



南京大學

本科毕业论文

院 系 计算机科学与技术系

专 业 计算机科学与技术

题 目 基于局部特征和上下文的多类物体检测研究

年 级 2007级 学号 071221101

学生姓名 王利民

指导老师 路通 职称 副教授

论文提交日期 2011年5月20日

学 号 : 071221101

论文答辩日期 : 2011 年 5 月 30 日

指导教师 : (签字)

南京大学本科生毕业论文中文摘要

毕业论文题目： 基于局部特征和上下文的多类物体检测研究
院系： 计算机科学与技术系
计算机科学与技术 专业 2007 级本科生生姓名： 王利民
指导教师（姓名、职称）： 路通 副教授

摘要

物体识别是计算机视觉领域热点问题之一，在现实中存在着极为广泛的应用前景。物体识别面临着诸多挑战和难点，典型问题包括物体所在场景一般较为复杂、物体类别较多，且易受尺度、视角等变化的影响。本文提出了一个新颖的、基于局部特征提取和上下文环境建模的多类物体检测算法框架，该框架能够同时检测出感兴趣的多类物体。首先，在单类霍夫森林模型基础上，本文提出了新的多类霍夫森林学习模型，并使用图像的局部特征，对物体可能存在的位置等进行概率投票。考虑到自然场景中多类物体之间的相对位置关系和几何约束，本文进一步提出了上下文模型，对霍夫森林多类检测结果加入相对位置关系约束，以进一步提高物体检测正确率。本文在两个基准数据集上进行了实验，实验结果表明本文算法可有效从复杂场景图像中检测多类物体，算法鲁棒性较好。

关键词： 物体识别；物体检测；多类霍夫森林；上下文模型

南京大学本科生毕业论文英文摘要

THESIS: Multiclass Object Detection by Combining Local Appearances
and Context

DEPARTMENT: Computer Science and Technology

SPECIALIZATION: Computer Science and Technology

UNDERGRADUATE: Limin Wang

MENTOR: Dr. Tong Lu

Abstract

Object recognition is one of the hot research topics in computer vision. It has a variety of applications in the real world, e.g. video surveillance, as well as some big challenges, e.g. scale and viewpoint variance. Object detection, one task of object recognition, has attracted a number of researchers in the past decades. Even though it has been mature enough for some usual object class such as pedestrian, car and so on, it still faces many problems for common object classes. In this paper, we propose a novel framework for multiclass object detection by combining local appearances and context. Firstly, based on the single class Hough forest, we construct our multiclass Hough forest and vote for possible object position using local patches. Then, considering the relative location constraints among objects, we present a context model. With this model, we can successfully avoid some wrong detection of multiclass Hough forest and improve the detection accuracy greatly. We conduct experiments on two data set and the results show that our method achieves state-of-the-art performance for multiclass object detection.

Keywords: Object Recognition; Object Detection; Multiclass Hough Forest; Context Model

目 录

目录	iii
第一章 绪论	1
1.1 物体识别概述	1
1.2 物体检测算法概述	3
1.2.1 滑动窗口方法	3
1.2.2 霍夫投票方法	4
1.2.3 多类物体检测	7
1.3 上下文模型概述	9
1.3.1 上下文关系分类	9
1.3.2 上下文关系建模	11
1.4 本文主要研究工作和组织结构	14
第二章 基于局部特征和上下文的多类物体检测算法	15
2.1 引言	15
2.2 霍夫森林模型	15
2.2.1 随机森林简介	16
2.2.2 单类霍夫森林	17
2.2.3 多类霍夫森林	20
2.3 上下文关系模型	22
2.3.1 位置关系建模	23
2.3.2 模型学习过程	24
2.3.3 模型预测过程	25
2.4 本章小结	27

第三章 实验结果与讨论	28
3.1 引言	28
3.2 9类数据库实验结果及分析	28
3.3 LabelMe 数据库实验结果及分析	31
3.4 本章小结	32
第四章 总结与展望	34
4.1 本文总结	34
4.2 将来工作	35
参考文献	37
科研成果与个人荣誉	42
致谢	43

表 格

2.1	Bagging算法框架	17
2.2	上下文模型学习算法框架	25
2.3	贪心搜索算法框架	26
3.1	16个通道特征列表	29
3.2	9类数据库图片来源列表	31

插 图

1.1 物体识别分类示意图	1
1.2 物体识别面临挑战示意图	2
1.3 物体识别具体任务示意图	2
1.4 滑动窗口检测示意图	3
1.5 Boosting 人脸检测采用特征和结果图	4
1.6 人脸检测级联示意图	5
1.7 霍夫投票检测示意图	5
1.8 ISM 模型的学习过程示意图	6
1.9 ISM 模型的检测过程示意图	7
1.10 上下文关系辅助物体识别示意图	9
1.11 全局上下文模型示意图	11
1.12 不同层次的局部相互关系示意图	13
2.1 多类物体检测算法框架示意图	15
2.2 物体相对位置关系示意图	23
3.1 9类数据库的实验结果统计图	29
3.2 9类数据库部分实验结果展示图	30
3.3 LebelMe 数据库多类物体检测混合矩阵	32
3.4 LableMe 数据库部分实验结果展示图	33
4.1 计算机视觉问题关系示意图	35

第一章 绪论

本章对多类物体检测的研究背景和现状作简要介绍，主要包括物体识别概述，物体检测概述和上下文关系模型概述，并且最后给出本文组织结构。

1.1 物体识别概述

物体识别(Object Recognition) [21] 是计算机视觉领域(Computer Vision) [36] 的核心问题之一，许多视觉问题的解决依赖于物体识别的结果。物体识别自身的应用也极为广泛，例如：基于内容的图像检索(Content Based Image Retrieval)、机器人导航(Navigation for Robots)、场景理解(Scene Interpretation)等等。

物体识别通常可分为两类：特定物体的识别(Specific Object Recognition) [28] 和物体类别识别(Generic Object Category Recognition) [12, 26]。前者在图像中寻找特定物体，例如：人脸、杂志封面、邻居家的车等。对于后者，则只需要识别出某一类物体的不同实例，例如：楼房，行人，汽车等等(见图 1.1)。

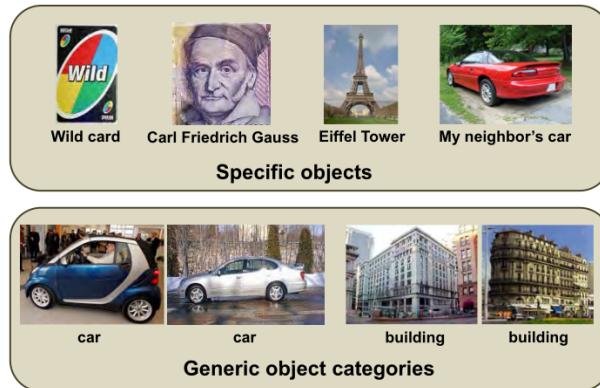


图 1.1: 物体识别分类示意图，摘自 [21] .

物体识别相对于人来说较为容易，一般幼儿就可以轻易地识别出不同类别的物体。然而，这个问题对计算机来说十分困难，面临着各种挑战，例如：尺度变化(Scale)、视角变化(Viewpoint)、光照变化(Illumination)、背景干扰(Background Clutter)、物体形变(Deformation)、部分遮挡(Occlusion)等等(见图 1.2)。此外，物体种类识别还需要处理同一类物体之间的变化(Intra Class

Variation), 例如图 1.1 中, 两辆汽车的外形可能相差很大, 但是它们仍然属于同一类物体, 因此, 物体种类识别相对更难一些。

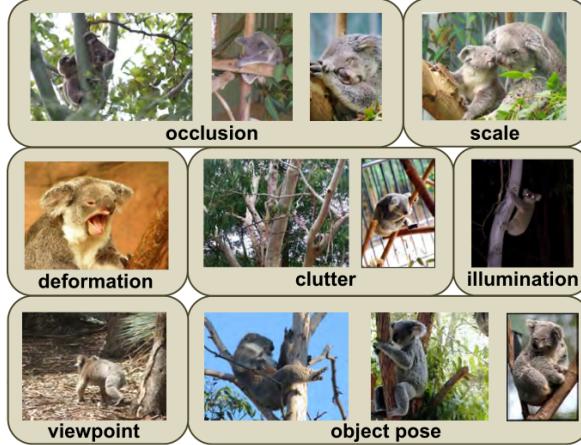


图 1.2: 物体识别面临挑战示意图, 摘自 [21].

在物体种类识别中, 根据最后识别结果的精细程度大体可以分为三种任务: 图像分类(Image Classification) [15, 27]、物体检测(Object Detection) [11, 46] 和物体分割(Object Segmentation) [20, 35], 具体见图 1.3。最粗糙的识别结果是图像分类, 即判断该幅图像中是否含有某类物体, 这种识别结果对基于内容的图像检索已经足够。有时候, 我们不仅需要判断是否存在某类物体, 同时我们还需要大概定位出物体出现的位置和物体的尺度, 这就是物体检测, 一般使用一个固定大小的窗口, 具体标出物体的位置和大小, 这种识别结果对机器人导航来说是必须的。但某些时候需要像素层次的识别结果(即物体分割), 此时需要将整幅图像的像素分割为物体部分和背景部分, 这类识别结果适合于对图像进行标注或对场景的理解。本文主要研究类别层次的物体检测, 故以下主要讨论物体检测。

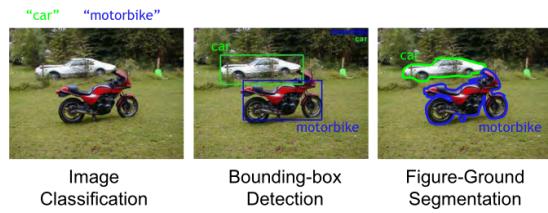


图 1.3: 物体识别具体任务示意图, 摘自 [21].

1.2 物体检测算法概述

经过过去几十年的发展，物体检测已经取得了很大的进步，特别在一些常见物体类别上，检测效果明显提高，例如：人脸检测(Face Detection) [46]，行人检测(Pedestrian Detection) [11]，汽车检测(Car Detection) [1] 等等。这些检测算法大体可以主要分为：滑动窗口方法(Sliding Window Method) [11, 12, 14, 46]和霍夫投票方法(Hough Voting Method) [17, 26, 31, 32]。本文下面具体介绍这两类检测算法的基本原理。

1.2.1 滑动窗口方法

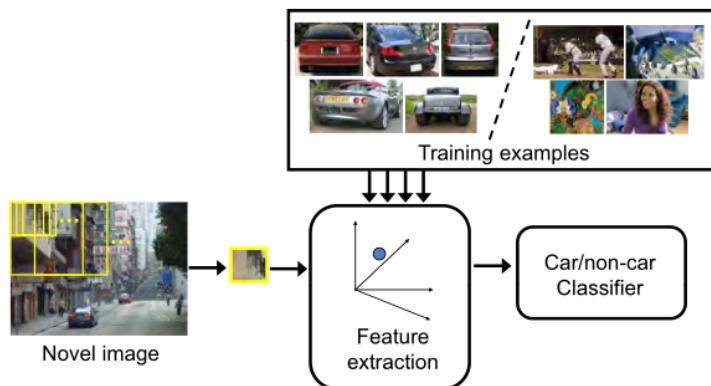
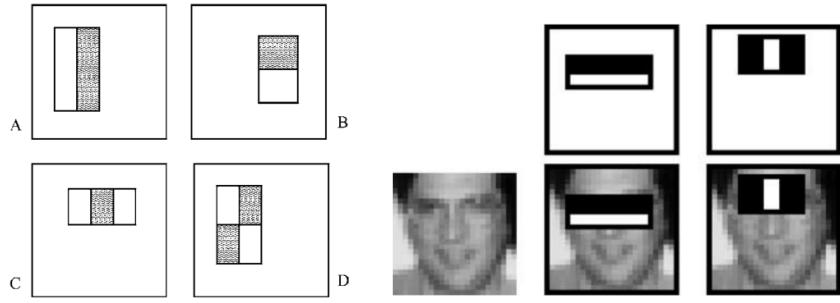


图 1.4: 滑动窗口检测示意图，摘自 [21].

滑动窗口方法首先对训练图像提取底层特征，例如：梯度方向直方图(Histogram of Oriented Gradients) [11]、尺度不变特征(Scale Invariant Feature Transform) [29] 等等。然后，使用这些特征训练一个两类分类器，例如：支持向量机(Support Vector Machine) [11]，Boosting 分类器(Boosting) [46]。接着，使用滑动窗口扫描整幅图像的各个位置和各个尺度，计算分类器分类的结果。最后，通过非极大值抑制(Non-Maxima Suppression) [11] 等后处理完善物体检测(见示意图 1.4)。下面以人脸检测为例，具体介绍使用滑动窗口技术的原理。

在 [46] 中，Viola *et al.* 提出了一种基于滑动窗口的人脸检测技术，该技术也是目前使用最广泛的人脸检测技术，它第一次把 Boosting 技术应用到了计算机视觉领域。首先，该方法使用了一种比较简单的特征(见图 1.5(a))，该特征被称为矩形框像素差值特征，它由2-4个矩形框组成，用灰色矩形框的像素减去白色



(a) Boosting 人脸检测使用的特征. (b) Boosting 优先选出的人脸特征.

图 1.5: Boosting 人脸检测采用特征和结果图, 摘自 [46].

矩形框内的像素。然后，基于这些特征，该方法构造了出一个 Boosting 的分类器 $h(\mathbf{x})$ ，这个分类器是由一系列若分类器构成：

$$h(\mathbf{x}) = \text{sign} \left[\sum_{j=0}^{m-1} \alpha_j h_j(\mathbf{x}) \right] \quad (1.1)$$

这些弱分类器 $h_j(\mathbf{x})$ 一般都是一些简单的阈值函数：

$$h_j(\mathbf{x}) = a_j[f_j < \theta_j] + b_j[f_j \geq \theta_j] = \begin{cases} a_j & \text{if } f_j < \theta_j \\ b_j & \text{otherwise.} \end{cases} \quad (1.2)$$

其中 \mathbf{x} 代表图像 patch， f_j 代表上述四种特征之一。Boosting 训练过程中，往往每次弱分类器，只选取特征的一维来作为分类标准，其中最先选出来的特征见图 1.5(b)。最后，该方法为了进一步加速检测效率，构造了一个级联分类器，每一步由一些简单分类起来(例如2-3项的 Boosting 分类器)，最后级联起来，这样可以在不降低正确率的情况下，很大提高检测效率，见图 1.6，具体细节可以参见原文 [46]。

虽然，滑动窗口检测技术在检测效率有了很多改进，例如：使用级联分类器 [46]，分支界定 [25] 等等，但是跟霍夫投票方法相比较，整体效率还是相对较低，因为滑动窗口需要扫描整幅图像，同时还需要变换原始图像的尺度或者检测窗口的尺度，进行循环扫描。

1.2.2 霍夫投票方法

霍夫投票方法首先在 [24] 中提出，然后在 [3] 中被应用到任意形状的检测，

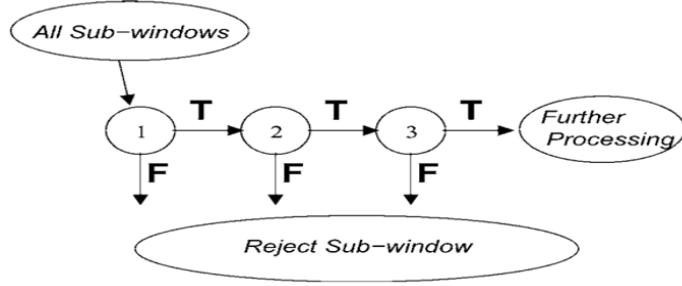


图 1.6: 人脸检测级联示意图, 摘自 [46].

最近被广泛应用到物体检测中 [17, 26, 28]。霍夫投票方法一般认为物体存在一个中心位置, 物体的其他部分相对物体中心存在一个偏移量。在学习阶段, 根据图像的局部 patch (可以是基于兴趣点的 [26], 也可以是稠密采样的 [17]), 构建物体模型, 建立物体部分和物体中心的关系。然后在检测时候, 使用检测图像局部 patch 在假设空间(Object Hypothesis Space)中对物体可能存在的位置和大小进行投票, 也称投票空间, 最后根据投票结果, 把投票空间中的一些极值点作为物体的位置, 同时根据投票可以反推出物体各个部分, 见示意图 1.7。以下以一种近期发表的霍夫投票模型为例对其作详细介绍。

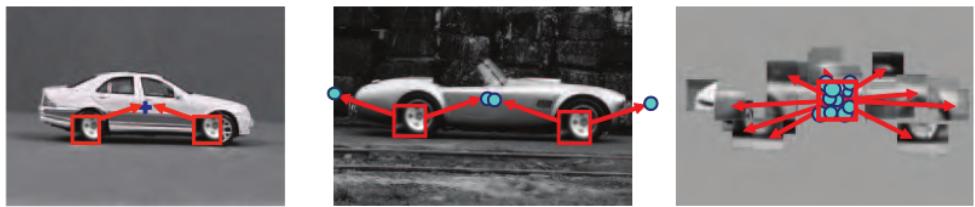


图 1.7: 霍夫投票检测示意图, 摘自 [21].

在 [26] 中, Leibe *et al.* 提出隐形状模型(Implicit Shape Model, ISM), 使用该模型同时来进行物体检测和物体分割。ISM 对每类物体生成一个编码本(Codebook), 编码本是通过训练图像的局部特征 f 聚类形成, 这些局部特征一般是先进过兴趣点提取, 确定 patch 的位置和大小, 例如: Harris-Laplace, Hessian-Laplace 等等。然后从这些局部区域中提取特征描述符, 即局部特征, 例如: Greyvalue Patches, SIFT 等等。这些聚类的中心称为编码本中的每一个单词(Visual Word), 每一个单词都存存在一个空间分布表(Spatial Occurrence Distribution): $\{(x_{occ}, y_{occ}, s_{occ})\}$, 列表记录了属于这一类的 patch 相对物体中心

的偏移量和尺度，具体生成过程见图 1.8。

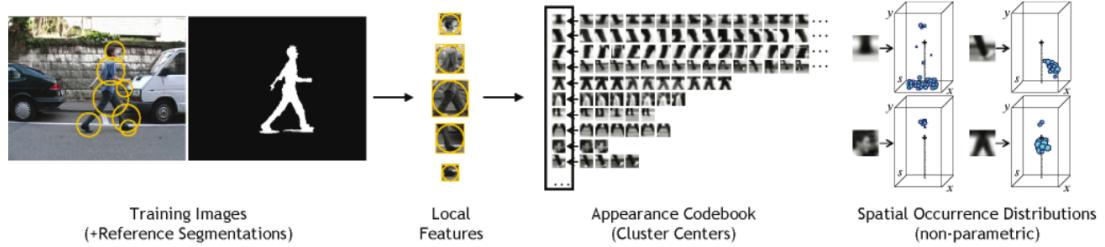


图 1.8: ISM 模型的学习过程, 摘自 [26].

在物体检测阶段, 该方法使用已经训练好的物体模型 (即编码本) 来对物体的位置和大小进行投票。与学习阶段类似, 对测试图像, 首先进行局部特征提取 (f, l), 其中 f 为特征, l 为特征所在的位置然后和编码本中的单词进行匹配, 得到概率 $p(C_i|f, l)$, 接着根据编码本中每个单词的偏移量列表得到在位置 x 检测到物体 o_n 的概率 $p(o_n, x|f, C_i, l)$, 最后对匹配的单词进行边缘化得到如下公式:

$$p(o_n, x|f, l) = \sum_i p(o_n, x|f, C_i, l)p(C_i|f, l) \quad (1.3)$$

根据一些概率独立关系, 例如: 局部特征 f 和编码本的匹配和局部特征所在的位置 l 无关, 于是公式 1.3 可以简化如下:

$$\begin{aligned} p(o_n, x|f, l) &= \sum_i p(o_n, x|C_i, l)p(C_i|f) \\ &= \sum_i p(x|o_n, C_i, l)p(o_n|C_i, l)p(C_i|f) \end{aligned} \quad (1.4)$$

当对物体中心进行计算的时候, 物体的尺度作为投票空间的第三维。如果测试图像的局部特征的位置为 $(x_{img}, y_{img}, s_{img})$, 它匹配的特征位置的空间分布为 $(x_{occ}, y_{occ}, s_{occ})$, 于是投票的坐标关系如下:

$$\begin{aligned} x_{vote} &= x_{img} - x_{occ}(s_{img}/s_{occ}) \\ y_{vote} &= y_{img} - y_{occ}(s_{img}/s_{occ}) \\ s_{vote} &= s_{img}/s_{occ} \end{aligned} \quad (1.5)$$

最后, 使用 Mean-Shift [10] 方法在投票空间中寻找极值点, 使用如下核密

度估计：

$$\hat{p}(o_n, x) = \frac{1}{V_b} \sum_k \sum_j p(o_n, x_j | f_k, l_k) K\left(\frac{x - x_j}{b}\right) \quad (1.6)$$

其中 K 一个圆对称的，非负中心为零的核函数， b 是核函数的带宽， V_b 是核函数的体积，具体检测流程见图示 1.8。

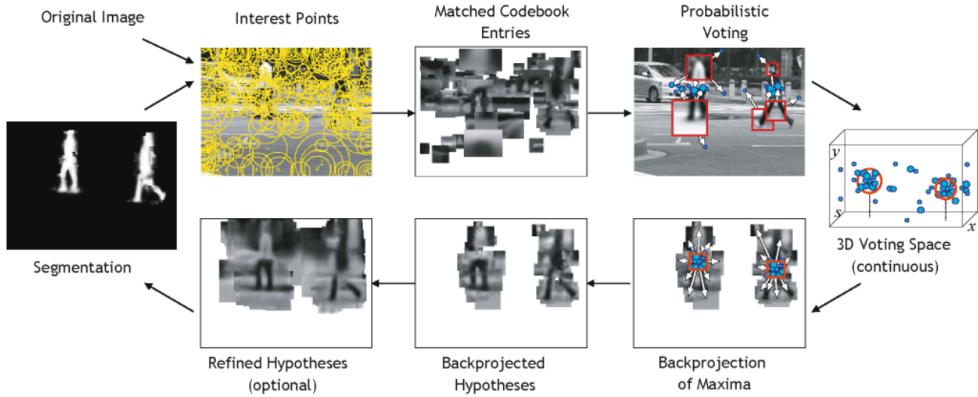


图 1.9: ISM 模型的检测过程示意图，摘自 [26] .

ISM 模型提出了一种基于霍夫变换的物体检测算法，对物体尺度变化存在一定的鲁棒性，同时检测效率相对较高。此外，作者还把物体检测和物体分割相结合，相互改进实验效果，在此不再介绍，具体细节可以参考原文 [26] 。

1.2.3 多类物体检测

前面主要介绍了两种主流的物体检测方法：滑动窗口方法和霍夫投票方法，然而在现有的工作中，这两种方法主要是用来做单类物体的检测，即一个模型只能检测一类物体，如果需要检测多类物体，需要训练多个模型，对测试图像进行重复多次检测，才能完成多类物体检测的任务。目前多类物体检测的工作相对较少 [12, 32, 43]，其中有些多类检测的算法，其实只是多类训练，检测的时候还是单类 [32, 43]，使用多类物体训练，又进行多类物体检测就相对更少了 [12]。

多类物体检测并不是简单的把单类物体检测推广到多类，多类检测需要考虑一些单类检测没有碰到的问题。学习时候，需要考虑类别之间的相似点，有些不同类别之间的一些特征是可以共享的(Sharing Features) [43]。考虑到学习的效率，可采用增量学习(Incremental Learning) [32]，这样不仅可以加快学习效

率，同时可以减少学习需要的样本，降低学习代价。多类物体检测还存在一个单类物体检测没碰到的问题，即多类物体之间是存在一定的关系的。物体不是孤立存在，它与周围的环境和其他物体总是存在某种联系，例如：鼠标和键盘经常出现，椅子经常出现在桌子的下面，街道场景很容易检测到行人和汽车等等，这种联系我们称为上下文联系(Context)。在多类物体检测的时候，除了要考虑物体本身一些属性，例如：形状，外表等等，物体的上下文联系也是可以帮助我们的检测 [12]。

在 [12] 中，Desai *et al.* 提出了一个多类物体检测的模型，该模型把一幅图像的多类物体检测问题归结为一个结构化预测问题(Structured Prediction Problem)，该文不仅考虑物体本身的外表属性，还考虑了物体之间的相对关系。总的来说，物体之间的相互关系是极其复杂的，这些关系对多类物体检测来说是一个极其重要的线索，既可以辅助一些正确物体的检测，又可以抑制一些错误物体的检测。

在其模型中，该文主要考虑了物体之间的相对位置关系，物体之间主要存在五种位置关系：Ontop、Above、Below、Nextto、Farnear。假设考虑 K 类物体，物体标签 $y_i \in \{0, \dots, K\}$ ，图像中的物体窗口标记为他的中心和尺度 $l_i = (x, y, s)$ ，同时该窗口的特征标记为 x_i ，一幅图像可以用一组特征来描述 $X = \{x_i : i = 1, \dots, M\}$ ，图像所对应的标签为： $Y = \{y_i : i = 1, \dots, M\}$ 。于是，一副图像 X 取标签 Y 建模如下：

$$S(X, Y) = \sum_{i,j} \omega_{y_i, y_j}^T d_{ij} + \sum_i \omega_{y_i}^T x_i \quad (1.7)$$

其中 $S(X, Y)$ 代表图像和标签的得分， ω_{y_i, y_j}^T 表示物体之间几何位置关系的权重， $\omega_{y_i}^T$ 表示物体本身的模板， d_{ij} 表示物体之间相对关系的特征。

该文把该模型的学习转化为凸优化(Convex Optimization)问题，使用一个割平面算法(Cutting Plane Optimization)来进行学习。同时，在最后的物体检测阶段，采用一种贪心的策略，使用该模型进行多类物体检测，具体细节参见原来文章 [12]。

通过上面的介绍，我们发现物体之间的上下文关系在多类物体检测中占有很重要的地位，下面本文对现有的上下文关系以及上下文模型做一个简单的概述。

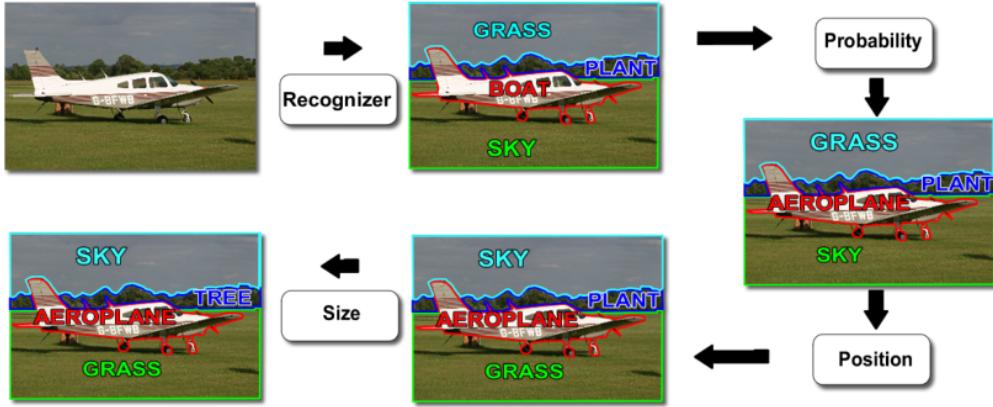


图 1.10: 上下文关系辅助物体识别示意图, 摘自 [18].

1.3 上下文模型概述

在现实世界中, 物体存在都不是孤立的, 它与它周围的环境(包括: 场景, 物体等等)总是存在某种联系, 这种联系被我们称为上下文联系(Context)。根据身理学家研究, 物体这种上下文联系在人类识别物体的时候具有很大作用[30]。在计算机视觉领域, 越来越多的人也开始考虑利用这种上下文联系, 来辅助物体识别和场景理解[13], 并且取得了一定的成果。下面将对上下文关系如何分类和建模作简单介绍。

1.3.1 上下文关系分类

Biederman 在 [5] 中将物体和他周围的环境的联系分为五种: Interposition、Support、Pobabilit、Posistion 和 Familiar Size。其中Pobability、Posistion 和 Familiar Size 被称为语义关系, 也被为上下文特征, 它们经常用来辅助物体识别, 见示意图 1.10。

Probability 也被称为语义上下文联系(**Semantic Context**)。现实世界中, 一个场景通常由一些常见的物体按照特定布局构成, 语义上下文联系是指一个物体会在某种场景中出现的可能性, 它通常被用来表示某种物体和其他一些物体或者场景同时出现(Cooccurrence)。例如图 1.10 中, 飞机和天空、草地同时出现的可能性比较大, 而船和天空和草地同时出现的可能性比较低, 于是我们通过语义上下文联系, 就可以排除船的错误识别, 把船纠正为飞机。这种语义上下

文联系一般来源于一个专家系统，图像数据库或者Google搜索引擎，通常表现为一些规则(Predined Rule)或者同时出现矩阵(Cooccurrence Matrix)，在过去十几年的发展中，有很多论文都使用语义上下文联系来辅助物体识别，具体可以参见文章 [35, 41, 47]。

Position 也被称为空间上下文联系(**Spatial Context**)。真实场景中，不同种类物体之间总是存在某种相对位置关系，空间上下文联系是指一类物体相对其他物体在某些位置出现的可能性。例如图 1.10 中，天空正常应该出现在飞机的上方，草地正常应该出现在飞机的下方，于是利用这种空间上下文联系，我们就可以纠正天空和草地的错误识别。早期，这些空间上下文主要来源于专家系统，这些专家系统的知识主要局限于某个领域，它不能很好地应用到现实的场景中去。近期，视觉领域中的这种空间上下文联系主要来源于图像数据库的标注，这种联系可以很好地应用到现实场景。在过去的研究中，也有相当多的计算机视觉论文研究如何使用这种空间上下文联系来辅助物体识别，具体可以参见文章 [12, 20, 41]。

Size 也被称为尺度上下文联系(**Scale Context**)。物理世界中，物体总是以某种大小尺度存在，不同种类物体之间的尺度总是存在某种联系，尺度上下文联系是指一类物体相对他周围其他物体的尺度大小关系。尺度上下文联系，不仅研究不同类物体同时出现个可能性，而且他还研究不同类物体之间的位置关系和深度关系。例如图 1.10 中，我们已经识别出来的飞机大小和位置，同时植物的大小和飞机也存在一个相对关系，植物的大小一般不可能会比飞机大，于是我们可以纠正植物的错误识别。总的来说，尺度上下文联系要比语义上下文联系和空间上下文联系相对复杂，因为它需要分析出图像中物体的细节信息。在过去的研究中，也出现了一些将尺度上下文联系加入到物体识别中的论文，具体可以参见文章 [41, 44, 45]。

这三类上下文联系，在过去计算机视觉领域被广泛地应用到物体识别中去，从而提高了物体识别的正确率。一个模型往往只是明确地使用一种或两种上下文联系，其中空间上下文联系和尺度上下文联系最为广泛使用，因为总的来说，空间上下文联系和尺度上下文联系已经包含了语义上下文联系。但是，语义上下文联系应该是最简单，最有效的上下文联系，它能够更有效地提高物体识别的效率。因为，现实世界的图片往往是复杂的，空间上下文联系和尺度上

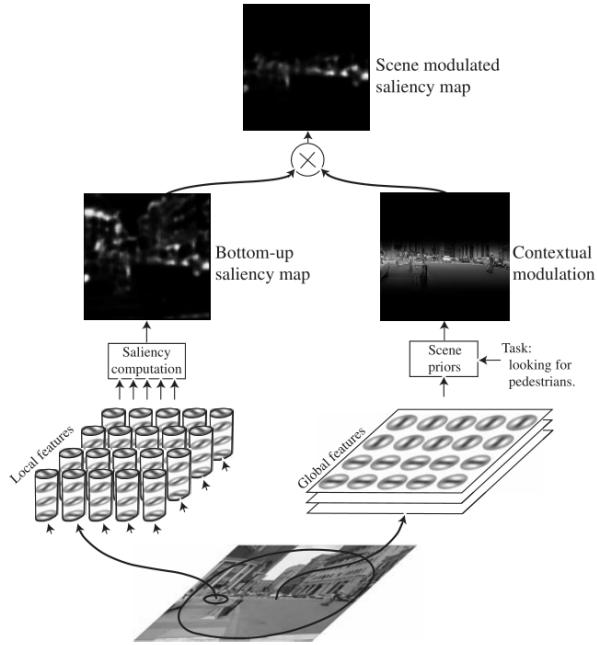


图 1.11: 全局上下文模型示意图, 摘自 [42].

下文联系变化比较大, 而语义上下文联系变化相对较少, 它能够提供的有效信息更多, 显得更为宝贵。

1.3.2 上下文关系建模

上面详细介绍了上下文联系的分类以及各自的含义, 下面本文将详细介绍上下文联系究竟如何被应用到物体识别中去。总的来说, 上下文模型主要可以分为全局上下文模型(Global Context Model)和局部上下文模型(Local Context Model) [34]。全局上下文模型主要是把整幅图像作为一个整体来考虑, 使用一些全局有用的信息来帮助物体识别, 这种模型主要考虑物体与图像场景之间的关系, 例如厨房中很可能出现火炉, 因此, 也称为基于场景的上下文模型(Scene Based Context Model, SBC)。局部上下文模型主要考虑物体周围的信息来辅助识别, 例如周围的像素、区域、物体, 这种模型其实主要是考图像中虑物体之间的相互关系, 例如桌子旁边很有可能出现椅子, 因此, 也被称为基于物体的上下文模型(Object Based Context Model, OBC)。

全局上下文模型使用场景先验来辅助物体识别, 它主要是用来建模物体和场景之间的关系(Object-Scene Interaction) [37, 41, 44, 45]。下面将介绍一个算法

框架，该框架使用全局上下文模型，使用图像的全局特征作为场景先验，辅助物体识别。在 [41] 中，Torralba 根据一个图像中出现的物体和图像中一些全局特征的相关性，提出一个基于这些全局特征的物体识别方法，见示意图 1.11。假设观察到得图像特征为 $\mathbf{v} = \{\mathbf{v}_l, \mathbf{v}_c\}$ ，其中 \mathbf{v}_l 为局部特征， \mathbf{v}_c 为全局特征，物体识别就是需要求概率 $p(O|\mathbf{v})$ ，其中 $O = \{o, \mathbf{x}, \sigma\}$ ，分别表示物体的种类，位置，尺度。在不考虑上下文联系的情况下，就认为物体识别只与物体周围的局部特征相关，于是一般作如下近似：

$$p(O|\mathbf{v}) \simeq p(O|\mathbf{v}_l) = \frac{p(\mathbf{v}_l|O)}{p(\mathbf{v}_l)} p(O) \quad (1.8)$$

但是，实际中上述近似是不合理的，物体的出现和它所在的环境是存在一定的联系，于是 Torralba 做了下面精确 Bayes 推导：

$$p(O|\mathbf{v}) = p(O|\mathbf{v}_l, \mathbf{v}_c) = \frac{p(\mathbf{v}_l|O, \mathbf{v}_c)}{p(\mathbf{v}_l|\mathbf{v}_c)} p(O|\mathbf{v}_c) \quad (1.9)$$

从上面可以看出，公式 1.9 中的每一项概率都比公式 1.8 中的每一项多了一个条件 \mathbf{v}_c 。这次推导主要将上述概率分为两项，第一项中 $p(\mathbf{v}_l|O, \mathbf{v}_c)$ 为在给定全局特征 \mathbf{v}_c 和局部特征 \mathbf{v}_l 的情况下，物体 O 的似然函数， $p(\mathbf{v}_l|\mathbf{v}_c)$ 为归一化常数：给定全局特征 \mathbf{v}_c ，出现局部特征 \mathbf{v}_l 的概率。第二项 $p(O|\mathbf{v}_c)$ ，为给定全局特征 \mathbf{v}_c 情况下，物体 O 的先验概率，这先验概率也被称为上下文先验(Context Based Prior)。这两项对物体识别来说都是比较重要的，在这个全局上下文模型中，他们重点考虑了第二项，可以拆分如下：

$$p(O|\mathbf{v}_c) = p(o, \mathbf{x}, \sigma|\mathbf{v}_c) = p(\sigma|\mathbf{x}, o, \mathbf{v}_c)p(\mathbf{x}|o, \mathbf{v}_c)p(o|\mathbf{v}_c) \quad (1.10)$$

这三项的具体含义如下：

- Object Priming: $p(o|\mathbf{v}_c)$ 表示给定全局特征 \mathbf{v}_c ，物体 o 存在的先验概率。
- Focus of Attention: $p(\mathbf{x}|o, \mathbf{v}_c)$ 表示给定全局特征 \mathbf{v}_c 和确定存在物体 o 的条件下，物体位置 \mathbf{x} 的概率。
- Scale Selection: $p(\sigma|\mathbf{x}, o, \mathbf{v}_c)$ 表示给定全局特征 \mathbf{v}_c 和物体类别 o 与物体位置 \mathbf{x} 的条件下，物体大小 σ 的概率。

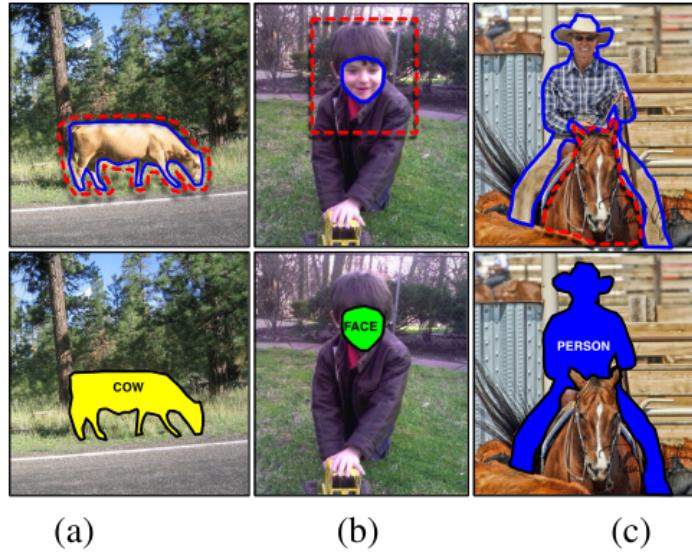


图 1.12: 不同层次的局部相互关系示意图, 摘自 [19].

Torralba 就是按照上述概率描述形式, 将全局的特征融入了物体识别, 提出了一个全局的上下文模型, 具体细节可以参见原来文章 [44]。

局部上下文模型主要是考虑图像一些局部的信息, 考虑物体与它周围信息的相互关系, 一般采用一种自底向上的方法来辅助物体识别。这种局部的相互关系可以分成不同层次: 像素层次(Pixel Interactions) [23], 区域层次(Region Interactions) [16] 和物体层次(Object Interactions) [35], 这些不同层次的关系可以单独用来建模, 也可以混合用来建模 [19]。像素层次的相互关系考虑物体周围像素之间的相关性, 是属于最底层的上下文关联, 例如图 1.12 中牛与它周围草和树的像素关联; 局域层次的相互关系考虑图像 patch、分割块、物体不同部分之间的相关性, 是属于中层的上下文关联, 例如图 1.12 中人脸和人的身体上部分之间的关联; 物体层次的相互关系主要考虑不同物体之间的相关性, 是属于高层的上下文关联, 例如图 1.12 中人和马之间的关联性。像素层次的相互关系计算复杂度相对较高, 因为像素的数目相对很多, 需要考虑的相互关系对的数量较多。物体层次的相互关系是最有效的, 因为它考虑的关系已经比较高层, 有比较明确地语义信息, 一般一幅场景中出现的物体数量不可能很多。区域层次的相互关系一般都需要一些预处理方法, 先分出一些简单区域, 例如一些分割算法, 在这个基础上考虑一些区域相互关系, 这样效率较

高。这些不同层次的局部相互关系，考虑起来相对复杂，一般都是通过一些随机场方法进行建模，例如马尔科夫随机场(Markov Random Fields, MRFs)，条件随机场(Conditional Random Fields, CRFs)，然后利用学习的方法对模型参数进行学习，最后使用这些模型将不同层次的局部上下文关系加入到物体识别中去[16, 19, 23, 35]。在[35]中，Rabinovich *et al.* 考虑物体层次之间的相互关系，建模如下：

$$p(c_1, \dots, c_k | S_1, \dots, S_k) = \frac{B(c_1, \dots, c_k) \prod_{i=1}^k A(i)}{Z(\Phi, S_1, \dots, S_k)} \quad (1.11)$$

其中， $A(i) = p(c_i | S_i)$ ， $B(c_1, \dots, c_k) = \exp\{\sum_{i,j=1}^k \phi(c_i, c_j)\}$ ，具体细节参见原来文章[35]。

1.4 本文主要研究工作和组织结构

本篇文章主要研究多类物体检测的课题，提出了一种基于局部特征和上下文关系的多类物体检测算法框架。物体识别目前是计算机视觉领域研究的热点问题之一，它是图像内容理解的关键部分，视觉中很多难题的解决都依赖于物体识别技术。物体检测是物体识别的一个特定任务，目前存在一系列单类物体检测的算法，特别一些常见物体的检测已经相对成熟，然而，多类物体同时检测的技术研究相对较少。于是，我们基于现在一些单类检测的算法，进行了一些扩充，推广到了多类检测，同时本文还融入一个上下文关系，从而进一步辅助多类检测，提高检测的正确率。围绕上述课题，本文的结构安排如下：

第一章介绍该课题的相关背景知识，包括：物体识别概述，物体检测概述，上下文关系概述，并概述本文的主要研究工作。

第二章详细介绍我们的算法框架和实现细节，包括：单类和多类霍夫森林模型，上下文关系模型，并介绍一些模型的学习方法和预测算法。

第三章主要介绍我们算法框架的实验结果和结果分析，包括：9类数据库的实验结果和LabelMe的实验结果以及相应的结果分析。

第四章总结全文，并本文的一些后续工作作出展望。

第二章 基于局部特征和上下文的多类物体检测算法

本章内容对本文提出的多类物体检测的算法框架作详细介绍，具体包括单类和多类霍夫森林模型、上下文关系模型，及相应的模型学习算法。

2.1 引言

通过前面绪论内容的介绍，单类物体的检测目前研究的较多，多类物体的检测研究的相对较少。同时多类物体的检测，并不是简单的单类物体检测的推广，还需要考虑一些单类物体检测没有碰到的问题，例如：多类物体训练时候的特征共享问题，多类物体之间的上下文关系等等。本文在已有单类霍夫森林的基础上提出了一个多类的霍夫森林，使用该多类霍夫森林对多类物体进行检测。同时在多类霍夫森林的投票空间中，考虑多类物体之间的相对位置关系，本文又提出了一个上下文模型，进一步提高物体检测的正确率，见示意图 2.1。下面将详细介绍算法框架的两个主体部分：霍夫森林模型和上下文关系模型。

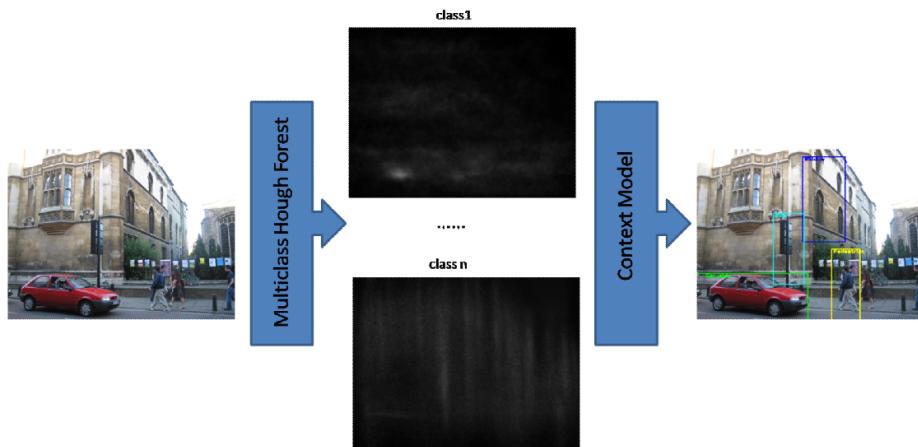


图 2.1: 多类物体检测算法框架示意图

2.2 霍夫森林模型

在 [17] 中，Gall *et al.* 提出了一种基于霍夫变换的判别式物体检测模型。他们使用该模型直接学习一个图像 Patch 块和它的霍夫投票之间的对应关系，而

不是先通过聚类生成一些编码本。图像 Patch 和它的霍夫投票之间的对应关系，既是一个分类问题，又是一个回归问题。首先，需要对图像 Patch 做个分类，判断它属于哪一些物体，然后，需要对该 Patch 的偏移量作预测，算出它与物体中心偏移多少。该文使用了随机森林模型(Random Forest) [9] 来解决上述问题，主要把霍夫投票检测物体的思想融入了随机森林模型，提出了霍夫森林模型(Hough Forest)。下面，首先介绍一下随机森林模型，然后再介绍霍夫森林模型。

2.2.1 随机森林简介

随机森林(Random Forest)是 Breiman 在 [9] 中提出的，它本质上是一种集成学习的方法(Ensemble Learning) [33]。集成学习主要思想用一系列的弱分类器(Weak Learner)构成一个强分类器(Strong Learner)，这个强分类器的泛化能力要强于每个弱分类器，能够进行更准确的预测。集成学习一般需要经过两个过程：首先是一列弱分类器的训练学习，这些分类器的学习方法大体分两类：并行训练(一个分类器的训练与其他分类器的训练是相互独立的，例如：Bagging [8])和串行训练(一个分类器的训练与其他分类器的训练是相关的，例如：Boosting [39])；然后这些弱分类器组合起来形成一个强的分类器，常见的组合方式有： Majority Voting (针对分类问题) 和 Weighted Averaging (针对回归问题)。

随机森林采用的是并行训练过程，其具体定义如下 [9]：

定义 2.1. 随机森林是一个强分类器，由一系列树形结构的分类器集成起来： $\{h(\mathbf{x}, \Theta_k), k = 1, \dots\}$ ，其中 $\{\Theta_k\}$ 是独立同分布的随机向量，每个弱分类器都对输入 \mathbf{x} 做一份预测，最后随机森林的预测结果是这些预测的组合。

在 [9] 中，随机森林主要是基于 Bagging [8] (见表格 2.1)训练框架，加入了随机特征选取(Random Feature Selection)：首先对每一个决策树(Decision Tree)，有重复地采样生成一批训练样本；然后使用该样本，结合随机特征选取训练每棵决策树，并且这些决策树是没有剪枝的；最后使用这些决策树组成随机森林。Breiman 主要考虑了两种随机特征选取的方法：随机输入选取(Random Input Selection)和输入线性组合(Linear Combinations of Inputs)，通过实验证明，

表 2.1: Bagging 算法框架, 摘自[4].

Algorithm: The Bagging Algorithm.

Input: Data Set \mathcal{D} , Learning Algorithm \mathcal{L} , Number of Learning Rounds T .

Process:

1. for $i = 1$ to T {
2. \mathcal{D}' = bootstrap sample from \mathcal{D} (i.i.d sample with replacement).
3. $h_i = \mathcal{L}(\mathcal{D}')$.
4. }
5. $H(\mathbf{x}) = \arg \max_{y \in \mathcal{Y}} \sum_{i=1}^T 1(y = h_i(\mathbf{x}))$.

Output: Classifier: $H(\mathbf{x})$.

这两种随机特征的选取方法提高预测的正确率, 达到了 Adaboost 的效果(有时甚至超过 Adaboost)。同时, 随机森林还存在一些其它优点:

- 它对噪声相对鲁棒。
- 它训练的速度比 Bagging 和 Boosting 快。
- 它的算法框架简单, 比较容易实现并行。
- 它可以提供比较有用的信息、相关性和强度分析。

总的来说, 随机森林是一个非常有效地监督学习的工具, 它给出了一个比较通用的集成学习的算法框架, 可以结合各种随机方法(例如: Random Inputs, Random Features)。只要我们选择适当的随机方法, 就可以构造非常高效的分类器, 实现比较精确的预测。

2.2.2 单类霍夫森林

在 [17] 中, Gall *et al.* 将随机森林的框架应用到了基于霍夫变换的物体检测中去, 提出了霍夫森林的模型。霍夫森林使用了随机森林的一些优点, 例如: 可以大数据量训练, 并且同时避免过拟合问题(Overfitting); 随机森林对噪声不敏感, 具有较好的鲁棒性等等, 使得霍夫森林模型在物体检测方面更加有效和

鲁棒。霍夫森林主要结合了两种随机化方法：每颗二叉树训练的数据样本是随机生成的，是整个数据库的一个子集；二叉树每个分叉节点，考虑的分类器也是随机生成的，是所有可能分类器的一个子集。下面我们将详细介绍霍夫森林的训练和检测。

霍夫森林是由一系列二叉决策树组成，这些二叉决策树并列关系，它们分别是有随机生成的训练样本子集学习而成。每颗树 T 都是根据一些图像 Patches : $\{\mathcal{P}_i = (\mathcal{I}_i, c_i, \mathbf{d}_i)\}$ 构造出来的，其中 \mathcal{I}_i 是图像 Patch 的特征， c_i 是图像 Patch 的类别标签， \mathbf{d}_i 是图像 Patch 相对物体中心的偏移量。我们的训练图像 Patch 是从一系列的训练图像中产生的，这些训练图像由正样本和负样本组成。正样本就是包含我们感兴趣物体的图像，物体一般是被矩形框确定的，这些图片的 Patch 的类别标签 $c_i = 1$ ，同时每个 Patch 相对物体中心是存在一个偏移量 \mathbf{d}_i 。负样本是指没有我们感兴趣物体的图像，一般就是一些背景图片，这些图片的 Patch 的类别标签 $c_i = 0$ ，同时这些 Patch 是不存在偏移量 \mathbf{d}_i 。

霍夫森林的每个叶子节点都可以被认为是一个判别式单词，可以用它直接对物体的中心投票。每个叶子节点存在一个比例因子 C_L ，记录到达该叶子节点，正样本 Patch 的比例，例如 $C_L = 1$ 说明到达该叶子的节点的所有 Patch 都属于物体的一个部分。同时每个叶子节点还存在一个偏移量列表 D_L ，记录到达该叶子节点，所有正样本 Patch 相对于物体中心的偏移量。这些信息，将在物体检测阶段，被直接用来对物体可能存在的位置进行投票。

霍夫森林是由一系列二叉决策树构成的，二叉决策树内部节点就是一个简单的分类器。考虑分类器的效率，选择的一种相对简单的分类方式，分类主要还是根据物体的外表属性来判断的，即图像 Patch 的局部特征 \mathcal{I} ，假设在学习和测试阶段，图像 Patch 大小是固定好的，例如： 16×16 ，物体的局部特征对应很多通道(Channel): $\mathcal{I}_i = (I_i^1, I_i^2, \dots, I_i^C)$ 。于是考虑一个简单的分类器 $t(\mathcal{I}) \rightarrow \{0, 1\}$ ，具体定义如下：

$$t_{a,p,q,r,s,\tau}(\mathcal{I}) = \begin{cases} 0 & \text{if } I^a(p, q) < I^a(r, s) + \tau \\ 1 & \text{otherwise.} \end{cases} \quad (2.1)$$

其中， $a \in \{1, 2, \dots, C\}$ ，对应于图像的某一个通道； (p, q) 和 (r, s) 对应于图像中 Patch 中的具体位置， τ 对应于一个阈值。

霍夫森林每棵二叉决策树的构建，都遵循一般二叉树的构建框架：递归构造二叉树，首先将根节点按照某种分类标准，分成左右子节点，然后对左右子节点，递归构造二叉树，一直满足递归停止条件为止，例如：树的高度达到最大深度 d_{max} 或者 Patch 数目达到最少数目 N_{min} 。二叉树构建的关键部分就是分类标准如何确定，也就是如何选出一个最优的分类器。根据前面的介绍，霍夫森林既是一个分类模型，我们需要给图像 Patch 进行分类，判断他属于哪个类型的物体，同时霍夫森林又是一个回归模型，需要计算出物体 Patch 相对物体中心的偏移量。为了同时达到这两个目的，定义了如下两种不确定性，来衡量分类器的性能。假设图像 Patch 集合 $A = \{\mathcal{P}_i = (\mathcal{I}_i, c_i, \mathbf{d}_i)\}$ ，首先是类别不确定度，定义如下：

$$U_1(A) = |A| \times \sum_{i=0}^1 p_i \log \frac{1}{p_i} \quad (2.2)$$

这是集合的大小和集合的信息熵的乘积。其次，偏移量不确定度定义如下：

$$U_2(A) = \sum_{i:c_i=1} (\mathbf{d}_i - \mathbf{d}_A)^2 \quad (2.3)$$

其中， \mathbf{d}_A 表示平均偏移量。于是，一个分类器 t 的性能评价指标定义如下：

$$Q(t) = \alpha[U_1(A_0) + U_1(A_1)] + (1 - \alpha)[U_2(A_0) + U_2(A_1)] \quad (2.4)$$

其中 $A_k = \{\mathcal{P}_i | t(\mathcal{I}^i) = k\}$, $k \in \{0, 1\}$ ， α 是调节这两种不确定性的权重。于是，在构建过程中，我们每次都是选一个最优的分类器 \hat{t} ，即：

$$\hat{t} = \arg \min_{t^k} Q(t^k) \quad (2.5)$$

霍夫森林在投票的时候，每棵树对物体可能出现的位置分别进行投票，然后森林的投票结果就是所有树投票结果的平均值。对于测试图像而言，考虑一个图像 Patch : $\mathcal{P}(\mathbf{y}) = (\mathcal{I}(\mathbf{y}), c(\mathbf{y}), \mathbf{d}(\mathbf{y}))$ ，该图像 Patch 的中心是 \mathbf{y} ，图像 Patch 的局部特征是 $\mathcal{I}(\mathbf{y})$ ，图像 Patch 的类别标签是 $c(\mathbf{y})$ ，图像 Patch 相对物体中心的偏移量是 $\mathbf{d}(\mathbf{y})$ 。物体检测就是考虑观察到图像 Patch 局部特征 $\mathcal{I}(\mathbf{y})$ 情况下，物体中心位置为 \mathbf{x} 的概率，即 $p(E(\mathbf{x})|\mathcal{I}(\mathbf{y}))$ 。同时，认为只有物体部分的 Patch 对物体中心可能存在的位置才能产生影响，于是，在观察到特征 $\mathcal{I}(\mathbf{y})$

情况下，物体中心在 \mathbf{x} 的概率具体如下：

$$\begin{aligned} p(E(\mathbf{x})|\mathcal{I}(\mathbf{y})) &= p(\mathbf{d}(\mathbf{y}) = \mathbf{y} - \mathbf{x}, c(\mathbf{y}) = 1|\mathcal{I}(\mathbf{y})) \\ &= p(\mathbf{d}(\mathbf{y}) = \mathbf{y} - \mathbf{x}|c(\mathbf{y}) = 1, \mathcal{I}(\mathbf{y}))p(c(\mathbf{y}) = 1|\mathcal{I}(\mathbf{y})) \end{aligned} \quad (2.6)$$

根据该图像到达的叶节点记录的信息，上述概率可以按如下展开：

$$p(E(\mathbf{x})|\mathcal{I}(\mathbf{y}); \mathcal{T}) = \left[\sum_{\mathbf{d} \in D_L} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|(\mathbf{y} - \mathbf{x}) - \mathbf{d}\|^2}{2\sigma^2}\right) \right] \times \frac{C_L}{|D_L|} \quad (2.7)$$

于是，针对整个森林而言，所有的投票结果进行简单的平均，结果如下：

$$p(E(\mathbf{x})|\mathcal{I}(\mathbf{y}); \mathcal{F}) = \frac{1}{T} \sum_{t=1}^T p(E(\mathbf{x})|\mathcal{I}(\mathbf{y}); \mathcal{T}_t) \quad (2.8)$$

最终，对一幅图像所有的 Patch 进行求和，就得到投票空间中位置 \mathbf{x} 的投票结果：

$$V(\mathbf{x}) = \sum_{\mathbf{y} \in B(\mathbf{x})} p(E(\mathbf{x})|\mathcal{I}(\mathbf{y}); \mathcal{F}) \quad (2.9)$$

上述整个检测过程中，并没有考虑到尺度变化问题，为了应对训练图像和测试图像中物体实例的尺度不一致问题，采用了一种多尺度的检测方法。对测试图像进行尺度变化，进行放大和缩小，然后对每个尺度的图像都进行投票，最后使用 MeanShift [10] 算法在三维空间中寻找极值点，确定物体可能出现的位置和尺度。

2.2.3 多类霍夫森林

前面已经详细介绍了单类霍夫森林的构造以及物体检测，然而这种模型只能适用于单类物体的检测，如果用来多类物体检测，必须对每一类物体都训练一个模型，同时检测的时候，需要分开单独对每一类物体进行检测，这样不仅效率低下，同时也没有考虑到不同类别物体之间，它们可能共享一些特征。于是，在单类霍夫森林的基础上，本文提出了多类霍夫森林模型，该模型在训练和检测阶段都和单类霍夫模型存在本质区别。在训练阶段，我们多类霍夫森林是同时对多类物体进行训练，这样不同物体之间共享一些他们公共的特征，在测试阶段使用多类霍夫森林，本文同时对多类物体进行检测，效率提高了。下面，详细介绍多类霍夫森林的构建和检测。

多类霍夫森林也是由一系列二叉决策树组成，每棵二叉树由一些图像 Patches : $\{\mathcal{P}_i = (\mathcal{I}_i, c_i, \mathbf{d}_i)\}$ ，这里 $c_i \in \{0, 1, \dots, |C|\}$ 表示物体类别：0 表示背景， $1, \dots, |C|$ 表示物体类别标签，本文关注的物体总类别数为 $|C|$ ，其他符号与单类森林一样。每棵树的叶子节点有两个数据结构不同类别比例列表 P_L 和偏移量矩阵 D_L ，其中 $P_L = \{p_i\}$ ， p_i 表示第 i 类物体到达该叶子节点的图像 Patch 数目所占的比例， $D_L = \{\mathbf{d}_{ij}\}$ ， \mathbf{d}_{ij} 表示第 i 类物体到达该叶子节点的第 j 个 Patch 相对物体中心的偏移量。

多类霍夫森林的构建和单类霍夫森林的构建类似，都是遵循决策树构建的递归框架。本文采用了和单类森林相同的特征 $\mathcal{I}_i = (I_i^1, I_i^2, \dots, I_i^{|C|})$ 和相同的分类器 $t(\mathcal{I}) \rightarrow \{0, 1\}$ ，具体定义参见公式 2.1。针对分类器的性能分析，本文还是考虑上述两种不确定性：类别不确定性 $U_1(A)$ 和偏移量不确定性 $U_2(A)$ ，但是需要将上述不确定性的定义进行修改，扩展到多类物体。首先，类别不确定性修改如下：

$$U_1(A) = |A| \times \sum_{i=0}^{|C|} p_i \log \frac{1}{p_i} \quad (2.10)$$

这个定义将公式 2.2 中关于信息熵的定义从两类扩展到了多类，其次，偏移量不确定性定义修改如下：

$$U_2(A) = \sum_{i=1}^{|C|} \sum_{j:c_j=i} (\mathbf{d}_j - \mathbf{d}_{A_i})^2 \quad (2.11)$$

这个定义将公式 2.3 中关于偏移量的变化描述从两类扩展到了多类，其中， \mathbf{d}_{A_i} 表示第 i 类图像 Patch 偏移量的均值。关于分类器 t 性能评估还是采用和公式 2.4 一样的定义：

$$Q_{multi}(t) = \alpha[U_1(A_0) + U_1(A_1)] + (1 - \alpha)[U_2(A_0) + U_2(A_1)] \quad (2.12)$$

其中 $A_k = \{\mathcal{P}_i | t(\mathcal{I}^i) = k\}$ ， $k \in \{0, 1\}$ ， α 是调节这两种不确定性的权重。在构建过程中，每次都是选一个最优的分类器 \hat{t} ，即：

$$\hat{t} = \arg \min_{t^k} Q_{multi}(t^k) \quad (2.13)$$

多类霍夫森林在物体检测阶段，可以同时对多类物体检测，每棵树对每类物体可能出现的位置分别进行投票，然后森林的对每类物体的投票结果

就是所有树投票结果的平均值。对于测试图像而言，考虑一个图像 Patch：
 $\mathcal{P}(\mathbf{y}) = (\mathcal{I}(\mathbf{y}), c(\mathbf{y}), \mathbf{d}(\mathbf{y}))$ ，该图像 Patch 的中心是 \mathbf{y} ，图像 Patch 的局部特征是 $\mathcal{I}(\mathbf{y})$ ，图像 Patch 的类别标签是 $c(\mathbf{y})$ 。与单类物体检测类似，多类物体检测也是考虑在观察到考虑局部特征 $\mathcal{I}(\mathbf{y})$ 的情况下，第 i 类物体出现在位置 \mathbf{x} 的概率，即 $p(E(\mathbf{x}), E(i)|\mathcal{I}(\mathbf{y}))$ 。同时，认为只有第 i 类物体部分的 Patch 才会对第 i 类物体中心的位置产生影响，于是，第 i 类物体中心位置为 \mathbf{x} 的概率可以展开如下：

$$\begin{aligned} p(E(\mathbf{x}), E(i)|\mathcal{I}(\mathbf{y})) &= p(\mathbf{d}(\mathbf{y}) = \mathbf{y} - \mathbf{x}, c(\mathbf{y}) = i|\mathcal{I}(\mathbf{y})) \\ &= p(\mathbf{d}(\mathbf{y}) = \mathbf{y} - \mathbf{x}|c(\mathbf{y}) = i, \mathcal{I}(\mathbf{y}))p(c(\mathbf{y}) = i|\mathcal{I}(\mathbf{y})) \end{aligned} \quad (2.14)$$

根据霍夫森林叶子节点记录的信息 P_L 和 D_L ，上述概率可以按如下展开：

$$p(E(\mathbf{x}), E(i)|\mathcal{I}(\mathbf{y}); \mathcal{T}) = \left[\sum_{j: \mathbf{d}_{ij} \in D_L} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|(\mathbf{y} - \mathbf{x}) - \mathbf{d}_{ij}\|^2}{2\sigma^2}\right) \right] \times \frac{p_i}{|D_L^i|} \quad (2.15)$$

其中， $|D^i|$ 表示到达该叶子节点第 i 类物体 Patch 的数目。针对整个森林而言，所有的投票结果进行简单的平均，结果如下：

$$p(E(\mathbf{x}), E(i)|\mathcal{I}(\mathbf{y}); \mathcal{F}) = \frac{1}{T} \sum_{t=1}^T p(E(\mathbf{x}), E(i)|\mathcal{I}(\mathbf{y}); \mathcal{T}_t) \quad (2.16)$$

最终，对一幅图像所有的 Patch 进行求和，就得到第 i 类物体的投票空间中位置 \mathbf{x} 的投票结果：

$$V_i(\mathbf{x}) = \sum_{\mathbf{y} \in B(\mathbf{x})} p(E(\mathbf{x}), E(i)|\mathcal{I}(\mathbf{y}); \mathcal{F}) \quad (2.17)$$

跟单类物体检测处理尺度变化的方法相同，我们对测试图像进行放大和缩小，然后对每个尺度的图像都进行投票，最后使用 MeanShift [10] 算法在三维空间中寻找极值点，确定物体可能出现的位置和尺度。

2.3 上下文关系模型

由绪论部分的内容，我们知道现实中物体不是孤立存在的，物体与它周围的环境(包括：场景，物体)总是存在某种关系，这种关系被称为上下文联系(Context)。通过大量的实验证明，这种上下文联系在物体识别中占据比较重

要的地位，它可以辅助物体识别，提高识别正确率。在本文中，我们考虑了多类物体之间的位置上下文关系，即相对位置关系，在多类霍夫森林模型的基础上提出了一个上下文模型，下面将详细介绍该模型。

2.3.1 位置关系建模

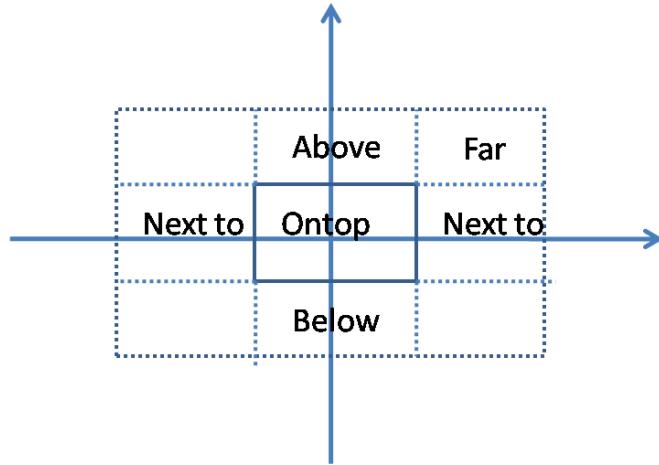


图 2.2: 物体相对位置关系示意图

为了考虑现实场景中物体之间的上下文联系，我们从 LabelMe [38] 数据库中选取了500张街道场景图片，在这些场景中，我们关注经常出现的六类物体：行人、正面汽车、侧面汽车、窗户、路面和柱子。通过图片的选取和观察，我们发现物体的相对位置关系大体可以分为五种位置关系：Ontop, Nextto, Above, Below 和 Far，具体见示意图 2.2。

针对每一种相对位置关系 $R^i, i \in \{1, \dots, 5\}$ ，我们考虑物体对 $\langle m, n \rangle$ 出现的概率，即一个第 m 类物体和第 n 类物体同时出现，且他们的相对位置关系满足 R^i 。不失一般性，我们把这种概率定义为如下指数形式：

$$p(\langle m, n \rangle; \Phi^i) = \frac{1}{Z(\Phi^i)} \exp\{\phi_{mn}^i\} \quad (2.18)$$

其中， Φ^i 为模型参数，它可以是关系 R^i 的关系矩阵， ϕ_{mn}^i 为关系矩阵中的元素， $Z(\Phi^i) = \sum_{m,n} \exp\{\phi_{mn}^i\}$ ，是归一化常数，被称为 Partition Function。我们之所以选择这种概率形式，一方面它不失一般性，可以表达很丰富的概率分布，另一方面，它的数学形式便于我们后面的数学推导。

2.3.2 模型学习过程

我们从LabelMe [38] 数据库中选取了500张街道场景图片，希望利用这些图片学习出上述模型的参数。我们主要考虑用极大似然估计的方法来对上述参数进行点估计，首先我们先从训练图像中，统计出五种关系不同物体对出现的频率矩阵 $L^i, i \in \{1, \dots, 5\}$ ，然后，假设所有图像之间的观测是独立的，于是最后的似然函数只与频率矩阵相关：

$$p(D; \Phi^i) = \frac{1}{Z(\Phi^i)^{M^i}} \exp \left\{ \sum_{m=1}^{|C|} \sum_{n=1}^{|C|} l_{mn}^i \phi_{mn}^i \right\} \quad (2.19)$$

其中， l_{mn}^i 为频率矩阵 L^i 中的元素， M^i 为频率矩阵 L^i 的所有元素之和。为了便于进一步推导，我们取如下对数似然函数：

$$\mathcal{L}(\Phi^i) = \log P(D; \Phi^i) = \sum_{m=1}^{|C|} \sum_{n=1}^{|C|} l_{mn}^i \phi_{mn}^i - M^i \times \log Z(\Phi^i) \quad (2.20)$$

最大似然函数估计，就是对上述对数似然函数进行最优化，即：

$$\hat{\Phi}^i_{MLE} = \arg \max_{\Phi^i} \mathcal{L}(\Phi^i) \quad (2.21)$$

然而对数函数中存在一项 $\log Z(\Phi^i)$ ，这一项为关于 Φ^i 的求和，比较复杂，于是我们采用采样方法(Sampling Method) [6] 来做一个近似，我们采用的是 Importance 采样方法：

$$\begin{aligned} \frac{Z_E}{Z_G} &= \frac{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-G(\mathbf{z}))} \\ &= \frac{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}) + G(\mathbf{z})) \exp(-G(\mathbf{z}))}{\sum_{\mathbf{z}} \exp(-G(\mathbf{z}))} \\ &= \mathbb{E}_{G(\mathbf{z})}[\exp(-E + G)] \\ &\simeq \frac{1}{L} \sum_l \exp(-E(\mathbf{z}^l) + G(\mathbf{z}^l)) \end{aligned} \quad (2.22)$$

其中， Z_E 和 $E(\mathbf{z})$ 分别为采样分布的 Partition Function 和 Energy Function， Z_G 和 $G(\mathbf{z})$ 分别为辅助分布的 Partition Function 和 Energy Function，我们采用的辅助分布就是各个物体对的频率分布，由频率矩阵可以计算得到。

在近似估算出 Partition Function 以后，我们采用梯度下降法寻找最优解：

$$\Phi_{t+1}^i = \Phi_t^i + \beta \nabla_{\Phi^i} \mathcal{L}(\Phi^i) \quad (2.23)$$

表 2.2: 上下文模型学习算法框架.

Algorithm: The Learning Algorithm for Context Model.

Input: Frequency Matrix L^i .

Process:

1. Set Initial Value: $\Phi^i = \Phi_0$, $m_{t-1} = m_t = m_0$.
2. While ($m_{t-1} \leq m_t$) {
3. $m_{t-1} = m_t$, $m_t = 0$.
4. for $i = 1$ to 10 {
5. $\Phi^i = \Phi^i + \beta \nabla_{\Phi^i} \mathcal{L}(\Phi^i)$.
6. $Z(\Phi^i) = \text{Sampling Method}(\Phi^i)$.
7. $m_t = m_t + \mathcal{L}(\Phi^i)$.
8. }
9. $m_t = m_t / 10$.
10. }

Output: Parameter: Φ^i .

其中, β 为更新步长, $\nabla_{\Phi^i} \mathcal{L}(\Phi^i)$ 为对数似然函数关于参数 Φ^i 的导数:

$$\nabla_{\Phi^i} \mathcal{L}(\Phi^i) = \begin{bmatrix} l_{11}^i & \cdots & l_{1|C|}^i \\ \vdots & \ddots & \vdots \\ l_{|C|1}^i & \cdots & l_{|C||C|}^i \end{bmatrix} \quad (2.24)$$

由于上述过程中采用了采样近似求解 Partition Function, 所以我们不能保证上述迭代收敛, 我们采用了一个简单的方法来判断收敛, 每迭代10次, 计算这10次似然函数的均值, 与上10次地均值进行比较, 判断是否增加, 没有增加, 则认为已经到达局部极值, 否则继续迭代, 具体见算法框架表 2.2。

2.3.3 模型预测过程

前面我们已经介绍了如何学习上下文模型, 现在我们考虑如何将霍夫投票的结果和上下文模型结合起来, 实现多类物体检测。根据两个模型, 我们定义

表 2.3: 贪心搜索算法框架.

Algorithm: Greedy Search for Multiclass Object Detection.

Input: Some candidates: $(\mathbf{x}_1, c_1), \dots, (\mathbf{x}_n, c_n)$ with appearance probability:

 $P_{app}(\mathbf{x}_1, c_1), \dots, P_{app}(\mathbf{x}_n, c_n).$
Process:

1. Set Initial Value: $R = \emptyset, P_{con}(\mathbf{x}_1, c_1) = \dots = P_{con}(\mathbf{x}_n, c_n) = 0, M = 0.$
2. $(\mathbf{x}_*, c_*) = \arg \max_{(\mathbf{x}_i, c_i) \notin R} P(\mathbf{x}_i, c_i).$
3. While $(P(\mathbf{x}_*, c_*) > \theta) \{ \% \theta$ is the threshold
4. $R = R \cup \{(\mathbf{x}_*, c_*)\}, M = M + 1.$
5. For the remaining candidates {
6. $P_{con}(\mathbf{x}_j, c_j) = \frac{1}{M} \left[(M - 1)P_{con}(\mathbf{x}_j, c_j) + \frac{1}{Z(\Phi^*)} \exp(\phi_{c_j c_*}^*) \right].$
7. }
8. $(\mathbf{x}_*, c_*) = \arg \max_{(\mathbf{x}_i, c_i) \notin R} P(\mathbf{x}_i, c_i).$
9. }

Output: The detection result $R = \{(\mathbf{x}_i, c_i)\}.$

如下概率形式:

$$P(\mathbf{x}, c) = \omega P_{app}(\mathbf{x}, c) + (1 - \omega) P_{con}(\mathbf{x}, c) \quad (2.25)$$

其中, $P(\mathbf{x}, c)$ 表示在位置 \mathbf{x} 出现一个 c 类物体的概率, $P_{app}(\mathbf{x}, c)$ 表示外表信息对上述概率贡献, 由多类霍夫森林模型产生, $P_{con}(\mathbf{x}, c)$ 表示上下文信息对上述概率的贡献, 由上下文模型产生, 它该物体与周围所有物体构成物体对的概率均值, ω 是这两项之间的一个权重。假设, 在多类霍夫森林投票的基础上, 我们在投票空间中已经得到了一些局部极值点: $(\mathbf{x}_1, c_1), \dots, (\mathbf{x}_n, c_n)$, 并且它们所对应的概率为: $P_{app}(\mathbf{x}_1, c_1), \dots, P_{app}(\mathbf{x}_n, c_n)$, 我们提出了一个贪心近似算法, 求解最后所有物体出现可能性, 总体思想: 我们每轮搜索, 总是寻找出当前概率最大的物体, 然后跟阈值比较, 如果小于, 则结束搜索; 如果大于, 则将该物体加入到结果中, 同时考虑该物体的对其他物体的上下文影响, 更新其他物体的上下文概率, 具体参见算法框架表 2.3。

2.4 本章小结

本章我们介绍一种新的多类物体检测算法框架，该框架将图像的局部特征和物体的上下文关系结合起来，共同辅助物体识别。

首先，在单类霍夫森林的基础上提出了多类霍夫森林，多类霍夫森林本质上一个判别式(Discriminative)模型，该模型直接将图像 Patch 映射到它的霍夫投票，这样该模型训练针对性比较强。在物体检测阶段，通过已经训练好的霍夫森林模型，使用局部特征对物体可能出现的位置进行投票，投票空间的局部极值点作为存在物体的候选点。

最后，在投票空间中我们提出了上下文关系，考虑多类物体之间的相对位置关系和几何约束，使用概率方法进行位置建模，然后通过一些近似方法，对模型的参数进行极大似然估计，利用已经学习好的模型，我们提出了一种贪心搜索策略，将上下文关系融入多类物体检测中去，进一步提高多类物体检测的效率。

第三章 实验结果与讨论

本章内容主要介绍多类物体检测的实验结果，包括：9类数据库的实验结果和 LabelMe 数据库的实验结果。同时，对实验结果进行了定性和定量分析，从中发现算法框架的优点和缺点。

3.1 引言

第二章本文主要介绍了基于局部特征和上下文多类物体检测的算法框架，本章主要介绍这个算法框架下的实现细节以及实验结果，并且通过实验结果分析这个算方法框架的优点和缺点。本次总共设计了两组实验：9类数据库实验和 LabelMe [38] 数据库实验。设计9类数据实验的目的是用于测试多类霍夫森林的鲁棒性，并验证该方法是否可以识别多类物体。LabelMe 数据库实验是本文实验的主要部分，因为 LabelMe 数据库中图片都是现实场景，我们可以同时检测多类物体，因此，该数据库可以验证基于局部特征上下文多类物体检测算法框架在真实场景中具有一定的鲁棒性。在两组实验中，霍夫森林的训练基本相同，使用的霍夫森林都是最大高度(d_{max})为20，最小节点数(N_{min})为20，使用了32个特征通道，主要由16个通道特征通过两个滤波器生成，具体16个通道特征见列表 3.1。这些特征通道既包括原始图像的绝对值，该特征对图像的纹理色彩比较敏感，又包括原始图像的一阶、二阶导数，该特征对图像的边缘和角点比较敏感。同时，还是用一些统计特征，例如梯度方向直方图，这类特征鲁棒性更强，表示的信息更丰富。

3.2 9类数据库实验结果及分析

9类数据库中，9类物体为：人脸、飞机、摩托车、正面汽车、侧面汽车、行人、牛、马和瓶子，它们都是来自以前有关物体检测论文使用的数据，具体来源参见表格 3.2，每类物体训练图像为100幅，测试图像为50幅。

图 3.1 中给出我们多类霍夫森林在9类数据上检测结果统计图，从统计的结果中可以看出，多类霍夫森林还是存在一定的鲁棒性的，绝大多数类别的正

表 3.1: 16个通道特征列表.

通道编号	使用特征
1-3	Lab颜色特征空间
4	X方向一阶导数绝对值
5	Y方向一阶导数绝对值
6	X方向二阶导数绝对值
7	Y方向二阶导数绝对值
8-16	HOG9维向量

确率(Precision)和召回率(Recall)都超过了百分之八十，其中，刚体类物体(Rigid Object)识别正确率都达到了百分九十以上，例如：正面汽车，侧面汽车，瓶子等等；形变物体(Articulated Object)识别正确率也达到了百分之八十，例如：行人，马，牛等等。本次实验中，飞机这一类的识别效果不太理想，主要是由于飞机包含了：民航机，直升机，战斗机等等，这些飞机之间的差别很大，同时飞机的形态也各异，有上升的飞机，有下降的飞机，有飞行中的飞机。图 3.2 给出了9类数据库中物体检测的几个例子，从图中可以看出，霍夫投票的中心比较集中，对于这9类物体检测比较准确。总的来说，多类霍夫森林的在物体检测方面存在一定优势，其识别效果达到了与一些高水平研究成果具有可比性。

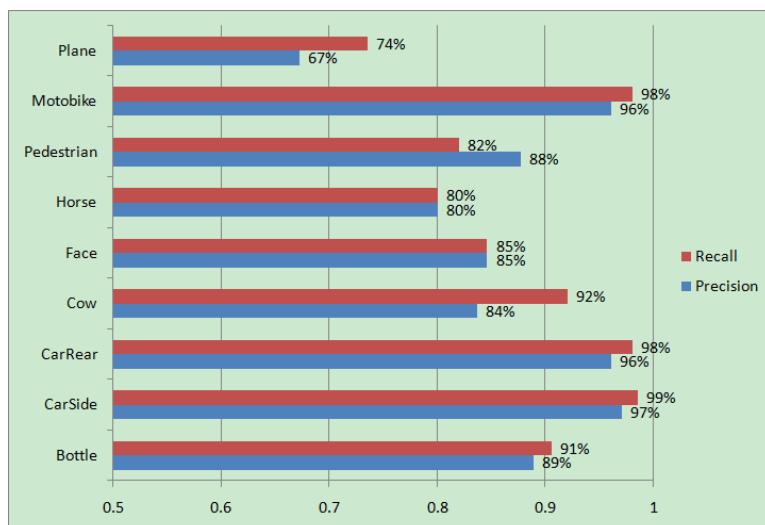
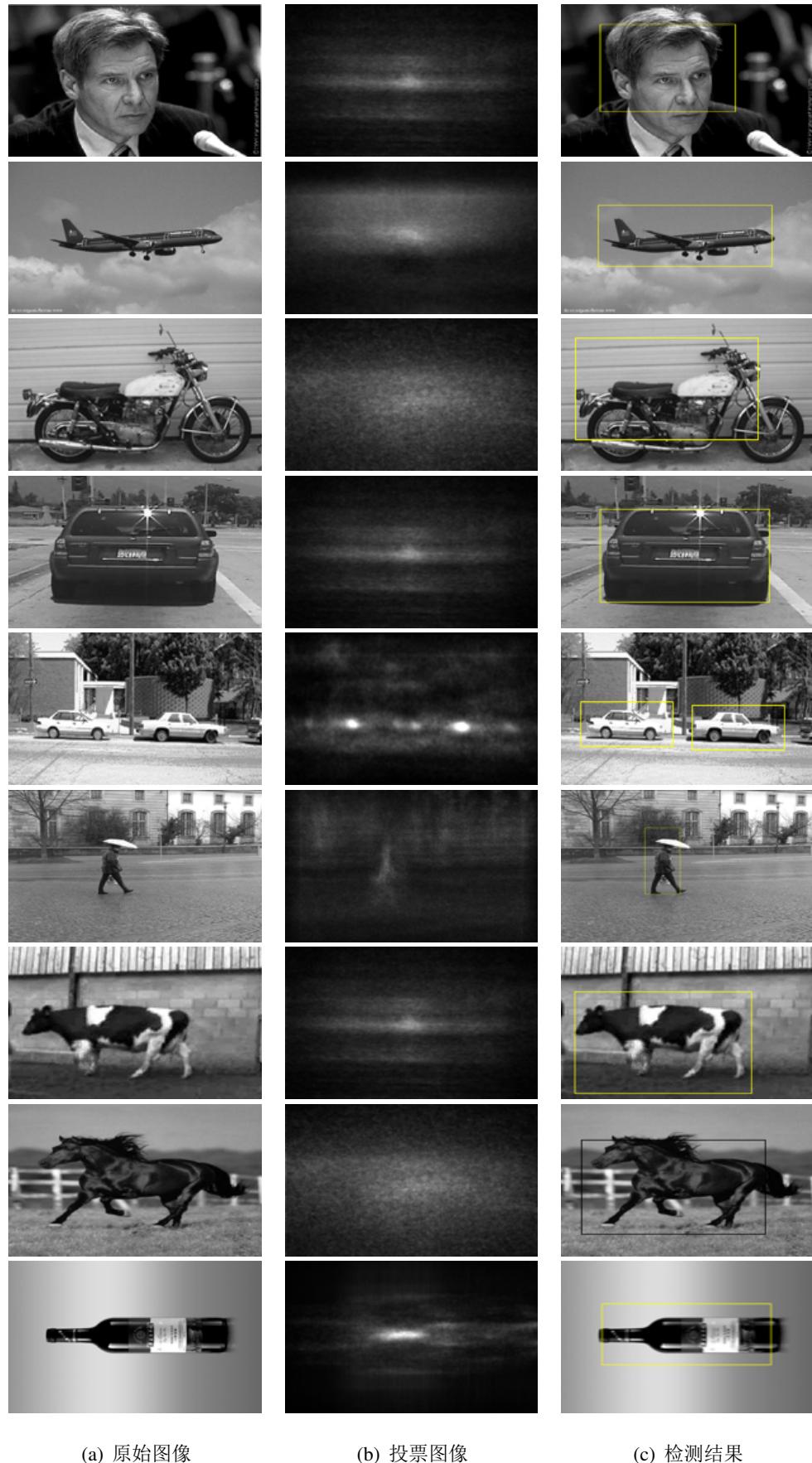


图 3.1: 9类数据库的实验结果统计图.



(a) 原始图像

(b) 投票图像

(c) 检测结果

表 3.2: 9类数据库图片来源列表.

物体类别	训练图片来源	测试图片来源
人脸	Caltech256 [22]	MIT Face Dataset [40]
飞机	Caltech256 [22]	Caltech256 [22]
摩托车	Caltech256 [22]	Caltech256 [22]
正面汽车	Caltech256 [22]	Caltech256 [22]
侧面汽车	UIUC Car [1]	UIUC Car [1]
行人	TUD Pedestrian [2]	TUD Pedestrian [2]
牛	CowSide [26]	CowSide [26]
马	Weizmann Horse [7]	Weizmann Horse [7]
瓶子	Google Image	Google Image

3.3 LabelMe 数据库实验结果及分析

LabelMe数据库实验是本文实验的主要部分。本文从LabelMe数据库 [38] 中搜集了200幅训练图片和100幅测试图片，均为室外街道场景。本文主要关注六类场景物体：行人，正面汽车，侧面汽车，窗户，马路和柱子(例如电线杆，路标排等等)。本文首先使用这些图片训练多类霍夫森林，这个和9类数据库做法类，然后统计出6类物体5种相对位置关系的频率矩阵，训练上下文模型，最后结合多类霍夫森林模型和上下文模型，进行多类物体检测。

本文主要做了两组实验，一组是没有上下文模型的多类物体检测，一组是含有上下文模型的多类物体检测，图 3.3 中，给出了我们六类结果的检测结果统计图。从实验结果中，我们可以看出多类霍夫森林在现实场景图片中，识别率虽然有所下降，但是对一些刚体物体还是一定鲁棒性，例如对窗户和道路识别率能够接近百分之八十，对侧面汽车识别率能够接近百分之七十，对于行人这种形变物体识别正确率相对较低。同时由于正面汽车好多跟侧面汽车相似，很多正面汽车被识别为侧面汽车，导致正面汽车识别率最低。然而，加入了上下文模型以后，本文利用物体之间的相互关系，我们可以避免好多错误的检测。通过图 3.3 中的对比，可以发现加入了上下文模型，物体识别的正确率几乎没有下降，同时错误检测都可以一定程度地避免，例如窗户被误检测为杆子，正

面汽车被误检测为侧面汽车。图 3.4 中给出了本文部分实现效果图，从实验效果图中证实了上述分析，例如中间几幅图像中，没有上下文模型，总是存在一些错误检测，例如行人和正面汽车的错误检测，但是加入了这些上下文模型以后，可以成功地避免这些错误，例如行人和正面汽车。

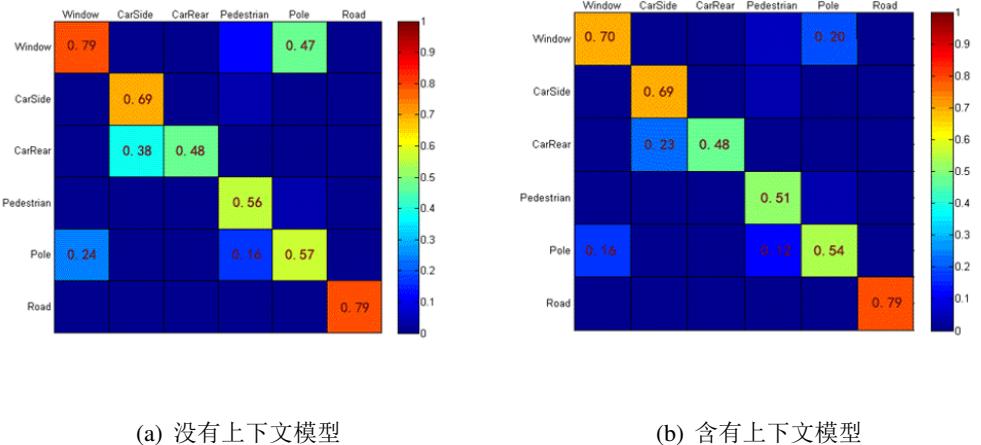


图 3.3: LabelMe 数据库多类物体检测混合矩阵.

当然，本模型还存在一定的局限性。目前上下文模型只能避免一些错误检测，然而不能增加一些新的检测，例如图 3.4 中由于部分遮挡，多类霍夫森林并没有识别出其中的汽车，然而我们的上下文模型也没能够辅助识别出其中汽车，因为我们的上下文模型主要是基于多类霍夫森林投票结果进行建模的，它很大程度上还是受限于多类霍夫森林的性能，以后我们还需要进一步考虑将上下文模型考虑融入霍夫森林投票过程中去，进一步提高我们模型的正确率。

3.4 本章小结

本章主要介绍了本文所提出的多类物体检测算法框架的实验设计及结果，实验主要分类两个部分：9类数据库实验和 LabelMe [38] 数据实验。通过实验效果，可以看出多类霍夫森林模型在多类物体识别中，具有一定的鲁棒性，特别是对一些刚性物体；此外，上下文模型可以辅助多类霍夫森林进行物体识别，进一步提高物体检测的效率，改善我们最后的实验效果。当然模型还是存在一定的局限性，需在后续工作中作进一步改进。



图 3.4: LableMe 数据库实验结果, 中间: 没有上下文模型, 右面: 含有上下文模型.
33

第四章 总结与展望

本章小结本文的工作，分析本文算法框架的优点和不足，同时对未来工作提出进一步展望。

4.1 本文总结

本文主要研究了物体识别课题，该课题是近年来计算机视觉领域比较热门的研究问题之一，在理论和应用方面都存在比较大的价值。在理论方面，物体识别技术的研究可以辅助我们进一步对人脑视觉原理的理解，从而促进人脑科学的进步；在应用方面，物体识别技术可以用于很多领域，例如：信息检索，视频监控，自动导航等等。物体识别技术还处于一个初级阶段，还需要面临各种挑战，还有许多问题有待进一步解决。

本文重点研究了物体识别中的多类物体检测问题，提出了一个基于局部特征和上下文关系的多类物体检测算法框架。本文创新点主要有两个：多类霍夫森林模型和上下文模型。多类霍夫森林模型是在单类霍夫森林基础上提出来的，该模型同时学习和检测多类物体，提高了训练和检测的效率。它对多类物体检测具有一定的鲁棒性，无论物体是刚体物体还是形变物体，多类霍夫森林检测正确率都能够达到现在方法的一般水平。同时，在多类物体检测中，物体之间的上下文关系始终是一个重要的线索。于是，本文在多类霍夫森林模型的基础上又加入了上下文模型，该模型主要考虑了多类物体之间的相互之间位置关系约束，对物体相对位置进行建模，最后使用上下文模型辅助多类霍夫森林进行物体检测，上下文模型能够避免一些错误检测，进一步提高检测的正确率。

实验部分我们总共做了两组实验来验证我们算法框架的鲁棒性。首先，我们在一个9类的数据上，单独考虑使用多类霍夫森林模型进行9类物体检测，实验结果证明我们多类霍夫森林物体检测正确率较高，特别是一些刚体物体，我们的正确率都达到百分之九十，同时对形变物体，正确率虽然有所下降，但是仍然存在达到当前平均水平。其次，为了验证我们算法在真实场景图片中的可行性，我们在数据库 LabelMe 上测试了我们算法框架，我们同时训练了霍夫森

林模型和上下文模型，实验结果证明，我们算法框架在真实场景中仍然是可行的，识别正确率能够达到百分之六十左右。

4.2 将来工作

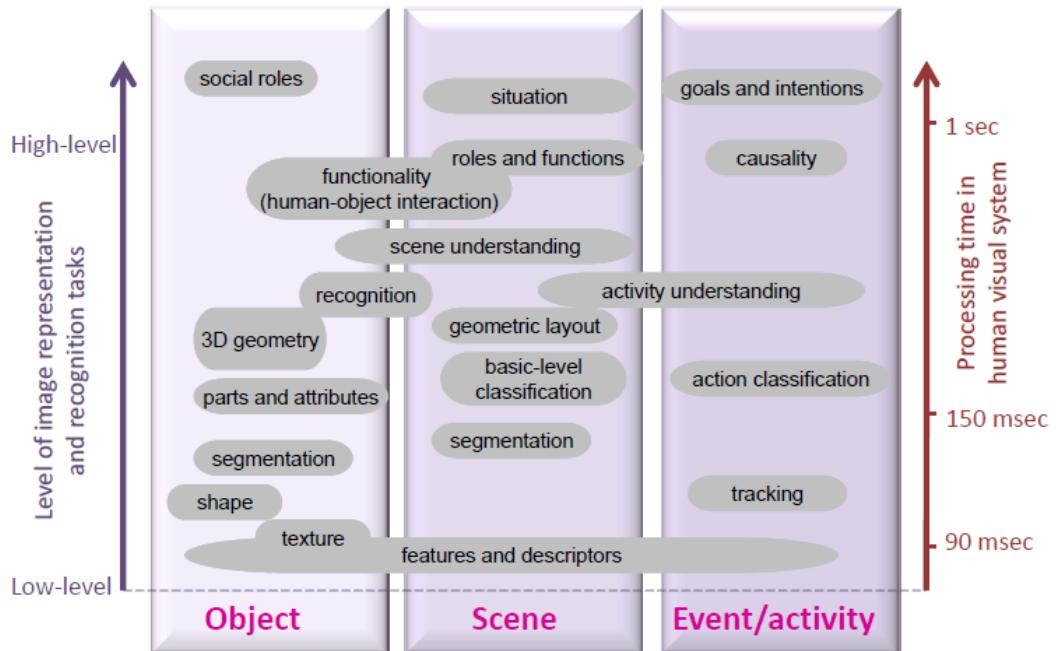


图 4.1: 计算机视觉问题关系示意图, 摘自 Fei-fei Li Slide.

本文提出算法框架还是存在一定的局限性，正如实验部分提到的，本文方法还很大程度上受限于多类霍夫森林投票的结果，如果投票结果不是很好，上下文模型也不能利用物体之间的相互关系找出未识别的物体实体。其实，还可以考虑融入各种上下文关系，不仅仅是物体层次(Object)的上下文联系，还可以是区域层次(Region)或者部分层次(Part)的上下文联系，这样就可以把这种上下文联系融入多类霍夫森林投票机制中去，考虑 Patch 对之间的投票，这样可以更加精确的定位物体的中心。同时，如果考虑了物体不同部分之间的联系，一个部分的检测结果对另外一个部分检测的产生影响，辅助一些被遮挡的物体部分的检测，这样就可以解决遮挡问题，发现更多物体实体，解决前面提出的不足之处。

本文方法还可以用来做场景理解方面的问题。物体和场景总的来说应该是不可分割的两部分，物体是场景中的元素，场景是物体的总结，物体识别和场景识别是可以相互促进。本文识别出一些特定的物体，可以辅助我们对场景类别的判断。同理，加入某种场景先验知识，就可以事先判断可能出现的物体类别和物体位置。今后，可以在我们算法框架中加入场景知识，使用把场景识别和物体识别作为一个整体问题来建模，这样应该既可以提高物体识别的准确率，又可以同时识别出场景的类别。

识别对于计算机视觉来说是一个很难的问题，它应该是计算机视觉高层问题的一个基础。识别不仅仅包括物体识别，它还包括：场景识别，行为识别，动作识别等等，具体见示意图 4.1。同时，这些不同识别之间存在相互联系，它们基于共同的基础，它们的识别结果又可以相互辅助，彼此改善识别效果。总之，这些识别都是为了一个共同目标，就是从图像中提取出符合人类认知的语义信息，让计算机模拟出人类的视觉能力，真正实现计算机的智能化。识别目前仍然处于一个初步阶段，还有许多问题亟待解决，因此，它存在十分广泛的研究前景。

参考文献

- [1] S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1475 –1490, 2004.
- [2] M. Andriluka, S. Roth, and B. Schiele. People tracking by detection and people detection by tracking. In *CVPR*, pages 1 – 8, 2008.
- [3] D. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.
- [4] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1-2):105–139, 1999.
- [5] I. Biederman. Perceiving real-world scenes. *Science*, 177(7):77 – 80, 1972.
- [6] C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.
- [7] E. Borenstein and S. Ullman. Learning to segment. In *ECCV*, pages 315–328, 2004.
- [8] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- [9] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [10] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603 –619, may 2002.
- [11] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.

- [12] C. Desai, D. Ramanan, and C. Fowlkes. Discriminative models for multi-class object layout. In *ICCV*, pages 229–236, 2009.
- [13] S. Divvala, D. Hoiem, J. Hays, A. Efros, and M. Hebert. An empirical study of context in object detection. In *CVPR*, pages 1271 –1278, june 2009.
- [14] P. F. Felzenszwalb, D. A. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In *CVPR*, pages 1–8, 2008.
- [15] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR*, pages 264–271, 2003.
- [16] M. Fink and P. Perona. Mutual boosting for contextual inference. In *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.
- [17] J. Gall and V. S. Lempitsky. Class-specific hough forests for object detection. In *CVPR*, pages 1022–1029, 2009.
- [18] C. Galleguillos and S. Belongie. Context based object categorization: A critical survey. *Computer Vision and Image Understanding*, 114(6):712 – 722, 2010.
- [19] C. Galleguillos, B. McFee, S. Belongie, and G. Lanckriet. Multi-class object localization by combining local contextual interactions. In *CVPR*, pages 113–120, 2010.
- [20] C. Galleguillos, A. Rabinovich, and S. Belongie. Object categorization using co-occurrence, location and appearance. In *CVPR*, pages 1–8, 2008.
- [21] K. Grauman and B. Leibe. *Visual Object Recognition*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers, 2011.
- [22] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report 7694, 2007.

- [23] X. He, R. Zemel, and M. Carreira-Perpinan. Multiscale conditional random fields for image labeling. In *CVPR*, volume 2, pages 695–702, june-2 july 2004.
- [24] P. Hough. Method and means for recognizing complex patterns. U.S. Patent 3.069.654, Dec. 1962.
- [25] C. H. Lampert, M. B. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *CVPR*, pages 1–8, 2008.
- [26] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, 77(1-3):259–289, 2008.
- [27] F.-F. Li, R. Fergus, and P. Perona. A bayesian approach to unsupervised one-shot learning of object categories. In *ICCV*, pages 1134–1141, 2003.
- [28] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999.
- [29] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [30] A. Oliva and A. Torralba. The role of context in object recognition. *Trends in cognitive sciences*, 11(12):520–527, 2007.
- [31] B. Ommer and J. Malik. Multi-scale object detection by clustering lines. In *ICCV*, pages 484–491, 2009.
- [32] A. Opelt, A. Pinz, and A. Zisserman. Learning an alphabet of shape and appearance for multi-class object detection. *International Journal of Computer Vision*, 80(1):16–44, 2008.
- [33] D. Opitz and R. Maclin. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11:169–198, 1999.

- [34] A. Rabinovich and S. Belongie. Scenes vs. objects: a comparative study of two approaches to context based recognition. In *CVPR, Workshops*, pages 92 –99, 2009.
- [35] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie. Objects in context. In *ICCV*, pages 1–8, 2007.
- [36] S. Richard. *Computer Vision: Algorithms and Applications*. Text in Computer Science. Springer, 2011.
- [37] B. C. Russell, A. Torralba, C. Liu, R. Fergus, and W. T. Freeman. Object recognition by scene alignment. In *Advances in Neural Information Processing Systems 19*, pages 1553–1560. MIT Press, Cambridge, MA, 2007.
- [38] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: a database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1-3):157–173, 2008.
- [39] R. E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, June 1990.
- [40] K. K. Sung and T. Poggio. Example-based learning for view based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1998.
- [41] A. Torralba. Contextual priming for object detection. *International Journal of Computer Vision*, 53(2):153–167, 2003.
- [42] A. Torralba. Contextual influences on saliency. In *Neurobiology of Attention*, pages 586 – 592. Academic Press, 2005.
- [43] A. Torralba, K. Murphy, and W. Freeman. Sharing visual features for multiclass and multiview object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):854 –869, may 2007.

- [44] A. Torralba, K. P. Murphy, and W. T. Freeman. Contextual models for object detection using boosted random fields. In *Advances in Neural Information Processing Systems 17*, pages 1401–1408. MIT Press, Cambridge, MA, 2005.
- [45] J. Verbeek and B. Triggs. Scene segmentation with crfs learned from partially labeled images. In *Advances in Neural Information Processing Systems 20*, pages 1553–1560. MIT Press, Cambridge, MA, 2008.
- [46] P. A. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [47] L. Wolf and S. Bilechi. A critical view of context. *International Journal of Computer Vision*, 69(2):251–261, 2006.

科研成果与个人荣誉

本科期间完成的学术成果

- [1] LiMin Wang, Yirui Wu, Tong Lu and Kang Chen, Multiclass Object Detection by Combining Local Appearances and Context, in *Proc. ACM Multimedia (ACM MM)* , 2011 (Top Conference, Submitted).
- [2] LiMin Wang, Yirui Wu, Zhiyuan Tian, Zailiang Sun and Tong Lu, A Novel Approach for Robust Surveillance Video Content Abstraction in *Proc. IEEE Pacific-Rim Conference on Multimedia (PCM)*, Lecture Notes in Computer Science, (EI): p348-356, 2010.
- [3] 路通, 巫义锐, 王利民, 田智源, 孙再亮. 基于监控视频内容提取车辆底盘图像的合成方法. 中国发明专利. (专利申请号20100264070.X)

本科期间参与的创新项目

1. 南京大学本科生创新项目, “基于鲁棒特征提取的多尺度景象匹配技术研究”(课题年限 2009.6~2010.9), 项目核心成员, 该项目结题被评为优秀 (全系共2项)。

本科期间获得的奖励

- | | |
|------------------|---------------------|
| 2010~2011 | Google优秀奖学金 (全系共1人) |
| 2009~2010 | 国家奖学金 (年级共3人) |
| 2009~2010 | 全国大学生数学建模比赛, 江苏省一等奖 |
| 2008~2009 | 董氏东方奖学金 (年级共3人) |

致 谢

首先，感谢我的指导老师路通副教授，从大三开始加入课题组以来，路老师在科研方面给了我很大的指导和帮助。路老师把带入了计算机视觉这个研究领域，这是一个相对比较年轻的领域，它诞生于1966年一个MIT本科生暑期项目，随着几十年的发展，这领域取得了一些进展，然而还有很多问题急需解决，特别是一些高层视觉问题。路老师给我提供了物体识别这个课题，这个课题应该算是计算机视觉领域最难的问题之一，至今这个问题到底能解决多少还没有定论。正是面临这样一个富有挑战性的问题，路老师给了我很多指导，和我一起讨论，经过很多次的想法交流，我们终于找到了一些突破口，找到一些已有方法的不足，然后提出了一些改进方法，并且最终成为了本篇毕业论文。

其次，我还要感谢巫义锐和陈康同学，在我们算法讨论阶段，他们提了很多宝贵的意见，在算法实现阶段，他们帮忙搜集数据和调节参数，他们对这篇毕业论文提供了很大的帮助。我还要感谢高荣军师兄，他给我提供了一个比较好的实验环境，感谢实验室里面所有的师兄师姐，他们给我提供了一个很好的科研氛围。

最后，我还要感谢我的父母，感谢他们23年对我的无私奉献。他们用自己的辛勤劳动，为我提供了一个读大学机会，他们在生活的方方面面为我付出了太多，在此向父母道一声谢谢。