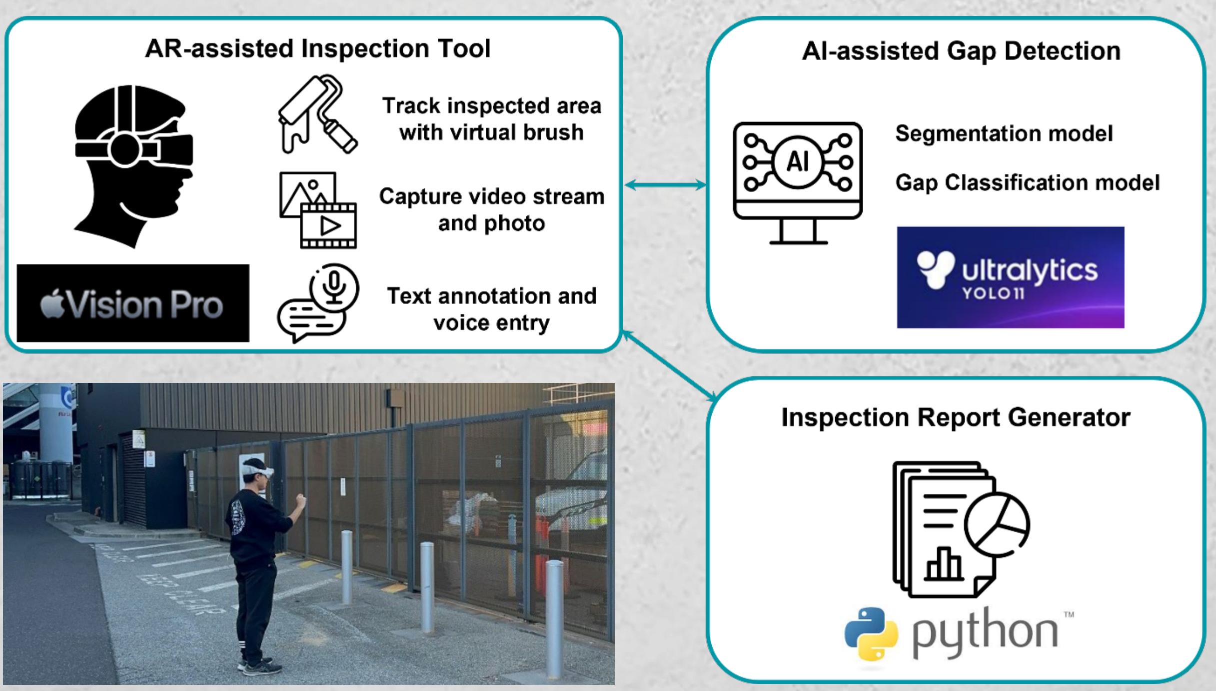
PROJECT #29: REAL-TIME EH&S INTERVENTION TO IMPROVE SITE SAFETY (SCOPING STUDY)

FINAL REPORT



MONASH University

THE UNIVERSITY OF MELBOURNE

lendlease

Australian Government
Department of Industry, Science and Resources

Cooperative Research Centres Program

TABLE OF CONTENTS

List of Tables

List of Figures

CONFIDENTIAL:	☐ Yes X No
Author/s of this report:	Dr Pari Delir Haghighi (Monash University), Dr Jiazhou Liu (Monash University), Dr Aravinda S. Rao (University of Melbourne), Fucai Ke (Monash University)
Date of this report:	27 June 2025
Project completion date:	30 June 2025
Program Leader reviewer:	Assoc. Prof. Duncan Maxwell
Project title:	Real-time EH&S Intervention to Improve Site Safety (Scoping Study)
Project duration:	2 years
Partners:	• Lendlease Digital • Monash University • The University of Melbourne
Project team members:	• Dr Pari Delir Haghighi (Project Lead) (Monash University) • Professor Tim Dwyer (Monash University) • Dr Hamid Rezatofighi (Monash University) • Dr Teresa Wang (Monash University) • Dr Benjamin Tag (Monash University) (University of New South Wales) • Dr Jiazhou Liu (Monash University) • Fucai Ke (Monash University) • Professor Tuan Ngo (University of Melbourne) • Dr Philip Christopher (University of Melbourne) • Dr Aravinda S. Rao (University of Melbourne)
Acknowledgements:	The Research team gratefully acknowledges the financial support from Building 4.0 CRC and industrial partners.

Disclaimer

ABBREVIATIONS

AI	Artificial Intelligence
AR	Augmented Reality
BIM	Building Information Modelling
CCTV	Closed Circuit Television
CNN	Convolutional Neural Network
CV	Computer Vision
DL	Deep Learning
EH&S	Environment, Health, and Safety
FFH	Fall from Height
IoT	Internet of Things
KPI	Key Performance Indicator
LLM	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing
PPE	Personal Protective Equipment
RNN	Recurrent Neural Network

EXECUTIVE SUMMARY

Fall from heights (FFH) is one of the most significant safety concerns in the construction industry, consistently identified as a leading cause of severe injuries and fatalities worldwide. The risks associated with FFH are particularly critical in high-rise construction projects, where the height of operations magnifies both the probability and the severity of incidents. Currently, the primary approach to identifying FFH risks involves physical site inspections by site personnel. These inspections typically identify potential hazards, such as gaps in perimeter safety screens, unprotected openings such as holes in floors, unstable scaffolding, or improperly secured safety equipment. While manual inspections are a critical component of workplace safety, they are inherently limited. They rely heavily on the skill and attentiveness of inspectors, are labour-intensive, and can be affected by time constraints, human error, or environmental factors. Moreover, construction sites are dynamic environments where hazards can emerge and change rapidly, making it difficult for manual inspections to provide comprehensive and timely coverage. Recognising these challenges, the project explores state-of-the-art and emerging technologies to develop innovative solutions that enhance hazard detection and support safer, more efficient inspection processes.

In collaboration with our industry partner, we first conducted an extensive literature review to examine recent research and technological developments related to FFH hazard detection in the construction sector. A key finding was the potential of Artificial Intelligence (AI), particularly AI-assisted vision models, to transform how FFH risks are identified and managed. These models can process and analyse visual data in real time, offering an effective means of automating hazard detection. This reduces reliance on manual inspections, enables quicker responses, and improves overall safety outcomes. We also see strong potential in Augmented Reality (AR) technology, which can enhance inspections through immersive, interactive experiences, providing real-time location and progress tracking, image and video capture, and automated report generation.

To demonstrate the feasibility and effectiveness of this approach, we developed proof-of-concept prototypes using AI vision and AR technologies. These prototypes show how the integration of AI and AR can automate inspection tasks, including gap detection in perimeter safety screens. Qualitative feedback from interviews with industry partners confirmed the strong potential of these prototypes to address key safety challenges in high-rise construction by enabling early risk detection, reducing reliance on manual inspections, and improving the accuracy and efficiency of reporting. The project also contributes to workforce readiness by enabling immersive training experiences for new staff and supports industry innovation by showcasing practical applications of emerging technologies.

PROJECT OVERVIEW

Construction sites play a critical role in the building industry, employing a significant portion of the workforce. However, they also represent environments with substantial risks, impacting workers' physical safety and mental well-being. Historically, the construction sector has consistently ranked among the industries with the highest serious injury claims, recently recording the third-highest fatality rate across all sectors (Australia, 2024). Common causes of severe injuries and fatalities in construction include falls from heights, impacts from falling objects, electrical incidents, and collisions involving moving plant and equipment.

Many of these incidents can be traced to underlying issues such as inadequate planning and ineffective workplace design. Notably, 'human error' is the most significant immediate contributing factor in construction accidents, reportedly accounting for up to 90% of all such incidents (Abdelhamid & Everett, 2000). Addressing these systemic and immediate issues through enhanced planning, improved communication strategies, and robust workplace designs could significantly mitigate these risks, thereby supporting safer and healthier work environments across the construction industry.

Data derived from past incidents, potential hazards, safety observations, and related reports offer valuable insights into the ongoing risks present on construction sites and the limitations of current safety practices. Analysing historical data from diverse sources can reveal recurring patterns, support the development of more effective interventions, inform training priorities, and redesign workflows to pre-emptively mitigate risk. The increasing availability of modern technologies for data collection and analysis presents a valuable opportunity for creating innovative, data-driven solutions that enhance safety outcomes across the construction industry.

For construction safety and health, continuous monitoring of conditions in a ‘live’ environment and the ability to proactively generate timely warnings and actions are essential to mitigating and managing potential hazards. Current reporting practices are mostly limited to manual entry of daily reports about incidents and risk events into the system. While current daily reports include useful data to identify meaningful patterns and relationships, such as correlations between variables like delays, level of changes, behaviour, workers’ mood and sentiment, and the Environmental Health & Safety (EH&S) performance and quality, they offer limited support for recording real-time information. By leveraging emerging AI, AR and IoT (including sensors and computer vision) technologies, there is an opportunity to automatically collect rich, real-time data about the physical environment, support tracking and inspection tasks, and provide predictive insights to mitigate incidents and eliminate risks.

Motivated by an envisioned shift toward autonomous buildings, this project aims to identify opportunities for automating safety inspection activities to enhance the Environmental Health & Safety (EH&S) performance and quality of the construction industry by applying modern and emerging technologies. By analysing existing safety records, industry practices, and properties of available technology, we will identify opportunities for improving the safety reporting process, implementing "safety by design" practices, and reducing the overall potential for human error or incompetence by introducing automated processes. Potential technologies of interest include:

- **Computer vision techniques and large visual models**, which can be used to collect, process, and analyse images and videos for identifying areas of concern and predicting safety hazards.
- **Augmented reality (AR) technologies**, which support the modelling, simulation and monitoring of physical environments in 3D, and enable intuitive, gesture-based interactions for capturing data, images, and voice comments.

The project began with a comprehensive literature review and market analysis to explore emerging technologies, assess their applicability to construction environments, and identify gaps in both evidence and practice. This review ensured that the proposed interventions are theoretically grounded and practically viable within the constraints of construction site operations. Following this, the project focused on defining a set of use case scenarios based on feedback from industry partners and then developing and implementing a concept prototype that combines the selected, most viable technologies to improve safety in the identified scenarios.

The project was structured around three main streams, described below.

Stream I: Data Exploration and Analysis – This stream explored existing datasets provided by the industry partner using advanced machine learning techniques and diverse visualisation methods. The goal was to identify and present meaningful trends and correlations related to Environmental Health & Safety (EH&S) issues. The stream aimed to uncover valuable insights by analysing variables such as the type of incidents, associated activities or environments, time, location, and weather conditions. It explored the most effective methods for visualising results and presenting analyses through interactive filters and options. This stream included a workshop with industry partners, followed by iterative cycles of data exploration and analysis.

Stream II: Literature Review and Market Analysis – This stream focused on examining the current research and market offerings to identify and assess the state-of-the-art technologies and data analysis methods for improving site safety. This stream evaluated the feasibility and suitability of existing technologies, such as computer vision or sensor-based systems, for enhancing real-time detection and intervention on construction sites. The stream provided a deeper understanding of the current landscape of technology adoption in construction health and safety and identifying both their strengths as well as key barriers to implementation. The findings from this stream informed the next phase of the project by identifying viable technologies for prototype development.

Stream III: Concept Prototype Design and Testing – This stream built on the findings from Streams I and II to select and test prototype technologies aimed at addressing safety challenges on construction sites. This stream first involved identifying high-risk use case scenarios in collaboration with the industry partner, and then developing concept prototypes that leverage selected modern technologies such as AI-assisted visual models and Augmented Reality (AR) to mitigate these risks. The prototype was qualitatively evaluated by industry partners through interviews, assessing its feasibility and potential for integration into current practices and workflows.

The goal of this research project was to pave the way for showcasing practical applications of emerging technologies to address key safety challenges in high-rise construction, enabling early risk detection, reducing reliance on manual data collection and reporting, and improving the accuracy and efficiency of safety inspections. By leveraging existing data, identifying suitable technologies, and validating them through prototype development and expert evaluation, the project aimed to deliver practical solutions that advance safety practices and contribute to workforce readiness. Ultimately, the project outcomes highlight the potential of integrating advanced technological solutions into everyday construction practices, laying the groundwork for more proactive, AI-assisted, and data-driven safety solutions across the industry.

STREAM I: DATA EXPLORATION AND ANALYSIS

Stream 1 of the project focused on exploring and analysing safety-related data within the construction industry, with a particular emphasis on the industry partner's ecosystem of tools, policies, and data platforms. This stream served as the foundation for understanding the current state of construction safety management, identifying data collection and usage gaps, and laying the groundwork for predictive analytics and digital interventions to enhance site safety.

To achieve this goal, a workshop was held online on July 13 2023, so researchers from Monash University and the University of Melbourne could better understand the industry partner's practices, procedures, and opportunities related to safety evolution. Topics discussed at the workshop included:

- evolution of the industry partner's safety practices over time
- tools and dashboards used for safety training and management at the time
- the methodology used at the time to determine the factors that were essential to the incident occurring
- opportunities and further goals for data analysis to support a safer work environment.

The research team analysed the existing data, including safety observations, incident events, individual behaviours, impacts, etc., via data visualisation, and then shared this analysis with the industry partner.

This information is not publicly available. Further, the industry partner's practices and procedures have continued to evolve and have been superseded by updated internal systems and strategies.

STREAM II: LITERATURE REVIEW AND MARKET ANALYSIS

Stream II of this research program focused on a comprehensive review of state-of-the-art technologies, industry practices, and emerging machine learning and AI techniques that could enhance construction site safety. This stream adopted a dual approach: an extensive literature review followed by a market analysis of leading companies. Together, the findings informed the development of a taxonomy of potential technologies for detecting and mitigating safety risks, as well as sensing, tracking, and interpreting worker movement, activities, and behaviour. The analysis examined technologies across key operational and technical dimensions, including cost, precision, safety risks, data management, and privacy. These findings were iteratively refined in collaboration with industry stakeholders to ensure practical relevance. Informed by insights from Stream I, Stream II lays the groundwork for selecting and designing the most viable technological solutions for Stream III, along with the development and evaluation of a concept prototype.

Literature review methodology

After analysing the data and conducting the workshop in the previous stream, we identified one main target scenario: fall from height, which includes falls of persons and falling objects. To understand the latest research, technologies, and proven implementations related to fall-from-height incidents in the construction industry up to 2024, we conducted a scoping review.

This literature review was guided by the following research questions (RQs):

1. What technologies and techniques are used to collect and analyse data related to construction site safety?
2. What are the main applications of these technologies in enhancing construction site safety?
3. What are the key strengths and limitations of these technologies in the context of construction site safety?

Keyword identification

We began by exploring a broad range of relevant academic publications using Google Scholar. Our initial strategy involved testing high-frequency, general terms commonly associated with safety incidents in construction, such as “Fall from Height” and “Fall of Person”. While these keywords returned a large volume of results, further inspection revealed a lack of specificity and uneven relevance to our study focus. In particular, when cross-referencing these results with a set of pre-selected benchmark articles, we observed that many of the most pertinent studies employed more precise and context-specific terminology, such as “Fall from Ladder”, “Scaffold Fall”, or “Fall Prevention Measures”. This led us to reconsider the breadth and granularity of our search parameters. We based our keyword selection on the structured approach proposed by (Soltanmohammadlou et al., 2019), adapting and refining their keywords to fit the specific focus of our review. Through this iterative process, we developed a tailored list of search terms that reflected both common and specialised language used in the domain. This customised keyword set became the basis for our article selection protocol, and the full list of terms is presented in Appendix I.

Database selection

After finalising the keyword set specific to Fall from Height incidents on construction sites, we proceeded to apply these terms systematically in a targeted literature search. Our primary search platform was Scopus[1] database, selected due to its comprehensive coverage of peer-reviewed engineering and occupational safety journals, as well as its robust search and filtering capabilities. Scopus also supports repeatable, structured queries and ensures consistency in search outputs, making it suitable for high-quality systematic reviews. In the first round of searching, our query returned 538 articles, which formed the initial dataset.

Paper screening

Following the initial retrieval of 538 articles, we first removed two duplicates, leaving a total of 536 unique records for screening. A manual screening process was then conducted by our research team to ensure consistent evaluation. Each article was assessed based on its title, abstract, and keywords, and where necessary, a brief scan of the full paper was performed to determine its relevance. We applied a series of predefined inclusion and exclusion criteria to refine the dataset. Specifically, we only retained articles that offered full-text availability and that were clearly situated within the context of the construction industry. Articles that consisted solely of statistical reporting, such as listings of risk factor frequencies without any discussion or analysis of interventions, were excluded. We also excluded studies that did not examine fall accidents through the lens of intervention measures, such as engineering controls, administrative policies, or safety training. This rigorous filtering process excluded 384 articles, leaving 152 articles that met all screening criteria.

We conducted a full-text review of the 152 previously shortlisted articles, with a focus on evaluating the methodological depth, technological relevance, and practical applicability of each study. Articles were included if they were peer-reviewed journal publications or review papers, explicitly addressed fall accidents in construction settings, and described technologies that had been deployed either in laboratory environments or on actual construction sites to prevent fall-related incidents.

Conversely, we excluded papers that only proposed theoretical models or conceptual frameworks without demonstrating a prototype or evaluation. Studies were also excluded if they did not provide any clear link to detection, prediction, or prevention of fall accidents, such as those focused purely on general occupational health without addressing falls specifically. As a result, 102 articles were excluded, yielding a final selection of 50 high-quality articles relevant to our safety technology review.

[1] Scopus Literature Database: https://www.scopus.com/

Following the multi-stage screening and exclusion process, as illustrated in Figure 1, we identified a final set of 50 articles (see Table 1 for details). To systematically examine these articles, we developed a structured analytical framework. Our analytical framework applied nine dimensions corresponding to both the technological and contextual factors relevant to fall prevention on construction sites. For each of these dimensions, we predefined common categories or keywords and used them to populate columns in the analysis table. Each article was analysed and its relevant attributes were coded across these columns. This structured coding process allowed us to systematically identify patterns, gaps, and opportunities across the selected literature set. The results of this qualitative analysis are presented in detail in the following section.

Figure 1. A PRISMA flow diagram showing the flow of examined papers through different phases of the literature review

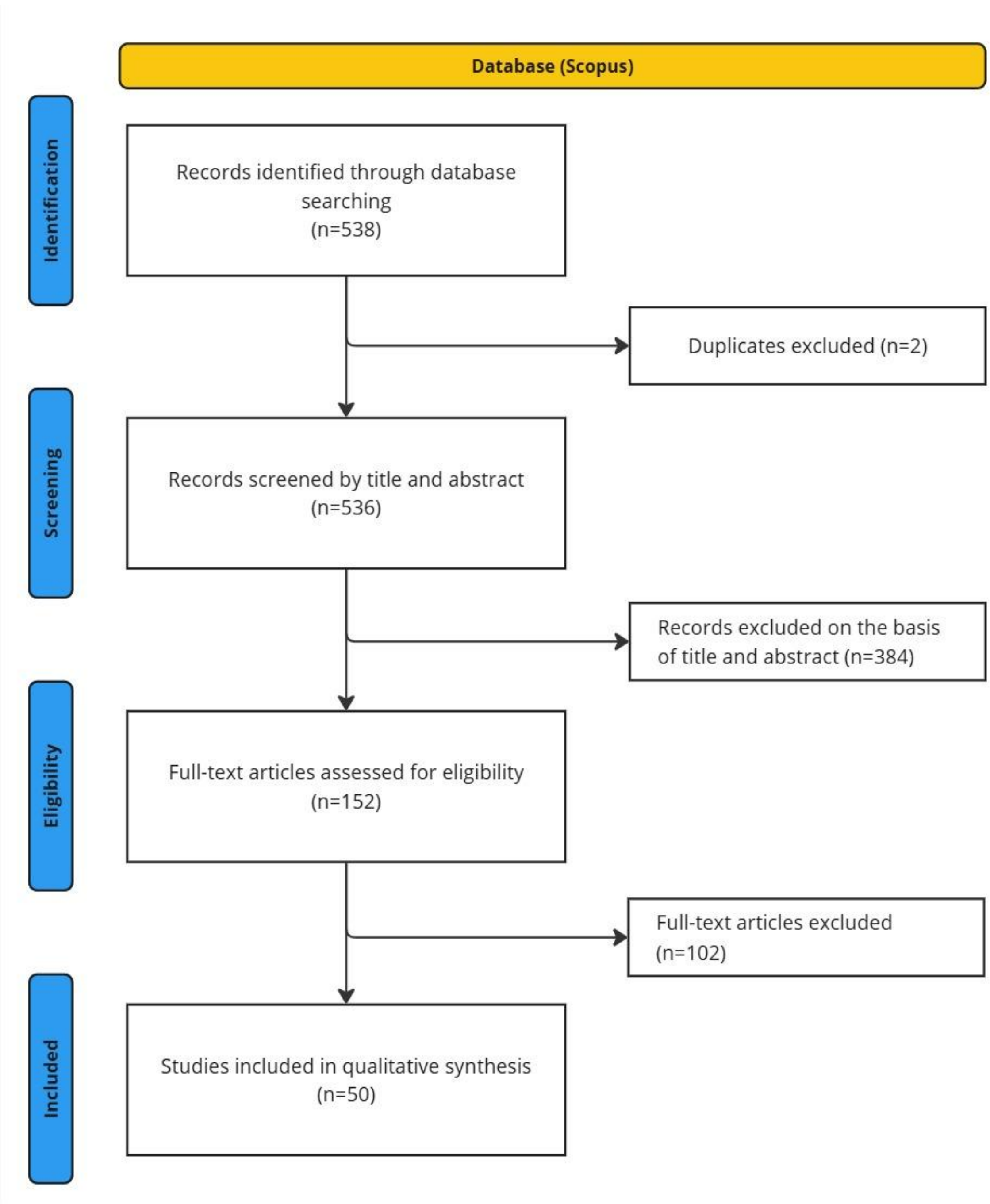


Table 1. Papers examined in the literature review

Analytics Methods	Sensor-based Technology	Camera-based Technology	Other Technology
Machine Learning Models	(Choo et al., 2023), (Rey-Merchán et al., 2022), (Dzeng et al., 2024), (Li et al., 2024)	(Liu et al., 2025), (Shetty et al., 2024)	
Deep Learning Models	(Song et al., 2023), (S. Lee et al., 2022), (H. Lee et al., 2022), (Zhang et al., 2019), (Khan, Khalid, Anjum, Khan, et al., 2022), (Khan, Khalid, Anjum, Tran, et al., 2022), (Ojha et al., 2023)	(Chen et al., 2023), (Park et al., 2023), (H. Guo et al., 2023), (Lee & Lee, 2023), (Huang et al., 2023), (Cheng et al., 2022), (Anjum et al., 2022), (Yang et al., 2022), (Lim et al., 2022), (Hu et al., 2022), (Zhang & Ge, 2022), (Khan et al., 2021), (Piao et al., 2021), (Chian et al., 2021), (Wu et al., 2021), (Shanti et al., 2021), (Fang et al., 2020), (Athidhi & Smitha Vas, 2023), (Yuhai et al., 2023), (Lin et al., 2024), (J. Li et al., 2022), (Shanti et al., 2022)	(R. Guo et al., 2023), (Chang et al., 2022),
Statistics Models	(Li et al., 2023), (Panuwatwanich et al., 2020), (Han & Wang, 2024),	(Duan et al., 2023), (Huang et al., 2024)	(Zhang et al., 2022)
Other	(Antwi-Afari et al., 2020), (Gomez-de-Gabriel et al., 2019), (Liu et al., 2020), (Jin & Gambatese), (Cheng et al., 2024)		(P. Li et al., 2022), (Johansen et al., 2024)

Note: Papers were categorised by main Data Source Technologies (Sensor-based, Camera-based, and Other) for each column and the main Data Analytics Methods (Machine Learning, Deep Learning, Statistics, and Other) for each row.

Findings from literature review

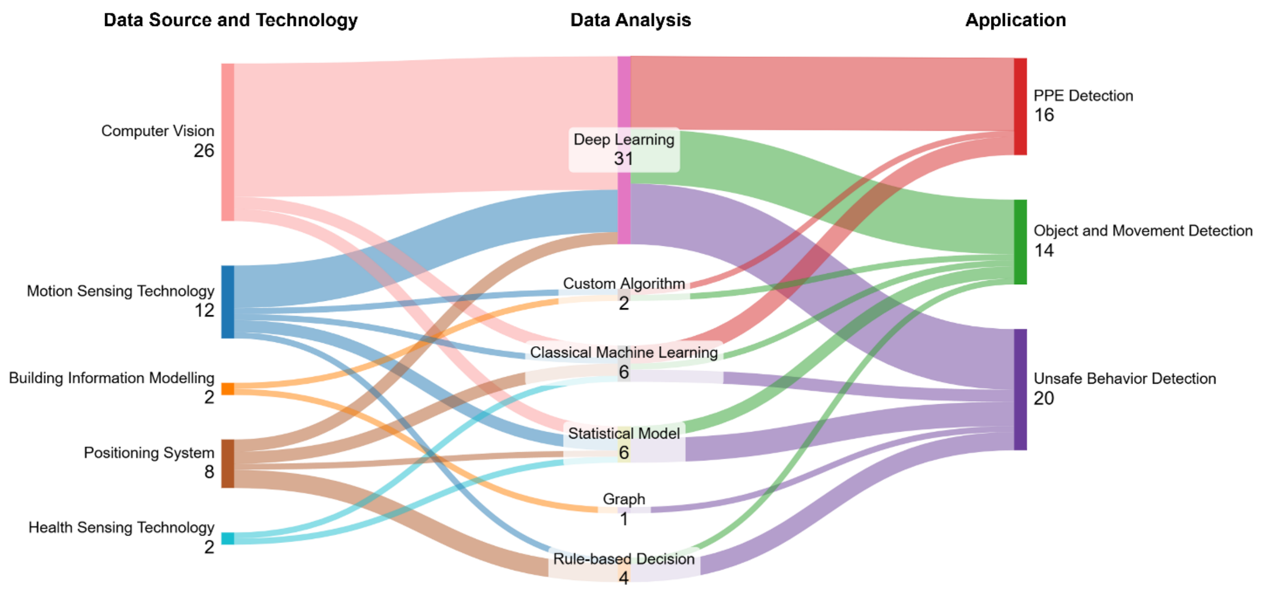
1. A review of technologies for fall from height prevention in construction

To synthesise the technological trends across the selected literature, we classified the technologies discussed in the 50 reviewed articles into three primary categories: Data Source and Technology, Data Analysis, and Application. This framework was designed to identify the technologies and methods used in each study to improve site safety. The results of this categorisation process are summarised in Figure 2.

The analysis revealed that Computer Vision emerged as the most widely used technology. Publications utilising computer vision often leveraged image or video feeds from construction sites to monitor worker activity and detect safety breaches. These data were predominantly analysed using Deep Learning models, which were applied to classify visual features, recognise patterns, and infer potential risks. In addition to computer vision, other frequently used technologies included Motion Sensing technologies (e.g., wearable accelerometers, inertial sensors) and Positioning Systems (e.g., GPS, RFID, UWB).

Overall, five categories of technologies were identified. These technologies have been applied in a wide range of safety-related applications on construction sites. Specifically, they have been used to detect compliance with Personal Protective Equipment (PPE) requirements, identify object locations, track worker and equipment movements, and detect unsafe or anomalous behaviours in real time. This review highlights a clear trend toward the use of computer vision and sensing technologies, integrated with advanced machine learning methods, to enable detection, prediction, and proactive safety interventions in dynamic site environments.

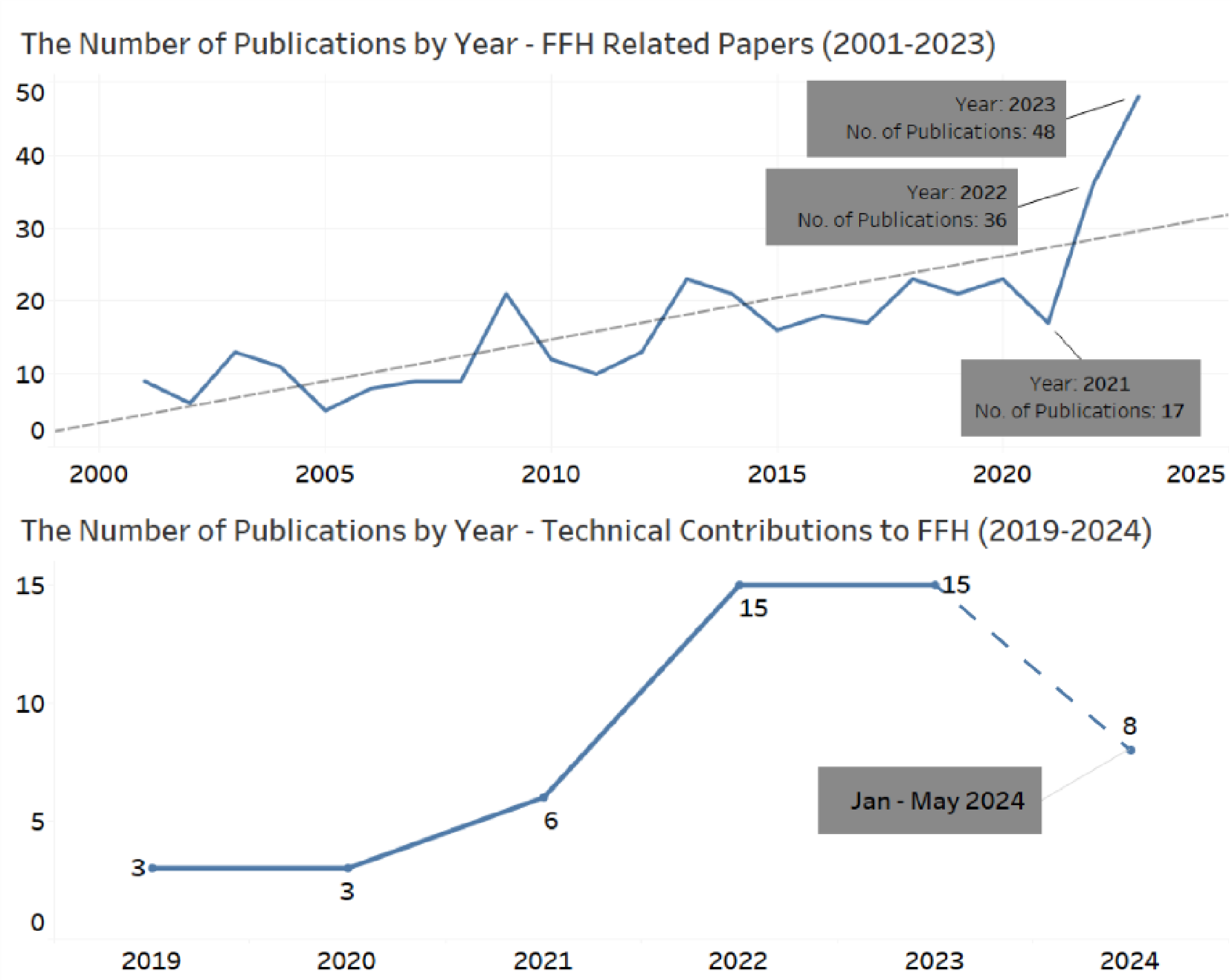
Figure 2. A Sankey diagram to show the distribution of selected papers with their categories in three aspects



2. Yearly distribution of publications related to falls from height

Historically, the volume of academic publications specifically addressing Falls from Height (FFH) in the construction industry has been relatively low. However, our review reveals a noticeable upward trend in recent years, as illustrated in Figure 3. This growth can be attributed to the increasing adoption of advanced, AI-assisted technologies that have significantly enhanced the ability to detect and analyse fall-related risks. The rise of these methods has been supported by parallel advances in Artificial Intelligence (AI) and Generative AI, improvements in computing infrastructure, including more affordable and powerful hardware and cloud-based platforms, and the broader availability of large-scale public datasets for training robust object detection and predictive models. This growing interest reflects a shift in both industry and research communities toward proactive, technology-driven approaches to construction safety.

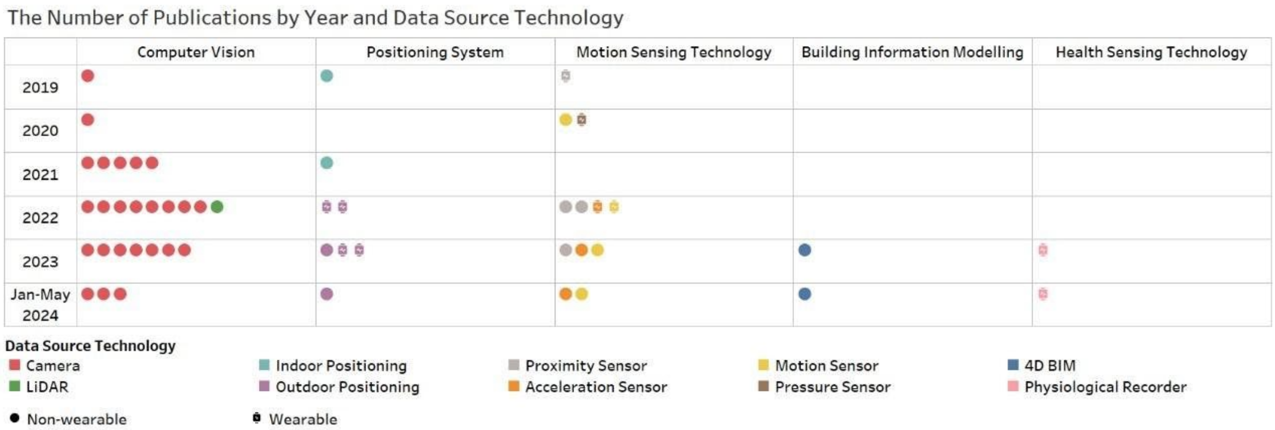
Figure 3. A line chart illustrates the yearly trend of the number of publications in the related field of Fall From Height by year



3. Data sources and technologies

As illustrated in Figure 4, our analysis of the selected publications highlights several notable trends in the use of technology for detecting and mitigating Fall from Height (FFH) incidents in construction settings.

Figure 4. A dot plot illustrates the distribution of publications in each Data Source Technology across years from 2019 to 2024



Camera-based computer vision technologies — Camera-based computer vision has emerged as the dominant data source for FFH detection. A significant increase in related publications has been observed since 2021, driven by the growing maturity of object detection and activity recognition algorithms, as well as the wider availability of affordable cameras and edge-computing devices. These technologies have been primarily used for monitoring PPE compliance, detecting hazardous postures, and tracking human-object interactions on-site.

Outdoor vs. indoor positioning systems — Our review also reveals that research on outdoor positioning systems, such as GPS and UWB (Ultra-wideband), is more prevalent than that on indoor systems. This disparity may stem from the logistical and financial complexity of implementing indoor positioning in high-rise or multi-storey buildings. In such environments, installing and maintaining location tracking infrastructure (e.g., sensors and cameras) on each floor is a resource-intensive process. Furthermore, the dynamic and constrained nature of indoor construction environments introduces further challenges in maintaining sensor accuracy and network stability.

Wearable and motion-based sensors — Several studies utilise motion, proximity, and acceleration sensors to monitor worker movement patterns, aiming to detect risky behaviours or pre-incident indicators related to falls. These devices are predominantly worn on the body (e.g., wrist, waist, or helmet) while some also include sensors deployed in the surrounding environment. Motion-based systems face similar challenges to positioning systems in terms of deployment and maintenance. Additionally, the use of wearables and motion sensors could raise privacy concerns.

Limited application of BIM— Although Building Information Models (BIMs) hold promise for integrated safety planning, the review shows that only a small number of studies apply BIM to Fall from Height (FFH) scenarios. This limited adoption is likely due to the proprietary nature of most BIM software, which imposes constraints on extensibility and interoperability.

The difficulty and cost of customising BIM platforms to support real-time safety analytics also pose significant barriers.

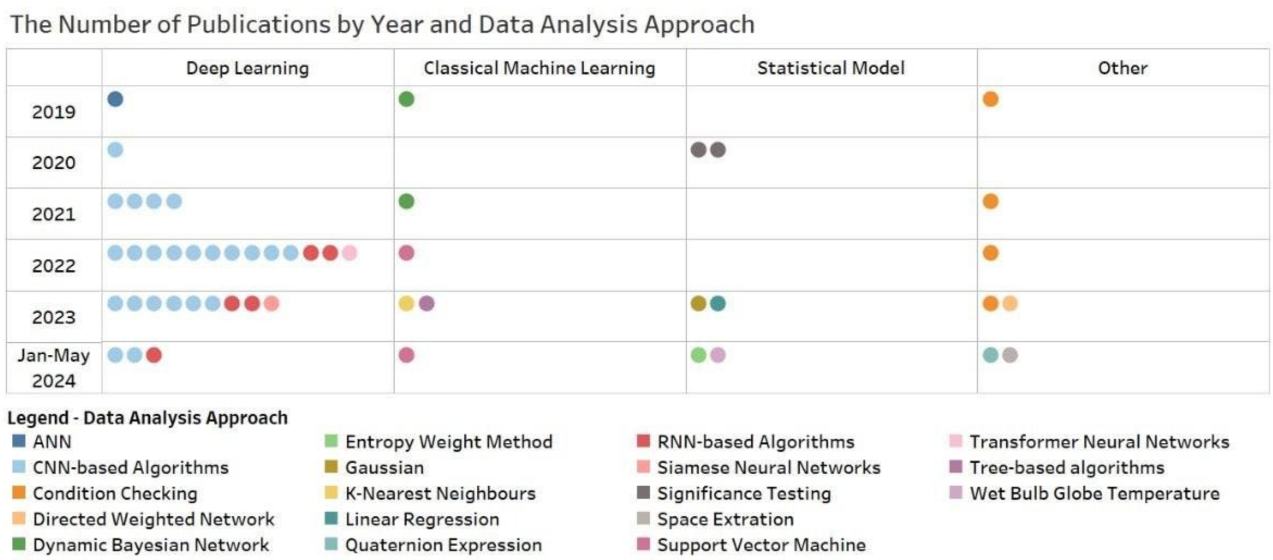
Health monitoring — Lastly, we identified a few publications involving health monitoring sensing technologies for FFH prevention. This scarcity may be due to significant privacy concerns, especially within the Australian context, where strict privacy regulations and strong worker protections could limit the adoption of such technologies, especially when it involves collecting biometric or physiological data, such as heart rate or fatigue levels.

4. Data analysis methods

The data analysis methods employed in studies related to Fall from Height (FFH) detection have undergone significant evolution over the past five years. As shown in Figure 5, Convolutional Neural Networks (CNNs) are the predominant method across the reviewed papers as they are well-suited for analysing image-based data collected from surveillance cameras, the most common data source in FFH studies.

In contrast, Recurrent Neural Networks (RNNs) and transformers (a type of neural network architecture) have been used less frequently in current FFH studies. These models are better suited for processing and modelling sequences and time-series data like motion sensor outputs. RNNs also typically require larger volumes of training data and greater computational resources, which may further limit their use.

Figure 5. Dot plot illustrating the distribution of publications in each Data Analysis Technology across years from 2019 to 2024



Our review did not identify any studies applying Natural Language Processing (NLP), which is primarily used with text data. There was also a notable absence of studies using Large Language Models (LLMs), which have emerged more recently. This gap reflects the relatively slow adoption of these technologies within the construction safety research domain. However, with the rising popularity of LLMs since 2022 and their growing ability to process multimodal data, including incident reports, safety manuals, and annotated images, we anticipate this landscape will evolve in the near future.

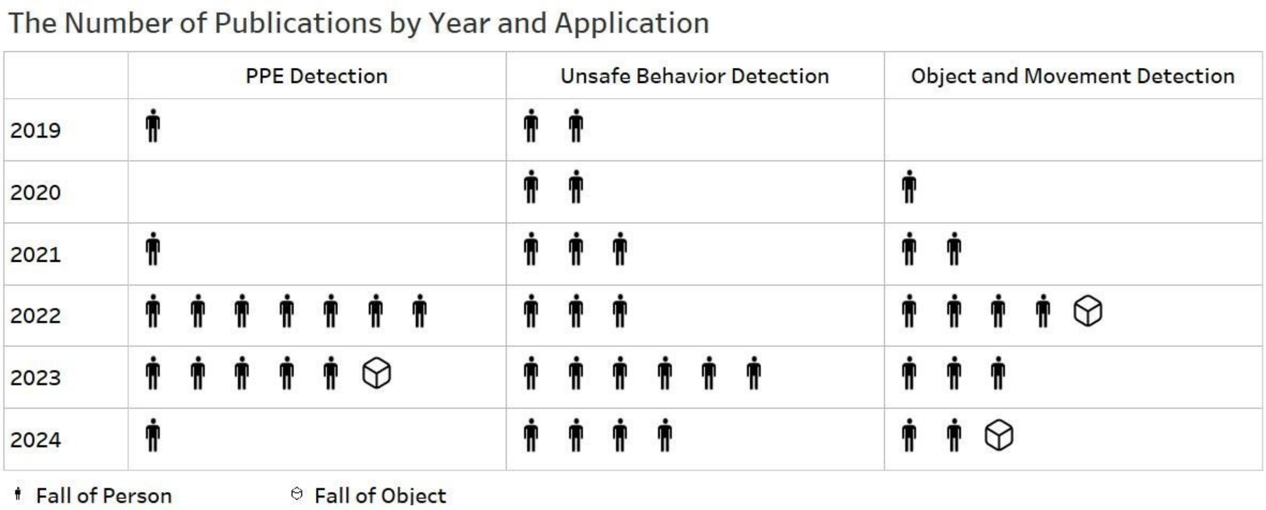
In summary, while CNN-based methods currently dominate FFH research, the industry has begun gradually adopting emerging technologies. At the same time, advances in research and technology continue to open pathways that may help overcome privacy challenges, paving the way for wider adoption of computer vision techniques. This adoption is expected to accelerate alongside the advancing capabilities of Large Language Models (LLMs) and Vision Language Models (VLMs), particularly in visual reasoning. Together, these developments promise more sophisticated and proactive approaches to addressing FFH risks and existing research gaps.

5. Application areas

An analysis of the application domains of the selected publications reveals distinct research priorities and emerging trends in fall prevention technologies within the construction industry (see Figure 6). Most studies focus on detecting or preventing falls of persons, while comparatively fewer publications address falls of objects. However, industry reports, supported by feedback from our industry partners, have increasingly emphasised that falls of objects are a common and serious hazard. Tools such as hammers or tape measures may fall through floor openings or gaps in temporary railings, potentially causing serious injuries.

This gap between the low number of research studies and the high industry concern may stem from technological and operational challenges. Studies on detecting falling objects, such as hand tools or debris, on multi-level construction sites is often limited by difficulties in installing cameras across different floors, temporary structures, and scaffolded areas.

Figure 6. Dot plot illustrating the distribution of publications in each Application Use Case Scenario across years from 2019 to 2024



Note: Two icons are used to distinguish two different types of applications (Fall of Person or Fall of Objects).

The review identified another popular application area in FFH safety, focusing on detecting PPE compliance. Over the last three years, researchers have begun to explore vision-based technologies and deep learning models to automatically detect whether workers are correctly wearing helmets, harnesses, or high-visibility vests. These technologies are often deployed as proactive systems, aiming to identify PPE violations before incidents occur.

This line of research aligns with broader industry efforts to strengthen behavioural safety and reduce human error.

While detecting unsafe behaviours, such as climbing without harnesses or entering restricted zones, remains a popular area of study, its adoption is limited by ethical and privacy concerns. Developing privacy-preserving approaches for sensing and monitoring human activity and behaviour could significantly benefit construction site safety research. Future work would benefit from defining and investigating various privacy-preserving and non-intrusive monitoring technologies to address both safety and privacy challenges in construction environments.

6. Scarcity of on-site evaluations

Among the 50 reviewed publications, only a small subset (7 studies) reported conducting evaluations directly on active construction sites. This highlights a significant gap between research-based validation and real-world, on-site evaluation. While many studies demonstrate promising results in controlled or simulated environments, field evaluations are critical for assessing the robustness, scalability, and practical challenges of deploying these technologies under the dynamic and unpredictable conditions of construction sites. This shortfall suggests a need for future research to prioritise on-site experiments and evaluations with industry partners to ensure that proposed solutions are both technically sound and operationally feasible in live settings.

Market analysis of construction site safety systems

As part of the market review, we examined 23 leading international companies that provide technology solutions for improving construction site safety (a detailed list can be found in Table 2). This review was based on publicly available sources, such as official websites and product documentation, to evaluate industry capabilities and identify the services and products offered in the construction sector.

Our review shows that CCTV is the most commonly used technology among these companies. This popularity is primarily driven by CCTV's affordability, widespread availability, and ease of deployment and integration into existing site infrastructure. However, while most companies capture video data, only about one-third offer data analytics services including real-time risk detection, behavioural analysis, and predictive insights. Most solutions are limited to manual monitoring or passive data storage, limiting the potential of these systems to deliver proactive safety interventions.

Across several companies offering advanced solutions, we identified five prevalent service offerings that characterise the current market landscape:

- Real-time Violation Detection – Automated systems to detect non-compliance with PPE requirements or unsafe proximity to moving equipment.
- Real-time Risk Alerts – Notification systems, typically tied to CCTV, that flag high-risk situations as they occur.
- Wearable Motion Sensors – Devices that monitor posture, movement, or sudden deceleration, often used to detect falls or unsafe behaviour.

- Digital Safety Administration – Tools for managing site inductions, checklist compliance, and daily safety reporting.
- Incident Tracking and Management – Platforms to log, investigate, and analyse safety incidents to support continuous improvement.

Among the companies reviewed, Sightdata[2], a leading Australian provider and partner of Building 4.0 CRC, offers fall detection for persons, along with services such as PPE detection, remote site inspection, exclusion zone monitoring, and loose object detection, powered by computer vision and machine learning models.

The review identified only a single provider, Benchmark Gensuite[3], offering Augmented Reality (AR) and Virtual Reality (VR) solutions. This finding indicates that immersive technologies are still in the early stages of development and adoption within the construction sector. Given their capabilities to enable virtual simulations and create highly immersive experiences for education, and training, this technology represents a promising opportunity for developing effective solutions to improve construction site safety, particularly in relation to falls from height.

Overall, these insights suggest that while CCTV remains the most commonly used technology in the construction industry, there is still a gap in advancing data-driven and AI-assisted systems to integrate emerging and potential technologies such as Large Language Models (LLMs), Visual Language Models (VLMs), and immersive virtual technologies. The next generation of innovative safety solutions in construction will require bridging this gap through the integration and convergence of these technologies, while ensuring the protection of worker privacy.

[2] Sightdata Australia: https://sightdata.ai/

[3] https://benchmarkgensuite.com/

Table 2. International companies that provide services for construction site safety

Company Name	Location	Website
Sightdata	Australia	https://sightdata.ai/
Sensera Systems	USA	https://www.senserasystems.com/sitecloud-analytics/
Blackline Safety	Global, Australia	https://www.blacklinesafety.com/solutions/software/blackline-live
Triax	USA	https://www.triaxtec.com/use-cases/emergency-management-software/
Guardhat	USA	https://www.guardhat.com/software/
Predictive Safety	USA	https://predictivesafety.com/alertmeter/
Predictive Safety-PRISM	USA	https://predictivesafety.com/prism-fatigue-management-system/
Zebra	USA	https://www.zebra.com/ap/en/solutions/industry/manufacturing.html
SmartCap	Canada	https://www.smartcaptech.com/life-smart-cap/
EcoOnline	Canada	https://www.ecompliance.com/construction/
Avetta	USA	https://www.avetta.com/en-au/workforce-management
SiteDocs	Global, Australia	https://www.sitedocs.com/
SoterAnalytics	Global, Australia	https://soteranalytics.com/the-soter-platform
Samsara	Global	https://www.samsara.com/products/site-visibility/
VelocityEHS	Global, Australia	https://www.ehs.com/solutions/safety/
SafetyCulture	Global, Australia	https://safetyculture.com/construction/
Benchmark Gensuite	USA	https://benchmarkgensuite.com/solutions/environmental-health-safety-software/
Near-Miss Management	USA	https://www.nearmissmgmt.com/products/
SafetyLine	Canada	https://safetylineloneworker.com/lone-worker-safety-devices
Inviglio AI	Singapore	https://invigilo.ai/
HammerTech	Australia	https://www.hammertech.com/
skytrust	Australia	https://skytrust.com.au/construction-risk-assessment-software/
Plinx	UK	https://plinx.io/

STREAM III: CONCEPT PROTOTYPE DESIGN AND TESTING

The project was extended by one year to include Stream III, which focused on designing and developing concept prototypes to address safety risks related to falls from height, identified as a priority area by the industry partner. This section begins with an overview of falls from height, followed by a definition of the use case scenarios identified by the industry partner. It then provides a comprehensive description of the prototype's design, development, components, and its qualitative evaluation through interviews.

Falls from height: use case context and statistics

Falls from height (FFH) continue to be one of the most persistent and hazardous risks on construction sites. Falls from height can be broadly classified into two categories: falls of persons and falls of objects. Falls of persons involve workers falling from elevated surfaces such as ladders, roofs, scaffolding, or platforms. Falls of objects refer to incidents where tools, materials, or equipment fall from a height and strike individuals below, posing significant injury risks to those working at ground level or in lower areas.

Data from the latest Safe Work Australia Work Health and Safety Statistics 2024 report[4] shows that falls from height were the second leading cause of workplace fatalities in 2023, marking a 71% increase compared to the previous year and a 32% rise above the five-year average. In 2023, the construction industry accounted for 45% of worker deaths caused by falls from height. According to the 2024 Construction Annual Health and Safety Statistics[5] for the UK, falls from height accounted for 52% of construction-related deaths over the same five-year period. Based on 2022-23 figures, being struck by falling objects was one of the leading causes of construction fatalities in the UK, accounting for 10% of deaths.

According to our industry partner, falling objects were proposed as a potential use case for developing the concept prototype, as they pose a significant hazard in construction environments, particularly in high-rise and vertical work settings where loads are frequently lifted, transported, or stored at elevated heights.

Use case scenarios

Our industry partner identified three specific use case scenarios related to perimeter screens that contribute to falling object incidents in high-rise building construction. These three scenarios informed the design of our concept prototype, ensuring it directly addressed the key safety challenges identified by our industry partner in real-world high-rise construction environments.

Safety nets are critical protective elements in construction sites, building maintenance, and other hazardous environments with a risk of falling objects or personnel. Regular inspection of these nets is essential to ensure their integrity and compliance with safety regulations.

[4] https://data.safeworkaustralia.gov.au/insights/key-whs-statistics-australia/latest-release

[5] https://www.hse.gov.uk/statistics/assets/docs/construction.pdf

Traditional manual inspection is time-consuming, subjective, and potentially hazardous, as inspectors often need to navigate risky areas to assess.

1. **Gaps between perimeter screens**

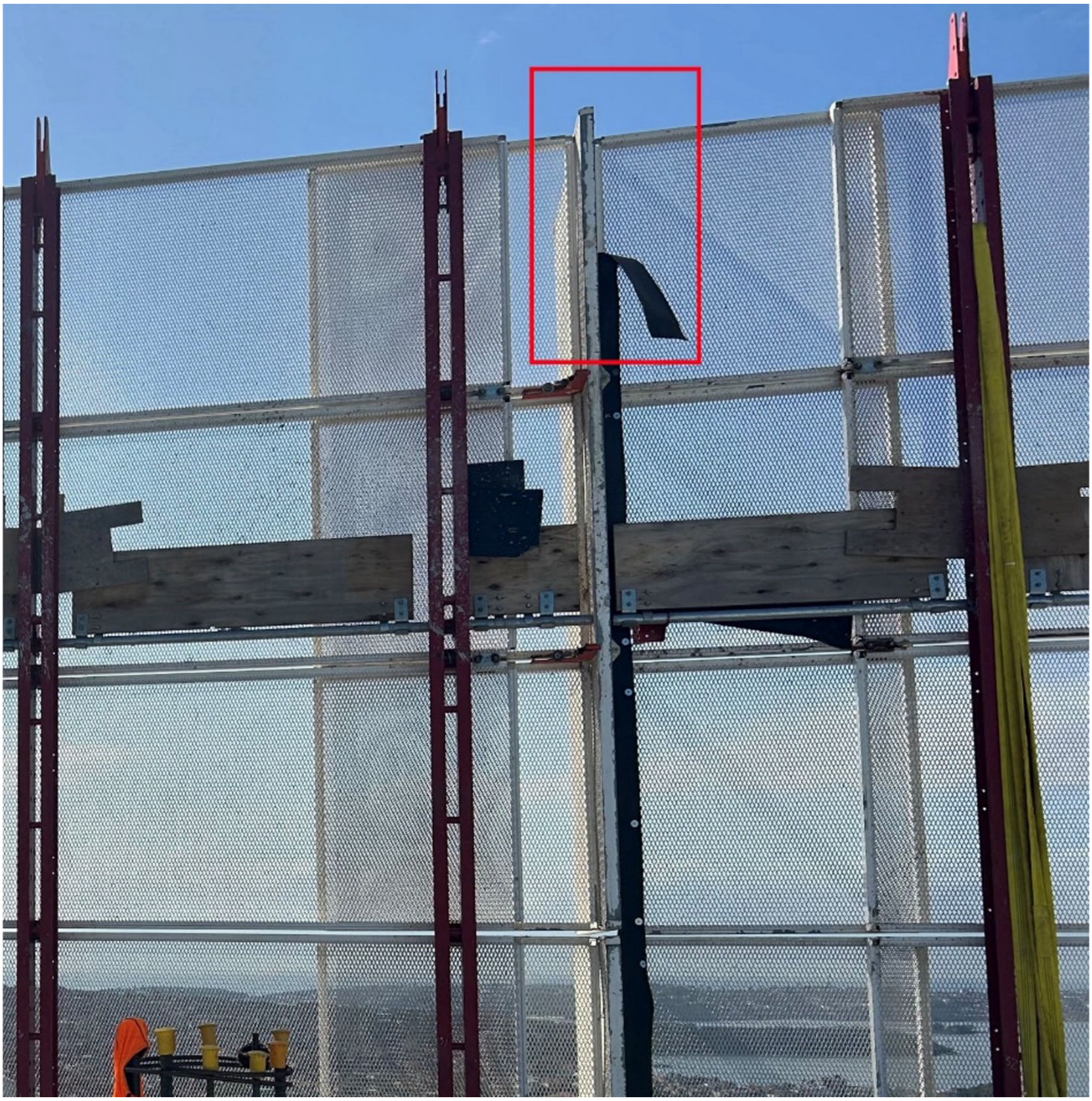
The first scenario involves falling objects passing through gaps between safety perimeter screens installed around the edges of elevated work zones (see Figure 7). These screens are designed to prevent tools, debris, and construction materials from falling; however, small gaps can occur during installation or due to environmental factors such as wind or thermal expansion. These openings can lead to falling object incidents, particularly involving smaller items such as hand tools, bolts, cable reels, and off-cuts. Despite their size, such objects pose serious risks to workers and pedestrians below. Current mitigation relies heavily on manual visual inspection, which is both labour-intensive and time-consuming. Automated detection of these gaps could support inspection activities, enhance efficiency and productivity, and reduce human error.

Figure 7. Photo showing a gap between perimeter screens

2. **Loose stripes on perimeter screens**

The second scenario focuses on the condition of black or infill stripes, which are commonly used on safety perimeter or mesh screens to seal minor gaps (see Figure 8). While these stripes serve an important function, they can become partially detached due to weather exposure or repeated handling. Once loosened, they may flap in the wind or detach completely, reintroducing gaps. Detecting loosely secured or missing stripes is especially important in high-traffic work areas or on floors where loose materials are frequently handled near the edge.

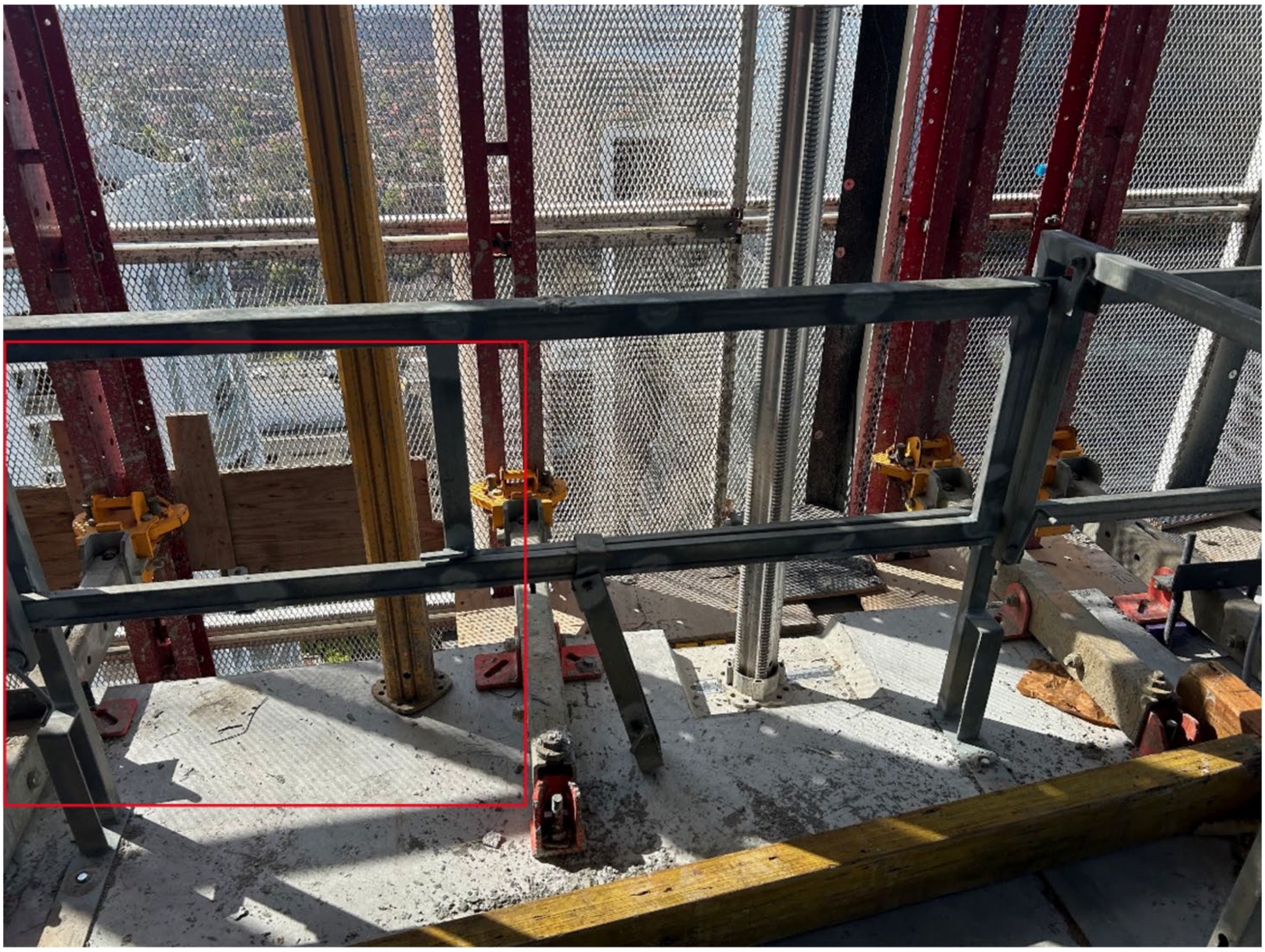
Figure 8. Photo showing a loose stripe attached to perimeter screens.



3. Holes or openings on the ground surface

The third scenario involves floor-level hazards, particularly openings or holes in the working surface (see Figure 9). These gaps are a common cause of falling object incidents, as tools or materials can drop through unprotected areas, and they also pose a significant trip and fall risk for workers. When located near active work zones, undetected floor openings increase the likelihood of objects falling to lower levels. An automated system capable of detecting such hazards could greatly enhance safety by providing real-time alerts when gaps are present or when floor covers are missing or displaced.

Figure 9. Photo showing a hole near the edge of the ground

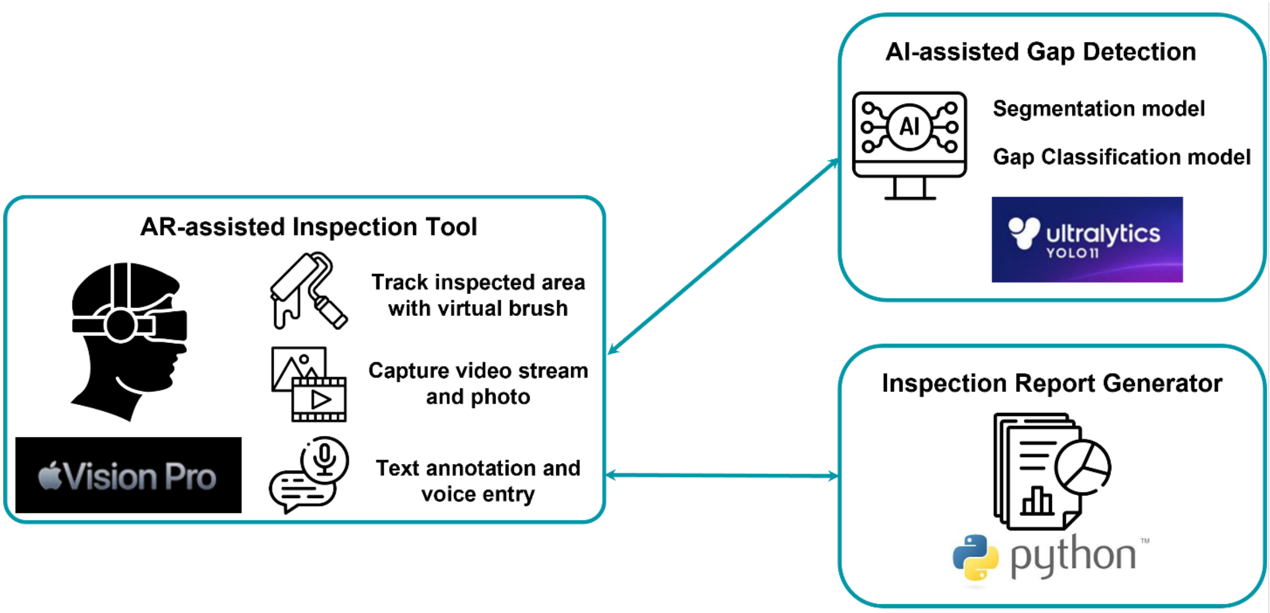


These three scenarios represent common and recurring safety concerns in multi-storey construction environments. By focusing on these use cases, we aimed to design and develop innovative solutions that are not only technically feasible but also closely aligned with the day-to-day operational needs of construction teams. Moreover, addressing these risks contributes to addressing a significant safety gap within the industry.

PROTOTYPE: INTEGRATION OF VISUAL MODELS INTO AUGMENTED REALITY FOR SAFETY

Building upon the outcomes of the preceding streams: Stream I (data-driven analysis of historical safety observations and incidents from the industry partner) and Stream II (scoping literature review and global market technology assessment), and the identification of three use case scenarios in Stream III, we developed an integrated solution prototype to improve site safety in high-rise building construction. This prototype comprises three interlinked modules: (1) an AI-assisted gap detection module designed to automate and classify safety-related gaps in perimeter screens using advanced computer vision models; (2) an AR-assisted inspection tool that supports on-site users through immersive interaction, real-time annotation, and hands-free documentation; and (3) an inspection report generator that consolidates all visual, textual, and analytical data into structured summaries for subsequent review and compliance tracking. The architecture of this prototype system is depicted in Figure 10, illustrating the functional interconnection between modules and the underlying technologies adopted in each.

Figure 10. Illustration of the three modules in our prototype, including AI-assisted Gap Detection, AR Assisted Inspection Tool, and Inspection Report Generator



Module 1 – AI-assisted gap detection

One of the key features of our prototype was its ability to automatically and proactively detect three types of safety risks, primarily related to gaps in the perimeter screens, as described in the three use case scenarios. To this end, we explored and developed two different solutions using computer vision technologies. Computer vision offers a promising approach to automate and enhance safety risk detection during inspections. Recent advances in deep learning, particularly in object detection and instance segmentation, have made it possible to develop systems that can accurately identify and classify objects in complex environments.

Module 1 aimed to address the need for automated, efficient, and reliable detection of safety nets through computer vision and deep learning techniques, providing a foundation for safer working environments and more effective safety compliance.

Module 1 consisted of two solutions. The first solution focused on a segmentation problem, while the second solution, the more comprehensive one, involved the detection of six types of safety risks, offering a more detailed classification of the three defined safety risk scenarios.

The primary objectives of this module were to:

1. Develop a robust computer vision model capable of identifying and classifying different types of safety nets in images with high accuracy
2. Evaluate the system's performance on real-world safety net images under various conditions
3. Create a practical tool that can assist safety inspectors in their assessment tasks by providing visual documentation and automated analysis.

Solution 1: Segmentation Model

The first solution involved the development and evaluation of an automated system for safety net segmentation using state-of-the-art computer vision model architectures. Employing the You Only Look Once (YOLO) model (YOLOv8[6]) as our segmentation model, we successfully identified and classified three types of safety nets in various real-world scenarios. The system achieved 62.3% mean Average Precision (mAP@0.5) in segmentation, demonstrating its potential for enhancing safety inspection processes through computer vision.

1.1 Overview

We selected YOLO as our architecture due to its state-of-the-art performance in object detection and instance segmentation tasks. YOLO developed by Ultralytics[7], represents the latest evolution in the YOLO family of models, offering improvements in both accuracy and inference speed compared to previous versions. YOLO employs several key components that contribute to its effectiveness. The model uses a multitask learning approach, jointly optimising for object detection and instance segmentation, which leads to more efficient training and better overall performance.

Our approach employed the YOLO instance segmentation model, which we trained on a custom dataset of safety net images. YOLO represents a family of real-time object detection algorithms that has evolved through several iterations, with recent versions offering

[6] https://yolov8.com/

[7] https://www.ultralytics.com/

significant improvements in accuracy and efficiency. The model was configured to identify three distinct safety nets (fence safety nets, blue safety nets and barricade safety nets) with pixel-level accuracy, enabling precise localisation and classification.

Instance segmentation extends beyond traditional object detection by identifying exact pixel boundaries of each object instance, rather than just providing bounding boxes. This capability is particularly valuable for safety net inspection, as it allows for more detailed analysis of net condition and coverage.

This module successfully delivered a trained YOLO segmentation model capable of identifying three types of safety nets with 0.623 mean Average Precision (mAP@0.5), demonstrating robust performance across various real-world scenarios. The visualisation system we developed highlights different types of safety nets using distinct colour coding, making it intuitive for safety inspectors to quickly identify and differentiate between net types. Additionally, this project lays a solid foundation for future work in automated safety inspection, including potential extensions for gap detection and integration with mobile applications for field use.

1.2 Dataset and annotations

The quality and characteristics of the training dataset significantly impact the performance of deep learning models. The data for building this model was collected using mobile phones with standard cameras from active construction sites, ensuring the dataset represents realistic field conditions. The collection was performed during daylight hours across multiple days to capture varying lighting conditions. The industry partner provided the dataset. Permission was obtained from site managers before data collection to ensure compliance with safety and privacy requirements.

The dataset consisted of 204 images of construction sites with safety nets in various configurations. The photos represented diverse real-world scenarios, including varying lighting conditions, camera angles, and net types. This diversity is crucial for ensuring the model can generalise well to new, unseen environments during deployment.

The annotation process involved marking three distinct classes of safety nets, each serving a specific safety function in construction environments:

1. **Fence Safety Nets:** These are safety nets installed at the edge of floors to prevent falls from heights. They typically appear as vertical installations along perimeters.
2. **Blue Safety Nets:** These cover equipment and materials. Their distinctive blue colour makes them relatively easier to identify visually than other net types.
3. **Barricade Safety Nets:** These are barricades installed on the floor before the fence safety nets, creating an additional safety zone. They serve as a visual and physical barrier to warn workers of proximity to edges.

Annotations were created using Computer Vision Annotation Tool (CVAT), providing pixel-level segmentation masks for each type of safety net. The annotation process followed the COCO format, including bounding box coordinates and segmentation polygons.

The dataset was organised in a standard format compatible with the YOLOv8 training pipeline to ensure efficient model training and evaluation:

A YAML configuration file defined the dataset structure, including paths to training and validation sets, and class definitions. Images were split into training (70%), validation (20%), and testing (10%) sets, ensuring balanced representation of all class types across the splits. Each image had corresponding annotation files with instance segmentation masks, and class mappings were defined in the configuration file to maintain consistency throughout the pipeline.

Figure 11. Dataset organisation for training, validation, and testing

```
dataset
├── data.yaml
├── images
│   ├── test  [21 jpg files]
│   ├── train  [143 jpg files]
│   └── val  [40 jpg files]
└── labels
    ├── test  [20 txt files]
    ├── train  [143 txt files]
    ├── val  [40 txt files]
```

Note: The dataset is organised into separate directories for training, validation, and testing, with each containing both images and their corresponding labels. The structure follows standard conventions for YOLO format.

1.3 Methods

The selection of an appropriate model architecture is critical for achieving high performance in instance segmentation tasks. We used the YOLOv8m-seg variant with the following configuration:

- Input resolution: 640×640 pixels
- Backbone: CSPDarknet
- Number of classes: 3 (corresponding to the three types of safety nets)
- Loss functions:
    - Classification: Binary Cross-Entropy (BCE)
    - Localisation: Complete Intersection over Union (CIoU)
    - Segmentation: Binary Cross-Entropy with Dice loss.

The training process is essential for developing a model that generalises well to unseen data. The model was trained using a single-stage approach that leveraged transfer learning to improve performance.

First, we initialised the model with pre-trained weights from the COCO dataset, which contained a wide variety of common objects and provided a strong foundation for feature extraction. The model was then fine-tuned on our safety net dataset, allowing it to adapt its learned features to the specific characteristics of safety nets. We employed a 70/20 train/validation split (143 images for training, 40 images for validation), ensuring that the validation set represented the same distribution of environments and net types as the training set to provide a realistic assessment of model performance.

This transfer learning approach significantly reduced the training time and improved performance compared with training from scratch, as the model could leverage general visual features learned from the larger COCO dataset.

The training was conducted on Ubuntu 24.04 LTS with a Quadro RTX 6000 with 24 GB Graphics Card, employing the Ultralytics YOLO implementation. The complete training process required approximately 1 hour and 7 minutes, demonstrating the efficiency of the YOLOv8 architecture and the Ultralytics implementation.

The hardware configuration provided sufficient computational power and memory to train the model efficiently, while the Ultralytics framework offered a streamlined implementation with comprehensive monitoring and visualisation capabilities.

The model was trained with the following hyperparameters:

- We used a batch size of 8 to balance memory usage and training stability.
- The learning rate was set to 0.01 with a cosine annealing schedule that gradually reduces the learning rate over time, helping the model converge to a better solution.
- The optimiser was Stochastic Gradient Descent (SGD) with momentum 0.937 and weight decay 0.0005, which helps prevent overfitting.
- Training ran for 100 epochs with early stopping patience of 50, meaning training would stop if no improvement was observed for 50 consecutive epochs.
- Automatic Mixed Precision (AMP) was enabled for faster training, allowing the model to use lower-precision calculations where appropriate without sacrificing accuracy.

These parameters were selected based on best practices for training YOLO models and adapted to our specific task and dataset characteristics.

We monitored the following metrics during training to assess model performance and ensure proper convergence:

Loss components (classification, localisation, segmentation) were tracked to ensure all aspects of the model were learning effectively. Mean Average Precision (mAP) at different IoU thresholds provided a comprehensive measure of detection and segmentation accuracy. Precision-Recall curves visualised the trade-off between these two important metrics. The F1 score offered a balanced measure of precision and recall, providing a single metric for overall performance assessment.

The model converged successfully, reaching stable performance on the validation set within the allocated training time.

1.4 Experiments

This section describes the experimental setup used to evaluate the model's performance, including evaluation metrics, testing protocols, and implementation details.

For segmentation performance, we used the following standard metrics to comprehensively evaluate our model:

- Mean Average Precision (mAP) at IoU thresholds of 0.5 and 0.5:0.95 were used as the primary metrics.
  - mAP@0.5 measures performance at a relatively lenient IoU threshold, while mAP@0.5:0.95 averages performance across multiple IoU thresholds (0.5 to 0.95 in 0.05 increments), providing a more stringent evaluation.
- Per-class Average Precision (AP) identified strengths and weaknesses in detecting specific net types, allowing for targeted improvements.
- Precision (P) and Recall (R) curves visualised the trade-off between these metrics across different confidence thresholds.
- The F1 score provided a balanced measure of precision and recall, combining them into a single metric.

These metrics are standard in segmentation tasks and provide a comprehensive view of model performance across different aspects, enabling thorough evaluation and comparison with other approaches.

We evaluated the model on the test set consisting of 21 images (approximately 10% of the dataset) not used during training or validation. The test set represented a diverse range of scenarios similar to those expected in real-world deployment, ensuring that our evaluation results would be indicative of actual performance in the field.

The evaluation was performed using the Ultralytics YOLO evaluation tools, which automatically calculate all the relevant metrics based on the model predictions and ground truth annotations. This approach

1.5 Results and discussion

This section presents the results of our experiments and discusses their implications, including both quantitative performance metrics and qualitative analysis of the model's behaviour.

The performance of our safety net segmentation model was assessed through both quantitative metrics and visual analysis, providing a comprehensive understanding of its capabilities and limitations.

Quantitative results

The YOLOv8 model achieved substantial segmentation performance on the test set, directly addressing our first project objective of developing a robust segmentation model capable of identifying different types of safety nets. These results demonstrate the model's ability to accurately detect and classify multiple safety net types across various real-world scenarios.

Figure 12. Class distribution across dataset splits

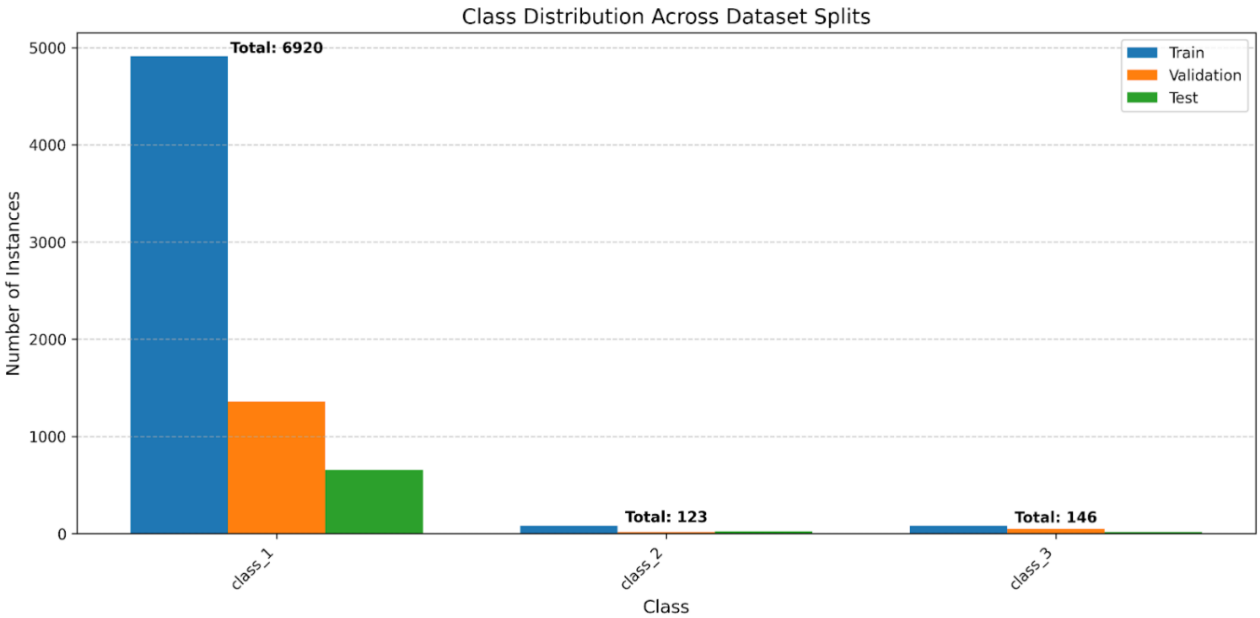


Figure 12 illustrates the class distribution in the test dataset, with Fence safety nets (class_1) being the most common with 653 instances across 20 images, followed by Blue safety nets (class_2) with 21 instances in 8 images, and Barricade safety nets (class_3) with 17 instances in 4 images.

The test dataset consisted of 21 images containing 691 instances of safety nets, with a significant imbalance in class distribution. Class_1 (Fence safety nets) dominates the dataset with 653 instances appearing in 20 images, while class_2 (Blue safety nets) and class_3 (Barricade safety nets) are much less represented with 21 and 17 instances respectively.

The detailed breakdown in Table 3 shows both bounding box (object detection) and mask (instance segmentation) performance metrics across all safety net classes, fulfilling our second objective of evaluating system performance on real-world safety net images. This comprehensive evaluation provides insights into which net types are more reliably detected and segmented, information that directly supports our third objective of creating a practical tool for safety inspectors.

In Table 3, P stands for Precision; R stands for Recall; mAP50 stands for Mean Average Precision at 0.5 IoU; mAP50-95 stands for mean average precision performance at IoU thresholds of 0.5 to 0.95 in 0.05 increment.

Table 3. YOLOv8 Object Detection and Instance Segmentation Performance on Safety Net Test Dataset

Class	No. of images	No. of instances	Box				Mask			
			P	R	mAP 50	mAP 50-95	P	R	mAP 50	mAP 50-95
All	21	691	0.64	0.59	0.58	0.36	0.69	0.63	0.62	0.38
class_1 (Fence)	20	653	0.68	0.77	0.79	0.61	0.69	0.78	0.79	0.59
class_2 (Blue)	8	21	0.80	0.58	0.64	0.33	0.80	0.59	0.64	0.38
class_3 (Barricade)	4	17	0.44	0.41	0.29	0.15	0.57	0.53	0.43	0.18

Table 4 summarises the key instance segmentation metrics relevant to practical deployment, highlighting the overall mean Average Precision of 0.623 at IoU 0.5. These results indicate that our model meets the performance threshold needed for a useful safety inspection tool, directly supporting our third objective of creating a practical tool that can assist safety inspectors. The F1 scores demonstrate a reasonable balance between precision and recall, which is critical for a safety application where both false positives and false negatives have important implications.

Table 4. YOLOv8 Instance Segmentation Performance on Safety Net Test Dataset

Class	AP@0.5	AP@0.5:0.95	Precision	Recall	F1 Score
class_1	0.797	0.585	0.697	0.783	0.737
class_2	0.642	0.375	0.804	0.585	0.677
class_3	0.431	0.177	0.567	0.529	0.547
Mean	0.623	0.379	0.689	0.632	0.659

From Table 3 and Table 4, we note that the model demonstrates strong overall performance, with a mean Average Precision (mAP@0.5) of 0.623 for segmentation. Performance varies significantly across classes, with the highest performance on class_1 (mAP@0.5 = 0.797) and the lowest on class_3 (mAP@0.5 = 0.431), which correlates with the class distribution in the dataset. The precision-recall balance also varies by class, with class_2 showing high precision (0.804) but lower recall (0.585), while class_1 has more balanced precision (0.697) and recall (0.783).

Figure 13. The precision-recall curves illustrate the trade-off between precision and recall at different confidence thresholds for each class

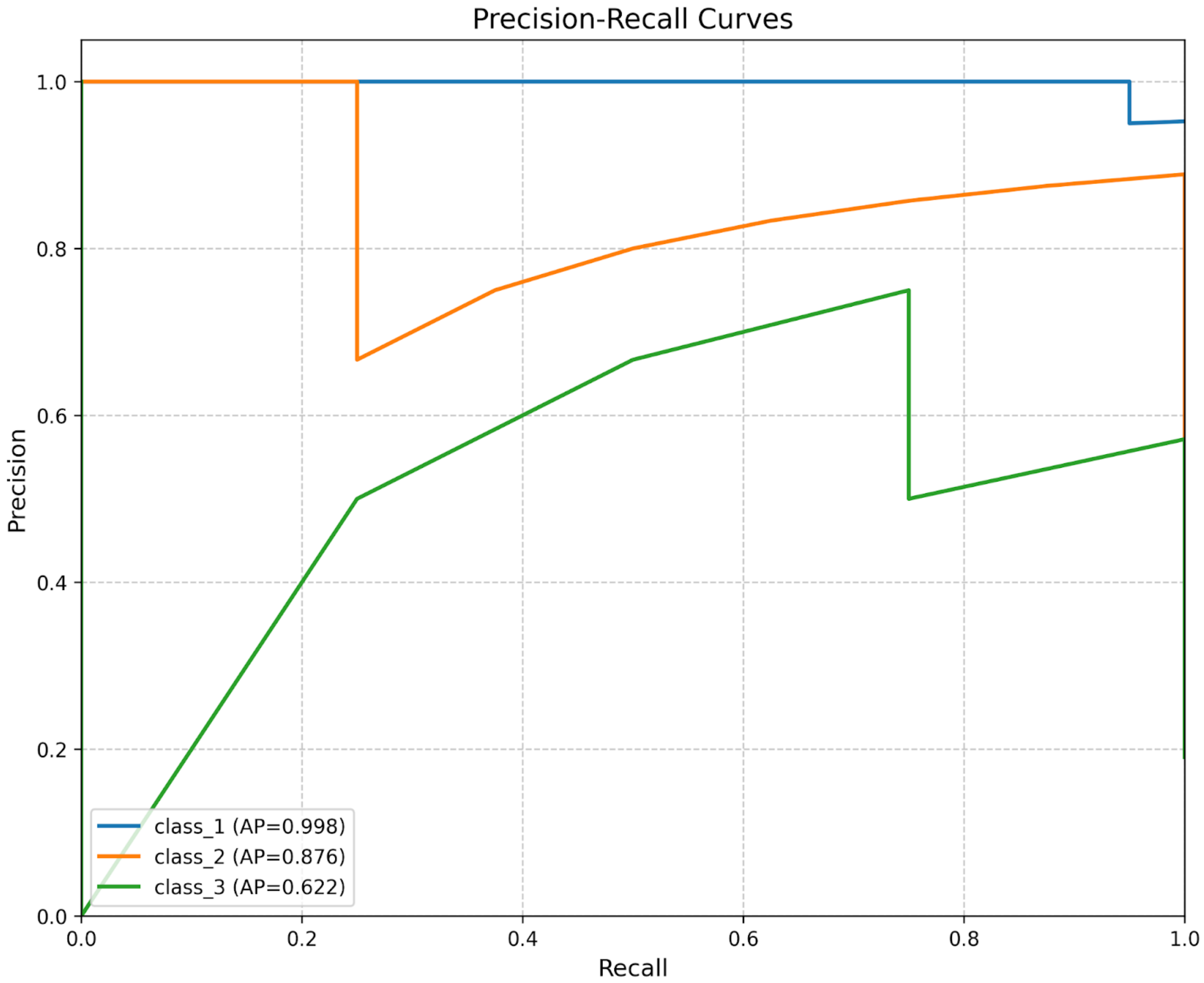


Figure 13 shows that class_1 (red line, Fence Safety Nets) maintains high precision even at high recall levels, class_2 (green line, Blue Safety Nets) shows high precision but drops more quickly as recall increases, and class_3 (blue line, Barricade Safety Nets) demonstrates lower overall performance with a steeper decline in precision as recall increases.

The precision-recall curves provide insight into model behaviour across different confidence thresholds. Class_1 (Fence Safety Nets) maintains relatively high precision even as recall increases, suggesting robust performance across various scenarios. Class_2 (Blue Safety Nets) starts with very high precision but shows a sharper drop as recall increases, indicating that while the model is very confident in its correct predictions, it misses more instances.

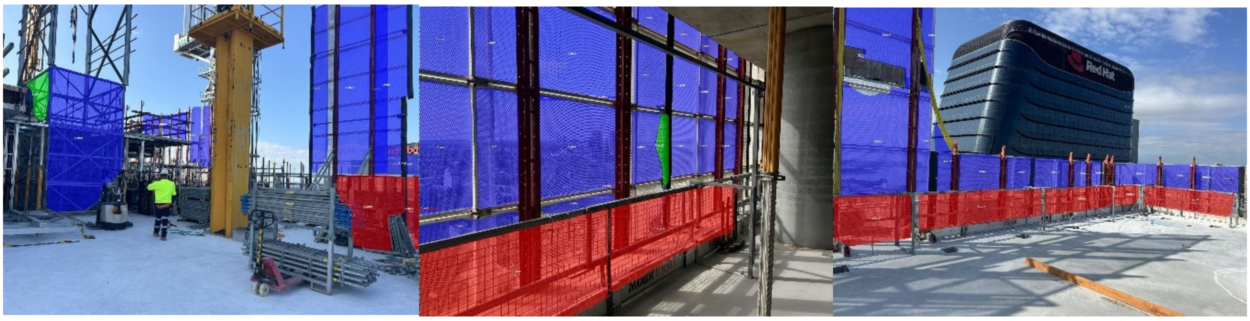
Class_3 (Barricade Safety Nets) shows the weakest curve, reflecting the challenges in detecting this class with limited training examples.

Visual analysis

Visual inspection of the model's predictions provides valuable insights into its real-world performance and limitations.

The model accurately identifies and segments Fence safety nets (red overlay), Blue safety nets (green overlay), and Barricade safety nets (blue overlay) with precise boundary detection even in complex environments with varying lighting conditions, as illustrated in Figure 14.

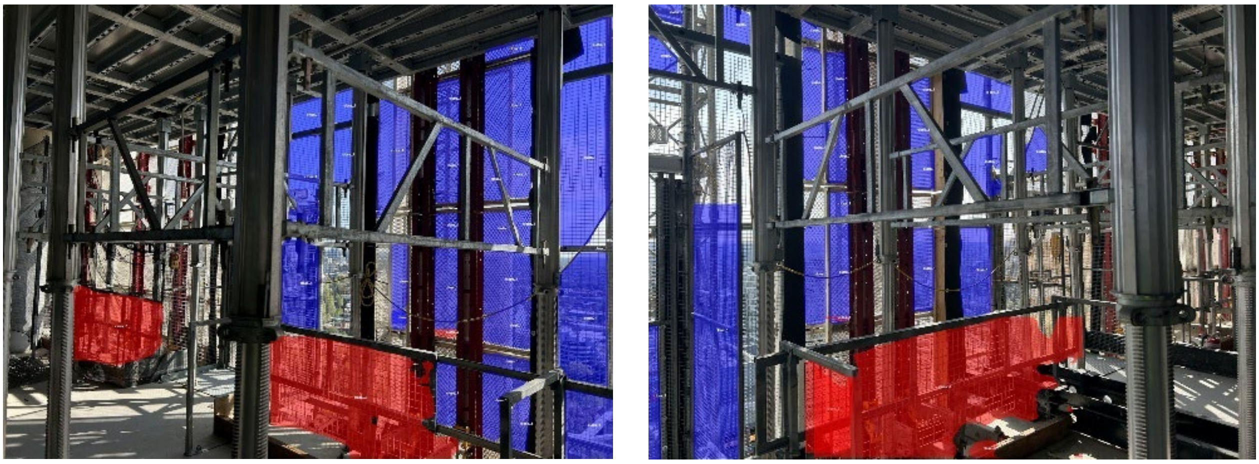
Figure 14. These images demonstrate successful segmentation of different types of safety nets in various construction site scenarios



In the left image of Figure 15, the model struggles with partially occluded safety nets and complex overlapping patterns. In the right image, unusual lighting conditions and atypical installation patterns present challenges for accurate segmentation.

From Figure 15, we note that the model faces challenges in scenarios with occlusion, unusual lighting conditions, or atypical installation patterns. In these cases, the segmentation masks may be incomplete or fragmented, and the model might misclassify certain regions. These limitations are expected given the complexity of construction sites and the relatively limited size of the training dataset.

Figure 15. Examples that illustrate scenarios where the model faced difficulties



Performance results

For practical deployment in safety inspection scenarios, computational efficiency is as important as accuracy. Our system demonstrated promising performance in this regard.

Table 5. Computational performance under different hardware

Hardware	Preprocess Time	Inference Time	Post-process Time	Total Time	Frames Per Second (FPS)
Quadro RTX 6000 (GPU)	2.55 ms	11.67 ms	35.11 ms	49.33 ms	24.66
CPU	2.48 ms	10.53 ms	34.03 ms	47.04 ms	25.63

Interestingly, CPU inference performed slightly faster than GPU inference in our tests. This suggests that the post-processing operations, which account for approximately 70% of the total inference time, isn't fully optimised for GPU acceleration with the current implementation. Both hardware configurations achieve around 25 frames per second, sufficient for near real-time safety inspection applications. The training was conducted on Ubuntu 24.04 LTS with Quadro RTX 6000 with 24 GB Graphics Card, employing the Ultralytics YOLO implementation. The complete training process required approximately 1 hour and 7 minutes for 100 epochs with a batch size 16.

Experiment results show that YOLO model effectively segments different types of safety nets with 62.3% mean Average Precision (mAP@0.5), demonstrating its potential for automating safety inspection tasks. This fulfils our objective of developing a robust segmentation model capable of identifying and classifying different types of safety nets.

Our system performs well across diverse environmental conditions, including varying lighting, camera angles, and installation patterns, showing its robustness for real-world applications.

The segmentation results provide detailed visual documentation of safety net locations and types, which can significantly enhance the inspection process and improve documentation consistency. This supports creating a practical tool for safety inspectors. Finally, our system achieves approximately 20 frames per second on both GPU and CPU hardware, making real-time inference feasible for field applications on standard equipment.

Despite the promising results, the model may struggle with highly occluded safety nets or unusual installation patterns not well-represented in the training data. This limitation could potentially be addressed with a more diverse training dataset. Performance degrades in extreme lighting conditions, such as strong shadows or glare, which are common challenges in outdoor construction sites. The model's performance varies significantly across classes, with the least-represented class (Barricade safety nets) showing considerably lower accuracy than the dominant class (Fence safety nets), reflecting the class imbalance in the training data.

Solution 2: Gap Detection and Classification Model

2.1 Overview

This detection model was designed to identify six types of gaps in safety perimeters on high-rise construction sites. As a backend or cloud-based system, it enables near-real-time detection during inspections. The system highlights potential risks (gaps) and classifies them into six categories.

Developing computer vision–based gap detection models typically follows one of two approaches: (1) deploying powerful pretrained large models without additional training, or (2) training a dedicated model for the specific task. In this project, we found that existing pretrained models performed poorly on gap detection, as they were not trained on relevant construction-site data and failed to generalise to such scenarios. Therefore, training a dedicated gap detection model was deemed a more effective and suitable solution.

Training a task-specific model typically requires a large volume of human-annotated images from similar environments. Expanding the dataset to include diverse scenarios with accurate labels helps the model learn more effectively and improves its performance on new, unseen images. In practice, building such a dataset often involves collecting and labelling several thousand images, each containing one or more gap instances, a process that demands considerable time and effort.

2.2 Dataset and annotations

In this project, only a limited number of annotated images (230) were available. The annotations, provided by domain experts, were broad and highlighted general gap issues without detailed instance-level labelling. When we directly used these images and annotations to train a model (e.g., YOLOv11), the resulting performance was poor, with a recall of only around 2%, indicating that the model fails to detect the majority of gaps accurately.

Therefore, the project explored class-aware approaches, incorporating a carefully designed classification scheme for different gap types. This allowed the model to learn more specific features associated with each gap category, leading to improved detection performance. After a series of experiments, six gap classes were defined, and the trained model achieved a precision of 81.5% and a recall of 95.1%.

2.3 Methods

In this project, the broad definition of a gap posed a challenge for the gap detection model. The wide variety of gap types on construction sites, combined with a limited number of training instances, made it difficult for the model to learn and recognise consistent visual features. For example, some gaps appear as narrow horizontal openings near the top of the image, while others may be square-shaped holes on the ground. With insufficient data, the model tends to detect only gaps that closely resemble those seen during training. This lack of generalisation likely contributed to the low recall observed when training the model without any annotation preprocessing.

Figure 16 shows four images, each annotated with red bounding boxes provided by safety investigation experts to highlight different types of gaps. While humans can easily recognise the presence of gaps using these visual cues, it remains challenging for the computer vision model to learn effectively from such a limited number of examples. The visual characteristics of the gaps vary significantly, making it difficult for the model to generalise and correctly identify all of them. Therefore, under this limited-data scenario, it is necessary to classify gaps into distinct categories to improve model learning and performance.

Figure 16. Examples of annotated images provided by industry partner

Note: The red annotated boxes indicate the hazard areas, including loose stripe (top-left), gaps between perimeters (top-right), and holes on the ground (bottom).

Based on our observations, study and experiment, we classified these gaps into six categories as shown in Figure 17.

Figure 17. Six types of gap classifications derived from the original three application scenarios

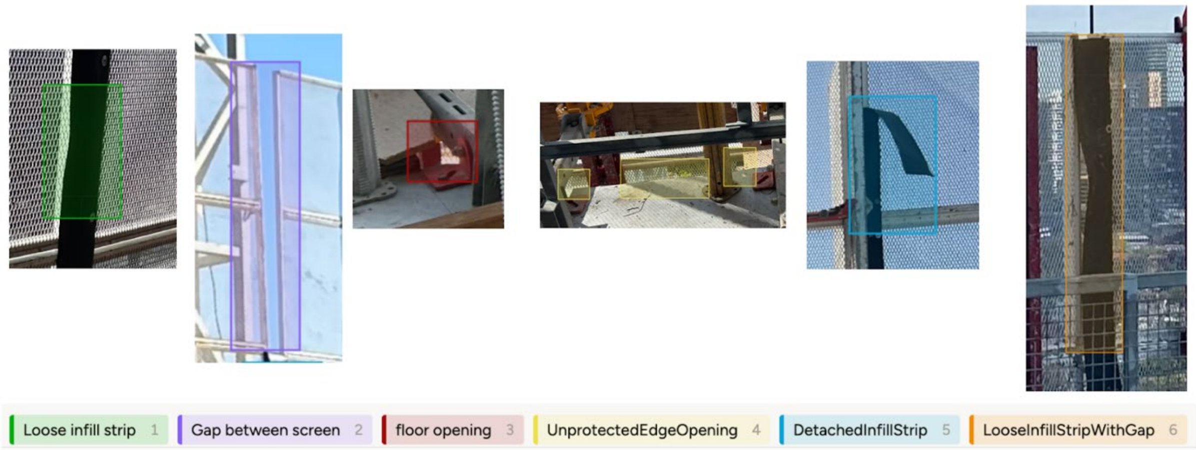


Table 6 summarises the six categories of gaps and characteristics of each gap based on its size, visibility and risk levels. There were two clearly visible types of gaps identified: the visible gap between safety screens without an infill strip (Type 2 in Figure 17), and the unprotected edge opening (Type 4). Both pose significant safety risks, as they may allow people or objects to pass through and fall from height.

In addition, there were more subtle and less noticeable types of gaps, such as floor openings (Type 3) and loosened infill strips with small gaps (Type 6), which can also lead to safety issues, especially if the gap size increases over time without proper intervention.

Detached infill strips (Type 5) often indicate the presence of a nearby gap, while loose infill strips (Type 1) suggest the potential for a gap that may not be visible from the current angle but could become apparent from another viewpoint.

Table 6. Classification and characteristics of six categories of gaps

Category	Visible gap	Size of gap	Risk level	Gap progression risk	Angle sensitivity
Class 1. Loose Infill Strip	❌	Small	Low	May loosen further over time	High
Class 2. Gap Between Screen	✅	Large	High	Stable but dangerous	Low
Class 3. Floor Opening	✅	Small	Medium	May be stable or widen	Medium
Class 4. Unprotected Edge Opening	✅	Large	High	Exposed and critical	Low
Class 5. Detached Infill Strip	❌	-	Low	Indicates nearby potential gaps	Medium
Class 6. Losse Infill Strip with Gap	✅	Small	Medium	May increase if not addressed	High

Computer vision model architecture

Due to dataset limitations, YOLOv11 was selected as the detection architecture for its enhanced performance in identifying complex, small-scale structural gaps and its efficiency in real-time applications. YOLOv11 introduces several architectural innovations that improve feature extraction and spatial awareness. It supports multi-task learning and demonstrates high adaptability across object detection, instance segmentation, pose estimation, and oriented object detection. It achieves improved mean Average Precision (mAP) with a reduced parameter count compared to earlier versions, making it suitable for deployment in both edge and high-performance computing environments.

Model configuration

We used the YOLOv11s variant for gap detection, selected for its balance of speed and accuracy on high-resolution construction site images. The configuration details are as follows:

- Input resolution: 2560×2560 pixels
- Model variant: YOLOv11s (small)
- Number of classes: 6 (corresponding to six types of construction-site gaps)
- Loss weights:
    - Bounding box regression: 6.0
    - Classification: 1.0
- Augmentation techniques:
    - Mosaic: 100% probability
    - Copy-paste: 30% probability
- IoU threshold for matching: 0.4.

Training process

The model was trained using a single stage fine-tuning approach:

- **Pretrained weights:** The YOLOv11s model was initialised with pretrained weights from the COCO dataset.
- **Dataset:** We used a custom gap detection dataset containing 186 training images and 44 test images, with a randomly curated split to ensure representation across different gap types and environments.
- **Training parameters:**
    - Epochs: 400
    - Batch size: 16
    - GPUs: 8 GPUs (device=0–7)

- Training Hardware:
  - 8 × NVIDIA RTX A6000 GPUs
- **Early stopping:** Training used a patience of 50 epochs to prevent overfitting.

This configuration enabled efficient training on limited data while leveraging transfer learning to adapt the pretrained model to our specific safety inspection scenario.

2.4 Experiments

To assess the detection performance of the trained model, we adopted the following standard object detection metrics:

- **Bounding Box Precision:** The proportion of predicted bounding boxes that correctly correspond to ground truth gap instances.
- **Bounding Box Recall:** The proportion of ground truth gaps that were successfully identified by the model.
- **mAP@50:** Mean Average Precision at an IoU threshold of 0.5, which reflects model accuracy under a more forgiving overlap requirement.
- **mAP@50–95:** Mean Average Precision averaged over IoU thresholds from 0.5 to 0.95 in 0.05 increments, providing a more stringent and comprehensive evaluation of detection quality.

These metrics collectively offer a detailed view of the model's precision, completeness, and robustness across a range of overlap conditions between predicted and actual gap regions.

Model evaluation was conducted on a validation set comprising 44 images that were excluded from the training process. This subset was manually curated to ensure a representative distribution of gap types, camera angles, and scene complexity, closely reflecting the conditions expected in real-world deployment. The testing protocol focused on assessing the model's generalisation ability to unseen data and its robustness across diverse environmental scenarios.

2.5 Results and discussion

Quantitative analysis

After fine-tuning the YOLOv11s model on our custom dataset, the Class-Aware Construction Gap Detection System demonstrated strong overall performance. As shown in Table 7, the model achieved a bounding box **precision of 0.815 and a recall of 0.951**, indicating high reliability in identifying gaps.

The mean Average Precision (mAP) values reached 0.936 at IoU 0.5 and 0.649 across IoU 0.5 to 0.95, reflecting robust localisation performance across various overlap thresholds.

The per-class breakdown provides further insights into the model's strengths and limitations:

- Detached Infill Strip and Floor Opening classes achieved perfect recall (1.0) and high precision (0.887 and 0.966 respectively), suggesting these gap types are well-represented and visually distinctive.

- Gap Between Screen, despite its perfect recall, showed the lowest precision (0.690) and the lowest mAP@50-95 (0.497), likely due to the very small number of examples (only 2 instances), which limits the model's ability to generalise.
- Loose Infill Strip showed lower recall (0.706) and moderate precision (0.795), possibly due to visual ambiguity or partial occlusion in real-world scenes.
- Loose Infill Strip With Gap and Unprotected Edge Opening achieved high recall (1.0) but slightly lower precision (0.768 and 0.782), reflecting the model's ability to detect these classes consistently, though sometimes with false positives.
- Overall, Floor Opening achieved the best combination of high mAP@50 (0.995) and strong mAP@50-95 (0.790), indicating consistent and precise detection performance.

Table 7. Class-Aware Construction Gap Detection System Performance

Class	Images	Instances	Box (Precision)	Box (Recall)	mAP50	mAP50-95
All classes	44	51	0.815	0.951	0.936	0.649
Detached Infill Strip	4	4	0.887	1	0.995	0.708
Gap between screen	1	2	0.690	1	0.995	0.497
Loose Infill Strip	12	17	0.795	0.706	0.756	0.529
Loose Infill Strip With Gap	7	8	0.768	1	0.928	0.609
Unprotected Edge Opening	3	4	0.782	1	0.945	0.762
Floor Opening	8	16	0.966	1	0.995	0.79

The evaluation results confirm that the class-aware approach improves detection performance across multiple gap types. However, gap types with very limited training instances (e.g., Gap Between Screen) remain challenging, leading to lower generalisation accuracy. This highlights the need for additional annotated data to further improve performance consistency across all classes.

Qualitative analysis

Figures 18-22 illustrate representative detection results for various gap types using the fine-tuned YOLOv11s model. The detection of a Detached Infill Strip with **a confidence score of 0.84** demonstrates the model's capability to accurately localise subtle structural anomalies. In scenes containing multiple Loose Infill Strips, the model consistently detects all relevant instances, reflecting strong intra-class generalisation.

Additionally, detections of Unprotected Edge Openings and Floor Openings show high spatial accuracy and confidence, even under varied lighting and viewpoint conditions.

Figure 18. Example detection of a Detached Infill Strip with a confidence score of 0.84

Figure 19. Example detections of different types of target objects

Figure 20. Example detections of two Loose Infill Strips

Figure 21. Example detections of Unprotected Edge Opening

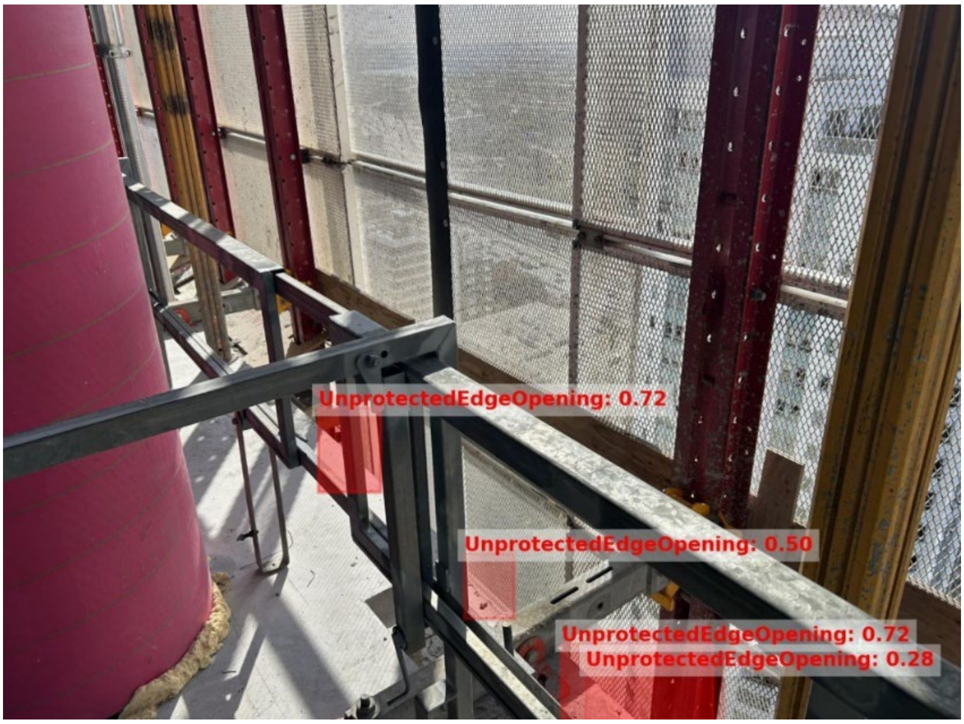
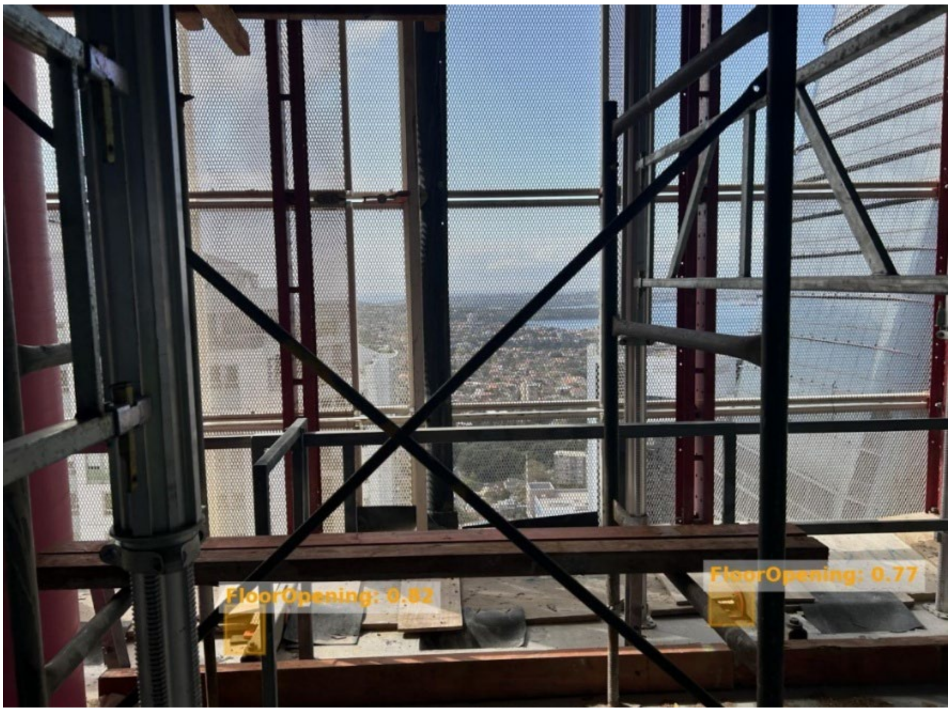


Figure 22. Example detections of Floor Opening



The figures showcasing different types of target objects further highlight the model's ability to distinguish between visually similar yet semantically distinct gap categories, underscoring the effectiveness of the class-aware design. These qualitative results align well with the quantitative performance metrics and confirm the model's robustness in real-world scenarios.

Inference efficiency

One of the key advantages of the Class-Aware Construction Gap Detection System applied in Solution 2 lies in its exceptional inference speed. On mainstream high-performance GPUs (e.g., NVIDIA RTX A6000), the model achieves single-image inference times of approximately **20 milliseconds**, enabling real-time detection at over 50 FPS. Even on consumer-grade hardware such as the MacBook Pro, the system maintains impressive efficiency, with inference latency consistently under 50 milliseconds per image.

This level of responsiveness makes the model not only suitable for offline analysis but also highly deployable in edge or on-site AR-assisted inspection scenarios, where immediate visual feedback is critical. The combination of lightweight architecture and optimised computation paths reflects the robustness of our design, striking a rare balance between speed, accuracy, and deployment flexibility.

Module 2 – AR-Assisted Inspection Tool

1. Introduction

Augmented Reality (AR) is an emerging interactive technology with the potential to significantly advance the construction industry. By bringing digital information "out of the display" and integrating it into the physical world, AR enables users to view data within the context of real-world objects and environments. Recent advances, such as low latency, wide fields of view, and pass-through headsets that allow users to see their external environment, may further increase AR adoption across a range of domains.

Additionally, AR systems are equipped with sensors, cameras, and internet connectivity, enabling them to perform geo-location, eye tracking, and data capture and recording directly through the headset. These capabilities, along with unique immersive interactions and user experiences, make AR systems a viable technology for automating and facilitating safety inspection activities on construction sites. According to industry experts, most high-rise building projects still rely heavily on manual inspection processes to detect and mitigate risks associated with falling objects. Furthermore, routine safety observations and inspections often require time-consuming report completion. With head-mounted AR systems, inspectors can easily track inspected areas and virtually annotate recorded scenes (Park & Kim, 2013). AR systems can also be integrated with Artificial Intelligence (AI) and visual models to automate the safety risk detection on construction sites (Chen et al., 2024).

Past studies have explored the benefits of applying AR in the construction industry, such as visualising subsurface utilities (Oke & Arowoiya, 2021) (Oke & Arowoiya, 2022), conducting in-situ virtual examinations of physical sites (Arowoiya et al., 2020), and providing safety training, including for heavy equipment operation (Wu et al., 2019). However, the application of AI-assisted AR systems to address safety risks related to falling objects remains underexplored. This prototype aimed to address this research gap, which relates to one of the highest-rated risks in the construction industry: falling objects in high-rise buildings.

2. System design

Our proposed system was developed based on requirements provided by our industry partners and offers the following features.

- **Capture the dynamic construction site environment**: We use the built-in spatial mapping features from the headset to scan the surrounding environment.
- **Indicate the progress and completeness of the inspection**: we design a “brushing” feature to highlight areas that have been inspected. The “brushing” interaction metaphor uses the inspector’s field of view to track the inspection progress.
- **Annotations**: Users can virtually annotate or add sticky notes to a specific location in 3D space. Voice-transcribed notes can also be generated and exported together with videos or photos.
- **Automatic report generation system**: The system automatically summarises and produces inspection reports based on user-entered notes and snapshots.
- **Integration of Large Vision models**: The system integrates the visual models (Module 1’s both Solution 1 and Solution 2) into our AR system to automatically detect safety issues (e.g., gaps on the perimeter safety screens)
- **Built-in inspection trails, scene recording and photo-capture**: Geolocation-based inspection trails, the inspector’s head position over time, and imagery recorded from their point of view are used to report.

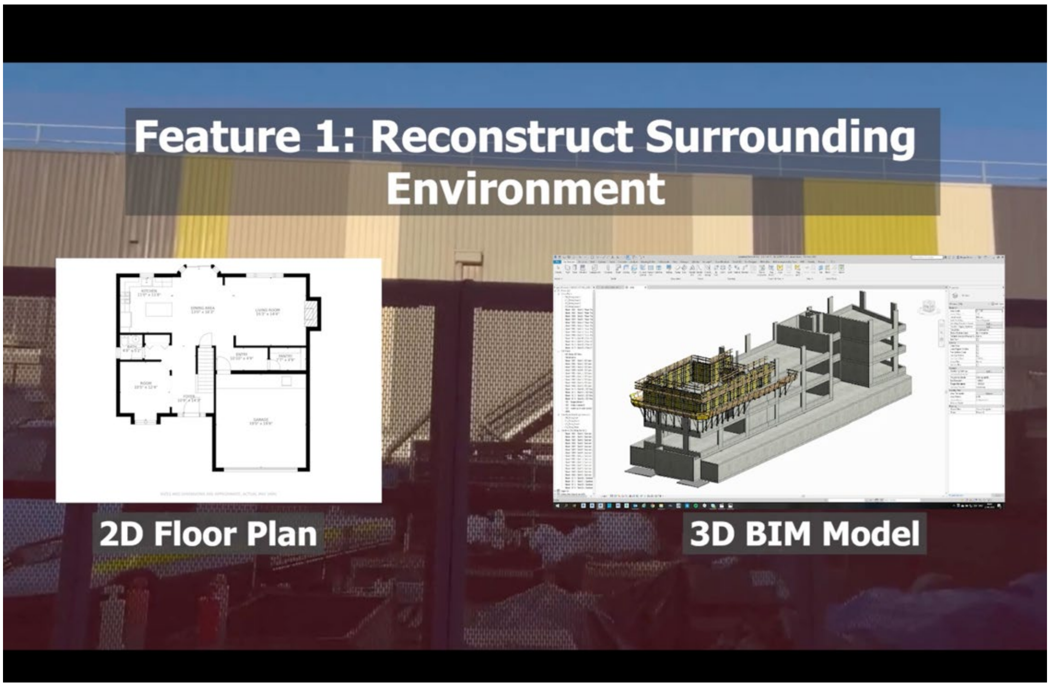
3. Implementation

The AI-assisted AR inspection prototype was developed using Xcode version 16.4 on an Apple MacBook Pro equipped with an M4 Pro chip, leveraging the latest development environment and processing capabilities optimised for spatial computing. The resulting application was subsequently deployed to the Apple Vision Pro headset, Apple’s premier augmented reality device, which supports immersive user interaction and precise spatial tracking. The Vision Pro’s integrated LiDAR sensor, high-resolution passthrough display, and spatial audio capabilities offer a robust platform for real-time construction safety inspection workflows. The deployment setup enables seamless integration of visual annotation, voice input, and environment mapping, all of which are essential for accurate and efficient on-site data collection within high-risk construction environments. The final code for the prototype is available on Microsoft Teams and shared with the industry partner.

Feature 1: Reconstruct the Surrounding Environment

At the start of the inspection task, the inspector needs to put on the headset. The headset then reconstructs the surrounding environment. Specifically, the system loads the 2D floor plan or 3D BIM model provided before the inspection and generates virtual surfaces overlaid with transparent colours on top of the physical surfaces, such as walls and floors. The coloured overlay on the physical surfaces is less noticeable, so it does not affect the regular inspection of the physical surfaces, such as inspecting perimeters for potential gaps.

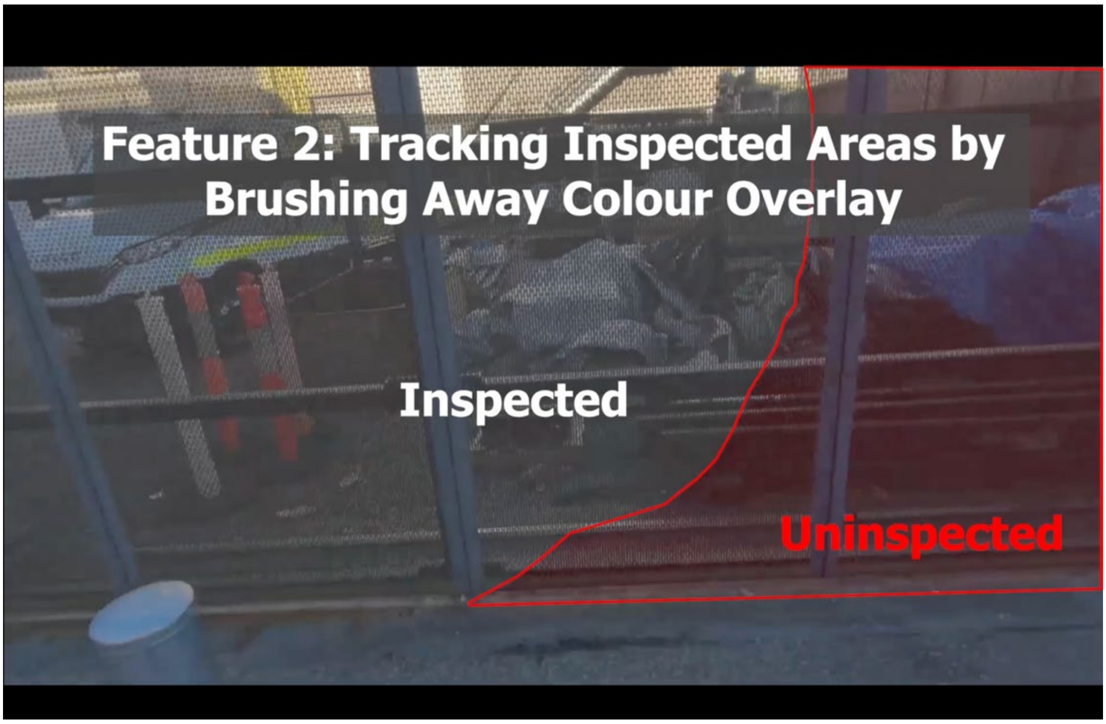
Figure 23. Feature 1 of the AR prototype: Reconstruct the Surrounding Environment



Feature 2: Tracking Inspected Areas by Brushing Away Colour Overlay

Once the reconstruction is complete, the inspector can begin the regular inspection. Our system allows tracking head movement and eye gaze information to brush away the generated colour overlay, which indicates the inspected areas with complete transparency. On the other hand, the remaining colour-overlaid surfaces indicate the uninspected areas. The inspection paths or trajectories could be saved and further analysed for optimisation and training purposes.

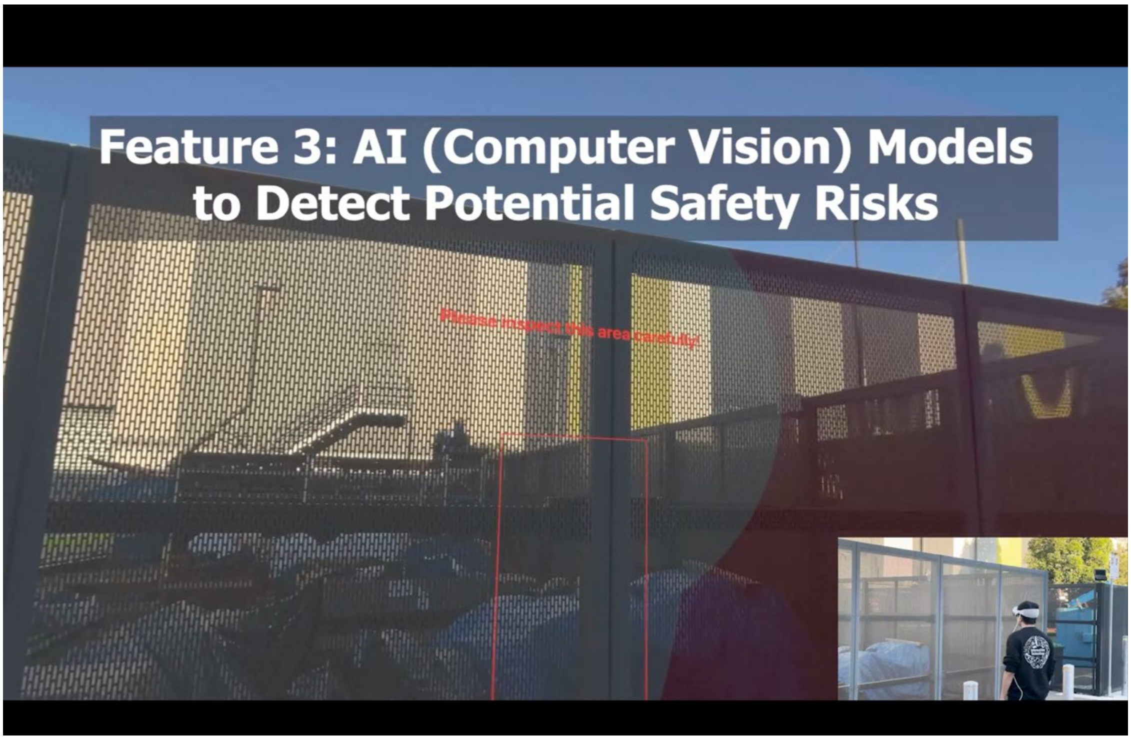
Figure 24. Feature 2 of the AR prototype: Tracking Inspected Areas by Brushing Away Colour Overlay



Feature 3: AI (Computer Vision) Models to Detect Potential Safety Risks

Feature 3 is a key component of the AR prototype as it allows the AR system to interact with server-side (external) components, which are the computer vision models we developed as Solution 1 and Solution 2 in Module 1. This feature is used to send the images to the visual models and receive the detection results, and if any gaps and risks are identified, they are highlighted as red boundaries in the user's view via the headset.

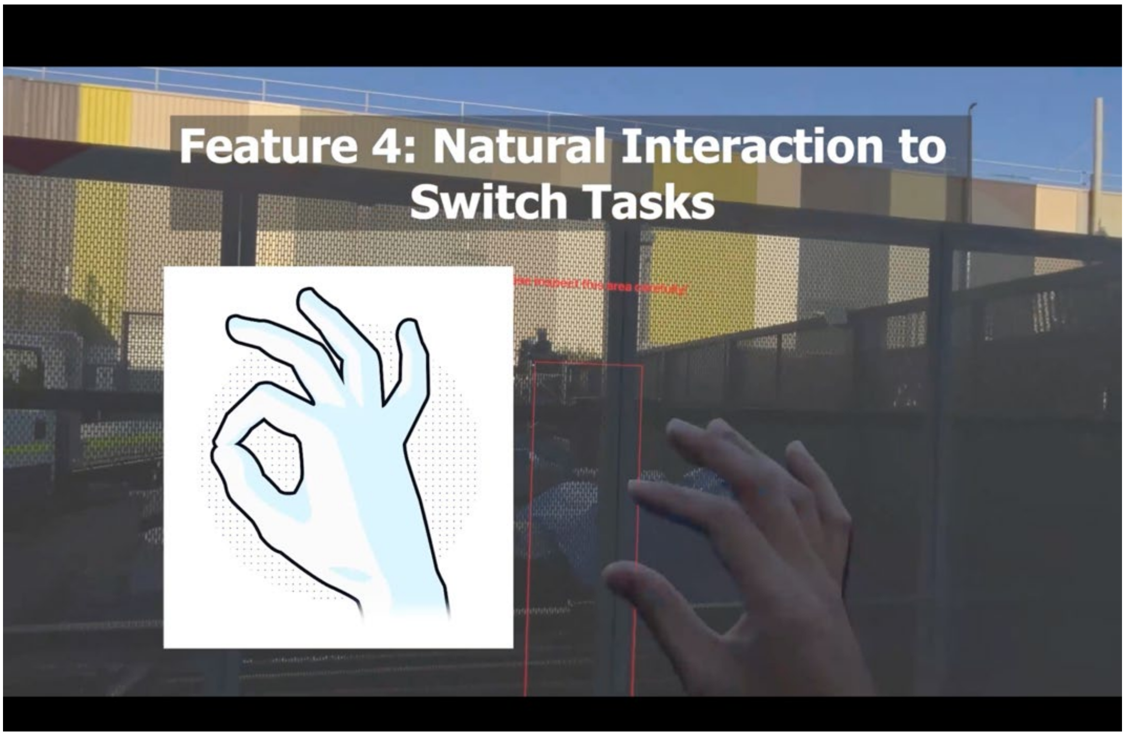
Figure 25. Feature 3 of the AR prototype: AI (Computer Vision) Models to Detect Potential Safety Risks



Feature 4: Natural Interaction to Switch Tasks

The fourth feature of the AR system allows the user to use natural hand gestures (swiping in the air) to switch inspection tasks during or after inspection, such as performing inspections or taking photos.

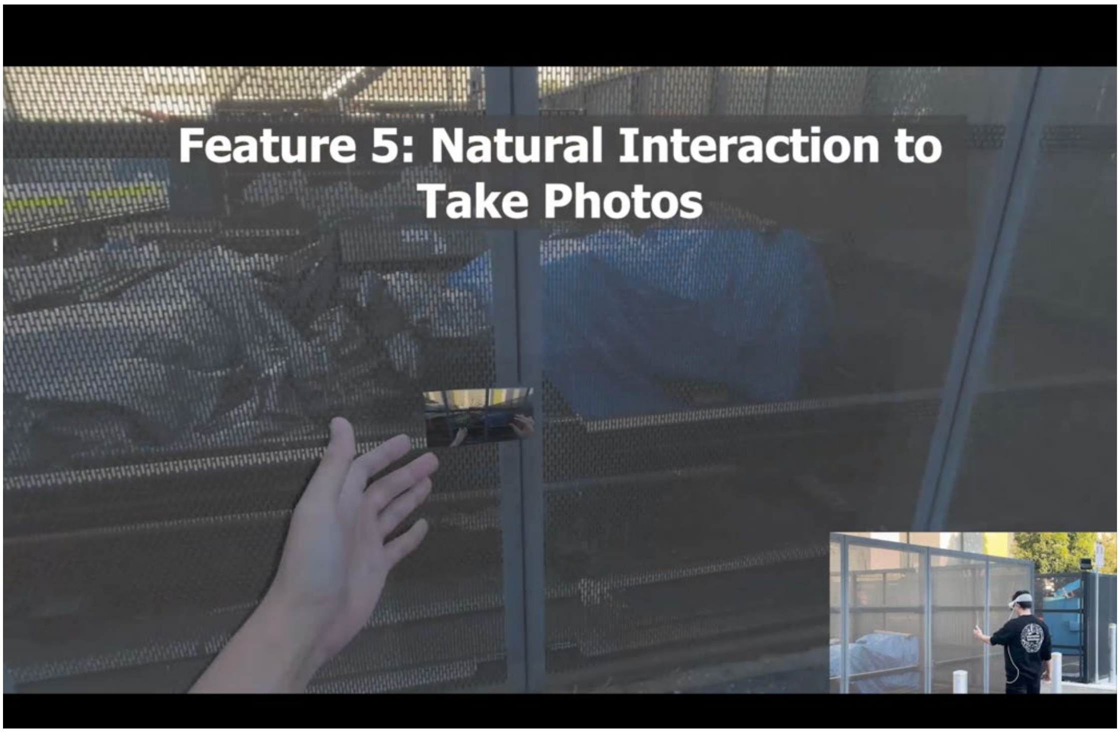
Figure 26. Feature 4 of the AR prototype: Natural Interaction to Switch Tasks



Feature 5: Natural Interaction to Take Photos

Feature 5 allows using natural interactions to easily take photos of interested areas. The inspector only needs to double-tap their fingers to photograph their field of view, significantly reducing the tedious process compared with other mobile devices. To provide an overview, thumbnails of the photos will be added to a virtual panel attached to the inspector's hand.

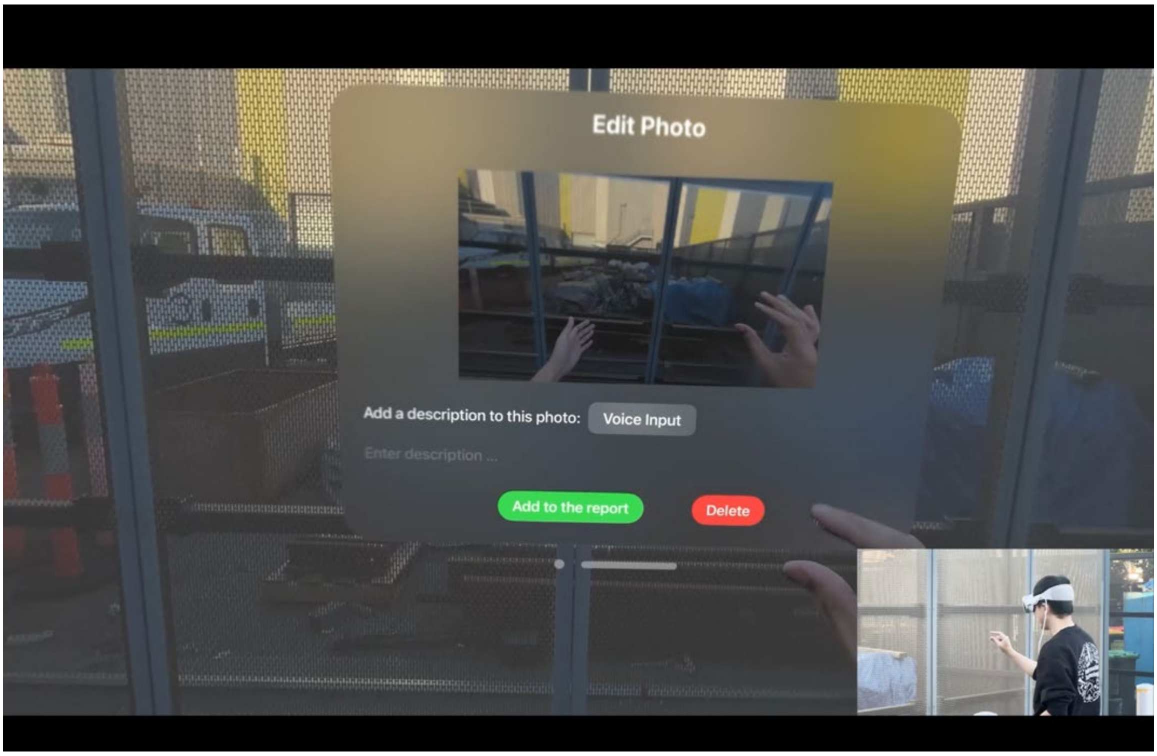
Figure 27. Feature 5 of the AR prototype: Natural Interaction to Take Photos



Feature 6: Edit Photos and Add Descriptions

Each photo could be selected on the virtual panel to bring up a Photo Edit window. In this window, inspectors can view full-resolution images and add a photo description or comments using voice input or a virtual keyboard. This voice input feature can significantly reduce text input time compared to traditional typing input. The inspector can also delete the photo or choose to include the photo in the final report by tapping a button.

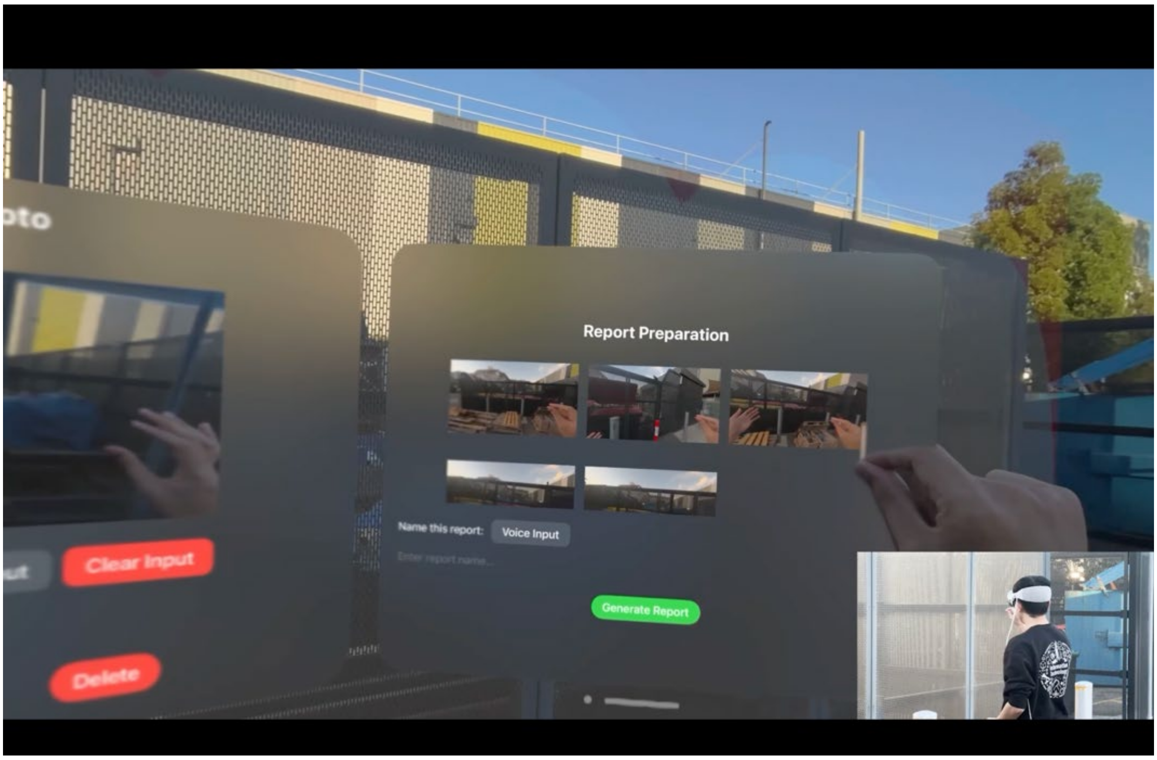
Figure 28. Feature 6 of the AR prototype: Edit Photos and Add Descriptions



Feature 7: Prepare Photos for the Report

Once inspectors select a photo to include in the final report, another window is triggered. In this Report Preparation window, the inspector can view and rearrange all selected photos if necessary. Finally, they can add a name for the report via voice input or virtual keyboard.

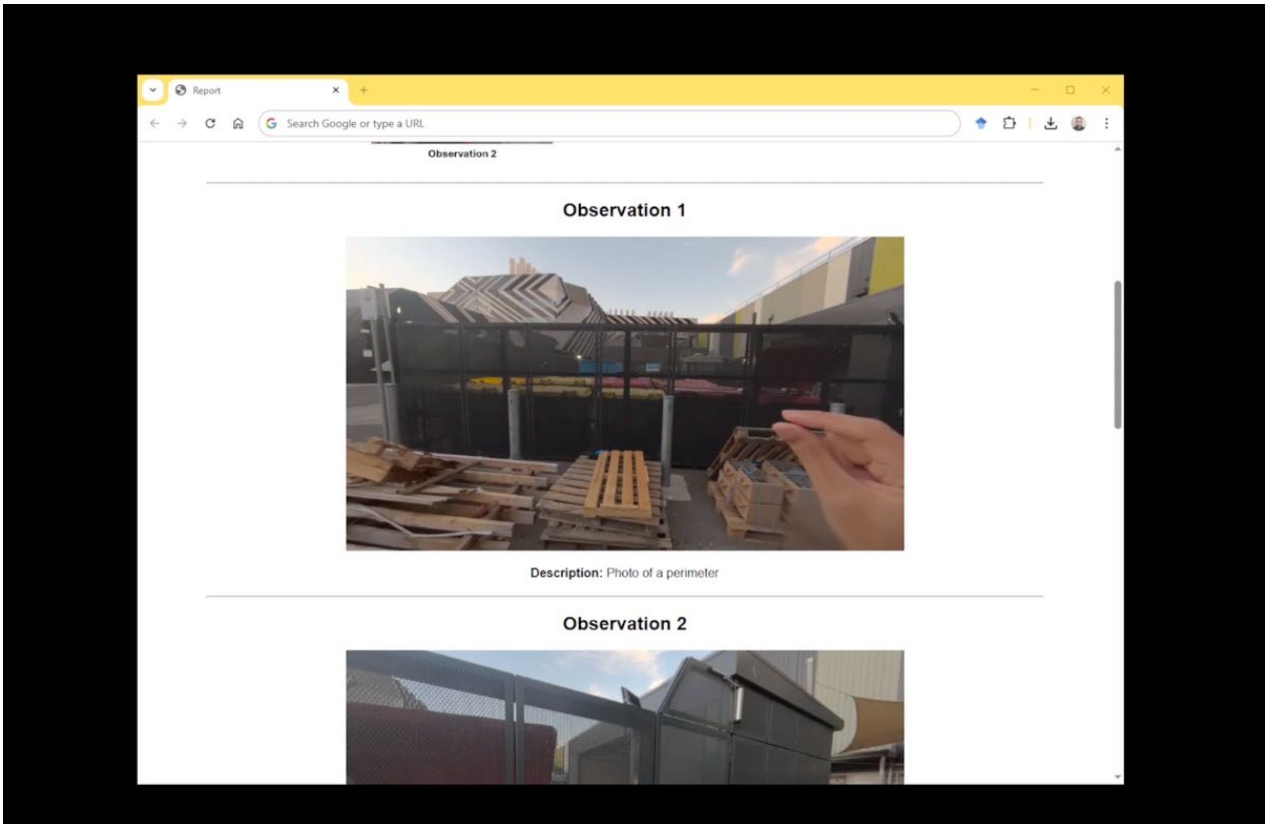
Figure 29. Feature 7 of the AR prototype: Prepare Photos for the Report



Feature 8: Generate a Final Report by Clicking a Button

After finalising the inspection process, the inspector can tap a button to send all photos with descriptions to a secure online server, which automatically processes the files and generates a final report. Because the files are uploaded to the server, the generation process can be repeated multiple times if needed. We expect the report to be viewed on a laptop or desktop computer after the inspection. The report will include the title, an overview of the geolocation of the photos on a floor plan indicating the hazard area, and full-resolution photos along with their descriptions.

Figure 30. Feature 8 of the AR prototype: Generate a Final Report by Clicking a Button



4. Limitations and future work

The AI-assisted AR system developed aims to focus on addressing the following tasks and contexts.

- Integration with computer vision models for proactively detecting gaps in safety parameters,
- Natural interactions to take photos/ videos from headsets,
- Geolocation-based tracking via the headset's built-in sensors,
- Automating safety inspection and generating a report, and
- Inspectors spend less time conducting site safety inspections.

While the prototype offers several innovative components to address key challenges in construction site safety, privacy concerns remain a critical barrier to the adoption of such systems. The headset continuously tracks the user's physical location, which can raise privacy issues. Additionally, snapshots and videos may capture workers or bystanders who have not provided consent to be recorded. To address these concerns, it is essential to implement privacy-preserving mechanisms that minimise potential risks. A number of these approaches are discussed below.

- **Segmenting and Blurring Bystanders' Faces**: One possible solution to protect the privacy of bystanders is to use AI segmentation techniques to blur their faces while taking videos or snapshots. In a system proposed by Cheng et al. (Cheng et al., 2022) for monitoring site safety compliance, the system can automatically blur workers' faces upon saving the video frames. With the rapid development of lightweight on-device deep learning techniques, future AR solutions could leverage these advancements to address bystander privacy concerns.

- **3D Human Skeleton Feature Extraction**: An alternative approach is to extract 3D human skeleton features from video recordings automatically. The extracted skeleton features can then be used to train an ergonomic posture recogniser, which can identify poor postures and perform posture-based stability analysis.
- **Context-aware Privacy-preserving Mechanisms**: Privacy can be protected by setting preferences based on factors such as location, scene, the presence of others, and hand gestures. These preferences can be dynamically adjusted in emergency situations or in areas requiring security clearance.
- **Federated Learning**: In federated learning, machine learning models are trained across multiple decentralised edge devices or servers while keeping data stored locally on each device. This approach is particularly useful for analysing data across multiple construction sites without centralising sensitive information.
- **Differential Privacy**: This technique allows the sharing of information derived and inferred from a dataset while preserving the confidentiality of user sensitive data. It can be explored when capturing data and generating reports related to worker productivity or safety compliance.

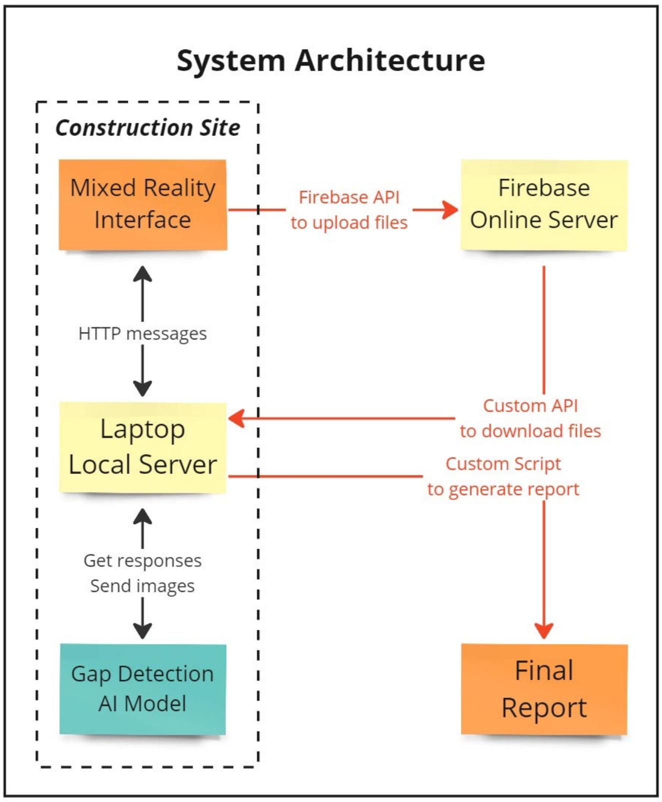
The data itself can be further protected by using more advanced techniques:

- **End-to-End Encryption**: Employing robust encryption protocols for data in transit and at rest to ensure that the data remains unreadable to unauthorised parties.
- **Blockchain technology**: Integrating blockchain technology will help create tamper-proof video access and modify audit trails, improving security and trust among stakeholders and workers.
- **Access Control and Authentication**: Multi-factor authentication and role-based access control systems ensure that only authorised personnel can access sensitive data (audio, images, videos and other data). Having tiered access, biometric authentication, and geolocation-aware systems will be useful.
- **Physical Security**: Implementing anti-tampering and anti-theft protection mechanisms to ensure monitoring system integrity.
- **Data Retention and Deletion Policies**: Implement policies and procedures for data retention periods and securely (and irreversibly) delete data after the retention period.

Module 3 – Inspection Report Generator

The **Inspection Report Generator** module serves as the final and essential component in our integrated prototype, transforming raw inspection data (e.g., photos and notes) collected via the AR-assisted tool into actionable safety documentation. This module bridges the immersive field data collection process with backend processing to generate consistent, structured, and auditable inspection outputs. It plays a crucial role in ensuring that observations made on-site are systematically captured, synthesised, and communicated to relevant stakeholders, including project managers, EH&S coordinators, and regulatory personnel.

Figure 31. A system architecture chart shows the information flow between different technologies



As illustrated in Figure 31, the system architecture for this module is designed to support seamless data flow between the on-site interface and cloud-based reporting services. During the inspection process, site personnel use the Augmented (or Mixed) Reality interface, implemented on a device such as the Apple Vision Pro, to capture photographs, voice annotations, and textual notes. These inputs are synchronised to a local laptop server via HTTP messaging. The server manages real-time coordination between the AR interface and a locally hosted AI-based gap detection model, which tags the captured images with classification outcomes and segmentation overlays.

Once the inspection is completed, all media and metadata are synchronised to a Firebase[8] online server. A custom Python script[9] then downloads the relevant files and parses image-annotation pairs, safety classifications, and location metadata. This information is

[8] https://firebase.google.com/
[9] The Python script is available on Microsoft Teams and shared with the industry partner.

programmatically assembled into a templated, HTML-based inspection report, which includes visual evidence, descriptive observations, and inspector notes. The generated report is archived and can be reviewed, versioned, and forwarded to project leads and compliance teams.

Importance and impact

Historically, the process of safety inspection reporting has relied heavily on manual entry, delayed documentation, and inconsistent formatting. This not only hampers timely response to identified risks but also introduces the possibility of human error and data loss. By automating the reporting process and embedding data capture directly into the inspection workflow, the proposed system significantly reduces administrative burden, enhances traceability, and ensures completeness of information.

Moreover, the integration with AR and AI technologies allows the report to reflect context-rich insights: images are marked with precise gap locations and types, while voice notes are transcribed and timestamped in line with visual evidence. This results in reports that are both highly detailed and efficient to review, supporting faster risk mitigation actions, improved compliance with site safety protocols, and more reliable audit trails.

Overall, the inspection report generator transforms fragmented observations into a single, integrated digital document. It demonstrates how real-time data collection, intelligent analysis, and automated documentation can be combined to improve both operational efficiency and safety outcomes in modern construction environments.

Expert evaluation through interviews

To evaluate the feasibility and usefulness of the developed prototype, we conducted qualitative interviews with two EH&S leads from the industry partner. These experts were selected for their operational experience and strategic oversight of safety practices across the industry partner's construction projects. They had knowledge and experience with inspection tasks in high-rise buildings, which was the context chosen for our prototype.

The interviews occurred on 4 April 2025, led by the lead CI and the Research Assistant, and were conducted via video conference. The average duration was 60 minutes. Before the interviews, both participants were provided with an explanatory statement and a consent form. Both participants provided signed consent before the interviews began.

The interview process comprised three parts. In the first part, we collected information about each participant's role, responsibilities, and prior experience with AR systems. In the second part, participants watched a five-minute demonstration video highlighting the key features and interactions of the AR-assisted inspection tool, the AI-based gap detection system, and the automated inspection report generator. In the final part, participants shared their feedback on the prototype's design, feasibility, usability, and potential areas of application.

The full list of interview questions is included in the Appendix II.

Ethics approval (attached in the Appendix III) for the study was obtained from Monash University Human Research Ethics approval committees (Project ID: 45953).

This information is not publicly available.

CONCLUSION AND FUTURE RESEARCH

Over the course of this two-year research project, we investigated the landscape of construction site safety—specifically in the context of falls from height—through a review of existing studies and market analyses, regular stakeholder meetings, workshops and interviews, access to collected data, and site observation reports and images. We also explored and tested ways to combine and leverage advanced technologies to develop innovative solutions for mitigating safety risks. Through collaboration with our industry partner we identified falling objects as the highest-rated risk in high-rise building construction and designed a prototype system to support both experienced site managers and newcomers. The developed prototype provides AI-assisted risk detection, automates documentation, enhances situational awareness, and facilitates real-time data capture about safety risks through voice and immersive interactions. Specifically, we list our contributions below for each project stream as part of this scoping study:

Stream I: Data exploration and visualisation for construction site safety management

This stream focused on understanding the industry partner's current practices in construction site safety management through a collaborative workshop, system demonstrations, and in-depth data exploration. Then, through a collaborative and iterative in-depth analysis of the industry partner's safety record data, we explored and identified patterns of incidents across different aspects, such as safety trends over time, environmental influences on incident severity, and reporting behaviours by role and region. We also identified gaps in data collection and proposed opportunities for enhancing predictive safety capabilities. Insights from this stream laid the groundwork for targeted improvements and informed the development of more advanced safety analytics in subsequent phases.

Stream II: A review of state-of-the-art technologies for construction site safety management

This stream focused on studying and analysing emerging technologies with the potential to improve safety management on construction sites. Through a structured literature review and market analysis, we examined 50 academic publications and 23 commercial solutions. Technologies were classified across dimensions such as data sources (e.g., vision-based cameras, motion-tracking sensors), analytical methods (e.g., deep learning, statistical models), and application areas (e.g., PPE detection, behaviour monitoring, object tracking). While camera-based systems and CNNs dominate current research, gaps remain in real-world deployment, falling object detection, and integration of advanced analytics. The market review highlighted a reliance on basic CCTV systems, with limited adoption of immersive tools such as AR/VR or predictive AI. This stream provided a foundational understanding of the technology landscape, identified key limitations, and outlined future opportunities for deploying intelligent, proactive safety solutions in construction environments.

Stream III: Design and development of a concept prototype

This stream focused on designing and developing a concept prototype to improve safety inspection by leveraging a combination of emerging technologies. Through a collaborative process with our industry partner, we first identified key fall-related risks (i.e., gaps between perimeter screens, loose edge-protection strips, and floor openings) and used these to build computer vision models for classifying different types of gap scenarios. We then formulated the prototype's functional goals to automate gap detection and enhance situational awareness through spatial context by integrating the computer vision models into the AR system, as well as to facilitate the creation of inspection reports through the AR system. The developed prototype comprised three main modules: an AI-assisted gap detection system, an AR-assisted inspection tool, and an automated inspection report generator. Each module was designed and developed based on feedback from domain experts, using real-world construction site data and scenarios to ensure contextual relevance and practical utility.

The prototype was evaluated through interviews with two industry partner EH&S leads. These interviews confirmed the prototype's potential to improve inspection efficiency through its role as a training tool for both novice and experienced users. While concerns were raised about privacy and headset usability, the overall response was positive, with several forward-looking suggestions for future system enhancements. These insights validate the prototype's relevance and offer clear directions for refinement and broader implementation.

In summary, this project explored construction site safety through three streams: understanding the industry partner's safety practices at the time, reviewing state-of-the-art technologies, and designing a concept prototype. The resulting prototype combines AI and AR to support proactive, immersive safety inspections by automatically detecting fall-related risks (such as perimeter screen gaps and floor openings), enhancing situational awareness, and enabling real-time inspection on a floor-by-floor basis in high-rise buildings. It also reduces time-consuming manual tasks by generating automated inspection reports and offers strong potential as a training tool for both novice and experienced users.

Ultimately, this project has laid the foundational groundwork for a new paradigm in construction safety, one that leverages modern technologies, real-time data, and is mindful of privacy concerns. However, open issues remain regarding the integration of such systems into existing practices and workflows, their compliance with regulations and standards, and the need to ensure their long-term sustainability and scalability.

Future research

Building on the successes and learnings of this project, several avenues for future research are both necessary and promising:

Privacy preserving mechanisms

Technology solutions involving on-site or wearable cameras and tracking technology do carry risks of compromising worker privacy. In principle, these can be mitigated through edge-computing anonymisation of data (e.g. blurring of faces). However, this barrier to technological acceptance must be considered in future deployments through transparent policies, consent-based design, and ongoing engagement with workers and stakeholders.

Data governance, ethics, and trust

As more visual and location-based data are captured, there is a growing need for research and collaborative solutions involving researchers, system developers, policymakers, and industry representatives. The goal is to advance construction workflows and practices through automation and the adoption of modern technologies, while ensuring transparency, privacy, and trust. Further work is required to design consent-driven, anonymised, and secure data protocols that align with regulatory requirements and meet the expectations of workers.

Real-time risk detection and feedback loops

Further investigation is needed into the automated detection of near-miss scenarios or high-risk behaviours using real-time video analytics and wearables. Future systems should not only observe but also initiate timely, context-aware alerts to workers or supervisors, closing the loop between detection and intervention.

Worker-centred design and interface adaptation

As mobile technology is increasingly integrated into safety workflows, there is a critical need to study how different user groups, such as construction workers, site managers and supervisors, and EH&S professionals, interact with digital inspection tools. Tailoring user interfaces to users' skill levels, cognitive loads, and task urgencies will be key to effective adoption.

Longitudinal impact studies

While our project provided immediate insight into utility and acceptance, long-term studies are needed to evaluate how such systems influence injury rates, safety culture, and regulatory compliance over time. Embedding evaluation mechanisms into future pilots and on-site deployment will support the scaling of evidence-based initiatives.

Continuing this research agenda can enable the construction industry to move toward a future where safety is not merely monitored but intelligently managed, anticipating hazards and proactively mitigating risks, empowering teams, and fostering a culture of continuous improvement.

REFERENCES

Abdelhamid, T. S., & Everett, J. G. (2000). Identifying Root Causes of Construction Accidents. *Journal of Construction Engineering and Management*, *126*(1), 52–60. https://doi.org/doi:10.1061/(ASCE)0733-9364(2000)126:1(52).

Aneziris, O., Papazoglou, I., Mud, M., Damen, M., Bellamy, L., Manuel, H., & Oh, J. (2014). Occupational risk quantification owing to falling objects. *Safety science*, *69*, 57–70.

Anjum, S., Khan, N., Khalid, R., Khan, M., Lee, D., & Park, C. (2022). Fall prevention from ladders utilizing a deep learning-based height assessment method. *IEEE Access*, *10*, 36725–36742.

Antwi-Afari, M. F., Li, H., Anwer, S., Yevu, S. K., Wu, Z., Antwi-Afari, P., & Kim, I. (2020). Quantifying workers' gait patterns to identify safety hazards in construction using a wearable insole pressure system. *Safety science*, *129*, 104855.

Arowoiya, V. A., Oke, A. E., Aigbavboa, C. O., & Aliu, J. (2020). An appraisal of the adoption internet of things (IoT) elements for sustainable construction. *Journal of Engineering, Design and Technology*, *18*(5), 1193–1208.

Athidhi, B., & Smitha Vas, P. (2023). YOLOv7-based model for detecting safety helmet wear on construction sites. International Conference on Intelligent Sustainable Systems.

Chang, C.-H., Chuang, M.-L., Tan, J.-C., Hsieh, C.-C., & Chou, C.-C. (2022). Indoor safety monitoring for falls or restricted areas using wi-fi channel state information and deep learning methods in mega building construction projects. *Sustainability*, *14*(22), 15034.

Chen, H., Hou, L., Wu, S., Zhang, G., Zou, Y., Moon, S., & Bhuiyan, M. (2024). Augmented reality, deep learning and vision-language query system for construction worker safety. *Automation in Construction*, *157*, 105158.

Chen, H., Li, Y., Wen, H., & Hu, X. (2023). YOLOv5s-gnConv: detecting personal protective equipment for workers at height. *Frontiers in public health*, *11*, 1225478.

Cheng, J. P., Wong, P. K.-Y., Luo, H., Wang, M., & Leung, P. H. (2022). Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. *Automation in Construction*, *139*, 104312.

Cheng, M.-Y., Vu, Q.-T., & Teng, R.-K. (2024). Real-time risk assessment of multi-parameter induced fall accidents at construction sites. *Automation in Construction*, *162*, 105409.

Chian, E., Fang, W., Goh, Y. M., & Tian, J. (2021). Computer vision approaches for detecting missing barricades. *Automation in Construction*, *131*, 103862.

Choo, H., Lee, B., Kim, H., & Choi, B. (2023). Automated detection of construction work at heights and deployment of safety hooks using IMU with a barometer. *Automation in Construction*, *147*, 104714.

Duan, P., Goh, Y. M., & Zhou, J. (2023). Personalized stability monitoring based on body postures of construction workers working at heights. *Safety science*, *162*, 106104.

Dzeng, R.-J., Watanabe, K., Hsueh, H.-H., & Fu, C.-K. (2024). A GRU-based model for detecting common accidents of construction workers. *Sensors*, *24*(2), 672.

Fang, W., Ma, L., Love, P. E., Luo, H., Ding, L., & Zhou, A. (2020). Knowledge graph for identifying hazards on construction sites: Integrating computer vision with ontology. *Automation in Construction*, *119*, 103310.

Gomez-de-Gabriel, J. M., Fernández-Madrigal, J. A., López-Arquillos, A., & Rubio-Romero, J. C. (2019). Monitoring harness use in construction with BLE beacons. *Measurement*, *131*, 329–340.

Guo, H., Zhang, Z., Yu, R., Sun, Y., & Li, H. (2023). Action recognition based on 3D skeleton and LSTM for the monitoring of construction workers' safety harness usage. *Journal of Construction Engineering and Management*, *149*(4), 04023015.

Guo, R., Li, H., Han, D., & Liu, R. (2023). Feasibility analysis of using Channel State Information (CSI) acquired from Wi-Fi routers for construction worker fall detection. *International journal of environmental research and public health*, *20*(6), 4998.

Halabi, Y., Xu, H., Long, D., Chen, Y., Yu, Z., Alhaek, F., & Alhaddad, W. (2022). Causal factors and risk assessment of fall accidents in the US construction industry: A comprehensive data analysis (2000–2020). *Safety science*, *146*, 105537.

Han, Z., & Wang, Y. (2024). Inertial Measurement Unit-and Time Series Approach-Based Motion Trajectory Reconstruction of the Safety Rope Fastening Behaviour. *Buildings*, *14*(1), 219.

Hsiao, H., & Simeonov, P. (2001). Preventing falls from roofs: a critical review. *Ergonomics*, *44*(5), 537–561.

Hu, Q., Bai, Y., He, L., Huang, J., Wang, H., & Cheng, G. (2022). Workers' unsafe actions when working at heights: Detecting from images. *Sustainability*, *14*(10), 6126.

Huang, H., Hu, H., Xu, F., & Zhang, Z. (2024). Kinesiology-Inspired Assessment of Intrusion Risk Based on Human Motion Features. *Journal of Construction Engineering and Management*, *150*(7), 04024072.

Huang, H., Hu, H., Xu, F., Zhang, Z., & Tao, Y. (2023). Skeleton-based automatic assessment and prediction of intrusion risk in construction hazardous areas. *Safety science*, *164*, 106150.

Huang, X., & Hinze, J. (2003). Analysis of construction worker fall accidents. *Journal of Construction Engineering and Management*, *129*(3), 262–271.

Jin, Z., & Gambatese, J. Development of a Cost-Effective Proximity Warning System for Fall Protection. In *Computing in Civil Engineering 2023* (pp. 375–382).

Johansen, K. W., Hong, K., Schultz, C., & Teizer, J. (2024). Automated quantification of construction workers' exposure to falling object hazards. *Proceedings of the Institution of Civil Engineers-Management, Procurement and Law*, *177*(4), 207–222.

Khan, M., Khalid, R., Anjum, S., Khan, N., Cho, S., & Park, C. (2022). Tag and IoT based safety hook monitoring for prevention of falls from height. *Automation in Construction*, *136*, 104153.

Khan, M., Khalid, R., Anjum, S., Tran, S. V.-T., & Park, C. (2022). Fall prevention from scaffolding using computer vision and IoT-based monitoring. *Journal of Construction Engineering and Management*, *148*(7), 04022051.

Khan, M., Nnaji, C., Khan, M. S., Ibrahim, A., Lee, D., & Park, C. (2023). Risk factors and emerging technologies for preventing falls from heights at construction sites. *Automation in Construction*, *153*, 104955.

Khan, N., Saleem, M. R., Lee, D., Park, M.-W., & Park, C. (2021). Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks. *Computers in Industry*, *129*, 103448.

Lee, H., Lee, G., Lee, S., & Ahn, C. R. (2022). Assessing exposure to slip, trip, and fall hazards based on abnormal gait patterns predicted from confidence interval estimation. *Automation in Construction*, *139*, 104253.

Lee, J., & Lee, S. (2023). Construction site safety management: a computer vision and deep learning approach. *Sensors*, *23*(2), 944.

Lee, S., Koo, B., Yang, S., Kim, J., Nam, Y., & Kim, Y. (2022). Fall-from-height detection using deep learning based on IMU sensor data for accident prevention at construction sites. *Sensors*, *22*(16), 6107.

Li, J., Ouyang, Y., & Luo, X. (2024). Towards an EEG-Based Approach for Detecting Falls from Height Hazards Using Construction Workers' Physiological Signals. Construction Research Congress 2024.

Li, J., Zhao, X., Zhou, G., & Zhang, M. (2022). Standardized use inspection of workers' personal protective equipment based on deep learning. *Safety science*, *150*, 105689.

Li, P., Wang, Q., Guo, Z., Mei, T., Li, Q., Qiao, S., & Zuo, W. (2022). Identifying falling-from-height hazards in building information models: A voxelization-based method. *Journal of Construction Engineering and Management*, *148*(2), 04021203.

Li, X., Long, Y., Yang, C., Li, Q., Lu, W., & Gao, J. (2023). Research on psychophysiological characteristics of construction workers during consciously unsafe behaviors. *Heliyon*, *9*(10).

Lim, J., Jung, D. G., Park, C., & Kim, D. Y. (2022). Computer vision process development regarding worker's safety harness and hook to prevent fall accidents: focused on system scaffolds in South Korea. *Advances in Civil Engineering*, *2022*(1), 4678479.

Lin, P.-H., Pal, A., Lin, J. J., & Hsieh, S.-H. (2024). Individual Component Detection of a Scaffolding Assembly for Vision-Based Safety Check. Construction Research Congress 2024.

Liu, D., Jin, Z., & Gambatese, J. (2020). Scenarios for integrating IPS–IMU system with BIM technology in construction safety control. *Practice Periodical on Structural Design and Construction*, *25*(1), 05019007.

Liu, X., Xu, F., Zhang, Z., & Sun, K. (2025). Fall-portent detection for construction sites based on computer vision and machine learning. *Engineering, Construction and Architectural Management*, *32*(3), 1499–1521.

Manzoor, B., Othman, I., & Manzoor, M. (2021). Evaluating the critical safety factors causing accidents in high-rise building projects. *Ain Shams Engineering Journal*, *12*(3), 2485–2492.

Nadhim, E. A., Hon, C., Xia, B., Stewart, I., & Fang, D. (2016). Falls from height in the construction industry: A critical review of the scientific literature. *International journal of environmental research and public health*, *13*(7), 638.

Nowobilski, T., & Hoła, B. (2023). Methodology based on causes of accidents for forcasting the effects of falls from scaffoldings using the construction industry in Poland as an example. *Safety science*, *157*, 105945.

Ojha, A., Liu, Y., Shayesteh, S., Jebelli, H., & Sitzabee, W. E. (2023). Affordable multiagent robotic system for same-level fall hazard detection in indoor construction environments. *Journal of Computing in Civil Engineering*, *37*(1), 04022042.

Oke, A. E., & Arowoiya, V. A. (2021). Evaluation of internet of things (IoT) application areas for sustainable construction. *Smart and Sustainable Built Environment*, *10*(3), 387–402.

Oke, A. E., & Arowoiya, V. A. (2022). An analysis of the application areas of augmented reality technology in the construction industry. *Smart and Sustainable Built Environment*, *11*(4), 1081–1098.

Oliveira, S. S., de Albuquerque Soares, W., & Vasconcelos, B. M. (2023). Fatal fall-from-height accidents: Statistical treatment using the Human Factors Analysis and Classification System–HFACS. *Journal of safety research*, *86*, 118–126.

Panuwatwanich, K., Roongsrisoothiwong, N., Petcharayuthapant, K., Dummanonda, S., & Mohamed, S. (2020). Ambient intelligence to improve construction site safety: case of high-rise building in Thailand. *International journal of environmental research and public health*, *17*(21), 8124.

Park, C.-S., & Kim, H.-J. (2013). A framework for construction safety management and visualization system. *Automation in Construction*, *33*, 95–103.

Park, M., Bak, J., Kulinan, A. S., & Park, S. (2023). Real-time monitoring unsafe behaviors of portable multi-position ladder worker using deep learning based on vision data. *Journal of safety research*, *87*, 465–480.

Piao, Y., Xu, W., Wang, T.-K., & Chen, J.-H. (2021). Dynamic fall risk assessment framework for construction workers based on dynamic Bayesian network and computer vision. *Journal of Construction Engineering and Management*, *147*(12), 04021171.

Rey-Merchán, M. d. C., López-Arquillos, A., & Soto-Hidalgo, J. M. (2022). Prevention of falls from heights in construction using an iot system based on fuzzy markup language and JFML. *Applied Sciences*, *12*(12), 6057.

SafeWork Australia. (2024, 02 Sep 2024). *Key Work Health and Safety Statistics Australia 2024*. Retrieved 20 Jun from https://data.safeworkaustralia.gov.au/insights/key-whs-statistics-australia/latest-release

Shanti, M. Z., Cho, C.-S., Byon, Y.-J., Yeun, C. Y., Kim, T.-Y., Kim, S.-K., & Altunaiji, A. (2021). A novel implementation of an AI-based smart construction safety inspection protocol in the UAE. *IEEE Access*, *9*, 166603–166616.

Shanti, M. Z., Cho, C.-S., de Soto, B. G., Byon, Y.-J., Yeun, C. Y., & Kim, T. Y. (2022). Real-time monitoring of work-at-height safety hazards in construction sites using drones and deep learning. *Journal of safety research*, *83*, 364–370.

Shetty, N. P., Himakar, J., Gnanchandan, P., & Jamadagni, S. (2024). Enhancing Construction Site Safety: A Tripartite Analysis of Safety Violations. 2024 3rd International Conference for Innovation in Technology (INOCON).

Soltanmohammadlou, N., Sadeghi, S., Hon, C. K., & Mokhtarpour-Khanghah, F. (2019). Real-time locating systems and safety in construction sites: A literature review. *Safety science*, *117*, 229–242.

Song, K.-S., Lim, J. G., Lee, D.-G., & Kwon, D.-S. (2023). Motion similarity-based safety hook fastening state recognition via deep siamese neural networks. *IEEE Sensors Letters*, *7*(10), 1–4.

Wu, H., Zhong, B., Li, H., Love, P., Pan, X., & Zhao, N. (2021). Combining computer vision with semantic reasoning for on-site safety management in construction. *Journal of Building Engineering*, *42*, 103036.

Wu, W., Hartless, J., Tesei, A., Gunji, V., Ayer, S., & London, J. (2019). Design assessment in virtual and mixed reality environments: Comparison of novices and experts. *Journal of Construction Engineering and Management*, *145*(9), 04019049.

Yang, B., Zhang, B., Zhang, Q., Wang, Z., Dong, M., & Fang, T. (2022). Automatic detection of falling hazard from surveillance videos based on computer vision and building information modeling. *Structure and Infrastructure Engineering*, *18*(7), 1049–1063.

Yuhai, O., Kim, H., Choi, A., & Mun, J. H. (2023). Deep learning-based slip-trip falls and near-falls prediction model using a single inertial measurement unit sensor for construction workplace. 2023 4th International Conference on Big Data Analytics and Practices (IBDAP).

Zhang, M., Cao, T., & Zhao, X. (2019). Using smartphones to detect and identify construction workers' near-miss falls based on ANN. *Journal of Construction Engineering and Management*, *145*(1), 04018120.

Zhang, M., & Ge, S. (2022). Vision and trajectory–based dynamic collision prewarning mechanism for tower cranes. *Journal of Construction Engineering and Management*, *148*(7), 04022057.

Zhang, X., Wang, X., & Lin, J. (2022). Building Fall Safety Early Warning System Based on Directed Weighted Network. International Conference on Big Data Analytics for Cyber-Physical System in Smart City.

APPENDIX I. LITERATURE DATABASE SEARCH QUERY

The below search query is used in the Scopus Advanced Search[10] page:

TITLE-ABS-KEY (("fall from height*" OR "fall* from elevation" OR "fall* injur*" OR "fall* accident*" OR "fall* incident*" OR "fall from roof*" OR "fall from scaffold*" OR "fall from ladder*" OR "fall from platform*" OR "fall to lower level" OR "fall to same level" OR "fall to open*" OR "fall* protection" OR "hazardous fall*" OR "risky fall*" OR "fatal fall*" OR "accidental fall*" OR "occupational fall*" OR "fall prevention" OR "fall hazard" OR "fall risk assessment" OR "fall-related mortality" OR "fall* object*" OR " fall* material*" OR "struck-by-falling-object*" OR "struck-by-falling-material*" OR "struck-by flying object*" OR "Struck-by flying material*" OR "Struck-by falling object*" OR "Struck-by falling material*" OR " Struck-by moving object*") AND ("construction * safety" OR ("construction site" OR "construction jobsite" OR "construction work zone" OR "construction industry" OR "construction workplace" OR "construction work*" OR "construction professional*" OR "construction labo*" OR "construction workforce*" OR "construction staff" OR "construction personnel*" OR "construction activit*") AND ("safety" OR "safety management" OR "risk" OR "risk management" OR "hazard" OR "accident" OR "accident prediction" OR "accident prevention"))) AND PUBYEAR >= 2019 AND PUBYEAR <= 2024 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SRCTYPE , "j"))*

[10] https://www.scopus.com/search/form.uri?display=advanced

APPENDIX II. INTERVIEW QUESTIONS

PART 1

Background Questions

1. What is your **highest level of education**?
2. What is your **current job title**?
3. **How many years** have you been working in this position?
4. How familiar are you with **virtual reality** (VR) and **augmented reality** (AR) technologies?
    a. Have you ever used VR or AR technologies? If yes, could you describe **your experience**?

Role and responsibilities

1. Can you describe your **main responsibilities** in your current role **when inspecting** safety hazards on construction sites?
    a. Would you describe your tasks as collaborative, individual, or a mix of both? Could you provide some examples?
2. How do you **keep track** of where an **inspection starts and ends** on a construction site, and ensure that all areas are **thoroughly covered**?
    a. Are there any checkpoints you follow?
    b. In multi-level buildings, what is the order in which you inspect each floor?
3. How do you **document and report** your inspections and findings?
    a. Do you follow a **standard template or guideline** for creating reports, or do you tailor them to each project?
    b. Are there any **floor plans or graphical markings** that you use during inspection?
    c. When submitting the final report, **how confident** are you that the procedures and technology used have provided a thorough and accurate inspection? Please rate your confidence on a scale from **0 to 100%.**

Technology

1. **What tools or technologies** do you currently use to assist with safety inspections?
    a. How **effective** do you find them?
    b. Are there any features or functionalities you feel are **missing** from the tools you currently use?

2. What are the **key challenges** you encounter in your role as a safety inspector on a daily basis?
    a. Can you share **an example** of a specific situation where you faced one of these challenges?
    b. In what ways do you think technology could help address these challenges, including automating manual tasks?

PART 2

The Demo

We will present a demo of an AR system designed for safety inspections. The demo will showcase key features, including tracking inspected areas, capturing photos of potential safety risks, analysing them remotely using visual models to detect parameter gaps, and adding comments through voice commands or a virtual keyboard.

PART 3

Post-demo feedback

1. How helpful do you find the 3D visualisation of the **surrounding environment** and **highlighting the areas that have been inspected** (tracking inspected areas)?
2. How helpful do you find the interaction techniques for **detecting potential safety hazards** from the photos taken (the Visual Model feature)?
3. How helpful do you find the interaction techniques for **capturing photos** and **entering comments** to **automatically generate reports**?
4. How **comfortable** would you feel using this AR system in a real-world setting?
5. Do you think this AR system could **improve the efficiency or accuracy** of safety inspections? Why or why not?
6. What kind of **training** or support do you think would be necessary to help you or your coworkers effectively adopt this AR system?
7. Do you see any **barriers** that might prevent this AR system from being adopted in your workplace? If so, what are they?
8. How well do you think the AR system can **fit into your current workflow**? Would it require any major adjustments?
9. Beyond gap detection in barriers and screens, what other **important applications** do you see for this type of AR technology?
10. Based on what you saw today, do you see any **specific use cases** for this technology in your work?
11. Are there any features of the AR system that you particularly **liked** in relation to your role? If so, why?
12. Are there any features of the AR system that you particularly **disliked** in relation to your role? If so, why?

APPENDIX III. ETHICS APPROVAL

MONASH University

Monash University Human Research Ethics Committee

Approval Certificate

This is to certify that the project below was considered by the Monash University Human Research Ethics Committee. The Committee was satisfied that the proposal meets the requirements of the *National Statement on Ethical Conduct in Human Research* and has granted approval.

Project ID:	45953
Application Type:	Human Ethics Low Risk
Project Title:	Context-Aware Fusion of AR, Vision Models, and LLMs for Safety Inspection
Chief Investigator:	Dr Pari Delir Haghighi
Approval Date:	06/01/2025
Expiry Date:	06/01/2030

Terms of approval - failure to comply with the terms below is in breach of your approval and the *Australian Code for the Responsible Conduct of Research.*

1. The Chief Investigator is responsible for ensuring that permission letters are obtained, if relevant, before any data collection can occur at the specified organisation.
2. Approval is only valid whilst you hold a position at Monash University.
3. It is the responsibility of the Chief Investigator to ensure that all investigators are aware of the terms of approval and to ensure the project is conducted as approved by MUHREC.
4. You should notify MUHREC immediately of any serious or unexpected adverse effects on participants or unforeseen events affecting the ethical acceptability of the project.
5. The Explanatory Statement must be on Monash letterhead and the Monash University complaints clause must include your project number.
6. Amendments to approved projects including changes to personnel must not commence without written approval from MUHREC.
7. Annual Report - continued approval of this project is dependent on the submission of an Annual Report.
8. Final Report - should be provided at the conclusion of the project. MUHREC should be notified if the project is discontinued before the expected completion date.
9. Monitoring - the project may be subject to an audit or any other form of monitoring by MUHREC at any time.
10. Retention and storage of data - The Chief Investigator is responsible for the storage and retention of the original data pertaining to the project for a minimum period of five years.

Kind Regards,

Professor William Sievert

Chair, MUHREC

CC: Mr Fucai Ke, Dr Joe Liu, Professor Tim Dwyer, Dr Hamid Rezatofighi, Dr Benjamin Tag, Dr Teresa Wang, Dr Aravinda Rao

List of approved documents:

Document Type	File Name	Date	Version
Consent Form	consent-form-Delir	20/12/2024	1
Supporting Documentation	Email-Invitation-Delir	20/12/2024	1
Explanatory Statement	explanatory-statement-Delir et al	30/12/2024	2
Supporting Documentation	InterviewQuestions	31/12/2024	1

Page 1 of 1

MONASH University

Monash University Human Research Ethics Committee

Approval Certificate

This is to certify that the project below was considered by the Monash University Human Research Ethics Committee. The Committee was satisfied that the proposal meets the requirements of the *National Statement on Ethical Conduct in Human Research* and has granted approval.

Project ID:	41751
Application Type:	Human Ethics Low Risk
Project Title:	Using Computer Vision to Reduce Risks and Improve Construction Site Safety
Chief Investigator:	Dr Pari Delir Haghighi
Approval Date:	29/02/2024
Expiry Date:	01/03/2029

Terms of approval - failure to comply with the terms below is in breach of your approval and the *Australian Code for the Responsible Conduct of Research*.

1. The Chief Investigator is responsible for ensuring that permission letters are obtained, if relevant, before any data collection can occur at the specified organisation.
2. Approval is only valid whilst you hold a position at Monash University.
3. It is the responsibility of the Chief Investigator to ensure that all investigators are aware of the terms of approval and to ensure the project is conducted as approved by MUHREC.
4. You should notify MUHREC immediately of any serious or unexpected adverse effects on participants or unforeseen events affecting the ethical acceptability of the project.
5. The Explanatory Statement must be on Monash letterhead and the Monash University complaints clause must include your project number.
6. Amendments to approved projects including changes to personnel must not commence without written approval from MUHREC.
7. Annual Report - continued approval of this project is dependent on the submission of an Annual Report.
8. Final Report - should be provided at the conclusion of the project. MUHREC should be notified if the project is discontinued before the expected completion date.
9. Monitoring - the project may be subject to an audit or any other form of monitoring by MUHREC at any time.
10. Retention and storage of data - The Chief Investigator is responsible for the storage and retention of the original data pertaining to the project for a minimum period of five years.

Kind Regards,

Emeritus Professor Napier (Nip) Thomson

Chair, MUHREC

CC: Mr Fucai Ke, Dr Joe Liu, Professor Tim Dwyer, Dr Benjamin Tag, Dr Teresa Wang, Dr Hamid Rezatofighi, Dr Philip Christopher, Dr Aravinda Rao, Prof Tuan Ngo

List of approved documents:

building 4.0 crc

info@building40crc.org
www.building4pointzero.org
/building-4-0-crc