中图分类号：TP391 论文编号：1028716 17-S001
学科分类号：080202

硕士学位论文

基于多示例学习的任意姿态人体检测

研究生姓名 蔡雅薇
学科、专业 计算机科学与技术
研 究 方 向 智能计算与机器学习
指 导 教 师 谭晓阳 教授

南京航空航天大学

研究生院 计算机科学与技术学院

二〇一七年一月

Nanjing University of Aeronautics and Astronautics

The Graduate School

College of Computer Science and Technology

Human Detection under Arbitrary Poses Based on Multiple Instances Learning

A Thesis in

Computer Science and Technology

By

Yawei Cai

Advised by

Prof. Xiaoyang Tan

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Engineering

January, 2017

承诺书

本人声明所呈交的硕士学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得南京航空航天大学或其他教育机构的学位或证书而使用过的材料。

本人授权南京航空航天大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本承诺书）

作者签名：_____

日　　期：_____

摘 要

现在大部分人体检测仅仅关注普通的直立姿态，但现实中的人体却呈现非常丰富的姿态（如弯曲的，躺着的，坐着的），我们将研究困难姿态（多视角或者任意姿态）下的弱监督人体检测问题。这不仅加大了人体检测的难度，而且令标注的工作更加困难，实际中通常只能获得弱标注样本。多示例学习方法放松了精准标注的要求，因此常常被用来解决此类问题。

我们在多示例框架下研究弱监督任意姿态人体检测问题。我们有四个重要贡献：（1）我们对多种多示例学习算法进行了实现和比较，并分析总结了所有算法的特点和适用情况，为多示例学习算法的使用提供参考；（2）我们认识到几个可能会对多示例检测性能产生重大影响的重要因素，例如，少量监督信息的帮助，对正示例比例（RoP）的较高需求等，这些因素有利于基于多示例学习的弱监督检测，但是在之前的文献中很少被研究；（3）我们第一个指出在弱监督学习环境下，当初始的示例层次检测器很弱或者训练正包中负示例数量较大时，一些常用的多示例学习打包工具（例如，Noisy-OR 和 ISR）会遭遇梯度消失的问题，从而导致对训练样本极低效率的使用。因此，在这种情况下，我们提倡使用更加鲁棒和简单的最大池规则；（4）我们提出了一种新的选择性弱监督检测算法（SWSD），并标注了一个新的大规模数据集叫做 LSP/MPII-MPHB，在这个数据集和其它著名的基准数据集上，我们证明了所提出的 SWSD 方法相对于目前最先进方法的优越性。

关键词：计算机视觉，弱监督学习，人体检测，选择性弱监督检测，多示例学习，多姿态人体数据集

ABSTRACT

Most current research on human body detection focuses only on a few common human body poses with human body in upright positions, while in the real world human bodies may exhibit very rich pose variations (e.g., when people are bending, sleeping, or sitting). The problem of weakly supervised human body detection under difficult poses (e.g., multi-view and/or arbitrary poses) is studied. This not only imposes great challenges on the task of human detection, but also makes the job of manual annotation even more difficult and usually only weak annotations are available in practice. The multi-instance learning method relaxes the requirements of accurate labeling and hence being commonly used to address the task.

In this paper we study the problem of weakly supervised human detection under arbitrary poses within the framework of multi-instance learning (MIL). Our contributions are four folds: (1) We carry out the implementation and comparison of many MIL algorithms, and analyze and summarize the characteristics and the applicable conditions of all algorithms to provide a reference for the use of MIL algorithms; (2) we identify several crucial factors that may significantly influence the performance, such as the usefulness of a small amount of supervision information, the need of relatively higher RoP (Ratio of Positive Instances), and so on - these factors are shown to benefit the MIL-based weakly supervised detector but are less studied in the previous literature; (3) we first show that in the context of weakly supervised learning, some commonly used bagging tools in MIL such as the Noisy-OR model or the ISR model, tend to suffer from the problem of gradient magnitude reduction when the initial instance-level detector is weak and/or when there exist large number of negative proposals, resulting in extremely inefficient use of training examples. We hence advocate the use of more robust and simple Max-Pooling rule under such circumstances; (4) we propose a new selective weakly supervised detection (SWSD) algorithm. We also annotate a new large-scale data set called LSP/MPII-MPHB, on which and another popular benchmark dataset we demonstrate the superiority of the proposed method compared to several previous state-of-the-art methods.

Keywords: Computer vision, Weakly supervised learning, Human detection, Selective weakly supervised detection (SWSD), Multi-instance learning (MIL), Multiple poses human body dataset

目 录

图表清单

注释表

D	多示例样本集合	r	在提议中选择的示例比例
d_{jk}	两个示例之间的距离	S	最高密度子图
G	多示例包构成的图	S_t	第 t 次迭代的训练集
g	子图个数	T	算法迭代次数
J	包中示例的个数	t_i	第 i 个训练包的标记
k	选择最高排名示例的比例	w_{jk}	两个示例之间的权重
l	搜索区域	x_i	第 i 个训练包
M	监督样本的个数	x_{ij}	第 i 个训练包的第 j 个示例
N	多示例训练包的个数	x_{ijk}	示例 x_{ij} 的第 k 个维度
n	选择最高排名示例的数量	y_i	包 x_i 的预测标记
p_i	包为正的概率	y_{ij}	示例 x_{ij} 的预测标记
p_{ij}	示例为正的概率	λ_1	监督信息的贡献程度
p_{ij}^{Ada}	MILBoost 的预测结果	λ_2	正则化系数
p_{ij}^{LR}	MILLR 的预测结果	η	搜索区域的收缩比率

缩略词

缩略词	英文全称
CNN	Convolutional Neural Network
R-CNN	Regions with CNN features
WSL	Weakly Supervised Learning
MIL	Multiple Instance Learning
ISR	Integrated Segmentation and Recognition
VOC	Visual Object Classes
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LSP	Leeds Sports Pose
MPHB	Multiple Pose Human Body
SWSD	Selective Weakly Supervised Detection
RoP	Ratio of Positive
APR	Axis-Parallel Rectangles
DD	Diverse Density
EM-DD	Expectation Maximization Diverse Density
SVM	Support Vector Machine
KNN	K-Nearest Neighbor
DPM	Deformable Parts Model
MCG	Multi-scale Combinatorial Grouping
HOG	Histogram of Oriented Gradient
SIFT	Scale Invariant Feature Transform
LBP	Local Binary Pattern
HSV	Hue Saturation Value
BPNN	Back Propagation Neural Networks
LR	Logistic Regression
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
IoU	Intersection-over-Union
AP	Average Precision
IoP	Instance-over-Proposals

缩略词	英文全称
PRLS	Posterior Regularized Latent SVM
MMIL	Multi-fold MIL
FMP	Flexible Mixtures-of-Parts
mAP	mean AP

第一章 绪论

1.1 引言

在图像和视频分析领域，人类是视觉对象中最常见的类型，而且在执法监测、医疗、娱乐等的访问控制中都有着广泛的相关应用，因此对图像和视频中人类的处理至关重要。检测人体通常是为更加深度和更高层次的人类分析奠定基础。至今有一个相关课题被研究的最多，那就是行人检测[1][2][3]，它是城市智能交通系统的重要组成部分。但是除了在一些特殊情况下的成功，行人检测主要关注直立姿态的人体，一般而言，现实中的人类往往会呈现多样的姿态，例如弯曲，坐着，躺着或者其它类型（见图 1.1 的说明，图片来自 MPII Human Pose 数据集[4]，边界框的标注来自我们的工作），这一问题强调了能够检测任意姿态人体的必要性。

图 1.1 不同姿态的人体说明

然而，任意姿态人体检测面临着多方面的挑战，主要来自于人体的外形变化，呈现出不同的姿态，光照变化，背景杂乱，遮挡等带来的困难[5]。为了解决这些问题，最先进的对象检测器，例如 R-CNN[6]和基于它的变形算法，挖掘了深度神经网络强大的建模能力，它们能够从大量的全监督样本中学习不变的特征集合。但是，想要获得如此大量的标记样本，手工标注的过程一般非常困难和费力。因此人们希望以最小的监督去定位对象，从本质上释放数据收集和标注的负担。这激励了所谓的弱监督学习（WSL）用于对象检测，它的目标是只依赖于仅有的弱监督信息（即这张图片中包含人类），而不是任何更深层次具体的监督信息，例如指明图中对象的数量和它们各自的位置。请注意这个设置不同于半监督学习算法，在半监督环境中，样本的标记部分是完全没有的，部分是完整的，而我们的弱监督学习的样本标记全部是不完整的。

1.2 多示例学习

1.2.1 问题的提出

Dietterich 等人在 1997 年研究了药物活性预测问题[7]。其目标是构建一个学习系统，通过对已知分子进行学习，这些分子可能适合制药，也可能不适合制药，然后尽可能准确地预测其它的分子是否适合制药。目前大家只能知道哪些分子适合制药，但由于每个分子都有很多种同分异构体，大家不知道其中的哪一种同分异构体起到了决定性作用。假如将适宜制药的分子的全

部同分异构体都作为正样本，像传统的全监督学习那样进行训练就会引入大量噪声。因此，人们提出了多示例学习问题[8]。

1.2.2 多示例学习的概念

在多示例学习问题中，数据集中的每个数据是一个包。而每个包中包含了若干个示例，我们可以知道包的标记，但无法知道包中示例的标记。若某个包的标记为正，则这个包中至少有一个示例为正；反之，若包的标记为负，则这个包中所有示例的标记均为负[9]。在本文中，我用 x_i 代表第 i 个包，x_{ij} 代表第 i 个包中第 j 个示例的特征向量，x_{ijk} 代表第个 i 包中第 j 个示例特征向量的第 k 个维度的值，t_i 表示第 i 个包的标记，y_i 表示第 i 个包的预测标记，y_{ij} 表示第 i 个包中第 j 个示例的预测标记。多示例学习的目标就是，通过对这些已标记的包进行学习，然后判断其它新的包和示例的标记。图 1.2 说明了多示例学习问题。

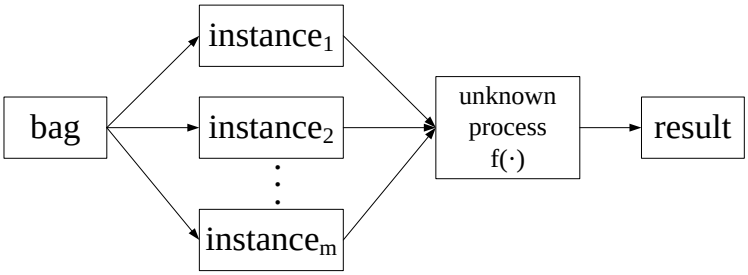

图 1.2 多示例学习的框架

1.3 本文的贡献

多示例学习（MIL）[10]的出现为弱监督检测带来了一个有希望的框架，它本质上释放了对示例标记的要求，并将包层次的弱监督信号后向传播。最近，Cinbis 等人[11]将多示例学习用于一般的对象检测中，并在 Pascal VOC 2007 数据集[12]上取得不错的检测效果，彰显了多示例学习方法的潜力。但是他们没有重点关注多姿态人体检测的问题以及多示例学习中的实践细节，例如他们的样本选择和参数设置至今仍未清楚。更重要的是，由于缺少强监督信息，之前的多示例学习方法不得不在每张图片上收集大量的示例以满足多示例学习的定义，即在正包中至少有一个正示例出现。不幸的是，这种做法不仅大大加重了计算代价，而且使得一些传统的多示例学习打包模型，打包模型是指融合多个独立的示例层次条件概率到一个包层次条件概率的方法，例如 Noisy-OR 规则和 ISR 规则[13]在训练中变得无效。我们第一个发现在正包中出现过多的负示例会在一定程度上损害学习性能，因为这会严重缩小梯度的量级，尤其在初始模型很弱的情况下。为了解决这个问题，我们提议使用更加鲁棒的打包模型，例如最大池规则。

对于对象检测，已经存在很多的公开数据集供研究者们使用，例如 Pascal VOC[14]和 ILSVRC[15]。但是，这些数据集都包含了很多类别的对象，在这些类别中并不着重强调人类这一类别。实际上，现在还不存在一个大规模的数据集适用于人体检测。幸运的是，已经存在了

很多用于姿态估计的数据集，例如 LSP 和 MPII Human Pose，在它们的基础上，我们标注了一个新的数据集叫做 LSP/MPII-MPHB，用于任意姿态人体检测这个特定任务。我们从 LSP[16]和 MPII Human Pose[4]数据集种选择了 26000 多张具有挑战性的图片，并且为每张被选择的图片标注了人体的边界框。

总之，本文的贡献主要在四个方面：首先，我们对多种多示例学习算法进行了实现，在多示例分类问题中比较了它们的性能差异，并且全面分析总结了每个算法优点，缺点和适用情况，为我们在弱监督检测中使用多示例学习算法提供铺垫。其次，我们发现了几个可能对性能产生重大影响的关键因素，例如少量监督信息的有效性，相对较高 RoP（正示例比例）的必要性等。这些因素有利于基于多示例学习的弱监督检测，但却在之前的文献中很少被研究。接着，我们第一个认识到在多示例学习中，Noisy-OR 和 ISR 模型容易被忽略的缺陷，它类似于神经网络中的梯度消失问题，严重削弱了多示例算法的训练效率。最重要的是，我们提出了一个新的选择性弱监督检测算法（SWSD）用于任意姿态下的人体检测，它采用了一种新的约束精英选择方法来选择排名最高的提议用于下一轮的检测器训练。这不仅能够防止模型训练过早的锁定在错误的位置，而且可以有效地提高学习的稳定性。最后，在我们最新标注得到的数据集 LSP/MPII-MPHB 和著名的 Pascal VOC 数据集上，我们证明了我们的 SWSD 方法比目前几个最好的弱监督对象检测方法更加有效。

1.4 本文的安排

本文剩下的安排是，在第二章简短的介绍了弱监督检测的相关工作之后，我们在第三章详细介绍多种多示例学习算法并进行了比较分析，在第四章呈现我们关于多示例检测深度评估的实验，探究会对多示例学习产生影响的因素。在第五章，我们解释了 Noisy-OR 和 ISR 模型中的缺陷，并在实验中证明了我们的发现。在第六章我们阐述了我们提出的 SWSD 算法以及对我们的 LSP/MPII-MPHB 数据集进行了具体描述，并通过实验从多个角度详细论证我们方法的有效性。最后在第七章，我们总结了全文。

第二章 多示例检测的研究现状

这一章我们将介绍多示例学习以及弱监督检测的研究现状，对于多示例学习，将说明算法的发展以及近两年在多个领域的应用，对于弱监督检测，将着重阐述其初始化和迭代方法的发展历程，并简单介绍我们所提出的方法与前人工作的不同点。另外我们还会介绍检测技术路线中使用的相关技术的研究现状，这些技术对于检测性能的提高至关重要。

2.1 多示例学习的发展

人们提出了多种用于多示例问题的算法。最先提出的解决多示例学习问题的方法是 1997 年 Dietterich 等人提出的 APR 算法[17]。接着，在 1998 年 Maron 等人提出了多样性密度（DD）框架用于解决多示例问题[18]。在 2001 年 Zhang 等人提出了 EM-DD 算法，它结合了期望最大化方法和多样性密度算法[19]。另外，多示例学习算法还可以通过对传统监督算法的改造来实现，例如 SVM，KNN，神经网络，逻辑回归，AdaBoost 等。在 2002 年 Andrews 等人提出了两种修改的 SVM 算法应用于多示例学习：mi-SVM 和 MI-SVM [20]，Wang 等人采用了豪斯道夫距离，扩展了最近邻算法（KNN）[21]。自从多示例学习被提出之后，在很多领域得到了广泛应用。其中有 Maron 和 Ratan 在 1998 年[22]和 Zhang 等人在 2002 年[23]研究的基于内容的图片检索，Maron 在 1998 年做出的股票预测[24]，Tong 等人在 2002 年用多示例学习算法进行了文本分类[25]，以及 Tao 等人在 2004 年研究了蛋白质家族建模等[26]。

多示例学习自提出已经过了二十年，近两年来依旧在理论和应用上都得到了不断发展，说明了多示例学习的无限潜力。Pathak 等人提出了全连接多类别的多示例学习[27]。Wu 等人将深度多示例学习用于图片分类和自动标注过程中[28]。Li 等人基于多示例学习对网络恐怖图片进行了识别[29]。Song 等人介绍了包差异多示例学习[30]。Melendez 等人提出将多示例学习与主动学习相结合用于计算机辅助的结核病的检测[31][32]。Vanwinckelen 等人对比了多示例学习中包层次和示例层次的准确度[33]。Papandreou 等人在深度学习的全局和局部变形中使用了多示例学习技术[34]。Li 等人介绍了通过高级示例来构建软的多示例包[35]。Chen 等人在脑小血管疾病的识别中应用了多示例学习[36]。Cano 等人提出使用 GPU 加速多示例学习的分类规则[37]。Bandyopadhyay 等人使用多示例学习预测微小目标中特定功能的集合位点[38]。Xu 等人通过结合多示例学习和 Fisher 信息实现了鲁棒的视觉追踪[39]。Zhang 等人提出了用于协同显著性检测的自学式多示例学习框架[40]。Wang 等人提出了基于在线加权的多示例学习的半监督视觉追踪方法[41]。Ruizmuñoz 等人基于多示例学习使用无监督记录分割实现鸟鸣的分类[42]。Rastegari 等人研究了多示例学习示例层次的区别力和一致相似性[43]。Shrivastava 等人介绍了多示例学习的广义字典[44]。Cheplygina 提出了基于差异的多示例学习[45]。Ren 等人使用多示例学习和包分裂方法进行弱监督大型对象定位[46]。Maken 等人将多示例学习用于乳腺癌磁共振成像[47]。

2.2 弱监督检测的发展

由于弱监督检测的发展时间较短，任意姿态的人体检测又很少被人们关注，所有现在只存在较少的致力于任意姿态人体检测的相关工作。但在普通对象检测领域中，一个最有影响力的方法是 Girshick 等人在 2015 年提出的 R-CNN 方法[6]。他们使用 Selective Search 算法生成对象检测提议，然后对每个提议提取 CNN 特征，最后学习感兴趣的对象模型。这个方法有效地避

开了滑动窗口耗时的扫描，而且利用了深度学习，在很多对象检测任务中实现了最先进的结果。然而，为了防止它复杂的似然模型过拟合，需要大量的全监督训练样本。这些全监督学习都建立在大量的精确注释和时间代价的基础上，因为在卷积神经网络 CNN 中存在太多的模型参数需要训练调整。我们的方法在某种意义上可以理解为 R-CNN 的一个弱监督学习版本，在 R-CNN 的后期阶段，我们用一个修改的多示例学习方法（SWSD）来学习对象检测模型。

可变形部件模型（DPM）是另一个用于人体检测的著名的全监督模型[48]。这个模型并不基于 CNN，因此它所需要的样本数量没有 R-CNN 那么多，但是它必须依靠不是很少的监督信息才能完成模型训练。

为了减轻对象层次手工标注的压力，人们希望以最少的监督信息定位对象，所以弱监督对象检测在提出之后得到了不断的发展。2011 年 Pandey 和 Lazebnik[49]提出使用 DPM 结合隐 SVM 方法训练一个用于弱监督学习的普通框架，他们在弱监督对象定位和场景识别中使用了这个模型。同年，Siva 和 Xiang[50] 在弱监督学习框架的基础上提出了一种新的初始化标记模型来启动检测器的迭代学习。2012 年 Russakovsky 等人[51]提出了一个对象中心空间池方法，从图片层次标记推断感兴趣的对象。

初始化对于很多对象检测都相当重要。在 2014 年 Song 等人[52]提出了一种基于图的初始化方法，定义为在候选窗口相似性图上差异子模块覆盖问题。他们也扩展了这个方法用于自动识别视觉图像的区别性结构[53]，在他们的方法中，通过搜索频繁部件结构和它们的边界框来确定初始化候选对象窗口。相反地，在我们的工作中，采用了一种使用极少量标记样本初始化对象检测器的监督方法。

由于强监督信息的稀缺，一种常见的处理方式是直接将这些信息视为不可见的但却对目标具有重要影响的隐变量。因此，迭代过程对于很多弱监督对象检测方法十分必要，它能够提高基于当前假设的隐变量的训练模型的性能。例如，在 2014 年 Wang 等人[54]提出的潜在类别学习方法，他们首先使用了潜在概率语义分析来学习潜在类别，然后他们采用类别选择方法评估每个类别的区别力。

在迭代过程中，阻止训练过早锁定在错误的对象位置上十分重要。2014 年 Bilen 等人[55]针对这个问题提出了两种启发式方法。第一个是迫使模型对正训练窗口和它们的水平镜像图像分配一个相似性得分。另一个是惩罚那些对于同一个候选窗口在多个类别上都得到了高分类得分的模型。此外，2015 年在 Cinbis 等人的工作中，提出了一个多重多示例学习方法，它将对象检测和定位融合到一个模型中，他们对正训练图片进行迭代训练并指出其中对象位置。他们将训练样本分成多组，当要定位一组样本中的对象位置时，采用剩下的其它样本训练检测模型，而不是像传统的多示例学习方法，用自己确定自己中的正示例。我们的方法和这个方法最大的区别在于，我们依旧采用了传统的多示例学习方式，但是和传统方法不同的是，我们采用了一种

新提出的约束精英选择方法来驱动学习过程，每一次迭代的过程中训练包中的示例组成都会不同，从而增加了样本的多样性。而且我们使用了极少量的真实数据作为正则项。并且，在学习过程中，我们利用了最具可能性的候选窗口周围的信息，即使他们包含了背景，取代了仅仅依赖于排名最高的候选窗口的估计方法。

2.3 检测技术的发展

一般检测问题的技术路线，主要包括四个部分，即检测提议的生成，特征表示，模型训练，以及后期窗口处理过程。

虽然滑动窗口[56][57][58]是最常用的检测提议（候选窗口）生成方法之一。滑动窗口始终要在整张图片上密集扫描，最后产生的候选窗口的数量可想而知非常庞大，需要花费大量的时间来一一处理这些窗口。在人们意识到滑动窗口的不足后，开始寻找新的方法提供候选窗口。人们假设感兴趣的对象都拥有相同的视觉特征，利用这些特征能够将它们从背景中分离出来。由于生成的检测提议比滑动窗口数量少很多，更加适用于复杂的算法，在我们的工作中，采用选择性搜索（Selective Search）算法[59][60][61]。它能产生很多的稀疏提议，但不会损失图片中的主要信息。Selective Search 算法结合了枚举搜索和图像分割的优点，利用图像结构来引导采样过程，但不放过任何可能包含对象的位置。不同于其它采用单一分离技术的算法，Selective Search 会根据颜色，纹理，结构，比例等不同的策略将对象从背景中分割出来，因此它能产生一组少量的高质量提议，适用于更加强大的分类器，而不会降低效率。在文献中还提到了很多其它的方法，例如 MCG[62]，Objectness[63]，随机种子[64]，它们都可以用于这个目的。

如何获取最具区别力的特征一直是模式识别领域的重要问题之一。已经有很多不错的特征表示方法供人们选择，例如 HOG 特征[65]，SIFT 特征[66]，LBP[67]，HSV 特征[68]等。但深度卷积神经网络[69]因为它在实际中的卓越性能，成为目前最受欢迎的特征表示方法，通过 CNN，很多解决计算机视觉问题的相关技术都得到了飞速发展。在 CNN 中，图像的一个小部分（局部感受区域）作为层级结构的最低层的输入，信息再依次传输到不同的层，每层通过一个数字滤波器去获得观测数据的最显著的特征。这个方法能够获取对平移、缩放和旋转不变的观测数据的显著特征。因为其突出的优越性，OverFeat[70]，AlexNet[71]，GoogleNet[72]，VGG[73]等已经训练好的网络被广泛应用于对象检测中。在我们的工作中，选择 VGG 网络，这是一个深度网络，它最初是在 ImageNet[74]数据集上训练得到。我们使用倒数第二层的输出作为最终的特征表示。也就意味着，每个提议都会映射成为一个 1000 维的向量，它将作为多示例学习模型的输入。注意到我们并没有在原有 VGG 网络的基础上进行进一步的微调，因为从 ImageNet 数据集上学到的特征空间足够描述人体图片。

人们提出了很多方法用于提高最终预测窗口的准确率，例如边缘调整法[11]和边界框回归法[6]。边缘调整法是基于窗口附近的像素信息，搜索对象的轮廓，确定合适的窗口位置。边界框

回归法则是建立预测窗口与真实窗口之间的回归方程，从而确定两者在位置和大小之间的关系，这也是一种全监督方法。本文中，我们使用一种新的高密度子图搜索算法，根据高密度子图中的候选窗口调整最终的预测窗口。

第三章 多示例学习算法

作为弱监督检测算法的框架，多示例学习算法值得我们关注。多示例学习算法有很多，有的是专门为多示例问题提出的，例如 APR，DD，EM-DD，有的则是从传统的全监督学习算法改造而来，例如 LR，AdaBoost。对于同一个全监督算法，改造的方式有很多种，我们只介绍其中的一部分。我们将分别介绍每个算法各自的特点，并且通过实验比较各个算法的准确度，从而深刻了解多示例学习问题的本质，为弱监督检测算法的设计奠定基础。

3.1 算法概述

表 3.1 简单介绍了每个多示例学习算法的特点，每一种算法的目标和内在原理都不一样。有些算法专注于找出训练正包中的正示例，然后根据这些正示例训练模型，例如 APR，EM-DD，mi-SVM，MI-SVM，BPNN，这些算法都涉及两层的迭代，外层的每一次迭代都是根据当前的示例标记重新训练一个新的模型，内层迭代则是根据误差调整模型的过程。有些算法则是直接根据当前模型所求得的训练正包为正的概率，采用梯度下降等方法不断调整当前的模型，例如 DD，LR，AdaBoost。有些算法则是根据训练样本直接对测试样本进行预测，没有训练模型，例如 KNN。

表 3.1 各种算法概述

算法	概述
APR	APR 算法最早提出的多示例学习算法，算法过程非常复杂，内到外算法主要包括构建最小轴平行矩形和特征选择两个过程。
DD	DD 算法以找到正示例的交点为目标，构建了一个学习框架，根据新示例与所求交点的距离衡量示例为正的可能性。
EM-DD	EM-DD 算法简化了 DD 算法的运算，在寻找交点的过程中，不再是训练集中的所有示例都参与计算，而是在包中选择一个与初始交点最近的示例替代整个包，迭代收敛条件为前后两次的训练差值小于某个可容忍的误差。
mi-SVM	mi-SVM 算法将全监督学习的 SVM 算法用于多示例问题，增加了正包中至少有一个示例为正，负包中所有示例为负的限制条件，所求的是最大示例边界间隔，迭代收敛条件为训练集中所有示例的标记不再变化。

表 3.1（续）各种算法概述

算法	概述
MI-SVM	不同于 mi-SVM 算法的是，MI-SVM 算法在每个正包中选择一个为正可能性最大的示例代表整个包，以包为数据点寻找分界面，所求的是最大包边界间隔，迭代收敛条件为代表正包的示例不再发生改变。
KNN	KNN 算法是惰性学习算法，使用豪斯道夫距离定义多示例问题中包与包之间的距离，在传统的 KNN 算法的基础上添加了引用的概念。
BPNN	BP 神经网络是非线性函数，在保证训练集每个正包中至少有一个示例的标记为正，每个负包中所有示例标记为负的基础上构建 BP 神经网络，迭代收敛条件为所有示例的标记不再变化。
LR	逻辑回归也是一种线性回归，引入逻辑方程来做归一化，使对示例的计算结果类似于概率，通过求损失函数的最小值来确定最优的分界面。
AdaBoost	AdaBoost 也是一种迭代算法，通过改变样本的分布训练出不同的弱分类器，根据分类器的训练误差定义分类器的权重，最终结果为各个弱分类器预测值的加权平均，迭代次数为所设置的弱分类器个数或者达到训练误差的要求。

3.2 轴平行矩形（APR）

3.2.1 算法简介

APR 算法是最早提出的多示例学习算法，当时 Dietterich 等人主要考虑了 3 种方法：

（1）容忍噪声的标准算法：简单的 APR 算法仅仅产生一个可以作为正示例边界的最小轴平行矩形。这是人们探索出的一种忽略多示例问题，容忍噪声的算法版本。

（2）外到内算法：此算法是标准 APR 算法的一个变形，首先构建一个作为正示例边界的最小轴平行矩形，然后排除假正性示例来缩小矩形。缩小的过程就作用于多示例问题。

（3）内到外算法：此算法开始于特征空间的一个种子点，然后寻找一个矩形，这个矩形要能够覆盖每个正包中至少一个示例，并且不能包括任何负包中的示例，以找到最小的这样一个矩形为目标，让种子点生长成为一个矩形。

他们的实验结果显示内到外算法优于其他两种算法，因为内到外算法识别相关特征的能力比外到内算法更好。而标准算法是最差的，恰恰说明了在设计算法的时候多示例问题是不能被忽略的。

3.2.2 内到外的多示例 APR 算法

如果我们假设第 k 个维度的下边界 lb_k 和上边界 ub_k 分别为：

$$lb_k = \min_{i \in positive} [\max_j x_{ijk}^+]$$
$$ub_k = \max_{i \in positive} [\min_j x_{ijk}^+] \quad (1)$$

那么，我们能保证在第k个维度上每个正包中至少有一个示例在边界以内。这个边界称为k维上的极小极大边界。在所有维度上构造极小极大边界就可以得到极小极大轴平行矩形，但这个矩形可能不包含任何正包，这是因为在不同的维度中选择了包中的不同示例。

然而，我们发现了如下定律：

任何一个轴平行矩形，如果它覆盖了每个正包中至少一个示例，那么它必然包含了极小极大轴平行矩形。

内到外轴平行矩形算法的目标是找到一个最小轴平行举行，它可以覆盖每个正包中至少一个正示例。我们用每个维度上边界宽度之和来定义轴平行矩形的大小。

$$Size(APR) = \sum_k ub_k - lb_k \quad (2)$$

算法过程：

1. 选择一个在欧几里得距离上与先前定义的极小极大轴平行矩形最近的正示例作为初始化轴平行矩形；

2. 在第l步中，$l = 2,3,...,n$，从没有被覆盖的正包中选出一个与轴平行矩形距离最近的正示例I_l加入轴平行矩形，使得轴平行矩形的大小增加的最少；

3. 回访先前每一个决定$I_1,...,I_l$，当先前的决定x_t被访问时，$t = 1,2,...,l$，算法构造覆盖$\{I_1,...,I_{t-1},I_{t+1},...,I_l\}$的轴平行矩形，把这个轴平行矩形记作$A^t$，从原$X_t$所在包中找出使$A^t$增大最少的示例记为$I_t$，取代原来的示例加入到轴平行矩形中；

4. 若所有正包都有一个示例在轴平行矩形中，即所有正包被覆盖，则算法终止，否则转到步骤 2。

3.2.3 特征选择算法

内到外算法在所有特征维度上构建了一个轴平行矩形的边界，也可以用来选择出具有识别力的特征。在维度k上，如果一个负示例与轴平行矩形边界的距离d大于设置的间隔，或者这个负示例在其它维度上与边界的距离都小于d，那么第k维的特征就具有识别力，这样的负示例个数越多，那么第k维特征的识别力就越强。

算法过程：

1. 初始化所有维度上的特征均不被选择；

2. 遍历所有的负示例，计算在第k维上，与轴平行矩形边界的距离d_k大于设置的间隔，或者在其它维度上与边界的距离都小于d_k的负示例个数C_k；

3. 选择C_k最大的维度k的特征被选择；

4. 更新负示例，如果在第 k 维上与轴平行矩形边界的距离 d_k 大于设置的间隔，或者在其它维度上与边界的距离都小于 d_k，那么这个负示例被排除；

5. 如果所有的特征被选择或者所有的负示例被排除，转到步骤 6，否则转到步骤 2；

6. 如果所有的特征被选择，算法结束，否则将所有被选择的特征作为新的特征范围，用内到外算法得到一个新的轴平行矩形，然后转到步骤 1。

预测方法：

通过内到外的多示例轴平行矩形算法和特征选择过程，最终得到一个所求的轴平行矩形。可以用这个轴平行矩形对未知标记的包进行预测。如果示例在轴平行矩形之内，那么这个示例的预测标记就为正，否则为负。

3.3 多样性密度（DD）

3.3.1 算法简介

多样性密度是对正包的交集减去负包的并集的测量。通过最大化多样性密度，我们可以找到正示例的交点和引起交集最大化的特征权重集。如果一个分子被标记为正包，那么在它的所有同分异构体中，至少有一种构造呈现了正确的形状适合于目标蛋白质。如果一个分子被标为负包，这意味着没有一种构造可以与目标蛋白绑定。如果假设只有一种形状能与目标蛋白绑定，那么它的位置就在特征空间中所有正包的交集处，同时不与任何一个负包相交。

多样性密度的定义，是在一个点处对拥有示例靠近这个点的正包的数量和负示例远离这个点的程度的测量。靠近这个点的正包的数量越多，负示例远离这个点的程度越大，这个点的多样性密度就越大。

3.3.2 多示例学习的 DD 算法

根据假设，所求的为一个单一的交点 X_t，可以对空间上的所有点 x 最大化 $\mathrm{P}(x = X_t \mid x_1, ..., x_N)$。根据贝叶斯法则，它就等价于最大化 $\mathrm{P}(x_1, ..., x_N \mid x = X_t)$。假设包与包之间相互独立，那么最大化 $\mathrm{P}(x_1, ..., x_N \mid x = X_t)$ 就相当于求：

$$\arg\max_{x_{ij}} \prod_{i \in pos} \mathrm{P}(x_i \mid x = X_t) \prod_{i \in neg} \mathrm{P}(x_i \mid x = X_t)$$

再使用贝叶斯法则，于是等价于求

$$\arg\max_{x_{ij}} \prod_{i \in pos} \mathrm{P}(x = X_t \mid x_i) \prod_{i \in neg} \mathrm{P}(x = X_t \mid x_i) \tag{3}$$

利用 Noisy-OR 模型，能够得到：

$$\mathrm{P}(x = X_t \mid x_i^+) = \mathrm{P}(x = X_t \mid x_{i1}, x_{i2}, ...) = 1 - \prod_j (1 - \mathrm{P}(x = X_t \mid x_{ij})) \tag{4}$$

$$\mathrm{P}(x = X_t \mid x_i^-) = \mathrm{P}(x = X_t \mid x_{i1}, x_{i2}, ...) = \prod_j (1 - \mathrm{P}(x = X_t \mid x_{ij})) \tag{5}$$

把单个示例作用于目标蛋白质的概率模型化为与它们之间的距离有关：

$$\mathrm{P}(x = X_t \mid x_{ij}) = \exp(-\| x_{ij} - x \|^2) \tag{6}$$

$$\| x_{ij} - x \|^2 = \frac{1}{k}\sum_k s_k^2 (x_{ijk} - x_k)^2 \tag{7}$$

直观地，如果正包中的一个示例靠近 x，那么 $\mathrm{P}(x = X_t \mid x_i^+)$ 就会比较大。如果每一个正包都有示例靠近 x，而且没有负包靠近 x，那么 x 就具有高多样性密度。

算法过程：

1. 随机选择一个正包中的一个示例作为初始点 x；

2. 计算所有示例的 $\mathrm{P}(x = X_t \mid x_{ij})$；

3. 计算所有包的 $\mathrm{P}(x = X_{ij} \mid x_i)$；

4. $f(x,s) = -\ln(\prod_{i \in pos} \mathrm{P}(x = X_t \mid x_i^+) \prod_{i \in neg} \mathrm{P}(x = X_t \mid x_i^-))$；

5. 用梯度下降法求 $f(x,s)$ 的最小值，$f(x,s)$ 取最小值的 x 和 s，即为目标交点和缩放比例。

其中：

$$\frac{\partial f}{\partial x} = \sum_{i \in pos} \left(\frac{1 - \mathrm{P}(x = X_t \mid x_i^+)}{\mathrm{P}(x = X_t \mid x_i^+)} \sum_j \frac{\mathrm{P}(x = X_t \mid x_{ij})}{1 - \mathrm{P}(x = X_t \mid x_{ij})} \frac{2}{k} \sum_k s_k^2 (x_{ijk} - x_k)\right)$$

$$- \sum_{i \in neg} \left(\sum_j \frac{\mathrm{P}(x = X_t \mid x_{ij})}{1 - \mathrm{P}(x = X_t \mid x_{ij})} \frac{2}{k} \sum_k s_k^2 (x_{ijk} - x_k)\right)$$

$$\frac{\partial f}{\partial s} = \sum_{i \in pos} \left(\frac{1 - \mathrm{P}(x = X_t \mid x_i^+)}{\mathrm{P}(x = X_t \mid x_i^+)} \sum_j \frac{\mathrm{P}(x = X_t \mid x_{ij})}{1 - \mathrm{P}(x = X_t \mid x_{ij})} \frac{2}{k} \sum_k s_k (x_{ijk} - x_k)^2\right)$$

$$- \sum_{i \in neg} \left(\sum_j \frac{\mathrm{P}(x = X_t \mid x_{ij})}{1 - \mathrm{P}(x = X_t \mid x_{ij})} \frac{2}{k} \sum_k s_k (x_{ijk} - x_k)^2\right)$$

预测方法：

通过训练得到了的目标交点 x 和最优缩放比例 s。令 $p_{ij} = \exp(-\| x_{ij} - x \|^2)$，即可求得每个示例为正的概率，根据设定的合理阈值，就能得到每个示例的标记。

3.4 期望最大化的多样性密度（EM-DD）

3.4.1 算法简介

EM-DD 结合了期望最大化（EM）和多样性密度（DD）算法。多示例学习之所以困难，一个原因在于不知道哪一个示例能起决定性作用。EM-DD 算法与 DD 算法一样从一个假设的初始交点 x 开始，然后不断执行以下两个步骤，结合期望最大化方法和多样性密度方法来搜索最大化的可能性假设。第一步中，用当前的假设 x 在每个包中选出一个示例，这个示例最有可能决

定这个包的标记。在第二步中，搜索标准的多样性密度算法，以求找到一个新的 x' 可以最大化 DD 算法的目标。一旦这个最大化过程完成，就用 x' 代替 x，重新进入第一步，直到整个算法收敛。

EM-DD 算法不仅提高了 DD 算法的准确度，还减少了计算时间。在搜索目标交点 X_t 的过程中，DD 算法使用了每个包中的所有点，因此最大值一定会被计算。EM-DD 算法在第一步中排除了每个包中除了一个点之外的所有点，把多示例数据转换为单示例数据，这个做法大大简化了搜索步骤。

3.4.2 具体过程

与 DD 算法一样，随机选择一个正示例点 x 作为起点，其它示例到 x 距离为：

$$\| x_{ij} - x \|^2 = \frac{1}{k} \sum_k s_k^2 (x_{ijk} - x_k)^2$$

那么 $\mathrm{P}(x = X_t \mid x_{ij}) = \exp(-\| x_{ij} - x \|^2)$，从每个包中选出与初始点 x 最近的点代表这个包，所以：

$$\mathrm{P}(x = X_t \mid x_i^+) = \max_j \mathrm{P}(x = X_t \mid x_{ij}) \tag{8}$$

$$\mathrm{P}(x = X_t \mid x_i^-) = 1 - \max_j \mathrm{P}(x = X_t \mid x_{ij}) \tag{9}$$

算法过程：

1. 随机选择一个正包中的一个示例作为初始点 x；
2. 计算所有示例的 $\mathrm{P}(x = X_t \mid x_{ij})$；
3. 计算所有包的 $\mathrm{P}(x = X_t \mid x_i)$；
4. $f(x,s) = -\ln(\prod_{i \in pos} \mathrm{P}(x = X_t \mid x_i^+) \prod_{i \in neg} \mathrm{P}(x = X_t \mid x_i^-))$；

5. 用梯度下降法求 $f(x,s)$ 的最小值，与上一次迭代的最小值的求差，若差小于限度值，算法结束，此时的 x 和 s，即为目标交点和缩放比例，否则转到步骤 2。

其中：

$$\frac{\partial f}{\partial x} = \sum_{i \in pos} \frac{2}{k} \sum_k s_k^2 (x_{ijk} - x_k) - \sum_{i \in neg} \frac{\mathrm{P}(x = X_t \mid x_{ij})}{\mathrm{P}(x = X_t \mid x_i^-)} \frac{2}{k} \sum_k s_k^2 (x_{ijk} - x_k)$$

$$\frac{\partial f}{\partial s} = \sum_{i \in pos} \frac{2}{k} \sum_k s_k^2 (x_{ijk} - x_k) - \sum_{i \in neg} \frac{\mathrm{P}(x = X_t \mid x_{ij})}{\mathrm{P}(x = X_t \mid x_i^-)} \frac{2}{k} \sum_k s_k (x_{ijk} - x_k)^2$$

EM-DD 算法的示例概率和标记预测方法与 DD 算法相同。

3.5 支持向量机（SVM）

3.5.1 SVM 算法简介

支持向量机是一种基于分类边界的方法，基于分类边界的分类目标是，通过训练，找到这些类别之间的边界。支持向量机是依靠线性区分的，但并不是全部数据都可以线性区分。支持向量机将低维空间中的点映射到高维空间中，这样就可以用一条直线或一个平面作为分类边界。如果数据只有二维，那么数据可以用一条直线区分，如果数据有三维及以上，就可以用一个平面区分。如果数据是线性不可分的，可以通过核函数将数据升维，使得数据能够线性可分。在高维空间中，这是一种线性划分，而在原来的数据空间中，它却是一种非线性划分。

不管用直线还是用平面，将数据分隔为两个类可以用最大间隔法。最大间隔法的目标是求一个分类边界，同时使分类边界的间隔最大。分类边界的间隔是指两个类中的点与分类边界的最小距离。

分类平面表示为：$w^T x+b=0$，分类间隔的倒数为：$\frac{1}{2}\|w\|^2$。

所以该最优化问题表达为：

$$\begin{aligned}&\min_{w,b}\ \frac{1}{2}\|w\|^2\\&s.t.\ \ y_i(w^T x_i+b)\geq 1,\ i=1,.....n\end{aligned}\tag{10}$$

但由于样本可能是线性不可分的，不能达到原来对间隔的要求。因此我们考虑放宽约束条件，引入了松弛变量ξ_i，但是，我们还是要尽量减少ξ_i，最好让$\xi_i=0$，这样就能满足原来的要求。于是，在优化过程中使用了惩罚参数C，来实现对ξ_i最小化的目标。

这样，该分类器的模型为：

$$\begin{aligned}&\min_{w,b}\ \frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i\\&s.t.\ \ y_i(w^T x_i+b)\geq 1-\xi_i,\ i=1,.....n\end{aligned}\tag{11}$$

3.5.2 多示例学习的 SVM 算法

在多示例问题中，示例的预测标记y_{ij}和包的真实标记t_i可以用一组线性约束条件表示：

$$\begin{aligned}&\sum_j \frac{y_{ij}+1}{2}\geq 1,\ \ \forall i\ \ s.t.\ t_i=1\\&y_{ij}=-1,\ \ \forall i\ \ s.t.\ t_i=-1\end{aligned}\tag{12}$$

对于多示例学习的 SVM 算法，人们提出了两种构想，分别是：最大示例边界间隔的构想和最大包边界间隔构想。

（1）最大示例边界间隔的构想（mi-SVM）

由于示例标记不可知，在初始化时假设示例的标记和包的标记一样。

分类模型为：

$$\begin{aligned}&\min_{w,b,\xi}\frac{1}{2}\|w\|^2+C\sum_i\xi_i\\&s.t.\ \forall x_{ij}:\ y_{ij}(w^Tx_{ij}+b)\geq 1-\xi_i,\ \xi_i>0,\ y_{ij}\in\{-1,1\}\end{aligned}\tag{13}$$

算法过程：

1. 初始化所有包中示例的标记 $y_{ij}=t_i$；

2. 根据所有示例以及示例的标记，用 SVM 算法计算出最优的 w 和 b；

3. 更新所有正包中示例 x_{ij} 的标记 $y_{ij}=\mathrm{sgn}(f(x_{ij}))$，$f(x_{ij})=w^Tx_{ij}+b$；

4. 在正包 x_i 中，若 $\sum_j(y_{ij}+1)/2=0$，则令 $\max_j f(x_{ij})$ 所对应的示例标记改为 1；

5. 比较所有示例的标记，若所有包中没有一个示例的标记发生变化，算法结束，否则转到步骤 2。

（2）最大包边界间隔构想（MI-SVM）

以示例为中心的 mi-SVM 构想，正包中每一个示例的边界间隔都起作用，但在以包为中心的构想中，每个包中只有一个示例起作用，它将决定整个包的边界间隔。

分类模型为：

$$\begin{aligned}&\min_{w,b,\xi}\frac{1}{2}\|w\|^2+C\sum_i\xi_i\\&s.t.\ \forall x_i:\ t_i(w^Tx_{ij}+b)\geq 1-\xi_i,\ \xi_i>0,t_i\in\{-1,1\}\end{aligned}\tag{14}$$

算法过程：

1. 对于正包，初始化 $x_i^+=\frac{1}{J}\sum_{j=1}^{J}x_{ij}$ 作为代表整个包的示例，对于负包，由于每个示例的标记可知，所有示例都可作为训练样本；

2. 根据所有样本以及样本的标记，用 SVM 算法计算出最优的 w 和 b；

3. 计算出所有正包中示例 x_{ij} 的标记 $y_{ij}=\mathrm{sgn}(f(x_{ij}))$，$f(x_{ij})=w^Tx_{ij}+b$；

4. 更新所有的 x_i^+，用 $\max_j f(x_{ij})$ 所对应的示例 x_{ij} 替代 x_i^+；

5. 检查所有的正包的 x_i^+ 是否发生改变，若没有改变，算法结束，否则转到步骤 2。

预测方法：

不管是 mi-SVM 还是 MI-SVM，算法最终得到的都是所求的 w 和 b。对于测试集中的 x_{ij}，可以求出 $p_{ij}=f(x_{ij})=w^Tx_{ij}+b$，示例的标记 $y_{ij}=\mathrm{sgn}(f(x_{ij}))$。

3.6 最邻近结点（KNN）

3.6.1 KNN 算法简介

最邻近结点算法的基本思路是：在针对测试集中的数据点进行预测时，在训练集中找到与该数据点距离最近的 K 个数据点，根据这 K 个数据点所属的种类来对测试数据点所属的种类进行估计，具体的算法过程为：

1. 在训练集中选出与测试数据点最相似的 K 个数据点，计算公式如下：

$$Sim(d_i,d_j)=\frac{\sum_{k=1}^{N} d_{ik}\times d_{jk}}{\sqrt{(\sum_{k=1}^{N} d_{ik}^2)(\sum_{k=1}^{N} d_{jk}^2)}} \tag{15}$$

2. 在与测试数据点最相近的 K 个数据点中，分别计算每个种类的权重：

$$\mathrm{P}(x,C_j)=\sum_{d_i} Sim(x,d_i)y(d_i,C_j) \tag{16}$$

其中，x 为测试数据点，$Sim(x,d_j)$ 为相似度计算公式，如果 d_i 属于种类 C_j，那么 $y(d_i,C_j)$ 的值为 1，否则为 0；

3. 比较各类的权重，将测试数据点划分为权重最大的类别。

3.6.2 多示例学习的 KNN 算法

使用豪斯道夫距离可以让 KNN 算法用于多示例问题。在多示例问题中，每个样本都是一个包，但一个包中包括了多个示例，因此无法对应于一个单一的数据点。为了定义包与包之间的距离，需要去描述如何计算两组示例之间的距离。豪斯道夫提供了度量空间集合之间距离的标准函数。根据定义，两个集合 A 和 B 之间的豪斯道夫距离为 d，当且仅当 A 中每一个点和 B 中至少一个点的距离在 d 以内，同时，B 中每一个点和 A 中至少一个点的距离在 d 以内。

正式地，已知两个点集 $A=\{a_1,...,a_m\}$ 和 $B=\{b_1,...,b_n\}$，豪斯道夫距离定义如下：

$$\begin{aligned} H(A,B)&=\max\{h(A,B),h(B,A)\}\\ h(A,B)&=\max_{a\in A}\min_{b\in B}\|a-b\| \end{aligned} \tag{17}$$

豪斯道夫距离对 A 和 B 中，即使一个边缘的点都很敏感。这意味着距离可能仅仅由这个边远的点决定。为了提高关于噪声距离的稳健性，修改了豪斯道夫距离。修改后的距离为第 k 大的距离而不是最大值：

$$h_k(A,B)=\underset{a\in A}{kth}\min_{b\in B}\|a-b\| \tag{18}$$

3.6.3 引用方法（Citation-KNN）

一种将 KNN 用于多示例问题的方法思想来自于图书馆和信息科学中的引用。在这个领域中，找到相关文献是一个重要的搜索主题。一个著名的方法是基于参考和引用，如果一篇文章引用了其它之前出版的文章，这篇文章就和参考者有关。相应的，如果一篇文章被后来的文章引用，这篇文章也与引用者有关。因此，引用和参考都被视为与给定文章有关的候选文章。

引用的概念暗示了不仅要考虑包 x_i 的邻近结点，而且要考虑那些把包 x_i 视为邻近结点的包。假设包的数量为 N， $BS=\{x_1,...,x_N\}$，对于一个 $x\in BS$，所有其它的包 $BS=\{x_1,...,x_N\}$，可以根据与 B 的相似度进行排序。对于 $x'\in BS/B$，若它的排名 $Rank(x',x)\le C$，那么 x' 就是 x 的引用者。

算法过程：

1. 计算测试集中包 x_i 与训练集中包 x_j 之间的豪斯道夫距离 d_{ij}；

2. 选择与包 x_i 距离最小的 R 个包的标记，作为参考者的标记；

3. 如果包 x_i 是与 x_j 最邻近的 k 个包中的一个，那么 x_j 就是 x_i 的引用者，它的标记也被选择；

4. 如果这些被选择的标记中，正标记数量 C_+ 比负标记数量 C_- 多，那么这个包 x_i 的标记就为正；

5. $p_i=\dfrac{C_+}{C_++C_-}$。

Citation-KNN 是直接在包层次上进行预测的惰性算法，没有一种具体框架可以用来预测包中示例的标记和概率。

3.7 BP 神经网络（BPNN）

3.7.1 BP 神经网络简介

BP 神经网络是一种多层神经网络。信号从输入层输入，经过隐含层的处理，在输出层输出预测值。如果在输出层不能得到理想输出，则将预测误差反向传递给各层，让它们调整网络权值和阈值，然后重新进行整个预测过程，最后使 BP 神经网络的预测输出逐渐接近理想输出。BP 神经网络是一个非线性函数。

图 2.1 表示了 BP 神经网络的结构。

在 BP 神经网络中，用 n 代表输入层结点数、l 代表隐含层结点数，m 代表输出层结点数。输入层和隐含层神经元之间的连接权值为 w_{ij}，隐含层和输出层神经元之间的连接权值为 w_{jk}，隐含层阈值为 a，输出层阈值为 b。

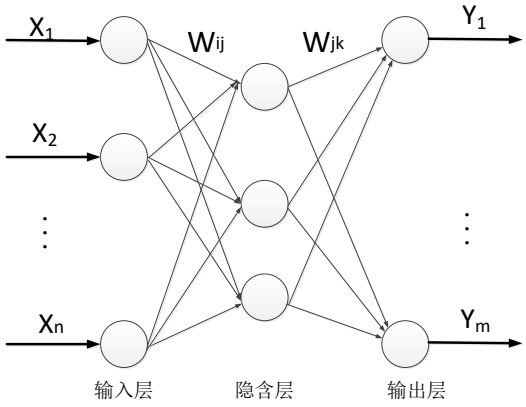

图 3.1 BP 神经网络的结构

算法过程的具体步骤为：

1. 初始化 BP 神经网络。确定 n，l，m，w_{ij}，w_{jk}，a，b 等参数值；

2. 计算隐含层输出 H，f 为激励函数；

$$H_j = f(\sum_{i=1}^{n} w_{ij}^{T} x_i + a_j) \quad j = 1,2,...,l \tag{19}$$

3. 计算输出层输出 O；

$$O_k = f(\sum_{j=1}^{l} H_j^{T} w_{jk} + b_k) \quad k = 1,2,...,m \tag{20}$$

4. 计算误差 e；

$$e_k = \frac{1}{2}(Y_k - O_k)^2 \quad k = 1,2,...,m \tag{21}$$

5. 更新权值和阈值；

6. 若迭代结束，算法终止，否则转到步骤 2。

3.7.2 多示例学习的 BP 神经网络

在我构建的网络中，输入层节点数 n 为训练数据的维度，隐含结点数 l 为 $\log_2 n$，输出层结点数 m 为 1。

隐含层和输出层激励函数，我选用的函数为：

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{22}$$

至于网络训练函数，我选用 Lenvenberg-Marquard 的 BP 算法训练函数。

算法过程：

1. 初始化所有训练包中示例的标记 $y_{ij} = t_i$；

2. 根据所有示例以及示例的标记，训练出 BP 神经网络；

3. 根据训练出的网络，对训练示例 x_{ij} 进行预测，输出结果为 $H(x_{ij})$；

4. 更新所有正包中示例 x_{ij} 的标记 $y_{ij}=\mathrm{sgn}(H(x_{ij}))$；

5. 在正包 x_i 中，若 $\sum_j (y_{ij}+1)/2=0$，则令 $\max_j H(x_{ij})$ 所对应的示例标记改为 1；

6. 比较所有示例的标记，若所有包中没有一个示例的标记发生变化，算法结束，否则转到步骤 2。

预测方法：

根据算法结束时所得到的 BP 神经网络，对测试数据集的所有示例 x_{ij} 进行预测，输出结果为 $H(x_{ij})$，那么示例 x_{ij} 的标记 $y_{ij}=\mathrm{sgn}(H(x_{ij}))$，$p_{ij}=H(x_{ij})$。

3.8 逻辑回归（LR）

3.8.1 LR 算法简介

逻辑回归由回归，线性回归和逻辑方程，三个部分组成。逻辑回归是线性回归的一种，很多情况下，我们需要对数据做归一化，引入了逻辑方程可以让回归产生一个类似概率值的 0~1 之间的数值。归一化的好处是让数值收敛在固定的边界之内，这样做也能使得它们之间具有可比性。

假设对于 n 维向量，$x=\{x_1,...,x_n\}$，$\mathrm{P}(y=1|x)$ 为某事件发生的概率，逻辑回归模型可用如下方式表示：

$$\mathrm{P}(y=1|x)=p=\frac{1}{1+e^{-g(x)}} \tag{23}$$

$\frac{1}{1+e^{-g(x)}}$ 就是 sigmoid 逻辑函数。其中 $g(x)=w^Tx+b$。

那么，事件不发生的概率为：

$$\mathrm{P}(y=0|x)=1-\mathrm{P}(y=1|x)=1-\frac{1}{1+e^{-g(x)}}=\frac{1}{1+e^{g(x)}} \tag{24}$$

假设有 n 个样本，观察值为 $y_1,...,y_n$，那么一个观测值的概率为：

$$\mathrm{P}(y_i)=\mathrm{P}(y_i=1|x_i)^{y_i}\,\mathrm{P}(y_i=0|x_i)^{1-y_i} \tag{25}$$

由于观测值之间相互独立，因此它们的联合分布为：

$$L(w)=\prod_{i=1}^{n} p_i^{y_i}[1-p_i]^{1-y_i} \tag{26}$$

上式称为 n 个观测值的似然函数，我们只要对这一似然函数的最大值做参数估计，就可以求出 w 的最优解。

对上述函数求对数 $\ln L(w)=\sum_{i=1}^{n}(y_i\ln p_i+(1-y_i)\ln(1-p_i))$。

于是损失函数定义为：

$$J(w)=-\sum_{i=1}^{n}(y_i \ln p_i+(1-y_i)\ln(1-p_i)) \tag{27}$$

如果采用牛顿法求解，参数的迭代公式为：

$$w^{(t+1)}=w^{(t)}-H^{-1}\nabla_w J \tag{28}$$

一阶导函数和黑塞矩阵表达式为：

$$\nabla_w J=\sum_{i=1}^{n}(\pi(x_i)-y_i)x_i \tag{29}$$

$$H=\sum_{i=1}^{n}[\pi(x_i)(1-\pi(x_i))x_i x_i^T] \tag{30}$$

3.8.2 多示例学习的 LR 算法

在多示例学习问题中，令 $g(x_{ij})=w^T x_{ij}+b$，那么 x_{ij} 标记为正的概率为：

$$p_{ij}=\frac{1}{1+e^{-g(x_{ij})}} \tag{31}$$

损失函数为：

$$J(w)=-\sum_{i=1}^{n}(t_i \ln p_i+(1-t_i)\ln(1-p_i)) \tag{32}$$

p_i 的计算方法就涉及到打包模型的选择，后面的章节我们会详细讨论这个问题。

算法过程：

1. 初始化 w 和 b；
2. 对训练集中所有示例计算 $g(x_{ij})$ 和 p_{ij}；
3. 计算所有包的 p_i；
4. 求损失函数的最小值，得到最优的 w 和 b。

预测方法：

根据训练得到的 w 和 b，可以用同样的方法对测试集中所有示例计算 p_{ij}。

3.9 AdaBoost

3.9.1 AdaBoost 算法简介

AdaBoost 算法是通过改变数据分布来实现的。针对同一个训练集，可以先采用各种方法训练出不同的分类器，但这些分类器的性能比较弱，根据训练集中的每个样本是否分类正确，计算整个训练集分类的准确度，更新每个样本的权值，然后对这样的新数据集，训练出下一个弱分类器。最后集合每次训练得出的不同弱分类器，构成最终的分类器。

AdaBoost 算法过程：

1. $D=\{x_1,y_1,...,x_n,y_n\}$，$t_{\max}$ 为最大迭代次数，第一步中 $t=0$，$W_1(x_i)=\frac{1}{n}$；

2. $t=t+1$，使用按照$W_t(x_i)$采样的D训练弱分类器c_t

3. 令E_t为C_t的训练误差；

4. $\lambda_t=\frac{1}{2}\ln\frac{1-E_t}{E_t}$；

5. $W_{t+1}(x_i)=\frac{W(x_i)}{Z_t}\times\begin{cases}e^{-\lambda_t}, if\ c_t(x_i)=y_i\\ e^{\lambda_t}, if\ c_t(x_i)\neq y_i\end{cases}$；

6. 如果$t=t_{\max}$，算法终止，否则转到步骤2。

迭代停止条件也可以根据实际情况调整，例如当前误差是否小于一个阈值时，算法终止。最终的分类判决是对每个弱分类器分类判决的加权平均：

$$H(x)=\mathrm{sgn}(C(x))$$
$$C(x)=\sum_{t=1}^{t_{\max}}\lambda_t c_t(x) \tag{33}$$

3.9.2 多示例学习的 AdaBoost 算法

我采用的弱分类器为决策树。决策树表示的是数据特征与数据类别之间的一种映射关系，它是一种不稳定的预测模型。

在 AdaBoost 算法的多示例版本中，示例层次的预测模型$C(x_{ij})$由多个弱分类器的输出线性组合构成，即$C(x_{ij})=\sum\lambda_t c_t(x_{ij})$。换言之，不同于多示例逻辑回归模型，输出标记$y_{ij}$对于相应的$x_{ij}$是非线性的。多示例 AdaBoost 的目标是在多示例框架下学习一组弱分类器$c_t(x_{ij})$，以及组合系数λ_t。

特别地，为了学习下一个弱分类器c_t，我们首先要固定目前已经学到的分类器，然后用它对每个x_{ij}估计y_{ij}。然后用 sigmoid 函数将y_{ij}过渡到p_{ij}，接着用合适的打包模型融合到p_i。最后，MIL_AdaBoost 的学习问题归结为最大化下面的似然函数：

$$\ln L(c_t)=\sum(t_i\ln p_i(c_t)+(1-t_i)\ln(1-p_i(c_t))) \tag{34}$$

在训练第t个弱分类器c_t时，已得到目前的分类器$\sum_{k=1}^{t-1}\lambda_k c_k(x_{ij})$，对于每一个$x_{ij}$，目前的得分为$y_{ij}=\sum_{k=1}^{t-1}\lambda_k c_k(x_{ij})$，现在的目标是求出$c_t$，使得似然函数值达到最大。按照梯度上升的思想，$y_{ij}=y_{ij}+\frac{\partial\ln(L)}{\partial y_{ij}}$。

对于正包的示例：梯度为：

$$\frac{\partial\ln(L)}{\partial y_{ij}}=\frac{t_i}{p_i}\frac{1-p_i}{1-p_{ij}}(1-p_{ij})p_{ij}=t_i\frac{1-p_i}{p_i}p_{ij}$$

对于负包的示例：梯度为：

$$\frac{\partial \ln(L)}{\partial y_{ij}} = -\frac{1-t_i}{1-p_i}\frac{1-p_i}{1-p_{ij}}(1-p_{ij})p_{ij} = (t_i - 1)p_{ij}$$

当加入第t个弱分类器c_t后，$y_{ij} = y_{ij} + \lambda_t c_t(x_{ij})$，所以$\lambda_t c_t(x_{ij}) = \frac{\partial \ln(L)}{\partial y_{ij}}$，又因为$c_t(x_{ij}) \in \{-1, +1\}$，若想$\lambda_t c_t(x_{ij}) = \frac{\partial \ln(L)}{\partial y_{ij}}$，$c_t(x_{ij})$的值就必须和$\frac{\partial \ln(L)}{\partial y_{ij}}$的符号相同。$|\frac{\partial \ln(L)}{\partial y_{ij}}|$的值越大，表示$c_t(x_{ij})$和$\frac{\partial \ln(L)}{\partial y_{ij}}$的符号相同的愿望越大。而$sign(\frac{\partial \ln(L)}{\partial y_{ij}}) = t_i$，因此令$w_{ij} = |\frac{\partial \ln(L)}{\partial y_{ij}}|$，就能按照愿望大小，求出新的$c_t$。

当求出c_t之后，还要确定λ_t的值，只要不断搜索，找出可以使似然函数达到最大的λ_t即可。

对于正包的示例，若p_{ij}较大，说明它是正示例的可能性很大，同时它的权重也较大，这样的示例所起的作用会越来越大；若p_{ij}较小，说明它为正的可能性很小，它的权重也较小，这样的示例就会逐渐被忽略。如果将为正概率很小的示例判做负例，重新开始下一轮训练，可以减少正包中的噪声。

预测方法：

最终得到的示例标记为$y_{ij} = \text{sgn}(\sum \lambda_t c_t(x_{ij}))$。

3.10 实验及结果分析

上面我们介绍了多种不同的多示例学习算法，接下来我们将通过实验来评估这些算法各自的性能，并进行比较。我们在药物活性预测和基于内容的图片检索，这两个经典的多示例分类问题中，使用我们实现的多种多示例学习算法，同时采用两种方式来评价这些算法。

3.10.1 实验数据

多示例学习已经应用于多个领域，例如药物活性预测，基于内容的图片检索，TrX 蛋白质的鉴定以及文本分类等。

多示例模型来自于药物活性预测问题，其中每一个示例都是一个分子可能的构造或形状。为了学习多示例模型的公共数据集都是用于概念学习，例如布尔型标记。分子与受体之间的绑定关系是定量描述的。

另一个多示例学习已经被应用的领域就是基于内容的图片检索。在这个领域中，一张图片被表示为一个包，图片的一部分对应于包中的一个示例。图片检索就是在图片数据数据库中找到包含感兴趣对象的图片。我们对于正包的假设是，图片的至少一个部分中含有感兴趣的对象。

在药物活性预测问题中，用于实验的数据集有：传统的数据集 Musk1 和 Musk2，以及 6 组不同维度的数据，分别为 30 维的 16.30.2 和 16.30.2-0.9，以及 166 维的 80.166.1，80.166.1-strong，

160.166.1，160.166.1-strong。在同样维度的数据中，相关特征的数量也不一样，例如 80.166.1 数据集在共 166 个特征属性中，80 个是相关特征，而在 160.166.1 数据集中，相关特征数量为 166 个。

在基于内容的图片检索问题中，用于实验的数据集有：230 维的 Elephant，Fox，Tiger，以及 15 维的 Desert，Mountains，Sea，Sunset，Trees。在 Desert，Mountains，Sea，Sunset，Trees 数据集中每个包中有 9 个示例，每个示例由 15 维的特征向量表示。1~3 维代表这个块的 R，G，B 三种颜色的灰度平均值，4~12 维代表这个块上下左右邻接块的 R，G，B 三种颜色的灰度平均值。

3.10.2 评价方法

要衡量分类的准确性，有两种常见的方法，一种是准确度 acc，另一种是 ROC 曲线下的面积 AUC。

假设测试集中正包的数量为 pos，负包的数量为 neg，预测结果为正的包中，实际上也为正包的数量为 tp，实际上为负包的数量为 fp，预测结果为负的包中，实际上为正包的数量为 fn，实际上也为负包的数量为 tn。那么，

$$\begin{aligned} pos &= tp + fn \\ neg &= fp + tn \end{aligned} \tag{35}$$

评价预测结果的准确度 acc 的方法是，计算预测正确的数量占总数量的比例，即：

$$acc = \frac{tp + tn}{pos + neg} \tag{36}$$

ROC 曲线是根据一系列不同的二分类阈值，以假正率 fpr 为横坐标，以真正率 tpr 为纵坐标，绘制的曲线。

$$\begin{aligned} tpr &= \frac{tp}{pos},\ fnr = \frac{fn}{pos} \\ fpr &= \frac{fp}{neg},\ tnr = \frac{tn}{neg} \end{aligned} \tag{37}$$

因此，ROC 曲线越靠近坐标点(0,1)，实验的准确性就越高。过 ROC 曲线上的所有点做斜率为 1 的直线，与纵轴截距最大的点即为最靠近坐标点(0,1)的点，相应的阈值也是错误最少的最合适的阈值。实验中可以根据这个最合适的阈值，调整算法的分类，提高预测的准确度。对于不同算法的实验结果，也可以通过分别计算各个实验的 ROC 曲线下的面积 AUC，并对它们进行比较，哪一种算法的 AUC 最大，则哪一种算法的准确性最佳。

acc 评价方法必须先根据阈值将预测样本唯一且明确的分为两类，然后才能据此进行分析。例如必须求出 tp 和 tn，才能计算出 acc 的值。而 ROC 曲线分析法却没有此要求，因为它是根据不同的阈值对预测样本进行多次分类，然后结合多次分类的结果。

3.10.3 结果与分析

（1）药物活性预测

在药物活性预测问题中，首先对于 Musk1 和 Musk2 数据集，各种算法的准确度 *acc* 和 ROC 曲线下面积 AUC 见表 3.2。在不同维度的数据集中做同样的实验，它们的准确度 *acc* 结果见表 3.3，ROC 曲线下的面积 AUC 结果见表 3.4。

表 3.2 Musk1 和 Musk2 数据集上实验结果

Method	Musk1		Musk2	
	acc	AUC	*acc*	AUC
APR	0.9240	0.9227	0.9520	0.9570
DD	0.9500	0.9568	0.9100	0.9354
EM-DD	0.9080	0.9266	0.9200	0.9544
mi-SVM	0.9500	**0.9892**	**0.9420**	**0.9801**
MI_SVM	**0.9660**	0.9876	0.9360	0.9754
KNN	**0.9660**	0.9853	0.9100	0.9349
BPNN	0.7740	0.7872	0.7180	0.7114
LR	0.9220	0.9219	0.9260	0.9225
AdaBoost	0.9340	0.9321	0.9380	0.9328

表 3.3 不同维度数据集的准确度 *acc*

Method	16.30.2	16.30.2-0.9	80.166.1	80.166.1-strong	160.166.1	160.166.1-strong
APR	**0.9500**	0.9833	0.6196	0.9674	0.9348	0.9565
DD	**0.9500**	**1**	0.9348	**1**	0.9674	**1**
EM-DD	**0.9500**	**1**	0.9348	**1**	0.9674	**1**
mi-SVM	0.8167	0.8333	**0.9891**	0.9891	0.9783	**1**
MI-SVM	0.8833	0.9000	**0.9891**	0.9891	0.9783	**1**
KNN	0.8833	0.9167	0.9565	**1**	0.9783	**1**
BPNN	0.8833	0.9167	0.9239	**1**	0.9783	**1**
LR	0.8333	0.9333	**0.9891**	0.9783	**0.9891**	**1**
AdaBoost	0.9333	0.9333	0.9348	0.9674	0.9613	0.9783

表 3.4 不同维度数据集的 ROC 曲线下面积 AUC

Method	16.30.2	16.30.2-0.9	80.166.1	80.166.1-strong	160.166.1	160.166.1-strong
APR	0.9501	0.9751	0.5228	0.9681	0.9118	0.9574
DD	0.9731	**1**	0.9536	**1**	0.9617	**1**
EM-DD	**0.9747**	**1**	0.9536	**1**	0.9617	**1**
mi-SVM	0.8866	0.8929	**0.9995**	0.9894	0.9985	**1**
MI-SVM	0.9338	0.9190	0.9850	0.9894	0.9843	**1**
KNN	0.8911	0.9161	0.9861	**1**	0.9967	**1**
BPNN	0.9338	0.9786	0.9582	**1**	0.9909	**1**
LR	0.8973	0.9641	0.9982	0.9863	**0.9992**	**1**
AdaBoost	0.9697	0.9520	0.9475	0.9872	0.9791	0.9877

（2）基于内容的图片检索

在基于内容的图片检索问题中，Elephant，Fox，Tiger 数据集上各种算法的准确度 *acc* 与 ROC 曲线下面积 AUC 结果见表 3.5。对于 Desert，Mountain，Sea，Sunset，Trees 数据集，各种算法的准确度 acc 结果见表 3.6，ROC 曲线下面积 AUC 结果见表 3.7。

表 3.5 Elephant，Fox，Tiger 数据集上的实验结果

	Elephant		Fox		Tiger	
	acc	AUC	*acc*	AUC	*acc*	AUC
APR	0.7680	0.7616	0.6680	0.6472	0.7300	0.7175
DD	**0.8880**	0.9235	0.7060	0.7051	0.8320	0.8612
EM-DD	0.8200	0.8639	0.6660	0.6806	0.8000	0.8170
mi-SVM	0.7580	0.8533	**0.7620**	**0.8562**	0.7200	0.8223
MI_SVM	0.7500	0.8392	0.7440	0.8495	0.7200	0.8179
KNN	0.8800	**0.9239**	0.7160	0.7376	0.8380	0.8784
BPNN	0.6820	0.6944	0.5580	0.5223	0.6480	0.6301
LR	0.8780	0.8952	0.7200	0.7022	**0.8920**	0.8940
AdaBoost	0.8660	0.9152	0.7400	0.7476	0.8840	**0.8951**

表 3.6 Desert，Mountains，Sea，Sunset，Trees 数据集上的准确度 *acc*

	Desert	Mountains	Sea	Sunset	Trees
APR	0.6400	0.5620	0.5200	0.5460	0.6520
DD	0.7300	**0.7360**	0.5960	0.7400	0.6780
EM-DD	0.7500	0.7280	0.5960	0.7400	0.6760
mi-SVM	0.6800	0.5540	0.5860	0.6840	0.5400
MI-SVM	0.7080	0.6960	0.6180	**0.7800**	0.6760
KNN	**0.7540**	0.7120	**0.6560**	0.7580	**0.6940**
BPNN	0.6960	0.7320	0.5920	0.7300	0.6520
LR	0.7120	**0.7360**	0.6240	0.7220	0.6720
AdaBoost	0.7440	0.7260	0.6340	0.7480	0.6700

表 3.7 Desert，Mountains，Sea，Sunset，Trees 数据集上 ROC 曲线下面积的 AUC

	Desert	Mountains	Sea	Sunset	Trees
APR	0.6294	0.5042	0.4145	0.4965	0.6566
DD	0.7837	0.7596	0.5664	0.7657	0.6649
EM-DD	0.7813	**0.7662**	0.5346	0.7591	0.6795
mi-SVM	0.6564	0.4432	0.5456	0.6206	0.3879
MI-SVM	0.7294	0.7157	0.6154	**0.8343**	0.6883
KNN	**0.7871**	0.7487	**0.6661**	0.7882	**0.7199**
BPNN	0.7096	0.7369	0.5686	0.7288	0.6538
LR	0.7356	0.7519	0.6211	0.7572	0.6826
AdaBoost	0.7259	0.7586	0.6052	0.7832	0.6791

经过以上两个实验，可以看出：

（1）在药物活性预测问题和基于内容的图片检测问题中，没有一个算法能够同时在这个两个问题的所有数据集上具有统治性优势，有些算法的性能波动较大，例如 APR，mi-SVM，有些算法则表现的比较稳定，虽然不是最好的，但永远不会是最差的，例如 LR 和 AdaBoost。

（2）药物活性预测的准确度比图片检索的准确度要高很多，可见基于内容的图片检索难度更大，这和我们主要研究的任意姿态人体检测一样，都是计算机视觉领域的问题。在基于内容的图片检索中，KNN 算法准确度较高，但不可忽略的是它的时间复杂度非常大。LR 和 AdaBoost 在诸多算法中性能处于较高水平，而且他们的时间复杂度较低，因此在下面将多示例学习用于弱监督检测时，我们主要考虑这两种算法。

（3）在不同维度的数据集上的实验结果表明，数据的相关特征个数越多，预测的准确度就越高。例如 166 维数据集上的准确度高于 30 维数据的准确度，有 160 个相关特征的数据集上的准确度也高于 80 个相关特征的准确度。

（4）各种算法在 Elephant，Fox，Tiger 数据集中的准确度比在 Desert，Mountain，Sea，Sunset，Trees 数据集的准确度较好。因为在 Desert，Mountain，Sea，Sunset，Trees 数据集中，15 维的数据特征太少。而且这 5 组数据集的用的是同样的包，只是在对不同的感兴趣的图像进行分类时，包的标记不一样。由于负示例的种类更多，学习难度也会提高，导致测试的准确度不理想。

3.11 本章小结

本章是对多示例学习算法的综合性介绍，并通过实验，实现了各种算法，在经典的多示例分类问题中比较了算法的准确度性能。表 3.8 总结了各种算法优点，缺点和适用情况。在和任意姿态人体检测同样属于计算机视觉问题的基于内容的图片检索中，我们发现了多示例学习用于计算机视觉问题的难度，同时也看出了 AdaBoost 和 LR 算法更适合用于弱监督检测问题。

表 3.8 各种算法分析总结

算法	优点	缺点	适用情况	分析
APR	对药物活性预测问题处理得较好。	算法太复杂，对样本的依赖性大。	因药物活性预测问题而生，最适合用于药物活性预测，但在其它情况下很难适用。	算法主要是用一个高维的轴平行矩形作为示例标记的分界面，这在很多数据分布情况下都很难做到用一个矩形区分所有的正示例，这个矩形外也包含了很多正示例。
DD	在所有算法中，准确度较高且稳定。	需要求出多个交点才能保证得到全局最优解，计算量大，收敛速度慢。	正示例在特征空间中分布集中，训练包中示例数量较少时。	准确度高是因为算法框架设计合理，速度慢不仅因为要计算的数据多，而且由于可能得到局部最优解，必须求出多个交点来确保得到真正的最优解。

表 3.8（续）各种算法分析总结

算法	优点	缺点	适用情况	分析
EM-DD	准确度较好	需要两层迭代，不断从零开始训练新模型，也需要求出多少个局部最优解，计算量大	和 DD 算法一样，要求正示例在特征空间中分布集中，但是 EM-DD 算法可以用于训练包中示例数量较大的情况。	用包中的一个示例替代整个包参与计算，减少了运算量，但也增大了偶然性和不稳定性，因为所选示例不一定就是那个能起决定性作用的示例，所以需要迭代来寻找最优示例，而且必须求出更多的局部最优解才能克服不稳定性。
mi-SVM	准确有时很高。	噪声太多，模型很难调整，时间复杂度太大。	不建议使用。	在训练时，需要明确的示例标记，但是这些标记通常很不准确，引入了很多噪声，而且如果样本的多样性太少，很难将错误的标记纠正。
MI-SVM	准确度较高，比 mi-SVM 稳定性好。	需要两层迭代，每次都要重新训练模型。	样本线性可分时，或者使用合适的核函数将样本升维后线性可分。	MI-SVM 算法之所以准确度比 mi-SVM 算法稳定，是因为用正包中一个示例代表整个正包，相当于只确定包中一个示例的标记，没有带来太大的噪声，计算量也变小了。
KNN	准确度较高，稳定性好。	时间复杂度较大	适合数据维度较小或数据个数较少时，但如果只追求准确度，数据量较多时也可以使用。	惰性学习算法不会利用训练集训练出一个用于预测的框架，而是在每一次预测新数据时都要访问训练集中的每个数据，但 KNN 算法基于距离的设计合理，还加入了引用的概念，所以准确度较好。

表 3.8（续）各种算法分析总结

算法	优点	缺点	适用情况	分析
BPNN	准确度有时很好。	大多数时候，准确度不理想，速度慢。	神经网络的这种多示例变形算法不宜采用，可以使用其它变形方法。	准确度较差一个原因是神经网络结构比较复杂，受初始化影响很大，不同初始化会导致不同的性能。另一个原因是和 mi-SVM 一样，假设了示例的标记，噪声太大。
LR	准确度高，稳定性好，速度快。	受初始化影响较大，需要求出多个分界线，确定最优解。	样本线性可分时，或者使用合适的核函数将样本升维后线性可分。	同样是线性分类器，多示例 LR 算法和 SVM 算法最大的不同在于不是根据样本标记训练，而是根据所求的正包为正的条件概率，计算误差，因此不需要重新训练模型，而是在当前模型上直接进行调整，速度快很多。
AdaBoost	准确度高，稳定性好。	当样本数量较少时，容易过拟合。	适用于样本数量较大的情况	AdaBoost 算法通过改变样本权重结合了多个弱分类器，属于非线性分类器，因此在样本较少时容易过拟合。虽然需要训练多个弱分类器，但每个弱分类器都很简单，训练时间很短，总体时间较快。

第四章 多示例检测深度评估

在研究弱监督检测算法之前，多示例检测作为弱监督检测的基础，我们有必要了解清楚其意义和机制，以及会对其产生重要影响的因素。在这一章，我们先给出将多示例学习作为一个框架融入弱监督检测的一般思想，随后我们将呈现对多示例检测的深度评估，着重研究在缺少监督样本的情况下全监督模型的性能变化，多示例训练包中正示例的纯度对性能的影响，以及少量监督信息的帮助等。

4.1 背景

在多示例学习[75]的设定中，数据集由 N 个包组成，可以表示为 $D=\{x_i,t_i\}_{i=1}^{N}$，其中 x_i 代表第 i 个包，t_i 是其相应的标记。假设一个包 x_i 包含了一组 J 个示例，即 $x_i=(x_{ij})_{j=1}^{J}$，一个包被视为多示例学习中一个完整的单元，就像普通全监督学习中一个单一的训练样本。一个包的标记主要由这个包中是否包含正示例决定。也就是说，只有不存在一个正示例的包才为负包（$t_i=0$），否则这就是一个正包（$t_i=1$）。在这个定义下，我们并不确定在正包中究竟哪一个示例是正示例，这加重了多示例学习算法的挑战，因为想要实现好的性能，它必须能够足够鲁棒地去对抗正包中的噪声数据。

多示例学习可以被用作弱监督对象检测的一个自然框架。原因如下：首先，一张图片可以被视为一个包，并被赋予了一个标记，但是我们没有更深层次的注释关于这张图片中对象位置的标记。其次，我们以无监督方式从这张图像生成一组提议。提议可以理解为一个特定包中的示例，为我们提供关于感兴趣的对象可能位置的初始假设服务。注意到我们没有关于这些示例的明确标记信息。在这些设置下，我们的目标是训练一个示例层次的检测器，它能够决定一个输入的提议是否是我们感兴趣的对象。如果我们将其视为一个给定的提议为正的条件概率，它遵循了一个好的检测器应该满足的条件，当把示例层次的预测融合到包层次时，应该与真实的包标记保持一致。这正是多示例学习算法的训练目标。

特别地，在我们的工作中，我们关心两个多示例学习算法，即多示例 AdaBoost（MILBoost）[76]和多示例逻辑回归（MILLR）[7]。之所以选择这两个算法，是因为 AdaBoost 是非线性分类器，而 LR 是典型的线性分类器，两者都是强分类器，而且在一定程度上可以形成互补。训练完成之后，我们也可以组合 MILLR 和 MILBoost。特别地，对于一个示例 x_{ij}，用 p_{ij}^{Ada} 和 p_{ij}^{LR} 分别表示它在两个算法下的得分输出，那么示例 x_{ij} 的最后得分定义为：

$$\hat{p}_{ij}=\max\{p_{ij}^{Ada},p_{ij}^{LR}\} \tag{38}$$

我们使用 Selective Search 方法[78]来生成检测提议，每张图片的提议数量在 1500 左右，然后用 VGG 网络[79]提取每个提议的特征，特征维度为 1000。当我们已经得到示例层次的检测器，在测试的时候，我们只要用它去给每个提议（示例）打分，表示它们为正的概率，输出为正得分最高的示例位置，经过窗口调整处理，最后确定我们的检测结果。

4.2 评估协议

检测性能的评估是基于检测器输出的边界框列表和它们相应的自信度得分。参照这个领域中的惯例（Pascal VOC[14]和 ILSVRC[15]），我们对每张测试图片根据公式（39）计算预测 B_p 和真实数据 B_{gt} 之间的 IoU（交集与并集之比）。和真实数据的 IoU 超过 0.5 的边界框被视为正。IoU 的定义如下：

$$IoU = \frac{B_p \cap B_{gt}}{B_p \cup B_{gt}} \tag{39}$$

和分类问题不同的是，检测问题需要对示例层次标记进行预测，因此无法从包层次进行检测性能评估。最后的性能由平均准确率（AP）机制来总结。对于一个特定的任务，准确率 $prec$ 和召回率 rec 曲线由一个算法的已排序的输出计算得到。对于检测算法，最后只输出预测为正的示例。在预测结果为正的示例中，实际上也为正示例的数量为 tp，实际上为负示例的数量为 fp。假设测试图片中正例的数量为 pos，召回率定义为所有正例中被预测为正的比例，即：

$$rec = \frac{tp}{pos} \tag{40}$$

准确率定义为所有预测为正的示例中预测正确的比例，即：

$$prec = \frac{tp}{tp + fp} \tag{41}$$

AP 总结了 $prec / rec$ 曲线，其值为这个曲线下的面积。

4.3 实验及结果分析

我们首先通过实验研究在全监督样本数量很少的情况下，全监督模型的性能变化，从而探究多示例检测存在的意义。接下来，我们进行一些实验来衡量对于多示例检测模型的有趣的影响因素，正包中正示例比例的影响和提高正示例比例的可行方法。这能够帮助我们快速认识弱监督多示例学习检测器，对于之后更深层次的研究非常重要。这一系列实验全都在 Pascal VOC 2007 数据集[12]上进行，其中训练集中包含了 2095 个正例样本和 2916 个负例样本，测试集的大小为 4952 张图片。

4.3.1 监督信息数量的影响

我们第一个想要研究的问题是在可得的监督信息的数量很少的情况会对最先进的全监督对象检测器的影响。为此，我们选择可变形部件模型（DPM），在 Pascal VOC 2007 数据集上，它的 AP 是 48.7%，虽然它的性能不及那些基于深度学习的方法，但是它所需要的已标记的样本的数量要少很多，因此它能够被有效的训练。

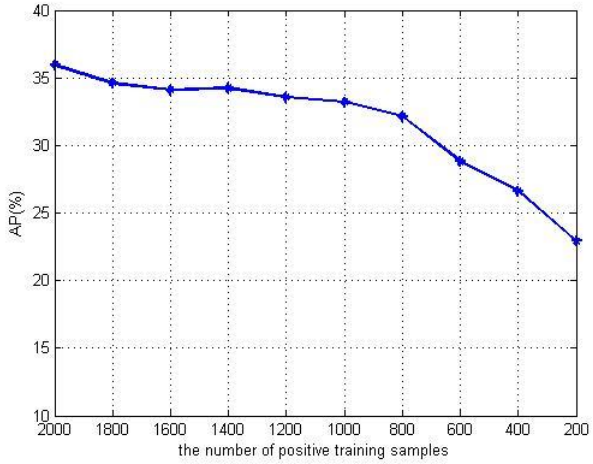

图 4.1 不同数量的监督样本下 DPM 的 AP

图 4.1 给出了随着可得的正训练样本数量不断减少得到的 DPM 的检测性能。为了获得变化趋势，我们逐渐减少已知标记的正例样本数量，负例样本数量和正例样本数量始终保持一致。我们重复了多次相同的实验，每次减少 200 个正例样本，样本数量从 2000 减少到 200。图中显示随着监督样本数量的减少，DPM 的 AP 性能也会逐渐降低。特别地，当正例训练样本的数量为 2000 时，AP 值最高为 35.96%。有趣的是，我们可以从图中发现，当已标记的正例样本数量少于 800 时，DPM 的性能下降较快，当正例样本数量很少到只有 200 时，AP 得分下降至只有 22.91%。

上述实验意味着对于 DPM 对象检测器可能存在一个断点，在这里它的性能会趋向于不理想。众所周知，一个全监督模型的最好的泛化能力需要模型复杂度和可得数据数量之间一个好的平衡实现，当监督样本的数量不是足够大的时候，一个复杂的模型的性能可能会由于过拟合而恶化，例如 DPM 模型。因此，在只有少量甚至没有全监督样本的情况下，全监督模型会失效，多示例学习的好处就得以体现，本章下面的时候可以说明多示例学习在少量样本的情况下相对于全监督模型的优势。

4.3.2 正示例比例的影响

接下来我们研究包中正示例纯度的影响。这里纯度通过正示例的比例（RoP）衡量，定义为正包中正示例的数量和示例总数之比。在标准的多示例学习设置下，我们不知道包中示例的标记，但是作为我们的实验基础，我们假设我们对这个信息已知，但是除了实验数据生成阶段，在检测器训练过程中，我们不会使用这些信息。有了这个条件，我们可以通过调整用于训练的包中的正示例和负示例的数量来自由的改变正示例的比例。正示例就是 IoU 值高于 0.5 的对象提议，负示例就是在同一个包中从非正示例随机采样得到的。在每一个比例下，我们训练三种多示例学习算法（没有使用任何示例标记信息），即 MILBoost，MILLR 和它们的集合版本

（MILBoost+MILLR）。这里，我们使用的融合示例层次条件概率到包层次条件概率的方法是最大池规则。作为比较，我们用全部全监督训练样本，同样训练了三种算法的全监督学习版本，分别表示为 FSLBoost，FSLLR 以及 FSLBoost+FSLLR。

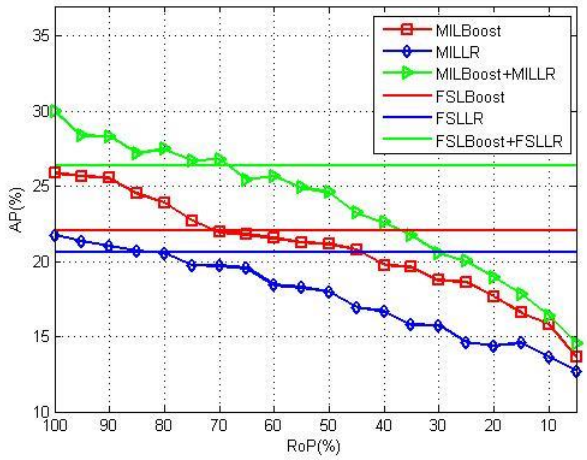

图 4.2 不同正示例比例下，多个多示例学习算法和它们的全监督版本的 AP

图 4.2 给出了实验结果，从这张图中可以发现几个有趣的现象：

（1）对于每个多示例学习算法，只要有一个足够高的 RoP，它们的 AP 得分可以高于全监督学习版本。例如，多示例学习的“MILBoost+MILLR”在 RoP 高于 70.0%时比“FSLBoost+FSLLR”更好。根据上一节的实验结果，我们可以推断在训练样本很少的情况下，全监督模型的性能还会下降，多示例学习的优势会更加明显。只要 RoP 达到 45%时，MILBoost+MILLR 的 AP 为 23.23%，这就会高于上一节中只使用 200 个正训练样本的 DPM 的 AP22.91%。

（2）图中显示，在正示例比例较高时，每种多示例学习算法都会有性能超越全监督学习的时候，所有正示例都是宽松定义的，这说明了宽松定义的正示例是有效的，对比于使用全部训练样本的纯粹的全监督学习，宽松定义的正示例可以提供更多的关于目标对象的信息。

（3）随着包中正示例纯度的减少，多示例学习的性能也会下降。此图显示，所有的多示例学习的性能会快速下降到低于它们相应的全监督版本。这个现象可能是示例层次模型恶化的结果。实际上，RoP 通常低于 30.0%，这也强调了为多示例学习算法选择高质量正示例的必要性。

（4）我们最后的发现是多示例学习的集合版本始终好于各自独立的版本。因此，在下面的实验中，我们选择“MILBoost+MILLR”作为我们的默认多示例算法。

4.3.3 示例与提议之比的影响

之前的实验中强调了提高正示例比例 RoP 的好处，假设关于示例层次标记的先验是已知的条件，虽然这在弱监督学习的环境下是不现实的。解决这个问题的一个方法是使用一些初始的示例层次检测器去为每个示例分配标记，这本质上也是一种 bootstrap 策略。我们可以使用极少

量的全监督样本来训练一个初始检测器，这些样本的获取并不需要耗费很大的代价，但这个初始检测器可能会发挥重要作用。

作为一种选择，我们可以通过改变示例与提议之比优化 RoP 的值。这个方法的关键思想是通过由极少量全监督样本训练得到的初始检测器在所有生成的提议中找到最可能为正的示例。特别地，我们在每个包中选择 n 个已生成的排名最高的提议，将它们作为那个包中的示例。这样，假如初始的对象检测器是可接受的，包中提议可能为正的范围就有效缩小了，相应的包中正示例的纯度也会提高。被选择的示例数量（n 个排名最高的）与所有生成的对象提议的比例称作 IoP。注意 IoP 可认为是弱监督学习环境中的一个操作机制，可能与 RoP 的值存在一种复杂的非线性关系，下面会说明这个问题。

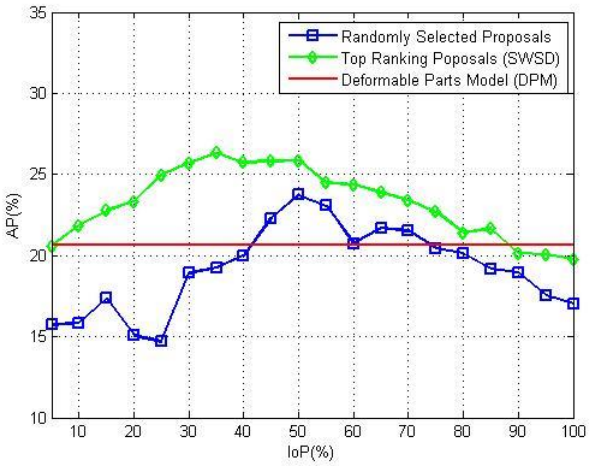

图 4.3 在不同 IoP 下提议选择方法的 AP 平均值

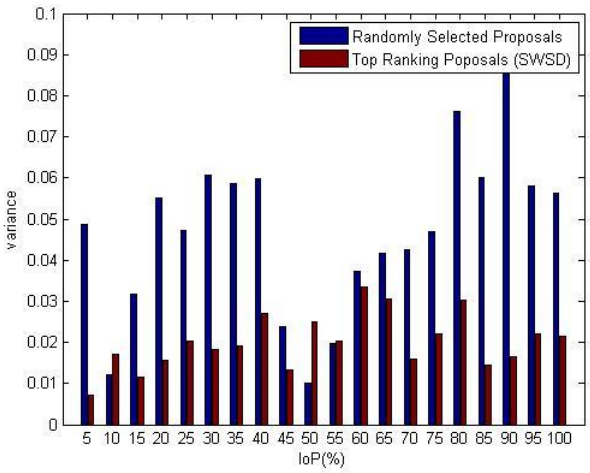

图 4.4 在不同 IoP 下提议选择方法的 AP 方法

为了研究不同的 IoP 对检测性能的影响，我们随机选择了 100 个正示例和 100 个负示例来训练一个初始的示例层次的对象检测器，然后用它在每个包生成的提议当中选择 n 个排名最高

的提议作为示例来重新训练多示例学习检测器。这个实验重复了 10 次，我们同时计算了 AP 的平均值和方差。作为比较，我们也通过随机选择相同数量的提议作为示例来训练多示例学习检测器。

图 4.3 和图 4.4 分别给出了不同 IoP 值下提议选择方法 AP 的平均值和方法。我们可以看到：

（1）选择得分排名最高的提议的方法，例如我们提出的 SWSD 算法，AP 准确度很大程度超过了随机的提议选择方法，而且 AP 方差更小，说明算法更加稳定，我们使用了很少的全监督信息训练得到的初始检测器来帮助选择提议，也说明了少量的全监督信息对于提高检测性能的帮助非常大。

（2）而且在缺少足够的监督信息的条件下，选择排名最高的提议的多示例检测算法也比全监督的 DPM 方法更好，它的性能几乎一直处于 DPM 的上方，可能的原因是，虽然两种方法使用的全监督信息数量都很少，但是弱监督检测还使用了大量额外的弱监督训练样本，虽然这些样本是带有很多不确定性，但在少量全监督信息的帮助下，可以消除部分噪声。

（3）此图揭示了既不是一个很高的 IoP 值也不是一个很低的 IoP 值会对性能有帮助。一个可能的解释是，一个大的 IoP 值可能包含了过多的噪声放入训练中，小的 IoP 值可能失去了很多有效的信息。一个好的 IoP 值与对象检测器的性能有潜在密切的关系，我们的实验结果显示 IoP 在 30.0%到 50.0%的范围内是对检测有帮助的。

通过比较此图和图 4.2，我们可以看到 IoP 和 RoP 之间的关系是非线性的。特别地，当 IoP 的值在 30.0%到 50.0%时，多示例检测算法的 AP 平均值达到最高，图 4.2 也揭示了 RoP 的值越高，性能就会越好。我们仔细看会发现，在被选择的示例 IoP 值在 40.0%时，它证明了相应的 RoP 值大约为 70.0%。这清楚的意味着这两者之间的关联。图 4.5 是 RoP 与 IoP 关系的完整说明。

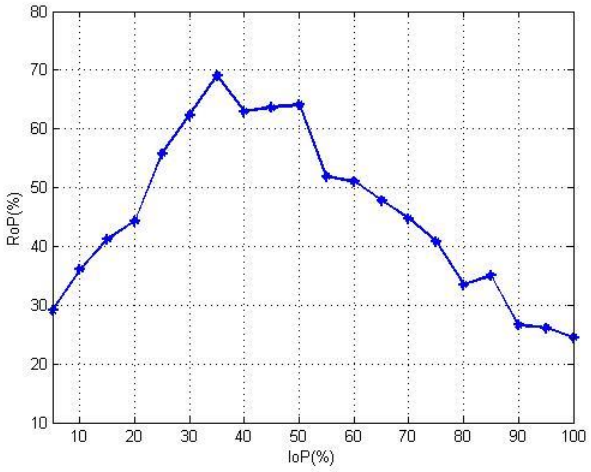

图 4.5 IoP 与 RoP 的之间的关系

4.4 本章小结

总而言之，这一章我们介绍了将多示例学习融入弱监督检测的思想，之后的多个基础实验给了我们一些重要的信息：（1）现在最先进的全监督对象检测方法，例如 DPM，在全监督信息不足的情况下性能会变得不理想；（2）当多示例学习的正训练包中正示例纯度足够大时，弱监督检测的性能可以高于全监督检测；（3）提高正训练包中正示例所占的比例对于弱监督多示例检测算法的成功十分必要；（4）使用少量全监督信息帮助我们选择得分排名最高的提议作为示例是提高正训练包中正示例纯度的有效方法；（5）在每张图片生成的所有提议中选择参与训练的示例数量也对检测性能也有较大影响。这些因素激励了我们去设计选择性弱监督检测算法。

第五章 Noisy-OR 和 ISR 模型的缺陷

除了上一章讨论的多示例检测影响因素，在多示例学习训练过程中，还有一个很重要的细节是将示例为正的条件概率过渡到包层次，从而根据已知的包标记计算误差，调整模型。常见的打包模型有最大池规则，平均规则，ISR 规则以及 Noisy-OR 规则。值得注意的是，上一章中的所有多示例检测实验，我们都采用最大池规则作为打包模型，但是这是否是最合适的方法，这一章我们将详细探究。对于多示例检测问题，我们第一个发现了 Noisy-OR 和 ISR 模型的缺陷。并且通过实验，证明了我们的想法。

5.1 打包模型

多示例学习一个重要的部件就是它的打包模型，即融合示例层次条件概率到包层次。正式地，我们表示示例 x_{ij} 为正的概率为 p_{ij}。为了估计包层次条件概率 p_i，可以采用不同的策略融合示例层次概率，典型的有：最大池（Max-Pooling），平均（Average），ISR 以及 Noisy-OR，分别如下：

Noisy-OR:

$$p_i = 1 - \prod_{j=1}^{J} (1 - p_{ij}) \tag{42}$$

ISR:

$$p_i = \frac{\sum_{j=1}^{J} \frac{p_{ij}}{1 - p_{ij}}}{1 + \sum_{j=1}^{J} \frac{p_{ij}}{1 - p_{ij}}} \tag{43}$$

Max-Pooling:

$$p_i = \max_j \{p_{ij}\} \tag{44}$$

Average:

$$p_i = \frac{1}{J}\sum_{j=1}^{J} p_{ij} \tag{45}$$

在这些模型之中，Noisy-OR 规则可能是最常用的，因为它与多示例学习的理念相匹配，即一个包被定义为正包当且仅当包中至少有一个示例为正。ISR 规则来自于文献[13]中的一个概率观点，可以理解为一种证据累积形式，类似于平均规则。最后，最大池规则可能是它们之中最简单且最直观的，它从本质上使得打包过程对于转换，旋转，移动保持不变。既然如此，一个自然的问题就是什么打包策略最适合于弱监督检测问题，下面我们将处理这个问题。

5.2 梯度消失问题

在这一节，我们将详细解释为什么在正包中存在过多负示例的情况下，应该避免使用 Noisy-OR 和 ISR 模型。这种情况在弱监督对象检测中经常遇到，由于缺少强监督信息，人们需要从一张图片中收集大量的示例来满足多示例学习的假设，即一个正包中至少存在一个正示例。然而，通常大部分示例为负，因为在一张单一的图片中，只有少量感兴趣的对象。

特别地，我们假设在一个正包中分别存在 10，100，1000 个示例，通过一个不是非常准确的示例层次模型，可以给每个示例分配一个很低的为正概率 0.1。注意到，由于假设大部分示例为负，上述分配是合理的，除了不知道正示例。表 3.1 给出了根据上述四种打包模型计算得到的 p_i 的值。可以看出，一旦示例数量超过 100，不论这个包是否为正包，Noisy-OR 和 ISR 模型都会趋向于对包为正的概率估计过高。在弱监督对象检测问题中，我们通常选择大量的图像碎片作为检测提议，数目通常超过 1000。

表 5.1 不同示例数量时，四种打包模型计算得到的 p_i 值

Method	10	100	1000
Max-Pooling	0.1	0.1	0.1
Average	0.1	0.1	0.1
ISR	0.5263	0.9174	0.9911
Noisy-OR	0.6513	1.0000	1.0000

为了更清楚地描述后果，我们采用多示例逻辑回归（MILLR）方法作为例子。这是一个线性分类器，示例层次的模型为 $y_{ij} = w^T x_{ij} + b$，其中 w 和 b 是学习得到的参数，p_{ij} 是使用 sigmoid 函数得到的示例为正概率。为了训练模型，采用负的似然作为损失函数：

$$L = -\sum_i (t_i \ln p_i + (1-t_i)\ln(1-p_i)) \tag{46}$$

w 的梯度为：

$$\frac{\partial L}{\partial w} = \sum_{i,j} \frac{\partial L}{\partial p_i}\frac{\partial p_i}{\partial p_{ij}}\frac{\partial p_{ij}}{\partial w} \tag{47}$$

其中，$\frac{\partial L}{\partial p_i}=-\frac{t_i}{p_i}+\frac{1-t_i}{1-p_i}$，$\frac{\partial p_{ij}}{\partial w}=(1-p_{ij})p_{ij}x_{ij}$。

注意到在学习参数的时候，梯度$\frac{\partial p_i}{\partial p_{ij}}$起到重要作用，它评估了示例层次概率$p_{ij}$的微小改变量如何影响了包层次概率$p_i$。在 Noisy-OR 和 ISR 模型下，分别有：

Noisy-OR:

$$\frac{\partial p_i}{\partial p_{ij}}=\prod_{k\neq j}(1-p_{ik})=\frac{\prod_k(1-p_{ik})}{1-p_{ij}}=\frac{1-p_i}{1-p_{ij}} \tag{48}$$

ISR:

$$\frac{\partial p_i}{\partial p_{ij}}=\frac{1}{(1+\sum_k\frac{p_{ij}}{1-p_{ij}})^2(1-p_{ij})^2}=\frac{p_i^2}{(\sum_k\frac{p_{ij}}{1-p_{ij}})^2(1-p_{ij})^2} \tag{49}$$

表 5.2 给出了和前面表 5.1 在同样的设定下，根据上述公式计算得到的分配给参数w的梯度值。可能会注意到，当示例数量超过 100 时，Noisy-OR 和 ISR 都会给w一个非常小的梯度。这背后的主要原因是，Noisy-OR 依赖于负示例之间相互独立的假设，使得模型更加趋向于为正，参考公式（42）。ISR 模型通过将乘积取代为求和，稍微缓解了这个问题，但是依旧将每个负示例和正示例平等对待，参考公式（43）。因此，一旦负示例数量是压倒性的，它也会遇到和 Noisy-OR 模型同样的数量问题。正如我们在训练深度神经网络时经常遇到的臭名昭著的梯度消失现象，这导致了对训练样本极低效率的使用。

表 5.2 不同示例数量时，四种打包模型对参数w分配的梯度值

Method	10	100	1000
Max-Pooling	$0.9\,x_{ij^*}$	$0.9\,x_{ij^*}$	$0.9\,x_{ij^*}$
Average	$\sum_k^J 0.09x_{ij}$	$\sum_k^J 0.009x_{ij}$	$\sum_k^J 0.0009x_{ij}$
ISR	$\sum_k^J 0.0474x_{ij}$	$\sum_k^J 8.2566*10^{-4}x_{ij}$	$\sum_k^J 8.9199*10^{-6}x_{ij}$
Noisy-OR	$\sum_k^J 0.0535x_{ij}$	0	0

另一方面，最大池和平均策略，表现的比另外两种模型更加稳定和鲁棒，参考表 5.1 和表 5.2。在一定程度上，它们表现的对于包中负示例的数量比较不敏感。但是简单的平均规则可能会低估p_i的值，最大池在所有示例上使用一个简单的非线性转换，提供了关于包中内容更加突

出和鲁棒的统计数据。最大池的梯度只依赖于为正概率最高的示例 x_{ij*}，无论在同一个包中存在多少的负示例，可根据 $p_{ij*}=\max_j\{p_{ij}\}$，$\frac{\partial p_{ij}}{\partial p_{ij*}}=1$ 证明这个观点。

最后，但同样重要的是，值得强调除了上述的缺陷，Noisy-OR 和 ISR 模型本身并没有问题。例如 Viola 等人成功使用了此模型用于训练基于多示例 AdaBoost(MILBoost)的对象检测器[76]。因为在他们的情境下，大部分示例（提议）确实为正，因为它们都是采样于感兴趣对象的已知位置。

5.3 实验及结果分析

最大池模型有效促进了包中的示例竞争。相反地，Noisy-OR 和 ISR 模型考虑了包中所有示例，但前提是假设它们之间相互独立。在这一节，我们将实际比较这四种在多示例学习中使用的经典打包策略，即最大池（Max-Pooling），平均（Average），ISR 和 Noisy-OR。 我们使用的多示例学习算法是 MILBoost 和 MILLR 的集合版本。

5.3.1 实验数据

为了更加全面的评估四种模型的性能，我们将同时在多示例检测和多示例分类问题中分别使用这四种模型。和上一章的多示例检测深度评估实验一样，对于多示例检测问题，我们使用的数据集是 Pascal VOC 2007，其中训练集中包含了 2095 个正例样本和 2916 个负例样本，测试集的大小为 4952 张图片。对于多示例分类问题，我们使用 Musk1，Musk2，Elephant，Fox 和 Tiger 数据集，和检测问题最大的区别在于，这几个数据集上每个包中示例的数量都是个位数的，而在检测中，每张图片产生上千的检测提议，即每个包中示例的个数是千位级别的。

5.3.2 结果与分析

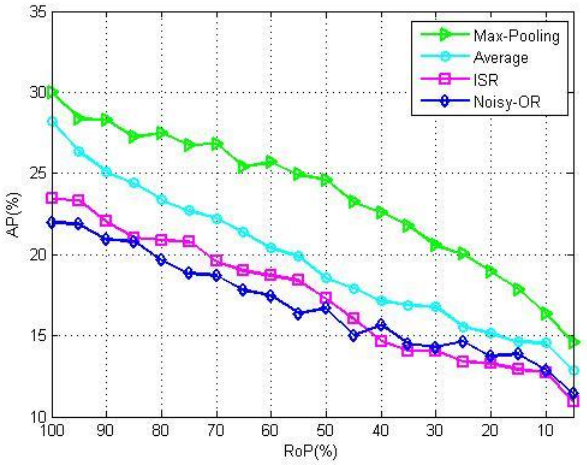

图 5.1 不同正示例比例下，四种打包模型的 AP

图 5.1 给出了在不同正示例比例 RoP 下四种模型的 AP 得分。可以看见 Noisy-OR 模型的性能始终低于最大池规则，实际上，最大池的最高 AP 为 29.85%，比 Noisy-OR 模型高了 7.8%。ISR 模型遭遇了和 Noisy-OR 相同的梯度消失问题，产生了和它差不多的性能。他们都比最大池和平均法表现的更差。实验结果与我们在 5.2 节的分析保持一致。

表 5.3 列出了四种模型在 Musk1，Musk2，Elephant，Fox 和 Tiger 数据集上的分类准确度，在包中示例数量只有个位数的情况下，Noisy-OR 和 ISR 模型相对于于最大池和平均规则的准确度差异并不像检测问题中表现的那么明显，说明 Noisy-OR 和 ISR 模型在示例数量过多的情况下才会失效，在示例数量很少时，依旧有效。

表 5.3 四种打包策略的分类准确度 *acc*

Method	Musk1	Musk2	Elephant	Fox	Tiger
Max-Pooling	0.7272	0.7894	0.8424	0.5853	0.8414
Nosiy-OR	0.7727	0.6315	0.8257	0.5466	0.6218
ISR	0.7272	0.6315	0.8227	0.5625	0.7451
Average	0.6818	0.7894	0.8285	0.6267	0.8263

5.4 本章小结

本章介绍了四种在多示例学习训练过程中常用的打包模型，并且详细解释了为什么 Noisy-OR 和 ISR 模型在包中示例数量较多的情况下会失效，所以它们不能适用于对象检测问题，我们提议使用最大池规则。另外，我们也通过实验说明了在检测问题中，Noisy-OR 和 ISR 模型的糟糕表现，同时也证明了它们在包中示例数量很少的情况下，依旧可以被采用。

第六章 弱监督人体检测

经过前面的介绍，我们已经深入了解了多示例学习应用于监督检测的细节问题，接下来我们将介绍我们提出的选择性弱监督检测算法（SWSD），它是在普通多示例学习算法的基础上进行的改造，采用了约束精英选择方法，使其更加适用于对象检测问题。此外，为了研究任意姿态人体的检测，我们自己标注了一个新的数据集 LSP/MPII-MPHB，并在其上检验了我们提出的算法的有效性。为了证明我们算法的通用性，我们还在 Pascal VOC 2007 和 Pascal VOC 2010 数据集的所有类别上进行了对象检测实验。

6.1 SWSD 方法概述

图 6.1 展示了我们所提出的选择性弱监督检测算法（SWSD）的总体框架。在训练中，对象检测提议作为示例层次检测器的输入来评估它们为正的得分。根据每个示例的得分和位置，我

们采用约束精英选择算法去调整搜索区域，然后选择排名最高的提议（精英）作为训练样本更新当前的检测器。在测试阶段，我们使用学习得到的检测器去给每个提议打分，然后预测边界框。主要思想是感兴趣的对象检测器可以理解为一个将包层次的监督信号反向传播给示例层次的传播者，基于这些信息，人们可以选择一些排名最高的提议（也可以叫做精英）反过来训练示例层次的对象检测器。

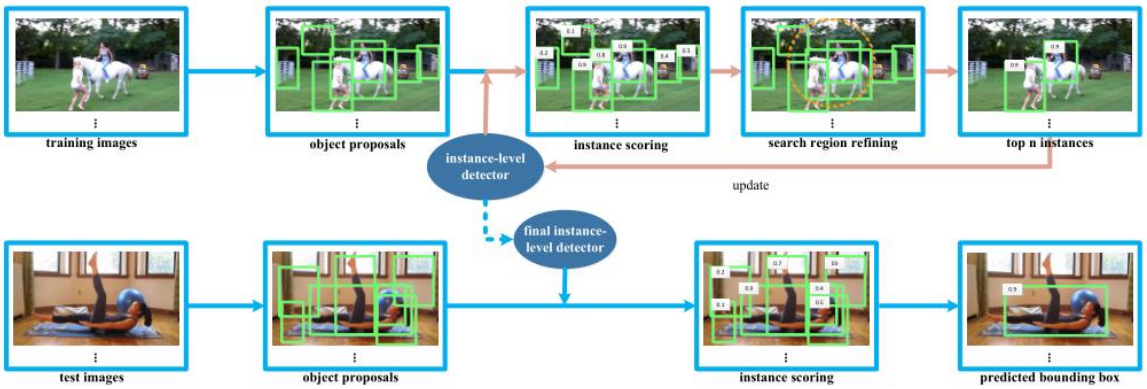

图 6.1 SWSD 算法的总体框架

为了提高初始的示例层次检测器的准确性，我们利用了在现实中可得的少量监督信息。另外，这些全监督示例会在整个训练过程中一直存在，担任多示例学习模型正则项的角色。

更正式一点，假设我们有 M 个全监督示例 $(x_j, t_j)_{j=1,\ldots,M}$，示例层次的分类器表示为 $y(x_{ij}, \theta)$，它用来预测每个输入的示例 x_{ij} 的标记。我们的目标是：（1）我们希望分类器在包层次产生好的预测；（2）它也应该可以正确预测我们那些少量的监督示例。这些要求归纳为下面的目标：

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}\log(1+\exp(-t_i \cdot p_i)) + \lambda_1 \frac{1}{M}\sum_{j=1}^{M}\log(1+\exp(-t_j \cdot y(x_j, \theta))) + \lambda_2 \| \theta \|_2 \quad (50)$$

这里 λ_1 和 λ_2 是两个用户自定义的参数，前者控制了监督信息对于模型训练贡献的程度，后者是一个常规的正则化系数，p_i 由上述的最大池规则计算得到：

$$p_i = \max_{j \in i} \sigma(y(x_{ij}, \theta)) \quad (51)$$

$\sigma()$ 是标准的标注的 sigmoid 函数（S 曲线函数）。

6.2 约束精英选择

既然以迭代的方式训练示例层次的预测模型 $y(x_{ij}, \theta)$，在文献[11]中却发现了一个潜在的问题，即训练过程可能会过早的停止，因为在反复选择训练样本的过程中所选择的精英缺少差异。他们提出使用一部分样本来训练模型，然后用它在另外一部分样本中选择最可能为正的示例作为训练样本。这种多重方法从本质上在包层次上注入了多样性。取而代之的是，我们提出了一

个新的示例层次的方法叫做约束精英选择，它限制了用于下一轮训练的可选择的候选提议的范围，以达到在使训练稳定化的过程中，减少训练噪声的效果。

特别地，在每一次迭代中，我们会通过简单的高密度子图搜索算法[77]调整候选提议的搜索区域。我们将每个包视为一张图G，包中每个示例是图上的节点。边的权重通过二维图像空间中示例的两两距离d_{jk}计算，这个二维空间的权重由相应示例的平均得分赋予，即第j和第k个示例之间的权重$w_{jk}=(p_{ij}+p_{ik})/2$。高密度子图搜索算法在图G中寻找最高密度的子图S，密度的定义为：

$$den(S)=\frac{\sum_{j,k\in S} w_{jk}\cdot d_{jk}}{|S|} \tag{52}$$

这个算法将排名最高的提议们以灵活的方式划分为g个小组（密度子图），我们采用最高密度子图中成员形成的最小二维凸包作为下一次迭代的搜索区域。另外，我们对每次迭代设置了一个收缩比率η来减少搜索区域的大小，随着训练过程的进行，当更多的信息被开发利用，我们的检测自信度也会不断提高。

在确定了搜索区域之后，它可以作为一个过滤器过滤掉在这个区域之外的对象提议。为了取代从剩下的提议中直接选择n个排名最高的提议，我们首先从这些剩下的提议中随机采样r比例的示例，然后选择k比例的最高排名示例。因此，最终选择的示例数量为$n=r\times k\times|l|$，$|l|$表示包中属于搜索区域l的示例数量。参数r控制了在寻找最好的候选提议的过程中，有多少示例应该被放弃，它是处理开发与利用的问题的一个内置机制，但是总体上，这种半随机策略能够帮助算法从坏的局部最小点（拥有高分的错误位置）中跳出，并且提高了对抗示例层次模型不确定性的鲁棒性。

在算法 1 中总结了整个训练过程，其中有三个主要参数，即迭代次数T，样本比例r，以及每次迭代中选择的最高排名的提议数量。在实验中所有参数都根据交叉验证设置，在实验中我们会研究它们对于算法的敏感度。

算法 1：选择性弱监督检测（SWSD）

1.初始化训练集S_0：由M个全监督样本和N个空的正训练包组成，其中全监督样本中，一半是真实数据，一半是正包中的负示例。同时会生成每张图片的检测提议；

2.使用S_0训练一个示例层次的检测器，并且基于每个提议的检测得分定位一个相对较大的搜索区域l；

3.开始迭代过程，$t=1\sim T$

（a）对于每个正包，在搜索区域l中随机选择r比例的示例（提议）；

（b）在这些随机选择的示例（提议）上运用当前的多示例学习检测器，并分别为每个示例分配一个得分；

（c）搜索最高密度子图，然后根据它定位下一个搜索区域；

（d）从最高密度子图中选择 n 个最高排名的提议，结合 S_0 构建一个新的训练集 S_t；

（e）使用 S_t 训练一个新的多示例学习检测器；

4.返回最终的检测器。

6.3 LSP/MPII-MPHB 数据集

这个数据集由我们自己构建，其中的图片都选自 LSP[16]和 MPII Human Pose 数据集[4]。在 LSP 数据集中，几乎所有图片中的人体都主要位于整个图片的中央，背景也没有很混乱，因此这些图片携带了关于人体更多的有效信息，这对我们的检测器训练很有帮助，见图 6.2。另一方面，MPII Human Pose 数据集中的图片包含了更多的姿态变化，而且可能是从不同的比例下捕获的，有些人体的尺寸非常小，一张图片中也可能包含了多个人体，参考图 6.3。

图 6.2 LSP 数据集中的人体说明

图 6.3 MPII Human Pose 数据集中的人体说明

最开始 LSP 和 MPII Human Pose 数据集都是为人类姿态估计创建的，因此伴随了人体关键部位的注释，例如，脚，腿等等。但不幸的是，这些注释不能直接作为人体边界框用于人体检测中，而且从它们推断得到的边界框噪声也会过多，这对于性能评估是不可信赖的。为此，我们基于 LSP 和 MPII Human Pose 数据集建造了一个新的数据集命名为 LSP/MPII-MPHB（多姿态人体）专门用于人体检测。我们选择了 26000 多张图片，并且对每张被选择的图片关于人体边界框做了必要的注释。

最后得到的 LSP/MPII-MPHB 数据集，包含了 26675 张图片和 29732 个人体。每张图片至少包含一个人体，有些图片可能包含了多个人体。在这些图片中，2000 张来自于 LSP 数据集，24675 张来自于 MPII Human Pose 数据集。我们对每个人体计算了真实数据的边界框对于整张图片的尺寸比例，并统计了频率直方图如图 6.4 所示。可以发现，大约 70%的真实数据的尺寸比例少于 10%，预示着检测 LSP/MPII-MPHB 数据集中的人体是个具有挑战性的工作。

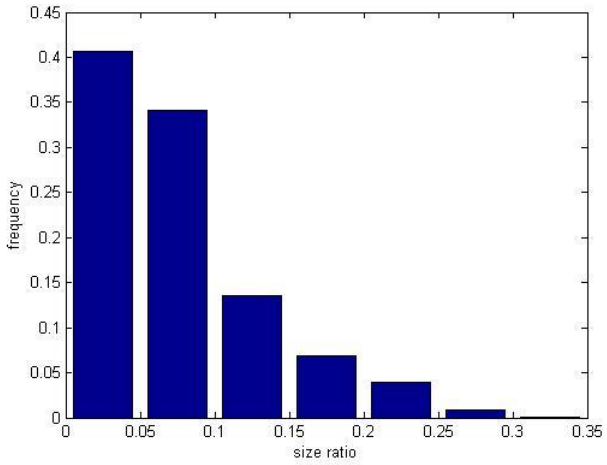

图 6.4 LSP/MPII-MPHB 数据集中人体尺寸比例的分布

LSP/MPII-MPHB 数据集中人体呈现多样的姿态，它们大致可分为六类，即弯曲，跪着，躺着，遮挡，坐着和直立，见图 6.5 给出的姿态说明，并且在表 6.1 中给出了更详细的描述。注意到每个姿态类别中图片的数量是不平衡的，最主要的类别是弯曲的，跪着和躺着的图片数量相对较少。

表 6.1 LSP/MPII-MPHB 数据集中六种典型的人体姿态的详细信息

姿态类型	描述	数量
弯曲（Bent）	至少有一个身体部分是弯的，例如弯腰或者马步	10229
跪着（Kneeling）	两个膝盖中至少有一个接触其它物体	1053
躺着（Lying）	睡觉或者游泳，呈现水平姿态	1123
遮挡（Occlusion）	只有身体的一部分	5739
坐着（Sitting）	臀部接触其它物体	4040
直立（Upright）	没有任何弯曲的站着	4492

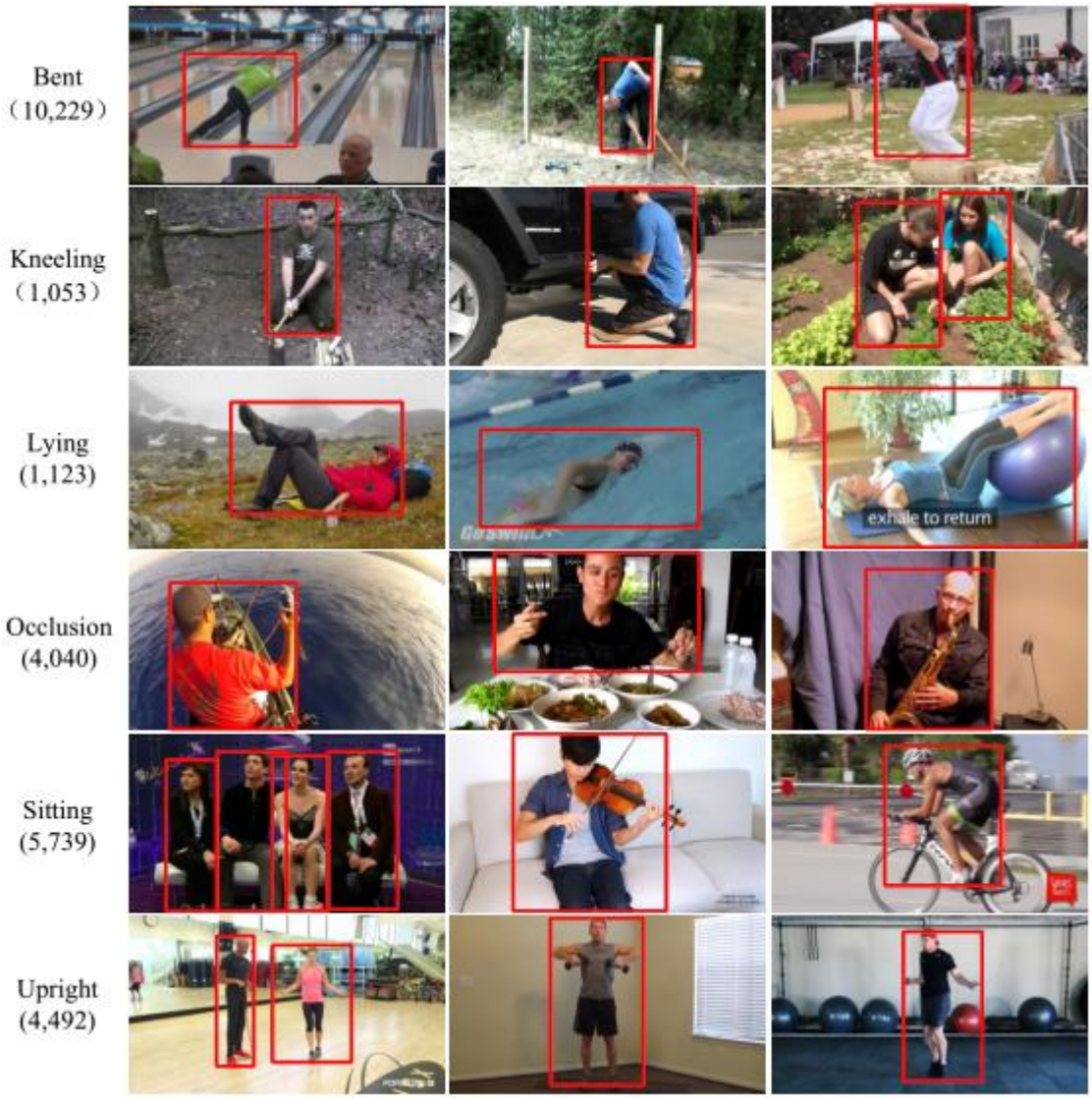

图 6.5 LSP/MPII-MPHB 数据集中六种典型的人体姿态说明

6.4 实验结果及分析

我们将 LSP/MPII-MPHB 数据集中的所有图片分为三部分，即训练集，验证集和测试集。具体地，图片数量分别为 8385，8110 和 10180，人体数量分别为 9732，8233 和 11767。

在本节所有实验中，我们使用下列默认的实验设置：在训练中，初始时我们使用 100 个全监督示例作为先验，算法中参数 r 设置为 95%，T 为 6，公式（50）中的正则系数分别设置 λ_1 为 1.0，λ_2 为 0.001。采用的多示例学习算法是 MILBoost 和 MILLR 的集合版本，所有的多示例学习算法采用最大池规则进行训练。

6.4.1 准确度性能评估

为了验证我们所提出的选择性弱监督检测方法（SWSD）的有效性，我们把它和两种最先进的弱监督对象检测方法，即 Cinbis 等人的多重多示例学习（MMIL[11]）和 Bilen 等人的后验正则化隐 SVM 方法（PRLS[55]），在我们最新标注的大规模人体数据集 LSP/MPII-MPHB 上进行比较，。两种算法均由我们自己实现，我们已经在 Pascal VOC 2007 数据集上验证了我们的程序的正确性，发现了和在原始报告的论文中相似的实验结果。为了说明全监督正则项的贡献，我们也评估了两种 SWSD 算法的变形体的性能，一个是仅仅使用 100 个全监督示例训练得到的检测器，另一个是不使用任何全监督信息的 SWSD 算法。

表 6.2 LSP/MPII-MPHB 数据集上多种人体检测方法的 AP

Method	AP(%)
PRLS[55]	10.97
MMIL[11]	16.61
SWSD	**35.37**
（a）标准的 MILBoost（Max-Pooling）	19.38
（b）标准的 MILBoost（Average）	15.54
（c）标准的 MILBoost（ISR）	11.91
（d）标准的 MILBoost（Noisy-OR）	10.81
（e）全监督检测器（100 个样本）	27.44
（f）SWSD（-）（没有全监督正则项的 SWSD）	21.26
（g）FMP[80]	24.33

表 6.2 给出了多种人体检测方法在 LSP/MPII-MPHB 数据集上的性能结果，可以看出以下几点：

（1）我们提出的 SWSD 算法在所有比较的方法中表现最好。正如表中所示，提出的方法远远超过了两种目前最先进的弱监督检测器，超出的性能分别为 18.8%和 24.6%。这说明在检测高度变形的人体时，我们的方法比这两种方法更加鲁棒。

（2）表中也说明了最大池规则比其它打包策略更加适合于弱监督对象检测。在 AP 方面，最大池策略比 Noisy-OR 提高了 8.5%。

（3）此表也揭示了即使在没有全监督正则项的情况下我们的性能优势，我们 SWSD（-）的 AP 为 21.26%，高于其它的纯粹的弱监督方法，标准的 MILBoost 最好的 AP 为 19.38%，PRLS 的 AP 为 10.97%， MMIL 的 AP 为 16.61%。虽然这个方法比使用 100 个样本的全监督方法低了 6.2%，但是在同样使用 100 个监督样本的情况下，我们的 SWSD 算法比这个全监督方法高了 8.0%，说明了我们提出的迭代的多示例学习过程是有效的。

（4）注意在表 6.2 中，我们也将我们的方法和最先进的姿态估计方法 FMP[80]进行了对比。特别地，我们是第一个使用 FMP 去估计一张图片中的人体部分位置，然后根据这些位置推断人体边界框来评估性能。表中的结果显示姿态估计方法的性能比我们的 SWSD 方法大约低了 10%，说明了人体解析方法可能不能直接应用于人体检测，这种方法可能依赖于人体检测的结果用于更深一步的部件解析，也可能不能处理人体尺寸很小的情况。

表 6.3 LSP/MPII-MPHB 数据集的不同姿态上的检测 AP

姿态类型	MILBoost[76]	PRLS[55]	MMIL[11]	SWSD(-)	SWSD
弯曲	10.90	11.61	14.41	17.43	20.68
跪着	14.58	15.33	15.91	16.42	22.52
躺着	6.97	7.30	9.53	9.36	10.68
遮挡	14.06	13.94	15.61	22.28	47.20
坐着	11.03	11.93	17.18	20.07	29.63
直立	23.57	25.53	27.50	29.56	48.28

表 6.3 给出了每种姿态的具体检测性能。这能够帮助我们更加清楚的理解我们提出的方法优势究竟在哪。可以看见我们的 SWSD 方法在所有姿态类型上以 5.0%~20.0%的比率超越了其它被比较的方法，除了躺着的姿态。尽管如此，此表也揭示了那些非直立的姿态，例如躺着，弯曲和跪着比其它的姿态挑战性更大。可能的原因是缺少训练样本（对于躺着或者跪着），也可能是因为大量的形变（例如弯曲）。

图 6.6 展示了我们的 SWSD 方法一些正确和错误的检测结果。正确的定位以黄色实线框显示，错误的定位以粉色实现框显示，不显示的表示缺失的预测，相应的正确位置用粉色虚线框表示。虽然在某些特定的姿态类别上检测性能有待提高，例如躺着，可以通过收集更多的训练数据来改善。但是这些图片上的检测结果也显示了在一般情况下，即使在部分遮挡，多姿态，光线暗和小尺寸的条件下，我们的 SWSD 方法也能够检测出人体，说明我们的方法是可信赖的。

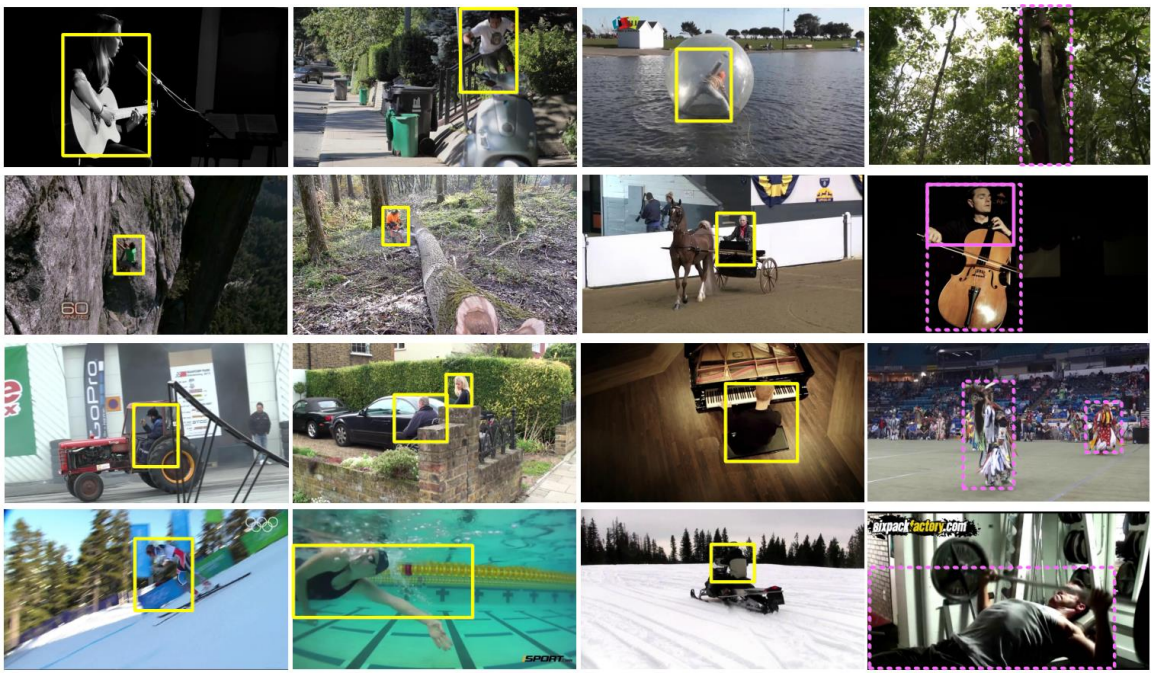

图 6.6 SWSD 算法在 LSP/MPII-MPHB 数据集上的检测结果说明

6.4.2 对象重定位行为

为了具体研究在处理一张训练图片时，我们的 SWSD 方法的对象重定位行为，我们将它与 Cinbis 等人的多重多示例学习方法进行对比（MMIL[11]），这个方法也是通过迭代不断调整粗略的对象位置估计。正如第二章描述的，MMIL 和我们的方法之间主要的不同点总结如下：（1）在初始化阶段，MMIL 采用保守的策略，从几乎整张图像空间开始搜索，但我们的方法开始于一组由弱对象检测器验证的对象提议；（2）为了注入多样性，MMIL 在每一次迭代都更换训练包，而我们使用约束精英选择策略；（3）MMIL 每一次迭代会挖掘负示例，而我们使用了少量的全监督正示例作为正则项。

这些不同点导致了两种算法不同的对象重定位行为，如图 6.7 所示。橙色虚线圆圈表示搜索区域，正确的定位结果用黄色显示，不正确的用粉色显示。这张图分别说明了两种方法的对象重定位过程。我们可以看到两种算法都成功定位到了感兴趣的对象，如图中第一个儿童例子所示，而且我们方法产生的边界框看上去比 MMIL 产生的更加紧致。在第二个例子中，多重多示例学习方法困于整个包围拖拉机的窗口。与之不用的是，我们的 SWSD 能够逐渐定位到人类，即使对象是部分遮挡的而且在一个非常小的区域内。注意到在这个例子中，我们的方法从一些明显不是很好的初始位置开始搜索，但是仍然可以收敛于一个不错的位置。在最后一个例子中，我们的方法过早地锁定在一个错误的位置。一个可能的原因是全监督正则项的影响，它对船的位置产生了一个强烈的反应，这压制了随后重新定位感兴趣的人类的尝试。这也说明了弱监督

检测器在某些特殊的情形下潜在的危害，可以解决这个问题的方法是在搜索过程中加入更多的多样性，这将是我们将来研究的重点。

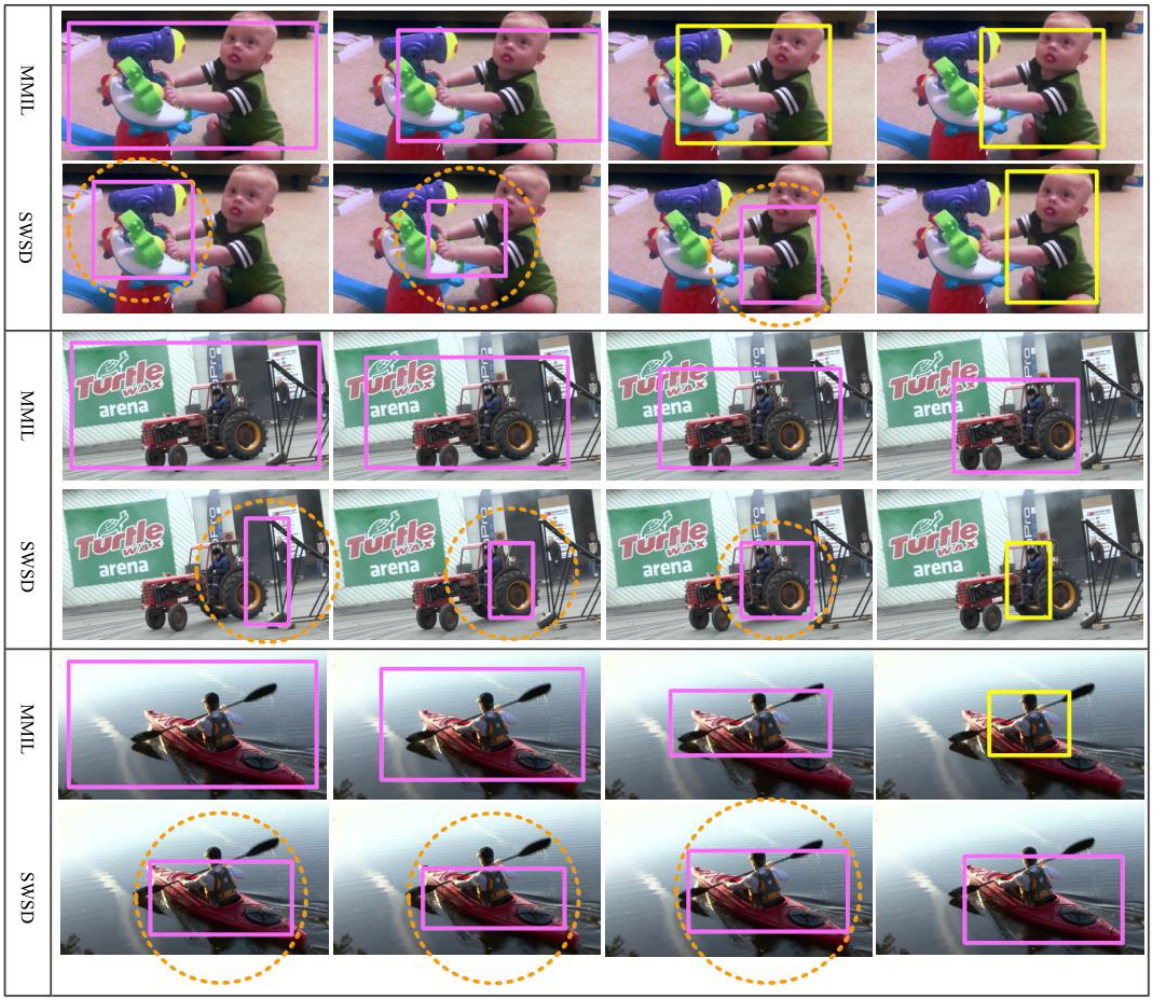

图 6.7 MMIL 和 SWSD 算法从初始到最终的对象重定位过程

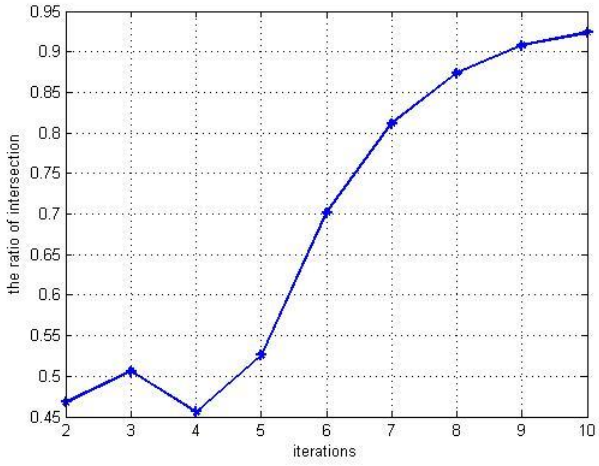

图 6.8 SWSD 算法两次连续的迭代中搜索区域交集的比例

为了研究约束精英选择方法的行为，图 6.8 说明了对于图 6.7 中第一个儿童例子中的图片，两次连续的迭代之间，搜索区域交集的改变比例。搜索区域的交集定义为两次连续的迭代步骤之间，在每个独立的搜索区域中，相同提议的数量。我们可以看到在第 2 至第 5 次迭代之间，交集的比例维持在一个相对较低的值 0.5，表明搜索区域改变的非常活跃。但是在第 6 次迭代之后，这个比例高于 0.7，而且增长迅速，说明了检测器的训练过程趋向于稳定和收敛。

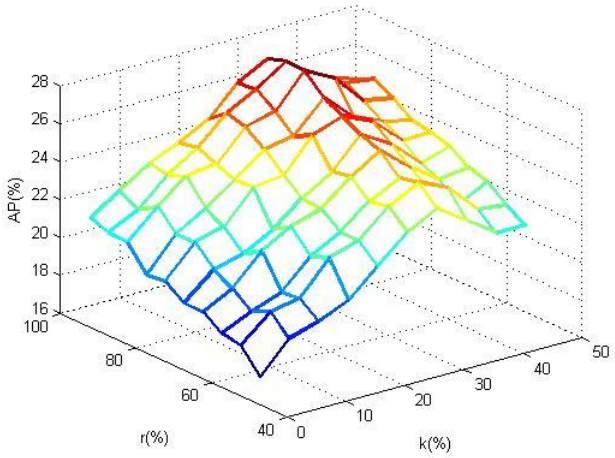

图 6.9 SWSD 算法对于不同的r值和k值，在 LSP/MPII-MPHB 数据集上的 AP

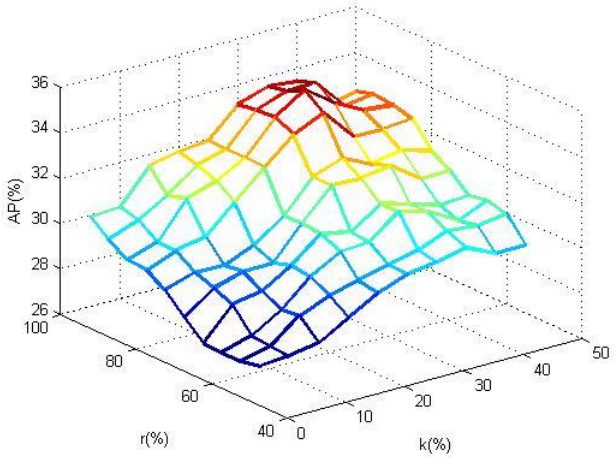

图 6.10 SWSD 算法对于不同的r值和k值，在 Pascal VOC 2007 数据集上的 AP

注意到我们的 SWSD 方法每一次的对象重定位行为受到两个参数的影响，（1）在搜索区域中提议被选择的比例r和（2）最后被选择作为包中示例的排名最高的提议的比例k。因此每个包中被选择的示例的数量为$n = r \times k \times |l|$，$|l|$表示属于包中搜索区域$l$的示例数量。图 6.10 和 6.11 分别给出了关于不同的r值和k值，算法得到的 AP 分数，图 6.11 为在 LSP/MPII-MPHB 数据集上的结果，图 6.12 为在 Pascal VOC 2007 数据集上的结果。在不用的设置下，两张图表现了相似的行为。特别地，我们可以看到算法的性能随着r值的减少而降低，说明了设置足够大

的候选集合的重要性。另一方面，k 值的影响更加复杂，这个和 IoP 的值具有亲密关系。总体来说，在两个被评估的数据集上，r 的最优值在 95.0%到 100.0%的区间内，k 的最优值大约在 35.0%到 40.0%的范围内。

6.4.3 独立阶段的贡献

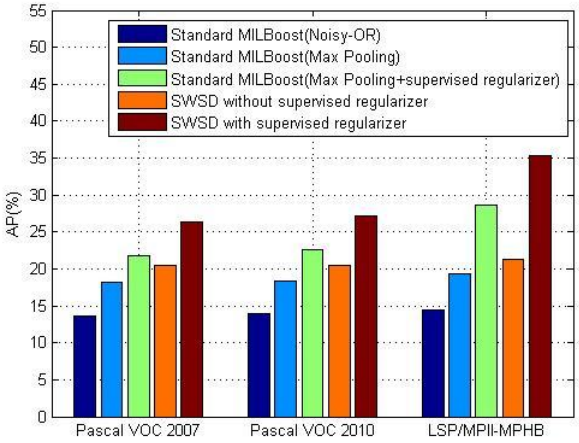

图 6.11 SWSD 算法中每个独立阶段的影响

为了探究我们所提出的 SWSD 方法在每个独立阶段的贡献，即全监督正则项，基于最大池的打包模型和约束精英的选择，我们在三个数据集上进行了一系列实验，通过轮流移除这三个主要阶段中的一个，在合适的位置留下剩下的阶段，对比项就是我们包含所有阶段的完整方法。图 6.11 给出了实验结果。一般来说，每个阶段都是有帮助的，最后的结果是各个阶段的累加。但是好像全监督正则项对于性能的提升贡献最多。特别地，如果我们除去这个阶段，在 LSP/MPII-MPHB 数据集上的性能从 35.37%剧烈下降至 21.26%，下降了 14%。类似地，只除去约束精英选择过程，我们将遭遇 8%的性能损失，当用 Noisy-OR 打包函数取代为最大池规则时，我们得到了 5%的性能提升。在 LSP/MPII-MPHB 数据集上表现可能不足够说明问题，但是我们在在 Pascal VOC 2007 和 Pascal VOC 2010 上也可以观察到全部类似的现象。

6.4.4 时间性能评估

我们对所提出的方法进行简要的时间性能分析。对于一张给定的测试图片，在得到结果之前，它经历了三个阶段，分别为提议生成，特征提取和得分评估。输入图片的大小从 480*816*3 到 1080*5760*3。采用 Selective Search 算法在我们的电脑上大约需要 4.9s 的时间为每张图片生成 1500 个检测提议，然后大概需要另外的 0.4s 来为每个提议提取 CNN 特征。最后，示例层次的分类器需要 0.06s 来评估这些提议并确定在这个提议中是否存在人类。我们可以看到，大部分的测试时间都用于特征提取，如果有时间要求，我们可以用更加有效的特征替代，例如 HoG

特征或者 SIFT 特征。注意到我们的实验基于 Matlab 平台，没有进行任何的代码优化。实际的运行时间效率可以通过其它低层次的编程语言例如 C，或者通过代码优化技术进行提高。

6.4.5 Pascal VOC 上的检测

在前面的实验中，我们已经在不同条件下证明了我们提出的方法在检测人类方面的有效性，但是我们也可以将它用于人类之外更加一般的对象检测中。为此我们在 Pascal VOC 2007 和 Pascal VOC 2010 数据集上分别进行了一组新的实验。Pascal VOC 2007 数据集已经被广泛用于评估弱监督对象检测算法的性能，但是目前只有很少的弱监督对象检测算法在 Pascal VOC 2010 数据集上进行测试。

表 6.4 各种弱监督检测方法在 Pascal VOC 2007 数据集上的 AP

Method	areo	bicy	bird	boa	bot	bus	car	cat	cha	cow	
Song[52]	27.6	41.9	19.7	9.1	10.4	35.8	39.1	33.6	0.6	20.9	
Song[53]	36.3	47.6	23.3	12.3	11.1	36.0	46.6	25.4	0.7	23.5	
Bilen[55]	42.2	43.9	23.1	9.2	**12.5**	44.9	45.1	24.9	8.3	24.0	
Wang[54]	48.8	41.0	23.6	12.1	11.1	42.7	10.9	35.5	11.1	36.6	
Wang[54]*	**48.9**	42.3	26.1	11.3	11.9	41.3	40.9	34.7	10.8	34.7	
Cinbis[11]	39.3	43.0	**28.8**	20.4	8.0	45.5	47.9	22.1	8.4	33.5	
MILBoost[76]	32.8	38.1	20.5	10.3	9.1	36.4	40.9	17.6	8.1	21.1	
SWSD(-)	39.5	45.3	22.5	20.6	11.2	42.7	49.3	30.6	12.0	31.3	
SWSD	41.1	**47.8**	23.7	**23.5**	11.4	**46.6**	**54.1**	**35.6**	**15.9**	**36.8**	
Method	dtab	dog	hors	mbik	pers	plnt	she	sofa	trai	tv	Av.
Song[52]	10.0	27.7	29.4	39.2	9.1	19.3	20.5	17.1	35.6	7.1	22.7
Song[53]	12.5	23.5	27.9	40.9	14.8	19.2	24.2	17.1	37.7	11.6	24.6
Bilen[55]	13.9	18.6	31.6	43.6	7.6	20.9	26.6	20.6	35.9	29.6	26.4
Wang[54]	18.4	**35.3**	34.8	51.3	17.2	17.4	26.8	32.8	35.1	45.6	30.9
Wang[54]*	18.8	34.4	35.4	**52.7**	19.1	17.4	**35.9**	**33.3**	34.8	**46.5**	31.6
Cinbis[11]	23.6	29.2	**38.5**	47.9	20.3	20.0	35.8	30.8	**41.0**	20.1	30.2
MILBoost[76]	13.3	17.8	28.9	41.2	13.6	16.4	24.8	20.4	34.1	19.8	23.2
SWSD(-)	22.6	26.6	31.9	46.1	20.5	22.9	28.3	24.1	37.4	32.0	29.9
SWSD	**26.1**	32.2	32.1	47.1	**26.3**	**24.1**	29.3	25.6	38.3	38.4	**32.8**

表 6.5 各种弱监督检测方法在 Pascal VOC 2010 数据集上的 AP

Method	areo	bicy	bird	boa	bot	bus	car	cat	cha	cow	
Cinbis[11]	44.6	42.3	**25.5**	14.1	11.0	44.1	36.3	23.2	12.2	**26.1**	
Bilen[55]	38.7	44.5	18.4	10.4	12.9	39.3	**42.8**	25.4	8.0	23.9	
MILBoost[76]	34.4	38.6	16.8	8.3	9.8	37.9	32.2	20.0	9.6	18.6	
SWSD(-)	39.7	48.1	23.5	14.0	14.7	48.9	33.7	30.5	14.8	23.2	
SWSD	**44.7**	**52.6**	24.5	**17.8**	**18.0**	**53.0**	34.4	**34.6**	**15.3**	24.5	
Method	dtab	dog	hors	mbik	pers	plnt	she	sofa	trai	tv	Av.
Cinbis[11]	**14.0**	**29.2**	36.0	**54.3**	20.7	12.4	**26.5**	20.3	31.2	23.7	27.4
Bilen[55]	11.0	22.9	29.9	46.2	9.9	22.6	24.9	18.5	32.1	**25.5**	25.4
MILBoost[76]	9.7	17.8	26.7	41.1	13.9	14.9	17.6	17.0	26.7	15.0	21.3
SWSD(-)	12.3	23.9	32.5	47.1	20.5	21.2	16.7	21.7	33.4	14.2	26.7
SWSD	12.5	24.7	**36.2**	48.0	**27.2**	**27.4**	20.5	**28.6**	**34.3**	17.7	**29.8**

表 6.4 和表 6.5 分别给出了在这两个数据集的测试集上全部 20 个类的对象检测性能。可以看到我们的方法在两个数据集上都实现了最高的总体性能，在 Pascal VOC 2007 数据集上的 mAP 为 32.8%，在 Pascal VOC 2010 上的 mAP 为 29.8%，mAP 是所有类上 AP 的平均值。在两个 Pascal VOC 数据集上，我们在人类检测的 AP 得分分别为 26.3%和 27.2%，这在所有被比较的方法中是最好的，而且与在 LSP/MPII-MPHB 数据集上的结果保持一致。除了人类，我们的方法也在几个其它类别上产生了最好的性能，例如猫，植物，椅子，公共汽车等。

图 6.12 展示了四种弱监督检测方法在 Pascal VOC 2007 数据集上的检测结果，黄色表示正确的检测结果，粉色表示错误的。可以看到，标准的 MILBoost[76]在一些混乱背景的图片上错过了检测目标，例如在瓶子类和沙发类。相反地，多重多示例学习方法[11]和我们的方法比 MILBoost 方法做的更好。但是在狗和瓶子的例子中，多重多示例学习方法可能对感兴趣的对象有些过检测。正如文献[11]中指出的，这可能归咎于大多数弱监督方法趋向于找到在正训练图片中出现的重复结构，但是我们的方法通过使用少量的全监督信息去正则化模型学习的行为，在一定程度上解决了这个问题。

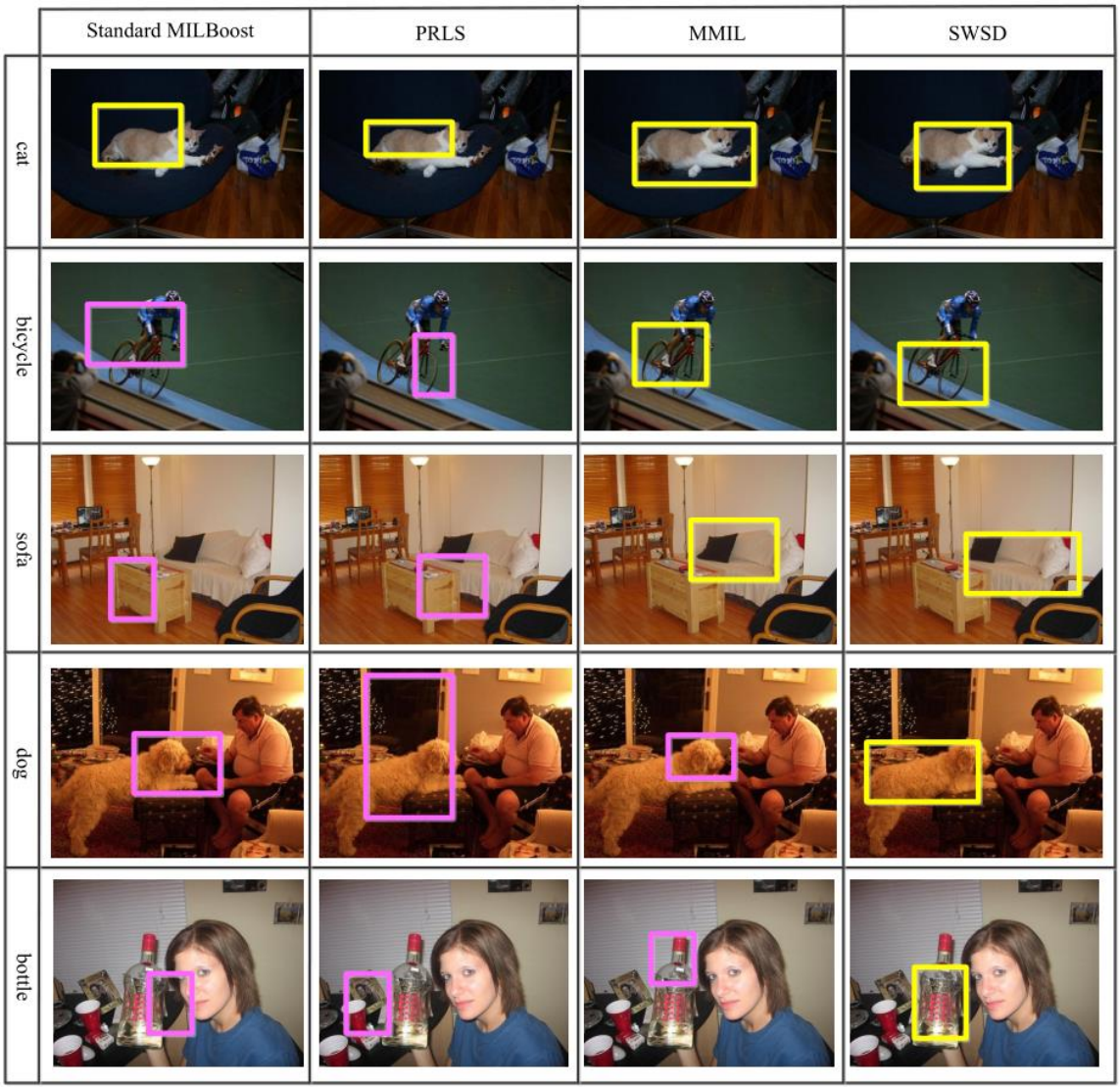

图 6.12 四种弱监督检测方法在 Pascal VOC 2007 数据集上的检测结果

6.5 本章小结

这一章我们详细阐述了我们提出的选择性弱监督检测算法以及约束精英选择方法，也介绍了我们构建的一个大规模的用于任意姿态人体检测的数据集。使用这个数据集，我们将我们提出的 SWSD 方法与最先进的弱监督检测算法在总体和各个姿态类别上进行了准确度对比，并分析了我们的人体检测结果，展示了 SWSD 算法的优越性。此外，我们还将 SWSD 方法与多重多示例检测的对象重定位行为进行了全过程的分析，探究了影响 SWSD 算法的参数设置。而且我们通过实验深度解析了 SWSD 方法的各个独立阶段对于总体性能的贡献程度，并计算了它完成一次测试图片检测所需要的时间。最后，我们在除人类之外的其它对象类别上，也证明了我们的选择性弱监督对象检测算法也是有效的。

第七章 总结与展望

7.1 工作总结

本文研究基于多示例学习的任意姿态人体检测，这是一项非常具有挑战性的工作。在弱监督学习环境下，虽然降低了对大量样本标记的要求，但也增加了训练检测模型的难度。此外，任意姿态的人体检测在前人的工作中很少被提及，专注于此的研究就更少。我们希望利用弱标记的样本，训练出一个有效的检测器，可以实现对现实中更加常见的呈现多种姿态的人体进行检测，而不是仅限于直立的人体。

我们首先要解决的问题是如何使用弱标记的样本进行训练。多示例学习提供了一个合理的思想，可以将其作为弱监督对象检测的基础，多示例学习算法自然值得我们研究。本文中，我们对多种多示例学习算法进行了分析，并总结出了每种算法的优点，缺点和适用情况。经过这些分析，我们不仅对多示例学习有了深刻的认识，也为多示例学习算法的使用提供了参考，同时为我们后面将多示例学习算法嵌入弱监督检测框架中奠定了基础。

将多示例学习用于弱监督检测中，可以将每张图片视为一个包，图片中的碎片就是包中的示例。但是为了满足多示例学习的定义，即这些碎片中至少有一个包含感兴趣的对象，我们必须提供非常大量的示例，这也为多示例学习问题带来了新的挑战。为此，我们进行了多示例检测深度评估实验，探究了多示例学习训练正包中示例的组成成分，即示例的数量和正示例的纯度，以及使用极少量的全监督信息对检测结果的影响。这对于我们设计弱监督对象检测算法至关重要，但在前人的工作中很少被研究。我们同时也发现了在全监督样本数量稀缺的情况下，全监督模型的性能会下降，多示例检测的价值和意义就得以体现。

多示例学习中一个很重要的过程是将示例层次的条件概率过渡到包层次。我们第一个发现了在示例数量较大的情况下，Noisy-OR 和 ISR 打包模型会失效，它们会引起训练过程中的梯度消失问题，从而导致对样本的使用效率降低。因此我们提议在多示例检测问题中，使用更加鲁棒的最大池规则。

深入了解了多示例检测的所有细节问题之后，我们提出了一种新的用于任意姿态下人体检测的选择性弱监督检测方法 SWSD。它是多示例学习方法的一种变形，我们使用少量的全监督样本加入到这个模型，为初始化和正则化服务。为了防止训练过程过早地锁定在错误的对象位置并提高结果检测器的鲁棒性，我们提出了一种约束精英选择方法，它可以在训练的早期阶段，在示例层次上加入更多的多样性，并且在后面的阶段，帮助检测器关注于最合适的示例。此外，我们标注了一个新的大规模多姿态人体数据集叫做 LSP/MPII-MPHB 用于任意姿态下的人体检测。

我们在新的LSP/MPII-MPHB和Pascal VOC 2007/2010数据集上评估了我们的SWSD方法，并将它与最先进的弱监督对象检测方法进行比较，证明了SWSD算法已经达到了目前最好的性能。我们也对检测算法的对象重定位行为进行了分析，甚至在目标被部分遮挡并且非常渺小的情况下，显示出我们的方法能够在训练图片中逐渐定位到感兴趣的对象。

7.2 未来展望

在弱监督对象检测领域，多示例学习是最有效的框架。但是多示例学习中，存在着太多的不确定性，关于如何消除这些疑惑，各种多示例学习算法采用了不同的方法，但不可避免的是迭代训练检测器的过程，而迭代的关键是通过当前决定下一步，如果初始化就是错误的，在过程中又无法纠正错误，就会导致最后的失败。在初始化时，往往很难找到正确的对象位置，否则迭代过程也就不需要了，所以最主要是在过程中不断纠正错误。这对样本的多样性提出了很高的要求，在检测器无法看到自身问题的时候，就需要新的训练样本的提醒。我们提出了约束精英选择来改变训练样本，未来，我认为就如何提高样本的多样性可以有更多的猜想。

对于任意姿态的人体检测，目前的研究还比较少，相应的多姿态人体图片也很稀缺。虽然我们提供了一个大规模的数据集，其中也包含了多种姿态的人体，但是各个姿态之间的数量分布不均匀，这对于那些图片数量较少的姿态而言，研究的难度比较大。在将来，可以考虑丰富这个数据集，如果可以的话，也可以为每个姿态单独训练一个姿态检测器，在对人体进行检测的同时也可以判断出人体的姿态。

参考文献

[1] Enzweiler M, Gavrila D M. Monocular pedestrian detection: survey and experiments.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009, 31(12):2179-2195.

[2] Peng P, Tian Y, Wang Y, et al. Robust multiple cameras pedestrian detection with multi-view Bayesian network[J]. Pattern Recognition, 2015, 48(5):1760-1772.

[3] Zhang Z, Tao W, Sun K, et al. Pedestrian detection aided by fusion of binocular information[J]. Pattern Recognition, 2016, 60:227-238.

[4] Andriluka M, Pishchulin L, Gehler P, et al. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis[C]// Computer Vision and Pattern Recognition. 2014:3686-3693.

[5] Zhou Y, Bai X, Liu W, et al. Similarity Fusion for Visual Tracking[J]. International Journal of Computer Vision, 2016, 118(3):337-363.

[6] Girshick R, Donahue J, Darrell T, et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 38(1):1-1.

[7] Ray S, Craven M. Supervised versus multiple instance learning: an empirical comparison[C]// International Conference on Machine Learning. ACM, 2005:697-704.

[8] Blum, A., & Kalai, A. (1998). A note on learning from multiple-instance examples. Machine Learning, 30(1), 23-29.

[9] Yang, J. (2005). Review of multi-instance learning and its applications. Tech. Rep.

[10] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles[J]. Artificial Intelligence, 1997, 89(1–2):31-71.

[11] Cinbis R G, Verbeek J, Schmid C. Weakly Supervised Object Localization with Multi-fold Multiple Instance Learning[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015:1-1.

[12] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes challenge 2007 (voc 2007) results (2007)[J]. 2008.

[13] Keeler J D, Rumelhart D E, Leow W K. Integrated segmentation and recognition of hand-printed numerals[C]// Conference on Advances in Neural Information Processing Systems. Morgan Kaufmann Publishers Inc. 1990:557-563.

[14] Everingham M, Gool L V, Williams C K I, et al. The Pascal, Visual Object Classes (VOC) Challenge[J]. International Journal of Computer Vision, 2010, 88(2):303-338.

[15] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[16] Johnson S, Everingham M. Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation[C]// British Machine Vision Conference, BMVC 2010, Aberystwyth, UK, August 31 - September 3, 2010. Proceedings. 2010:1-11.

[17] Dietterich T G, Lathrop R H, Lozano-Perez T. Solving the multiple instance problem with axis-parallel rectangles[J]. Artificial intelligence, 1997, 89(1): 31-71.

[18] Maron O, Lozano-Perez T. A framework for multiple-instance learning[J]. Advances in neural information processing systems, 1998: 570-576.

[19] Zhang Q, Goldman S A. EM-DD: An Improved Multiple-Instance Learning Technique[C]//Advances in Neural Information Processing Systems. 2001: 1073-1080.

[20] Andrews S, Tsochantaridis I, Hofmann T. Support Vector Machines for Multiple-Instance Learning[C]//NIPS. 2002: 561-568.

[21] Wang J, Zucker J D. Solving multiple-instance problem: A lazy learning approach[J]. 2000.

[22] Maron O, Ratan A L. Multiple-Instance Learning for Natural Scene Classification[C]//ICML. 1998, 98: 341-349.

[23] Zhang Q, Goldman S A, Yu W, et al. Content-based image retrieval using multiple-instance learning[C]//ICML. 2002, 2: 682-689.

[24] Maron O. Learning from ambiguity[D]. Massachusetts Institute of Technology, 1998.

[25] Tong, S., & Koller, D. (2002). Support vector machine active learning with applications to text classification. The Journal of Machine Learning Research, 2, 45-66.

[26] Tao Q, Scott S, Vinodchandran N V, et al. SVM-based generalized multiple-instance learning via approximate box counting[C]//Proceedings of the twenty-first international conference on Machine learning. ACM, 2004: 101.

[27] Pathak D, Shelhamer E, Long J, et al. Fully Convolutional Multi-Class Multiple Instance Learning[J]. Computer Science, 2015.

[28] Wu J, Yu Y, Huang C, et al. Deep multiple instance learning for image classification and auto-annotation[C]// Computer Vision and Pattern Recognition. IEEE, 2015:3460-3469.

[29] Li B, Xiong W, Hu W. Web Horror Image Recognition Based on Context-Aware Multi-instance Learning.[J]. Image Processing IEEE Transactions on, 2015, 24(12):1158-1163.

[30] Song H, Zhu Z, Wang X. Multiple Instance Learning with Bag Dissimilarities[J]. Computer Science, 2015.

[31] Melendez J, Van G B, Maduskar P, et al. On Combining Multiple-Instance Learning and Active Learning for Computer-Aided Detection of Tuberculosis.[J]. IEEE Transactions on Medical Imaging, 2015, 35(4):1-1.

[32] Melendez J, Van G B, Maduskar P, et al. A novel multiple-instance learning-based approach to computer-aided detection of tuberculosis on chest x-rays.[J]. IEEE Transactions on Medical Imaging, 2015, 34(1):179-92.

[33] Vanwinckelen G, Vinicius T D O, Fierens D, et al. Instance-level accuracy versus bag-level accuracy in multi-instance learning[J]. Data Mining and Knowledge Discovery, 2016, 30(2):313-341.

[34] Papandreou G, Kokkinos I, Savalle P A. Modeling local and global deformations in Deep Learning: Epitomic convolution, Multiple Instance Learning, and sliding window detection[C]// Computer Vision and Pattern Recognition. IEEE, 2015:390-399.

[35] Li W, Vasconcelos N. Multiple instance learning for soft bags via top instances[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2015:4277-4285.

[36] Chen L, Tong T, Ho C P, et al. Identification of Cerebral Small Vessel Disease Using Multiple Instance Learning[M]// Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015. Springer International Publishing, 2015.

[37] Cano A, Zafra A, Ventura, Sebasti&#. Speeding up multiple instance learning classification rules on GPUs[J]. Knowledge and Information Systems, 2015, 44(1):127-145.

[38] Bandyopadhyay S, Ghosh D, Mitra R, et al. MBSTAR: multiple instance learning for predicting specific functional binding sites in microRNA targets.[J]. Scientific Reports, 2015, 5:8004-8004.

[39] Xu C, Tao W, Meng Z, et al. Robust visual tracking via online multiple instance learning with Fisher information[J]. Pattern Recognition, 2015, 48(12):3917-3926.

[40] Zhang D, Meng D, Li C, et al. A Self-Paced Multiple-Instance Learning Framework for Co-Saliency Detection[C]// IEEE International Conference on Computer Vision. IEEE, 2015:594-602.

[41] Wang Z, Yoon S, Xie S J, et al. Visual tracking with semi-supervised online weighted multiple instance learning[J]. The Visual Computer, 2016, 32(3):1-14.

[42] JF Ruizmuñoz，MO Alzate，G Castellanosdominguez, et al. Multiple instance learning-based birdsong classification using unsupervised recording segmentation[C]// International Conference on Artificial Intelligence. AAAI Press, 2015.

[43] Rastegari M, Hajishirzi H, Farhadi A. Discriminative and consistent similarities in instance-level Multiple Instance Learning[C]// Computer Vision & Pattern Recognition. IEEE, 2015:740-748.

[44] Shrivastava A, Patel V M, Pillai J K, et al. Generalized Dictionaries for Multiple Instance Learning[J]. International Journal of Computer Vision, 2015, 114(2):288-305.

[45] Cheplygina V. Dissimilarity-Based Multiple Instance Learning[J]. Machine Learning, 2015.

[46] Ren W, Huang K, Tao D, et al. Weakly Supervised Large Scale Object Localization with Multiple Instance Learning and Bag Splitting.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(2):405-416.

[47] Maken F A, Gal Y, Mcclymont D, et al. Multiple Instance Learning for Breast Cancer Magnetic Resonance Imaging[C]// International Conference on Digital Lmage Computing: Techniques and Applications. IEEE, 2015:1-8.

[48] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object Detection with Discriminatively Trained Part-Based Models[J]. IEEE Transactions on Software Engineering, 2014, 32(9):1627-45.

[49] Pandey M, Lazebnik S. Scene recognition and weakly supervised object localization with deformable part-based models[J]. 2011, 23(5):1307-1314.

[50] Siva P, Tao X. Weakly supervised object detector learning with model drift detection[C]// IEEE, 2011:343-350.

[51] Russakovsky O, Lin Y, Yu K, et al. Object-Centric Spatial Pooling for Image Classification[C]// European Conference on Computer Vision. Springer-Verlag, 2012:1-15.

[52] Song H O, Girshick R, Jegelka S, et al. On learning to localize objects with minimal supervision[J]. Eprint Arxiv, 2014:1611-1619.

[53] Song H O, Yong J L, Jegelka S, et al. Weakly-supervised Discovery of Visual Pattern Configurations[J]. Advances in Neural Information Processing Systems, 2014, 2:1637-1645.

[54] Wang C, Ren W, Huang K, et al. Weakly Supervised Object Localization with Latent Category Learning[M]// Computer Vision – ECCV 2014. Springer International Publishing, 2014:431-445.

[55] Bilen H, Pedersoli M, Tuytelaars T. Weakly Supervised Object Detection with Posterior Regularization[C]// The British Machine Vision Conference. 2014:1997-2005.

[56] Papageorgiou C, Poggio T. A trainable system for object detection[J]. International Journal of Computer Vision, 2000, 38(1): 15-33.

[57] Viola P, Jones M J. Robust real-time face detection[J]. In-ternational journal of computer vision, 2004, 57(2): 137-154.

[58] Felzenszwalb P F, Girshick R B, McAllester D. Cascade object detection with deformable part models[C]//Computer vision and pattern recognition (CVPR), 2010 IEEE confer-ence on. IEEE, 2010: 2241-2248.

[59] Van de Sande K E A, Uijlings J R R, Gevers T, et al. Seg-mentation as selective search for object recogni-tion[C]//Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 1879-1886.

[60] Uijlings J R R, van de Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154-171.

[61] Hosang J, Benenson R, Dollár P, et al. What makes for ef-fective detection proposals?[J]. arXiv preprint arXiv:1502.05082, 2015.

[62] Arbelaez P, Pont-Tuset J, Barron J, et al. Multiscale combi-natorial grouping[C]//Computer Vision and Pattern Recog-nition (CVPR), 2014 IEEE Conference on. IEEE, 2014: 328-335.

[63] Alexe B, Deselaers T, Ferrari V. Measuring the objectness of image windows[J]. Pattern Analysis and Machine Intelli-gence, IEEE Transactions on, 2012, 34(11): 2189-2202.

[64] Van den Bergh M, Roig G, Boix X, et al. Online video seeds for temporal window objectness[C]//Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013: 377-384.

[65] Zhu Q, Yeh M C, Cheng K T, et al. Fast human detection using a cascade of histograms of oriented gradi-ents[C]//Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, 2: 1491-1498.

[66] Ke Y, Sukthankar R. PCA-SIFT: A more distinctive repre-sentation for local image descriptors[C]//Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, 2: II-506-II-513 Vol. 2.

[67] Tan X, Triggs B. Fusing Gabor and LBP Feature Sets for Kernel-Based Face Recognition[M]// Analysis and Modeling of Faces and Gestures. Springer Berlin Heidelberg, 2007:235-249.

[68] Howarth P, Rüger S. Evaluation of texture features for con-tent-based image retrieval[M]//Image and Video Retrieval. Springer Berlin Heidelberg, 2004: 326-334.

[69] Schmidhuber J. Deep learning in neural networks: An over-view[J]. Neural Networks, 2015, 61: 85-117.

[70] Sermanet, Pierre, et al. "Overfeat: Integrated recognition, localization and detection using convolutional networks." arXiv preprint arXiv:1312.6229 (2013).

[71] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[72] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// Computer Vision and Pattern Recognition. IEEE, 2014:1-9.

[73] Simonyan K, Zisserman A. Very deep convolutional net-works for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[74] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 248-255

[75] Babenko B, Yang M H, Belongie S. Robust Object Tracking with Online Multiple Instance Learning[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 33(8):1619-1632.

[76] Viola P A, Platt J C, Zhang C. Multiple Instance Boosting for Object Detection.[J]. Advances in Neural Information Processing Systems, 2005, 18:1419--1426.

[77] Lee V E, Ning R, Jin R, et al. A Survey of Algorithms for Dense Subgraph Discovery[M]// Managing and Mining Graph Data. 2010:303-336.

[78] Uijlings J R R, Sande K E A V D, Gevers T, et al. Selective Search for Object Recognition[J]. International Journal of Computer Vision, 2013, 104(2):154-171.

[79] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[80] Yang Y, Ramanan D. Articulated Human Detection with Flexible Mixtures of Parts[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(12):2878-90.

致 谢

两年半的硕士研究生学习生活转眼间已进入了尾声，我南京航空航天大学的学习生活了六年半的时间，这里的每一个人，每一处风景都使我印象深刻。值此毕业论文成稿之际，我谨向所有关心、帮助、爱护我的人表示最诚挚的感谢和最忠心的祝福。

首先感谢这些年来，教给我知识的老师们，是你们让我从一个对计算机一无所知的人逐渐变成一名程序员。尤其要感谢我的研究生导师谭晓阳教授，感谢谭老师选择了我当他的研究生，能够让我踏入学术的殿堂。研究生阶段，谭老师指引了我的研究方向，在学术上给予了我极大的帮助。尤其在指导我写小论文期间，谭老师付出了极大的心血，让我学会并成长了许多。在此真诚的感谢谭老师的悉心指导，祝愿谭老师工作顺利，身体健康。

感谢 ParNeC 研究组具有敏锐学术眼光的 Leader 陈松灿教授、执着思辨的张道强教授以及认真勤勉的刘学军教授，你们的言传身教激励着我，让我受益匪浅。感谢刘大琨，金鑫，王冬，张衡，陈骁师兄和秦晓倩师姐，你们是我在学术路上的标杆，感谢你们为我答疑解惑，带领我一步步前进。感谢同门王宇辉和宋歌，和你们一同起学习一同科研的日子十分难忘。感谢孙强、刘程、张文师弟和赵轩师妹，有你们的日子十分快乐，怀念和你们一起科研一起玩耍的时光。

当然还要感谢我的家人和同学，因为你们的支持和帮助，我才能生活的无忧无虑，克服所有困难，顺利完成学业。

在学期间的研究成果及发表的学术论文

攻读硕士学位期间发表（录用）论文情况

1. 蔡雅薇，谭晓阳，弱监督任意姿态人体检测，计算机科学与探索，2016

2. Yawei Cai, Xiaoyang Tan, Weakly Supervised Human Body Detection under Arbitrary Poses, in International Conference on Image Processing. IEEE, 2016