基于自适应叠合分割与深度神经网络的人数统计方法

郭文生 包 灵 钱智成 曹万里
(电子科技大学信息与软件工程学院 成都 610054)

摘 要 基于监控视频的人数(人群)统计是人群行为的分析、资源的优化配置、现代安防、商业信息的采集以及智能管理等重要任务的基础,具有较高的研究意义与应用价值。近年来,数字图像处理技术以及深度学习理论的不断完善和发展,极大地促进了基于监控视频的人数统计的研究,但仍然无法很好地解决监控场景中人数统计准确率较低、高清图片耗时的问题。针对在待检对象尺度变化较大的情况下,基于对象检测的人数统计方法的准确率大幅下降的问题,提出一种基于自适应叠合分割与深度神经网络的人数统计方法。该方法的思想来源于注意力机制,同时充分利用了叠合分割块内人头对象的尺度信息和人数信息。实验结果表明,自适应叠合分割算法能够与现有深度神经网络对象检测模型相结合,并且相较于直接利用深度神经网络对象检测模型进行人数统计的方法,该结合方法可以大幅提高人数统计的准确率。

关键词 人数统计,自适应叠合分割,深度神经网络,对象检测,非最大值抑制

中图法分类号 TP391.41 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.08.041

People Counting Method Based on Adaptive Overlapping Segmentation and Deep Neural Network

GUO Wen-sheng BAO Ling QIAN Zhi-cheng CAO Wan-li
(School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China)

Abstract People counting based on surveillance camera is fundamental for analyzing behavior of counting, resource optimization and resource allocation, modern security and protection, collecting commerce information as well as intelligent management. Therefore, it has significant meaning of study and application value. Recently, technology of digital image processing and theory of deep learning are constantly improved and developed, extremely promoting the study of people counting based on surveillance camera. However, there exist some problems, such as low accuracy of people counting and time-consuming of high definition, which are unable to be solved. In the wide range of object scale, accuracy of people counting method based on object detection decreases significantly. Aiming at this problem, this paper proposed a people counting method based on adaptive overlapping segmentation and deep neural network. The idea of this method comes from attention mechanism, and makes full use of information of the scales and numbers of head object in overlapping segmentation. The experimental results show that the adaptive overlapping segmentation algorithm can combine existing object detection model based on neural network. What's more, compared with the method of counting people by directly using object detection model based on neural network, the combination algorithm of adaptive overlapping segmentation and deep neural network can greatly improve the accuracy of people counting.

Keywords People counting, Adaptive overlapping segmentation, Deep neural network, Object detection, NMS

1 引言

基于监控视频的人数(人群)统计是人群行为的分析、资源的优化配置、现代安防、商业信息的采集以及智能管理等重要任务的基础,具有较高的研究意义和应用价值。近年来,数字图像处理技术以及深度学习理论的不断完善和发展,极大地促进了基于监控视频的人数统计的研究,但仍然无法很好地解决监控场景中人数统计准确率较低、计算成本高和耗时多的问题。Kowcika 等[1]指出基于监控视频的人数统计面临如下亟待解决的问题:1)在高遮挡、高密度、同时存在静止和运动目标、背景复杂以及光照变化条件下,人数统计准确率较低;2)计算成本高且耗时长。

现有人数统计方法有:1)基于对象检测的方法。该方法能同时得到人数信息和位置信息,但对高遮挡、高密度人群以

到稿日期:2017-06-06 返修日期:2017-09-14 本文受国家自然科学基金(61272175,61572109),中央高校基本业务费(ZYGX2015J066)资助。
郭文生(1976—),男,博士,副教授,CCF 会员,主要研究方向为人工智能、嵌入式系统及应用、软件测试、形式化方法等,E-mail:gws@uestc.edu.cn(通信作者);**包 灵**(1993—),男,硕士生,主要研究方向为深度学习、计算机视觉;**钱智成**(1995—),男,主要研究方向为深度学习;**曹万里**(1995—),男,主要研究方向为深度学习。

及待检对象尺度变化较大的情形，漏检率较高。Venkatesh 等[2]通过图片梯度信息和背景相减技术获取感兴趣点，并利用 Adaboost 分类器对子窗口进行分类以实现人头对象检测。该方法提高了获取感兴趣点的质量，但误检率较高且无法解决人群遮挡的问题。Tuan-Hung 等[3]提出了上下文相关的基于 CNN 的人头对象检测模型，在 TVHI 和 Casablanca 数据集上不仅提高了准确率，而且减少了时间消耗。2)基于人数回归的方法。该方法利用前景像素或感兴趣点的特征(如纹理、梯度等)与人数的对应关系实现人数统计，适用于高遮挡、高密度人群的人数统计，但现有方法的准确率均不高，常用的人数回归方法包括线性回归、高斯过程回归以及神经网络等。Zhang 等[4]提出了 MCNN 模型，用于将单张图片映射为密度图，同时进行人数统计，该模型可以估计任意密度及不同视角图片的人数。3)基于目标跟踪的方法。该方法利用多帧图片或目标轨迹信息实现人数统计，可以有效解决人群遮挡问题，但具有计算耗时的缺点。Li 等[5]实现了一种基于人头对象检测与跟踪的人数统计方法，在垂直监控以及人群密度较低的情形下能获得较高的准确率。因此，在人群存在高遮挡、高密度以及待检人头对象尺度变化较大时，现有人数统计方法的准确率均不高。

针对待检对象尺度变化较大的人数统计任务，基于深度神经网络对象检测的人数统计方法具有准确率大幅下降的缺点。对象检测的框架主要包括区域提取、特征提取、分类器、回归器 4 个部分，现有文献针对框架的不同部分提出了各种改进方法。Faster R-CNN[6]基于注意力(Attention)机制提出了 RPN 区域提取模型，并将该模型和特征提取模型相结合以实现特征共享，这不仅提高了区域提取的质量，而且提高了目标检测精度和时间效率，但难以检测小尺寸对象。Zeng 等[7]提出了一种双向门模型 GBD-Net 以利用图片局部区域和上下文信息，该方法可以有效提高小尺寸对象的检测率。为了降低对象检测的时间消耗，YOLO[8]和 SSD[9]通过直接从预定义的滑动窗口中估计对象类别，可以完成实时对象检测。因此，现有提高对象检测准确率的方法主要包括提高区域提取的质量、充分利用图片的上下文信息、设计高质量的特征提取模型等。这些方法能够解决图片解析度较低、小尺寸对象难检测的问题，但无法解决待检对象尺度变化较大从而导致人数统计准确率大幅降低的问题。

现有完成多尺度问题的方法主要包括获取高质量的区域提取以及改变输入图片的尺度。传统对象检测算法的输入图片固定不变，采用不同尺度的区域提取来支持不同尺度对象的检测；基于 CNN 的对象检测模型完成多尺度对象检测的方法是将区域提取算法整合到模型中，然后固定滑动窗口大小，通过改变输入图片的尺寸来实现。然而，改变整张图片的尺度时，因无法确定适当的尺度或缩放比例，而无法解决人数统计过程中待检测对象尺度变化较大的问题。

本文的思想来源于如下事实：当人们查看图片时，首先需要定位包含对象的区域，然后适当缩放该区域图片以保证可识别度，最后基于该区域内的对象完成更详细的任务，如人数统计、语义分割等，即注意力机制的思想。针对待检对象尺度变化较大从而导致人数统计准确率大幅下降的问题，本文利用分割块内待检对象尺度信息和人数信息，分别确定叠合区域参数以及叠合分割终止条件，自适应地对图片进行叠合分割，提出了一种自适应叠合分割与深度神经网络对象检测模型相结合的人数统计方法，能够将计算资源引导到可能包含对象的子区域。为了验证本文方法的有效性，我们制作了 LargeScale 数据集，该数据集包含 1150 张图片，其中训练与验证数据集 1000 张，具有人头区域尺度变化小且人数少的特点；测试数据集 150 张，具有人头区域尺度变化大且人数多的特点。通过实验表明，该方法不仅大幅提高了单张图片的人数统计准确率，而且在保证准确率的前提下，相对于区域叠合等分分割与深度神经网络对象检测模型相结合的方法，降低了人数统计的时间消耗。为进一步验证本文提出的自适应叠合分割算法与现有深度神经网络对象检测模型相结合的有效性，我们将自适应叠合分割算法与比较常用的对象检测模型 YOLO，SSD，Faster RCNN 进行结合以实现人数统计。实验结果表明，相较于直接利用对象检测模型进行人数统计的方法，结合后方法的准确率有较大幅度的提升。

2 自适应叠合分割思想

2.1 区域缩放方法

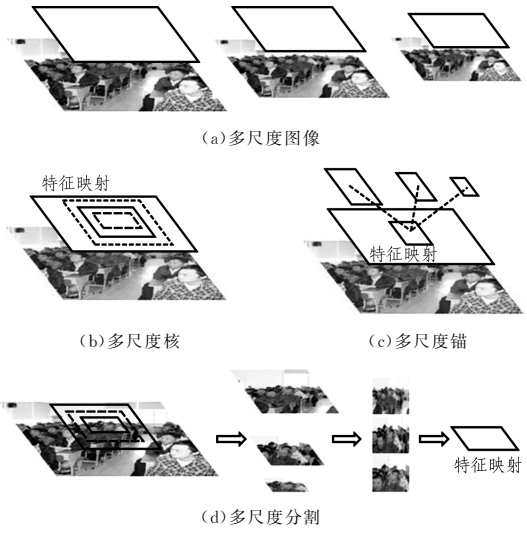
为了处理不同尺度的对象检测问题，Overfeat[10]和 SPP-Net[11]将原始图片转换成不同尺度的图片，并利用区域提取算法(如 Selective Search[12]或 EdgeBox[13])提取高质量的区域提取，然后使这些区域提取独立地通过 CNN 模型，最终把这些不同尺度的图片的检测结果进行综合汇总以完成对象检测任务。该处理方法的本质是对整张图片进行缩放以适应不同尺度范围的检测对象。随后的研究将区域提取整合到神经网络模型内，并利用不同尺度的输入图片和锚(Anchors)描述不同尺度的检测对象。基于该思想，Lu 等[14]针对 Faster RCNN 固定锚策略的 RPN 区域提取模型的不足，提出基于邻域信息及自适应缩放锚策略的 AZ-Net 模型来代替原有 RPN 模型；Xia 等[15]针对待检对象尺度变化较大和对象难以精确定位的问题，提出了能自动适应对象及其组成部分尺度的 HAZN 模型，即该模型能以不同的尺度值自适应地处理图片的不同区域，故而不需要浪费计算资源来缩放图片。因此，针对同一图片中待检对象尺度变化较大的问题，上述处理方法存在以下局限性。

1)传统对象检测算法的输入图片大小固定不变，通过不同尺度的滑动窗口来支持多尺度对象检测，因此该方法严重依赖训练数据以及高质量的区域提取算法。

2)基于 CNN 的对象检测模型支持多尺度对象检测的方法是固定滑动窗口的大小，通过改变输入图片的尺寸来实现。然而，改变整张图片的尺度，即进行一定程度的缩放，因无法确定适当的尺度或缩放比例而具有一定盲目性；除此之外，大多数已训练的目标检测模型(如 Faster R-CNN)，虽然可以输入任意大小的图片，但是受训练集图片大小的限制，过大或过小的缩放图片对对象检测性能的提升很有限，并且放大图片会增加额外的时间消耗。

3）输入图片经过对象检测模型的卷积、池化层后，用于分类和对象检测的特征映射层会被缩小，即不同尺度的锚只能处理待检对象尺度在很小范围内变化的情况。同时，由于受到特征映射层最小感受野的限制，若不对原始图片进行适当缩放或减少特征映射层对原始图片的缩小程度，将无法检测和识别小尺寸对象。

原始图片中不同区域的待检对象具有不同的尺度变化范围，因此需要针对不同区域进行不同程度的缩放。对比现有解决多尺度问题的方法，如：多尺度图片[10]（见图 1(a)）、多尺度核[11]（见图 1(b)）以及多尺度锚[6]（见图 1(c)），本文提出的多尺度分割（见图 1(d)）能针对图片的不同区域进行不同程度的分割，并利用已训练对象检测模型需输入固定尺寸图片的限制（如：Faster R-CNN 输入图片的短边尺寸为 600 像素，YOLO 为 448×448 像素，SSD 为 300×300 像素），实现图片不同区域的缩放。但盲目地对图片进行分割存在以下两方面的缺点：1）容易将图片中的待检对象切碎；2）在图片中待检对象尺度变化不大的区域并不需要进行过多的分割，因为分割尺寸变小将使得分割块数量增加，不仅会增加整张图片的检测时间，而且会降低人数统计的准确率。



(a)多尺度图像　(b)多尺度核　(c)多尺度锚　(d)多尺度分割

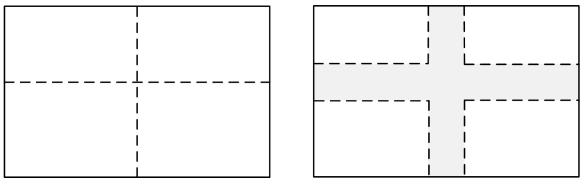
图 1　多尺度支持

Fig. 1　Multi-scale support

2.2　多尺度支持方法的对比

Overfeat[10] 和 SPP-Net[11] 将原始图片缩放成不同的大小以适应不同尺度的对象检测；Faster R-CNN[6] 通过将 RPN 区域提取模型整合到 Fast R-CNN 模型内来实现端到端的对象检测，同时由于 RPN 区域提取模型产生的提取区域相较于传统的 Selective Search[12] 或 EdgeBox[13] 具有更高的质量，因此其可以在一定程度上适应不同尺度的对象检测并降低时间消耗。针对待检测图片是否采用缩放实现人数统计，实验 1 利用 Faster R-CNN 模型在 LargeScale 测试数据集上进行了对比，实验结果表明，在缩放条件下，Faster R-CNN 可以提升 2.36% 的准确率。

考虑待检测图片中因不同区域远近景关系不一致而导致的对象尺度变化较大的问题，一个很自然的做法是基于缩放的思想，针对不同区域进行不同程度的缩放以适应对象检测模型的检测。在实验 1 中，我们对图片进行了不同等级的等分分割，图 2(a)给出了 2×2 等分分割的示意图；然后将这些等分分割块直接送入 Faster R-CNN 模型进行人头对象检测并汇总各分割块人数，以得到最终的统计结果。实验表明，对待检测图片进行等分分割能显著提高人数统计准确率，但随着等分数的增加，误检率以及检测时间也均随之增加。因此，如何降低误检率以及检测时间是图片等分分割方法最需要解决的问题。2.3 节将对该问题进行详细分析与研究，并提出自适应叠合分割算法。



(a)等分分割　(b)等分叠合分割

图 2　等分分割示意图

Fig. 2　Sketch map of equal segmentation

2.3　自适应叠合分割思想的提出

通过分析实验 1 的结果发现，误检率以及检测时间增加的主要原因包括两个方面：1）等分数的增加使分割块尺寸变小，从而造成待检测图片中的人头对象区域被切碎；2）等分数的增加使分割数目增加，从而造成 Faster R-CNN 对象检测模型的检测次数也随之增加。

针对误检率以及检测时间增加的原因的第一方面，本文提出了图片叠合分割的思想，从而有效避免了人头区域被切碎的问题。实验 1 在等分的基础上加入区域叠合思想，图 2(b)给出了 2×2 区域等分叠合分割示意图。实验结果表明，相较于等分分割方法，等分叠合分割方法能有效降低人数统计的误检率。

不同待检测图片的远近景关系存在差异，使得不同图片中人头对象的尺度变化的程度不同。因此，对待检测图片进行区域叠合分割时，需要利用待检测图片不同区域的人头对象尺度信息，确定待检测图片不同区域的区域叠合参数以及分割块大小；同时也需要利用该区域人数信息确定叠合分割区域的分割终止条件。本文针对待检测图片等分分割数过大导致的误检率以及检测时间增加的问题，提出了一种自适应确定叠合分割参数以及分割块大小的待检测图片分割算法，该算法被称为自适应叠合分割。第 3 节将对自适应叠合分割算法的设计与实现进行详细阐述。

3　算法设计与实现

3.1　总体流程

本文提出的基于自适应叠合分割与深度神经网络对象检测模型的人数统计方法的总体流程主要包括自适应叠合分割算法、深度神经网络对象检测模型、检测结果集合以及结果汇总算法等几个部分。其中，自适应叠合分割算法利用待检人头对象的尺度信息和人数信息自适应地分割图片；深度神经网络对象检测模型对叠合分割块进行人头对象检测；检测结果集合用于存放各叠合分割块的人头对象检测结果；结果汇

总算法用于对最终的检测结果集进行去重及汇总处理。详细流程如图 3 所示。

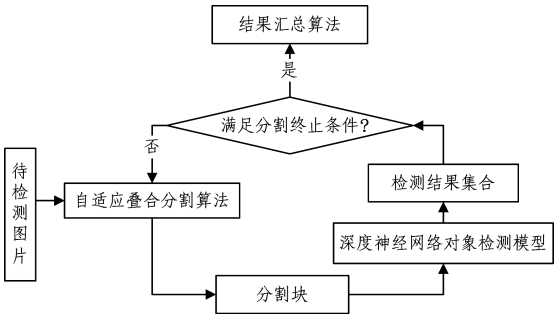


图 3 总体流程

Fig. 3 Overall flow of process

3.2 自适应叠合分割算法的设计

自适应叠合分割算法的思想来源于注意力机制。本文利用已训练对象检测模型需输入固定尺寸的图片的条件,将自适应叠合分割得到的区域分割块与区域缩放结合起来,即在区域分割块被送入对象检测模型时,如果区域分割块的尺寸小于模型固定的输入尺寸,则该区域分割块被放大且区域分割块尺寸越小其放大的程度越大,反之则该区域分割块被缩小且区域分割块尺寸越大其缩小的程度也越大。本文基于 2.3 节的自适应叠合分割的思想,提出了自适应叠合分割算法,以解决以下问题:1)当对象检测模型的输入为固定尺寸的图片时,直接缩放待检测图片存在无法根据特定区域进行不同程度的缩放以及过度放大图片不能提升人数统计准确率的缺点;2)对待检测图片进行等分分割易将待检测人头对象区域切碎,从而使误检率升高;3)在一定程度上对待检测图片进行等分叠合分割虽能避免待检测人头对象区域被切碎,但无法根据图片不同区域的人头对象尺度变化程度信息进行自适应叠合分割。自适应叠合分割算法如算法 1 所示。

算法 1 自适应叠合分割算法

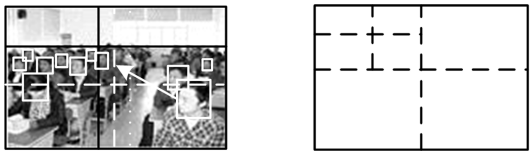
输入:待检测图片 P_1

输出:人头对象检测框集合 $H=\bigcup_{k=1}^{D}\bigcup_{l=1}^{L_k}\bigcup_{m=1}^{M_l}h_k^m$(其中,D 是深度;$L_k$ 是第 k 层含有的叠合分割块数目,最大为 4^{k-1};M_l 是第 l 个叠合分割块中所检测到的人头对象数目;$h_k^m\{x_{center},y_{center},w,h,score\}$为人头对象检测框)

参数:可识别为人头对象的最小阈值 λ,最大切割深度 D(默认为 4),叠合分割参数 $O(p^0)=\{w,h\}$(默认为$\{P_1$ 宽/4,P_1 高/4$\}$)

1. 初始化 $H=\varnothing$,$d=2$(当前分割深度);
2. 将图片 P_1 送入对象检测神经网络模型进行人头对象检测,并将检测框 score 大于阈值 λ 的人头对象检测框 $Det(P_1)=\bigcup_{l=1}^{L_1}\bigcup_{m=1}^{M_l}h_1^m$ 加入集合 H;
3. 如果 $Det(P_1)=\varnothing$,则将叠合分割参数 $O(p^1)$设置为 $O(p^0)$,否则将叠合分割参数 $O(p^1)$设置为 $Max\{Det(P_1)\}$;
4. 结合叠合分割参数 $O(p^1)$对 P_1 进行初始四分叠合分割(横切割线为 P_1 纵向 1/3 处,纵切割线为 P_1 横向 1/2 处),得到叠合分割块集合 $P_2=\{p_1^1,\cdots,p_1^{L_2}\}$,并将其分别送入对象检测神经网络模型进行人头对象检测,将阈值大于 λ 的人头对象检测框 $Det(P_2)=\bigcup_{l=1}^{L_2}\bigcup_{m=1}^{M_l}h_2^m$ 加入集合 H;
5. For $d<D$
6. For $p^d\in P_d$
7. 将 p^d 送入对象检测神经网络模型进行人头对象检测,并将阈值大于 λ 的人头对象检测框 $Det(p^d)=\bigcup_{m=1}^{M_l}h_d^m$ 加入集合 H;
8. If $Det(p^d)\neq Det(P_d^d)$
// $Det(P_d^d)$是叠合分割块 p^d 在 P_d 所对应的区域在第 d−1 深度检测到的人数
9. 如果 $Det(p^d)=\varnothing$,则将叠合分割参数 $O(p^d)$设置为$\{p^d$宽/4,p^d高/4$\}$,否则将叠合分割参数 $O(p^d)$设置为 $Max\{Det(p^d)\}$;
10. 结合叠合分割参数 $O(p^d)$对 p^d 进行四分叠合分割,得到 4 个叠合分割块并将其加入分割块集合 P_{d+1};
11. Endif
12. Endfor
13. $d=d+1$;
14. Endfor

该算法的示意图如图 4 所示,其中图 4(a)是利用待检测图片的人头对象尺度变化程度信息确定叠合分割参数,箭头所指区域为叠合中心所在位置,其尺寸为待检测原图送入对象检测模型所检测出的最大人头对象区域的尺寸。图 4(b)是自适应叠合分割示意图,为了使绘图简洁,图中虚线代表自适应叠合分割的叠合区域;在自适应确定叠合分割块大小时,使用了注意力机制思想以及不同区域的人数信息,即待检测图片中人数较多的区域需要运用注意力机制,将计算资源引导到该区域。具体做法为:通过人数信息设置自适应叠合分割区域的分割终止条件,对满足分割条件的自适应叠合区域进行图 4(b)左上角所示的四分分割。需要注意的是,对待检测图片进行第一次四分分割时,考虑到实际监控场景和摄像头所在位置的限制,图片最上面区域包含的人数较少并且大致属于远景区域。因此,在确定横切割线时应使图片上部区域小于下部区域,本文选取图片纵向 1/3 处为该横切割线所在位置。



(a)获取叠合分割参数 (b)自适应切割

图 4 自适应叠合分割示意图

Fig. 4 Sketch map of adaptive overlapping segmentation

3.3 结果汇总算法

由自适应叠合分割算法得到的人头对象检测框集合 H 中存在大量重复的人头对象检测框,因此需要利用非极大值抑制(NMS)算法对其进行处理。非极大值抑制算法的本质是在人头对象检测框集合 H 中搜索局部极大值并且抑制非极大值元素,从而达到消除多余交叉重复窗口以找到人头对象最佳检测框的目的。

4 实验与分析

4.1 数据集

由于现有数据集不完全适用于本研究工作中针对待检对

象尺度变化较大情况下人数统计任务的评估，因此我们建立了 LargeScale 数据集，并利用该数据集对所提出的自适应叠合分割算法进行评价。

为了保证 LargeScale 的可用性，该数据集参考 Pascal Voc 以及 ImageNet 目标检测数据集标签的制作方式对图片中的人头对象区域进行标签制作。LargeScale 由训练数据集和测试数据集两部分组成，图片数据全部来源于网上图片或视频，主要涉及教室、会议及广场等场景。其中，训练集和验证集包含 1000 张图片，图中人头尺度变化较小且多数图片的人数不超过 10；测试集包含 150 张图片，图中人头尺度变化较大且多数图片的人数大于 80。

4.2　模型训练

本实验利用 GTX1060 显卡在 LargeScale 训练数据集上对 YOLO，SSD，Faster R-CNN 模型进行训练，初始参数的设置主要参考文献[6，8-9]。由于 YOLO 和 SSD 的训练相对简单，此处仅对 Faster R-CNN 模型训练相关的内容进行简要阐述，主要包括样本参数、损失函数、模型参数初始化、训练参数、数据增广以及模型训练等几个部分。Faster R-CNN 模型主要由共享特征网络层、区域提取 RPN 模型、ROI 池化层以及分类回归等部分组成，如图 5 所示。

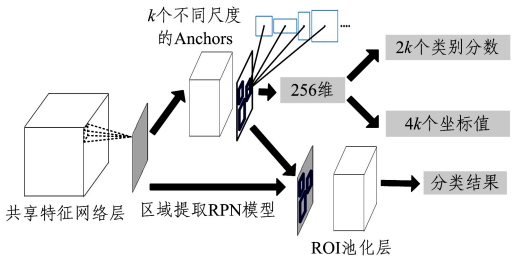


图 5　Faster R-CNN 模型

Fig. 5　Model of Faster R-CNN

1）样本参数。对于训练样本集中的每张图片，利用真值候选区域与 Faster R-CNN 区域提取 RPN 模型所产生的 Anchor 的重叠比例来进行训练集正负样本的提取，重叠比例最大的 Anchor 或重叠比例大于 0.7 的 Anchor 为正样本，小于 0.3 的 Anchor 为负样本，跨越图片边界的 Anchor 或重叠比例大于 0.3 且小于 0.7 的 Anchor 舍弃不用。

2）损失函数。与文献[6]相同，分类误差利用对数损失函数，预测窗口偏差利用 Fast RCNN 中提出的平滑 L1 损失，即损失函数同时最小化分类误差与预测窗口偏差。

3）模型参数初始化。共享特征网络层参数利用在 ImageNet 上已训练的 VGG16 模型进行参数初始化，其余新增网络层参数从标准差为 0.01 的零均值高斯分布中随机抽取。

4）训练参数。采用了 Momentum 优化方法，动量因子取 0.9，权重衰减因子取 0.0005；每个小批量（mini-batch）包含从一张图片中提取的 256 个锚（Anchors），正负样本比例为1∶1；首先在学习率为 0.001 的条件下进行 50000 次迭代，然后在学习率为 0.0001 的条件下进行 30000 次迭代。

5）数据增广。LargeScale 训练集的数据相对较少，为防止数据不足而导致过拟合问题，需要对训练数据进行增广。我们对图片进行随机加噪、改变亮度、尺度缩放以及水平翻转等处理以获得更多训练数据，最终将训练集增广到 10000 张图片。

6）模型训练。Faster R-CNN 模型的训练主要包括 4 个阶段：第一阶段，使用在 ImageNet 数据集上预训练的 VGG16 模型对共享特征网络层参数进行初始化，其余新增网络层参数从标准差为 0.01 的零均值高斯分布中随机抽取，再利用 LargeScale 训练数据集对区域提取 RPN 模型进行训练；第二阶段，同样使用在 ImageNet 数据集上预训练的 VGG16 模型对共享特征网络层参数进行初始化，然后利用第一阶段已训练的区域提取 RPN 模型对每一张训练图片进行正负样本提取，最后将这些样本用于共享特征网络层参数的训练；第三阶段，使用第二阶段训练好的共享特征网络层参数重新初始化区域提取 RPN 模型的共享特征网络层参数，然后固定共享特征网络层参数并对区域提取 RPN 模型进行微调；第四阶段，将第三阶段中区域提取 RPN 模型的卷积层参数固定并使用该模型提取正负样本，然后对共享特征网络层参数进行微调。经过上述 4 个阶段实现了共享特征网络层与区域提取 RPN 模型的交替训练，同时在每次训练过程中需要冻结一部分参数，然后训练另一部分参数，这样可以同时提高区域提取 RPN 模型以及 Faster R-CNN 分类回归的整体性能。

4.3　性能评价指标

针对人数统计方法的常用性能评价指标主要包括：平均绝对误差（MAE）[2]、平均相对误差（MRE）[2,5]、平均均方误差（MSE）[5]以及平均准确率（MPR）等。同时，有别于基于人数回归以及目标跟踪的方法，基于对象检测的人数统计方法能同时得到人数及其位置信息，因此，也可以使用平均误检率（MFPR）对基于对象检测的人数统计方法进行评价。

本文采用平均准确率以及平均误检率对相关人数统计方法进行性能评价，两种指标的定义如下：

$$MPR=\frac{1}{N}\sum_{i=1}^{N}\frac{D(i)-D_E(i)}{T(i)} \tag{1}$$

$$MFPR=\frac{1}{N}\sum_{i=1}^{N}\frac{D_E(i)}{T(i)} \tag{2}$$

其中，$D(i)$代表在第 i 张图片中所检测到的人数，$D_E(i)$代表 $D(i)$中被误检的人数，$T(i)$代表第 i 张图片中实际存在的人数，N 代表总图片数。

4.4　实验 1

不同图片中不同区域的远近景关系存在差异，导致人头尺度不同，现有解决多尺度问题的方法包括多尺度图片、多尺度核以及多尺度锚，其中用于多尺度图片的方法之一是保持区域提取模型的尺度模板不变，然后对图片进行缩放。Faster R-CNN 对象检测模型是多尺度锚（见图 1（b））的代表方法；Faster R-CNN＋缩放是多尺度锚与多尺度图片（见图 1(a)）相结合的人数统计方法；本文涉及到的 Faster R-CNN＋等分分割、Faster R-CNN＋等分叠合分割与 Faster R-CNN＋自适应叠合分割均是多尺度锚与多尺度分割（见图 1(d)）相结合的人数统计方法。

在人数统计任务中，为了比较 Faster R-CNN、Faster R-

CNN＋缩放、Faster R-CNN＋等分分割、Faster R-CNN＋等分叠合分割与 Faster R-CNN＋自适应叠合分割 5 种人数统计方法的性能，本实验在 LargeScale 测试数据集上对这 5 种方法进行了测试。接下来将分别对各方法的实验设置以及实验结果进行阐述。

1)Faster R-CNN＋缩放

Faster R-CNN＋缩放相结合的方法用于考查图片缩放是否能够提升 Faster R-CNN 模型的识别准确率。该方法对图片进行等比例缩放，然后将缩放图片依次送入 Faster R-CNN 模型中进行检测，最后将各缩放图片的检测结果进行去重处理，得到最终检测结果。图片缩放的取值范围为 0.1～2.0，取值间隔为 0.05，因此，总共可以得到 38 张缩放图片。相较于直接利用 Faster R-CNN 进行人头对象检测的人数统计方法，Faster R-CNN＋缩放相结合的方法能提升 2.36％的人数统计准确率，结果如表 1 所列。

2)Faster R-CNN＋等分分割

针对不同图片、不同区域的人头对象尺度变化程度不同的问题，使用 Faster R-CNN＋等分分割相结合的方法对人头对象进行检测和识别的实验。实验过程中，首先对待检测图片进行等分分割，如图 2(a)所示，除了图 2(a)所示的 2×2 等分分割外，还包括 3×3，4×4，…，14×14 几种等分分割方式；然后，将这些等分分割块分别送入 Faster R-CNN 模型中进行人头对象检测并进行去重处理，从而得到最终的检测结果。实验结果表明，当等分分割数为 4×4 时在 LargeScale 测试数据集上能够达到最佳检测识别率，如图 6 和表 1 所示。

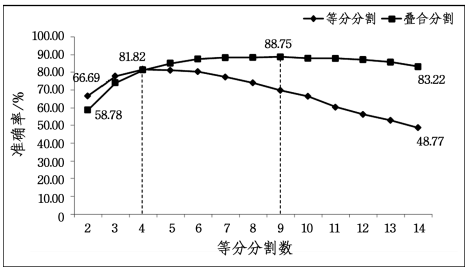


图 6 等分数与准确率的关系

Fig. 6 Relation between numbers of segmentation and accuracy

3)Faster R-CNN＋等分叠合分割

针对图片等分分割存在待检对象被切碎的缺点，使用了 Faster R-CNN＋等分叠合分割相结合的方法对人头对象进行检测识别的实验。图片的等分分割与等分叠合分割的不同在于是否对分割线区域进行叠合处理，如图 2 所示。在实验过程中，首先对待检测图片进行等分叠合分割，如图 2(b)所示，等分叠合分割方式与 Faster R-CNN＋等分分割中所设置的分割方式相同；然后将这些等分叠合分割块分别送入 Faster R-CNN 模型中进行人头对象检测并进行去重处理，以得到最终的检测结果。实验结果表明，等分分割数为 9×9 时在 LargeScale 测试数据集上能够达到最佳检测识别率，如图 6 和表 1 所示。

4)Faster R-CNN＋自适应叠合分割

对图片进行等分分割能大幅提高待检测图片的识别率，但不同的等分分割方式得到的结果存在差异，并且不同图片达到最优检测识别率的等分分割方式也不相同。例如，待检测图片中人头对象尺度变化较大，则达到最优检测识别率的等分数越大，反之则等分数越小。实验过程中，为了探求分割块大小与 Faster R-CNN 模型检测识别率之间的关系，对图片进行等大小分割，主要包括 100×100，150×150，…，800×800 等几种分割块大小。从实验结果可以发现，随着分割块尺寸的变小，分割块数量增加，Faster R-CNN 模型进行人头对象检测的准确率先缓慢增加到最大值然后快速减小，而误检率和检测识别时间则会一直增加，如图 7 所示。

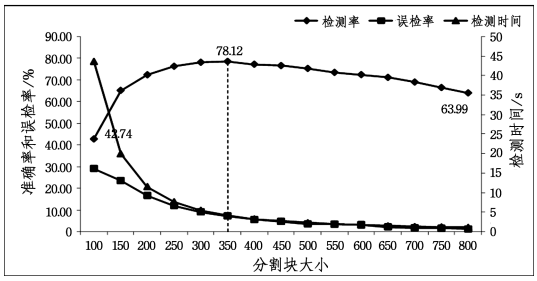


图 7 分割块大小与 Faster R-CNN 识别率的关系

Fig. 7 Relation of recognition rate between segmentation block size and Faster R-CNN

为了解决在待检对象尺度变化较大的情况下，基于深度学习的目标检测方法的准确率大幅下降的问题，本文实验分析了等分数与分割块大小对人数统计准确率的影响，实验结果如图 6、图 7 所示，并在该实验结果的基础上提出了自适应叠合分割算法。为了验证该算法的有效性，采用 Faster R-CNN＋自适应叠合分割相结合的方法对人头对象进行了检测识别的实验。可以发现，相较于直接利用 Faster R-CNN 进行人头对象检测，该方法能提升 44.66％的准确率，并且相较于 Faster R-CNN＋叠合等分分割的方法，误检率与耗时均得到了降低，实验结果如表 1 所列。

表 1 多尺度解决方法的对比

Table 1 Comparison of multi-scale solution methods

人数统计方法	准确率/％	误检率/％	耗时/s
Faster R-CNN	42.58	0.60	0.40
Faster R-CNN＋缩放	44.94	1.55	12.53
Faster R-CNN＋等分分割	81.82	4.32	1.91
Faster R-CNN＋叠合等分分割	88.75	10.15	8.81
Faster R-CNN＋自适应叠合分割	87.24	4.76	8.01

4.5 实验 2

为进一步探求本文提出的自适应叠合分割算法与其他目标检测模型相结合的有效性，本实验在 LargeScale 测试数据集上，对 YOLO、SSD、YOLO＋自适应叠合分割、SSD＋自适应叠合分割以及 Faster R-CNN＋自适应叠合分割等几种人数统计方法的性能进行了实验对比。

实验结果表明，YOLO＋自适应叠合分割、SSD＋自适应叠合分割相结合的方法相较于直接利用 YOLO 和 SSD 进行对象检测的方法，检测率均有大幅度提升。同时还可以发现，Faster R-CNN＋自适应叠合分割相结合的方法能获得较高的准确率，但耗时也最多；而 SSD＋自适应叠合分割相结合

的方法耗时较少，并且准确率仅次于 Faster R-CNN＋自适应叠合分割相结合的方法；YOLO＋自适应叠合分割相结合的方法的准确率最低，但耗时最少。具体如表 2 所列。

表 2　不同人数统计方法的对比

Table 2　Comparison of different people counting methods

人数统计方法	准确率/%	误检率/%	耗时/s
YOLO	25.84	1.24	0.09
SSD	38.12	0.16	0.27
YOLO＋自适应叠合分割	76.59	37.92	1.93
SSD＋自适应叠合分割	86.49	2.26	6.15
Faster R-CNN＋自适应叠合分割	87.24	4.76	8.01

5　实验结论

在 LargeScale 测试数据集上对 Faster R-CNN、Faster R-CNN＋缩放、Faster R-CNN＋等分分割、Faster R-CNN＋等分叠合分割与 Faster R-CNN＋自适应叠合分割 5 种人数统计方法进行了实验对比，证实了本文提出的 Faster R-CNN＋自适应叠合分割相结合的方法能够大幅提升待检对象尺度变化较大情况下的人数统计的准确率。为进一步验证本文提出的自适应叠合分割算法的有效性，对 YOLO，SSD，YOLO＋自适应叠合分割、SSD＋自适应叠合分割以及 Faster R-CNN＋自适应叠合分割等几种人数统计方法的性能进行了实验对比，实验结果表明，相比于直接利用目标检测模型进行人头对象检测的方法，目标检测模型＋自适应叠合分割相结合的方法的人数统计准确率有大幅提升。

结束语　用于解决多尺度问题的方法主要包括获取高质量的区域提取以及改变输入图片的尺度，常用的多尺度支持方法包括多尺度图片、多尺度核、多尺度锚以及本文提出的多尺度分割。针对待检对象的尺度变化较大导致人数统计准确率大幅下降的问题，本文基于注意力机制的思想，充分利用了叠合分割块内待检对象尺度信息和人数信息，提出了一种自适应叠合分割与深度神经网络对象检测模型相结合的人数统计方法，能够将计算资源引导到可能包含对象的子区域，从而达到提高待检测图片人头对象的检测率以及减少检测时间的目的。相较于直接利用目标检测模型进行对象检测的方法，目标检测模型＋自适应叠合分割相结合的方法具有耗时比较严重的缺点。针对此问题，我们将在下一步研究工作中考虑把自适应叠合分割算法融入到对象检测模型中，通过减少不必要的重复计算和对象检测次数，来解决人数统计耗时的问题。

参 考 文 献

[1] KOWCIKA A. A Literature Study on Crowd (People) Counting With the Help of Surveillance Videos[J]. International Journal of Information Technology & Decision Making, 2015, 3(4): 2353-2361.

[2] SUBBURAMAN V B, DESCAMPS A, CARINCOTTE C. Counting people in the crowd using a generic head detector[C]// 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS). IEEE, 2012: 470-475.

[3] VU T H, OSOKIN A, LAPTEV I. Context-Aware CNNs for Person Head Detection[C]// IEEE International Conference on Computer Vision. IEEE, 2015.

[4] ZHANG Y, ZHOU D, CHEN S, et al. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 589-597.

[5] LI B, ZHANG J, ZHANG Z, et al. A people counting method based on head detection and tracking[C]// 2014 International Conference on Smart Computing (SMARTCOMP). IEEE, 2014: 136-141.

[6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[7] ZENG X, OUYANG W, YAN J, et al. Crafting gbd-net for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(99): 1-9.

[8] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Computer Vision and Pattern Recognition. IEEE, 2016: 779-788.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]// European Conference on Computer Vision. Springer, Cham, 2016: 21-37.

[10] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. arXiv preprint arXiv: 1312.6229, 2013.

[11] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916.

[12] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[13] ZITNICK C L, DOLLÁR P. Edge boxes: Locating object proposals from edges[C]// European Conference on Computer Vision. Springer, Cham, 2014: 391-405.

[14] LU Y, JAVIDI T, LAZEBNIK S. Adaptive object detection using adjacency and zoom prediction[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2351-2359.

[15] XIA F, WANG P, CHEN L C, et al. Zoom better to see clearer: Human and object parsing with hierarchical auto-zoom net[C]// European Conference on Computer Vision. Springer, Cham, 2016: 648-663.