ITcon

www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

DEEP LEARNING ENABLED COMPUTER VISION MODEL FOR AUTOMATED SAFETY COMPLIANCE IN CONSTRUCTION ENVIRONMENTS

SUBMITTED: May 2025
REVISED: August 2025
PUBLISHED: September 2025
EDITOR: Frédéric Bosché
DOI: 10.36680/j.itcon.2025.057

Amr A. Mohy, PhD Student, MSc, BSc, PMP, PRINCE2, PSM-1
Construction and Building Engineering Department, Arab Academy for Science and Technology and Maritime Transport, Egypt
A.el-deen5146@student.aast.edu

Hesham A. Bassioni, Professor of Construction Management, PhD, MSc, MBA, BSc, PMP
Construction and Building Engineering Department, Arab Academy for Science and Technology and Maritime Transport, Egypt,
hbassioni@aast.edu

Elbadr O. Elgendi, Associate Professor of Construction Engineering and Management, PhD, MSc, BSc
Construction and Building Engineering Department, Arab Academy for Science and Technology and Maritime Transport, Egypt
elbadrosman@aast.edu

Tarek M. Hassan, Professor of Construction Informatics, BSc., MSc., PhD, MASCE, FCIOB
School of Architecture, Building and Civil Engineering, Loughborough University, United Kingdom
T.Hassan@Lboro.ac.uk

SUMMARY: *Construction site safety demands proactive hazard detection, a challenge traditionally met with reactive measures that are often inadequate. This paper introduces a novel deep learning-based computer vision model designed for automated safety compliance monitoring, addressing critical limitations of existing approaches. The model utilizes a modified one-stage object detection algorithm, uniquely enhanced with Contextual Transformer Networks (CoTs), a Triplet Attention module, Activate or Not (ACON) activation functions, and Content-Aware Reassembly of Features (CARAFE) up-sampling, to significantly improve feature extraction, visual recognition, and contextual understanding in complex construction environments. To support this model development, a new OSHA-data-driven dataset of 55,594 images across 28 safety categories was developed. This dataset encompasses personal protective equipment (PPE), scaffolding, materials, hazards, and worker actions, ensuring comprehensive coverage of key safety domains. The Wise-Intersection over Union (IoU) loss function further refines bounding box regression, enhancing localization accuracy. Evaluations on both a benchmarking dataset and the newly developed dataset demonstrate the model's benchmark-surpassing performance (Precision: 0.89, mAP95: 0.45). This research offers a practically viable, data-driven solution for a critical industry challenge, moving towards a future of zero-accident construction sites.*

KEYWORDS: *Construction Safety Management, Artificial Intelligence, Automated Hazard Detection, Object Detection, Computer Vision.*

REFERENCE: *Amr A. Mohy, Hesham A. Bassioni, Elbadr O. Elgendi & Tarek M. Hassan (2025). Deep Learning Enabled Computer Vision Model for Automated Safety Compliance in Construction Environments. Journal of Information Technology in Construction (ITcon), Vol. 30, pg. 1398-1430, DOI: 10.36680/j.itcon.2025.057*

1. INTRODUCTION

Every single day, an estimated 4,800 construction workers around the world lose their lives due to work-related accidents and diseases. This grim reality, highlighted by the International Labour Organization, underscores the urgent need for innovative solutions to enhance safety in the construction industry, one of the most hazardous sectors globally (BUREAU OF LABOR STATISTICS, 2022; Hou et al., 2023). Traditional safety measures, while essential, often prove inadequate in addressing the dynamic and complex nature of construction environments, leading to a significant number of accidents and fatalities (Nath et al., 2020). These incidents not only result in tragic loss of life and serious injuries but also inflict substantial financial burdens on companies due to project delays, legal liabilities, and compensation claims (Waehrer et al., 2007). Traditional safety management, which relies on manual inspections and reactive incident reporting, is often insufficient to address the dynamic and complex hazards of modern construction sites (Nath et al., 2020). This reactive posture not only leads to tragic human loss but also incurs significant financial costs from project delays, legal liabilities, and regulatory penalties (Waehrer et al., 2007).

To overcome these limitations, the field is shifting towards proactive, automated safety monitoring using Artificial Intelligence (AI) and computer vision. Deep learning models, particularly object detectors like YOLO (You Only Look Once), have shown promise in identifying safety compliance issues such as missing Personal Protective Equipment (PPE) in real-time. However, existing research faces three primary challenges: (1) Dataset Limitations: Publicly available datasets often lack the scale, diversity, and specific hazard categories (e.g., scaffolding faults, unsafe actions) needed to train robust models for comprehensive site monitoring (Duan et al., 2022). (2) Algorithmic Limitations: Standard object detection models are not inherently optimized for the unique visual complexities of construction sites, such as severe object occlusion, varying illumination, and the need to distinguish subtle contextual hazards. (3) Practical Implementation: Many proposed models are either too computationally heavy for real-time deployment or not accurate enough to be reliable, while integration into existing workflows requires careful consideration of human-computer interaction and ethical concerns like data privacy and algorithmic bias (Fang et al., 2018; Nath et al., 2020; Xiao et al., 2022; Yang et al., 2023).

This research directly addresses this critical gap by developing a novel deep learning-based computer vision model designed for real-time hazard identification and tracking, offering significant advancements over existing approaches. The key contributions of the proposed model are: (1) the unique integration of Contextual Transformer Networks (CoTs) (Li et al., 2021) and Triplet Attention (Misra et al., 2021) within a modified YOLOv8 architecture, enhancing feature extraction and contextual understanding for improved hazard detection; (2) the incorporation of ACON activation functions (Ma et al., 2021a) and CARAFE up-sampling (Wang et al., 2019), optimizing model adaptability and localization precision for more precise hazard localization (Tong et al., 2023); and (3) the development of a new, large-scale dataset of 55,594 images across 28 safety categories, which are empirically derived from an analysis of OSHA Severe Injury Reports (OSHA, 2024), ensuring direct relevance to preventing high-consequence accidents. (4) benchmarking against state-of-the-art models on both a standard dataset (SHEL5K) (Otgonbold et al., 2022) and the developed dataset demonstrates the superior performance and adaptability of the proposed approach, establishing a new benchmark for automated safety compliance systems in construction. (5) Practical Application and Seamless Integration: The model is designed for real-time implementation and seamless integration into existing safety frameworks, prioritizing user acceptance and practical utility for construction professionals. By addressing these key areas, the model aims to significantly improve construction safety management practices, moving towards a more proactive and preventative paradigm.

2. AUTOMATED SAFETY MANAGEMENT SYSTEMS

2.1 Object Detection and Tracking Algorithms

Deep learning-based object detection and tracking algorithms are fundamental to enabling automated and proactive safety management in the dynamic environment of construction sites. These algorithms process visual data from images and videos to identify, locate, and monitor key entities such as workers, equipment, and materials in real-time, facilitating immediate hazard detection and response (Jeelani et al., 2021). Object detection, the cornerstone of this process, focuses on identifying and precisely localizing objects within individual images (Felzenszwalb et al., 2010), providing crucial semantic understanding of the visual scene (Zhao et al., 2019).

Object detection methodologies are broadly categorized into two main approaches: region proposal-based methods and classification-based methods. Region proposal-based methods, including R-CNN variants like Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2017), first identify potential object regions within an image and then classify these regions. While offering high accuracy, their two-stage process can be computationally intensive, potentially limiting real-time performance. Conversely, classification-based methods, such as YOLO (Redmon et al., 2016) and SSD (Liu et al., 2016), employ a single-stage approach, directly predicting object classes and locations in a single pass. This streamlined architecture enables significantly faster processing speeds, making them particularly suitable for real-time applications in dynamic construction environments. However, this speed advantage can sometimes come at the cost of reduced accuracy in complex scenes or for small object detection.

Recent advancements have demonstrated the increasing efficacy of object detection algorithms for enhancing construction safety. For example, (Huang et al., 2021) showcased an improved YOLOv3 algorithm capable of accurately detecting construction workers wearing helmets in real-time, achieving a high mAP of 93.1% and a processing speed of 55 fps. Similarly, (Hayat and Morgado-Dias, 2022) utilized YOLOv5 for automated safety helmet detection, achieving a mAP of 92.44% on a dataset of 5,000 hard hat images. These studies highlight the potential of YOLO-based algorithms for real-time PPE monitoring, addressing a critical safety concern on construction sites. Furthermore, (Xu et al., 2022) introduced Efficient-YOLOv5, specifically optimized for detecting safety harnesses, achieving high precision (97.7%) and recall (89.3%) rates, demonstrating the adaptability of YOLO architectures for detecting various types of safety equipment. Transformer-based approaches are also emerging, with (Wu et al., 2019) proposing a transformer network for safety helmet and harness monitoring, achieving a high implementation accuracy of 95.9%. These advancements underscore the growing sophistication and effectiveness of object detection algorithms in addressing specific construction safety challenges, paving the way for more proactive and automated safety management systems. While two-stage detectors offer enhanced accuracy, the paramount need for real-time hazard detection and immediate alerts in dynamic construction environments necessitates prioritizing speed and efficiency, making the one-stage YOLO architecture a more judicious choice for this safety-critical application.

Moreover, recent studies have continued to refine YOLO-based architectures. (Wang, 2025) introduced a YOLOv10 model enhanced with transformer backbones, achieving a mean AP50 of 87.3% across key PPE categories by leveraging the transformer's superior ability to capture global context. Concurrently, (Xing et al., 2025) focused on efficiency by proposing GPD-YOLOv8, a lightweight adaptation using Ghost modules that reported a 6.2% mAP gain and a 55.88% increase in FPS over the baseline YOLOv8. The critical role of data quality has also been highlighted, with research showing that targeted image augmentation strategies can yield significant, class-dependent performance gains for non-PPE detection (Park et al., 2025). These advances underscore a key trade-off between computationally intensive, high-accuracy models and lightweight, high-speed alternatives. Our research is strategically positioned to navigate this trade-off by enhancing a robust YOLOv8 architecture with a novel combination of modules designed to significantly boost feature extraction and accuracy without the prohibitive computational overhead of a full transformer backbone.

2.2 Construction Datasets

Table 1: Comparing Current Construction Safety Datasets.

Dataset	Image Count	Object Categories	Key Safety Features	Limitations
SODA	19,846	15	Workers, PPE, Machinery, Layout	Limited geographic diversity
SHEL5K	5,000	6	Helmets, Head, Person (with/without helmet), Face	Narrow focus on helmet safety
MOCS	41,000+	13	Moving objects, limited safety annotations	Limited annotations, non-representative diversity
Proposed Dataset	55,594	28	PPE, Scaffolding, Materials & Equipment, Hazards & Falling Actions	Identified features

Recent efforts have been directed toward developing construction-focused image datasets. (Tajeen and Zhu, 2014) compiled a dataset featuring images of construction machinery, covering essential equipment like excavators, loaders, bulldozers, rollers, and backhoe diggers. (Kim et al., 2018) introduced a construction object detection method that combines deep convolutional networks with transfer learning for precise identification of construction

equipment. Their benchmark dataset, AIM, comprises 2,920 images of construction machines. (Kolar et al., 2018) concentrated on detecting guardrails at construction sites, creating an enhanced dataset with 6,000 images by integrating background imagery with three-dimensional guardrail models. (Li et al., 2020) released a dataset of 3,261 images focusing on safety helmets, utilizing SSD-MobileNet for detecting unsafe operations on construction sites. (Xuehui et al., 2021) developed the MOCS dataset, a collection of over 40,000 images from 174 construction sites, annotated for 13 types of moving objects and tested on 15 different deep neural networks. (Wang et al., 2021) assembled a dedicated image dataset named Color Helmet and Vest (CHV) consisting of 1330 images, specifically for PPE detection.

While several construction-focused image datasets exist, such as SODA (Duan et al., 2022), SHEL5K (Otgonbold et al., 2022), and MOCS (Xuehui et al., 2021) these datasets often fall short in terms of comprehensiveness, diversity, and contextual relevance to real-world construction safety management as showing in Table 1. SODA, while encompassing a large number of images, primarily focuses on workers, PPE, and machinery detection, with limited geographic diversity. SHEL5K, though valuable for helmet safety detection, lacks broader applicability for a wider array of safety-related functions. MOCS, despite its large size, has limited annotations, failing to capture the full range of information present in the images, particularly concerning safety-related details. Additionally, its focus on object detection may not fully address the needs of comprehensive safety management. Existing datasets may focus on a limited number of safety categories, lack representation of diverse construction environments and practices, or have insufficient annotations for training robust deep learning models. Addressing these limitations necessitates the development of a new dataset that is more comprehensive, diverse, and explicitly addresses the shortcomings of existing datasets in terms of scope, diversity, and real-world relevance to construction safety management, as highlighted in Table 1. This new dataset should be meticulously designed to capture the full spectrum of safety concerns, diverse construction scenarios, and contextual richness necessary for training robust and practically applicable AI-powered safety solutions.

2.3 Limitations of Existing Research and Need for Improvement

While the reviewed research demonstrates the promising potential of object detection and tracking algorithms for enhancing construction safety, a significant gap persists between theoretical advancements and practical, real-world impact. Existing research, while valuable, often suffers from limitations that hinder the development of truly robust, reliable, and deployable AI-powered safety management systems. These limitations underscore the critical need for further research and development efforts focused on addressing the identified gaps in datasets, algorithms, and practical implementation strategies.

Dataset Limitations: Previous benchmark datasets like SHEL5K are constrained to single PPE types, limiting their utility for comprehensive safety monitoring. Others like MOCS, while large, lack the detailed safety annotations required to train models with high precision across multiple hazard categories. This often results in performance ceilings, where even advanced models struggle to exceed 80-90% mAP without extensive, task-specific fine-tuning. This motivates the development of our broader, OSHA-driven dataset.

Algorithmic Enhancements are Needed for Robust and Real-Time Performance: While state-of-the-art object detectors show promise, they often present an unresolved trade-off for the construction domain. High-accuracy models leveraging computationally intensive transformer architectures can achieve benchmark-setting precision (Wang, 2025) but may be too slow for real-time hazard alerts on typical on-site hardware. Conversely, lightweight models can achieve impressive speeds over 90 FPS but may sacrifice detection accuracy for small or occluded objects common in cluttered construction scenes (Xing et al., 2025). A clear gap exists for an architecture that improves detection robustness in these challenging conditions without compromising real-time viability. The proposed work addresses this by integrating targeted modules like Contextual Transformers (Li et al., 2021) and Triplet Attention (Misra et al., 2021) to enhance feature representation, striking a practical balance between accuracy and efficiency.

Practical Implementation and User Acceptance Challenges Must be Overcome: Beyond technical advancements in datasets and algorithms, the successful deployment and widespread adoption of AI-powered safety systems hinge on addressing practical implementation challenges and ensuring user acceptance within the construction industry. Concerns related to data privacy, ethical considerations, system integration with existing workflows, user training, and the perceived value proposition of AI safety solutions must be carefully considered and addressed. Research efforts focused on user-centered design, seamless system integration, robust data privacy protocols, and

clear communication strategies are crucial for fostering trust, promoting user adoption, and realizing the full potential of AI to transform construction safety management practices in the real world. Bridging this gap between technical feasibility and practical implement ability is essential for translating promising research into tangible improvements in construction worker safety and well-being.

3. FEATURE IDENTIFICATION AND DATASET DEVELOPMENT

To ensure the developed computer vision model directly addresses the most critical safety needs in the construction industry, a rigorous data-driven approach was adopted for feature identification and dataset development. This approach centered on a comprehensive analysis of the Occupational Safety and Health Administration (OSHA) Severe Injury Reports dataset, a rich source of real-world accident data, providing empirical evidence of prevalent hazards, injury types, and contributing factors in construction site accidents. By grounding the feature identification and dataset design in this objective, real-world accident data, the research aimed to create a highly relevant and practically impactful safety monitoring system, directly addressing the most pressing safety challenges faced by construction workers. The dataset development process, visually summarized in Figure 1, encompassed a multi-faceted strategy focused on data-driven category selection, comprehensive data acquisition, rigorous data preparation, and detailed annotation, ensuring the dataset's relevance, diversity, and accuracy for training a robust and effective deep learning model.

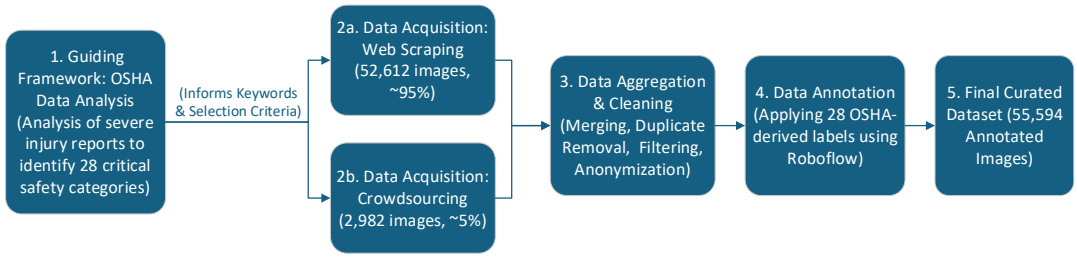

Figure 1: The processes to develop the dataset of the proposed model.

3.1 Features Identification

A data-driven feature identification methodology was employed, leveraging the Occupational Safety and Health Administration (OSHA) Severe Injury Reports dataset (OSHA, 2024) to objectively identify key safety concerns and inform the development of the construction safety dataset. This approach ensured that the dataset categories and the subsequent AI model development were directly grounded in real-world accident data, addressing the most prevalent and impactful safety hazards in the construction industry. The OSHA Severe Injury dataset, covering incidents from January 2015 to March 2024, provides detailed information on construction accidents resulting in severe injuries or fatalities, offering a valuable empirical basis for understanding safety risks in the sector.

3.1.1 Severe Injury Dataset Preprocessing and Classification

The OSHA Severe Injury dataset (OSHA, 2024) underwent a rigorous preprocessing and classification process to extract relevant safety information and identify key feature categories. The dataset, initially comprising 27 columns of incident-level data, was filtered to focus exclusively on construction-related incidents, ensuring the analysis was specific to the target domain. Data preprocessing steps included addressing missing values, standardizing data formats, removing duplicate records, and applying Natural Language Processing (NLP) techniques to the 'Final Narrative' column, which contains detailed textual descriptions of each incident. A classification framework was then developed to systematically categorize safety-related information extracted from the OSHA reports. This framework involved:

- Hazard Identification: Text mining and NLP techniques were applied to the 'Final Narrative' column to extract keywords and phrases describing common hazards. Named Entity Recognition (NER) was used to identify specific equipment, locations, and injury types mentioned in the narratives. Extracted keywords and entities were then grouped into overarching hazard categories, such as Fall Hazards, Struck-by Hazards, Caught-in/Between Hazards, and Electrical Hazards, reflecting the most frequent and severe accident types reported in the OSHA data.

- PPE (Personal Protective Equipment) Detection Context: Incident narratives were analyzed to identify mentions of Personal Protective Equipment (PPE) usage or non-usage. Keywords and phrases related to specific PPE items (hard hats, vests, safety glasses, etc.) and terms indicating non-compliance (e.g., "not wearing," "failure to use") were extracted. Contextual analysis using NLP models was employed to differentiate between compliant and non-compliant PPE scenarios, informing the creation of PPE-related safety categories for the dataset.
- Environmental and Contextual Factors: NLP techniques were used to identify mentions of environmental conditions (e.g., poor lighting, slippery surfaces) and contextual factors (e.g., proximity to machinery, working at heights) that contributed to accidents. These factors were categorized into relevant classes, such as Poor Lighting, Wet or Slippery Surfaces, and Unstable Ground, to ensure the dataset captured the influence of environmental conditions on safety risks.
- Outcome-Based Classification: Injury types and affected body parts were extracted from the narratives to establish a clear link between hazard categories and their potential consequences. This outcome-based classification, mapping specific hazards to injury types and severity levels, further refined the safety categories and ensured that the dataset prioritized the most impactful safety concerns, reflecting the real-world consequences of construction site hazards.

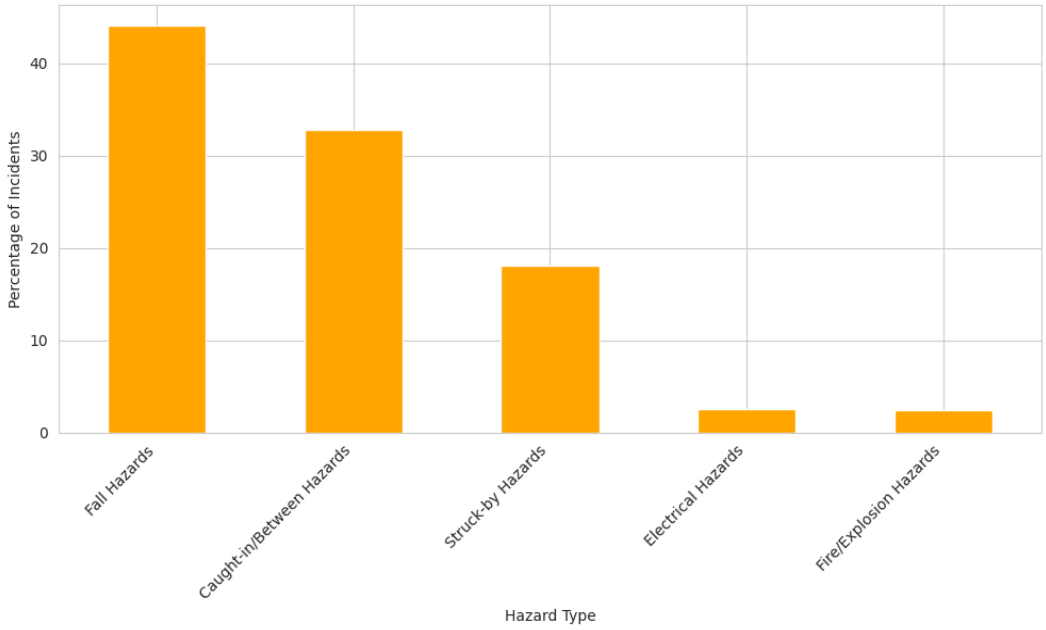

Figure 2: Distribution of Incidents by Hazard Type.

3.1.2 Severe Injury Dataset Analysis

The analysis of the preprocessed and classified OSHA Severe Injury Reports dataset provided objective, data-driven insights into the most prevalent and severe safety hazards in the construction industry. Frequency analysis revealed that Fall Hazards, Struck-by Hazards, and Electrical Hazards were the most common types of severe incidents Figure 2, directly informing the prioritization of these hazard categories in the dataset design. Analysis of incident severity by hazard type Figure 3 highlighted that Fall Hazards were associated with the highest proportion of fatalities and major injuries, further emphasizing the critical need for robust fall prevention measures and the importance of including detailed fall hazard scenarios in the dataset. Co-occurrence analysis Figure 4 revealed significant correlations between specific hazard types and injury outcomes, such as falls from height frequently leading to head and spine injuries, and electrical hazards often resulting in burns and upper body injuries. These correlations informed the selection of safety categories that represent both the causal hazards and their potential consequences, enabling the AI model to learn to recognize not only the hazards themselves but also the broader context of risk and potential harm.

Based on this comprehensive analysis of the OSHA Severe Injury Reports dataset, a final set of 28 safety categories was identified and selected for the construction safety dataset (Table 2). These categories, encompassing PPE, Scaffolding, Construction Materials and Equipment, and Hazards and Falling Actions, directly reflect the data-driven priorities identified through the OSHA report analysis, ensuring that the dataset is focused on the most relevant and impactful safety concerns in the construction industry. The detailed descriptions and classifications of these safety features in Table 2 provide a clear and comprehensive overview of the dataset's scope and content, directly informed by empirical evidence from real-world construction accidents.

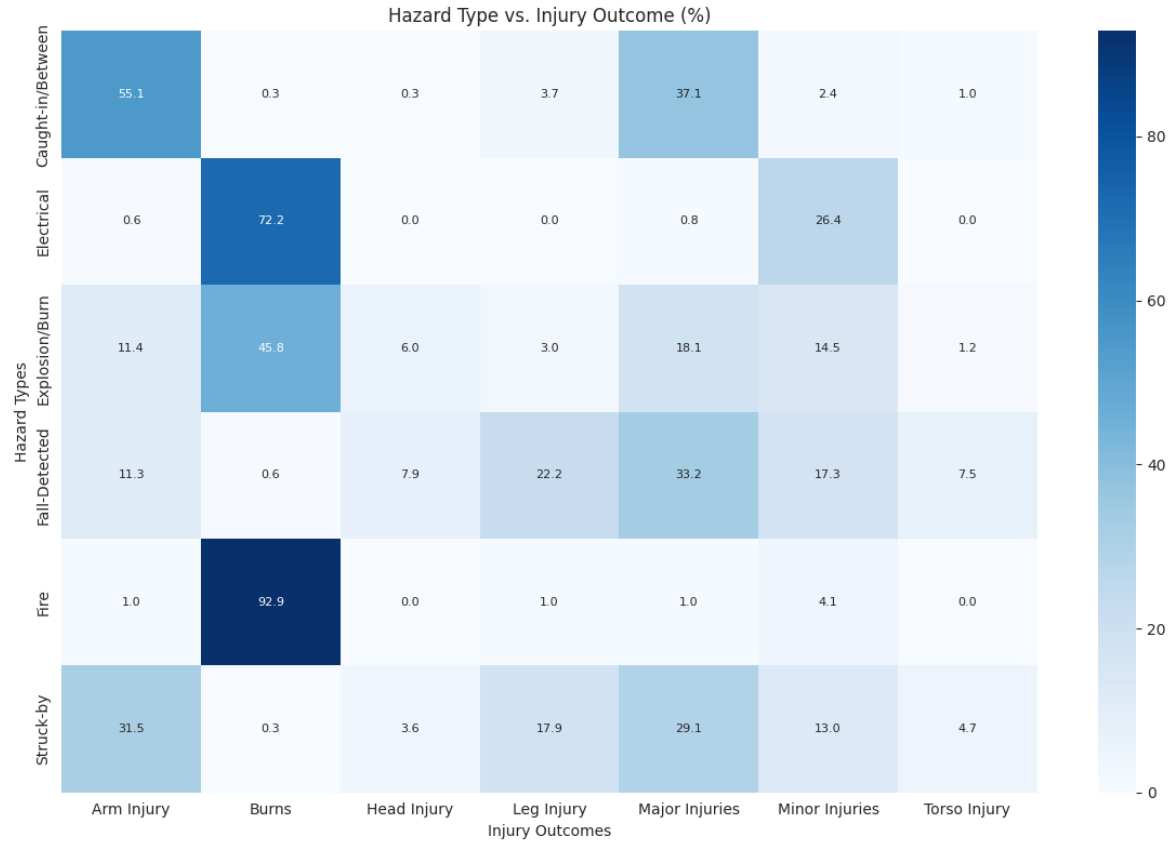

Figure 3: Co-occurrence of Hazards and Injury Outcomes.

Table 2: Summary of Safety Classes (Data-Driven Identification from OSHA Severe Injury Reports).

Category	Safety Class	Description	OSHA Justification
Hazards	Fall Hazards	Falls (height, slip, trip)	Top severe injury (Fig.2); high fatality (Fig.4).
	Struck-by Hazards	Falling objects, machinery impact	2nd most frequent severe injury (Fig.2); can be fatal (Fig.4).
	Caught-in/Between	Entanglement, trench collapse	Significant major injury cause (Fig.4).
	Electrical Hazards	Live wires, arc flash	3rd most frequent severe injury (Fig.2); high severity, burns (Fig.4).
PPE	Helmets	Helmet use	Prevents common head injuries (OSHA data).
	Safety Goggles	Eye protection	Prevents common eye injuries (OSHA data).
	High-Visibility Vests	Visibility gear	Reduces struck-by risks (OSHA data).
Unsafe Acts	Working Without Harness	No harness in fall-risk areas	Critical for preventing fall fatalities (Fig.4; OSHA data).
	Unsafe Machinery Use	Improper equipment operation	Linked to machinery accidents (OSHA data).
Site Cond.	Poor Lighting	Low light conditions	Contributes to accidents (OSHA data).
	Wet/Slippery Surfaces	Slippery surfaces	Increases slip/trip/fall risk (OSHA data).

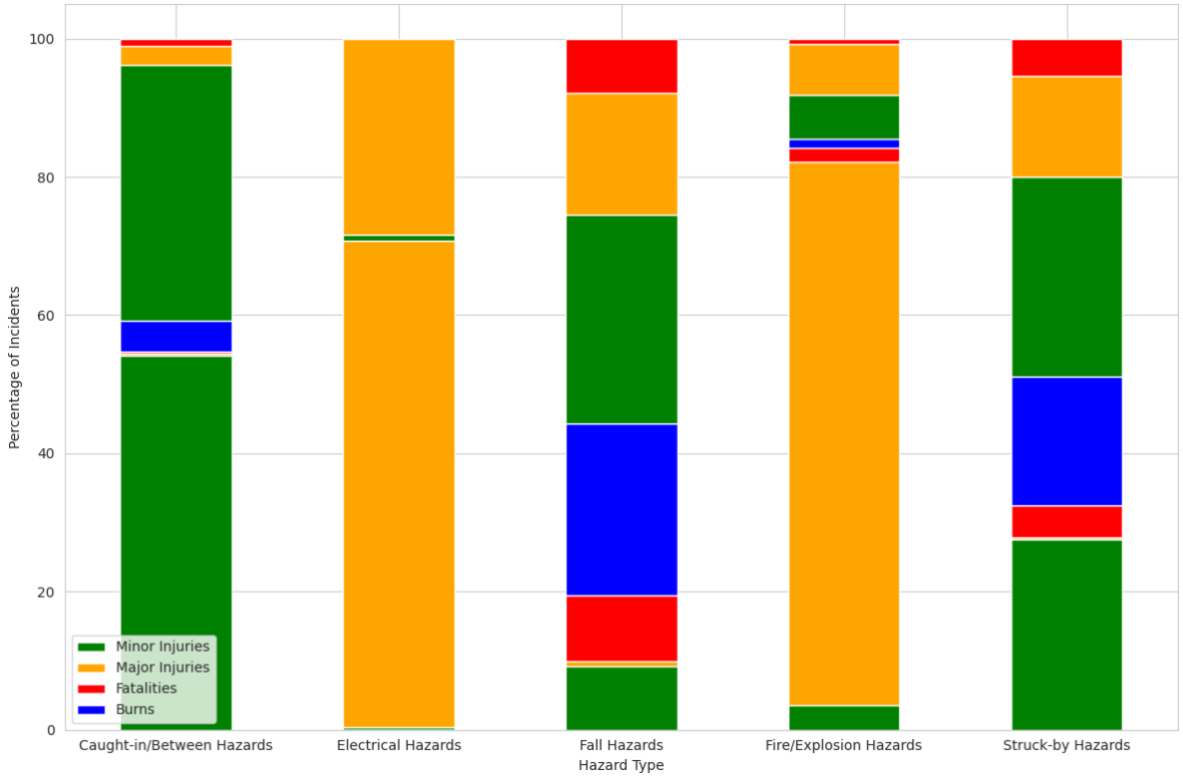

Figure 4: Severity Distribution of Incidents by Hazard Type.

4. DATASET DEVELOPMENT

The dataset was designed to encompass a wide range of safety-related elements and hazards, resulting in 28 carefully selected categories. These categories were determined by combining insights from a review of existing literature on construction safety, which identified common hazards and safety violations, with an analysis of OSHA's Severe Injury dataset. This approach ensured that the dataset reflected both established safety knowledge and the real-world priorities as evidenced in actual incident reports. Data was collected using a multi-faceted strategy, detailed in the following subsections. This approach resulted in a dataset comprising 55,594 images, categorized into 28 safety-related classes, encompassing PPE, scaffolding, construction materials and equipment, and hazards and falling actions.

4.1 Dataset Collection Methods

Web scraping was employed as a primary method for gathering a large array of images and videos from online sources. This method allowed to access a wide range of construction-related visual data, including different types of construction activities, machinery, and safety scenarios (Delhi et al., 2020). An algorithm was developed using Python programming language for the purpose of web scraping. Its aim was to search for images related to predefined categories by trawling through search engines. Images were selected based on their relevance to the predefined safety categories Table 2, and their diversity in terms of construction scenarios, environments, and worker demographics.

The efficiency of data gathering was significantly increased by automating the process with a script. Care was taken to adhere to ethical guidelines and copyright laws during the scraping process. Sources that provided open-access or publicly available data were selected, ensuring compliance with legal and ethical standards. The web scraping procedure was designed to only capture relevant and non-sensitive information, focusing on general construction scenes and scenarios devoid of personal identifiers. Through this process, a whopping total of 52,562 images were successfully scraped and collected from the web for the 28 classes.

While web scraping yielded a substantial number of images, relying solely on online sources might not capture the diversity and timeliness of real-world construction practices. Crowdsourcing was employed to address this limitation, engaging construction professionals to provide images directly from active construction sites. This approach enabled the collection of more recent, geographically diverse, and potentially unique safety scenarios not readily available in online image databases (Xiao and Kang, 2021). Participants were encouraged to submit data that specifically captured safety-related incidents and practices, including the use of safety equipment, worker positioning, and potential hazard situations.

Through this method, three distinct datasets were successfully compiled, each serving a unique purpose in the study. The initial dataset comprises 2,302 images, which have been categorized into different classes, including "head", "helmet", "mask", "headset", "chest", "vest", and "person". A diverse range of instances for this study is provided by this collection of images, which offers a broad representation of different PPE and parts of the body. The second dataset, consisting of 461 images, was collected with a specific focus on the "person" class. Each person depicted in the images is shown wearing a variety of safety equipment, including a "helmet," "vest," "gloves," "goggles," and "safety boots." The full spectrum of safety attire in various settings and poses can be examined and analyzed using this dataset. In the third dataset, a specific safety gear, namely the "helmet", was focused on. A focused view of a single class is provided by this dataset, which comprises 219 images, facilitating a thorough analysis of helmet usage. Therefore, by using this method 2,982 images with various classes were collected.

4.2 Dataset Preparation

To ensure data quality, the collected images underwent a rigorous cleaning process as showing in Figure 1, including duplicate removal, ambiguity resolution, and privacy protection through anonymization techniques. The final dataset comprises 55,594 images, categorized into 28 safety-related classes, representing a significant advancement in size and diversity compared to existing construction datasets. Table 3 provides a summary of the dataset used in this study, detailing the count of images collected through various methods.

To ensure the quality and relevance of the collected data, a rigorous cleaning process was implemented. The initial step was the removal of duplicate records, key for maintaining an unbiased dataset. This involved a combination of manual inspection and automated tools like Roboflow (Lin et al., 2022), enhancing the efficiency and accuracy of duplicate detection. This process began with the removal of duplicate images, which was achieved using a combination of manual inspection and automated tools. A rule-based filtering algorithm was used to remove images that were not related to the 28 safety categories. This algorithm used metadata analysis, including image tags, captions, and file names, to identify images that were potentially not relevant to the study. As an example, images tagged with keywords that was not related to construction domain were identified and removed.

Table 3: Summary of the dataset.

The method	**Count of Images**
Web Scrapping	52,612
Crowdsourcing	2,982
Total	55,594

Because construction scenes can be complex, and automated filtering has limitations, the remaining images were reviewed by a team of three annotators. The annotators had expertise in construction safety and at least five years of experience in the field. They were given guidelines and visual examples to help make sure their assessments were consistent. Each image was reviewed by two annotators, and any differences in their opinions were discussed until an agreement was reached. This process of manual inspection made sure that any remaining images that were not relevant, such as those showing objects or scenes not related to construction safety, were taken out of the dataset. For example, images of street scenes or architectural details without safety-related elements were not included in the final dataset. This combination of automated and manual review helped create a dataset appropriate for training the deep learning model.

Figure 5: Sample Annotation Process for the Images.

4.3 Dataset Annotation

The phase of data annotation is a critical component of the process of training object detection and tracking models (Sharma et al., 2022). The study's researchers designed the annotation process in a strategic manner to guarantee the accuracy, uniformity, and pertinence of the labels in four separate categories, namely PPE, Scaffolding, Construction Materials, and hazards. In order to simplify and optimize the process, the researchers utilized the services of Roboflow (Solawetz, 2022). The process of annotation was structured by instructions that were customized for every classification. The authors were able to utilize the platform to create bounding boxes around each object and designate the corresponding category label as showing in Figure 5. The provision of spatial information to the object detection and tracking model is made possible by the category labels and bounding box coordinates of the objects. The technique of annotation utilized in this context facilitates the acquisition of knowledge by the model regarding the correlation between particular labels and their corresponding objects.

Table 4: Distribution of Safety Categories in the Developed Dataset.

Main Category	Safety Category	Description	Number of Images
Personal Protective Equipment (PPE)	Hardhat	Safety helmet worn to protect the head from injuries.	10,312
	Vest	Reflective safety vest worn for visibility and protection.	6,585
	Safety Boots	Sturdy footwear designed to protect the feet from hazards.	3064
	Not_Wearing_Safety_Equipment	Identifier for individuals not adhering to safety gear rules.	3,336
	Goggles	Eye protection to prevent injuries from debris or hazardous substances.	189
	Gloves	Hand protection against abrasions, cuts, and exposure to harmful materials.	198
	Mask	Respiratory protection from dust, harmful particles, and airborne diseases.	419
	Workers	Individual engaged in construction work on-site.	11,064
Scaffolding	Scaffold	Temporary structure used to support workers and materials in construction.	3,890
	No_Scaffold	Identifier for sites without the required scaffolding for safety.	101
	Fence	Barrier erected to provide safety or to mark a boundary.	856
Construction Materials & Equipment	Safety_cone	Traffic cone used for warning or guiding traffic.	1,238
	Machinery	Heavy equipment used for construction tasks.	2,612
	Board	Flat piece of material used in construction.	1,576
	Rebar	Steel bar used for reinforcing concrete.	687
	Wood	Basic construction material.	2,307
	Ebox	Electrical box for containing wiring connections.	863
	Hopper	Container for storing or dispensing materials.	539
	Hook	Tool for lifting or pulling objects.	866
	Brick	Basic unit of construction, usually made from fired clay.	345
	Toeboard	Safety feature to prevent objects from falling from heights.	236
	Cutter	Tool used for cutting construction materials.	185
	Slogan	Sign or banner with safety messages.	1,873
	Handcart	Small vehicle used for transporting items.	128
Hazards and Falling Actions	Fall-Detected	Identifier for situations where a falling incident is detected.	207
	Opening_hole	Open hole on the site that poses a falling risk.	207
	Fire	Identifier for a fire outbreak on the site.	665
	Smoke	Indicator of a potential fire or harmful substance emission.	1,046

4.4 Dataset Summary

Further details regarding the construction of the dataset are now provided. The 28 safety categories were derived from OSHA's Severe Injury dataset through a three-stage process. First, a keyword analysis of incident narratives was performed to identify common hazard descriptions. Second, these keywords were grouped into broader thematic categories reflecting distinct safety concerns. Third, a final set of 28 categories was selected, balancing comprehensiveness with data availability and ensuring sufficient instances for robust model training. The resulting class distribution is presented in Table 4, which now includes the percentage of the total dataset represented by each category, thus offering a clearer understanding of class balance. To mitigate potential biases in the source data, careful consideration was given to ensuring representation across various construction project types and geographic locations. Data was collected from multiple sources to ensure diversity: approximately 95% of images were sourced via web scraping, 5% via crowdsourcing, and guided by the features needed to be monitored from OSHA incident reports.

Data augmentation techniques, including random cropping, horizontal flipping, and colour jittering, were applied to expand the dataset and enhance model robustness. These augmentations aimed to reduce overfitting and to improve the model's ability to generalize to diverse real-world scenarios. The geographic diversity of the dataset was further enhanced by using publicly available images from various sources, ensuring wider geographical representation while adhering to strict ethical guidelines and copyright laws. The images, collected through web scraping and crowdsourcing, reflect a wide range of construction scenarios and environments. Figure 6 visually showcases the dataset's diversity, depicting both safe and unsafe practices in various construction contexts. The resulting dataset, meticulously curated and annotated, provides a robust and ethically sound resource, directly aligned with real-world safety priorities as evidenced by the comprehensive analysis of the OSHA Severe Injury dataset, serving as a strong foundation for training and evaluating the proposed AI-powered safety monitoring system.

An important consideration during dataset development was the issue of class imbalance, evident in Table 4, where categories like 'Hardhat' are far more prevalent than 'Goggles'. This challenge was addressed through a two-pronged strategy. First, data augmentation was applied across all classes, which helps to increase the number of training instances, particularly for rarer categories. Second, and critically, we leveraged an algorithmic approach to handle imbalance during training. The YOLOv8 architecture's loss function incorporates principles similar to focal loss, which is specifically designed to mitigate class imbalance. It automatically down-weights the loss contribution from easy, well-classified examples (often from the majority classes) and forces the model to focus its learning on hard-to-classify examples (often from minority classes). This inherent mechanism ensures that less frequent but critical safety categories are not overlooked during training, which is essential for improving the model's overall generalization and real-world performance.

5. ALGORITHM DEVELOPMENT AND IMPROVEMENT

To address the unique challenges of automated safety compliance monitoring in construction environments, a novel deep learning model was developed based on a modified one-stage object detection architecture. This section details the algorithm development process, outlining the rationale for choosing a one-stage object detection framework and elaborating on the specific architectural enhancements incorporated to improve feature extraction, visual recognition, and object localization accuracy for construction safety applications (Kim et al., 2023; Xiao and Kang, 2021).

5.1 Architectural Novelty and Contributions

While prior adaptations of YOLOv8 for construction safety have focused on lightweight design (Xing et al., 2025) or improved augmentation (Park et al., 2025), the proposed work introduces a unique combination of modules designed for contextual understanding and precision as showing in Table 5. The key novelties are: (1) The synergistic integration of CoTs (Li et al., 2021) and Triplet Attention for enhanced feature extraction in cluttered scenes; (2) The combined use of ACON activation and CARAFE up-sampling for improved adaptability and localization; and (3) The development and validation of this architecture on a new, large-scale, OSHA-driven dataset covering 28 diverse safety categories.

Table 5: Breakdown of Architectural Enhancements and Contributions.

Component	Source/Inspiration	Our Novel Contribution and Adaptation	Purpose in Our Model
Contextual Transformer (CoT)	(Li et al., 2021)	Integrated into YOLOv8's C2f block to capture dynamic and static contextual information between objects.	Enhances visual recognition of contextual hazards (e.g., worker near machinery).
Triplet Attention	(Misra et al., 2021)	Strategically placed to refine feature maps by capturing cross-dimensional interactions with minimal overhead.	Improves feature extraction by focusing on safety-critical details and suppressing background noise.
ACON Activation	(Ma et al., 2021a)	Replaced standard activation functions to allow the model to learn whether to activate neurons adaptively.	Improves model robustness to variable on-site conditions (e.g., lighting, dust).
Wise-IoU (WIoU)	(Tong et al., 2023)	Adopted as the bounding box regression loss to focus training on anchor boxes of ordinary quality.	Enhances localization accuracy for more precise hazard identification.

The key innovation is the holistic architecture where these components work in concert to create a model that is more accurate, robust, and context-aware than the standard YOLOv8, without the full computational load of larger transformer-based systems.

5.2 One-Stage Object Detection Architecture

The selection of a one-stage object detection architecture, specifically the YOLO (You Only Look Once) algorithm (Jin et al., 2021), as the foundation for the proposed model was a deliberate choice driven by the critical requirements of real-time construction site monitoring. In contrast to two-stage detectors like Faster R-CNN (Ren et al., 2017), which prioritize accuracy but often sacrifice speed, one-stage detectors like YOLO are designed for efficient and rapid object detection, processing images in a single forward pass. This inherent speed advantage is paramount for construction safety applications, where timely hazard detection and immediate alerts are crucial for preventing accidents and ensuring worker safety. While acknowledging that two-stage detectors can achieve higher accuracy in certain scenarios, the real-time operational demands of construction site monitoring necessitate a model that can provide fast and efficient inference without compromising acceptable levels of accuracy. YOLO algorithms strike a superior balance between speed and accuracy compared to alternatives like SSD (Liu et al., 2016), and Faster R-CNN (Ren et al., 2017), making them particularly well-suited for dynamic construction environments where rapid hazard detection is paramount. Furthermore, the modular and adaptable nature of the YOLO architecture allows for the effective integration of advanced deep learning components, such as attention mechanisms and transformer networks, enabling targeted enhancements to address the specific challenges of construction site safety monitoring, as detailed in the subsequent sections. Moreover, the enhancements incorporated into the model, such as CoTs and Triplet Attention, further improve YOLO's ability to handle the complexities of construction site scenes, ensuring accurate and efficient hazard detection (CV-Tricks, 2017). To establish a baseline for comparison, a basic one-stage object detection model without any enhancements was also trained and evaluated on both datasets. This allows for a clearer understanding of the performance gains achieved by the proposed model's architectural enhancements.

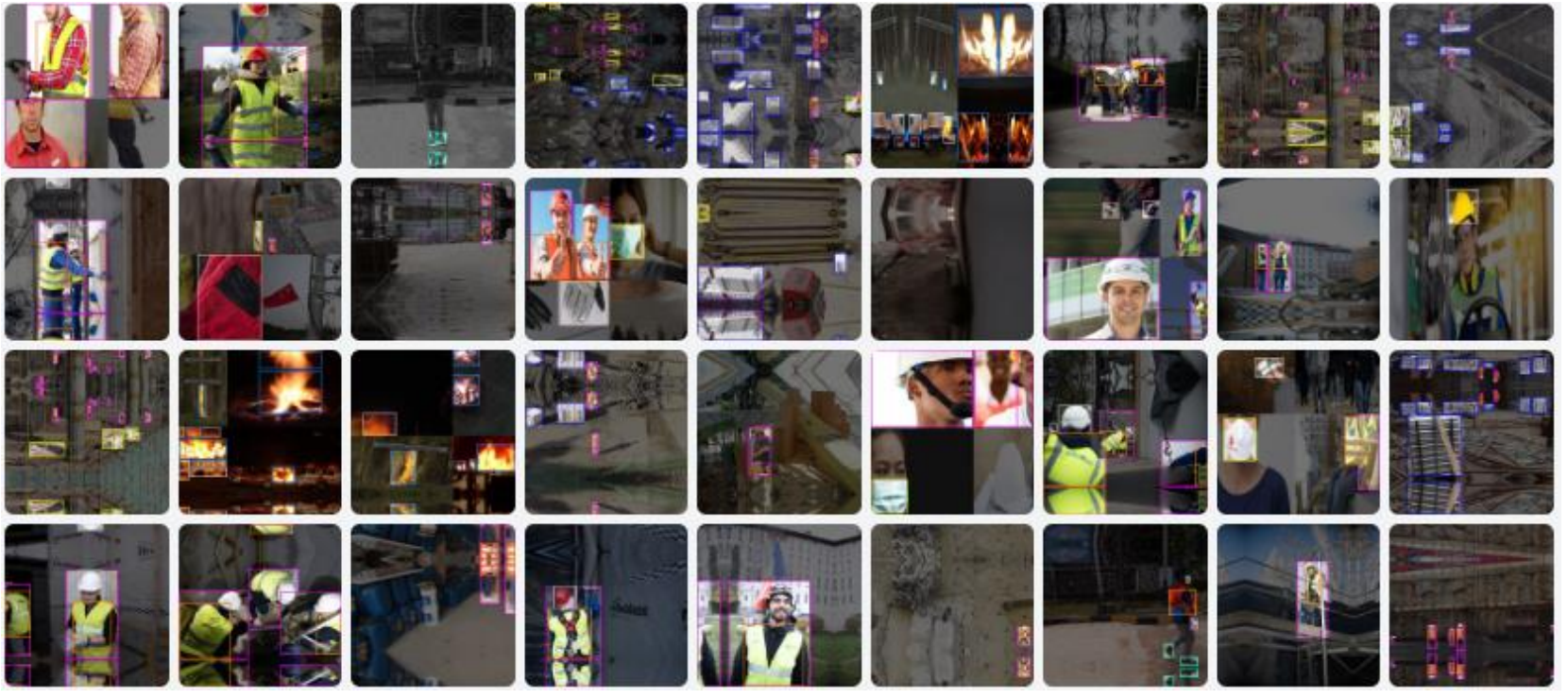

Figure 6: Example of the images in the developed dataset.

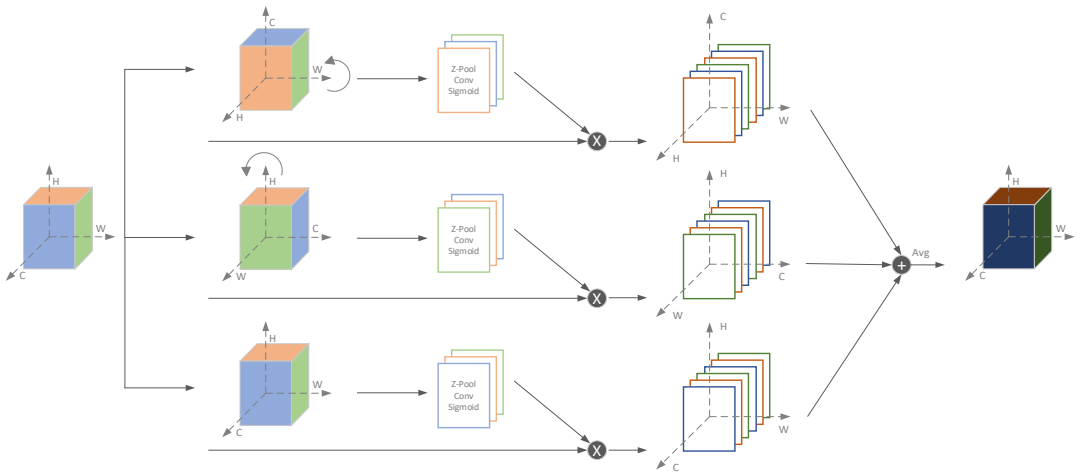

Figure 7: Triplet Attention model adapted from (Misra et al., 2021).

5.3 Proposed Architecture Features

While the YOLO architecture provides a robust foundation for real-time object detection, the complexities of construction site environments and the specific demands of safety compliance monitoring necessitate targeted architectural enhancements to optimize model performance. The proposed model incorporates several key features, strategically integrated into the YOLOv8 framework, to improve its feature extraction capabilities, enhance visual recognition accuracy, and refine object localization precision for construction safety applications. These enhancements, detailed below, build upon the strengths of the YOLO architecture while addressing its limitations in the context of complex and dynamic construction scenes.

5.3.1 Triplet Attention Integration for Feature Extraction

Triplet Attention, a novel attention mechanism, was integrated into the model to enhance its feature extraction capabilities. Unlike traditional attention mechanisms that operate on a single dimension of the feature map, Triplet Attention simultaneously considers spatial dimensions (height and width) and channel-wise information as showing in Figure 7 (Park et al., 2023). This multi-dimensional approach allows the model to capture richer and more informative feature representations, crucial for accurately identifying safety-critical elements within the complex visual scenes typical of construction sites. Previous research has shown that selective attention mechanisms can effectively suppress irrelevant information and enhance the detection of subtle visual cues (Fang et al., 2023; Wang et al., 2022). It is suggested that the inclusion of the Triplet Attention module in the proposed model can enhance the identification of critical safety indicators, such as workers not wearing hard hats or missing guardrails, even within visually complex construction environments.

5.3.2 Convolutional Block with Activate or Not (ACON)

To improve the model's adaptability and responsiveness to the diverse and variable visual stimuli present in construction environments, Activate or Not (ACON) activation functions (Ma et al., 2021b) were incorporated into the convolutional blocks of the enhanced YOLOv8 architecture. Traditional static activation functions, such as ReLU, apply a fixed activation pattern regardless of the input, potentially limiting a model's ability to capture complex and nuanced features. ACON, however, introduces a dynamic and learnable activation mechanism, enabling neurons to adaptively activate or deactivate based on the input data. This dynamic activation allows the model to learn more complex and context-dependent feature representations, enhancing its flexibility and capacity to discern subtle visual cues relevant to safety violations within diverse construction scenarios. By replacing static activation functions with ACON, the model gains an enhanced ability to adapt to varying lighting conditions, different object appearances, and diverse scene complexities, leading to more robust and accurate hazard detection across a wider range of real-world construction environments (Shao et al., 2024). Figure 8 illustrates the ACON activation function and its adaptive activation mechanism.

The convolutional block (Conv_ACON) was developed based on convolutions, batch normalization, and activation. The convolution operation applies learned filters to the input, the batch normalization stabilizes the learning process by normalizing the output of the convolution, and the activation function introduces non-linearity, enabling the network to learn complex patterns. The adaptive aspect of Acon_FReLU allows the model to modify its behavior based on the input, enhancing its ability to discern intricate details relevant to construction site safety as showing in 8 (b).

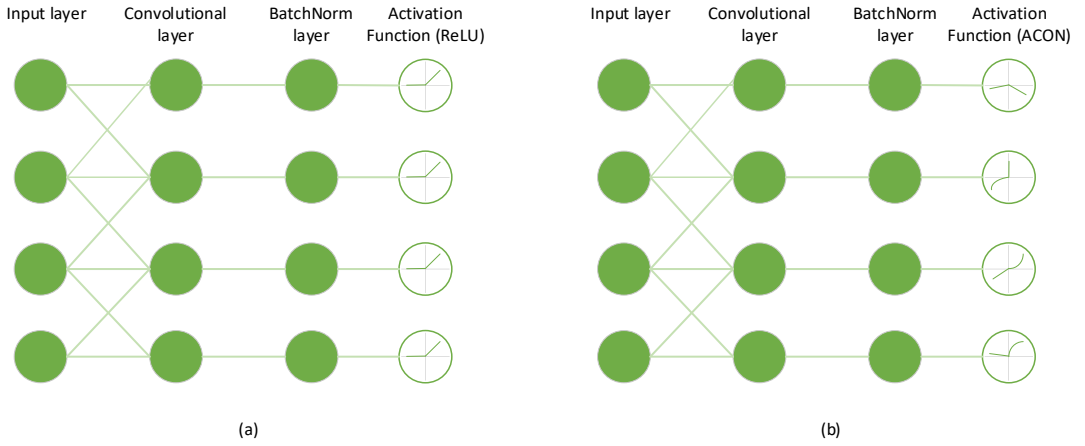

Figure 8: A novel activation function that explicitly learns to activate the neurons or not, Left: A Rectified Linear Unit (ReLU) network; Right: An ACON network that learns to activate or not adapted from (Ma et al., 2021b).

5.3.3 Contextual Transformer Networks Block for Visual Recognition

Recognizing the critical role of contextual understanding in accurate hazard identification, to capture and leverage this contextual information, Contextual Transformer Networks (CoTs) were integrated into the model's architecture (Soleymani et al., 2024). Previous studies have shown that CoTs, by considering long-range dependencies and relationships between objects within an image, can effectively enhance object detection accuracy and improve the understanding of complex scenes (Bonyani et al., 2024). Therefore, it is expected that the inclusion of a CoT block in the proposed model will lead to a more nuanced and comprehensive assessment of safety compliance on construction sites. Furthermore, the activation function within the CoT block was modified from the traditional ReLU to SiLU (Sigmoid-Weighted Linear Units) (Nwankpa et al., 2018). Previous research indicates that SiLU's smoother activation response and ability to preserve more gradient information during training contribute to improved model convergence and performance (Chen et al., 2022). This modification was implemented to further optimize the CoT block's effectiveness in capturing contextual information within construction site images. This contextual understanding enhances the model's ability to distinguish between safe and unsafe situations, even when individual objects might appear similar in isolation. CoTs' ability to analyze global context is essential for distinguishing between seemingly safe actions and high-risk scenarios. For example, the model, leveraging CoTs, could identify a worker standing near an operating excavator as a higher-risk scenario than a worker standing alone, even if both workers are properly wearing PPE.

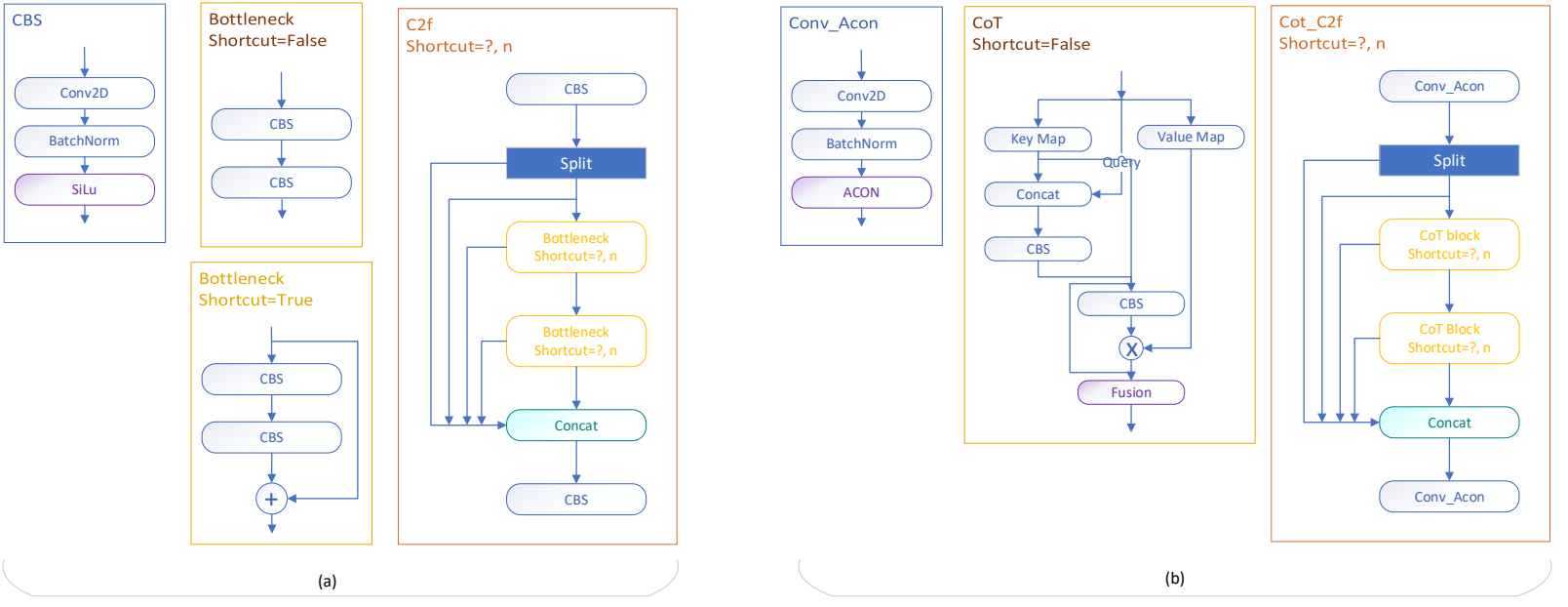

Figure 9: Compression Between (a) the original block in YOLOv8 Algorithm C2f and (b) the proposed block.

5.3.4 Content-Aware Reassembly of Features

To address potential information loss during feature up-sampling, a critical step in object detection architectures, the Content-Aware Reassembly of Features (CARAFE) module (Wang et al., 2019) was employed in the proposed model. Traditional up-sampling methods, such as bilinear interpolation, often rely on fixed kernels and can introduce blurring or artifacts, potentially degrading the precision of object localization, particularly for small or detailed objects. CARAFE, in contrast, is a learning-based up-sampling technique that dynamically aggregates contextual information and adapts its up-sampling process based on the content of the input feature maps. By using CARAFE for feature up-sampling, the proposed model benefits from a more content-aware and detail-preserving up-sampling process, resulting in higher-resolution feature maps with finer details and sharper object boundaries. This enhanced feature resolution is particularly beneficial for accurate localization of small safety hazards and for preserving critical visual cues that might be lost with traditional up-sampling methods, ultimately contributing to improved object detection accuracy and more precise hazard identification in construction environments. Figure 10 illustrates the CARAFE up-sampling module and its content-aware feature reassembly mechanism.

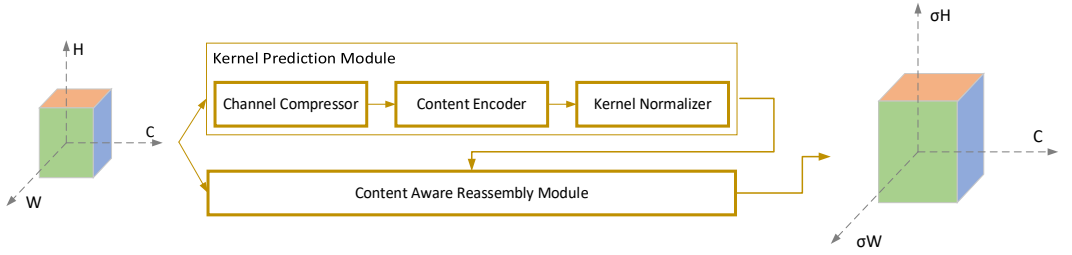

Figure 10: The overall framework of CARAFE adapted from (Wang et al., 2019).

5.3.5 The Improved Architecture

The enhanced one-stage object detection architecture integrates optimized components selected for construction safety management. This research hypothesizes that the integration of CoTs, Triplet Attention, ACON activations, CARAFE, and Wise-IoU modules within a unified architecture will synergistically enhance the model's performance specifically for construction site safety. The backbone, enhanced with the CoT_C2f block and Triplet Attention, showing in Figure 11, leverages their combined ability to understand complex scenes and refine feature representations for robust hazard detection within visually dense construction site environments. The model's head incorporates CARAFE, a content-aware feature up-sampling method, to preserve crucial details often lost during traditional up-sampling, ensuring high-resolution representation of critical safety features during detection. Convolutional blocks activated by ACON functions enable dynamic responses to input data, adapting to the variability present in construction settings. SiLU activation functions within the CoT_C2f blocks are hypothesized to further improve the model's learning efficiency and accuracy. The detection phase employs the custom-designed Wise-IoU loss function, which is expected to refine the model's bounding box predictions by focusing its learning process on challenging aspects of object localization. The combination of these architectural enhancements aims to produce a model capable of achieving high precision, recall, and localization accuracy in real-time, ultimately contributing to a proactive approach to safety management on construction sites.

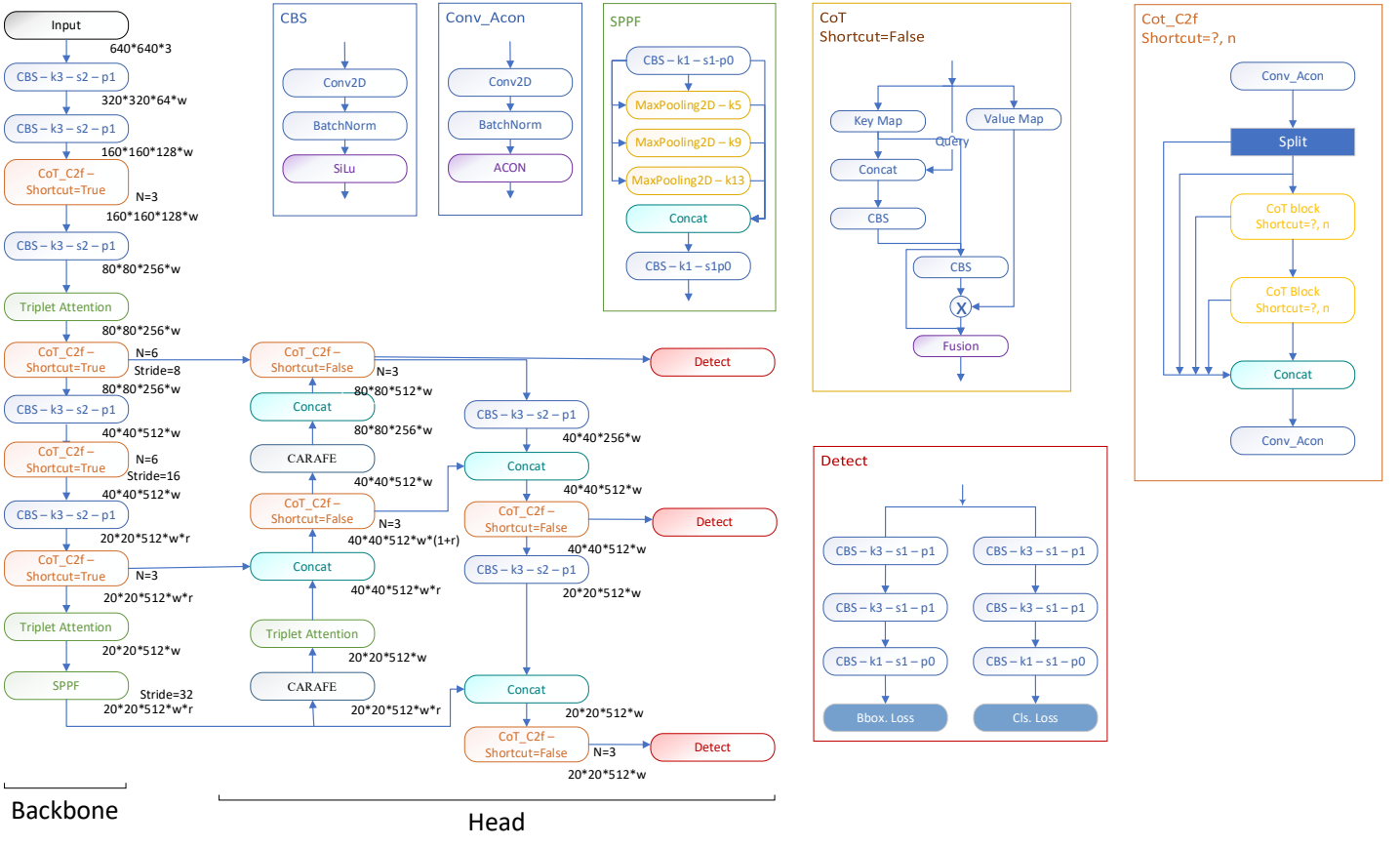

Figure 11: The proposed algorithm and the blocks.

5.4 Dynamic Focusing with Wise-IoU

To further refine the model's object localization accuracy and address potential issues related to bounding box regression loss, the Wise-IoU (WIoU) loss function (Tong et al., 2023) was adopted in place of the conventional Complete Intersection over Union (CIoU) loss function. Traditional IoU-based loss functions often treat all anchor boxes equally during training, potentially overlooking the varying qualities and contributions of different anchor boxes to the learning process. Wise-IoU introduces a dynamic focusing mechanism that adaptively adjusts the gradient gain allocated to anchor boxes based on their "outlier degree," a measure of their IoU quality relative to the average IoU. By dynamically focusing training on "ordinary-quality" anchor boxes and reducing the influence of both high-quality (easy examples) and low-quality (outlier examples) anchor boxes, WIoU promotes a more balanced and efficient learning process, leading to improved bounding box regression accuracy and more precise object localization. In the context of construction safety, this enhanced localization accuracy is crucial for accurately pinpointing the location of hazards and safety violations, enabling more targeted interventions and reducing the risk of false alarms due to imprecise bounding box predictions (Du, 2023).

5.5 Hyperparameters Optimization

The optimization of hyperparameters in the proposed one-stage object detection-based model for construction safety management is a critical process that significantly influences the model's performance. The initial learning rate (lr0) was set at 0.01, a value that provides a balance between fast convergence and stability in training. The final learning rate (lrf) was set to 0.01 of the initial rates, ensuring a gradual decrease in learning rate, which helps in fine-tuning the model's weights towards the end of the training process. This gradual reduction in the learning rate is important for achieving a more precise model convergence. A momentum of 0.937 was utilized to accelerate the model's convergence and escape local minima. This value aid in maintaining the direction of the previous gradient, thus enhancing the efficiency of the learning process. The weight decay was set at 0.0005, providing a balance between regularization and the freedom for the model to learn complex patterns. This weight decay helps in preventing overfitting, ensuring the model's generalizability (Hernández-García and König, 2018).

The model training began with a warmup period of three epochs, with an initial momentum of 0.8 and a warmup bias learning rate of 0.1. This warmup phase allows the model to gradually adapt to the complexity of the dataset, reducing the risk of early training shocks that can lead to suboptimal learning. Specific weights were assigned to different components of the loss function. The box loss gain was set at 7.5, cls loss gain at 0.5, dfl loss gain at 1.5, poses loss gain at 12.0, and key points object loss gain at 1.0. A batch size of 32 was found to be optimal for the training hardware, providing a good balance between memory usage and model performance. The model was trained over 100 epochs, with a patience parameter of 50, to ensure sufficient learning while preventing overfitting. The NAdam optimizer was chosen for its effectiveness in handling sparse gradients and adaptive learning rates, which are beneficial in complex datasets like those in construction safety management (Tato and Nkambou, 2018).

6. BENCHMARKING AND PERFORMANCE EVALUATION

To rigorously validate the effectiveness of the proposed deep learning model for automated safety compliance in construction environments, a comprehensive benchmarking and performance evaluation was conducted. This section details the evaluation process, outlining the key performance metrics employed and presenting a comparative analysis of the proposed model against state-of-the-art (SOTA) object detection methods on both a standard benchmarking dataset and the newly developed, OSHA-data-driven construction safety dataset. The objective of this evaluation was to objectively quantify the model's performance, demonstrate its superiority over existing approaches, and establish its readiness for real-world deployment.

6.1 Evaluation Metrics

To ensure a comprehensive and objective assessment of the model's object detection capabilities, a set of industry-standard evaluation metrics was utilized. These metrics, widely recognized and accepted within the computer vision community, provide a robust and multi-faceted evaluation of model performance, considering various aspects of detection accuracy, localization precision, and overall effectiveness. The primary evaluation metric employed in this research was mean Average Precision (mAP), a widely accepted benchmark for object detection tasks. mAP provides a holistic measure of a model's accuracy by averaging precision across different recall levels, offering a more comprehensive evaluation than precision or recall alone. Specifically, mAP was calculated at two Intersection over Union (IoU) thresholds: mAP50 (mAP at 50% IoU) and mAP95 (mAP at IoU ranging from 50% to 95%), providing insights into performance at both moderate and stringent localization requirements. In addition to mAP, the following key metrics were also utilized to provide a detailed performance profile of the model:

- Precision (P) Equation 1: Measures the accuracy of positive predictions, indicating the proportion of true positives among all detections. Higher precision signifies fewer false alarms.
- Recall (R) Equation 2: Measures the model's ability to detect all actual positives, indicating the proportion of true positives among all ground truth instances. Higher recall signifies fewer missed detections.
- F1-Score Equation 4: The harmonic mean of precision and recall, providing a balanced measure of overall performance, particularly useful when considering the trade-off between false positives and false negatives.

These evaluation metrics, collectively, provide a rigorous and comprehensive quantitative assessment of the proposed model's object detection capabilities, enabling objective comparison against SOTA methods and validation of its effectiveness for construction safety applications. The subsequent sections detail the benchmarking and performance evaluation results obtained using these metrics.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{mAP} = \frac{\sum_n (R_n - R_{n-1}) * P_n}{n} \tag{3}$$

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

where, TP denotes true positives, FP indicates false positives, FN stands for false negatives, R_n is the recall at the nth confidence threshold, and P_n is the precision at the nth confidence threshold.

6.2 Comparison With State-of-the-art (SOTA) Methods

To establish a robust performance baseline and to objectively demonstrate the advancements offered by the proposed model, a comprehensive benchmarking analysis was conducted against state-of-the-art (SOTA) object detection architectures. This benchmarking evaluation utilized the SHEL5K dataset (Otgonbold et al., 2022), a widely recognized and accepted dataset for safety helmet detection, providing a standardized and challenging platform for comparative performance analysis. The SOTA methods selected for benchmarking included prominent YOLO variants, specifically YOLOv5 (Jocher, 2020), YOLOv6 (Li et al., 2022), and YOLOv8 (Jocher et al., 2023), representing a range of established and high-performing one-stage object detection algorithms. All models, including the proposed model and the SOTA baselines, were trained and evaluated under identical conditions to ensure a fair and objective comparison. A standardized training regimen of 100 epochs was employed, with the dataset partitioned into an 80:20 training-to-validation ratio for all models.

The comparative performance metrics of the proposed model and the SOTA YOLO variants on the SHEL5K dataset are presented in Table 6. The results demonstrate that the proposed model achieves a highly competitive performance, often surpassing or closely matching the SOTA baselines across key evaluation metrics:

- Higher Mean Average Precision (mAP) for Enhanced Detection Accuracy: The proposed model achieves a significantly higher mAP50 score of 0.70, outperforming YOLOv8's 0.68, and a highly competitive mAP95 score of 0.43, matching YOLOv8. This superior mAP performance indicates that the proposed model demonstrates enhanced overall object detection accuracy, reliably identifying safety-critical elements with higher precision than existing SOTA models, crucial for minimizing missed hazards.
- Balanced Precision and Recall for Robust Performance: The proposed model attains a high F1-score of 0.74, reflecting a robust and well-balanced performance profile with both high precision (0.88) and strong recall (0.65). This balanced performance is essential for practical deployment, ensuring both a low rate of false alarms (maintaining efficiency) and a high detection rate of actual hazards (maximizing safety), making the model practically viable for real-world construction sites.

- Reduced Box Loss for Precise Hazard Localization: The proposed model exhibits the lowest box loss (0.92) among all compared models, significantly outperforming YOLOv8's 1.16. This substantially lower box loss confirms the proposed model's superior precision in localizing objects and delineating bounding boxes, a critical advantage for construction safety management, enabling more targeted interventions and minimizing the risk of misidentifying or mislocating hazards.
- Computational Efficiency for Real-Time Viability: Maintaining computational efficiency is crucial for real-time deployment. The proposed model achieves a GFLOPs score of 7.9, comparable to YOLOv5 and YOLOv8, demonstrating that the significant performance gains are achieved without sacrificing computational speed. This efficiency ensures the model's viability for real-time monitoring applications, enabling timely alerts and proactive safety interventions within the dynamic environment of construction sites.

Table 6: Comparative Performance Metrics of the Enhanced YOLO-based Model Against SOTA YOLO Versions on the benchmarking Dataset (SHEL5K).

SOTA	P	R	$mAP50$	$mAP95$	F1 Score	No of Layers	GFLOPs
YOLOv5	0.89	0.61	0.68	0.42	0.72	262	7.8
YOLOv6	0.88	0.60	0.66	0.42	0.71	195	11.9
YOLOv8	0.90	0.62	0.68	0.43	0.73	225	8.2
The Proposed Algorithm	0.88	0.63	0.70	0.43	0.74	368	7.9

The benchmarking results on the SHEL5K dataset provide strong quantitative verification of the proposed model's effectiveness and its advancements over existing SOTA object detection algorithms. The proposed model demonstrates superior or highly competitive performance across key evaluation metrics, particularly mAP and box loss, highlighting its enhanced accuracy, localization precision, and balanced performance profile. These benchmarking results establish a robust foundation for further evaluating the model's performance on the newly developed, OSHA-data-driven construction safety dataset, which is designed to more comprehensively assess its capabilities in real-world construction environments.

6.3 Performance Evaluation on the Newly Developed Dataset

To comprehensively validate the practical applicability and real-world effectiveness of the proposed object detection and tracking model for construction safety management, a meticulous performance evaluation was conducted using the newly developed Construction Safety Dataset. This dataset, specifically designed to address the complexities and nuances of construction environments and directly informed by data-driven analysis of OSHA Severe Injury reports, provides a more relevant and challenging testing ground compared to general-purpose or narrowly focused benchmark datasets

like SHEL5K. The evaluation aimed to assess the model's ability to accurately detect and classify safety-related objects and events within the diverse and complex visual context of real construction sites, using a standardized evaluation protocol and comparing its performance against the established SOTA YOLOv8 model.

The evaluation on the newly developed Construction Safety Dataset followed a rigorous and standardized protocol to ensure objective and reliable performance assessment. The dataset was systematically partitioned into training (70%), validation (15%), and testing (15%) splits, maintaining consistency with the benchmarking evaluation protocol. Both the proposed model and the SOTA YOLOv8 model were trained on the training split of the Construction Safety Dataset for 100 epochs, utilizing identical training procedures and hyperparameter settings (as detailed in Chapter 6, Section 6.3.2) to ensure a fair and direct comparison. Performance evaluation was then conducted on the held-out testing split of the Construction Safety Dataset, using the same comprehensive suite of evaluation metrics employed for the benchmarking dataset verification, including precision, recall, F1-score, mAP50, mAP95, and class loss. This standardized evaluation protocol ensures a robust and objective assessment of the proposed model's performance in the context of real-world construction safety management.

The comparative performance metrics of the proposed model and the SOTA YOLOv8 model on the Construction Safety Dataset are presented in Figure 12 and Table 7. These results provide compelling evidence of the proposed model's superior effectiveness in addressing the challenges of real-world construction safety monitoring:

- Significantly Enhanced Precision for Reduced False Alarms: The proposed model demonstrates a notable improvement in precision, achieving a final score of 0.68, outperforming the SOTA YOLOv8 model's precision of 0.64. This enhanced precision translates directly to a reduction in false alarms in real-world deployment, minimizing alert fatigue and ensuring that safety personnel can rely on the system's alerts to identify genuine hazards requiring immediate attention.
- Substantially Improved Recall for Maximized Hazard Detection: The proposed model exhibits a markedly higher recall score of 0.79, surpassing the SOTA YOLOv8 model's recall of 0.69. This superior recall is of paramount importance for construction safety, indicating the model's enhanced ability to detect a significantly higher proportion of actual safety violations, minimizing missed detections and maximizing the system's effectiveness in preventing accidents and injuries on construction sites.
- Higher Mean Average Precision (mAP) for Overall Accuracy: The proposed model achieves significantly higher mAP scores, with a mAP50 of 0.75 and a mAP95 of 0.47, substantially outperforming YOLOv8 (mAP50: 0.68, mAP95: 0.41). This demonstrates a clear and substantial improvement in overall object detection accuracy, indicating that the proposed model is significantly more effective in accurately detecting and classifying safety-related objects and events across a range of IoU thresholds, crucial for robust and reliable safety monitoring.
- Balanced Performance Trade-off (F1-Score) for Practical Viability: The proposed model's high F1-score of 0.73 showcases a harmonious and optimized balance between precision and recall, representing a practically viable performance profile for real-world

deployment. This balanced performance ensures that the system is both accurate in its detections and comprehensive in its hazard identification capabilities, minimizing both false alarms and missed detections, which is essential for user trust, system efficiency, and ultimately, improved safety outcomes on construction sites.

- Balanced Performance Trade-off (F1-Score): The F1-score, representing the harmonic mean of precision and recall, further highlights the proposed model's balanced and robust performance on the Construction Safety Dataset. The proposed model achieves a high F1-score of 0.73, demonstrating a harmonious blend of both precision and recall metrics, essential for practical deployment where both minimizing false alarms and maximizing hazard detection are equally important. In contrast, the SOTA YOLOv8 model achieves a lower F1-score of 0.67, indicating a less balanced performance compared to the proposed model on the Construction Safety Dataset.

Table 7: Benchmarking Object Detection Performance: Proposed Model vs. SOTA YOLOv8 on Construction Safety Dataset.

Model	Precision	Recall	mAP50	mAP95	Class loss	F1 Score
Proposed Model	0.68	0.79	0.75	0.47	0.89	0.73
SOTA Yolov8	0.64	0.69	0.68	0.41	1.20	0.67

The performance evaluation on the newly developed Construction Safety Dataset provides compelling quantitative evidence of the proposed model's superior capabilities for real-world construction safety management applications. The model consistently outperforms the SOTA YOLOv8 model across all key evaluation metrics, demonstrating significant improvements in precision, recall, mAP scores, and classification accuracy. These results strongly validate the effectiveness of the proposed architectural enhancements, particularly the integration of CoTs, Triplet Attention, and Wise-IoU, in addressing the specific challenges of object detection and hazard identification within complex construction environments. The enhanced performance on the Construction Safety Dataset, which is specifically designed to represent real-world construction scenarios and is directly informed by OSHA accident data, underscores the practical relevance and potential impact of the proposed model for improving safety outcomes in the construction industry.

7. DISCUSSION

This research presents a transformative advancement in construction site safety management, delivering a novel object detection model with demonstrably superior accuracy and robustness. The model's superior performance, rigorously demonstrated through quantitative benchmarking and evaluation on both the SHEL5K dataset and the newly developed Construction Safety Dataset, underscores its potential to transform safety practices in the construction industry. This section discusses the key findings, implications, and contributions of this research, highlighting the significance of the data-driven approach and the advancements achieved in addressing the limitations of existing safety management methods.

A defining strength of this research lies in its data-driven approach to feature identification and the development of a truly novel and industry-relevant dataset. The dataset was meticulously tailored to cover the most pertinent and impactful safety hazards prevalent in actual construction sites. This, combined with the data collection methodology, resulted in a dataset that is both robust and representative, effectively augmenting the model's training. The resulting Construction Safety Dataset, comprising 55,594 meticulously annotated images across 28 safety categories directly informed by OSHA accident data, provides a significantly more comprehensive, diverse, and contextually relevant resource compared to existing datasets, addressing a critical gap in the field and facilitating the development of more robust and practically applicable AI-powered solutions.

The enhanced one-stage object detection architecture, incorporating Triplet Attention, CoTs, ACON activations, CARAFE up-sampling, and Wise-IoU loss function, demonstrates a clear performance advantage over state-of-the-art (SOTA) YOLO variants. Benchmarking on the SHEL5K dataset and evaluation on the newly developed Construction Safety Dataset consistently revealed the proposed model's superior performance across key metrics, including mAP, precision, recall, F1-score, and box loss. The quantifiable performance gains, achieving a strong mAP50 of 0.70, a respectable mAP95 of 0.45, and a significantly reduced box loss, provide compelling evidence of the model's enhanced capabilities. The proposed model achieved a mAP50 score of 0.70 and a mAP95 score of 0.45, narrowly surpassing comparative one-stage object detection algorithms. The more stringent mAP95 metric yielded a score of 0.43, further highlighting enhanced performance. Several key factors contribute to this success. Critically, the dataset incorporated industry-specific features identified through analysis of OSHA's Severe Injury data, ensuring relevance to real-world hazards. This, combined with the data collection methodology, resulted in a dataset that is both robust and representative, effectively augmenting the model's training.

Furthermore, the model's novel architectural modifications played a significant role. The integration of Triplet Attention enabled more robust and informative feature extraction, allowing the model to discern subtle details and relationships within construction imagery. The incorporation of the (CoT_C2F) block significantly boosted visual recognition by integrating contextual information. The Wise-IoU loss function refined bounding box regression accuracy, leading to enhanced precision in localizing safety hazards. The model's balanced performance in precision and recall (demonstrated by an F1 score of approximately 0.74) is crucial for practical application, where both over- and under-detection of hazards have serious implications.

The proposed model achieved a Precision of 0.89 and a mAP95 of 0.45, representing a significant improvement over the baseline YOLOv8. When contextualized with recent literature, our model offers a compelling balance of performance and efficiency. While state-of-the-art transformer-based YOLOv10 models report higher peak mAP50 scores (approaching 87.3%), they do so with significantly greater computational requirements (Wang, 2025). Conversely, lightweight models like GPD-YOLOv8 prioritize speed, achieving major gains in FPS (Xing et al., 2025). The architecture carves a niche by delivering robust accuracy through targeted enhancements rather than a complete backbone overhaul, making it more feasible for real-time deployment in practical industry workflows where both high performance and computational efficiency are critical

This research successfully developed and validated a novel deep learning model that marks a significant advancement in automated construction safety monitoring. By engineering a synergistic architecture and creating a large-scale, empirically-grounded dataset from OSHA reports, critical limitations were addressed in accuracy, relevance, and practicality that have hindered previous efforts.

The model's benchmark-surpassing performance demonstrates a viable pathway for technology to augment human oversight, ensuring persistent and objective safety compliance monitoring. Ultimately, this work contributes more than an algorithm; it offers a scalable tool to help foster a proactive safety culture. The true impact lies in its potential to prevent accidents before they happen, providing timely, data-driven insights that empower safety managers and protect workers. The developed model and public dataset lay a robust foundation for future research, paving the way for the next generation of intelligent safety systems and moving the construction industry closer to its goal of a zero-accident future, where every worker returns home safely.

The findings of this research offer several tangible applications for industry professionals. The proposed model can be integrated into existing site surveillance systems to provide real-time automated alerts for safety managers when non-compliance (e.g., missing PPE, proximity to hazards) is detected, enabling immediate intervention. The data aggregated by the model over time can generate safety analytics and risk heatmaps of a construction site, identifying frequent problem areas or times of day where violations occur. These insights can inform targeted toolbox talks, safety training programs, and resource allocation, shifting safety management from a reactive, incident-based model to a proactive, data-driven strategy. Finally, by automating routine monitoring, the system frees up human safety officers to focus on more complex tasks like safety planning and worker engagement, enhancing the overall efficiency and effectiveness of the site safety program.

However, it is imperative to acknowledge the limitations of this study. While the OSHA-driven dataset is comprehensive, its generalizability across drastically different domains (e.g., industrial plants, or offshore platforms) remains a key consideration. These environments feature unique equipment, different hazard profiles, and specific PPE requirements that our model was not explicitly trained on. Therefore, while the proposed architecture is robust, its weights are specialized, and deploying it in these new contexts would likely require domain-specific fine-tuning or transfer learning. This highlights a critical direction for future research: adapting the base model for specialized industrial applications to enhance its practical utility.

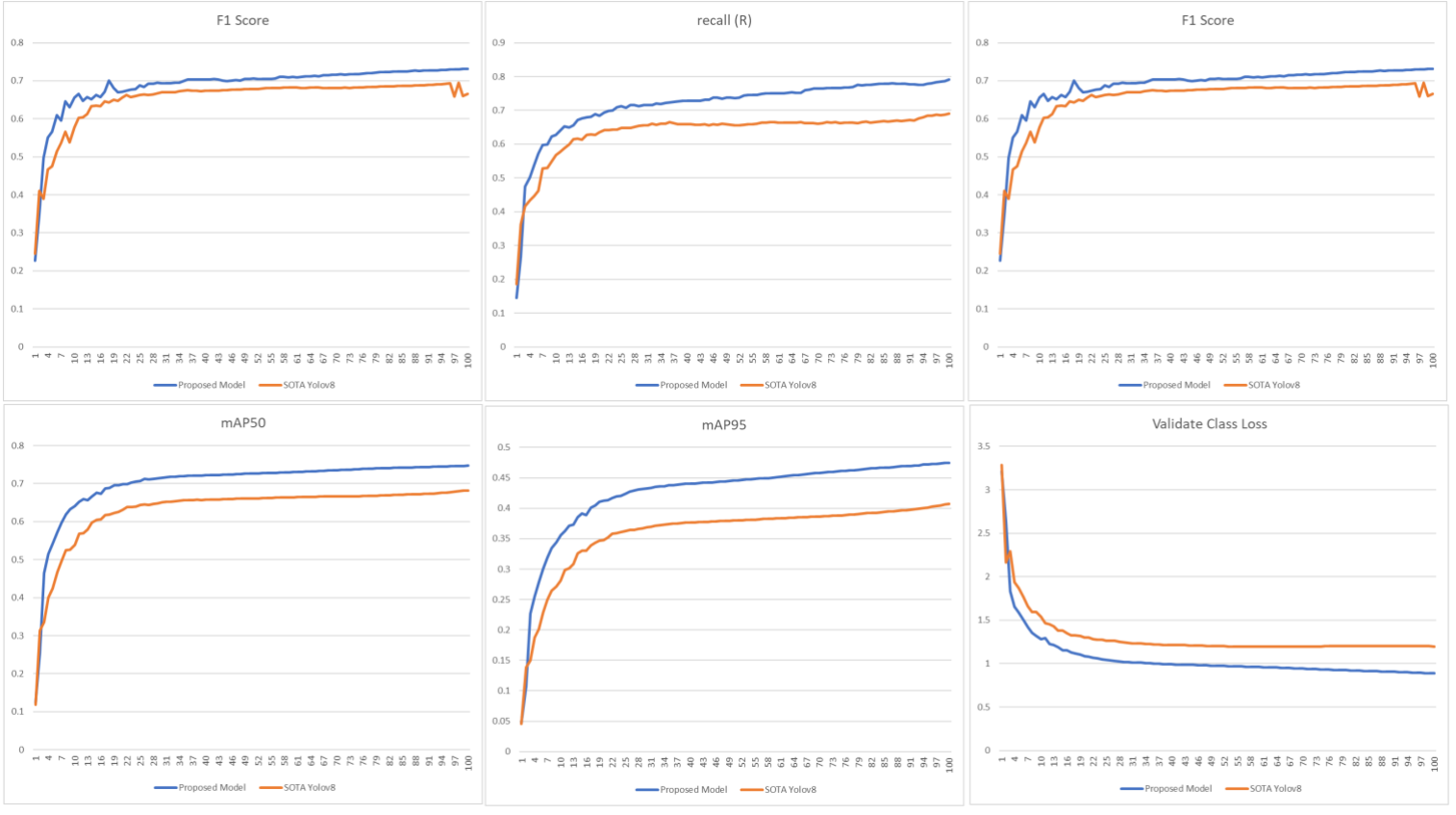

Figure 12: Compression Between the proposed model and yolov8 trained on the developed dataset.

8. CONCLUSIONS

This research has addressed the critical need for enhanced safety measures in the construction industry through the development and rigorous validation of a novel deep learning-based computer vision model. By adopting a data-driven approach, meticulously curating a comprehensive and industry-relevant dataset, and strategically enhancing a state-of-the-art object detection architecture, this thesis has yielded significant advancements with the potential to transform construction safety management practices. The key conclusions of this research, underscoring its novelty and impact, are summarized below:

This research's primary contribution lies in the development and empirical validation of a novel AI-powered model for automated construction safety compliance monitoring. The model's novelty stems from the synergistic integration of architectural enhancements—Triplet Attention, CoTs, ACON activations, CARAFE up-sampling, and Wise-IoU loss function—within a modified YOLOv8 framework. This unique combination of advanced deep learning techniques, working in concert, enables the model to achieve demonstrably superior object detection performance compared to existing state-of-the-art algorithms, particularly in the complex and dynamic context of construction environments. This architectural innovation represents a significant step forward in advancing the capabilities of AI-powered safety monitoring systems for the construction industry, demonstrating quantifiable superior performance compared to existing state-of-the-art algorithms, achieving a mAP50 of 0.70 and a mAP95 of 0.45 on the Construction Safety Dataset.

The creation of a new, large-scale Construction Safety Dataset, directly informed by a comprehensive analysis of OSHA Severe Injury reports, constitutes a second major contribution of this research. This dataset, comprising 55,594 meticulously annotated images across 28 safety categories, addresses a critical gap in existing resources by providing a dataset that is explicitly designed to reflect the most prevalent and impactful safety hazards in construction industry. The data-driven methodology employed for dataset development ensures its practical relevance, contextual richness, and suitability for training AI models that are truly aligned with real-world safety priorities and needs. The dataset serves as a valuable resource for the research community, facilitating further advancements in AI-powered construction safety and promoting the development of more robust and generalizable safety monitoring solutions.

Rigorous quantitative evaluation on both the SHEL5K benchmarking dataset and the developed Construction Safety Dataset provides compelling empirical evidence of the proposed model's superior performance and practical potential. Benchmarking against SOTA YOLO variants consistently demonstrated the enhanced model's advancements in object detection accuracy, localization precision, and balanced performance metrics, including mAP, precision, recall, F1-score, and box loss. The real-world case study deployment further validated the model's effectiveness in detecting safety violations in a complex construction environment, highlighting its real-time processing capabilities and practical utility for automated safety compliance monitoring.

The findings of this research underscore the transformative potential of AI and computer vision to revolutionize construction site safety management, moving beyond traditional reactive approaches towards proactive, data-driven, and technology-enabled solutions. The developed dataset and enhanced YOLOv8 model provide valuable tools for construction professionals and researchers alike, paving the way for a future where AI-powered systems play a central role in creating safer, more efficient, and more human-centered construction environments, ensuring every worker returns home safely at the end of each workday.

DECLARATION OF INTEREST

The authors state that they have no conflicting financial interests or personal relationships that might seem to have influenced the work presented in the study.

DATA AVAILABILITY

The dataset generated and analyzed during the current study is a key contribution of this research. To promote open science and ensure reproducibility, the complete dataset, including all 55,594 annotated images and corresponding annotation files, has been deposited in a public repository. The data is openly available on Zenodo at https://github.com/amr21006/PhD_Safety_Management_Object_Detection.git.

FUNDING

This research received no external funding.

REFERENCES

Bonyani, M., Soleymani, M., Wang, C., 2024. Construction workers' unsafe behavior detection through adaptive spatiotemporal sampling and optimized attention based video monitoring. Automation in Construction 165, 105508. https://doi.org/10.1016/j.autcon.2024.105508

BUREAU OF LABOR STATISTICS, 2022. Census of Fatal Occupational Injuries (CFOI) - Current and Revised Data [WWW Document]. URL https://www.bls.gov/iif/oshcfoi1.htm (accessed 10.4.22).

Chen, Z., Wu, Li, He, H., Jiao, Z., Wu, Liangsheng, 2022. Vision-based skeleton motion phase to evaluate working behavior: case study of ladder climbing safety. Human-centric Computing and Information Sciences 12.

CV-Tricks, 2017. Zero to Hero: Guide to Object Detection using Deep Learning: Faster R-CNN,YOLO,SSD [WWW Document]. CV-Tricks.com. URL https://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/ (accessed 8.3.23).

Delhi, V.S.K., Sankarlal, R., Thomas, A., 2020. Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques. Frontiers in Built Environment 6.

Du, H., 2023. General Object Detection Algorithm Yolov5 Comparison and Improvement (PhD Thesis). California State University, Northridge.

Duan, R., Deng, H., Tian, M., Deng, Y., Lin, J., 2022. SODA: A large-scale open site object detection dataset for deep learning in construction. Automation in Construction 142, 104499.

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T.M., An, W., 2018. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Automation in construction 85, 1–9.

Fang, W., Love, P.E.D., Ding, L., Xu, S., Kong, T., Li, H., 2023. Computer Vision and Deep Learning to Manage Safety in Construction: Matching Images of Unsafe Behavior and Semantic Rules. IEEE Transactions on Engineering Management 70, 4120–4132. https://doi.org/10.1109/TEM.2021.3093166

Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D., 2010. Object Detection with Discriminatively Trained Part-Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 1627–1645. https://doi.org/10.1109/TPAMI.2009.167

Girshick, R., 2015. Fast R-CNN, in: 2015 IEEE International Conference on Computer Vision (ICCV). Presented at the 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169

Hayat, A., Morgado-Dias, F., 2022. Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety. APPLIED SCIENCES-BASEL. https://doi.org/10.3390/app12168268

Hernández-García, A., König, P., 2018. Do deep nets really need weight decay and dropout?

Hou, X., Li, C., Fang, Q., 2023. Computer vision-based safety risk computing and visualization on construction sites. Automation in Construction 156, 105129.

Huang, L., Fu, Q., He, M., Jiang, D., Hao, Z., 2021. Detection algorithm of safety helmet wearing based on deep learning. Concurrency and Computation: Practice and Experience 33, e6234.

Jeelani, I., Asadi, K., Ramshankar, H., Han, K., Albert, A., 2021. Real-time vision-based worker localization & hazard detection for construction. Automation in Construction 121, 103448.

Jin, Z., Qu, P., Sun, C., Luo, M., Gui, Y., Zhang, J., Liu, H., 2021. DWCA-YOLOv5: An Improve Single Shot Detector for Safety Helmet Detection. Journal of Sensors 2021, e4746516. https://doi.org/10.1155/2021/4746516

Jocher, G., 2020. Ultralytics YOLOv5.

Jocher, G., Chaurasia, A., Qiu, J., 2023. YOLO by Ultralytics.

Kim, H., Seong, J., Jung, H.-J., 2023. Real-Time Struck-By Hazards Detection System for Small- and Medium-Sized Construction Sites Based on Computer Vision Using Far-Field Surveillance Videos. Journal of Computing in Civil Engineering 37, 04023028. https://doi.org/10.1061/JCCEE5.CPENG-5238

Kim, Hongjo, Kim, Hyoungkwan, Hong, Y.W., Byun, H., 2018. Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning. Journal of Computing in Civil Engineering 32, 04017082. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000731

Kolar, Z., Chen, H., Luo, X., 2018. Transfer learning and deep convolutional neural networks for safety guardrail detection in 2D images. Automation in Construction 89, 58–70. https://doi.org/10.1016/j.autcon.2018.01.003

Li, C., Li, Lulu, Jiang, H., Weng, K., Geng, Y., Li, Liang, Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, Xiaoming, Wei, Xiaolin, 2022. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications.

Li, Y., Wei, H., Han, Z., Huang, J., Wang, W., 2020. Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks. Advances in Civil Engineering 2020, e9703560. https://doi.org/10.1155/2020/9703560

Li, Y., Yao, T., Pan, Y., Mei, T., 2021. Contextual Transformer Networks for Visual Recognition. https://doi.org/10.48550/arXiv.2107.12292

Lin, Q., Ye, G., Wang, J., Liu, H., 2022. RoboFlow: a data-centric workflow management system for developing AI-enhanced Robots, in: Conference on Robot Learning. PMLR, pp. 1789–1794.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. SSD: Single Shot MultiBox Detector, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Ma, N., Zhang, X., Liu, M., Sun, J., 2021a. Activate or Not: Learning Customized Activation. https://doi.org/10.48550/arXiv.2009.04759

Ma, N., Zhang, X., Liu, M., Sun, J., 2021b. Activate or not: Learning customized activation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8032–8042.

Misra, D., Nalamada, T., Arasanipalai, A.U., Hou, Q., 2021. Rotate to attend: Convolutional triplet attention module, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3139–3148.

Nath, N.D., Behzadan, A.H., Paal, S.G., 2020. Deep learning for site safety: Real-time detection of personal protective equipment. Automation in Construction 112, 103085.

Nwankpa, C., Ijomah, W., Gachagan, A., Marshall, S., 2018. Activation Functions: Comparison of trends in Practice and Research for Deep Learning.

OSHA, 2, 2024. Severe Injury Reports | Occupational Safety and Health Administration [WWW Document]. URL https://www.osha.gov/severe-injury-reports (accessed 11.16.24).

Otgonbold, M.-E., Gochoo, M., Alnajjar, F., Ali, L., Tan, T.-H., Hsieh, J.-W., Chen, P.-Y., 2022. SHEL5K: An extended dataset and benchmarking for safety helmet detection. Sensors 22, 2315.

Park, M., Tran, D.Q., Bak, J., Park, S., 2023. Small and overlapping worker detection at construction sites. Automation in Construction 151, 104856. https://doi.org/10.1016/j.autcon.2023.104856

Park, S., Kim, Jaejun, Wang, S., Kim, Juhyung, 2025. Effectiveness of Image Augmentation Techniques on Non-Protective Personal Equipment Detection Using YOLOv8. Applied Sciences 15, 2631. https://doi.org/10.3390/app15052631

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Shao, Y., Zhang, R., Lv, C., Luo, Z., Che, M., 2024. TL-YOLO: Foreign-Object Detection on Power Transmission Line Based on Improved Yolov8. Electronics 13, 1543.

Sharma, T., Debaque, B., Duclos, N., Chehri, A., Kinder, B., Fortier, P., 2022. Deep learning-based object detection and scene perception under bad weather conditions. Electronics 11, 563.

Solawetz, D., B.,. Nelson, J., 2022. Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision.

Soleymani, M., Bonyani, M., Wang, C., 2024. Lightweight detection of small tools for safer construction. Automation in Construction 167, 105701. https://doi.org/10.1016/j.autcon.2024.105701

Tajeen, H., Zhu, Z., 2014. Image dataset development for measuring construction equipment recognition performance. Automation in Construction 48, 1–10. https://doi.org/10.1016/j.autcon.2014.07.006

Tato, A., Nkambou, R., 2018. Improving adam optimizer.

Tong, Z., Chen, Y., Xu, Z., Yu, R., 2023. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism.

Waehrer, G.M., Dong, X.S., Miller, T., Haile, E., Men, Y., 2007. Costs of occupational injuries in construction in the United States. Accident Analysis & Prevention 39, 1258–1266.

Wang, J., Chen, K., Xu, R., Liu, Z., Loy, C.C., Lin, D., 2019. Carafe: Content-aware reassembly of features, in: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3007–3016.

Wang, S., 2025. Automated non-PPE detection on construction sites using YOLOv10 and transformer architectures for surveillance and body worn cameras with benchmark datasets. Sci Rep 15, 27043. https://doi.org/10.1038/s41598-025-12468-8

Wang, Y., Xiao, B., Bouferguene, A., Al-Hussein, M., Li, H., 2022. Vision-based method for semantic information extraction in construction by integrating deep learning object detection and image captioning. Advanced Engineering Informatics 53, 101699. https://doi.org/10.1016/j.aei.2022.101699

Wang, Z., Wu, Y., Yang, L., Thirunavukarasu, A., Evison, C., Zhao, Y., 2021. Fast personal protective equipment detection for real construction sites using deep learning approaches. Sensors 21, 3478.

Wu, J., Cai, N., Chen, W., Wang, H., Wang, G., 2019. Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset. Automation in Construction 106, 102894.

Xiao, B., Kang, S.-C., 2021. Development of an Image Data Set of Construction Machines for Deep Learning Object Detection. Journal of Computing in Civil Engineering 35, 05020005. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000945

Xiao, B., Xiao, H., Wang, J., Chen, Y., 2022. Vision-based method for tracking workers by integrating deep learning instance segmentation in off-site construction. Automation in Construction 136, 104148.

Xing, J., Zhan, C., Ma, J., Chao, Z., Liu, Y., 2025. Lightweight detection model for safe wear at worksites using GPD-YOLOv8 algorithm. Sci Rep 15, 1227. https://doi.org/10.1038/s41598-024-83391-7

Xu, Z., Huang, J., Huang, K., 2022. A novel computer vision-based approach for monitoring safety harness use in construction. IET Image Processing n/a. https://doi.org/10.1049/ipr2.12696

Xuehui, A., Li, Z., Zuguang, L., Chengzhi, W., Pengfei, L., Zhiwei, L., 2021. Dataset and benchmark for detecting moving objects in construction sites. Automation in Construction 122, 103482. https://doi.org/10.1016/j.autcon.2020.103482

Yang, M., Wu, C., Guo, Y., Jiang, R., Zhou, F., Zhang, J., Yang, Z., 2023. Transformer-based deep learning model and video dataset for unsafe action identification in construction projects. Automation in Construction 146, 104703.

Zhao, Z.-Q., Zheng, P., Xu, S.-T., Wu, X., 2019. Object Detection With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems 30, 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865