

分类号: TP330

UDC: D10621-407-(2022)3141-0

密级: 公开

编号: 2019075071

成都信息工程大学 学位论文

基于视觉技术的人数统计系统设计与实现

论文作者姓名:	汪洋
申请学位专业:	机器人工程
申请学位类别:	工学学士
指导教师姓名(职称):	王伟(讲师)
论文提交日期:	2023年5月17日

基于视觉技术的人数统计系统设计与实现

摘要： 基于视觉技术的人数统计是利用视频监控实时采集的图像数据进行人数检测和分析的一种技术。传统的人数统计方法主要通过设置人工计数器或使用红外线、超声波等传感器进行人数监测，存在覆盖范围较小、成本高、易受环境影响等问题。

基于视觉技术的人数统计是一种实时性高、准确度高的人流量监测方式。视觉技术的底层是基于深度学习，它可以通过摄像头进行捕捉。具体来说，该技术首先需要在视频图像处理方面实现滤波、去噪、背景建模、目标检测等多种技术，并对图像进行预处理以提高检测性能。接着需要使用相关的神经网络模型对视频流进行实时分析和处理。通过对检测到的物体进行分类和跟踪，最终可以达到人数统计的目的。总的来说，基于视觉技术的人数统计是一种应用前景广泛的监控场景下统计人流量的方式。除了在商业、安全等领域中发挥重要作用外，在其他领域，如人员管理、公共交通等方面还有很多潜力可以挖掘。

本研究的 UI 设计主要是利用 PyQt5 编写软件界面，将各种后端参数的配置方式集成在界面上，对系统进行优化和美化，这有利于使用者快速的发现视频中的人员变化情况。

关键词： 视觉技术；实时性；UI 设计

Design and Implementation of a Person Count System Based on Visual Technology

Abstract : Visual technology based population counting is a technology that utilizes real-time image data collected by video surveillance for population detection and analysis. The traditional method of counting the number of people mainly uses manual counters or sensors such as infrared and ultrasound to monitor the number of people, which has problems such as small coverage, high cost, and susceptibility to environmental impact.

The number of people statistics based on visual technology is a real-time and accurate method for monitoring pedestrian flow. The underlying layer of visual technology is based on deep learning, which can be captured through cameras. Specifically, this technology first needs to implement various technologies such as filtering, denoising, background modeling, and object detection in video image processing, and preprocess the image to improve detection performance. Next, it is necessary to use relevant neural network models for real-time analysis and processing of video streams. By classifying and tracking the detected objects, the ultimate goal of counting the number of people can be achieved. Overall, visual technology-based person count is a widely used method for measuring traffic in monitoring scenarios. In addition to playing an important role in fields such as commerce and safety, there is still a lot of potential to be tapped in other fields such as personnel management and public transportation

The UI design of this study mainly utilizes PyQt5 to write software interfaces, integrating various backend parameter configuration methods into the interface, optimizing and beautifying the system, which is beneficial for users to quickly detect personnel changes in videos.

Key words: Visual technology; real-time; UI design

目 录

论文总页数：36 页

1 绪论	6
1.1 研究背景及意义	6
1.2 国内外研究现状	6
1.3 研究内容	7
1.3.1 数据采集	7
1.3.2 数据处理	8
1.4 本章小结	8
2 系统的总体设计	9
2.1 总体设计	9
2.2 系统总体设计	9
2.3 本章小结	10
3 理论分析	11
3.1 计算机视觉基础理论	11
3.1.1 概念	11
3.1.2 图像分割	11
3.1.3 图像分类	12
3.1.4 图像检测	12
3.2 目标检测综述	13
3.2.1 Bottom-up 算法	13
3.2.2 CNN 模型	13
3.2.2.1 损失函数	14
3.2.2.2 Multiple Stage (MS)	15
3.2.3 关键点检测	15
3.2.3.1 Part confidence maps 算法	15
3.2.3.2 Part Affinity Fields 算法	15
3.2.3.3 实时分帧处理算法	16
3.2.3.4 头部识别	17
3.2.4 深度学习算法来进行无监督特征提取和动作模式分析	17
3.2.4.1 Hourglass 网络	17
3.2.4.2 无监督特征提取	17
3.2.4.3 动作模式分析	18
3.2.5 CNN 在 OpenPose 算法实现	18
3.2.6 Json 数据提取	21

3.3 人数统计技术综述	21
3.4 本章小结	22
4 算法实现	23
4.1 Openpose 环境配置	23
4.2 OpenPose 框架介绍	24
4.3 摄像头调用实时检测人数	25
4.4 通过本地视频进行人数统计	26
5 软件界面设计	27
5.1 前端工具介绍	27
5.2 登录页面设计	27
5.3 主页面选择及功能展示	28
6 实验结果测试与分析	30
6.1 实验结果测试	30
结论	33
参考文献	34
致谢	35
声明	36

1 绪论

1.1 研究背景及意义

人数统计技术是了解事件发生时所具有人次数据，能合理的调配活动时所需的人员安排，活动资源。在客流量较大的公共场所，人数统计的意义是管理者根据客流量变化，可以用于监测车站、机场、地铁等公共交通系统中乘客数量，以便调度更多或更少的车辆和列车，优化交通资源分配，缓解拥堵情况。利用数据分析为监控区域的管理，工作运营及规划决策提供了科学性的指导。因此，人数统计具有广泛的应用前景，如医疗控制，交通车辆管理。

最近几年，伴随着人工智能的快速发展，机器视觉在世界范围内的生产制造中的作用变得越来越重要，因此，对人的姿态、动作的识别和行为的理解也渐渐变成了计算机视觉领域的研究热点^[1]。机器视觉属于一种非接触式的自动识别技术，它是一种利用图像处理及相关技术来实现计算机代替人眼进行判断和测量的技术。利用机器视觉，将被摄取的目标转化为图像数据，并将其存储在数据库中，视觉处理系统对这些数据进行各种运算，从而提取出目标的特征，进而根据识别的结果，完成人员统计。在信息技术飞速发展的今天，以可视化技术为基础的人口统计系统已成为目前国内外研究的热点。

因此，考虑到日益增长的公共场所人流需求如客流量掌控和教室、会堂等出席人数统计的需要，本题目计划设计基于机器视觉的人数统计方法，实时识别摄像头区域内的人数个数，有利于公共安全管理部的应急管理需要。

1.2 国内外研究现状

人数统计可用于公共区域的管理和规划，应对突发事件^[2]，优化市场策略，实现安全监管，进行学术研究等。对这种情况下的人流量进行统计，对市场安排、政府统筹调度有着重要的意义^[3]。将这种技术运用到城市公交系统中，能够让公交公司对每条线路上的车辆资源进行合理的调度，从而缓解城市交通拥堵的情况，从而带来巨大的社会和经济效益。在这一领域，西方发达国家对此进行了较早的研究，并取得了一定的成效。比如，日本在很久之前就已经将乘客数量计算技术应用到了城市的交通管理系统中，而美国则将乘客数量计算技术应用到了 ITS 上^[4]。近年来，以视觉为基础的人群统计方法因其具有较高的计算精度和易于实现等优点而受到广泛关注。美国的“拜默”公司使用一种以机器视觉为基础的技术，对与之有关的人物体进行追踪和探测，并对其状况进行分析。在试验过程中，通过对数百位客户的数据进行统计，最终的结果误差只有 1%。南非的 headcounter 系统，是根据人的脚步声来计算交通量的，但因为体积太小，在处理海量数据方面，效率并不是很高，所以计算出来的数据并不准确。

陈先等研制出一套可在长途汽车上使用的内嵌式屏幕监视系统^[5]，它可以自动地采集汽车内部的影像，并对所采集的影像进行分析和处理，从而实现对旅客人数的统计。对于背景较为复杂的图像，传统的基于肤色的椭圆聚类算法无法实现人种肤色与背景的有效分离，并且无法实现对人种区域的准确定位，这极大地限制了后续的人种识别。该项目将重点研究基于颜色的椭圆聚类算法用于人群识别中存在的问题。在此基础上，该项目拟研究一种融合皮肤颜色与几何特征的人

头识别算法,并在此基础上,研究皮肤颜色与几何特征的人头识别算法,以提升人头识别的准确性。马丽娟等在 DSP 的基础上^[6],利用 TI 公司的 TMS320VC5402 作为核心处理器,利用高速视频译码电路,对模拟视频信号进行数字化采集,并对采集到的数据进行图象处理,从而达到对旅客流量的实时监测。张英、陈临强、杨礼坤采用线性 SVM 训练得到的头部分类器识别头部,并将头部中心点作为运动人体的特征点,利用最近邻匹配法进行数据关联,完成行人的跟踪计数^[7]。不同场景下的视频测试结果表明,该方法能较准确地实现人数统计^[8]。兰州理工大学完成了基于卷积神经网络的密集人数算法研究,采用基于目标检测的方法和基于密度估计的方法,从精度和速度两个方面入手进行改进,提出了一种改进的人头检测方法和一种基于深度可分离卷积的人群计数算法^[9]。虽然各式系统存在较多,但大多数这类系统仍然只适用于特定的区域,存在一定的局限性。因此,针对人数统计的系统种类存在很多,不同的场景需要使用不同的方法进行系统处理操作的现象,本次研究将基于 openpose 尝试做一个相对良好的人数统计系统。

1.3 研究内容

本文将深入研究视觉技术的相关原理,了解其在底层数据流的相关处理方式,掌握视觉技术后端处理格式、内容,练习基于视觉技术的实现方式与调用方法,针对当前业务场景。

再次,针对当前互联网流行的 PYQT 交互框架、python 编程语言进行应用,研究人数统计系统和机器视觉相关方面的知识,考虑到深度学习、机器学习的应用场景,本文考虑到实际应用的可能性,综合了视频分帧速度、人数统计检测效率等技术之后,最终考虑利用 openpose 算法的作为视觉技术的后端算法进行人数统计系统的研究。

基于视觉技术的人数统计系统研究工作主要分为视频输入流的实现,视频转为图像进行预处理,去噪等优化图像,分帧控制等算法进行实现,以及利用这些技术进行数据采集和数据处理,数据呈现等工作。

1.3.1 数据采集

基于视觉技术的人数统计系统中,视频数据采集是实现准确统计目标区域内人数的关键因素。以下是几种常见的基于视觉技术的人数统计系统中视频数据采集的方法:

固定摄像头采集:将固定的摄像头安装在目标区域内,实时采集图像数据,可以用于对室内场景下的人数进行统计。

可移动摄像头采集:通过搭载可移动的人脸识别摄像机或者无人机等可独立飞行的设备,对大型场馆和广场等空旷区域内进行实时监测,并确定目标位置及数量来估算人数。

红外线或激光雷达传感器采集:利用红外线或激光雷达传感器等设备对目标区域内的人数进行检测和计数,不依赖于照片清晰度、光照强度等条件,精度高且速度快。

智能手机和移动设备采集:利用智能手机拍摄照片和录制视频,然后使用人

人工智能技术运用分析算法对照片和视频进行处理，以实现人数的统计。移动设备可以随时携带，灵活方便。

1.3.2 数据处理

将采集到的视频进行图像预处理，例如视频去噪、图像增强、背景差分等操作，以提高后续人物检测和跟踪的准确性。利用深度学习模型，Faster R-CNN 算法，在每一帧中检测出所有的行人目标，并获得其图像区域和位置信息。将前一帧和当前帧中的行人目标进行匹配，保持各自的唯一标识符不变，从而获得相应的轨迹信息。去除重复的计数，消除明显的误差，比如计数器抖动或者遮挡问题，确保数字计数准确可靠。汇总所有已处理的数据，进行统计分析，包括每分钟，每小时，每天的人流量，以及通过数据可视化手段呈现在界面上。也可以结合其他数据进行更深入的分析，例如与天气、节假日等相关联的客流效果。并且在数据处理完成后，进行备份和存储，同时保证数据的安全性和隐私性。

1.4 本章小结

基于视觉技术的人数统计系统旨在通过计算机视觉技术，实现对行人的检测、跟踪等操作，并综合运用数据处理和分析方法，对人流量进行准确、实时的统计分析。本篇论文的研究目的，分析各类人口流动和聚集趋势，了解不同时间段、不同地点的人流特征和变化情况，从而更好地规划和管理公共场所。同时，还能够根据所得数据帮助商业招商策划、配套综合服务设施等决策。此外，在事件发生或者突发情况下，可以利用人数统计技术，推断出当前场所的人员密度和状态，给予明确的现场指挥，并辅助救援部门进行有效的决策，在减少伤害和损失方面起到关键的作用。同时也是为了解决传统的手工计数过程中存在的诸多问题，提高客流量统计的自动化、准确性和效率性。

在本论文的第一章中，我们首先阐述了本研究的背景和意义，即在教室人员人数统计及客运车实时车载人员统计等方面，基于视觉技术的人数统计系统有着广泛的应用前景。其次介绍了国内外在该领域的研究现状及最新进展，包括研究方法、主要技术和知名案例等，指出目前行业普遍存在的一些挑战和瓶颈。最后，本文针对当前的研究空白和发展趋势，阐明了我们本研究的主要内容。

以此作为切入点，接下来将详细阐述关键技术原理，系统架构设计和算法实现等各个层面的研究内容，结合实验数据和应用效果，为读者们提供更加全面和深入的了解和启示。

2 系统的总体设计

2.1 总体设计

数据收集：该模块通过摄像头录取视屏，并利用视觉技术实现对目标数据的统一收集和管理。将目标与背景分别处理，且还要对拍摄到的图像进行预处理，例如 REID 特征处理，这样才能获取正确的图像。

检测数据：将目标视频内容输入处理单元，将采集到的视频源进行预处理，比如视频去噪、图像增强、背景差分等操作，以提高后续人物检测和跟踪的准确性。通过深度学习算法 Faster R-CNN 处理视频内容，在每一帧中检测出所有的行人目标，并获得其图像区域和位置信息。将前一帧和当前帧中的行人目标进行匹配，保持各自的唯一标识符不变，从而获得相应的轨迹信息。去除重复的计数，消除明显的误差，比如计数器抖动或者遮挡问题，确保数字计数准确可靠。汇总所有已处理的数据，进行统计分析，包括每分钟，每小时，每天的人流量变化，以及通过数据可视化手段呈现在界面上。也可以结合其他数据进行更深入的分析，例如与天气、节假日等相关联的客流效果。最终将结果显示在交互界面上，同时将结果视频和目标视频都存放于数据库中。

统计人数：该模块的核心功能是实现业务流的处理，将检测数据部分读的结果通过接口的方式进行输出，调用结果集，实现统计人口的目的。

数据可视化：由于算法处理的视频流结果是以 json 文件的格式进行后台保存，所以对最终结果进行提取，最后通过数据可视化技术进行提取，实现对数据处理的结果的统一管理。通过界面实现人员数量的可视化管理。

2.2 系统总体设计

此小节是基于视觉技术的人数统计系统的系统流程图主要内容即为：打开系统，采集图像，预处理、特征提取、人体检测与跟踪、计数统计。系统流程图仅为示意图，仅为方便用户操作。如下图 2.1

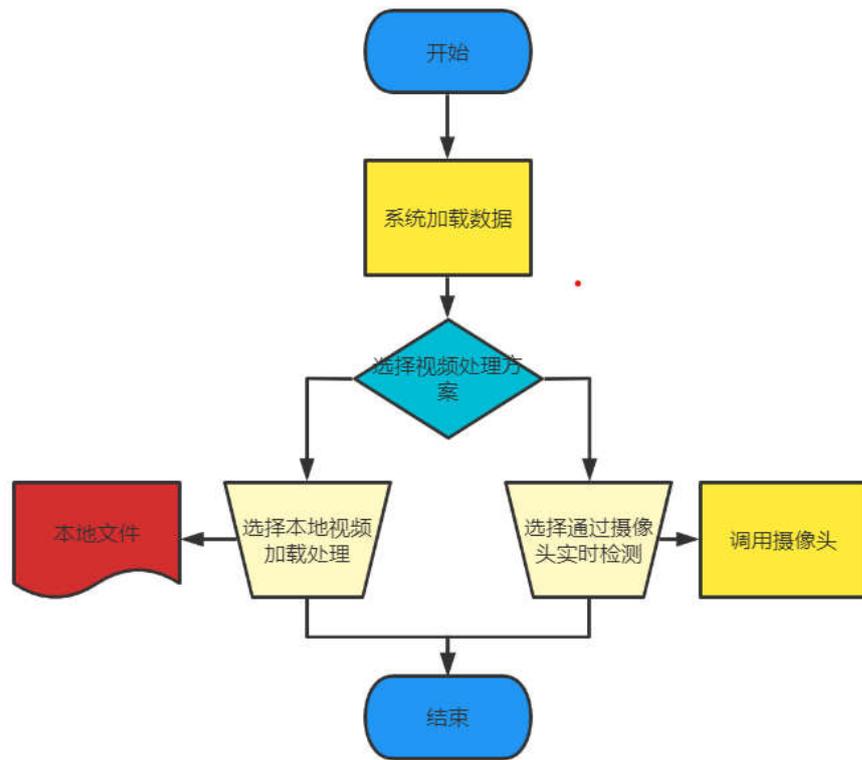


图 2.1 系统流程图

2.3 本章小结

本章主要讲述了此系统完成的应当具有的功能，以及此系统的整体框架，通过对本章节的理解，可以对此系统的整体设计和工作流程有一个大致的了解，并且因为本系统采用了逐渐深入方式，也能够对此系统的各部分功能有一个初步的了解，后续将对各部分功能进行详细讲述。

3 理论分析

3.1 计算机视觉基础理论

3.1.1 概念

计算机视觉研究的主要目标是，先尝试了解动态图像和静态图像的内容，然后利用深度学习算法收集实用图片内的信息，以便解决越来越多现实中出现的问题。简单来说，机器视觉技术解决的主要问题是：针对图像的噪声、失真、模糊等问题，需要采用图像处理技术对图像进行修复和增强，并通过图像分割算法来将图像中的不同物体或区域进行区分。通过各种算法，如深度学习算法和传统的基于特征的方法来实现不同物体的从图像中获取有效地信息，通过视频监控系统或图片获取数据，然后通过算法对这些数据进行识别、挖掘和分析，从而尽可能详细地描述该图像的信息。

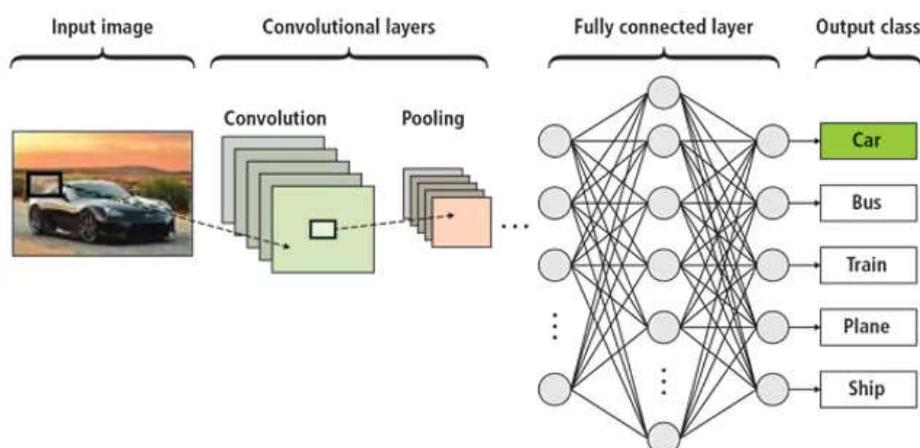


图 3.1 计算机视觉概念分析图

作为人工智能 (AI) 和深度学习的子领域，计算机视觉可训练卷积神经网络 (CNN)，以便针对各种应用场合开发仿人类视觉功能。使用计算机科学和视觉感知的原理可以自动化分析和解释数字图像和视频数据。计算机视觉领域的主要任务包括图像处理、物体识别、场景理解、运动分析、三维重建、人脸识别等。这项技术可以用于智能驾驶、人脸识别、智能制造、安全监控等众多领域。机器视觉包括对 CNN 进行特定训练，以便利用算法对图像和视频进行数据分割、分类和检测。如图 3.1 所示，卷积神经网络 (CNN) 能够针对多种应用场合执行分割、分类和检测。

3.1.2 图像分割

图像分割指的是将一幅图像分成不同的区域，每个区域包含具有相似特征的像素。这个过程旨在提取图像中的相似对象或区域，并为后续的图像分析和处理

提供更丰富的信息。

图像分割的应用场景非常广泛，如医学影像分析、自动驾驶、虚拟现实、计算机视觉等等。它可以对图像进行语义级别的解释和理解，为电脑深度学习等提供判断和识别基础。

常见的图像分割方法包括：目标追踪：在视频序列中跟踪目标的运动轨迹，比如将一段视频中的汽车跟踪并标记出来；图像分割：将图像分成若干个区域，每个区域包含相似的像素，比如将一张地图中的不同地区分割出来；视频分析：对视频进行分析，包括目标跟踪、动作识别、事件检测等；神经网络：一种模拟人脑结构的计算模型，广泛应用于计算机视觉中的深度学习；GPU 加速：使用 GPU 加速计算机视觉算法，提高计算效率和速度；数据增强：通过对数据进行变换和扩充，增加数据的样本量，提高算法的泛化能力。以及其及光流分析，语义分割等基础技术。

3.1.3 图像分类

计算机视觉中的图像分类是指将数字图像自动分成不同已知类别的过程。这个过程通常基于计算机学习，通过从大量带有标签的图像数据中学习特征，并根据新输入图像的相似性将其分为已知类别。

通常，图像分类可以使用传统的机器学习算法，如支持向量机（SVM）、随机森林和 K 最近邻等算法。在传统机器学习算法中，需要对图像数据进行预处理和特征提取，以降低维度和构建分类模型。

另一方面，深度学习技术也在图像分类任务中表现出色，尤其是使用卷积神经网络（CNN）。CNN 通常具有多层结构，在每一层中，它们会提取更高级别、更抽象的特征表示。通过训练模型，可以获得逐渐更复杂和有用的特征，以使模型能够更好地处理各种图像。

3.1.4 图像检测

图像检测技术通常可以分为两种类型：基于图像特征的方法和基于深度学习的方法。

基于图像特征的方法主要使用传统的机器学习方法，如支持向量机（SVM），利用手动设计的人工特征进行目标检测。尽管这些方法已经取得了很好的效果，但它们通常需要大量的时间和精力来设计和优化特征提取器，并且由于复杂场景下特征不稳定而存在一定的局限性。

基于深度学习的方法则采用卷积神经网络（CNN）对图像进行处理和特征提取，在学习足够的样本后从数据中自动学习有效的特征并进行端到端的训练。通过使用深度学习，特征分类器可以获得更好的泛化表现，并有效地解决了基于图像特征的方法的局限性。

图像检测技术已经在许多领域得到应用，如智能交通系统、安防监控、工业

制造等。通过图像检测，计算机可以定位对象的位置。在许多应用中，CNN 会在相关区域周围设置矩形边界框，将对象完全包含在内。检测器也可以接受训练，以便检测图像中汽车或人员的位置。

3.2 目标检测综述

本次采用的视觉技术是一种基于深度学习技术的的目标检测和姿态估计算法，它可以通过分析视频或图像中的人体姿态信息来实现人数统计。该技术是一种基于卷积神经网络的姿态估计算法，它可以同时检测多个人的身体部位，包括头部、手臂、腿部等，从而实现对人体姿态的准确估计。

在本文研究的统计系统中，首先需要对视频或图像进行预处理，包括图像的裁剪、缩放等操作，以便于后续的姿态估计。然后，通过视觉技术的算法对图像中的人体姿态进行检测和估计，得到每个人的身体部位坐标信息。接着，根据身体部位的坐标信息，可以计算出每个人的身高和体积等信息，从而实现人数统计。

在实际应用中，基于视觉技术的人数统计系统可以应用于人流量统计、安防监控等领域。它具有检测速度快、准确率高、可扩展性强等优点，可以有效地提高人数统计的效率和准确性

3.2.1 Bottom-up 算法

Bottom-up 算法指的是一种从小尺度开始逐步推导的方法，通常用于解决计算机视觉中的目标检测和分割等问题。

在目标检测方面，传统的 top-down 算法会先预设一个整体模板，再逐步细化地融合各部位的特征以识别目标物体。这种方法需要事先设计有效的目标模板，并耗费大量的计算时间和存储空间。

而 Bottom-up 算法则是直接从图像空间中的像素点出发，逐层递增的检测出可能的目标，并最终将它们合并成完整的对象。此类算法可以充分利用庞大的数据集和海量的计算能力，避免了人为模板的建立过程，并且具有较好的鲁棒性和泛化性能。

在分割问题上，Bottom-up 算法也经常被应用。该算法同样是在不涉及全局信息的情况下，通过不断迭代的方式从单个像素开始，递增地生成超像素或区域并执行聚类。通过这种方式可以快速准确地实现对图像中不同目标物体的分割。

3.2.2 CNN 模型

卷积神经网络（Convolutional Neural Network, CNN）是一种在深度学习中广泛应用的神经网络结构。它主要体现在对于图像和视频等数据的处理和分析。CNN 由多层神经网络和卷积层、池化层、全连接层等多个模块组成的。其中卷积层可以使用一组可训练的滤波器对输入进行局部连接和参数共享运算，以捕获

输入中的局部模式和特征；而池化层则可以通过选取最大或者平均值等方式，进一步压缩并减小前一层输出的尺寸，从而降低模型的计算复杂度，全连接层是从高层折叠特征中提取有用的信息，最终目的是产生输出，如下图 3.2 所示。

CNN 模型是一个具有多个卷积层、池化层和全连接层的深度神经网络。CNN 模型可以处理空间相关数据，例如可以对图像、语音和自然语言等数据进行分类、

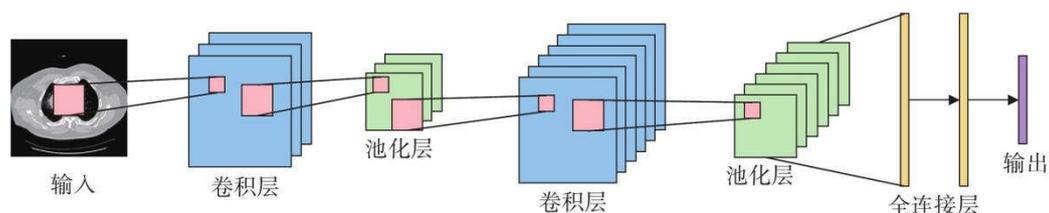


图 3.2 cnn 模型图

识别、分割、特征提取等操作。这种模型具有许多优点，如参数共享和局部相关性等，可以有效减少参数量和计算速度。

3.2.2.1 损失函数

损失函数（loss function）又叫做代价函数（cost function），是机器学习中的一种重要概念，用来度量模型预测结果与真实结果之间的差距。简单地说，损失函数表示了算法的预测值与实际值之间的误差。

在训练模型的过程中，人们通常需要确定一个适当的损失函数来评估模型的表现，并据此来优化模型的参数。不同的应用场景中适合使用不同的损失函数。例如，在分类问题中，常见的损失函数有交叉熵损失函数、对数损失函数等；而在回归问题中，我们可以使用均方误差损失函数、平均绝对误差损失函数等。选择正确的损失函数可以直接影响模型的性能和准确度。因此，在进行机器学习任务时，选择适当的损失函数是非常重要的。而神经网络训练或者优化的过程就是最小化损失函数的过程，损失函数越小，说明模型的预测值就越接近真是值，模型的健壮性也就越好。

卷积神经网络（Convolutional Neural Network, CNN）的损失函数通常采用交叉熵损失函数（Cross-Entropy Loss）。交叉熵损失函数是一种常用的分类损失函数，它可以衡量模型输出的概率分布与真实标签的差异。

对于一个样本，假设它的真实标签为 y ，模型的输出为 p ，则交叉熵损失函数的表达式如（3.2.1）：

$$L = \sum_{i=0}^1 (C)y^i / \log_p i \quad (3.2.1)$$

其中， C 表示分类的类别数， y^i 表示样本属于第 i 类的真实标签， p^i 表示模型输出样本属于第 i 类的概率。

在卷积神经网络中，通常采用 softmax 函数将模型的输出转化为概率分布，

然后再计算交叉熵损失函数。同时，为了防止过拟合，通常还会在损失函数中加入正则化项，例如 L 正则化。

3.2.2.2 Multiple Stage (MS)

Multiple Stage (MS) 技术是机器视觉中常使用的一种技术。它通过引入多阶段处理，以提高算法在不同数据集和场景下的鲁棒性和精度。

在视频处理的第一个阶段中，该算法采用了 **Faster R-CNN** 等传统目标检测算法来为输入图像中可能包含的人体定位生成候选区域，然后在每个这样的区域内进行精确定位和分类来判断其中是否包含人体。在第二个阶段中，由于机器视觉技术采用卷积神经网络 (CNN) 模型来预测人体关键点的位置，并得到各个关键点的置信度热力图。在第三个阶段中，视觉处理模块采用 **PAF** 技术来建模关节之间的连线关系，并使用 CNN 模型来联合预测所有的关键点和连接。所有这些阶段都被视为单个任务，以提高效率和准确性。

但是，针对大量而复杂的场景，例如如此众多和相互遮挡得非常严重的人员之间，需要更加精细的处理方法。因此，在 **MS** 技术的帮助下，机器视觉处理单元可以对每一个步骤进行了多次迭代和优化，将结果传递给后续阶段，从而逐步提高算法的鲁棒性和准确性。这个多阶段的过程可以使用树形结构来描述，其中有一个主干网络以及若干支路 (Branch) 网络用于处理每个关键点或关节之间的连接，并且将结果传递到下一层。

3.2.3 关键点检测

3.2.3.1 Part confidence maps 算法

Part confidence maps (关节点置信图)，表示像素在关节点的高斯响应^[10]，离关节点越近的像素，响应值越大。在目标实例中，每个关键点都有对应的热图，表示这个关键点可能存在的概率，热图中像素值越高则表明此处出现该关键点的概率越大，反之亦然。

在创建关键点置信图时，机器视觉处理图片依赖于决策树森林算法进行多层次多分支的处理，使得关键点位置更加准确。同时，还采用了递归神经网络 (RNN) 来捕捉时间序列信息，以便更好地对视频序列进行处理。

因此，关键点置信图是机器视觉技术中至关重要的组成部分，它为精确检测人体姿态提供了基础，并保证了在复杂场景下的准确性和鲁棒性。

3.2.3.2 Part Affinity Fields 算法

在计算机视觉的应用中，**PAF** (Part Affinity Fields) 算法是一种基于 CNN 的方法，用于检测人体关键点之间的关联关系。PAF 算法的主要思想是将人体关键点之间的连接关系建模为一个向量场，每个向量表示两个关键点之间的连接关系。在训练阶段，PAF 算法通过学习大量的人体姿态数据，学习如何从输入图像中提取这些向量场。在测试阶段，PAF 算法将这些向量场应用于输入图像，

以检测人体关键点之间的连接关系。



图 3.3 关键点与热力示意图

图 3.3 中上部分是在不同阶段是右手腕的关键点预测效果，下半部分是在不同阶段的右大臂的预测效果。可以看出随着阶段增加，预测的效果越来越准确。红色圈圈是从假阳性变为了真阴性，橙色的圈圈是从假阴性变为了真阳性。

PAF 算法的优点是能够检测出人体关键点之间的连接关系，从而提高了人体姿态检测的准确性。缺点是需要大量的训练数据和计算资源，训练和测试时间较长。

3.2.3.3 实时分帧处理算法

实时分帧处理算法是一种针对连续视频数据流进行处理的技术，其主要目标是在保持系统实时性和高效性的同时，对视频进行分割，便于后续的计算机视觉任务算法如下（3.2.2）：

$$\begin{aligned} \text{Frame_num} &= (\text{N-win} + \text{inc}) / \text{inc} \\ &= (\text{N-overloop}) / \text{inc} \\ &= (\text{n-win}) / \text{inc} + 1 \end{aligned} \quad (3.2.2)$$

其中各类数据的定义如下 frame_num 为视频的帧数 N 为视频的采样点数 采样点数 = 采样率 * 视频时长 win 为窗口长 inc 为帧移 $\text{overloop} = \text{win} - \text{inc}$

主要包括两个部分：预处理和后处理。在预处理部分，首先对原始图像进行物理操作，以适应 OpenPose 的模型输入大小，并且可以降低计算复杂度^[1]。接着，通过使用高斯滤波器和直方图均衡化等技术来减少噪声和增强图像对比度，从而帮助检测更准确。

在后处理阶段，输出结果集合（如关键点分布、连线关系和姿态数据）需要经过一些过滤和优化方法，以更好地提高准确性。例如，可以利用多目标追踪算法和贪心匹配算法来消除重复和错误检测，或者使用基于空间约束的姿态估计方法来纠正可能的姿态异常。

同时，为了保持实时性，OpenPose 还采用了一些卓越的 GPU/CPU 并行加速策略和具体的深度学习推理优化措施。这些特性使得 OpenPose 能够更快地处理并检测出更多人体关键点的信息，同时具有很好的扩展性和可移植性，以适应不同硬件和场景需求。

3.2.3.4 头部识别

机器视觉的发展应用是一种基于卷积神经网络（CNN）的人体姿态估计方法，从图像中直接预测出人体的关键点坐标，并通过关键点之间的连接关系来估计人体姿态。

具体而言，本次系统设计的视觉处理算法采用了用于人体 2D 姿态估计的 Hourglass 网络，该网络结构可以对特征提取和姿态回归两个任务同时进行优化，并充分考虑到不同分辨率级别上的语义信息。这使得处理后的结果具有很好的姿态度量性能，在处理遮挡或多人情况下也可保持较高的准确性。

总之，本次系统采用的视觉处理单元并没有利用传统的边缘检测算法来定位头部等人体部位，而是利用深度学习算法来进行无监督特征提取和动作模式分析，因此在姿态估计任务上表现良好。

3.2.4 深度学习算法来进行无监督特征提取和动作模式分析

本次使用的机器视觉算法采用的深度学习算法主要是 Hourglass 网络，该网络结构可以对特征提取和姿态回归两个任务同时进行优化。下面将介绍 Hourglass 网络在机器视觉中的应用以及无监督特征提取和动作模式分析的过程。

3.2.4.1 Hourglass 网络

Hourglass 网络结构由 Newell 等人于 2016 年提出，它具有自顶向下和自底向上的两个通道，并且包括一个降采样卷积模块和一个上采样卷积模块（称为残差模块）。整个网络堆叠多个 Hourglass 模块，并最终输出人体关键点的坐标作为姿态估计结果。

本次系统使用了方法论中尺寸为 4 的 Hourglass（HG-4）网络，该网络结构包含了 16 层卷积层和 5 个池化操作，总共有 4 个 Hourglass 模块。每个模块包含编码、解码和跨步连接等三部分，其中编码器通过降采样卷积层提取是否存在人体的二进制 mask 和相应的特征图，解码器通过反卷积层重建隐藏信息并生成目标物体的 3D 位置估计。此外，这里还引入了中间级别 block 的 shortcut 来帮助梯度传递以及损失函数的施加（常用 L 均方误差函数）。

3.2.4.2 无监督特征提取

本次采用机器视觉算法中一般使用未标注的数据集对网络进行训练，通过大量数据的特征提取来预训练模型并学习不同姿态下的关键点分布规律。Hourglass 网络中利用卷积神经网络对输入图像进行多次处理和池化操作，产生

多个尺度上不同抽象程度的特征表示。这些特征映射会被传递到解码器中进行上采样运算，并与相应编码层的信息进行残差连接，从而得到人体关键点位置的估计结果。

在机器视觉中，为了提高模型性能并减少训练时间，可以利用了 MPII Human Pose 数据集以及 COCO Keypoint Challenge 数据集上的公开数据作为额外的训练数据，从而实现了迁移学习的效果。更进一步，在机器视觉算法中还支持动态调整 batch size 大小，从而优化模型性能和训练速度。

3.2.4.3 动作模式分析

当机器视觉的关键点识别确定了关键点的位置和连线关系之后，可基于此进行动作模式分析。例如，可以使用基于 SVM 或深度学习分类器的方法来识别某个动作类型，也可以使用基于物理建模的方法来进一步推断身体的运动轨迹和肌肉受力等细节信息。

总之，通过无监督特征提取和动作模式分析，能够捕捉到人体关键点的空间位置、运动状态和运动轨迹等多个方面的信息，并为下游应用提供更加鲁棒和全面的数据支持。

3.2.5 CNN 在 OpenPose 算法实现

在 opepose 实现整个任务时，我们可以先使用姿态估计模型对输入图像中的每个姿态进行估计，然后对每个姿态进行关节检测。最终，我们将所有的关键点组合起来，最终得到最后的检测结果。

CNN 模型在 OpenPose 中被用来检测人体的关键点。具体来说，CNN 模型被用来实现以下两个模块：

1. Part Confidence Maps 模块：该模块实现了 cnn 模型的应用，实现了对人体姿态估计的深度学习模型中常用的一种组件。其主要功能为生成每个人体关键点位置的置信度图，即针对每个关键点生成一个热力图，用于表示关键点的位置和置信度。PCM 图像包含了 18 个通道（channel），每个通道代表一种人体关键点。

2. Part Affinity Fields 模块：该模块也使用 CNN 模型对输入的图像进行处理，生成一个 Part Affinity Fields (PAF) 图像，表示两个 Parts 之间的关联性 (affinity)。PAF 图像包含了 19 个通道，其中前 18 个通道代表一个 Parts 与它相邻的 16 个部位的亲和度，如下图 3.4 所示，第 19 个通道代表两个相邻 Parts 之间的向量方向和模长信息。

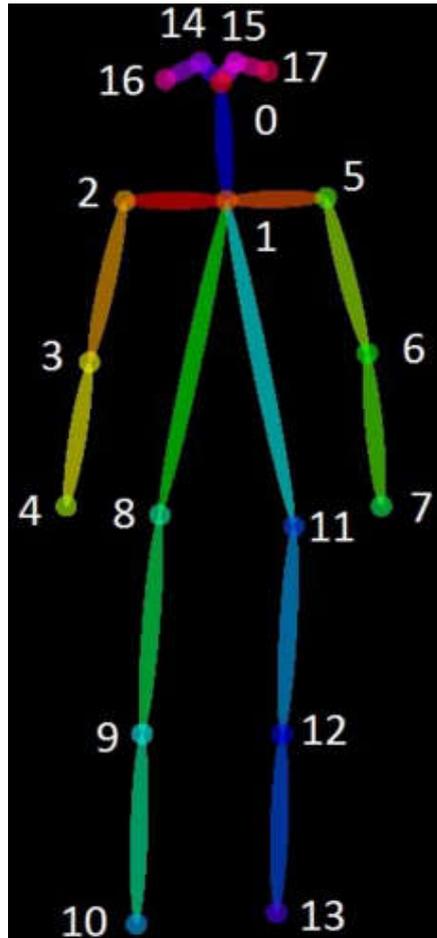


图 3.4 人体关键点图示

在 Openpose 中使用预先训练的 CNN 网络，可以对每个关节进行检测和识别。这个过程涉及到大量的数据学习和预处理，因此 Openpose 的速度非常快，并且可以实时检测和识别多人姿态信息。

人体姿态估计 (Human Posture Estimation)，是视觉系统算法将图片中已检测到的人体 (如下图的图 3.5) 关键点，通过机器视觉算法将关键点正确的联系起来， (如图 3.6 所示) 从而显示估计的人体姿态。

人体关键点是指用于描述人体姿态的一些特定关键部位点，如头部、颈部、肩部、手肘、腕部、髋部、膝盖和脚踝等。它们通常被用作人体姿态估计、动作捕捉等任务的数据特征。如图 3.7 所示。

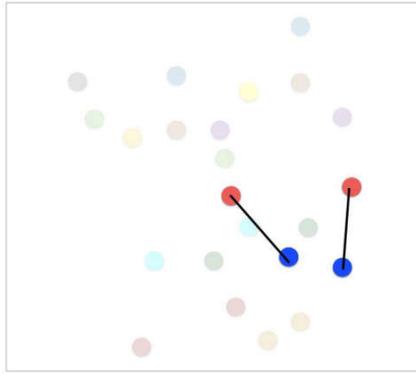


图 3.5 关键点连接图

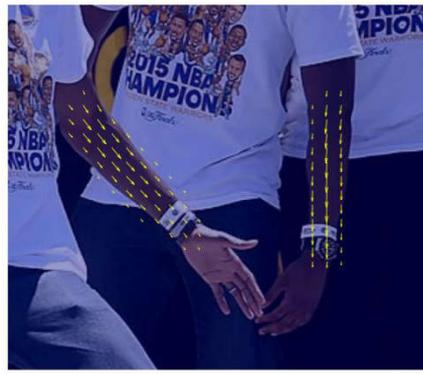


图 3.6 热力图

通过对人体关键点在三维空间相对位置的计算，来估计人体当前的姿态。进一步，增加时间序列，看一段时间范围内人体关键点的位置变化，可以更加准确的检测姿态，估计目标未来时刻姿态，以及做更抽象的人体行为分析，例如判断一个人是否在打电话等。

人体姿态检测是一个重要而具有挑战性的计算机视觉任务，面临以下几个主要挑战：人类身体本身就具有多样化和复杂度，不同体型、肤色、年龄、性别等都会影响到人体姿态估计的精度，因此需要对不同人群和各种人体姿态进行充分训练和测试。

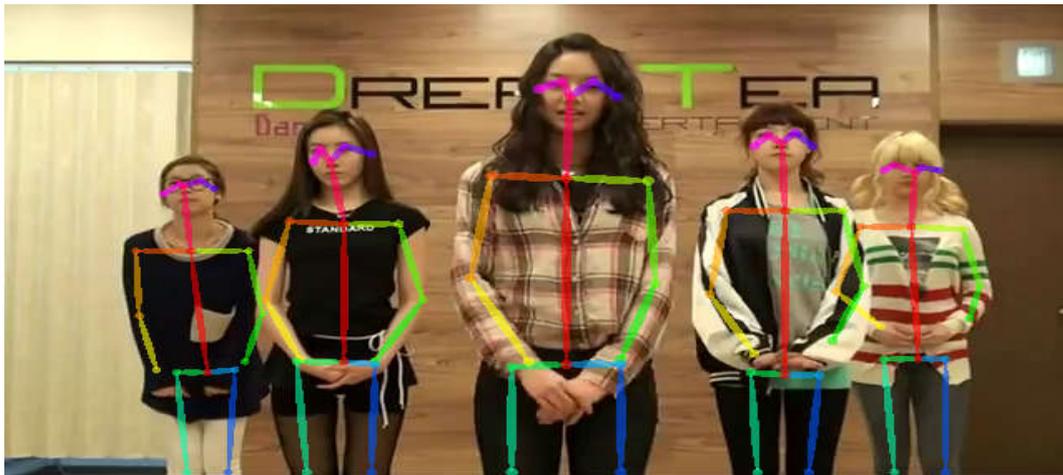


图 3.7 多人人体姿态识别图

人体姿态有很多种表达形式，如平静站立、行走、跑步、跳跃等，这些不同动作之间的变化也会增加算法的难度和复杂度。在实际应用中，人体可能被遮挡、部分缺失或出现模糊等情况，这些都会影响关键点的检测和跟踪精度。环境的光照和背景也会对人体姿态的检测造成非常大的影响，例如弱光照、阴影、杂乱的背景等等。根据不同领域的需求，人体姿态检测往往需要实时性，即时反馈人体动作和状态的变化，因此算法的速度和效率也是一个重要挑战点。

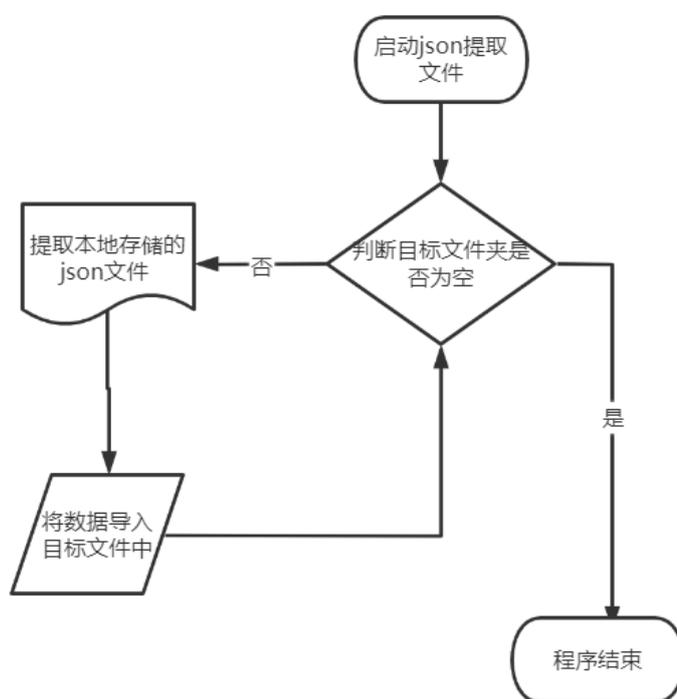
总之，虽然人体姿态检测在学术和商业领域都有广泛应用，但仍需要进一步研究和优化算法，以提高精度、鲁棒性和实时性，并适应不同的应用场景。

3.2.6 Json 数据提取

由于前台的数据需要实时进行更新，因此数据在处理 JSON 数据文件时，采用如下步骤进行数据提取：

加载 JSON 文件：首先使用相关编程语言中的 API 读取 JSON 文件内容，并将其转换成字符串类型。解析 JSON 数据：使用相应的 JSON 解析器将 JSON 字符串解析成对应的数据结构。提取数据：通过字典或列表索引的方式获取所需数据。

流程图如下图 3.8



3.8 json 文件循环提取流程图

3.3 人数统计技术综述

人数统计技术是指利用图像、视频和传感器等技术手段，对目标区域内的人数进行自动识别、计数和估计的一类技术。以下是几种常见的人数统计技术：

目前是市面上常见的技术主要是。基于视频图像分析的技术，基于热红外成像的技术，基于深度学习的技术，基于计算机视觉的技术：这类技术结合了基于图像分析和深度学习的方法，通过对摄像头捕捉到的图像进行处理和分析，实现对复杂场景下的人数估计和统计，如体育馆内的观众量监测。

总之，不同的人数统计技术各有特点，可根据具体应用场景选择适宜的方法，并针对不同问题进行优化和改进。通过视频监控采集的场景中的人体关键点坐标与先前采集的参考图像中人体关键点坐标进行匹配比对，从而实现对场景中的人数进行实时统计。与传统的人数统计技术相比，基于视觉技术的人数统计系统具有诸多优势。首先，该系统能够从视频监控中实时提取关键点信息，实现人数统计的同时还能掌握人员行为特征、分析人员流动方式等信息，对人员管理和场所设计提供参考和指导。其次，该系统可以实现数据的快速采集和准确分析，真正实现了“实时监控、辅助决策”的目标。

因此，基于视觉技术的人数统计系统已经成为人数统计技术的新趋势，具有广阔的应用前景。

3.4 本章小结

本章主要介绍了目标计算机视觉发展现状及理论基础，涉及到了机器视觉中热力图等目标检测算法。在基于视觉的人数统计算法中 CNN 模型、Part confidence maps 和 Part Affinity Fields 是其中比较重要的核心算法，在机器视觉的数据的处理、特征提取和目标检测都有所创新。而基于视觉处理融合的算法中，实时分帧则更加重视视频帧率控制利用和如何提高视频实时性，小勇快慢的问题。同时，文章也指出了目标机器视觉在应用中仍存在一些问题和挑战，如人数规模及人的位置不确定，人与人之间存在干扰，使关节关联变得困难；人数越多复杂度越高，实时性能越差。未来的研究和应用方向主要是提高算法的鲁棒性、准确性和普适性，为人流统计、车辆监控和物流运输等领域的应用提供支持。

4 算法实现

4.1 Openpose 环境配置

1、数据集下载准备

下载 openpose 源码:

```
git clone https://github.com/CMU-Perceptual-Computing-Lab/openpose.git
```

运行 getModels.bat

运行 ①getCaffe.bat | ②getCaffe3rdparty.bat | ③getFreeglut.bat | ④getOpenCV.bat | ⑤getSpinnaker.bat 文件完成数据集环境配置。

使用 cmake 生成工程

如图 4.1, 图 4.2

Name	Value
BOOST_FILESYSTEM_LIB_DEBUG	D:/bishe/mmpose/openpose/3rdparty/windows/caffe3rdparty/lib/boost_filesystem-vc142-mt-gd-x64-1_74.lib
BOOST_FILESYSTEM_LIB_RELEASE	D:/bishe/mmpose/openpose/3rdparty/windows/caffe3rdparty/lib/boost_filesystem-vc142-mt-x64-1_74.lib
BOOST_SYSTEM_LIB_DEBUG	D:/bishe/mmpose/openpose/3rdparty/windows/caffe3rdparty/lib/boost_system-vc142-mt-gd-x64-1_74.lib
BOOST_SYSTEM_LIB_RELEASE	D:/bishe/mmpose/openpose/3rdparty/windows/caffe3rdparty/lib/boost_system-vc142-mt-x64-1_74.lib
BUILD_BIN_FOLDER	<input checked="" type="checkbox"/>
BUILD_CAFFE	<input checked="" type="checkbox"/>
BUILD_DOCS	<input checked="" type="checkbox"/>
BUILD_EXAMPLES	<input checked="" type="checkbox"/>
BUILD_PYTHON	<input checked="" type="checkbox"/>
BUILD_SHARED_LIBS	<input checked="" type="checkbox"/>
BUILD_UNITY_SUPPORT	<input type="checkbox"/>
CMAKE_BUILD_TYPE	Release
CMAKE_CONFIGURATION_TYPES	Release;Debug
CMAKE_INSTALL_PREFIX	C:/Program Files (x86)/OpenPose
CUDA_ARCH	Auto
CUDA_HOST_COMPILER	\$(VCInstallDir)/MSVC/\$VCToolsVersion/bin/Host\$(Platform)/\$(PlatformTarget)
CUDA_SDK_ROOT_DIR	CUDA_SDK_ROOT_DIR-NOTFOUND
CUDA_TOOLKIT_ROOT_DIR	C:/Program Files/NVIDIA GPU Computing Toolkit/CUDA/v12.0
CUDA_USE_STATIC_CUDA_RUNTIME	<input checked="" type="checkbox"/>
Caffe_INCLUDE_DIRS	3rdparty/windows/caffe/include;3rdparty/windows/caffe/include2
Caffe_LIB_DEBUG	D:/bishe/mmpose/openpose/3rdparty/windows/caffe/lib/caffe-d.lib
Caffe_LIB_RELEASE	D:/bishe/mmpose/openpose/3rdparty/windows/caffe/lib/caffe.lib
Caffe_Proto_LIB_DEBUG	D:/bishe/mmpose/openpose/3rdparty/windows/caffe/lib/caffeproto-d.lib
Caffe_Proto_LIB_RELEASE	D:/bishe/mmpose/openpose/3rdparty/windows/caffe/lib/caffeproto.lib
DL_FRAMEWORK	CAFFE
DOWNLOAD_BODY_25_MODEL	<input checked="" type="checkbox"/>
DOWNLOAD_BODY_COCO_MODEL	<input type="checkbox"/>
DOWNLOAD_BODY_MPL_MODEL	<input type="checkbox"/>
DOWNLOAD_FACE_MODEL	<input checked="" type="checkbox"/>

图 4.1 cmake 生成目录

```
Adding Example 02_whole_body_from_image_default
Adding Example 03_keypoints_from_image
Adding Example 04_keypoints_from_images
Adding Example 05_keypoints_from_images_multi_gpu
Adding Example 06_face_from_image
Adding Example 07_hand_from_image
Adding Example 08_heatmaps_from_image
Adding Example 09_keypoints_from_heatmaps
Adding Example 10_asynchronous_custom_input
Adding Example 11_asynchronous_custom_input_multi_camera
Adding Example 12_asynchronous_custom_output
Adding Example 13_asynchronous_custom_input_output_and_datum
Adding Example 14_asynchronous_custom_input
Adding Example 15_asynchronous_custom_preprocessing
Adding Example 16_synchronous_custom_postprocessing
Adding Example 17_synchronous_custom_output
Adding Example 18_synchronous_custom_all_and_datum
Download the models.
Downloading BODY_25 model...
Model already exists.
Not downloading body (COCO) model
Not downloading body (MPI) model
Downloading face model...
Model already exists.
Downloading hand model...
Model already exists.
Models Downloaded.
Configuring done (6.0s)
```

图 4.2 cmake 完成标志

2、环境工程准备:

将生成的文件再次 visual studios 进行编译。如下图 4.3:

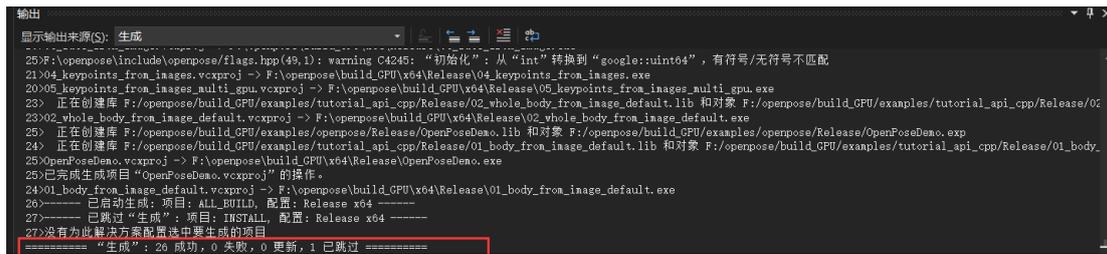


图 4.3 系统文件生成图

将编译完成的文件进行手动添加命令行参数。使用的为 176×176（4g 显存）配置图片参数。

3、虚拟环境安装

搭建 cuda 和 pytorch 适配环境方便 GPU 训练

配置 Pytorch 环境

进入环境进行安装，安装命令：

```
conda install pytorch==1.12.1 torchvision==0.13.1 torchaudio==0.12.1
  cudatoolkit=11.3 -c pytorch
```

4.2 OpenPose 框架介绍

OpenPose 算法是一种基于深度学习的人体关键点估计算法。在 OpenPose 中，身体部位的置信度图是 2D 热图，其表示在输入图像中的每个像素位置处存在相应身体部位的概率。

例如，如果我们考虑图像中的人的左肘，则左肘的置信度图将是 2D 热图，其中每个元素对应于肘部出现在该特定位置的可能性。地图中的值是归一化概率，因此它们的范围从 0 到 1，较高的值表示指定身体部位存在的概率较高。

这些置信图是使用卷积神经网络（CNNs）生成的，卷积神经网络已经在标记的人体姿势图像的大型数据集上进行了训练。神经网络结合了特征提取和图像分析技术来检测和识别人体的各个部位。

最后，这些置信度图与其他信息（如肢体连接和关节角度）一起使用，以估计图像中的完整人体姿势。这一过程构成了许多涉及人类检测和跟踪的计算机视觉应用的基础，包括活动识别、手势识别、动作识别和虚拟现实。OpenPose 算法使用了卷积神经网络（CNN）模型来实现人体关键点检测。该模型在训练集上进行了大量的训练，并且使用了一种称为 Part Affinity Fields (PAF)的技术来对关节之间的连线进行建模，同时还使用了另外一种称为 Multiple Stage (MS)的技术来加强模型的鲁棒性和提高准确率。

具体而言，OpenPose 算法采用了三个主要的阶段：人物检测、关节定位和姿态矫正。首先，该算法使用一类称为“候选区域”的技术来构建可能包含人体的区域，然后在这些区域内进行精确定位并判断其中是否包含人体。接下来，算法使

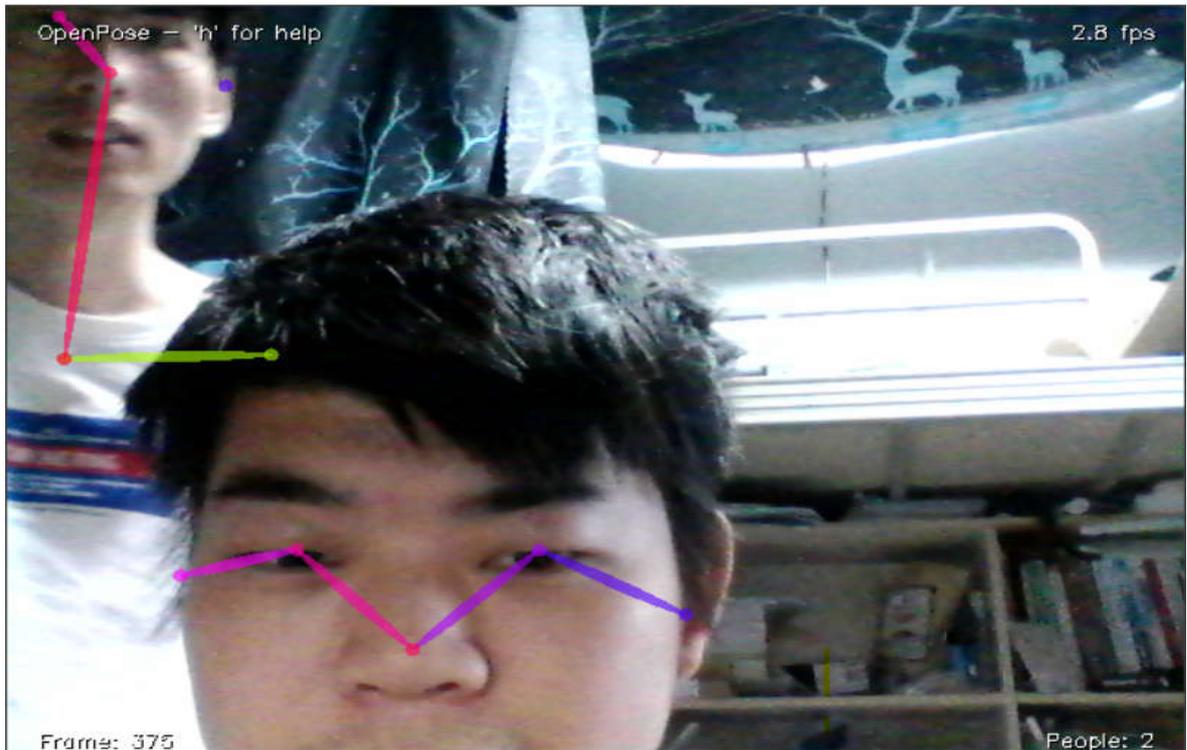


图 4.5 摄像头识别人数效果图

4.4 通过本地视频进行人数统计

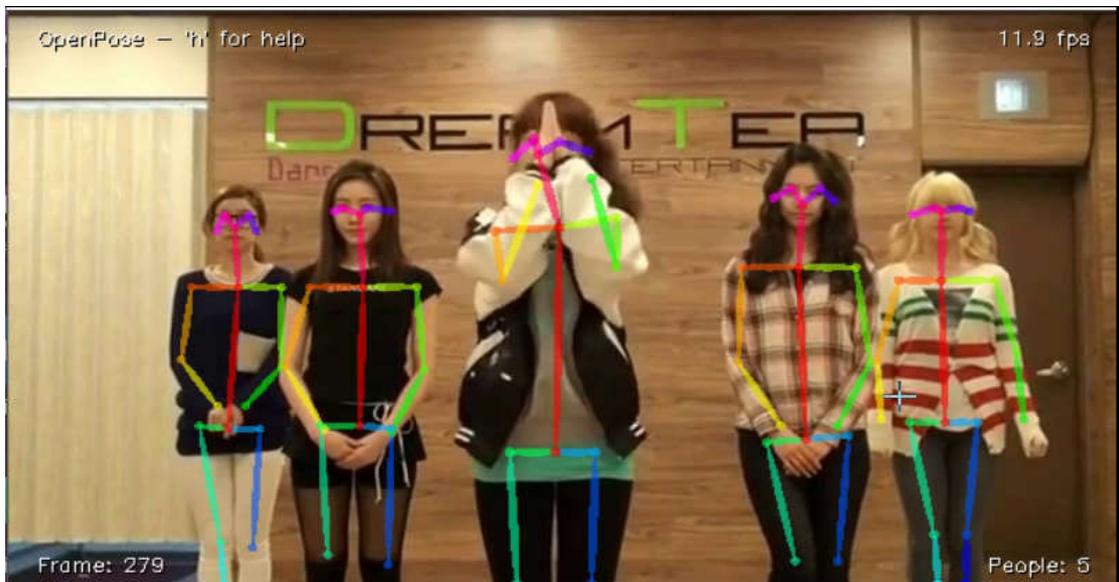


图 4.6 被识别效果图

通过准备好的本地视频，将视频文件导入算法中，从而进行识别，如上图 4.6 所示。

5 软件界面设计

上一个章节已经介绍了软件主要的处理，并将处理的结果展示了出来，由于本文的系统基础是机器视觉，需要一个外界的交互来接受后台的界面因此，设计了一个界面来体现结果。它提供了与用户交互的方式和入口，使得用户可以利用软件完成一些任务。

5.1 前端工具介绍

PyQt 是一个用于创建 GUI 桌面应用程序的 Python 模块，它是基于 C++ GUI 库 Qt 的封装。QT 是一套跨平台图形用户界面开发框架，可用于开发本地、嵌入式和移动设备应用程序。PyQt 与 PySide 一样，提供了使用 Qt 库的 Python 绑定，使得 Python 程序员能够利用 C++编写的 Qt 组件和类来制作 GUI 应用程序。

Qt 有一个理念是 Write once, compile everywhere（只写一次，到处运行），这就意味着无论是在什么系统上开发出来的程序都可以在其他的系统上迅速、轻松地移植，并且可以实现原生的操作系统界面。

PyQt 为 Python 程序员提供了良好的 API 接口，程序员可以使用 Python 语言，减少繁琐的 C++代码编写，并且能够利用 Python 的简洁性和灵活性，同时拥有 Qt 强大和美观的图形界面设计。PyQt 除了提供常规的控制件 PyQt 是 Python 编程语言和 Qt 应用程序框架的绑定，它可以让 Python 程序员使用 Qt 库中的类和方法。PyQt 提供了许多插件功能，包括：Qt Designer、QScintilla、Qwt、QGIS、PyQtGraph、PyQtWebEngine 等插件，这些插件功能可以帮助 Python 程序员更轻松地创建各种类型的应用程序，从简单的 GUI 界面到复杂的科学计算和数据可视化应用程序。

5.2 登录页面设计

本小节将介绍基于 PyQt 设计的登录页面。该页面主要部分是用户名和密码，除了以上基础构造，登录界面还考虑安全性设计和 UI 美化等方面设计了忘记密码的功能，简单设计了界面 UI，总直，为达到高效良好的效果，本次创建登录页面结合实际业务需求和用户体验设计。如下图 5.1



图 5.1 登录界面 UI

主要代码如下：

```
self.label.setText(_translate("Form", "用户名: "))  
self.label_2.setText(_translate("Form", "密码 : "))  
self.pushButton.setText(_translate("Form", "登录"))  
self.pushButton_2.setText(_translate("Form", "忘记密码"))
```

5.3 主页面选择及功能展示



图 5.2 主页面示意图

软件的主要界面主要展示两个功能，分别是本地文件导入和开启摄像头。如上图 5.2。本地视频导入可以导入事先准备好的视频进行处理，开启摄像头则为通过摄像头进行实时的统计人数，

主要按钮代码为

点击本地视频导入后就会跳转进入下图 5.3

```
Form.setWindowTitle(_translate("Form", "Form"))
self.pushButton.setText(_translate("Form", "本地视频导入"))
self.pushButton_2.setText(_translate("Form", "开启摄像头"))
```

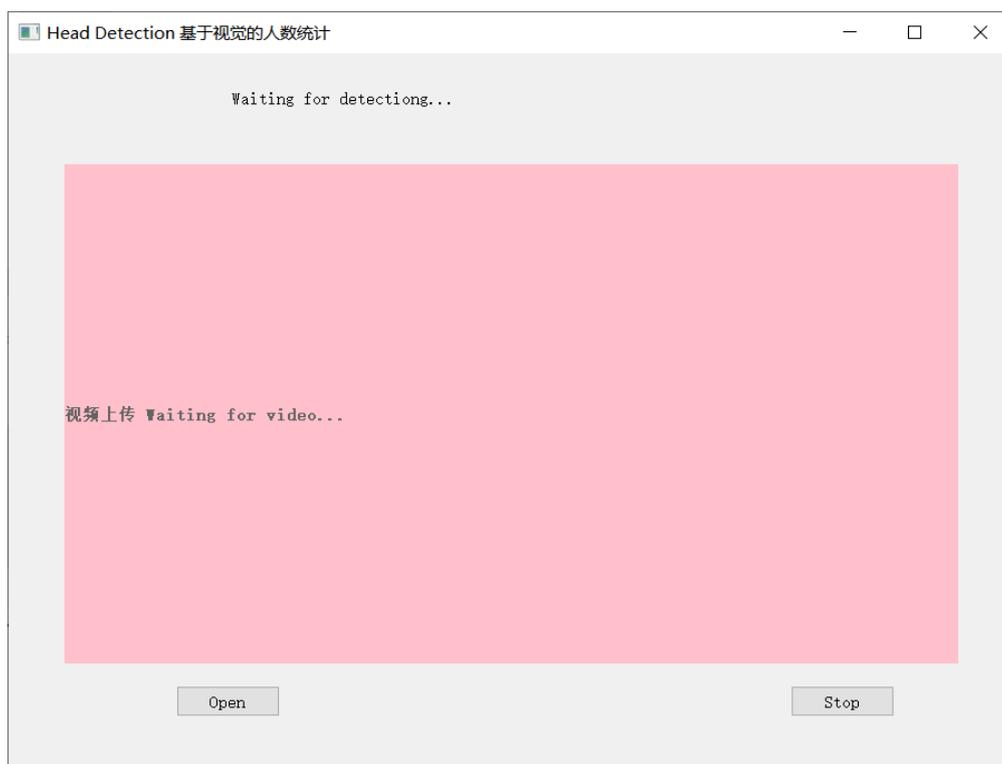


图 5.3 本地视频导入图

主要代码如下：

```
self.setWindowTitle("Head Detection 基于视觉的人数统计")
self.label.setText("视频上传 Waiting for video...")
self.label_num = QLabel(self)
self.btn.setText("Open")
self.btn_stop.setText("Stop")
```

6 实验结果测试与分析

主要包括两个方面，一方面是软件功能的实现，另一方面是 UI 界面的设计，本章节将不再赘述软件的功能了，接下来将使用系统对不同的环境目标进行评估，测试系统的准确度。

6.1 实验结果测试

由于本次研究是一种利用计算机视觉技术，对图像或视频进行分析和处理，实现自动检测和统计人数的系统。所以需要系统的人数识别能力进行必要的实验测试，基于视觉技术的人数统计系统具有以下特点：系统的实时性检测，检测系统是否能够能够快速地对输入的图像或视频进行处理，实时监测人员数量变化，并即时更新数据。系统的精度，能否使用算法和模型对人体行为进行检测和跟踪对不同场景下的实际人数进行精确计数。

但是视觉技术仍然存在着局限性，如需要考虑检测环境的光照变化、遮挡等问题，而且算法和模型也需要需要训练和调试才能达到较好效果。

基于以上考虑，本次研究选取了一段具备背景环境具有光影变换，以及人物具有复杂肢体动作的一段 1 分钟舞蹈视频:选取其中的 30s 秒进行总结观察部分图像如下所示：

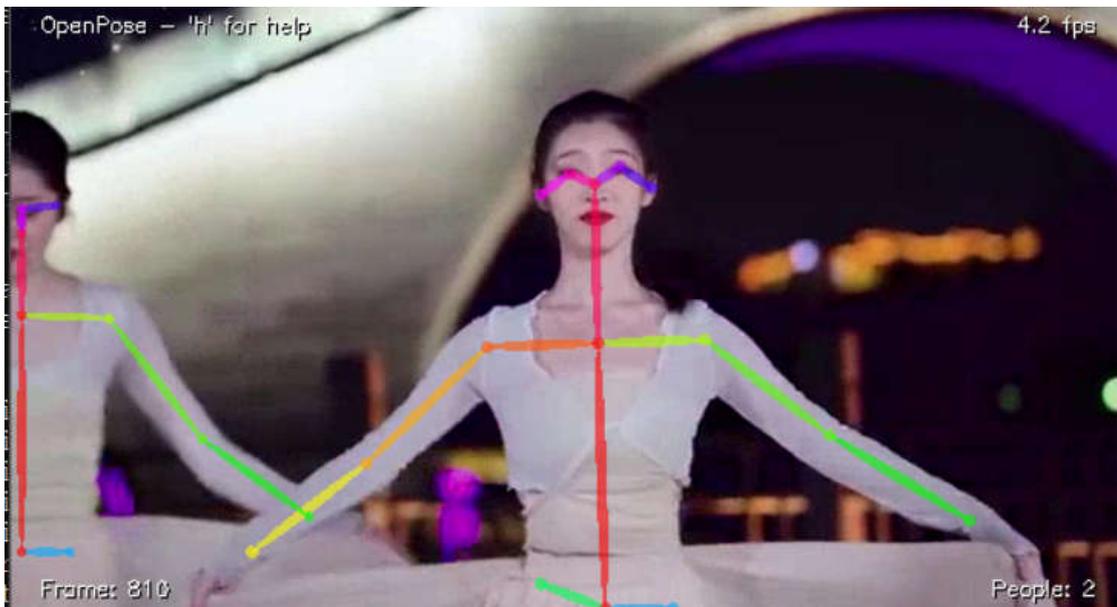


图 6.1 稀疏人数识别效果图

通过上传本地视频，后台调用视觉处理算法，当人数较少时能进行精确识别。如上图 6.1，可以看出右下角显示出了 2 个人，即使左边的舞蹈者被遮挡一大半

的身体，系统仍然可以精确地识别人。

当人数增多，或者人员仅有小部分身体处于视频检测的范围时，系统仅能检测出部分人员信息，如下图所示：

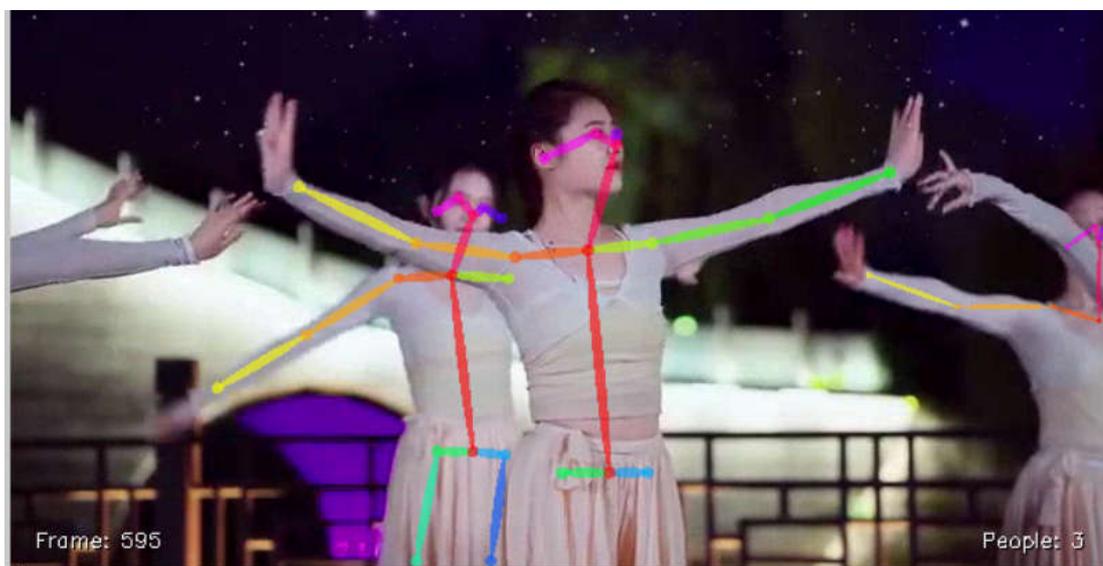


图 6.2 遮挡人数识别效果

如图 21，图片中的人员显示为 3 人，但根据图片可以看出有 6 人，有 3 人只露出了半只手臂，受限于设备和环境中光线变化不断，人员处于运动中的原因，识别率 3 人；

当增加视频中的人数，使环境变得更复杂，多人进行运动，且舞台光线不断变换的场景时，受限于设备的影响，人员的识别就变得更为困难，

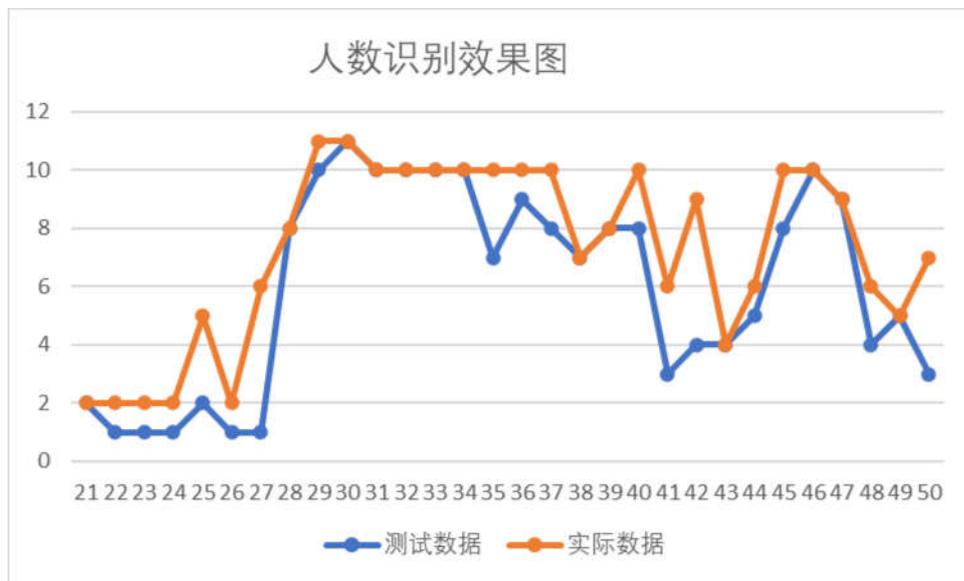


图 6.3 多人识别效果图



图 6.4 多人识别效果图

如上图 6.3 所示，一人仅露出半条胳膊，剩下多人处于运动时，人物识别了 6 人；在测试过程中，受限于测试环境的影响，可以明显感觉到人数的识别是波动性。当人物完全被识别时，如图 6.4 可以看出人物被识别为 9 人，根据选取材料截取 21 秒到 50 秒人数识别效果最终进行分析制作成折线图如下：



6.6 系统效果折线图

根据图示可知大多数时候能高于 80%的准确率，在适应光影变换后甚至可以达到 100%的程度。

结论

本课题完成了一种基于视觉技术的人数统计系统的设计与实现。该系统通过对视频流数据进行分析和处理，能够实现对目标的检测和识别，并提供不同功能选择，并制作了 GUI 界面方便算法评估者的使用。

基于视觉技术的人数统计系统可以有效地实现对人群数量的自动化统计和监控。通过对视频流进行处理，可以实现对人群数量、密度和流动情况的实时监控和分析。本文所提出的基于视觉技术的人数统计系统在实验中取得了较好的效果，并且本文对实验结果进行了评估：人流密度、背景环境、拍摄角度、相机像素等因素会对人数统计结果产生很大影响，在采取实验时需要综合考虑。同时，在测试过程中，需要特别注意数据的采集，并分析数据的质量。

本研究实现了该课题的要求，但是碍于自身水平的技术储备有限，没有设计一个更美观的界面，以及更多功能的集成，虽然能够完成基本的视频处理，但是没有提供数据集在的接口和算法的优化，仍然给系统评估者的使用带来了一些不便。

总之，基于视觉技术的人数统计系统尽管存在一些局限性，但其仍然具有很大的应用潜力，能在各个领域得到了广泛应用。而本人也将在后续的学习与生活中不断完善该课题。

参考文献

- [1]胡佳琪 王成军 杨超宇 基于深度学习的 3D 人体姿态估计研究综述
- [2]寿翌辰 基于机器视觉技术的排队人数识别系统
- [3]基于计算机视觉的电梯轿厢内人数统计研究
应俊（导师：李强）；杭州电子科技大学，硕士（专业：计算机应用技术）；2012
- [4] onuoi ror eievaorsJ mnte1596. Yu Shengsheng,Chen Xiaoping,Sun Weiping,Xie Deping. A robust method for detectingand counting people [C]/International Conference on Audio,Language and lmageProcessing.Shanghai: IEEE, 2008:1545-1549.
- [5]陈先,王民. 基于图像的客车乘员状态监测[D].北京工业大学硕士学位论文.2008.
- [6]马丽娟,曲仕茹.长途客车客流统计系统设计[D].西北工业大学硕士学位论文.2007.
- [7]张英; 陈临强; 杨礼坤, 基于目标分割与 SVM 的人数统计
- [8]基于计算机视觉的电梯轿厢内人数统计研究
应俊（导师：李强）；杭州电子科技大学，硕士（专业：计算机应用技术）；2012
- [9] 兰州理工大学 ， 基于卷积神经网络的密集人数算法研究
- [10] OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affifinity Fields, Zhe Cao , Student Member, IEEE, Gines Hidalgo , Student Member, IEEE, Tomas Simon , Shih-En Wei, and Yaser Sheikh
- [11]Multi person pose estimation based on improved openpose model Jiayuan XingJun ZhangChenxing Xue School of Electronic Engineering,Tianjin University of Technology and Education

致谢

在此，在本文完成的过程中，我受到了许多人的帮助和支持，同时也感谢成都信息工程大学的培养，使我获得了足够的知识，在我个人的成长路上起到了非常重要的作用。

首先，我要感谢我的导师。他在整个研究过程中给予了我无私的指导和支持，不仅在学术上给予了我很多宝贵的建议和指导。没有他的支持，我无法完成这篇论文。本人要对我的指导老师表达最真诚的谢意。首先，我要感谢老师给予我的宝贵机会，使我有机会能够独立进行科学研究，并从中学到了许多实用和有益的知识。

其次，我要感谢我的家人和朋友。他们在我完成论文的过程中给予了我很多的鼓励和支持，让我能够克服困难和挫折，坚持完成这篇论文。最后，我要感谢学科老师教授的多门课程，为我的整个研究工作提供了坚实的理论基础，并为我的毕业论文做出了巨大的贡献。

最后，我要感谢评委老师们，感谢你们的耐心、认真和细心，为我们的项目提出了宝贵的意见和建议。没有他们的支持和帮助，我将可能无法顺利的完成这篇论文。

作者简介：

姓名：汪洋

性别：男

出生年月：1998年12月

民族：汉族

E-mail:3298404175@qq.com

声明

本论文的工作是 2022 年 12 月至 2023 年 5 月在成都信息工程大学自动化学院完成的。文中除了特别加以标注地方外，不包含他人已经发表或撰写过的研究成果，也不包含为获得成都信息工程大学或其他教学机构的学位或证书而使用过的材料。

关于学位论文使用权和研究成果知识产权的说明：

本人完全了解成都信息工程大学有关保管使用学位论文的规定，其中包括：

(1) 学校有权保留并向有关部门递交学位论文的原件与复印件。

(2) 学校可以采用影印、缩印或其他复制方式保存学位论文。

(3) 学校可以学术交流为目的复制、赠送和交换学位论文。

(4) 学校可允许学位论文被查阅或借阅。

(5) 学校可以公布学位论文的全部或部分内容（保密学位论文在解密后遵守此规定）。

除非另有科研合同和其他法律文书的制约，本论文的科研成果属于成都信息工程大学。

特此声明！

作者签名：汪洋

2023 年 5 月 14 日