

基于卷积特征的非平衡人群密度估计方法

曲 佳 时增林 叶阳东

(郑州大学信息工程学院 郑州 450001)

摘 要 人群密度估计在智能监控领域具有重要的应用价值。大量理论和经验研究表明,基于数据驱动的深度神经网络往往优于传统的基于手工特征的方法。但是人群样本的数据规模很小,深层次的网络很难得到较优解。鉴于此,提出了 3 种解决方法:训练较浅的神经网络,使用预训练深度模型的全连接层特征和使用预训练深度模型的卷积-FV (Fisher Vector)特征。针对样本的不平衡性问题,提出了使用多个分类评估标准的解决方案。在标准数据集 PETS2009 上的实验结果表明,相比于现有的手工特征,卷积特征具有更好的效果。其次,相比于训练一个全新的卷积模型,基于迁移学习的深度卷积特征是更好的选择。另外,通过层数较少的深度模型获得的较低层特征的迁移性更好。

关键词 人群密度估计,深度卷积神经网络,迁移学习,支持向量机,纹理特征

中图法分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.08.042

Unbalanced Crowd Density Estimation Based on Convolutional Features

QU Jia SHI Zeng-lin YE Yang-dong

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract Crowd density estimation plays a central role in intelligent monitoring. Deep neural network usually outperforms conventional approaches based on manual features owing to its data-driven superiority. However, deep neural networks are still far from optimal solution because of the scarceness of large-scale datasets. To address this problem, this paper investigated the feasibility of several solutions which are training shallow neural network from scratch, using fully connected layer features of pretrained deep neural network and aggregating convolutional features by way of fisher vector(FV). Aiming at the problem of unbalanced distribution, this paper further proposed several classification evaluation criteria. Comprehensive experiments were carried out on benchmark PETS2009 dataset. Results show that convolutional features outperform existing hand-crafted ones. Moreover, utilizing deep convolutional features based on transfer learning usually leads to better performance than the models trained from scratch. Finally, simpler pretrained models such as AlexNet can generalize better mobility of the lower layer features than more complicated ones such as VGGNet.

Keywords Crowd density estimation, Deep convolutional neural network, Transfer learning, SVM, Texture feature

近年来,参加庆典、集会等公共活动的人越来越多,所造成的公共安全问题也逐渐受到人们的重视。上海外滩踩踏事件和云南昆明明通小学踩踏事件等的发生都是由于人们没有足够地察觉或预防局部人群密度过高的地点。因此建立大规模公共活动监控系统,自动识别具体场景的人群密度,当某个局部场景人群密度超过失控阈值时能够自动报警,对公共安全具有重要意义^[1-2]。

在计算机视觉和智能监控领域,研究者们已经做了大量的相关研究,提出了一些有效的解决方法^[1-8]。这些方法总体上可以分为两类:基于像素统计的人群密度估计^[3-4]和基于纹理分析的人群密度估计^[5-8]。第一种方法首先需要进行背景分割,然后利用人群前景和人群数目所占像素的线性关系进行人群密度估计。这种方法简单、实时性强,但是对于背景复杂、人群密度高和遮挡严重的场景则易产生较大误差^[6]。第二种方法是把人群看作一种特殊的纹理,将人群密度估计问

题转化为纹理识别问题:假定纹理模式较粗的人群图像的人群密度等级较低,纹理模式较细的人群图像的人群密度等级较高^[5]。根据这个特点,可提取人群图像的纹理特征来进行人群密度估计。

基于纹理分析的人群密度估计方法的关键是提取有效的纹理特征。当前常用的纹理特征提取方法有:GLCM^[6] (Gray-Level Co-Occurrence Matrix), LBP^[7-8] (Local Binary Pattern)和小波变换^[9] (Wavelet Transform)。Maran 等^[10]首次将 GLCM 用于人群密度估计,从全局图像中提取特征,针对全局图像,给出一个整体的密度等级估计。Wu 等^[5]将原始图像分成许多个子单元,对每个子单元提取 GLCM 特征,进而得到子单元的密度级别,然后综合所有子单元的密度级别得到原图像的密度级别。针对纹理分析, LBP 是一个很有效的描述子并且已经被广泛地应用在人群密度估计之中。例如, Ma 等^[7]提出了一种 DH-LBP 用于人群密度估计;为了更

到稿日期:2017-06-28 返修日期:2017-08-13 本文受国家自然科学基金(61170223, 61502434),河南省科技攻关项目(172102210011)资助。

曲 佳(1993—),女,硕士生,主要研究方向为深度学习、计算机视觉;时增林(1992—),男,硕士生,主要研究方向为机器学习、深度学习;叶阳东(1962—),男,博士,教授,博士生导师,CCF 高级会员,主要研究方向为机器学习、智能系统, E-mail:ieydy@zzu.edu.cn(通信作者)。

好地描述行人的动态纹理,Yang 等^[1]提出了 SST-LBP;Wang 等^[8]提出了一种融合 LBP 与 GLCM 的方法,提出在 LBP 图上而不是在原始图像上计算 GLCM,充分利用了这两种方法在提取纹理特征时的优点。由于小波变换在纹理识别问题上有很好的表现,Marana 等^[9]将该方法应用于人群密度估计问题,并取得了很好的估计效果。

基于纹理分析的密度估计方法在高密度人群密度估计问题上有很好的效果。然而,当存在透视现象、人群密度分布不均匀和人群背景较复杂等问题时,对中低密度人群估计的误差较大。一个好的人群特征应该能够描述或刻画人群的本质。深度学习能够直接利用大数据自动学习数据的特征,具有很强的数据描述能力,避免了人工设计特征存在的局限性^[11-12]。卷积神经网络(Convolutional Neural Network, CNN)是第一个真正意义上的深度学习模型,也是最成功的深度模型之一。研究者们已经提出了一些优秀的 CNN 模型,如 AlexNet^[13],VGGNet^[14],GoogleNet^[15]等。这些模型已经被广泛应用到图像识别、场景分类、目标检测和图像检索等计算机视觉问题中。考虑到人群密度数据的规模很小,本

文提出一个较浅的 CNN 模型用于人群密度估计。

大量的研究表明,在大规模数据集上训练过的 CNN 模型可以作为通用的特征描述符使用。这种特征描述符具有较好的可分性和迁移性,并对类内变化具有较好的鲁棒性^[16-18]。为了更有效地解决复杂类内变化条件下的小样本人群密度估计问题,提出将深度卷积特征作为输入,采用支持向量机作为分类器进行人群密度分类。基于迁移学习的思想,通过两种方式获得深度卷积特征:1)直接将预训练深度模型的全连接层值作为人群特征;2)将预训练深度模型的卷积层值作为初始人群特征,然后采用 Fisher Vector(FV)生成最终的人群特征。该方法的流程如图 1 所示。准确率通常被用来评价一个分类器建立的分类模型的好坏。不同人群密度等级的样本数据易出现分布不平衡的现象,并且在非平衡数据集的学习中,使用准确率评估时往往更侧重于少数目标类的正确分类率,因此仅使用准确率进行评估并不合适。为解决此问题,本文引入了以下评估标准:召回率(Recall)、精度(Precision)、F-measure 和 MCC^[19]。

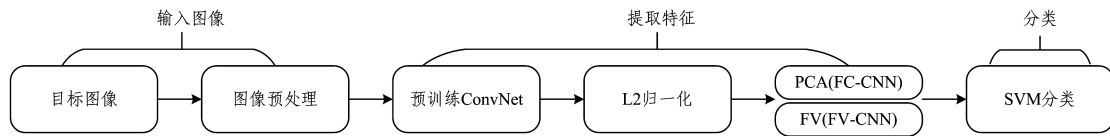


图 1 本文方法流程图

Fig. 1 Flowchart of proposed method

相比于已有的方法,本文的贡献主要有以下几点:1)考虑到人群样本数据的规模太小,提出一个较浅的 CNN 模型用于人群密度估计;2)提出基于迁移学习的思想,将深度卷积特征用于人群密度估计问题,提出使用预训练深度卷积模型全连接层的值和将卷积层与 FV 相结合两种方式来获取深度特征;3)考虑到人群密度样本数据分布不平衡的问题,提出使用多个分类评估标准的解决方案。在标准数据集 PETs2009 上的实验结果表明,所提方法在多个评估标准上都优于已有方法。

1 人群密度估计的卷积方法

与已有的基于像素统计的人群密度估计方法和基于纹理分析的人群密度估计方法不同,本文提出了一种基于卷积特征的人群密度估计方法。由于人群样本数据规模较小,无法训练一个有效的深度卷积模型,因此提出一个较浅的人群密度估计卷积模型。理论和经验研究表明,深度卷积特征在处理复杂问题时更有效。基于迁移学习的思想,本文提出两种获取和使用深度卷积特征的方法。

1.1 人群密度估计卷积模型

深度卷积神经网络是一种有监督的深度学习模型,已成为当前语音分析和图像识别领域的研究热点。Lecun 等^[20]早在 1998 年就提出了经典的 LeNet 网络,Hinton 等^[13]使用的大规模深度 CNN 在 Imagenet 的 1000 类的分类问题上取得了非常好的结果。CNN 的核心是局部感受野、权值共享(或者权值复制)以及时间或空间亚采样。CNN 获得的特征具有较好的对类内变化的不变性。CNN 能够学习数据的多层表征,高层特征被认为更能反映数据的本质,这是以往手工选取特征和浅层学习特征所不具备的。大量研究成果表明,在图

像识别等大量应用领域中,相比于传统浅层特征,高层特征能够获得非常明显的性能提升。

理论和经验研究表明,卷积模型的层次越深,处理问题的能力就越强。然而,较深模型的学习需要大量的训练样本。当前存在的人群样本规模都很小,无法训练一个很深的卷积模型用于人群密度估计。因此,提出一个较浅的卷积模型。所提模型与 LeNet 网络比较相似,模型的网络结构如图 2 所示。网络包括 $6 \times 5 \times 5$ 和 $16 \times 4 \times 4$ 两个卷积层,每个卷积层后有一个 2×2 的最大池化层。在第二个最大池化层后有一个 $500 \times 1 \times 1$ 的全连接层。在全连接层使用了 Dropout 机制。使用 ReLU 非线性函数作为激活函数。

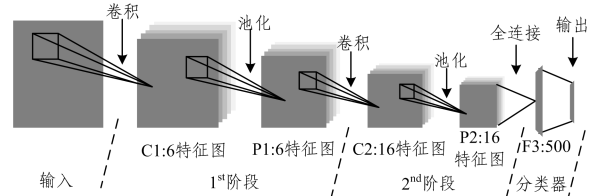


图 2 人群密度识别所使用的卷积神经网络结构图

Fig. 2 Architecture of convolutional neural network for crowd density estimation

1.2 人群密度估计的深度卷积特征

当数据规模比较小或者硬件能力较差时,无法从头开始训练一个有效的深度模型,但是可以通过迁移学习的方法使用深度神经网络来解决实际的问题。最近的研究表明,在大规模数据集上学习过的深度神经网络模型的较低层可以被视为朴素的特征提取器。由这种方式获得的特征已被证明是非常有效的。

通常使用预训练深度卷积模型的全连接层的值作为迁移特征(Fully-connected CNN-based Feature, FC-CNN Feature),然而通过这种方式获取迁移特征有两方面的不足:1)由于全连接层需要固定尺度的输入,因此需要对原始图像进行尺度调整,这会使原始图像信息有一定程度的损失;2)全连接层在深度模型中属于较深层,具有领域任务特殊性,因此其迁移性不如较浅层。使用卷积层和 Fisher Vector 获取的特征(Fisher Vector CNN-based Feature, FV-CNN Feature)可以有效地克服 FC-CNN 特征的问题。其思想是:将预训练深度卷积模型的最后一个卷积层作为原始特征,然后使用 FV 进行编码和聚合操作来构建一个无序的表示。卷积层不要求固定尺度的输入,因此不需要有代价性的尺度调整操作。FV-CNN 特征具有更强的迁移能力和人群表达能力。本文选用 3 种经典的深度 CNN 模型,即 AlexNet, VGGNet-16 和 VGGNet-19,来提取人群特征并比较它们的表现。

AlexNet 网络: AlexNet 是包含 5 个卷积层(conv1—conv5)和 3 个全连接层(fc6—fc8)的一个卷积神经网络模型,如图 3 所示(下方的数字表示同样的卷积层的数量,上方的文字是对该层参数设置的描述)。其相比于原始的 LeNet, AlexNet 拥有更多的层数,因此具有更强的学习能力。而且 AlexNet 使用了很多新技术,如 ReLU 激活函数、Dropout、数据集扩增(Data Augmentation)、局部响应归一化(Local Res-

ponse Normalization)、重叠池化(Overlapping Pooling)等,进一步增强了网络的学习能力。AlexNet 为以后其他新的优秀深度卷积模型的提出奠定了深厚的基础。

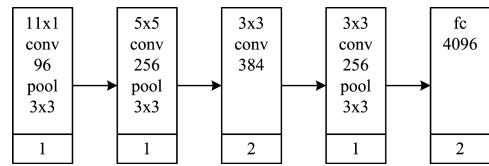


图 3 AlexNet 模型

Fig. 3 Architecture of AlexNet model

VGGNet 网络:相比于 AlexNet, VGGNet 有着更多的层数。VGGNet 使用几个小的感受野来模拟一个大的感受野,例如,用 2 个 3×3 的感受野模拟 1 个 5×5 的感受野,用 3 个 3×3 的感受野模拟 1 个 7×7 的感受野。这样做有 3 个方面的好处:1)增加了模型的深度,使模型具有更强的学习能力;2)使用小的感受野更能捕捉到目标的细节信息;3)降低模型的参数规模,更易于拟合,加快了训练速度。根据模型深度的不同, VGGNet 分为 VGGNet-16 和 VGGNet-19。这两个网络的模型结构如图 4 所示(下方数字表示同样的卷积层的数量,当该值取 2 时表示 VGGNet-16 模型,取 3 时表示 VGGNet-19 模型,上方的文字是对该层参数的设置)。

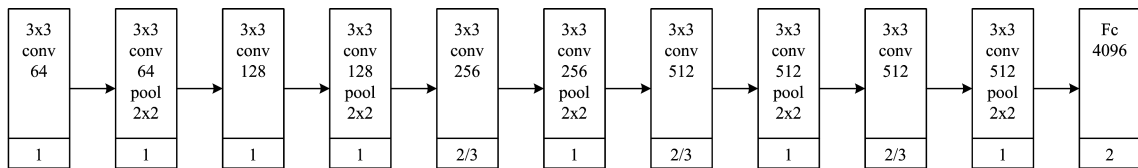


图 4 VGGNet 模型

Fig. 4 Architecture of VGGNet model

2 不平衡人群密度估计

训练数据的大小对于分类器的选择至关重要。人群密度估计是一个复杂的分类问题,需要大量的训练数据和泛化能力很强的学习算法。然而可供训练的人群密度数据集的规模非常小,并且通过深度卷积模型获取的特征向量的维度很高。鉴于 SVM 在解决小样本、非线性和高维模式识别问题时有着特有的优势,本文选择 SVM 作为分类器。

在真实的人群场景中,人群密度低和中等的情况比较多,而人群密度很高的情况比较少。因此,在获取用于人群密度估计的训练样本时,各个密度等级的样本数量差异较大,是非平衡的。当前的人群密度估计方法没有考虑人群样本的不平衡性,只关注整体的估计准确率而忽视了少数类样本的准确率。在非平衡数据集的学习中,不仅要关注整体的预测准确率,而且要侧重于少数的目标类的正确分类率。因此,仅仅使用整体的准确率来评估所提分类模型的优劣是不恰当的。针对此问题,本文提出使用能更加准确评估不平衡分类效果的评估标准:召回率(Recall)、精度(Precision)、F-measure 和 MCC^[19]。

假设仅考虑正常和异常两类实例的二元分类问题,假设评估所用的测试集 D 有 m 条数据记录,其中 C 表示异常类, NC 表示正常类,那么可以根据分类器得出,测试的预测的类

标和和实际的类标号是一个含混矩阵(Confusion Matrix)(见表 1)。其中异常类的样本总数表示被正确分类的异常样本数; FP 表示被错误分类的样本数, $n_c = n_{00} + n_{01}$;正常类的样本总数是 $n_{nc} = n_{10} + n_{11}$ 。 TP 表示正常样本数, FN 表示被错误分类为正常样本的异常样本数。在表 1 中 n_{ij} 的不同组合构成了评价分类器性能的不同标准。

表 1 含混矩阵

Table 1 Confusion Matrix

	预测为 C 类	预测为 NC 类	
实际为 C 类	$(TP)n_{00}$	$(FN)n_{01}$	n_c
实际为 NC 类	$(FP)n_{10}$	$(TN)n_{11}$	n_{nc}
总数	$n_0 = n_{00} + n_{10}$	$n_1 = n_{01} + n_{11}$	n

其中,本文采用的分类评估标准的定义如下:

准确率(Accuracy):所有样本数量中,被正确分类的样本数量所占的百分比。

$$A_c = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}} \quad (1)$$

召回率(Recall):在实际为少数目标类的所有样本数量中,被正确分类的样本数量所占的百分比。

$$R_c = \frac{n_{00}}{n_{00} + n_{01}} \quad (2)$$

精度(Precision):在预测为少数目标类的所有样本数中,

被正确分类的样本数所占的百分比。

$$P_c = \frac{n_{00}}{n_{00} + n_{10}} \quad (3)$$

F-measure:在 F-measure 中, $0 \leq \lambda \leq 1$, λ 就是对目标类的召回率 R_c 和精确率 P_c 的关注权重因子,一般取等同的权重,即 $\lambda = 0.5$ 。

$$F_{c\lambda} = \frac{1}{\lambda \frac{1}{R_c} + (1-\lambda) \frac{1}{P_c}} \quad (4)$$

MCC(Matthews's Correlation Coefficient):MCC 被视为一种平衡的度量标准。MCC 的值介于 -1 和 1 之间。当 MCC 为 1 时表示最好的预测;当 MCC 为 -1 时表示最坏的预测。

$$MCC = \frac{n_{00} \cdot n_{11} - n_{10} \cdot n_{01}}{\sqrt{(n_{00} + n_{10})(n_{01} + n_{11})(n_{00} + n_{01})(n_{10} + n_{11})}} \quad (5)$$

3 实验结果及分析

本文在人群密度估计标准数据集 PETs2009 上进行实验。Polus^[21]提出了经典的人群密度理论:当人群密度超过阈值($0.15 \text{ m}^2/\text{人}$)时,场景中的人群很有可能会失控,从而引起严重的人群安全事故。这一理论为后续关于人群密度的研究工作提供了很好的理论基础。尽管人群密度的定义是每平方单元的行人数量,但是对于人群密度分类来说,并不需要知道确切的人数。Polus 等提出将人群密度量化为 5 个等级:很低、低、中等、高和很高。这 5 个等级的划分标准和示例图如表 2 和图 5 所示。由于本文使用的 PETs2009 数据集规模较小,为了与文献[2]和文献[8]保持一致,我们将很低和低这两个等级合为一个低等级。PETs2009 由 S0, S1, S2 和 S3 4 个部分组成,本文选取用于人群计数和人群密度估计的 S1, L1, Time_13-57, S1, L1, Time_13-59 和 S1, L2, Time_14-06 组成实验的数据集。

表 2 密度等级划分标准

Table 2 Classification standard of density levels

密度等级	密度范(people/m ²)	人数范围(for 13 m ²)
很低	<0.5	<7
低	0.5~0.8	7~10
中等	0.81~1.26	11~16
高	1.27~2.0	17~26
很高	>2.0	>26



图 5 各密度等级示例图

Fig. 5 Illustration of various density levels

3.1 图像预处理

R_0, R_1, R_2 是人群密度估计中感兴趣的区域(见图 6),其中, R_0 用于人群计数, R_1 和 R_2 用于人群密度估计。先对图像进行灰度化,将 R_1 和 R_2 从图像中分割出来。当提取全连接层特征时,需要将分割出来的图像从原尺度调整到 224×224 像素尺度,在提取卷积层特征时保留原图像尺度。然后对图像进行均值归一化操作:计算所有图像的均值,减去均值后得到平均图像。最终获取的数据集的各个密度等级的样本数量如表 3 所列。将各个等级的样本一分为二,一半作为训练集,另一半作为测试集。



图 6 人群密度估计中感兴趣的区域

Fig. 6 Region of interest for crowd density estimation

表 3 各个密度等级的样本数量

Table 3 Number of samples for various density levels

兴趣区域	密度等级				总数
	低	中等	高	很高	
R_1	309	151	100	40	600
R_2	455	90	50	5	600

3.2 提取深度卷积特征

根据第 2 节的描述,本文使用 3 个经典 CNN 模型(AlexNet, VGGNet-16 和 VGGNet-19)来提取深度卷积特征。研究表明,深度模型低层提取的是边、角特征,中间层提取的是局部的目标特征,高层提取的是整体的目标特征^[22]。因此,高层特征更能表达目标的本质信息。在迁移学习中,最高层的信息更多地学习原任务目标的信息,而次高层的特征迁移性更好,更能反映目标任务的信息^[23]。本文中,对每个模型提取 5 种特征:fc6 层特征、relu6 层特征、fc7 层特征、relu7 层特征和 conv5 层特征。全连接层特征的维度是 4096 维,先进行 L2 范数归一化,然后对其使用 PCA 降维后得到 500 维的特征,然后将其输入到 SVM 进行分类。对卷积层特征,采用 conv5 层的值作为原始特征,然后使用具有 64 个高斯分量的 Fisher Vector 进行编码和聚合操作,最后得到 32768 维的特征。经实验验证,对卷积层特征进行 PCA 降维不利于识别效果的提升。

3.3 人群密度估计

基于 Matconvnet^[24]深度学习框架进行本文的人群密度估计实验。将 SoftmaxLoss 作为损失函数,使用随机梯度下降方法训练提出的卷积网络。批量的大小为 10,动量项的大小为 0.9。使用一个固定大小为 0.0001 的学习率。训练周期是 30。

为验证本文提出的深度卷积特征在人群密度估计上的有效性,提取了 4 种传统的纹理特征(GLCM、LBP、小波变换和 LBP+GLCM)作为对比实验。对比方法特征向量的描述如下:

GLCM:计算 $d=1, \theta$ 分别为 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 4 个方向共生矩阵的能量、对比度、均匀度、相关,进而获得一个 16 维的特征向量。

LBP:将检测窗口划分为 3×3 的小区域(cell),对于每个 cell 中的一个像素,对其环形邻域内的 8 个点进行顺时针比较,如果中心像素值比该邻点大,则将邻点赋值为 1,否则赋值为 0;然后计算每个 cell 的直方图,并对该直方图进行归一化处理;最后连接得到的每个 cell 的统计直方图,得到一个 59 维的特征向量。

WAVELET:进行四层小波分解,求分解后每个子带小波系数的均值和标准差,获得一个 26 维的特征向量。

LBP+GLCM:在 LBP 的编码图上计算一个 16 维的 GLCM 特征。

如表 4 和表 5 所列,所提卷积模型在多个指标上都优于传统方法。然而,卷积模型的识别结果不如基于迁移学习的方法获取的结果。Alexnet 提取的深度特征在多个评价指标上都优于其他几种方法;conv5+fv 获得的特征获得了最优的识别结果。对于全连接层,相比于 fc7 层,fc6 层具有更好的迁移性,获得的特征更能表达人群,识别效果最好。在不同的任务中,relu 层所表现出的效果不一样,在 R_1 区域,fc6 层比 relu6 层具有更好的效果;然而在 R_2 区域,relu6 层比 fc6 层具有更好的效果。VGGNet-16 和 VGGNet-19 模型特征获得的结果在多个评价指标上也都优于其他几种传统方法。就不同深度模型的识别表现而言,Alexnet 最优,VGGNet-16 优于 VGGNet-19。

表 4 各方法对 R_1 兴趣区域的密度估计结果

Table 4 Density estimation results of various methods in R_1 region of interest

(单位:%)					
模型	层	准确率	精度	召回率	F-measure
VGG19	fc6	96.00	94.49	94.50	94.48
	fc7	94.66	93.15	92.74	92.88
	relu6	95.66	94.25	94.06	94.14
	relu7	95.66	94.68	94.06	94.34
	Conv5_4+FV	97.04	95.38	95.47	95.32
VGG16	fc6	97.00	95.90	96.05	95.91
	fc7	96.66	94.68	96.05	95.23
	relu6	97.33	96.13	96.49	96.26
	relu7	96.66	94.68	96.05	95.23
	Conv5_3+FV	98.11	96.78	97.21	96.87
AlexNet	fc6	97.33	96.31	97.15	96.71
	fc7	96.00	95.21	96.05	95.61
	relu6	96.00	94.67	95.83	95.18
	relu7	95.66	94.98	95.61	95.27
	Conv5+FV	98.52	97.43	98.41	97.84
CNN	—	92.33	90.89	92.16	91.48
GLCM ^[6]	—	72.33	55.11	57.95	51.87
LBP ^[2]	—	84.33	79.51	79.08	79.28
WAVELET ^[9]	—	83.66	79.14	76.61	77.62
GLCM+LBP ^[8]	—	84.66	80.26	79.31	79.66

表 5 各方法对 R_2 兴趣区域的密度估计结果

Table 5 Density estimation results of various methods in R_2 region of interest

(单位:%)					
模型	层	准确率	精度	召回率	F-measure
VGG19	fc6	95.66	91.45	94.18	92.73
	fc7	94.00	89.04	92.84	90.74
	relu6	95.66	91.45	94.18	92.73
	relu7	92.00	86.13	89.66	87.81
	Conv5_4+FV	96.57	92.42	95.37	93.68
VGG16	fc6	96.66	92.52	95.84	94.05
	fc7	94.00	88.89	92.08	90.33
	relu6	97.00	93.10	95.99	94.46
	relu7	93.66	87.71	92.08	89.73
	Conv5_3+FV	97.48	93.51	96.92	95.41
AlexNet	fc6	97.33	94.24	96.71	95.45
	fc7	96.33	92.38	96.27	94.22
	relu6	98.00	94.66	98.23	96.37
	relu7	96.00	91.80	96.13	93.83
	Conv5+FV	99.12	95.79	99.31	97.54
CNN	—	95.33	92.24	78.40	81.10
GLCM ^[6]	—	84.33	74.89	69.22	70.25
LBP ^[2]	—	87.00	78.77	83.66	80.98
WAVELET ^[9]	—	87.00	80.60	79.02	79.76
GLCM+LBP ^[8]	—	86.00	75.45	73.65	73.99

从该实验可以看出:1)当数据样本规模较小时,相比于从头训练一个卷积模型,基于迁移学习的思想使用深度卷积特征是更优的选择;2)深度特征的迁移性与模型的深度并不成正比关系;3)相比于高层特征,较低层特征的迁移性更好;4)相比于全连接层特征,卷积层结合 FV 获得的特征的表达能力更强。为进一步比较这些深度特征,图 7 和图 8 展示了这些特征在各个密度等级上的识别表现。从图中可以看出,相比于其他几种方法,Alexnet 在各个密度等级上都取得了最好的识别结果。然而,其他几种传统方法的识别表现并不稳定。

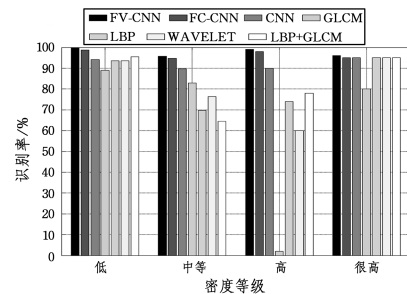


图 7 R_1 区域各个密度等级的识别率

Fig. 7 Recognition accuracy of various density levels in R_1

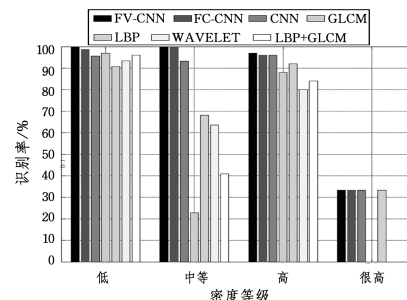


图 8 R_2 区域各个密度等级的识别率

Fig. 8 Recognition accuracy of various density levels in R_2

结束语 本文提出了一种基于卷积特征的人群密度估计方法。针对小样本人群数据,提出了一个较浅的卷积模型。此外,为了应用深度卷积特征,通过预训练深度卷积模型提取人群特征,然后使用 SVM 进行人群密度估计。针对人群密度数据不平衡问题,提出使用多个评价指标,能更准确地评价人群密度估计方法的优劣。在 PETs2009 数据集的 R_1, R_2 区域上的实验结果表明,所提方法在多个评价指标上都优于已有方法。实验结果还表明,面对小样本数据时,相比于训练一个全新的卷积模型,基于迁移学习的思想使用深度卷积特征是更好的选择;通过层数较少的深度模型获得的较低层特征的迁移性更好;深度卷积特征在各个密度等级上都取得了最好的且鲁棒的识别结果。

在进行人群密度估计时,本文仅使用了一种深度卷积特征,在未来的工作中可以将多种深度卷积特征进行融合。当前人群样本的规模很小,深度模型很难得到最优解,未来的工作将提出大规模的人群数据,进一步展示深度模型的优势。

参考文献

- [1] YANG H, SU H, ZHENG S, et al. The large-scale crowd density estimation based on sparse spatiotemporal local binary pattern[C] // Proceedings of IEEE International Conference on Multimedia and Expo. IEEE, 2011: 1-6.
- [2] FRADI H, ZHAO X, DUGELAY J L. Crowd density analysis using subspace learning on local binary pattern[C] // IEEE International Conference on Multimedia and Expo Workshops. IEEE, 2013: 1-6.
- [3] DAVIES A C, YIN J H, VELASTIN S A. Crowd monitoring using image processing[J]. Electronics & Communication Engineering Journal, 1995, 7(1): 37-47.
- [4] XU L Q, ANJULAN A. Crowd behaviours analysis in dynamic visual scenes of complex environment[C] // IEEE International Conference on Image Processing. IEEE, 2008: 9-12.
- [5] WU X, LIANG G, LEE K K, et al. Crowd density estimation using texture analysis and learning[C] // IEEE International Conference on Robotics and Biomimetics. IEEE, 2006: 214-219.
- [6] MA W, HUANG L, LIU C. Crowd density analysis using co-occurrence texture features[C] // International Conference on Computer Sciences and Convergence Information Technology. IEEE, 2010: 170-175.
- [7] MA W, HUANG L, LIU C. Advanced local binary pattern descriptors for crowd estimation[C] // IEEE International Conference on Computational Intelligence and Industrial Application Pacific-Asia Workshop. IEEE, 2008: 958-962.
- [8] WANG Z, LIU H, QIAN Y, et al. Crowd density estimation based on local binary pattern co-occurrence matrix[C] // IEEE International Conference on Multimedia and Expo Workshops. IEEE, 2012: 372-377.
- [9] MARANA A N, VERONA V V. Wavelet packet analysis for crowd density estimation[OL]. <http://www.oalib.com/references/15205544>.
- [10] MARANA A N, VELASTIN S A, COSTA L F, et al. Estimation of crowd density using image processing[C] // Proceedings of IEEE Colloquium on Image Processing for Security Applications. 1997: 11/1-11/8.
- [11] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [12] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: A review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828.
- [13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C] // Proceedings of Advances in Neural Information Processing Systems Nevada, USA: NIPS, 2012: 1097-1105.
- [14] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [15] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C] // IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.
- [16] ZHOU Y, CHEUNG N M. Vehicle Classification using Transferable Deep Neural Network Features[J]. arXiv preprint arXiv:1601.01145, 2016.
- [17] SARKAR S, PATEL V M, CHELLAPPA R. Deep feature-based face detection on mobile devices[C] // IEEE International Conference on Identity, Security and Behavior Analysis. IEEE, 2016: 1-8.
- [18] CIMPOI M, MAJI S, VEDALDI A. Deep filter banks for texture recognition and segmentation[C] // IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3828-3836.
- [19] KANDASWAMY K K, CHOU K C, MARTINETZ T, et al. AFP-Pred: A random forest approach for predicting antifreeze proteins from sequence-derived properties[J]. Journal of Theoretical Biology, 2011, 270(1): 56-62.
- [20] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [21] POLUS A, SCHOEFER J L, USHPIZ A. Pedestrian flow and level of service[J]. Journal of Transportation Engineering, 1983, 109(1): 46-56.
- [22] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C] // European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014: 818-833.
- [23] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [C] // Proceedings of Advances in Neural Information Processing Systems. 2014: 3320-3328.
- [24] VEDALDI A, LENC K. MatConvNet: Convolutional Neural Networks for MATLAB[C] // Proceedings of ACM International Conference on Multimedia. ACM, 2015: 689-692.