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**

# 摘要

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

КЛЮЧЕВЫЕ СЛОВА: 神经网络；行人检测；目标跟踪；行人重识别；客流量统计

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](#_Toc134346174)

[绪论（Ведение） 5](#_Toc134346175)

[ГЛАВА 1 Обзор литературы （文件综述） 5](#_Toc134346176)

[目标检测 6](#_Toc134346177)

[基于传统方法的目标检测 6](#_Toc134346178)

[基于深度学习的目标检测 6](#_Toc134346179)

[多目标跟踪 7](#_Toc134346180)

[基于目标检测的多目标跟踪 7](#_Toc134346181)

[联合检测和跟踪的多目标跟踪 7](#_Toc134346182)

[行人重识别 8](#_Toc134346183)

[单域场景下的行人重识别 8](#_Toc134346184)

[跨域场景下的行人重识别 9](#_Toc134346185)

[ГЛАВА 2 Обучение модели для отследования людей в нескольких видеопотоках （训练模型） 10](#_Toc134346186)

[ГЛАВА 3 Система оценки трафика магазина основа на нейтронных сетей （客流量统计系统） 10](#_Toc134346187)

[ГЛАВА 4 Анализ полученных результатов （结果分析） 10](#_Toc134346188)

[结论 ЗАКЛЮЧЕНИЕ 10](#_Toc134346189)

[参考文献 10](#_Toc134346190)

# 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，用以掌握商场客流信息。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，这些方法还有非常明显的缺点：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，以实现超市客流量的实时统计和计算。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）

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，其次回顾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

多目标跟踪也可以简称为多目标跟踪，其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，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，真正做到了检测和跟踪同时进行。

## 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。

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

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

# ГЛАВА 4 Анализ полученных результатов （结果分析）

# 结论 ЗАКЛЮЧЕНИЕ

# 参考文献