REPORT

D5.5 Device Selection Report

Title of the work package: AI driven buildings

Responsible partner: Optima Ideas

Partners: IT4I@VSB, CVTI, CANEX, Asseco CE

Version: 3.0
Date: November 2025

Author/Authors: Richard Hollý
Confidentiality: Consortium

www.innovaite.sk

Financované Európskou úniou
NextGenerationEU

PLÁN OBNOVY

ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU

VAIA VÝSKUMNÁ A INOVAČNÁ AUTORITA

Content

List of Figures

List of Tables

Revision history			
Version	Date	Comment	Author
v1	02.10.2025	First draft	Optima Ideas - Richard Hollý and team
v2	17.11.2025	Second draft	Optima Ideas – Richard Hollý and team
v3	24.11.2025	Third draft	Optima Ideas – Richard Hollý and team
v 4 final	27.11.2025	Final version	Optima Ideas – Richard Hollý and team

1 Executive summary

This document forms part of Deliverable D5.5 within Work Package 5 (AI-Video driven Building and Road Safety: Guarding the Future) of the InnovAIte project. The purpose of this deliverable is to provide a comprehensive overview of the devices for Edge AI devices in Smart Buildings regarding typical use cases in this area.

The amount of computing power available in edge devices limits the possibilities of AI software solutions that can run on them. The report describes key aspects of finding a compromise between software requirements and hardware capabilities. It also describes trends in AI software and, on the other hand, in hardware development, especially edge devices.

The results of research focused on understanding and selecting suitable edge AI devices are important for further research tasks. They provide objective input on which edge devices should be used in the context of the latest trends and possible alternatives. They also indicate their appropriate placement in the solution architecture. The recommended edge AI device is Jetson AGX Orin as a highly capable embedded AI compute module built on the NVIDIA Ampere architecture.

This device will be primarily used in conjunction with the AInalytics platform of our partner Optima Ideas to solve other research tasks.

2 Introduction

The research and analysis presented in this document were conducted as part of the InnovAIte project, within deliverable D5.5: Device Selection Report in Work Package 5 (AI-Video driven Building and Road Safety: Guarding the Future), coordinated by the IT4Innovations and VSB-TU Ostrava, in collaboration with project partners, Optima Ideas, CANEX, Asseco Central Europe, CVTI and DFKI. Optima Ideas is leading the preparation of the document, with all consortium partners collaborating on its creation.

The aim of this initial phase is to analyse the needs of building owners, facility managers, energy providers and end users to enhance safety and security in buildings through AI-based video intelligence with local processing. Target outcomes include faster reaction times, improved energy efficiency and stronger data protection. Algorithms and tools for object detection, human activity recognition and anomaly detection will be designed and tested. A common element is the use of synthetic image data from virtual 3D scenes to train on complex scenarios that are hard or impossible to recreate; the supercomputing centre provides the capacity to generate and process these datasets.

This report defines the strategic orientation and methodological framework for WP5, together with the devices and tools used in the subsequent research and validation phase. Building on these goals, WP5 adopts an edge-centric video intelligence pipeline for multi-camera smart-building environments. At the acquisition layer, heterogeneous streams (codecs, resolutions, frame rates) are normalised and distributed via a lightweight framework with time synchronisation and topology-aware routing. Above this transport layer, multi-view fusion—geometric calibration and image stitching—forms a global scene view for downstream analytics, enabling consistent region-of-interest management across overlapping fields of view and better robustness to occlusions.

The analytical core relies on convolutional neural networks, with real-time YOLO-family detectors as the primary workhorse for object detection and classification under varying conditions. Detected trajectories are maintained by online multi-object tracking to support counting, occupancy and path analysis. For activity recognition, short temporal windows of per-frame CNN features capture motion cues while preserving the low latency required on edge devices. Anomaly and intrusion detection use reconstruction/prediction-based models and lightweight one-class learners operating on scene-level indicators (flow, density, dwell anomalies) to flag deviations from learned normal patterns. Synthetic video data generated from virtual 3D scenes—using parameterised layouts, lighting, sensor noise and actor behaviours with domain randomisation—complements pilot recordings and supports repeatable benchmarks on high-performance computing resources.

Device selection and software choices are driven by measurable edge hardware constraints. For representative use cases, we profile end-to-end latency, throughput (FPS), accuracy and power consumption across candidate devices and runtime stacks (e.g., optimised CNN inference with hardware accelerators). The report compiles (i) a comparative overview of suitable edge platforms for building scenarios, (ii) recommended deployment architectures for single-site and multi-site roll-outs (ingest, fusion, analytics, secure backhaul), and (iii) data-protection procedures for on-device processing, retention and encrypted transport. Together, these elements form the foundation for the research and validation phases, culminating in algorithm suites, device recommendations and an integrated prototype demonstrator for smart-building safety and security.

3 Project Context and Objectives of WP5

The ***InnovAIte Slovakia: Illuminating Pathways for AI-Driven Breakthroughs*** project (TIC call 09I02-03-V2) is funded by the **Recovery and Resilience Plan – NextGenerationEU**. It lasts 30 months (M1 → M30) and brings together 11 partners from academia, the private and public sector, led by Gratex International, a.s. The main objective is to *"advance the state of knowledge in the field of artificial intelligence applications across multiple domains and deliver measurable innovations that will support economic growth, sustainability, and quality of life during the project implementation period."*

This document focuses on describing suitable edge devices in the context of WP5 research objectives in the field of buildings. The content of the report therefore fulfils the objectives of O5.4 goal of work package 5. To provide context, all work package objectives are listed below:

O5.1: Generate high-quality synthetic image/video data through rendered 3D scenes.
O5.2: Develop an AI-based algorithm that can increase the size and variability of synthetically generated datasets.
O5.3: Develop an AI-based algorithm to detect critical phenomena in road traffic from camera images.
O5.4: Evaluate state-of-the-art devices suitable for edge vision in smart buildings with an emphasis on their security and data protection capabilities.
O5.5: Develop and optimize Edge AI algorithms for object detection, human activity recognition, and anomaly detection in time series data, tailored to the safety and security contexts of smart buildings.

Through the realization of these objectives, WP5 ensures that the InnovAIte project advances the security and safety of buildings, fostering a more intelligent, resilient, and sustainable built environment in Europe.

The implementation of the research task of selecting suitable edge AI devices fulfills objective O5.4. This involves creating outputs/selections for two purposes. One is operational, namely the need to deploy real devices in the field as part of the research, collect data, and conduct research. The second is to update business opportunities as the field develops. This area will certainly need to be updated once again within the research project, namely within WP8.

These objectives will be achieved by fulfilling the tasks listed below.

Table 1: Tasks in WP05

Code	Title	Month (M)
T5.1	Specification for the 3D scene generation	M1-M4
T5.2	AI methods for extending variability of generated datasets	M1-M6
T5.3	Development of the 3D virtual environments, generation of synthetic datasets	M5-M24
T5.4	Development of AI methods for extending dataset variability	M7-M24
T5.5	Development of AI algorithms for increasing traffic safety from camera imagery	M1-M24
T5.6	Optimization of AI models and algorithms for multi-source video stream processing adapted to traffic and smart building domain	M1-M24

The task of selecting suitable edge devices, which is the subject of this deliverable, is related to the solution of several WP5 tasks. In order to collect and research the possibilities and limits of data evaluation, especially in the subdomain of buildings, it is necessary to have the appropriate research infrastructure. In this case, edge AI devices. The selection was influenced by the direction of the research subject. Their solution was formed and will be further refined **in tasks T5.1 and T5.3.**

The report fulfills the requirements for deliverables 5.5, which are listed below in the context of other expected deliverables of the work package.

Table 2: Deliverables in WP05

Code	Name	Due date (M)
D5.1	Tools and methods for synthetic data generation	M8
D5.2	Frameworks for synthetic data generation	M24
D5.3	Algorithms, tools and methods for increasing traffic safety	M24
D5.4	Journal publication - multi-source processing model for efficient AI evaluation	M24
D5.5	**Device Selection Report**	**M8**
D5.6	Safety Algorithm Suite and Data Protection Protocol	M24
D5.7	Anomaly Detection and Prediction Tool for Time Series	M24

4 Updated detail scope definition

Approach/methodology for solving the task

The following approach was chosen to address the task of selecting suitable edge devices and architecture:

1. In the first phase, an update on trends in video technology was implemented.
2. Definition of the set of use cases for which the research task
3. Overview of edge devices and their suitability for AI-based video processing was carried out, including the establishment of selection criteria for edge devices from the perspective of the needs of the research tasks being launched.
4. Market and Technology Review
5. Suitable equipment was selected
6. Next steps were determined.

A specific aspect of the task was that the team had to place great emphasis on the ability to take ownership of the chosen solution and be able to operate it for data collection and research tasks. The ability to operate their own research platform thus became key.

The approach described above ensured that all currently available innovative equipment was taken into consideration. In addition, the expected tasks to be solved were defined. Emphasis is placed on ensuring that the selected equipment allows for solving a variety of research tasks.

5 Description of Deliverables achievement

The report describes the method of selecting suitable edge AI devices for research tasks. The leader of the entire activity is Optima Ideas, which drew on its many years of experience in the field when developing the solution. The experience of project partner CANEX, which also has extensive experience in this area, was also utilized in developing the solution. These two partners have a leading position in WP5 in the area of HW deployment and utilization.

Other partners involved in WP5 also collaborated on the creation of the output. All partners were involved in the implementation of T5.1 and T5.3: IT4Innovations and VSB-TU Ostrava, Asseco Central Europe, CVTI, and DFKI. Their involvement was not directly related to HW research, but to related areas within T5.1 and T5.3.

The necessary mutual coordination took place at regular weekly WP5 status meetings. The individual activities carried out to achieve the output were supervised by The following activities were carried out as part of the research task:

- Research activities in the field of HW
- Research activities in the field of SW
- Analytical activities
- Defining a suitable subject for research tasks in terms of safety and security
- Defining a suitable subject for research tasks in terms of WP4 needs
- Classification of buildings from the perspective of research task needs
- Laboratory description and testing of selected devices
- HW openness for implementation and research with the AInalytics platform
- Possible architectures with the inclusion of edge AI devices
- The relationship between HW and SW edge devices and the need to use artificial data for tool training

6 Deliverable

6.1 Basic update on trends in video technology

The integration of Artificial Intelligence (AI) into video surveillance systems is fundamentally transforming building security, moving away from reactive recording to proactive and predictive risk management in real-time. [1]

6.1.1 Current trends in the use of AI and video technology in building security and the scope of the research project

Advanced Deep Learning (DL) models, primarily based on Convolutional Neural Networks (CNNs) and recurrent architectures, are essential for executing critical safety functions:

Object Detection - To immediately identify and track critical entities such as people, unauthorized vehicles, or sensitive equipment in the video stream. This capability is fundamental for implementing real-time intrusion detection and access control (e.g., biometric verification). Systems utilize optimized Convolutional Neural Networks (CNNs). The You Only Look Once (YOLO) architecture is widely recognized in academic literature as the industry standard, favoured for its speed and accuracy, which is necessary for processing video feeds directly on edge devices. [2]

Anomaly Detection - To automatically identify unusual, suspicious, or dangerous behaviours (e.g., unauthorized lingering, or abandoned objects) by detecting deviations from established 'normal' activity patterns, thereby serving as an early warning system. This application heavily relies on Generative Models and Recurrent Neural Networks, such as Autoencoders and Recurrent Autoencoders. These models are trained exclusively on normal data, and an event is flagged as an anomaly if the system cannot accurately reconstruct or predict the observed video sequence. [3]

Human Activity Recognition (HAR) - To synthesize and understand complex human actions over a period (not just a single pose) to classify non-standard behaviours that require intervention, such as fights or a potential fall. HAR systems combine spatial feature extraction (via CNNs) with temporal sequence analysis (via 3D CNNs or Recurrent Units) to successfully process and classify actions across a continuous flow of video frames.

6.1.2 Architectural Imperatives: Privacy and Ethics

The widespread adoption of AI video technology is shaped by non-functional requirements related to ethics and data protection. These imperatives accelerate the shift toward Edge computing architectures. [4]

Privacy and Data Protection: Centralized surveillance systems storing raw video data present high risks for data leakage and unauthorized access. The primary architectural solution is to process data at the source on Edge devices. This strategy minimizes risks by ensuring sensitive raw data never leaves the local hardware.

Ethical Deployment and Integrity: A key ethical risk is the potential for AI algorithms to inherit and amplify bias from real-world training data. Models are also vulnerable to sophisticated cyberattacks like data poisoning. The InnovAIte project addresses both ethical and integrity challenges by focusing on synthetic video data generation from virtual 3D scenes. This method allows researchers to safely simulate complex, rare, or sensitive scenarios, providing a controlled environment for algorithm training while mitigating the risks associated with collecting and using sensitive real-world data.

6.1.3 Current trends in the development of HW edge device capabilities in the field of video technology

The successful shift of complex video analytics from cloud-based servers to local Edge devices, such as embedded cameras, gateways, and single-board computers (SBCs) is entirely dependent on advancements in hardware acceleration and power efficiency. The primary trend is the ubiquitous integration of specialized accelerators designed to execute Deep Neural Network (DNN) inference in a resource-constrained, low-power environment.

Evolution of Edge AI Accelerators

Current hardware development is defined by heterogeneous computing, where general-purpose processors are augmented by dedicated hardware for AI acceleration.

Dataflow Architectures: Most commercial Edge processors today utilize dataflow architectures, such as Graphics Processing Units (GPUs), Neural Processing Units (NPUs), and Vision Processing Units (VPUs). These are optimized for the massive parallel operations required by neural networks used in video analytics. [5]

High-Performance Embedded GPUs (e.g., NVIDIA Jetson): Systems like the NVIDIA Jetson Orin family provide substantial computing power for complex, multimodal AI inference at the Edge. These platforms are often used in demanding applications, such as processing multiple high-

resolution video streams or running large transformer models, and are crucial for real-time video analytics that demand high throughput and low latency.

Figure 1: NVIDIA Jetson Orin nano

Specialized Vision Processors (e.g., Intel Movidius, NXP i.MX): Dedicated Vision Processing Units (VPUs), such as those from Intel Movidius, are designed specifically for low-power, small-form-factor integration into cameras and network video recorders (NVRs). Similarly, application processors like the NXP i.MX series integrate high-performance CPUs (Arm® Cortex®-A53/A72/A78AE) alongside a dedicated NPU or GPU optimized for machine learning inference, which allows for advanced computer vision tasks while maintaining high power efficiency.

Figure 2: Intel Movidius

Figure 3: NXP i.MX 8X

Energy Efficiency and Hardware-Aware Model Optimization

The practical limitations of Edge hardware, particularly concerning power and cooling, necessitate rigorous co-design between hardware and software. Since the energy consumption of AI is a rapidly growing concern, **energy efficiency**, often quantified as TOPS (Tera Operations Per Second) per Watt, has become a critical performance metric for hardware vendors. [6]
This focus on efficiency drives simultaneous innovation in two reciprocal areas: low-power chip architectures and hardware-aware model compression techniques. Hardware manufacturers are increasingly exploring **non-von Neumann architectures** to drastically reduce power consumption.

Research is progressing in **Processing In-Memory (PIM)** processors, which minimize the energy-intensive data transfer between the processor and external memory, thereby bypassing the traditional von Neumann bottleneck.
The hardware limitations directly impose constraints on the size and complexity of the AI models that can be deployed. This mandates a trend towards **hardware-aware compression** to reduce the computational and memory load of Deep Learning models. Key optimization techniques include **pruning** (removing non-essential connections or channels from the network architecture) and **quantization** (reducing the numerical precision of the data - e.g., from 32-bit floating point to 8-bit integer). [7]

6.1.4 Current developments in the architecture of video technology implementation with integrated AI tools

The basic direction of physical and application architecture in the field of video technology implementation follows from the context of the facts presented in the previous two chapters.

This chapter only briefly summarizes everything.

- The prevailing architectural trend is a mandatory shift from monolithic, centralized Video Management Systems (VMS) to a distributed Edge-Fog-Cloud model.
- This trend is driven by the critical need for enhanced data security and low latency in real-time threat response.

- **The architecture utilizes a tiered hierarchy:**
  - **Edge** devices perform immediate, low-latency inference (e.g., using optimized YOLO, ViT or DETR models),
  - while a middle **Fog** layer handles resource-intensive integration tasks. This Fog layer is essential for synchronizing and fusing multiple heterogeneous video streams to create a comprehensive "global scene perspective" for advanced analysis.
  - **Cloud** - High-Performance Computing environments are reserved solely for model training and synthetic data generation.
  - This strategy effectively balances computational load across the network to maximize responsiveness and integrity. [8]

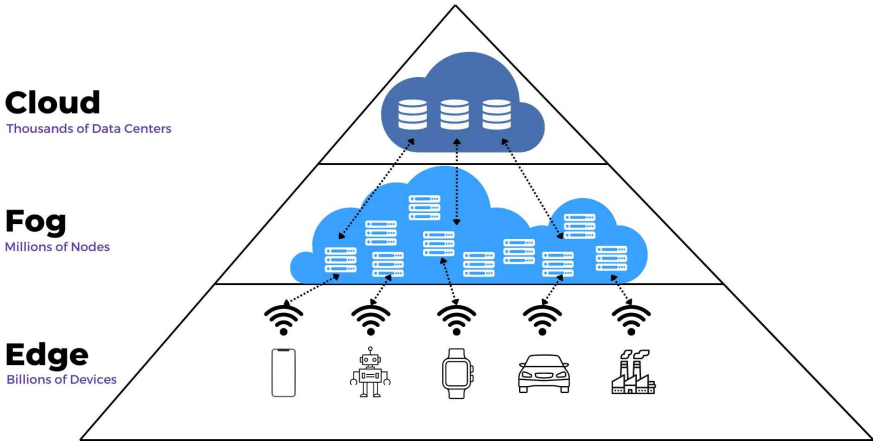

Figure 4: Edge-Fog-Cloud model

6.2 Overview of use cases and description of research tasks

In this chapter, we will be running through the possible use cases implied by the use of AI-video analysis. We will solely focus on cases with regards to monitoring areas in the building sector.

The text outlines how recent advancements in AI-driven video analytics—particularly Edge-AI—enable a wide range of safety, security, operational, and energy-efficiency applications across indoor and outdoor building environments. It highlights use cases such as people- and equipment-monitoring, anomaly detection, occupancy and queue management, healthcare-related monitoring, vehicle and parking control, and privacy-preserving analytics.

Furthermore, it introduces a classification of building types to illustrate how specific operational requirements shape the deployment of video technologies and distinguishes between general and specialized applications for both exterior and interior camera systems.

6.2.1 Overview of use cases for video technology with AI tools in the building sector

The newest innovations within the field of AI and video processing are central to the broader digital transformation of the built environment, where AI video tools ensure safety, security, operational intelligence and also support sustainable and efficient facility management. The field of Edge-AI video analytics, in particular, allows the processing to occur locally at cameras or edge devices, where in turn the latency is reduced, privacy is enhanced, and the reliance on cloud resources is minimized. The integration of artificial intelligence in video technology enables a rich set of use cases that apply to both indoor and outdoor building environments.

People & crowd–related safety/security

Perimeter and lobby intrusion / presence detection. Edge video analytics can identify unauthorized entry into restricted zones. Systems analyse real-time video feeds and raise alerts when individuals cross predefined boundaries. Such AI-enabled intrusion detection enhances perimeter security and allows immediate response without constant human observation. [12]

Loitering, tailgating, and unauthorized access patterns. Combining person detection with simple temporal rules (such as time-in-zone, two entries on one badge, etc.) raises alarms locally to inform the personnel about the potentially suspicious activity. [12]

Equipment monitoring/management. For example at construction sites or in assembly halls, by identifying the types of equipment being used, combined with other types of algorithms, it is possible to identify anomalies in equipment usage patterns, predict maintenance needs, and optimize equipment utilization. [11]

People counting & flow. Intelligent algorithms continuously monitor video streams to detect and track each person in the monitored area. This supports both security applications and operational insights, such as crowd flow in commercial spaces, counting of entries/exits to manage the volume of the crowd, estimate crowding to plan evacuation. [11]

Anomaly and Behaviour Detection. AI systems trained on behavioural data can detect unusual motion or postures, enabling rapid response to emergencies such as falls, fights, or accidents, which can significantly improve the chances of resolving the situation. [12]

Occupancy, utilization & operations

Room/zone occupancy & density for space management. Within large buildings or campuses, AI video tools measure crowd density, detect congestion, and identify potential hazards like stampedes or overcrowding. These capabilities are vital for evacuation planning, event management, and optimizing building occupancy for energy efficiency. [12]

Queue monitoring at reception/security desks. Track queue length and wait-time to trigger staff dispatch or open additional lanes, relying on the same people-detection/tracking primitives. [12]

Healthcare & assisted-living areas inside buildings

Fall detection and elder/patient safety monitoring. Smart healthcare monitoring makes it possible to continuously analyse the movement of an elderly person in his/hers home and send alerts to the remote caretaker in case of falling or other safety concerns. [9]

Behavioural-pattern alerts (wandering, inactivity, distress). Likewise in medical facilities or home environments, systems can identify abnormal inactivity or agitation (e.g., prolonged "not moving" after a fall) and notify the caregivers. [9]

Vehicles and parking

Vehicle detection near restricted areas. Detect and classify vehicles (pedestrians, bicycles, cars, motorcycles, trucks, buses) for access control at gates, at campuses, and safety around dangerous or restricted areas. [10]

Parking occupancy and violation detection. Determine stall occupancy and trigger alerts for unauthorized parking, fire-lane blocking and violation of parking in restricted or pedestrian zones. [12]

Privacy-preserving monitoring

Privacy-tuned analytics in sensitive zones. Several methods enable presence/people-counting without high-fidelity images—appropriate for rest areas, wellness rooms, or certain compliance regimes. A critical consideration in deploying AI-based video systems is privacy protection. Modern architectures increasingly integrate privacy-preserving mechanisms such as on-device anonymization, face blurring, and federated learning, ensuring compliance with regulations like the GDPR while maintaining analytics accuracy. [11]

6.2.2 Classification of buildings in terms of use and business requirements for video technology

This chapter contains descriptions of using video technologies from slightly different perspectives than the description given in the previous chapter.

The proposed classification organizes buildings based on their primary use and operational requirements. This framework helps identify specific needs for security, video surveillance, and energy management. By categorizing buildings into commercial, industrial, residential, educational, healthcare, public/cultural, and infrastructure types, it becomes easier to define appropriate solutions and optimize resources efficiently. This classification serves as a guide for strategic planning, technology deployment, and energy efficiency measures.

1. Commercial (offices, malls, hotels, coworking)
2. Residential (apartment buildings, housing complexes, gated communities)
3. Educational (schools, universities, kindergartens)
4. Healthcare (hospitals, clinics, labs)
5. Public & Cultural (museums, libraries, stadiums, theatres)
6. Infrastructure (airports, train stations, power plants, dams)
7. Industrial & Logistics (warehouses, factories, distribution centres)

Exterior Camera Use

From the perspective of exterior video technology, there are common uses that apply across all building categories, such as monitoring entrances, perimeters, parking areas, and public outdoor spaces for safety and security purposes. In addition to these shared functions, some uses are specific to certain building types. For example, industrial facilities may require monitoring of loading docks and machinery areas, healthcare buildings may focus on ambulance and patient entry zones, and public or cultural buildings may need crowd monitoring during events. This classification highlights both the general and specialized applications of exterior video technology for each building type.

List of basic general use cases for video technology outside buildings:

Security: perimeter security, security of building exterior, entrance Gate (ANPR), violent behaviour, leftover items, weapons

Safety: fire, fallen people

Monitoring: parking lot occupancy, vehicle and people monitoring, crowd monitoring

Using video technology outputs for energy management: Exterior lighting optimization, smart control of outdoor HVAC for semi-enclosed areas, energy consumption of shared outdoor utilities.

Interior Camera Use

From the perspective of interior video technology, there are common uses that apply across all building categories, such as monitoring hallways, lobbies, elevators, shared facilities, and other internal spaces for safety and security purposes. In addition to these shared functions, some uses are specific to certain building types. For example, industrial facilities may require monitoring of production lines and machinery areas, educational institutions may focus on classroom and corridor monitoring, and healthcare buildings may need surveillance in critical zones and patient care areas. This classification highlights both the general and specialized applications of interior video technology for each building type.

List of basic general use cases for video technology insight buildings:

Security: access control, monitoring of hallways, lobbies, elevators, sensitive areas, suspicious behaviour detection, violent behaviour, leftover items, weapons

Safety: fire, fallen people, emergency incidents

Monitoring: occupancy counting, movement tracking, activity monitoring in common area

Using video technology outputs for energy management: lighting optimization, HVAC control based on occupancy, energy monitoring for elevators and shared facilities, and so on.

6.3 Overview of Edge devices and their suitability for use

Building on the previous chapter—which explored the broad spectrum of AI-driven video use cases across various building types—this section shifts the focus toward the underlying technological foundations that make such applications possible. While earlier discussions highlighted *what* AI-video analytics can achieve, the following chapter concentrates on *how* these capabilities are enabled through the choice of hardware architectures, processing approaches, and deployment strategies.

AI-based video processing leverages machine learning and deep learning algorithms to interpret video content, whether in real time or through post-processing. These technologies enable a broad array of capabilities—from object detection and motion tracking to scene understanding and human activity recognition—making video feeds a rich source of actionable insights. However, due to the high data volumes and computational demands, such systems require specialized hardware to maintain reliable performance, accuracy, and responsiveness.

In the context of modern buildings and their surroundings, edge devices play a critical role in enabling efficient AI-driven video analytics. Smart cameras, embedded GPU modules, and IoT gateways can execute inference tasks directly at or near the point of data capture. This shift toward local processing reduces latency, strengthens privacy protection, minimizes dependence on cloud resources, and enhances the resilience of monitoring systems. By bringing real-time perception and first-line decision-making closer to where events actually occur, edge-enabled analytics create opportunities for faster, more context-aware applications.

To ensure these systems operate at high speed while maintaining accuracy, security, and modularity, a wide spectrum of design considerations must be addressed. As part of our research project, we examined the key aspects involved in selecting and structuring the right technologies—including hardware capabilities, software architectures, deployment models, and

integration requirements. The following chapter outlines these considerations and frames the technological landscape necessary for effective AI-video analytics within the building sector.

Detector selection (detection tool)

The first task revolves around choosing the detectors that meet real-time constraints on edge devices without sacrificing too much accuracy on critical (safety) tasks. The most notable family of detectors are YOLO or DETR alternatives, for object and people detection, which provide speed and accuracy on smaller devices.

Other backbones, such as SSD or other neural networks compromise on speed with accuracy. Another option is the use of various model sizes, but all in all it is about trade-offs and other characteristics of the system to meet the latency targets. More options include whether it is possible to employ offload-strategies for a more distributed system or to delegate some processing to the cloud and so on. [13]

Device selection

Device selection focuses on choosing and standardizing edge hardware and accelerators that can meet the requirements (latency/FPS/accuracy) while fitting power, cost, and privacy constraints. In essence, this means mapping workloads to different computational combinations. (e.g., SBCs with USB NPUs versus embedded GPU SoCs etc.), and validating that the chosen platform sustains real-time inference for the target models and resolutions. Beyond raw compute, device selection must account for the model/toolchain fit (e.g., OpenVINO, TensorRT, TFLite paths from training to deployment) and hardware decoders/I-O needed for your cameras and local actuators.[13]

System architecture

We have touched on the architecture, so the next task is to partition the workload. In order to achieve efficiency the optimal setup is to run the first-pass detection and filtering on the edge device, while only streaming the key frames and metadata upstream, where a heavier analytics is employed. The strategy is decided by elements like bandwidth, local inference speed, and filtering, multimodal analysis or other. The processing in general is split into edge, fast inference and processing and cloud, cross-camera and heavy analytics while archiving the data for review. [14]

It is also imperative to keep in mind model governance. Edge provides the possibility of OTA updates, so tracking versions and having immediate roll-backs available brings support and compliance.

Filtering

Filtering allows for efficient bandwidth control. It is not plausible to fully analyse every frame, not on the cloud or locally. Some profiling shows that accuracy falls gently with reduced frame rate when targets move slowly and with reduced resolution when the objects are large. Keyframe selection or event driven publishing/actuation is important in order to ship what matters, while discarding or down-sampling the rest. End–edge–cloud pipelines commonly process a full stream on the device, select keyframes that preserve task accuracy, and transmit detections with sparse imagery to the cloud. This approach dramatically reduces traffic without degrading the results.

Privacy

Preserving privacy is key when deploying these edge solutions into public spaces, the need to balance security with GDPR or data protection. Keeping raw frames only on-premises, forwarding only the minimum necessary evidence or face-blurring lets the system remain in accordance with the privacy enforcements and data governance.

Description of the link to synthetic video data generation tasks

Synthetic data is a common pattern used with AI development, namely because it tackles three persistent constraints at once: privacy, rarity of safety incidents, and the need for continual model improvement.

From a model-development perspective, video analytics is inherently iterative—new data is repeatedly acquired, models are retrained/evaluated, and the cycle repeats. However, many high-risk events that matter for building safety (falls, intrusions, violence, stampedes) are rare in the wild; collecting balanced, labelled footage raises ethical and legal issues. Reviews of edge video analytics for public safety explicitly note the importance of detecting “rarely occurring activities,” which strengthens the argument for synthetic scenario generation to fill these class gaps and stress-test pipelines. Synthetic data can be rendered to match building layouts, camera viewpoints, illumination cycles, and occlusion patterns, creating controlled coverage of edge cases that are underrepresented in operational feeds. For example the usage of synthetic face generators in order to avoid privacy infringements, or the use of avatars to synthesize required situations without the need to use actual real actors. [12][15][16]

Finally, synthetic generation coincides with the privacy-preserving methods and distributed learning patterns favoured at the edge. When paired with on-device processing, encryption, anonymization, and privacy, synthetic video clips can replace or supplement sensitive footage in data pipelines and analytics queries. Synthetic video data supports safer, faster iteration on building safety use cases—training detectors and re-identification models, rehearsing rare incident responses, and validating end-to-end edge pipelines. [12][15]

6.3.1 Overview of hardware requirements for AI-based video processing

Computational power (GPU, CPU, edge AI devices)

Central Processing Unit CPUs

The Central Processing Unit (CPU) serves as the primary controller of the system, managing overall coordination between hardware components, scheduling tasks, and handling general-purpose computing operations. In AI-based video processing systems, the CPU oversees data flow between cameras, memory, storage, and AI accelerators, ensuring synchronization and system stability.

To ensure smooth performance and real-time responsiveness, multi-core processors with higher clock speeds are recommended. In embedded or edge systems, energy-efficient ARM-based CPUs (such as those found in NVIDIA Jetson platforms) are often used to balance computational performance with low power consumption.

In modern architectures, many video-specific workloads, including encoding, decoding, are increasingly offloaded to the GPU or to dedicated media processing units integrated within the hardware. Examples include NVIDIA NVENC/NVDEC, Intel Quick Sync Video, and AMD Video Core Next (VCN), which provide efficient hardware acceleration for video compression and decompression. Offloading these tasks reduces CPU utilization and enables smoother real-time performance.

Graphics Processing Unit (GPU)

The Graphics Processing Unit (GPU) serves as the primary computational engine for AI inference.

Modern neural networks, based on transformers such as RT-DETR, RF-DETR, or YOLO family. Also older ones such as ResNet, and MobileNet, perform a large number of parallel mathematical operations, something GPUs are specifically designed to handle efficiently. As already mentioned in the previous section, GPUs not only accelerate AI model inference but also often integrate hardware-accelerated video encoding and decoding.

When selecting a GPU for AI-based video analysis, several hardware characteristics are critical:

- **Compute Performance:** Indicates the GPU's raw processing power. TOPS (Trillions of Operations Per Second) measures the performance of AI hardware, focusing on integer operations, while TFLOPS (Tera Floating Point Operations Per Second) measures

floating-point operations

- **Parallel Processing Cores:** Define the level of parallelism available for deep learning computations. A higher number of processing cores, such as CUDA cores, tensor cores, (depending on the GPU architecture), allows faster execution of the parallel operations, essential for AI inference and video analysis.

- **Video Memory (VRAM):** Defines how much video and model data the GPU can store and process at a given time. Sufficient VRAM capacity ensures smooth handling of high-resolution inputs and complex AI models.

- **Media Engine Capabilities**: Integrated hardware video encoders/decoders.

- **Power Efficiency:** Critical for edge deployments where devices must operate continuously under limited power and thermal constraints. Embedded GPUs, such as the Jetson Orin or Xavier, deliver strong AI performance while maintaining low power consumption within 10-30 W ranges.

- **Software Ecosystem and Framework Support:** Refers to the range of software tools, drivers, libraries, and AI frameworks that are compatible with the GPU and enable efficient model development.

Edge AI Accelerators (TPU, VPU, NPU)

Edge AI accelerators are specialized processors designed to execute artificial intelligence workloads efficiently at the edge. Unlike general-purpose GPUs, which are flexible but power-hungry, these accelerators are optimized for real-time inference, low latency, and energy-efficient computation in compact embedded systems.

Tensor Processing Units (TPUs) are application-specific integrated circuits developed by Google, optimized for matrix and tensor operations commonly used in deep learning. TPUs are often found in both data center environments and smaller variants such as the Google Coral Edge TPU, designed specifically for embedded inference.

Vision Processing Units (VPUs) is an emerging class of microprocessor. It is a specific type of AI accelerator, designed to accelerate machine vision tasks. Vision processing units are distinct from graphics processing units (which are specialised for video encoding and decoding) in their suitability for running machine vision algorithms such as CNN (convolutional neural networks). An example is the *Intel Movidius Myriad X*, a third-generation VPU in Intel's Myriad product line.

A Neural Processing Unit (NPU) is a specialized hardware accelerator integrated into modern processors to enhance the execution of neural network inference tasks. Although the term “NPU” is not strictly defined and is often used for marketing, it generally refers to dedicated AI processing units integrated into consumer and embedded devices, such as the *Huawei Kirin NPU, Samsung Exynos NPU,* and *Apple Neural Engine.*

Camera Requirements

The constraints of Edge AI, specifically the need for low latency and high energy efficiency make the technical specifications of the camera sensor a critical determinant of overall system performance and cost. The selection of superior sensor hardware represents an upstream optimization, as providing high-quality source data reduces the necessity for computationally intensive preprocessing steps on the constrained Edge device, freeing up cycles for core AI inference.

Resolution (Spatial Detail)

Resolution dictates the spatial detail available for feature extraction by Convolutional Neural Networks (CNNs). It is paramount for the efficacy of **Small Object Detection (SOD)**, where distant or small targets are represented by a minimal number of pixels. Although CNN models require small, standardized input resolutions (e.g., 512 x 512) for inference speed, the camera's native capture resolution must be high to preserve critical object features. Downscaling an image captured at a lower native resolution can cause small objects to lose essential pixel integrity, severely degrading detection accuracy upon model input. Therefore, while the final processing input is small, the minimum native capture should be **1080p (FHD)**. Resolutions up to **4 Megapixels** are recommended for applications involving complex or distant targets (such as traffic monitoring) to ensure maximum feature resilience during preprocessing and normalization. [17]

Frame Rate (Temporal Detail)

Frame rate determines the temporal resolution, which is essential for tasks requiring movement analysis, such as **real-time object tracking and temporal anomaly detection**. While security operations utilizing human observers require a minimum of **8 FPS** to maintain detection confidence, AI systems generally require higher rates between **15 to 30 FPS**, to robustly track objects, handle occlusions, and analyse fast micro-situations. This higher frame rate is necessary for applications demanding rapid temporal analysis (like human activity recognition or anomaly prediction). Given that video processing is estimated to be approximately 30 times more energy-intensive than processing static images, this increased FPS must be justified by the critical nature of the application, as it directly impacts power consumption. Applications focused primarily on **general object detection** may utilize lower frame rates to maintain energy efficiency. [18]

Dynamic Range and Low-Light Performance

The system's ability to operate 24/7 requires specialized capabilities to handle environmental extremes. **Wide Dynamic Range (WDR)** is crucial for managing high-contrast scenes (e.g., strong backlight) where severe contrast can cause silhouetting and degrade detection performance. However, WDR can amplify sensor noise in dark regions, which requires subsequent computational compensation on the Edge device. Similarly, for low-light performance, high sensor sensitivity (low LUX requirement) and ideally **Near-Infrared (NIR) capability** are needed to counter the effects of noise, which significantly deteriorates object detection accuracy, especially for small objects. High-sensitivity hardware reduces the dependency on computationally expensive software denoising filters, maximizing the cycles available for core AI inference. Selecting hardware that features superior WDR and high sensitivity is a requirement for all 24/7 security applications, as it provides inherently cleaner source data, thereby ensuring the efficiency and sustainability of the downstream Edge AI processing. [19]

6.3.2 Data Storage and Transmission Needs

In order to manage and process the video streams, it is essential to choose the transmission and storage strategy, which is mainly driven by the volume of the video streams and the system architecture setup.

Storage

On edge devices, storage is as much an algorithmic choice as it is a hardware one. A general pattern is to treat local media (eMMC/UFS on smart cameras; NVMe-class SSDs on gateways) as “evidence buffers,” while pushing only condensed outputs upstream. Concretely, on-device analytics (motion filtering, detection/tracking, summarization) let systems persist compact keyframes/clips and rich metadata (boxes, tracks, IDs/embeddings) instead of full streams; this keeps write loads sequential and predictable and reduces capacity pressure on flash with limited endurance. Some frameworks designed for resource-constrained sensors explicitly extract salient frames and discard redundancy, with the summarized set encrypted before any off-device movement— an approach aligned with edge computing’s goal of preserving utility while shrinking the dataset. [20]

From a codec standpoint, H.264/AVC remains the baseline due to encode/decode blocks, but H.265/HEVC typically delivers about a 50% bitrate reduction at similar subjective quality, at the cost of higher compute and is broadly available in contemporary SoCs, encoders, and recorders. AV1 promises further efficiency, but live/edge adoption depends on available hardware decode/encode and power budgets. For systems intended to be field-upgradable, a pragmatic storage plan pairs (i) the best hardware-accelerated codec available on the device

(often HEVC today; AV1 when acceleration exists) with (ii) content-aware reduction upstream of the encoder (frame/ROI filtering on device) to keep the stored corpus proportional to information, not frame rate. [21]

Transmission

In general, most edge solutions are limited in the upstream side of things therefore multi-tier edge designs push perception near the camera and exchange results or intermediate features rather than raw frames, explicitly reducing backhaul sensitivity and stabilizing latency under variable networks (Wi-Fi/4G/5G). [22]

In general-purpose video pipelines, standards-based media paths remain the safest option. RTP/RTSP under ONVIF Media/Replay, including RTSP-over-TCP and RTSP-over-WebSocket for traversal, are widely supported in cameras, NVRs, and gateways. For challenging or lossy links, modern contribution transports such as SRT (ARQ on UDP with optional AES) or WebRTC (interactive, ultra-low-latency) are increasingly viable at the edge, chosen according to whether reliability over the public Internet (SRT) or sub-second interactivity (WebRTC) is the dominant constraint. Finally, when video must traverse wirelessly, store-and-forward behaviours are essential. Edge nodes buffer locally, prioritize metadata/summaries, and backfill clips when unattended. Many edge frameworks formalize this by encrypting salient frames before transmission and treating multi-camera deployments as collaborative producers that ingest once and distribute locally. [20][22]

6.4 Market and Technology Review

This section provides an overview of current hardware technologies enabling AI-based video analysis at the network edge. The review focuses on the computational platforms that form the foundation of edge AI systems, examining their architectural principles, performance characteristics, and suitability for research use. Key parameters considered throughout the analysis include processing performance and power consumption, together with qualitative aspects such as openness for custom software development, flexibility of integration within broader system architectures, and availability of compatible AI tools and frameworks. These criteria establish the basis for selecting appropriate devices to be used in other phases of the project.

6.4.1 Review of available technologies and devices for AI-based video analysis

The deployment of AI models at the network edge requires specialized hardware capable of handling high computational and memory demands efficiently. The field of AI accelerators based on custom hardware is still emerging and lacks a dominant design. [23] Existing solutions

differ widely in their physical architectures, hardware features, and software capabilities, making it difficult for users to select appropriate devices for their computational requirements.

Current AI accelerators can be categorized into several **physical types**:

- **Coprocessor:** A supplemental processor connected to a host CPU to accelerate specific AI tasks. For example, the Google Coral USB Accelerator includes an Edge TPU coprocessor and can be plugged into a host system for inference acceleration.
- **System-on-a-Chip (SoC):** A compact chip integrating CPUs, GPUs, memory controllers, and AI coprocessors. The HiSilicon Kirin 970, for instance, combines CPU and GPU cores with two dedicated Neural Processing Units (NPUs) for efficient on-device inference.
- **System-on-Module (SoM)**: A modular unit that incorporates a processor, memory, and I/O but requires a carrier board for full functionality. The NVIDIA Jetson Nano exemplifies this type, providing a flexible and upgradeable edge AI platform.
- **Single-Board Computer (SBC):** A complete computing system on a single circuit board, including all components required for standalone operation. The Google Coral Dev Board is an SBC featuring an integrated Edge TPU alongside CPU, GPU, memory, and networking interfaces.

These device types illustrate the diversity of hardware architectures currently available for AI-based video analysis and other edge applications, underscoring the need for careful evaluation based on performance, power efficiency, and integration requirements. [23]

6.4.2 Overview of relevant manufacturers and models

The market for AI-based video analysis hardware includes both traditional video surveillance manufacturers integrating on-device AI capabilities and technology companies developing specialized edge computing platforms for AI inference, as those mentioned in the previous subsection. These solutions differ in design philosophy, ranging from fully integrated smart cameras to modular edge computing devices capable of processing multiple video streams.

Axis Communications, Dahua Technology, and Hikvision are among the leading producers of AI-enabled IP cameras designed for security and smart building applications. Their latest models incorporate embedded AI accelerators for real-time analytics.

While integrated smart cameras represent one end of the AI video analysis spectrum, this section focuses primarily on the hardware platforms that serve as the computational foundation for such systems. These platforms, typically designed as SoC, SoM, or SBC configurations, enable flexible deployment of AI workloads across a wide range of edge environments. However, all-in-one smart camera systems, while convenient to deploy, often present a key disadvantage. They are closed and proprietary ecosystems. Their firmware, AI models, and software stacks are usually locked by the manufacturer, limiting access to low-level configurations, customization, or

integration with third-party algorithms. This restricts experimentation, model retraining, and adaptation to specific research or privacy requirements.

Directly comparing different categories of edge AI hardware (SoC/ SoM/ SBC) is inherently challenging. These device types vary not only in their physical form factors but also in their intended use, integration level, and performance metrics. For instance, an SoC may be optimized for power-efficient inference within an embedded sensor, while an SBC includes full peripheral interfaces and operating systems for standalone deployment. Likewise, SoMs occupy an intermediate position, offering high flexibility but relying on external carrier boards for operation. As a result, performance indicators such as processing speed, power consumption, and thermal efficiency are often not directly comparable across categories.

Examples of Edge AI Hardware

Google Coral USB Accelerator

The Google Coral USB Accelerator is a compact Edge TPU coprocessor designed to accelerate AI inference on host systems. It delivers up to 4 TOPS (tera-operations per second, int8) with an efficiency of approximately 2 TOPS per watt. The device connects via USB 3.0 Type-C, providing both data transfer and power. Measuring roughly 65 mm × 30 mm × 8 mm, it is compatible with Linux (including Debian-based distributions), macOS, and Windows. The accelerator works seamlessly with TensorFlow Lite models, which can be compiled to run on the Edge TPU, enabling high-performance on-device AI inference in a small form factor. [24]

Figure 5: Google Coral USB Accelerator

Google Coral Dev Board

The **Coral Dev Board** is a compact single-board computer designed for high-speed machine learning inference directly at the edge. It integrates the **Coral System-on-Module (SoM)**, which combines an **NXP i.MX 8M SoC** (quad-core Arm Cortex-A53 with Cortex-M4F), **1–4 GB of LPDDR4 RAM**, and **8–16 GB of eMMC storage**. Its core advantage lies in the dedicated **Google Edge TPU coprocessor**, a custom ASIC capable of executing modern vision models such as MobileNet v2 at nearly **400 FPS** with minimal power consumption. [25]

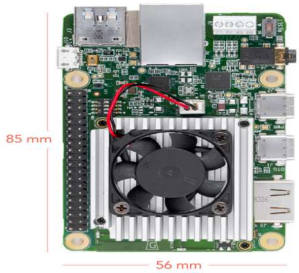

Figure 6: Google Coral Dev Board

NVIDIA Jetson Orin

NVIDIA Jetson Orin family (AGX, NX, Nano) family represents a highly scalable set of embedded AI computers built on the NVIDIA Ampere architecture, offering solutions ranging from high-performance industrial gateways (AGX) to power-efficient modules (Nano, NX). The Jetson Orin modules offer up to 275 TOPS (Tera Operations Per Second) for multimodal AI inference, providing a performance increase of up to 8X over the previous generation. The flagship Jetson AGX Orin features a high-performance GPU with 2048 CUDA cores and 64 Tensor Cores, alongside an Arm Cortex-A78AE CPU, operating in a power envelope between 15W and 40W. The more compact Orin NX and Orin Nano modules offer lower power consumption and smaller footprints while retaining access to the same powerful Jetson software stack, which includes the optimized DeepStream SDK and the TensorRT inferencing engine for accelerated video analytics. This architectural uniformity and robust software environment are critical for rapidly developing and deploying complex, multi-stream AI applications. [26]

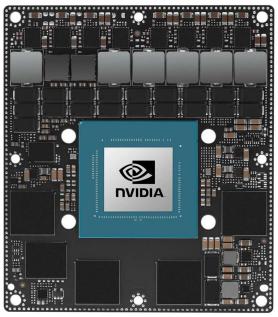

Figure 7: NVIDIA Jetson AGX

ArmSoM-AIM7

ArmSoM-AIM7 uses Rockchip RK3588, a new generation flagship eight-core 64-bit processor with a main frequency of up to 2.4GHz, 6 TOPS computing power NPU, and can be equipped with up to 32GB of large memory. The interface is fully compatible with Jetson Nano. [27]

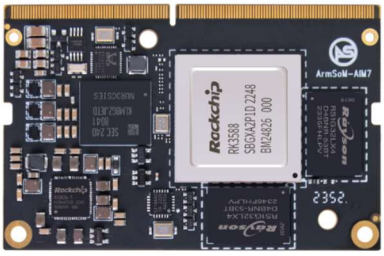

Figure 8: ArmSoM-AIM7

Geniatech AIM-M2

The Geniatech AIM-M2 is a compact, high-efficiency AI accelerator designed for edge servers, industrial PCs, and embedded systems requiring enhanced neural-network performance without redesigning the main compute platform. Powered by the Kinara Ara-2 NPU, the module delivers up to 40 TOPS of dedicated AI compute through a standard M.2 interface, enabling seamless integration into existing hardware architectures. [28]

Figure 9: Geniatech AIM-M2

6.5 Methodology of Evaluation

Determining the suitability of Edge AI hardware is a complex task because the landscape comprises heterogeneous solutions: dedicated accelerators (e.g., Google Coral, Geniatech AIM-M2), full System-on-Chips (e.g., NVIDIA Jetson, ArmSoM-AIM7), and general-purpose embedded systems. This diversity makes direct comparison based on single hardware metrics (like TOPS or power consumption) challenging, as performance varies significantly based on architecture and software optimization. Therefore, the evaluation methodology must integrate both standard quantitative metrics and critical qualitative criteria to assess the device's viability for a specific research and deployment environment.

Quantitative Metrics: Performance and Efficiency

The primary quantitative metrics are processing speed and power consumption. While TOPS indicates potential computational capability, the real measure of performance is the

achievable **processing speed (**latency and throughput, typically measured in frames per second—FPS) for the project's specific target models (e.g., YOLOv7). For example, Neural Processing Units (NPUs) often achieve superior efficiency for standard inference tasks, while integrated GPUs (such as in the Jetson series) offer higher peak throughput for large batch sizes, albeit often at greater **power consumption.** For smart building deployments, **TOPS per Watt** (energy efficiency) is often the most critical quantitative criterion, as it directly impacts cooling requirements and long-term operational costs, forcing a trade-off against raw speed.

Qualitative and Systemic Criteria for Research Utilization

Beyond raw speed, the successful utilization of hardware in a dynamic research environment depends on systemic factors. The **openness for custom software** and the **flexibility of hardware integration** are paramount. High-performance accelerators from manufacturers like NVIDIA often come with highly optimized, proprietary toolkits (like the **TensorRT** inferencing engine and the DeepStream SDK), which deliver unmatched performance but restrict deployment to that specific hardware ecosystem. Conversely, devices leveraging broader architectures (like Arm, NXP) and utilizing frameworks such as **OpenCV** or vendor-neutral runtimes (e.g., OpenVINO, ONNX) offer greater **flexibility of placement** within the Edge-Fog-Cloud architecture. The evaluation must assess the ease of integrating the core project tools with the proposed hardware, factoring in the maturity of the software development kits, documentation, and long-term market support. This qualitative assessment ensures the chosen devices not only meet minimum speed requirements but also provide a stable, manageable, and adaptable platform for the project's ongoing research and eventual commercial utilization.

6.6 Description of recommended architecture and edge devices - results of evaluation

6.6.1 Edge devices

The evaluation of Edge AI hardware for the InnovAIte project led to the selection of the NVIDIA Jetson Orin AGX module as the primary edge device. This choice is rooted in the device's substantial computational throughput and energy management capabilities, which are essential for managing the project's most demanding video processing and multi-stream analysis requirements.

The Jetson AGX Orin is a highly capable embedded AI compute module built on the NVIDIA Ampere architecture. It integrates a powerful GPU with 2048 CUDA cores and 64 Tensor Cores, supported by an Arm Cortex-A78AE CPU. Quantitatively, the module delivers a maximum throughput of 275 TOPS for multimodal AI inference. Critically, its power consumption is highly

configurable, operating in modes ranging from 15W up to 40W, ensuring a very strong performance-per-watt ratio. This power efficiency is necessary for meeting the project's energy management objectives in smart building environments. The AGX Orin provides the robust power needed to reliably execute complex models like YOLOv11 at over 30 FPS on multiple (20 and more) cameras, a requirement for real-time security applications and central data fusion.

This selection is further validated by the great scalability offered across the entire Jetson Orin family, which ensures flexibility for different application loads. The AGX Orin serves as the high-performance core, while its scaled-down variants are selected for less demanding applications:

- The Jetson Orin NX provides reliable intermediate performance, delivering up to 157 TOPS, while operating within a power envelope of 10W to 25W.
- The highly energy-efficient Jetson Orin Nano offers accessible performance of up to 67 TOPS while consuming only 5W to 10W of power, making it an excellent choice for localized, simpler object detection tasks at the network edge.

Beyond raw hardware metrics, the qualitative advantages of the Jetson platform were decisive. The organization has established internal expertise and reliability working with previous generations, such as the Jetson Nano, ensuring a seamless transition and minimizing time spent on basic system integration. Furthermore, the platform's robust and extensive software ecosystem is irreplaceable for accelerating research goals. The unified JetPack SDK, along with specialized tools like the TensorRT inferencing engine and the DeepStream SDK, guarantees that the project's complex models can be quickly optimized and deployed for real-time performance. This strong software support simplifies multi-platform development across the entire Orin family. Finally, the platform's wide market presence and the availability of third-party components (carrier boards, peripherals, and software partners) secure the supply chain and provide exceptional architectural flexibility. This modularity future-proofs the investment, allowing the compute module to be upgraded without redesigning the entire embedded system.

6.6.2 Architecture

Following the depiction of the edge device, we move on to the architecture. We evaluate two complementary architectures for AI-enabled video analytics in buildings. **Fully local** (all-on-edge) processing and split **edge–cloud processing**. Both approaches aim to minimize end-to-end latency for safety/security events, reduce bandwidth, and preserve privacy by default, while remaining operable across heterogeneous camera estates. Research shows that pushing perception close to cameras improves resilience to connectivity outages and reduces exposure of raw imagery.

Our design targets include: fast local alerting for critical events, bounded uplink (events/metadata over continuous video), privacy-by-design, graceful degradation during

connectivity loss and fleet scalability via adaptive configuration and remote orchestration. These goals reflect lessons from large-scale surveillance and building-safety deployments that rely on edge-side filtering (keyframe/usefulness) and selective publishing to maintain utility under constrained resources.

For **fully local** processing each camera or small cluster attaches to an edge node (Jetson) that executes the complete pipeline: decode → quality/usefulness checks → keyframe selection → detection/tracking → rule evaluation and local actuation → short-horizon ring buffer. Only compact events and optional evidence frames are forwarded to a central dashboard. This placement delivers deterministic latency for safety actions (e.g. violation alarms, restricted-area breaches) and sustains operation during network outages. It also confines raw video to the premises, lowering privacy risk. Evaluations of edge systems and task-aligned detectors on modest hardware motivate this design, showing feasible real-time inference and direct, on-device alerting when models are sized for the platform.

The data flow would resemble the following:

- Input: RTSP/H.264/H.265 into hardware decode.
- Filtering: frame-level "video usefulness" (blur, occlusion, brightness) and low-cost keyframe selection to suppress content with little information.
- Inference: quantized, edge-suitable detectors; post-processing and evaluation.
- Output: GPIO/fieldbus for alarms/interlocks; event logging with short clips.

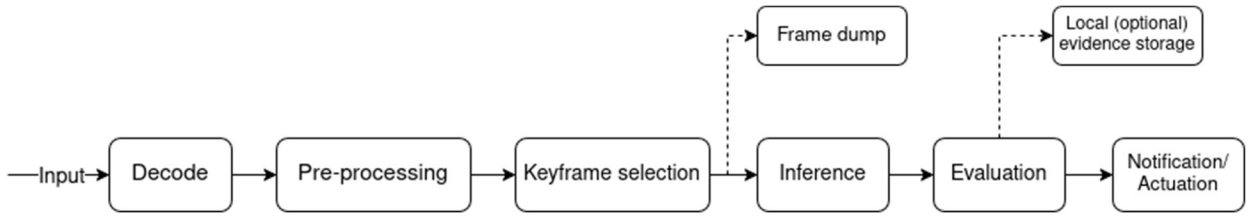

Figure 10: The Data Flow

The end-result would attain low and stable latency, low uplink (events + sparse keyframes), privacy-preserving by default, but it is performance bounded by its local compute, which can be on the other hand scaled-out via additional nodes.

Considering the **edge-cloud** setup (edge-fog-cloud model), the edge side of things executes fast, first-pass analytics (decode, keyframe selection, lightweight detection/tracking) and publishes selected frames/snippets with rich metadata to the cloud for second-stage compute-heavy tasks (e.g., cross-camera correlation, longer analytics, model A/B testing and aggregation, search).

This partitioning preserves local responsiveness while enabling higher reasoning and rapid model iteration. To respect resource and link variability, we would incorporate online adaptation that jointly tunes per-stream frame rate, resolution, and even model choice, coupled with bandwidth allocation across cameras while maintaining accuracy/latency under shared constraints.

The data path would look similar to the previous architecture type with some different caveats:

- Edge Input: RTSP/H.264/H.265 into hardware decode.
- Edge Filtering: frame filtering (blur, occlusion, brightness) and keyframe selection
- Edge/Cloud Inference: model inference, post-processing and evaluation.
- Cloud storage: long-term storage; model registry and A/B harness
- Edge Output: notifying, event logging with short clips.

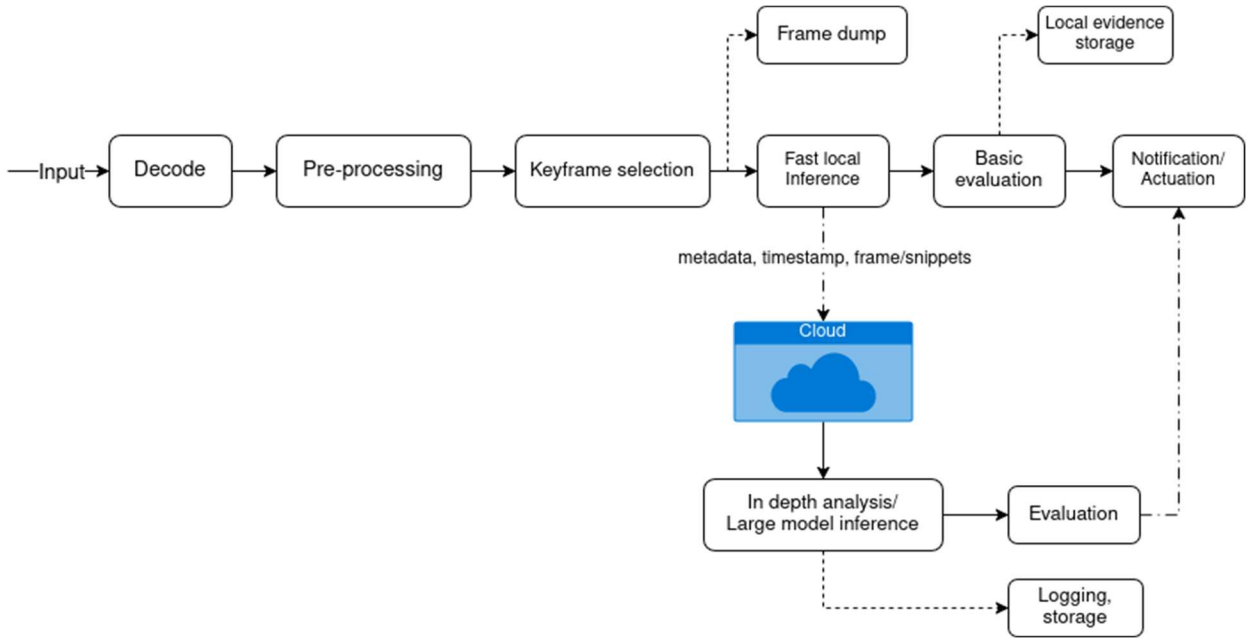

Figure 11: The Data Path

This setup would again include low local latency, bounded uplink via selective publishing, higher analytic expressiveness centrally, graceful degradation when the network fails (edge continues local alerting, queues uploads).

All in all, both architectures satisfy the design goals on different frontiers. The first architecture maximizes privacy and offline resilience with deterministic local performance, while the second one opts for local responsiveness with an online in-depth analysis. Both are anchored by the same edge foundation, keyframe selection, toolchain alignment and task-appropriate models.

6.6.3 Software and communication - edge devices

To implement the recommended architectures, we adopt a unified software stack that runs end-to-end from model training to on-device inference and, when applicable, cloud-side enrichment. At the core are modern one-stage object detectors from the YOLO family (lightweight models, n/s/tiny variants for constrained devices and larger backbones for gateways). Models are trained in PyTorch using the Ultralytics pipeline, with task-specific class sets and synthetic augmentation to reflect different situations. We export trained models to ONNX as the interchange format, which lets us produce hardware-specific inference engines without maintaining multiple training codebases.

As mentioned before, we build the systems on top of the Nvidia's Metropolis platform, its application framework and SDKs, such as the NVIDIA DeepStream SDK, which we use as our primary runtime for video ingest, analytics, and message publishing. DeepStream's GStreamer-based pipeline gives us zero-copy hardware decode, multi-stream batching, high-throughput inference via TensorRT, built-in tracking, and on-screen display. For system integration, we rely on DeepStream's message plugins to publish compact events and camera frames to MQTT/HTTP backends, while keeping raw streams local by default. This lets us implement both architectures with the same building blocks. For edge execution, we compile ONNX models into device-optimized engines - TensorRT for NVIDIA Jetson. We also aim to push post-processing into the engine where it is possible to reduce CPU overhead and stabilize latency.

On the device, the runtime is intentionally simple and resilient. The RTSP streams are ingested via GStreamer/FFmpeg with hardware-accelerated H.264/H.265 decode, then the use of OpenCV for lightweight image operations and frame access as a part of the prefilter stage, before inference, which applies keyframe selection and quick quality checks (blur, brightness, tilt/occlusion) so we avoid spending cycles—or bandwidth—on redundant or unusable frames. The detector runs on the compiled engine (TensorRT/ONNX RT), optionally followed by multi-object tracking and low-latency monitored-area-rule obedience. Events are emitted immediately with compact payloads (timestamps, classes, boxes, confidences), and when needed or justified, an evidence image or a short clip is exported. Local actuation uses simple GPIO/IO paths and a ring buffer stores a few hours to days of video for retrospective analysis.

Jetson itself offers a plethora of communication options, from general-purpose buses like USB, UART, SPI, I2C, CAN, PCIe (for extra storage peripherals) or a simple GPIO to higher level such as MIPI for camera integration or ethernet for accessing a network. The device can be connected to a network router via ethernet. to access wireless communication in general or to enable over-the-air updates.

Deployment and operations are containerized for consistency. Edge services run as Docker images with signed images and models distributed via a registry. Nodes export metrics (FPS, per-stage latency, engine utilization, drop rates, quality scores) to a central time-series store

and dashboard. Logging is essential to help diagnose camera or pipeline issues, even remotely. Security and privacy are defaults - mutual-TLS for control, per-device credentials, encrypted local storage when feasible, and local analysis keeps the raw streams except under explicit incident workflows.

6.7 Plan for implementing edge device integration into research tasks

Building on the previous chapters—where a selected subset of use cases, deployment environments, and appropriate Edge-AI devices were identified—the next phase of the project will use this framework in practical experimentation.

The outputs of this report provide the foundation for executing the respective research tasks using the selected devices, focusing on the monitoring-oriented use cases. This ensures that the conceptual analysis can be directly applied to real-world deployment, supporting both experimentation and performance evaluation within building environments.

Deployment Objectives and Research Focus

The planned research tasks will utilize the selected Edge-AI devices for two main purposes:

- collecting reference data

- deploying trained neural networks

First, the devices will gather real-world video data to form high-quality datasets for the selected subset of use cases.

Second, they will serve as execution platforms for on-device inference, enabling the team to test and validate models in real-time scenarios. The outputs of this report guide the selection and configuration of devices to ensure these tasks can be effectively addressed.

A central focus of the upcoming work is to evaluate the feasibility and utility of synthetic data. Comparative experiments will determine how well artificially generated data can supplement or partially replace real-world data for training neural networks corresponding to the selected use cases. The outputs of this report directly inform the design of these experiments, ensuring that real, synthetic, and hybrid datasets can be leveraged efficiently to assess model accuracy, robustness, and generalization.

In addition, the research will assess the computational performance of the selected Edge-AI devices when executing multiple neural networks simultaneously for the chosen use-case subset—such as people detection, occupancy estimation, anomaly recognition, or vehicle monitoring. Metrics including inference speed, resource utilization, and scalability will be evaluated. The findings derived from the outputs of this report will enable the team to plan and implement these experiments reliably, ensuring that the devices are used optimally for the respective research tasks.

Implementation Conditions and Next Steps

The detailed schedule of research tasks will be defined within the relevant work packages, with execution contingent upon real-world installation feasibility, access permissions, and compliance with safety and privacy requirements. Once these conditions are met, the selected Edge-AI devices will be deployed to collect data and perform inference for the chosen subset of use cases. The outputs of this report provide the necessary guidance to ensure that the research team can successfully carry out these tasks, generating results that advance understanding of Edge-AI applications in the building sector.

7 Bibliography

[1] Dardour, A.; El Haji, E.; Begdouri, M.A. Video Surveillance and Artificial Intelligence for Urban Security in Smart Cities: A Review of a Selection of Empirical Studies from 2018 to 2024. Comput. Sci. Math. Forum 2025, 10, 15. https://doi.org/10.3390/cmsf2025010015

[2] Bilous, N.; Malko, V.; Frohme, M.; Nechyporenko, A. Comparison of CNN-Based Architectures for Detection of Different Object Classes. AI 2024, 5, 2300-2320. https://doi.org/10.3390/ai5040113

[3] Duong HT, Le VT, Hoang VT. Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. Sensors (Basel). 2023;23(11):5024. Published 2023 May 24. Doi: 10.3390/s23115024

[4] El Husseini, F.; Noura, H.N.; Salman, O.; Chahine, K. Machine Learning in Smart Buildings: A Review of Methods, Challenges, and Future Trends. Appl. Sci. 2025, 15, 7682. https://doi.org/10.3390/app15147682

[5] Iqbal, U.; Davies, T.; Perez, P. A Review of Recent Hardware and Software Advances in GPU-Accelerated Edge-Computing Single-Board Computers (SBCs) for Computer Vision. Sensors 2024, 24, 4830. https://doi.org/10.3390/s24154830

[6] Wang, Xubin, and Weijia Jia. "Optimizing edge AI: a comprehensive survey on data, model, and system strategies." arXiv preprint arXiv: 2501.03265 (2025).

[7] Balaskas, Konstantinos, et al. "Hardware-aware DNN compression via diverse pruning and mixed-precision quantization." IEEE Transactions on Emerging Topics in Computing 12.4 (2024): 1079-1092.

[8] Mohan, Nitinder & Kangasharju, Jussi. (2017). Edge-Fog Cloud: A Distributed Cloud for Internet of Things Computations. 10.1109/CIOT.2016.7872914.

[9] Rajavel, R., Ravichandran, S.K., Harimoorthy, K. et al. IoT-based smart healthcare video surveillance system using edge computing. J Ambient Intell Human Comput 13, 3195–3207 (2022). https://doi.org/10.1007/s12652-021-03157-1

[10] Y. -Y. Chen, Y. -H. Lin, Y. -C. Hu, C. -H. Hsia, Y. -A. Lian and S. -Y. Jhong, "Distributed Real-Time Object Detection Based on Edge-Cloud Collaboration for Smart Video Surveillance Applications," in IEEE Access, vol. 10, pp. 93745-93759, 2022, doi: 10.1109/ACCESS.2022.3203053

[11] Cob-Parro, A.C.; Losada-Gutiérrez, C.; Marrón-Romera, M.; Gardel-Vicente, A.; Bravo-Muñoz, I. Smart Video Surveillance System Based on Edge Computing. Sensors 2021, 21, 2958. https://doi.org/10.3390/s21092958

[12] E. Badidi, K. Moumane and F. E. Ghazi, "Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review," in IEEE Access, vol. 11, pp. 80543-80572, 2023, doi: 10.1109/ACCESS.2023.3300658.

[13] G. Gallo, F. D. Rienzo, F. Garzelli, P. Ducange and C. Vallati, "A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge," in IEEE Access, vol. 10, pp. 110862-110878, 2022, doi: 10.1109/ACCESS.2022.3215148

[14] H. Sun, W. Shi, X. Liang and Y. Yu, "VU: Edge Computing-Enabled Video Usefulness Detection and its Application in Large-Scale Video Surveillance Systems," in IEEE Internet of Things Journal, vol. 7, no. 2, pp. 800-817, Feb. 2020, doi: 10.1109/JIOT.2019.2936504

[15] Q. Zhang, H. Sun, X. Wu and H. Zhong, "Edge Video Analytics for Public Safety: A Review," in Proceedings of the IEEE, vol. 107, no. 8, pp. 1675-1696, Aug. 2019, doi: 10.1109/JPROC.2019.2925910.

[16] Ferone, A.; Maratea, A.; Camastra, F.; Ciaramella, A.; Staiano, A.; Lettiero, M.; Polizio, A.; Lombardi, F.; Spoleto, A.J. AiWatch: A Distributed Video Surveillance System Using Artificial Intelligence and Digital Twins Technologies. Technologies 2025, 13, 195. https://doi.org/10.3390/technologies13050195

[17] Qihan Feng, Xinzheng Xu, Zhixiao Wang. Deep learning-based small object detection: A survey[J]. Mathematical Biosciences and Engineering, 2023, 20(4): 6551-6590. doi: 10.3934/mbe.2023282

[18] Harjanto, Fredro & Wang, Zhiyong & Lu, Shiyang & Feng, David Dagan Feng. (2012). Evaluating the impact of frame rate on video based human action recognition. ACM International Conference Proceeding Series. 376-381., doi: 10.1145/2425836.2425909.

[19] Rodríguez-Rodríguez, José & López-Rubio, Ezequiel & Ángel-Ruiz, Juan & Molina-Cabello, Miguel A.. (2024). The Impact of Noise and Brightness on Object Detection Methods. Sensors. 24. 821. doi: 10.3390/s24030821.

[20] K. Muhammad, T. Hussain, J. J. P. C. Rodrigues, P. Bellavista, A. R. L. de Macêdo and V. H. C. de Albuquerque, "Efficient and Privacy Preserving Video Transmission in 5G-Enabled IoT Surveillance Networks: Current Challenges and Future Directions," in IEEE Network, vol. 35, no. 2, pp. 26-33, March/April 2021, doi: 10.1109/MNET.011.1900514

[21] Yutong Liu, Linghe Kong, Guihai Chen, Fangqin Xu, Zhanquan Wang, Light-weight AI and IoT collaboration for surveillance video pre-processing, Journal of Systems Architecture, Volume 114, 2021, 101934, ISSN 1383-7621, https://doi.org/10.1016/j.sysarc.2020.101934.

[22] H. Sun, Y. Yu, K. Sha and B. Lou, "mVideo: Edge Computing Based Mobile Video Processing Systems," in IEEE Access, vol. 8, pp. 11615-11623, 2020, doi: 10.1109/ACCESS.2019.2963159

[23] Li, Wenbin, and Matthieu Liewig. "A survey of AI accelerators for edge environment." *World Conference on Information Systems and Technologies*. Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-45691-7_4

[24] Google Colar USB accelerator. Available from: https://www.coral.ai/static/files/Coral-USB-Accelerator-datasheet.pdf [Online; accessed 12-11-2025].

[25] Coral Dev Board . Available from: https://www.coral.ai/static/files/Coral-Dev-Board-datasheet.pdf [Online; accessed 12-11-2025].

[26] Jetson Orin Modules. Available from: https://developer.nvidia.com/embedded/jetson-modules [Online; accessed 12-11-2025].

[27] ArmSoM AIM7. Available from: https://docs.armsom.org/armsom-aim7 [Online; accessed 12-11-2025]

[28] Geniatech AIM-M2. Available from: https://www.geniatech.com/product/aim-m2/ [Online; accessed 12-11-2025].

8 Annex

Annex 1 - Comparison of devices - link to annex

- **Accessibility classification:** Public (P)
- Content
  - A table containing an overview of selected devices on the market in recent years. The table includes both brand new devices and devices that have been on the market for several years. Relevant parameters are listed for each device.

About the project

The InnovAIte Slovakia project was launched in response to the call "Transformation and Innovation Consortia" announced by the Government Office of the Slovak Republic under the Recovery and Resilience Plan of the Slovak Republic (Component 9 – Investment 2: Supporting cooperation between companies, academia, and research and development organizations). The call aims to drive systemic transformation and increase the added value of key sectors of the Slovak economy through intensive collaboration among research institutions, innovative enterprises, the public sector, and internationally renowned partners.

InnovAIte Slovakia seeks to build a dynamic and sustainable innovation ecosystem in the field of artificial intelligence (AI), effectively linking cutting-edge research with practical applications. The project emphasizes the development of AI solutions that are not only technologically advanced but also ethical, environmentally sustainable, and socially beneficial. The consortium brings together leading research centers, universities, and businesses from Slovakia, Germany, and the Czech Republic.

Key focus areas include the development of AI algorithms for improving building energy efficiency, enhancing traffic safety through video analytics, automating software development, driving digital transformation in the insurance sector, and validating functional prototypes. Special attention is also given to AI education, talent development, and the incubation of startups.

The project's outcomes will significantly strengthen Slovakia's position in research and innovation, reduce environmental burdens, increase the competitiveness of the national industry, and modernize public services. By bridging the public and private sectors with international research excellence, InnovAIte Slovakia stands as a key instrument in addressing the social and economic challenges of today.

"Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I02-03-V01-00029"

InnovAite
SLOVAKIA

www.innovaite.sk

Financované Európskou úniou
NextGenerationEU

PLÁN OBNOVY

ÚRAD PODPREDSEDU VLÁDY
SLOVENSKEJ REPUBLIKY
PRE PLÁN OBNOVY
A ZNALOSTNÚ EKONOMIKU

VAIA VÝSKUMNÁ A INOVAČNÁ AUTORITA