

编码-解码多尺度卷积神经网络人群计数方法

孟月波, 纪拓, 刘光辉, 徐胜军, 李彤月

(西安建筑科技大学信息与控制工程学院, 710055, 西安)

摘要: 针对基于多列卷积神经网络的人群计数方法存在的多尺度特征信息丢失、融合不佳以及密度图质量不高等问题,提出了一种编码-解码结构的多尺度卷积神经网络人群计数方法。编码器采用多列卷积捕获多尺度特征,通过空洞空间金字塔池化扩大感受野并减少参数量,保留尺度特征和图像的上下文信息;解码器对编码器输出进行上采样,实现高层语义信息和编码器前端低层特征信息有效融合,从而提升了密度图的输出质量。为增强网络对计数的敏感性,在以往像素空间损失的基础上考虑了计数误差,提出了一种新型损失函数。采用 Shanghai Tech、Mall 以及自建数据集进行了对比实验,结果表明:与之前最优方法相比,所提方法在 Shanghai Tech 数据集 Part_A 部分的平均绝对误差和均方误差分别降低了 8.3% 和 21.3%, Part_B 部分分别降低了 12.9% 和 12.0%, Mall 数据集分别降低了 15.1% 和 23.8%, 自建数据集分别降低了 13.5% 和 7.1%;在不同人群场景下,所提方法的人群计数准确性和鲁棒性均优于其他对比方法的。

关键词: 人群计数;编码-解码结构;多尺度;空洞空间金字塔池化;计数误差;损失函数

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtuxb202005020 **文章编号:** 0253-987X(2020)05-0149-09



OSID 码

Encoding-Decoding Multi-Scale Convolutional Neural Network for Crowd Counting

MENG Yuebo, JI Tuo, LIU Guanghui, XU Shengjun, LI Tongyue

(School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: Aiming at the problems of multi-scale feature information loss, poor fusion and low quality of density map in the crowd counting method based on multi-column convolutional neural network, a new crowd counting method is proposed based on encoding-decoding multi-scale convolutional neural network. The encoder part adopts multi-column convolution to capture multi-scale features, expands the receptive field and reduces the amount of calculation via atrous space pyramid pooling, and retains the multi-scale feature and the context information of the image. The decoder part upsamples the encoder output to achieve effective fusion of the features with rich high-level semantic information and the features with rich low-level detail information to improve the output quality of the density map. To enhance the sensitivity of the network to counting, a new loss function is proposed by considering the previous pixel space loss and the counting error. Contrast experiments with previous methods on Shanghai Tech, Mall, and self-built datasets are conducted, and it is found that the mean absolute error and mean square error of

收稿日期: 2019-11-20。 作者简介: 孟月波(1979—),女,副教授;刘光辉(通信作者),男,副教授。 基金项目: 国家自然科学基金资助项目(51678470);陕西省教育厅专项科研计划资助项目(18JK0477);西安建筑科技大学基础研究基金资助项目(JC1703)。

网络出版时间: 2020-01-14

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.T.202001113.1748.002.html>

this method on part_A of Shanghai Tech dataset are 8.3% and 21.3% lower than the previous optimal method, and 12.9% and 12.0% lower in part_B of Shanghai Tech dataset. The mean absolute error and mean square error decrease by 15.1% and 23.8% for the Mall dataset, and decrease by 13.5% and 7.1% for the self-built dataset. The experimental results on Shanghai Tech, Mall and self-built datasets show the higher accuracy and better robustness of the proposed method than the traditional methods.

Keywords: crowd counting; encoding-decoding; multi-scale; atrous space pyramid pooling; counting error; loss function

随着国民经济迅猛发展及城市化进度不断加快,城市人口数量急剧增加,由此带来的社会问题也不断增加,例如:人员拥挤导致的踩踏事故的发生,候车大厅人员的增加对交通调度带来的压力等。解决这些问题需要准确预测出实际场景下的人群数量,而图像可以清晰直观地反映出实际场景中的人群变化情况,因此基于图像信息的人群密度估计与计数具有重要的研究意义。

目前,人群计数方法可以分为基于目标分析、基于纹理分析、基于像素统计共3类。基于目标分析的人群计数方法依赖场景中个体的识别,其主要思想是通过一定的技术将图像中的人体进行分割,然后对分割得到的个体进行统计,从而得到最终结果^[1-3]。基于纹理分析的人群计数方法是纹理特征作为密度特征,不同密度的人群图像具有不同的纹理特征,根据图像的纹理特征就可以对人群进行密度分类^[4-6]。基于像素统计的人群计数方法基本思想是人群数与像素数之间存在较强的正比例关系,人数越多对应的像素数也就越多,故可通过计算人群目标的像素数来估计人群数量^[7-8]。这些方法所提取的特征单一,在低中密度且背景相对比较简单的环境下有较好的人群计数效果。随着科技的进步,许多领域追求更精确的人群计数结果,这些方法已无法满足人群计数任务的精度要求。

卷积神经网络(CNN)模型因具有获取深层次特征的能力,在语义分割、目标检测与识别等领域发展迅速^[9-11],研究人员也将其应用于人群计数领域并取得了较好的效果。Wang等首次提出适用于高密度人群的端到端CNN回归模型,将AlexNet网络结构^[12]作为基础架构,网络末端通过单一卷积核的全连接层输出人群计数结果,但该模型应用于新场景时,计数准确性下降明显,鲁棒性较差^[13]。Zhang C等提出基于CNN的跨场景人群计数方法,通过一种数据驱动的方法来微调目标场景的受训CNN模型,使其适应一个未知的人群场景,该方法

通过两个相关的学习目标(人群密度和人群数量)交替训练,以获得更好的人群计数效果,但难以提取尺度不变特征,在拥挤场景中效果不佳,同时在训练和测试时对透视图的需求也限制了该方法在实际人群场景中的应用^[14]。至此,如何提取有效的多尺度特征,进而获得精确的人群计数结果,成为人群计数问题的难点。为解决此问题,研究者们提出了许多不同结构的CNN模型。Zhang Y等首次提出一种多列卷积神经网络(MCNN)模型,通过多列CNN结构并行提取多尺度信息,一定程度上解决了多尺度特征提取问题,大大提升了人群计数的精度^[15]。受此多列结构的启发,Sam等提出Switching CNN网络,该网络先将图像分块,然后通过VGG16网络^[16]将图像块按人群密集程度分成高中低共3类,最后针对不同密度类别采用不同CNN回归模型估计图像的人群密度图,回归计算人员数量,该网络精度比MCNN高,但结构庞大,计算复杂度较高^[17]。Zeng等提出一种基于多尺度块的人群密度估计(MSC-NN)模型,侧重解决多模型组合和多栏组合模型存在的模型优化困难问题,简化了模型训练步骤,但多卷积核提取到的特征经 1×1 的卷积融合,损失了底层特征,进而影响了计数准确度^[18]。

由上述研究可知,多列结构为提取多尺度特征提供了一种有效手段,在一定程度上解决了人群计数存在的尺度变化问题,但不同CNN学习到的多尺度人群特征如何在保证信息不丢失的情况下充分融合利用,提高输出密度图质量,仍是多列结构没有解决的难题。有鉴于此,本文提出一种用于人群计数任务的编码-解码结构的多尺度卷积神经网络。编码器采用多列卷积捕获多尺度特征,通过空洞卷积和空间金字塔池化结构扩大感受野并减少参数量,保留尺度特征和图像的上下文信息;解码器对编码器输出进行上采样,实现高层语义信息和编码器前端低层特征信息的有效融合,从而提升密度图的输出质量。为进一步提升计数精度,提出一种包括

人群计数误差描述的新型损失函数,提升了网络对计数的敏感性。为使模型充分训练,采用数据增强来扩展训练集。在 Shanghai Tech^[15]和 Mall^[1]以及一个自建数据集上的实验结果表明,本文方法较好地提高了计数精度。

1 基于编码-解码的多尺度卷积神经网络

1.1 密度图

完成人群计数任务的前提是获得输入图像的人群密度图。密度图是以特殊高亮的形式显示人群分布位置的图示,其积分值代表图像中的人数。若图像 x 中第 i 个人头标记点的坐标为 x_i ,则密度图可以表示为函数 $\delta(x-x_i)$ 。对于一张具有 N 个人头标记点的图像,密度图可表示为

$$H(x) = \sum_{i=1}^N \delta(x-x_i) \tag{1}$$

使用高斯核滤波器 G_σ ^[19]与式(1)进行卷积,得到密度方程

$$F(x) = \sum_{i=1}^N \delta(x-x_i) * G_\sigma \tag{2}$$

式中 $*$ 代表卷积运算。

由透视原理可知,在三维实际人群场景被二维彩色图像表示的过程中,每一个像素点所表示的人头所占面积会存在较大差异,使用式(2)的生成方式会导致生成的密度图不够准确。本文将图像的视角扭曲考虑到密度图的生成中,采用自适应高斯滤波器与式(1)进行卷积,得到

$$F(x) = \sum_{i=1}^N \delta(x-x_i) * G_{\sigma_i(x)}, \sigma_i(x) = \beta \bar{d}_i \tag{3}$$

式中 $\bar{d}_i$ 是标记点 x_i 与其最近的 K_0 个人头之间的平均距离。经大量实验可知,当参数 $\beta=0.3$ 时生成

的密度图质量最好^[15]。

1.2 编码-解码多尺度卷积神经网络的结构

由于池化层的存在,输入图像在卷积的过程中会周期性的进行下采样操作,而下采样会损失部分低级特征,降低图像分辨率。以往用于人群计数的卷积神经网络没有考虑低级特征对于人群预测密度图生成的影响。编码-解码结构在目标检测中的应用^[20-21]证明了其有助于高低级特征融合,从而实现更精确的预测。有鉴于此,本文提出通过编码-解码结构实现人群图像的高低级特征充分融合,获取更高质量的预测密度图,进而得到准确的人群计数结果。

本文网络结构如图 1 所示。编码器通过多尺度卷积模块捕获多尺度特征,采用空洞空间金字塔池化(ASPP)^[22]聚合尺度特征,避免细节信息的丢失。解码器对编码器的输出(高级特征)进行上采样,并与前半层的输出(低级特征)经 1×1 的卷积来匹配通道数,之后进行融合,再经过一个 3×3 的卷积微调后,获得预测密度图。

1.3 多尺度卷积

由于视角失真,人群图像常存在较大的尺度变化,而单列网络难以应对人群尺度变化。因此,本文设计了一种基于多列结构的多尺度卷积模块,结构如图 2 所示。在多尺度卷积模块中,沿用 MSCNN 模型思想,采用多尺度卷积块从原始图像中学习与尺度相关的特征,实现多尺度信息的有效获取。多尺度卷积块结构如图 3 所示。

MSCNN 模型多尺度卷积块采用 4 个并列的 $3\times 3, 5\times 5, 7\times 7, 9\times 9$ 卷积核来获取尺度信息,实验发现,将 9×9 的卷积核替换为 1×1 的卷积核可以更好地聚合上下文语义信息并减少参数量,故本文多尺度卷积块选用 $1\times 1, 3\times 3, 5\times 5, 7\times 7$ 的卷积核,从而避免了信息丢失并减少了参数量。

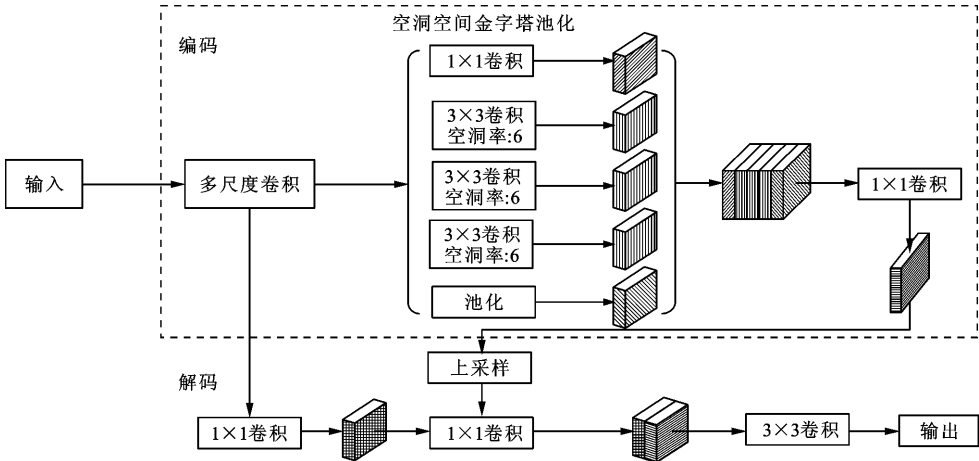


图 1 本文网络结构

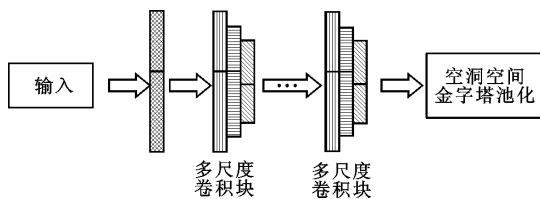


图2 多尺度卷积模块结构

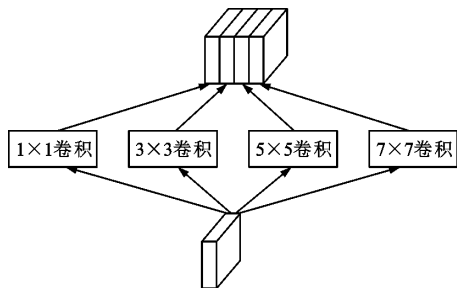


图3 多尺度卷积块结构

1.4 空洞空间金字塔池化

在密集人流量场景中,人群遮挡情况严重,卷积网络池化层的下采样操作会造成图像分辨率降低,丢失人群部分信息,从而导致预测密度图质量下降,影响人群计数准确性。空洞卷积^[23]可以通过在标准卷积核中添加空洞,即在空洞的位置添加0,达到扩大卷积核尺度,增大感受野的目的,且不增加参数运算量。由此,本文构建了一个ASPP模块,利用4个并行的不同空洞率的空洞卷积核对不同尺度特征进行重采样,揭示图像的上下文信息。在空洞卷积尾部串联空间金字塔结构^[24],将不同大小的特征图映射为相同维度的空间信息,然后融入图像表示中,以更好地完成密度图预测任务。

空洞卷积定义为

$$(a *_l w)[i] = \sum_{k=1}^K a[i + kl]w[k] \quad (4)$$

式中: w 表示卷积核; K 表示卷积核尺寸; $w[k]$ 表示大小为 k 的卷积核; $a[i]$ 表示第 i 个输入; $*_l$ 表示空洞卷积运算; l 表示空洞率,描述卷积核处理数据时采样的步幅,通过调整 l 即可调整感受野大小。

不同空洞率下感受野的变化如图4所示。图4a表示标准的 3×3 卷积,其感受野仅为 3×3 ;图

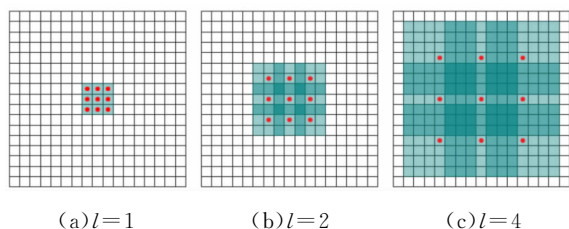


图4 不同空洞率下感受野的变化

4b表示空洞率为2的 3×3 扩张卷积,其感受野可达 7×7 ;图4c表示空洞率为4的 3×3 扩张卷积,其感受野可达 15×15 。

2 损失函数

像素空间损失是人群计数方法的常用损失函数^[25-26],其表达式为

$$L_D(\theta) = \frac{1}{M} \sum_{i=1}^M \|F_D(K_i, \theta) - D_i\|^2 \quad (5)$$

式中: D_i 表示第 K_i ($i=1, \dots, M$)幅训练样本图像的真实值密度图; $F_D(K_i, \theta)$ 表示第 K_i ($i=1, \dots, M$)幅训练样本图像的网络输出,即预测密度图; θ 表示网络的可学习参数; M 表示训练图像数。

由式(5)可知,像素空间损失函数仅反映了估计的密度图像素级别的准确程度,而人群计数任务的最终目的是要回归计算得到图像中的总人数。因此,本文提出了一种新型损失函数,在以往像素空间损失的基础上考虑了计数误差,以此来增强网络对计数的敏感性。考虑计数误差的计数损失为

$$L_Y(\theta) = \frac{1}{M} \sum_{i=1}^M \left| \frac{F_Y(K_i, \theta) - Y_i}{Y_i} \right|^2 \quad (6)$$

式中: $F_Y(K_i, \theta)$ 表示第 K_i ($i=1, \dots, M$)幅训练样本图像经由预测密度图 $F_D(K_i, \theta)$ 积分求和后得到的预测人数; Y_i 表示第 K_i ($i=1, \dots, M$)幅训练样本图像的真值人数。

由此,得到本文的总损失函数为

$$L(\theta) = (1 - \lambda)L_D(\theta) + \lambda L_Y(\theta) \quad (7)$$

式中: λ 表示计数损失的权值,本文根据实验将其设为0.5。

3 实验及分析

在Shanghai Tech^[15]、Mall^[11]及自建数据集进行了实验与分析。对比方法采用文献[14]方法、MCNN^[15]、Switch-CNN^[16]以及MSCNN^[18]。

实验在Ubuntu系统下进行,GPU型号为TitanV,环境配置为CUDA9.0+anaconda3+python3+tensorflow1.8.0。本文方法所有层均使用标准差为0.01的高斯分布进行初始化,网络初始训练学习率为0.01,迭代次数为100 000。

此外,为使模型充分训练,本文采用数据增强方法,对样本图像随机进行裁剪、旋转、放缩等操作,扩充数据集样本数,增强CNN模型的鲁棒性。

3.1 评价指标

由于文献[14]方法、MCNN^[15]、Switch-CNN^[16]

以及 MSCNN^[18]均采用平均绝对误差和均方误差作为评价指标,所以本文方法也采用这些评价指标以保持一致性。

平均绝对误差为网络预测人数与图像真实人数之间的误差,其定义为

$$f_{MAE} = \frac{1}{N_1} \sum_{i=1}^{N_1} |C_i - C_i^{GT}| \tag{8}$$

式中: N_1 为测试图像数; C_i 为第 i 幅测试图像的预测人数; C_i^{GT} 为第 i 幅测试图像的真实人数。

均方误差为网络预测人数与图像真实人数之间的差异程度,其定义为

$$f_{MSE} = \sqrt{\frac{1}{N_1} \sum_{i=1}^{N_1} (C_i - C_i^{GT})^2} \tag{9}$$

3.2 Shanghai Tech 数据集实验与分析

Shanghai Tech 数据集共包含 1 198 幅图像,共 330 165 个已标记人头,是目前已知标记人数最多的数据集。数据集分为两部分:①Part_A,包含 482 幅图像,来源于互联网;②Part_B,包含 716 幅图像,来源于上海街道。将 Part_A 的 300 幅和 Part_B 的 400 幅图像用于训练,其他的用于测试。在 Shanghai Tech 数据集中,本文方法单幅图像实验结果、测试集计数结果分别如图 5 和图 6 所示,不同方法性能指标对比结果如表 1 所示。

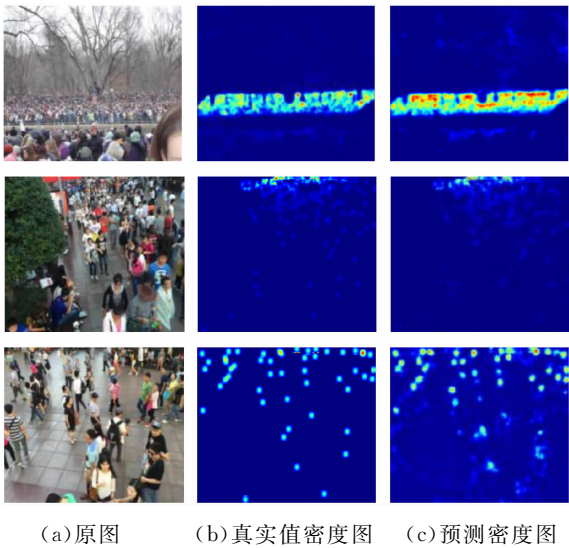


图 5 Shanghai Tech 数据集的本文方法单幅图像实验结果

从图 6 可以看出,单幅图像中若人数超过 1 500,则本文方法性能有所下降,分析其原因主要为训练集中高密度图像数较少,模型未能学到足够的特征。由表 1 可知,本文方法与目前准确率较高的 MSC-NN 方法相比, f_{MAE} 提升了 7.3, f_{MSE} 提升了 27.2,表明本文方法具有较高准确度和鲁棒性。

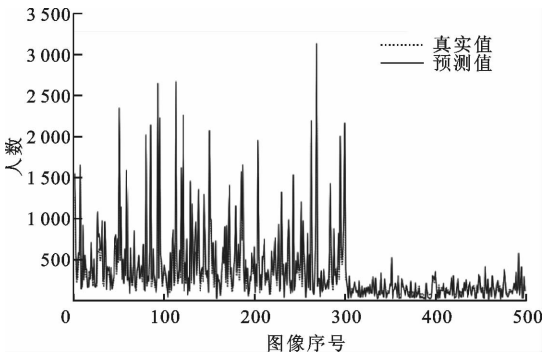


图 6 Shanghai Tech 数据集测试集本文方法计数结果

表 1 Shanghai Tech 数据集的不同方法性能指标对比结果

方法	Part_A		Part_B	
	f_{MAE}	f_{MSE}	f_{MAE}	f_{MSE}
文献[14]方法	181.8	277.7	32.0	49.8
MCNN ^[15]	110.2	173.2	26.4	41.3
Switch-CNN ^[16]	90.4	135.0	21.6	33.4
MSCNN ^[18]	83.8	127.4	17.7	30.2
本文方法	76.5	100.2	15.4	26.3

3.3 Mall 数据集实验与分析

Mall 数据集是使用安装在购物中心的监视摄像机收集到的数据集。该数据集具有人群密度变化大、活动模式多、透视畸变以及遮挡严重的特点。实验选用 1 000 幅图像作为训练集,1 000 幅图像作为测试集。在 Mall 数据集中,本文方法单幅图像实验结果、测试集计数结果分别如图 7 和图 8 所示,不同方法性能指标对比结果如表 2 所示。

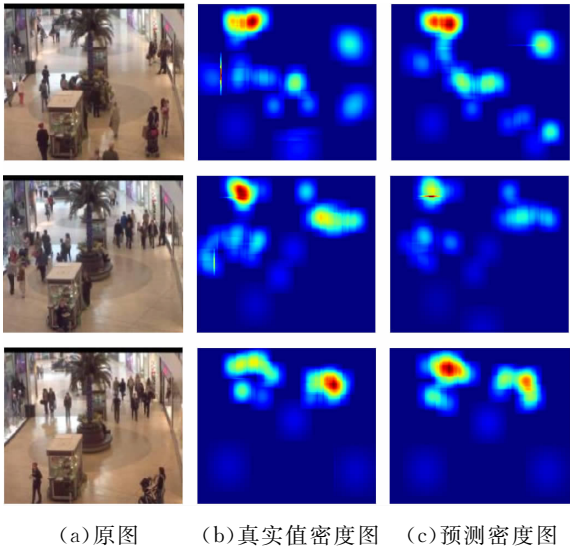


图 7 Mall 数据集的本文方法单幅图像实验结果

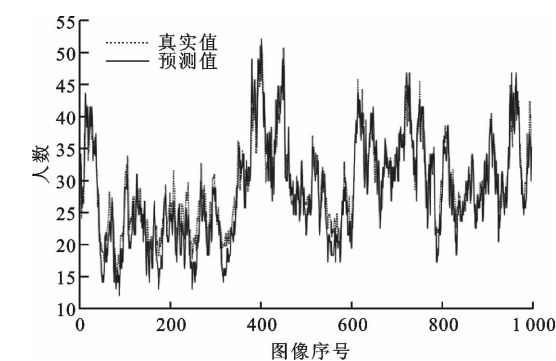


图 8 Mall 数据集测试集的本文方法计数结果

表 2 Mall 数据集的不同方法性能指标对比结果

方法	f_{MAE}	f_{MSE}
文献[14]方法	3.15	15.7
MCNN ^[15]	2.21	7.33
Switch-CNN ^[16]	2.01	6.25
MSCNN ^[18]	2.12	7.04
本文方法	1.80	5.36

由表 2 可以看出,在 Mall 数据集中,本文方法的 f_{MAE} 与 MSCNN 相比有了 0.32 的提升,表明本文方法在复杂多变的室内场景下也有较好的精度。同时,本文方法的 f_{MSE} 优于表中其他经典方法的,证明本文方法具有较高的鲁棒性。

3.4 自建数据集实验与分析

现有人群数据集多为室外场景或人群密度较大场景,Mall 数据集虽是室内场景人群数据集,但缺少场景变化,背景单一。因此,本文自建了一个具有背景复杂多变、人员流动性大的特点的商场数据集。

该数据集来源于万达商场内 3 个不同场景(电梯、大厅和商场区)。针对这 3 个场景,分别采集多天多时段(9:00—22:00)视频信息,每个场景选取 400 幅图像,共计选取 1 200 幅图像构成本文自建数据集。本文自建数据集的 3 个场景如图 9 所示。自建数据集图像未进行预处理,保留了原始图像所具有的噪声以及亮度变化干扰。

本文自建数据集与 Shanghai Tech 及 Mall 数据集采用一致的标注方式。首先,手动标注数据集每幅图像的人头坐标点,图像边缘的不完整人体不做标注,人头标记点数即为图像的真实人数;之后,根据式(1)和式(3),对人头坐标点标注图像进行计算,获得图像对应的真实值密度图。

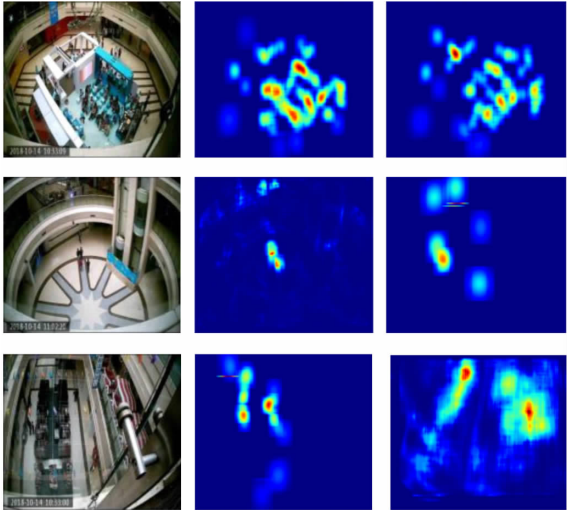
数据集共有 1 200 幅图像,实验选用其中的 900 幅(3 种场景各 300 幅)图像用来训练网络模型,其他的 300 幅(3 种场景各 100 幅)图像用来进行测试。



(a)大厅 (b)商场区 (c)电梯

图 9 自建数据集的 3 个场景

试。在自建数据集中,本文方法单幅图像实验结果、测试集计数结果分别如图 10 和图 11 所示,不同方法性能指标对比结果如表 3 所示。



(a)原图 (b)真值密度图 (c)预测密度图

图 10 自建数据集的本文方法单幅图像实验结果

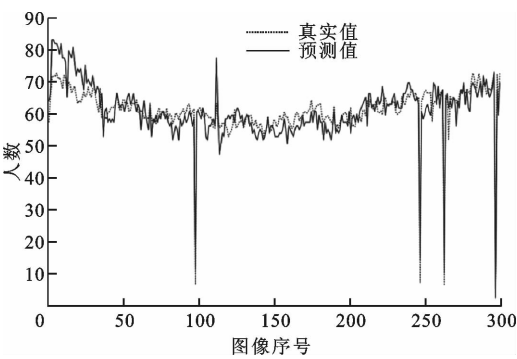


图 11 自建数据集测试集的本文方法计数结果

表 3 自建数据集的不同方法性能指标对比结果

方法	f_{MAE}	f_{MSE}
文献[14]方法	12.38	20.39
MCNN ^[15]	7.69	12.36
Switch-CNN ^[16]	5.23	10.20
MSCNN ^[18]	6.35	10.25
本文方法	4.52	9.15

从图 11 可以看出,部分数据误差较大(第 1~

30 幅图像),原因是此部分图像遮挡非常严重,且有假人模特等干扰,后续会跟进此问题并进行相关改进。由表 3 可以看出,在自建数据集中,本文方法与表 3 中其他经典算法相比也具有较高的准确度和泛化能力。综合各方法性能可知,本文方法能适应复杂背景,鲁棒性较好。

由 3.2~3.4 节中 3 个数据集上的实验结果可看出,本文方法无论在 f_{MAE} 指标还是 f_{MSE} 指标上均优于现有方法,分析原因有 2 点:①对比方法虽都是多列结构,但采用的融合方式会导致尺度信息丢失,而本文方法通过 ASPP 模块在融合多尺度信息时较好地保留了尺度特征;②对比方法使用高级特征,对人群有更本质的刻画,但损失了纹理、轮廓等底层特征,而本文方法采用的编码-解码结构保留了底层特征,生成了更精确的预测密度图。

3.5 模型大小与运行时长分析

为全面评估本文方法,本小节从训练模型大小和运行时长两方面进行了对比实验,结果如表 4 所示。

表 4 模型大小和运行时长对比

方法	模型大小 /MB	平均单幅测试图像运行时长/s		
		Shanghai Tech	Mall	自建
文献[14]方法	7.1	2.36	0.32	1.22
MCNN ^[15]	19.2	2.31	0.32	1.21
Switch-CNN ^[16]	32.2	2.41	0.34	1.30
MSCNN ^[18]	22.2	2.34	0.32	1.24
本文方法	13.6	2.32	0.31	1.21

由表 4 可以看出:Switch-CNN^[16]模型最大,运行时长也最长;文献[14]方法模型较小,但其采用了全连接层导致运行时长较长;MCNN^[15]模型采用了多列结构,MSCNN^[18]模型采用了尺寸较大的卷积核,导致模型较大,运行时间较长。

相较而言,本文方法模型较小,且运行时长较短,分析原因有 2 点:①本文方法运用空洞卷积,在扩大感受野的基础上大大减少了参数量;②本文方法的多尺度卷积块采用 1×1 的卷积核,在保留上下文信息的同时降低了参数量。

3.6 验证性实验与分析

为验证本文所提出的编码-解码结构、ASPP、新型损失函数以及数据增强的有效性,在 3 个数据集上进行了验证性实验,结果如表 5 和表 6 所示。

由表 1~3 和表 5~6 可看出,编码-解码结构、ASPP 和新型损失函数皆对人群计数任务有贡献。其中,编码-解码结构以及 ASPP 对模型性能提升最

为明显,新型损失函数效果次之,当三者组合使用时,方法性能最好。同时可以看出,当没有数据增强扩充训练数据量时,模型极易陷入过拟合,导致方法性能达不到最佳。本小节实验结果表明,本文方法各模块有效地提高了人群计数的准确性。

表 5 Shanghai Tech 数据集验证性实验结果

方法	Part_A		Part_B	
	f_{MAE}	f_{MSE}	f_{MAE}	f_{MSE}
无计数损失、无 ASPP 的 本文方法	79.9	136.2	17.3	28.4
无计数损失的本文方法	77.3	126.2	16.3	27.9
无 ASPP 的本文方法	79.3	139.0	16.9	28.5
无数据增强的本文方法	116.7	192.0	32.5	40.9
本文方法	76.5	100.2	15.4	26.3

表 6 Mall 及自建数据集验证性实验结果

方法	Mall		自建数据集	
	f_{MAE}	f_{MSE}	f_{MAE}	f_{MSE}
无计数损失、无 ASPP 的 本文方法	1.96	6.32	5.36	10.10
无计数损失的本文方法	1.89	6.14	5.02	9.25
无 ASPP 的本文方法	1.92	6.22	5.16	9.38
无数据增强的本文方法	3.62	9.14	10.36	18.31
本文方法	1.80	5.36	4.52	9.15

4 结 论

本文提出了一种基于编码-解码结构的多尺度人群密度估计网络模型,解决了多列卷积神经网络人群计数方法存在的多尺度特征信息丢失、融合不佳以及密度图质量不高的问题。本文网络采用改进的多列结构来提取多尺度特征,并在其尾部串联了一个空洞空间金字塔模块来聚合尺度特征,有效地避免了细节信息的丢失。通过编码-解码结构实现了高层语义信息和编码器前端低层特征信息的有效融合,进而提升了输出密度图质量。本文还提出了一种包含人群计数误差的新型损失函数,提升了网络对计数的敏感性。

为使模型充分训练,实验部分采用数据增强来扩展训练集。针对商场具有背景复杂多变、人员流动性大的特点,自建了新的商场数据集用以进行方法验证。Shanghai Tech、Mall 以及自建数据集上的实验结果表明:①本文方法在不同人群场景下都具有很好的密度估计和人群计数性能,泛化能力较

强;②本文方法各模块对人群计数任务都有贡献作用;③本文方法模型较小,且运行时长较短。

综上可知,本文方法在人群计数任务上具有较优越的性能,提高了计数精度,但在处理高遮挡场景时效果依然不理想,这也是后续改进的重点。

参考文献:

- [1] CHEN K, LOY C C, GONG S, et al. Feature mining for localised crowd counting [C]//Proceedings of the 2012 British Machine Vision Conference (BMVC). Guildford, UK: BMVA Press, 2012: 3-13.
- [2] CAI Zebin, YU Zhu Liang, LIU Hao. Counting people in crowded scenes by video analyzing [C]//Proceedings of the 2014 IEEE 9th Conference on Industrial Electronics and Applications (ICIEA). Piscataway, NJ, USA: IEEE, 2014: 1841-1845.
- [3] CHEN T Y, CHEN C H, WANG D J, et al. A people counting system based on face-detection [C]//Proceedings of the 4th International Conference on Genetic and Evolutionary Computing (ICGEC). Piscataway, NJ, USA: IEEE, 2011: 699-702.
- [4] MARANA A N, VELASTIN S A, COSTA L F, et al. Estimation of crowd density using image processing [C]//Proceedings of the 1997 IEE Colloquium on Image Processing for Security Applications. Stevenage, UK: IEE, 1997: 11-1-11-8.
- [5] KILAMB P, RIBNICK E, JOSHI A J, et al. Estimating pedestrian counts in groups [J]. Computer Vision and Image Understanding, 2008, 110(1): 43-59.
- [6] CHEN K, GONG S, XIANG T, et al. Cumulative attribute space for age and crowd density estimation [C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2013: 2467-2474.
- [7] CHANGE L C, GONG S, XIANG T. From semi-supervised to transfer counting of crowds [C]//Proceedings of the 2013 IEEE International Conference on Computer Vision (CVPR). Piscataway, NJ, USA: IEEE, 2013: 2256-2263.
- [8] CHO S Y, CHOW T W S, LEUNG C T. A neural-based crowd estimation by hybrid global learning algorithm [J]. IEEE Transactions on Cybernetics, 1999, 29(4): 535-541.
- [9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Proceedings of the 2015 Annual Conference on Neural Information Processing Systems (NIPS). Vancouver, Canada: NIPS, 2015: 91-99.
- [10] 曹玉良, 明廷锋, 贺国, 等. 基于深度学习的离心泵空化状态识别 [J]. 西安交通大学学报, 2017, 51(11): 165-172.
CAO Yuliang, MING Yanfeng, HE Guo, et al. Artificial recognition of centrifugal pump cavitation status based on deep learning [J]. Journal of Xi'an Jiaotong University, 2017, 51(11): 165-172.
- [11] 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络 [J]. 自动化学报, 2016, 42(9): 1300-1312.
CHANG Liang, DENG Xiaoming, ZHOU Mingquan, et al. Convolutional neural networks in image understanding [J]. Journal of Automation, 2016, 42(9): 1300-1312.
- [12] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [13] 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. New York, USA: ACM, 2015: 1299-1302.
- [14] ZHANG C, LI H, WANG X, et al. Cross-scene crowd counting via deep convolutional neural networks [C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 833-841.
- [15] ZHANG Y, ZHOU D, CHEN S, et al. Single-image crowd counting via multi-column convolutional neural network [C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 589-597.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C/OL]//Proceedings of the 3rd International Conference on Learning Representations (ICLR). London, UK: ICLR, 2015. [2019-07-01]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [17] SAM D B, SURYA S, BABU R V. Switching convolutional neural network for crowd counting [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 4031-4039.
- [18] ZENG L, XU X, CAI B, et al. Multi-scale convolutional neural networks for crowd counting [C]//Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2017: 465-469.

- [19] LEMPITSKY V, ZISSERMAN A. Learning to count objects in images [C]//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada; NIPS, 2010: 1324-1332.
- [20] JOSEPH R, SANTOSH D, ROSS G. You only look once: unified, real-time object detection [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA; IEEE, 2016: 779-788.
- [21] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA; IEEE, 2017: 7263-7271.
- [22] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]// Proceedings of the European Conference on Computer Vision (ECCV). Berlin, Germany; Springer, 2018: 801-818.
- [23] 张焯林, 赵建伟, 曹飞龙. 构建带空洞卷积的深度神经网络重建高分辨率图像 [J]. 模式识别与人工智能, 2019, 32(3): 259-267.
- ZHANG Zhuolin, ZHAO Jianwei, CAO Feilong. Building deep neural networks with dilated convolutions to reconstruct high-resolution image [J]. Pattern Recognition and Artificial Intelligence, 2019, 32(3): 259-267.
- [24] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [25] 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 Zenglin, YE Yangdong, WU Yunpeng, et al. Crowd counting using rank-based spatial pyramid pooling network [J]. Journal of Automation, 2016, 42(6): 866-874.
- [26] 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.

(编辑 陶晴)

(上接第148页)

- [18] 文新宇, 王子豪, 王宝光, 等. 一种反向递推正弦干扰观测器的设计方法 [J/OL]. 控制与决策. [2019-10-01]. <https://doi.org/10.13195/j.kzyjc.2019.0769>.
- WEN Xinyu, WANG Zihao, WANG Baoguang, et al. Design method of reverse recursive sinusoidal disturbance observer [J/OL]. Control and Decision. [2019-10-01]. <https://doi.org/10.13195/j.kzyjc.2019.0769>.
- [19] 韩乃玉, 李志刚. 基于干扰观测器的弹仓趋近自适应滑模控制 [J/OL]. 兵器装备工程学报. [2019-10-01]. <http://kns.cnki.net/kcms/detail/50.1213.TJ.20190910.1512.020.html>.
- HAN Naiyu, LI Zhigang. Approach adaptive sliding mode control of magazine based on disturbance observer [J/OL]. Journal of Ordnance Equipment Engineering. [2019-10-01]. <http://kns.cnki.net/kcms/detail/50.1213.TJ.20190910.1512.020.html>.
- [20] 王东委, 富月. 基于高阶观测器和干扰补偿控制的模型预测控制方法 [J/OL]. 自动化学报. [2019-10-01]. <https://doi.org/10.16383/j.aas.c180697>.
- WANG Dongwei, FU Yue. Model predict control method based on higher-order observer and disturbance compensation control [J/OL]. Acta Automatica Sinica. [2019-10-01]. <https://doi.org/10.16383/j.aas.c180697>.
- [21] 贺同福, 吴忠. 基于复合干扰观测器的CMG框架系统高精度转速控制 [J]. 系统科学与数学, 2019, 39(4): 495-506.
- HE Tongfu, WU Zhong. High-precision speed control of CMG frame system based on compound disturbance observer [J]. Journal of Systems Science and Mathematical Sciences 2019, 39(4): 495-506.

(编辑 陶晴)