激光与光电子学进展
先进成像

基于计算机视觉的目标计数方法综述

蒋妮，周海洋，余飞鸿*
浙江大学光电科学与工程学院，浙江 杭州 310027

摘要 目标计数作为一项基础的技术，在许多领域都有广泛的应用，如人群计数、细胞计数、车辆计数等。随着互联网时代的信息爆炸，视频数据呈指数级增长，如何快速、准确地获得目标的数量是用户普遍关心的主要问题之一。得益于计算机视觉技术的快速发展，基于传统机器学习的计数方法正逐步向基于深度学习的方法转变，并在计数的准确性上取得了实质性的进展。介绍了目标计数的研究背景和应用领域，根据模型任务分类，归纳了三类常用的计数模型框架，并从不同的角度分别介绍了近 10 年来基于计算机视觉技术的模型方法。然后介绍了在人群计数、细胞计数和车辆计数领域中常用的几种公开数据集，并横向比较了各个模型之间的性能。最后总结了现阶段的目标计数模型还存在的不足，并对未来的研究方向进行了展望。

关键词 图像处理；目标计数；神经网络；机器学习；密度图
中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202158.1400002

Review of Computer Vision Based Object Counting Methods

Jiang Ni, Zhou Haiyang, Yu Feihong*
College of Optical Science & Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, *China*

Abstract As a fundamental technique, object counting has broad applications, such as crowd counting, cell counting, and vehicle counting. With the information explosion in the internet era, video data has been growing exponentially. How to obtain the number of objects efficiently and accurately is one of the problems that most users care about. By virtue of the great development of computer vision, the counting methods are gradually turned from the traditional machine learning based methods to deep learning based methods, and the accuracy has been improved substantially. First, this paper introduces the background and applications of object counting. Then according to the model task classification, three counting model frameworks are summarized and the computer vision based counting methods in the recent 10 years are introduced from different aspects. Some public datasets in the fields of crowd counting, cell counting, and vehicle counting are introduced and the performance of various models is compared horizontally. Finally, the challenges to be solved and the prospects for future research are summarized.

Key words image processing; object counting; neural network; machine learning; density map
OCIS codes 100.4996; 100.2960; 150.1135

1 引 言

目标计数任务是对给定的图像/视频进行分析，从而估计出图像/视频中的目标数量。在许多图像/视频语义分析中，目标计数任务作为其中的一个子任务，扮演着十分重要的角色。通过目标数量的空间分布信息，可以推测场景中关注度高的区域，进而对图像/视频进行合理的分析和解读。如在商场营销中，掌握人群流动情况可以更好地分析消费者的喜好，使利益最大化；在大范围的游行、聚会活动中，了解人群的分布情况可以及时、有效地疏散人群，防止踩踏事故的发生；在交通管理系统中，掌握实时路

收稿日期：2020-10-10；**修回日期**：2020-11-06；**录用日期**：2020-12-03
通信作者：*feihong@zju.edu.com

况能快速、合理地调配交通资源，及时引导和疏散车流量；在临床医学中，细胞的密集分布有可能是细胞恶性增殖的表现，准确的细胞数量分布有助于医生快速地诊断病人的健康状况和确定病灶；在野外活动中，对野生动物进行监控，了解它们的活动轨迹和生活习惯，有助于专家进行相关课题的研究。传统的计数方法主要依赖于人工，但在长时间的工作情况下，人工计数的结果受主观因素的影响较大。为了解决这个问题，研究人员结合计算机视觉技术的优势，提出了一系列的计数模型，逐渐将目标计数任务由手动变为半自动，再到如今的全自动。

人群计数一直以来都是目标计数中的热点，近些年也不断地有文献对人群计数模型的研究现状进行总结和概括。Zhan 等[1]从不同研究领域的角度对人群分析技术进行详细的阐述。Saleh 等[2]总结了人群密度估计和计数的方法，包含直接法（基于模型的方法、基于轨迹聚类的方法）和间接法（基于像素的方法、基于纹理的方法、基于角点的方法）。张君军等[3]分别从浅层学习和深层学习两个层面对人群统计和人群密度估计技术进行了详细的介绍。Sindagi 等[4]将常用的计数方法分为了三类：基于检测的方法、基于回归的方法和基于密度估计的方法，并分别对其进行了简单的总结和概括，重点介绍了若干种基于卷积神经网络（CNNs）的方法，比较了这些方法在特定数据集上的表现。Gao 等[5]研究分析了 220 多项工作，以时间轴为线索，详细、系统地介绍了基于卷积神经网络的人群密度估计方法的发展历程。尽管这些文献对人群计数方法的发展历史和现状进行了比较全面的总结和概括，但是缺少对其他类型目标的计数方法的总结。为弥补这一方面的缺失，考虑到不同目标类型之间计数任务的相似性，本文将在总结人群计数模型的基础上，对细胞计数和车辆计数的研究现状及进展进行补充。

根据目标计数模型的任务属性，本文将现有的模型归纳为以下三类：基于回归的计数模型、基于密度图估计的计数模型和多任务模型（即模型有多种输出），对这三类模型（涵盖了人群、细胞、车辆多个应用场景）进行系统、全面的介绍。然后介绍人群计数、细胞计数和车辆计数中常用的几种公开数据集和模型性能的评估指标，对多种模型进行比较，并总结了当前流行的计数方法中面临的挑战和待解决的问题，提出未来可能的发展方向。

2　目标计数模型综述

本节对三类输出类型的模型进行介绍，图 1 是三类模型的示意图。图 1（a）表示模型通过学习输入图像直接输出待计数目标的数量。图 1（b）输出的是一张密度图，其中每一个元素的值反映的是当前位置上所包含目标的数量，同时密度图整体上反映了目标的空间分布。对于输出为密度图的模型，通过对密度图中任意区域内的元素进行积分，可得到该区域内的目标数量。图 1（c）表示模型有多个任务，以密度图估计任务为主，其他任务[如目标数量估计、密度等级（如高密度、中高密度、中密度、中低密度、低密度等）划分、图像分割等任务]为辅来帮助提高计数任务的准确性。

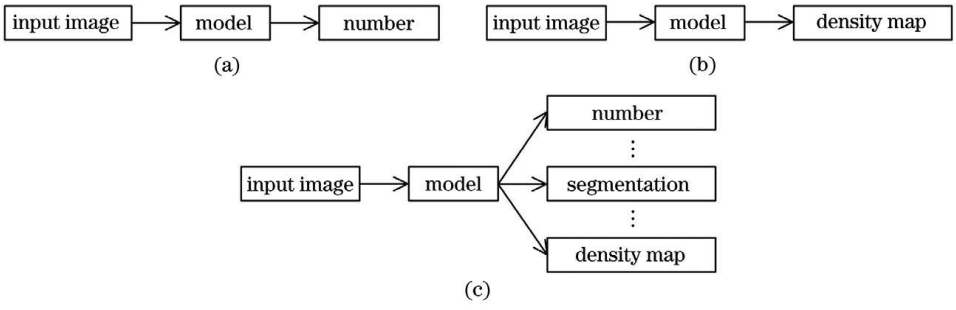

图 1　三种类型的模型示意图。（a）基于回归的目标计数模型；（b）基于密度估计的目标计数模型；（c）多任务模型

Fig. 1　Schematic diagrams of three models. (a) Regression based object counting model; (b) density estimation based object counting model; (c) multi-task model

2.1　基于回归的目标计数模型

Shang 等[6]使用 GoogLeNet[7]作为特征提取器，将生成的特征图输入到长短时记忆网络（LSTM）进行解码，通过卷积层和全连接层分别获得输入图像的局部和全局计数结果。神经网络联合优化局部损失和全局损失，降低了整体的计数误差。CNN Boosting[8]使用级联 CNN 来增强特征的表达，除第一个 CNN 以外，后面的每一个 CNN 都学

习前一个 CNN 的预测结果和真实值之间的残差，整个集成模型不断地逼近残差，使得预测值尽可能地接近真实值。此外该文献还提出了一种样本选择的方法，作者认为那些计数误差大的样本对模型的训练更有帮助，因此周期性地调整训练数据中具有不同计数误差的样本比例，不仅能提高模型的性能，还能加快模型的收敛。为了适应不同场景之间的差异变化，Marsden 等[9]通过多条网络支路分别对人群计数、车辆计数、细胞计数等多个场景进行判断，并自动选择合适的场景进行推断。图 2 展示了网络自动选择场景的示意图。

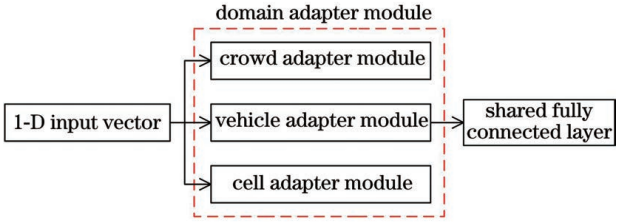

图 2　多场景判断的网络结构

Fig. 2　Architecture of multi-scene judgment

以上这些方法只能提供场景内目标的数量信息，但往往目标数量在空间上的分布情况能提供更复杂的信息。因此，越来越多的研究倾向于通过估计密度图的方式来间接获得目标的数量，或者将目标数量的估计作为一项辅助任务来更好地推测密度图。与此相反，Zhang 等[10]将估计的密度图作为中间结果来辅助模型获得最终的计数结果，实现了从输入图像到计数结果的端到端训练模式。模型（FCN-rLSTM）的网络结构如图 3 所示。首先全卷积网路（FCN）负责输出密度图，然后以残差连接的形式引入 LSTM，利用视频在时间序列上的连续性对最终的计数结果进行进一步的优化。相比以往的一些计数模型[11-13]，FCN-rLSTM 在人群计数和车辆计数方面的表现均有所提升。

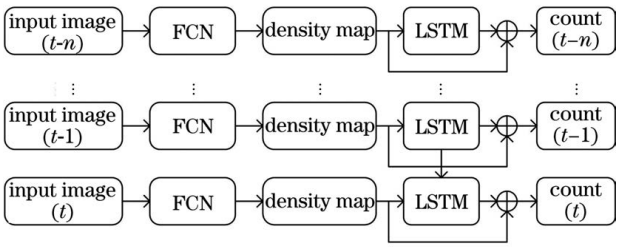

图 3　FCN-rLSTM 网络结构

Fig. 3　Architecture of FCN-rLSTM

2.2　基于密度图估计的目标计数模型

在总结基于密度图估计的目标计数模型之前，本文先回顾一下密度图真实值的生成过程。给定一张包含目标的图像，对每个目标都用一个靠近目标（如人头、细胞等）中心的点进行标注，记为 $I(x)$，然后采用归一化高斯函数 $G(\mu,\sigma)$ 对其进行卷积，重叠区域输出的密度图是每个目标经卷积后的叠加结果，可表示为

$$F(x)=\sum_{p\in P}I(p)\otimes G(\mu,\sigma),\tag{1}$$

式中：P 表示标注点的坐标集合；$G(\mu,\sigma)$ 表示均值为 μ、方差为 σ 的高斯函数。

密度图的生成过程如图 4 所示。图 4(a)是输入的原图像数据，图 4(b)中的 $I(x)$ 是对应的点标注信息，$F(x)$ 则是经过高斯核 $G(\mu,\sigma)$ 卷积之后得到的密度图。

图 4　输入图像和密度图生成。(a)输入图像；(b)密度图生成

Fig. 4　Input image and generation of density map. (a) Input image; (b) generation of density map

基于传统图像处理的目标计数方法所面临的最棘手的两个问题就是复杂背景的干扰和目标重叠。尽管已有一些文献提出相应的方法[14-16]来克服这类问题，但这些方法的鲁棒性不够高，限制了模型的推广使用。自从 Lempitsky 等[17]创造性地提出了采用密度图来模拟目标的空间密度分布的方法后，大量基于密度图估计的方法[12-13, 18-19]相继被提出，并取得了显著的进步。

为了在避免过拟合的同时保证局部区域计数的准确度，Lempitsky 等[17]定义了一个距离函数来度量真实值和预测值之间的差异。该方法不仅在人群计数上获得了成功，也同样适用于细胞计数和车辆计数，这一工作为后来目标计数领域的发展指明了方向。在 Lempitsky[17]工作的基础上，Ma 等[20]提出了对估计的密度图添加整数规划求解的方法，在计数的同时实现了对小目标的定位。Fiaschi 等[21]

通过随机森林逐块学习输入图像，并对重叠的输出区域取平均，实现了一种简洁、高效的计数模型，该方法已成功应用于一款专业计数软件[22]。生成密度图的前提是需要对目标样本逐个进行标记，这是一项十分重要而又繁琐的任务。为了减轻标注的工作量，Borstel 等[23]基于贝叶斯模型提出了一种只需要对目标区域子集内的目标数量进行标记的弱标记训练方法。

Hydra CNN[11]和 MCNN[19]是两种十分经典的目标计数模型。两者都是通过多列网络分支来学习多尺度目标，不同的是 Hydra CNN 是对输入图像的不同尺度进行处理，而 MCNN 采用不同大小的卷积核来获取不同大小的感受野。图 5 展示了 Hydra CNN 的结构示意图，将经过不同尺度缩放后的目标分别输入到网络中，使网络能同时学习多种尺度的变化。MCNN 将图像输入到三列具有不同卷积核大小的并行网络中，每条网络支路中具有不同大小的感受野，最后将这些携带多尺度信息的特征图进行合并。其网络结构如图 6 所示。此外，为了能让生成的真实密度图和目标空间分布更匹配，MCNN 还根据目标之间的平均距离来确定高斯函数的参数，但使用该方法的前提是目标分布密集。

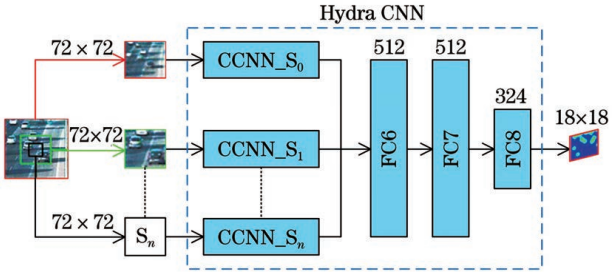

图 5　Hydra CNN 结构

Fig. 5　Architecture of Hydra CNN

在人群计数和车辆计数问题中，由于透视效应的存在，不同位置的目标存在显著的尺度差异。为了更进一步地学习这种由透视效应引起的尺度差异性，Deb 等[24]采用了更宽的网络结构，且每一路网络的感受野大小逐渐递增。以上模型均是在网络的头部采用多分支的结构来获得多尺度信息，而 Wang 等[25]则是在网络的末端通过多个并行的空洞卷积[26]来提取多尺度信息，这种结构参数量更小，网络更容易训练。但 Li 等[27]认为这种多分支网络结构的设计使网络变得冗余，且训练过程变得更加复杂，因而在网络前端采用 VGG16[28]预训练模型提取特征，在网络后端使用空洞卷积[26]来获得更大的感受野。

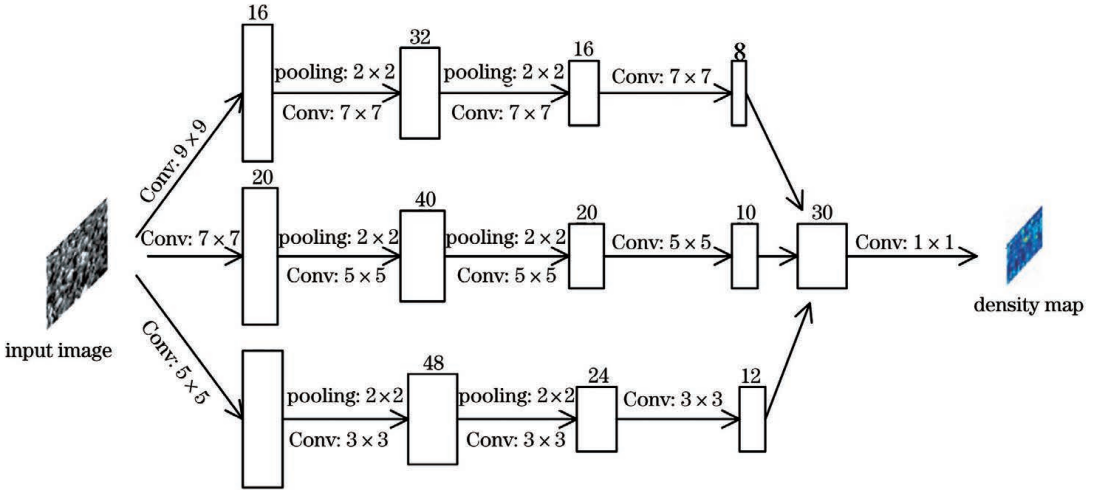

图 6　MCNN 结构

Fig. 6　Architecture of MCNN

Marsden 等[29]在全卷积神经网络(FCN)上连续采用大尺度的卷积核来获得更大的感受野，网络通过两次下采样操作来减少计算量，最终获得的密度图大小是原图的 1/16。在生成密度图时，参考了 MCNN[19]的做法，每个目标的高斯核参数与当前目标到邻域内其他目标的平均距离有关。但当目标分布比较稀疏时，利用目标之间的平均距离来度量估计人头大小的方法便不再适用了。为了能更好地根据目标大小自适应地改变局部区域内密度图的生成参数，A-CCNN[30]通过检测目标来估计密度图的超参。由于不同区域内目标的尺度差异比较大，所以需要对图像分块进行训练和估计，最后再对所有图像块的预测结果进行拼接，得到最终的估计密度图。

和 Fiaschi 早期的工作[21]类似，Xie 等[31]也是对重复预测的密度值取平均，然后通过确定目标中

心点的方式来计算目标个数。Xue 等[32]首次将深度残差网络[33]的应用从分类、检测、分割等领域拓展到目标计数领域。和直接估计密度图不同，以滑动窗口的形式逐块对目标进行计数，在获得所有图像块的计数结果之后，通过双线性插值对计数结果进行放大，并以热力图的形式反映目标的空间分布。为了提高视频中目标计数的实时性，刘旭[34]提出一种直接在视频压缩域中进行目标计数的模型 HCR，并针对压缩域相比像素域信息量不足的缺点，结合时间维度和空间维度对信息进行约束，提高在压缩域中计数的准确度，同时还针对单一图片的目标计数问题提出了一种金字塔目标计数网络 POCNet。通过分析在低密度人群和高密度人群数据上的对比实验结果，Liu 等[35]发现基于检测的计数方法在低密度人群上表现更好，相反地，基于密度图估计的计数方法在高密度人群上表现更好。于是，在此分析的基础上提出了 DecideNet，其网络结构如图 7 所示。图 7 中上面一条网络支路负责估计密度图，下面一条网络支路负责检测稀疏人群目标，然后再将检测的结果转换为密度图的形式，中间的网络则是对来自上下两条网络支路的密度图进行权值加和，输出最后得到的密度图。大多数模型在处理多尺度问题时常采用的方法就是用多列具有不同卷积核大小的网络来捕获不同大小的感受野，如 MCNN[19]。而 Chen 等[36]则认为这种方法结构复杂，计算资源消耗大，因此在 MCNN 的基础上进行了两点改进以减少计算量：1）由于每一列提取的底层特征都是相似的，因此只对深层特征在多条网络支路上进行处理；2）在多条并行的网络支路上用空洞卷积[26]代替常规卷积，形成空间金字塔的结构。

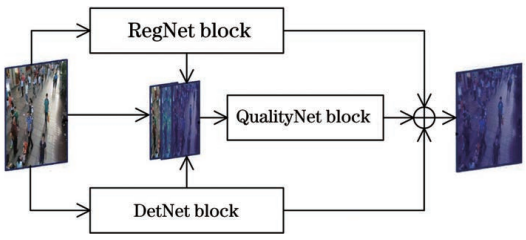

图 7　DecideNet 结构

Fig. 7　Architecture of DecideNet

RD-UNet[37]是一种专门用于人类胚胎细胞的计数网络，该网络结合了 U-Net[38]、空洞卷积[26]和残差网络[39]的优势，能较好地对大尺度、重叠的胚胎细胞进行检测。同样是用于细胞计数，考虑到细胞个体的尺寸相比整张图像小很多，深层次的网络结构就非必要了，因此 Xie 等[18]侧重于提出一个极其简洁的网络（甚至不包含跨层连接），该方法不仅在仿真图像上取得了较好的计数结果，在真实图像上也获得了较高的准确度。Count-ception[40]的网络结构由 Inception v2[41]改进而来，同样是对输入图像逐块进行估计并通过热力图对细胞计数结果进行可视化，相比于 Xue 等的工作[32]，Count-ception 网络具有更高的准确度。Rad 等[42]提出了带有残差连接的空间金字塔结构和渐进式上采样网络的计数模型 Cell-Net，能够更好地提取全局信息和重建高分辨率特征图，还设计了一个和图像纹理有关的损失函数，来克服样本不平衡带来的问题。

近年来，注意力机制在计算机视觉领域引起了研究者们的广泛关注，而这一项技术也正逐渐被应用在目标计数领域中。在 MCNN[19]的基础上，AM-CNN[43]在网络末端添加了一层注意力掩模来强化密度图中和目标相关的区域，以弱化背景噪声。AM-CNN 只是通过一个掩模对特征图的空间信息进行了过滤，而 SCAR[44]则同时考虑了空间维度上各个像素之间的相互关系和通道维度上各个通道之间的关系来实现空间注意力和通道注意力，这种形式更能加强特征的表达。在细胞计数领域，SAU-Net[45]首次将空间注意力模块应用在细胞计数上，并提出了一种在线的批量归一化方法来提升网络性能。不同于以往注重度量不同位置之间像素关系的注意力机制[44-45]，Hossain 等[46]通过注意力来关注图像内和图像间的目标密度变化。全局注意力关注的是图像间目标的差异性，局部注意力则关注同一图像内目标的差异性。注意力机制的提出是为了考虑两两像素之间的相互影响，避免估计结果中出现噪声。一般图像内的所有像素都参与了空间注意力的计算，而 RANet[47]对局部区域和全局区域都分别应用了空间注意力模块，并探讨了在局部注意力和全局注意力的结果中像素之间的关联性。Jiang 等[48]为了解决人群分布不均匀和尺度差异带来的问题，提出了一种能对密度值自适应缩放的计数网络 ASNet，通过对注意力掩模和缩放因子的共同约束提高网络的计数性能。Wang 等[49]分别用两个网络分支来估计不同尺度大小的目标，并各自生成注意力掩模以对密度图进行调整，最终将这两种关注不同尺度的密度图进行融合，得到最终的密度图估计结果。

为了使计数模型能够满足实际的应用需求，Liu 等[50]首次提出将现有的人群计数网络所学到的知识迁移到轻量化模型上，构成复杂度低的高效计数

模型。目前有关计数模型的轻量化网络研究还比较少，但随着人们对计数模型的研究深入和实际应用需求的提高，模型轻量化势必是将来的一个研究热点。

2.3　多任务模型

除了上述两类具有单一训练任务的模型之外，研究者们也尝试通过多任务的训练方式来提升模型性能。

Zhang 等[13]提出了一种具有人数估计和密度估计两种任务的网络结构 Cross-scene，通过交替优化这两种任务的目标函数，使网络整体获得更好的局部最优解，并通过选择和测试场景类似的训练样本对网络进行微调，使网络能适用于不同的场景。此外，该文献还公开了一个新的行人数据集 WorldExpo'10。FF-CNN[51]参考 MCNN[19]的方法生成对应的密度图，并给网络设置了两个损失函数——密度图有关的损失函数和目标数量有关的损失函数，联合优化这两个损失函数有助于提高网络的性能。为了克服同一张图像内目标密度变化带来的问题，MMCNN[52]首先将输入图像均等地划分成若干个块，并同时对密度图、前景/背景分割图和密度等级这三个任务进行训练，测试阶段只选择密度图作为网络输出。这种以密度估计为主、其他任务为辅的训练方式有助于网络更好地表达特征。Jiang 等[53]给网络设置密度等级划分和密度图估计两个任务，并将与密度等级划分任务有关的特征嵌入到用于估计密度图的特征中，使两个任务相互关联，从而进一步提升了密度图的准确性。

Idrees 等[54]分析了目标数量、密度图和位置信息之间的关系，认为这三者之间存在特定的内在联系。在此分析的基础上，Idrees 在网络中使用了 4 种不同的损失函数，分别用来度量目标数量的误差（L_c loss），两种尺度下的密度图误差（L_1 loss 和 L_2 loss）和位置信息（L_∞ loss）的误差，这种具有组合损失函数的网络结构如图 8 所示。

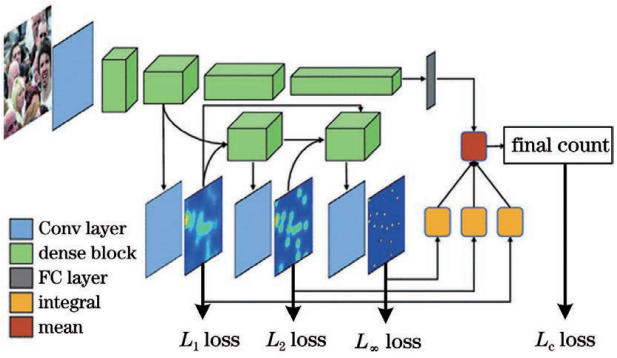

图 8　具有组合损失函数的网络结构

Fig. 8　Structure of network of combined loss function

与多列网络结构相反，Zhang 等[55]采用单一尺度的小卷积核构建深层网络 SaCNN，其网络结构如图 9 所示。该网络通过融合不同深度的特征图来适应目标尺度大小的变化，这种单列多尺度网络的训练参数更少，训练速度更快。同时网络也设置了密度图估计和目标数量估计两个任务，这两个任务与计数任务直接相关，更有利于降低计数的误差。一般来说，基于密度图估计的误差（L_D）和基于目标数量估计的误差（L_Y）都是采用欧氏距离度量，但是考虑到当场景中的目标数量较少时，直接采用欧氏距离度量目标数量的误差是不合适的，因此改用欧氏距离度量目标数量的相对误差，而非绝对误差。实验结果表明相对误差项对稀疏目标场景的计数性能有显著提升。而 Sang 等[56]则是在 SaCNN 结构的基础上，采用 MCNN[19]中生成密度图的方式重新对网络进行训练。由于采用的密度图更匹配目标的尺度变化，该方法在 Shanghai Tech 数据集[19]上的表现优于 SaCNN。

Zhang 等[57]提出了一种以密度图估计任务为主、密度等级估计任务为辅的多分辨率注意力网络结构 MRA-CNN，并在网络的顶部使用空洞卷积来扩大感受野并使用生成的注意力图对空间特征图进

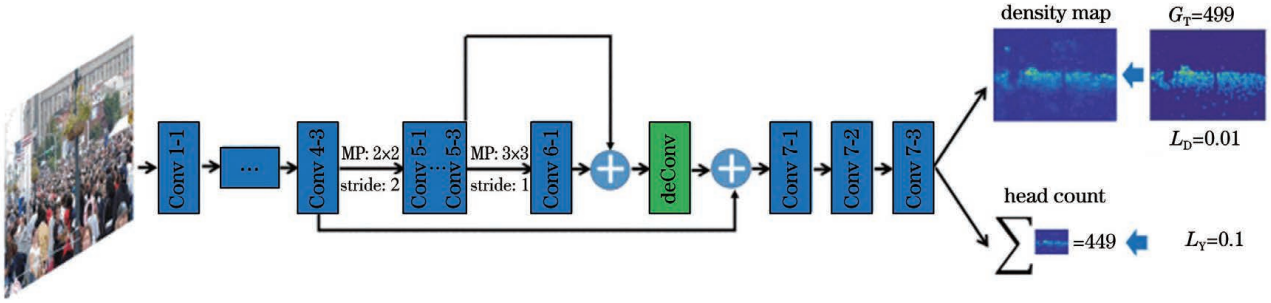

图 9　SaCNN 网络结构

Fig. 9　Architecture of SaCNN

行过滤。从图 10 可以看出，与 MRA-CNN 在网络中嵌套注意力图不同，SFANet[58] 将注意力掩模的生成当作一项子任务来训练（也可看作是图像分割任务）。密度图任务和注意力掩模任务的网络结构十分相似，可在网络的前半部分实现参数共享。和 SFANet 类似，ACCNet[59] 也是采用 VGG 网络[28] 来提取特征，并借助语义分割的结果引导密度图的生成。不同的是 ACCNet 网络后端生成密度图和分割结果时，网络设计更为复杂一些。

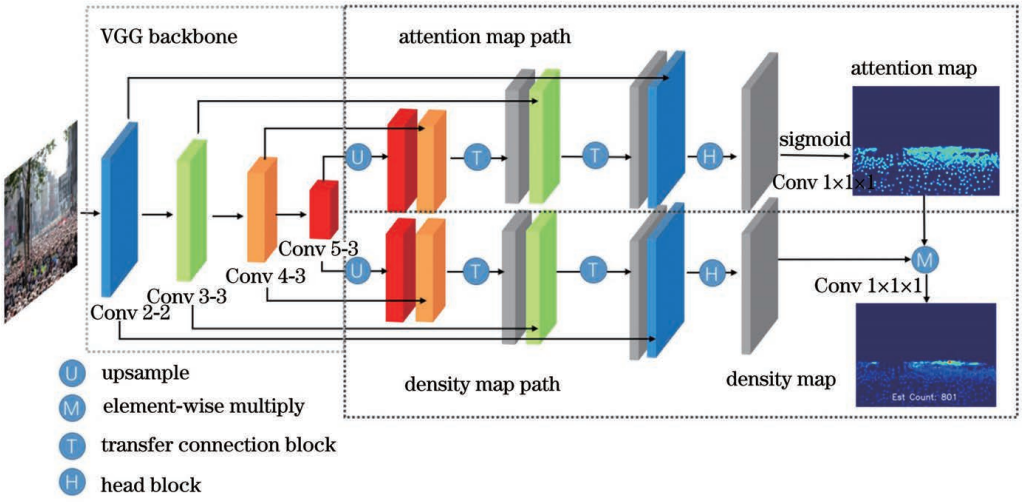

图 10　SFANet 结构

Fig. 10　Architecture of SFANet

由于场景的复杂性，经常会有一些物体被错误估计为人群，为了解决这个问题，Chen 等[60] 提出了一种结构较为复杂的人群注意力网络 CAT-CNN（图 11）。网络主体由 4 个部分组成：用于提取特征的多信息处理模块、置信度模块、密度图估计模块和融合模块。这 4 个部分分别对应 4 个任务，网络的总损失函数也由 4 个损失函数组成。多信息处理模块除了提取特征外，还负责对输入图像内的人群密度类别进行估计；置信度模块负责完成图像分割任务；密度图估计模块和融合模块则分别计算与置信图融合之前和融合之后的密度图损失值。Cao 等[61] 通过多列网络来获得具有多种感受野的特征，通过联合训练图像分割和密度图估计两种任务来优化模型，提高输出密度图的质量。

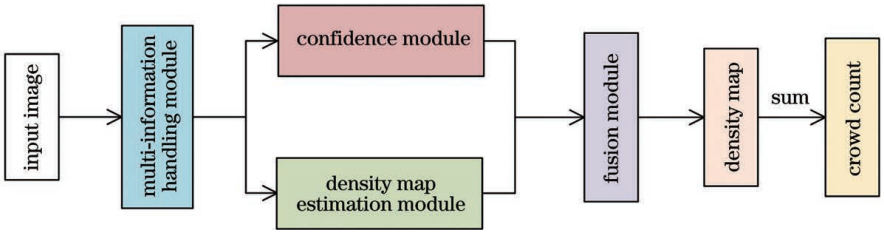

图 11　CAT-CNN 结构

Fig. 11　Architecture of CAT-CNN

如图 12 所示，与 FCN-rLSTM[10] 最后的残差连接类似，多任务全卷积网络（FCN-MT）[12] 也是通过 FCN 预测密度图，然后在密度图估计的基础上推测出目标数量。不同的是 FCN-MT 直接对密度图估计的残差进行拟合，而 FCN-rLSTM 通过 LSTM 获得了连续图片的时序信息，提升了对密度图的估计质量。

2.4　其他类型模型

以上方法都只能对特定的目标类别进行估计，为了能对任意类别的目标都能进行计数，GMN[62] 利用图像的自相似性来对图像中的相似目标进行匹配，将目标的计数问题转化为目标的匹配问题。利用目标检测的方法，Akram 等[63] 提出先对细胞的回归框进行预测，把可能的候选区域输入到下一层网络中进行分割，最终得到细胞分割的结果，其附属结果就是细胞的数量，图 13 是相应的网络结构。类似地，刘晓平[64] 也是先提取细胞的候选区域，然后对候选区域内的密度分布进行估计，不同的是该项工

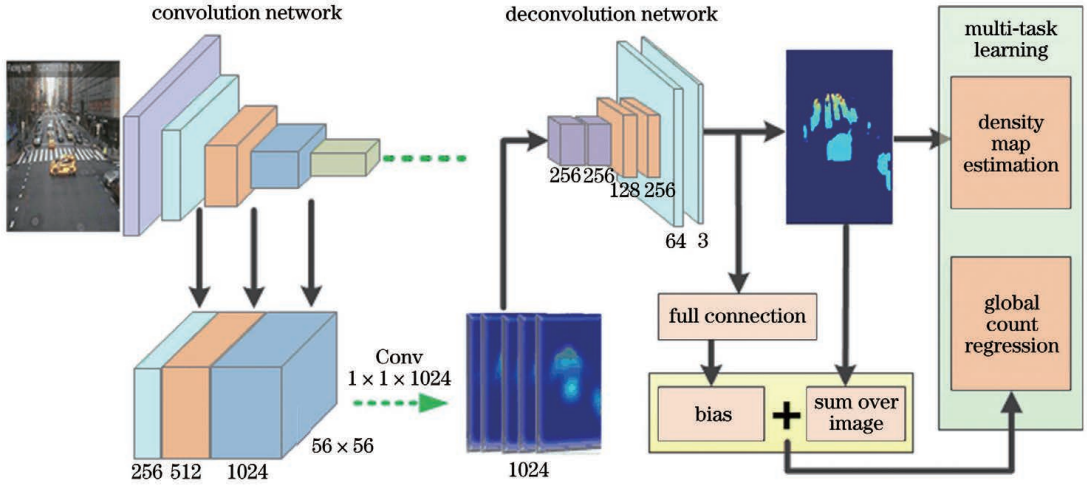

图 12　FCN-MT 结构

Fig. 12　Architecture of FCN-MT

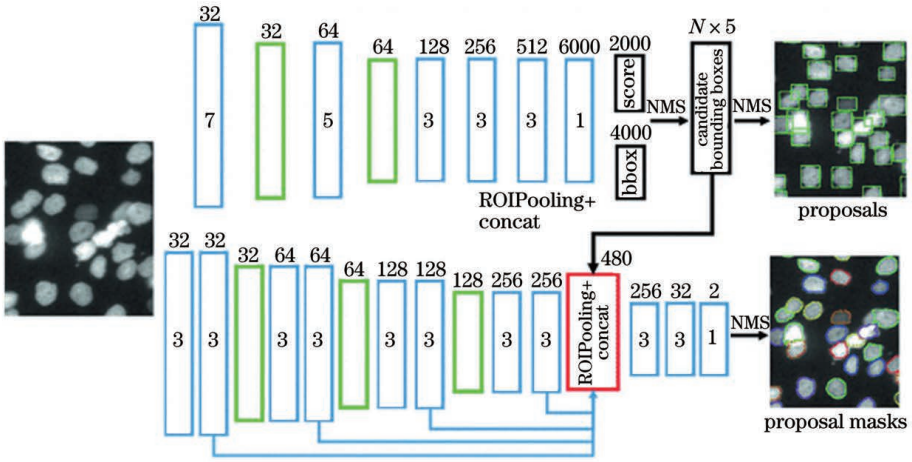

图 13　细胞分割网络结构

Fig. 13　Architecture of cell segmentation network

作创造性地在细胞计数模型中引入了推荐系统的召回排序模型，经过召回模型筛选后的图像区域所携带的噪声大大减少，模型的计算量也减少。

3　常用的数据集和评价指标

目标计数的应用范围十分广泛，在人群计数[13,46,52,57-58]、细胞计数[17-18,21,32,42]、车辆计数[10,12,24,27]等方面均有相应的应用。为了便于横向比较各个模型之间的差异，本文将在人群、细胞、车辆这个三类目标的公开数据集上展示各个模型的计数误差。

3.1　人群计数常用数据集

在人群计数任务中，常用的公开数据集有以下 5 种：

1）UCSD[65]。UCSD 数据集是第一个用于人群计数的数据集，它包含 2000 张分辨率为 158×238 的图片，每隔 5 张图片对人群进行一次标记，剩余图片的标记信息则由插值得到。数据集共包含 49885 个行人，选择第 600 张到第 1399 张这 800 张图片作为训练集，剩余的 1200 张图片作为测试集。该数据集记录了固定场景下的人群流动情况。

2）Mall[66]。Mall 数据集来源于一个商场公共区域的监控视频，包含 2000 张分辨率大小为 240×320 的图片，一共有 62325 个行人。相比于 UCSD 数据集，Mall 数据集记录的场景相对复杂一些：光照会随一天内时间的推移发生变化，人群分布密度不一致，静止人群和活动人群同时存在，透视畸变更严重，经常存在物体遮挡问题。

3）UCF_CC_50[67]。前两种数据集记录的都是固定场景下的数据，UCF_CC_50 覆盖了许多不同的场景，如音乐会、抗议游行、体育场，对人群计数任务提出了更多的挑战。UCF_CC_50 只包含 50 张

图片，每张图片的人群数量最少为 94，最多为 4543，平均每张图片包含 1280 个人。

4) WorldExpo'10[13]。WorldExpo'10 数据集源于 2010 年上海世界博览会，它包含了 108 个监控摄像头的 1132 个视频序列，分辨率大小为 576×720，覆盖了大量不同的场景。作者团队标记了 3980 张图片，共计 199923 人，所有图片都是通过从视频序列中均匀采样得到的。

5) Shanghai Tech[19]。Shanghai Tech 数据集有 1198 张图像，共计 330165 个人。整个数据集由两部分组成，Part A 和 Part B。Part A 数据是通过在网络上搜集得到的，包含 482 张大小不一致的图片，每张图片人数最少为 33，最多为 3139 人，一共包含 241677 人。Part B 拍摄的是上海主城区的繁华街头场景，包含 716 张分辨率大小为 768×1024 的图片，每张图片人数最少为 9 人，最多为 578 人，共计 88488 人。

通过上面的描述可知，近年来随着技术研究的发展和深入，人群数据集在场景、密度分布、分辨率、数量上的种类越来越多，这对人群计数任务提出了更多的挑战。表 1 是 5 个人群数据集相关信息的总结。图 14 展示了部分样例图片。

表 1　5 个公开人群数据集的总结

Table 1　Summary of five public pedestrian datasets

Dataset		Scene	Resolution	Range	Total number of people	Image No.
UCSD[65]		Same	158×238	11－46	49885	2000
Mall[66]		Same	240×320	13－53	62325	2000
UCF_CC_50[67]		Different	Different	99－4543	63974	50
WorldExpo'10[13]		Different	576×720	1－253	199923	3980
Shanghai Tech[19]	Part A	Different	Different	33－3139	241677	482
	Part B	Different	768×1024	9－578	88488	716

图 14　6 个人群数据集的部分样例图片。(a) UCSD；(b) Mall；(c) UCF_CC_50；(d) WorldExpo'10；(e) Shanghai Tech Part A；(f) Shanghai Tech Part B

Fig. 14　Samples from six crowd datasets. (a) UCSD; (b) Mall; (c) UCF_CC_50; (d) WorldExpo'10; (e) Shanghai Tech Part A; (f) Shanghai Tech Part B

3.2　细胞计数常用数据集

在细胞计数任务中，常用的公开数据集有以下三种：

1) VGG Cells[17]。VGG Cells 是一个模拟的荧光细胞数据集，它逼真地模拟了真实图片中经常会出现的一系列问题，如细胞重叠、细胞形状变化、离焦模糊、细胞边缘渐晕、光照不均等。数据集一共包含 200 张分辨率大小为 256×256 的图片，每张图片中最少包含 74 个荧光细胞，最多包含 317 个细胞。

2) MBM Cells[40]。MBM Cells 初次在文献[68]中被引入，它记录了 8 个不同病人的健康骨髓细胞，分辨率大小为 1200×1200，共 11 张。图片中包含许多残留的细胞组织和染色杂质，导致有些细胞标记不确定。Cohen 等[40]对错误的标记进行了修正，并将图片裁剪为 44 张分辨率大小为 600×

600 的图片。

3) Adipocyte Cells[69]。Adipocyte Cells 是人体皮下脂肪细胞数据集，它是通过一个大小为 1700×1700 的窗口滑动扫描得到的，将每张图片下采样至 150×150。脂肪细胞的尺寸在 20～200 μm 范围内变化，尺度差异十分明显，且细胞之间粘连紧密，难以区分。

表 2 是三个细胞数据集的相关信息总结。细胞数据集不同于人群数据集，若没有专业知识的支撑，一般无法对从网络上搜集的数据或任意视频数据进行标注，所以细胞数据集的数量相对较少，在实验时常采用交叉验证法。图 15 对三种数据集进行了样例展示，可以看到这三种数据集所面临的主要挑战各不相同。

表 2　三个公开细胞数据集的总结

Table 2　Summary of three public cell datasets

Dataset	Resolution	Range	Total number of cells	Image No.
VGG Cells[17]	256×256	74—317	35192	200
MBM Cells[40]	600×600	65—193	5446	44
Adipocyte Cells[69]	150×150	48—299	31017	200

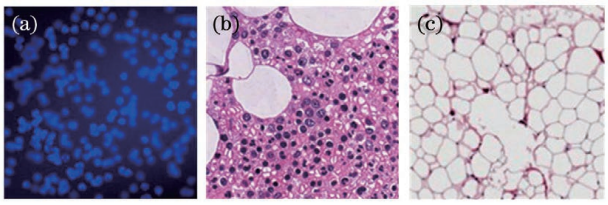

图 15　三个细胞数据集的部分样例图片。(a) VGG Cells；(b) MBM Cells；(c) Adipocyte Cells

Fig. 15　Samples from three cell datasets. (a) VGG Cells; (b) MBM Cells; (c) Adipocyte Cells

3.3　车辆计数常用数据集

在车辆计数任务中，常用的公开数据集有以下两种(图 16)：

1) WebCamT[12]。WebCamT 的特点是低分辨率、低帧率、车辆严重遮挡。多台不同的交通摄像机记录了在不同天气条件下、不同时刻的路况。整个数据集被划分为包含 45850 张图片的训练集和包含 14150 张图片的测试集，场景可分为市区道路和公园道路两类。

2) TRANCOS[11]。TRANCOS 包含 1244 张图片，其中记录了多种不同的道路交通场景，一共有 46796 辆标记车辆。由于这些图片来自不同的拍摄场景和视角，因此透视参数也不固定。此外，每张图片还提供了一个感兴趣区域(ROI)供模型性能评估用。

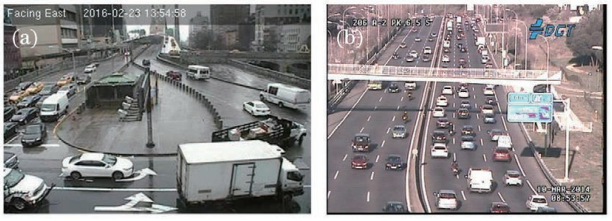

图 16　两种数据集的部分样例图片。(a) WebCamT；(b) TRANCOS

Fig. 16　Samples from two datasets. (a) WebCamT; (b) TRANCOS

3.4　评价指标及模型结果比较

近年来，平均绝对误差(MAE，可用 E_{MAE} 表示)和均方根误差(RMSE，可用 E_{RMSE} 表示)常用来度量目标计数模型的性能。人群计数中常用 MAE 和 RMSE 度量目标计数模型的性能，细胞计数和车辆计数中常用 MAE 度量目标计数模型的性能。MAE 反映计数的准确性，RMSE 反映模型的鲁棒性，可分别表示为

$$E_{\mathrm{MAE}}=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i|, \tag{2}$$

$$E_{\mathrm{RMSE}}=\left[\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2\right]^{\frac{1}{2}}, \tag{3}$$

式中：N 表示测试集的大小；y_i 和 $\hat{y}_i$ 分别表示第 i 张图中目标的真实数量和估计数量。

表 3 展示了近年来部分人群计数模型在 5 个主流人群数据集上的结果。方法 1～3 是基于回归的模型，方法 4～19 是基于密度估计的模型。经过对比之后可以发现，直接估计目标数量的模型表现会优于部分密度图估计的模型，但是以密度图的形式来表示计数结果可以传递更多的信息。方法 13～19 是在密度估计的基础上添加了注意力模块，增强了网络的特征表达，可以看到包含注意力模块的网络计数精度普遍有所提升，其中方法 18 和 19 由于特别关注了不同尺度下对应密度图的生成情况，其整体的计数误差更低。方法 20～30 是多任务模型，其中方法 26～29 引入了注意力模块。通过纵向对比可以推测，多任务模型优于单任务模型，包含注意力机制的模型优于不包含注意力机制的模型。在表 3 所列举的模型中，SFANet[58] 的整体表现最好，图 17 展示了 SFANet[58] 在 Shanghai Tech 数据集上的预测结果以及网络中间产生的注意力图。

表 3　人群计数模型比较

Table 3　Comparison of crowd counting models

Number	Method	UCSD[65]		Mall[66]		UCF_CC_50[67]		WorldExpo'10[13]		SHT A[19]		SHT B[19]	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
1	Shang *et al*.[6]					270.3		11.7					
2	CNN boosting[8]	1.10		2.01		364.4							
3	Marsden *et al*.[9]									85.7	131.1	17.7	28.6
4	Lempitsky *et al*.[17]					493.4	487.1						
5	Fiaschi *et al*.[21]												
6	MCNN[19]	1.07	1.35			377.6	509.1	11.6		110.2	173.2	26.4	41.3
7	Hydra CNN[11]					333.7	425.3						
8	Wang *et al*.[25]					264.9	382.1	8.6		83.7	124.5	17.9	32.4
9	FCN[29]					338.6	424.5			126.5	173.5	23.8	33.1
10	A-CCNN[30]	1.35				367.3							
11	POCNet[34]	1.24	1.50	1.82	5.48							12.1	20.3
12	DecideNet[35]			1.52	1.90			9.23				20.8	29.4
13	SPN[36]	1.03	1.32			259.2	335.9			61.7	99.5	9.4	14.4
14	AM-CNN[43]					279.5	377.8	7.84		87.3	132.7	15.6	26.4
15	SCAR[44]					259.0	374.0			66.3	114.1	9.5	15.2
16	Hossain *et al*.[46]			1.28	1.68	271.6	391.0					16.9	28.4
17	RANet[47]					239.8	319.4			59.4	102.0	7.9	12.9
18	ASNet[48]					174.8	251.6	6.6		57.8	90.1		
19	Wang *et al*.[49]					170.1	232.4	6.5		57.7	99.7	7.4	11.1
20	Cross-scene[13]	1.60	3.31			467.0	498.5	10.7		181.8	277.7	32.0	49.8
21	FF-CNN[51]									81.8	138.8	16.5	26.2
22	MMCNN[52]	1.02	1.18	1.98	5.68	320.6	323.8	9.1		91.2	128.6	18.5	29.3
23	DensityCNN[53]					244.6	341.8	6.9		63.1	106.3	9.1	16.3
24	SaCNN[55]					314.9	424.8	8.5		86.8	139.2	16.2	25.8
25	Sang *et al*.[56]									75.8	124.9	11.0	18.6
26	MRA-CNN[57]					240.8	352.6	7.5		74.2	112.5	11.9	21.3
27	SFANet[58]	0.82	1.07			219.6	316.2			59.8	99.3	6.9	10.9
28	ACCNet[59]	1.00	1.27			201.6	282.1			64.3	104.1	8.7	13.6
29	CAT-CNN[60]					235.5	324.8	7.2		66.7	101.7	11.2	20.0
30	MSMT-CNN[61]					319.5	358.1	9.3					
31	GMN[62]									95.8	133.3		

Notes: SHT is the abbreviation for Shanghai Tech

表 4 展示了部分模型在三个常用的公开细胞数据集上的结果。在细胞计数领域，模型通常用不同的训练集训练 N 次，并计算 N 次测试结果(MAE)的平均值和标准差。方法 1 直接估计细胞的数量，方法 8 通过估计样例和目标之间的相似性来确定目标的数量。其他方法均是基于密度图估计的模型，

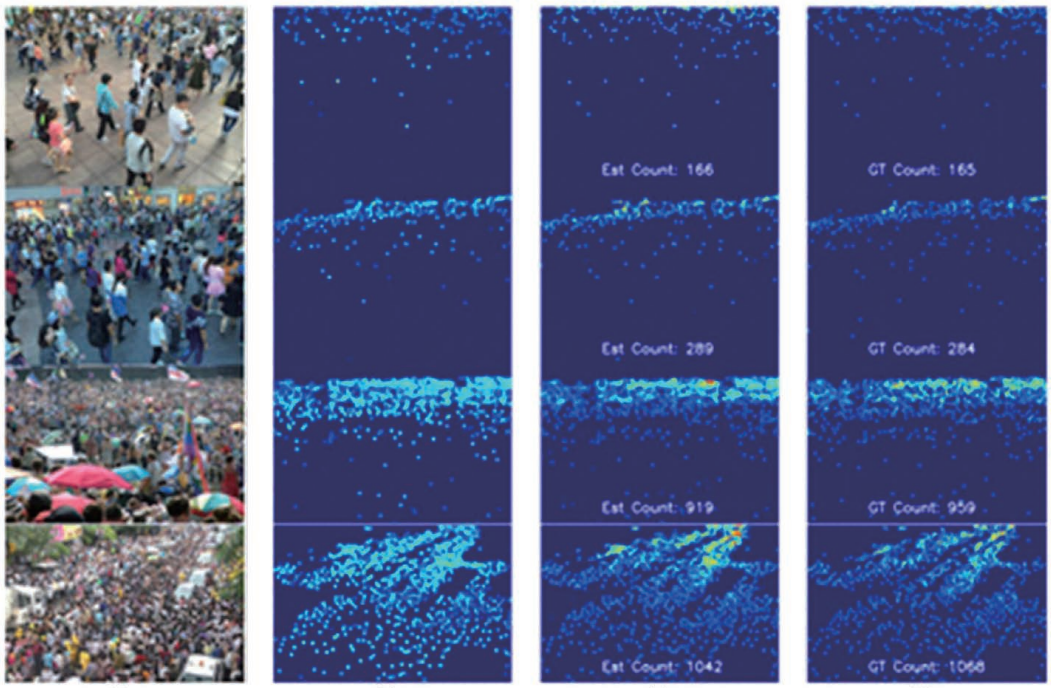

图 17　SFANet 在 Shanghai Tech 数据集上的估计结果，前两行为 Part B，后两行为 Part A[58]。(a) 输入图像；(b)注意力图；(c)密度图；(d)真实地物

Fig. 17　Estimation results on Shanghai Tech dataset generated by SFANet. The first two rows belong to Part B, and the last two rows belong to Part A[58]. (a) Input images; (b) attention maps; (c) density maps; (d) ground truths

表 4　细胞计数模型比较

Table 4　Comparison of cell counting models

Number	Method	VGG Cells[17]		MBM Cells[40]		Adipocyte Cells[69]	
		$N=32$	$N=50$	$N=10$	$N=15$	$N=25$	$N=50$
1	Marsden *et al.*[9]			21.5±4.2	20.5±3.5		
2	Lempitsky *et al.*[17]	3.5±0.2					
3	Fiaschi *et al.*[21]	3.2±0.1					
4	FCRN-A[18]	2.9±0.2	2.9±0.2	22.2±11.6	21.3±9.4		
5	Count-ception[40]	2.4±0.4	2.3±0.4	10.7±2.5	8.8±2.3	21.9±2.8	19.4±2.2
6	Cell-Net[42]		2.2±0.5	9.8±3.2			
7	SAU-Net[45]		2.6±0.4		5.7±1.2		14.2±1.6
8	GMN[62]	3.6±0.3					

其中方法 2 和 3 采用传统的机器学习方法，而方法 4～7 采用深度学习方法。通过对比在 VGG Cells 上的数据可以发现，方法 4～7 的表现优于方法 2～3，说明合适的网络结构也能在小数据集上取得好的效果。Cell-Net[42]采用了金字塔残差连接的形式，因此对重叠的细胞群和单个细胞之间的这种尺度差异比较敏感。而 SAU-Net[45]包含了空间注意力模块，具有更丰富的全局信息，因此在另外两个数据集上表现更好。

表 5 展示了近些年部分计数模型在 WebCamT[12]和 TRANCOS[11]两个数据集上的结果。在车辆计数领域，通常采用 MAE 和 RMSE 来评估模型性能。在这两个数据集上都使用 MAE 来度量模型性能，但不同的是在 TRANCOS 数据集上定义了一种新的度量方式 GAME(Grid Average Mean absolute Error)，其具体计算式[11]为

$$\mathrm{GAME}(L)=\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{l=1}^{4^L}\left|D^l_{I_n}-D^l_{I^{gt}_n}\right|\right),\quad(4)$$

式中：$D^l_{I_n}$ 和 $D^l_{I^{gt}_n}$ 分别对应第 n 张密度图中区域 l 内目标数量的估计值和真实值。假设给定一个具体的等级 L，GAME(L)将密度图划分为 4^L 个非重叠的子区域。图像的总误差则是这些子区域误差之和，这种度量方式能分别反映模型在全局区域和局部区域的计数误差。GAME(0)即为一般情况下使用的 MAE，即全局范围内的计数误差。

在 WebCamT 数据集上，利用了视频序列的时序特性的 FCN-rLSTM[10] 表现最佳，具有密度图估计和目标数量估计两种任务的多任务模型 FCN-MT[12] 的性能次之。在 TRANCOS 数据集上，DensityCNN[53] 表现最佳，它融合了用于密度等级划分的特征和用于密度图估计的特征，进一步增强了特征的表达能力。其次，CSRNet[27] 通过空洞卷积来获得大的感受野，从而降低模型的计数误差，说明了感受野大小对计数模型性能的重要性。

表 5　车辆计数模型比较

Table 5　Comparison of vehicle counting models

Number	Method	WebCamT[12]		TRANCOS[11]			
		Downtown	Parkway	GAME 0	GAME 1	GAME 2	GAME 3
1	Lempitsky *et al*.[17]	5.91	5.19	13.76	16.72	20.72	24.36
2	Fiaschi *et al*.[21]			17.77	20.14	23.65	25.99
3	Marsden *et al*.[9]			9.70			
4	FCN-rLSTM[10]	1.53	1.63	4.38			
5	CCNN[11]			12.49	16.58	20.02	22.41
6	Hydra-CNN[11]	3.55	3.64	10.99	13.75	16.69	19.32
7	AMDCN[24]			9.77	13.16	15.00	15.87
8	CSRNet[27]			3.56	5.49	8.57	15.04
9	DensityCNN[53]			3.17	4.78	6.30	8.26
10	FCN-MT[12]	2.74	2.52	5.31			

4　结束语

从模型任务属性的角度对目标计数方法进行了分类，模型任务属性主要分为单任务模型和多任务模型，而单任务模型又可分为基于回归的计数模型和基于密度图估计的计数模型。由于密度图既能反映目标的总数量，又能反映目标的空间分布，因此基于密度图的估计方法逐渐成为目标计数模型的主流方向。重点介绍了近些年来基于密度图估计的计数模型的发展，并对各个模型的性能进行了比较和分析。总的来看，基于计算机视觉的目标计数方法已经取得了较大的进步，但还存在以下几个方面的问题有待解决：

1）密度图的设置。对于目标个体大小存在差异的场景，尤其是监控镜头下透视效应导致的目标尺度的显著差异，尽管有文献[19]已经提出可利用目标个体与周围人群之间的平均距离来自适应地计算密度函数参数，且这种方法的计数结果的准确性的确得到提高，但是这种密度计算方法只适用于密集分布的人群，对于稀疏的人群分布无效。如何设计一个和目标尺寸相匹配的密度函数是提升模型性能的一个关键影响因素。

2）通过估计密度图来推测目标数量的方法只关注图像内的目标总数量是否接近真实值，但对于密度图本身是否能正确反映目标的空间分布却很少加以讨论。例如，在目标总数量估计正确的情况下，局部区域内的目标数量估计可能存在较大偏差，此时密度图就无法正确反映目标真实的空间分布。因此，对于通过密度图推断目标数量的方法，如何在保证全局估计数量准确性的同时提高密度图估计的准确性是提升计数模型性能的一个重要改进方向。

3）计数目标的变化。目标计数有许多不同的应用领域，目前大多数研究都只针对单个应用领域。虽然已有文献[9]提出了能够同时适用于多个领域的模型，但相比专门针对某个应用领域提出的模型，该模型在计数准确度上还略逊一筹。在不同的应用

领域中，模型的任务属性有所差异，但是模型的总体目标是一致的，即估计出图像中目标的数量。因此，如何同时学习多种类型目标的密度分布也是将来要攻克的一个难点。

4）数据集的大小。首先，人群数据集的标注是一项十分重要而又繁琐的工作，尤其是对于密集的人群分布。其次，细胞数据集本身较难获取，且细胞标注对研究人员的专业知识的要求也比较高，这也进一步导致了有关细胞计数的数据集不充足。其他应用领域的数据集也同样存在类似的问题。因此，建立一个公开、全面的数据集是当下最迫切的任务。

5）实时性。极少有文献会讨论计数模型的实时性问题，大多数研究是在图像层面上对模型进行比较。人群计数网络的结构一般比较复杂，其实时性有待考究。在保证计数准确度的同时，降低网络复杂度、提高网络实时性是目标计数未来的一个发展方向。

参　考　文　献

[1] Zhan B B, Monekosso D N, Remagnino P, et al. Crowd analysis: a survey[J]. Machine Vision and Applications, 2008, 19(5/6): 345-357.

[2] Saleh S A M, Suandi S A, Ibrahim H. Recent survey on crowd density estimation and counting for visual surveillance[J]. Engineering Applications of Artificial Intelligence, 2015, 41: 103-114.

[3] Zhang J J, Shi Z G, Li J C. Current researches and future perspectives of crowd counting and crowd density estimation technology[J]. Computer Engineering & Science, 2018, 40(2): 282-291.
张君军, 石志广, 李吉成. 人数统计与人群密度估计技术研究现状与趋势[J]. 计算机工程与科学, 2018, 40(2): 282-291.

[4] Sindagi V A, Patel V M. A survey of recent advances in CNN-based single image crowd counting and density estimation[J]. Pattern Recognition Letters, 2018, 107: 3-16.

[5] Gao G S, Gao J Y, Liu Q J, et al. CNN-based density estimation and crowd counting: a survey[EB/OL]. (2020-03-01)[2020-10-10]. https://arxiv.org/abs/2003.12783.

[6] Shang C, Ai H Z, Bai B. End-to-end crowd counting via joint learning local and global count[C]//2016 IEEE International Conference on Image Processing (ICIP), September 25-28, 2016, Phoenix, AZ, USA. New York: IEEE Press, 2016: 1215-1219.

[7] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1-9.

[8] Walach E, Wolf L. Learning to count with CNN boosting[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9906: 660-676.

[9] Marsden M, McGuinness K, Little S, et al. People, penguins and petri dishes: adapting object counting models to new visual domains and object types without forgetting[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8070-8079.

[10] Zhang S H, Wu G H, Costeira J P, et al. FCN-rLSTM: deep spatio-temporal neural networks for vehicle counting in city cameras[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3687-3696.

[11] Oñoro-Rubio D, López-Sastre R J. Towards perspective-free object counting with deep learning[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9911: 615-629.

[12] Zhang S H, Wu G H, Costeira J P, et al. Understanding traffic density from large-scale web camera data[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4264-4273.

[13] Zhang C, Li H S, Wang X G, et al. Cross-scene crowd counting via deep convolutional neural networks[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 833-841.

[14] Xu C Y. Research on automated cervical cytological smears interpretation method[D]. Chongqing: Chongqing University, 2014.
徐传运. 宫颈细胞学涂片自动判读方法研究[D]. 重庆: 重庆大学, 2014.

[15] Maitra M, Gupta R K, Mukherjee M. Detection and counting of red blood cells in blood cell images using Hough transform [J]. International Journal of Computer Applications, 2012, 53(16): 13-17.

[16] Kothari S, Chaudry Q, Wang M D. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques[C]//2009 IEEE International Symposium on Biomedical

Imaging: From Nano to Macro, June 28-July 1, 2009, Boston, MA, USA. New York: IEEE Press, 2009: 795-798.

[17] Lempitsky V, Zisserman A. Learning to count objects in images[C]//Advances in neural information processing systems. December 6-9, 2010, Vancouver, BC. Canada: Curran Associates Inc., 2017: 1324-1332.

[18] Xie W D, Noble J A, Zisserman A. Microscopy cell counting and detection with fully convolutional regression networks[J]. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2018, 6(3): 283-292.

[19] Zhang Y Y, Zhou D S, Chen S Q, et al. Single-image crowd counting via multi-column convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 589-597.

[20] Ma Z, Yu L, Chan A B. Small instance detection by integer programming on object density maps[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3689-3697.

[21] Fiaschi L, Koethe U, Nair R, et al. Learning to count with regression forest and structured labels [C]//Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), November 11-15, 2012, Tsukuba, Japan. New York: IEEE Press, 2012: 2685-2688.

[22] Sommer C, Straehle C, Köthe U, et al. Ilastik: interactive learning and segmentation toolkit[C]// 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, March 30-April 2, 2011, Chicago, IL, USA. New York: IEEE Press, 2011: 230-233.

[23] Borstel M, Kandemir M, Schmidt P, et al. Gaussian process density counting from weak supervision[M]// Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 365-380.

[24] Deb D, Ventura J. An aggregated multicolumn dilated convolution network for perspective-free counting[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 308-309.

[25] Wang Y J, Hu S Y, Wang G D, et al. Multi-scale dilated convolution of convolutional neural network for crowd counting[J]. Multimedia Tools and Applications, 2020, 79(1/2): 1057-1073.

[26] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23)[2020-10-10]. https://arxiv.org/abs/1511.07122.

[27] Li Y H, Zhang X F, Chen D M. CSRNet: dilated convolutional neural networks for understanding the highly congested scenes[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1091-1100.

[28] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-01)[2020-10-10]. https://arxiv.org/abs/1409.1556.

[29] Marsden M, McGuinness K, Little S, et al. Fully convolutional crowd counting on highly congested scenes[C]// Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, February 27-March 1, 2017, Porto, Portugal. Setúbal: SCITEPRESS-Science and Technology Publications, 2017: 27-33.

[30] Amirgholipour S, He X J, Jia W J, et al. A-CCNN: adaptive CCNN for density estimation and crowd counting[C]//2018 25th IEEE International Conference on Image Processing (ICIP), October 7-10, 2018, Athens, Greece. New York: IEEE Press, 2018: 948-952.

[31] Xie Y P, Xing F Y, Kong X F, et al. Beyond classification: structured regression for robust cell detection using convolutional neural network[M]// Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 358-365.

[32] Xue Y, Ray N, Hugh J, et al. Cell counting by regression using convolutional neural network[M]// Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9913: 274-290.

[33] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[34] Liu X. Object counting in surveillance video[D]. Hefei: University of Science and Technology of China, 2018.
刘旭. 视频监控中的目标计数方法研究[D]. 合肥: 中国科学技术大学, 2018.

[35] Liu J, Gao C Q, Meng D Y, et al. DecideNet:

counting varying density crowds through attention guided detection and density estimation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5197-5206.

[36] Chen X Y, Bin Y R, Sang N, et al. Scale pyramid network for crowd counting[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV), January 7-11, 2019, Waikoloa, HI, USA. New York: IEEE Press, 2019: 1941-1950.

[37] Rad R M, Saeedi P, Au J, et al. Blastomere cell counting and centroid localization in microscopic images of human embryo[C]//2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), August 29-31, 2018, Vancouver, BC, Canada. New York: IEEE Press, 2018: 1-6.

[38] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. New York: Springer, 2015: 234-241.

[39] He K M, Zhang X Y, Ren S Q, et al. Identity mappings in deep residual networks[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9913: 630-645.

[40] Cohen J P, Boucher G, Glastonbury C A, et al. Count-ception: counting by fully convolutional redundant counting[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 18-26.

[41] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2818-2826.

[42] Rad R M, Saeedi P, Au J, et al. Cell-net: embryonic cell counting and centroid localization via residual incremental atrous pyramid and progressive upsampling convolution[J]. IEEE Access, 2019, 7: 81945-81955.

[43] Zhang Y M, Zhou C L, Chang F L, et al. Attention to head locations for crowd counting[M]//Zhao Y, Barnes N, Chen B Q, et al. ICIG 2019: image and graphics. Lecture notes in computer science. Cham: Springer International Publishing, 2019, 11902: 727-737.

[44] Gao J Y, Wang Q, Yuan Y. SCAR: spatial-/channel-wise attention regression networks for crowd counting[J]. Neurocomputing, 2019, 363: 1-8.

[45] Guo Y, Stein J, Wu G R, et al. SAU-net: a universal deep network for cell counting[C]//Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Niagara Falls, NY, USA. New York: ACM, 2019: 299-306.

[46] Hossain M, Hosseinzadeh M, Chanda O, et al. Crowd counting using scale-aware attention networks[C]//2019 IEEE Winter Conference on Applications of Computer Vision (WACV), January 7-11, 2019, Waikoloa, HI, USA. New York: IEEE Press, 2019: 1280-1288.

[47] Zhang A R, Shen J Y, Xiao Z H, et al. Relational attention network for crowd counting[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6787-6796.

[48] Jiang X H, Zhang L, Xu M L, et al. Attention scaling for crowd counting[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 4705-4714.

[49] Wang Y J, Zhang W, Liu Y Y, et al. Two-branch fusion network with attention map for crowd counting[J]. Neurocomputing, 2020, 411: 1-8.

[50] Liu L B, Chen J Q, Wu H F, et al. Efficient crowd counting via structured knowledge transfer[EB/OL]. (2020-08-11)[2020-10-10]. https://arxiv.org/pdf/2003.10120.pdf.

[51] Luo H L, Sang J, Wu W Q, et al. A high-density crowd counting method based on convolutional feature fusion[J]. Applied Sciences, 2018, 8(12): 2367.

[52] Yang B, Cao J M, Wang N, et al. Counting challenging crowds robustly using a multi-column multi-task convolutional neural network[J]. Signal Processing: Image Communication, 2018, 64: 118-129.

[53] Jiang X H, Zhang L, Zhang T Z, et al. Density-aware multi-task learning for crowd counting[J]. IEEE Transactions on Multimedia, 2021, 23: 443-453.

[54] Idrees H, Tayyab M, Athrey K, et al. Composition loss for counting, density map estimation and localization in dense crowds[M]//Ferrari V, Hebert

M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11902: 544-559.

[55] Zhang L, Shi M J, Chen Q B. Crowd counting via scale-adaptive convolutional neural network[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV), March 12-15, 2018, Lake Tahoe, NV, USA. New York: IEEE Press, 2018: 1113-1121.

[56] Sang J, Wu W Q, Luo H L, et al. Improved crowd counting method based on scale-adaptive convolutional neural network[J]. IEEE Access, 2019, 7: 24411-24419.

[57] Zhang Y M, Zhou C L, Chang F L, et al. Multi-resolution attention convolutional neural network for crowd counting [J]. Neurocomputing, 2019, 329: 144-152.

[58] Zhu L, Zhao Z J, Lu C, et al. Dual path multi-scale fusion networks with attention for crowd counting [EB/OL]. (2019-02-01) [2020-10-10]. https://arxiv.org/abs/1902.01115v1.

[59] Yu S Y, Pu J. Aggregated context network for crowd counting [J]. Frontiers of Information Technology Electronic Engineering, 2020, 21 (11): 1626-1638.

[60] Chen J W, Su W, Wang Z F. Crowd counting with crowd attention convolutional neural network [J]. Neurocomputing, 2020, 382: 210-220.

[61] Cao J M, Yang B, Nan W, et al. Robust crowd counting based on refined density map[J]. Multimedia Tools and Applications, 2020, 79(3/4): 2837-2853

[62] Lu E, Xie W D, Zisserman A. Class-agnostic counting[M]//Jawahar C V, Li H D, Mori G, et al. Computer vision-ACCV 2018. Lecture notes in computer science. Cham: Springer, 2019, 11363: 669-684.

[63] Akram S U, Kannala J, Eklund L, et al. Cell segmentation proposal network for microscopy image analysis[M]//Carneiro G, Mateus D, Peter L, et al. Deep learning and data labeling for medical applications. Lecture notes in computer science. Cham: Springer, 2016, 10008: 21-29.

[64] Liu X P. A research on automatic cell counting method in fluorescence microimaging based on deep learning [D]. Chengdu: University of Electronic Science and Technology of China, 2020.
刘晓平. 基于深度学习的荧光显微成像中细胞自动计数方法研究[D]. 成都: 电子科技大学, 2020.

[65] Chan A B, Liang Z S, Vasconcelos N. Privacy preserving crowd monitoring: counting people without people models or tracking [C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008: 1-7.

[66] Chen K, Loy C C, Gong S G, et al. Feature mining for localised crowd counting [EB/OL]. [2020-10-10]. http://www.bmva.org/bmvc/2012/BMVC/paper021/index.html.

[67] 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, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 13824453.

[68] Kainz P, Urschler M, Schulter S, et al. You should use regression to detect cells [M] // Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 276-283.

[69] Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project [J]. Nature Genetics, 2013, 45(6): 580-585.