

可伸缩模块化 CNN 人群计数方法

李云波¹ 唐斯琪¹ 周星宇² 潘志松¹

(中国人民解放军理工大学指挥信息系统学院 南京 210000)¹

(中国人民解放军理工大学通信工程学院 南京 210000)²

摘要 本文目标是根据任意视角、任意人群密度的图像信息,估计真实场景中的人群密度。但三维空间景物投影到二维空间时会造成透视失真和人群遮挡问题,导致难以区分个体与个体、个体与背景的差异。为此,提出一种灵活高效的伸缩模块化卷积神经网络(CNN)的架构,允许直接输入任意大小和分辨率的图像,不额外计算视角变化信息,通过生成密度图的方式来估计人群数量。架构的每个模块采用不同卷积核的多列结构,可以拟合不同远近的个体信息;并结合前后两层的特征信息,减少了梯度消失造成的精度下降损失。实验证明,在 ShanghaiTech PartA 和 PartB 数据集上,所提方法的准确率比之前最好的 MCNN 方法分别提高了 14.58%,40.53%,均方根误差分别降低了 23.89%,33.90%。

关键词 人群计数,卷积神经网络,可伸缩模块,密度图,特征融合

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.08.004

Crowd Counting Method via Scalable Modularized Convolutional Neural Network

LI Yun-bo¹ TANG Si-qi¹ ZHOU Xing-yu² PAN Zhi-song¹

(Institute of Command Information System, PLA University of Science and Technology, Nanjing 210000, China)¹

(College of Communication Engineering, PLA University of Science and Technology, Nanjing 210000, China)²

Abstract The purpose of this paper is to accurately estimate the crowd density in real scenes based on image information from arbitrary perspective and arbitrary crowd density. However, crowd counting on static images is a challenging problem. Due to the perspective distortion and the crowd crushes caused by the projection from 3D space into 2D space, it is difficult to distinguish the difference between individual and individual and the difference between individual and background. To this end, this paper proposed a flexible and efficient scalable modularized convolutional neural network (CNN) architecture. The network allows to directly input images with arbitrary size and resolution and it does not require additional computational changes in view information. Each module of the architecture employs a multiple column structure with different convolution kernels, which can be used to fit individual information of different distances. The proposed module also combines the feature information of the front and rear two layers, reducing the decrease loss of the accuracy caused by the vanishing of the gradient. Experiments show that the accuracy of proposed method is increased by 14.58% and 40.53%, and the root mean square error is reduced by 23.89% and 33.90% respectively on ShanghaiTech PartA and PartB datasets compared with the state-of-the-art MCNN methods.

Keywords Crowd counting, Convolutional neural network, Scalable module, Density maps, Feature fusion

1 引言

密集人群在狭小空间内的拥挤移动,容易导致交通阻塞,对公共安全产生威胁。传统的人工监控效率低下,不仅浪费人力资源,而且难以对拥挤情况及时做出反应。随

着安全监控行业的发展,各种监控设备几乎覆盖了所有公共场所。计算机视觉和图像技术的发展和广泛应用,使得可以对人群密度实现智能估计,及时发现人群活动的拥塞情况。人群计数,即通过图像等实时信息准确估计实际场景的人群密度。通过量化人群数量,实时监控并提前预

到稿日期:2017-10-24 返修日期:2017-12-16 本文受属性学习及其应用研究(61473149),2017 年国家重点研发计划“网络空间安全”重点专项(2017YFB0802800)资助。

李云波(1994—),男,硕士,主要研究方向为机器学习、深度学习,E-mail:18252059269@163.com;唐斯琪(1993—),女,硕士,主要研究方向为深度学习、图像视觉;周星宇(1985—),男,硕士,讲师,主要研究方向为机器学习、模式识别;潘志松(1973—),男,博士,教授,主要研究方向为模式识别与人工智能,E-mail:panzs@nuaa.edu.cn(通信作者)。

警,实现流量控制和公共安全决策。

此外,人群计数在相关领域的推广也取得了有效的成果,比如在显微镜下的微生物计数、密集小生物如蜜蜂和蚂蚁的种群活动及野外环境下的物种迁移等方向均有相似的应用。

1.1 研究挑战

相机采集的图片存在透视效果,会造成透视失真。在将3D镜像映射到2D图片的过程中,人群中各个个体的大小都发生了变化,距摄像头近的人所占的像素比重大,而距离远的人所占的像素比重小;同时,由于拥塞和遮挡,近处的个体会显示全身,远处的个体往往只有头部被捕捉到。以上因素给我们的特征提取增加了难度。在低像素条件下,个体之间的差异性降低,有用的特征会减少,与背景、环境噪音的区别也会减小,难以进行背景分割。

1.2 相关工作

主流的人群计数方法有两种:检测框法和密度图回归法。

检测框法^[1-3]采用检测框准确定位个体位置,基于外观和行为特征,将每一个个体与其他个体分隔开,然后累计全部个体数目,该方法在低密度人群条件下取得了很好的效果。但是,受限于密集环境下拥塞人群的相互遮挡,随着分辨率的降低、人群数量的剧增、视角的变化,准确率逐渐下降,其鲁棒性较差^[4-6]。

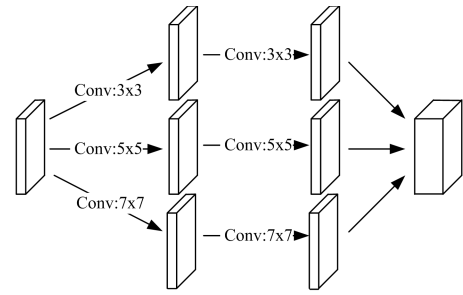
Lempitsky等^[7]提出了通过整体估计密度图的方式实现人群计数。通过生成密度图,可以有效地减少图像带来的噪声。Zhang等^[8]提出了一种不同场景下基于CNN的计数方法,他们首先在该场景下预训练一部分图片,得到该场景网络模型。当在一个新场景测试时,只需要根据新场景的透视图微调已经训练好的网络模型即可在测试集上表现出不错的性能;但该方法需要进行预训练,并且需要新场景下的透视图,这极大地限制了其实际应用。

2 模型构建

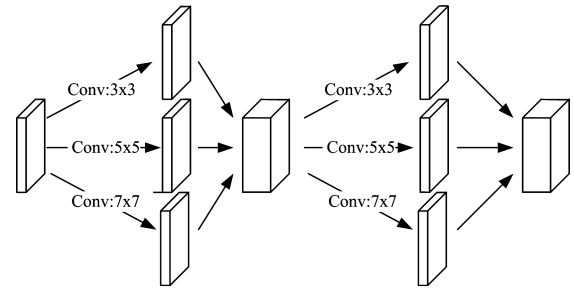
2.1 模型思路

Wang等^[9]采用了单列结构的深度网络来进行学习,但很难解决透视失真的问题。为了解决实际图片中个体头部大小不等的情况,借鉴Boominathan^[10]的工作,选取不同尺度的卷积核(大小分别为 $7 \times 7, 5 \times 5, 3 \times 3$)来拟合不同远近个体的头部大小。通过采用大、中、小3列不同卷积核的方式达到一种自适应的效果,可以解决不同角度和分辨率的问题。Zhang等^[11]提出的多列结构的深度算法也采用了相似的解决方案。

这些改进算法可以解决规模变化的问题,但是多列网络需要训练基于全局优化的单列网络,在端到端的过程中实用性不强;同时,直连网络不能获取组合特征,只能获取单一卷积核大小的特征图,不能够将不同尺度的特征图结合起来。如图1所示,图1(a)采用多列结构,仅在网络的最后设置merge层,因而在隐层阶段没有特征融合;图1(b)采用层融合结构,使得深层网络可以获得更多的信息。考虑到隐层是特征的特征,将深层特征结合起来也是必要的。



(a) 多列结构



(b) 层融合结构

图1 网络结构对比

Fig. 1 Comparison of network structure

为了更好地解决实际问题,输入任意大小的图片,以减少在图片处理方面造成的特征损失。但因为层融合的结构,我们在每层的结束位置将特征图融合起来,以保证每层结束时特征图的大小相同。

很多方法都是对图片进行预处理,即通过剪裁或者添补的方式使得图片的大小相同,这实际上是人工添加了噪音,或者浪费了图片资源,因此实际应用的效果并不理想。

我们采用添加pad的方式,并在非池化层设置stride为1,不改变图片的大小,这样可以很好地解决因为卷积核的大小不同造成特征图不一致的问题。

借鉴vgg16中通过增加特征图的方式来获取更多特征的方法,但是为了减小网络的规模,去除了最后的全连接层,根据Han等^[12-14]的剪枝操作,通过修剪网络中的全连接层并进行再训练,可以大规模降低计算成本而不损失训练精度。例如,vgg16全连接层占总参数的90%,但对准确率的提升不足1%。

网络采用了两次最大池化层来减小特征图,最后的特征图大小为原图片的1/4。为了更好地解决回归密度图的问题,借鉴Lin^[15]的方法,使用全卷积层作为最后一层,采用多个 1×1 的卷积核回归密度图。使用Hinton等^[16]介绍的ReLU函数作为激活函数,可以减少因为网络过深而造成的梯度消失的损失。

2.2 可伸缩模块

由VGG经典网络可知,在网络不太深的情况下,深层网络的效果比浅层网络好得多。考虑到梯度消失的情况,借鉴了He等^[17]残差网络残差映射的思想。残差网络ResNet采用将浅层特征图残差计算后叠加到深层特征图的方式,实现了浅层和深层的特征图融合。

因为较深的神经网络往往会导致梯度消失,即根据链式

法则,如果每一层神经元对上一层的输出的偏导乘上权重的结果都小于 1,在经过足够多层传播之后,即使这个结果是 0.99,那么 n 个 0.99 相乘,误差对输入层的偏导也会趋于 0。残差网络是解决梯度消失的一种方案。但由于残差网络比较深,连接数众多,收敛速度较慢,而且每条路径是等可能的,而我们的目标是尽量使较深的网络起到较好的作用,尽量使网络通过深层反向传播,因此我们构建了一种可伸缩比例模块。

我们借鉴了 VGG 增加特征图以获取更多特征的方式来获取更多特征图。函数公式为:

$$F(X) = \{a \times f_1(x), b \times f_2(x)\} \quad (1)$$

函数通过调节 a 与 b 的比例来实现深、浅层特征图的伸缩。当 $a:b < 1$ 时,浅层特征图的数量少于深层特征图的数量,使得 Module 中的特征主要依靠深层网络来提取。

Module 模块由两层组成,如图 2 所示。Scalable Module 每层采用层融合模型,在两个 Merge 层之间有一条直连通道,将浅层与深层的特征融合在一起。两层均由 3 个卷积层和 1 个 Merge 层组成,在第一层和第二层的 Merge 层之间有一条直连通道,通过调节 a 与 b 的比值来调节浅层特征和深层特征图的数目,当 $a:b=1:2$ 时可以较好地提取深层特征。因此,当第一层的特征图数目为 12:10:8 时,第二层的特征图数目为 24:20:16。同时,使用 Relu 函数作为激活函数,可以减轻深层网络造成的梯度消失问题。

Module 每层都采用层融合结构,采用不同尺度的卷积核

进行拟合,卷积核的大小分别为 $3 \times 3, 5 \times 5, 7 \times 7$ 。针对人群中低像素数量较多的问题,采用不同尺度的卷积核,并使输出对应的特征图的数量不同,使得卷积核小的卷积网络产生的特征图相应较多,而卷积核大的网络产生的特征图相应较少。

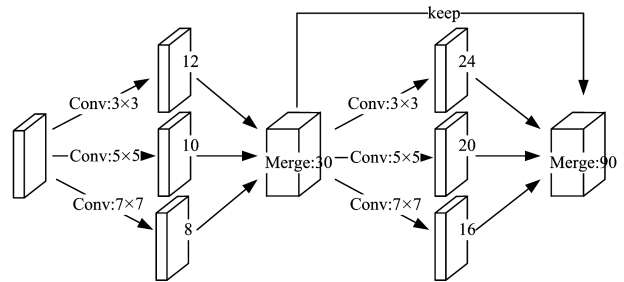


图 2 可伸缩模块

Fig. 2 Scalable module

2.3 网络的构建

如图 3 所示,我们分别构建了由 3 个可伸缩模块和 5 个可伸缩模块组成的卷积神经网络。上面是 3 个 Scalable Module 组合的 CNN 训练网络,网络中使用了两个 Pooling 层对特征图进行降维。左边是输入图片,右边是模型产生的密度图。可以清晰地看到,左图中密集的人群可以准确地回归到密度图中。下面是 5 个 Scalable Module 组合的 CNN 训练网络。其中使用了两个最大池化层进行特征图降维,特征图的最终大小为输入图片的 $1/4$ 。

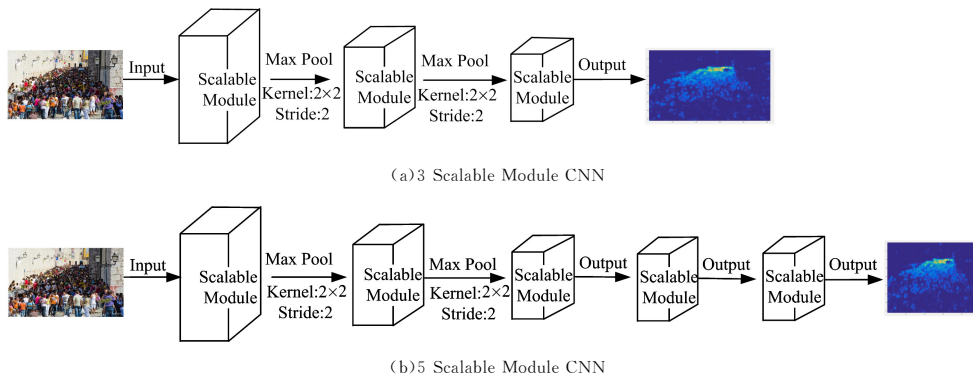


图 3 不同可伸缩模块组成的卷积神经网络

Fig. 3 Convolution neural network composed of different scalable modules

在最后的 Output 层使用了全卷积层,采用多个 1×1 的卷积核回归密度图,不会改变图片的大小,能够很好地将多张特征图回归为一张密度图,效果如图 4 所示。其中,第一列是原始图片,第二列是由真实标签产生的生成标注,第三列是 3 Scalable Module CNN 回归的密度图,第四列是 5 Scalable Module CNN 生成的密度图。深色代表背景,人群密度低;浅色代表人群密度适中,左上角人群密度高。从图 4 中也可以看出,3 Scalable Module CNN 中的噪音点多一些,因此效果略差。

我们使用两种不同深度的网络进行对比实验,以检测网络效果。

总的人群计数是由预测的密度图求和得到,但在训练过程中,损失函数为预测的密度图和真实标签生成标注的差值,采用 l_2 损失函数。

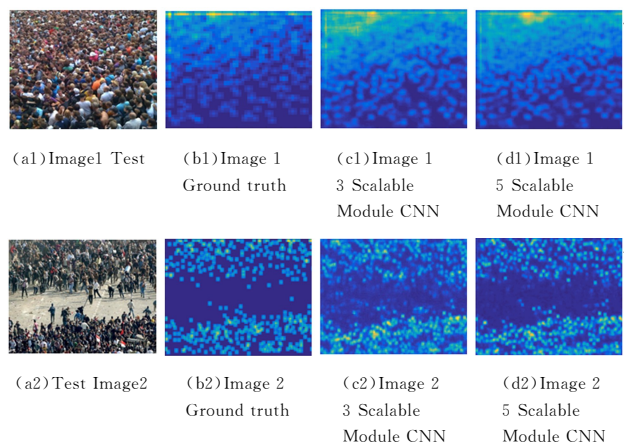


图 4 实验效果

Fig. 4 Experimental effects

3 实验

实验在两个公开数据集上进行,分别是适用性最广的 ShangHai Tech 数据集和具有挑战性的 UCF 数据集。GPU 芯片为泰坦 X,使用随机梯度下降(SGD)算法,超参数设置:学习率(lr)采用 $1E-7$,动量(moment)采用 0.9,CNN 网络基于 caffe 的框架进行训练。

3.1 评测标准

我们采用 MAE, RMSE 作为检测的指标。

$$MAE = \frac{1}{N} \sum_1^N |y_i - \tilde{y}_i| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_1^N (y_i - \tilde{y}_i)^2} \quad (3)$$

MAE 用于计算平均绝对误差,表示预测结果和实际结果之间的平均差值。RMSE 用于计算均方根误差,表示差值之间的离散程度。其中, N 表示测试数据集中图片的数量, y_i 和 \tilde{y}_i 分别代表图片中实际的人数和模型计算的人数。

3.2 生成标注

数据集提供了每个人群个体的二值标注,在每个人的头部中心有一个中心点,但直接使用其作为标签训练非常困难,预测效果很差。因此,采用高斯核回归的方式来产生生成标注。通过高斯核进行拟合,将每个标记点的像素作为中心,生成高斯核,使得这种模糊生成的生成标注总数与实际总数一致,并能很好地发现人群密集的区域。

我们采用 k 近邻的方式确定生成的个体头部大小,尽量与原个体头部大小保持一致。假设点 X_i 周围小范围内的人群均匀地分布在地面上,那么 X_i 和它周围 8 个点的平均距离 d_i 可以利用透视图表示(见式(4)),其中 N 为人头总数, a 的经验性取值为 0.3。

$$F(x) = \sum_{i=1}^N t(x-x_i) \times G_{a_i}, a_i = u \times \bar{d}_i \quad (4)$$

从图 4(b1) 可以看到,每个标记产生了 9 个像素,即 X_i 与周围的 8 个点。本文模型产生的密度图 4(c1) 与真实标签产生的生成标注基本一致。

3.3 ShanghaiTech 数据集

ShanghaiTech 数据集是一个标准的大型数据集,目前该数据集中的标注人群总数最多。该数据集分为两个部分,即 PartA 和 PartB。

3.3.1 数据集处理

PartA 是从互联网上选取的图片,共包含 482 个注释的图像,每张图片的分辨率不同。表 1 列出了 3 个公开数据集的相关信息。

表 1 3 个公开数据集的数据信息

Dataset	ShanghaiTech		UCF
	PartA	PartB	
Number	482	716	50
Max	3139	578	4543
Min	33	9	94
Average	501.4	123.6	1297.5
Total	241677	88488	63974
Resolution	different	768 * 1024	different

表 1 中,Number 是该数据集的图片总数,Max 是单张图片最多人数,Min 是单张图片最少人数,Average 是数据集的平均单张图片人数,Total 是该数据集的全部图片人群总数,Resolution 是该数据集中图片的分辨率。

从表 1 可以看到,平均每张图片的人群总数为 501.4,最少为 33 人,最多为 3139 人,总的人群数量达到 241677,人群密度较高。数据集已经由发布者分割为训练数据集和测试数据集,分别是 300 张训练图片和 182 张验证图片。

为了获取更多的数据而不造成过拟合,我们采用了经典的对称增强的方法,只进行左右对称,图片数量加倍。然后进行模块切割,将 1 张图片切割成 9 张,每张图片的大小为原来的 $1/4$,间隔为图片的 $1/6$,因此原来的 1 张图片增强为 18 张。从而,PartA 的训练集有 5400 张图片,测试集有 3276 张图片;PartB 的训练集有 7200 张图片,测试集有 5688 张图片。

测试集的增强仅用于模型训练,实际效果的检测过程采用原图片,不做改变。

3.3.2 实验结果

1) PartA 分析

将已有方法与可伸缩模块化计数方法进行比较,结果如表 2 所列。

表 2 不同方法在 PartA 数据集上的性能比较

Table 2 Performance comparison of different methods on PartA

Method	ShanghaiTech PartA	
	MAE	RMSE
LBP+RR	303.2	371.0
Zhang 等 ^[8]	181.8	277.7
MCNN ^[11]	110.2	173.2
3 Scalabe Module	96.0	139.8
5 Scalabe Module	93.2	134.3

LBP+RR 方法同样基于回归来实现人群计数,从原始图片中提取局部二值模式(LBP)特征,然后使用岭回归(RR)来预测每张图片的人群数量;但由于实验必须将图片切割成相同大小的小块,因此造成了信息的损失。

Zhang 等^[8]的方法是基于 CNN 在某场景下预训练一部分图片来得到该场景的网络模型,然后根据新场景的透视图微调已经训练好的网络模型,在当时表现出了较好的性能;但预训练和透视图极大地限制了该方法的应用。

MCNN 是已有方法中表现最好的结构,基于如图 1(a)所示的多列结构,采用单列结构端到端学习的方式进行预训练,然后将最终特征融合,并进行微调。该方法实现了列与列之间特征的融合,但没有考虑前后层之间以及每层内部不同卷积核的特征融合。

从表 2 可以看出,本文的 Scalabe Module CNN 模型取得了意料之外的效果,其中 5 Scalabe Module 的准确率提高了 15.43%,均方根误差降低了 23.89%,从而证明了所提方法是最有效的,其可以很好地融合不同尺度卷积核和不同层次的特征。

3 Scalabe Module 实现了每层网络不同尺度卷积核的特征融合,可以更好地拟合人群中不同个体的头部大小,进而通过不同的卷积核尺度达到一种自适应的效果,从而解决不同

角度和分辨率的问题,并减少透视失真带来的损失。

5 Scalabe Module 进一步降低了 MAE 和 RMSE,取得了更优异的效果,表明 Scalabe Module 的多模块融合确定实现了深浅层的特征交互,进一步降低了梯度消失造成的精度下降损失。

2) PartB 实验分析

表 3 列出了不同方法在 PartB 上的效果。可以看出,3 Scalabe Module 和 5 Scalabe Module 的性能都有压倒性优势,说明深层的网络具有较好的效果,其中 5 Scalabe Module 可以较好地减少梯度消失带来的问题,准确率提高了 40.53%,均方根误差降低了 33.90%。

表 3 不同方法在 PartB 数据集上的性能比较

Table 3 Performance comparison of different methods on PartB

Method	ShanghaiTech PartB	
	MAE	RMSE
LBP+RR	59.1	81.7
Zhang 等 ^[8]	32.0	49.8
MCNN ^[11]	26.4	41.3
3 Scalabe Module	16.4	28.0
5 Scalabe Module	15.7	27.3

通过对比 PartA 和 PartB 两个数据集,可以发现 PartA 的难度更大,因为 PartA 中的很多场景都是不相关的,因此不论训练集还是测试集的误差都很高,但这并不影响我们的评估标准。

3.4 UCF 数据集

UCF 数据集包含 50 张图片,是目前最具挑战性的数据集。

3.4.1 数据集处理

UCF 数据集图片取自互联网,挑战性不仅在于图片数量有限,还在于每张图片的人群数量急剧变化也会带来很大困扰。如表 1 所列,单张图片从 4543 人变化到 94 人,数据难度极高。

我们采取了 Zhang 等^[11]的 5 折交叉验证将数据集分成 5 份,分别作为测试集来检测模型的效果,并计算 MAE 和 RMSE 的平均值。

鉴于图片数量有限,将图片切割为原大小的 1/4,切割间距为 1/12,因此图片被增强为 36 张。

测试集的增强仅用于模型训练,实际效果检测过程仍采用原图片,不做改变。

3.4.2 实验结果

在 UCF 数据集上对 8 种方法进行比较。其中,Rodriguez 等^[18]采用密度图的方式对场景中的人群实现检测;Lempitsky 等^[7]采用随机选取块上的手工 SIFT 特征和距离度量来实现密度图回归;Idrees 等^[19]采用多元特征的方式回归密度图;其他方法同上文。实验结果如表 4 所列。

对比之前同类算法的实验效果,5 Scalabe Module 方法的准确率提高了 18.19%,均方根误差降低了 22.45%,取得了最好的效果,从而证明 Scalabe Module 可以自适应地拟合不同大小的个体,并能很好地提取深层特征而不面临梯度消失的问题。

表 4 不同方法在 UCF 数据集上的性能比较

Table 4 Performance comparison of different methods on UCF dataset

Method	UCF	
	MAE	RMSE
Rodriguez 等 ^[18]	655.7	697.8
Lempitsky 等 ^[7]	493.4	487.1
Idrees 等 ^[19]	419.5	541.6
Zhang 等 ^[8]	467.0	498.5
CCNN ^[20]	488.7	646.7
MCNN ^[11]	377.6	509.1
3 Scalabe Module	318.3	401.0
5 Scalabe Module	308.9	394.8

结束语 本文提出了一种可伸缩模块化 CNN 人群计数方法,该方法可以根据任意视角、任意人群密度、任意分辨率的图像信息,准确估计真实场景中的人群密度。

根据实验和理论分析,所提模型有两个主要特点:1)算法的适用性强,不需要传统摄像机标定计算图像透视关系,也不需要背景建模去掉图像背景;2)模型性能较好,在准确性和鲁棒性上都取得了较好的效果。我们在 ShanghaiTech 数据集 PartA 的准确率比之前最好的 MCNN 方法提高了 14.58%,在 ShanghaiTech 数据集 PartB 的准确率比 MCNN 方法提高了 40.53%。均方根误差分别降低了 23.89%,33.90%,其很好地融合了不同尺度卷积核和不同层次的特征。

未来我们计划将残差网络(ResNet)应用到人群计数问题上,并借鉴检测(Detection)领域的应用。

参 考 文 献

- [1] LIN S F, CHEN J Y, CHAO H X. Estimation of number of people in crowded scenes using perspective transformation[J]. IEEE Transactions on Systems, Man & Cybernetics Part A Systems & Humans, 2001, 31(6): 645-654.
- [2] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2005: 886-893.
- [3] WANG M, WANG X. Automatic adaptation of a generic pedestrian detector to a specific traffic scene[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2011: 3401-3408.
- [4] GE W, COLLINS R T. Marked point processes for crowd-counting[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2009(CVPR 2009). IEEE, 2009: 2913-2920.
- [5] IDREES H, SOOMRO K, SHAH M. Detecting Humans in Dense Crowds Using Locally-Consistent Scale Prior and Global Occlusion Reasoning[M]. IEEE Computer Society, 2015.
- [6] LIN Z, DAVIS L S. Shape-Based Human Detection and Segmentation via Hierarchical Part-Template Matching[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(4): 604-618.
- [7] LEMPITSKY V S, ZISSERMAN A. Learning To Count Objects in Images[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc., 2010: 1324-1332.

- 恢复[J]. 科学技术与工程, 2016(5):177-181.
- [7] XIAN G, YAN C X, ZHANG X J. Motion-blurred image point spread function estimation method based on spectral analysis [J]. Chinese Journal of Liquid Crystals and Displays, 2014, 29(5):751-755. (in Chinese)
贤光, 颜昌翔, 张新洁. 运动模糊图像点扩散函数的频谱估计法[J]. 液晶与显示, 2014, 29(5):751-755.
- [8] LOKHANDE R, ARYA K V, GUPTA P. Identification of parameters and restoration of motion blurred images[C]// ACM Symposium on Applied Computing. 2006:301-305.
- [9] SHI M Z, XU T F, ZHANG K. Total variation image restoration for mixed blur in moving image[J]. Optics and Precision Engineering, 2011, 19(8):1973-1981. (in Chinese)
石明珠, 许廷发, 张坤. 运动成像混合模糊的全变分图像复原[J]. 光学精密工程, 2011, 19(8):1973-1981.
- [10] SHI M Z, XU T F, LIANG J, et al. PSF estimation via gradient cepstrum analysis for single blurred image[J]. Acta Physica Sinica, 2013, 62(17):174204. (in Chinese)
石明珠, 许廷发, 梁炯, 等. 单幅模糊图像点扩散函数估计的梯度倒谱分析方法研究[J]. 物理学报, 2013, 62(17):174204.
- [11] DESHPANDE A M, PATNAIK S. A novel modified cepstral based technique for blind estimation of motion blur[J]. Optik - International Journal for Light and Electron Optics, 2014, 125(2):606-615.
- [12] LI H S, ZHANG Y N, YAO R, et al. Parameter estimation of linear motion blur based on principal component analysis[J]. Optics and Precision Engineering, 2013, 21(10):2656-2663. (in Chinese)
李海森, 张艳宁, 姚睿, 等. 基于主成分分析的直线运动模糊参数估计[J]. 光学精密工程, 2013, 21(10):2656-2663.
- [13] YAN R M, SHAO L. Blind Image Blur Estimation via Deep Learning[J]. IEEE Transactions on Image Processing, 2016, 25(4):1910-1921.
- [14] DASH R, MAJHI B. Motion blur parameters estimation for image restoration[J]. Optik - International Journal for Light and Electron Optics, 2014, 125(5):1634-1640.
- [15] ZHANG M Q. Research on parameters estimation and restoration of motion-blurred images[D]. Chengdu: University of Electronic Science and Technology of China, 2009. (in Chinese)
张明谦. 运动模糊图像的运动参数估计及复原研究[D]. 成都: 电子科技大学, 2009.
- [16] SHI M Z. Motion Blur Image Resroration Theory and Key Technology based on Total Variation[D]. Beijing: Beijing Institute of Technology, 2014. (in Chinese)
石明珠. 运动模糊图像全变分复原理论及关键技术研究[D]. 北京: 北京理工大学, 2014.
-
- (上接第 21 页)
- [8] ZHANG C, LI H, WANG X, et al. Cross-scene crowd counting via deep convolutional neural networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015:833-841.
- [9] WANG C, ZHANG H, YANG L, et al. Deep People Counting in Extremely Dense Crowds[C]// ACM International Conference on Multimedia. ACM, 2015:1299-1302.
- [10] BOOMINATHAN L, KRUTHIVENTI S S S, BABU R V. CrowdNet: A Deep Convolutional Network for Dense Crowd Counting[C]// Proceedings of ACM Conference on Multimedia (ACMMM) - 2016. 2016:640-644.
- [11] ZHANG Y, ZHOU D, CHEN S, et al. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network[C]// Computer Vision and Pattern Recognition. IEEE, 2016:589-597.
- [12] HAN S, POOL J, TRAN J, et al. Learning both Weights and Connections for Efficient Neural Networks[C]// NIPS 2015. 2015:1135-1143.
- [13] HAN S, LIU X, MAO H, et al. EIE: Efficient Inference Engine on Compressed Deep Neural Network[C]// ACM/IEEE International Symposium on Computer Architecture. IEEE, 2016:243-254.
- [14] HAN S, MAO H, DALLY W J. Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding[J]. Fiber, 2015, 56(4):3-7.
- [15] LIN M, CHEN Q, YAN S. Network In Network[C]// International Conference on Learning Representations. 2013.
- [16] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]// International Conference on International Conference on Machine Learning. Omnipress, 2010:807-814.
- [17] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:770-778.
- [18] RODRIGUEZ M, LAPTEV I, SIVIC J, et al. Density-aware person detection and tracking in crowds[C]// International Conference on Computer Vision. IEEE Computer Society, 2011:2423-2430.
- [19] IDREES H, SALEEMI I, SEIBERT C, et al. Multi-source Multi-scale Counting in Extremely Dense Crowd Images[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013:2547-2554.
- [20] OÑORO-RUBIO D, LÓPEZ-SASTRE R J. Towards Perspective-Free Object Counting with Deep Learning[C]// European Conference on Computer Vision. Springer, Cham, 2016:615-629.