

# 基于多任务学习及由粗到精的卷积神经网络人群计数模型



陈训敏 叶书函 詹瑞

四川大学电子信息学院 成都 610065

**摘要** 人群计数是指计算单张图像或单个视频帧中人的数目,为了解决人群任务的计数不够准确的问题,提出了一种基于多任务学习及由粗到精的卷积神经网络人群计数模型。首先,多任务学习是指引入与原始任务相关的辅助任务,指导主要任务的学习,人群密度估计是人群计数模型的主要任务,人群分割任务作为辅助任务以提高网络性能。其次,由粗到精策略表明人群计数模型预测密度图是一个由粗糙到精细的过程,即生成粗糙且不准确的人群密度图,结合人群分割图后得到准确的人群密度图。在 Shanghai Tech 数据集 Part A 部分、Part B 部分和 UCF\_CC\_50 数据集上的实验表明,所提人群计数模型相比之前最好的 CSRNet 模型绝对误差分别降低了 4.55%,14.15%,19.09%,均方误差分别降低了 10.00%,19.09%,19.47%,显著提高了人群计数模型的准确性和鲁棒性。

**关键词:** 人群计数;卷积神经网络;人群密度估计;人群分割;多任务学习

**中图法分类号** TP391

## Crowd Counting Model of Convolutional Neural Network Based on Multi-task Learning and Coarse to Fine

CHEN Xun-min, YE Shu-han and ZHAN Rui

College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China

**Abstract** Crowd counting refers to counting the number of people in a single image or a single video frame. In order to solve the problem of insufficient counting of crowd tasks, a crowd counting model based on multi-task learning and coarse to fine convolutional neural network is proposed. Firstly, multi-task learning means introducing auxiliary tasks related to the original task to guide the learning of the main tasks. The crowd density estimation is the main task of the crowd counting model, and the crowd segmentation task is used as an auxiliary task to improve network performance. Secondly, the proposed crowd counting model is able to predict the density map from coarse to fine. A rough and inaccurate crowd density map is generated, which is combined with the crowd segmentation map to obtain an accurate crowd density map. Experiments on the Shanghai Tech dataset Part A and Part B, and UCF\_CC\_50 dataset show that the proposed crowd counting model outperforms the state of the art CSRNet models by 4.55%, 14.15% and 19.09% respectively, and the mean square error is reduced by 10.00%, 19.09% and 19.47% respectively compared with the SOTAs. The proposed model significantly improves the accuracy and robustness of the crowd counting model.

**Keywords** Crowd counting, Convolutional neural network, Crowd density estimation, Crowd segmentation, Multi-task learning

## 1 引言

改革开放后,我国的城市化进程明显加快,人口从农村逐渐向城市聚集,各种人群活动也明显增加,比如外出旅行、开展演唱会、进行体育竞赛等。而对于这些活动的安全监控,正是社会治安的重要部分<sup>[1]</sup>,因此有必要对各种场景进行准确的人群计数。人群计数旨在根据视频帧或图像获取人群的总数和人群密度分布情况,此外鲁棒性强的人群计算方法可以推广至车辆计算<sup>[2]</sup>和微生物计数<sup>[3]</sup>等领域。

人群计数的传统方法主要是基于检测和回归的方法<sup>[4]</sup>。早期的人群计数方法一般参考行人检测算法,需要检测出一张图像上的所有行人,并以此统计出图像上的人群总数。首先手工提取图像中的相关特征,如 Haar wavelets<sup>[5-6]</sup>、方向梯度直方图特征<sup>[7]</sup>或边缘特征<sup>[8]</sup>。然后使用 SVM<sup>[9-10]</sup>,Boos-

ting<sup>[11]</sup>等方法对特征进行分析以识别个体目标。这种基于检测的方法容易受到背景噪声、背景遮挡、个体重叠等因素的严重影响,在人群密度高的区域检测效果表现得不够优异。为解决这类问题,Chan 等<sup>[12-13]</sup>提出了基于回归的人群计数方法。基于回归的人群计数方法又可以细分为回归到人群总数的方法<sup>[14]</sup>和回归到人群密度图的方法<sup>[15]</sup>。前者是通过学习得到输入图像与图像所含有目标总数的映射关系,具体做法是先对底层特征,如纹理特征<sup>[16]</sup>,局部特征和全局特征<sup>[12]</sup>进行提取,再应用线性回归<sup>[17]</sup>等算法进行模型回归;后者则是学习输入图像与密度之间的映射关系。Pham 等<sup>[18]</sup>应用随机森林算法来学习这种非线性映射关系,实现对密度图的估计。

随着硬件性能的提高,深度学习取得爆发性的进展,在图像分类、物体检测与识别、超分辨率重建等多个领域均相比传

统方法都有着显著的提升。Zhang 等<sup>[19]</sup>采用了 3 列全卷积神经网络模型(Multi-column Convolutional Neural Network, MCNN),每一列卷积层有着不同大小的卷积核,尺寸小的卷积核旨在提取较小人头的特征信息,尺寸大的卷积核旨在提取较大人头的特征信息。Sam 等<sup>[20]</sup>提出一种 Switch-CNN 卷积神经网络模型,先将输入图像进行预处理,分割成 9 个图像块,对每个图像先粗略估计密度级别,如高、中、低 3 个级别,再根据密度级别选用特定的分支网络进行更为精确的密度估计。文献[21]提出的 CSRNet 分为两个部分:前端网络和后端网络。前端网络使用 VGG-16 的前 10 层,后端网络使用空洞卷积网络,空洞卷积可以扩大感受野区域,生成较为准确的人群密度图。

本文分别对 MCNN 和 CSRNet 进行了改进,提出一种基于多任务学习的由粗到精的卷积神经网络模型,即模型一方

面输出人群分割图,另一方面则输出基于人群分割图得到的人群分布密度图,并在 Shanghai Tech 数据集和 UCF\_CC\_50 数据集上进行了测试,相比 MCNN 与 CSRNet,人群计数效果均有显著提升,并且优于目前大部分现有的人群计数方法。

## 2 卷积神经网络人群计数模型的构建

本文方法的基本思想是首先由骨干模型生成粗糙的人群密度图;其次对粗糙的密度图提取人群分割信息,作为新的任务来监督模型对特征的提取;最后融合粗糙密度图与分割图的特征,生成最终预测的人群分布密度图。

### 2.1 密度标签图

采用基于几何自适应高斯核<sup>[19]</sup>的方法,生成的密度标签图(ground truth)如图 1(b)所示。

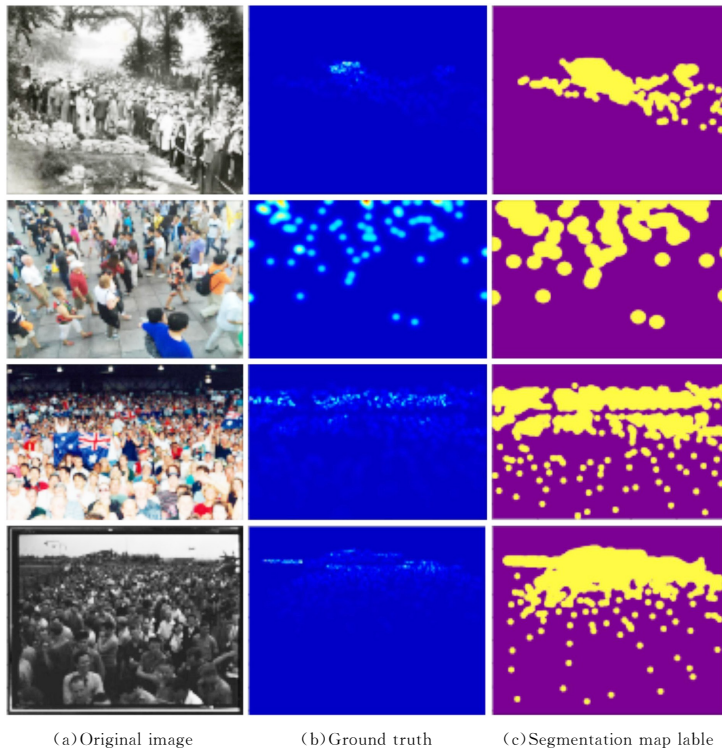


图 1 标签图展示

Fig. 1 Label image display

由于人群数据集通常提供人头中心位置的坐标点信息,直接使用这个坐标信息进行训练会导致人群总数估计效果差,因此通过高斯核对标记点进行模糊化处理,步骤如下:设一张人群图像的总人数为  $N$ ,其中第  $i$  个人头中心位置坐标为  $x_i$ ,用单位冲击函数  $\delta(x-x_i)$  表示,则这张人群图像可用式(1)表示,将  $H(x)$  与高斯卷积核相乘得到人群密度图  $D(x)$ ,如式(2)所示。

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

$$D(x) = \sum_{i=1}^N \delta(x-x_i) * G_{\sigma^i}(x), \sigma^i = \overline{\beta d_i} \quad (2)$$

$G(x)$  是一个二维的高斯核,  $\sigma^i$  是高斯核的标准差,根据文献[19]设定,  $\beta=0.3$ ,  $d_i$  由  $k$  最近邻算法给出,表示与  $x_i$  最接近的  $k$  个相邻点的平均距离,在本文中  $k$  设置为 3。

### 2.2 分割标签图

根据已有的标注信息进一步进行处理得到分割标签图(segmentation map label)。下文将借助人群分布密度图分析

这一过程。采用几何自适应高斯核算法,由原图人群图像生成得到人群分布密度图,如图 1(b)所示。再根据给定的阈值对此分布密度图进行二值化处理:如果分布密度图中的像素的值大于阈值,则映射为标签图上的像素 1;反之则映射为标签图上的像素 0,即可得到人群分割标签图。分割标签图显示人群分布位置,如图 1(c)所示。

准确来说,此处提及的人群分割与文献[22]所阐述的语义分割有所不同。文献[22]中的语义分割需详细地描绘每个物体(人)的准确轮廓,而实验部分采用分割标签图是为了显示人群主要分布的区域,并不需要准确地了解某一个人的轮廓信息。

实验结果也表明,即便是这样简单的策略也能够显著提高人群计数任务的准确性。

### 2.3 网络结构

基于多任务学习的卷积神经网络人群计数模型如图 2 所

示,主要由3部分构成:粗糙密度图生成模块、人群分割图生成模块和特征融合模块。正如前文所述,将最初骨干网络的单一任务(人群计数)拓展为多任务(人群计数与人群分割),并采取由粗到精的策略来提高人群计数的准确性。通常来说,人群密度估计是一个编码解码的过程,先由前端编码器将输入的人群图像映射到高维特征,再由后端解码器将高维特征解码为像素级别的密度图。本文采用的编码器与解

码器由骨干网络实现,将骨干网络作为粗糙人群密度图生成模块,并生成输入图像的粗糙人群密度图(coarse density map),再将粗糙人群密度图送入人群分割图生成模块,生成人群分割图(segmentation map)。最后通过融合粗糙人群密度图与人群分割图以得到更为准确而细致的人群密度图(density map)。再接下来的章节中,本文将逐一介绍模型的3个主要模块。

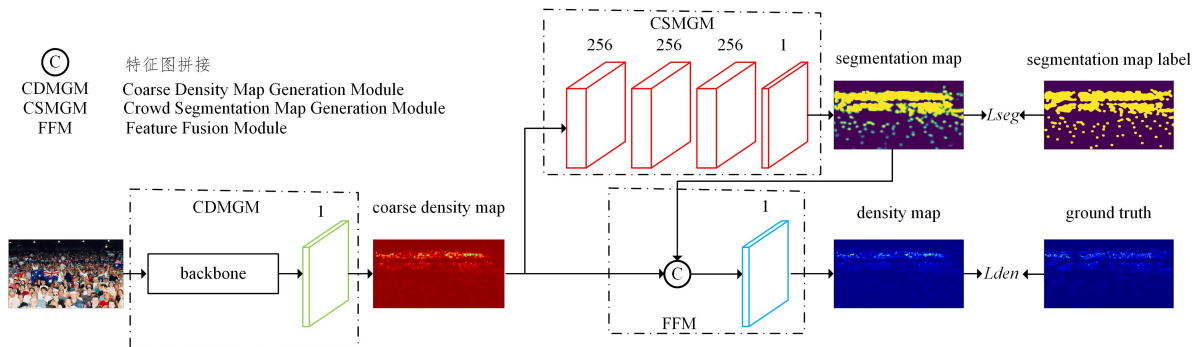


图2 基于多任务学习及由粗到精的人群计数模型结构

Fig. 2 Structure of crowd counting model based on multi-task learning and coarse to fine

### 2.3.1 粗糙人群密度图生成模块

为了验证基于多任务的回归思想的正确性,本文采用了两个骨干网络:MCNN和CSRNet,由骨干网络配上 $1 \times 1$ 的卷积层作为粗糙人群密度图生成模块(Coarse Density Map Generation Module, CDMGM),即用于生成粗糙的人群密度图。MCNN采用3列体系结构的网络模型,每一列卷积神经网络通过不同大小的卷积核,可以适应不同大小的人头,最后将3列卷积神经网络的输出进行融合,得到最终的人群密度图。人群密度图为输入图像的 $1/4$ 。CSRNet网络模型的主要组成部分包括前端网络与后端网络。前端网络采取VGG16的前13层,后端网络采取空洞卷积网络以提取非相邻像素点的相互关系,最终的人群密度图为输入图像的 $1/8$ 。骨干网络信息可详见文献[19,21]。

### 2.3.2 人群分割图生成模块

设计人群分割图生成模块(Crowd Segmentation Map Generation Module, CSMGM),该模块主要用于生成人群分割图,可以提供输入图像中人群的主要分布位置信息。本文针对不同的骨干网络设计略有不同的分割图生成模块。对于MCNN网络模型,该模块主要由4层卷积层构成,前3个卷积层输出通道数均为64,最后一个卷积层输出通道数为1,以生成人群分割图。此外,第一个卷积层的卷积核大小为 $3 \times 3$ ,后3个卷积层的卷积核大小为 $1 \times 1$ 。该模块以粗糙密度图作为输入,模块的输出经过sigmoid函数处理后,得到最终的人群分割密度图。对于CSRNet网络模型的分割图生成模块,仅仅将前3个卷积层的输出通道数改成256,其余均不变。这样做的好处在于为不同网络模型设计分割图模块时,仅需根据需要修改的网络模型结构大小,在通道数细节处修改即可。模块结构如表1所列,其中卷积层参数由C(卷积核大小-输出通道数-步长-边缘填充大小)表示。

表1 人群分割图生成模块

Table 1 Crowd segmentation map generation module

Backbone	MCNN	CSRNet
CSMGM	C(3-64-1-1)	C(3-256-1-1)
	C(1-64-1-0)	C(1-256-1-0)
	C(1-64-1-0)	C(1-256-1-0)
	C(1-1-1-0)	C(1-1-1-0)

### 2.3.3 特征融合模块

为了将粗糙密度图和人群分割图进行特征融合,我们设计了一个简单的特征融合模块(Feature Fusion Module, FFM),采取将两者拼接成两通道的形式再作为输入信息输入模块最后一层卷积层,如图2所示。相比直接将人群分割图与粗糙密度图进行像素级别相乘的方法,这种通道拼接的方法可以更加灵活地融合两者的信息,即由卷积核大小为 $1 \times 1$ ,输出通道数为1的卷积层来动态学习融合两者信息。

## 2.4 由粗到精

2.3.1节详细介绍了如何生成粗糙人群密度图。需要注意的是,这个生成过程并没有监督信号监督密度图的生成,因此,若对该粗糙的密度图做积分操作得出人数,这个数值其实是不准确的。粗糙的密度图虽然可以反映一定的人群分布信息,但无法得到合理的人群总数。为了得到合理的预测人数,本文提出一个由粗到精生成密度图的方法,如图2所示。将粗糙的密度图送入人群分割图生成模块,提取出人群分割图信息,人群分割图反映图像中人群分布的大致信息。紧接着通过特征融合模块将粗糙的密度图和人群分割图进行通道拼接,并经过卷积核大小为 $1 \times 1$ 的卷积层得到最终较为精确的密度图,然后对密度图进行积分操作,则可得到人群总数。

## 3 实验

为了验证人群计数模型的有效性,本节将本文改进方法与其他相关算法在两个数据集(Shanghai Tech, UCF\_CC\_50)上进行对比。实验过程统一采用CSRNet<sup>[21]</sup>中的数据增强方

式,从原图中裁出9个大小为原图1/4的图像块。前4个图像块为原图的4个互不重叠的1/4的图像块,其余5个随机从原图上裁剪。数据集的详细信息如表2所列。

表2 标准数据集的数据信息

数据集	Shanghai Tech		UCF_CC_50
	Part A	Part B	
总图数	482	716	50
尺寸	Different	768 * 1024	Different
最大人数	3139	578	4543
最小人数	33	9	94
平均人数	501.4	123.6	1297.5
总人数	241677	88488	63974

### 3.1 实验环境

实验基于联想集群 LT-5038A 工作站,操作系统为 Ubuntu16.04,64 位操作系统,选择深度学习框架 PyTorch 0.4.0,采用运算显卡 NVIDIA-1080Ti(12G 显存),进行训练和测试。

### 3.2 评价指标

根据已有的研究工作<sup>[19,21]</sup>,将平均绝对误差(MAE)与均方误差(MSE)作为评估人群计数方法的评价指标,其定义如式(3)、式(4)所示。

$$MAE = \frac{1}{N} \sum_{i=1}^N |R_i - R_i^T| \quad (3)$$

$$MSE = \sqrt{\frac{1}{N} \sum_{i=1}^N |R_i - R_i^T|^2} \quad (4)$$

其中, $N$ 代表测试集中图像的总数量, $R_i$ 表示由预测密度图积分得到的估计人数, $R_i^T$ 表示图像上实际的人数。MAE 值越小,说明该方法的准确性较高;MSE 值越小,则说明该方法的鲁棒性越好。

### 3.3 损失函数与优化

参考已有的方法<sup>[19,21]</sup>,使用 L2 范数的平方来计算预测密度图与真实密度图之间的差异,损失函数如式(5)所示。

$$L_{den} = \frac{1}{2N} \sum_{i=1}^N \|F_d(X_i; \theta) - F_i^{GT}\|_2^2 \quad (5)$$

另外对于人群分割这一任务,我们采用交叉熵作为分割图损失函数  $L_{seg}$ ,如式(6)所示。

$$L_{seg} = -\frac{1}{N} \sum_{i=1}^N \|F_s(X_i; \theta) \log(F_i^{seg}) + (1 - F_s(X_i; \theta)) \log(1 - F_i^{seg})\|_1 \quad (6)$$

在网络训练过程中,我们将两个任务的损失函数融合成一个损失函数,定义如式(7)所示:

$$Loss = L_{den} + \lambda L_{seg} \quad (7)$$

$\lambda$  作为超参数用于平衡两个  $loss$ ,在本文实验中,将其值设为 0.008。

### 3.4 实验过程

数据处理:从 Shanghai Tech 数据集 Part A 中选取 300 张图像作为训练集,从 Part B 中选取 400 张图像作为训练集。UCF\_CC\_50 数据集包含 50 张图像,故采用 5 折交叉验证法,每次将 40 张图像用于训练,将 10 张图像用于测试。使用几何自适应高斯核<sup>[19]</sup>的方法生成真值密度图,再对其进行二值化,得到真实分割图。采用 CSRNet<sup>[21]</sup>中的数据增强方式以扩充训练样本。此外对数据集中所有样本进行归一化处理,即每个 RGB 通道减去对应的均值,再除以相应的方差,保证每个像素点的值都在  $(-1,1)$  区间内。

参数设置:实验设置数据集遍历次数为 200,每个批量训

练只包含一个样本。训练过程采用 Adam 优化器<sup>[23]</sup>,以  $1 \times 10^{-3}$  作为初始学习率进行训练,动量因子设为 0.95,权重衰减设为  $5 \times 10^{-4}$ 。随着训练的进行,当损失曲线趋于稳定,则将学习率乘以系数 0.3。

网络结构:人群计数模型结构如图 2 所示。在粗糙人群密度图生成模块中采用预训练的 VGG16,对于其他模块的卷积层均采用均值为 0,标准差为 0.001 的正态分布进行初始化。除了密度回归层和分割回归层以外,每个卷积层之后都会加上 Rule 层,以防止出现过拟合,并保证输出的特征图非负。

## 3.5 实验结果分析

### 3.5.1 Shanghai Tech 数据集

Shanghai Tech 数据集共包含 1198 个带有标签的图像,总共标记人数为 330165。其中 Part A 部分包含 482 张人群分布较为密集的人群图像,Part B 部分包含 716 张人群分布较为稀疏的人群图像。表 3 列出了本文方法与其他方法在 Shanghai Tech 数据集上结果的对比,并对 MAE 与 MSE 两个方面进行了比较。在表格中带“++”标志的算法为本文提出的改进算法。

表3 不同算法在 Shanghai Tech 数据集上的表现

Table 3 Performance of different algorithms on Shanghai Tech dataset

Algorithm	Part A		Part B	
	MAE	MSE	MAE	MSE
Zhang <sup>[24]</sup>	181.8	277.7	32.0	49.8
MCNN <sup>[19]</sup>	110.2	173.2	26.4	41.3
Switch-CNN <sup>[20]</sup>	90.4	135.0	21.6	33.4
CSRNet <sup>[21]</sup>	68.2	115.0	10.6	16.0
SANet <sup>[25]</sup>	67.0	104.5	<b>8.4</b>	<b>13.6</b>
MCNN++	88.4	145.8	25.9	46.2
CSRNet++	<b>65.1</b>	<b>103.5</b>	9.1	14.3

由表 3 可知,改进方法使得 MCNN 算法在 Shanghai Tech 数据集 Part A 上的 MAE 降低了 19.78%,MSE 降低 15.82%;在 Shanghai Tech 数据集 Part B 上的 MAE 降低了 1.89%,但 MSE 没有降低。同时,使得 CSRNet 算法在 Shanghai Tech 数据集 Part A 上的 MAE 降低了 4.55%,MSE 降低了 10.00%;在 Shanghai Tech 数据集 Part B 上的 MAE 降低了 14.15%,MSE 降低了 10.63%。

由此数值可以看出,在骨干网络的基础上结合本文提出的改进方法,人群计数模型预测人数的准确率都一定的提高,说明了多任务学习及由粗到精的设计思路具有一定的有效性。在 Shanghai Tech 数据集 Part B 上的 MSE 没有降低,表明对 MCNN 算法进行改进时,准确率提升了但鲁棒性有所损失,故针对这类平均人数较少的数据集,未来可以继续开展研究,争取做到在不牺牲鲁棒性的同时提高准确性。

### 3.5.2 UCF\_CC\_50 数据集

在人群计数领域中,UCF\_CC\_50 数据集是一个极具挑战性的数据集,仅仅包含来自互联网的 50 张图像。这个数据集不仅数量十分有限,而且图像的人群数据分布极为分散,人群范围在 94~4543 之间。针对这个数据集,本文采用 5 折交叉验证法进行实验,将 50 张图像分割成 5 个子样本,保留其中一个子样本作为验证模型的数据,其他 4 个样本用于训练模型,重复完成此过程 5 次即可。表 4 列出了本文方法与其他方法在 UCF\_CC\_50 数据集上结果的对比,带“++”的算法为本文改进而来的算法。

表4 不同算法在UCF\_CC\_50数据集上的表现

Table 4 Performance of different algorithms on UCF\_CC\_50 dataset

Algorithm	MAE	MSE
Zhang <sup>[24]</sup>	467.0	498.0
MCNN <sup>[19]</sup>	377.6	509.1
Switch-CNN <sup>[20]</sup>	318.1	439.2
CSRNet <sup>[21]</sup>	266.1	397.5
SANet <sup>[25]</sup>	258.4	334.9
MCNN++	336.9	454.2
CSRNet++	215.3	320.1

由表4可知,改进方法使得MCNN算法在UCF\_CC\_50数据集上的MAE降低了10.78%,MSE降低了10.78%。同时,也使得CSRNet算法在UCF\_CC\_50数据集上的MAE降低了19.09%,MSE降低了19.47%。由此可以看出,本文改进的人群计数模型在UCF\_CC\_50这一具有挑战性的数据集上,算法的准确率和鲁棒性全都得以提升。

### 3.5.3 有效性验证实验

本节将对CSRNet模型在Shanghai Tech数据集Part A上进行实验,探究多任务学习及由粗到精两项工作的有效性,分析其对骨干网络的影响。

如表5所列,CSRNet表示本文采用的骨干网络,未做任何改进。CSRNet+表示骨干网络只增加了人群分割图生成模块,CSRNet++为最终的实验模型,不仅增加了人群分割图生成模块,还通过特征融合模块对前两个模块的输出进行融合。观察表5可以发现,MAE与MSE均呈现下降趋势,可见这两项工作在一定程度上能够提高人群计数模型的性能。

表5 有效性验证

Table 5 Verification of effectiveness

Algorithm	CSMG	FFM	Part A		Part B	
			MAE	MSE	MAE	MSE
CSRNet <sup>[21]</sup>			68.2	115.0	10.6	16.0
CSRNet+	✓		66.9	110.8	9.5	15.2
CSRNet++	✓	✓	65.1	103.5	9.1	14.3

### 3.5.4 人群密度图可视化

为了直观地展示人群计数模型对人群进行密度估计以及人群分割的性能,本节将展示来自Shanghai Tech数据集和UCF\_CC\_50数据集中4张示例测试图像的预测密度图以及人群分割图,如图3所示。在图3中,我们分别展示了测试图像、密度标签图、预测密度图和预测分割图。实际人数与预测人数分别显示在密度标签图和预测密度图的右下角。

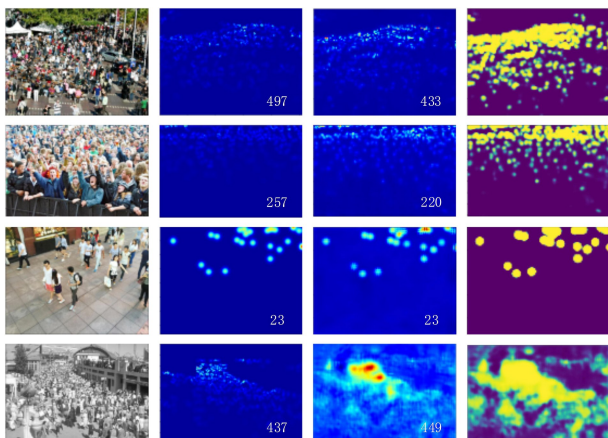


图3 实验效果展示

Fig. 3 Experimental effect display

从图3可以看出,在不同类型的数据集上,依然可以取得较为精确的效果。前两张图像来自Shanghai Tech数据集Part A部分,图像包含人数较多;第三张图像来自Shanghai Tech数据集Part B部分,图像包含人数较少;最后一张测试图像来自UCF\_CC\_50,这个数据集中各图像包含的人数普遍极多。将第一张输入与第四张输入进行分析,发现图像中若出现树叶等,在一定程度上可能会影响对人数的预测。再观察第三张图像可以发现,人数较少时,可以比较精准地定位图像中每个人的位置,预测人数也比较精确。再分析测试图像、密度标签图以及预测分割图,可以认为分割模块输出的人群分割图指示着人群所在的大致位置。

**结束语** 为了提高人群计数任务的准确性,本文提出基于多任务学习及由粗到精的卷积神经网络人群计数模型。首先,根据骨干网络生成粗糙的密度图,再将其输入分割模块生成人群分割图,最后融合粗糙的密度图与人群分割图以得到细致的密度图。此外,通过我们设计的方法,本文不需要人工表标注即可得到分割标签图,并构造新的损失函数来实现多任务学习。在Shanghai Tech数据集和UCF\_CC\_50数据集上进行实验,实验数据显示,本文提出的算法可以大幅提高人群计数的准确性,相比其他的人群计数算法具有更好的性能,但在平均人数较少的数据集上的表现有待提高。未来考虑结合残差网络设计思想提出新的具有强鲁棒性的人群计数模型,以进一步提高人群计数的准确性。

## 参考文献

- [1] FU H, MA H, XIAO H. Scene adaptive accurate and fast vertical crowd counting via joint using depth and color information [J]. Multimedia Tools and Applications, 2014, 73(1): 273-289.
- [2] WEI WU, ZHANG Q S, WANG M J, et al. Detection of traffic parameters based on computer vision and image processing [J]. Information and Control, 2001, 30(3): 257-261.
- [3] FRENCH G, FISHER M, MACKIEWICZ M, et al. Convolutional neural networks for counting fish in fisheries surveillance video [C]// British Machine Vision Conference, 2015: 23-32.
- [4] RYAN D, DENMON S, SRIDHARAN S, et al. An evaluation of crowd counting methods, features and regression models [J]. Computer Vision and Image Understanding, 2015, 130: 1-17.
- [5] VIOLA P, JONES M J. Robust Real time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [6] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2005: 886-893.
- [7] HAAR A. Zur Theorie der orthogonalen Funktionen systeme [J]. Mathematische Annalen, 1911, 71(1): 38-53.
- [8] WU B, 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, 2005 (ICCV 2005). IEEE, 2005: 90-97.
- [9] HEARTS M A, DUMAIS S T, OSMAN E, et al. Support vector machines [J]. IEEE Intelligent Systems, 1998, 13(4): 18-28.