D5.3 State of the Art Analysis

Security of Critical Infrastructure by Multi-Modal Dynamic Sensing and AI

01 August 2024

ITEA Project No: 22006

sintra-ai.eu

TABLE OF CONTENTS

ACRONYMS

ACM	Access Control Manager
ADAS	Advanced Driver-Assistance Systems
AI	Artificial Intelligence
AoA	Angle of Arrival
AP	Average Precision
AR	Augmented Reality
ASR	Automatic Speech Recognition
BACnet	Building Automation and Control network
BiLSTM	Bidirectional long short-term memory
CaaS	Container as a Service
CCTV	Closed-Circuit Television
CER	Character Error Rate
CMOS	Complementary Metal-Oxide-Semiconductor
CNN	Convolutional Neural Networks
COCO	Common Objects in Context
DALI	Digital Addressable Lighting Interface
DAS	Depth Sensing Activity Recognition System
DeepSORT	Deep Simple Online and Realtime Tracking
FBRRAS	Food and beverage, retail, shop and airport safety and security solution
FCN	Fully Convolutional Networks
FFT	Fast Fourier Transform
FIR	Far Infrared
FOV	Field-of-view
GIS	Geographic Information Systems
GS	Gas Chromatography
HAR	Human Action Recognition
HCI	Human–Computer Interaction
HOG	Histogram of Oriented Gradients
HTTP	Hypertext Transfer Protocol
ICAO	International Civil Aviation Organization
IMS	Ion Mobility Spectrometry
IoT	Internet of Things
IP	Internet Protocol
LSTM	Long Short-Term Memory
MIL	Multi-instance Learning

ML	Machine Learning
MOT	Multiple Object Tracking
MOTA	Multiple Object Tracking Accuracy
MOTP	Multiple Object Tracking Precision
MQTT	Message Queuing Telemetry Transport
MS	Mass Spectrometry
MSE	Mean Squared Error
NIR	Near Infrared
NIRS	Near Infrared Spectroscopy
OP	Operating System
OPC/UA	Open Platform Communications Unified Architecture
PCK	Percentage of Correct Keypoints
PLC	Programmable Logic Controller
RCE	Resonant-Cavity-Enhanced
RNN	Recurrent Neural Networks
SHGM	International Civil Aviation Organization
SINTRA	Security of Critical Infrastructure by Multi-Modal Dynamic Sensing and Artificial Intelligence
SME	Small and Medium-Sized Enterprises
SORT	Simple Online and Realtime Tracking
SSD	Single Shot Multi Box Detector
SVM	Support Vector Machines
TCP	Transmission Control Protocol
ToF	Time-of-Flight
UDP	User Datagram Protocol
VAE	Variational Auto Encoder
VR	Virtual Reality
WER	Word Error Rate
YOLO	You Only Look Once

TABLE OF FIGURES

TABLE OF TABLES

1. INTRODUCTION

Welcome to the D5.3 State-of-the-Art Gap Analysis document of the SINTRA project, an ambitious initiative that aims to improve the resilience and protection of the critical infrastructures by developing an open data-streaming AI platform that enables interoperability, information sharing, and privacy protection.

Within the scope of the SINTRA project, the focus is on solving safety and security problems innovatively and collaboratively for different critical infrastructures like airports, harbours, construction sites and railways.

2. STATE OF THE ART ANALYSIS

The integration of multi-modal sensing technologies and AI-powered data analysis has revolutionized the field of infrastructure security and safety. By combining data from various sensor modalities and existing data sources, a comprehensive and nuanced understanding of security and safety situations can be achieved. This approach enhances the detection of anomalies, maps them to potential threats, and facilitates coordinated responses. This literature study explores state-of-the-art technologies and methodologies in this domain, focusing on the fusion of diverse sensor data and AI-based analysis techniques. The following subsections provide the state-of-the-art analysis of important topics such as human action recognition (HAR), emotion detection, object detection, speech detection, anomaly detection, localization analysis, human heat map, CCTV which are essential to the scope of the project and its key technological areas.

2.1 State of the Art Technologies and Literature Studies

2.1.1 Artificial Intelligence

Artificial Intelligence (AI) has emerged as a powerful tool in the detection of hidden, complex, or context-dependent anomalies. By leveraging advanced machine learning algorithms and deep learning techniques, AI systems can identify patterns and irregularities that may not be apparent through traditional methods. These capabilities are crucial for mapping detected anomalies to potential threats and coordinating timely responses. This literature study explores the state-of-the-art technologies and methodologies in AI-based anomaly detection and threat mapping.

2.1.1.1 Edge Computing

Cloud Computing is an emerging technology that allows machines/people to access data anywhere and everywhere. Edge Computing is a new paradigm that moves computing application and services from central units to logical endpoints or locations closest to the source and provides data processing power there[1].

Edge computing architecture is typically categorized into three components: the front-end, near-end, and far-end.

Front-end: End devices such as sensors and actuators are situated at the front-end of the edge computing framework. This environment enhances interaction and responsiveness for users. The computational power available from numerous nearby end devices enables edge computing to deliver real-time services for certain applications. However, the limited capabilities of these

[1] Yu, W., Liang, F., He, X., Hatcher, W.G., Lu, C., Lin, J., Yang, X. A Survey on the Edge Computing for the Internet of Things, Department of Computer and Information Sciences, Towson University, MD, USA, School of Electronic and Information Engineering, Xi’an Jiaotong University, Shaanxi, P.R. China.

devices mean that not all demands can be met at the front-end. Consequently, these end devices often need to pass on their resource requirements to servers for further processing.

Near-end: In a near-end environment, gateways will handle the majority of network traffic. Edge or cloudlet servers in this setting have extensive resource requirements, including real-time data processing, data caching, and computation offloading. In the context of edge computing, much of the data computation and storage tasks will be shifted to this near-end environment. This shift enables end users to experience significantly improved performance in data computing and storage, albeit with a minor increase in latency.

Far-end: Cloud servers situated at a greater distance from end devices contribute to considerable transmission latency within the networks. Despite this, far-end cloud servers offer superior computing power and data storage capabilities. For instance, they can support massive parallel data processing, big data mining, big data management, and machine learning. The architecture of edge computing networks is illustrated in the figure below.

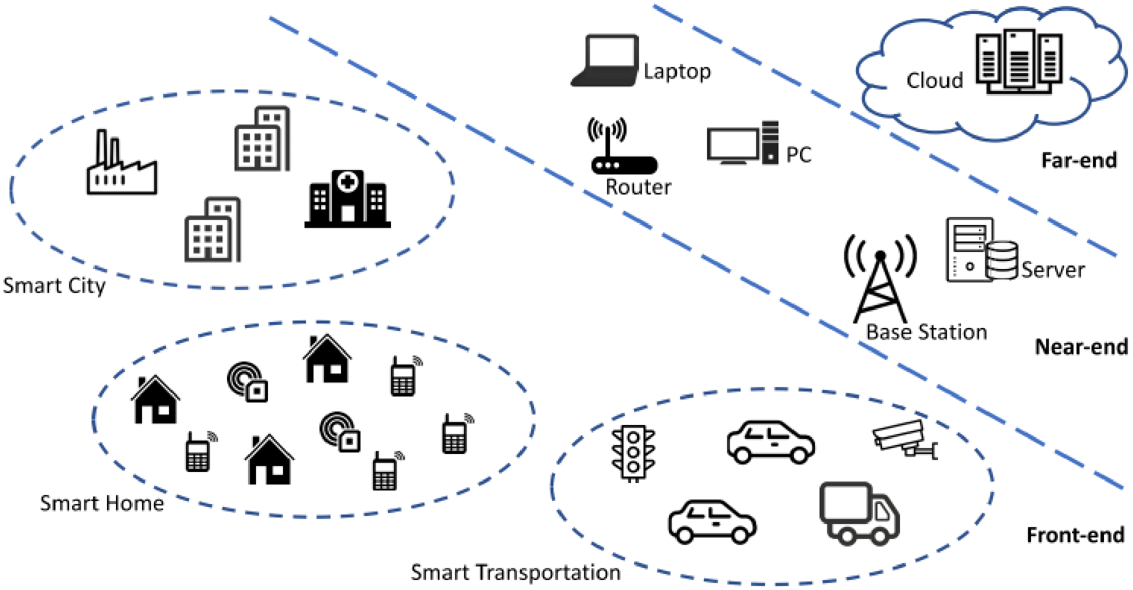


Figure 1. Edge computing network architecture[2]

Edge Computing integrates a wide array of technologies, bringing them together to create a cohesive system. Within this field, it leverages various technologies, such as wireless sensor networks (WSN), mobile data acquisition, mobile signature analysis, Fog/Grid Computing,

[2] Yu, W., Liang, F., He, X., Hatcher, W.G., Lu, C., Lin, J., Yang, X. A Survey on the Edge Computing for the Internet of Things, Department of Computer and Information Sciences, Towson University, MD, USA, School of Electronic and Information Engineering, Xi’an Jiaotong University, Shaanxi, P.R. China.

distributed data operations, and remote Cloud services. Additionally, it incorporates the following protocols and terms[3]:

- 5G communication: This is the fifth-generation wireless system designed to offer higher capacity, lower power consumption, and reduced latency compared to its predecessors. As the volume of data and the number of connected devices grow, 5G is expected to address traffic issues.
- PLC protocols: Object Linking and Embedding for Process Control Unified Architecture (OPC-UA) is a protocol created for industrial automation. It is highly valued in industries such as oil and gas, pharmaceuticals, robotics, and manufacturing due to its openness and robustness.
- Message queue broker: Protocols like MQTT and TCP/IP are popular among smart sensors and IoT devices. By supporting these message brokers, Edge Computing can connect more devices. To enhance MQTT security, AMQP is used for communication with Cloud Computing servers.
- Event processor: When IoT messages reach the Edge server, the event processor analyses them and generates semantic events based on predefined rules. Examples of this enabler include EsperNet, Apache Spark, and Flink.
- Virtualization: Cloud services are deployed as virtual machines on Cloud servers or clusters, allowing multiple operating system instances to run on the same server.
- Hypervisor: In addition to virtual machines, performance evaluation and data handling are managed by the hypervisor, which controls virtual machines on the host computer.
- OpenStack: Managing multiple resources can be challenging. OpenStack is a Cloud operating system that simplifies the management of computing and storage resources through a control panel and monitoring tools.
- AI platform: Rule-based engines and machine learning platforms support local data analysis. As mentioned in Section IV, this is crucial for achieving one of the goals of Edge Computing: gathering, analysing, and initially filtering data.
- Hyperledger: Blockchain technology is commonly used in highly sensitive areas, such as digital currencies like Bitcoin. It is also valuable for data protection in Cloud Computing, enabling secure data sharing with external parties and servers.
- Docker: Unlike virtual machines that require installing operating systems, Docker is a Container as a Service (CaaS) that uses a single shared operating system to run software in an isolated environment. It only needs the software libraries, making it a lightweight system without concerns about the software's deployment location.

[3] Gezer, V., Um, J., & Ruskowski, M. (2017). An extensible edge computing architecture: Definition, requirements and enablers. Proceedings of the UBICOMM.

2.1.1.2 Fog Computing

The Internet of Things (IoT) encompasses on-device computing at the intermediate layer between edge devices and the cloud. Transferring data to the cloud in a Fog Computing setup involves multiple stages, addressing complexities and data transformations along the way. Fog computing is particularly suitable for critical applications such as data collection and pre-processing, condition monitoring, rule-based decision-making, and short-term data storage that require real-time responsiveness and are time-sensitive. For instance, autonomous vehicles have a maximum latency tolerance of about 10 milliseconds, but this tolerance is even lower in scenarios where human safety is critical and the economic stakes are high. In high-frequency stock market trading, institutions have a latency tolerance of 0.25 seconds, whereas ensuring worker safety in a mine requires a latency tolerance of just 1 millisecond. Therefore, fog computing is essential in these critical situations to minimize the risk of communication failures, enhance real-time analysis and decision-making speed, and reduce the costs associated with transmitting, processing, and storing data in the cloud. In essence, fog computing aims to bring computational power closer to data sources to minimize response times without compromising throughput. In a fog computing model, computing tasks are distributed efficiently between the data source and the cloud. While fog computing is locally closer to devices compared to the cloud, it acts as an intermediary layer, forwarding information even when some decisions are made locally.[4]. Fog architecture is given in the figure below.

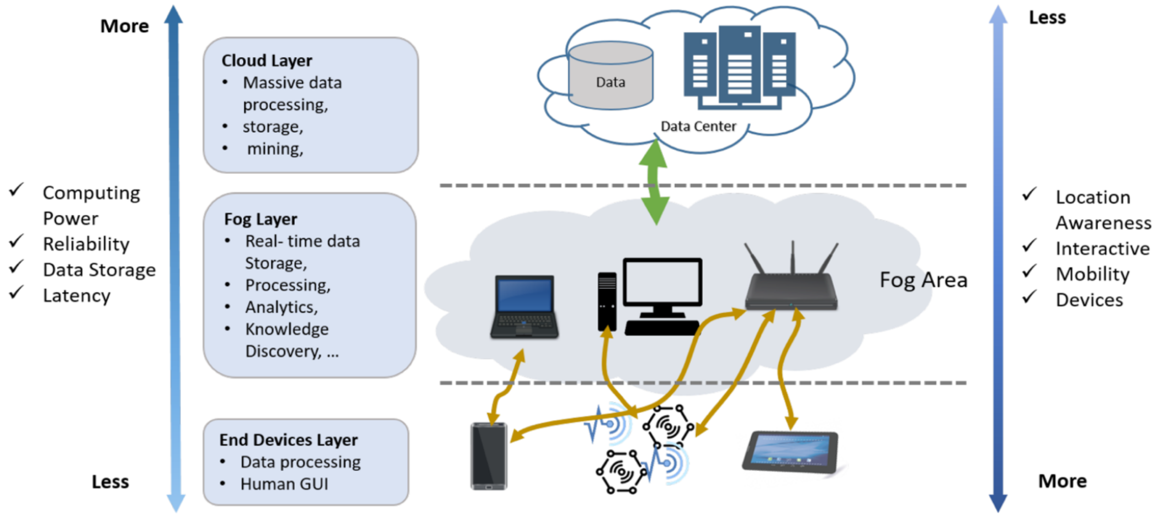


Figure 2. Fog computing architecture[5]

[4] Fog Computing: Current Research and Future Challenges, March 2018. Conference: 1. GI/ITG KuVS Fachgespräche Fog ComputingAt: Darmstadt, Germany.

[5] Fog Computing: Current Research and Future Challenges, March 2018. Conference: 1. GI/ITG KuVS Fachgespräche Fog ComputingAt: Darmstadt, Germany.

2.1.1.3 Human Action Recognition (HAR)

Human Action Recognition (HAR) is a field within computer vision and artificial intelligence dedicated to identifying and categorising human actions from various observations. This technology analyses data captured from sources like videos, wearable sensors, or environmental sensors to recognize and predict human activities. HAR integrates aspects of computer vision, machine learning, and data analytics to understand human movements and behaviours, with applications spanning across human-computer interaction (HCI), surveillance, virtual reality (VR), and elder care.

HAR systems gather data through cameras or sensors that capture human movements. The raw data is then processed to extract significant features that represent different human actions, such as identifying body postures, movements, or gestures. Machine learning algorithms use these extracted features to classify actions into categories like walking, running, or jumping[6]. Some HAR systems also consider the context, such as the environment or interaction with objects, to enhance accuracy.

The technology of human action recognition[7] employs data from specific algorithmic sensors to identify types of human actions. It has become crucial in various fields, including artificial intelligence, due to its technical applications and potential for development. HAR shows promise in safety state monitoring[8], behaviour feature analysis, and network video image restoration. Through video surveillance, the technology of recognizing human motion can intelligently detect abnormal behaviours, such as fighting and illegal tracking. These behaviours may cause harm to personal safety, so monitoring can be used for timely detection and early warning. In the aspect of behaviour analysis, the most common technology to identify human movement[9] is to compare the athlete’s movement with that of standard athletes and to improve the accuracy of movement.

In behaviour analysis, HAR technology is used to compare an athlete’s movements with those of standard athletes to improve accuracy. Despite the high status of deep learning in HAR, ongoing research has revealed limitations in the accuracy and recognition rates of deep learning-based human action recognition. The attention mechanism aims to address these issues by allocating limited computing resources to high-priority information. With the abundance of raw information, computational difficulties can arise, leading to decreased accuracy. Researchers

[6] Zehua Sun, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, Gang Wang, Jun Liu, Human Action Recognition from Various Data Modalities: A Review, Computer Vision and Pattern Recognition.

[7] Qiong H., Lei Q., Qingming H.. 2013.Overview of Human Action Recognition Based on Vision. Chinese Journal of Computers,12(12),2512-2524.

[8] Bin F., Xin F., Jianguo C.. 2021. A MEMS sensor-based human body gesture recognition method for the elderly-aiding mechanism. Journal of Harbin University of Commerce (Natural Sciences Edition),37(05),590-594.

[9] Zhaole D., Kang W., Shenglong L.. 2021. Human action recognition based on deep learning. Command Informatipn System and Technology,12(04),70-74.

suggest using the attention mechanism to improve the accuracy and efficiency of human action recognition.

2.1.1.3.1 HAR Current Technologies and Trends

HAR has a wide range of applications, including in healthcare for patient monitoring, in security systems for surveillance, in sports for performance analysis, and in entertainment for interactive gaming[10].

- Daily activities, such as walking, running, jumping, sitting, standing, etc.
- Sports activities, such as basketball, soccer, tennis, etc.
- Exercise activities, such as weightlifting, yoga, aerobics, etc.
- Medical activities, such as gait analysis for patients with mobility impairments.
- Industrial activities, such as assembly line work, machine operation, etc.
- Interpersonal activities, such as handshaking, hugging, pointing, etc.
- Artistic activities, such as dancing, playing musical instruments, etc.
- Household activities, such as cooking, cleaning, etc.
- HAR can be used for patient monitoring, fall detection for the elderly, and assisting in physical therapy by analysing movements to ensure exercises are performed correctly.
- Surveillance and Security: In security systems, HAR can detect unusual or suspicious activities, enhancing public safety and security monitoring.
- By recognizing the actions of residents, HAR can automate home appliances and lighting, contributing to energy savings and personalized living experiences.
- Coaches and athletes can use HAR to analyse movements for improving performance and technique in sports training.
- HAR enables interactive gaming experiences where players' physical actions are translated into game movements[11].
- In manufacturing, HAR can ensure safety by detecting improper actions that could lead to accidents, and it can also assist in automating repetitive tasks[12].
- Teachers can use HAR to analyse student engagement and participation during classes.
- HAR can help in understanding customer behaviour and preferences, leading to improved customer service and targeted marketing[13].

[10] Md Golam Morshed,Tangina Sultana, Aftab Alam and Young-Koo Lee, Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities.

[11] Shuchang Zhou, Computer Vision and Pattern Recognition, A Survey on Human Action Recognition.

[12] Zehua Sun, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, Gang Wang, Jun Liu, Human Action Recognition from Various Data Modalities: A Review, Computer Vision and Pattern Recognition.

[13] Md Golam Morshed,Tangina Sultana,Aftab Alam andYoung-Koo Lee, Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities.

- In advanced driver-assistance systems (ADAS), HAR can monitor driver behaviour to detect signs of drowsiness or distraction[14].

2.1.1.3.2 HAR Implementation Challenges

Despite having received great attention from various researchers from various groups of interests, the challenges in solving real-world HAR problems still remain. Some of these challenges are due to background clutter, changes in lighting and illumination, occlusions, be it self-occlusion or with other objects, camera view-dependent, frame resolution, differences in scale and appearance, as well as the nature of the action itself. Additionally, the inter- and intra-variations in human action add another dimension of complexity to the problem. The challenges listed highlight the need for continued research and development in the field of HAR to create more accurate, efficient and ethical recognition systems. The challenges in solving HAR problems still remain[15],[16],[17],[18],[19],[20],[21],[22],[23],[24].

Challenge 1: Natural variability in human actions is one of the important factors affecting HAR processes. Human actions can vary greatly in speed, style, and execution, making it difficult for algorithms to recognize and classify them accurately.

Challenge 2: Viewpoint and biometric variability are another challenge. Changes in camera, radar and sensor viewport, as well as different body shapes, sizes and appearances, can lead to significant differences, affecting the performance of recognition algorithms.

[14] Gupta, N., Gupta, S.K., Pathak, R.K. et al. Human activity recognition in artificial intelligence framework: a narrative review.

[15] Shuchang Zhou, Computer Vision and Pattern Recognition, A Survey on Human Action Recognition.

[16] Zehua Sun, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, Gang Wang, Jun Liu, Human Action Recognition from Various Data Modalities: A Review, Computer Vision and Pattern Recognition.

[17] Md Golam Morshed,Tangina Sultana,Aftab Alam andYoung-Koo Lee, Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities.

[18] Gupta, N., Gupta, S.K., Pathak, R.K. et al. Human activity recognition in artificial intelligence framework: a narrative review.

[19] Muhammad Haseeb Arshad,Muhammad Bilal andAbdullah Gani, Human Activity Recognition: Review, Taxonomy and Open Challenges.

[20] Pareek, P., Thakkar, A. A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications.

[21] Othman, N.A., Aydin, I. (2021). Challenges and limitations in human action recognition on unmanned aerial vehicles: A comprehensive survey.

[22] Kumar, P., Chauhan, S. & Awasthi, L.K. Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies.

[23] Progress and Future Research Directions. Arch Computat Methods Eng.

[24] Singh, P.K., Kundu, S., Adhikary, T. et al. Progress of Human Action Recognition Research in the Last Ten Years: A Comprehensive Survey. Arch Computat Methods Eng

Challenge 3: In real-world scenarios, the presence of multiple people, obstacles, and dynamic backgrounds increases the complexity of accurately recognizing actions. Data obtained with cameras, radar and sensors may provide incorrect or incomplete information due to the complexity of the environment. High-quality, labelled datasets are crucial for training recognition systems, but they can be scarce or unstable, hindering the development of robust HAR models. People perform the same action in different ways. HAR systems must be resilient to changes in action execution due to factors such as clothing, lighting conditions, and occlusions.

Challenge 4: Combining data from different sources, such as videos, sensors, and still images, can improve recognition accuracy but also increases the complexity of the system[25].

Challenge 5: The current approaches employ multiple branches, analysing different features to produce richer and more robust information. On the other hand, some methods employ backbone networks for the initial feature extraction (temporal or regional), dividing both the training and inference process into a two-stage process each. Despite its high effectiveness, the inference time is sacrificed.

Challenge 6: The video annotation process becomes an extremely exhausting task concerning the unpredictable number of video hours needed to successfully train a model. Therefore, there is a need for semi-supervised and unsupervised learning algorithms to recognize human actions. The increasing number of action classes becomes even more challenging due to the higher overlapping between classes.

Challenge 7: One of the challenges in HAR is the intra-class variation of an action. This is to say that the same action performed by the same person but viewed at different camera angles may result in different extracted features, resulting in dissimilar feature descriptors. In order to circumvent this problem, some researchers have resorted to multi-view, also known as cross-view approach.

Challenge 8: An image-based recognition system's primary challenge is lighting fluctuation, which has an impact on the quality of pictures and, thus, on the information that is processed.

Challenge 9: While deep learning has significantly advanced HAR, there is still a need to develop models that can handle the vast amount of data and complex patterns associated with human actions[26].

[25] Pareek, P., Thakkar, A. A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications.

[26] Othman, N.A., Aydin, I. (2021). Challenges and limitations in human action recognition on unmanned aerial vehicles: A comprehensive survey.

Challenge 10: As HAR systems often involve surveillance, there are ethical considerations and privacy concerns that need to be addressed, ensuring that such technologies are used responsibly[27].

2.1.1.4 Emotion Detection

Emotion detection technologies have rapidly evolved, significantly impacting various sectors including transportation and customer service. The state-of-the-art review specifically explores recent advancements in AI-based emotion detection within airport settings, focusing on enhancing customer satisfaction and operational efficiency.

Recent research demonstrates a shift towards integrating AI technologies to analyse facial expressions, speech patterns, and even contextual behaviours to gauge customer emotions in real-time. For instance, Gerard Deepak and A Santhanavijayan[28] have developed a semantic AI model that incorporates the Normalized Pointwise Mutual Information (NPMI) measure for intelligent response generation in dynamic environments like airports, which could significantly enhance customer service interactions. Raj Deshmukh et al.[29] employed temporal logic learning to monitor anomalies in terminal airspace operations, indirectly facilitating emotion detection by ensuring smoother operations and reducing passenger stress caused by delays or other irregularities. In an innovative approach, Asad Abbas and Stephan Chalup[30] explored the affective analysis of visual scenes using face pareidolia (seeing faces in inanimate objects) and scene context, an approach that can be adapted to monitor passenger satisfaction in visually complex environments such as airports.

The role of speech in emotion detection cannot be understated, as demonstrated by Md. Zia Uddin and Erik G Nilsson[31], who applied neural structured learning to enhance emotion recognition from passenger speech. This technology could be pivotal in customer service desks and announcements in airports, providing real-time emotional feedback to service providers.

[27] Kumar, P., Chauhan, S. & Awasthi, L.K. Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies.

[28] Deepak, G., & Santhanavijayan, A. (2020). A Novel Semantic Approach for Intelligent Response Generation using Emotion Detection Incorporating NPMI Measure. Procedia Computer Science, 168, 126-133

[29] Deshmukh, R., Sun, D., Kim, K., & Hwang, I. (2021). Temporal logic learning-based anomaly detection in metroplex terminal airspace operations. Transportation Research Part C: Emerging Technologies, 124, 102955.

[30] Abbas, A., & Chalup, S. (2021). Affective analysis of visual scenes using face pareidolia and scene-context. Neurocomputing, 423, 634-645.

[31] Uddin, M. Z., & Nilsson, E. G. (2020). Emotion recognition using speech and neural structured learning to facilitate edge intelligence. Engineering Applications of Artificial Intelligence, 94, 103789.

Lastly, understanding passenger hesitancy and decision-making is crucial for improving airport services, a challenge tackled by Jing Lu et al.[32] through machine learning models. These models predict passenger choices based on emotional and rational factors, providing insights into enhancing overall passenger. These studies collectively indicate a growing trend of employing advanced AI techniques to not only detect but also respond to emotional cues in a high-stake environment like airports, ultimately aiming to boost passenger satisfaction and streamline airport operations.

2.1.1.5 Object Detection

Object detection focuses on identifying specific instances of objects within real-world data, presenting a significant challenge in computer vision. This process involves two primary components: object localization and classification, which together facilitate the extraction of objects from images. Essentially, computer vision accomplishes this by distinguishing specific objects from the background, determining their respective classes, and outlining the proposed object boundaries.

Object detection builds upon object classification, which solely aims to recognize objects within an image. The objective of object detection is to identify all instances of predefined classes and provide a rough localization in the image using axis-aligned bounding boxes. The detection system should be capable of recognizing all instances of the object classes and drawing a bounding box around each. This task is typically approached as a supervised learning problem. Contemporary object detection models are trained using large sets of labelled images and are assessed on various standard benchmarks.

Early object detection models were built as an ensemble of hand-crafted feature extractors such as Viola-Jones detector[33], Histogram of Oriented Gradients (HOG)[34] etc. These models were slow, inaccurate and performed poorly on unfamiliar datasets. The usage of the convolutional neural network (CNNs) and deep learning for image classification changed the landscape of visual perception. Its use in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012

[32] Lu, J., Meng, Y., Timmermans, H., & Zhang, A. (2021). Modeling hesitancy in airport choice: A comparison of discrete choice and machine learning methods. Transportation Research Part A: Policy and Practice, 146, 102-117.

[33] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, IEEE Comput. Soc., 2001, pp. I-511–I-518.

[34] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, ISSN1063-6919, 2005-06, pp.886–893.

challenge by AlexNet[35] inspired further research on its application in computer vision. Today, object detection finds applications from self-driving cars and identity detection to security and medical uses. In recent years, it has seen exponential growth with the rapid development of new tools and techniques.

The development of object detection algorithms can be divided into two stages: traditional object detection algorithms and deep-learning-based object detection algorithms. Deep-learning-based object detection algorithms are further divided into two main technical routes: one-stage and two-stage algorithms[36]. Figure 2 shows the development of object detection from 2001 to 2023[37].

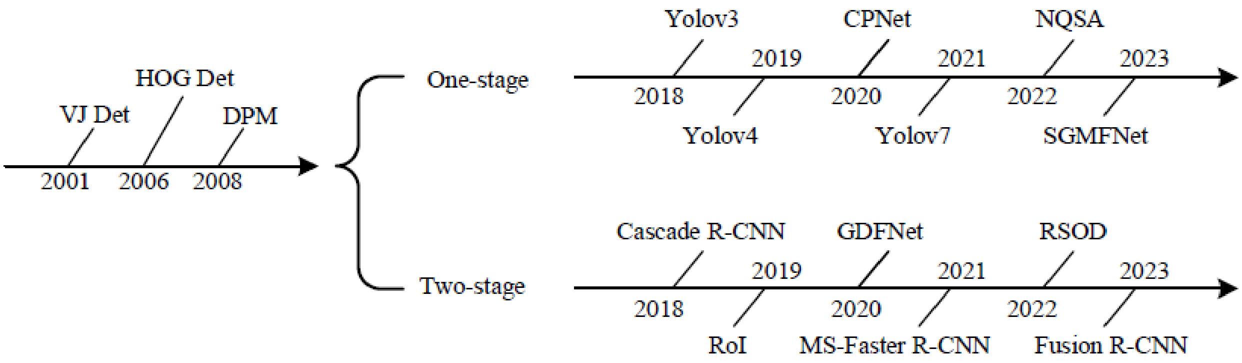


Figure 3. The development of object detection from 2001 to 2023.

Traditional object detection algorithms primarily rely on sliding window techniques and manual feature extraction methods, typically involving three steps: region proposal, feature extraction, and classification regression. The region proposal step identifies regions of interest where objects are likely located. During the feature extraction phase, manual methods are used to convert images in candidate regions into feature vectors. Finally, a classifier categorizes objects based on the extracted features. However, these algorithms are often hampered by high computational complexity, limited feature representation capability, and optimization challenges.

Object detection offers extensive real-world applications. It involves localizing and classifying objects within images or videos, driving research into several key areas:

[35] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc., 2012, p.9.

[36] Xu, X.; Dong, S.; Xu, T.; Ding, L.; Wang, J.; Jiang, P.; Song, L.; Li, J. FusionRCNN: LiDAR-Camera Fusion for Two-Stage 3D Object Detection. Remote Sens. 2023, 15, 1839.

[37] Guangyi T., Jianjun N., Yonghao Z., Yang G., Weidong C., A Survey of Object Detection for UAVs Based on Deep Learning, Remote Sensing, 2024, 16, 149.

- Object Localization[38]: Crucial for precise object positioning, enabling applications like autonomous driving and pedestrian detection.
- Object Classification[39]: Essential for recognizing object types, aiding tasks such as automatic product tagging in e-commerce applications.
- Counting of Objects: Extending detection to quantify instances within images or videos, valuable for crowd management and inventory tracking.
- Object Tracking and Monitoring[40]: Tracking objects across frames in videos, vital for surveillance and movement analysis.
- Improving Accuracy and Efficiency[41]: Ongoing efforts enhance detection algorithms' precision and speed through novel architectures and optimization techniques.
- Adapting to Challenging Conditions: Algorithms are being fortified to perform well under adverse conditions like varying lighting and occlusions, critical for applications in unpredictable environments like robotics[42].
- Deployment in Edge Devices: Focus on deploying efficient models on resource-constrained devices, requiring advancements in compression and quantization techniques[43].
- Domain Adaptation and Transfer Learning[44]: Techniques explore adapting models to new domains with limited data, leveraging pre-trained models to improve performance.

2.1.1.5.1 Object Detection Challenges and Considerations

Object detection typically addresses two fundamental questions: "What is the object?" and "Where is the object?" Initially, object classification and localization posed significant challenges, but advancements in computer vision have enabled digital devices to identify image contents. However, despite notable progress, object detection encounters persistent hurdles, including dual priorities, limited data, class imbalances, variations in size, speed constraints, environmental factors, and handling multiple scales.

[38] BRESSON, Guillaume, et al. Simultaneous localization and mapping: A survey of current trends in autonomous driving. IEEE Transactions on Intelligent Vehicles, 2017, 2.3: 194-220.

[39] KEJRIWAL, Mayank, et al. An evaluation and annotation methodology for product category matching in e-commerce. Computers in Industry, 2021, 131: 103497.

[40] ELHARROUSS, Omar; ALMAADEED, Noor; AL-MAADEED, Somaya. A review of video surveillance systems. Journal of Visual Communication and Image Representation, 2021, 77: 103116.

[41] ZHAO, Zhong-Qiu, et al. Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 2019, 30.11: 3212-3232.

[42] MARTINEZ-MARTIN, Ester; DEL POBIL, Angel P. Object detection and recognition for assistive robots: Experimentation and implementation. IEEE Robotics & Automation Magazine, 2017, 24.3: 123-138.

[43] SHUVO, Md Maruf Hossain, et al. Efficient acceleration of deep learning inference on resource-constrained edge devices: A review. Proceedings of the IEEE, 2022, 111.1: 42-91.

[44] KAMATH, Uday, et al. Transfer learning: Domain adaptation. Deep learning for NLP and speech recognition, 2019, 495-535.

Numerous researchers have dedicated efforts to overcoming these obstacles, yielding notable results, yet challenges persist.

Challenge 1: Real-world images present diverse variations such as lighting conditions, occlusions, noise, camera distortions, and background clutter, making the detection of small objects against complex backgrounds particularly demanding.

Challenge 2: Techniques employing image pyramids[45] facilitate the effective detection of objects of varying sizes. Despite recent strides, accurately detecting small objects remains a challenge in object detection.

Challenge 3: Other factors impacting detection quality include training strategies, backbone model selection, improving loss functions, and addressing imbalances between positive and negative samples. While numerous architectures have been proposed to tackle these challenges, achieving real-time performance comparable to human capability remains elusive. Extensive research efforts[46],[47],[48],[49] have been directed toward mitigating these challenges across various application domains in object detection.

Challenge 4: Intra-class variation between the instances of the same object is relatively common in nature. This variation could be due to numerous reasons like occlusion, illumination, pose, viewpoint, etc. These unconstrained external factors can have a dramatic effect on the object's appearance[50]. It is expected that the objects could have non-rigid deformation or be rotated, scaled, or blurry. Some objects could have inconspicuous surroundings, making the extraction difficult.

Challenge 5: The sheer number of object classes available to classify makes it a challenging problem to solve. It also requires more high-quality annotated data, which is hard to come by. Using fewer examples to train a detector is an open research question.

[45] MEER, Peter. Stochastic image pyramids. Computer Vision, Graphics, and Image Processing, 1989, 45.3: 269-294.

[46] Tamilselvi M, Karthikeyan S (2022, Elsevier) An ingenious face recognition system based on HRPSM_CNN under unrestrained environmental condition. Alexandria Eng J 61(6):4307–4321.

[47] Naiemi F, Ghods V, Khalesi H (2021, Elsevier Ltd) A novel pipeline framework for multi oriented scene text image detection and recognition. Expert Syst Appl 170(2020):114549.

[48] Ma C, Sun L, Zhong Z, Huo Q (2021) ReLaText: exploiting visual relationships for arbitrary-shaped scene text detection with graph convolutional networks. Pattern Recogn 111:107684.

[49] Lu X, Ji J, Xing Z, Miao Q (2021) Attention and feature fusion SSD for remote sensing object detection. IEEE Trans Instrum Meas 70

[50] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: a survey, Version: 1, arXiv:1809 .02165, 2018.

Challenge 6: Present-day models need high computation resources to generate accurate detection results. With mobile and edge devices becoming commonplace, efficient object detectors are crucial for further development in the field of computer vision.

2.1.1.6 Speech Recognition

The integration of advanced speech recognition technologies in airport security frameworks has significantly enhanced the efficiency and effectiveness of threat detection and access control systems. This detailed review presents a comprehensive analysis of the current technical and technological state of speech recognition applications within airport security, supported by recent scholarly contributions.

Speech recognition technologies are employed in critical airport operations, including biometric voice authentication for access controls and real-time monitoring of communications in air traffic control. Cornacchia et al.[51] explored the application of voice biometrics in secure area access management, highlighting the technical process where voice samples are analysed using spectral feature extraction techniques to match live or recorded utterances against a database of known voice prints, thus securing access to sensitive areas. Addressing vulnerabilities within automatic speech recognition (ASR) systems is crucial due to the sensitive nature of their application. Chen et al.[52] delved into the modular security measures necessary to fortify ASR systems against adversarial attacks, including the use of noise-injection and signal distortion techniques that enhance system robustness by reducing the efficacy of mimicry and spoofing attacks. Abdullah et al.[53] further analysed the attack vectors specific to ASR systems, discussing the deployment of cryptographic voice encapsulation to safeguard data transmission between users and security systems.

Technological advancements focus on enhancing the security layers within voice-driven systems. Zhang et al.[54] introduced a method for embedding cryptographic hashes in transient speech tokens, which prevents the unauthorized reuse of captured speech, ensuring that voice commands cannot be replayed to gain illicit access. Continuous voice authentication systems, as

[51] Cornacchia, M., Papa, F., & Sapio, B. (2020). User acceptance of voice biometrics in managing the physical access to a secure area of an international airport. Technology Analysis & Strategic Management, 32(5), 585-598.

[52] Chen, Y., Zhang, J., Yuan, X., Zhang, S., Chen, K., Wang, X., & Guo, S. (2022). SoK: A Modularized Approach to Study the Security of Automatic Speech Recognition Systems. arXiv preprint arXiv:2103.10651v2.

[53] Abdullah, H. Z., Warren, K., Bindschaedler, V., Papernot, N., & Traynor, P. (2021). SoK: The Faults in our ASRs: An Overview of Attacks against Automatic Speech Recognition and Speaker Identification Systems. arXiv preprint arXiv:2007.06622v3.

[54] Zhang, Y., Arora, S. S., Shirvanian, M., Huang, J., & Gu, G. (2021). Practical Speech Re-use Prevention in Voice-driven Services. arXiv preprint arXiv:2101.04773v1.

explored by Zhang and Yang[55], implement real-time voice pattern analysis using dynamic time warping and machine learning algorithms to detect anomalies in speech that may indicate a security breach. Looking forward, the enhancement of speech recognition accuracy in noisy environments remains a critical research area. Fan et al.[56] investigated the application of convolutional neural networks (CNNs) for real-time language identification, which adapts speech recognition algorithms based on the detected language to improve accuracy in the linguistically diverse environment of international airports. Additionally, ongoing research by Li et al.[57] into securing voice assistant applications emphasizes the development of decentralized ASR systems that utilize blockchain technology to ensure data integrity and prevent unauthorized access.

2.1.1.7 Anomaly Detection

AI techniques, such as machine learning and deep learning, are employed to detect anomalies in the fused data. These methods can identify hidden, complex, or context-dependent patterns indicative of potential threats[58],[59]. Once anomalies are detected, they are mapped to specific threats using AI models. The system can then coordinate timely responses, including alerts, lockdowns, and dispatching emergency services[60],[61]. Machine learning (ML) techniques are widely used for anomaly detection due to their ability to learn from data and improve over time. Common ML methods include clustering, classification, and regression[62],[63].

- Clustering: Techniques like k-means and density-based spatial clustering of applications with noise (DBSCAN) are used to identify outliers in data by grouping similar data points and detecting those that do not fit into any cluster.

[55] Zhang, L., & Yang, J. (2021). A Continuous Liveness Detection for Voice Authentication on Smart Devices. arXiv preprint arXiv:2106.00859v1.

[56] Fan, P., Guo, D., Zhang, J., Yang, B., & Lin, Y. (2023). Enhancing multilingual speech recognition in air traffic control by sentence-level language identification. arXiv preprint arXiv:2305.00170v1.

[57] Li, J., Chao, C., Pan, L., Azghadi, M. R., Ghodosi, H., & Zhang, J. (2023). Security and Privacy Problems in Voice Assistant Applications: A Survey. arXiv preprint arXiv:2304.09486v1.

[58] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1-58.

[59] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.

[60] He, W., Yan, G., & Xu, L. D. (2014). Developing vehicular data cloud services in the IoT environment. IEEE Transactions on Industrial Informatics, 10(2), 1587-1595.

[61] Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. Business Horizons, 58(4), 431-440.

[62] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1-58.

[63] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.

- Classification: Supervised learning algorithms such as support vector machines (SVM) and random forests classify data points into normal and anomalous categories.
- Regression: Algorithms like linear regression and decision trees predict values and detect deviations from expected patterns.

Deep learning, a subset of machine learning, involves neural networks with multiple layers that can model complex data patterns. Techniques such as autoencoders, convolutional neural networks (CNNs), and recurrent neural networks (RNNs) are particularly effective in anomaly detection[64],[65].

- Autoencoders: Used for unsupervised anomaly detection by learning a compressed representation of normal data and identifying anomalies based on reconstruction errors.
- CNNs: Effective in detecting spatial anomalies, particularly in image and video data.
- RNNs: Suitable for detecting temporal anomalies in time-series data due to their ability to capture sequential dependencies.

2.1.2 Multi-Modal Sensing Technologies

2.1.2.1 Acoustic Sensors

Acoustic sensors detect sound waves and are used in various security applications, including gunshot detection, intrusion detection, and environmental monitoring. Advanced signal processing algorithms analyse acoustic data to identify specific sounds and their sources[66],[67].

2.1.2.1.1 Acoustic Data Analysis

Acoustic data analysis involves extracting meaningful information from sound waves, which can be crucial in various fields such as product design, environmental monitoring, and speech recognition.

Sound level: Sound level, often measured in decibels (dB), quantifies the intensity or loudness of a sound. It provides insight into how much energy a sound wave carries. The A-weighted decibel

[64] Kwon, D., Kim, K., & Lee, H. (2019). A survey of deep learning-based network anomaly detection. Cluster Computing, 22(1), 1-13.

[65] Zhao, Y., Nasrullah, Z., & Li, Z. (2019). PyOD: A python toolbox for scalable outlier detection. Journal of Machine Learning Research, 20(96), 1-7.

[66] Ntalampiras, S. (2017). Audio pattern recognition of acoustic events for surveillance applications. Pattern Recognition Letters, 85, 5-11.

[67] Valenzise, G., Gerosa, L., Tagliasacchi, M., Antonacci, F., & Sarti, A. (2007). Scream and gunshot detection and localization for audio-surveillance systems. IEEE Conference on Advanced Video and Signal Based Surveillance, 21-26.

scale (dBA) is commonly used to account for human perception of sound. It emphasizes frequencies relevant to human hearing.

Spectrum: The frequency spectrum represents the distribution of energy across different frequencies in a sound signal. Fast Fourier Transform (FFT) is a common technique used to convert a time-domain signal into its frequency-domain representation. The resulting spectrum shows the contribution of each frequency component.

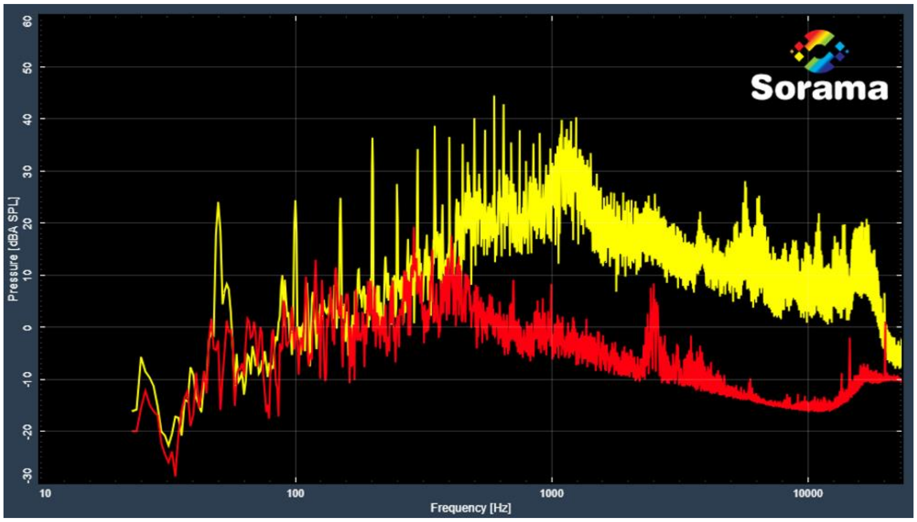


Figure 4. Comparison of the spectra between a background noise measurement (in red) and a coffee machine in operation (in yellow). The coffee machine emits sharp tones below 800 Hz (as seen from the sharp peaksm, e.g. at 100 Hz). From 800 Hz onwards, the coffee machine also exhibits broadband noise that is consistently approximately 20dB higher level than background noise. Analysis made using the Sorama Portal.

Spectrogram: A spectrogram is a visual representation of how the frequency content of a signal changes over time. It displays the spectrum of short segments of a signal (usually overlapping) as a function of time. In a spectrogram, time is on the horizontal axis, frequency on the vertical axis, and colour intensity represents the energy at each frequency.

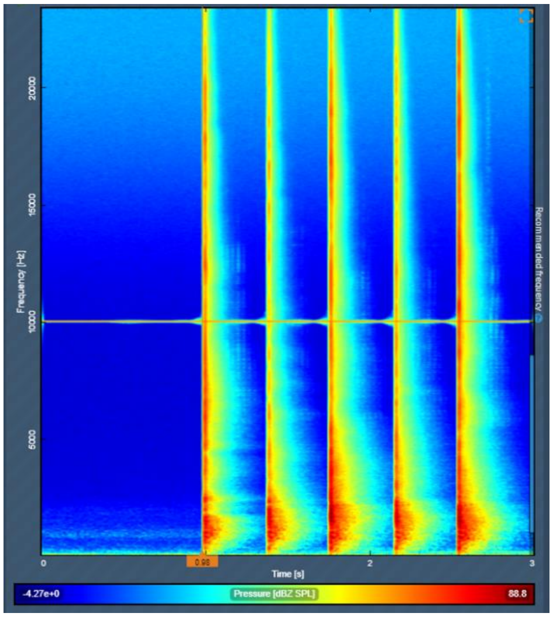


Figure 5. Example of a spectrogram demonstrating a steady tone at 10 kHz (indicated by the horizontal line at 10 kHz) that persists throughout the 3 seconds, and also five impulses (indicated by the vertical features that begins from 1 second). The colorbar at the bottom displays the sound level strength. Notably, the energy of the impulse decays slower in the low frequencies as it stays red for longer compared to the higher frequencies' components. Analysis made using the Sorama Portal.

Far Field Beamforming: Beamforming is a technique used to enhance the directivity of microphones or loudspeakers. In the far field, where the distance from the sound source is significant compared to the wavelength, beamforming aims to focus on a specific direction. Algorithms like Delay-and-Sum Beamforming adjust the phase and amplitude of microphone signals to create a "beam" pointing toward the desired source.

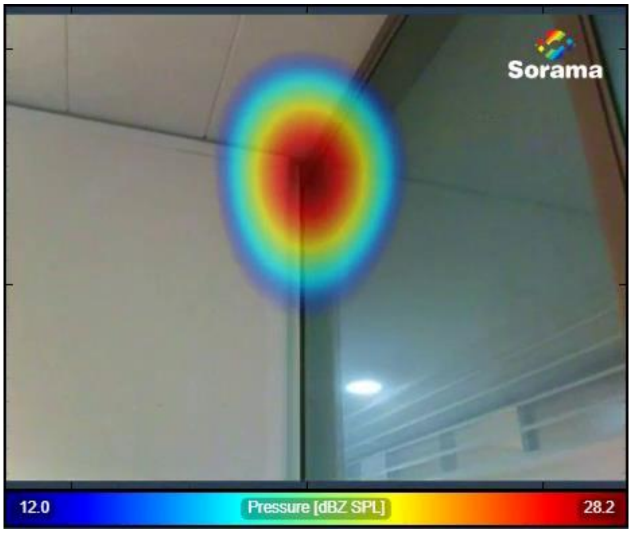


Figure 6. A beamformed acoustic power map overlaid on top of a visual camera demonstrates that there is a sound coming from the corner of the room. This shows that the sound isolation is poor in the corner, causing sound of adjacent room to leak through that spot. Analysis made using the Sorama Portal with recording made by the Sorama CAM iV64 acoustic camera.

Near field holography reconstructs the sound field close to a source (near field) using measurements taken on a surface (such as a microphone array). It's particularly useful for analysing sound radiation from complex structures (e.g., car engines, musical instruments). By solving the inverse problem, near field holography provides insights into the spatial distribution of sound sources.

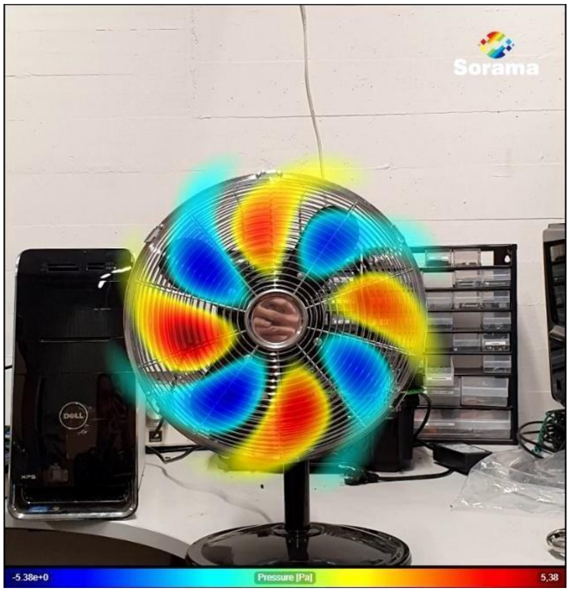


Figure 7. Holography image which reveals the sound field caused by pressure fluctuation at the blade-pass frequency of the table fan. Analysis made using the Sorama Portal.

2.1.2.1.2 Acoustic Surveillance

Acoustic surveillance involves the use of auditory sensory information to monitor and detect specific events or activities. By analysing sound waves captured through acoustic sensors, this technology plays a crucial role in various domains, including environmental noise monitoring and public safety.

2.1.2.1.2.1 Detecting Noise Disturbances

Numerous products, from cars to industrial machinery, must comply with specific noise level standards. Traditional decibel meters can measure overall noise levels but fail to pinpoint the exact cause of noise disturbances. Current state of the art relies on the use of acoustic cameras, such as those from Sorama. They combine microphone arrays with intelligent software, including Artificial Intelligence (AI) algorithms, to localize and identify sound sources, providing insights to law enforcement authorities and relevant agencies for further actions. For instance, in a lively nightlife area like Stratumseind in Eindhoven, Sorama's acoustic cameras could visualize noise levels in each pub, helping authorities manage potential conflicts[68].

[68] Sorama. (n.d.). Acoustic Cameras and Sound Imaging. Retrieved from Sorama Portal

2.1.2.1.2.2 Aggression Detection

Imagine a bustling city street or a crowded bar. Acoustic cameras, such as the Sorama L642, can distinguish between various sounds, including quiet conversations and arguments. When an argument occurs, the camera links the aggressive sound to specific coordinates. Law enforcement can then respond promptly without recording the actual sound, ensuring privacy[69].

2.1.2.2 Visual Sensors

Visual sensors, such as CCTV cameras, provide critical visual data for surveillance. Recent advancements include high-resolution imaging, night vision, and thermal imaging, enhancing the capability to monitor and analyse activities in various lighting conditions[70],[71].

2.1.2.2.1 CCTV Systems

Smartening old CCTV systems involves integrating advanced technologies into existing security frameworks to enhance their capabilities, efficiency, and usability. This process often includes upgrading hardware components, incorporating AI and machine learning algorithms, and improving connectivity features to enable real-time monitoring, analysis, and data management.

The current state of the art in enhancing traditional CCTV systems from a hardware perspective involves various advancements. Researchers have proposed modifications to deep learning architectures like YOLO to process multiple input sources simultaneously, significantly increasing efficiency and practical frame rates[72]. Additionally, there is a shift towards smart cameras that can process information locally, reducing the reliance on cloud infrastructure for real-time threat detection[73]. Smart surveillance systems now incorporate features like background image extraction, object image analysis, and region of interest extraction to enhance monitoring capabilities[74]. Furthermore, the integration of smart devices with CCTV systems enables wireless communication for video transmission and remote management, improving overall system

[69] Sorama. (n.d.). Acoustic Surveillance Solutions. Retrieved from Sorama Portal

[70] Singh, D., Chitranshi, P., Kumar, A., & Pahuja, R. (2020). Intelligent video surveillance using deep learning and AI: A review. Journal of Information Security and Applications, 52, 102500.

[71] Hu, W., Xiao, T., Xie, D., Wu, Y., & Maybank, S. (2004). Traffic accident prediction using 3-D model-based vehicle tracking. IEEE Transactions on Vehicular Technology, 53(3), 677-694.

[72] Alam, F., Alraeesi, A. (2022). Computer Vision Based Smart CCTV Solution. Paper presented at the ADIPEC, Abu Dhabi, UAE, October 2022. doi: 10.2118/211842-ms

[73] Oguzhan, Can., Sezai, Burak, Kantarci., Gozde, Unal. (2021). A New Approach to Use Modern Object Detection Methods More Efficiently on CCTV Systems. doi: 10.1109/UBMK52708.2021.9558899

[74] Ume, Habiba., Muhammad, Awais., Milhan, Khan., Abdul, Jaleel. (2020). An Inexpensive Upgradation of Legacy Cameras Using Software and Hardware Architecture for Monitoring and Tracking of Live Threats. IEEE Access, doi: 10.1109/ACCESS.2020.2964778

functionality and maintenance[75]. These advancements collectively aim to improve the performance, efficiency, and intelligence of traditional CCTV systems through innovative hardware enhancements.

Traditional CCTV systems primarily offer real-time video surveillance and recorded footage for later review. However, these systems often lack the ability to analyse video content in real time, respond to dynamic situations, or integrate seamlessly with other security technologies. Smartening these systems involves integrating advanced hardware and software that introduce intelligence and connectivity into the mix.

The key components of smartening old CCTV systems encompass a series of strategic upgrades aimed at enhancing the functionality and efficiency of existing surveillance infrastructure. The process begins with hardware upgrades, which involve not just the addition of higher-resolution cameras for improved image quality but also the integration of sensors capable of motion detection. Moreover, these upgrades include implementing network capabilities that enable remote access and control, a crucial feature for modern surveillance needs. Some hardware enhancements may extend to installing edge computing devices at the camera site, which allows for local data processing and reduces the dependency on central servers, thereby minimizing latency and bandwidth usage.

Software enhancements form another critical pillar in the smartening process. The deployment of new software capabilities, such as AI algorithms for facial recognition, object detection, and behaviour analysis, significantly boosts the surveillance system's ability to identify and respond to security incidents intelligently. Furthermore, these software upgrades often include advanced data encryption methods, ensuring the secure transmission and storage of surveillance footage, addressing potential cybersecurity vulnerabilities.

Connectivity and integration represent the third cornerstone of upgrading old CCTV systems. By interfacing with modern technologies like the Internet of Things (IoT) devices, cloud computing platforms, and mobile applications, old CCTV systems are transformed into sophisticated, integrated security networks. This level of integration not only enables smarter alert systems and remote monitoring capabilities but also enhances the ability to perform comprehensive data analysis, leveraging the full potential of the collected surveillance data.

The objectives behind these upgrades are clear and multifaceted. The primary goal is to shift from passive surveillance setups to proactive security solutions that can accurately identify and alert on security breaches, unusual behaviours, and specific individuals of interest without the need for constant human oversight. This shift promises not only enhanced surveillance accuracy but also significant improvements in operational efficiency by automating surveillance processes,

[75] Lee, Donghyeok., Park, Namje. (2019). CCTV video smart surveillance system and method thereof.

thereby reducing the workload on security personnel and minimising the risk of human error. Furthermore, the adoption of advanced data analytics and cloud storage solutions facilitates more effective data management, allowing for the efficient handling and analysis of the vast amounts of data generated by these systems. Finally, modernising old CCTV systems with scalable technologies paves the way for future expansions and ensures adaptability to evolving security challenges, offering a path towards sustainable and flexible security infrastructure.

However, the journey towards smarter CCTV systems is not without its challenges. The cost of upgrading existing systems can be prohibitive, especially for large installations or those requiring extensive hardware modifications. Compatibility issues may arise, demanding technical expertise to ensure that new components integrate seamlessly with the old system without compromising its functionality. Moreover, as surveillance capabilities expand, especially with the integration of features like facial recognition and behaviour tracking, privacy concerns and ethical considerations come to the forefront, necessitating a careful and thoughtful approach to implementation.

In conclusion, the initiative to smarten old CCTV systems is more than a mere upgrade; it represents a comprehensive effort to leverage cutting-edge technology to create more intelligent, responsive, and integrated surveillance solutions. Achieving this requires a balanced approach that not only focuses on technical upgrades but also considers important considerations such as privacy, cost, and scalability. As technology continues to advance, staying informed and adaptable will be crucial for maximising the benefits of these smarter CCTV systems, ensuring they can meet the evolving demands of security and surveillance in the modern world.

2.1.2.2.2 New Technologies to be Adapted to CCTV Systems

2.1.2.2.2.1 Artificial Intelligence (AI) and Machine Learning (ML)

Artificial Intelligence (AI) and Machine Learning (ML) are revolutionising the way CCTV systems operate, turning traditional surveillance into intelligent monitoring solutions. These technologies enable systems to learn from the data they collect, making them smarter over time.

Applications:

- Facial Recognition: AI algorithms can identify and verify individuals in real-time, enhancing security measures and enabling personalised services.
- Anomaly Detection: ML models are trained to recognize normal behaviour patterns and alert operators to anomalies or suspicious activities, significantly reducing false positives.
- Behaviour Analysis: Analysing crowd dynamics, traffic flows, and individual behaviours to improve public safety, retail experiences, and workplace security.

Challenges: Implementing AI and ML requires significant processing power and data, raising concerns about privacy and data protection. Ensuring the accuracy and fairness of these systems, particularly in facial recognition, is also a major focus.

2.1.2.2.2.2 Edge Computing

Edge computing involves processing data near the source of data generation (i.e., the CCTV cameras themselves) rather than relying on a central data centre. This approach reduces latency and bandwidth requirements, enabling faster decision-making and action.

Applications:

- Real-Time Monitoring: Immediate processing on the edge allows for real-time security monitoring and quick response to incidents without the delay of sending data to a remote server.
- Data Optimization: By analysing and filtering data locally, only relevant information is sent to the cloud or central servers, optimizing storage and bandwidth usage.

Challenges: Edge computing demands more sophisticated hardware and software at the camera level, potentially increasing the cost and complexity of CCTV systems.

2.1.2.2.2.3 Cloud Storage and Computing

Cloud technologies provide scalable storage solutions and powerful computing capabilities for advanced analytics, remote access, and data sharing.

Applications:

- Scalable Storage: Cloud platforms offer flexible storage options, accommodating the vast amounts of data generated by high-definition and 24/7 surveillance cameras.
- Advanced Analytics: Leveraging cloud computing for advanced analytics allows for sophisticated data analysis, such as trend analysis and predictive modelling, to improve security and operational efficiency.
- Remote Access and Control: Cloud-enabled CCTV systems can be accessed and managed remotely, providing flexibility and ease of use for operators and security personnel.

Challenges: Cloud solutions depend on reliable internet connectivity and raise concerns about data security and privacy. The ongoing costs of cloud services also need to be considered.

2.1.2.2.2.4 4k and Higher Resolution Cameras

The adoption of 4K and higher resolution cameras significantly improves image quality, which is critical for identification purposes and detailed analysis.

Applications:

- Enhanced Clarity: Higher resolution cameras capture more detail, making it easier to identify individuals, read license plates, and observe suspicious activities.
- Digital Zoom: Improved resolution allows for better quality when zooming in on recorded footage, maintaining clarity where lower-resolution cameras would blur.

Challenges: High-resolution video generates large amounts of data, requiring more storage capacity and bandwidth. This can also increase the strain on processing hardware for analytics.

2.1.2.2.2.5 IoT Connectivity

The Internet of Things (IoT) connectivity enables CCTV systems to integrate and communicate with other devices and systems, creating a more interconnected and intelligent security ecosystem.

Applications:

- System Integration: CCTV systems can be integrated with alarm systems, access controls, and other security measures for comprehensive security management.
- Smart Alerts: IoT connectivity allows for smart alerts and actions, such as automatically locking doors or turning on lights in response to detected movements or identified threats.

Challenges: IoT connectivity introduces complexity in system integration and management. Security vulnerabilities in interconnected devices can also pose new risks.

2.1.2.2.3 CCTV Systems Challenges and Considerations

The process of upgrading old CCTV systems to incorporate modern, smart technologies comes with a set of challenges and considerations that can significantly impact the feasibility, implementation, and effectiveness of such initiatives. Understanding these hurdles is crucial for any organisation looking to enhance its surveillance capabilities.

Challenge 1: Cost

The financial aspect of upgrading CCTV systems is one of the most significant barriers. High-resolution cameras, advanced processing units for edge computing, and sophisticated software for data analysis and artificial intelligence (AI) functionalities can entail considerable expenses. This is particularly true for large installations or systems that require extensive modifications to accommodate new hardware and software. Organisations must carefully assess the cost versus benefit of such upgrades, considering not only the initial outlay but also the ongoing operational costs, such as cloud storage fees and maintenance expenses[76],[77],[78].

Challenge 2: Compatibility

Integrating new technologies with existing CCTV infrastructure poses technical challenges, especially when the old systems were not designed with future upgrades in mind. Ensuring compatibility between various components—such as cameras, storage solutions, and analytics software—requires a thorough technical evaluation and possibly the development of custom solutions. This can involve significant time and expertise to manage successfully. The risk is that incompatibilities can lead to system malfunctions, data losses, or security vulnerabilities, undermining the effectiveness of the surveillance system[79],[80].

Challenge 3: Privacy and Ethics

As CCTV systems become smarter, incorporating capabilities like facial recognition and behavioural analysis, they raise significant privacy and ethical concerns. The ability of these systems to identify individuals and track their movements can be perceived as invasive, leading to public backlash and legal challenges. Organizations must navigate a complex landscape of regulations and ethical considerations, ensuring that their use of advanced surveillance technologies respects individual privacy rights and complies with data protection laws. This involves implementing strict data management policies, securing informed consent when necessary, and maintaining transparency about how surveillance data is collected, used, and stored[81],[82].

Challenge 4: Cybersecurity

The increased connectivity and complexity of smart CCTV systems also make them more vulnerable to cyberattacks. As these systems often handle sensitive data and are integral to

[76] Kalbo, N., Mirsky, Y., Shabtai, A., & Elovici, Y. (2020). The Security of IP-Based Video Surveillance Systems. Sensors, 20(17), 4806.

[77] CCTV Security Pros. (n.d.). The Drawbacks of Old CCTV Security Cameras - How to Upgrade. Retrieved June 4, 2024, from https://www.cctvsecuritypros.com

[78] Chris Lewis Group. (n.d.). Why You Should Upgrade Your CCTV Cameras. Retrieved June 4, 2024, from https://www.chrislewis.co.uk

[79] Gupta, P., & Margam, M. (2021). CCTV as an efficient surveillance system? An assessment from 24 academic libraries of India. Global Knowledge, Memory and Communication, 70(4/5), 355-376.

[80] Defend Security Group. (n.d.). When Should You Upgrade Your Existing CCTV System To Newer Technology? Retrieved June 4, 2024, from https://www.defendsecuritygroup.com.au

[81] Kalbo, N., Mirsky, Y., Shabtai, A., & Elovici, Y. (2020). The Security of IP-Based Video Surveillance Systems. Sensors, 20(17), 4806.

[82] Chris Lewis Group. (n.d.). Why You Should Upgrade Your CCTV Cameras. Retrieved June 4, 2024, from https://www.chrislewis.co.uk

security operations, breaches can have serious consequences. Protecting against such threats requires robust cybersecurity measures, including secure data encryption, regular software updates, and vigilant monitoring for potential vulnerabilities. Organizations must be proactive in their cybersecurity efforts, recognizing that the smartening of CCTV systems introduces new attack vectors that must be defended against[83],[84].

Challenge 5: Technical Expertise

Successfully upgrading and managing smart CCTV systems demand a higher level of technical expertise than traditional surveillance setups. From the initial installation and integration of new technologies to the ongoing management and troubleshooting of advanced analytical software, organizations need access to skilled professionals. This may necessitate training existing staff, hiring new experts, or contracting with specialized service providers, adding to the overall cost and complexity of the project[85],[86],[87],[88].

Challenge 6: Scalability and Future-Proofing

Finally, ensuring that upgraded CCTV systems are scalable and adaptable to future technological advancements is crucial. Surveillance needs and technologies evolve rapidly, and systems that are not designed with flexibility in mind can quickly become obsolete. This requires careful planning and the selection of modular, upgradable components that can accommodate new features and capabilities as they become available[89],[90],[91].

[83] Kalbo, N., Mirsky, Y., Shabtai, A., & Elovici, Y. (2020). The Security of IP-Based Video Surveillance Systems. Sensors, 20(17), 4806.

[84] Defend Security Group. (n.d.). When Should You Upgrade Your Existing CCTV System To Newer Technology? Retrieved June 4, 2024, from https://www.defendsecuritygroup.com.au

[85] Kalbo, N., Mirsky, Y., Shabtai, A., & Elovici, Y. (2020). The Security of IP-Based Video Surveillance Systems. Sensors, 20(17), 4806.

[86] Gupta, P., & Margam, M. (2021). CCTV as an efficient surveillance system? An assessment from 24 academic libraries of India. Global Knowledge, Memory and Communication, 70(4/5), 355-376.

[87] Defend Security Group. (n.d.). When Should You Upgrade Your Existing CCTV System To Newer Technology? Retrieved June 4, 2024, from https://www.defendsecuritygroup.com.au

[88] Ratcliffe, J. (2018, January 2). Upgrade or Repair? Assessing an Old CCTV System. CCTV.co.uk. Retrieved June 4, 2024, from https://www.cctv.co.uk

[89] Yao, Q., Huang, Y., & Wang, X. (2020). Integrating AI into CCTV Systems: A Comprehensive Evaluation of Smart Video Surveillance in Community Space. arXiv preprint arXiv:2312.02078.

[90] Vector Security Networks. (n.d.). The Evolution of Closed-Circuit Television (CCTV) Systems. Retrieved June 4, 2024, from https://www.vectorsecuritynetworks.com

[91] Defend Security Group. (n.d.). When Should You Upgrade Your Existing CCTV System To Newer Technology? Retrieved June 4, 2024, from https://www.defendsecuritygroup.com.au

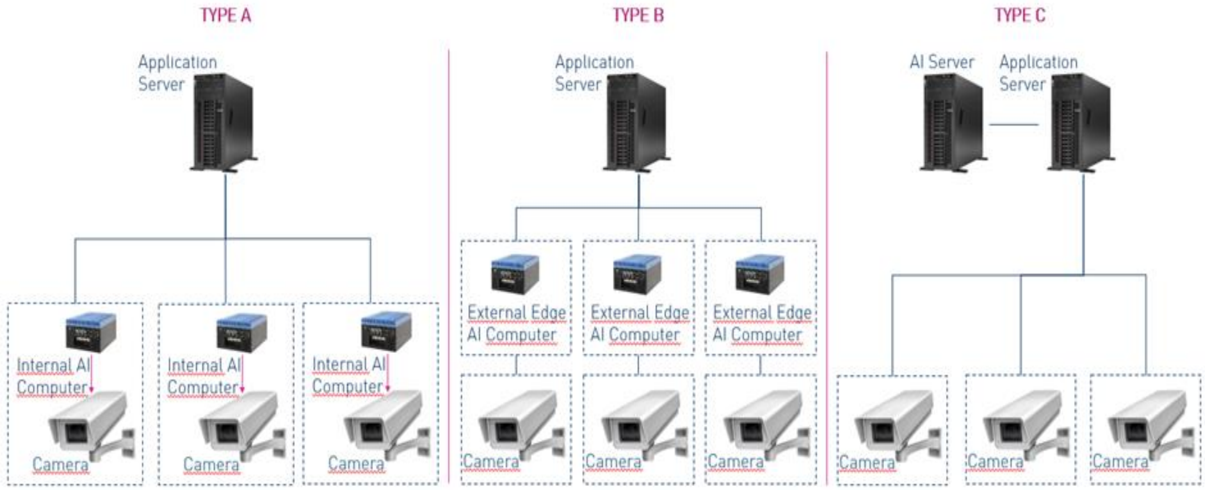


Figure 8. Hardware architectures of smart CCTV systems

2.1.2.2.4 Visual-Thermal Imaging

A visible image sensor captures incident light in the visible spectrum (wavelength between 380nm and 700nm) and converts it into electrical signals which in turn can be used to create images and videos. In good lighting conditions, the signal to noise ratio (SNR) of the signals generated by the sensor are high, leading to high quality images. However, the SNR drops significantly with decreasing light intensity, which ultimately leads to poor quality images. This problem becomes particularly prominent in real world surveillance scenarios, where monitoring becomes challenging in dark areas due to the loss of relevant information in the images or videos (see the following figure-left). This is undesirable, especially in high security premises. To address this problem, different methods have been proposed such as filtering with morphological operators, histogram-based equalization or altering brightness and contrast levels[92]. These methods are mostly focused on removal of noise and adjusting the dynamic range of the images. These methods can be effective up to a certain limit, beyond which the traditional image processing methods become ineffective and unreliable. This is where far infrared (FIR) sensors come into play.

Thermal cameras capture the energy in the FIR spectrum (wavelength between 15µm and 1mm) and converts it into a visible light display. Its behaviour is based on the fact that all objects above absolute zero emit thermal infrared energy due to the temperature and the emissivity of the object, and due to radiation, that is reflected on the object. Although thermal camera images can see all objects regardless of the ambient light, their image has different properties than what we

[92] T. Acharya and A. Ray. Image Processing: Principles and Applications. Wiley, 2005.

are used to in the visible spectrum with RGB images, in the sense that they lack information about texture and colour (see the following figure-centre).

The use of FIR or thermal cameras is a common practice for surveillance in locations that require high security. RGB-thermal imaging fusion aims at combining the available colour and texture information from the visible spectrum, and the relevant information from the thermal spectrum, to create a more complete scene (see the following figure-right).

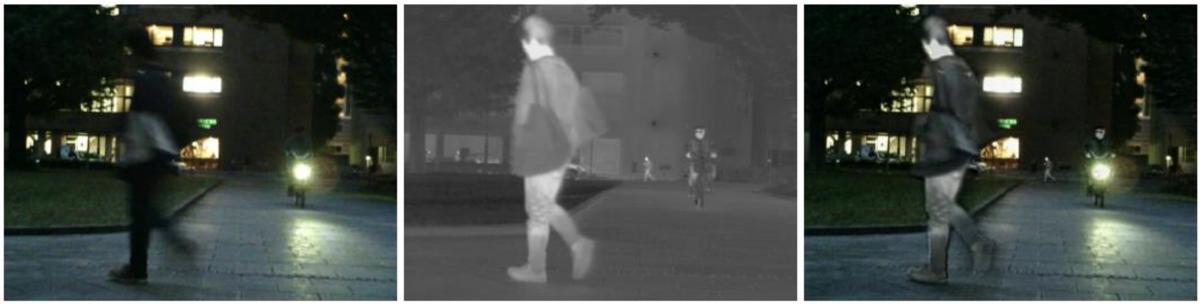


Figure 9. An example of visual and thermal imaging fusion in low light conditions, with (left) visual input, (centre) thermal input and (right) a resulting fused visual-thermal image

Multi-modal image fusion includes, but is not limited to, medical scans (i.e. MRI, CT or PET), near infrared (NIR) and visible, and FIR/thermal and visible image fusion. To solve the problem of fusing images from the visible and thermal domain, different methods have been proposed. These methods are considered SOTA on a plethora of applications. However, the performance of these methods is limited for surveillance, as they are not optimized for this specific application. The resulting images of these methods contain artefacts, are noisy, low in contrast and lack sharpness and textures, which are critical features in low-light surveillance applications.

The existing multi-modal image fusion methods are built on a variety of neural network architecture principles. The networks consist of convolutional neural networks (CNN), autoencoders, generative adversarial networks (GAN), and vision transformers, or an ensemble of these methods.

- CNN-based methods

As the name suggests, CNN-based fusion methods utilize CNN, where the input consists of a concatenated array of multiple images. CNN performs a pixel-level fusion between the two input images[93]. There are a multitude of methods performing CNN based image fusion from which we will mention those that have better performance. The fusionDN method[94] uses a linear DenseNet,

[93] S. Kalamkar et al. Multimodal image fusion: A systematic review. Decision Analytics Journal, p. 100327, 2023.

[94] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo. Fusiondn: A unified densely connected network for image fusion. Proceedings of the AAAI Conference on Artificial Intelligence, 34(07):12484–12491, Apr. 2020. doi: 10.1609/aaai.v34i07.6936

which is a CNN with skip-connections. The same authors proposed a continuation of this method called U2Fusion[95]. The architecture is largely the same with some minor improvements. The first improvement is a changed loss function to induce similar lighting intensity compared to the input image. The second change includes training the model for inference on medical data. The core of the FusionDN and U2Fusion network is the use of a pretrained VGG-16 network to calculate information-based weights in training for use in a weighted loss function. The MetaFusion method[96] focuses on a combined fusion and detection network. Through meta-feature embedding, this model produces a fusion output with detection masks. This research is of interest when involving downstream tasks such as object detection.

- GAN-based methods

GAN-based methods consist of a generator to create a fused output and one or more discriminators. These discriminators learn the differences between the input thermal and the fused output, or the visible input image and the fused output of the generator, to create a feedback loop which trains the generator. This method has the drawback of focusing its discriminators on comparing the output back to the inputs, which are not ideal representations for this work. An example of a model employing the GAN-based fusion strategy is the work done by Ma et al.[97].

- Autocenter based methods

Autoencoder based models include an encoder and decoder part of the network where the encoder learns the feature embedding of the different input modalities and the decoder translates the fused features to a fused output. The LadleNet method[98] uses a cascaded U-net structure where the first U-net, “the handle”, transfers thermal images into visible images (through style-transferring). The U-net being an autoencoder with bridged connections between the encoder and decoder[99]. This “handle” style-transferred output can then be used to map the

[95] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling. U2fusion: A unified unsupervised image fusion network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1):502–518, 2020.

[96] W. Zhao, S. Xie, F. Zhao, Y. He, and H. Lu. Metafusion: Infrared and visible image fusion via metafeature embedding from object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13955–13965, 2023.

[97] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang. Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing, 29:4980–4995, 2020.

[98] T. Zou and L. Chen. Ladlenet: Translating thermal infrared images to visible light images using a scalable two-stage u-net, 2023.

[99] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

transferred thermal images to the visible images for alignment, or for fusion itself. CDDFuse[100] is one of the SOTA autoencoder methods. It concerns an attention-based model with promising quantitative and qualitative results. DIVFusion[101] is an autoencoder-based method similar to CDDFuse. From the interpretation of the quantitative results on the DIVFusion model, it appears to perform as one of the SOTA methods. In the qualitative assessment, the SOTA performance is challenged, with the DIVFusion method boosting the luminance, inducing an unrealistic bright output image where contrast is sacrificed. The DATFuse model[102] employs a dual attention transformer to focus on thermal and visible image fusion in multiple heavy weather conditions. The work of Tang et al. claims real-time inference performance running on a NVIDIA GeForce RTX 3090 GPU. However, the paper does not include performance regarding low-light scenes and, thus, a further inquiry on its performance on this category of data needs to be performed. Swinfusion[103] operates using the principles of SwinIR[104] transformer blocks. The SwinIR feature extraction method has good performance according to the work of Li et al.[105]. In the fusion method proposed by Ma et al.[106], the architecture of its attention block is unchanged from the original work by Vaswani et al.[107]. The novelty is in the application of cutting up the input image into 6 windows and shifting these windows in between attention blocks. This alteration forces the model to learn attention relations within the image, from multiple perspectives. The downside of this method is the computational disadvantage in SwinIR and therefore SwinFusion. Additionally, the need for fixed square patches in inference is deemed inefficient. ReCoNet's model[108] is

[100] Z. Zhao, H. Bai, J. Zhang, Y. Zhang, S. Xu, Z. Lin, R. Timofte, and L. Van Gool. Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5906–5916, June 2023.

[101] L. Tang, X. Xiang, H. Zhang, M. Gong, and J. Ma. Divfusion: Darkness-free infrared and visible image fusion. Information Fusion, 91:477–493, 2023.

[102] W. Tang, F. He, Y. Liu, Y. Duan, and T. Si. Datfuse: Infrared and visible image fusion via dual attention transformer. IEEE Transactions on Circuits and Systems for Video Technology, 2023.

[103] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma. Swinfusion: Cross-domain longrange learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica, 9(7):1200–1217, 2022. doi: 10.1109/JAS. 2022.105686.

[104] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte. Swinir: Image restoration using swin transformer, 2021.

[105] Y. Li, Y. Zhang, R. Timofte, L. Van Gool, Z. Tu, K. Du, H. Wang, H. Chen, W. Li, X. Wang, et al. Ntire 2023 challenge on image denoising: Methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1904–1920, 2023.

[106] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma. Swinfusion: Cross-domain longrange learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica, 9(7):1200–1217, 2022. doi: 10.1109/JAS. 2022.105686.

[107] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, 2023.

[108] Z. Huang, J. Liu, X. Fan, R. Liu, W. Zhong, and Z. Luo. Reconet: Recurrent correction network for fast and efficient multi-modality image fusion. In European Conference on Computer Vision, pp. 539–555. Springer, 2022.

optimized for speed rather than performance. A point of interest in this method is its use of the parallax problem[109] of multi-modal cameras for depth estimation.

2.1.2.2.4.1 Feature extraction via vision transformers

The feature extraction is an intricate part of the autoencoder based fusion models. Due to their method of fusing features rather than pixels. Inspired from the denoising challenge of NTIRE 2023[110], several leading methodologies have been proposed, using vision transformers to extract features while simultaneously learning to suppress unwanted patterns such as noise or artifacts. The results of showcase the power of novel transformers architectured like SwinIR, NAFNet, and Restormer as the SOTA. Each of these methods employs an improved performance alternative to the traditional transformer block architecture[111], which becomes computationally expensive with increased image resolution. The SwinIR[112], NAFNet[113], and Restormer[114] blocks exhibit improved image-removal capability while suppressing noise and artifacts. The Restormer method is used in fusion models such as CDDFuse. More recently, the new SOTA of the work by Gao et al.[115] has been published, which combines the functionalities of the previous SOTA.

Further exploration of the literature on these SOTA transformer methods uncovers some additional advancements showing incremental improvements. The first improvement relates to patch size during training. Attention based models trained on patches optimize the fusion to the patch dimension. During inference time, the full image that is put through the model is often much larger than the patch dimensions. This difference in training and inference creates a drop-in performance. To improve on this, there are multiple possible solutions. Training a model on patches of images requires less computational power, making it more efficient and advantageous.

[109] J. Kang, I. Cohen, G. Medioni, and C. Yuan. Detection and tracking of moving objects from a moving platform in presence of strong parallax. In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 1, pp. 10–17 Vol. 1, 2005. doi: 10.1109/ICCV.2005.72

[110] Y. Li, Y. Zhang, R. Timofte, L. Van Gool, Z. Tu, K. Du, H. Wang, H. Chen, W. Li, X. Wang, et al. Ntire 2023 challenge on image denoising: Methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1904–1920, 2023.

[111] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, 2023.

[112] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte. Swinir: Image restoration using swin transformer, 2021.

[113] L. Chen, X. Chu, X. Zhang, and J. Sun. Simple baselines for image restoration, 2022.

[114] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang. Restormer: Efficient transformer for high-resolution image restoration, 2022.

[115] H. Gao and D. Dang. Learning enriched features via selective state spaces model for efficient image deblurring, 2024.

One workaround used by many models in the NTIRE 2023 challenge[116], involves the use of incrementally increasing patch sizes during their training, starting at a size of 128x128 and moving up to 512x512. A second method proposed in[117] as Test-time Local Converter (TLC), brings the size of the training and inference data closer together. This method proposes cutting up inference-time images to the same size as the training patch size to extract the performance in the model. It should be considered that, it is more effective when training with a patch size of 256X256 or higher.

2.1.2.2.4.2 Challenges of visual and thermal imaging

There are several challenges in the fusion of visual and thermal imaging for low-light surveillance applications.

Challenge 1:

Spatial and temporal alignment between visual and thermal sensors represent a challenge in low-light conditions, where the lack of details in the visual images makes it difficult to find matching features in the two modalities. For the temporal alignment, the difficulty also lies in the different frame rate of the two sensors together with the possibility of frame drops in either sensor. There is lack of work in the literature related to self- or semi-supervised thermal and visible image alignment methods, so existing methods rely on finding matching features in the two modalities or using bounding box coordinates, which have been previously annotated by another method. In either way, some manual annotation is needed to, for example, dispose of misaligned pairs or delete duplicate bounding boxes in one of the modalities.

Challenge 2:

The fused image with existing methods still presents artefacts, noise, lack of contrast, sharpness and textures/details. Additionally, fusion methods need to learn to not introduce unneeded information from one of the modalities that would hamper the information you get from the other modality.

Challenge 3:

Deployment of the models for edge fusion. Current methods are based on large scale models that contains millions of parameters and are not optimized to be used at the edge.

[116] Y. Li, Y. Zhang, R. Timofte, L. Van Gool, Z. Tu, K. Du, H. Wang, H. Chen, W. Li, X. Wang, et al. Ntire 2023 challenge on image denoising: Methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1904–1920, 2023.

[117] X. Chu, L. Chen, , C. Chen, and X. Lu. Improving image restoration by revisiting global information aggregation. arXiv preprint arXiv:2112.04491, 2021.

2.1.2.2.5 Video-Based Anomaly Detection

In the domain of video surveillance, anomaly detection has been subject to a progressive evolution. Initially, attempts have been made to devise expert systems characterized by complex sets of rules, aiming at emulating the intricacies of behavioural dynamics. However, the implementation and scalability of such systems are impeded by the inherent complexity and variability of people behaviour, often necessitating a disproportional number of exceptions to accommodate diverse scenarios. Recent advancements in computer vision are bolstered by the proliferation of sensor data from surveillance monitoring infrastructure, which allow for a paradigm shift in anomaly detection methodologies. These modern approaches leverage the computational capabilities of neural networks to analyse data streams obtained from surveillance cameras. While exhibiting promising performance on standardized benchmarks, these methodologies have concurrently shown critical aspects such as privacy preservation, algorithmic biases, and false-positive mitigation. In recent years, the approaches proposed to the task of anomaly detection in traffic can be split in three main groups: unsupervised, weakly-supervised and fully-supervised.

2.1.2.2.5.1 Video Feature Extraction

The first step to many video understanding tasks, such as action recognition and anomaly detection, is extracting features from given videos. Over the past two decades, convolutional neural network (CNN) based architectures (such as 2D-CNN methods[118],[119],[120], C3D[121], I3D[122], SlowFast[123], VGG[124], ResNet[125], and DenseNet[126]) have been extensively studied to understand spatio-temporal representation. In recent years, Vision Transformer (ViT) and its variants,

[118] Wang L, Xiong Y, Wang Z, Qiao Y, Lin D, Tang X, et al. Temporal segment networks: Towards good practices for deep action recognition. In: European conference on computer vision. Springer; 2016. p. 20–36.

[119] Lin J, Gan C, Han S. Temporal shift module for efficient video understanding. CoRR abs/1811.08383 (2018) 1811.

[120] Fan Q, Chen CFR, Kuehne H, Pistoia M, Cox D. More is less: Learning efficient video representations by big-little network and depthwise temporal aggregation. Advances in Neural Information Processing Systems. 2019; 32.

[121] Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 4489–4497.

[122] Carreira J, Zisserman A. Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 6299–6308.

[123] Feichtenhofer C, Fan H, Malik J, He K. Slowfast networks for ideo recognition. In: Proceedings of the IEEE/CVF international conference on computer vision; 2019. p. 6202–6211.

[124] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.

[125] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770–778.

[126] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–4708.

including TimeSformer[127], ViViT[128], MViT[129], MViTv2[130], and Video Swin Transformer[131] have achieved outstanding performance in video action recognition. Moreover, large language model-based solutions like InternVideo2[132] are gaining importance due to high potential.

2.1.2.2.5.2 Supervised Anomaly Detection

The methods belonging to the supervised paradigm are trained with frame-level annotated videos. Classical works in this field involved algorithms, such as Support Vector Machine (SVM)[133],[134], that focus on distinguishing anomalous trajectories from normal ones in n-dimensional spaces. In later works, neural networks trained with direct supervision[135],[136] proved capable of handling a more diverse set of anomalous actions. Semi-supervised and self-supervised methods have shown some promise as well[137],[138],[139]. However, these methods have two main drawbacks: they require large amounts of annotated data to obtain efficient models, and the fact that they are constrained by the classes of anomalies on which they are trained, preventing them

[127] Bertasius G, Wang H, Torresani L. Is space-time attention all you need for video understanding? In: ICML. vol. 2; 2021. p. 4.

[128] Arnab A, Dehghani M, Heigold G, Sun C, Lučić M, Schmid C. Vivit: A video vision transformer. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 6836–6846.

[129] Fan H, Xiong B, Mangalam K, Li Y, Yan Z, Malik J, et al. Multiscale vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 6824–6835.

[130] Li Y, Wu CY, Fan H, Mangalam K, Xiong B, Malik J, et al. Mvitv2: Improved multiscale vision transformers for classification and detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 4804–4814.

[131] Liu Z, Ning J, Cao Y, Wei Y, Zhang Z, Lin S, et al. Video Swin transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022. p.3202–3211.

[132] Wang Y, Li K, Li X, Yu J, He Y, Chen G, et al. InternVideo2: Scaling Video Foundation Models for Multimodal Video Understanding. arXiv preprint arXiv:240315377. 2024.

[133] Batapati P, Tran D, Sheng W, Liu M, Zeng R. Video analysis for traffic anomaly detection using support vector machines. Proceedings of the World Congress on Intelligent Control and Automation (WCICA). 2015 03;2015:5500–5505.

[134] Piciarelli C, Micheloni C, Foresti GL. Trajectory-Based Anomalous Event Detection. IEEE Transactions on Circuits and Systems for Video Technology. 2008;18(11):1544–1554.

[135] Sultani W, Chen C, Shah M. Real-world anomaly detection in surveillance videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 6479–6488.

[136] Sarker MI, Losada-Gutiérrez C, Marron-Romera M, Fuentes-Jiménez D, Luengo-Sánchez S. Semi-supervised anomaly detection in video-surveillance scenes in the wild. Sensors. 2021;21(12):3993.

[137] Wu JC, Hsieh HY, Chen DJ, Fuh CS, Liu TL. Self-supervised Sparse Representation for Video Anomaly Detection. In:Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors. Computer Vision – ECCV 2022. Cham: Springer Nature Switzerland; 2022. p. 729–745.

[138] Georgescu MI, Barbalau A, Ionescu RT, Khan FS, Popescu M, Shah M. Anomaly detection in video via self-supervised and multi-task learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. p.12742–12752.

[139] Huang C, Wen J, Xu Y, Jiang Q, Yang J, Wang Y, et al. Selfsupervised attentive generative adversarial networks for video anomaly detection. IEEE transactions on neural networks and learning systems. 2022.

from generalizing to unseen anomalies. The latter is a crucial point: anomalous actions are unpredictable, meaning that any dataset collected will not contain all the anomalies that can happen in real-world scenarios. Furthermore, neural networks are inherently not explainable [109, 110, 111, 140], which heavily hinders the real-world applications in which they can be safely deployed.

2.1.2.2.5.3 Weakly-Supervised Anomaly

One of the most popular approaches to anomaly detection problems is weakly-supervised anomaly detection. In this approach, train-set annotations only include a class label. In contrast, test-set annotations contain a video class label, the number of frames, and the starting and ending frame positions of an abnormal event in a video. Recently, a lot of studies such as[141],[142] are conducted on weakly-supervised approach by employing the multi-instance learning (MIL). Recent works have shown that the transformer architecture generally outperforms previous designs in this context, likely due to their temporal and spatial inductive biases[143],[144]. A weakly supervised solution brings a major advantage over fully supervised anomaly detection methods due to faster annotation than labelling in each frame. A disadvantage of weakly-supervised anomaly detection is that it only learns to recognize anomalies that occurred in the training set. Also, background pixels could influence the final prediction in unexpected ways, even if the anomalous action is contained in the training dataset but in a different scenario.

2.1.2.2.5.4 Unsupervised Anomaly Detection

The main idea behind unsupervised methods is to model the distribution of regular data and solve the anomaly detection task by detecting data points that fall outside the learned model. In the context of anomaly detection in traffic, unsupervised methods model how different types of agents normally behave in a given context, such as vehicles or pedestrians in an intersection, and detect an agent deviating from this model as anomalous. Often, these methods rely on a

[140] Nguyen QP, Lim KW, Divakaran DM, Low KH, Chan MC. GEE: A gradient-based explainable variational autoencoder for network anomaly detection. In: 2019 IEEE Conference on Communications and Network Security (CNS). IEEE; 2019. p.91–99.

[141] Tian Y, Pang G, Chen Y, Singh R, Verjans JW, Carneiro G. Weakly-Supervised Video Anomaly Detection with Robust Temporal Feature Magnitude Learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021. p. 4975–4986.

[142] Sultani W, Chen C, Shah M. Real-world anomaly detection in surveillance videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 6479–6488.

[143] Li S, Liu F, Jiao L. Self-training multi-sequence learning with transformer for weakly supervised video anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36; 2022. p. 1395–1403.

[144] Chen Y, Liu Z, Zhang B, Fok W, Qi X, Wu YC. Mgfn: Magnitude contrastive glance-and-focus network for weakly-supervised video anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37; 2023. p. 387–395.

clustering algorithm in high-dimensional spaces[145],[146],[147]. Deep learning-based approaches were developed to leverage the capabilities of neural networks, with encouraging results on a large set of publicly available datasets[148],[149]. More recently, Variational AutoEncoders (VAEs)[150],[151] proved to be effective in anomaly detection tasks[152],[153],[154], leveraging the capabilities of such models to learn latent representations of normal distributions even in complex contexts such as traffic analysis[155],[156],[157]. A similar line of research focuses on frame reconstruction within videos, where anomalies are detected based on discrepancies between generated and original frames[158],[159],[160]. In general, unsupervised methods are useful when the exact nature of the anomalies is unknown or uncertain. However, these methods may fail when considering subtle differences that result in a low image reconstruction error. In fact, an action can often be detected as anomalous only when

[145] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Comput Surv. 2009 jul;41(3). https://doi.org/10.1145/1541880.1541882.

[146] Liu SWTT, Ngan HYT, Ng MK, Simske SJ. Accumulated Relative Density Outlier Detection for Large Scale Traffic Data. Electronic Imaging. 2018;30(9):239–1–239–1. https://library.imaging.org/ei/articles/30/9/art00010.

[147] Santhosh KK, Dogra DP, Roy PP. Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey. ACM Comput Surv. 2020 dec;53(6). https://doi.org/10.1145/3417989.

[148] Zhou JT, Du J, Zhu H, Peng X, Liu Y, Goh RSM. AnomalyNet: An Anomaly Detection Network for Video Surveillance. IEEE Transactions on Information Forensics and Security. 2019 10;14(10):2537–2550.

[149] Singh P, Pankajakshan V. A deep learning based technique for anomaly detection in surveillance videos. In: 2018 Twenty Fourth National Conference on Communications (NCC). IEEE; 2018. p. 1–6.

[150] Kingma DP, Welling M. An Introduction to Variational Autoencoders. CoRR. 2019;abs/1906.02691. http://arxiv.org/abs/1906.02691.

[151] Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:13126114. 2013.

[152] Lin S, Clark R, Birke R, Schönborn S, Trigoni N, Roberts S. Anomaly detection for time series using vae-lstm hybrid model. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Ieee; 2020. p. 4322–4326.

[153] Zhou Y, Liang X, Zhang W, Zhang L, Song X. VAE-based deep SVDD for anomaly detection. Neurocomputing. 2021; 453:131–140.

[154] Niu Z, Yu K, Wu X. LSTM-based VAE-GAN for time-series anomaly detection. Sensors. 2020; 20(13):3738.

[155] Roy PR, Bilodeau GA. Road user abnormal trajectory detection using a deep autoencoder. In: Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13. Springer; 2018. p. 748–757.

[156] Santhosh KK, Dogra DP, Roy PP, Mitra A. Vehicular trajectory classification and traffic anomaly detection in videos using a hybrid CNN-VAE Architecture. IEEE Transactions on Intelligent Transportation Systems. 2021; 23(8):11891–11902.

[157] Kumaran SK, Dogra DP, Roy PP, Mitra A. Video trajectory classification and anomaly detection using hybrid CNN-VAE. arXiv preprint arXiv:181207203. 2018.

[158] Ionescu RT, Khan FS, Georgescu M, Shao L. Object-centric Auto-encoders and Dummy Anomalies for Abnormal Event Detection in Video. CoRR. 2018; abs/1812.04960. http://arxiv.org/abs/1812.04960.

[159] Chen W, Xu H, Li Z, Pei D, Chen J, Qiao H, et al. Unsupervised anomaly detection for intricate kpis via adversarial training of vae. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications. IEEE; 2019.

[160] Gong D, Liu L, Le V, Saha B, Mansour MR, Venkatesh S, et al. Memorizing Normality to Detect Anomaly: Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection. CoRR. 2019; abs/1904.02639.

considering the context in which it is happening, making the anomaly detection task more difficult.

2.1.2.3 Radar Sensors

Radar sensors are utilized for their ability to detect objects and measure their speed, distance, and direction. They are effective in various environmental conditions and are often used in perimeter security and traffic monitoring[161],[162].

2.1.2.3.1 MmWave Radar Technology

Microwave (mmWave) radar sensors[163] developed by Texas Instruments are powerful detection devices commonly used for detecting and tracking objects at long distances. Microwave radar sensors utilize high-frequency radio waves to detect objects in the surroundings using electromagnetic waves. These sensors typically operate in the frequency range of 57 GHz to 81 GHz and have a wide range of applications. MmWave radar sensors are used in various fields such as autonomous vehicles, industrial automation, security systems, and medical imaging, providing high-resolution and accurate detection.

MmWave radar sensors have various technical specifications including operating frequency, antenna structure, output power, detection range, resolution, and data processing capabilities. These features determine the sensor's performance and application areas.

The IWR series is one of the leading models of mmWave radar sensors. Sensors in this series are available in different frequency ranges and technical specifications (maximum detection range, resolution, power consumption, etc.). In the literature, models and variants such as IWR1443, IWR1843, IWR1642, IWR6432, and IWR6843 can be found. Additionally, these sensors can be available in board form such as IWR1443BOOST, IWR1843BOOST, IWR6843ISK, and IWR6843ISK-ODS, in addition to being sold as chips in the market[164]. Each model has different application areas and performance characteristics.

MmWave radar sensor can provide 5D (x, y, z, velocity, intensity) point cloud data and 3D (range, doppler velocity and intensity) range-doppler data. The figures for two different types of data are provided in the following.

[161] Amin, M. G. (Ed.). (2017). Radar for indoor monitoring: Detection, classification, and assessment. CRC Press.

[162] Patole, S. M., Torlak, M., Wang, D., & Ali, M. (2017). Automotive radars: A review of signal processing techniques. IEEE Signal Processing Magazine, 34(2), 22-35.

[163] https://www.ti.com/sensors/mmwave-radar/products.html

[164] https://dev.ti.com/tirex/explore/node?node=A__AMKSv8im74YjT7cmO9jVHg__com.ti.mmwave_industrial_toolbox__VLyFKFf__4.9.0

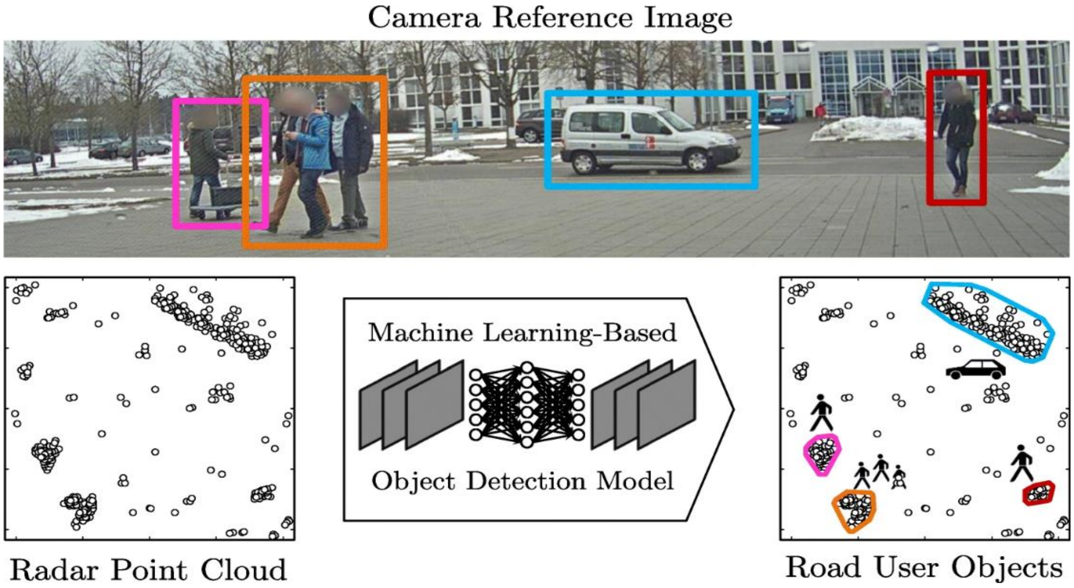


Figure 10. Point cloud data which can be detected from mmWave radar sensors and an example of usage of them with AI[165]

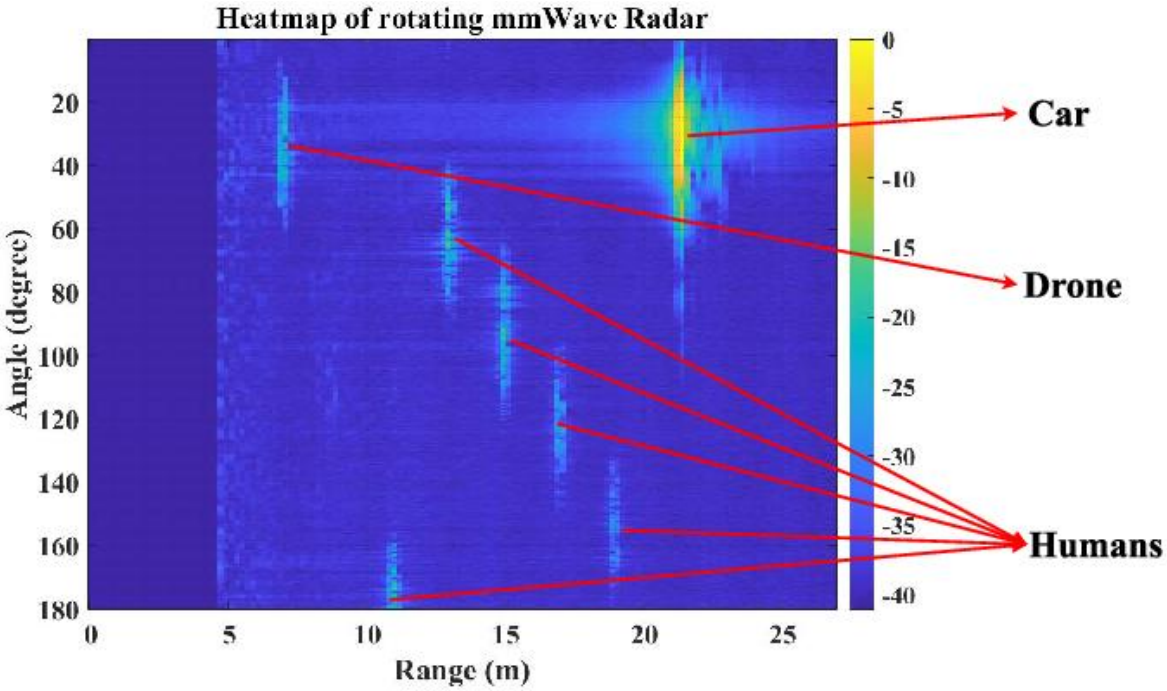


Figure 11. Range-Doppler data in an article[166]

[165] https://aiperspectives.springeropen.com/articles/10.1186/s42467-021-00012-z

[166] https://ieeexplore.ieee.org/document/9465137

There are many ML and DL studies where mmWave radar sensors are used.

- MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare (*2021)

In this study, a new approach is developed to enable patients with motor impairments to perform prescribed exercises at home and alleviate their transportation needs, expert shortages, and healthcare costs. A mmWave radar sensor was preferred based on criteria such as high cost, serious privacy concerns, and lighting for the detection of patients' 3D joint points. The aim is to obtain valuable visualization and feedback based on body movements by detecting human joint points. Therefore, a millimetre-wave (mmWave) based supportive rehabilitation system (MARS) is proposed for pose detection to identify motor impairments in patients.

Ten typical exercise movements were performed against a mmWave radar sensor (IWR1443) and a Kinect-v2 sensor labelling radar data, and a dataset was created. Point cloud data containing features for each point (x, y, z, velocity, intensity) from the mmWave radar sensor were obtained, while point cloud data representing 19 basic joint points on the human body were obtained from the Kinect-v2 sensor, which labels the point cloud data using its built-in camera. A CNN-based deep learning approach was used to detect joint points in the human body.

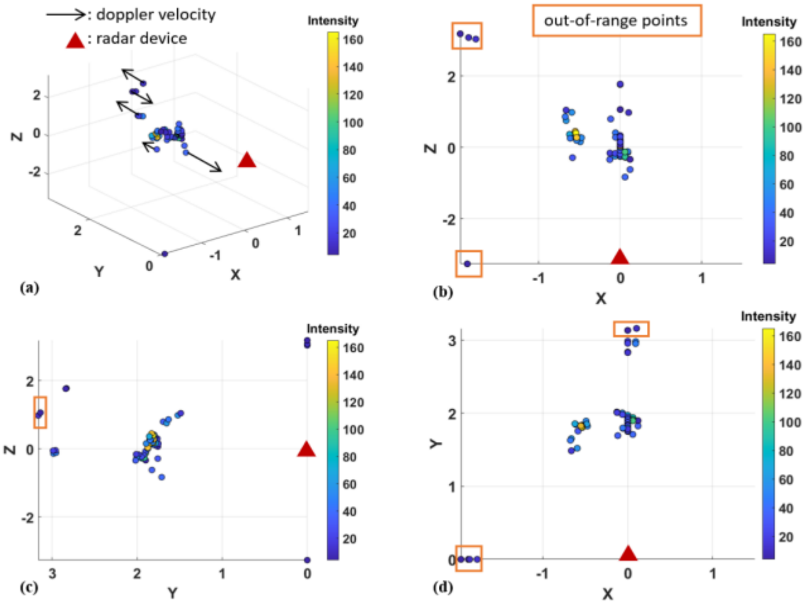


Figure 12. 3D and 2D Visualization of radar point cloud data which is used in this paper. (a), (b), (c), and (d) shows the 3D view, front view, side view and top view respectively.[167]

[167] Sizhe An and Umit Y. Ogras. 2021. MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare. ACM Trans. Embedd. Comput. Syst. 1, 1, Article 1 (January 2021), 22 pages. Fig. 3, p. 9. https://doi.org/10.1145/3477003

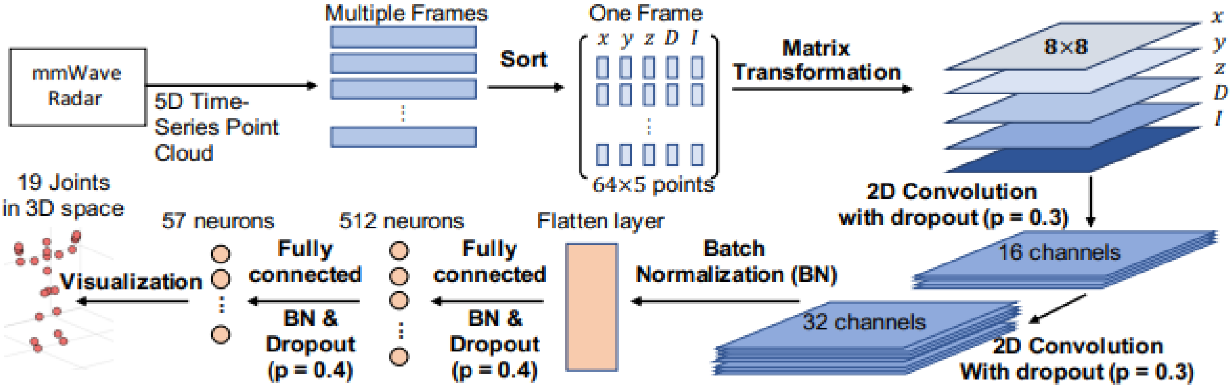


Figure 13. Data pre-processing and CNN stages of this paper[168]

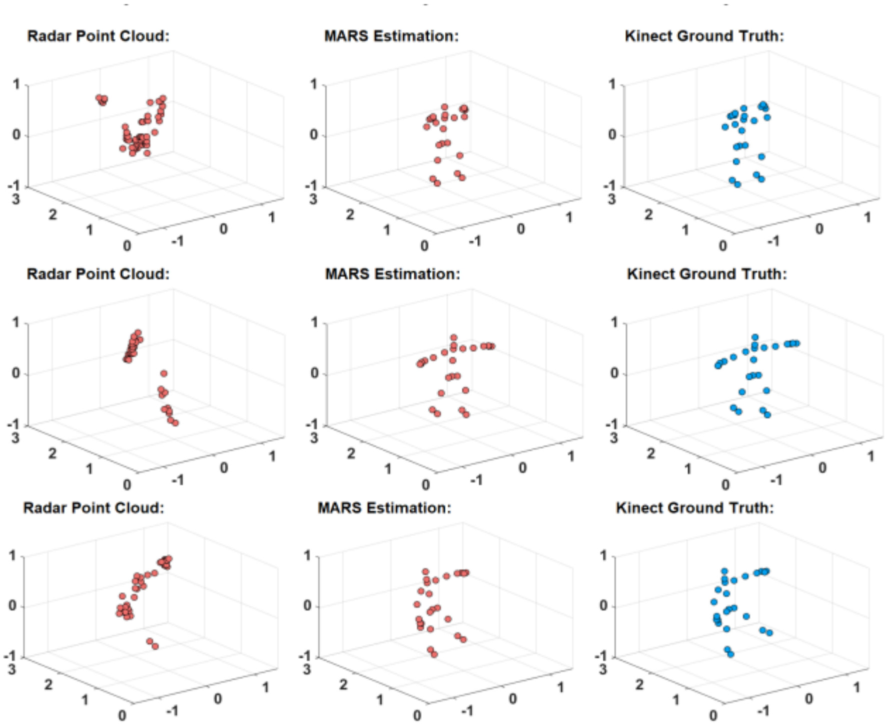


Figure 14. Respectively, radar point cloud data derived from mmWave radar sensor, predictions of the MARS model, label data derived from Kinect sensor[169]

[168] Sizhe An and Umit Y. Ogras. 2021. MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare. ACM Trans. Embedd. Comput. Syst. 1, 1, Article 1 (January 2021), 22 pages. Fig. 5, p. 11. https://doi.org/10.1145/3477003

[169] Sizhe An and Umit Y. Ogras. 2021. MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare. ACM Trans. Embedd. Comput. Syst. 1, 1, Article 1 (January 2021), 22 pages. Fig. 6, p. 12. https://doi.org/10.1145/3477003

- Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-Trained 3D Human-Joint Coordinates

Similar to the study mentioned above, in this study[170], the mmWave radar sensor (IWR1443) and the Kinect-v2 sensor were used to label radar point cloud data. Unlike this study, Doppler velocity and intensity features were not used in the radar point cloud data, only (x, y, z) features were used in the training process. Additionally, 25 human joint points were detected from the Kinect-v2 sensor, and thus the dataset was created.

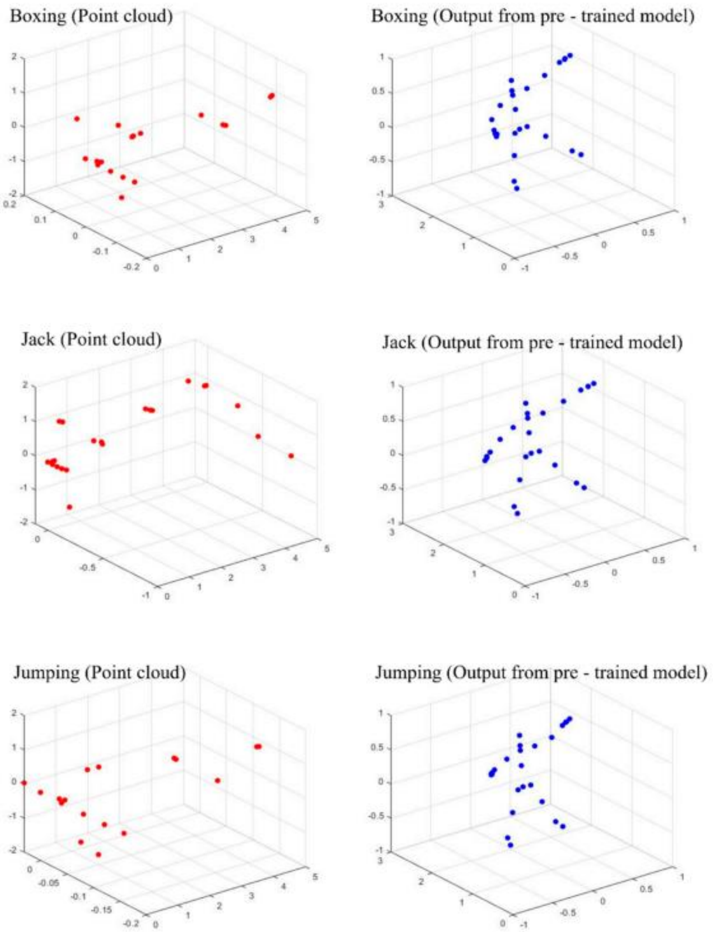


Figure 15. Radar point cloud data as input and output of joint points from pre-trained CNN model[171]

[170] Lee, G.; Kim, J. Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-Trained 3D Human-Joint Coordinates. Appl. Sci. 2022, 12, 2168. https://doi.org/10.3390/app12042168

[171] Lee, G.; Kim, J. Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-Trained 3D Human-Joint Coordinates. Appl. Sci. 2022, 12, 2168. Fig.6, p. 12. https://doi.org/10.3390/app12042168

This study involves a two-stage training process. In the first stage, a CNN model was trained with radar and Kinect data for the purpose of pose detection task. This model was used to detect 25 human joint points from radar point cloud data. In the second stage, the obtained 25 joint points were classified based on the type of movement using a Graph Neural Network-based model.

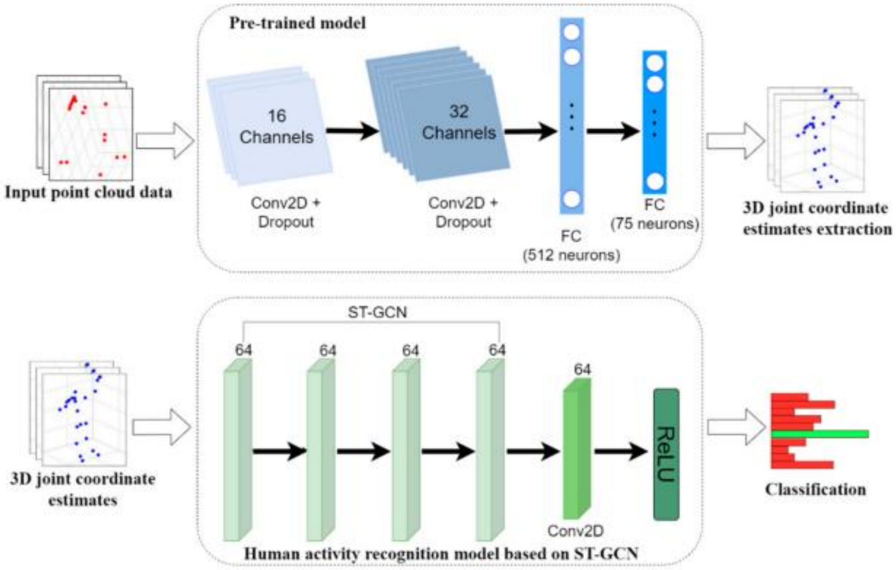


Figure 16. CNN and GNN architecture used in this paper[172]

- Indoor Detection and Tracking of People Using mmWave Sensor

This article[173] proposes a new indoor human detection and tracking system using a millimetre-wave (mmWave) radar sensor. Static Clutter Removal is performed on radar point cloud data obtained from the mmWave radar sensor (IWR1642) to remove stationary points. The remaining point cloud data is clustered using DBmeans or DBmedoids algorithms, and the locations of humans inside the indoor space are determined.

[172] Lee, G.; Kim, J. Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-Trained 3D Human-Joint Coordinates. Appl. Sci. 2022, 12, 2168. Fig.1, p. 4. https://doi.org/10.3390/app12042168

[173] Huang, X., Cheena, H., Thomas, A., Tsoi, J. K. P., & Gao, B. (2021). Indoor detection and tracking of people using mmWave sensor. Journal of Sensors, 2021, Article 6657709. https://doi.org/10.1155/2021/6657709

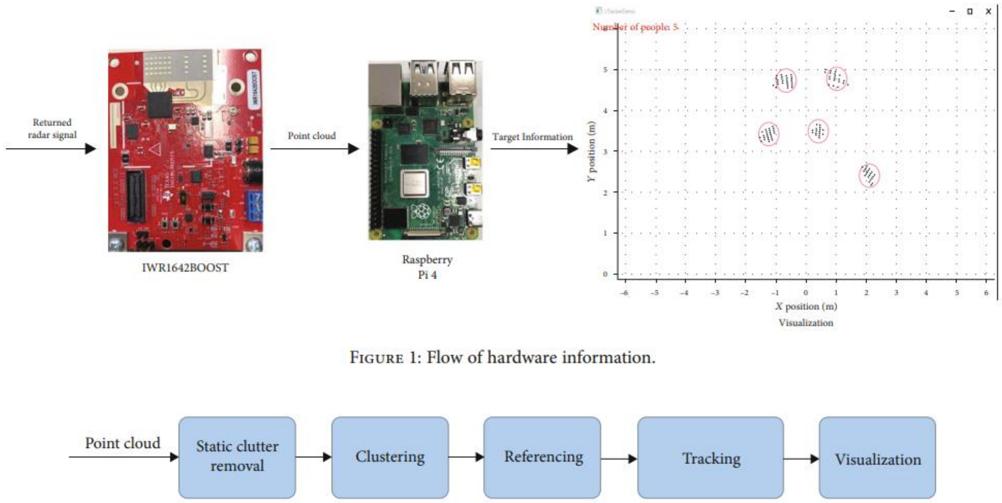


FIGURE 1: Flow of hardware information.

Figure 17. Workflow and Framework of the data process of this paper[174]

- Activity Recognition Based on Millimetre-Wave Radar by Fusing Point Cloud and Range–Doppler Information

This paper[175] proposes a multi-model deep learning approach that combines the features of both point cloud and Range-Doppler to classify six activities (boxing, jumping, squatting, walking, circling, and high-knee lifting) based on millimetre-wave radar. A CNN-LSTM model is used to extract time-series features from the point cloud, and a CNN model is utilized to obtain features from Range-Doppler. Then, they merge the two features and input the fused feature into the fully connected layer for classification. A dataset consisting of 17 volunteers is created. This study, which uses both point cloud and Range-Doppler data, also demonstrates higher accuracy compared to using each type of information separately.

[174] Huang, X., Cheena, H., Thomas, A., Tsoi, J. K. P., & Gao, B. (2021). Indoor detection and tracking of people using mmWave sensor. Journal of Sensors, 2021, Article 6657709. Fig. 1, p. 2. https://doi.org/10.1155/2021/6657709

[175] Huang, Y.; Li, W.; Dou, Z.; Zou, W.; Zhang, A.; Li, Z. Activity Recognition Based on Millimeter-Wave Radar by Fusing Point Cloud and Range–Doppler Information. Signals 2022, 3, 266-283. https://doi.org/10.3390/signals3020017

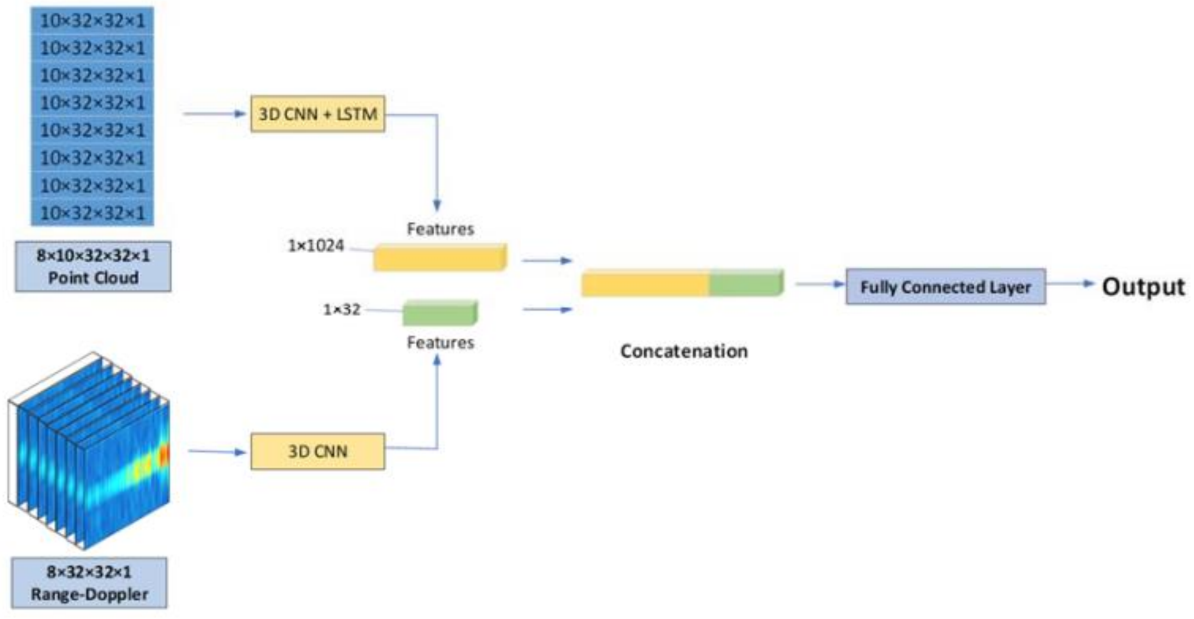


Figure 18. Deep learning architecture which used in this article[176]

2.1.2.4 Multispectral and LiDAR Sensors

Multispectral sensors capture data across multiple wavelengths, enabling the analysis of material properties and conditions. LiDAR sensors provide high-resolution 3D mapping, crucial for terrain analysis and object detection[177],[178].

2.1.2.5 Time-of-Flight (ToF) Sensors

ToF sensors measure the time it takes for a light pulse to travel to an object and back, enabling accurate distance measurement and 3D imaging. They are widely used in robotics, gesture recognition, and industrial automation[179],[180].

[176] Huang, Y.; Li, W.; Dou, Z.; Zou, W.; Zhang, A.; Li, Z. Activity Recognition Based on Millimetre-Wave Radar by Fusing Point Cloud and Range–Doppler Information. Signals 2022, 3, 266-283. Fig. 8, p. 10. https://doi.org/10.3390/signals3020017

[177] Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. Precision Agriculture, 13(6), 693-712.

[178] Liu, W., Xia, T., Wang, D., & Fu, H. (2019). LiDAR point cloud data processing and analysis in urban environments: Methods and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 149, 59-72.

[179] Hansard, M., Lee, S., Choi, O., & Horaud, R. (2013). Time-of-flight cameras: Principles, methods, and applications. Springer.

[180] Kolb, A., Barth, E., Koch, R., & Larsen, R. (2010). Time-of-flight sensors in computer graphics. Eurographics State-of-the-Art Report, 119-134.

2.1.2.6 Environmental Sensors

Environmental sensors monitor parameters such as temperature, humidity, air quality, and radiation levels. They are essential for assessing environmental conditions and detecting hazardous substances[181],[182].

2.1.2.7 Internet of Things (IoT)

IoT can be considered as a paradigm where most physical devices such as smartphones, vehicles, sensors, actuators, and all other embedded devices connect, stay in communication and exchange information with data centres[183]. Typically, there exist three types of components in an IoT network: sensors/devices, IoT gateways/local network, and backhaul network/cloud.

Sensors / Devices: The basis of IoT, sensors are structures that allow measuring various types of data in the network. For end users, devices serve as human-computer interfaces to generate users' requirements and communicate them to the IoT. All these sensors and end devices are interconnected so that they can exchange data with each other and provide additional services.

IoT Gateways: IoT gateways collect measurement data from sensors/devices and transmit it to cloud servers. Although sensors/devices can set up a network to transmit the data they produce, data pre-processing is required before being transmitted to cloud servers. Generally speaking, IoT gateways often perform data pre-processing to reduce redundancy and unnecessary overhead. Also, IoT gateways will transmit the results of data processing back to end users from cloud servers.

Cloud/Core Network: Through backhaul networks, cloud servers will receive data and requirements from end users. To support IoT applications, cloud servers have significant capacity for computing and storage. Thus, cloud servers can meet the resource requirements of different applications. When the data processing is complete, the cloud servers send the results back to the end users.

Overall, IoT can benefit from both edge and cloud computing due to the characteristics of the two structures (i.e. high computing capacity and large storage). However, despite having more limited computing capacity and storage, edge computing has more advantages for IoT over cloud computing. In particular, IoT requires fast response rather than high computational capacity and

[181] Akyildiz, I. F., Pompili, D., & Melodia, T. (2005). Underwater acoustic sensor networks: Research challenges. Ad Hoc Networks, 3(3), 257-279.

[182] Burdick, A., & Szalay, A. (2012). The case for cloud-based sensor networks. IEEE Internet Computing, 16(6), 63-67.

[183] Internet of Things (IOT): Confronts and Applications. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653.

large storage. Edge computing offers acceptable computing capacity, sufficient storage and fast response time to meet IoT application requirements[184]. On the other hand, edge computing can also take advantage of IoT by extending the edge computing structure to deal with edge computing nodes that are distributed and dynamic. IoT devices or devices with residual computational power can be used as end nodes to provide services. Importantly, several research efforts have tried to leverage cloud computing to aid the IoT, but in many cases, edge computing can provide much more competitive performance. Due to the increasing number of IoT devices, IoT and edge computing are likely to become inseparable. Characteristic of the IoT, edge and cloud computing are given in the following table.

Table 1. Comparison between characteristic of IoT, edge and cloud computing[185]

Characteristic	IoT	Edge	Cloud
Deployment	Distributed	Distributed	Centralized
Components	Physical devices	Edge nodes	Virtual resources
Computational	Limited	Limited	Unlimited
Storage	Small	Limited	Unlimited
Response Time	NA	Fast	Slow
Big data	Source	Process	Process

Layer architecture of the edge computing-based IoT is given in the figure below.

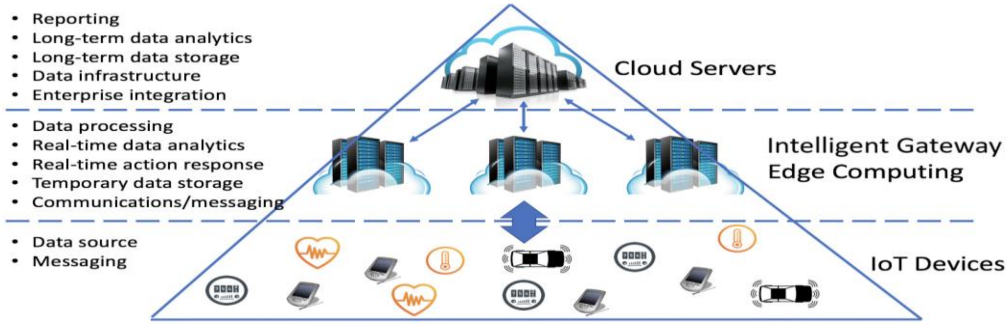


Figure 19. Layer architecture of edge computing based IoT[186]

[184] Yu, W., Liang, F., He, X., Hatcher, W. G., Lu, C., Lin, J., & Yang, X. (2017). A survey on the edge computing for the Internet of Things. IEEE access, 6, 6900-6919.

[185] Yu, W., Liang, F., He, X., Hatcher, W.G., Lu, C., Lin, J., Yang, X. A Survey on the Edge Computing for the Internet of Things, Department of Computer and Information Sciences, Towson University, MD, USA, School of Electronic and Information Engineering, Xi'an Jiaotong University, Shaanxi, P.R. China.

[186] Yu, W., Liang, F., He, X., Hatcher, W.G., Lu, C., Lin, J., Yang, X. A Survey on the Edge Computing for the Internet of Things, Department of Computer and Information Sciences, Towson University, MD, USA, School of Electronic and Information Engineering, Xi'an Jiaotong University, Shaanxi, P.R. China.

2.1.2.7.1 Communication Protocols

The integration of software and hardware components from different brands and manufacturers can be very challenging and can cause problems when different brands of hardware have different communication protocols between the devices. To solve this problem multiple connection drivers can be implemented where devices can be connected, and we can use several communication protocols such as: HTTP, MQTT, Modbus/RTU, Modbus/TCP, OPC/UA, BACnet, TCP, UDP, and DALI[187].

Hypertext Transfer Protocol (HTTP), is a protocol that allows obtaining resources, is the basis of any data exchange on the Web, and a client-server protocol, which means that requests are initiated by the recipient, usually made through a web browser. Unlike a data stream, the client and server communicate through individual message exchanges. It is not the ideal protocol for IoT device integration because of its cost, huge power consumption, and weight issues, but it is still used because of the large amounts of data it can publish.

Message Queuing Telemetry Transport (MQTT) is a lightweight IoT data protocol. It features a publisher-subscriber messaging model and allows simple data flow between different devices. The main advantage of MQTT is its architecture, its composition is basic and lightweight and can provide low power consumption for devices. It works on top of a TCP/IP protocol. IoT data protocols are designed to connect with unreliable communication networks. This has become a necessity in the IoT world due to the increasing number of objects appearing on the network in recent years. Despite the wide adaptation of MQTT, it does not support a defined data representation and device management framework mode.

Modbus/RTU defines a way of interpreting data that is sent and received by a device, so it is a communication protocol. The communication model is of the master-slave type. Thus, a slave should not initiate any type of communication in the physical environment until it has been requested by the master. This communication allows the connection of several devices at the same time, one of which is the master, which coordinates the communication, and the rest are the slaves. It is very common when communicating with a PLC to use Modbus/RTU.

Modbus/TCP is a simple and easy to implement protocol, it ends up being applied in most industrial equipment that uses some networked technology, it is an open protocol so it can be implemented freely in any equipment. To apply this protocol in architecture in the physical environment, Modbus has a TCP communication protocol, one of the main advantages of this communication is the ease of implementation of infrastructure, through switches or industrial hubs, communication can reach very high speeds. The protocol can integrate devices installed in

[187] Lee, J., Bagheri, B., & Kao, H. A. (2015). A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems.

the field and allows the exchange of information between them without restrictions, that is, each user can connect directly to the servers.

Open Platform Communications Unified Architecture (OPC/UA) is a protocol for industrial automation, it enables information and data exchange on devices inside machines, between machines, and from machines to systems. OPC/UA circumvents the division between information technology and operational technology. In other words, we will not be able to have the benefits of the Internet of Things and Industry 4.0 without OPC/UA. It is designed to allow manufacturers to leverage all modern technology to help create a smart factory, being able to make use of mobile devices, big data, machine learning, artificial intelligence, and predictive maintenance. OPC/UA is the means to connect machines and devices.

Building Automation and Control networks (BACnet) is a network protocol for centralized technical management. It makes it possible to establish seamless communication between field devices and control technology with a standard that is free to use. The strengths are worldwide standardization as an open and vendor-neutral protocol. This protocol uses an object-oriented approach to standardize the representation of processes and data within a device, provides standard services to access the data within a device, and provides more than a physical interface to accommodate small, medium, and large systems.

Transmission Control Protocol (TCP) is the most common communication protocol used on the internet, responsible for dividing the message into datagrams, reassembling them, and retransmitting lost datagrams. It is connection-oriented, which means data can be sent bidirectionally once a connection is established. It includes an automatic error-checking system to ensure that each packet is delivered as requested. The Internet Protocol (IP) is responsible for routing datagrams which is no easy task on the internet as the connection may want the datagram to traverse several networks until it reaches its destination.

User Datagram Protocol (UDP) is a communication protocol that has as an essential characteristic, unreliability. This means that, by using this protocol, it is possible to send datagrams from one machine to another, but with no guarantee that the data sent will arrive intact and in the correct order, unlike TCP. UDP is a protocol that does not follow a connection, this means that it is not necessary to establish communication, this way with UDP it is possible to send, through the same output, data to several different machines without any problem. UDP becomes an advantageous protocol when you want to deal with services where speed is something fundamental and the minimal loss of data is not very disadvantageous.

Digital Addressable Lighting Interface (DALI) is a communication protocol for building lighting applications and is used for communication between lighting control devices, such as light sensors or motion detectors. This protocol maximizes flexibility by simply adjusting the light control to new conditions. It is a protocol that has many advantages, such as being an open protocol, any user can use it, interoperability between manufacturers is guaranteed by mandatory certification procedures. The installation is simple, the communication is digital, not analogical, so the same darkening values can be received by multiple devices, resulting in a more stable and accurate darkening performance and all devices have their unique address in the system, opening many possibilities for flexible control.

With the use of all these protocols, it is possible to guarantee the interoperability of the system.

2.1.2.7.2 MQTT

MQTT is a publish-subscribe-based messaging protocol used in the internet of Things. It works on top of the TCP/IP is designed for connections with remote locations where a "small code footprint" is required or the network bandwidth is limited. The goal is to provide a protocol, which is bandwidth-efficient and uses little battery power[188].

HTTP can serve as a transport mechanism between devices and the IoT Agent, utilizing a request/response model where each device connects directly to the IoT Agent. In contrast, MQTT operates on a publish-subscribe model, which is event-driven and pushes messages to clients. MQTT requires a central communication point, known as the MQTT broker, responsible for dispatching messages between senders and receivers. When a client publishes a message to the broker, it includes a topic in the message, which serves as routing information for the broker. Clients wishing to receive messages subscribe to a specific topic, and the broker delivers all messages with the matching topic to those clients. This architecture allows for highly scalable solutions without dependencies between data producers and data consumers, as clients communicate solely through the topic.

[188] Vermesan, O., & Friess, P. (Eds.). (2014). Internet of Things: From Research and Innovation to Market Deployment.

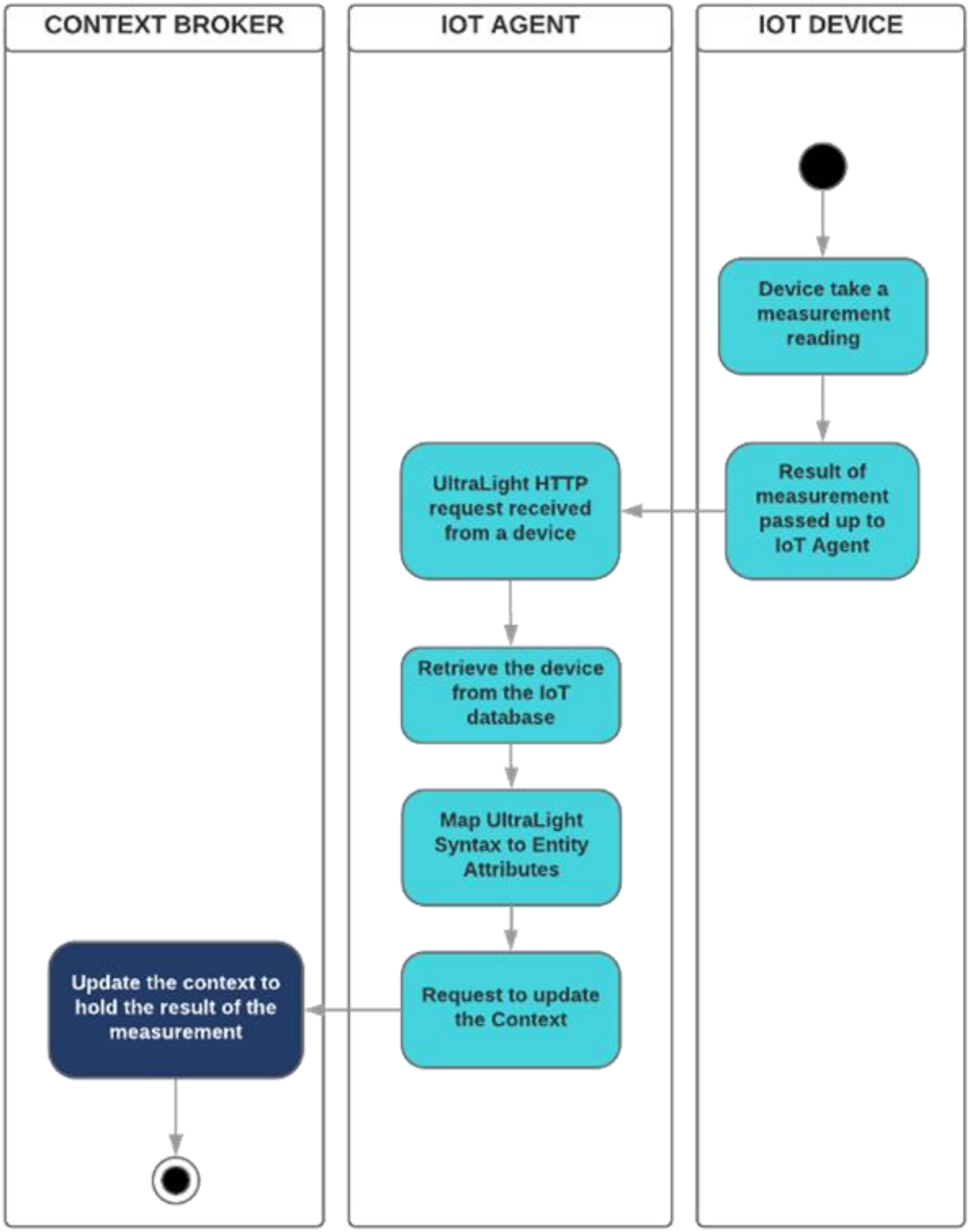

CONTEXT BROKER
IOT AGENT
IOT DEVICE
Device take a measurement reading
Result of measurement passed up to IoT Agent
UltraLight HTTP request received from a device
Retrieve the device from the IoT database
Map UltraLight Syntax to Entity Attributes
Request to update the Context
Update the context to hold the result of the measurement

Figure 20. Sample Sensor Measurement and pass to the IoT Agent

2.1.2.7.3 Fireware

FIWARE[189] is a platform aims to manage context data in a generalized set of standards with the use of its APIs to implement in smart solutions. Context data are the virtual representations of the real-world objects, people, and relationships between them. FIWARE components are open-source, and the middleware for the platform is the Orion Context Broker. Orion Context Broker provides an API for managing context data that is called NGSIv2 API.

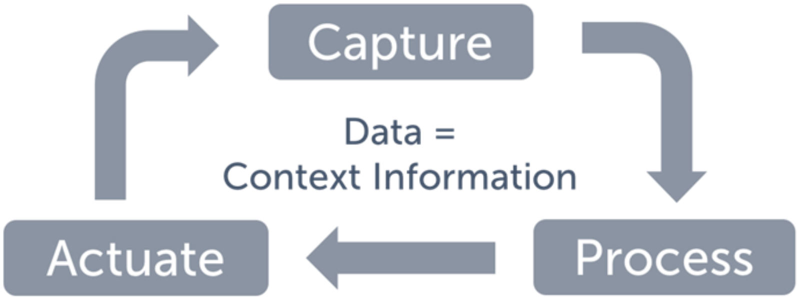


Figure 21. Context Data Flow (FIWARE, 2020)

FIWAREs other components support the context broker in terms of:

- supplying context data from various sources (IoT, social networks, robots),
- managing context data,
- processing, analysing and visualization of context data,
- accomplishing complex event processes,
- authorization, access control and monetization.

Following services can be used for receiving-sending data, recording, visualizing and analysing data. Additional services may be added later. The services and operating systems mentioned below are all open-source and community driven which requires no purchase.

- Orion Context Broker (FIWARE), middleware for holding the latest state of the virtual entities and sending updates to other services, databases with subscriptions.
- MongoDB, no-sql database that will be used by Orion and Draco. Orion will store virtual entities and subscriptions. Draco will store past data of these virtual entities.
- Draco, an alternative data persistence mechanism for managing the history of virtual entities.

[189] https://www.fiware.org/.

- Mosquitto, message broker that will implement the MQTT protocol for the electrical motor.
- ROS, robotics middleware that will be used in robot.
- FIROS, tool for translating ROS messages into NGSI to publish them in Orion.
- IoT-Agent-Ultralight, an IoT Agent that will translate MQTT messages into NGSI to publish them in Orion.
- User Interface: Interface for making http requests on Orion Context Broker

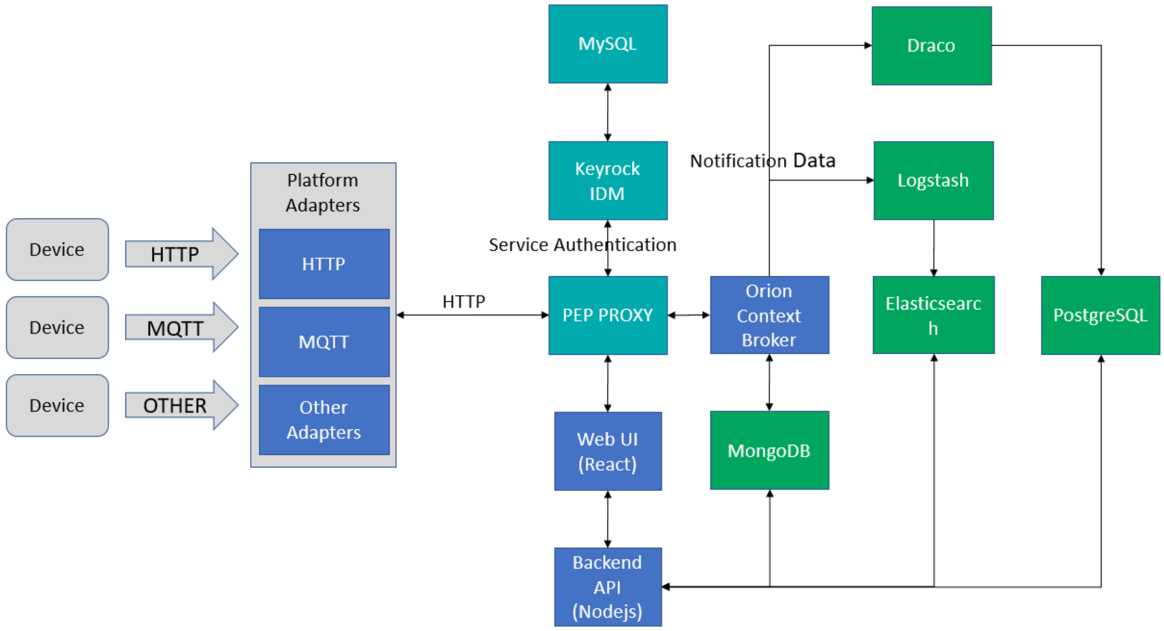


Figure 22. General data flow Fiware infrastructure.

2.1.2.7.4 Arrowhead

Arrowhead[190] is a framework consisting of local clouds, devices, systems, and services. Its primary objective is to ensure interoperability among heterogeneous systems by utilizing existing protocols to manage legacy systems. Arrowhead has been implemented in various IoT automation scenarios, including the efficient deployment of numerous IoT sensors, monitoring of programmable logic controller (PLC) devices, replacement of devices, energy optimization, and maintenance.

[190] https://www.arrowhead.eu/.

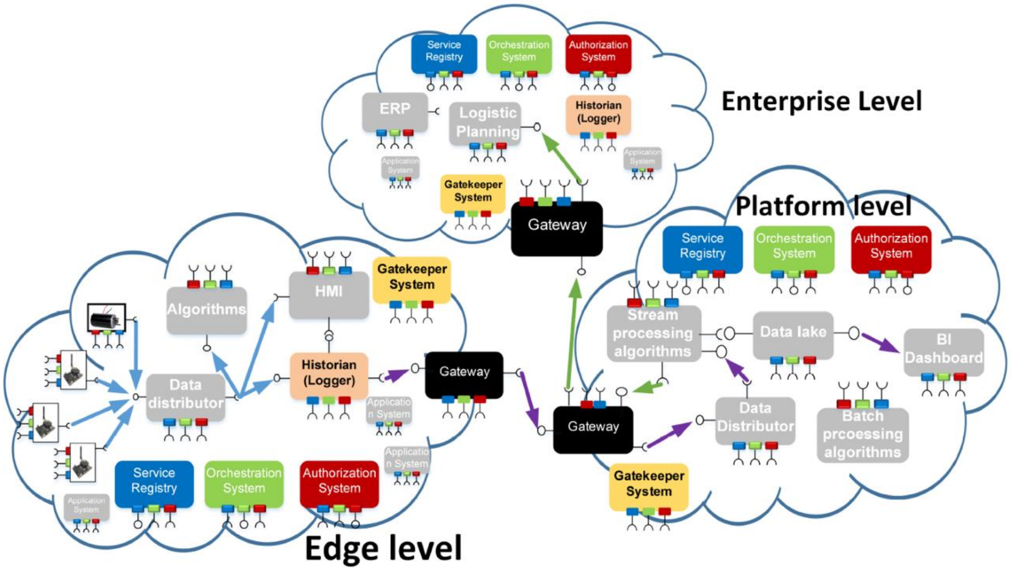


Figure 23. Example of arrowhead structure.

2.1.2.7.5 Communication Models

Three different communication models are used in IoT[191].

Machine to Machine Communication: In this communication model, machines are directly connected to each other without the aid of any intermediary hardware. This model makes sense for systems that communicate with each other by sending small data packets and have relatively low data rate requirements.

Machine to Cloud Communication: This communication model relies on requesting services from a cloud application service provider or keeping data in cloud storage due to the limitations of the computing capability of the devices. Although this model solves the problems of the M2M model, traditional networking and bandwidth and network resources limit the performance of this communication model.

Machine to Gateway Communication: In this communication model, the device-to-application layer gateway model is considered a proxy or middleware box. At the application layer, some software-based security control schemes, or other functions, such as data or protocol translation

[191] A. Botta, W. D. Donato, V. Persico and A. Pescapé, "Integration of cloud computing and Internet of Things: A survey," Future Gener. Comput. Syst., vol. vol. 56, p. pp. 684–700, 2016.

algorithms, operate on a gateway or other network device that acts as an intermediate bridge between IoT devices and cloud application services.

Examples of the above communications methods are given in figure below.

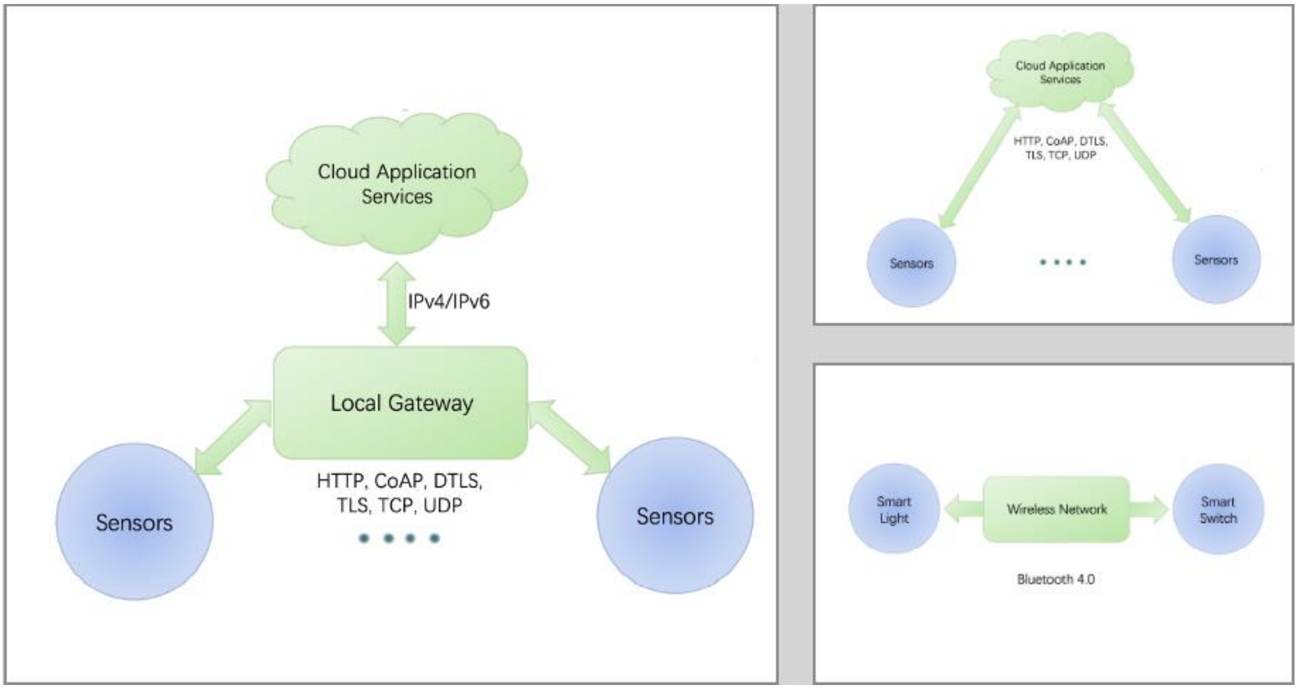

Cloud Application Services
IPv4/IPv6
Local Gateway
Sensors
HTTP, CoAP, DTLS, TLS, TCP, UDP
Sensors
Cloud Application Services
HTTP, CoAP, DTLS, TLS, TCP, UDP
Sensors
Sensors
Smart Light
Wireless Network
Smart Switch
Bluetooth 4.0

Figure 24. Examples of different communication models (a) M2M, (b) M2C, (c) M2G

2.1.3 Data

2.1.3.1 Localization

Combining pixel location with real-world location information is a common task in computer vision and robotics, and enhancing this process with deep learning techniques can provide more accurate and robust results. Before mapping pixel locations to real-world coordinates, it's essential to calibrate the camera. Camera calibration involves estimating intrinsic parameters (such as focal length, principal point, and distortion coefficients) and extrinsic parameters (position and orientation) of the camera relative to the scene. Techniques like Zhang's method or Tsai's[192] method are commonly used for camera calibration.

Deep learning models, particularly convolutional neural networks (CNNs), are adept at extracting features from images. Techniques like transfer learning, where pre-trained CNN models (e.g.,

[192] LI, Wei, et al. A practical comparison between Zhang's and Tsai's calibration approaches. In: Proceedings of the 29th International Conference on Image and Vision Computing New Zealand. 2014. p. 166-171.

VGG, ResNet, etc.) are fine-tuned on specific tasks, can be employed to extract features[193] relevant to the mapping between pixel locations and real-world coordinates.

Semantic segmentation[194] is the task of classifying each pixel in an image into a specific category. By segmenting objects in the image, it becomes easier to associate pixel locations with real-world objects or regions. Deep learning models like Fully Convolutional Networks[195] (FCNs) or U-Net are commonly used for semantic segmentation tasks.

Object detection is another technique that can be employed to detect and locate objects of interest in an image. Models like Faster R-CNN, YOLO[196] (You Only Look Once), or SSD[197] (Single Shot Multi Box Detector) can detect objects and provide bounding boxes, which can then be used to estimate their real-world positions.

Once pixel locations are associated with real-world coordinates, geometric transformations such as perspective transformation[198] or homography[199] can be used to map between the two spaces accurately. Deep learning models can be used to learn these transformations directly from the data.

Data augmentation[200] techniques such as rotation, scaling, translation, and flipping can be applied to both pixel locations and real-world coordinates to augment the training data and improve the robustness of the model.

Instead of separating the tasks of feature extraction, object detection, and geometric transformation, end-to-end learning approaches can be employed where a single deep learning model is trained to directly map pixel locations to real-world coordinates[201]. This approach can

[193] NAMATĒVS, Ivars. Deep convolutional neural networks: Structure, feature extraction and training. Information Technology and Management Science, 2017, 20.1: 40-47.

[194] HAO, Shijie; ZHOU, Yuan; GUO, Yanrong. A brief survey on semantic segmentation with deep learning. Neurocomputing, 2020, 406: 302-321.

[195] OZTURK, Ozan; SARITÜRK, Batuhan; SEKER, Dursun Zafer. Comparison of fully convolutional networks (FCN) and U-Net for road segmentation from high resolution imageries. International journal of environment and geoinformatics, 2020, 7.3: 272-279.

[196] FAN, Jiayi, et al. Improvement of object detection based on faster R-CNN and YOLO. In: 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). IEEE, 2021. p. 1-4.

[197] NING, Chengcheng, et al. Inception single shot multibox detector for object detection. In: 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2017. p. 549-554.

[198] DUBROFSKY, Elan. Homography estimation. Diplomová práce. Vancouver: Univerzita Britské Kolumbie, 2009, 5.

[199] DUBROFSKY, Elan. Homography estimation. Diplomová práce. Vancouver: Univerzita Britské Kolumbie, 2009, 5.

[200] MUMUNI, Alhassan; MUMUNI, Fuseini. Data augmentation: A comprehensive survey of modern approaches. Array, 2022, 16: 100258.

[201] CHEN, Changhao, et al. A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence. arXiv preprint arXiv:2006.12567, 2020.

potentially capture complex relationships between the input image and the real-world environment more effectively.

Designing appropriate loss functions is crucial for training deep learning models to perform pixel-to-real-world mapping tasks[202]. Loss functions such as mean squared error (MSE), smooth L1 loss, or custom loss functions tailored to the specific requirements of the task can be used to train the model effectively[203].

Challenge 1: A key challenge in camera-based localization is eliminating interference caused by background. Some powerful techniques from the computer vision domain have opened up the potential of obtaining occupancy information from CCTV videos[204]. Some early efforts applied pattern recognition technologies (e.g., filtering algorithms, classification, and clustering methods) to subtract background information from videos, but these background subtraction-based approaches can fail if occupants remain static for extended periods.

Challenge 2: A challenge emerged as the extracted feature points were prone to change due to factors such as variations in illumination from day to night or weather or seasonal changes. To address this issue, recent efforts have focused on data-driven approaches using deep learning for feature-point extraction[205],[206].

[202] KAKANI, Vijay, et al. Feasible self-calibration of larger field-of-view (FOV) camera sensors for the advanced driver-assistance system (ADAS). Sensors, 2019, 19.15: 3369.

[203] EBERT-UPHOFF, Imme, et al. CIRA Guide to Custom Loss Functions for Neural Networks in Environmental Sciences--Version 1. arXiv preprint arXiv:2106.09757, 2021.

[204] C. Feng, A. Mehmani, and J. Zhang, “Deep learning-based real-time building occupancy detection using AMI data,” IEEE Trans. Smart Grid, vol. 11, no. 5, pp. 4490–4501, Sep. 2020.

[205] Dusmanu, M.; Rocco, I.; Pajdla, T.; Pollefeys, M.; Sivic, J.; Torii, A.; Sattler, T. D2-net: A trainable cnn for joint description and detection of local features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8092–8101.

[206] DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-supervised interest point detection and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 224–236.

The positioning of cameras and radar plays a crucial role in Human Action Recognition (HAR) analysis. Positioning is of great importance in order to obtain accurate data and to minimize the negative effects of environmental factors on the system[207],[208],[209],[210],[211],[212],[213],[214],[215].

Proper camera placement ensures a wide and unobstructed field of view, capturing the necessary details for action recognition. The camera angle and height can affect the visibility of actions and the ability to distinguish between different types of movements. Cameras need to be positioned to minimize glare and shadows that could obscure or distort the actions being recorded. Using multiple cameras at different angles can help overcome occlusions and provide a more comprehensive view of the actions.

Unlike cameras, radar systems can recognize actions without capturing visual images, thus preserving individuals' privacy.

Radar devices have fewer installation requirements compared to cameras, offering more flexibility in positioning. Radars can detect and recognize actions even without a direct line of sight, making them useful in various environments. Radar systems are less affected by lighting conditions and can operate in complete darkness or through smoke and fog.

In HAR analysis, the synergy between camera and radar positioning can significantly enhance the system's ability to accurately recognize and analyse human actions, especially in complex and dynamic environments. This is essential for applications ranging from surveillance and security to healthcare and human-computer interaction.

The challenges of camera and radar positioning in Human Action Recognition (HAR) projects are multifaceted and can significantly impact the effectiveness of the system. The challenges listed

[207] Pareek, P., Thakkar, A. A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications.

[208] Othman, N.A., Aydin, I. (2021). Challenges and limitations in human action recognition on unmanned aerial vehicles: A comprehensive survey.

[209] Kumar, P., Chauhan, S. & Awasthi, L.K. Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies.

[210] Progress and Future Research Directions. Arch Computat Methods Eng.

[211] Singh, P.K., Kundu, S., Adhikary, T. et al. Progress of Human Action Recognition Research in the Last Ten Years: A Comprehensive Survey. Arch Computat Methods Eng.

[212] Hieu H. Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, Sergio A. Velastin, Computer Vision and Pattern Recognition Video-based Human Action Recognition using Deep Learning: A Review

[213] Beddiar, D.R., Nini, B., Sabokrou, M. et al. Vision-based human activity recognition: a survey.

[214] Kong, Y., Fu, Y. Human Action Recognition and Prediction: A Survey. Int J Comput Vis

[215] Saleem, G., Bajwa, U.I. & Raza, R.H. Toward human activity recognition: a survey. Neural Comput & Applic

below require careful planning, and ongoing evaluation to ensure HAR systems are effective and reliable[216],[217],[218],[219],[220],[221],[222],[223],[224].

Challenge 1: Finding the optimal location for cameras and radars to ensure comprehensive coverage and minimize blind spots is challenging, especially in complex environments. Cameras and radars must be positioned to cope with environmental factors such as lighting, weather conditions, and physical obstructions that can affect data quality.

Challenge 2: Especially for cameras, positioning must be considered carefully to respect privacy while still capturing necessary data for action recognition.

Challenge 3: The cost of equipment and the need for supporting infrastructure can limit the number of devices installed, affecting the system's overall performance. Regular calibration and maintenance are required to keep the system accurate, which can be logistically challenging and costly.

Challenge 4: The angle and elevation of cameras and radars can affect the detection range and the ability to distinguish between different actions or individuals. For real-time action recognition, the positioning must facilitate quick data transmission and processing without delays.

2.1.3.2 Human Heat-Map

The integration of human heat-map technology into security systems at airports has seen significant advancements in recent years, propelled by increasing demands for enhanced surveillance and threat detection capabilities. This state-of-the-art analysis reviews the latest

[216] Pareek, P., Thakkar, A. A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications.

[217] Othman, N.A., Aydin, I. (2021). Challenges and limitations in human action recognition on unmanned aerial vehicles: A comprehensive survey.

[218] Kumar, P., Chauhan, S. & Awasthi, L.K. Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies, Progress and Future Research Directions. Arch Computat Methods Eng

[219] Singh, P.K., Kundu, S., Adhikary, T. et al. Progress of Human Action Recognition Research in the Last Ten Years: A Comprehensive Survey. Arch Computat Methods Eng

[220] Hieu H. Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, Sergio A. Velastin, Computer Vision and Pattern Recognition Video-based Human Action Recognition using Deep Learning: A Review

[221] Beddiar, D.R., Nini, B., Sabokrou, M. et al. Vision-based human activity recognition: a survey.

[222] Kong, Y., Fu, Y. Human Action Recognition and Prediction: A Survey. Int J Comput Vis

[223] Saleem, G., Bajwa, U.I. & Raza, R.H. Toward human activity recognition: a survey. Neural Comput & Applic

[224] Giovanni Diraco,Gabriele Rescio,Andrea Caroppo,Andrea Manni andAlessandro Leone Human Action Recognition in Smart Living Services and Applications: Context Awareness, Data Availability, Personalization, and Privacy

developments in the field, emphasizing novel methodologies and their implications for airport security.

Recent studies have focused on utilizing advanced imaging and machine learning techniques to improve the detection and analysis of human activities within airport environments. For example, extended motion diffusion-based methods have been developed for more effective surveillance, especially in detecting subtle movements or behaviours indicative of potential threats[225]. The fusion of data from multiple sources, such as thermal imaging and standard CCTV footage, enhances the detection capabilities. A notable advancement in this area includes the use of multimodal semantic segmentation to delineate airport runways, which, while not directly related to threat detection, demonstrates the potential for multi-source integration in enhancing overall airport security operations[226].

The application of machine learning algorithms, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has been pivotal in interpreting the data obtained from heat-maps. These algorithms facilitate the recognition of patterns and anomalies in human behaviour that may indicate security threats[227]. Another critical application of human heat-maps in airports is thermal passenger screening, which has been widely adopted as a measure to prevent the spread of infectious diseases. The effectiveness of this technique, however, has been varied, with studies suggesting improvements are needed to catch a higher percentage of infected travellers[228]. Innovative researches have also been directed toward using heat-map data for behavioural analysis. The ability to detect unusual or erratic human behaviour through changes in heat signatures provides a non-invasive way to enhance monitoring and ensure passenger safety[229].

[225] Zhang, X., Wu, H., Wu, M., & Wu, C. (2020). Extended Motion Diffusion-Based Change Detection for Airport Ground Surveillance. IEEE. Retrieved from IEEE Xplore

[226] Datla, R., Chalavadi, V., & Chalavadi, K. M. (2022). A multimodal semantic segmentation for airport runway delineation in panchromatic remote sensing images. International Conference on Machine Vision. Retrieved from Semantic Scholar

[227] Guo, H., Fan, X., & Wang, S. (2017). Human attribute recognition by refining attention heat map. Pattern Recognition Letters. Retrieved from Pattern Recognit. Lett.

[228] Quilty, B., Clifford, S., Flasche, S., & Eggo, R. (2020). Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. Retrieved from Eurosurveillance

[229] Ahmad, N., & Yoon, J. (2021). StrongPose: Bottom-up and Strong Keypoint Heat Map Based Pose Estimation. International Conference on Pattern Recognition. Retrieved from ICPR 2020

2.1.4 Chemical Analysis of Suspected Materials

Chemical material analysis is essential in various fields, including security, law enforcement, environmental monitoring, and pharmaceuticals. Accurate identification of suspected materials, such as explosives and illegal drugs, ensures safety, regulatory compliance, and public health. Among the numerous analytical techniques available, spectroscopy methods, especially Near Infrared Spectroscopy (NIRS), have gained significant attention due to their potential for rapid, non-destructive, and on-site analysis. This document delves into the state of the art in chemical material analysis, focusing on drugs and explosive substances, the potential of spectroscopic methods, and the potential of NIRS in miniaturization and field applications.

2.1.4.1 Suspected Drug and Explosives Identification

Fast and accurate on-scene identification of suspected materials, particularly drugs and explosives, is crucial for safety and efficient resource use. Underestimating the threat can lead to severe injuries or proliferation of drug abuse, whereas overestimating the threat can enable criminal exploitation of hoax materials and lead to a general waste of resources. The ability to identify the suspected material directly at the scene-of-crime can steer the investigation process and provide important information for decisions such as search warrants, arrests, and requests for laboratory analysis[230]. Additionally, early information on substance identity increases the safety of investigators. On-site analysis without sending samples to a lab is ideal, requiring reliable techniques that provide admissible evidence. The diverse chemical nature of drugs and explosives complicates visual identification, necessitating portable technology capable of precise identification.

Traditionally, investigation officers use chemical spot tests for presumptive drug or explosive testing. In these so-called colorimetric tests, a colour can be observed after reaction with a specific substance, such as a blue colour for the cobalt(II)thiocyanate complex with cocaine in the Scott test. Unfortunately, colorimetric tests are only available for a small range of drugs, are prone to false positive reactions, require manual handling of the suspect material, destruction of the evidence and require single-use consumables and chemicals[231]. Moreover, the interpretation of the colour formation is somewhat subjective. For on scene detection of explosives, these tests detect classes of compounds and can indicate the possible presence of an explosive, but lack selectivity and are typically unable to identify an explosive within a class[232]. In the case of potential explosives, manual sampling also introduces significant risks to the officer. In any case, a single

[230] Kranenburg, R. F., Ramaker, H. J., & van Asten, A. C. (2022). Portable near infrared spectroscopy for the isomeric differentiation of new psychoactive substances. *Forensic Science International*, *341*, 111467.

[231] M. Philp, S. Fu, A review of chemical "spot" tests: a presumptive illicit drug identification technique, Drug Test. Anal 10 (2018) 95–108, https://doi.org/10.1002/dta.2300.

[232] Almog, J.; Zitrin, S. Colorimetric Detection of Explosives. In Aspects of Explosives Detection, 1st ed.; Marshall, M., Oxley, J., Eds.; Elselvier: Oxford, UK, 2009; pp. 41–58. ISBN 978-0-12-374533-0.

colorimetric test covers only a sub-range of drugs or explosives, and it may not be possible to perform a series of multiple tests given limited evidence samples.

Gas chromatography – mass spectrometry (GC-MS) is currently the default technique for unambiguously identifying common drugs of abuse in forensic samples[233]. However, this reliable technique is expensive, requires experienced operators, is not portable due to the requirement of delicate stable vacuum systems and is thus intended for use only in dedicated laboratory facilities rather than on-site testing. A widely used method for rapid explosives detection is ion mobility spectrometry (IMS)[234]. Given its high sensitivity, it is used primarily by aviation security to detect trace amounts on luggage. IMS is less favourable for identifying bulk amounts due to the potential overloading of the instrument, which can lead to false-positive results in subsequent analyses[235].

2.1.4.2 Spectroscopic Methods

Spectroscopic techniques are well-suited for on-site evaluation of suspected samples due to their ability to deliver highly-specific chemical information with minimal or no need for sample preparation. These methods allow for rapid and accurate analysis, making them invaluable in scenarios where timely decisions are critical. Furthermore, the non-destructive nature of many spectroscopic techniques preserves the integrity of the sample for potential further testing or evidence collection. This combination of detailed chemical insights, efficiency, and preservation makes spectroscopy an essential tool for field-based investigations and real-time assessments.

Raman spectroscopy is non-invasive but can be affected by fluorescence, and Raman laser sources have sufficient power to potentially burn samples and pose ignition risks[236]. On the other hand, Fourier-transform infrared spectroscopy (FT-IR) is safer but requires sampling and more extensive sample preparation[237].

In contrast, near-infrared (NIR) spectroscopy is non-invasive, not influenced by fluorescence, and poses minimal ignition risks. Additionally, NIR analysers are relatively inexpensive and can be

[233] R.F. Kranenburg, A.R. García-Cicourel, C. Kukurin, H.-G. Janssen, P. J. Schoenmakers, A.C. van Asten, Distinguishing drug isomers in the forensic laboratory: GC-VUV in addition to GC-MS for orthogonal selectivity and the use of library match scores as a new source of information, Forensic Sci. Int. (2019) 109900, https://doi.org/10.1016/j.forsciint.2019.109900.

[234] Ewing, R.G.; Atkinson, D.A.; Eiceman, G.A.; Ewing, G.J. A critical review of Ion Mobility Spectrometry for the detection of explosives and explosive related compounds. Talanta **2001**, 54, 515–529.

[235] Madhusudhan, P.; Latha, M.M. Ion Mobility Spectrometry for the detection of explosives. Int. J. Eng. Res. Technol. 2013, 2, 1369–1372.

[236] Harvey, S.; Peters, T.J.; Wright, B.W. Safety considerations for sample analysis using a Near-Infrared (785 nm) Raman laser source. Appl. Spectrosc. **2003**, 57, 580–587.

[237] Benson, S.; Speers, N.; Otieno-Alego, V. Portable explosive detection instruments. In Forensic Investigation of Explosions, 2nd ed.; Beveridge, A., Ed.; CRC Press: Boca Raton, FL, USA, 2011; pp. 691–724. ISBN 978-0-367-77820-0.

miniaturized, making them suitable for routine field use. It measures the absorption of near-infrared light by molecular overtones and combinations of vibrational modes, primarily involving C-H, N-H, and O-H bonds. However, NIR spectra alone are often insufficient for structural elucidation because the bands in the NIR region (780–2500 nm) are weak and result from complex combined vibrations and overtone absorptions[238],[239]. Nonetheless, this lack of signal interpretability can be overcome by pre-processing the raw data and applying chemometric methods (multivariate data analysis) to extract informative features from the NIR spectral measurements[240].

2.1.4.3 Near Infrared Spectral Sensing

For decades, near-infrared spectroscopy (NIRS) has played an important role in countless applications, ranging from monitoring industrial processes to assessing the chemical composition and quality of organic-based materials. However, traditionally, spectrometers are large, expensive, complex and include moving parts, making them sensitive to vibrations and shocks. The challenge today lies in reducing the size and cost of these spectroscopic devices while maintaining their robustness and sensitivity. This is essential to expand their application beyond dedicated stations in industrial settings and analytical labs, into the hands of non-specialists working on-site and in the field.

The design of current portable NIR sensor systems is mainly focused on the miniaturization of conventional spectrometers using gratings or interferometers. While they represent valid options for portable NIR spectroscopy, the size and cost of most commercially available systems are still relatively large. The level of miniaturization, cost and production scalability required for consumer applications can only be reached with wafer-scale integration – analogous to how complementary metal-oxide-semiconductor (CMOS) cameras came to pervade industrial and consumer applications. There has been substantial progress in this direction for the visible spectral region (c. 400-700 nm) and up to 1100 nm, utilizing mature silicon technologies[241]. However, progress in the integration of NIR spectral sensors has been relatively slow.

Recently, a novel approach to NIR spectral sensing was proposed, using a miniaturized fully-integrated multipixel array of resonant-cavity-enhanced (RCE) InGaAs photodetectors [13]. Their mm-scale footprint and wafer-scale fabrication make them appealing for portable and embedded

[238] Ozaki, Y.; Morisawa, Y. Principles and Characteristics of NIR spectroscopy. In Near-Infrared Spectroscopy, 1st ed.; Ozaki, Y., Huck, C., Tsuchikawa, S., Engelsen, S.B., Eds.; Springer: Singapore, 2021; pp. 11–36. ISBN 978-981-15-8647-7.

[239] Small, G.W. Chemometrics and Near-Infrared Spectroscopy: Avoiding the pitfalls. TrAC **2006**, 25, 1057–1066.

[240] Kranenburg, R.F.; Verduin, J.; Weesepoel, Y.; Alewijn, M.; Heerschop, M.; Koomen, G.; Keizers, P.; Bakker, F.; Wallace, F.; van Esch, A.; et al. Rapid and robust on-scene detection of cocaine in street samples using a handheld Near-Infrared spectrometer and machine learning algorithms. Drug Test Anal. **2020**, 12, 1404–1418.

[241] R.A. Crocombe, *Appl. Spectrosc.*, **72**(12), 1701–1751 (2018). DOI: 10.1177/0003702818809719.

NIR sensing. The approach utilized by this multipixel sensor is different from conventional spectrometry, as the sensor does not endeavour to measure the full spectrum, but rather a limited number of spectral regions with limited resolution (~50 nm). The target biochemical information is directly extracted from the power reflected/transmitted by the sample in these regions without any intermediate step for full spectral reconstruction. The multipixel sensor had a footprint of 1.8-2.2 mm^2 and consists of an array of 16 pixels with tailored spectral responses in the 850-1700 nm wavelength range. Each pixel is fabricated within a single monolithic element, having a thin absorbing layer and a tuning layer inside an optical cavity. In this approach, the detector and filter elements are directly co-integrated at the wafer-level, providing a robust system which can be fabricated at high volumes using standard semiconductor processing methods[242].

2.1.5 Sensitive Data Privacy

Based on the project requirements and identified use cases, sensitive data privacy preserving topics related to SINTRA are “Distributed privacy preservation” and “Video-based privacy protection”.

2.1.5.1 Distributed Privacy Preservation

The authors in [1] provide a systematic review of sensitive data privacy-preserving methods in the context of cloud computing. Most of the solutions that have been reviewed involve masking sensitive data so that only protected values are stored in the cloud, and only the data owner can unmask it. However, manipulation of masked data is challenging because the masking method should be made compatible with computations needed for exploration. From a technical point of view, there are three types of data protection techniques with respect to privacy, i.e., i) data splitting, ii) data anonymization, and iii) cryptographic.

Data splitting is based on fragmenting sensitive data [2]. In this technique, a clear form of each fragment is stored in a separate location based on standard mechanisms like RAID, and computations are outsourced on split data. Although storing data fragments in clear form makes it possible to support exploration seamlessly, there exist two main challenging conditions for this technique that must be considered. First, each fragment should neither provide re-identification of the specific individuals nor disclose confidential information. Second, storage locations must be independent and unlinked to prevent collusion attacks.

Data anonymization masks data in a privacy-preserving way. This masking method must be irreversible so that protected data stay analytically useful for exploration but do not disclose

[242] K.D. Hakkel, M. Petruzzella, F. Ou, A. van Klinken, F. Pagliano, T. Liu, R.P.J. van Veldhoven and A. Fiore, *Nat. Commun.* **13**(1), 1–8 (2022), DOI: 10.1038/s41467-021-27662-1.

information that can be linked to a subject. Generally, a downside of the data anonymization method is that the result of computations on masked data is approximate rather than exact. Masking methods used for data anonymization are classified as *non-perturbative* and *perturbative* [1]. Perturbative masking, based on the "original values plus some noise" approach, may preserve the statistical properties of the original data better than non-perturbative masking, which reduces the data accuracy. K-anonymity is one of the data anonymization models proposed in [3] that is robust against identity disclosure; however, attribute disclosure can happen since it cannot prevent attacks that combine multiple records in the anonymized data set. A prevalent privacy model in recent years is the Differential Privacy model [4], which provides strong privacy guarantees. This method is based on adding a special type of noise to the attribute values so that the presence or absence of any subject in a data set does not significantly influence the exploration results.

Cryptographic techniques towards privacy-preserving solutions proposed in the state of the art provide *searching on encrypted data* and *computing on outsourced data*. Searching on encrypted data is possible by employing Searchable Encryption (SE) schemes [5]. SE encrypts the data before outsourcing to allow secure search over the outsourced data. However, SE does not support computation on encrypted data, and having SE with strong security is inefficient. Among cryptographic techniques, homomorphic encryption (HE) and secure multi-party computation (MPC) are well-known approaches for distributed domains that can provide secure computation on outsourced data. A fully homomorphic encryption (FHE) scheme is a category of HE that allows the computation of any Boolean circuit on the encrypted inputs [6], [7], [8], [9]. As a result, FHE is able to be employed for AI algorithms [10]. Another cryptographic technique is secure MPC, in which parties jointly compute a function over their private inputs without disclosing their inputs. Compared to the HE and FHE approach, MPC requires several rounds of communication between all parties to compute the final result. As mentioned, non-cryptographic solutions are more efficient than cryptographic solutions, but they do not offer the same level of formal security.

2.1.5.2 Progress beyond the-State-of-the-Art

To summarize, the main shortcomings of the state-of-the-art are:

1. There is a trade-off between security and efficiency when proposing a privacy-preserving solution for specific use cases taking into account the available computational resources. Mobile sensors, drones typically, have limited computational power. On the other hand, many UAV applications, such as surveillance or emergency response require real-time decision-making based on the collected data. Sensors on the other side are very battery dependent.
2. The diversity of the data (because of the heterogenous multi-stakeholder environment) in all SINTRA use cases makes the design challenging.

SINTRA will innovate and improve the current state-of-the-art in the following directions:

1. Investigate on possibilities for utilizing both non-cryptographic and cryptographic schemes to propose distributed privacy preservation solutions that are efficient enough to apply to edge computing applications, particularly in resource-constrained devices and environments, including power consumption, low-latency, and high-throughput.
2. Investigate robust anonymization techniques, privacy-enhancing technologies, secure data sharing protocols, and contextual data handling methods to tackle privacy-preserving challenges of SINTRA that has multi-modal data (e.g., static cameras, live video from drones, body cams, mobile asset tracking).

2.1.5.3 Visual-based Privacy Protection

Visual-based privacy protection either protects the full video area or specific region-of-interest (RoI). In addition, privacy protection could be against a computer vision (CV) adversary, a human vision (HV) adversary, or both. Besides the type of adversaries, there are three types of privacy sensitive data that need to be protected: biometric identity (e.g., face identity), attributes (e.g., age, expression, race, and gender), and physiological signals of the person (e.g., heart rate, respiratory rate). From an adversary perspective, a face will provide a person's identity information; attributes will provide a way to spy on the person more accurately; and finally physiological signals will provide a way of gaining advantages in negotiation and analysing the health status of the person. Despite the importance of visual privacy in several domains, such as camera surveillance videos, there is a trade-off between data privacy and usability due to the complexity and diversity of the video content. There are two types of entities (machines/humans with and without consent) that wish to access personal privacy data. The entities without the necessary consent, CV or HV adversaries, should not have access to any privacy sensitive information. On the other hand, entities with consent can access all or part of privacy-sensitive data (e.g., biometric identity, attributes, or physiological signals) depending on their security level in the system. However, the aggregation of accessed information may inadvertently expose sensitive data beyond the intended scope. This risk persists even with the introduction of new data streams in the future.

There are several state-of-the-art approaches proposing visual privacy [11], [12], [13], [14], [15], [16], which are based on three main approaches: (i) *synthesis* (i.e., replacing privacy-sensitive content), (ii) *filtering* (e.g., blurring and pixelation), and (iii) *encryption*. In [11], a deep forgery technique based on Generative Adversarial Networks (GAN) is proposed which is lightweight and robust to different facial expressions. In [12], the authors used masking filters for surveillance videos captured by UAVs to protect privacy regions. The authors of [13] proposed a solution for the human face area that relies on encryption, which allows surveillance of a general nature while improving privacy issues, and full access only with the use of a decryption key, maintained by a court or other third party in the event of an accident. In [14], the authors proposed a solution to protect RoI through video encryption where protection depends on an encryption key. The process proposed in [14] is fully reversible for authorized entities. A new security scheme, Securecam, has been proposed by the authors in [15] for the protection of privacy in video surveillance systems in which the sensitive content (i.e., ROI) is encrypted using the lightweight Chacha 20 cipher. In [16], the authors developed a privacy-preserving approach that encrypts both sensitive and non-sensitive parts of the video. In this solution, the authors pursue the approach of compressive sensing (CS)-encryption to accomplish both compression and cryptographic security on the whole data, and data hiding technology. However, when handling large datasets, the secret matrix of a CS-based cryptosystem may cause substantial data storage and involve high computational complexity.

2.1.5.4 Progress beyond the-State-of-the-Art

To summarize, the main shortcomings of the state-of-the-art are:

1. Lack of an efficient solution to encrypt and hide sensitive data in the video especially in the case of drone cameras with limited battery life streaming live images to a backend. In message hiding techniques proposed for images or videos, such as steganography, a high PSNR (Peak Signal-to-Noise Ratio) value is desirable, indicating good quality in the reconstructed image/video. The quality of the image/video is closely related to the hiding capacity of the steganography algorithm, which is calculated using the Bits Per Pixel (BPP). BPP represents the number of bits that are hidden in every pixel of the image to produce the stego image. There exists a trade-off between PSNR and BPP, as increasing BPP for a higher hiding capacity can potentially reduce PSNR, affecting the visual quality of the stego image/video [17]. For instance, the method proposed in [18] can achieve stego-image quality exceeding 42 dB with a payload of 3 BPP, however, when the BPP is increased to 4 in this method, the PSNR decreases to 36, illustrating the impact of higher hiding capacity on the visual quality.
2. Need for both authenticity and encryption of the data.
3. Lack of long-term track, trace and audit of accessed data and insights as aggregation of new data streams can still expose sensitive information that was not expected.

SINTRA will innovate and improve the current state-of-the-art in the following directions:

1. Investigate lightweight data hiding techniques to embed and protected encrypted sensitive data within visual context using efficient synthesis or filtering techniques (visual privacy protection). The scheme can simultaneously protect the video against CV and HV, allowing entities with different levels of security to access privacy-sensitive data (multi-level privacy protection scheme).

2. Investigate multi-level security and lightweight encryption, as well as access control solutions to design an efficient and secure multi-level privacy protection scheme. Since the integrity and authenticity of the data is vitally important in the SINTRA use cases, the encryption algorithm should provide data authenticity as well as confidentiality. Thus, we will also investigate the use of lightweight authenticated encryption (AE) algorithms such as ASCON [19] that can provide both authenticity and confidentiality. ASCON-128a is an excellent choice for SINTRA, especially for resource-constrained devices that require lightweight yet secure cryptographic algorithms. What sets ASCON-128a apart is its consistent use of 128-bit security across key, nonce, tag, and data block parameters. This adherence to 128-bit security aligns with the BSI - Technical Guideline[243], which stipulates that cryptographic applications should utilize block ciphers with a block size of at least 128 bits. This guideline emphasizes the importance of maintaining a uniform security level of ≥ 128 bits for all system components, even exceeding the minimum requirements. For SINTRA's specific needs, ASCON-128a's combination of lightweight design, 128-bit security across all parameters, and alignment with established security standards makes it an ideal choice to ensure both the safety and efficiency of resource-constrained devices in SINTRA.

2.1.6 Secure and Trusted Data Transmission & Exchange

Based on the SINTRA requirements and use cases, secure and trusted data exchange aspects focus on “Zero Trust data exchange and data governance”, “BLE secure data transmission” and “UAV secure data transmission and logging system”.

2.1.6.1 Zero Trust Data Exchange and Data Governance

Traditional data exchange models between multiple stakeholders are based on *perimeter-based security* models where external access requests are protected with a firewall, an intrusion prevention system (IPS), etc. In these models, data consumers located on external networks must be authenticated and authorized before being trusted. However, data consumers located in the internal network are considered trusted by default and can access internal enterprise resources. As a result, an adversary can access resources without any restrictions just by compromising an internal consumer. Additionally, once external data consumers are authenticated, they are trusted for a long time and can access internal enterprise resources if authorized. Furthermore, an adversary who has gained access to the internal network is able to move laterally throughout the network and compromise other critical hosts, and servers, and gain data (VMware Report 2022: Lateral movement was seen in 25% of all attacks[244]).

[243] https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TG02102/BSI-TR-02102-1.pdf?__blob=publicationFile

[244] https://news.vmware.com/releases/vmware-report-warns-of-deepfake-attacks-and-cyber-extortion

Conversely, *Zero Trust* (ZT) is a significant shift in modern cybersecurity based on the concept of “never trusting and always verifying”. In this way, ZT is a data-centric approach that secures stakeholders’ data by removing implicit trust in perimeter-based tools and applying the same security checking and access control to internal and external users. ZT also limits internal lateral movement by controlling the data bridge and validating every access per session.

In contrast, stakeholders will not be able to adequately prioritize the controls needed to protect critical assets if they do not have a good understanding of data and potential threats. Thus, data discovery, governance, and classification as well as risk assessment of data assets are critical in such a data-centric trust approach. In [20], the authors introduced the existing *Zero Trust architecture* (ZTA) and stated that the important challenge in proposing ZTA for a system is how to apply it to the real enterprise network environment. Moreover, they stated that to design a ZTA, identity authentication, access control, and trust evaluation algorithms should be well thought of. As a result, the access control model used in ZT should employ not only static policies but also dynamic policies considering users, accessed resources and environmental attributes such as the location where the access request originated. Additionally, it is essential to match the risk with the level of trust assigned to a particular consumer that wants to access specific resources. Thus, the design of the access control model for data consumers that is appropriate for ZTA is required in SINTRA. More precisely, the SINTRA ZTA needs to monitor different consumers trying to access the data assets and ensures that each consumer is authorized. The ZTA proposed by NIST in 2020 [21] has been primarily applied to enterprise systems (most commercial offerings). However, ZT's deployment in SINTRA domains, such as construction sites, is not addressed or is still at an early stage with many challenges regarding its principles, architecture, and implementation remaining. A case in point is [3], where the authors mentioned that ZT is more than just a security product that can be placed in the infrastructure and confirmed to be secure. Rather, it is more of a concept that covers all aspects of security and trustworthy solutions with a variety of security products that are employed in accordance with the system environment that is in use. In [22], the authors fully explored the Zero Trust model for the Smart Manufacturing Industry and explained its principles, architecture, and implementation procedure.

Specifically, in SINTRA using either local servers or cloud systems to store information, data consumers can access this data with some static and dynamic access control rules and perform specific actions. A well-known access control model widely used recently is *Attribute-Based Access Control* (ABAC) which is an excellent alternative solution to Role-Based Access Control (RBAC). With ABAC, access is granted to a consumer if and only if it meets the model's attributes. There are several efforts to employ ABAC in ZTA [23]. However, the major problem with employing ABAC is that all elements need to be described in the form of attributes [24]. In addition to what is mentioned above, the access control model used for ZT must consider the consumer's or network's recent history into account during access requests evaluation. Given the fact that the

ZT system requires real-time decision-making capabilities, a *risk-aware access control model* can be used to solve grant access problems based on action risk levels [25]. Many factors can be considered in calculating risk. These include contextual and environmental factors as well as the trustworthiness of a consumer who requests the data. Generally, the risk access control model consists of “risk factors”, “risk estimation”, and “access control module” elements [26].

In addition, when designing ZTA, it is crucial to ensure that it is easily maintainable and scalable, with adaptable building components (e.g., adaptable access control mechanism). These should take into account the unique requirements and specificities of each company's platform, as they may vary across different application domains and services beyond the scope of the SINTRA objectives.

2.1.6.2 Data Spaces

Data Spaces can be described as an emerging decentralized network, wherein multiple connected data sources collectively provide a wide range of valuable services and resources over a network infrastructure. It can create the conditions for a competitive marketplace among participants or a collaborative environment among diverse, interconnected participants who are dependent on each other for their mutual benefit for as long as they are interconnected. However, it is necessary to conduct a substantial amount of research and design before integrating multiple data sources together. Data Spaces mainly should provide:

- *Data Interoperability* where all participants can effectively exchange data.
- *Sovereignty and Trust* in which parties accessing data can be verified and access control policies enforced.
- *Governance* that adopts agreements for business, operational, and organizational aspects.

The envisioned approach in SINTRA can support the trusted and secure sharing and trading of commercial data assets.

Industrial Data Spaces such as various ongoing initiatives are exploring information sharing solutions, including GAIA-X [27], IDS [28] [29], MyData [30], and iSHARE [31], to address challenges in this domain. GAIA-X tackles the problem through decentralized identifiers and data services, empowering companies to control data storage, processing, and access based on the concept of self-sovereign identity (SSI) [32]. MyData also utilizes SSI with operator support, separating authorization from data flows. The adoption of SSI promotes user control, facilitates compliance with GDPR regulations, and enables selective data disclosure. The International Data Spaces Association (DSA) reference architecture involves IDS connectors to integrate diverse partner data sources. iSHARE acts as a trust framework for cross-sector data spaces, collaborating with GAIA-X, and IDSA. The envisioned approach in SINTRA can support the trusted and secure sharing and trading of commercial data assets.

2.1.6.3 Progress beyond the-State-of-the-Art

To summarize, the main shortcomings of the state-of-the-art are:

1. Evaluation of access requests based on the recent history of data consumers and past activities and events that have occurred within the network in ZT deployment has always been challenging. This requires the utilization of techniques like risk-aware access control to dynamically assess trustworthiness and risk levels based on factors such as user activity, device, behaviour, network traffic, and security events.
2. Developing a secure and trustworthy data exchange mechanism is a complex process that faces many technical, organizational, legal, and commercial challenges. Key considerations include identifying best practices for creating privacy-aware solutions for sharing sensitive personal and industrial data, enabling enterprises to access shared data and supporting technologies, managing the trade-off between privacy and data analysis, and addressing standardization challenges for data sharing, including interoperability.

SINTRA will innovate and improve the current state-of-the-art in the following directions:

1. By integrating a risk-aware access control model with an ABAC model where the risk factor is treated as an attribute, it is possible to propose a ZTA that can be used as a practical trust model for SINTRA.
2. The proposed ZTA not only addresses the challenges of developing a secure and trustworthy data exchange mechanism but also offers a comprehensive solution to support trust in data sharing. By implementing granular and dynamic access control scheme (e.g., ABAC), robust authentication and encryption techniques, and privacy-aware practices, organizations can establish trust between data providers and recipients while maintaining privacy and compliance, thereby fostering confidence in data sharing initiatives.

2.1.6.4 BLE Secure Data Transmission

The study conducted in [33] examines the security and privacy properties of different Bluetooth Low Energy (BLE) versions. Upon investigation of BLE versions 5.0, 5.1, and 5.2, it is evident that they remain vulnerable to pairing method confusion, passkey reuse, and BlueMirror attacks. Furthermore, devices already paired using these BLE versions (i.e., 5.0, 5.1, and 5.2) are also susceptible to BLURtooth CTKD key overwrite, BLESA spoofing, and SCO mode downgrade attacks. Numerous open-source projects are actively monitoring BLE connections, including the Ubertooth project, which initially served as a Bluetooth classic test tool[245]. In [34], the author expanded upon the Ubertooth project to create a BLE sniffer capable of monitoring both existing and upcoming BLE connections. Another open-source project named BtleJack was introduced in [35], offering the capability to sniff ongoing connections. While sniffing entails eavesdropping packets, a Man-in-The-Middle (MITM) attack involves an attacker intercepting, manipulating, or dropping the connection. In this scenario, all messages pass through malicious devices, which then relay the data to legitimate recipients. Two open-source tools capable of implementing MITM attacks on BLE are GATTacker [36] and BTLEjuice [37]. GATTacker was presented at Black Hat USA[246], while BTLEjuice was published by the authors at the DefCon 24 conference[247].

Designing a secure BLE device or analyzing its security is a complex task due to the extensive range of possible configurations. The BLE specification consistently introduces new features and subtle changes related to privacy and security, further complicating this task. In [38], the authors propose and implement a lightweight digital certificate-based authentication mechanism for BLE devices that utilizes the Just Works model. The proposed model can be easily integrated into the existing BLE stack as an extension to the pairing mechanism. To mitigate the risks associated with MITM attacks and device spoofing in the Just Works pairing scenario, their model leverages the Public Key Infrastructure (PKI). This enables the establishment of peer entity authentication and ensures a secure cryptographic tunnel for communication purposes.

2.1.6.5 Progress beyond the-State-of-the-Art

To summarize, the main shortcomings of the state-of-the-art are:

1. BLE devices often have limited computational power, memory, and energy resources. This makes it challenging to implement strong security mechanisms without compromising device performance or battery life.
2. Although, I/O capabilities mechanisms are a solution for pairing, BLE devices without keyboard or display mechanism used in SINTRA (and hence using the Just Works pairing) are still vulnerable.

SINTRA will innovate and improve the current state-of-the-art in the following directions:

1. Investigate secure mutual authentication and key agreement schemes to design an authentication mechanism for BLE devices. The proposed model is an add-on to the already existing pairing mechanism and therefore can be easily incorporated in the existing BLE stack. To counter the existing Man-in-The-Middle attack scenario in Just

[245] https://ubertooth.sourceforge.net

[246] https://infocondb.org/con/black-hat/black-hat-usa-2016/

[247] https://defcon.org/html/defcon-24/dc-24-venue.html

Works pairing (device spoofing), the proposed model allows the scanner and tag to establish peer entity authentication and a secure cryptographic tunnel for communication.

2. Furthermore, investigation into how we can adopt energy-efficient cryptographic algorithms and protocols with the proposed solution can reduce the computational overhead and energy consumption associated with encryption and decryption operations. A lightweight mutual authentication and key agreement protocol should be explored, aiming to provide data integrity and confidentiality with a minimum 128-bit security. The solution should be meticulously designed, ensuring cryptographic algorithms, keys, and encryption methods meet or exceed the minimum recommended bit length requirements to guarantee end-to-end security on every communication link.

2.1.6.6 UAV Secure Data Transmission and Logging System

Various advancements are aimed at addressing the unique challenges associated with ensuring the security and integrity of data transmission and logging in the UAV ecosystem. In [39] survey, the authors present a systematic division of privacy and security issues by categorizing them into software, hardware, and communication classes. Secure communication protocols and robust encryption techniques have been the focus of most of the research. Encryption algorithms, such as Advanced Encryption Standard (AES), are widely adopted to protect data confidentiality during transmission and storage [40], [41]. Additionally, secure communication protocols, including Transport Layer Security (TLS), have been implemented to establish secure connections between UAVs and ground control systems, mitigating the risk of unauthorized access or data interception [42]. While AES and TLS are widely used for encryption and secure communication protocols, they introduce latency issues that can be problematic in the UAV ecosystem.

Authentication and secure onboarding of drones are critical challenges that must be addressed effectively in the UAV ecosystem, particularly for 5G flying drones, which require seamless re-authentication and secure communication protocols during dynamic flight scenarios and handovers (transferring the connection and communication from one base station to another as the drone moves within the coverage area). These challenges encompass various aspects, including mitigating potential vulnerabilities in authentication mechanisms to prevent unauthorized access and tampering, ensuring the secure storage and transmission of sensitive data during the onboarding process, and establishing standardized practices for secure authentication and onboarding across different drone platforms.

Moreover, tamper-resistant data logging mechanisms, such as cryptographic digital signatures and blockchain-based approaches, are being explored to ensure the integrity and immutability of logged data [43], [44]. Employing a blockchain that relies on a distributed consensus mechanism also may introduce a latency issue in the UAV ecosystem. In [45] the authors proposed a solution (called DASLog) to overcome the latency issue of the UAV logging system. The solution is based on generating verifiable proofs for the log records based on efficient Merkle tree and hash chains. However, this solution also relies on blockchain.

2.1.6.7 Progress beyond the-State-of-the-Art

To summarize, the main shortcomings of the state-of-the-art are:

1. Existing solutions for UAV secure communications often struggle with the trade-off between security and efficiency, as implementing strong encryption and authentication mechanisms can introduce significant overhead and latency, impacting real-time communication requirements for UAVs.
2. In scenarios with many UAVs operating at the same time, scalability remains a challenge. Quite often, the overhead of secure channels and managing cryptographic keys can become burdensome.
3. The dynamic nature of UAV networks such as 5G, with UAVs entering and leaving the network frequently, poses challenges for maintaining secure communication sessions and ensuring continuous authentication (i.e., re-authentication).
4. The process of securely onboarding UAVs into the network, including authentication, and establishing trusted connections, poses challenges in terms of mitigating potential vulnerabilities and ensuring secure integration.
5. Ensuring non-repudiation using an efficient and secure logging system, which prevents UAVs, operators, or users from denying their actions, is a challenge in UAV communication systems. This requires robust mechanisms to provide evidence of communication and transaction integrity.

SINTRA will innovate and improve the current state-of-the-art in the following directions:

1. Investigating lightweight cryptographic algorithms specifically designed for resource-constrained UAVs can help strike a better balance between security and efficiency. These algorithms should also provide both integrity (including message authenticity) and confidentiality since the transferred messages are the combination of command-and-control data and sensitive environmental information.
2. Investigating how to design scalable key management mechanisms that can handle a large number of UAVs while minimizing the overhead of key establishment and distribution is essential for secure and efficient communication.
3. Investigating how to develop dynamic authentication and access control mechanisms that adapt to the dynamic nature of UAV networks such as 5G. This will allow UAVs to join and leave the network (needs re-authentication) seamlessly while maintaining secure communication sessions.
4. Investigating how to establish robust secure onboarding procedures that ensure UAV integrity and authenticity during the onboarding process. This includes secure device registration, identity verification, and trusted connection establishment.

5. Investigation of non-repudiation mechanisms, such as digital signatures or trusted and secure logging systems, to provide evidence of communication and transaction integrity, preventing UAVs or users from denying their actions. The solution will be based on efficient proof generation based on efficient algorithms such as Merkle tree and hash chains.

2.1.7 Multimodal Data Integration

Multi-modal data integration is the process of combining data from disparate sources, possibly consisting of different types and with different volumes with the goal of providing users with a single, unified view. It is pivotal to improve the performance and robustness of predictive and analysis models. Information fusion from different sources is also quite crucial for understanding the current context and further historical and future evolution analysis. Both classical and recent deep learning-based techniques were used in the past extensively to tackle various challenges [46]. However mere data type heterogeneity (e.g., images, text, audio) itself makes the feature extraction, fusion and analysis challenging. In addition, state-of-the-art research also finds it difficult to tackle the following challenges.

- Spatial and temporal alignment: Accurately aligning data from different sensors, both in space and time, is crucial for effective fusion. This often involves correcting for sensor calibration errors, synchronizing timestamps, and handling different sensor sampling rates.
- Occlusions and missing data: Sensors may have different fields of view and certain objects may be occluded in one sensor but visible in another. Handling occlusions and missing data during fusion is a complex problem to solve.
- Scalability: Scalability is a significant challenge in multi-data fusion models because as the number of data sources and the volume of data increase, the complexity of the fusion process exponentially increases, demanding highly efficient algorithms and computational resources. Traditional data fusion models may suffer from computational inefficiency due to the requirement for pairwise comparisons between data items, which often leads to a quadratic increase in computational complexity as data size grows. The situation is also similar with state-of-the-art deep learning models. For instance, fully connected deep neural networks such as Deep Belief Networks (DBNs) and Stacked Auto-Encoders (SAEs) involve a great number of connections between neurons that require extensive training objects, making it computationally intensive. These networks often struggle with high-dimensional data, particularly large image, and audio data, affecting their scalability.
- Computational complexity: From a computational complexity perspective, dealing with the uncertainty in the fusion of cross-modal observations presents another challenge in multi-data fusion models. Traditional robust particle filtering (RPF) methods focus on managing uncertainties in the state-transition function or modelling the observation noise, but the introduction of cross-modal data adds another layer of complexity. This

complexity arises from the need to find an "ideal way" to fuse cross-modal observations that aligns with the true data generating mechanism and leads to the most accurate state estimations. This requires new methodologies that can handle cross-modal data fusion in nonlinear non-Gaussian dynamic systems, as existing trust models often focus on co-modal data or fusion of information from multiple sources in a static setting [47].

Based on the stage of data fusion, multimodal data fusion techniques can be mainly categorised into [48]:

1. *Early/data level fusion* combines raw data from multiple modalities at an early stage, leveraging inherent correlations to generate a comprehensive feature representation. This approach enables models to exploit rich intermodal relationships and potentially uncover latent patterns but may suffer from increased computational complexity and challenges in handling heterogeneous data types.
2. *Late/decision level fusion*, on the other hand, focuses on aggregating predictions from separate models trained on individual modalities, effectively leveraging the strengths of each modality while mitigating risks of information loss or distortion from early fusion. However, it may overlook important intermodal relationships that could have been captured through joint processing.
3. *Intermediate fusion* methods strike a balance between these two extremes by merging information at a higher abstraction level, such as fusing feature representations or combining model predictions with context-aware weighting. This enables intermediate fusion methods to harness the benefits of both early and late fusion while mitigating their respective drawbacks, providing a versatile framework for multi-modal data fusion in various machine learning applications [49].

Data fusion at all the stages will be explored based on the use case requirements and amount of reasoning and explainability required for the use cases. Further, the listed shortcomings will be addressed as a part of SINTRA keeping the security and privacy aspects of data as a priority.

2.1.7.1 Progress beyond the-State-of-the-Art

To summarize, the main shortcomings of the state-of-the-art are:

1. Algorithmic inefficiency in identifying and incorporating multi-modal contextual data sources in a scalable manner.
2. Shortage of effective heterogeneous data integration methods with data generalization across different locations with minimal additional data requirements.
3. Context extrapolation and prediction based on incomplete and partially available data.
4. Lack of interpretable and explainable models for data aggregation and context-generation: State-of-the-art solutions make use of black box deep learning models with low interpretability.

SINTRA will innovate and improve the current state-of-the-art in the following directions.

1. Research possibilities to incorporate multi-modal contextual data sources with different level of (temporal and spatial) granularities and inaccuracies.
   Goal: enable multi-source data integration, aggregation and inference at the edge addressing scalability challenges.

2. Innovate on long and short-term multi-modal prediction models for rare but possibly re-occurring events to inform within specified time windows with given accuracies.
   Goal: Ensure pre-defined accuracy levels for context evolution predictions at different time scales by considering contextual historic as well as real-time data.

3. Explore data extrapolation approaches based on (partially missing) multi-source data streams.
   Goal: generate realistic monitoring site evolution statistics which can improve the resource planning and result in better safety measures.

4. Investigate approaches for improving interpretability of model context analysis and recommendations.
   Goal: develop robust indicators such as alarm root cause (categorical variable), predicted resource requirements (continuous variable) that provide insights to context and requirement analysis, applied to site monitoring use cases.

2.1.8 Multi-View Clustering for Context Analysis

Once proper data integration techniques are in place, the next step is to extract context from these multiple sources. *Multi-view clustering* is an essential tool to address the increasing complexity of data collected from various sources and perspectives. It leverages the complementary and supplementary information inherent in these diverse views to yield more accurate and robust cluster results than traditional single-view methods. By harnessing the power of multiple perspectives, multi-view clustering unveils the multi-faceted nature of complex data, allowing for a richer understanding of underlying patterns and structures.

Multi-view analysis is hot research problem and recently a few papers have addressed some of the main challenges faced in analysing multiple views. For example, in [50] the authors propose a joint contrastive triple-learning framework that seeks to improve multi-view representation for deep clustering. This innovative approach is tripartite, involving feature-level alignment-oriented and commonality-oriented contrastive learning (CL), and cluster-level consistency-oriented CL. The authors introduce a deep learning-based framework designed to mitigate performance degradation caused by view increase in multi-view clustering in [51]. The model is trained to concurrently extract complementary information and disregard meaningless noise through automatic feature selection. There are other papers in literature which makes use of advanced deep learning based autoencoders. For instance, in [52] the authors present a deep multi-view clustering algorithm known as MVC-MAE, which is based on multiple auto-encoders. It combines representation learning and clustering into a unified framework, enabling these two tasks to be jointly optimized.

2.1.8.1 Progress beyond the-State-of-the-Art

To summarize, the main shortcomings of the state-of-the-art are:

1. Incorporating prior information: Utilizing prior information such as pairwise constraints that describe the relationship between data instances can be beneficial for multi-view clustering. None of the existing methods fully leverage prior information for error-robust multi-view clustering in a joint manner.
2. Dealing with incomplete views: Data instances within views may be missing due to the nature of the monitored data or data collection costs. There is a need for a comprehensive clustering algorithm for multiple views where views may be both incomplete and erroneous. Existing solutions for partial multi-view clustering often neglect possible errors in views, and there is a need for better methods that jointly handle view completion and clustering.
3. User-relevant context generation: Extracting user-relevant context from these multiple views is quite challenging mainly due to the variations in user preferences. The state-of-the-art papers do not address the user-centric nature of multi-view analysis which is important in SINTRA use cases because of the context importance requirements from different partners.

SINTRA will innovate and improve the current state-of-the-art in the following directions.

1. Research possibilities to incorporate prior information available from the monitoring sites to improve clustering and context generation.
   Goal: enable privacy-aware static prior information integration to the context generation models to improve the relevance of context extraction.

2. Explore and innovate in developing techniques for dealing with missing views by incorporating information from other available sources (both historical and real-time)
   Goal: Get a complete working picture of the monitored site even with interruptions in views over time and space.
3. Incorporate user preferences into the current context generation algorithms without affecting the performance of the models.
   Goal: Extract user-relevant contexts based on the query inputs as different stakeholders are looking into different contexts based on their area of interest.

2.1.9 Multi-Sensor Anomaly Detection and Proactive Resource Planning

Timely detection of system failures, security breaches, and performance issues of activities concerning site management is inevitable to make automated site monitoring a reality. Multi-sensor powered anomaly detection harnesses the power of multiple sensors to monitor a system or environment, collecting diverse data types to detect unusual events or changes to enable this. Concurrently, in resource planning, multi-sensor data is invaluable in predicting resource demand and optimizing allocation. It allows for proactive management, facilitating the prediction of future needs based on past and real-time data, and tailoring resource distribution to match the forecasted requirements. In essence, multi-sensor-powered anomaly detection and resource planning combine to provide a robust, responsive system that can identify potential issues before they escalate and dynamically manage resources for optimum efficiency and productivity.

Various state-of-the-art papers have investigated multi-sensor anomaly detection, mainly due to the increase in monitoring sensor deployments in different domains [53] [54] [55]. In [56] The authors fused multi-sensor signals to provide robust anomaly detection in the presence of sensor occlusion. Further, they developed a proactive anomaly detection network (PAAD) for enabling planned robot navigation in uncertain environments. In [57], the authors proposed a novel Deep Convolutional Autoencoding Memory network (CAE-M) to characterize spatial dependence of multi-sensor data. A deep convolutional autoencoder is used to capture spatial correlations and a bi-directional LSTM is used to model temporal dependencies. A semi-supervised anomaly detection module for wireless spectrum sensors was developed in [58] where the authors included techniques to incorporate user preferences into the entire anomaly detection process. In [59] a coupled attention-based neural network framework (CAN) for anomaly detection is proposed for multivariate time series data. The framework addresses the challenges of dependencies among sensors and variables that often change over time by generating a global-local graph that represents both global correlations and dynamic local correlations among sensors.

Video-based monitoring and anomaly detection solutions, while increasingly effective, face numerous challenges. For instance, these solutions often struggle in poor or varying lighting conditions. Algorithms can misinterpret shadows or reflections as anomalies or fail to detect anomalies in low light. In addition, image-based monitoring raises significant privacy concerns [60], particularly in public or semi-public spaces which makes the algorithms complex and real-time processing difficult [61]. Further, these algorithms fail to accurately monitor and detect anomalies in busy or complex environments, such as crowded public spaces due to the high level of detail and fast-changing conditions. These complex scenarios also make it difficult to have a deeper semantic understanding of the scene. In addition, as these systems become more common, they could also become targets for adversarial attacks, where the system is deliberately manipulated or fooled [62]. All these challenges mainly point to the use of a multi-sensor solution for enabling a robust context aware anomaly detection.

2.1.9.1 Progress beyond the-State-of-the-Art

To summarize, the main shortcomings of the state-of-the-art are:

1. Lack of robust context awareness anomaly detection techniques with multiple views: Existing solutions often struggle to extract context from situations with multiple operational views. Anomaly in one context might not be anomalous in another. Lack of deep scene semantic understanding in complex scenarios: Current state-of-the-art solutions typically use pre-defined rules or patterns for anomaly detection, which can fail in complex or dynamic situations where context and understanding of the scene are critical. This is especially true in situations involving crowded public spaces, dense traffic, or various weather conditions, where the semantics of the scene dramatically affect what is considered "normal" or "anomalous".
2. Interoperability issues: Currently, there are challenges in integrating and making sense of data from different types of sensors. Each sensor type may have its own data format, resolution, and accuracy, making seamless integration difficult.
3. Robustness to sensor failures: Present solutions often do not account for potential sensor failures or inaccuracies. Building robustness to sensor noise, malfunctions, and failures should be an integral part of future multi-sensor anomaly detection systems.
4. Inadequate privacy protection: With the increasing use of multi-sensor networks, particularly cameras, the issue of privacy becomes more critical. The state-of-the-art often lacks effective mechanisms to protect individuals' privacy while conducting anomaly detection while fusing multimodal sensor information.

SINTRA will innovate and improve the current state-of-the-art in the following directions.

1. Enhanced contextual understanding: SINTRA will aim to improve deep scene semantic understanding in complex scenarios, enabling accurate anomaly detection regardless of the situation's context. This will involve developing advanced machine learning models that can learn and adapt over time, adjusting to new normal patterns as the scene evolves. Goal: enable generalizable and easily deployable context understanding for evolving site monitoring use cases in SINTRA.
2. Robustness to sensor failures: We will work on developing models that are resilient to sensor noise, malfunctions, and failures. Strategies such as redundancy, self-diagnosis, and error correction will be explored to ensure consistent performance even in the event of individual sensor failures.

   Goal: Sensor failure is realistic in the monitoring site use cases mentioned in SINTRA. The goal is to improve the consistency of context understanding even with failing sensors.
3. Robustness against adversarial attacks: Recognizing the growing threat of adversarial attacks, the project will aim to develop defenses against these types of attacks. Research will be conducted on the detection and mitigation of adversarial inputs, ensuring that the system remains reliable and trustworthy even in the face of targeted attacks.

Goal: The algorithms in the use cases will be prone to adversarial attacks by the mere nature of the parties involved. The goal is to alert the stakeholders on time if the models are facing sensor data manipulation and other adversarial inputs.

2.2 State-of-the-Art of Business Operations

2.2.1 AI Market

2.2.1.1 Market and Business Implications of HAR Systems

The future of object detection coupled with Human Action Recognition (HAR) is a rapidly advancing field of research that promises to bring significant improvements to a variety of applications. The combination of object detection and HAR will enable us to create more intelligent and responsive systems that can better interpret and interact with the world around us.

It is possible that the integration of deep learning techniques with HAR will continue, resulting in more complex models that can perform complex object detection tasks with higher accuracy.

The multimodal approach, which is open to development, will provide a more comprehensive understanding of the scene and the actions taking place. Developing real-time processing capabilities for both object detection and HAR is of great importance for applications such as surveillance and interactive systems. The use of 3D sensors to support 2D images is one of the important aspects open to research. This includes improving depth estimation and temporal sequence analysis for a better understanding of the real world. This is an important area for development as future HAR systems will need to be resilient to environmental changes, including changing lighting conditions, weather and occlusions, to maintain high accuracy in object detection.

The ability to detect unusual or abnormal actions and objects will be an important aspect of future HAR systems, especially for security and safety applications.

Object detection within HAR will play a role in improving human-robot collaboration, allowing robots to better understand and respond to human actions and the surrounding environment.

2.2.1.2 Market and Business Implications of Object Detection

2.2.1.2.1 Application Areas

Combining pixel location with real-world location facilitates a wide range of applications spanning from entertainment and gaming to critical fields such as healthcare, transportation, and environmental monitoring.

AR applications overlay virtual objects onto the real world seen through a device's camera. Combining pixel locations with real-world coordinates enables accurate placement of virtual objects within the camera view, creating immersive AR experiences.

In robotics, mapping pixel locations to real-world coordinates is essential for tasks such as robot localization, object detection and manipulation, path planning, and navigation. Robots can use this information to perceive and interact with their environment effectively.

Autonomous vehicles rely on sensors, including cameras, to perceive their surroundings. By mapping pixel locations from camera images to real-world coordinates, autonomous vehicles can identify objects, lanes, traffic signs, and other important features on the road for safe navigation.

Geographic Information Systems (GIS): GIS applications use pixel locations from satellite or aerial imagery to map features such as land use, vegetation cover, urban development, and more onto real-world coordinates. This helps urban planners, environmental scientists, and government agencies in decision-making processes.

Medical Imaging: In medical imaging, combining pixel locations with real-world coordinates assists in tasks like image registration (aligning images from different modalities or time points), tumour localization, and surgical navigation, leading to more accurate diagnoses and treatments.

Photogrammetry: Photogrammetry techniques use pixel locations from overlapping images to reconstruct 3D models of objects or terrain in the real world. Accurate mapping between pixel locations and real-world coordinates is crucial for generating precise 3D reconstructions.

Surveillance and Security: Surveillance systems utilize pixel locations to track objects or individuals in real time. By associating pixel locations with real-world coordinates, security personnel can monitor and respond to events effectively.

Archaeology and Cultural Heritage: Archaeologists and cultural heritage specialists use pixel-to-real-world mapping to analyse and document artifacts, monuments, and archaeological sites captured in images. This aids in preservation efforts and historical research.

2.2.1.2.2 Future Potential of Object Detection

Object detection techniques can be explored more deeply based on application areas and resource requirements. Besides many advancements, still, object detection has future directions.

Detecting objects in videos presents greater challenges compared to still images due to the diverse appearance variations in video frames, including defocus, motion blur, truncation, occlusion, and fast motion. While extensive research has been conducted using video data, further enhancements are required in detection capabilities. Despite efforts to enhance accuracy, there remains a need for more effective and efficient feature extraction techniques and motion estimation networks[248]. Researchers worldwide should focus on addressing more dynamic targets and handling more complex data for future advancements in object detection within video streams.

3D sensors provide supplementary depth information to enhance the utilization of 2D images, bridging the gap between digital imagery and real-world understanding. Despite the rapid advancements in object detection, there are areas that warrant further analysis and exploration. Depth estimation, temporal sequence analysis, and generalization are among the pertinent domains in 3D object detection[249], serving as key directions for future research endeavours in this field.

Small object detection has been a challenge in a large or real-time environment. Some of the applications of OD, such as small vehicles from real-time CCTV cameras, detecting some important targets state of the military, ship detects from remote sensing images etc., are the research direction. Some of the other research directions may include the design of lightweight networks and visual attention mechanisms.

2.2.1.3 Advantages and Disadvantages of Edge Computing Based on IoT

There are several advantages about the edge computing based IoT that are listed below[250],[251]:

- Efficiency: an edge device takes full advantage of the available resources by allocating storage, computing, and control functions to available resources in any place between the end-user and cloud.
- Cognition: an edge device is conscious of customer requirements.
- Agility: it is quicker and inexpensive to experiment with edge devices and clients because data processing and storage are done close to the end user.
- Latency: edge computing supports time-critical applications by enabling data analysis and data processing near the end-user, which grants IoT applications the ability to make decisions faster and better.

There are several disadvantages about the edge computing based IoT that are listed below:

[248] Liu D, Cui Y, Chen Y, Zhang J, Fan B (2020, Elsevier B.V.) Video object detection for autonomous driving: motion-aid feature calibration. Neurocomputing 409:1–11.

[249] Qian R, Lai X, Li X (2021) 3D object detection for autonomous driving: A Survey 14(8), 1–24, [Online].

[250] Gezer, V., Um, J., & Ruskowski, M. (2017). An extensible edge computing architecture: Definition, requirements and enablers. Proceedings of the UBICOMM.

[251] T. Lin, B. Park, H. Bannazadeh, and A. Leon-Garcia, ‘‘Demo abstract: End-to-end orchestration across SDI smart edges,’’ in Proc. IEEE/ACM Symp. Edge Comput. (SEC), Oct. 2016, pp. 127–128.

- Heterogeneity: Heterogeneity in the IoT-based edge computing environment exists in computing and communication technologies. Computing platforms can have different operating systems and hardware architectures, whereas communication technologies can be heterogeneous regarding the data rate, transmission range, and bandwidth. One of the challenges in edge computing is to develop a solution in software space that is portable across different environments. This challenge is crucial because various applications are deployed in edge devices.
- Standard Protocols and Interfaces: Edge computing is an emerging technology in the IoT field. In this heterogeneous environment, different devices and sensors connect and communicate with one another and with the edge server via communication protocols. These devices have their own interfaces and thus demand specific communication protocols. Considering that different vendors manufacture different devices in the IoT environment, standard protocols and interfaces should be developed to enable communication among these heterogeneous devices. The development of standard protocols and interfaces in the IoT environment is challenging because of the rapid development of new devices.
- Availability: Availability in the IoT-based edge computing environment includes hardware-level and software-level provision of resources and services anywhere and anytime for subscribed IoT devices. Usually, availability comprises three factors, namely, mean time between failure, failure probability, and mean time to recovery. Ensuring the availability of resources and services for the growing number of IoT devices is a challenging research perspective. However, availability can be optimized by maximizing the mean time between failures and minimizing the failure probability and mean time to recovery.
- Data Abstraction: With IoT, several data-generating devices are connected in the network, and all of these data generators report tremendously large raw data to the edge device. For the edge device, analysing such big data is computationally difficult. Security risks are also involved. Therefore, the data should be pre-processed at the gateway level, such as noise/ low-quality removal, event detection, and privacy protection. The processed data will be sent to the upper layer for future service provision. However, many challenges may occur in this process. For privacy and security purposes, applications running on edge devices should be blind to these raw data. Therefore, the details of the data should be removed during data pre-processing. However, the usability of the data can be affected by hiding the details of sensed data. Defining the extent to which the raw data should be filtered out is also a challenge because several applications cannot obtain accurate results from such data.
- Security and Privacy: Edge computing acts as a boon to cybersecurity because data do not travel over a network. However, a highly dynamic environment at the edge of a network makes the network unprotected. Given that different devices are connected in IoT, a large array of potential security threats can be generated. Many applications are running at the

network edge, so the data provided to these applications should be in a hidden form. Otherwise, any intruder can use the open data for illegal purposes. For example, if a home is connected to IoT, then private data, such as individual health data, can be stolen. In this case, how to support the service without harming privacy is a challenge. Applications running on edge devices should be blind to the raw data. Personal data can be removed before reaching the edge device.

2.2.2 CCTV Market

2.2.2.1 Market Perspective and Implications

The market for advanced airport CCTV analytics has grown significantly due to increasing demands for enhanced security and operational efficiency at airports worldwide. This growth is driven by the rising number of air passengers, heightened security threats, and the need for airports to optimize operations such as passenger flow and baggage handling.

The current state of market analysis regarding Advanced Airport CCTV Analytics indicates a shift towards utilizing video surveillance beyond security purposes[252],[253] . This evolution involves integrating real-time data from ground radars with IP cameras to enhance aircraft management on airport aprons[254]. Innovative systems like ASEV are being developed to automatically assess airport surveillance situations, improving operator performance through real-time event assessment and privacy protection[255]. Airports are seen as ideal environments for developing digital marketplaces that offer context-aware services, such as personalized marketing campaigns and streamlined airport processes, leveraging CCTV analytics[256]. The focus is on combining security and operational needs, utilizing existing camera networks for both security and efficiency monitoring in critical infrastructure like airports.

The demand for advanced CCTV analytics in airports is driven by the critical need for enhanced security protocols, operational efficiencies, and improved passenger experiences. The aviation industry's pivot towards Total Airport Management underlines the necessity for integrated, intelligent systems capable of not only surveillance but also providing actionable insights across

[252] Arppitha, Krishna., Neha, Pendkar., Shruti, Kasar., Umesh, Mahind., Shridhar, Desai. (2021). Advanced Video Surveillance System. doi: 10.1109/ICSPC51351.2021.9451694

[253] Ezequiel, Roberto, Zorzal., Ariel, Fernandes., Bruno, Castro. (2017). Using Augmented Reality to overlapping information in live airport cameras. doi: 10.1109/SVR.2017.53

[254] Simon, Denman., Tristan, Kleinschmidt., David, Ryan., Paul, Barnes., Sridha, Sridharan., Clinton, Fookes. (2015). Automatic surveillance in transportation hubs. Expert Systems with Applications. doi: 10.1016/J.ESWA.2015.08.001

[255] Simon, Denman., Tristan, Kleinschmidt., David, Ryan., Paul, Barnes., Sridha, Sridharan., Clinton, Fookes. (2015). Automatic surveillance in transportation hubs. Expert Systems with Applications. doi: 10.1016/J.ESWA.2015.08.001

[256] Eli, Katsiri., George, Papastefanatos., Manolis, Terrovitis., Timos, Sellis. (2014). Airport Context Analytics. doi: 10.1007/978-3-319-11113-1_13

various operational facets. The global uptick in air travel, alongside heightened security concerns, underscores the urgency for such advanced analytics solutions.

Companies like Hikvision, Dahua, Axis Communications, and Bosch Security Systems are at the forefront of integrating smart technologies into CCTV systems. Innovations include AI-powered cameras, thermal imaging for health and safety applications, and integrated analytics platforms.

2.2.2.2 Regulations

The adoption and implementation of advanced CCTV analytics in airports are heavily influenced by regulatory standards and privacy laws, which vary significantly across regions. Key regulatory considerations include:

- Data Protection and Privacy Laws: In jurisdictions such as the European Union, stringent data protection laws (e.g., GDPR) impose strict guidelines on the collection, processing, and storage of personal data. These regulations mandate clear consent mechanisms, data minimization, and the implementation of substantial security measures to protect personal data.
- Aviation Security Regulations: International and national aviation authorities (e.g., ICAO, SHGM, EASA) set comprehensive security standards that include the use of surveillance technologies. Compliance with these standards is mandatory for airports and can drive the adoption of advanced CCTV analytics solutions.
- Emerging Technologies Regulation: As technologies like 3D sensing and computer vision evolve, regulatory bodies are increasingly focusing on setting guidelines that ensure their responsible and ethical use, especially concerning biometric data and surveillance.

2.2.2.3 Competition & Growth Opportunities

The market is populated with a mix of established technology vendors such as SITA, Amadeus, and Thales, and emerging companies focused on computer vision and 3D sensing technologies. While traditional vendors have a stronghold in airport operations technologies, the influx of companies specializing in advanced analytics and 3D sensing solutions introduces a new dimension to the competition, fostering innovation and potentially reshaping market dynamics.

- Beyond Security: Increased security threats in both public and private sectors drive the demand for smarter surveillance systems. Extending the application of CCTV analytics from security to operational efficiency and passenger experience is crucial.
- Technological Advancements: Innovations in AI, machine learning, and IoT (Internet of Things) are key enablers for upgrading old CCTV systems. Leveraging ongoing developments in 3D imaging and computer vision to offer more sophisticated, accurate analytics are considered.

- Metaverse Applications: Exploring the use of 3D sensing and CCTV analytics in creating virtualized environments and experiences, catering to the burgeoning interest in the metaverse is a point of view.
- Regulatory Compliance: Governments and regulatory bodies are mandating more stringent security measures, pushing for the adoption of advanced surveillance technologies.
- Computer Vision and 3D Sensing

Computer vision technologies are at the heart of advanced CCTV analytics, enabling sophisticated surveillance capabilities that include facial recognition, behaviour analysis, and anomaly detection. The advent of 3D sensing technology aims to mimic the human visual system, offering depth perception and enhancing the accuracy of analytics. These technologies find applications beyond security, aiding in crowd management, and facilitating immersive customer experiences.

- Integration with Other Systems

The vision for Total Airport Management emphasizes the integration of advanced CCTV analytics with other operational systems within airports. This integration promises a synergistic approach to airport management, where insights derived from CCTV analytics can inform decisions across logistics, retail, and customer service operations.

2.2.2.4 Market and Business Implications

The regulatory environment has profound implications for the market and businesses operating within it:

- Compliance Costs: Compliance with diverse and evolving regulatory standards can significantly increase operational costs for companies. This includes investments in technology that ensures data protection, privacy compliance, and security standards.
- Innovation vs. Regulation Balance: Businesses must navigate the fine line between innovation and compliance, ensuring that new solutions adhere to regulatory requirements while pushing the boundaries of what's technologically possible.
- Market Entry Barriers: Stringent regulations can act as barriers to entry for new players, particularly small and medium-sized enterprises (SMEs) that may lack the resources to meet compliance demands. This can affect the competitive landscape, potentially limiting the diversity of solutions available in the market.
- Global Market Dynamics: The regulatory landscape influences the global dynamics of the market. Companies may find it easier to deploy solutions in regions with less stringent regulations, leading to uneven market development and adoption rates worldwide.
- Trust and Adoption: Compliance with regulations, especially those related to privacy and data protection, can enhance trust among users and regulatory bodies, facilitating the adoption of advanced CCTV analytics in sensitive environments like airports.

- Collaboration Opportunities: The complexity of the regulatory environment may encourage collaborations between technology providers, regulatory consultants, and legal experts to develop solutions that not only push technological boundaries but also adhere to the highest standards of privacy and data protection.

2.2.3 Sensor Market

2.2.3.1 Applications of mmWave Radars and Machine Learning Techniques

MmWave radars are used in a variety of applications include automotive applications, industrial applications, military applications, medical applications, robotics and automation applications, civilian applications, and security and surveillance applications[257].

[257] https://www.mdpi.com/1424-8220/23/21/8901

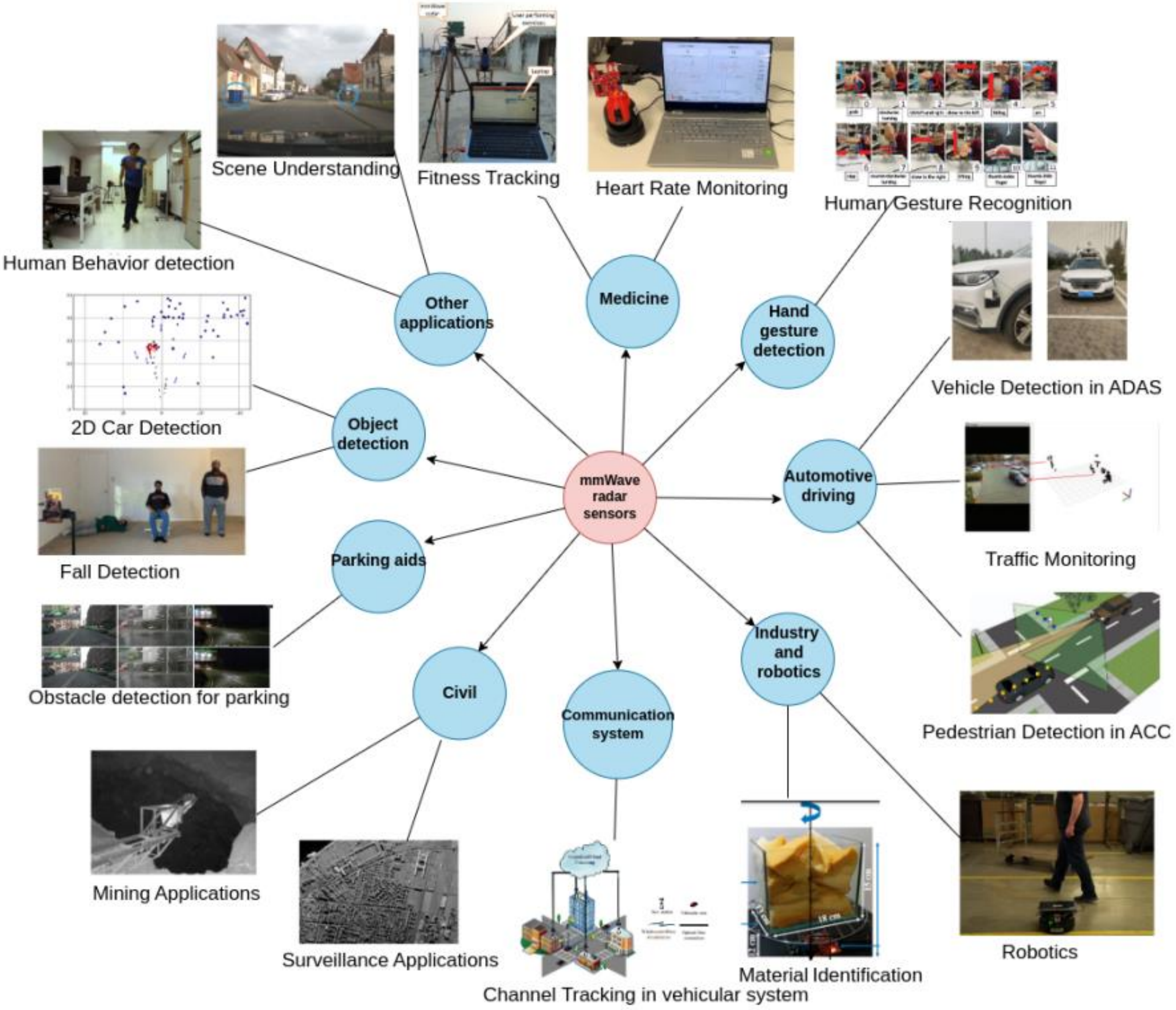
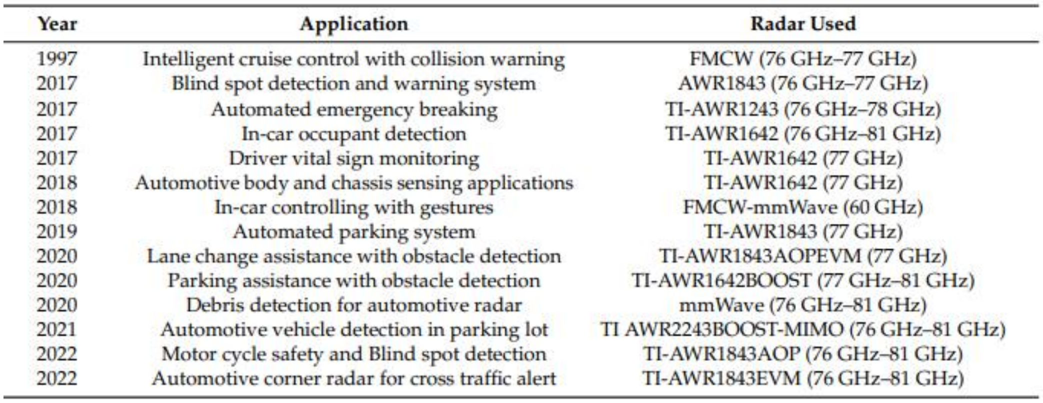


Figure 25. MmWave radar sensor applications cited in this article

- Automotive Applications

In automotive applications, there are numerous studies utilizing mmWave radar sensors to accurately determine the radial distance, velocity, and Angle of Arrival (AoA) of moving objects. These applications include adaptive cruise control, autonomous emergency braking, blind spot detection, lane change assistance, front and rear cross-traffic alert, automated parking, and in-cabin detection applications (such as gesture recognition and passenger location detection).

Table 2. Automotive applications table on this article

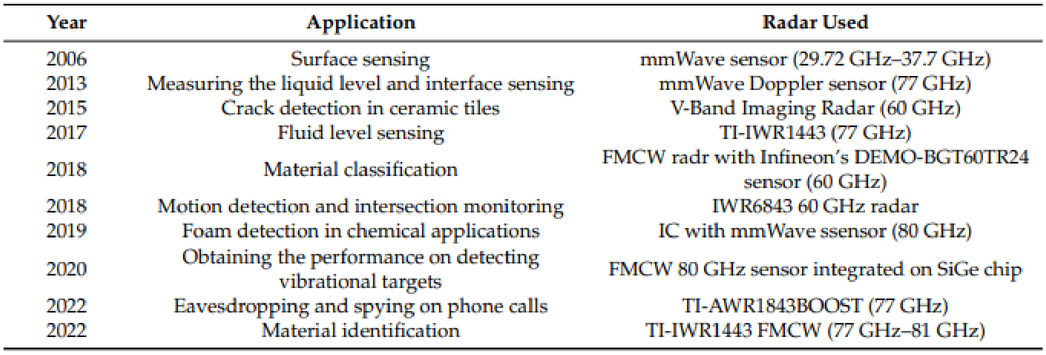


Year	Application	Radar Used
1997	Intelligent cruise control with collision warning	FMCW (76 GHz–77 GHz)
2017	Blind spot detection and warning system	AWR1843 (76 GHz–77 GHz)
2017	Automated emergency breaking	TI-AWR1243 (76 GHz–78 GHz)
2017	In-car occupant detection	TI-AWR1642 (76 GHz–81 GHz)
2017	Driver vital sign monitoring	TI-AWR1642 (77 GHz)
2018	Automotive body and chassis sensing applications	TI-AWR1642 (77 GHz)
2018	In-car controlling with gestures	FMCW-mmWave (60 GHz)
2019	Automated parking system	TI-AWR1843 (77 GHz)
2020	Lane change assistance with obstacle detection	TI-AWR1843AOPEVM (77 GHz)
2020	Parking assistance with obstacle detection	TI-AWR1642BOOST (77 GHz–81 GHz)
2020	Debris detection for automotive radar	mmWave (76 GHz–81 GHz)
2021	Automotive vehicle detection in parking lot	TI AWR2243BOOST-MIMO (76 GHz–81 GHz)
2022	Motor cycle safety and Blind spot detection	TI-AWR1843AOP (76 GHz–81 GHz)
2022	Automotive corner radar for cross traffic alert	TI-AWR1843EVM (76 GHz–81 GHz)

- Industrial Applications

Among the industrial applications utilizing mmWave radar sensors are level sensing of fluids, volume identification of solids, infrastructure systems, surface quality assessment in production industries, and vibration monitoring.

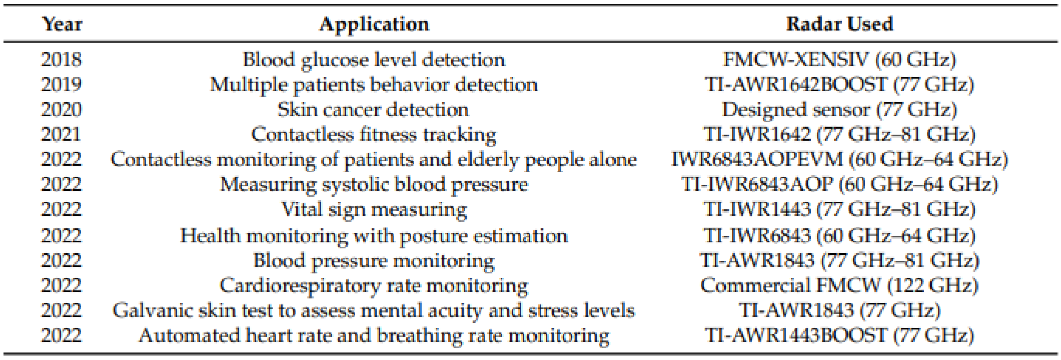
Table 3. Industrial applications table on this article



Year	Application	Radar Used
2006	Surface sensing	mmWave sensor (29.72 GHz–37.7 GHz)
2013	Measuring the liquid level and interface sensing	mmWave Doppler sensor (77 GHz)
2015	Crack detection in ceramic tiles	V-Band Imaging Radar (60 GHz)
2017	Fluid level sensing	TI-IWR1443 (77 GHz)
2018	Material classification	FMCW radr with Infineon's DEMO-BGT60TR24 sensor (60 GHz)
2018	Motion detection and intersection monitoring	IWR6843 60 GHz radar
2019	Foam detection in chemical applications	IC with mmWave ssensor (80 GHz)
2020	Obtaining the performance on detecting vibrational targets	FMCW 80 GHz sensor integrated on SiGe chip
2022	Eavesdropping and spying on phone calls	TI-AWR1843BOOST (77 GHz)
2022	Material identification	TI-IWR1443 FMCW (77 GHz–81 GHz)

- Medical Applications

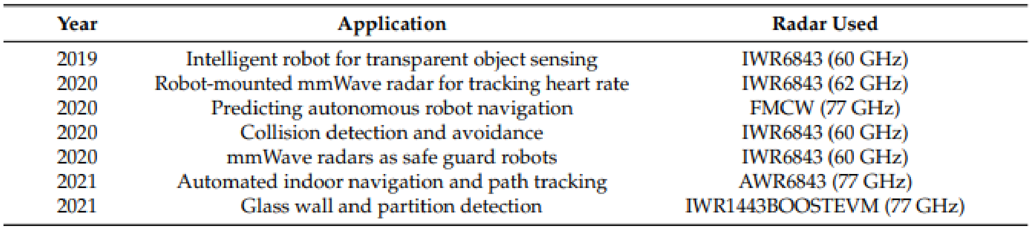
Table 4. Medical applications table on this article



Year	Application	Radar Used
2018	Blood glucose level detection	FMCW-XENSIV (60 GHz)
2019	Multiple patients behavior detection	TI-AWR1642BOOST (77 GHz)
2020	Skin cancer detection	Designed sensor (77 GHz)
2021	Contactless fitness tracking	TI-IWR1642 (77 GHz–81 GHz)
2022	Contactless monitoring of patients and elderly people alone	IWR6843AOPEVM (60 GHz–64 GHz)
2022	Measuring systolic blood pressure	TI-IWR6843AOP (60 GHz–64 GHz)
2022	Vital sign measuring	TI-IWR1443 (77 GHz–81 GHz)
2022	Health monitoring with posture estimation	TI-IWR6843 (60 GHz–64 GHz)
2022	Blood pressure monitoring	TI-AWR1843 (77 GHz–81 GHz)
2022	Cardiorespiratory rate monitoring	Commercial FMCW (122 GHz)
2022	Galvanic skin test to assess mental acuity and stress levels	TI-AWR1843 (77 GHz)
2022	Automated heart rate and breathing rate monitoring	TI-AWR1443BOOST (77 GHz)

- Robotics and Automation Applications

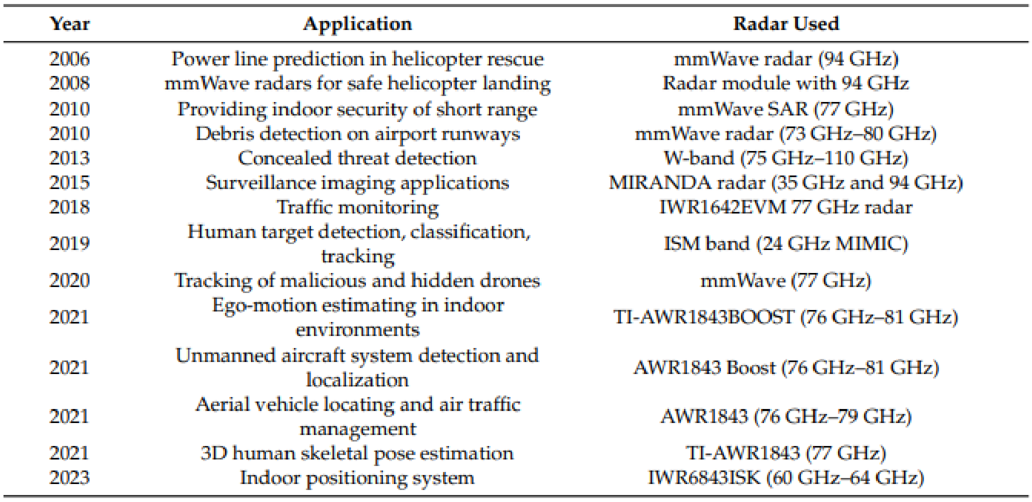
Table 5. Robotics and automation applications table on this article



Year	Application	Radar Used
2019	Intelligent robot for transparent object sensing	IWR6843 (60 GHz)
2020	Robot-mounted mmWave radar for tracking heart rate	IWR6843 (62 GHz)
2020	Predicting autonomous robot navigation	FMCW (77 GHz)
2020	Collision detection and avoidance	IWR6843 (60 GHz)
2020	mmWave radars as safe guard robots	IWR6843 (60 GHz)
2021	Automated indoor navigation and path tracking	AWR6843 (77 GHz)
2021	Glass wall and partition detection	IWR1443BOOSTEVM (77 GHz)

- Security and Surveillance Applications

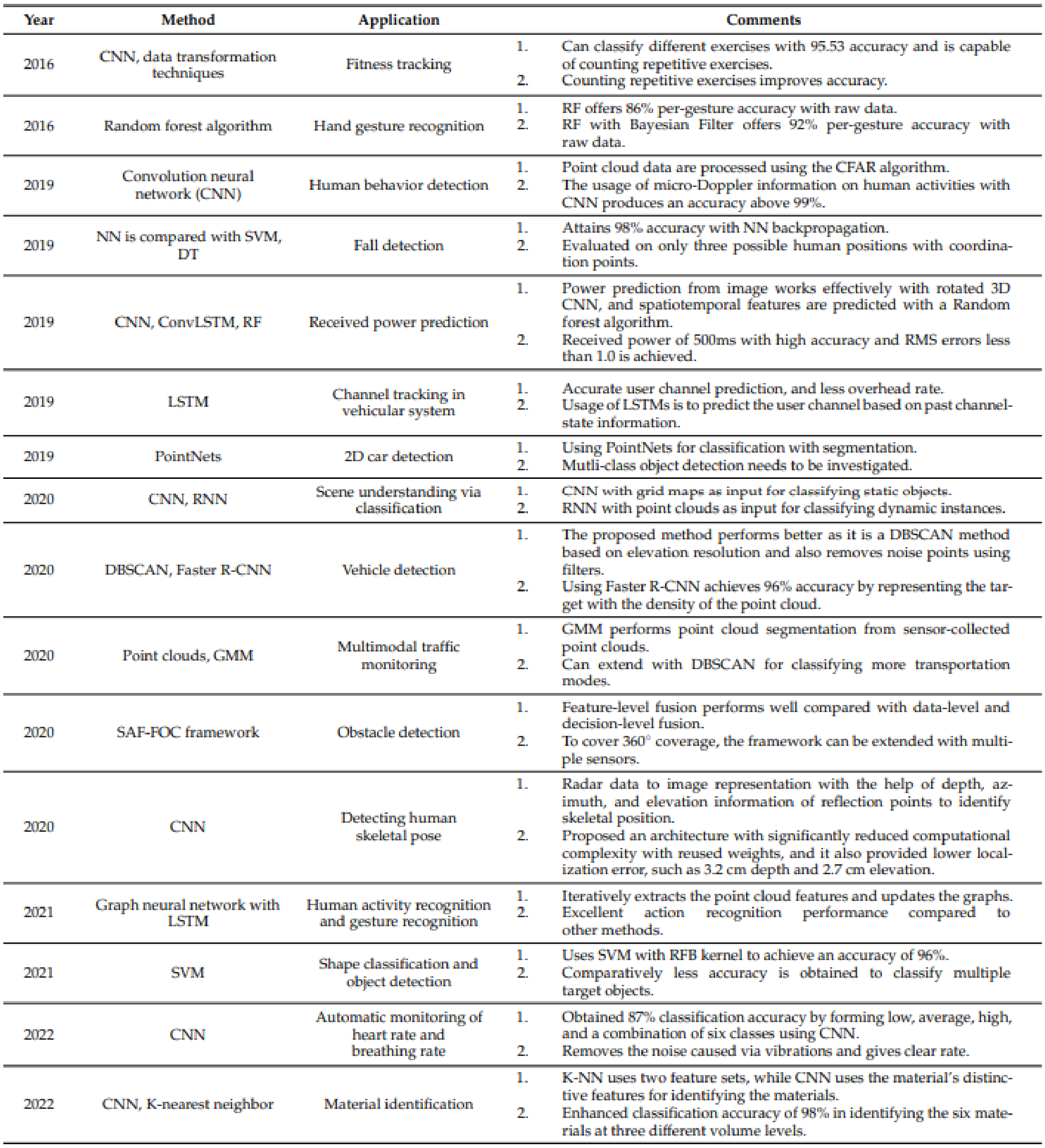
Table 6. Security and Surveillance applications table on this article



Year	Application	Radar Used
2006	Power line prediction in helicopter rescue	mmWave radar (94 GHz)
2008	mmWave radars for safe helicopter landing	Radar module with 94 GHz
2010	Providing indoor security of short range	mmWave SAR (77 GHz)
2010	Debris detection on airport runways	mmWave radar (73 GHz–80 GHz)
2013	Concealed threat detection	W-band (75 GHz–110 GHz)
2015	Surveillance imaging applications	MIRANDA radar (35 GHz and 94 GHz)
2018	Traffic monitoring	IWR1642EVM 77 GHz radar
2019	Human target detection, classification, tracking	ISM band (24 GHz MIMIC)
2020	Tracking of malicious and hidden drones	mmWave (77 GHz)
2021	Ego-motion estimating in indoor environments	TI-AWR1843BOOST (76 GHz–81 GHz)
2021	Unmanned aircraft system detection and localization	AWR1843 Boost (76 GHz–81 GHz)
2021	Aerial vehicle locating and air traffic management	AWR1843 (76 GHz–79 GHz)
2021	3D human skeletal pose estimation	TI-AWR1843 (77 GHz)
2023	Indoor positioning system	IWR6843ISK (60 GHz–64 GHz)

- Other Applications

Table 7. Other applications table on this article



Year	Method	Application	Comments
2016	CNN, data transformation techniques	Fitness tracking	1. Can classify different exercises with 95.53 accuracy and is capable of counting repetitive exercises. 2. Counting repetitive exercises improves accuracy.
2016	Random forest algorithm	Hand gesture recognition	1. RF offers 86% per-gesture accuracy with raw data. 2. RF with Bayesian Filter offers 92% per-gesture accuracy with raw data.
2019	Convolution neural network (CNN)	Human behavior detection	1. Point cloud data are processed using the CFAR algorithm. 2. The usage of micro-Doppler information on human activities with CNN produces an accuracy above 99%.
2019	NN is compared with SVM, DT	Fall detection	1. Attains 98% accuracy with NN backpropagation. 2. Evaluated on only three possible human positions with coordination points.
2019	CNN, ConvLSTM, RF	Received power prediction	1. Power prediction from image works effectively with rotated 3D CNN, and spatiotemporal features are predicted with a Random forest algorithm. 2. Received power of 500ms with high accuracy and RMS errors less than 1.0 is achieved.
2019	LSTM	Channel tracking in vehicular system	1. Accurate user channel prediction, and less overhead rate. 2. Usage of LSTMs is to predict the user channel based on past channel-state information.
2019	PointNets	2D car detection	1. Using PointNets for classification with segmentation. 2. Mutli-class object detection needs to be investigated.
2020	CNN, RNN	Scene understanding via classification	1. CNN with grid maps as input for classifying static objects. 2. RNN with point clouds as input for classifying dynamic instances.
2020	DBSCAN, Faster R-CNN	Vehicle detection	1. The proposed method performs better as it is a DBSCAN method based on elevation resolution and also removes noise points using filters. 2. Using Faster R-CNN achieves 96% accuracy by representing the target with the density of the point cloud.
2020	Point clouds, GMM	Multimodal traffic monitoring	1. GMM performs point cloud segmentation from sensor-collected point clouds. 2. Can extend with DBSCAN for classifying more transportation modes.
2020	SAF-FOC framework	Obstacle detection	1. Feature-level fusion performs well compared with data-level and decision-level fusion. 2. To cover 360° coverage, the framework can be extended with multiple sensors.
2020	CNN	Detecting human skeletal pose	1. Radar data to image representation with the help of depth, azimuth, and elevation information of reflection points to identify skeletal position. 2. Proposed an architecture with significantly reduced computational complexity with reused weights, and it also provided lower localization error, such as 3.2 cm depth and 2.7 cm elevation.
2021	Graph neural network with LSTM	Human activity recognition and gesture recognition	1. Iteratively extracts the point cloud features and updates the graphs. 2. Excellent action recognition performance compared to other methods.
2021	SVM	Shape classification and object detection	1. Uses SVM with RFB kernel to achieve an accuracy of 96%. 2. Comparatively less accuracy is obtained to classify multiple target objects.
2022	CNN	Automatic monitoring of heart rate and breathing rate	1. Obtained 87% classification accuracy by forming low, average, high, and a combination of six classes using CNN. 2. Removes the noise caused via vibrations and gives clear rate.
2022	CNN, K-nearest neighbor	Material identification	1. K-NN uses two feature sets, while CNN uses the material's distinctive features for identifying the materials. 2. Enhanced classification accuracy of 98% in identifying the six materials at three different volume levels.

2.2.3.2 Advantages and Disadvantages of mmWave Sensors Compared to Other Sensors

MmWave radar sensors have advantages and disadvantages compared to other detection sensors (such as cameras, LiDAR, ultrasonic sensors, etc.).

Advantages:

Performance in Adverse Weather Conditions: mmWave radar sensors can effectively operate even in adverse weather conditions such as rain, fog, snow, or dust that do not affect visibility. This is particularly important for autonomous vehicles and security systems.

Compact Design: With internally integrated patch antennas on the board, mmWave radar sensors take up very little space.

3D Detection: Radar sensors typically have the capability of 3D object and point velocity detection, allowing for more accurate determination of object position and movement. It also provides Range-Doppler data which includes both range and velocity information about objects around the sensor in 2D form.

Operation in Low-Light Conditions: Compared to optical sensors like camera systems, mmWave radar sensors can perform better in low-light conditions or darkness.

Privacy: In comparison to camera-based systems, radar sensors provide privacy as they do not require the collection of personal images or details.

Disadvantages:

Less Precise Positioning and Less Detailed Object Recognition: mmWave radar sensors do not have as detailed object recognition capabilities as sensors such as cameras or LIDAR. Especially distinguishing and identifying small objects may be difficult. Radio waves emitted by radar have low accuracy and produce very sparse data. In addition to wavelength issues, inherent noise is also a cause of sparsity in radar data. Therefore, many studies have investigated effectively combining radar with camera sensors to achieve more accurate detection and object identification that are not possible with radar alone[258].

Low Resolution: Compared to optical sensors, radar sensors generally have lower resolution, which can result in less detailed shape and structure information.

High Cost: In some cases, radar sensors can be more expensive compared to other detection sensors.

Type of Obstacle Detection: In some cases, radar sensors may be less effective in determining the types of obstacles (such as trees, signs, etc.) compared to sensors like LIDAR.

Poor Detection for Some Materials and Small Objects: There are some objects that it is not good at detecting by mmWave radar sensors. Also, some objects are difficult to detect. These are

[258] https://www.mdpi.com/2076-3417/12/4/2168

relatively small objects and objects with low reflectivity to radio waves, such as cardboard boxes[259].

2.2.3.3 Business and Marketing View

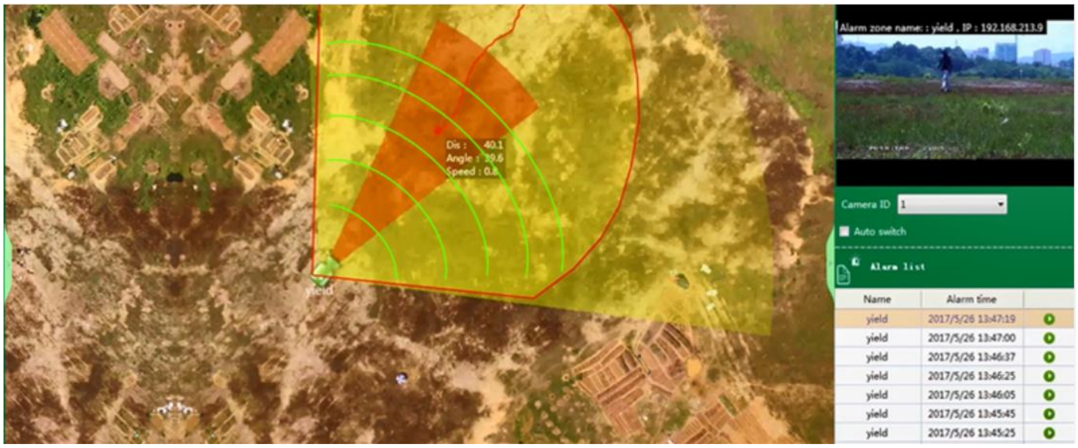
In crowded environments, security measures are becoming increasingly important, which in turn increases the demand for mmWave radar sensors. Particularly in places like airports, shopping malls, and other crowded areas, early detection and prevention of potential threats are of great importance. In such environments, features such as high sensitivity, fast response capabilities, and resistance to variable environmental conditions like weather and light may be necessary. mmWave radar sensors can meet these demands with their wide coverage and rapid detection capabilities.

Moreover, the privacy of customers or users is important to them. mmWave radar sensors can be developed in response to potential future concerns regarding data security and privacy. However, detecting small objects and objects with low reflection characteristics can be a significant challenge in such environments, so these factors should be considered in the selection and placement of sensors.

Many companies in the industry, especially in the security sector, have been able to respond to problems using mmWave radar sensors. For example, the company NANORADAR[260] in China has developed a solution called Radar Video Surveillance System by combining mmWave radar sensors and HD PTZ cameras. This system is equipped with radars that actively detect targets. The radar sensors used trigger PTZ cameras for automatic tracking. By employing video analysis technology and artificial intelligence algorithms for dual identity verification, the system sends alarms to the security monitoring center. This system can adapt to all kinds of adverse weather conditions such as rain, snow, fog, dust, and smoke, while providing 3D protection to lock targets in real time and providing access to the control center by recording alarm videos.

[259] https://us.metoree.com/categories/2759/

[260] http://en.nanoradar.cn/Article/detail/id/289.html



Alarm zone name: : yield . IP : 192.168.213.9

Camera ID 1

□ Auto switch

Alarm list

Name	Alarm time	
yield	2017/5/26 13:47:19	
yield	2017/5/26 13:47:00	
yield	2017/5/26 13:46:37	
yield	2017/5/26 13:46:25	
yield	2017/5/26 13:46:05	
yield	2017/5/26 13:45:45	
yield	2017/5/26 13:45:25	

Figure 26. Ground Surveillance Security System of NanoRadar Comp[261]

MINEW company[262], which operates with mmWave radar sensors, provides businesses with unique advantages by preferring mmWave radar sensors. With products containing mmWave radar sensors like MSR01, they offer services to meet the security and operational efficiency needs of businesses by providing solutions to problems such as Presence Detection, People Flow Management, Redefining Safety with Advanced Sensing, and Children/Adults/Pets Differentiation. Additionally, they provide significant advantages for businesses such as the ability to provide fast and accurate information, the ability to operate independently of lighting conditions, and protection of personal privacy.

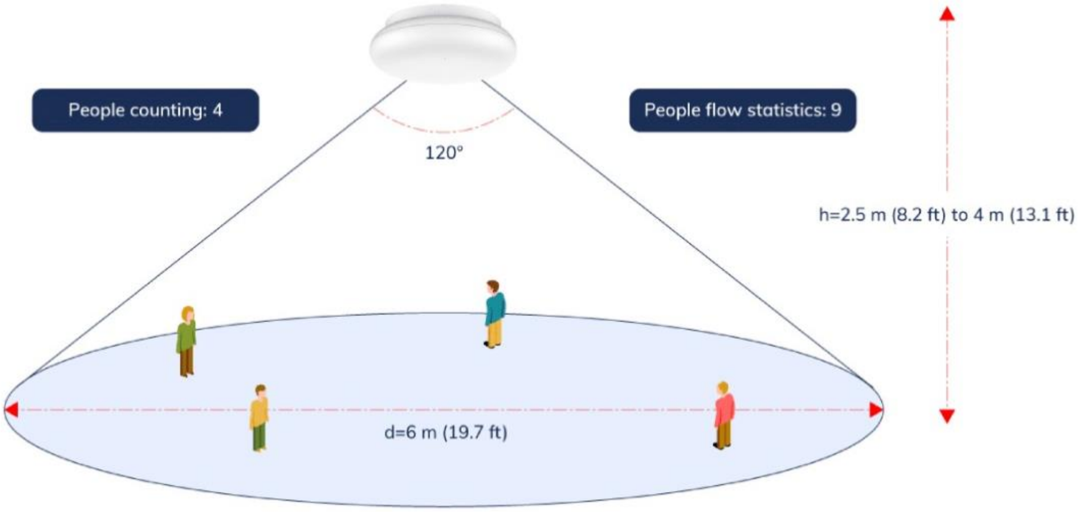


Figure 27. One of use cases of MSR01 product

[261] http://en.nanoradar.cn/Article/detail/id/369.html

[262] https://www.minew.com/product/msr01-millimeter-wave-radar-sensor

In addition, NoraSens company[263] has utilized mmWave radar sensors in applications such as 360° Short-Range Perception (ASPER), Automotive In-Cabin Monitoring (ACAM), Human Presence Detection (RIOT100), and Industrial Zone Monitoring (ZORM). These, along with many other companies, have employed mmWave radar sensors successfully for various purposes.

2.2.4 Market State of the Art in Construction Site Safety

According to a report from Mordor Intelligence, the artificial intelligence market in construction was valued at USD 2.74 billion in the previous year, and it is expected to reach USD 9.53 billion by the next five years, registering a CAGR of 24.30% during the forecast period.[264]

This growth would be mainly driven by AI's ability to transform and optimise laborious, repetitive, and common tasks across construction projects.

Construction sites are notorious for frequent workplace-related fatalities due to their dangerous and risk prone environment. Sadly, in the construction sector, as well as transportation and logistics, workers are five times more likely to die on the job when compared with workers in other sectors.[265] Tracking the real-time interactions of workers, machinery, and objects on site to signal any potential safety issues would be an important step in ensuring safety at each stage of the construction project. By monitoring the day-to-day site activity, managers and other stakeholders can assess safety compliance, as well as address potential safety risks that stem from transiting workers, equipment, and materials. There is also the tracking of onsite presence and entry into restricted access zones.

Many companies in the field of construction site safety are investing significantly in extracting analytical data from images and sensors. However, there is currently a gap in the market for a comprehensive, multi-modal platform that integrates IoT sensors, tags, live video feeds from drones and robots, and images from static cameras. The development of such a platform would represent a significant leap forward, providing a substantial competitive advantage over existing solutions.

Direct Competitors in Belgium:

- Buildevolution[266]: time-lapse solutions for marketing, management, & follow-up of construction projects.

[263] https://www.novelic.com/norasens-radar-sensor-technology

[264] https://www.mordorintelligence.com/industry-reports/artificial-intelligence-in-construction-market

[265] https://www.equipmentworld.com/business/article/15304920/construction-worker-death-rate-declines-22

[266] https://www.buildevolution.be/

- TowerEye[267]: security cameras, lighting, and alarm system all in a single solution.
- Time-Lapse Factory[268]: time lapse movies of complex projects within the construction and film industries that use specially designed and built time-lapse camera systems.
- Visuatech[269]: distributor of innovative camera solutions.
- VNI (View and Integrate)[270]: transforming camera images into reliable management information.
- Elapse[271]: time-lapse video retraces the entire construction or demolition project to keep a visual and aesthetic trace through a promotional film.
- AICON[272]: Data Analytics on construction sites. (AI and ML)

Direct competitors globally:

- Oxblue US[273]: OxBlue's construction cameras and technology bring together all aspects of a project via image monitoring, time-lapse video, and an intuitive interface accessible from any location. They work with AI that can estimate activities on the construction site such as weather downtime, security, activity, and construction equipment tracking.
- Earthcam US[274]: live-streaming video of jobsite activity with high quality images used for documentation and marketing. AI alerts (recognizing objects and activity on construction sites).
- Sensera Systems US[275]: real-time site intelligence solution using integrated compact solar/wireless cameras, sensors, and software in a single platform. The information is stored in Sensera SiteCloud which offers a range of features for viewing the archived camera images and video.
- Evercam IRL[276]: advanced camera software that enables clients to concentrate on better management and speedy completion.

[267] https://towereye.be/en

[268] http://www.timelapses.be/

[269] https://www.visuatech.be/

[270] https://www.viewandintegrate.be/

[271] http://www.elapse.be/

[272] https://www.aicon.construction/

[273] https://www.oxblue.com/

[274] https://www.earthcam.com/

[275] https://www.senserasystems.com/

[276] https://evercam.io/

- Devisubox FR[277]: time-lapse cameras that capture HD images of the construction site, which are transmitted to a cloud platform and accessible in real time. AI Features: AI (face blurring, helmet detection, people counting, attractive images) 3D integration with ioT trigger (motion detection)
- Baucamera GER[278]: suppliers of rental cameras for construction supervision and documentation.
- Enlaps FR[279]: time-lapse video used for sharing construction project monitoring on social media or broadcast during a grand opening. AI Features: activity monitoring, object detection and 3D visualisation.
- Others: PhotoSentinel[280] | US, Camdo[281] | US, Spot-r[282] | US, Bouwcam[283] | NL, Boxcam[284] | FR.

Competitors' use cases:

- Smartvid.io[285], US (Now, Newmetrix) : Smartvid.io uses a user-friendly AI platform called Vinnie (Very Intelligent Neural Network for Insight and Evaluation) to mainly detect project risk and improve worker safety.
- Indus.ai[286], US: Uses a combination of AI, computer vision, and machine learning to analyze project progress, monitor all site activity and scan for safety concerns. Instead of a human eye monitoring everything, Indus.ai sends alerts. [recently taken over by Procore]

[277] https://www.devisubox.com/
[278] https://bau.camera/
[279] https://enlaps.io/fr/
[280] https://photosentinel.com/
[281] https://cam-do.com/
[282] https://www.spotr.ai/
[283] https://bouwcam.live/
[284] https://boxcam.fr/
[285] https://www.newmetrix.com/
[286] https://www.procore.com/en-gb

- Others: Doxel.ai[287] | US, Cad42[288] | FR, Airsquire.ai[289] | NL, Pixelvision[290] | B, Firmus | ISR, Avvir[291] | US, OpenSpace[292] | US , Pillar Technologies[293] | US, Pix4D[294] | CH.

2.2.5 Market State of the Art on the Use of Drones in the Construction Industry

Drones in the construction industry already add a lot of value today. They are used to speed up surveys and to measure the progress on construction sites. The use of drones is common at many large construction companies where they are operated by certified drone pilots. Drones that are used often for this purpose are DJI Phantom 4 RTK[295] or the DJI Mavic 3[296] Enterprise with RTK module[297]. These drones weigh respectively 1,4kg and 915g so they must be operated in category A2. This means that they are legally not allowed to fly over people who are not involved in the operation, and they have to keep a 5m horizontal distance from them. To be involved in the operation, a person needs to know what the drone is used for and what its flight plan is, what the safety instructions are. Additionally, involved persons need to always know where the drone is and be ready to act in case of an unexpected problem with the drone. On a construction site with workers from many contractors and subcontractors moving around, it is difficult to inform everyone at the right moment.

A second limitation construction companies face is that they have a limited number of drone pilots who often also have other responsibilities (e.g., surveyors). This means that it's difficult to ‘quickly do a scan’ to e.g., measure progress of a certain part of the site. For example, if the drone pilot is working on a site near Antwerp, it is impractical to quickly drive to a second site near Kortrijk to quickly measure the volume of a pile of sand that needs to be transported the next day.

Automated drones are the solution to provide regular updates without having to send a certified drone pilot on site. Additionally, they can be used for security purposes. There are several

[287] https://doxel.ai/
[288] https://cad42.com/
[289] https://www.airsquire.ai/
[290] https://www.pixelvision.ai/
[291] https://www.avvir.io/
[292] https://www.openspace.ai/
[293] https://pillar.tech/
[294] https://www.pix4d.com/
[295] https://enterprise.dji.com/phantom-4-rtk
[296] https://enterprise.dji.com/mavic-3-enterprise
[297] https://www.easa.europa.eu/en/domains/civil-drones/drones-regulatory-framework-background/open-category-civil-drones

companies working on a ‘drone-in-a-box’ solution. Examples are the DJI Dock[298] with a DJI Matrice 30 drone (3,95kg)[299], the Percepto dock with drone (8,5kg)[300] or the Dronematrix Yacob (±6kg)[301]. All drones currently being developed are relatively heavy and the price of a system ranges from 30-50kEuro. This means that the ground risk and dealing with uninvolved people will be difficult to handle. Additionally, the price tag is a limitation for many smaller companies to adopt the technology.

The innovation **Airobot** wants to pursue is to use low-cost and lightweight drones for the same purpose - combined with smart cloud software to control them remotely. If successful, this will offer 2 large benefits to potential users. Firstly, they'll be able to more easily use the situation in areas with uninvolved people (cities…). Secondly, the purchase price of these drones is much lower (< 1.000 Euro), which makes them accessible to many more companies. However, compared to the larger platforms, these drones have more limited resources (camera quality, anti-collision technology, GPS accuracy…) which makes it more difficult to automate them. Optimized flight planning and threat detection algorithms, that can efficiently work at the edge, should be developed by combining information from multiple sensors to overcome the challenges posed by resource limitation. Moreover, exploring lightweight cryptographic algorithms and non-repudiation mechanisms, such as secure logging systems, will enhance both security and operational efficiency in the context of automated drones, while also providing valuable evidence in the event of accidents or system issues.

2.2.6 Market State of the Art of IIoT for Asset Monitoring

The Industrial Internet-of-Things (IIoT) market is a steadily growing market[302]. The creation of a digital twin of each object in the supply chain creates value for different stakeholders. This datafication of the supply chain has been going on for some years but is still not finalized. One of the main reasons for this is that the Return-on-Investment calculation only fitted for high value assets in the beginning and is now slowly adopting for more and more assets. In addition, the available technology was still expensive and did come with a high operational cost due to the need for replacement of the batteries.

The “asset pyramid”, Figure 5, shows the evolution of which assets were datafied first (the top in grey) and the evolution towards the current situation (the bottom in green). Initially only high-

[298] https://enterprise.dji.com/dock

[299] https://enterprise.dji.com/matrice-30

[300] https://percepto.co/drone-in-a-box/percepto-base/

[301] https://www.dronematrix.eu/product

[302] https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/iot-value-set-to-accelerate-through-2030-where-and-how-to-capture-it

value assets with a battery could be datafied. With the evolution of technology and the lower TCO cost, it is becoming possible to also connect low-value assets without battery to the Internet. The figure reflects well the general market trend, namely going to lower value assets, but higher volumes.

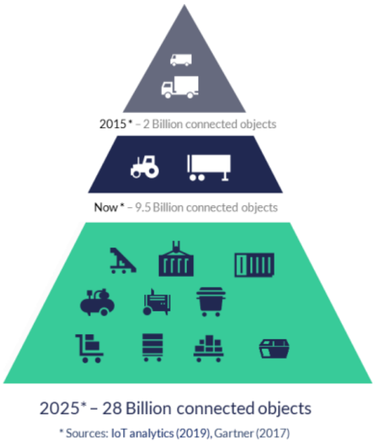

2015* – 2 Billion connected objects
Now* – 9.5 Billion connected objects
2025* – 28 Billion connected objects
* Sources: IoT analytics (2019), Gartner (2017)

Figure 28. Asset pyramid[303]

The growth of the business of Sensolus will depend on if we can answer to this trend. Connecting more assets, in higher volumes, at lower connection costs per asset is key to further dataficate more asset flows and thus further optimize the supply chain processes.

The most important competition of the BLE tag-based solution Sensolus presented in SINTRA are the players in the RFID market. (Asset infinity, Zebra solutions, ...). Today, some of the existing RFID players are moving up in the chain, and some international innovators extend their portfolio with BLE tags also lowering the cost/asset, all with the focus on close-loop supply chain.

2.2.7 Market State of the Art of Logistic Hubs and Harbours Safety/Security

Video analytics is one of the important components of logistics hub safety and security solutions. The global market for video analytics is estimated to be 7.1 billion USD in 2022 and is prognosed to grow into 20.3 billion USD by 2027[304]. Asia Pacific is expected to record the highest growth over this period. One of the biggest drivers behind this is thought to be the increased focus by governments to improve public safety and reduce crime rates. The competence of analytics to incorporate way more relations into the analyses as a human and the exclusion of human error and fatigue are main drivers for a technology solution. Limitations in labour cost as well as shortage at the labour market increase this focus. Generating predictive information using video

[303] https://www.grandviewresearch.com/industry-analysis/industrial-internet-of-things-iiot-market

[304] https://www.marketsandmarkets.com/Market-Reports/intelligent-video-analytics-market-778.html

analytics is thought to be an important opportunity within video analytics, using statistics, modelling, and data mining, using ML and AI techniques. Important constraint for this technology to develop is the compliance to GDPR. This brings limitations to the use of analytics identifying individuals which limits the possibility to individualise patterns of behaviour and distinguish these patterns from anomalies. Also, GDPR requires an organisation to have valid reasons to install video surveillance. This tends to lead to a situation where video surveillance is only present in the area of immediate risk. The absence of cameras (and other data sources) for a wider area limits the possibilities to find patterns that lead to an infringement.

According to AllTheResearch, the global AI for surveillance and security market will see substantial growth by USD 4.46 billion in 2023. This market is expanding at a faster rate to a wider range of countries. At least 43% of 176 countries are actively using AI for surveillance and security. This includes smart city/safe city platforms, facial recognition systems, and smart policing. Software companies like Avigilon Control Center (ACC), Avigilon Access Control Manager (ACM), Hitachi Video Management Platform (VMP), Dell Technologies IoT Solution for Surveillance, Eagle Eye Cloud VMS, and DMI EndZone are providing various AI for surveillance and security solutions and platforms for different camera manufactures like Manufacturers Axis, Milestone, Vivotek, QNAP, Optica, Mobotix, ACTi, Arecont Vision, Avigilon, Bosch, Canon, Cisco, and Extreme CCTV.

The second component used in logistics hub security solutions is drug detection. For years, investigation officers use chemical spot tests for presumptive drug testing. However, such tests are only available for a small range of drugs, are prone to false positive reactions, require manual handling of, and reaction with, the suspect material, and require single-use consumables and chemicals thus impacting the environment. As a more advanced solution, NIR (near-infrared) sensors are very promising for fast and reliable on-scene drug detection. Currently, NIRS is widely used in many industries including pharmaceuticals, petrochemical, agri-food, and recycling. While this method has existed for decades, the cost and complexity of existing spectrometer analysers hinder the full adoption of these systems and limit the number of measuring nodes, locations, and use cases. The large dimensions of spectrometer systems make their suitability for in-field use scenarios prohibitive. Furthermore, the high barrier to adoption makes them beyond reach for small and medium sized enterprise users, who are forced to send samples to external analytical labs. Relying on external laboratories for chemical analysis is expensive, time-consuming and may cause disruptions to critical processes.

The third major technology focus of logistics hub security is on drones. Mobile robots are already being used with some frequency. This usually concerns relatively simple drones with simple cameras. Until now, drone inspection almost always took place with a pilot on site and within the visible range. Autonomous flying and certainly flying beyond visual line of sight (BVLOS) is rare (strict legislation also plays a role in this). Reliance on a pilot severely limits the scalability of UAV based solutions and services. The bulk of the drone applications for inspection concerns applications that are aimed at collecting regular camera images (RGB image). More fundamental innovation steps are expected in the coming years in the field of increasing the autonomy of drones, self-coordination, and interaction with the physical world. The final publication of the new ISO approved drone standards is expected to have a massive impact on the future growth of the global drone industry. Several recent reports have attempted to forecast the economic impact of air drones globally. For instance, in its report Drones Reporting for Work, Goldman Sachs has estimated that the size of the global drone industry will reach $100 billion by 2020[305]. Most recently, analysts at Barclays estimate that the global commercial drone market will grow tenfold from $4bn in 2018 to $40bn in five years[306]. Differing estimations these may be, but it seems all are predicting rapid growth in the sector.

2.2.8 Market State of the Art in the Drone Sector

Currently, drone inspections of industrial sites predominantly rely on Visual Line-of-Sight (VLOS) flights. However, this approach is marred by inefficiencies, primarily due to the reliance on subcontracted pilots who are typically only engaged for planned inspections. These planned inspections are geared toward expected events e.g, inspection of leaking pipes. Nevertheless, for robust security monitoring of vital infrastructure such as ports, there is an imperative need for an integrated aerial and ground surveillance system like drones and robot dogs, capable of instantaneous deployment at any given moment.

Effective security monitoring of critical infrastructure necessitates the combination of diverse data sources, including conventional Internet of Things (IoT) sensors, such as localization sensors, fixed cameras, and a high-performance wireless network. Currently most of these data sources work in parallel of each other, they all have their own data platform, data resolution and formats. This data needs to be combined with multimodal data integration algorithms, without which we believe it is not possible to ensure comprehensive security and surveillance of big critical infrastructure.

2.2.9 Market State of the Art in the Security Sector

Today, security primarily consists of one sensor - one alarm - one alarm response, lacking an integrated approach to full multi-sensor security. We are going to explore how we can evolve towards integrated security handling. This would lead us to a proactive approach with a reactive response.

[305] https://dronedj.com/2019/01/28/drones-reporting-for-work-goldman-sachs/

[306] Patrick McGee, “How the commercial drone market became big business,” The Financial Times Limited 2020, https://www.ft.com/content/cbd0d81a-0d40-11ea-bb52-34c8d9dc6d84

This reactive response would visualize the entire on-site situation comprehensively and allow us to holistically evaluate it in the SOC (Security Operations Centre) through a digital twin. This, in turn, would elevate the awareness, thinking, and operational patterns of our operators to a higher level.

Through Integration of data platforms and systems, organizations can become more accurate at identifying and remediating risks, responding to incidents quickly, reporting on trends, and reducing costs

This investment is not exclusively focused on the need to identify cost-reduction strategies but rather to improve risk posture and to address corporate responsibility to promote a culture of security awareness and protection of people and assets. This wil also enlarge the ability to protect organizations from operations risks that may disrupt the business and overall revenue creation

Securitas invests in a continuous transformation journey towards technology-based solutions. Our strategy adds resilience and creates a significant platform for innovation, as we have the ambition to drive and redefine the future of the security industry. This project is completely in line with our strategy and will enable us to differentiate in the Belgian security market and maintain a competitive edge against other security companies.

2.2.10 Market State of the Art on Visual Inspection

Currently, depending on the type of visual inspection, they are carried out manually by operators, professional climbers and certified inspectors. The challenge that these inspectors face in the current market state of the art is that a manual inspection is inefficient and requires extra tools (e.g. scaffolding, aerial platforms, or even diving equipment). Visual inspections in the industry are costly as they take a long time due to the high costs of the inspectors, equipment and operational losses linked to the downtime of the assets. The inspectors are struggling with a lot of administrative work (very often on paper) to create the inspection reports. These inspections are also not without danger. Some areas which need to be inspected on an industrial yard or infrastructure are hard to reach (e.g.: heights, confined spaces, underwater inspections). Additionally, in the scenario of an incident, the inspector is required on site for the visual inspection and risks being injured when the incident unexpectedly escalates. Due to these constraints which are linked to the market state of the art on visual inspections, industrial companies and inspection service providers are looking for new technologies like robotic inspections to improve the inspection efficiency and overall costs. By doing so the plan is to improve the overall inspection safety by reducing the exposure of the inspector to external risks. They are also looking towards the use of Artificial intelligence for support during the inspection itself and the creation of the inspection reports.

3. REFERENCES

3.1 Bibliography

1. A. Botta, W. D. Donato, V. Persico and A. Pescapé, "Integration of cloud computing and Internet of Things: A survey," Future Gener. Comput. Syst., vol. vol. 56, p. pp. 684–700, 2016.
2. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, Cur-ran Associates, Inc., 2012, p.9.
3. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need, 2023.
4. Abbas, A., & Chalup, S. (2021). Affective analysis of visual scenes using face pareidolia and scene-context. Neurocomputing, 423, 634-645.
5. Abdullah, H. Z., Warren, K., Bindschaedler, V., Papernot, N., & Traynor, P. (2021). SoK: The Faults in our ASRs: An Overview of Attacks against Automatic Speech Recognition and Speaker Identification Systems. arXiv preprint arXiv:2007.06622v3.
6. Ahmad, N., & Yoon, J. (2021). StrongPose: Bottom-up and Strong Keypoint Heat Map Based Pose Estimation. International Conference on Pattern Recognition. Retrieved from ICPR 2020.
7. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.
8. Akyildiz, I. F., Pompili, D., & Melodia, T. (2005). Underwater acoustic sensor networks: Research challenges. Ad Hoc Networks, 3(3), 257-279.
9. Almog, J.; Zitrin, S. Colorimetric Detection of Explosives. In Aspects of Explosives Detection, 1st ed.; Marshall, M., Oxley, J., Eds.; Elselvier: Oxford, UK, 2009; pp. 41–58. ISBN 978-0-12-374533-0.
10. Amin, M. G. (Ed.). (2017). Radar for indoor monitoring: Detection, classification, and assessment. CRC Press.
11. Arnab A, Dehghani M, Heigold G, Sun C, Lučić M, Schmid C. Vivit: A video vision transformer. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 6836–6846.
12. Arppitha, Krishna., Neha, Pendkar., Shruti, Kasar., Umesh, Mahind., Shridhar, Desai. (2021). Advanced Video Surveillance System. doi: 10.1109/ICSPC51351.2021.9451694
13. Batapati P, Tran D, Sheng W, Liu M, Zeng R. Video analysis for traffic anomaly detection using support vector machines. Proceedings of the World Congress on Intelligent Control and Automation (WCICA). 2015 03;2015:5500–5505.
14. Beddiar, D.R., Nini, B., Sabokrou, M. et al. Vision-based human activity recognition: a survey.
15. Benson, S.; Speers, N.; Otieno-Alego, V. Portable explosive detection instruments. In Forensic Investigation of Explosions, 2nd ed.; Beveridge, A., Ed.; CRC Press: Boca Raton, FL, USA, 2011; pp. 691–724. ISBN 978-0-367-77820-0.
16. Bertasius G, Wang H, Torresani L. Is space-time attention all you need for video understanding? In: ICML. vol. 2; 2021. p. 4.
17. Bin F., Xin F., Jianguo C. 2021. A MEMS sensor-based human body gesture recognition method for the elderly-aiding mechanism. Journal of Harbin University of Commerce (Natural Sciences Edition),37(05),590-594.
18. BRESSON, Guillaume, et al. Simultaneous localization and mapping: A survey of current trends in autonomous driving. IEEE Transactions on Intelligent Vehicles, 2017, 2.3: 194-220.
19. Burdick, A., & Szalay, A. (2012). The case for cloud-based sensor networks. IEEE Internet Computing, 16(6), 63-67.
20. C. Feng, A. Mehmani, and J. Zhang, “Deep learning-based real-time building occupancy detection using AMI data,” IEEE Trans. Smart Grid, vol. 11, no. 5, pp. 4490–4501, Sep. 2020.

21. Carreira J, Zisserman A. Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 6299–6308.
22. CCTV Security Pros. (n.d.). The Drawbacks of Old CCTV Security Cameras - How to Upgrade. Retrieved June 4, 2024, from https://www.cctvsecuritypros.com
23. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys (CSUR), 41(3), 1-58.
24. Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Comput Surv. 2009 jul;41(3). https://doi.org/10.1145/1541880.1541882.
25. CHEN, Changhao, et al. A survey on deep learning for localization and mapping: Towards the age of spatial machine intelligence. arXiv preprint arXiv:2006.12567, 2020.
26. Chen W, Xu H, Li Z, Pei D, Chen J, Qiao H, et al. Unsupervised anomaly detection for intricate kpis via adversarial training of vae. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications. IEEE; 2019.
27. Chen Y, Liu Z, Zhang B, Fok W, Qi X, Wu YC. Mgfn: Magnitude contrastive glance-and-focus network for weakly-supervised video anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37; 2023. p. 387–395.
28. Chen, Y., Zhang, J., Yuan, X., Zhang, S., Chen, K., Wang, X., & Guo, S. (2022). SoK: A Modularized Approach to Study the Security of Automatic Speech Recognition Systems. arXiv preprint arXiv:2103.10651v2.
29. Chris Lewis Group. (n.d.). Why You Should Upgrade Your CCTV Cameras. Retrieved June 4, 2024, from https://www.chrislewis.co.uk
30. Cornacchia, M., Papa, F., & Sapio, B. (2020). User acceptance of voice biometrics in managing the physical access to a secure area of an international airport. Technology Analysis & Strategic Management, 32(5), 585-598.
31. Datla, R., Chalavadi, V., & Chalavadi, K. M. (2022). A multimodal semantic segmentation for airport runway delineation in panchromatic remote sensing images. International Conference on Machine Vision. Retrieved from Semantic Scholar.
32. Deepak, G., & Santhanavijayan, A. (2020). A Novel Semantic Approach for Intelligent Response Generation using Emotion Detection Incorporating NPMI Measure. Procedia Computer Science, 168, 126-133.
33. Defend Security Group. (n.d.). When Should You Upgrade Your Existing CCTV System To Newer Technology? Retrieved June 4, 2024, from https://www.defendsecuritygroup.com.au
34. Deshmukh, R., Sun, D., Kim, K., & Hwang, I. (2021). Temporal logic learning-based anomaly detection in metroplex terminal airspace operations. Transportation Research Part C: Emerging Technologies, 124, 102955.
35. DeTone, D.; Malisiewicz, T.; Rabinovich, A. Superpoint: Self-supervised interest point detection and description. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 224–236.
36. DUBROFSKY, Elan. Homography estimation. Diplomová práce. Vancouver: Univerzita Britské Kolumbie, 2009, 5.
37. Dusmanu, M.; Rocco, I.; Pajdla, T.; Pollefeys, M.; Sivic, J.; Torii, A.; Sattler, T. D2-net: A trainable cnn for joint description and detection of local features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8092–8101.
38. EBERT-UPHOFF, Imme, et al. CIRA Guide to Custom Loss Functions for Neural Networks in Environmental Sciences--Version 1. arXiv preprint arXiv:2106.09757, 2021.
39. ELHARROUSS, Omar; ALMAADEED, Noor; AL-MAADEED, Somaya. A review of video surveillance systems. Journal of Visual Communication and Image Representation, 2021, 77: 103116.
40. Eli, Katsiri., George, Papastefanatos., Manolis, Terrovitis., Timos, Sellis. (2014). Airport Context Analytics. doi: 10.1007/978-3-319-11113-1_13.
41. Ewing, R.G.; Atkinson, D.A.; Eiceman, G.A.; Ewing, G.J. A critical review of Ion Mobility Spectrometry for the detection of explosives and explosive related compounds. Talanta 2001, 54, 515–529.

42. Ezequiel, Roberto, Zorzal., Ariel, Fernandes., Bruno, Castro. (2017). Using Augmented Reality to overlapping information in live airport cameras. doi: 10.1109/SVR.2017.53.
43. Fan H, Xiong B, Mangalam K, Li Y, Yan Z, Malik J, et al. Multiscale vision transformers. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 6824–6835.
44. Fan, P., Guo, D., Zhang, J., Yang, B., & Lin, Y. (2023). Enhancing multilingual speech recognition in air traffic control by sentence-level language identification. arXiv preprint arXiv:2305.00170v1.
45. FAN, Jiayi, et al. Improvement of object detection based on faster R-CNN and YOLO. In: 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). IEEE, 2021. p. 1-4.
46. Fan Q, Chen CFR, Kuehne H, Pistoia M, Cox D. More is less: Learning efficient video representations by big-little network and depthwise temporal aggregation. Advances in Neural Information Processing Systems. 2019; 32.
47. Feichtenhofer C, Fan H, Malik J, He K. Slowfast networks for ideo recognition. In: Proceedings of the IEEE/CVF international conference on computer vision; 2019. p. 6202–6211.
48. Fog Computing: Current Research and Future Challenges, March 2018. Conference: 1. GI/ITG KuVS Fachgespräche Fog ComputingAt: Darmstadt, Germany.
49. Georgescu MI, Barbalau A, Ionescu RT, Khan FS, Popescu M, Shah M. Anomaly detection in video via self-supervised and multi-task learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. p.12742–12752.
50. Gezer, V., Um, J., & Ruskowski, M. (2017). An extensible edge computing architecture: Definition, requirements and enablers. Proceedings of the UBICOMM.
51. Giovanni Diraco, Gabriele Rescio, Andrea Caroppo, Andrea Manni and Alessandro Leone Human Action Recognition in Smart Living Services and Applications: Context Awareness, Data Availability, Personalization, and Privacy.
52. Gong D, Liu L, Le V, Saha B, Mansour MR, Venkatesh S, et al. Memorizing Normality to Detect Anomaly: Memory-augmented Deep Autoencoder for Unsupervised Anomaly Detection. CoRR. 2019; abs/1904.02639.
53. Guangyi T., Jianjun N., Yonghao Z., Yang G., Weidong C., A Survey of Object Detection for UAVs Based on Deep Learning, Remote Sensing, 2024, 16, 149.
54. Guo, H., Fan, X., & Wang, S. (2017). Human attribute recognition by refining attention heat map. Pattern Recognition Letters. Retrieved from Pattern Recognit. Lett.
55. Gupta, N., Gupta, S.K., Pathak, R.K. et al. Human activity recognition in artificial intelligence framework: a narrative review.
56. Gupta, P., & Margam, M. (2021). CCTV as an efficient surveillance system? An assessment from 24 academic libraries of India. Global Knowledge, Memory and Communication, 70(4/5), 355-376.
57. H. Gao and D. Dang. Learning enriched features via selective state spaces model for efficient image deblurring, 2024.
58. H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling. U2fusion: A unified unsupervised image fusion network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1):502–518, 2020.
59. H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo. Fusiondn: A unified densely connected network for image fusion.
60. Hansard, M., Lee, S., Choi, O., & Horaud, R. (2013). Time-of-flight cameras: Principles, methods, and applications. Springer.
61. HAO, Shijie; ZHOU, Yuan; GUO, Yanrong. A brief survey on semantic segmentation with deep learning. Neurocomputing, 2020, 406: 302-321.
62. Harvey, S.; Peters, T.J.; Wright, B.W. Safety considerations for sample analysis using a Near-Infrared (785 nm) Raman laser source. Appl. Spectrosc. 2003, 57, 580–587.
63. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 770–778.
64. He, W., Yan, G., & Xu, L. D. (2014). Developing vehicular data cloud services in the IoT environment. IEEE Transactions on Industrial Informatics, 10(2), 1587-1595.

65. Hieu H. Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, Sergio A. Velastin, Computer Vision and Pattern Recognition Video-based Human Action Recognition using Deep Learning: A Review.
66. Huang C, Wen J, Xu Y, Jiang Q, Yang J, Wang Y, et al. Selfsupervised attentive generative adversarial networks for video anomaly detection. IEEE transactions on neural networks and learning systems. 2022.
67. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–4708.
68. Huang, X., Cheena, H., Thomas, A., Tsoi, J. K. P., & Gao, B. (2021). Indoor detection and tracking of people using mmWave sensor. Journal of Sensors, 2021, Article 6657709. Fig. 1, p. 2. https://doi.org/10.1155/2021/6657709
69. Huang, Y.; Li, W.; Dou, Z.; Zou, W.; Zhang, A.; Li, Z. Activity Recognition Based on Millimeter-Wave Radar by Fusing Point Cloud and Range–Doppler Information. Signals 2022, 3, 266-283. https://doi.org/10.3390/signals3020017.
70. Internet of Things (IOT): Confronts and Applications. International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653.
71. Ionescu RT, Khan FS, Georgescu M, Shao L. Object-centric Auto-encoders and Dummy Anomalies for Abnormal Event Detection in Video. CoRR. 2018; abs/1812.04960. http://arxiv.org/abs/1812.04960.
72. J. Kang, I. Cohen, G. Medioni, and C. Yuan. Detection and tracking of moving objects from a moving platform in presence of strong parallax. In Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 1, pp. 10–17 Vol. 1, 2005. doi: 10.1109/ICCV.2005.72.
73. J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte. Swinir: Image restoration using swin transformer, 2021.
74. J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang. Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing, 29:4980–4995, 2020.
75. J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma. Swinfusion: Cross-domain longrange learning for general image fusion via swin transformer. IEEE/CAA Journal of Automatica Sinica, 9(7):1200–1217, 2022. doi: 10.1109/JAS. 2022.105686.
76. KAKANI, Vijay, et al. Feasible self-calibration of larger field-of-view (FOV) camera sensors for the advanced driver-assistance system (ADAS). Sensors, 2019, 19.15: 3369.
77. Kalbo, N., Mirsky, Y., Shabtai, A., & Elovici, Y. (2020). The Security of IP-Based Video Surveillance Systems. Sensors, 20(17), 4806.
78. KAMATH, Uday, et al. Transfer learning: Domain adaptation. Deep learning for NLP and speech recognition, 2019, 495-535.
79. K.D. Hakkel, M. Petruzzella, F. Ou, A. van Klinken, F. Pagliano, T. Liu, R.P.J. van Veldhoven and A. Fiore, *Nat. Commun.* 13(1), 1–8 (2022), DOI: 10.1038/s41467-021-27662-1.
80. KEJRIWAL, Mayank, et al. An evaluation and annotation methodology for product category matching in e-commerce. Computers in Industry, 2021, 131: 103497.
81. Kingma DP, Welling M. An Introduction to Variational Autoencoders. CoRR. 2019;abs/1906.02691. http://arxiv.org/abs/1906.02691.
82. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:13126114. 2013.
83. Kranenburg, R. F., Ramaker, H. J., & van Asten, A. C. (2022). Portable near infrared spectroscopy for the isomeric differentiation of new psychoactive substances. *Forensic Science International*, *341*, 111467.
84. Kolb, A., Barth, E., Koch, R., & Larsen, R. (2010). Time-of-flight sensors in computer graphics. Eurographics State-of-the-Art Report, 119-134.
85. Kong, Y., Fu, Y. Human Action Recognition and Prediction: A Survey. Int J Comput Vis.
86. Kranenburg, R.F.; Verduin, J.; Weesepoel, Y.; Alewijn, M.; Heerschop, M.; Koomen, G.; Keizers, P.; Bakker, F.; Wallace, F.; van Esch, A.; et al. Rapid and robust on-scene detection of cocaine in street samples using a handheld Near-Infrared spectrometer and machine learning algorithms. Drug Test Anal. 2020, 12, 1404–1418.

87. Kumar, P., Chauhan, S. & Awasthi, L.K. Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies.
88. Kumaran SK, Dogra DP, Roy PP, Mitra A. Video trajectory classification and anomaly detection using hybrid CNN-VAE. arXiv preprint arXiv:181207203. 2018.
89. L. Chen, X. Chu, X. Zhang, and J. Sun. Simple baselines for image restoration, 2022.
90. Lin J, Gan C, Han S. Temporal shift module for efficient video understanding. CoRR abs/1811.08383 (2018) 1811.
91. L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: a survey, Version: 1, arXiv:1809 .02165, 2018.
92. L. Tang, X. Xiang, H. Zhang, M. Gong, and J. Ma. Divfusion: Darkness-free infrared and visible image fusion. Information Fusion, 91:477–493, 2023.
93. Lee, G.; Kim, J. Improving Human Activity Recognition for Sparse Radar Point Clouds: A Graph Neural Network Model with Pre-Trained 3D Human-Joint Coordinates. Appl. Sci. 2022, 12, 2168. Fig.6, p. 12. https://doi.org/10.3390/app12042168.
94. Lee, I., & Lee, K. (2015). The Internet of Things (IoT): Applications, investments, and challenges for enterprises. Business Horizons, 58(4), 431-440.
95. Lee, J., Bagheri, B., & Kao, H. A. (2015). A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems.
96. Liu D, Cui Y, Chen Y, Zhang J, Fan B (2020, Elsevier B.V.) Video object detection for autonomous driving: motion-aid feature calibration. Neurocomputing 409:1–11.
97. Li, J., Chao, C., Pan, L., Azghadi, M. R., Ghodosi, H., & Zhang, J. (2023). Security and Privacy Problems in Voice Assistant Applications: A Survey. arXiv preprint arXiv:2304.09486v1.
98. Li S, Liu F, Jiao L. Self-training multi-sequence learning with transformer for weakly supervised video anomaly detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36; 2022. p. 1395–1403.
99. Li Y, Wu CY, Fan H, Mangalam K, Xiong B, Malik J, et al. Mvitv2: Improved multiscale vision transformers for classification and detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 4804–4814.
100. Liu, W., Xia, T., Wang, D., & Fu, H. (2019). LiDAR point cloud data processing and analysis in urban environments: Methods and applications. ISPRS Journal of Photogrammetry and Remote Sensing, 149, 59-72.
101. Lin S, Clark R, Birke R, Schönborn S, Trigoni N, Roberts S. Anomaly detection for time series using vae-lstm hybrid model. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Ieee; 2020. p. 4322–4326.
102. Liu SWTT, Ngan HYT, Ng MK, Simske SJ. Accumulated Relative Density Outlier Detection for Large Scale Traffic Data. Electronic Imaging. 2018;30(9):239–1–239–1. https://library.imaging.org/ei/articles/30/9/art00010.
103. Liu Z, Ning J, Cao Y, Wei Y, Zhang Z, Lin S, et al. Video Swin transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022. p.3202–3211.
104. Lu, J., Meng, Y., Timmermans, H., & Zhang, A. (2021). Modelling hesitancy in airport choice: A comparison of discrete choice and machine learning methods. Transportation Research Part A: Policy and Practice, 146, 102-117.
105. Lu X, Ji J, Xing Z, Miao Q (2021). Attention and feature fusion SSD for remote sensing object detection. IEEE Trans Instrum Meas 70.
106. M. Philp, S. Fu, A review of chemical "spot" tests: a presumptive illicit drug identification technique, Drug Test. Anal 10 (2018) 95–108, https://doi.org/10.1002/dta.2300.
107. Ma C, Sun L, Zhong Z, Huo Q (2021). ReLaText: exploiting visual relationships for arbitrary-shaped scene text detection with graph convolutional networks. Pattern Recogn 111:107684.
108. Madhusudhan, P.; Latha, M.M. Ion Mobility Spectrometry for the detection of explosives. Int. J. Eng. Res. Technol. 2013, 2, 1369–1372.

109. MARTINEZ-MARTIN, Ester; DEL POBIL, Angel P. Object detection and recognition for assistive robots: Experimentation and implementation. IEEE Robotics & Automation Magazine, 2017, 24.3: 123-138.
110. Md Golam Morshed,Tangina Sultana,Aftab Alam andYoung-Koo Lee, Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities.
111. MEER, Peter. Stochastic image pyramids. Computer Vision, Graphics, and Image Processing, 1989, 45.3: 269-294.
112. Muhammad Haseeb Arshad,Muhammad Bilal andAbdullah Gani, Human Activity Recognition: Review, Taxonomy and Open Challenges.
113. MUMUNI, Alhassan; MUMUNI, Fuseini. Data augmentation: A comprehensive survey of modern approaches. Array, 2022, 16: 100258.
114. N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 1, ISSN1063-6919, 2005-06, pp.886–893.
115. Naiemi F, Ghods V, Khalesi H (2021, Elsevier Ltd) A novel pipeline framework for multi oriented scene text image detection and recognition. Expert Syst Appl 170(2020):114549.
116. NAMATĒVS, Ivars. Deep convolutional neural networks: Structure, feature extraction and training. Information Technology and Management Science, 2017, 20.1: 40-47.
117. Nguyen QP, Lim KW, Divakaran DM, Low KH, Chan MC. GEE: A gradient-based explainable variational autoencoder for network anomaly detection. In: 2019 IEEE Conference on Communications and Network Security (CNS). IEEE; 2019. p.91–99.
118. NING, Chengcheng, et al. Inception single shot multibox detector for object detection. In: 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2017. p. 549-554.
119. Niu Z, Yu K, Wu X. LSTM-based VAE-GAN for time-series anomaly detection. Sensors. 2020; 20(13):3738.
120. O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
121. Othman, N.A., Aydin, I. (2021). Challenges and limitations in human action recognition on unmanned aerial vehicles: A comprehensive survey.
122. Ozaki, Y.; Morisawa, Y. Principles and Characteristics of NIR spectroscopy. In Near-Infrared Spectroscopy, 1st ed.; Ozaki, Y., Huck, C., Tsuchikawa, S., Engelsen, S.B., Eds.; Springer: Singapore, 2021; pp. 11–36. ISBN 978-981-15-8647-7.
123. OZTURK, Ozan; SARITÜRK, Batuhan; SEKER, Dursun Zafer. Comparison of fully convolutional networks (FCN) and U-Net for road segmentation from high resolution imageries. International journal of environment and geoinformatics, 2020, 7.3: 272-279.
124. Qian R, Lai X, Li X (2021) 3D object detection for autonomous driving: A Survey 14(8), 1–24, [Online].
125. Qiong H., Lei Q., Qingming H. 2013.Overview of Human Action Recognition Based on Vision. Chinese Journal of Computers,12(12),2512-2524.
126. Quilty, B., Clifford, S., Flasche, S., & Eggo, R. (2020). Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. Retrieved from Eurosurveillance
127. P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, IEEE Comput. Soc., 2001, pp. I-511–I-518.
128. Pareek, P., Thakkar, A. A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications.
129. Patole, S. M., Torlak, M., Wang, D., & Ali, M. (2017). Automotive radars: A review of signal processing techniques. IEEE Signal Processing Magazine, 34(2), 22-35.
130. Patrick McGee, “How the commercial drone market became big business,” The Financial Times Limited 2020, https://www.ft.com/content/cbd0d81a-0d40-11ea-bb52-34c8d9dc6d84.
131. Piciarelli C, Micheloni C, Foresti GL. Trajectory-Based Anomalous Event Detection. IEEE Transactions on Circuits and Systems for Video Technology. 2008;18(11):1544–1554.

132. Proceedings of the AAAI Conference on Artificial Intelligence, 34(07):12484–12491, Apr. 2020. doi: 10.1609/aaai.v34i07.6936
133. Progress and Future Research Directions. Arch Computat Methods Eng.
134. R.A. Crocombe, *Appl. Spectrosc.*, 72(12), 1701–1751 (2018). DOI: 10.1177/0003702818809719.
135. R.F. Kranenburg, A.R. García-Cicourel, C. Kukurin, H.-G. Janssen, P. J. Schoenmakers, A.C. van Asten, Distinguishing drug isomers in the forensic laboratory: GC-VUV in addition to GC-MS for orthogonal selectivity and the use of library match scores as a new source of information, Forensic Sci. Int. (2019) 109900, https://doi.org/10.1016/j.forsciint.2019.109900.
136. Ratcliffe, J. (2018, January 2). Upgrade or Repair? Assessing an Old CCTV System. CCTV.co.uk. Retrieved June 4, 2024, from https://www.cctv.co.uk
137. Roy PR, Bilodeau GA. Road user abnormal trajectory detection using a deep autoencoder. In: Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13. Springer; 2018. p. 748–757.
138. S. Kalamkar et al. Multimodal image fusion: A systematic review. Decision Analytics Journal, p. 100327, 2023.
139. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang. Restormer: Efficient transformer for high-resolution image restoration, 2022.
140. Saleem, G., Bajwa, U.I. & Raza, R.H. Toward human activity recognition: a survey. Neural Comput & Applic.
141. Santhosh KK, Dogra DP, Roy PP. Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey. ACM Comput Surv. 2020 dec;53(6). https://doi.org/10.1145/3417989.
142. Santhosh KK, Dogra DP, Roy PP, Mitra A. Vehicular trajectory classification and traffic anomaly detection in videos using a hybrid CNN-VAE Architecture. IEEE Transactions on Intelligent Transportation Systems. 2021; 23(8):11891–11902.
143. Sarker MI, Losada-Gutiérrez C, Marron-Romera M, Fuentes-Jiménez D, Luengo-Sánchez S. Semi-supervised anomaly detection in video-surveillance scenes in the wild. Sensors. 2021;21(12):3993.
144. Shuchang Zhou, Computer Vision and Pattern Recognition, A Survey on Human Action Recognition.
145. SHUVO, Md Maruf Hossain, et al. Efficient acceleration of deep learning inference on resource-constrained edge devices: A review. Proceedings of the IEEE, 2022, 111.1: 42-91.
146. Simon, Denman., Tristan, Kleinschmidt., David, Ryan., Paul, Barnes., Sridha, Sridharan., Clinton, Fookes. (2015). Automatic surveillance in transportation hubs. Expert Systems with Applications. doi: 10.1016/J.ESWA.2015.08.001
147. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.
148. Singh, P.K., Kundu, S., Adhikary, T. et al. Progress of Human Action Recognition Research in the Last Ten Years: A Comprehensive Survey. Arch Computat Methods Eng.
149. Singh P, Pankajakshan V. A deep learning based technique for anomaly detection in surveillance videos. In: 2018 Twenty Fourth National Conference on Communications (NCC). IEEE; 2018. p. 1–6.
150. Sizhe An and Umit Y. Ogras. 2021. MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare. ACM Trans. Embedd. Comput. Syst. 1, 1, Article 1 (January 2021), 22 pages. Fig. 3, p. 9. https://doi.org/10.1145/3477003
151. Small, G.W. Chemometrics and Near-Infrared Spectroscopy: Avoiding the pitfalls. TrAC 2006, 25, 1057–1066.
152. Sorama. (n.d.). Acoustic Cameras and Sound Imaging. Retrieved from Sorama Portal
153. Sultani W, Chen C, Shah M. Real-world anomaly detection in surveillance videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 6479–6488.
154. T. Acharya and A. Ray. Image Processing: Principles and Applications. Wiley, 2005.
155. T. Lin, B. Park, H. Bannazadeh, and A. Leon-Garcia, ‘‘Demo abstract: End-to-end orchestration across SDI smart edges,’’ in Proc. IEEE/ACM Symp. Edge Comput. (SEC), Oct. 2016, pp. 127–128.
156. T. Zou and L. Chen. Ladlenet: Translating thermal infrared images to visible light images using a scalable two-stage u-net, 2023.

157. Tamilselvi M, Karthikeyan S (2022, Elsevier) An ingenious face recognition system based on HRPSM_CNN under unrestrained environmental condition. Alexandria Eng J 61(6):4307–4321.
158. Tian Y, Pang G, Chen Y, Singh R, Verjans JW, Carneiro G. Weakly-Supervised Video Anomaly Detection with Robust Temporal Feature Magnitude Learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021. p. 4975–4986.
159. Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 4489–4497.
160. Uddin, M. Z., & Nilsson, E. G. (2020). Emotion recognition using speech and neural structured learning to facilitate edge intelligence. Engineering Applications of Artificial Intelligence, 94, 103789.
161. Vector Security Networks. (n.d.). The Evolution of Closed-Circuit Television (CCTV) Systems. Retrieved June 4, 2024, from https://www.vectorsecuritynetworks.com
162. Vermesan, O., & Friess, P. (Eds.). (2014). Internet of Things: From Research and Innovation to Market Deployment.
163. W. Tang, F. He, Y. Liu, Y. Duan, and T. Si. Datfuse: Infrared and visible image fusion via dual attention transformer. IEEE Transactions on Circuits and Systems for Video Technology, 2023.
164. W. Zhao, S. Xie, F. Zhao, Y. He, and H. Lu. Metafusion: Infrared and visible image fusion via metafeature embedding from object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13955–13965, 2023.
165. Wang L, Xiong Y, Wang Z, Qiao Y, Lin D, Tang X, et al. Temporal segment networks: Towards good practices for deep action recognition. In: European conference on computer vision. Springer; 2016. p. 20–36.
166. Wang Y, Li K, Li X, Yu J, He Y, Chen G, et al. InternVideo2: Scaling Video Foundation Models for Multimodal Video Understanding. arXiv preprint arXiv:240315377. 2024.
167. Wu JC, Hsieh HY, Chen DJ, Fuh CS, Liu TL. Self-supervised Sparse Representation for Video Anomaly Detection. In: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, editors. Computer Vision – ECCV 2022. Cham: Springer Nature Switzerland; 2022. p. 729–745.
168. X. Chu, L. Chen, , C. Chen, and X. Lu. Improving image restoration by revisiting global information aggregation. arXiv preprint arXiv:2112.04491, 2021.
169. Xu, X.; Dong, S.; Xu, T.; Ding, L.; Wang, J.; Jiang, P.; Song, L.; Li, J. FusionRCNN: LiDAR-Camera Fusion for Two-Stage 3D Object Detection. Remote Sens. 2023, 15, 1839.
170. Yao, Q., Huang, Y., & Wang, X. (2020). Integrating AI into CCTV Systems: A Comprehensive Evaluation of Smart Video Surveillance in Community Space. arXiv preprint arXiv:2312.02078.
171. Y. Li, Y. Zhang, R. Timofte, L. Van Gool, Z. Tu, K. Du, H. Wang, H. Chen, W. Li, X. Wang, et al. Ntire 2023 challenge on image denoising: Methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1904–1920, 2023.
172. Yu, W., Liang, F., He, X., Hatcher, W.G., Lu, C., Lin, J., Yang, X. A Survey on the Edge Computing for the Internet of Things, Department of Computer and Information Sciences, Towson University, MD, USA, School of Electronic and Information Engineering, Xi'an Jiaotong University, Shaanxi, P.R. China.
173. Yu, W., Liang, F., He, X., Hatcher, W. G., Lu, C., Lin, J., & Yang, X. (2017). A survey on the edge computing for the Internet of Things. IEEE access, 6, 6900-6919.
174. Z. Huang, J. Liu, X. Fan, R. Liu, W. Zhong, and Z. Luo. Reconet: Recurrent correction network for fast and efficient multi-modality image fusion. In European Conference on Computer Vision, pp. 539–555. Springer, 2022.
175. Z. Zhao, H. Bai, J. Zhang, Y. Zhang, S. Xu, Z. Lin, R. Timofte, and L. Van Gool. Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5906–5916, June 2023.
176. Zehua Sun, Qiuhong Ke, Hossein Rahmani, Mohammed Bennamoun, Gang Wang, Jun Liu, Human Action Recognition from Various Data Modalities: A Review, Computer Vision and Pattern Recognition.
177. Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. Precision Agriculture, 13(6), 693-712.

178. Zhang, L., & Yang, J. (2021). A Continuous Liveness Detection for Voice Authentication on Smart Devices. arXiv preprint arXiv:2106.00859v1.
179. Zhang, X., Wu, H., Wu, M., & Wu, C. (2020). Extended Motion Diffusion-Based Change Detection for Airport Ground Surveillance. IEEE. Retrieved from IEEE Xplore.
180. Zhang, Y., Arora, S. S., Shirvanian, M., Huang, J., & Gu, G. (2021). Practical Speech Re-use Prevention in Voice-driven Services. arXiv preprint arXiv:2101.04773v1.
181. Zhaole D., Kang W., Shenglong L. 2021. Human action recognition based on deep learning. Command Information System and Technology,12(04),70-74.
182. ZHAO, Zhong-Qiu, et al. Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems, 2019, 30.11: 3212-3232.
183. Zhou JT, Du J, Zhu H, Peng X, Liu Y, Goh RSM. AnomalyNet: An Anomaly Detection Network for Video Surveillance. IEEE Transactions on Information Forensics and Security. 2019 10;14(10):2537–2550.
184. Zhou Y, Liang X, Zhang W, Zhang L, Song X. VAE-based deep SVDD for anomaly detection. Neurocomputing. 2021; 453:131–140.
185. https://www.fiware.org/
186. https://aiperspectives.springeropen.com/articles/10.1186/s42467-021-00012-z
187. https://ieeexplore.ieee.org/document/9465137
188. https://dronedj.com/2019/01/28/drones-reporting-for-work-goldman-sachs/
189. https://www.ti.com/sensors/mmwave-radar/products.html
190. https://dev.ti.com/tirex/explore/node?node=A__AMKSv8im74YjT7cmO9jVHg__com.ti.mmwave_industrial_toolbox__VLyFKFf__4.9.0
191. https://www.arrowhead.eu/.
192. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TG02102/BSI-TR-02102-1.pdf?__blob=publicationFile
193. https://news.vmware.com/releases/vmware-report-warns-of-deepfake-attacks-and-cyber-extortion
194. https://ubertooth.sourceforge.net
195. https://infocondb.org/con/black-hat/black-hat-usa-2016/
196. https://defcon.org/html/defcon-24/dc-24-venue.html
197. https://www.mdpi.com/2076-3417/12/4/2168
198. https://us.metoree.com/categories/2759/
199. http://en.nanoradar.cn/Article/detail/id/289.html
200. https://www.novelic.com/norasens-radar-sensor-technology
201. https://www.mordorintelligence.com/industry-reports/artificial-intelligence-in-construction-market
202. https://www.equipmentworld.com/business/article/15304920/construction-worker-death-rate-declines-22
203. https://www.buildevolution.be/
204. https://towereye.be/en
205. http://www.timelapses.be/
206. https://www.visuatech.be/
207. https://www.viewandintegrate.be/
208. http://www.elapse.be/
209. https://www.aicon.construction/
210. https://www.oxblue.com/
211. https://www.earthcam.com/
212. https://www.senserasystems.com/
213. https://evercam.io/
214. https://www.devisubox.com/
215. https://bau.camera/
216. https://enlaps.io/fr/
217. https://photosentinel.com/

218. https://cam-do.com/
219. https://www.spotr.ai/
220. https://bouwcam.live/
221. https://boxcam.fr/
222. https://www.newmetrix.com/
223. https://www.procore.com/en-gb
224. https://doxel.ai/
225. https://cad42.com/
226. https://www.airsquire.ai/
227. https://www.pixelvision.ai/
228. https://www.avvir.io/
229. https://www.openspace.ai/
230. https://pillar.tech/
231. https://www.pix4d.com/
232. https://enterprise.dji.com/phantom-4-rtk
233. https://enterprise.dji.com/mavic-3-enterprise
234. https://www.easa.europa.eu/en/domains/civil-drones/drones-regulatory-framework-background/open-category-civil-drones
235. https://enterprise.dji.com/dock
236. https://enterprise.dji.com/matrice-30
237. https://percepto.co/drone-in-a-box/percepto-base/
238. https://www.dronematrix.eu/product
239. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/iot-value-set-to-accelerate-through-2030-where-and-how-to-capture-it
240. https://www.grandviewresearch.com/industry-analysis/industrial-internet-of-things-iiot-market
241. https://www.marketsandmarkets.com/Market-Reports/intelligent-video-analytics-market-778.html
242. https://dronedj.com/2019/01/28/drones-reporting-for-work-goldman-sachs/