



研究与开发

基于深度时空特征卷积—池化的视频人群计数方法

李强, 康子路

(中国电子科技集团公司信息科学研究院, 北京 100086)

摘要: 由于摄像机角度、背景、人群密度分布和遮挡的限制, 传统的基于底层视觉特征的视频人群计数方法往往难以实现理想的效果。利用视频的时空特征和卷积—池化方法形成高层的视觉特征, 采用局部特征聚合描述符进行量化和码本计算, 实现了对视频人群信息的精准描述; 该方法充分利用了视频的运动和外观信息, 基于卷积神经网络和池化方法提升了对视频本征属性和特征的描述能力。实验结果表明, 所提方法比传统的视频人群计数方法具有更高的精度和更好的顽健性。

关键词: 人群计数; 卷积神经网络; 深度时空特征; 卷积—池化

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018161

Video crowd counting method based on conv-pooling deep spatial and temporal features

LI Qiang, KANG Zilu

Information Science Academy, China Electronics Technology Group Corporation, Beijing 100086, China

Abstract: Due to angle of camera, background, population density distribution and occlusion limitations, traditional video crowd counting methods based on underlying visual features are often difficult to achieve ideal results. Using the temporal and spatial features of video and conv-pooling method, high-level visual features were formed, local feature aggregation descriptors were used for quantization and codebook calculation to achieve accurate description of video crowd information. This method made full use of video motion and appearance information. Based on convolutional neural networks and pooling methods, the ability to describe video intrinsic attributes and features was improved. Experimental results show that the proposed method has higher precision and better robustness than traditional video crowd counting methods.

Key words: crowd counting, convolutional neural network, deep spatial and temporal feature, conv-pooling

1 引言

视频人群计数具有重要的现实意义和应用价值, 可以为公众提供安全、预警服务, 还可以通过

对人的行为分析, 优化公共资源配置、提高服务质量等。视频人群计数问题已成为模式识别和智能视频处理领域的重要研究内容。近年来, 研究人员提出一系列人群计数方法, 这些方法大体可

收稿日期: 2017-11-23; 修回日期: 2018-04-18

以分为 3 类^[1]：基于检测（detection-based）的方法、基于聚类（clustering-based）的方法和基于回归（regression-based）的方法。基于检测的方法通过检测和跟踪视频中的人体或头部、上肢等人体部分实现人数统计，同时实现对个体的定位，但当人群密度较高或场景复杂时，识别效果往往不理想；基于聚类的方法建立在“人的运动轨迹或视觉特征相对规整”这一假设的基础上，用连续轨迹代表移动个体，通过轨迹聚类实现人群计数。高密度人群环境下的目标检测和识别属于非平凡问题（non-trivial problem）。与前两种方法不同，基于回归的方法忽略单一的个体，而将人群视为一个整体，通过人群模式的整体性描述估计人群密度，利用视频帧（图像）特征和人群之间的回归关系实现人群计数，避免对单一个体的跟踪识别，这类方法通常用于大规模人群计数。

人群计数方法的关键在于视觉特征的选择，常见的特征包括边缘形状特征、关键点（兴趣点^[2]、角点^[3]）特征、方向梯度直方图（histogram of oriented gradient, HOG）^[4]等，这些特征对人群的表征能力有限，加上人群遮挡、透视效果的影响，往往难以达到理想的效果。视频帧前景分割和分块也是人群计数方法中常用的前置计算步骤，而这种计算本身属于计算机视觉领域的难点问题，计算精度直接影响人群计数的准确性。

充分利用视频中的时域和空域特征，并进行高层次抽象有助于实现对视频内容的准确描述。因此，本文提出一种基于深度时空特征卷积—池化（conv-pooling）的视频人群计数方法，该方法使用两种深度卷积网络（deep ConvNet）VGG-net^[5]和 3D ConvNet^[6]分别提取视频数据的时域（spatial）、空域（temporal）和时—空（spatial-temporal）特征构建特征图（feature map），对特征图池化（pooling），使用局部特征聚合描述符（vector of locally aggregated descriptor, VLAD）^[7]

量化编码以获得对视频人群数据的精准表达。传统算法与本文算法流程对比如图 1 所示。本文提出的深度时空特征卷积—池化的计算流程如图 2 所示。与传统方法相比，本文的贡献主要有以下 3 点：提取代表视频运动（motion）和外观（appearance）信息的时域和空域特征，通过卷积—池化方法获取的高层特征相比于底层特征实现对视频的更精准描述；提出的方法不依赖于前景分割，有效避免了传统方法中使用视频分割作为前置处理所带来的不准确的问题；在 UCSD 和 Mall 两个数据集上的实验结果表明，本文方法相比于传统基于视频的人群计数方法具有更高的准确率和更好的顽健性。

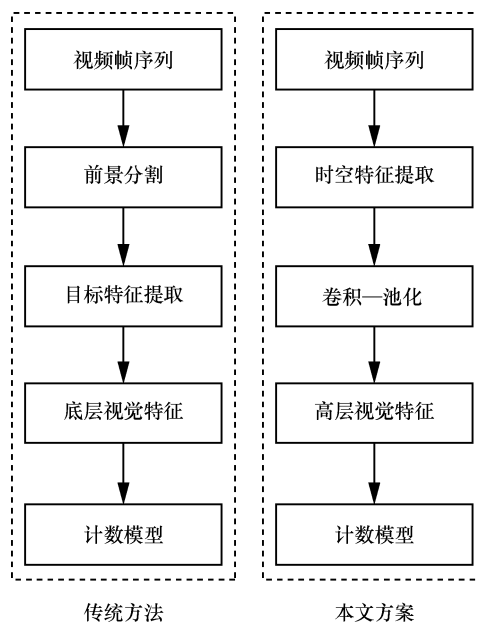


图 1 传统算法与本文算法的流程对比

2 相关研究

视频特征提取是影响人群计数方法的关键。例如，在基于检测的方法中，在低密度情况下使用 Haar 小波特征^[8]、edgelet 特征^[9]和 shapelet 特征^[10]等检测完整的行人，在高密度、存在大量遮挡情况下使用几何形状轮廓、3D 形状检测行人头部、肩部等；在使用特征的同时，利用线性分类

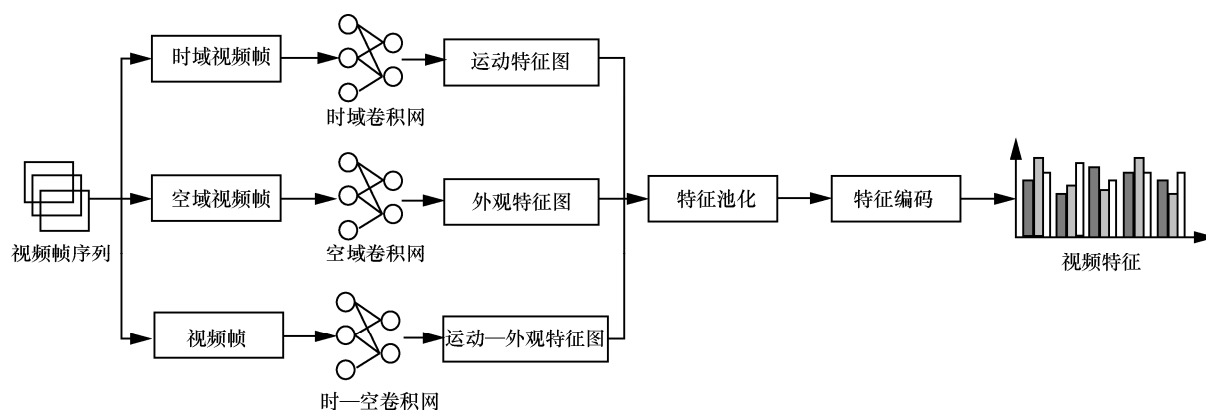


图2 本文提出的深度视频特征卷积—池化方法

器,如支持向量机、随机数等^[11]提升计算效果。基于回归的方法通过在人群特征和视频帧之间建立映射关系实现计数,这些特征包括背景、HOG、局部二值模式(local binary pattern, LBP)^[12]、灰度共生矩阵(gray level co-occurrence matrix, GLCM)^[13]等。

深度学习(deep learning)^[14]自提出以来,由于其优异的特征表达和模型泛化能力,在语音识别、计算机视觉等领域得到广泛应用;它通过多层神经网络和非线性特征结构对底层特征进行抽象转换,利用深层次模型获取高层特征。卷积神经网络(convolutional neural network, CNN)将深度学习的思想引入神经网络中,模拟人脑视觉皮层和神经元工作机理,通过卷积—池化运算进行特征提取和采样,使得特征的描述和判别能力更强,对平移、缩放和旋转等具有很好的顽健性。

近年来,学者们利用 CNN 提出一系列人群计数方法。Zhang 等人^[15]提出一种跨场景人群计数方法,利用人群密度和人群总数交替训练实现人数估计,该方法表明,与人工特征相比,深度学习得到的模型具有更好的描述能力;同时,为了更好地适应未知场景,提出了一种数据驱动的方法,在目标场景的基础上,对预训练获得的模型进行微调,使其能够适应新的场景,从而避免对模型的重复训练,实现跨场景计算;然而,该

方法在训练过程中需要大量使用难以获得的透视图(perspective map),这限制了其广泛适用性。参考文献[16]提出一种简单有效的多列卷积神经网络(multi-column convolutional neural network, MCNN)结构,将视频帧映射到人群密度图上;该方法允许输入大、中、小3种尺寸或分辨率的视频帧,每列 CNN 能够适应不同大小人体或头部的变化,具有良好的顽健性,并能在不需要输入视图的透视先验情况下通过几何自适应核(geometry-adaptive kernel)精确计算人群密度,该方法的主要作用是对人群密度进行估计,在常规密度情况下的计数效果与现有方法并没有明显区别。参考文献[17]提出一种基于序的空间金字塔池化网络的人群计数方法。该方法将原始图像分为不同尺度的子块,利用基于序的空间金字塔池化不同网络获得子图像块的人数,最后将所有子图像块人数相加得出图像总人数,这种方法在图像分块、池化操作时引入多个依赖特定场景的参数,使得算法的适应性不足。参考文献[18]通过深层网络(deep network)和浅层网络(shallow network)的融合,获取待检测图片的特征图(能量图、密度图)并进行积分获得人群计数结果,然而在计算两个特征图的过程中需要对大量的有标记数据进行训练,而这种标记数据往往是难以获得的。上述基于 CNN 的方法普遍建立在分层或分块的基础上,这种计算前置步骤产生的不

确定性会对后续计算产生难以预测的影响。本文所提方法无需前景分割,提取包含视频时、空要素的深度特征,相比传统方法能够获得更好的计算结果,下面将详细介绍计算流程。

3 深度时空特征处理

本节将给出卷积网中的卷积层选择(特征提取)、特征池化以及特征编码方法。

3.1 提取深度卷积网

在卷积网络中,不同深度的卷积层代表不同的信息,例如,全连接层(the fully-connected layer)代表了全局特征,如场景、目标等,而浅卷积层(shallower convolutional layer)则包含更多的局部特征,如线型、边缘等。

视频中主要包含运动和外观两种信息,其中运动信息依赖于时域特征,外观信息依赖于空间特征,而时空特征可以同时表达两种信息。下面将使用两种不同的 ConvNet 分别提取上述信息。

对于时域特征,使用 VGG-net。VGG-net 有两种深度结构:VGG-16 和 VGG-19,分别代表拥有 16 层和 19 层的 VGG-net,这里使用 VGG-19,原因在于更深层的卷积网络包含更多具有高区别性的局部深度特征。时域 VGG-19 的输入是光流簇(optical flow cluster),每一光流为一个视频帧的垂直和水平方向的运动信息,使用 OpenCV 中的 TVL1 算法^[19]抽取光流簇。提取 VGG-19 中的最深卷积层 pool5, pool5 的输出为一个大小为 7×7 像素和 512 个通道的特征图。为了获取特征图的局部深度特征,分别抽取每个空间位置信息并将 512 个通道的值直接连接,获得一个维度为 512 的局部深度特征向量。因此,对每个视频帧,都能够抽取 49 个深度时域特征。

对于空域特征,仍然使用 VGG-19,此时,它的输入是包含 3 个通道颜色信息的视频帧。使用新的输入数据重新训练 VGG-19,以获得新的网络初始化结构、更低的学习率、更好的增强泛化能

力、更高的丢弃率(防止过拟合)等。为了抽取视频帧的空间信息,依旧提取 VGG-19 中 pool5。空域 pool5 特征图的大小为 7×7 像素,包含 512 个通道,最后获得的局部深度特征由 pool5 特征图的每个局部特征连接而成,即 49 个深度空域特征。

对于时空特征,使用 3D ConvNet^[6],该网络使用 3D 卷积核同时捕获视频的外观和运动信息。3D ConvNet 的输入是长度为 16 帧的视频数据,通过视频迭代采样创建输入数据。由于 3D ConvNet 的最后一层包含空间信息的特征图的分辨率为 4×4 像素,为了与前面两个特征图大小相匹配,采用这一低分辨率特征图之前的一层,命名为 conv5。conv5 与之前两个特征图具有相同的大小和通道数量,但包含两个大小为 $7 \times 7 \times 512$ 的特征图,进一步通过 conv5 两个特征图中位置信息的最大值构建一个特征图,这使得深度时空特征图能够与前两个特征图保持一致。

在提取时域、空域特征之外再提取时空特征,最大的优点在于同时计算两种特征可以实现对视频数据的全局性表达,避免了传统方法中视频帧特征简单叠加导致的时域特征丢失问题,如图 3 所示。

3.2 池化深度卷积网

获得卷积特征图后,使用轨迹池化策略(trajjectory pooling strategy)对特征进行池化。轨迹池化策略建立在密集轨迹(dense trajectory)的基础上。为了提取密集轨迹,从多个尺度空间采样特征点。在密集光流场 $\omega=(u_t, v_t)$ 中,使用中值滤波追踪第 t 帧的点 $P_t=(x_t, y_t)$ 在 $t+1$ 帧的位置:

$$\begin{aligned} P_{t+1} &= (x_{t+1}, y_{t+1}) \\ &= (x_t, y_t) + (M * \omega) | \hat{x}_t, \hat{y}_t \end{aligned} \quad (1)$$

其中,“*”是卷积运算, M 是中值滤波器, $(\hat{x}_t, \hat{y}_t)$ 是 (x_t, y_t) 的近似位置。当计算得出密集光流场后,其后续帧的点轨迹就可以确定并连接形成轨迹 $(P_t, P_{t+1}, P_{t+2}, \dots)$ 。为了避免轨迹漂移问题,

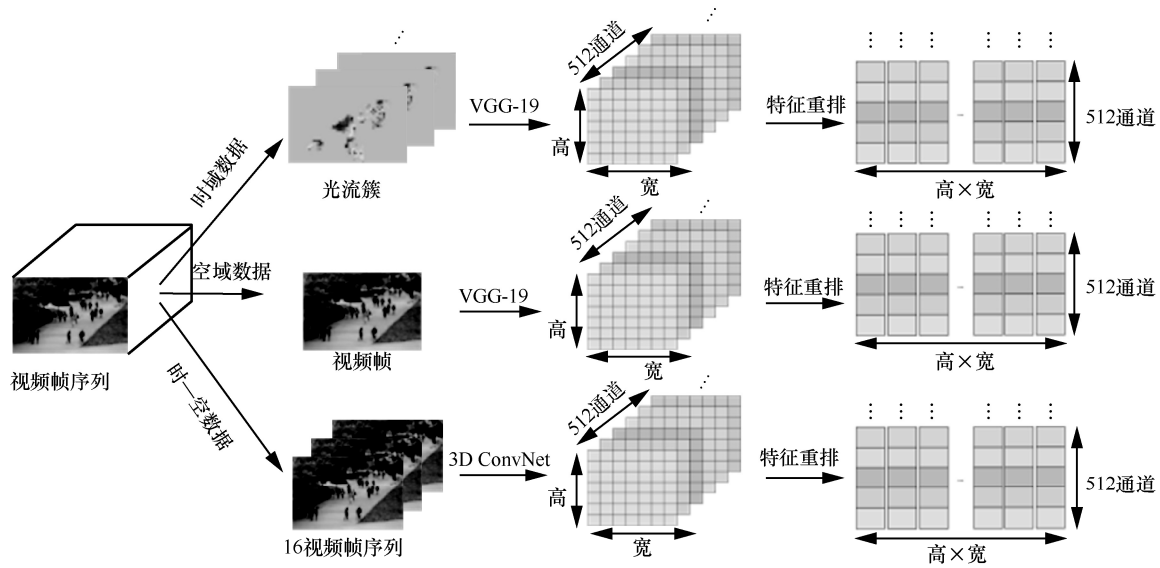


图3 提取深度卷积网

将轨迹的最大长度设置为 16 帧。在改进的密集轨迹方法中,忽略了光流中相机的运动信息以提升性能,它假定连续两个动态帧的全局背景是单应性矩阵(homography matrix)相关的。为了估计这一单应性,使用两种互补的匹配策略计算两帧之间的一致性: SURF^[20]特征匹配和光流匹配。通过 RANSAC^[21]算法纠正视频帧消除相机运动带来的影响来估计单应性。

提取的轨迹可以表示为:

$$T^k = \{(x_1^k, y_1^k, t_1^k), \dots, (x_{15}^k, y_{15}^k, t_{15}^k)\} \quad (2)$$

其中, T^k 是全部 K 个点中的第 k 个轨迹。为了使用轨迹池化卷积图,需要计算从视频特征点到特征图的映射比 α :

$$\alpha = \frac{H_f^l}{H_v} = \frac{W_f^l}{W_v} \quad (3)$$

其中, H_f^l 和 W_f^l 表示第 l 个卷积特征图的高度和宽度, H_v 和 W_v 表示视频帧的高度和宽度。从第 l 层的第 i 个轨迹描述符 TD_i^l 可以计算为:

$$TD_i^l = \sum_{i=1}^N C_i^l (x_i^k \times \alpha, y_i^k \times \alpha) \quad (4)$$

3.3 特征编码

VLAD 是一种有效的局部描述符的编码方

法。与 BOF (bag-of-feature) 类似^[7], 它首先使用 k 聚类 (k -means) 方法学习获得包含 K 个视觉单词 (visual word) 的码本 (codebook) $C = \{c_1, c_2, \dots, c_K\}$, 每个局部特征符都会用距其最近的视觉单词来表示。VLAD 的思想在于累积向量 x 和其最近邻 c_i 的差 $x - c_i$, 表示向量在视觉单词上的分布特征。

给定 $X = (x_1, x_2, \dots, x_N)$ 为池化的描述符, 从中心 k 计算的差分向量为:

$$\mu_k = \sum_{i:c_k} (x_i - c_k) \quad (5)$$

VLAD 通过级联获得 μ_k 在所有 K 个中心 $\mu = \{\mu_1, \mu_2, \dots, \mu_K\}$ 的表示, 使用幂律 (power) 和 2 范数归一化 (L2-normalization) 进行后续处理, 最后生成一个长的特征描述子。

VLAD 在考虑离其最近的聚类中心 (码本或单词) 的同时, 保留每个特征点到离其最近的聚类中心的距离, 并用所有聚类中心的线性组合表示特征点, 建模过程没有信息损失。

3.4 计算模型

上述步骤从 3 个不同维度对视频时空特征进行卷积—池化、特征融合、聚类计算以生成视频的特征表达, 最后使用线性支持向量机 (liner

SVM) 作为回归模型输出视频人群数目, 计算模型如图 4 所示。

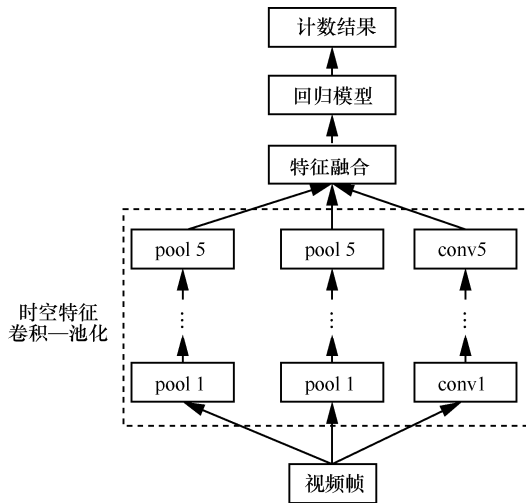


图 4 人群计数模型

4 实验结果与分析

使用平均绝对误差 (mean absolute error, MAE) 和均方误差 (mean squared error, MSE) [22] 评价性能, MAE 和 MSE 的定义为:

$$MAE = \frac{1}{N} \sum_1^N |G_n - E_n| \quad (6)$$

$$MSE = \sqrt{\frac{1}{N} \sum_1^N (G_n - E_n)^2} \quad (7)$$

其中, N 是总帧数, G_n 是第 n 帧的实际人数, E_n 是第 n 帧的预测人数。MAE 和 MSE 用来量化人群计数效果的客观标准, MAE 表示预测的准确性, MSE 表示预测的顽健性。通常, MAE 和 MSE 的值越小, 代表准确性和顽健性越好。

使用 UCSD 和 Mall 两个基准数据集, 两个数据集分别取自室外和室内场景。UCSD [16] 数据取自可俯瞰美国加州大学圣迭戈分校一条步行街的摄像头, Mall [1] 数据取自一个购物中心的监控摄像头。两个数据集的信息见表 1 和图 5, N_f 为视频帧数, R 为分辨率, FPS 为每秒帧数, D 为人群密度 (兴趣区域内最少和最多人数), T_p 为全部视频

帧中所含的总人数。

表 1 数据集属性

数据集	N_f	R	FPS	D	T_p
UCSD	2 000	238×158 像素	10	11~46	49 885
Mall	2 000	640×480 像素	< 2	13~53	62 325



(a) UCSD

(b) Mall

图 5 数据集样例

在 VLAD 编码中, 视觉单词 K 的数目决定了最终的特征维度。通过实验在 UCSD 和 Mall 两个数据集上寻求最好的 K , 见表 2, 这里, 准确度是指计数正确的视频帧数占全部视频帧数的百分比。当 $K=256$ 时, 生成特征的辨别能力最强, 因此实验中将 VLAD 中 K 的数量修改为 256。

表 2 VLAD 采用不同的聚类中心时的计数准确度

数据集	$K=32$	$K=64$	$K=128$	$K=256$	$K=512$
UCSD	47%	50%	52%	58%	55%
Mall	51%	57%	61%	66%	62%

对于 UCSD 和 Mall, 使用后 800 帧用于训练, 前 1 200 帧用于测试。将本文方法与已有的 3 个视频人群计数的方法 [16-17,22] 相对比, 结果见表 3。图 6 给出了 Mall 测试数据的计数结果, 可以看出, 估计人数与实际人数具有很高的相似度。

表 3 数据集对比实验

方法	UCSD		Mall	
	MAE	MSE	MAE	MSE
参考文献[16]方法	2.24	7.97	2.88	13.15
参考文献[17]方法	1.89	5.43	2.54	10.09
参考文献[22]方法	2.08	7.25	2.69	12.10
本文方法	1.81	5.25	2.38	9.29

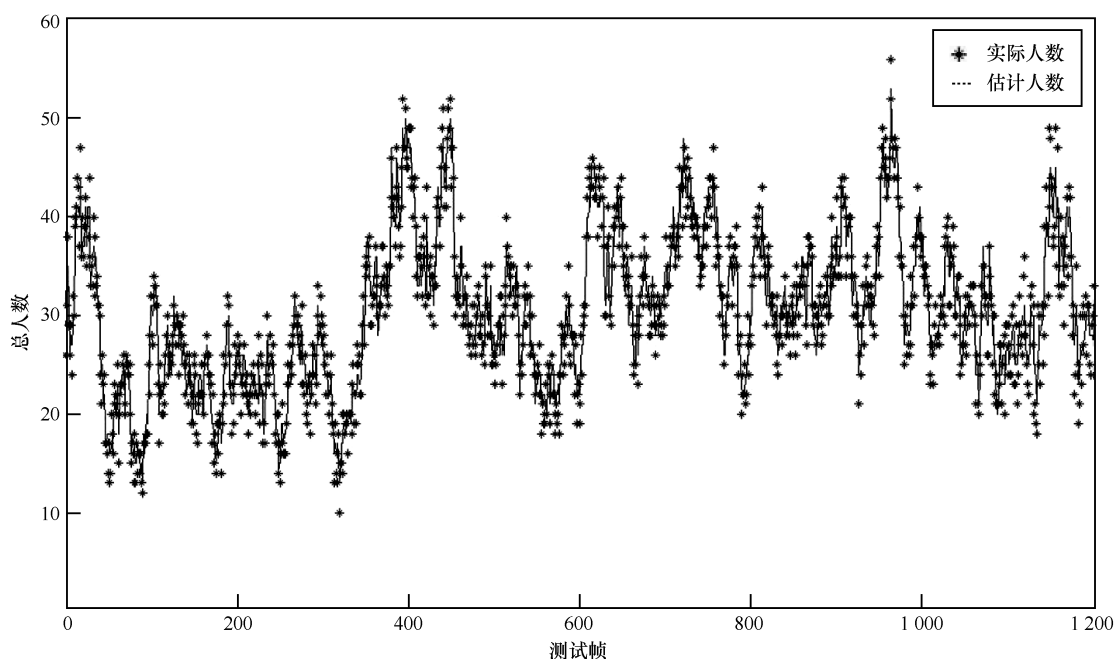


图6 Mall数据集的计数结果

表3给出不同方法在两个数据集上的对比结果,相比现有方法,本文的方法获得了更好的准确度和顽健性,原因在于:建立在前景分割基础上的人群计数方法受光照、人群密集程度、背景复杂度等的影响,对分块、面积计算等方法的使用也明显受限,准确度难以保证;特征提取是影响人群计数方法的重要步骤,提取视频的深度时空特征并对其进行有效编码是提升特征判别能力的有效方法,这是本文方法与对比方法最根本的区别。实验中给出使用不用特征时的MAE和MSE值,具体见表4。可以看出,使用时空特征能够更好地挖掘视频帧数据间表示视频时空要素发展变化趋势、规律以及本质属性的时空耦合、时空相关与时空异质特征,与单一特征相比具有更好的可靠性。本文中提取的深度时空特征具有更好的自主辨识能力和内容描述能力,对光照、背景变化、遮挡等的顽健性好;特征提取难度小,避免了前景分割这一复杂操作,且计算过程简单具有更好的适应性。

5 结束语

本文提出一种基于深度时空特征卷积—池

表4 数据集使用不用特征的结果

特征类型	UCSD		Mall	
	MAE	MSE	MAE	MSE
时域特征	2.67	8.52	3.37	14.70
空域特征	2.31	8.13	2.99	13.38
时空特征	1.81	5.25	2.38	9.29

化的视频人群计数方法,使用VGG-net和3D ConvNet分别提取深度时空卷积特征,使用轨迹池化策略进行特征池化,最后采用VLAD进行特征编码。与传统的视频人群计数方法相比,本文提出的方法在UCSD和Mall两个数据集上获得了良好的计数性能,这表明基于深度时空特征卷积—池化的方法能够更有效地描述视频的本质信息,这种特征提取、编码方法还可以扩展到更广的视频模式识别领域。未来将进一步训练和利用深度网络处理复杂的视频语义概念和内容、探索利用在不同卷积层上的池化方法。

参考文献:

- [1] LOY C C, CHEN K, GONG S G, et al. Crowd counting and profil-

- ing: methodology and evaluation[M]. New York: Springer, 2013: 347-382.
- [2] IDREES H, SALEEMI I, SEIBERT C, et al. Multi-source multi-scale counting in extremely dense crowd images[C]//The 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'13), June 23-28, 2013, Portland, OR, USA. Piscataway: IEEE Press, 2013: 2547-2554.
 - [3] ROSTEN E, PORTER R, DRUMMOND T. Faster and better: a machine learning approach to corner detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(1): 105-119.
 - [4] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//The 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, New York, NY, USA. Piscataway: IEEE Press, 2005: 886-893.
 - [5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014(9).
 - [6] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatio-temporal features with 3D convolutional networks[C]//The 2015 IEEE International Conference on Computer Vision (ICCV'15), Dec 7-13, 2015, Santiago, Chile. Piscataway: IEEE Press, 2015: 4489-4497.
 - [7] J'EGOU H, DOUZE M, SCHMID C, et al. Aggregating local descriptors into a compact image representation[C]//The 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10), June 13-18, 2010, San Francisco, CA, USA. Piscataway: IEEE Press, 2010: 3304-3311.
 - [8] VIOLA P, JONES M J. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
 - [9] WU B, NEVATIA R. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors[C]//The 2005 IEEE International Conference on Computer Vision (ICCV'05), Oct 17-21, 2005, Beijing, China. Piscataway: IEEE Press, 2005: 90-97.
 - [10] SABZMEYDANI P, MORI G. Detecting pedestrians by learning shapelet features[C]//The 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07), June 18-23, 2007, Minneapolis, Minnesota, USA. Piscataway: IEEE Press, 2007: 1-8.
 - [11] GALL J, YAO A, RAZAVI N, et al. Hough forests for object detection, tracking, and action recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(11): 2188-2202.
 - [12] HUANG D, SHAN C, ARDABILIAN M, et al. Local binary patterns and its application to facial image analysis: a survey[J]. IEEE Transactions on Systems Man & Cybernetics Part C Applications & Reviews, 2011, 41(6): 765-781.
 - [13] SULOCHANA S, VIDHYA R. Texture based image retrieval using framelet transform-gray level co-occurrence matrix(GLCM)[J]. International Journal of Advanced Research in Artificial Intelligence, 2013, 2(2).
 - [14] HINTON G E, SALAKHUTDINOV R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
 - [15] ZHANG C, LI H, WANG X, et al. Cross-scene crowd counting via deep convolutional neural networks[C]//The 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15), June 8-10, 2015, Boston, Massachusetts, USA. Piscataway: IEEE Press, 2015: 833-841.
 - [16] CHAN A, LIANG Z, Vasconcelos N. Privacy preserving crowd monitoring: counting people without people models or tracking[C]//The 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), June 24-26, 2008, Anchorage, Alaska, USA. Piscataway: IEEE Press, 2008: 1-7.
 - [17] 时增林, 叶阳东, 吴云鹏, 等. 基于序的空间金字塔池化网络的人群计数方法[J]. 自动化学报, 2016, 42(6): 866-874.
SHI Z L, YE Y D, WU Y P, et al. Crowd counting using rank-based spatial pyramid pooling network[J]. Acta Automatica Sinica, 2016, 42(6): 866-874.
 - [18] BOOMINATHAN L, KRUTHIVENTI S, BABU R. Crowdnet: a deep convolutional network for dense crowd counting[C]//The 2016 ACM Conference on Multimedia Conference (MM'16), Oct 15-19, 2016, Amsterdam, The Netherlands. New York: ACM Press, 2016: 640-644.
 - [19] ZACH C, POCK T, BISCHOF H. A duality based approach for realtime TV-L1 optical flow[M]. Berlin: Springer, 2007: 214-223.
 - [20] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
 - [21] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395.
 - [22] ZHANG Z X, WANG M, GENG X. Crowd counting in public video surveillance by label distribution learning[J]. Neurocomputing, 2015(166): 151-163.

[作者简介]



李强 (1984-), 男, 博士, 中国电子科技集团公司信息科学研究院物联网技术研究所工程师, 主要研究方向为视频/图像处理、模式识别、机器学习。



康子路 (1972-), 男, 中国电子科技集团公司信息科学研究院物联网技术研究所高级工程师, 主要研究方向为物联网、数据架构。