

论文引用格式:

# 车辆检测与识别: 基于 DPM 的整合计算方法

吴俊, 周志敏, 王磊, 赵旭

上海交通大学 自动化系, 上海, 200240

**摘要:** 物体检测与识别是计算机视觉领域的长期研究热点。在特征提取、检测、识别的一般化流程中, 检测与识别通常被作为两个相互独立的模块分别处理。本文提出一种能够同时实现物体检测和类型识别的整合算法, 并给出参数选择的实现方法和细节。该算法基于目前最有效的物体检测方法之一——可变形部件模型(Deformable Part Models, DPM), 通过增加得分偏置, 进而融合多个不同类别模型的检测结果, 最终结果不仅含有检测信息, 还包含子类识别信息, 可同时实现物体检测和类型识别。本文在 SJTUVehicle 数据库中进行了车辆检测和车型识别实验。实验结果显示, 该方法的车辆检测精度与车型识别准确率均可达 95% 以上。本文提出方法具有一般性, 可扩展用于其它物体的检测和子类识别。

**关键词:** 车辆检测; 车型识别; 可变形部件模型; 得分偏置; 整合计算

## An integrated framework for simultaneously vehicle detection and recognition

Wu Jun, Zhou Zhimin, Wang Lei, Zhao Xu

Department of Automation, Shanghai Jiao Tong University, Shanghai, 200240, China

**Abstract:** Object detection and recognition are one of the hottest topics in computer vision. Generally, detection and recognition are separated into two independent parts, while recognition follows detection and feature extraction. An integrated framework for simultaneously vehicle detection and recognition is described in this paper. Our method is based on one of the most effective ways of object detection—DPM (Deformable Part Models). Score bias are used to balance the scores from different models and we mix them all. Not only the detection but also the recognition results are acquired finally. We performed experiments on SJTUVehicle dataset. Experiments show that both the precision of vehicle detection and the over-all accuracy of vehicle recognition reach higher than 95%. Our method can easily be applied to other similar applications.

**Key words:** vehicle detection; vehicle recognition; DPM (Deformable Part Models); score bias; integrated framework

## 0 引言

物体检测与识别是计算机视觉领域的长期研究热点, 在工业制造、安全监控、信息检索等方面具有广泛的应用<sup>[1]</sup>。受光照、视角、尺度等众多因素影响, 真实、不受控环境下的通用物体检测与识别仍然是挑战性难题<sup>[2]</sup>。

通常情况下, 视觉特征的提取和描述、物体检测、类型识别可构成一个完整的检测—识别流程框架。在此过程中, 物体检测和子类型识别常被作为两个独立的模块分别处理, 在检测完成后再训练一个单独的分类器用于物体子类识别。例如, 在车辆检测与识别的应用中, 检测完成后, 仅回答了“车辆在哪儿”的问题, 而要回答“是什么车辆”则需要训练一个单独的分类器。近年来, 出现了一些效果较理想的物体检测方法, 如 HOG + SVM<sup>[3]</sup>、DPM<sup>[4-6]</sup>、regionlets<sup>[7]</sup>、Bing<sup>[8]</sup>等, 这些方法在给出满意的物体检测结果同时, 其实也给出了可用于

再识别的潜在丰富信息。如何利用这些信息, 在一个检测框架内同时进行检测和识别, 是本文方法的主要着眼点。

本文在 DPM 物体检测基础上, 利用该方法能很好地刻画物体局部细节的优势, 挖掘检测过程中获得的有关物体类型的信息, 并将其用于类型识别。首先利用 DPM 训练框架, 分别对各类待识别物体单独训练检测模型。对获得的各类模型分别进行检测效果评价, 根据 F-score 指标, 确定各模型得分偏置的数值, 用于修正模型得分。在检测与识别环节, 利用 DPM 检测框架, 获得各类模型带得分的检测矩形框。再对各模型得分增加相应的得分偏置, 对得分进行整体平移。通过增加得分偏置, 可平衡不同 DPM 模型检测结果的得分, 使它们能够融合为带有类型标记信息的、得分可比较的统一结果, 在完成物体检测的同时, 实现了物体类型分类。

本文给出了方法中相关参数求取的方法和细节, 并以车辆检测与识别作为典型应用, 通过实验

**基金项目:** 国家自然科学基金 (61273285, 61375019); 国家重点基础研究发展计划(973)(2011CB302203)

**收稿日期:** ; 改回日期:

**第一作者简介:** 吴俊 (1991-), 男, 研究生在读, 主要研究方向为计算机视觉, 模式识别。E-mail: zhugetian@sjtu.edu.cn

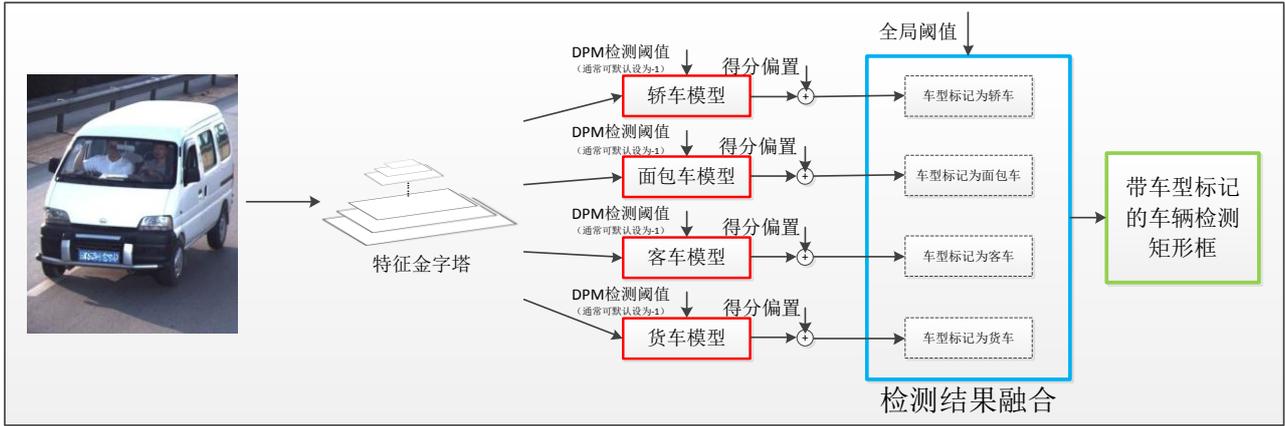


图 1 检测识别流程图

Fig. 1 Flow chart of detection and recognition

验证了方法的有效性，并可扩展应用于其他物体类别的检测和识别中。

## 1 DPM 介绍

DPM 具体实现细节见[6]，以下对 DPM 方法的检测过程进行简要回顾。

DPM 通过多组整体和局部滤波器的协作，能很好刻画被检测物体整体轮廓和内部细节，对于不同类型物体有一定区分能力。该特点为本文类型信息的挖掘提供了可能性。

假设一个完整的 DPM 模型由  $m$  个子模型构成记为  $(M_1, M_2, \dots, M_m)$ ，每个子模型又由  $l$  个根滤波器（记为  $F_0$ ）和  $n$  个局部滤波器（记为  $F_1, F_2, \dots, F_n$ ）组成。将改进的方向梯度直方图<sup>[3]</sup>作为图像特征，在待检测图像多尺度上分别提取特征，形成总层数为  $L$  的特征金字塔  $H$ 。滤波器  $F_i$  在特征金字塔中的位置可用  $p_i = (x, y, l)$  表示，代表其位于特征金字塔第  $l$  层，平面坐标为  $(x, y)$  处。进一步可用  $\phi(H, p)$  表示点  $p$  处特征变形后的特征列向量。此外，用  $d_i$  表示滤波器  $F_i$  与根滤波器锚点之间的平面偏移， $\phi_d(dx_i, dy_i)$  表示偏移对得分的影响系数，用  $b$  表示得分的偏置。单个子模型  $M_c$  在金字塔某处的得分为各个滤波器与特征卷积之和减去由于局部滤波器偏移产生的损失，再加上偏置，故可表示为：

$$\begin{aligned} & score^{M_c}(p_0, \dots, p_n) \\ &= \sum_{i=0}^n F'_i \cdot \phi(H, p_i) - \sum_{i=1}^n d_i \cdot \phi_d(dx_i, dy_i) + b \end{aligned} \quad (1)$$

其中， $F'_i$  表示滤波器  $F_i$  参数变形后的滤波器列向

量。

对子模型各个局部滤波器做变形处理，得到根滤波器  $p_0$  处的最大得分，该得分可表示为：

$$score^{M_c}(p_0) = \max_{p_1, \dots, p_n} score^{M_c}(p_0, \dots, p_n) \quad (2)$$

每个完整的 DPM 模型由多个子模型组成，故 DPM 完整模型的最终得分为：

$$score(p_0) = \max_{c=1 \dots m} score^{M_c}(p_0) \quad (3)$$

再通过阈值  $T$  滤除得分较小的检测框，构成物体检测结果集合  $D$ ：

$$D = \{p_0 \mid score(p_0) > T\} \quad (4)$$

最后，对集合  $D$  中的检测结果使用非最大抑制方法，去除任意两个检测矩形框重叠面积大于 50% 中得分较小的检测框。

## 2 检测与识别的整合计算框架

本系统检测和识别流程如图 1 所示。本文在 DPM 的基础上，对各模型检测结果增加不同的得分偏置，将多个模型检测结果进行融合，输出带车型标记的检测框，从而在完成物体检测任务的同时，实现了物体子类别的分类。作为典型应用，本文主要关注车辆检测和车型识别，不失一般性，本文方法可扩展到其他的物体类别。

### 2.1 基于 DPM 的检测

利用 DPM 训练得到的轿车模型、面包车模型、客车模型以及货车模型对图片特征金字塔进行检测。为了保留较多的检测信息，该环节中的 DPM 检测阈值设置为较小的值，通常为  $-1$ ，这能保留绝大部分的有效检测结果。

### 2.2 检测结果的融合

每类 DPM 模型输出带有得分信息的矩形框。但由于每类模型是分别训练得到的，在类与类之间，得分不具有可比性。因此对各类得分，需要加上一个数值不等的得分偏置进行修正，使其具有可比性，

并融合其结果。对于车型  $t$  而言,即在公式(4)基础上,增加得分偏置  $b_t$ 。修正过程可表示为:

$$D'_t = \left\{ \begin{array}{l} p'_0, \text{score}(p'_0) = \\ \text{score}(p_0) + b_t | p_0 \in D_t \end{array} \right\} \quad (5)$$

对于融合后的结果,利用全局阈值,滤除得分较低检测结果,滤除过程可表示为:

$$D = \left\{ \begin{array}{l} p_0 | \text{score}(p_0) > T_{\text{全局}} \\ , p_0 \in D'_{\text{轿车}} \cup D'_{\text{面包车}} \cup D'_{\text{客车}} \cup D'_{\text{货车}} \end{array} \right\} \quad (6)$$

对上述得到的检测框集合  $D$ ,再次使用非最大抑制方法<sup>[6]</sup>,去除集合中重复的检测矩形框。

### 2.3 车型识别

融合结果后的每个车辆检测框不仅含有坐标和得分信息,还能获知该检测框由哪个 DPM 模型获得,根据该信息可以直接确定该车辆的车型。

通过上述步骤,即可在实现车辆检测的同时完成车型识别。

## 3 实验设置

### 3.1 SJTUVehicle 数据库介绍

本文使用 SJTUVehicle 数据库。该数据库由上海交通大学计算机视觉实验室收集、整理和维护。共收集了近 27 万张车辆图片,从中选取约 5500 张图片(单张图片可能包含多辆车),并标注车辆位置和车型(轿车、面包车、客车及货车)等真值信息建立数据库。图 2 为该数据库图片示例。



图 2 SJTUVehicle 数据库示例  
Fig. 2 Example images of SJTUVehicle dataset

随机选取数据库中三分之一图片作为本文实验训练集,其余三分之二作为测试集。

### 3.2 DPM 模型的训练

首先在 DPM 框架下,分别训练轿车、面包车、客车以及货车的检测模型。考虑到根滤波器与局部滤波器之间的大小尺寸关系,将局部滤波器的数目设置为 6。此外,考虑到不同车辆在图像中的长宽比之间存在差异,结合 DPM 方法的特性,将各类

检测模型的子模型数目设置为 4。

为了强化各类模型区分车型的能力,上述训练环节中,把非本类车辆的样本作为训练负样本。例如在轿车模型的训练过程中,将一定数量的包含面包车、客车和货车的图片作为负样本。通过该方法,可以使各类模型对于非本类车辆的误检做出惩罚,减少误检测。

### 3.3 车辆检测的评价方法

按照 PASCAL<sup>[9]</sup>的评价标准,主要计算精度、召回率以及 F-score 等指标。

对单类模型检测效果评价时,测试集中仅本类车辆图片作为正样本。对最终的车辆检测效果评价时,测试集中所有的车辆图片均作为正样本。

### 3.4 得分偏置的确定

在 3.2 节,可得到对应轿车、面包车、客车和货车,共 4 个模型。本文采用如下方法确定各类模型得分偏置数值——对各类模型单独进行车辆检测效果的评价,对不同阈值分别计算精度和召回率,进而得到 F-score,找到 F-score 最大点对应的阈值,将该阈值的相反数作为模型的得分偏置。

### 3.5 车型识别的评价方法

对于车型分类效果的评价,本文只对检测环节中被正确检测车辆的分类结果进行评价,不考虑误检和漏检车辆。使用准确率、精度、召回率、F-score 以及混淆矩阵等指标进行评价。

## 4 实验结果

### 4.1 各类模型偏置值选取

训练得到各个车型的模型后,再对各类车型单独进行检测效果评价。本实验中,各类模型的检测阈值除面包车设为 -1.5 外,其余均设为 -1。

#### 1) 轿车模型

图 3 所示是对轿车模型检测效果的评价,图 4 是 F-score 曲线。选取 F-score 最大值所对应阈值的相反数,作为轿车模型的偏置值,其偏置值可选为 0.27。

#### 2) 面包车模型

与轿车模型偏置选取类似,面包车模型的偏置值可选为 0.72。

#### 3) 客车模型

与轿车模型偏置选取类似,客车模型的偏置值可选为 0.65。

#### 4) 货车模型

与轿车模型偏置选取类似,货车模型的偏置值可选为 0.45。

### 4.2 车辆检测

对各个模型的检测得分增加上述偏置,融合各类模型检测结果后,我们对整个数据库所有车型检测效果进行了评价。如图 5 和图 6 所示,其中横坐标表示的是全局阈值。

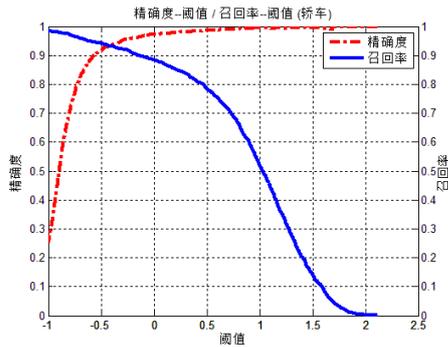


图 3 轿车检测精度、召回率与阈值关系  
Fig. 3 Relationship between threshold and precision, recall for car detection

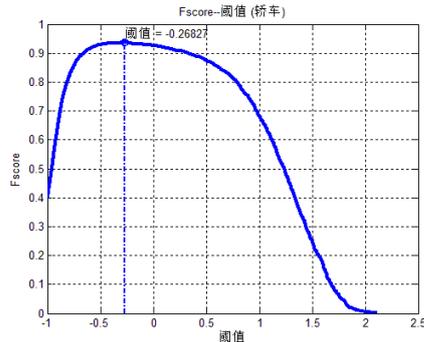


图 4 轿车检测 F-score 与阈值关系  
Fig. 4 Relationship between threshold and F-score for car detection

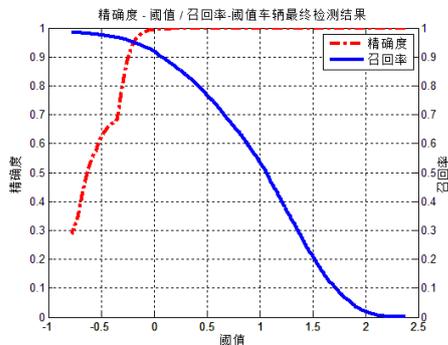


图 5 车辆检测精度、召回率与阈值关系  
Fig. 5 Relationship between threshold and precision, recall for vehicle detection

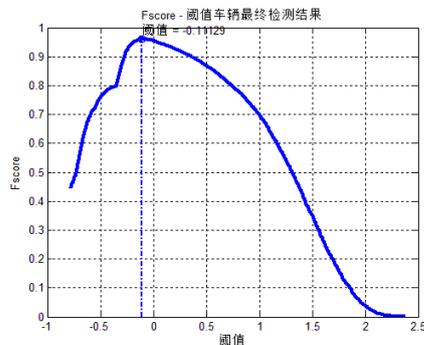


图 6 车辆检测 F-score 与阈值关系  
Fig. 6 F- Relationship between threshold and F-score for vehicle detection

当阈值选为 $-0.10$ 左右时,最终的车辆检测结果能够较好地兼顾精度和召回率。此时,车辆检测的精度为 $98.32\%$ ,召回率为 $93.94\%$ 。

### 4.3 车型识别

#### 4.3.1 全局阈值为 $-0.10$

该情况下,车型识别的整体准确率为 $96.78\%$ ,表 1 是在此阈值下各类车型的识别结果,车型分类混淆矩阵如图 7(a)所示。

表 1 各类车型识别准确率、精度、召回率和 F-score (全局阈值为 $-0.10$ )

Table 1 Accuracy, Precision, Recall and F-score of each type

	准确率	精度	召回率	F-score
轿车	<b>0.9831</b>	<b>0.9706</b>	<b>0.9933</b>	<b>0.9818</b>
面包车	<b>0.9742</b>	<b>0.9660</b>	<b>0.9120</b>	<b>0.9382</b>
客车	<b>0.9882</b>	<b>0.9432</b>	<b>0.9689</b>	<b>0.9559</b>
货车	<b>0.9900</b>	<b>0.9801</b>	<b>0.9685</b>	<b>0.9743</b>

#### 4.3.2 全局阈值为 $0$

在该情况下,车型识别整体准确率为 $97.03\%$ ,结果见表 2,车型分类混淆矩阵如图 7(b)所示。车辆检测的精度为 $99.35\%$ ,召回率为 $92.07\%$ 。

表 2 各类车型识别准确率、精度、召回率和 F-score (全局阈值为 $0$ )

Table 2 Accuracy, Precision, Recall and F-score of each type

	准确率	精度	召回率	F-score
轿车	<b>0.9839</b>	<b>0.9725</b>	<b>0.9933</b>	<b>0.9828</b>
面包车	<b>0.9763</b>	<b>0.9682</b>	<b>0.9202</b>	<b>0.9436</b>
客车	<b>0.9899</b>	<b>0.9489</b>	<b>0.9738</b>	<b>0.9612</b>
货车	<b>0.9906</b>	<b>0.9822</b>	<b>0.9690</b>	<b>0.9755</b>

#### 4.3.3 不设全局阈值

即不滤除融合后结果中得分较小的检测框。

在该情况下,车型识别整体准确率为 $95.40\%$ ,结果见表 3,车型分类混淆矩阵如图 7(c)所示。车辆检测的精度仅为 $28.67\%$ ,召回率为 $98.47\%$ 。

表 3 各类车型识别准确率、精度、召回率和 F-score (不设全局阈值)

Table 3 Accuracy, Precision, Recall and F-score of each type

	准确率	精度	召回率	F-score
轿车	<b>0.9752</b>	<b>0.9666</b>	<b>0.9799</b>	<b>0.9732</b>
面包车	<b>0.9681</b>	<b>0.9517</b>	<b>0.8966</b>	<b>0.9233</b>
客车	<b>0.9800</b>	<b>0.9083</b>	<b>0.9477</b>	<b>0.9276</b>
货车	<b>0.9847</b>	<b>0.9592</b>	<b>0.9604</b>	<b>0.9598</b>

图 8 展示了数据库中部分图片车辆检测和车型识别结果。图中红框代表车辆检测结果,框中的文字信息为车型识别结果以及增加偏置后的检测得分。

## 5 结论

由图 5 和图 6 可见,全局阈值对车辆检测效果有较大影响,在整个变化范围内与召回率始终保持负相关,当全局阈值大于一定值后,精度一直保持在较高水平。从 4.3 节可以看出,车型识别结果

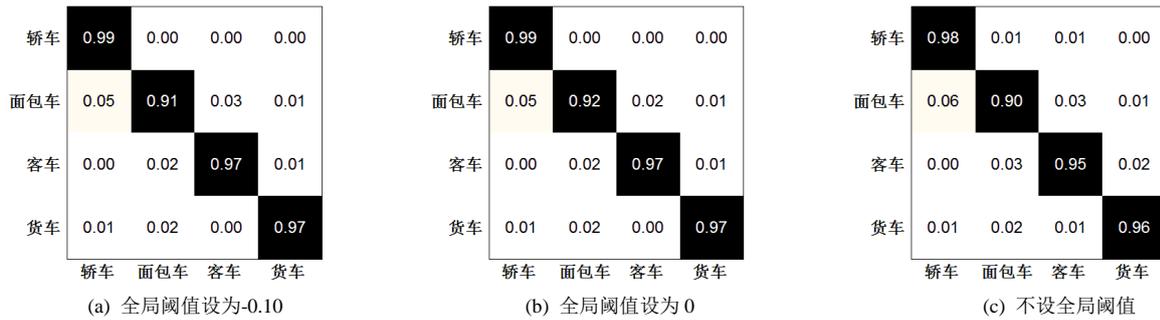


图 7 车型识别混淆矩阵

Fig. 7 Confusion matrix for recognition



图 8 车辆检测及车型识别效果示例

Fig. 8 Example images of experiment results

不易受到全局阈值的影响，虽然随着全局阈值的增大，车型识别效果变好，但变化较小。因此，全局阈值的选取主要取决于对车辆检测的要求，全局阈值越低，车辆检测召回率越高，但精度越低。

混淆矩阵中，面包车的分类效果相对较差，有较多面包车被错分为轿车。其原因是，SJTUVehicle 数据库将别克商务车类车辆定义为面包车，但系统更加倾向于将其分类为轿车。与之类似，SJTUVehicle 数据库将依维柯类车辆定义为客车，而系统倾向于将其划分为面包车。此外，由于微型货车和面包车的前脸十分相似，两者之间较容易出现错分情况。

下一步我们将把得分偏置的计算融合到 DPM 的训练过程中，利用数据驱动，自动寻找合适的模型得分偏置。

## 参考文献 (References)

[1] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition: IEEE, 2001: I-511-I-518.

[2] Yilmaz A, Javed O, Shah M. Object tracking: A survey[J]. Acm computing surveys (CSUR), 2006, 38(4):13.

[3] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005:886-893

[4] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK: IEEE, 2008:1-8.

[5] Felzenszwalb P, Girshick R, McAllester, D. Cascade object detection with deformable part models[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA: IEEE, 2010:2241-2248.

[6] Felzenszwalb P, Girshick R, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models[J]. Pattern Analysis and Machine Intelligence, 2010, 32(9):1627-1645.

[7] Wang X, Yang M, Zhu S, et al. Regionlets for generic object detection[C] // Proceedings of International Conference on Computer Vision. Sydney, NSW: IEEE, 2013:17-24.

[8] Cheng M M, Zhang Z, Lin W Y, et al. BING: Binarized normed gradients for objectness estimation at 300fps[C]. IEEE CVPR, 2014.

[9] Everingham M, Van Gool L, Williams C K I, et al. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results. [EB/OL]. <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2010/>