2022 年华中科技大学电气与电子工程学院

证券投资训练营

课程项目设计报告

Final Report

作品题目：基于 YOLOv5 和 Deepsort 的物体识别与跟踪

团队成员：

李瑞昊

姚佳琦

俞晨欣

鲍林奕

魏子健

基于 YOLOv5 和 Deepsort 的物体识别与跟踪

摘要

视频图像中的运动物体识别与跟踪技术是计算机视觉、计算机科学、视频监控等学术领域广泛关注的一个重要课题。本文本论文主要基于计算机视觉领域下目标跟踪方向中的多目标跟踪领域进行探索，围绕目前主流的基于检测的多目标跟踪策略进行研究，并应用于实际路况下行人和车辆的识别和跟踪任务。首先是物体的识别和标记，通过训练 YOLOv5 网络对目标视频每一帧的对象进行识别处理，将视频中的对象标记出来并打上标签，传递到 Deepsort 网络中进行跟踪处理。在 Deepsort 模型中将 YOLOv5 中得到的检测框与预测的跟踪框输入到匈牙利算法中进行线性分配来关联每帧间的 ID，同时将目标的外观信息加入到帧间的匹配中，避免在 ID 受到遮挡后出现丢失。模型调试完成后将其合并打包成相应函数，完成项目目标。本文结合多目标检测算法，选择 YOLOv5 单阶段目标检测算法和 Deepsort 多目标跟踪算法，提高了算法在多目标跟踪数据集的识别和跟踪效果。

关键词： YOLOv5 Deepsort 物体识别跟踪 python

Object recognition and tracking based on YOLoOv5 and Deepsort

Abstract

The recognition and tracking technology of moving objects in video images is an important topic that is widely concerned in academic fields such as computer vision, computer science, and video surveillance. This text paper is mainly based on the multi-target tracking field in the field of target tracking in the field of computer vision, focusing on the current mainstream detection-based multi-target tracking strategy, and applied to the identification and tracking tasks of people and vehicles under actual road conditions. The first is the recognition and labeling of objects, by training the YOLOv5 network to identify and process the objects of each frame of the target video, the objects in the video are marked and labeled, and passed to the Deepsort network for tracking processing. In the Deepsort model, the detection box obtained in YOLOv5 and the predicted tracking box are entered into the Hungarian algorithm for linear allocation to correlate the ID between each frame, and the appearance information of the target is added to the inter-frame match to avoid loss after the ID is obscured. After the model is debugged, it is merged and packaged into the corresponding functions to complete the project objectives. In this paper, YOLOv5 single-stage object detection algorithm and Deepsort multi-target tracking algorithm are selected in combination with the multi-object detection algorithm, which improves the recognition and tracking effect of the algorithm in the multi-target tracking dataset.

Key words： YOLOv5 Deepsort Object recognition and tracking python

目录

一、 研究背景和项目目标

1.1 选题依据

计算机视觉（Computer Vision）是研究如何使计算机对图像数据产生智能化感知的一门学科。运动物体识别、检测与跟踪有广阔的运用前景，是计算机视觉研究的热点和重要领域。运动物体检测与跟踪算法不仅可以作为主要算法应用于视频监视、交通监管等方面，又能作为其他算法的基础应用于机器人技术、认证系统、人机交互等领域。同时，对运动物体检测与跟踪的研究能为更高层的计算机视觉分析提供基础的信息。因此，运动物体检测与跟踪一直是计算机视觉研究的热点和重要方向。

视频运动目标识别与跟踪的主要工作可以分为运动目标的识别和运动目标的跟踪两个方面，这两方面工作是一个承接的关系，相互依赖，同时也相互影响[1]。

运动目标识别作为视频运动目标识别与跟踪技术的第一部分，目的是在场景中识别出运动目标，并将其提取出来。视频图像序列中的运动目标跟踪一直是数字视频、数字图像处理和模式识别中一个重要的研究课题，运动目标跟踪同样也是衔接运动目标识别和运动目标行为分析和理解的一个重要环节。所谓运动目标跟踪，是在运动目标识别的基础上，利用目标的特征，选择使用适当的模板匹配等算法，在视频序列图像中寻找与目标模板最为相似图像的位置，最终达到跟踪的目的。在实际应用中，运动目标跟踪可以提供目标的运动轨迹，也为运动物体目标行为分析与理解提供了可靠的数据来源。运动目标识别是将运动目标背景图像中分离出来，是智能监控和运动图像分析的重要处理步骤，通过运动识别可以初步得到图像中的运动信息，提取视频序列图像中的运动目标并对目标进行初步定位，简化了后续的运动跟踪、分析的难度。运动目标识别是运动目标跟踪的基础和前提。

本项目要求我们能够基于 YOLOv5 和 Deepsort 实现视频或摄像头内的物体识别与跟踪；设计 GUI 界面，实时展示视频/摄像头中物体识别与跟踪的画面；能够在 GUI 界面中选中所有识别对象的其中一个或若干个，并跟踪被选定的对象；能够根据目标对象的移动调整摄像头或视频画面位置/放大对象所在区域，使得目标对象始终在画面中央。

通过本项目的研究，我们能够实现基于 YOLOv5 和 Deepsort 物体识别与跟踪，并结合现实场景构建一个智能交通监控系统，以解决现实生活中的相关问题，为人们的生产生活提供便利。而且随着智能视频监控、人脸识别门禁系统、自动驾驶、机器人视觉导航等贴近人们日常生活的视频数据的暴增，视频目标检测的研究具有更大的现实研究意义与应用价值。

1.2 业界现状介绍

目标识别与跟踪是指用计算机实现人的视觉功能，它的研究目标就是使计算机具有从一幅或多幅图像或者是视频中认知周围环境的能力（包括对客观世界三维环境的感知、识别与理解）。目标识别与跟踪作为视觉技术的一个分支，就是对视场内的物体进行识别与跟踪，如人或交通工具，先进行检测，检测完后进行识别，然后分析他们的行为。目前，国内外上许多高校和研究所，都专门设立了针对目标检测和识别的研究组或者研究实验室。

对于静止背景下视频序列图像的变化检测，主要有三种常用的方法：连续帧间差分法、背景差方法以及光流场法[2]。

（1）连续帧间差分法对于动态环境具有良好的适应性，即在运动目标和背景有着

较明显区别且目标运动较慢的时候比较有效，该算法计算简单，算法复杂度低，但是不能将目标的所有相关点提取出来。

（2）背景差分法能够比较完整地提取目标相关运动点，对于背景较为稳定的情形，识别效果比较好。而对于背景变化剧烈的场景，例如背景的抖动、光照、背景中新物体的加入等情况比较敏感。

（3）基于光流的运动变化检测采用了运动目标随时间变化的光流特性这--特 点，虽然该方法能够直接用于摄像机运动下的目标识别，但大多数光流算法运算量较大，如果没有特定的硬件支持，一般很难满足实时识别处理的要求。在实时的识别跟踪系统中，研究者们更热衷于使用帧差法或背景差分法来获得识别和跟踪的结果。

动态背景下，由于存在着目标与摄像机之间较为复杂的相对运动，所以动态背景下运动目标识别算法要比静态背景下运动目标识别算法复杂的多，常用的动态背景下运动识别算法是图像匹配法、光流估计法、全局运动估计等方法。

这些目标搜索算法包括两个方面，一种是通过预测目标在后续图像序列中可能出现的位置缩小目标搜索范围，另一种是为了提高搜索匹配的精度和搜索匹配的速度，从而缩短目标搜索的时间。常用的目标跟踪方法有：基于特征的跟踪，基于模型的跟踪，基于区域的跟踪，基于活动轮廓的跟踪[2]。

（1）基于特征的跟踪

视频序列图像中，由于相邻的两帧图像序列间的采样时间间隔小，可以认为这些个体的特征在运动形式上具有一定的平滑性，因此可以用点、直线、曲线等特征对运动目标进行跟踪。基于特征的跟踪方法具有算法简单的优点，在运动分析时可以不必区分运动物体是刚体或非刚体，但是该算法对复杂运动的跟踪效果差。

（2）基于模型的跟踪

基于模型的跟踪方法利用点、线以及区域将被跟踪目标拟合成几何模型，运动目标的跟踪过程变成了运动目标识别问题。与其它跟踪方法相比，这种方法含有高层的语义知识和描述，而且这种优势在复杂背景环境下尤为突出。该算法的缺点是计算量大，且需要大量关于拟定跟踪目标的先验知识。

（3）基于区域的跟踪

基于区域的跟踪方法是利用高斯分布建立目标模型和背景模型，目标的像素点被分配于不同的区域部分，通过跟踪各个小块区域来完成对整个目标的跟踪。这种方法绝大部分是通过基于光流的匹配模板来实现的，跟踪误差逐步积累，因此随着跟踪图像帧数的增加，误差也随之增大产生错误的匹配，一旦出现错误匹配，就不易进入正确的匹配中去。

（4）基于活动轮廓的跟踪

基于活动轮廓的跟踪方法是利用封闭的曲线轮廓描述运动目标，并且该曲线轮廓能够自动连续地更新。轮廓表达方法可以大大降低计算复杂度，如果跟踪初始阶段能够合理地分开每个运动目标，实现轮廓初始化，既使在出现部分遮挡的情况下也能连续地进行跟踪但初始化过程通常很困难。

视觉目标跟踪的主要任务是在一组图像序列中寻找目标，起初跟踪聚焦于单一运动目标的视觉任务需求，即单目标跟踪（Single Object Tracking，简称 SOT）。单目标跟踪 SOT 是指在视频首帧给出目标，根据上下文信息，在后续帧定位出目标位置，建立跟踪模型对目标的运动状态进行预测。解决 SOT 问题主要有两种方法：判别式跟踪及生成式跟踪，随着深度学习在图像分类、目标检测等机器视觉相关任务中的成功应用，深度学习也开始大量应用于目标跟踪算法中。另一类目标跟踪为多目标跟踪（Multi Object Tracking，简称 MOT）[3]。多目标跟踪（MOT）是指对视频中每一帧

的物体都赋予一个 ID，并将每个 ID 的行为轨迹画出来。多目标的跟踪策略主要有两种，一是基于检测的跟踪（Tracking by Detection， TBD），另一种是基于初始框的跟踪（Detection Free Tracking，DFT）。DFT 与单目标跟踪有相似之处，都需要在初始化目标时由人工标记视频第一帧中的目标，然后在检测的同时进行跟踪。由于人工初始化的方式无法标记第一帧中没有出现过的目标，而多目标跟踪本身包含新旧目标消失出现的场景，因此在跟踪过程中出现的未经人工初始化的新目标将无法被跟踪。TBD 是指基于检测进行跟踪，基于 TBD 策略的 MOT 包括一个独立的检测过程、一个检测结果和跟踪器轨迹连接的过程。TBD 跟踪目标的数量和类型都与检测算法的结果相关，通常检测结果具有一定的不可预测性，所以该方法的性能基本取决于检测成果的好坏。简单的在线和实时深度关联度量跟踪是基于 TBD 策略的 MOT 算法，通过设计检测结果和跟踪预测结果的关联策略实现跟踪[4]。此外基于 TBD 的算法还有降低检测不稳定性影响的基于深度学习候选人选择与再识别的实时多跟踪[5]。由于多目标跟踪任务的景像繁杂性，它的建模困难得多，目前遇到的最大挑战就是遮挡（Occlusion），即目标之间的彼此遮挡或环境对目标产生的遮挡。

1.3 本项目的目标

YOLO 是一种运行速度很快的目标检测 AI 模型， YOLOv5最大可处理 1280 像素的图像。当我们检测出图像中目标后，把视频分解成多幅图像并逐帧执行时，可看到目标跟踪框随目标移动。Deepsort 是实现目标跟踪的算法，从 sort （simple online and realtime tracking）演变而来，其使用卡尔曼滤波器预测所检测对象的运动轨迹，匈牙利算法将它们与新的检测目标相匹配。借助跟踪器 Deepsort 与检测器 YOLOv5，可以打造一个高性能的多目标跟踪模型。

本项目计划基于 YOLOv5 和 Deepsort 实现了一个多功能智能交通监控系统。首先，能够实现视频内的物体识别与跟踪，并设计 GUI 界面，实时展示视频中物体识别与跟踪的画面，同时能够在 GUI 界面中选中所有识别对象的其中一个或若干个，并跟踪被选定的对象。其次，本项目可以区分行人和车辆，并对不同方向行驶的车辆实现计数功能。

本项目可以应用到实际场景中，通过统计经过十字路口、丁字路口车辆流动繁忙的交通场合的道路情况，可以合理安排交通警察或保安人员的工作时间和工作额度，大大提高城市通勤效率，同时对行人车辆的识别追踪，也可为交通事故的追踪调查提供便利。

二、 项目总体设计

2.1 问题分解

本项目分为物体的识别和追踪，分别由 YOLOv5 和 Deepsort 网络完成，再进行整合实现相应功能。

首先是物体的识别和标记，通过训练 YOLOv5 网络对目标视频每一帧的对象进行识别处理，将视频中的对象标记出来并打上标签，传递到 Deepsort 网络中进行跟踪处理。

在 Deepsort 模型中将 YOLOv5 中得到的检测框与预测的跟踪框输入到匈牙利算法中进行线性分配来关联每帧间的 ID，同时将目标的外观信息加入到帧间的匹配中，避免在 ID 受到遮挡后出现丢失

模型调试完成后将其合并打包成相应函数，完成项目目标

2.2 模块划分

（1） YOLOv5

YOLOv5 模块在本项目中负责对视频中的对象进行识别和标注，从而在下一步中进行预测和追踪。

（2） Deepsort

Deepsort 在本项目中将上一步标注好的对象输入网络中进行预测和分析，得到对应物体的移动轨迹，从而实现对物体的追踪功能。

三、 项目关键技术

3.1 YOLOv5

YOLOv5 是一种单阶段目标检测算法，该算法主要是在 YOLOv4 的基础上添加了一些新的改进思路，使其速度与精度 都得到了极大的性能提升。整个 YOLOv5 网络结构空语分成四部分：输入端、Backbone、Neck、Prediction。

输入端的主要操作任务也就是在用户输入多个数据的同时需要进行各种数据增强，因此 YOLOv5 继承了 YOLOv4 所使用的 Mosaic 数据增强方式，对图片进行随机组合缩放、随机组合裁剪、随机组合排布等多种方式进行随机拼接，实现了既对数据集进行增强又同时解放了对于 GPU 的依赖。 YOLOv5 针对不同的数据集，采用自定义不同长宽锚框的自适应锚框，同时运用自适应图片缩放，达到数据增强的目的。

Backbone 主要被划分为了两大结构：Focus 和 CSP，Focus 结构就是 YOLOv5 新提出的一种结构，可以将原始的设定为 608*608*3 的图片大小，经过切片后改变为 304*304*12 的特征结构图，利用 32 个卷积核的卷积，得到 304*304*32 的图片特征结构图。CSP 模块借鉴之前 CSPNet 的网络结构，由卷积层和 X 个 Res unint 模块通过 concate 构成，每个 CSP 模块前面的卷积核的大小都被设定成是 3*3，strIDe=2，因此它们可以起到对图片进行下采样的作用，降低了内存的使用成本。YOLOv5 中的 CSP 结构 如 YOLOv4 类似，但分别划分了两处不同的使用范围，CSP1_ X 类型结构广泛应用于现在 Backbone 主干网络，另一种 CSP2_X 类型结构则广泛应用于 Neck 中。

YOLOv5 中的 Neck 网络采用 FPN+PAN 的结构，FPN 是自顶向下的，将高层的特征信息通过上采样的方式进行传递融合，得到进行预测的特征图，而 PAN 特征金字塔则自底向上传达强定位特征，两两联手，从不同的主干层对不同的检测层进行参数聚合。相较于 YOLOv4 其改进的部分在于，通过借鉴 CSPnet 网络而设计 CSP2 网络，进一步加强网络特征融合。

目标检测任务的损失函数一般由分类损失回归函数和回归损失函数两个子部分组合构成，YOLOv5 的 Prediction 中的端口损失采用了 GIOU_Loss 函数作为回归损失函数，计算不同状态下的 GIOU 的数值 ，解决了在边界框不完全重合的问题 ；同时利用加权 DIOU_nms 实现非极大值抑制，抑制冗余框，只保留我们所需要的框，可以对被遮挡的物体进行更为有效地识别。

3.2 Deepsort

Deepsort 是在整个 Sort 算法在对目标追踪基础上的改进，整体设计框架没有大的修改，还是完全延续了卡尔曼滤波加匈牙利算法的设计思路，在这个算法基础上又新增加了一个 Deep Association Metric，此外还重新加入了外观图像信息以便于实现

了如何在较长时间内对被遮挡的目标进行跟踪时的问题。其主要特点之处在于，Deepsort 加入了更多外观特征信息，借用了新的 ReID 应用领域特征模型来快速提取外观特征，减少了 REID 特征转换的发生次数。

在跟踪方面，Deepsort 采用的级联匹配算法，可以针对每一个检测器都会分配一个跟踪器，每个跟踪器会设定一个 time_since_update 参数；对于运动信息和外观信息的变换和模糊问题，利用马氏距离与余弦距离计算 ；添加了深度学习 ReID 的模块，有利于更好地区别不同的人物或物体。

Deepsort 的运算完成流程，第一步依赖卡尔曼滤波器预测轨迹 Tracks；第二步 ，使用匈牙利算 法将通过预测轨迹得到的帧中轨 迹数据 tracks 和当前帧中的轨迹 detections 组合进行匹配（ 包括级联匹配和 IOU 匹配 ）；第三步，卡尔曼算法滤波器的更新。Deepsort 算法通过将使用目标跟踪检测的算法（如 YOLO） 计算得到的目标检测框与 之前预测的目标跟踪框的 iou（称为交并比）进行输入组合。输入到匈牙利这个算法中进行线性分配来直接关联这个帧中的 ID，目标的物体外观位置信息可以加入连接到帧间匹配的过程计算中，这样在目标被物体遮挡但后续目标再次出现的实际情况下，还可以正确的来匹配这个帧间 ID，在进行实时检测目标追踪过程中，可以改善在有遮挡目标情况下的实时目标自动 追踪检测效果。同时，也大大减少了目标 ID 之间跳变频繁的问题，达到持续跟踪的目的。

四、 项目实现

detector 模块：下载训练好的 YOLOv5m.pt 模型，对图片内容进行预测，可以区分六个类；

tracker 模块：主要实现画框、更新框、修正框的误差三个功能（Deepsort 进行的预测非常准确，对比简单的 cv2 里 KCF 的追踪效果十分明显，可以查看为了对比添加的“cv2_mouse.py”）；

Deepsort_keyboard.py/Deepsort_YOLOv5_all/cv2_mouse 模块：就是主程序，“Deepsort_YOLOv5_all”是源程序，稍微修改了判定条件、判定内容，是先画两个判断线，用 detector 得到追踪目标的 bbox 信息，然后再用 tracker 进行每一帧 bbox 的更新，最后根据同一个 ID 撞两条线的顺序进行方向检测和计数。另外两个程序是根据源程序修改出来的，Deepsort_keyboard 是对画面内的 ID 可以通过键盘按键进行选择，就是添加了标注单一目标进行追踪的功能，通过键盘反馈与 ID 进行对比实现。cv2_mouse 是根据鼠标画的 bbox 进行追踪（没有使用 Deepsort），就是简单的使用 kcf 实现，画框也是自己简单的“预测”画框。

五、 项目测试

1) 源程序修改后能正确运行任意视频，并且标注画面中所有目标的类别和序号，在食堂到宿舍视频中能识别人、自行车、电瓶车、汽车；

图 1 标注目标类别序号功能测试图

2) Deepsort_keyboard 能够根据键盘按键选择标号 0~9 的单一目标追踪，但是由于技术原因，识别类型显示不出来，但是能显示 ID。如图是从所有目标中选择 9 号；

图 2 单一目标追踪功能测试图

3) 单一的 cv2 主要是用作对比，能够发现不用 Deepsort 后会出现追踪框移动缓慢，容易被目标“甩开”，并且遇到遮挡后会丢失目标。

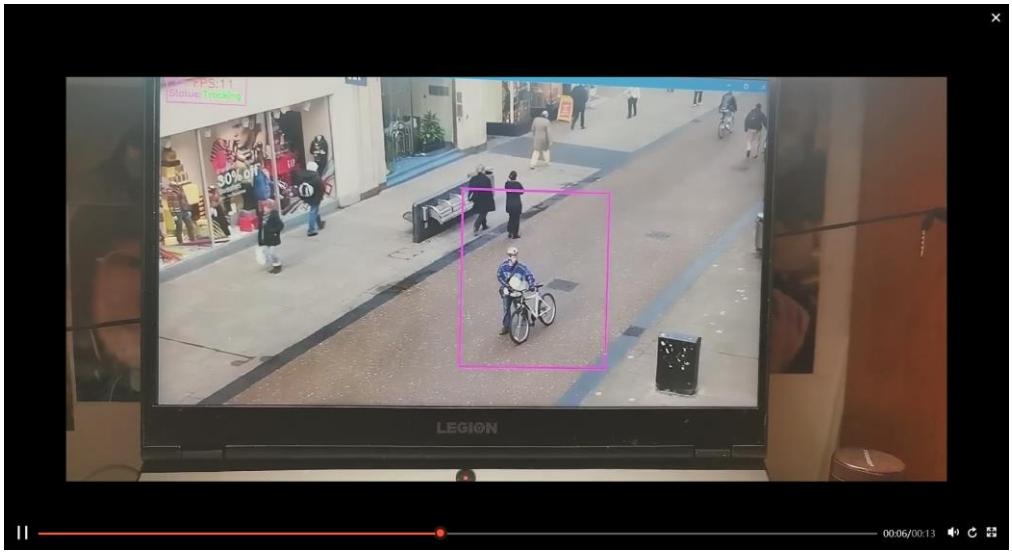

图 3 Deepsort 功能测试图

六、 项目管理

6.1 团队人员组成

表 1 人员组成

姓名	班级	学号
李瑞昊	电气 2009 班	U202012480
姚佳琦	电气 2012 班	U202010069
俞晨欣	电气 2012 班	U202011296
鲍林奕	电气 2005 班	U202012349
魏子健	电气 2005 班	U202012350

6.2 任务分工

俞晨欣：搜集项目代码资源，制作结题答辩 PPT

魏子健：进行项目代码的调试，分工撰写课题设计报告

李瑞昊：进行选题答辩，收集项目相关资料，分工撰写课题设计报告

姚佳琦：制作选题 PPT，收集项目相关资料，分工撰写课题设计报告

鲍林奕：学习其他队员找到的源代码，并修改添加内容，实现部分计划的功能，进行结题答辩

七、 总结与反思

7.1 遇到的困难

1、对 github 使用不熟练，github 上传大文件时遇到困难，用了 lfs 还是不行；

2、用键盘选择追踪目标时只能使用 0~9 简单按键，比如 11 这种两位数按键不会实现；

3、想用鼠标选定单一目标框，但是过程中遇到很多不懂得问题；

4、无法自己游刃有余的将 YOLOv5 和 Deepsort 的具体功能进行修改，比如针对学习的源码，只能做一些皮毛上的修改，想要修改 detect 文件里的具体内容就容易出错，这也就导致只能修改到对 0~9 的目标进行追踪。

7.2 收获与感悟

1、熟练使用 github 非常重要；

2、Deepsort 预测很厉害，在学习 Deepsort 过程中还了解了一些 strongsort 这些更厉害的技术；

3、要善于寻找数据，比如别人训练好的 YOLOv5 权重，同时还发现了 yolov7 等更高级的东西；

4、集诸家之长必有长进，多看看别人的代码是最快的学习方式；

5、熟练掌握 debug 功能，能帮助我们更快的发现错误；

6、代码一定要标注清楚，命名要主流要有含义，在读代码的时候深深感受到这一点的重要性，会让同伴非常愉快。

八、 参考文献

[1]鹿雪娇.基于视频图像的运动物体识别与跟踪技术研究[D].大庆石油学院，2009.

[2]何斌，马天予，王运坚等.VC++数字图像处理[M].北京：人民邮电出版社，2001：15~18.

[3]Luo W， Xing J， Zhang x， et al. Multiple object tracking： A literature review [J]. Artificial intelligence， 2021， 293： 103448.

[4]Wojke N， Bewley A， Paulus D. Simple online and real-time tracking with a deep association metric [C]. Proceedings of the 2017 IEEE International Conference on Image Processing， Beijing， China， 2017：3645- 3649.

[5]Chen L， Ai H， Zhuang Z， et al. Real-time multiple people tracking with deeply learned candIDate selection and person re -IDentification [C].Preceedings of the IEEE International Conference on Multimedia and Expo， San Diego， CA， USA， 2018：1-6.