

基于卷积神经网络和密度分布特征的人数统计方法

郭继昌, 李翔鹏

(天津大学电气自动化与信息工程学院 天津 南开区 300072)

【摘要】在行人监控视频中, 由于行人遮挡、场景光照变化, 人群分布不均等因素的影响使得现有方法难以准确统计视频中人数。针对该问题, 提出一种基于卷积神经网络和密度分布特征的人数统计方法。该方法首先将场景中的人群依据密度进行划分; 对稀疏人群, 使用Retinex算法将场景去噪后转换至HSV空间中对行人位置进行预判, 并使用栅极损失函数分块训练卷积神经网络提取行人特征, 实现对遮挡行人局部位置的识别; 对密集人群, 提取人群密度分布特征并使用多核回归函数估计人群数量。该算法在PETS2009、UCSD等数据集上进行了测试, 实验结果表明所提算法具有更好的统计精度。

关键词 Caffe; 卷积神经网络; 人数统计; 密度分布特征

中图分类号 TN919.8

文献标志码 A

doi:10.3969/j.issn.1001-0548.2018.06.002

A Crowd Counting Method Based on Convolutional Neural Networks and Density Distribution Features

GUO Ji-chang and LI Xiang-peng

(School of Electrical and Information Engineering, Tianjin University Nankai Tianjin 300072)

Abstract Crowd counting is difficult to get accurate statistics due to shading, shadows and changes in crowd density. This paper presents an approach to combine the convolutional neural networks and density features map legitimately. We segment the crowd scene into many blobs according to the density. For low-density blobs, Retinex algorithm is used to denoise the scene and then the scene is transformed into HSV color space to locate the pedestrian. Convolutional neural networks are used to extract the pedestrian features with grid loss function to avoid the occlusion issue. For high density blobs, crowd density distribution features are extracted to train the multiple kernel regression models to estimate the numbers. Experiments are conducted on datasets PETS2009, UCSD. The experimental results show that the proposed method improves the accuracy to some extent in comparison with other algorithms.

Key words Caffe; convolutional neural networks; crowd counting; density distribution features

近年来, 随着计算机技术的不断发展, 智能视频监控已在商场、学校、火车站等公共场所大量应用, 以保障社会有序稳定运行。人数统计是智能视频监控领域的一个具有现实意义的研究方向, 也是计算机视觉领域的研究热点和难点之一。准确统计监控场景中的人数在公安防控、商业信息采集以及配置社会资源和设施上具有重要意义。

行人统计使用的方法主要有基于目标检测的人数统计方法和基于特征回归的人数统计方法, 这两类方法均用到了有监督的机器学习思想, 此外还有基于无监督学习的跟踪轨迹聚类^[1]方法等。在有监督学习方法中, 基于HOG^[2]算法检测行人是被广泛使用的方法之一, 该方法通过计算和统计图像局部

区域的梯度方向直方图来构成行人特征, 此外还有提取行人头部、面部特征和模板匹配的检测方法, 如LBP算法^[3], DPM算法^[4], 再使用机器学习中SVM分类器以及Adaboost级联分类器^[3]训练出行人分类器进而识别和检测行人。此类方法在行人遮挡严重以及光照不足时, 精度较差; 基于特征回归的方法则通过提取区域的纹理等特征, 然后采用核函数实现纹理特征到人数的回归映射。此类方法可以有效的降低行人互相遮挡对检测的影响, 但是人群分布特征很难使用数学特征完全描述, 影响统计准确性。

近几年, 随着深度学习理论的日趋成熟以及硬件设备性能的提升, 使得卷积神经网络成为计算机视觉与模式识别领域的一个有力工具。文献[5]将卷

收稿日期: 2017-04-26; 修回日期: 2017-12-04

基金项目: 国家重点基础研究计划项目(2014CB340400)

作者简介: 郭继昌(1966-), 男, 博士, 教授, 主要从事数字图像处理、模式识别方面的研究。

积神经网络结构优化后应用于目标分类, 在 ImageNet 图像数据库上的测试中取得了令人满意的结果。文献[6]提出了 R-CNN(region proposal-CNN) 算法并创建了 GPU 上运行的 Caffe 框架, 成为近年深度学习中实现目标检测的经典算法之一。文献[7]采用卷积神经网络将人群的分布特征提取后投入训练, 用网络生成的模型估计视频中行人数量, 成为深度学习应用于行人检测的一个实例。这些算法虽然提取到了有效的行人特征并建立了精确的预测模型, 但仍然不能很好地解决遮挡和光照变化以及人群分布不均等因素对检测的影响。

为了解决上述问题, 在提取监控视频中行人特征之前需要考虑到实际场景中的情况, 如人群分布聚集不规律、行人间存在相互遮挡、光照度偏暗以及雨雾天气等复杂的室外环境因素的影响。本文针对以上问题对现有算法做出选择和改进, 提出结合卷积神经网络和密度分布特征的人数统计算法, 将

不同的特征提取算法应用在其对应的场景中, 从而有效解决了上述干扰因素, 增加了算法统计精度。

1 方法原理及流程

行人统计的关键在于寻找含有行人的区域并提取其特征。由于人群分布的随意性使得特定时间段内场景中人群不规律聚集, 稠密度不均。为了提高精度, 本文提出在不同稠密度的场景下分别使用不同的特征提取方法: 在人数密集的场景下, 人和人之间存在遮挡, 人群呈团块区域状分布, 此时采用提取区域纹理、密度分布特征的方法来估计行人数量; 在人群稀少的场景中, 使用统计回归方法来估计人数误差会很大, 此时采用行人识别的方法可以简单快捷的来统计人数。因此, 本文的行人统计算法分为行人识别和特征回归两部分, 以应对不同复杂度的人群场景。

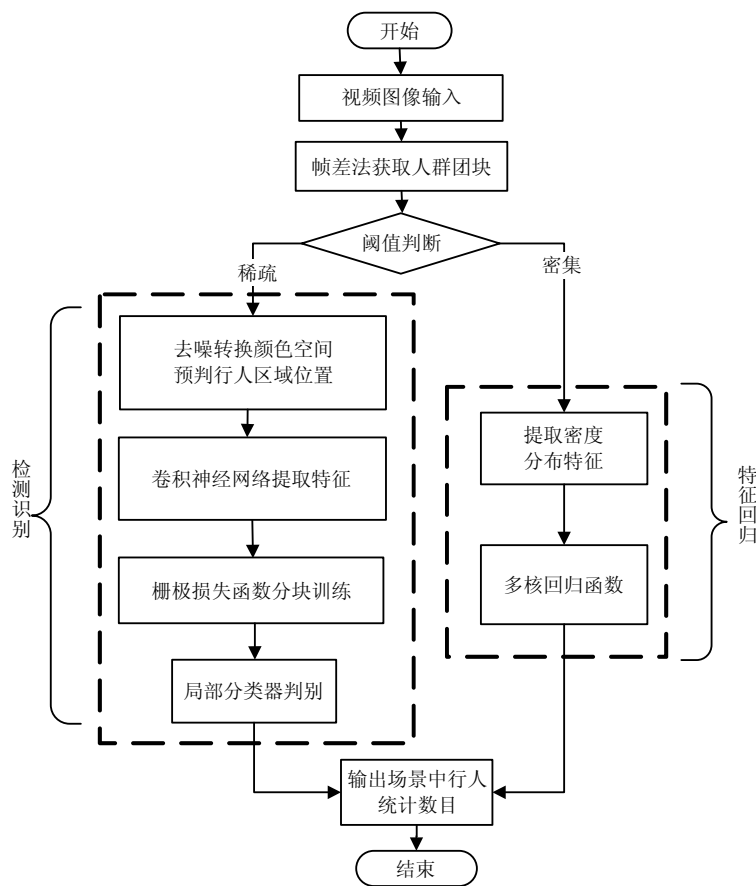


图1 人数统计算法框架结构示意图

本文算法将视频中的人群依据稠密度分割成若干团块, 统计各团块的像素数和周长, 取中值作为阈值 θ , 像素数和周长小于该阈值则记为稀疏人群团块, 反之为密集人群团块。对于稀疏人群场景, 在

使用卷积神经网络提取行人特征之前, 加入选择搜索算法对行人所处区域进行预判, 剔除冗余特征。为了避免光照变化对行人定位造成干扰, 首先采用 Retinex 算法^[8]对场景去噪, 并利用 HSV 颜色空间中

各个分量相互分离对光照变化不敏感的特性，将行人场景由RGB颜色空间转换为HSV颜色空间后再判断行人位置和特征提取。在卷积神经网络训练时，采用栅极损失函数^[9]分块训练网络，可以增加算法对行人局部位置的识别，从而有效减少遮挡对识别的影响。对于密集人群场景，在训练中设置描述子来标注行人目标，再选择线性映射将特征向量转换为密度值并构建人群密度分布图，借助其良好的描述人群分布信息的能力以提高回归模型统计人数的精确度。算法的流程图如图1。

2 稀疏场景中行人识别算法

在稀疏场景中，人与人之间有间隔距离或者轻微遮挡，此时适用于对单个行人目标进行的检测和识别算法，该场景下的行人识别算法分为3个步骤：

- 1) 使用Retinex算法对场景去噪后转换至HSV颜色空间，并采用选择搜索算法^[10]找到包含行人的区域和位置；
- 2) 使用栅极损失函数分块训练卷积神经网络提取行人局部区域特征；
- 3) 使用局部分类器判断提取出的特征是否为行人目标。

2.1 对行人区域的预判

在行人检测识别中，输入卷积神经网络的训练数据精确与否决定了提取出的行人特征质量的好坏，本文借鉴目标识别中的R-CNN算法，在训练卷积神经网络之前采用选择搜索算法对行人区域进行预判和选择，使得训练目标更具针对性。考虑到行人场景容易受到光照变化的影响，本文将选择搜索算法进行了改进。从表1^[11]可以看出，HSV颜色空间对光强，阴影和高光变化的稳定性最强，因为HSV颜色空间是将白色光区域沿着RGB模型对角向黑色光区域投射得到，在这个空间中色调H和饱和度S分量与明度V是分离开来的，因而对光照变化影响很小。图2是行人区域定位结果图。

表1 不同颜色空间对光照变化的稳定性

颜色空间类型	RGB	I	Lab	HSV	rgb	C	H
光强	-	-	+/-	2/3	+	+	+
阴影/遮挡	-	-	+/-	2/3	+	+	+
高光	-	-	-	1/3	-	+/-	+

(注：“+/-”表示部分颜色分量不变，1/3表示某一通道颜色分量不变)

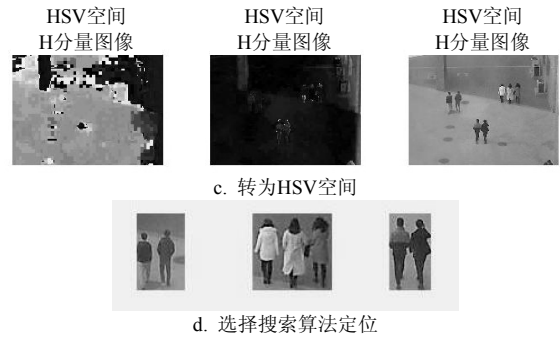
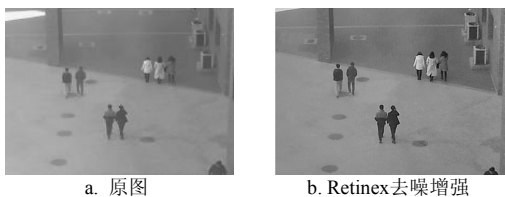


图2 行人区域预判结果示意图

2.2 卷积神经网络结构设计

有监督卷积神经网络可以进行多层卷积运算，通过反复迭代训练以及反向传播等方式调整权值和参数。网络结构通常包含卷基层、池化采样层和全连接层。其中对卷积层和池化采样层模块的设计是特征提取的关键步骤，此外为了保证有效完成对卷积神经网络的回归训练，损失函数的选择也至关重要。使用它提取目标区域中的行人特征，在大量行人数据集上训练后可以建立精确的预测模型。

2.2.1 栅极损失函数

传统的损失函数是直接根据整个图片的信息特征计算损失函数(global loss)，使得训练出的神经网络适于全局目标的识别。而栅极损失函数整合局部与整体特征，将最后一个卷积层的特征图划分为 $n \times n$ 的小栅极，每个小栅极看成一个单独的区域(blocks)，单独按同样的方式计算每一个小区域的损失函数，将整个区域的损失函数加和作为最终损失函数，以强化每个小栅极区域的独立判别能力。如果行人的身体或者脸的一部分被遮挡住，其余子块可以将该区域没有遮挡住的部分输入检测器，这样降低了遮挡对检测的影响。

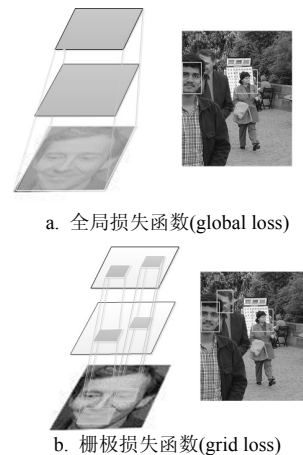


图3 栅极损失函数原理图

设定 x 表示 $f \times r \times c$ 维的最后一层卷积层 r 列 c

行的特征向量图, f 表示滤波器的数量, 本文将 x 分割为 $f \times n \times n$ 格非重叠的小栅极区域 f_i , $i=1, 2, \dots, N$, 其中 $N = \left\lfloor \frac{r}{n} \right\rfloor \cdot \left\lfloor \frac{c}{n} \right\rfloor$, 并选用铰链损失函数(hinge loss):

$$l(\theta) = \sum_{i=1}^N \max(0, m - y(\omega_i^T f_i + b_i)) \quad (1)$$

式中, $\theta = [\omega_1, \omega_2, \dots, \omega_N, b_1, b_2, \dots, b_N]$; m 为常数 $1/N$, 表示每个栅极区域对分类贡献相等; $y \in \{-1, 1\}$ 是分类标签; ω_i 和 b_i 是第 i 块的权值向量和偏置参数。为了避免每个局部子分类器输出相似冗余的分类结果, 把每个分类器赋予相应权重将每个局部区域的损失函数联结起来, 通过局部分类器共享权值的算法。设权值 $\omega = [\omega_1, \omega_2, \dots, \omega_N]$, 偏置 $b = \sum_i b_i$, 损失函数变为:

$$l(\theta) = \max(0, 1 - y \cdot (\omega^T x + b)) + \lambda \sum_{i=1}^N \max(0, m - y \cdot (\omega_i^T f_i + b_i)) \quad (2)$$

式中, N 代表栅极的个数; ω_i 与 b_i 是最后一个特征图对应的整体权值参数; $b = b_1 + b_2 + \dots + b_N$ 为其对应偏置项; 公式的第一项代表了特征图上的整个损失; 第二项代表了每一个栅极的损失; λ 是一个平衡系数, 权衡全局与局部损失的大小。训练卷积神经网络时使用梯度下降算法对该损失函数进行优化。

使用该损失函数训练时可以让神经网络获得更

加丰富和差异化的局部中层特征, 建立高效的行人模型。该算法细化了卷积神经网络的检测部位, 增加了行人的局部位置检测和识别, 有效提升了对部分遮挡行人的检测率。同时, 相比常规全局损失函数, 栅极损失函数虽然增加了 $N-1$ 层的权重和偏置参数, 但因为权重 ω 在全局和局部分类器中共享, 且全局权重向量是由局部权重向量串联起来组建的, 全局偏置也是通过加和局部偏置来获得, 因而没有额外的增加整体计算时间。

2.2.2 卷积神经网络的结构

本文在行人数据集上对神经网络进行反复迭代训练并调整权值参数后确定了针对提取行人特征的网络结构。该网络结构包含了2个 5×5 卷积层, 分别有64和128个卷积核, 1个 4×4 池化层, 在每层卷积后利用纠正线性单元(rectified linear unit, ReLU)^[12] 函数激活, 增加模型训练时的收敛速度。选用高斯零均值标准差为0.01的随机数初始化权重, 在卷积层后加入约束上限为0.1的弃权算法(dropout)^[13], 将不满足权值 L_2 范数上限约束的隐含层节点舍弃, 以减弱神经元节点间的联合适应性, 增强网络的泛化能力; 引入栅极损失函数作为神经网络的损失函数来增加检测对遮挡的鲁棒性, 从PETS2009和UCSD原始数据集中随机抠取大小为 64×64 的行人局部子图像块, 用以训练卷积神经网络。本文采用的卷积神经网络框架如图4所示。

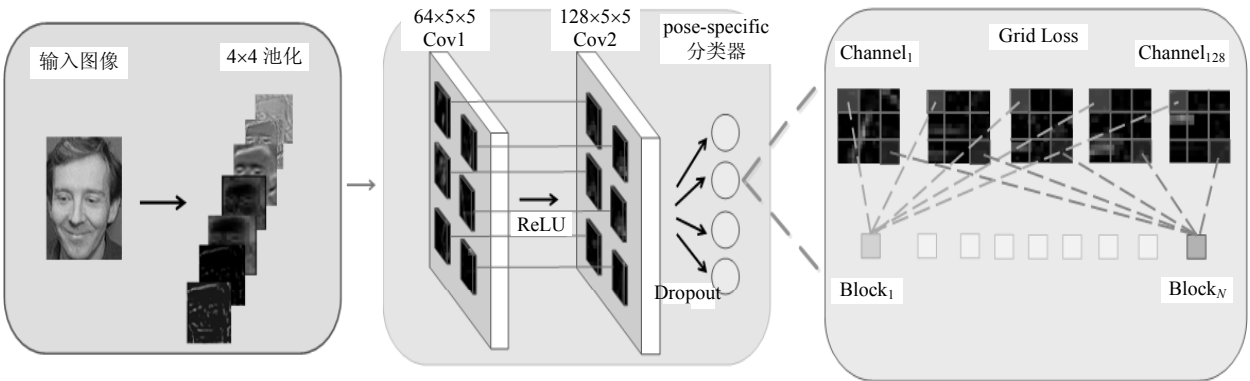


图4 行人识别卷积神经网络框架图

2.2.3 分类器的选择

算法的最后一个步骤是将深度神经网络提取的特征投入到分类器中进行判别, 以确定该输入区域是否为行人样本。本文将传统的SVM分类器改为基于局部的姿势分类器(part-based pose-specific classifiers)^[14], 以配合栅极损失函数检测出的行人脸或躯干等局部区域, 以提高网络的分类精度, 取得了良好的效果。

在网络训练结束后, 即可使用该神经网络算法生成的模型对低密度行人团块进行检测, 以得到该团块中人数。

3 密集场景中人群特征回归算法

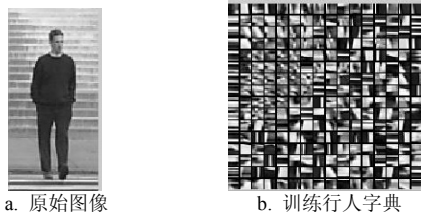
当在行人数量较多的密集场景下, 人与人之间的遮挡重叠现象会很严重, 通过检测每个行人来计数的方法将不再适用, 这时就需要寻求回归的方法

来解决这个问题。研究表明,低密度和高密度人群在图像纹理特征上表现为粗、细两种不同模式。因此对于密集人群团块,可通过分析图像纹理特征来获得相应人群的信息。灰度共生矩阵GLCM(gray-level co-occurrence matrix)是常用的纹理特征之一^[15],一般计算其0°、90°和135°方向上的对比度、熵、方差等特征,然后结合主成分分析进行选取。最后将提取的多维特征采用核函数建立回归模型^[16]以估计行人数量。为了更加精确描述人群分布信息,本文提取了场景中的行人密度分布特征并将其融合至纹理特征中,建立从数学特征到实际人数的映射模型,取得了良好的统计效果。

3.1 行人密度分布特征

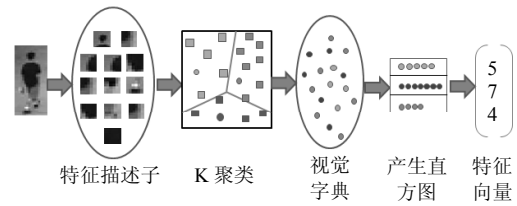
文献[17]提出了一种通过统计图像中目标像素值来生成密度图的方法。在训练中设置描述点来标注目标,使目标的每个像素都有一个实值特征向量来描述其局部外观,在本文中,这些特征用SIFT视觉词典定义,再选择一个线性映射,将每一个像素的特征向量变换为密度值,该点的密度值大小通过像素密度函数值得到。然后采用线性映射进行系数优化,以便使映射出的密度函数与真实密度函数尽可能一致,并对映射系数采用二次正则化约束,防止过拟合。学习完成后,只需要提取每个像素的特征向量乘以学到的对应线性系数即可得到该点密度值,并在任意形状的区域中进行密度集成,即可得到行人的密度分布图。

在构建行人密度分布特征图的过程中采用了图像检索中构建行人视觉词典的算法,训练中设置了特征描述子把图像中的行人描述为一组关键点的集合,使用K-means聚类算法对局部特征描述子聚类,聚类的中心是视觉词汇,将聚类中心特征向量化表示,生成对应码字。各个视觉词汇组成视觉词典,对应码字集合组成码书,即构成视觉词典,词典容量的大小由所含单词数量决定。行人的每个局部特征均会被对应到视觉词典中某个词汇,这种映射关系通过衡量局部特征和词汇间的欧氏距离来实现,然后通过直方图统计词典中视觉词汇出现的频次,生成与原矩阵大小相同的特征描述向量,即词袋,图5描述了行人视觉词典的构建过程。



a. 原始图像

b. 训练行人字典



c. 行人特征向量生成步骤

图5 行人词典构建方法

为了方便将特征向量值与行人密度分布的对应关系直观表达出来,本文将其转换为可视化的密度特征图。使用文献[18]定义的点密度分布函数PSF(point spread function),将行人像素点的特征向量值转换为点密度数值,公式如下:

$$\text{PSF}(x, y; p, q) = \exp\left(-\frac{0.5}{\sigma_g^2(p, q)}((x-p)^2 + (y-q)^2)\right) \quad (3)$$

式中, (p, q) 表示要计算的像素点所在位置; (x, y) 为其周围邻域的点; $\sigma_g^2(p, q)$ 为高斯核; σ_g^2 代表了点 (x, y) 周围的行人特征点的密集程度,即邻域像素点对应的特征向量值的大小。转换完成后将不同大小的密度值使用不同深浅的色度——对应构成行人密度分布图,图6为在UCSD数据集上建立的行人密度分布特征图。



图6 行人密度分布特征图

由式(3)将行人图像中每个像素的特征向量值转换为密度分布函数值,因此该密度分布函数值可以与纹理特征融合构成人群密度分布特征。

3.2 多核回归函数

核函数的选取需要根据数据的分布信息和规律来选取,使用单一的核函数显然不能精确地对数据特征进行描述。本文描述行人的纹理特征以及密度分布特征入手,从不同方面提取多层次多类型的特征描述子,这些特征分别独立提供了行人的信息,因此核函数的选择需要适应多种特征。本文采用 l_p

准则多核回归函数^[19]来实现从多类数学特征到实际人数的估计。 l_p 准则多核支持向量机的数学模型可表达如下:

$$\min_{\omega, b, \theta, \theta \geq 0} C \sum_{i=1}^n V \left(\sum_{m=1}^M \omega_m \cdot \phi_m(X_i) + b, y_j \right) + \frac{1}{2} \sum_{m=1}^M \frac{\|\omega_m\|^2}{\theta_m} \quad (4)$$

$$\text{s.t. } \|\theta\|_p \leq 1$$

式中, $V(f(x), y) = \max\{0, 1 - yf(x)\}$; $\|\theta\|_p = \left(\sum_{m=1}^M \theta_m^p \right)^{1/p}$ 。C为常数, 表示惩罚系数; X_i 为特征向量; b 为标量; 代表偏移量; ω_m 代表垂直于分类超平面的法向量; θ_m 为多核加权值; M 是核的数目; n 为训练样本数。解式(4)即寻找最优分类面 $\sum_{m=1}^M \omega_m \phi_m(X_i) + b = 0$ 。

本算法提取了密集人群的纹理特征, 实验选择了300幅PETS2009数据集中的行人场景图像, 当 θ 取 0° 、 90° 、 135° 方向时计算每个密集团块的灰度共生矩阵, 并求出对比度、熵、能量3个特征值, 构成大小为2 700的数据样本矩阵。经过主成分分析后, 选取了包含信息量最多, 对图像纹理特征贡献最大的水平、竖直、对角线方向的对比度特征作为人群纹理特征, 并与密度分布特征融合构成描述密集人群分布信息的4维特征指标。在多核回归函数中, 针对算法提取的不同行人特征种类设置了4个核函数, 经过训练后输出核函数的多核加权值, 建立起从人群分布特征到实际人数的映射, 完成密集场景下的人数统计。

4 实验结果与分析

4.1 实验设置

为了验证所提的结合卷积神经网络和轨迹预测的行人统计算法在复杂环境下的有效性, 本文在UCSD和PETS2009监控视频中获取的图像序列上进行了实验, 这两个行人数据集中的人群密度各不相同且分布具有随时间变化的性质, 此外摄像头采集视频的角度各不相同, 是理想的用以测试算法精度的数据集。此外, 本文还利用自己采集的雾天行人视频数据集来验证算法对光照和恶劣环境的适应性。实验结果验证了所提算法的有效性和鲁棒性。

实验的硬件平台为i7×10核, 3.4 GHZ工作站, 内存为64 GB, 显卡为GTX-Titan X, 内含GPU, 系统为Ubuntu14.04, 卷积神经网络在Caffe开源框架下实现。文献[20]中方法的对比实验在MATLAB 2015b

环境下实现, 文献[21]方法的对比实验在Visual Studio 2010环境下实现, 并调用OpenCV库函数。

UCSD数据集的图片分辨率为 258×158 , 如图7所示; PETS2009数据集的图片分辨率为 768×576 , 如图8所示。为了保证卷积神经网络训练和学习的有效性, 对两个场景数据分别抽取1 000幅图像采样归一化为 256×128 统一大小, 手工标注其中的500张并采用本文算法构建人群密度分布图特征, 用以训练回归模型。之后从这500幅已标注的图像中随机抽取1 000张 64×64 规格的子图像, 用以训练栅极损失函数提取行人局部特征, 用剩下500幅图像进行测试。

4.2 实验结果与分析

算法在UCSD、PETS2009数据集上的测试效果如图7和图8所示, 在自己采集雾天数据集上测试效果如图9所示, 实验结果证明所提算法在不同场景下均具有准确的人数统计精度。

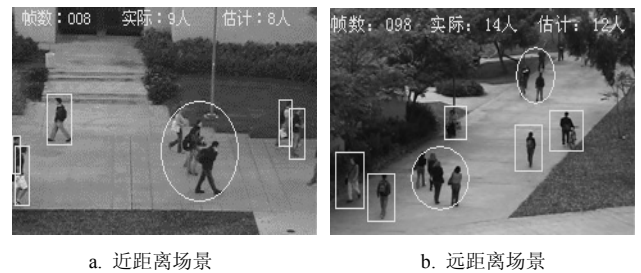


图7 UCSD数据集测试效果



图8 PETS2009数据集测试效果 图9 雾天数据集测试效果

通过图7a和图8可以看出, 本文算法识别出了最左侧、最右侧被遮挡住的行人, 当行人相互重叠时常规算法无法检测出整体行人, 但采用本文所提出的栅极损失函数以及分块训练卷积神经网络的方法后, 即可针对遮挡行人局部进行检测识别统计。在雾天数据集上的实验结果表明, 所提Retinex去噪算法和转换颜色空间算法消除了场景中雨雾和光照度不足等因素对人数统计的干扰。对密集人群团块, 提取团块灰度共生矩阵的水平、竖直和对角线方向的对比度作为人群纹理特征指标, 融合3.1节中构建出如图6的人群密度分布特征, 使用多核回归函数进行估计。最终, 加上不同密度团块中人数即为场景中统计出的总人数。算法在UCSD和PETS2009数据

集上的统计结果如图10、图11所示。

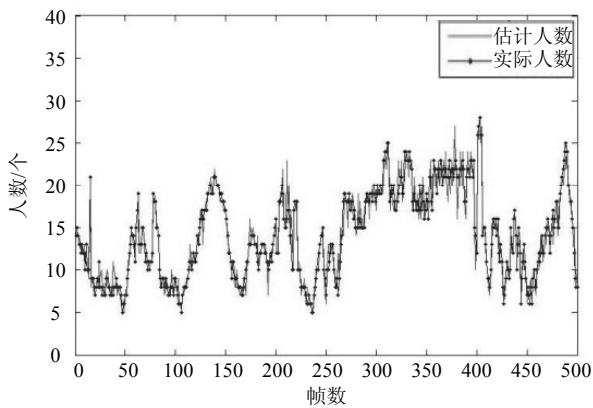


图10 UCSD数据集统计结果

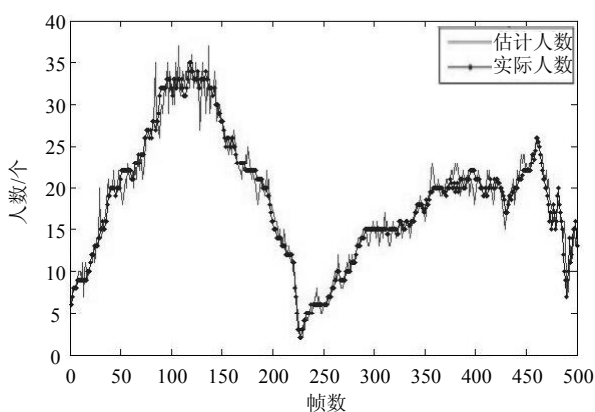


图11 PETS2009数据集统计结果

行人统计的评价指标用准确性衡量，其中准确性使用平均绝对误差(mean absolute error, MAE)和平均均方误差(mean squared error, MSE)和计算。表2和表3是现有的其他算法和本文算法在这两个数据集上测试的比较结果。

表2 UCSD数据集对比实验指标评价

算法	MAE	MSE
Cumulative Attribute Regression ^[20]	2.07	0.08
clustering motion cues ^[23]	2.97	0.12
文献[7]	1.60	0.10
本文算法	1.40	0.07

表3 PETS2009数据集对比实验指标评价

算法	MAE	MSE
文献[22]	4.00	0.24
clustering motion cues ^[23]	2.78	0.21
Head detection ^[21]	2.40	0.12
本文算法	2.10	0.08

通过对比实验结果图可以得出，相比较近年人数统计领域内其他学者提出的算法，本文算法在PETS2009和UCSD数据集上的测试结果在准确率上

有显著的提升，相比较文献[7]的研究结果，精度提高了12%，验证了所提算法的有效性。可以看出当人群密度由低到高变化时，使用本算法得到的人数统计结果均与场景中真实人数相差较小，折线图位置较接近，证明了本文构建的密度分布图有良好的描述人群分布信息的能力，增加了特征回归模型的精度。在不同数据集上进行测试时算法性能并没有随之下降或改变，也证明了算法对不同密度的行人场景均有适应性，能应对较为复杂的环境和人群分布情况。相比文献[21]的算法，在大规模人群场景中行人形状轮廓不明显，难以剔除周围行人造成的干扰，不适用于人头检测的算法；反之在低密度人群场景中，稀疏行人并不具备纹理密度等概率分布特征，此时识别检测算法准确迅速，其精度优于文献[23]中概率估计算法。

5 结束语

在行人监控视频中，由于行人遮挡、场景光照变化、人群分布不均等因素的影响使得现有算法难以准确统计视频中人数。针对以上问题，本文将场景中分布不均的行人根据密度划分并提出基于卷积神经网络识别和密度特征回归相结合的人数统计算法。为了避免光照变化和雨雾天气对算法造成干扰，本文将场景去噪增强处理后，转换到HSV颜色空间中预判行人位置并提取特征。提出了栅极损失函数分块训练卷积神经网络的算法，实现了对遮挡行人局部位置的识别。提出了融合行人密度分布特征的回归算法以增加统计的精度。

实验证明在相同场景下，本文所提方法优于其他同类方法，但本文未完全考虑距离远近对团块大小的影响以及非行人目标出现时的情况，在未来的工作中将寻求更优的方法来提升算法性能。

参 考 文 献

- [1] ANTONINI G, THIRAN J P. Counting pedestrians in video sequences using trajectory clustering[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2006, 16(8): 1008-1020.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE Conference on Computer Vision & Pattern Recognition. [S.l.]: IEEE, 2005.
- [3] YANG S, LIAO X, BORASY U K. A pedestrian detection method based on the HOG-LBP feature and gentle AdaBoost[J]. International Journal of Advancements in Computing Technology, 2012, 4(19): 553-560.
- [4] FORSYTH D. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(9): 1627-45.

- [5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems. Doha, Qatar: Curran Associates Inc, 2012: 1097-1105.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Ohio, USA: IEEE, 580-587.
- [7] ZHANG C, LI H, WANG X, et al. Cross-scene crowd counting via deep convolutional neural networks[C]//Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2015: 833-841.
- [8] RAHMAN Z U, JOBSON D J, WOODDELL G A. Retinex processing for automatic image enhancement[J]. Human Vision and Electronic Imaging VII, 2002, 13(1): 100-110.
- [9] OPITZ M, WALTNER G, POIER G, et al. Grid loss: detecting occluded faces[M]//Computer Vision – ECCV. [S.l.]: Springer International Publishing, 2016.
- [10] UIJLINGS J R R, SANDE K E A V D, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
- [11] GEUSEBROEK J M, VAN D B R, SMEULDERS A W M, et al. Color invariance[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(12): 1338-1350.
- [12] ANDREW L M, AWNI Y H, ANDREW Y N. Rectifier nonlinearities improve neural network acoustic models[C]//Proceedings of the 30th International Conference on Machine Learning. Atlanta, Georgia, USA: IMLC, 2013.
- [13] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [14] VILLAMIZAR M, GRABNER H, MORENO-NOGUER F, et al. Efficient 3D object detection using multiple pose-specific classifiers[C]//Proceedings of the British Machine Vision Conference. Dundee, Scotland, UK: BMVA Press, 2011.
- [15] CHAN A B, VASCONCELOS N. Counting people with low-level features and Bayesian regression[J]. IEEE Transactions on Image Processing, 2012, 21(4): 2160-2177.
- [16] 薛陈. 复杂场景下的人数统计系统[D]. 天津: 天津大学, 2012.
- XUE Chen. People counting system in complex scenario[D]. Tianjin: Tianjin University, 2012.
- [17] LEMPITSKY V S, ZISSERMAN A. Learning to count objects in images[C]//Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc, 2010: 1324-1332.
- [18] LEHMUSSOLA A, RUUSUVUORI P, SELINUMMI J, et al. Computational framework for simulating fluorescence microscope images with cell populations[J]. Medical Imaging IEEE Transactions on, 2007, 26(7): 1010-1016.
- [19] KLOFT M, BREFELD U, SONNENBURG S. lp-Norm multiple Kernel learning[J]. Journal of Machine Learning Research(S1533-7928), 2011, 12(3): 953-997.
- [20] CHEN K, GONG S, XIANG T, et al. Cumulative attribute space for age and crowd density estimation[C]//Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2013: 2467-2474.
- [21] SUBBURAMAN V B, DESCAMPS A, CARINCOTTE C. Counting people in the crowd using a generic head detector[C]//IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance. [S.l.]: IEEE, 2012: 470-475.
- [22] CONTE D, FOGGIA P, PERCANNELLA G, et al. A method for counting moving people in video surveillance videos[J]. EURASIP Journal on Advances in Signal Processing, 2010, doi: 10.1155/2010/231240.
- [23] RAO A S, GUBBI J, MARUSIC S, et al. Estimation of crowd density by clustering motion cues[J]. The Visual Computer, 2015, 31(11): 1533-1552.