HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

GRADUATION THESIS

Hybrid Edge-Server Person Re-ID Scalable Microservices with Metadata-Enhanced Features

Duong Minh Quan
quandm.210710@sis.hust.edu.vn

Major: Data Science and Artificial Intelligence

Supervisor: Dr. Dang Tuan Linh _____

Signature

Department: Computer Engineering

School: School of Information and Communications Technology

HANOI, 07/2025

ACKNOWLEDGMENT

Firstly, I would like to express my gratitude to Dr. Dang Tuan Linh for all of his assistance, not just in helping me finish this thesis but also during my four years of university study. I have learned so much from him and those values are extremely valuable to me, which have greatly assisted me in both work and study. Additionally, I also want to sincerely thank Mr. Hung and Ms. Quynh, without whom it's likely I would not have been able to successfully complete this thesis. I am also grateful to other friends, seniors, and contributors, whether in small or large ways, tangible or intangible, for their contributions to this completed thesis today. Finally, I would like to thank my family for always supporting me unconditionally.

ABSTRACT

In today's fast-paced world, effective management and analysis are crucial for maintaining security and enhancing business productivity, particularly as the number of enterprises and the size of enterprises rise. According to a recent report from the General Statistics Office of Vietnam, between 2016 and 2019, there was an average annual increase of 9.8% in the number of enterprises, which is higher than the average annual growth rate of 8.1% observed between 2011 and 2015. Moreover, the behavior of employees can be difficult to manage, and employers can increase productivity if they have valuable information from human behavior. Therefore, it is necessary to implement modern technology, specifically AI, into monitoring systems in order to reduce costs and increase productivity.

However, the majority of current AI implementations rely on centralized servers, making scaling difficult. This thesis proposes a novel AI module that can be installed on edge devices, as a means of overcoming this obstacle. Human detection, human tracking, and human feature extraction are the three primary components of the proposed module. All of these components are directly executed on edge devices. This AI module can be utilized to monitor individuals and collect data that can be used to enhance the productivity of businesses.

I aim to achieve efficient human management and analysis by implementing AI models on edge devices that are readily scalable. The algorithms utilized in this thesis have been successfully implemented on Jetson Nano devices with low computational capability while maintaining above 10 FPS for less than seven people in a single frame. A prototype of this module has been put into practical use and examined in room 405, B1 building at Hanoi University of Science and Technology. The proposed module has the potential to revolutionize office human resource management and analysis, thereby enhancing office security and productivity while reducing costs.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AI	Artificial Intelligence
CNN	Convolutional Neural Networks
CPU	Central Processing Unit
EER	Equal Error Rate
FPS	Frames Per Second
GPU	Graphics Processing Unit
ID	Identification
IoT	Internet of Things
IoU	Intersection over Union
mAP	mean Average Precision
NMS	Non-maximum Suppression
Re-ID	Re-identification
ROI	Region Of Interest
USB	Universal Serial Bus

CHAPTER 1. INTRODUCTION

This chapter explains why small businesses need affordable person tracking systems even though they have limited budgets. I will present the research goals, introduce a practical system design that combines edge devices and servers to make advanced customer monitoring cost-effective for small businesses, and describe the main contributions including a lightweight detection model and improved database system designed specifically for small business use.

1.1 Motivation

Small and Medium Enterprises (SMEs) across various industries are facing an extraordinary challenge in today's competitive market. Customer expectations have fundamentally shifted from simple product transactions to demanding rich, personalized experiences. This change is particularly evident in sectors like Food & Beverage (F&B) and retail, where 65% of customers report that positive experiences influence their purchasing decisions more than traditional advertising [1].

SMEs operate under much tighter financial constraints than large corporations. In the F&B sector alone, 45% of businesses report that raw materials account for over 30% of their selling prices, leaving little room for major technology investments [2]. Over 60% of F&B businesses have experienced revenue decreases while facing rising operational costs including rent, labor, and materials [3].

This creates an "innovation deadlock" where SMEs:

- Recognize the critical need for better customer experience solutions.
- Understand that technology could provide competitive advantages.

Modern AI technologies like Person Re-identification (Re-ID) offer powerful solutions for understanding customer behavior, optimizing store layouts, and creating personalized experiences. Re-ID systems can seamlessly track customer movements across different areas of a store, measure how long customers spend in specific sections, and identify popular pathways and bottlenecks. This technology enables businesses to provide tailored assistance, highlight relevant promotions based on customer interests, and optimize staff allocation in real-time.

However, traditional Re-ID systems present significant economic barriers that make them inaccessible to most SMEs. Conventional implementations require expensive GPU-powered edge devices. When scaled across multiple cameras needed for comprehensive coverage, these costs become prohibitive.

The high computational requirements of traditional Re-ID systems also demand powerful central servers for data processing and storage, further inflating the total cost of ownership. For SMEs already struggling with thin profit margins, these substantial upfront investments often exceed their entire annual technology budgets.

Therefore, there is an urgent need for cost-effective, scalable Re-ID solutions specifically designed for SME deployment. Such systems should significantly reduce hardware costs by leveraging efficient CPU-based processing at the edge, minimize complex infrastructure requirements, and provide meaningful customer experience improvements that allow SMEs to compete on service quality rather than just price. By democratizing access to intelligent customer interaction technologies, we can enable businesses of all sizes to enhance customer satisfaction, build loyalty, and drive sustainable growth in an increasingly experience-driven marketplace.

1.2 Objectives and scope of Thesis

1.3 Tentative solutions

1.4 Contributions

This thesis presents two main contributions:

1. An application is deployed on hybrid edge-server devices and uses a microservices architecture, allowing for easy system scaling (increasing the number of cameras). It includes:
    - A custom-trained, lightweight human detection model specifically designed for CPU-based, resource-constrained edge devices.
    - A vector database optimization algorithm for efficient identity retrieval. This uses a person's metadata (gender) to reduce the search space, improving retrieval speed and accuracy.
    - This thesis also provides an interactive web application. It lets users monitor the system, view live camera streams, and search for people using their metadata.

1.5 Organization of Thesis

CHAPTER 2. LITERATURE REVIEW

Chapter 1 has discussed the problems which lead to the motivations of this thesis. It has also presented the objectives, the tentative solutions, and the contributions of this thesis. There are two main parts in this chapter. They are presentations about (i) related works in Section 2.1, and (ii) the foundation theory in Section 2.2. Specifically, in Section 2.2, models, and algorithms contributing to the making of the lightweight module on edge devices will be introduced in detail. For the purpose and scope of the thesis, four main components that make up the proposed human monitoring module on edge devices will be introduced including: (i) object detection with YOLOv5 in Section 2.2.1, (ii) object tracking algorithms including SORT and DeepSORT in Section 2.2.2, (iii) deep metric learning with particular emphasis on the hard triplet loss function in Section **??**, and (iv) MobileNetV2 architecture in Section **??**.

2.1 Related works

2.1.1 Person Re-Identification

2.1.2 Edge Computing in AI

Edge computing has emerged as a transformative paradigm in artificial intelligence, bringing computational capabilities closer to data sources and end users. The global edge AI market size was estimated at USD 20.78 billion in 2024 and is anticipated to grow at a CAGR of 21.7% from 2025 to 2030 [4]. This rapid growth reflects the increasing demand for real-time processing, reduced latency, and enhanced privacy in AI applications. In the context of retail and customer experience, edge computing offers significant advantages. Edge computing allows retail IT teams to optimize cloud costs by being more strategic about the data they send to the cloud, processing only the essential information instead of all the raw data [5]. This selective data processing is particularly relevant for person Re-ID systems, where massive amounts of video data can be filtered and processed locally before sending relevant features to central servers.

Recent research has focused on optimizing AI models for edge deployment. Novel person Re-ID networks integrate pedestrian edge features into the representation and utilize edge information to guide global context feature extraction [6]. This approach demonstrates the feasibility of running sophisticated Re-ID algorithms on resource-constrained edge devices.

The trend toward edge AI democratization is evident in various applications.

Edge Intelligence or Edge AI moves AI computing from the cloud to edge devices, where data is generated, representing a key to AI democratization [7]. This democratization is particularly important for SMEs that cannot afford cloud-based AI solutions with high operational costs.

2.1.3 Microservices and Distributed Systems

Microservices architecture has become increasingly important in developing scalable AI systems. Microservices are generally characterized by a focus on modularity, with each service designed around a specific business capability. These services are loosely coupled, independently deployable, and often developed and scaled separately [8].

In the context of person Re-ID systems, microservices offer several advantages for distributed deployment. Person Re-ID microservices over artificial intelligence Internet of Things edge computing gateways address privacy issues while enabling identification of the same person from multiple different angles across multiple cameras [9].

The distributed nature of modern Re-ID systems presents unique challenges. Current computer vision algorithms on person Re-ID mainly focus on performance, making them unsuitable for distributed systems. For distributed systems, computational complexity becomes a critical consideration [10]. This highlights the need for lightweight, distributed-friendly Re-ID algorithms that can operate efficiently across multiple edge nodes.

Recent work has explored distributed frameworks for deep learning applications. Distributed microservice deep-learning frameworks for object detection in edge computing involve analyzing images and videos to extract information about object classes and their locations [11]. These frameworks provide a foundation for building scalable Re-ID systems that can handle multiple cameras and locations.

Cloud-based distributed approaches have also been investigated. Video-based person Re-ID based on distributed cloud computing uses distributed data storage methods, storing pedestrian datasets and parameters in cloud nodes with data redundancy mechanisms to increase fault tolerance [12]. However, such approaches may not be suitable for cost-sensitive SME deployments due to ongoing cloud costs.

The evolution toward microservices in AI applications reflects broader trends in software architecture. While there are similarities between microservices and distributed systems, microservices are an approach to design where an application is broken into multiple smaller services that can be deployed independently [13].

This independence is crucial for Re-ID systems deployed across multiple retail locations, where individual components may need updates or maintenance without affecting the entire system.

2.2 Foundation theory

2.2.1 Object detection

Object detection technology finds applications across numerous domains including automatic traffic violation systems, identification of unfamiliar persons, digital attendance systems, and autonomous robotic vehicles. The advent of deep learning has dramatically enhanced object detection capabilities. Region-Based Convolutional Neural Networks (R-CNN) [14] represented one of the pioneering breakthroughs in this area, combining CNN architectures [15] with region proposal mechanisms to achieve accurate object localization and classification in images. Subsequent iterations, Fast R-CNN [16] and Faster R-CNN [17], were developed to enhance both processing speed and detection precision compared to the original model. Despite these improvements in detection performance, the multi-step processing pipeline made these approaches impractical for real-time applications.

Modern frameworks such as Detectron2 [18] and EfficientDet [19] have pushed object detection forward considerably. Detectron2 offers flexible deployment of high-performing models but demands extensive setup and typically involves computationally heavy architectures, making it unsuitable for real-time or resource-limited environments. EfficientDet provides a more practical option for devices with constrained resources, though it may struggle to meet strict real-time performance criteria.

You Only Look Once (YOLO) addresses the limitations of multi-stage detection approaches by reformulating object detection as a single regression problem. YOLO processes the entire image in one forward pass, directly predicting bounding boxes and class probabilities from full images. This unified architecture enables real-time performance while maintaining reasonable accuracy for many applications.

The YOLO family has evolved through multiple iterations, with each version improving upon speed-accuracy trade-offs. YOLOv11 [20], in particular, offers several model variants ranging from nano (yolo11n) to extra-large (yolo11x) configurations. The nano variant is specifically designed for resource-constrained environments, featuring significantly reduced parameters and computational requirements while preserving essential detection capabilities.

A significant advancement in YOLOv11 is the integration of thhe C2PSA (Convolutional block with Parallel Spatial Attention) component, which enhances spatial attention capabilities beyond previous YOLO iterations. The C2PSA blockenables the model

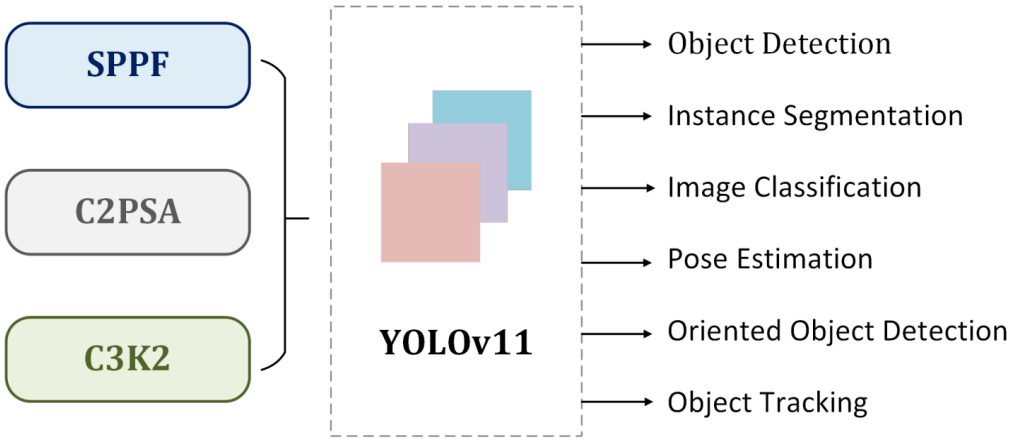

Figure 2.1: Key architectural modules in YOLO11 [20].

to focus more effectively on critical regions within images by inaplementing parallel spatial attention mechanisms. This enhancement is particularly beneficial for detecting objects of varying sizes and positions, addressing common challengees in complex visual environments with partially occluded or small objects. The retention of the Spatial Pyramid Pooling - Fast (SPPF) blockfrom previous versions, combined with the new C2PSA component, creates a compreehensive feature processing pipeline that balances computational efficiency with enhancced spatial awareness.

For edge-based human monitoring applications, YOLOv11n provides an optimal balance between detection performance and computational efficiency. Its lightweight architecture enables deployment on edge devices for real-time person detection, serving as the foundation for subsequent tracking and Re-ID processes in distributed camera networks.

2.2.2 Object tracking

Person Re-ID systems face significant challenges when relying solely on frame-by-frame analysis. Individuals frequently lose their visual identity due to various factors including occlusions from other people or objects, rapid movement causing motion blur, and temporary disappearance from camera coverage areas. While deep learning-based feature extraction models can effectively capture contextual information and compute discriminative identity embeddings, frame-based matching approaches often suffer from identity fragmentation—where the same person receives multiple different identities across consecutive frames.

To address these limitations, tracking mechanisms play a crucial role in maintaining identity consistency over temporal sequences. Unlike existing methods that perform Re-ID independently for each frame, tracking-based approaches maintain continuous identity associations across time. This temporal continuity significantly outperforms

computationally expensive alternatives such as query-driven region proposals [21] and graph-based retrieval methods [22], which become prohibitively costly in large-scale deployment scenarios. By leveraging tracking, our system can efficiently associate multiple detections of the same person, substantially reducing redundant identity searches while improving real-time processing capabilities.

The development of multi-object tracking (MOT) algorithms has evolved through several generations, each addressing specific limitations of previous approaches.

Multi-object tracking has evolved through several approaches, each with distinct trade-offs. SORT [23] provides efficient tracking by integrating object detection with motion prediction, but struggles with complex movement patterns and fast-paced scenarios. To address these limitations, DeepSORT [24] incorporates appearance features via a pre-trained Siamese network, improving performance in dense environments where motion alone is inadequate. However, this enhancement introduces dependency on embedding quality and computational complexity, making it susceptible to visual disturbances. FairMOT [25] advances this paradigm by merging detection and tracking into a unified architecture that simultaneously produces detection outputs and Re-ID features for enhanced multi-object tracking. This unified approach, while effective, demands significant computational resources and requires careful optimization between detection and Re-ID objectives, ultimately compromising processing speed. Alternative solutions include MMTracking [26], which provides a versatile framework supporting multiple advanced algorithms but necessitates substantial parameter optimization.

ByteTrack [27] represents a breakthrough in tracking methodology, delivering exceptional performance without requiring dedicated appearance models, thereby maintaining high processing speeds particularly in crowded environments. This approach achieves an optimal trade-off between real-time processing and tracking reliability. Consequently, our framework employs ByteTrack for maintaining pedestrian identity consistency across video frames, significantly enhancing ID assignment precision. This capability proves essential in dense scenarios where overlapping persons create significant challenges for camera-based identification systems.

2.2.3 Feature extraction

Feature extraction serves as the cornerstone of person Re-ID systems, converting raw image data into discriminative descriptors that enable robust identity matching. Various research efforts have focused on enhancing feature extraction efficiency, particularly for deployment on resource-constrained edge devices.

Several approaches have explored image preprocessing techniques to enhance

feature extraction quality under challenging conditions. Yan et al. [28] proposed a weighted object detection framework that selectively emphasizes salient image regions to generate more discriminative global representations. Building upon this work, Yan et al. [29] developed a hybrid neural network architecture incorporating specialized preprocessing modules, demonstrating robust performance under varying illumination conditions and pose variations. Further extending these preprocessing strategies, Yan et al. [30] investigated Generative Adversarial Network (GAN)-based image enhancement techniques as preprocessing steps, effectively addressing common surveillance challenges such as poor lighting conditions and low-resolution imagery that typically compromise Re-ID performance.

While these preprocessing approaches demonstrate effectiveness in improving feature quality, they present significant limitations for edge-based deployment scenarios. The additional computational overhead and memory requirements introduced by these techniques exceed the processing capabilities of typical edge hardware. GAN-based enhancement methods, in particular, demand substantial computational resources that are incompatible with edge device constraints. Furthermore, the increased processing latency introduced by these preprocessing stages renders them unsuitable for real-time applications requiring immediate response.

Contemporary research has also explored Transformer-based architectures for enhanced feature extraction, particularly Vision Transformers (ViT) [31], which leverage self-attention mechanisms to capture long-range spatial dependencies within images. Despite their demonstrated effectiveness, these models typically require substantial computational resources, making them challenging to deploy on edge devices without extensive optimization. Multi-scale feature extraction approaches, exemplified by Feature Pyramid Networks (FPN) [32], have shown promise in capturing both low-level and high-level semantic information. This approach proves particularly beneficial for Re-ID scenarios involving significant variations in object scale and camera distance. However, when implemented with large backbone networks, these methods remain computationally prohibitive for edge deployment.

To address these computational constraints, our research emphasizes lightweight feature extraction methodologies. Mobile-optimized architectures such as MobileNetV2 [33] and MobileNetV3 [34] have gained prominence in edge-based applications due to their depth-wise separable convolutions that significantly reduce computational complexity while maintaining competitive accuracy. Squeeze-and-Excitation Networks (SE-Net) [35] further enhance CNN architectures by implementing adaptive channel-wise feature recalibration, enabling models to focus on the most informative image regions without substantially increasing model complexity.

However, these generic lightweight models lack specific optimization for person Re-ID tasks. This limitation has motivated the development of specialized lightweight architectures tailored specifically for Re-ID applications.

OSNet (Omni-Scale Network) [36] represents a significant advancement in lightweight Re-ID architectures. OSNet uses only 2.2 million parameters, substantially fewer than traditional ResNet-50-based approaches which employ 24 million parameters, making it particularly suitable for edge deployment. OSNet introduces innovative omni-scale feature learning that captures multi-scale information through a novel building block design, enabling effective feature extraction across different spatial scales without the computational overhead of traditional multi-scale approaches. The architecture employs depth-wise separable convolutions and efficient channel shuffling operations to maintain computational efficiency while preserving discriminative capability.

LightMBN (Lightweight Multi-Branch Network) [37] further advances edge-optimized Re-ID by introducing a multi-branch architecture specifically designed for resource-constrained environments. LightMBN incorporates efficient attention mechanisms and feature fusion strategies that maximize discriminative power while minimizing computational requirements. The network architecture employs channel-wise and spatial attention modules that adaptively emphasize important features without introducing significant computational overhead.

Despite these advances, current lightweight models still face challenges in balancing computational efficiency with discriminative capability, particularly in scenarios involving significant pose variations, occlusions, and lighting changes. Our research addresses these limitations by investigating custom-trained lightweight architectures that are specifically optimized for the unique requirements of SME deployment scenarios, emphasizing both computational efficiency and robust performance across diverse operational conditions.

2.2.4 Message queue

2.2.5 Containerization

2.2.6 Vector database

CHAPTER 3. METHODOLOGY

Building upon the theoretical foundations established in Chapter 2, this chapter presents a comprehensive examination of (i) the person Re-ID module architecture and its integration within the hybrid edge-server management system, (ii) detailed specifications of the proposed lightweight Re-ID module optimized for SME deployment, (iii) edge device hardware implementation strategies, (iv) the deployment of microservices frameworks and (v) how the system utilizes person metadata (gender) to enhance the efficiency of identity retrieval processes.

3.1 Overview

3.2 The proposed AI module

3.2.1 Human detection

a, Pre-processing

b, Detecting with YOLOv5n

c, Non-maximum suppression

3.2.2 Human feature extraction

a, Pre-processing

CHAPTER 4. EXPERIMENTAL RESULTS

CHAPTER 5. CONCLUSIONS AND FUTURE WORKS

REFERENCE

[1] S. PRO, *Hành trình nâng tầm trải nghiệm khách hàng ngành f&b*, 2024. **url**: https://soipro.vn/hanh-trinh-nang-tam-trai-nghiem-khach-hang-nganh-fb/.

[2] *Rising costs threaten vietnam's f&b profitability*. **urlseen** 2025. **url**: https://www.vietdata.vn/post/the-f-b-industry-is-seeing-its-profits-disappear-due-to-rising-costs, TheLEADER.

[3] V. News, *Nearly 60 per cent of food and beverage companies reported decline in revenue in 2024*, 2024. **url**: https://vietnamnews.vn/economy/1694177/nearly-60-per-cent-of-food-and-beverage-companies-reported-decline-in-revenue-in-2024.html.

[4] G. V. Research, *Edge ai market size, share & growth | industry report, 2030*, 2024. **url**: https://www.grandviewresearch.com/industry-analysis/edge-ai-market-report.

[5] "2024 tech trends: How to reduce friction in the retail experience," *BizTech Magazine*, 2024. **url**: https://biztechmagazine.com/article/2024/03/2024-tech-trends-how-reduce-friction-retail-experience.

[6] "Person re-identification network based on edge-enhanced feature extraction and inter-part relationship modeling," *Applied Sciences*, **journvol** 14, **number** 18, **page** 8244, 2024. **url**: https://www.mdpi.com/2076-3417/14/18/8244.

[7] Viso.ai, *Edge intelligence: Edge computing and ml (2025 guide)*, 2024. **url**: https://viso.ai/edge-ai/edge-intelligence-deep-learning-with-edge-computing/.

[8] Wikipedia, *Microservices*, 2024. **url**: https://en.wikipedia.org/wiki/Microservices.

[9] "Person re-identification microservice over artificial intelligence internet of things edge computing gateway," *Electronics*, **journvol** 10, **number** 18, **page** 2264, 2021. **url**: https://www.mdpi.com/2079-9292/10/18/2264.

[10] "Distributed implementation for person re-identification," **in***IEEE Conference Publication* IEEE, 2015. **url**: https://ieeexplore.ieee.org/document/7288501/.

[11] "Mded-framework: A distributed microservice deep-learning framework for object detection in edge computing," *Sensors*, **journvol** 23, **number** 10, **page** 4712, 2023. **url**: https://www.mdpi.com/1424-8220/23/10/4712.

[12] "Video-based person re-identification based on distributed cloud computing," *Journal of Artificial Intelligence and Technology*, 2022. **url**: https://ojs.istp-press.com/jait/article/view/13.

[13] Splunk, *What are distributed systems?* 2024. **url**: https://www.splunk.com/en_us/blog/learn/distributed-systems.html.

[14] R. Girshick, J. Donahue, T. Darrell **and** J. Malik, *Rich feature hierarchies for accurate object detection and semantic segmentation*, 2014. arXiv: 1311.2524 [cs.CV]. **url**: https://arxiv.org/abs/1311.2524.

[15] K. O'Shea **and** R. Nash, *An introduction to convolutional neural networks*, 2015. arXiv: 1511.08458 [cs.NE]. **url**: https://arxiv.org/abs/1511.08458.

[16] R. Girshick, *Fast r-cnn*, 2015. arXiv: 1504.08083 [cs.CV]. **url**: https://arxiv.org/abs/1504.08083.

[17] S. Ren, K. He, R. Girshick **and** J. Sun, *Faster r-cnn: Towards real-time object detection with region proposal networks*, 2016. arXiv: 1506.01497 [cs.CV]. **url**: https://arxiv.org/abs/1506.01497.

[18] G. Merz **andothers**, "Detection, instance segmentation, and classification for astronomical surveys with deep learning implementation and demonstration with hyper suprime-cam data," *Monthly Notices of the Royal Astronomical Society*, **jourvol** 526, **number** 1, 1122–1137, **september** 2023, ISSN: 1365-2966. DOI: 10.1093/mnras/stad2785. **url**: http://dx.doi.org/10.1093/mnras/stad2785.

[19] M. Tan, R. Pang **and** Q. V. Le, *Efficientdet: Scalable and efficient object detection*, 2020. arXiv: 1911.09070 [cs.CV]. **url**: https://arxiv.org/abs/1911.09070.

[20] R. Khanam **and** M. Hussain, *Yolov11: An overview of the key architectural enhancements*, 2024. arXiv: 2410.17725 [cs.CV]. **url**: https://arxiv.org/abs/2410.17725.

[21] B. Munjal, S. Amin, F. Tombari **and** F. Galasso, *Query-guided end-to-end person search*, 2019. arXiv: 1905.01203 [cs.CV]. **url**: https://arxiv.org/abs/1905.01203.

[22] Z. Zhu, T. Huang, K. Wang, J. Ye, X. Chen **and** S. Luo, *Graph-based approaches and functionalities in retrieval-augmented generation: A comprehensive survey*, 2025. arXiv: 2504.10499 [cs.IR]. **url**: https://arxiv.org/abs/2504.10499.

[23] A. Bewley, Z. Ge, L. Ott, F. Ramos **and** B. Upcroft, "Simple online and realtime tracking," **in***2016 IEEE International Conference on Image Processing*

(ICIP) IEEE, **september** 2016. DOI: 10.1109/icip.2016.7533003. **url**: http://dx.doi.org/10.1109/ICIP.2016.7533003.

[24] N. Wojke, A. Bewley **and** D. Paulus, *Simple online and realtime tracking with a deep association metric*, 2017. arXiv: 1703.07402 [cs.CV]. **url**: https://arxiv.org/abs/1703.07402.

[25] Y. Zhang, C. Wang, X. Wang, W. Zeng **and** W. Liu, "Fairmot: On the fairness of detection and re-identification in multiple object tracking," *International Journal of Computer Vision*, **jourvol** 129, **number** 11, 3069–3087, **september** 2021, ISSN: 1573-1405. DOI: 10.1007/s11263-021-01513-4. **url**: http://dx.doi.org/10.1007/s11263-021-01513-4.

[26] C. Lin, C. Yu, X. Xu **and** R. Wang, *Mmtracking: Trajectory tracking for uplink mmwave devices with multi-path doppler difference of arrival*, 2025. arXiv: 2503.16909 [eess.SP]. **url**: https://arxiv.org/abs/2503.16909.

[27] Y. Zhang **andothers**, *Bytetrack: Multi-object tracking by associating every detection box*, 2022. arXiv: 2110.06864 [cs.CV]. **url**: https://arxiv.org/abs/2110.06864.

[28] L. Yan, K. Li, R. Gao, C. Wang **and** N. Xiong, "An intelligent weighted object detector for feature extraction to enrich global image information," *Applied Sciences*, **jourvol** 12, **number** 15, 2022, ISSN: 2076-3417. DOI: 10.3390/app12157825. **url**: https://www.mdpi.com/2076-3417/12/15/7825.

[29] L. Yan, M. Sheng, C. Wang, R. Gao **and** H. Yu, "Hybrid neural networks based facial expression recognition for smart city," **jourvol** 81, **number** 1, 319–342, **january** 2022, ISSN: 1380-7501. DOI: 10.1007/s11042-021-11530-7. **url**: https://doi.org/10.1007/s11042-021-11530-7.

[30] L. Yan, F. Jiarun, C. Wang, Z. ye, H. Chen **and** H. Ling, "Enhanced network optimized generative adversarial network for image enhancement," *Multimedia Tools and Applications*, **jourvol** 80, **april** 2021. DOI: 10.1007/s11042-020-10310-z.

[31] A. Dosovitskiy **andothers**, *An image is worth 16x16 words: Transformers for image recognition at scale*, 2021. arXiv: 2010.11929 [cs.CV]. **url**: https://arxiv.org/abs/2010.11929.

[32] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan **and** S. Belongie, *Feature pyramid networks for object detection*, 2017. arXiv: 1612.03144 [cs.CV]. **url**: https://arxiv.org/abs/1612.03144.

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov **and** L.-C. Chen, *Mobilenetv2: Inverted residuals and linear bottlenecks*, 2019. arXiv: 1801.04381 [cs.CV]. **url**: https://arxiv.org/abs/1801.04381.

[34] A. Howard **andothers**, *Searching for mobilenetv3*, 2019. arXiv: 1905.02244 [cs.CV]. **url**: https://arxiv.org/abs/1905.02244.

[35] J. Hu, L. Shen, S. Albanie, G. Sun **and** E. Wu, *Squeeze-and-excitation networks*, 2019. arXiv: 1709.01507 [cs.CV]. **url**: https://arxiv.org/abs/1709.01507.

[36] K. Zhou, Y. Yang, A. Cavallaro **and** T. Xiang, *Omni-scale feature learning for person re-identification*, 2019. arXiv: 1905.00953 [cs.CV]. **url**: https://arxiv.org/abs/1905.00953.

[37] F. Herzog, X. Ji, T. Teepe, S. Hormann, J. Gilg **and** G. Rigoll, "Lightweight multi-branch network for person re-identification," **in***2021 IEEE International Conference on Image Processing (ICIP)* IEEE, **september** 2021, 1129–1133. DOI: 10.1109/icip42928.2021.9506733. **url**: http://dx.doi.org/10.1109/ICIP42928.2021.9506733.