

Improved CoTNet-SE Model for Smart Helmet Positioning Error Correction Under UWB and Inertial Sensor Sequences

Damian Kaczmarek^{1,*}, Hubert Kasprzyk¹ and Alicja Gruszka¹

¹ Faculty of Electrical and Automatic Control, University of Warmia and Mazury in Olsztyn, Olsztyn, 10-719, Poland

*Corresponding author: damian.ka@uwm.edu.pl

Abstract. This study develops an improved CoTNet-SE model for smart helmet positioning error correction under UWB and inertial sensor sequences. The task addresses construction-site localization errors caused by temporary barriers, steel structures, material stacks, moving equipment, worker posture changes, non-line-of-sight ranging, inertial drift, and brief trajectory interruptions. A simulated dataset is constructed with 42 work zones, 18 UWB anchors, 96 helmet terminals, and 1.48 million synchronized UWB-IMU samples collected from open spaces, scaffold corridors, stair entrances, underground passages, and material storage yards. The model builds UWB range residuals, IMU motion windows, anchor-visibility masks, signal-confidence descriptors, zone-distance features, and trajectory context features. Contextual Transformer Convolution learns local-to-global error patterns, while Squeeze-and-Excitation recalibration selects reliable sensing channels before coordinate correction and safety-zone scoring. Compared with raw UWB, Kalman filtering, LSTM fusion, temporal convolution, and attention-only baselines, the model reduces mean positioning error from 0.74 m to 0.39 m, decreases 95th-percentile error from 1.82 m to 0.91 m, and lowers safety-zone false alarms by 31.6%. The average inference latency is 8.7 ms per helmet window on a desktop setting, supporting edge-side updates. The method remains limited by simulated data and requires real-site validation, anchor calibration, privacy protection, weather testing, and map-update integration.

Keywords: *UWB-IMU Fusion, Smart Helmet Localization, Positioning Error Correction, CoTNet-SE, Construction Safety,*

Received on 04 April 2026, Accepted on 05 July 2026, Published on 13 July 2026

Copyright © 2026 Author(s), licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Smart helmets are important sensing carriers for construction safety that can integrate worker identification, movement path, local environment and risk-zone access into a single digital record. [1]. Workers move through lifting zones, material yards, underground passages, reinforcing cages, formwork areas and temporary routes in busy construction sites, and Nkrow et al. [2] found that time-based indoor localisation is easily disrupted by non-line-of-sight conditions. A positioning error of less than one metre can still change the alert status when a worker is near a hoisting boundary, edge-protection line, stair entrance or equipment turning radius. Wang et al. [3] have shown that UWB accuracy is significantly affected by NLOS/LOS classification. Wearable positioning becomes more reliable when inertial cues are combined with UWB measurements, and Liu et al. [4] have demonstrated improved indoor positioning under dynamic feasible-region constraints through IMU/UWB fusion. Therefore, this study views smart helmet localisation as an error-correction and safety-warning problem rather than simple coordinate acquisition.

The main difficulty is the coupled error behavior of UWB and inertial sensing. UWB ranging can provide high precision in open indoor environments, but construction sites are rarely clean radio spaces. Steel columns, dense materials, crane components, concrete walls, temporary lifts, and stacked pipelines can produce multipath delay and non-line-of-sight measurements. IMU sequences preserve short-term motion continuity between UWB updates, yet acceleration integration and heading estimates drift when the helmet is shaken, tilted, or affected

by worker posture. Earlier localization research has discussed NLOS instability, and inertial fusion research has emphasized time-accumulated drift. Safety harness detection [5] and IoT-based worker monitoring [6] further show that sensing errors can directly affect field safety decisions. The model must therefore judge whether a coordinate jump comes from real motion, UWB disturbance, or inertial compensation failure.

Deep sequence models can learn error patterns that are difficult to express through fixed thresholds. Recurrent models, temporal convolutions, graph-based range correction, and attention methods have all been applied to localization and sensor-fusion problems, but helmet-side and edge-side systems require high throughput and interpretable outputs. The model should be insensitive to normal worker motion while remaining sensitive to sudden signal loss, anchor blockage, and safety-zone boundary approach. Channel attention research suggests that sensing dimensions should not contribute equally in all windows [7], and contextual feature extraction can help distinguish open-area drift, steel-structure jumps, and brief occlusion [8]. Digital construction studies [9] and wearable safety studies [10] also show that corrected locations must be connected with operational data and understandable warnings. These needs motivate the improved CoTNet-SE model.

This is how the remainder of the paper is structured. Inertial sequence correction, attention-based lightweight modelling, and construction-site localisation are introduced in Section 2. The UWB-IMU data creation, CoTNet-SE correction method, training aim, and safety-zone output are all introduced in Section 3. The simulated data set, position accuracy, trajectory stability, robustness, ablation analysis, and warning application are displayed in data figures and one comparison table in Section 4. The primary findings and deployment issues are displayed in Section 5.

Related Work

UWB-Based Smart Helmet Localization

Due to the brief pulses' relative immunity to narrow-band interference and ability to accomplish high-precision time-of-flight range, UWB localisation is frequently used in indoor and semi-indoor conditions. To connect a worker's location to a safety zone and process data, a position terminal is frequently fastened to a hard helmet, vest, or cell phone on the construction site. Anchor geometry, sample frequency, range methodology, and ambient reflection all have an impact on accuracy, according to studies of indoor localisation surveys [11]. Studies on construction monitoring [12] have also demonstrated that when workers enter temporarily restricted areas or move behind machinery, the sensing system must maintain a consistent identity-location association. By connecting movement with job tasks and warning responsibilities, smart helmets enhance the utility of positioning. However, UWB comparative study [13] demonstrates that device-level variations may still alter the ultimate mistake pattern. Geometric correction and behavior-aware filtering are therefore necessary for UWB Helmet localisation.

Inertial Sensor Sequence Correction

IMU data are utilised to preserve trajectory continuity between UWB updates and give acceleration, angular velocity, and attitude changes. Short-term motion cues can lessen coordinate discontinuities in the presence of a weak radio signal, according to research on inertial navigation [14]. Minimise the IMU's drift. The helmet's position may be altered by a worker's routine activities, such as bending, turning, climbing stairs, carrying items, gazing around, etc., without requiring a substantial amount of body movement. Thus, temporal filtering, motion-state identification, and confidence allocation across sources are the main focus of sensor fusion studies [15], and review work on worker tracking [16] has further demonstrated that location continuity has to be assessed in practice under site mobility. UWB measurements are not replaced by the IMU sequence in the suggested job. To assist the model in determining if the present UWB value is physically reasonable based on the recent velocity, direction, and movement rhythm, set a time restriction.

Attention Recalibration in Lightweight Models

Sequences of sensors may be modelled using attention methods, and depending on the work context, various feature channels are given varied weights. In an outside yard, IMU correction will be minimal and UWB range may be rather precise. The model must give inertial continuity and anchor visibility greater weight since residuals

in a steel corridor are often bigger. In order to recognise short-window patterns and longer motion contexts, Contextual Transformer Networks [17] combine convolutional localisation with attention-based context aggregation. Another compact technique for allocating various weights to UWB distance residuals, inertial increments, signal quality, and safety-zone distance is SE-style channel recalibration [18]. This choice is supported by selective kernel modelling [19] and efficient channel attention [20], as helmet localisation must satisfy edge-device latency constraints and cannot employ a hefty cloud model for each correction step.

Methodology

Multi-Source Positioning Signal Construction

The input sequence is created by combining the synchronous UWB and IMU data captured by the smart helmet terminal. There are 24 samples after resampling to 10 Hz, with each location window lasting 2.4 seconds. The raw UWB coordinate, the distance between the helmet and visible anchors, the anchor signal quality, acceleration, angular velocity, yaw variation, and the most recent corrected coordinate are all saved by the system for every sample. Open ground, scaffold corridor, reinforced concrete barrier, material storage area, stair entry, temporary road, and equipment interaction area are the 42 zones of the construction site. Before the network topology is presented, the UWB-IMU relationship may be understood since Figure 1 depicts the entire signal gathering process to safety-zone output.

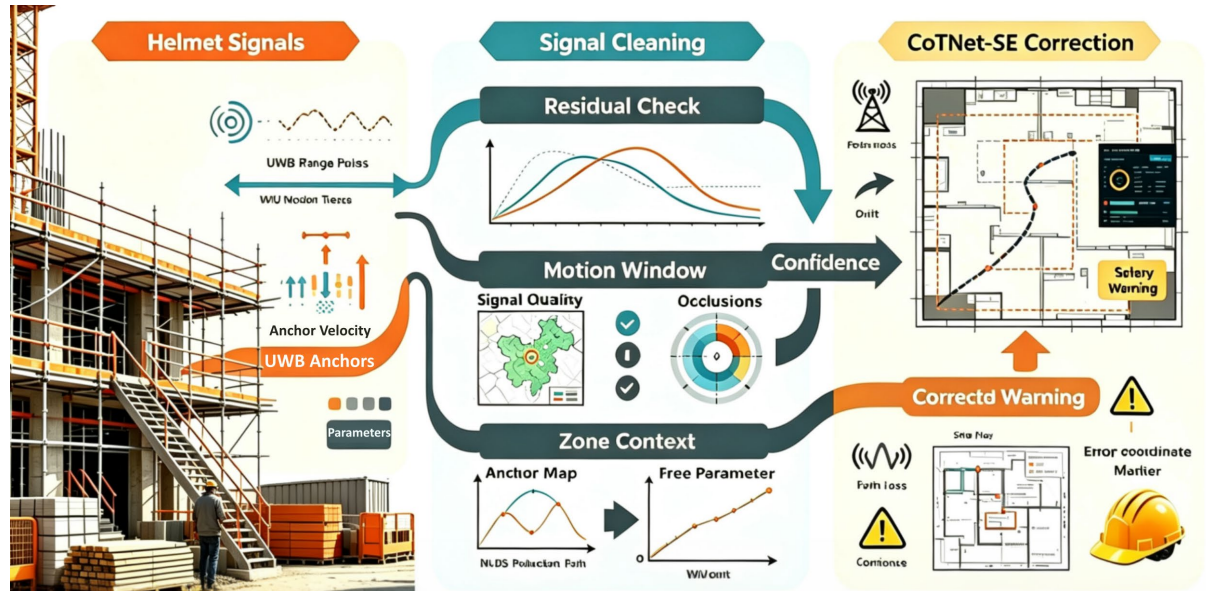


Figure 1. UWB-IMU smart helmet positioning error correction workflow.

The difference between the observed distance and the helmet state's geometric distance to each anchor is used to compute the UWB ranging residual. The residual is a direct indicator of non-line-of-sight interference and indicates how far the present UWB signal deviates from the anchor geometry.

$$r_{i,t} = d_{i,t}^{uwb} - \|p_t - b_i\|_2 \quad \text{Eq.(1)}$$

Short-term motion continuity is demonstrated by carrying over the inertial state inside the same window. The acceleration term is a displacement tendency that may be connected to changes in UWB coordinates rather than an absolute-position source.

$$s_t^{\text{imu}} = [a_t, \omega_t, \theta_t, v_{t-1} + a_t \Delta t] \quad \text{Eq.(2)}$$

UWB residuals, inertial increments, anchor visibility, and local safety-zone distance make up the multi-source input vector. In the simulated dataset, every vector has 38 normalised dimensions. Missing anchor observations are portrayed as visibility masks and confidence ratings rather than being consistently assigned to a coordinate, and the missingness itself might function as a feature.

$$x_t = \text{concat}(r_{1,t}, \dots, r_{m,t}, s_t^{\text{imu}}, q_t, z_t) \quad \text{Eq.(3)}$$

Before training the model, raw records are filtered using a window-level validity check. A window is retained when at least four anchors are visible in general areas, or when three anchors are visible in restricted corridors where site geometry constrains the worker route. The timestamp difference between IMU and UWB records must not exceed 35 milliseconds. If the difference is larger, the sample is treated as asynchronous and excluded from the main analysis so that the model does not learn false jumps caused by recording delay. Each trajectory is tagged with scene type, occlusion state, worker motion state, and closest safety-zone category. The simulated coordinate ground truth is built from calibrated reference trajectories with 0.05 m uncertainty.

CoTNet-SE Error Correction Model

After receiving the window sequence, the improved CoTNet-SE model generates a coordinate correction vector rather than predicting absolute position from scratch. Figure 2 shows the model structure. UWB residual streams, IMU sequence streams, anchor-confidence maps, and zone-distance features enter the left side of the network. Contextual convolution learns temporal-neighbor associations in the central block, and the SE branch dynamically recalibrates feature channels according to global window statistics. The right side outputs corrected coordinates, trajectory confidence, vertical-level correction flags, and safety-zone warning scores. This design preserves useful UWB geometry while allowing IMU continuity and scene context to correct unreliable windows.

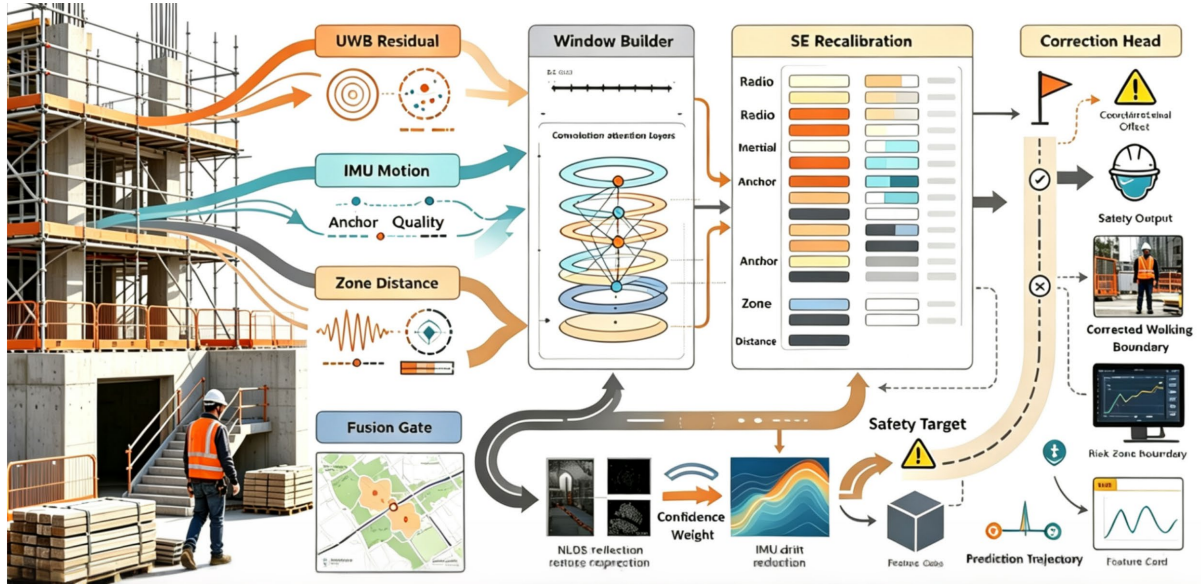


Figure 2. Improved CoTNet-SE error correction network.

The contextual feature block maps local sequence features into a context-aware representation. This operation is useful when a short UWB jump should be interpreted together with several previous stable points instead of being corrected in isolation.

$$h_t = \phi_c \left(W_k x_t + \sum_{j \in N(t)} \alpha_{t,j} W_v x_j \right) \quad \text{Eq.(4)}$$

The attention coefficient measures the relation between the current state and neighbouring states in the same window. A window with stable movement should produce smooth attention, while a window with occlusion or anchor loss should produce stronger contrast between reliable and unreliable moments.

$$\alpha_{t,j} = \text{softmax} \left(\frac{(W_q x_t)^T (W_k x_j)}{\sqrt{d}} \right) \quad \text{Eq.(5)}$$

The SE branch recalibrates feature channels by using global window statistics. It gives higher weights to signal dimensions that remain stable in the current scene and lower weights to disturbed or redundant channels.

$$w_c = \sigma \left(W_2 \delta(W_1 \text{mean}_t(h_t)) \right) \quad \text{Eq.(6)}$$

The recalibrated sequence feature is then obtained through channel-wise multiplication. This step is important because UWB residual, IMU direction, and zone-distance features have different reliability in different construction areas.

$$g_t = w_c \odot h_t \quad \text{Eq.(7)}$$

Instead of using a deep fully connected stack, the corrected representation is passed to a shallow regression head so that many helmets can be processed at the edge gateway. The regression head contains two hidden layers with 128 and 64 units, and dropout is used only during training. The output includes a horizontal coordinate correction, a confidence score, and a vertical-level correction flag. The vertical flag is necessary because the same horizontal map location can have different safety meanings on stairs or platforms. For example, a worker near a stair entrance may appear close to a restricted area in two-dimensional space, but the alert depends on whether the worker is on the same level or moving between levels.

Training Objective and Warning Output

The coordinate correction vector is estimated from the recalibrated representation. The corrected coordinate is added to the raw UWB coordinate rather than predicted from scratch, because raw UWB still carries useful geometric information even when it is noisy.

$$\hat{p}_t = p_t^{uwb} + W_o g_t \quad \text{Eq.(8)}$$

The training loss combines point error, temporal smoothness, and safety-zone consistency. Point error penalizes the distance between corrected and reference coordinates. Temporal smoothness reduces implausible jumps across neighboring windows, and safety-zone consistency penalizes corrections that change warning status incorrectly near restricted areas. In the simulated training setup, zones and neighboring trajectories are separated to reduce leakage between similar routes. The corrected coordinate is added to the raw UWB coordinate rather than predicted independently because raw UWB still carries useful anchor geometry even under noise. The safety-zone score is then computed from corrected position, confidence, motion direction, and distance to restricted boundaries.

$$L = \text{mean}(\|\hat{p}_t - p_t^*\|_1) + \lambda_s \text{mean}(\|\Delta \hat{p}_t\|_2) + \lambda_z L_{\text{zone}} \quad \text{Eq.(9)}$$

The safety-zone score is computed after coordinate correction. The score increases when a worker approaches a restricted zone, equipment interaction boundary, or temporary edge-protection line. It is not a substitute for site rules; it converts corrected positioning into a readable warning signal for managers.

$$R_t = \sigma(\beta_1 \text{dist}(\hat{p}_t, Z) + \beta_2 c_t + \beta_3 u_t) \quad \text{Eq.(10)}$$

Results and Discussion

Dataset and Positioning Error Source Analysis

The simulated dataset contains 1.48 million synchronous UWB-IMU samples from 18 UWB anchors and 96 smart helmets. Five representative construction settings are included: open ground, scaffold corridor, steel-structure area, stair entry, and material storage yard. Figure 3 presents the dataset and positioning error sources. Figure 3(a) reports scene type proportions: open ground, scaffold corridors, steel-structure regions, stair entrances, and material yards account for 29.5%, 21.0%, 18.7%, 14.6%, and 16.2% of samples. Figure 3(b) shows that low-confidence UWB ranging is most common in steel and material zones. Figure 3(c) shows IMU drift intervals, and Figure 3(d) reports occlusion types including temporary walls, piled materials, equipment bodies, and human crowding.

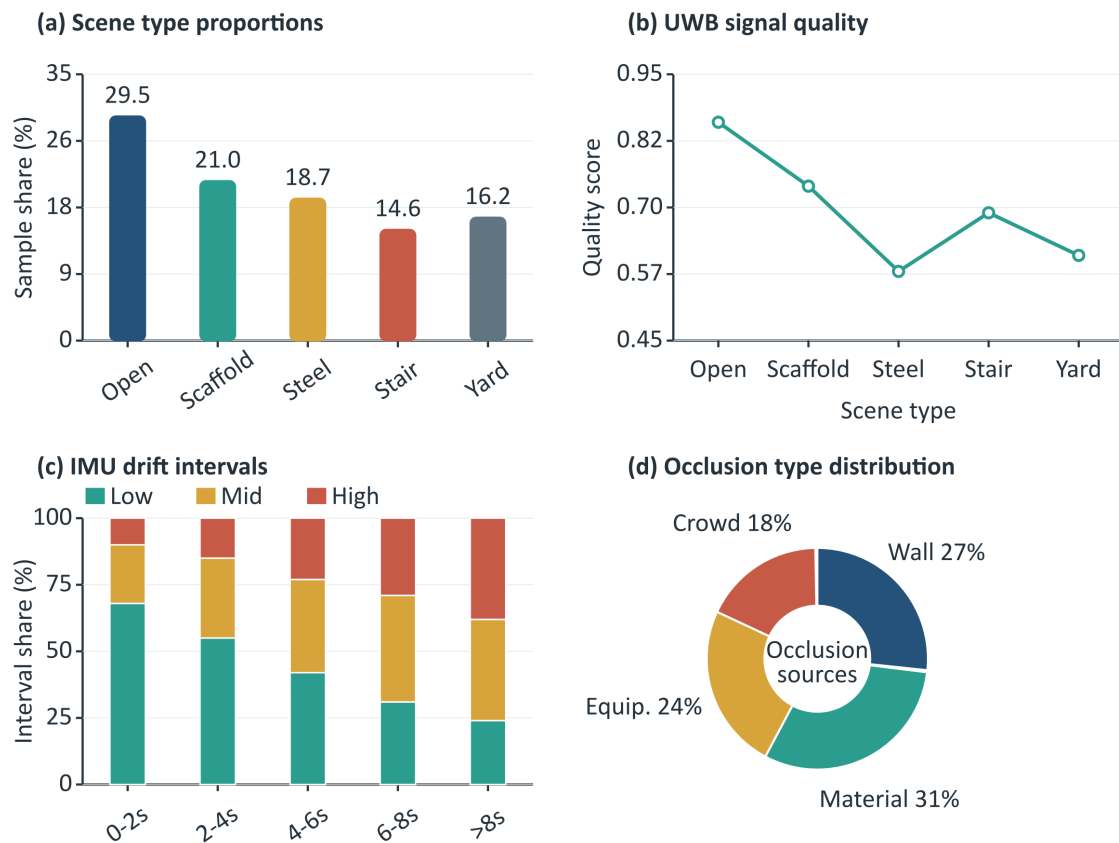


Figure 3. Dataset and positioning error sources. (a) Scene type proportions. (b) UWB signal quality. (c) IMU drift intervals. (d) Occlusion type distribution.

The error distribution demonstrates that while UWB and IMU faults are complimentary, they are not always consistent. When a worker turns suddenly or changes the carrying posture, IMU direction will still fluctuate even when UWB residuals are modest in open terrain. The IMU trajectory is often smooth since scaffold corridor workers follow a set course; nevertheless, UWB signals may bounce off metal pipes. The head and body are briefly out of alignment at the stair entry due to a shift in vertical direction. Even when there has been a slight range error, existing indoor positioning analysis [21] shows that it is impossible to determine whether the trajectory is truly viable without taking signal quality into account. Consequently, anchor visibility, residual magnitude, inertial increments, and zone distance are not fused in a single channel by the created features before fusing. To aid in error debugging, separate them. A significant residual can indeed occur during high-speed travel between two anchors, but if the fused vector is created too early, the model might learn that a large residual always corresponds to an unreliable location. The adjusted path will be smooth but may differ from the real walking path if inertial increments are given too much weight. Compare radio error, inertial tendency, and scene constraint in the same window after keeping the source-specific evidence visible until the contextual block. Therefore, compared to a purely statistical smoothing approach, the latter adjustment is better suited for building safety.

One common finding is that scenes with the highest average coordinate errors do not usually have the most severe warning faults. Even after adjustment, the location along a forklift route may still be inside the tight safety zone due to a somewhat substantial localisation mistake in some material-yard windows. However, because the closest danger zone is far away, a reasonably substantial mistake in an open transfer route might not be immediately apparent. The smart helmet will be impacted by this distinction. Safety-zone sensitivity should be evaluated alongside positioning accuracy, and worker-safety research [22] has demonstrated that context-aware warning outperforms raw location reporting. Thus, mean error, 95th-percentile error, trajectory discontinuity, false alarm rate, and inference latency will also be included in the subsequent trials.

The dataset also includes missing and disturbed observations. Anchor packets are randomly removed in 18.0% of severe occlusion windows and 6.5% of regular windows. Short signal bursts appear near lifts and stacked steel

components, and IMU jitter is introduced into 12.4% of fast-turning windows. These conditions are included because construction monitoring systems often contain partial or asynchronous records. Sensor-fusion research [23] shows that robustness requires a model to use one source to compensate for another without over-relying on either. CoTNet-SE is therefore expected to help because it can suppress unreliable feature channels while learning short-window context.

Comparing sample completeness with the distance to the closest danger threshold reveals another kind of pattern. Just 18.5% of the high-risk windows have been thoroughly cleaned by UWB ranging, although nearly 34.2% of them being within 1.2 meters of a prohibited area. In other words, the riskiest times are sometimes also the most fragile. As a result, the model shouldn't assume that the warning must be postponed until the signal is clear. Although a delayed warning is operationally poor for personnel near machinery or temporary edges, it is technically cautious. The dataset is configured to illustrate this trade-off: if a worker is approaching a tight danger zone, a window with a comparatively high coordinate uncertainty may nevertheless require a warning. As a result, both spatial inaccuracy and warning quality are included in the experiment report. Even while a model that just minimises the mean square error of coordinates would look good on a graph, it still falls short of the smart helmet system's real safety criteria. Both the tail error and the mean error are shown for the same reason. Although they make up a small percentage of the data, a few significant coordinate jumps might cause issues with real-world warning danger. In the simulated data, steel-frame corners, makeshift lifts, and thick material racks account for the poorest 5% of sample records. Additionally, the site manager is more inclined to establish careful or limited sections in these locations. P95 error and false alarm rate are therefore analysed more carefully since they are seen to be significant indications of a model's performance in a chaotic environment as opposed to perfect settings.

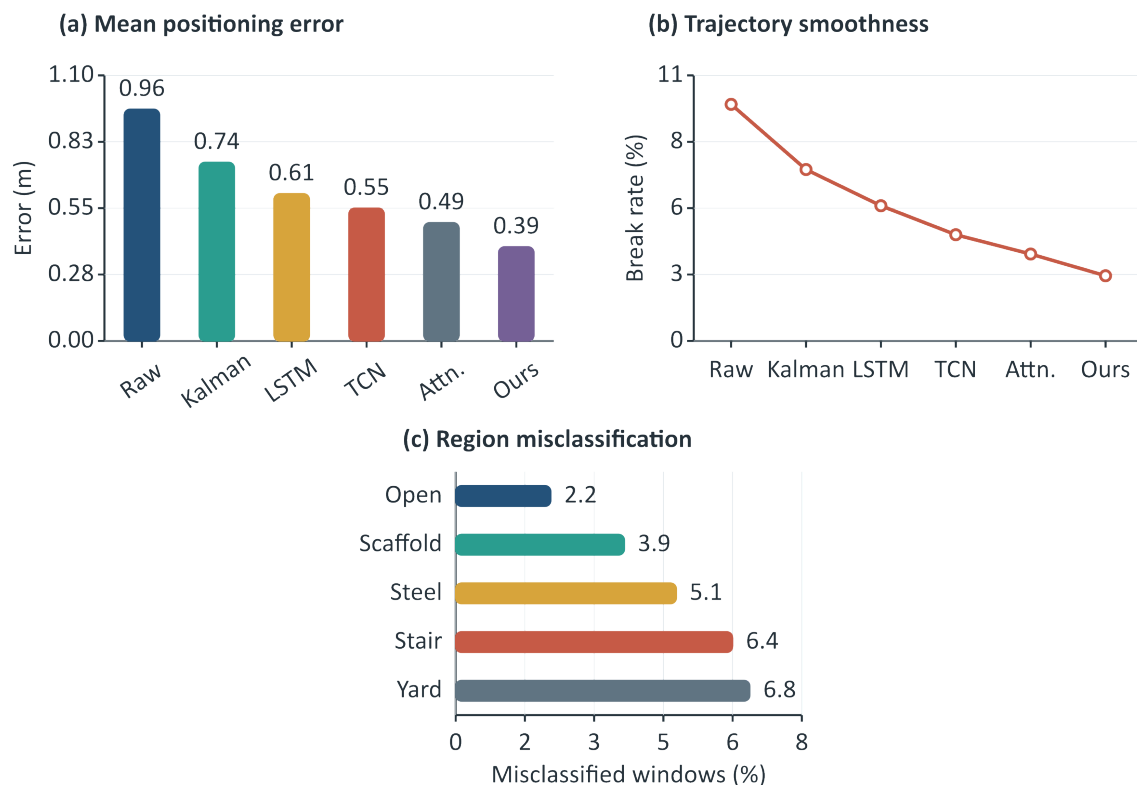


Figure 4. Positioning correction performance. (a) Mean positioning error. (b) Trajectory smoothness. (c) Region misclassification.

Correction Accuracy and Trajectory Stability

The first positioning correction test is shown in Figure 4. Figure 4(a) compares the mean absolute positioning error of raw UWB, Kalman filtering, LSTM fusion, temporal convolution, attention-only fusion, and the proposed model. Raw UWB reaches 0.96 m because NLOS windows remain uncorrected. Kalman filtering reduces jitter but remains weak under prolonged blockage. LSTM fusion and temporal convolution improve sequence use, while the proposed model reaches 0.39 m by combining context modeling with SE recalibration. Figure 4(b)

reports trajectory smoothness, where trajectory break rate falls from 9.8% in raw UWB to 2.7%. Figure 4(c) reports region misclassification, showing that zone-level accuracy can matter more than point-level error near material yards and stairs [24].

Table 1 shows the related numbers. Instead of presenting accuracy, tail error, false alarm, and latency separately, the table is utilised to provide them collectively for integrated understanding. While a model with high accuracy but sluggish inference is unsuitable for helmet-side deployment, a model with a low average error but a significant tail error is nonetheless hazardous for safety management. Although the suggested method may not have the lowest latency, it operates quickly and with decent correction accuracy. This balance is better than a single best-value choice since smart helmet monitoring is an ongoing service. The edge gateway must maintain the responsiveness of the warning interface, integrate zone map updates, and employ a large number of workers. By the time of the alert, the corrected coordinate can already be outdated if the model is operating too slowly. A quick result might nevertheless be incorrect if the model is very simplistic. Consequently, a lightweight contextual model rather than a hefty end-to-end localisation network will be chosen using the table. The trials mentioned above also show that the raw accuracy is insufficient. The identification of a worker entering a limited area may be delayed by a model that smoothes all trajectories to improve the average error. Another model produces a lot of false alarms in the vicinity of reflecting steel buildings, but it may react quickly to all coordinate changes. The suggested approach is beneficial since it simultaneously lowers the quantity of high-risk boundary errors and overall drift.

Table 1. Quantitative comparison of positioning correction models.

Model	MAE (m)	RMSE (m)	P95 Error (m)	Latency (ms)
Raw UWB	0.96	1.24	2.46	1.1
Kalman Filter	0.74	0.98	1.82	2.4
LSTM Fusion	0.56	0.73	1.36	6.8
Temporal Conv.	0.51	0.69	1.28	5.9
Attention Fusion	0.47	0.63	1.17	9.5
CoTNet-SE	0.39	0.52	0.91	8.7

An error curve that demonstrates trajectory continuity is shown in Figure 5. Open-ground and scaffold-corridor approaches are combined in Figure 5(a). Although raw UWB still shows spikes close to the equipment parking area, open-ground movement is typically steady. Even when UWB measurements momentarily diverge, the corrected locations in scaffold corridors should be around the accessible walking route. Stair-entrance and material-yard pathways are combined in Figure 5(b). Larger IMU orientation shifts are caused by stair movement, while cement sacks and piled pipes can obstruct material-yard anchors. As a result, anchor loss, residual surge, walking direction, and scene class should all be taken into account at the same time in the corrected trajectory. Evaluation of temporal continuity under various signal settings is supported by research on deep sequence models [25].

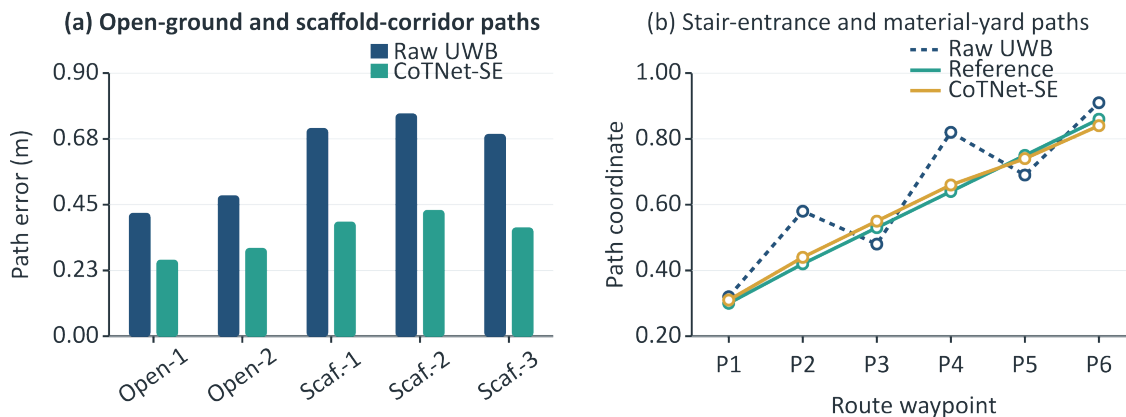


Figure 5. Trajectory continuity under representative construction scenes. (a) Open-ground and scaffold-corridor paths. (b) Stair-entrance and material-yard paths.

The two locations for the remaining mistakes are: The first is a packed, short-distance passage where many helmets are in close proximity to one another and there are simultaneous reflections in the UWB channel. The

second is a sudden shift in posture that occurs when an employee bends to pick up stuff and then rapidly turns. Although there is little real displacement, the IMU sequence in these windows indicates motion. The suggested approach still employs simulated labels and lacks visual verification, so while the inaccuracy has decreased, it has not completely disappeared. The location of a safety choice may be enhanced by taking site geometry and process status into account, according to research on construction digital twins [26]. This is consistent with the present model, and the next stage will be to more closely integrate equipment movement data with map updates. According to the comparison above, the warning module does not need to have great accuracy or low latency. Although attention-only fusion has a high latency and is less reliable when packet loss occurs, it has a respectable error level. Although Kalman filtering is quick, it is unable to recognise recurring disturbance patterns in the material and steel zones. The suggested model has greater trajectory consistency, a reduced false alarm rate, and is within the edge update budget, while not being the quickest method.

Ablation, Robustness, and Safety-Zone Application

The ablation and robustness tests are shown in Figure 6. Figure 6(a) shows that removing the context block increases mean error from 0.39 m to 0.48 m because the model has difficulty distinguishing chronic NLOS drift from brief isolated residuals. Figure 6(b) shows that removing the SE branch leaves unreliable channels active in steel-structure occlusion, increasing false alarms. Figure 6(c) reports UWB packet-loss robustness; the proposed model performs better than LSTM fusion when packet loss rises from 5% to 20% because anchor visibility is modeled as an explicit feature. Figure 6(d) reports IMU jitter robustness. Attention-based fusion research supports adaptive weighting to prevent one disrupted modality from dominating the output [27].

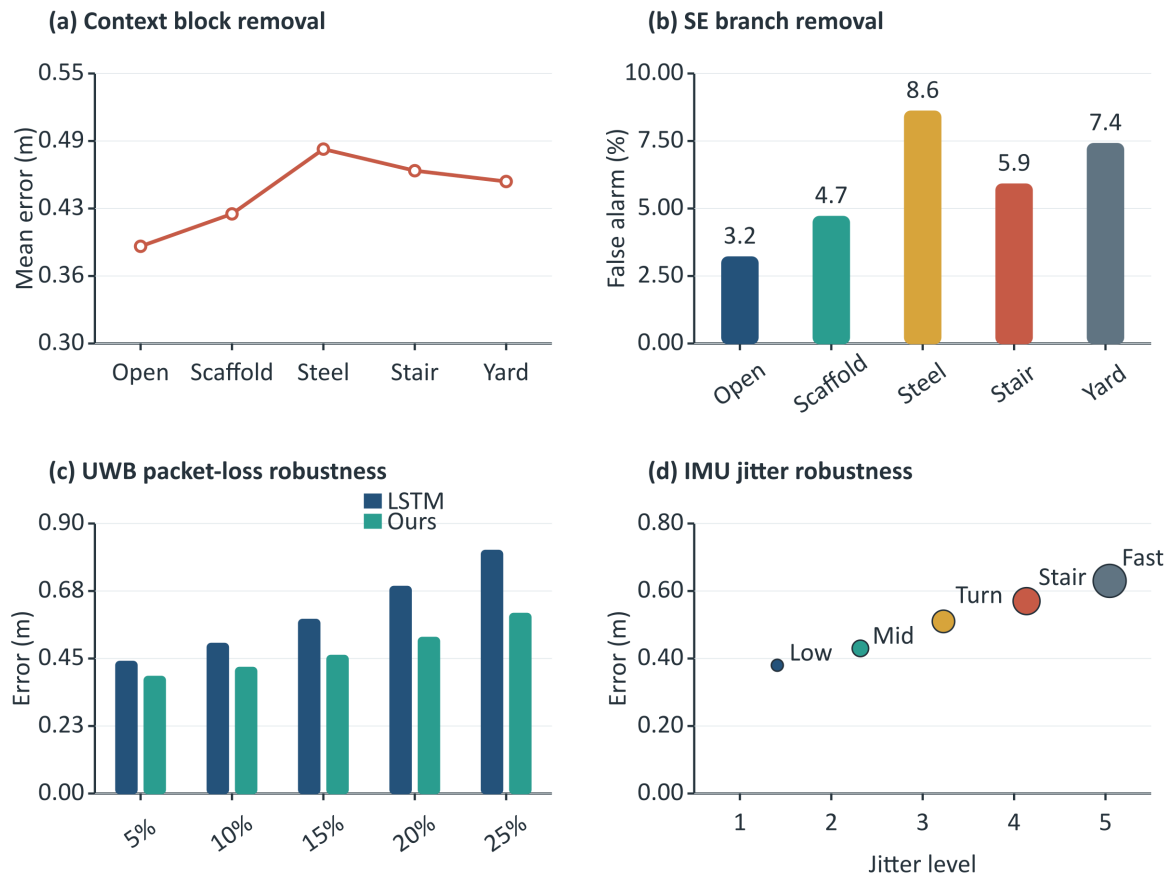


Figure 6. Ablation and robustness analysis. (a) Context block removal. (b) SE branch removal. (c) UWB packet-loss robustness. (d) IMU jitter robustness.

Figure 7 displays the output for safety warnings and correction accuracy. The false alarm rate for equipment-turning, edge-protection, and lifting zones is shown in Figure 7(a). Because fewer corrected points have breached the safety barrier as a result of coordinate jitter, the overall number of false alarms in the new model has decreased by 31.6%. The warning lead time is shown in Figure 7(b). The management can ascertain whether

a worker has entered the restricted area with a mean lead time of 3.6 seconds for equipment contact. Since a late notice will not allow for intervention and an early warning may be disregarded, the lead time should not be prolonged forever. Alarm quality is correlated with timing, confidence, and context, according to safety-management study [28].

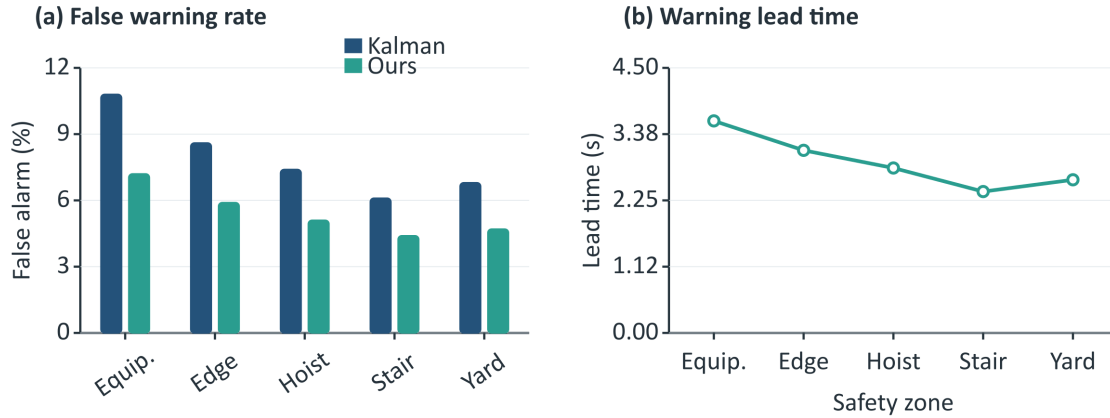


Figure 7. Safety-zone warning performance. (a) False warning rate. (b) Warning lead time.

The model can be applied at the edge, according on runtime outcomes. For a single helmet window, the average inference latency is 17.9 ms on an integrated edge controller and 8.7 ms on a desktop. The system can manage several helmets via batching because the value is smaller than the simulation's 100 ms update budget. There are other deployment conditions than latency. Address the issues of worker privacy, sampling synchronisation, helmet battery status, and anchor calibration. Machine-learning deployment study [29] suggests that operational factors can change the perception of model confidence, and edge AI experiments [30] have demonstrated that a model with excellent offline accuracy may not work well in reality if data synchronisation and maintenance protocols are insufficient. As a result, the suggested approach will include model-confidence output, missing-packet recording, and a signal-quality dashboard. In an actual deployment, the site manager should be able to identify if the alert was caused by a worker in close proximity to a risk region, a low-confidence UWB condition, a quick inertial change, or an inaccurate coordinate. Otherwise, after several false alarms, people won't know what is being warned of and would ignore the alert. Therefore, rather of being a comprehensive automatic decision-maker, the suggested output is to be used as a monitoring tool. For the purpose of reviewing the worker's movement following an incident, it saves the corrected location, confidence value, and warning cause together.

A single coordinate is insufficient; therefore, it is more convenient for managers to see the continuous path of a worker. Corrected Trajectories can support safety-zone inspection, route replay, worker congestion analysis and incident review. Human-centred monitoring research [31] proposes that managers use a decision-support system to process information and make the final decision. Research on explainable AI [32], anomaly detection surveys [33], hierarchical attention research [34], and rule-based explanation research [35] also support providing consistent reasons for automated warnings. Therefore, the model stores the corrected position, confidence level, warning reason and recurrent blind spot in anchor layout. A staged observation mode is used prior to real-time notification to compare the threshold with a manual safety record.

Conclusion

This paper introduces an improved CoTNet-SE model for smart helmet positioning error correction under UWB and inertial sensor sequences. The method integrates UWB range residuals, IMU motion windows, anchor visibility, zone-distance features, contextual transformer convolution, SE channel recalibration, coordinate-offset regression, and a safety-zone warning output. Kalman filtering, LSTM fusion, temporal convolution and attention-only baselines were compared with raw UWB; the model improved trajectory continuity, reduced false safety-zone alarms and lowered the mean positioning error in the simulated construction-site dataset.

The treatment of helmet localisation as a context-aware safety issue is the primary contribution. IMU drift is not solely associated with motion estimation, and UWB faults are not restricted to geometric problems. Both are

related to the workers' locations, anchor visibility, temporary construction facilities, and the worker's distance from a hazardous region. The corrected noisy coordinates will still be interpretable for site managers' alerts if the feature channels are stored separately before adaptive recalibration in the model.

Some Deficiencies Remain. The dataset is simulated, and real construction projects may have more anchor failure types, weather effects, signal reflections, helmet mounting variations, battery degradation, worker behaviour differences, and daily route changes. Future deployment will add long-term field data, site-specific calibration, anchor relocation tests, visual verification, equipment movement integration, map updates and privacy-preserving management rules. The above extensions are required before the system can be used for actual smart construction safety monitoring in a controlled position-correction experiment. Different site layouts, anchor positions, work routes and daily schedules will change the reliability of the learned correction model otherwise.

Author Contributions

Damian Kaczmarek contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Hubert Kasprzyk contributes to analysis, investigation, data collection, draft preparation. Alicja Gruszka contributes to software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized QuillBot for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

References

- [1] Khazen, M., Nik-Bakht, M., & Moselhi, O. (2024). Monitoring workers on indoor construction sites using data fusion of real-time worker's location, body orientation, and productivity state. *Automation in Construction*, 159, 105327. <https://doi.org/10.1016/j.autcon.2024.105327>
- [2] Nkrow, R. E., Silva, B., Boshoff, D., Hancke, G., Gidlund, M., & Abu-Mahfouz, A. (2024). NLOS identification and mitigation for time-based indoor localization systems: Survey and future research directions. *ACM Computing Surveys*, 57(1), Article 4. <https://doi.org/10.1145/3663473>
- [3] Wang, F., Shui, L., Tang, H., & Wei, Z. (2024). Enhancing UWB indoor positioning accuracy through improved snake search algorithm for NLOS/LOS signal classification. *Sensors*, 24(15), 4917. <https://doi.org/10.3390/s24154917>
- [4] Liu, J., Zhang, L., Xu, J., & Shi, J. (2024). Dynamic feasible region-based IMU/UWB fusion method for indoor positioning. *IEEE Sensors Journal*, 24(15), 24895-24904. <https://doi.org/10.1109/jsen.2024.3398789>
- [5] Chen, H., Hou, L., Wu, S., Zhang, G., Zou, Y., Moon, S., & Bhuiyan, M. (2024). Augmented reality, deep learning and vision-language query system for construction worker safety. *Automation in Construction*, 158, 105158. <https://doi.org/10.1016/j.autcon.2023.105158>
- [6] Zhu, H., & Hwang, B.-G. (2025). Real-time safety and worker self-assessment: Sensor-based mobile system for critical unsafe behaviors. *Automation in Construction*, 172, 105879. <https://doi.org/10.1016/j.autcon.2024.105879>
- [7] Wang, H., He, J., & Cui, L. (2025). An advanced indoor localization method based on xLSTM and residual multimodal fusion of UWB/IMU data. *Electronics*, 14(13), 2730. <https://doi.org/10.3390/electronics14132730>

- [8] Muthineni, K., Artemenko, A., Vidal, J., Nájjar, M., Catalan, M., & Paradells, J. (2025). Deep learning-based UWB-IMU data fusion for indoor positioning in industrial scenario. *IEEE Open Journal of Vehicular Technology*, 6, 892-905. <https://doi.org/10.1109/ojvt.2025.3566888>
- [9] Zhang, Y., Zhan, Q., & Ma, Z. (2024). EfficientNet-ECA: A lightweight network based on efficient channel attention for class-imbalanced welding defects classification. *Advanced Engineering Informatics*, 62, 102737. <https://doi.org/10.1016/j.aei.2024.102737>
- [10] Zhang, M., Ma, S., Xu, R., Chen, T., Ding, Y., & Luo, X. (2025). Evaluating the impact of proactive warning systems on worker safety performance: An immersive virtual reality study. *Safety Science*, 183, 106774. <https://doi.org/10.1016/j.ssci.2024.106774>
- [11] Lin, Y., Yu, K., Zhu, F., Bu, J., & Dua, X. (2024). The state of the art of deep learning-based Wi-Fi indoor positioning: A review. *IEEE Sensors Journal*, 24(19), 30013-30032. <https://doi.org/10.1109/jsen.2024.3432154>
- [12] Altheimer, J., & Schneider, J. (2024). Smart-watch-based construction worker activity recognition with hand-held power tools. *Automation in Construction*, 166, 105684. <https://doi.org/10.1016/j.autcon.2024.105684>
- [13] Choi, J., Lee, B., Son, S., Choi, J., & Oh, K.-Y. (2026). High-fidelity virtual modeling framework for outdoor UWB localization via real-world measurements. *International Journal of Precision Engineering and Manufacturing-Smart Technology*, 4, 1-12. <https://doi.org/10.57062/ijpem-st.2026.00122>
- [14] Han, C., Xue, S., Long, L., & Xiao, X. (2024). Research on inertial navigation and environmental correction indoor ultra-wideband ranging and positioning methods. *Sensors*, 24(1), 261. <https://doi.org/10.3390/s24010261>
- [15] He, Z., Tang, X., Li, M., & Zhang, F. (2026). Enhanced indoor mobile robot localization via Lie-group IMU-UWB fusion and dual-stage Kalman filtering. *Sensors*, 26(9), 2686. <https://doi.org/10.3390/s26092686>
- [16] Lin, X., Guo, Z., Jin, X., & Guo, H. (2025). Digital twin-enabled safety monitoring system for seamless worker-robot collaboration in construction. *Automation in Construction*, 175, 106147. <https://doi.org/10.1016/j.autcon.2025.106147>
- [17] Huang, H., Lu, Y., Zhou, S., Zhang, X., & Li, Z. (2024). CoTNeT: Contextual transformer network for encrypted traffic classification. *Egyptian Informatics Journal*, 27, 100475. <https://doi.org/10.1016/j.eij.2024.100475>
- [18] Ocal, H. (2024). scSEETV-Net: Spatial and channel squeeze-excitation and edge attention guidance V-shaped network for skin lesion segmentation. *Advanced Intelligent Systems*, 7(1), 2400438. <https://doi.org/10.1002/aisy.202400438>
- [19] Yu, X., & Al-qaness, M. A. A. (2025). ASK-HAR: Attention-based multi-core selective kernel convolution network for human activity recognition. *Measurement*, 239, 115981. <https://doi.org/10.1016/j.measurement.2024.115981>
- [20] Lu, J., Huang, T., Zhang, Q., Chen, X., & Zhou, J. (2024). A lightweight vehicle detection network fusing feature pyramid and channel attention. *Internet of Things*, 27, 101166. <https://doi.org/10.1016/j.iot.2024.101166>
- [21] Jiang, S., Zhang, R., Xue, J., & Gao, H. (2026). NLOS identification and error mitigation algorithms for indoor positioning: A review. *Complex & Intelligent Systems*, 12, 124. <https://doi.org/10.1007/s40747-026-02335-y>
- [22] Pan, Z., & Yu, Y. (2024). Learning multi-granular worker intentions from incomplete visual observations for worker-robot collaboration in construction. *Automation in Construction*, 158, 105184. <https://doi.org/10.1016/j.autcon.2023.105184>
- [23] Liu, J., Kong, W., & Shrestha, A. (2024). Development of UWB NLOS model aided by IMU in indoor positioning systems. *2024 IEEE International Conference on Future Machine Learning and Data Science*, 190-195. <https://doi.org/10.1109/fmlds63805.2024.00037>
- [24] Gufran, D., & Pasricha, S. (2025). Towards explainable indoor localization: Interpreting neural network learning on Wi-Fi fingerprints using logic gates. *2025 International Conference on Indoor Positioning and Indoor Navigation*. <https://doi.org/10.1109/ipin66788.2025.11213392>
- [25] Zhou, Z., Xu, Z., & Xia, J. (2025). Deep learning optimization positioning algorithm based on UWB/IMU fusion in complex indoor environments. *Physical Communication*, 72, 102702. <https://doi.org/10.1016/j.phycom.2025.102702>
- [26] Saif, W., RazaviAlavi, S., & Kassem, M. (2024). Construction digital twin: A taxonomy and analysis of the application-technology-data triad. *Automation in Construction*, 165, 105715. <https://doi.org/10.1016/j.autcon.2024.105715>

- [27] Wang, Y., Bai, W., Hu, F., Tang, J., & Cheng, Z. (2024). Adaptive multi-sensor fusion localization method based on filtering. *Mathematics*, 12(14), 2225. <https://doi.org/10.3390/math12142225>
- [28] Lu, D. B., & Ergan, S. (2025). Principal attributes of wearable warning alarms to promote roadway worker safety. *Advanced Engineering Informatics*, 67, 103481. <https://doi.org/10.1016/j.aei.2025.103481>
- [29] Mittal, P. (2024). A comprehensive survey of deep learning-based lightweight object detection models for edge devices. *Artificial Intelligence Review*, 57, 242. <https://doi.org/10.1007/s10462-024-10877-1>
- [30] Zalfani, M., Süle, E., & Bakar, M. (2026). Human-centered AI for decision support systems: Enhancing usability and trustworthiness. *Systems*, 14(6), 651. <https://doi.org/10.3390/systems14060651>
- [31] Coussement, K., Abedin, M. Z., Kraus, M., Maldonado, S., & Topuz, K. (2024). Explainable AI for enhanced decision-making. *Decision Support Systems*, 184, 114276. <https://doi.org/10.1016/j.dss.2024.114276>
- [32] Baniecki, H., & Biecek, P. (2024). Adversarial attacks and defenses in explainable artificial intelligence: A survey. *Information Fusion*, 107, 102303. <https://doi.org/10.1016/j.inffus.2024.102303>
- [33] Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C., & Salehi, M. (2024). Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys*, 57(1), Article 10. <https://doi.org/10.1145/3691338>
- [34] Wei, R., & Wang, J. (2024). TCN-attention-HAR: Human activity recognition based on attention mechanism time convolutional network. *Scientific Reports*, 14, 6964. <https://doi.org/10.1038/s41598-024-57912-3>
- [35] Li, Y., Chan, J., Peko, G., & Sundaram, D. (2024). An explanation framework and method for AI-based text emotion analysis and visualisation. *Decision Support Systems*, 180, 114121. <https://doi.org/10.1016/j.dss.2023.114121>