

基于深度神经网络的“弱监督”密集场景人群计数算法



刘 砚 雷印杰 宁 芊

四川大学电子信息学院 成都 610065

(yanliu27@stu.scu.edu.cn)

摘 要 目前,在密集场景人群计数任务中,标注真实密度图的方法是对行人头部的中心位置进行标注,并利用高斯卷积生成真实的密度分布图作为监督信息。但是,对于密集场景而言,这样的标注方式是费时、费力的,并且密集场景图片中有诸多“非受控”因素,如低分辨率、背景噪声、目标遮挡和尺度变化等。针对这一问题,提出了一种新的标注方法,即只需要知道图片中包含多少个物体,以图片中行人的数量作为监督信息。与传统的真实密度图相比,所提出的标记方法中以真实目标的数值为“弱监督”信息。实验结果表明,对于人群回归任务,利用弱监督信息对神经网络进行训练得到的模型能够较为准确地回归出图片中所包含目标的数量,从而证明了该方法的有效性。

关键词:深度学习;神经网络;人群计数;弱监督

中图法分类号 TP391.413

Study of Crowd Counting Algorithm of “Weak Supervision” Dense Scene Based on Deep Neural Network

LIU Yan, LEI Yin-jie and NING Qian

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

Abstract At present, in the crowd counting task of dense scenes, the method of annotating true density is to annotate the central position of pedestrian's head. Gaussian convolution is used to generate the ground-truth density map as the supervision information. However, for dense scenes, such labeling method is time-consuming and laborious, and there are many “uncontrolled” factors in the images of dense scenes, such as low resolution, background noise, heavy occlusion and scale change. To solve this problem, we proposed a new annotation method, that is, we only need to know how many persons are included in the picture, and the total count of pedestrians in the picture is used as the supervision information. Compared with the traditional real density map, in proposed labeling method, the real target value is used as the “weak supervision” information. The experimental results show that the model obtained by training neural network with weak supervisory information can accurately regress the number of targets in the image for crowd regression task, indicating the effectiveness of this method.

Keywords Deep learning, Neural network, Crowd counting, Weak supervision

1 引言

密集场景人群计数任务的目的是获得图片中所包含物体的数目。随着人口的增加,各类踩踏事故在全球各地层出不穷,造成了大量的人员伤亡。密集场景人群计数任务在公共安全、交通监测和流量分析等领域具有广泛的应用前景。近年来,国内外诸多学者对密集场景计数任务提出了推动性的解决办法。但是,目前的解决方法均使用人群的真实密度分布图作为真实标签,这样的标注方式对每一个计数目标进行点标记,往往需要投入大量的人力和物力。在密集场景图片中,低分辨率、物体遮挡和尺度变化等“非受控”因素,导致对每个物体进行实际标注并非易事。因此,我们利用“弱监督”

信息对图片进行标注,即用图片中包含的目标物体数目作为监督信息。

密集场景中人群计数任务的算法大致可以分为3类:1)基于目标检测的方法;2)基于回归的方法;3)利用深度学习和机器学习算法生成预测密度图的方法。

在基于目标检测的方法中,需要人为提取出图片中的低层次特征,并将其用特征框标记出来,特征框的数量即为图片中包含行人的数量。然而,由于目标遮挡和尺度变化等原因,该类方法在处理密集场景人群计数任务时,其预测值往往小于实际值。

基于回归的方法又可以分为两类:1)直接回归数值的方法,该方法通过自动特征提取回归出图片中包含的物体数目;

收稿日期:2019-07-31 返修日期:2019-11-07 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:四川省重点研发项目(2019YFG0409)

This work was supported by the Key Research and Development Program of Sichuan Province (2019YFG0409).

通信作者:宁芊(ningq@scu.edu.cn)

2) 回归密度的方法,该方法能够反映出物体在图片中的分布情况。在基于密度回归的方法中,对预测的密度图进行求和便可以得到物体的预测值。与目标检测的方法相比,基于回归的方法已经能够利用物体的空间分布信息与语义信息,为后续使用深度学习的方法奠定了理论基础。

上述方法都是利用传统的目标检测和特征回归等方法进行密集场景人群的计数。近年来,随着人工智能的飞速发展,在密集人群计数领域也出现了更具有挑战性的数据集。传统方法由于自身的局限性,不适宜处理较为复杂的人群分布情况。

随着深度学习技术在密集场景人群计数任务中取得巨大进展,许多传统方法尚且无法处理的问题(如尺度变化、物体遮挡等)都得到了一定程度的解决。但是,还有一些亟待解决的问题:1)在当前密集场景物体计数任务中,标记真实密度图的难度太大,且需要投入大量的人力物力;2)在当前的标记方式下,当获取物体分布的手段发生变化,如改变拍摄视角或人群本身的空间分布特征发生变化时,需重新进行标记。

针对上述问题,本文提出了一种使用“弱监督”的真实标记信息来训练深度神经网络的算法,并构造出了相应的深度神经网络模型。文中提出了两种方法:1)只使用少量点标记的“强监督”样本去训练深度神经网络;2)使用少量点标记的“强监督”样本与大量“弱监督”样本共同训练深度神经网络。实验结果表明,对少量的“强监督”标记样本加入大量的“弱监督”标记样本,能够大幅度地提升深度网络的泛化性能。

2 相关工作

密集场景人群计数最初是针对公共安全领域展开的,早期出现的主要是基于目标检测的方法。文献[1]使用方向梯度直方图对场景中的行人进行检测;文献[2]利用哈尔小波变换对场景中行人的脸部进行检测。在上述方法中,检测到的行人数目即为场景中所包含的人群数。但是,基于目标检测的人群计数算法无法处理存在严重遮挡和行人尺度变化的场景。对此,文献[3]利用随机森林提取图片中的相关特征,据此对场景图片中所包含的人群数目进行统计;文献[4]则利用高斯随机过程对场景中的行人数目进行回归。

随着深度学习的崛起,使用深度卷积神经网络处理密集场景人群计数问题吸引了国内外学者的注意。文献[5]设计了一个由3条支路组成的深度卷积神经网络。为了提取不同层次的上下文特征,设置每条支路的卷积核大小各不相同,以便处理不同的人群密度分布;将不同支路提取的特征图拼接起来输入到一个回归网络中,最终得到预测的人群分布密度图。为了加快多支路深度神经网络的运算速度,文献[6]预训练了一个分类器。分类器会对图片中所包含的物体数目进行分类,包括“密集”“适中”“稀疏”3个等级,并根据分类结果指导模型的训练过程。文献[7]提出利用上下文信息的方法提取图片的全局信息和局部信息,以生成较高质量的人群密度分布图。文献[8]使用VGG16^[9]作为“骨干网络”,利用深度神经网络能够提取高层次语义信息的特点,结合空洞卷积提取非相邻像素点之间的上下文关系,进而准确地对密集场景中的人群分布情况进行估计。

以上方法在深度学习网络的设计上对密集人群计数问题进行了研究。但是在现实情况中,标记好的高质量人群密度分布图难以获得。利用“半监督”学习的方法去构造深度学习模型,是处理当前人群密度计数任务数据集稀少问题的重要方法。文献[10]以现有的数据集为基础,在互联网上获得了一系列未被标注的密集人群图片;对这些未被标记的图像进行切片,保证小的切片包含在大的切片中,利用小的切片包含更少的人群而大的切片包含更多的人群这一原理,辅助现有的数据集对深度神经网络进行训练。文献[11]利用“GTA5”游戏中可以构造虚拟场景并获得虚拟场景中所布置“行人”坐标的原理,构造出一个进行了点标记的合成图像;并利用生成对抗网络^[12]探究了合成图像与真实场景图像的转化关系。文献[13]提出了一种自动编码器对特征图进行稀疏处理,再通过一个解码器将解码特征与原始输入信息进行自监督学习。

然而,在密集场景人群计数这一领域中,目前还没有对训练图片使用“弱监督”标记训练深度神经网络开展系统的研究。与全监督学习不同,在使用“弱监督”标记信息学习的条件下,只有小部分数据能够获得真实人群分布密度图,大批量的数据使用图片中包含的人群数目进行标记。在半监督学习算法中,只有少部分图片具有“强监督”标注信息,大量的图片不具备监督信息。在现实情况下,一个范围内人群的总体数量容易获得,但是要获得人群的具体分布情况绝非易事。

因此,本文提出了一种基于深度神经网络的“弱监督”密集场景人群统计算法,其利用“弱监督”标记的方式来减少科研人员对密集场景人群数据进行标记的工作量。

3 弱监督人群计数算法

3.1 问题定义与训练样本输入策略

当前,有关人群计数的数据集还相对较少,并且现有的“强监督”标记方法需要对每一个行人出现的位置进行标记(通常是标记在行人的头部中心位置)。为了减少标记真实密度图所消耗的精力,我们提出了一种基于“弱监督”标记方式学习的人群密度计数算法。在实际情况中,对于某一帧的监控图像,如果我们知道其中包含的人数(以该帧为参考帧),就可以根据当前帧与参考帧中(如在当前帧中有行人离开监控区域,有其他行人进入监控区域,或者监控场景中的人群在自由走动)人群的行为来推测后续帧中行人的数目。这样,我们便可以在不做“强监督”点标记的情况下获得大量以“弱监督”真值标记的图像。这种“弱监督”行人标记方法也可以轻松地推广到物体计数应用中。

为了评估本文提出的“弱监督”学习方法的鲁棒性,从原始训练集中选择少部分样本保留原有的“强监督”信息,即以图片的人群密度分布图为真值信息;更多的训练样本采用“弱监督”标记方法标记真值,即以图片中所包含的人群数目为真值信息。

本文方法基于深度神经网络,由于对数据标记的方式进行了改变,因此将训练集分成了两个子集: A_F 与 A_W 。其中, $A_F = \{(X_i, D_i)\}_{i=1}^m$, $A_W = \{(X_j, C_j)\}_{j=1}^n$ 。 A_F 代表使用“强监督”标记的训练样本构成的集合,其中图片 X_i 使用真实人群

密度分布图 D_i 作为监督信息; A_w 代表使用“弱监督”标记的训练样本构成的集合, 其中图片 X_i 使用图片中所包含的人群数目真值 C_i 作为监督信息。在本文的问题定义中: 大量数据以图片中包含的行人数目进行标记, 少量数据以真实的密度图进行标记。

3.2 人群密度特征提取

本文采用文献[8]中使用的空洞卷积神经网络作为骨干网络。该网络为主流人群计数网络之一, 其具有较高的预测精度。

空洞卷积神经网络以密集场景图片为输入, 以对应的预测人群密度分布图为输出, 预测过程如图 1 所示。

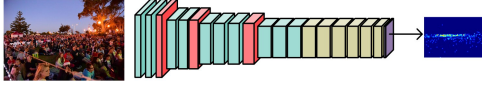


图 1 空洞卷积网络结构(电子版为彩色)

Fig. 1 Structure of dilation convolutional neural network

图 1 中, 青色标注的卷积层为卷积核为 3、步长为 1 的普通卷积层(Normal Convolution Layer); 红色标注的层为步长为 2 的最大池化层(Max Pooling Layer); 黄色标注的层为空洞卷积层(Dilated Convolution Layer), 其卷积核大小为 3, 步长为 1, 卷积率为 2; 紫色部分为卷积核大小为 1, 步长为 1, 输出为 1 的普通卷积层。场景图片经过图 1 所示的网络结构, 便可生成相应图片的人群密度预测图。

在传统的“强监督”标记下, 图片中每个行人的头部中间位置都会被打上一个标记, 从而得到一个使用二值化标注的人群分布图——行人出现的位置被标注为 1, 而没有行人出现的区域, 如背景环境(树木、天空等区域)则被标注为 0。对原始的二值化标记图像进行高斯模糊, 从而得到人群密度的分布函数 $D_i(x)$ 。

$$D_i(x) = \sum_{x_i \in H} G\left(\frac{\|x - x_i\|^2}{\sigma^2}\right) \quad (1)$$

其中, H 为一系列被标注点的集合, x_i 为对应行人被标注的坐标, $G\left(\frac{\|x - x_i\|^2}{\sigma^2}\right)$ 为一个以 x_i 为均值矢量、 σ^2 为方差的二维高斯核。对高斯核进行积分: $\int G\left(\frac{\|x - x_i\|^2}{\sigma^2}\right) = 1$ 。因此, 对真实密度图进行积分便可以获得图片中包含的真实人群数目。

使用“强监督”对原始图片进行标注与生成密度图的方式如图 2 所示。在原始图片中, 对每一个行人出现的位置进行点标记, 利用高斯过程生成如图 2 右图所示的密度图。密度图可以反映密集场景中人群分布的情况。



图 2 使用“强监督”方式对密集场景图片进行标记的示例图
(电子版为彩色)

Fig. 2 Example of annotating dense scene images with “strong supervision” method

在训练过程中, “强监督”标记的图片 X_F 在经过深度神经网络模型后, 得到预测的密度分布图 D_F 。

$$D_F(x) = \Phi(X_F; \theta) \quad (2)$$

其中, $\Phi(\cdot)$ 为深度神经网络模型, 在文本中使用空洞卷积神经网络; θ 为神经网络中的模型参数。针对“强监督”标记的图片, 本文采用欧氏距离度量作为损失函数, 以减小预测密度图与真实密度图的差距 L_{MSE} 。

$$L_{MSE} = \|D_F(x) - D_i(x)\|^2 \quad (3)$$

使用“弱监督”标记的图片无法获得其密度分布图 $D_i(x)$, 但可以获得其包含的人群数量 C_i , 因此可以利用“弱监督”信息对神经网络进行训练。使用“弱监督”方式标记的图片如图 3 所示。因此, 在“弱监督”的标记条件下, 本文使用每张图片包含的人群数目作为图片的“弱监督信息”。



图 3 使用“弱监督”方式对密集场景图片进行标记的示例图
Fig. 3 Example of annotating dense scene images with “weak supervision” method

使用“弱监督”标记的图片 X_w 生成预测密度图的方式与使用“强监督”标记的图片生成预测密度图的方式相似, 将原始场景图片送入网络即可获得:

$$D_w(x) = \Phi(X_w; \theta) \quad (4)$$

对预测密度图 D_w 中包含的行人数目与图片中包含的真实人群数目 C_i 进行比较。损失函数为:

$$L_{count} = \left| \int D_w(x) dx - C_i \right| \quad (5)$$

在密集场景人群统计问题中, 已有学者对图片中包含的人群数量进行了回归。与已有工作不同, 本文假定无法通过对每张图片进行“点标注”获得图片中每个人的位置。但是, 在连续的视频帧中, 本文通过为小部分图片进行点标记来获得密度分布图, 这些帧将被作为参考帧。以参考帧为基础, 利用后续视频帧中行人的行为, 来获得后续视频帧中包含的行人数目。这种获取数据的方法更加灵活。

4 实验及结果分析

4.1 实验数据集与评估标准

4.1.1 数据集

为了证明文中提出的“弱监督”学习算法的有效性, 本文在公开的上海科技大学 Shanghai Tech 数据集上进行了实验。Shanghai Tech 数据集一共包括 1 198 张密集场景图片。该数据集又被划分为两个子数据集: Shanghai Tech PartA 和 Shanghai Tech PartB。

Shanghai Tech PartA 一共包含 482 张图片(被标记好的行人累计有 241 667 个), 其中, 300 张图片作为训练集, 剩下的 182 张图片作为测试集。

Shanghai Tech PartB 一共包含 716 张图片(累计有 88 498 个被标记好的行人)。其中, 400 张图片作为训练集, 剩

下的 316 张图片作为测试集。

为了评估“弱监督”学习算法的有效性,随机选取少量样本进行“强监督”标记。对于这部分样本,可以获得其真实密度分布图。而对于更多的数据,则对其进行“弱监督”标记。对于这些样本,采用图片中包含的人群数目作为监督信息用于网络参数的学习。数据标记的方式参考图 2 和图 3。

4.1.2 评价指标

本文采用绝对平均误差 (Mean Absolute Error, MAE) 和绝对均方误差 (Mean Square Error, MSE) 对“弱监督”学习算法进行评估。

$$MAE = \frac{1}{N} \sum_{i=1}^N |Pre_i - GT_i|$$

$$MSE = \sqrt{\frac{1}{N} \sum_{i=1}^N |Pre_i - GT_i|^2}$$

其中, Pre_i 为预测密度图中所包含的行人数目, GT_i 为图片中真实包含的行人数目, N 为测试集中的图片样本数目。

4.2 实验配置

开发所使用的硬件设备为:PC 端为联泰集群 LT-5038A 工作站,配置了一块英特尔酷睿 i5-1650(12 核)处理器;运算显卡为 NVIDIA-1080Ti(12GB 显存)。使用的系统环境为 Ubuntu16.04;实验软件环境为 Anaconda3,pycharm2017,python2.7;采用的深度学习框架为 Pytorch 0.4.0。

4.3 实验细节

(1) 数据处理。对于数据集中的所有样本,为了保证其具有原始的分辨率,对于每一个 RGB 通道,按照 ImageNet 数据集^[14]的处理方法进行归一化处理。

(2) 网络结构。对于空洞卷积神经网络,图 1 已给出了大致的描述。在进入空洞卷积之前的网络部分,使用预训练的 VGG16 进行初始化;对于后续的空洞卷积层与密度回归层,采用均值为 0、标准差为 0.01 的正态分布进行初始化。除密度回归层外,每一个卷积层都会采用“Relu”操作来保证输出的特征图为非负值。在整个训练阶段,采用随机梯度下降法 (Stochastic Gradient Descent, SGD) 对网络参数进行优化。由于在空洞卷积神经网络中使用了 3 个最大池化层,因此回归出的密度图的分辨率为原始输入图片分辨率的 $\frac{1}{8} * \frac{1}{8}$ 。对于“强监督”标记的数据,由于预测密度图需要与真实密度图做欧氏距离度量,因此对这一部分的真实密度图进行了缩放处理,以保证预测密度图与对应的真实密度图都具有相同的分辨率。

(3) 参数设置。数据集遍历次数为 300,每一个小批量训练只使用 1 个样本。对于 SGD,初始化其学习率为 1×10^{-7} 。在训练过程中,每遍历 50 次样本,学习率减半。动量因子为 0.9。在 Shanghai Tech PartA 中,从训练集中随机选出 60 张图片作为验证集,用于评估模型的性能。对于剩下的 240 张图片,30 张图片作为“强监督”样本,以真实分布密度图作为监督信息;210 张图片作为“弱监督”样本,以图片中包含的行人数目为监督信息。而在 Shanghai Tech PartB 中,我们选取 80 张图片作为验证集,在余下的 320 张图片中选取 20 张图片以“强监督”信息进行标注,剩下的 300 张图片以图片中的人群数目进行标注。标注实例如图 2 和图 3 所示。

4.4 实验结果

4.4.1 实验方案概述

本文提出的方法基于“弱监督”学习,仅对非常有限的样本进行“强监督”标记,对大部分样本进行“弱监督”标记。为了证明所提算法的有效性,本文也报告了只对少部分数据使用传统训练方法的结果,这部分实验可以视为从原始数据集中随机抽选了一个子数据集对网络进行训练。本文主要是讨论一种“弱监督”标记下的学习方式,因此与传统的“强监督”学习算法相比并不公平,但文中也列出了一些使用“强监督”学习算法的结果以供参考。

(1) Baseline。在每次迭代中,只选取进行“强监督”标记的数据送入网络直接进行训练,损失函数如式(3)所示。

(2) 本文方案。在每次迭代中,选取所有数据送入网络进行训练。其中“强监督”标记的少量数据利用预测密度图与真实密度图的欧氏距离进行损失度量;而进行“弱监督”标记的数据则通过对预测密度图进行积分获得预测人数,损失函数如式(5)所示。与 Baseline 方法相比,本文方法利用了“弱监督”信息对人群分布特征进行学习。

4.4.2 实验结果分析

为了进一步评估利用“弱监督”信息对人群分布特征进行学习的有效性,本文选取了较小规模的训练集样本进行“强监督”标记,更多的数据采用“弱监督”标记。在两个基准数据库 Shanghai Tech PartA 和 Shanghai Tech PartB 上进行测试,结果如表 1 和表 2 所列。从表中的数据可知,引入使用“弱监督”信息标注的样本后,网络的泛化能力得到了较大提升。

表 1 Shanghai Tech PartA 测试集上各方法 MAE 与 MSE 的比较
Table 1 Comparison of MAE & MSE of different algorithms on testing set of Shanghai Tech PartA

方法类型	实验方法	骨干网络结构	MAE	MSE
弱监督	Baseline	空洞卷积网络	106.5	167.4
	本文方案	空洞卷积网络	89.3	135.8
强监督	MCNN ^[5]	多列神经网络	110.2	173.2
	Switch-CNN ^[6]	多列神经网络	90.4	135.0

表 2 Shanghai Tech PartB 测试集上各方法 MAE 与 MSE 的比较
Table 2 Comparison of MAE & MSE of different algorithms on testing set of Shanghai Tech PartB

方法类型	实验方法	骨干网络结构	MAE	MSE
弱监督	Baseline	空洞卷积网络	16.4	25.7
	本文方案	空洞卷积网络	13.5	19.9
强监督	MCNN ^[5]	多列神经网络	26.4	41.3
	Switch-CNN ^[6]	多列神经网络	21.6	33.4

表 1 和表 2 的数据结果表明,当“强监督”标记样本较少时,与使用整个原始训练集进行训练相比,仅使用这些“强监督”标记的样本进行训练会造成较大的性能损失。在公开文献报道中,对于 Shanghai Tech PartA,对训练集中所有图片进行“强监督”标记后,训练得到的模型在测试集上的 MAE 为 68.2;相应地,PartB 中公开报道的 MAE 为 10.6。

仅使用少部分“强监督”标记图片按传统方式训练会造成模型的性能损失,因此本文方法引入了大量使用“弱监督”标记的图片对网络进行训练。本文方法与 Baseline 相比,在

Shanghai Tech PartA 上, MAE 从 106.51 降到了 89.28; 而在 Shanghai Tech PartB 上, MAE 从 16.41 降到了 13.51。由此证明, 引入“弱监督”标记的图片对网络参数进行学习可以提升网络的泛化性能。

结束语 本文提出了一种解决密集场景人群计数问题中“强监督”图片受限问题的新方法, 主要贡献如下:

(1) 探索了只对少量样本进行“强监督”标记时, 模型泛化性能的变化;

(2) 提出了一种利用“参考帧”和后续帧中行人行为的变化来获得大量使用“弱监督”标记数据的方法;

(3) 在密集场景人群分布特征学习过程中, 引入大量“弱监督”标记的密集场景人群图片, 利用其弱监督信息提升网络对人群计数任务的泛化性能, 减少了对密集场景中行人的人为点标注。

同时, “弱监督”学习在密集人群计数问题中还有一些需要解决的问题:

(1) 本文提出了一种搜集“弱监督”标记数据的方法, 但是目前还没有使用这种搜集方法的公开数据集;

(2) 本文利用图片中包含的人群数目作为“弱监督”信息, 后续将挖掘密集场景计数问题中的其他信息来提升模型的性能。

参 考 文 献

- [1] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego: IEEE Press, 2005: 886-893.
- [2] VIOLA P, JONES M J. Robust Real-Time Face Detection[J] International Journal of Computer Vision, 2004, 57(2): 137-154.
- [3] FIASCHI L, KÖTHE U, NAIR R, et al. Learning to Count with Regression Forest and Structured Labels[C] // Proceedings of International Conference on Pattern Recognition. Tsukuba: IEEE Press, 2012: 2685-2688.
- [4] CHAN A B, VASCONCELOS N. Bayesian Poisson Regression for Crowd Counting[C] // Proceedings of IEEE International Conference on Computer Vision. Tokyo: IEEE Press, 2009.
- [5] ZHANG Y, ZHOU D, et al. Single-image Crowd Counting via Multi-column Convolutional Neural Network[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. LAS VEGAS: IEEE Press, 2016: 589-597.
- [6] SAM D B, SURYA S, BABU R V, et al. Switching Convolutional Neural Network for Crowd Counting[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.

Honolulu: IEEE Press, 2017: 5744-5752.

- [7] 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. Venice: IEEE Press, 2017: 1861-1870.
- [8] LI Y, ZHANG X, CHEN D. Csrnet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake: IEEE Press, 2018: 1091-1100.
- [9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C] // Proceedings of International Conference on Learning Representations. San Diego: IEEE Press, 2015.
- [10] LIU X, VAN DE WEIJER J, et al. Leveraging Unlabeled Data for Crowd Counting by Learning to rank[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake: IEEE Press, 2018: 7661-7669.
- [11] WANG Q, GAO J, et al. Learning from Synthetic Data for Crowd Counting in The Wild[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Long Beach: IEEE Press, 2019.
- [12] GOODFELLOW I, POUGET-ABADIE J, et al. Generative Adversarial Nets[C] // Proceedings of International Conference on Neural Information Processing Systems Lake Tahoe: MIT Press, 2014.
- [13] SAM D B, SAJJAN N N, et al. Almost Unsupervised Learning for Dense Crowd Counting[C] // Proceedings of American Conference on Artificial Intelligence. Honolulu: AAAI Press, 2019.
- [14] KRIZHEVSKY A, SUTSKEVER I, et al. Imagenet Classification With Deep Convolutional Neural Networks[C] // Proceedings of International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2012: 1097-1105.



LIU Yan, born in 1995, postgraduate. His main research interests include deep learning and crowd counting.



NING Qian, born in 1969, associate professor. Her main research interests include computer application and intelligent control.