Computer Science and Application 计算机科学与应用, 2024, 14(12), 11-21
Published Online December 2024 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2024.1412235

Hans汉斯

基于改进YOLOV3的商场人员类型及流量分析

杨贤玉

上海理工大学光电信息与计算机工程学院，上海

收稿日期：2024年10月30日；录用日期：2024年11月26日；发布日期：2024年12月4日

摘　要

随着人口数量在全世界范围内快速增长，对特定场景如商场中人口数量的控制就成为了重要的研究课题。近年来，公共场所当中因为人流密集和大量拥挤而产生的对社会稳定有所影响的事故屡见不鲜。传统的人力控制方法费时费力，近年来随着人工智能和高清摄影机技术的不断完善，利用计算机对商场实行监控成为了主流趋势。然而目前的探头技术仍然只能实现观察和监测的作用，无法对一段监控视频中的人流进行数量上的计算。本文意在实现对商场人员的人流进行数据化以及可视化的监测。与此同时，尽可能地对商场的特定时段的不同人员类型进行分析，这样不仅能很好地维持商场秩序和人流量控制，同时还为商场管理人员提供了思路，可以通过对不同时段的经营策略进行针对化的调整，来更好地对商场进行管理和经营。YOLO算法是一种经典的目标检测算法，本文将使用较新的YOLOV3算法对商场的人员类型和流量进行检测，建模以及分析。并给出了多个维度的模型参数的性能指标图，以此来表明该算法的可行性。

关键词

人流检测，类型检测，人工智能，YOLOV3算法，模型参数

Shopping Mall Personnel Type and Traffic Analysis Based on Improved YOLOV3

Xianyu Yang

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Oct. 30th, 2024; accepted: Nov. 26th, 2024; published: Dec. 4th, 2024

Abstract

With the rapid growth of population around the world, the control of population in specific scenarios such as shopping malls has become an important research topic. In recent years, accidents

文章引用: 杨贤玉. 基于改进 YOLOV3 的商场人员类型及流量分析[J]. 计算机科学与应用, 2024, 14(12): 11-21.
DOI: 10.12677/csa.2024.1412235

affecting social stability in public places due to the dense flow of people and mass crowding are not uncommon. The traditional manual control method is time-consuming and laborious. In recent years, with the continuous improvement of artificial intelligence and high-definition camera technology, the use of computers to monitor shopping malls has become the mainstream trend. However, the current probe technology can only realize the role of observation and monitoring, and can not calculate the number of people in a surveillance video. The purpose of this paper is to realize the digital and visual monitoring of the flow of shopping mall personnel, at the same time, as far as possible to analyze the different types of personnel in the specific period of the shopping mall, so as not only to maintain the order of the shopping mall and the control of the flow of people, but also to provide ideas for the shopping mall management personnel, through the adjustment of the business strategy of different periods of time, to better manage and operate the shopping mall. YOLO algorithm is a classic target detection algorithm. In this paper, the relatively new YOLOV3 algorithm will be used to detect, model and analyze the personnel types and traffic of shopping malls. The performance index graphs of model parameters with multiple dimensions are given to show the feasibility of the proposed algorithm.

Keywords

Flow Detection, Type Checking, Artificial Intelligence, YOLOV3 Algorithm, Model Parameters

1. 引言

每逢假期、节日或大型活动，公众场所的人流量激增。为有效避免群体性活动中意外事故的发生，基于人工智能的视频监控技术已被广泛应用于车站、机场、地铁及大型的购物中心等场所。人流量检测是智能监控中的一个重要应用，能为公共场所的管理提供重要的数据。在实际的应用场景中，该技术仍面临着许多挑战。例如，行人之间的遮挡、行人行走轨迹的随机性和监控视频的低分辨率都将影响检测的精度[1]。尽管如此，在行人流量较为密集的大型购物中心，各商家要想能够取得最高的利润，就必须对不同时刻途经商场的人员进行检测和类别分析，据此来推出最为优良的经营策略。因此，目标检测技术一直以来都备受国内外研究人员的关注。除开传统的目标检测技术，基于深度学习的目标检测算法在精度和速度上都有大幅度提升。在此背景下，本文设计了一种基于深度学习的人员类型分析和流量统计系统，能够对监控视频当中的行人进行检测和类型声明，并且生成人流量统计图，管理者可以据此统计数据推出合适的管理策略。

计算机视觉隶属于人工智能范畴，目标检测为计算机视觉领域基本的视觉识别问题之一。人是构成社会的主体，行人检测的主体目标是众行人，因此以目标检测技术为基础的行人检测在计算机视觉领域中扮演着重要的角色[2]，在智能交通、智能安全防范视频监控、图像检索等传统领域得到广泛应用。一些新兴应用市场如基于航拍图像的行人及受害者检测、Google 街景图中的行人定位、家用服务机器人等，场景中包含行人、车辆、道路、交通标志牌等多样化目标。在这些繁杂的情境中，对人体行为的分析并不能像之前那样对人体位置进行手动标记，而必须要求程序自动在视频或者图片中给出人体的位置。因此，行人检测的技术研究逐渐得到了科研人员的重视。目前，行人检测的应用场景有自动驾驶领域，智能机器人领域、安防监控领域等[3]。

目标检测(Object Detection)是对图像中的感兴趣目标进行识别和定位的技术，解决了图像中物体在哪

里和是什么的问题[4]。已有研究表明，可靠的目标检测算法是实现对复杂场景进行自动分析与理解的基础。因此，图像目标检测是计算机视觉领域的基础任务，其性能好坏将直接影响后续的目标跟踪、动作识别以及行为理解等中高层任务的性能，进而决定了人脸检测、行为描述、交通场景物体识别、基于内容的互联网图像检索等后续 AI 应用的性能[5] [6]。随着这些应用渗透到人们生活的方方面面，目标检测技术在一定程度上减轻了人们的负担，极大程度上改变了人们的生活方式。目标检测的基本任务是需要判别图片中被检测的目标类别，同时需要使用矩形边界框来确定目标所在的位置及大小，并给出相应的置信度。作为计算机视觉领域的一个基本问题，目标检测也是许多计算机视觉任务如图像分割、目标追踪、图像描述的基础。

2. 系统方案

本论文通过深度学习有关算法对商场中行人类别和流量进行分析，故总体方案分为两大部分：行人检测和人流量统计两大板块。两部分的开展思路和方法有共同点，两者都需要实现制作行人检测和流量统计有关的数据集，随后对数据集进行多次训练，经过训练，验证，检测等步骤得出深度学习模型，随后各自开展功能。在这当中，由于人流量统计除了需要人流量检测之外，还需要实时行人跟踪，同时需要利用卡尔曼滤波提高统计精度，于是该部分将花费较多时间。

2.1. YOLOV3 算法

YOLO (You Only Look Once)算法是一种典型的单阶段(One-Stage)目标检测算法。其主要意义是没有真正去掉双阶段算法当中所使用的候选区域，而是将候选区和目标分类两步合二为一，YOLO 算法的主要优势在于对图片进行以此观察就能知道有哪些对象并确定它们的位置[7] [8]。这其中，YOLO 算法经过了多次的迭代，直到现在较为常用的 YOLOV5。

YOLOV3 将输入图像变为 416 × 416 大小后输入到 Darknet-53 主干特征提取网络中，通过 Darknet 网络得到三种不同尺度的预测结果，每个尺度都对应 N 个通道，包含着需要预测的信息。对比 YOLOV1 和 YOLOV2，YOLOV3 明显有更多的预测结果。同时与传统的卷积神经网络相比，取消了池化层和最后的全连接层，采用步长为 2 的卷积下采样，显著减少计算量的同时，还保留了图像更多的相关信息[9] [10]。将输入图像分成 S × S 的网格，每当目标的中心点落入网格当中，该网格就将用于预测物体的先验框，这时这个网格的权重为 1，其他网格为 0 [11]。每个网格可以预测三个先验框，每个框包含了 5 + N 个值，这其中包括了先验框的中心位置(x, y)，先验框的宽和高(w, h)，以及先验框的置信度 C。而最后的 N 则代表了数据集的类别数[12] [13]。经过一系列下采样、卷积等操作，可以获得图片不同层次的位置和语义信息。网络输出层采用 3 层进行预测，最终使用非极大值抑制确定结果[14]。

2.2. 维度聚类算法——K-means++算法

实际应用当中，可能会出现目标提取不明显，或者预测结果与实际结果相差较大的情况。而维度聚类算法的提出就是为了解决这类问题。所谓聚类就是将数据集根据某种或某几种维度划分为若干个不相交的子集，以提取目标的特征[15]。YOLOv2 已经开始采用 K-均值聚类算法(K-means clustering algorithm)聚类得到先验框的尺寸。

means 算法是最常用的聚类算法，其主要思想为：假定给定数据样本 X，包含了 n 个对象 $X=\{X_1, X_2, X_3, \cdots, X_n\}$，其中每个对象都具有 m 个维度的数据。K-means 算法的目标是将 n 个对象依据对象间的相似性聚集到指定的 k 个类簇中，每个对象属于且仅属于一个其到类簇中心距离最小的类簇中。这里的距离用欧式距离表示，其计算公式如下：

$$\mathrm{dis}\left(X_j, C_j\right)=\sqrt{\sum_{i=1}^{m}\left(X_{it}-C_{jt}\right)^2} \tag{1}$$

其中，X_i 表示第 i 个对象$(1\le i\le n)$，C_j 表示第 k 个聚类中心$(1\le j\le k)$，X_{it} 表示第 i 个对象的第 t 个属性，C_{jt} 表示第 j 个聚类中心的第 t 个属性$(1\le t\le m)$ [16]。

依次比较每一个对象到每一个聚类中心的距离，将对象分配到距离最近的聚类中心的类簇中，得到 k 个类簇$\{S_1, S_2, S_3, \cdots, S_k\}$。K-means 算法用中心定义了类簇的原型，其类簇中心就是类簇内所有对象在各个维度的均值。

在使用的过程当中需要优先确定若干组尺寸的先验框，一般的均值聚类算法选择了 9 组维度尺不同的先验框，它们分别为：

(10, 13), (16, 30), (33, 23), (30, 61), (62, 45), (59, 119), (116, 90), (156, 198), (373, 326).

但这些用于 COCO 和 VOC 数据集的 anchor 无法用于本文中的系统。与此同时，K-means 聚类算法在使用的过程中会出现初始中心点选择较为随机导致效果不好的情况。因此，本文选择了更为先进的 K-means++算法并且选择了九组不同的 anchor [17]。它们分别是：

(66, 13), (136, 22), (162, 30), (218, 30), (217, 38), (220, 47), (212, 64), (347, 40), (444, 62).

相比于传统方式的手动选择先验框，聚类可以促进网络结构的收敛并且能够有效改善训练过程中的梯度下降的现象。集群评估标准如下：

$$d\left(\text{bbox, center}\right)=1-\mathrm{IOU}\left(\text{bbox, center}\right) \tag{2}$$

其中 $d\left(\text{bbox, center}\right)$ 是边界盒和中心盒之间的距离，$\mathrm{IOU}\left(\text{bbox, center}\right)$ 是两个盒子之间的交点除以并点。这种聚类方法可以产生较大的交集比和较小的聚类边界框之间的距离。

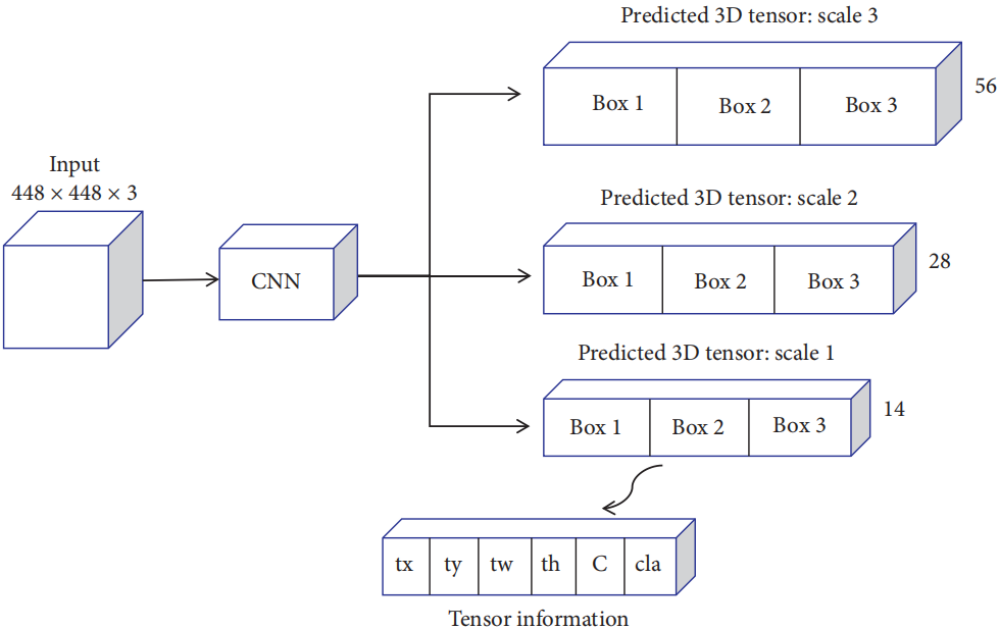

Figure 1. Tensor information in different dimensions
图 1. 不同维度下的预测信息

如图 1 所示，每一个比例尺的特征图中的每个网格预测了三个先验帧。它包含了$(4 + 1 + N)$维向量，用来表示边界框的中心坐标(x, y)，宽度和高度(w, h)和置信度 C，N 表示数据集中类别的数量：

$$b_x = \sigma(t_x) + c_x \tag{3}$$

$$b_y = \sigma(t_y) + c_y \tag{4}$$

$$b_w = p_w e^{tw} \tag{5}$$

$$b_h = p_h e^{th} \tag{6}$$

$$P_r(\text{object}) * \text{IOU}(b, \text{object}) = \sigma(t_0) \tag{7}$$

$$\text{IOU}(b, \text{object}) = \frac{\text{area}(b) \cap \text{area}(g)}{\text{area}(b) \cup \text{area}(g)} \tag{8}$$

式中 b_x 和 b_y 为集合框架检测结果，t_x 和 t_y 为各目标边界盒先验预测中心坐标。在网络中激活函数后，得到与边界盒中心坐标对应的偏移量 $\sigma(t_x)$ 和 $\sigma(t_y)$，b_w 和 b_h 分别为检测到的框的宽度和高度。p_w 和 p_h 分别为宽度和高度的权重矩阵，e^{tw} 和 e^{th} 为边界框宽度和高度的比率。$\sigma(t_0)$ 为目标预测评分的偏移量，由目标预测概率与 IOU 相乘得到。面积 b 为预测框面积，面积 g 为实地面积，交点面积之比为 IOU [18]。如图 2 所示。

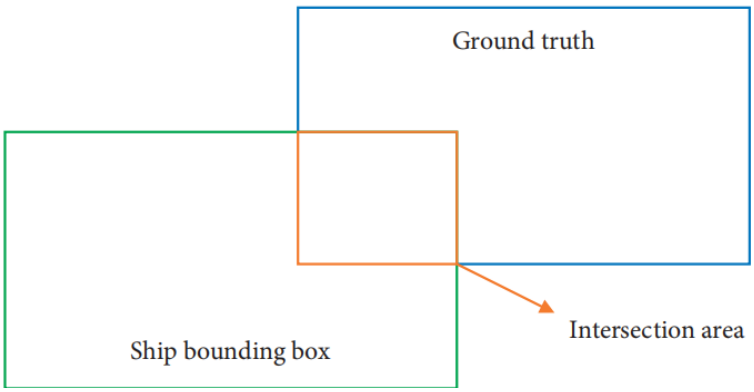

Figure 2. Target prediction principle
图 2. 目标预测原理

2.3. 网络结构优化

对于在不同环境下的目标检测，特征提取是人员分类当中一个重要的步骤。特征提取网络的卷积层可以有效分析目标的特征。如图所示，将残差网络添加到 YOLOV3 网络中，并多次使用残差的跳连接。残差网络可以解决特征提取过程中梯度难以下降的问题，加快收敛速度，减少误差。图片的大小增加到 448 × 448。步长为 2 的卷积在网络中用来对图像进行下采样，分别为 32，16，8 倍。最后一层的输出是轻量级改进的。对原输出预测层 3 × 3 和 1 × 1 的卷积进行修剪，只有 3 × 3 的卷积可以用于预测，这样可以减少网络运行，避免过拟合，使模型具有更好的泛化能力[19]。最后，得到三个尺度上的特征图，该特征信息在矩阵中的表示形式如下：

$$f_a^b = f\left(\sum_t S_t^{b-1} * P_{ta}^b + W_a^b\right) \tag{9}$$

其中 f_a^b 为第 b 个网络和第 a 个输出之间的对应关系，f 为卷积层 b 的激活函数，S_t^{b-1} 为 $b-1$ 卷积网络层对 t 目标的映射，P_{ta}^b 为第 t 层与第 a 个特征层之间的权重矩阵，参数 W_a^b 为第 a 个输出特征在第 b 层卷积网络层的偏置。改进后的网络结构如图 3 所示。

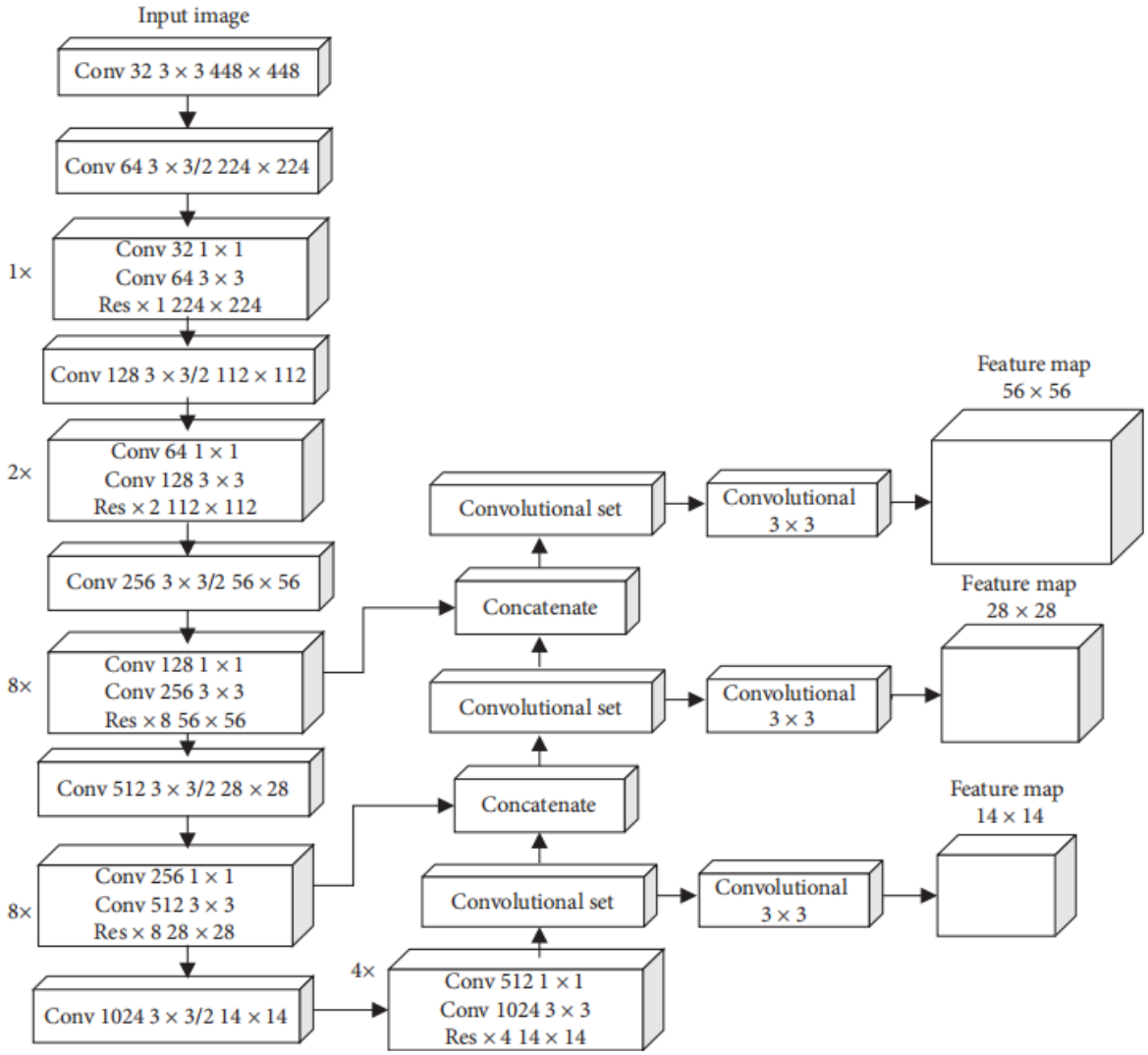

Figure 3. Structure of network
图 3. 网络结构

2.4. 行人跟踪

人流统计的目的是要检测出一段视频当中的行人流量情况，仅仅将该段视频中的行人检测出来显然不行，要达到计数功能，就需要设置行人跟踪系统。行人检测模型在检测出行人之后需要将输出结果输入到目标跟踪算法中，行人检测的输出为具体哪些区域有行人，需要通过跟踪算法来确定两张相邻图片的行人。例如，在第二帧图片当中判断有哪些行人在上一帧当中出现，哪些人是新出现的。这是行人跟踪最主要的目的。最简单的判断方法就是利用 ID 号，如果出现新的行人，就给其分配一个新的 ID，如果没有就仍是老 ID，若上一帧的行人在该帧当中消失了，就需要标记为跟踪丢失，其 ID 号不再使用。行人跟踪的效果好坏很大程度上决定了行人检测效果的好坏。

本项目拟采用 SORT 进行目标的跟踪，可根据跟踪目标数目多少分为单目标跟踪和多目标跟踪。而 SORT 属于多目标跟踪算法。SORT 算法的速度非常快，能够实现实时跟踪的效果。虽然只使用了卡尔曼滤波和匈牙利算法的跟踪组件的基本技术组合，但是这种方法达到了较高的精度。多目标跟踪问题类似于一个数据关联问题，将每一帧中被检测到的目标进行关联，以某一指标计算其相似度，判断不同帧上

的两个目标是否是同一个目标，其目标就是为了实现视频中帧间的连续性的检测。SORT 模型在跟踪目标时忽略检测组件的颜色分布纹理等外观特征，只使用边界框的位置和尺寸进行目标跟踪。鉴于目标跟踪的原理为上一帧图片检测的目标与下一帧目标数据相匹配，故可将每个目标的状态建模为：

$$X = \left[u, v, s, r, \dot{u}, \dot{v}, \dot{s}\right]^T \tag{10}$$

其中 u 和 v 代表的是目标中心水平和垂直的坐标，s 代表目标的大小，r 代表目标边界框的长宽比，这个值假设恒定不变，三者的导数则表示目标在水平方向、垂直方向和目标大小的变化。当进行目标关联时，使用卡尔曼滤波器，其原理是使用上一帧中目标的位置信息预测下一帧中的这个目标的位置。若上一帧中没有检测到下一帧中的某个目标，那么对这个目标重新初始化一个新的卡尔曼滤波器，在关联完成了之后，卡尔曼滤波器需要通过下一帧中该目标的位置来更新。SORT 算法中的代价矩阵为上一帧的 M 个目标与下一帧当前帧的 N 各目标两辆目标之间的 IOU (Intersection over Union)，若连续若干帧都没有实现两者之间的匹配，则认为目标消失。

3. 模型训练

完成数据预处理工作以后，需对数据集进行划分，共分为训练集(Train)，验证集(Val)和测试集(Test)。训练集的作用是帮助使用者训练模型，通过训练集的数据来确定拟合曲线的参数。此处的拟合曲线主要记录了多次训练的结果，以此来确定性能最优的模型。后文将介绍用来判定模型性能好坏的各类参数。验证集可用来做模型选择(Model Selection)，即做模型的最终优化及驱动，用来辅助模型的构建。测试集则是用来测试已经训练好的模型的精确度，在训练模型的时候，参数全是根据现有训练集里的数据进行修正、拟合，有可能会出现过拟合的情况，即这个参数仅对训练集里的数据拟合较准确，如果出现一个新数据需要利用模型预测结果，准确率可能会很差。测试集的作用是为了测试最终的检测模型对新样本的判别能力。

值得注意的是，如果在划分数据集的时候将测试集的比重划分的较小，则在最终评估的时候会出现准确度较低的问题，因此在划分数据集的时候需要进行权衡。在当前的大数据时代，百万级别的数据集，常见的比例最多可以达到 99.5:0.3:0.2，常见的比例一般是 98:1:1，但是在本次项目当中，数据集属于小规模数据集，数量级仅仅在万级，一般分配比例为训练集和测试集的比例为 7:3 或是 8:2，为了进一步降低信息泄露同时更准确反应模型的性能，更为常见的划分比例是训练，验证，测试的比例为 6:2:2，本次检测也选用这个比例。将处理好的数据集按该比例划分以后，即可开始训练。

现介绍判定模型性能好坏的参数。首先介绍混淆矩阵，混淆矩阵(Confusion Matrix)也被称作误差矩阵，是表示精度评价的一种标准格式，用 $n \times n$ 的矩阵来表示，是一种被广泛用于监督学习当中的可视化工具。主要用于比较分类结果和实际测得值，可以把分类结果的精度显示在一个混淆矩阵里[20]。

对于一个二分类系统，将实例分为正类(Positive)，负类(Negative)，则按照排列组合，模式分类器共有四种分类结果：*TP* (*True Positive*)：正确的正例，一个实例是正类并且也被判定成正类；*FN* (*False Negative*)：错误的反例，本为正类但判定为假类，简而言之就是发生了漏报的现象；*FP* (*False Positive*)：错误的正例，本为假类但判定为正类，也就是误报的情况出现；*TN* (*True Negative*)：正确的反例，一个实例是假类并且也被判定成假类。机器学习当中的大部分参数都与这四类结果有关[21]。

1) 准确率：所有预测正确(正类负类)的占总的比例，计算公式如下：

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{11}$$

2) 精确率：也叫查准率，即正确预测为正的占全部预测为正的比例，也就是传统意义上的预测正确的比例，计算公式如下：

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

准确率和精确率都可以用来表示深度学习模型的精确程度，但本次研究将主要以观测精确率为主。这其中的原因在于，虽然准确率能够判断总的正确率，但是在样本不均衡的情况下，这并不是一项很好的来衡量结果的指标，因为在样本不平衡的情况下所得到的高准确率没有任何意义。精确率则更能体现模型检测准确率，该指标代表着对正样本结果中的预测准确程度，尽管可能会遗漏一些样本中的个体，但能够最大限度保证现有的预测保持精确，这是本研究当中希望得到的结果，因此将精准率 P 参数作为一个重要观测指标。精确率和准确率的数值都应在 0~1 之间变化，越接近 1 则代表模型性能越好。

3) 召回率：另一个重要指标，表示正确预测为正的占全部实际为正的比例。

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

P 参数和 A 参数主要是为了观测检测的准确度，R 参数召回率是为了保证每个样本都已被找到，虽然可能会有更多的误检，但在某些场景中，该参数也十分重要。应用场景例如癌症筛查、排查安全隐患。由于本产品应用于商场当中，有一定的排查安全隐患的需求，所以召回率 P 也是重要观测指标。和 A 参数，P 参数相同，召回率也是越接近于 1 表示模型的性能越好。

4) mAP (mean Average Precision)：均值平均精度，另一个衡量目标检测模型好坏的常用指标。通过计算得出召回率 R 和精确率 P，根据两者的数值绘制出 PR 曲线，经过多次实验不难看出两者负相关，绘制出来的图像与 P 轴，R 轴围成的面积称为 AP 即平均精度，每次检测都能计算出一个 AP 值，对所有 AP 值求平均则可得到模型的 mAP 值，该值能够更精确地表明模型的性能好坏，该数值也是越接近 1 越好。该指标为本研究中一个非常重要的观测指标，用来观测训练得出模型的性能好坏。

5) 损失率(Loss)：损失率主要用来表现预测与实际数据的差距程度。一般目标检测任务当中的损失包括两个部分，一部分为回归损失，衡量目标实际的空间与预测位置之间的距离。第二部分为分类损失，衡量的是一张图中每个正负样本分类是否正确，这两部分进行一定的运算结合以后一起作为最终的损失函数。本论文的研究结果当中将会比较现实损失率的变化趋势，并以此为据来判断模型的好坏程度。

$$L_{reg} = \frac{1}{n_{\text{postives}}}\left(\sum_{\text{postives}} \text{Smooth}_{L1}\left(y_{\text{p\_loc}}, y_{\text{t\_loc}}\right)\right) \tag{14}$$

通过对以上几项指标的观测，得出性能最好的模型，利用该模型进行调试。在 PC 端上模拟现实情况，实现人员检测。在这一部分当中应当分为多种检测模式，如读取图片，读取视频以及通过检测摄像头实现检测。至此，行人检测部分方案设计完成，以下是人流统计部分。

4. 训练结果

该部分将展示训练所得的模型及相关性能指标，PR 参数的关系及混淆矩阵可见图 4，图 5。

4.1. P-R 曲线

如下图所示，该曲线中相对两条较细的曲线为本次训练中所选取的类别，该曲线中的每个点都有对应的坐标(R, P)，通过利用 mAP 的计算方法，得到对应的精度，即为曲线当中较粗的蓝色曲线。

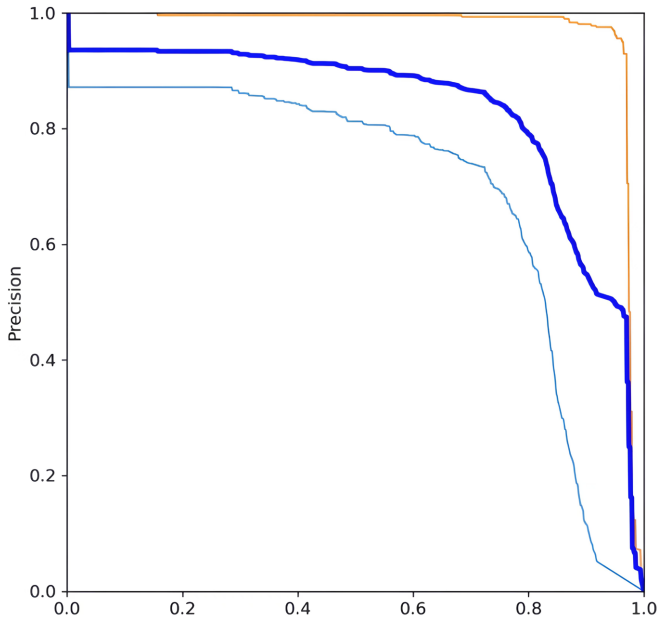

Figure 4. P-R curve
图 4. P-R 曲线

4.2. 混淆矩阵

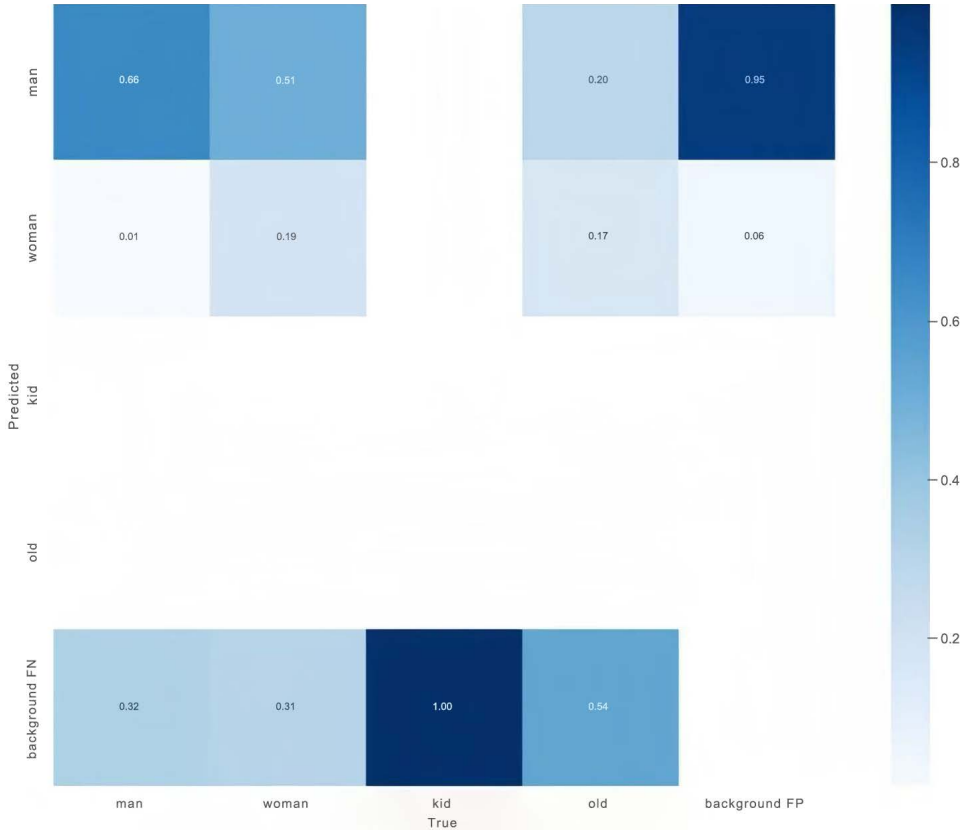

Figure5. Confusion matrix
图 5. 混淆矩阵

本文通过调整训练轮数及相关参数，得出了四种不同的模型，其结果见表 1。

Table 1. Model parameters
表 1. 模型参数

模型序号	训练轮数(Epoch)	训练规模(Batch-Size)	均值平均精度(mAP)	损失率(Loss)	训练时间
模型 1	10	4	0.721	0.05 附近	4 时 23 分
模型 2	20	4	0.813	0.04 附近	6 时 37 分
模型 3	50	10	0.922	稳定在 0.04 以下	7 时 20 分
模型 4	100	10	0.942	0.03 附近	8 时 47 分

5. 结论

本文以改进之后的 YOLOV3 算法为核心算法，设计了一套能够用于对商场人员类型及流量进行检测和分析的系统。首先，针对传统的 YOLO 算法中存在的问题，本文在 K-means 聚类算法的基础之上，采用了升级之后的 K-means++均值聚类算法，更好地实现了分类的功能。其次，传统的网络结构在运行的时候会产生诸多问题，本文对网络结构进行了改进。最后，本文在模型训练的过程中，关注了较多的性能参数，可以用来观测模型性能的好坏，最后通过比较，得出了一个性能较为良好，优越的模型，并最终实现了对人员类型和流量进行检测分析的功能。本项目还存在较多需要改进的地方，如对人员流量的分析较为不直观，同时算法的功能还未得到完全开发，还有一定需要继续努力的地方。

致　谢

由衷感谢上海理工大学及上海理工大学光电信息与计算机工程学院对本人的栽培，是他们的支持让我完成了这篇论文。

参考文献

[1] 夷德. 基于 YOLO 的目标检测优化算法研究[D]: [硕士学位论文]. 南京: 南京邮电大学, 2021.

[2] 张政. 基于 YOLOv3 的密集行人检测算法研究[D]: [硕士学位论文]. 大庆: 东北石油大学, 2021.

[3] 董利兵. 基于区域全卷积网络的行人检测研究[D]: [硕士学位论文]. 阜新: 辽宁工程技术大学, 2021.

[4] 李祥兵, 陈炼. 基于改进 Faster-RCNN 的自然场景人脸检测[J]. 计算机工程, 2021, 47(1): 210-216.

[5] 黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述[J]. 计算机学报, 2014, 36(12): 1-18.

[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(2): 1-12.

[7] Turk, M.A. and Pentland, A.P. (1991) Recognition in Face Space. *SPIE Proceedings*, Vol. 1381, 43-54. https://doi.org/10.1117/12.25133

[8] Viola, P. and Jones, M.J. (2004) Robust Real-Time Face Detection. *International Journal of Computer Vision*, **57**, 137-154. https://doi.org/10.1023/b:visi.0000013087.49260.fb

[9] Vedaldi, A., Gulshan, V., Varma, M. and Zisserman, A. (2009) Multiple Kernels for Object Detection. 2009 *IEEE* 12*th International Conference on Computer Vision*, Kyoto, 29 September-2 October 2009, 606-613. https://doi.org/10.1109/iccv.2009.5459183

[10] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. https://doi.org/10.1023/b:visi.0000029664.99615.94

[11] Lienhart, R. and Maydt, J. (2002) An Extended Set of Haar-Like Features for Rapid Object Detection. *Proceedings. International Conference on Image Processing*, Rochester, 22-25 September 2002, 1. https://doi.org/10.1109/icip.2002.1038171

[12] Kuang, H., Chan, L.L.H. and Yan, H. (2015) Multi-Class Fruit Detection Based on Multiple Color Channels. 2015 *International Conference on Wavelet Analysis and Pattern Recognition* (*ICWAPR*), Guangzhou, 12-15 July 2015, 1-7.

https://doi.org/10.1109/icwapr.2015.7295917

[13] 李柯泉, 陈燕, 刘佳晨, 牟向伟. 基于深度学习的目标检测算法综述[J]. 计算机工程, 2022, 48(7): 1-12.

[14] Nauata, N., Hu, H., Zhou, G.T., *et al*. (2019) Structured Label Inference for Visual Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 1257-1271.

[15] 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(8): 10-25.

[16] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 2117-2125. https://doi.org/10.1109/cvpr.2017.106

[17] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. https://doi.org/10.1109/cvpr.2014.81

[18] Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T. and Smeulders, A.W.M. (2013) Selective Search for Object Recognition. *International Journal of Computer Vision*, **104**, 154-171. https://doi.org/10.1007/s11263-013-0620-5

[19] 吕璐, 程虎, 朱鸿泰, 等. 基于深度学习的目标检测研究与应用综述[J]. 电子与封装, 2022, 22(1): 010307.

[20] 包晓敏, 王思琪. 基于深度学习的目标检测算法综述[J]. 传感器与微系统, 2022, 41(4): 5-9.

[21] 张少伟. 基于深度学习的人流量统计[D]: [硕士学位论文]. 东营: 中国石油大学(华东), 2016.