圣彼得堡国立理工大学

学士学位论文

论文题目：

**《使用神经网络进行商店客流量统计》**

**ПРИМЕНЕНИЕ НЕЙТРОННЫХ СЕТЕЙ ДЛЯ ОЦЕНКИ ТРАФИКА МАГАЗИНА**

**作者 ：李逸嘉**

# 摘要

На - с., - рисунков, - таблицы, - приложение

*（*关键词：计算机视觉；神经网络；目标检测；多目标跟踪；行人重识别；

KEY WORDS: Computer vision, Neural networks, Object detection, Multiple object tracking, Pedestrian re-identification）

КЛЮЧЕВЫЕ СЛОВА: КОМПЬЮТЕРНОЕ ЗРЕНИЕ, НЕЙРОННЫЕ СЕТИ, ОБНАРУЖЕНИЕ ОБЪЕКТОВ, ОТСЛЕЖИВАНИЕ НЕСКОЛЬКИХ ОБЪЕКТОВ, ПЕРЕИДЕНТИФИКАЦИЯ ПЕШЕХОДОВ

Выпускная квалификационная работа на тему: «Применение нейронных сетей для оценки трафика магазина».

Данная статья посвящена реализации алгоритмов нейронных сетей для оценки трафика в магазине. Для решения поставленных задач были использованы методы глубокого обучения и компьютерного зрения.

Задачи, решаемые в ВКР:

1.Для обнаружения пешеходов был использован алгоритм YOLOv5

2.Для установления связи между персонажами при работе с несколькими камерами был реализован алгоритм повторного распознавания (ReID).

3.Для отслеживания локальных зон был реализован алгоритм DeepSORT он основе повторного распознавания.

4.На основе методики, описанных в данной статье, была разработана система подсчета посетителей для супермаркета.

В результате проведенных исследований была разработана система подсчета посетителей для супермаркета, которая демонстрирует точность \*\*\* и производительность.

Abstract

目录

[摘要 2](#_Toc134490196)

[绪论（Ведение） 6](#_Toc134490197)

[ГЛАВА 1 Обзор литературы （文件综述） 7](#_Toc134490198)

[目标检测 7](#_Toc134490199)

[基于传统方法的目标检测 8](#_Toc134490200)

[基于深度学习的目标检测 8](#_Toc134490201)

[多目标跟踪 9](#_Toc134490202)

[基于目标检测的多目标跟踪 9](#_Toc134490203)

[联合检测和跟踪的多目标跟踪 9](#_Toc134490204)

[行人重识别 10](#_Toc134490205)

[单域场景下的行人重识别 10](#_Toc134490206)

[跨域场景下的行人重识别 11](#_Toc134490207)

[本文的主要研究内容 12](#_Toc134490208)

[ГЛАВА 2 Обучение модели для отследования людей в нескольких видеопотоках （训练模型） 12](#_Toc134490209)

[ReID模型的训练 12](#_Toc134490210)

[数据集准备 12](#_Toc134490211)

[训练原理 14](#_Toc134490212)

[构建模型 14](#_Toc134490213)

[损失函数和优化算法 14](#_Toc134490214)

[训练步骤 15](#_Toc134490215)

[训练过程 17](#_Toc134490216)

[YOLOv5+DeepSORT模型的训练 18](#_Toc134490217)

[数据集准备和数据预处理 19](#_Toc134490218)

[构建模型 21](#_Toc134490219)

[构建损失函数和优化算法 21](#_Toc134490220)

[训练过程 22](#_Toc134490221)

[ГЛАВА 3 Система оценки трафика магазина основа на нейтронных сетей （商店客流量统计系统） 24](#_Toc134490222)

[项目背景 24](#_Toc134490223)

[商店客流量统计系统简介 24](#_Toc134490224)

[系统整体架构 24](#_Toc134490225)

[行人跟踪流程 24](#_Toc134490226)

[开发环境 24](#_Toc134490227)

[ГЛАВА 4 Анализ полученных результатов （结果分析） 25](#_Toc134490228)

[ReID模型的评价指标 25](#_Toc134490229)

[DeepSORT模型的评价指标 25](#_Toc134490230)

[商店客流量统计系统展示与测试 25](#_Toc134490231)

[系统展示 25](#_Toc134490232)

[测试环境 25](#_Toc134490233)

[跨摄像头行人跟踪功能 25](#_Toc134490234)

[准确性、效率、稳定性 25](#_Toc134490235)

[ЗАКЛЮЧЕНИЕ （结论） 25](#_Toc134490236)

[参考文献 25](#_Toc134490237)

# 绪论（Ведение）

（研究的主要背景）

随着零售业的发展和竞争加剧，目前国内很多的商场已经意识到了客流数据对于决策分析的重要性。客流量是指单位时间进出商场的顾客人数，通过对客流量的深层分析，可以对商场方方面面的管理提供科学依据。特别是对于提高商场日常经营决策的科学性、购物环境舒适度、人力资源调配的合理性等方面起着重要的作用。

(现有研究的不足)

最早出现的是人工统计方法，商场派若干员工在出入口，在一段时间内，连续地对进入商场的顾客进行目视计数。该方法的优点在于不需要进行信息采集设备的投资，调查资料比较全面、灵活，但缺点是不仅耗费人力、物力，并且容易出现漏记的情况。

随着机械触碰设备、红外传感设备和监控摄像头设备的出现，人流量统计方法依次经历了物理触碰统计和红外统计**两个阶段**。**物理触碰统计**多用在进出口，行人依次通过，但有设备安装要求较高、使用起来不方便、通行速度慢以及不能大范围使用等弊端； **红外统计**安装要求低、 使用相对方便， 但其统计误差大、不能用于大区域场景；除此之外，这些方法都有非常明显的缺点：无法知道每个顾客出入商店的时间。其次，客流量是动态的，某一时间段的客流量并不代表所有时间的客流量。最后，人工或机械统计后，资料整理的工作量也很大，统计的数据在使用之前必须经过编辑整理、数据提炼等过程。用这种方法得到的数据，只可以用于定性的了解，没有实际的统计分析价值。在经济竞争如此激烈的今天，这些统计方法已完全不满足市场的需求。

传统的人工检测方法效率低下、成本高昂，因此需要一种更加高效、准确且自动化的方法来满足快速变化的市场需求。基于此，我们开展了本研究，旨在探索利用神经网络方法对超市客流量进行监控与计算。

本文的主要目的是提出一种基于 YOLOv5 算法、DeepSORT 跟踪算法和行人重识别算法相结合的方法，以实现商店客流量的统计。我们通过对行人检测、跟踪和重识别算法的优化，实现了对不同角度和光照条件下的顾客进行准确的跟踪与计数，并开发了一个面向超市的客流量统计系统。

在可行性分析方面，我们认为本文提出的方法具有较高的可行性。首先，所使用的算法均为当前深度学习领域内较为成熟和广泛应用的技术，已经得到了充分实验验证。其次，在硬件和软件环境上，本文所需的配置相对较低，能够在普通 PC 上运行。最后，我们还对所提出方法的准确性、效率和稳定性进行了全面的实验评估，证明了其可行性。

为了实现本文目标，我们需要完成以下任务：

1）采集超市摄像头监控画面，并进行预处理；

2）设计并实现基于 YOLOv5 算法的行人检测模块；

3）设计并实现基于行人重识别算法的 DeepSORT 跟踪算法，应用于局部区域的跟踪；

4）设计并实现行人重识别算法以进行多摄像头下制定顾客的人物关联；

5）存储记录出入顾客的图片和出入时间，并进行客流量统计。

在此基础上，我们开发了一个实用的面向超市的客流量统计系统，可为超市提供有效的管理和决策支持。

# ГЛАВА 1 Обзор литературы （文件综述）

【todo:增加英语】

人流量统计是一个集目标检测、多目标跟踪、行人重识别技术于一体的综合性任务。鉴于此，本文首先叙述目标检测和目标跟踪的研究现状，其次考察跨摄像头多目标跟踪的研究现状。另外，考虑到行人重识别是解决跨摄像头轨迹关联问题的核心技术，在本文最后也总结了当前主流的行人重识别方法以及现存的一些不足。

## 目标检测

目标检测是多目标跟踪任务的基础，其主要目的是把需要被检测的目标从每一帧的静态图片中用矩形框标记出来，得到目标在图片中的具体位置。如果目标检测中出现了漏检，那么后期也就不会对目标进行跟踪，所以目标检测在多目标跟踪任务中起着非常关键的作用。纵观目标检测的整个发展过程，可将其分为基于传统方法的目标检测和基于深度学习的目标检测。

### 基于传统方法的目标检测

在深度学习发展之前，传统的目标检测算法通常遵循基于手工特征设计的思想，此类方法大致包括三个步骤，分别为区域选择、特征提取和分类器分类。该时期主流的检测算法包括 Viola-Jone 检测器[2]、HOG 检测器[3]和变形部件模型[4]（Deformable Parts Model, DPM）。Viola 等人提出的 Viola-Jone 检测器采用积分图像、特制选择和检测级联三种有效的策略，并通过从左到右、从上到下的滑动方式，判断是否有窗口包含目标，该检测器的提出使得目标检测的速度大大提高。HOG 检测器最初由 Dalal 等人在 2005 年提出，其核心思想是局部物体形状和外观能够被局部梯度或边缘的密度分布所描述，利用方向梯度直方图特征描述子来构建特征，并与 SVM 分类器结合，有效地解决了行人检测任务。另外，DPM 检测器由 Pedro 等人在 2008 年提出，遵循“分而治之”的理论对目标进行检测， 通过检测目标包含的各个部件从而完成对目标整体的检测。

### 基于深度学习的目标检测

随着传统目标检测算法的性能趋于饱和，目标检测的研究也进入了一个平稳的发展期，直到卷积神经网络的出现，使得目标检测技术的性能获得了显著的提高。此类算法大致可分为以下两类，一类是两阶段检测算法，先在图像中选择候选区域，然后再完成这些区域的分类任务，该类算法的特点是精度高但速度慢。如 Ross 等人[5]所提出的 R-CNN，首先以选择性搜索操作预先定位图片中的待检测区域，然后通过卷积神经网络提取这些区域上的特征，最后进行分类。虽然 R- CNN 算法大大提高了目标检测精度，但其存在计算冗余的问题。He 等人[6]提出的 SPPNet 算法，通过加入空间金字塔池化层解决了 R-CNN 反复缩放图像造成的影响。此外，Girshick 等人[7]提出 Fast R-CNN 算法，通过加入区域池化层进一步优化了 R-CNN 与 SPPNet 算法。Ren 等人[8]提出 Faster R-CNN 算法，将选择性搜索用区域提案网络代替，对网络进行端到端地训练，极大地提高了检测器的速度。另一类是一阶段检测算法，在产生候选区域的同时进行回归输出检测对象的位置和类别，其代表性算法就是 YOLO 系列算法[9]和 SSD 算法[10]。YOLO 系列算法首先由 Redmon 等人于 2016 年提出，其抛弃了“提案检测+验证”的两阶段检测范式，通过将图像分割成多个网格，直接预测每个网格所属的类别概率和位置信息。针对 YOLO 算法检测精度低的缺点，Liu 等人提出了 SSD 算法，引入了多尺度特征图检测技术，使得检测器的精度得到了极大提高。

## 多目标跟踪

多目标跟踪的主要目的是对视频中的多个目标同时进行轨迹跟踪，并维持对应的身份信息。随着近年来多目标跟踪研究的不断增多，其相关方法层出不穷，当前主流的方法可分为以下两类：基于目标检测的多目标跟踪和联合检测和跟踪的多目标跟踪。

### 基于目标检测的多目标跟踪

在基于目标检测的多目标跟踪算法中，目标检测和特征提取被分为两步，首先对当前帧的图像执行目标检测，然后提取检测框的特征表示，最后将检测框与已有的运动轨迹进行关联。在 2016 年，Alex 等人[11]提出了一种在线和实时跟踪方法（Simple Online and Realtime Tracking, SORT），利用卡尔曼滤波预测每条跟踪轨迹在当前帧中的运动状态，将预测值与当前帧中候选检测框的交并比作为相似度矩阵得分，最后通过匈牙利算法解决轨迹和候选检测框间的关联任务，这一创新性工作为多目标跟踪研究找到了新的方向。SORT 在数据关联时只使用目标的运动信息，因此在遇到频繁遮挡时容易出现身份信息交换的情况。Wojke 等人[12]提出的 DeepSORT 算法解决了该问题，在 DeepSORT 中使用结合运动和外观信息的联合度量来完成数据关联，大大减少了行人身份信息变化的次数，还引入级联匹配解决遮挡导致的概率弥散问题。另外，为解决检测器漏检、误检等问题， Chen 等人[13]设计了一个基于卷积神经网络的评分函数，利用该函数从检测器和跟踪器的输出中生成最优的候选框，进一步拓展了基于检测的多目标跟踪算法。

### 联合检测和跟踪的多目标跟踪

联合检测和跟踪的多目标跟踪算法采用了多任务联合学习的思想，该类算法通常训练一个可以同时输出目标检测框位置和目标外观特征的模型，提高了多目标跟踪的速度。Wang 等人[14]提出将行人重识别模型合并到 YOLOv3 检测器中， 并设计一个联合检测框回归、anchor 分类以及嵌入学习的损失函数完成同时学习外观特征表示和目标检测的任务。作为第一种联合检测和特征提取的方法，该算法存在很多不足，比如外观信息中缺少多层特征的融合、基于锚框的检测器不适合提取目标外观特征等问题。为了解决这些问题，Zhang 等人[15]提出了 FairMOT，使用深层聚合网络来实现不同尺度的检测，以便提取多层次的行人重识别信息， 同时使用了基于关键点的检测方式，缓解了目标中心偏移问题。此外，研究者们还提出将数据关联也统一到单个网络中，如 Zhou 等人[16]提出的 CenterTrack，通过将检测模型应用于相邻图像帧，结合前一帧的目标检测结果来预测当前帧的目标运动状态并完成数据关联，巧妙地将数据关联问题转为相邻帧之间的检测框回归问题。Peng 等人[17]提出了链式跟踪算法（Chained-tracker），业内首创统一目标检测、特征提取、数据关联三个任务的端到端多目标跟踪网络，并设计了联合注意力模块，相较于其他算法集成度更高。但链式跟踪算法的一个主要缺点是对目标个数的预估不够准确。由于该算法在跟踪期间动态地添加和删除节点，因此需要事先设置一个节点阈值来确定最大目标个数。如果该阈值设置过小，则可能会漏掉某些目标；而如果该阈值设置过大，则可能会将噪声或无关目标误判为跟踪对象，从而影响跟踪效果。DeepSORT采用了一些特殊的技巧来处理大量目标的跟踪问题，例如运动模型的加速度偏差、卡尔曼滤波器的参数设置等。这些技巧可以帮助DeepSORT处理复杂的场景和大量的目标，从而提高其跟踪的准确性和稳定性。

## 行人重识别

在跨摄像头多目标跟踪任务中，当目标离开当前摄像头的监控区域后，需要在其他摄像头视野中将目标重新激活，从而实现对目标的持续跟踪。当前的方法通常使用行人重识别来解决该问题。根据设定场景的不同，可以将行人重识别大致分为两个研究方向：一个是单域场景下的行人重识别，另一个是跨域场景下的行人重识别。在单域行人重识别方法中，训练集和测试集图像之间的风格类似， 数据分布差异不大。而跨域行人重识别则使用不同场景的数据进行训练和测试。

### 单域场景下的行人重识别

近年来，随着深度学习的发展，基于卷积神经网络的行人重识别成为了该领域的主流方法，这些方法克服了传统方法不能充分发掘行人图像中丰富判别信息的局限性，使得单域行人重识别技术获得了显著的性能改善。Zheng 等人[31]设计了一个身份嵌入网络，在该网络中行人被当作分类网络中的一个具体类别，同时使用行人的身份标签来训练网络。Xiao 等人[32]设计了一个深度分类网络模型， 该模型利用交叉熵损失学习数据集中行人图像的特征表达。此外，还有的研究者使用基于度量学习的方法来解决行人重识别问题。Hermans 等人[33]提出了一种困难三元组损失函数，其通过挖掘训练数据中的困难样本对实现模型的深度度量学习。Chen 等人[34]第一次提出四元组损失函数的概念，与使用三元组损失函数训练模型不同，该类损失使得不同行人特征之间产生的差异更大，而在相同行人特征之间产生的差异更小。Sun 等人[35]提出 VPM 网络，以自监督训练为主要方法， 同时使用交叉熵损失对网络进行区域分类训练，从而取得更高的准确率。

### 跨域场景下的行人重识别

尽管单域场景下的行人重识别方法在近几年获得了极大的性能提升，但是由于此类方法受限于单一场景下标注的数据集，当场景中背景、光照等因素发生变化时，其有效特征的提取能力会显著下降。因此，研究者提出跨域行人重识别方法，利用有标注的源域数据集和无标注的目标域数据集，提高源域训练的行人重识别模型在目标域上的性能。此类方法更加符合现实场景下行人重识别模型的部署，即在一个有标注的源域中训练好模型后，将该模型推广到新的目标域场景中。

现有的跨域行人重识别方法可以分为基于生成对抗网络（ Generative Adversarial Networks, GAN）的风格迁移法和基于特征相似性的伪标签估计法。风格迁移法使用 GAN 将源域图像转换成目标域风格，从像素级别上缓解源域和目标域间的数据分布差异。Deng 等人[36]提出了 SPGAN 模型，在图像风格迁移过程中，该模型通过构建自相似性和域相异性尽可能保留源域图像的身份特征。Zhong 等人[37]提出异构同质学习方法，考虑同质学习中的相机不变性，以解决不同摄像头配置引起的域内图像变化。但风格迁移法生成的数据与目标域真实的数据分布差距较大，导致其最终性能较弱。伪标签估计法使用聚类算法为目标域数据标注伪标签，从而以监督学习的方式优化网络。Zhai 等人[38]提出 AD-Cluster， 利用图像生成器在目标域增加聚类点，这些扩张数据可以在样本空间最大化聚类样本的多样性，大大提高模型的判别能力。Lin 等人[39]提出一种递进的聚类方法为未标记的目标域训练样本分配伪标签，然后使用标注的伪标签训练网络。Ge 等人[40]提出“同步平均教学”框架对伪标签进行精炼，核心思想是通过同行网络的输出来减轻伪标签噪声，同时利用模型的互补性来对彼此进行优化。Fu 等人[41] 提出自相似性分组，将行人图片的自相似空间分成三个，从全局到局部挖掘图像之间的潜在相似性。为了减轻伪标签噪声的影响，Zheng 等人[42]提出了一种不确定性估计策略，根据不确定性重新加权分类损失、三元组损失和对比损失。伪标签估计法通过探索目标域图像之间的特征相似性，为模型训练提供伪标签信息作为指导，在跨域行人重识别任务中显示出了优越性。除了上述两类方法外，一些新的方法也被提出。Dai 等人[43]提出了一种中间域模块（Intermediate Domain Module, IDM），通过使用适当的中间域连接源域和目标域，将源域知识更好地传递到目标域。Li 等人[44]提出了一种相关感知专家混合方法，动态地利用了源域的多样性特征，并使用基于投票的混合机制来提高模型的泛化能力。Zheng 等人[45]提出了一种基于标签迁移的组感知特征学习策略，通过将表示学习和聚类算法与标签迁移方法相结合，精炼在线训练方案中的伪标签。

## 本文的主要研究内容

本文的主要研究内容是设计一种YOLOv5 算法、DeepSORT 跟踪算法和行人重识别算法相结合的方法，以实现商店客流量的统计。我们通过对行人检测、跟踪和重识别算法的优化，实现了对不同角度和光照条件下的顾客进行准确的跟踪与计数，并开发了一个面向超市的客流量统计系统。

# ГЛАВА 2 Обучение модели для отследования людей в нескольких видеопотоках （训练模型）

## ReID模型的训练

### 数据集准备

#### MSMT17数据集

在CVPR2018会议上，提出了一个新的更接近真实场景的大型数据集**MSMT17**，即**Multi-Scene Multi-Time**，涵盖了多场景多时段。

MSMT17数据集采用了安防在校园内的15个摄像头网络，其中包含12个户外摄像头和3个室内摄像头。为了采集原始监控视频，在一个月里选择了具有不同天气条件的4天。每天采集3个小时的视频，涵盖了早上、中午、下午三个时间段。因此，总共的原始视频时长为180小时。

图片包含 游戏机

描述已自动生成

基于Faster RCNN作为行人检测器，三位人工标注员用了两个月时间查看检测到的包围框和标注行人标签。最终，得到4101个行人的126441个包围框。和其它数据集的对比以及统计信息如下图所示。

图形用户界面, 应用程序, 表格

描述已自动生成

MSMT17与其他数据集相比的优势如下：

（1）数目更多的行人、包围框、摄像头数；

（2）复杂的场景和背景；

（3）涵盖多时段，因此有复杂的光照变化；

（4）更好的行人检测器（faster RCNN），bounding box更精准

**评估协议**

按照训练-测试为1：3的比例对数据集进行随机划分，而不是像其他数据集一样均等划分。这样做的目的是鼓励高效率的训练策略，由于在真实应用中标注数据的昂贵。

最后，训练集包含1041个行人共32621个包围框，而测试集包括3060个行人共93820个包围框。对于测试集，11659个包围框被随机选出来作为query，而其它82161个包围框作为gallery.

测试指标为CMC曲线和mAP. 对于每个query, 可能存在多个正匹配。

### 训练原理

Reid模型的训练原理如下：

1. 数据预处理：reid首先将数据集中的图片进行0.5概率的裁剪、缩放、对称和数据增强等预处理操作，以便提高模型对不同姿态和光照条件下的人员图像的鲁棒性。
2. 特征提取：Reid使用训练好的ResNet系列卷积神经网络作为特征提取器，从输入图像中提取出512维特征向量。同时fastreid还引入一些新的特征提取方法，例如Scale-Aware Alignment (SAA) 和 random-erasing等，以进一步提高特征表达能力。
3. 度量学习：通过度量学习的方式将特征向量映射到一个低维空间中，并采用triplet loss和Cross Entropy Loss损失函数来优化模型，使得同一身份的样本在特征空间中距离更近，不同身份的样本在特征空间中距离更远。
4. 模型优化：使用Adam优化器对模型参数进行优化，以提供模型的训练效率和精度

### 构建模型

为了加速模型的收敛和提高模型性能，我们使用了在ImageNet数据集上预训练的卷积神经网络模型ResNet50。ImageNet是一个广泛使用的图像分类数据集，其中包含超过1000个类别的140万张图像。ResNet50模型使用了深度残差网络结构，可以有效地处理大规模图像分类任务。在训练完成后，改模型的权重可以被用于其他计算机视觉任务中以提高任务的性能和效率，比如我们的行人重识别。

预训练的ResNet50模型输出特征维度是2048。我们在主干网络后面继续接入了一个平均池化层和一个线性分类器。线性分类器的输出维度是N。N 表示训练数据的 ID 数目。

日程表

描述已自动生成

### 损失函数和优化算法

损失函数：

我们的reid模型会产生两个输出： features f 进行预测 logits p。features f 被用于计算 triplet loss，logits p 用于计算交叉损失熵。

**Cross Entropy Loss**

CrossEntropyLoss是一个常用的分类损失函数，通常用于训练深度神经网络进行图像分类任务。具体地，对于一个输入样本x和其真实标签y，CrossEntropyLoss的计算公式如下：

图形用户界面

描述已自动生成

其中C是类别数，yi表示第i个类别的概率（即真实标签），yi^表示模型预测x属于第i个类别的概率。

CrossEntropyLoss的目标是最小化预测与真实标签之间的差距，使得模型能够更准确地预测出每个样本所属的类别。

**Triplet Loss**

Triplet Loss是在人脸识别领域中广泛应用的损失函数之一，其目的是将同一个人的图像嵌入向量（embedding）拉近，不同人的图像嵌入向量拉远。具体来说，Triplet Loss会为每个样本学习一个嵌入向量，该向量能够使得同一个人的图像距离更近，不同人的图像距离更远。

对于一个triplet包括anchor、positive和negative三张图片，Triplet Loss的计算公式如下：

徽标, 公司名称

描述已自动生成

其中，d(a,p)表示anchor和positive之间的欧几里得距离，d(a,n)表示anchor和negative之间的欧几里得距离，m为margin，是一个预先设定的超参数，通常为正数。这个公式的含义是：如果当前的anchor和positive之间的距离减去anchor和negative之间的距离加上margin小于等于0，则说明当前的嵌入向量已经足够好，不需要再进行优化；否则，需要更新模型参数以获取更好的嵌入向量。

Triplet Loss通常与Batch Hard Triplet Mining结合使用，即在每个batch中选择最难的triplet进行训练，以提高模型的性能。

### 训练步骤

1.使用 ResNet50（初始化权重来自于 ImageNet 的预训练模型），然后改变其全链接层为N。N 表示训练数据的 ID 数目。

2.我们随机采样 P 个身份ID，并且对每个ID采集 K 张，最后一个batch size B = P \* K，在这篇论文中，我们设置P=16,K=4

3.我们改变每张图像的大小为 256 × 128，并且使用 0 值填充 10 个像素，然后使用随机剪切的方式，重新剪切 256 × 128 大小的图像。

4. 每张图像以 0.5 的概率值，随机进行水平反转。

5. 每幅图像都被解码为[0,1]中32位浮点原始像素值，然后我们通过减去0.485,0.456,0.406，再除以来对RGB通道进行归一化分别为0.229、0.224、0.225。

6.模型输出的ReID features f 进行预测 logits p。

7.ReID features f 被用于计算 triplet loss，logits p 用于计算交叉损失熵， triplet loss 的 margin m 设置为0.3

8.采用Adam方法对模型进行优化。初始学习率设为0.00035，在第40个epoch和第70个epoch分别降低0.1。总共有120个epoch

图示

描述已自动生成

### 训练过程

电脑屏幕截图

描述已自动生成

图形用户界面, 文本

描述已自动生成

文本

描述已自动生成

结果可视化：

图片包含 图示

描述已自动生成

## YOLOv5+DeepSORT模型的训练

**YOLOv5算法**

**【**韩晓冰**】**

YOLOv5 是由作者 Glenn Jocher 提出，一共有五种大小模型，分别为YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x。 其 中 ：YOLOv5n 的 卷 积 层 数 最少、检测速率最快、检测精度最低；另外的 4 个模型，卷积层数从小到大依次排列，随着模型复杂度增大检测速率逐渐降低，检测精度逐渐提高。YOLOv5 结构包括四部分：Input、Backbones、Neck、Prediction，如图1所示

图示

描述已自动生成

Input端包含 Mosaic 增强、动态锚框、自适应图片处 理等；Backbones 包括跨阶段局部网络 CSP、空间金字塔 池化 SPPF，前者有助于减少计算量，后者提高检测精确 度；Neck 采用了 FPN+PAN 的结构，下采样增强语义信 息，上采样增强定位信息；Prediction 是最后检测输出端。

为了追求更快的速度，我们选用预训练好的yolov5s模型(链接)。

**DeepSORT算法**

**【**韩晓冰**】**

DeepSort是对 Sort追踪算法的改进，跟踪过程中，利用卡尔曼滤波对目标下一帧位置进行预测，再进行级联匹配，最后利用匈牙利算法进行数据关联，提升跟踪效果，减少身份切换问题，整体流程如图 2 所示，输入视频到检测网络得到行人位置信息，再传输到跟踪网络中进行数据关联并对前后帧的人匹配，得到跟踪结果。

图示, 示意图

描述已自动生成

### 数据集准备和数据预处理

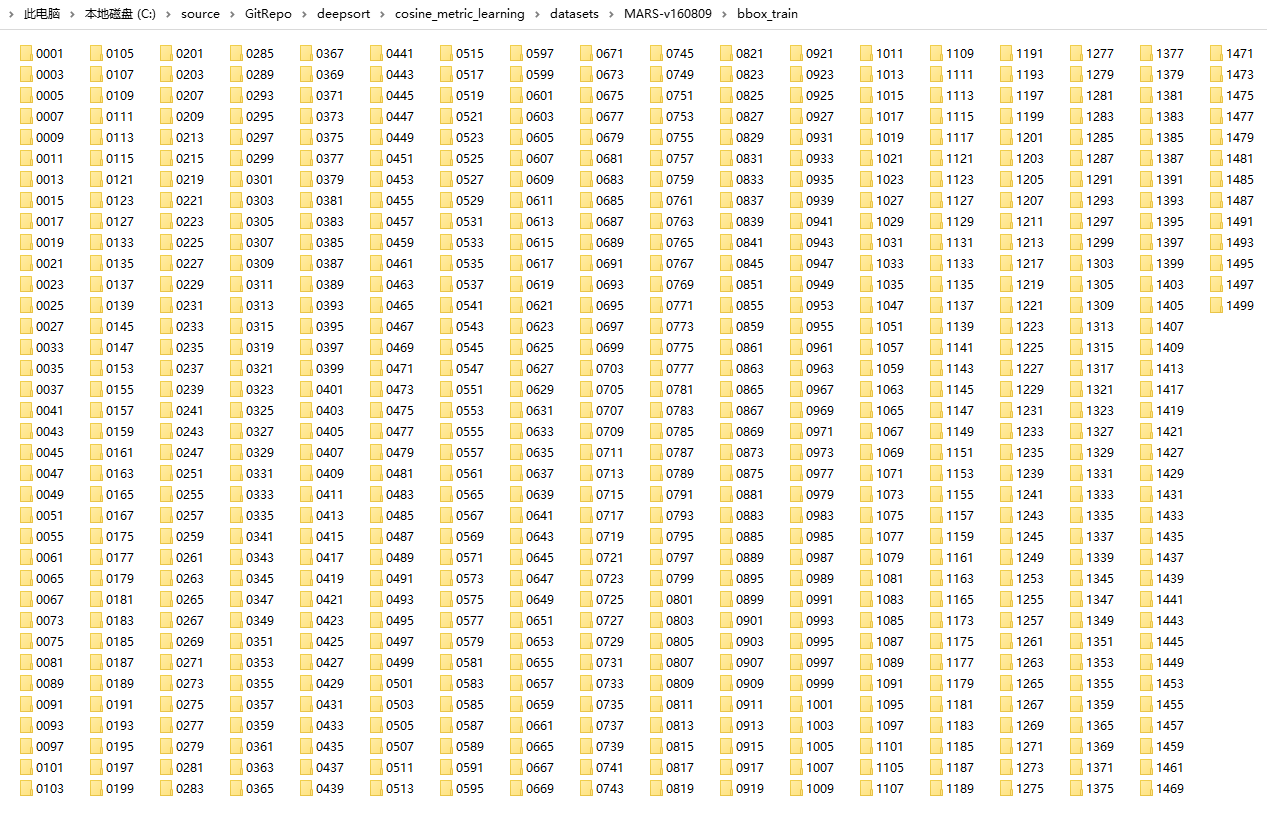
MARS 全称 Motion Analysis and Re-identification Set，是一个大规模的行人重识别视频数据集，也是 Market-1501 数据集的扩充。

围栏里的人群

描述已自动生成

数据是由六个近乎同步的摄像机收集的。数据集包含 1,261 个不同的行人，每人至少被 2 个摄像机拍摄。行人的姿势、着装颜色和光照的差异，以及较差的图像质量，使其很难产生高匹配精度。此外，该数据集包含 3,248 个干扰因素，以使其更加切合实际情况。DPM 模型和 GMMCP 跟踪器被用来自动生成短轨迹（大多为 25 – 50 帧长）。

bbox\_train文件夹中，有625个子文件夹（代表着625个行人id），共包含了8298个小段轨迹（tracklets），总共包含**509,914**张图片。



bbox\_test文件夹中共有636个子文件夹（代表着636个行人id），共包含了12180个小段轨迹（tracklets），总共包含**681,089**张图片。在实验中这个文件夹被划分为图库集（gallery）+ 查询集（query）。

背景图案

描述已自动生成

名称为**00-1**子文件夹表示无用的图片集，他们对应的行人id被设为\*\*-1\*\*，一般在算法中直接无视**pid = -1**的图片。

### 构建模型

1. 使用YOLO目标检测器来检测每帧图像中的人物
2. 使用DeepSORT算法来跟踪每个人物并将其与之前的轨迹相关联
3. 将fastreid用于提取特征，使得DeepSORT可以区分不同的人物
4. 确定模型超参数

DEEPSORT:

REID\_CKPT: "./fast-reid/checkpoint/model-final.pth"

MAX\_DIST: 0.2

MIN\_CONFIDENCE: 0.3

NMS\_MAX\_OVERLAP: 0.5

MAX\_IOU\_DISTANCE: 0.7

MAX\_AGE: 140

N\_INIT: 3

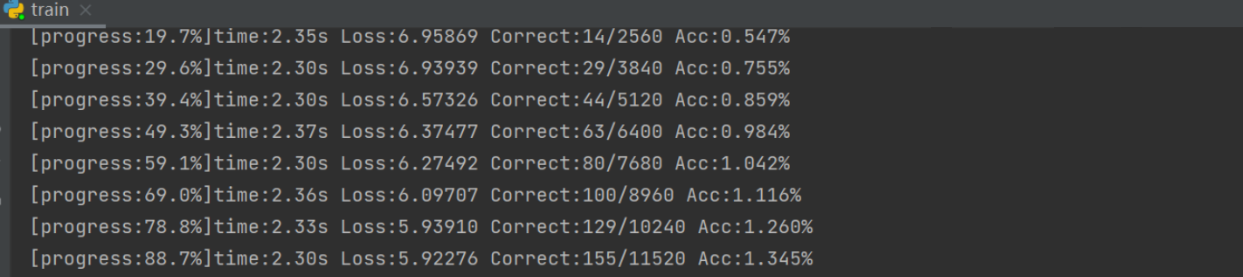
NN\_BUDGET: 100

### 构建损失函数和优化算法

DeepSORT的训练主要是针对特征提取模块进行训练。原文作者采用的损失函数是交叉熵损失函数（CrossEntropyLoss），我们前面在fastreid的训练中也使用了这个损失函数。

优化算法为带动量的随机梯度下降算法（SGD with momentum），权重衰减系数为5e-4。

### 训练过程



经过50代的训练，得到结果：

背景图案

描述已自动生成

图形用户界面, 应用程序

描述已自动生成

图表

描述已自动生成

最终结果：train\_acc=92.07%, test\_acc=91.34。

# ГЛАВА 3 Система оценки трафика магазина основа на нейтронных сетей （商店客流量统计系统）

## 项目背景

基于监控视频的人流量统计是资源合理分配、收集商业信息和进行智能管理等任务的基础，它也是现代视频智能监控重要的功能组成部分。本项目将采用YOLOv5、DeepSORT和Fastreid等先进的计算机视觉和深度学习技术，实现对商店内顾客的自动识别、追踪和分析。具体而言，该系统将通过摄像头获取商店内的视频流，利用YOLOv5对视频中出现的顾客进行自动检测和识别，然后使用DeepSORT对顾客进行目标跟踪，最终使用fastreid对顾客进行身份识别和行为分析，从而实现对商店客流量的自动统计和分析。

通过本项目的研究和实践，可以帮助商家更加精准地了解客户的行为习惯和偏好，制定更加科学有效的销售策略，提升商店的盈利能力和市场竞争力。同时，该系统还可以为商家提供实时的客流量监控和预测功能，帮助商家进一步优化运营管理，提高服务水平和用户体验。

## 商店客流量统计系统简介

### 系统整体架构

本文在第三章提出的跨域行人重识别方法和第四章提出的跨摄像头多目标跟踪方法的基础上，设计了一个面向社区的跨摄像头行人跟踪系统。该系统由四个模块组成：行人检测模块、特征提取模块、单摄像头跟踪模块以及跨摄像头轨迹关联模块。系统的整体架构如图 5.1 所示。

图示

描述已自动生成

（1）

图示

描述已自动生成

（2）

#### 行人检测模块

行人检测模块的输入为单个摄像头中的视频流，在实现过程中本文采用YOLO目标检测算法。利用YOLO目标检测算法对输入的图像序列依次进行行人检测，将其中的行人用检测框标记出来，并返回检测框的具体位置信息，定义如式（5.1）所示：

𝑏 = [𝑥, 𝑦, 𝑤, ℎ] （5.1）

其中，(𝑥, 𝑦)为检测框左上角像素坐标，(𝑤, ℎ)为检测框的宽度和高度。

#### 行人跟踪模块

行人跟踪模块采用了基于检测的跟踪策略，其输入为行人检测模块提供的检测框和特征提取模块提取的行人特征，目的是将当前帧中的行人检测框与前一帧中现存的运动轨迹进行关联。在得到行人检测模块提供的检测框后，首先利用特征提取模块提取当前帧检测框和现存运动轨迹的外观特征、姿态特征以及运动特征，然后计算它们之间的距离矩阵，最后使用卡尔曼滤波器和匈牙利算法将当前帧中的候选检测框与前一帧中现存的运动轨迹进行关联，完成行人在单摄像头下的跟踪任务。

#### 人流量统计模块

图示

描述已自动生成

在考虑了现实作业需求后，本研究使用“双线法”来完成人流量计数。双线法的具体原理如下：在图像中标记出两条计数线，分别记做 A、B。以身体运动轨迹为基础，通过轨迹与线相交的情况来表示从此经过的具体人数。

举个例子，我们假设商店的门位于B线的下方。那么顾客的轨迹会先接触A线再接触B，这被称为进门行为。反之，如果行人的轨迹先接触B线再解除A线，则被称为出门行为。每发生一次进门行为，行人检测、行人追踪、特征提取模块会依次工作，以获得顾客截图和特征，并在log中记录进入时间。当出门行为发生时，获取顾客的截图和特征，与数据库中的特征集进行比对，获取顾客id，并根据id记录顾客离开的时间。于是我们就完成了记录顾客进入和离开超市具体时间的工作。

人流量统计模块的输入为单摄像头下的人物轨迹。在进行特征比对的时候，需要用到先前存储的所有顾客的特征集合，可以通过访问数据库获取。

#### 特征提取模块

特征提取模块旨在提取行人的外观特征以及姿态信息等。在第x章中详细介绍了特征的提取方法。

特征提取模块的输入为行人检测模块中输出的行人检测框，经过ReID模型后，得到行人的外观特征。外观特征将会为后面的单摄像头跟踪模块

#### 存储模块

### 

### 跨摄像头行人跟踪流程

图示

描述已自动生成

在

### 开发环境

# ГЛАВА 4 Анализ полученных результатов （结果分析）

## ReID模型的评价指标

## DeepSORT模型的评价指标

## 商店客流量统计系统展示与测试

### 系统展示

### 测试环境

### 跨摄像头行人跟踪功能

### 准确性、效率、稳定性

# ЗАКЛЮЧЕНИЕ （结论）

# 参考文献