

## 基于支持向量机的目标检测算法综述

郭明玮, 赵宇宙, 项俊平, 张陈斌, 陈宗海

(中国科学技术大学 自动化系, 合肥 230027)

**摘要:** 目标检测的目的在于从静态图片或视频中检测并定位设定种类的目标物体, 已有研究大都将目标检测问题简化为一个二分类问题. 鉴于支持向量机在模式识别领域尤其是解决二分类问题中所表现出来的优越性, 如何将其应用于目标检测已成为当今计算机视觉领域关注的重点. 对此, 从支持向量机原理、目标特征模型构建、学习训练和目标检测框确定等角度, 综述了基于支持向量机的目标检测算法的研究现状, 并就进一步的发展进行了展望.

**关键词:** 支持向量机; 目标检测; 局部可变模型; 目标特征模型; 学习训练

**中图分类号:** TP273

**文献标志码:** A

## Review of object detection methods based on SVM

GUO Ming-wei, ZHAO Yu-zhou, XIANG Jun-ping, ZHANG Chen-bin, CHEN Zong-hai

(Department of Automation, University of Science and Technology of China, Hefei 230027, China. Correspondent: CHEN Zong-hai, E-mail: chenzh@ustc.edu.cn)

**Abstract:** The purpose of object detection is to detect and locate the object with a certain class from the static image or videos, and many studies simplify the object detection as a binary classification problem. For the reason that the support vector machine(SVM) can solve the pattern recognition problem well, especially the binary classification problem, how to use the SVM in computer vision becomes a hot point of many researchers. The status of object detection methods based on SVM is reviewed by introducing the concept and theory of SVM, the building of object feature model, training process and location of detection box. Finally, the future work of object detection methods based on SVM is discussed.

**Key words:** SVM; object detection; deformable part models; object feature model; training

## 0 引言

目标检测是计算机视觉中最基础的任务之一, 其研究主要包括基于视频图像的目标检测和基于静态图片的目标检测. 本文主要讨论基于静态图片的目标检测, 即在静态图片中检测并定位所设定种类的目标物体. 基于静态图片的目标检测的应用十分广泛, 其在监控系统、图片搜索、身份认证、自动导航等多个领域扮演着重要的角色. 支持向量机在语音识别、文字识别、3D 识别<sup>[1]</sup>等领域表现出一定的优越性, 其在目标检测中的应用具有重要的理论与实际研究价值, 近年来引起了人们的广泛兴趣.

基于静态图片的目标检测的难点主要在于图片中的目标会因光照、视角以及目标内部等变化而产生变化. 针对目标检测的任务及其难点, 国内外学者进行了很多尝试. Itti 等<sup>[2]</sup>构建了选择性注意机

制, 用于提取图片中的显著性区域; 而 Borji 等<sup>[3]</sup>和 Navalpakkam 等<sup>[4]</sup>基于选择性注意机制构建了目标检测算法. Viola 等<sup>[5-7]</sup>提出了基于 Haar-Like 特征的目标检测算法, 该算法可以较好地解决人脸检测问题; 而 Lienhart 等<sup>[8]</sup>则扩展了 Haar-Like 特征, 改进了目标检测算法, 获得了较好的检测效果. Leibe 等<sup>[9]</sup>基于图像分类和分割算法构建了一种鲁棒的目标检测算法, 该算法可以较好地解决行人及车辆的检测问题. 上述目标检测算法在解决静态图片的目标检测问题时存在一些缺陷: 1) 特征模型仅在解决某一类目标检测时效果较好, 例如对行人检测效果较好, 但对车辆检测效果很差; 2) 检测准确率较低, 很多目标检测算法在针对设定的图像测试集时具有较好的效果, 但对于实际图片, 如网络图片等, 则很难保证较高的检测准确率; 3) 训练图像集大小难以确定, 训练图像集中的训练图片过少, 会导致构建的特征模型不具有代表性,

收稿日期: 2013-07-02; 修回日期: 2013-11-05.

基金项目: 国家自然科学基金项目(61005091, 61375079).

作者简介: 郭明玮(1986—), 男, 博士生, 从事计算机视觉与模式识别的研究; 陈宗海(1963—), 男, 教授, 博士生导师, 从事复杂系统的建模仿真与控制、机器人与智能系统、汽车新能源技术等研究.

而如果过多,则目标特征模型更为复杂,可能会导致过学习情况的发生;4) 目标检测时间较长,上述目标检测算法检测目标的时间大多在 3 s 以上,使其在很多实际场景难以应用。

随着 Vapnik<sup>[10-11]</sup>以及 Trevor 等<sup>[12]</sup>提出的统计学习理论在模式识别领域的广泛应用,支持向量机(SVM)<sup>[13-17]</sup>在解决小样本、非线性及高维模式识别中表现出很多特有的优势<sup>[18-23]</sup>,如:结构简单、全局最优、推广能力强、学习和预测时间短,目前已广泛应用于文字识别、语音识别和生物特征识别等实际问题中。由此产生了基于 SVM 的目标检测方法,它将目标检测任务简化为一个二分类问题:通过训练学习得到的目标特征模型,判断待检测图片中滑动窗口内是否存在该类目标,如果存在,则定位该目标位置。Dalal 等<sup>[24]</sup>和 Felzenszwalb 等<sup>[25-30]</sup>以 SVM 作为分类器,分别构建了目标检测算法,其目标检测准确率和检测时间等都达到了较好的指标,是基于 SVM 的目标检测算法中的典范。

本文重点综述基于 SVM 的目标检测算法。首先介绍了支持向量机的相关理论基础;然后按照构建目标检测算法的顺序:目标特征模型构建→学习训练→目标检测框确定分别进行了介绍;进而介绍了目标检测常用的数据库以及评价目标检测算法的常用指标;最后对这类目标检测算法存在的问题进行分析,并对进一步的工作进行了展望。

## 1 支持向量机

### 1.1 统计学习理论

SVM 是建立在统计学习理论中 VC 维理论和结构风险最小原理基础上的,根据有限的样本信息在模型的复杂性和学习能力之间寻求最佳折中,以获得最好的推广能力。经验风险  $R_{\text{emp}}(w)$  和实际风险  $R(w)$  之间至少以  $1 - \eta$  的概率满足如下关系:

$$R(w) \leq R_{\text{emp}} + \sqrt{\left| \frac{h(\ln(2n/h)) + 1 - \ln(\eta/4)}{n} \right|}. \quad (1)$$

其中:  $h$  为函数集的 VC 维,  $n$  为样本数,  $\eta$  为满足  $0 \leq \eta \leq 1$  的参数。式(1)可简单地表示为

$$R(w) \leq R_{\text{emp}} + \Phi(h/n), \quad (2)$$

其中  $\Phi(h/n)$  即为置信风险。置信风险反映了实际风险和经验风险的上确界,同时反映了结构复杂所带来的风险,它与学习机器的 VC 维  $h$  及训练样本数  $n$  有关。置信风险表明,在训练样本有限的情况下,学习机器的 VC 维越高,置信风险越大,从而导致实际风险与经验风险之间可能的差别越大,其推广能力越差。

### 1.2 SVM 理论

SVM 的核心思想是寻找一个满足分类要求的最优分类超平面<sup>[19-22]</sup>,该超平面在保证分类精度的同时,应使分类间隔最大化,如图 1 所示。

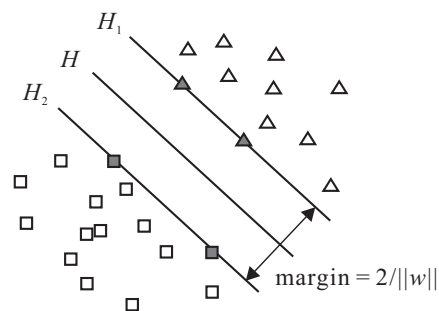


图 1 支持向量机原理示意

目标检测问题可以简化为一个二分类问题,这里主要讨论支持向量机如何解决二分类问题。给定训练样本集  $(x_i, y_i), i = 1, 2, \dots, l, x \in R^n, y \in \{\pm 1\}$ , 超平面为  $(w \cdot x + b = 0)$ 。为使分类面对所有样本正确分类并具备分类间隔,要求其满足

$$y_i[(w \cdot x_i) + b] \geq 1, i = 1, 2, \dots, l. \quad (3)$$

由此可得分类间隔为  $2/\|w\|$  (图 1 中 margin 的值),于是构造最优超平面的问题便转化为如下带约束的最小值问题:

$$\min \Phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w' \cdot w), \quad (4)$$

约束条件为式(1)。引入 Lagrange 函数

$$L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l a_i y_i (w \cdot x_i + b) + \sum_{i=1}^l a_i, \quad (5)$$

其中  $a_i > 0$  为 Lagrange 系数。约束最优化问题由 Lagrange 函数的鞍点决定,并且最优化问题的解在鞍点处满足对  $w$  和  $b$  的偏导为 0。将该 QP (quadratic programming) 问题转化为如下相应的对偶问题 ( $a = (a_1, a_2, \dots, a_l)$ ):

$$\begin{aligned} \max Q(a) &= \sum_{j=1}^l a_j - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j (x_i \cdot x_j); \\ \text{s.t. } \sum_{j=1}^l a_j y_j &= 0, \\ a_j &\geq 0, j = 1, 2, \dots, l. \end{aligned} \quad (6)$$

经计算,最优权值向量  $w^*$  和最优偏置  $b^*$  分别为

$$w^* = \sum_{j=1}^l a_j^* y_j x_j, \quad (7)$$

$$b^* = y_i - \sum_{j=1}^l y_j a_j^* (x_j \cdot x_i), \quad (8)$$

其中  $j \in \{j | a_j^* > 0\}$ 。因此得到最优分类面  $(w^* \cdot x) + b^* = 0$ , 而最优分类函数为

$$f(x) = \text{sgn}\{(w^* \cdot x) + b^*\} = \text{sgn}\left\{\left(\sum_{j=1}^l a_j^* y_j (x_j \cdot x_i)\right) + b^*\right\}, \quad (9)$$

其中  $x \in R^n$ .

### 1.3 SVM 解决分类问题的优势

SVM 在解决分类问题时具有一定的优势<sup>[18-23]</sup>, 可以解决目标检测算法中出现的一些问题, 主要表现在以下几个方面.

1) 系统结构简单. 分类器仅由支持向量决定, 即分类器的复杂程度与训练集样本的数量无关, 仅与支持向量的数量有关, 而实际情景中支持向量的数量远小于训练样本的数量. 在目标检测构建目标特征模型时, 训练样本集的样本数量决定分类器的效果, 过多或过少的样本数量都会导致分类效果不佳, 而 SVM 仅由支持向量的数量决定分类器的复杂程度. 系统的结构简单, 保证了分类器具有较好的鲁棒性. 鲁棒性体现在: ① 增、删非支持向量样本对模型没有影响; ② 支持向量样本集具有一定的鲁棒性; ③ SVM 对核函数的选取不敏感.

2) 全局最优. SVM 是通过求解最优超平面进行学习的, 在高维特征空间中的超平面对应于原始模式空间中的非线性分类面, 而寻找最优超平面的问题是利用 Lagrange 优化方法转化为二次规划问题, 能够保证 SVM 算法得到全局最优解. 全局最优可以保证训练学习得到的目标检测分类器在分类时具有较高的准确率.

3) 推广能力强. SVM 基于统计学习理论, 采用结构风险最小化原则, 能在经验风险与模型复杂度之间进行合理的折中, 尽量提高学习机的推广性能, 即使由有限训练样本得到的决策规则, 对于独立的测试集仍能得到较小的误差. SVM 的推广能力强, 使得基于 SVM 的目标检测算法在特征模型构建时, 不仅仅对某一类目标有效, 其适用性较广.

4) 学习和预测时间短. 文献[19]对比了支持向量机和神经网络的训练时间和预测时间, 神经网络用时为支持向量机的 3~4 倍, 且当训练样本数目增加时, 支持向量机在算法复杂度方面的优势将更加明显. SVM 的学习和预测时间较短, 可以使目标检测算法更加高效, 其应用也将具有更广阔的前景.

## 2 目标特征模型

本文讨论的目标检测方法采用基于支持向量机的分类方法, 而如何提取目标的特征模型作为分类处理的数据已成为该类目标检测算法所讨论的重点. 由统计学习理论可知, 在训练样本有限的情况下, 目标的特征模型越复杂, 导致训练模型 VC 维越高, 此时

虽然可以记住训练样本的每一个样本, 在样本集中能轻易达到 100% 的正确率, 但在真实分类时却效果很差. 其原因由式 (2) 可以看出: 当 VC 维越高时, 置信风险越大, 导致实际风险与经验风险之间可能的差别越大, 其推广能力越差, 这便是过学习的情况. 另外, 显而易见的是, 如果特征模型过于简单, 则很难较好地表示目标物体, 检测的正确率也会较低. 如何选择合适的特征模型来描述目标物体是目标检测算法的关键步骤之一. 目前特征模型的构建主要有 3 种思路: 整体特征模型、局部特征模型和上下文特征模型. 而目标检测算法的发展趋势则是沿着整体特征模型到局部特征模型再到上下文特征模型. 下面分别介绍这 3 种特征模型.

### 2.1 整体特征模型

整体特征模型是指对目标的整体进行建模, 并未单独考虑目标的局部特征<sup>[31-35]</sup>. 整体特征模型以 Dalal 等<sup>[24]</sup>提出的以梯度方向直方图 (HOG) 特征为基础的行人检测算法为代表, 采用 HOG 特征来描述目标物体, HOG 特征表示了梯度的统计信息, 而梯度主要存在于边缘位置. 文献[24]所提出的方法构建了目标特征模型的主要过程, 即: 1) 输入目标图像, 例如一个  $40 \times 40$  的训练图像; 2) 标准化 Gamma 空间和颜色空间; 3) 梯度计算, 包括梯度大小计算公式

$$R(X, Y) = ((I(X+1, Y) - I(X-1, Y))^2 + (I(X, Y-1) - I(X, Y+1))^2)^{1/2}, \quad (10)$$

梯度方向计算公式

$$\text{Ang}(X, Y) = \arccos(I(X+1, Y) - I(X-1, Y) / R), \quad (11)$$

其中  $I(X, Y)$  表示在像素点  $(X, Y)$  处的像素值; 4) 将梯度投影到单元格 (例如为  $8 \times 8$  的像素) 的梯度方向 (直方图取 9 个方向); 5) 将所有单元格在块 (例如为  $2 \times 2$  单元格) 上进行归一化; 6) 收集检测空间所有块的 HOG 特征, 最后得到一个维数为  $4 \times 4 \times 2 \times 2 \times 9 = 576$  维的特征向量.

### 2.2 局部特征模型

局部特征模型<sup>[36-37]</sup>不仅对目标的局部特征建模, 而且也考虑了目标的整体特征模型. 其中以 Felzenszwalb 等<sup>[25-30]</sup>所做的工作最具代表性, 他们在文献[25-30]中提出了局部可变特征模型, 以 HOG 特征为基础构建了目标的局部可变特征模型. 该方法对滑动窗口内的图像构建图像金字塔, 对金字塔每层生成 HOG 特征. 目标的特征模型包括一个根滤波器和若干局部滤波器. 其中: 根滤波器获取全局特征, 但特

征相对粗略,其基本覆盖了目标的整体;局部滤波器置于HOG金子塔的下几层,覆盖了目标物体的局部位置.用 $F_0$ 表示根滤波器,用 $(P_1, P_2, \dots, P_n)$ 表示局部模型.其中: $P_i = (F_i, v_i, d_i)$ ,  $F_i$ 为第 $i$ 个局部滤波器,  $v_i$ 为位置坐标,  $d_i$ 为局部形变参数.对于滤波器(包括根滤波器和局部滤波器)在检测窗口内的得分可由下式计算得到:

$$F \cdot \phi(H, p, w, h). \quad (12)$$

其中: $H$ 为HOG特征金字塔,  $p$ 为金子塔中一层的像素块,而 $w$ 和 $h$ 为检测窗口的宽和高.

另外, Ott等<sup>[38]</sup>提出了一种允许共享局部模型的局部特征模型思路,在PSACAL数据库中取得了很好的检测效果; Pedersoli等<sup>[39]</sup>提高了局部特征模型完成目标检测的速度; Hyunggi等<sup>[40]</sup>将局部特征模型应用于3D自行车跟踪,得到了不错的效果; Parikh等<sup>[41]</sup>则将局部特征模型较好地应用于行人检测中.

### 2.3 上下文特征模型

上下文特征模型<sup>[42]</sup>考虑了目标上下文的相关性,包括:局部区域上下文、目标上下文和场景上下文.在解决目标检测问题时,前两类特征模型仅考虑了目标的特征,而忽略了上下文特征对目标检测的影响,存在一定的局限性.为了获得更好的目标检测效果,合理地利用上下文特征是重要途径之一.

1) 局部区域上下文.待检测目标图像中包含很多局部区域,每个局部区域与其周围都有一种关系,描述这种关系的信息称为局部区域上下文特征,包括局部区域的邻域上下文和局部区域之间的几何上下文(位置关系).这类特征主要考虑的是目标本身的特征,前面介绍的局部可变模型即是利用这类上下文特征的典型例子.

2) 目标上下文.待检测目标与周围目标都有一定的联系,例如这些目标是否出现,以及这些目标与待识别目标的位置和尺度关系,这些信息称为目标上下文.在文献[43]中, Sadeghi等提出了基于视觉短语的目标识别方法,在构建特征模型时就利用了目标的上下文关系,例如人与马的关系等.

3) 场景上下文.目标所处的场景对于目标的检测与识别有很大的帮助,该场景信息称为场景上下文.例如在网球场上会出现网球、网球拍等目标.在文献[44]中, Murphy等构建目标模型时,强调了目标与所在场景之间的关系,在行人检测等问题上得到了较好的效果.

## 3 学习训练

在目标的特征模型结构确定后,需要根据训练样本集进行学习训练以获得该目标特征模型的相关参

数,并作为分类的重要依据.本文所介绍方法的特点在于采用SVM作为训练学习的数学工具.

### 3.1 SVM训练过程

文献[24]所述方法采用了线性支持向量机,利用SVMLight<sup>[23]</sup>进行训练学习;文献[45-47]也以线性支持向量机作为学习训练的工具构建目标检测算法.训练学习的基本思想是:已知目标的特征模型结构便可从训练集中获取到特征向量,然后根据训练图片是否为正样本将特征向量表示为 $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ .其中: $x_i$ 表示按特征模型获取的特征向量;  $y_i$ 表示该图片是否为正样本,如果该训练图片为正样本则 $y_i$ 为1,若为负样本则 $y_i$ 为-1;  $n$ 表示训练样本的个数.对于二分类问题,容易得到损失函数<sup>[15]</sup>

$$L(y, f(x, w)) = \begin{cases} 0, & y = f(x, w); \\ 1, & y \neq f(x, w). \end{cases} \quad (13)$$

其中: $f(x, w)$ 为预测函数集,  $w$ 为函数的广义参数.基于支持向量机的训练学习目标在于获得使结构风险最优化的分类面,训练过程如式(6)~(9)所示.

### 3.2 隐性支持向量机训练过程

在文献[25-26]中, Felzenszwalb等提出的局部可变模型采用了基于隐性支持向量机的训练过程,其将目标位置信息作为隐性向量,获得了较好的检测结果.其分类器为

$$f_\beta = \max \beta \cdot \Phi(x, z), \quad z \in Z(x). \quad (14)$$

其中

$$\beta = (F_0, \dots, F_n, a_1, b_1, \dots, a_n, b_n), \quad (15)$$

$$\Phi(H, z) = (\phi(H, p_0), \phi(H, p_1), \dots, \phi(H, p_n)). \quad (16)$$

式(14)中的 $z$ 即为隐性变量,式(15)和(16)中出现的符号同第1节和第2节中符号定义.代价函数可以写为

$$L_D(\beta) = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i f_\beta(x_i)). \quad (17)$$

## 4 目标检测框的确定

目标检测框的大小和位置的确定是目标检测的最后步骤.滑动窗口法是目前研究的一个趋势,其思想是将检测窗口置于不同尺度的图像中的每个位置,然后判断窗口覆盖的区域能否通过检测,这种穷搜的方式可以保证不漏掉任何一个可能的位置.除了滑动窗口法,也有研究针对目标检测框进行改进的算法.在文献[48]中,杨杨等对可变局部模型进行了改进,针对滑动窗口目标检测方法需要穷搜、检测速度相对较慢的问题,提出了一种基于候选点检测的可变局部模型目标检测算法.

## 5 数据库和评估参数

### 5.1 数据库

目前, 目标检测研究常用的数据库有PASCAL VOC数据库<sup>[49-52]</sup>、UIUC汽车数据库、LabelMe数据库和INRIA行人数据库. 其中PASCAL VOC数据库(图2)是近几年目标检测算法最权威的数据库之一. 而INRIA行人数据库(图3)则是目标应用较多的行人数据库, 其主要应用于直立行人目标检测的研究.



图2 PASCAL 数据库图片



图3 INRIA 数据库示例图片

### 5.2 评估参数

1) 准确性. 准确性是指目标检测算法对于不同种类目标的检测准确率, 是考察目标检测算法优劣的一项重要指标. 为了衡量目标检测算法的准确性, PASCAL VOC采用平均精度(AP)进行评估. 目标检测算法AP通过以下步骤计算得到: ① 计算随着精度单调递减的精度/召回率曲线; ② 对精度/召回率曲线下方的区域进行数值积分, 得到AP值.

表1列出了在PASCAL VOC 2008数据集<sup>[49]</sup>中几种目标检测算法的结果. 其中: DMP<sup>[25]</sup>、UCT<sup>[26]</sup>、MPI<sup>[60]</sup>和LPC<sup>[58]</sup>均是基于SVM的目标检测算法, 而另外几种则未采用SVM作为分类器, 如基于高斯过程的分类器<sup>[59]</sup>. 从结果可以看出, 基于SVM的目标检测算法在平均精度这项重要指标上具有较大的优势. 其中DMP和UCT的效果相对较好, 二者均为基于Felzenszwalb等提出的局部可变模型构建的目标检测算法. 近几年在PASCAL VOC目标检测竞赛中效果较好的算法大都采用了SVM作为分类器, 其

表1 目标检测算法平均精度对比

类别	DMP	UCT	CD	Jena	LPC	MPI	XRCE
Aero	0.351	0.326	0.252	0.048	0.365	0.259	0.264
Bike	0.402	0.420	0.146	0.014	0.343	0.080	0.105
Bird	0.117	0.113	0.098	0.003	0.107	0.101	0.014
Boat	0.114	0.110	0.105	0.002	0.114	0.056	0.045
Bottle	0.284	0.282	0.063	0.001	0.221	0.001	0.000
Bus	0.251	0.232	0.232	0.010	0.238	0.113	0.108
Car	0.334	0.320	0.176	0.013	0.366	0.106	0.040
Cat	0.188	0.179	0.090	—	0.166	0.213	0.076
Chair	0.166	0.146	0.096	0.001	0.111	0.003	0.020
Cow	0.114	0.111	0.100	0.047	0.177	0.045	0.018
Table	0.087	0.066	0.130	0.004	0.151	0.101	0.045
Dog	0.078	0.102	0.055	0.019	0.090	0.149	0.105
Horse	0.347	0.327	0.140	0.003	0.361	0.166	0.118
Mbik	0.395	0.386	0.241	0.031	0.403	0.200	0.136
Pers	0.431	0.420	0.112	0.020	0.197	0.025	0.090
Plant	0.117	0.126	0.030	0.003	0.115	0.002	0.015
Sheep	0.181	0.161	0.028	0.004	0.194	0.093	0.061
Sofa	0.166	0.136	0.030	0.022	0.173	0.123	0.018
Train	0.256	0.244	0.282	0.064	0.296	0.236	0.073
Tv	0.347	0.371	0.146	0.137	0.340	0.015	0.068

相关资料见文献[49-52].

2) 鲁棒性. 目标检测算法的鲁棒性主要体现在该算法所采用分类器的鲁棒性, 鲁棒性通常指对假设较小的偏差不敏感<sup>[53]</sup>. 分类器的鲁棒性主要受两个方面的影响: ① 训练样本集的改变; ② 核函数的选取. 而SVM对于增、减非支持向量样本和核函数的选取不敏感, 具有较好的鲁棒性.

3) 实时性. 实时性主要考察目标检测算法从一张图片中检测出目标所需要的时间长短, 目标检测算法的实时性决定了该目标检测算法的应用前景. 目前基于SVM的目标检测算法, 由于采用滑窗的搜索策略, 其实时性仍较差, 检测目标的时间大约为每张图片3 s.

## 6 存在问题与研究展望

基于SVM的目标检测算法在解决目标检测问题上展现出许多优势, 但也存在很多问题, 主要体现在以下几个方面.

1) 检测准确率. 基于SVM的目标检测算法在检测准确率上达到了较高的标准, 但在面对一些实际场景的图片时, 会出现误检和漏检的情况(包括检测框重合), 即目标检测高准确率还不具有普遍适用性.

2) 目标检测时间. 基于SVM的目标检测算法完成目标检测时间较长, 很难满足实时性的要求. 如果能够减少这类算法目标检测的时间, 则其在智能导航、视频监控等领域将有更多的应用前景.



3) 目标特征表达. 该算法在目标特征表达上经历了由整体特征模型到局部特征模型, 再到上下文特征模型的发展<sup>[54-55]</sup>. 目前这类算法在特征表达上所用到的上下文特征还较为有限, 而人类的视觉系统在完成目标检测时, 上下文特征扮演着重要的角色.

4) SVM 的局限性. SVM 分类算法对大规模训练样本难以实施, 由于 SVM 是借助于二次规划(QP)求解支持向量, 而求解二次规划将涉及  $m$  阶矩阵的计算( $m$  为样本的个数), 当  $m$  数目很大时该矩阵的存储和计算将耗费大量的机器内存和运算时间<sup>[56]</sup>. 另外, SVM 在解决多分类问题时存在一定难度.

5) 智能性. 计算机视觉的最终目标在于使计算机视觉具有和人类视觉系统同样的功能, 目标检测的目的也在于使计算机具有人类视觉系统检测目标的能力, 而目前基于支持向量机的目标检测算法距离这一目标还相差甚远.

针对上述存在的问题, 基于 SVM 的目标检测算法的进一步研究应从以下 3 个方面展开:

1) 构建目标具有唯一性和不变性的特征模型. 目标的特征模型的优劣决定了该目标检测算法性能的高低, 而且目标的特征模型构建也是整个计算机视觉中的基础问题, 其在目标跟踪和行为分析等研究领域都扮演着重要的角色. 在构建目标特征模型的过程中, 需要充分利用目标的上下文特征, 模拟人类视觉系统的特征学习与特征联想的工作机理<sup>[57]</sup>.

2) 快速的目标检测框确定策略. 搜索策略决定了目标检测算法的处理时间, 基于滑窗算法的搜索策略过于耗时, 而人类的视觉系统对于视野中图像信息并不是全部进行检测处理, 人类会首先对其感兴趣区域内的图像进行处理, 这便是人类的选择性注意机制<sup>[2]</sup>. 如何模拟这一机制, 构建快速的目标检测框搜索策略是提高目标检测算法实时性的重要途径之一.

3) 分类器算法的改进. 根据目前分类器研究的新进展, 使目标检测的分类器更加准确、稳定和快速, 并向着多分类器发展, 从而更接近于人类的视觉系统的功能.

## 7 结 论

本文重点介绍了基于 SVM 的目标检测算法, 从算法的出发点、理论基础、工作流程和研究展望等对其进行了详细的介绍. 基于支持向量机的目标检测算法是以支持向量机分类算法为出发点和基础, 目标特征表达为核心而构建的一类目标检测算法, 在目标检测的适用性、准确性和实时性上都具有一定的优势, 但还有很大的上升空间. 从目标特征表达、搜索策略和分类器的改进入手, 提高基于支持向量机的目标检

测算法的智能性是这类算法未来研究的重点.

## 参考文献(References)

- [1] Pontil M, Verri A. Support vector machines for 3-d object recognition[J]. Pattern Analysis Machine Intelligence, 1998, 20(6): 637-646.
- [2] Itti L, Koch C. Computational modelling of visual attention[J]. Nature Reviews Neuroscience, 2001, 2(3): 194-203.
- [3] Borji A, Sihite D N, Itti L. Salient object detection: A benchmark[C]. Proc of European Conf on Computer Vision. Florence: Springer Berlin Heidelberg, 2012: 414-429.
- [4] Navalpakkam V, Itti L. An integrated model of top-down and bottom-up attention for optimal object detection[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. New York: IEEE Press, 2006, 2: 2049-2056.
- [5] Viola P, Jones M. Rapid object detection using a boosted cascade of simple feature[C]. Computer Vision and Pattern Recognition. Kauai: IEEE Press, 2001, 1: 511-518.
- [6] Viola P, Platt J, Zhang C. Multiple instance boosting for object detection[C]. Proc of Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2006: 1417-1424.
- [7] Viola P, Jones M. Robust real-time face detection[J]. Int J of Computer Vision, 2004, 57(2): 137-154.
- [8] Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection[C]. Proc of the 2002 Int Conf on Image Processing. Rochester: IEEE Press, 2002, 1: 900-903.
- [9] Leibe B, Leonardis A, Schiele B. Robust object detection with interleaved categorization and segmentation[J]. Int J of Computer Vision, 2008, 77(1): 259-289.
- [10] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995: 37-69.
- [11] Vapnik V N. An overview of statistical learning theory[J]. IEEE Trans on Neural Networks, 1999, 10(5): 988-999.
- [12] Trevor Hastie, Robert Tibshirani, Jerome Friedman, et al. The elements of statistical learning: Data mining, inference and prediction[J]. The Mathematical Intelligencer, 2005, 27(2): 83-85.
- [13] Cortes C, Vapnik V N. Support vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [14] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-168.
- [15] Chang C C, Lin C J. Libsvm — A library for support vector machines[EB/OL]. [2013-6-25]. [www.csie.ntu.edu.tw/~cjlin/libsvm](http://www.csie.ntu.edu.tw/~cjlin/libsvm).

- [16] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42.  
(Zhang X G. Introduction to statistical learning theory and support vector machines[J]. Acta Automatica Sinica, 2000, 26(1): 32-42.)
- [17] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报, 2011, 40(1): 2-10.  
(Ding S F, Qi B J, Tan H Y. An overview on theory and algorithm of support vector machines[J]. J of University of Electronic Science and Technology of China, 2011, 40(1): 2-10.)
- [18] 许建华, 张学工, 李衍达. 支持向量机的新发展[J]. 控制与决策, 2004, 19(5): 481-493.  
(Xu J H, Zhang X G, Li Y D. Advances in support vector machines[J]. Control and Decision, 2004, 19(5): 481-493.)
- [19] 彭璐. 支持向量机分类算法研究与应用[D]. 长沙: 电器与信息工程学院, 湖南大学, 2007.  
(Peng L. Research on classification algorithm of support vector machine and its application[D]. Changsha: College of Electrical and Information Engineering, Hunan University, 2007.)
- [20] Maji S, Berg A C, Malik J. Classification using intersection kernel support vector machines is efficient[C]. Computer Vision and Pattern Recognition. Anchorage: IEEE Press, 2008: 1-8.
- [21] Scholkopf B, Sung K, Burges C, et al. Comparing support vector machines with Gaussian kernels to radial basis function classifiers[J]. Signal Processing, 1997, 45(11): 2758-2765.
- [22] Weston J, Watkins C. Multiclass support vector machines[R]. London: Department of Computer Science, University of London, 1998.
- [23] Joachims T. Making large-scale svm learning practical[C]. Advances in Kernel Methods — Support Vector Learning. Cambridge: The MIT Press, 1999: 169-184.
- [24] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. San Diego: IEEE Press, 2005, 1: 886-893.
- [25] Felzenszwalb P, Girshick R, McAllester D, et al. Object detection with discriminatively trained part based models[J]. Pattern Analysis and Machine Intelligence, 2009, 32(9): 1627-1645.
- [26] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. Anchorage: IEEE Press, 2008: 1-8.
- [27] Felzenszwalb P, Girshick R, McAllester D. Cascade object detection with deformable part models[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. San Francisco: IEEE Press, 2010: 2241-2248.
- [28] Felzenszwalb P, Huttenlocher D. Pictorial structures for object recognition[J]. Int J of Computer Vision, 2005, 61(1): 55-79.
- [29] Felzenszwalb P, McAllester D. The generalized  $A^*$  architecture[J]. J of Artificial Intelligence Research, 2007, 29: 153-190.
- [30] Ross Girshick, Pedro Felzenszwalb, David McAllester. Object detection with grammar models[C]. The 25th Annual Conf on Neural Information Processing Systems. Granada: Curran Associate Inc, 2011: 145-154.
- [31] Fergus R, Perona P, Zisserman A. Object class recognition by unsupervised scale-invariant learning[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. Madison: IEEE Press, 2003, 2: 264-271.
- [32] Schneiderman H, Kanade T. Object detection using the statistics of parts[J]. Int J of Computer Vision, 2004, 56(3): 151-177.
- [33] Gavrila D M, Giebel J, Munder S. Vision-based pedestrian detection: The protector system[C]. Proc of IEEE Intelligent Vehicles Symposium. Parma: IEEE Press, 2004: 13-18.
- [34] Freeman W T, Roth M. Orientation histograms for hand gesture recognition[C]. Int Workshop on Automatic Face and Gesture-Recognition. Zurich: IEEE Press, 1995: 296-301.
- [35] Mikolajczyk K, Schmid C, Zisserman A. Human detection based on a probabilistic assembly of robust part detectors[C]. Proc of European Conf on Computer Vision. Prague: Springer, 2004, 1: 69-81.
- [36] Yuille A, Hallinan P, Cohen D. Feature extraction from faces using deformable templates[J]. Int J of Computer Vision, 1992, 8(2): 99-111.
- [37] Ramanan D, Sminchisescu C. Training deformable models for localization[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. New York: IEEE Press, 2006, 1: 206-213.
- [38] Ott P, Everingham M. Shared parts for deformable part-based models[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. Providence: IEEE Press, 2011: 1513-1520.
- [39] Pedersoli M, Vedaldi A, González J. A coarse-to-fine approach for fast deformable object detection[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. Providence: IEEE Press, 2011: 1353-1360.
- [40] Hyunggi Cho, Paul E Rybski, Wende Zhang. Vision-based 3D bicycle tracking using deformable part model and interacting multiple model filter[C]. Proc of IEEE Int Conf

- on Robotics and Automation. Shanghai: IEEE Press, 2011: 4391-4398.
- [41] Parikh D, Zitnick C L. Finding the weakest link in person detectors[C]. Proc of IEEE Int Conf on Robotics and Automation. Shanghai: IEEE Press, 2011: 1425-1432.
- [42] Torralba A. Contextual priming for object detection[J]. Int J of Computer Vision, 2003, 53(2): 169-191.
- [43] Sadeghi M, Farhadi A. Recognition using Visual phrases[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. Providence: IEEE Press, 2011: 1745-1752.
- [44] Murphy K, Torralba A, Freeman W. Using the forest to see the trees: A graphical model relating features, objects and scenes[C]. Proc of Advances in Neural Information Processing Systems. Vancouver: MIT Press, 2003: 1499-1507.
- [45] Lowe D. Distinctive image features from scale-invariant keypoints[J]. Int J of Computer Vision, 2004, 60(2): 91-110.
- [46] Vedaldi A, Gulshan V, Varma M, et al. Multiple kernels for object detection[C]. IEEE 12th Int Conf on Computer Vision. Kyoto: IEEE Press, 2009: 606-613.
- [47] Papageorgiou C, Poggio T. A trainable system for object detection[J]. Int J of Computer Vision, 2000, 38(1): 15-33.
- [48] 杨杨, 李善平. 分割位置提示的可变形部件模型快速目标检测[J]. 自动化学报, 2012, 38(4): 540-548.  
(Yang Y, Li S P. Fast object detection with deformable part models and segment locations' hint[J]. Acta Automatica Sinica, 2012, 38(4): 540-548.)
- [49] Everingham M, Van Gool L, Williams C K I, et al. The PASCAL Visual object classes challenge 2009 (VOC 2008) Results[EB/OL]. [2013-6-25]. <http://www.pascal-network.org/challenges/VOC/voc2008/>. 2008.
- [50] Everingham M, Van Gool L, Williams C K I, et al. The PASCAL Visual object classes challenge 2010 (VOC 2009) Results[EB/OL]. [2013-6-25] <http://www.pascal-network.org/challenges/VOC/voc2009/>. 2009.
- [51] Everingham M, Van Gool L, Williams C K I, et al. The PASCAL Visual object classes challenge 2011 (VOC 2011) Results[EB/OL]. [2013-6-25] <http://www.pascal-network.org/challenges/VOC/voc2011/>. 2011.
- [52] Everingham M, Van Gool L, Williams C K I, et al. The PASCAL Visual object classes challenge 2012 (VOC 2011) Results[EB/OL]. [2013-06-25]. <http://www.pascal-network.org/challenges/VOC/voc2012/>. 2011.
- [53] Andrews D F. Robust estimates of location: Survey and advances[M]. Princeton: Princeton University Press, 1972: 31-43.
- [54] Belongie S, Malik J, Puzicha J. Matching shapes[C]. The 8th ICCV. Vancouver: IEEE Press, 2001: 454-461.
- [55] Torralba A, Murphy K P, Freeman W T. Sharing Visual features for multiclass and multiview object detection[J]. Pattern Analysis and Machine Intelligence, 2009, 29(5): 854-869.
- [56] Olivier Chapelle, Patrick Haffner, Vladimir N Vapnik. Support vector machines for histogram-based image classification[J]. IEEE Trans on Neural Networks, 1999, 10(5): 1055-1064.
- [57] Xu X, Wang Z, Chen Z. Visual tracking model based on feature-imagination and its application[C]. Proc of the 2nd Int Conf on Multimedia Information Networking and Security. Nanjing: IEEE Press, 2010: 370-374.
- [58] Harzallah H, Jurie F, Schmid C. Combining efficient object localization and image classification[C]. Int Conf on Computer Vision. Kyoto: IEEE Press, 2009: 237-244.
- [59] Erik Rodner, Joachim Denzler. One-shot learning of object categories using dependent Gaussian processes[C]. Proc of the 32nd Annual Symposium of the German Association for Pattern Recognition. Darmstadt: Springer Berlin Heidelberg, 2010, 6376: 232-241.
- [60] Lampert, Christoph H, Blaschko M B, et al. Beyond sliding windows: Object localization by efficient subwindow search[C]. Proc of IEEE Conf on Computer Vision and Pattern Recognition. Anchorage: IEEE Press, 2008: 1-8.

(责任编辑: 李君玲)