DOI:10.19651/j.cnki.emt.2108231

基于深度学习的人群计数算法综述

田月媛[1,2] 邓淼磊[1,2] 高 辉[1,2] 张德贤[1,2]
(1.河南工业大学信息科学与工程学院 郑州 450001;2.河南省粮食信息处理国际联合实验室 郑州 450001)

摘 要: 人群计数在视频监控、公共安全、智能商业等许多领域都有广泛的应用,近年来,随着深度学习的不断发展,人群计数已经成为计算机视觉领域研究的热点之一。本文根据提取特征方式的不同,将人群计数分为两类一类是传统方法,另一类是基于深度学习的方法,对基于卷积神经网络的方法进行重点分析和介绍;进一步介绍了人群计数领域的基准数据集和其他代表性数据集,实验结果表明,在人群密集和尺度变化较大的场景,基于卷积神经网络的方法优于传统方法,在尺度变化较大、人群较复杂的场景中多列网络比单列网络计数更加准确,效果更好;最后讨论了算法的未来发展方向。

关键词: 人群计数;卷积神经网络;深度学习;计算机视觉
中图分类号: TP183 **文献标识码:** A **国家标准学科分类代码:** 510.1050

Review of crowd counting algorithms based on deep learning

Tian Yueyuan[1,2] Deng Miaolei[1,2] Gao Hui[1,2] Zhang Dexian[1,2]
(1. School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China;
2. Henan International Joint Laboratory of Grain Information Processing, Zhengzhou 450001, China)

Abstract: Crowd counting is widely used in video surveillance, public security, intelligent business and many other fields. In recent years, with the continuous development of deep learning, crowd counting has become one of the hot topics in the field of computer vision. In this paper, according to the different feature extraction methods, crowd counting is divided into two categories: one is traditional method, the other is based on deep learning method, and the method based on convolutional neural network is analyzed and introduced. Further introduces the population count in the field of benchmark data sets and other representative data sets, the experimental results show that the larger changes in the crowded and scale, based on the convolution of the neural network method is superior to the traditional method, the scale change is bigger, more complex scenarios crowd more columns than a single network count more accurate, more effective. Finally, the future development direction of the algorithm is discussed.

Keywords: crowd counting;convolutional neural network;deep learning;computer vision

0 引 言

人群计数属于智能监控领域,主要是估计人群密集场景中人群密度或人群分布[1]。随着人群活动的不断增多,在车站、景区[2]等公共场所经常会出现人群聚集。为了避免因人群拥挤而发生踩踏等事故,提高安全防范能力,人群计数得到了广泛的研究。

人群计数发展至今根据其提取特征方式的不同主要分为两类,一个是传统机器学习的方法,另一个是基于深度学习的方法[3]。传统方法主要依靠手工提取特征,导致计数精度无法满足实际需求。得益于卷积神经网络(convolutional neural network, CNN)在计算机视觉其他领域的成功应用,此方法也逐渐被应用到计数领域中来提取特征。2015 年,Wang 等[4]首次提出将卷积神经网络用于人群计数中,和传统手工提取特征的方法相比,卷积模型包含了更多的细节信息[5],卷积神经网络作为一种端到端的自学习方法,在性能方面表现的更加优异,因此,基于 CNN 的人群计数方法也成为研究的热点。

本文第 1 节介绍了人群计数的主要方法,根据特征提取方式的不同,主要分为传统方法和基于深度学习的方法,对基于卷积神经网络的方法重点分析;第 2 节介绍了人群计数基准数据集与其他代表性数据集和模型性能评价指标,基于实验数据比较了不同模型的计数结果。最后,对人群计数未来可能的研究方向和趋势进行

收稿日期:2021-10-31

展望与总结。针对人群计数算法进行总结与分析,对不同场景中高效识别人群,大幅图提升计数精度具有重要的现实意义。

1 相关工作

1.1 传统人群计数方法

Loy 等[6]提出早期传统的人群计数算法大致可以分为3类:检测法、回归法和密度估计的方法。

基于检测的方法主要是检测视频中每帧图像的个体并统计人数。2012 年 Dollar 等[7]使用滑动窗口来检测行人或身体部位,并计算人数。训练一个学习算法的分类器,如 Boosting 和随机森林[8]等,最后基于分类器进行人数统计。上述方法在稀疏场景中得到了良好的计数结果,当人群密度较高或者场景复杂时,基于检测的人群计数算法无法得到准确的计数结果。

基于回归的方法主要是将人视为一个整体,建立手工提取特征到图像人数之间的映射来完成计数任务。线性回归、岭回归[9]和贝叶斯回归[10]等是经常用到的回归模型。在人群密集的场景中单一特征无法精确计算出人数,Idress 等[11]提出用傅里叶变换和尺度不变的特征变换(SIFT)的方法来提取特征,用马尔科夫随机场建立回归模型,最后获得图像中人群数量的估计值。由于回归法依然使用手工提取特征[12],并没有从根本上解决拥挤场景中的计数问题。

基于密度估计的人群计数算法主要是学习提取特征与密度图之间的映射关系,包括线性映射[13]和非线性映射[14]。传统人群计数方法使用手工提取特征,计算复杂影响了人群计数的准确度。

1.2 基于深度学习的人群计数方法

近年来,卷积神经网络具有对图像深层次特征出色的提取和学习能力,被广泛应用于人群计数领域中[15-17]。Wang 等[4]首次使用 CNN 来估计人群数目,适用于较密集的人群场景。Fu 等[18]对人群按密集程度分类,但是此方法只能粗略估计人数,导致计数结果不够准确。越来越多的研究者使用深度学习的方法来研究人群计数,图 1 所示为人群计数的发展路线。

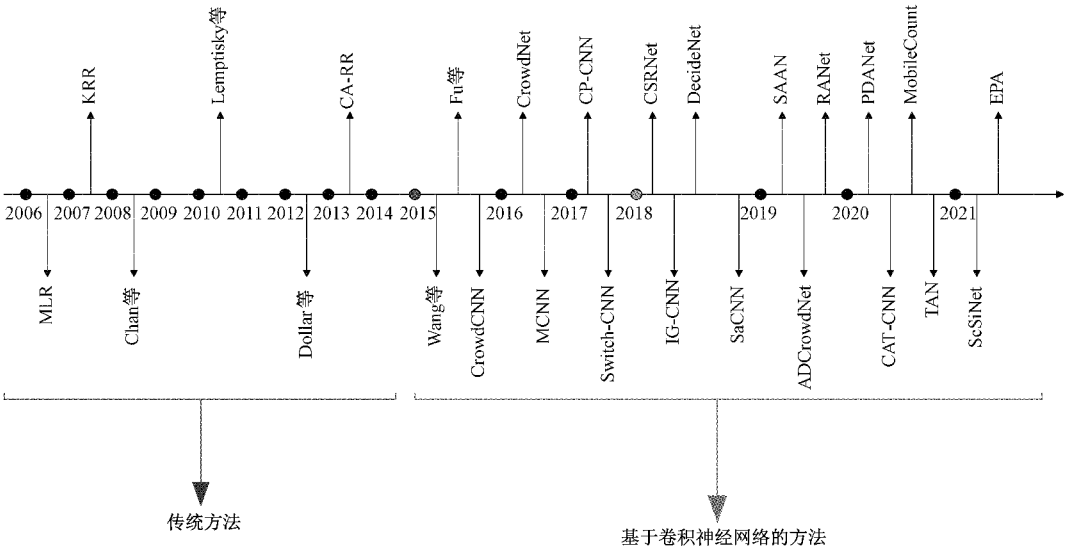

图 1 人群计数发展路线

根据网络结构的不同,将人群计数分为多列 CNN 模型和单列 CNN 模型[19]。

1) 多列 CNN 模型

多列网络是指用不同的列对应于不同感受野的多尺度信息。由于监控所处的位置和角度的不同,导致拍摄的人群图像中行人目标的尺度存在较大的差异。为了解决多尺度问题,Boomianthan 等[20]提出双分支计数网络,深层网络用来捕获高级语义信息,浅层网络检测小的头部斑点,将深层网络和浅层网络结合,有利于不同场景不同尺度的人群计数。

同年,Zhang 等[21]提出使用多列卷积神经网络(MCNN)来解决尺度自适应问题,使用大小不同的感受野来提取人群多尺度信息,在多尺度场景中计数准确性明显提高。Zhang 等还收集并标注了 ShanghaiTech 人群计数数据集,目前此数据集已经成为人群计数领域的基准数据集之一。

由于 MCNN 具有较多的参数,计算较复杂,Sam 等[22]在此基础上提出转换网络(switch convolution neural network,Switch-CNN),利用图像中人群密度的变化来提高预测人群数量的质量和定位,在高遮挡和尺度变化剧烈的场景中,计数准确性和鲁棒性都有明显提升。

同年,Sindagi 等[23]提出上下文金字塔计数模型 CP-CNN,为了获得更好的计数结果和高质量密度图,他们提出融合局部和全局的人群特征,共同生成密度图,还引

入对抗损失克服欧氏距离损失函数的不足。

现实世界中人群计数场景经常面临巨大的密度变化问题，无论是使用检测的方法还是回归的方法都无法对整个密度范围进行精确估计。Liu 等[24]提出了一种检测和密度估计网络 DecideNet，根据人群密度的变化自适应地调整权重，将检测法和回归法结合来估计人群密度。

受 Chen 等[25]在语义分割工作方面的启发，2019 年 Hossain 等[26]提出了多分支的尺度感知注意力网络(scalc-aware attention network，SAAN)，在 CP-CNN 的基础上引入注意力机制，用注意力来关注特定的尺度，通过使用注意力机制可以自动选择全局特征和局部上下文信息，得到更准确的计数精度。

密度估计是人群计数中的重要策略，传统密度估计方法主要是像素回归，不明确考虑像素之间的相互依赖，导致独立的像素级预测可能会有噪声。为了解决此问题，Zhang 等[27]提出了一种用于人群计数的关系注意网络(RANet)，该网络结合了全局和局部自我注意机制，捕获像素的长距离和短距离相互依赖关系。RANet 集成了注意力机制和关系模块，增强了人群计数的特征表示。

2020 年 Amirgholipour 等[28]也注意到密度估计在人群计数中的重要作用，他们提出了基于金字塔密度感知注意网络(PDANet)，PDANet 结合了多尺度特征提取、密度分类和自适应密度估计 3 个模块，有助于处理不同图像之间以及每个输入场景中的密度变化。PDANet 的提出解决了拥挤场景的巨大的密度变化，提升了计数的准确性。

大部分现有的方法使用不同感受野来解决尺度变化，但是和连续尺度相比，现有方法实现了有限的尺度多样性。2021 年，为了有效地分析任意场景的透视信息处理尺度的变化，Yang 等[29]提出了一种透视感知计数网络(EPA)，该方法从密度图中挖掘透视信息，并将透视分析嵌入到计数网络。他们还提出了变换扩张卷积，打破了固定的采样结构，提高了效率和估计的精度。此方法的提出，对未来的研究工作提供了更多的思路，未来在人群计数领域，将研究卷积层来获得更灵活的感受野。表 1 是对多列网络模型下主流人群计数算法的总结与分析。

表 1　人群计数主流算法分析与总结

方法类别	主流算法	算法特点	优点	缺点
多列网络	CrowdNet	深层网络和浅层网络相结合，有利于不同场景不同尺度的人群计数。	采用不同的列来对应于不同感受野的多尺度信息，提高了人群计数的性能和目标尺度变化的鲁棒性。	网络模型参数较多，训练困难，导致计数实时性较差；网络结构冗余度较高。
	MCNN	用大小不同的感受野提取人群多尺度信息。		
	CP-CNN	将上下文信息和高维特征图相结合，生成高分辨率的密度图。		
	SAAN	使用注意力机制自动选择全局特征和局部上下文信息。		
	RANet	全局和局部自我注意机制结合，集成了注意机制和关系模块，增强人群计数的特征表示。		
	PDANet	将多尺度特征提取、密度分类和自适应密度估计结合，有助于处理不同图像的密度变化。		

2) 单列 CNN 模型

虽然多列网络的研究取得了很大的进展，但是由于其参数较多、训练困难、网络结构冗余度较高，导致计数实时性较差，所以许多研究人员开始开发更简单、更高效的网络。和多列网络相比，单列网络仅存在单个深度网络，不增加网络的复杂性，结构简单，模型训练容易。

2016 年，为了解决跨场景的问题 Zhang 等[30]提出了一个用于交叉场景的人群计数模型 CrowdCNN，与早期的方法相比，Zhang 等提出将密度图和透视图结合，降低透视形变的不良影响，提升密度图质量。此外，他们还引入了一个用于评估跨场景计数的数据集 WorldExpo'10。

2018 年，Li 等[31]提出了空洞卷积神经网络模型 CSRNet，首次将空洞卷积用于人群计数，扩大了感受野，更利于人头特征的提取。CSRNet 的成功为人群计数提供了新的思路，随后许多学者开始效仿采用空洞卷积进行人群计数的研究[32]。

现有方法均使用欧氏距离作为损失函数导致生成的密度图图像模糊，Cao 等[33]提出了一种尺度聚合网络(SANet)，借鉴了 Inception 的架构思想，通过堆叠 4 个不同尺寸并联的卷积核对 网络进行编码，利用反卷积层进行解码，从而生成高分辨率的密度图。

由于背景杂波、严重遮挡和人群分布多样化等因素导致人群计数的结果不准确，Liu 等[34]提出了一种融合注意力机制的可变形卷积网络 ADCrowdNet，通过引入注意力机制强调了人群区域，使用可变形卷积保证了高度拥挤场景中密度图的准确性。同年，DADNet[35]也使用可变形卷积进行人群计数，取得了良好的计数效果。

之前的研究广泛采用均方误差(MSE)损失假设像素独立，忽略了密度图中的局部相关性和空间相关性，不足以促进高质量密度图的生成。为了解决现有编码器-解码器网络中的这些问题，提高计数性能。Jiang 等[36]提出了网格编码器-解码器网络(TEDnet)，利用跨层连接和多尺

度特征融合,在语义和空间上融合多尺度特征,还提出了空间抽象损失,弥补欧几里得损失的缺点,感知人群的尺度变化,提升密度图质量。

受 Li 等[31]在单列 CNN 在人群计数领域取得的良好性能的影响,2020 年 Wang 等[37]提出了一种轻量级的编码器-解码器体系结构(MobileCount),以达到精度和速度之间的最佳平衡,用 MobilenetV2 作为主干网络既不会降低精度同时减少了参数的运算量,并减少了内存占用,提高了精度。

2021 年,Wang 等[38]提出了一种单列尺度不变网络(ScSiNet),该网络通过层间多尺度积分和层内尺度不变变换相结合来提取复杂的尺度。为了扩大密度的多样性,提出了一种随机积分的损失方法,有效地减轻了人群数据集中巨大的密度变化的影响。大量实验表明,该方法在计数精度方面始终优于现有的方法,并具有显著的可转移性和尺度不变性。此方法的提出有效地解决了拥挤场景中尺度连续变化的问题,避免了过度拟合。

在实际场景中,由于行人不均匀分布、遮挡、光照等条件,可靠有效地计算人群数量仍然是一项挑战。由于人群计数的数据集通常由监控视频采集,和静止的图像相比,视频帧之间有一定的重叠,现有的方法忽略了相邻帧之间的时间关系,影响了计数的性能。Wu 等[39]提出了一种时间感知网络(TAN)用于动态模拟人群计数连续帧的时间特性,使用轻量级卷积神经网络处理计数任务加快了计算的速度保证了计数的准确性。此模型的提出为以后在研究视频中的人群计数提供了参考。表 2 是对单列网络结构下人群计数的分析与总结,与多列网络结构相比,单列网络结构简单,训练容易但对尺度变化处理效率较低。

表 2 人群计数算法分析与总结

方法类别	主流算法	算法特点	优点	缺点
单列网络	CrowdCNN	将密度图和透视图结合降低了透视形变的影响。	削减了网络参数的数量,降低了训练难度,网络结构简单训练效率更高。	单列网络不能有效分析任意场景的透视信息,对尺度变化的处理效率低。
	CSRNet	借助空洞卷积保持分辨率扩大了感受野。		
	SANet	并联不同尺寸的卷积核提取多尺度特征,使用反卷积生成高分辨率的密度图。		
	ADCrowdNet	使用可变形卷积保证了高度拥挤场景中密度图的准确性。		
	TEDnet	利用跨层连接和多尺度特征融合,在语义和空间上融合多尺度特征。		
	ScSiNet	通过层间多尺度积分和层内尺度不变变换相结合来提取复杂尺度不变特征,提高了计数性能。		

2 数据集和评价指标

2.1 人群计数基准数据集

在过去的几年中,出现了各种各样的数据集,早期的数据集包含低密度人群图像,最近的数据集则关注高密度人群,因此也带来了许多问题,如遮挡等。本节主要介绍如下 5 个基准数据集:UCSD[40]数据集、Mall[41]数据集、UCF_CC_50[42]数据集、WorldExpo'10[30]数据集、Shanghai Tech[21]数据集,表 3 所示为人群计数基准数据集,由于拍摄角度和场景的不同,呈现出多样化的特点。

表 3 基准数据集信息统计

数据集	图像数量	分辨率	总人数	平均人数	训练集图像	测试集图像
UCSD	2 000	238×158	49 885	15	800	1 200
Mall	2 000	320×240	62 325	31	800	1 200
UCF_CC_50	50	2 101×2 888	63 075	1 280	—	—
WorldExpo'10	3 980	576×720	199 923	50.2	—	—
Shanghai Tech A	482	589×868	241 677	501.4	300	182
Shanghai Tech B	716	768×1 024	88 488	123.6	400	316

UCSD[40]数据集是最早用于人群计数的数据集,是从人行道的摄像头中收集的,它包含 2 000 帧图像,每帧大小 238×158,每 5 帧标注一次每个行人的地面实况,此数据集一共有 49 885 个行人,其中测试集包含 1 200 个图像,训练集包含 800 个图像,该数据集场景单一,人群密度相对较低。

由于 UCSD 数据集的场景单一,所以 Chen 等[41]收集了一个具有不同照明条件和人群密度的 Mall 数据集,该数据集是在一家购物中心的监控摄像头中收集的,它包含 2 000 帧图像,每帧大小 320×240,此数据集一共有 62 325

个行人，其中测试集图像 1 200 个，训练集图像 800 个。

UCF_CC_50[42]数据集是第 1 个具有挑战性的大规模人群计数数据集，该数据集是从公开课用的 web 图像创建的，为了捕获到不同的场景，作者收集了带有不同标签的图像，如音乐会、抗议、体育场和马拉松等，它一共有 50 幅图像，每帧大小 2 101×2 888，每张图像平均有 1 280 个人，共有 63 075 个被标记的个体，每张图像的个体数量从 94～4 543 个不等。

由于早期的数据集的场景比较单一，因此 Zhang 等[30]引入了一个用于跨场景计数的数据集，该数据集共包含 3 980 帧图像。每帧图像大小为 576×720，每张图像平均有 50.2 个人，有 199 923 个个体被标记，该数据集的测试集只有 5 个不同场景，场景数量少，所以此数据集不足以评估极密集人群。

Shanghai Tech[21]数据集，早在 2016 年 Zhang 等提出了多列卷积神经网络 MCNN 的同时，还介绍了一个新的大规模数据集，此数据集是标注人数最多的数据集之一，一共有 1 198 幅图像，共标记了 330 165 个个体，根据密度的不同，数据集分为两部分：A 部分有 482 张从互联网上随机选择的图像，每帧图像大小 589×868，训练集包含 300 张图像，测试集包含 182 张图像。B 部分的图像是从上海大都市的一条繁忙街道上拍摄的，每帧图像大小为 768×1 024，训练集包含 400 张图像，测试集包含 316 张图像，两者相比，A 部分的图像密度比 B 部分大，此数据集成功创建了跨不同场景类型和密度的具有挑战性的数据集。

2.2　人群计数其他数据集

本节主要介绍近几年新出现的数据集：Smartcity[43]数据集、UCF-QNRF[44]数据集、Crowd Surveillance[45]数据集、GCC[46]数据集和 NWPU-Crowd[47]数据集。

Smartcity[43]数据集，一共有 50 幅图像，每帧图像大小为 1 920×1 080，分别来自人行道、办公室入口和购物中心等 10 个场景，总共标记了 369 个个体，每幅图像平均有 7.4 个个体，此数据集主要用于验证在稀疏场景下模型的泛化能力。

UCF-QNRF[44]数据集，一共有 1 535 幅图像，每帧图像大小为 2 013×2 902，该数据集总共标记了 12 865 个个体，每幅图像平均有 815 个个体，其中测试集包含 334 幅图像，训练集包含 1 201 幅图像。由于该数据集中的部分图像分辨率太高，导致在训练时会出现 GPU 内存不足的问题，是一个具有挑战性的数据集。

Crowd Surveillance[45]数据集，由 13 945 幅图像组成，共 386 513 个个体，每帧图像大小为 1 342×840，是目前为止人群计数的最大和最高的平均分辨率，与此同时该数据集还提供了 ROI 注释，用于过滤训练和测试图像中模糊的部分。

GCC[46]数据集，一共包含 15 212 张图像，每张图像大小为 1 080×1 920，共 7 625 843 个个体，平均每幅图像有 501 个个体，和现有的数据集相比 GCC 有如下 4 个优点：免费注释、数据量大，分辨率高、多样化的场景。

NWPU-Crowd[47]数据集，一共有 5 109 幅图像，每帧图像大小为 2 311×3 383，共有 213 328 个个体被标注，平均每幅图像包含 418 个个体，该数据集和之前的数据集相比具有更高的分辨率，有利于提升计数的准确性，同时还加入了一些负样本，有利于提升训练模型的鲁棒性。

2.3　性能评价指标

平均绝对误差 MAE 用来计算数据集中所有图像的实际人数和预测人数的绝对差值的平均值。均方误差 RMSE 用来计算数据集中所有图像的实际人数和预测人数的绝对差值的平方平均值。定义如下：

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|x_i-\hat{x}_i| \tag{1}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i-1}^{N}|x_i-\hat{x}_i|} \tag{2}$$

其中，N 为测试数据集中图像的总数，x_i 为图像人数的实际值，$\hat{x}_i$ 为图像人数的估计值。

本文所述人群计数算法在 UCF_CC_50、Shanghai Tech 等基准数据集上的实验结果对比如表 4 所示，和传统方法相比，基于 CNN 的方法在拥挤场景下计数效果更准确，性能更好。在基于卷积神经网络的方法中，大多数算法通过增加模型复杂性减少计数误差；多列网络与单列网络相比，多列网络模型在复杂场景上的表现效果更好。

3　展　　望

在尺度变化、遮挡等场景中基于卷积神经网络的人群计数算法具有优越的性能，因此使用卷积神经网络来计数也逐渐成为主流发展方向[48]，随着人群计数与密度估计在公共安全、城市规划[49-50]等领域的应用日益增多，人群计数已成为国内外计算机视觉领域的研究热点[51]，但仍面临着许多挑战。未来可以从如下方面继续探索。

严重遮挡问题。密集的集的人群之间遮挡严重，混乱程度高，部分行人只能展现局部特征，导致计数结果不准确，所以如何在高度遮挡的场景中获得准确的计数结果是未来的重要研究方向。

尺度变化是人群计数面临的主要挑战之一，一些研究者使用多列网络模型来学习不同尺度的特征[52]，在一定程度上可以解决视角变化问题[49]，大多数研究者引入注意力机制[53]，使用空洞卷积[54]、对抗生成网络[55]和可变形卷积[56]等来解决多尺度问题，提升了密度图的质量，未来多尺度问题仍是人群计数的一个研究方向。

视频中的人群计数。在现有的方法中，大多数都是在静止图像上计数，在视频上的探索是有限的，因此在视频中，如何实时处理数据并具有较高准确率，是人群计数未来的一个重要发展方向。

背景外界因素干扰。在复杂的背景中，和行人头部相

表4 基准数据集在不同网络上的性能对比

数据集	UCSD		Mall		UCF_CC_50		World Expo'10		SHT A		SHT B	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Idress 等[11]	—	—	—	—	468.00	590.30	—	—	—	—	—	—
Pham 等[14]	1.61	4.40	2.50	10.00	—	—	—	—	—	—	—	—
Wang 等[4]	1.98	1.82	2.74	2.10	—	—	—	—	—	—	—	—
MCNN[21]	1.07	1.35	—	—	**377.60**	**509.10**	11.60	—	**110.20**	**173.20**	**26.40**	**41.30**
Switch-CNN[22]	**1.62**	**2.10**	—	—	318.10	439.20	9.40	—	90.40	135.00	21.60	33.40
CP-CNN[23]	—	—	—	—	295.80	320.90	8.86	—	73.60	106.40	20.10	30.10
DecideNet[24]	—	—	1.52	1.90	—	—	9.23	—	—	—	21.53	31.98
SAAN[26]	—	—	1.28	1.68	271.60	391.00	—	—	—	—	16.86	28.41
PDANet[28]	0.93	1.21	—	—	119.80	159.00	6.00	—	58.50	93.40	7.10	10.90
EPA[29]	1.02	1.21	—	—	250.10	342.10	7.40	—	62.30	100.90	8.30	12.60
CSRNet[31]	1.16	1.47	—	—	266.10	397.50	8.60	—	68.20	115.00	10.60	16.00
SANet[33]	1.02	1.29	—	—	258.40	334.90	8.20	—	67.00	104.50	8.40	13.60
ADCrowdNet[34]	1.09	1.35	—	—	273.60	362.00	7.30	—	70.90	115.20	7.70	12.90
TEDNet[36]	—	—	—	—	249.40	354.50	8.00	—	64.20	109.10	8.20	12.80
MobileCount[37]	—	—	—	—	321.70	437.10	12.20	—	98.60	162.90	9.10	15.10
ScSiNet[38]	—	—	—	—	157.90	199.40	—		55.77	90.23	6.79	10.95
TAN[39]	1.08	1.41	2.03	2.60	262.00	358.60	8.30	—	93.30	157.00	15.10	23.30

似的部分容易被识别为人群导致计数结果过估计,未来如何解决这些外界因素造成的计数误差是一个具有挑战性的问题[57]。

4 结 论

本文根据提取特征方式的不同,将人群计数分为两类:一类是传统方法,另一类是基于深度学习的方法,对基于卷积神经网络的方法进行重点分析和介绍;进一步介绍了人群计数领域的基准数据集和其他代表性数据集,实验结果表明,在人群密集和尺度变化较大的场景,基于卷积神经网络的方法优于传统方法,在尺度变化较大、人群较复杂的场景中多列网络比单列网络计数更加准确,效果更好;最后讨论了算法的未来发展方向。

参考文献

[1] 余思悦,浦剑.聚合上下文信息的人群计数(英文)[J].Frontiers of Information Technology & Electronic Engineering,2020,21(11):1626-1639.

[2] 徐涛,段仪浓,杜佳浩,等.基于多尺度增强网络的人群计数方法[J].电子与信息学报,2021,43(6):1764-1771.

[3] 孟月波,陈宣润,刘光辉,等.多特征信息融合的人群密度估计方法[J].激光与光电子学进展,2021,58(20):276-287.

[4] WANG C, ZHANG H, YANG L, et al. Deep people counting in extremely dense crowds[C]. Proceedings of the 23rd ACM international conference on Multimedia, 2015: 1299-1302.

[5] 朱阳光,刘瑞敏,黄琼桃.基于深度神经网络的弱监督信息细粒度图像识别[J].电子测量与仪器学报,2020,34(2):115-122.

[6] LOY C C, CHEN K, GONG S, et al. Crowd counting and profiling: Methodology and evaluation[M]. Modeling, Simulation and Visual Analysis of Crowds, Springer, New York, NY, 2013: 347-382.

[7] DOLLAR P, WOJEK C, SCHIELE B, et al. Pedestrian detection: An evaluation of the state of the art[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(4): 743-761.

[8] GALL J, YAO A, RAZAVI N, et al. Hough forests for object detection, tracking, and action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2188-2202.

[9] 张宇倩,李国辉,雷军,等.FF-CAM:基于通道注意机制前后端融合的人群计数[J].计算机学报,2021,44(2):304-317.

[10] CHAN A B, VASCONCELOS N. Counting people with low-level features and Bayesian regression[J]. IEEE Transactions on Image Processing, 2011, 21(4): 2160-2177.

[11] IDREES H, SALEEMI I, SEIBERT C, et al. Multi-source multi-scale counting in extremely dense crowd images[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013:

2547-2554.

[12] 余鹰,潘诚,朱慧琳,等.融合通道与空间注意力的编解码人群计数算法[J/OL].计算机科学与探索,2022:1-10[2022-01-04]. http://kns.cnki.net/kcms/detail/11.5602.TP.20210624.0840.004.html.

[13] WANG Y, ZOU Y. Fast visual object counting via example-based density estimation[C]. 2016 IEEE International Conference on Image Processing(ICIP), IEEE, 2016: 3653-3657.

[14] PHAM V Q, KOZAKAYA T, YAMAGUCHI O, et al. Count forest: Co-voting uncertain number of targets using random forest for crowd density estimation[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 3253-3261.

[15] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.

[16] CHANG L, DENG X M, ZHOU M Q, et al. Convolutional neural networks in image understanding[J]. Acta Automatica Sinica, 2016, 42(9): 1300-1312.

[17] 蓝金辉,王迪,申小盼.卷积神经网络在视觉图像检测的研究进展[J].仪器仪表学报,2020,41(4):167-182.

[18] FU M, XU P, LI X, et al. Fast crowd density estimation with convolutional neural networks[J]. Engineering Applications of Artificial Intelligence, 2015, 43: 81-88.

[19] GAO G, GAO J, LIU Q, et al. Cnn-based density estimation and crowd counting: A survey[J]. ArXiv Preprint,2020, ArXiv:2003.12783.

[20] BOOMINATHAN L, KRUTHIVENTI S S S, BABU R V. Crowdnet: A deep convolutional network for dense crowd counting[C]. Proceedings of the 24th ACM International Conference on Multimedia, 2016: 640-644.

[21] ZHANG Y, ZHOU D, CHEN S, et al. Single-image crowd counting via multi-column convolutional neural network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 589-597.

[22] SAM D B, SURYA S, BABU R V. Switching convolutional neural network for crowd counting[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5744-5752.

[23] SINDAGI V A, PATEL V M. Generating high-quality crowd density maps using contextual pyramid cnns[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 1861-1870.

[24] LIU J, GAO C, MENG D, et al. Decidenet: Counting varying density crowds through attention guided detection and density estimation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5197-5206.

[25] CHEN L C, YANG Y, WANG J, et al. Attention to scale: Scale-aware semantic image segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 3640-3649.

[26] HOSSAIN M, HOSSEINZADEH M, CHANDA O, et al. Crowd counting using scale-aware attention networks[C]. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019: 1280-1288.

[27] ZHANG A, SHEN J, XIAO Z, et al. Relational attention network for crowd counting[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6788-6797.

[28] AMIRGHOLIPOUR S, HE X, JIA W, et al. Pdanet: Pyramid density-aware attention net for accurate crowd counting[J]. ArXiv Preprint, 2021, ArXiv:2001.05643.

[29] YANG Y, LI G, DU D, et al. Embedding perspective analysis into multi-column convolutional neural network for crowd counting[J]. IEEE Transactions on Image Processing, 2020, 30: 1395-1407.

[30] ZHANG C, LI H, WANG X, et al. Cross-scene crowd counting via deep convolutional neural networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 833-841.

[31] LI Y, ZHANG X, CHEN D. Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1091-1100.

[32] LIU W, SALZMANN M, FUA P. Context-aware crowd counting[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5099-5108.

[33] CAO X, WANG Z, ZHAO Y, et al. Scale aggregation network for accurate and efficient crowd counting[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 734-750.

[34] LIU N, LONG Y, ZOU C, et al. Adcrowdnet: An attention-injective deformable convolutional network for crowd understanding[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3225-3234.

[35] GUO D, LI K, ZHA Z J, et al. Dadnet: Dilated-attention-deformable convnet for crowd counting[C]. Proceedings of the 27th ACM International Conference on Multimedia, 2019: 1823-1832.

[36] JIANG X, XIAO Z, ZHANG B, et al. Crowd counting and density estimation by trellis encoder-decoder networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 6133-6142.

[37] WANG P, GAO C, WANG Y, et al. MobileCount: An efficient encoder-decoder framework for real-time crowd counting[J]. Neurocomputing, 2020, 407: 292-299.

[38] WANG M, CAI H, ZHOU J, et al. Interlayer and intralayer scale aggregation for scale-invariant crowd counting[J]. Neurocomputing, 2021, 441: 128-137.

[39] WU X, XU B, ZHENG Y, et al. Fast video crowd counting with a temporal aware network[J]. Neurocomputing, 2020, 403: 13-20.

[40] SINDAGI V A, PATEL V M. A survey of recent advances in cnn-based single image crowd counting and density estimation[J]. Pattern Recognition Letters, 2018, 107: 3-16.

[41] CHEN K, LOY C C, GONG S, et al. Feature mining for localised crowd counting[C]. Bmvc, 2012, 1(2): 3.

[42] IDREES H, SALEEMI I, SEIBERT C, et al. Multi-source multi-scale counting in extremely dense crowd images[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2547-2554.

[43] ZHANG L, SHI M, CHEN Q. Crowd counting via scale-adaptive convolutional neural network[C]. 2018 IEEE Winter Conference on Applications of Computer Vision(WACV), IEEE, 2018: 1113-1121.

[44] IDREES H, TAYYAB M, ATHREY K, et al. Composition loss for counting, density map estimation and localization in dense crowds[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 532-546.

[45] YAN Z, YUAN Y, ZUO W, et al. Perspective-guided convolution networks for crowd counting[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 952-961.

[46] WANG Q, GAO J, LIN W, et al. Learning from synthetic data for crowd counting in the wild[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8198-8207.

[47] WANG Q, GAO J, LIN W, et al. NWPU-crowd: A large-scale benchmark for crowd counting and localization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(6): 2141-2149.

[48] 蒋妮,周海洋,余飞鸿.基于计算机视觉的目标计数方法综述[J].激光与光电子学进展,2021,58(14):43-59.

[49] COŞAR S, DONATIELLO G, BOGORNY V, et al. Toward abnormal trajectory and event detection in video surveillance[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(3): 683-695.

[50] 柳长源,王琪,毕晓君.多目标小尺度车辆目标检测方法[J].控制与决策,2021,36(11):2707-2712.

[51] 刘丹,汪慧兰,曾浩文,等.VoVNet-FCOS道路行人目标检测算法研究[J].国外电子测量技术,2021,40(11):64-71.

[52] ZHANG Y, ZHOU D, CHEN S, et al. Single-image crowd counting via multi-column convolutional neural network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 589-597.

[53] HOSSAIN M, HOSSEINZADEH M, CHANDA O, et al. Crowd counting using scale-aware attention networks[C]. 2019 IEEE Winter Conference on Applications of Computer Vision(WACV), IEEE, 2019: 1280-1288.

[54] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. ArXiv Preprint, 2015, ArXiv:1511.07122.

[55] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014.

[56] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 764-773.

[57] 余鹰,朱慧琳,钱进,等.基于深度学习的人群计数研究综述[J].计算机研究与发展,2021,58(12):2724-2747.

作者简介

田月媛,硕士研究生,主要研究方向为计算机视觉、模式识别与智能系统。

E-mail:15832963805@163.com

邓淼磊,工学博士,教授,硕士生导师,主要研究方向为信息安全、物联网技术。

E-mail:dmlei2003@163.com

高辉,博士研究生,主要研究方向为计算机视觉、模式识别与智能系统。

E-mail:ghshow@139.com

张德贤,工学博士,教授,博士生导师,主要研究方向为模式识别与智能信息处理技术研究。

E-mail:zdx@haut.edu.cn