DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Robotics, Cognition, Intelligence

Private 5G Networks for Real-Time Object Detection in Autonomous Systems

Barış Sözüdoğru

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Robotics, Cognition, Intelligence

Private 5G Networks for Real-Time Object Detection in Autonomous Systems

Private 5G-Netzwerke für die Echtzeit-Objekterkennung in autonomen Systemen

Author:	Barış Sözüdoğru
Supervisor:	Prof. Dr.-Ing. Johannes Fottner
Advisor:	Leonhard Feiner, M.Sc.
Submission Date:	July 15, 2024

I confirm that this master's thesis in robotics, cognition, intelligence is my own work and I have documented all sources and material used.

Munich, July 15, 2024 Barış Sözüdoğru

Abstract

This study investigates the integration and performance of private 5G networks for real-time object detection in autonomous systems, focusing on their application in industrial automation and intralogistics. The study begins with a comprehensive review of Wi-Fi and 5G technologies, emphasizing core advancements and deployment in industrial settings. The experimental phase benchmarks 5G and Wi-Fi performance in a controlled environment, measuring latency, jitter, throughput, and object detection accuracy. A practical implementation involves a small forklift equipped with real-time object detection capabilities.

Results reveal that private 5G shows a higher peak throughput, with an average of 1.582 Mbps and a maximum of 23.458 Mbps, compared to Wi-Fi's average of 1.574 Mbps and a maximum of 12.913 Mbps. Throughput has a significant positive effect on Intersection over Union, indicating that higher throughput correlates with better object detection accuracy, demonstrating the benefits of private 5G for real-time object detection. In terms of latency and jitter, Wi-Fi displays lower average jitter and latency compared to private 5G, with mean values of 8.971 milliseconds and 3.597 milliseconds respectively. However, Wi-Fi's peak values for these metrics are significantly higher than those of private 5G, indicating greater instability. In contrast, private 5G demonstrates higher but more consistent averages, with jitter at 15.696 milliseconds and latency at 13.46 milliseconds. This consistency in private 5G may contribute to more stable performance despite the higher average values. Despite similar Intersection over Union values between the two networks, the higher throughput of private 5G may offer a potential advantage for enhancing object detection performance. These findings suggest that private 5G networks have the potential to improve real-time data processing in industrial operations.

Keywords: Private 5G, Wi-Fi, Object Detection, Industrial Applications, Latency, Jitter, Throughput

Table of Contents

List of Abbreviations

5G - 5th Generation
5GC - 5G Core Network
AGV - Autonomous Guided Vehicle
AR - Augmented Reality
AUC - Area Under the Curve
AP - Average Precision
AR - Augmented Reality
ANOVA - Analysis of Variance
API - Application Programming Interface
CDN - Content Delivery Network
CNN - Convolutional Neural Network
COCO - Common Objects in Context
dBm - Decibel Milliwatts
EDGE - Enhanced Data Rates for GSM Evolution
eMBB - Enhanced Mobile Broadband
EPC - Evolved Packet Core
FDMA - Frequency Division Multiple Access
FPR - False Positive Rate
FMS - Fixed Mobile Substitution
FPS - Frames Per Second
GPIO - General-Purpose Input/Output
GSM - Global System for Mobile Communications
HSD - Honestly Significant Difference
HLS - HTTP Live Streaming
HTTP - Hypertext Transfer Protocol
ICE - Interactive Connectivity Establishment
IEEE - Institute of Electrical and Electronics Engineers
IOU - Intersection over Union
IoT - Internet of Things
IP - Internet Protocol
ITU - International Telecommunication Union
LTE - Long Term Evolution
MAC - Media Access Control
MEC - Multi-Access Edge Computing
MIMO - Multiple Input Multiple Output
mMTC - Massive Machine-Type Communication
MU-MIMO - Multi-User Multiple Input Multiple Output
NSA - Non-Standalone
NR - New Radio
OFDMA - Orthogonal Frequency-Division Multiple Access

OME - Oven Media Engine
PHY - Physical Layer
QoS - Quality of Service
RAN - Radio Access Network
RFID - Radio-Frequency Identification
ROC - Receiver Operating Characteristic
RSRP - Reference Signal Received Power
RSSI - Received Signal Strength Indicator
RTMP - Real-Time Messaging Protocol
RTP - Real-Time Transport Protocol
RTSP - Real-Time Streaming Protocol
RTT - Round-Trip Time
SA - Standalone
SDO - Standards Development Organizations
SMS - Short Message Service
SSD - Single Shot MultiBox Detector
TDMA - Time Division Multiple Access
TPR - True Positive Rate
TSN - Time-Sensitive Networking
TUM - Technical University of Munich
UFMC - Universal Filtered Multi Carrier
UMTS - Universal Mobile Telecommunications System
URLLC - Ultra-Reliable Low-Latency Communication
VR - Virtual Reality
W-CDMA - Wideband Code Division Multiple Access
WebRTC - Web Real-Time Communication
Wi-Fi - Wireless Fidelity
WiMAX - Worldwide Interoperability for Microwave Access
XR - Extended Reality
YOLO - You Only Look Once
YT-BB - YouTube-Bounding Boxes Dataset
3GPP - The 3rd Generation Partnership Project

1 Introduction

The advancement of wireless communication technologies has introduced significant innovations across various industries. Among these, the introduction of 5th generation (5G) technology represents a major leap forward, promising substantial enhancements in data transfer speed, reliability, and connectivity. This evolution is expected to affect numerous fields by enabling more efficient and effective real-time data processing, automation, and communication. Despite the widespread anticipation surrounding 5G, its practical benefits and limitations, particularly in comparison to established technologies like Wi-Fi, require thorough investigation.

One critical application area where 5G holds promise is in the domain of autonomous systems, particularly those used in logistics and industrial automation. Automated Guided Vehicles (AGVs), which rely on real-time data processing for navigation and object detection, stand to benefit immensely from the high throughput and low latency capabilities of 5G networks. However, the practical performance gains achievable through 5G, especially in comparison to traditional Wi-Fi, remain under-explored.

This thesis aims to address this gap by conducting a comprehensive evaluation of 5G private networks versus traditional Wi-Fi in facilitating real-time object detection tasks within autonomous systems. The study is conducted within the 5G-enabled campus network at the Technical University of Munich (TUM) and O2 Tower, providing a controlled environment to assess the performance of these technologies under comparable conditions.

Central to this research is the performance comparison of object detection systems in both 5G and Wi-Fi environments. Key performance indicators such as throughput, latency, jitter, and object detection accuracy are measured to determine the efficacy of each network technology. These metrics are crucial for ensuring the seamless operation of AGVs in dense and dynamic settings, where rapid and accurate object detection is important.

One of the significant challenges in evaluating object detection performance in live streaming scenarios is the absence of a definitive ground truth. To address this, a framework was developed to assess the performance of object detection systems under varying network parameters. This framework is essential for understanding how different network conditions impact the accuracy and reliability of object detection, thus providing a more nuanced evaluation of 5G and Wi-Fi capabilities.

In addition to theoretical and empirical analyses, a practical use case was developed as part of this research. This use case involves deploying a small forklift equipped with object detection capabilities in a 5G environment. By comparing performance metrics across these settings, the study aims to provide tangible insights into the real-world applicability and benefits of 5G technology.

2 Related Work

Private 5G networks represent a transformative advancement in wireless communications, offering dedicated on-site network infrastructure with enhanced control over data security, reliability, and customization [1]. These networks provide significant benefits for industrial applications by allowing businesses to tailor connectivity to specific requirements. For instance, a study explains that private 5G networks support multiple industrial use cases under a unified network umbrella, ensuring efficient management of diverse traffic classes, including critical and non-critical services, through Ultra-Reliable Low-Latency Communication (URLLC) capabilities [2]. This capability to maintain high-quality service for each type of traffic is crucial for industrial automation.

Several studies have explored the effectiveness of public and private 5G networks across various application areas through specific use cases. A study demonstrates how retrofitting collaborative robots with 5G network capabilities using joysticks and virtual reality (VR) can enhance performance and efficiency for small and medium-sized enterprises, highlighting the benefits of low latency in private 5G networks [3]. Another critical application area where latency and network performance are of importance is telesurgery. For example, a study demonstrates that 5G technology enables precise, low-latency remote control of surgical robots, successfully performing transoral laser microsurgeries on a cadaver [4]. Additionally, research on 5G-connected drones underscores the importance of reliable connectivity for real-time high-resolution video transfer in public road safety and traffic analysis, demonstrating the benefits of public 5G over traditional networks in such applications [5].

To highlight private 5G use cases in manufacturing, notable examples include Ericsson's 5G smart factory, which demonstrated a 120% improvement in output per employee and a 65% reduction in manual material handling compared to a traditional factory, utilizing over 200 operational robots [6]. Additionally, Mitsubishi developed a smart warehouse using a private 5G network [7]. The US Department of Defense also established a smart warehouse at the Marine Corps Logistics Base [8]. Furthermore, the application of private 5G extends to ports, exemplified by Nokia's collaboration with the Hamburg Port Authority and Deutsche Telekom in a successful 5G field trial at the Port of Hamburg in Germany [9]. These examples underscore the significant impact of private 5G networks on enhancing the efficiency and productivity of warehouses and indoor manufacturing facilities.

However, a study shows that private 5G meets the delay and reliability needs of mobile robots, but performance degrades significantly with video streaming, high cross-traffic, and more devices [10]. Moreover, another study examines the efficiency of low-latency video streaming in a standalone 5G environment, revealing that 5G supports low-latency streaming, but uplink performance is adversely affected by network congestion [11]. In addition, a study evaluates 5G live video streaming performance for low-latency use cases, identifying uplink bottlenecks with delays averaging 12.5 ms, ensuring reliable transmission for applications such as vehicle communication and healthcare with end-to-end latency under 200 ms [12]. On the other hand, a paper investigates the impact of challenging environments on VR and Augmented Reality (AR) video streaming, highlighting that 5G significantly outperforms WiFi in terms of bitrate, packet loss, and overall video quality [13]. Another study evaluates the performance of Wi-Fi 6 and 5G networks in an industrial Internet of Things (IoT) setting, demonstrating that while Wi-Fi 6 is effective under low load conditions, 5G, especially with licensed spectrum, outperforms Wi-Fi 6 [14].

Network parameters such as latency, throughput, and jitter are crucial for ensuring an optimal user experience during live streaming, requiring seamless interaction between applications and the network infrastructure [15]. Especially for VR video live streaming, the requirements for lower latency and higher throughput are critical. To ensure an immersive user experience, 360° videos require high resolutions (over 3840 × 2160), high frame rates (over 40 FPS), and high bitrates (over 10 Mbps), resulting in significantly larger data sizes than fixed-viewpoint videos [16]. On the other hand, studies have shown that packet loss significantly degrades multimedia streaming more than latency, with a 0.75% loss rate severely impacting gaming user experience and a 1% loss at 200ms, making the game nearly unplayable [17].

The implementation of Web Real-Time Communication (WebRTC) for peer-to-peer live video streaming impacts network performance by enabling direct browser-to-browser communication, by that reducing server loads and optimizing bandwidth usage through the efficient distribution of video content among peers using a pull-based protocol [18]. For instance, a paper presents an end-to-end system utilizing WebRTC for low-latency streaming, enabling real-time object detection and other advanced applications. This system demonstrated latency as low as <200ms for standard video streams and <1s for 360-degree video streams, highlighting the importance of low latency for effective video streaming and processing [19]. Another study explores using VR and WebRTC to provide remote assistance for self-driving vehicles via 4K 360° live streams, with experiments confirming that sub-second glass-to-glass latency allows timely decisions by remote human operators in critical situations [20]. Also, a paper proposes a method for end-to-end latency measurement in edge computing with black-box components, validated with WebRTC and private 5G networks for extended reality (XR) video streaming, showing latency values mostly below 20ms [21]. Consequently, the combination of WebRTC with 5G technology holds significant potential for enabling high-quality, low-latency live multimedia services in mobile environments [22].

Additionally, a paper presents a mobile surveillance system leveraging WebRTC for real-time peer-to-peer communication to capture live feeds. The system performs object detection using the You Only Look Once (YOLO) [23] algorithm, achieving low latency and accurate surveillance with intruder alerts via Android devices [24]. Another paper evaluates a real-time face detection system using WebRTC, showing that it enables secure, low latency, platform-independent communication, effectively detecting faces in various light conditions [25].

One of the most challenging aspects of real-time object detection is benchmarking. Studies have been conducted to develop frameworks for online streaming perception [26]. This paper introduces a new metric and framework for evaluating real-time perception systems in autonomous agents, emphasizing the integration of latency and accuracy to enhance performance in dynamic environments. Furthermore, alternative approaches such as Stream YOLO present methods for streaming perception. The paper introduces an approach for streaming perception in autonomous driving, combining real-time object detection with future prediction to improve latency and accuracy, resulting in a 4.9% increase in Average Precision (AP) on the Argoverse-HD dataset [27]

3 Background

The background section provides foundational information on Wi-Fi technologies and the development of 5G networks, particularly private 5G networks, and their relationship to intralogistics. It also covers essential details about streaming services, object detection tasks, and provides informative insights into the hardware used for the use cases and benchmarks.

3.1 Wi-Fi

Wireless Fidelity (Wi-Fi) has reshaped the way people connect to the internet and communicate wirelessly. Since its beginning, Wi-Fi has evolved significantly, driven by the need for higher data rates, better efficiency, and more reliable connectivity. This section provides a comprehensive overview of the historical context, theoretical frameworks, and foundational concepts of Wi-Fi, highlighting key milestones and the current state of research.

3.1.1 Historical Context and Evolution

Wi-Fi technology, based on the Institute of Electrical and Electronics Engineers (IEEE) 802.11 [28] family of standards, has undergone several iterations since its first standardization in 1997. The initial versions, IEEE 802.11 and 802.11b [29] provided basic wireless connectivity but were limited in terms of speed and range. The subsequent introduction of IEEE 802.11g improved data rates up to 54 Mbps, laying the groundwork for widespread adoption in residential and commercial settings.

The next significant step came with IEEE 802.11n [30], which introduced MIMO (Multiple Input Multiple Output) technology, drastically enhancing throughput and reliability. This was followed by IEEE 802.11ac (Wi-Fi 5) [31], which offered even higher speeds and better performance in crowded environments by utilizing wider channel bandwidths and advanced modulation schemes. Below are the key milestones listed:

- IEEE 802.11 (1997): The first standard, offering data rates up to 2 Mbps.
- IEEE 802.11b (1999): Enhanced speeds up to 11 Mbps.
- IEEE 802.11g (2003): Further improved speeds up to 54 Mbps.
- IEEE 802.11n (2009): Introduced MIMO technology, increasing speeds up to 600 Mbps.
- IEEE 802.11ac (2014): Provided Gigabit speeds, better range, and improved performance in dense environments.
- IEEE 802.11ax (Wi-Fi 6, 2019): Focused on efficiency, capacity, and performance, also in dense environments. It supports a range of advanced features, including OFDMA (Orthogonal Frequency Division Multiple Access) and MU-MIMO (Multi-User MIMO).

- IEEE 802.11be (Wi-Fi 7): Already introduced, it aims to provide Extremely High Throughput (EHT) with speeds up to 46.1 Gbps, supporting advanced applications such as 4K/8K video streaming, AR/VR, and industrial IoT [32].

3.1.2 Theoretical Frameworks and Foundational Concepts

Wi-Fi operates on the principle of using radio waves to transmit data over the air, utilizing the Industrial, Scientific, and Medical (ISM) bands, specifically the 2.4 GHz and 5 GHz frequency bands. The key components of a Wi-Fi network include the internet, provided by the internet service provider; the modem, which converts the signal from the internet service provider to a signal usable in the network; the router, which manages local network traffic and assigns IP (Internet Protocol) addresses to devices; the access point, which broadcasts the Wi-Fi signal and allows devices to connect; and Wi-Fi clients, such as computers or phones with wireless network interface cards that connect to the network. The structure of the Wi-Fi network is illustrated in Figure 1 below. The lines represent the connections between each component and are multidirectional. For instance, a request from a Wi-Fi client travels from the client to the access point, then to the router, and finally to the modem and out to the internet service provider. In modern home setups, the modem, router, and access points are often combined into a single device.

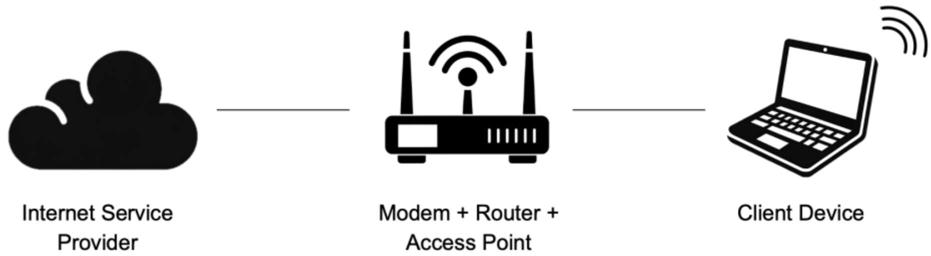

Figure 1 Wi-Fi Network Structure

Within these components, the MAC (Media Access Control) and PHY (Physical Layer) are essential for communication. The PHY layer handles the physical connection between devices, including the modulation and transmission of data over the air. The MAC layer, on the other hand, manages access to the network, controls data packet transmission, and ensures data integrity and security. Wi-Fi establishes and maintains connections using scan, authentication, and association procedures. The scan procedure identifies nearby devices and their MAC addresses, authentication validates their identity, and association establishes the connection for data transmission [33].

Today's Wi-Fi networks leverage several advanced technologies to enhance performance, range, and reliability:

1. **MIMO (Multiple Input Multiple Output)**: This technology uses multiple antennas to send and receive more data simultaneously, significantly improving data throughput and channel capacity [34].

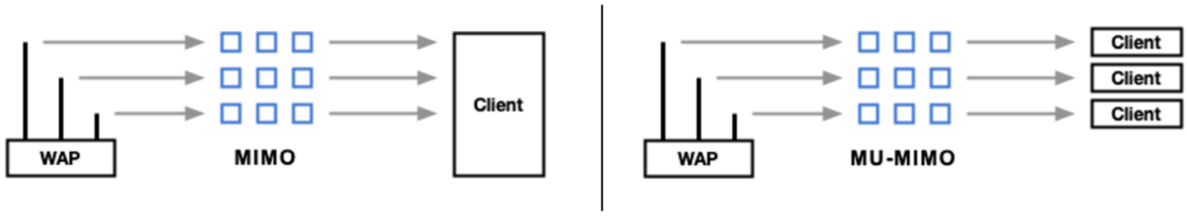

Figure 2 MIMO and MU-MIMO Architecture Comparison [35]

2. **MU-MIMO (Multi-User MIMO)**: An extension of MIMO, this technology allows the router to communicate with multiple devices simultaneously, increasing the system throughput and lowering the degree of network congestion [36]. Figure 2 illustrates the difference between the technologies. The lines represent the data transmission pathways.

3. **OFDMA (Orthogonal Frequency-Division Multiple Access)**: OFDMA, utilized in Wi-Fi 6, enhances network performance by dividing channels into smaller, overlapping sub-channels to serve multiple devices simultaneously. This approach reduces latency and increases efficiency in high-density environments [37].

These technologies, integrated into modern Wi-Fi standards, enhance the user experience by providing less congestion, high throughput, and greater reliability.

3.1.3 Current State of Research and Future Directions

Wi-Fi 6 introduces significant enhancements for high-density environments, including improved latency, increased throughput, and reduced congestion. These advancements enable seamless applications of AR, VR, and IoT technologies and real-time video streaming. Parallel to these technical improvements, Wi-Fi usage has consistently grown year after year. As shown in Figure 3, the proportion of IP traffic for Wi-Fi has been increasing [38, 39]. The proportion of IP traffic for Wi-Fi (combining Wi-Fi only and mobile/Wi-Fi) increased from approximately 43% in 2017 to about 51% by 2022. During this period, the total traffic volume increased from 120 exabytes to 400 exabytes. Additionally, Figure 3 shows that in 2022, approximately 71% of total IP traffic was wireless.

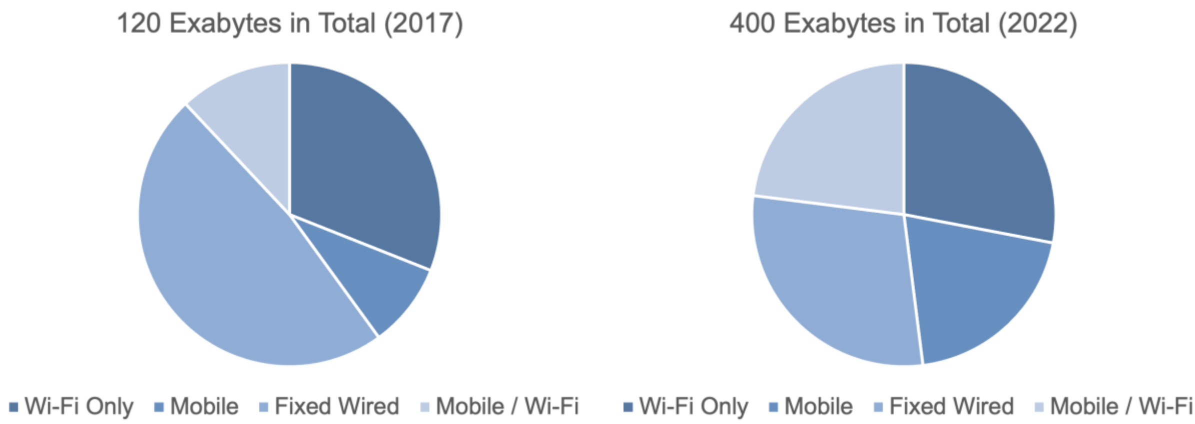

Figure 3 Approximate Global Monthly IP Traffic for Different Internet Access Methods in 2017 and 2022 [38, 39]

The recently introduced Wi-Fi 7 proposes even higher throughput, enhanced spectral efficiency, and greater reliability. Key features include support for 320 MHz channel bandwidth, 4K-QAM modulation, and enhanced multi-link operations, which enable simultaneous data transmission across multiple bands and channels [32]. Although the technology is new and requires further examination, it sets the stage for Wi-Fi 8. Future Wi-Fi 8 standards are anticipated to address current technical challenges by focusing on increased network capacity, improved QoS, and higher reliability through collaborative multi-device operation modes and innovative transmission schemes [40]. Additionally, use-case-specific modes of operation may be defined for Wi-Fi 8, optimized for managed environments while ensuring compatibility with legacy devices [40].

3.2 5G Technology

3.2.1 The Evolution of Mobile Networks: From 1G to 5G

The historical evolution of mobile communications spans from the early days of 1G to the advent of 5G. Each generation brought technological advancements and addressed various limitations, progressively building upon the achievements of its predecessors. The transition from the analog systems of 1G, through the introduction of digital technologies in 2G, the enhancement of data services in 3G, and the high-speed connectivity of 4G, set the stage for the advanced capabilities of 5G. This trajectory highlights the continuous innovation and development that have led to the high-performing mobile networks people rely on today.

3.2.1.1 First Generation (1G)

1G, introduced in the 1980s, marked the pioneering phase of mobile telecommunications by establishing the first commercially available cellular network. The primary technology underpinning 1G was analog voice communication, specifically designed for voice transmission without native support for data communication, which would become standard in later generations. However, 1G had several limitations. The analog system had low capacity and was capable of handling only a small number of simultaneous conversations due to the high bandwidth consumption of analog signals. Additionally, 1G suffered from poor voice quality, as analog signals were highly susceptible to noise and interference, resulting in degraded voice clarity. Such disturbances were typically introduced by various environmental factors and electronic interference. Furthermore, the lack of encryption in 1G networks posed significant security concerns, as voice calls transmitted over these networks could be easily intercepted by unauthorized parties using basic radio scanners [41].

3.2.1.2 Second Generation (2G)

In the early 1990s, the arrival of the Second Generation (2G) of mobile telecommunications introduced a technological shift from the analog systems used in 1G to digital networks. This transition allowed the implementation of various digital communication standards, most notably Global System for Mobile Communications (GSM), followed by Enhanced Data Rates for GSM Evolution (EDGE) which further boosted data transmission speeds.

- **GSM**, introduced in the 1980s, altered mobile telecommunications with digital cellular networks, ISDN (Integrated Services Digital Network) compatibility, and worldwide roaming. It employs technologies such as Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA).

- **EDGE**, as an extension to GSM, EDGE provided faster data rates and is often referred to as a 2.5G technology and was considered a pre-3G technology. By employing higher-order modulation schemes, EDGE increased the data rate per time slot to 384 kbps.

The switch to digital with 2G brought several advancements over the first generation, particularly in terms of security, data services, and communication capabilities. One significant improvement was the introduction of encryption standards for voice and data transmissions, enhancing security and privacy compared to the unencrypted transmissions of 1G. Additionally, 2G networks allowed for data services over mobile devices for the first time, including basic internet access, opening the way for mobile internet applications that were fully realized in subsequent generations. Another notable feature was the introduction of Short Message Service (SMS) and Multimedia Messaging Service (MMS), aiding personal and business communication by enabling text and multimedia messaging across the network without a voice call. Furthermore, the digital architecture of 2G networks enhanced voice call quality. The use of digital encoding methods reduced noise and allowed for clearer conversations compared to the analog transmissions of 1G.

3.2.1.3 Third Generation (3G)

The early 2000s marked the deployment of the Third Generation (3G) of mobile telecommunications, substantially enhancing data transfer rates compared to its predecessors. This advancement was primarily enabled through the introduction of new technological frameworks such as Universal Mobile Telecommunications System (UMTS) and Wideband Code Division Multiple Access (W-CDMA).

- **UMTS**: As the successor to GSM and EDGE, UMTS employed W-CDMA radio access technology, which significantly increased bandwidth and data transfer rates. This advancement allowed for the smoother transmission of large data packets necessary for multimedia and internet services.

- **W-CDMA:** It uses a bandwidth of 5 MHz or more per carrier, providing enhanced coverage, greater capacity, and support for higher data rate services compared to narrow-band Code Division Multiple Access systems [42].

3G considerably expanded the capabilities of mobile devices, introducing several new services and features that were not feasible with 2G networks. Initially, 3G networks aimed to achieve data rates of up to 2 Mbps in controlled indoor environments and up to 144 kbps in outdoor environments [43]. However, advancements have since pushed these rates well beyond the initial targets, often exceeding several Mbps in real-world applications.

- **Video Calls**: One of the hallmark features of 3G was the ability to conduct real-time video calls. This utilized the much higher data bandwidth available through 3G networks, allowing users to send and receive live video feeds alongside voice data.

- **Enhanced GPS and Localization Services:** With 3G, GPS, and other location services provided more accurate and reliable positioning data. This improvement facilitated the development of a wide range of location-based services and applications.

- **Support for Multimedia**: Beyond just improved speeds, 3G networks support a wide range of multimedia services, including mobile TV and video on demand [44]. These capabilities transformed the mobile phone from a mere communication device into a powerful multimedia tool.

3.2.1.4 Fourth Generation (4G)

The deployment of fourth-generation (4G) networks, which began actively in 2009, is an important milestone in mobile communication technology. This generation introduced two major standards for broadband cellular networks:

- **LTE (Long-Term Evolution)**: As the most widely adopted 4G technology, LTE offers substantial improvements in speed and efficiency over 3G networks. It utilizes a simplified, flat network architecture that reduces the transfer latency significantly, facilitating speeds up to ten times faster than 3G [45].

- **WiMAX (Worldwide Interoperability for Microwave Access)**: Though less commonly used than LTE, WiMAX provides similar benefits in terms of high-speed data transfer and reduced latencies. It was particularly significant in early 4G deployments and in providing broadband internet access in previously underserved areas [46].

The enhanced capabilities of 4G have had a profound impact on both consumer and business practices by enabling a range of high-demand applications:

- **Streaming High-Quality Video and Audio**: With 4G, users experience significantly smoother and higher-quality video streaming due to increased bandwidth and faster data speeds. This has played an important role in the growth of streaming services that provide high-definition video and audio content.

- **Improved Mobile Gaming**: 4G's lower latency and higher data rates have greatly enhanced mobile gaming, allowing for real-time, multiplayer gaming experiences that were not previously possible on mobile networks.

- **Reliable High-Speed Mobile Internet**: The speed and reliability of 4G have improved mobile internet access, enabling more robust mobile applications and allowing for smoother, more consistent access to cloud services.

3.2.1.5 Fifth Generation (5G)

The introduction of Fifth Generation (5G) technology marks a significant advancement in wireless communication, offering capabilities that surpass those of its predecessors. Distinguished by several key technological innovations, 5G technology includes:

- **Higher Bandwidth**: 5G networks provide significantly increased bandwidth compared to 4G, enabling faster data speeds. Statistics indicate that 5G technology represents a tenfold improvement over 4G [47].

- **Extremely Low Latency**: A notable advancement of 5G is its sub-millisecond latency, typically around 1 ms, which is significantly lower than the 10 ms latency achieved by 4G networks. This improvement is critical for applications requiring real-time feedback.

- **Extended Network Capacity**: 5G technology enhances network capacity significantly, allowing a substantial increase in the number of connected devices per square kilometer. This capability is essential for supporting the rapid growth of IoT devices and the development of densely populated smart city initiatives [48].

Table 1 shows the improvements gained with 5G compared to 4G. As shown in Table 1, 5G offers peak data rates up to 20 Gbps, a twentyfold increase over 4G's 1 Gbps, and experienced data rates of 100 Mbps, ten times higher than 4G. It supports mobility up to 500 km/h, reduces radio latency to ~1 ms from 4G's ~10 ms, and can handle 1,000,000 devices per km², compared to 4G's 10,000 devices per km².

Table 1 Performance Requirements of 5G NR [49]

Technology	Peak Data Rates	Experienced Data Rates	Mobility	Radio Latency	Connection Density
4G-LTE	1 Gbps	10 Mbps	Up to 350 km/h	~10 ms	10000 devices/km2
5G-NR	20 Gbps	100 Mbps	Up to 500 km/h	~1 ms	1000000 devices/km2

The unique attributes of 5G unlock a variety of new and emerging services, as defined by ITU M.2083-0 [50], contributing to the future of networked society beyond 2020:

- **Enhanced Mobile Broadband (eMBB)**: 5G improves the speed and stability of mobile broadband, offering fiber-like speeds wirelessly. This enhancement is crucial for fulfilling the bandwidth requirements of data-intensive IoT applications, such as immersive AR, industrial video surveillance, and VR [51].

- **Ultra-Reliable Low-Latency Communications**: URLLC targets applications demanding immediate communication with strict requirements for high throughput, low latency, and high availability [52]. This functionality is essential for industrial automation, and mission-critical operations.

- **Massive Machine Type Communications (mMTC)**: mMTC supports scenarios with a large number of connected devices transmitting low-volume, non-delay-sensitive data [53]. This is vital for smart cities, industrial applications, large-scale sensor networks, smart utilities, smart grids, and smart homes/buildings, facilitating the widespread connectivity of IoT devices without overloading the network.

3.2.2 Impact of 5G on Industries

The 5G technology promises to significantly transform multiple industries by providing high-speed, stable, and efficient communication networks. This technological advancement is expected to transform:

Smart Cities: Leveraging 5G's massive capacity, speed, and ultra-low latency, smart cities can enhance real-time traffic management, emergency response, and energy distribution while driving economic growth and innovation through IoT and Intelligent Transportation Systems (ITS) [54].

Autonomous Driving: The ultra-reliable and low-latency communication of 5G is fundamental for the safety and efficiency of autonomous driving systems, which require immediate data transmission to navigate complex environments [55].

Internet of Things: 5G's extensive network capacity and enhanced bandwidth enable the seamless connection of millions of IoT devices across various sectors, including smart cities, factories, agriculture, and healthcare. This high-speed, massive connectivity supports advanced applications while addressing challenges in security, data management, and scalability [56].

Industrial Automation: 5G drives the digital transformation of manufacturing by enabling sophisticated, connected automation processes, real-time monitoring, and control systems, fostering distributed production, collaborative robots, and integrated logistics under the Industry 4.0 paradigm [57].

3.2.3 Core Technological Advancements in 5G

Enhanced Spectral Efficiency: 5G introduces enhanced spectral efficiency, enabling more effective use of the spectrum to transmit greater amounts of data while maintaining high quality and speed. This is achieved through advanced modulation techniques and encoding, allowing the flexible and efficient use of fragmented, unused spectrum for various deployment scenarios. In intralogistics, this enhanced spectral efficiency allows for more devices to operate concurrently within a warehouse or manufacturing facility without interference, supporting a higher density of IoT sensors and mobile devices. The use of new waveforms like OFDM and Universal Filtered Multi Carrier (UFMC) makes better use of available spectrum and ensures they work without interfering with older systems, thereby improving how efficiently the spectrum is used [58].

Network Slicing: Network slicing enables the creation of multiple virtual networks within a single physical 5G infrastructure, each tailored with specific capabilities to meet diverse service requirements. This capability supports the vision of 5G as a multi-service network accommodating a wide range of verticals, from enhanced mobile broadband to critical machine-type communications. In intralogistics, network slicing can effectively segregate various data traffic types, such as operational data, real-time machine communications, and security systems, ensuring that critical communications occur without delay or data integrity loss. This flexibility is crucial for accommodating the heterogeneous requirements of different services, enabling the efficient and reliable operation of a vast array of applications over a common physical infrastructure [59].

Beamforming: Beamforming is a technique in 5G networks that focuses wireless signals toward specific receiving devices to enhance signal strength and reduce interference. This method is particularly vital in densely packed industrial environments where numerous obstacles can obstruct signals [60]. By directing the signal accurately, beamforming significantly enhances the reliability and efficiency of wireless communications within intralogistics, ensuring stable connections for AGVs and drones.

Massive MIMO: MIMO technology leverages a large number of antennas at both the transmitter and receiver to significantly boost communication performance by increasing network capacity, throughput, and efficiency through the simultaneous handling of multiple data signals [61]. In intralogistics, Massive MIMO supports the high volumes of data traffic generated by sensors, wearables, and mobile devices in real time, enabling complex operations such as inventory tracking and real-time monitoring of logistic processes.

3.2.4 5G Architectures

The deployment of 5G networks involves various architectures designed to optimize performance, compatibility, and cost-efficiency. These architectures enable mobile network operators to gradually transition from existing 4G infrastructure to advanced 5G capabilities. The two primary deployment architectures are Non-Standalone (NSA) and Standalone (SA).

3.2.4.1 Non-Standalone Architecture

NSA architecture utilizes existing 4G LTE infrastructure as support for control signaling while incorporating new 5G radio access capabilities. This approach allows operators to leverage their current 4G network to provide 5G services, making it a faster and more cost-effective deployment method compared to building a completely new network from scratch [62]. In NSA deployments, the 4G LTE Core Network, also referred to as the Evolved Packet Core (EPC), handles control plane functions and mobility management, while 5G New Radio (NR) provides high-speed data connections [63]. This architecture ensures compatibility with existing 4G infrastructure and services.

Figure 4 illustrates the deployment methods defined by The 3rd Generation Partnership Project (3GPP). Option 1 (SA LTE) represents the existing 4G LTE network without 5G integration, using the EPC and LTE RAN. Option 2 (SA 5G NR) is a pure 5G deployment with 5G NR operating independently on its own 5G core network (5GC). Option 3 (NSA 5G NR with EPC) deploys 5G NR alongside LTE, using the EPC. Option 4 (NSA 5G NR with 5GC) deploys 5G NR alongside LTE, both connected to the new 5G. Option 5 (SA LTE with 5GC) upgrades the LTE network to connect to the 5GC, benefiting from its new capabilities without deploying 5G NR. Option 7 (NSA LTE with 5GC and 5G NR) allows LTE to connect to the 5GC while using 5G NR for dual connectivity.

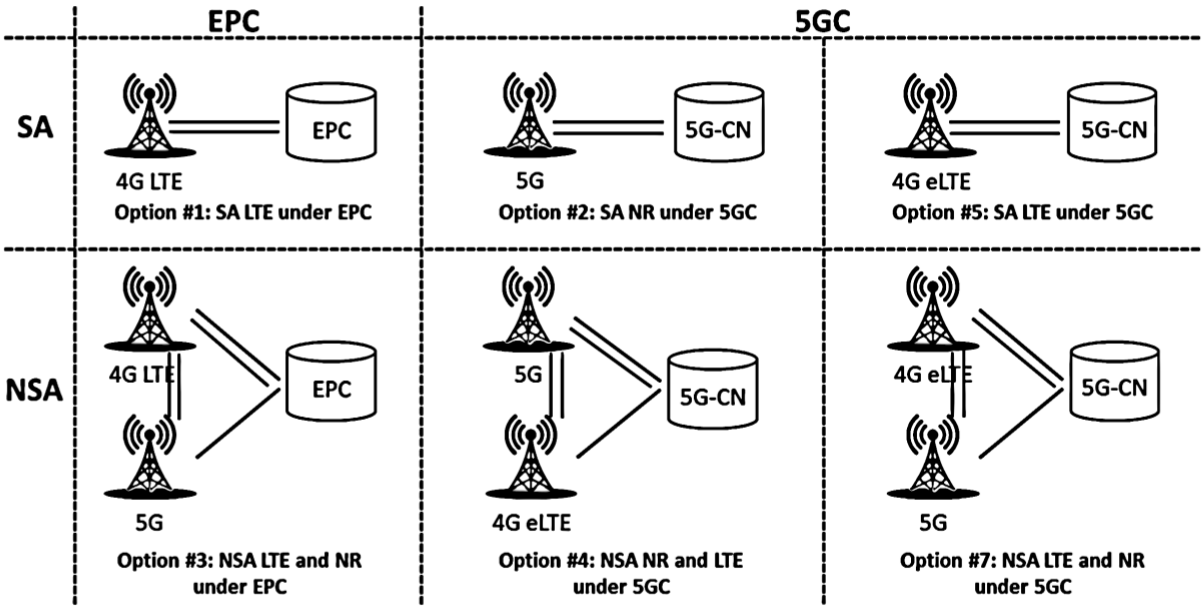

Figure 4 Deployment options 1, 2, 3, 4, 5, and 7 being defined by 3GPP [63]

3.2.4.2 Standalone Architecture

SA architecture involves deploying a new, dedicated 5G network that includes both the 5GC and 5G NR, unlike NSA architecture, which relies on existing 4G infrastructure. The 5GC supports all 5G capabilities, including network slicing. SA supports URLLC and mMTC, which are ideal for applications like autonomous vehicles and industrial automation. Trials comparing SA and NSA 5G networks show that SA slightly outperforms NSA in uplink data rates, with performance increasing monotonically with NR bandwidth [64].

3.2.5 Spectrum Allocations and Usage in 5G

The allocation and usage of spectrum in 5G are critical to achieving the high-speed, low latency, and mass connectivity goals of this new technology. 3GPP [65] has standardized a flexible spectrum approach that includes:

- **Low-Band Spectrum (< 1 GHz)**: Provides extensive coverage and indoor penetration at lower speeds, suitable for rural and suburban areas.

- **Mid-Band Spectrum (1-6 GHz)**: Offers a balance between coverage and capacity, with faster speeds and lower latency than low-band spectrum. This band is ideal for urban and suburban coverage.

- **High-Band Spectrum (Above 24 GHz)**: Known as millimeter waves, these frequencies offer the highest capacity and speeds, essential for dense urban areas but limited in coverage and penetration.

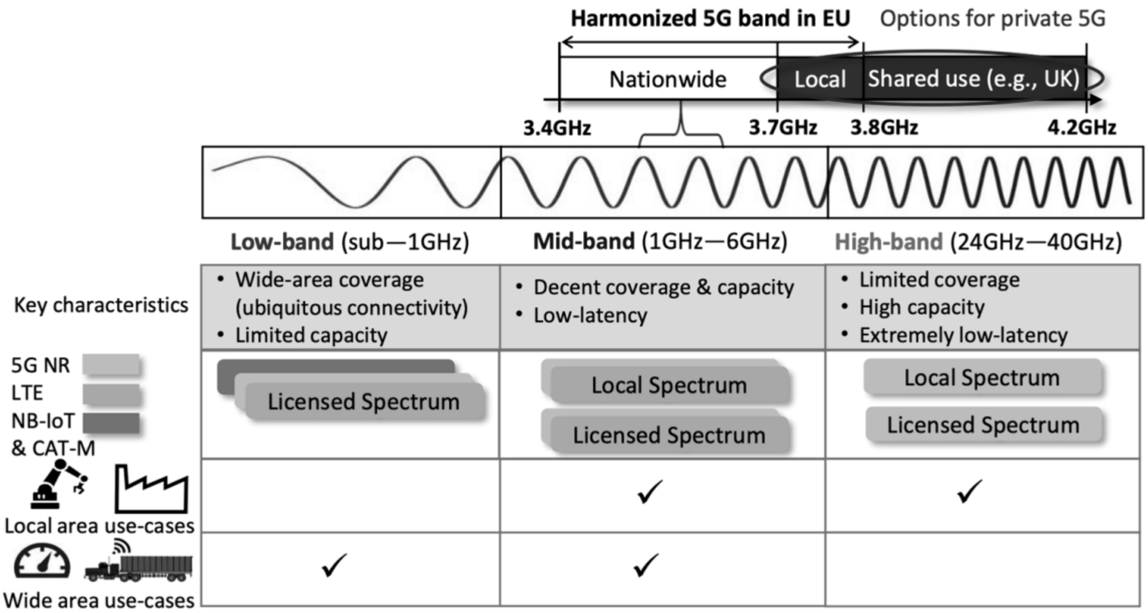

Figure 5 5G spectrum allocation in different bands [66]

Figure 5 shows a comparison of different 5G spectrums. These spectrum bands are vital for addressing the varied use cases of 5G, from widespread IoT deployments to ultra-reliable, high-speed mobile communications.

Even though 5G has many advantages, most security models from pre-5G networks cannot be directly utilized in 5G due to its new architecture and services, necessitating additional work [67]. On the other hand, 5G and Wi-Fi differ mainly in resource management and performance. For instance, 5G uses centralized scheduling, eliminating contention and ensuring efficient spectrum use, while Wi-Fi's Carrier Sense Multiple Access (CSMA) approach can lead to inefficiencies and interference [68]. CSMA is a Wi-Fi protocol where devices check the channel for availability before transmitting. In the context of autonomous devices, 5G is considered ideal for diverse use cases, while Wi-Fi is seen as optimal for fixed locations, where competition is also present. This brings the discussion to private 5G networks.

3.2.6 Private 5G Networks

Private 5G networks, also known as non-public networks (NPNs) [69], represent a significant advancement in industrial wireless communication, providing dedicated, reliable, and secure connectivity tailored to specific organizational needs. Unlike public 5G networks, private 5G networks are designed for exclusive use by an enterprise, offering greater control, customization, and security .

3.2.6.1 Architecture and Deployment Models

Private 5G networks, based on 5G NR technology, offer a variety of deployment models. As defined by 3GPP, these models include two primary categories: Standalone Non-Public Networks (SNPN) and Public Network Integrated Non-Public Networks (PNI-NPN). However, these can be further extended into three distinct deployment models to suit different enterprise needs [68]. Figure 6 illustrates these models, which are explained in detail below:

1. **Standalone Deployment**: The private 5G network operates independently from public networks. This configuration ensures that all data flows and network functions remain within the enterprise premises, providing maximum security and control. The network includes base stations, local 5G core, user data management, control plane, user plane function, and other necessary infrastructure.

2. **Shared RAN Deployment**: The private 5G network shares the Radio Access Network (RAN) with a public network while maintaining separate core network functions. This model allows for resource sharing without compromising the privacy of data flows. Base stations and spectrum are shared between public and private networks.

3. **Public-Private Shared RAN and Control Plane**: In this hybrid model, both the RAN and some control plane functions are shared with a public network. This setup uses network slicing to ensure isolated traffic and dedicated resources for the private network.

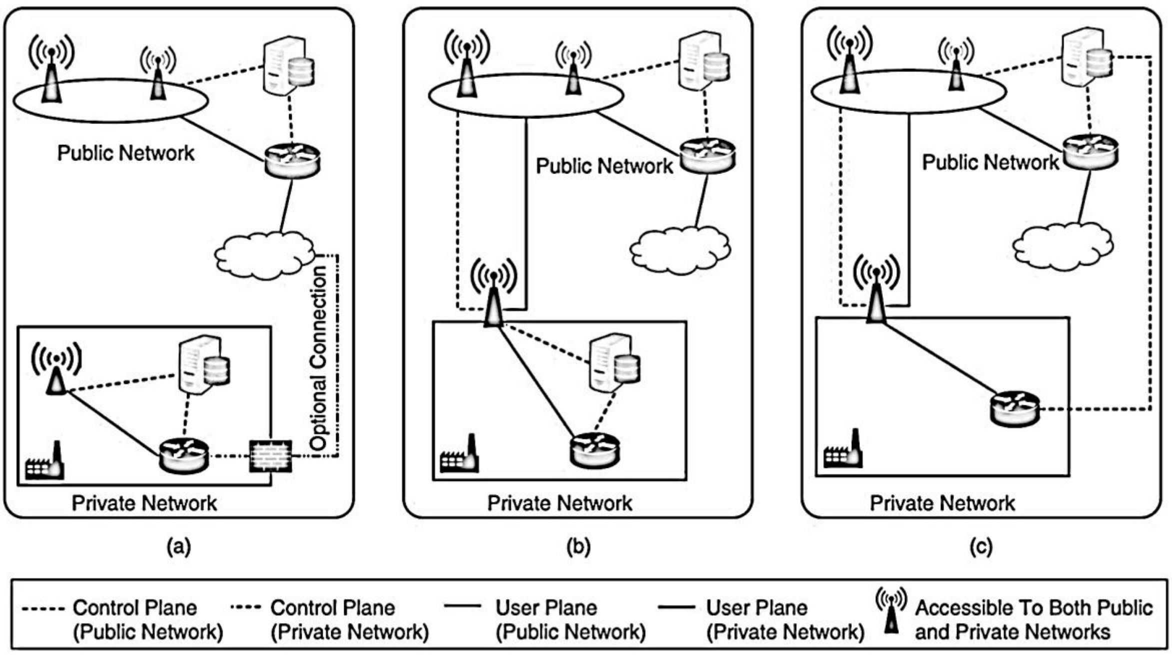

Figure 6 The functional architecture of private 5G networks [70]

(a) Standalone private deployment. (b) Public–private shared RAN deployment. (c) Public–private shared RAN and control plane deployment.

3.2.6.2 Technical Features and Innovations

In addition to the previously mentioned technical features of 5G in 3.2.3, the following features play a crucial role in the effectiveness of private 5G networks:

Multi-Access Edge Computing (MEC): MEC brings computation capabilities closer to the user, significantly reducing latency and enabling real-time data processing. This is essential for applications such as AR and industrial automation. By decentralizing computing resources and placing them at the edge of the network, MEC allows for more efficient data processing and management, thus enhancing the performance and reliability of time-sensitive applications [71].

Time-Sensitive Networking (TSN): TSN integrates with 5G to provide precise time synchronization across industrial devices, essential for automation and control systems. This integration enhances deterministic communication, ensuring low latency and high reliability, which are critical for industrial applications. TSN's ability to work with both Ethernet-based and 5G networks allows it to support diverse industrial automation requirements efficiently [72].

Given the critical nature of industrial data, private 5G networks incorporate advanced security protocols designed to meet restrictive requirements. These include end-to-end encryption, which ensures data protection throughout its journey across the network, mitigating risks associated with data interception [73]. Furthermore, private 5G networks offer organizations control over network traffic and data. By maintaining control over their data, organizations can ensure that sensitive information remains on-premises or within their control, which is essential for sectors with strict data privacy requirements [68]. This capability is important in adhering to national and international data protection regulations, such as the General Data Protection Regulation (GDPR). Moreover, the monitoring and management capabilities of private 5G networks provide organizations with real-time visibility into network performance, threat landscapes, and rapid response to potential issues. The flexibility of private 5G networks supports customized network configurations tailored to the specific compliance needs of various industries [70].

In summary, private 5G networks offer several key benefits. They provide dedicated coverage, ensuring robust connectivity in areas where public networks are unavailable or insufficient, particularly in remote or industrial locations. These networks also offer exclusive capacity, guaranteeing bandwidth and performance free from contention with public network users, which is essential for mission-critical applications [70]. Enhanced security is another significant advantage, as enterprises can implement their security policies, ensuring that sensitive data remains within the network and is protected against external threats. Additionally, private 5G networks offer customizable services, allowing networks to be tailored to meet the specific requirements of various industrial applications, from high-bandwidth needs to URLLC [1].

3.2.7 Standardisation for 5G

3GPP and the International Telecommunication Union (ITU) play crucial roles in the development and standardization of telecommunication technologies. 3GPP is a union of seven telecommunications standards development organizations (SDOs), including the European Telecommunications Standards Institute (ETSI), the Association of Radio Industries and Businesses (ARIB) in Japan, and the Telecommunications Standards Development Society, India (TSDSI). These SDOs have worked together to create a cohesive and globally applicable framework for 5G. 3GPP is responsible for creating detailed technical specifications for telecommunication technologies, which are then propagated within the industry. The ITU, particularly through its Radiocommunication Sector (ITU-R) and

Telecommunication Standardization Sector (ITU-T), evaluates these specifications for global applicability. ITU undertakes the task of spectrum allocation, harmonization, and regulatory frameworks necessary for international adoption. The interaction between 3GPP and ITU involves a process of submission, review, and iterative feedback. 3GPP submits its specifications to ITU, which conducts thorough evaluations and may recommend modifications to align with global regulatory and operational requirements.

3.2.8 Integration with Intralogistics

The integration of 5G technologies into intralogistics has the potential to make operations smarter, more efficient, and significantly more reliable. Enhanced spectral efficiency and massive MIMO provide the backbone for high-density device environments, essential for IoT and smart warehouse solutions. Network slicing and beamforming ensure that critical tasks can be carried out with minimal latency and maximum reliability, which are vital for autonomous systems and real-time decision-making in logistics operations.

The replacement of wired communications with 5G wireless technology significantly enhances flexibility and reduces maintenance costs in manufacturing environments. Common wireless technologies such as Bluetooth, Wi-Fi, and ZigBee fall short of meeting the stringent requirements of timeliness, reliability, data rates, scalability, and availability needed for cyber-physical production systems (CPPS) [74]. By providing eMBB, URLLC, and mMTC, 5G meets the diverse needs of industrial applications, supporting a wide range of use cases from AR for workers to the operation of autonomous guided vehicles and mobile robots on the shop floor [75]. Furthermore, 5G can also be a game changer for human-machine interactions in manufacturing by supporting enhanced Quality of Service (QoS) and enabling advanced applications such as AR-assisted operations and digital twin interactions [74]. These use cases require high data transmission rates, low latency, and secure connections, which 5G can efficiently provide.

3.3 Intralogistics

Logistic operations are divided into internal and external areas, with intralogistics focusing on the internal aspects. Intralogistics involves the management of materials and goods flow within a facility, encompassing internal logistics activities such as handling, transport, storage, and picking of goods. The Machinery and Equipment Manufacturers Association (VDMA) defines it as the organization, control, implementation, and optimization of the internal flow of materials, information, and handling of goods in industry, retail, and public facilities [76]. Digitalization and automation are key trends transforming work characteristics in this field, with automation having a stronger impact on manual tasks than digital technologies [77]. Significant changes are expected due to the possibilities for supporting or automating these tasks, enhancing efficiency, responsiveness, and flexibility in supply chain operations [78]. Additionally, there are competitions like the Amazon Picking Challenge (APC) [79] and the RoboCup@Home Challenge [80], which focus on warehouse automation and service robots for the distribution of objects.

3.3.1 Review of Traditional Technologies in Intralogistics

Traditional technologies have been integral to the development and operational efficiency of intralogistics. Below is an overview of some key technologies:

- **Barcoding**: Barcodes remain a fundamental technology in intralogistics for tracking inventory as goods move through various stages of the logistics chain. They are cost-effective and easy to implement, providing a reliable and straightforward method for identifying and recording goods.

- **RFID:** Radio-Frequency Identification (RFID) technology is used in intralogistics for tracking goods throughout the supply chain. This technology, characterized by its ability to read multiple tags simultaneously, larger data storage capacity, and read/write capabilities without the need for a line of sight, provides automatic identification and data capture capabilities [81]. RFID facilitates the automation of verification activities during shipping processes, thus reducing potential errors and improving real-time inventory management [82]. The RFID technology replaces the Barcode technology because barcode scanners are high in cost, and security is less [83].

- **Wi-Fi and Wireless Communication Technologies:** These technologies have facilitated communication within warehouses and distribution centers, supporting systems like mobile data terminals and warehouse management systems (WMS). Wi-Fi, in particular, is essential for high bandwidth applications and is widely used in modern warehouses [84] for connecting wireless scanners and other devices, enhancing operational efficiency.

3.3.2 Role of Automation and Digitalization in Intralogistics

Automation and digitalization improve intralogistics by enhancing efficiency, accuracy, and safety in warehouse operations. Through the integration of technologies such as AGVs, digital twins, and cloud computing, intralogistics can achieve higher levels of automation, from basic support to fully autonomous systems. These technologies contribute to more effective operations:

- **Automated Guided Vehicles**: AGVs are crucial for automating transport tasks within warehouses, offering efficient material handling without human intervention. They enhance flexibility and precision by integrating with robotic cells and other storage machinery. This integration not only increases efficiency but also reduces labor costs and improves safety by minimizing human interaction with heavy loads [85].

- **Digital Twins:** Digital twins create dynamic virtual representations of physical systems, enabling real-time monitoring and optimization. In warehouses, digital twins enhance decision-making and process efficiency by providing accurate simulations and predictive analytics [86].

- **Cloud Computing:** Cloud technology supports intralogistics by offering scalable resources for data storage and analytics, which allows better decision-making and integration across the supply chain. It achieves this, for example, by expanding the capabilities of IoT and providing the flexibility needed to manage the vast amounts of data generated by IoT devices.

In Table 2 below, the intralogistics operations are classified as automation [87]. Here, the levels mean different stages of automation for various tasks, ranging from manual execution (Level 0) to full autonomy (Level 5). This classification includes motorized support (Level

1), basic sensors for object identification (Level 2), autonomous vehicles for simple routes (Level 3), highly automated processes for standardized goods (Level 4), and fully autonomous systems handling all processes (Level 5). As seen in Table 2, no operations are listed under Stage 5. This is because, at present, there are no realized operations representing full autonomy with no human intervention.

Table 2 Classification of Tasks in Intralogistics [87]

Automation Stage	Transport	Storage	Order Picking	Handling	Packaging
5					
4	Non track-guided AGV	Automated storage and retrieval system	Autonomously driving picking robot	Robotic (de-) palletizer for cubic elements	Packaging robot and Automated packaging machine
3	Track-guided AGV	Automatically following vehicles	Automatically following picking vehicles and stationary picking robots	Robotic (de-) palletizer for repeating problems	Packaging machine for standardized products
2	Forklift or tugger train with assistance systems	Forklift or reach truck with assistance systems	Pick-by-voice systems	Manipulators and scissor lifts	Carton erectors
1	Forklift and tugger train	Forklift and powered manipulators	Forklift and order pickers	Powered manipulators and scissor lifts	Powered manipulators
0	Mechanically driven platform trolley	Mechanically driven platform trolley	Mechanically driven platform trolley	Mechanically driven manipulators for heavy goods	Mechanically driven manipulators for heavy goods

Digitalization is a major trend in intralogistics, not only in academia but also in the industry [88], further motivating advancements in this field. With its ultra-reliable, high-bandwidth, and low-latency communication, 5G technology holds transformative potential for intralogistics, enabling real-time data processing, automation, and machine-to-machine communication.

3.4 Live Streaming Technologies

3.4.1 Streaming Protocols

Streaming protocols are the rules and methods used to deliver multimedia content over the internet in real-time or on-demand. These protocols are designed to ensure efficient and reliable delivery of video and audio streams, accommodating different network conditions and client capabilities. The classification of streaming protocols can be broadly divided into two main categories: Hypertext Transfer Protocol (HTTP) based streaming protocols and real-time streaming protocols.

3.4.1.1 HTTP-Based Streaming Protocols

HTTP-based streaming protocols are widely used because they operate over standard web protocols (HTTP/HTTPS), making them highly compatible with existing internet infrastructure and easy to deploy through content delivery networks (CDNs) [89].

- **HLS (HTTP Live Streaming):** Developed by Apple, HLS is one of the most commonly used streaming protocols. It works by breaking down the video into a sequence of small HTTP-based file segments, each containing a short chunk of the video stream. The client then plays these segments in order, which allows for adaptive bitrate streaming by switching between different quality streams based on available network conditions [90].

- **MPEG-DASH (Dynamic Adaptive Streaming over HTTP):** MPEG-DASH is an adaptive bitrate streaming technique similar to HLS but is an international standard. DASH provides high quality and efficient streaming by adapting the bitrate of the multimedia content to the current network conditions and capabilities of the device [91].

- **HTTP Dynamic Streaming:** Developed by Adobe, HTTP Dynamic Streaming is similar to HLS and enables high-quality media streaming over the HTTP network protocol. It dynamically segments media files and delivers them over standard HTTP protocols [92].

- **Microsoft Smooth Streaming:** An adaptive streaming protocol that adjusts video quality in real time based on the user's network and device performance. It's similar to HLS and DASH but was specifically designed for use with Microsoft's Silverlight framework [93].

3.4.1.2 Real-Time Streaming Protocols

Real-time streaming protocols are designed to deliver live content with minimal latency, making them ideal for applications such as industrial automation or real-time monitoring.

- **RTSP (Real-Time Streaming Protocol):** RTSP is used for establishing and controlling media sessions between endpoints. It allows the client to remotely control a media server, fetching content as needed without downloading the entire stream. This is particularly useful for surveillance systems [94].

- **RTP (Real-Time Transport Protocol):** Often used in conjunction with RTSP, RTP is designed for end-to-end, real-time transfer of streaming media. It helps manage the data transfer of audio and video by packetizing media streams [95].

- **RTMP (Real-Time Messaging Protocol):** Originally developed by Macromedia and later owned by Adobe, RTMP was designed to transmit audio, video, and data over the Internet between a Flash player and a server. It uses a persistent connection and allows low-latency communication [96].

- **WebRTC (Web Real-Time Communication):** Unlike other streaming protocols, WebRTC is designed for peer-to-peer connections and provides capabilities for streaming audio, video, and arbitrary data between browsers without the need for an intermediary [97]. It's ideal for applications that require minimal latency, such as remote control of industrial equipment.

Each streaming protocol serves different use cases and has its strengths and weaknesses. HTTP-based protocols like HLS and MPEG-DASH are excellent for scalable, on-demand video delivery but usually have higher latency, making them less suitable for real-time interaction. In contrast, real-time protocols like WebRTC offer lower latency but can be more complex to scale for large audiences without additional infrastructure. The choice of protocol largely depends on the specific requirements of the streaming application, such as the need for real-time delivery, compatibility with devices and browsers, scalability, and the network environment.

In this part, the focus will be more on the RTMP and WebRTC protocols as they are the selected protocols for the thesis use case.

3.4.1.3 Real-Time Messaging Protocol

RTMP was initially developed by Macromedia, which was later acquired by Adobe. This protocol was primarily used to stream audio, video, and data over the Internet between a Flash player and a server. However, now RTMP is also widely used for live streaming applications across various platforms. RTMP maintains a persistent, TCP-based connection to ensure stable delivery of streams. The protocol is divided into multiple channels, allowing data to be segmented and sent concurrently for high-performance transmission. RTMP streams are encapsulated in small packets that can carry audio, video, or other data types.

RTMP offers several advantages that make it a preferred protocol choice for specific applications. Designed for low-latency communication, RTMP is particularly suitable for live streaming applications such as broadcasts, live events, and gaming. Studies have confirmed the low latency of RTMP live streams over 5G networks, as illustrated in Figure 7 [12]. Figure 7 shows that 5G averages around 12 ms delay for uplink and 8 ms for downlink.

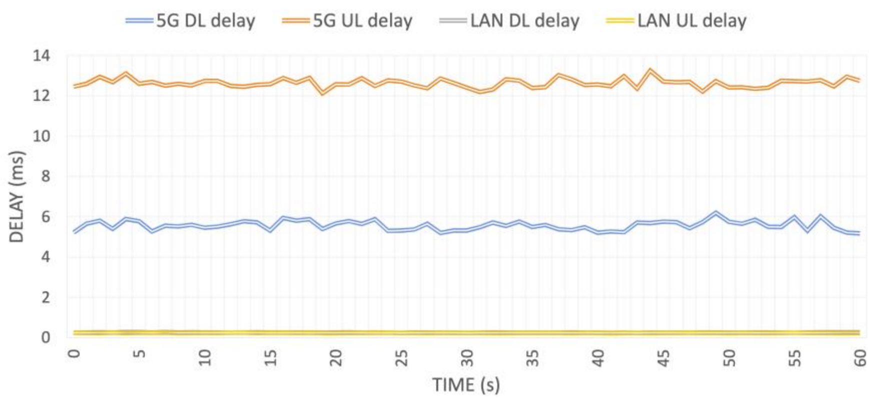

Figure 7 Average RTMP Delay for Live Stream [12]

3.4.1.4 Web Real-Time Communication

Web Real-Time Communication (WebRTC) is an open-source project that provides web browsers and mobile applications with real-time communication via simple application programming interfaces (APIs). It supports video, voice, and generic data to be sent between peers, eliminating the need for external plugins or third-party software. The quality of video performance in WebRTC is influenced by packet losses and round trip times, underscoring the critical importance of network quality for optimal functionality [98]. Figure 8 shows the architecture of WebRTC. The architecture has two layers: the browser layer (including WebRTC API, transport/session, RTP stack, STUN/ICE, and codecs) and the web app layer (providing a Web API for video/audio chat applications).

WebRTC implements several advanced technologies to facilitate real-time communication:

- **Peer-to-Peer Architecture:** Unlike traditional streaming protocols that typically use a server to relay data, WebRTC allows direct data transfer between browsers or devices. This significantly reduces latency and increases the efficiency of data transmission.

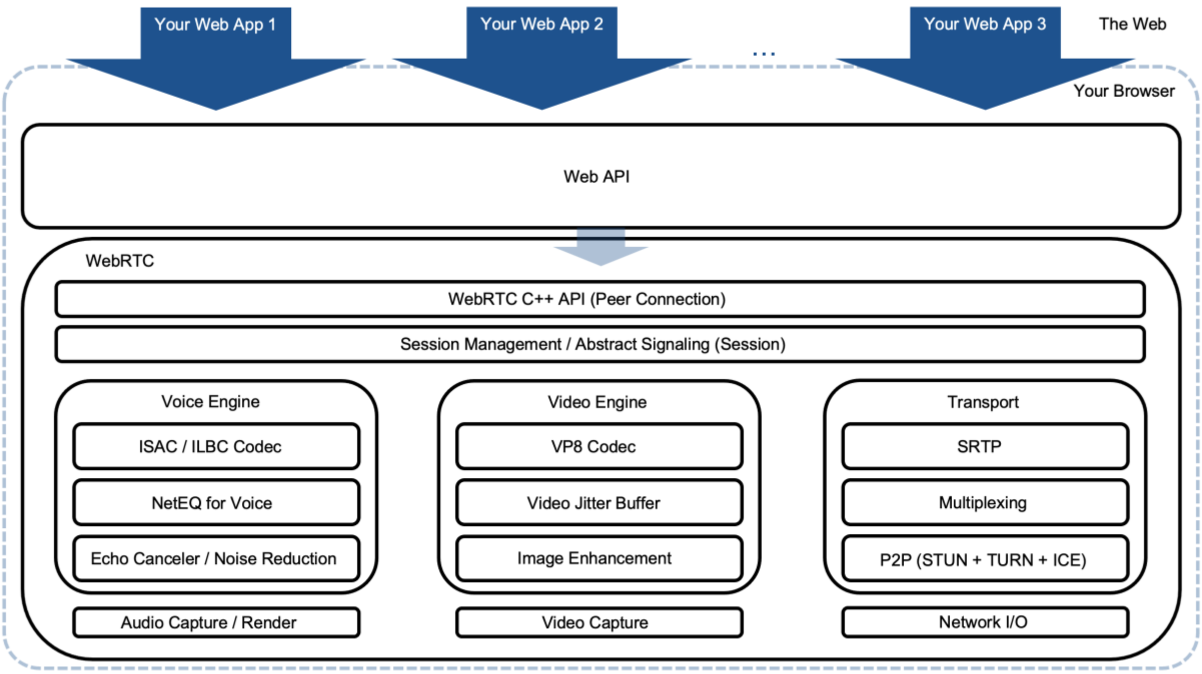

Figure 8 WebRTC Architecture [99]

- **NAT Traversal:** WebRTC uses Interactive Connectivity Establishment (ICE) to overcome network address translator (NAT) and firewall restrictions, improving compatibility and connectivity between different networks.

- **Media Capture:** Browsers can directly capture media from the user's device (camera and microphone) using the Media Capture and Streams API embedded in WebRTC.

WebRTC offers several key features and benefits that make it an attractive choice for modern communication solutions. Firstly, its low latency enables real-time communication with minimal delays [100]. Secondly, WebRTC ensures secure communication by encrypting all components, including signaling, data transfer, and media, through protocols such as Datagram Transport Layer Security (DTLS) and Secure Real-time Transport Protocol (SRTP) while employing Interactive Connectivity Establishment (ICE) checks for media consent verification to protect against unauthorized access [101]. Furthermore, WebRTC is supported by major browsers, including Chrome, Firefox, Safari, and Microsoft Edge, allowing developers to implement complex communication solutions without relying on platform-specific apps.

In summary, RTMP continues to be valuable for backend processes in streaming workflows, whereas WebRTC excels in front-end real-time communication within web applications. Each has distinct advantages that make them suitable for different aspects of modern streaming and communication solutions.

3.4.2 Oven Media Engine

Oven Media Engine (OME) [102] is designed to deliver high-quality, real-time video streaming, with a particular focus on low-latency performance. The engine's modular architecture supports customization and scalability for specific project requirements. OME's transcoder ensures efficient video delivery by converting streams to various codecs and formats for diverse devices and bandwidths.OME also supports hardware acceleration technologies (e.g., NVIDIA CUDA, Intel Quick Sync) to enhance performance and reduce server load.

The overall pipeline for streaming using OME is detailed below. Figure 9 illustrates the technical workflow.

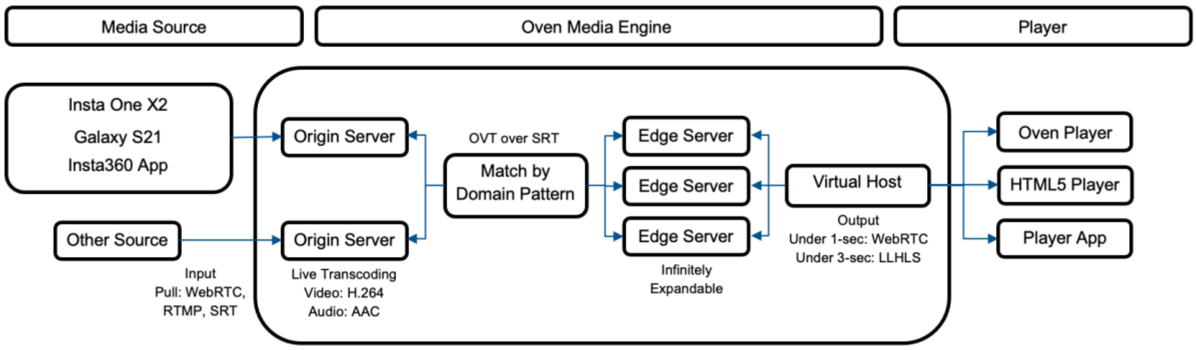

Figure 9 Oven Media Engine Pipeline [103]

1. **Ingestion:** The streaming process starts with the ingestion of a live video source through RTMP or other supported protocols. This source can originate from live encoder software or hardware that transmits the live feed directly to the OME server.

2. **Processing:** Upon ingestion, the stream undergoes processing according to the configured profiles. This processing includes transcoding, packaging, and preparing the stream for delivery. The transcoder optimizes the video stream by adjusting the bitrate, resolution, and codec to ensure compatibility with the target delivery channels.

3. **Delivery:** OME supports several delivery methods:

    - **WebRTC for ultra-low latency streaming**, is ideal for real-time interactions and applications requiring sub-second latency.

- **HLS and MPEG-DASH for HTTP-based streaming** are more suitable for broad distribution and compatibility with standard HTTP servers and CDNs.

- **Low Latency MPEG-DASH (LL-DASH)** and **Low-Latency HLS (LL-HLS) [104]** for scenarios where a balance between low latency and broad compatibility is needed.

4. **Playback:** On the client side, OME streams can be accessed through compatible web players or native applications that support the chosen streaming protocols. WebRTC streams, for example, can be played in web browsers with no additional plugins required.

3.5 Object Detection

Object detection is a frequently misunderstood term, often confused with other concepts such as image classification, object localization, instance segmentation, and semantic segmentation. Image classification assigns a class label to the entire image, while object localization involves drawing bounding boxes around the objects in the image. Object detection merges these two tasks by drawing bounding boxes and assigning class labels to each object of interest. In addition, there is segmentation, which differs from object detection by taking a step further to define object boundaries and regions on a pixel-wise level, providing more granular results than object detection. Segmentation can be divided into two categories: instance segmentation and semantic segmentation. Semantic segmentation only predicts the category of each pixel and does not distinguish between object instances. Instance segmentation, on the other hand, predicts both the category of each pixel and the object instances. Object recognition can be broadly classified as encompassing all these terms [105].

Numerous deep-learning-based computer vision solutions are deployed across industries, including product quality inspection in manufacturing, computer-aided medical diagnosis, agricultural monitoring and automation, and traffic and road condition monitoring in transportation [106]. Specifically, these applications often utilize object detection technologies to enhance precision and efficiency in tasks. But also, by detecting pedestrians, vehicles, and obstacles, these algorithms boost the safety and reliability of autonomous vehicles, opening the way for secure and efficient future transportation [107]. However, several challenges persist in achieving successful object detection. For instance, in warehouse automation, challenges include handling diverse object types, such as textureless or reflective items undetectable by depth sensors and limitations in sensor range [108].

The field of object detection is predominantly influenced by Convolutional Neural Network (CNN) based architectures. Contemporary object detection models include R-CNN [109], Fast R-CNN [110], Faster R-CNN [111], YOLO [23], and Single Shot MultiBox Detector (SSD) [112]. These detectors generate outputs in the form of bounding boxes, object classes, and confidence scores.

R-CNN employs a selective search [113] algorithm to produce region proposals, each processed through a CNN, achieving high accuracy but at the cost of slow speeds. Faster R-CNN enhances this by incorporating a region proposal network, which improves speed and efficiency but does not match YOLO and SSD's real-time capabilities [114]. YOLO processes the entire image in a single pass through a neural network, delivering real-time performance with high speed, albeit with occasionally lower accuracy. SSD also targets real-time detection by utilizing multiple convolutional filters on feature maps at various scales, thus balancing speed and accuracy more effectively than YOLO [115].

In summary, YOLO and SSD are ideal for applications demanding high-speed detection, while Faster R-CNN is more suitable for tasks that require higher accuracy. Given our focus on the impact of network conditions, prioritizing higher detection speed is crucial for the application at hand.

Frameworks used for the use case:

1. **TensorFlow and TensorFlow.js:** TensorFlow [116] is an open-source framework developed by Google for machine learning and neural network research. TensorFlow.js [117] is a library for handling machine learning in JavaScript, allowing the implementation of machine learning directly in the browser. It allows for on-device object detection, which helps improve privacy and reduces latency.

2. **ONNX and ONNX.js:**

    - **ONNX (Open Neural Network Exchange):** ONNX [118] is an open format built to represent machine learning models. It allows models to be transferred between different frameworks, making moving models from one tool to another easier.

    - **ONNX Runtime:** ONNX Runtime [119] is a high-performance runtime for machine learning models in the ONNX format, designed to optimize the execution of machine learning models across various hardware platforms.

    - **ONNX.js:** ONNX.js [120] is a JavaScript library for running ONNX models directly in the browser, facilitating the use of ONNX models in web applications and enabling real-time object detection on the client side.

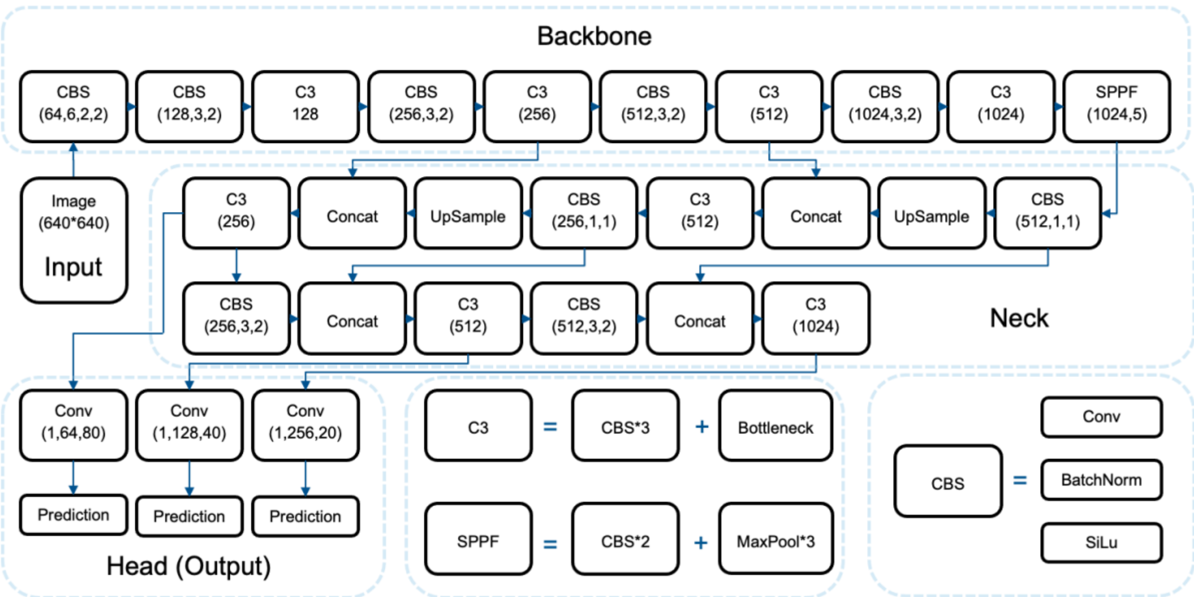

Figure 10 YOLO v5 Architecture [121]

Specific object detection models used for the use case:

1. **COCO-SSD:**

    - **Model Description:** COCO-SSD is a pre-trained object detection model based on the SSD framework, which is capable of detecting multiple objects in an image. It uses a base architecture of MobileNetV1 [122] or MobileNetV2, optimized for mobile and web applications due to its efficiency and relatively small size.

    - **Utility:** This model is trained on the Common Objects in Context (COCO) [123] dataset, which includes 80 object categories. It is well-suited for applications where multiple objects need to be detected in real time within diverse environments.

2. **YOLOv5s, YOLOv8s, YOLOv8x, and YOLOv9c:**

    - **Model Description:** YOLOv5 [124], YOLOv8 [125], and YOLOv9 [126] are state-of-the-art, real-time object detection models known for their speed and accuracy. The "s" variants, YOLOv5s and YOLOv8s, are one of the smallest and fastest models in their respective families, making them particularly suitable for applications requiring quick response times and efficient use of computational resources. The "x" variants represent the largest and most computationally intensive models in their families. YOLOv9 includes a unique naming convention, with the "c" variant being the second largest in its family. The architecture of YOLO v5 is shown in Figure 10.

    - **Utility:** All of the YOLO models are trained on the COCO dataset and are capable of precisely detecting multiple objects in real time. Their design allows for deployment on edge devices, enhancing their applicability in environments with limited computational resources. YOLOv9, the latest iteration, offers improved performance over YOLOv8, making it suitable for more complex environments requiring enhanced precision.

Table 3 illustrates a comparison of the YOLO models used in the experiments. The metrics include mAPval 50-95, which measures precision and recall across various Intersection over Union (IoU) thresholds (0.50 to 0.95), image size indicating pixel dimensions, the number of parameters representing trainable weights, and FLOPS for computational complexity.

Table 3 Performance Metrics for YOLO Models [124, 125, 127]

Model	Size (pixels)	Number of Parameters	mAPval 50-95	FLOPs
YOLOv5s	640	7.2M	37.4	16.5
YOLOv9c	640	25.5M	53.0	102.8
YOLOv8s	640	11.2M	44.9	28.6
YOLOv8x	640	68.2M	53.9	257.8

Integrating object detection models with live video streaming enables real-time detection and analysis of objects, offering significant benefits across various domains. For instance, in intralogistics, such models can be used in warehouses to track inventory in real time, optimize the placement of goods, and monitor the movement of items, thus improving the efficiency and accuracy of warehouse operations. However, several challenges and considerations must be addressed. Performance and accuracy are critical factors, as models like COCO-SSD and YOLO, while optimized for speed, may sacrifice some accuracy compared to more complex models. This trade-off needs careful management based on application requirements. Additionally, object detection models must be robust enough to handle environmental variability, such as different lighting conditions, occlusions, and object variations, which are common in dynamic environments. Furthermore, deploying advanced machine learning models on edge devices or in browsers presents constraints on memory and computational power, necessitating the use of models that are both accurate and resource-efficient.

3.6 Datasets

3.6.1 YouTube-Bounding Boxes Dataset

The YouTube-Bounding Boxes (YT-BB) [128] dataset, shown in Figure 11, is a large-scale benchmark dataset designed for object detection in video. It consists of a wide variety of labeled video segments from YouTube, providing annotations for object detection tasks.

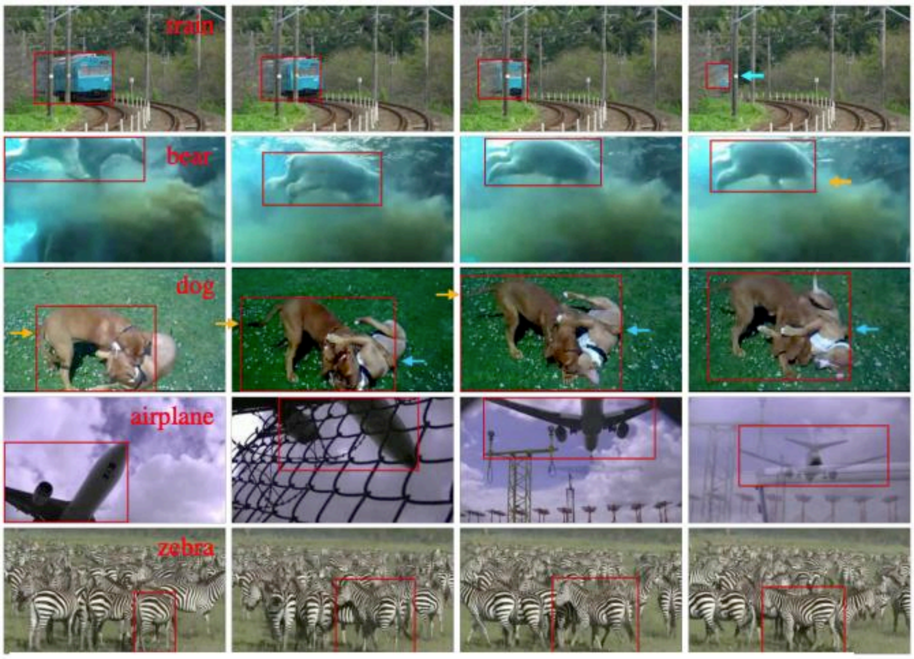

Figure 11 Youtube-Bounding Boxes Dataset

The dataset features 23 object categories, capturing a range of common objects in diverse environments and activities, all of which are a subset of the COCO dataset. It provides bounding box annotations for each frame, allowing localization of objects throughout the video sequences. The dataset contains approximately 380,000 video segments, each about 15-20 seconds long. In the use case, the YT-BB dataset is utilized to benchmark the object detection task under different network conditions.

3.6.2 COCO Dataset

The COCO dataset is one of the most widely used datasets for training and evaluating object detection, segmentation, and captioning models. The dataset includes rich annotations for object detection, segmentation, and key points across 80 object categories. It contains over 330,000 images, offering a wide range of real-world scenarios and object interactions.

In this use case and benchmarking, pre-trained models on the COCO dataset, such as COCO-SSD, YOLOv5s, and YOLOv8s, are utilized. These models, developed by other researchers, offer accurate real-time object detection capabilities. They benefit from the annotations of the COCO dataset.

3.7 Control and Processing Platform

3.7.1 Raspberry Pi

The Raspberry Pi [129] is a compact single-board computer developed by the Raspberry Pi Foundation in the United Kingdom. Originally intended for educational purposes, the Raspberry Pi quickly found a multitude of uses beyond the classroom due to its affordability, small size, and versatility. It has become especially prominent in IoT applications, where its capability to handle computing operations with minimal power consumption makes it an excellent choice. Figure 12 illustrates the Raspberry Pi 4B.

The Raspberry Pi is equipped with a CPU that balances cost and performance efficiently. It supports wireless connections and can accommodate a wide range of input and output peripherals. One of its defining features is the General-Purpose Input/Output (GPIO) pins, which facilitate the connection of a variety of external electronic devices. The GPIO pins are used to control the forklift in the use case.

Figure 12 Raspberry Pi 4B

As a fully functional computer system, the Raspberry Pi includes microprocessors, memory, input/output options, and other necessary components on a single Printed Circuit Board (PCB). This integration, typical of system-on-chip architectures, includes the CPU, RAM, GPU, and various interface cores on a single chip. This architecture is comparable to that used in modern smartphones and tablets.

The Raspberry Pi's reliance on the ARM architecture means it is inherently energy-efficient, with a low power footprint suitable for mobile and battery-powered applications. While software compatibility issues with x86 architecture-based applications remain, the extensive support from the open-source community helps bridge most gaps.

In the context of this thesis, the Raspberry Pi was chosen due to its capability to provide a reliable, low-cost, and energy-efficient platform and convenient integration with the RM520N-GL 5G HAT. This integration facilitates high-speed, low-latency communication essential for real-time object detection tasks.

3.8 5G Connectivity and Communication

3.8.1 5G Modem and 5G Hat

The Quectel RM520N-GL and Quectel RM502Q-AE modems were both used as 5G modems, yielding similar test results. Based on compatibility with TUM's private 5G network and previous testing, as well as recommendations from Telefonica experts for this use case, the Quectel RM502Q-AE was chosen. Quectel RM502Q-AE modem is well-suited for IoT operations due to its 5G capabilities. It supports peak data rates of 4.2 Gbps downlink, and 450 Mbps uplink in 5G SA mode. It is compatible with the following 5G bands in SA mode: n1, n2, n3, n5, n7, n8, n12, n20, n25, n28, n38, n40, n41, n48, n66, n71, n77, n78, and n79. In terms of power consumption, it uses 80 µA in shutdown mode, 4.2 mA in hibernate mode, 39 mA when idle with USB 2.0, and 54.5 mA when idle with USB 3.0 [130]. The modem provides favorable throughput levels, low latency, and compatibility with various 5G bands. Its power consumption is reasonable, which is crucial for mobile use cases.

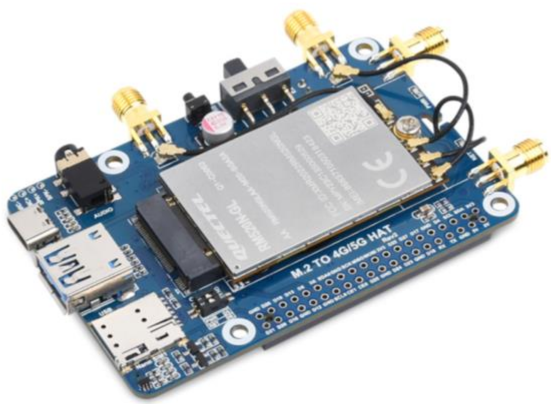

Figure 13 5G Hat for Raspberry Pi and Quectel 5G Modem

On the other hand, the Waveshare 5G HAT allows the RM502Q-AE to connect to the Raspberry Pi 4B. With multiple interfaces, including USB, PCIe, and dual SIM slots, the Waveshare 5G hat provides versatile connection between Raspberry Pi and 5G modem. Furthermore, the hardware connection is straightforward: the RM502Q-AE connects to the Raspberry Pi via USB 3.0, with additional power supplied through a USB-C connector to ensure stable operation. The 5G hat also allows the Raspberry Pi to serve as the power supply for the 5G modem, which is the configuration used in this specific use case. For the software setup, it is compatible with Raspberry Pi OS, Ubuntu, and OpenWRT, requiring no additional drivers for recent OS versions. The network setup supports multiple dial-up methods, which are configurable through AT commands. Thus, the RM502Q-AE enhances the efficiency, and connectivity of IoT systems, making it an optimal choice for applications requiring computer vision and real-time data processing. Figure 13 depicts the 5G HAT and the 5G modem.

3.8.2 ASKEY 5G Dongle

The ASKEY 5G dongle, model NDQ 1300, is a 5G USB modem. Equipped with the Qualcomm SDX55 multimode chip, it supports GSM, UMTS, LTE, and 5G networks. It operates in 5G mode using both Non-Standalone and Standalone technologies. With four antennas, it receives up to 2.7 Gbps and sends up to 800 Mbps [131]. The NDQ 1300 works on the following 5G NR bands: n48, n77, n78, and n79, making it compatible with many mobile networks worldwide.

3.8.3 ASKEY FMS Box

The ASKEY Fixed Mobile Substitution (FMS) Box, ASKEY RTL 6310, is used to connect VR devices to private 5G networks, as VR devices cannot directly connect to 5G. The FMS box connects to the private 5G network and emits Wi-Fi signals, allowing VR devices to connect seamlessly. It supports Wi-Fi 6, offering WLAN speeds of 2,4 Gbit/s on the 5 GHz band and 1,2 Gbit/s on the 2.4 GHz band, and can accommodate up to 64 users. The FMS box supports multiple bands: n1, n3, n7, n8, n41, n77, n78, and n79, and utilizes the SDX55 modem for optimal performance [132].

3.9 Real-Time Video Streaming and Data Collection

3.9.1 Insta360 One X2

The Insta360 One X2 is a 360-degree camera utilized for real-time video streaming in the 5G-enabled intralogistics application. It captures high-resolution 5.7K dual-lens video and offers a single-lens mode for standard shots. The camera is equipped with four integrated microphones. For the use case, the Insta360 One X2, when paired with a Samsung Galaxy S21, streams high-definition video up to 4K over a 5G network. Additionally, the Insta360 One X2 features internal stitching, which seamlessly combines footage from its dual lenses, providing a smooth and immersive viewing experience without the need for post-processing. The camera is compact, measuring 4.6 x 11.3 x 3 cm and weighing only 149 grams.

3.9.2 Samsung Galaxy S21

The Samsung Galaxy S21 facilitates the real-time streaming of 360-degree video from the Insta360 One X2 to a web server using the RTMP. The device is equipped with both Wi-Fi and 5G modules. The Wi-Fi module, powered by the Qualcomm FastConnect 6900 chipset, ensures high-speed and reliable local network connections essential for initial setup and local data transfer. It supports Wi-Fi 6 (802.11ax) with dual-band operation on 2.4 GHz, 5 GHz, and 6 GHz, achieving peak data rates of up to 3.6 Gbps [133]. The 5G module, powered by the Qualcomm Snapdragon X60 modem, supports both Sub-6 GHz and mmWave bands, allowing the Galaxy S21 to leverage the benefits of the 5G network, such as low latency and high bandwidth, to maintain smooth and uninterrupted live streams and control operations. The X60 modem achieves maximum downlink speeds of up to 7.5 Gbps and uplink speeds of up to 3 Gbps [134].

4 Metrics

The metrics section is divided into three subsections: metrics for network benchmarking, metrics for object detection, and metrics for evaluation of the results.

4.1 Metrics for the Network Benchmarking

In this study, critical network performance metrics are focused on jitter, latency, throughput, and frames per second (FPS). These metrics are essential for evaluating the performance of the 5G network in supporting real-time video streaming and object detection.

Table 4 outlines the requirements for network parameters necessary to achieve the Quality of Experience (QoE) for users in a multi-modal stream. The table shows that haptics demand very low jitter (≤2 ms) and a high update rate (≥1000 Hz) to ensure responsiveness. Video requires low jitter (≤30 ms) and high throughput (2500 - 40000 kbit/s) to maintain quality.

Table 4 Quality of Service Requirements [15]

Medium	Jitter (ms)	Delay (ms)	Packet loss (%)	Update rate (Hz)	Packet size (bytes)	Throughput (kbit/s)
Haptics	≤2	≤50	≤10	≥ 1000	64 - 128	512 - 1024
Video	≤30	≤400	≤1	≥ 30	≤ MTU	2500 - 40000
Audio	≤30	≤150	≤1	≥ 50	160 - 320	64 - 128

Consequently, the study focused on these metrics to effectively assess benchmarking and real-use cases. Below, a detailed explanation of how each metric is calculated is provided.

4.1.1 Jitter

Jitter measures the variation in packet arrival times, which can impact the smoothness of video playback. It is calculated using the WebRTC statistics API, specifically from the inbound-rtp report for video streams. The algorithm steps to obtain it are as follows:

1. Initialize periodic retrieval of network statistics.
2. For each statistical report, check if the report type is inbound-rtp and the media type is video.
3. Extract the jitter value from the report.

4.1.2 Latency

Latency is the time it takes for a data packet to travel from the source to the destination. It is calculated using the candidate-pair report in the WebRTC statistics API. The following algorithmic steps are used to calculate it:

1. Initialize periodic retrieval of network statistics.

2. Calculate the round-trip time (RTT) by summing the CurrentRoundTripTime values from candidate-pair reports and counting the number of such reports.

3. Compute the average RTT and determine latency as half of the average RTT.

Equation 1 provides the mathematical representation of average RTT and latency, where "n" represents the number of reports:

$$RTT_{average} = \frac{\sum_{k=1}^{n} CurrentRoundTripTime}{n} \qquad Latency = \frac{RTT_{average}}{2}$$

Equation 1 RTT and Latency Calculation

4.1.3 Frames Per Second

FPS measures the number of frames displayed per second in the video stream, which impacts the smoothness of video playback and real-time monitoring. It is calculated using the following steps:

1. Initialize a counter for the number of frames processed.

2. For each frame processed, increment the frame counter.

3. After one second, calculate the FPS by dividing the frame count by the time interval (in seconds).

FPS can be represented by the following Equation 2:

$$FPS = \frac{Number\ of\ Frames}{Time\ Interval\ in\ Seconds}$$

Equation 2 FPS Calculation

4.1.4 Throughput

Throughput measures the amount of data successfully transmitted over the network in a given time period. It is calculated using data from the WebRTC statistics API. Since the metric is not provided directly by the API, it is calculated as follows:

1. Initialize periodic retrieval of network statistics.

2. Calculate the difference in bytes received between the current and previous timestamps.

3. Compute the throughput as the number of bits transmitted per second.

Its mathematical representation can be seen in Equation 3.

$$Throughput\ (bps) = \frac{bytesSent \cdot 8}{Time\ Interval\ in\ Seconds}$$

Equation 3 Throughput Calculation

4.2 Metrics for the Object Detection Performance

4.2.1 Intersection over Union

IoU is a key metric used to evaluate the localization accuracy of object detection models. It measures the overlap between the predicted bounding box and the ground truth bounding box.

The IoU is calculated as follows in Equation 4:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

Equation 4 IOU Calculation

Where:

- Area of Overlap is the area where the predicted bounding box and the ground truth bounding box overlap.
- Area of Union is the total area covered by both the predicted bounding box and the ground truth bounding box.

4.2.2 Confidence Score

The "score" refers to the confidence score or probability that the model assigns to a detected object of a particular class. This score is a floating-point value between 0 and 1, where:

- A score close to 1 indicates a high confidence that the detected object belongs to the predicted class.
- A score close to 0 indicates a low confidence in the prediction.

In Equation 5, confidence score s for a detected object is given by the output of the classification layer of the SSD model. It represents the probability $P(c|x)$, where c is the class and x is the detected object.

$$s = P\ (c|x)$$

Equation 5 Confidence Score Calculation

4.3 Metrics for the Classification

This section will detail the metrics used for evaluating the classification performance of object detection models. These metrics are essential for understanding how well the models detect and classify objects in the given dataset. The metrics include Receiver Operating Characteristic (ROC) Curve, Area Under the Curve (AUC), Precision-Recall Curve, Average Precision (AP), F1 Score, Precision, and Recall. Here is an explanation of these metrics:

4.3.1 Receiver Operating Characteristic Curve and Area Under the Curve

The ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The calculation of TPR and FPR is shown in Equation 6:

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN}$$

Equation 6 TPR and FPR Calculation

In Equation 7, AUC represents the degree or measure of separability, with higher AUC indicating better performance.

$$AUC = \int_0^1 TPR(FPR)d(FPR)$$

Equation 7 AUC Calculation

4.3.2 Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positives. It is a measure of the accuracy of the positive predictions. The mathematical representation is shown in Equation 8.

$$Precision = \frac{TP}{TP + FP}$$

Equation 8 Precision Calculation

4.3.3 Recall

Recall is the ratio of correctly predicted positive observations to all observations in the actual class. It is a measure of the ability of the model to find all relevant cases within a dataset. It is calculated as shown in Equation 9.

$$Recall = \frac{TP}{TP + FN}$$

Equation 9 Recall Calculation

4.3.4 F1 Score

The F1 Score is the harmonic mean of precision and recall, providing a single metric that balances both concerns. It is particularly useful when both false positives and false negatives need to be considered. Equation 10 shows its calculation:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Equation 10 F1 Score Calculation

4.3.5 Precision-Recall Curve and Average Precision

The Precision-Recall curve plots precision against recall. AP is the average of precision scores calculated at different threshold levels and provides a single-value summary of the precision-recall curve. It is calculated using

Equation 11, where "n" represents the number of thresholds used in the calculation.

$$AP = \sum_{k=0}^{k=n-1} (Recall_k - Recall_{k-1}) \cdot Precision_k$$

Equation 11 Average Precision Calculation

5 Methodology

The methodology encompasses two distinct sections: a benchmarking setup and a real-world use case demonstration. The benchmarking section is designed to assess the performance of object detection tasks under varying network conditions and model selection, focusing on the comparative analysis of Wi-Fi and private 5G networks. This evaluation is crucial for understanding the impact of network characteristics on real-time video streaming and object detection accuracy.

Meanwhile, the second section showcases the practical implementation of the proposed system in a real-world intralogistics scenario. This demonstration aims to highlight the benefits and potential applications of integrating technologies such as 5G, edge computing, and real-time object detection into industrial processes. By showcasing the system's capabilities in a real-world context, this phase also validates the findings from the benchmarking phase

5.1 Research Design

As previously mentioned, the research is divided into two main phases: benchmarking and real-world implementation. In the benchmarking phase, the objective is to empirically compare the performance of 5G and Wi-Fi networks in supporting real-time object detection tasks. This is done by running object detection tasks on live streaming in a controlled environment under both network types, measuring key performance indicators such as latency, throughput, jitter, and object detection accuracy. In the real-world implementation phase, the objective is to demonstrate the practical benefits and scalability of 5G technology in an industrial intralogistics scenario. This involves implementing the developed system in an industrial setting using a small forklift equipped with real-time object detection to handle logistics tasks and collecting performance data and feedback to assess operational efficiency and scalability.

5.1.1 Alternative Approaches and Rationale for Current Setup

Prior to the final system configuration, several alternative approaches were explored, each with varying degrees of success. These explorations provided valuable insights into the challenges and trade-offs involved in designing a robust and efficient system for real-time object detection.

Initially, HLS was chosen for its widespread support and reliability. HLS breaks streams into small HTTP-based file downloads, enabling adaptive streaming based on network speed. However, its high latency (15 to 30 seconds) was unsuitable for real-time applications. To meet real-time data needs, the system transitioned to WebRTC. WebRTC supports video, voice, and data transmission between peers, enhancing real-time interactivity. Additionally, WebRTC works securely with the HTTPS protocol. Its peer-to-peer architecture also provides scalability and reduces server dependence, ideal for large deployments.

Oven Media Engine was adopted for its optimized transcoders and real-time streaming capabilities. The Insta360 camera was selected for its optimized 360-degree view. Custom software to convert videos from the Insta360 camera was unsuccessful due to issues with stitching videos from its dual cameras, leading to the use of the Insta360 camera with a phone for optimal performance.

Object detection could theoretically be performed on the device itself, but this would significantly limit processing power. Instead, the aim is to achieve object detection centrally, allowing the use of smaller devices. This approach requires excellent connectivity to ensure fast video transmission. Additionally, it is crucial to process the stream without adding latency, ensuring that the stream reaches the end user without delay before performing object detection. This prevents additional processing time related to latency, throughput, and jitter, maintaining optimal performance.

The Samsung Galaxy S21 was chosen over the Raspberry Pi with a Quectel modem for benchmarking. Despite several configurations, the Raspberry Pi with Quectel modem could not achieve real-life downlink speeds beyond 300 Mbps and uplink speeds of 100 Mbps. In contrast, the Galaxy S21 demonstrated superior performance with approximately 900 Mbps downlink and 200 Mbps uplink speeds.

The bandwidth demand for lossless transmission proved to be too high for stable streaming of full HD videos. Using lossless H.264 was found to be computationally too expensive, especially on devices such as the Raspberry Pi or Samsung Galaxy S21. This codec's high computational requirements for encoding and decoding lossless video strained the processing capabilities of these devices. Consequently, the decision was made to use the H.264 codec. Moreover, it is uncommon in the industry to use fully lossless transmission due to these high bandwidth and computational demands.

5.2 Experimental Setup

5.2.1 Experimental Setup for the Benchmarking

The system is designed to address real-world scenarios where object detection inference is computationally demanding for edge devices. By separating the video capture and inference processes, resource-constrained devices are enabled for video acquisition, while object detection inference is performed on the user interface device. This architecture is particularly relevant for applications requiring immediate processing to reduce latency and improve response times. The benchmarking experiments are primarily conducted within the private 5G network at the O2 Tower.

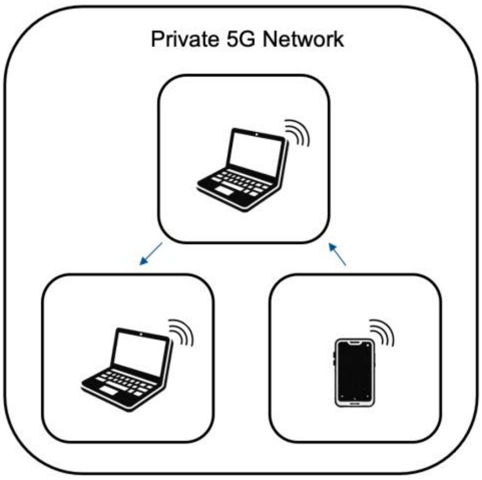

Figure 14 Structure of the Benchmarking Experiment

Figure 14 shows the overall structure of the benchmarking experiment, which includes a Galaxy S21 and two computers. The arrows indicate the data transmission between devices. The data flow starts from the Galaxy S21 to the main server and then proceeds from the main server to the user interface. The benchmarking results are recorded at the user interface. Further details are explained in 5.2.1.2. The details of the devices are as follows:

- **Streaming Device:** Samsung Galaxy S21 smartphone with Qualcomm Snapdragon X60 modem

- **Main Server:** ASUS UX490UA computer (detailed specifications in Table 5) with ASKEY 5G dongle, running an OME Docker container for media streaming and web server for the web application.

- **User Interface:** MSI Vector GP68 HX 12V (detailed specifications in Table 5) with ASKEY 5G dongle, providing a web-based interface for viewing video visualization, model selection, result export, and object detection model execution (YOLOv5s, COCO-SSD).

Table 5 Computer Specifications

Computer	Processor	Graphics	Memory	Operating System	Wifi Module
ASUS UX490UA	Intel Core i5-7200U	Intel HD Graphics 620	8GB DDR3	Ubuntu 22.04	Intel Wireless-AC 8260
MSI Vector GP68HX 12VH	Intel Core i9-12900HX	NVIDIA GeForce RTX 4080	16GB DDR4	Windows 11	Intel Wi-Fi 6E AX211

5.2.1.1 Benchmark Dataset Preparation

A custom benchmark dataset was constructed from the YT-BB dataset to evaluate the system performance. To achieve this, videos with 1080p quality were filtered and merged. A total of 232 videos were combined to create a single video lasting 6 hours and 23 minutes.

Specifically, the following steps were taken:

1. The validation set videos were extracted from the YT-BB dataset using a Python script.

2. Videos were trimmed to the timeframes annotated in the YT-BB validation set.

3. Videos were concatenated, with 40-second black screens inserted at the beginning and 20-second black screens between videos to facilitate scene detection.

5.2.1.2 Video Streaming and Object Detection Pipeline

The following steps describe the video streaming and object detection pipeline used in the benchmarking experiment.

1. **Streaming:** Samsung Galaxy S21 streams video data to the OME server using FFmpeg over RTMP on port 1935.

2. **WebRTC Delivery:** On the main server, OME receives the stream on port 1935 and configures WebRTC streaming on port 3333 for the user interface.

3. **Object Detection:** On the user interface, the selected object detection model processes the video stream upon user initiation.

4. **Data Collection:** Detection results and network statistics are continuously logged in a CSV file.

5. **Post-Processing:** The CSV data undergoes post-processing to align timestamps with ground truth labels from the YT-BB dataset. Considering video transitions, a custom algorithm matches predicted bounding boxes to ground truth.

For the YT-BB dataset, post-processing converts the bounding boxes to use the top-left corner as the origin and pixel coordinates for dimensions. This ensures that both the object detection output and the YT-BB ground truth bounding boxes are consistently represented, allowing for direct comparison during evaluation.

5.2.2 Web-Based Control System

The web-based control system has been developed to facilitate real-time interaction with a forklift in a 5G environment. The system integrates several technologies, including Flask for the backend, HTML and JavaScript for the frontend, Docker for containerization, and OME for media streaming. The application comprises multiple functionalities, such as WebRTC streaming, live network benchmarking, HLS streaming (deprecated due to high streaming latency), forklift control integrated with 360-degree streaming, and object detection benchmarking. These functionalities are organized into separate pages accessible through the web application interface.

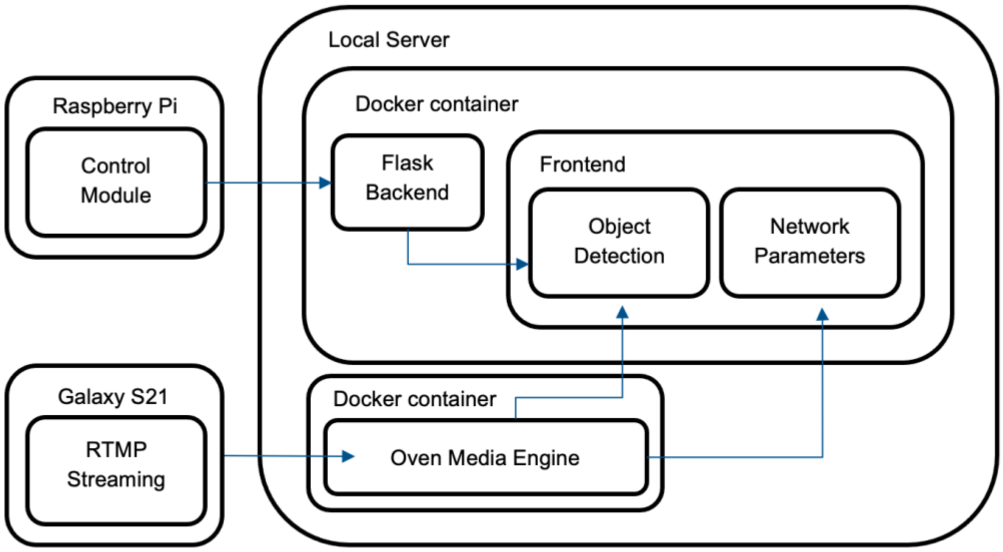

Figure 15 Software Architecture

The general architecture of the software is depicted in Figure 15. The system utilizes several key software and technologies. Flask serves as the backend framework, chosen for its simplicity and flexibility in handling HTTP requests and managing API endpoints that control forklift operations and stream video feeds. The frontend interface is built using HTML, CSS, and JavaScript. HTML displays the control interface and video stream. JavaScript enables dynamic interactions and data fetching from Flask backend APIs. Docker is employed for containerization, ensuring consistency across different computing environments and simplifying deployment and scalability. Docker containers encapsulate the Flask app, OME, and other services. OME delivers essential real-time streaming capabilities and performance metrics for both the network and streaming. However, the primary source for network performance metrics is the WebRTC API.

In parallel, the Raspberry Pi with the control module receives control signals from the Flask backend, which originate from the user interface. These signals are converted into commands for motor drivers via the GPIO connection of the Raspberry Pi, which then controls the motors of the forklift. The streaming workflow begins with a camera on the forklift capturing video footage, which is sent to the OME on the main server via port 1935. OME configures the video for WebRTC delivery on port 3333. A Flask application manages HTTP requests and API endpoints, interacting with a web interface. Users control the forklift and view the video stream through this interface. The Flask backend processes control commands and communicates with the forklift in real time.

5.2.3 Private 5G Network

The private 5G network is a critical component of this experimental setup, providing the necessary connectivity for high throughput, low-latency communication required for real-time video streaming and remote control operations. It supports both licensed and shared spectrum and has a high device-density capacity. Ericsson is the company that provides the infrastructure for private 5G networks.

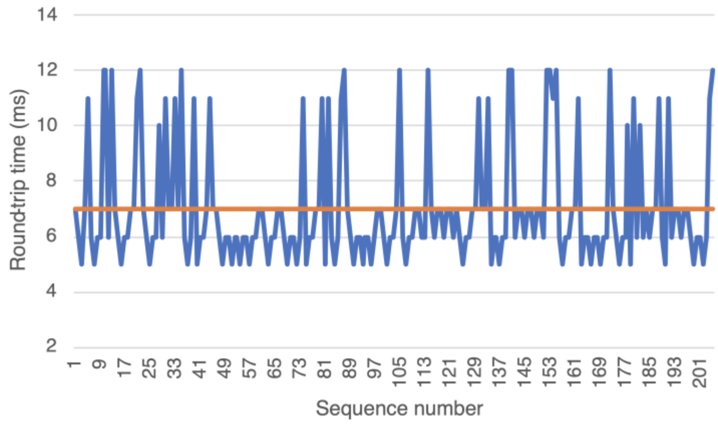

Figure 16 5G System-Latency Test Results from Ericsson [135]

System-latency lab tests conducted with Ericsson private 5G measured the round-trip time between a single device and the network controller. The average round-trip time observed was below 10 ms for 5G and below 20 ms for 4G in an unloaded system. These tests were performed using the Radio Dot 4479 in Time Division Duplex (TDD) mode 0 [135]. The results are shown in Figure 16.

5.2.3.1 Network Architecture

The private 5G network structure illustrated in Figure 17 connects the TUM testing hall and the O2 Tower to the core network and cloud servers. In this architecture, devices within the TUM testing hall communicate with the core network via high-speed fiber optic backhaul links. Similarly, the O2 Tower, the main control station, connects to the core network using high-speed fiber optic backhaul. Both locations employ advanced features such as network slicing and QoS management to optimize network performance for various applications. Here is an overview of the components:

- **Core Network:** The private 5G network is built around a dedicated 5G core, which manages all network functions, including authentication, session management, and data routing. This core network leverages cloud-native architecture, ensuring scalability and flexibility. Key components include the Access and Mobility Management Function (AMF), Session Management Function (SMF), User Plane Function (UPF), and the Unified Data Management (UDM) system.

- **Radio Access Network:** The RAN consists of multiple small cells strategically placed to ensure seamless coverage and capacity. These small cells, such as Radio Dots and Micro Radios, communicate with the 5G core via high-speed backhaul links.

- **Backhaul:** The backhaul network utilizes fiber optic links to connect the RAN to the 5G core, ensuring minimal latency and high data throughput. The network supports advanced features like network slicing and QoS management to optimize performance for various industrial applications.

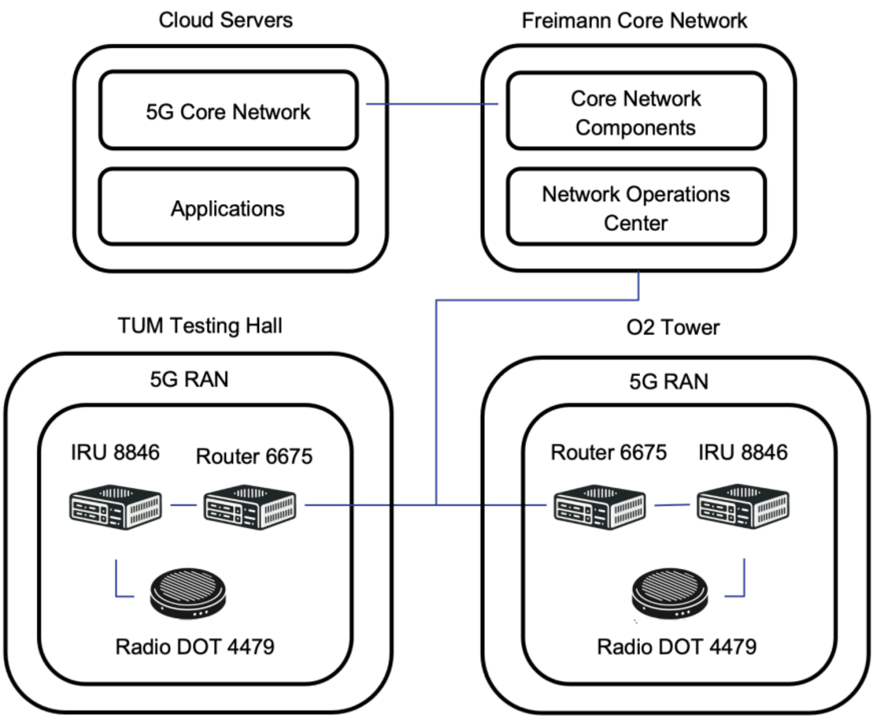

Figure 17 Network Structure of Private 5G

Further key hardware components include the Router 6675, which manages data traffic and ensures secure, reliable connections between network components. The IRU 8846 is a centralized radio unit that connects multiple Radio Dots to enhance signal coverage and capacity. The Radio Dot 4479 provides enhanced indoor coverage, ensuring strong signal strength and high data throughput within buildings.

Ericsson also offers cloud management capabilities for the private 5G network to remote control and monitor network infrastructure. The network management portal provides a user-friendly interface for installation, configuration, monitoring, and maintenance tasks. Additionally, API integration allows for integration with customized cloud applications and other management functions.

5.2.4 Data Collection Procedures

The data collection procedures are divided into two sets. The first set focuses on network performance metrics, which include methods for fetching metrics such as bitrate, frame rate, jitter, throughput, latency, and packet loss. The second set refers to object detection data collection, capturing details like detection confidence score, FPS, object classes, and bounding box coordinates and dimensions. Although divided for explanation purposes, both sets of data are accumulated simultaneously.

5.2.4.1 Network Performance Metrics Collection

Two primary methods are used to fetch network performance metrics: the getStats function of WebRTC and the OvenMediaEngine API. Additionally, time counters are utilized in the code base. Here's an overview:

- **Get Stats Function of WebRTC:** The getStats function in WebRTC provides statistics regarding the performance of media streams. This function collects various metrics, both directly and indirectly, including bitrate, frame rate, jitter, throughput, latency, and packet loss. Algorithm 1 is an example pseudocode for utilizing getStats in a WebRTC application:

```
Algorithm 1: WebRTC Stats Gathering
  Input: RTCPeerConnection pc
  Output: Round-Trip Time (RTT), Video Bitrate
pc.oniceconnectionstatechange ← OnIceConnectionStateChange
OnIceConnectionStateChange() if pc.iceConnectionState == 'connected' then
  stats ← pc.getStats()
  foreach report in stats do
    if report.type == 'candidate-pair' ∧ report.state == 'succeeded' then
      Record RTT: report.currentRoundTripTime
    end
    if report.type == 'outbound-rtp' ∧ report.mediaType == 'video' then
      Record Bitrate: report.bytesSent / (report.timestamp - report.startTime) * 8
    end
  end
end
```

Algorithm 1: WebRTC Stats Gathering

- **OvenMediaEngine API:** The API also provides codec information and stream resolution. Additionally, it enables remote operation of the streaming server. Scripts integrated within the Flask backend periodically invoke this API to pull data during live streaming sessions.

5.2.5 Object Detection and Benchmarking Pipeline

The object detection process involves loading a specific model, capturing video frames, and using the loaded model to detect objects within these frames. Algorithm 2 illustrates the process employed for both object detection and benchmarking. The following is an explanation of each step involved in the process.

1. **Model Loading**: Load the selected model (e.g., COCO-SSD, YOLO) based on user input.

2. **Video Stream Initialization**: Initialize the video stream using WebRTC.

3. **Object Detection Loop**: Video frames are first captured and then preprocessed to fit the model's requirements. The preprocessed frames are fed into the model to perform object detection. The raw detection data is then postprocessed to format bounding boxes, class labels, and confidence scores for easy interpretation.

4. **Data Logging**: Log the detection results, including timestamps, network type, bounding boxes, class labels, confidence scores, jitter, latency, throughput, and FPS.

5. **Export Results**: Export the results through the user interface.

Algorithm 2: Real-Time Object Detection

```
Input: Pre-trained object detection model and video frames
Output: Continuous detection and log of detection results
Initialize video stream
Load object detection model
while video stream is active do
    Capture video frame
    Preprocess frame (resize, normalize, etc.)
      Apply object detection model to frame
    foreach detected object do
        Extract object information (class, location, confidence)
    end
    Log the detection results along with corresponding network parameters
end
```

Algorithm 2: Real-Time Object Detection

After the object detection and network performance measurements, the collected data logs require post-processing to facilitate accurate analysis and comparison. The details of the post-processing steps include:

1. **Bounding Box Conversion**: The bounding box coordinates, initially in relative values, are converted to pixel values based on the frame dimensions.

2. **Assigning Video IDs to Predictions**: Video IDs are assigned to the predictions based on timestamp resets, indicating new video segments.

3. **Filtering and Adjusting Predictions:** The initial data is filtered to retain only relevant video IDs. Timestamps in the prediction data are adjusted relative to the first occurrence timestamps, and predictions are filtered based on the time ranges for each video ID.

4. **Matching and Calculating IOU**: Predictions are matched with the ground truth data, and the IOU metric is calculated to evaluate object detection accuracy. The process is illustrated in Algorithm 3.

```
Algorithm 3: Match and Calculate IOU
Input: CSV files for predictions and ground truths
Output: Matched and filtered predictions with IOU and confidence scores
data1 ← read rows from ground truth; data2 ← read rows from predictions;
foreach row1 in data1 do
    Initialize best IOU ← 0, best row2 ← None, highest score row2 ← None
    foreach row2 in data2 do
        if timestamps of row1 and row2 are within 500 ms then
            Calculate IOU between bbox1 and bbox2
            if IOU > best IOU then
                best IOU ← IOU
                best row2 ← row2
            end
            Update highest score row2 if score is higher
        end
    end
    if best row2 exists then
        Write row1 and best row2 with IOU to output;
    end
    else if highest score row2 exists then
        Write row1 and highest score row2 with IOU ← 0 to output;
    end
    else
        Write row1 with placeholders to output;
    end
end
```

Algorithm 3: Match and Calculate IOU

After the post-processing step, the output appears as shown in Table 6. The latency and jitter values are presented in seconds, while throughput values are in bits per second (bps). The prediction bounding box coordinates are measured in pixels. This table represents only a slice of the data and includes additional columns not shown here.

Table 6 Final Output After Post-Processing

timestamp	Mode	Prediction BBox_x	Prediction BBox_y	Prediction Class	Score	Latency	Jitter	Model	Throughput
118000	Private 5G	39.48	4.7	train	0.717	0.009	0.014	yolov5s	1370147.99
14000	Private 5G	0.77	270.29	airplane	0.98	0.015	0.017	yolov5s	1399402.09
26000	Private 5G	436.22	474.16	airplane	0.983	0.012	0.014	yolov9c	1440000.0
70000	Private 5G	2.22	291.03	person	0.768	0.013	0.014	yolov8s	1843188.81
188000	Wi-Fi	684.96	394.29	bus	0.932	0.004	0.005	yolov5s	1246616.29
340000	Wi-Fi	1302.99	339.05	car	0.764	0.002	0.009	cocoSsd	1305849.67
344000	Wi-Fi	935.92	393.93	zebra	0.847	0.002	0.012	yolov9c	1573975.22
28000	Wi-Fi	770.69	429.24	person	0.545	0.003	0.009	yolov8x	1386415.71

5.2.6 Experimental Setup for the Real-World Use Case Demonstration

A real-world use case was implemented to demonstrate the practical application and advantages of the proposed system. This involved the remote operation of a small forklift within the TUM testing hall, controlled from the O2 Tower.

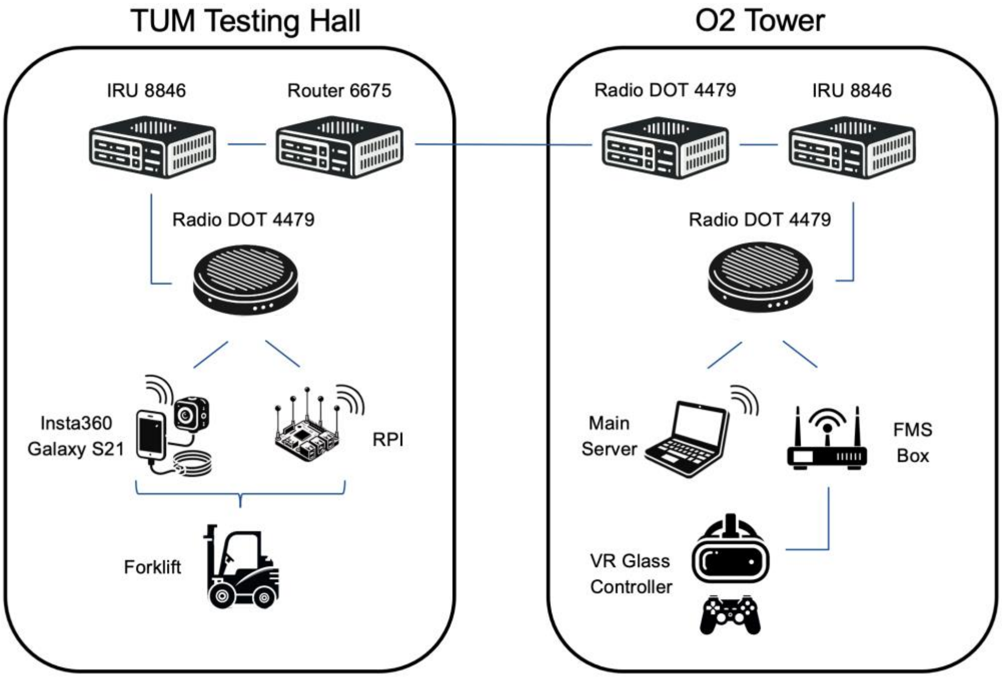

Figure 18 Overall Structure of the Real-World Usecase

Two private 5G networks were established at the O2 Tower and the TUM testing hall. OME was deployed on the main server to manage real-time video streaming from the forklift's camera to the operator's web interface. The forklift was equipped with a Raspberry Pi control unit and an Insta360 camera connected to a mobile phone for a 360-degree view. At the O2 Tower, the operator used a web interface, an Xbox One controller, and a VR device for control, with additional flexibility provided by mobile phone access to the control interface. The overall structure of the real-world use case is illustrated in Figure 18. The lines represent bidirectional data transmission between components. The complete structure of the private 5G networks is illustrated in Figure 17.

From the user's perspective, the workflow begins by accessing the control interface via a web browser at the O2 Tower. The operator uses an Xbox One controller and a VR device to navigate and operate the forklift, viewing the live 360-degree video feed from the Insta360 camera mounted on the forklift. Commands are sent in real-time through the private 5G network to the Raspberry Pi control unit on the forklift. For added flexibility, the operator can also use a mobile phone to control the forklift through the same web interface.

6 Results and Discussion

This section presents the results of the experiments and analyses performed to evaluate the performance and reliability of the 5G-enabled system. The overall goal of these experiments is to empirically assess the advantages of using 5G technology over traditional Wi-Fi in real-time object detection and video streaming applications within an industrial setting. Furthermore, the experiments aim to review the effect of network parameters and detection models on object detection performance.

The experiments were designed to measure various performance metrics, including latency, jitter, and throughput, as well as the accuracy of object detection models. By systematically comparing these metrics across different network setups, this study aims to provide a clear understanding of the potential benefits and limitations of 5G technology in enhancing the efficiency and reliability of autonomous logistics systems.

During the experiments, the devices were connected to the n78 band for the private 5G network. The Wi-Fi module of the devices was connected to a Wi-Fi access point using the 802.11ac standard. Private 5G signal power measurements are provided in Reference Signal Received Power (RSRP) format, while those for Wi-Fi are in Received Signal Strength Indicator (RSSI) format. For Wi-Fi, the main server had a signal strength of -53 dBm, the user interface device had -40 dBm, and the streaming device had -44 dBm. For private 5G, the main server had a signal strength of -74 dBm, the user interface device had -67 dBm, and the streaming device had -72 dBm. All devices were positioned at the same distance from the Wi-Fi router and the private 5G radio dot. The signal levels for both networks were within a range that provides reliable connectivity, confirming that signal strength did not significantly impact performance.

For the data analysis, the following methods were used: statistical significance tests, Analysis of Variance (ANOVA), Tukey's Honestly Significant Difference (HSD) test, correlation coefficients, multiple linear regression, and standard statistical metrics.

6.1 Descriptive Statistics and Comparative Analysis

In this section, the descriptive statistics of the gathered data used in the experiments are presented. These statistics provide a comprehensive overview of key performance metrics, including latency, jitter, and throughput, as well as the accuracy of the object detection models measured by the IOU. Statistical significance tests and further analyses are also conducted on the data. This analysis enables a detailed comparison between Wi-Fi and private 5G networks, highlighting their respective impacts on the performance of object detection models.

Table 7 and Table 8 below summarize the key statistical measures for the dataset for private 5G and Wi-Fi, including the count, mean, median, standard deviation, minimum, and maximum values for each metric. The metrics include jitter, latency, confidence score, IOU, throughput, packet loss, and inference time. The data consists of 31,920 data points for Wi-Fi and 31,860 data points for private 5G. Excluding the zero values, the minimum throughput is 0.0133 Mbps for private 5G and 0.0117 Mbps for Wi-Fi.

Table 7 Descriptive Statistics of Network Performance Metrics (Wi-Fi)

	Jitter (milliseconds)	Latency (milliseconds)	Score	IOU	Throughput (megabits per second)	Packet Loss	Inference Time (milliseconds)
mean	8.971	3.597	0.742	0.646	1.574	0.0	31.744
std	2.559	2.328	0.157	0.235	0.744	0.0	3.213
min	4.0	1.0	0.5	0.0	0.0	0.0	24.4
25%	7.0	2.5	0.597	0.497	1.273	0.0	29.4
50%	9.0	3.5	0.735	0.685	1.471	0.0	31.3
75%	10.0	4.0	0.884	0.83	1.703	0.0	33.3
max	53.0	160.5	0.999	0.997	12.913	0.0	97.7

* Statistical Terms: count = Number of observations, mean = Average value, std = Standard deviation, min = Minimum value, 25% = 25th percentile, 50% = Median (50th percentile), 75% = 75th percentile, max = Maximum value.

Table 8 Descriptive Statistics of Network Performance Metrics (Private 5G)

	Jitter (milliseconds)	Latency (milliseconds)	Score	IOU	Throughput (megabits per second)	Packet Loss	Inference Time (milliseconds)
mean	15.696	13.46	0.74	0.645	1.582	0.0	31.696
std	2.041	2.309	0.157	0.235	0.783	0.0	3.145
min	10.0	7.5	0.5	0.0	0.0	0.0	24.9
25%	14.0	11.5	0.595	0.495	1.269	0.0	29.5
50%	16.0	13.0	0.732	0.688	1.47	0.0	31.3
75%	17.0	15.0	0.883	0.827	1.709	0.0	33.3
max	27.0	29.5	1.0	0.997	23.458	0.0	103.0

Wi-Fi displays a lower average jitter and latency, with mean values of 8.971 milliseconds and 3.597 milliseconds respectively, but these metrics also show higher maximum values, indicating occasional significant variations. In contrast, private 5G demonstrates higher but more consistent averages, with jitter at 15.696 milliseconds and latency at 13.46 milliseconds, which suggests a more stable performance albeit at higher base levels. The analysis reveals that private 5G displays significantly higher jitter and latency compared to Wi-Fi ($p < 0.001$ for both), indicating that private 5G may introduce more delay in data transmission.

Both networks achieve similar scores and IOU values, with Wi-Fi averaging a score of 0.742 and an IOU of 0.646, while private 5G averages a score of 0.74 and an IOU of 0.645. This indicates that the overall performance in these areas is comparable between the two networks. Furthermore, they also share a similar distribution for IOU values, which can be seen in Figure 19. The analysis indicates that there is no statistically significant difference between private 5G and Wi-Fi networks in terms of IOU for object detection, as evidenced by a p-value greater than 0.05. Throughput, a critical measure of network capacity, shows slightly better results for private 5G, with an average of 1.582 megabits per second (Mbps) and a maximum reaching 23.458 Mbps, compared to Wi-Fi's average of 1.574 Mbps and a maximum of 12.913 Mbps. Both network environments recorded no packet loss, reflecting reliability in data transmission.

Figure 19 illustrates the distribution of IOU values for object detection in private 5G and Wi-Fi modes. Both modes exhibit a right-skewed distribution, indicating that most detections have high IOU values, suggesting a good overlap between predicted and actual bounding boxes. The majority of IOU values for both modes are concentrated in the higher range, peaking around 0.95 for private 5G and 0.90 for Wi-Fi, signifying good average accuracy. While private 5G shows a slightly wider spread of IOU values, indicating more variability in its detection accuracy, both modes perform well. A few outliers with very low IOU values exist in both modes, potentially indicating challenging detection scenarios.

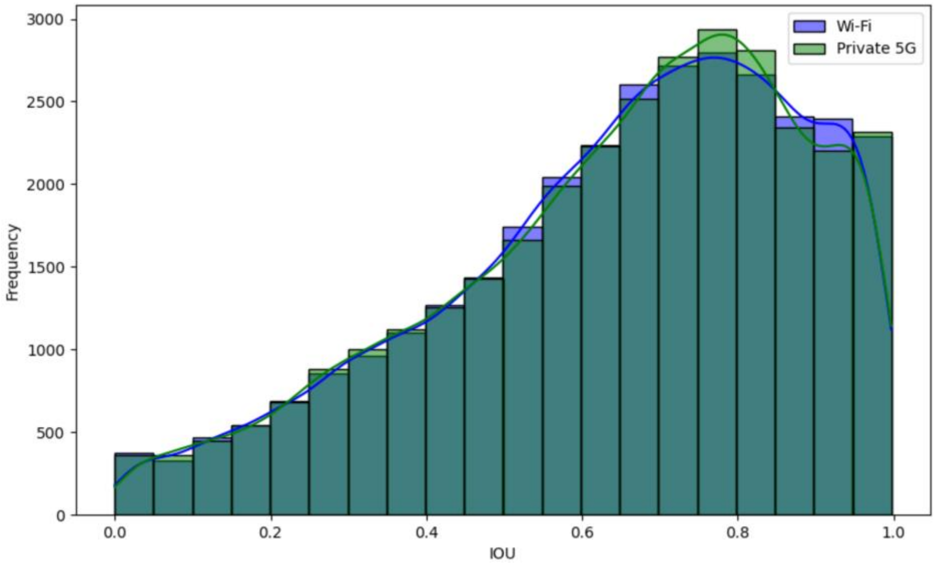

Figure 19 Histogram of IOU Values

The graphs in Figure 20, Figure 21, and Figure 22 illustrate network performance metrics over a 10-minute timeframe for private 5G and Wi-Fi modes. When the entire dataset is focused on, for private 5G, latency is stable at around 13.46 ms, with a standard deviation of 2.309 ms and a range from 7.5 to 29.5 ms. In contrast, Wi-Fi shows lower latency, around 3.597 ms, with a standard deviation of 2.328 ms, but with a much higher maximum value of 160.5 ms. Throughput for private 5G fluctuates between 0.0132 and 23.458 Mbps, with a mean of 1.582 Mbps and a standard deviation of 0.783 Mbps, indicating its capability to handle higher peak data rates. Wi-Fi throughput ranges from 0.0117 to 12.913 Mbps, with a mean of 1.574 Mbps and a standard deviation of 0.744 Mbps. This indicates that private 5G can achieve significantly higher peak throughput, providing better performance for data-intensive applications, while Wi-Fi has lower average latency but can experience significant spikes.

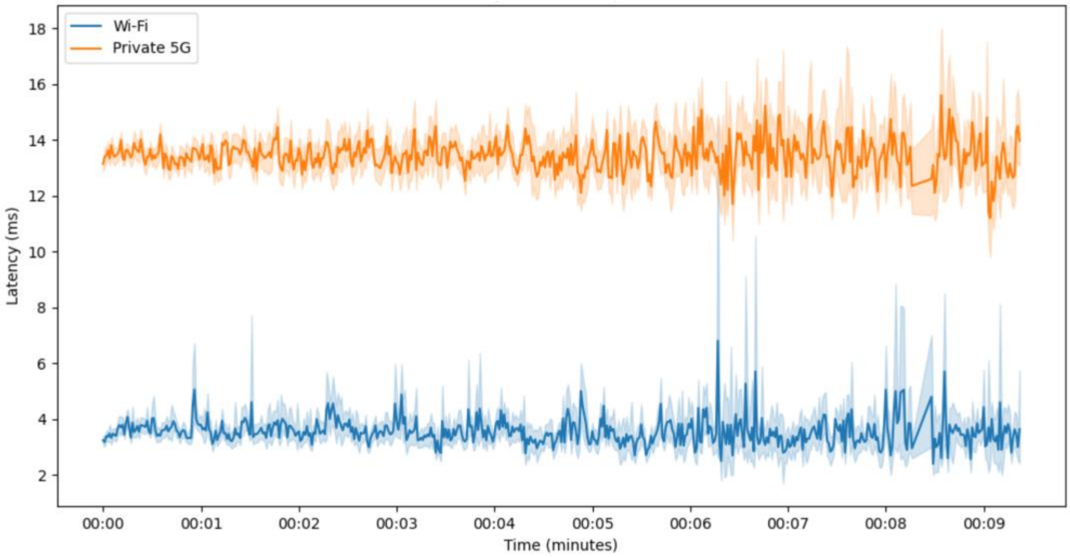

Figure 20 Latency over Time

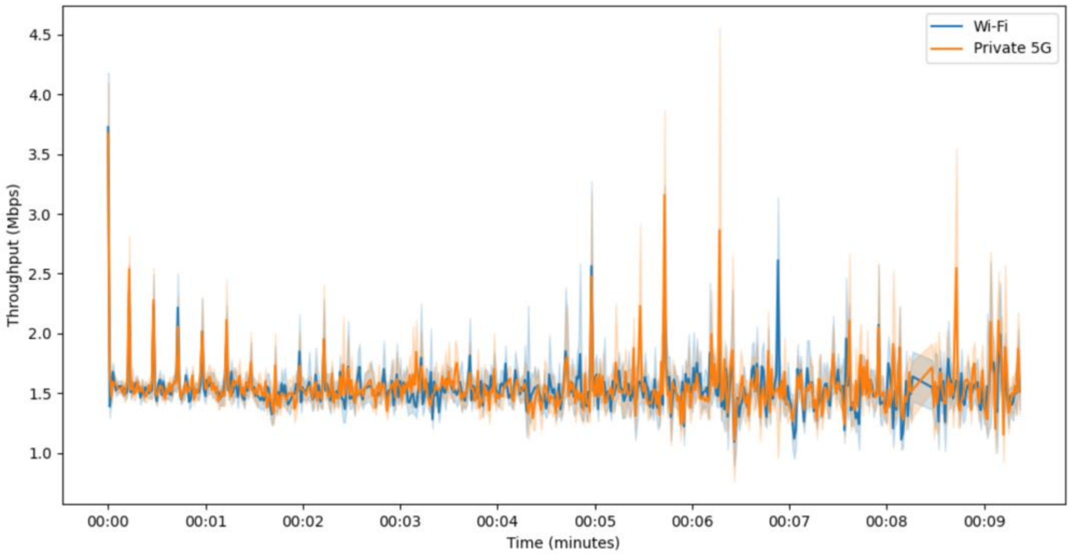

Figure 21 Throughput over Time

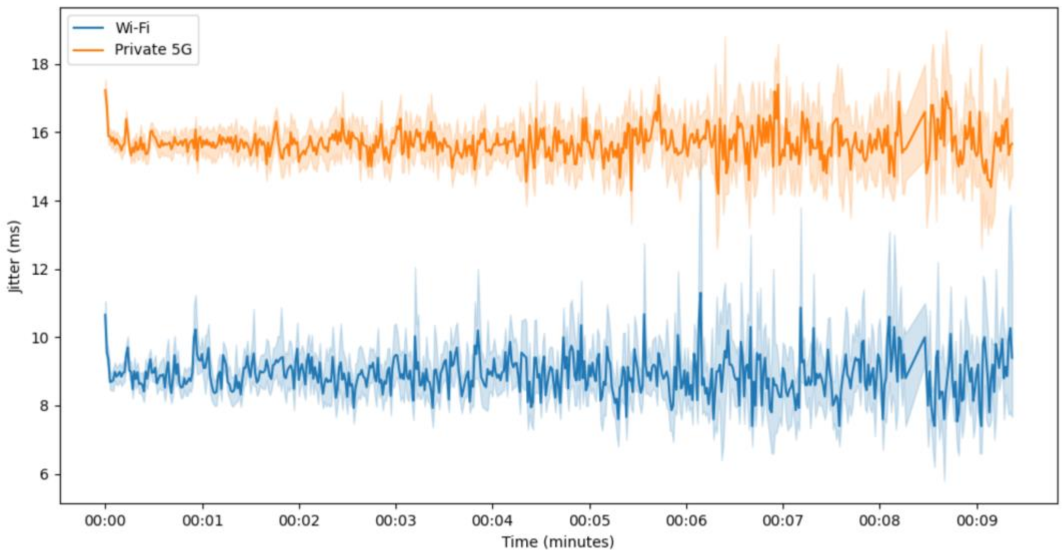

Figure 22 Jitter over Time

Figure 23 below shows the average IOU values for each detected class for Wi-Fi and private 5G.

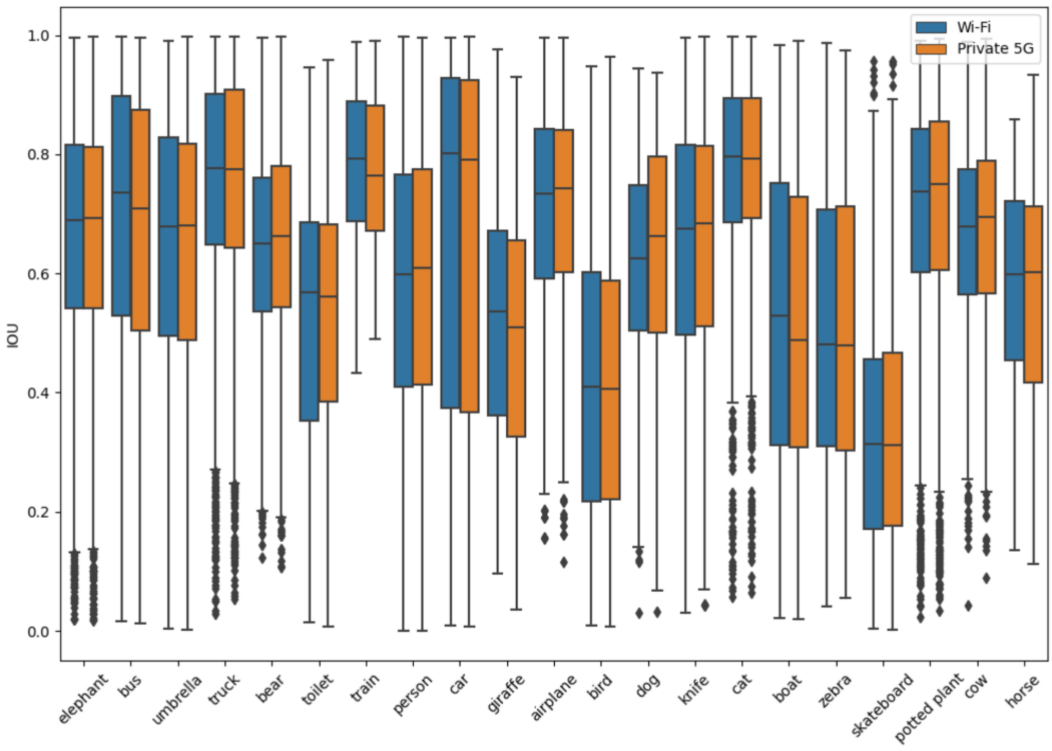

Figure 23 IOU by Object Class and Network Mode

The data contains information about 21 unique object classes, with 'bus' being the most frequent (9900 instances) and 'horse' being the least frequent (190 instances). The ANOVA test reveals a significant effect of object class on IOU (F-statistic = 545.032, p-value < 0.001) and on confidence score (F-statistic = 77.863, p-value < 0.001). Tukey's HSD test identifies numerous significant pairwise comparisons between classes, indicating that the accuracy of object detection varies substantially depending on the type of object (see Appendix 2). This implies that the type of object significantly influences both the IOU and confidence score of object detection. Certain object classes are more challenging to detect than others. For example, the model performs worse at detecting 'bird' objects compared to 'airplane' objects, as suggested by the negative mean difference and the rejection of the null hypothesis (see Appendix 2).

Both networks perform well in detecting certain objects, such as trains, cats, and trucks, with high IOU values (Wi-Fi: 0.781, 0.77, 0.753; Private 5G: 0.772, 0.774, 0.751). These results indicate strong and consistent detection capabilities for these classes across both network types. In contrast, both networks struggle with categories like birds, skateboards, and giraffes, which have lower IOU values, indicating these objects are more challenging to detect. For example, the IOU for birds is 0.423 on Wi-Fi and 0.421 on private 5G; and for skateboards, it is 0.329 on Wi-Fi and 0.334 on private 5G. For the justification, it can be stated that objects with higher IOU values generally possess distinct shapes, features, and sizes, which enhance detection accuracy. In contrast, objects with lower IOU values tend to show a wide range of sizes and patterns and are detected in more variable environments, contributing to the observed reduction in detection performance.

Figure 24 shows the distribution of object sizes for different network modes, along with their corresponding average IOU values. The data demonstrate that as the size of the object increases, the mean IOU values also rise. This observation is further substantiated by the strong positive correlation (0.706) between bounding box area and IOU, with a p-value < 0.001. These findings indicate that larger objects tend to achieve higher IOU values, suggesting they are easier to detect accurately. Overall, the performance differences across network types for various object sizes are minimal.

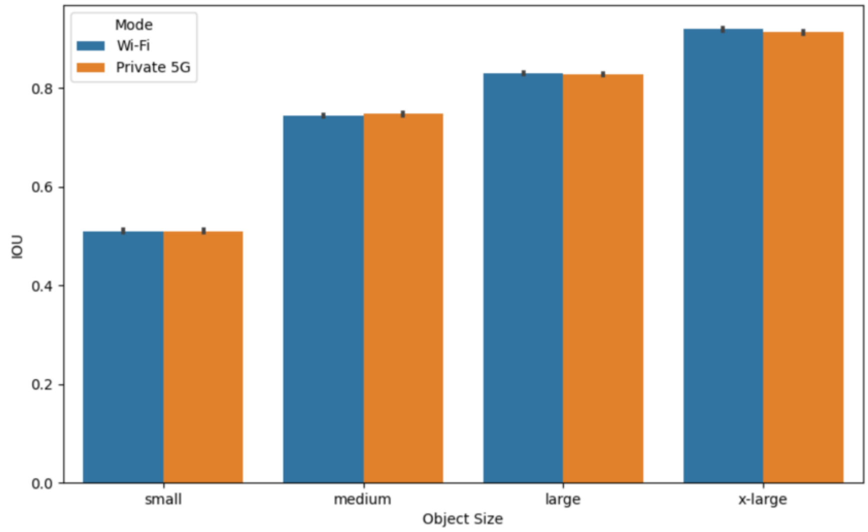

Figure 24 IOU by Object Size and Network Mode

Figure 25 below illustrates the distribution of bounding box centers detected by the object detection models. The plot demonstrates an even distribution, with a notable concentration along the edges and corners, which is expected for dynamically moving video streams.

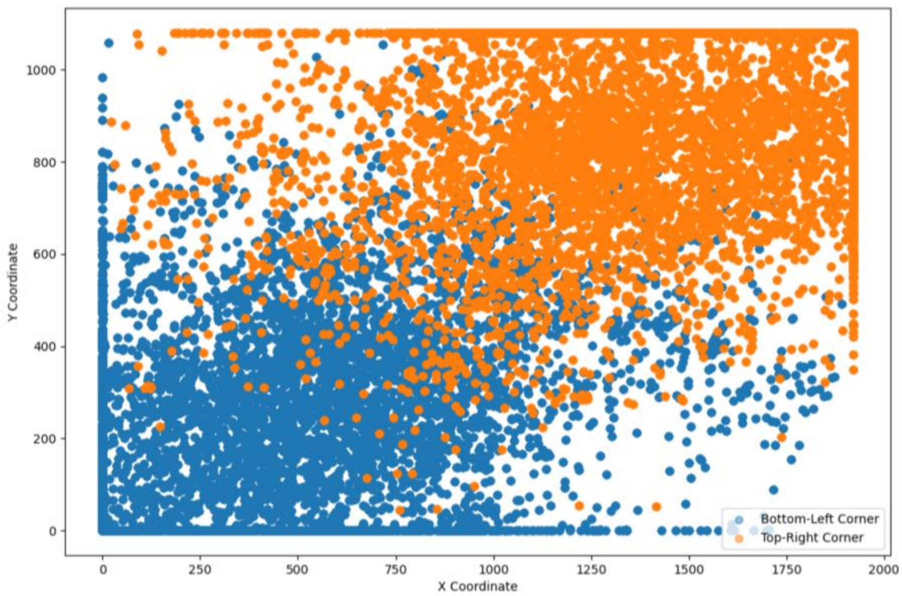

Figure 25 Scatter Plot of Bounding Boxes

6.2 Jitter, Latency, Throughput on IOU and Confidence Score

Table 9 analyzes the effect of latency, jitter, and throughput on the IOU using multiple linear regression. The table presents the output values. The regression model evaluates the impact of the independent variables (throughput, jitter, latency) on IOU. The coefficient indicates the change in IOU for a one-unit change in an independent variable. A p-value less than 0.05 suggests statistical significance. The 95% confidence interval indicates the range within which the true parameter value is likely to fall. The constant (const) represents the baseline IOU when all independent variables are zero.

Table 9 Regression Model Coefficients, P-Values, and 95% Confidence Intervals for IOU

Metric	Coefficient	P-Value	Confidence Interval (95%)
const	0.6411	0.0	(0.634, 0.6483)
Throughput	0.0	0.0025	(0.0, 0.0)
Jitter	-0.4241	0.2566	(-1.1568, 0.3086)
Latency	0.4194	0.1338	(-0.1289, 0.9677)

The analysis indicates that jitter does not have a significant effect on IOU, with a p-value of 0.2566. Throughput has a significant positive effect on IOU, with a p-value of 0.0025, indicating that higher throughput correlates with higher IOU values. Latency does not significantly affect IOU, as evidenced by a p-value of 0.1338. Similar results can be observed in the pair plots shown in Figure 26.

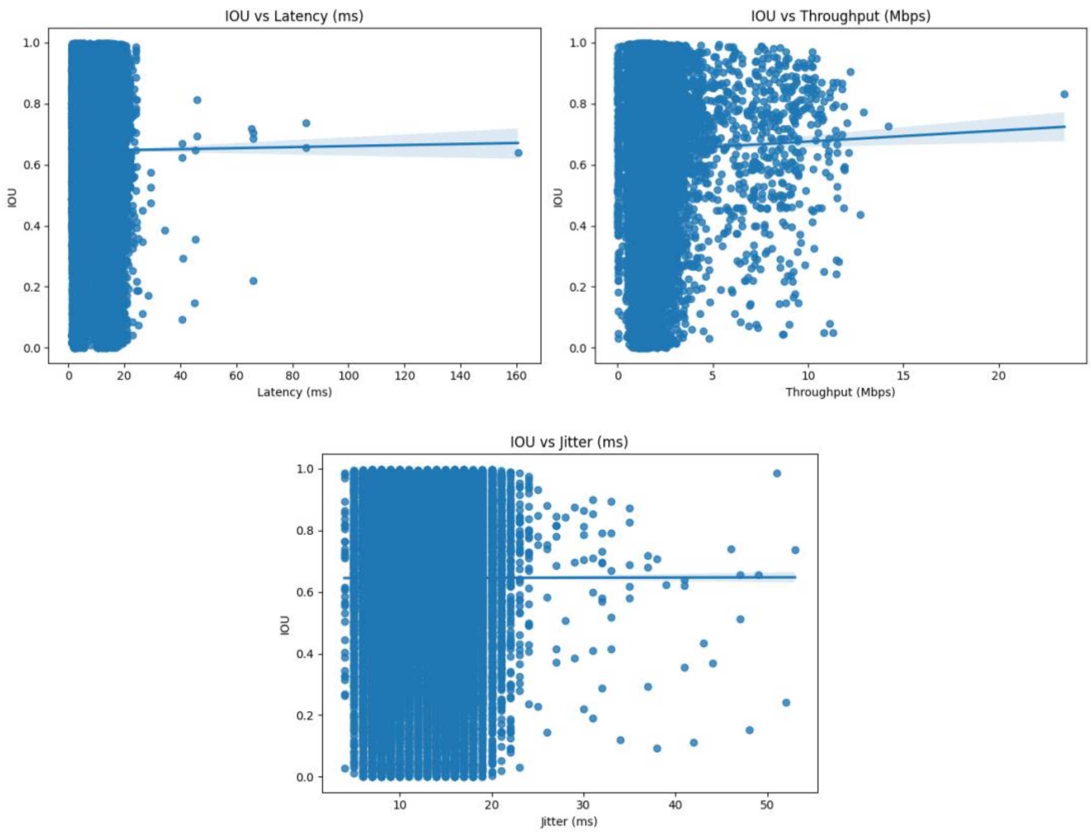

Figure 26 Pair Plot for Throughput, Latency, Jitter versus IOU

Figure 26 above illustrates the relationships between throughput, latency, jitter, and IOU. The blue line represents the regression line. The pair plot of throughput versus IOU shows a positive correlation, suggesting that higher throughput is associated with increased IOU values, indicating better model performance with improved network throughput. On the other hand, the plots for jitter and latency versus IOU display more scattered data points, indicating weaker or less consistent correlations. This suggests that changes in jitter and latency do not clearly or strongly impact the IOU values, making their effects on the model's performance less predictable.

Table 10 Regression Model Coefficients, P-Values, and 95% Confidence Intervals for Score

Metric	Coefficient	P-Value	Confidence Interval (95%)
const	0.737	0.0	(0.7322, 0.7418)
Throughput	-0.0	0.7625	(-0.0, 0.0)
Jitter	0.7982	0.0014	(0.3086, 1.2878)
Latency	-0.6467	0.0005	(-1.0131, -0.2803)

Table 10 analyzes the effect of latency, jitter, and throughput on the confidence score using multiple linear regression. The regression analysis results show that jitter has a statistically significant positive effect on confidence score, with a coefficient of 0.7982 (p-value = 0.0014). Latency has a statistically significant negative effect on confidence score, with a coefficient of -0.6467 (p-value = 0.0005). Throughput, however, does not significantly affect Score, with a near-zero coefficient and a high p-value (0.763).

Overall, while some individual effects are statistically significant, the cumulative impact of jitter, throughput, and latency on IOU and confidence score appears to be limited in this data. Further research incorporating additional variables may be needed to better understand the factors influencing object detection performance.

6.3 Effect of Different Models

This section examines the impact of various models on object detection performance, with a focus on IOU, confidence score, and inference time, in the context of Wi-Fi and private 5G networks.

The ANOVA test reveals a statistically significant difference in mean IOU among the models (F-statistic = 2.833, p-value = 0.023) and no significant difference in mean confidence score among the models (F-statistic = 2.374, p-value = 0.05). These results indicate that the choice of model can significantly impact IOU values in object detection. Tukey's HSD test compares two pairwise models for IOU and score. Tukey's HSD test reveals that the only significant difference in IOU is between the YOLOv5s and YOLOv8x models (p = 0.0125). The YOLOv5s model has a significantly higher IOU than the YOLOv8x model (see Appendix 2). While there is a significant difference in IOU between YOLOv5s and YOLOv8x, there are no significant differences in confidence score between any of the models. This suggests that while the choice of the model may influence the localization accuracy (IOU), it does not substantially affect the confidence score of the detection.

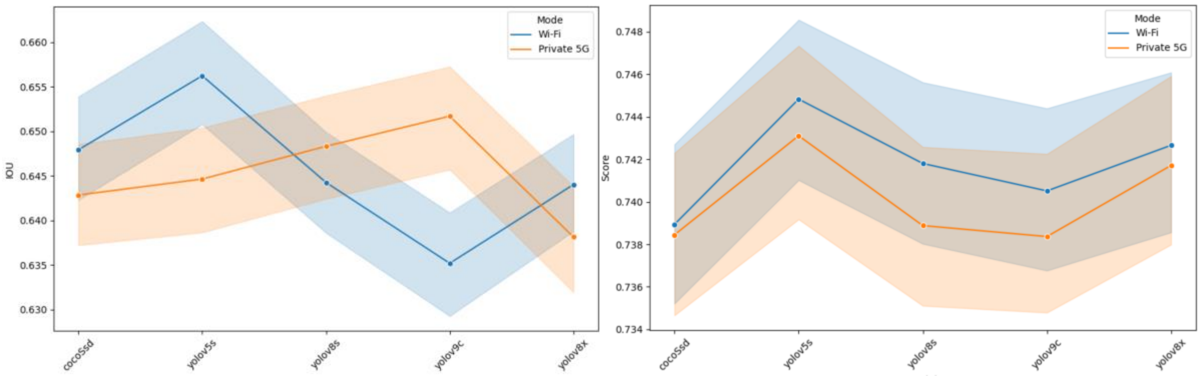

Figure 27 IOU and Score by Network Mode and Model

Figure 27 above illustrates the performance of various object detection models with respect to IOU and confidence score, arranged from least complex to most complex from left to right. YOLOv5s consistently outperforms other models, achieving the highest average IOU in Wi-Fi (0.656), which indicates better accuracy. The standard deviations for IOU are relatively small, reflecting consistent performance within each model and mode. All models exhibit comparable average scores across both network modes, with YOLOv5s slightly leading in Wi-Fi (0.745) and private 5G (0.743). The small standard deviations for scores suggest consistent confidence levels across models and modes. The model with the highest IOU for private 5G is YOLOv9c. Although YOLOv9c is the most accurate model under private 5G, YOLOv5s maintains high accuracy in both network modes, making it a versatile and reliable choice for object detection tasks.

Figure 28 below depicts the differences in inference time for different object detection models and comparisons between private 5G and Wi-Fi networks.

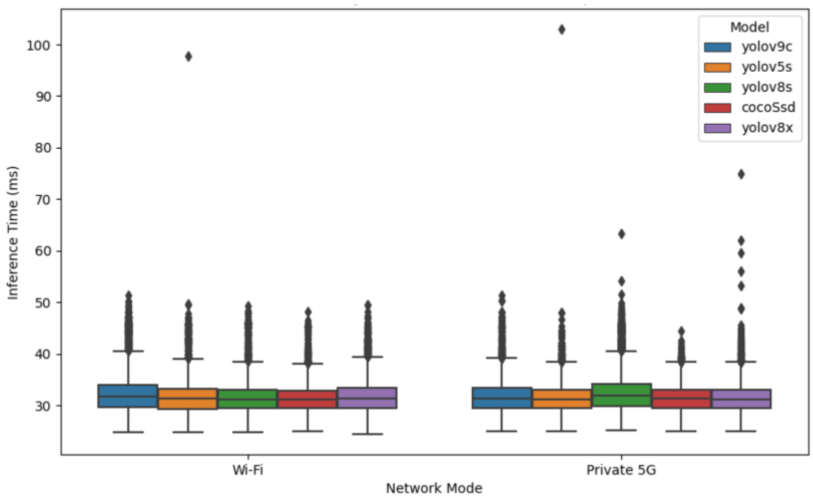

Figure 28 Inference Time by Network Mode and Model

The analysis shows a weak negative correlation between inference time and latency (-0.007, $p < 0.093$), a very weak positive correlation with throughput (0.001, $p = 0.822$), and a very weak positive correlation with jitter (0.002, $p = 0.535$). These results indicate that latency, throughput, and jitter have minimal influence on inference time, with other factors such as device performance likely playing a more significant role. Specifically, the mean inference times for the models are cocoSsd at 31.372 ms, YOLOv5s at 31.536 ms, YOLOv8x at 31.554 ms, YOLOv9c at 31.995 ms, and YOLOv8s at 32.141 ms.

Under both private 5G and Wi-Fi conditions, the five object detection models (cocoSsd, YOLOv5s, YOLOv8s, YOLOv8x, YOLOv9c) demonstrated comparable performance, with mean IOU values ranging from 0.641 to 0.650. While slight variations in inference times were observed across models and network modes, these differences were not substantial.

Table 11 Performance Metrics of Various Object Detection Models

Model	ROC AUC	Average Precision	F1 Score	Mean IOU
yolov9c	0.601	0.806	0.852	0.643
yolov5s	0.619	0.82	0.858	0.65
yolov8s	0.608	0.811	0.857	0.646
cocoSsd	0.607	0.816	0.856	0.645
yolov8x	0.612	0.806	0.849	0.641

Table 11 above shows the performance metrics for the models YOLOv9c, YOLOv5s, YOLOv8s, cocoSsd, and YOLOv8x. YOLOv5s consistently outperforms the others. It has the highest ROC AUC (0.619), indicating the best class distinction, the highest average precision (0.82), suggesting better precision-recall trade-off, the highest F1 score (0.858), reflecting the best balance between precision and recall, and the highest mean IOU (0.65), indicating better localization accuracy. In contrast, YOLOv9c and YOLOv8x exhibit lower performance across these metrics, with YOLOv9c having the lowest ROC AUC (0.601) and average precision (0.806), and YOLOv8x having the lowest F1 score (0.849) and mean IOU (0.641). Overall, YOLOv5s demonstrates the most effective performance for object detection tasks in this evaluation. Figure 29 illustrates the average precision curve.

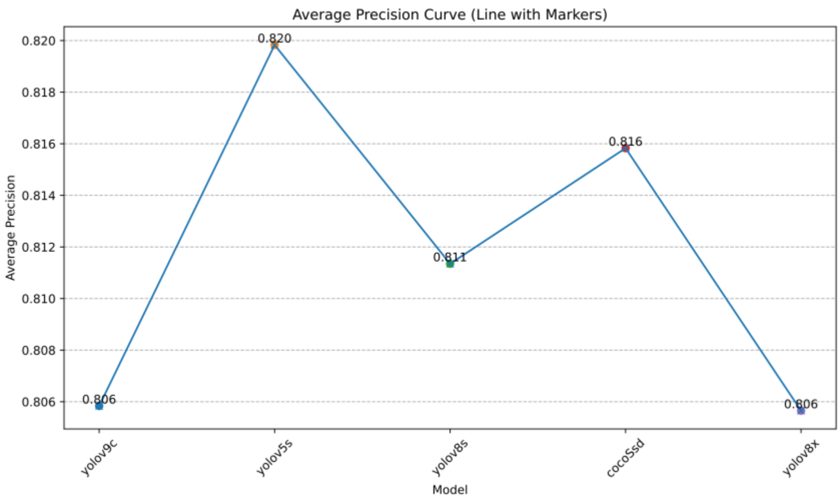

Figure 29 Average Precision Curve

The Precision-Recall Curve in Figure 30 indicates that all models show similar performance, with YOLOv5s performing slightly better at higher recall values. This suggests that YOLOv5s is slightly more effective at detecting a higher proportion of true positives while maintaining reasonable precision.

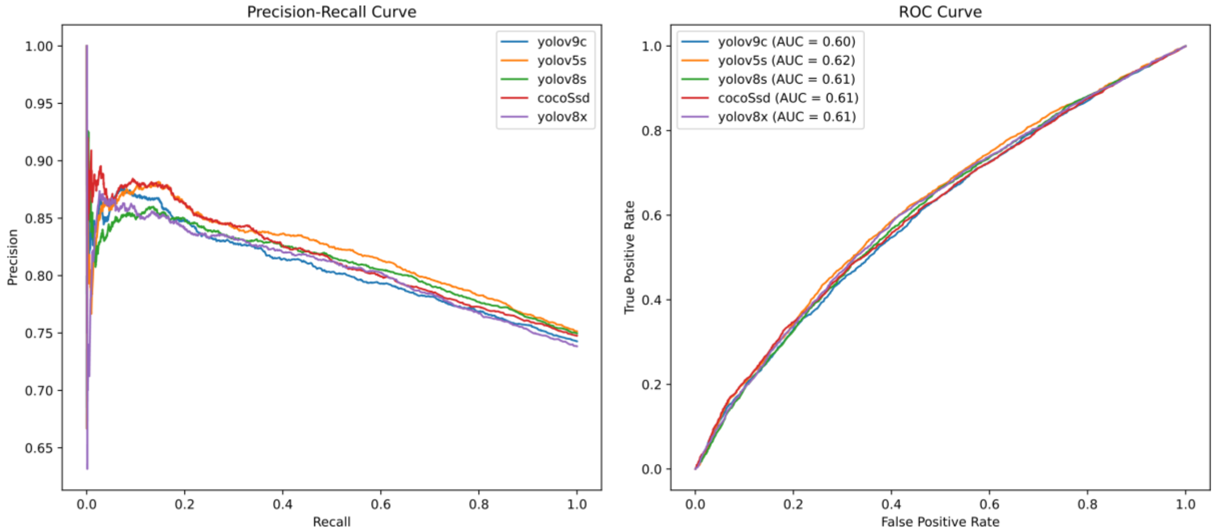

Figure 30 Precision-Recall and ROC Curve

In Figure 30, the ROC curve plots the true positive rate against the false positive rate at various threshold levels for each model. The ROC curve shows that all models perform comparably, as they exhibit similar closeness to the top-left corner. No model significantly outperforms the others, indicating similar abilities to distinguish between positive and negative instances.

6.4 Discussion of Results Section

The experiments evaluated the performance and reliability of private 5G and Wi-Fi in real-time object detection and video streaming applications within an industrial setting. Both networks showed similar detection accuracy, with Wi-Fi averaging an IOU of 0.646 and a score of 0.742 and private 5G averaging an IOU of 0.645 and a score of 0.74. Furthermore, the comparable throughput performance of private 5G and Wi-Fi suggests that both technologies can effectively support data-intensive applications in industrial settings. However, the potential for higher peak data rates in private 5G may be advantageous for applications with fluctuating or unpredictable data demands. Specifically, private 5G shows a mean throughput of 1.582 Mbps with a standard deviation of 0.783 Mbps and a maximum of 23.458 Mbps. In contrast, Wi-Fi shows a mean throughput of 1.574 Mbps with a standard deviation of 0.744 Mbps and a maximum of 12.913 Mbps. Regarding latency and jitter, Wi-Fi exhibits higher peaks compared to private 5G, but the standard deviations are similar. Wi-Fi shows a mean latency of 3.597 ms, a standard deviation of 2.3 ms, and a maximum of 160.5 ms, whereas private 5G has a mean latency of 13.46 ms, a standard deviation of 2.3 ms, and a range from 7.5 to 29.5 ms. While Wi-Fi's lower and more stable mean latency may be preferable for applications that require real-time responsiveness and are sensitive to delays, private 5G provides more consistent performance with less extreme variations. Figure 31 shows the performance comparison between private 5G and Wi-Fi based on the key parameters highlighted in this study.

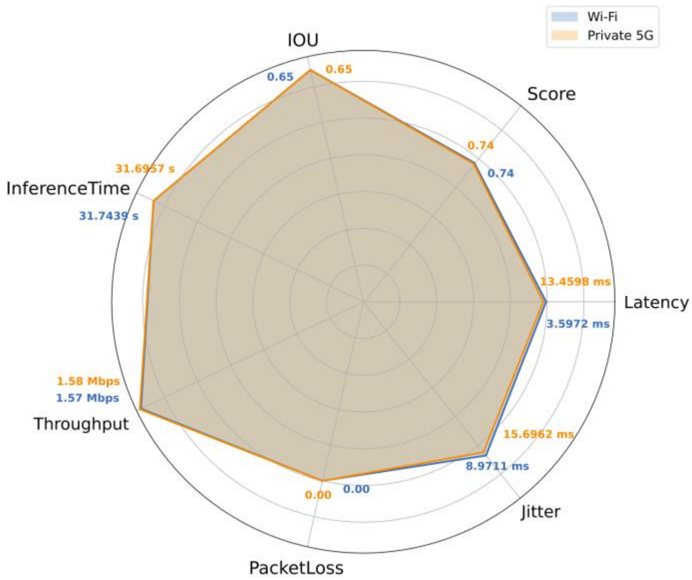

Figure 31 Performance Chart of Private 5G and Wi-Fi

The impact of the object detection model on accuracy underscores the importance of careful model selection. The superior performance of YOLOv5s in terms of IOU and other metrics suggests that it may be a more suitable choice for real-world applications compared to other models. YOLOv5s achieved the highest average IOU in Wi-Fi (0.656) and private 5G (0.645), as well as the highest average precision (0.82), F1 score (0.858), and ROC AUC (0.619). In comparison, YOLOv9c showed the lowest ROC AUC (0.601) and average precision (0.806), while YOLOv8x had the lowest F1 score (0.849) and mean IOU (0.641). When considering both Wi-Fi and private 5g, YOLOv5s achieved the highest average IOU (0.65). In comparison, YOLOv8s had the second highest average IOU (0.646), followed by cocoSsd (0.645), YOLOv9c (0.643), and YOLOv8x had the lowest average IOU (0.641). These results highlight the trade-off between model complexity and inference time in relation to object detection accuracy for live streaming. The difference in performance could be attributed to inference speed, suggesting that simpler models like YOLOv5s can sometimes outperform more complex ones in live stream conditions. Following cocoSsd, which had the lowest mean inference time of 31.372 ms, YOLOv5s had a mean inference time of 31.536 ms, making it the next quickest model. However, these conclusions need further investigation to fully understand the underlying factors and ensure the findings are robust.

Throughput has a significant positive effect on IOU, with a p-value of 0.0025, indicating this correlation between higher throughput and higher IOU values. However, a similar effect on the confidence score is not observed. Moreover, latency has a significant negative effect on confidence score, with a p-value of 0.0005. The limited influence of other network parameters on IOU highlights the need to consider other factors, such as object characteristics and environmental conditions when optimizing the performance of autonomous systems. The strong positive correlation between bounding box area and IOU (correlation coefficient of 0.706, p-value < 0.001) suggests that larger objects are generally easier to detect, which could inform the design of object placement and handling strategies in warehouses and factories.

The varying detection accuracy across different object classes emphasizes the importance of training object detection models on diverse datasets that adequately represent the range of objects encountered in real-world intralogistics environments. This could involve collecting data from various camera angles, lighting conditions, and object orientations to ensure robust performance across different scenarios. For example, trains, cats, and trucks have high IOU values (Wi-Fi: 0.781, 0.77, 0.753; Private 5G: 0.772, 0.774, 0.751), while birds and skateboards have lower IOU values (Wi-Fi: 0.423, 0.329; Private 5G: 0.421, 0.334).

During the 63 hours and 57 minutes of network parameter recordings for both Wi-Fi and private 5G, no packet loss was experienced. This is an important observation, as packet loss is a critical parameter for data transmission. The absence of packet loss during this extensive period highlights the stability and reliability of both networks, which is essential for accurate and efficient object detection in real-time applications.

Since there is no stable video dataset available for 360-degree video object detection, evaluating the performance of object detection with 360-degree videos is challenging. The network-induced latency is imperceptible to humans. However, a slight latency is observed when using the forklift with a 360-degree camera, primarily due to the processing of the 360-degree video. This latency is less pronounced when using conventional cameras. Additionally, latency is introduced by the FMS box, which also uses Wi-Fi. Since VR devices do not currently support a direct 5G connection, this additional layer is necessary for viewing the 360-degree video stream. Therefore, they must connect to the private 5G network through the FMS box, meaning the VR device is not directly connected to the private 5G network. Addressing this issue could further reduce latency. This was tested by viewing the stream from another phone with 5G capabilities, where reduced latency was observed. Moreover, there is also a small latency from the VR devices themselves because they must process 360-degree video in real time to provide a fully immersive experience. To enhance control intuitiveness during streaming, accounting for this latency difference ensures a smoother steering experience. These adjustments do not require any changes to the software or setup but would significantly improve latency.

If these small latency differences can be eliminated from the equation, a much more real-time and smooth experience can be achieved, considering the cumulative effect of all latencies. With advancements in technology, such as more effective algorithms for stitching and post-processing of 360-degree video, and the availability of 5G in VR devices, private 5G would be the optimal choice for implementing these types of applications, supported by state-of-the-art network infrastructures.

7 Future Work and Challenges

The development of a real-time object detection system faced challenges, including high latency with HLS, leading to a switch to WebRTC for low-latency streaming. The complexities of video conversion and post-processing with the Insta360 camera were resolved by utilizing the native software and smartphone for optimal performance. Limited processing power on devices necessitated centralized object detection with good connectivity. The system relies on the user interface for processing rather than the edge device itself. Network performance tests favored the Samsung Galaxy S21 over the Raspberry Pi in terms of both throughput and latency. Additionally, the computational demands of lossless video transmission led to adopting the more efficient lossy H.264 codec. These challenges highlighted the critical trade-offs in achieving a robust system.

For broader applicability, extending the scope of datasets beyond YT-BB is essential. Utilizing the framework established in this study, it should be feasible to achieve this extension. The LOCO [136] dataset, developed by researchers at TUM, offers a promising resource specifically designed for industrial logistic settings. Training YOLO and SSD models on LOCO and conducting further tests with this dataset could enhance the generalizability of the findings to real-world intralogistics environments. Additionally, exploring the utilization of more advanced object detection models and deploying the web application on enhanced hardware could potentially improve overall system performance and generalizability.

The current study employed "best effort" settings for SIM cards within the private 5G network. While sufficient for scenarios without strict traffic prioritization, alternative radio link control modes like real-time video and automation could potentially improve latency and throughput. However, these benefits might come at the cost of increased packet loss risk, which could have a more significant negative impact on object detection accuracy. Future experiments exploring the trade-off between these unacknowledged radio link control modes and "best effort" settings are necessary to find the optimal network configuration for industrial applications.

By addressing these challenges and pursuing these advancements, 5G holds promise for improving intralogistics efficiency, reliability, and security. However, to successfully deploy 5G-based object detection systems in real-world industrial settings, it is crucial to investigate and mitigate the impact of indoor environmental factors, such as lighting variations and object occlusion, on object detection accuracy.

8 Conclusion

This study investigated the potential of 5G technology in enhancing real-time object detection tasks, with a particular focus on industrial applications. The research revealed a nuanced landscape where private 5G networks demonstrated superior throughput and, under specific conditions, improved IOU for object detection. However, these networks also exhibited higher jitter and latency compared to Wi-Fi. This trade-off between speed and stability highlights the need for careful consideration when selecting network technology for real-time applications.

The choice of object detection model significantly influenced performance, highlighting the importance of selecting models that align with the specific requirements and constraints of real-world scenarios. The analysis further revealed the complex interplay between network conditions, object characteristics, and model performance, emphasizing the need for tailored network optimization strategies.

A notable finding of this study is the potential benefit of conducting object detection inference on the frontend in WebRTC-enabled applications. This approach can help maintain real-time performance and minimize latency. By performing inference directly on the client side, delays are reduced because video data doesn't need to be transmitted to a backend server for processing and then returned with the results. This can enhance the user experience, particularly for applications requiring instantaneous feedback, such as autonomous systems and real-time monitoring.

Moreover, this study suggests potential applications of 5G technology in autonomous vehicle use cases, as demonstrated by the development and successful operation of a VR-controlled small forklift equipped with object detection capabilities, from the O2 Tower to the TUM testing hall. Furthermore, a framework was developed to evaluate the performance of object detection models on live streams in relation to network parameters. This tool could be helpful for future research and development in this area.

In conclusion, while private 5G networks offer promising advantages for real-time object detection, their successful implementation requires careful consideration of network optimization, model selection, and the specific challenges posed by real-world environments. Future research should address these challenges, explore the integration of 5G with other technologies, and develop specialized object detection models tailored to the unique demands of various applications. By doing so, the potential of 5G in creating more efficient, reliable, and intelligent systems across diverse domains can be better understood and realized.

Bibliography

[1] M. Wen *et al.*, "Private 5G networks: Concepts, architectures, and research landscape," *IEEE Journal of Selected Topics in Signal Processing,* vol. 16, no. 1, pp. 7-25, 2021.

[2] A. Alabbasi, T. Dudda, Z. Zou, and J. Kronander, "5G toolbox for realizing industrial automation," in *2019 IEEE 2nd 5G World Forum (5GWF)*, 2019: IEEE, pp. 512-515.

[3] M. Peterhansl, "Remote Control of a Collaborative Robot with Virtual Reality and Joystick in a 5G Network," in *2023 13th International Conference on Advanced Computer Information Technologies (ACIT)*, 2023: IEEE, pp. 473-478.

[4] A. Acemoglu *et al.*, "5G robotic telesurgery: Remote transoral laser microsurgeries on a cadaver," *IEEE Transactions on Medical Robotics and Bionics,* vol. 2, no. 4, pp. 511-518, 2020.

[5] M. Jurčević and I. Vitas, "5G-Connected Drone for Public Road Safety-Research Challenges and Future Research Roadmap," in *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, 27 Sept.-1 Oct. 2021 2021, pp. 417-422, doi: 10.23919/MIPRO52101.2021.9597065.

[6] Ericsson. "Ericsson USA 5G Smart Factory." https://www.ericsson.com/en/about-us/company-facts/ericsson-worldwide/united-states/5g-smart-factory (accessed 2024).

[7] M. Electric. "Mitsubishi Electric Begins Demonstrating Local 5G System in Japan." https://www.mitsubishielectric.com/news/2020/0518.html (accessed 2024).

[8] U. S. DepartmentofDefense. "Department of Defense Successfully Demonstrates a 5G Network for Smart Warehouses." https://www.defense.gov/News/Releases/Release/Article/2650242/department-of-defense-successfully-demonstrates-a-5g-network-for-smart-warehous/ (accessed 2024).

[9] Nokia. "Nokia, Deutsche Telekom and Hamburg Port Authority collaborate in 5G research in industrial environment." https://www.nokia.com/about-us/news/releases/2018/02/02/nokia-deutsche-telekom-and-hamburg-port-authority-collaborate-in-5g-research-in-industrial-environment/ (accessed 2024).

[10] P. Sossalla, J. Rischke, F. Baier, S. Itting, G. T. Nguyen, and F. H. P. Fitzek, "Private 5G Solutions for Mobile Industrial Robots: A Feasibility Study," in *2022 IEEE Symposium on Computers and Communications (ISCC)*, 30 June-3 July 2022 2022, pp. 1-6, doi: 10.1109/ISCC55528.2022.9912791.

[11] M. Uitto and A. Heikkinen, "Evaluating 5G uplink performance in low latency video streaming," in *2022 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 7-10 June 2022 2022, pp. 393-398, doi: 10.1109/EuCNC/6GSummit54941.2022.9815703.

[12] M. Uitto and A. Heikkinen, "Evaluation of live video streaming performance for low latency use cases in 5g," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 2021: IEEE, pp. 431-436.

[13] J. M. Batalla, "On Analyzing Video Transmission Over Wireless WiFi and 5G C-Band in Harsh IIoT Environments," *IEEE Access,* vol. 8, pp. 118534-118541, 2020, doi: 10.1109/ACCESS.2020.3005641.

[14] R. Maldonado *et al.*, "Comparing Wi-Fi 6 and 5G Downlink Performance for Industrial IoT," *IEEE Access,* vol. 9, pp. 86928-86937, 2021, doi: 10.1109/ACCESS.2021.3085896.

[15] G. Americas, "3GPP Technology Trends," January 2024. Accessed: 2024. [Online]. Available: https://www.5gamericas.org/3gp-technology-trends/

[16] X. Wei, M. Zhou, and W. Jia, "Toward Low-Latency and High-Quality Adaptive 360° Streaming," *IEEE Transactions on Industrial Informatics,* vol. 19, no. 5, pp. 6326-6336, 2023, doi: 10.1109/TII.2022.3192398.

[17] V. Clincy and B. Wilgor, "Subjective Evaluation of Latency and Packet Loss in a Cloud-Based Game," in *2013 10th International Conference on Information Technology: New Generations*, 15-17 April 2013 2013, pp. 473-476, doi: 10.1109/ITNG.2013.79.

[18] F. Rhinow, P. P. Veloso, C. Puyelo, S. Barrett, and E. O. Nuallain, "P2P live video streaming in WebRTC," in *2014 World Congress on Computer Applications and Information Systems (WCCAIS)*, 17-19 Jan. 2014 2014, pp. 1-6, doi: 10.1109/WCCAIS.2014.6916588.

[19] R. Sacoto-Martins, J. Madeira, J. P. Matos-Carvalho, F. Azevedo, and L. M. Campos, "Multi-purpose Low Latency Streaming Using Unmanned Aerial Vehicles," in *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, 20-22 July 2020 2020, pp. 1-6, doi: 10.1109/CSNDSP49049.2020.9249562.

[20] O. E. Marai, T. Taleb, and J. Song, "AR-Based Remote Command and Control Service: Self-Driving Vehicles Use Case," *IEEE Network,* vol. 37, no. 3, pp. 170-177, 2023, doi: 10.1109/MNET.119.2200058.

[21] A. Muslim and S. Recker, "End-to-End Latency Measurements with Black Box Components," in *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 2023, vol. 1: IEEE, pp. 73-78.

[22] H. W. Kao and E. H. K. Wu, "QoE Sustainability on 5G and Beyond 5G Networks," *IEEE Wireless Communications,* vol. 30, no. 1, pp. 118-125, 2023, doi: 10.1109/MWC.007.2200260.

[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[24] A. Baretto, N. Pudussery, V. Subramaniam, and A. Siddiqui, "Real-Time WebRTC based Mobile Surveillance System," *International Journal of Engineering and Management Research,* vol. 11, 06/07 2021, doi: 10.31033/ijemr.11.3.4.

[25] M. Phankokkruad and P. Jaturawat, "An evaluation of technical study and performance for real-time face detection using Web Real-Time Communication," in *2015 International Conference on Computer, Communications, and Control Technology (I4CT)*, 21-23 April 2015 2015, pp. 162-166, doi: 10.1109/I4CT.2015.7219558.

[26] M. Li, Y.-X. Wang, and D. Ramanan, "Towards streaming perception," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, 2020: Springer, pp. 473-488.

[27] J. Yang, S. Liu, Z. Li, X. Li, and J. Sun, "Real-time object detection for streaming perception," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 5385-5395.

[28] B. P. Crow, I. Widjaja, J. G. Kim, and P. T. Sakai, "IEEE 802.11 Wireless Local Area Networks," *IEEE Communications Magazine,* vol. 35, no. 9, pp. 116-126, 1997, doi: 10.1109/35.620533.

[29] "IEEE Standard for Information Technology - Telecommunications and information exchange between systems - Local and Metropolitan networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher Speed Physical Layer (PHY) Extension in the 2.4 GHz band," *IEEE Std 802.11b-1999,* pp. 1-96, 2000, doi: 10.1109/IEEESTD.2000.90914.

[30] "IEEE Standard for Information technology-- Local and metropolitan area networks-- Specific requirements-- Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput," *IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009),* pp. 1-565, 2009, doi: 10.1109/IEEESTD.2009.5307322.

[31] "IEEE Standard for Information technology-- Telecommunications and information exchange between systemsLocal and metropolitan area networks-- Specific requirements--Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications--Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz," *IEEE Std 802.11ac-2013 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012, IEEE Std 802.11aa-2012, and IEEE Std 802.11ad-2012),* pp. 1-425, 2013, doi: 10.1109/IEEESTD.2013.6687187.

[32] C. Chen, X. Chen, D. Das, D. Akhmetov, and C. Cordeiro, "Overview and performance evaluation of Wi-Fi 7," *IEEE Communications Standards Magazine,* vol. 6, no. 2, pp. 12-18, 2022.

[33] E. Ferro and F. Potorti, "Bluetooth and Wi-Fi wireless protocols: a survey and a comparison," *IEEE Wireless Communications,* vol. 12, no. 1, pp. 12-26, 2005.

[34] D. W. Bliss, K. W. Forsythe, and A. M. Chan, "MIMO wireless communication," *Lincoln Laboratory Journal,* vol. 15, no. 1, pp. 97-126, 2005.

[35] Elektronik-Kompendium. "WLAN-Übertragungstechnik." www.elektronik-kompendium.de (accessed 2024)

[36] K.-H. Lee, "Using OFDMA for MU-MIMO user selection in 802.11 ax-based Wi-Fi networks," *IEEE Access,* vol. 7, pp. 186041-186055, 2019.

[37] S. Avallone, P. Imputato, G. Redieteab, C. Ghosh, and S. Roy, "Will OFDMA improve the performance of 802.11 WiFi networks?," *IEEE Wireless Communications,* vol. 28, no. 3, pp. 100-107, 2021.

[38] K. Pahlavan and P. Krishnamurthy, "Evolution and impact of Wi-Fi technology and applications: A historical perspective," *International Journal of Wireless Information Networks,* vol. 28, pp. 3-19, 2021.

[39] G. Forecast, "Cisco visual networking index: global mobile data traffic forecast update, 2017–2022," *Update,* vol. 2017, p. 2022, 2019.

[40] E. Reshef and C. Cordeiro, "Future directions for Wi-Fi 8 and beyond," *IEEE Communications Magazine,* vol. 60, no. 10, pp. 50-55, 2022.

[41] K. Santhi, V. Srivastava, G. SenthilKumaran, and A. Butare, "Goals of true broad band's wireless next wave (4G-5G)," in *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No. 03CH37484)*, 2003, vol. 4: IEEE, pp. 2317-2321.

[42] S. Dehghan, D. Lister, R. Owen, and P. Jones, "W-CDMA capacity and planning issues," *Electronics & communication engineering journal,* vol. 12, no. 3, pp. 101-118, 2000.

[43] I.-R. Recommendation, "Guidelines for evaluation of radio transmission technologies for IMT-2000," *Rec. ITU-R M. 1225,* 1997.

[44] E. Ezhilarasan and M. Dinakaran, "A review on mobile technologies: 3G, 4G and 5G," in *2017 second international conference on recent trends and challenges in computational models (ICRTCCM)*, 2017: IEEE, pp. 369-373.

[45] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, 2012, pp. 225-238.

[46] R. binti Ramle and J. N. bin Hamid, "Investigating the distance effects on performance degradation of mobile WiMAX technology using NCTUns tool," in *2010 International Conference on Science and Social Research (CSSR 2010)*, 2010: IEEE, pp. 933-937.

[47] G. Intelligence, "Understanding 5G: Perspectives on future technological advancements in mobile," *White paper,* pp. 1-26, 2014.

[48] S.-Y. Shin, D. Kim, and S. A. Chun, "Digital divide in advanced smart city innovations," *Sustainability,* vol. 13, no. 7, p. 4076, 2021.

[49] M. M. d. Silva and J. Guerreiro, "On the 5G and Beyond," *Applied Sciences,* vol. 10, no. 20, p. 7091, 2020.

[50] M. Series, "IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond," *Recommendation ITU,* vol. 2083, no. 0, 2015.

[51] B. S. Khan, S. Jangsher, A. Ahmed, and A. Al-Dweik, "URLLC and eMBB in 5G Industrial IoT: A Survey," *IEEE Open Journal of the Communications Society,* vol. 3, pp. 1134-1163, 2022, doi: 10.1109/OJCOMS.2022.3189013.

[52] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh, "5G URLLC: Design Challenges and System Concepts," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, 28-31 Aug. 2018 2018, pp. 1-6, doi: 10.1109/ISWCS.2018.8491078.

[53] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Communications Surveys & Tutorials,* vol. 22, no. 2, pp. 905-929, 2020.

[54] A. Gohar and G. Nencioni, "The role of 5G technologies in a smart city: The case for intelligent transportation system," *Sustainability,* vol. 13, no. 9, p. 5188, 2021.

[55] J. Wang, J. Liu, and N. Kato, "Networking and communications in autonomous driving: A survey," *IEEE Communications Surveys & Tutorials,* vol. 21, no. 2, pp. 1243-1274, 2018.

[56] L. Chettri and R. Bera, "A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems," *IEEE Internet of Things Journal,* vol. 7, no. 1, pp. 16-32, 2019.

[57] M. Wollschlaeger, T. Sauter, and J. Jasperneite, "The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0," *IEEE industrial electronics magazine,* vol. 11, no. 1, pp. 17-27, 2017.

[58] R. Gerzaguet *et al.*, "The 5G candidate waveform race: a comparison of complexity and performance," *EURASIP Journal on Wireless Communications and Networking,* vol. 2017, pp. 1-14, 2017.

[59] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE communications magazine,* vol. 55, no. 5, pp. 94-100, 2017.

[60] L. Rao, M. Pant, L. Malviya, A. Parmar, and S. V. Charhate, "5G beamforming techniques for the coverage of intended directions in modern wireless communication: in-depth review," *International Journal of Microwave and Wireless Technologies,* vol. 13, no. 10, pp. 1039-1062, 2021.

[61] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE communications magazine,* vol. 52, no. 2, pp. 74-80, 2014.

[62] M. Agiwal, H. Kwon, S. Park, and H. Jin, "A survey on 4G-5G dual connectivity: Road to 5G implementation," *Ieee Access,* vol. 9, pp. 16193-16210, 2021.

[63] A. El Rhayour and T. Mazri, "5G Architecture: Deployment scenarios and options," in *2019 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)*, 2019: IEEE, pp. 1-6.

[64] R. Mohamed, S. Zemouri, and C. Verikoukis, "Performance evaluation and comparison between SA and NSA 5G networks in indoor environment," in *2021 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*, 2021: IEEE, pp. 112-116.

[65] r. G. P. P. (3GPP), "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (Release 17)," 3GPP, 3GPP TS 36.101 V17.2.0, 2021.

[66] A. Mahmood, S. F. Abedin, T. Sauter, M. Gidlund, and K. Landernäs, "Factory 5G: a review of industrial-centric features and deployment options," *Authorea Preprints,* 2023.

[67] D. N. K. Jayakody, K. Srinivasan, and V. Sharma, *5G Enabled Secure Wireless Networks*. Springer, 2019.

[68] S. Eswaran and P. Honnavalli, "Private 5G networks: a survey on enabling technologies, deployment models, use cases and research directions," *Telecommunication Systems,* vol. 82, no. 1, pp. 3-26, 2023.

[69] A. Jerichow, B. Covell, D. Chandramouli, A. Rezaki, A. Lansisalmi, and J. Merkel, "3GPP non-public network security," *Journal of ICT Standardization,* vol. 8, no. 1, pp. 57-76, 2020.

[70] A. Aijaz, "Private 5G: The Future of Industrial Wireless," *IEEE Industrial Electronics Magazine,* vol. 14, no. 4, pp. 136-145, 2020, doi: 10.1109/MIE.2020.3004975.

[71] S. Kekki *et al.*, "MEC in 5G networks," *ETSI white paper,* vol. 28, no. 2018, pp. 1-28, 2018.

[72] L. L. Bello and W. Steiner, "A perspective on IEEE time-sensitive networking for industrial communication and automation systems," *Proceedings of the IEEE,* vol. 107, no. 6, pp. 1094-1120, 2019.

[73] J. Prados-Garzon, P. Ameigeiras, J. Ordonez-Lucena, P. Muñoz, O. Adamuz-Hinojosa, and D. Camps-Mur, "5G non-public networks: Standardization, architectures and challenges," *IEEE Access,* vol. 9, pp. 153893-153908, 2021.

[74] F. Longo, A. Padovano, G. Aiello, C. Fusto, and A. Certa, "How 5G-based industrial IoT is transforming human-centered smart factories: A Quality of Experience model for Operator 4.0 applications," *IFAC-PapersOnLine,* vol. 54, no. 1, pp. 255-262, 2021.

[75] G. ACIA, "5G for Connected Industries and Automation," ed: zVEI-German Electrical and Electronic Manufacturers' Association, 2019.

[76] VDMA, Verband Deutscher Maschinen- und Anlagenbau (VDMA), , "Forum Intralogistik," 2024.

[77] S. Winkelhaus, E. H. Grosse, and C. H. Glock, "Job satisfaction: An explorative study on work characteristics changes of employees in Intralogistics 4.0," *Journal of Business Logistics,* vol. 43, no. 3, pp. 343-367, 2022.

[78] S. Winkelhaus and E. H. Grosse, "Logistics 4.0: a systematic review towards a new logistics system," *International Journal of Production Research,* vol. 58, no. 1, pp. 18-43, 2020.

[79] C. Eppner *et al.*, "Lessons from the amazon picking challenge: Four aspects of building robotic systems," in *Robotics: science and systems*, 2016, vol. 12.

[80] T. Wisspeintner, T. Van Der Zant, L. Iocchi, and S. Schiffer, "RoboCup@ Home: Scientific competition and benchmarking for domestic service robots," *Interaction Studies,* vol. 10, no. 3, pp. 392-426, 2009.

[81] X. Su, C.-C. Chu, B. Prabhu, and R. Gadh, "On the creation of automatic identification and data capture infrastructure via RFID," *The Internet of Things: from RFID to the next-generation pervasive networked systems. Auerbach,* 2007.

[82] S. Fosso Wamba and A. Takeoka Chatfield, "The impact of RFID technology on warehouse process innovation: A pilot project in the TPL industry," *Information Systems Frontiers,* vol. 13, pp. 693-706, 2011.

[83] B. S. S. Tejesh and S. Neeraja, "Warehouse inventory management system using IoT and open source framework," *Alexandria engineering journal,* vol. 57, no. 4, pp. 3817-3823, 2018.

[84] Y. Zhang and M. Li, "Wireless Communication Management Model of Workshop and Warehouse Manufacturing System," in *2009 Second International Conference on Intelligent Computation Technology and Automation*, 10-11 Oct. 2009 2009, vol. 4, pp. 121-123, doi: 10.1109/ICICTA.2009.746.

[85] L. Custodio and R. Machado, "Flexible automated warehouse: a literature review and an innovative framework," *The International Journal of Advanced Manufacturing Technology,* vol. 106, pp. 533-558, 2020.

[86] M.-J. Bélanger, R. Pellerin, and S. Lamouri, "A literature review on digital twins in warehouses," *Procedia Computer Science,* vol. 219, pp. 370-377, 2023.

[87] J. Fottner *et al.*, "Autonomous systems in intralogistics–state of the art and future research challenges," *Journal Logistics Research,* 2021.

[88] W. Kersten, M. Seiter, B. von See, N. Hackius, and T. Maurer, "Trends and strategies in logistics and supply chain management–digital transformation opportunities," *BVL International,* 2017.

[89] T. Ruether. "Streaming Protocols and Latency." https://www.wowza.com/blog/streaming-protocols (accessed 2024)

[90] R. Pantos and W. May, "HTTP live streaming," 2070-1721, 2017.

[91] Cloudflare. "What is MPEG-DASH?" https://www.cloudflare.com/learning/video/what-is-mpeg-dash/ (accessed 2024)

[92] A. H. D. Streaming. http://www.adobe.com/products/httpdynamicstreaming/ (accessed 2024)

[93] M. S. Streaming. http://www.iis.net/download/ (accessed 2024)

[94] H. Schulzrinne, A. Rao, and R. Lanphier, "RFC2326: Real time streaming protocol (RTSP)," ed: RFC Editor, 1998.

[95] A.-V. T. W. Group, H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RFC1889: RTP: A transport protocol for real-time applications," ed: RFC Editor, 1996.

[96] X. Lei, X. Jiang, and C. Wang, "Design and implementation of streaming media processing software based on RTMP," in *2012 5th International Congress on Image and Signal Processing*, 2012: IEEE, pp. 192-196.

[97] B. Sredojev, D. Samardzija, and D. Posarac, "WebRTC technology overview and signaling solution design and implementation," in *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, 2015: IEEE, pp. 1006-1009.

[98] B. Jansen, T. Goodwin, V. Gupta, F. Kuipers, and G. Zussman, "Performance evaluation of WebRTC-based video conferencing," *ACM SIGMETRICS performance evaluation review,* vol. 45, no. 3, pp. 56-68, 2018.

[99] WebRTC. https://webrtc.github.io/webrtc-org/architecture/ (accessed 2024).

[100] B. Garcia, L. Lopez-Fernandez, F. Gortazar, and M. Gallego, "Analysis of video quality and end-to-end latency in WebRTC," in *2016 IEEE Globecom Workshops (GC Wkshps)*, 2016: IEEE, pp. 1-6.

[101] E. Rescorla, "WebRTC security architecture," *Work in Progress,* 2013.

[102] *OvenMediaEngine*. AirenSoft. [Online]. Available: https://github.com/AirenSoft/OvenMediaEngine (accessed 2024)

[103] . Oven Media Engine. [Online]. Available: https://www.ovenmediaengine.com (accessed 2024)

[104] Apple. "Enabling Low-Latency HTTP Live Streaming (HLS)." https://developer.apple.com/documentation/http-live-streaming/enabling-low-latency-http-live-streaming-hls (accessed 2024).

[105] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision,* vol. 115, pp. 211-252, 2015.

[106] D. Wang, J.-G. Wang, and K. Xu, "Deep learning for object detection, classification and tracking in industry applications," vol. 21, ed: MDPI, 2021, p. 7349.

[107] N. Rane, "YOLO and Faster R-CNN object detection for smart Industry 4.0 and Industry 5.0: applications, challenges, and opportunities," *Available at SSRN 4624206,* 2023.

[108] C. Rennie, R. Shome, K. E. Bekris, and A. F. De Souza, "A dataset for improved rgbd-based object detection and pose estimation for warehouse pick-and-place," *IEEE Robotics and Automation Letters,* vol. 1, no. 2, pp. 1179-1185, 2016.

[109] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.

[110] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.

[111] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems,* vol. 28, 2015.

[112] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, 2016: Springer, pp. 21-37.

[113] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International journal of computer vision,* vol. 104, pp. 154-171, 2013.

[114] M. Maity, S. Banerjee, and S. S. Chaudhuri, "Faster R-CNN and YOLO based Vehicle detection: A Survey," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 8-10 April 2021 2021, pp. 1442-1447, doi: 10.1109/ICCMC51019.2021.9418274.

[115] M. Phadtare, V. Choudhari, R. Pedram, and S. Vartak, "Comparison between yolo and ssd mobile net for object detection in a surveillance drone," *Int. J. Sci. Res. Eng. Man,* vol. 5, pp. 1-5, 2021.

[116] J. Dean and R. Monga'TensorFlow, "Large-Scale Machine Learning on Heterogeneous Distributed Systems'," *TensorFlow. org,* 2015.

[117] D. Smilkov *et al.*, "Tensorflow. js: Machine learning for the web and beyond," *Proceedings of Machine Learning and Systems,* vol. 1, pp. 309-321, 2019.

[118] T. L. F. AI. "ONNX: Open Neural Network Exchange." https://onnx.ai (accessed 2024)

[119] O. R. developers. "ONNX Runtime." https://onnxruntime.ai/ (accessed 2024)

[120] T. L. F. AI, "ONNX.js," 2021. [Online]. Available: https://github.com/microsoft/onnxjs. (accessed 2024)

[121] S.-H. Tsang, "Brief Review: YOLOv5 for Object Detection." [Online]. Available: https://sh-tsang.medium.com/brief-review-yolov5-for-object-detection-84cc6c6a0e3a (accessed 2024)

[122] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861,* 2017.

[123] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 2014: Springer, pp. 740-755.

[124] *YOLOv5 by Ultralytics*. (2020). [Online]. Available: https://github.com/ultralytics/yolov5 (accessed 2024)

[125] *Ultralytics YOLO v8*. (2023). [Online]. Available: https://github.com/ultralytics/ultralytics (accessed 2024)

[126] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616,* 2024.

[127] Ultralytics. "YOLOv9: A Leap Forward in Object Detection Technology." https://docs.ultralytics.com/models/yolov9/ (accessed 2024).

[128] E. Real, J. Shlens, S. Mazzocchi, X. Pan, and V. Vanhoucke, "Youtube-boundingboxes: A large high-precision human-annotated data set for object detection in video," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5296-5305.

[129] R. P. Foundation. "Raspberry Pi 4 Model B Specifications." https://www.raspberrypi.org/products/raspberry-pi-4-model-b/specifications/ (accessed 2024)

[130] Waveshare. "RM500Q-GL 5G HAT." https://www.waveshare.com/wiki/RM500Q-GL_5G_HAT (accessed 2024).

[131] H. Magazine. (2022) 5G-Mobilfunk per USB.

[132] ASKEY, "o2 HomeSpot 5G Handbuch." [Online]. Available: https://www.askey.com.tw/file/germany_o2/ (accessed 2024)

[133] Qualcomm. "FastConnect 6900." https://www.qualcomm.com/products/technology/wi-fi/fastconnect/fastconnect-6900 (accessed 2024)

[134] Qualcomm. "Snapdragon X60 5G Modem-RF System." https://www.qualcomm.com/products/technology/modems/snapdragon-x60-5g-modem (accessed 2024).

[135] Ericsson, "Technical Product Description Ericsson Private 5G 24.10," 2021.

[136] C. Mayershofer, D. M. Holm, B. Molter, and J. Fottner, "LOCO: Logistics Objects in Context," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 14-17 Dec. 2020 2020, pp. 612-617, doi: 10.1109/ICMLA51294.2020.00102.

List of Figures

List of Tables

Appendix 1

Below are some examples of object detection results displayed on the dataset video. The green bounding boxes represent the ground truth, while the red bounding boxes indicate the predictions of the object detection models. Figure 1 shows the detection of an elephant, and Figure 2 shows the detection of a bus.

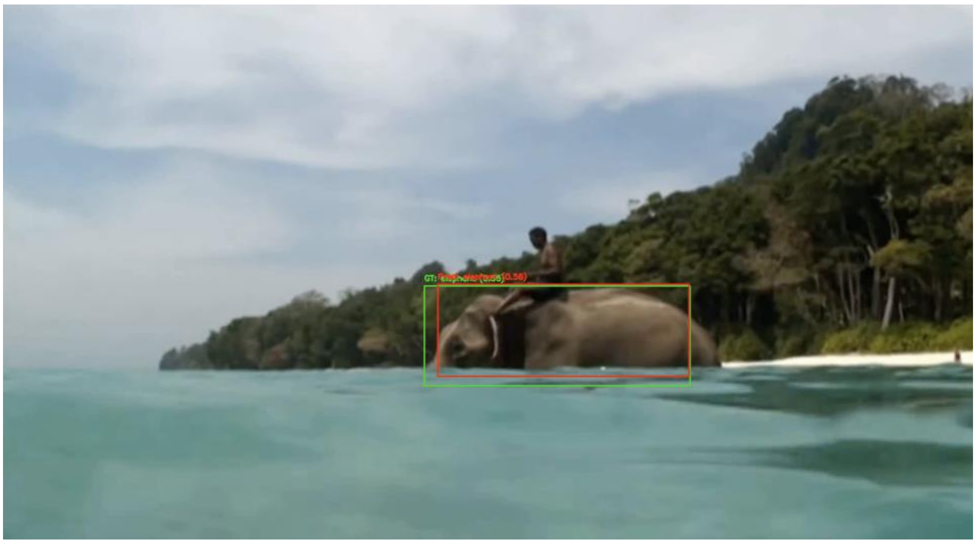

Figure 1

Figure 2

Appendix 2

The following are the results of Tukey's HSD analysis:

Group 1	Group 2	Mean Difference	P-Adjusted	Lower	Upper	Reject
airplane	bird	-0.29	0.0	-0.313	-0.266	True
airplane	skateboard	-0.38	0.0	-0.403	-0.358	True
bear	bird	-0.23	0.0	-0.252	-0.208	True
bear	skateboard	-0.321	0.0	-0.342	-0.3	True
bird	boat	0.104	0.0	0.077	0.13	True
bird	bus	0.255	0.0	0.234	0.276	True
bird	skateboard	-0.091	0.0	-0.118	-0.064	True
bird	train	0.355	0.0	0.311	0.399	True
bird	truck	0.33	0.0	0.308	0.352	True
cat	skateboard	-0.44	0.0	-0.463	-0.418	True

Tukey's HSD Results for IOU by Class with Meaningful Combinations

Group 1	Group 2	Mean Difference	P-Adjusted	Lower	Upper	Reject
bear	skateboard	-0.107	0.0	-0.122	-0.092	True
bird	train	0.101	0.0	0.07	0.133	True
boat	skateboard	-0.058	0.0	-0.077	-0.04	True
bus	skateboard	-0.101	0.0	-0.115	-0.086	True
car	skateboard	-0.075	0.0	-0.091	-0.059	True
cat	skateboard	-0.079	0.0	-0.095	-0.063	True
dog	skateboard	-0.102	0.0	-0.128	-0.075	True
elephant	skateboard	-0.098	0.0	-0.113	-0.083	True
knife	skateboard	-0.088	0.0	-0.103	-0.073	True
skateboard	train	0.135	0.0	0.103	0.166	True

Tukey's HSD Results for Score by Class with Meaningful Combinations

Group 1	Group 2	Mean Difference	P-Adjusted	Lower	Upper	Reject
cocoSsd	yolov5s	0.005	0.422	-0.003	0.013	False
cocoSsd	yolov8s	0.001	0.998	-0.007	0.009	False
cocoSsd	yolov8x	-0.004	0.584	-0.012	0.004	False
cocoSsd	yolov9c	-0.002	0.961	-0.01	0.006	False
yolov5s	yolov8s	-0.004	0.619	-0.012	0.004	False
yolov5s	yolov8x	-0.009	0.012	-0.017	-0.001	True
yolov5s	yolov9c	-0.007	0.116	-0.015	0.001	False
yolov8s	yolov8x	-0.005	0.388	-0.013	0.003	False
yolov8s	yolov9c	-0.003	0.862	-0.011	0.005	False
yolov8x	yolov9c	0.002	0.935	-0.006	0.01	False

Tukey‘s HSD Results for IOU by Model

Group 1	Group 2	Mean Difference	P-Adjusted	Lower	Upper	Reject
cocoSsd	yolov5s	0.005	0.056	-0.0	0.011	False
cocoSsd	yolov8s	0.002	0.918	-0.004	0.007	False
cocoSsd	yolov8x	0.004	0.388	-0.002	0.009	False
cocoSsd	yolov9c	0.001	0.995	-0.005	0.006	False
yolov5s	yolov8s	-0.004	0.347	-0.009	0.002	False
yolov5s	yolov8x	-0.002	0.894	-0.007	0.004	False
yolov5s	yolov9c	-0.004	0.146	-0.01	0.001	False
yolov8s	yolov8x	0.002	0.882	-0.004	0.007	False
yolov8s	yolov9c	-0.001	0.991	-0.006	0.004	False
yolov8x	yolov9c	-0.003	0.633	-0.008	0.003	False

Tukey‘s HSD Results for Score by Model

The following descriptive statistics are from the live streaming of a 6 hour and 23 minutes recording for private 5g network at the TUM Testing Hall with the best-performing model, YOLOv5s. The RSRP measured was -72 dBm.

	Jitter (milliseconds)	Latency (milliseconds)	Score	IOU	Throughput (megabits per second)	Packet Loss	Inference Time (seconds)
mean	16.781	15.843	0.738	0.644	1.567	0.0	35.208
std	2.273	2.657	0.158	0.235	0.725	0.0	6.217
min	10.0	10.5	0.5	0.0	0.015	0.0	24.6
25%	15.0	14.0	0.59	0.498	1.273	0.0	30.3
50%	17.0	15.5	0.727	0.685	1.466	0.0	33.0
75%	18.0	17.5	0.881	0.826	1.694	0.0	40.9
max	28.0	76.5	0.999	0.997	12.665	0.0	56.7

Results for Private 5G Network at TUM Testing Hall