



# 基于尺度自适应卷积神经网络的人群计数算法

翟 强,王陆洋,殷保群,彭思凡,邢思思

(中国科学技术大学 自动化系,合肥 230027)

**摘 要:**为解决单幅图像中的人群遮挡和尺度变化问题,提出一种基于多列卷积神经网络的人群计数算法。利用具有不同尺寸感受野的卷积神经网络(CNN)和特征注意力模块自适应提取多尺度人群特征,引入可变形卷积增强 CNN 网络空间几何形变学习能力并优化特征图,从而生成高质量的密度图。Shanghai Tech 和 UCF\_CC\_50 数据集上的实验结果表明,该算法能学习输入图和人群密度图之间的映射关系,且计数准确性高、鲁棒性强。

**关键词:** 人群计数;卷积神经网络;可变形卷积;特征图;密度图

开放科学(资源服务)标志码(OSID):



**中文引用格式:**翟强,王陆洋,殷保群,等.基于尺度自适应卷积神经网络的人群计数算法[J].计算机工程,2020,46(2):250-254,261.

**英文引用格式:**ZHAI Qiang, WANG Luyang, YIN Baoqun, et al. Crowd counting algorithm based on scale adaptive convolutional neural network[J]. Computer Engineering, 2020, 46(2): 250-254, 261.

## Crowd Counting Algorithm Based on Scale Adaptive Convolutional Neural Network

ZHAI Qiang, WANG Luyang, YIN Baoqun, PENG Sifan, XING Sisi

(Department of Automation, University of Science and Technology of China, Hefei 230027, China)

**[Abstract]** In order to solve the problem of crowd occlusion and scale change in a single image, this paper proposes a crowd counting algorithm based on multi-column convolution neural network. The algorithm uses Convolutional Neural Network(CNN) with receptive fields of different sizes and the feature attention module to adaptively extract multi-scale crowd features. The deformable convolution is introduced to enhance the learning ability of spatial geometric deformation of the network and optimize the feature map, so as to generate a high quality density map. Experimental results on the Shanghai Tech and UCF\_CC\_50 datasets show that the algorithm can learn the mapping relationship between input images and crowd density maps, and has high counting accuracy and robustness.

**[Key words]** crowd counting; Convolutional Neural Network(CNN); deformable convolution; feature map; density map  
**DOI:**10.19678/j.issn.1000-3428.0053842

### 0 概述

近些年来,发生在世界各地的大型人群踩踏事故屡见不鲜。及时检测人群急剧变化,优化公共交通调度可以避免类似安全事故的发生,因此准确估计人群数量具有重要的研究价值,而透视效果导致的人群遮挡、尺度变化问题仍是人群计数面临的重大挑战。随着深度学习的发展,卷积神经网络(Convolutional Neural Network, CNN)模型<sup>[1-2]</sup>在图像处理上表现优异,使得研究人员将卷积神经网络应用于人群计数中<sup>[3]</sup>。文献[4]提出使用卷积神经网络

统计不同场景中的人群数目,先为某些场景预先训练一个网络,当给定来自新场景的测试图像时,根据透视信息和密度图中的相似性选择相似训练数据对预训练网络进行微调。该方法在大多数现有数据集上表现出良好的性能,但在训练和测试时对透视图的需求,限制了该方法在实际人群场景中的适用性和扩展性。文献[5]提出一个端到端的 CNN 架构,将整个图像作为输入并直接输出计数结果。但由于难以解决图像尺度多变的问题,基于单列 CNN 的方法的计算精度和鲁棒性效果均不佳。文献[6]提出 Switching CNN 网络,先将图像分块,通过 VGG16 网

**基金项目:**装备预研领域基金(61403120201)。

**作者简介:**翟 强(1992—),男,硕士研究生,主研方向为图像处理;王陆洋,博士研究生;殷保群,教授;彭思凡、邢思思,硕士研究生。

**收稿日期:**2019-01-28 **修回日期:**2019-03-05 **E-mail:**565617821@qq.com

络把图像块密度分成高中低 3 类,选择针对特定人群密度设计的 CNN 回归模型来估计图像块的人群密度图。然而人群在图像中的分布通常是不均匀的,对图像的简单分块影响了计数的准确性。文献[7]提出多列卷积网络,其能够适应部分人群的尺度变化,在一定程度上减少由于透视而导致的计数误差,但其使用一个  $1 \times 1$  卷积对多列密度图进行简单平均融合,会给最终结果带来一定的误差。

本文提出一种用于图像人群计数的新型卷积神经网络模型。考虑到经典 VGG16 网络<sup>[8]</sup>具有良好的特征提取能力,该网络模型前半部分采用 VGG16 的前 13 层提取图像特征,后半部分利用具有 3 列不同感受野的卷积神经网络进一步提取多样化的图像特征。在此基础上,构建特征注意力模块提取判别信息,对 3 列卷积神经网络得到的特征图进行评判打分,生成与特征图尺寸相同的注意力特征图,通过加权求和得到最终密度图,并在 Shanghai Tech 和 UCF\_CC\_50 两个数据集上训练并测试人群计数算法。

## 1 基于尺度自适应网络的人群计数算法

### 1.1 密度图

本文旨在准确估计输入图像的人群密度图,其积分值代表图中的人数。人群计数数据库只提供人头标记的坐标点,在网络训练前首先需要生成训练图片的密度图。对于图片中坐标为  $x_i$  的人头标记点,可将其表示为  $\delta(x - x_i)$ ,那么对于一张具有  $N$  个人头标记点的图片可以表示为:

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

然后结合高斯核滤波器  $G_\sigma$  与式(1)进行卷积操作得到密度方程,表示为:

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

对于人群均匀分布的图片,可以使用自适应高斯滤波器<sup>[4]</sup>进行卷积,密度图生成过程中需考虑图片视角扭曲因素。在此情况下的密度图方程可以表示为:

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

其中,  $\bar{d}_i$  是标记点  $x_i$  与其最近的  $k$  个人头之间的平均距离,  $\beta$  根据经验取值为 0.3。

### 1.2 网络结构

针对人群计数中的人群遮挡和尺度变化问题,本文提出一个用于人群计数的尺度自适应网络。在图 1 中,尺度自适应网络分为 3 个模块,具体为:第一模块引入经典 VGG16 网络的前 13 层对图片提取特征,由于 VGG 4 个池化层的存在,导致输出变为原图的  $1/16$ ,而本文采用  $1/8$  密度图作为输出结果,因此加入上采样层将特征图变为原图尺寸的  $1/8$ ;第二模块基于多列卷积神经网络思想,分别用  $3 \times 3, 5 \times 5, 7 \times 7$  不同感受野的卷积核对第一模块输出提取多尺度特征,在此阶段本文引入在目标检测、语义分割领域的可变形卷积,凭借其良好的空间几何形变学习能力,进一步改善特征图;第三模块为特征注意力模块,其对 3 列卷积神经网络得到的特征图经 Softmax 层进行评判打分及优化,从而生成高质量的密度图。

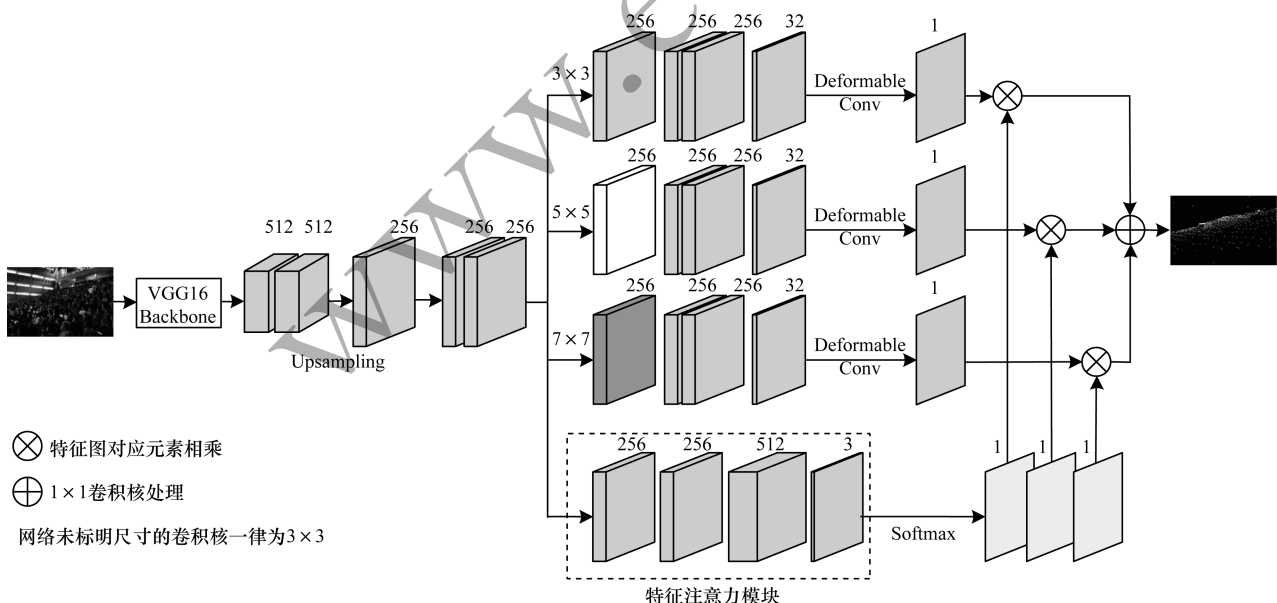


图 1 尺度自适应网络结构

Fig. 1 Scale adaptive network structure

### 1.2.1 可变形卷积

近年来卷积神经网络在视觉领域表现优异,其良好的特征提取能力和端对端的学习方式,避免了传统方法的弊端。但网络模型的几何形变能力来自于数据本身的多样性,即卷积操作。该操作在输入图像的每个位置都会进行基于对应位置的采样,然后对于采样到的特征区域做卷积并作为该位置的输出。以 $3 \times 3$ 卷积核为例,设 $R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$ , $R$ 代表卷积核覆盖特征图的感受野区域。对于特征图上的像素点 $p_0$ ,标准卷积方式表示为:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \quad (4)$$

但该方式会导致网络自身难以适应尺度多变的人群计数问题,因此本文引入可变形卷积<sup>[9]</sup>。可变形卷积是对卷积核的每个采样点的位置都增加一个偏移的变量,通过这些变量,卷积核可以在当前位置附近采样,而不再局限于之前对应的位置采样。可变形卷积方式表示为:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (5)$$

其中, $\Delta p_n$ 通过对网络训练得到。标准卷积加上该偏移量的学习后,可变形卷积核的大小和位置可以根据当前需要识别的图像内容进行动态调整,即不同位置的卷积核采样点位置会根据图像内容发生自适应变化,从而适应不同场景人群形状、大小、尺度变化等几何形变。分析上述计算方式可知,若可变形卷积的偏移量 $\Delta p_n$ 为0,则可变形卷积退化成普通卷积,本文算法也能达到普通卷积的算法性能。当可变形卷积起作用时,特别是卷积核在处理人群与树木、马路、房屋等干扰背景交界的区域时,普通卷积只能对卷积核相应尺寸大小的特征进行计算,这样无形中把非人群的无用特征计算进来,会给结果带来一定的误差。而可变形卷积可以有效地解决这个问题,当算法在训练过程中遇到上述场景时,通过训练卷积核附带的偏移量参数 $\Delta p_n$ ,将普通卷积核部分对应计算的非人群特征偏移到人群特征中,这样整个卷积核提取的都是人群特征,有效避免了非人群特征的干扰,从而提高人群计数性能。

### 1.2.2 特征注意力模块

人脑每个时刻接收的外界输入信息非常多,当人脑在接收这些信息时会自觉或不自觉地利用注意力机制获取重要的信息。目前,这种注意力机制被引入自然语言处理、物体检测、语义分割等领域,并具有很好的应用效果。本文将注意力机制作为一个模块加入卷积神经网络,旨在优化第二模块的3列特征图,因此本文将命名为特征注意力模块。为解决图片中人群尺度变化问题,该注意力模块将第一模块结果作为输入,对经过可变形卷积处理的3列多尺度特征图逐像素点生成具有判别能力的高级信息,即3列多尺度特征图对应的每个像素点都生成相应的权重,以此衡量3列多尺度特征图对应像素

点的重要程度。总体来看,注意力模块是对3列多尺度特征图进行打分,评判3列多尺度特征图所有对应像素点的重要程度。计算方式如下:假设 $F_k$ 为注意力模块输出的3列特征图中的某一列, $S_k$ 为经过可变形卷积处理的多尺度特征图中的某一列,则网络经Softmax层生成相应的注意力特征图(与特征图相同尺寸的概率矩阵)。

$$P_k = \frac{\exp(F_k)}{\sum_{i=1}^m \exp(F_k)}, m=3 \quad (6)$$

注意力特征图与特征图相对应的像素点相乘,再对其求和生成最终的密度图。

$$O = \sum_{k=1}^m P_k \cdot S_k, m=3 \quad (7)$$

通过该方式使网络在3列多尺度特征图上从像素点级别寻找与真实密度图相关的有用信息,对密度图生成有益的像素点获得更高的权重,对密度图生成无益的像素点获得更低的权重,即评判特征图的重要程度,从而提高最终密度图的生成质量。

### 1.3 网络结构优化

本文采用自适应学习率的Adam优化算法进行网络优化,使用欧氏距离作为目标函数,计算预测得到的密度图与真实密度图之间的差距。损失函数的定义如下:

$$L(\Theta) = \frac{1}{2N} \sum_{i=1}^N \|F(X_i; \Theta) - F_i\|^2 \quad (8)$$

其中, $\Theta$ 是网络参数, $N$ 是训练集图片总数, $X_i$ 是第 $i$ 张图片, $F(X_i; \Theta)$ 和 $F_i$ 分别表示第 $i$ 张图片的估计密度图和真实密度图。

## 2 实验结果与分析

本文在Shanghai Tech和UCF\_CC\_50数据集上训练并测试人群计数算法,借鉴经典网络VGG<sup>[8]</sup>、MCNN<sup>[7]</sup>、SAnet<sup>[10]</sup>等的参数设置方法,若没有特殊声明,卷积核尺寸默认为 $3 \times 3$ ,卷积核的激活函数均为ReLU非线性激活函数。网络对VGG提取的特征(原图片的1/16)进行上采样后,特征图尺寸变为原来的2倍(原图片的1/8),则卷积核通道变为原来的1/2,数目为256。本文基于1/8密度图估计人群数目,因此卷积层通道数均采用256通道。在设置可变形卷积通道超参数时,由于该卷积方式的复杂性,实验过程中发现对内存的消耗较大,训练时间较长,因此对每列特征图只设置32通道的可变形卷积。本文在Pytorch深度学习环境下进行训练,在网络训练权值初始化方面,VGG特征提取器使用已训练好的模型参数进行初始化,其他层使用标准差为0.01的高斯分布进行初始化,网络的初始训练学习率为 $1e-5$ ,共训练300个周期。

### 2.1 评价标准

本文使用平均绝对误差(Mean Absolute Error, MAE)和均方误差(Mean Squared Error, MSE)评价

算法性能。MAE 和 MSE 的定义如下:

$$M_{MAE} = \frac{1}{N} \sum_{i=1}^N |z_i - z'_i| \quad (9)$$

$$M_{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_i - z'_i)^2} \quad (10)$$

其中,  $z_i$  和  $z'_i$  分别表示第  $i$  张图片的真实人数和算法估计人数。

## 2.2 Shanghai Tech 数据集

Shanghai Tech 数据集共包含 1 198 张图片,标记人数达到 330 165 人。该数据集由 Part\_A 和 Part\_B 两部分组成。两个数据集无论是人群规模还是人群密度都有很大的区别。Part\_A 人群密度较大,Part\_B 人群密度中等但分布变化较大。

在数量上,Part\_A 由 482 张从网络抓取的图片组成,Part\_B 包含从上海街头拍摄的 716 张图片,同时它们都分为训练集和测试集两部分。Part\_A 中 300 张作为训练集,182 张作为测试集。Part\_B 中训练集和测试集图片分别为 400 张和 316 张。

由于 Part\_A 和 Part\_B 数据集的训练集分别是 300 张和 400 张图片,图片数量太少,因此采用数据增强方式,从每一张训练图片中截取 36 张图片,同时过滤单幅少于 5 人的图片,每一张图片尺寸都为原图片尺寸的 1/4,采用该方式增加训练图片的数量。在训练过程中,手动将初始学习率降低 10 倍再次训练,使本文算法达到更好效果。

表 1 为本文算法与 8 种当前主流人群计数算法的性能比较结果,由此可知:在 Part\_A 数据集上,本文算法 MAE 和 MSE 均为最优;在 Part\_B 数据集上,本文算法与其他算法性能接近。

表 1 Shanghai Tech 数据集上各算法性能对比结果  
Table 1 Performance comparison of different algorithms on the Shanghai Tech dataset

算法	Part_A		Part_B	
	MAE	MSE	MAE	MSE
ZHANG <sup>[4]</sup>	181.8	277.7	32.0	49.8
MCNN <sup>[7]</sup>	110.2	173.2	26.4	41.3
Switching CNN <sup>[6]</sup>	90.4	135.0	21.6	33.4
CP-CNN <sup>[11]</sup>	73.6	106.4	20.1	30.1
ACSCP <sup>[12]</sup>	75.7	102.7	17.2	27.4
CSRNet <sup>[13]</sup>	68.2	115.0	10.6	16.0
IG-CNN <sup>[14]</sup>	72.5	118.2	13.6	21.1
SANet <sup>[10]</sup>	67.0	104.5	8.4	13.6
本文算法	66.8	100.0	11.6	18.4

本文算法不仅在人群密集场景的 Part\_A 数据集上有效,对于人群密度中等且分布变化较大的 Part\_B 数据集依然表现良好,说明本文算法的适用性强和准确性高。图 2 分别展示了两个数据集上测试图片的估计密度图和真实密度图。

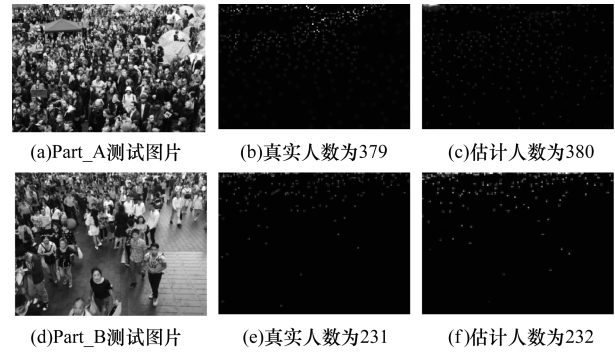


图 2 测试图片的估计密度图与真实密度图 1

Fig. 2 Estimated and true density images of test image 1

## 2.3 UCF\_CC\_50 数据集

UCF\_CC\_50 数据集<sup>[15]</sup>是一个复杂的数据集,该数据集人群密度大,仅有 50 张图片,但却有 63 974 人。同时,该数据集图片的人群数目从 94 人到 4 543 人不等,平均每张图片有 1 280 人。本文采用与当前主流人群计数算法相同的方式,对原数据集进行五折交叉验证,并采取与 Shanghai Tech 数据集相同的方式进行数据增强,最后用本文算法分别在 5 个数据集上进行训练和测试。表 2 为本文算法与该数据集上表现良好的 9 种当前主流人群计数算法的性能比较结果,数据显示本文算法的 MAE 和 MSE 两项指标均优于其他算法,说明本文算法具有良好的泛化能力和较高的准确性。

表 2 UCF\_CC\_50 数据集上各算法性能对比结果  
Table 2 Performance comparison of different algorithms on the UCF\_CC\_50 dataset

算法	MAE	MSE
MCNN <sup>[7]</sup>	377.6	509.1
Switching CNN <sup>[6]</sup>	318.1	439.2
TDF-CNN <sup>[16]</sup>	354.7	491.4
CP-CNN <sup>[11]</sup>	298.8	320.9
ACSCP <sup>[12]</sup>	291.0	404.6
IG-CNN <sup>[14]</sup>	291.4	349.4
CSRNet <sup>[13]</sup>	266.1	397.5
ic-CNN <sup>[17]</sup>	260.9	365.5
SANet <sup>[10]</sup>	258.4	334.9
本文算法	234.1	317.7

图 3 展示了两张测试图片的估计密度图和真实密度图。综合来看,本文算法在 Shanghai Tech 和 UCF\_CC\_50 数据集上的实验结果均已超越目前大部分计数算法,具有较高的准确性和较好的鲁棒性。数据和特征决定了机器学习的上限,而模型和算法只是逼近该上限且在深度学习领域同样适用,算法对图片特征的提取能力决定算法性能的优劣。鉴于以上思想,本文算法设计以提取和优化图片特征为目的,基于 VGG 前 13 层及使用多列不同尺寸的卷积核提取特征,并从以下方面优化特征:1)引入可变形卷积改进 3 列多尺度特征;2)构建注意力

模块对3列多尺度特征进行打分评判,使对密度图生成有益的特征获得更高权重,对密度图生成无益的特征获得更低权重,从而优化所得特征。通过图2、图3估计密度图与真实密度图的比较,算法生成的密度图人群分布与真实密度图接近,由此可知算法输出了较高质量的密度图,从而提升了计数准确度。而目前典型网络大多只从多列网络、更深层次网络进行设计,并未考虑影响模型性能的特征优化这一因素,而上述两方面保障了本文网络结构具有良好的特征提取能力,因此整体性能优于现有大部分经典网络。

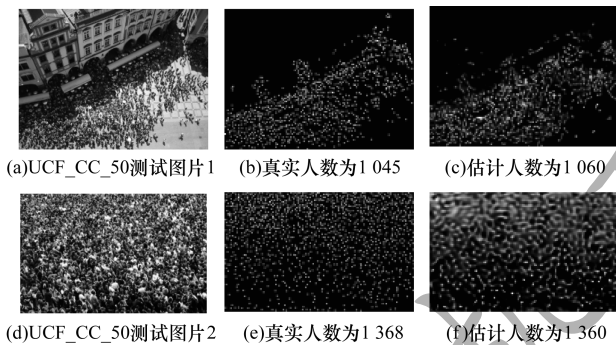


图3 测试图片的估计密度图与真实密度图2

Fig.3 Estimated and true density images of test image 2

为验证本文算法引入可变形卷积和特征注意力模块的有效性,并测试其对本文算法的性能影响,在Shanghai Tech数据集 Part\_A 上进行实验,结果如表3所示。在验证过程中,基于无可变形卷积网络和无注意力模块网络的算法与本文算法的训练方法一致,手动将初始学习率降低10倍再次训练,可使算法达到更好效果。

表3 Shanghai Tech数据集 Part\_A 验证模块对算法性能的影响

Table 3 Impact of the Part\_A verification module in the Shanghai Tech dataset on algorithm performance

算法	MAE	MSE
基于无可变形卷积网络的算法	69.6	110.0
基于无注意力模块网络的算法	68.0	101.6
本文算法	66.8	100.0

### 3 结束语

本文提出一种尺度自适应的基于多列卷积神经网络的人群计数算法。为解决图片中的人群遮挡和尺度变化问题,引入具有良好空间几何形变学习能力的可变形卷积,以提取多尺度特征。在此基础上,构建特征注意力模块自适应地融合多尺度特征并生成密度图。在Shanghai Tech和UCF\_CC\_50数据集上的实验数据显示,本文算法具有较强的鲁棒性和较高的准确性。后续将通过提高密度图质量和改进损失函数等技术进一步优化人群计数算法。

### 参考文献

- [1] CAO Lu, CHEN Ming, QIN Yufang. Pedestrian detection algorithm based on multi-channel and sparse convolution neural network [J]. Transducer and Microsystem Technologies, 2018, 37(10): 130-132, 136. (in Chinese) 曹璐, 陈明, 秦玉芳. 基于多通道稀疏卷积神经网络的行人检测算法[J]. 传感器与微系统, 2018, 37(10): 130-132, 136.
- [2] Pucuo Cairen, HONG Jianchao. Research on remote sensing image detection based on deep convolution neural network and significant image [J]. Automation and Instrumentation, 2018(12): 50-53, 57. (in Chinese) 普措才仁, 洪建超. 基于深度卷积神经网络与显著性图像的遥感图像检测研究[J]. 自动化与仪器仪表, 2018(12): 50-53, 57.
- [3] TANG Siqi, TAO Wei, ZHANG Liangliang, et al. Research and implementation of crowd counting based on convolutional neural network [J]. Journal of Zhengzhou University (Natural of Science Edition), 2018, 50(2): 1-6. (in Chinese) 唐斯琪, 陶蔚, 张梁梁, 等. 一种多列特征图融合的深度人群计数算法[J]. 郑州大学学报(理学版), 2018, 50(2): 1-6.
- [4] ZHANG Cong, LI Hongsheng, WANG Xiaogang, et al. Cross-scene crowd counting via deep convolutional neural networks [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 833-841.
- [5] SHANG Chong, AI Haizhou, BAI Bo. End-to-end crowd counting via joint learning local and global count [C]//Proceedings of IEEE Conference on Image Processing. Washington D. C., USA: IEEE Press, 2016: 1215-1219.
- [6] SAM D B, SURYA S, BABU R V. Switching convolutional neural network for crowd counting [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1-5.
- [7] ZHANG Y, ZHOU D, CHEN S, et al. Single-image crowd counting via multi-column convolutional neural network [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 589-597.
- [8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-01-15]. <https://www.cnblogs.com/yinheyi/p/6233950.html>
- [9] DAI Jifeng, QI Haozhi, XIONG Yuwen, et al. Deformable convolutional networks [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 3.
- [10] CAO Xinkun, WANG Zhipeng, ZHAO Yanyun, et al. Scale aggregation network for accurate and efficient crowd counting [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 757-773.
- [11] SINDAGI V A, PATEL V M. Generating high-quality crowd density maps using contextual pyramid CNNs [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 1861-1870.

(下转第261页)

(上接第 254 页)

- [12] SHEN Z, XU Y, NI B, et al. Crowd counting via adversarial cross-scale consistency pursuit [ C ]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2018; 5245-5254.
- [13] LI Yuhong, ZHANG Xiaofan, CHEN Deming. CSRNet: dilated convolutional neural networks for understanding the highly congested scenes [ C ]// Proceedings of Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2018; 1-6.
- [14] BABU S D, SAJJAN N N, VENKATESH B R, et al. Divide and grow: capturing huge diversity in crowd images with incrementally growing CNN [ C ]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2018; 3618-3626.
- [15] IDREES H, SALEEMI I, SEIBERT C, et al. Multi-source multi-scale counting in extremely dense crowd images [ C ]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA; IEEE Press, 2013; 2547-2554.
- [16] 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. Palo Alto, USA; AAAI Press, 2018; 1-8.
- [17] RANJAN V, LE H, HOAI M. Iterative crowd counting [ C ]// Proceedings of European Conference on Computer Vision. Berlin, Germany; Springer, 2018; 278-293.