IJRCS

International Journal of Robotics and Control Systems
Vol. 6, No. 1, 2026, pp. 528-539
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

ASCEE

Adaptive Feature Concatenation in YOLOv8n for Construction Site Safety Detection

Daffa Muhamad Azhar [a,1], Rihan Farih Bunyamin [a,2], Alissa Velia Rohyatul Jannah [a,3], Nanik Suciati [a,4,*], Muhammad 'Arif Faizin [a,5]

[a] Department of Informatics, Institut Teknologi Sepuluh Nopember, Jl. Raya ITS, Surabaya, 60111, Indonesia
[1] 6025241057@student.its.ac.id; [2] 6025241057@student.its.ac.id; [3] 6025241057@student.its.ac.id; [4] nanik@its.ac.id;
[5] 6025231092@student.its.ac.id
* Corresponding Author

ARTICLE INFO

Article history
Received December 02, 2025
Revised January 30, 2026
Accepted February 12, 2026

Keywords
Feature Fusion;
Object Detection;
Personal Protective Equipment;
YOLO

ABSTRACT

Monitoring Personal Protective Equipment (PPE) compliance is critical for construction safety, yet traditional lightweight detectors often struggle with severe occlusion and scale variations due to static feature fusion mechanisms. The main research contribution is the development of an enhanced YOLOv8n architecture integrating a novel Adaptive Feature Concatenation Module (AFCM) to address these limitations. Unlike standard concatenation, the proposed method employs AFCM in the neck network to dynamically recalibrate input features using learnable scalar weights normalized via Softmax. This mechanism allows the model to selectively emphasize semantically rich features while suppressing background noise without increasing channel dimensions. Experimental validation on the Construction Site Safety Image Dataset (CSSID) demonstrates that the model achieves an mAP@50:95 of 51.42%, surpassing the baseline YOLOv8n by 3.45%. Comparative analysis confirms that the proposed method outperforms state-of-the-art lightweight detectors, including YOLO-GP and MKD-YOLO, particularly in recovering small and occluded targets. Crucially, these improvements are achieved while maintaining 3.00 million parameters and 8.1 GFLOPs, identical to the baseline, ensuring no additional computational overhead. Consequently, the proposed framework offers a viable and effective solution for real-time, automated safety monitoring in resource-constrained edge devices.

1. Introduction

The construction industry plays a strategic role in infrastructure development and economic growth, yet it remains one of the most hazardous sectors due to the dynamic nature of labor, tools, and materials [1]-[3]. Construction safety management is therefore a priority to mitigate significant socioeconomic impacts [4]-[6], where the primary line of defense is the strict monitoring of Personal Protective Equipment (PPE) compliance [7]-[10]. However, non-compliance remains a major factor in accidents, often driven by behavioral factors and operational constraints [11]-[13]. Traditional monitoring methods, reliant on manual inspections, are labor-intensive, costly, and prone to human error [14]-[16]. Consequently, there is an urgent need for automated monitoring solutions capable of detecting PPE usage reliably in real-time [17]-[20], ensuring objective evidence generation [21]-[24].

While early automated approaches utilized sensor-based technologies like GPS and RFID [25], [26], these systems are less practical for dynamic sites due to device maintenance requirements [27]. To overcome these limitations, researchers have shifted toward computer vision-based detection [28]. Early vision methods depended on handcrafted features or pose estimation [29], [30], but modern deep learning has superseded these with dominant architectures: two-stage detectors (e.g., Faster R-CNN) and one-stage detectors (e.g., YOLO) [31]-[34]. While two-stage models excel in accuracy [35], [36], one-stage detectors are preferred for their inference speed [37]. Numerous studies have optimized YOLO variants (v3, v5) [38]-[41] and other architectures like SSD [42] for this domain. Subsequent research introduced enhancements such as face-helmet regression [43], anchor-free mechanisms [44], and MobileNet backbones [45]. To improve robustness, recent works have integrated aggregated attention [46], pose estimation [47], and edge-tailored optimizations [48], alongside refinements for class separation [49], dynamic convolution [50], and unsafe behavior detection [51], [52]. Advanced attention mechanisms have also been explored to address small object detection [53], [54].

Recently, several lightweight detectors have been specifically tailored for the Construction Site Safety Image Dataset (CSSID). Notable examples include YOLO-GP with Grouped Pointwise Convolution [55], MKD-YOLO utilizing knowledge distillation [56], YOLO-DS for low-light conditions [57], and GSO-YOLO for global optimization [58]. However, a critical research gap remains: these state-of-the-art lightweight models typically rely on static feature fusion mechanisms in the neck network. This approach assigns equal importance to all input features, leading to suboptimal performance when detecting small or occluded objects where semantic information must be distinguished from background noise.

To address this limitation, the research contribution is the development of an Adaptive Feature Concatenation Module (AFCM) integrated into the YOLOv8n architecture. Unlike static fusion, AFCM dynamically recalibrates feature importance using learnable scalar weights, allowing the network to emphasize relevant features while suppressing noise. This study aims to significantly enhance detection robustness in complex environments without sacrificing the computational efficiency required for real-time deployment.

2. Method

2.1. Research Procedure

The systematic workflow of this research is illustrated in Fig. 1. The procedure begins with the acquisition of the Construction Site Safety Image Dataset (CSSID), which is utilized in its pre-processed and pre-split state to ensure standardization. The core development phase involves the design of the proposed method, specifically integrating the Adaptive Feature Concatenation Module (AFCM) into the neck of the YOLOv8n architecture. Following this, the model undergoes a rigorous training process configured with the SGD optimizer, a learning rate of 0.01, and a duration of 300 epochs. Finally, the model is evaluated and compared against state-of-the-art (SOTA) methods using comprehensive metrics, including mAP, Precision, Recall, F1-score, as well as computational efficiency indicators such as Parameters and GFLOPs.

2.2. Baseline Model

Baseline model the YOLOv8n architecture is selected as the baseline framework due to its optimal balance between inference speed and detection accuracy [46], [48]. Briefly, the architecture comprises a CSPDarknet53 backbone for feature extraction, a Path Aggregation Network (PANet) neck for multi-scale feature fusion, and a decoupled head for object prediction. However, a critical limitation in the standard neck is its reliance on a static concatenation mechanism, which assigns equal importance to all input channels regardless of their semantic content. This lack of adaptability is suboptimal for construction safety monitoring, where upsampled feature maps often introduce background noise that suppresses the semantic details of small or occluded objects [7]. To address this specific deficiency, this study replaces the standard fusion layers with the proposed adaptive mechanism.

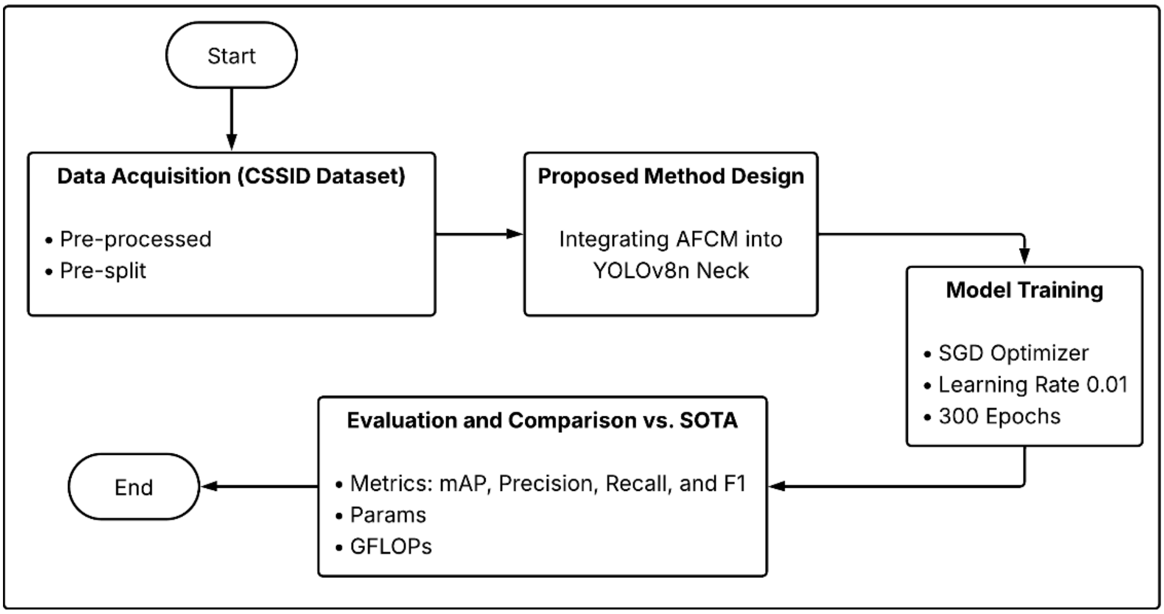

Fig. 1. The systematic research methodology flowchart illustrating the sequential phases from data acquisition to performance evaluation

2.3. Adaptive Feature Concatenation Module (AFCM)

The core contribution of this study is the Adaptive Feature Concatenation Module (AFCM). Unlike element-wise summation methods (e.g., BiFPN) which cause information loss through channel compression, or complex attention mechanisms (e.g., SE, CBAM) that introduce significant computational overhead, AFCM employs a lightweight, learnable weighting mechanism directly on the concatenation operation. The structure of the proposed AFCM is illustrated in Fig. 2.

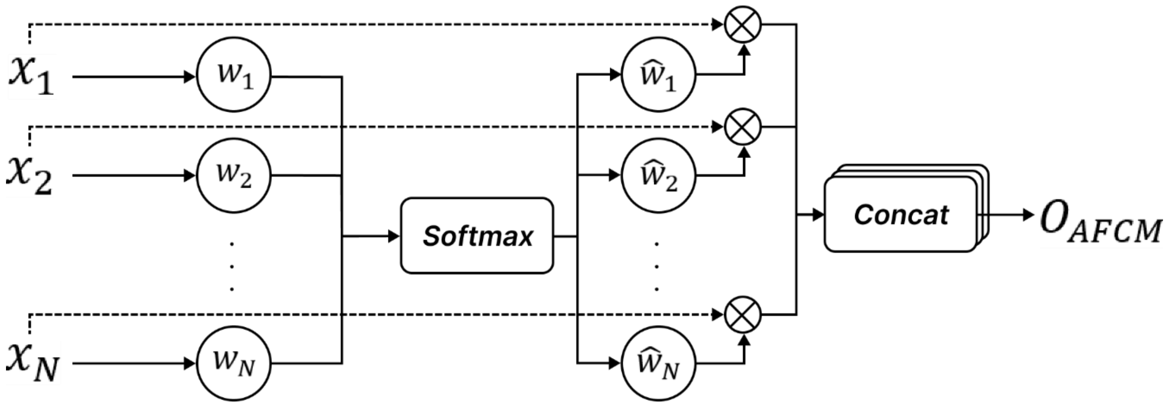

Fig. 2. Structure of the proposed adaptive feature concatenation module (AFCM). The module takes multi-scale feature maps as input, applies learned weights normalized via Softmax, and fuses them through a weighted concatenation operation

Let $X = \{x_1, x_2, \dots, x_N\}$ denote a set of input feature maps from different scales. We define a corresponding set of learnable scalar weights $w = \{w_1, w_2, \dots, w_N\}$. To ensure numerical stability and restrict the weights to a normalized range of $(0,1)$, a Softmax normalization is applied, as defined in (1).

$$\hat{w}_i = \frac{e^{w_i}}{\sum_{j=1}^{N} e^{w_j} + \varepsilon} \tag{1}$$

where ε is a small constant (1×10^{-4}) to prevent division by zero, and e is the Euler's number, ensuring that the resulting weights are always positive and differentiating the importance of features non-linearly. Critically, the raw parameters w_i are initialized to values that ensure a balanced initial contribution from all inputs, mimicking standard concatenation at the start of training to ensure optimization stability. The final output O_{AFCM} is obtained by scaling each input feature map with its normalized weight, as shown in (2).

$$O_{AFCM} = Concat(\widehat{w}_1 \cdot x_1, \widehat{w}_2 \cdot x_2, \dots, \widehat{w}_N \cdot x_N) \tag{2}$$

Through end-to-end backpropagation, the network learns to adjust $\widehat{w}_i$ dynamically. This mechanism allows the model to selectively suppress inconsistent features (such as background noise in upsampled layers) and amplify strong semantic cues from the backbone without altering the output channel dimensions or increasing inference latency.

2.4. Integration into YOLOv8n

The proposed AFCM is seamlessly integrated into the YOLOv8n architecture by replacing the standard concatenation layers within the PANet neck. This modification is strategically applied at two distinct fusion stages to address specific detection challenges as shown in Fig. 3.

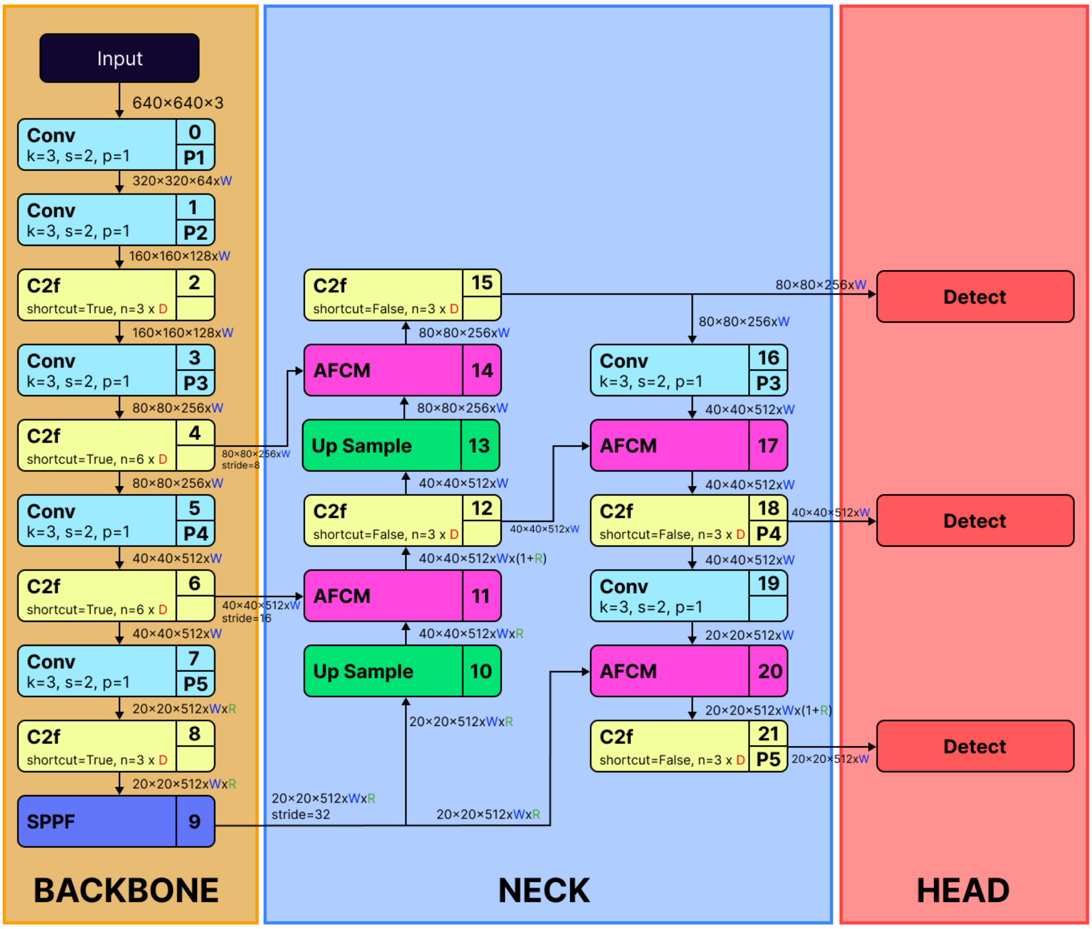

Fig. 3. Integrated AFCM into YOLOv8n

1. Top-Down Pathway: Fuses high-level features from the backbone with upsampled features. At this stage, AFCM is critical for enhancing semantic context, helping the model correctly classify objects even when they are partially occluded.
2. Bottom-Up Pathway: Fuses low-level features with the semantic-rich streams. Here, AFCM prioritizes the preservation of fine-grained spatial details, which is essential for localizing small objects such as distant hardhats or safety vests.

Crucially, since AFCM operates using simple scalar weights rather than convolutional layers, it maintains the output channel dimensions consistent with the original architecture. Consequently, it introduces zero additional parameters and negligible GFLOPs, ensuring that the enhanced model remains lightweight and suitable for real-time deployment on edge devices.

2.5. Dataset and Implementation Details

To evaluate the effectiveness of the proposed method, this study utilizes the Construction Site Safety Image Dataset (CSSID) sourced from the Roboflow platform. The dataset consists of images standardized to a resolution of 640×640 pixels to ensure consistent input dimensions. To enhance generalization capability, the dataset includes variations generated through geometric and photometric augmentations, such as random rotation and brightness adjustments. The specific distribution of object instances across the training, validation, and testing sets is detailed in Table 1.

Table 1. Distribution of object instances per class across training, validation, and testing sets in CSSID

Class	Number of Instances			Total
	Train	Validation	Test	
hardhat	3,145	79	110	3,334
mask	1,651	21	28	1,700
no-hardhat	2,317	69	41	2,427
no-mask	3,097	74	79	3,250
no-safety vest	3,962	106	90	4,158
person	9,532	166	174	9,872
safety cone	3,366	44	92	3,502
safety vest	3,033	41	61	3,135
machinery	5,247	55	44	5,346
vehicle	1,545	42	41	1,628
Total	36,895	697	760	38,352

All experiments were conducted on the Kaggle cloud platform utilizing an NVIDIA Tesla P100 GPU with 16 GB of VRAM. The software environment was configured with Python 3.10 and the PyTorch framework. The models were trained using the Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of 0.01, momentum of 0.937, and weight decay of 5×10^{-4}. To ensure efficiency, the training duration was set to 300 epochs, incorporating an early stopping mechanism with a patience of 25 epochs to prevent overfitting.

3. Results and Discussion

3.1. Comparison with State-of-the-Art Model

To rigorously validate the efficacy of the proposed YOLOv8-AFCM, a comparative analysis was conducted against the baseline and recent lightweight detectors tailored for safety detection, including YOLO-GP [55], MKD-YOLO [56], YOLO-DS [57], and GSO-YOLO [58]. The results are summarized in Table 2.

Regarding the fairness of the experimental settings, it is important to note that all comparative models listed in Table 2 were evaluated under identical conditions. Specifically, all models utilized the exact same training and testing splits of the CSSID dataset and were trained with a standardized input resolution of 640×640 pixels. This standardization ensures that the observed performance differences are solely attributable to the architectural innovations rather than discrepancies in data distribution or input scaling.

As demonstrated in Table 2, the proposed method establishes a new benchmark for the CSSID dataset. The YOLOv8n-AFCM achieves an mAP@50:95 of 51.42%, surpassing the baseline YOLOv8n (47.97%) by a substantial margin of 3.45%. Crucially, in terms of computational efficiency, the proposed model maintains an extremely lightweight profile with only 3.00 M parameters and 8.1 GFLOPs. This stands in sharp contrast to GSO-YOLO, which requires significantly higher resources (9.44 M parameters, 13.2 GFLOPs) to achieve lower accuracy

(44.25%). Even compared to the highly optimized MKD-YOLO, our method delivers a +6.82% improvement in mAP@50:95, confirming that the dynamic feature fusion of AFCM is more effective than knowledge distillation for this specific task.

Table 2. Performance comparison with state-of-the-art methods on CSSID

Model	Params (M)	GFLOPs	Precision (%)	Recall (%)	F1-Score (%)	mAP@50 (%)	mAP@50:95 (%)
YOLOv8n	3.00	8.1	87.73	70.22	77.64	76.57	47.97
YOLO-GP [55]	7.27	15.1	81.30	59.80	68.90	66.30	35.00
MKD-YOLO [56]	**2.43**	12.1	-	-	-	78.20	44.60
YOLO-DS [57]	3.14	20.4	83.90	64.91	72.70	70.81	39.49
GSO-YOLO [58]	9.44	13.2	82.20	68.48	74.15	73.50	44.25
Proposed	3.00	**8.1**	**90.83**	**72.99**	**80.46**	**78.96**	**51.42**

3.2. Ablation Study

To isolate the contribution of AFCM, we evaluated four configurations: Baseline (Concat), Weighted Add (Summation), Attention Concat, and the proposed AFCM. The results are detailed in Table 3. The proposed YOLOv8n-AFCM is the only configuration that improves performance over the baseline. Conversely, the “Weighted Add” [59] strategy resulted in a performance drop to 44.74%, confirming that channel compression causes “destructive interference,” where valuable spatial cues are lost. Similarly, adding complex attention mechanisms (Attention Concat) [60] degraded mAP to 42.43%, suggesting that excessive complexity introduces optimization difficulties on noisy construction images. These findings validate that the simple, learnable scalar weighting of AFCM provides the optimal balance between adaptability and stability.

Table 3. Ablation study of different feature fusion mechanisms on CSSID.

Model	Precision (%)	Recall (%)	F1-Score (%)	mAP@50(%)	mAP@50:95(%)
YOLOv8n	87.73	70.22	77.64	76.57	47.97
YOLOv8n-Weighted Add [59]	84.54	67.26	74.49	73.70	44.74
YOLOv8n-Attention Concat [60]	82.73	65.82	72.82	71.63	42.43
YOLOv8-Weighted Concat (proposed)	**90.83**	**72.99**	**80.46**	**78.96**	**51.42**

To further substantiate the improvement attributed specifically to the AFCM module, a statistical significance test (Welch’s t-test) was conducted comparing the confidence scores of the Baseline and the proposed model across the test set. The analysis revealed that the AFCM achieved a mean confidence score of 0.7135, significantly surpassing the Baseline’s mean of 0.6674. With a calculated p-value of 2.51×10^{-5} ($p < 0.001$), the results confirm that the performance gain demonstrated in the ablation study is statistically significant, proving that the AFCM systematically reduces prediction uncertainty.

3.3. Qualitative Analysis

To provide intuitive evidence of the model's robustness, Fig. 4 visualizes detection results on a challenging test sample characterized by varying lighting conditions and object scales. First, regarding object recovery, the baseline model (Fig. 4 a) exhibits a critical False Negative by failing to detect the black vehicle located in the background on the left side. Visual inspection suggests that this failure is primarily caused by poor contrast and visual camouflage, where the dark texture of the vehicle blends seamlessly with the wet, shadowed asphalt. In standard convolution operations, such low-gradient boundaries are often lost during downsampling. In contrast, the proposed YOLOv8-AFCM (Fig. 4 c) successfully recovers this difficult object with a confidence score of 0.56. This recovery demonstrates that the AFCM's adaptive weighting mechanism effectively amplifies weak spatial features, such as the vehicle's silhouette against the building, which allows the model to distinguish the object from the background noise despite the severe lack of contrast.

Second, regarding confidence enhancement, both models successfully detect the prominent white dump truck on the right. However, the baseline achieves a confidence score of 0.87, whereas the YOLOv8n-AFCM improves this to 0.97. This specific instance aligns with the statistical validation presented in the ablation study (Section 3.2), serving as a visual confirmation that the confidence gain is consistent across the dataset.

(a) (b) (c)

Fig. 4. Visual comparison of detection results. (a) Ground Truth showing annotations for vehicles and workers. (b) Baseline YOLOv8n results. (c) Proposed YOLOv8n-AFCM results

3.4. Discussion

The primary finding of this study is that replacing static concatenation with the proposed Adaptive Feature Concatenation Module (AFCM) significantly enhances detection performance without incurring additional computational overhead. The experiments validate that the dynamic feature recalibration effectively addresses the challenge of occlusion in construction sites, achieving a 3.45% improvement in mAP@50:95 compared to the baseline. This performance gain is not merely improved accuracy but also increased prediction certainty, as evidenced by the statistically significant rise in mean confidence scores ($p < 0.001$).

In the context of existing literature, this study distinguishes itself by prioritizing fusion efficiency over architectural complexity. Unlike previous works such as YOLO-GP [55] and GSO-YOLO [58], which rely on complex convolutional modules to improve accuracy, often at the expense of parameter efficiency, our approach optimizes how features are combined. Critically, while some SOTA methods achieve high accuracy by increasing network depth or input resolution, our comparison reveals that YOLOv8n-AFCM outperforms them (e.g., MKD-YOLO) even when evaluated under identical experimental settings (640×640 pixels). This confirms that for lightweight models, refining the feature fusion process is more effective than simply adding heavy architectural blocks.

These findings have significant implications for real-time construction safety monitoring. The model maintains an identical GFLOPs count (8.1) to the baseline, ensuring it can be deployed on resource-constrained edge devices without requiring hardware upgrades. Furthermore, the improved Recall, specifically the ability to recover low-contrast objects like the black vehicle discussed in the qualitative analysis, directly translates to fewer missed safety violations, potentially reducing accident rates in dynamic and visually cluttered site environments.

Despite these strengths, particularly the AFCM's ability to filter background noise via Softmax-normalized weights, certain limitations remain. The current evaluation focused on standard daylight and overcast conditions. The model's robustness in extreme weather scenarios, such as heavy rain or dense fog, has not been fully stress-tested. Future work should address these environmental variables, potentially by incorporating temporal information from video sequences to further stabilize detection in adverse conditions.

4. Conclusion

This study presented YOLOv8n-AFCM, a lightweight object detector that introduces a theoretical shift from conventional feature concatenation to dynamic channel recalibration. By effectively emphasizing relevant features and suppressing background noise, the proposed Adaptive Feature Concatenation Module (AFCM) addresses the critical challenge of occlusion in construction site safety monitoring. The experimental results confirm that this approach establishes a new state-of-the-art on the CSSID benchmark, achieving a 3.45% improvement in mAP@50:95 compared to the baseline. Critically, these gains are achieved with identical computational complexity (8.1 GFLOPs), contributing new knowledge to the domain by demonstrating that optimizing feature fusion is a far more efficient strategy for real-time edge deployment than increasing architectural depth.

Despite these advancements, the study has limitations. The evaluation was strictly confined to static images within the CSSID dataset and standard lighting conditions; the model's robustness in extreme environments, such as heavy rain or dense fog, remains to be fully quantified. Furthermore, the current implementation does not yet exploit temporal correlations found in video streams.

Future work will directly address these limitations by extending the AFCM mechanism to video-based architectures. Specifically, we aim to integrate temporal attention modules to stabilize detections across frames, thereby enhancing reliability in adverse weather conditions. Additionally, we plan to validate the generalizability of the AFCM by applying it to other lightweight architectures beyond the YOLO family.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding: This research received no external funding

Acknowledgement: The authors gratefully acknowledge financial support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2025.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] R. T. Mohona, S. Nawer, M. S. I. Sakib and M. N. Uddin, "A YOLOv8 Approach for Personal Protective Equipment (PPE) Detection to Ensure Workers' Safety," *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, pp. 1-6, 2024, https://doi.org/10.1109/ICAEEE62219.2024.10561752.

[2] S. Cucos and R. Turcan, "The Role of the Construction Industry in Economic Growth and Sustainable Development," *Journal of Social Sciences*, vol. 8, no. 1, pp. 25-38, 2025, https://doi.org/10.52326/jss.utm.2025.8(1).02.

[3] Z. Wang, Z. Cai, and Y. Wu, "An improved YOLOX approach for low-light and small object detection: PPE on tunnel construction sites," *Journal of Computational Design and Engineering*, vol. 10, no. 3, pp. 1158-1175, 2023, https://doi.org/10.1093/jcde/qwad042.

[4] B. H. W. Guo, Y. Zou, Y. Fang, Y. M. Goh, and P. X. W. Zou, "Computer vision technologies for safety science and management in construction: A critical review and future research directions," *Safety Science*, vol. 135, p. 105130, 2021, https://doi.org/10.1016/j.ssci.2020.105130.

[5] A. Gapur, D. Wahiddin, T. Al Mudzakir, and J. Indra, "Personal Protective Equipment Detection for Occupational Safety and Health using YOLOv8 in Manufacturing Companies," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 1175–1182, 2024, https://doi.org/10.52436/1.jutif.2024.5.5.2619.

[6] Q. Salsabila, "Health and Safety PPE Compliance Tracking," *Journal of Islamic Global Network for Information Technology and Entrepreneurship*, vol. 2, no. 2, pp. 36-44, 2024, https://doi.org/10.59841/ignite.v2i2.1049.

[7] S. Wang, “Automated non-PPE detection on construction sites using YOLOv10 and transformer architectures for surveillance and body worn cameras with benchmark datasets,” *Scientific Reports*, vol. 15, no. 1, p. 27043, 2025, https://doi.org/10.1038/s41598-025-12468-8.

[8] M. I. B. Ahmed *et al*., “Personal Protective Equipment Detection: A Deep-Learning-Based Sustainable Approach,” *Sustainability*, vol. 15, no. 18, p. 13990, 2023, https://doi.org/10.3390/su151813990.

[9] N. M. A. Ambarwati, D. Aryani, and Asriyadi, “Usage of Safety Helmet Warning System Using Deep Learning Method,” *Jurnal Teknologi Elekterika*, vol. 21, no. 2, pp. 93-99, 2024, https://doi.org/10.31963/elekterika.v21i2.5103.

[10] Y. Gao, V. A. González, and T. W. Yiu, “Exploring the relationship between construction workers’ personality traits and safety behavior,” *Journal of Construction Engineering and Management*, vol. 146, no. 3, p. 04019111, 2020, https://doi.org/10.1061/(ASCE)CO.1943-7862.0001763.

[11] A. Jalil Al-Bayati, A. T. Rener, M. P. Listello, and M. Mohamed, “PPE non-compliance among construction workers: An assessment of contributing factors utilizing fuzzy theory,” *Journal of Safety Research*, vol. 85, pp. 242-253, 2023, https://doi.org/10.1016/j.jsr.2023.02.008.

[12] R. Sehsah, A.-H. El-Gilany, and A. M. Ibrahim, “Personal protective equipment (PPE) use and its relation to accidents among construction workers,” *La Medicina del Lavoro*, vol. 111, no. 4, pp. 285-295, 2020, https://doi.org/10.23749/mdl.v111i4.9398.

[13] M. Namian, F. Taherpour, E. Ghiasvand, and Y. Turkan, “Insidious safety threat of fatigue: Investigating construction workers’ risk of accident due to fatigue,” *Journal of Construction Engineering and Management*, vol. 147, no. 12, p. 04021162, 2021, https://doi.org/10.1061/(ASCE)CO.1943-7862.0002180.

[14] Y. T. Lim, W. Yi, and H. Wang, “Application of Machine Learning in Construction Productivity at Activity Level: A Critical Review,” *Applied Sciences*, vol. 14, no. 22, p. 10605, 2024, https://doi.org/10.3390/app142210605.

[15] R. Azizi, M. Koskinopoulou, and Y. Petillot, “Comparison of Machine Learning Approaches for Robust and Timely Detection of PPE in Construction Sites,” *Robotics*, vol. 13, no. 2, p. 31, 2024, https://doi.org/10.3390/robotics13020031.

[16] N. A. N. M. Nazli, N. Sabri, R. Aminuddin, S. Ibrahim, S. Yusof, and S. D. N. M. Nasir, “A real-time system for detecting personal protective equipment compliance using deep learning model YOLOv5,” *Procedia Computer Science*, vol. 245, pp. 647-656, 2024, https://doi.org/10.1016/j.procs.2024.10.291.

[17] Y. Pan and L. Zhang, “Roles of artificial intelligence in construction engineering and management: A critical review and future trends,” *Automation in Construction*, vol. 122, p. 103517, 2021, https://doi.org/10.1016/j.autcon.2020.103517.

[18] F. Alsakka, S. Assaf, I. El-Chami, and M. Al-Hussein, “Computer vision applications in offsite construction,” *Automation in Construction*, vol. 154, p. 104980, 2023, https://doi.org/10.1016/j.autcon.2023.104980.

[19] S. Setyawan and E. P. Widiyanto, “Analisis Kinerja Model YOLOv8 Untuk Monitoring Kepatuhan Penggunaan Sepatu Safety pada Petugas Pemadam Kebakaran,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3S1, pp. 1296-1301, 2025, https://doi.org/10.23960/jitet.v13i3S1.8065.

[20] L. Liu *et al*., “Multi-Task Intelligent Monitoring of Construction Safety Based on Computer Vision,” *Buildings*, vol. 14, no. 8, p. 2429, 2024, https://doi.org/10.3390/buildings14082429.

[21] J.-H. Lo, L.-K. Lin, and C.-C. Hung, “Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm,” *Sustainability*, vol. 15, no. 1, p. 391, 2022, https://doi.org/10.3390/su15010391.

[22] B. Tong, G. Li, X. Bu, Y. Wang, and X. Yu, “A deep learning-based algorithm for the detection of personal protective equipment,” *PLoS One*, vol. 20, no. 5, p. e0322115, 2025, https://doi.org/10.1371/journal.pone.0322115.

[23] F. Formisano, M. Dellutri, E. Massera, A. Del Giudice, L. Barretta, and G. Di Francia, “Wearable Sensor Node for Safety Improvement in Workplaces: Technology Assessment in a Simulated Environment,” *Sensors*, vol. 24, no. 15, p. 4993, 2024, https://doi.org/10.3390/s24154993.

[24] A. M. Vukicevic, M. Petrovic, P. Milosevic, A. Peulic, K. Jovanovic, and A. Novakovic, "A systematic review of computer vision-based personal protective equipment compliance in industry practice: advancements, challenges and future directions," *Artificial Intelligence Review*, vol. 57, no. 12, p. 319, 2024, https://doi.org/10.1007/s10462-024-10978-x.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv*, 2015, https://doi.org/10.48550/arXiv.1512.03385.

[26] M. Arabshahi *et al*., "Review on Sensing Technology Adoption in the Construction Industry," *Sensors*, vol. 21, no. 24, p. 8307, 2021, https://doi.org/10.3390/s21248307.

[27] S. Chen and K. Demachi, "A Vision-Based Approach for Ensuring Proper Use of Personal Protective Equipment (PPE) in Decommissioning of Fukushima Daiichi Nuclear Power Station," *Applied Sciences*, vol. 10, no. 15, p. 5129, 2020, https://doi.org/10.3390/app10155129.

[28] A. Barlybayev *et al*., "Personal protective equipment detection using YOLOv8 architecture on object detection benchmark datasets: a comparative study," *Cogent Engineering*, vol. 11, no. 1, 2024, https://doi.org/10.1080/23311916.2024.2333209.

[29] Y.-R. Lee, S.-H. Jung, K.-S. Kang, H.-C. Ryu, and H.-G. Ryu, "Deep learning-based framework for monitoring wearing personal protective equipment on construction sites," *Journal of Computational Design and Engineering*, vol. 10, no. 2, pp. 905-917, 2023, https://doi.org/10.1093/jcde/qwad019.

[30] A. M. Vukicevic, M. Djapan, V. Isailovic, D. Milasinovic, M. Savkovic, and P. Milosevic, "Generic compliance of industrial PPE by using deep learning techniques," *Safety Science*, vol. 148, p. 105646, 2022, https://doi.org/10.1016/j.ssci.2021.105646.

[31] H. Choo, B. Lee, H. Kim, and B. Choi, "Automated detection of construction work at heights and deployment of safety hooks using IMU with a barometer," *Automation in Construction*, vol. 147, p. 104714, 2023, https://doi.org/10.1016/j.autcon.2022.104714.

[32] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustainable Cities and Society*, vol. 65, p. 102600, 2021, https://doi.org/10.1016/j.scs.2020.102600.

[33] A. A. Protik, A. H. Rafi and S. Siddique, "Real-time Personal Protective Equipment (PPE) Detection Using YOLOv4 and TensorFlow," *2021 IEEE Region 10 Symposium (TENSYMP)*, pp. 1-6, 2021, https://doi.org/10.1109/TENSYMP52854.2021.9550808.

[34] R. Xiong and P. Tang, "Pose guided anchoring for detecting proper use of personal protective equipment," *Automation in Construction*, vol. 130, p. 103828, 2021, https://doi.org/10.1016/j.autcon.2021.103828.

[35] Z. Wang, Y. Zhang, and S. Zhang, "Real-time personal protective equipment detection and classification with YOLOv8 multi-scale fusion," *Journal of Real-Time Image Processing*, vol. 22, no. 3, p. 131, 2025, https://doi.org/10.1007/s11554-025-01715-w.

[36] B. Kong, Z. Cao, T. Sun, C. Qi, and Y. Zhang, "Safety hazards in coal mines of Guizhou China during 2011–2020," *Safety Science*, vol. 145, p. 105493, 2022, https://doi.org/10.1016/j.ssci.2021.105493.

[37] J. Zhang *et al*., "Multi-class object detection using faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting," *Computers and Electronics in Agriculture*, vol. 173, p. 105384, 2020, https://doi.org/10.1016/j.compag.2020.105384.

[38] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: Real-time detection of personal protective equipment," *Automation in Construction*, vol. 112, p. 103085, 2020, https://doi.org/10.1016/j.autcon.2020.103085.

[39] X. Ji, F. Gong, X. Yuan, and N. Wang, "A high-performance framework for personal protective equipment detection on the offshore drilling platform," *Complex & Intelligent Systems*, vol. 9, no. 5, pp. 5637–5652, 2023, https://doi.org/10.1007/s40747-023-01028-0.

[40] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao, "Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches," *Sensors*, vol. 21, no. 10, p. 3478, 2021, https://doi.org/10.3390/s21103478.

[41] M. M. Alateeq, F. R. P.P., and M. A. S. Ali, "Construction Site Hazards Identification Using Deep Learning and Computer Vision," *Sustainability*, vol. 15, no. 3, p. 2358, 2023, https://doi.org/10.3390/su15032358.

[42] Y. Li, H. Wei, Z. Han, J. Huang, and W. Wang, "Deep Learning-Based Safety Helmet Detection in Engineering Management Based on Convolutional Neural Networks," *Advances in Civil Engineering*, vol. 2020, no. 1, pp. 1-10, 2020, https://doi.org/10.1155/2020/9703560.

[43] J. Shen, X. Xiong, Y. Li, W. He, P. Li, and X. Zheng, "Detecting safety helmet wearing on construction sites with bounding-box regression and deep transfer learning," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 2, pp. 180-196, 2021, https://doi.org/10.1111/mice.12579.

[44] Md. Ferdous and Sk. Md. M. Ahsan, "PPE detector: a YOLO-based architecture to detect personal protective equipment (PPE) for construction sites," *PeerJ Computer Science*, vol. 8, p. e999, 2022, https://doi.org/10.7717/peerj-cs.999.

[45] K. Huang and M. B. Abisado, "Lightweight construction safety behavior detection model based on improved YOLOv8," *Discover Applied Sciences*, vol. 7, no. 4, p. 326, 2025, https://doi.org/10.1007/s42452-025-06766-z.

[46] X. Li, M. Hu, B. Li, and R. Tong, "OAM-YOLO: A real-time small object detection framework for PPE compliance monitoring in industrial environments," *Process Safety and Environmental Protection*, vol. 204, p. 108058, 2025, https://doi.org/10.1016/j.psep.2025.108058.

[47] Y. Zhang *et al.*, "Detection of helmet use among construction workers via helmet-head region matching and state tracking," *Automation in Construction*, vol. 171, p. 105987, 2025, https://doi.org/10.1016/j.autcon.2025.105987.

[48] D. Kim and S. Xiong, "Enhancing Worker Safety: Real-Time Automated Detection of Personal Protective Equipment to Prevent Falls from Heights at Construction Sites Using Improved YOLOv8 and Edge Devices," *Journal of Construction Engineering and Management*, vol. 151, no. 1, 2025, https://doi.org/10.1061/JCEMD4.COENG-14985.

[49] H. Samma, S. Al-Azani, H. Luqman, and M. Alfarraj, "Contrastive-based YOLOv7 for personal protective equipment detection," *Neural Computing and Applications*, vol. 36, no. 5, pp. 2445-2457, 2024, https://doi.org/10.1007/s00521-023-09212-6.

[50] H. Wang *et al.*, "Exploring human factors of major chemical accidents in China: Evidence from 160 accidents during 2011–2022," *Journal of Loss Prevention in the Process Industries*, vol. 89, p. 105279, 2024, https://doi.org/10.1016/j.jlp.2024.105279.

[51] M. Shahin *et al.*, "Enhanced safety implementation in 5S + 1 via object detection algorithms," *The International Journal of Advanced Manufacturing Technology*, vol. 125, no. 7-8, pp. 3701-3721, 2023, https://doi.org/10.1007/s00170-023-10970-9.

[52] X. Jiao *et al.*, "Detection Method for Safety Helmet Wearing on Construction Sites Based on UAV Images and YOLOv8," *Buildings*, vol. 15, no. 3, p. 354, 2025, https://doi.org/10.3390/buildings15030354.

[53] S. Malaikrisanachalee, N. Wongwai, and E. Kowcharoen, "ESPCN-YOLO: A High-Accuracy Framework for Personal Protective Equipment Detection Under Low-Light and Small Object Conditions," *Buildings*, vol. 15, no. 10, p. 1609, 2025, https://doi.org/10.3390/buildings15101609.

[54] Y. Li *et al.*, "SOD-YOLO: Small-Object-Detection Algorithm Based on Improved YOLOv8 for UAV Images," *Remote Sensing*, vol. 16, no. 16, p. 3057, 2024, https://doi.org/10.3390/rs16163057.

[55] B. Liu, C. Yu, B. Chen, and Y. Zhao, "YOLO-GP: A Multi-Scale Dangerous Behavior Detection Model Based on YOLOv8," *Symmetry*, vol. 16, no. 6, p. 730, 2024, https://doi.org/10.3390/sym16060730.

[56] J. Zan, Y. Fang, Q. Liu, U. Khairuddin, Y. Li and K. Sun, "MKD-YOLO: Multi-Scale and Knowledge-Distilling YOLO for Efficient PPE Compliance Detection," *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1-5, 2025, https://doi.org/10.1109/ICASSP49660.2025.10889626.

[57] C. Zhang and J. Shen, "YOLO-DS: Application of One-Stage Instance Segmentation in the Dark at Construction Sites," *2024 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, 2024, https://doi.org/10.1109/IJCNN60899.2024.10650361.

[58] Y. Zhang, D. Guan, S. Zhang, J. Su, Y. Han, J. Liu, "GSO-YOLO: Global Stability Optimization YOLO for Construction Site Detection," *arXiv*, 2024, https://doi.org/10.48550/arXiv.2407.00906.

[59] M. Tan, R. Pang, and Q. V Le, "EfficientDet: Scalable and Efficient Object Detection," *arXiv*, 2020, https://doi.org/10.48550/arXiv.1911.09070.

[60] L. Xiaojuan, "Research on Multi-Feature Fusion of Improved Concat Module Based on Yolov11 in Wheat Spike Recognition," *International Journal of Cognitive Informatics and Natural Intelligence*, vol. 19, no. 1, pp. 1-18, 2025, https://doi.org/10.4018/IJCINI.386848.