基于轻量化卷积神经网络的人数估计算法研究

林园园[1,2],杨会成[1,2],胡耀聪[1,2]
1. 安徽工程大学 电气工程学院,安徽 芜湖 241000
2. 高端装备先进感知与智能控制教育部重点实验室,安徽 芜湖 241000

摘 要:目的 目前人群计数模型中存在两种问题:复杂的重型计数模型虽然计数性能较强,但模型参数量和计算量过大,因此实用性不高;当前的轻量化模型虽然降低了模型的复杂度,但计数性能不佳。针对以上问题,提出一种有效均衡计数性能和计数效率的基于轻量化卷积神经网络的人群计数模型。**方法** 该方法分为两个模块:特征提取模块和密度图回归模块。首先,在特征提取模块打破以往提取特征时丢弃高度相似信息的思想,更加注重本征特征和相似特征的融合,设计了一个轻量化线性映射单元,在减少网络参数和计算成本的同时,提高了计数精度;然后,由多个线性映射单元组成轻量化线性映射块,并串行多个线性映射块组成特征提取模块;接着,将特征提取模块提取到的特征馈送到密度图回归模块,密度图回归模块不再使用较少的标准卷积来回归密度图,而是使用扩张卷积来替代标准卷积,利用堆叠的扩张卷积来增加感受野从而得到更加精确的回归密度图;最后将回归密度图求和得到估计人数。**结果** 所提方法的参数量仅有 0.12 MB(Mbyte),计算量仅有 9.23 GFLOPS(Giga Floating-point Operations per Second),与其余轻量化人群计数模型相比均有降低,且在 3 个人群计数数据集,即 Shanghai Tech 数据集、UCF-QNRF 数据集、NWPU-Crowd 数据集都取得了较为优异的计数性能。**结论** 模型在保证计数性能的同时也保证了计数效率,实现了两者的最佳平衡,并实现了实时快速精确的人群计数,相较于其他轻量级人群计数算法,拥有更高的计数性能和计数效率,更具备实用性。

关键词:轻量级卷积神经网络;人群计数;特征融合;密度图估计
中图分类号:TP391.41;TP183 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2026.0001.004

Research on Crowd Counting Algorithm Based on Lightweight Convolutional Neural Network

LIN Yuanyuan[1,2], YANG Huicheng[1,2], HU Yaocong[1,2]
1. School of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, Anhui, China
2. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Wuhu 241000, Anhui, China

Abstract: Objective There are two problems in current crowd counting models. Complex heavy-weight counting models have strong counting performance, but their large number of model parameters and high computational complexity result in low practicality. Current lightweight models reduce the model complexity but have poor counting performance. To address these issues, a crowd counting model based on a lightweight convolutional neural network is proposed to effectively balance counting performance and counting efficiency. **Methods** The method consisted of two modules: the feature extraction

收稿日期:2023-11-29 **修回日期:**2024-02-08 **文章编号:**1672-058X(2026)01-0028-11
基金项目:国家自然科学基金青年项目资助 (62203012).
作者简介:林园园(1999—),女,安徽芜湖人,硕士研究生,从事深度学习和人群统计研究.
通信作者:杨会成(1970—),男,安徽来安人,教授,从事信号处理和计算机视觉研究. Email:18315388656@163.com.
引用格式:林园园,杨会成,胡耀聪. 基于轻量化卷积神经网络的人数估计算法研究[J]. 重庆工商大学学报(自然科学版),2026,43(1):28-38.
LIN Yuanyuan, YANG Huicheng, HU Yaocong. Research on crowd counting algorithm based on lightweight convolutional neural network[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(1): 28-38.

module and the density map regression module. First, in the feature extraction module, instead of discarding highly similar information during feature extraction as in the past, more attention was paid to the fusion of intrinsic features and similar features. A lightweight linear mapping unit was designed, which improved the counting accuracy while reducing network parameters and computational costs. Then, multiple linear mapping units formed a lightweight linear mapping block, and multiple linear mapping blocks were serially connected to form the feature extraction module. Next, the features extracted by the feature extraction module were fed into the density map regression module. Instead of using a small number of standard convolutions to regress the density map, the density map regression module used dilated convolutions to replace standard convolutions. Stacked dilated convolutions were utilized to increase the receptive field, and a more accurate regression density map was obtained. Finally, the estimated number of people was obtained by summing the regressed density map. **Results** The proposed method had a model size of only 0. 12 MB and a computational cost of only 9. 23 GFLOPS (Giga Floating-point Operations per Second), both of which were lower than those of other lightweight crowd counting models. It also achieved excellent counting performance on three crowd counting datasets: the Shanghai Tech dataset, the UCF-QNRF dataset, and the NWPU-Crowd dataset. **Conclusion** The model ensures both counting performance and counting efficiency, achieving the best balance between the two. It realizes real-time, fast, and accurate crowd counting. Compared with other lightweight crowd counting algorithms, it has higher counting performance and efficiency, and is more practical.

Keywords: lightweight convolutional neural network; crowd counting; feature fusion; density map estimation

人群计数在图像处理领域一直都是重点研究任务,其目的是自动估计监控场景中的人数。随着城镇化进程的加快,会迎来很多大型人群聚集活动。随之而来的由人群过于密集导致的踩踏事故层出不穷[1]。因此对密集场景下人群数量进行估计成为一个重要的研究课题。

近年来,基于卷积神经网络的人群计数受到更多学者的关注,并提出了很多高质量的计数模型。SINDAGI 等[2]提出了一种上下文金字塔卷积神经网络模型,有效结合了人群输入图像的全局特征和局部特征,使其在人群密度变化较大的场景下也能够生成高质量的密度图;OH 等[3]采用重型经典卷积神经网络 ResNet-50[4]作为骨干网络来提取特征,用多个并行卷积来生成密度图,它通过不确定性地量化对模型的输出进行概率解释,提高了预测质量,明确处理了不确定输入和特殊情况;Li 等[5]采用复杂骨干网络进行特征提取和扩张卷积作为后端生成密度图;Liu 等[6]在 CSRNet 中插入情境感知模块(Context-Aware),使得不同人头大小的信息可以在特征图中反映;Meng 等[7]提出空间不确定性感知师生框架来侧重于高置信度区域信息;Wang 等[8]采用尺度树多样性增强器和一个多级辅助器来减轻现有方法因尺度水平不足造成的局限性。然而,为了提取更多特征,它们一般使用复杂的多尺度卷积作为主干,虽然成功捕获了不同规模的特征,但是不可避免地会产生大量网络参数和巨大的计算负担。综上,虽然目前这些方法都能生成高质量的密度图,达到很好的计数性能,但是模型的参数量和计算量非常大,在模型的训练过程中会消耗很多资源。在目前移动计算和边缘计算时代,对于边缘设备或嵌入式系统而言,在有限的计算预算下获得优秀的计数性能和计数效率几乎同等重要。

因此,如何获得一个高效的轻量化人群计数模型成为一个热点研究问题。目前研究中,也提出了一系列轻量级人群计数模型。Zhang 等[9]每一列都使用较少的具有不同大小卷积核的卷积来预测人群密度图;Sam 等[10]提出一个自顶向下的反馈卷积神经网络,利用两列不同卷积核的少量标准卷积生成密度图。然而,少量标准卷积层提取的人群特征非常有限,会导致最终计数的准确性不理想;Cao 等[11]采用转置卷积生成密度图;Gao 等[12]提出前/背景分割(Fore-/Background Segmentation ,FBS)对中间特征进行编码,以分割前景和背景;Ma 等[13]提出一个轻量级端到端网络用于人群计数,该方法利用尺度感知模块提取多尺度特征,然后将这些特征回归到密度图中,使用具有较小空间滤波器的级联卷积降低模型的复杂度;Liang 等[14]使用所提出的轻量级金字塔卷积模块进行多尺度特征提取;Yi 等[15]选择 MobileViT 模块作为网络骨干,以减少网络参数的数量和计算成本。然而这些模型为了降低复杂度,大多只注重特征的提取而忽视特征融合,这必然会导致计数精度的损失。综上,目前的轻量化人群计数网络在参数量和计算量上都有所降低,然而为了获得更快的运行速度,现有的轻量级网络常会使用较少的标准卷积或忽略特征的融合,避免一些相似的特征图,丢弃一些高度相似的特征。虽然降低了

模型的复杂度,但是计数精度也受到了影响。

为了有效解决上述问题,本文对轻量级人群计数模型进行了更深入的分析,设计了一种高效的基于轻量化卷积神经网络的人群计数模型。模型主要由两部分组成,分别为特征提取模块和密度图回归模块。特征提取模块由轻量化线性映射块搭建,线性映射块由多个线性映射单元组成,它打破了以往提取特征时丢弃高度相似信息的思想,更加注重特征的融合。在线性映射单元,首先通过少部分标准卷积来生成本征特征,再经过一系列简单线性映射生成更多的相似特征,最后再将两种特征进行融合。这样的操作不仅能保留更多的特征信息,提高计数精度,也大幅度减少了模型的参数量和计算量。在密度图回归模块,本文不再使用较少的标准卷积来回归密度图,而是使用扩张卷积来替代标准卷积,扩张卷积以较小的卷积核来获取更大的感受野,不仅可以减少参数量还能提取更多的人群特征,回归出更加精确的密度图。本文提出的模型与其余轻量化模型相比在参数和计算量上均有所减少,且计数性能在 3 个人群计数数据集上都取得了优异的成绩,取得了计数性能和计数效率的最佳平衡。

1 轻量化人群计数模型构建

模型搭建分为两个模块:特征提取模块和密度图回归模块。特征提取模块由本文设计的轻量化线性映射块搭建,密度图回归模块由多个扩张卷积组成。特征提取模块中线性映射块由多个线性映射单元组成,该映射单元首先使用少部分标准卷积提取本征特征,再通过一些简单线性映射得到更多的有效特征,然后将两者结合。这种线性映射单元在没有改变输入输出特征的情况下,大大减少了模型的参数量和计算量。在密度图回归模块,使用多个 3×3 小卷积核的扩张卷积来替代标准卷积。扩张卷积扩大了感受野且不增加参数数量,如果使用标准卷积想要达到相同的感受野,需要增加更多的卷积数量,从而参数量也会随之增加。模型结构如图 1 所示。下面将对两个模块进行详细介绍。

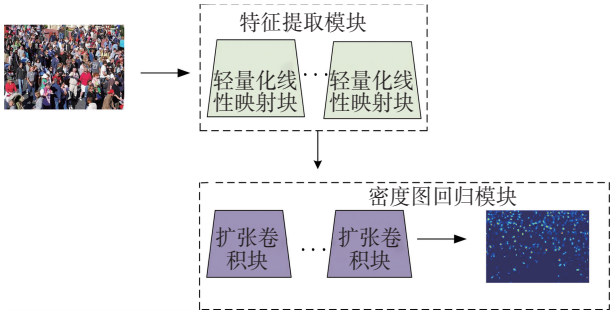


图 1 轻量化人群计数模型结构

Fig. 1 Structure of lightweight crowd counting model

1.1 特征提取模块

1.1.1 轻量化线性映射单元

轻量化线性映射块由多个轻量化线性映射单元搭建而成,首先介绍轻量化线性映射单元,结构如图 2 所示。由图 2 可知,映射单元一共分为 3 个步骤,首先是标准卷积生成本征特征,然后利用线性映射生成相似特征,最后将两种特征进行拼接。

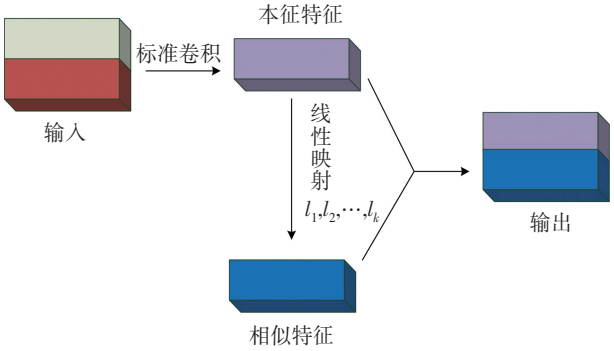


图 2 轻量化线性映射单元模型图

Fig. 2 Diagram of lightweight linear mapping unit model

(1) 本征特征生成。在深度学习网络中特征提取时都会包括本征特征和相似特征。由于内在的本征特征较小,只需用少量标准卷积来生成本征特征即可。假设总特征映射为 p 个,内在本征特征为 q 个($q<p$)。对于给定的输入图像 $I\in\mathbf{R}^{h\times w\times c}$($h$、$w$、$c$ 分别为输入图像的长宽以及通道数),q 个内在本征特征映射 $Q\in\mathbf{R}^{h'\times w'\times q}$ 是通过一次标准卷积生成的本征特征图,如式(1)所示:

$$Q=I*f \tag{1}$$

$f\in\mathbf{R}^{c\times k\times k\times q}$ 为使用的滤波器,$k\times k$ 为滤波核大小,"$*$"代表卷积,不包含偏置项。

(2) 相似特征生成。对于相似特征而言,没有必要使用大量的计算来逐个生成。

对上述生成的 Q,为了进一步得到想要的 p 个特征映射,对 Q 中的每个本征特征进行一系列线性操作生成 z 个相似特征:

$$z_{i,j}=l_{i,j}(q_i);\forall i=1,2,\cdots,q;j=1,2,\cdots,z \tag{2}$$

其中,式(2)的 q_i 为 Q 中第 i 个内在本征特征映射,$l_{i,j}$ 为上述函数中生成第 i 个相似特征映射的第 j 次线性操作,$z_{i,j}$ 表示第 i 个内在本征特征映射生成的第 j 次相似特征。也就是说,q_i 可以有一个或多个相似特征。

(3) 特征图拼接。最后就是对标准卷积生成的本征特征和线性映射的相似特征进行拼接生成最终的输出。

1.1.2 轻量化线性映射块

轻量化线性映射块主要由两个串联的轻量化线性映射单元组成。为了增加提取特征的速率,设计了步长不同的两个线性映射块,如图 3 所示,上面为步长为

1 的线性映射块,下面为步长为 2 的线性映射块。第一个轻量化线性映射单元用于扩大通道的数量。第二个轻量化线性映射单元减少了通道的数量以增加速率。批归一化(Batch Normalization, BN)[16] 和 ReLU(Rectified Linear Unit)非线性激活函数在每一层之后应用[17]。在图 3 的下半张图添加了步长为 2 的深度卷积。

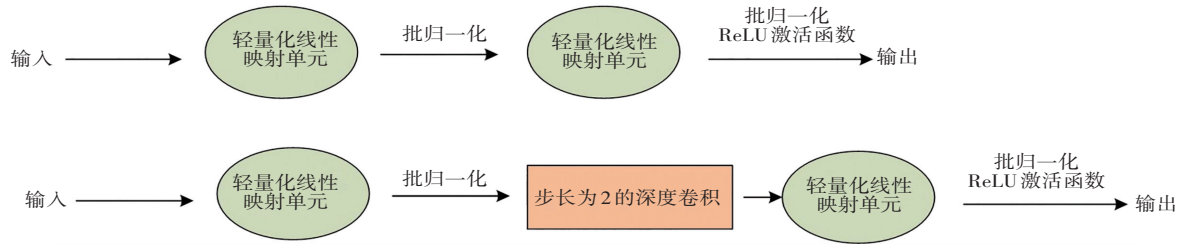


图 3　轻量化线性映射块模型图

Fig. 3　Diagram of lightweight linear mapping block model

1.1.3　特征提取模块模型搭建

本文所提出的特征提取模块模型基于轻量化线性映射块搭建而成,特征提取模块模型配置如表 1 前 7 层所示。表 1 中, SE[18](Squeeze-and-Excitation)表示挤压和激励模块,扩张因子是指轻量化线性映射块内部对通道数的扩张倍数,步长指卷积核在图片中每一步移动的距离。标准卷积层 1 表示卷积核为 3×3 的标准卷积,标准卷积层 2 和标准卷积层 3 表示卷积核为 1×1 的标准卷积,扩张卷积层 1 和扩张卷积层 2 表示卷积核为 3×3,扩张率为 2 的扩张卷积。

表 1　轻量化人群计数模型的结构配置

Table 1　Structural configuration of a lightweight crowd counting model

输入尺寸	操作方式	卷积核大小	扩张因子	输出通道	步　长	是否添加 SE[18] 模块
224×224×3	标准卷积层 1	3	无	16	2	否
112×112×16	轻量化线性映射块	3	1	16	1	否
112×112×16	轻量化线性映射块	3	2	24	2	否
56×56×24	轻量化线性映射块	3	3	24	1	否
56×56×24	轻量化线性映射块	5	1.8	40	2	是
28×28×40	轻量化线性映射块	5	3	40	1	是
28×28×40	轻量化线性映射块	3	3	80	1	否
28×28×80	标准卷积层 2	1	无	40	1	否
28×28×40	上采样层	无	无	无	无	否
56×56×40	扩张卷积层 1	3	无	40	2	否
56×56×40	上采样层	无	无	无	无	否
112×112×40	扩张卷积层 2	3	无	20	2	否
112×112×20	上采样层	无	无	无	无	否
224×224×20	标准卷积层 3	1	无	1	1	否

1.2　密度图回归模块

本文的密度图回归模块采用多个扩张卷积进行堆叠。从特征提取模块提取到的特征再送入密度图回归模块来生成估计密度图,最后对估计密度图求和得到最终的估计人数。

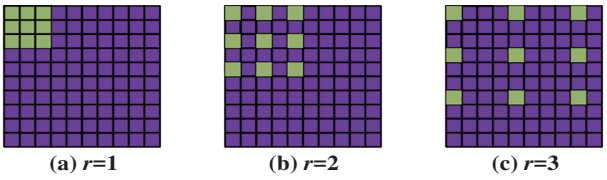
1.2.1　扩张卷积

与标准卷积相比,扩张卷积扩大了感受野[19]。能够捕获更深层次的位置导向信息。二维扩张卷积定义如式(3)所示,其中 $F(a,b)$ 是输入 $f(a,b)$ 与滤波器 $w(i,j)$ 扩张卷积的输出,a、b 分别为长度和宽度,r 表示扩张卷积的扩张速率,如果 $r=1$,则扩张卷积变为标准卷积。

$$F(a,b)=\sum_{i=1}^{M}\sum_{j=1}^{N}f(a+r\times i,b+r\times j)w(i,j) \quad (3)$$

目前,池化层如最大池化层和平均池化层,大多被用于进行特征压缩和防止过拟合,但它们也极大损耗了特征图的分辨率,导致提取到的特征映射图一部分空间信息丢失。然而,反卷积层虽然可以减轻信息丢失,但它的复杂度和执行时延在有些情况下也许并不适用。近年来,扩张卷积层在计算机视觉任务[20]上的准确率显著提高,它利用较小的卷积核就可以替代大卷积核的标准卷积或池化层,如图 4 所示。这种特性不仅减少了模型的参数量和计算量,还扩大了接受域。虽然可以使用更多的卷积层来产生更大的接受域,但是引入的操作也会更多。在扩张卷积中,一个具有 $s\times s$ 滤波器的小尺寸核被扩大到 $s+(s-1)(r-1)$,扩张率为 r,从而可以在保持相同分辨率的情况下灵活聚合多尺度

上下文信息。如图 4 所示,使用不同扩张率下的 3×3 卷积核得到的感受野也不同,正常卷积得到 3×3 感受野,$r=2$ 和 $r=3$ 的扩张卷积分别得到 5×5 和 7×7 的感受野。



(a) r=1 (b) r=2 (c) r=3

图 4 不同扩张率下的 3×3 卷积

Fig. 4 3×3 convolutions with different dilation rates

1.2.2 人群密度图生成

目前在人群计数领域,基于卷积神经网络的人数估计模型在回归人数阶段,包括直接估计出输入图像的人数和回归出密度图再求和得到人数这两种方法。前者输入的是图片,输出的是预测人数。后者输入图片,输出人群密度图,然后通过求和得到人数。在本文中,使用卷积神经网络对人群密度进行预测时,选择第二种方法,这一选择主要有两个原因:

首先,和直接生成预测人数相比,生成的密度图会保留更多有效的直观信息。例如,在密度图中可以直观观察到输入图像中的人群分布情况。这些分布情况在实际场景中有着重要的意义。如果在输出密度图中观察到某一小区域的人群密度远高于其他区域,则可能表明该区域发生了异常[21]。

其次,在使用卷积神经网络学习密度图的过程中,会加强卷积网络对不同大小人头的适应性。因此,即使输入的图片透视效果变化较大,也可以保证计数的准确性。

假设所有人群都提供逐点标注,即图像中的个体都进行了点标注(每个头部一个点),标记了 M 个头像的输入图像可以表示为一个函数:

$$P(x)=\sum_{i=1}^{M}\delta(x-x_i) \tag{4}$$

式(4)中,x 表示人群图像中的图像坐标,x_i 表示 x 中第 i 个注释点的坐标,δ 表示冲激函数。

为了将式(4)中的函数转换为连续密度函数,可以将该函数与高斯核 G_σ[22] 进行卷积,使密度 $F(x)=P(x)*G_\sigma(x)$。然而,在真实的三维场景中,这些图像中的人头并不都是独立存在的。在密集场景下,一个小区域内都可能存在很多标注人头。并且由于透视失真,在近处的人头所占区域会大一些,远处人头所占的区域会小一些。如果使用高斯核 G_σ 来生成密度图,会造成计数精度的损失。因此本文使用几何自适应核来生成密度图。

对于给定图像中的每个头像 x_i,将其与 c 个最近的人头距离表示为$\{l_1^i,l_2^i,\cdots,l_c^i\}$,平均距离为 $\bar{l}^i$,如式(5)所示。因此与 x_i 所对应的区域大小可以大致表示为图片地面上的一个半径为 $\bar{l}^i$ 的圆。

$$\bar{l}^i=\frac{1}{c}\sum_{j=1}^{c}l_j^i \tag{5}$$

为了估计邻近 x_i 的人群数量,需要将 $P(x)$ 与方差为 σ_i 的高斯核进行卷积。密度公式如式(6)所示,$\beta=0.3$,“$*$”表示卷积。图 5 是本文使用数据集中的部分图片利用几何自适应高斯核得到的密度图。

$$F(x)=P(x)*G_{\sigma_i}(x),\sigma_i=\beta\bar{l}^i \tag{6}$$

(a) Shanghai Tech 数据集部分图片 (b) 由(a)图生成的地面真实密度图 (c) UCF-QNRF 数据集部分图片 (d) 由(c)图生成的地面真实密度图

图 5 部分数据集图片几何自适应高斯核得到的地面真实密度图

Fig. 5 Ground truth density maps obtained with geometry-adaptive Gaussian kernels for some images in the dataset

1.2.3 密度图回归模块模型搭建

在密度图回归模块,使用两层扩张率为 2 的 3×3 卷积核的扩张卷积和两层 1×1 卷积核的标准卷积。在每层卷积后都添加 ReLU 激活函数。图 6 是密度图回归模块的结构示意图,其中 H 表示图片长度,W 表示图片高度,C 和 C'表示通道数。

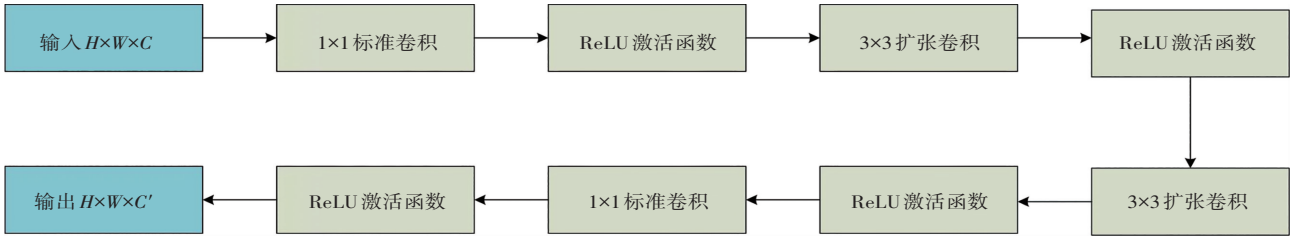


图 6 密度图回归模块的结构图

Fig. 6 Structure diagram of the density map regression module

1.2.4　轻量化人群计数模型结构配置

上述两节介绍了轻量级人群计数特征提取模块和密度图回归模块。为了提高计数精度,在密度图回归模块,要使输出的密度图恢复至原始输入分辨率大小,需要加入上采样操作。图 7 为本章设计的轻量化人群计数模型网络结构图。

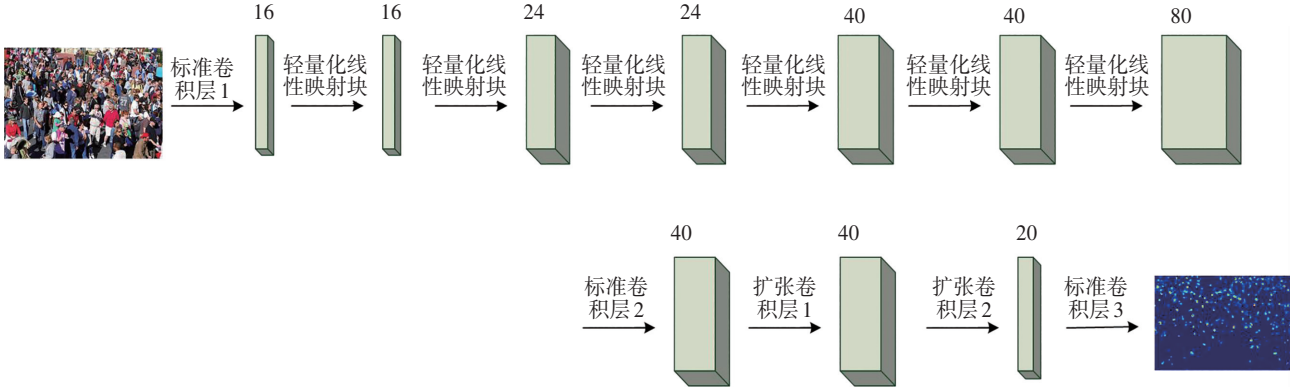


图 7　轻量化人群计数模型网络结构图

Fig. 7　Network structure diagram of lightweight crowd counting model

1.3　损失函数设计

本文模型在训练时采用欧几里得距离[23-24]来度量估计的密度图与地面真实密度图之间的差值,损失函数定义如式(7):

$$L_1(\theta)=\frac{1}{2N}\sum_{i=1}^{N}\|D(X_i;\theta)-D_g\|_2^2 \tag{7}$$

其中,θ 是模型中的一组可学习参数,N 为训练图像的个数,L_1 为估计密度图与地面真值密度图之间的损失,X_i 是输入图像,D_g 是图像 X_i 的地面真实密度图,$D(X_i;\theta)$ 表示由模型生成的估计密度图。

由于该模型训练的最终结果是希望预测人数能够接近于真实人数,因此还提出一个真实人数和估计人数之间的损失。首先将估计密度图 $D(X_i;\theta)$ 求和得到估计人数 C_e,定义如下:

$$C_e=\sum_{i=0}^{W}\sum_{j=0}^{H}D(X_i;\theta) \tag{8}$$

式(8)中,W、H 分别表示估计密度图的宽和长。真实人数和估计人数之间的损失 L_2 如式(9)所示:

$$L_2(\theta)=\frac{1}{2N}\sum_{i=1}^{N}\|C_e-C_g\|_2^2 \tag{9}$$

式(9)中,L_2 为估计人数与真实人数之间的损失,C_g 是真实人数,C_e 表示估计人数。

最终损失函数由上述两个损失函数加权求和得到,权重因子为 γ,γ 值为 0.1,如式(10)所示:

$$L_{\text{loss}}=L_1(\theta)+\gamma L_2(\theta) \tag{10}$$

2　仿真实验与结果分析

2.1　实验设置

本文研究的人群计数所用工作站是两个 NVIDIA RTX3090 显卡,处理器为 Intel Core-I7。在软件上使用的编程语言为 Python,版本为 3.7。模型的训练和测试阶段,全部都是在 pytorch 深度学习框架上进行的,版本为 1.6。在训练阶段,为了减少内存和增强模型的鲁棒性和多样性,首先使用随机裁剪从源人群图像中生成训练补丁。随机裁剪是使用一个固定大小为 224×224 的矩形方框在输入图片中随机位置处生成,是一种纯随机裁剪方式。由于裁剪时位置是随机的,所以每次裁剪的结果都不一样。如图 8 所示,白色的方框是每次迭代后在图片中随机位置处生成的,每迭代一次,就会产生一个随机矩形方框,从而就会产生一个随机子图。

这种随机裁剪方式因为增加了训练的多样性,因此要比顺序裁剪的效果更好。顺序裁剪是将输入图像按照固定尺寸,有顺序地进行裁剪,如图 9 所示,每次迭代生成一个规律的子图。为了证明训练过程中使用随机裁剪生成的训练模型比使用顺序裁剪生成的训练模型在测试阶段计数效果更好,本文 3.3 节中,在 ShanghaiTech[25]数据集上进行了实验对比分析。

图 8　随机裁剪示意图

Fig. 8　Schematic diagram of random cropping

图 9 顺序裁剪示意图

Fig. 9 Diagram of sequential cropping

接下来,使用 Adam 算法对本章提出的模型框架进行优化,默认学习率为 l_r 为 1e-5,权值衰减为 1e-4,批大小为 8,训练轮数为 500。损失函数为式(10)定义的 L_{loss}。本节在 ShanghaiTech、UCF-QNRF[26] 和 NWPU-Crowd[27] 3 个主流人群统计数据集上,对所提出模型的框架性能进行了评估。

评估的性能指标包括模型的复杂度和计数精确度。模型复杂度由参数量和计算量表示,模型计数精确度利用平均绝对误差(Mean Absolute Error, M_{ae})和均方误差(Mean-square Error, M_{se})[28]进行评估。

$$M_{\text{ae}} = \frac{1}{N}\sum_{i=1}^{N} \| C_e - C_g \|_1 \tag{11}$$

$$M_{\text{se}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \| C_e - C_g \|_2^2} \tag{12}$$

其中,M_{ae} 表示平均绝对误差,M_{se} 表示均方误差,C_e 和 C_g 分别表示第 i 个人群图像的预测值和真值,N 为测试数据集中的图像总数。

2.2 实验训练过程

本文提出的模型没有任何预训练模型可以使用,所以需要重新训练。训练时输入为选定的人群数据集,输出为模型的权重 λ,训练的大致过程如下:

(1) 初始化训练轮数 $i=0$,设置训练轮数为 500,当 i 达到 500 或达到设定的 M_{ae} 或 M_{se} 时,训练结束。

(2) 每执行一次第一步,将会得到预测密度图 $D(X_i;\theta)$ 和真实密度图 D_g、预测人数 C_e、真实人数 C_g。

(3) 根据第 2.3 节定义的损失函数计算出损失 $L_{\text{loss}}(D(X_i;\theta), D_g, C_e, C_g)$,$\Delta\lambda = L_{\text{loss}}$。

(4) 计算当前训练的 M_{ae} 和 M_{se}。

(5) 当训练的次数达到一个指定更新学习率的轮次时,更新学习率 $l_r = 0.5l_r$。

(6) 更新模型的权重 $\lambda = \lambda - l_r * \Delta\lambda$。

(7) 重复上述除(1)以外的所有步骤,当训练达到步骤(1)的条件时,停止训练。

2.3 实验结果和对比分析

本节的实验对比分析分为 3 个部分:第一部分对比 ShanghaiTech 数据集上,随机裁剪生成的训练模型和顺序裁剪生成的训练模型在测试阶段的计数能力;第二部分对比本文提出的轻量化计数模型与其他轻量化计数模型的参数量和计算量;第三部分验证本文提出的轻量化人群计数模型的计数能力,在经典的 3 个数据集上进行实验,并对实验结果与多个经典人数估计算法模型进行对比。下面对涉及的 3 个数据集进行简单介绍。

2.3.1 数据集介绍

(1) ShanghaiTech 数据集。ShanghaiTech 数据集中有 1 198 张图片,其中对 330 165 人的头部中心进行了注释[29],是一个数量较大的人群数据集。ShanghaiTech 数据集分为 A、B 两部分。A 部分(Part A)从互联网上随机抓取 482 张图片,B 部分(Part B)从上海大都市繁忙的街道上抓取 716 张图像。两部分之间的人群密度差异很大,A 部分数据集的人群密度明显高于 B 部分。此外,A 部分的场景变化较大,B 部分的场景较为固定。这使得人群计数的任务更加具有挑战性。A 部分和 B 部分都分为训练和测试两部分,其中 A 部分 300 张用于训练,剩下 182 张用于测试;B 部分 400 张用于训练,316 张用于测试。

(2) UCF-QNRF 数据集。UCF-QNRF 数据集包含 1 535 张图像,其中包含 1 201 张训练集和 334 张测试集。UCF-QNRF 数据集拥有高计数人群图像和注释,该数据集图片包含多样化的视角、不同的密度和不同的光照情况以及多变的场景[30]。另一方面, UCF-QNRF 数据集存在于野外拍摄的真实场景中,例如建筑、天空、植物等,也使得这个数据集更加逼真和具有挑战性。

(3) NWPU-Crowd 数据集。NWPU-Crowd 数据集是目前为止人群计数任务中数量最大的数据集,该数据集拥有 5 109 张图片和 2 133 238 个标注实体。数据集被分为训练集 3 109 张图片、验证集 500 张图片和测试集 1 500 张图片[31]。该数据集包含一些负样本,比如人群密度极高的图片,这样可以增强模型的鲁棒性。而且该数据集的分辨率较其他数据集相比也更高。该数据集对模型的计数性能提出了更高的要求。

2.3.2 实验对比分析

在本小节中,首先对比在 ShanghaiTech 数据集上随机裁剪生成的训练模型和顺序裁剪生成的训练模型在测试阶段的计数能力,实验对比结果如表 2 所示。由表 2 可知:在 ShanghaiTech 数据集上,训练过程中对图片使用随机裁剪方式生成的训练模型在测试阶段中的 M_{ae} 和 M_{se} 都更低,因为在训练过程中对图片进行随机裁剪相比顺序裁剪,增加了训练模型的多样性和鲁棒性。其次,评估本文提出的计数模型框架的性能,并与其他经典人数估计算法进行比较。比较的经典人数

估计算法分为两大类,一类是采用较为复杂的网络结构,另一类是采用轻量级或简单的网络结构,实验结果如表 3 所示。由于 UCF-QNRF 数据集和 NWPU-Crowd 数据集出现得较晚,且图片人群密度跨度较大,对于人群计数任务有很大的挑战性,因此一些人群计数模型还未在这两个数据集上进行实验。而复杂的网络结构一般不会在实验中进行参数量和计算量的计算,且一些轻量化计数模型也仅给出了参数量的计算。因此在表 3 中仅列出在对比模型中进行了实验的数据。

表 2　不同图片裁剪方式在测试阶段的实验结果(Shanghai Tech 数据集)

Table 2　Experimental results of different image cropping methods during the testing phase (Shanghai Tech dataset)

训练过程图片裁剪方式	Part A		Part B	
	M_{ae}	M_{se}	M_{ae}	M_{se}
随机裁剪	**68.1**	**106.3**	**9.3**	**15.9**
顺序裁剪	70.3	115.6	12.4	19.1

表 3　不同人数估计模型方法在 3 个数据集上的实验结果

Table 3　Experimental results of different crowd counting model methods on three datasets

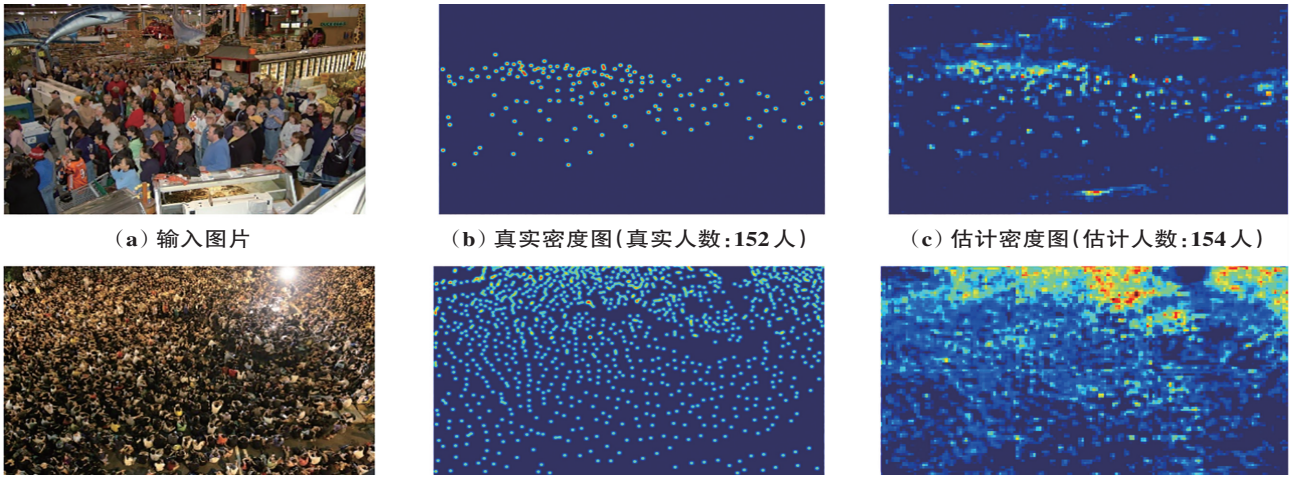
人数估计模型方法	地点及年份	Part A		Part B		UCF-QNRF		NWPU-Crowd		arams (M)	计算量 (GFLOPS)
		M_{ae}	M_{se}	M_{ae}	M_{se}	M_{ae}	M_{se}	M_{ae}	M_{se}		
CP-CNN[2]	ICCV17	73.6	106.4	20.1	30.1	—	—	—	—	68.4	—
CSRNet[5]	CVPR18	68.2	115.0	10.6	16	120.3	208.5	121.1	**387.8**	16.26	325.3
CAN[6]	CVPR19	**62.3**	**100**	**7.8**	**12.2**	**107.0**	**183.0**	**106.3**	**386.5**	18.1	193.58
DUB-Net[3]	AAAI20	**64.4**	106.8	7.7	12.5	**105.6**	**180.5**	—	—	18.05	—
SUA-Fully[7]	ICCV21	**66.9**	125.6	12.3	17.9	119.2	213.3	—	—	15.85	—
STNet[8]	TMM22	**52.9**	**83.6**	**6.3**	**10.3**	**87.9**	**166.4**	—	—	15.56	
MCNN[9]	CVPR16	110.2	173.2	26.4	41.3	277.0	426.0	232.5	714.6	0.13	11.87
TDF-CNN[10]	AAAI18	97.5	145.1	20.7	32.8	—	—	—	—	0.13	—
SANet[11]	ECCV18	75.3	122.5	10.5	17.9	152.6	247.0	190.6	491.4	0.91	71.45
LCNet[13]	ICIP19	93.3	149.0	15.3	25.2	—	—	—	—	0.86	—
PCC-Net[12]	TCSVT20	73.5	124.0	11.0	19.0	148.7	247.3	167.4	566.2	0.55	72.80
PDDNet[14]	APIN22	72.6	112.2	10.3	17.0	130.2	246.6	—	—	1.1	—
LMSFFNet[15]	TGRS23	85.85	139.9	**9.2**	**15.1**	112.8	201.6	—	—	4.58	14.9
Our model	—	68.1	106.3	9.3	15.9	108.6	184	115.4	450.6	**0.12**	**9.23**

由表 3 可知:本文提出的轻量级人群计数模型的参数量仅有 0.12 MB,计算量仅有 9.23 GFLOPS,对比其余轻量级的计数模型是最低的;在计数性能上,与轻量化计数模型相比,在 Part B 数据集上,本文模型的 M_{ae} 仅比最新的 LMSFFNet 模型高 0.1,M_{se} 仅高 0.8;而本文模型比 LMSFFNet 模型的参数量低 4.46,计算量低 5.67;在其余数据集上,本文模型与其他轻量化人群计数模型相比都取得了最佳的 M_{ae} 和 M_{se}。

与复杂的网络结构相比,在 Part A 数据集上,本文模型的 M_{ae} 比模型 CAN、DUB-Net、SUA-Fully 和 STNet 高 5.8、3.7、1.2 和 15.2,M_{se} 比模型 CAN 和 STNet 高 6.3 和 22.7。在 Part B 数据集上,本文模型的 M_{ae} 分别比模型 CAN、DUB-Net 和 STNet 高 1.5、1.6 和 3,M_{se} 分别比模型 CAN、DUB-Net 和 STNet 高 3.7、3.4 和 5.6。在 UCF-QNRF 数据集上,本文模型的 M_{ae} 分别比模型 CAN、DUB-Net 和 STNet 高 1.6、3 和 20.7,M_{se} 分别比模型 CAN、DUB-Net 和 STNet 高 1.0、3.5 和 17.6。在 NWPU-Crowd 数据集上,本文模型的 M_{ae} 比模型 CAN 高 9.1,M_{se} 分别比模型 CAN 和 CSRNet 高 67.8 和 64.1。

虽然和复杂的网络结构相比,在 3 个数据集上没有取得最佳的 M_{ae} 和 M_{se},但是,本文提出模型的参数量和计算量与这些复杂的网络结构相比大大减少,且对比其他轻量化的人群计数模型,计数性能最优异,模型的参数量和计算量最小。因此实验结果证明本文提出的模型在计数性能和计数效率上得到了最佳的均衡。

3 个数据集的可视化结果如图 10—图 13 所示。从左到右依次为输入的测试集图片、地面真实密度图和估计密度图。

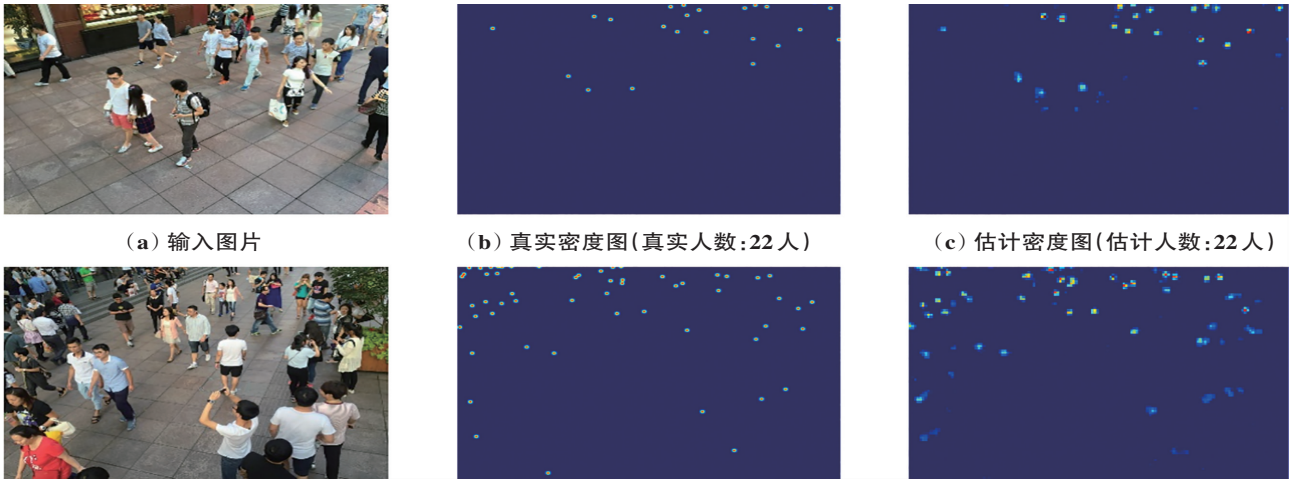


(a)输入图片 (b)真实密度图(真实人数:152 人) (c)估计密度图(估计人数:154 人)

(d)输入图片 (e)真实密度图(真实人数:1 174 人) (f)估计密度图(估计人数:1 171 人)

图 10 Shanghai TechA 数据集上部分实验结果

Fig. 10 Selected experimental results on Shanghai TechA dataset

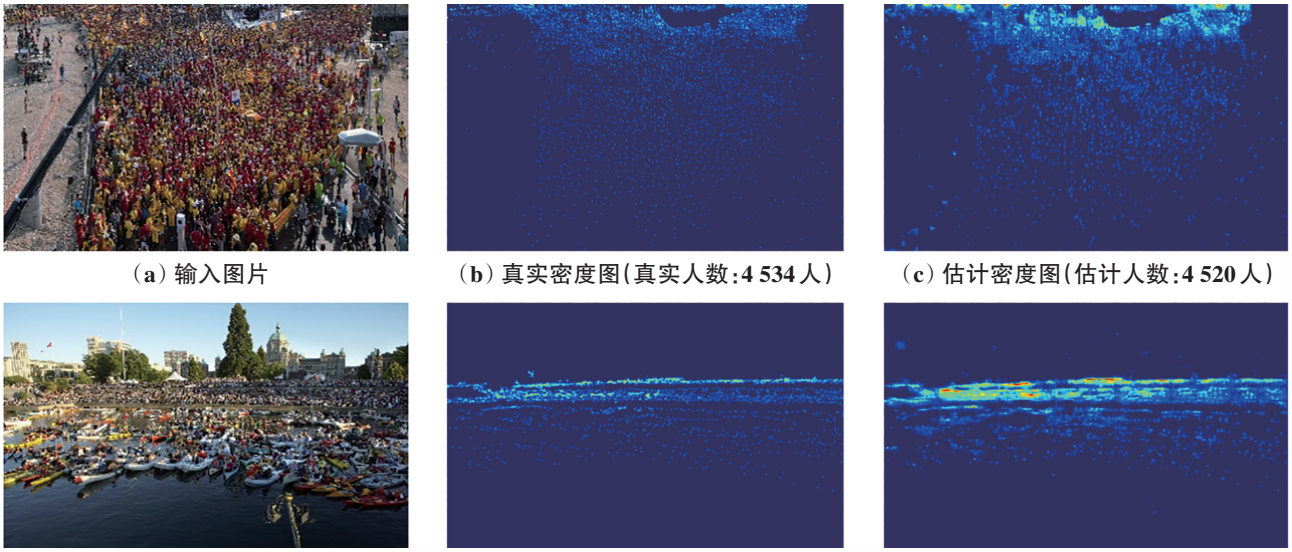


(a)输入图片 (b)真实密度图(真实人数:22 人) (c)估计密度图(估计人数:22 人)

(d)输入图片 (e)真实密度图(真实人数:55 人) (f)估计密度图(估计人数:55 人)

图 11 Shanghai TechB 数据集上部分实验结果

Fig. 11 Selected experimental results on Shanghai TechB dataset

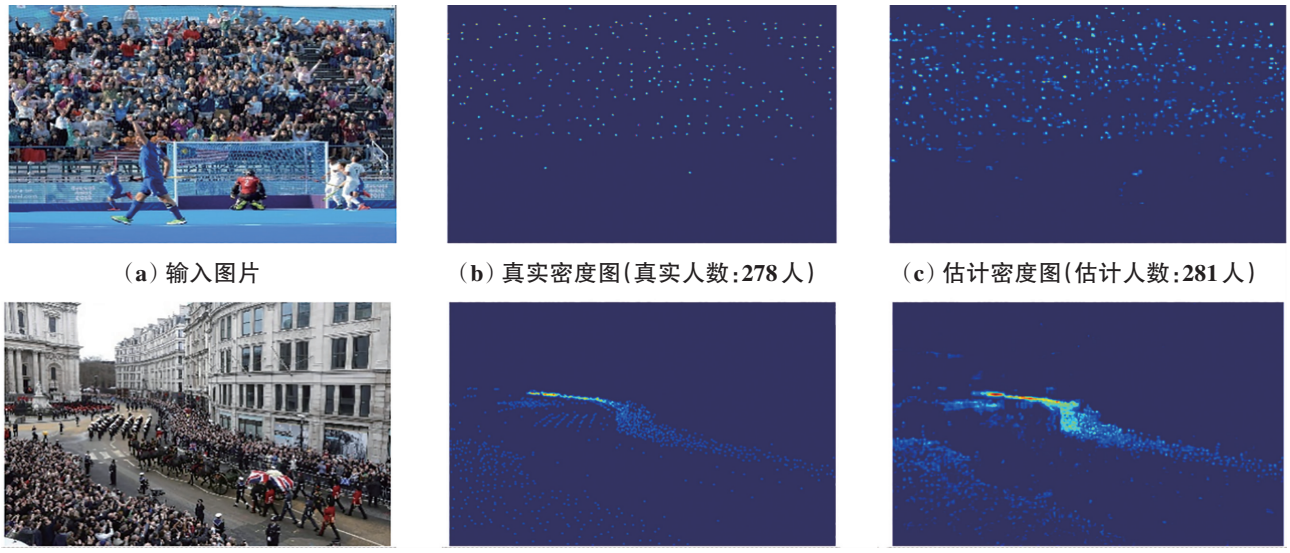


(a)输入图片 (b)真实密度图(真实人数:4 534 人) (c)估计密度图(估计人数:4 520 人)

(d)输入图片 (e)真实密度图(真实人数:2 978 人) (f)估计密度图(估计人数:2 981 人)

图 12 UCF-QNRF 数据集上部分实验结果

Fig. 12 Selected experimental results on UCF-QNRF dataset



(a) 输入图片　(b) 真实密度图(真实人数:278 人)　(c) 估计密度图(估计人数:281 人)

(d) 输入图片　(e) 真实密度图(真实人数:1 201 人)　(f) 估计密度图(估计人数:1 196 人)

图 13　NWPU-Crowd 数据集上部分实验结果

Fig. 13　Selected experimental results on NWPU-Crowd dataset

3　结　论

提出一种基于卷积神经网络的轻量化人群计数模型,在保证模型复杂度降低的情况下保留了模型优异的计数性能。针对目前人群计数大模型参数量和计算量过大而轻量化小模型计数性能又较差的问题,设计了一个由轻量化线性特征映射块构建的特征提取模块和基于扩张卷积网络组成的密度图回归模块。特征提取模块通过线性映射块利用简单的线性计算就可以提取更多的相似特征,更加注重特征的融合,在保证计算精度的同时大大降低了参数量和计算量。密度图回归模块利用扩张卷积以使用更小的卷积核来获得更大的感受野,在降低计算量的同时不减少有效的特征提取。最后在实验阶段,将本文提出的轻量化计数模型在 3 个主流数据集上进行了实验,并与多个经典人群计数模型进行对比。由实验指标可知:本文提出的轻量化计数模型与其他轻量级计数模型相比,参数量和计算量是最低的,取得的计数性能除 Part B 数据集外均是最优异的;与重型计数模型相比,虽然计数能力不是最优异的,但是最佳地平衡了计数性能和计数效率,更具备实用性。未来的工作将考虑加入 Vision Transformer 来加强全局特征的提取,将计数性能进一步优化。

参考文献(References):

[1] 班玉冰. 大型群体活动密集人群: 突发公共安全事件的诱因与治理[J]. 福州党校学报, 2019(5): 30-34.
BAN Yu-bing. The dense crowds in the large scale-group activities: The cause and the management of the public safety emergency[J]. Journal of the Party School of Fuzhou, 2019(5): 30-34.

[2] 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. Piscataway: IEEE Press, 2017: 1879-1888.

[3] OH M, OLEN P A, RAMAMURTHY K N. Crowd counting with decomposed uncertainty[C]// AAAI Conference on Artificial Intelligence(AAAI). New York, NY, USA: AAAI, 2020: 799-806.

[4] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

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

[6] LIU W, SALZMZNN M, FUA P. Context aware crowd counting[C]// Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 5099-5108.

[7] MENG Y, ZHANG H, ZHAO Y, et al. Spatial uncertainty-aware semi-supervised crowd counting[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2022: 529-539.

[8] WANG M, CAI H, HAN X F, et al. STNet: Scale tree network with multi-level auxiliator for crowd counting[J]. IEEE Transactions on Multimedia, 2023, 25: 2074-2084.

[9] 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. Piscataway: IEEE Press, 2016: 589-597.

[10] SAM D B, BABU R V. Top-down feedback for crowd counting convolutional neural network[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence and 30th Innovative Applications of Artificial Intelligence Conference and 8th AAAI Symposium on Educational Advances in Artificial Intelligence. New York: ACM, 2018: 7323-7330.

[11] CAO X, WANG Z, ZHAO Y, et al. Scale aggregation network for accurate and efficient crowd counting[C]//Computer Vision-ECCV 2018. Cham: Springer, 2018: 757-773.

[12] GAO J, WANG Q, LI X. PCC net: Perspective crowd counting via spatial convolutional network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(10): 3486-3498.

[13] MA X, DU S, LIU Y. A lightweight neural network for crowd analysis of images with congested scenes[C]//Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2019: 979-983.

[14] LIANG L, ZHAO H, ZHOU F, et al. PDDNet: Lightweight congested crowd counting via pyramid depth-wise dilated convolution[J]. Applied Intelligence, 2023, 53(9): 472-484.

[15] YI J, SHEN Z, CHEN F, et al. A lightweight multiscale feature fusion network for remote sensing object counting[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5902113.

[16] IOFFE S, SZEGED C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning (ICML). Lille, France: ACM, 2015: 448-457.

[17] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4510-4520.

[18] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[19] 张倩倩. 基于知识蒸馏的轻量级网络图像人群统计[D]. 武汉: 华中科技大学, 2020.
ZHANG Qian-qian. The image crowd counting of light-weight network based on knowledge distillation [D]. Wuhan: Huazhong University of Science and Technology, 2020.

[20] 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. Piscataway: IEEE Press, 2015: 833-841.

[21] LEMPITSKY V, ZISSERMAN A. Learning to count objects in images[C]//In Advances in Neural Information Processing Systems(NIPS). Vancouver, Canada: IEEE, 2010: 1324-1332.

[22] LIU Y B, JIA R S, XU Z F. Crowd counting algorithm based on scale space pyramid network[J]. Chinese Science and Technology Papers, 2021, 16(3): 276-280.

[23] 付宇豪. 基于卷积神经网络的人群计数算法研究与应用[D]. 北京: 北京工业大学, 2019.
FU Yu-hao. Research and application of crowd counting algorithm based on convolutional neural network[D]. Beijing: Beijing University of Technology, 2019.

[24] 陆金刚. 面向人群数量估计的多尺度卷积神经网络模型研究[D]. 苏州: 苏州大学, 2020.
LU Jin-gang. Multi-scale convolutional neural network models for crowd count estimation[D]. Suzhou: Soochow University, 2020.

[25] LIU C, WENG X, MU Y. Recurrent attentive zooming for joint crowd counting and precise localization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 1217-1226.

[26] IDREES H, TAYYAB M, ATHREY K, et al. Composition loss for counting, density map estimation and localization in dense crowds[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 544-559.

[27] 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, 2021, 43(6): 2141-2149.

[28] 李诚. 基于轻量级卷积神经网络的人群计数算法研究[D]. 哈尔滨: 哈尔滨工程大学, 2021.
LI Cheng. Research on crowd counting algorithm based on light-weight convolutional neural network[D]. Harbin: Harbin Engineering University, 2021.

[29] 陆金刚, 张莉. 基于多尺度多列卷积神经网络的密集人群计数模型[J]. 计算机应用, 2019, 39(12): 3445-3449.
LU Jin-gang, ZHANG Li. Crowd counting model based on multi-scale multi-column convolutional neural network[J]. Journal of Computer Applications, 2019, 39(12): 3445-3449.

[30] 贾云舒. 基于深度学习的人群计数方法研究[D]. 成都: 电子科技大学, 2022.
JIA Yun-shu. Crowd counting method based on deep learning[D]. Chengdu: University of Electronic Science and Technology of China, 2022.

[31] 夏殷锋. 基于卷积神经网络的人群密度估计方法研究[D]. 合肥: 中国科学技术大学, 2021.
XIA Yin-feng. Crowd density estimation method based on convolutional neural network[D]. Hefei: University of Science and Technology of China, 2021.

责任编辑:李翠薇