Journal of Image and Signal Processing 图像与信号处理, 2020, 9(4), 202-210
Published Online October 2020 in Hans. http://www.hanspub.org/journal/jisp
https://doi.org/10.12677/jisp.2020.94024

Hans汉斯

监控视频密集人群的人数统计系统设计

沈礼文[1]，邱　钊[1*]，彭贵超[2]，黄　萍[3]，李　超[1]，蔡金晔[1]

[1]海南大学计算机与网络空间安全学院，海南 海口
[2]国家计算机网络应急技术处理协调中心海南分中心，海南 海口
[3]海南大学理学院，海南 海口
Email: 269774968@qq.com, *5454734@qq.com, 446419342@qq.com, 1197648483@qq.com

收稿日期：2020年9月9日；录用日期：2020年10月2日；发布日期：2020年10月9日

摘　要

随着社会的不断发展，人群密集的场所随处可见。对监控视频下的人员进行统计分析，实现人数统计算法，可以为城市公共资源优化配置、安保人员调度、安全管理等提供有效的技术手段。本文基于YOLO V4平台，采用深度学习算法来识别监控视频的人，并加入统计算法对识别出的人进行统计计算，实现视频监控下的人数统计。

关键词

密集人群，视频监控，人群计数，YOLO V4

The Design of the People Counting System for Monitoring Video-Intensive Crowds

Liwen Shen[1], Zhao Qiu[1*], Guichao Peng[2], Ping Huang[3], Chao Li[1], Jinye Cai[1]

[1]School of Computer and Cyberspace Security, Hainan University, Haikou Hainan
[2]National Computer Network Emergency Technology Coordination Center Hainan Branch, Haikou Hainan
[3]School of Science, Hainan University, Haikou Hainan
Email: 269774968@qq.com, *5454734@qq.com, 446419342@qq.com, 1197648483@qq.com

Received: Sep. 9th, 2020; accepted: Oct. 2nd, 2020; published: Oct. 9th, 2020

Abstract

With the continuous development of society, crowded places can be seen everywhere. Performing

*通讯作者。

文章引用: 沈礼文, 邱钊, 彭贵超, 黄萍, 李超, 蔡金晔. 监控视频密集人群的人数统计系统设计[J]. 图像与信号处理, 2020, 9(4): 202-210. DOI: 10.12677/jisp.2020.94024

statistical analysis on the personnel under surveillance video and realizing the number counting algorithm can provide effective technical means for the optimal allocation of urban public resources, security personnel scheduling, and safety management. Based on the YOLO V4 platform, this article uses a deep learning algorithm to identify people monitoring video, and adds a statistical algorithm to perform statistical calculations on the identified people to realize the number of people under video surveillance.

Keywords

Dense Crowd, Video Surveillance, Crowd Count, YOLO V4

Copyright © 2020 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

随着社会的发展，人们外出活动人数快速增长，人员密集的场所随处可见，意外事故发生的概率也随之增高，比如 2014 年 12 月 31 日，上海跨年夜活动，因众多游客市民聚集在上海外滩迎接新年，上海市黄浦区外滩陈毅广场东南角通往黄浦江观景平台的人行通道阶梯处底部有人失衡跌倒，继而引发多人摔倒、叠压，致使拥挤踩踏事件发生，造成 36 人死亡，49 人受伤的重大意外事故。因此，研究监控视频下的人数统计方法，对于疫情防控、共场所人数控制以及防止意外事故等有重要意义。统计特定场所下的人员数量，不仅能给公共场所管理人员进行人员调度，而且也能给相关职能部门人数信息，可以做出相应措施，对密集人群进行遣散，防止意外事故的发生，对社会安全意义重大。

国内的人数统计大多使用传统的人工统计方式。如果需要知道某个场所的人员情况，必须派人进行手工统计或者在通过监控视频逐一进行人员统计。这种人工统计方法不仅费时费力，而且只能知道某一时段的人数总数，不能得到某一时刻的人数数量。

在当今社会，视频监控基本覆盖了公共区域，而且监控设备可以 24 小时不间断的对目标区域进行实时监控，同时基于当今相关的成熟的图像算法，监控视频人数统计系统可以实现实时的人数统计功能，为社会稳定和预警防止意外的发生提供了数据的支持。随着这几年深度学习技术在计算机视觉方面取得的快速发展，越来越多的方法开始应用于人数统计当中。

本文在目标检测的基础上，采用当前最新的检测算法，利用深度学习的方式，用监控视频下的人群数据集训练模型，实现监控视频下密集人群人数的实时统计系统。

2. 相关研究

作为图像领域中的热点问题之一，视频监控下的人数统计方法的研究也成为当下研究的热点之一，特别是在全球疫情期间。

2.1. 基于检测的方法

基于检测的方法是人群数量统计研究的主要方法之一。Piotr 等人[1]采用一个滑动窗口检测器，选择性的从图像中裁剪出不同大小的图像块的方式来检测场景中人群，并统计相应图像内的人群数量，进而统计出全图内的人数。其中，基于检测的方法又大致可以分为基于人的整体的检测和基于人的身体的某

些部分的检测两类。[2] [3]中采用 SVM 方法训练分类器，利用从视频图像中提取的 HOG 特征去检测行人。[4]中采用 boosting 的方法训练一个分类器，利用从视频图像中行人的全身提取的小波特征去检测行人。[5]中采用随机森林的方法训练一个分类器，利用从视频图像中行人的全身提取的边缘等特征去检测行人。稀疏的人群因人相互遮挡相对较少，适合应用人的整体检测，随着监控视频内人数的提升，行人与行人之间的肢体相互遮挡会得越来越严重，导致识别的准确性也随之降低。为了解决这一问题，[6] [7]中利用人的身体某些部位来识别统计人群计量，这种方式主要通过检测身体的头部、肩膀、躯干等部位来识别人。这种方法在一定程度上解决了行人间相互遮挡的问题，在识别效果上比人的整体识别方法有略微的提升。

2.2. 基于回归的方法

[8] [9]等人实现人数统计的方法收是采取了学习一种图像特征到人群数量的映射来实现。这种方法首先提取前景特征、边缘特征、纹理特征和梯度特征等低级别的特征；然后利用线性回归，分段线性回归，岭回归和高斯过程回归等方法训练回归模型，来学习先前提取的低级别的特征到人群数量的映射关系，从而完成人数的统计。

2.3. YOLO 算法的提出与发展

2015 年，[10]中提出了一种一步到位的速度相对较快的目标检测方法，明显的提升了目标检测速度，但因其不支持拥挤物体的检测以及对小物体的检测效果差，且对新的宽高比物体检测效果不好等缺点，2016 年，[11]中对其进行改进，进一步提升的 YOLO (You Only Look Once)模型的运行速度。为了使网络可以输入不同大小的图片，YOLO V2 使用卷积操作代替全链接层，输入不同大小图片的训练也使得 YOLO V2 网络模型对不同大小的图片的检测更具有鲁棒性。为了更有效的检测拥挤物体和小物体，YOLO 网络的每个 cell 采用 5 个 anchor box 对图片进行预测，但未专注于人物识别。2018 年，[12]中作者使用一种多个尺度融合的添加残差网络的混合新模型，利用多标签的独立的逻辑分类器进行预测，检测速度和检测准确率进一步提升。为了提升网络在实际应用中的快速检测的目的，2020 年，Alexey 等[13]设计了一个简单且高效的目标检测算法，即 YOLOV4 算法，进一步提升了目标检测的速度和准确率，并减少了模型的计算量。

相比于 2.1、2.2 中所提检测算法，YOLO V4 算法在检测准确率以及检测速度上均有较大优势，并且该算法可以检测拥挤人群，在一定程度上解决了拥挤人群相互遮掩的这一难点。本文在 YOLO V4 的基础上，主要使用监控视频下的人群数据集，用这些人群数据集训练 YOLO V4 网络模型，达到快速准确的识别人，进而实现监控视频下人数统计。

3. 基于 YOLO V4 的人数统计系统架构

本文提出的监控视频下基于 YOLO V4 算法模型的人数统计系统主要包括如图 1 所示的几个部分。

1) 视频获取

该部分的工作为接收监控摄像头的传回的视频图像，根据系统调取摄像头的命令调取相应的摄像头监控视频，将视频暂时存入缓存中，方便后续的调取。

本文采用的是海康威视 SDK 接口来完成监控视频调用来获取监控视频图像，作为我国乃至全球领先的监控摄像头的提供商的海康威视，许多监控设备都是使用其产品，用其接口来完成视频获取，减少了开发时间，提高了系统稳定性、橹棒性。

2) 人数统计

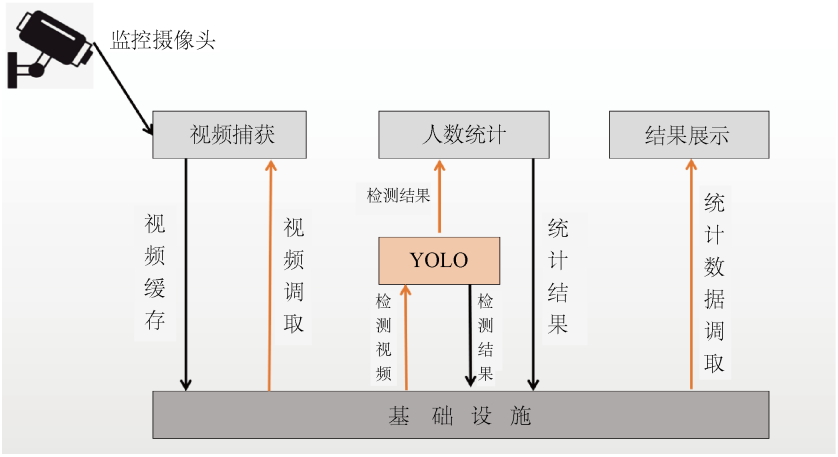

Figure 1. System architecture diagram
图 1. 系统架构图

该部分为系统核心部分，首先，将电脑硬盘中缓存的视频图像输入到 YOLO 网络模型中；然后，YOLO 会对输入的视频进行抽帧处理；最后，将处理后的结果存入电脑硬盘中，将处理后的结果 txt 文件送入人数统计中进行统计计算，然后将结果保存于电脑硬盘中,通过输出的图像图片以及统计结果进行有序的合成为视频图像，完成视频图像的人数统计。

由于 YOLO V4 速度可以达到 120 fps 左右，通过对视频进行抽帧处理，基本可以实现视频监控图像的实时识别，所以将输入图片改为实时摄像头视频图像的输入，能够实现快速的检测并得出结果并进行存储，使得 YOLO V4 能很好的应用于实际的视频监控场景当中，并且能实时的得出检测结果。

3) 系统基础设施

该部分主要是指能满足本系统运行所需的各种基础的软硬件设施，包括：监控摄像头、高性能电脑、用于数据展示的显示设备等。

4) 结果

该部分为结果展示，调取储存于电脑硬盘的统计结果，根据 YOLO 输出的图像加上统计的结果进行展示，方便用户的查看。

4. 系统功能模块说明

本文提出了一种监控视频下基于 YOLO v4 的人数统计系统设计方案，通过调取监控摄像头获取视频图像，然后通过 YOLO 算法模型进行图像抽帧检测，根据检测结果加入统计算法，得到最终的人数统计结果。本系统的各个功能模块如图 2 所示：

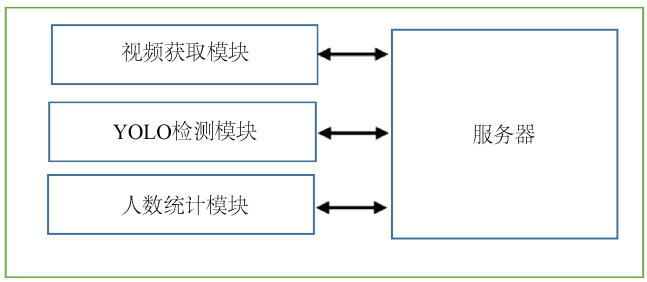

Figure 2. System function module
图 2. 系统功能模块

4.1. 视频获取模块

该模块主要是接收监控摄像头回传的视频图像，然后将视频缓存至服务器缓存中，方便模型调取数据。

许多监控都是使用海康威视的产品，应用其接口来调取视频监控，满足系统在日常情况下的应用，符合大部分所需。在本文中我们使用海康威视 SDK 接口来完成监控视频调用来获取监控视频图像。下图 3 为有关监控视频调用的流程。

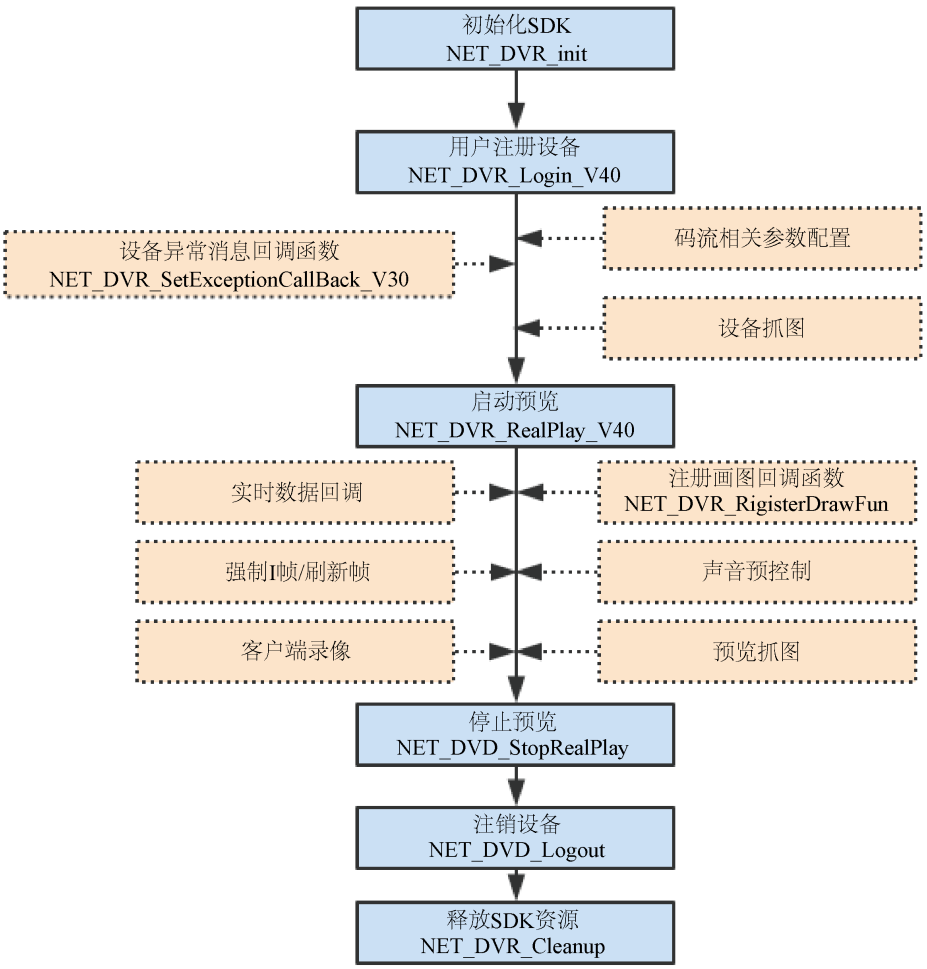

Figure 3. Hikvision video call flow chart
图 3. 海康威视视频调用流程图

1、初始化 SDK：对调用监控视频的 SDK 进行初始化操作，预分配内存。

2、用户注册设备：实现视频监控设备使用用户的注册功能，注册成功后，用户 ID 作为唯一标识，用这唯一 ID 才能进行其他功能的操作。

3、启动预览：实现视频监控的预览，视频预调取存入缓存中。

4、停止预览：释放相应缓存资源，停止视频监控的预览。

5、注销设备：注销相应设备。

6、释放 SDK：释放 SDK 资源，结束整个过程。

4.2. YOLO V4 检测模块

该模块主要完成视频监控图像中行人的检测，通过监控视频获取模块获取的视频进行处理，对视频进行抽帧处理，送入模型进行识别，得到一个输出图像。图 4 是 YOLO V4 整体网络构造：

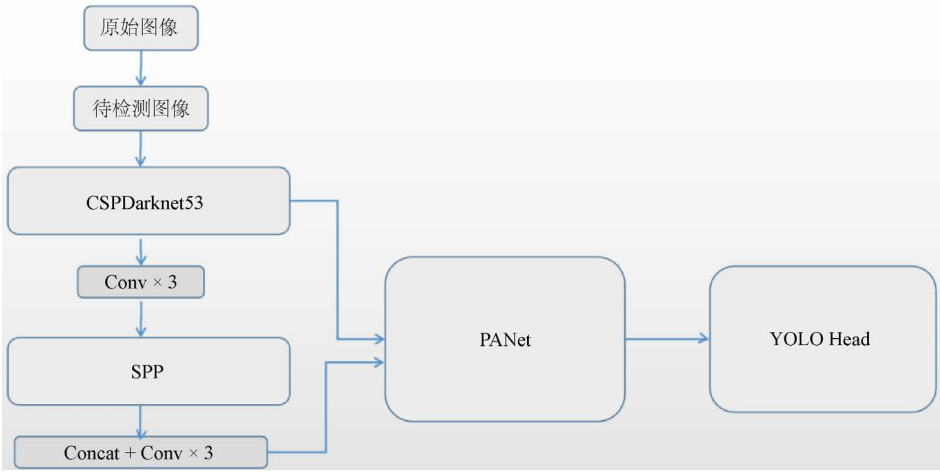

Figure 4. YOLO V4 overall network structure
图 4. YOLO V4 整体网络结构

该网络大致由 4 个部分组成，YOLO V4 的 backbone 采用的是 Darknet53 + CSPResnet (图 5(a))构成的网络模块，Neck 采用的是 PANet (图 6)网络，Head 采用的是 YOLO V3。首先将原始视频图像缩放到为 416 * 416 大小，最后得到 608 * 608 * 3 的大小的待检测图像，将待检测图像输入到网络中。经过由 Darknet53 + CSPResnet 组成的 CSPDarknet53 网络得到 1024 个 13 * 13 的图像，将得到的结果先用 3 * 3 的卷积核进行卷积操作，再将结果送入 SPP 中，SPP 采用的是最大池化操作，然后将结果连接并用 3 * 3 卷积核采用，送入 PANet 中。PANet 网络接收 SPP 输入的数据，加上 CSPDarknet 网络倒数第二层和第三层网络输出数据分别进行连接操作并上采用会原来大小，将得到的结果相连接送入 YOLO Head 中，YOLO Head 采用的是 YOLO V3 网络结构，最终得到结果。YOLO V4 的网络模型简洁，主要利用 Darknet 和 CSPResnet 提取输入图像的特征，送入 YOLO V3 中得到最终的目标检测结果。

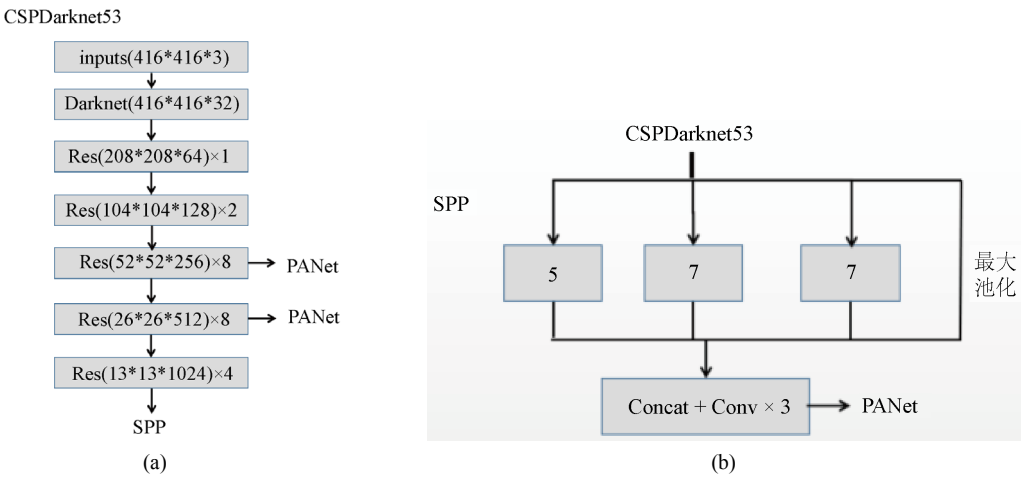

Figure 5. CSPDarknet53 (a) and SPP network structure (b)
图 5. CSPDarknet53(a)和 SPP(b)网络结构

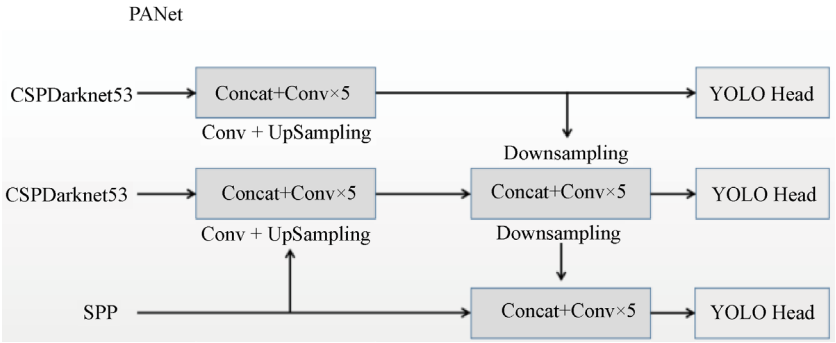

Figure 6. PANet network structure
图 6. PANet 网络结构

YOLO V4 模型可以检测多个分类物体，但本系统只需识别人，因此我们修改该模型只保留检测人即 Person 这一分类。同时为了让 YOLO 模型专注于识别人，我们选用 Cheng 等在[14]中的新 Mall 数据集，该数据集收集的人群数据集来自安装在购物中心中的监视摄像头，拥有不同光照条件和不同的人群密度。该数据除了有不同密度的人群外，它还包括了行人的不同的姿态，不仅有行走的行人、静止站立的行人，还有坐着的行人。该数据集还呈现了由场景对象(例如沿着行走路径的室内植物)引起的严重遮挡的挑战。该数据由两千帧 320 * 240 大小的图像构成的视频序列组成，其中有 6000 个标记为行人的实例。我们用前 800 帧用于训练我们的模型，用剩余的 1200 帧用于评估测试。

4.3. 人数统计模块

该模块完成最终的人数统计，并输出相应的结果。YOLO V4 模型可以快速的识别出输入图像并输出检测结果，YOLO V4 的输出为一个类别对应一个 txt 文件，有多少类别，就产生多少个 txt 文件，因本系统只需输出 person 这一分类，所以只有一个输出 txt 文件，人数统计模块只需处理这一文件得出统计结果即可。图 7 是人数统计模块的处理过程：

Figure 7. The processing process of the people counting module
图 7. 人数统计模块的处理过程

4.4. 实验结果

我们在 Mall 数据集上展示人数统计效果。该数据集收集的是安装在一个商城中监控摄像头获取的监控视频图像，符合我们的系统设计要求，并且该数据集图像中的行人姿态、人数均各不相同，具有一定的挑战性的代表性。我们在人群检测结果上加入统计算法，进行结果统计，我们准确识别率可以达到 90%以上，统计效果较好，满足日常情况下的监控视频里的人群统计。下图 8 是我们加入统计算法后进行人群统计的结果。

下图 9 是分别我们在网络上截取的大巴车和公交车内图像，因为图像清晰度及角度等原因，我们的统计结果准确率也在 85%左右。

5. 结束语

随着社会的不断发展，以及疫情还未完全消除，通过视频监控得到某区域内的人数，方便相关职能

部门根据结果及时处理、遣散密集人群，也可以为一些特定场所提供数据支持，特别是商场、地铁等公共区域，可以根据人数统计结果做出相应的人员调度，防止公共区域发生意外事件。本文使用 YOLO V4 算法模型进行行人检测，同时加入统计算法完成监控视频下的人数统计，YOLO V4 能快速运算出结果，可以及时的得到结果，符合系统的时效性。

Figure 8. Mall datasets show
图 8. Mall 数据集展示

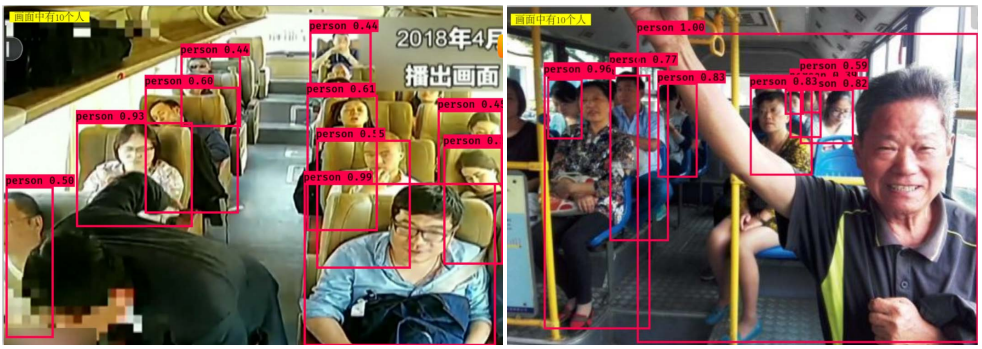

Figure 9. The people counting of bus result show
图 9. 大巴车和公交车统计结果

致　谢

本文的撰写十分感谢我的研究生导师即通讯作者邱钊教授的悉心指导，感谢相关项目使我有幸参与到监控人数统计系统的设计研发工作。同时，本文的完成离不开师兄师姐、师弟师妹的支持。本文提出的系统架构以及模块的划分都是在师兄师姐，师弟师妹共同帮助下努力的结果，在此，感谢以上提到了每一个人为本文所做出的贡献。

基金项目

海南省自然科学基金项目“基于 YOLO V2 的监控视频人数统计研究(No. 618MS028)”。

参考文献

[1] Piotr, D., Christian, W., Bernt, S. and Pietro, P. (2012) Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**, No. 4. https://doi.org/10.1109/TPAMI.2011.155

[2] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. IEEE Computer Society, **1**, 886-893.

[3] Leibe, B., Seemann, E. and Schiele, B. (2005) Pedestrian Detection in Crowded Scenes. *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, **1**, 878-885.

[4] Markus, E. and Gavrila Dariu, M. (2009) Monocular Pedestrian Detection: Survey and Experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**, No. 12. https://doi.org/10.1109/TPAMI.2008.260

[5] Oncel, T., Fatih, P. and Peter, M. (2008) Pedestrian Detection via Classification on Riemannian Manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**, No. 10. https://doi.org/10.1109/TPAMI.2008.75

[6] Felzenszwalb Pedro, F., Girshick Ross, B., McAllester, D. and Ramanan, D. (2010) Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, No. 9. https://doi.org/10.1109/TPAMI.2009.167

[7] Nevatia, R. and Wu, B. (2007) Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet Based Part Detectors. *International Journal of Computer Vision*, **75**, 247-266. https://doi.org/10.1007/s11263-006-0027-7

[8] Bayesian Poisson Regression for Crowd Counting (2009) *IEEE* 12*th International Conference on Computer Vision* (*ICCV* 2009), 545-551.

[9] Ryan, D., Denman, S., Fookes, C., *et al.* (2009) Crowd Counting Using Multiple Local Features. 2009 *Digital Image Computing*: *Techniques and Applications* (*DICTA* 2009), **1**, 81-88. https://doi.org/10.1109/DICTA.2009.22

[10] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 779-788.

[11] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 6517-6525.

[12] Redmon, J. and Farhadi, A. (2018). YOLOv3: An Incremental Improvement. ArXiv, abs/1804.02767.

[13] Bochkovskiy, A., Wang, C.-Y. and Liao, H.-Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. ArXiv, abs/2004.10934.

[14] Chen, K., Loy, C.C., Gong, S. and Xiang, T. (2012) Feature Mining for Localised Crowd Counting. In *Proceedings of British Machine Vision Conference*, (*BMVC*), **1**, 1-11. https://doi.org/10.5244/C.26.21