

多卷积特征融合的HOG行人检测算法

高琦煜 方虎生

(陆军工程大学 南京 210007)

摘 要 行人检测是计算机视觉领域中的经典问题,HOG 结合 SVM 的方法是解决这一问题的有效途径,HOG 对行人特征的有效描述起到了重要作用。卷积神经网络(CNN)作为一种有效的特征提取方法,通过特征图可以实现对特征更好的描述。提出将卷积神经网络(CNN)与传统的 HOG+SVM 算法相结合的方法。首先利用 CNN 在下采样层中可以使用不同的卷积核对数据进行不同角度特征描述的特点,对样本进行多角度浅层特征提取;然后用 HOG 对得到的浅层特征进行进一步的提取;最后采用支持向量机(SVM)完成训练、分类。实验表明,该方法对于行人检测具有很高的识别率,优于传统方法。

关键词 行人检测,卷积神经网络,多卷积特征,HOG

中图法分类号 TP139.41 **文献标识码** A

HOG Pedestrian Detection Algorithm of Multiple Convolution Feature Fusion

GAO Qi-yu FANG Hu-sheng

(PLA Army Engineering University, Nanjing 210007, China)

Abstract Pedestrian detection is utilized as the fundamental of various computer vision applications. A typical and effective solution of pedestrian detection is combining histogram of oriented gradient (HOG) with support vector machine (SVM). In this paper, we proposed a novel pedestrian detection method, which uses convolutional neural network (CNN) in HOG+SVM to obtain more comprehensively feature description through a variety of convolutional kernels in CNN. Firstly, it extracts shallow features from data set by CNN, and uses CNN's characteristics of displacement, and scale and deformation invariance. Then, it merges strongly relevant features by analyzing correlation coefficient of shallow features. After this, it extracts HOG features from CNN shallow features which are weakly correlated with each other. Finally, it uses SVM to complete the training and classification. The experimental results show that the proposed method can obtain higher accuracy than the existing method in pedestrian detection.

Keywords Pedestrian detection, Convolutional neural network, Multi-convolutional features, Histogram of oriented gradient

1 引言

行人检测是计算机视觉中的一个重要课题,广泛地应用于国防工程、公共安全防护、自动驾驶安全等领域,具有很大的研究意义。目前,常用的行人检测方法有两种:1)基于全局特征的行人检测技术,其核心思想是将高维的行人图像特征通过线性或者非线性变化映射到一个低维子空间中,使得原始样本特征在该低维子空间中更容易分类^[1];2)基于局部特征的行人检测技术,局部特征侧重于反映行人的细节特征,其中应用较广泛的有尺度不变特征变换(Scale Invariant Feature Transform, SIFT)^[2]以及梯度方向直方图(Histogram of Oriented Gradient, HOG)^[3-4]等。

梯度方向直方图特征(HOG)是一种在计算机视觉和图像处理中用来进行物体检测的特征描述子。边缘是图像最基本的特征之一,携带着大量的图像信息,图像的边缘通常是图

像深度不连续处、图像梯度朝向不连续处、图像光照强度不连续处和纹理变化处^[3]。而 HOG 正是描述图像方向梯度的特征,因此 HOG 具有很好的分辨能力。

近年来,随着深度学习概念的提出,其强大的特征提取能力、高效的识别效率引起了广泛关注。作为深度学习模型的一种,卷积神经网络不需要通过人工手动设置特征,它可以通过建立深度神经网络的训练机制来模拟人脑的学习过程,进行无监督学习以及有监督学习,改变传统机器学习中人为设计特征提取方法的方式,自主地进行特征提取,减少人为干预^[5-6]。本文提出一种行人检测算法,首先利用 CNN 的浅层特征图实现对图像的多特征描述;然后利用 HOG 对各特征图分别进行特征提取,并对得到的特征进行相关性分析,将相关性较低的特征组合为新的 HOG 特征;最后通过 SVM 进行分类识别。实验证明,本算法在行人检测中比现有的 HOG+SVM 方法具有更高的识别率。

本文受国家 973 计划资助项目(2009CB723803),国家自然科学基金资助项目(60873120)资助。

高琦煜(1992—),男,硕士生,主要研究方向为机器视觉、模式识别,E-mail:540473772@qq.com;方虎生(1982—),男,博士,副教授,主要研究方向为计算机技术,E-mail:fhssxxll@sohu.com。

2 卷积神经网络的原理与特点

卷积神经网络(Convolutional Neural Network, CNN)是一种特殊的前馈神经网络,它的人工神经元可以影响一部分覆盖范围内的周围单元,对于大型图像处理有着出色的表现,它包括卷积层(alternating convolutional layer)和池化层(pooling layer)。卷积神经网络的特殊性体现在两个方面:1)其神经元间是稀疏连接的;2)同一层神经元之间的连接权值(卷积核)是相同的^[7-8]。稀疏连接和权值共享的网络结构使卷积神经网络大大降低了网络模型的复杂度。

稀疏连接:图像的空间联系是局部的,一个神经元不需要感受全局图像,只需感受局部的图像区域,在更高层将这些感受不同局部区域的神经元综合起来就可以得到全局的信息^[9]。因而减少了连接的数目,减少了神经网络需要训练的权值参数的个数。

权值共享:卷积神经网络是一种多层的监督学习网络,卷积层和下采样层交替出现。卷积层由多个区域(map)组成,每个map由多个神经单元组成,同一个map中的所有神经单元共用一个权值,即权值共享。这样大大降低了网络的复杂度,但是由于共享一个权值,相当于只提取了一种特征,因此会限制检测效果。我们通常会设置多个不同的卷积核进行多角度的特征提取,这些卷积核之间存在着一定的相关性。

卷积神经网络采用对函数核影响小的sigmoid函数作为卷积网络的激活函数,使得特征映射具有位移不变性^[10]。此外,一个map上的神经元共享权值,减少了网络自由参数的个数,降低了网络参数选择的复杂度。卷积神经网络中的每一个特征提取层都紧跟着一个用来求局部平均与二次提取的计算层,这种特有的两次特征提取结构使网络对输入样本有较高的畸变容忍能力。

卷积层的计算公式为:

$$x_l^j = f(\sum_{i \in M_j} x_{l-1}^i \times k_{ij}^l + b_l^j) \quad (1)$$

其中, l 代表层数, k 代表卷积核, M_j 代表输入层的感受野, b 代表偏置。

池化(Pooling)的计算公式为:

$$h_{i+1,x,y} = \sqrt{\sum_{(j,k) \in N(x,y)} h_{i,j,k}^2} \quad (2)$$

3 HOG 的原理与特点

HOG特征是一种在计算机视觉和图像处理中用来进行物体检测的特征描述子。它通过计算和统计图像局部区域的梯度方向直方图来构成特征。

特征提取过程如下。

(1)将整个图像进行归一化。在图像的纹理强度中,局部的表层曝光贡献的比重较大,因此适当的压缩处理能够有效减弱图像阴影和光照变化带来的影响^[15]。

伽马函数(Gamma)压缩公式为:

$$I(x,y) = I(x,y)^{\text{gamma}} \quad (3)$$

通常取 $\text{gamma} = 1/2$ 。

(2)计算图像横坐标和纵坐标方向的梯度,并据此计算每个像素位置的梯度方向值。

图像中像素点 (x,y) 的梯度为:

$$\begin{aligned} G_x(x,y) &= H(x+1,y) - H(x-1,y) \\ G_y(x,y) &= H(x,y+1) - H(x,y-1) \end{aligned} \quad (4)$$

其中, $G_x(x,y), G_y(x,y), H(x,y)$ 分别表示图像中像素点的方向梯度和像素值。

(3)用 $[-1,0,1]$ 梯度算子对原图像做卷积运算,得到 x 方向的梯度分量 G_x ,用 $[1,0,-1]^T$ 梯度算子对原图像做卷积运算,得到 y 方向的梯度分量 G_y 。然后利用式(5)计算该像素点的梯度大小和方向。

像素点 (x,y) 处的梯度幅值和梯度方向分别为:

$$\begin{aligned} G(x,y) &= \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \\ \alpha(x,y) &= \tan^{-1} \left(\frac{G_y(x,y)}{G_x(x,y)} \right) \end{aligned} \quad (5)$$

(4)将图像分成若干个单元格(cell),对单元格内的每个像素用梯度方向在直方图中进行加权投影,可以得到这个单元格的梯度方向直方图。而梯度大小就是计算投影的权值。

把cell单元组合成大的、空间上连通的区间。将一个区间内所有cell的特征向量串联起来即可得到该区间的HOG特征。这些区间是互有重叠的,这就意味着:每一个单元格的特征会以不同的结果多次出现在最后的特征向量中。我们将归一化之后的块描述符称为HOG描述符,即为所要提取的HOG特征。

4 基于多卷积特征的 HOG 行人检测

前文提到卷积神经网络的核心思想是局部感受野、权值共享以及时间或者空间降采样。基于这3种结构思想,卷积神经网络获得了某种程度的位移、尺度、形变不变性。因此卷积神经网络具有强大的学习能力和表达能力,能够很好地提取出数据的有效特征。通常,每一卷积层会设置多个卷积核,这些不同的卷积结果是对原图像不同的特征描述,CNN也是通过这种方式获得了对图像多样性的特征描述。将CNN多特征描述与HOG相结合,来改变经典HOG只对原图进行特征提取的方式,这将会提高HOG特征描述的能力;卷积神经网络的学习过程是对原始数据不断抽象的过程,它能够从像素级原始数据到抽象的语义概念逐层提取信息,因此随着卷积神经网络层数的增加,其提取的特征的抽象程度逐渐增加,同时其表达能力也越来越强大。但是对于HOG来说,其特征提取的对象是像素级的梯度特征,对于高度抽象的数据难以提取出有效的特征信息。我们进行了大量的对比实验,通过对比发现,在经过一次卷积后得到的特征图可以很好地被HOG使用,而如果再进行一次卷积计算,特征变得更加抽象,用HOG已经不能完整地提取出所需要的特征。本文提出采用CNN浅层多样性特征与HOG相结合的方法,改善经典HOG从单一原图像获取特征的方式,赋予HOG更强的特征描述能力,从而提高行人检测的效果。

本文方法的实现步骤如图1所示,将第一卷积层的每组特征映射都作为HOG的输入,经过HOG提取得到多组特征,将它们分别输入到分类器SVM中,得到分类结果。

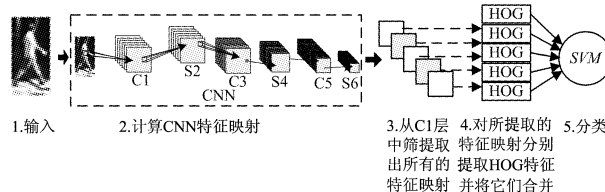


图1 实验框架1

在 CNN 应用中,为了使网络获得丰富的特征描述,对每个卷积层一般都会设置较多的特征映射(feature map),这些特征映射提供了对特征的多样性描述,同时也可能产生对特征描述的冗余。在利用 HOG 对第一层卷积结果进行特征提取时,冗余的特征映射不但不能提供有效的特征描述,反而会使分类器陷入“维度灾难”,从而降低分类器性能。为了避免上述情况,本文对第一层卷积核进行相关系数分析,强相关卷积核所产生的卷积结果必然存在信息冗余,因而只保留其中一个特征映射,因为弱相关卷积核所产生的卷积结果是对原图像不同的特征表达,所以予以保留。由于单一的特征映射对图像特征的描述是单一的,因此本文将相关性较弱的特征映射分别通过 HOG 进行特征提取,然后将其 HOG 特征列向量进行相加,得到最终的 HOG 特征,并交由 SVM 进行分类。实验框架如图 2 所示。

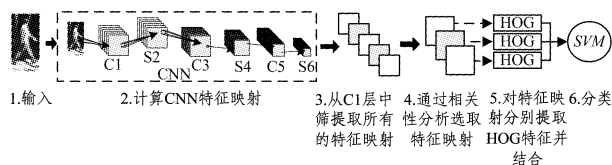


图 2 实验框架 2

5 实验与分析

5.1 样本集

为了使实验数据更加准确、具有代表性及说服力,在样本集的选择上采取了标准样本集与自制样本集相结合的模式,标准样本集更加具有代表性及说服力,自制样本集则针对一些特殊背景环境针对性地选取行人样本图片。标准样本集使用 DC-ped-dataset 数据集与 CVC 数据集,DC-ped-dataset 包含 4000 幅行人图像与 5000 幅背景图像,所有图像都是从录像中截取得到的。CVC 数据集中包含 3678 幅行人图片与 4048 幅背景图片。为了使样本集更加丰富,通过手动截取行人图片来自制样本集。样本集中共有正样本 23000 幅,负样本 25000 幅,其中选取正样本 19000 幅、负样本 21000 幅作为训练集。将其余正样本 4000 幅、负样本 4000 幅作为测试集。样本集部分示例如图 3 所示。



图 3 样本集部分示例

5.2 实验参数设置

特征提取的关键因素是卷积核尺寸的选择,其直接决定了感受野的大小,感受野过大则会超出卷积核的表征范围,过小则无法有效提取局部特征。因此,卷积核尺寸的大小应根据检测对象的大小和复杂程度来决定。本文分别采用 3×3 、

5×5 、 7×7 的卷积核进行实验,可视化结果如图 4 所示。

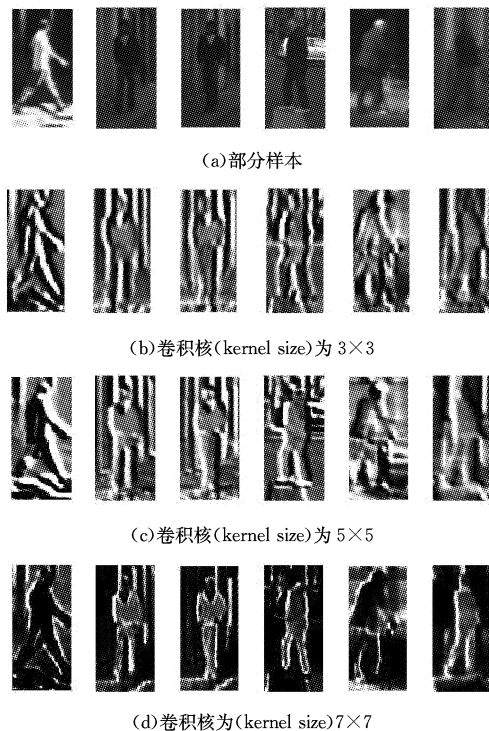


图 4 卷积核的效果对比

由图 4(d)可以看出,当卷积核为 7×7 时效果最好,因此本文实验采用 7×7 的卷积核。

通过卷积计算得到的特征映射即为我们所需要的浅层特征,由图 2 得知,在卷积神经网络的第一卷积层中有 6 个特征映射,因此得到 6 组 HOG 特征,并分别用 SVM 对这 6 组 HOG 特征进行分类。其结果如表 1 所列。

表 1 不同特征映射的实验效果对比/%

编号	1	2	3	4	5	6
识别率	94.96	95.15	95	95.18	95.45	94.82

从表 1 中可以看出,使用 6 组特征映射进行行人检测的识别率全部在 95% 左右,效果并不理想。因此,在接下来的实验中,本文分别对 6 种不同的卷积核进行相关性分析,结果如表 2 所列。

表 2 特征映射的相关性分析

编号	1	2	3	4	5	6
1	1	0.4181	0.1746	0.4534	0.2546	0.7162
2		1	0.5274	0.8279	0.9345	0.3793
3			1	0.2598	0.2879	0.1929
4				1	0.7650	0.5868
5					1	0.6468

本文对表 2 中的数据进行相关性分析,试图找出其中相关性最低的 3 组特征映射。第 1 组特征映射与第 3 组、第 5 组特征映射的相关性最低,分别为 0.1746 与 0.2546,并且第 3 组与第 5 组特征映射的相关性也较低,为 0.2876。因此选取 1,3,5 共 3 组特征映射作为实验对象。接下来,本文使用 HOG 分别对这 3 组特征映射进行特征提取,得到 3 个 HOG 特征向量,然后将这 3 个 HOG 向量进行列向量相加,组合成最终的 HOG 分类特征列向量,最后采用支持向量机对图像进行分类。

(下转第 232 页)

- gram thresholding-Fuzzy C-means hybrid approach[J]. Pattern Recognition, 2011, 44(1): 1-15.
- [6] WANG H, ZHANG H, RAY N. Adaptive shape prior in graph cut image segmentation[J]. Pattern Recognition, 2013, 46(5): 1409-1414.
- [7] 郑庆庆, 桑农, 高常鑫, 等. 改进的基于区域合并的纹理图像分割方法[J]. 华中科技大学学报(自然科学版), 2011(5): 109-112.
- [8] 张竹伟, 周海英. 基于区域色彩空间和 LBP 算法的图像分割[J]. 微电子学与计算机, 2015(2): 84-87.
- [9] 何坤, 郑秀清, 张永来. 纹理模糊的图像分割[J]. 四川大学学报(工程科学版), 2015, 47(4): 111-117.
- [10] 刘毅, 黄兵, 孙怀江, 等. 结合 CS-LBP 纹理特征的快速图像分割算法[J]. 计算机科学, 2013, 40(5): 300-302.
- [11] 汤顺青. 色度学[M]. 北京: 北京理工大学出版社, 1990.
- [12] SUN Y, XU R, CHEN L, et al. Image compression and encryption scheme using fractal dictionary and Julia set[J]. Iet Image

- Processing, 2015, 9(3): 173-183.
- [13] DELP E J, MITCHELL O R. Image Compression Using Block Truncation Coding[J]. IEEE Transactions on Communications, 1979, 27(9): 1335-1342.
- [14] OJALA T, PIETIK, INEN M, et al. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7): 971-987.
- [15] LIU L, LAO S, FIEGUTH P W, et al. Median Robust Extended Local Binary Pattern for Texture Classification[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2016, 25(3): 1368-1381.
- [16] TANG Z, SU Y, MENG J E, et al. A local binary pattern based texture descriptors for classification of tealeaves[J]. Neurocomputing, 2015, 68(C): 1011-1023.

(上接第 201 页)

5.3 实验结果及分析

本文的实验结果是对 DC-ped-dataset 与 CVC 标准数据集测试得到的, 本文在实验过程中将正、负测试集样本进行标号排序, 并设定测试集中正样本标签为 1, 负样本标签为 0。在实验过程中算法若将正样本识别为 0 或者将负样本识别为 1 皆为误识, 则识别率 = 正确识别数 / 测试集总数。本文在 CVC 标准测试集中选取 3000 幅行人图片作为训练正样本, 选取 3400 幅非行人图片作为训练负样本。选取正、负测试集样本各 600 幅, 分别采用经典 CNN 行人检测算法、经典 HOG+SVM 算法与本文算法进行实验。实验结果如表 3 所列, 本文算法的识别率为 97.825%, 优于经典 CNN 算法(96.937%)与传统 HOG+SVM 算法(96.0125%)。在 DC-ped-dataset 标准测试集中选取 3000 幅行人图片作为训练正样本, 4000 幅非行人图片作为训练负样本。选取正、负测试集样本各 1000 幅。实验结果如表 4 所列, 经典的 CNN 算法识别率为 97.223%, 高于经典 HOG+SVM 行人检测算法, 而本文算法的识别率为 98.325%, 远高于其他两种经典算法。

表 3 CVC 数据集实验效果对比/%

检测方法	CNN+HOG+SVM	HOG+SVM	经典 CNN 算法
识别率	97.825	96.0125	96.937

表 4 DC-ped-dataset 实验效果对比/%

检测方法	CNN+HOG+SVM	HOG+SVM	经典 CNN 算法
识别率	98.325	96.768	97.223

结束语 随着人工智能概念进入公众的视野, 越来越多的研究人员关注其相关的领域, 不仅许多类似于深度学习的新兴算法被提出, 而且一些传统算法也在不停地发展。本文提出的利用 CNN+HOG 的特征提取方式就是一种“新, 老”算法的结合, 使用了多特征融合方法, 首先使用 CNN 在下采样层对数据进行多角度的特征描述, 然后用 HOG 对得到的特征图进行特征提取, 将得到的特征矩阵进行相关性分析, 取

其中相关性低的 3 组特征映射矩阵, 将它们组成一个新的特征矩阵, 最后交由 SVM 进行分类。实验证明, CNN+HOG 的多特征融合的特征描述比传统的单一的 HOG 特征更优秀, 识别率更高。在行人检测领域中, CNN+HOG 的特征表征能力更强, 比经典 CNN 算法更适合用来进行行人检测。

参考文献

- [1] HYVARINEN A. Fast and robust fixed-point algorithms for independent component analysis [J]. IEEE Transactions on Neural Networks, 1999, 10(3): 626-634.
- [2] LOWE D. Distinctive image features from scale Invariant keypoints[J]. IJCV, 2004, 60(2): 91-110.
- [3] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//CVPR. 2005.
- [4] CHENG H S, HU X F. The human body detection based on HOG and SVM technology research in static image[J]. Chinese Journal of Scientific Instrument, 2012, 19(5): 20-23.
- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Image net classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. 2012: 1097-1105.
- [6] HINTON G E, SALAKHUTDINOV R. Reducing dimensionality of data with neural networks[J]. Science, 2006(504): 313.
- [7] BENGIO Y, COURVILLE A C, VINCENT P. Unsupervised feature learning and deep learning: A review and new perspectives [J]. CoRR abs/1206.2538.
- [8] SERMANET P, KAVUKCUOGLU K, CHINTALA S, et al. Pedestrian detection with unsupervised and multi-stage feature learning[C]//CVPR. 2013.
- [9] ENZWEILER M, EIGENSTETTER A, Schiele B. Multi-cue pedestrian classification with partial occlusion handling[C]//CVPR. 2010.
- [10] VIOLA P, JONES M J, SNOW D. Detecting pedestrians using patterns of motion and appearance[C]//IJCV. 2005: 153-161.
- [11] 田仙仙, 鲍泓, 徐成. 一种改进 HOG 特征的行人检测算法[J]. 计算机科学, 2014, 41(9): 237-259.