

全国大学生物联网设计竞赛

基于低质量红外摄像的 跨模态行人重识别系统



基于低质量红外摄像的 跨模态行人重识别系统

random 团队

队员 1: 王亿宽

队员 2: 冉丰源

队员 3: 忻颖杰

队员 4: 覃威勋

摘 要

在当前快速发展的城市化进程中，智能监控与公共安全成为城市管理的关键议题。为了应对传统监控在低光照及恶劣天气条件下识别能力受限的挑战，本设计提出了一种基于低成本红外摄像的跨模态行人重识别系统。

系统采用火山引擎边缘计算开发板作为硬件基础，通过部署先进的深度神经网络模型，实现跨模态图像特征对齐，确保即使在低质量红外图像条件下，也能保持稳定的识别效果。数据增强技术被用来扩充训练数据集，优化了模型在红外图像的泛化能力和鲁棒性。系统还利用了云边协同架构，结合云端的大数据处理能力与边缘设备的实时响应优势，构建了一个高效、灵活的监控系统。在云端，系统可以进行大规模的数据分析和模型训练，而边缘设备则负责实时数据处理和本地响应，从而减少网络延迟，增强整体系统的稳定性和响应速度。

综上，本设计不仅能够克服低成本红外设备图像质量的局限，还能够平衡隐私保护与识别精度，通过红外图像减少对个体面部特征的直接捕捉，从而在保障公共安全的同时，尊重个人隐私。此外，系统还实现了本地数据处理和实时响应，减少了网络延迟，增强了整体系统的稳定性和响应速度。

关键词：可见光-红外跨模态行人重识别、智能监控、云边协同架构

目 录

一、 设计需求分析.....	1
1.1 设计背景	1
1.2 课题背景	2
1.3 研究现状	3
1.3.1 技术现状.....	3
1.3.2 应用现状.....	4
1.4 课题研究意义及任务	6
1.4.1 课题研究的意义.....	6
1.4.2 课题研究的任务.....	7
1.5 报告内容组织形式	7
二、 特色与创新.....	9
2.1 系统特色	9
2.2 系统实现难点	10
2.3 系统实现创新点	10
三、 功能设计—系统总体设计.....	12
3.1 设计原则	12
3.2 设计目标	14
3.2.1 服务器设计.....	14
3.2.2 网页前端设计.....	15
3.3 系统硬件选择	16
3.3.1 英伟达 Jetson Orin 开发板.....	16
3.3.2 锐创微纳 Raytron 摄像头.....	17
3.4 系统软件平台	18
3.4.1 开源深度学习库 PyTorch.....	18
3.4.2 Vue.js.....	19
3.4.3 MySQL 数据库	20
3.4.4 Epoll 高并发数据传输模型	21

四、 功能设计—系统架构设计	23
4.1 总体架构设计	23
4.2 功能模块设计	24
4.2.1 数据采集终端模块.....	24
4.2.2 数据接收节点模块.....	25
4.2.3 数据网骨干节点模块.....	26
4.2.4 服务器信息处理模块.....	27
4.3 系统总体运作流程	28
五、 系统实现.....	30
5.1 行人图像采集	30
5.2 基于 CA（随机通道可交换增强）和 CRE（通道级随机擦除）策略 的数据增强技术.....	30
5.2.1 随机通道可交换增强策略.....	31
5.2.2 通道级随机擦除策略.....	32
5.3 基于 ECML（增强型渠道混合学习）和 CAJL（通道增强联合学 习）的跨模态度量算法.....	34
5.3.1 增强型渠道混合学习.....	34
5.3.2 通道增强联合学习.....	36
六、 结果展示.....	38
6.1 应用系统展示	38
6.1.1 登录界面展示.....	38
6.1.2 强光照时界面展示.....	39
6.1.3 弱光照时界面展示.....	39
6.1.4 比对界面展示.....	40
6.2 算法结果分析	40
七、 参考文献.....	43

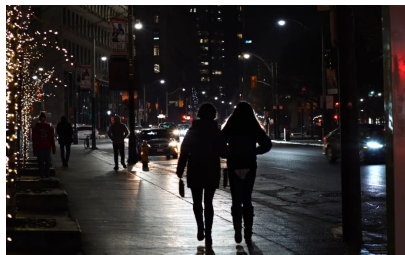
一、设计需求分析

内容提要	
可见光-红外跨模态行人重识别	智能安防
面向复杂场景识别	火山引擎开发板

1.1 设计背景

随着城市扩张步伐的加速和智能化水平的提升，公共安全和智能监控成为现代城市管理者的首要关切。与此同时，民众对于个人安全日益增长的需求，正推动着家庭监控系统的普及与升级。然而，传统监控摄像头在低光照或恶劣气候条件下表现不尽如人意，其模糊不清的画面质量不仅阻碍了人眼对于画面中个体特征的辨识，更难以达到现今计算机视觉技术对精准人脸识别的要求。

红外摄像头相较于传统摄像头，对光线条件要求更低，可以保证全天候的监控需求。它们在诸如雨雪、雾霾等极端气候环境下，依旧能够维持高效运作，保证监控网络的连续性和可靠性。尤为重要是，在隐私保护层面，红外摄像头能够在不对个人隐私造成侵扰的前提下，采集到高质量的图像数据，这使其特别适合于对隐私高度敏感的公共区域使用。因此 random 团队通过引入新的计算机识别技术开发出可见光-红外跨模态行人重识别系统，并将其应用于家庭和公共环境，充分体现智能化管理的思想，减轻了人力识别监控图像的负担。



夜晚可见光图像



热红外夜晚图像

图 1

1.2 课题背景

当前，人们的物质生活水平越来越高，对居住环境以及社会稳定的要求也越来越高，伴随着智能化城市的创建，人们摄像监控的安防系统的要求也会更高。随着科学技术的不断进步，特别是物联网技术的普及应用，智能安防已成为当前的发展趋势。智能安防技术涵盖了信息化、图像传输和存储、数据存储和处理等方面，为现代城市的综合管理体系提供了强有力的技术支持。智能安防系统的主要组成部分包括门禁系统、报警系统和视频监控系统，这些系统通过传感器和摄像头设备，实现了对环境的实时监控和异常情况的及时预警。

但是在低光和复杂环境下，传统摄像头表现的局限性逐渐凸显，尤其在夜晚和恶劣天气下传统可见光摄像头难以提供高质量的监控图像。这一问题靠红外摄像头可以得到解决，基于红外摄像头开发行人重识别技术以提高监控系统的全天候能力和智能化水平变得尤为重要。

行人重识别（Person Re-Identification, ReID）作为智能监控与视频分析领域内的一项核心议题，其核心目标在于跨摄像头网络识别并匹配同一行人。以往的方法多侧重于利用可见光谱图像，然而这种方法对成像设备与外部环境条件有着苛刻的要求。相比之下，红外摄像头因其对环境因素的低敏感度以及技术的成熟度，能在成本效益较高的情况下捕获清晰的热成像图，这为跨模态行人重识别的研究开辟了新路径。鉴于此，我们有必要对跨模态的行人重识别进行研究，针对红外图像提出新的解决方案。

1.3 研究现状

1.3.1 技术现状

目前红外摄像头技术已经非常成熟，高灵敏度的红外传感器可以在极低光照条件下捕捉到清晰图像。加之现代热成像技术充分利用了红外线的热辐射特性，确保了在无光照条件下仍能生成可视化的热辐射图像。近年来，深度学习的飞速发展，尤其是卷积神经网络（CNN）的广泛应用，显著增强了行人重识别（ReID）系统的精确度。时下，ResNet、DenseNet 等深层网络结构成为主导，它们通过从可见光图像中抽取高维特征向量，强化了行人重识别技术的鲁棒性和区分性。但是，经过市场调研，公共安全及家庭安防领域所采用的行人重识别技术普遍基于可见光图像。在遭遇恶劣气候或低照度环境时，这种依赖性会导致监控视频丧失有效的特征识别能力。值得注意的是，尽管红外摄像机能够产生适应各种照明条件的

图像，但在行人重识别系统中的应用却相对有限，特别是在成本控制严格的场景下，低成本红外设备几乎未被纳入 ReID 研究的考量范畴。



图 2 可见光图像的行人重识别

1.3.2 应用现状

我国的智能安防系统起步相对较晚，过半的市场份额都被美国、日本、韩国等国的企业主导。但自 2008 年北京奥运会及后续的冬奥会、世博会等重大国际活动举办以来，中国智能安防产业经历了快速发展，预计到 2026 年，其市场规模将突破 2500 亿元人民币。截至 2020 年，智能门锁与视频监控为代表的物联网设备已占据了智能安防产业超过 55% 的份额。同年 7 月，随着《关于开展 2020 年网络安全技术应用试点示范工作的通知》的出台，对智能家庭、

智能安防等场景中的网络安全需求得到了进一步的强化，这标志着相关领域正展现出巨大的市场潜力。智能安防的终端应用可划分为城市级、行业级以及消费级。近年来，“平安城市”、“雪亮工程”和“智慧城市”等工程陆续开展，城市级应用约占智能安防市场份额的30%-40%。行业级方面，智能安防广泛应用于公安、交通、金融、能源、文教卫等行业。消费级方面，智能安防产品应用推广至社区、家庭，随着安防系统成本的降低和智能化的推进，民用市场的潜在需求较大。而结合红外摄像头的跨模态行人重识别系统具有广泛的适用性，不仅适用于城市夜间重点区域监控，还能用于智能交通管理，帮助迅速复原事故现场，支持事故责任判定。在智能小区和工业园区入口，这一系统能够即时探测潜在威胁，确保区域安全。在监狱等特定环境中，红外行人重识别系统可在夜间提供连续监控，防止暴力事件发生，提升监管效率和安全性。



图 3 中国智能安防硬件市场规模变化情况

1.4 课题研究意义及任务

1.4.1 课题研究的意义

虽然当今社会经济发展飞速，人们的生活水平普遍提高，并且法律体系日益完善，犯罪率大大下降，但仍有犯罪分子伺机而动，尤其偏爱在光线昏暗的街巷实施盗窃或抢劫，以此规避传统监控系统的监视。另外，夜间光线条件不佳也是导致交通事故频发的一大因素，此时，常规监控设备往往难以清晰捕捉车内人员的动作细节，更难以精准辨识面部特征，这对事故责任的判定造成了困扰。因此，对红外监控及其识别系统的深入研究显得尤为紧迫和关键。在构建监控系统时，可靠性和成本效益是两个核心考量因素。可靠性保证监控系统对人的特征识别的准确性，确保能在最大程度上准确的识别出人脸。成本控制则是决定系统能否广泛部署的关键，高昂的价格会限制其使用范围，即便技术性能卓越，也难以普及至更广泛的区域。唯有在功能与成本之间找到平衡点，才能确保该系统在社会各领域得到广泛应用。

本文研究的红外摄像头行人重识别系统，基于跨模态的算法实现了 RGB 可见光图像与 IR 红外图像之间的有效映射，补齐了红外图像特征。由于采用了低成本的设计理念，该系统能够在不同角度的摄像头下准确辨认行人，既保证了识别的可靠性，又兼顾了经济效益，这使得其在社会安全防护、保障民众生命财产安全方面展现

出了巨大潜力。现有的安防系统普遍为视频监控系统，视频监控存在成本高、功耗大、信息储存量庞大、容易躲避、容易侵犯无关人员隐私等缺陷。本作品提出的基于对抗生成式网络的跨模态行人重识别系统相较于传统的监控系统具有易部署、低成本、保护行人隐私、存储便利、可靠性好等优势，对于提升总体的安防效果具有重大意义。

1.4.2 课题研究的任务

基于跨模态的图像识别技术，结合已有的研究成果，本作品的主要任务是建立一个能在光线不足和天气恶劣的情况下也普遍适用的红外监控系统，需要完成的具体任务如下：

- 构建可见光-红外映射网络，基于深度学习实现跨模态图像特征对齐
- 基于火山引擎边缘计算开发板，部署监控系统
- 完成识别效果的可视化，完成 Web 页面设计
- 实现跨模态行人重识别系统总体设计

1.5 报告内容组织形式

本文共七章，各个章节内容安排如下：

第一章：设计需求分析。主要介绍 random 团队开发的可见光-红外跨模态行人重识别系统的背景以及研究现状，并说明了本课题的研究意义和研究任务。

第二章：特色与创新。主要指出我们开发的应用系统在算法设计和功能设计上的创新点和难点，讲出了我们的特色。

第三章：系统总体设计。从设计原则、设计目标硬件选择和软件选择上讲解了 random 团队的设计选择和出发点。

第四章：系统架构设计：详细介绍了系统架构和系统各层的技术实现，讲解了系统运作流程，对我们的应用做出了基本介绍。

第五章：系统实现。主要讲解我们的可见光-红外跨模态行人重识别技术的算法，是整个系统的核心部分。

第六章：结果展示。对我们的应用运行结果和优越性进行展示，总结全文的研究工作，分析工作的实施进展情况，并提出后期工作的展望。

第七章：参考文献。

二、 特色与创新

内容提要	
集成深度学习	低成本红外兼容
数据增强技术	云边多协同开发

2.1 系统特色

- 跨模态识别技术：本系统创造性地将可见光图像与红外图像进行融合，解决了单一模态图像识别在特定环境下的局限性，实现了全天候、全环境下的高精度行人识别。
- 低成本红外设备兼容：通过优化算法和模型训练策略，系统成功克服了低成本红外设备的图像质量限制，使得系统不仅高效而且经济，适用于大规模部署。
- 隐私保护机制：红外图像的使用减少了对个体面部特征的直接捕捉，从而在保障公共安全的同时，最大限度地尊重和保护了个人隐私。
- 智能化数据分析：系统集成深度学习算法，自动分析和识别图像中的行人，减轻了人工监控的压力，提高了识别效率和准确性。
- 边缘计算部署：利用火山引擎边缘计算开发板，实现了本地数据处理和实时响应，减少了网络延迟，增强了系统的稳定性和响应速度。

2.2 系统实现难点

- 跨模态特征提取：红外图像与可见光图像在特征表达上有本质差异，如何在不同模态间建立有效映射，是本系统设计中的一大挑战。
- 数据集构建：缺乏足够的跨模态行人图像数据集，需要自行采集和标注大量样本，以训练和验证模型的性能。
- 模型泛化能力：确保模型在不同环境、不同光照条件和不同人群中的通用性和稳定性，需要大量的测试和调整。
- 隐私保护与识别精度权衡：在提高识别精度的同时，必须确保不侵犯个人隐私，找到两者间的最佳平衡点。

2.3 系统实现创新点

- 数据增强技术：利用生成对抗网络（GAN）生成合成的跨模态图像，扩充了训练数据集，提升了模型在低质量红外图像上的识别效果。

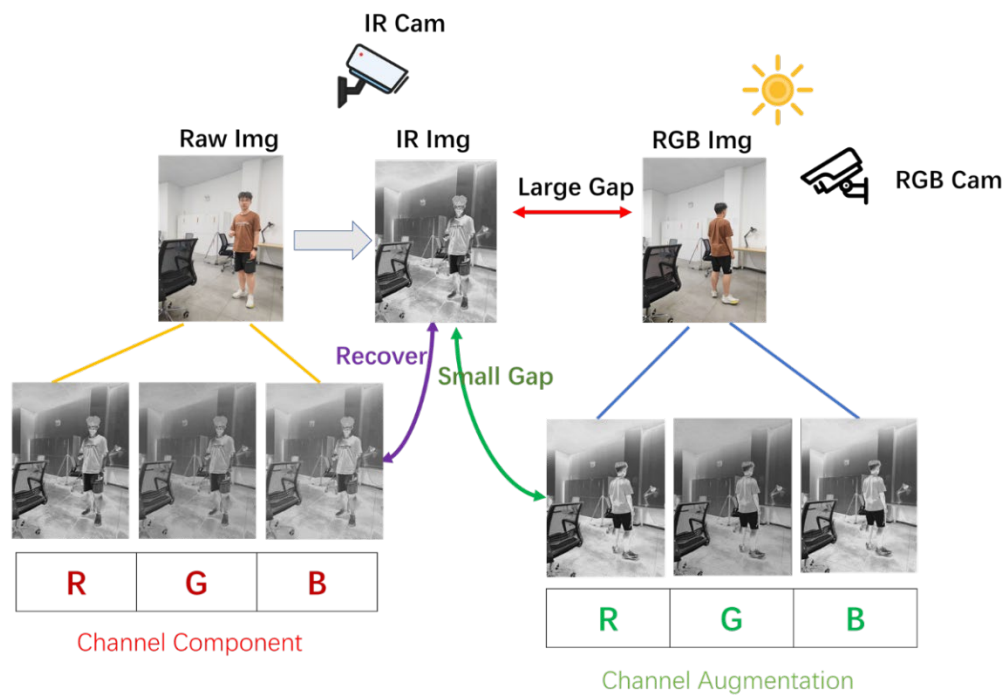


图 4 通道增强算法

- 云边协同架构：结合火山的云端大数据处理能力和火山引擎提供的边缘设备的实时响应优势，构建了一个高效、灵活的监控系统。



图 5 云边协同架构

三、 功能设计—系统总体设计

内容提要	
网页前端开发	系统硬件
服务器设计	系统软件

3.1 设计原则

本作品的核心设计理念紧密遵循物联网系统设计的五大原则：

- 多样性原则：物联网设计的体系结构必须根据不同的物联网节点类型，分为多种类型的体系结构，确保全面覆盖和互联互通；
- 时空性原则：物联网设计的体系结构必须兼顾时间响应与空间分布，同时优化能耗，以满足物联网在动态环境中的高效运作需求；
- 互联性原则：物联网设计的体系结构必须能够无缝对接互联网，确保数据的实时传输与快速响应，强化物联网与互联网的融合度；
- 安全性原则：物联网设计的体系结构必须构筑牢固的防御体系，能够做到防范大范围网络攻击、保障用户隐私及数据安全；
- 坚固性原则：物联网设计的体系机构必须具备坚固性和可靠性，确保在复杂环境下持续运行无虞。

遵循上述原则，我们构建了一个可见光与红外热成像结合的跨模态行人重识别系统，其流程可概括为：设计物联网应用系统：旨在创建一个直观的在线监控平台，能根据夜晚的作案现场准确地识别白天出现过的可疑人员，并且提供用户友好的交互界面，提升用户体验。

构建物联网应用平台：运用火山引擎开发板，搭建完整的前后端架构，连接各类物联网终端，完成网络配置，打造集用户管理、数据传输于一体的综合性 Web 应用。

捕捉行人图像：分别利用可见光摄像头和红外摄像头捕获行人图像，将获取的图像资料传输至服务器，借助生成对抗网（GAN）进行跨模态数据分析，实现精准匹配与识别。

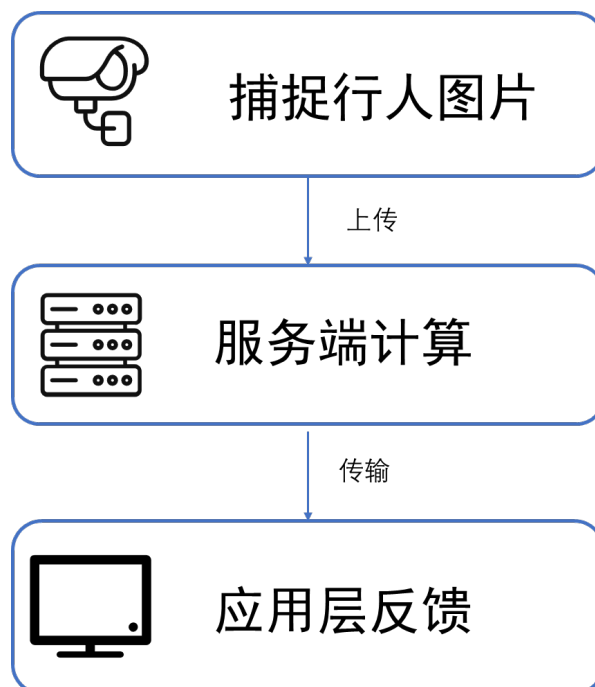


图 6 应用设计示意图

3.2 设计目标

鉴于实际应用场景与物联网技术的独特优势，我们计划利用火山引擎开发板作为技术支撑，整合前沿的研究成果，致力于构建一套先进的行人重识别监控体系。该体系将聚焦于实现可见光与红外热成像的跨模态识别，以期在各种光照条件下都能精准地辨识与追踪行人的身份，从而显著提升公共安全领域的监控效能。

3.2.1 服务器设计

本项目采用现代技术栈进行开发，旨在打造一个高效、灵活且易于扩展的系统架构。系统核心功能包括平台间的信息交流、数据的有效管理以及直观的数据展示，确保数据流的畅通无阻。用户仅需通过浏览器即可轻松接入 WEB 界面，实现对系统的便捷操作，涵盖数据录入、检索及深度分析等全方位功能。整个设计流程严格遵守软件即服务（SaaS）的理念，系统能够依据用户需求动态调整资源，保证了出色的可维护性和扩展性。

项目采用前后端分离的开发模式进行 WEB 应用开发，前端开发采用 Vue.js 框架，利用其灵活的组件化设计和强大的响应式能力，构建用户友好的交互界面。前后端之间的数据通信基于行业标准的 HTTP 协议，确保了数据传输的安全与稳定。在数据交换上，我们遵循键值对（KVP）规范，并选用 JSON 作为数据传输格式。JSON 格式不仅轻量级，而且易于解析和生成，适合前后端的数据交换。

后端开发基于 Python 的 Django 框架，使用 MySQL 数据库进行数据存储和管理。Django 提供了强大的管理界面和 ORM 功能，使得数据操作更加便捷和安全。为了实现高效的请求转发和负载均衡，使用 Traefik 作为反向代理。整个应用程序运行在 Docker 容器中，通过容器化技术确保环境的一致性和部署的简便性。借助 Docker 和 Traefik，大大简化了环境配置和服务管理，提升了开发效率和系统的稳定性。



图 7 Django 框架

3.2.2 网页前端设计

网页设计采用 Vue 框架。Vue.js 是一个开源前端 JavaScript 框架，用于构建用户界面和单页面应用。Vue.js 以其轻量级、易用性和高性能，迅速赢得了开发者的青睐。它的设计理念是通过渐进式的方式，提供核心库的基础功能，并通过生态系统中的各种工具和库扩展其功能。

网页前端设计的主要功能是依托用户可操作性强、用户使用友好、观感上乘的观念设计的。保证监控管理员能按照管理员权限对监控视频实施监控，能借助可见光-红外跨模态识别应用快速找到可疑人员。前端数据通信基于 HTTP 协议，遵循 KVP 标准，并通过 JSON 格式传输给后端实现计算。

3.3 系统硬件选择

3.3.1 英伟达 Jetson Orin 开发板

火山引擎提供的英伟达 Jetson Orin 开发板是 NVIDIA 推出的新一代嵌入式 AI 计算平台，旨在提供强大的计算能力和高效的能耗比，支持从边缘设备到自主机器等各种应用场景。Jetson Orin 系列开发板基于 NVIDIA Orin 系统级芯片（SoC），包含 7 个具有相同架构的模组，可提供每秒高达 275 万亿次浮点运算 (TOPS) 的算力，其性能是上一代多模态 AI 推理的 8 倍，并支持高速接口。集成了 NVIDIA Ampere 架构的 GPU 和第八代 NVIDIA Carmel ARM CPU，提供了强大的 AI 性能。

核心特性：

- 高性能计算： Jetson Orin 配备了大量的 CUDA 核心和 Tensor 核心，提供高达 200TOPS（Tera Operations Per Second）的 AI 计算能力。
- 高效能耗比： Jetson Orin 开发板不仅拥有强大的 AI 计算能

力，而且其在设计之初还注重能效，因此该开发板十分适合低功耗的嵌入式应用。

- 先进的 AI 和机器学习能力：支持 NVIDIA TensorRT、cuDNN、CUDA 等软件栈，优化深度学习推理和训练性能。
- 丰富的 I/O 接口：包括 USB、PCIe、HDMI、以太网连接器、MIPI 相机连接器等接口，支持多种传感器和外设的连接。

英伟达 Jetson Orin 开发板作为新一代嵌入式 AI 计算平台，以其强大的计算能力、高效的能耗比和丰富的 I/O 接口，成为边缘计算和自主机器等应用场景的理想选择。Jetson Orin 不仅支持复杂的 AI 和机器学习任务，还为开发者提供了全面的软件支持和灵活的硬件拓展能力。

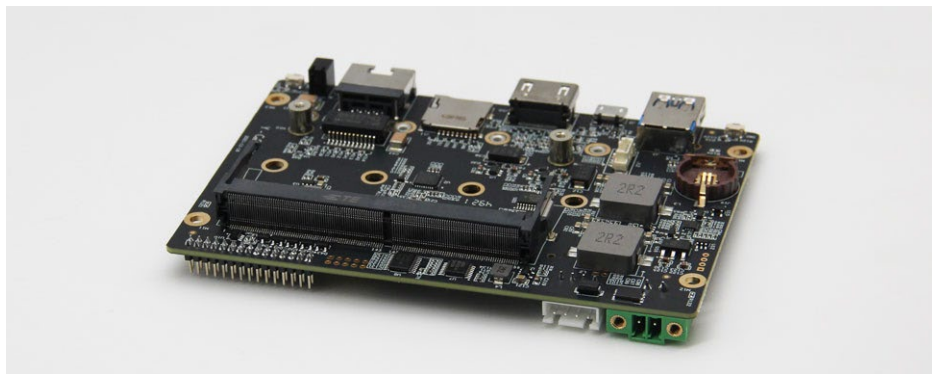


图 8 英伟达 Jetson Orin 开发板

3.3.2 锐创微纳 Raytron 摄像头

锐创微纳（Raytron Technology Co., Ltd.）是一家专业从事红外探测器及热成像系统研发、生产和销售的高科技企业。其摄像头产品广泛应用于工业检测、安防监控、医疗诊断、科研教育等多个领

域。该摄像头采用先进的红外探测器技术，具有高灵敏度的侦测能力，提供高分辨率的热成像，确保图像清晰，细节丰富；并且支持宽动态范围成像，具备快速响应能力，适用于长期监控和自动化检测系统，提供稳定的热成像数据输出。



图 9 锐创微纳 Raytron 摄像头

3.4 系统软件平台

3.4.1 开源深度学习库 PyTorch

PyTorch 是一个由 Facebook's AI Research lab (FAIR) 开发的开源深度学习库，它于 2016 年首次发布。PyTorch 以其灵活性和动态计算图的特性迅速成为了深度学习研究和应用的热门工具。它可以运行在 Linux、Windows 和 Mac OS 操作系统上，并且提供了 Python 和 C++ 接口，支持从研究到生产的各种深度学习任务。

PyTorch 采用动态计算图 (Dynamic Computational Graph)，使得模型构建和调试更加灵活和直观，方便处理复杂的模型结构和动态输入，并且内置强大的自动微分 (Autograd) 机制，轻松实现梯

度计算，支持复杂的神经网络训练。

Pytorch 提供了高层次的 API，如 `torch.nn`、`torch.optim`、`torch.utils.data` 等，简化了模型构建、优化和数据处理。并且它天然支持 CUDA 加速，能够充分利用 NVIDIA GPU 的计算能力，加速深度学习模型的训练和推理，与常用的深度学习工具和框架（如 TensorBoard、ONNX 等）深度集成，支持模型可视化、导出和跨平台部署。在经过适当的修改后，基于 Pytorch 的代码 经过修改后可以正常的运行在物联网系统中，因此我们移植 Pytorch 库用于智能监控。

3.4.2 Vue. js

Vue 是一个开源前端 JavaScript 框架，用于构建用户界面和单界面应用。Vue.js 以其轻量级、易用性和高性能的特点，通过渐进式的方式，提供核心库的基础功能，并通过生态系统中的各种工具和库扩展其功能。



图 10 Vue 框架

核心特性：

- 渐进式框架：Vue.js 可以逐步集成到现有项目中，也可以用于构建复杂的单页面应用。
- 声明式渲染：通过简洁直观的模板语法，将数据绑定到 DOM，使得开发更加高效。
- 组件化：Vue.js 的组件系统使得代码复用性高，维护性强，适合构建复杂的应用。
- 响应式数据绑定：Vue.js 采用响应式系统，自动追踪数据变化并更新视图，无需手动操作 DOM。
- 生态系统丰富：Vue.js 有强大的生态系统，包括 Vue Router、Vuex、Vue CLI 等，支持复杂应用的开发。

通过创建一个 Vue 实例来创建 Vue 应用，使用一系列内置指令（如 v-bind、v-if 等）来处理常见的 DOM 操作，使得 Vue 开发变得更为简介简洁方便，代码 s 可复用性更强。无论是小型项目还是大型应用，Vue.js 都能提供灵活和高效的解决方案，帮助开发者提升开发效率和用户体验。因此本应用基于 Vue 开发前端页面，减少开发的时间成本，增强应用的用户观感。

3.4.3 MySQL 数据库

MySQL 是一个关系型数据库管理系统，由 C 和 C++语言编

写。旨在为 Web 应用提供高性能和可靠的数据存储解决方案。它的特点是高性能、易部署、易使用，存储数据非常方便。MySQL 是高性能开源关系型数据库，也是目前最受关注的数据库技术之一，以其成熟、稳定和对企业应用友好（支持事务，一致性和数据完整性保证，有大量企业应用案例）而著称。主要功能特性有：

- 关系型数据库：MySQL 遵循关系型数据库的设计原理，数据以表格形式存储，支持数据之间的关系和完整性约束。
- 高性能：MySQL 通过多种优化技术，如索引、查询缓存等，提供高效的数据存储和查询性能。
- 高可用性：支持复制和集群技术，确保数据的高可用性和故障恢复能力。
- 事务支持：支持 ACID（原子性、一致性、隔离性、持久性）事务特性，确保数据一致性和完整性。
- 多语言支持：提供多种编程语言的 API 接口，包括 C、C++、Java、Python、PHP、Ruby 等。
- 开放源码：MySQL 是开源软件，用户可以自由获取和修改源代码，灵活定制数据库系统。

3.4.4 Epoll 高并发数据传输模型

Epoll 是 Linux 内核为处理大批量文件描述符而作了改进的 poll，是 Linux 下多路复用 IO 接口 select/poll 的增强版本，它能显著提高程序在大量并发连接中只有少量活跃的情况下的系统 CPU 利用

率。另一点原因就是获取事件的时候，它无须遍历整个被侦听的描述符集，只要遍历那些被内核 IO 事件异步唤醒而加入 Ready 队列的描述符集合就行了。Epoll 除了提供 select/poll 那种 IO 事件的水平触发（Level Triggered）外，还提供了边缘触发（Edge Triggered），这就使得用户空间程序有可能缓存 IO 状态，减少 `epoll_wait`/`epoll_pwait` 的调用，提高应用程序效率。相比起其他模型，Epoll 主要体现出的优点如下：

- 支持一个进程打开大数目的 socket 描述符；
- IO 效率不随 FD 数目增加而线性下降；
- 使用 `mmap` 加速内核与用户空间的消息传递；

在本项目中，将使用 Epoll 模型来处理服务器高并发状态，提高服务器处理大量视频数据传输的速度。

四、 功能设计--系统架构设计

ASSF 架构	内容提要	功能模块设计
---------	------	--------

4.1 总体架构设计

可见光-红外跨模态行人重识别系统总体采用 ASSF(access—service-standard-fundation)架构模式进行系统设计。系统整体架构图如图所示。



图 11 系统整体架构图

对各层说明如下：

- a. 基础层：仅包含负责信息的传输以及交换的网络，设备以及信息交换的协议。图像传感网负责针对行人的图像处理，进行可见光图像以及红外图像采集；数据骨干网负责传输图像传感网传输给服务器的信息以及视频监控信息。信息交换协议保证图像传感网与数据骨干网之间信息交换的正确性，完整性。
- b. 标准层：负责对于基础层采集的数据进行建模分析，处理储存，进行可见光->红外图像的学习，姿态检测，行人特征检测。
- c. 功能层：负责对用户进行管理，为用户行人重识别服务，将标准层处理后的数据去除冗余无用的信息并转化为于用户有用的形式。并处理来自用户的请求，做出相应的功能。
- d. 访问层：即用户层，包含用户进入的 Web 应用和后台管理人员 PC 端管理软件。

4.2 功能模块设计

4.2.1 数据采集终端模块

数据采集终端包含两种类型：可见光图像与红外测温图像。其任务是负责采集照明充足状态下的可见光图片和光线不足时的红外图片，采集流程如图所示：

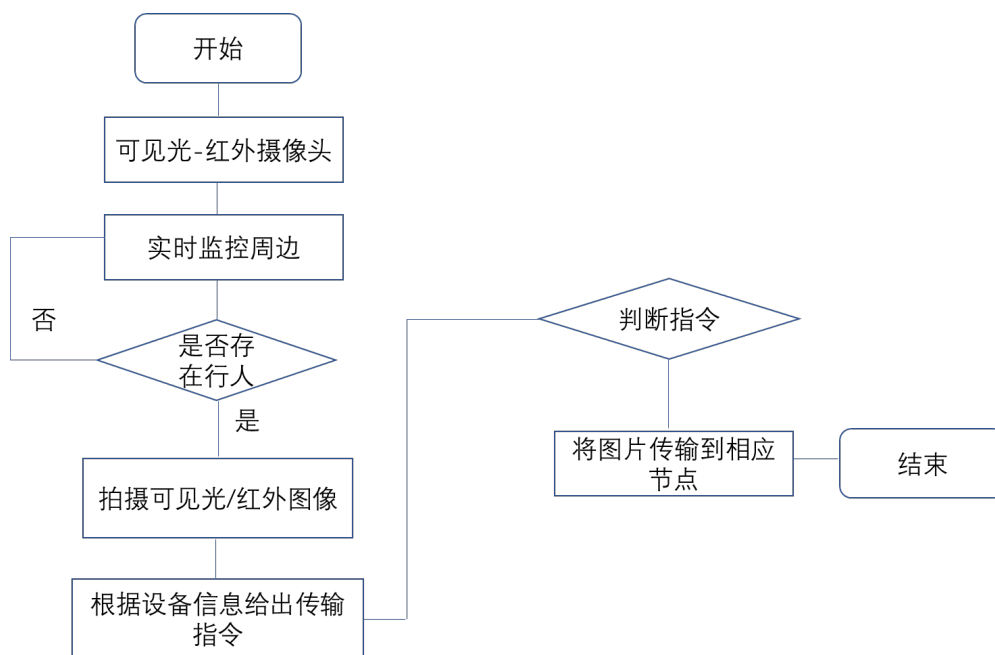


图 12 数据采集流程图

4.2.2 数据接收节点模块

数据接收节点是图像传感网与数据骨干网的网关，其负责将图像传感网的信息传输到数据骨干网中。其处理流程如图所示。

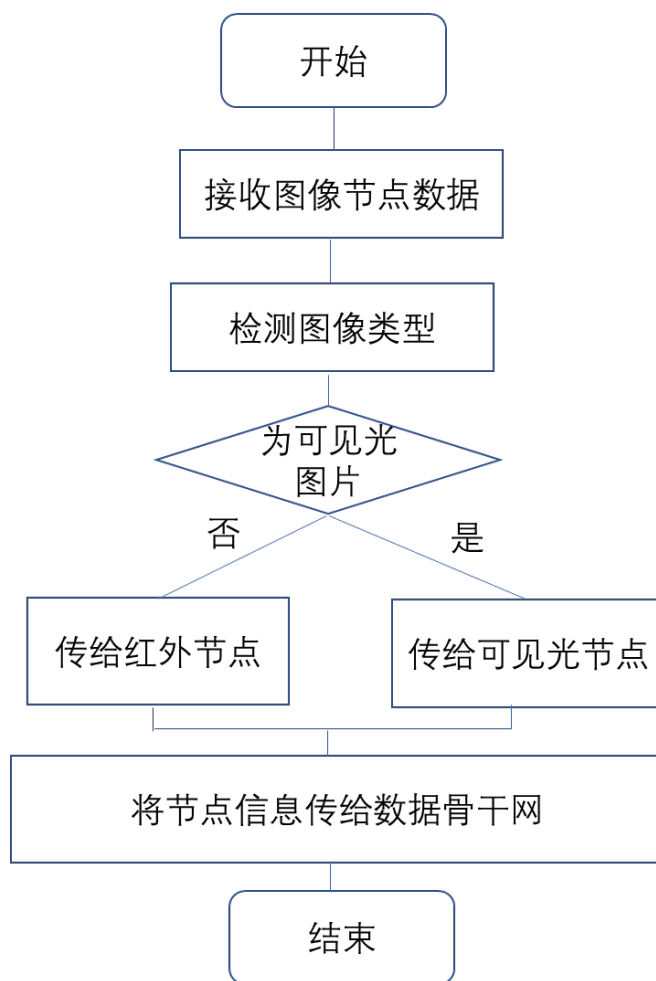


图 13 数据接收流程图

4.2.3 数据网骨干节点模块

数据网骨干节点负责初步处理信息数据，调用摄像头进行视频监控并且将数据发送至信息服务引擎，流程图如图所示：

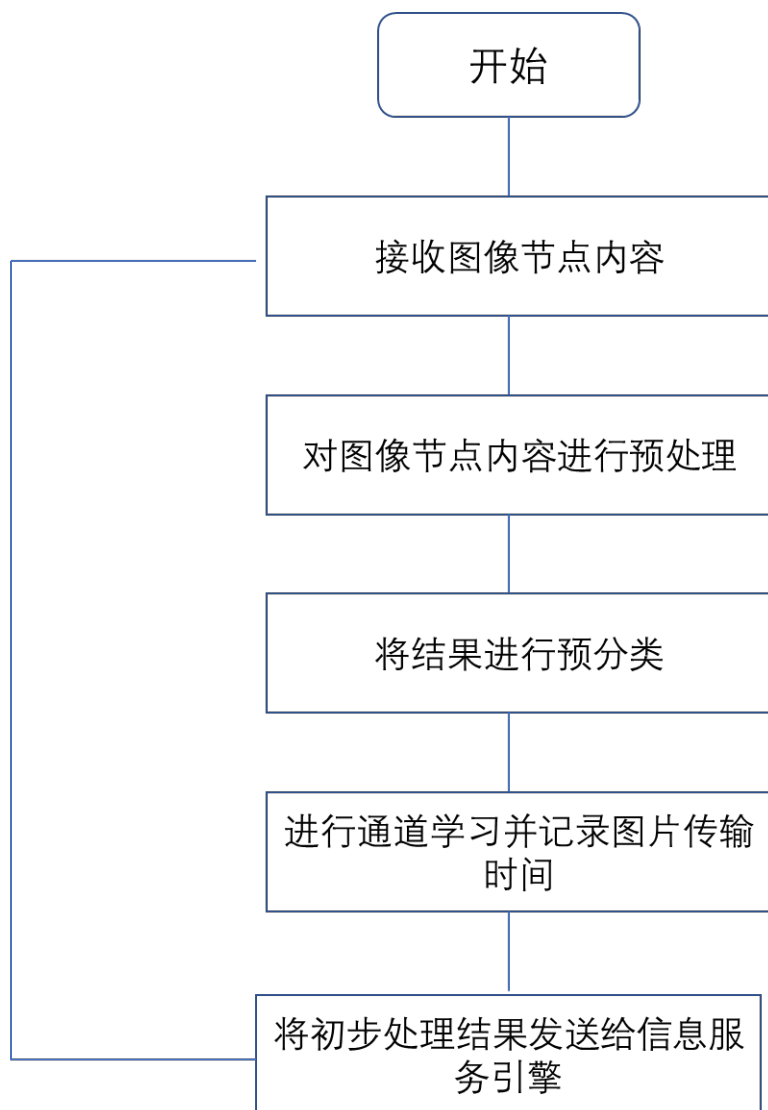


图 14 数据骨干网节点信息处理图

4.2.4 服务器信息处理模块

服务器信息要对可见光与红外光图像进行跨模态学习训练，得出行人重识别结果，流程图如下：

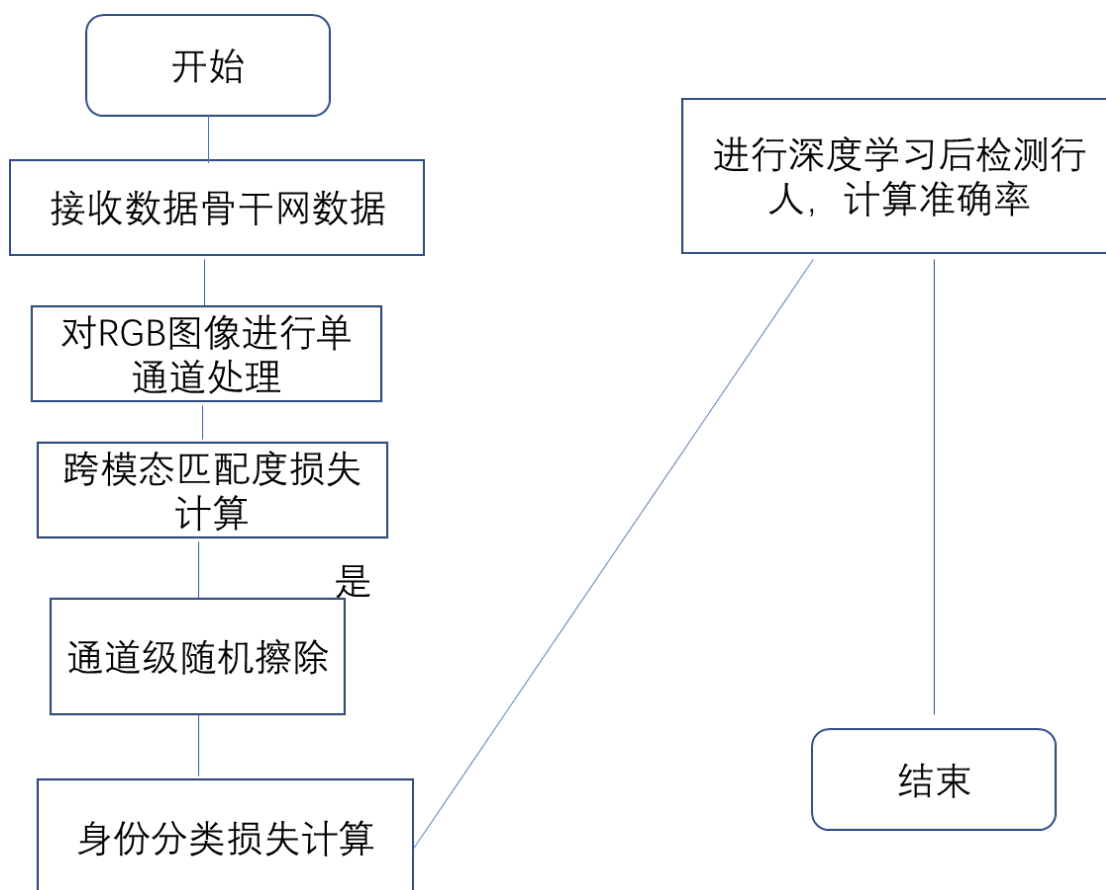


图 15 服务器信息处理流程图

4.3 系统总体运作流程

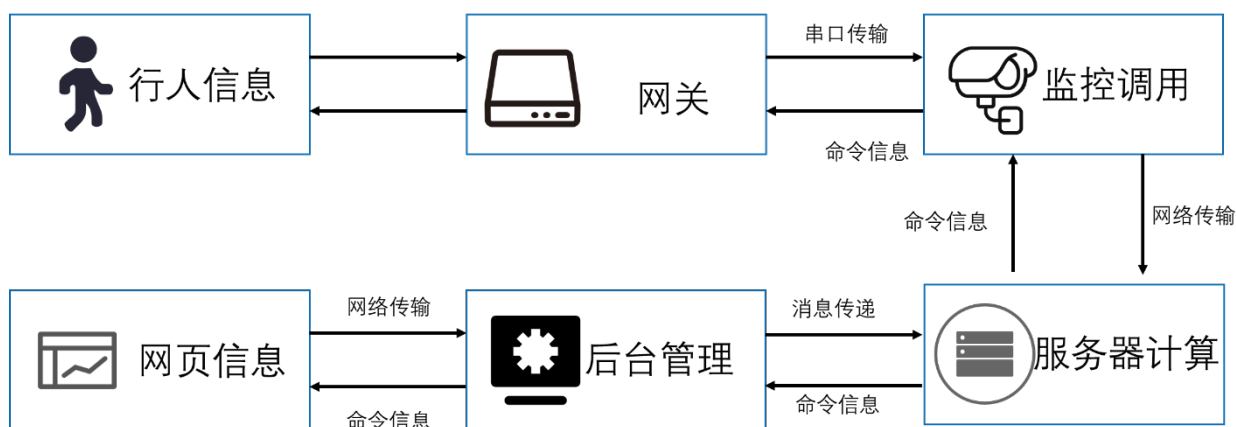


图 16 系统运作流程图

在进行完系统设计之后，我们进行了相应部署，如图所示：

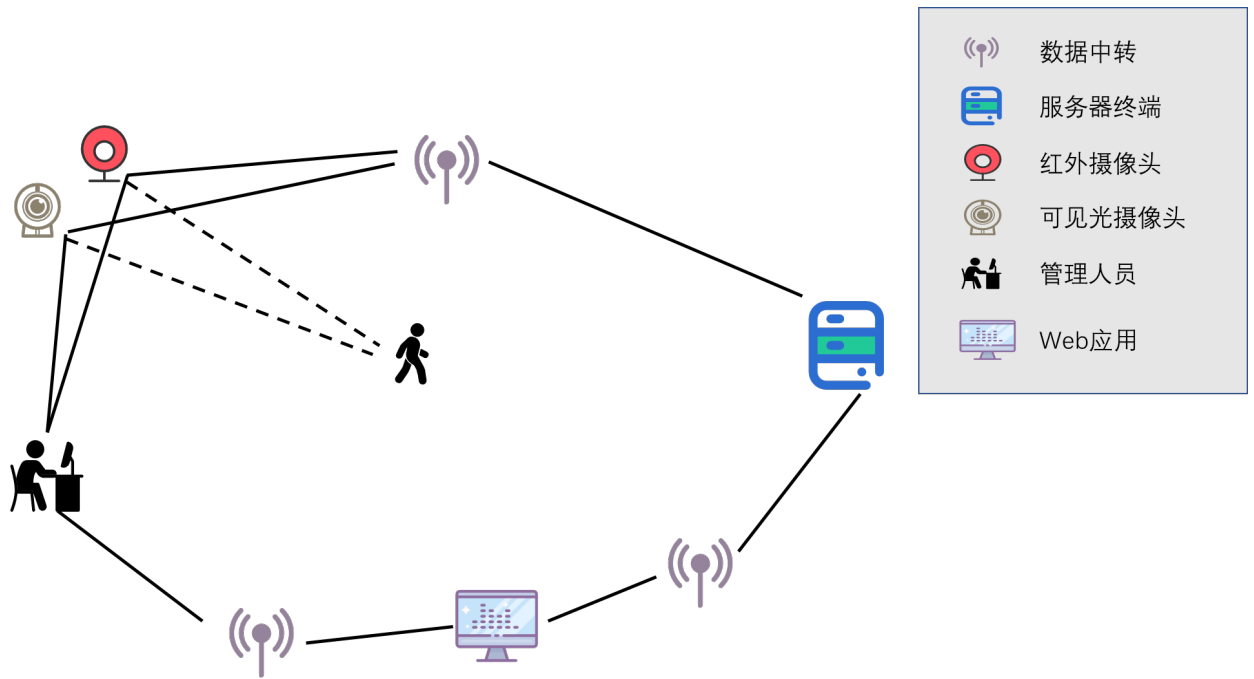


图 17 系统部署图

说明：

- a. 后台管理系统由 Python 语言所写，基于 Django 框架，使用 MySQL 数据库进行数据存储和管理，使用的操作系统为 Ubuntu，推荐内存大小为 4GB，处理器基于火山引擎提供的英伟达 Jetson Orin 开发板所搭载。
- b. 前端页面由 Vue 编写，提供了操作性良好的用户页面。与后端的 API 请求交互采用 Axios，Axios 是一个基于 Promise 的 HTTP 请求库，用于与后端 Django API 进行通信。
- c. 跨模态学习：random 团队完成了跨模态学习算法的设计与改善，在现有技术的基础上，在低质量红外摄像头的前提下提高了可见光-红外图像的跨模态学习识别的准确率。

五、系统实现

内容提要

CA 策略

CRE 策略

ECML 策略

CAJL 策略

5.1 行人图像采集

行人监控系统集成两种图像采集技术，以适应不同光照条件下的监控需求。系统包含可见光成像和红外热成像两个模块，分别在不同的光照环境下发挥作用。在白天或光照充足的环境中，可见光摄像头捕捉清晰的彩色图像；而当夜间或光线较暗时，红外摄像头则接手，利用红外辐射形成图像，确保全天候的监控效果。这些图像数据经由网络实时传输至服务器，服务器会对收到的图像进行处理分析，处理后的数据及原始图像信息封装在 JSON 中进行传输。

5.2 基于 CA（随机通道可交换增强）和 CRE（通道级随机擦除）策略的数据增强技术

由于红外热成像图像固有的单通道特性，直接将可见光图像与红外图像输入到同一学习网络中会遇到兼容性问题。为此，我们将 RGB 图片转为统一的三元组，基于 CA 策略精妙地对齐两种模态的数据，使得它们能够在特征空间中相互映射。此外，为缓解数据稀缺带来的影响，我们实施了 CRE 策略，有效地扩大了数据集规模，

增加了模型训练时的多样性和鲁棒性，从而提升了跨模态行人重识别的准确性与泛化能力。

5.2.1 随机通道可交换增强策略

RGB 三通道的可见光图像蕴含着丰富的颜色信息，对于实现可见光与红外图像的匹配至关重要。然而，从单通道的红外热成像图像中直接恢复成对应的三通道可见光图像是一项艰巨的任务。因为红外图像缺乏颜色信息，导致从红外到可见光的匹配过程本质上是不可逆的。

为应对这一行人重识别领域中的难点，random 团队创造性地引入了随机通道可交换增强（CA）策略。这一策略的核心在于对 RGB 图像的每个颜色通道进行分离和独立处理。通过这一转换，我们得以实施基于色彩的分割和数字图像增强技术，提升了跨模态匹配的准确率，为解决不可逆匹配问题提供了可行的路径。

对于可见光和红外图像的匹配，学习目标（跨模态匹配度损失函数）可以表示为：

$$L = \sum \ell(f^v(x_i^{R,G,B}), f^r(x_j^r), y_i, y_j) \quad (1)$$

其中， f_v 和 f_r 分别用于提取可见光和红外图像特征的网络； $x_i^{R,G,B}$ 和 x_j^r 是来自不同模态的图像； y^i 和 y^j 是对应的标签； ℓ 是损失函数。

这里我们引入了一种通道增强策略，通过随机选择一个颜色通

道（R、G 或 B）来替换其他两个通道，生成一个新的训练图像，这个图像会专注于单个颜色通道：

$$\begin{aligned}\tilde{x}_i^{v,R} &= (x_i^R, x_i^R, x_i^R) \\ \tilde{x}_i^{v,G} &= (x_i^G, x_i^G, x_i^G) \\ \tilde{x}_i^{v,B} &= (x_i^B, x_i^B, x_i^B).\end{aligned}\tag{2}$$

其中 v 代表向量空间，后面的值代表是向量空间中的 R，G，B 的位置。

这样处理的好处是可以隔离和操作单个颜色通道，将每个 RGB 通道转换为一个统一的三元组，基于颜色的分割数字图像增强。它本质上是一种通道增强策略，通过挖掘个别通道(R、 G 或 B)与单通道红外图像。主要思想是运行选择一个通道(R、 G 或 B)代替另一个通道通过一个新的训练图像生成一个新的训练通道，且只在一个通道上对比。 $\tilde{x}_i^{v,R}$ ， $\tilde{x}_i^{v,G}$ ， $\tilde{x}_i^{v,B}$ 分别代表将 RGB 图像中的绿色和蓝色通道替换为红色通道，红色和蓝色通道替换为绿色通道，以及红色和绿色通道替换为蓝色通道后得到的图像。

接下来我们由公式（1）、公式（2）得到替换后的损失函数，公式如下：

$$\mathcal{L} = \sum \ell(f^v(\tilde{x}_i^v), f^r(x_j^r), y_i, y_j)\tag{3}$$

5.2.2 通道级随机擦除策略

随机擦除技术作为一种增强策略，已在诸多视觉识别任务中得到广泛应用，其核心目标在于提升模型在未知数据上的泛化能力。

为了进一步增加训练样本的多样性，我们创新性地融合了通道增强技术，实施了通道级随机擦除（CRE）策略。

具体而言，我们在 RGB 三个通道中分别随机选取擦除区域，这意味着每个通道都有机会经历局部信息的随机抹除。对于被选中的每个像素点，我们会用该通道在整个图像上的平均值来替代其原值。

random 团队采用的 CRE 擦除策略如下：

$$\tilde{x}_i^{v,*}(m,n) = \begin{cases} \alpha^*, & (m,n) \in S_e^* \\ \tilde{x}_i^{v,*}(m,n), & otherwise \end{cases} \quad (4)$$

其中， $\tilde{x}_i^{v,*}$ 表示经过通道随机擦除处理后的像素值， m, n 定义了像素的具体坐标位置，如果某像素恰好落在随机选定的擦除区 S_e^* 内，则其原有值会被替换为平均值，否则不变。

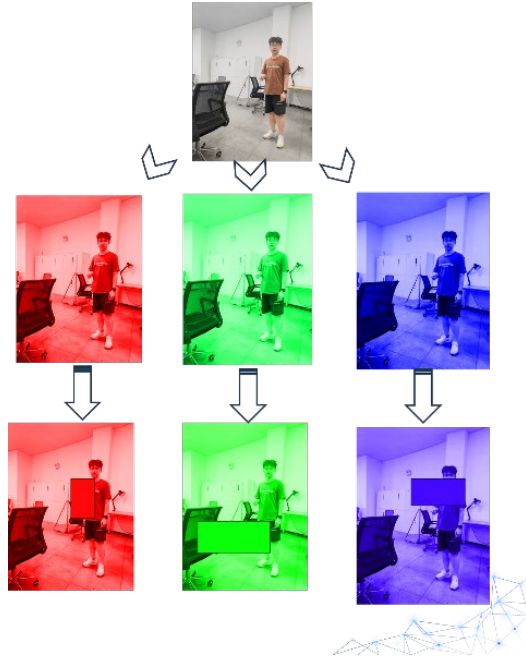


图 18 通道交换增强擦除策略

这一做法促使模型在面对局部信息缺失时，仍然能够依靠剩余特征进行准确的预测，从而显著增强了模型的鲁棒性和对新场景的适应能力。通过 CRE 策略，我们不仅丰富了训练集，还确保了模型在复杂多变的真实环境中能够表现出色。

5.3 基于 ECML（增强型渠道混合学习）和 CAJL（通道增强联合学习）的跨模态度量算法

5.3.1 增强型渠道混合学习

传统的跨模态匹配模型往往依赖于双向三元组变体来指导特征学习，旨在优化不同模态下正负样本对之间的相对距离。然而，这类常规方法未能有效排除那些差异过于明显，早已超出预设阈值的容易三元组（easy triplets），导致模型难以胜任跨模态红外行人重识别的高要求任务。为此，我们引入了一种名为增强型渠道混合学习（Enhanced Channel-Mixed Learning, ECML）的方法，通过引入平方差（Squared Difference）作为度量指标，优化跨模态学习过程，以更精细地处理模态内及模态间的差异。采用平方差度量能够更贴近大边距度量学习的准则，从而增强模型对特征的区分能力，提升其识别精度。

更重要的是，ECML 特别关注那些处于优化边缘的硬三元组（hard triplets），即那些尚未完全符合理想状态的样本对。通过降低原始值极大的容易三元组的权重，模型优化过程得以更集中地针对这些硬三元组，从而显著提升模型的整体性能和泛化能力。这种策

略确保了模型能够更加稳健地应对复杂多变的跨模态识别挑战。

ECML 具体流程如下：

首先我们需要计算身份分类损失

$$\mathcal{L}_{id} = -\frac{1}{N} \sum_{i=1}^N \log(p(y_i | f(x_i); \theta^0)) \quad (5)$$

其中， \mathcal{L}_{id} 表示算出来的结果（身份分类损失）， N 是样本数， y_i 是第 i 个样本的标签， $f(x_i)$ 是从图像 x_i 中提取的特征， θ^0 是共享的身份分类器参数，用于优化模型识别不同身份的表现。

接下来我们要计算加权正则化三元组损失，算法公式如下：

$$\mathcal{L}_{wrt} = \frac{1}{N} \sum_{i=1}^N \log \left(1 + \exp(\sum_{ij} w_{ij}^p d_{ij}^p - \sum_{ik} w_{ik}^n d_{ik}^n) \right) \quad (6)$$

$$w_{ij}^p = \frac{\exp(d_{ij}^p)}{\sum_{d_{ij}^p \in P_i} \exp(d_{ij}^p)}$$

$$w_{ik}^n = \frac{\exp(-d_{ik}^n)}{\sum_{d_{ik}^n \in N_i} \exp(-d_{ik}^n)}$$

其中， \mathcal{L}_{wrt} 是结果，目的是优化正例和负例之间的一个相对距离。而 w_{ij}^p 和 w_{ik}^n 分别是正例和负例对的权重， d_{ij}^p 和 d_{ik}^n 是正例和负例对的距离。

再接下来，为了进一步提升跨模态匹配的准确性、减少因样本偏差带来的影响，我们需要计算增加平方差损失，公式如下：

$$\mathcal{L}_{sq} = \frac{1}{N} \sum_{i=1}^N \log \left(1 + \exp(\varphi[\sum_{ij} w_{ij}^p d_{ij}^p - \sum_{ik} w_{ik}^n d_{ik}^n \mu_i]) \right)$$

在这里，我们定义 μ_i 为正例和负例距离之差，引入改进的平方差函数用来替代 L1 范数（曼哈顿距离），定义为：

$$\varphi[\mu_i] = \begin{cases} \mu_i^2, & \mu_i > 0, \\ -\mu_i^2, & \mu_i < 0. \end{cases} \quad (7)$$

至此，增强型渠道混合学习对常规型跨模态匹配进行了改进，减少了容易三元组的影响、优化了边际硬三元组，使得模型具有更强的判别力。

增强型渠道混合学习算法流程如图所示：

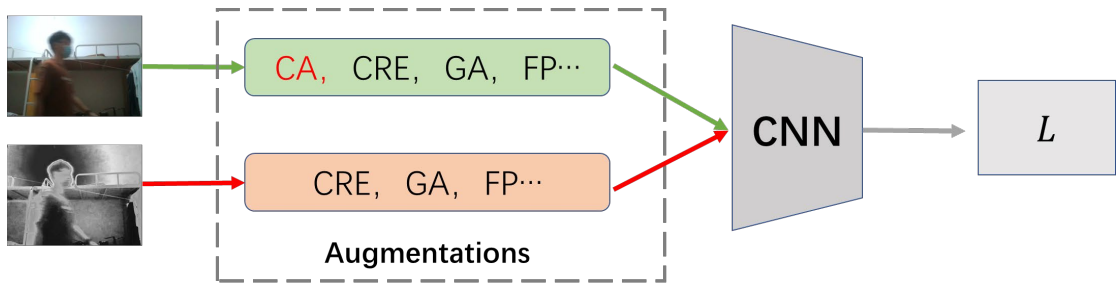


图 19 增强型渠道混合学习

5.3.2 通道增强联合学习

将经过通道增强处理的可见光图像作为一种补充模态，与未经处理的可见光图像和红外图像协同工作，共同构成了一个三模态的融合学习体系。具体而言，我们将经过随机颜色通道交换得到的可见光图像作为额外的信息来源，与原始可见光图像及红外图像并行用于训练，从而构建起三模态的协同学习环境。这种方法不仅极大

地丰富了训练数据的变异性，还促使模型更加专注于提取那些在不同模态间保持一致性的特征。通过将身份损失（ L_{id} ）与强化的平方损失（ L_{sq} ）结合，得到作为优化目标的复合损失函数：

$$L = L_{id} + L_{sq}$$

通道增强联合学习算法流程如图所示：

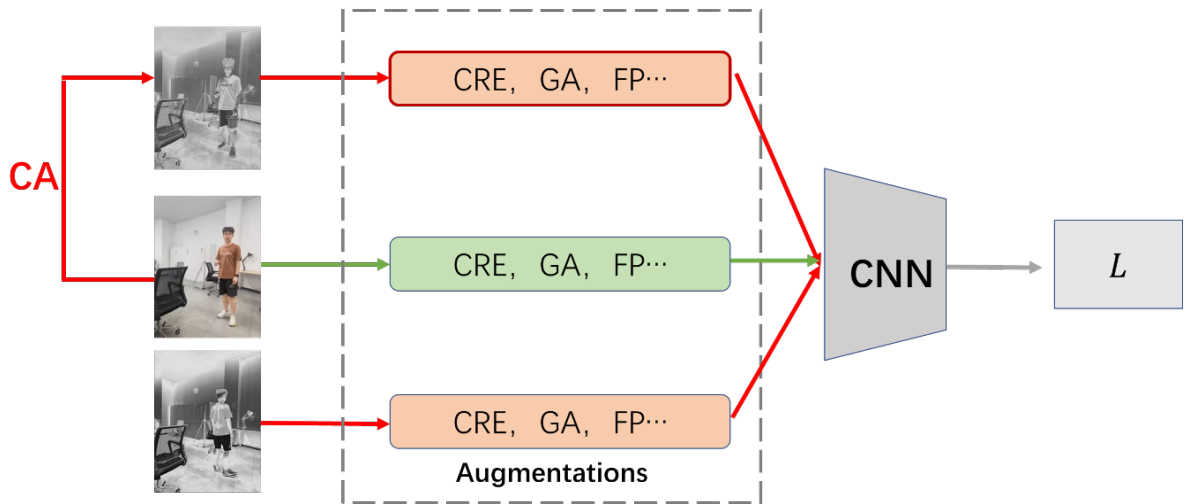


图 20 通道增强联合学习

这一融合学习架构充分利用了通道增强技术生成的图像，而无需对神经网络架构做出调整。该技术允许模型从扩充的数据集中汲取更为广泛和深层次的特征表达，显著增强了模型对色彩变换的适应能力。

六、结果展示

6.1 应用系统展示

6.1.1 登录界面展示

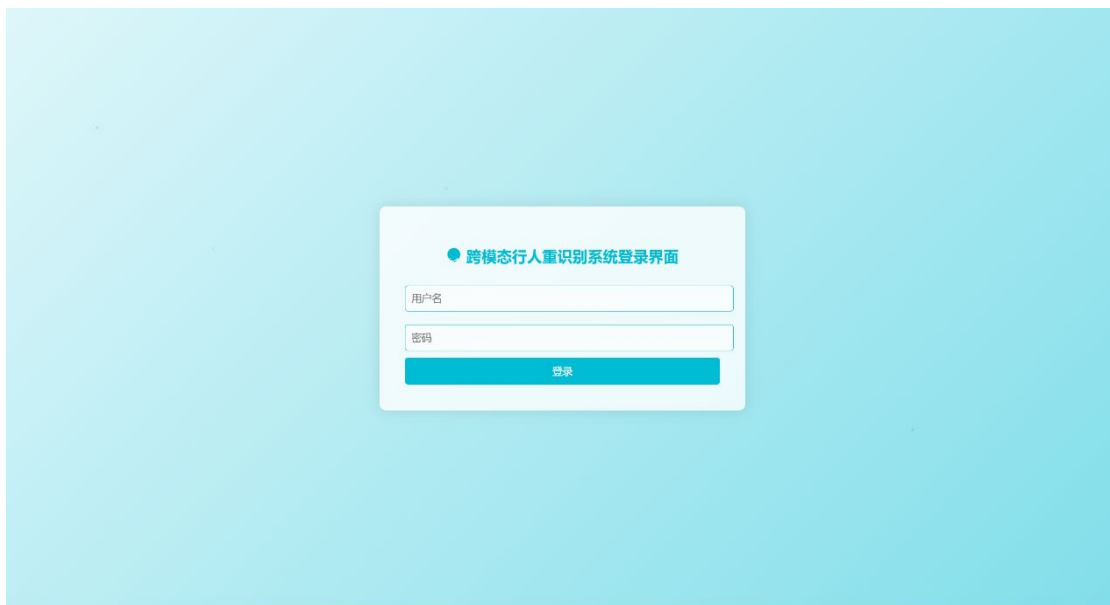


图 21 应用登录界面图

random 团队设计的 Web 应用登录界面如图，输入用户名和密码之后进入到监控界面。登录界面整体色调为蓝色渐变，体现出设计的科技感，冷色调突出冷静，符合监控应用的设计要求。与火山引擎的前端页面设计理念不谋而合。

在输入预设的用户名 `admin` 和密码 `password` 后，我们进入到监控界面进行实际应用。

6.1.2 强光照时界面展示

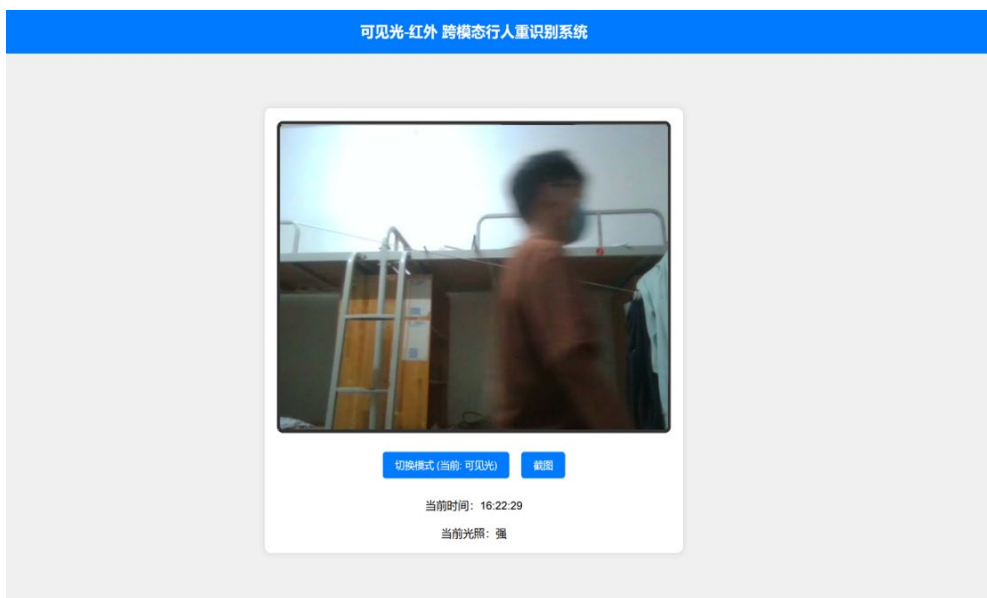


图 22 强光照应用界面图

界面设计显示当前时间，可以准确捕捉到拍摄到目标行人的行为时间轨迹。在光学传感器侦测到当前光照较强时，自动切换到可见光摄像头，管理员也可以进行手动切换。

6.1.3 弱光照时界面展示



图 23 弱光照应用界面图

同上，当检测到当前光照较弱时，摄像头切换为红外摄像头，拍摄行人图像。

6.1.4 比对界面展示



图 24 比对界面图

将需要对比的红外行人图像同拍摄到的可见光图像数据进行行人特征比对，找出该行人对应的标签信息，完成行人重识别功能。

6.2 算法结果分析

算法对比结果表

Settings		Accuracy			
Method	Venue	r=1	r=10	r=20	mAP
Zero-Pad	ICCV-17	16.80	50.12	74.33	19.95
HSME	AAAI-19	18.68	34.74	74.95	27.12
D ² RL	CVPR-19	24.9	72.6	84.4	33.2

续表

Settings		Accuracy			
Method	Venue	r=1	r=10	r=20	mAP
AlignGAN	ICCV-19	38.4	88.4	89.7	41.9
Hi-CMD	CVPR-20	30.9	73.6	93.7	31.9
AGW	ArXiv-20	51.50	80.39	95.14	51.65
HAT	TIFS-22	55.29	93.14	93.36	57.89
Ours	-	63.42	95.11	93.86	61.54

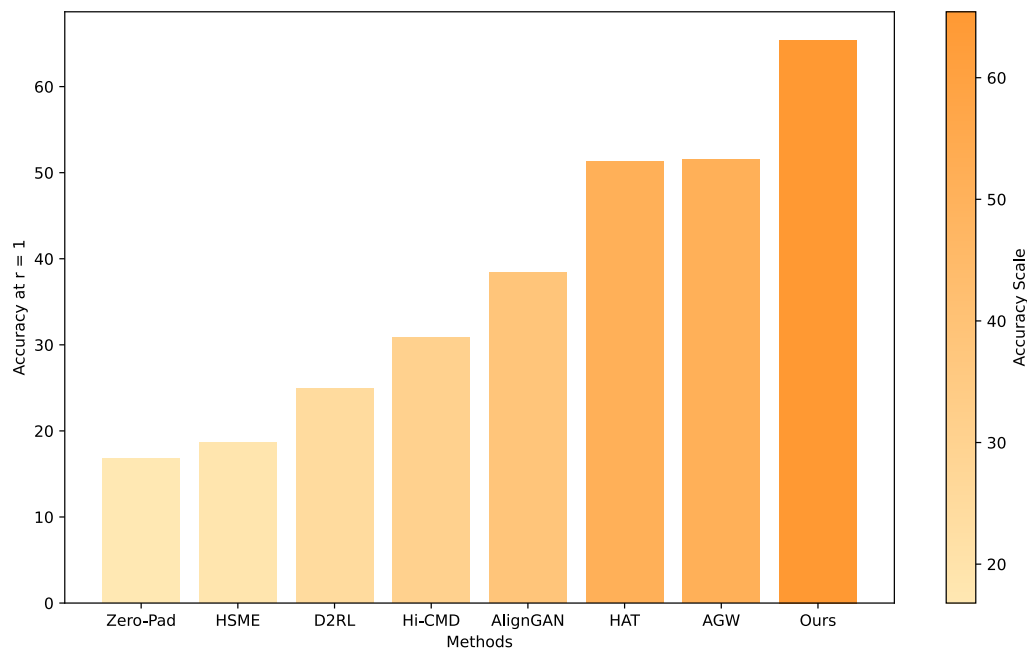


图 25 算法结果对比条形图

我们首先利用一个公共数据集 SYSU-MM01 在可见光-红外行人重识别（VI-ReID）任务上评估我们的模型。使用 rank-k 匹配精度和

平均精度（mAP）作为评估指标。其中，rank-k 匹配精度直接衡量查询图像的真实身份在前 k 个返回结果中的准确性，非常直观，易于理解和解释，做到了快速评估模型在前 k 个结果中的表现，适用于需要快速判断模型性能的场景；平均平均精度（mAP）则综合衡量整个返回结果列表的排序准确性，考虑到所有相关结果的位置，能够全面评价模型在整个检索列表上的性能，而不仅仅是前几个结果，从而更好地反映模型的排序质量。

SYSU-MM01 是最大的 VI-ReID 数据集，它是在 SYSU 校园内在白天和夜间的室内和室外环境中采集的。训练集包含 395 个身份的 22,258 张可见光图像和 11,909 张红外图像。该数据集包含全搜索和室内搜索测试模式，其中前者更具挑战性，我们的应用仍保持高精度识别。

综合上述两个评判指标和 SYSU-MM01 数据集，我们 random 团队的方法相比于其他的跨模态识别算法表现出对于行人特征识别的强大优势。

七、参考文献

- [1] Bai, Song, Xiang Bai, and Qi Tian. "Scalable person re-identification on supervised smoothed manifold." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [2] Chen, Jie, et al. "Learning mappings for face synthesis from near infrared to visual light images." 2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009.
- [3] Cheng, Ziang, et al. "Non-local intrinsic decomposition with near-infrared priors." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
- [4] Cubuk, Ekin D., et al. "Autoaugment: Learning augmentation policies from data." arXiv preprint arXiv:1805.09501 (2018).
- [5] Dai, Pingyang, et al. "Cross-modality person re-identification with generative adversarial training." IJCAI. Vol. 1. No. 3. 2018.
- [6] Duan, Boyan, et al. "Cross-spectral face hallucination via disentangling independent factors." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [7] Hou, Ruibing, et al. "Vrsrc: Occlusion-free video person re-identification." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
- [8] Li, Diangang, et al. "Infrared-visible cross-modal person re-identification with an x modality." Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 04. 2020.
- [9] Li, He, Mang Ye, and Bo Du. "Weperson: Learning a generalized re-identification model from all-weather virtual data." Proceedings of the 29th ACM international conference on multimedia. 2021.
- [10] Lu, Yan, et al. "Cross-modality person re-identification with shared-specific feature transfer." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [11] Peng, Chunlei, et al. "Re-ranking high-dimensional deep local representation for NIR-VIS face recognition." IEEE Transactions on Image Processing 28.9 (2019): 4553-4565.
- [12] Sun, Yifan, et al. "Circle loss: A unified perspective of pair similarity optimization." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [13] Wang, Guan'an, et al. "RGB-infrared cross-modality person re-identification via joint pixel and feature alignment." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
- [14] Wang, Zhixiang, et al. "Learning to reduce dual-level discrepancy for infrared-visible person re-identification." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

致 谢

本课题是在我们 random 团队的指导老师艾浩军副教授的悉心指导下以及全体团队成员的共同努力下完成的。艾浩军副教授严谨的治学态度和科学的教育方法给了我们极大的帮助和影响。在此，我们衷心感谢近一年来艾浩军副教授对我们团队的关心和指导。

感谢火山引擎提供了强大的边缘计算开发板以及火山引擎的技术支持团队，在我们的项目开发过程中给予了宝贵的技术支持和建议。这些支持使我们能够顺利地将理论转化为实践，并有效地解决了项目开发中遇到的各种技术难题。

同时，我们要感谢开源社区的贡献者们，特别是那些为深度学习框架和计算机视觉库做出贡献的人们。这些优秀的开源工具和算法为我们项目的开发提供了坚实的基础，极大地加速了我们算法的研发进程。

最后，我们要感谢全国大学生物联网设计竞赛组委会，感谢你们为我们提供了这样一个宝贵的展示和锻炼机会。

由于 random 团队学术与技术水平有限，所编写的文档与所制作的系统难免有不足之处，恳请组委会各位老师和专家不吝赐教，给予批评指正！