

论文题目	基于深度学习的多目标距离检测方法研究
课题来源	辽宁省教育厅面上项目（LJKZ0584）
<p>报告要求：</p> <p>开题报告字数应不少于 5000 字，阅读的参考文献应不少于 20 篇，其中外文文献不少于三分之一。</p> <p>内容提示：</p> <ol style="list-style-type: none"> 1. 论文选题的背景、理论意义或应用价值 2. 国内外研究现状及发展趋势 3. 主要研究内容 4. 理论依据或技术方案 5. 创新点和预期结论 <p>一、论文选题的背景，理论意义或应用价值</p> <p>自 2019 年末新冠疫情在武汉爆发以来，据世界卫生组织(WHO)统计，全球已经累计感染五亿人并且因为新冠病毒死亡的人数也已经达到了六百万人^[1]，尽管现在已经拥有了系统的治疗防疫体系和研制出了有效的疫苗，但是新冠病毒的传染性强同时新冠病毒也处于不断变异的过程，因此导致新冠疫情一直无法根除。此次疫情对各行各业产生了不同程度的影响，尤其是对交通运输业、餐饮娱乐业、制造业以及教育培训业的打击巨大，对人民的生命产生巨大的威胁，对中国以及全球各国都带来了巨大的影响。为了以最低的社会成本在最短时间内控制住疫情，最大限度保护人民生命安全和身体健康，最大限度减少对经济社会发展和群众正常生活的影响，一定的疫情防控措施是很有必要的。我国国务院在就新冠病毒对公众的建议中提到坚持预防为主，做到坚持佩戴口罩、减少人员聚集、加强通风消毒，并且强调在人员密集的场所与他人保持 1 米以上社交距离^[2]，从而降低被感染的风险。在人员密集的场所特别是封闭环境中保持一定的社交距离对于预防新冠病毒是必不可少的措施，比如医院、学校、银行以及核酸检测的场所。为了在科学疫情防控的基础上降低人力物力成本，通过计算机视觉实现目标之间的距离检测的算法模型研究是势在必行的。</p> <p>实现距离检测的方法很多，可以通过激光传感器、超声波测距雷达等方法实现，但是以上距离测量方法的成本较高，且对使用环境有一定要求，且对于多个目标以及移动目标进行距离检测时容易出现精度降低的问题。为了实现一种成本较低，且相对更容易部署的距离检测方法，本研究提出一种基于深度学习的多目标距离检测方法。人类对距离最直观</p>	

的判断来自于视觉，通过最直观的视觉感受判断目标的远近程度，因此可以通过运用计算机视觉来实现距离检测。计算机视觉是一门研究如何使机器“看”世界的科学，确切来说，就是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量，并进一步做图形处理，使电脑处理后的图像更适合人眼观察或传送给仪器检测。运用计算机视觉领域内的目标检测技术对目标完成识别后，再进一步采用距离判断算法等即可完成距离检测。目标检测分为传统的目标检测和基于深度学习的目标检测，传统的目标检测算法例如 Viola Jones 检测算法^[3]、HOG+SVM 检测算法^[4]、DPM 检测算法^[5]等提取特征时主要采取人工提取，具有一定的局限性和鲁棒性低的缺点，同时传统的目标检测算法基于滑动窗口的区域选择策略会产生大量冗余候选框，导致检测速度变慢。由于人体距离识别需要较高的检测速度保证实时性，导致传统目标检测算法并不适用于距离检测。

近些年来，深度学习随着其相关技术变得越来越成熟，它也逐渐成为了人工智能领域中的研究热点，并在语音识别、计算机视觉等多个应用场景中取得了突破性的进展。深度学习旨在通过模拟人类大脑的神经连接机制建立网络模型，网络模型再对学习样本中的数据特征进行提取处理，这种学习方式与人类感知事物的原理比较相符，所以当有足够多的学习样本时，深度学习网络模型的检测效果也更佳^[6]。基于深度学习的目标检测算法不断发展，其检测速度和精度上已经远远超过传统目标检测算法，并且一些基于深度学习的网络模型在 COCO^[7]、VOC^[8]等常用目标检测数据集上的检测精度不亚于人工识别。因此选用基于深度学习的目标检测算法进行距离检测的研究更为合适。

本课题通过基于深度学习的方法完成目标检测并计算被检测目标之间距离的研究，不仅可以进行人与人的安全距离检测，还可以进行人与其他类别的物品、物品与物品之间的距离检测，例如对人与车辆之间的距离进行检测、对电气设备与可燃物品之间的距离检测等，因此本研究除了提供一种在新冠疫情之下计算人与人之间安全社交距离检测的算法，更提供了一种可以应用于工业制造、消防安全以及自动驾驶等各领域的多目标距离检测算法。

二、国内外研究现状及发展趋势

目标检测分为传统的目标检测和基于深度学习的目标检测，传统的目标检测算法首先通过滑动窗口从左到右、从上到下遍历整张图片并产生一定数量的候选框，然后提取候选框的特征，最后利用支持向量机（SVM）等分类方法对提取到的特征进行分类进而得到结果。2001 年 Viola Jones 等人在特征抽取部分使用 Haar 特征作为特征表示的算子，使用 Adaboost 算法训练分类器，最后运用级联结构这种高效的分类策略，解决了计算量庞大导致时间复杂度极高的问题；2006 年，Dalal 等人提出了 HOG(Histogram of Oriented Gradients)检测器，通过计算图像某一区域的不同方向梯度累积形成直方图来表示局部对象特征，在

图像局部细节信息的提取上具有较好的稳定性，是当时尺度特征不变性和形状上下文的重要改进，使用 SVM 作为分类器进行行人检测；2008 年，Felzenszwalb 等人提出了 DPM 检测算法，相比于 HOG 检测算法作了很多改进，可以说是 HOG 检测算法的延申，该算法由一个主过滤器和多个辅过滤器组成，通过硬负挖掘，边框回归和上下文启动技术改进检测精度。

随着卷积神经网络（convolutional neural network, CNN）的广泛应用，使得基于深度学习的目标检测算法得以实现并迅速发展。基于深度学习的目标检测算法目前可分为两大类：两阶段（two-stage）检测和一阶段（one-stage）检测^[9]，RCNN、Fast RCNN、Faster RCNN 等算法都是典型的两阶段检测算法，典型的一阶段算法有 SSD、YOLO 系列等。

2014 年，Girshick 等人提出了 RCNN^[10]，在 VOC 2007 数据集中取得出色的效果，比之前其他方法有近 50% 的性能提升，mAP 达到了 58.5%，RCNN 算法使用选择性搜索算法来评估相邻图像子块的特征相似性，给合并后的相似图像区域打分，选择感兴趣区域的候选框作为卷积神经网络样本输入，由标定框与候选框组成的正负样本特征形成相应的特征向量，采用 SVM 对特征向量进行分类，最后返回标定框与候选框，以达到目标检测的目的；2015 年，Girshick 等人提出了 Fast RCNN^[11]，改进了 ROI Pooling，将不同大小候选框的特征图采样成大小固定的特征，大大提升了检测速度；Ren 等人于 2015 年提出了 Faster RCNN^[12]，Faster RCNN 在 Fast RCNN 算法的主干网络中添加了 RPN（Region Proposal Network），实现了网络端到端的训练。

Liu 等人于 2015 年提出的一阶段检测算法 SSD^[13]，使用 VGG16 基础网络作为特征提取网络，为检测不同尺度的目标，在前面卷积层输出的特征图中检测小的目标，大的目标则在后面卷积层传递的特征图中被检测，在准确率和检测速度上都有提高；2015 年，Redmon 等人提出了 YOLO（You Only look Once）^[14]算法，利用回归的思想直接在原始图像多个位置回归出定位框和目标类别，其最大的优势就是检测速度快，检测速度更是达到 45FPS，但是相对来说检测精度并不高，YOLOv2 相比前一版本，主要改进点是提出了联合训练算法，其基本思想是使用两种数据集同时对检测器进行训练，检测数据集和分类数据集，在检测数据集上来定位物体的位置，而分类数据集则用来增加检测器的识别的物体种类，因此 YOLOv2 在保持速度的基础上定位更准确，YOLOv3^[15]的特点是引入了特征金字塔结构来实现多尺度预测，YOLOv4^[16]在 YOLOv3 的基础之上 YOLOv4 引入了 CSPDarknet-53 提取特征，让其成为 YOLO 系列中一个重要的里程碑。

三、主要研究内容与目标

1 研究目标

本课题的研究目标是在对基于深度学习的目标检测算法分析对比的基础上，选取合适

的目标检测算法实现对人的检测和定位，并对算法模型进行轻量化改进，提升检测的速度，同时对模型进行精度提升改进，解决检测模型对被遮挡目标以及小目标检测效果不佳的问题，达到检测精度与检测速度兼具的效果。然后采取实时跟踪算法技术对人体进行跟踪，解决视频检测时目标丢失的问题，再利用逆透视映射视角转换算法和距离判断算法，实现对各个被检测目标之间的距离检测。本课题在目标检测的基础上，开展基于深度学习的多目标距离检测方法研究，实现一种成本较低、检测速度更快且易于部署的距离检测方法。

2 研究内容

(1) 自建数据集的制作

① 为了满足人体距离检测的实际要求，不仅要考虑不同场景、不同角度以及不同视角下对人的检测效果，同时需要考虑人在不同姿态下对检测的影响，例如站立、行走、打伞遮挡以及骑行等。利用开源行人数据集 CUHK Occlusion Dataset^[17]、Mall Dataset^[18]来丰富数据集的场景和拍摄视角，同时利用网络爬虫工具爬取网上不同姿态的人体图片，丰富数据集中不同姿态人体的数据，使数据集数量达到 2871 张。

② 为了防止目标检测算法在数据集上训练出现过拟合导致最终检测效果不佳的情况，在自建数据集上利用旋转一定角度、水平翻转、高斯模糊等数据增强方法对数据集进行扩充使数据集数量达到 8607 张，增加模型的鲁棒性和抗干扰能力。

③ 使用 LabelImg 人工标注工具对数据集中人体进行标注并保存，以便今后对模型进行训练，标注时场景如图 1 所示。

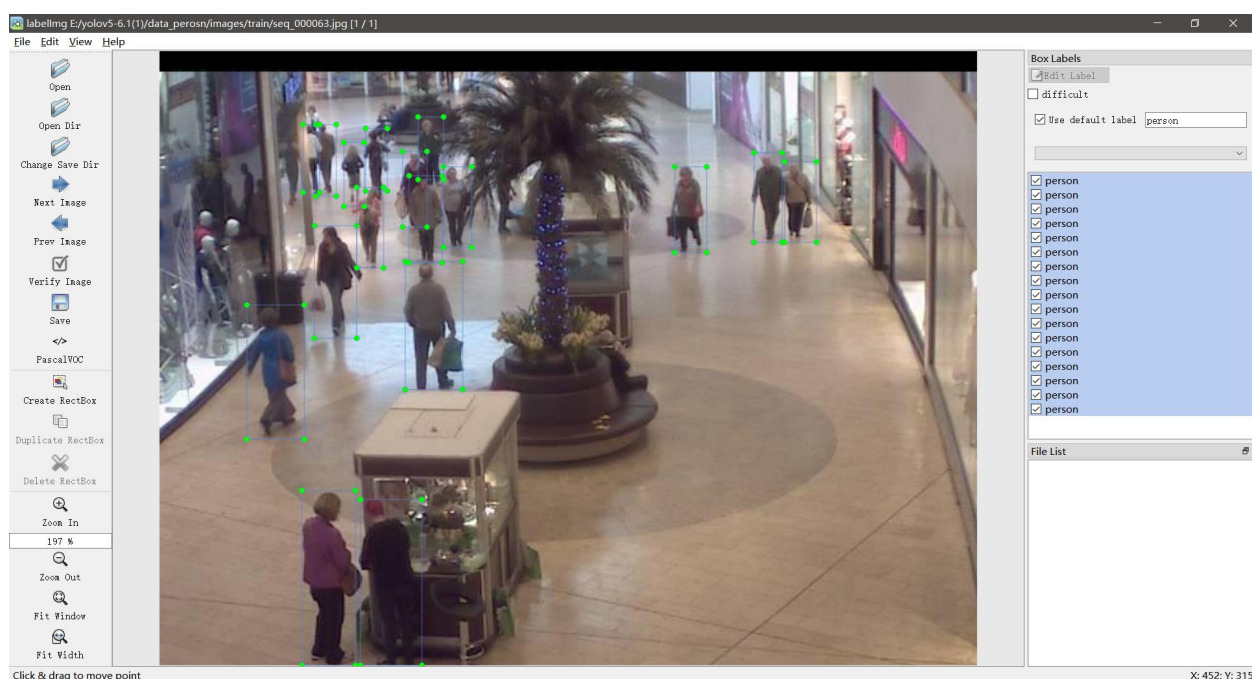


图 1 使用 LabelImg 对数据集进行标定

(2) 人体的检测识别

- ① 利用脚本将自建数据集按照 8: 2 的比例分为训练集和测试集。
- ② 将划分好的数据集在 YOLOv5 网络模型上进行训练, 并保存训练好的模型。
- ③ 将训练好的模型在测试集上进行测试, 计算出模型的精度 AP 和速度 FPS。

(3) 人体的距离检测

① 通过训练好的模型完成人体检测后, 利用简单在线实时跟踪算法, 对人体进行跟踪和 ID 分配。

② 利用 opencv 库中视角变换函数将视角从摄像机视角转换到俯瞰视角。

③ 在俯瞰视角下, 利用距离判断算法完成人与人之间的距离计算, 最后把俯瞰视角中完成的距离检测转换到摄像机视角。

四、理论依据或技术方案

1 YOLOv5 目标检测算法

在人体距离检测中, 第一步便是对影像序列中的人体进行检测并定位, 所以检测器的选择对于最后的结果有着巨大的影响。YOLOv5 的轻量化网络设计使得其便于部署到移动设备, 且检测速度快。YOLOv5 包括 YOLOv5s、YOLOv5m、YOLOv5l 以及 YOLOv5x, 这四种网络结构相同, 网络深度和宽度依次增加^[19]。因为需要考虑实际场景中对检测速度和运算量的要求, 所以选择 YOLOv5s 作为人体检测的基础模型。

YOLOv5s 的网络结构如图 2 所示, 分为输入端、Backbone 主干特征提取网络、Neck 特征融合网络以及 YOLO Head 输出端。输入端实现对输入的预处理, 包括 Mosaic 数据增强对多个图像进行随机缩放、裁剪、排布后进行拼接, 以及自适应图片缩放将原始图片统一缩放到一个标准尺寸, 还有自适应锚框计算针对不同的数据集自动设定初始锚框的长宽。Backbone 主干特征提取网络通过采用 Focus 下采样、Conv 卷积、C3 结构以及 SPP 池化金字塔结构对输入的图片进行卷积操作并提取出有效特征层^[20]。Neck 特征融合网络采用特征金字塔网络 FPN^[21]与路径聚合网络 PAN^[22]对不同尺寸有效特征层的特征信息进行特征融合, 实现不同尺寸特征信息之间的传递。YOLO Head 作为输出端, 采用三种损失函数对类别、位置、置信度进行计算, 并通过非极大抑制 NMS^[23]对结果进行筛选。

整个 YOLOv5 网络的工作内容就是对输入进网络的图片进行特征提取、特征融合与加强、预测特征点对应的特征信息, 最后完成对物体类别、位置、置信度的预测。

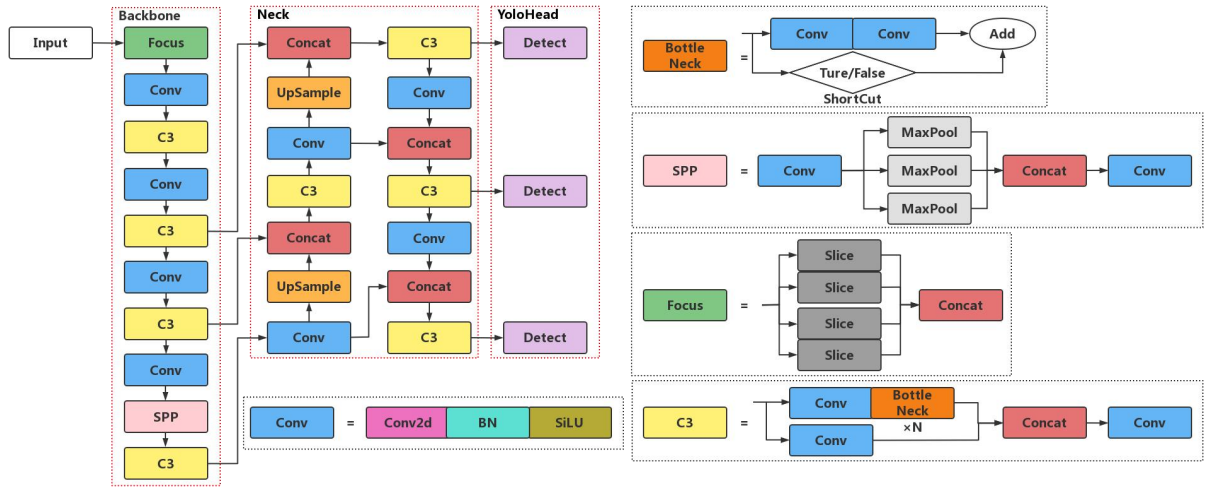


图 2 YOLOv5s 网络模型结构

2 SORT 目标跟踪算法

在完成人体的检测识别后的通过简单在线实时跟踪算法^[24](Simple Online and Realtime Tracking, SORT) 对视频序列中的检测到的人体进行跟踪。SORT 算法利用卡尔曼滤波和匈牙利算法完成对检测器输入进来的检测目标的跟踪和 ID 分配，实现了一种简单高效的跟踪范式^[25]，其实现步骤主要分为人体状态估计、数据关联和跟踪器更新三个步骤，如图 2 所示。

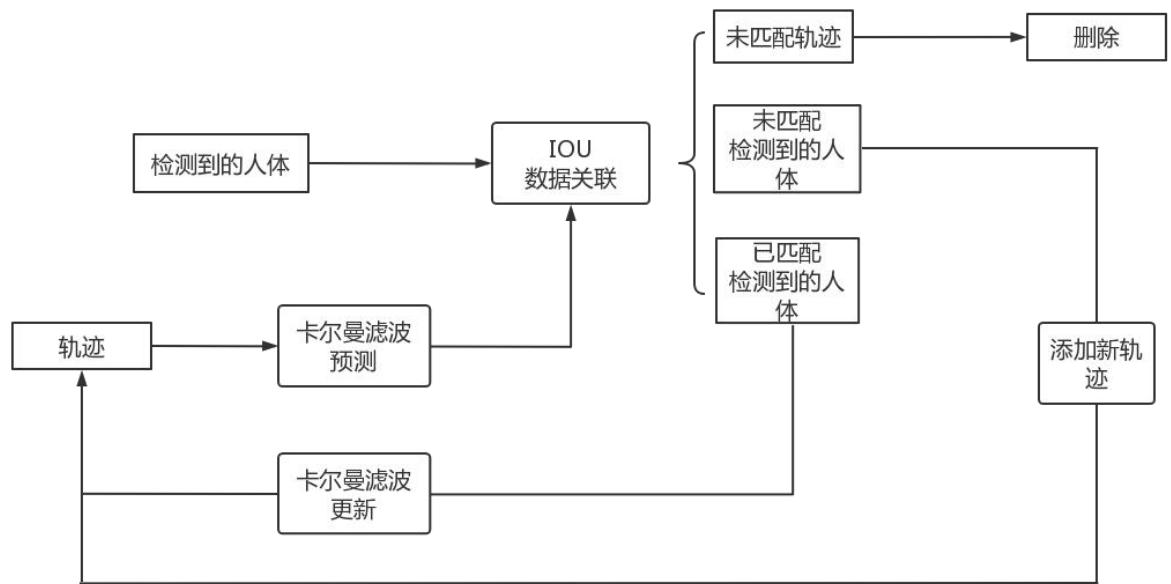


图 3 SORT 算法原理图

(1) 人体状态估计

对被检测到的每个人体使用七维状态空间表示其状态，如式 (1)。

$$x = [u, v, s, r, u', v', s']^T \quad (1)$$

式中 u 和 v 分别对应目标检测框中心点的横坐标和纵坐标， s 表示检测框的面积， r 表示检测框的宽高比， u' 、 v' 、 s' 分别是 u 、 v 、 s 的导数，表示检测框的速度信息。

通过卡尔曼滤波完成对人体的状态估计，分为预测和更新两个阶段。预测阶段通过对前一帧被跟踪人体的位置来预测当前帧人体的位置；更新阶段基于当前帧检测到的人体的位置更新预测阶段的人体的位置。状态预测式如下：

$$\begin{cases} X_t = FX_{t-1} \\ P_t = FP_{t-1}F^T + Q_{t-1} \end{cases} \quad (2)$$

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

式中 X_t 表示预测第 t 帧人体的状态， F 表示状态转移矩阵， X_{t-1} 表示第 $t-1$ 帧人体的状态， P_t 表示第 t 帧的误差协方差矩阵， P_{t-1} 表示第 $t-1$ 帧的误差协方差矩阵， Q 表示系统的噪声矩阵。

人体的状态更新式如下：

$$\begin{cases} K_{t+1} = P_t H^T (H P_t H^T + R)^{-1} \\ X_{t+1} = X_t + K_{t+1} (Y_{t+1} - H X_t) \\ P_{t+1} = (I - K_{t+1} H) P_t \end{cases} \quad (4)$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (5)$$

式中 K_{t+1} 为第 $k+1$ 帧时的卡尔曼增益， H 表示测量矩阵， R 表示检测器的噪声矩阵， X_{t+1} 表示更新后第 $t+1$ 帧人体的位置， Y_{t+1} 表示第 $t+1$ 帧人体的检测位置， P_{t+1} 表示为更新后第 $t+1$ 帧的误差协方差矩阵， I 表示单位矩阵。

(2) 数据关联

通过匈牙利算法对改进的 YOLOv5s 检测到的人和卡尔曼滤波预测的人进行进行关联，实现帧与帧之间的同一个人的编号保持一致性。设第 k 帧跟踪到的人有 n 个，记为 $T = \{T_1, T_2, \dots, T_n\}$ ， $k+1$ 帧检测到的人有 m 个，则记为 $D = \{D_1, D_2, \dots, D_m\}$ 。则通过计算 T 与 D 之间的交并比得到关联矩阵 C ，公式如下：

$$C = IOU(T, D) \quad (6)$$

设置阈值 $S=0.3$ ，处理关联矩阵使其得到一个只有 0 和 1 的矩阵 R ，处理过程如下：

$$R_{ij} = \begin{cases} 0, C_{ij} < S \\ 1, C_{ij} > S \end{cases} \quad (7)$$

在上式中， i 表示第 i 个被跟踪到的人， j 表示第 j 个被检测到的人，则 C_{ij} 表示第 i 个跟踪框与第 j 个检测框之间的交并比的值， R_{ij} 等于 1 时表示第 i 个被跟踪的人与第 j 个被检测到的人成功匹配。同时根据匈牙利算法要求，每个被跟踪的人只能与一个被检测到的人匹配，即满足：

$$\begin{cases} \sum_{i=1}^n R_{ij} = 1 \\ \sum_{j=1}^m R_{ij} = 1 \end{cases} \quad (8)$$

(3) 跟踪器更新

在完成数据关联后，跟踪器则需要进行更新。对于与被检测到的人匹配成功的跟踪器，该跟踪器以及其编号则会由与之匹配成功的被检测到的人继承，并利用其检测框的状态信息预测下一帧人的位置；对于未能与被检测到的人匹配的跟踪器，暂时保留该跟踪器和编号，若在接下来的连续 5 帧都未能成功匹配则删除该跟踪器和编号；对于未能与跟踪器匹配的被检测到的人，则为其创建一个新的跟踪器并分配新的编号，利用该被检测到的人的状态信息对下一帧进行预测。

3 距离检测算法

在完成对人体的检测与跟踪后，需要对人与人之间的距离进行检测并判断距离是否超过安全距离。由于单个相机将三维世界投影到二维平面上，会导致物体之间的像素距离不能正确反映三维世界中的实际距离，因此利用 `opencv` 库中的视角转换函数将摄像机视角转换到俯瞰视角。在俯瞰视角下，通过距离判断算法实现人与人的距离检测。

(1) 视角转换

首先使用 `cv2.getPerspectiveTransform(source, target)` 获得透视变换矩阵 M ，其中 `source` 和 `target` 分别为变换前后图中呈矩形的四个点的坐标；然后使用 `cv2.warpPerspective(image, M)` 根据透视变换矩阵 M 对摄像机拍摄的图像 `image` 做透视变换；再使用 `cv2.perspectiveTransform(point, M)` 根据透视变换矩阵 M 对坐标点 `point` 进行透视变换^[26]。

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (9)$$

其中矩阵 M 代表透视变换矩阵， (x, y) 代表在俯瞰视角下的像素点坐标， (X, Y) 代表现实世界摄像机视角下的像素点坐标。

(2) 距离判断

在俯瞰视角下，将被检测到的人的检测框的底部中心作为一个人的参考坐标 (x_i, y_i) ，若任意两个人 P_i 、 P_j 之间参考坐标 (x_i, y_i) 与 (x_j, y_j) 的距离小于安全社交距离 S ，则视为距离过近违反安全社交距离^[27]。因此定义一个函数 $Y=V(P_i, P_j, S)$ ， $Y=1$ 时表示违反安全社交距离， $Y=0$ 时表示符合安全社交距离， S 表示距离阈值，具体根据摄像机视角设置，如下公式所示：

$$Y=V(P_i, P_j, S) = \begin{cases} 1, & \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq S \\ 0, & \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} > S \end{cases} \quad (10)$$

五、创新点与预期结论

1 创新点

(1) 针对自建数据集图片数量较小，采用旋转一定角度、水平翻转、高斯模糊等数据增强方法对数据集进行扩充，丰富数据集，防止过拟合。

(2) 针对检测视频中出现的同一人在不同帧下检测丢失以及检测框定位不准的问题，采用 SROT 跟踪算法，减小视频检测中漏检、定位框不准的情况。

(3) 针对小目标检测准确率不高以及 YOLOv5s 模型参数量过大问题，对 YOLOv5s 模型进行优化改进，提高准确率与检测速度。

2 预期结论

(1) 采用 SORT 算法，可以有效解决被检测对象在移动过程中检测丢失以及定位框不准的问题，从而使距离检测更精确。

(2) 经过优化改进后的 YOLOv5s 检测算法与原算法相比，模型参数量和计算量大大下降，检测精度也有提升，兼具检测速度与检测精度。

六、论文的工作计划

第一阶段：阅读深度学习、计算机视觉以及目标检测和目标跟踪的相关文献资料，同时学习 python 编程语言和 pytorch 深度学习框架，学习相关的目标检测算法，为后续的课

题发展打下基础，历时 7 个月，已完成。（2022.6-2022.12）

第二阶段：深入学习 YOLO 系列目标检测算法与 SORT 目标跟踪算法，理解 YOLOv5s 与 SORT 算法的代码并进行运行调试，并对算法的代码进行改进，同时学习视角转换算法和距离判断算法，理解其代码并进行改进。预计 8 个月完成。（2022.12-2022.7）

第三阶段：制作数据集，并学习数据增强算法对数据集进行扩充，将自制数据集应用到 YOLOv5s 模型进行训练，并对 YOLOv5s 模型进行改进。将改进后的 YOLOv5s 与 SORT 结合运用到人体距离检测当中。撰写一篇基于 YOLOv5s 与 SORT 的人体距离检测小论文。预计 6 个月完成。（2023.7-2023.12）

第四阶段：对课题整体进行总结，撰写毕业论文，最终形成书面材料，装订成册，并准备参加毕业答辩。预计 7 个月完成。（2023.12-2024.6）

七、主要参考文献

- [1] World Health Organization. WHO Coronavirus (COVID-19) Dashboard[EB/OL]. (2022-10-14)[2022-10-14]. <https://covid19.who.int/table/>.
- [2] 国务院应对新型冠状病毒感染肺炎疫情联防联控机制. 国务院应对新型冠状病毒感染肺炎疫情联防联控机制关于做好新冠肺炎疫情常态化防控工作的指导意见[EB/OL]. (2020-05-08)[2022-10-14]. http://www.gov.cn/zhengce/content/2020-05/08/content_5509896.htm.
- [3] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2001: I.
- [4] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'05). New York: IEEE Press, 2005, 1: 886-893.
- [5] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.
- [6] 谢富, 朱定局. 深度学习目标检测方法综述[J]. 计算机系统应用, 2022, 31(02): 1-12.
- [7] Lin T-Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context[A]. D. Fleet, T. Pajdla, B. Schiele, et al. Computer Vision – ECCV 2014[C]//Cham:Springer International Publishing, 2014: 740–755.

- [8] Everingham M, Eslami S M A, Van Gool L, et al. The Pascal Visual Object Classes Challenge: A Retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98–136.
- [9] 许德刚, 王露, 李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57(08): 10–25.
- [10] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] //2014 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2014: 580-587.
- [11] Girshick R. Fast R-CNN[A]. 2015: 1440–1448.
- [12] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [13] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [14] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 779-788.
- [15] Redmon J, Farhadi A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2021-06-04]. <https://arxiv.org/abs/1804.02767> .
- [16] Bochkovskiy A, Wang C Y, Liao H Y MARK. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2022-10-14]. <https://arxiv.org/abs/2004.10934> .
- [17] Ouyang W, Wang X. A Discriminative Deep Model for Pedestrian Detection with Occlusion Handling[J]. in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). 2012: 3256-3265.
- [18] Chen K, Gong S, Xiang T, et al. Feature Mining for Localised Crowd Counting[J]. In Proceedings of British Machine Vision Conference 2012, 21: 1-11.
- [19] Ultralytics. YOLOv5[EB/OL]. (2020-6-26)[2022-10-14]. <https://github.com/ultralytics/yolov5> .
- [20] 邱天衡, 王玲, 王鹏等. 基于改进 YOLOv5 的目标检测算法研究[J]. 计算机工程与应用, 2022, 58(13): 63-73.