系统需要传送的数据量缩小了 12.2085%，从 55 级到 60 级，系统需要传送的数据量缩小了 8.8096%。在图像通过 Adaboost 模型 60 级后得到识别结果时，需要通过无线网络传输的数据量仅占原图像的 27.7677%。可见，新型级联网络在不同带宽下体现出了非常强的可伸缩性。假设原图像大小为 1MB，在通信状况较好时，带宽充足（如 4Mbps）能够传送较多数据，假设可以省去 20% 需要传输的数据量，则需要 1.6s 可将数据传输完成。当带宽状况不好时，Adaboost 通过运行更多的级数来减少需要传送的数据量，最大可减少 80% 的数据量，假设数据传输速率为 1Mbps 时，同样只需 1.6s 就能够传输完全部数据，可见，新型系统在不同的带宽之下展现了良好的动态性和可伸缩性，在带宽不稳定的情况下，能够顺利将数据从前端传送回服务器端。



(a) 高清航拍原图像

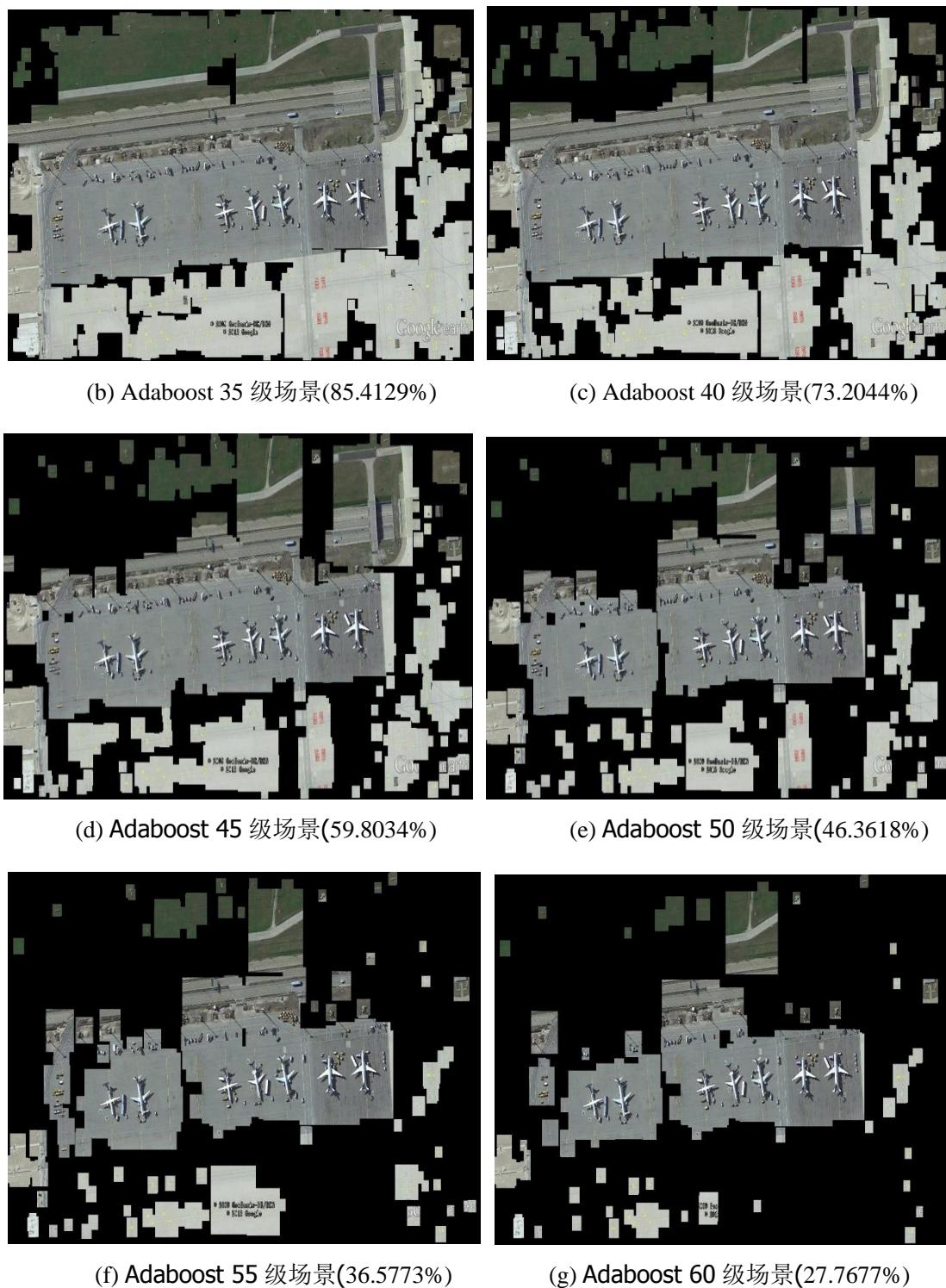


图 4-14 原高清图像与服务器端收到的不同带宽下的图像

图 4-15 表示模型级数与节省带宽百分比的关系。横轴表示级数，纵轴表示

节省带宽的百分比。图 4-15 通过在模型运行不同的级数之下，输入 100 幅图像，将各幅图像节省带宽百分比取均值后，绘制成带宽与级数的关系。由图可见，在只加载 36 级模型的时候，占用带宽量占直接发送原图的 68% 左右，而加载 61 级的时候，占用带宽量为发送原图的 36% 左右就可以达到直接传送原图的效果。随着加载模型级数越来越多，传输的数据量占原图的百分比逐步下降，基本上与级数呈现线性关系。在这里，我们之所以从 36 级开始计算是因为只加载前 36 级节省的信息量不多，对于整体系统的意义并不大。所以，前端系统可调的级数在 36 级~61 级之间。这一点在下面要介绍的窗口数统计中同样得到证明。运行的级数越多，系统节省的带宽量越大。

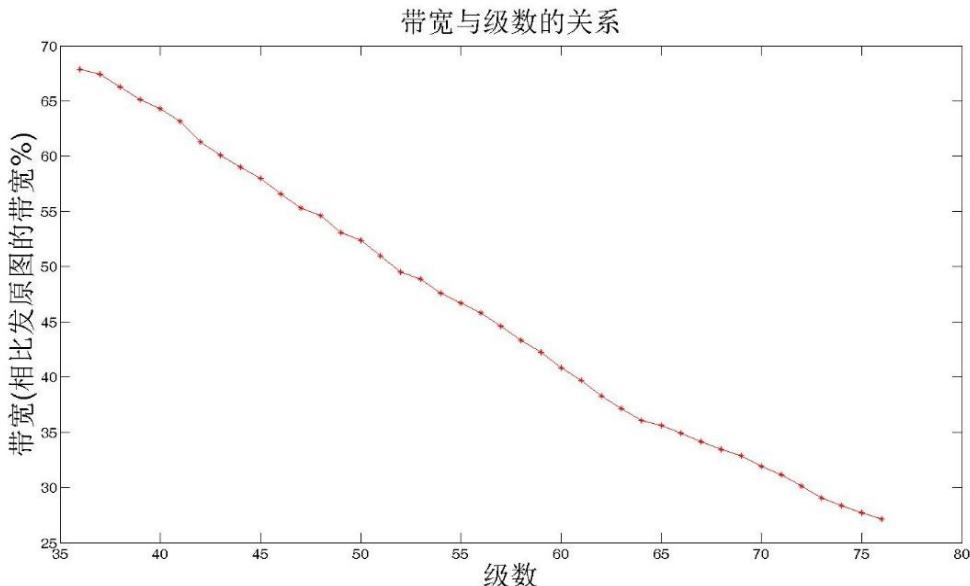


图 4-15 Adaboost 模型各级与带宽百分比之间的关系

4.4.3 前端模型各级完成检测后需要的计算量

在前端目标识别系统当中，由于嵌入式板卡的计算资源有限，计算量过大的模型无法在嵌入式板卡上运行。即使针对可以在嵌入式板卡上流畅运行的模型，统计其计算量也是非常有必要的。在本文的研究当中，通过增加 Adaboost 运行的级数来减少需要传输的数据量，其实质是通过增加计算量来减少需要传输的数据量。所以，掌握前端系统运行在不同的级数时计算量的具体情况是十分必要的。

由图 4-16 可见，随着级数的增加，计算量也在逐步上升，数量级基本都在 10^7 ，原因是最开始几级都是百万级的窗口，虽然逐级排除掉一些非目标的窗口，

但是仍有较大的计算量，从第36级到第61级，计算量相差大概一百万，所以，Adaboost 级联级数越少，检测速度越快。

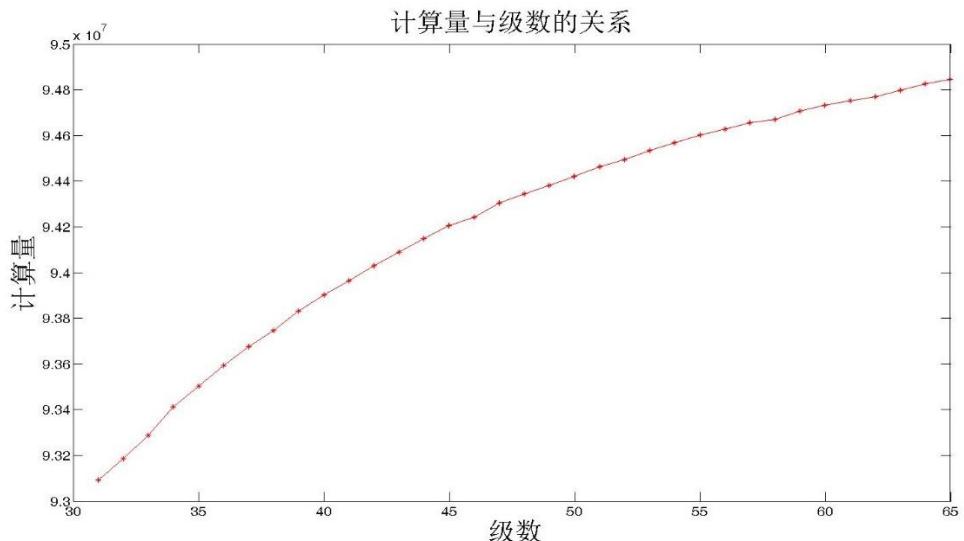


图 4-16 前端系统各级与计算量之间的关系

4.4.4 通信资源与计算资源相互转换情况

级联目标识别框架的实质其实就是计算资源与通信资源相互转换。当通信带宽不够充分时，通过增加目标识别精度来缓解对带宽的需求，而目标识别的精度往往与计算量成正比；当通信资源充分时，适当减小模型的精度，保证召回率，把一切有可能是目标的物体都传回服务器端。所以，掌握计算资源与通信资源相互转换的情况是至关重要的。

图 4-17 表示计算资源与通信资源相互转换的结果。横轴表示发送的数据量占直接发送原图的百分比，纵轴表示与发送数据相对应的计算量。由图可知，在传送数据量只有传送原图的 36% 时，系统的计算量在 9.47×10^7 左右，而在传送数据量为原图 68% 时，系统的计算量在 9.32×10^7 左右，在减小 32% 传输数据量的同时，增加了 100 万次左右的计算量，是通信资源到计算资源的转化。这个过程反过来，就是通过加大计算资源的利用来减少通信资源的压力。这一指标，将在下一步工作中，即嵌入式模拟平台上进行深入的研究。

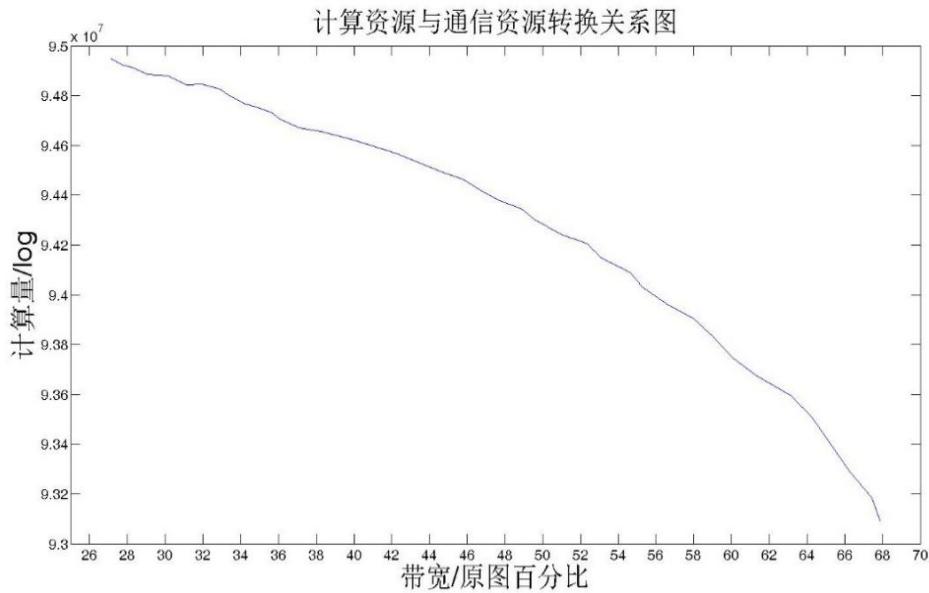


图 4-17 计算资源与通信资源相互转换结果展示

4.4.5 可调的级数

对于 Adaboost 而言，并非所有级都是可调的。掌握 Adaboost 中有多少级属于可调节区域对前端系统的可调节性而言非常重要。

图 4-18 表示经过检测后各级所剩余的窗口数对数与级数之间的关系。图中横坐标表示级数，纵坐标表示窗口数的对数。图中虚线表示一个阈值，这个阈值表示所要传输的数据量与传输原图数据量持平。所以，在蓝线下方的那些级属于可调的级数。实验中随机取 100 幅图像的平均值来测试结果，由图可见，在一个 61 级的模型中，从第 38 级开始，传输数据量要小于原图，可以调节的级数在总级数的 37% 左右，是一个比较大的可调节的范围，而通信资源与计算资源的转化过程，就是通过改变级数来实现的。在这个范围之内，可以通过实时监测带宽状况来决定模型实际运行的级数。

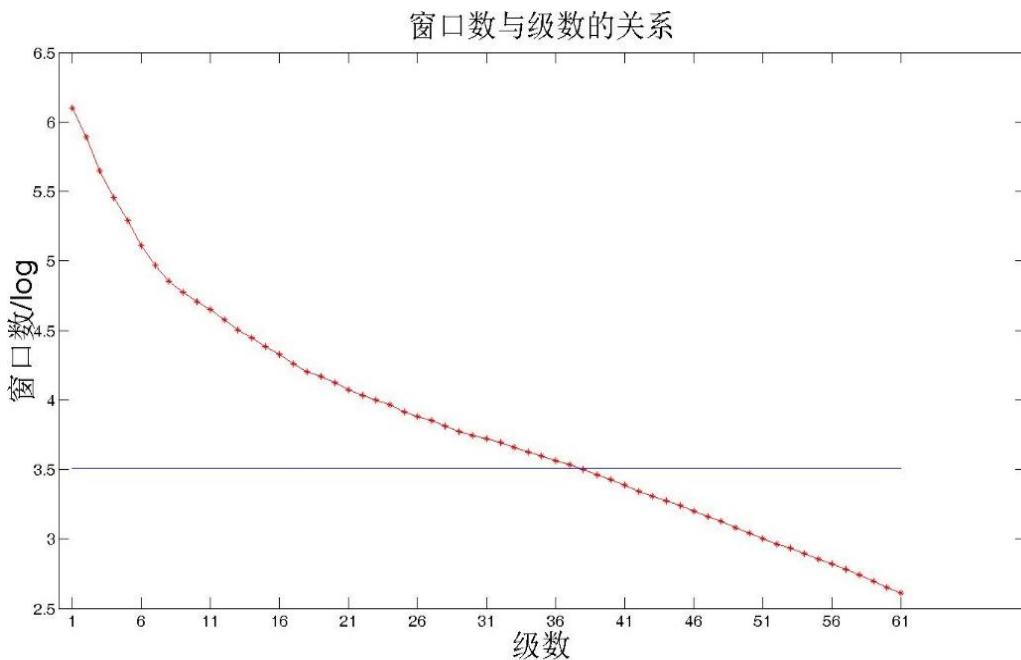


图 4-18 Adaboost 模型各级可调级数

4.5 模拟系统升级为实际系统

本文之前实现了模拟系统并分析模拟系统的检测结果，验证了框架之前的推论。现在，本文将前文介绍的模拟系统升级为实际系统，以 Jetson TK1 为前端系统硬件平台，通过实际的信号衰减器来模拟带宽的变化。下面分别介绍实际系统中的前端目标识别系统和服务器端目标识别系统。

4.5.1 前端目标识别系统

在实际目标识别系统中，将原本模拟系统中运行在 PC 机上的前端系统移植到 Jetson TK1 嵌入式平台上。由模拟目标识别系统升级成为实际目标识别系统的过程中，主要解决了以下几点问题：

1) 添加摄像头实时获取数据模块

在模拟系统当中，获取输入数据的来源是 SDL 高清航拍数据集，该数据集的飞机部分共有 1000 幅航拍图像，本文使用了 700 幅用来训练模型，余下 300 幅存放在指定路径，模拟目标识别系统从指定路径下读取数据进行目标识别工作。

在实际的目标识别系统中，需要设备从外接的摄像头获取的视频流中截取其中一帧图像作为输入数据。在本文的工程实现中，本文通过 Logitech 的高清网络

摄像头 C310 来获取数据，获取到视频图像的大小是 640*480 像素的，在工程实现方面，本文通过 OpenCV 的提供的方法从视频流中获取某一帧图像。

2) 升级网络模块

在模拟目标识别系统当中，通过 socket 套接字实现了数据的接收发送功能。由于模拟系统主要作用是分析结果，因此以有线传输的方式通过局域网来传输数据。具体做法是：接收端与发送端通过无线网卡连接到同一个路由器制成的 AP 当中，双方监听同一个端口，服务器不间断监听指定的端口，前端系统将数据发送到指定的 IP 和端口下即可。另外，socket 套接字在不同平台下使用的 API 有所不同，因此需要重新编写网络传输模块。本文实现了 Windows 平台和 Linux 平台两个版本。

在传输数据方面，输入图像经过前端目标识别系统的识别后得到一系列目标区域的坐标，然而检测结果得到的目标区域必然包含很多重复的区域，该方法通过判定两个窗口的重合区域占整个区域的百分比来判定窗口是否重叠，如果重叠，则将这两个窗口合并。这个方法最大的弊端在于消耗了过多的时间，当窗口数目很多时，合并时间甚至远远超过了目标检测运行的时间。因此，本文从工程角度对去重工作做了优化，本文不断将目标区域复制到一个与图像大小的模板当中，模板中被复制过的区域就是要传送的图像区域，将模板与原图像对应起来即可得到待传输的图像区域。在传送过程中，将检测得到的数据以二进制文件的方式发送出去即可。

3) 改进带宽实时获取模块

在新型动态级联目标识别系统框架中，需要根据当前带宽情况来确定待传输的数据量，即确定 Adaboost 实际运行的级数。因此，带宽获取模块是整个系统中非常重要的一个环节。在模拟系统当中，通过 srand() 和 rand() 函数来生成随机数，通过判断随机数在哪一个范围内来确定模型需要运行的级数。

在实际系统当中，本文通过调节信号衰减器来实现在不同程度上干扰无线信号，进而实现模拟无人机与基站之间距离的变化。在解决如何获取当前带宽情况的问题中，本文通过获取 BitRate 确定当前的带宽状况，如图 4-19 所示。Bit Rate 的单位是 Mb/s，换算成 MB/s 后，就可以确定当前无线网络传输数据的能力。

```
wlan0      IEEE 802.11bg  ESSID:"UAVproject"
          Mode:Managed  Frequency:2.437 GHz  Access Point: 00:12:17:32:01:4A
          Bit Rate=54 Mb/s  Tx-Power=20 dBm
          Retry short limit:7  RTS thr:off  Fragment thr:off
          Power Management:off
          Link Quality=70/70  Signal level=-26 dBm
          Rx invalid nwid:0  Rx invalid crypt:0  Rx invalid frag:0
          Tx excessive retries:5  Invalid misc:71  Missed beacon:0
```

图 4-19 BitRate 示意图

4) 程序对不同平台的兼容性问题

在第四章实现的模拟系统当中，模拟系统的前端系统是运行在 Windows 平台下，而实际系统是运行在 NVIDIA 官方给出的 Linux4Tegra 操作系统平台下。由于实际系统与模拟系统运行的平台不同导致了一些兼容性问题，如：网络传输模块、性能统计模块和一些涉及到操作系统 API 的模块。对于这些部分，本文对它们进行了重写。因此，本文根据不同平台提供了两套前端系统，包括：Windows 平台下的前端系统模型、嵌入式 Linux4Tegra 平台下的前端系统模型。

其次，在 Linux 平台以及 Linux4Tegra 平台上开发 C++ 程序主要涉及到了三款辅助工具软件：Cmake、pkg-config、Eclipse。Cmake 是一款跨平台编译工具，OpenCV 在 Linux 平台下的安装主要通过 Cmake 来完成，Cmake 避免了开发者编写 MakeFile，开发者只需要编写指定格式的 MakeList 即可，为开发者节省了大量的时间。pkg-config 是一款编译辅助工具，帮助开发者检查库的版本号以及头文件和宏定义的位置等等，有效防止了编写好的程序只能在某台机器上运行的窘境。Eclipse 是一个开放源代码的、基于 Java 的可扩展开发平台，就其本身而言，它只是一个框架和一组服务，用于通过插件组件构建开发环境。本文通过安装 CDT 插件使得 Eclipse 可以编译、调试、运行 C/C++ 开发的程序。

前端系统的硬件实物图如图 4-20 所示。Jetson TK1 外接无线网卡来发送数据。

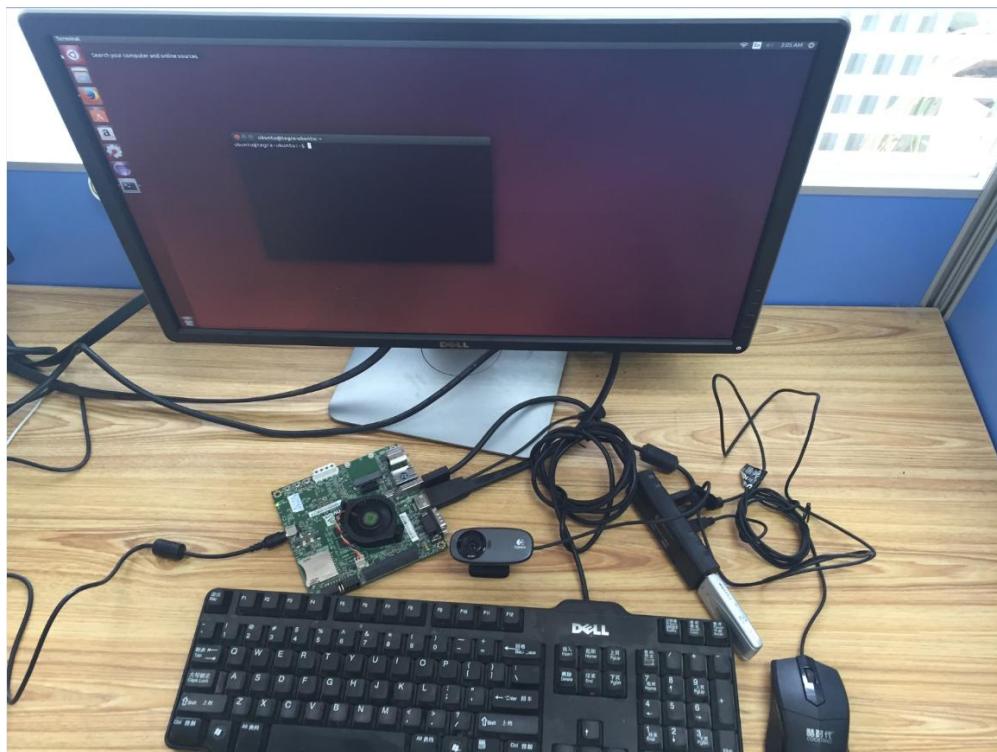


图 4- 20 实际系统的硬件实物图

在本文将模拟系统升级为实际系统后,本文用航模和机场图纸来模拟飞机场,通过摄像头在不同角度拍摄,作为实际目标识别系统的输入,如图 4-21 所示。



图 4- 21 航模与机场模拟真实机场场景

4.5.2 服务器端目标识别系统

在整个目标识别系统中，前端目标识别系统和服务器端目标识别系统在检测速度上的匹配对整个系统至关重要，在 Linux4Tegra 操作系统下，前端系统检测一幅 640*480 图像耗时 1s 左右，而在服务器端，传统的 CNN 检测一幅图像需要几十秒钟，CNN 加上 Selective Search 后检测时间还要增加十几秒，服务器端的检测速度与前端系统的检测速度不匹配。本文前端系统检测后提取到目标区域，通过各个目标区域的得分排序，将目标区域的坐标保存下来发送给服务器端，以此来代替 Selective Search 的结果，和传送的图像一起作为 CNN 模型的输入。因此，本文以前端系统得到的识别结果作为 CNN 中 Selective Search 的输入，很大程度上缩短了服务器端检测时间。但是，服务器端检测时间与前端仍然不能够匹配，所以，本文采用前文提到的 Fast CNN 作为服务器端的目标识别模型。Fast CNN 检测一幅图像需要 1s~2s，且模型准确率和召回率等性能指标也很好，基本符合系统对服务器端目标识别系统的要求。Fast CNN 留给开发人员的接口是由 Python 编写的。本文之前分别在 Windows 平台下和 Linux 平台下实现了 C++ 版本的服务器端目标识别系统，内置目标识别系统由 Adaboost 和 RCNN。在本章实际系统的搭建过程中，本文以 Python 为编程语言重新搭建服务器端目标识别系统。

4.6 实际目标识别系统的检测结果

将模拟目标识别系统升级为实际目标识别系统，4.4 节中部分模拟系统的结果需要我们重新考虑。其中，模型的性能、节省带宽的百分比以及计算资源与通信资源相互转化的关系相对来说比较固定。我们需要重新考虑系统消耗的带宽、计算量和可调的级数。

4.6.1 带宽与级数之间的关系

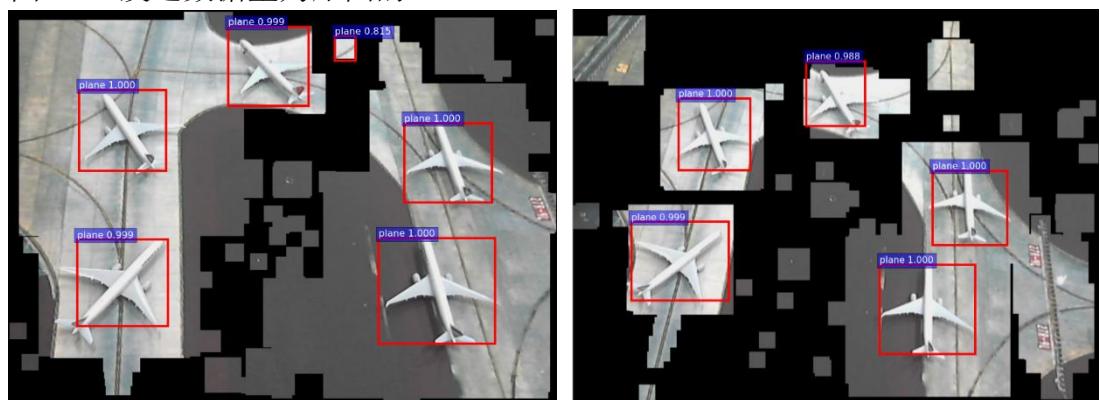
本文通过系统在一段时间内稳定的运行，分别在 30 级、36 级、42 级、48 级、54 级和 60 级之下获取 100 幅图像，对传送 100 幅图像占用的带宽求平均值，结果如表 4-1 所示。

表 4-1 带宽与级数之间的关系

级数	30	36	42	48	54	60
带宽 (%)	77.2293	63.8243	52.8939	40.1517	36.6042	34.3180

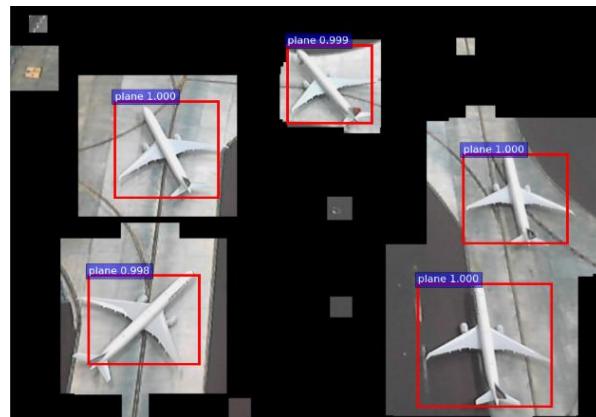
由表 4-1 可见，随着级数的增加，系统传输图像需要消耗的带宽越来越小。系统运行在 30 级时，传送图像的数据量占传送原图的 77.2293%，系统运行在 60 级时，传送图像的数据量占传送原图的 34.3180%。

图 4-22 (a) ~ (c) 分别表示前端系统运行在 40 级、50 级和 60 级的结果，图中非黑色区域为前端目标识别系统的识别结果，图中被方框框出来的部分为系统最终识别结果。由图可见，系统运行在不同级的时候，图像中的目标都被识别出来了，节省带宽的情况与级数相关，但也与图像中目标的数目相关。在图 4-22 中，图(a)发送数据量为原图的 73.3271%，图(b)发送数据量为原图的 49.8757%，图 (c) 发送数据量为原图的 40.4964%。



(a) 前端系统运行在 40 级的结果

(b) 前端系统运行在 50 级的结果



(c) 前端系统运行在 60 级的结果

图 4-22 动态级联目标识别结果演示

4.6.2 计算量与级数之间的关系

在模拟系统中，本文用到的 SDL 高清航拍检测数据集中图像的大小一般在

1280*659 像素大小。而实际系统中，通过摄像头获取到的图像数据大小为 640*480。原始数据大小缩小了近四分之三。因此，在计算资源受限问题上，实际系统需要的计算量远小于模拟系统。本文统计的计算量通过窗口数乘以当前窗口能够通过的级数来确定级数与计算量的关系。由于计算量主要消耗在前 20 级，表 4-2 列出了模拟系统与实际系统前 20 级的窗口数。通过对比数据可以看出，模拟系统各级窗口数与实际系统各级窗口数比例始终保持在 1:4 与 1:5 之间。因此，实际系统级数与计算量的关系基本与模拟系统相一致。

表 4-2 模拟系统与实际系统窗口数对比

级数	1	2	3	4	5
窗口数（模拟）	1281199	813591	491891	284719	191347
窗口数（实际）	407927	222330	134099	87513	55305
级数	6	7	8	9	10
窗口数（模拟）	108543	89729	77625	62598	53483
窗口数（实际）	35993	23600	16437	13992	11938
级数	11	12	13	14	15
窗口数（模拟）	47619	39823	33040	27705	23994
窗口数（实际）	10640	9207	7472	6620	6137
级数	16	17	18	19	20
窗口数（模拟）	21809	19641	17262	15231	14232
窗口数（实际）	5202	4382	3855	3544	3215

4.6.3 可调的级数

由于输入数据有所变化且在实际系统中容易受到光照等因素的影响，我们需要重新确认模型可调的级数。如图 4-23 所示，图中蓝色直线为发送原图的窗口数，红色曲线为各级运行之后剩余的窗口数。由图可见，可调的级数在 38 级到 61 级之间，基本与模拟系统测得可调的级数保持一致。

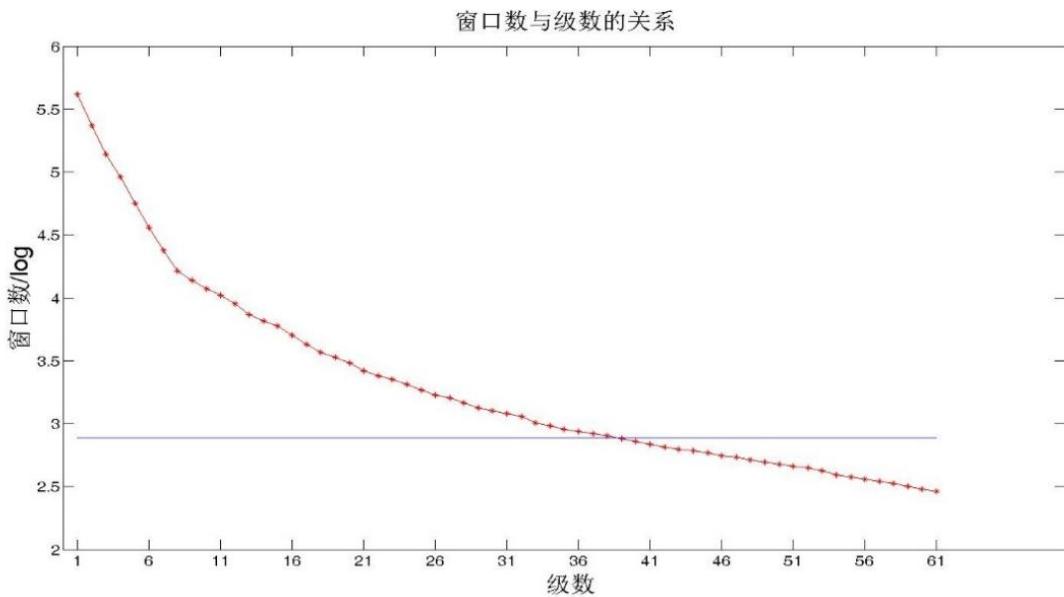
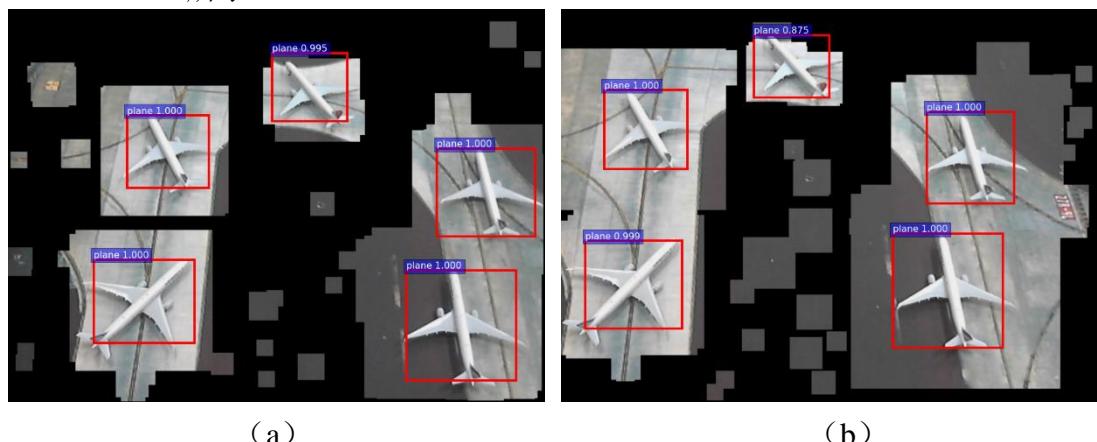


图 4-23 可调的级数

4.6.4 实际系统检测结果演示

本文通过沙盘和航模模拟真实的航拍检测场景，摄像头实时获取航拍数据，动态级联目标识别系统的演示结果如图 4-24 所示。系统在运行过程中，对角度的旋转，距离的远近，以及轻微的模糊等具有一定的鲁棒性，以 4-24 (a) 为基准，分别对摄像头做角度的调整、与目标距离的调整以及模糊情况，如图 4-24 (b) ~ (d) 所示。



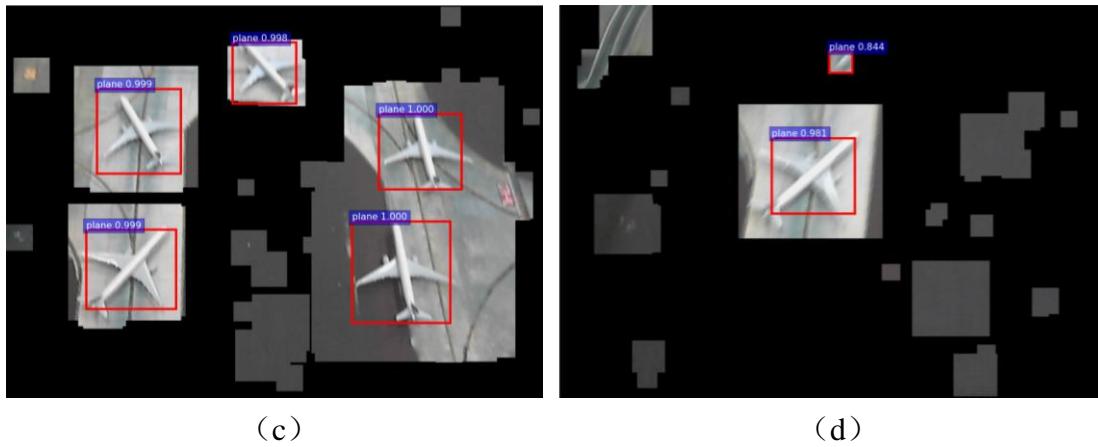


图 4-24 角度、尺度变化以及模糊情况对目标识别系统的影响

4.7 本章小结

本章以无人机航拍检测为实验背景, 分别实现了第三章提出的动态级联目标识别系统框架模拟系统和实际系统。分析了前端目标识别系统的性能和服务器端目标识别系统的性能, 并且还分别分析了前端系统各级节省的带宽情况、各级需要的计算量、计算资源与通信资源的相互转化的情况、可调的级数。经过本章的分析, 第三章提出的动态级联目标识别系统框架得到验证。最后, 本章将模拟目标识别系统升级到实际目标识别系统, 列举了由模拟系统升级为实际系统需要完成的工作, 展示了实际动态级联目标识别系统的结果, 分析了实际系统的性能。

第5章 结论与展望

传统的目标识别系统在部分领域应用比较成熟，如：车牌识别、人脸识别等等。但是在航拍检测、应急搜救等灵活性比较高的领域还没有得到广泛应用。在这一类场景之下，目标识别技术需要同其他技术相互协作，建立一个相对复杂的系统才能完成目标识别工作。本文通过无线网络与目标识别技术相结合，同时将目标识别系统移植到嵌入式平台，增加了系统的灵活性，使得目标识别系统在这一类场景下具备了实现的可能性。但是随着无线网络和嵌入式平台的引入，出现了通信资源与计算资源的受限的问题。本文主要就带宽资源受限情况下的目标识别系统给出了解决方案。

本文首先论述了课题的研究背景及意义、国内外研究现状和目前已存在的研究方法，介绍了本文的研究内容；在第二章里详细介绍了本文研究工作中使用的相关算法和技术，为第四章中动态级联目标识别系统的实现进行了铺垫；第三章介绍了动态级联目标识别系统框架的设计，以目前目标识别系统在灵活性较高场景下的应用现状为出发点，通过不断提出改进方案并分析改进方案的可能性，最终提出了动态级联目标识别系统框架。第四章以无人机航拍检测为实验背景，将抽象的框架具体化，实现了模拟动态级联系统和实际动态级联系统。在上述的研究工作中，总结本文的主要贡献如下：

1) 本文针对目前传统目标识别系统面临的问题，提出动态级联目标识别系统框架，并且对在传统目标识别系统基础上提出的两种改进方案进行了分析。论证了在计算资源和带宽资源受限的情况下，全部目标识别系统设置在嵌入式端或者服务器端并不可行。

2) 将抽象框架具体实现，设计一套可用于航拍检测的动态级联目标识别系统。在 PC 机上实现了模拟的动态级联目标识别系统，搭建系统的软硬件平台，并且对 DPM、CNN 和 Adaboost 三种算法进行了详细的分析和比较并将 Adaboost 作为前端目标识别系统的内置算法。在系统性能、级数与带宽关系、级数与计算量的关系、计算资源与通信资源相互转化的关系以及可调的级数五个方面来验证框架的可行性和有效性。

3) 将模拟的目标识别系统移植到嵌入式平台，实现了实际的动态级联目标识别系统。以航模和机场图纸搭建模拟的飞机场，通过高清摄像头实时获取数据，

将前端模型移植到嵌入式平台，并且通过实时获取带宽情况动态决定 Adaboost 运行的级数，使得整个系统运行在接近真实无人机航拍检测的场景之下。

虽然本文提出了动态级联目标识别系统框架，并在模拟航拍检测场景下实现了目标识别系统，但是因为时间关系，系统仍然存在以下不足：

系统的性能还可以更进一步的提高。将嵌入式平台搭建在无人机上可能还会遇到一潜在问题需要根据实际情况改进。

1) 将模拟系统升级为实际系统后，由于获取数据的来源变为通过摄像头实时获取数据，训练模型的数据与进行检测的数据来源不一致以及受到光照等因素的影响，实际系统的检测效果比模拟系统有所下降。

2) Adaboost 算法在航拍检测场景下的性能有待提高，虽然系统内置了通过目标区域得分二次筛选待传送区域的模块，但是由于 Adaboost 算法的性能不足，难以得到应用。

3) 前端目标识别系统在工程上用 C++ 实现，而服务器端目标识别系统在工程上用 Python 实现，前端与服务器端相互配合的稳定性还稍有不足。

参 考 文 献

- [1] 大疆创新科技有限公司 [J]. <http://www.dji.com/cn/product/inspire-1-pro-and-raw>.
- [2] 环球军事--全球鹰 [J]. http://weapon.huanqiu.com/rq_4.
- [3] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network; proceedings of the Image Processing (ICIP), 2015 IEEE International Conference on, F, 2015 [C]. IEEE.
- [4] Jetson Tk1 官方资料 [J]. http://elinux.org/Jetson_TK1.
- [5] ROWE A, ROSENBERG C, NOURBAKHSH I. A Second Generation Low Cost Embedded Color Vision System; proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition-workshops-volume, F, 2005 [C].
- [6] ROWE A G, GOODE A, GOEL D, et al. CMUcam3: an open programmable embedded vision sensor [J]. 2007,
- [7] RYMER J, RENNO J, GREENHILL D, et al. Adaptive eigen-backgrounds for object detection; proceedings of the Image Processing, 2004 ICIP'04 2004 International Conference on, F, 2004 [C]. IEEE.
- [8] ROWE A, GOEL D, RAJKUMAR R. Firefly mosaic: A vision-enabled wireless sensor networking system; proceedings of the Real-time systems symposium, 2007 RTSS 2007 28th IEEE international, F, 2007 [C]. IEEE.
- [9] KANDHALU A, ROWE A, RAJKUMAR R, et al. Real-time video surveillance over IEEE 802.11 mesh networks; proceedings of the Real-Time and Embedded Technology and Applications Symposium, 2009 RTAS 2009 15th IEEE, F, 2009 [C]. IEEE.
- [10] RAHIMI M, BAER R, IROEZI O I, et al. Cyclops: in situ image sensing and interpretation in wireless sensor networks [J]. Sensys, 2005, 192-204.
- [11] ATAN O, ANDREOPoulos Y, TEKIN C, et al. Bandit Framework for Systematic Learning in Wireless Video-Based Face Recognition [J]. IEEE Journal of Selected Topics in Signal Processing, 2014, 9(1): 704-8.
- [12] UIJLINGS J R, VAN DE SANDE K E, GEVERS T, et al. Selective search for object recognition [J]. International journal of computer vision, 2013, 104(2): 154-71.
- [13] FELZENSZWALB P F, HUTTENLOCHER D P. Efficient graph-based image segmentation [J]. International Journal of Computer Vision, 2004, 59(2): 167-81.
- [14] !!! INVALID CITATION !!!,
- [15] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection; proceedings of the Computer Vision and Pattern Recognition, 2005 CVPR 2005 IEEE Computer Society Conference on, F, 2005 [C]. IEEE.
- [16] DALAL N, TRIGGS B, SCHMID C. Human detection using oriented histograms of flow and appearance [M]. Computer Vision–ECCV 2006. Springer. 2006: 428-41.
- [17] WANG X, HAN T X, YAN S. An HOG-LBP human detector with partial occlusion handling;

- proceedings of the Computer Vision, 2009 IEEE 12th International Conference on, F, 2009 [C]. IEEE.
- [18] PAPAGEORGIOU C, POGGIO T. A trainable system for object detection [J]. International Journal of Computer Vision, 2000, 38(1): 15-33.
- [19] LOWE D G. Object recognition from local scale-invariant features; proceedings of the Computer vision, 1999 The proceedings of the seventh IEEE international conference on, F, 1999 [C]. Ieee.
- [20] REED S, SOHN K, ZHANG Y, et al. Learning to disentangle factors of variation with manifold interaction; proceedings of the Proceedings of the 31st International Conference on Machine Learning (ICML-14), F, 2014 [C].
- [21] KAVUKCUOGLU K, SERMANET P, BOUREAU Y-L, et al. Learning convolutional feature hierarchies for visual recognition; proceedings of the Advances in neural information processing systems, F, 2010 [C].
- [22] TUZEL O, PORIKLI F, MEER P. Pedestrian detection via classification on riemannian manifolds [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2008, 30(10): 1713-27.
- [23] BURGES C J. A tutorial on support vector machines for pattern recognition [J]. Data mining and knowledge discovery, 1998, 2(2): 121-67.
- [24] CHANG C-C, LIN C-J. LIBSVM: a library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.
- [25] 张学工. 关于统计学习理论与支持向量机 [J]. 自动化学报, 2000, 26(1): 32-42.
- [26] YAN J, LEI Z, WEN L, et al. The fastest deformable part model for object detection; proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, F, 2014 [C].
- [27] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model; proceedings of the Computer Vision and Pattern Recognition, 2008 CVPR 2008 IEEE Conference on, F, 2008 [C]. IEEE.
- [28] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D. Cascade object detection with deformable part models; proceedings of the Computer vision and pattern recognition (CVPR), 2010 IEEE conference on, F, 2010 [C]. IEEE.
- [29] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of computer and system sciences, 1997, 55(1): 119-39.
- [30] SCHAPIRE R E. The strength of weak learnability [J]. Machine learning, 1990, 5(2): 197-227.
- [31] DIETTERICH T G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization [J]. Machine learning, 2000, 40(2): 139-57.
- [32] HUNT E B, MARIN J, STONE P J. Experiments in induction [J]. 1966,
- [33] QUINLAN J R. Induction of decision trees [J]. Machine learning, 1986, 1(1): 81-106.
- [34] QUINLAN J R. C4. 5: programs for machine learning [M]. Elsevier, 2014.

参考文献

- [35] BREIMAN L, FRIEDMAN J, STONE C J, et al. Classification and regression trees [M]. CRC press, 1984.
- [36] KHAMMARI A, NASHASHIBI F, ABRAMSON Y, et al. Vehicle detection combining gradient analysis and AdaBoost classification; proceedings of the Intelligent Transportation Systems, 2005 Proceedings 2005 IEEE, F, 2005 [C]. IEEE.
- [37] FREUND Y, SCHAPIRE R, ABE N. A short introduction to boosting [J]. Journal-Japanese Society For Artificial Intelligence, 1999, 14(771-780): 1612.
- [38] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks; proceedings of the Advances in neural information processing systems, F, 2012 [C].
- [39] DONAHUE J, JIA Y, VINYALS O, et al. Decaf: A deep convolutional activation feature for generic visual recognition [J]. arXiv preprint arXiv:13101531, 2013,
- [40] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2014 [C].
- [41] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [M]. Computer vision–ECCV 2014. Springer. 2014: 818-33.
- [42] GIRSHICK R. Fast r-cnn; proceedings of the Proceedings of the IEEE International Conference on Computer Vision, F, 2015 [C].
- [43] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks; proceedings of the Advances in Neural Information Processing Systems, F, 2015 [C].
- [44] CHEN M, MAO S, LIU Y. Big data: A survey [J]. Mobile Networks and Applications, 2014, 19(2): 171-209.
- [45] FONSECA R, GNAWALI O, JAMIESON K, et al. Four-Bit Wireless Link Estimation; proceedings of the HotNets, F, 2007 [C].
- [46] QIN F, DAI X, MITCHELL J E. Effective-SNR estimation for wireless sensor network using Kalman filter [J]. Ad Hoc Networks, 2013, 11(3): 944-58.
- [47] LINUX4TEGRA. http://elinux.org/Tegra/Downstream_SW/Linux4Tegra [J]. .

致 谢

在中国科学院大学攻读硕士学位的学习生活中，我经历了诸多坎坷，也得到了非常大的锻炼和收获。在毕业论文即将完成之际，由衷的感谢那些给予我无私帮助的老师、同学、朋友和家人。

本课题的研究工作是在焦建彬教授、秦飞副教授、叶齐祥教授、韩振军副教授的信息指导下完成的。感谢您在我硕士三年的科研学习中孜孜不倦的指导，感谢您在我三年生活中对我的无私帮助和大力支持，感谢您在我遇到挫折时给予的关怀和鼓励。

感谢秦飞老师在我攻读硕士学位期间的辛勤指导和无私帮助。您对我工作不懈的指导让我受益甚多，您对待工作认真和严谨的态度是我今后学习和工作的榜样，您在工作中处理问题的方法将使我在今后的生活中不断受益。

感谢叶齐祥老师在我攻读硕士学位期间对我的科研提供的许多中肯的建议。您的建议让我在科研中不断得到提升，您耐心的指导让我在科研中开阔了眼界、丰富了经验。

感谢韩振军老师在我攻读硕士学位期间对我无私的帮助，您对我的科研、工作、生活的大力支持，让我顺利完成了学业。同时感谢您对我毕业论文给出非常宝贵的意见。

我还要由衷的感谢模式识别与智能系统开发实验室的同学们在学习生活中给我的无私帮助。是你们的陪伴，让我在欢乐中度过了研究生三年的科研生活。是你们的陪伴，让我更加快速的成长。

最后，感谢我的父母、我的亲人、还有那些默默支持我的挚友。感谢你们的支持和帮助，我会在今后的生活中继续努力，拼搏！

王忻雷

2016年4月

个人简介

个人简介

姓名：王忻雷 性别：男 出生日期：1989.08.02 政治面貌：共青团员

教育经历

- 2008年9月至2012年7月 合肥工业大学 电子信息工程 学士
- 2013年9月至2016年7月 中国科学院大学 计算机技术 硕士

