2020 年 10 月 第 10 期
Vol. 41 No. 10 2020

多通道融合分组卷积神经网络的人群计数算法

严芳芳,吴 秦

(江南大学 江苏省模式识别与计算智能工程实验室,江苏 无锡 214122)
E-mail: qinwu@jiangnan.edu.cn

摘 要: 相机视角引起的头部尺度多变性和人群分布的多样性是图像人群计数中存在的两个主要挑战,很多方法试图通过采用多列或者多分支网络来解决这些问题,但由于受列数或分支数的限制,提取的特征尺度有限. 本文提出一种面向人群计数的多通道融合分组卷积神经网络,该网络主要由两个部分组成: 采用预训练的 VGG 网络前 10 层作为基础主干网络,以多通道融合分组卷积模块作为网络的第二部分,多通道融合分组卷积模块是本网络的关键组成部分,该模块中每个分组卷积模块都与其他层之间密集相连以获得不同层次的特征,同时,我们引入分组卷积来减少网络参数. 在三个公开的数据集(ShanghaiTech,UCF_CC_50,UCF_QNRF)上进行验证,实验结果证明了本文所提方法的有效性.

关 键 词: 人群计数; 尺度多变性; 分布多样性; 特征融合; 分组卷积
中图分类号: TP391 **文献标识码:** A **文 章 编 号:** 1000-1220(2020)10-2200-06

Crowd Counting Algorithm Based on Multi-channel Fusion Group Convolutional Neural Network

YAN Fang-fang, WU Qin
(Jiangsu Provincial Engineering Laboratory for Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi 214122, China)

Abstract: The diversity of head scales and the diversity of crowd distribution caused by camera angle are the two main challenges in crowd counting. Many methods tried to use multi-column or multi-branch networks to solve these problems, but due to the number of columns or branches, the extracted feature scales are limited. This paper proposes a multi-channel fusion group convolutional neural network for crowd counting. The network is composed of two major components: the first 10 layers of VGG network are used as the basic backbone network, and the multi-channel fusion group convolutional module is proposed to be the second part of network, the multi-channel fusion group convolutional module is the key components of this network. Each group convolution module in this module is densely connected with other layers to obtain different levels of features. Meanwhile, we introduce group convolution to reduce network parameters. This approach is evaluated on three benchmark datasets (ShanghaiTech, UCF_CC_50, UCF_QNRF), and the experimental results demonstrate the effectiveness of the proposed method.

Key words: crowd counting; scale variability; distribution diversity; feature fusion; group convolution

1 引 言

计算机视觉中人群计数工作是通过学习图片或者视频得到其中包含的人数. 在公共集会、体育赛事等场景中,为了帮助控制人群和公共安全,需要精确的人数信息. 另外参与人数或人群密度是未来活动规划和空间设计的重要信息.

人群计数也存在着很多其他视觉领域同样存在的问题,密集场景图片中人群计数存在遮挡,尺度变化以及背景噪声等问题.

目前,人群计数领域两类主要方法分别是基于传统的人群计数方法和基于深度学习的人群计数方法. 早期,主要是通过一些传统的方法来完成人群计数,例如通过检测的方式[1,2]得到人数,或者通过回归的方式[3,4]得到人数,但是这些方法在严重拥挤的场景下性能较差. 近年来,基于深度学习的方法常被用来完成人群计数任务. 例如,通过卷积神经网络(Convolutional Neural Networks, CNN)进行精确的人群密度图生成或人群计数[5,6]. Zhang 等人[7]提出 MCNN(Multi-Column Convolutional Neural Network)网络,利用三列不同大小的卷积核提取不同尺度的特征,在一定程度上解决了尺度变化问题. Li 等人[8]利用预训练的 VGG16 网络前 10 层以及结合空洞卷积(Dilated Convolution)得到了较高分辨率的密度图. 多列或者多分辨率[9-11]的方法在一定程度上缓解了尺度多变问题,但仍然受到卷积核大小和多列结构的限制,同时多列结构带来两个显著缺点: 大大增加网络的训练时间和冗余的分支结构.

针对上述问题,本文提出一个端到端的多通道融合分组卷积神经网络,避开多列结构获取不同特征的方式,多通道融合分组卷积神经网络跨层将不同网络深度连接起来,融合不同网络深度的特征得到更加丰富的特征信息. 综上所述,本文有以下两个主要贡献:

收稿日期: 2019-11-11 收修改稿日期: 2020-04-07 基金项目: 国家自然科学基金项目(61972180)资助; 江苏省自然科学基金项目(BK20181341)资助. 作者简介: 严芳芳,女,1994 年生,硕士研究生,CCF 会员,研究方向为机器学习、计算机视觉; 吴 秦,女,1978 年生,博士,副教授,CCF 会员,研究方向为计算机视觉、机器学习、人工智能.

1) 提出了一个新的人群计数算法,多通道融合分组卷积神经网络. 在不同网络深度之间均建立网络通路,得到丰富的网络特征.

2) 多通道之间跨层连接导致网络参数增加,为了缓解这一问题,我们采用两种措施: ①在网络中加入 1×1 卷积层实现特征降维; ②在多通道融合分组卷积模块中,引入分组卷积替代普通的卷积操作.

与已有方法相比较,本文在三个公开数据集 ShanghaiTech[7],UCF_CC_50[12],UCF_QNRF[13] 上的实验结果均有所提升.

2　相关工作

基于深度学习强大的学习能力,近年来在很多计算机视觉领域取得了很好的成果,例如图像识别,图像分割以及目标检测等. 同样,人群计数领域目前效果较好的方法也多数是基于深度学习方法.

传统的方法通常先对图片进行分割,然后分别对分割后的图像块进行特征提取,再通过回归[14]或者分类[15]得到人数,最后图像块人数相加得到图片人数. 选择合适的特征是传统方法的关键,然而,CNN 训练过程中使用很多特征图,可以被训练来自动提取合适于特定任务的特征. 基于 CNN 的人群计数方法则不需要人为选择手工特征以及前背景分割,直接将图片输入网络,由网络学习得到高层特征,最后生成密度图或者通过回归得到人数. 相比传统的方法,基于 CNN 的方法结果更加有竞争性. Cong 等人[16]提出一个六层的卷积网络进行密度图生成和人数估计,训练的时候两个任务交替优化,完成跨场景的人群计数问题. Zhang 等人[7]提出一个多列卷积神经网络(MCNN),该网络有三个不同的分支网络,每个分支的深度相同但是采用不同大小的卷积核,三个分支分别得到不同的感受野,来抓取不同尺寸目标的特征. 最后三个分支网络提取的特征图通过 1×1 的卷积融合生成人群密度图. Sam 等人[17]的设计包括三个子网络和一个分类器,首先将一张图片裁剪成几个图像块,使用分类网络分成不同密度级别,然后让不同密度等级的图像块通过相应的子网络得到人数,最后由图像块人数之和得到图片人数. 该结构的训练方式采用先用所有的训练数据对三个分支进行预训练,再利用分类网络完成密度级别划分,最后根据分类结果再次将训练图片送到对应的分支网络完成子网络的训练. Sindagi 等人[18]通过三个网络共同完成计数任务: 全局分类网络,局部分类网络和特征提取网络. 和其他工作不同,除了特征提取网络,该设计还结合了图片的全局和局部的密度信息来辅助解决图片中人群分布不均的问题. Li 等人[8]指出多列网络的冗余性,并摒弃了多列的网络结构,使用简单的单列网络结构,选取 VGG16 的前 10 层作为网络第一部分,网络的第二部分采用空洞卷积(Dilated Convolution)替代了普通卷积. 空洞卷积通过填充卷积核扩大了网络的感受野(Receptive Field),因此网络中不需要过多的池化层,从而得到了较高分辨率的密度图. Cao 等人[19]受到图像识别领域的 Incenption[20]网络结构启发,编码器使用尺度聚合模块(类 inception 结构)提高网络的表达能力,并且能提取多种尺度的特征,解码器由卷积和转置卷积组成,可以生成与输入图片相同分辨率的密度图,在网络的损失函数部分使用欧式距离损失函数和局部一致性(SSIM)损失函数的结合. 利用预测密度图和真实密度图之间的结构相似性(structural similarity index; SSIM)来构成网络的局部一致性损失. SSIM 和人类视觉系统(human visual system,HVS)类似,考虑了人类视觉感知,得到的结果比欧式距离损失函数得到的结果包含更多的细节信息.

3　多通道融合分组卷积神经网络

3.1　多通道融合分组卷积神经网络结构

相机视角引起的头部尺度多变性和人群分布的多样性是人群计数中存在的两个主要挑战. 针对这一问题,本文设计了多通道融合分组卷积神经网络,网络框架如图 1 所示. 我们在网络中加入大量的跨层连接,使得特征和梯度在不同层之间高效流通,从而达到特征复用的功效,通过融合不同层的特征得到丰富的多尺度特征.

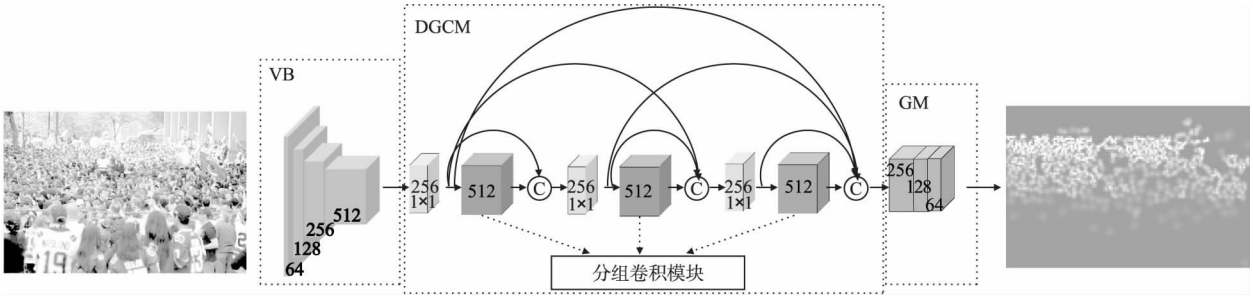

图 1　多通道融合分组卷积神经网络框架

Fig. 1　Framework of mutil-channel fusion group convolution network

在密集场景下的人群图片中单个目标较小,选择更深层的网络可能丢失小目标的信息,因此我们选择 VGG 网络前 10 层(如图 1 中 VB(VB: VGG Backbone))作为我们的基础主干网络.

识别不同大小的目标是计算机视觉领域的一个挑战,同样,人群图片中也存在头部尺寸多变、人群分布多样等问题. 主流的卷积神经网络由卷积层或池化层顺序连接而成,获取网络的高层特征来进行下一步的处理,使用高层特征对于分类或者检测等问题,可以得到较好的结果. 但是,由于密集人群图片中目标个数众多,每个目标较小且尺度多变,浅层特征

对于人群计数来说也很重要. 根据文献[21]提出的DenseNet,网络不同层的特征融合有助于学习到更多的信息. 我们通过结合多种层次的特征来处理多尺度和小目标问题. 基于这一原因,我们提出了多通道融合分组卷积模块(如图1中DGCM(DGCM: dense group convolution module)). 为了增加网络中层与层之间的信息流,将模块中所有层两两相连,使得网络中每一层都接受它前面所有层的输入. 跨层连接通过特征通道上的连接(concatenate)操作来实现特征复用. 同时为了避免特征维度增加过快,在每一次特征连接之后,通过1×1的卷积将通道数恢复到原值.

在网络的最后,我们通过生成模块(如图1中GM(GM: generating module))得到密度图,生成模块由两层3×3卷积和一层1×1卷积组成. 3×3卷积逐步减少特征图的通道数,最后,用通道数为1的1×1卷积作为输出. 所以网络的输出为一张长宽各为原图1/8的单通道密度图.

结合多种层次特征图的方式在解决多尺度问题的同时也增加了网络参数. 为了解决这一问题,我们设计了分组卷积模块,如图2所示. 在该模块中,首先通过1×1卷积降维,然后用分组卷积替代了普通卷积,最后设计了一个跨层连接进行特征融合,保持输入输出特征维度一致. 与普通卷积操作相比,分组卷积参数更少(分组卷积细节见3.2节),而跨层连接可以获得更加丰富的特征.

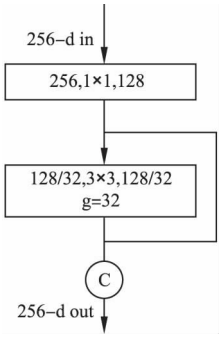

图2 分组卷积模块

Fig. 2 Group convolution module

根据文献[22]的实验结果:网络结构中使用更多的小卷积核比使用更少的大卷积核要有效,且更加节省参数. 因此我们的网络中卷积操作均选用3×3的卷积核. 此外,我们在网络中多次使用1×1卷积,它是一个非常好的结构,可以跨通道组织信息,提高网络的表达能力,完成特征通道升维或降维.

我们的网络脱离了原有的加深变宽的思想,缓解网络加深或变宽之后带来的一系列问题,从特征角度出发,结合多层次的特征图来处理多尺度的问题. 这样的设置结合信息流和特征复用两大优势既大幅的减少了网络的参数量,又在一定程度上缓解了梯度消失的问题.

3.2 分组卷积

分组卷积的思想最早出现在AlexNet[23]中. 与普通的卷积网络相比,相同的卷积操作分组卷积所需计算的参数更少,不容易过拟合. 因此,我们在网络的多通道融合分组卷积模块中引入分组卷积替代普通卷积.

分组卷积首先将输入数据分组,然后对每组数据分别进行卷积操作. 假设输入数据的尺寸是$W\times H\times C_1$,输出数据尺寸是$W\times H\times C_2$,卷积核大小均为$K\times K$. 普通卷积与分组卷积的差异如图3所示,图3上方是普通卷积,下方是分组卷积(图中组数为2). 对于分组卷积,若设定分组数为g组,则每组的输入数据尺寸为$W\times H\times(C_1/g)$,输出数据尺寸为$W\times H\times(C_2/g)$,卷积核尺寸为$K\times K\times(C_1/g)$,个数为C_2/g,每组卷积核只与同组的输入数据卷积,而不与其他组的输入数据卷积,最后所有组输出共同组成输出数据. 在一次卷积操作中,普通卷积的参数个数为:$C_1\times C_2\times K\times K$,而分组卷积的参数个数为:$(C_1/g)\times(C_2/g)\times K\times K\times g$. 普通卷积的参数是分组卷积的g倍.

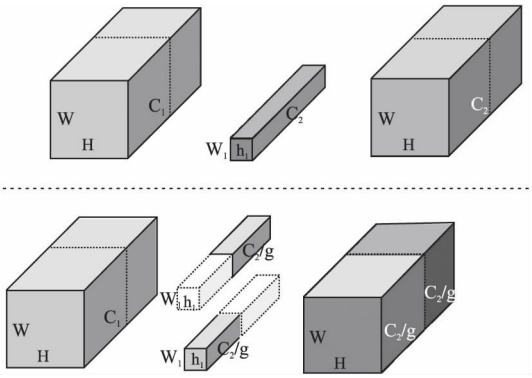

图3 普通卷积(上)与分组卷积(下)

Fig. 3 Normal convolution(top) and group convolution(bottom)

3.3 密度图生成

不同于传统的基于检测和回归的方法,对于稠密的人群图片,通过密度图来计数可以获得更准确、全面的信息. 在我们的方法中,通过卷积神经网络学习图像的局部特征和其相应的密度图之间的映射,从而将图像中包含的空间信息加入计数的过程中.

由于密度图遵循逐像素预测,因此输出密度图必须包含空间相关性,这样才能在最近的像素之间呈现平滑的过渡. 现有的数据集中仅提供了人头位置坐标,我们需要根据人头位置坐标信息进一步处理得到对应密度图. 这里我们采用文献[7]中提到的方法,使用高斯分布去替换人头的位置. 首先,我们用$\delta(z-z_i)$表示在像素点z_i的人头标签值,$\delta(z-z_i)=1$表示像素点z_i处有一个人,$\delta(z-z_i)=0$则表示像素点z_i处没有人. 则含有V个人头标签的密度图的计算方式如公式(1)所示:

$$E(z)=\sum_{i=0}^{V}\delta(z-z_i) \tag{1}$$

我们用二维高斯分布函数去替换每一个人头位置坐标,将密度图转化成一个连续密度图,相对应的标签密度图$D(z)$计算方式如公式(2)所示:

$$D(z)=\sum_{i=0}^{V}\delta(z-z_i)\times G_{u,\rho^2}(z) \tag{2}$$

V表示人群图片中包含的人数,z_i表示图片中第i个人头标签的坐标,$G_{u,\rho^2}(z)$表示均值为u,方差为ρ^2的二维高斯函数.

3.4 损失函数

我们采用端到端的方式进行网络训练,由于网络输出的

预测密度图为原图的 1/8,我们将真实密度图长宽分别缩放为原图的 1/8,保持预测密度图和真实密度图分辨率大小一致. 然后采用欧式距离来评估预测密度图与真实密度图之间的相似性. 欧式距离损失函数定义如公式(3):

$$loss = \frac{1}{M}\sum_{l=0}^{M} \| \hat{D}(X_l - D(X_l) \|_2^2 \tag{3}$$

M 表示一个训练批次的图片张数,$D(X_l)$ 表示第 l 张输入图片的预测密度图,$D(X_l)$ 表示第 l 张输入图片的真实密度图.

4　实　验

4.1　数据集

我们在三个公开的数据集上验证我们的实验方法,下面简单介绍这三个数据集.

ShanghaiTech[7]: ShanghaiTech 数据集由 partA 和 partB 组成,其中 partA 由 482 张网络中随机选择的图片组成,partB 由不同时间段随机拍摄于上海街道上的图片组成. 这两部分又分别划分成训练数据集和测试数据集,partA 的训练数据集和测试数据集分别有 300 和 182 张图片,而 partB 的训练数据集和测试数据集分别有 400 和 316 张图片.

UCF_CC_50[12]: UCF_CC_50 数据集为 50 张不同场景下各种密度以及视角失真的图片,每一张图片中包含的人数从 94 到 4543 不等. 但由于数据集图片数量较少,本文采用交叉验证协议[12]进行训练和测试,其中数据集被均分成 5 组,并进行 5 次交叉验证.

UCF-QNRF[13]: UCF-QNRF 数据集拍摄于不同的野外真实场景,并拥有最多样化的视角、密度和光照变化的场景下的密集人群图片,克服了以往数据集中标注不准确,图片像素低,数据集图片少等缺点. 整个数据集包含 1535 张图片与 1251642 个人头位置注释,单张图片人数从 95 到 12865 不等. 通过图片的标注点进行排序,每 5 张图片中选择一张作为测试图片,生成训练数据集和测试数据集,训练数据集由 1201 张图片组成,测试数据集由 334 张图片组成.

4.2　实验细节

4.2.1　数据增强

由于人群图片标注困难,人群数据集图片数量有限,为了更好的训练网络,我们采用两种方式对训练数据集进行数据增强操作. 对于每一张训练图片,我们以原图1/4大小裁剪成

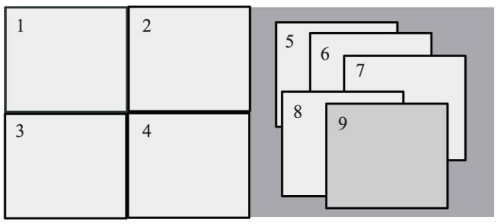

图 4　裁剪示意图

Fig. 4　Sample of cropping

9 张,前四张选择原图不重复位置的 1/4 大小,后面五张以原图 1/4 大小随机裁剪,如图 4 所示. 然后,对图片进行随机翻转,获得更多的训练图片.

4.2.2　实验设置

我们基于 pytorch 深度学习框架实现多通道融合分组卷积神经网络,基于 Imagenet[24] 预训练的 VGG16 进行网络前 10 层的参数初始化,网络的其他部分参数利用均值为 0 方差为 0.01 的高斯函数随机初始化. 网络训练时使用动量为 0.9 的随机梯度下降(Stochastic gradient descent,SGD)作为我们模型的优化器,学习速率设置为 1e-7,随迭代次数自适用调整. UCF-QNRF 数据集中的图片平均分辨率为 2013×2902,为了方便训练,我们将 UCF-QNRF 数据集中所有图片裁剪成尺寸为 1024×1024. 我们以训练数据集的绝对误差作为衡量模型收敛的标准,当训练数据集的绝对误差不再下降时停止训练.

4.2.3　评价指标

模型的性能通过预测人数与真实标注人数的绝对误差(Mean Absolute Error,MAE)和均方误差(Mean Square Error,MSE)来衡量,其值越小越表示模型误差越小,即性能越好. MAE,MSE 计算公式如式(4)、式(5):

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | \hat{C}(X_i) - C(X_i) | \tag{4}$$

$$MSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{C}(X_i) - C(X_i))^2} \tag{5}$$

N 表示测试图片张数,$C(X_i)$ 表示第 i 张测试图片的标注人数,$\hat{C}(X_i)$ 表示第 i 张测试图片的预测人数,计算方式如公式(6):

$$\hat{C}(X_i) = \sum_{w=1}^{W}\sum_{h=1}^{H} \hat{D}(w,h) \tag{6}$$

X_i 表示输入图片,w,h 分别表示生成的预测密度图的长和宽,$\hat{D}(w,h)$ 表示预测密度图中(w,h)位置的像素值.

4.3　对比实验

为了验证多通道融合分组卷积神经网络的有效性,我们在 ShanghaiTech 数据集上做了 4 组实验.

通过控制网络的深度,我们分别进行了两组实验,对比深度为 13(实验 1)和 10(实验 2)的两种网络,并在其后接入生成模块,通过对比 MSE 和 MAE 两个指标,选择较好的主干网络结构.

为了验证提出的多通道融合分组卷积模块的有效性,我们设计了另外两组不同设置的实验:

实验 3 使用 VGG16 的前 10 层作为基础主干网络,后接密集卷积模块(卷积层是正常卷积),最后接生成模块.

实验 4 使用 VGG16 的前 10 层作为基础主干网络,后接多通道融合分组卷积模块模块(卷积层是使用分组卷积模块替代正常卷积),最后接生成模块.

实验 1-实验 4 的对比实验结果如表 1 所示,实验 2 的 MAE 和 MSE 的值均比实验 1 的低,表明深度为 10 的网络比深度为 13 的网络作为主干网络的效果好,因此我们实验选择深度为 10 的网络作为我们的主干网络. 由于高密度的人群图片中,每个目标人头占据很小的分辨率,而相比 VGG10,VGG13 网络更深,反而会丢失图片中的小目标,不利于人群图片中的人头特征提取,所以 VGG10 作为主干网络更有利于人群图片的特征提取. 实验 3 的结果优于实验 2,证明我们设计的密集卷积模块是有效的. 实验 4 的结果优于实验 3,证明选择分组卷积模块替代正常卷积在我们的网络中减少网络参数的同时,不影响网络的计数误差,并进一步减少误差.

4.4 与其他算法的比较

为了验证本文方法的有效性,我们在三个常用的数据集(4.1 节中介绍)上进行实验,并与一些当前领先的结果[6-8,11,17,18]作比较.

表 1 在 ShanghaiTech 数据集上对比实验结果

Table 1 Results of ablation experiments on ShanghaiTech

设 置	partA		partB	
	MAE	MSE	MAE	MSE
实验 1(VGG13 + 生成模块):	67.6	103.5	10.1	16.8
实验 2(VGG10 + 生成模块):	65.4	98.6	9.7	16.0
实验 3(VGG10 + 密集卷积模块 + 生成模块):	64.2	96.9	9.0	14.7
实验 4(VGG10 + 多通道融合分组卷积模块模块 + 生成模块):	63.3	95.8	8.3	13.6

ShanghaiTech[7]:在 ShanghaiTech 的实验结果如表 2 所示. 在 partA 数据集上,本文所提方法的比其他方法中最优的结果相比,MAE 降低 5.5%(两种算法的相对误差),MSE 降低 8.2%. 在 partB 数据集上,本文的结果与当前最优结果不相上下.

表 2 ShanghaiTech 上实验结果

Table 2 Results of ShanghaiTech

方法	partA		partB	
	MAE	MSE	MAE	MSE
MCNN[7]	110.2	173.2	26.4	41.3
CMTL[6]	101.3	152.4	20.0	31.1
Switch-CNN[17]	90.4	135.0	20.1	30.1
CSRNnet[8]	68.2	115	10.6	16.0
SaNet[19]	67.0	104.5	8.4	13.6
Ours	63.3	95.8	8.3	13.6

UCF_CC_50[12]:在 UCF_CC_50 上的实验结果如表 3 所示. 本文方法的结果与其他方法中最优的结果相比,MAE 降低 22.2%,MSE 降低 19.7%.

表 3 UCF_CC_50 上实验结果

Table 3 Results of UCF_CC_50

UCF_CC_50	MAE	MSE
MCNN[7]	377.6	509.1
CMTL[6]	322.8	397.9
Switch-CNN[17]	318.1	439.2
CSRNet[8]	266.1	397.5
SaNet[19]	258.4	334.9
Ours	200.9	268.9

UCF-QNRF[13]:在 UCF-QNRF 上的实验结果如表 4 所示. 本文方法的结果与其他方法中最优的结果相比,本文方法的 MAE 降低 20.5%,MSE 降低 10.7%.

表 4 UCF-QNRF 上实验结果

Table 4 Results UCF-QNRF

UCF-QNRF	MAE	MSE
Idrees et al[11]	315.0	508.0
MCNN[7]	277.0	426.0
CMTL[6]	252.0	514.0
Switch-CNN[17]	228.0	445.0
CL-CNN[13]	132.0	191.0
Ours	105.0	170.5

与原始标签相比,密度图更加直观更有利于视觉上的对比,图 5 展示了本文算法和对比算法 CSRnet[8] 在一些测试图片上生成的密度图. 从左到右,分别是测试图片、真实密度图、CSRnet算法生成的预测密度图、本文算法生成的预测密度图;从上到下,测试图片分别取自 ShanghaiTech partA、ShanghaiTech partB、UCF_CC_50 和 UCF-QNRF.

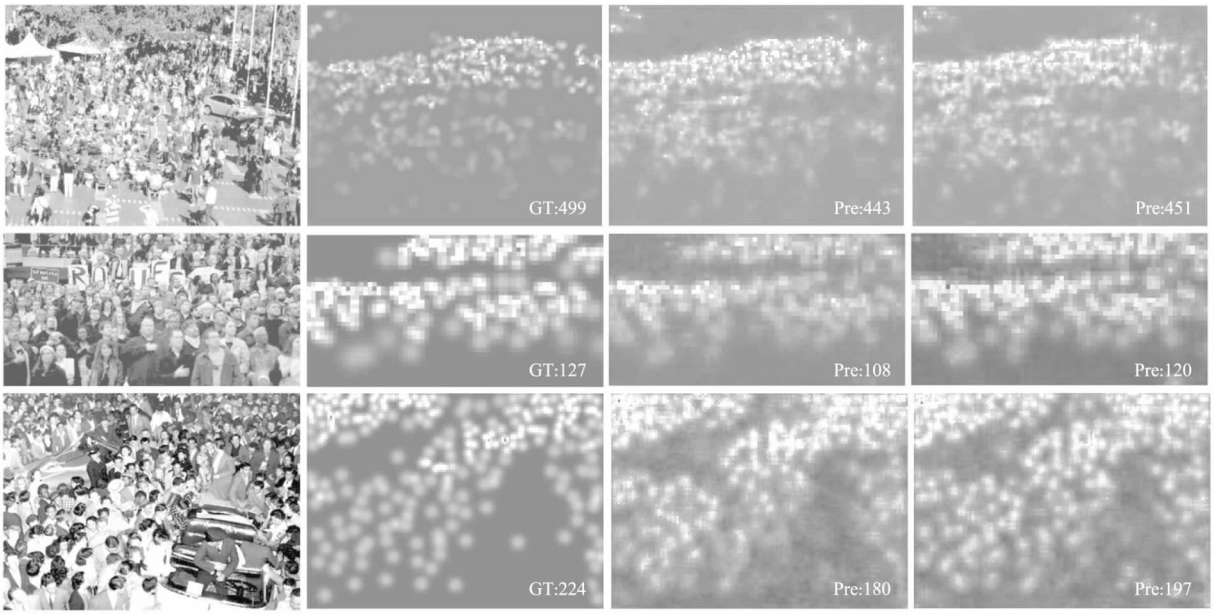

图 5 一些测试图片及其密度图实例

Fig. 5 Some samples of testing images and their density maps

5　结束语

人群计数在灾害控制、空间规划等方面有着广泛的应用. 由于相机视角引起的头部尺度变化大和人群分布多样等问题,精确完成图片人群计数任务仍然存在很大的挑战. 本文提出的多通道融合分组卷积神经网络建立不同层之间的连接,通过通道上的密集连接来融合不同层的特征,特征和梯度通过密集连接的形式来实现更加有效的传递,从而使得网络训练更加简单,此外,跨层连接可以避免网络过深带来梯度消失问题. 同时在网络中引入分组卷积模块,减少参数的同时充分利用特征信息. 在三个公开的数据集上的实验结果显示,本文方法的绝对误差和均方误差比其他方法有所下降,验证了本文方法的有效性.

在实验中,我们还发现,通过密度图获得的位置信息并不十分精确,这在一定程度上将影响到人群计数的精确性. 在接下来的工作中,我们将研究如何在计数的同时得到精确的位置信息,以进一步提升人群计数正确率.

References:

[1] Lin Z,Davis L S. Shape-based human detection and segmentation via hierarchical part-template matching [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2010,32(4):604-618.

[2] Bo Wu,Nevatia R. Detection of multiple,partially occluded humans in a single image by Bayesian combination of edgelet part detectors [C]//Tenth IEEE International Conference on Computer Vision(ICCV'05) Volume 1,Beijing,2005:90-97.

[3] Ryan D,Denman S,Fookes C,et al. Crowd counting using multiple local features [C]//Digital Image Computing:Techniques and Applications,Melbourne,VIC,2009:81-88.

[4] Qin Xun-hui,Wang Xiu-fei,Zhou Xi,et al. Counting people in various crowed density scenes using support vector regression [J]. Journal of Image and Graphics,2013,18(4):392-398.

[5] 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,New York,NY,USA,2016:640-644.

[6] Sindagi V A,Patel V M. Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting [C]//2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance(AVSS),IEEE,Lecce,Italy,2017:1-6.

[7] Zhang Y,Zhou D,Chen S,et al. Single-image crowd counting via multi-column convolutional neural network [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Las Vegas,NV,USA,2016:589-597.

[8] Li Yu-hong,Zhang Xiao-fan,Chen De-ming. CSRNet:dilated convolutional neural networks for understanding the highly congested scenes [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,Salt Lake City,UT,2018:1091-1100.

[9] Ma Hao,Yin Bao-qun,Peng Si-fan. Population counting algorithm based on characteristic pyramid network [J]. Computer Engineering,2019,45(7):203-207.

[10] Zhang Xiu-ling,Dong Xiao-peng. Mutil-channel cross fusion deep resnet for offline handwritten chinese character recognition [J]. Journal of Chinese Computer Systems,2019,40(10):2232-2235.

[11] Idrees H,Saleemi I,Seibert C,et al. Multi-source multi-scale counting in extremely dense crowd images [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition,Portland,OR,USA,2013:2547-2554.

[12] Oñoro-Rubio D,López-Sastre R J. Towards perspective-free object counting with deep learning [C]//Proceedings of the European Conference on Computer Vision(ECCV),Amsterdam,Netherlands,2016:615-629.

[13] 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),Munich Germany,2018:532-564.

[14] Tuzel O,Porikli F,Meer P. Pedestrian detection via classification on riemannian manifolds [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2008,30(10):1713-1727.

[15] Wang M,Wang X. Automatic adaptation of a generic pedestrian detector to a specific traffic scene [C]//2011 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Colorado Springs,CO,USA,2011:3401-3408.

[16] Cong Zhang,Hongsheng Li,Wang X,et al. Cross-scene crowd counting via deep convolutional neural networks [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Boston,MA,USA,2015:833-841.

[17] Sam,Deepak Babu,Shiv Surya,et al. Switching convolutional neural network for crowd counting [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Honolulu,HI,2017:4031-4039.

[18] Sindagi V A,Patel V M. Generating high-quality crowd density maps using contextual pyramid CNNs [C]//2017 IEEE International Conference on Computer Vision(ICCV),Venice,2017:1879-1888.

[19] Xinkun Cao,Zhipeng Wang,Yanyun Zhao,et al. Scale aggregation network for accurate and efficient crowd counting [C]//Proceedings of the European Conference on Computer Vision(ECCV),Munich Germany,2018:757-773.

[20] Szegedy C,Wei Liu,Yangqing Jia,et al. Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Boston,MA,USA,2015:1-9.

[21] Huang G,Liu Z,Van Der Maaten L,et al. Densely connected convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:4700-4708.

[22] Szegedy C,Vanhoucke V,Ioffe S,et al. Rethinking the inception architecture for computer vision [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,Las Vegas,NV,2016:2818-2826.

[23] Krizhevsky A,Sutskever I,Hinton G E. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems,Lake Tahoe,2012:1097-1105.

[24] Jia Deng,Wei Dong,Richard Socher,et al. ImageNet:a large-scale hierarchical image database [C]//2009 IEEE Conference on Computer Vision and Pattern Recognition,Miami,FL,2009:248-255.

附中文参考文献:

[4] 覃勋辉,王修飞,周　曦,等. 多种人群密度场景下的人群计数[J]. 中国图象图形学报,2013,18(4):392-398.

[9] 马　皓,殷保群,彭思凡. 基于特征金字塔网络的人群计数算法[J]. 计算机工程,2019,45(7):203-207.

[10] 张秀玲,董逍鹏. 多通道交叉融合的深度残差网络脱机手写汉字识别 [J]. 小型微型计算机系统,2019,40(10):2232-2235.