# MedLSVM 及其在结构化物体检测中的应用

（申请清华大学工学硕士学位论文）


培 养 单 位 ： 计算机科学与技术系

学 科 ： 计算机科学与技术

研 究 生 ： 王 鹏

指 导 教 师 ： 张 钹 教 授


二〇一二年五月

MedLSVM及其在结构化物体检测中的应用

王鹏

# MedLSVM And Its Application To Structural Object Detection Models

Thesis Submitted to

**Tsinghua University**

in partial fulfillment of the requirement

for the degree of

**Master of Science**

in

**Computer Science and Technology**

by

**Wang Peng**

Thesis Supervisor： Professor Zhang Bo

**May, 2012**

# 关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：

清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容。

本人保证遵守上述规定。

**（保密的论文在解密后应遵守此规定）**

作者签名：_____　　导师签名：_____

日　　期：_____　　日　　期：_____

# 摘 要

本文提出了MedLSVM分类器模型及其学习算法。它为含有隐变量的最大间隔学习问题提供了统一的概率框架。据我们所知它是第一个使用概率方法和最大熵原则来处理最大间隔分类器中的隐变量的尝试。我们发现了现有含隐变量分类器的重要区别，提出了对称与非对称含隐变量分类器的概念。我们尝试了多种设计思路，设计了一套图像像素与物体部件隐变量的概率产生模型。我们也为模型参数引入了概率分布，并设计了一种方向分布与径向分布分离的概率分布形式，使得MedLSVM的最优化问题在数学上可解。我们将这个新的模型使用在了基于部件的结构化物体检测问题上，并且为了适应物体检测问题的速度要求，而推导出了快速softmax算法和快速冲击分布算法两种快速实现。它们在当前最具挑战性的物体检测数据集上得到了与当前最好方法可比拟的性能。

本文的创新点主要有：

- 提出了MedLSVM，首次使用概率方法和最大熵原则处理最大间隔方法中的隐变量；
- 设计了观测数据与隐变量的概率产生模型以及方向与径向分离的参数分布形式；
- 提出了两种用于结构化物体检测的MedLSVM的快速算法。

**关键词：**物体检测 部件模型 MED SVM LSVM MedLSVM

# Abstract

In the paper we designed the MedLSVM classification model and its learning algorithm. It provides a unified probabilistic framework for maximum-margin classification models which have latent variables. To our knowledge it is the first attempt to employ probabilistic methods and maximum entropy principle to deal with latent variables in maximum-margin classifiers. We recognized an important difference among existing classifiers with latent variables, and introduced the concepts of Symmetric LSVM and Asysmmetric LSVM. We explored many design options, proposed a probabilistic generating model below image pixels and the latent variables in the object model. We also introduced probabilistic distributions for the classifier's model parameters, and designed a orientation-radius separated distribution form for them, in order to make the optimization problem of MedLSVM trackable. We used this new model in the part-based structural object detection problem, and in order to meet the speed requirement in object detection tasks, we developed two fast algorithms for it, one called Fast-Softamx and the other called Fast-Impulse-Distribution. We achieved comparable performance with regard to the state-of-the-art method on the most challenging object detection dataset.

**Key words:** Object Detection Part-based Model MED SVM LSVM MedLSVM

# 目 录

# 主要符号对照表

| | |
|---|---|
| $\mathbb{E}(X)$ | 随机变量$X$的期望 |
| $\mathcal{N}(X\|\mu,\sigma^2)$ | 随机变量$X$服从期望为$\mu$标准差为$\sigma$的高斯分布 |
| $\|\|$ | 向量2范数或向量的拼接 |
| $\mathrm{KL}(q(x)\|\|p(x))$ | 概率分布密度函数$q(x)$和$p(x)$的KL距离 |
| $u \cdot v$ | 向量u和v的点击 |
| $A \otimes B$ | 矩阵A与B的反向二维卷积 |
| $A \odot B$ | 矩阵A与B的元素级相乘 |
| $H(p(x))$ | 概率分布$p(x)$的熵 |

# 第 1 章 序言

## 1.1 问题描述

本文主要研究物体检测问题。所谓物体检测，是指在一幅图片或视频帧中找到待检测物体的位置和大小。为完成这种检测，一般需要预先设计检测器模型和训练出检测器，然后在待检测图片上运行。在本文中，我们所研究和设计的检测器模型为一般物体检测器模型，即模型本身不针对特定物体或物体类别，而是可以用来训练出各种物体类别的检测器。我们在本文中将在一种称为基于部件的可形变物体模型的基础上，设计出我们称为MedLSVM的检测器模型及其学习算法。这种模型结合了最大熵方法与带隐变量的SVM（LSVM）两条研究思路，是第一个用概率方法处理最大间隔分类中隐变量的尝试。

## 1.2 研究意义

物体检测问题是最重要的计算机视觉问题之一。它的用途非常广泛，例如可以用于视频监控、自动驾驶、机器人、自动对焦等等任务中。可以说，作为人类最普通的视觉任务，物体检测是一切拟人系统的基石之一，是最近似于人的信息输入方式。但目前，要将计算机视觉的物体检测方法用于实用系统中，还存在着许多困难。视觉算法的不稳定性，使得它在自动驾驶、机器人等任务中还无法取代激光雷达技术；它的速度与运算量，使得它难以实现在嵌入式设备上；即使在预先选定的数据集中，如果物体变化较大，一般物体检测模型很难达到令人满意的正确率。一些特定物体的检测算法已经达到了可实用的水平，例如人脸检测、车辆检测等。但对于其它大部分的物体检测问题，其研究需求仍非常迫切，尤其是需要处理变化较大、形变较大的情况。

物体检测的现有研究成果已显现出了巨大的经济价值。例如，现在的主流数码相机厂商都将人脸检测辅助对焦功能作为旗下产品的重要卖点，这项功能产生的商业附加值，考虑到数码相机的市场规模，将十分巨大。又例如，北京全市道路交通系统现在已安装数万只视频摄像头，正在规划中的车辆自动检测系统如果在这数万只摄像头的数据流上实时运行，将为北京提供实时、准确、

全面且细至街道的交通流量状态。再例如，微软公司推出的结合了人体检测、姿势识别等技术的游戏主机外设Kinect在上市首月就达到了千万级的销量，盈利丰厚。同时，物体检测技术对国防安全也有着重大意义。机器人技术、无人驾驶技术（飞机及车辆）、导弹制导技术等国防领域的关键先进技术都可以从视觉物体检测的研究成果中获益。

## 1.3 相关工作

### 1.3.1 物体检测

在计算机视觉领域里，所谓“物体检测”是指通过计算机视觉方法，在一幅图片或视频帧中，找到待检测物体的位置和大小（一般以包围矩形bounding box）来表示。物体检测算法有特定物体检测算法和一般物体检测算法之分。特定物体检测算法，是指该算法只用于（或只适用于）检测一种类别的物体，它往往使用了这一类物体的某个特有属性，使得它一般不能用于检测其它类别的物体。一般物体检测算法，是指该算法可以被用于检测许多种类别的物体。它往往只依靠一个一般化的模型，而不使用某一类特定物体的特征。在使用时，依训练数据的不同，该算法可以被用于检测不同的物体。例如，同为基于HOG特征和SVM分类器的物体检测算法，当使用一组含行人的图片作为训练正例时，可以得到一个行人检测器；而同样的训练算法，当使用一组含车辆的图片作为训练正例时，则可以得到一个车辆检测器。

近15年来，应用于几个重要物体类别的特定物体检测算法得到了长足的发展，不仅有效地解决了其对应的特定物体类别的检测问题，而且其中提出的思想和技术也被广泛地应用到了其它类别物体的检测，以及一般物体检测算法领域。取得巨大进展的几个重要物体类别包括人脸、行人、车辆等。在人脸检测领域，早期的算法主要使用主分量分析（PCA）技术，从训练用的人脸图片集中得到几个有代表性的人脸模板（称为“特征脸（Eigen Face）”），然后使用模板匹配或分类技术，使用这几个代表性人脸模板在待检测的图片中寻找人脸出现的位置和大小。这种方法在速度和准确度上都难以令人满意。[1]提出了基于Haar小波的Cascade检测器算法，使得人脸检测的速度和准确度（尤其是速度）得到了巨大提升。基于Haar小波的Cascade检测器使用一种层级结构，将能

力强弱不同的多个检测器组成一个流水线，能力弱的检测器位于流水线的前端，能力强的检测器位于流水线的后端。当一个待检测的图片窗口被送交给流水线时，它会一级一级地尝试通过各个层级的检测器。一旦它被某个层级的检测器判定了负例，它将直接被判定为负例而无法到达下一级。这种结构的好处是，能力弱的检测器往往速度也较快，于是许多负例可以在流水线的早期阶段被快速地排除，只有最终会被判定为正例的那些图片窗口，才会完整地通过整个流水线而花费完整的运行时间。而且，在物体检测中，往往使用滑动窗口技术，这使得将会被判定为负例的图片窗口的数量远远多于将会被判定为正例的窗口的数量。所以，缩短了排除一个负例窗口平均所用的时间，将会大大提高检测器在整张图片上的运行速度。Viola Jones等人提出的这种层级分类器的思想，后来被广泛地应用到了各种物体检测任务中。

行人检测领域，对整个物体检测研究社区的最大贡献，则是该领域的研究人员提出的Histogram Of Gradient（HOG）特征。Dalah在2005年提出了这一特征[2]，并首次将它用于行人检测。HOG特征的基本思想是使用图片像素值梯度的局部统计量来表征图像的特征。它将图片划分为固定大小的、不相交的网格。在每个格子内，它首先计算每个像素位置的像素值梯度（像素值梯度可以由计算机视觉算法近似得到），然后统计这些梯度向量的方向的分布。具体来说，它首先将整个方向空间（360度或180度）量化为固定数量（如8个）的桶，然后统计每个桶（方向区间）有多少个方向向量落入其中。每个方向向量在为它应属于的桶投票时，以它的向量长度作为投票权重。当投票结束时，我们得到了n个实数（n为桶数），表示每个桶落入了多少个方向向量（受向量长度加权）。于是，对于每个HOG特征的格子，我们得到了一个长度为n的实向量。如果整个图片被划分成了m个格子，我们将这m个n维向量拼接在一起，就得到了一个长度为mn的向量。这个向量就是这幅图片的HOG特征。 HOG特征的特点，一是它使用了像素值梯度的信息，而不是像素值本身，从而使得它不会随着整个图片的亮度变化而变化，从而具有了一定程度的光照不变性；二是它使用了局部的梯度向量的统计量，而不是梯度向量本身，从而能够抵御一定程度的误差和形变，具有了一定程度的形变不变性；三是它是局部特征，即从图片的各个局部提取特征，然后拼接在一起组成整幅图片的特征，而不是对整幅图片做统计从而得到一个特征向量（整体特征)，如颜色直方图特征等，从而

在一定程度上反应了图片的空间布局，可以用来刻画物体的形状、外貌等与空间布局有关的信息。HOG特征在提出后就被广泛应用到了各个领域，而且，由于它的提取和计算速度较快（相较于SIFT[3] 等特征），HOG特征成为了物体检测这种对速度要求较高的领域里的通用特征。

在一般物体检测领域，近年来最重要的工作是[4]提出的基于部件的、以最大间隔原则训练的结构化物体模型。该模型将待检测物体表示为主模板与多个子部件组成的星形结构，主模板描述物体的整体外观，子部件捕捉物体的局部细节。子部件相对于主模板的位置可以变化，从而实现了可形变物体模型。本文的工作将以[4]为基础，后文中还会详细介绍该方法。还有很多其它的一般物体检测的模型。Leo Zhu等作者[5]提出了使用概率语法和与或树来描述可形变的物体模型，并使用同现次数等指标来无监督地学习出该模型。 Songchun Zhu等作者[6]提出了使用可局部移动的Gabor线段来描述可形变的物体。[7]在[4]的基础上做了三层的部件模型。 [8]在 [4]的基础上提出了对两个物体的组合做联合学习和检测的思路。另外，在特征的层面上，[9]尝试了同时使用HOG特征和LBP特征。另外还有一些研究工作[10]尝试利用上下文信息来帮助物体检测任务。

### 1.3.2 最大间隔学习

在分类器设计和学习领域，最大间隔分类成为了近年来最主流的研究方向，也成为了其它领域里使用最多的分类器方法。最大间隔分类器研究的起点和方法得名源自支持向量机（SVM）方法[11]。最简单的情况下，它是一个线性二分类器。“最大间隔”的称法，源于它试图以这样一种方法决定分类面，使得分类面能够完全分开两类的样本，而且分类面距离最近的样本点（称为“支持向量机”）的距离（称为“间隔”）最大。在样本点无法被线性分开的情况下，SVM又引入了两条重要的扩展。一是使用松弛向量方法放松了训练样本必须被正确分开的约束。二是将向量的点积推广为一大类满足一定条件的二元泛函，称为“核”[12]。

在SVM被提出后，有一些对于SVM的功能上的扩展被陆续提出。例如SVM是一个二分类器，无法（直接）进行多分类。[13]提出了用于多分类的Multi-class SVM。[14]提出了能够给训练样例赋予重要性权重的cost-sensitive

SVM。并且，在扩展SVM的研究中，研究者们逐渐将SVM设计背后的最大间隔原则提取出来并形成了Hinge Loss的概念以及如何使用最大间隔原则为任何一种分类准则设计最大间隔分类器的一般方法[15]。在Multi-class SVM以及Hinge Loss研究工作的基础上，[16]提出了一种加速SVM训练的方法，使得Multi-class SVM能用于类别非常非常多（与某一个参数呈指数关系）的情况，从而可以处理类别标签是一种有结构的数据的情况，称为Struct-SVM。

为了处理分类问题中经常出现的隐变量，即在分类准则中使用到、但在训练数据中没有提供其值的变量，[4]和[17]从不同的研究背景出发都提出了所谓Latent SVM。但它们实际上并不等价并有着重要的区别，这是本文将要重点提出的一条重要观察。但从形式上，它们都试图在分类准则中使用max运算来隐藏掉隐变量，使它不以自由变量的形式出现在分类准则和学习最优化问题中。这样做的代价是原本凸的优化问题变得不凸，从而必须引入迭代优化策略来寻找局部最优值。

在另外一条道路上，从概率和统计学习背景出发的研究者，试图为原本纯粹是最优化问题而没有任何概率背景的SVM方法找到一种概率背景，从而使得最大间隔方法和概率统计学习这两条重要的研究路线得以结合。这种努力的最早成果是Jaakkola等作者提出的最大熵辨别式学习（MED）[18]。它为SVM的模型参数引入概率分布，将SVM对参数的点估计改变为对其的概率分布估计，将SVM学习问题中原本是对参数的优化改变为对参数的概率分布的优化。为了正则化约束这个待优化的概率分布以防其陷入过拟合，MED方法使用待优化概率分布和一个预先给定的分布（先验分布）的KL距离来约束待优化分布。由于KL距离又称为相对熵，MED的名称由此而来。

以MED框架为基础，Jun Zhu等作者发展出了多种针对不同问题的概率最大间隔学习模型[19–21]。它们的主要区别是使用了不同的分布家族来作为MED中的先验分布。与本文联系最紧密的是其中的iSVM[19]。它使用Dirichlet过程（DP）[22]作为MED中的先验分布。DP过程可以随机产生一个和为1的无限非负数列，这个数列可以作为一个离散分布的概率分布函数。当我们把这个离散分布作为从无限多个混合模型进行选取时使用的选取概率时，我们就得到了一个有无限多个部件的混合模型。当它使用在MED中时，它试图将样本分为多个聚类，对每个聚类学习出分类面。它相比于朴素地先聚类再训练SVM的方法的

优势是将聚类与分类联合优化，使得分类器会对聚类结果有影响，而且整个问题有一个定义明确的优化目标。它相比于有限混合模型SVM[23,24]的优势是不需要预先指定混合模型个数，这个个数会被自动决定，而且没有上限。

## 1.4 研究方法与本文结构

本文的研究使用变分估计、贝叶斯分析、拉格朗日乘子等数学方法，为我们要设计的分类器模型建立一套严格的概率框架及其最优化方法。然后结合具体问题，即我们要解决的结构化物体检测问题，使用动态规划、分治法、快速傅里叶变换等技术来设计出我们的数学模型的快速算法。最后在公开数据集上进行试验得到经验结果，并与当前最好的结果做比较。

本文的结构如下。我们将首先介绍我们要处理的物体检测任务，并详细描述我们要使用的基于部件的物体模型。我们要主要研究的隐变量将定义自该模型，而且后文要推导出的各种算法都是为了学习该物体模型（但这些算法可以被推广）。随后我们将介绍本文提出的MedLSVM分类器模型及其学习算法。之后我们将介绍一系列的实验结果，以从各个方面考察MedLSVM方法的性能和特点。

# 第 2 章　基于部件的物体检测模型

## 2.1　本章概述

基于部件的物体检测模型由Pedro Felzenszwalb等人于2008年提出[4]。它的核心思想是用一个主模板和多个部件模板来描述一个物体。当用于滑动窗口检测时，对于每一个图像窗口，该窗口在该模型下的得分，由主模板与整个窗口的内积值、各个部件模板与它们所在位置的图片局部的内积值、以及描述各部件位置与其期望位置（锚点）相对距离的形变惩罚量相加得到。对于每一个窗口，检测算法试图为每一个部件选择一个最好的位置，使得该窗口在该模型下的得分最高。最终，图像窗口的得分被与一个训练所得的阈值做比较，来决定将它分类为正例（待检测物体）还是负例。这种模型的优点是，当主模板与部件模板运行在图片的不同分辨率上，尤其是当部件模板运行在更高的分辨率上，则部件模板可以抓住物体细节特征，而主模板则用于刻画物体的整体外观。而且该模型允许部件自由移动，从而允许物体具有一定程度的形变，实现了检测模型的形变不变性。本章的模型、学习算法以及快速检测算法均由[4]提出，本文对模型进行了重新数学表述，提出了两种隐变量视角，以融入后文提出的MedLSVM框架。

## 2.2　模型描述

### 2.2.1　图像的特征表示

表示一幅图像，既可以用它的原始像素值，也可以使用某一种经过计算后提取出的特征，以期达到某种不变性。[4]使用了HOG特征。它首先用计算机视觉相关方法计算图片各像素位置的像素值梯度向量，然后将图片划分成固定大小的不相交的格子，在每个格子内，它统计各个像素位置的梯度向量的方向的直方图。统计时，它将整个方向空间（圆周或半圆周）划分为固定数量的桶，然后将每个梯度向量根据其方向放入对应的桶中，放入时以其向量长度作为其对该桶贡献的大小。假设方向空间被划分成了p个桶，那么经过统计后，

每个格子得到了p个实数，代表各个桶接收到的梯度向量个数（经过长度加权）。也就是说，在每个格子上，我们提取出了一个p维的特征向量。如果以格子标号作为下标，那么在一幅图像上，我们会得到一个Feature Map。假设一幅图片在高度方向被划分成m格，在宽度方向被划分成了n格，即总共被划分成了mn格，那么在这幅图像上可以得到一个Feature Map，表示为$F_{m,n}(i,j)$（或简记为$F(i,j)$），表示第i行第j列（编号从1开始）的格子上提取得到的p维HOG特征向量。

本文在此引入一个向量化表示方法。对于一个Feature Map $F(\cdot,\cdot)$，我们以格子$(i,j)$为左上角，在高度方向取h个格子，在宽度方向取w个格子，得到一个含有hw个格子的局部Feature Map，记为$\hat{F}(i,j,h,w)$。我们再将这些格子对应的各个p维HOG特征向量依次首尾拼接，得到这个局部Feature Map的向量化表示，记为$\Phi_F(i,j,h,w)$。即：

$$\begin{aligned}\Phi_F(i,j,h,w) \triangleq\ & F(i,j)\|F(i,j+1)\|\cdots\|F(i,j+w-1)\|F(i+1,j) \\ & \|F(i+1,j+1)\|\cdots\|F(i+h-1,j+w-1)\end{aligned} \tag{2-1}$$

其中||表示向量的拼接。

### 2.2.2　检测模板

一个用于检测的物体模板，可以视为是一个固定大小的Feature Map。当把它放置于图像的Feature Map的某一位置时，它在该图像该位置的得分（或者说该图像在该位置该模板下的得分），是将它的向量化表示，点乘他所覆盖的图像局部Feature Map的向量化表示，得到的实数值。形式化地，假设一个模板的高为h，宽为w，记为$T_{h,w}(\cdot,\cdot)$，一幅图像的Feature Map的高为m，宽为n，记为$F_{m,n}(\cdot,\cdot)$，那么当把模板的左上角放置在图像Feature Map的$(y,x)$号格子时，该模板在该图像的该位置的得分为：

$$K(F,T,y,x) \triangleq \sum_{i=1}^{h}\sum_{j=1}^{w} F(y+i-1,x+i-1)\cdot T(i,j)$$

如果用向量化表示方法，则表示为：

$$K(F,T,y,x) = \Phi_F(y,x,w,h) \cdot \Phi_T(1,1,w,h)$$

若想表示模板在图像各个位置的得分，则引入卷积记法：

$$F \otimes T(y,x) \triangleq K(F,T,y,x) \qquad 1 \le y \le m-h+1,\ 1 \le x \le n-w+1$$

### 2.2.3　滑动窗口检测法与多分辨率的图像表示

所谓“滑动窗口”检测法，是指以固定大小的窗口，依次放置于图像的各个位置，在每个位置上提取出窗口内的局部图像，交给一个分类器进行正/负例分类。若该窗口在该位置的局部图像被分类为正例，则输出此处有一个检测结果，检测结果以包围矩形（Bounding Box）的形式表示，矩形的左上角即为窗口在当前位置的左上角，矩形的大小即为窗口的大小。在使用检测模板的情况下，我们可以认为窗口的大小就是模板的大小，滑动的步长不是一像素而是一格，分类器就是模板及其对应的得分算法（配合某一阈值用于分类）。

通常我们无法预先知道待检测图像中待检测物体的大小，所以我们需要尝试所有可能的窗口大小（假设窗口长宽比固定）。等效地，我们也可以固定窗口的大小，而将它在图像的不同分辨率上进行滑动。如果使用HOG特征，我们需要在图像的不同分辨率上计算出HOG特征的Feature Map。这些不同分辨率下计算出的Feature Map按分辨率从高到低组成了一个特征金字塔（Pyramid），记为$H(i,j,l)$，其中$l$表示分辨率等级，也即特征金字塔的层号。$l=1$表示最高分辨率（位于金字塔底层），$l=L$表示最低分辨率，其中$L$为金字塔的总层数。若想表示第$l$层的Feature Map，则可以记为$H_l$。如果层与层之间的分辨率以一个固定的倍数变化，那么我们可以把这个变化系数记为$\lambda_H$，即若最底层的分辨率为$c$像素/厘米，则次底层的分辨率为$c/\lambda$像素/厘米。

### 2.2.4　部件模型

为了抓住物体的细节特征，并允许这些细节特征相对于物体整体的位置发生变化，[4]引入了可形变的部件模型。一个部件模型由一个大小固定的主模板$R = T^0_{w_0,h_0}(\cdot,\cdot)$和若干个部件构成。每个部件由以下参数组成：（1）

一个大小固定的检测模板$T_{w,h}(\cdot,\cdot)$；（2）它与主模板的分辨率层级差$\Delta l$；（3）它相对于主模板的期望位置（锚点）$a=(a_x,a_y)$；（4）它的形变惩罚系数$v=(v_1,v_2,v_3,v_4)$。设模型有C个部件，则每个部件可以表示为：

$$P_c=(T^{(c)}_{w_c,h_c},\Delta l^{(c)},a^{(c)},v^{(c)}) \qquad 1\le c\le C$$

整个模型则可以表示为：

$$M=(R,P_1,\cdots,P_C)$$

对于一个部件$P=(T_{w,h},\Delta l,a,v)$，若图像的特征金字塔为H，使用滑动窗口检测，主模板现在滑动到了第$l$层的$(i,j)$位置，则该部件需要在第$l-\Delta l$层（更高分辨率）选择一个位置放置。假设它选择了$l-\Delta l$层的$(i',j')$位置进行放置，那么该部件在这一摆放位置下对该模型在当前滑动窗口的得分贡献为：

$$S_P(H,i,j,l,i',j')\triangleq K(H_{l-\Delta l},T,i',j')+v\cdot d \tag{2-2}$$

其中$d$定义为：

$$d\triangleq(dx,dx^2,dy,dy^2) \tag{2-3}$$

$$dx\triangleq i'-i\lambda_H^{\Delta l}-a_x$$

$$dy\triangleq j'-j\lambda_H^{\Delta l}-a_y$$

代表了这个部件与主模板的相对位置偏离它本应该在的相对位置（锚点）的程度。其中$(i'-i\lambda_H^{\Delta l},j'-j\lambda_H^{\Delta l})$表示部件所在的位置与主模板的相对位置，主模板的位置坐标$(i,j)$被乘以了系数$\lambda_H^{\Delta l}$，是因为主模板的位置坐标$(i,j)$在$l$层，而部件的位置坐标$(i',j')$在$l-\Delta l$层，中间相差了$\Delta l$层，从而必须把主模板的位置坐

标$(i, j)$等效地转换到它对应的$l - \Delta l$层的坐标，也就是$(i\lambda_H^{\Delta l}, j\lambda_H^{\Delta l})$，其中$\lambda_H$是层与层之间的分辨率相差倍数。

可以看出，一个部件在某一放置位置上的得分分为两部分。第一部分是它的模板与那个区域的图像Feature Map的点积，即$K(H_{l-\Delta l}, T, i', j')$，我们可以把它称作“外观得分”；第二部分是因它偏离了它本应所在的锚点位置$(i\lambda_H^{\Delta l} + a_x, j\lambda_H^{\Delta l} + a_y)$而得到的惩罚分，我们可以把它称作“形变得分”。一般情况下，形变参数$v$的值会使得$v \cdot d$项始终是一个负数，而且部件位置偏离锚点越多，这一惩罚项越负。描述惩罚项大小与形变的关系，可以使用多种自由度。这里使用了4自由度模型，那么形变惩罚项的大小在形变空间$(dx, dy)$上的形状，是一个中心可以移动的、长短轴与坐标平行的椭圆。注意“中心可以移动”这个性质，意味着形变惩罚项并不一定在形变为0（也就是部件正好位于锚点所指定的位置）时最小，这样就等效地实现了对锚点的修正。在后文的学习算法中可以看到，锚点是通过手工设定得到的，而形变惩罚系数$v$是通过学习得到的。使得由学习得到的参数$v$具有对手工设定的锚点进行修正的能力，非常重要。

对于主模板的得分，因为滑动窗口的大小就是主模板的大小，所以主模板的得分就是主模板与窗口内的Feature Map的点积，即：

$$S_R(H, i, j, l) \triangleq K(H_l, T^0, i, j) \tag{2-4}$$

在各个部件都选定了自己的放置位置后，整个模型在当前滑动窗口位置、当前部件摆放布局下的得分，定义为主模板得分与各个部件得分的和，即：

$$S_M(H, i, j, l, i_1, j_1, i_2, j_2, \cdots, i_C, j_C) \triangleq S_R(H, i, j, l) + \sum_{c=1}^{C} S_{P_c}(H, i, j, l, i_c, j_c) \tag{2-5}$$

可以看出，一个模型在一幅图像上的得分，既与模型整体的窗口位置$(i, j, l)$有关，也与模型各部件的摆放位置$(i_1, j_1, \cdots, i_C, j_C)$有关。等效地，我们可以说，要描述一个被该模型检测到的物体实例，既要描述它在整个图像中的位置$(i, j, l)$，又要描述它的各个部件所处的位置。

## 2.3　学习算法

### 2.3.1　两种隐变量视角

所谓隐变量，是指在模型的学习算法中需要作为观测值给出，但在训练数据中又不存在的变量，即“缺失的观测数据”。例如，对于以上所描述的部件模型，若想学习得到描述模型的各个参数（如主模板和部件模板、部件形变系数等），我们的训练数据应该包括能够完整描述一个物体的全部观测变量，以及它对应的正负标签。而前文已经指出，要完整地描述一个物体实例，既要描述它在整个图像中的位置，又要描述它的各个部件所处的位置。物体检测领域通常所使用的训练数据，往往包含了物体实例在图像中的位置（以Bounding Box的形式在训练数据中标出），但通常不包含物体实例的各个部件的位置。事实上，什么是物体的部件，往往是由检测算法所使用的模型定义的，所以通用的训练数据根本不可能标注出物体实例的各个部件所处的位置。因此，显然各部件的摆放位置$(i_1, j_1, \cdots, i_C, j_C)$是[4]模型中的隐变量。而物体实例整体的位置$(i, j, l)$，由于已经在训练集中标出，所以不属于隐变量。

但也存在另外一种视角。如果我们暂时先忽略训练数据中标注出的Bounding Box信息，而认为训练数据只提供每一幅图像是否含有待检测物体这一正/负标签，那么物体实例在图像中的位置也成为了隐变量。采用这种视角的原因，是因为训练数据中的Bounding Box标注往往很不准确，噪音很大。如果过分信任它，认为它就是物体实例在图像中的准确位置、真实位置，则训练结果和最终性能往往会受到标注噪音的干扰。相反地，如果我们不硬性地依赖它，而只将它作为在学习算法的某一环节中对某些隐变量取值的一种提示，则既可以利用到它所提供的信息，又可以避免它含有的噪音。例如，以训练数据中的Bounding Box标注作为对物体位置这一隐变量的提示，我们在决定该隐变量的取值时，可以限制它的取值必须与训练数据中的Bounding Box标注有足够的重叠面积。这样可以大大缩小该隐变量的可能取值范围，而又不将它定死在一个唯一的取值上。近年来的实践证明了这种将Bounding Box也视为隐变量的方法可以有效地降低Bounding Box标注噪音的影响以及提高准确率。

在后续的模型和算法描述中，[4]采用后一种视角，即认为物体整体在图像中的位置$(i, j, l)$和物体各个部件的位置$(i_1, j_1, \cdots, i_C, j_C)$都是隐变量，而训练数据

提供的信息只包括图像本身以及在该图像内是否存在待检测物体这个二值标签。

### 2.3.2　模型的线性表示

设一个部件模型由主模板$R = T^0$和C个部件组成，每个部件的参数为$P_c = (T^{(c)}_{w_c,h_c}, \Delta l^{(c)}, a^{(c)}, v^{(c)})$，$1 \le c \le C$。设图像的HOG特征金字塔为$H$。设模型在图像上的当前滑动位置为$(i, j, l)$，各部件在各自所应该在的层（$l - \Delta l^{(c)}$层）上被摆放的位置为$(i_1, j_1), \cdots, (i_C, j_C)$。定义$z \triangleq (i, j, l, i_1, j_1, \cdots, i_C, j_C)$。由式2-2到2-5定义的模型得分计算方法，可得如下关系：

$$S_M(H, z) = \Phi(H, z) \cdot w(M) \tag{2-6}$$

其中，$\Phi(H, z)$定义为：

$$\Phi(H, z) \triangleq \Phi_{H_l}(i, j, h_0, w_0) \| \Phi^{(1)}(H, z) \| \cdots \| \Phi^{(C)}(H, z)$$

$$\Phi^{(c)} \triangleq \Phi_{H_{l-\Delta l^{(c)}}}(i_c, j_c, h_c, w_c) \| d^{(c)} \qquad 1 \le c \le C$$

回忆$\Phi_F(i, j, w, h)$表示对Feature Map $F$以$(i, j)$为左上角、面积为$hw$的局部的向量化，$\|$为向量的拼接，$d^{(c)}$的定义同式2-3。

$w(M)$的定义为：

$$w(M) \triangleq \Phi_{T^0}(1, 1, h_0, w_0) \| w^{(1)}(P_1) \| \cdots \| w^{(C)}(P_C)$$

$$w(P_c) \triangleq \Phi_{T^{(c)}}(1, 1, h_c, w_c) \| v^{(c)}$$

将模型的得分定义重写为如式2-6的线性形式后，我们可以看出，模型的得分实际上为两个向量乘积：一个向量为模型的参数所定义的权重向量$w$；

一个向量为由图像HOG特征金字塔$H$和隐变量$z$所决定的特征向量$\Phi(H,z)$。（注意$\Phi(H,z)$同样也与模型的参数有关，这些参数是：模型中各模板的面积$(h_0,w_0),(h_1,w_1),\cdots,(h_C,w_C)$，各部件的锚点$\{(a_x^{(c)},a_y^{(c)})\}$，各部件与主模板的层级差$\{\Delta l^{(c)}\}$，以及部件总数$C$。后文将会看到，这些参数都通过手工设定，而不是由学习产生。）

### 2.3.3　最优化问题

当有了模型得分的线性表示后，[4]从这个线性表示出发来设计对权重向量$w$的学习算法。设计学习算法前，首先要确定，假如已经得到了一个训练好的部件模型，应该怎样将它用于物体检测。给定一幅图像$x$，检测任务是判断该图像内是否存在一个待检测的物体。这个判断可以通过比较模型在该图像上的得分与阈值0来实现（若要使用其它阈值，可以为特征向量$\Phi(H,z)$添加一个常数维，最后还是可以写成得分与0做比较的形式）。但在计算得分前，需要先确定隐变量$z$的值。在这里，[4]穷举$z$的所有可能取值，最终将$z$的值取为那个使模型在整个图像上得分最大的。直观上，这意味着它搜索物体在整个图像上所有可能的位置、大小，以及各个部件所有可能的摆放位置，来找到那个最好的摆放方式，使得在这种摆放方式下图像最有可能含有一个待检测物体。这时，只要这种最好的摆放方式的得分能够超过阈值，就认为图像中存在一个待检测物体；如果连这个最好的摆放方式的得分都无法超过阈值，就判定该图像中不存在待检测物体。

形式化地，设待检测图像为$x$，其对应的HOG特征金字塔为$H(x)$。简记$\Phi(H(x),z)$为$\Phi(x,z)$。分别用1和-1来表示图像中存在或不存在待检测物体的判定结果。则在得到一个训练好的模型尤其是它的权重向量$w$后，判定一幅图像$x$是否含有待检测物体的判定准则为：

$$y^*(w)=\operatorname{sgn}(\max_z w^\top\Phi(x,z))$$

或等价地：

$$y^*(w) = \underset{y\in\{1,-1\}}{\operatorname{argmax}} \left(y \max_z w^\top \Phi(x,z)\right) \tag{2-7}$$

当有了判定准则后，使用3.2节介绍的Hinge Loss技术可导出其最大间隔学习算法。针对[4]的部件模型的分类准则式2-7，我们可以得到它诱导出的Hinge Loss：

$$\mathcal{R}(w,x,y_d) = \max\{0, 1 - y_d \max_z w^\top \Phi(x,z)\}$$

有了Hinge Loss后，可写出分类准则对应的优化问题。设训练数据包含$D$个样本，每个样本由特征向量$x_d$和实际类别$y_d$构成，即训练数据表示为$\mathcal{D} = \{(x_d, y_d)\}_{d=1}^{D}$。那么最大间隔学习算法定义为如下优化问题：

$$\min_w \frac{1}{2}\|w\|^2 + C\sum_{d=1}^{D} \max\{0, 1 - y_d \max_z w^\top \Phi(x_d,z)\} \tag{2-8}$$

其中常数$C$控制正则项与分类误差的权衡。

下面来考察此最优化问题的凸性。对于$w$，$w^\top\Phi(x_d,z)$是线性的，所以$\max_z w^\top\Phi(x_d,z)$是凸的。但当$y_d = 1$时，$1 - y_d \max_z w^\top\Phi(x_d,z)$是凹的，因此$\max\{0, 1 - y_d \max_z w^\top\Phi(x_d,z)\}$非凸也非凹。因此整个优化问题不是凸的。但从分析过程可以看到，凸性丧失的原因在于$-y_d$这一因子，因此与$y_d$的取值有关。如果我们把训练数据划分为正例集$P$和负例集$N$，即：

$$\begin{aligned} P &\triangleq \{d | 1 \le d \le D, y_d = 1\} \\ N &\triangleq \{d | 1 \le d \le D, y_d = -1\} \end{aligned}$$

那么优化问题的目标函数可以重写为：

$$L(w) \triangleq \frac{1}{2}\|w\|^2 + C\sum_{d=1}^{D} \max\{0, 1 - y_d \max_z w^\top \Phi(x_d,z)\}$$

$$= \underbrace{\frac{1}{2}\|w\|^2 + \sum_{d\in N} \max\{0, 1 + \max_z w^\top \Phi(x_d, z)\}}_{g(x),\ \text{convex}} + \underbrace{\sum_{d\in P} \max\{0, 1 - \max_z w^\top \Phi(x_d, z)\}}_{h(x),\ \text{non-convex}} \tag{2-9}$$

分为两部分重写后，$g(x)$部分为一个凸函数，$h(x)$不是凸函数。

但进一步观察可以发现，当$h(x)$中$z$的可能取值只有一个时，$1 - \max_z w^\top \Phi(x_d, z)$项退化为$w$的线性函数，$h(x)$也就成为了凸函数。

### 2.3.4　LSVM

LSVM（Latent SVM）算法使用了上文所指出性质。它将正例集中的$z$的可能取值集合限定为只有一个元素，即事先为正例集的$z$确定取值，然后求解凸优化问题2-8。求解得到模型参数后，它再使用准则$z^* = \text{argmax}_z\ w^\top \Phi(x_d, z)$为正例集重新确定$z$的取值，以此循环。形式化地，LSVM算法为：

**while** 没有收敛或到达指定次数 **do**
    使用准则$z^* = \text{argmax}_z\ w^\top \Phi(x_d, z)$为正例集确定$z$的取值。
    求解凸优化问题2-8。
**end**

**Algorithm 1:** LSVM学习算法

为证明算法收敛性，定义辅助目标函数：

$$L(w, \{z_d\}_{d\in P}) \triangleq \frac{1}{2}\|w\|^2 + \sum_{d\in N} \max\{0, 1 + \max_z w^\top \Phi(x_d, z)\} + \sum_{d\in P} \max\{0, 1 - w^\top \Phi(x_d, z_d)\}$$

可以看出LSVM算法的每一步都使得目标函数$L(w, \{z_d\}_{d\in P})$下降。而$L(w, \{z_d\}_{d\in P})$具有下界：

$$L(w, \{z_d\}_{d\in P}) \geq L(w) \geq 0$$

因此LSVM算法的迭代最后会收敛。但由于原问题非凸，收敛结果为局部最优，受初始值影响。

## 2.4　快速检测算法

在检测一幅图像时，如果对于每一个滑动窗口，都重新去寻找相对于这个窗口的各个部件的最佳摆放位置（之所以说相对这个窗口是因为部件的摆放得分包含了部件与锚点的相对位置），会造成很多重复计算。例如，部件模板在图像上各个位置的匹配得分（内积），是与滑动窗口位置无关的，它们可以被各个滑动窗口的搜索过程共用，从而只需要在进行滑动窗口前预先计算好即可。这相当于使用动态规划技术避免了子问题中的重复计算。对于部件c，我们记它的模板在图像HOG特征金字塔第$l$层各个位置进行匹配得到的得分矩阵为$R_{c,l}(i,j)$，即：

$$R_{c,l} \triangleq H_l \otimes T^{(c)}$$

我们称它为“响应矩阵”。

有了响应矩阵后，下一步就是要对每一个滑动窗口位置寻找最好的部件摆放位置。[4]使用两个原则来优化这一寻找过程：第一，由于每个部件的摆放位置选取与其它部件无关，所以[4]对每个部件单独寻找最优摆放位置；第二，尽量把滑动窗口和部件摆放两个穷举过程融合在一起，以期能够找到优化机会降低运算复杂度层级。为此，针对每一个部件c，在每一个金字塔层级$l$上，我们定义它的最优得分矩阵$D_{c,l}(i,j)$。它的含义是，当滑动窗口滑动到某一个位置，使得此时部件c的锚点位于图像（的Feature Map）的$(i,j)$位置时，该部件最优摆放位置的得分可以直接由$D_{c,l}(i,j)$获得。因此，如果我们已经计算得到了矩阵$D_{c,l}(i,j)$，则在做滑动窗口时，在每一个滑动窗口位置，我们通过$C$（部件个数）次查表即可获得模型在此滑动窗口处的得分。要计算$D_{c,l}(i,j)$，我们需要搜索各种偏离锚点位置$(i,j)$的偏移量$(d_i,d_j)$，考虑进它带来的形变惩罚分数，然后得到最好者。对于每一个部件c和每一个金字塔层级，$D_{c,l}(i,j)$的计算式可以写为：

$$D_{c,l}(i,j) = \max_{d_i,d_j} R_{c,l}(i+d_i, j+d_j) + v_1 d_i + v_2 d_i^2 + v_3 d_j + v_4 d_j^2$$

其中$v = (v_1, v_2, v_3, v_4)$为形变惩罚系数，$R_{c,l}$为部件c在l层的响应矩阵。在假设$v_2 \leq 0$和$v_4 \leq 0$的条件下，该矩阵可以使用分治法在$O(|i||j|\log(|d_i||d_j|)$的时间复杂度下计算出来，而不需要$O(|i||j||d_i||d_j|)$的时间复杂度（$|i|$指i的取值个数）。

## 2.5　本章小结

本章中，我们在数学上重描述（reformulate）了[4]提出的基于部件的可形变物体检测模型，及其学习算法和快速检测算法。后续章节将以本章的模型和数学表述为基础，提出新的学习算法和快速检测算法。

# 第 3 章 MedLSVM

## 3.1 本章概述

本章提出MedLSVM框架。MedLSVM是求解带有隐变量的最大间隔分类问题的概率模型框架。它在MED（Maximum Entropy Discriminative）的基础上，为分类器引入隐变量，并为隐变量引入概率分布，以期得到数学性质更好的最优化问题，并且使分类性能得到提高。

上文所描述的LSVM，也将隐变量引入了最大间隔分类器学习。它对隐变量的处理方式，是在分类准则中将隐变量用max操作消去，从而使得隐变量不作为自由变量出现在最优化问题中。这样带来的后果一是使得其最优化问题非凸，从而必须设计如上文所述的LSVM迭代算法；二是max是对隐变量的点估计，在有些情况下点估计不足以描述隐变量的可能分布情况，这时对隐变量的分布进行完整估计可能更好。

## 3.2 最大间隔准则与Hinge Loss技术

在标准线性SVM（后文详述）中，最大间隔学习算法的目标是在训练样本的约束下使分类面（平面）距离最近的训练样本的距离（称为“间隔”）尽量大。这是适用于标准SVM的几何解释。更一般地，对于一个普通的多分类问题，在给定了判断准则的情况下，最大间隔方法使用一种称为Hinge Loss的误差估计来作为当前训练所得参数在训练数据上的经验误差的上界，并试图通过最小化这个上界来最小化训练所得参数在训练数据上的经验误差。形式化地，假设我们有一个K分类问题，并如同对多分类问题的通常处理方式，我们使用判别函数来定义分类规则。一个判别函数是一个定义在样本空间上的实值函数。设待分类样本为$x$，K个类别的每个类别分别对应了一个判别函数，记为$f_k(x)$，$1 \le k \le K$。使用判别函数进行多分类时的分类准则为：

$$y^*(f) = \underset{1 \le y \le K}{\operatorname{argmax}} \ f_y(x) \tag{3-1}$$

设当真实类别为$y_d$时，将类别判定为$y$所产生的误差为$\Delta(y_d, y)$。设待分类样本$x$对应的真实类别为$y_d$。使用当前的判别函数集$f = \{f_y\}$，依分类准则3-1得到的分类结果记为$y^*(f)$。那么当前的判别函数集（即待训练的模型参数）在训练样本$(x, y_d)$上经验误差即为：

$$\Delta(y_d, y^*(f))$$

根据式3-1对$y^*(f)$的定义，以下不等式关系可以很容易得到证明：

$$\begin{aligned} \Delta(y_d, y^*(f)) &\leq \Delta(y_d, y^*(f)) + f_{y^*(f)}(x) - f_{y_d}(x) \\ &\leq \max_{1 \leq y \leq K} (\Delta(y_d, y) + f_y(x)) - f_{y_d}(x) \end{aligned} \tag{3-2}$$

我们称不等号最右边的式子为由分类准则3-1诱导出的、当前模型参数$f$在训练数据$(x, y_d)$上的Hinge Loss，将其记为$\mathcal{R}(f, x, y_d)$，即：

$$\mathcal{R}(f, x, y_d) \triangleq \max_{1 \leq y \leq K} (\Delta(y_d, y) + f_y(x)) - f_{y_d}(x)$$

由式3-2可知，Hinge Loss是当前模型参数在训练数据上的分类误差的上界。

在二分类线性判别函数情况下，Hinge Loss还有另一个等价的形式。二分类问题的类别一般表示为1和-1。不失一般性地，二分类问题的线性判别函数总可以写成：

$$\begin{aligned} y_1(x) &= \frac{1}{2} w^\top x \\ y_{-1}(x) &= -\frac{1}{2} w^\top x \end{aligned}$$

二分类问题的分类误差$\Delta(y_d, y)$一般定义为0-1误差，即：

$$\Delta(y_d, y) = \begin{cases} 0 & y_d = y \\ 1 & y_d \neq y \end{cases}$$

在这些设定下，Hinge Loss可以改写为：

$$\begin{aligned}\mathcal{R}(w,x,y_d) &= \max_{y\in\{1,-1\}}(\Delta(y_d,y)+f_y(x))-f_{y_d}(x)\\ &= \max\{\Delta(y_d,y_d)+f_{y_d}(x)-f_{y_d}(x),\Delta(y_d,-y_d)+f_{-y_d}(x)-f_{y_d}(x)\}\\ &= \max\{0,1-2f_{y_d}(x)\}\\ &= \max\{0,1-y_d w^\top x\}\end{aligned} \tag{3-3}$$

有了Hinge Loss的定义后，最大间隔方法试图通过最小化Hinge Loss这一上界来最小化模型在训练数据上的分类误差。因此，最大间隔学习算法是解一个优化问题，优化目标是Hinge Loss，以及一个为防止过拟合而加入的对模型参数的正则化项。

## 3.3 标准SVM形式化描述

要描述MedLSVM，我们首先从标准SVM的形式化入手。SVM的推导和表述可以从多种角度进行，本文在此以其中一种形式化表示来描述标准SVM，以期读者能够熟悉本文在后续章节使用的符号和形式。

一个线性二分类SVM，其模型参数就是一个用于分类的权重向量$\eta$。在已经训练得到$\eta$的情况下，当输入一个待分类向量$x$时，它的分类准则（用1和-1表示二分类的类别号）是：

$$y^*(\eta)=\operatorname*{argmax}_{y\in\{1,-1\}} y\eta^\top x \tag{3-4}$$

有了分类准则后，根据3.2描述的Hinge Loss方法，我们可以由这个分类准则诱导出最大间隔学习算法。这个算法是求解一个最优化问题。设训练数据包含$D$个样本，每个样本由特征向量$x_d$和实际类别$y_d$构成，训练数据表示为$\mathcal{D}=\{(x_d,y_d)\}_{d=1}^D$。则这个优化问题表述为：

$$\min_\eta \frac{1}{2}\|\eta\|^2+C\sum_{d=1}^{D}\max\{0,1-y_d\eta^\top x_d\} \tag{3-5}$$

其中的$\max\{0, 1 - y_d\eta^\top x_d\}$项就是由式3-4诱导出的、模型参数$\eta$在训练样本$(x_d, y_d)$上Hinge Loss。它是$\eta$对$(x_d, y_d)$的实际分类误差（使用0-1误差定义）的上界。使用松弛变量，这个线性二分类SVM还可以表述成另一种等价的形式。为各个训练样本$(x_d, y_d)$引入实数$\xi_d$，使得：

$$\xi_d \geq \max\{0, 1 - y_d\eta^\top x_d\} \quad \forall d$$

或等价地：

$$\begin{cases} y_d\eta^\top x_d \geq 1 - \xi_d \\ \xi_d \geq 0 \end{cases} \quad \forall d$$

则式3-5可以等价地改写为：

$$\begin{aligned} &\min_{\eta,\xi} \tfrac{1}{2}\|\eta\|^2 + C\sum_d \xi_d \\ \text{s.t.} \quad &\xi_d \geq \max\{0, 1 - y_d\eta^\top x_d\} \quad \forall d \end{aligned} \tag{3-6}$$

或等价地：

$$\begin{aligned} &\min_{\eta,\xi} \tfrac{1}{2}\|\eta\|^2 + C\sum_d \xi_d \\ \text{s.t.} \quad &y_d\eta^\top x_d \geq 1 - \xi_d \quad \forall d \\ &\xi_d \geq 0 \quad \forall d \end{aligned} \tag{3-7}$$

由最优化理论知，式3-5与式3-7具有相同的最优解$\eta$。变量$\xi$可以认为是对训练样本最大间隔约束的一种松弛，即允许某些样本在一定程度上进入间隔区，但这些破坏间隔约束的样本数量应该尽量少。

## 3.4 MED框架

标准SVM是对参数$\eta$的点估计。但如果$\eta$很明显地服从一种分布，或我们拥有$\eta$的先验分布的知识，那么我们希望能够对$\eta$的整个分布做出估计，而不只是

单点估计。为参数$\eta$引入概率分布，使得最大间隔学习与统计学习两条脉络结合到了一起，使最大间隔学习方法可以使用统计学习理论中的方法和工具，如贝叶斯估计、变分推理等。MED（Maximum Entropy Discriminative）模型就是这样一种为最大间隔学习引入概率背景和概率框架的方法。它为要在优化问题中求解的模型参数引入一个概率分布，把原来求解最优参数值的问题转换成了求解最优参数分布的问题。它使用KL距离（或称相对熵Relative Entropy）作为对待求解的分布的一种正则化约束。KL距离衡量待求解的分布与一个预先给定的先验分布的相差程度。我们可以利用这个先验分布来提供我们对参数的先验知识。使用不同的先验分布将使得优化问题以及求解出的参数概率分布具有非常不同的特性。

在MED模型中，我们仍使用分类权重向量$\eta$对样本做线性分类。所不同的是，我们不再拥有一个确定的参数$\eta$的值，而是拥有一个估计出的$\eta$的分布$q(\eta)$。当已经训练得到这个$q(\eta)$后，在用它做分类时，我们将所有的$\eta$的可能值都用来进行分类，并用$\eta$的概率分布$q(\eta)$对这些分类结果进行加权平均。由此，我们有了MED模型的分类准则：

$$y^*(q) = \underset{y\in\{1,-1\}}{\operatorname{argmax}}\ y\mathbb{E}_{q(\eta)}[\eta^\top x]$$

注意其中原先SVM中$\eta$的地位现在已经被$\mathbb{E}_{q(\eta)}[\cdot]$所取代。有了这个分类准则后，我们可以使用与3.2相同的技术来诱导出最优化学习问题。但在原先的最优化问题中为防止过拟合对待优化参数$\eta$的正则化约束由$\|\eta\|^2$项扮演，而现在我们的待优化参数是$q(\eta)$，我们必须找到一种方式来对这个以概率分布的形式出现的参数做正则化约束。KL距离就是这样一种方法。KL距离是定义在两个概率密度函数上的泛函，其定义为：

$$\mathrm{KL}(q(x)\|p(x)) \triangleq -\int q(x)\ln\frac{p(x)}{q(x)}\mathrm{d}x \quad q,p\in\mathcal{P}_{\mathrm{prop}}$$

其中$\mathcal{P}_{\mathrm{prop}}$指所有概率密度函数组成的集合。KL距离反应了两个概率分布的相差程度（但它不是“距离”，尤其它不是可交换的）。对于$q(\eta)$，如果我们已经有了一个$\eta$的先验分布$p_0(\eta)$，我们就可以使用KL 距离控制$q(\eta)$使得$q(\eta)$不

会距离$p_0(\eta)$过远，这样就防止了$q(\eta)$过分向训练数据靠拢从而产生过拟合。而且，我们还可以以$p_0(\eta)$为入口向$\eta$的分布提供一些先验知识以作为在估计它的分布时的一种提示。使用KL距离作为对$q(\eta)$的正则化约束后，我们就得到了MED模型的最优化问题。同样假设训练数据包含$D$个样本，每个样本由特征向量$x_d$和实际类别$y_d$构成，训练数据表示为$\mathcal{D} = \{(x_d, y_d)\}_{d=1}^{D}$。假设已知模型参数$\eta$的先验分布为$p_0(\eta)$，那么为求解MED模型的最优参数$q(\eta)$而需要求解的最优化问题为：

$$\begin{aligned} &\min_q \mathrm{KL}(q(\eta)\|p_0(\eta)) + C\sum_d \max\{0, 1 - y_d\mathbb{E}[\eta^\top x_d]\} \\ \text{s.t.} \quad &q \in \mathcal{P}_{\text{prop}} \end{aligned} \tag{3-8}$$

由于$\mathrm{KL}(q\|p_0)$和$\max\{0, 1 - y_d\mathbb{E}[\eta^\top x_d]\}$相对于$q(\eta)$都是凸的（泛函意义下），所以问题3-8是一个凸优化问题。

## 3.5 MedLSVM

参照MED的思路，在处理带有隐变量的最大间隔分类器学习问题时，我们也希望将对隐变量的单点估计（如max操作）替换为对隐变量概率分布的整体估计。但在这之前，我们需要再进一步观察已有的对隐变量的处理方式，以及其所导致的分类器的行为。

### 3.5.1 两种LSVM及其比较

回忆2.3.3所定义的带有隐变量的最大间隔分类器，我们在此将其称为LSVM-A。使用本章所用的符合$\eta$，LSVM-A的分类准则可以重写如下：

$$y^*(\eta) = \operatorname*{argmax}_{y\in\{1,-1\}} (y \max_z \eta^\top \Phi(x, z))$$

但对于同样的隐变量，实际上还存在另一种处理方式。[17]提出了另一种带有隐变量的最大间隔分类器，在此我们将它称为LSVM-S。它的分类准则如下：

$$y^*(\eta) = \underset{y\in\{1,-1\}}{\operatorname{argmax}} \max_{z} (y\eta^\top \Phi(x,z))$$

仔细观察两种分类准则，我们发现他们的唯一区别在于$y$因子与$\max_z$操作的位置。交换两者的位置会导致结果的变化吗？我们可以通过如下两个计算结果来说明：

$$\underset{y\in\{1,-1\}}{\operatorname{argmax}} \max_{z\in\{-2,1\}} zy = 1 \tag{3-9}$$

$$\underset{y\in\{1,-1\}}{\operatorname{argmax}} (y \max_{z\in\{-2,1\}} z) = -1 \tag{3-10}$$

可以看出，在$y$可以取负值的情况下，乘$y$和取$\max_z$两者的顺序不能交换。更仔细分析，可以看到它们在本质上的区别。如果两者的顺序为$y \max_z f(z)$，那么$\max_z$不会考虑$y$的取值（尤其是$y$的符号），而只顾让$f(z)$达到最大。如果$y$的取值范围为$\{1,-1\}$，那么最后乘$y$（并取$\operatorname{argmax}_y$）的操作只是相当于为这个最大的$f(z)$取sgn。而如果两者的顺序为$\max_z yf(z)$，那么$\max_z$的行为会受$y$的符号的影响：如果$y$为正，那么$\max_z$会使$f(z)$达到最大（或者最正）；如果$y$为负，那么$\max_z$会使$f(z)$达到最小（或者最负）。最终$\operatorname{argmax}_y$的结果，取决于使$f(z)$最大的努力和使$f(z)$最小的努力哪个达到的效果更好。

如果$z$表示对样例的某种观察方式，类型号1表示第一类，类型号-1表示第二类，那么当乘$y$和$\max_z$两者的顺序为$y \max_z$时，意味着分类器会尽量尝试去寻找有利于第一类的观察方式，如果找到的最有利的观察方式满足第一类的要求，那么就报告为第一类；而如果连这个最有利的观察方式都无法满足第一类的要求，则只好报告为第二类。这种行为方式无疑是偏向于第一类的。而如果两者的顺序是$\max_z y$，则分类器会分别去试图寻找最有利于第一类的观察方式和最有利于第二类的观察方式。最终结果会报告为哪一类，取决于这两类的最优观察方式哪一个得分更高。也可以说，这种分类器会突显出两类最具有各自特征的地方，然后使用这两个最显著部位来比拼谁更显著。

我们将第一种分类器称为“非对称LSVM（LSVM-Asymmetric）”，将第二种分类器称为“对称LSVM（LSVM-Symmetric）”。总结起来，它们的分类原则是：对于非对称LSVM，如果有任何途径能够将样例判为第一类，则输出为第

一类；对于对称LSVM，如果最有利于第一类的观察方式好于最有利于第二类的观察方式，则输出第一类。其它情况下两者均输出第二类。

它们各自适用于什么场合，可以从如下角度看。如果要使用严重偏向于第一类的非对称LSVM，意味着只要样例中出现了（足够的）第一类的特征，就应该被判为第一类，而只有当样例中不含有第一类的特征时，才应该被判为第二类。那么从本质上讲，第二类这个类别的定义应该是“不含有第一类特征的”、“非第一类的”。所以更符合习惯地，第一类和第二类应该被分别称作正例类和负例类。正例具有自己的特征，而负例不具有自己的特征，它们只需要“不是正例”。例如，在物体检测中，如果检测对象是人脸，则所有正例都应该具有人脸的特征，但各个负例可以来自各种背景，而不需要具有共有的特征。因此，非对称LSVM适合于进行正/负例分类。而物体检测中的分类就是正/负例分类。

对称LSVM则适合于两类都拥有各自的特征，并因自己的特征而被定义的情况。例如分类猫的图片和汽车的图片。对称LSVM的一个附带好处是可以找出两类各自的最具有区分性的地方，所以可以用来实现例如检测-分类联合学习。

从上文分析可看出，我们要处理的物体检测问题更适合于使用非对称LSVM。所以，下文的推导将从LSVM-A出发，而且如未经特别指出，后文所指的LSVM均指LSVM-A。

### 3.5.2 为隐变量引入概率分布

用max操作在分类准则中消去隐变量$z$，会带来两个问题：一是它使得分类准则诱导出的最优化问题非凸；二是max是对$z$的点估计，但有时$z$的各种可能取值的概率分布更能反映$z$的信息。借鉴MED模型，我们为隐变量$z$ 引入概率分布$q(z)$。在我们的物体检测问题中，$z$是离散变量。但为了一般化，并为了记法简洁统一，我们在后文的推导中会统一使用积分运算。实际计算时，可视情况将积分改为求和。引入$q(z)$后，分类结果不再根据 $\max_z$，而是$z$在$q(z)$分布下各种可能取值造成的分类结果的平均值。形式化地，引入概率

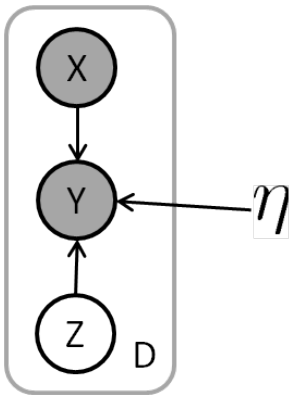


图 3.1　MedLSVM模型3-12的概率图表示

分布$q(z)$后，LSVM的分类准则变为：

$$y^*(\eta, q) = \underset{y\in\{1,-1\}}{\operatorname{argmax}} \left(y\mathbb{E}_{q(z)}[\eta^\top \Phi(x, z)]\right) \tag{3-11}$$

注意，为使问题简单，我们暂时还没有为模型参数$\eta$引入概率分布。有了分类准则后，由3.2介绍的Hinge-Loss技术，我们可以由分类准则诱导出其对应的最优化问题。设训练数据包含$D$个样本，每个样本由特征向量$x_d$和实际类别$y_d$构成，训练数据表示为$\mathcal{D} = \{(x_d, y_d)\}_{d=1}^D$。由于每个训练样本$(x_d, y_d)$都有自己对应的隐变量$z_d$，我们记$z = \{z_d\}_{d=1}^D$，则$q(z)$表示所有$z_d$的联合分布。分类准则3-11诱导出的最优化问题为：

$$\begin{aligned} &\min_{\eta,q} \tfrac{1}{2}\|\eta\|^2 + C_1 \operatorname{KL}(q(z)\|p_0(z)) + C_2 \textstyle\sum_d \max\{0, 1 - y_d \mathbb{E}_q[\eta^\top \Phi(x_d, z_d)]\} \\ \text{s.t.} \quad &q \in \mathcal{P}_{\text{prop}} \end{aligned} \tag{3-12}$$

为$z$引入了概率分布后，我们实际上拥有了一个概率模型。将它用有向概率图表示，为图3.1

### 3.5.3　两点不足

引入概率分布$q(z)$后，当前的分类器模型存在两点不足。一是难以对新样本进行分类。在分类器已经训练好要使用时，当输入一个新的样本$x_{\text{new}}$，我们需要推理出与它配套的隐变量$z_{\text{new}}$的分布。形式化地，要对新样本$x_{\text{new}}$ 进行分

类，表示为概率的形式，即要求解$p(y_{\text{new}}|x_{\text{new}},\mathcal{D})$，其中$\mathcal{D}$为训练数据。（对于没有引入概率分布的条件关系，可以看做单值概率为1的分布。）使用贝叶斯公式，我们有：

$$\begin{aligned} p(y_{\text{new}}|x_{\text{new}},\mathcal{D}) &= \int p(y_{\text{new}},z_{\text{new}}|x_{\text{new}},\mathcal{D})\mathrm{d}z_{\text{new}} \\ &= \int p(z_{\text{new}}|x_{\text{new}},\mathcal{D})p(y_{\text{new}}|z_{\text{new}},x_{\text{new}},\mathcal{D})\mathrm{d}z_{\text{new}} \end{aligned} \tag{3-13}$$

为此，我们需要求解$p(z_{\text{new}}|x_{\text{new}},\mathcal{D})$。直观地想，$z$表示了部件的摆放位置等信息，它应该是与图像有着直接关系的（看到了图像才能决定哪个部件应该摆放在哪里）。但由图3.1可以看出，$x_{\text{new}}$和$z_{\text{new}}$之间并没有直接的条件概率关系。这既违反了直觉，也使得推理$p(z_{\text{new}}|x_{\text{new}},\mathcal{D})$变得困难。因此我们需要在$x$与$z$之间建立概率关系。

第二点，为了使3-12最小化，$q(z)$会尽量迎合训练数据以使得Hinge Loss最小。迎合的方式将会是，对于正例（$y_d = 1$），$q(z_d)$会把概率集中在使得$\eta^\top\Phi(x_d,z_d)$尽量大的地方，因为这样$\max\{0,1-y_d\mathbb{E}_q[\eta^\top\Phi(x_d,z_d)]\}$会较小；同样的，对于负例（$y_d = -1$），$q(z_d)$会把概率集中在使得$\eta^\top\Phi(x_d,z_d)$尽量小的地方，因为这样$\max\{0,1-y_d\mathbb{E}_q[\eta^\top\Phi(x_d,z_d)]\}$也会较小。可以看出，这种行为实际上非常类似于对称LSVM的行为，即将隐变量的值设成最能突出各个类别各自特征的地方。但前文已说过，对于物体检测问题，我们希望它的分类器的行为更接近于一个非对称LSVM，即无论是正例还是负例，都将隐变量的取值集中在最像正例特征的地方。负例中接近正例的样例或样例观察方式，才是正/负例分类中常说的“困难负例”。而只有使用困难负例进行训练，才能期望正/负例分类器有足够的区分能力。对称LSVM的行为会使降低训练误差变得简单，但也使得训练得不到什么实际进展。

### 3.5.4　加入产生模型

由于以上两点不足，我们为观测数据$x$和隐变量$z$之间引入概率关系，以期同时解决这两点不足。要引入$x$和$z$的概率关系，既可以从定义$p(z|x)$出发，也可以从定义$p(x|z)$出发。由于$x$是观测数据，$z$是其背后的隐含信息，依照习惯，我

们一般定义$p(x|z)$，表示我们观测到的数据是怎样由其背后的决定因素决定的。有了$p(x|z)$后，前文所述的我们需要的$p(z|x,\eta)$可由贝叶斯公式得到：

$$\begin{aligned} p(z|x,\eta) &\propto p(z|\eta)p(x|z,\eta) \\ &= p(z)\prod_d p(x_d|z_d) \end{aligned} \tag{3-14}$$

所以我们的任务是要定义似然函数$p(x_d|z_d)$（或$p(x_d|z_d,\eta)$）。

### 3.5.5 选择似然函数

观察LSVM的分类准则2-7，它与我们的分类准则3-11的差别在于$\max_z f(z)$与$\mathbb{E}_{q(z)}[f(z)]$。要让$\mathbb{E}_{q(z)}[f(z)]$的行为与$\max_z f(z)$相似，我们希望$q(z)$能够让概率集中在那些使$f(z)$较大的$z$的取值上，对于我们的问题，也就是使$\eta^\top\Phi(x,z)$较大的$z$的取值上。由于$p(z|x,\eta) \propto p(x|z,\eta)p(z)$，我们定义似然函数$p(x|z,\eta)$来达到这个目的。要让使$\eta^\top\Phi(x,z)$越大的$z$对应的$p(x|z,\eta)$越大，看上去我们可以定义$p(x|z,\eta)$为：

$$p(x|z,\eta) \propto \eta^\top\Phi(x,z)$$

但很可惜这不是合法的概率分布或似然函数，因为$\eta^\top\Phi(x,z)$的值可能为负。一个简单的处理方法，是将这个可能为负的值放在指数函数的肩上，成为：

$$p(x|z,\eta) \propto \exp(\eta^\top\Phi(x,z)) \tag{3-15}$$

从直观上看，使值$\eta^\top\Phi(x,z)$较大的$\Phi(x,z)$是那些与$\eta$本身较为相像的（如果$\Phi(x,z)$的模相差不大），所以从相似或距离的角度，可供候选的似然函数形式也可以是：

$$p(x|z,\eta) \propto \exp(-\frac{1}{2}(\eta - \Phi(x,z))^2) \tag{3-16}$$

或

$$p(x|z,\eta) \propto \exp(-|\eta - \Phi(x,z)|)$$

实际上，高斯形式的似然函数定义3-16与指数形式的似然函数3-15有着紧密的联系，这可以从如下推导看出：

$$\exp(-\frac{1}{2}(\eta - \Phi(x,z))^2) = A\exp(\eta^\top \Phi(x,z) - \frac{1}{2}\|\Phi(x,z)\|^2) \tag{3-17}$$

高斯形式的似然函数与指数形式的似然函数只相差一个$\exp(-\|\Phi(x,z)\|^2)$系数。但这个系数的影响大小则只能通过经验实验观察。后文4.2节的实验结果显示，式3-15定义的指数形式的似然函数行为和性质最好。所以我们选择式3-15作为我们的似然函数定义。

确定似然函数的形式后，还必须给出详细的定义。所谓详细的定义，是指明在给定$z$和$\eta$的条件下，$x$是以何种方式产生的。结合具体问题，由于$z$在我们的物体检测问题中表示部件的摆放位置，即表示选择图像的 HOG特征的哪一块来形成样本向量，所以式3-15的含义是：在给定$z$的条件下，在$z$框出的区域，图像的HOG特征以概率$\exp(\eta^\top \Phi(x,z))$（再除以某一归一化因子）随机生成，而在$z$框选范围以外的地方，图像的HOG特征以[0,1]的均匀分布概率随机生成（归一化因子为1）。另外，如果允许$x$无限趋向正无穷或负无穷，式3-15不是一个合法的分布，因为积分结果可能不是有限值。所以我们必须限制观测数据（即图像的HOG特征）的取值范围。幸运的是，HOG特征的提取过程有归一化的步骤，所以HOG特征的各个维天然就在[0,1]之间。综上，我们的准确的给定$z$和$\eta$条件下的$x$的条件概率分布为：

$$p(x|z,\eta) = \begin{cases} \frac{1}{A(\eta)}\exp(\eta^\top \Phi(x,z)) & 0 < x_i < 1,\ \forall i \\ 0 & \text{otherwise} \end{cases}$$

到这里，我们将要面临计算归一化系数的问题。因为$z$只起到选择特征区域的作用，所以它的取值$z$不影响归一化因子。而一般情况下，归一化因子的值

会和$\eta$有关，我们把它记为$A(\eta)$。在受限指数分布的情况下， $A(\eta)$能够由积分得到：

$$\begin{aligned} A(\eta) &= \int \exp(\eta^\top \Phi(x,z)) \mathrm{d}\Phi(x,z) \\ &= \int_{x_1,\cdots,x_M} \exp(\eta_1 x_1 + \cdots + \eta_M x_M) \mathrm{d}x_1 \cdots \mathrm{d}x_M \\ &= \int_0^1 \exp(\eta_1 x_1) \mathrm{d}x_1 \cdots \int_0^1 \exp(\eta_M x_M) \mathrm{d}x_M \\ &= \prod_{k=1}^{M} \frac{\eta_k}{\exp(\eta_k) - 1} \end{aligned} \tag{3-18}$$

其中$\eta_k$表示$\eta$的各维，$x_k$表示$\Phi(x,z)$的各维，M为$\eta$的维数。有了归一化因子$A(\eta)$后，$x$的条件分布可以具体化为：

$$p(x|z,\eta) = \begin{cases} \prod_k \frac{\eta_k}{\exp(\eta_k)-1} \exp(\eta^\top \Phi(x,z)) & 0 < x_i < 1,\ \forall i \\ 0 & \text{otherwise} \end{cases} \tag{3-19}$$

### 3.5.6 简化归一化因子

在后文的推导中我们将会看到，式3-19的归一化因子同$\eta$的关系过于复杂，使得对$\eta$的学习变得困难。有一种对3-15的简单的修改方法可以使$A(\eta)$变得简单。我们将式 3-15修改为：

$$p(x|z,\eta) \propto \exp(\frac{1}{\|\eta\|} \eta^\top \Phi(x,z)) \tag{3-20}$$

并将$\Phi(x,z)$限制在一个球内（即将$\Phi(x,z)$的模长限制）。由于$p(x|z,\eta)$是关于$x$的分布，自变量是$x$，所以式3-20并不改变$x$的函数形式，而只是改变了指数函数的上升速度，而且是跟随 $\eta$自动改变。可以证明，在式3-20的函数形式下，归一化因子将与$\eta$无关。也就是：

$$p(x|z,\eta) = \frac{1}{A} \exp(\frac{1}{\|\eta\|} \eta^\top \Phi(x,z))$$

直观的看，这是因为$\exp(\frac{1}{\|\eta\|}\eta^\top\Phi(x,z))$的积分结果只与$\eta$的方向有关，而积分是在一个球内进行（即$\Phi(x,z)$的取值范围），所以无论$\eta$是什么方向，$\exp(\frac{1}{\|\eta\|}\eta^\top\Phi(x,z))$的积分结果都相同。也可以说，$\exp(\frac{1}{\|\eta\|}\eta^\top\Phi(x,z))$在$\Phi(x,z)$的取值被限制在球内的情况下，刻画的是一个方向空间（或球面）的指数分布。其分布具有旋转不变性，无论主方向是什么方向，分布的形状都相同，归一化因子也相同。

### 3.5.7 完整MED框架

到目前为止，我们的分类器模型还有一个问题需要解决。在引入了$x$和$z$、$\eta$之间的产生模型后，式3-12中原先起到对$q(z)$约束作用的$z$的先验$p_0(z)$应该反映出因观测到$x$而带来的新的信息，因此先验分布$p_0(z)$应该被取代为观测到$x$后的$z$的后验分布$p(z|x,\eta)$（因$\eta$没有概率分布，所以作为参数写在条件列表中；也可认为$\eta$有一个单值概率为1的分布，于是$p(z|x)$等同于$p(z|x,\eta)$）。这样，式3-12可以改写为：

$$\begin{aligned} &\min_{\eta,q} \tfrac{1}{2}\|\eta\|^2 + C_1\,\mathrm{KL}(q(z)\|p(z|x,\eta)) + C_2\sum_d \max\{0, 1 - y_d\mathbb{E}_q[\eta^\top\Phi(x_d,z_d)]\} \\ \text{s.t.}\quad & q \in \mathcal{P}_{\text{prop}} \end{aligned} \tag{3-21}$$

而其中的$p(z|x,\eta)$可以用贝叶斯公式展开为：

$$\begin{aligned} p(z|x,\eta) &= \frac{p(z|\eta)p(x|z,\eta)}{p(x|\eta)} \\ &= \frac{p(z)\prod_d p(x_d|z_d)}{p(x|\eta)} \end{aligned} \tag{3-22}$$

其中的$p(x|\eta)$可以进一步展开为：

$$p(x|\eta) = \int p(x|z,\eta)p(z)\mathrm{d}z \tag{3-23}$$

从式3-23和$p(x|z,\eta)$的定义式3-19或式3-20可以看出，$p(x|\eta)$对$\eta$的依赖关系将会非常复杂。具体地说，它会是多个指数运算的和的形式。在之后的学习和

最优化问题中，为了最优化$\eta$，我们会遇到对$\ln p(x|\eta)$求导并求解导数零点等问题，$p(x|\eta)$对$\eta$的这种复杂的依赖关系使得对$\eta$的最优化变得无法进行。

为了解决这个问题，即要避免直接对$p(z|x,\eta)$的计算，我们最终也为$\eta$引入了概率分布。这样，分类准则中使用到的参数$z$和$\eta$都以概率分布的形式出现并以概率分布的形式被优化，我们就得到了一个完整的 MED框架。为$\eta$引入概率分布后，我们的分类准则公式变成了如下形式：

$$y^*(q) = \underset{y\in\{1,-1\}}{\operatorname{argmax}} \left(y\mathbb{E}_{q(\eta,z)}[\eta^\top \Phi(x,z)]\right) \tag{3-24}$$

其中原先$\eta$的作用被$\mathbb{E}_{q(\eta)}[\cdot]$所取代。注意我们在这里并没有做$\eta$和$z$的独立性假设，所以我们将$\eta$和$z$的概率分布合写为联合概率分布的形式。有了分类准则3-24 后，使用3.2介绍的Hinge Loss约束，并以观测到$x$后的$\eta$和$z$的联合后验概率分布来约束待求解的分布$q(\eta,z)$（使用KL距离），我们有了分类准则3-24诱导出的最优化问题：

$$\begin{aligned} &\min_q \mathrm{KL}(q(\eta,z)\|p(\eta,z|x)) + C\sum_d \max\{0, 1 - y_d\mathbb{E}_{q(\eta,z)}[\eta^\top\Phi(x_d,z_d)]\} \\ \text{s.t.}\quad &q \in \mathcal{P}_{\text{prop}} \end{aligned} \tag{3-25}$$

式3-25即是我们最终得到的分类器模型及其学习算法。将这个模型表达的概率模型用有向概率图表示出来，得到图3.2。我们将其命名为MedLSVM，表示MED框架下带有隐变量的最大间隔学习。它为含有隐变量的最大间隔学习模型，尤其是非对称、适用于正/负例分类的含隐变量最大间隔学习模型，提供了一个统一的概率背景和学习框架。

### 3.5.8 求解MedLSVM

最优化问题3-25中的$\mathrm{KL}(q(\eta,z)\|p(\eta,z|x))$项对于要求解的$q(\eta,z)$是凸的（泛函意义下），项$1 - y_d\mathbb{E}_{q(\eta,z)}[\eta^\top\Phi(x_d,z_d)]$对$q(\eta,z)$是线性的，因此$\max\{0, 1 - y_d\mathbb{E}_{q(\eta,z)}[\eta^\top\Phi(x_d,z_d)]\}$对于$q(\eta,z)$是凸的，因此整个问题3-25对于$q(\eta,z)$是凸的。该问题可以用采样方法求解，但采样方法的采样空间与$\eta$和$z$的维数成指数

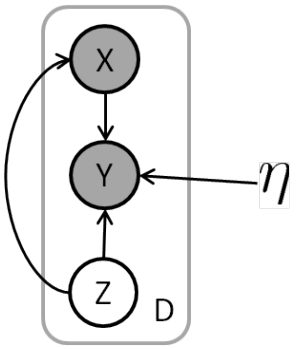


图 3.2 MedLSVM模型3-25的概率图表示

关系，所以采样方法的速度和精度可能都不理想。在这里本文用变分近似（Variational Approximation）方法求解。

使用变分近似方法，我们需要对待求解的$q(\eta, z)$做出一些假设。首先，我们需要对它做Mean Field假设。所谓Mean Field假设，是假设分布$q(\eta, z)$可以分解为几个因子的乘积的形式，即可分解性假设。在这里，我们做出假设：

$$q(\eta, z) = q(\eta)q(z) = q(\eta)\prod_d q(z_d)$$

注意，这个假设会造成求解最优化问题的结果不再是最优，而是近似最优。虽然从概率图3.2来看，$z$和$\eta$以及各个$z_d$是互相独立的，联合概率$p(\eta, z)$可以分解为 $p(\eta, z) = p(\eta)\prod_d p(z_d)$的形式，而没有引入任何近似。但事实上我们要求解的分布$q(\eta, z)$并不是先验分布$p(\eta, z)$，而更多地是一种对后验分布$p(\eta, z|x)$的逼近。后验分布$p(\eta, z|x)$ 并不能分解成$p(\eta|x)\prod_d p(z_d|x)$的形式（因为从图3.2看$z$和$\eta$并不相对于$x$条件独立）。在完整的搜索空间，即不对$q(\eta, z)$的形式做任何假设的函数搜索空间中，$q(\eta, z)$ 的最优解可能不是$q(\eta)\prod_d q(z_d)$的形式。但当我们假定$q(\eta, z)$为$q(\eta)\prod_d q(z_d)$的形式后，我们实际上相当于缩小和限制了$q(\eta, z)$的搜索空间。因此$q(\eta, z)$在受限制的搜索空间中找到最优解很可能不是完整搜索空间中的最优解，所以只能作为那个真正的最优解的一种近似。

在引入了Mean Field假设后，我们使用坐标下降法（Coordinate Descent），轮流对$q(\eta)$和各个$q(z_d)$进行最优化，并迭代进行。为此，我们先将Mean Field假设的形式带入问题3-25中。令问题3-25的目标函数为$L(q)$，我们得到：

$$
\begin{aligned}
L(q) &\triangleq \mathrm{KL}(q(\eta,z)\|p(\eta,z|x)) + C\sum_d \max\{0, 1 - y_d\mathbb{E}_{q(\eta,z)}[\eta^\top\Phi(x_d,z_d)]\} \\
&= -\int q(\eta)q(z)\ln\frac{p(x|\eta,z)p(\eta)p(z)}{p(x)}\mathrm{d}\eta\mathrm{d}z - H(q(\eta)) - H(q(z)) \\
&\quad + C\sum_d \max\{0, 1 - y_d\int q(\eta)\eta^\top\mathrm{d}\eta\int q(z_d)\Phi(x_d,z_d)\mathrm{d}z_d\} \\
&= -\int q(\eta)q(z)\ln p(x|\eta,z)\mathrm{d}\eta\mathrm{d}z + \mathrm{KL}(q(\eta)\|p(\eta)) + \mathrm{KL}(q(z)\|p(z)) \\
&\quad + C\sum_d \max\{0, 1 - y_d\int q(\eta)\eta^\top\mathrm{d}\eta\int q(z_d)\Phi(x_d,z_d)\mathrm{d}z_d\} + \mathrm{const} \\
&= -\sum_d\int q(\eta)q(z_d)\frac{\eta^\top\Phi(x_d,z_d)}{\|\eta\|}\mathrm{d}\eta\mathrm{d}z + \mathrm{KL}(q(\eta)\|p(\eta)) + \mathrm{KL}(q(z)\|p(z)) \\
&\quad + C\sum_d \max\{0, 1 - y_d\int q(\eta)\eta^\top\mathrm{d}\eta\int q(z_d)\Phi(x_d,z_d)\mathrm{d}z_d\} + \mathrm{const} \qquad (3\text{-}26)
\end{aligned}
$$

其中，$H(q(x))$表示概率分布$q(x)$的熵，其定义为：

$$
H(q(x)) \triangleq -\int q(x)\ln q(x)\mathrm{d}x
$$

在此，我们引入一些记号以简化公式。定义：

$$
\Phi_d^q \triangleq \mathbb{E}_{q(z_d)}[\Phi(x_d,z_d)]
$$

我们记式3-26中只与$q(\eta)$有关的部分为$L(q(\eta))$，我们可以得到：

$$
\begin{aligned}
L(q(\eta)) &= -\sum_d \Phi_d^{q\top}\mathbb{E}_q[\frac{\eta}{\|\eta\|}] + \mathrm{KL}(q(\eta)\|p(\eta)) \\
&\quad + C\sum_d \max\{0, 1 - y_d\Phi_d^{q\top}\mathbb{E}_q[\eta]\} \qquad (3\text{-}27)
\end{aligned}
$$

在此我们先尝试不对$q(\eta)$的形式做任何假设，而是用变分求导的方法，解出$q(\eta)$的最优形式。首先，我们写出$L(q(\eta))$对应的最优化问题：

$$\begin{aligned} &\min_q -\sum_d \Phi_d^{q\top}\mathbb{E}_q[\tfrac{\eta}{\|\eta\|}] + \mathrm{KL}(q(\eta)\|p(\eta)) + C\sum_d \max\{0, 1 - y_d\Phi_d^{q\top}\mathbb{E}_q[\eta]\} \\ \text{s.t.} \quad &q \in \mathcal{P}_{\text{prop}} \end{aligned} \tag{3-28}$$

然后将3-28等价地转换为引入松弛变量的形式（参考3.3节）：

$$\begin{aligned} &\min_q -\sum_d \Phi_d^{q\top}\mathbb{E}_q[\tfrac{\eta}{\|\eta\|}] + \mathrm{KL}(q(\eta)\|p(\eta)) + C\sum_d \xi_d \\ \text{s.t.} \quad &q \in \mathcal{P}_{\text{prop}} \\ &\xi_d \geq 1 - y_d\Phi_d^{q\top}\mathbb{E}_q[\eta] \quad \forall d \\ &\xi_d \geq 0 \quad \forall d \end{aligned} \tag{3-29}$$

再将问题3-29中的所有约束放入用拉格朗日方法放入目标中。为此，引入拉格朗日乘子$\omega = \{\omega_d\}_{d=1}^D$、$\alpha = \{\alpha_d\}_{d=1}^D$和$\lambda$。引入拉格朗日乘子后的拉格朗日函数形式为：

$$\begin{aligned} &L(q(\eta), \omega, \alpha, \lambda) \\ = \ &-\sum_d \Phi_d^{q\top}\mathbb{E}_q[\frac{\eta}{\|\eta\|}] + \mathrm{KL}(q(\eta)\|p(\eta)) + C\sum_d \xi_d \\ &-\sum_d \omega_d(\xi_d - 1 + y_d\Phi_d^{q\top}\mathbb{E}_q[\eta]) - \sum_d \alpha_d\xi_d - \lambda(\int q(\eta)\mathrm{d}\eta - 1) \end{aligned} \tag{3-30}$$

并有约束：

$$\begin{aligned} &\omega_d \geq 0 \quad \forall d \\ &\alpha_d \geq 0 \quad \forall d \end{aligned} \tag{3-31}$$

然后求$L(q(\eta), \omega, \alpha, \lambda)$对$q(\eta)$的变分极值点。首先整理出$L(q(\eta), \omega, \alpha, \lambda)$中与$q(\eta)$有关的项：

$$L(q(\eta), \omega, \alpha, \lambda)$$

$$
\begin{aligned}
= & -\sum_{d} \Phi_d^{q\top} \mathbb{E}_q[\frac{\eta}{\|\eta\|}] + \mathrm{KL}(q(\eta)\|p(\eta)) \\
& -\sum_{d} \omega_d y_d \Phi_d^{q\top} \mathbb{E}_q[\eta] - \lambda \int q(\eta)\mathrm{d}\eta + \mathrm{const}
\end{aligned} \tag{3-32}
$$

为使用Euler-Lagrange方程[25]，我们首先将$L(q(\eta), \omega, \alpha, \lambda)$写成

$$
L(q(\eta), \omega, \alpha, \lambda) = \int G(q(\eta), q'(\eta), \eta)\mathrm{d}\eta \tag{3-33}
$$

的形式：

$$
\begin{aligned}
& L(q(\eta), \omega, \alpha, \lambda) \\
= & \int q(\eta)(-\sum_{d} \Phi_d^{q\top} \frac{\eta}{\|\eta\|} - \ln p(\eta) + \ln q(\eta) - \sum_{d} \omega_d y_d \Phi_d^{q\top} \eta - \lambda)\mathrm{d}\eta
\end{aligned} \tag{3-34}
$$

对比3-33和3-34可以写出$G(q(\eta), q'(\eta), \eta)$来：

$$
\begin{aligned}
& G(q(\eta), q'(\eta), \eta) \\
= & \ q(\eta)(-\sum_{d} \Phi_d^{q\top} \frac{\eta}{\|\eta\|} - \ln p(\eta) + \ln q(\eta) - \sum_{d} \omega_d y_d \Phi_d^{q\top} \eta - \lambda)
\end{aligned} \tag{3-35}
$$

所谓Euler-Lagrange方程，是指如下偏微分方程：

$$
\frac{\partial G}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial G}{\partial q'}\right) = 0
$$

它的解$q^*(\eta)$，就是使$L(q(\eta), \omega, \alpha, \lambda)$达到极值的$q(\eta)$的解。因为我们得到的$G$与$q'(\eta)$无关，所以Euler-Lagrange方程可以简化为：

$$
\frac{\partial G}{\partial q} = 0 \quad \forall \eta
$$

注意其中的“对于任意$\eta$”的约束。我们令$G$对$q(\eta)$求偏导，并令其对所有$\eta$为0，得：

$$\frac{\partial}{\partial q(\eta)}q(\eta)(-\sum_d \Phi_d^{q\top}\frac{\eta}{\|\eta\|} - \ln p(\eta) + \ln q(\eta) - \sum_d \omega_d y_d \Phi_d^{q\top}\eta - \lambda) = 0$$

$$-\sum_d \Phi_d^{q\top}\frac{\eta}{\|\eta\|} - \ln p(\eta) + \ln q(\eta) - \sum_d \omega_d y_d \Phi_d^{q\top}\eta - \lambda + q(\eta)\frac{1}{q(\eta)} = 0$$

$$\ln q(\eta) = \ln p(\eta) + \sum_d \Phi_d^{q\top}\frac{\eta}{\|\eta\|} + \sum_d \omega_d y_d \Phi_d^{q\top}\eta + \lambda - 1$$

$$q(\eta) = \exp(\lambda - 1)p(\eta)\exp(\sum_d \Phi_d^{q\top}\frac{\eta}{\|\eta\|} + \sum_d \omega_d y_d \Phi_d^{q\top}\eta) \tag{3-36}$$

$\lambda$的值可由对$q(\eta)$归一化得到，所以我们将对$q(\eta)$的最终求解结果写为：

$$q(\eta) \propto p(\eta)\exp(\frac{\eta^\top}{\|\eta\|}\sum_d \Phi_d^q + \eta^\top \sum_d \omega_d y_d \Phi_d^q) \tag{3-37}$$

这是$q(\eta)$的最优函数形式，我们下一步的任务就是对$q(\eta)$积分，求出其归一化因子（可能与$\omega_d$、$\Phi_d^q$等的值有关），并将求解出的$q(\eta)$的完整形式带入式3-30以转化为对于$\omega$、$\alpha$等拉格朗日乘子进行最优化的对偶问题。但在这里我们遇到了困难。对式3-37求积分极其困难，要解出它的与$\omega_d$、$\Phi_d^q$有关的归一化因子$A(\omega,\alpha)$几乎不可能，更不用说带回 3-30写出对偶问题。我们只好放弃使用$q(\eta)$的最优函数形式，转而对它的函数形式作出假设并假设成由参数控制，以使得对$q(\eta)$的泛函最优化变为对参数的最优化。

### 3.5.9 径向、角度分离

仔细观察式3-27，我们发现，除了KL距离项外，$L(q(\eta))$与$q(\eta)$的关系体现在期望$\mathbb{E}_q[\frac{\eta}{\|\eta\|}]$和期望$\mathbb{E}_q[\eta]$。因为$\frac{\eta}{\|\eta\|}$表示$\eta$的单位方向向量，所以我们可以将$\mathbb{E}_q[\frac{\eta}{\|\eta\|}]$和$\mathbb{E}_q[\eta]$视作$\eta$的“平均方向”和“平均向量”，其中“平均向量”既包含了“平均方向”的信息，又包含了“平均长度”的信息。对于一般的分布形式$q(\eta)$，“平均方向”很难显示地写出来，但如果分布$q(\eta)$可以拆成两部分的乘积，一部分控制概率在方向空间的分布，一部分控制概率在半径上的分布，并假设这两者独立，那么“平均方向”、“平均长度”、“平均向量”都可以很简单地算出来。

沿着这个思路我们将向量$\eta$拆成两部分，一部分表示方向，一部分表示长度。即令：

$$\vec{\eta} = r\vec{e},\ \left\|\vec{e}\right\| = 1,\ r > 0$$

其中$e$是$\eta$的同方向单位向量，$r$是$\eta$的模。将$\eta$表示成这样后，我们假设$\eta$的分布$p(\eta)$可以分解为方向分布和径向分布两部分，即假设：

$$q(\eta) = \frac{1}{A_2} f_{\mu_e}(e) f_{\mu_r}(r)$$

其中$f_{\mu_e}(e)$表示$\eta$在方向空间的概率分布，分布由参数$\mu_e$控制。$f_{\mu_r}(r)$表示$\eta$在径向空间的概率分布，分布由参数$\mu_r$控制。注意$f_{\mu_e}(e)$和$f_{\mu_r}(r)$并不需要归一化，它们只表示$q(\eta)$可分解为两个因子。此式同时还对$f_{\mu_e}(e)$和$f_{\mu_r}(r)$的形式做了一个要求，即要求$f_{\mu_e}(e)$和$f_{\mu_r}(r)$的积分结果与参数$\mu_e$和$\mu_r$无关，即使得$p(\eta)$的归一化因子可以写为常数$A_2$。这个要求很容易达到，只需要求出$f_{\mu_e}(e)$和$f_{\mu_r}(r)$的积分结果（可能和参数$\mu_e$和$\mu_r$有关），然后将和参数$\mu_e$和$\mu_r$有关的部分并入$f_{\mu_e}(e)$和$f_{\mu_r}(r)$即可。 $f_{\mu_e}(e)$和$f_{\mu_r}(r)$的形式可以有很多选择。对于$f_{\mu_e}(e)$，可以选择的形式可以是类如式3-15的指数函数：

$$f_{\mu_e}(e) = \exp(\mu_e^{\top} e)$$

因为两个向量的内积表达了它们之间的夹角的余弦，所以我们也可以直接以两个向量的夹角作为指数函数的输入。实现方式是使用反余弦函数作用在内积上，$f_{\mu_e}(e)$的形式定义为：

$$f_{\mu_e}(e) = \exp(-\arccos \mu_e^{\top} e) \tag{3-38}$$

因为反余弦函数是单调减函数，所以我们取了负号。对于$f_{\mu_r}(r)$，它表达了概率密度在半径上的分布。它的形状并不重要，这是因为，在线性SVM中，作为过零点的分类面（的法向量），$\eta$的方向才是决定分类正确与否的最重要

因素，而至于$\eta$的大小，则只起到能够使训练样例被推到间隔区之外的作用。相似的，$q(\eta)$在半径方向上的分布也只是起到将训练样例推到间隔区之外的效果。特别地，如果最后使用到的只是这个径向分布的期望值，那么这个期望值的作用将和SVM中$\eta$的模一样。

因为$\eta$的径向分布不重要，我们可以选择几个容易处理的分布形式。可选的分布包括高斯形式：

$$f_{\mu_r}(r) = \frac{1}{\mu_r}\exp(-\frac{r^2}{2\mu_r^2}) \tag{3-39}$$

以及Weibull分布：

$$f_{\mu_r}(r) = \frac{1}{\mu_r^2} r \exp(-\frac{r^2}{2\mu_r^2})$$

后者的峰值可以移动，从而不需要像高斯形式那样峰值固定在原点。但就像前文所说，径向分布的形状并不重要，重要的是它的期望值。注意，两者都包含了归一化因子中带有分布参数的部分，以使得$q(\eta)$的归一化因子是一个常数。

我们以$f_{\mu_e}(e) = \exp(-\arccos \mu_e^\top e)$和$f_{\mu_r}(r) = \frac{1}{\mu_r^n}\exp(-\frac{r^2}{2\mu_r^2})$为例子来展示对参数$\mu_e$和$\mu_r$的最优化。将式3-38和式3-39带入 3-27，并假设$p(\eta) \sim \mathcal{N}(0,1)$，我们有：

$$\begin{aligned} L(\mu_e, \mu_r) &= -\frac{1}{A_3}\sum_d \Phi_d^{q\top}\mu_e - n\ln\mu_r + \frac{1}{A_4}\mu_r^2 \\ &\quad + C\sum_d \max\{0, 1 - \frac{1}{A_5} y_d \Phi_d^{q\top}\mu_e\mu_r\} \end{aligned} \tag{3-40}$$

其中$A_3$、$A_4$、$A_5$都是与待优化参数无关的常数（只与$\eta$的维数有关）。这个优化问题同样可以通过循环迭代求解$\mu_e$和$\mu_r$实现。对于求解$\mu_r$，因为$\mu_r$只是一个单维实变量，而且式3-40 对于$\mu_r$是凸的（因此是单峰的），因此对$\mu_r$的最优化可以通过黄金分割搜索（Golden Section Search）[26]来以$-\ln\epsilon$时间复杂度实

现，$\epsilon$为要求的精度。对于最优化$\mu_e$，则可以通过在方向空间的梯度下降（即每一步向着目标值下降最快的方向转动一个小角度）来实现。

### 3.5.10 求解$q(z)$

对$q(z)$的求解，相对来说比较简单。因为$z$是离散变量，所以我们可以不加任何假设地将它的分布全部用参数表示出来。我们将$z_d$的分布$q(z_d)$表示为：

$$q(z_d = i) = \phi_d^i \quad 0 \le \phi_d^i \le 1 \quad i = 1, 2, \cdots, K_d \quad \sum_i \phi_d^i = 1$$

其中$K_d$表示$z_d$的所有可能取值个数。为简化公式，我们引入一些简化记号，记：

$$\Psi_{d,z_d}^q \triangleq \Phi(x_d, z_d)^\top \mathbb{E}_q[\eta]$$

以及

$$\hat{\Psi}_{d,z_d}^q \triangleq \Phi(x_d, z_d)^\top \mathbb{E}_q[\frac{\eta}{\|\eta\|}]$$

将问题3-25整理出只与$q(z_d)$有关的部分，我们得到：

$$\begin{aligned} L(q(z_d)) &= -\mathbb{E}_{q(z_d)}[\hat{\Psi}_{d,z_d}^q] + \mathrm{KL}(q(z_d)\|p(z_d)) \\ &\quad + C\sum_d \max\{0, 1 - y_d \mathbb{E}_{q(z_d)}[\Psi_{d,z_d}^q]\} \end{aligned} \tag{3-41}$$

将它表示成与各参数$\phi_d^i$有关的形式，我们得到：

$$\begin{aligned} L(\phi_d^i) &= -\phi_d^i \hat{\Psi}_{d,z_d}^q + \phi_d^i \ln \phi_d^i - \phi_d^i \ln p(z_d = i) \\ &\quad + C \max\{0, 1 - y_d \phi_d^i \Psi_{d,z_d}^q\} \end{aligned} \tag{3-42}$$

式3-42用梯度下降法求解。由于该目标函数受约束$\sum_i \phi_d^i = 1$和$0 \le \phi_d^i \le 1$限制，所以还必须考虑约束。但因为约束为线性约束，所以我们可以使用可行方向法[27]做梯度下降。

## 3.6 本章小结

本章中，我们逐步提出并完善了MedLSVM模型，讨论了设计最大间隔模型的几项关键技术、两种LSVM及其比较、为隐变量引入概率分布的方法、概率产生模型的必要性及其设计方法、完整MED框架的好处、求解MedLSVM的方法等。围绕MedLSVM模型发展出了一套完整的概率模型和数学方法。

# 第 4 章 实验

## 4.1 本章概述

在本章，我们通过多种实验来从各方面考察MedLSVM框架的特性和行为，以发现它的长处和弱点。在本章中，我们将依次考察各种似然函数的行为、效果和最终选取结果，MedLSVM的一种softmax快速实现在完整物体检测任务上的性能，以及MedLSVM的另一种冲击分布快速实现在完整物体检测任务上的性能。

## 4.2 似然函数的选取

如3.5.5所描述，我们有多种似然函数的形式可以用来描述在给定$z$和$\eta$的条件下$x$的随机产生方式。这些可选的形式包括

$$p(x|z,\eta) \propto \exp(\eta^{\top}\Phi(x,z)) \tag{4-1}$$

$$p(x|z,\eta) \propto \exp(-\frac{1}{2}(\eta - \Phi(x,z))^2) \tag{4-2}$$

$$p(x|z,\eta) \propto \exp(-|\eta - \Phi(x,z)|) \tag{4-3}$$

我们现在通过一个简化模型下的实验，来考察几种似然函数的行为。在该模型下，$x$表示一幅图像的HOG特征金字塔，$\eta$表示分类器参数，同时也是一个HOG特征检测模板。例如，我们以自行车检测器为例，一幅示例的含有自行车的待检测图像如图4.1

对它的各个分辨率抽取HOG特征，得到HOG特征金字塔，如图4.2。$\eta$表示的HOG特征检测模板可以可视化为图4.3。对HOG特征的可视化方法，是在每

图 4.1　含有自行车的图像示例

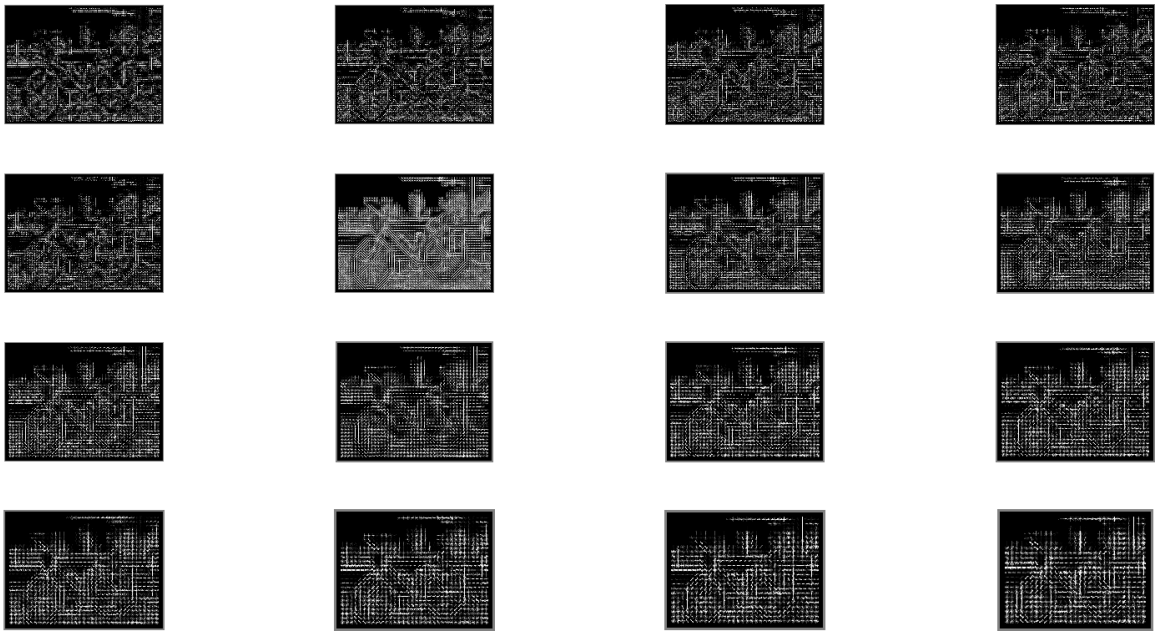

图 4.2　示例图片的HOG特征金字塔

个HOG格中，按各个方向桶的获得梯度向量个数的多少，以不同亮度将各个方向桶用那个方向的小线段表示出来。

隐变量$z$，在我们的简化问题中，表示待检测物体在这幅图像中所出现的位置，即$(x, y, l)$，既包含了位置$(x, y)$，又包含了分辨率层次。于是，我们可以画出在隐变量$z$的各个取值下，似然函数值$p(x|z, \eta)$的大小。尤其是，由于$z$的取值代表各个位置，我们可以将$p(x|z, \eta)$的大小画成伪彩色图。我们观察似然函数$p(x|z, \eta)$的值在各个$z$取值下的分布图，来考察$p(x|z, \eta)$的性质是否如我们所愿，即值大的地方是$z$应该取的值，即物体应该出现的位置。

使用式4-1，我们画出在$z$的各个取值下$p(x|\eta, z)$的伪彩色图，得到图4.4。

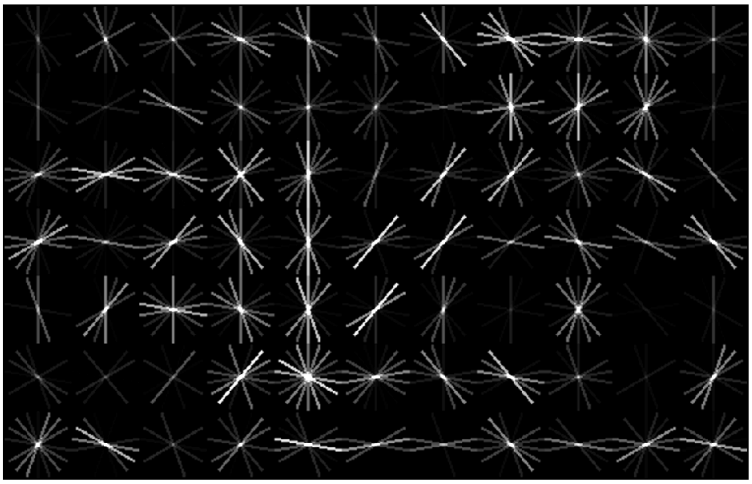

图 4.3　自行车的检测模板

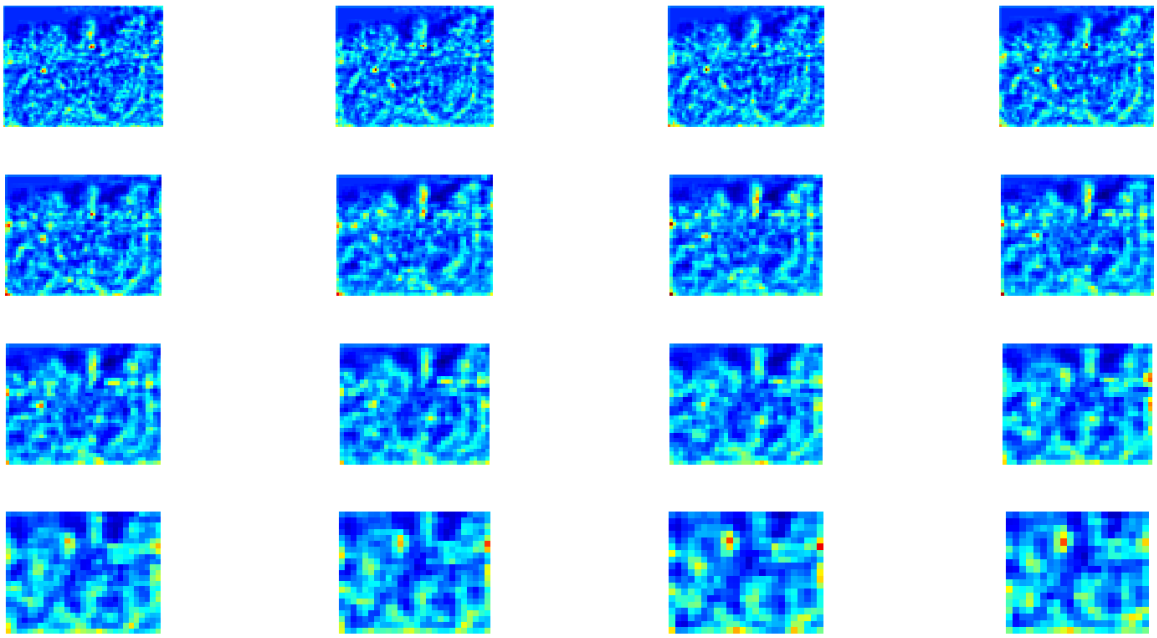

图 4.4　似然函数4-1的伪彩色图

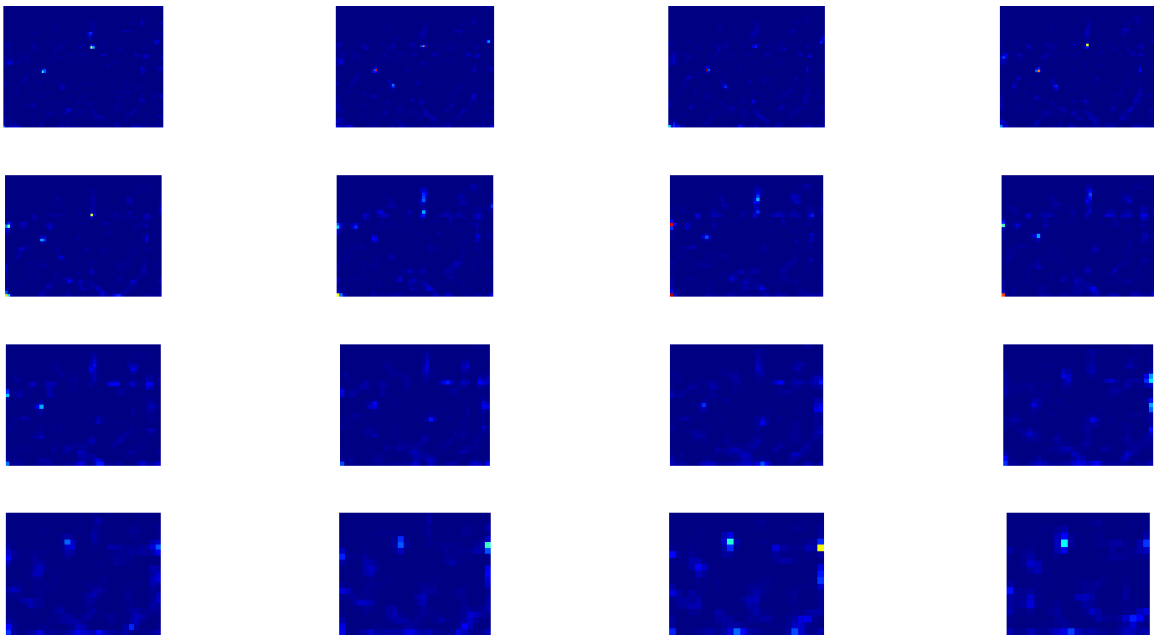

图 4.5　当$c$ = 5时似然函数式4-4的伪彩色图

在图4.4中，红色区域表示值较大的地方，蓝色区域表示值较小的地方。像素位置代表隐变量$z$的取值（物体位置及分辨率层次）。物体位置是指当模板的左上角与该点对齐时，模板在图像上的得分值就填在该点。从图4.4中可以看出，一些外观比较像自行车的位置，会出现突出的红点，表示此处的值远远高于它的邻居，有较好的选择性。但图4.4显示的背景区域的得分依然很嘈杂，如果以此概率对$z$取平均值，背景区域可能会对得分有较大影响。为此，我们可以通过参数控制似然函数中指数函数的陡峭程度。也就是，我们把式4-1扩充为：

$$p(x|z,\eta) \propto \exp(c\eta^{\top}\Phi(x,z)) \tag{4-4}$$

它引入了控制指数函数陡峭程度的参数c。当c越大时，指数函数越陡峭，$z$的概率分布就越集中，得分值高处就突出地越明显。例如，当c的值取5时，得到的似然函数取值伪彩色图如图4.5。

图4.5中的非零取值基本只出现在了几个得分较大的地方。但是注意提高分布陡峭程度后也并没有导致只有一个最大点非零的情况，这说明为$z$引入概率分布和取期望的操作的结果与max不相同，并没有退化成max操作，从而可以带来新的信息。尤其是，相同的位置，在不同的分辨率层次上，可能出现多个得分

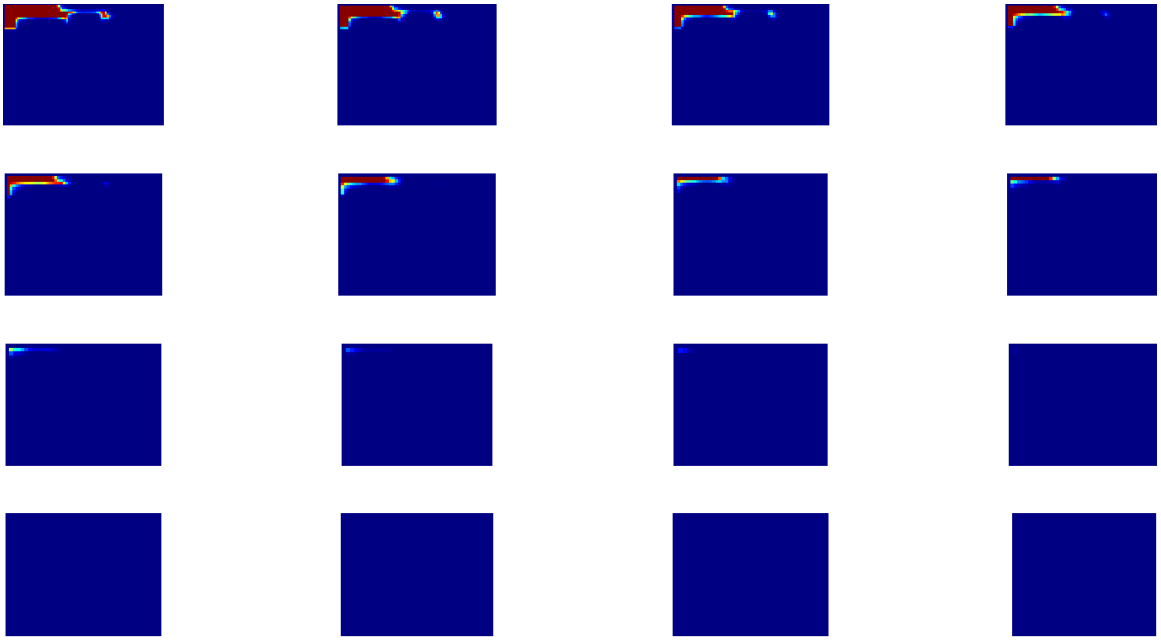

图 4.6　似然函数4-2的伪彩色图

近似的最优点，从而达到使用多个分辨率层次做平均的效果，可以避免只使用一个分辨率而带来的误差（分辨率不能连续变化，而只能被人为地分成几个离散层次，所以只使用某一个分辨率层次可能带来量化误差）。

使用高斯形式的似然函数4-2，我们得到伪彩色图图4.6。

图4.6可以明显地体现出高斯形式似然函数的问题。图中得分最高的区域，集中在图像的左上部。而查看原图像图4.1，我们发现图像左上部是玻璃窗户。这部分图像的特点是几乎没有任何纹理或边缘，也就是像素梯度几乎都为0，也就是HOG特征几乎都为0。观察对似然函数4-2的展开和分析3-17，我们可以看到高斯形式的似然函数等效于在指数形式似然函数4-1的基础上乘上一个因子$\exp(-\frac{1}{2}\|\Phi(x,z)\|^2)$，这个因子会放大$\|\Phi(x,z)\|$小的地方，而缩小$\|\Phi(x,z)\|$大的地方。所以高斯形式的似然函数4-2会突出图像中空白一片而没有任何边缘和纹理的区域，例如图4.1的左上部。这也就意味着，高斯形式的似然函数不适合作为我们的似然函数形式。

指数使用1范数的似然函数4-3的伪彩色图为图4.7。

同样的现象，该似然函数也会突出图像中没有任何变化的区域。它的伪彩色图与4-2的图很像，说明使用2范数距离作为指数与作为1范数作为指数在似然

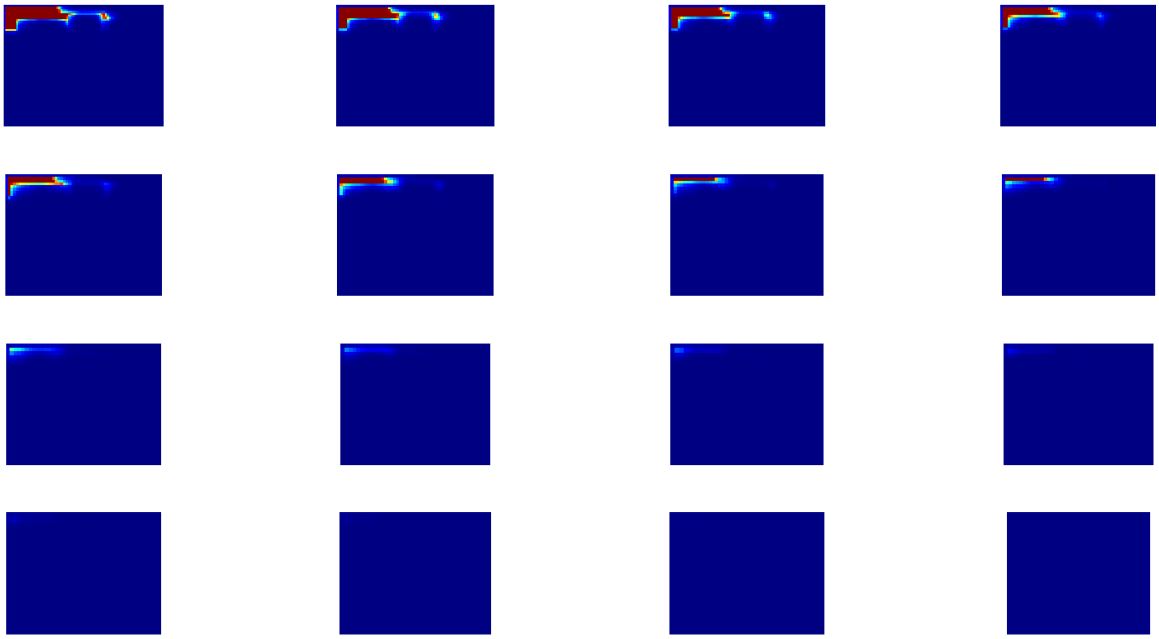

图 4.7　似然函数4-3的伪彩色图

函数选取这个问题上效果差不多。

综上，我们通过经验实验选取x和z、$\eta$之间似然函数的结论是，指数函数的似然函数4-1在该问题中更合适。

## 4.3　MedLSVM用于基于部件的物体检测

当面临完整的基于部件的物体检测时，因为对于每幅图像，既要做滑动窗口穷举，又要在每个滑动窗口内对每个部件的所有位置做穷举，计算量过大。所以要将MedLSVM模型用于物体检测任务，必须引入一些简化以及设计快速算法。我们在这里使用两种从完整的MedLSVM框架推导出的快速算法来进行物体检测实验。

### 4.3.1　实验数据集描述

下面，我们把我们发展出的MedLSVM模型及其几个快速实现用在完整的基于部件物体检测任务上。首先，我们描述一下我们使用的数据集。

我们使用公开数据集Pascal VOC 2007物体检测子任务数据集。该数据集包含20个物体类别，各类别名称参见表4.1和表4.2。数据集被提供方分成了不相

图 4.8　VOC Pascal 2007数据集bird类图片示例

交的训练集和测试集。对于每一类，训练集包含300到1000张不等的正例（包含该类别物体的）图片和2000张左右的负例图片；测试集包含600到2000张不等的正例图片和4000张左右的负例图片。类与类之间的训练集和测试集图片都有重合，每个类的训练和测试可以视为与其他类无关的任务，比较结果时，既可以用在某一类上的性能进行比较，也可用所有20类进行比较。在用所有20类进行比较时，整体的得分是取20类各自得分的评价值（不使用图片数量加权）。

对于某一类别的检测任务的得分，定义为平均精度（AP）或称为ROC面积。首先，使用训练好的检测器在测试图片上检测时，伴随每一个输出的检测结果（包围框），检测器会汇报一个可信度得分（例如可使用2.2.4节定义的模型得分）。每当对这个可信度得分取一个阈值，我们就得到了一组汇报为正例的检测结果。我们可以在这组检测结果上计算准确率（Precison）和召回率（Recall）。准确率定义为检测结果中正确的个数除以检测结果的个数，召回率定义为测试集中标注的物体被命中的个数除以标注的物体个数。“正确”和“被命中”都定义为检测结果矩形和标注矩形的重叠部分面积除以各自的面积都不小于0.7。

数据集图片的示例如图4.8和图4.9。

用于同我们的模型进行比较的对象，是[4]中提出的LSVM部件模型，我们此后将它简称为“基准模型”。基准模型是当前在Pascal VOC 2007数据集上性能最好的模型之一。由于它实现了多视角混合模型的功能，我们只与它的混合个

图 4.9　VOC Pascal 2007数据集sofa类图片示例

数为1的情况作比较。在最终输出检测结果包围框时，它使用了两种方法：一种是直接输出主模板对应的矩形边框；一种是利用训练集中标注的包围框，用回归方法拟合出这个包围框与检测结果中各部件位置的关系，然后在测试时，通过各部件的位置推测出包围框。我们把这两种包围框输出方法对应的评测结果分别称为AP1和AP2。基准模型得到的各个类的评测结果如表4.1和表4.2

表 4.1　基准模型的检测性能

| | aeroplane | bicycle | bird | boat | bottle | bus | car | cat | chair | cow |
|---|---|---|---|---|---|---|---|---|---|---|
| AP1 | 0.267 | 0.441 | 0.094 | 0.012 | 0.228 | 0.360 | 0.462 | 0.128 | 0.141 | 0.172 |
| AP2 | 0.273 | 0.458 | 0.095 | 0.012 | 0.240 | 0.374 | 0.481 | 0.127 | 0.139 | 0.171 |

表 4.2　基准模型的检测性能（续）

| table | dog | horse | motor | person | plant | sheep | sofa | train | tv | avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.173 | 0.105 | 0.390 | 0.365 | 0.297 | 0.124 | 0.148 | 0.221 | 0.218 | 0.386 | 0.237 |
| 0.172 | 0.105 | 0.395 | 0.374 | 0.306 | 0.127 | 0.148 | 0.223 | 0.220 | 0.389 | 0.241 |

### 4.3.2　使用softmax实现

#### 4.3.2.1　MedLSVM的另一视角

到目前为止，我们通过为隐变量引入分布，并合适地选取隐变量与观测

变量之间的概率关系，发展出了MedLSVM模型。从另一个视角看，我们也可以发现我们推导出的MedLSVM模型有着直接而自然的物理含义。如果我们选用$p(x|z,\eta) \propto \exp(\eta^\top \Phi(x,z))$作为我们的似然函数形式，并且假设$z$的先验分布为均匀分布，即$p(z) = \text{const}$，那么我们可以得到在已知$x$和$\eta$的条件下$z$的后验概率分布：

$$p(z|x,\eta) = \frac{\exp(\eta^\top \Phi(x,z))}{\sum_{z'} \exp((\eta^\top \Phi(x,z'))}$$

3.5.8节已经说过，$q(z)$是一种对后验概率概率$p(z|x,\eta)$的逼近。在$q(z)$完全等于$p(z|x,\eta)$的情况下，我们的MedLSVM的分类准则（在为引入$\eta$的分布之前）3-11将成为：

$$\begin{aligned} y^*(\eta) &= \underset{y\in\{1,-1\}}{\text{argmax}} \left(y \sum_z \frac{\exp(\eta^\top \Phi(x,z))}{\sum_{z'} \exp((\eta^\top \Phi(x,z'))} \eta^\top \Phi(x,z)\right) \\ &= \underset{y\in\{1,-1\}}{\text{argmax}} \left(y \underset{z}{\text{softmax}}\ \eta^\top \Phi(x,z)\right) \end{aligned} \tag{4-5}$$

其中的softmax定义为：

$$\underset{z}{\text{softmax}}\ (f(z)) \triangleq \sum_z \frac{exp(f(z)}{\sum_{z'} exp(f(z'))} f(z) = \frac{\sum_z exp(f(z)) f(z)}{\sum_z exp(f(z))} \tag{4-6}$$

将式4-5与LSVM的分类准则2-7相比较，我们可以发现我们的MedLSVM是用softmax代替LSVM中max的地位。

#### 4.3.2.2　softmax的快速实现

仿照2.4节介绍的快速检测算法，我们也可以设计出适用于softmax操作的快速算法。使用相同的$R_{c,l}$和$D_{c,l}$定义，对于我们的模型，根据式 4-6，我们可以得到$D_{c,l}$的计算式：

$$D_{c,l}(i,j) = \frac{\sum_{d_i,d_j} \exp(R_{c,l}(i+d_i, j+d_j) + v^\top \phi(d_i,d_j))(R_{c,l}(i+d_i, j+d_j) + v^\top \phi(d_i,d_j))}{\sum_{d_i,d_j} \exp(R_{c,l}(i+d_i, j+d_j) + v^\top \phi(d_i,d_j))} \tag{4-7}$$

其中$v$仍表示形变惩罚系数，$\phi(d_i, d_j)$的定义为：

$$\phi(d_i, d_j) \triangleq (d_i, d_i^2, d_j, d_j^2)$$

我们引入几个简化记号，简记$D_{c,l}$、$R_{c,l}$为D、R，定义：

$$\psi(d_i, d_j) \triangleq v^\top \phi(d_i, d_j)$$

并引入几个矩阵运算：

$$A_{m,n} \odot B_{m,n} \triangleq (a_{ij} b_{ij})_{m,n}$$

表示元素级乘法，

$$e^{A_{m,n}} \triangleq (e_{ij}^a)_{m,n}$$

表示元素级指数运算，

$$A \otimes B(i, j) \triangleq \sum_{d_i, d_j} A(i + d_i, j + d_j) B(d_i, d_j)$$

表示二维（反向）卷积。那么式4-7可以写为：

$$D = \frac{(e^R \odot R) \otimes e^\psi + e^R \otimes (e^\psi \odot \psi)}{e^R \otimes e^\psi}$$

重写为卷积运算与矩阵元素级运算的目的，是因为许多数值计算库可以快速地（并行地）实现这两种基本运算，尤其是卷积运算，可以通过快速傅里叶变换转换到频域空间进行，可大大加快运算速度。

#### 4.3.2.3 实验结果

我们将MedLSVM模型的快速softmax实现在VOC Pascal 2007数据集全部20个类上的实验结果列在了表4.3和表4.4，并将其与基准模型的性能做比较。表中加粗数字表示在对应包围框输出方法（AP1 或AP2）下的较好结果。

表 4.3 MedLSVM模型的快速softmax实现在VOC Pascal 2007数据集全部20个类上的测试结果及其与基准模型的比较

| | aeroplane | bicycle | bird | boat | bottle | bus | car | cat | chair | cow |
|---|---|---|---|---|---|---|---|---|---|---|
| AP1(base) | 0.267 | 0.441 | 0.094 | 0.012 | **0.228** | 0.360 | **0.462** | **0.128** | **0.141** | 0.172 |
| AP2(base) | **0.273** | 0.458 | 0.095 | 0.012 | **0.240** | 0.374 | 0.481 | **0.127** | **0.139** | 0.171 |
| AP1(softmax) | 0.240 | **0.448** | **0.096** | **0.033** | 0.187 | **0.364** | 0.461 | 0.087 | 0.133 | **0.174** |
| AP2(softmax) | 0.247 | **0.458** | **0.097** | **0.033** | 0.190 | **0.380** | **0.483** | 0.089 | 0.134 | **0.175** |

表 4.4 MedLSVM模型的快速softmax实现在VOC Pascal 2007数据集全部20个类上的测试结果及其与基准模型的比较（续）

| table | dog | horse | motor | person | plant | sheep | sofa | train | tv | avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.173 | **0.105** | 0.390 | **0.365** | **0.297** | **0.124** | 0.148 | **0.221** | **0.218** | 0.386 | **0.237** |
| 0.172 | **0.105** | 0.395 | 0.374 | **0.306** | **0.127** | 0.148 | **0.223** | **0.220** | 0.389 | **0.241** |
| **0.255** | 0.051 | **0.404** | 0.360 | 0.295 | 0.119 | **0.167** | 0.151 | 0.179 | **0.386** | 0.229 |
| **0.251** | 0.052 | **0.411** | **0.374** | 0.303 | 0.118 | **0.168** | 0.148 | 0.183 | **0.390** | 0.233 |

AP1的柱形图比较见图4.10。AP2的柱形图比较见图4.11。

从结果来看，我们的算法在11个类别上超过了基准模型，平均正确率略低于基准模型。观察实验比较结果，我们发现我们的模型在diningtable这一类领先较多。diningtable这一类的示例图片见图4.12。这一类中的桌子与桌子之间差别很大，但由于桌子的部件的变化范围不大，所以对部件位置的概率平均可以使带来的好处大于坏处。反观dog类，我们的模型在这一类落后比较多，这一类的实例图片见图4.13。dog 与dog之间的差别并不比桌子与桌子之间大，但dog的部件的变化范围很大（四肢可以处于各个位置，而且会伴有旋转，我们的模型只能处理部件平移，不能处理旋转），所以对部件位置取概率平均可能会完全抹平掉信息，拉低检测性能。另外，由于各个类别都含有视角非常不

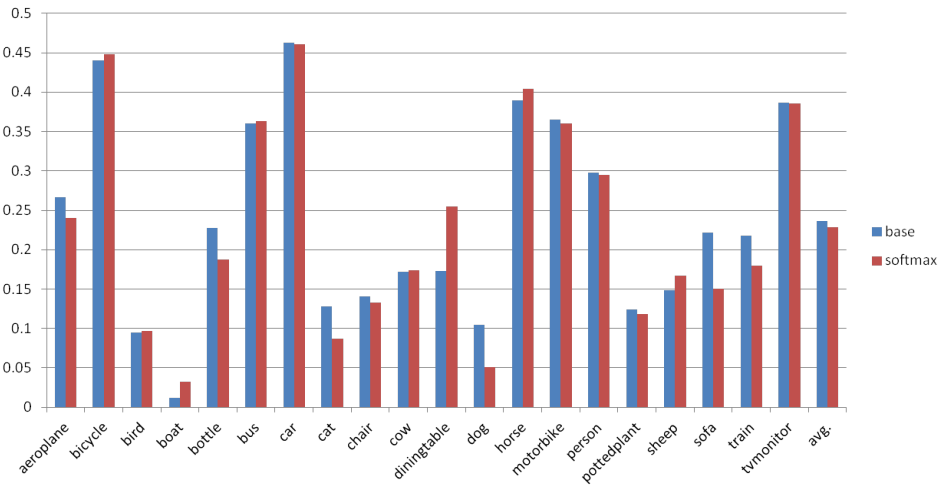


图 4.10　MedLSVM模型的快速softmax实现与基准模型在AP1方法下测试性能的比较

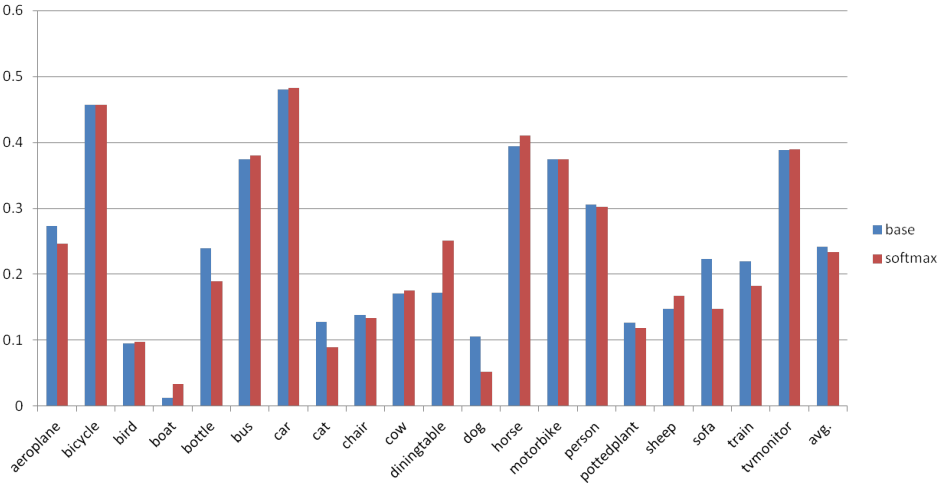


图 4.11　MedLSVM模型的快速softmax实现与基准模型在AP2方法下测试性能的比较

图 4.12　VOC Pascal 2007数据集diningtable类图片示例

图 4.13　VOC Pascal 2007数据集dog类图片示例

同的物体，所以很难用同一个物体模型来描述各个视角下的物体（例如正面的自行车与侧面的自行车完全不同）。因此，在没有办法为每一个视角单独训练一个物体模型的情况下，单视角模型在多视角数据集上的性能优劣比较难以得到直观的解释。

作为例子，bicycle类和person类的ROC曲线如图4.14和图4.15。在这两类上训练出的检测器模型的可视化表示见图4.16和图4.17。在可视化表示图中，左侧为主模板的HOG可视化表示，表示方法参见4.2节。中间为各部件模板的可视化表示，部件模板的位置反应了锚点位置。右侧为形变惩罚系数的可视化表示，它用像素亮度表示了在不同程度地偏离锚点（部件矩形中心）时惩罚分数

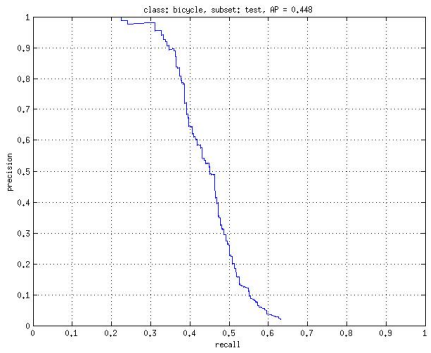


图 4.14　MedLSVM模型的快速softmax实现在bicycle类上的测试结果的ROC曲线

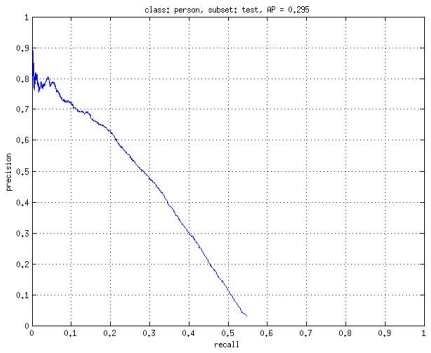


图 4.15　MedLSVM模型的快速softmax实现在person类上的测试结果的ROC曲线

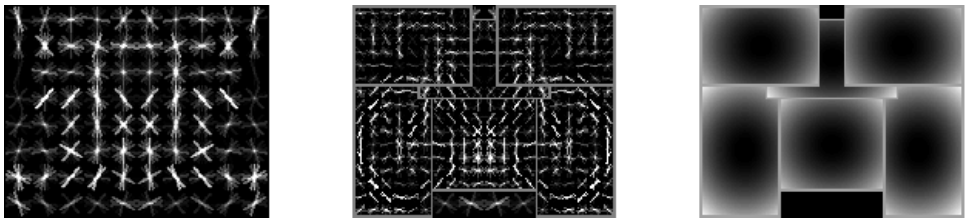

图 4.16　MedLSVM模型的快速softmax实现在bicycle类上训练得到的模型的可视化表示

的大小。可以看到部件模板可以在较细节的层面上捕捉到车轮、头等部位的形状。

bicycle类和person类的示例检测结果如图4.18和图4.19。其中红色矩形表示输出的包围框，蓝色矩形表示各个部件的最佳摆放位置。从对person类的检测结果可以看到因为我们使用了滑动窗口，所以图像中出现的多个目标物体都可以被检测到。

### 4.3.3　使用冲击分布实现

#### 4.3.3.1　使用冲击分布简化模型

为了减少模型的参数个数和运算量，我们可以将优化问题中待优化的分布$q(\eta)$的取值空间限制在冲击（单点）分布集合中，从而将对$q(\eta)$的优化简化为对单个参数的优化。为说明冲击分布的效果，我们先在MED模型上进行考察。在MED模型中，如果我们令$q(\eta) = \delta_{\mu}(\eta)$，其中$\delta_{\mu}(\eta)$是以$\mu$为中心的冲击分布，

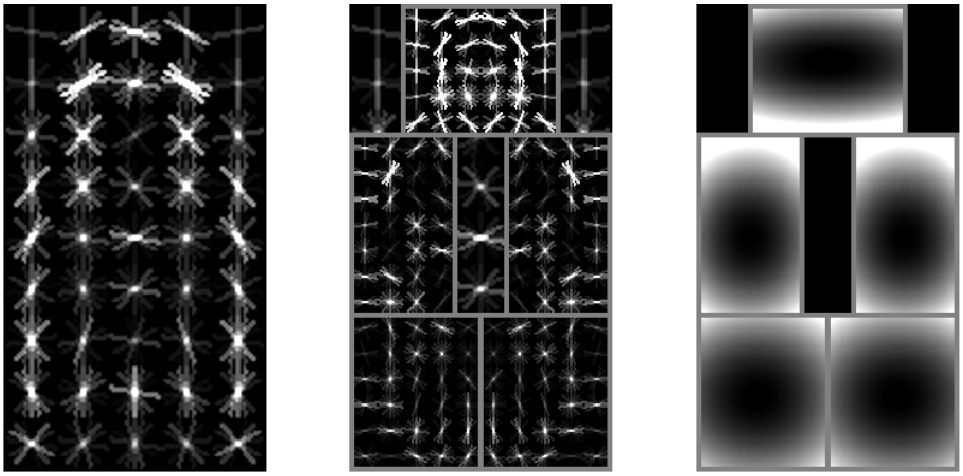

图 4.17　MedLSVM模型的快速softmax实现在person类上训练得到的模型的可视化表示

它定义为：

$$\delta_\mu(\eta) = 0, \forall \eta \neq \mu \text{ and } \int \delta_\mu(\eta) \mathrm{d}\eta = 1$$

而且令MED模型中的$p_0(\eta) \sim \mathcal{N}(0,1)$，那么学习MED模型的最优化问题将会成为：

$$\min_\mu \frac{1}{2} \|\mu\|^2 - H(q(\eta)) + C \sum_d \max\{1 - y_d \mu^\top x_d\}$$

对于冲击分布，熵$-H(q(\eta))$会成为负无穷。但对于一个普通分布，只要分布的形状不变，如果分布被平移，熵的值不变。由于冲击分布可以看做矩形或高斯分布的极限，所以我们可以将冲击分布的熵看做一个与其参数$\mu$ 无关的大常数，从而在对$\mu$的优化问题中忽略它。在这之后，使用冲击分布的MED模型将会变成一个标准SVM。

相似的，在MedLSVM中，令$q(\eta) = \delta_\mu(\eta)$，设$p(\eta) \sim \mathcal{N}(0,1)$，并忽略$H(q(\eta))$。如果使用似然函数$p(x|z,\eta) \propto \exp(\eta^\top \Phi(x,z))$，我们将得到最优化问题：

$$L(q(\eta)) = \frac{1}{2} \|\mu\|^2 - \mu^\top \sum_d \Phi_d^q + \sum_k \ln \frac{\exp(\mu) - 1}{\mu}$$

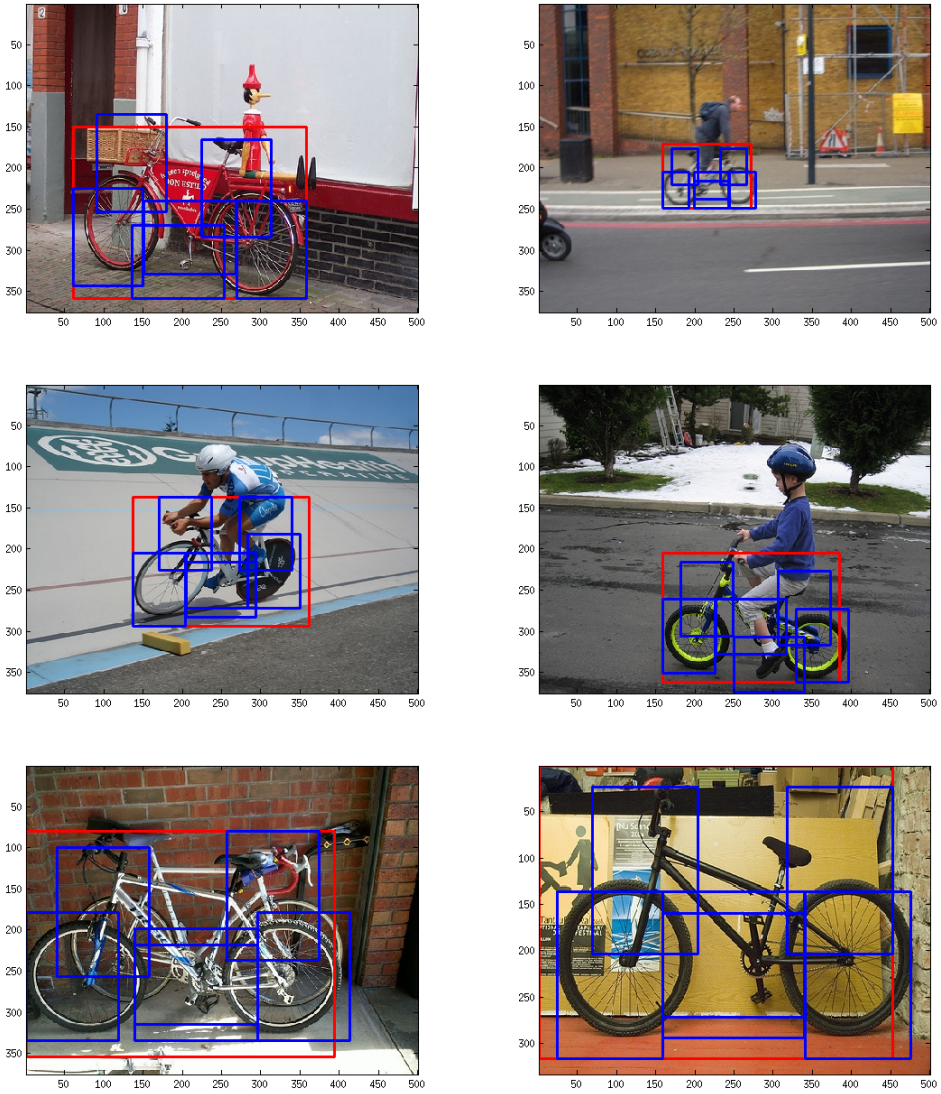

图 4.18　MedLSVM模型的快速softmax实现在bicycle类的检测结果。其中红色矩形表示输出的包围框，蓝色矩形表示各个部件的最佳摆放位置。

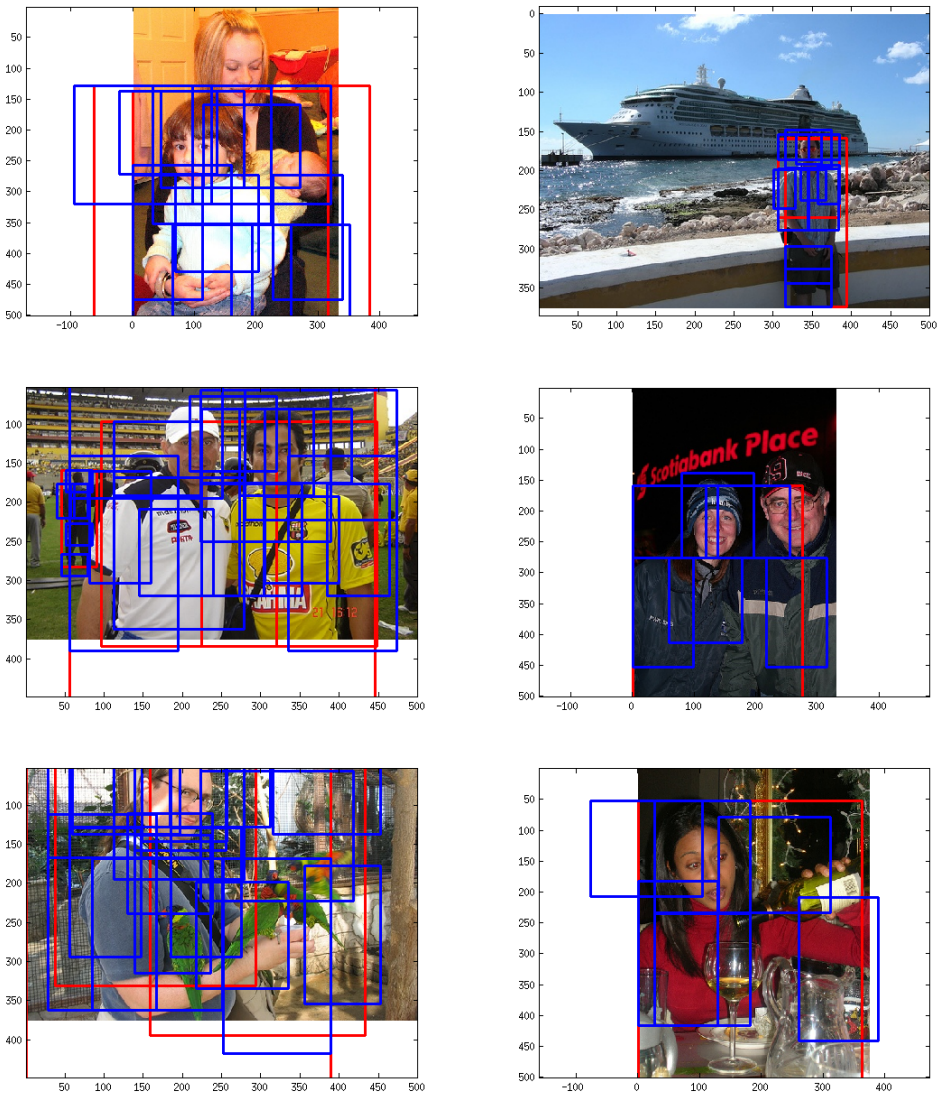


图 4.19　MedLSVM模型的快速softmax实现在person类的检测结果。其中红色矩形表示输出的包围框，蓝色矩形表示各个部件的最佳摆放位置。

$$+C\sum_{d}\max\{1-y_d\mu^{\top}\Phi_d^q\} \tag{4-8}$$

如果使用似然函数$p(x|z,\eta)\propto\exp(\frac{1}{\|\eta\|}\eta^{\top}\Phi(x,z))$，我们将得到最优化问题：

$$L(q(\eta)) = \frac{1}{2}\|\mu\|^2-\frac{\mu^{\top}}{\|\mu\|}\sum_{d}\Phi_d^q+C\sum_{d}\max\{1-y_d\mu^{\top}\Phi_d^q\} \tag{4-9}$$

下面的实验中，我们使用第一种似然函数，即问题4-8。因为函数$\ln\frac{\exp(\mu)-1}{\mu}$对于$\mu$是凸的，所以问题4-8是凸优化问题，可以使用梯度下降求解。

#### 4.3.3.2 实验结果

使用冲击函数实现的MedLSVM模型在Pascal VOC 2007全部20个类别上的测试结果及其与基准模型的比较见表4.5和表4.6。

表 4.5 使用冲击分布的MedLSVM模型实现在VOC Pascal 2007数据集全部20个类上的测试结果及其与基准模型的比较

| | aeroplane | bicycle | bird | boat | bottle | bus | car | cat | chair | cow |
|---|---|---|---|---|---|---|---|---|---|---|
| AP1(base) | 0.267 | 0.441 | 0.094 | **0.012** | 0.228 | **0.360** | **0.462** | 0.128 | 0.141 | 0.172 |
| AP2(base) | 0.273 | **0.458** | **0.095** | **0.012** | 0.240 | 0.374 | 0.481 | 0.127 | 0.139 | 0.171 |
| AP1(impulse) | **0.267** | **0.448** | **0.095** | 0.003 | **0.250** | 0.357 | 0.454 | **0.155** | **0.145** | **0.201** |
| AP2(impulse) | **0.275** | 0.456 | 0.094 | 0.003 | **0.254** | **0.377** | **0.481** | **0.152** | **0.143** | **0.200** |

表 4.6 使用冲击分布的MedLSVM模型实现在VOC Pascal 2007数据集全部20个类上的测试结果及其与基准模型的比较（续）

| table | dog | horse | motor | person | plant | sheep | sofa | train | tv | avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.173 | 0.105 | **0.390** | **0.365** | 0.297 | 0.124 | **0.148** | 0.221 | 0.218 | 0.386 | **0.237** |
| 0.172 | 0.105 | **0.395** | **0.374** | **0.306** | 0.127 | **0.148** | 0.223 | 0.220 | 0.389 | **0.241** |
| **0.233** | **0.107** | 0.367 | 0.358 | **0.298** | **0.132** | 0.092 | **0.231** | **0.231** | **0.390** | 0.234 |
| **0.228** | **0.107** | 0.372 | 0.363 | 0.305 | **0.130** | 0.092 | **0.234** | **0.232** | **0.396** | 0.238 |

AP1的柱形图比较见图4.20。AP2的柱形图比较见图4.21。

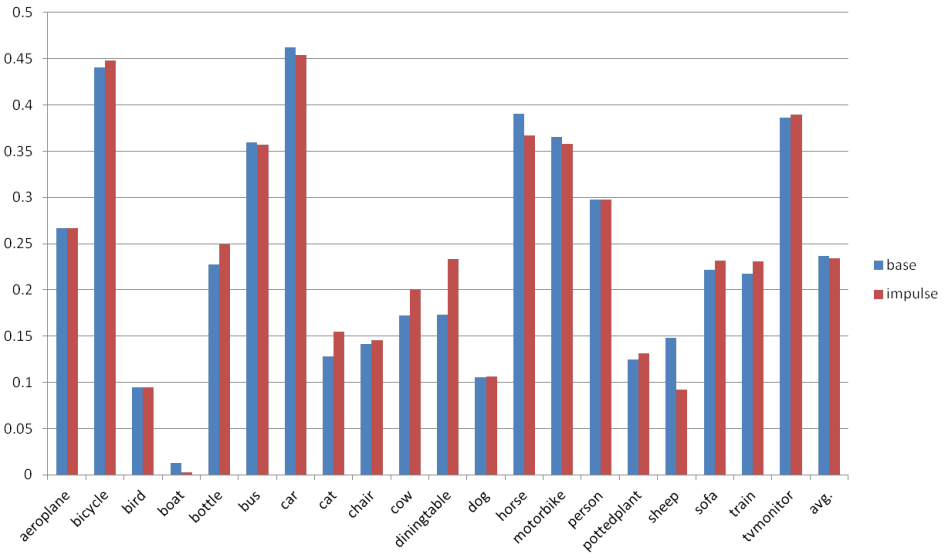


图 4.20　使用冲击分布的MedLSVM模型实现与基准模型在AP1方法下测试性能的比较

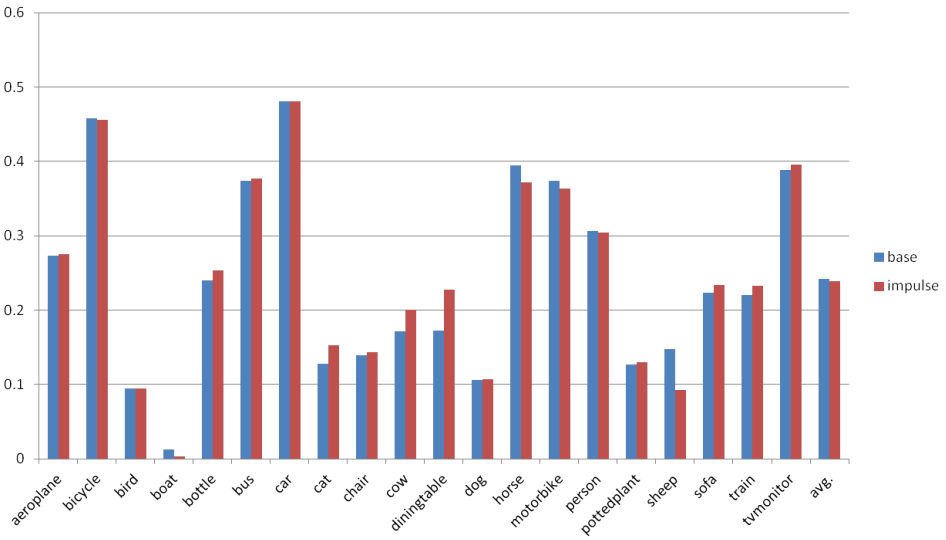


图 4.21　使用冲击分布的MedLSVM模型实现与基准模型在AP2方法下测试性能的比较

图 4.22　VOC Pascal 2007数据集cow类图片示例

可以看出，我们的算法在20个类别中的16个类别上都超过了基准模型（或至少有一项超过），但因为在sheep类别上落后较多而使得20个类上平均得分略微低于基准模型。我们的模型在cow和diningtable这两类上领先较多，cow类的示例图片可见图4.22。我们的模型在sheep这一类上落后较多，sheep类的示例图片可见图4.23。分析sheep类落后较多的原因，可能是因为sheep类的图片中较经常出现多个目标物体（羊经常以羊群出现），而我们的模型会误认为在多只羊上的概率极大点出自同一只羊，从而对这多只羊进行概率平均。如果这些羊的姿势很不同（例如一个向左一个向右）时，概率平均就会带来破坏性的后果。当然，如前所说，在没有办法为每一个视角单独训练一个物体模型的情况下，单视角模型在多视角数据集上的性能优劣比较难以得到直观的解释。

作为例子，我们的模型在car类和horse类上检测结果的ROC曲线如图4.24和图4.25。训练得到的模型的可视化表示如图4.26和图 4.27。注意，因为我们在训练中为了减少模型参数，限制了我们的模型必须左右对称。所以马的模型类似于左右对称的双头马。两类的检测结果的示例如图4.28和图4.29。

## 4.4　本章小结

本章中，我们以实际效果探讨了MedLSVM模型中似然函数的选择，为MedLSVM模型在物体检测任务中的应用开发了两种快速实现方法，并通过这两种快速算法将MedLSVM模型用在了完整的物体检测任务中，获得了与当

图 4.23　VOC Pascal 2007数据集sheep类图片示例

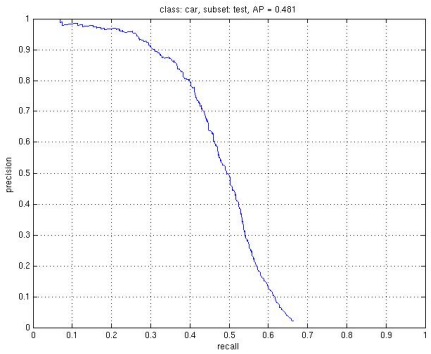


图 4.24　使用冲击分布的MedLSVM模型实现在car类上的测试结果的ROC曲线

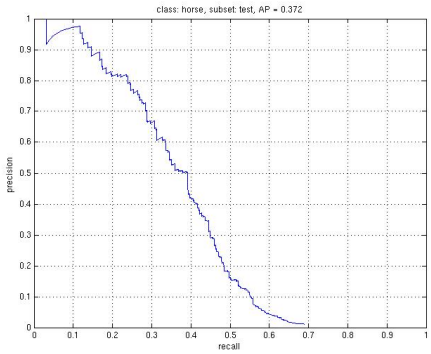


图 4.25　使用冲击分布的MedLSVM模型实现在horse类上的测试结果的ROC曲线

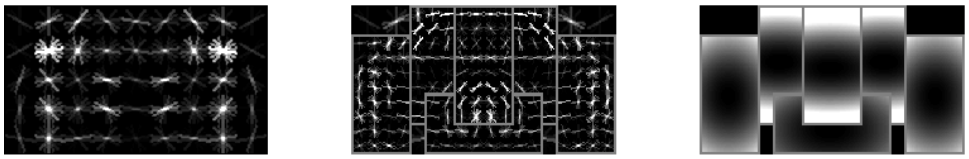

图 4.26　使用冲击分布的MedLSVM模型实现在car类上训练得到的模型的可视化表示

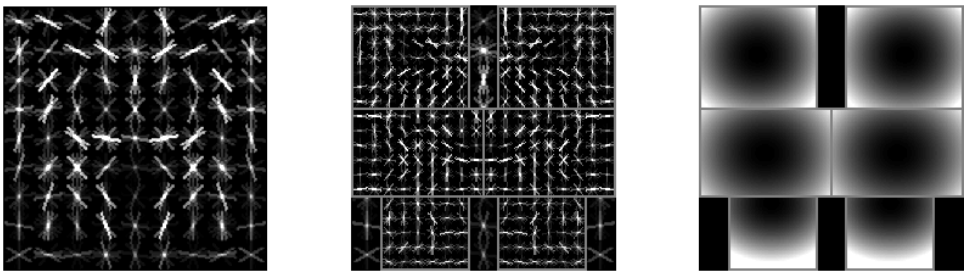

图 4.27　使用冲击分布的MedLSVM模型实现在horse类上训练得到的模型的可视化表示

前最好方法可比拟的实验性能。

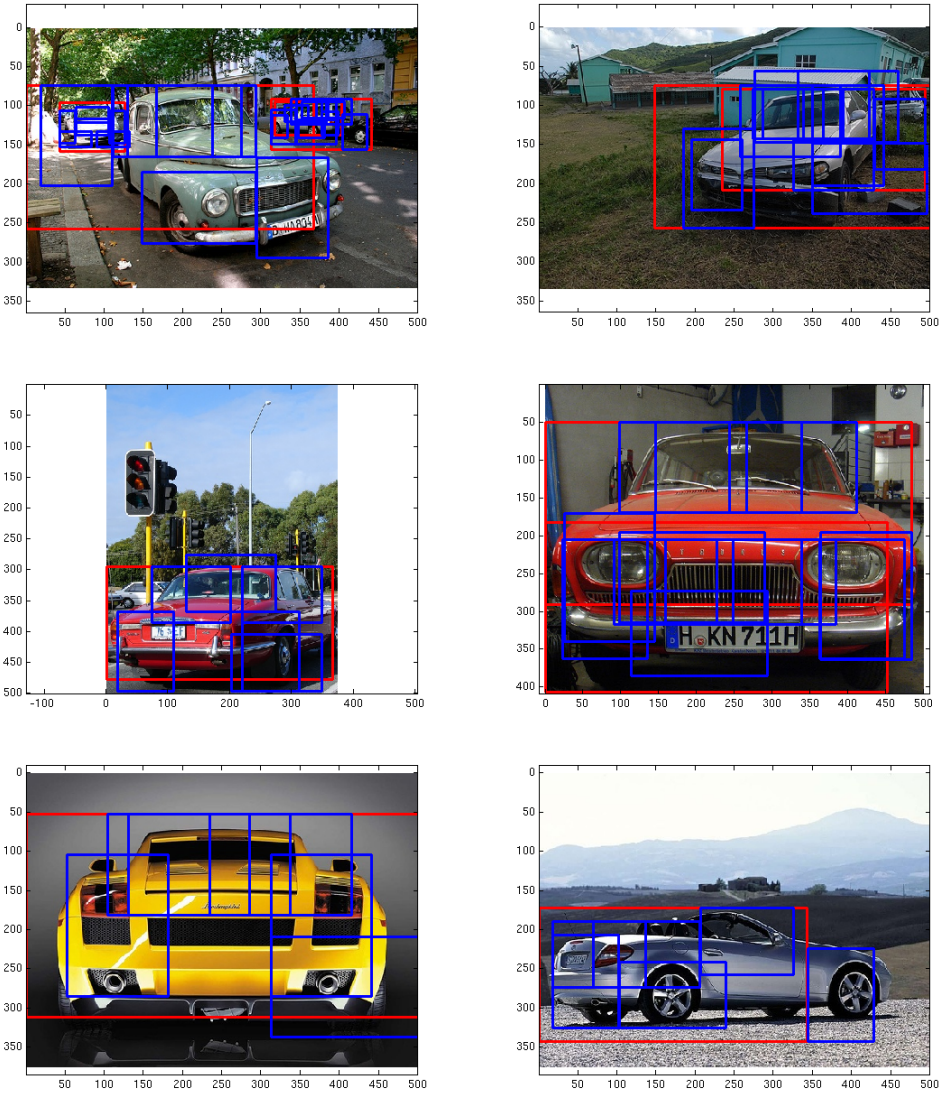

图 4.28　使用冲击分布的MedLSVM模型实现在car类的检测结果。其中红色矩形表示输出的包围框，蓝色矩形表示各个部件的最佳摆放位置。

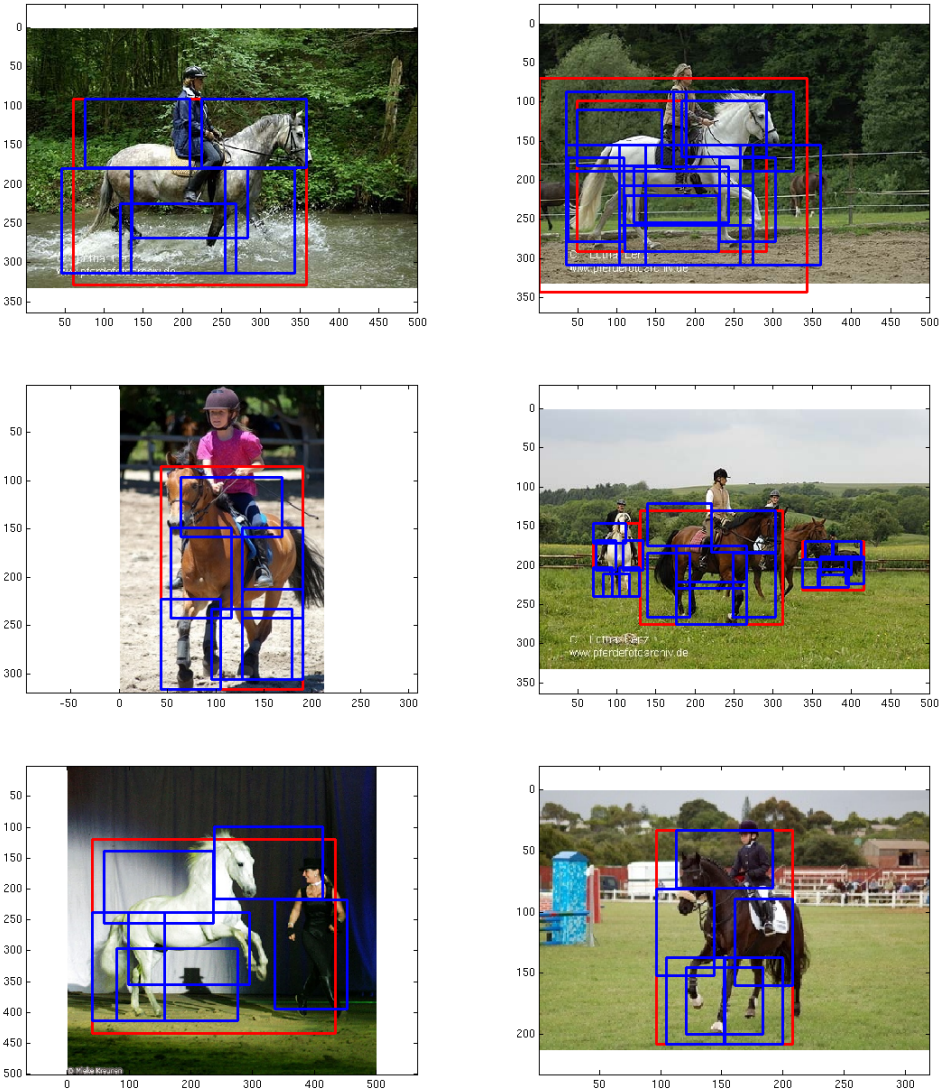

图 4.29　使用冲击分布的MedLSVM模型实现在horse类的检测结果。其中红色矩形表示输出的包围框，蓝色矩形表示各个部件的最佳摆放位置。

# 第 5 章 总结

在本文中，为了处理难度很大的一般物体检测问题，我们借助于基于组件的结构化物体模型，在此基础上提出了MedLSVM分类器模型及其学习算法。它首次利用概率和最大熵的技术来处理分类器设计中的隐变量问题，将MED和LSVM两条重要的最大间隔分类器研究路线融合在了一起。我们尝试了在MedLSVM设计过程中观察变量与隐变量之间的概率产生模型的多种选择，以及对模型参数$\eta$的分布的多种选择，尽可能全面地考察了MedLSVM的设计可能选项。我们提出了 MedLSVM的求解技术。我们在完整MedLSVM理论框架的基础上设计出了两种快速实现，以使得它在物体检测这类对速度要求较高的场景下的使用变成可能。我们的实验结果显示我们的方法在当前最具挑战的数据集上能够获得和当前最好方法相当的性能。

在设计MedLSVM的过程中，我们首先详细考察了[4]提出的基于部件的可形变物体检测模型，然后介绍了设计最大间隔分类器模型的几项关键技术。在选择我们要以之为基础的分类器隐变量处理方法时，我们考察了现有的几种号称Latent SVM的分类器设计思路，深入挖掘出了它们的本质不同，明确了我们的模型应该具有的行为和设计目标。接着我们从一种LSVM出发，为隐变量引入了概率分布，并从此开始一步步完善我们的模型。我们首先发现了这样简单地为隐变量加入概率分布后带来的两点缺点或者说困难，然后发现我们可以通过为观测数据和隐变量建立概率产生模型来同时克服这两项困难。在为观测数据和隐变量建立概率产生模型时，我们面临了多种模型设计的可能选项。我们以经验结果来实证地做出决定，而不是纯粹从数学性质做考察。在选择似然函数的过程中，我们还提出了一种简化归一化因子的数学技巧。之后，由于仍存在的概率计算困难，我们又为分类器参数引入了概率分布，从而得到了完整的MedLSVM概率框架。为求解这个完整的MedLSVM模型，我们详细介绍了其求解技术，包括变分近似技术、Mean Field独立性假设、坐标下降迭代优化等。特别地，为处理分类器参数概率分布的最优化问题，我们创造性地使用了径向、角度分离的概率分布形式，并指出这实际上非常符合SVM分类面参数的

物理意义。

我们的工作仍需改进的地方，包括：（1）进一步设计快速算法，使得MedLSVM理论模型的更多能力能够被用于物体检测这类实际问题。当前我们还没有充分利用MedLSVM提供的对先验信息的控制能力，没有能够将物体检测问题的先验知识用于模型学习中。（2）扩展我们的概率框架来处理待检测物体多视角的问题，以期能够用一种类似于混合模型的统一概率框架来同时检测坐着的人、站立的人、正面的车辆、侧面的车辆等。（3）将朱军等作者提出的iSVM模型与我们的MedLSVM进行融合，以利用前者强大的处理混合模型（多视角、多版本等）的能力，结合我们的MedLSVM模型良好的隐变量概率框架，完整地解决物体检测任务的几项难点。这项工作包含两个方面，一是将iSVM模型与MedLSVM模型在理论上进行整合，提出一个统一的概率模型来包含两者，使两者都成为这一统一模型的特例；二是针对物体检测问题，找出一种方法，使得这个统一模型的Inference步骤（针对于物体检测是“检测”步骤）能够非常快速地、批量地完成（因为物体检测任务使用滑动窗口，有大量的窗口需要Inference）。我们在第二项工作上已有了一些初步的尝试。它还不具有一个统一的理论框架，而是在实现的层面进行。为使得iSVM能快速地用于检测，我们借用Cascade的思想，在检测时首先将iSVM模型训练出的各子分类器当作一个独立的分类器在图像上进行检测，然后将各子分类器得到的得分较高的若干窗口合起来作为候选集，输入给完整的iSVM 分类器进行重评分，以此作为iSVM分类器在图像上的检测结果。只要控制候选集的大小，iSVM在图像上的运行时间只比普通SVM慢常数倍。但由于SVM本身并不适合作为基于部件的可形变物体模型的分类器（没有考虑隐变量），单纯地使用iSVM来训练基于部件的可形变物体模型也遗传性地得不到很好的实验结果。结合iSVM和MedLSVM的努力能有许多工作要做。

# 参考文献


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

[2] Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1. Ieee, 2005. 886–893.

[3] Lowe D. Object recognition from local scale-invariant features. Proceedings of Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2. Ieee, 1999. 1150–1157.

[4] Felzenszwalb P, Girshick R, McAllester D, et al. Object detection with discriminatively trained part-based models. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010, 32(9):1627–1645.

[5] Zhu L, Lin C, Huang H, et al. Unsupervised structure learning: Hierarchical recursive composition, suspicious coincidence and competitive exclusion. 2008..

[6] Wu Y, Si Z, Gong H, et al. Learning active basis model for object detection and recognition. International journal of computer vision, 2010, 90(2):198–235.

[7] Zhu L, Chen Y, Yuille A, et al. Latent hierarchical structural learning for object detection. Proceedings of Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010. 1062–1069.

[8] Sadeghi M, Farhadi A. Recognition using visual phrases. Proceedings of Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011. 1745–1752.

[9] Wang X, Han T, Yan S. An HOG-LBP human detector with partial occlusion handling. Proceedings of Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009. 32–39.

[10] Song Z, Chen Q, Huang Z, et al. Contextualizing object detection and classification. Proceedings of Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011. 1585–1592.

[11] Vapnik V. The nature of statistical learning theory. Springer-Verlag New York Inc, 2000.

[12] Schölkopf B, Smola A. Learning with kernels: Support vector machines, regularization, optimization, and beyond. the MIT Press, 2002.

[13] Crammer K, Singer Y. On the algorithmic implementation of multiclass kernel-based vector machines. The Journal of Machine Learning Research, 2002, 2:265–292.

[14] Morik K, Brockhausen P, Joachims T. Combining Statistical Learning with a Knowledge-Based Approach-A Case Study in Intensive Care Monitoring. Proceedings of Proceedings of the Sixteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc., 1999. 268–277.

[15] Rosasco L, Vito E, Caponnetto A, et al. Are loss functions all the same? Neural Computation, 2004, 16(5):1063–1076.

[16] Tsochantaridis I, Hofmann T, Joachims T, et al. Support vector machine learning for interdependent and structured output spaces. Proceedings of Proceedings of the twenty-first international conference on Machine learning. ACM, 2004. 104.

[17] Yu C, Joachims T. Learning structural SVMs with latent variables. Proceedings of Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009. 1169–1176.

[18] Jaakkola T, Meila M, Jebara T. Maximum entropy discrimination. 1999..

[19] Zhu J, Chen N, Xing E. Infinite SVM: a Dirichlet Process Mixture of Large-margin Kernel Machines. Proceedings of Proceedings of the 28th International Conference on Machine Learning, 2011.

[20] Zhu J, Chen N, Xing E. Infinite Latent SVM for Classification and Multi-task Learning. Advances in Neural Information Processing Systems, 25.

[21] Zhu J, Ahmed A, Xing E. MedLDA: maximum margin supervised topic models for regression and classification. Proceedings of Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009. 1257–1264.

[22] Ferguson T. A Bayesian analysis of some nonparametric problems. The annals of statistics, 1973. 209–230.

[23] Collobert R, Bengio S, Bengio Y. A parallel mixture of SVMs for very large scale problems. Neural computation, 2002, 14(5):1105–1114.

[24] Fu Z, Robles-Kelly A, Zhou J. Mixing linear SVMs for nonlinear classification. Neural Networks, IEEE Transactions on, 2010, 21(12):1963–1975.

[25] Bishop C, ligne) S S. Pattern recognition and machine learning, volume 4. springer New York, 2006.

[26] Press W, Teukolsky S, Vetterling W, et al. Golden section search in one dimension. Numerical Recipes in Fortran, 1992. 390–395.

[27] Rosen J. The gradient projection method for nonlinear programming. Part I. Linear constraints. Journal of the Society for Industrial and Applied Mathematics, 1960, 8(1):181–217.

# 致　谢

衷心感谢导师张钹院士和李建民副教授对本人的精心指导。他们的言传身教将使我终生受益。本工作从朱军副教授的相关工作中获益良多，在此对朱军老师表示衷心感谢。本课题承蒙国家自然科学基金资助，特此致谢。

# 声　明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签　名：_____ 日　期：_____

# 个人简历、在学期间发表的学术论文与研究成果

## 个人简历

1987年06月14日出生于河北省保定市蠡县。

2006年 9 月考入清华大学精密仪器与机械学系精密仪器与机械学专业，2007年9月转入清华大学计算机科学与技术系计算机科学与技术专业，2010 年 7 月本科毕业并获得工学学士学位。

2010年 9 月免试进入清华大学计算机科学与技术系攻读工学硕士学位至今。

## 发表的学术论文

[1] Wang P, Li J M, Zhang B. A Real World Detection System: Combining Color, Shape and Appearance to Enable Real-time Road Sign Detection. International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP), 2012.